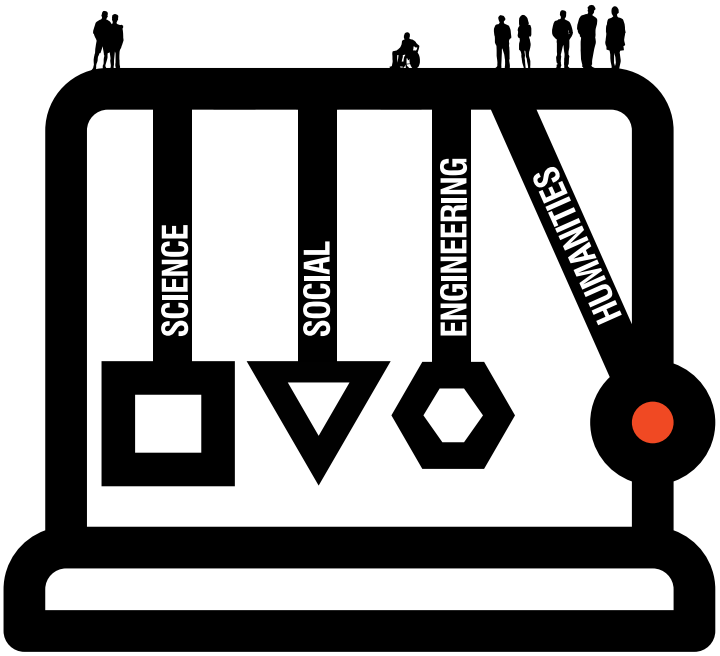
El profesional de la información
The information professional
INTERNATIONAL JOURNAL ON INFORMATION AND COMMUNICATION

eISSN 1699-2407
CODEN: PINFF2

EPI

ISSN: 1699-2407
2019, v. 28, n. 1



NÚMERO MULTIDISCIPLINAR DE INFORMACIÓN Y COMUNICACIÓN

El profesional de la información

Revista bimestral fundada en 1992 por
Tomàs Baiget y Francisca García-Sicilia

El profesional de la información es una revista de la editorial Ediciones Profesionales de la Información S.L.

NIF: B63664544

Apartado 13 - 24080 León, España

Tel.: +34 - 608 491 521

http://www.elprofesionaldelainformacion.com

Redacción

El profesional de la información

Apartado 13

24080 León, España

Tel.: +34 - 934 250 029

epi@elprofesionaldelainformacion.com

Publicidad

Tel.: +34 - 608 491 521

publici@elprofesionaldelainformacion.com

Suscripciones

El profesional de la información

Apartado 13

24080 León, España

suscripciones@elprofesionaldelainformacion.com

http://www.elprofesionaldelainformacion.com/suscripciones.html

Servicios online

Roxana Dinu: *roxadinu@gmail.com*

Diseño

MASmedios, *http://www.masmedios.com*

Director artístico: Moisés Mañas

Maquetación

Isabel Olea
isabel.iolea@gmail.com

Distribución online

Recyt. Fundación Española para la Ciencia y la Tecnología

http://recyt.fecyt.es/index.php/EPI

Depósito legal: B. 12.303-1997

Los trabajos publicados en EPI son aprobados según el sistema tradicional "peer review" en doble ciego: son revisados al menos por dos expertos en el tema, del Consejo Asesor de la revista y/o externos.

Para conseguir que los trabajos no pierdan actualidad, la dirección y los evaluadores de esta revista ponen especial esfuerzo en revisar los artículos con gran rapidez, consiguiendo un tiempo medio de aceptación o rechazo de los trabajos de sólo unas pocas semanas.

La revista *El profesional de la información* está referenciada en:

BASES DE DATOS

Academic search (Ebsco)
https://www.ebsco.com/products/research-databases/academic-search

Business source (Ebsco)
https://www.ebsco.com/products/research-databases/business-source-complete

Communication source (Ebsco)
https://www.ebsco.com/products/research-databases/communication-source

Dialnet (Universidad de La Rioja)
http://dialnet.unirioja.es/servlet/revista?&clave_revista=469

Social sciences citation index, Social SCI, WoS (Clarivate Analytics)
http://wokinfo.com/products_tools/multidisciplinary/webofscience/ssci

Journal Citations Reports (JCR) (Clarivate Analytics)
https://clarivate.com/products/journal-citation-reports
Impact Factor 2017 = 1,318

Inspec, Information services in physics, electronics and computing (IET, The Institution of Engineering and Technology)
http://www.theiet.org/publishing/inspec

ÍnDICEs CSIC
https://indices.app.csic.es

ISTA, Information science and technology abstracts (Ebsco)
https://www.ebsco.com/products/research-databases/information-science-technology-abstracts

Lisa, Library and information science abstracts (CSA)
http://www.csa.com/factsheets/lisa-set-c.php

Lista, Library, information science & technology abstracts (Ebsco)
http://www.libraryresearch.com

Scopus (Elsevier) Scimago Journal Rank 2017 = 0,652
http://www.scopus.com

ÍNDICES

Catálogo de Latindex
http://www.latindex.unam.mx/latindex/ficha?folio=23967

ERIH Plus
https://dbh.nsd.uib.no/publiseringskanaler/erihplus/periodical/info.action?id=449601

ACCESO A LOS TEXTOS COMPLETOS

Recyt (2000-)
http://recyt.fecyt.es/index.php/EPI

Digitalia Publishing (2011-)
http://www.digitaliapublishing.com

Ebscohost Electronic Journals Service (2000-)
http://ejournals.ebsco.com/direct.asp?JournalID=105302

e-Libro (2011-)
http://www.e-libro.net

Informe Académico (2000-)
https://www.cengage.com

Library, information science & technology abstracts with full text (2000-)
http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=584

Academic search complete (2000-)
https://www.ebsco.com/products/research-databases/academic-search

Business source complete (2000-)
https://www.ebsco.com/products/research-databases/business-source-complete

Communication source (2000-)
https://www.ebsco.com/products/research-databases/communication-source

El profesional de la información (1992-embargo 3 años)
http://elprofesionaldelainformacion.com/contenidos.html

EPI EN REDES SOCIALES

http://www.facebook.com/elprofesionaldelainformacion
http://twitter.com/revista_EPI
http://www.linkedin.com/company/el-profesional-de-la-informacion-epi-
http://pinterest.com/source/elprofesionaldelainformacion.com

PLATAFORMA DE PRODUCCIÓN OJS

Recyt, Repositorio español de ciencia y tecnología (Fecyt)
http://recyt.fecyt.es/index.php/EPI

Sumario

EPI, 2019, v. 28, n. 1

Número multidisciplinar de Información y Comunicación

OBSERVATORIO

REVIEW ARTICLES / ARTÍCULOS DE REVISIÓN

RESEARCH ARTICLES / ARTÍCULOS DE INVESTIGACIÓN

ANALYSIS / ANÁLISIS

OBSERVATORIO

Muchos cambios y algunas certezas para las bibliotecas de investigación, especializadas y centros de documentación

Many changes and some certainties for research and special libraries, and documentation centers

Lluís M. Anglada

Cómo citar este artículo:

Anglada, Lluís M. (2019). "Muchos cambios y algunas certezas para las bibliotecas de investigación, especializadas y centros de documentación". *El profesional de la información*, v. 28, n. 1, e280113.

https://doi.org/10.3145/epi.2019.ene.13

Artículo recibido el 05-02-2019
Aceptación definitiva: 12-02-2019

Lluís M. Anglada ✉
http://orcid.org/0000-0002-6384-4927

Consorci de Serveis Universitaris de Catalunya (*CSUC*).
Àrea de Ciència Oberta
Gran Capità, 2-4, edif. Nexus, 4ª pl. 08034 Barcelona,
España
langlada@gmail.com

Resumen

El artículo argumenta que a pesar de los muchos cambios que están experimentando las bibliotecas de investigación, especializadas y los centros de documentación, que les ocasionan dudas sobre su futuro, existen algunas certezas que permiten que estos centros puedan dirigir adecuadamente sus actividades. Se propone distinguir entre cambios menores y cambios profundos y se afirma que, para los centros de información, el cambio fundamental es el haber pasado de una situación en la que la información era escasa y el tiempo para buscarla abundante a otro en el que el tiempo para buscar información es escaso y la información abundante. Esto debería reconducir las actividades de los centros de información para –siguiendo a M. Buckland- dejar de centrarse en la 'información-objeto' para hacerlo en la 'información-proceso'. Se fijan como certezas donde apoyarse diferentes aspectos de dos movimientos: el de la Ciencia Abierta y el de las Humanidades Digitales. Se examinan algunos productos documentales que reflejan un cambio en las maneras de enfocar y usar la información por parte de las personas: *La Colección* del *Museo del Prado*, *Transcribe Bentham*, *eBird* y *Mapa Literari Català*. Para finalizar, se recomienda a los centros de documentación, bibliotecas especializados y de investigación que orienten sus servicios a digitalizar de forma enriquecida, a vehicular la participación ciudadana, y a crear y reforzar comunidades.

Palabras clave

Bibliotecas especializadas; Centros de documentación; Gestión del cambio; Evolución; Futuro; Ciencia abierta; Humanidades digitales; Nuevos enfoques; Actividades; Sociedad de la información; Participación ciudadana.

Abstract

It is argued that despite the many changes being experienced by research and specialized libraries, and documentation centers, which cause them doubts about their future, there are some certainties that allow these centers to properly address their activities. It is proposed to distinguish between minor and profound changes and it is affirmed that, for information centers, the fundamental change has been to go from a situation in which information was scarce and the time to look for it abundant, to another in which time to look for information is scarce and information is abundant. This should help redirect the activities of the information centers -following M. Buckland- to stop focusing on the 'information-as-an-object' to concentrate on the 'information-as-a-process'. Two consolidated movements should help them to guide their future: the Open Science and the Digital Humanities. Some information products that reflect a change in the ways people approach and use information are examined: *Museo del Prado*'s *The Collection*, *Transcribe Bentham*, *eBird* and *Mapa Literari Català*. Finally, the author recommends to documentation centers, and specialized and research libraries to guide their services to digitize their heritage collections in an enriched manner, to promote citizen participation, and to create and strengthen communities.

Keywords

Special libraries; Documentation centers; Change management; Evolution; Future; Open science; Digital humanities; New approaches; Activities; Information society; Citizen participation; Croudsourcing.

1. Cambios y más cambios[1]

Vivimos una época de cambios constantes y acelerados, y esto crea en nuestra subjetividad la sensación de que cualquier cambio es permanente y bueno y que, en los tiempos que corren, gestionar es adaptarse al cambio. Pero, por mucho que el cambio caracterice la época en la que nos ha tocado vivir, no todos los cambios son iguales; los cambios pueden clasificarse por sus efectos, y así tendríamos cambios aparentes, circunstanciales, pasajeros, temporales, definitivos, profundos, fundamentales… Además, uno tiene la impresión de que la adaptación al cambio está consistiendo en dejarlo pasar a la espera del siguiente, cuando lo propio sería adoptarlo por lo que esto supone de sintonizar con el momento.

La gestión efectiva no es la adaptación inmediata a cualquier cambio, sino la diferenciación entre los muchos cambios que nos sacuden para discriminar entre modas y tendencias, y para seleccionar entre estas últimas las que debemos adoptar (ver, por ejemplo, el uso del análisis estratégico para direccionar el futuro de las bibliotecas en: **Wiorogórska**, 2016). Así, una respuesta cortoplacista de adaptación al cambio es debatir a través de qué canal comunicaremos (ayer *Second Life*, después *Facebook*, hoy quizá mejor *Twitter*, pero mañana... quién sabe), pero lo fundamental es ver, por debajo del cambio novedoso (y constante) una tendencia de profundidad que nos aconseja adoptar las redes sociales como vehículo para difundir las actividades y servicios de nuestro centro y, sobre todo, para establecer vínculos con nuestros usuarios o 'clientes'.

> La gestión efectiva no es la adaptación inmediata a cualquier cambio, sino la diferenciación entre lo que simplemente son modas y lo que son tendencias

El problema para categorizar desapasionadamente los elementos nuevos que aparecen en el escenario profesional es su abundancia. A esto se añade el hecho de que hoy cualquier cambio tecnológico va acompañado de una valoración positiva. Muchos cambios en un marco mental de valoración alta conllevan a que no sepamos o queramos distinguir entre lo que tendrá consecuencias (sólo) a corto plazo y lo que afectará (de forma lenta pero inexorable) a nuestro futuro a partir de ahora. Miremos 40 años atrás y tomemos el CD-ROM como ejemplo. Con su aparición se inició un ciclo de cambios constantes con respecto al almacenamiento de información y fue tarea profesional estudiar las mejoras que trajeron los instrumentos que terminaron sustituyéndolo, pero al final estos aspectos no han sido más que momentos menores que ejemplifican la tendencia que importa: cada vez hay más información accesible con más facilidad desde cualquier lugar y momento (y por lo tanto, cada vez es menor el valor de la biblioteca dispensadora de información). Siguiendo a Dempsey:

> "...los cambios en los comportamientos en la investigación y el aprendizaje y en las expectativas son más importantes para la biblioteca académica que cualquier otro cambio en la tecnología bibliotecaria en sí mismo (**Dempsey**, 2012)"

Dejemos de lado pues los cambios que podríamos denominar 'tácticos', por ser sus efectos pequeños o grandes pero poco sostenidos en el tiempo, y pasemos a fijarnos en los que podríamos denominar 'profundos', por ser sus efectos continuados a lo largo de un cierto período. Hay más de los que podemos analizar en detalle: la accesibilidad de grandes cantidades de información digital, la ubicuidad en el acceso, la tendencia de la información a ser libre, o el hecho de que la facilidad en el acceso esté sustituyendo a la calidad como criterio de selección de la información…

Pero al final, las demandas generadas por nuevos artilugios, lo que nos parecen gustos cambiantes del público, sus aproximaciones variables al consumo de información, dependen en parte de los medios a través de los que se accede a la

información, pero sobre todo de su posición en el contexto vital de las necesidades humanas. El cambio fundamental y director es que hemos pasado de un contexto en el que la información era escasa y el tiempo para buscarla abundante, a otro en el que el tiempo para buscar información es escaso y la información abundante. En un contexto de *infoabundancia* para los centros de información, lo importante no es tener sino enseñar, y, para el consumidor, lo determinante es menos obtener información que incorporarla a la vida cotidiana a través sus procesos habituales y el vivirla como experiencia (**Anglada**, 2016).

Podemos entender mejor este cambio recordando los distintos significados que tiene la palabra 'información' en el artículo seminal de **Buckland** (1991). Éste distingue entre la información como objeto (*information-as-thing*), la información como proceso (*information-as-process*) y la información como conocimiento (*information-as-knowledge*). A pesar de que la 'información' que nos interesa es el conocimiento, no hay conocimiento sin información procesada, ni proceso sin información que deglutir.

> En el cambio de paradigma que estamos viviendo, la información pasa a ser abundante y se *comodiza*

Casi toda la biblioteconomía hecha hasta ahora se ha centrado en la información objeto. Lógico, en un contexto de información escasa, lo primero es recoger, y luego –claro- organizar. El documento como objeto (cómo lo guardamos, cómo lo clasificamos, cómo lo hacemos accesible…) ha construido nuestros conocimientos profesionales y ha sido totalmente normal que la biblioteconomía moderna centrara su atención en maximizar los usos (presentes y futuros) de los documentos en la información como objeto. Cuando el primer y principal obstáculo para acceder a la información es conocer de su existencia y acceder físicamente a ella, facilitar el proceso de estar informado no es una prioridad.

Pero en el cambio de paradigma que estamos viviendo, la información pasa a ser abundante y se *comodiza*. La dedicación a guardar, indexar y difundir documentos pasa de las personas a las máquinas y la biblioteconomía debe dirigir su atención hacia cómo facilitar que la información se incorpore a los flujos de trabajo y procesos cotidianos de investigadores y ciudadanos. El uso de información dependerá menos de su tratamiento físico que del trabajo que puedan hacer los centros de información para adaptar sus técnicas a los procesos intelectuales tal como nos vienen modulados por la época que vivimos.

Así pues, si ha lugar redefinir la biblioteca para el futuro –cosa que ciertamente tiene sentido hacer-, no ha de ser a partir de un nuevo marco tecnológico sino a partir de un nuevo marco sociológico –e incluso me atrevería a decir antropológico-. Olvidemos los soportes documentales y preguntémonos qué necesidades cubría y qué necesidades debe cubrir la biblioteca o centro de documentación y qué 'lenguaje' debe emplear para hablar con sus usuarios. En una fase de estabilidad social, el evolucionismo tecnológico y el inmovilismo organizativo son la receta; en fases de cambios sociales profundos, debemos hacer reajustes en la manera en que hacemos las cosas para garantizar el poder seguir haciéndolas con la finalidad para la que nacimos.

Archiveros, bibliotecarios, documentalistas e incluso museólogos tenemos el objetivo profesional común de procesar información para que ésta pueda ser encontrada y utilizada con facilidad por el ciudadano y el investigador. Para ello nos hemos valido en el pasado, y lo estamos haciendo en el presente, de diferentes avances tecnológicos (muchos de ellos creados fuera de nuestro ámbito profesional) que aplicamos a nuestras actividades. En las últimas décadas los avances están siendo de tal magnitud que a veces monopolizan nuestra atención y alejan los debates profesionales del objeto mismo de la profesión.

El *tecnosolucionismo* (**Morozov**, 2015) no es una enfermedad exclusiva de los *infonomistas*, (término acuñado por Alfons Cornella, que quizá es el que mejor describe esta "multiprofesión", ver **Cobarsí**; **Canals**; **Ortoll**, 2016), pero nosotros no hemos desarrollado poderosas defensas ni tan siquiera prevenciones contra la misma. Por experiencia propia vemos como la tecnología soluciona hoy problemas que eran insolubles hace pocos años y quedamos admirados hasta la fascinación en el 'cómo se hace', dejando algo de lado el 'por qué lo queríamos hacer'. Mientras que el 'cómo lo hacemos' subdivide al sector en múltiples subdisciplinas profesionales, el 'por qué lo hacemos' afecta por igual a todas ellas. Hay tendencias que creo que afectan y afectarán a todas las organizaciones gestoras de información, aunque en estos momentos puedan parecer alejadas de las preocupaciones principales de la subespecialidad profesional en la que cada uno anda metido. La adaptación organizacional de archivos, bibliotecas y centros de documentación a las necesidades de información de la cambiante sociedad que los ha creado y mantiene debería ocupar un lugar bastante central en nuestros debates y preocupaciones.

2. Algunas certezas

Planteado así, los cambios en los gustos o preferencias de los usuarios de información deberíamos buscarlos en movimientos más amplios, en tendencias que están reconfigurando la forma de trabajar y los centros de interés de los investigadores y también de los ciudadanos de a pie. Detengámonos un momento a describir brevemente las motivaciones y las consecuencias de la Ciencia Abierta y de las Humanidades Digitales. Las dos tienen efectos importantes para las bibliotecas de investigación, especializadas y los centros de documentación, si es que queremos que éstos 'hablen' el lenguaje de la época.

2.1. Ciencia Abierta

El término Ciencia Abierta (**Anglada**; **Abadal**, 2018) tiene una gestación larga e interesante. En marzo de 2013, la *Dirección General de Communications, Networks, Content and Technology* (*Connect*) de la *Comisión Europea*, publicó un documento conceptual sobre el momento actual de la ciencia. En él se exponían de forma embrionaria los conceptos que la *Comisión* ya no ha dejado de usar en documentos posteriores, a saber, que internet y otras tecnologías han cambiado radicalmente la forma en que se crea y difunde el nuevo conocimiento, que este cambio necesita sustentarse en un entorno científico colaborativo y abierto, y que la nueva manera de hacer ciencia permite que ésta se acerque a la sociedad.

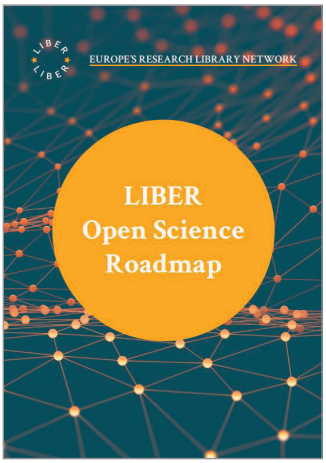
https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe

Los motivos de la *Comisión Europea* para preocuparse por los cambios en la manera de hacer ciencia son político-sociales y quedan bien explicados con unas palabras de Jean-Claude Juncker, su presidente:

> "La investigación y la innovación crean oportunidades de inversión para nuevos y mejores productos y servicios y, por tanto, incrementan la competitividad y el empleo" (*European Commission*, 2016).

Pero el movimiento de la Ciencia Abierta tiene también motivaciones científicas ya que quiere adecuar la comunicación científica a las posibilidades tecnológicas actuales y aumentar su eficacia haciendo que la ciencia sea más rápida, certera y reutilizable. Lo que el concepto de Ciencia Abierta pone de manifiesto es que, debido a los cambios tecnológicos recientes, la ciencia, sin haber modificado sus motivaciones y objetivos, sí lo ha hecho con su manera de operar. Esta nueva manera de hacer ciencia se fundamenta en tres criterios: la ciencia debe ser:

- abierta,
- colaborativa, y
- hecha con y para la sociedad.



https://zenodo.org/record/1303002

La llamada europea a la Ciencia Abierta ha tenido un amplio efecto en diferentes ámbitos. Por ejemplo, algunos países han creado sus propios planes nacionales para fomentarla. Así lo han hecho Finlandia (2014), Eslovenia (2015), Portugal (2016), Holanda (2017), Francia (2018) y Serbia (2018) (**Lasauca-Cisa**, 2018). Pero ha tenido también efecto en entidades científicas como la *European University Association* o la *Asociación Europea de Bibliotecas de Investigación* (*Liber*). Vale la pena fijarse en el contenido del *Liber open science roadmap* (*Liber*, 2018) ya que en él no sólo se describen las siete áreas focales en las que podemos dividir la Ciencia Abierta, sino que en cada una de ellas se señalan las oportunidades y retos para las bibliotecas, a las que se les dan también recomendaciones para que se preparen para los efectos que tendrá este desarrollo.

2.2. Humanidades Digitales

El otro gran movimiento que afecta la vida y la evolución de los centros de información es el denominado Humanidades Digitales. Como cualquier conceptualización de las actividades humanas, el concepto de Humanidades Digitales es controvertido. A nuestros efectos, nos quedaremos con la definición de humanidades que da el *Diccionario de la RAE*[2]:

> "conjunto de disciplinas que giran en torno al ser humano, como la literatura, la filosofía o la historia".

No hay una definición canónica de Humanidades Digitales[3], pero nos valdrá decir que no es una disciplina propiamente dicha, sino una manera de trabajar, en este caso la aplicación de métodos y herramientas computacionales a problemas de investigación en humanidades[4]. Los métodos digitales o computacionales comparten dos características:

- a través de ordenadores tratan grandes conjuntos de datos (*big data*) imposibles de gestionar de forma manual, a partir de los cuales,
- construyen modelos o patrones de comportamiento, establecen correlaciones, crean simulaciones, construyen representaciones gráficas...

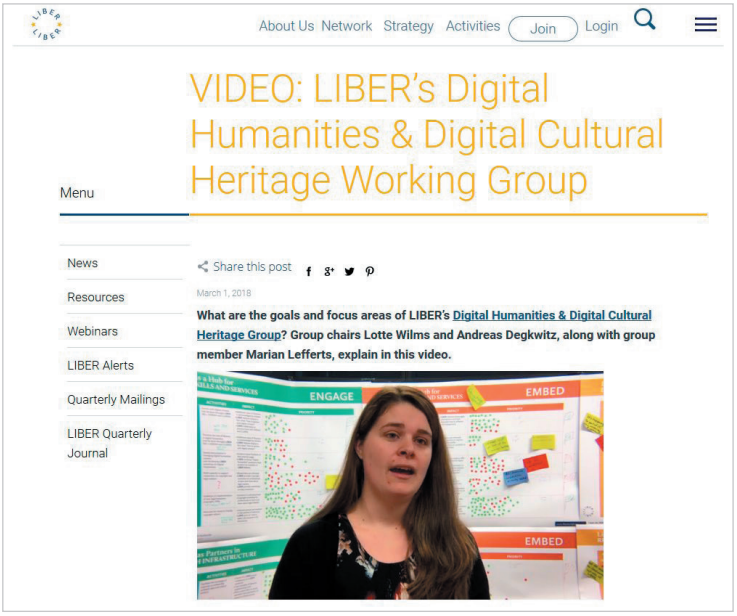
Las Humanidades Digitales necesitan pues objetos digitales (o digitalizados) para que éstos puedan ser tratados con instrumentos computacionales. Pero los ordenadores no pueden tratar datos si previamente no han sido documentados. Los registros catalográficos en formato MARC[5] podían ser 'leídos' por un ordenador porqué los elementos bibliográficos estaban etiquetados. De forma parecida, las máquinas no pueden analizar los datos ni mostrar patrones a no ser que los datos se presenten en formatos conocidos, de forma normalizada, con la información suficiente que permita su preservación y reutilización.

Vale la pena destacar también que muchas de las actividades que en Europa consideramos de Ciencia Abierta o de Humanidades Digitales en Norte América se engloban bajo el concepto '*Digital scholarship*'. Ciertamente, la afirma-

ción anterior requiere muchas explicaciones y matices, ya que las diferencias se encuentran más en los orígenes de los conceptos respectivos que en las prácticas documentales que los mismos incluyen.

2.3. Consecuencias para las bibliotecas

Los movimientos de la Ciencia Abierta y de las Humanidades Digitales son un reflejo de los profundos cambios producidos en los procesos de la investigación, cambios que a su vez producen o deberían producir otros en la orientación de las bibliotecas especializadas y centros de documentación. Destaquemos los aspectos de cada movimiento con más influencia en nuestro ejercicio profesional.



https://libereurope.eu/blog/2018/03/01/video-libers-digital-humanities-digital-cultural-heritage-working-group

Hemos visto que uno de los postulados base de la Ciencia Abierta es que la investigación debe ser hecha con y para la sociedad. Podemos parafrasear a Clemenceau y decir que la ciencia es un asunto demasiado serio como para dejarla en manos de los científicos. Si la investigación pasa a ser el motor reconocido de desarrollo económico y social, ésta debe sustentarse en los valores éticos y sociales, dirigirse a dar respuesta a los retos de la sociedad del siglo XXI e implicar a los ciudadanos.

No se trata sólo de hacer más divulgación científica. La ciencia abierta observa que, en el contexto tecnológico actual, es posible establecer una nueva relación entre las personas y la ciencia, una relación de participación. Los ciudadanos no sólo pueden interesarse más por lo que hacen los científicos, sino que pueden contribuir a la ciencia con su participación y aportaciones.

Por parte de las Humanidades Digitales los efectos son aún más claros y directos: los investigadores necesitan objetos digitales y necesitan que éstos vayan acompañados de la información sobre los procesos que los hace tratables por ordenadores, compartibles entre diferentes colectivos y reutilizables.

3. Bibliotecas de investigación, especializadas y centros de documentación

Después de tanto preludio, quizá sea ahora el momento de parar y examinar algunos productos documentales que reflejan este cambio en las maneras de acercarse y usar la información por parte de los individuos. Vamos a comentar brevemente cuatro iniciativas:

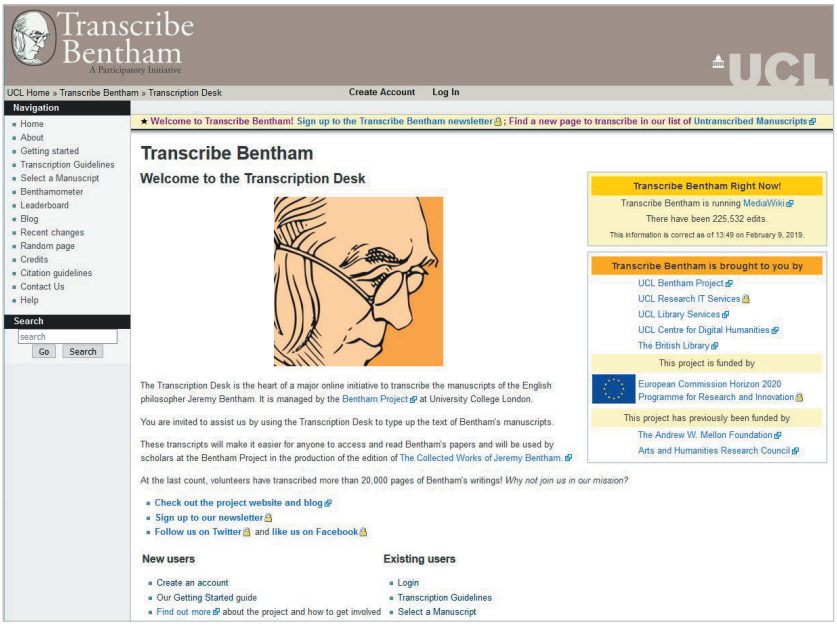
- *La colección* del *Museo del Prado*,
- *Transcribe Bentham*,
- *eBird*, y
- *Mapa Literari Català*.

Museo del Prado

La web del *Museo del Prado* tiene el apartado *La colección* en el que su director describe las colecciones del mismo y da entrada a explorar a distancia una selección de más de 16.000 obras[6]. No vamos a comentar aquí la presentación y las capacidades del servicio de consulta online de las obras –que es magnífica-, pero sí algunos de sus detalles que vienen a cuento de lo tratado hasta ahora. El visitante online puede interactuar con cada obra (ampliar o reducir la imagen, fijarse en un detalle…), usarla para comunicarse a través de redes sociales (*Twitter*, *Facebook* y correo electrónico), re-

https://www.museodelprado.es/coleccion

lacionarla con otras de la colección (a través de enlaces y etiquetas creados por los conservadores), y, finalmente, descargarla sin restricciones con lo que puede reutilizarla para lo que crea mejor (se distingue entre uso privado y comercial)[7].



http://blogs.ucl.ac.uk/transcribe-bentham/about

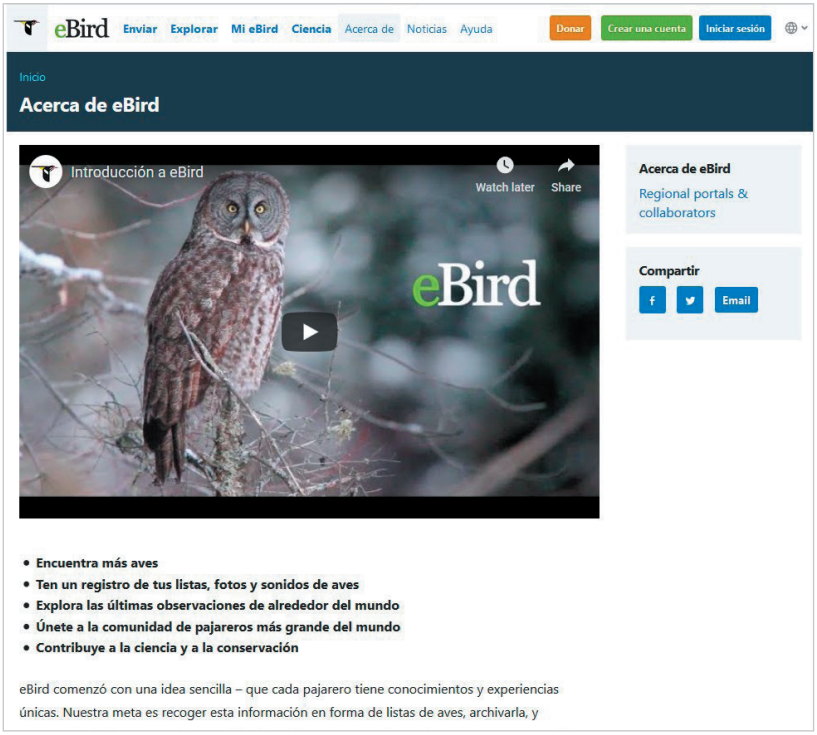
Transcribe Bentham

Es una iniciativa participativa basada en un proyecto de investigación de la *University College London*. Su objetivo es transcribir documentos manuscritos originales y no estudiados del filósofo, economista, pensador y escritor inglés Jeremy Bentham (1748-1832) a través de la participación ciudadana. Según el último recuento, los voluntarios han transcrito así más de 20.000 páginas de los escritos de Bentham[8]. El objetivo final del proyecto es científico y quiere facilitar la edición de las obras completas del filósofo, que dejó muchos manuscritos sin publicar. Pero esto lo hace a través de la participación ciudadana que se motiva de la forma siguiente:

> "se contribuye así a la investigación sobre Bentham, se preserva y difunde su obra, se aprende y se desarrollan capacidades paleográficas, y… uno se lo pasa bien haciéndolo"[9].

eBird

Este proyecto fue puesto en marcha por el *Cornell Lab of Ornithology* y la *National Audubon Society* con el objetivo de recoger información sobre las migraciones de pájaros. Esto lo consigue no con las observaciones de científicos específicamente preparados para ello, sino usando la información que proporcionan los aficionados al avistamiento de pájaros. Más de 300.000 aficionados aportan al año unos 100 millones de registros de avistamiento, una cantidad difícil de conseguir sólo con participantes profesionales. *eBird* no es una iniciativa diletante: la base de datos ha sido usada por científicos para estudiar la conexión entre las migraciones de aves y las lluvias monzónicas en la India, o el cambio climático, por ejemplo. Las personas que aportan información al proyecto ven que su afición contribuye a la ciencia.



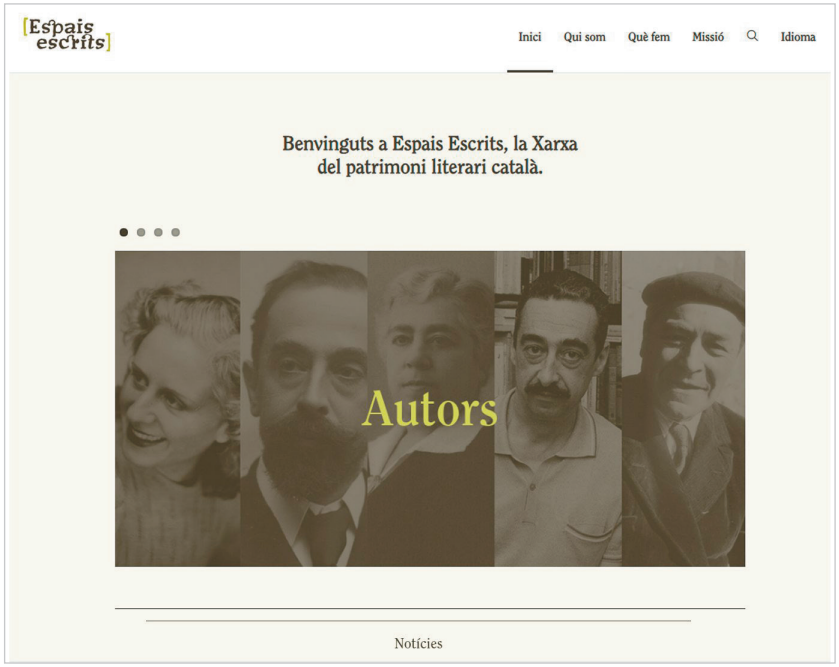
https://ebird.org/about

Mapa Literari Català (MLC)

Es una aplicación web que permite consultar puntos de la geografía mundial asociados con textos de la literatura catalana. Cada localización geográfica o espacio escrito está formado por un texto relacionado con el territorio, así como toda una serie de contenidos multimedia que enriquecen la experiencia del usuario-lector (fotografías de la época de los escritores, vídeos explicativos o con el propio escritor recitando sus textos, enlaces a manuscritos y ediciones varias, traducciones…). Los escritos se completan con unas contextualizaciones para explicar los vínculos de esa obra y de su autor con el territorio. *Mapa Literari Català* está producido por la asociación *Espais Escrits*, una entidad sin afán de lucro que aglutina instituciones que conservan y promueven la lectura y estudios de los escritores patrimoniales de la literatura catalana. *MLC* ofrece unas cincuenta rutas literarias que permiten seguir la trayectoria vital o literaria de los escritores, fo-

menta clubes de lectura sobre obras de autores incluidos en el mapa, mantiene una agenda de actividades literarias, ofrece citas de los autores… En resumen, fomenta experiencias alrededor de la literatura.

Desconozco el grado de participación de archiveros, bibliotecarios o documentalistas en la gestación y mantenimiento de estos productos, pero los cuatro reflejan el cambio que se ha producido recientemente en la forma en como los ciudadanos e investigadores se acercan a los documentos. La información como objeto pierde importancia mientras lo gana la información como proceso. La información en sí misma tiene valor en tanto en cuanto de ella se deriva una experiencia emocional e intelectual.



https://www.espaisescrits.cat/que-fem/mapa-literari-catala

4. Direcciones de futuro

Estamos viviendo un momento que tiende a hacer que toda la información que una persona tendrá en cuenta esté en la Red. Esto es así por dos motivos: porqué cada vez hay más información en internet, y porqué lo que no está en ella no cuenta a efectos prácticos. A medida que ese momento llega, la biblioteca como agregador de contenidos escasos pierde valor. La manera que bibliotecas y centros de documentación tienen de ganar valor es desplazar su atención y servicios desde el documento hacia el proceso de usarlo, y cuando decimos usarlo no nos referimos sólo a facilitar su acceso sino a facilitar su reutilización y a favorecer que este uso permita y fomente la implicación, participación y empoderamiento del usuario[10].

Vistos los cambios en las técnicas documentales y en el hacer diario de los usuarios de información, los documentalistas o bibliotecarios especializados, creo que tendrán el viento en la espalda –empujando- si orientan sus servicios a hacer tres cosas:

- Digitalizar de forma enriquecida
- Vehicular (fomentar) la participación ciudadana, y,
- Crear (reforzar) comunidades.

> Para ganar valor, la bibliotecas y centros de documentación tienen que desplazar su atención desde el documento hacia el proceso de usarlo

Las colecciones de las bibliotecas especializadas han sido su fortaleza, pero hemos visto que esto va a dejar de ser un valor en poco tiempo. Lo que va a continuar teniéndolo van a ser las colecciones especiales o únicas. Diversos autores (**Dempsey**; **Malpas**; **Lavoie**, 2014; **Dempsey**, 2016) consideran que el enfoque bajo el que se tratan las colecciones debe cambiar en este nuevo entorno de información digital en red. Debemos digitalizar tanto como podamos, teniendo por prioritario lo más escaso, pero ya no basta con tener imágenes de alta resolución, sino que debemos digitalizar de forma enriquecida. Reutilización es la música que suena (o la palabra clave, si lo prefieren). Reutilización por parte de los investigadores, que necesitan información detallada de los objetos digitales, reutilización por parte de los usuarios, que quién sabe lo que quieren hacer con las imágenes, y, reutilización por parte de los ordenadores que serán los que en muchos casos usaran los objetos digitales y que no pueden 'verlos', pero sí conocer sus características si estas están debidamente descritas y codificadas.

A partir de un cierto nivel de bienestar, las personas no nos conformamos con el mero disfrute de un bien, reclamamos la participación en la gestión del mismo. En muchos aspectos las bibliotecas han sido creaciones que una autoridad ilustrada ofrecía a sus administrados, pero la gente reclama (reclamamos) formar parte del movimiento. En el pasado hemos visto que la ciencia ha tendido a justificarse por si misma, pero ahora debe hacerlo mostrando que sintoniza con los valores y preocupaciones de la sociedad que la sufraga. Las bibliotecas y centros de documentación ocupan una posición de privilegio para ser quién haga de enlace entre una institución y el público al que se dirige. Este enlace puede tomar la forma de ciencia ciudadana (contribución de los ciudadanos a la ciencia aportando datos) o la de experimentación (refuerzo de los procesos de aprendizaje por co-creación), pero en cualquier caso necesita tener un agente neutral y próximo que lo facilite, y proximidad y neutralidad son características que definen a las bibliotecas (**Rooney-Browne**; **McMenemy**, 2010).

Finalmente, nuestro enfoque profesional en los objetos portadores de información nos ha alejado del hecho que detrás de los mismos hay personas y que a estas les une un nexo que es el que justamente les conduce a visitar la biblioteca. El alto sentido social de los humanos se manifiesta en la necesidad de sentirse parte de diversos colectivos o comunidades (territoriales, de gustos o dedicaciones, de intereses…) y las bibliotecas son organizaciones con capacidad y disponibilidad de reforzar estas comunidades. Quizá nadie lo ha expresado mejor que R. David Lankes cuando redefinió la misión de las bibliotecas como la de mejorar la sociedad a través de facilitar la creación de conocimiento en sus comunidades, o cuando afirma que:

> Uno de los postulados base de la Ciencia Abierta es que la investigación debe ser hecha con y para la sociedad

> "Las bibliotecas malas construyen colecciones, las buenas bibliotecas construyen servicios, las grandes bibliotecas construyen comunidades" (**Lankes**, 2011; 2012).

Vivimos momentos de cambios transformadores, momentos en los que parece que adherirse a cualquier novedad tecnológica es un acierto; pero la más nueva novedad tecnológica se verá (más pronto que tarde) sustituida por una novedad más nueva. No es momento para adhesión a modas y sí para la adopción de tendencias de largo alcance. No hay otra forma de acertar en las múltiples disyuntivas con las que nos encontraremos que tener nuestra razón de ser como mecanismo discriminador. Y esta razón de ser nuestra no es coleccionar. Coleccionar es (pronto diremos, fue) el medio para la finalidad de facilitar a las personas el proceso de consumir información. Lo dice así de bien David W. Lewis a propósito de las bibliotecas universitarias:

> Lo que va a continuar teniendo valor son las colecciones especiales o únicas

> "...tenemos que reconocer que... lo que debería tratarse no es salvar la biblioteca. Más bien... debería tratarse de ofrecer productos o servicios que puedan ayudar a los estudiantes y profesores a realizar de manera más efectiva, conveniente y económica el trabajo que han estado tratando de hacer en sus vidas académicas" (**Lewis**, 2016).

5. Notas

1. Este artículo tiene su origen en una intervención en el *IX Encuentro de Centros de Documentación de Arte Contemporáneo* que, bajo la coordinación de Elena Roseras Carcedo, tuvo lugar en el *Museo Artium* de Vitoria-Gasteiz los días 24 y 25 de octubre de 2018.
http://www.artium.org/es/explora/actividades/item/60805-ix-encuentros-de-centros-de-documentacion-de-arte-contemporaneo

2. *https://dle.rae.es/?id=KnPZIuF*

3. Quienes tengan curiosidad pueden consultar:
https://whatisdigitalhumanities.com
donde aparece una definición diferente de Humanidades Digitales cada vez que refresquen la página. Las definiciones provienen de las aportaciones a la iniciativa de "*Day of DH*".

4. Uso la definición que dio Núria Bel en el curso "*Humanitats Digitals*" (*CSUC*, 15-03-2018). El concepto incluiría las reflexiones críticas de los efectos de las tecnologías digitales en la cultura, pero esta acepción no tiene aplicabilidad al caso que nos ocupa.

5. El desarrollo de las siglas MARC es *MAchine-Readable Cataloging*.

6. El número cada vez mayor de museos que permiten consultar sus obras a distancia ha ido aumentando en los últimos años. Quizá ahora lo encontremos 'normal', pero hasta no hace mucho, los museos no daban facilidades para ver y usar reproducciones de sus obras: la digitalización de las colecciones estaba gobernada por el paradigma predigital de que poner a disposición de los usuarios copias de las obras detraería visitantes e ingresos al museo. Pero nuestra sociedad es cada vez más 'abierta' y no se concibe ya la función museística como atesoramiento patrimonial y sí como instrumento de compartición pública.

7. Con posterioridad a la conferencia dada en Vitoria veo que el *Museu Nacional d'Art de Catalunya* ha creado una app que permite explorar algunas de sus obras en imágenes de alta resolución. Significativamente el *MNAC* presenta la app así: "Disfruta de una nueva experiencia y redescubre el museo a través de un *storytelling* interactivo":
https://www.museunacional.cat/es/articulo/una-nueva-forma-de-explorar-el-museo-la-app-second-canvas

8. Un proyecto más cercano, parecido (y a su vez diferente) al comentado es "Transcriu-me!", una iniciativa cooperativa impulsada por la *Biblioteca de Catalunya* con la que se han transcrito unos 4.000 documentos, entre ellos 351 pergaminos:
http://transcriu.bnc.cat

9. *http://blogs.ucl.ac.uk/transcribe-bentham/about*

10. Otro ejemplo, la decisión del *Cleveland Museum* de ofrecer en acceso abierto imágenes digitales de sus obras que

están en el dominio público. El museo no sólo las deja 'ver', sino que "Usted las puede usar como quiera": *https://www.cleveland.com/expo/life-and-culture/g66l-2019/01/fe82a74cbf1054/cleveland-museum-of-art-launches-nextgeneration-open-access-to-artworks-and-data-online-.html*

6. Referencias

Anglada, Lluís M. (2016). "Rellenando espacios: las bibliotecas como tejido conector en una sociedad densa". En: *Bibliotecas 2029: Documentos de las Jornadas*. Murcia: Anabad, p. 79-83.
http://hdl.handle.net/10760/29123

Anglada, Lluís M.; **Abadal, Ernest** (2018). "¿Qué es la ciencia abierta?". *Anuario ThinkEPI*, v. 12, pp. 292-298.
https://doi.org/10.3145/thinkepi.2018.43

Buckland, Michael K. (1991). "Information as thing". *Journal of the American Society for Information Science*, v. 42, n. 5, pp. 351-360.
https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3

Cobarsí-Morales, Josep; **Canals, Agustí**; **Ortoll, Eva** (2016). "Infonomista: una discusión actualizada". *El profesional de la información*, v. 25, n. 2, pp. 255-261.
http://dx.doi.org/10.3145/epi.2016.mar.12

Dempsey, Lorcan (2012). "Libraries and the informational future: some notes". En: Marchionini, Gary; Moran, Barbara B. (eds.). *Information professionals 2050: Educational possibilities and pathways*. Chapel Hill: School of Information and Library Science, University of North Carolina at Chapel Hill, pp. 113-126. ISBN: 978 1 300 20486 2
http://sils.unc.edu/sites/default/files/publications/Information-Professionals-2050.pdf

Dempsey, Lorcan (2016). "Library collections in the life of the user: two directions". *Library quarterly*, v. 24, n. 4.

Dempsey, Lorcan; **Malpas, Constance**; **Lavoie, Brian** (2014). "Collection directions: The evolution of library collections and collecting". *Portal: Libraries and the academy*, v. 14, n. 3, pp. 393-423.
http://www.oclc.org/research/news/2014/10-14.html
https://doi.org/10.1353/pla.2014.0013

European Commission (2016). *Open innovation, open science, open to the world: A vision for Europe*. Brussels: European Commission. Directorate General for Research and Innovation, p. 5.
https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe

Lankes, R. David (2011). *The atlas of new librarianship online*. MIT Presses.
https://davidlankes.org/new-librarianship/the-atlas-of-new-librarianship-online

Lankes, R. David (2012). *Expect more: Demanding better libraries for today's complex world*.
https://davidlankes.org/wp-content/uploads/2014/01/ExpectMoreOpen.pdf

Lasauca-Cisa, Xavier (2018). "Una aproximació al concepte de ciència oberta (i 25 recursos per aprofundir-hi)". *L'ase quàntic,* 31 julio.
https://lasequantic.wordpress.com/2018/07/31/una-aproximacio-al-concepte-de-ciencia-oberta-i-15-recursos-per-aprofundir-hi

Lewis, David W. (2016). *Reimagining the academic library*. Rowman & Littlefield Publishers, p. 91. ISBN: 978 1 442238589

Liber (2018). *Open science roadmap*.
https://zenodo.org/record/1303002

Morozov, Evgeny (2015). *La locura del solucionismo tecnológico*. Móstoles: Katz Editores. ISBN: 978 84 15917 19 9

Rooney-Browne, Christine; **McMenemy, David** (2010). "Public libraries as impartial spaces in a consumer society: possible, plausible, desirable?". *New library world*, v. 111, n. 11/12, pp. 455-467.
https://doi.org/10.1108/03074801011094831

Wiorogórska, Zuza (2016). "¿Real o imaginario? ¿Qué escenarios contemplamos para las bibliotecas? El caso de Polonia". En: *Bibliotecas 2029: Documentos de las jornadas*. Murcia: Anabad, pp. 11-16.
http://hdl.handle.net/10760/29123
http://eprints.rclis.org/29123

Digital journalism: 25 years of research. Review article

Ramón Salaverría

Nota: Este artículo se puede leer en español en:
http://www.elprofesionaldelainformacion.com/contenidos/2019/ene/01_es.pdf

How to cite this article:

Salaverría, Ramón (2019). "Digital journalism: 25 years of research. Review article". *El profesional de la información*, v. 28, n. 1, e280101.

https://doi.org/10.3145/epi.2019.ene.01

Article received on September 29th, 2018
Approved on December 13th, 2018

Ramón Salaverría ✉
http://orcid.org/0000-0002-4188-7811

University of Navarra, School of Communication
Journalism Projects Department
Edificio Bibliotecas, 31009 Pamplona, Spain
rsalaver@unav.es

Abstract

In 1994, the first web online media outlets were introduced in several countries around the world. Twenty-five years later, digital or online journalism is a confirmed reality and common practice in professional and academic circles. Based on an extensive bibliographic review, this article examines the main areas of academic research related to digital media at the global level in the last quarter of a century. It shows the lines of research on the history of journalism on the Internet, the forms of digital media, their languages and economic challenges. It also reviews the most widespread research theories and methods. The analysis confirms that research on digital journalism is a strong, ongoing discipline, despite the fact that several methodological and thematic challenges will need to be addressed in the next few years.

Keywords

Journalism; Digital journalism; Digital media; Academic research; Communication theory; Research methods; History of communication; Review article.

1. Introduction

What is digital journalism? This question may seem too simple or even inappropriate for a scientific journal. However, the most basic questions are often the most persistent and difficult to answer. And the most important.

If long-standing researchers have posed questions in recent years such as "What is news?" (**Harcup**; **O'Neill**, 2001; 2017) and "What is 21st-century journalism?" (**Deuze**, 2005; **McNair**, 2006; **Broersma**; **Peters**, 2013), it cannot be out of place to discuss one of its most successful versions in the last twenty-five years: digital journalism. Besides, the answer is not all that obvious, no matter how long you think about it. This is confirmed by the huge number of research studies carried out on the topic since the first online media outlets were launched twenty-five years ago in the mid-1990s.

The aim of the next few pages is to review not only the theoretical contributions made to digital journalism (a highly elusive, changing, multifaceted concept), but also the main areas of academic research on the topic. The year 2019 will mark the 25th anniversary of the first news media outlets on the World Wide Web, which makes it the ideal time to evaluate the developments, findings and unfinished business of research on this discipline.

Rather than providing a comprehensive review, this article will discuss the main points. It is literally impossible to include everything published on this subject in a single article. Despite the lengthy bibliography at the end of this article (with over 200 references), this list of publications is merely a condensed selection of the towering piles of content researched and written in recent years on digital journalism. It is impossible to mention all the authors who have written on this topic, but the article discusses most of the important ones. Despite the inevitable omissions and oversights, the aim is to give readers a general idea of the main research areas in digital journalism and online media outlets in the last 25 years, as well as their achievements and current challenges.

In fact, this is not the first assessment of the progress of research on digital journalism and communication in the last few decades. Several authors have helped define the limits of this discipline through extensive reviews that have outlined the research community's main areas of study in recent years (**Fernández-Quijada**; **Masip**, 2013; **Mercier**; **Pignard-Cheynel**, 2014; **Reese**, 2016). These works are warmly welcomed because the accumulation of specific empirical studies often makes it difficult for researchers to distinguish passing trends from basic essentials.

For example, **Steensen** and **Ahva** (2015) carried out a study of this kind on the 20th anniversary of the appearance of online media outlets. Their analysis, based on the abstracts of all the articles published between 2000 and 2013 in two of the most distinguished academic journals on the discipline, *Journalism* and *Journalism Studies*, determined that lines of research on journalism had increased dramatically in this period. They observed that studies had moved away from the typical political science approach and were turning increasingly towards a sociological perspective, both in society as a whole and within media outlets. They also perceived that matters of an ethical and philosophical nature such as ethics and objectivity in journalism were gaining importance in journalism studies. In terms of theory, they identified an evolution from pure empiricism to increased attention to theoretical aspects, where the dominant model was grounded theory.

Other authors carried out similar review works and focused specifically on online journalism (**Masip** *et al.*, 2010). The most comprehensive review was done by **Mitchelstein** and **Boczkowski** (2009), who classified international research in this discipline in five main areas:

- the historical and market context,
- the innovation process,
- changes in journalistic practices,
- challenges in established professional dynamics, and
- the role of user-generated content.

More recently, other research reviews addressed specific aspects such as the more than 300 research studies on the impact of digital media on political participation (**Boulianne**, 2018).

All reviews are incomplete by nature. They force you to summarize, which means that taking into account all references and approaches is simply impossible. In this article, I decided to leave out bibliometric analysis, which examines research production based on essentially quantitative parameters. For those interested in this type of approach, several recent academic studies are available (**Masip**, 2005; **Castillo**; **Carretón**, 2010; **Caffarel-Serra**; **Ortega-Mohedano**; **Gaitán-Moya**, 2017). I have opted for a descriptive approach in this article with the aim of exploring the paths followed by research on digital journalism: history, languages, economic challenges and related topics. However, before we begin, we must first decide on some concepts. Let's begin with our opening question: what is digital journalism?

2. Looking for a term (and its definition)

It may seem surprising that more than 20 years since journalism in digital media took its first steps, there is still no established common name for it used by international researchers. In fact, as we shall see, agreement has not been reached on the definition or even the actual name of this type of journalism. Even today, professionals and academics are not in agreement when it comes to choosing between expressions such as "digital journalism" (**Kawamoto**, 2003), "cyber journalism" (**Díaz-Noci**; **Salaverría**, 2003), "online journalism" (**Deuze**, 2001; **Steensen**, 2011), "multimedia journalism" (**Deuze**, 2004), and a long list of similar labels (**Karlsen**; **Stavelin**, 2014).

> More than 20 years since journalism in digital media took its first steps, there is still no established common name

The most widespread expression among professionals in most countries is definitely "digital journalism." Internationally, however, academics do not appear to be in such agreement and, depending on each case and country, they opt for one word or another. For example, in the Spanish- and Portuguese-speaking research communities, the term "cyber journalism" (*ciberperiodismo* in Spanish and *ciberjornalismo* in Portuguese) is the most widely used. By contrast, in the French-speaking world, the form "digital journalism" (*journalisme numérique*) prevails. Among English and German speakers, the preferred expressions for most of the last 25 years were "online journalism" in English and *online journalismus* in German. However, in the last 10 years "digital journalism" has been gradually taking over and is now the most popular form.

Spanish readers will notice that this article, originally written in Spanish, uses the expression "digital journalism" (*periodismo digital*) in the title, whereas in the past the author would have used "cyber journalism" (*ciberperiodismo*) in Spanish (**Salaverría**, 2005a). This apparent change of criteria should be explained. For the reasons of linguistic accuracy and expressive economy I discussed at the time (**Salaverría**, 2005a, p. 20), I still maintain that the term *ciberperiodismo* is more accurate and appropriate in Spanish, despite its current limited use in the professional world. It defines the specialty of journalism that uses cyberspace to investigate, produce and, above all, disseminate news content (p. 21). It therefore identifies the journalism carried out "on" digital networks. But the expression *periodismo digital* is broader, as it refers to all journalism carried out "with" digital technologies. The difference is subtle, but relevant. In broad terms, the expression "digital journalism" encompasses all forms of journalism that use digital resources; it therefore includes not only Internet resources and mobile networks, but also digital television (**Robin**; **Poulin**, 2000) and digital radio (**Martínez-Costa**, 1998). Although digital television and radio are rarely considered part of cyber journalism, they are indisputably digital expressions of journalism. The term *ciberperiodismo* is therefore more specific than *periodismo digital* in that it refers exclusively to what is done by media outlets in cyberspace, i.e., the Internet. Therein lies the subtle difference between the two concepts.

This clarification helps explain why I used the expression "digital journalism" in the title of this article: I aim to include all digital forms of journalism, not only those related to interactive networks. In fact, since computer technologies began to be embraced by the media, relevant academic contributions have been made on digital forms of journalism that have not been confined to Internet media (**Pool**, 1983). This article will reflect on all digital forms of journalism while paying particular attention to those that have focused specifically on Internet media outlets. The research done on all of these digital forms of journalism is discussed on the next few pages.

3. Historic perspectives on digital journalism

Journalism's initial encounter with computing predates the advent of the Internet. In fact, it dates back to the 1950s, when some media outlets began to experiment with using computers in media coverage. The experiment by *CBS* on November 4, 1952 is often cited as a pioneering example. On the occasion of the US presidential elections, which were held on that day, the television network used a primitive computer, the gigantic *Univac*, to predict (correctly) the final result of the vote based on the count of the first votes (**Bohn**, 1980).

"Journalism's initial encounter with computing predates the advent of the Internet"

Over the next four decades, until the early 1990s, media and digital technology continued on a course of mutual approach. Meanwhile, an increasingly scientific and technological approach was under way toward handling information, under the label of precision journalism (**Meyer**, 1973; 1991). While this idea achieved relative success among journalists, it had a considerable impact on journalism researchers and professors in the following decades. Indeed, its echo can still be heard. There is a clear connection between the precision journalism formulated by **Meyer** in the 1970s and one of the most successful aspects of digital journalism in our era: data-driven journalism.

Up until the late 1980s, during the years in which digitization became gradually assimilated by the media, theoretical treatises (**Bagdikian**, 1970; **Washburn**, 1981) and empirical research (see, for example, **Reese**, 1988; **Reagan**, 1989) on the adoption of the "new technologies" in journalism were common. As **Armand Mattelart** (2001) explained, this was also when the influential contemporary concept of the information society was definitively forged (**Masuda**, 1981). At that time, most of the research on technological change focused on television, which was then considered to be the main electronic medium (**Meyrowitz**, 1985, p. 5). To a much lesser extent, some studies were also being published on the use of computers and databases in journalistic information (**Aumente**, 1987) and on the vague potential of developing journalism in interactive networks (**Koch**, 1991). A few authors, however, including apocalyptic authors such as **Neil Postman** (1992) and integrated authors such as **Nicholas Negroponte** (1995) and **Roger Fidler** (1997), glimpsed the breadth of the change that was being triggered in the media. In the early 1990s, a genuine technological revolution was knocking at the door of journalism.

David Carlson (2003) points to January 19, 1994 (25 years ago) as the date when the first news publication appeared on the World Wide Web. According to that author, the honor of inaugurating the history of Web-based journalism belongs to the *Palo Alto Weekly*, a humble weekly newspaper published in the San Francisco bay area. Before this, however, other larger media outlets had launched online editions on the commercial networks of the time, such as *America OnLine* (*AOL*), *Compuserve* and *Prodigy*. The first to do so was the *Chicago Tribune*, which published an edition on *AOL* in March 1992 (**Díaz-Noci**; **Meso-Ayerdi**, 1998). This newspaper was followed by big newspapers such as *The New York Times*, *The Washington Post*, the *Los Angeles Times* and *USA Today*. Alongside these big dailies, one of the most influential newspapers of the time was the Californian *San Jose Mercury News*, with its advanced digital edition called *Mercury Center* (**Chyi**; **Sylvie**, 1998). The proliferation of digital media was so rapid that, almost immediately, those publications became known as the fourth medium (**Bonnington**, 1995), in competition with the press, radio and television.

Other countries soon followed the trend that had begun in the United States. Between 1994 and 1996, most countries in the world saw the birth of their first digital media outlets, which were almost always driven by newspapers. Historical research on online media still has much to learn because there are few countries with monographs on the origin and initial development of online media in their respective territories. Nevertheless, many studies of this kind have been published on countries all over the world. Indeed, historical reviews on the evolution of online media can be found in many different countries such as Argentina (**Bergonzi** *et al.*, 2008), Bolivia (**Cabrero-Rubio**, 2005; **Banegas-Flores**; **Barba-Sanjínez**, 2014), Cameroon (**Ngounou**, 2010), Canada (**Perigoe**, 2009), China (**Lagerkvist**, 2008), Costa Rica (**Bravo-Herrera**, 2003), Ecuador (**Rivera-Rogel**, 2012), Estonia (**Tähismaa**, 2003), France (**Charon**; **Le-Floch**, 2011; **Atheaume**, 2013), Israel (**Caspi**, 2011), Italy (**Gazoia**, 2013), Mexico (**Crovi-Druetta**; **Toussaint**; **Tovar**, 2006), Nigeria (**Kperogi**, 2012), Peru (**Yezers'ka**, 2008), Portugal (**Bastos**, 2010), South Africa (**Bosch**, 2010), Spain (**Díaz-Noci**, 2005; **Salaverría**, 2005b; **Salaverría**, 2016, pp. 169-209), United States (**Carlson**, 2003; **Scott**, 2005) and Venezuela (**Rojano**, 2006).

> There are few countries with monographs on the origin and initial development of online media in their respective territories

Coincidentally, the first online media in most of these countries emerged at around the same time. With the exception of the most economically and technologically limited countries and countries under authoritarian regimes and severe restrictions on freedom of the press, by the mid-1990s, virtually every country in the world had been hit by that wave. Thanks to these pioneering media, which appeared first on commercial networks and, shortly afterward, on the World Wide Web, the history of online journalism had just begun. And with it came the exuberant development of academia's areas of research.

The spectacular proliferation of online media spurred the sudden interest by researchers. By the year 2000, the main online media directory of the time, which was maintained by the professional American journal *Editor & Publisher*, listed 4,400 online publications worldwide. Of these, 3,161 were based in North America, 1,634 in Europe, 269 in South America and 214 in Asia (**Chyi**; **Sylvie**, 2000). The phenomenon very rapidly went global (**Siapera**; **Veglis**, 2012; **Dragomir**; **Thompson**, 2014). As a result of this globalization, national studies have been carried out such as the ones listed above, and comparative analyses have also been done of the evolution of digital news media outlets on entire continents. Examples of this perspective include several monographs on online journalism in Europe (**Van-der-Wurff**; **Lauf**, 2005; **Oggolder** *et al.*, 2019), Africa (**Mabweazara**; **Mudhai**; **Whittaker**, 2014) and Latin America (**Salaverría**, 2016).

This national and international research has encountered several as yet unsolved methodological difficulties. As **Díaz-Noci** (2013) points out, researching the history of digital journalism poses major methodological challenges due to the heterogeneity and evanescence of online documents, the lack of systematic archives in most online media and, I would add, the vastness of the object of research. Because of the difficulty of covering the complete framework of online journalism, there have been far more studies in the last 25 years that focus on specific media outlets (e.g., **Brinca**, 2006; **Molinos**; **Marques**; **Ferreira**, 2006; **Smyrnaios**; **Bousquet**, 2011), often as part of doctoral theses, than on national online journalism markets as a whole. In other words, there have been far more studies of trees than of forests.

> Although the number of digital publications worldwide is unknown, it can be estimated in hundreds of thousands

4. Models of digital journalism

The tendency to study specific media outlets in particular is not surprising, given that the forest of online media publications is not a small one. It might be better described as a rapidly growing jungle. As mentioned above, there were an estimated 4,400 online media publications in the whole world in 2000. And yet, a study carried out in 2018 found more than 3,000 active online publications in Spain alone (**Salaverría**; **Martínez-Costa**; **Breiner**, 2018). We do not know how many digital publications are active worldwide, as there are no studies of this size. Nevertheless, a conservative extrapolation of the census for Spain makes it possible to venture with some confidence that the figure will be at least in the hundreds of thousands.

Given this large size, the diversity of online media is clearly enormous. More than two decades after the appearance of the first digital publications, most of them online editions of daily newspapers, the current diversity of online media is extraordinary: there are digital outlets linked to traditional journalism brands as well as native digital media outlets. Although publications for multiple devices are the norm, there continue to be publications designed solely for viewing on personal computers, while publications designed exclusively for tablets or smartphones are also emerging. As with traditional media, general and specialized media coexist on the Internet; free publications with paid publications; global online publications with national, local (**López-García**, 2008) and even hyperlocal publications (**Tenor**, 2018). And the list keeps growing.

There have been stages in this development, which continues today. As **Scolari** (2013a) explains, the first stage is always that of transition between what exists and what is in the process of being born because

> "each new medium that appears fills its space with content from other media" (p. 16).

Several authors (**Fidler**, 1997; **Bolter**; **Grusin**, 2000) have described this connection between traditional media —press, radio, television, news agencies— and "new media" (**Pavlik**, 2001), a label that is now out of date. In the 1990s and in the early years of this century, the predominant model in online media was based on that of traditional media, particularly newspapers. However, over time, digital media have become increasingly independent in both form and content. Most of the first digital publications used the shovelware approach (**Bardoel**; **Deuze**, 2001), a publishing model that consists of simply reproducing on the Internet the content originally created for another platform. Although that model can still be found in some primitive online publications, it has been surpassed by another model, which is based on producing original content for the Internet. Just as has occurred with journalism genres, digital media show an evolution in four stages: repetition, enrichment, renovation and innovation (**Salaverría**; **Cores**, 2005, pp. 148-149). Twenty-five years after the birth of the first online publications, digital media are finally entering the fourth stage.

Along the way, however, there have been some major shake-ups. One of the most disruptive of these, for journalism, was unquestionably the success of content generated and distributed by users. In the early days of this century, the first signs of a phenomenon appeared: the active participation of citizens in public communication, which, over time, has transformed the traditional role of the media. The appearance of blogs, hailed as a revolution by some authors (**Orihuela**, 2006), signaled the start of this phenomenon, which rapidly incorporated the voice of citizens into public communication in general and also, more specifically, into journalism. Of course, media publications had long ago timidly opened their doors to public opinion through features such as letters to the editor (**Grey**; **Brown**, 1970). But the arrival of blogs was a qualitative leap. This new form of publishing allowed for public communication that was free and simple and had tremendous reach. And it was also enthusiastically embraced by millions of users throughout the world. Suddenly, a new divide arose: professional journalism on one side and user-published content on the other.


JOIN THE DISCUSSION IN THE NEW Forums

The New York Times ON THE WEB

The '96 Yankees On Parade

Gallery in the Round

"All the News That's Fit to Print"

Monday, December 30, 1996

SECTIONS

Front Page • CyberTimes
Politics • Business
Editorials/Letters • Op-Ed
Arts & Leisure • Travel
Real Estate • Job Market
Diversions • Web Specials

NEWS BY CATEGORY

CLASSIFIEDS | FORUMS

SERVICES | SEARCH

TABLE OF CONTENTS

Remains of North Koreans Returned

LATE NEWS UPDATE

Pace Quickens on Efforts to Wind Up Hebron Withdrawal

IN CYBERTIMES

Apple Seeing Mixed Results From Cloning

The New York Times (December 30th, 1996)
https://web.archive.org/web/19961230230427/http://nytimes.com

This dichotomy gave rise to what has been one of the most fertile areas of research in online journalism in all these years: the exploration of the basic tenets, modalities and consequences of user-generated content (UGC) (**Van-Dijck**, 2009; **Leung**, 2009) and journalists' tortuous assimilation of it into their daily professional routines (**Hermida**; **Thurman**, 2008; **Paulussen**; **Ugille**, 2008). Early research on user-generated content also coincided with a growing criticism of traditional media, which, some predicted, would be unable to adapt to the new context. For example, Jakob Nielsen, known for his publications on website usability, predicted in 1998 that

> "most current media formats will die and be replaced with an integrated Web medium in five to ten years" (**Nielsen**, 1998).

That prediction was made 20 years ago and yet we still have books, newspapers, magazines, radio and television.


The New York Times

Thursday, January 17, 2019

Today's Paper

Steady Raises and a Pension: These Workers Had a Good Deal Until the Pay Stopped

"We are getting crushed!" Despite his public bravado, President Trump is said to be frustrated with aides over the shutdown.

White House Redefines Who Is Essential to Get Parts of Government Moving Again

Four Americans Among Dead in ISIS Attack in Syria

Democrats Fall Short in Russia Sanctions Vote

Vice President Mike Pence said North Korea had failed to take any substantive steps to give up its nuclear weapons.

Opinion

The Malign Incompetence of the British Ruling Class

Kamala Harris Was Not a 'Progressive Prosecutor'

There's Nothing Wrong With Open Borders

The Kind of Judge We Need

The Rudderless West

The Fight Over Men Is Shaping Our Political Future

Marching Toward a Massacre

Listen to 'The Argument': How Trump Loses the G.O.P.

Take the Trump-Russia Quiz

Germany's China Problem

The Bond Market Comes to Its Senses

The New York Times (January, 17th, 2019)
https://www.nytimes.com

Some predicted not just the disappearance of certain media, but even the death of professional journalism as a whole. Indeed, in the mid-2000s, when ideas about the technological evolution toward Web 2.0 were in vogue (**O'Reilly**, 2009), the concept of citizen journalism was popularized as a presumptive alternative to journalism produced by professionals. **Dan Gillmor** (2004) was the main proponent of these ideas, although many other authors spoke out in their favor (**Bowman**; **Willis**, 2003). According to their view, the days of media produced by professionals and managed by companies were essentially numbered due to the rising tide of content shared by ordinary citizens. As Manuel Castells put it, there was a move from mass communication to "mass self-communication" (**Castells**, 2007).

> Research examining journalism in relation to social media have become one of the main areas of studies on digital journalism

While some authors reviewed the theoretical bases of citizen journalism critically (**Lewis**; **Kaufhold**; **Lasorsa**, 2010), many empirical studies tested its real reach in many different news markets (**Allan**; **Thorsen**, 2009). Many of those studies downplayed the triumphant tone of those who were announcing a radical shift in journalism; the new participatory forms of producing news content were more of a complement to the media than a substitute. Indeed, it was soon seen that, as had happened in the past with other technological innovations, the media gradually assimilated these new phenomena into their daily professional routines. Digital media incorporated blogs, opened up spaces for readers' comments and, in general, offered many new ways for the public to participate in and contribute to the news, thus giving rise to a trend called crowdsourcing (**Howe**, 2006).

Despite the fanfare that accompanied the launch of certain media dedicated to user-generated information (the most famous case undoubtedly being the South Korean *OhmyNews* (**Gillmor**, 2004, pp. 125-129)), citizen journalism soon proved to be nothing more than a fad and, as such, quickly declined. Nevertheless, some of its essential traits survived. While it became clear that active audiences were not about to cause the collapse and immediate replacement of the media, they had triggered an irreversible underlying transformation in the relationship between the media and the public. The public no longer consisted of mere passive recipients and were now active interlocutors. "Participatory journalism" (**Domingo** *et al.*, 2008) was born, also labeled collaborative journalism (**Bruns**, 2005), with its own dynamics and new challenges.

Academic interest in this trend increased with the expansion of social media. The global success of services such as *Facebook*, *Twitter* and *Instagram*, each with hundreds of millions of users, attracted the interest of researchers because of their rapid and profound impact on the media (**Newman**, 2009; **Hermida**, 2012). There was a proliferation of studies examining journalism in relation to social media. At the end of the period analyzed in this article, this research is, in fact, one of the main areas of studies on digital journalism. Based on the key value achieved by the act of sharing (**Noguera-Vivo**, 2018), these studies deal with such wide-ranging aspects as the phenomena of public opinion and political participation in the context of social media (**Gil-de-Zúñiga**; **Jung**; **Valenzuela**, 2012), their impact on the media business (**Qualman**, 2009) and journalistic narratives used on these platforms (**García-Avilés**; **Arias Robles**, 2016).

> Researching digital journalism goes beyond researching its media

5. Online media and platforms

As we have just seen, researching digital journalism goes beyond researching its media. However, as would be expected, one of the most common areas of research on this discipline has been types of online media (**Salaverría**, 2017).

Over these first 25 years, the evolution of digital media has been very rapid and intense. As with humans, who experience their fastest growth in the early stages of life, followed by slower growth, online media have shown considerable mutability in these early years. Researchers have tried to explain this diversity through studies of different kinds.

One of the most frequent approaches has focused on exploring the forms and strategies of digital media on the successive publication platforms:

- the World Wide Web (**Canavilhas**, 2003),
- tablets (**Scolari**, 2013a) and
- smartphones (**Mitchell**; **Rosenstiel**; **Christian**, 2012).

As a result of the popularization of smartphones and app-based communication systems, a large number of studies have focused on this trend in the last ten years, giving rise to the development of a new discipline: mobile journalism or mojo (**Westlund**, 2013; **Wolf**; **Schnauber**, 2015).

Besides studying platforms, researchers have also paid attention to different technologies and services that have been launched on the market over the years with varying degrees of success. These services have often had some connection with journalism, which made them susceptible to analysis of their modalities and potential development as news platforms. This is the focus of studies on journalism in *Second Life* (**Brennen**; **De-la-Cerna**, 2010), journalism with drones

(**Holton**; **Lawson**; **Love**, 2015) and locative journalism, based on geolocation technologies (**Nyre** *et al.*, 2012). Today, one of the areas attracting most interest among researchers is that associated with immersive journalism, a form that is on the rise thanks to augmented-reality and virtual-reality technologies (**De-la-Peña** *et al.*, 2010; **Domínguez-Martín**, 2015; **Shin**; **Biocca**, 2018).

The study of online media models has also been approached from a publishing perspective. In this case, the division has been established between publications derived from media on other platforms and native digital publications. In the 1990s, digital media linked to traditional media, known as legacy media (**Arrese**; **Kaufmann**, 2016), made up almost the entire market; over the years, however, the number of native online outlets has greatly increased. In Spain, media founded on the Internet accounted for one third of all online publications in 2018 (**Salaverría**; **Martínez-Costa**; **Breiner**, 2018). Because of this rapid development, native digital media have begun to be the subject of monographic studies (**Nicholls**; **Shabbir**; **Nielsen**, 2016).

Another different factor of study of online media is their geographic scope. Digital media have shown tremendous versatility in this regard, giving rise to publications with a global reach (*The Huffington Post*, *Quartz*, *Politico*, etc.), as well as hyperlocal publications (**Williams**; **Harte**, 2016). Since the first digital media appeared, it became clear that one of their distinctive traits was their ability to ignore geographic limits. As Nicholas Negroponte explained in the early days of the digital era,

> "In the same ways that hypertext removes the limitations of the printed page, the post-information age will remove the limitations of geography. Digital living will include less and less dependence upon being in a specific place at a specific time, and the transmission of place itself will start to become possible" (**Negroponte**, 1995, p. 165).

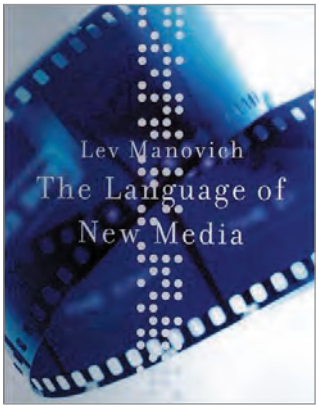
Indeed, unlike the press, radio and television, Internet media have, since their beginning, shown a hitherto unknown quality: while they allowed for maximum personalization of content, they facilitated the transmission of their content to global audiences, beyond the physical territory of the different media. In recent years, this trans-territorial quality has been somewhat curtailed for legal reasons, as restrictions on publishing rights have proliferated, particularly in relation to certain content such as sports videos, audiovisual fiction series and other similar content. This content can be consumed by users in one country, but is blocked for users outside that country. Despite these legal limitations, all digital media are, in essence, global.

Furthermore, as mentioned, digital media are, or can be, personalized. In the 1990s, different prototypes of digital media had already been designed as personalized news services (**Haake**; **Hüser**; **Reichenberger**, 1994). Over the years, however, that promise of à la carte digital media gradually lost steam, as the big news brands opted for digital publications with a limited degree of personalization. In the last ten years, the development of social media based on personal profiles and the refinement of algorithms that can assign content increasingly adapted to the interests and preferences of each user have brought the old idea of personalized media back to life. The new wave of technologies using artificial intelligence, which can produce, organize and distribute "their" information to each user, poses several professional and ethical challenges for journalism that are being actively analyzed by the academic community (**Carlson**, 2017).

Current development in the direction of the Internet of Things, a new generation of technologies that will bring connectivity to any electronic device, is a harbinger of the arrival of a new set of online media. In the last quarter century, online media have been limited to computers and mobile devices. However, in the near future, which, in some cases, is already here, online media will extend to devices such as voice assistants, wearables, home automation objects and a whole host of elements we use in our daily lives that we would currently find hard to imagine. News content will have to adapt to this emerging technological paradigm by overcoming the convergence and multiplatform concepts that professionals and academics have been using in recent years (**Doyle**, 2010; **Quandt**; **Singer**, 2009).

6. Language of digital journalism

Another fertile area of research on digital media focuses on narrative forms or modes of communication; that is, on the languages of digital journalism. In reality, it would be more appropriate to expand the frame and, as explained by Lev Manovich, talk about the language of new media in general. Indeed, in his influential book *The language of new media* (**Manovich**, 2002), this American professor presented a systematic theory of digital media at the beginning of this century, in which he connects them to visual media, most especially cinema. Manovich explained that, despite their novelty, digital media really make use of the narrative conventions of old media.



The language of new media (**Manovich**, 2002)

This debt with the past can be seen in the language of online media. Since the first steps of publications on the Internet, it was seen that their genres and formats were and continued to be directly inspired by nondigital news media. However, shortly afterwards, online media outlets began to exhibit characteristics of their own, which pointed to a gradual disconnection from the languages used by previous media. Time

has shown that this separation is slower and more gradual than expected. Even today, the genres and communication codes of many online media outlets continue to owe a direct debt to legacy media. It is clear that, 25 years after their birth, many online publications, or at least the most advanced ones, show specific forms of expression that have been the subject of many studies.

The first studies in this area focused on establishing the basic concepts. On this topic, many authors agreed on identifying three main traits in the language of digital journalism: the use of hypertext, multimedia and interactivity (**Nielsen**, 1995; **Salaverría**, 2005a, pp. 21-28).

The term "hypertext" was coined more than half a century ago by **Theodor H. Nelson** (1965) and, since the 1990s, the concept has given rise to a wealth of theoretical contributions regarding its application in journalism (**Huesca**; **Dervin**, 1999; **Díaz-Noci**; **Salaverría**, 2003, pp. 81-139). The ability to connect texts, understood in the broadest sense as units of information in any format, is considered to have been one of the fundamental bases of the language of digital journalism. Compared to the static information of the press, radio and television (which do not allow users to expand that initial information), the hypertext links in digital media have taught users how to actively access the news. The news is no longer just read, listened to or watched; it is also browsed. Linear reading of texts is now often replaced with nonlinear consultation.

In theory, the potential for hypertext development in journalism was seen as enormous (**Fredin**, 1997). However, empirical studies have shown that its real application in online media has been far more modest than expected and, in many cases, has been limited to repeating models inherited from pre-Internet journalism (**Himelboim**, 2010). Online journalism that is a slave to breaking news and easy clicks, and is more concerned with traffic than with quality has ended up turning its back on many of the narrative and documentary possibilities of hypertext, and barely uses it as a resource for increasing views. Furthermore, links have been observed to work as an additional gatekeeping resource: the media outlet not only decides what is news based on the sources and content it chooses to publish, but also on the links it includes in its news (**Dimitrova** *et al.*, 2003).

Multimedia is the second key element in the language of Internet journalism (**Guallar**; **Rovira**; **Ruiz**, 2010; **Salaverría**, 2014). Unlike hypertext, its real development in online media is much more advanced and diverse. The success of online video, a format which, according to different consulting firms, accounted for three quarters of global Internet traffic in 2018, has encouraged online publications to explore multimedia narratives. In this case, academic research has focused on identifying the types of multimedia packages in journalism (**Jacobson**, 2012) and their effect on news consumption by the public.

Closely related to multimedia is transmedia, another important concept that has impacted not only the academic world, but also professional practice. This idea became popular in 2003, when Henry Jenkins used it in an article published in *MIT*'s *Technology review*. In the article, he explained that we had entered

> "an era of media convergence that makes the flow of content across multiple media channels almost inevitable" (**Jenkins**, 2003).

Transmedia, as described by Jenkins and by the authors who followed him (**Scolari**, 2013b), refers to the trend of producing and consuming content in a supplemental manner through different channels. It is therefore a synthesis of the concepts of multimedia, convergence and participation, which results in cultural expressions that are perfectly suited to the culture of mixing and matching that is so successful on the Internet. This culture is most highly developed in expressions such as the fandom phenomenon, communities of film buffs and fans of cult fiction series. To a lesser extent, it has also appeared in journalism content (**Renó**, 2014; **Rost**; **Bernardi**; **Bergero**, 2016).

> "Some new stroytelling categories studied by academia include long-form journalism, slow journalism, immersive journalism, and newsgames"

The third major concept associated with the language of digital journalism is interactivity. As mentioned above, interactivity has been a key factor in the development of online journalism in recent years because it has driven participative or collaborative journalism. More specifically, interactivity is also an essential element of the language of digital journalism. And it is also one of the elements most studied by the academic community, even more so than multimedia. Many of these studies focus simply on identifying and analyzing the interactive resources used by news websites (**Stroud**; **Scacco**; **Curry**, 2016). Other studies, however, go further and explore the effects of these interactive qualities on how information is interpreted by the public (**Chung**; **Nah**, 2009).

The combination of all these ingredients has given rise to increasingly developed ways of presenting online news content. In the early years, digital media barely left room for anything other than simple news and certain classical journalism genres (interviews, reportage, chronicles, columns, etc.). However, over time, they explored much richer and more innovative ways of presenting information content. For example, feature articles have undergone a clear metamorphosis (**Larrondo**, 2009) through the development of multimedia resources and the exploration of novel hypertext and interactive formulas. Other genres, such as live news updates and interactive infographics (**Cairo**, 2012), have also contributed to renewing the genres and formats of digital journalism.

Besides the classical paradigm of journalism genres, new ways of classifying content have emerged from online media. These renewed categories order news information based on criteria such as its narrative features, its length and how it combines multimedia formats. Some of the most successful categories and those most studied by academia include

- long-form journalism (**Wolf**; **Godulla**, 2016; **Hiippala**, 2017),
- slow journalism (**Rosique-Cedillo**; Barranquero-Carretero, 2015),
- immersive journalism (**De-la-Peña** *et al.*, 2010; **Domínguez-Martín**, 2015) and
- newsgames (**Burton**, 2005; **Bogost**; **Ferrari**; **Schweizer**, 2012).

7. Economics of digital media

One of the most important points in the last 25 years in regard to digital media has been its economic impact on the communication industry and, by extension, on the long-term viability of quality journalism.

Economic uncertainty has become a primary concern, first and foremost for media outlet directors and professionals, who have been forced to deal with the radical transformation of their market. It has also been a concern for the academic community, which has almost exclusively analyzed this transformation by describing it rather than making proposals. Although some authors have made suggestions on how to solve the media's serious economic problem (**Vaccaro**; **Cohn**, 2004; **Cagé**, 2015), the vast majority of studies have merely quantified and analyzed the effects of this transformation (**Grueskin**; **Seave**; **Graves**, 2011).

In the 1980s and early 1990s, before the digital impact, the news media business was highly lucrative and relatively stable. Naturally, there were differences between different media. Television was unquestionably the most profitable format, thanks to its enormous income from advertising. It was followed by the press and radio, with more modest but nevertheless substantial levels of profitability. It is true that the publishing industry watched with concern as young people moved away from reading newspapers (**Smith**, 1980; **Bogart**, 1989), but it is also true that those same young people spent their time watching television, a medium that, in the end, often belonged to the same newspaper owners. In each journalism market, media outlets competed with their rivals, but did not feel threatened by external competitors. It is symptomatic that a book published in 1992, i.e., just before the start of the digital-journalism revolution, had the following title: *The Future of news: Television, newspapers, wire services, newsmagazines* (**Cook**; **Gomery**; **Lichty**, 1992). It is significant that online media were not even mentioned when listing the media of the future. Few people foresaw the breadth of the change that was about to take place. And from one day to the next, everything changed.

The future of news (**Cook**; **Gomery**; **Lichty**, 1992)

In reality, it was not even necessary to wait for the rise of today's Internet giants (e.g., *Google*, *Facebook*), which have cornered the advertising market since the 2000s. In the 1980s, companies had already appeared that foreshadowed this new external competition for journalism. The first commercial Internet providers began to connect homes to the Internet and, with that, they became the first platforms providing citizens with access to the incipient digital media. Although their focus was not on journalism, companies like *AOL*, *Compuserve* and *Prodigy* became the first online providers of news content in the late 1980s. Few were aware of it at the time, since the World Wide Web had not yet been invented, but it was the first step toward a change that would destabilize the entire news media industry.

Although the first economic studies of digital media found a limited substitution effect with respect to traditional media (**Bromley**; **Bowles**, 1995), it was not long before other studies began to warn of the power of the Internet to disrupt the media business (**Chyi**; **Sylvie**, 1998). The change that had started to take place in information-consumption habits was, in fact, a serious threat to the future of news companies. Digital transformation had made journalism content so accessible and omnipresent that it virtually lost all value. Information had become a commodity, a ubiquitous raw material with barely any added value.

> "In 1990s, few researchers foresaw the breadth of the change that was about to take place to media business"

The tremendous economic success of *Google*, thanks to its *AdWords* service, a search-based advertising platform, which was launched in 2000, was a major step in the decline of the media business. Whereas before, a large part of the audience had begun to migrate to the Internet attracted by the growing range of free content, the advertising model invented by *Google* meant that the media also began to lose their advertisers. News companies entered a death spiral and continue falling today: traditional publications lost readers and, although business increased through their respective online editions, it did so with very little financial return. From that moment, the problem with the media's business model became a priority matter, not only for media executives but also for academic researchers.

Since the turn of the 21st century, a plethora of studies, monographs and reports has been published analyzing the transformation of the media business. One of the pioneers and most influential researchers in this discipline is the American Robert G. Picard. He is the author and director of many studies on the economic impact of the Internet on the media and was one of the first to explore the "emerging business models" (**Picard**, 2000, p. 66) of digital media. This type of study began to show not only the erosion suffered by the traditional business of news companies, but, more profoundly, the keys to the radical change in the rules of the business. Similar to the fate of the publishing, cinema and music industries (**Vaccaro**; **Cohn**, 2004), the news business saw its traditional models collapse while new opportunities arose for new actors with innovative ideas. News media are facing an ambivalent reality. So far, many outlets have lost, but some are beginning to win thanks to their innovative strategies; in the words of Cham-Olmsted,

> "the Internet provides an alternative distribution channel for traditional media products and strengthens the position of the existing media with their readership and audience, while competing with legacy media for the consumer's attention and resources." (**Cham-Olmsted**, 2004, p. 2).

Finding business models that are sustainable over the long term has become a key challenge for the media (**Van-der-Wurff**, 2012). Different studies have examined the sources of funding that the media have begun to apply in their search for that rare alchemy that is profitability (**Kaye**; **Quinn**, 2010; **Picard**, 2011), with special attention to paid content (**Herbert**; **Thurman**, 2007). To the traditional sources of advertising, sales and subscriptions, online media have added an increasingly diverse catalogue of revenue sources (**Cea-Esteruelas**, 2013) such as sponsored content, event organization, sales intermediation and online service provision. The combination of these traditional and modern sources has given rise to an ever-lengthening list of business models: free (advertiser-supported), paid or subscription-based (paywall) (**Myllylahti**, 2014), soft paywall (metered paywall), membership-based, mixed (freemium), donation-based (crowdfunding) (**Casero-Ripollés**, 2010; **Cerezo**, 2017; **Palacio**, 2018).

In light of this diversity, one of the emerging areas of research in recent years is the comparison of the business models of native digital media and online media derived from traditional markets (**Arrese**; **Kaufmann**, 2016).

8. Other topics and disciplines

As we have seen, research on digital news media has been carried out in many areas. Each of these areas also has its own branches, which makes the whole field so complex that it is impossible to cover fully in an article of this nature. Nevertheless, before looking at other theoretical, methodological and general matters, it is necessary to provide a concise description of some other research topics and disciplines that the academic community has shown particular interest in over the years. I will highlight four of them:

1) the changing technological context,

2) multimedia convergence and the evolution of journalists' professional routines,

3) innovation models in news organizations and

4) research on new audiences of online media.

Given that digital journalism arose thanks to a technological leap, it is not surprising that this has been one of the areas researchers have paid the most attention to. Understanding online journalism required understanding its changing technologies. The Internet not only led to the evolution of the tools of journalism, but also produced a comprehensive change in its paradigms (**Orihuela**, 2004).

Research on the technologies of digital journalism have been divided into two main categories: studies of journalistic production and studies of news consumption.

Studies on production have analyzed the way in which digital technologies have changed journalists' work habits and environments. Digital tools quickly showed their ability to transform work processes in newsrooms, but not always for the better (**Cottle**; **Ashton**, 1999), and forced journalists to master multiple skills (**García-Avilés** *et al.*, 2004). There is a well-founded concern regarding the increasingly difficult working conditions of journalists. The harm this does to the quality of published information is also of concern. Nevertheless, it cannot be denied that digital tools have also had a positive side: they have broadened the media's range of activities and coverage. Disciplines such as data-driven journalism, multimedia news storytelling and advanced analysis of audience metrics would be impossible without the digital production technologies that have been developed in the last quarter century.

> "Mobile journalism is one of the fastest-growing areas of research in the 2010s"

Studies on digital information-consumption technologies have analyzed the specific features and opportunities of the devices launched on the market during this time. Starting in the 1980s, the first studies on digital journalism technologies analyzed the potential for developing platforms such as videotex (**Carey**, 1982; **Heikkinen**; **Reese**, 1986) and teletext

(**Henke**; **Donohue**, 1986) and others. With the arrival of the Internet and especially the popularization of the first online media outlets, the focus of research on technology shifted to computers. Topics such as the strengths and weaknesses of reading on-screen began to be studied through the use of experimental tools such as eye-tracking devices (**Outing**, 2004).

The launch of the *iPhone* by *Apple* in 2007 was the next technological leap for news consumption. For some years before that, different mobile phones had made it possible to consult journalism content using primitive systems such as WAP (Wireless application protocol). However, the *iPhone* revealed a new form of multimedia interaction with content to its millions of buyers throughout the world. In addition to traditional text, new formats were now available at all times such as photography and, especially, audio and video. Other technology companies soon brought their own alternative smartphones to market, while telecommunications companies drove the connectivity of these devices with rapid progression from the most rudimentary standards [2G (GPRS) and 3G (UMTS)] to more advanced standards [4G (LTE)] and, at the time of writing, 5G. At the same time, there was a proliferation of devices for reading electronic books, known as e-readers. Although some newspapers developed editions adapted for these devices, the initiative was soon shelved, particularly as a result of the launch of the *iPad* by *Apple* in 2010. Tablets became another successful way of consuming information on the go. We all know the result of that process: citizens enthusiastically adopted smartphones (in different sizes) and now spend hours and hours using them every day.

The technological leap from the personal computer to the mobile device opened the way for mobile journalism or mojo, one of the fastest-growing areas of research on digital journalism in the 2010s (**Westlund**, 2013). Much of this research on mobile communication and on other technologies is based on uses and gratifications theory in an attempt to use it to explain why people replace certain devices with others launched on the market.

Besides strictly technological research, another area of interest to researchers is multimedia convergence and the evolution of journalists' professional routines. With a focus taken from the sociology of organizations, these studies have explored the internal change undergone by the media and media professionals: how work spaces, journalist profiles, decision-making dynamics and the relationship with sources and the public have changed, as well as other topics. A large part of these studies has focused on the concept of convergence, with special attention to the change in work dynamics in newsrooms and an examination of the difficult coexistence (often mere cohabitation) of traditional journalists and digital journalists (**Boczkowski**, 2005; **Quinn**, 2005; **Salaverría**; **Negredo**, 2008).

Another important area of research that is linked to these phenomena of multimedia convergence and the integration of newsrooms focuses on innovation models in journalism organizations (**Gynnild**, 2014). As discussed, media outlets have faced a technological, social and economic scenario of abrupt changes and have had little time to react. In order to meet these challenges and take advantage of the opportunities of the new market, some media outlets have adopted innovation strategies of different kinds. The most common ones have focused on product innovation by experimenting with multimedia narratives and thinking up new content and services. Others have focused on internally restructuring the company with the aim of promoting new projects and keeping the organization up to date with market innovations. Finally, much innovation has been based on distribution and sales systems, driven by the new devices and platforms launched by technology companies. Although there are differences between countries, it has been shown that the media organizations most inclined toward innovation have not exactly been the big ones; in Spain at least, more innovation has been seen in small online media organizations, particularly native digital media outlets (**García-Avilés** *et al.*, 2018). It has happened in many other fields and journalism is no different: the weakest have been forced to embrace the fact that necessity is the mother of invention (**Sádaba-Chalezquer**, 2016).

Another research topic focuses on investigating the audiences of digital media. Since the beginning, online media outlets displayed a unique ability to attract an audience that was different from the audience of the press, radio and television. In the early years of the online press, in the 1990s, that almost experimental audience consisted of people with mid-level or advanced computer skills and, especially, young people. But this situation soon evolved. As digital media became established, their audiences diversified so that, after the first 25 years, they now constitute an undeniably intergenerational journalistic phenomenon. The fact that digital media outlets continue to attract more young people does not mean that they are of no interest to adult audiences. Instead, it highlights the increasing difficulty of legacy media (including television) to attract young people. Today, digital media attract young and old in equal measure, whereas legacy media are consumed by an increasingly older audience.

> "If the overhaul of research methodologies has been intense, a more modest adjective would be needed to describe the overhaul of theories"

There is a wealth of studies on the profile, distribution, interests and habits of digital media news consumers. Many of these studies and reports have a more professional than academic focus, as they are often aimed at providing strategic

http://www.pewresearch.org/topics/state-of-the-news-media

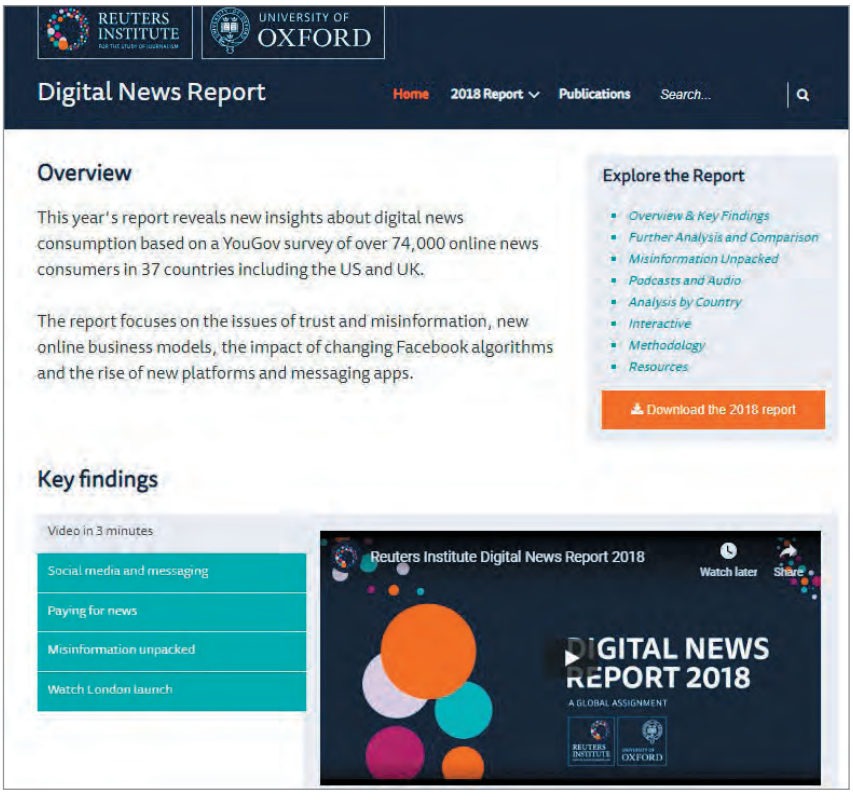
references for decision-making by news companies. One of the most highly respected reports is the *State of the news media* produced every year since 2004 by the *Pew Research Center*.
http://www.pewresearch.org/topics/state-of-the-news-media

The report analyzes the profiles of the media and their audiences in the United States. Together with this type of market report, there is also a long list of purely academic studies that use quantitative and qualitative methodologies to examine the transformation of the news-consuming public and the novel ways in which it behaves on the Internet (see, for example, **Tewksbury**, 2005; **Masip** *et al.*, 2015). They also study such topics as the impact of audience metrics on journalists and specifically their effect on how journalists select, write and hierarchize the news (**MacGregor**, 2007).

In the area of audience, the broadest and most prominent study on news consumption on the Internet and mobile devices is, without doubt, the *Digital news report.*
http://www.digitalnewsreport.org

This study has been produced annually since 2014 by the *Reuters Institute for the Study of Journalism*, *University of Oxford*, with the contribution of a dozen other universities in other countries. In its 2018 edition, the latest edition at the time of writing, the report made a comparative analysis of the consumption of online news in 37 countries, based on a survey of more than 74,000 Internet users.

As we have seen, research on digital journalism has given rise to many developments and areas of interest. Besides the disciplines mentioned, other studies focus on journalism ethics in the digital setting, factors related to selective attention to news (filter bubbles and echo chambers) and, particularly in recent years, the serious threat of deliberate disinformation over online networks — the much-vaunted fake news. Indeed, there is much that this article has not covered. But what has been discussed is sufficient to show the undeniable diversity and notable maturity of research on digital journalism.



http://www.digitalnewsreport.org

9. New methods, old theories

So far, we have looked at some of the main areas on which research on digital journalism has focused in recent years and mentioned the areas researchers have paid the most attention to, with a summary of their main contributions. This enormous corpus of academic research in the field of the social

sciences would have little validity if it had not been carried out with due methodological thoroughness. The following lines will deal with this: What are the characteristics of research on digital journalism and online media from the point of view of the methodologies used?

This is an essential matter for the research community and, in the last ten years, it has begun to appear as the subject of monographic studies (**Palacios**; **Díaz-Noci**, 2009; **Karlsson**; **Sjøvaag**, 2016; 2019). In the last 25 years, researchers have not only been faced with the challenge of exploring a growing reality (that of digital news media outlets), but have also been forced to design and apply suitable study methods and tools.

Some methodologies, especially those based on observation, have not undergone a major evolution compared to how they were applied previously with nondigital media. For example, the many studies published during these years on professional routines in the newsrooms of digital media (see **Boczkowski**, 2004; **Domingo**; **Paterson,** 2011, among others) based on ethnographic methodologies are not substantially different from those that were carried out (and are still carried out today) on nondigital media. Ethnographic research has given rise to specific methodologies on the Internet, as part of the emerging concept of netnography (**Kozinets**, 2015), but the reality is that most ethnographic studies on the media are still anchored in traditional research models.

Among qualitative research methodologies, studies based on in-depth interviews have also not varied substantially. The application of this research method essentially follows the models that have been applied for decades in many social science studies. Nevertheless, subsequent analysis of the content of these interviews has begun to benefit from much more sophisticated tools. In recent years, researchers in the social sciences have incorporated different analytical applications that have been brought to market (*NVivo*, *ATLAS.ti* and similar tools), which facilitate the exploration of linguistic frequencies in interview transcriptions and help identify relationships between concepts. These computer programs make it possible to analyze those aspects not only in texts, but also in many different multimedia elements, which makes them especially useful for researchers studying digital news media outlets. These applications are able to translate the qualitative raw material, which is often unstructured, into well-ordered quantitative magnitudes that can be analyzed and represented graphically.

These same tools, together with others that can analyze large volumes of data (*R*, *Python*, etc.), have given a major boost to another of the principal research methodologies in digital journalism: content analysis (**Flaounas** *et al.*, 2013; **Lewis**; **Zamith**; **Hermida**, 2013). Before these tools became available, research based on content analysis was limited to small samples, as it was often impossible to analyze very large populations or datasets. Today, thanks to content digitization and to these powerful computational calculation tools, research is carried out on data corpuses that would have been unthinkable barely ten years ago.

> Digital media and their community of researchers have lived parallel lives: from being almost nobodies, they have become the undisputed protagonists

Perhaps the most telling example of this quantitative leap (in this case, the term quantitative is more appropriate than qualitative) is the increasingly large number of studies on journalism and social media (**Newman**, 2009; **Hermida** *et al.*, 2012). Owing to the nature of these networks, made up of millions of users spread all over the world, studying them involves working with huge volumes of content in the realm of big data. Without powerful analysis and visualization tools, analyzing these data would be impossible. However, the computational applications developed in recent years make it possible to analyze massive datasets, not only using traditional statistical procedures, but also by applying analysis models based on machine-learning technology (**Bouwmans**; **Trilling**, 2016).

If the overhaul of research methodologies has been intense, a more modest adjective would be needed to describe the overhaul of theories. These first 25 years of digital media have not been particularly fertile in terms of the creation of new theories for the general interpretation of journalism. The ideas of the great thinkers of the 20th century in social communication, journalism and public opinion (authors as diverse as Walter Lippmann, Marshall McLuhan, Paul Lazarsfeld, Harold Lasswell, Umberto Eco, Jürgen Habermas and others) essentially continue to form the theoretical interpretation framework that holds up a large part of the research on the journalism of our times. In the words of **Mitchelstein** and **Boczkowski** (2009, p. 575),

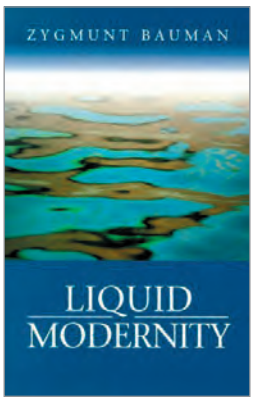
> "most studies continue to use existing lenses to look at new phenomena."

Although the ideas of those authors continue to determine the theoretical framework, this does not mean that they have not been revised. On the contrary, the flood of empirical studies in the last thirty years, presented at innumerable conferences and in an endless list of articles in the growing number of scientific journals on communication, has helped extend the scope of these theories to include new settings.

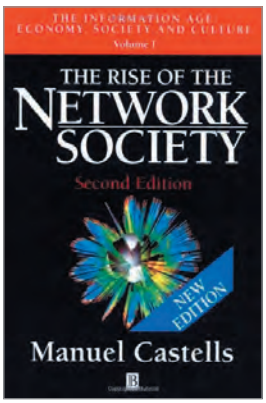
To better explain this idea, it may be useful to look at an example, in this case, the agenda-setting theory. This popular theory was put forward almost half a century ago by the American professors **Maxwell E. McCombs** and **Donald L. Shaw** (1972) and explains how the media determine not only the opinions of citizens but, especially, the topics of debate

among public opinion. The first studies to test the agenda-setting theory, which were based mainly on content-analysis methodologies, focused on information published by newspapers. Later, in the 1980s and 1990s, these studies spread to audiovisual media, particularly television. Since the 2000s, we have witnessed a new step forward for this theory; in this case, in the context of online media and social media. The most recent studies on influential persons on social media, known as influencers (**Meraz**, 2009), are really just another link in the chain of empirical research based on this theory.

Convergence culture (**Jenkings**, 2006)



Liquid modernity (**Bauman**, 2013)



The rise of the network society (**Castells**, 1996)

Like the agenda-setting theory, other classical theories in journalism (framing theory, uses and gratifications theory, information theory and many others) are being revisited and, in some cases, revised by contemporary researchers on digital journalism.

Of course, not all research is set in the context of these theories. The phenomenon of the Internet already has its own thinkers, some of whom are still emerging, while others are firmly established. The second group includes authors such as the sociologist **Manuel Castells** (1996), thanks to his influential theory on the network society, the philosopher and sociologist **Zygmunt Bauman** (2013), with his theory of liquid modernity, and **Henry Jenkins** (2006), mentioned above, for his theory of convergence and transmedia culture.

10. Conclusion and pending challenges

The first 25 years of research on digital journalism have been prolific. A discipline that had a humble birth in the shadow of more classical areas of media research has forged ahead and now constitutes the most dynamic and vigorous area of research on journalism. In reality, digital media and their community of researchers have lived parallel lives: from being almost nobodies, they have become the undisputed protagonists.

Like digital journalists, researchers on this topic have come up with their own concepts and methods for dealing with a reality in rapid transformation. The digital journalism academic community took some pre-existing theories and tools and adapted them while it developed new tools, with which it has studied hitherto unknown realities and innovative phenomena.

The general conclusion is positive: a fertile and diverse field of research has been established with multiple areas and increasingly more stable research methodologies. Studies on digital news media boast firmly established scientific and professional conferences (Austin, Cardiff, Perugia, Huesca) and respected specialist journals (*New media and society*, *Journal of computer-mediated communication*, *Digital journalism*). Other high-impact journals such as this one (*El profesional de la información*) periodically issue monographs and special editions to digital journalism. Leading publishers have also released collections of books that focus on digital media, based on which they regularly publish collected works, monographs and manuals. The main international communication research associations (*ICA*, *IAMCR*, *AEJMC*, *ECREA*) have divisions and working groups that deal with phenomena and trends in digital media from different perspectives. Numerous national and international research projects are dedicated to exploring this discipline, which encourages the creation and strengthening of specialist and interdisciplinary research teams. Regardless of the elements considered, it cannot be denied that research on digital journalism is now a firmly established discipline.

> More than data and diagnostics, journalism as a profession demands ideas and solutions

But it still faces challenges. As I mentioned some years ago in this journal (**Salaverría**, 2015), there are some aspects in which research on digital media needs an overhaul. At that time, I mentioned five aspects:

1) moving forward on native digital research, understood as research that goes beyond comparative studies with other platforms and focuses exclusively on studying digital media;

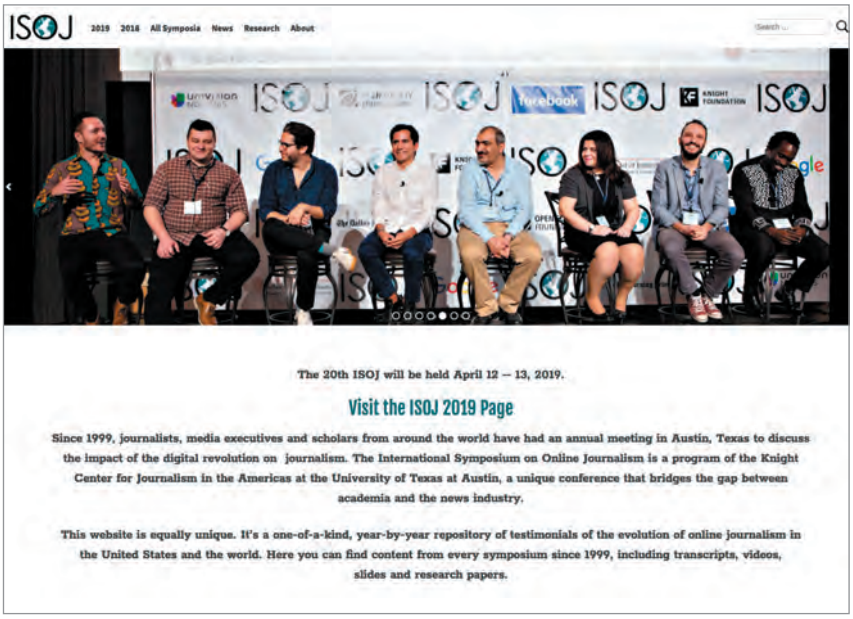
2) using advanced research technology;

3) committing to innovation-oriented research;

4) strengthening the analysis of phenomena by providing more than simple case descriptions; and

5) opening up research to topics and areas that receive little attention.

In the few years since those thoughts were published, some steps have been taken that I consider to be in the right direction.



International Symposium on Online Journalism (*The University of Texas at Austin*)
https://isoj.org

Indeed, research focusing on purely digital media is gaining ground. An increasing number of studies and projects are looking at publications created for the Internet, which are given the same treatment (or almost the same) as that given to traditional journalistic brands. The phenomenon of social media, which has exploded in recent years, has contributed to this evolution in interests, given that a large part of the most important journalistic phenomena worthy of study today occurs outside traditional media. As native digital media move rapidly toward becoming firmly established and many traditional names in journalism continue their decline, a myriad of studies can be expected in coming years focusing on purely Internet-based media. Who knows whether this will lead to a phenomenon like the one triggered 25 years ago: at that time, the focus of academic research shifted from nondigital media to digital media. Perhaps in the coming years, we will see a new shift, in this case, from non-native digital media to native digital media.

There can also be no doubt that the use of advanced research methodologies and technologies has grown in recent years. Big data has ceased to be merely a subject for study for journalism academics and has become an everyday work tool. Quantitative studies proliferate in scientific journals on digital journalism, thus reducing the space dedicated to exclusively qualitative research. Now, data accuracy and size are as important as the shrewdness applied in data interpretation. This evolution has a positive side, as it reinforces the scientific credibility of studies, which are increasingly carried out on complete populations or highly representative samples. However, this extreme empiricism also has its drawbacks: among the avalanche of individual studies, there is a need for more holistic analyses that can deal with the whole and not just the parts. Perhaps for this reason so much progress has been made in recent years in empirical research, but not as much has been made in the theories and interpretative models.

Another pending task is improving the transfer of results from the scholar community to the professional world in order to spread sorely needed innovation. In the academic community, there are those who, with good reason, state that universities should not be at the service of industry. But they should not turn their backs on it either. Unfortunately, that is what tends to happen. In studies on digital journalism, there is a dearth of academic contributions that are of use for solving the very serious problems of the media and journalists. After the disappointing evolution of the profession of journalism in recent years, it would seem justified for researchers to adopt a new approach by not simply observing, but making any proposals they can. More than data and diagnostics, journalism as a profession demands ideas and solutions.

Digital media have completed a key phase in their history: the initial stage. It has been 25 years of triumphs and failures, of trial and error. The academic community has followed these steps closely and analyzed their phenomena and processes. On a global scale, a research discipline has become firmly established and boasts shared concepts, tools and methods. The next quarter century will likely see major technological, social and professional innovations that will continue the profound transformation of journalism, quite possibly at an even faster pace than in recent years. It will be interesting to see how academic research responds to this challenge.

11. References

Allan, Stuart; **Thorsen, Einar** (eds.) (2009). *Citizen journalism: Global perspectives*. New York: Peter Lang. ISBN: 978 1 433102950

Antheaume, Alice (2013). *Le journalisme numérique*. Paris: Presses de Sciences Po. ISBN: 978 2 724612646

Arrese, Ángel; **Kaufmann, Jürg** (2016). "Legacy and native news brands online: do they show different news consumption patterns?". *International journal on media management*, v. 18, n. 2, pp. 75-97.
https://doi.org/10.1080/14241277.2016.1200581

Aumente, Jerome (1987). *New electronic pathways: Videotex, teletext and online databases*. Newbury Park, CA: Sage Publications. ISBN: 978 0 803927841

Bagdikian, Ben H. (1971). *The information machines. Their impact on men and the media.* New York: Harper & Row. ISBN: 978 0 060902582

Banegas-Flores, Cecilia (coord.); **Barba-Sanjínez, Iván** (ed.) (2014). *Periodismo digital en Bolivia 2012-2013*. La Paz: Fundación UNIR. ISBN: 978 9995496050
https://www.academia.edu/9985683/Periodismo_Digital_en_Bolivia_2012-2013

Bardoel, Jo; **Deuze, Mark** (2001). "Network journalism: converging competences of media professionals and professionalism". *Australian journalism review*, v. 23, n. 2, pp. 91-103.
https://scholarworks.iu.edu/dspace/bitstream/handle/2022/3201/BardoelDeuze+NetworkJournalism+2001.pdf?sequence=1

Bastos, Helder (2010). *Origens e evolução do ciberjornalismo em Portugal: Os primeiros quinze anos (1995-2010)*. Oporto: Edições Afrontamento. ISBN: 978 9723611298

Bauman, Zygmunt (2013). *Liquid modernity*. New York: John Wiley & Sons. ISBN: 978 0 745674414

Bergonzi, Juan-Carlos; **Rost, Alejandro**; **Bergero, Fabián**; **Bernardi, María-Teresa**; **García, Viviana**; **Pugni-Reta, María-Emilia**; **Apesteguía, Ezequiel** (2008). *Periodismo digital en la Argentina*. Río Negro: Publifadecs. ISBN: 978 987 1549054
https://www.researchgate.net/publication/272168151_Periodismo_digital_en_la_Argentina

Boczkowski, Pablo J. (2004). "The processes of adopting multimedia and interactivity in three online newsrooms". *Journal of communication*, v. 54, n. 2, pp. 197-213.
https://doi.org/10.1111/j.1460-2466.2004.tb02624.x

Boczkowski, Pablo J. (2005). *Digitizing the news: Innovation in online newspapers*. Cambridge, MA: MIT Press. ISBN: 0 262 02559 0

Bogart, Leo (1989). *Press and public: Who reads what, when, where, and why in American newspapers*. Hillsdale, NJ: Lawrence Erlbaum. ISBN: 978 0 805804317

Bogost, Ian; **Ferrari, Simon**; **Schweizer, Bobby** (2012). *Newsgames: Journalism at play*. Cambridge, MA: MIT Press. ISBN: 978 0 262518079

Bohn, Thomas W. (1980). "Broadcasting national election returns, 1952-1976". *Journal of communication*, v. 30, n. 4, pp. 140-153.
https://doi.org/10.1111/j.1460-2466.1980.tb02026.x

Bolter, Jay-David; **Grusin, Richard** (2000). *Remediation: Understanding new media*. Cambridge, MA: MIT Press. ISBN: 978 0 262522793
https://monoskop.org/images/a/ae/Bolter_Jay_David_Grusin_Richard_Remediation_Understanding_New_Media_low_quality.pdf

Bonington, Paul (1995). "The fourth media". *Internet world*, v. 6, n. 4, p. 6.

Bosch, Tanja (2010). "Digital journalism and online public spheres in South Africa". *Communicatio: South African journal for communication theory and research*, v. 36, n. 2, pp. 265-275.
https://doi.org/10.1080/02500167.2010.485374

Boulianne, Shelley (2018). "Twenty years of digital media effects on civic and political participation". *Communication research*, 1-20.
https://doi.org/10.1177/0093650218808186

Boumans, Jelle W.; **Trilling, Damian** (2016). "Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars". *Digital journalism*, v. 4, n. 1, pp. 8-23.
https://doi.org/10.1080/21670811.2015.1096598

Bowman, Shayne; **Willis, Chris** (2003). *We media. How audiences are shaping the future of news and information*. Reston, VA: The American Press Institute.
http://www.hypergene.net/wemedia

Bravo-Herrera, Vanessa (2003). *The Costa Rican press on the internet: The first ten years* [tesis de maestría]. Gainesville, FL: University of Florida.

Brennen, Bonnie; **Dela-Cerna, Erika** (2010). "Journalism in Second Life". *Journalism studies*, v. 11, n. 4, pp. 546-554.
https://epublications.marquette.edu/cgi/viewcontent.cgi?article=1065&context=comm_fac
https://doi.org/10.1080/14616701003638418

Brinca, Pedro (2006). "Setúbal na rede: o caso de um projecto pioneiro". *Comunicação e sociedade*, v. 9, n. 10, pp. 147-150.
https://doi.org/10.17231/comsoc.9(2006).2458

Broersma, Marcel J.; **Peters, Chris** (2013). "Introduction. Rethinking journalism: The structural transformation of a public good". In: Peters, Chris; Broersma, Marcel J. (eds.). *Rethinking journalism: Trust and participation in a transformed news landscape*. Oxon: Routledge, pp. 1-12. ISBN: 978 0 415697019

Bromley, Rebekah V.; **Bowles, Dorothy** (1995). "Impact of internet on use of traditional news media". *Newspaper research journal*, v. 16, n. 2, pp. 14-27.
https://doi.org/10.1177/073953299501600203

Bruns, Axel (2005). *Gatewatching: Collaborative online news production*. New York: Peter Lang. ISBN: 978 0 820474328

Burton, Jonathon (2005). "News-game journalism: History, current use and possible futures". *Australian journal of emerging technologies and society*, v. 3, n. 2, pp. 87-99.
http://pandora.nla.gov.au/pan/36357/20061206-0000/www.swinburne.edu.au/sbs/ajets/journal/V3N2/V3N2-3-Burton.html

Cabrero-Rubio, Cristina (2005). *Periodismo digital en Bolivia*. La Paz: Plural Editores. ISBN: 978 9990575989

Caffarel-Serra, Carmen; **Ortega-Mohedano, Félix**; **Gaitán-Moya, Juan-Antonio** (2017). "Investigación en comunicación en la universidad española en el período 2007-2014". *El profesional de la información*, v. 26, n. 2, pp. 218-227.
https://doi.org/10.3145/epi.2017.mar.08

Cagé, Julia (2015). *Sauver les médias. Capitalisme, financement participatif et démocratie*. Paris: Seuil. ISBN: 978 2 021219555

Cairo, Alberto (2013). *The functional art: An introduction to information graphics and visualization*. Berkeley, CA: New Riders. ISBN: 978 0 321834737

Canavilhas, João-Messias (2003). "Webjornalismo: considerações gerais sobre jornalismo na web". In: Fidalgo, António; Serra, Joaquim-Paulo (orgs.). *Informação e comunicação online (vol. I): Jornalismo online*. Covilhã: Universidade da Beira Interior, pp. 63-73. ISBN: 9728790074
http://www.bocc.ubi.pt/pag/canavilhas-joao-webjornal.pdf

Carey, John (1982). "Videotex: The past as prologue". *Journal of communication*, v. 32, n. 2, pp. 80-87.
https://doi.org/10.1111/j.1460-2466.1982.tb00498.x

Carlson, David (2003). "The history of online journalism". In: Kawamoto, Kevin (ed.). *Digital journalism: Emerging media and the changing horizons of journalism*. Lanham, MD: Rowman & Littlefield Publishers, p. 31-55. ISBN: 0 7425 2681 X

Carlson, Matt (2018). "Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism". *New media & society*, v. 20, n. 5, pp. 1755-1772.
https://doi.org/10.1177/1461444817706684

Casero-Ripollés, Andreu (2010). "Prensa en internet: nuevos modelos de negocio en el escenario de la convergencia". *El profesional de la información*, v. 19, n. 6, pp. 595-601.
https://doi.org/10.3145/epi.2010.nov.05

Caspi, Dan (2011). "A revised look at online journalism in Israel: Entrenching the old hegemony". *Israel affairs*, v. 17, n. 3, pp. 341-363.
https://doi.org/10.1080/13537121.2011.584663

Castells, Manuel (1996). *The information age: Economy, society and culture. Vol. I, The rise of the network society*. Cambridge, MA: Blackwell. ISBN: 978 1 557866165
https://deterritorialinvestigations.files.wordpress.com/2015/03/manuel_castells_the_rise_of_the_network_societybookfi-org.pdf

Castells, Manuel (2007). "Communication, power and counter-power in the network society". *International journal of communication*, v. 1, n. 1, pp. 238-266.
https://ijoc.org/index.php/ijoc/article/view/46

Castillo, Antonio; **Carretón, Carmen** (2010). "Investigación en comunicación: Estudio bibliométrico de las revistas de comunicación en España". *Communication & society*, v. 23, n. 2, pp. 289-327.
https://www.unav.es/fcom/communication-society/es/resumen.php?art_id=372

Cea-Esteruelas, María-Nereida (2013). "Economía de los cibermedios: modelo de ingresos y fuentes de financiación". *El profesional de la información*, v. 22, n. 4, pp. 353-361.
https://doi.org/10.3145/epi.2013.jul.12

Cerezo, Pepe (2017). *El auge de los modelos de pago. Trump y el Brexit, el efecto de los fenómenos altamente imprevistos en la prensa*. White Paper Evoca, marzo.
http://evocaimagen.com/dosieres/dosier-evoca-04-paywalls.pdf

Chan-Olmsted, Sylvia M. (2004). "Introduction: Traditional media and the internet: The search for viable business models". *International journal on media management*, v. 6, n. 1-2, pp. 2-3.
https://doi.org/10.1080/14241277.2004.9669376

Charon, Jean-Marie; **Le-Floch, Patrick** (2011). *La presse en ligne*. Paris: La Découverte. ISBN: 978 2 707157744

Chung, Deborah S.; **Nah, Seungahn** (2009). "The effects of interactive news presentation on perceived user satisfaction of online community newspapers". *Journal of computer-mediated communication*, v. 14, n. 4, pp. 855-874.
https://academic.oup.com/jcmc/article/14/4/855/4583549

Chyi, Hsiang-Iris; **Sylvie, George** (1998). "Competing with whom? Where? And how? A structural analysis of the electronic newspaper market". *Journal of media economics*, v. 11, n. 2, pp. 1-18.
https://doi.org/10.1207/s15327736me1102_1

Chyi, Hsiang-Iris; **Sylvie, George** (2000). "Online newspapers in the U.S.: Perceptions of markets, products, revenue, and competition". *International journal on media management*, v. 2, n. 2, pp. 69-77.
https://doi.org/10.1080/14241270009389924

Cook, Philip S.; **Gomery, Douglas**; **Lichty, Lawrence W.** (eds.) (1992). *The future of news: Television, newspapers, wire services, newsmagazines*. Washington: Woodrow Wilson Center Press. ISBN: 978 0 943875347

Cottle, Simon; **Ashton, Mark** (1999). "From BBC newsroom to BBC newscentre: On changing technology and journalist practices". *Convergence*, v. 5, n. 3, pp. 22-43.
https://doi.org/10.1177/135485659900500304

Crovi-Druetta, Delia-María; **Toussaint, Florence**; **Tovar, Aurora** (2006). *Periodismo digital en México*. México: UNAM y Sitesa Ediciones. ISBN: 978 970 6291899

De-la-Peña, Nonny; **Weil, Peggy**; **Llobera, Joan**; **Giannopoulos, Elias**; **Pomés, Ausiàs**; **Spanlang, Bernhard**; **Friedman, Doron**; **Sánchez-Vives, Maria-V.**; **Slater, Mel** (2010). "Immersive journalism: immersive virtual reality for the first-person experience of news". *Presence: Teleoperators and virtual environments*, v. 19, n. 4, pp. 291-301.
https://doi.org/10.1162/PRES_a_00005

Deuze, Mark (2001). "Online journalism: Modelling the first generation of news media on the World Wide Web". *First Monday*, v. 6, n. 10.
https://doi.org/10.5210/fm.v6i10.893

Deuze, Mark (2004). "What is multimedia journalism?". *Journalism studies*, v. 5, n. 2, pp. 139-152.
https://www.researchgate.net/publication/238688450_What_is_multimedia_journalism1
https://doi.org/10.1080/1461670042000211131

Deuze, Mark (2005). "What is journalism? Professional identity and ideology of journalists reconsidered". *Journalism*, v. 6, n. 4, pp. 442-464.
https://doi.org/10.1177/1464884905056815

Díaz-Noci, Javier (2005). "Historia de los cibermedios en España". In: Salaverría, Ramón (coord.). *Cibermedios. El impacto de internet en los medios de comunicación en España*. Sevilla: Comunicación Social Ediciones y Publicaciones, pp. 21-38. ISBN: 978 84 96082 33 5

Díaz-Noci, Javier (2013). "A history of journalism on the internet: A state of the art and some methodological trends". *Revista internacional de historia de la comunicación*, v. 1, n. 1, pp. 253-272.
https://doi.org/10.12795/RiHC.2013.i01.12

Díaz-Noci, Javier; **Meso-Ayerdi, Koldo** (1998). "Desarrollo del periodismo electrónico". *El profesional de la información*, v. 7, n. 12, pp. 4-11.
http://www.elprofesionaldelainformacion.com/contenidos/1998/diciembre/desarrollo_del_periodismo_electronico.html

Díaz-Noci, Javier; **Salaverría, Ramón** (coords.) (2003). *Manual de redacción ciberperiodística*. Barcelona: Ariel. ISBN: 84 344 1297 7

Dimitrova, Daniela V.; **Connolly-Ahern, Colleen**; **Williams, Andrew-Paul**; **Kaid, Lynda-Lee**; **Reid, Amanda** (2003). "Hyperlinking as gatekeeping: Online newspaper coverage of the execution of an American terrorist". *Journalism studies*, v. 4, n. 3, pp. 401-414.
https://doi.org/10.1080/14616700306488

Domingo, David; **Paterson, Chris** (eds.) (2011). *Making online news (Volume 2). Newsroom ethnography in the second decade of internet journalism*. New York: Peter Lang ISBN: 978 1 4331 1064 1

Domingo, David; **Quandt, Thorsten**; **Heinonen, Ari**; **Paulussen, Steve**; **Singer, Jane B.**; **Vujnovic, Marina** (2008). "Participatory journalism practices in the media and beyond: An international comparative study of initiatives in online newspapers". *Journalism practice*, v. 2, n. 3, pp. 326-342.
http://openaccess.city.ac.uk/3457
https://doi.org/10.1080/17512780802281065

Domínguez-Martín, Eva (2015). "Periodismo inmersivo o cómo la realidad virtual y el videojuego influyen en la interfaz e interactividad del relato de actualidad". *El profesional de la información*, v. 24, n. 4, pp. 413-423.
https://doi.org/10.3145/epi.2015.jul.08

Doyle, Gillian (2010). "From television to multi-platform: Less from more or more for less?". *Convergence*, v. 16, n. 4, pp. 431-449.
http://www.oecd.org/tad/services-trade/47559455.pdf
https://doi.org/10.1177/1354856510375145

Dragomir, Marius; **Thompson, Mark** (eds.) (2014). *Mapping digital media. Global findings*. New York: Open Society Foundations.
https://www.opensocietyfoundations.org/reports/mapping-digital-media-global-findings

Fernández-Quijada, David; **Masip, Pere** (2013). "Tres décadas de investigación española en comunicación: hacia la mayoría de edad". *Comunicar*, n. 41, pp. 15-24.
https://doi.org/10.3916/C41-2013-01

Fidler, Roger (1997). *Mediamorphosis: Understanding new media*. Thousand Oaks, CA: Pine Forge Press. ISBN: 978 0 803990869

Flaounas, Ilias; **Ali, Omar**; **Lansdall-Welfare, Thomas**; **De-Bie, Tijl**; **Mosdell, Nick**; **Lewis, Justin**; **Cristianini, Nello** (2013). "Research methods in the age of digital journalism: Massive-scale automated analysis of news-content - topics, style and gender". *Digital journalism*, v. 1, n. 1, pp. 102-116.
https://doi.org/10.1080/21670811.2012.714928

Fredin, Eric S. (1997). "Rethinking the news story for the internet: Hyperstory prototypes and a model of the user". *Journalism & mass communication monographs*, n. 163, pp. 1-47.
http://www.aejmc.org/home/wp-content/uploads/2012/09/Eric-S.-Fredin.Rethinking-the-News-Story-for-the-Internet.September-1997.pdf

García-Avilés, José-Alberto; **Arias-Robles, Félix** (2016). "Géneros periodísticos en los formatos visuales de Twitter: una propuesta de tipología". *Textual & visual media*, v. 9, pp. 101-132.
http://textualvisualmedia.com/index.php/txtvmedia/article/view/52/42

García-Avilés, José-Alberto; **Carvajal-Prieto, Miguel**; **De-Lara-González, Alicia**; **Arias-Robles, Félix** (2018). "Developing an index of media innovation in a national market: The case of Spain". *Journalism studies*, v. 19, n. 1, pp. 25-42.
https://bit.ly/2GOexgQ
https://doi.org/10.1080/1461670X.2016.1161496

García-Avilés, José-Alberto; **León, Bienvenido**; **Sanders, Karen**; **Harrison, Jackie** (2004). "Journalists at digital television newsrooms in Britain and Spain: workflow and multi-skilling in a competitive environment". *Journalism studies*, v. 5, n. 1, pp. 87-100.
https://bit.ly/2R2b8ju
https://doi.org/10.1080/1461670032000174765

Gazoia, Alessandro (2013). *Il web e l'arte della manutenzione della notizia: il giornalismo digitale in Italia*. Roma: Minimum Fax. ISBN: 978 8 875215033

Gil-de-Zúñiga, Homero; **Jung, Nakwon**; **Valenzuela, Sebastián** (2012). "Social media use for news and individuals' social capital, civic engagement and political participation". *Journal of computer-mediated communication*, v. 17, n. 3, pp. 319-336.
https://doi.org/10.1111/j.1083-6101.2012.01574.x

Gillmor, Dan (2004). *We the media: Grassroots journalism by the people, for the people*. Sebastopol, CA: O'Reilly Media. ISBN: 978 0 596007331
http://library.uniteddiversity.coop/Media_and_Free_Culture/We_the_Media.pdf

Grey, David L.; **Brown, Trevor R.** (1970). "Letters to the editor: Hazy reflections of public opinion". *Journalism quarterly*, v. 47, n. 3, pp. 450-471.
https://doi.org/10.1177/107769907004700302

Grueskin, Bill; **Seave, Ava**; **Graves, Lucas** (2011). *The story so far: What we know about the business of digital journalism*. New York: Columbia University Press.
https://www.cjrarchive.org/img/posts/report/The_Story_So_Far.pdf

Guallar, Javier; **Rovira, Cristòfol**; **Ruiz, Sara** (2010). "Multimedialidad en la prensa digital. Elementos multimedia y sistemas de recuperación en los principales diarios digitales españoles". *El profesional de la información*, v. 19, n. 6, pp. 620-629.
https://doi.org/10.3145/epi.2010.nov.08

Gynnild, Astrid (2014). "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets". *Journalism*, v. 15, n. 6, pp. 713-730.
https://doi.org/10.1177/1464884913486393

Haake, Anja; **Hüser, Christoph**; **Reichenberger, Klaus** (1994). "The individualized electronic newspaper: An example of an active publication". *Electronic publishing*, v. 7, n. 2, pp. 89-111.
https://pdfs.semanticscholar.org/288b/1b7a833cf4a97288fd7e02dcdb37d3877e97.pdf

Harcup, Tony; **O'Neill, Deirdre** (2001). "What is news? Galtung and Ruge revisited". *Journalism studies*, v. 2, n. 2, pp. 261-280.
http://eprints.hud.ac.uk/id/eprint/27381
https://doi.org/10.1080/14616700118449

Harcup, Tony; **O'Neill, Deirdre** (2017). "What is news? News values revisited (again)". *Journalism studies*, v. 18, n. 12, pp. 1470-1488.
http://eprints.hud.ac.uk/id/eprint/27268
https://doi.org/10.1080/1461670X.2016.1150193

Heikkinen, Kalle J.; **Reese, Stephen D.** (1986). "Newspaper readers and a new information medium: Information need and channel orientation as predictors of videotex adoption". *Communication research*, v. 13, n. 1, pp. 19-36.
https://doi.org/10.1177/009365028601300103

Henke, Lucy L.; **Donohue, Thomas R.** (1986). "Teletext viewing habits and preferences". *Journalism quarterly*, v. 63, n. 3, pp. 542-545.
https://doi.org/10.1177/107769908606300314

Herbert, Jack; **Thurman, Neil** (2007). "Paid content strategies for news websites. An empirical study of British news papers' online business models". *Journalism practice*, v. 1, n. 2, pp. 208-226.
http://openaccess.city.ac.uk/118
https://doi.org/10.1080/17512780701275523

Hermida, Alfred (2012). "Social journalism: Exploring how social media is shaping journalism". In: Siapera, Eugenia; Veglis, Andreas (eds.). *The handbook of global online journalism*. Malden, MA: Wiley-Blackwell, pp. 309-328. ISBN: 978 1 4443 3855 3
https://doi.org/10.1002/9781118313978.ch17

Hermida, Alfred; **Fletcher, Fred**; **Korell, Darryl**; **Logan, Donna** (2012). "Share, like, recommend: Decoding the social media news consumer". *Journalism studies*, v. 13, n. 5-6, pp. 815-824.
https://doi.org/10.1080/1461670X.2012.664430

Hermida, Alfred; **Thurman, Neil** (2008). "A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites". *Journalism Practice*, v. 2, n. 3, pp. 343-356.
http://openaccess.city.ac.uk/53/2/hermida_thurman_a_clash_of_cultures.pdf
https://doi.org/10.1080/17512780802054538

Hiippala, Tuomo (2017). "The multimodality of digital longform journalism". *Digital journalism*, v. 5, n. 4, pp. 420-442.
https://jyx.jyu.fi/handle/123456789/53553
https://doi.org/10.1080/21670811.2016.1169197

Himelboim, Itai (2010). "The international network structure of news media: An analysis of hyperlinks usage in news

web sites". *Journal of broadcasting & electronic media*, v. 54, n. 3, pp. 373-390.
https://doi.org/10.1080/08838151.2010.499050

Holton, Avery E.; **Lawson, Sean**; **Love, Cynthia** (2005). "Unmanned aerial vehicles: Opportunities, barriers, and the future of 'drone journalism'". *Journalism practice*, v. 9, n. 5, pp. 634-650.
https://doi.org/10.1080/17512786.2014.980596

Howe, Jeff (2006). "The rise of crowdsourcing". *Wired*, v. 14, n. 6, pp. 1-4.
https://www.wired.com/2006/06/crowds

Huesca, Robert; **Dervin, Brenda** (1999). "Hypertext and journalism: Audiences respond to competing news narratives". In: Media in Transition Conference 1999. Boston: MIT
http://web.mit.edu/m-i-t/articles/huesca.html

Jacobson, Susan (2012). "Transcoding the news: An investigation into multimedia journalism published on nytimes.com 2000-2008". *New media & society*, v. 14, n. 5, pp. 867-885.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.833.2400&rep=rep1&type=pdf
https://doi.org/10.1177/1461444811431864

Jenkins, Henry (2003). "Transmedia storytelling". *MIT technology review*, 15 January.
https://www.technologyreview.com/s/401760/transmedia-storytelling

Jenkins, Henry (2006). *Convergence culture: Where old and new media collide*. New York: New Your University Press. ISBN: 978 0 814742815
https://bit.ly/2hInf10

Karlsen, Joakim; **Stavelin, Eirik** (2004). "Computational journalism in Norwegian newsrooms". *Journalism practice*, v. 8, n. 1, pp. 34-48.
https://doi.org/10.1080/17512786.2013.813190

Karlsson, Michael; **Sjøvaag, Helle** (2016). "Introduction. Research methods in an age of digital journalism". *Digital journalism*, v. 4, n. 1, pp. 1-7.
https://doi.org/10.1080/21670811.2015.1096595

Karlsson, Michael; **Sjøvaag, Helle** (eds.) (2019). *Rethinking research methods in an age of digital journalism*. London: Routledge. In press.

Kawamoto, Kevin (ed.) (2003). *Digital journalism: Emerging media and the changing horizons of journalism*. Lanham, MD: Rowman & Littlefield Publishers. ISBN: 978 0 7425 2681 5

Kaye, Jeff; **Quinn, Stephen** (2010). *Funding journalism in the digital age: Business models, strategies, issues and trends*. New York: Peter Lang. ISBN: 978 1 433106859

Koch, Tom (1991). *Journalism for the 21st century: Online information, electronic databases and the news*. New York: Greenwood. ISBN: 978 0 313277504

Kozinets, Robert V. (2015). *Netnography: redefined*. New York: SAGE Publications, 2nd ed. ISBN: 978 1 4462 8574 9

Kperogi, Farooq. A. (2012). "The evolution and challenges of online journalism in Nigeria". In: Siapera, Eugenia; Veglis, Andreas (eds.). *The handbook of global online journalism*. Malden MA: Wiley-Blackwell, pp. 445-461. ISBN: 978 1 4443 3855 3
https://doi.org/10.1002/9781118313978.ch24

Lagerkvist, Johan (2008). "Online journalism in China: constrained by politics, spirited by public nationalism". In: Paterson, Chris A.; Domingo, David (eds.). *Making online news: The ethnography of new media production*, New York: Peter Lang, pp. 127-141. ISBN: 978 1 433102141

Larrondo, Ainara (2009). "La metamorfosis del reportaje en el ciberperiodismo: concepto y caracterización de un nuevo modelo narrativo". *Communication & society*, v. 22, n. 2, pp. 59-88.
http://www.unav.es/fcom/communication-society/en/resumen.php?art_id=317

Leung, Louis (2009). "User-generated content on the internet: An examination of gratifications, civic engagement and psychological empowerment". *New media & society*, v. 11, n. 8, pp. 1327-1347.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.895.690&rep=rep1&type=pdf
https://doi.org/10.1177/1461444809341264

Lewis, Seth C.; **Kaufhold, Kelly**; **Lasorsa, Dominic L.** (2010). "Thinking about citizen journalism: The philosophical and practical challenges of user-generated content for community newspapers". *Journalism practice*, 2010, v. 4, n. 2, pp. 163-179.
https://doi.org/10.1080/14616700903156919

Lewis, Seth C.; **Zamith, Rodrigo**; **Hermida, Alfred** (2013). "Content analysis in an era of big data: A hybrid approach to computational and manual methods". *Journal of broadcasting & electronic media*, v. 57, n. 1, pp. 34-52.
https://doi.org/10.1080/08838151.2012.761702

López-García, Xosé (2008). *Ciberperiodismo en la proximidad*. Sevilla: Comunicación Social. ISBN: 978 84 96082 72 4

Mabweazara, Hayes-Mawindi; **Mudhai, Okoth-Fred**; **Whittaker, Jason** (eds.) (2014). *Online journalism in Africa: trends, practices and emerging cultures*. London: Routledge. ISBN: 978 1 138689190

MacGregor, Phil (2007). "Tracking the online audience: Metric data start a subtle revolution". *Journalism studies*, v. 8, n. 2, pp. 280-298.
https://doi.org/10.1080/14616700601148879

Manovich, Lev (2002). *The language of new media*. Cambridge, MA: MIT Press. ISBN: 978 0 262133746
https://dss-edit.com/plu/Manovich-Lev_The_Language_of_the_New_Media.pdf

Martínez-Costa, María-del-Pilar (1997). *La radio en la era digital*. Madrid: El País Aguilar. ISBN: 978 84 03596023

Masip, Pere (2005). "European research in communication during the years 1994-2004: a bibliometric approach". In: *First European communication conference* [CD-ROM]. Amsterdam: European Communication Research and Education Association (Ecrea).

Masip, Pere; **Díaz-Noci, Javier**; **Domingo, David**; **Micó-Sanz, Josep-Lluís**; **Salaverría, Ramón** (2010). "Investigación internacional sobre ciberperiodismo: hipertexto, interactividad, multimedia y convergencia". *El profesional de la información*, v. 19, n. 6, pp. 568-576.
https://doi.org/10.3145/epi.2010.nov.02

Masip, Pere; **Guallar, Javier**; **Suau, Jaume**; **Ruiz-Caballero, Carlos**; **Peralta, Miquel** (2015). "News and social networks: audience behavior". *El profesional de la información*, v. 24, n. 4, pp. 363-370.
https://doi.org/10.3145/epi.2015.jul.02

Masuda, Yoneji (1981). *The information society as post-industrial society*. Bethesda, MD: World Future Society. ISBN: 978 0 930242152

Mattelart, Armand (2001). *Histoire de la societé de l'information*. Paris: Éditions La Découverte. ISBN: 978 2 707134158

McCombs, Maxwell E.; **Shaw, Donald L.** (1972). "The agenda-setting function of mass media". *Public opinion quarterly*, v. 36, n. 2, pp. 176-187.
https://doi.org/10.1086/267990

McNair, Brian (2006). "What is journalism?". In: De-Burgh, Hugo (ed.). *Making journalists*. London: Routledge, pp. 41-59. ISBN: 978 1 134377558

Meraz, Sharon (2009). "Is there an elite hold? Traditional media to social media agenda setting influence in blog networks". *Journal of computer-mediated communication*, v. 14, n. 3, pp. 682-707.
https://doi.org/10.1111/j.1083-6101.2009.01458.x

Mercier, Arnaud; **Pignard-Cheynel, Nathalie** (2014). "Mutations du journalisme à l'ère du numérique: un état des travaux". *Revue française des sciences de l'information et de la communicatio*n, n. 5.
https://doi.org/10.4000/rfsic.1097

Meyer, Philip (1973). *Precision journalism*. Bloomington: Indiana University Press. ISBN: 978 0 253345974

Meyer, Philip (1991). *The new precision journalism*. Bloomington: Indiana University Press. ISBN: 978 0 253337900

Meyrowitz, Joshua (1985). *No sense of place: The impact of electronic media on social behavior*. New York: Oxford University Press. ISBN: 0 19 503474 0

Mitchell, Amy; **Rosenstiel, Tom**; **Christian, Leah** (2012). *Mobile devices and news consumption: Some good signs for journalism*. Washington, DC: Pew Research Center.
http://www.journalism.org/2012/03/18/mobile-devices-and-news-consumption-some-good-signs-for-journalism

Mitchelstein, Eugenia; **Boczkowski, Pablo J.** (2009). "Between tradition and change: A review of recent research on online news production". *Journalism*, v. 10, n. 5, pp. 562-586.
https://doi.org/10.1177/1464884909106533

Molinos, Manuel; **Marques, Nuno**; **Ferreira, Paulo** (2006). "Dez anos de jornalismo digital no Jornal de Notícias". *Comunicação e sociedade*, v. 9, n. 10, pp. 141-146.
https://doi.org/10.17231/comsoc.9(2006).1162

Myllylahti, Merja (2014). "Newspaper paywalls -the hype and the reality. A study of how paid news content impacts on media corporation revenues". *Digital journalism*, v. 2, n. 2, pp. 179-194.
https://doi.org/10.1080/21670811.2013.813214

Negroponte, Nicholas (1995). *Being digital*. New York: Alfred A. Knopf. ISBN: 0 679 43919 6

Nelson, Theodor-Holm (1965). "Complex information processing: A file structure for the complex, the changing and the indeterminate". In: *ACM Proceedings of the 1965 20th National conf.*, pp. 84-100.
https://dl.acm.org/citation.cfm?id=806036

Newman, Nic (2009). *The rise of social media and its impact on mainstream journalism*. Oxford: Reuters Institute for the Study of Journalism.
https://reutersinstitute.politics.ox.ac.uk/our-research/rise-social-media-and-its-impact-mainstream-journalism

Ngounou, Ingrid-Alice (2010). *Internet et la presse en ligne au Cameroun: Naissance, évolution et usages*. Paris: Editions L'Harmattan. ISBN: 978 2 296124424

Nicholls, Tom; **Shabbir, Nabeelah**; **Nielsen, Rasmus-Kleis** (2016). *Digital-born news media in Europe*. Oxford: Reuters Institute for the Study of Journalism.
https://ora.ox.ac.uk/objects/uuid:0037eea3-8f97-4bf3-a336-f86e21abfd0b

Nielsen, Jakob (1995). *Multimedia and hypertext: The internet and beyond*. Boston: AP Professional. ISBN: 978 0 125184083

Nielsen, Jakob (1998). "The end of legacy media (newspapers, magazines, books, TV networks)". *NN/g Nielsen Norman Group*.
https://www.nngroup.com/articles/the-end-of-legacy-media-newspapers-magazines-books-tv-networks

Noguera-Vivo, José-Manuel (2018). "You get what you give: Sharing as a new radical challenge for journalism". *Communication & society*, v. 31, n. 4, pp. 147-158.
http://www.unav.es/fcom/communication-society/en/articulo.php?art_id=711

Nyre, Lars; **Bjørnestad, Solveig**; **Tessem, Bjørnar**; **Øie, Kjetil-Vaage** (2012). "Locative journalism: Designing a location-dependent news medium for smartphones". *Convergence*, v. 18, n. 3, pp. 297-314.
https://doi.org/10.1177/1354856512441151

Oggolder, Christian; **Brügger, Niels**; **Metykova, Monika**; **Salaverría, Ramón**; **Siapera, Eugenia** (2019). "The emergence of the internet and the end of journalism?". In: Arnold, Klaus; Preston, Paschal; Kinnebrock, Susanne (eds.). *Handbook of European communication history.* Oxford: Wiley Blackwell (In press).

O'Reilly, Tim (2009). *What is web 2.0? Design patterns and business models for the next generation of software*. Sebastopol, CA: O'Reilly Media. ISBN: 978 1 449391706

Orihuela, José-Luis (2004). "eCommunication: the 10 paradigms of media in the digital age". In: Salaverría, Ramón; Sádaba, Charo (eds.). *Towards new media paradigms. Content, producers, organisations and audiences*. Pamplona: Ediciones Eunate, pp. 129-135. ISBN: 84 7768 150 3

Orihuela, José-Luis (2006). *La revolución de los blogs: cuando las bitácoras se convirtieron en el medio de comunicación de la gente*. Madrid: La Esfera de los Libros. ISBN: 978 84 97344982

Outing, Steve (2004). "Eyetrack III: What news websites look like through readers' eyes". *Poynter Institute*, 24 August.
https://www.poynter.org/news/eyetrack-iii-what-news-websites-look-through-readers-eyes

Palacio, Luis (2018). "Modelos de negocio de los medios digitales: diversificar ingresos para financiar el periodismo". *Cuadernos de periodistas*, n. 36, pp. 9-22.
http://www.cuadernosdeperiodistas.com/media/2018/07/9-22-Luis-Palacio.pdf

Palacios, Marcos; **Díaz-Noci, Javier** (eds.) (2009). *Online journalism: research methods. A multidisciplinary approach in comparative perspective*. Bilbao: Servicio Editorial de la Universidad del País Vasco. ISBN: 978 84 9860 191 6
http://hdl.handle.net/10810/15569

Paulussen, Steve; **Ugille, Pieter** (2008). "User generated content in the newsroom: Professional and organisational constraints on participatory journalism". *Westminster papers in communication & culture*, v. 5, n. 2, pp. 24-41.
https://doi.org/10.16997/wpcc.63

Pavlik, John V. (2001). *Journalism and new media*. New York: Columbia University Press. ISBN: 978 0 23111483

Perigoe, Ross (2009). "Ten-year retrospective: Canada and the United States in the age of digital journalism". *Journal of media practice*, v. 10, n. 2-3, pp. 247-253.
https://www.tandfonline.com/doi/abs/10.1386/jmpr.10.2-3.247_3

Picard, Robert G. (2000). "Changing business models of online content services: Their implications for multimedia and other content producers". *International journal on media management*, v. 2, n. 2, pp. 60-68.
https://doi.org/10.1080/14241270009389923

Picard, Robert G. (2011). *Mapping digital media: Digitization and media business models*. London: Open Society Foundations.
https://www.opensocietyfoundations.org/reports/digitization-media-business-models

Pool, Ithiel-de-Sola (1983). *Technologies of freedom*. Cambridge, MA: Harvard University Press. ISBN: 978 0 674872325

Postman, Neil (1992). *Technopoly: The surrender of culture to technology*. New York: Alfred A. Knopf. ISBN: 978 0 394582726
https://www.collier.sts.vt.edu/1504/pdfs/technopoly-neil-postman.pdf

Qualman, Erik (2009). *Socialnomics: How social media transforms the way we live and do business*. Hoboken, NJ: John Wiley & Sons. ISBN: 978 0 470477236

Quandt, Thorsten; **Singer, Jane B.** (2009). "Convergence and cross-platform content production". In: Wahl-Jorgensen, Karin; Hanitzsch, Thomas (eds.). *The handbook of journalism studies*. London: Taylor and Francis, pp. 130-144. ISBN: 978 1 315167497

Quinn, Stephen (2005). "Convergence's fundamental question". *Journalism studies*, v. 6, n. 1, pp. 29-38.
https://doi.org/10.1080/1461670052000328186

Reagan, Joey (1989). "New technologies and news use: Adopters vs. nonadopters". *Journalism quarterly*, v. 66, n. 4, pp. 871-877.
https://doi.org/10.1177/107769908906600414

Reese, Stephen D. (1988). "New communication technologies and the information worker: The influence of occupation". *Journal of communication*, v. 38, n. 2, pp. 59-70.
https://doi.org/10.1111/j.1460-2466.1988.tb02047.x

Reese, Stephen D. (2016). "The new geography of journalism research: Levels and spaces". *Digital Journalism*, n. 4, v. 7, pp. 816-826.
https://doi.org/10.1080/21670811.2016.1152903

Renó, Denis (2014). *Periodismo transmedia: miradas múltiples*. Barcelona: Editorial UOC. ISBN: 978 84 90642399

Rivera-Rogel, Diana-Elizabeth (2012). *Los cibermedios en Ecuador: evolución, estructura y ciberparticipación*. Tesis doctoral. Santiago de Compostela: Universidad de Santiago de Compostela.
https://minerva.usc.es/xmlui/handle/10347/3712

Robin, Michael; **Poulin, Michel** (2000). *Digital television fundamentals: Design and installation of video and audio systems*. New York: McGraw-Hill. ISBN: 978 0 071355810

Rojano, Miladys (2006). *Diez años de periodismo digital en Venezuela 1996-2006*. Caracas: Universidad Católica Andrés Bello.

Rosique-Cedillo, Gloria; **Barranquero-Carretero, Alejandro** (2015). "Periodismo lento (*slow journalism*) en la era de la inmediatez. Experiencias en Iberoamérica". *El profesional de la información*, v. 24, n. 4, pp. 451-462.
https://doi.org/10.3145/epi.2015.jul.12

Rost; **Alejandro**; **Bernardi, María-Teresa**; **Bergero, Fabián** (2016). *Periodismo transmedia. La narración distribuida de la noticia*. Neuquén: Publifadecs. ISBN: 978 987 46421 1 0

Sádaba-Chalezquer, Charo (2016). "Innovación en el sector de los medios". In: Salaverría, Ramón (ed.). *Ciberperiodismo en Iberoamérica*. Madrid: Fundación Telefónica & Editorial Ariel, pp. 423-433. ISBN: 978 84 08 15452 5

Salaverría, Ramón (2005a). *Redacción periodística en internet*. Pamplona: Eunsa. ISBN: 84 313 2259 4

Salaverría, Ramón (coord.) (2005b). *Cibermedios. El impacto de internet en los medios de comunicación en España*. Sevilla: Comunicación Social Ediciones y Publicaciones. ISBN: 978 84 96082 33 5

Salaverría, Ramón (2014). "Multimedialidade: informar para cinco sentidos". In: Canavilhas, João (org.). *Webjornalismo: 7 caraterísticas que marcam a diferença*. Covilhã: Livros LabCom, pp. 25-52. ISBN: 978 989 654 145 3

Salaverría, Ramón (2015). "Ideas para renovar la investigación sobre medios digitales". *El profesional de la información*, v. 24, n. 3, pp. 223-226.
https://doi.org/10.3145/epi.2015.may.01

Salaverría, Ramón (ed.) (2016). *Ciberperiodismo en Iberoamérica*. Madrid: Fundación Telefónica & Editorial Ariel. ISBN: 978 84 08 15452 5

Salaverría, Ramón (2017). "Tipología de los cibermedios periodísticos: bases teóricas para su clasificación". *Revista mediterránea de comunicación*, v. 8, n. 1, pp. 19-32.
https://doi.org/10.14198/MEDCOM2017.8.1.2

Salaverría, Ramón; **Cores, Rafael** (2005). "Géneros periodísticos en los cibermedios hispanos". In: Salaverría, Ramón (coord.). *Cibermedios. El impacto de internet en los medios de comunicación en España*. Sevilla: Comunicación Social Ediciones y Publicaciones, pp. 145-185. ISBN: 978 84 96082 33 5

Salaverría, Ramón; **Martínez-Costa, María-del-Pilar**; **Breiner, James** (2018). "Mapa de los cibermedios de España en 2018: análisis cuantitativo". *Revista latina de comunicación social*, n. 73, pp. 1034-1053.
https://doi.org/10.4185/RLCS-2018-1295

Salaverría, Ramón; **Negredo, Samuel** (2008). *Periodismo integrado: Convergencia de medios y reorganización de redacciones*. Barcelona: Sol90 Media. ISBN: 978 84 9820 954 9

Scolari, Carlos A. (2013a). "De las tablillas a las tablets: evolución de las emagazines". *El profesional de la información*, v. 22, n. 1, pp. 10-17.
https://doi.org/10.3145/epi.2013.ene.02

Scolari, Carlos A. (2013b). *Narrativas transmedia. Cuando todos los medios cuentan*. Barcelona: Deusto. ISBN: 978 84 234 1336 2

Scott, Ben (2005). "A contemporary history of digital journalism". *Television & new media*, v. 6, n. 1, pp. 89-126.
https://doi.org/10.1177/1527476403255824

Shin, Donghee; **Biocca, Frank** (2017). "Exploring immersive experience in journalism". *New media & society*, v. 20, n. 8, pp. 2800-2823.
https://doi.org/10.1177/1461444817733133

Siapera, Eugenia; **Veglis, Andreas** (eds.) (2012). *The handbook of global online journalism*. Malden, MA: Wiley-Blackwell. ISBN: 978 1 4443 3855 3

Smith, Anthony (1980). *Goodbye Gutenberg: The newspaper revolution of the 1980's*. New York: Oxford University Press. ISBN: 978 0 195027099

Smyrnaios, Nikos; **Bousquet, Franck** (2011). "The development of local online journalism in South-Western France: The case of 'La Dépêche du Midi'". In: Salaverría, Ramón (ed.). *Diversity of Journalisms. Proceedings of ECREA/CICOM conference*, Pamplona, 4-5 July 2011. Pamplona: Eunsa, pp. 347-358. ISBN: 84 8081 082 3

Steensen, Steen (2011). "Online journalism and the promises of new technology: A critical review and look ahead". *Journalism studies*, v. 12, n. 3, pp. 311-327.
https://doi.org/10.1080/1461670X.2010.501151

Steensen, Steen; **Ahva, Laura** (2015). "Theories of journalism in a digital age: An exploration and introduction". *Digital journalism*, v. 3, n. 1, pp. 1-18.
https://doi.org/10.1080/17512786.2014.928454

Stroud, Natalie-Jomini; **Scacco, Joshua M.**; **Curry, Alexander L.** (2016). "The presence and use of interactive features on news websites". *Digital journalism*, v. 4, n. 3, pp. 339-358.
https://doi.org/10.1080/21670811.2015.1042982

Tähismaa, Inno (2003). *Online uudised Eesti ajalehtedes* [Online news in Estonian newspapers]. Tesis doctoral. Tartu: University of Tartu.
http://dspace.ut.ee/bitstream/handle/10062/15450/Bakt_Tahismaa.pdf

Tenor, Carina (2018). "Hyperlocal news and media accountability". *Digital journalism*, v. 6, n. 8, pp. 1064-1077.
https://doi.org/10.1080/21670811.2018.1503059

Tewksbury, David (2005). "The seeds of audience fragmentation: Specialization in the use of online news sites". *Journal of broadcasting & electronic media*, v. 49, n. 3, pp. 332-348.
https://doi.org/10.1207/s15506878jobem4903_5

Vaccaro, Valerie L.; **Cohn, Deborah Y.** (2004). "The evolution of business models and marketing strategies in the music industry". *International journal on media management*, v. 6, n. 1-2, pp. 46-58.
https://doi.org/10.1080/14241277.2004.9669381

Van-der-Wurff, Richard (2012). "The economics of online journalism". In: Siapera, Eugenia; Veglis, Andreas (eds.). *The handbook of global online journalism*. Malden, MA: Wiley-Blackwell, pp. 231-250. ISBN: 978 1 4443 3855 3

Van-der-Wurff, Richard; **Lauf, Edmund** (eds.) (2005). *Print and online newspapers in Europe. A comparative analysis in 16 countries*. Amsterdam: Het Spinhuis Publishers. ISBN: 90 5589 238 6

Van-Dijck, José (2009). "Users like you? Theorizing agency in user-generated content". *Media, culture & society*, v. 31, n. 1, pp. 41-58.
https://doi.org/10.1177/0163443708098245

Washburn, Pat (1981). "Electronic journalism, computers and privacy". *Computer/Law journal*, v. 3, n. 1, pp. 189-209.
https://repository.jmls.edu/jitpl/vol3/iss1/8

Westlund, Oscar (2013). "Mobile news: A review and model of journalism in an age of mobile media". *Digital journalism*, v. 1, n. 1, pp. 6-26.
https://doi.org/10.1080/21670811.2012.740273

Williams, Andy; **Harte, David** (2016). "Hyperlocal news". In: Witschge, Tamara; Anderson, C. W.; Domingo, David; Hermida, Alfred (eds.). *The SAGE handbook of digital journalism*. London: SAGE, pp. 280-293. ISBN: 978 1 473955066

Wolf, Cornelia; **Godulla, Alexander** (2016). "Potentials of digital longforms in journalism. A survey among mobile Internet users about the relevance of online devices, Internet-specific qualities, and modes of payment". *Journal of media business studies*, v. 13, n. 4, pp. 199-221.
https://doi.org/10.1080/16522354.2016.1184922

Wolf, Cornelia; **Schnauber, Anna** (2015). "News consumption in the mobile era: The role of mobile devices and traditional journalism's content within the user's information repertoire". *Digital journalism*, v. 3, n. 5, pp. 759-776.
https://doi.org/10.1080/21670811.2014.942497

Yezers'ka, Lyudmyla (2008). *Ciberperiodismo en Perú. Análisis de los diarios digitales*. Lima: Editorial San Marcos y Universidad de Piura, 2008. ISBN: 978 997 2 38 594 0

Scholarly reputation building in the digital age: An activity-specific approach. Review article

Eti Herman; David Nicholas

How to cite this article:

Herman, Eti; **Nicholas, David** (2019). "Scholarly reputation building in the digital age: An activity-specific approach. Review article". *El profesional de la información*, v. 28, n. 1, e280102.

https://doi.org/10.3145/epi.2019.ene.02

Article received on November 16, 2018
Approved on December 11, 2018

Eti Herman ✉
https://orcid.org/0000-0001-8526-9081

Ciber Research Ltd, Newbury, UK.
eherman@univ.haifa.ac.il

David Nicholas
https://orcid.org/0000-0001-8046-2835

Tomsk State University, Russia
Dave.Nicholas@CIBER-research.eu

Abstract

Seeking to understand how today's scholars may, indeed should go about building, maintaining and showcasing their professional reputation, the literature review presented here explores the reputational opportunities available to them in the increasingly open-values based, digital and networked environment of Science 2.0. Using a conceptual framework developed with the help of the *European Commission* and specifically designed for conducting analytical reviews and audits of the reputational value of scholarly activities, this study examines in some detail the practices –more than 30 of them- that comprise the present-day scientific undertaking from a reputation-accruing angle.

Keywords

Scholarly reputation; Reputation building; Communication; Scholars; Careers; Profession; Professional activities; Literature review; Bibliography.

1. Introduction

As it has long been shown, peer recognition of one's professional accomplishments, as it builds up over time to culminate in a good scientific reputation[1], is a primary goal and central motivator for members of the scholarly[2] community (**Becher**, 1989; **Becher**; **Trowler**, 2001; **Merton**, 1968; 1973; **Storer**, 1966).

Inevitably so, of course, for the cultivation of science is a highly communal undertaking, which has scholars extraordinarily dependent on informed scholars for a good opinion of their achievements. Indeed, it is only when a scientific contribution has been assessed, validated and its value confirmed by the scholarly community that it can confer on its originator a high intellectual standing (**Blackmore**; **Kandiko**, 2011; **Latour**; **Woolgar**, 1986; **Hagstrom**, 1964; 1974; **Storer**, 1963). As the achievement of a good reputation is translated into many concrete rewards for the scholar, some of which can be 'cashed in' for money –employment, tenure, promotions, resources, publications, prizes- it is hardly surprising that they strive relentlessly to build and enhance their prestige (**Blackmore**; **Kandiko**, 2011; **Reif**, 1961).

Acknowledgements

The *European Commission*'s *Institute for Prospective Technological Studies* for funding the *Analysis of emerging reputation, mechanisms for scholars* project, which has informed this paper, as well as *Tomsk State University Competitiveness Improvement Program* for their support.

Indeed, as Merton suggests, the hunger for recognition, so characteristic of scholars, is hardly the manifestation of vanity it may seem to be at a first glance; rather, it is

> "the outer face of the inner need for assurance that one's work really matters, that one has measured up to the hard standards maintained by the community of scientists" (**Merton**, 1963, p. 270).

Thinking much along the same lines, albeit going one step further, **Storer** (1963) contends that peer recognition is frequently interpreted by scholars not only as sanctioning the validity and significance of their work but, more generally, as an affirmation of their own personal worth.

If reputation is traditionally everything for a scholar, this would seem to be all the more so nowadays, when the increasingly marketised and entrepreneurial higher education system world-wide is driven by an intense rivalry among institutions forever competing for resources and recognition (**Altbach**; **Reisberg**; **Rumbley**, 2009; **Blackmore**, 2016a; **Clark**, 1998; **Delanty**, 1998; **Etzkowitz**; **Leydesdorff**, 2000; **Frost**; **Brockmann**, 2014; **Gibbons** *et al.*, 1994; **Nedeva**; **Boden**; **Nugroho**, 2012; **Waaijer** *et al.*, 2018; **Winter**, 2017). In fact, with universities vying with each other for students, star professors, funding and their share of the state's limited budget, prestige-affording recognition of their scholarly achievements becomes the key for winning the competition. So much so, that today the scholarly world seems to have at its heart a very elaborate prestige economy, a kind of celebrity system, whose linchpin are the scholars (**Barbour**; **Marshall**, 2012; **Blackmore**, 2016a; 2018; **Blackmore**; **Kandiko**, 2011).

> "Peer recognition is frequently interpreted by scholars not only as sanctioning the validity and significance of their work but, more generally, as an affirmation of their own personal worth"

Not only have scholars thus become 'managed professionals' (**Rhoades**, 1998) and 'state-subsidized entrepreneurs' (**Slaughter**; **Leslie**, 2001) in an incessant race for attaining commercially attractive purposes and marketable outcomes (**Gibbons** *et al.*, 1994; **Rinne**; **Koivula**, 2009), but their professional priorities, too, look as if they may be undergoing changes, with an attendant transformation in what is seen as a reputation-accruing scholarly accomplishment. Traditionally, narrowly defined research attainments –the volume of papers published in high-ranking journals and the number of citations they obtain- have been given disproportionate reputational weight above all other scholarly activities (**Boyer**, 1990; **Harley** *et al.*, 2010; **Van-Dalen**; **Henkens**, 2012). However, as **Blackmore** (2016a) contends, when pressures to produce particular outputs intensify, existing tensions in what is valued (for example, between research and teaching, or between pure and applied research) are likely to increase.

Thus, now that the open, democratised, technology- and collaboration-centered paradigms of Science 2.0 (*European Commission*, 2014; **Lasthiotakis**; **Kretz**; **Sá**, 2015; **Shneiderman**, 2008; **Vicente-Sáez**; **Martínez-Fuentes**, 2018; **Cronin**, 2017) are fast becoming the foundations on which today's scholarly realities are built (Weller, 2011); now that the 'triple helix' model of academic-government-industry collaboration (**Leydesdorff**; **Etzkowitz**, 1996; **Etzkowitz**; **Leydesdorff**, 2000) is well established; and now that the future in a globalised knowledge society is seen by policy makers as hinging not only on research and innovation, but also on education for all (*European Commission*, 2013; *European Parliament*, 2012), a more wide-ranging, inclusive and representative view of reputation-building scholarly achievement is called for. Seeking to understand how, in these circumstances, scholars may go about building, maintaining and showcasing their prestige, the state-of-the-art, literature-based study, reported here, takes an activity-specific approach to the exploration of the reputation-accruing components of today's scholarly practices.

> "Traditionally, narrowly defined research attainments –the volume of papers published in high-ranking journals and the number of citations they obtain– have been given disproportionate reputational weight above all other scholarly activities"

2. Aims and objectives

The overarching aim of the study is establishing, through a conceptual analysis and audit of the pertinent literature, how today's scholars do/might go about building, sustaining and enhancing their reputation as part and parcel of the great variety of activities that comprise their work-life.

The specific objects of the investigation are therefore:

- Identifying the range of traditional and novel, online and offline activities which, taken together, form the present-day scholarly undertaking.
- Identifying the reputational potentials and affordances of each activity.
- Identifying the ways and means at a scholar's disposal for taking advantage of the reputational opportunities found to be on offer in the increasingly open-values based, digital and networked environment of the Science 2.0 age.

3. Scope and definitions

The study reported here builds on the findings of an exploratory investigation (**Jamali**; **Nicholas**; **Herman**, 2016; **Nicholas**; **Herman**; **Jamali**, 2015a; 2015b; **Nicholas** *et al.*, 2015a), commissioned by the *European Commission* through its *Joint Research Centre* to investigate novel developments in the field of scholarly reputation building. The project, calling as it did for developing a conceptual framework for conducting analytical reviews and audits of the reputational affordances of scholarly activities (for full details see the Methodology section), yielded comprehensive and systematically structured evidence on the reputational components of scholars' professional practices. It is this unique body of evidence that is recapped, extended and brought up to date in the literature-based analytical examination of scholarly reputation building that follows.

However, first an examination of the term 'scholarly/scientific reputation' is in order, for, curiously enough, with all its above-noted centrality to the scholarly endeavor and the frequent reference to it in academic discourse, people seem to use the phrase without giving much thought to what it really means (**Bourne**; **Barbour**, 2011; **Parra** *et al.*, 2011). Neither is the term 'reputation' distinctly differentiated in everyday parlance from related concepts, such as recognition, visibility and impact. The definition adopted for the purposes of this study is Herman's literature-derived one:

> "Scholarly reputation is the expert appraisal of a scholar's standing in their collegial reference group, which is collectively determined on the basis of their research achievements in terms of productivity –high quantity and high-quality scholarly output, and impactfullness[3]- the effects attributable to their thinking and work over time. Thus, successful scholarly reputation building is contingent upon making one's research visible in the 'scholarly marketplace' in an effort to capture the attention of those who are capable of judging its over-all value" (**Herman**, 2018).

Also, although traditionally theoreticians have tended to treat the terms *prestige* and *reputation* as synonyms, when they talked of scientists' gain from the perceived scientific value of their contributions –see, for example, **Andersen**'s (2000) review of the literature on the subject-, more recent studies do propose a distinction between the two concepts, at least on the institutional level. Thus, for example, **Blackmore** (2016a), building on the work of **Brewer**, **Gates**, & **Goldman** (2001) on the topic, notes quite a few distinctions between the two terms, the gist of which being that reputation, unlike prestige, is an absolute term that is not measured in relation to others at all, indeed, does not necessarily have to be gained at the expense of a competitor.

"Scholarly reputation is the expert appraisal of a scholar's standing in their collegial reference group, which is collectively determined on the basis of their research achievements in terms of productivity –high quantity and high-quality scholarly output, and impactfullness– the effects attributable to their thinking and work over time"

4. Methodology

4.1. The conceptual basis

The point of departure for this literature-based exploration of current and emerging scholarly practices and their reputation building purposes and mechanisms was **Boyer**'s (1990) seminal mapping of the broad territory of scholarly activity, which, although a product of the last century, remains valid in its basic observations and contentions[4]. Indeed, **Boyer**'s (1990) model, which defines scholarship as capturing the whole range of scholarly activities in an attempt to present 'a more exclusive view of what it means to be a scholar', has been shown in previous studies to be suitable for providing a sound basis for exploring scholarly behaviours (see, for example, **Braxton**; **Luckey**; **Helland**, 2002; **Crow** *et al.*, 2018; **Garnett**; **Ecclesfield**, 2012; **Greenhow**; **Gleason**, 2014; 2015; **Heap**; **Minocha**, 2012; **Pearce** *et al.*, 2010; **Scanlon**, 2014; **Weller**, 2011).

However, as any consideration of contemporary scholarly practices obviously needs to address their digitally changed and fluctuating nature, **Boyer**'s (1990) model could not have served our purposes had it not been updated and extended to reflect the realities of the Science 2.0 age. Thus, the conceptual basis of this paper for examining contemporary scholars' professional activities was Boyer's well-established, four-dimensional model of scholarship, updated by **Garnett** & **Ecclesfield** (2012) to include a fifth facet (co-creation):

1) The scholarship of research (discovery), the individual or collaborative creation of new knowledge;

2) The scholarship of integration, the arraying of extant knowledge into larger intellectual patterns, often within a wider, cross-disciplinary context;

3) The scholarship of application, the application of disciplinary knowledge and skill to societal/practical problems;

4) The scholarship of teaching, the conveying of the human store of knowledge to new generations;

5) The scholarship of co-creation, the participation of teachers, students and practitioners in the increasingly converging processes of knowledge production and transmission.

4.2. Data collection and analysis

The methodology employed for this paper is desk research and expert evaluation, with the gathering and analysis of the evidence undertaken through the utilisation, as a template, of a framework specially developed for the purpose (see next section). Using **Boyer**'s (1990) classification as a benchmark against which present-day scholarly practices could be compared, as suggested by **Scanlon** (2014) and **Weller** (2011), the study was thus conducted in six stages, each one feeding into the next:

- The published literature was searched in order to compile a comprehensive list of the range of scholarly activities, both online and offline, which comprise the work-life of scholars.
- Each activity identified was defined/described to denote its precise nature and procedures.
- Each of the activities was then analysed to discern its scientific purposes. This enabled the classification of the various activities by the main scientific purpose they serve into the five scholarly categories.
- Each of the activities was further analysed to determine its reputational purposes (if any).
- Each activity found to have reputational purposes was evaluated to discern the specific fit for purpose reputational mechanism(s) it utilised. This, on the basis of our literature-based awareness of the ways and means at the disposal of scholars for achieving visibility and obtaining peer recognition and esteem, which allowed for 'matching' the processes and mechanisms that could be useful in each case with the hoped-for reputational outcome.
- The picture that had emerged from steps 1-5 was further analysed in an attempt to compare the overall reputational potential of the five categories of scholarly activities.

It is important to note here that whilst the literature provides a wealth of information as to the specific practices encompassing the scholarly undertaking, surprisingly, very little of it had been examined from a reputational angle, the only exceptions being discussions of the reputational effects of excelling (or not) in research and the reputational affordances of scholarly social networking platforms (SSNs). Plainly then, many of the findings presented in this study are based on a careful analysis of the literature on how things work in academe, rather than specific empirical evidence on scholarly reputational building practices. Exactly how the paper went about finding answers to the questions asked at the outset is perhaps best explained through the example which follows.

"Many of the findings presented in this study are based on a careful analysis of the literature on how things work in academe, rather than specific empirical evidence on scholarly reputational building practices"

The relatively mundane activity of requesting/providing help in locating research literature is certainly an activity that scholars often undertake in the course of their work. However, can it possibly have reputation building potential? There appears to be no concrete evidence as to how this activity may be of help in reputation building. Still, online sharing/accepting of help can certainly afford visibility, which, by definition, is conducive to enhanced reputation, for it is through exposure of their work that scholars can capture the attention of those who are capable of judging their scientific accomplishments (**Franck**, 1999). Thus, it may be said that scholars' sharing of research literature on a peer-to-peer basis, and from a reputational point of view even more effectively via SSNs, can contribute to their achieving visibility-afforded reputation.

4.3. The framework

Originally developed as part and parcel of the aforementioned *European Commission* commissioned exploratory investigation (**Nicholas**; **Herman**; **Jamali**, 2015a; 2015b), the framework adopted was aired at interviews, focus groups and in a questionnaire with 251 academics from a wide-range of European countries (**Jamali**; **Nicholas**; **Herman**, 2016; **Nicholas** *et al.*, 2015a). Since its employment for the purposes of the original *EC* project that serves as the basis for the present undertaking, the framework has also been rolled-out in a study that assessed the support provided by *ResearchGate* (*RG*) to scholars' reputation building endeavours (**Nicholas**; **Herman**; **Clark**, 2016).

The framework is represented and summarised in eight tables, delineating the activities that comprise the scholarly undertaking in each of the above-noted 5 categories of scholarship: research, integration, application, teaching and co-creation. Each category is summarised separately, although in point of fact the entire range of research associated activities is dispersed amongst three categories (research, integration and application). This, because all three have as their aim the creation of new knowledge, albeit with a different focus. Inevitably, then, some of the research activities are typically undertaken in all of these categories.
In an attempt to avoid redundancy, the first four tables, summarising the scholarship of research, offer a full description of the different research activities typically undertaken in a scholarly investigation. The tables summarising the two remaining research categories focus,

"Scholars' sharing of research literature on a peer-to-peer basis can contribute to their achieving visibility-afforded reputation"

therefore, only on the activities unique to the scholarships of integration and application, respectively.

For each broad category the following data are provided:

- Specific activities: practices performed either online or offline by scholars for work-related purposes. Thus, for example, a major research activity in the work-life of scholars is the producing of research output.
- Scientific purpose: the anticipated contribution of a scholarly activity towards the advancement of science and the achievement of its goals for benefiting humankind. Thus, for example, the expected contribution of producing a research output is advancing science via discovering new knowledge and/or achieving enhanced understanding in order to solve a theoretical or practical problem for the public good.
- Reputational purpose: the anticipated contribution of a scholarly activity towards building/ maintaining/ enhancing a scholar's standing among their peers and, at times, the general public. Thus, for example, producing a research output has as its reputational purpose the obtaining of peer recognition and esteem.
- Fit for purpose reputational mechanism: the specific purpose-relevant process used to build reputation. Thus, for example, in order to achieve the reputational purpose of producing a research output the results of a scientific investigation need to be formally-written up and communicated in a manner suitable for presenting to peers for their evaluation and use.

5. Findings

5.1. The scholarship of research (discovery)

The scholarship of research, the pursuit of knowledge for its own sake and the benefit of humankind, is universally held to be the principal professional endeavour and focal point of the scholarly enterprise (**Harley** *et al.*, 2010; **Van-Dalen**; **Henkens**, 2012; **Wilson**, 1942; **Wolff-Eisenberg**; **Rod**; **Schonfeld**, 2016a; 2016b). Indeed, as this study's findings re-affirm yet again, there can be little doubt that in these days, too, as **Boyer** (1990, p. 2) said more than a quarter of a century ago, "to be a scholar is to be a researcher". The evident primacy of the scholarship of research over other aspects of the scientific undertaking is obviously associated with the importance accorded to its stated goal of extending the stock of human knowledge. Still, the centrality of research, 'disinterested' a pursuit as it should be (**Merton**, 1973), undeniably stems from its aforementioned reputation building capacities, too, for research achievements are used as the yardstick by which scholarly success is measured (**Borrego**; **Anglada**, 2016; **Boyer**, 1990; **De-Rond**; **Miller**, 2005; **Dewett**; **Denisi**, 2004; **Fanelli**; **Larivière**, 2016; **Harley** *et al.*, 2010; **Mabe**; **Mulligan**, 2011; **Miller**; **Taylor**; **Bedeian**, 2011; **Mulligan**; **Hall**; **Raphael**, 2013; **Ponte**; **Simon**, 2011; **Van-Dalen**; **Henkens**, 2012; **Waaijer** *et al.*, 2018; **Wilson**, 1942). Thus, scholars are greatly concerned with the impact of their research upon the wealth of human knowledge for the sake of the scholarly endeavour and society, certainly, but no less for the sake of their professional prestige.

> "Scholars are greatly concerned with the impact of their research upon the wealth of human knowledge for the sake of the scholarly endeavour and society, certainly, but no less for the sake of their professional prestige"

With research achievements seen as being synonymous with scholarly success, it is hardly surprising to find, as Åkerlind (2008) concludes from a review of a host of studies, that an important underlying intention in being a researcher and undertaking research is establishing oneself in the field and gaining, thereby, standing amongst scholarly peers. Indeed, according to **Blackmore** & **Kandiko** (2011), the extrinsic motivator of money and the intrinsic motivator of interest in the field of activity intersect for scholars to culminate in a blended motivation of prestige benefits awarded for pursuing, creating, and sharing knowledge. Thus, as **Brew** (2001) finds, one perception among scholars of the concept of research is that it is a kind of social marketplace, where the products of research (publications, grants, and networks) are exchanged for money, prestige or recognition. By the same token, **Bazeley** (2010) sees scholarly reputation as not merely a by-product of the research process but, alongside publications and impact, one of its three main outcomes.

Plainly then, the quest for reputation is literally 'built into' research work. Indeed, the portrayal of the range of traditional and novel activities comprising the scholarship of research in today's knowledge-driven era, presented below, shows them all to have a strong reputational focus alongside their scientific one. This holds true whether a research activity is performed individually or in collaboration with others, whether it is specifically aimed at the actual producing of an original contribution to human knowledge, the dissemination of the by-products and outputs of research work, the networking with colleagues or the evaluation of others' research outputs.

> "The quest for reputation is literally 'built into' research work"

5.1.1. Producing research output

Producing a new input to the extant body of certified knowledge is comprised of stages that follow a reliable, if not always consciously or rigorously adhered to progressive order (**Garvey**, 1975). This generic workflow is very much with us still, despite the aforementioned societal-demands-driven transformations in the scholarly environment and the te-

chnology-afforded changes in the research process itself (**Weller**, 2011), which brought about a widening of the range of acceptable research outputs and distribution channels. The procedure involves a series of activities, each of which has been found to have reputation building potentials (for a full list see **Nicholas**; **Herman**; **Jamali**, 2015a). A representative selection of the key activities are presented in Table 1.1.

Table 1.1. Producing research output

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Identifying a researchable topic, planning the research project and obtaining funding	Finding a scientifically significant research topic and establishing its viability	Producing evidence of scholarly ability to identify the significance of the topic and conduct the research as proposed; achieving visibility for one's ideas	Constructing a proposal for interested collaborators and for persuading funders that the proposed project can yield the best research on an important topic
Reviewing the pertinent previous knowledge	Anchoring a research undertaking in its theoretical base	Obtaining peer recognition and esteem	Selecting appropriate research content and presenting it as an analytic review of the literature
Requesting/providing help for locating pertinent previous knowledge	Same as above	Achieving disciplinary and trans-disciplinary visibility	Sharing literature peer-to-peer or via social media based scholarly platforms
Producing a research output individually or in collaboration with peers or even committed amateur experts	Discovering new knowledge and/or achieving enhanced understanding	Obtaining peer recognition and esteem; achieving visibility among one's peers	Presenting the results of a scientific investigation in a manner suitable for peer use and evaluation

Having seen how various activities aimed at producing a research output can contribute towards enhancing scholarly reputation, it is important to single out one activity that plays an especially vital role in the process: obtaining external research funding. After all, there can be little doubt that beyond providing scholars with the essential financial resources to conduct research, grants are also purveyors of prestige, especially if the funder is a blue-chip organisation. As **Laudel** (2005) explains, the decision of a grant-giving agency to fund a research, based as it is on peer review, represents a vote of confidence in a scholar by their peers, and, of course, the more competitive the grant, and the more rigorous the peer review system of the funder, the higher it is weighted. Indeed, a host of studies attests to the importance accorded in academe to the acquisition of research grants as a measure of successful research performance, which, as already noted, is seen as a reputation enhancing achievement (**Anderson**; **Slade**, 2016; **Auranen**; **Nieminen**, 2010; **Bloch**; **Graversen**; **Pedersen**, 2014; **Boyer**; **Cockriel**, 1998; 2001; **Nicholas** *et al.*, 2018; **Van-Arensbergen**; **Van-der-Weijden**; **Van-den-Besselaar**, 2014). So much so, that the rigorous directives of the 'publish-or-perish' mentality in academe have long been joined by the no less compelling behavioural rules of the 'get-grants-or-perish' ideology (**Vannini**, 2006; **Waaijer** *et al.*, 2018).

> Obtaining external research funding is an activity that plays an especially vital role in the process of enhancing scholarly reputation through the production of a research output

Another activity of especially far-fetching reputational ramifications is the collaborative producing of a research output[5]. Research collaboration has long been considered central to the scholarly enterprise by virtue of its singular ability to address in a comprehensive manner and to a synergetic effect complex, critical problems in an era of increasing specialisation (**Leahey**, 2016; **Sonnenwald**, 2007; **Wray**, 2006). Often it is also an inescapable necessity, for the costs of gaining access to expensive instruments, unique scientific data and expertise, scarce natural and social resources, and large amounts of scientific funding prohibit their being borne single-handedly by any one researcher or institution, at times even one country (**Bukvova**, 2010; **Sonnenwald**, 2007). Collaborative knowledge production has been gaining all the more importance in the open science environment, where greater collaboration is seen by its advocates as the key for the future success of research (**Shneiderman**, 2008).

Spurred on by the unprecedented opportunities for cooperative work in today's digital, web-based, socio-technical environment (**Leahey**, 2016), on the one hand, and by many public and private funding agencies' policies that promote collaborative, preferably also interdisciplinary and/or international programmes, on the other (**Breschi**; **Cusmano**, 2004; **Corley**; **Boardman**; **Bozeman**, 2006; **Defazio**; **Lockett**; **Wright**, 2009; **Hoekman** *et al.*, 2012), the last few decades have seen a veritable paradigm shift in scientific research from a singular enterprise into an expanding social endeavour (**Benavent-Pérez** *et al.*, 2012; **Bukvova**, 2010; **Cronin**; **Shaw**; **La-Barre**, 2003; **Freeman**; **Ganguli**; **Murciano-Goroff**, 2014; **Hsieh**, 2013; **Larivière** *et al.*, 2015; **Leahey**, 2016; **Sonnenwald**, 2007; **Wuchty**; **Jones**; **Uzzi**, 2007).

This virtual explosion in collaborative activity over the past decades has had strong individual-level, career- and reputa-

tion- related benefits for the researcher. On the most basic level, collaborating researchers are better placed to achieve visibility in the scholarly community, which is the essential prerequisite of reputation building, simply because they are more people to share information about their work and enlist support for it through the distinct set of relationships that each group member can use for the purpose (**Bikard**; **Murray**; **Gans**, 2015).

Furthermore, collaborative research has repeatedly been found to be associated with greater productivity (**Abramo**; **D'Angelo**; **Di-Costa**, 2009, 2010; **Adams** *et al.*, 2005; **Beaver**; **Rosen**, 1978; **Bordons** *et al.*, 1996; **Price**; **Beaver**, 1966 **Glänzel**; **De-Lange**, 2002; **Landry**; **Traore**; **Godin**, 1996; **Lee**; **Bozeman**, 2005; **Mairesse**; **Turner**, 2005) and wider impact (**Beaver**, 2004; **Bikard**, **Murray**; **Gans**, 2015; **Freeman**; **Ganguli**; **Murciano-Goroff**, 2014; **Katz**; **Martin**, 1997; **Larivière** *et al.*, 2015; **Leahey**, 2016; **Singh**; **Fleming**, 2010; **Van-Raan**, 1998; **Wuchty**; **Jones**; **Uzzi**, 2007), although as **Leahey** (2016) suggests, the coordination costs that accompany collaboration for the individual researcher may at times compromise real productivity. As research productivity coupled with research impactfullness are seen as the main composites of scholarly reputation, and, as such, the goals to be pursued to achieve a stellar standing as a scientist (**Herman**, 2018), scholars have very good reasons for perceiving collaborative research as helpful for realising their aspirations for attaining greater prestige.

Viewed from a reputational angle, collaborative research is also beneficial because a kind of 'reflected glory' can be gained from collaboration with renowned scholars (**Becher**; **Trowler**, 2001; **Kling**; **McKim**, 1999; **Lee**; **Bozeman**, 2005; **Lindgren**, 2011; **Van-Dalen**; **Henkens**, 2001). Indeed, collaboration with better known colleagues within a discipline can also gain stronger 'sponsorship' in proposals to funders, editors and among referees, as well among colleagues who would then be more likely to cite the collaborative publication (**Abramo**; **D'Angelo**; **Murgia**, 2014). No wonder then, as **Martín-Sempere**, **Rey-Rocha**, & **Garzón-García** (2002) find, that researchers belonging to established research groups (unlike those who are affiliated to non-established groups or belong to no group) show higher propensity to international collaboration and to participation in international projects.

Moreover, the 'Matthew-effect'[6] governed systems of science reward the collaborating scholar's above-noted improved research accomplishments and enhanced visibility with additional work and heightened reputation; it is a virtuous circle. Still, as **Abramo**, **D'Angelo** & **Murgia** (2014) contend, the 'Matthew effect' also implies that the merit for a co-authored article will go primarily to the most famous of the scientists credited in the byline, a state of affairs that may have career-related repercussions for a researcher, especially an early career researcher (ECR). However, as a recent longitudinal study into ECRs' scholarly behaviours and attitudes finds, the important reputational asset of being first author is, on the whole, not that difficult for them to attain (**Nicholas**; **Rodríguez-Bravo**; **Watkinson** *et al*., 2017).

5.1.2. Communicating, sharing and networking

As basic tenet of the scholarly world is the interdependence of specialised scientists, who contribute information to one another and receive in exchange the recognition of their colleagues (**Hagstrom**, 1964). Obviously then, scholars accord great importance to the building of a network of connections in order to communicate with likeminded colleagues. It is, in point of fact, as **Becher** (1989) suggests, an inescapable imperative for a scholar, as both the promotion of knowledge (the main cognitive concern) and the establishment of reputation (the key social consideration) are necessarily dependent on communication.

Inevitably then, communicating, sharing and networking –all collaborative sub-activities, of course-, have always been a vital part of the work-life of a scholar, and courtesy of the Web 2.0 enabled possibilities for scholars to congregate virtually in order to share their work, ideas and experiences, the common-interests based bonds among scholars are more easily forged and maintained (**White**; **Le-Cornu**, 2011). Concurrently, as **Copiello** & **Bonifaci** (2018) contend, the advent of Web 2.0 and the growing use of social media by scholars are changing, perhaps even revolutionising our capability to identify those who most contribute to the advancement of science. In consequence, the present-day digital and intercommunicative environment affords more effective ways and means of achieving and managing scholarly reputation than ever. This is all the more important given that all the sharing/communicating/networking activities scholars undertake in the course of their research work have been found in to possess reputation building capabilities, as the selective list of the main activities in Table 1.2 demonstrates (for a full list see **Nicholas**, **Herman**, & **Jamali**, 2015a).

> "Communicating, sharing and networking have always been a vital part of the work-life of a scholar"

A closer look at the prestige-enhancing capabilities of the scholarly communication practices that emerge from the above examples, indicates that they can be, if not already, greatly enriched through scholars' growing propensity to harness the web to engage more openly and in novel ways with colleagues and interested community groups. Traditional communication opportunities, such as face-to-face meetings, telephone conversations or email exchanges, will all support scholars' reputation building efforts, provided that they target their close circle of peers. However, how much more effective could it be, reputation-wise, if the net is spread wider to include the broader scholarly community and even the general public.

Moreover, the spirit of openness, generosity and helpfulness, which is at the heart of the communication-targeted scholarly activities delineated here, is inherently conducive to reputational gains, for being reputable involves, as **Willinsky**

Table 1.2. Communicating, sharing and networking

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Sharing research data with the scholarly community	Enabling other researchers to use extant data for discovering new knowledge faster; inviting collaboration	Achieving disciplinary and trans-disciplinary visibility; obtaining peer recognition and esteem; networking; enhancing one's digital identity	Disseminating data sets – peer to peer or via institutional websites, data centres or repositories
Sharing methodologies, research tools and protocols with the scholarly community	Enabling other researchers to use tried and proven methods for discovering new knowledge; promoting scholarly rigour and scrutiny	Same as above	Making one's working practices transparent and accessible over the web
Providing help for solving problems arising in the course of others' research	Enabling other researchers to discover new knowledge	Achieving disciplinary visibility; obtaining peer recognition and esteem; networking; enhancing one's digital identity	Disseminating information, 'tips', resources etc., peer-to-peer or on social media based scholarly platforms
Sharing research ideas, opinions and interim research findings with disciplinary peers and the wider scholarly community	Obtaining peer feedback and review of one's work, both explicit and implicit[7]; influencing scholarly thinking	Achieving disciplinary and trans-disciplinary visibility; obtaining peer recognition and esteem; networking; enhancing one's digital identity	Interacting with peers in conferences or on social media based scholarly platforms; live tweeting from conferences; blogging

(2010) points out, the possession of precisely such attributes. Indeed, proponents of the open and sharing ideologies of Science 2.0 cite its proven reputation-enhancing outcomes, such as citations, funding and media attention, as the major incentives for its adoption (**McKiernan** *et al.*, 2016). Take, for example, data sharing, which has been shown to attract reputational rewards: in a study which examined 10,555 studies, robust citation benefit was found in the case of those that made data publicly available, compared to those that did not (**Piwowar**; **Vision**, 2013).

Perhaps the best proof of the reputational potentials of communicating and networking online is the meteoric rise of scholarly social networks (SSNs) over the past few years. SSNs, enabling scholars to make new connections, maintain existing ones and showcase achievements, thus extend traditional practices of converting interactions and outputs into reputational terms, and more directly and quickly, too (**Desrochers** *et al.*, 2018; **Hammarfelt**; **De-Rijcke**; **Rushforth**, 2016). Beyond that, SSNs contribute directly to scholarly prestige building via their much-appreciated identity management and profiling functions (**Barbour**; **Marshall**, 2012; **Donelan**, 2016; **Duffy**; **Pooley**, 2017; **Hammarfelt**; **De-Rijcke**; **Rushforth**, 2016; **Jordan**, 2017; **Menéndez**; **De-Angeli**; **Menestrina**, 2012; **Van-Noorden**, 2014).

5.1.3. Disseminating and publishing research findings

The dissemination of research findings is accorded a critical role in the scholarly enterprise, laying the essential foundations for the cooperative, cumulative generation of eventually reliable additions to the stock of human knowledge (**David**; **Den-Besten**; **Schroeder**, 2010). Indeed, the norm calling for the open disclosure of the outcomes of scientific enquiry is one of the basic creeds of the scientific ethos (**Merton**, 1973). Traditionally, the dissemination of scholarly outputs was seen a two-staged process: first the preliminary results of work underway were reported semi-formally or informally to restricted audiences, and, as such, were typically rendered ephemeral and non-retrievable; then the finalised results were reported formally, predominantly published as journal articles, monographs or chapters in edited books, so that the information was made publicly available and remained in permanent storage (**Garvey**; **Griffith**, 1972; **Meadows**, 1998).

> "The dividing lines between formal/informal dissemination of scholarly work have crumbled"

No longer, though: now that a scholar's more or less finalised research outputs are showcased in institutional/disciplinary repositories and on personal websites and SSN profile pages, which serve as complementary distribution channels to publishing in pay-walled or open access scholarly journals and books, the dividing lines between formal/informal dissemination of scholarly work have crumbled. However, as **Kjellberg** & **Haider** (2018) find, rather than challenging the status of the formal scientific publication, the new forms of informal scholarly distribution channels, providing additional indicators of value, afford it more credibility and merit and thereby, reinforce its importance and stabilise its value as the scholarly world's chief currency.

This state of affairs has far fetching reputational ramifications, for, given the primacy of research achievements amongst the measures of scholarly success, the dissemination-fueled showcasing of the results of a scientific investigation has key reputation building roles. In fact, in today's digital and social media focused scholarly realities this showcasing can be undertaken more effectively, as **Herman** (2018) concludes, thanks to the greater visibility, quantifiable information on

research performance and networking and even increased citations afforded by the new dissemination channels, most notably SSNs. No wonder then that scholarly dissemination practices, as identified in the literature, have all been found to possess reputation-building purposes and potentials, as the representative selection of the key activities, presented in Table 1.3, exemplifies (for a full list see **Nicholas**; **Herman**; **Jamali**, 2015a).

Table 1.3. Disseminating and publishing research findings

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Disseminating research results formally	Reporting the results of research for scholarly peers to verify/critique and use	Securing priority for a new contribution; achieving visibility; obtaining recognition and esteem; achieving scholarly impact; enhancing one's digital identity	Publishing research articles in peer reviewed and highly regarded scholarly journals; publishing books with well-regarded publishing houses
Disseminating research results informally to colleagues and disciplinary peers	Same as above	Establishing priority of a new contribution; achieving visibility and obtaining peer recognition and esteem; networking; enhancing one's digital identity	Disseminating manuscripts, pre- or post-prints peer to peer or via repositories, personal websites and SSNs; giving a talk/paper/poster in a conference; blogging; live tweeting from a conference
Disseminating research findings informally to both the disciplinary and the wider scholarly community	Same as above	Same as above; additionally, reaching multiple audiences	Making research findings openly accessible in repositories and on personal websites and SSNs; blogging
Disseminating research findings informally to the public	Popularising science	Achieving public visibility; reaching multiple audiences; enhancing one's digital identity	Posting recorded lectures, pictures or video trailers on popular social media (i.e. *Facebook*); blogging

As the above analysis demonstrates, now that research findings in digital form are so conveniently shared and made visible on the web, realising the reputation building potential of disseminating activities has become easier to achieve. No surprise then that scholars, having become increasingly cognisant of the potential benefits of utilising web and social media afforded ways of showcasing their work (**Dermentzi** *et al.*, 2016; **Haustein** *et al.*, 2014; **Lupton**, 2014; **Nicholas** *et al.*, 2018; **Rowlands** *et al.*, 2011) and ever more aware of the need to build their digital identities (**Duffy**; **Pooley**, 2017; **Meishar-Tal**; **Pieterse**, 2017; **Van-Noorden**, 2014), manifest a much greater degree of readiness to embrace novel platforms and techniques (**Sugimoto** *et al.*, 2017). True, scientists have traditionally been reluctant to engage in public communication for fear of the 'sagan Effect' –the professional stigma attached to spending too much time translating one's research to the broader public and thereby creating 'dumbed-down' science (**Ecklund**; **James**; **Lincoln**, 2012)-, and especially for fear of its coming at the expense of focusing on academic productivity (**Dunwoody**; **Ryan**, 1985). However, according to **Liang** *et al.* (2014) there is scholarly impact enhancing, and therefore reputational value in utilising social media as well as legacy mass communication channels to reach public audiences.

> "The dissemination-fueled showcasing of the results of a scientific investigation has key reputation building roles"

The importance accorded these days to adopting new ways and means of scholarly reputation building is perhaps best exemplified by the diffusion of SSNs in academia. As of November 2018, *ResearchGate* (*RG*) reports to have over 15 million registered users (*ResearchGate*, 2018), whilst *Academia.edu*'s corresponding figure is over 69 million registered users (*Academia.edu*, 2018). Still, active use of SSNs seems to lag (far) behind readiness to register, with studies finding quite low rates of actual use (**Mas-Bleda** *et al.*, 2014; **Nentwich**; **König**, 2014; **Ortega**, 2015). Take, for example, a relatively recent study into *RG* usage and perceptions among academics (particularly in the USA and Europe), which suggests that academics are not very active users of the site, with less than 4 percent reporting usage on a daily basis (daily use is a very high bar, though) and a majority reporting using it only once a week or on a monthly basis (**Muscanel**; **Utz**, 2017).

As **Sugimoto** *et al.* (2017, p. 2052) argue, despite the evident popularity and interest among scholars in social media-based tools, platforms and indicators, only

> "time will tell whether social media and altmetrics are an epiphenomenon of the research landscape, or if they become central to scholars' research dissemination and evaluation practices".

It is important to note here that among ECRs active SSN use has become much more commonplace (*Ciber Research*, 2018), indeed, a mainstream activity, which can arguably foreshadow future developments in the wider scholarly community.

Thus, it is by no mistake that we are talking here of taking up novel ways and means as aids in, rather than alternatives to, the traditional dissemination of research. There can be little doubt that scholars are wary of relinquishing their tried and true formal disseminating practices (**Nicholas** *et al.*, 2017; 2014; *Ciber Research*, 2018; **Tenopir** *et al.*, 2015; **Watkinson** *et al.*, 2016; **Wolff-Eisenberg**; **Rod**; **Schonfeld**, 2016a; 2016b), which, of course, is hardly surprising. Faced with the greater competition resulting from the massive explosion of content and players and well-aware of the career-advancing reputational strengths of traditional dissemination norms and behaviours, scholars cannot but tread carefully where reputation building is concerned.

5.1.4. Evaluating research

With research based on trusted sources, channels and metrics that serve as widely-accepted proxies of the quality and reliability of the knowledge communicated, evaluative activities by necessity have always formed an essential part of its processes (**Becher**, 1989; **Latour**; **Woolgar**, 1986). However, determining the value of research has become ever more crucial with advent of the publish-and-perish driven, competitive and pressured scholarly information environment, in which the quality and dependability of some of the knowledge produced might be questionable (**Bauerlein** *et al.*, 2010; **Casadevall**; **Fang**, 2012; **Colquhoun**, 2011; **Ness**, 2014; **Truex** *et al.*, 2011; **Voas** *et al.*, 2011). Indeed, as producers of information, intent upon making sure that their message is the one attended to, researchers are well-aware that their contributions must first pass muster with their peers (**Franck**, 1999). As keen consumers of information, they are well-aware of the need to assess carefully others' research outputs in order to sift out the wheat from the chaff in the discovery process, in the information management process, and in the citation process (**Tenopir** *et al.*, 2015; **Watkinson** *et al.*, 2016). However, being the avid pursuers of prestige that they are, they are bound to be particularly mindful of the importance of evaluating research performance, both their own and their scholarly peers', with an eye to comparing their endeavours to those of their colleagues. Unsurprisingly, therefore, the evaluative practices that form such an inseparable part of scholarly work have all been found to possess reputation-conferring potential (Table 1.4).

Table 1.4. Evaluating research

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Peer reviewing	Maintaining and improving research quality and rigour	Obtaining peer recognition and esteem	Demonstrating scholarly proficiency and expertise as referee by: appearing on the list of a journal's reviewers; noting reviewing experience on one's CV and website; keeping a validated track-record of contributions as a reviewer on *Publons*
Participating in open peer reviewing	Same as above	Same as above; additionally, achieving visibility; enhancing one's digital identity	Demonstrating scholarly proficiency and expertise via posting reviews of others' research on dedicated sites
Monitoring one's impact	Accruing tangible evidence that one's research work is high quality and trustworthy	Obtaining peer recognition and esteem	Showcasing (for example, on one's website) the scores achieved in: citations-based metrics; usage-based metrics; SSN ratings

Researchers' evaluative practices thus all have reputation-building potentials, although, here again, adherence to traditional perceptions and behavioural rules characterise them. This can be best seen in the importance accorded to peer review, the process in which professional experts (peers) are invited to critically assess the quality, novelty, theoretical and empirical validity, and potential impact of research (**Tennant** *et al.*, 2017). Universally held to be essential for safeguarding the quality and reliability of human knowledge, peer review has been found central to the scholarly enterprise (see, for example, **Harley** *et al.*, 2010; **Mulligan**; **Hall**; **Raphael**, 2013; **Nicholas** *et al.*, 2015b; 2015c; *Nature Publishing Group*, 2015; *Publishing Research Consortium*, 2016; *Research Information network* (*RIN*), 2010; **Rodríguez-Bravo** *et al.*, 2017; *Sense about Science*, 2009; *Taylor & Francis*, 2016; **Ware**; **Monkman**, 2008).

Despite its well-documented pitfalls, among which problematic scientific gatekeeping, reviewer bias, ineffective filtering of error or fraud and the suppression of innovation are arguably the most notable (**Becher**; **Trowler**, 2001; **Bornman**, 2011; **Egghe**; **Bornman**, 2013; **Lee** *et al.*, 2013; *Research Information network* (*RIN*), 2010; **Shatz**, 2004; **Siler**; **Lee**; **Bero**, 2015; **Souder**, 2011; **Weller**, 2001), peer review remains a vital component of the publication-based reward and incentive system of the scholarly enterprise (**Fyfe** *et al.*, 2017; **Moore** *et al.*, 2017; **Tennant**, 2018). Calls to modify the present system to align with web 2.0 technologies abound, with reputational gains, such as greater visibility and digital identity enhancement, cited among the incentives to do so (**Ross-Hellauer**; **Deppe**; **Schmidt**, 2017; **Tennant**, 2018). However, whilst there is considerable scope for new initiatives to be developed, their success is contingent on a significant change of incentives in research environments (**Tennant** *et al.*, 2017). Importantly from a reputational point of view, by now scholars can validate and showcase their contributions to peer review via a dedicated platform – *Publons*. Thus, over time they can build up a public profile of their activity, both as a reviewer and as an editor for different academic journals, which may be used to claim credit –in job or promotion (**Curry**, 2017).

Another area where the reputation building potentials of the present-day scholarly practices, whilst very much evident, are less easily realised, is that of monitoring one's 'impactfullness'. Measuring 'impactfullness', in its original sense, would mean counting the number of people who change their thinking or practice because of some research achievement (**Allen**; **Stanton**; **Di-Pietro** *et al.*, 2013), a manifestly impossible task. Obviously then, determining the impact of a scholar's professional undertakings has to rely on surrogate indicators, with all their limitations –as a recent study demonstrates–, peer judgements of the importance and significance of scholarly work differ from metrics-based measurements (**Borchardt** *et al.*, 2018). As a result, the scholarly world, constantly preoccupied with reputation, has become governed by a 'culture of counting', culminating in a so called 'metric tide' or 'metric deluge' (**Wilsdon** *et al.*, 2015).

After decades of scholarly impact being viewed through the narrow prism of paper productivity and citation-based metrics, none of which seems to be ideally suited to capture scientific impact (**Agarwal** *et al.*, 2016; **Bornmann**; **Daniel**, 2008; **Cronin**, 2013; **Tahamtan**; **Afshar**; **Ahamdzadeh**, 2016; **Waltman**, 2016), researchers are being offered a wider range of metrics, known as altmetrics. Touted as presenting alternative indicators of impact (**Priem** *et al.*, 2010), and, as such, capable of counter-balancing the obsession with and influence of citation-based indicators (**Haustein**, 2016), altmetrics capture impact beyond citations to include speedier measures of the overall usage of more diverse types of scholarly work and even ideas aired in conversations or teaching, as well as their social effects (**Bornmann**, 2014; **Erdt** *et al.*, 2016; **Halevi**; **Schimming**, 2018; **Haustein**; **Larivière**, 2015; **Moed**; **Halevi**, 2015; **Priem**, 2014; **Wouters** *et al.*, 2015).

However, with all these alternative forms of impact measurement do allow for a degree of self-assessment (**Wouters**; **Costas**, 2012), and, indeed, scholars are growingly aware of the added value that altmetrics may have in their scholarly undertakings (**Aung**; **Erdt**; **Theng**, 2017; **Desrochers** *et al.*, 2018; **Haustein** *et al.*, 2014; **Sugimoto** *et al.*, 2017), their behaviour continues to be guided by traditional, productivity- and citation-based metrics (**Kjellberg**; **Haider**, 2018; **Nicholas** *et al.*, 2015a; **Nicholas** *et al.*, 2018; **Tenopir** *et al.*, 2015; **Watkinson** *et al.*, 2016; **Zheng**; **Erdt**; **Theng**, 2018). Inevitably so, of course, with academics typically still recruited, promoted and funded exclusively on the basis of their publication record and citation scores-based reputation (**Alperin** *et al.*, 2018).

5.2. The scholarship of integration

The scholarship of integration, defined as the assembling of extant knowledge into larger intellectual patterns within a wider, often cross-disciplinary context (**Boyer**, 1990), sets out to combine perspectives, concepts, theories, information and data to achieve thorough explorations of complex problems from novel angles. Thus, it aims specifically at producing research outputs that critically analyse, interpret and bring new insight to bear on original research, for example, a review article surveying the salient developments in a field, a textbook, or an article that reports on multi-faceted investigations of a specific topic.

Thus, if in original research the question is 'What is to be known, what is yet to be found'? in integrative research it is rather 'Is it possible to interpret what's been discovered in ways that provide more comprehensive understanding?' (**Boyer**, 1990). Nevertheless, the synthesising research tradition represented by the integrative mode of scholarship is just as much concerned with creating knowledge as the investigative tradition represented by the scholarship of research. In fact, its approach to problem solving is especially suitable for tackling complex, societal often global challenges, which cannot be solved by a single disciplinary approach (**Weller**, 2011). Of course, many of the research activities described in the preceding sections, inclusive of their reputation building capabilities, characterise the scholarship of integration, too. Those practices, unique to the scholarship of integration, which are listed in Table 2, have all been found to have reputational potentials (for a full list see **Nicholas**; **Herman**; **Jamali**, 2015a).

Table 2. Conducting integrative research

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Identifying a complex topic in need of a more wide-ranging understanding and planning the research project to investigate it	Finding a scientifically significant research question and establishing how cross-fertilisation of knowledge can answer it	Producing evidence of scholarly ability to identify the significance of the problem and conduct the research as proposed	Constructing a proposal for interesting collaborators and editors/ publishers
Producing and disseminating an integrative research output using traditional strategies	Discovering and sharing wider- and novel-perspectives afforded new knowledge	Obtaining peer recognition and esteem; achieving disciplinary and trans-disciplinary visibility; achieving scholarly impact	Presenting the results of integrative interpretation of the extant knowledge on a topic in a manner suitable for peer use and evaluation
Producing and disseminating an integrative research output using open and participatory strategies	Same as above; additionally, updating and complementing extant knowledge by current informed opinion	Same as above; additionally, networking; reaching multiple audiences; enhancing one's digital identity	Same as above; additionally, crowd-sourcing and interacting with peers on social media based scholarly platforms

Engaging in the integrative mode of scholarship, effectively showcasing as it does scholarly expertise and proficiency, can thus serve scholars' reputation-building goals. The analytical synthesising of (discipline-spanning) knowledge, which is the raison d'être of integrative scholarship, may give rise to fresh theoretical insights (**Conole** *et al.*, 2010; **Carayol**; **Thi**, 2005), with all the reputation-enhancing implications such an achievement is bound to have. Nevertheless, integrative scholarship has its reputational costs, too, primarily because, almost by definition, it necessitates taking a multi-disciplinary approach.

> Academics still are recruited, promoted and funded exclusively on the basis of their publication record and citation scores-based reputation

The managing of the transition between disciplines can be challenging, as is mastering more than one discipline (**Conole** *et al.*, 2010; **Weller**, 2011). Thus, coordination problems are especially rife in interdisciplinary research, prone as it is to epistemological or methodological conflicts between members of different disciplines (**Leahey**; **Beckman**; **Stanko**, 2012). Complicating things further, peer review, standards of validity and effective criteria of excellence in academe are essentially based on disciplinary standards (**Mallard**; **Lamont**; **Guetzkow**, 2009; **Rafols** *et al.*, 2012). Add to this that prestigious journals tend to be strongly disciplinary (**Weller**, 2011), and interdisciplinary publications are seen as less prestigious (**Conole** *et al.*, 2010), and it becomes obvious why researchers claim that crossing research boundaries comes at a price (**Rhoten**; **Parker**, 2004).

Also, perhaps most notably from a reputational point of view, opting for interdisciplinary projects may bring on a 'production penalty' in a world where success is measured in terms of getting published and cited. Indeed, scholars with greater interdisciplinary research experience have been found to have lower levels of productivity (**Leahey**; **Beckman**; **Stanko**, 2012), possibly because the aforementioned epistemological or methodological conflicts can slow progress toward publication (**Murray**, 2010). The findings as to the relationship between interdisciplinary research and its citation impact are also suggestive, if less decisive: bibliometric studies come up with mixed findings on the topic, possibly because different studies use different operational definitions of interdisciplinarity or because interdisciplinary work can have broad societal and economic impacts that are not captured by citations (**Larivière**; **Gingras**, 2010; **Van-Noorden**, 2015; **Wang**; **Thijs**; **Glänzel**, 2015). Thus, for example, **Larivière** & **Gingras** (2010) found no clear correlation between the level of interdisciplinarity of articles and their citation rates in general, although there were some disciplines in which a higher level of interdisciplinarity was related to a higher citation rates whilst for other disciplines, citations declined as interdisciplinarity grew. It is hardly surprising to find then that disciplinary collaborations contribute more to career development –and hence to reputation- than interdisciplinary collaborations (**Van-Rijnsoever**; **Hessels**, 2011).

5.3. The scholarship of application

Setting out to meet its express aim of informing practice, the scholarship of application (**Boyer**, 1990) utilises disciplinary knowledge and skill to address societal and industrial/organisational challenges. It sees scholars partnering with practitioners, policymakers and community leaders to design application oriented, albeit no less rigorously treated research-based solutions that fruitfully bring together theory and practice. Thus, although scientists and policy-makers alike insist that the distinction between basic research and applied research is increasingly irrelevant and based on misconceptions about modern knowledge production (**Gulbrandsen**; **Kyvik**, 2010), in the context of accelerated international competition, financial austerity and governmental steering of university research towards 'useful' knowledge generation, the scholarship of application has gained an unprecedented significance (**Koryakina**; **Sarrico**; **Teixeira**, 2015). Indeed, as **Blackmore** & **Kandiko** (2011) point out, the work of many scholars includes a wider range of activities –notably 'third stream' or highly applied research that is more closely linked with industry-. In this way universities manage to comply with the imperative of the day, becoming more entrepreneurial and market-oriented in order to obtain legitimacy and to conform to outside pressures, whilst maintaining and supporting the highly valued traditional research activities (**Koryakina**; **Sarrico**; **Teixeira**, 2015).

> Disciplinary collaborations contribute more to career development –and hence to reputation- than interdisciplinary ones

The scholarship of application can be seen as encompassing the area of service work and academic administration, too. This is less surprising than it might seem, for serving the scholarly community, i.e., sitting on committees, fulfilling editorial roles, heading professional organisations, has the practical aspect of furthering the scholarly aims of one's discipline and its local manifestation, to use **Blackmore** & **Kandiko**'s (2011) words, the department. Nevertheless, the ultimate goal of this application-oriented mode of scholarship is the creation of new knowledge, which is why quite a few of the activities comprising the research enterprise are typical of it, too. There are, of course, activities that more uniquely characterise the scholarship of application, as exemplified in the representative list of the key ones among them, presented in Table 3, all of which have been found to possess reputation building potential (for a full list see **Nicholas**; **Herman**; **Jamali**, 2015a).

Linking research-based insights to practice through dynamic interaction, the scholarship of application thus opens up the boundaries between academia and the real world (**Pearce** *et al.*, 2010). Indeed, in these days of Science 2.0 supported initiatives that break down traditional binaries like research/practice, scholar/participant, inside/outside and contribu-

Table 3. Engaging in application-aimed scholarship

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Identifying a societal/ industrial challenge in need of a theory-based practical solution, planning the research project and obtaining funding	Finding a scientifically significant application-oriented research question and establishing its viability	Producing evidence of scholarly ability to identify the significance of the topic and conduct the research as proposed; achieving visibility for one's ideas	Constructing a proposal for interesting peer and practitioner collaborators and for persuading funders that the proposed project can yield the best research on an important topic
Producing and disseminating an application-oriented research output	Discovering new knowledge that offers solutions to a practical problem	Achieving scholarly and public visibility; obtaining peer and public recognition and esteem; achieving scholarly and societal impact	Presenting the results of an application-aimed investigation both in a manner suitable for peer use and evaluation and for mass media dissemination
Serving government or industry as an external consultant	Devising scholarly expertise afforded solutions to societal/ industrial problems	Same as above	Reporting the solutions both in a manner suitable for peer use and evaluation and for mass media dissemination
Serving the scholarly community (i.e., sitting on committees, fulfilling editorial roles, heading professional organisations)	Furthering the aims of one's professional community to better enable the pursuit of scientific goals	Same as above; additionally, networking	Demonstrating proficiency and expertise in scholarly leadership roles via personal and institutional websites; reporting achievements in community functions and publications

tor/user (**Greenhow**; **Gleason**, 2014), application-oriented undertakings are increasingly becoming joint, rather than individual ventures (see also the forthcoming section on the scholarship of co-creation). The ensuing dialogue between scholars and representatives of practitioner/public interests can prove to be advantageous for both parties. For the former, it is the opportunities to open up fresh interconnections between public, scientific, institutional, political and ethical visions of change. For the latter, it is the opportunities for 'sustained dialogue' among groups normally excluded from decision making (**Irwin**, 2008).

From a reputational point of view the great strength of taking on application-aimed scholarly projects is the socio-economic impact they can afford, especially now that furnishing evidence of impact beyond academia is often a requirement in governmental research assessment exercises, as exemplified by the *Research Excellence Framework* (*REF*) in the UK (**Penfield** *et al.*, 2014). By the same token, application-targeted scholarly undertakings can enhance public visibility, which can go a long way towards enhancing scholarly prestige, too. No surprise then that in a survey amongst members of the *American Association for the Advancement of Science* (*AAAS*), the vast majority (87%) supported the idea that participation in policy debates and engagement with citizens was necessary to further their work and careers (**Rainie**; **Funk**; **Anderson**, 2015).

Not that application-aimed, professional/non-professional alliances pose no problems for scholars; rather the contrary. They may have apprehensions about failure for lack of shared language with lay collaborators; they may be concerned about time taken away from 'real' research work, when a lack of time is seen as the most insurmountable barrier to doing more outreach anyway; they may be worried that publicly transparent undertakings may lead to their being 'scooped'. They may also confront a lack of encouragement at the institutional level or a lack of funding for more extensive engagement in the scholarship of application (**Ecklund**; **James**; **Lincoln**, 2012; **Jensen** *et al.*, 2008). However, above all, a major discourager for scholars to take on community-interest driven, application-oriented projects is that the outcomes may remain unpublished (**Braxton**; **Luckey**; **Helland**, 2002). In the scholarly world, where success is measured by publications and citations, such a project is likely to be regarded as too costly in reputational terms.

> "The scholarship of application, with its considerable potential for enhancing public visibility and thereby scholarly reputation, has gained an unprecedented significance"

As to the prestige-conferring capabilities of engaging in the scholarship of application via its service work and academic administration aspect: it is widely held that the holding of managerial/leadership/headship positions in one's collegial community, constituting as it does a very strong source of personal power, influence and respect (**Kekäle**, 2003; **Kogan**, 2007; **Winter**, 2017), serves to enhance a scholar's reputation. Indeed, academic managers, who normally have been appointed to a leadership role in virtue of their superior scholarly achievements and professional competence (**Moodie**; **Eustace**, 1974), are greatly visible and well-known figures, certainly in their own institutions, but very possibly outside of

it, too. As a result, they are more likely to be invited to hold offices in professional organisations, serve on committees, and undertake public-spirited tasks, all of which may serve to further increase their prestige in a virtuous circle brought about by the aforementioned Matthew effect (**Merton**, 1968).

5.4. The scholarship of teaching

Readily understood to refer to the conveying of the human store of knowledge to new generations, the scholarship of teaching, as **Boyer** (1990) sees it, is a more expansive concept than commonly held perceptions indicate. It requires that scientists take a studied approach to their pedagogy in order to achieve evidence-based 'best' teaching practices that can transform, extend and enhance students' learning (**Greenhow**; **Gleason**, 2014). In fact, **Boyer**'s vision of the scholarship of teaching sounds more appealing these days: it is wholly in line with current pedagogical thinking, which puts the student at the heart of the teaching/learning process (**Anderson**, 2016; **Brew**, 2012; **Robson**, 2017; **Weller**, 2016), and readily facilitated by Science 2.0 principles- and technologies-afforded participatory strategies (**Veletsianos**, 2016). The problem is, of course, that the rhetoric with regard to the fundamental importance of both teaching and research has not (yet?) been transported to reality: as **Blackmore** (2016b) puts it, excellence in research attracts prestige, but excellence in teaching does not. As it is only research that is rewarded, there is no incentive for a scholar to spend any more than the minimally required time on teaching and student advising (**Melguizo**; **Strober**, 2007). Indeed, teaching is perceived as simply taking time away from the all-important research, to the extent that ECRs, as a rule, are advised to focus on publishing and avoid spending too much time on any other scholarly pursuit (**Harley**; **Acord**; **Earl-Novell** *et al.*, 2010).

> "A major discourager for scholars to take on community-interest driven, application-oriented projects is that the outcomes may remain unpublished"

However, now that recent policy-level decisions regard the teaching component of scholarly undertakings a global/national/regional priority and call for a sharper focus on teaching and for granting teachers the same professional recognition and opportunities that researchers get (*European Commission*, 2013; *European Parliament*, 2012; **French**, 2017), things might change. Not according to **Blackmore** (2016b), though, who contends that such well-intentioned governmental proposals to recognise and reward teaching excellence may not be sufficient to change the current state of affairs. Instead, he proposes adopting the current pedagogical theories that have at their heart the concept of linking research and teaching closely together at all levels (**Brew**, 2012; **Robson**, 2017; **Weller**, 2016; **Wood**, 2017), so that students take a more active and engaged part in their learning and research-informed teaching can be effectively evaluated in terms of learning success and positive societal effects. The change of climate brought about might culminate in the disproportionate reputational weight given to research above teaching becoming a thing of the past. It is certainly a development to hope for, as scholarly teaching activities, especially those fueled by novel approaches, do seem to possess reputational potentials, as the representative list of the key activities in Table 4 demonstrates (for a full list see **Nicholas**; **Herman**; **Jamali**, 2015a).

Table 4. Engaging in the scholarship of teaching

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Designing a course/learning programme	Establishing how extant knowledge may best be transmitted/shared to promote and support an effective learning process	Producing evidence of disciplinary and pedagogical ability to teach the course/programme as proposed	Constructing a proposal for peer evaluation of its potential effectiveness and for attracting students
Producing and delivering a course using traditional strategies	Achieving effective learning	Obtaining peer and student recognition and esteem	Demonstrating scholarly and pedagogical proficiency and expertise as teacher; excelling in peer monitoring/student ratings of teaching quality
Producing and delivering a course using open and participatory strategies	Same as above	Achieving scholarly and public visibility; obtaining peer, student and public recognition and esteem; enhancing one's digital identity	Same as above; additionally, excelling in public feedback on teaching quality
Engaging in classroom research to advance learning theory	Discovering new pedagogical knowledge and/or achieving enhanced understandings of instructional design	Obtaining peer recognition and esteem	Presenting the results of a scientific investigation in a manner suitable for peer use and evaluation

For the time being teaching activities may be marginalized in the scholarly quest for prestige, but the analysis reported here does show them all to have reputational potentials. Of course, where the specific activity of engaging in classroom research to advance learning theory is concerned, this is hardly surprising; for all practical purposes, classroom research is no different from any other research, affording as it does expert achievements-based eligibility for peer recognition

and esteem through publications. However, the existence of reputation building potentials holds true even where the activities conform to traditional teaching strategies, which, based on the notion of the teacher as the focal point of teaching and centred on lectures delivered either face-to-face or online, are more likely to attract local attention (although in the age of social media word of mouth can spread quickly and widely). Still, it is when teaching is approached, as **Boyer** (1990) suggests and the aforementioned current theories and policies advocate, in a manner similar to research-focused scholarly work, as a disciplinary- and pedagogical-knowledge based, peer-authorised and formally reported undertaking, the outcomes can become all the more reputation accruing.

The affordances of open science 2.0 facilitate the changes in this direction, enabling a shift to learner-centred, qualitatively different, open and participatory practices of teaching, which break out of the confines of the four walls of the classroom to reach multiple and diverse audiences. Thus, for example, the ubiquitous access to an unprecedented wealth of digitised learning resources, brought about by the adoption of open educational resources (OER) policies by a wide variety of governmental, institutional and philanthropic organisations (**Veletsianos**; **Kimmons**, 2012), must have further bolstered the increasingly more prevalent practice of creating open courses and/or making openly available course materials to the public, as well as the many, social-media afforded networked spaces that invite participatory engagement in learning (**Couros**; **Hildebrandt**, 2016; **Cronin; MacLaren**, 2018; **Koseoglu**; **Bozkurt**, 2018; **Veletsianos**, 2016).

The opportunities for reaping the prestige-accruing rewards of excelling as a teacher have grown immeasurably now that web-based, open and participatory teaching strategies, focusing on expert-facilitated dialogue and knowledge exchange among all participants, have come into vogue. This is demonstrated most clearly by MOOCs (massive open online courses) –social networks based, crowd-sourcing technologies enabled, participatory online courses (**Moe**, 2016)-. As **Daniel** (2012) argues, institutions that place their MOOCs in the public domain for a worldwide audience, inevitably have to do more than pay lip service to the importance of teaching and put it at the core of their missions. If so, scholars conducting MOOCs stand to gain twice: their teaching achievements will be taken into career-related consideration, whilst the massive, globe-spanning visibility, which is an inherent feature of MOOCs, will contribute significantly to their scholarly and public visibility driven prestige.

5.5. The scholarship of co-creation

Taking the notions driving much of the current discourse on the nature of contemporary scholarship one step further, **Garnett** and **Ecclesfield** (2012) update **Boyer**'s (1990) model by proposing the addition of a fifth dimension, the scholarship of co-creation. A timely undertaking, indeed, for Boyer's framework, which considers research and teaching as two distinct spheres of activity, and sees the producing of knowledge as a linear process, no longer wholly reflect the realities of the digital and interactive world. The dimension of co-creation refers to the increasingly converging processes of knowledge discovery and knowledge transmission and the resultant blurring of the distinction between the roles of researcher and teacher (**Brew**, 2012; **Robson**, 2017; **Weller**, 2016; **Wood**, 2017).

Beyond the changing face of higher education teaching, as delineated in the previous section, it is public participation in scientific research (PPSR) projects –intentional collaborative endeavors between science researchers and public participants, including but not limited to amateur experts-, concerned community members and/or students, that best embody the spirit of the scholarship of co-creation. Typically designed and led by scientists, with members of the public primarily gathering and analysing data (**Bonney** *et al.*, 2009; **Shirk** *et al.*, 2012), PPSR projects (also known as citizen science projects and community-based participatory research projects), have been gaining traction for the past two decades (**Willyard**; **Scudellari**; **Nordling**, 2018).

The analysis of the activities involved in the participatory and collaborative discovery of new knowledge, as exemplified by PPSR undertakings, shows them to have a strong reputation building capacity, as Table 5 below demonstrates. This is perhaps not surprising: with scholars' various activities in the course of both their research and teaching clearly possessing reputation-accruing potentials, a synergetic effect of their combination is only to be expected.

Table 5. Engaging in the scholarship of co-creation

Activity	Scientific purpose	Reputational purpose	Fit for purpose reputational mechanism
Collaborating in a PPSR (public participation in scientific research) project	Discovering new knowledge that can resolve local concerns; promoting learning about science concepts and processes	Achieving scholarly and public visibility; obtaining peer and public recognition and esteem; achieving scholarly and societal impact	Presenting the results of a PPSR investigation both in a manner suitable for peer use and evaluation and as a societal publication
Leading a PPSR project in a course/learning programme	Same as above; additionally, achieving effective learning about science concepts and processes	Achieving scholarly and public visibility; obtaining peer, student and public recognition and esteem	Same as above; additionally, demonstrating scholarly and pedagogical proficiency and expertise as teacher; excelling in peer monitoring/student ratings of teaching quality

The above examination of what seems to be the most obvious instance of co-creation, the increasingly widespread trend of public participation in scientific research (**Williamson** *et al.*, 2016) demonstrates their strengths in this area. If nothing else, PPSR projects, inviting as they do amateur experts and informed citizens to join the scholarly net and opening the entire process of research to the scrutiny of public collaborators and audiences, can bring about increased visibility-afforded prestige for the scholar.

However, it goes beyond that: since such projects may yield both conventional scientific papers and popular media publications, the scholar stands to gain both peer recognition and esteem and reputation-enhancing societal impact. As **Rotman** *et al.* (2012) have shown, scientists see the opportunity to obtain data on a scale that otherwise might not have been obtainable to support their publishing endeavours as their primary motivation (aside from advancing science) for participating in such co-created endeavors. Obviously, then, they mainly appreciate the reputational, career-advancing potentials of such undertakings. Still, it is interesting to note in this context that according to **Chikoore** *et al.* (2016), the majority of UK academics questioned in their study were opposed to the notion of mandating public engagement with research as part of appraisal systems.

> "Teaching is perceived as simply taking time away from the all-important research"

6. Conclusions

The evidence produced by more than 200 relevant and authoritative papers on scholarly reputation and related matters has been gathered and assessed in the writing of this review. The data has been categorised utilizing a powerful reputational framework designed for the purpose, with the framework providing a lens through which we can examine, almost line by line, the reputational potential of all the practices (nearly 30 of them) that comprise scholars' work-life. In addition, the review produced is advisory in that it demonstrates how scholars may, indeed should, go about building, maintaining and showcasing their reputation.

It clearly emerges from the review that scholarly reputation is still very much associated with research activities. Hardly surprisingly, of course, with recruitment of staff, their career advancements and their further work opportunities widely seen as contingent on proven research achievements, most notably as measured by the quantity of papers published in high-ranking journals and the number of citations they obtain. Thus, although the reputation building component of the scholarly undertaking is potentially very well-supported indeed in this era of Science 2.0, there are still challenges to be faced. The scholar may have strong incentives to embrace more inclusive scholarly goals and to pursue them via open and participatory ways of working, which can provide more encompassing means of achieving and showcasing scholarly reputation, but the reputational price to be paid may be too high.

Thus, as we have seen, the integrative mode of scholarship, effectively showcasing scholarly expertise and proficiency as it does, can successfully serve scholars' reputation-building goals, but this, not without reputational risk. By the same token, both the scholarship of application and the scholarship of co-creation, with their potential to create today's much sought after socio-economic impact, are certainly conducive to prestige, but they, too, can be costly in reputational terms. However, it is the scholarship of teaching, which is the ultimate proof that the reputational price to be paid for participating in novel scholarly undertakings may be too high: with all that creating open courses and/or making openly available course materials to the public can have considerable potential for enhancing a scholar's standing, as these activities cannot be readily translated into conventional research outputs and their effects are mainly felt locally, their reputational value is seen as very limited indeed

> "Activities involved in the participatory and collaborative discovery of new knowledge have a strong reputation building capacity"

It is to be hoped that this analysis of the prestige-accruing potential of scholarly practices can fill the somewhat surprising gap in what we know about this truly vital aspect of the scholarly undertaking. The 'matching' of the hoped-for reputational outcome of an activity with the ways and means at the disposal of scholars for achieving visibility and obtaining peer recognition and esteem, as these emerge from the literature, thus hopefully resulted in a move towards untangling the complex picture of scholarly reputation building.

7. Notes

1. For an exploration of the concept of reputation, roughly definable as the overall judgment of a scholar's standing as determined by experts in their field, see the **Scope and definitions** section.

2. The term 'scholarly' refers in this paper to the characteristics of the research undertaking, the systematic utilising of observation, analysis and/or experimentation to describe and explain social, cultural, medical, natural and agricultural phenomena, in any branch of knowledge. Thus, for all intents and purposes the terms 'scientific' and 'scholarly' are treated as synonyms.

3. So dubbed by **Cronin** (2013) 'for want of a better word'.

4. For example, *IEEE Transactions on education* accepts manuscript submissions under three areas of scholarship, based on Boyer's categories.
https://ieeexplore.ieee.org/xpl/aboutJournal.jsp?punumber=13

5. Collaborating is the action of working with someone to produce/create something; it is cooperation for the sake of achieving something together, whereas networking is the cultivation of relationships in order to enable the ongoing exchange of useful information or services.

6. **Merton**'s (1968) terminology for denoting the pattern of a misallocation of credit for scientific work, whereby greater increments of recognition for scientific contributions are accorded to scientists of considerable repute and such recognition is withheld from scientists who have not (yet) made their mark.

7. Explicit review is the process whereby work is made openly accessible and the audience invited to scrutinise, comment or rate it. Implicit review is the capturing and integrating of usage metadata (page views and downloads, *Twitter* counts, *Facebook* comments, science blog postings, bookmarking and reference sharing), to provide immediate feedback about the performance of a journal, an author or an article.

> "The integrative, application-oriented and co-creation aimed modes integrative mode of scholarship can successfully serve scholars' reputation-building goals, but this not without reputational risk"

8. References

Abramo, Giovanni; **D'Angelo, Ciriaco-Andrea**; **Di-Costa, Flavia** (2009). "Research collaboration and productivity: Is there correlation?". *Higher education*, v. 57, n. 2, pp. 155-171.
https://doi.org/10.1007/s10734-008-9139-z

Abramo, Giovanni; **D'Angelo, Ciriaco-Andrea**; **Di-Costa, Flavia** (2010). "Testing the trade-off between productivity and quality in research activities". *Journal of the Association for Information Science and Technology*, v. 61, n. 1, pp. 132-140.
https://arxiv.org/abs/1811.01815
https://doi.org/10.1002/asi.21254

Abramo, Giovanni; **D'Angelo, Ciriaco-Andrea**; **Murgia, Gianluca** (2014). "Variation in research collaboration patterns across academic ranks". *Scientometrics*, v. 98, n. 3, pp. 2275-2294.
https://arxiv.org/abs/1810.13352
https://doi.org/10.1007/s11192-013-1185-3

Academia.edu (2018). *About*.
http://www.academia.edu/about

Adams, James D.; **Black, Grant C.**; **Clemmons, J. Roger**; **Stephan, Paula E.** (2005). "Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999". *Research policy*, v. 34, n. 3, pp. 259-285.
https://doi.org/10.1016/j.respol.2005.01.014

Agarwal, Ashok; **Durairajanayagam, Damayanthi**; **Tatagari, Sindhuja**; **Esteves, Sandro D.**; **Harlev, Avi**; **Henkel, Ralf**; **Roychoudhury, Shubhadeep**; **Homa, Sheryl**; **Garrido-Puchalt, Nicolás**; **Ramasamy, Ranjith**; **Majzoub, Ahmad**; **Dao-Ly, Kim**; **Tvrda, Eva**; **Assidi, Mourad**; **Kesari, Kavindra**; **Sharma, Reecha**; **Banihani, Saleem**; **Ko, Edmund**; **Abu-Elmagd, Muhammad**; **Gosálvez, Jaime**; **Bashiri, Asher** (2016). "Bibliometrics: Tracking research impact by selecting the appropriate metrics". *Asian journal of andrology*, v.18, n. 2, pp. 296-309.
https://doi.org/10.4103/1008-682X.171582

Åkerlind, Gerlese S. (2008). "An academic perspective on research and being a researcher: An integration of the literature". *Studies in higher education*, v. 33, n. 1, pp. 17-31.
https://formamente.guideassociation.org/wp-content/uploads/2009_1_2_akerlind.pdf
https://doi.org/10.1080/03075070701794775

Allen, Heidi G.; **Stanton, Tasha R.**; **Di-Pietro, Flavia**; **Moseley, G. Lorimer** (2013). "Social media release increases dissemination of original articles in the clinical pain sciences". *PloS one*, v. 8, n. 7, e68914.
https://doi.org/10.1371/journal.pone.0068914

Alperin, Juan-Pablo; **Fischman, Gustavo E.**; **McKiernan, Erin C.**; **Muñoz-Nieves, Carol**; **Niles, Meredith T.**; **Schimanski, Lesley** (2018). "How significant are the public dimensions of faculty work in review, promotion, and tenure documents?". *Humanities commons* [preprint].
https://doi.org/10.17613/M6W950N35

Altbach, Philip G.; **Reisberg, Liz**; **Rumbley, Laura E.** (2009). *Trends in global higher education: Tracking an academic revolution*. A report prepared for the Unesco 2009 World conference on higher education. Chestnut Hill, MA: Boston College Center for International Higher Education.
http://www.cep.edu.rs/public/Altbach,_Reisberg,_Rumbley_Tracking_an_Academic_Revolution,_UNESCO_2009.pdf

Andersen, Heine (2000). "Influence and reputation in the social sciences – how much do researchers agree?". *Journal of documentation*, v. 56, n. 6, pp. 674-692.
https://doi.org/10.1108/EUM0000000007132

Anderson, Derrick M.; **Slade, Catherine P.** (2016). "Managing institutional research advancement: Implications from a university faculty time allocation study". *Research in higher education*, v. 57, n. 1, pp. 99-121.
https://doi.org/10.1007/s11162-015-9376-9

Anderson, Terry (2016). "Theories for learning with emerging technologies". In: Veletsianos, George (ed.). *Emergence and innovation in digital learning: Foundations and applications.* Edmonton: Athabasca University Press, pp. 35-50. ISBN: 978 1 771991490
https://doi.org/10.15215/aupress/9781771991490.01

Aung, Htet-Htet; **Erdt, Mojisola**; **Theng, Yin-Leng** (2017). "Awareness and usage of altmetrics: A user survey". *Proceedings of the Association for Information Science and Technology*, v. 54, n. 1, pp. 18-26.
https://doi.org/10.1002/pra2.2017.14505401003

Auranen, Otto; **Nieminen, Mika** (2010). "University research funding and publication performance — An international comparison". *Research policy*, v. 39, n. 6, pp. 822-834.
https://doi.org/10.1016/j.respol.2010.03.003

Barbour, Kim; **Marshall, David** (2012). "The academic online: Constructing persona through the World Wide Web". *First Monday*, v. 17, n. 9.
https://doi.org/10.5210/fm.v0i0.3969

Bauerlein, Mark; **Gad-el-Hak, Mohamed**; **Grody, Wayne**; **McKelvey, Bill**; **Trimble, Stanley W.** (2010). "We must stop the avalanche of low-quality research". *The chronicle of higher education*, June 13.
http://chronicle.com/article/We-Must-Stop-the-Avalanche-of/65890

Bazeley, Pat (2010). "Conceptualising research performance". *Studies in higher education*, v. 35, n. 8, pp. 889-903.
https://doi.org/10.1080/03075070903348404

Beaver, Donald (2004). "Does collaborative research have greater epistemic authority?". *Scientometrics*, v. 60, n. 3, pp. 399-408.
https://doi.org/10.1023/B:SCIE.0000034382.85360.cd

Beaver, Donald De B.; **Rosen, Richard** (1978). "Studies in scientific collaboration. Part I. The professional origins of scientific co-authorship". *Scientometrics*, v. 1, n. 1, pp. 65-84.
https://doi.org/10.1007/BF02016840

Becher, Tony (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Stony Stratford: Society for Research into Higher Education & Open University Press. ISBN: 978 0 335092215

Becher, Tony; **Trowler Paul** (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Second ed. Buckingham, UK: The Society for Research into Higher Education & Open University Press. ISBN: 978 0 335206278
https://www.mheducation.co.uk/openup/chapters/0335206271.pdf

Benavent-Pérez, María; **Gorraiz, Juan**; **Gumpenberger, Christian**; **De-Moya-Anegón, Félix** (2012). "The different flavors of research collaboration: A case study of their influence on university excellence in four world regions". *Scientometrics*, v. 93, n. 1, pp. 41-58.
https://doi.org/10.1007/s11192-012-0638-4

Bikard, Michaël; **Murray, Fiona**; **Gans, Joshua S.** (2015). "Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward". *Management science*, v. 61, n. 7, pp. 1473-1495.
https://www.nber.org/papers/w18958
https://doi.org/10.1287/mnsc.2014.2052

Blackmore, Paul (2016a). *Prestige in academic life: Excellence and exclusion*. New York: Routledge. ISBN: 978 1 138884944

Blackmore, Paul (2016b). "Why research trumps teaching and what can be done about it". In: Blackmore, Paul; Blackwell, Richard; Edmondson, Martin (eds.). *Tackling wicked issues: Prestige and employment outcomes in the teaching excellence framework*. Oxford: Higher Education Policy Institute.
https://www.hepi.ac.uk/wp-content/uploads/2016/09/Hepi_TTWI-Web.pdf

Blackmore, Paul (2018). "What can policy-makers do with the idea of prestige, to make better policy?". *Policy reviews in higher education*, v. 2, n, 2, pp. 227-254.
https://doi.org/10.1080/23322969.2018.1498300

Blackmore, Paul; **Kandiko, Camille B.** (2011). "Motivation in academic life: A prestige economy". *Research in post-compulsory education*, v. 16, n. 4, pp. 399-411.
https://doi.org/10.1080/13596748.2011.626971

Bloch, Carter; **Graversen, Ebbe-Krogh**; **Pedersen, Heide-Skovgaard** (2014). "Competitive research grants and their impact on career performance". *Minerva*, v. 52, n. 1, pp. 77-96.
https://doi.org/10.1007/s11024-014-9247-0

Bonney, Rick; **Ballard, Heidi**; **Jordan, Rebecca**; **McCallie, Ellen**; **Phillips, Tina**; **Shirk, Jennifer**; **Wilderman, Candie C.** (2009). *Public participation in scientific research: Defining the field and assessing its potential for informal science education*. A Caise Inquiry Group Report.
http://files.eric.ed.gov/fulltext/ED519688.pdf

Borchardt, Rachel; **Moran, Cullen**; **Cantrill, Stuart**; ***Chemjobber***; **Oh, See-Arr**; **Hartings, Matthew R.** (2018). "Perception of the importance of chemistry research papers and comparison to citation rates". *PloS one*, v. 13, n. 3, e0194903.
https://doi.org/10.1371/journal.pone.0194903

Bordons, María; **Gómez-Caridad, Isabel**; **Fernández-Bajón, María-Teresa**; **Zulueta, María-Ángeles**; **Méndez, Aida** (1996). "Local, domestic and international scientific collaboration in biomedical research". *Scientometrics*, v. 37, n. 2, pp. 279-295.
https://doi.org/10.1007/BF02093625

Bornmann, Lutz (2011). "Scientific peer review". *Annual review of information science and technology*, v. 45, n. 1, pp. 197-245.
https://doi.org/10.1002/aris.2011.1440450112

Bornmann, Lutz (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, v. 8, n. 4, pp. 895-903.
https://doi.org/10.1016/j.joi.2014.09.005

Bornmann, Lutz; **Daniel, Hans-Dieter** (2008). "What do citation counts measure? A review of studies on citing behavior". *Journal of documentation*, v. 64, n. 1, pp. 45-80.
https://doi.org/10.1108/00220410810844150

Borrego, Ángel; **Anglada, Lluís** (2016). "Faculty information behaviour in the electronic environment: Attitudes towards searching, publishing and libraries". *New library world*, v. 117, n. 3/4, pp. 173-185.
https://doi.org/10.1108/NLW-11-2015-0089

Bourne, Philip E.; **Barbour, Virginia** (2011). "Ten simple rules for building and maintaining a scientific reputation". *PLoS comput biol.*, v. 7, n. 6, e1002108.
https://doi.org/10.1371/journal.pcbi.1002108

Boyer, Ernest L. (1990). *Scholarship reconsidered: Priorities of the professoriate. A special report*. The Carnegie Foundation for the Advancement of Teaching. San Francisco, California: Jossey-Bass. ISBN: 0787940690
http://www.hadinur.com/paper/BoyerScholarshipReconsidered.pdf

Boyer, Patricia; **Cockriel, Irv** (1998). "Factors influencing grant writing: Perceptions of tenured and nontenured faculty". *SRA Journal*, v. 29, n. 3-4, pp. 61-68.
https://link.galegroup.com/apps/doc/A53643801/AONE?u=googlescholar&sid=AONE&xid=f759dff6

Boyer, Patricia; **Cockriel, Irv** (2001). "Grant performance of junior faculty across disciplines: Motivators and barriers". *Journal of research administration*, v. 2, n. 1, pp. 19-23.
https://bit.ly/2RkHQvr

Braxton, John M.; **Luckey, William**; **Helland, Patricia** (2002). *Institutionalizing a broader view of scholarship through Boyer's four domains*. Ashe-Eric higher education report. San Francisco: Jossey-Bass. ISBN: 0787958417
https://files.eric.ed.gov/fulltext/ED468779.pdf

Breschi, Stefano; **Cusmano, Lucia** (2004). "Unveiling the texture of a European research area: Emergence of oligarchic networks under EU framework programmes". *International journal of technology management*, v. 27, n. 28, pp. 747-772.
https://doi.org/10.1504/IJTM.2004.004992

Brew, Angela (2001). "Conceptions of research: A phenomenographic study". *Studies in higher education*, v. 26, n. 3, pp. 271-285.
https://doi.org/10.1080/03075070120076255

Brew, Angela (2012). "Teaching and research: New relationships and their implications for inquiry-based teaching and learning in higher education". *Higher education research & development*, v. 31, n. 1, pp. 101-114.
https://doi.org/10.1080/07294360.2012.642844

Brewer, Dominic J.; **Gates, Susan M.**; **Goldman, Charles A.** (2001). *In pursuit of prestige: Strategy and competition in U.S. higher education*. New Brunswick, NJ: Transaction Publishers.
https://www.rand.org/pubs/drafts/DRU2541.html

Bukvova, Helena (2010). "Studying research collaboration: A literature review". *Working papers on information systems*, v. 10, n. 3. Sprouts.
https://pdfs.semanticscholar.org/1126/d981037d4a640ac92b3b7e81fe4111f26a6a.pdf

Carayol, Nicolas; **Thi, Thuc-Uyen-Nguyen** (2005). "Why do academic scientists engage in interdisciplinary research?". *Research evaluation*, v. 14, n. 1, pp. 70-79.
http://carayol.u-bordeaux4.fr/interdisciplinarity.pdf
https://doi.org/10.3152/147154405781776355

Casadevall, Arturo; **Fang, Ferric C.** (2012). "Reforming science: Methodological and cultural reforms". *Infection and immunity*, v. 80, n. 3, pp. 891-896.
https://doi.org/10.1128/IAI.06183-11

Chikoore, Lesley; **Probets, Steve**; **Fry, Jenny**; **Creaser, Claire** (2016). "How are UK academics engaging the public with their research? A cross-disciplinary perspective". *Higher education quarterly*, v. 70, n. 2, pp. 145-169.
https://doi.org/10.1111/hequ.12088

Ciber Research (2018). *Harbingers: Third year interim results*.
http://ciber-research.eu/download/20180630-Harbingers_year3_interim_report-early.pdf

Clark, Burton (1998). *Creating entrepreneurial universities: Organizational pathways of transformation*. Oxford, MA: Pergamon Press. ISBN: 978 0 080433547

Colquhoun, David (2011). "Publish or perish: Peer review and the corruption of science". *The guardian*, 5 September.
http://www.theguardian.com/science/2011/sep/05/publish-perish-peer-review-science

Conole, Gráinne; **Scanlon, Eileen**; **Mundin, Paul**; **Farrow, Robert** (2010). *Interdisciplinary research - Findings from the technology enhanced learning research programme*. TLRP, UK.
http://oro.open.ac.uk/35300/1/TELInterdisciplinarity.pdf

Copiello, Sergio; **Bonifaci, Pietro** (2018). "A few remarks on ResearchGate score and academic reputation". *Scientometrics*, v. 114, n. 1, pp. 301-306.
https://doi.org/10.1007/s11192-017-2582-9

Corley, Elizabeth A.; **Boardman, P. Craig**; **Bozeman, Barry** (2006). "Design and the management of multi-institutional research collaborations: Theoretical implications from two case studies". *Research policy*, v. 35, n. 7, pp. 975-993.
https://doi.org/10.1016/j.respol.2006.05.003

Couros, Alec; **Hildebrandt, Katia** (2016). "Designing for open and social learning". In: Veletsianos, George (ed.). *Emergence and innovation in digital learning: Foundations and applications.* Edmonton: Athabasca University Press, pp. 143-161. ISBN: 978 1 771991490
https://doi.org/10.15215/aupress/9781771991490.01

Cronin, Blaise (2013). "Metrics à la mode". *Journal of the Association for Information Science and Technology*, v. 64, n. 6, pp. 1091.
https://doi.org/10.1002/asi.22989

Cronin, Blaise; **Shaw, Debora**; **La-Barre, Kathryn** (2003). "A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy". *Journal of the Association for Information Science and Technology*, v. 54, n. 9, pp. 855-871.
https://doi.org/10.1002/asi.10278

Cronin, Catherine (2017). "Openness and praxis: Exploring the use of open educational practices in higher education". *The international review of research in open and distributed learning*, v. 18, n. 5.
https://doi.org/10.19173/irrodl.v18i5.3096

Cronin, Catherine; **MacLaren, Iain** (2018). "Conceptualising OEP: A review of theoretical and empirical literature in Open Educational Practices". *Open praxis*, v. 10, n. 2, pp. 127-143. Open Education Global Conference Selected Papers.
https://www.learntechlib.org/p/183580

Crow, Robert; **Cruz, Laura**; **Ellern, Jill**; **Ford, George**; **Moss, Hollye**; **White, Barbara-Jo** (2018). "Boyer in the middle: Second generation challenges to emerging scholarship". *Innovative higher education*, v. 43, n. 2, pp. 107-123.
https://doi.org/10.1007/s10755-017-9409-8

Daniel, John (2012). "Making sense of MOOCs: Musings in a maze of myth, paradox and possibility". *Journal of interactive media in education*, v. 2012, n. 3, art. 18.
https://doi.org/10.5334/2012-18

David, Paul A.; **Den-Besten, Matthijs**; **Schroeder, Ralph** (2010). "Will e-science be open science?". In: Dutton, William; Jeffreys, Paul (eds.). *World wide research: Reshaping the sciences and humanities in the century of information*. Siepr Discussion paper N. 08-10. Stanford, CA: Stanford Institute for Economic Policy Research, Stanford University.
http://siepr.stanford.edu/sites/default/files/publications/08-10_1.pdf

Delanty, Gerard (1998). "The idea of the university in the global era: from knowledge as an end to the end of knowledge?". *Social epistemology*, v. 12, n. 1, pp. 3-25.
https://doi.org/10.1080/02691729808578856

De-Rond, Mark; **Miller, Alan N.** (2005). "Publish or perish: Bane or boon of academic life?". *Journal of management inquiry*, v. 14, n. 4, pp. 321-329.
https://pdfs.semanticscholar.org/a640/8c13c67ccb4319e8bf61e83460669381f64a.pdf
https://doi.org/10.1177/1056492605276850

Defazio, Daniela; **Lockett, Andy**; **Wright, Mike** (2009). "Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program". *Research policy*, v, 38, n. 2, pp. 293-305.
https://doi.org/10.1016/j.respol.2008.11.008

Dermentzi, Eleni; **Papagiannidis, Savvas**; **Osorio-Toro, Carlos**; **Yannopoulou, Natalia** (2016). "Academic engagement: Differences between intention to adopt social networking sites and other online technologies". *Computers in human behavior*, v. 61, pp. 321-332.
https://doi.org/10.1016/j.chb.2016.03.019

Desrochers, Nadine; **Paul-Hus, Adèle**; **Haustein, Stefanie**; **Costas, Rodrigo**; **Mongeon, Philippe**; **Quan-Haase, Anabel**; **Bowman, Timothy D.**; **Pecoskie, Jen**; **Tsou, Andrew**; **Larivière, Vincent** (2018). "Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science". *Social science information*, v. 57, n. 2, pp. 223-248.
https://doi.org/ 10.1177/0539018417752089

Dewett, Todd; **Denisi, Angelo S.** (2004). "Exploring scholarly reputation: It's more than just productivity". *Scientometrics*, v. 60, n. 2, pp. 249-272.
https://doi.org/10.1023/B:SCIE.0000027796.55585.6

Donelan, Helen (2016). "Social media for professional development and networking opportunities in academia". *Journal of further and higher education*, v. 40, n. 5, pp. 706-729.
http://oro.open.ac.uk/42255/
https://doi.org/10.1080/0309877X.2015.10143

Duffy, Brooke-Erin; **Pooley, Jefferson D.** (2017). "'Facebook for academics': The convergence of self-branding and social media logic on Academia.edu". *Social media + society*, v. 3, n. 1.
https://doi.org/10.1177/2056305117696523

Dunwoody, Sharon; **Ryan, Michael** (1985). "Scientific barriers to the popularization of science in the mass media". *Journal of communication*, v. 35, n. 1, pp. 26-42.
https://doi.org/10.1111/j.1460-2466.1985.tb01882.x

Ecklund, Elaine H.; **James, Sarah A.**; **Lincoln, Anne E.** (2012). How academic biologists and physicists view science outreach. *PloS one*, v. 7, n. 5, e36240.
https://doi.org/10.1371/journal.pone.0036240

Egghe, Leo; **Bornmann, Lutz** (2013). "Fallout and miss in journal peer review". *Journal of documentation*, v. 69, n. 3, pp. 411-416.
https://doi.org/10.1108/JD-12-2011-0053

Erdt, Mojisola; **Nagarajan, Aarthy**; **Sin, Sei-Ching-Joanna**; **Theng, Yin-Leng** (2016). "Altmetrics: An analysis of the state-of-the-art in measuring research impact on social media". *Scientometrics*, v. 109, n. 2, pp. 1117-1166.
https://doi.org/10.1007/s11192-016-2077-0

Etzkowitz, Henry; **Leydesdorff, Loet** (2000). "The dynamics of innovation: from national systems and "Mode 2" to a triple helix of university–industry–government relations". *Research policy*, v. 29, n. 2, pp. 109-123.
http://www.oni.uerj.br/media/downloads/1-s2.0-S0048733399000554-main.pdf
https://doi.org/10.1016/S0048-7333(99)00055-4

European Commission (2013). *Opening up education: Innovative teaching and learning for all through new technologies and open educational resources*. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*.
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52013DC0654&from=EN

European Commission (2014). *Public consultation 'science 2.0': Science in transition*. Directorates-General for Research and Innovation (RTD) and Communications Networks, Content and Technology (Connect). Background document.
http://ec.europa.eu/research/consultations/science-2.0/background.pdf

European Parliament (2012). *Modernising Europe's higher education systems*. *European Parliament resolution of 20 April 2012 on modernising Europe's higher education systems (2011/2294(INI)*.
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012IP0139&rid=7

Fanelli, Daniele; **Larivière, Vincent** (2016). "Researchers' individual publication rate has not increased in a century". *PloS one*, v. 11, n. 3, e0149504.
https://doi.org/10.1371/journal.pone.0149504

Franck, Georg (1999). "Scientific communication. A vanity fair?". *Science*, v. 286, n. 5437, pp. 53-55.
https://doi.org/10.1126/science.286.5437.53

Freeman, Richard B.; **Ganguli, Ina**; **Murciano-Goroff, Raviv** (2014). *Why and wherefore of increased scientific collaboration* (Working paper no. w19819). National Bureau of Economic Research.
http://www.nber.org/papers/w19819.pdf

French, Amanda (2017). "Contextualising excellence in higher education teaching: Understanding the policy landscape". In: French, Amanda; O'Leary, Matt (eds.). *Teaching excellence in higher education*. Bingley, UK: Emerald Publishing, pp. 5-38. ISBN: 978 1 78714 762 1

Frost, Jetta; **Brockmann, Julia** (2014). "When qualitative productivity is equated with quantitative productivity: Scholars caught in a performance paradox". *Zeitschrift für Erziehungswissenschaft*, v. 17, n. 6, pp. 25-45.
https://doi.org/10.1007/s11618-014-0572-8

Fyfe, Aileen; **Coate, Kelly**; **Curry, Stephen**; **Lawson, Stuart**; **Moxham, Noah**; **Rostvik, Camila** (2017). *Untangling academic publishing: A history of the relationship between commercial interests, academic prestige and the circulation of research*. A briefing paper.
https://doi.org/10.5281/zenodo.546100

Garnett, Fred; **Ecclesfield, Nigel** (2012). Towards a framework for co-creating open scholarship. *Research in learning technology*, v. 19.
https://journal.alt.ac.uk/index.php/rlt/article/view/724

Garvey, William D. (1975). "The dynamic scientific-information user". In: Debons, Anthony; Cameron, William J. (eds). *Perspectives in information science*. NATO Advances study institutes series (Series E: Applied science), v. 10. Dordrecht: Springer.
https://doi.org/10.1007/978-94-011-7759-7_32

Garvey, William D.; **Griffith, Belver** (1972). "Communication and information processing within scientific disciplines: Empirical findings for psychology". *Information storage and retrieval*, v. 8, n. 3, pp. 123-136.
https://doi.org/10.1016/0020-0271(72)90041-1

Gibbons, Michael; **Limoges, Camille**; **Nowotny, Helga**; **Schwartzman, Simon**; **Scott, Peter**; **Trow, Martin** (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies.* London: Sage. ISBN: 0 8039 7794 8

Glänzel, Wolfgang; **De-Lange, Cornelius** (2002). "A distributional approach to multinationality measures of international scientific collaboration". *Scientometrics*, v. 54, n. 1, pp. 75-89.
https://doi.org/10.1023/A:1015684505035

Greenhow, Christine; **Gleason, Benjamin** (2014). "Social scholarship: Reconsidering scholarly practices in the age of social media". *British journal of educational technology*, v. 45, n. 3, pp. 392-402.
https://doi.org/10.1111/bjet.12150

Greenhow, Christine; **Gleason, Benjamin** (2015). "The social scholar: Re-interpreting scholarship in the shifting university". *On the horizon*, v. 23, n. 4, pp. 277-284.
https://doi.org/10.1108/OTH-10-2014-0035

Gulbrandsen, Magnus; **Kyvik, Svein** (2010). "Are the concepts basic research, applied research and experimental development still useful? An empirical investigation among Norwegian academics". *Science and public policy*, v. 37, n. 5, pp. 343-353.
https://doi.org/10.3152/030234210X501171

Hagstrom, Warren O. (1964). "Anomy in scientific communities". *Social problems*, v. 12, n. 2, pp. 186-195.
https://doi.org/10.2307/798981

Hagstrom, Warren O. (1974). "Competition in science". *American sociological review*, v. 39, n. 1, pp. 1-18
https://doi.org/10.2307/2094272

Halevi, Gali; **Schimming, Laura** (2018). "An initiative to track sentiments in altmetrics". *Journal of altmetrics*, v. 1, n. 1, p. 2.
https://doi.org/10.29024/joa.1

Hammarfelt, Björn; **De-Rijcke, Sarah**; **Rushforth, Alexander D.** (2016). "Quantified academic selves: The gamification of science through social networking services". *Information research*, v. 21, n. 2, paper SM1.
http://www.informationr.net/ir/21-2/SM1.html

Harley, Diane; **Acord, Sophia-Krzys**; **Earl-Novell, Sarah**; **Lawrence, Shannon**; **King, C. Judson** (2010). *Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines*. Berkeley: University of California Center for Studies in Higher Education. ISBN: 978 0 615 35834 5
https://escholarship.org/uc/item/15x7385g

Haustein, Stefanie (2016). "Grand challenges in altmetrics: Heterogeneity, data quality and dependencies". *Scientometrics*, v. 108, n. 1, pp. 413-423.
https://doi.org/10.1007/s11192-016-1910-9

Haustein, Stefanie; **Larivière, Vincent** (2015). "The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects". In: Welpe, Isabell M.; Wollersheim, Juta; Ringelhan, Stefanie; Osterloh, Margit (eds). *Incentives and performance*. Cham: Springer, pp. 121-139. ISBN: 978 3 319 09785 5
https://doi.org/10.1007/978-3-319-09785-5_8

Haustein, Stefanie; **Peters, Isabella**; **Bar-Ilan, Judit**; **Priem, Jason**; **Shema, Hadas**; **Terliesner, Jens** (2014). "Coverage and adoption of altmetrics sources in the bibliometric community". *Scientometrics*, v. 101, n. 2, pp. 1145-1163.
https://doi.org/10.1007/s11192-013-1221-3

Heap, Tania; **Minocha, Shailey** (2012). An empirically grounded framework to guide blogging for digital scholarship. *Research in learning technology*, v. 20 (Supp.), pp. 176-188.
https://doi.org/10.3402/rlt.v20i0.19195

Herman, Eti (2018). "Scholarly reputation". *FEMS microbiology letters*, v. 365, n. 18, fny200.
https://doi.org/10.1093/femsle/fny200

Hoekman, Jarno; **Scherngell, Thomas**; **Frenken, Koen**; **Tijssen, Robert** (2012). "Acquisition of European research funds and its effect on international scientific collaboration". *Journal of economic geography*, v. 13, n. 1, pp. 23-52.
https://doi.org/10.1093/jeg/lbs011

Hsieh, David (2013). *Organization and role of international collaboration in research production*. Doctoral dissertation. The University of Arizona.
http://arizona.openrepository.com/arizona/bitstream/10150/281178/1/Hsieh,+David.pdf

Irwin, Alan (2008). "Risk, science and public communication: Third-order thinking about scientific culture". In: Bucchi, Massimiano; Trench, Brian (eds.). *Handbook of public communication of science and technology*. London, UK: Routledge, pp. 199-212. ISBN: 978 1 135049478

Jamali, Hamid R.; **Nicholas, David**; **Herman, Eti** (2016). "Scholarly reputation in the digital age and the role of emerging platforms and mechanisms". *Research evaluation*, v. 25, n. 1, pp. 37-49.
https://doi.org/10.1093/reseval/rvv032

Jensen, Pablo; **Rouquier, Jean-Baptiste**; **Kreimer, Pablo**; **Croissant, Yves** (2008). "Scientists who engage with society perform better academically". *Science and public policy*, v. 35, n. 7, pp. 527-541.
https://arxiv.org/abs/0810.4672
https://doi.org/10.3152/030234208X329130

Jordan, Katy (2017). *Understanding the structure and role of academics' ego-networks on social networking sites*. PhD thesis. Milton Keynes: The Open University.
http://oro.open.ac.uk/48259/1/PhDThesis_KLJ.pdf

Katz, J. Sylvan; **Martin, Ben R.** (1997). "What is research collaboration?". *Research policy*, v. 26, n. 1, pp. 1-18.
https://doi.org/10.1016/S0048-7333(96)00917-1

Kekäle, Jouni (2003). "Academic leaders as thermostats". *Tertiary education & management*, v. 9, n. 4, pp. 281-298.
https://doi.org/10.1023/A:1025822504242

Kjellberg, Sara; **Haider, Juta** (2018). "Researchers' online visibility: Tensions of visibility, trust and reputation". *Online information review*.
https://doi.org/10.1108/OIR-07-2017-0211

Kling, Rob; **McKim, Geoffrey** (1999). "Scholarly communication and the continuum of electronic publishing". *Journal of the American Society for Information Science*, v. 50, n. 10, pp. 890-906.
https://pdfs.semanticscholar.org/2c49/f1dca89d3258c919cbe20c2f6145ebf86982.pdf
https://doi.org/10.1002/(SICI)1097-4571(1999)50:10<890::AID-ASI6>3.0.CO;2-8

Kogan, Maurice (2007). "The academic profession and its interface with management". In: Kogan, Maurice; Teichler, Ulrich (eds.). *Key challenges to the academic profession*. Paris and Kassel: Unesco Forum on Higher Education Research and Knowledge.
https://goo.gl/EiJWD4

Koryakina, Tatyana; **Sarrico, Cláudia S.**; **Teixeira, Pedro N.** (2015). "Universities' third mission activities". In: *The transformation of university institutional and organizational boundaries*. Rotterdam: SensePublishers, pp. 63-82. ISBN: 978 94 6300 178 6
https://doi.org/10.1007/978-94-6300-178-6_4

Koseoglu, Suzan; **Bozkurt, Aras** (2018). "An exploratory literature review on open educational practices". *Distance education*, pp. 1-21.
https://doi.org/10.1080/01587919.2018.1520042

Landry, Réjean; **Traore, Namatie**; **Godin, Benoît** (1996). "An econometric analysis of the effect of collaboration on academic research productivity". *Higher education*, v. 32, n. 3, pp. 283-301.
https://doi.org/10.1007/BF00138868

Larivière, Vincent; **Gingras, Yves** (2010). "On the relationship between interdisciplinarity and scientific impact". *Journal of the American Society for Information Science and Technology*, v. 61, n. 1, pp. 126-131.
https://arxiv.org/pdf/0908.1776.pdf
https://doi.org/10.1002/asi.21226

Larivière, Vincent; **Gingras, Yves**; **Sugimoto, Cassidy R.**; **Tsou, Andrew** (2015). "Team size matters: Collaboration and scientific impact since 1900". *Journal of the Association for Information Science and Technology*, v. 66, n. 7, pp. 323-1332.
https://doi.org/10.1002/asi.23266

Lasthiotakis, Helen; **Kretz, Andrew**; **Sá, Creso** (2015). "Open science strategies in research policies: A comparative exploration of Canada, the US and the UK". *Policy futures in education*, v. 13, n. 8, pp. 968-989.
https://doi.org/10.1177/1478210315579983

Latour, Bruno; **Woolgar, Steve** (1986). *Laboratory life: The construction of scientific facts*. Princeton, New Jersey: Princeton University Press. ISBN: 978 0 691028323

Laudel, Grit (2005). "Is external research funding a valid indicator for research performance?". *Research evaluation*, v. 14, n. 1, pp. 27-34.
https://doi.org/10.3152/147154405781776300

Leahey, Erin (2016). "From sole investigator to team scientist: Trends in the practice and study of research collaboration". *Annual review of sociology*, v. 42, pp. 81-100.
https://doi.org/10.1146/annurev-soc-081715-074219

Leahey, Erin; **Beckman, Christine**; **Stanko, Taryn** (2012). *The impact of interdisciplinarity on scientists' careers*. Paper presented at the annual meeting of the American Sociological Association, Denver, CO.
http://faculty.chicagobooth.edu/workshops/orgs-markets/past/pdf/LeaheySpring2013Impact.pdf

Lee, Carole J.; **Sugimoto, Cassidy R.**; **Zhang, Guo**; **Cronin, Blaise** (2013). "Bias in peer review". *Journal of the American Society for Information Science and Technology*, v. 64, n. 1, pp. 2-17.
https://doi.org/10.1002/asi.22784

Lee, Sooho; **Bozeman, Barry** (2005). "The impact of research collaboration on scientific productivity". *Social studies of science*, v. 35, n.5, pp. 673-702.
https://doi.org/10.1177/0306312705052359

Leydesdorff, Loet; **Etzkowitz, Henry** (1996). "Emergence of a triple helix of university-industry-government relations". *Science and public policy*, v. 23, n. 5, pp. 279-286.
https://doi.org/10.1093/spp/23.5.279

Liang, Xuan; **Su, Leona-Yi-Fan**; **Yeo, Sara K.**; **Scheufele, Dietram A.**; **Brossard, Dominique**; **Xenos, Michael**; **Nealey, Paul**; **Corley, Elizabeth A.** (2014). "Building buzz: (Scientists) communicating science in new media environments". *Journalism & mass communication quarterly*, v. 91, n. 4, pp. 772-791.
https://goo.gl/MSbvLg
https://doi.org/10.1177/1077699014550092

Lindgren, Lena (2011). "If Robert Merton said it, it must be true: A citation analysis in the field of performance measurement". *Evaluation*, v. 17, n. 1, pp. 7-19.
https://doi.org/10.1177/1356389010389908

Lupton, Deborah (2014). *'Feeling better connected': Academics' use of social media*. Report. Canberra: News and Media Research Centre, University of Canberra.
https://www.canberra.edu.au/about-uc/faculties/arts-design/attachments2/pdf/n-and-mrc/Feeling-Better-Connected-report-final.pdf

Mabe, Michael; **Mulligan, Adrian** (2011). "What journal authors want: Ten years of results from Elsevier's author feedback programme". *New review of information networking*, v. 16, n. 1, pp. 71-89.
https://doi.org/10.1080/13614576.2011.574495

Mairesse, Jacques; **Turner, Laure** (2005). *Measurement and explanation of the intensity of co-publication in scientific research: An analysis at the laboratory level* (NBER Working paper n. 11172). Revised in 2010. National Bureau of Economic Research.
https://www.nber.org/papers/w11172
https://doi.org/10.3386/w11172

Mallard, Grégoire; **Lamont, Michèle**; **Guetzkow, Joshua** (2009). "Fairness as appropriateness. Negotiating epistemological differences in peer review". *Science, technology & human values*, v. 34, n. 5, pp. 573-606.
https://goo.gl/wLsFkZ
https://doi.org/10.1177/0162243908329381

Martín-Sempere, María-José; **Rey-Rocha, Jesús**; **Garzón-García, Belén** (2002). "The effect of team consolidation on research collaboration and performance of scientists. Case study of Spanish university researchers in Geology". *Scientometrics*, v. 55, n. 3, pp. 377-394.
https://doi.org/10.1023/A:1020462712923

Mas-Bleda, Amalia; **Thelwall, Mike**; **Kousha, Kayvan**; **Aguillo, Isidro F.** (2014). "Do highly cited researchers successfully use the social web?". *Scientometrics*, v. 101, n. 1, pp. 337-356.
https://doi.org/10.1007/s11192-014-1345-0

McKiernan, Erin C.; **Bourne, Philip E.**; **Brown, C. Titus**; **Buck, Stuart**; **Kenall, Amye**; **Lin, Jennifer**; **McDougall, Damon**; **Nosek, Brian A.**; **Ram, Karthik**; **Soderberg, Courtney K.**; **Spies, Jeffrey R.**; **Thaney, Kaitlin**; **Updegrove, Andrew**; **Woo, Kara H.**; **Yarkoni, Tal** (2016). "How open science helps researchers succeed". *Elife*, n. 5, e16800.
https://doi.org/10.7554/eLife.16800

Meadows, Arthur-Jack (1998). *Communicating research*. San Diego, CA: Academic Press. ISBN: 978 0 124874152

Meishar-Tal, Hagit; **Pieterse, Efrat** (2017). "Why do academics use academic social networking sites?". *The international review of research in open and distributed learning*, v. 18, n. 1, pp. 1-22.
https://doi.org/10.19173/irrodl.v18i1.2643

Melguizo, Tatiana; **Strober, Myra H.** (2007). "Faculty salaries and the maximization of prestige". *Research in higher education*, v. 48, n. 6, pp. 633-668.
https://doi.org/10.1007/s11162-006-9045-0

Menéndez, Maria; **De-Angeli, Antonella**; **Menestrina, Zeno** (2012). "Exploring the virtual space of academia". In: Dugdale, Julia; Masclet, Cédric; Grasso, Maria-Antonietta; Boujut, Jean-François;, Hassanaly, Parina (eds.). *From research to practice in the design of cooperative systems: Results and open challenges*, 49-63. London, UK: Springer. ISBN: 978 1 4471 4093 1
https://doi.org/10.1007/978-1-4471-4093-1_4

Merton, Robert K. (1963). "Resistance to the systematic study of multiple discoveries in science". *European journal of sociology*, v. 4, n. 2, pp. 237-282.
https://doi.org/10.1017/S0003975600000801

Merton, Robert K. (1968). "The Matthew effect in science". *Science*, v. 159, n. 3810, pp. 56-63.
http://www.garfield.library.upenn.edu/merton/matthew1.pdf
https://doi.org/10.1126/science.159.3810.56

Merton, Robert K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: The University of Chicago. ISBN: 978 0 226520926

Miller, Alan N.; **Taylor, Shannon G.**; **Bedeian, Arthur G.** (2011). "Publish or perish: Academic life as management faculty live it". *Career development international*, v. 16, n. 5, pp. 422-445.
https://doi.org/10.1108/13620431111167751

Moe, Rolin (2016). "The phenomenal MOOC". In: Veletsianos, George (ed.). *Emergence and innovation in digital learning: Foundations and applications.* Edmonton: Athabasca University Press, pp. 163-178. ISBN: 978 1 771991490
https://doi.org/10.15215/aupress/9781771991490.01

Moed, Henk F.; **Halevi, Gali** (2015). "Multidimensional assessment of scholarly research impact". *Journal of the Association for Information Science and Technology*, v. 66, n. 10, pp. 1988-2002.
https://arxiv.org/abs/1406.5520
https://doi.org/10.1002/asi.23314

Moodie, Graeme C.; **Eustace, Rowland** (1974). *Power and authority in British universities*. London: George Allen & Unwin. ISBN: 978 1 138006409

Moore, Samuel; **Neylon, Cameron**; **Eve, Martin-Paul**; **O'Donnell, Daniel-Paul**; **Pattinson, Damian** (2017). "'Excellence R Us': University research and the fetishisation of excellence". *Palgrave communications*, n. 3, 16105.
https://doi.org/10.1057/palcomms.2016.105

Mulligan, Adrian; **Hall, Louise**; **Raphael, Ellen** (2013). "Peer review in a changing world: An international study measuring the attitudes of researchers". *Journal of the American Society for Information Science and Technology*, v. 64, n. 1, pp. 132-161.
https://doi.org/10.1002/asi.22798

Murray, Fiona (2010). "The oncomouse that roared: Hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions". *American journal of sociology*, v. 116, n. 2, pp. 341-388.
https://doi.org/10.1086/653599

Muscanell, Nicole; **Utz, Sonja** (2017). "Social networking for scientists: An analysis on how and why academics use ResearchGate". *Online information review*, v. 41, n. 5, pp. 744-759.
https://goo.gl/kHy29A
https://doi.org/10.1108/OIR-07-2016-0185

Nature Publishing Group (2015). *Author insights 2015 survey*.
https://figshare.com/articles/Author_Insights_2015_survey/1425362

Nedeva, Maria; **Boden, Rebecca**; **Nugroho, Yanuar** (2012). "Rank and file: Managing individual performance in university research". *Higher education policy*, v. 25, n. 3, pp. 335-360.
https://doi.org/10.1057/hep.2012.12

Nentwich, Michael; König**, René** (2014). "Academia goes Facebook? The potential of social network sites in the scholarly realm". In: Bartling, Sönke; Friesike, Sascha (eds.). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Cham: Springer. ISBN: 978 3 319 00026 8
https://doi.org/10.1007/978-3-319-00026-8_7

Ness, Roberta B. (2014). *The creativity crisis: Reinventing science to unleash possibility*. New York, NY: Oxford University Press. ISBN: 978 0 199375387

Nicholas, David; **Herman, Eti**; **Clark, David** (2016). "Scholarly reputation building: How does ResearchGate fare". *International journal of knowledge content development and technology*, v. 6, n. 2, pp. 67-92.
https://doi.org/10.5865/IJKCT.2016.6.2.067

Nicholas, David; **Herman, Eti**; **Jamali, Hamid R.** (2015a). *Emerging reputation mechanisms for scholars: A literature-based theoretical framework of scholarly activities and a state-of-the-art appraisal of the social networking services used by scholars, to build, maintain and showcase their reputation*. European Commission, Joint Research Centre, Institute for Prospective Technological Studies. ISBN: 978 92 79 47225 1
http://publications.jrc.ec.europa.eu/repository/bitstream/JRC94955/jrc94955.pdf
https://doi.org/10.2791/891948

Nicholas, David; **Herman, Eti**; **Jamali, Hamid R.** (2015b). "Analysis of emerging reputation mechanisms for scholars". In: Vuorikari, Riina; Punie, Yves (Eds). *Analysis of emerging reputation and funding mechanisms in the context of open Science 2.0*. European Commission, Joint Research Centre, Institute for Prospective Technological Studies, pp. 3-72.
http://www.ciber-research.eu/download/20150521-Reputation_Mechanisms-Final_report-JRC94952.pdf
https://doi.org/10.2791/84669

Nicholas, David; **Herman, Eti**; **Jamali, Hamid R.**; **Rodríguez-Bravo, Blanca**; **Boukacem-Zeghmouri, Cherifa**; **Dobrowolski, Tom**; **Pouchot, Stephanie** (2015a). "New ways of building, showcasing, and measuring scholarly reputation". *Learned publishing*, v. 28, n. 3, pp. 169-183.
https://doi.org/10.1087/20150303

Nicholas, David; **Herman, Eti**; **Xu, Jie**; **Boukacem-Zeghmouri, Cherifa**; **Abrizah, Abdullah**; **Watkinson, Anthony**; **Świgoń, Marzena**; **Rodríguez-Bravo, Blanca** (2018). "Early career researchers' quest for reputation in the digital age". *Journal of scholarly publishing*, v. 49, n. 4, pp. 375-396.
https://doi.org/10.3138/jsp.49.4.01

Nicholas, David; **Jamali, Hamid R.**; **Watkinson, Anthony**; **Herman, Eti**; **Tenopir, Carol**; **Volentine, Rachel**; **Allard, Suzie**; **Levine, Kenneth** (2015b). "Do younger researchers assess trustworthiness differently when deciding what to read and cite and where to publish?". *International journal of knowledge content development and technology*, v. 5, n. 2.
https://doi.org/10.5865/IJKCT.2015.5.2.045

Nicholas, David; **Watkinson, Anthony**; **Jamali, Hamid R.**; **Herman, Eti**; **Tenopir, Carol**; **Volentine, Rachel**; **Allard, Suzie**; **Levine, Kenneth** (2015c). "Peer review: Still king in the digital age". *Learned publishing*, v. 28, n. 1, pp. 15-21.
https://doi.org/10.1087/20150104

Nicholas, David; **Rodríguez-Bravo, Blanca**; **Watkinson, Anthony**; **Boukacem-Zeghmouri, Cherifa**; **Herman, Eti**; **Xu, Jie**; **Abrizah, Abdullah**; **Świgoń, Marzena** (2017). "Early career researchers and their publishing and authorship practices". *Learned publishing*, v. 30, n. 3, pp. 205-217.
https://doi.org/10.1002/leap.1102

Nicholas, David; **Watkinson, Anthony**; **Volentine, Rachel**; **Allard, Suzie**; **Levine, Kenneth**; **Tenopir, Carol**; **Herman, Eti** (2014). "Trust and authority in scholarly communications in the light of the digital transition: Setting the scene for a major study". *Learned publishing*, v. 27, n. 2, pp. 121-134.
http://ciber-research.eu/download/20140406-Learned_Publishing_27_2-Trust.pdf
https://doi.org/10.1087/20140206

Ortega, José-Luis (2015). "Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members". *Journal of informetrics*, v. 9, n. 1, pp. 39-49.
https://doi.org/10.1016/j.joi.2014.11.004

Parra, Cristhian; **Casati, Fabio**; **Daniel, Florian**; **Marchese, Maurizio**; **Cernuzzi, Luca**; **Dumas, Marlon**; **Kungas, Peep**; **García-Bañuelos, Luciano**; **Kisselite, Karina** (2011). "Investigating the nature of scientific reputation". In: *13th Intl Society for Scientometrics and Informetrics Conf.*, Durban, South Africa.
http://www.floriandaniel.it/papers/ParraISSI2011.pdf

Pearce, Nick; **Weller, Martin**; **Scanlon, Eileen**; **Ashleigh, Melanie** (2010). "Digital scholarship considered: How new technologies could transform academic work". *In education*, v. 16, n. 1.
http://ineducation.ca/ineducation/article/view/44/508

Penfield, Teresa; **Baker, Matthew J.**; **Scoble, Rosa**; **Wykes, Michael C.** (2014). "Assessment, evaluations, and definitions of research impact: A review". *Research evaluation*, v. 23, n. 1, pp. 21-32.
https://doi.org/10.1093/reseval/rvt021

Piwowar, Heather A.; **Vision, Todd J.** (2013). "Data reuse and the open data citation advantage". *PeerJ*, 1, e175.
https://doi.org/10.7717/peerj.175

Ponte, Diego; **Simon, Judith** (2011). "Scholarly communication 2.0: Exploring researchers' opinions on Web 2.0 for scientific knowledge creation, evaluation and dissemination". *Serials review*, v. 37, n. 3, pp. 149-156.
https://doi.org/10.1080/00987913.2011.10765376

Price, Derek-John-De-Solla; **Beaver, Donald** (1966). "Collaboration in an invisible college". *American psychologist*, v. 21, n. 11, pp. 1011-1018.
https://www.researchgate.net/publication/17263883_Collaboration_in_an_Invisible_College
https://doi.org/10.1037/h0024051

Priem, Jason (2014). "Altmetrics". In: Cronin, Blaise; Sugimoto, Cassidy R. (eds.). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*. Cambridge, Mass: The MIT Press. ISBN: 978 0 262525510
https://arxiv.org/abs/1507.01328

Priem, Jason; **Taraborelli, Dario**; **Groth, Paul**; **Neylon, Cameron** (2010). *Altmetrics: A manifesto*.
http://altmetrics.org/manifesto

Publishing Research Consortium (2016). *Publishing research consortium peer review survey 2015*. London: Mark Ware Consulting.
http://publishingresearchconsortium.com

Rafols, Ismael; **Leydesdorff, Loet**; **O'Hare, Alice**; **Nightingale, Paul**; **Stirling, Andy** (2012). "How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management". *Research policy*, v. 41, n. 7, pp. 1262-1282.
https://doi.org/10.1016/j.respol.2012.03.015

Rainie, Lee; **Funk, Cary**; **Anderson, Monica** (2015). *How scientists engage the public*. Pew Research Center.
http://www.pewinternet.org/2015/02/15/how-scientists-engage-public

Reif, Fred (1961). "The competitive world of the pure scientist". *Science*, v. 134, n. 3494, pp. 1957-1962.

ResearchGate (2018). *About*.
https://www.researchgate.net/about

Research Information network (2010). *Peer review: A guide for researchers*. Research Information Network.
http://www.rin.ac.uk/system/files/attachments/Peer-review-guide-screen.pdf

Rhoades, Gary (1998). *Managed professionals*. Albany: State University of New York Press. ISBN: 978 0 7914 3716 2

Rhoten, Diana; **Parker, Andrew** (2004). "Risks and rewards of an interdisciplinary research path". *Science*, v. 306, n. 5704, pp. 2046.
https://doi.org/10.1126/science.1103628

Rinne, Risto; **Koivula, Jenni** (2009). "The dilemmas of the changing university". In: Shattock, Michael (ed.). *Entrepreneurialism in universities and the knowledge economy. Diversification and organisational change in European higher education.* London: Open University Press & Paris: IIEP, Unesco, pp. 183-199.

Robson, Sue (2017). "Developing and supporting teaching excellence in higher education". In: French, Amanda; O'Leary, Mat (eds.). *Teaching excellence in higher education*. Bingley, UK: Emerald Publishing, pp. 109-136. ISBN: 978 1 78714 762 1

Rodríguez-Bravo, Blanca; **Nicholas, David**; **Herman, Eti**; **Boukacem-Zeghmouri, Chérifa**; **Watkinson, Anthony**; **Xu, Jie**; **Abrizah, Abdullah**; **Świgoń, Marzena** (2017). "Peer review: The experience and views of early career researchers". *Learned publishing*, v. 30, n. 4, pp. 269-277.
https://doi.org/10.1002/leap.1111

Ross-Hellauer, Tony; **Deppe, Arvid**; **Schmidt, Birgit** (2017). "Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers". *PloS one*, v. 12, n. 12, e0189311.
https://doi.org/10.1371/journal.pone.0189311

Rotman, Dana; **Preece, Jennifer**; **Hammock, Jennifer**; **Procita, Kezee**; **Hansen, Derek**; **Parr, Cynthia**; **Lewis, Darcy**; **Jacobs, David** (2012). "Dynamic changes in motivation in collaborative citizen-science projects". In: *Procs of the ACM 2012 conf on computer supported cooperative work*, pp. 217-226.
http://www.cs.umd.edu/hcil/trs/2011-28/2011-28.pdf

Rowlands, Ian; **Nicholas, David**; **Russell, Bill**; **Canty, Nicholas**; **Watkinson, Anthony** (2011). "Social media use in the research workflow". *Learned publishing*, v. 24, n. 3, pp. 183-195.
https://doi.org/10.1087/20110306

Scanlon, Eileen (2014). "Scholarship in the digital age: Open educational resources, publication and public engagement". *British journal of educational technology*, v. 45, n. 1, pp. 12-23.
https://doi.org/10.1111/bjet.12010

Sense about science (2009). *Peer review survey 2009*.
http://senseaboutscience.org/activities/peer-review-survey-2009

Shatz, David (2004). *Peer review: A critical inquiry.* Lanham, MD: Rowman & Littlefield. ISBN: 978 0 742514355

Shirk, Jennifer L.; **Ballard, Heidi L.**; **Wilderman, Candie C.**; **Phillips, Tina**; **Wiggins, Andrea**; **Jordan, Rebecca**; **McCallie, Ellen**; **Minarchek, Matthew**; **Lewenstein, Bruce V.**; **Krasny, Marianne E.**; **Bonney, Rick** (2012). "Public participation in scientific research: A framework for deliberate design". *Ecology and society*, v. 17, n. 2.
https://doi.org/10.5751/ES-04705-170229

Shneiderman, Ben (2008). "Science 2.0". *Science*, v. 319, n. 5868, pp. 1349-1350.
https://doi.org/10.1126/science.1153539

Siler, Kyle; **Lee, Kirby**; **Bero, Lisa** (2015). "Measuring the effectiveness of scientific gatekeeping". *Proceedings of the National Academy of Sciences*, v. 112, n. 2, pp. 360-365.
https://doi.org/10.1073/pnas.1418218112

Singh, Jasjit; **Fleming, Lee** (2010). "Lone inventors as sources of breakthroughs: Myth or reality?". *Management science*, v. 56, n. 1, pp. 41-56.
https://doi.org/10.1287/mnsc.1090.1072

Slaughter, Sheila; **Leslie, Larry L.** (2001). "Expanding and elaborating the concept of academic capitalism". *Organization*, v. 8, n. 2, pp. 154-161.
https://doi.org/10.1177/1350508401082003

Sonnenwald, Diane H. (2007). "Scientific collaboration". *Annual review of information science and technology*, v. 41, n. 1, pp. 643-681.
https://doi.org/10.1002/aris.2007.1440410121

Souder, Lawrence (2011). "The ethics of scholarly peer review: A review of the literature". *Learned publishing*, v. 24, n. 1, pp. 55-74.
https://doi.org/10.1087/20110109

Storer, Norman W. (1963). "Institutional norms and personal motives in science". In: *The annual meetings of the Eastern Sociological Society*, April 6, New York.

Storer, Norman W. (1966). *The social system of science*. New York: Holt, Rinehart and Winston. ISBN: 978 0 030568657

Sugimoto, Cassidy R.; **Work, Sam**; **Larivière, Vincent**; **Haustein, Stefanie** (2017). "Scholarly use of social media and altmetrics: A review of the literature". *Journal of the Association for Information Science and Technology*, v. 68, n. 9, pp. 2037-2062.
https://arxiv.org/abs/1608.08112
https://doi.org/10.1002/asi.23833

Tahamtan, Iman; **Afshar, Askar-Safipour**; **Ahamdzadeh, Khadijeh** (2016). "Factors affecting number of citations: A comprehensive review of the literature". *Scientometrics*, v. 107, n. 3, pp. 1195-225.
https://doi.org/10.1007/s11192-016-1889-2

Taylor & Francis (2016). *Peer review in 2015: A global view*.
https://authorservices.taylorandfrancis.com/peer-review-global-view

Tennant, Jonathan P. (2018). "The state of the art in peer review". *FEMS Microbiology letters*, v. 365, n. 19, fny204.
https://doi.org/10.1093/femsle/fny204

Tennant, Jonathan; **Dugan, Jonathan M.**; **Graziotin, Daniel**; **Jacques, Damien C.**; **Waldner, François**; **Mietchen, Daniel**; **Elkhatib, Yehia**; **Collister, Lauren B.**; **Pikas, Christina K.**; **Crick, Tom**; **Masuzzo, Paola**; **Caravaggi, Anthony**; **Berg, Devin R.**; **Niemeyer, Kyle E.**; **Ross-Hellauer, Tony**; **Mannheimer, Sara**; **Rigling, Lilian**; **Katz, Daniel S.**; **Greshake-Tzovaras, Bastian**; **Pacheco-Mendoza, Josmel**; **Fatima, Nazeefa**; **Poblet, Marta**; **Isaakidis, Marios**; **Irawan, Dasapta-Erwin**; **Renaut, Sébastien**; **Madan, Christopher R.**; **Matthias, Lisa**; **Kjaer, Jesper-Norgaard**; **O'Donnell, Daniel-Paul**; **Neylon, Cameron**; **Kearns, Sarah**; **Selvaraju, Manojkumar**; **Colomb, Julien** (2017). "A multi-disciplinary perspective on emergent and future innovations in peer review". *F1000 Research*, n. 6, 1151.
https://doi.org/10.12688/f1000research.12037.3

Tenopir, Carol (2016). "Changes in the digital scholarly environment and issues of trust: An exploratory, qualitative analysis". *Information processing & management*, v. 52, n. 3, pp. 446-458.
https://doi.org/10.1016/j.ipm.2015.10.002

Tenopir, Carol; **Levine, Kenneth**; **Allard, Suzie**; **Christian, Lisa**; **Volentine, Rachel**; **Boehm, Reid**; **Nichols, Frances**; **Nicholas, David**; **Jamali, Hamid R.**; **Herman, Eti**; **Watkinson, Anthony** (2015). "Trustworthiness and authority of scholarly information in a digital age: Results of an international questionnaire". *Journal of the Association for Information Science and Technology*, v. 67, n. 10, pp. 2344-2361.
https://doi.org/10.1002/asi.23598

Truex, Duane; **Cuellar, Michael**; **Vidgen, Richard**; **Takeda, Hirotoshi** (2011). "Emancipating scholars: Reconceptualizing scholarly output". In: *7th Intl critical management studies conf* (Cms7 2001). Naples, Italy, July 11-13. Naples: CMS7.
http://www.diva-portal.org/smash/get/diva2:459506/FULLTEXT02

Van-Arensbergen, Pleun; **Van-der-Weijden, Inge**; **Van-den-Besselaar, Peter** (2014). "The selection of talent as a group process. A literature review on the social dynamics of decision making in grant panels". *Research evaluation*, v. 23, n. 4, pp. 298-311.
https://goo.gl/rfaL4t
https://doi.org/10.1093/reseval/rvu017

Van-Dalen, Hendrik P.; **Henkens, Kène** (2001). "What makes a scientific article influential? The case of demographers". *Scientometrics*, v. 50, n. 3, pp. 455-482.
https://doi.org/10.1023/A:1010510831718

Van-Dalen, Hendrik P.; **Henkens, Kène** (2012). "Intended and unintended consequences of a publis-or-perish culture: A worldwide survey". *Journal of the American Society for Information Science and Technology*, v. 63, m. 7, pp. 1282-1293.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1983205
https://doi.org/10.1002/asi.22636

Van-Noorden, Richard (2014). "Online collaboration: Scientists and the social network". *Nature*, v. 512, n. 7513, pp. 126-129.
https://doi.org/10.1038/512126a

Van-Noorden, Richard (2015). "Interdisciplinary research by the numbers". *Nature news*, v. 525, n. 7569, pp. 306-307.
https://goo.gl/BGRVf2

Van-Raan, Anthony F. J. (1998). "The influence of international collaboration of the impact of the research results". *Scientometrics*, v. 42, n. 3, pp. 423-428.
https://doi.org/10.1007/BF02458380

Van-Rijnsoever, Frank J.; **Hessels, Laurens K.** (2011). "Factors associated with disciplinary and interdisciplinary research collaboration". *Research policy*, v. 40, n. 3, pp. 463-472.
https://doi.org/10.1016/j.respol.2010.11.001

Vannini, Phillip (2006). "Dead poets' society: Teaching, publish-or-perish, and professors' experiences of authenticity". *Symbolic interaction*, v. 29, n. 2, pp. 235-57.
https://doi.org/10.1525/si.2006.29.2.235

Veletsianos, George (2016). "The defining characteristics of emerging technologies and emerging practices in digital education". In: Veletsianos, George (ed.). *Emergence and innovation in digital learning: Foundations and applications.* Edmonton: Athabasca University Press, pp. 3-16. ISBN: 978 1 771991490
https://doi.org/10.15215/aupress/9781771991490.01

Veletsianos, George; **Kimmons, Royce** (2012). "Assumptions and challenges of open scholarship". *The international review of research in open and distance learning*, v. 13, n. 4, pp. 166-189.
https://doi.org/10.19173/irrodl.v13i4.1313

Vicente-Sáez, Rubén; **Martínez-Fuentes, Clara** (2018). "Open science now: A systematic literature review for an integrated definition". *Journal of business research*, v. 88, pp. 428-436.
https://doi.org/10.1016/j.jbusres.2017.12.043

Voas, Jeffrey; **Hurlburt, George F.**; **Miller, Keith W.**; **Laplante, Phillip A.**; **Michael, Bret** (2011). "Thoughts on higher education and scientific research". *IT professional*, v. 13, n. 2, pp. 6-9.
https://doi.org/10.1109/MITP.2011.34

Waaijer, Cathelijn J.; **Teelken, Christine**; **Wouters, Paul F.**; **Van-der-Weijden, Inge C.** (2018). "Competition in science: Links between publication pressure, grant pressure and the academic job market". *Higher education policy*, v. 31, n. 2, pp. 225-243.
https://doi.org/10.1057/s41307-017-0051-y

Waltman, Lutz (2016). "A review of the literature on citation impact indicators". *Journal of informetrics*, v. 10, n. 2, pp. 365-91.
https://arxiv.org/abs/1507.02099
https://doi.org/10.1016/j.joi.2016.02.007

Wang, Jian; **Thijs, Bart**; **Glänzel, Wolfang** (2015). "Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity". *PloS one*, v. 10, n. 5, e0127298.
https://doi.org/10.1371/journal.pone.0127298

Ware, Mark; **Monkman, Mike** (2008). *Peer review in scholarly journals. An international study into the perspective of the scholarly community*. Mark Ware Consulting.
https://goo.gl/KZR9vp

Watkinson, Anthony; **Nicholas, David**; **Thornley, Clare**; **Herman, Eti**; **Jamali, Hamid R.**; **Volentine, Rachel**; **Allard, Suzie**; **Levine, Kenneth**; **Tenopir, Carol** (2016). "Changes in the digital scholarly environment and issues of trust: An exploratory, qualitative analysis". *Information processing & management*, v. 52, n. 3, pp. 446-458.
https://doi.org/10.1016/j.ipm.2015.10.002

Weller, Ann (2001). *Editorial peer review: It's strengths and weaknesses.* Medford, NJ: Information Today. ISBN: 978 1 573871006

Weller, Martin (2011). *The digital scholar: How technology is transforming academic practice*. Basingstoke: Bloomsbury Academic. ISBN: 978 1 84966 617 6
http://blog.edtechie.net/wp-content/uploads/2018/02/The-Digital-Scholar_-How-Technology-Is-T-Martin-Weller.pdf

Weller, Saranne (2016). *Academic practice: Developing as a professional in higher education*. London, UK: Sage. ISBN: 978 1 446274231

White, David S.; **Le-Cornu, Alison** (2011). "Visitors and residents: A new typology for online engagement". *First Monday*, v. 16, n. 9.
https://doi.org/10.5210/fm.v16i9.3171

Williamson, Kirsty; **Kennan, Mary-Anne**; **Johanson, Graeme**; **Weckert, John** (2016). "Data sharing for the advancement of science: Overcoming barriers for citizen scientists". *Journal of the Association for Information Science and Technology*, v. 67, n. 10, pp. 2392-2403.
https://researchoutput.csu.edu.au/ws/portalfiles/portal/8971975/1000005591Accepted+Manuscript.pdf
https://doi.org/10.1002/asi.23564

Willinsky, John (2010). "Open access and academic reputation". *Annals of library and information studies*. v. 57, pp. 296-302.
http://nopr.niscair.res.in/bitstream/123456789/10242/4/ALIS%2057%283%29%20296-302.pdf

Willyard, Cassandra; **Scudellari, Megan**; **Nordling, Linda** (2018). "How three research groups are tearing down the ivory tower: The people who should benefit from research are increasingly shaping how it is done". *Nature*, n. 562, pp. 24-28.
https://doi.org/10.1038/d41586-018-06858-4

Wilsdon, James; **Allen, Liz**; **Belfiore, Eleonora**; **Campbell, Philip**; **Curry, Stephen**; **Hill, Steven**; **Jones, Richard**; **Kain, Roger**; **Kerridge, Simon**; **Thelwall, Mike**; **Tinkler, Jane**; **Viney, Ian**; **Wouters, Paul**; **Hill, Jude**; **Johnson, Ben** (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*.
https://doi.org/10.13140/RG.2.1.4929.1363

Wilson, Logan (1942). *The academic man: A study in the sociology of a profession*. New York: Oxford University Press. ISBN: 978 1 560008101

Winter, Richard-Philip (2017). *Managing academics: A question of perspective*. Cheltenham, UK: Edward Elgar Publishing. ISBN: 978 1 781006689

Wolff-Eisenberg, Christine; **Rod, Alisa B.**; **Schonfeld, Roger C.** (2016a). *Ithaka S+R US faculty survey 2015.* New York, NY.: Ithaka S+R.
https://doi.org/10.18665/sr.277685

Wolff-Eisenberg, Christine; **Rod, Alisa B.**; **Schonfeld, Roger C.** (2016b). *UK survey of academics 2015: Ithaka S+R | Jisc | RLUK.* New York, NY.: Ithaka S+R.
https://doi.org/10.18665/sr.2827

Wood, Phil (2017). "From teaching excellence to emergent pedagogies: A complex process alternative to understanding the role of teaching in higher education". In: French, Amanda; O'Leary, Mat (eds.) *Teaching excellence in higher education*. Bingley, UK: Emerald Publishing, pp. 39-74. ISBN: 978 1 78714 762 1

Wouters, Paul; **Costas, Rodrigo** (2012). *Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century*. Utrecht: SURF Foundation.
http://sticonference.org/Proceedings/vol2/Wouters_Users_847.pdf

Wouters, Paul; **Thelwall, Mike**; **Kousha, Kayvan**; **Waltman, Ludo**; **De-Rijcke, Sarah**; **Rushforth, Alex**; **Franssen, Thomas** (2015). *The metric tide: Literature review (Supplementary Report I to the independent review of the role of metrics in research assessment and management)*. The Higher Education Funding Council for England.
https://doi.org/10.13140/RG.2.1.5066.3520

Wray, K. Brad (2006). "Scientific authorship in the age of collaborative research". *Studies in history and philosophy of science. Part A*, v. 37, n. 3, pp. 505-514.
https://doi.org/10.1016/j.shpsa.2005.07.011

Wuchty, Stefan; **Jones, Benjamin F.**; **Uzzi, Brian** (2007). "The increasing dominance of teams in production of knowledge". *Science*, v. 316, n. 5827, pp. 1036-1039.
https://doi.org/10.1126/science.1136099

Zheng, Han; **Erdt, Mojisola**; **Theng, Yin-Leng** (2018). "How do scholars evaluate and promote research outputs? An NTU case study". In: Erdt, Mojisola; Sesagiri-Raamkumar, Aravind; Rasmussen, Edie; Theng, Yin-Leng (eds.). *Altmetrics for research outputs measurement and scholarly information management*. Arosim 2018. Communications in Computer and Information Science, v. 856. Singapore: Springer.
https://doi.org/10.1007/978-981-13-1053-9_6

Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article

Marcia Lei Zeng

How to cite this article:

Zeng, Marcia Lei (2019). "Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article". *El profesional de la información*, v. 28, n. 1, e280103.

https://doi.org/10.3145/epi.2019.ene.03

Article received on December 13th, 2018
Approved on December 28th, 2018

Marcia Lei Zeng ✉
https://orcid.org/0000-0003-0151-5156

Kent State University, School of Information
1125 Risman Drive, 314 University Library
Kent, Ohio, OH 44242-0001, USA
mzeng@kent.edu

Abstract

With the rapid development of the digital humanities (DH) field, demands for historical and cultural heritage data have generated deep interest in the data provided by libraries, archives, and museums (LAMs). In order to enhance LAM data's quality and discoverability while enabling a self-sustaining ecosystem, "semantic enrichment" becomes a strategy increasingly used by LAMs during recent years. This article introduces a number of semantic enrichment methods and efforts that can be applied to LAM data at various levels, aiming to support deeper and wider exploration and use of LAM data in DH research. The real cases, research projects, experiments, and pilot studies shared in this article demonstrate endless potential for LAM data, whether they are structured, semi-structured, or unstructured, regardless of what types of original artifacts carry the data. Following their roadmaps would encourage more effective initiatives and strengthen this effort to maximize LAM data's discoverability, use- and reuse-ability, and their value in the mainstream of DH and Semantic Web.

Keywords

Semantic enrichment; Libraries, archives, and museums; LAMs; Digital humanities; DH; Smart data; Metadata; Structured data; Semi-structured data; Unstructured data; Knowledge discovery; Entity-centric modeling and information access; Data integration and interoperation; Literature review.

1. Introduction

The role of libraries, archives, and museums (LAMs) in supporting digital humanities (DH) research and education has been widely recognized in recent years. In DH research, the difficult part is not typically at the stages of data's cleaning, analysis, visualization, and synthesizing. The most challenging stage is essentially at the beginning, when deciding what and how data can be collected while dealing with historical materials. These items can be documents, artifacts, and other kinds of information-bearing objects. They might be digitized or not-digitized, be textual or non-textual, and exist in all kinds of formats and media. For those researchers in need of historical data that cannot be obtained through web crawling or scraping, the data and information resources provided by LAMs have extraordinary value. The last two deca-

Acknowledgements

The author would like to thank Dawn Sedor for providing valuable feedback and editorial assistance.

des have witnessed a huge investment in digitizing, documenting, and making LAM resources accessible online. In order to enhance LAM data's quality and discoverability while enabling a self-sustaining ecosystem, "semantic enrichment" becomes a strategy increasingly used during recent years.

This article introduces a number of semantic enrichment methods and efforts that can be applied to LAM data at various levels. After the primer explanation of a set of key concepts, the key methods and approaches are explained through the types of data to be enhanced, mainly categorized as structured, semi-structured, and unstructured data. More specifically, these include such methods as: enhancing existing metadata's quality and discoverability with more contextualized meanings by using knowledge organization systems (KOS) vocabularies and other resources that have embraced Linked Open Data (LOD); transforming semi-structured data into structured data through entity-based semantic analysis and annotation to extend access points; digging into unstructured data and generating knowledge bases to support knowledge discovery; enabling one-to-many usages of LAM data across data silos while delivering intuitive user interfaces online; and making the heterogeneous contents from diverse providers semantically interoperable via shared ontology infrastructure. Each section ends with a discussion of representative approaches and additional resources devoted to semantic enrichment. The article concludes with the benchmarks recommended by the *W3C* (2017) in *Data on the Web best practices* which identify the ultimate goals for LAM data: comprehension, processability, discoverability, reuse possibility and effectiveness, trustiness, linkability, accessibility, and interoperability.

2. Key concepts

This review article focuses on semantic enrichment for enhancing LAM data and supporting Digital Humanities. Several key concepts need to be explained before the key approaches and methods are introduced.

Digital humanities (DH) have commanded increasing attention worldwide over the past several years. Although the definitions are being debated and the multifaceted landscape is yet to be fully understood, most agree that initiatives and activities in DH are at the intersection of the disciplines of the humanities and digital information technology. It is at this junction where digital technology will generate a paradigm shift in the near future, enabling scholars to identify major patterns in history, literature, and in the arts. DH refers to new modes of scholarship and institutional units for collaborative, transdisciplinary, and computationally engaged research, teaching, and publication (**Svensson**, 2010; **Burdick** *et al.*, 2012; **Van-Ruyskensvelde**, 2014). It is important to point out that the mere use of digital tools for the purpose of humanistic research and communication does not qualify as DH; nor is DH to be understood as the study of digital artifacts, new media, or contemporary culture in place of physical artifacts, old media, or historical culture (**Burdick** *et al.*, 2012).

One example of the growing DH movement is the *Digging into Data Challenge* program, which has funded dozens of projects aimed at research questions in the humanities and/or social sciences.
https://diggingintodata.org

Figure 1. Key concepts expressed in the project descriptions of *Digging into Data Challenge*.
Source: Created by the author based on project descriptions available at:
https://diggingintodata.org/awards

Since 2009, the number of participating funding organizations and nations has expanded dramatically.[1] Based on the project descriptions of the four rounds (2009, 2011, 2013, and 2016), the resources (see central column "Resources" in Figure 1) vary widely, ranging from unstructured data assets originating in ancient times to structured datasets created in the digital age. Methodologically, the projects are interdisciplinary and strive to show how best to tap into data in large-scale and diverse formats in order to search for key insights, while also ensuring access to such data by researchers through new technology-supported tools. These approaches demonstrate the essential efforts to semantically enrich the data. [Figure 1]

DATA is a concept that needs to be agreed upon when putting data into the context of digital humanities. In the digital age, it is common for people to think of data only in terms of digitally available formats. The connection between digital data and data analytics is correct, but we need to fully understand that the terms "data" and "digital data" are not equivalent. Types of data are also not limited to quantitative data.

The *Reference model for an Open Archival Information System* (*OAIS*) defines data as a

> "reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing"

while offering examples of data as: a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen. This definition of "data" was given within the context of "information," which is

> "Any type of knowledge that can be exchanged. In an exchange, it is represented by data" (*Consultative Committee for Space Data Systems*, 2012, 1-10 and 1-12).

In the book *Information: A very short introduction*, Luciano Floridi defines data at its most basic level as the absence of uniformity, whether in the real world or in some symbolic system. Only once such data have some recognizable structure and are given some meaning can they be considered information (**Floridi**, 2010, pp. 22-25).

After a comprehensive review of the definitions and terminology for "data" in her book titled *Big data, little data, no data: Scholarship in the networked world*, Christine Borgman summarized that

> "data are representations of observations, objects, or other entities used as evidence of phenomena for the purpose of research or scholarship" (**Borgman**, 2015, p. 28).

LAM DATA is a broad term used in this article to refer to all data provided by LAMs and other information institutions. They provide tremendous opportunities for humanities researchers to unearth nuggets of gold from the available data. LAM data can be categorized in three main groups based on their type:

- *Structured data* in LAMs include bibliographies, indexing and abstracting databases, citation indexes, catalogs of all kinds, special collections portals, metadata repositories, curated research datasets, and name authorities. Structured data are typically held in databases in which all key/value pairs have identifiers and clear relations and follow an explicit data model (**Schöch**, 2013).
- *Semi-structured data* in LAMs comprise the unstructured sections within metadata descriptions (e.g., notes in bibliographic records, the rich content descriptions contained in archival finding aids, table-of-contents and abstracts of reports in digital repositories), archival documentation not carried in *Encoded Archival Description* (EAD) or other digital finding aids, intellectual works encoded following the *Text Encoding Initiative* (TEI) guidelines (main text body excluding the Header), value-added or tagged resources that exist in all kinds of formats, and the unstructured portions of otherwise structured datasets.
- *Unstructured data* are

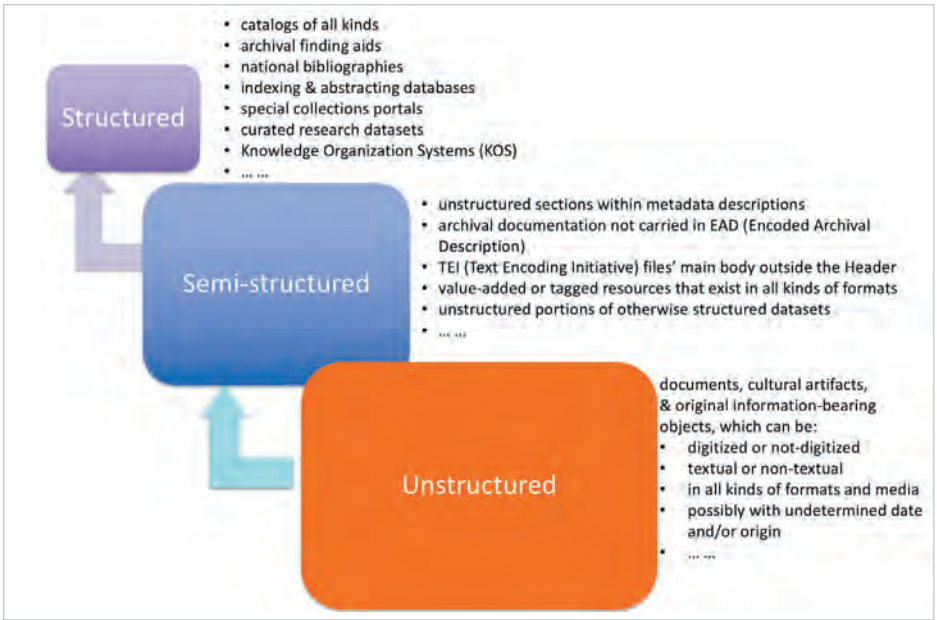


Figure 2. LAM data examples.

characterized as "everything else." In LAMs, they can be found in documents and other information-bearing objects (textual or non-textual, digitized or non-digitized), in all kinds of formats. In data resources that are served through LAMs, unstructured data are usually available in the largest quantity in comparison with structured data, have the most diversity in type, nature, and quality, and are the most challenging to process. [Figure 2]

The primary LAM data assets might be held in special collections, archives, oral history files, annual reports, provenance indexes, and inventories, to name just a few. During the last two decades, many digital collections were born in LAMs and have exhibited strong outcomes. These projects have digitized and delivered integrated resources (metadata, representative images, original documents, and media) on the web. The creation of these digital products involved very complicated digitization and documentation processes requiring tremendous investments from government and other institutions. For example, from 2002 to 2011, in the United States, the federal agency *Institute of Museum and Library Services' IMLS Grants to States* program supplied $980 million to support increased access to digital information, including $67 million toward the digitization of local history and special collections (*IMLS*, 2018, p. 9). Having invested so much, it is critical for LAMs to extend the values of these digital collections beyond being just retrospective resource warehouses in order to have them be better shared, linked, enriched, and reused.

Thus, the next challenge is to move from digitizing to datafying. To "datafy" the unstructured data means to turn the heritage materials into not only machine-readable but also machine-processable resources, and reconstruct materials through digitization pipelines. The demand for datafication might explain why, for digital humanities, the Smart Data approach emphasizes the processes to transform unstructured data to structured and semi-structured data (**Schöch**, 2013; **Mayer-Schönberger**; **Cukier** 2013; **Kaplan**, 2015).

SMART DATA is a concept embraced by humanities research, and underlines the organizing and integrating processes from unstructured data to structured and semi-structured data, making the big data smarter (**Kobielus**, 2016; **Schöch**, 2013). The concept should be understood in the context of Big Data. Among the many "V"s of Big Data (volume, velocity, variety, variability, veracity, and value), the "V"alue of data relies on the ability to achieve big insights from such data on any scale, great or small (**Kobielus**, 2016).

> "[I]n its raw form, data is just like crude oil; it needs to be refined and processed in order to generate real value. Data has to be cleaned, transformed, and analyzed to unlock its hidden potential." (*TiECON East*, 2014).

Only after it has been tamed through the organization and integration processes is such data turned into Smart Data that reflects the research priorities of a particular discipline or field. As Smart Data inquiries, these tamed results can then be used to provide comprehensive analyses and generate new products and services. (**Gardner**, 2012; **Mukerjee**, 2014; **Schöch**, 2013). [Figure 3]

Schöch concluded in his monumental article *Big? Smart? Clean? Messy? Data in the humanities* that we need Smart Big Data because it both adequately represents a sufficient number of relevant features of humanistic objects of inquiry to enable the necessary level of precision and nuance required by humanities scholars, and it provides us with a sufficient amount of data to enable quantitative methods of inquiry, helping us surpass the limitations inherent in methods based on close reading strategies (**Schöch**, 2013). Researchers in the humanities have incorporated the data-driven environment where advanced digital technologies have created the possibility of novel and hybrid methodologies. In the processes that transform unstructured data to structured and semi-structured data, the Smart Data strategy drives data service providers to aim at supporting DH by:

- creating machine-understandable, -processable, and -actionable (instead of merely machine-readable) data;
- providing accurate data in the processes of interlinking, citing, transferring, rights-permission management, use and reuse;
- enabling both one-to-many usages and high efficiency processing of data (**Zeng**, 2017).

The SEMANTIC ENRICHMENT strategy represents one of the major steps to align with the aim of Smart Data. Semantic enrichment is directly applied

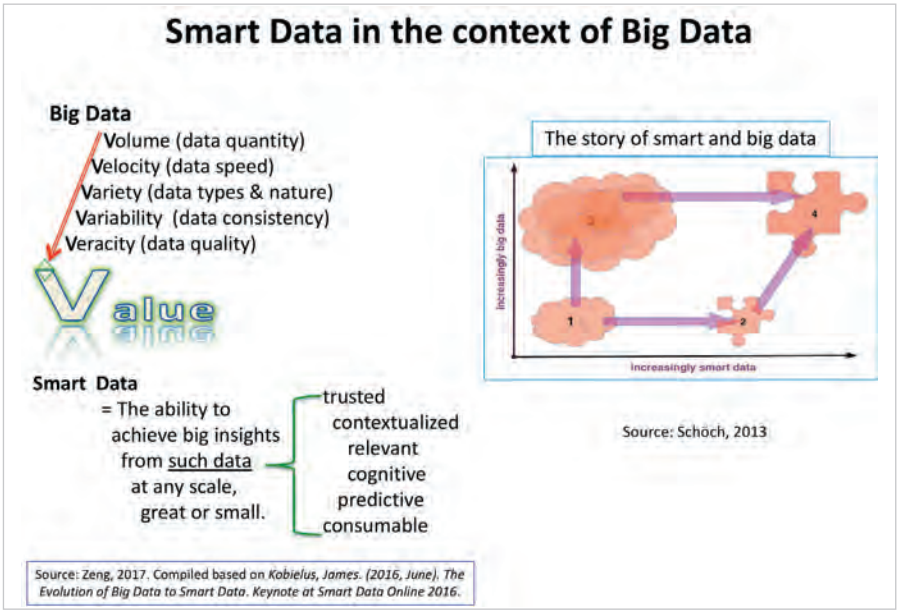


Figure 3. Smart Data in the context of Big Data.

to the enhancement of LAM data (structured, semi-structured, and unstructured) by using semantic technologies. In a broad sense, semantic enrichment in the context of data may aim at different targets. **Damjanovic** *et al.* (2011) surveyed various semantic enhancement approaches and techniques and presented four categories.

- *Semantic search and browsing*, which includes distinct research directions such as (i) augmenting traditional keyword search with semantic techniques, (ii) basic concept location (e.g. multi-facet search, semantic auto-completion, search behavior research), (iii) complex constraint queries for creating query patterns as intuitively as possible, (iv) problem solving, and (v) connecting path discovery.
- *Semantic mediation: merging and mapping*. Merging unifies two or more ontologies with overlapping parts into a single ontology that includes all information from the sources. Mapping builds the mapping statements that define relationships between concepts of ontologies and rules that specify transformations between two ontologies.
- *Semantic annotation*, which formally identifies concepts and relations between concepts in documents, and is intended for use by machines.
- *Semantic analytics and knowledge discovery*. Semantic analytics is a process of analyzing, searching, and presenting information by using explicit semantic relationships between known entities. Both federated and centralized approaches to processing queries on Linked Open Data have been used. (**Damjanovic** *et al.*, 2011)

The article also anticipates the revolutions of Web-based Content Management Systems (WCMS), which replaces in-house-developed CMS for intranet sites and integrates firmly within the Web and document-oriented environments.

Also addressing the differences between knowledge organization in the bibliographic domain and requirements for resource discovery in a web environment, **Prasad**, **Giunchiglia** and **Devika** (2017) presented the *DERA* model (Domain, Entity, Relations, Attributes) featuring the transition from document-centric to entity-centric knowledge modeling. The authors believe that any domain following *DERA* can be seen as formalized by the structure *D<E,R,A>*, where:

- Domain (D) is a particular area of knowledge or field of interest being studied.
- Entity (E) consists of a set of facets where each facet represents a group of terms denoting the entity classes of the real-world entities (instances) having perceptual correlates or only conceptual existence.
- Relation (R) consists of a set of facets where each facet represents a group of terms denoting the relations between entities. Each relation term establishes a semantic relation between two entities.
- Attribute (A) consists of a set of facets where each facet represents a group of terms denoting the qualitative and/or quantitative properties of entities (**Prasad**; **Giunchiglia**; **Devika**, 2017).

It is important to refer to the *International Federation of Library Associations and Institutions* (*IFLA*)'s new *Library Re-*

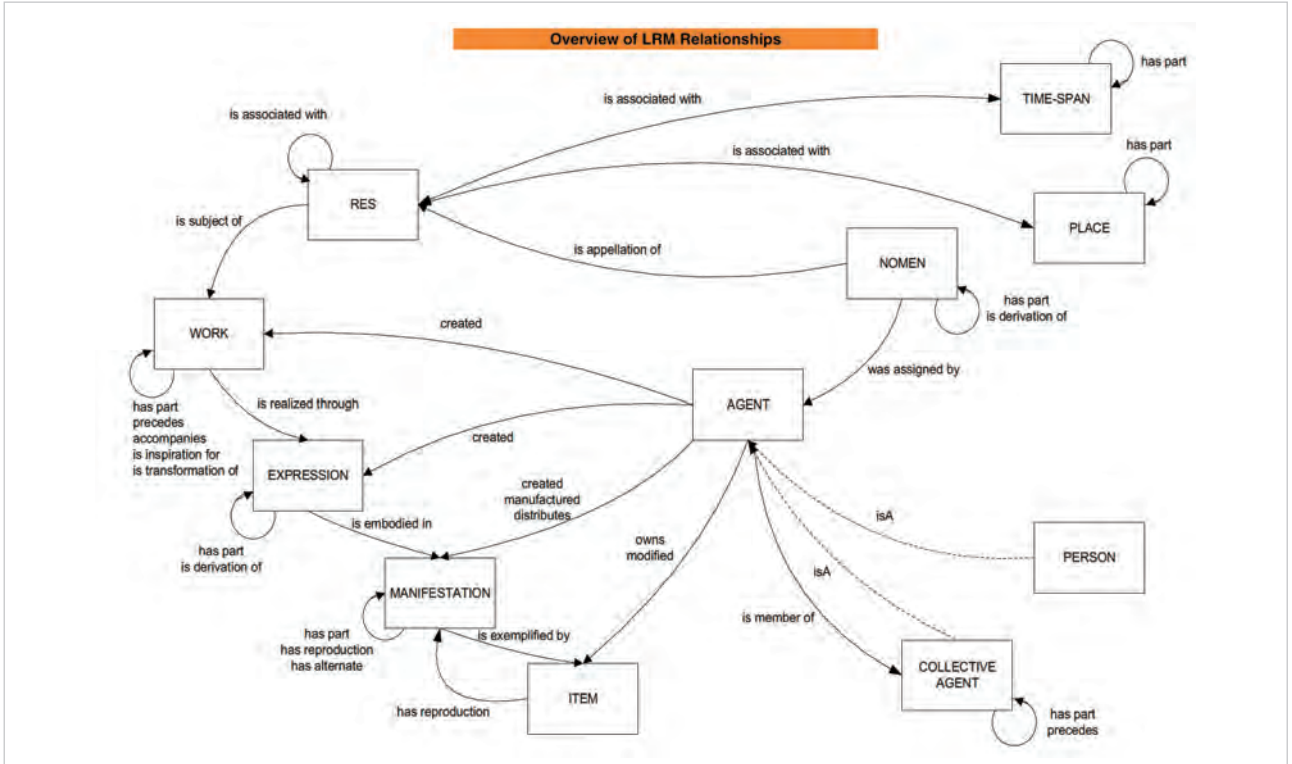


Figure 4. Overview of *IFLA Library Reference Model* (*LRM*) relationships.
Source: **Žumer** and **Riva**, 2017.

ference Model (*IFLA LRM*) in this obvious transition from document-centric to entity-centric modeling trend. *IFLA LRM* consolidated the three models of the *FRBR* Family (*FRBR*,[2] *FRAD*,[3] and *FRSAD*[4]) and *LRM* was formally adopted by the *IFLA Professional Committee* in August 2017 (**Riva**; **LeBoeuf**; **Žumer**, 2017). *LRM* has presented a complete model of the bibliographic universe using entity-relationship modeling. In contrast with the *FRBR* Family, where all entities are at the same level, a hierarchical structure of entities is introduced in *LRM* by declaring entities within a structure of superclasses and subclasses. That one entity is a subclass of another entity can be expressed using the *isA* relationship. For example, *Agent* (an entity capable of deliberate actions, of being granted rights, and of being held accountable for its actions) is declared as a superclass which brings a hierarchical structure to handle the former *FRBR Group 2* of entities. This powerful mechanism enables considerable simplification of the model, because attributes and relationships can be declared on the higher level and do not have to be repeated on lower levels. *Res* (any entity in the universe of discourse) is the superclass of all entities in the model. *Nomen* (an association between an entity and a designation that refers to it) is an entity itself, being the appellation used to refer to an instance of *Res*. In order to model more precisely the temporal and spatial aspects, *LRM* introduces two additional entities, *Place* and *Time-span*. All of these entities presented by *LRM*, as well as the more semantically meaningful attributes and relationships, make *LRM* foundational for the development of cataloguing rules and bibliographic formats (**Žumer**, 2018; **Žumer**; **Riva**, 2017). [Figure 4]

IFLA LRM defines five user tasks (Table 1) and lists the goals users want to reach when performing the tasks. The term resource is used in its broadest sense, meaning an instance of any entity defined in the model. The tasks of *find*, *identify*, *select*, and *obtain* are the same as defined in *FRBR*, with slightly modified and broader definitions; *explore* was first introduced by *FRSAD* (**Žumer**, 2018).

Table 1. A summary of user tasks defined by *LRM*

Find	To bring together information about one or more resources of interest by searching on any relevant criteria
Identify	To clearly understand the nature of the resources found and to distinguish between similar resources
Select	To determine the suitability of the resources found, and to be enabled to either accept or reject specific resources
Obtain	To access the content of the resource
Explore	To discover resources using the relationships between them and thus place the resources in a context

From the perspective of user tasks, it is clear that most of the metadata semantic enrichment efforts extend the initial functions of bibliographic control and enable the *explore* task to be accomplished meaningfully and effectively.

Semantic enrichment process pertaining to LAM data reflects the transformation from document-centric to entity-centric knowledge modeling. The process is distinguished by three main stages in the *Europeana Semantic Enrichment Framework* documentation:

- Analysis: the pre-enrichment phase focuses on the analysis of the metadata fields in the original resource descriptions, the selection of potential resources to be linked to, and derives rules to match and link the original fields to the contextual resource.
- Linking: the process of automatically matching the values of the metadata fields to values of the contextual resources and adding contextual links (whose values are most often based on equivalent relationships) to the dataset.
- Augmentation: the process of selecting the values from the contextual resource to be added to the original object description. This might not only include (multilingual) synonyms of terms to be enriched but also further information, for example broader or narrower concepts (**Isaac** *et al.*, 2015, pp. 8-9; **Manguinhas**, 2016).

> “Data is a concept that needs to be agreed upon when putting data into the context of digital humanities. The connection between digital data and data analytics is correct, but the terms “data” and “digital data” are not equivalent”

The “contextual resources” referred by this documentation and the real cases to be introduced in the next chapter of this article (Section 3.1.1.) signify the selected vocabularies and datasets. For all types of vocabularies and schemes for organizing information and promoting knowledge management, “knowledge organization systems (KOS)” is a broad term that can be used. Examples of KOS include pick lists, authority files, gazetteers, synonym rings, taxonomies and classification schemes, lists of subject headings, thesauri, and ontologies. An instance might be referred to as a “controlled vocabulary,” a “value vocabulary,” or, in a broader sense, a “taxonomy,” by different communities. In this article, when an instance of knowledge organization systems is mentioned, the term KOS vocabulary will be used. For those KOS vocabularies that have been published as Linked Open Data (LOD), they will be referred to as LOD KOS (**Zeng**; **Mayr**, 2018). In the semantic enrichment projects, LOD KOS and other LOD datasets are essential.

Core agents in the semantic enrichment processes are explained according to the concepts identified in the report of the *Europeana Task Force on Enrichment and Evaluation* (**Isaac** *et al.*, 2015). Even though the project focused on structured data, the concepts can also be applied to a wide range of semantic enrichment of LAM data, structured or not.

- At the beginning of each enrichment process, there is the source data that will be enriched. This data comes from different data providers.
- The agent in charge of selecting the different datasets and processing them for enrichment is the enricher - the one who handles the process of enrichment.
- The user of the services made possible or enhanced through enrichment is the end user.
- Sometimes, enrichments can be crowdsourced. In these cases, the volunteers using the crowdsourcing tool and annotating data are the annotators.

3. Key approaches / Methods for LAM data's semantic enrichment

After getting the key concepts explained above, this chapter presents the key approaches and methods for LAM data's semantic enrichment, using cases and experiments reported in the 2010s. These will be explained through the types of data to be enhanced, mainly categorized as structured, semi-structured, and unstructured data.

3.1. Semantic enrichment of structured data

Enriching structured data (often referring to metadata) has become a common initiative in LAM data enhancement efforts, in order to overcome challenges relating to data quality and discoverability in the digital age, while providing more context and multilingual information for cultural heritage (CH) objects. The term "enrichment" may refer to the process, e.g., the application of an enrichment tool, or to its results, such as the new metadata created at the end of the process (**Isaac** *et al.*, 2015). In literature, various terms may reference such methods as reconciliation, mapping, alignment, matching, massaging, merging, interlinking, etc. The overall result is clearly the enrichment of existing metadata, with more contextualized meanings.

Methods and cases discussed in this section have a starting point: the structured LAM data, especially the components where data values are in a controlled /normalized form and are expected to be the access points. As demonstrated by these cases, such data values can reside in original metadata descriptions for CH objects in a LAM data digital platform (e.g., metadata in *Europeana* and *Swissbib*), or exist in agent-centered information clusters (e.g., web pages delivered by the *Museum of Modern Art* (*MoMA*) and the website of the *Museums and Collections with Maya Inscriptions*). The initial "source" and "target" substances involved in the alignment process can be any of these types: metadata descriptions (e.g., *Europeana* metadata), KOS vocabularies and other contextual resources [e.g., *GeoNames*, *VIAF*, *Faceted Application of Subject Terminology* (*FAST*), *Wikidata*, *DBpedia*, etc.], or information resources (e.g., *Wikipedia* entries, biographies, geo-maps) where the focused subjects are the entities in metadata descriptions or KOS vocabularies. Their positions in the alignment (as a "source" or a "target") and the directions of linking can be switched according to a project's design.

3.1.1. Example: structured data in original metadata descriptions for CH objects

Case: *Europeana*

Considered first as an experiment, metadata enrichment has become part of the strategy of *Europeana* and its data providers. The semantic enrichment intends to enrich data providers' metadata by automatically linking text strings found in the metadata to contextual resources from selected LOD datasets or KOS vocabularies. (**Stiller** *et al.*, 2014; **Isaac** *et al.*, 2015) The method comprises augmenting the source metadata with additional terms, seen in two steps:

- Matching the metadata of *Europeana* objects to external semantic data results in links between these objects and resources from external datasets.
- The created links point to additional data such as translated labels or broader labels. A record might be supplemented with all the translated labels of the *DBpedia* concept as well as linking to a broader concept in *DBpedia* and all its translated labels. (**Isaac** *et al.*, 2015)

An example of these processes and results are provided on the *Europeana* semantic enrichment website[5] and demonstrated in Figure 5 in the subsection below.

This automatic linking method is effective when applied to those metadata values that are in a controlled form, including place, agent, concept, and time period. For instance, it may enrich the names of places with values from *GeoNames*, while person names and concepts are enriched with values from *DBpedia* and other vocabularies. The pattern can be simply interpreted as:

"*Europeana* enriches names of … with values from [xxx]", where … and [xxx] can be:

- PLACE [*GeoNames*, *Gemeinsame Normdatei* (*GND*)]
- AGENT [*Virtual International Authority File* (*VIAF*), *The Getty Union List of Artist Names* (*ULAN*), *MIMO Instrument makers*, *GND*, *DBpedia*, etc.]
- CONCEPT [*Art & Architecture Thesaurus* (*AAT*), *Unesco Thesaurus*, *WWI Concepts* from *Library of Congress Subject Headings* (*LCSH*), *Universal Decimal Classification* (*UDC*), *MIMO Concepts*, *IconClass*, *GND*, *Europeana Sounds Genres*, *DBpedia*, etc.]
- TIME PERIOD [*Semium Time*]

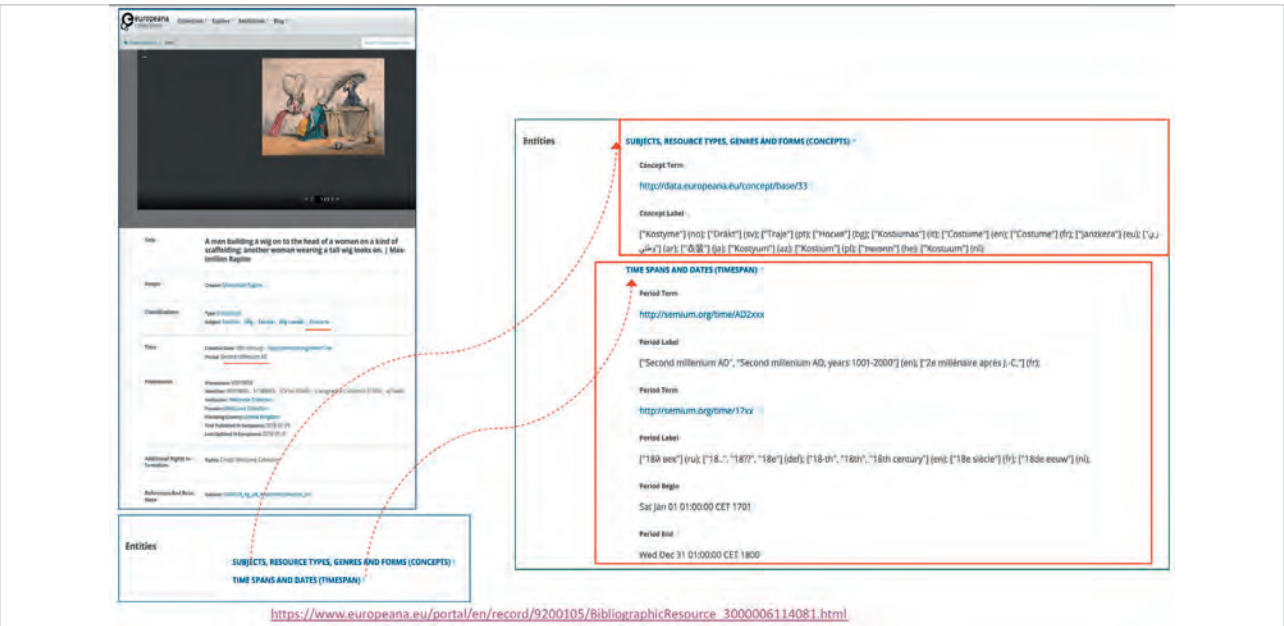


Figure 5. Example of a *Europeana* record semantically enriched with additional data such as translated labels.
Source: Image captured from *Europeana* website at
https://www.Europeana.eu/portal/en/record/9200105/BibliographicResource_3000006114081.html

Those [xxx] represent the external datasets or KOS vocabularies currently chosen by *Europeana* for the automatic alignment process. They are internationally established initiatives or more specific projects whose vocabularies are published as LOD (refer to list of *Europeana Dereferenceable vocabularies*).[6] *Europeana* has developed a tool that 'dereferences' the URIs, i.e., that fetches all the multilingual and semantic data that are published as LOD for vocabulary concepts and other contextual resources on third-party services. *Europeana* encourages participants to use them while also accommodating the participant's own LOD KOS vocabulary/vocabularies.

An enrichment may manifest links that were already implicitly present in the data, as in the case of metadata 'massaging' in *Europeana*. It can be done through advanced mapping or by using tools such as *OpenRefine* where the (string) label of a concept used in an object's metadata is replaced by the identifier of this concept used in its original KOS vocabulary (**Isaac** *et al.*, 2015, p. 11.) Figure 5 shows additional data, such as multilingual labels of the concept from external KOS vocabularies, that result from the enriched metadata section of Entities.
http://openrefine.org

Great progress has led to millions of semantically enriched metadata. Based on experiments conducted around 2015, *Europeana* performed quantitative and qualitative evaluations of seven enrichment services on the same subset of a *Europeana* dataset containing 17300 records. The quantitative overview of the results of the semantic enrichment have been updated by the team (**Manguinhas**, 2016). Millions more have already been added for concepts, places, agents, and time spans since 2017.[7]

Contextualization implies creating "typed relationships" between resources of different types. This process of contextualization involves matching between two objects, two places, or two concepts, e.g., considering whether the match of two concepts are *exactMatch* or *closeMatch*, whether a concept is *broader* or *narrower* than the one aligned to, or whether the two identifiers from two sources actually represent the same place. The *Europeana Data Model* (*EDM*) is the core for the defined properties which express various types of relationships in the alignment results. The Figure 6 illustrates some of the types of links specified. [Figure 6]

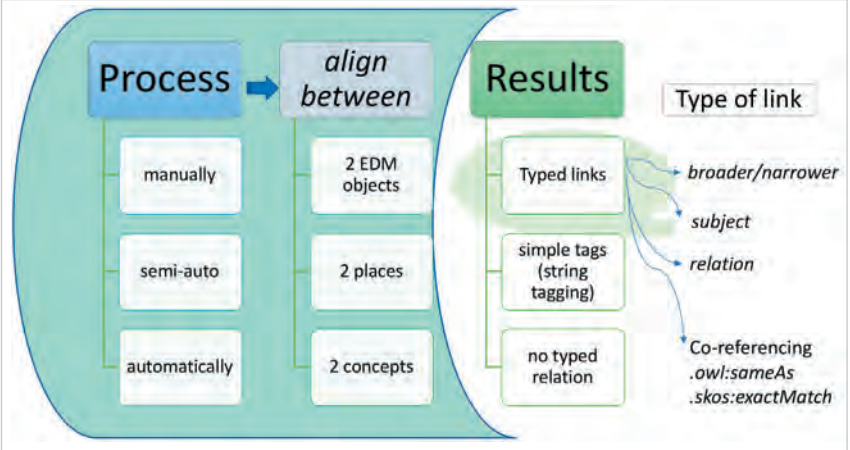


Figure 6. Contextualization - creating typed relationships between resources of different types.
Source: Image based on **Isaac** *et al.*, 2015.

Case: *Swissbib*

The enrichment of structured data may be centralized on a specific type of entity, for example, authors. *Swissbib*, a provider for bibliographic data in Switzerland, reported a process for extracting and shaping the data into a more suitable form (see Figure 8). In one example, data available in MARC21 XML were extracted from the *Swissbib* system and transformed into an RDF/XML representation. In another exercise, author information was extracted from approximately 21 million monolithic records and interlinked with authority files from the *VIAF* and *DBpedia*. [Figure 7]
http://linked.swissbib.ch

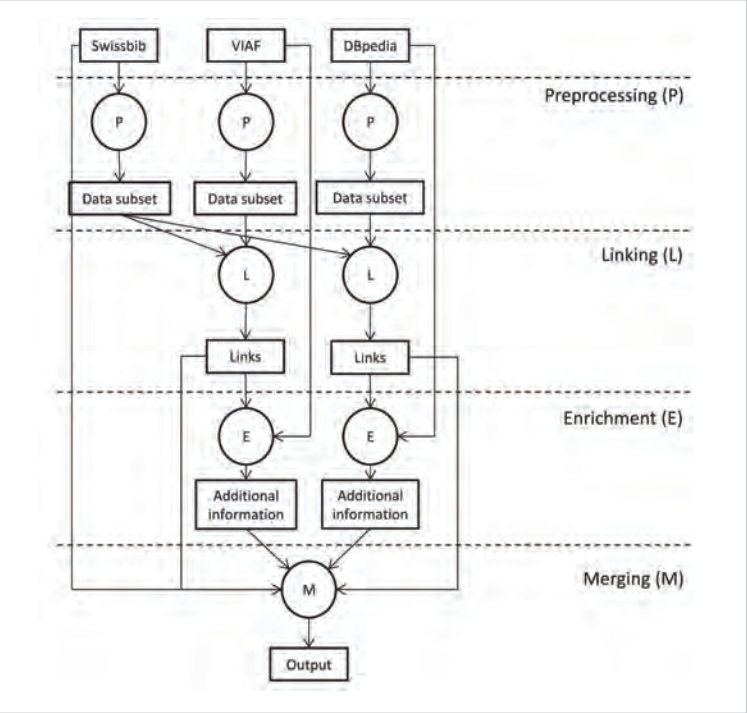


Figure 7. Data flow diagram of the interlinking procedure in the *Swissbib* project
Source: **Bensmann**; **Zapilko**; **Mayr**, 2017, p. 8.

In the enrichment process, the researchers take the links <swissbib> <owl:sameAs> <viaf>, sort them by their object and align them with the respective external corpus. In this way, selected statements about the referenced author are extracted from the external corpus and rewritten to make them statements of the *Swissbib* author resource. Particular statements may refer to further resources instead of literals, e.g., locations. In order to be able to display these resources on the user interface, these resources are summarized into a single literal that represents the resource in a suitable manner, e.g. using labels or descriptions. The resulting literal is additional to the original property, added to the person description using a new extended property (*dbp:birthPlace->swissbib:dbpBirthPlaceAsLiteral*). Finally, all persons, together with the links and extracted data, are deposited at an agreed location for indexing. The approach established 30,773 links to *DBpedia* and 20,714 links to *VIAF*, and both link sets show high precision values and were generated in reasonable expenditures of time, according to the authors (**Bensmann**; **Zapilko**; **Mayr**, 2017).

3.1.2. Example: structured data in agent-centered information clusters on the Web

Case: *Museum of Modern Art* (*MoMA*)

A unique named-entity-centered case is the information cluster on website of the *Museum of Modern Art* (*MoMA*) in New York City, United States. The *Museum*'s website features 72,706 artworks from 20,956 artists (*MoMA*, 2017). In addition to web pages, two types of datasets are openly available:[8]

(1) The artworks dataset (artworks.csv) contains more than 130,000 records, including basic metadata for each work, such as title, artist, date, medium, dimensions, and date acquired by the *Museum*.

(2) The artists dataset (artists.csv) contains more than 15,000 records, representing all artists who have work featured in *MoMA*'s collection and have been cataloged in the database. The structured data for each artist includes name, nationality, gender, birth year, and death year. By mapping the artists dataset to the *Union List of Artist Names* (*ULAN*) data through *Getty Vocabularies*: LOD,[9] the majority of the artists in the *MoMA* dataset obtained http URIs from the *ULAN*. [Figure 8]

From the front-end of the webpage which shares information about the artist as well as related exhibitions (Figure 8, center), it is clear that the artist's information is greatly enriched by the high-quality *ULAN* name authority data, which include the information about an artist, such as: name, bio, nationality, role, type, and multilingual appellations (Figure 8 right). If one chooses to "View the full Getty record"[10], a user can find not just identity information, but also rich data, such as: the artist's associative relationships with related people or corporate bodies, sources of relevant information, aligned entries with other name authorities, and many publications and databases where the artist is the subject.

> With the rapid development of the digital humanities field, demands for smarter and bigger historical and cultural heritage data, which usually cannot be obtained through web crawling or scraping, have generated deep interest in LAM data, the treasure of all society. The semantic enrichment strategy represents one major step and directly enhances LAM data by using semantic technologies

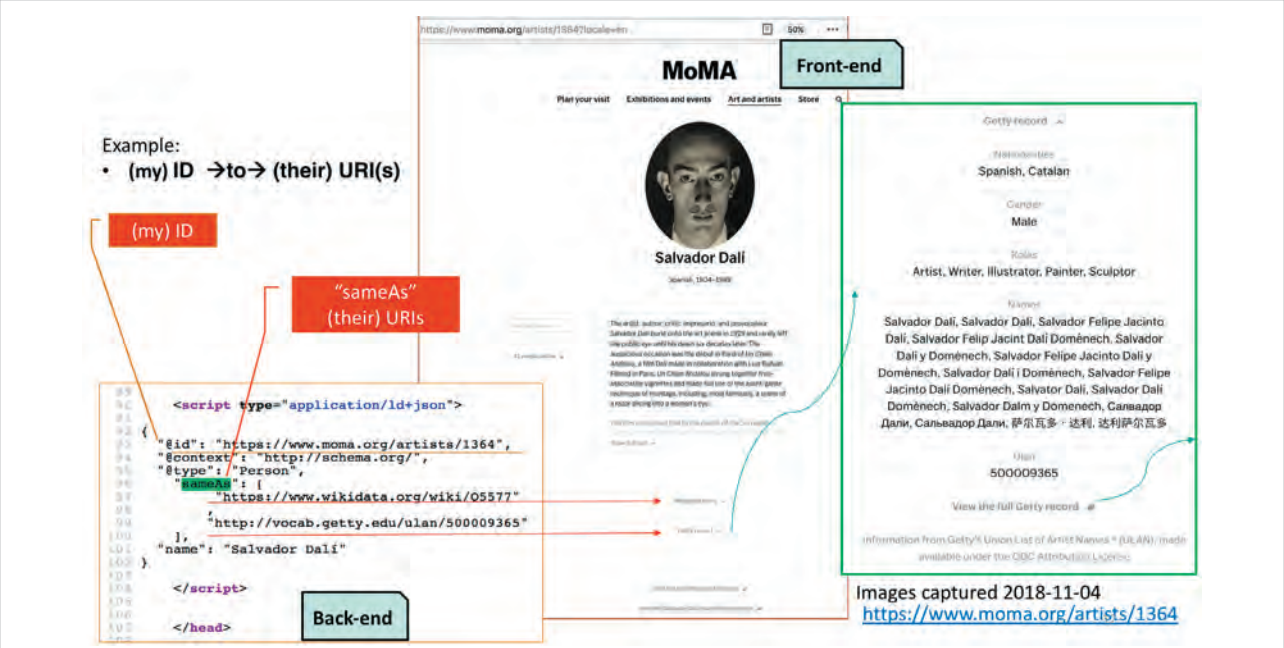


Figure 8. Front-end and back-end of the "massaged" results, using a *MoMA* artist webpage as an example.
Source: Generated by the author based on *MoMA* webpage and its source code view.
https://www.moma.org/artists/1364

The aligned contents are presented not just in the front-end webpages (explained above). At the back-end (Figure 8 left), the artist's URIs from *Wikidata* and *ULAN* are embedded in the <head> section of the HTML page, coded with schema.org property *sameAs*. The impact of such an enrichment also exposes resources about the focused entity of interest (e.g., the artist) to the web. For example, searching the artist's *ULAN* identifiers through search engines (e.g., *Google*) would bring highly relevant links on the first two results pages. In this way, the ranks of significant cultural heritage institutions such as museums and libraries are actually greatly increased, placing them ahead of hundreds of other results pushed out by search engines.

It has become a common method to use tools such as *OpenRefine* to consolidate unstructured data, where the (string) label of a concept used in an object's metadata (e.g., *MoMA*'s artist records available in a CSV file) is aligned with the identifier of this concept in its source KOS vocabulary (e.g., *ULAN*). In the *MoMA* case, the process is described in the guidelines provided by the *Getty Vocabularies LOD service*.[11]

Case: A checklist of ***Museums and Collections with Maya Inscriptions***

A similar agent-centered information cluster resource is the *Museums and Collections with Maya Inscriptions* website, developed by the *Interdisciplinary Dictionary of Classic Mayan* (*Textdatenbank und Wörterbuch des Klassischen Maya*) research center at the *University of Bonn*.[12] One of the corpus databases has been constructed for objects that are now housed in museums and collections. The website provides a resource listing all museums and collections with Mayan inscriptions worldwide. Each page provides the museum or collection's name, location, contact information, and links to the museum's website, catalogs and databases. When possible, identifiers from *GeoNames*, the *Getty Thesaurus of Geographic Names* (*TGN*), and the *Union List of Artist Names* (*ULAN*) are included. A map (*Google Maps*) showing the exact location of the relevant museum or collection is also given for

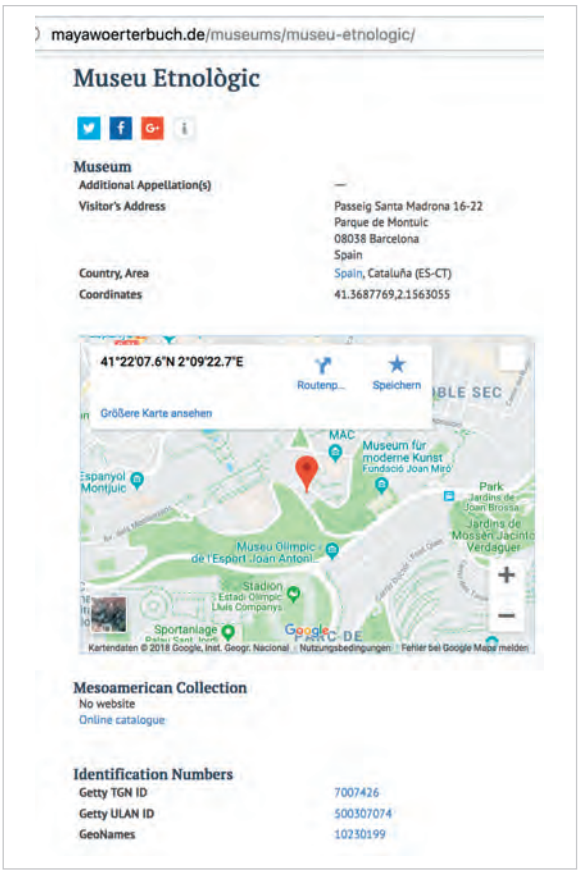


Figure 9. A webpage about a museum including the "massaged" results including the identifiers of *Getty TGN*, *ILAN*, and *GeoNames*.
Source: Entry of *Museu Etnològic*
http://mayawoerterbuch.de/museums/museu-etnologic

each record (**Wagner** *et al.*, 2014). For example, the page for *Museu Etnològic* in Barcelona, Spain[13] provides a direct link to *ULAN* ID 500307074[14] for the museum's name authority record, the *Getty TGN* ID 7007426[15] for Barcelona, and *GeoNames* 10230199[16] for *Museu Etnològic*, plus a geographic map showing the exact location of the relevant museum or collection. [Figure 9]
http://vocab.getty.edu/page/ulan/500307074
http://vocab.getty.edu/page/tgn/7007426
http://www.geonames.org/10230199/museu-etnologic.html

> Enriching structured data has become a common initiative in LAM data enhancement efforts. These structured data have been developed and sustainably maintained; they exist and are ready to be enriched

3.1.3. Example: structured data in subject authority entries

Case: *Faceted Application of Subject Terminology* (*FAST*)

Conversely, this kind of contextualization or "massage" can also be conducted to enrich an existing KOS vocabulary by aligning to external contextual resources. What is enriched is not a metadata record for an object, but rather the subject authority entry. The important concept to be revisited here is the typed relationships that are applied in the massaged process. *Faceted Application of Subject Terminology* (*FAST*) has used *schema:sameAs*, *owl:sameAs*, and *foaf:focus* to allow *FAST*'s controlled terms (representing instances of *skos:Concept*) to be connected to URIs that identify real-world entities specified at *VIAF*, *GeoNames*, and *DBpedia*.
https://www.oclc.org/research/themes/data-science/fast.html

In the following example, the highlighted codes indicate the relevant coding of properties that connects the artist entry in *Wikipedia* through *foaf:focus* property, while the next set of codes shows how the data from *VIAF* is connected through *schema:sameAs*.

- through *foaf:focus*, the *Wikipedia* URI allows a *FAST* concept to connect with information about the concept, which is usually excluded in authority records;
- through schema:sameAs, the identifier of *VIAF* lets *FAST* take advantage of all of the various string values included in *VIAF* (containing dozens of multilingual name authorities) without having to manually include the values in the RDF triples for the specific entry in *FAST*.

With these accurate typed relationships, machines can understand (and reason) that a *FAST*'s controlled term is related to a real-world entity, and allows humans to gather more information about the entity that is being described (**O'Neill**; **Mixter**, 2013). [Figure 10]

FAST Authority File example

Entry 39278 for artist Gaudí, Antoni, 1852-1926 is enriched through the alignment with external identifiers using typed-relationships, foaf:focus and owl:sameAs

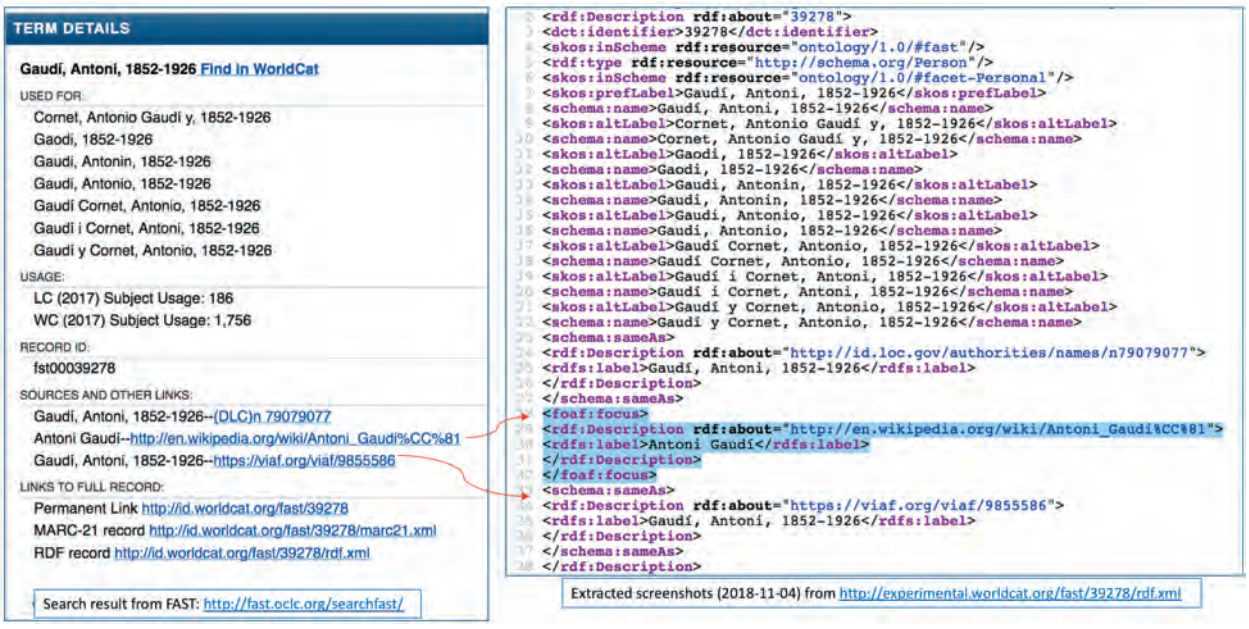


Figure 10. Examples from *Faceted Application of Subject Terminology* (*FAST*) showing the machine understandable coding of linkages to external KOS vocabularies and contextual resources.
Source: screenshots from *FAST*, 2018-11-01.

3.1.4. Discussion

Enriching structured data has become a common initiative in LAM data enhancement efforts, and more reports can be found in publications and on the web. The successful cases presented so far have proved that, when semantic enrichment is applied to the structured data that are in a controlled/normalized form, including entities for place, agent, concept, and time period, the results and impacts are significant. These structured data have been developed and sustainably maintained; they exist and are ready to be enriched. This is a major difference of such data from the others to be discussed in the following sections (on semi-structured and unstructured data), when more complicated semantic enrichment workflows (including model developing, batch processing, validating, disseminating, etc.) might need significant additional investments and resources.

The cases introduced in this section also revealed the usage of external contextual resources for enriching those controlled values in structured data. What are such resources, then? An *International Linked Data Survey for Implementers* conducted by the *OCLC Research* in 2018 reported top ranked Linked Data sources:
1) *id.loc.gov*
2) *VIAF*
3) *DBpedia*
4) *GeoNames*
5) *Wikidata*.

In comparison with the previous 2015 survey, the biggest change was the rise in *Wikidata* as a linked data source. *Wikidata* was elevated to the 5th-ranked data source utilized by linked data projects/services in the 2018 survey, compared to 15th in the 2015 survey. Eighty-one institutions responded to the 2018 survey, describing 104 linked data projects (**Smith-Yoshimura**, 2018).

In addition to these commonly used resources, other well-established LOD KOS vocabularies are similarly invaluable for specialized areas. The cases introduced in this section demonstrated the use of some of them. On a larger scale, a substantial activity that should be noticed is *Mix'n'Match* which has brought the largest mash-up effort forward in order to fully use established, reliable vocabularies and datasets.
https://tools.wmflabs.org/mix-n-match/#/

As a tool, *Mix'n'Match* lists entries of hundreds of external databases in a variety of categories and allows volunteers to manually match them against *Wikidata* items. An exceptional feature of this resource is the number and variety external datasets: for example, dozens in the Heritage category and over 500 in the Biography category are all sourced from different countries. The *Authority control for people* has reached over 480 catalogs as of the end of 2018, including, for example, *Comité Olímpico Argentino* (*Argentinian Olympians*), *RANM* (identifiers of members of the *Spanish Royal Academy of Medicine*), *Royal Swedish Academy of Letters*, Who's *Who in France*, *Database of Classical Scholars*, etc. Taking the general "Authority control" datasets (over 100) as another example, it includes well-known vocabularies such as *GeoNames*, *FAST*, *Unesco Thesaurus*, and *MeSH* (*Medical Subject Headings*) sub-lists, as well as other specialized vocabularies such as *DoS* (*Dictionary of Sydney*), *Inran Italian Food Nutrient profiles*, *ISO 15924* numeric code, *Gran Enciclopèdia Catalana*, *Europeana Fashion Thesaurus*, *MIMO Music Instruments*, *Great Russian Encyclopedia*, etc. More than half of these vocabularies have over 70% of entries manually matched by contributors. These resources can be investigated when a semantic enrichment project task force starts defining the target datasets, as a part in the process of enrichment (refer to the 2nd step in the next paragraph).

The report of the *Europeana Task Force on Enrichment and Evaluation* (**Isaac** *et al.*, 2015) contains the most comprehensive and relevant guidelines for the whole workflow and project design. It recommends ten steps for developing and maintaining a successful enrichment strategy:

1. Define the enrichment goals (annotation guidelines) that will guide your enrichment strategy.
2. Choose the right components for your enrichment workflow: enrichment solution and target datasets.
3. Define the enrichment workflow.
4. Make sure your enricher has sufficient knowledge.
5. Test your enrichment workflow.
6. Assess the quality of your enrichment and have an evaluation strategy.
7. Choose the right evaluation method.
8. Apply user-initiated enrichment workflows.
9. Document your enrichment process and learnings.
10. Monitor your enrichment process and re-assess.

3.2. Semantic enrichment of semi-structured data, expanding access points

In the previous section on semantic enrichment of structured data, methods discussed are normally applied to components in metadata records where data values are available in a controlled form and are expected to be the primary access points. What about those uncontrolled, not-normalized, and free-text parts within metadata records, or the

unstructured segments of otherwise structured datasets? For example, although agent names in a music bibliographic record are available access points, especially those fields in a controlled form, the responsibility or role of the agents might be hidden in non-controlled components within the same record. Uncovering potentially valuable, yet hidden information and access points in the semi-structured data leads to another major category of semantic enrichment actions.

This section is based on research experiments and pilot studies addressing the semantic enrichment of semi-structured data, parsing data that are in non-controlled/not-normalized form, turning them into access points, and providing useful contextual information. Semi-structured data waiting to be parsed and semantically enriched include, for example, the text in certain MARC bibliographic records' fields, the summary section in a photograph collection metadata description, and the detailed descriptive information in archival finding aids like EAD (*Encoded Archival Description*) records, just to name a few. They are different from unstructured data because these semi-structured data contents are the results of information processing or documentation workflow and are recorded in metadata records. They assemble invaluable resources not represented in other structured data fields, and usually contain important contextual information. Collectively, the examples included in this section illustrate methodologies and findings resulting from the extensive exploration and analysis conducted by the investigators.

3.2.1. Example: semi-structured data within MARC bibliographic records

In **Weitz** *et al.*'s article *Mining MARC's hidden treasures: Initial investigations into how notes of the past might shape our future* (2016), the researchers at *OCLC* (*Online Computer Library Center*) reported their study on finding, interpreting, and manipulating the rich trove of data already present in MARC bibliographic records. The following sub-sections are all based on this article.

As cataloging moves from AACR2 to Resource Description and Access (RDA), and MARC 21 gains the explicit level of detail that will allow cataloging to move into a post-MARC environment, the manipulation of existing data grows in importance. Finding, interpreting, and manipulating the data already present in MARC bibliographic records to create systematized forms is an invaluable step in moving MARC toward the Linked Data future. The semi-structured data's creation could depend on the original metadata structure and format, or the restrictions of the practices such as before implementation of RDA.

The *OCLC* researchers have been investigating the means by which to find names and their associated role phrases, in order to match those names to authorized forms, and to match role terms and phrases to controlled vocabularies. Approximately 19 million records for musical resources in *WorldCat* were analyzed in 2016. The records were generated during the 45-year history of *WorldCat*, and comprised both musical sound recordings and musical scores. The analysis of these 19 million records determined that they contain approximately 2.5 million names that can be identified as distinct. Roughly 50% of those 2.5 million names have a role designation in the form of a subfield $e (Relator Term), subfield $4 (Relator Code), or both, as well as additional descriptive data that could be mined to further refine the data coded in name fields.

The process can be summarized as:

- from uncontrolled occurrences in notes and/or statements of responsibility in records [e.g., 508 (Creation/Production Credits Note) or 511 (Participant or Performer Note)];
- find named entities and their associated role phrases;
- match them to the corresponding 7XX fields in the same record;
- match names to authorized forms;
- match role terms and phrases to controlled vocabularies.

An encouraging observation is that RDA elements can be incorporated in hybrid records without complete re-cataloging. In this case, it is accomplished by the addition of role designations in access points for music-related metadata.

In matching the identified role phrases to appropriate controlled vocabularies, multiple KOS vocabularies have been used, including the *Library of Congress Subject Headings* (*LCSH*), the *Library of Congress Demographic Group Terms*, and the *Dictionary of Occupational Titles* (*DOT*). Extended work has been conducted on multiple languages for the performer roles, medium of performance terms, associating the name of an instrument with its performer, and more. The experiment also led to the activities of compiling new value vocabularies and mapping to other sources, encompassing the entries in *OCLC*'s *Faceted Application of Subject Terminology* (*FAST*), *LCSH*, *Wikipedia*/*Wikidata*, *Library of Congress Medium of Performance Thesaurus for Music* (*LCMPT*), *Dewey Decimal Classification* (*DDC*), and other multilingual controlled vocabularies.

OCLC's experiments with parsing that data in order to both associate performer names with authority data and identify role terms and phrases with controlled vocabularies have proven fruitful. The potential opportunities for Linked Data across bibliographic and authority data, across vocabularies, and across languages are vast (**Weitz** *et al.*, 2016).

3.2.2. Example: bibliographic metadata in MARC and beyond

Going beyond MARC, the enhancement of library catalogs towards Linked Data will empower library users to discover many more information resources, providing them not just with access to basic descriptive information about works,

but also the context surrounding works' creation, distribution, and use. As pointed out in the *W3C Linked Library Data Incubator Group* (*LLD-XG*)'s final report (*W3C*, 2011), in our current document-based ecosystem, data is always exchanged in the form of entire records, each of which is presumed to be a complete description. In order to integrate library metadata with the Semantic Web, **Dunsire** and **Willer** (2011) laid out detailed examples of how traditional library bibliographic records may be disaggregated into catalog records consisting of RDF triples, as well as what benefits LAMs may receive from including such data. **Alemu** *et al.* (2012) called for a conceptual shift from document-centric to data-centric metadata, moving from MARC-based to RDF-based description, and presented methods to achieve this goal without disrupting current library metadata operations.

In researching metadata's linkability, a team at *Kent State University* in the United States showed how to align metadata structures and properties from diverse communities, specifically how to relate metadata vocabularies familiar in the library community to the unfamiliar resources of the LOD universe (**Gracy**; **Zeng**; **Skirvin,** 2013; **Zeng**; **Gracy**; **Skirvin**, 2013). In this case study, the research team collected, analyzed, and mapped properties used in describing and accessing music recordings, scores, and music-related information across a variety of music genres. The team evaluated 11 music-related Linked Data datasets, 20 collections of digitized music recordings and scores, and 64 representative MARC records for sound recordings and musical scores (and their extended schema.org records). A number of crosswalks were created to align all the data structures, not only by classes and properties, but also indicating the levels of mapping (such as exact match, broad match, narrow match, close match, and no match). The researchers then randomly selected 280 MARC records from each genre to verify the crosswalking results. The process resulted in a unified crosswalk that aligns these properties according to nine common groups of bibliographic data. These include: title information, responsible bodies, physical characteristics, location, subject, description of content, intellectual property, usage, and relation. Many of these properties are usually hidden in the semi-structured data portion of records.

In the areas of performance and recording, the researchers believe that three MARC note (5xx) fields in particular contain valuable data points that could be useful as links to other data sources external to the library catalog, although they are not usable as Linked Data in their current form. Depending on the application and practices of cataloging for sound recordings and musical scores, the linkability could vary, as illustrated by the Figures 11a and 11b.

Other collections with non-MARC metadata reviewed for this study included such information as instrumentation; performance medium; meter; tempo; duration; notes about strains, rendition, phrase structure, assorted musicological details; region and language of music; digital reproduction history; file size; description of the physical container; identifiers such as plate, publisher, label, matrix, and take numbers; and rights information. It is promising that by making bibliographic data shareable, extensible, and reusable, libraries are able to aggregate data based on the pieces/chunks of information they need from a dataset without integrating a whole database or converting full metadata records. They can mash up metadata statements (not whole records) from different namespaces or aggregate data from a variety of resources, based solely on what is needed. (**Gracy**; **Zeng**; **Skirvin**, 2013; **Zeng**; **Gracy**; **Skirvin**, 2013).

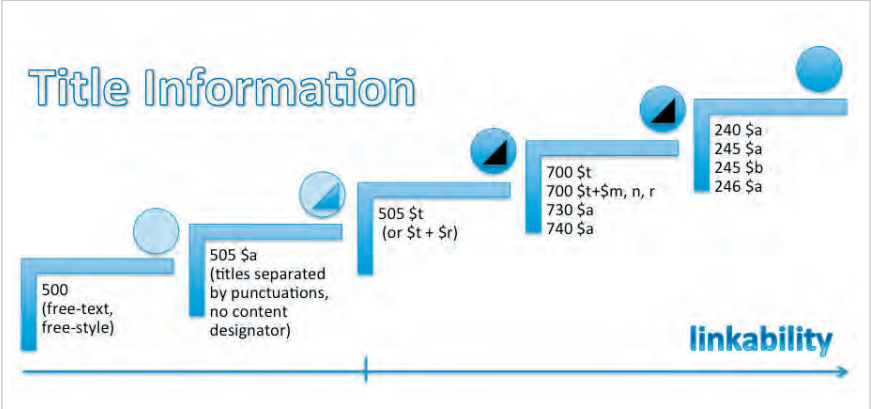


Figure 11a. Illustration of linkability of title information based on the study samples.
Source: **Zeng**; **Gracy**; **Skirvin**, 2013.

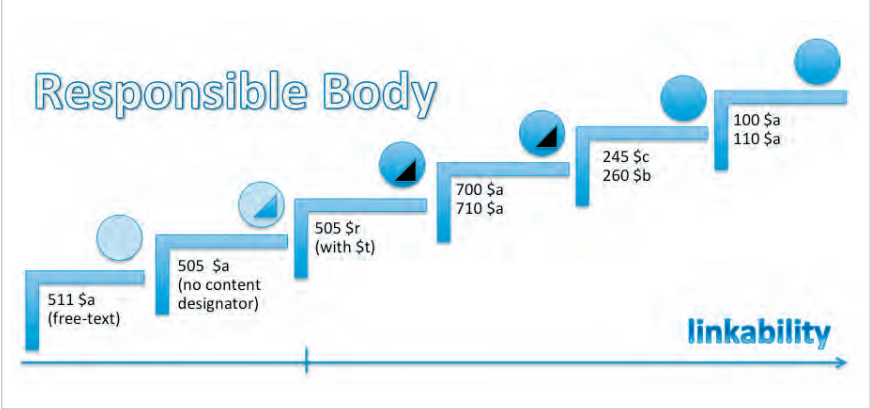


Figure 11b. Illustration of linkability of Responsible Body information based on the study samples.
Source: **Zeng**; **Gracy**; **Skirvin**, 2013.

It should be noted that the cataloging practices and standards have changed during the past five years, hence some of the situations illustrated above might have improved already and the linkability measurement could be updated. Collectively, LAM data as Linked Data is perceived as mainstream since the *W3C*

LLD-XG 2011 reports' release. Today, the number of registered LOD datasets in the old *DataHub* is over 11,000, including 1,100+ named with KOS types (thesaurus, taxonomy, classification, terminology, and ontology) and 1,500+ that corresponded to search by LAMs and special collections, as of October 2018.[14] More research on their role in semantic enrichment of semi-structured will be interesting.

3.2.3. Example: semi-structured data in archive finding aid descriptions

In current archival information systems, useful linkable information is embedded in narrative descriptions known as finding aids. Finding aids are tools that help a user find information in a specific record group, collection, or series of archival materials. Examples of finding aids include published and unpublished inventories, container and folder lists, card catalogs, calendars, indexes, registers, and institutional guides (*National Archives*, 2016). Finding aids can be found in many different formats, ranging from handwritten documents, *Word*, and spreadsheets to HTML and XML. MARC 21 has also been used to encode archival metadata. Increasingly, creating XML-tagged finding aids enabled by *Encoded Archival Description* (EAD) schemas has become a popular practice in the archival community. EAD is a pioneer of metadata schemas developed for the LAM community, with its beta version issued in 1996. EAD has always been among the first to implement the newest markup language standards in the schemas, including from SGML DTD to XML DTD to XML Schema (in Relax NG and in XSD). The semantics/structure and syntax in EAD schemas are integrated as one unit. In other words, EAD as a metadata standard keeps both the specification of the element set and encoding schema in one document.

Finding aids provide contextual information to understand the nature and scope of archival collections. The format of the genre is mainly semi-structured, characterized by large text blocks in which many types of information are intermingled and unlabeled. In the blocks of histories about records creators and the scope notes for collections, many names of persons, organizations, places, and events, as well as topical terms might be mentioned. Apart from full-text searching, the lack of semantic distinction among the different entities and topics hinders efficient and effective information retrieval, which also restricts the ability of information systems to create the links that would gather widely dispersed information about the same person, organization, or thing into one place.

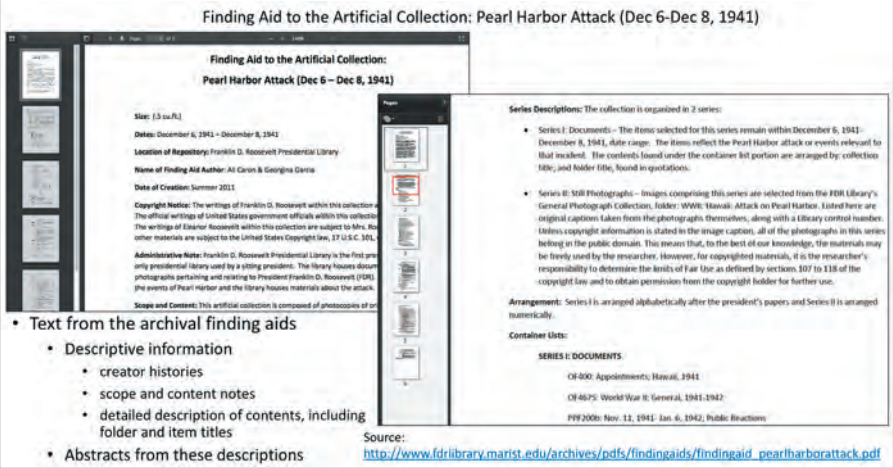


Figure 12a. Portions of a finding aid and explanation of the text used in the semantic analysis process.

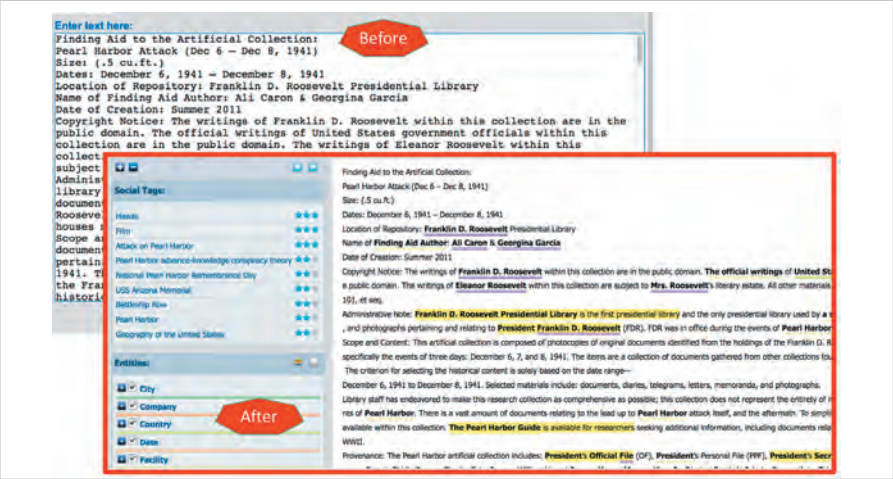


Figure 12b. Example from the semantic analysis results using *OpenCalais* demo tool, indicating the entities and social tags generated.

Inspired by the possible semantic enrichment technology for entity analysis and extracting, as well as the potential of using Linked Data principles to enhance the discoverability and linkability of the rich information in finding aids, Karen Gracy led a research team in conducting an experiment using a sample of 43 archival record groups. The finding aids are from 16 institutions, including university archives, government records archives, and manuscript/special collections repositories in various LAMs. Using the semantic analysis tool *OpenCalais*' free version, descriptive information contained in the archival finding aids (such as creator histories, and scope and content notes sections), as well as abstracts from these descriptions, were used to generate extracted access point candidates (**Gracy**; **Davidson** 2014). [Figure 12a and 12b] *http://www.opencalais.com/opencalais-demo*

The analysis resulted in dozens and, at times, hundreds of potential entities and social tags that could be used to provide additional points of entry to

Findings	Results of semantic analysis by OpenCalais (example)
Entities correctly identified via *Calais* analysis included: - personal names (Person), - corporate names (Company, Facility, Organization), - geographic names (City, Continent, Country, Natural Feature, Province Or State, Region), and - events (Holiday, Political Event). The scores for each identified entity may be used as a valuable clue about the importance of that entity to the overall scope of the archival collection. The "Social Tags," "Industry Term," and "Products" categorizations were the least reliable in terms of accuracy. Analysis of detailed description areas was most likely to lead to incorrect identification of text strings because the descriptions have the physical location information intermingled. Targeted analysis of particular areas of the finding aids may result in better accuracy for topical analysis.	 Figure 13. Findings of the results of automatic semantic analysis using *OpenCalais* demo tool.

these archival records (refer to the Findings in the Figure 13). The dataset of the structured data in RDF/XML format can also be obtained directly from the same output. [Figure 13]

The research team developed a program to allow batch processing. The software automatically obtained the archival records and sent them to the semantic analysis service supported by *OpenCalais*. The output, which was in the JSON format, was then converted directly into a CSV file. The resulting database contained the following fields: Entity-type, Entity-name, Relevance-ratio, and File-source. Using the *OpenRefine* tool, the data were clustered automatically to allow the researchers to clean up the data manually (e.g., merge the synonyms and delete incorrect extractions) and validate the names and topical terms against various controlled vocabularies, such as the *Library of Congress Name Authority File*, *LCSH*, and the *Getty* vocabularies. Figure 14 illustrates this multi-step process. [Figure 14]
http://openrefine.org

The researchers also experimented with other tools to identify named entities and topical terms from finding aids, including *Cogito Intelligence API*, *MachineLinking*, and *Zemanta* (**Gracy**; **Zeng**, 2015).
http://www.intelligenceapi.com
http://www.machinelinking.com/wp
http://www.zemanta.com/api

Figure 15a is an example from the semantic analysis results using *Cogito Intelligence API* demo tool for the selected archival finding aids, showing the initial result preview. The top row circle indicates the options of display, including tagging, categorization, text mining, semantic reasoning, and fact mining, plus by entities such as People, Organizations, and Places (see Figure 15b). The tool also provides Wri-

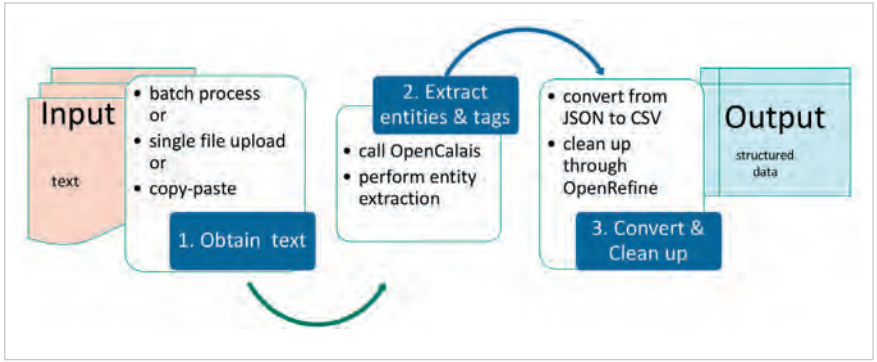


Figure 14. Illustration of the process of entity analysis and extracting

teprint analysis which estimates the readability level of a provided document, collecting and forging a full set of readability indexes as well as grammatical, lexical and semantic analysis scores.

Both *OpenCalais* and *Cogito Intelligence API* are powered by multiple taxonomies and domain ontologies, and feature automatic processes of text mining, categorization, semantic tagging, fact mining, extraction of entities and relationships, and visualization of the entity relationships and geographic locations. In addition to the functions similar to *OpenCalais*, *Cogito Intelligence API* also conducts semantic reasoning. Although neither of these tools have been developed for cultural heritage or humanities domains, meaning that their taxonomies and ontologies are not in these domains, the research results strongly suggest that it would be well worth the effort for institutions to experiment with semantic analysis methods as an initial step to suggest key entities and topics, or as a final check to ensure that important concepts or entities have not been overlooked. For certain types of records, particularly those for which subject indexing is not common, semantic analysis may provide entry points to archival records that were not previously available. Such techniques will enhance subject analysis at the levels of description and identification, but are unlikely to be useful for interpretation of the material. These findings were confirmed through a second case study the research team conducted based on 44 philosophy theses from *KentLINK* and *OhioLINK* (**Zeng**; **Gracy**; **Žumer**, 2014).

3.2.4. Example: semi-structured data in item-centered information cluster webpages

Comparable to the processes for enhancing the semi-structured data in archival finding aids records, many information cluster webpages also show the infinite potential of semantic enrichment for expanded access points through the semi-structured portions. The uniqueness of the examples to be discussed next demonstrates that, ultimately, the enrichment process can be implemented case-by-case, collectively or independently, with or without significant project funding.

By testing various types of descriptive information clusters carried by webpages, it is clear that the semi-structured data that have been created for information access and exposure can be semantically enriched effectively. Examples of item-centered information cluster webpages that were tested by the author include metadata descriptions about photograph special collections, table-of-contents included in metadata records, back-of-book indexes for transcribed oral history materials, and captions and other descriptive information for cultural objects presented on their digital representation webpages. They demonstrate a limited number of types, while denoting that similar results can be obtained by many other types of semi-structured data resources.

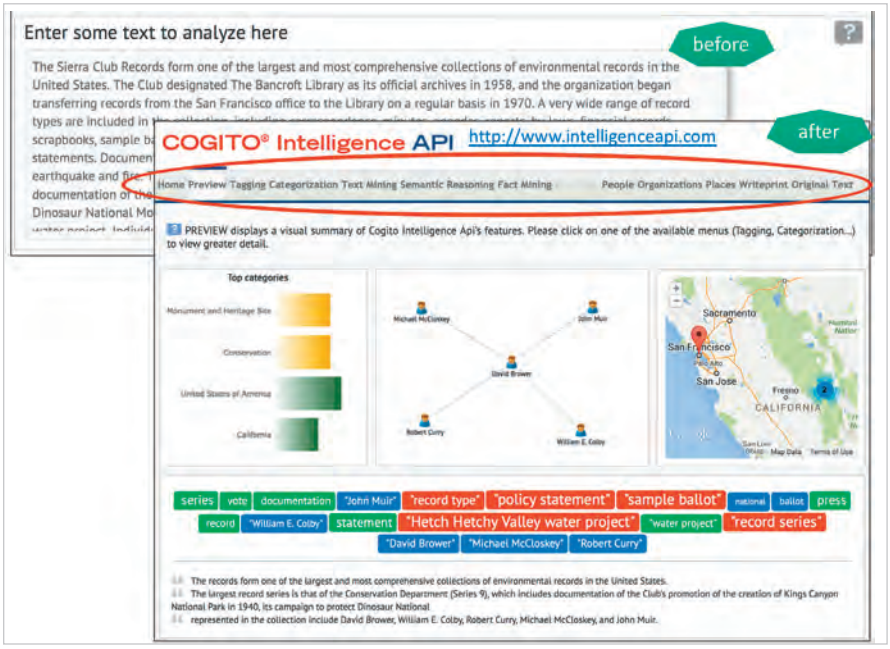


Figure 15a. Example from the semantic analysis results using *Cogito Intelligence API* demo tool.

Figure 15b. Example from the semantic analysis results using *Cogito Intelligence API* demo tool, indicating the text mining and named entity (places) generated.

The results of newly generated structured data can be embedded in individual webpages or integrated into databases.

Metadata and curator's notes for a special collection

Named entities can be extracted from free-text segments in metadata descriptions about a special collection. In this example, entity extraction generated expanded access points based on place names of a special collection of thousands of historical photographs. [Figure 16a]

Table of contents included in metadata records or linked by metadata records

Table of contents (TOC) included in or linked by a metadata record of an edited book could be sources for entity extraction. In this example, shown in the Figure 16b, the TOC as entered in the original record (left in Figure) or generated by machine (center in Figure) provides expanded access to the authors and themes of this item (right in Figure). [Figure 16b]

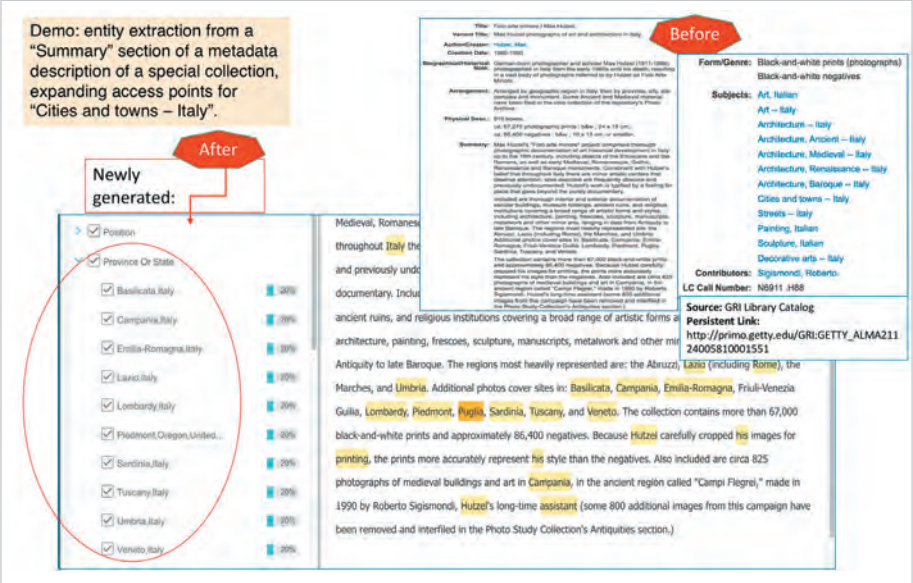


Figure 16a. Example of entity extraction (place) from a Summary section of a metadata description.

Back-of-book indexes for transcribed oral history materials

Similar to the TOC, back-of-the-book style indexes for oral history materials are precious sources for generating new access points.

Due to widespread digitization efforts over the last 20 years, many of the oral history materials hosted by LAMs have transcripts available in PDF or other digital formats, according to the copyright and privacy conditions. The oral history transcripts files might be managed at the collection level only, or they may be indexed using back-of-the-book style and kept together with PDF files that are downloadable. The indexes are usually established with high quality and involve collaborations between different units and institutions. At the highest level these indexes are processed for text-based searching. There is great potential for applying semantic enrichment method to these information products.

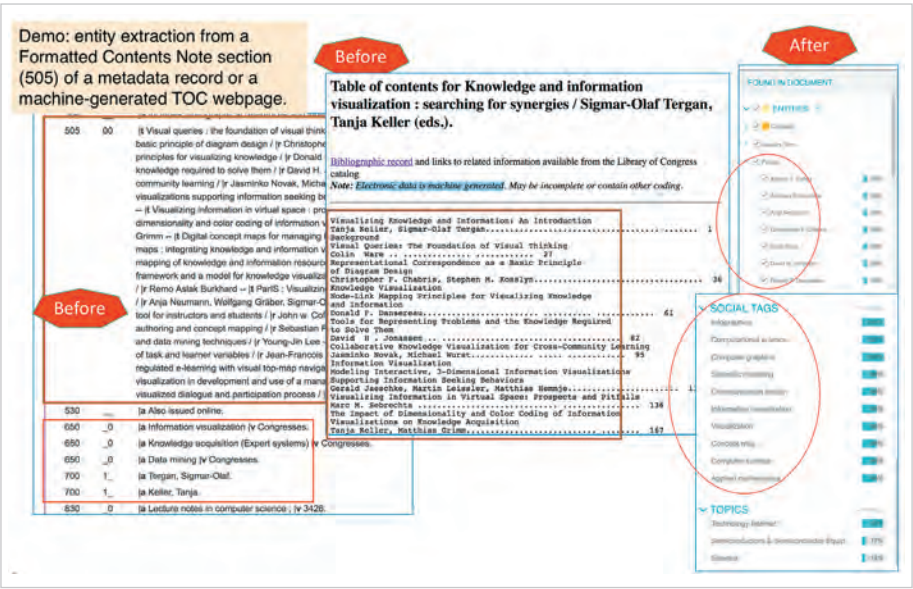


Figure 16b. Example of entity extraction from TOC provided by the metadata record of a publication.

Captions and other descriptive information for cultural objects presented on their digital representation webpages

Captions and other informative descriptions for cultural objects featured on the individual webpages of cultural institutions could also benefit from entity extraction. In this example, the text (see <Before> in the Figure 17) about four ancients and their favorites depicted on a vase made during the Yuan Dynasty (1271-1368) are recognized by machines (see <After> in Figure 17) through entity extraction (using *Boson*, *OpenCalais*, and *Cogito*). [Figure 17] *https://bosonnlp.com/demo*

Although this test was performed using free tools and resulted in different degrees of completeness, the potential of generating information access and discovery based on the entities and keywords is promising. Such benefits might be

especially meaningful to institutions with relatively small digitally represented collections, including those in non-English-speaking locations or less digital-driven places.

3.2.5. Discussion

An important feature of semi-structured data resources that should be recognized actually resides in their nature of being the products of information processing. These semi-structured data represent the accumulated time, knowledge, and experience of the creators who generated these metadata through formal workflow which conforms to professional standards and best practices. With semantic enrichment processes, the data values in semi-structured data are contextualized through the metadata elements/fields; hence, the function and meaning are clearly implied. By parsing these data through advanced information technologies, these LAM data are dramatically enriched and are converted into new access points.

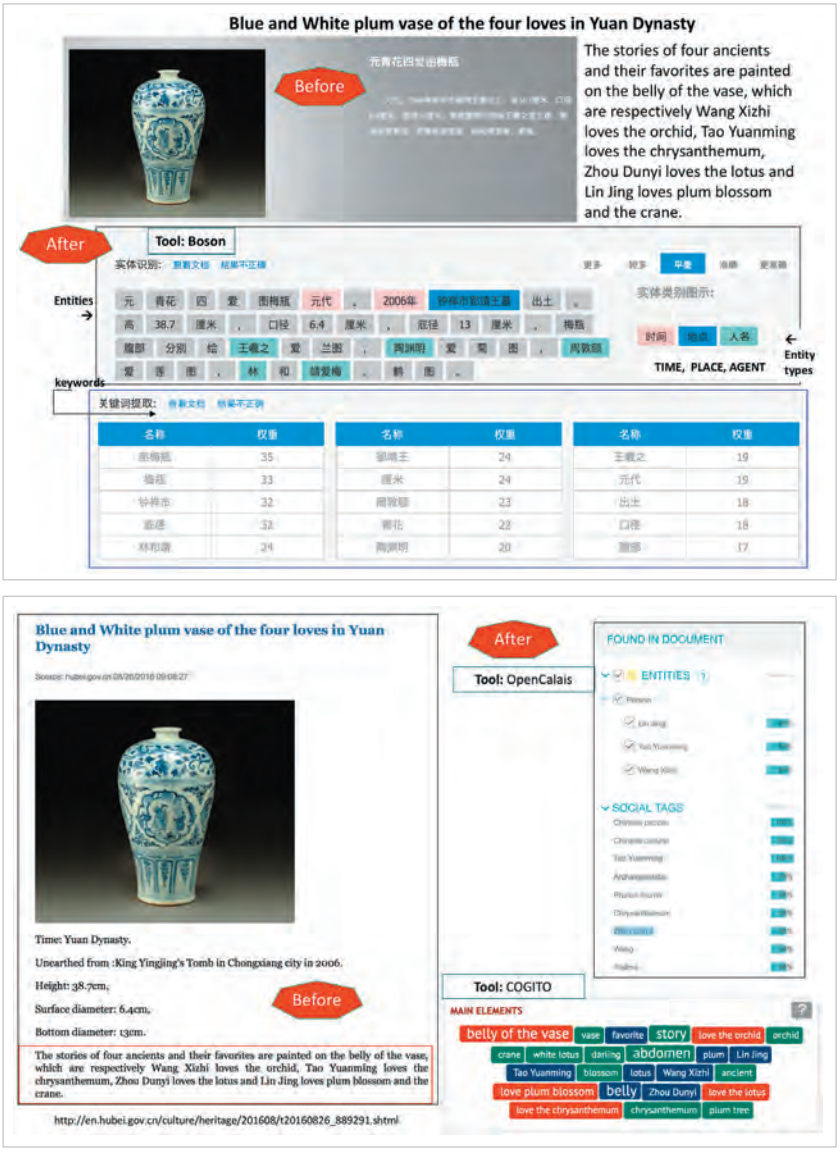


Figure 17. Example of entity extraction from captions and other descriptive information for a cultural object using three different semantic analysis tools.

The *OCLC*'s experiments on bibliographic records revealed potential opportunities for Linked Data across bibliographic and authority data, across vocabularies, and across languages. Other pilot studies shared in this section also prove that, by making bibliographic data shareable, extensible, and reusable, LAMs will be able to aggregate data based on the pieces/chunks of information they need from a dataset without integrating a whole database or converting full metadata records.

The examples presented in this semi-structured data enrichment section reveal that, ultimately, additional useful data can be derived from large digital collections as well as from individual item-centered information clusters. These activities can be managed case-by-case, from the top-down or the bottom-up, collectively or independently, with or without significant project funding.

Semantic analytics, one of the advanced semantic enrichment methods, has been used for analyzing, searching, and presenting information by using explicit semantic relationships between known entities. It is a major method used in the semi-structured data processing presented above. The tools used in the experiments discussed in this section, such as *OpenCalais* and *Cogito Intelligence API*, are powered by multiple taxonomies and domain ontologies, and benefit from machine learning and other new artificial intelligence (AI) technologies, far beyond normal natural language processing. The APIs classify entities using different taxonomies and disambiguate them with different knowledge bases. Processes include recognizing named entities mentioned in text, assigning them as pre-defined types, and linking them with their matching entities in a knowledge base. "Entity" has been a hot keyword ubiquitous in semantic technology related conferences, such as the most recent 2018 *International Semantic Web Conference*, where "entity"-related research tracks range from entity extraction, annotation, recognition, disambiguation, to relation linking and embedding, while the entity of interest could be varying.
https://link.springer.com/book/10.1007%2F978-3-030-00671-6

Countless semantic analysis and machine learning experiments and tools have been reported and can be found in literature. Taking entity disambiguation as an example, numerous algorithms can be applied to measure text string si-

milarity, semantic or structural similarity, contextual similarity, and commonness. Similar to what *Cogito* has also included as one of its semantic analysis results' delivery, Knowledge Graphs (KG) are widely used abstractions in representing entity-centric knowledge. On top of rule-based systems, embedding-based systems for Knowledge Graph completion has become a dominating focus in research and development during recent years, also revealed by those reported at the *2018 International Semantic Web Conference*. The role of LAM data created for cultural objects will be both the consumer and the contributor of semantic technologies and the web of data.

> "An important feature of semi-structured data resources actually resides in their nature of being the products of information processing. These semi-structured data represent the accumulated time, knowledge, and experience of the creators who generated these metadata through formal workflow which conforms to professional standards and best practices. Their semantic enrichment activities can be managed case-by-case, from the top-down or the bottom-up, collectively or independently, with or without significant project funding"

3.3. New structured data generated from unstructured data, supporting knowledge discovery

The semantic enrichment approaches and issues discussed in this section will be about information discovery and re-discovery from unstructured data. In data resources that are served through LAMs, unstructured data are usually available in the largest quantity in comparison with the structured data; have the most diversity in type, nature, and quality; and are the most challenging to process. In LAMs, they can be found in documents and other information-bearing objects (textual or non-textual, digitized or non-digitized), in all kinds of formats. This section will focus on certain types of LAM data hosted at retrospective resource warehouses. The mixed or heterogeneous contents from distributed data providers will be discussed in the last section of this chapter (Section 4).

In terms of text-based information processing and retrieval, for a long time, text-matching (especially word matching) had been the primary way to do full-text searching. Only recently, content-based searching (according to semantic meanings) has been implemented and used on a large scale, as symbolized by Knowledge Graphs. To advance discovery from text-based resources, instead of relying on full-text searching, methods such as semantic-based analysis, extraction, mining, and tagging are used to improve the information discovery from unstructured data. [Refer to the examples introduced in the previous section (Section 2) on content-based new approaches and tools].

For non-text-based resources, semantic annotation and ontology-based knowledge bases have revealed great potential for digging into data and supporting digital humanities research. In general, "annotating" is the act of expressing knowledge about a resource, attaching names, attributes, comments, descriptions, etc., to a document or to a selected part in a resource. Annotating provides additional information (metadata) about an existing resource. "Semantic annotation" goes one level deeper: it enriches the unstructured data with a context that is further linked to the structured knowledge of a domain.

Tools such as *Oasis* (*Open Annotation Semantic Imaging System*), *Pundit*, *Mirador*, *Brat Rapid Annotation Tool*, *Recogito*, and *MapHub* enable users not only to comment, bookmark, and tag, but also to create semantically structured data while annotating. With effective application of *W3C* standards, these annotation results are semantically marked up and expressed with typed relationships, teaching a computer how data items are related and how these relations can be evaluated automatically. In computerized systems, further semantic enrichment is based on the annotations. From there, the decision can be made according to logic, to come up with knowledge insertion. It also helps to link other same entities or related entities, which would be very useful for semantic analysis, as illustrated by **Daniel Mayer**'s video (2011), *Mainstream Semantic Enrichment*.
https://www.synaptica.com/image-annotation-indexing
http://thepund.it
http://projectmirador.org
http://brat.nlplab.org
https://recogito.pelagios.org
http://maphub.github.io

The following subsections introduce cases representing semantic enrichment on various types of LAM data, including oral history transcripts, OCR-ed materials, digital counterparts of cultural objects, maps, and images. These cases and research experiments are among the pioneering works related to Linked Open Data and digital humanities. Inspiring news of similar projects are reported more and more worldwide.

3.3.1. Example: oral history transcripts

The last five years have witnessed the widespread use of "Linked xyz" titles in cultural heritage discovery in digital environments, after the *W3C Library Linked Data Incubator Group Final Report* was released (*W3C*, 2011). Among these titles are highly recognized examples acknowledged by the LODLAM community such as: projects (*Linked Jazz*, *Linked Heritage*, *Linked Taiwan Artists*, *Linked Maps*); models (e.g., the *Linked Art Data Model* from the *Linked Art Community*);

and ontologies (e.g., a chain of *Linked Building Data* ontologies by the *Linked Building Data Community Group*).
https://linkedjazz.org
http://www.linkedheritage.eu
http://linkedart.ascdc.tw
http://usc-isi-i2.github.io/linked-maps
https://linked.art/index.html
https://www.w3.org/community/lbd

The *Pratt Institute*'s *Linked Jazz* project can be regarded as a forerunner in the field, presented as a finalist in the *2013 LODLAM Challenge* competition.
http://summit2013.lodlam.net

The project draws on jazz history materials in digital format to uncover meaningful connections between documents and data related to the personal and professional lives of jazz artists, and expose relationships between musicians that reveal their community network (**Pattuelli**, 2012). The 50+ interview transcripts were from various resources (the *Rutgers Institute for Jazz Studies Archives*, *Smithsonian Jazz Oral Histories*, the *Hamilton College Jazz Archive*, *UCLA*'s *Central Avenue Sounds series*, and the *University of Michigan*'s *Nathaniel C. Standifer Video Archive of Oral History*).[15] The documents were in PDF and text format, ranging from 12 to 187 pages in length.

The *Linked Jazz* project provides clear roadmaps for others to follow:

- digging into unstructured data, applying semantic analysis and annotation;
- mashing-up using *DBpedia*;
- establishing name authorities based on *VIAF* and *DBpedia*;
- developing an ontology for relationships (e.g., *knows*, *mentorOf*, *isInfluencedBy*, *collaboratedWith*) after consulting *Friend of A Friend (FOAF) ontology* and *Music Ontology*;
- crowdsourcing on assigning more granular terms to describe the relationship between an interviewee and the person mentioned (for the *Linked Jazz 52nd Street* sub-project); and
- visualizing the networks with images, videos, and short biographies of jazz musicians within the networks. [Figure 18]

The *Linked Jazz* project sits at the intersection of three important domains:

(1) LOD concepts and semantic technologies to be used in LAMs;

(2) the roles and contributions of knowledge organization methods in LOD-enabled products; and

(3) the applicable areas beyond bibliographic data and conventional resources' management, i.e., the unstructured data that LAMs have managed for years but could be better used in discovery.

It utilized solid research methodologies, implemented international standards, and applied innovative technologies to the digitized oral history transcripts from jazz archives for the discovery, visualization, and use of primary sources.

The efforts of digging into these oral history transcripts have revealed new discoveries to the world that are highly regarded by the jazz education community. Extended research projects are adding facets representing various professional

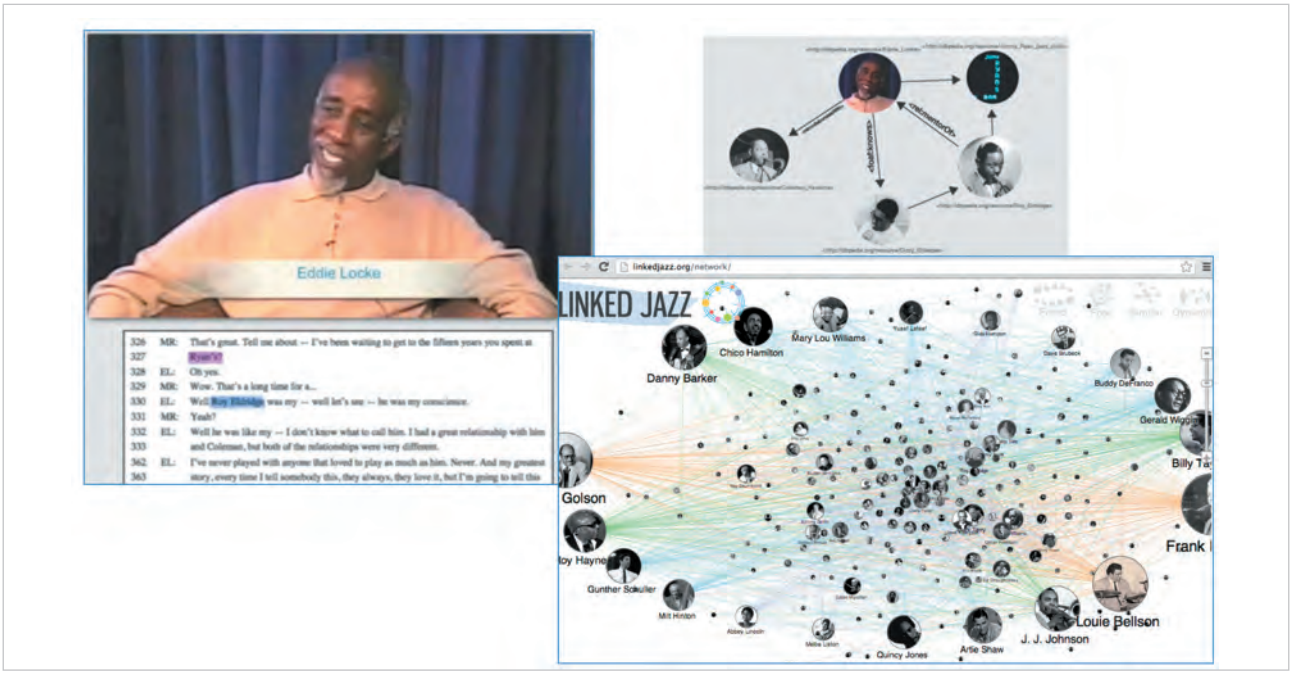


Figure 18. Demo of *Linked Jazz* project outcomes. *http://linkedjazz.org*

and social aspects of the jazz community, and expanding data sources to include other document types and music-related datasets (**Thorsen**; **Pattuelli**, 2016; **Pattuelli**; **Hwang**; **Miller**, 2016).

3.3.2. Example: OCR-ed documents

Analogous to the transcribed documents are a large number of historical documents that have been obtained through Optical Character Recognition (OCR), which involves translating the documents into machine processable text.

Before any data mining, entity extracting, and analysis can be done for OCR-ed documents, one has to face the challenges for character recognition, due to various reasons such as font disparities across different materials, lack of orthographic standards where the same word might be spelled differently, material quality, and the unavailability of lexicons of known historical spelling variants (**Mutuvi** *et al*., 2018; **Lin** *et al.*, 2012). Yet, such aspects and their impacts on semantic analysis have been less explored so far in comparison with other topics on digital libraries (**Hinze** *et al.*, 2015; **Bainbridge** *et al.*, 2016). The two research projects introduced next represent the efforts to fill such a gap.

Approaches for OCR post-processing can be seen through three main categories: manual error correction, dictionary-based error correction, and context-based error correction (**Nguyen** *et al.,* 2018). Obtaining thematic patterns from large unstructured collections of text by grouping documents into coherent topics is another major approach, among which common topic modeling techniques are the *Latent Dirichlet Allocation* (LDA) and the *Non-negative Matrix factorization* (NMF). **Mutuvi** *et al.*'s research examines the effect of noise on unsupervised topic modeling algorithms, through comparison of performance of both the LDA and NMF topic models in the presence of OCR errors. The research is supported by the *NewsEye* project, which is funded by the European Union's *Horizon 2020* research and innovation program. *NewsEye* is aimed at tens of millions of newspaper pages from European libraries that have been digitized and made available online during the last decade.
https://www.newseye.eu

In its poster reporting the inner testing of newspaper datasets from three partner libraries, the processes are laid out as:

- text recognition & article separation;
- semantic text enrichment;
- dynamic text analysis;
- personal research assistant; and
- user interface,

where the "semantic text enrichment" processes include: named entity recognition, stance detection, novelty detection, and event detection.
https://www.newseye.eu/fileadmin/user_upload/Poster_white.png

The research team of the *Capisco* project at *University of Waikato*, New Zealand, conducted research to address OCR-ed documents (**Bainbridge** *et al.*, 2016). The rationale is that lexicographic search in large collections, such as the *HathiTrust*'s 13 million volumes with 4.6 billion pages, often returns large sets of unrelated documents (due to homographs —same spelling, different meaning-being included), while relevant sources may remain undetected unless the right keyword is found. The problem is exacerbated in documents that have been obtained through OCR, as recognition errors may lead to misidentification of terms, which are then either mistakenly included or omitted from the search results. Meanwhile, the research team is motivated to avoid major changes to the document retrieval mechanism and indexing strategy, query processing, and interface to which users and technical teams are accustomed.

The *Capisco* system developed by the research team focuses on semantic indexing and search. It uses a knowledge base containing information about concepts in context, initially created by mining *Wikipedia* and potentially further enriched by domain experts. Each concept is identified by an ID, and also carries a human-readable concept label. There are several key steps:

- Concept labels are derived from *Wikipedia* article titles.
- Synonymous terms for a concept are stored with reference to the context in which they appear.
- The context of a term refers to the main area in which this term is used for this concept.
- Because contexts are also concepts, the knowledge base forms an interlinked Concepts in Context (CiC) network.

The knowledge base is used to disambiguate a term (i.e., identifying its semantic concept) and identify potentially matching concepts, which would lead to the identification of significant topics within a document. For each context, an index entry is created with references to the pages on which the term appears.

The *Capisco* project team reported an analytical approach to explore five strategies for low-cost semantic enhancement to large digital collection's metadata and indexing, especially the OCRed historical materials, showing through examples how using semantic concepts can help identify OCR errors. In testing the results, five approaches were used:

- Approach 1. Concept labels added to metadata
- Approach 2. Concepts and synonyms added to metadata
- Approach 3. Concept labels indexed at page-level

- Approach 4. Concept label and synonyms indexed
- Approach 5. Concepts and synonyms added at page metadata (only possible for digital library implementations that support page-level metadata fields)

Three of these approaches address adding information about semantic concepts to the metadata (1, 2, 5) and two others concern adding information about semantic concepts to the full-text index (3 and 4). A case study of four documents showed the differences in result sets for lexical search, semantic search, and each of the five approaches. The study also established four collections (with nearly five thousand pages and thousands of tokens and concepts indexed) in which the performance implications of enriching the full-text index with concept labels was experimentally determined. Based on semantic enrichment, search is now possible via both the simple interface and the advanced interface for filtering and search in metadata-based enhancements. The researchers concluded that even though the work test focused on simply enhancing the lexical search capabilities of traditional digital libraries, the most powerful solution would be a combination between lexical-based search and semantic search as offered in *Capisco* (**Bainbridge** *et al.*, 2016).

3.3.3. Example: historical maps

Maphub, another pioneer project, is an online application for exploring and annotating digitized, high-resolution historic maps, developed at *Cornell University*'s *Department of Information Science*. All user-contributed annotations are shared via the *Maphub Open Annotation API*. The first demo was bootstrapped with approximately 6,000 public domain maps taken from the *Library of Congress Historic Map Division*.
http://maphub.github.io
http://maphub.github.io/api

The *Maphub Open Annotation API* follows the *W3C Open Annotation* specification and uses *Apache Solr* for map full-text search.
https://www.w3.org/TR/annotation-model
http://lucene.apache.org/solr

Annotations can easily be added by creating overlays on top of map images. When a user opens a map to annotate zoomable raster images, *Maphub* suggests potentially relevant *Wikipedia* tags. The semantic tag enrichments are retrieved from *DBpedia*. By aligning with *DBpedia*, it is possible to exploit those connections to enrich annotations and their tags with additional information, such as the ability to search for a map by its content and not its title, and translations of terms in other languages.

The distinctive feature of this project is that, after adding at least three control points to a map, it is possible to calculate real world locations for any point on the map. This allows users to create different views through *Google Maps*. The historical map can also be laid over the *Google Earth* map, creating an overlay of the historic map onto its current day location. The *Google Earth* file can also be downloaded to a computer for later viewing.[16]

3.3.4. Example: images of cultural heritages

Over the last two decades, the web welcomed all kinds of new digital collections, domain-specific information resource portals, online exhibitions, and other products that expose and provide access to the not-born-digital resources hosted in LAMs. Digital images are a container for much of the information content in web-based delivery of images, books, newspapers, manuscripts, maps, scrolls, single sheet collections, and archival materials. Many of them are not good candidates for OCR, and therefore usually remain as digital images. In the digital age, access to these image-based resources is fundamental to research, scholarship and the transmission of cultural knowledge. On the other hand, cultural heritages objects, though steadily being digitized, are still difficult to find, reuse, cite, exchange, and compare.

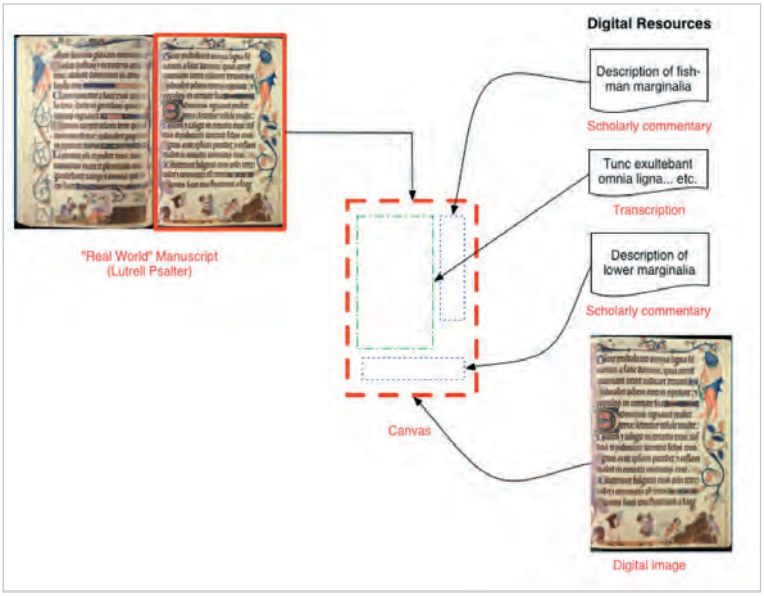


Figure 19. *Shared Canvas Data Model*.
Source: **Albritton**, 2013. Slide 12.

Image exchange APIs

As pointed out by the community that developed *IIIF* (*International Image Interoperability Framework*) APIs, much of the Internet's image-based resources are locked up in silos, with access restricted to bespoke, locally built applications.[17] API is the

abbreviation for "application programming interface", and is compiled with a set of routines, protocols, and tools for building software applications. To address the issue of interoperability and provide access to high quality image resources across all kinds of digital collections, four *IIIF* APIs have been developed, tested, and revised during the last five years and are already supported by major software for image servers, image clients, and image apps. They are: Image API, Presentation API, Content Search API, and Authentication API. An API for Audio/Visual materials is in progress.

One of the important fundamental concepts of *IIIF* is the standardized URI patterns. For example, the Image API specifies a web service that returns an image in response to a standard HTTP or HTTPS request. The URI can specify the region, size, rotation, quality characteristics, and format of the requested image. A complication for Image APIs is the varying quality, size, and format of the source images involved in the viewing, zooming, searching, annotating, exchange, etc. The *IIIF* Image API URI for requesting an image *must* conform to the following URI Template (**Appleby** *et al.*, 2012a): *https://iiif.io/api/image/2.1/#image-request-uri-syntax*

{scheme}://{server}{/prefix}/{identifier}/{region}/{size}/{rotation}/{quality}.{format}

For example:

http://www.example.org/image-service/abcd1234/full/full/0/default.jpg

Using the supported tools, annotations by experts or other contributors can be added to the canvas. Annotations, annotation lists, and content all have unique identifiers that can be processed by machines without confusion. [Figure 19]

Different resource types defined by the framework follow the recommended URI patterns (**Appleby** *et al.*, 2012b): [Figure 20]

Hundreds of successful cases of digital collections have been benefited from *IIIF* APIs. One example is *The National Library of Wales - Welsh Newspapers Online* which currently has 1.1 million *IIIF* images. Implementations of *IIIF* at many LAMs can be found via the demo links from the *IIIF* website as well as from the *IIIF* conferences held worldwide. *http://newspapers.library.wales*

Deep image annotation

Deep image annotation (DIA) is becoming an accepted new concept in treating images for semantic enrichment, aiming to improve the efficiency of image retrieval and access, enhance users' understanding of an image, and support automatic image resources integration and knowledge discovery.

It is widely recognized that a knowledge gap exists between access to knowledge in textual content versus image-based content. A vast amount of information resides inside images, photographs, paintings, diagrams and drawings which can be seen but not searched. It is waiting to be discovered, but has been inaccessible to traditional query methods. In addressing knowledge organization and discovery objectives for images, **David Clarke** (2015) provided a line of objectives:

1. Every image should have a URI and an extensible set of metadata.
2. Visual features within images should have URIs and an extensible set of metadata.

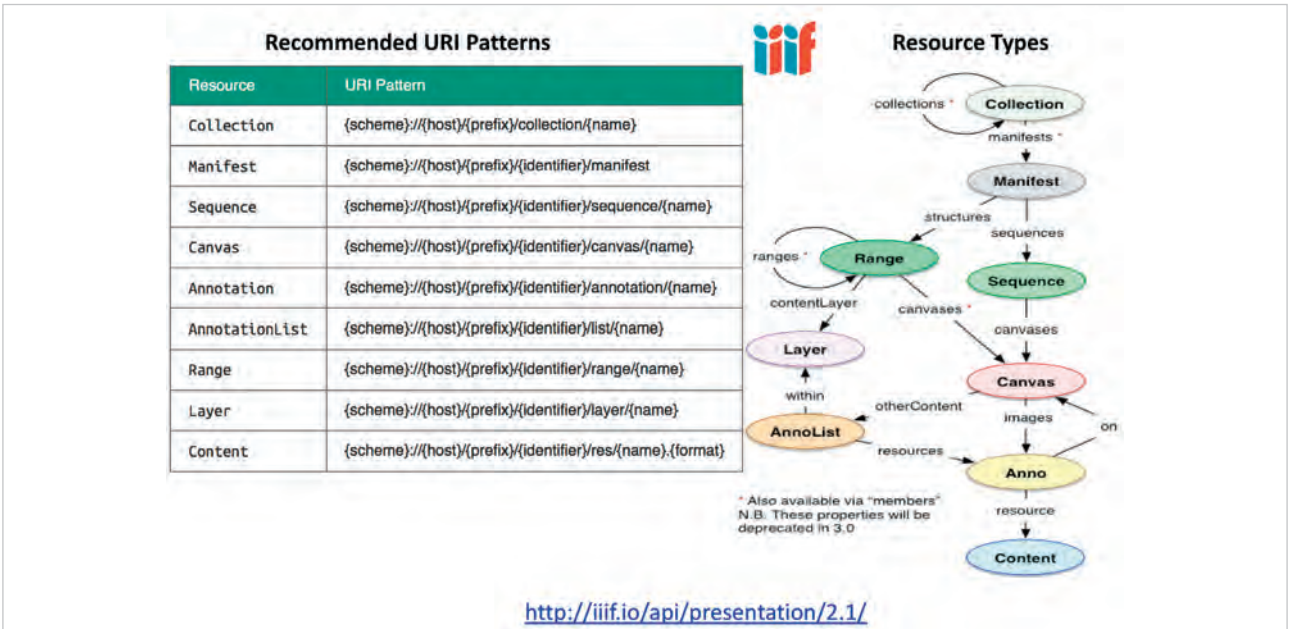


Figure 20. Recommended URI Patterns for *IIIF* resource types.
Source: Compiled based on the images from *IIIF* Implementation API specification.

3. Visual features should be identifiable in-line with the image, i.e., by using point-to- bound-area markers.
4. Visual features should be presentable as alphabetical lists.
5. Visual features should be presentable as ordered lists or hierarchical structures.
6. Images and visual features should be indexable using controlled vocabularies.
7. All image, annotation, and indexing metadata should be searchable.
8. Users should be able to pan-and-zoom to specific parts of an image from search results or browse lists.
9. Users should be free to pan-and-zoom anywhere on an image and discover the visual features in view and their related concepts.
10. The system should support the discovery of images and parts of an image that are conceptually related to any other image or part of an image. (**Clarke**, 2015, slide 12).

The *IIIF* APIs have enabled some of these goals to be reached. Other objectives spelled out in this list have been realized by tools like *Oasis*, *Mirador*, *Luna*, *Universal Viewer*, and others.

The following example demonstrates the deep semantic annotating results using *Oasis* for the *Dunhuang Mogao Caves*' "*Nine-colored deer*" painting (**Wang**; **Liu**; **Xia**, 2017). *Oasis* enables the visual features of an image to be individually identified and expressed as Linked Data URIs. These features can then be semantically indexed to internally or externally curated KOS vocabularies. The annotation documented on the left side of the figure represents hierarchical and ordered list structure, expressed automatically in SKOS properties to form a new LOD KOS vocabulary. [Figure 21a]

The narrative story painting of the *Nine-colored deer Jataka*, on the wall of the *Mogao Grottoes Cave* 257,[18] consists of eight episodes, in which the *Nine-colored deer* and other major characters occur multiple times. It reminds us of the value of semantic image annotating (with contextual information) which generates machine-processable data (not just machine-readable data). [Figure 21b]

Existing methods for indexing features and themes appearing in images of works are usually found at the metadata level. The existing metadata standards and cataloging rules provide detailed guidance about external characteristics of images. On the other hand, content-based semantic analysis and annotation models and practices have not been standardized; they are usually project-based, and are therefore less useful in discovering images and parts of an image that are conceptually related to any other image or part of an image. The impact goes to the granular resource aggregation and knowledge discovery of cultural heritages. For this reason, researchers of the Mogao Caves at *Wuhan University* in China established a workflow for image representation and annotation related tasks, with three semantic annotation models: (1) the macroscopic concept model, (2) the information hierarchy model, and (3) the structured image annotation model. The testing results prove that semantic enrich-

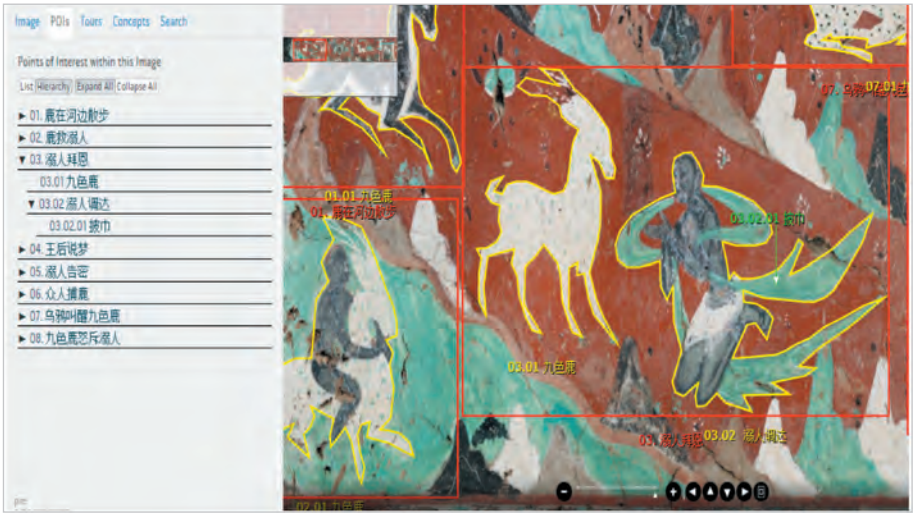


Figure 21a. A section of the semantic deep image annotation result.
Source: **Wang**, **Liu**, and **Xia** (2017).

Figure 21b. Image of the *Nine-colored deer Jataka*.
Source: *Dunhuang Academy*. *Digital Dunhuang*. Mogao Grottoes Cave 257, West Wall.
https://www.e-dunhuang.com/cave/10.0001/0001.0001.0257

ment can be significant in both the development of image information organization methodology and digital humanities research (**Wang**; **Liu**; **Xia**, 2017).

3.3.5. Example: cultural objects

Memory institutions and information services are increasingly embracing the new information technologies in order to meet the needs of the changing digital age. The challenges of three-dimensional objects are unique, in comparison with text-based unstructured data and two-dimensional materials. Innovation is needed because "data are for discovery and inspiration, not just management." (*Opencontext.org*)
https://opencontext.org

Online Coins of the Roman Empire (*OCRE*), was initiated as a joint project of the *American Numismatic Society* and the *Institute for the Study of the Ancient World* at *New York University*.
http://numismatics.org/ocre

This revolutionary new service is designed to help in the identification, cataloging, and research of the rich and varied coinage of the Roman Empire. The original goal of *OCRE* was to make available a digital corpus of all published Roman Imperial coin types. At the same time, the project aimed to expand the ability of all external contributors interested in linking any collection-based online catalog (**Reinhard** *et al.*, 2017). Since the first edition of *OCRE* launched in late 2012, the project has grown tremendously with international collaborations. All coin types from Augustus to Zeno (representing five

Built with Linked Data technology, *OCRE* becomes a knowledge base, much more than a traditional website.

While the searching, browsing, querying and visualization are supported by the Sparql queries, triple stores, and various apps, its ontology-based design make it an easy-to-use digital corpus, with downloadable catalog entries, incorporating over 43,000 types of coins. Users are provided with the multiple browsing, searching, and refining options.

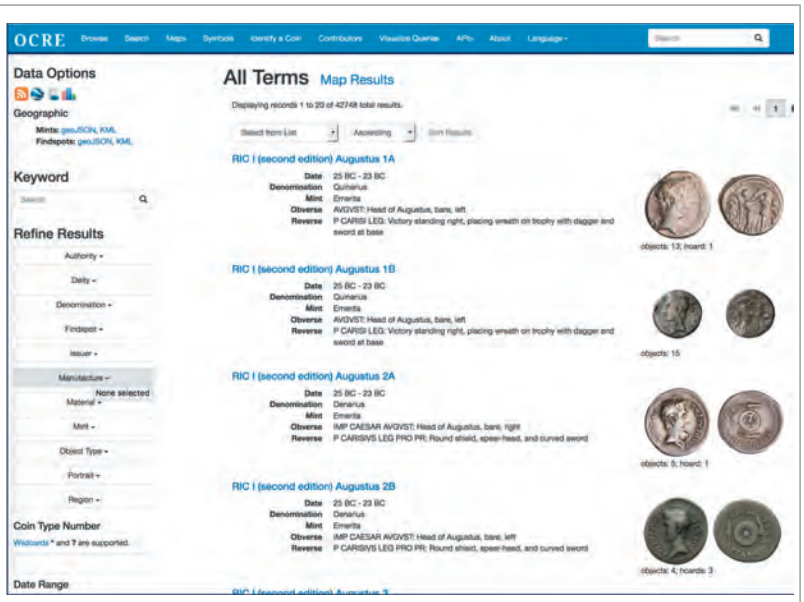


Figure 22a. *OCRE* browse page.
http://numismatics.org/ocre/results

OCRE offers detailed structured data about objects' typological descriptions, geo maps, and examples, as well as Quantitative Analysis options.

In this example, in the Quantitative Analysis section, data about the average measurements for this coin type are provided. A user can also select measurements (by axis, diameter, or wright), choose a chart type (bar or column), and request results of comparison according to the selected categories (Denomination, Mint, Region, Manufacture, Material, Authority, Portrait, and Deity).

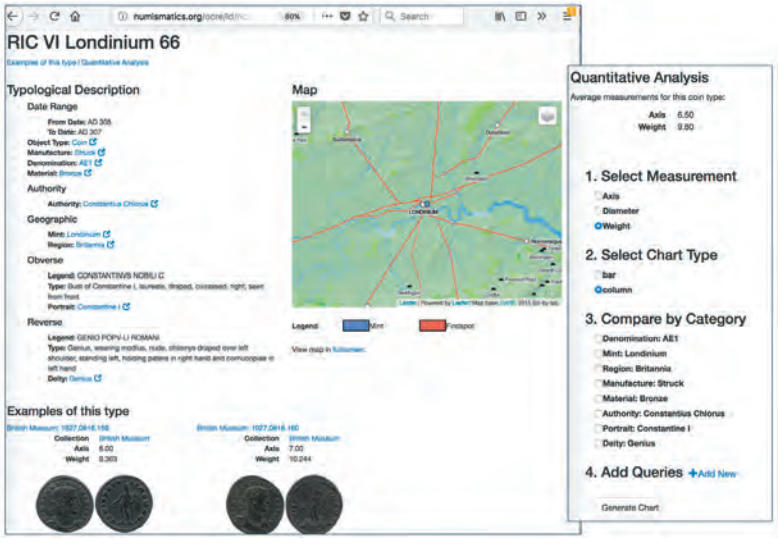


Figure 22b. *OCRE* individual coin type page.
http://numismatics.org/ocre/id/ric.6.lon.66

centuries of Roman imperial numismatics) have been published. *OCRE* incorporated more than 107,000 physical coins related to these coin types from 21 different datasets. These datasets originate from large collections as well as smaller civic or university museums, archaeological databases, and the Domuztepe excavations published through *OpenContext* which publishes research data on the web (**Gruber**, 2017). [Figure 22a and 22b]
https://opencontext.org

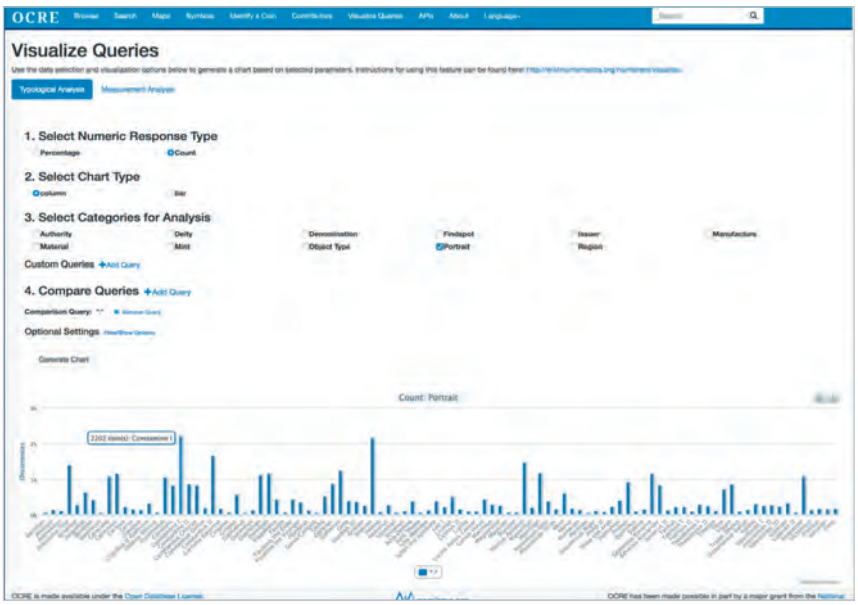


Figure 22c. *OCRE* visualize queries page.
http://numismatics.org/ocre/visualize

Even more useful to researchers is the visualized querying and analyzing across the whole dataset. Using the data selection and visualization options provided, a user can generate a chart based on selected parameters for typological analysis or measurement analysis. (Refer to the next figure.) [Figure 22c]

It is important to point out the meaning of this innovative resource to humanities researchers. End users are typically untrained in using LOD datasets through data dumps and Sparql endpoints (the most popular LOD deliverables). To master a query, one has to understand the syntax, forms, operators, result set modifiers, variables, and functions of the Sparql query language. Hence, easy access by end users becomes critical to the effective use of the LOD products. In order to better reuse digital resources while maximizing the outcomes for both machine and human users, easy-to-use LOD services have direct impacts on consuming these datasets as knowledge bases.

In the *OCRE* case, the visualize queries page (Figure 22c) allows the user to render queries in the form of charts and graphs, enabled by its JavaScript library at the back-end. Chart parameters are passed RESTfully by URL parameters, making it possible to bookmark and share charts over the web. Another powerful and useful analytic feature of *Numishare* is the maps interface. For example, the map page for hoard collections differs in one key way from other types of collections: The *Simile* timeline is incorporated into the interface through the TimeMap library. The points of the map correspond to find spots, and points in the timeline are created when hoard record contents contain datable coins.
http://wiki.numismatics.org/numishare:maps

Overall, users of *OCRE* are liberated from unfamiliar query languages and are provided effective analytics outcomes. This is the best example of how LAM data are used to support digital humanities researches. As **Gruber** (2017) pointed out, the study of coins has long been seen as an esoteric sub-discipline within history and archaeology, but the introduction of numismatics through intuitive user interfaces online serves as a bridge in exposing what is typically viewed as highly specialized information to a more general audience of archaeologists and classicists.

3.3.6. Discussion

It is obvious that there are limitless potentials for those [so far] un-structured data in LAMs to be used in supporting digital humanities research and education. This section showcased the pioneer products that have already brought LAM data to a totally new level of significance. It is also clear that Linked Data is a major concept that has been making semantic enrichment of a variety of LAM data in a distributed content creation environment possible and effective. Linked Data is about using the web to connect related data that wasn't previously linked, or using the web to lower the barriers to linking data currently linked using other methods.
http://linkeddata.org

The cases also imply that, semantic annotation, which formally identifies concepts and relations between concepts in documents, relies on both human and machine actions. Meanwhile, ontology-based approaches have already formed new mainstreams in digging into unstructured data and bringing unbelievable new discoveries to the front-end.

As pointed out by Robert Allen, some of the initial steps toward highly structured repositories may be relatively easy. However, scaling these will be difficult and requires active community engagement. Community involvement will also be required for policies and procedures in determining thresholds for consensus about results (**Allen**, 2017). It is obvious that the *IIIF* community introduced above is addressing such needs successfully and opportunely. Allen further suggested

that, for the representation of digital collections, rich semantic structures such as systems, models, simulations, events, and frames are needed (**Allen**, 2017). Each of the cases and research projects presented in this section so far have established their rich semantic structures at the back-end, delivering user-friendly, highly specialized information products to a more general audience at the front-end.

> "In data resources that are served through LAMs, unstructured data are usually available in the largest quantity in comparison with structured data, have the most diversity in type, nature, and quality, and are the most challenging to process. It is obvious that there are limitless potentials for those [so far] un-structured data in LAMs to be used in supporting digital humanities research and education"

Products such as *OCRE*, *Maphub*, and *Linked Jazz* can be considered as the representatives of Smart Data because they adequately represent a sufficient number of relevant features of humanistic objects of inquiry to enable the necessary level of precision and nuance required by humanities scholars, and also provide users with a sufficient amount of data to enable quantitative methods of inquiry, helping researchers to surpass the limitations inherent in methods based on close reading strategies (refer to **Schöch**, 2013). It is thrilling that LAM data providers and researchers in the humanities are incorporating the data-driven environment where advanced digital technologies have created the possibility of novel and hybrid methodologies.

3.4. Unifying heterogeneous contents in a distributed data creation environment

In digital humanities research, integration and interoperation of cultural heritage (CH) data are constantly in demand. An example was given by Prof. Eero Hyvönen in an invited talk at *VIII Encounter of documentation centres of contemporary art: open linked data and integral management of information in cultural centres*, held at the *Artium Museum*, Vitoria-Gasteiz, Spain, October 2016:

> "For example, if metadata about a painting created by Picasso comes from an art museum, it can be enriched (linked) with, e.g., biographies from *Wikipedia* and other sources, photos taken of Picasso, information about his wives, books in a library describing his works of art, related exhibitions open in museums, and so on. At the same time, the contents of any organization in the portal having Picasso related material get enriched by the metadata of the new artwork entered in the system. This is a win-win business model for everybody to join such a system; collaboration pays off. The linking can be established correctly only if unambiguous URI identifies are constantly used." (**Hyvönen**, 2016)

In order to achieve such a win-win result, it is important to realize the inherent needs and challenges. The heterogeneous contents could come with all kinds of formats and media, languages, cultural backgrounds, often companioned with provenance records and other contextual information. Yet another complicated situation is that their documentation has followed special professional standards and best practices defined and implemented in different professions involving libraries, archives, museums, and other cultural heritage institutions for a long time. The interoperability efforts of accommodating another community's model and data structure in one project might enable moving a step ahead, yet the uncertainty of quality and possible redundant tasks would be beyond prediction when diverse institutions are involved.

As a result, a fundamental semantic question in dealing with CH data is how to make the heterogeneous contents semantically interoperable, so that they can be searched, linked, and presented in a harmonized way across the boundaries of the datasets and data silos.

3.4.1 CASE: the *Sampo* portals

Hyvönen summarized three major semantic agreements that are needed for interoperability: (1) Domain neutral semantic model; (2) Metadata alignment model; and (3) Shared domain ontologies. The shared domain ontologies refer to the agreement of sharing domain ontologies (places, persons, etc.) whose concepts are used for populating the metadata models (**Hyvönen**, 2016).

In Figure 23, the data publication system is illustrated by a circle. A shared semantic ontology infrastructure is situated in the middle. It includes mutually aligned metadata and shared domain ontologies, modeled using Semantic Web standards. If content providers outside of the circle provide the system with metadata about a CH object, the data is automatically linked and enriched with each other and forms a *Giant Global Graph* (*GGG*) (**Hyvönen**, 2016).

The *Sampo* Model that has been tested and used in three cultural heritage case studies in Finland demonstrates such a win-win situation. *Sampo* is a mythical object in Finnish folklore that gives the holder wealth and good fortune. *https://en.wikipedia.org/wiki/Sampo*

In this instance, the *Sampo* model is a generic name for the model applied in various CH projects that created large-scale aggregated data sets for digital humanities applications from heterogeneous sources using Linked Data.

- *WarSampo*, the most notable and award-winning case,[19] was released in November 2015 and is the first large-scale system for serving and publishing WW2 LOD on the Semantic Web. The data draws eight different major datasets

from different organizations, originally totaling 7.6 million triples in a Sparql endpoint. The portal allows both historians and laymen to study war history and the destinies of their family members in the war from different interlinked applications such as Events, Persons, Army Units, Places, *Kansa Taisteli* magazine articles, Casualties, Photographs, and War Cemeteries. The data is annotated using a set of domain ontologies, including: 1) an ontology of the troops and their hierarchies, 2) persons with their ranks and roles, 3) place ontology of historical places, 4) event ontology of battles, politics, and other war time incidents, 5) an ontology of time periods, 6) ontology of weapons, 7) ontology of vessels, and 8) a subject matter ontology (**Hyvönen** *et al.*, 2016).
https://www.sotasampo.fi/en

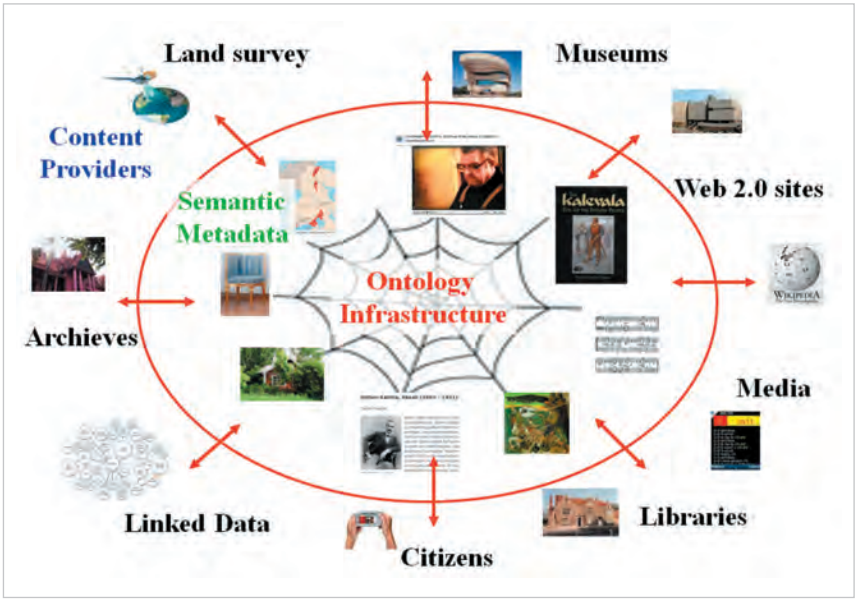


Figure 23. *Sampo* model for Linked Data publishing is based on a shared ontology infrastructure in the middle. Image source: **Hyvönen**, 2016

- *BiographySampo* is based on extracting knowledge from the underlying biographical texts —over 13,000 short biographies published by the *Finnish Literature Society*— using language technology, and by enriching the data through linking it to various external biographical databases, *Wikipedia*/*Wikidata*, collection databases of memory organizations, semantic web data services, etc. Similar to the *WarSampo* portal, users can find information via multiple different interlinked applications such as Persons, Places, Life maps, Statistics, Networks (of filtered people), Relations, and Language (**Hyvönen** *et al.*, 2018).
  http://biografiasampo.fi
- *NameSampo* for toponomastic research, is based on over two million places names collected in Finland and beyond. The objective of the project was to convert the place name entry slips, collection maps, and the attributes and metadata related to them into digital format (**Ikkala** *et al.*, 2018).
  https://seco.cs.aalto.fi/projects/nimisampo

The *Sampo* model has been applied to a series of semantic *Sampo* portals, resulting in more break-through LOD products. The portals include *CultureSampo* (2009), *BookSampo* (2011), *TravelSampo* (2011), *WarSampo* (2015), *NameSampo* (2018), and *BiographySampo* (2018).

3.4.2 Discussion

The *Sampo* model and the semantic portals demonstrated the win-win situation in dealing with CH data and making the heterogeneous contents of LAM data semantically interoperable, so that they can be searched, linked, and presented in a harmonized way across the boundaries of the datasets and data silos. They reveal a bright direction for the LAM data's next step in the semantic enrichment movement.

Parallel to the *Giant Global Graph* (*GGG*) featured by the *Sampo* model for Linked Data publishing —which is based on a shared ontology infrastructure—, integrating and reusing high level upper-ontologies and knowledge bases has led to another new product: *KBpedia*. Combining artificial intelligence (AI) with formal ontologies, the knowledge-based AI is pushing the speed of higher-level content analysis, entity recognition and categorization, and semantic network construction. A recently released open-source *KBpedia* built its knowledge structure by integrating seven core public knowledge bases — *Wikipedia*, *Wikidata*, *schema.org*, *DBpedia*, *GeoNames*, *OpenCyc*[20], and *Umbel*[21]. *KBpedia*'s upper structure, or knowledge graph, is the *KBpedia Knowledge Ontology* (*KKO*).
http://kbpedia.org

Written primarily in OWL 2, *KBpedia* includes 55,000 reference concepts, about 30 million entities, and 5,000 relations and properties, all organized according to about 70 modular typologies that can be readily substituted or expanded. Such a product characterizes the advanced semantic technologies that can be used for concept analysis and entity annotation, mapping, and data integration, plus automatic support for AI machine learning and semantic searching (*KBpedia*, 2018).

4. Summary and conclusion

With the rapid development of the digital humanities field, demands for smarter and bigger historical and cultural heri-

tage data, which usually cannot be obtained through web crawling or scraping, have focused attention toward LAM data, the treasure of all society. On the technology side, the semantic technologies and new artificial intelligence (AI) have brought innovative and implementable applications to the data-driven solutions. On the LAM side, the decade long digitization investment has formed LAM data as the most reliable resources for DH research (**Varner**; **Hswe**, 2016). Semantic enrichment is a strategy increasingly used during recent years, directly applied to the enhancement of LAM data (structured, semi-structured, and unstructured). In the "Discussion" sections dedicated to each of these data types in section 3, typical features and approaches are summarized and further discussions and resources are given. It is clear that there is no limit to the types of LAM data that can be semantically enriched. The real cases, research projects, experiments, and pilot studies shared in this article demonstrate endless potential for LAM data, whether they are structured, semi-structured, or unstructured, regardless of what types of original artifacts carry the data.

> Smart Data is a concept embraced by humanities research, and underlines the organizing and integrating processes from unstructured data to structured and semi-structured data, making the big data smarter (**Kobielus**, 2016; **Schöch**, 2013). Activities pertaining to LAM data enrichment have proved to be the initiator that enables LAMs to advance their data into smart data, supporting deeper and wider exploration and use of data in digital humanities research

Activities pertaining to LAM data enrichment have proved to be the initiator that enables LAMs to advance their data into smart data, supporting deeper and wider exploration and use of data in digital humanities research. The semantic enrichment strategy represents one major step and directly enhances LAM data by using semantic technologies. The cases and examples used in this article are representative of such activities. Following these roadmaps would encourage more effective initiatives and strengthen this effort to maximize LAM data's discoverability, use- and reuse-ability, and their value in the mainstream of DH and the Semantic Web.

Information technologies and new semantic technologies supporting each of the processes of semantic enrichment are advancing quickly. Reports of diverse experiments, tools, and issues can be found in publications across a wide range of domains, and in extraordinary depth. In addition to those discussed in the Discussion sections in section 3, AI-assisted processing is also becoming a norm for entity recognition, auto-recommendation, mapping, and verification. The fast extension of semantic technologies has been bringing astonishing news daily. Another closely related topic is increasing the findability of LAM data by exposing them to search engines, a benefit of semantic enrichment. Other articles in this journal will be dedicated to these topics, hence they are not covered in this article.

No matter what end products results from the semantic enrichment of LAM data, they will most likely join data on the web. Thus, this article concludes with the benchmarks recommended by the *W3C* in *Data on the web best practices*. The document focuses mainly on publishing data rather than consumption of data, and is geared toward data available through the web. It provides best practices related to the publication and usage of data on the web, so that data will be discoverable and understandable by humans and machines, while a self-sustaining ecosystem is facilitated. It is noteworthy that "provide metadata" is always recommended throughout the document. Each of the benchmarks also identify the ultimate goals for LAM data: comprehension, processability, discoverability, reuse possibility and effectiveness, trustiness, linkability, accessibility, and interoperability (**Farias-Lóscio**; **Burle**; **Calegari**, 2017). [Figure 24]

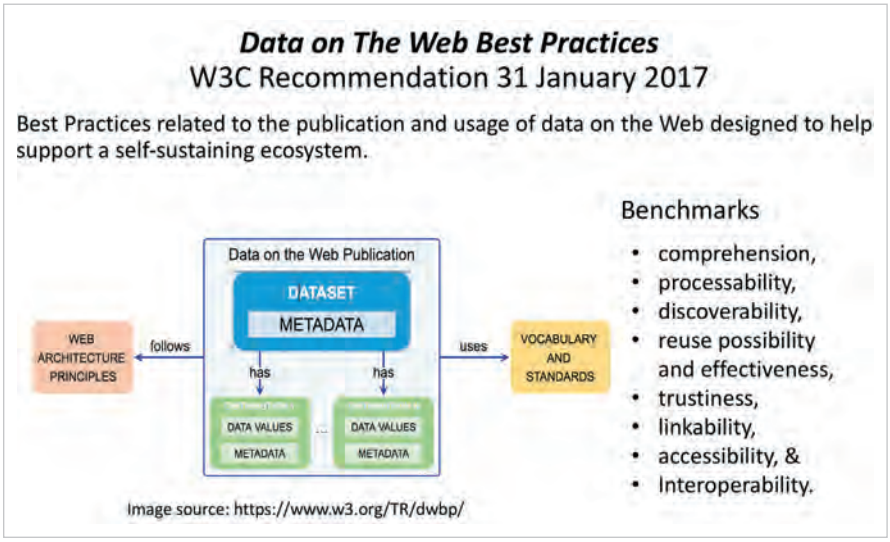


Figure 24.*W3C*'s data on the *Web best practices* recommendations and benchmarks.
Source: Created by the author based on the *W3C*, 2017.

Notes

1. As of 2016, participating nations and funding organizations include: Argentina (*MINCyT*); Brazil (*Fapesp*); Canada (*Sshrc*, *Nserc*, *FRQ*); Finland (*AKA*); France (*ANR*); Germany (*DFG*); Mexico (*Conacyt*); Netherlands (*NWO*); Portugal (*FCT*); United Kingdom (*AHRC*, *ESRC*), and United States (*NEH*, *NSF*, *IMLS*). Source: *https://diggingintodata.org/awards/2016/news/winners-round-four-t-ap-digging-data-challenge*

2. (*FRBR*) *Functional requirements for bibliographic records*

3. (*FRAD*) *Functional requirements for authority data*

4. (*FRSAD*) *Functional requirements for subject authority data*

5. *Europeana semantic enrichment* (November 5, 2015)
https://pro.europeana.eu/page/europeana-semantic-enrichment

6. Available from *Europeana* semantic enrichment website's link to "several vocabularies.
https://pro.europeana.eu/page/europeana-semantic-enrichment

7. According to April 19, 2018 note and updated Table 3 in **Manguinhas**, 2016.
http://shorturl.at/hyzL0

8. Data are available at:
https://www.kaggle.com/dorami/museum-project/data

9. *Getty Vocabularies: LOD* website
http://vocab.getty.edu

10. By clicking "View the full Getty record", a user is led to
http://vocab.getty.edu/page/ulan/500009365

11. *OpenRefine Reconciliation Service* (explanations and sample query templates).
http://vocab.getty.edu/queries#OpenRefine_Reconciliation_Service

12. *Museums and Collections with Maya Inscriptions*
http://mayawoerterbuch.de/museumscollections

13. Entry of *Museu Etnològic*
http://mayawoerterbuch.de/museums/museu-etnologic

14. Data collected on Oct. 23, 2018 from
https://old.datahub.io/dataset

15. "Data Sources. Oral History Transcripts."
https://linkedjazz.org/data-sources

16. Watch the video
http://maphub.github.io

17. About *IIIF*
https://iiif.io/about

18. Mogao Grottoes Cave 257, West Wall.
https://www.e-dunhuang.com/cave/10.0001/0001.0001.0257

19. *WarSampo* wins the *Open Data Prize* in the *2017 Lodlam Challenge*

20. *OpenCyc* (2002-2017) was a part of *Cyc* (/ˈsaɪk/), the world's longest-lived artificial intelligence project.
http://www.cyc.com

21. *Umbel* (*Upper Mapping and Binding Exchange Layer*)
http://umbel.org

5. References

Albritton, Benjamin (2013). *Digital manuscript interoperability: Shared canvas and IIIF in practice*.
https://slideplayer.com/slide/5840185

Alemu, Getaneh; **Brett, Stevens**; **Ross, Penny**; **Chandler, Jane** (2012). "Linked data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models". *New library world*, v. 113, n. 11/12, pp. 549-570.
http://eprints.rclis.org/17523
https://doi.org/10.1108/03074801211282920

Allen, Robert B. (2017). "Rich semantics and direct representation for digital collections". In: *ACM/IEEE Joint conference on digital libraries* (*JDCL*), pp. 348-349.
https://doi.org/10.1109/JCDL.2017.7991623

Appleby, Michael; **Crane, Tom**; **Sanderson, Robert**; **Stroop, Jon**; **Warnet, Simeon** (2012a). "IIIF Image API 2.1.1". IIIF.
https://iiif.io/api/image/2.1

Appleby, Michael; **Crane, Tom**; **Sanderson, Robert**; **Stroop, Jon**; **Warnet, Simeon** (2012b). "IIIF Presentation API 2.1.1". IIIF.
https://iiif.io/api/presentation/2.1

Bainbridge, David; **Hinze, Annika**; **Cunningham, Sally-Jo**; **Downie, J. Stephen** (2016). "Low-cost semantic enhancement to digital library metadata and indexing: Simple yet effective strategies". In: *2016 ACM/IEEE Joint conference on digital libraries* (*JDCL*), pp. 93-102.
https://core.ac.uk/download/pdf/44290466.pdf

Baker, Thomas; **Bermès, Emmanuelle**; **Coyle, Karen**; **Dunsire, Gordon**; **Isaac, Antoine**; **Murray, Peter**; **Panzer, Michael**; **Schneider, Jodi**; **Singer, Ross**; **Summers, Ed**; **Waites, William**; **Young, Jeff**; **Zeng, Marcia Lei** (2011). *Library linked data Incubator Group Final Report*. W3C Incubator Group Report 25.
http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025

Bensmann, Felix; **Zapilko, Benjamin**; **Mayr, Philipp** (2017). "Interlinking large-scale library data with authority records". *Frontiers in digital humanities*, n. 4, p. 5.
https://doi.org/10.3389/fdigh.2017.00005

Borgman, Christine L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press. ISBN: 978 0 262529914

Burdick, Anne; **Drucker, Johanna**; **Lunenfeld, Peter**; **Presner, Todd**; **Schnapp, Jeffrey** (2012). *Digital_Humanities*. Cambridge, MA: MIT Press. ISBN: 978 0 262528863

Clarke, David (2015). "Deep image annotation: Making a difference in knowledge organization". *Fourth ISKO-UK Biennial conference of the International Society for Knowledge Organization*.
http://docplayer.net/13812285-Deep-image-annotation-making-a-difference-in-knowledge-organization.html

Consultative Committee for Space Data Systems (2012). *Reference model for an Open Archival Information System*. Washington DC: CCSDS.
https://public.ccsds.org/Pubs/650x0m2.pdf

Damjanovic, Violeta; **Kurz, Thomas**; **Westenthaler, Rupert**; **Behrendt, Wernher**; **Gruber, Andreas**; **Schaffert, Sebastian** (2011). "Semantic enhancement: The key to massive and heterogeneous data pools". In: *Proceedings of the 20th intl IEEE ERK* (*Electrotechnical and Computer Science*) conference, pp. 413-416.
https://www.researchgate.net/publication/266603290_Semantic_Enhancement_The_Key_to_Massive_and_Heterogeneous_Data_Pools

Dunsire, Gordon; **Willer, Mirna** (2011). "Standard library metadata models and structures for the semantic web". *Library hi tech news*, v. 28, n. 3, pp. 1-12.
https://doi.org/10.1108/07419051111145118

Farias-Lóscio, Bernadette; **Burle, Caroline**; **Calegari, Newton** (2017). *Data on the web best practices. W3C Recommendation 31 January 2017*.
http://www.w3.org/TR/dwbp

Floridi, Luciano (2010). *Information: A very short introduction*. Oxford: Oxford University Press. ISBN: 978 0 199551378

Gardner, Dan (2012). "An ocean of data [Introduction]". In: Smolan, Rick; Erwitt, Jennifer (eds.). *The human face of big data*. Sausalito, CA: Against All Odds Productions, pp. 14-17. ISBN: 978 1 454908272

Gracy, Karen; **Davidson, Sammy** (2014). "Helping users find the 'good stuff': Using the semantic analysis method (SAM) tool to identify and extract potential access points from archival finding aids". In: *SAA Research Forum*, Society of American Archivists.
http://files.archivists.org/pubs/proceedings/ResearchForum/2014/posters/GracyDavidson-ResearchForumPoster2014.pdf

Gracy, Karen; **Zeng, Marcia Lei** (2015). "Creating linked data within archival description: Tools for extracting, validating, and encoding access points for finding aids". *Digital humanities conference of the Alliance of Digital Humanities Organizations* (*ADHO*).

Gracy, Karen; **Zeng, Marcia Lei**; **Skirvin, Laurence** (2013). "Exploring methods to improve access to music resources by aligning library data with linked data: A report of methodologies and preliminary findings". *Journal of the American Society for Information Science and Technology* (*JASIS&T*), v. 64, n. 10, pp. 2078-2099.
https://doi.org/10.1002/asi.22914

Gruber, Ethan (2017). "Final report to the NEH for online coins of the Roman Empire". *Day of archaeology*, July 28.
http://www.dayofarchaeology.com/final-report-to-the-neh-for-online-coins-of-the-roman-empire

Hinze, Annika; **Taube-Schock, Craig**; **Bainbridge, David**; **Matamua, Rangi**; **Downie, J. Stephen** (2015). "Improving access to large-scale digital libraries through semantic-enhanced search and disambiguation". In: *Proceedings of the 15th ACM/IEEE-CS Joint conference on digital libraries*. Association for Computational Linguistics, pp. 147-156.
https://doi.org/10.1145/2756406.2756920

Hyvönen, Eero (2016). "Cultural heritage linked data on the semantic web: Three case studies using the sampo model". *VIII Encounter of documentation centres of contemporary art: open linked data and integral management of information in cultural centres*. Artium, Vitoria-Gasteiz, Spain, October 19-20.
https://seco.cs.aalto.fi/publications/submitted/hyvonen-vitoria-2017.pdf

Hyvönen, Eero; **Heino, Erkki**; **Leskinen, Petri**; **Ikkala, Esko**; **Koho, Mikko**; **Tamper, Minna**; **Tuominen, Jouni**; **Mäkelä, Eetu** (2016). "Publishing Second World War history as linked data events on the semantic web". In: *Proceedings of the digital humanities conference*, pp. 571-573.
https://seco.cs.aalto.fi/publications/2016/hyvonen-et-al-warsa-dh2016.pdf

Hyvönen, Eero; **Leskinen, Petri**; **Tamper, Minna**; **Rantala, Heikki**; **Tuominen, Jouni**; **Keravuori, Kirsi** (2018). "Demonstrating BiographySampo in solving digital humanities research problems in biography and prosopography" [Submitted paper].
https://seco.cs.aalto.fi/publications/submitted/hyvonen-et-al-bs-2019.pdf

Ikkala, Esko; **Tuominen, Jouni**; **Raunamaa, Jaakko**; **Aalto, Tiina**; **Ainiala, Terhi**; **Uusitalo, Helinä**; **Hyvönen, Eero** (2018). "NameSampo: A linked open data infrastructure and workbench for toponomastic research". In: *GeoHumanities 18*, *Proceedings of the 2nd ACM SIG Spatial workshop on geospatial humanities*, Seattle, WA, USA, November 06-09, pp. 2:1-2:9, ACM.
https://doi.org/10.1145/3282933.3282936

IMLS (2018). *Transforming communities: IMLS strategic plan (2018-2022)*. Washington DC: Institute of Museum and Library Services.
https://www.imls.gov/sites/default/files/publications/documents/imls-strategic-plan-2018-2022.pdf

Isaac, Antoine; **Manguinhas, Hugo**; **Stiller, Juliane**; **Charles, Valentine** (2015). *Report on enrichment and evaluation*. The Hague, Netherlands: Europeana Task Force on Enrichment and Evaluation.
http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf

Kaplan, Frederic (2015). "A map for big data research in digital humanities". *Frontiers in digital humanities*, n. 2.
https://doi.org/10.3389/fdigh.2015.00001

KBpedia (2018). *KBpedia is now open source*, October 23.
http://kbpedia.org/resources/news/kbpedia-is-open-source

Kobielus, James (2016). "The evolution of big data to smart data". In: *Smart data online*, July 13 [PowerPoint slides].
http://smartdata2016.dataversity.net

Lin, Yuri; **Michel, Jean-Baptiste**; **Lieberman-Aiden, Erez**; **Orwant, Jon**; **Brockman, Will**; **Petrov, Slav** (2012). "Syntactic annotations for the Google Books Ngram corpus". In: *Proceedings of the ACL 2012 System demonstrations*. Association for Computational Linguistics, pp. 169-174.
http://aclweb.org/anthology/P12-3029

Manguinhas, Hugo (ed.) (2016). *Europeana semantic enrichment framework*. Version 17, Nov.
https://pro.europeana.eu/page/europeana-semantic-enrichment
http://shorturl.at/pEIQ5

Mayer, Daniel (2011). *Mainstream semantic enrichment* [*YouTube* video]. December 26.
http://www.youtube.com/watch?v=YVxvQ7UpqI0

Mayer-Schönberger, Viktor; **Cukier, Kenneth** (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Eamon Dolan/Mariner Books. ISBN: 978 0 544227750

MoMA (2017). *Museum of Modern Art Collection -- Title, artist, date, and medium of every artwork in the MoMA collection*, Last Updated: 2017 (Version 1).
https://www.kaggle.com/dorami/museum-project/data

Mukerjee, Prithwis (2014). "Introduction to data science" [*PowerPoint* slides], January 12.
http://www.slideshare.net/prithwis/01-intro2-datascienceyantrajaalblog

Mutuvi, Stephen; **Doucet, Antoine**; **Odeo, Moses**; **Jatowt, Adam** (2018). "Evaluating the impact of OCR errors on topic modeling". In: *Maturity and innovation in digital libraries. 20th Intl conf on Asia-Pacific digital libraries*, *ICADL 2018*, Ha-

milton, New Zealand, November 19-22, Proceedings, pp. 3-14. ISBN: 978 3 030 04257 8

National Archives (2016). "Finding aid type". *The national archives catalog*.
https://www.archives.gov/research/catalog/lcdrg/elements/findingtype.html

Nguyen, Thi-Tuyet-Hai; **Coustaty, Mickael**; **Doucet, Antoine**; **Jatowt, Adam**; **Nguyen, Nhu-Van** (2018). "Adaptive edit-distance and regression approach for post-OCR text correction". In: *Maturity and innovation in digital libraries. 20th Intl conf on Asia-Pacific digital libraries*, *ICADL 2018*, Hamilton, New Zealand, November 19-22, Proceedings, pp. 278-289. ISBN: 978 3 030 04257 8

O'Neill, Ed; **Mixter, Jeff** (2013). "Maximizing the usage of value vocabularies in the linked data ecosystem". In: *76th Annual meeting of the American Society for Information Science and Technology* (*ASIS&T*), Montreal, Canada, November.
http://nkos.slis.kent.edu/ASIST2013/ONeill-Mixter.pptx

Pattuelli, M. Cristina (2012). "Personal name vocabularies as linked open data: A case study of jazz artist names". *Journal of information science*, v. 38, n. 6, pp. 558-565.
https://doi.org/10.1177/0165551512455989

Pattuelli, M. Cristina; **Hwang, Karen**; **Miller, Matthew** (2016). "Accidental discovery, intentional inquiry: Leveraging linked data to uncover the women of jazz". *Digital scholarship in the humanities*, v. 32, n. 4, pp. 918-924.
https://doi.org/10.1093/llc/fqw0

Prasad, A. R. D.; **Giunchiglia, Fausto**; **Devika, P. Madalli** (2017). "DERA: from document centric to entity centric knowledge modelling". In: *Proceedings of the International UDC seminar 2017. Faceted classification today*. London, September, pp. 169-179.
http://seminar.udcc.org

Riva, Pat; **LeBoeuf, Patrick**; **Žumer, Maja** (2017). *IFLA library reference model: A conceptual model for bibliographic information*. Netherlands: IFLA.
https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf

Schöch, Christof (2013). "Big? Smart? Clean? Messy? Data in the Humanities". *Journal of digital humanities*, v. 2, n. 3, pp. 2-13.
http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities

Smith-Yoshimura, Karen (2018). "The rise of Wikidata as a linked data source". In: *Hanging together. The OCLC research blog*, August 6.
http://hangingtogether.org/?p=6775

Stiller, Juliane; **Petras, Vivien**; **Gäde, Maria**; **Isaac, Antoine** (2014). "Automatic enrichments with controlled vocabularies in Europeana: Challenges and consequences." In: *Euro-Mediterranean conf.*, pp. 238-247. Springer, Cham.
https://doi.org/10.1007/978-3-319-13695-0_23

Svensson, Patrik (2010). "The landscape of digital humanities". *Digital humanities quarterly*, v. 4, n. 1.
http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html

Thorsen, Hilary K.; **Pattuelli, M. Cristina** (2016). "Linked open data and the cultural heritage landscape". In: Jones, Ed; Seikel, Michele (eds.). *Linked data for cultural heritage*. Chicago, IL: Alcts Publishing. ISBN: 978 1 783301621

TiECON East (2014). *Data is the new oil*.
https://tieconeast.wordpress.com/page/2

Reinhard, Andrew; **Van-Alfen, Peter**; **Bransbourg, Gilles**; **Gruber, Ethan**; (2017). "Wishes granted: the ANS and the NEH". In: *National Endowment for the Humanities. Announces. New grant recipients*.
http://numismatics.org/pocketchange/wp-content/uploads/sites/3/NEH-Article-ANS-Magazine.pdf

Van-Ruyskensvelde, Sarah (2014). "Towards a history of e-ducation? Exploring the possibilities of digital humanities for the history of education". *Paedagogica historica*, v. 50, n. 6, pp. 861-870.
https://doi.org/10.1080/00309230.2014.955511

Varner, Stewart; **Hswe, Patricia** (2016). *Special report: Digital humanities in libraries*. American Libraries.
https://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries

W3C (2011). Library Linked Data Incubator Group Final Report
https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025

W3C (2017). Data on the Web best practices.
https://www.w3.org/TR/dwbp

Wagner, Elisabeth; **Matsumoto, Mallory**; **Kiel, Nikolai**; **Gronemeyer, Sven** (2014). *A checklist of museums with Maya Art*.
http://mayawoerterbuch.de/museumscollections

Wang, Xiaoguang; **Liu, Xuemei**; **Xia, Shengping** (2017). "Design and implementation of deep semantic indexing on digital cultural heritage images". *Journal of library and information science*, v. 43, n. 1, pp. 98-121.
http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/view/716

Weitz, Jay; **Toves, Jenny**; **Vizine-Goetz, Diane**; **Naught, Nannette**; **Bremer, Robert** (2016). "Mining MARC's hidden treasures: Initial investigations into how notes of the past might shape our future". Journal of library metadata, v. 16, n. 3-4, pp. 166-180.
https://doi.org/10.1080/19386389.2016.1262653

Zeng, Marcia Lei (2017). "Smart data for digital humanities". *Journal of data and information science*, v. 2, n. 1, pp. 1-12.
https://doi.org/10.1515/jdis-2017-0001

Zeng, Marcia Lei; **Gracy, Karen**; **Skirvin, Laurence** (2013). "Navigating the intersection of library bibliographic data and linked music information sources: A study in the identification of useful metadata elements for interlinking". *Journal of library metadata*, v. 13, n. 2-3, pp. 254-278.
https://doi.org/10.1080/19386389.2013.827513

Zeng, Marcia Lei; **Gracy, Karen F.**; **Žumer, Maja** (2014). "Using a semantic analysis tool to generate subject access points: A study using Panofsky's theory and two research samples". *Knowledge organization*, v. 41, n. 6, pp. 440-451.
https://pdfs.semanticscholar.org/bbeb/42b931fd32520a03167770d2b5de694128e6.pdf

Zeng, Marcia Lei; **Mayr, Philipp** (2018). "Knowledge organization systems (KOS) in the semantic web". *International journal on digital libraries*.
https://rdcu.be/PgZW
https://doi.org/10.1007/s00799-018-0241-2

Žumer, Maja (2018). "IFLA library reference model (IFLA LRM): Harmonisation of the FRBR family". *Knowledge organization*, v. 45, n. 4, pp. 310-318.

Also available in **Hjørland, Birger** (ed.). *ISKO Encyclopedia of knowledge organization*.
http://www.isko.org/cyclo/lrm

Žumer, Maja; **Riva, Pat** (2017). "IFLA LRM-Finally here". In: *Intl conf on Dublin Core and metadata applications*, Washington, D.C., USA, 26-29 October, pp. 13-23.
http://dcpapers.dublincore.org/pubs/article/download/3852/2037

Reimagining the academic library: What to do next. Review article

David W. Lewis

How to cite this article:

Lewis, David W. (2019). "Reimagining the academic library: What to do next. Review article". *El profesional de la información*, v. 28, n. 1, e280104.

https://doi.org/10.3145/epi.2019.ene.04

Article received on November 07, 2018
Approved on December 10, 2018

David W. Lewis ✉
https://orcid.org/0000-0001-9711-5565

Dean Emeritus, *University Library*
Indiana University Purdue University Indianapolis
522, Race Avenue, Lancaster, PA 17603, USA
dlewis@iupui.edu

Abstract

This article reviews the conclusions of the author's 2016 book, *Reimagining the academic library* and considers changes in scholarly communication and academic libraries that have taken place since its publication. Recommendations for alterations in the practice of individual libraries are provided. The problem of created integrated community-controlled open infrastructure is considered at length, especially the collective action problem that the library community must overcome.

Keywords

Academic libraries; Academic library futures; Scholarly communications infrastructure; Collective action problem; Open scholarly commons; Literature review.

1. Introduction

In May of 2016, my book, *Reimagining the academic library*, was published (**Lewis**, 2016). I finished the writing it sometime in the late fall of 2015. Over three years has passed since then and in those three years there have been significant changes to the world of scholarly communication and academic libraries. In this paper I will review the main points of my book in light of what has happened in the last three years, and lay out what I believe is required of academic libraries going forward.

To do this I will use the structure of the book's concluding chapter, "Conclusion: Ten things to do now."

As was the case with my book, my perspective will be primarily from the United States as this is what I know best. I will though endeavor to include perspectives and developments from other parts of the world.

2. The world we want: The open scholarly commons

To start, it is important to consider the goal. For me the goal is clear. We, the library community and others, need to create the open scholarly commons. We need to create a world where all scholarship and all of the world's cultural heritage is discoverable, openly available to everyone in the world, and preserved for future generations.

Doing so means that academic libraries will need to flip what they do. In the past they were primarily concerned with bringing knowledge from the world to their institutions or communities. In the future their focus needs to be on making the knowledge produced in or of interest to their institutions or communities available to the world by making it discove-

rable, accessible, and preserving it. This part individual libraries can do on their own. This flip was a major focus of my book. It is what Lorcan Dempsey has called moving from the "outside-in" library to the "inside-out" library (**Dempsey**, 2016). Libraries will be joined by museums and other cultural heritage organizations that will provide additional content. Government agencies, scholarly societies, and foundations will also contribute. The holders and stewards of the world's knowledge need to make it available to the world.

> "Academic libraries will need to flip what they do"

But for this work to have impact it needs to be paired with the collective creation of open integrated community-controlled infrastructure to bring all of the separate local contributions together to create a comprehensive world-wide resource. Jean-Claude Guédon describes the combination this way:

> "In the end, libraries can point out the fact that their future role actually points in two, apparently opposite, yet deeply complementary directions: on the one hand, they plunge deeply into the local production scenes since they aim at systematically sweeping, storing, preserving, and curating all that is produced in their hosting institution; at the same time, the libraries, with their sister institutions, are involved in the task of ensuring a vibrant knowledge-nurturing life for their documents: they will circulate, be discoverable, be interoperable, be evaluated, etc. With the first function, each library ensures it safe and strong function within its host institution; with the second function, the libraries connect to bring the knowledge infrastructure that we all really need" (**Guédon**, 2017).

What has become clearest to me in the past three years is that what is most important is not changing individual libraries, which was what I was focused on in my book. Changing individual libraries will matter, but the most important issue is the need for collective action to create the open integrated community-controlled infrastructure needed to support the open scholarly commons. As Alejandro Posada and George Chen put it after reviewing the monopoly tactics of the large commercial journal publishers, particularly *Elsevier* and *Wiley*,

> "To ensure the global resolutions of inequalities in discourse and scholarly representation there is thus a clear need for a community-driven integration of scholarly infrastructure, one that is aware of the potential of inequality preparation within the community rather than one which only seeks to add value and co-opt the process for the objective of rent" (**Posada**; **Chen**, 2018).

Library's contribution to funding such infrastructure was my book's ninth thing to do. Both creating and funding the infrastructure will be discussed at length below. It is quite clear that without robust infrastructure the content we all curate will be less useful and our efforts much less effective.

In attempting to create the open scholarly commons we face many challenges as the vested interests that currently control large portions of the scholarly communications ecosystem are well resourced and will not easily give up their positions of profit and control. Librarians as a profession, are often overly timid and we are too often unwilling to abandon the long-established practices we have carried with us from the print era. We also too often underestimate the resources we control and our ability to be the masters of our own fate. I am however optimistic. I believe that there are three fundamentals that work in our favor.

> "Librarians too often underestimate the resources they control and their ability to be the masters of their own fate"

2.1. Fundamental one: The money flows through academic libraries

Most of the money that funds scholarly communications flows through academic libraries. The most recent figures from 2014 put the total budgets for all U.S. academic libraries at over $8 billion with more $3 billion being spent on content and preservation (*Almanac*, 2018). This is a lot on money.

From an economic perspective the most efficient way to provide scholarly content requires subsidy. When costs are passed on to users, the content is not used to the extent that would provide the most societal benefit. Understanding this, institutions and communities created libraries to provide content free of charge to users. One of the key questions we face today is: given the changes in technologies, are libraries the best way to use the subsidy? The key thing to understand is that today the subsidy comes to the library and in general libraries are trusted to use the subsidy wisely. That is, libraries have the money and the ability to spend it.

In most cases libraries don't think of themselves as well-resourced because they have established commitments for all to their budgets and then some. The excessive cost increases of journals published by the large commercial publishers has exacerbated these budget pressures. What is often not considered is the opportunity cost of these commitments. Libraries have it within their power to redirect their budgets should they choose to do so. There are of course challenges in changing practice, not the least of these being our own reluctance to take risks and the general conservatism of the academy and faculty who are reluctant to change their practice, but should we choose to do so, redirecting our resources is an option. The example of the German and Swedish negotiations with *Elsevier* demonstrate that when the groundwork is done this is possible.

2.2. Fundamental two: The nature of digital content on the network is different from print

In the print world the economics of scholarly communication were based on scarcity and ownership, and too often monopoly control and monopoly rent taking. In the early digital world publishers, particularly large commercial and society publishers, carried over the practices from print as they generated significant profits. But digital content on the Web is very different from paper-based content. It has the following characteristics:

- A copy can be instantaneously delivered anywhere in the world.
- A copy is the same as the original.
- A copy can be made at zero marginal cost.

That is to say, as Andrew McAfee and Erik Brynjolfsson do, digital content should be, "Free, perfect and instant" (**McAfee**; **Brynjolfsson**, 2017, pp. 135-137).

This is how it can and should be, but as we all know this is not how it yet is. We can though see it moving in this direction. Information may not want to be free, but the economics argue that it should be as cheap as possible. Economics also argues that the most efficient way to deliver digital content is not by charging the user, or the user's library. Rather the most efficient way is to cover the first copy cost in some way and then distribute the content to users at no cost. That is theoretically the most economically efficient business model is open access. I have described in some detail why gold open access journals are more efficient and why they are likely to disrupt the subscription-based journal model (**Lewis**, 2012). Good data on the extent to which open access publishing has grown and what portion of the published literature is open access today has not been consistently collected, but a recent methodological strong study by Heather Piwowar and her colleagues suggests that about 50% of the journal literature published in 2015 was in one way or another open access (**Piwowar** *et al.*, 2018). The growth of open access is shown in Figure 1 from their study.

> "When costs are passed on to users, the content is not used to the extent that would provide the most societal benefit"

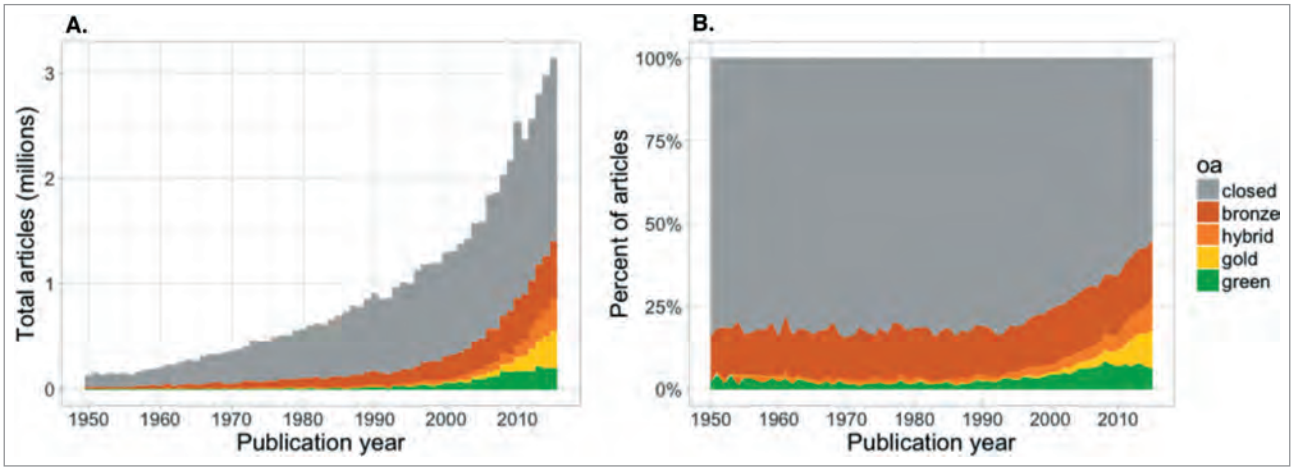


Figure 1. Total and % of articles available in open access 1950-2010 (**Piwowar** *et al.*, 2018)
https://peerj.com/articles/4375/#fig-2

If open access is the most efficient business model for publishing scholarly content on the network then the trends we have seen to date will continue, and, I believe, they will accelerate. Economics tells us that when the margin cost of an item is zero, and as noted above the margin cost of digital content on the network is zero, then the price that will maximize social good is zero. This of course means that the funding of the first copy cost and the cost of infrastructure need to be covered in some way besides sales of the content. This in turn means that subsidies are required, usually provided with taxes of some sort. How to accomplish this is a key piece of the problem we face. The important flip side of this is that when the cost to cover the first copy cost and the cost of infrastructure are passed on to users of the content, there is a significant dead weight loss. This means that societal benefit that should have accrued did not. Users who could have benefited from the content would not or could not pay to use it, and so the uses they might have made of the content did not get made. Scholarship and science slow, and society loses as a result.

Sci Hub exists, at least in part, because the dead weight loss in the current economics of scholarly communications is clear and in the minds of many justifies the seemingly petty theft that *Sci Hub* enables. Thus, we have what has sometimes has been call "Robinhood OA."

2.3. Fundamental three: Moore's law and cost reduction

We also need to recognize that Moore's Law is not going away, at least not any time soon. Technically, Moore's Law predicts

the number of transistors that can be put onto an integrated circuit chip. But Moore's Law is part of the larger phenomenon of digitization. Ray Kurzweil looked at earlier digital technologies and when he did it became clear that Moore's Law is really the continuation of a trend that began with the first use of digital technology on electromechanical devices, like punched cards, in the late 19th century (**Kurzweil**, 2010). It is important to recognize that the change Moore observed was not really about transistors on silicon, rather it was about representing information with ones and zeros so that machines could process it. Using machines to manage information began over 125 years ago and the capacity of this machine processing continues to grow at an accelerating rate.

> "The most efficient way to deliver digital content is to cover the first copy cost in some way and then distribute the content to users at no cost"

If we generalize Moore's Law to say that the capacity of digital technology doubles every 18 months, then in a decade that capacity increases 100 times. But this might understate the extent of the change. The chart below is from the *U.S. National Institutes of Health* and shows the decrease in the cost of sequencing a human genome (**Wetterstrand**, 2017). Note that this chart is on a log scale. The white line is the price decrease we would expect from improvements based on Moore's Law. We would have expected a decline in the price between 2000 and 2017 of two orders of magnitude. In fact, the decline was five orders of magnitude from nearly $100 million in 2000 to about $1,000 in 2017. If we translate this into "library units", this is the difference between building a good-sized university library and purchasing a dozen books.

So, how do we think about Moore's Law and the increasing capacity of technology? One way is to ask the economic question posed by Ajay Agrawal who says, when talking about artificial intelligence,

> "When looking at artificial intelligence from the perspective of economics, we ask the same, single question that we ask with any technology: What does it reduce the cost of?" (**Agrawal**, 2018).

Artificial intelligence says Agrawal reduces the price of prediction. Computing Agrawal argues reduces the price of arithmetic. Because price of arithmetic has decrease at the rate of Moore's Law for several decades, we can now do things with arithmetic that don't seem at all like arithmetic problems —like music or photography. As prediction gets cheap we can do things with prediction that don't seem like prediction problems, like self-driving cars, or maybe peer review. The later possibility is suggested by *Meta*, a machine learning system, with what the developers call *Bibliometric Intelligence*. As they state in a white paper,

> "The results of this large-scale trial demonstrate that *Meta* is able to perform 2.7x better than the best baseline estimator at predicting article-level impact for new manuscripts prior to publication. Additionally, it performed 2x better than the baseline at identifying 'superstar articles' –those that represent the top 1% of high impact papers, prior to publication" (**Yang** *et al.*, 2016, p. 6; *Aries Marketing*, 2016).

The baseline estimator was expert reviewers. Today this system is being marketed to journals to do preliminary screening of manuscripts and to match manuscripts to the most appropriate journal.

In the library world *Google* reduces the price of searching, the Web and digital documents on it, as we noted above, reduces the price of publishing. As Clay Shirky has said,

> "It makes increasingly less sense even to talk about a publishing industry, because the core problem publishing solves —the incredible difficulty, complexity, and expense of making something available to the public— has stopped being a problem" (**Shirky**, 2009).

Shirky likely overstates the case as there are still first copy costs in the academic publishing, but the technology should certainly reduce them. *Amazon* has reduced the cost and time needed to acquire physical objects, most importantly for us, books.

So, the cost of many of the inputs to the scholarly communication system and academic libraries are being reduced and this should create the opportunities to reduce the cost or increase the quality of what we do. In my view, these three fundamental changes will allow academic libraries collectively to reimagine and revise their practice and in doing so create the open scholarly commons.

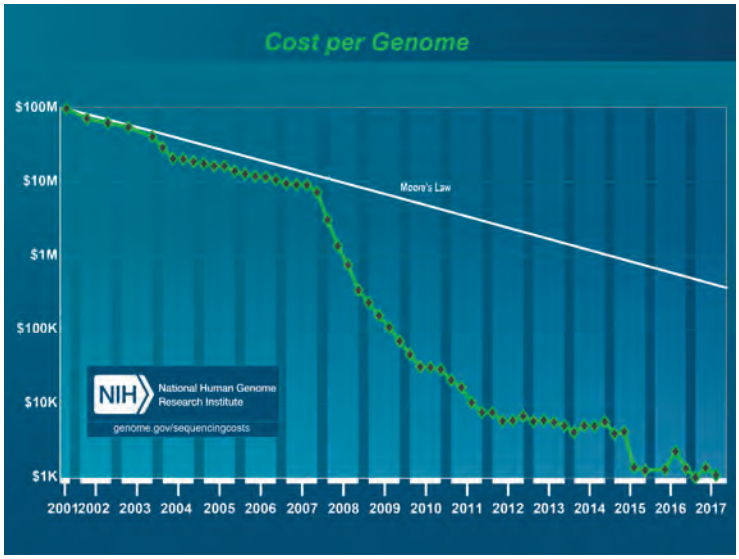


Figure 2. Decrease in the cost of sequencing a human genome (**Wetterstrand**, 2017) *https://www.genome.gov/images/content/costpergenome_2017.jpg*

That there is some wind at our backs does not necessarily mean that it will be smooth sailing. There are several real problems that must be faced and overcome. The three most important and difficult are:

> Economic tells us that when the margin cost of an item is zero then the price that will maximize social good is zero

- the general difficulty of creating infrastructure;
- the difficulty of combating monopoly rent taking by for-profit publishers given the current intellectual property regime; and
- the collective action problem that makes it very difficult for large groups to act together, even when it is in their own interest.

These problems will be addressed at some length later in the article.

3. Nine of the ten things to do now

3.1. Nine of ten introduction

Nine on the ten things I suggested in my book that needed doing had to do with how an individual library should change its practice. As we review them I would encourage you to keep the end goal in mind. Each library is doing its part to create a world-wide resource. In doing so we serve our institutions, but in a way that is different than it was in the past. We need to learn to tell a new story about the value of libraries and about the importance of this work.

As I did with my book, I want to start with an assertion.

> "Libraries have always done three things:
>
> - They have kept documents for the long haul.
> - They have provided the knowledge and information that the communities and institutions that fund them need.
> - They have assisted individuals in finding and using information.
>
> They have done so to assure that communities and individuals are productive and so that civilizations are long-lasting" (**Lewis**, 2016, p. xi).

Many things change when content moves from print on paper to digital documents on the network, but these long-standing functions remain a good place to start in thinking about academic libraries in today.

First, the documents that will be kept for the long haul will mostly be locally produced by faculty and students or will be of special interest to the campus or the local community. Most, though not all, of these documents will be digital and most will be unique or special in some way. The responsibility for this content will include making it discoverable, primarily through network level tools, and preserving it.

> The role of the library will be to assist researchers in making their work part of the open scholarly commons

Second, while the library will continue to purchase some content and make it available to students and faculty, much of the academic literature and much of the primary information that is the grist of research will be openly available on the Web, and so providing them to users will not be a library function. Rather the role of the library will be to assist researchers in making their work part of the open scholarly commons. This assistance will make the work available to all those who can benefit from it and thus will enhance the reputation of the researcher and that of the library's institution.

Finally, the library will continue to assist users in finding and using information, but this will be not about using the local library. Rather it will mostly be about productively and ethically using that open scholarly commons that will large and complex.

I expect that for the foreseeable future there will be buildings called libraries. They will, as they are now, be in the middle of the campus and will be the primary non-classroom academic space on campus. They will in most case continue to house some paper book collections, especially special collections, but their primary role will be as a place for study and for librarians and other experts to meet with and assist students and faculty with a variety of tasks related to finding, using, and maybe most importantly creating information. Many of these experts will likely be called librarians and many will have traditional librarian credentials, though we can expect that both the titles of the experts and the training and credentials they bring to their positions will become more diverse.

3.2. One: Retire the legacy print collection now

We have known since the Kent study of the use of library materials at the *University of Pittsburgh* in 1979 that a large portion of the books in most academic libraries will never be used (**Kent**, 1979). We have also known since the Fussler and Simon study of book use at the *University of Chicago*, published in 1961, that the only reasonable predictor of future book use is past use (**Fussler**; **Simon**, 1961). Libraries have had for several decades an available strategy for culling books

from their collections and thus reducing the space need for collections. As Paul N. Courant and Matthew "Buzzy" Nielsen have shown this cost is not trivial. Their estimate was $4.26 per year in 2009 dollars (**Courant**; **Nielsen**, 2010, p. 91). Daniel Gore outlined this strategy in his 1976 well-reasoned essay "Farewell to Alexandria" (**Gore**, 1976), but no one paid any attention. At the time building large collections of unused books made some sense. Books, especially scholarly monographs, went out of print quickly and the ordering process and interlibrary loan were both slow. Buying and holding as many books as a library could afford was a good insurance policy against the possibility of use, even if that possibility was not large. It was also the case that the main measure of library quality was library size, so there was limited incentive to examine or change the practice on filling stacks with as many books as possible.

That was then. We are now in an environment where most scholarly work is available in digital form that is almost always acquirable instantaneously. Print book use is declining and print books are often available from *Amazon* or other providers in a day or two. Print-on-demand technology means that books don't go out of print. *GoogleBooks* and *HathiTrust* provide electronic access to most public domain books. This makes user-driven purchase strategies for both digital and print books a good option. Simply put, it means never again buying a book that will not be used by at least one person. In addition, interlibrary lending is much faster and easier. With *WorldCat* you can locate available copies of books across the world and *FedEx* or other delivery services can move them quickly and at reasonable cost. Only partly in jest I have suggested that libraries should allow users to select he titles they need for purchase and then the library would give the book to the user to save on the cost of storage (**Lewis**, 2010).

All of this means that countries or regions can begin to think of their library's collections as one, what Lorcan Dempsey and his colleagues have called the "collective collection." Their work suggests,

> "If the current growth trajectory of the *HathiTrust Digital Library* is sustained, we can project that more than 60% of the retrospective print collections held in *ARL*[1] libraries will be duplicated in the shared digital repository by June 2014."

For *ARL* libraries, there could be cost avoidance of $500,000 to $2 million per year and space savings of more than 45,000 assignable square feet could be achieved through shared print provision (**Dempsey**; **Lavoie**; **Malpas**, 2013, pp. 80-81).

For these savings to be achieved two things need to happen. First, there needs to be a national or regional strategy to assure that some appropriate number of use and preservation copies of all print books are preserved, preferably in high quality storage facilities. Second, individual libraries need to be prepared to deal with the potential objections of faculty who would be unhappy with the removal of books from their local library collections. Fortunately, at least in the United States, there is more than enough high-quality storage capacity. While somewhat dated, you can use the data on the capacity of storage facilities (**Payne**, 2007) and data on the number of book and copies of these books (**Lavoie**; **Schonfeld**, 2006) to demonstrate that the capacity to store the United States nation book collection exists (**Lewis**, 2016, p. 108).

OCLC's *Sustainable Collection Services* unit provides a service that many libraries and library consortiums have used to determine which books in their collections are not used and what other libraries hold them. Thus, it is possible for a library to identify and withdraw titles and be assured that copies of these titles are available both nationally, and if the study has been done by a consortium, which partners will continue to hold the title. Several initiatives of this sort are already underway (**O'Gara**; **Osterman**, 2015; **Levenson**, 2015). This reduces the need to be overly concerned with faculty pushback on withdrawals as access to the withdrawn title can be assured. And, as we will see in the next section, the ability to repurpose library stacks when print volumes are removed is generally inexpensive way to acquire space.

3.3. Two: Develop a space plan

Academic library space is often the most valuable space on campus. It is generally centrally located. The building is generally open long hours. For most library building built after the Second World War, the space is often reasonably inexpensive to renovate. It is also the case that space is the most valuable resource on a university campus, so many people will propose many uses for "excess" library space.

Given the high value of space, it is important for the library to have a plan for the use of space as book stacks are removed. An important part of such a plan is the justification keeping the space in the library's hands. Such a justification is likely to have two parts. First, the library needs space for new services, especially those related to assisting faculty and students in creating and managing digital scholarly content. Second, on most campuses it should be possible to make a case for expanding student study space, both the variety of the space and the absolute number of seats. In my experience, making the second argument is more powerful as it ties the physical library to student academic success, which can be especially powerful on many campuses.

There is no doubt that libraries are special places on campus. Using methods drawn from the psychology of religion, Heather Lea Jackson and Trudi Bellardo Hahn make a case for the library as sacred space. As they put it,

> "This empirical study affirmed our hypothesis that spaces deemed as 'sacred' or 'sanctified' produce affective benefits for people that extend beyond attitudes and into the realm of behavior (projected library use). Circulation statistics do not measure these benefits; students may not actually use the books on the shelves, but they 'sanctify' the books —being around the books makes them feel more scholarly and connected to the institution's educational mission" (**Jackson**; **Hahn**, 2011, p. 436).

Academic libraries can be seen as "third spaces." The concept of "third spaces" comes from Ray Oldenburg's book *The great good place* (Oldenburg, 1989). Oldenburg describes a place between work and home that is neutral and informal, where everyone is welcome, but the tone and behavior expectations are set to be "regulars." Third space tend to be plane rather than flashy and the mood is often playful. I have argued that for students because the library has the characteristics of Oldenburgh's third space it becomes their home away from home (Lewis, 2017a, pp. 169-170).

> "For students because the library has the characteristics of Oldenburgh's 'third space' it becomes their home away from home"

There has been a push over the past decade or two to create space for group study in libraries and to add computers both for individual and group use. While they are expensive to build, where they exist, enclosed group study rooms are inevitably the most used space in the library, and often the most used space on campus. This push for group spaces is important, both because students are increasingly being give group assignments and they need somewhere to work together. There is also a trend for students to study alone in groups, each student doing their own work, but in the company of friends.

The design firm *Gensler* found, after a survey of over 1,200 students that while group study space is important, most students still look to the library for individual study. As the report on their study says, "While students report the library as a preferred space for both individual and group activities, their conception of the ideal library skews toward the solitary. Their top-ranked library qualities center around space to complete focused, individual work, and when asked what resources are most important to libraries today, 'quiet space for students' ranks first. When asked about the future, the rising importance of digital resources and connectivity becomes clear, while quiet space shows its continued importance" (*Gensler Research*, 2015, p. 3).

A number of other studies have similar findings (**Regalado**; **Smale**, 2015; **Applegate**, 2009).

The *Gensler Research* report makes three recommendations for designing library space:

- Focus on pragmatic, individual student needs.
- Don't sacrifice quiet in pursuit of collaboration.
- Integrate technology as a companion to traditional tools (*Gensler Research*, 2015).

The last recommendation means, at the very least, that you should put electrical outlets and USB ports everywhere you can.

An interesting study by John K. Stemmer and David M. Mahan provides a nuanced view of the impacts of library use, both of space and resources, on student academic success and how these relationships change over the course of an undergraduate's time at the university. While in most cases library use is positively correlated measures of student academic success, in some cases it is not. For example, studying in the library late at night is not (**Stemmer**; **Mahan**, 2016).

Given the value of space on every university campus, libraries need to make the case for using the space they create as books are removed from the collection. There are any number of good studies that document both the impact of library space on student success and the best ways to design this space. It will be important for library leaders to begin making arguments based on these studies and of input from their own students, to campus leaders. This should begin before the books are gone so that when the time comes a strong case for the library's perspective on how the space should be used has been presented and hopefully the arguments accepted.

3.4. Three: Have a materials budget strategy to manage the transition from traditional publishing models to open access

The budget strategy of most libraries, to the extent that they have one at all, is to buy what they bought last year if they can, or more likely cover price increases of journal subscriptions by making whatever cuts they can manage.

This is far from what is currently required.

As has been discussed above, digital documents differ in important ways from their print predecessors. Library practice based on digital documents should and will soon differ as well. In order to make appropriate contributions to the open scholarly commons, individual libraries will need to change what they do and what they spend their money on. It is unreasonable to expect that universities will increase the funds they provide libraries so libraries will need to reallocate their existing budgets. They will need to stop spending as much as they now do to purchase content, particularly expensive commercially published science and technology journals, and move that money to supporting access, discovery, and preservation of locally produced content, supporting open access publishing though article processing fees and other means, and importantly supporting the development and maintenance of the infrastructure needed to support the commons. We will discuss the latter at length later in the article. For now, it is enough to know that together these expenditures will make up a sizable portion of a library's budget once the transition is complete. I speculated that today the infrastructure portion of this expenditure should be 2.5% of the library's total budget (**Lewis**, 2017c). I now believe that it is likely to need to be three to five times that at least, and it will have to grow at a noticeable rate in future years. In five to ten years I believe support for activities to develop the open scholarly commons will need to consume over half of library expenditures.

> Given the value of space on every university campus, libraries need to make the case for using the space they create as book are removed from the collection

What is important now is to change the library's budget so that there is some surplus available today to invest in what will be important in the future. These funds only become available if libraries are aggressive in changing the way they spend their money. If libraries only make cuts only to match journal price increase there will be no surplus and no investment, and thus no escape from our current dependence on exploitive commercial publishers.

Fortunately, changes in the market place for books and journal articles provide opportunities for the required cuts. I believe that the strategy that can be implemented by most libraries that will provide a surplus that can be invested in the kinds of initiatives need to move to an open scholarly commons. I have outlined this strategy in more detail in my book and in a 2013 article (**Lewis**, 2013; 2016). The strategy has four parts:

- Moving to a patron-drive acquisitions (PDA) model for book purchases.
- Purchase individual articles as an alternative to journal subscriptions.
- Replace subscriptions with open access.
- Replace databases with free web content.

The first two parts are critical and can be implemented by any library as a way of saving money. The latter two require that open content continues to become available and so libraries are dependent on developments that are largely beyond their direct control. However, libraries can be assertive in taking advantage of these developments as they occur. Let's now look at each part of the strategy in turn.

A patron-drive acquisitions (PDA) system for books involves loading records into the library's catalog for the books that the library is prepared to purchase, but not actually purchasing them until a user requests the book. This approach has been widely used for electronic books for nearly 20 years. For electronic books this strategy is particularly effective because delivery can be instantaneous. It can now also be reasonable applied to print books given the speed at which many titles can be attained from *Amazon* and other providers. Param Bedi and Jason Snyder document *Bucknell University*'s library adopted an exclusive PDA model for books. In two years, materials expenditures dropped nearly 75% and circulations remained constant (**Bedi**; **Snyder**, 2015). One approach would be for a library to use an approval plan as a way of selecting bibliographic records. It would also seem possible where there copies of both the electronic and print book are available and the print version is what the user wants to provide access to the electronic version immediately so the user has access to it for the several days the print book is in transit.

Purchasing articles as an alternative to journal subscriptions works when subscriptions are expensive and use is relatively small. There are a variety of tactics for determining how best to make journal cuts. The method used by the *Université de Montréal* is often cited as best practice (**Gagnon**, 2017). However, the large commercial publishers are masters at making cancelling titles difficult, most effectively through the "big deal". The "big deal" was never a good deal. Kenneth Frazier's 2001 predictions have all come true:

> "In the longer run, these contracts will weaken the power of librarians and consumers to influence scholarly communication systems in the future. Librarians will lose the opportunity to shape the content or quality of journal literature through the selection process. Those who follow us will face the all-or-nothing choice of paying whatever publishers want or giving up an indispensable resource. The largest publishers will not only have greater market power to dictate prices" (**Frazier**, 2001).

The first step is to get out of the "big deal" so that the library can cancel individual titles. This is a huge hurtle for many libraries. While there is much discussion about getting out of "big deals", the number of libraries doing so has been relatively small. The *SPARC Big deal cancellation tracking tool* recorded only 13 "big deal" cancellations in 2017 (*Sparc*, sf).

Once libraries can make title-by-title cancellations, tradeoffs can be made. In general, individual articles can be purchased for between $25 and $40 on the publisher's website or for slightly less from services like the *Copyright Clearance Center*'s *Get it now*. It is straight forward to calculate the cost of purchasing articles against the cost of the subscription, but it needs to be remembered that in most cases libraries do not lose access to the backfiles they have subscribed to in the past so those articles will remain available at no additional cost. Over time the value the paid for backfile will decline and more individual document purchasing will be required as the backfile ages, but at the outset, this makes a big difference in the number of articles that must be purchased on-demand.

There are some indications that the journal usage statistics (*Counter*) reported by publishers may be inflated. Ted Bergstrom analyzed *Counter* data from over 5,000 journals from the *University of California* system and found publisher-based discrepancies. As he says,

> "some publishers recording significantly more downloads than would be predicted by the characteristics of their journals… this raises the question of what causes such substantial differences across publishers once journal and discipline characteristics are accounted for" (**Bergstrom**, 2018).

Following on this work Bergstrom, Richard Uhrig and Kristin Antelman examined publisher weblogs from four publishers

that documented over 435,00 downloads. They found that often users looked at both the HTML and PDF versions of an article and there was often repeated use of the same article by the same user over several months. Bulk downloads were also included in the *Counter* totals. They conclude,

> "the *Counter* JR1 and JR5 reports conceal information on duplicate downloads that is crucial for evaluating subscriptions" (**Bergstrom**; **Uhrig**; **Antelman**, 2018, p. 7).

This may mean that for some journals use is less than we believe and replacing subscriptions with article purchase will be less expensive than expected. In addition, it is possible to limit the article purchase service to faculty or graduate students and ask others to rely on interlibrary loan. Or, all users could be presented with the choice of having the library pay $10 to $15 for an interlibrary loan that might take 24 hours or having the library pay $25 to $40 to immediate access. I suspect many users can afford to wait and will take the cheaper option.

It is important to note that moving to purchase-on-demand saves not only money that might have been spent on materials. It also saves staff time. Records for PDAs are generally batch loaded and this saves cataloging staffing that would have gone to adding records one at a time. Acquisitions staff time is also saved as most ordering can be automated. There may be some increase in interlibrary loan work, but it should be possible to automate article purchase through the library's link resolver. As will be discussed below, this staffing can be reallocated to other tasks that support inside-out library functions.

The third and fourth parts of the strategy require that more content be open access or generally available on the Web. The trick will be to recognize when this is the case and to cancel purchased items promptly. To take the easy example, how long did it take to accept that *Wikipedia* was the definitive source and cancel subscriptions to other encyclopedias. Is *PubMed Central* or *Europe PMC* an adequate source of the medical literature if you are not a medical school? How many math journals can the *ArXiv* replace? Asking these questions, especially about science and technology resources, where much of the spending is, should be done regularly.

In the *University of California*'s *Pathways to open access* report a working group drawn from across the *University of California* system review a variety of strategies for directing funds away from paywalled subscriptions and towards open access publishing (*University of California Libraries*, 2018). They review a number of approaches to open access, Green OA, APC based Gold OA, and non-APC Gold OA, and suggest strategies and their challenges and opportunities.

The particulars of a library's strategy will vary, but something like what has been outlined above is possible for everyone. What is important is to develop your particular strategy and discuss it on campus. Managing the price increases of for-profit science and technology journals should make a strong case for change, but it is equally important to talk about the need to make investments that lead to the ultimate goal. Managing the present is only half of what needs to be done, we also need to create a future that better serves our institution and scholarship at large.

Finally, these changes will be easier if there are metrics that demonstrates the library's progress in advancing its new strategy. I have suggested several, including the percentage of the library spent on on-demand purchasing and the portion of the library's budget allocated to inside-out activities versus outside-in activities (**Lewis**, 2017b). Measuring allows the library to set tangible goals and know whether or not they are being met. The strategies suggested above will be a challenge to deeply embedded past practice and without tangible and measurable goals the organizational changes necessary to implement them will be much more difficult.

3.5. Four: Support the creation of, access to, and preservation of the scholarly content created on your campus

Here I am talking about digital content. There are several things that are necessary to support an effective program. It can be built over time, but the library should have a plan for putting all of the pieces in place.

The first requirement is the technical infrastructure to manage the content. This would usually mean one or more repository systems and might mean journal or even monograph hosting systems. These resources can either be provided with local technical support on local technology or the capacity can be acquired from a service provider.

The second requirement is the human resources to support faculty and students. There are a number of roles that are useful:

- Librarians who can assist with copyright and how and where works can and should be deposited (in either local or disciplinary repositories).
- Librarians to assist with data management.
- Metadata expertise to assure that the content is discoverable.
- Expertise in digitization of all sorts of formats.
- Staff who can do much of the required work for faculty. This could be clerical staff or students.
- While not directly related, expertise in metrics, especially how faculty can use them to make cases for promotion and tenure, can be very useful.

Depending on the size of the library, the number of staff involved in these functions could be one or two, or in a larger

institution, a dozen or more. Building the staff expertise for these functions will take time and require examining how to use open positions when opportunities arise. In some cases, existing staff can be trained to take on new responsibilities. In other cases, it will be necessary to bring the expertise into the organization with new hires. But as noted above, the changes in collection building strategies should free positions that can be used to fill these needs.

The third requirement is a robust preservation strategy for digital content. The first step is good backup of servers, but more is required. Unfortunately, while some progress has been done this is still an area without good established practice. Long term preservation requires technical systems for managing bits and assuring that the bits that go into the systems are the bits that come out. This has to be done with multiple and changing formats and all of this needs to be documented with stable metadata. The institutional structures that support the technical structures also need to be robust with some assurance of longevity. Even if it is housed in a technically sound system, without long-term institutional support for content can easily be abandon, at which point it is likely lost. Finally, there needs to be a financial model that can support these functions for at least decades. Balancing financing so that it is both robust enough to assure the long-term preservation of the content and is at the same time affordable is a considerable challenge. The uncertainty of the rate of technological change is a wildcard. If Moore's Law remains in effect financial structures can be built on the assumption of cheaper storage and processing, but how certain is this over decades? Either you need to make assumptions about future costs and create an endowment to cover these costs, which is the approach taken by the *Digital Preservation Network*. Or, you trust that the library will have the long-term budget and institutional commitment to continue paying annual fees. There is always the risk that preservation organization will fail or be forced to cut corners in financially difficult times. *Lockss* succession plan is an attempt to manage these risks and is a good development (**Schonfeld**, 2018). The general point is that currently digital preservation carries inherent risks.

There are currently no obviously superior solutions. As a recent *Ithaka S+R Issue brief* states:

> "There is a significant level of confusion about the purpose and business models of existing and nascent preservation services. How do such services fit together in creating a comprehensive preservation service framework? What are the gaps and redundancies? Do they compete with each other or work towards a complementary ecosystem? How do they interact with institutional digital asset management systems? Such questions indicate that the community would benefit from a deeper understanding of the available tools vis-à-vis preservation, discovery, and access stages" (**Rieger**, 2018, pp. 8-9).

My perspective at this point is that libraries should triage the digital content to identify what is most vulnerable, either because it was born digital and the library holds the only copy or because it was digitized and the cost of digitizing it again is prohibitive. The variety of preservation options for this content should be explored and implemented. It is also important to begin to build a budget for digital preservation as it will be a crucial function for the library going forward, even if we don't fully understand how best to do it today. Remember, one of the core functions of the library is to keep documents for the long haul.

The final thing that is necessary, or at least very helpful, is to have a policy environment on campus that supports the library's efforts to capture content so that it can be made accessible, discoverable, and that it is preserved. At a minimum, thesis and dissertations should be archived by the library as a matter of policy, whether or not this work is also provided to a commercial vendor. Next, institutions should have an open access policy so there is an expectation that faculty will deposit articles and similar work in the institutional repository. There are currently nearly 100 institutions in with such policy who are members of the *Coalition of Open Access Policy Institutions* (*Coapi*), most in the United States. This is a significant group, though open access policies are far from universal. Overlaying any campus policy will be national policies such as the *National Institutes of Health Public access policy* or the coming European *Plan S*. Helping researchers negotiate these policies can be a valued serve that the library should offer.

> "Libraries should triage the digital content to identify what is most vulnerable, either because it was born digital and the library holds the only copy or because it was digitized and the cost of digitizing it again is prohibitibe"

To the extent that there are opportunities libraries should work to make open access part of campus policies relating to promotion and tenure. Today this is rare, a recent study of promotion and tenure guidelines of 129 institutions in the United States and Canada found that only 5% (six institutions) had any explicit mention of open access in their guidelines (**Alperin** *et al.*, 2018, p. 18). However, it is possible as has been shown by *Indiana University—Purdue University Indianapolis* (**Odell**; **Coates**; **Palmer**, 2016) and *Harvard's School of Engineering and Applied Sciences* (*Harvard Library*, 2014) or more forcefully at the *University of Liège* (**Rentier**, 2015).

Good policy is important, but not sufficient. Ruth Kitchin Tillman studied faculty self-deposit in institutional repositories and found that it rarely happened (**Tillman**, 2017). Clifford Lynch, who in 2003 declared institutional repositories to be "essential infrastructure for the scholarship in the digital age," has recently expressed doubt about the possibility of successful implementation of open access policies and success in giving items into institutional repositories (**Lynch**, 2003;

2017). But, as demonstrated by *University—Purdue University Indianapolis* (*Iupui*), it is possible to get substantial levels of deposits in institutional repositories. In 2017, *Iupui* reported a 70% level of deposit —1,986 articles deposited of 2,824 eligible articles (**Odell**, 2018, p. 3).

3.6. Five: Commit to the special collections your library will support and make the required investment

In the future a large part of what will make a library special, what its reputation will rest on, are the items it holds that are rare or unique. Many of these items will be the works produced by students and faculty as was discussed above. For many libraries, especially research libraries, these will also be special collections of both tangible and digital items.

Fortunately, libraries know how build these collections, it is mostly a matter of reallocating resources. Some of the savings, both money and staff, from altering the collections strategy should be directed to enhancing special collections.

To the extent possible tangible items should be digitized and records for these items should be added to national directories such as the *Digital Public Library of America*, the *Canadian National Digital Heritage Index* or *Europeana*. The advantages of doing so are clear. Everyone in the world has access to what previously required visiting a specific library, which would have been prohibitive for most people. Secondly, items in digitized collections can brought together digitally in ways that were previously impossible allowing comparison of items physically located in far flung collections. Finally, in most cases the digital copy is adequate for the user's needs and the original copy need not be consulted. This saves ware on the original tangible item. In 2013, *Ithaka S+R* and the *Association of Research Libraries* (*ARL*) conducted a study of eight particularly successful programs. Among its findings were that the ability to have access to shared infrastructure was important, that a diversity of funding sources provided greater long-term stability of projects, and that a clear understanding of how the resource will be used leads to a more valued product (**Maron**; **Pickle**, 2013).

“Some of the savings, both money and staff, from altering the collections strategy should be directed to enhancing special collections”

Since few foundations now offer grants to digitize collection, one strategy for getting collections digitized is to work with a commercial vendor to have the collection digitized in exchange for the vendor having the right to sell access for a period of time, usually ten years. The institution providing the content gets access to the digital version at no cost. The *ARL* has developed a useful set of principles to help libraries who choose this option (*ARL*, 2010).

Managing special collections in the digital environment will require the rethinking of some practices. In the past items in special collections were generally available to researchers, but putting them on the Web for the world to see is something else. As Peter B. Hirtle, Anne R. Kenney, and Judy Ruttenberg put it,

> “Collections that were once made available to scholarly researchers under the watchful eye of special collections librarians in a physical reading room can now be made readily accessible to the entire world via the Internet. Donors who were willing to allow access to materials in a controlled setting could be taken aback by the trajectory of increased access” (**Hirtle**; **Kenney**; **Ruttenberg**, 2012).

For gifts acquired in the future can be accommodated with more explicit deeds of gift, but for items donated in the past it will not always be simple. Even if the library can make a good case for their legal right to digitize and make an item available on the Web this might be inappropriate if it is clear that the donor expected more restricted access.

As is the case with preservation systems the current mix of systems that can be used in special collections and archives workflows is not integrated and the organizational capacity of the systems providers is not always robust. Work is being done in this area, but investments from libraries beyond what they currently do will be required if there is to be a full suite of integrated and adequately funded systems to support the required workflows.

3.7. Six: Infuse the curriculum with the skills necessary to create and consume information productively and ethically

The *Ithaka S+R 2015 Faculty survey* found that 54% of the U.S. faculty surveyed said that their undergraduate students possessed

> “poor skills related to locating and evaluating scholarly information.”

This was a noticeable increase from the 47% of faculty who said this in the 2012 survey. About two-thirds strongly agreed that improving their undergraduate students'

> “research skills related to locating and evaluating scholarly information”

is an important goal for the courses they teach. About half of respondents strongly agreed that

> “librarians at my college or university library contribute significantly to my students' learning” and that “librarians at my college or university library contribute significantly to my students' learning by helping them to develop their research skills.” (**Wolff-Eisenberg**; **Rod**; **Schonfeld**, 2015, p. 58).

While librarians might hope for stronger support, there is a clear understanding among many faculty that their student need better information and research skills and that librarians have a role to play in enhancing them.

In the U.S. the *ACRL Framework for Information Literacy for Higher Education* was adopted on 2016. The Framework replaced the earlier *Information Literacy Competency Standards for Higher Education* and was needed, as the *Framework* introduction states because,

> "the rapidly changing higher education environment, along with the dynamic and often uncertain information ecosystem in which all of us work and live, require new attention to be focused on foundational ideas about that ecosystem" (*ACRL*, 2016, p. 2).

The *Framework* is made up of six concepts that anchor the frames are:

- Authority is constructed and contextual
- Information creation as a process
- Information has value
- Research as inquiry
- Scholarship as conversation
- Searching as strategic exploration

As the introduction goes on the state,

> "At the heart of this *Framework* are conceptual understandings that organize many other concepts and ideas about information, research, and scholarship into a coherent whole… each library and its partners on campus will need to deploy these frames to best fit their own situation, including designing learning outcomes" (*ACRL*, 2016, p. 2).

In the U.K., using somewhat different terminology a *JISC* report, *Developing digital literacies*, states,

> "Digital literacies are those capabilities which fit an individual for living, learning and working in a digital society" (*JISC*, 2014).

Figure 3 from the *JISC* report shows the seven digital literacies.

Both the *ACRL Framework* and *JISC*'s *Digital literacies* are complex, rich and ambitious. The challenge, and it is a large one, is to marshal the library's resources so that these literacies are a significant part of the university's curriculum. Heidi Julien, Melissa Gross, and Don Latham report on a national online survey of over 600 instructional librarians they identified challenges that are familiar to anyone engaged in this work:

- A lack of time and conflicts with other job responsibilities.
- The prevalence of the "one-shot" model and lack of a structured program that builds skills through the student's career.
- Student's lack of understanding of what they don't know and are not motivated.
- Lack of understand and support from faculty.
- Lack of support from library administration.
- Difficulties with assessment (**Julien**; **Gross**; **Latham**, 2018, pp. 187-189).

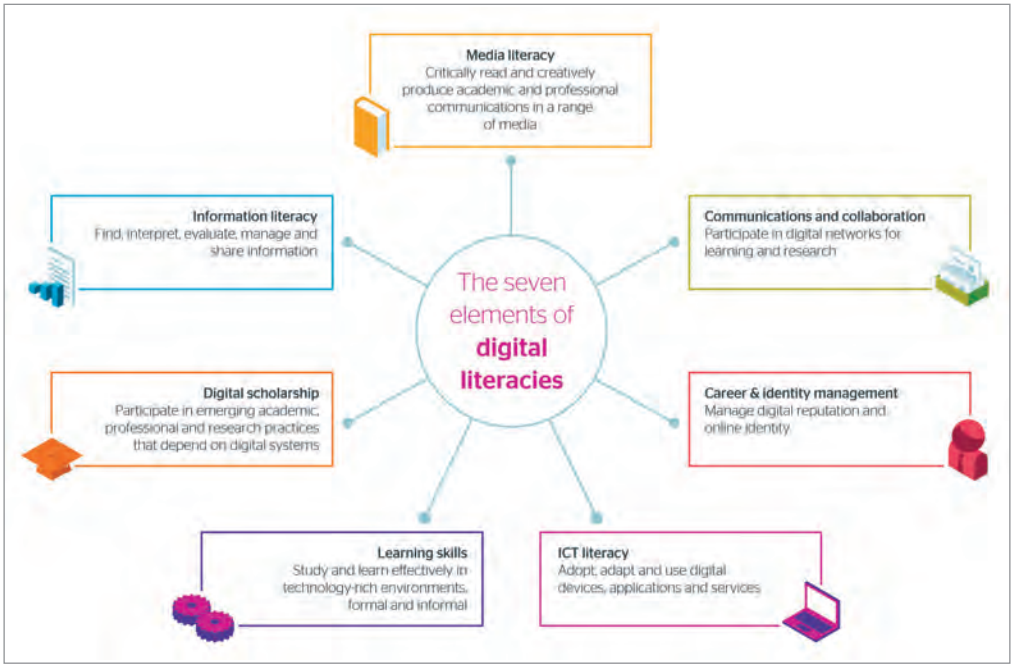


Figure 3. The seven digital literacies (*JISC*, 2014)
https://www.jisc.ac.uk/guides/developing-digital-literacies

The library literature is filled with reports of various approaches to information literacy and research skills instruction, but even the most successful rarely achieve impact across the curriculum and from the freshman experience to graduate programs. It is inevitable that individual institutions will need to allocate instruction resources to meet the most pressing needs of their campus. Most libraries have librarians who are passionate about this part of the library's mission. In my view the key is to provide support for this passion and to build a structure program across the student's academic career. It will also be important to assess the results of instructional efforts both to improve them and to demonstrate to campus administration the importance and success of this work.

3.8. Seven: Understand the demographics of your organization and have a plan to hire or develop the expertise your library will need

The simple truth is that the academic libraries will need to navigate the difficult technological and service changes that this article describes, at least at the beginning with the people they have today. You cannot wait until you have the ideal personnel in place to begin the work. Unfortunately, the people you have today will not have all the skills and expertise that are required. This will be a challenge. Opportunities to add new staff with new types of expertise will be limited and training current staff will only be able to accomplish so much.

The first thing to understand is the structural change in the type of staff in academic libraries. There has been an overall decline in staff in academic libraries. Clerical positions and the use of student workers both declined notably. At the same time there has been a modest increase in the number of librarians and a significant increase in the number of other professionals. Between 1998 and 2012, U.S. academic library staff declined by 4.7%. The number of clerical staff declined about 10% and the number of FTE student workers declined over 17%. The number of librarians increased over 7% and the number of other professionals increased nearly 50% (*National Center for Educational Statistics*). *https://nces.ed.gov/surveys/SurveyGroups.asp?Group=5*

One way to think about this is that your library will trade two clerical positions, as the work they have done is eliminated, and replace them with one professional. The cost will be about the same, but the library's headcount will decline.

The trends outlined above are likely to continue or even accelerate as digital content become predominate and labor-intensive paper content declines. Figure 4 shows this data.

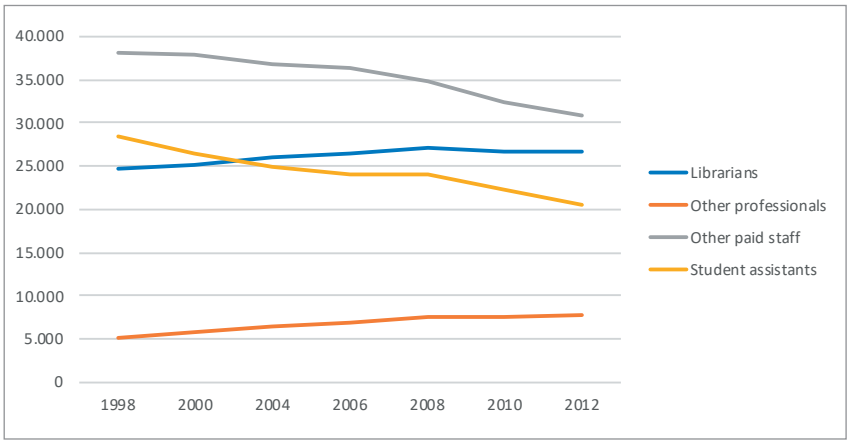


Figure 4. U.S. library workforce make-up based on *NCES* data

Kindra Orr and I have look at library staffing through the lens of generational change and described two particular challenges. We believe that:

> "Stated most simply, the challenges are:
> - Academic libraries need to exploit new technologies using new service strategies to be effective. They will need to do so with a librarian labor force that consists of a large number of baby boomers; many of whom will remain in the workforce for nearly another decade. At the same time millennium librarians, who will replace the baby boomers, need to be attracted and provided an environment that will allow them to develop and grow, and that will productively use their skills and energy. It will be important that both groups be productive and make contributions, but creating an organization with a structure and culture to do both will be difficult because in many cases aspirations and needs of the two groups will differ and be at cross-purposes.
> - As baby boomer leaders retire, Gen X and millennium leaders will need to replace them. It is unclear if there are a sufficient number of seasoned leaders in these generations. The result could easily be a large number of inexperienced new academic library leaders who will face steep learning curves" (**Lewis**; **Orr**, 2018, p. 1).

These challenges are driven by the fact that the baby boomer generation (born 1946-1964) was very large and the Generation X (born 1965-1980) was smaller. Baby boomer librarians held, and held on to, most leadership positions so that Generation X librarians could often not develop leadership experience until recently.

With the large numbers of baby boomers retiring it will be necessary to attract large numbers of millennium librarians (born 1981-2000). In my experience the pool of talented new librarians has been strong, but there are questions about the pipeline and whether traditional sources of librarians, that is library schools, will be adequate. I expect many libraries will look for the talent they need from other sources and will create positions that require expertise and credentials other than the MLS. As noted above, we can already see this happening in library staffing statistics.

Finally, there is the issue of diversity. In the U.S., academic librarians are notably whiter than the general population, the student body, and the faculty (**Lewis**, 2016, pp. 69-70). As Hui-Fen Chang documents, there has been little progress between 1980 and 2010 in diversifying academic librarianship in the U.S. (**Chang**, 2013). Unfortunately, the MLS pipeline is not much more diverse than the current population of practicing librarians (**Albertson**; **Spetka**; **Snow**, 2015, p. 93). This means increasing the diversity of the staff of any given library will be a challenge. A comprehensive study by *Mellon Foundation*/*Ithaka S+R* looked at inclusion, diversity, and equity in *ARL* libraries (**Schonfeld**; **Sweeney**, 2017). The findings where similar to other earlier studies. Interestingly the study asked library directors if there were barriers to achieving diversity and 81% said there where. This was particularly true with respect to race/ethnicity and particularly true with attracting a diverse pool of applicants. The study considered diversity at all levels of staffing and found that because non-MLS staff tended to be more diverse than librarians that pathways for these staff to enter professional positions might be an opportunity.

> "In relation to diversity, in the U.S., academic librarians are notably whiter than the general population, the student body, and the faculty"

All of this means that it is essential that academic libraries have a clear understanding of the demographics of their current staffing and formulate strategies to acquire the new expertise they require both through training and hiring. Academic libraries, even large ones, are relatively small organizations, and as such they have a narrow margin for error when making personnel decisions. Academic libraries also tend to make long term commitments to staff so that opportunities for hiring new staff tend to be infrequent and so bad hiring decisions have lasting impact. Moving forward without a clear understanding of their current situation and a plan for getting to where they need to be is not a good approach.

3.9. Eight: Get the culture right

So here is the situation. The library will need to change its practice and do different things that will require its staff to have new skills and expertise. Many long serving librarians will see what once made them expert professionals become unimportant. They will have to learn how to do new things. Newer and younger librarians will want to move quickly and will often find the library's usual deliberate pace of change demoralizing. Clerical staff will see that their role in the organization diminish and their numbers grow smaller. Faculty and administrators will express concern with many of the changes that need to be made. Money will be tight.

On the other hand, these are very interesting times and the work of librarians can be challenging and exciting. We are inventing libraries, actually scholarly communications, for the digital age. In many ways, there has never been a better time to be a librarian or to do library work.

The challenge for library leadership, and actually everyone in the organization, is to create a library culture where everyone works together for understood goals, where trust and respect are the norm, and where accomplishments lead to pride and celebration. This is difficult task one that will take time and effort.

There are many articles and books in the business literature that address organizational culture and managing change. The issues libraries face are not unique, and so dipping into this literature is useful for library leaders. Rather than attempting to summarize this literature, I will review my experience and make recommendations based on it.

It continues to surprise me how little time and effort most academic libraries spend on organizational development. At *Iupui*, twice a year we held a three day "organizational week" in which all staff participated. The dean would give a state of the library talk and then there would be a variety or training and/or planning activities sometimes with internal resources, but more often with the help of outside consultants. There was always a meal or two together. These activities allowed staff from different parts of the library to be together and get to know each other better. It focused on how to make the organization better and more effective. It provided the dean a regular chance to let all of the library staff know what issues the library faced and how the library was attempting to deal with them. This commitment to spend time and money on organizational development activities had clear payoffs. It led to shared sense of both the libraries goals and how we would reach them.

Many look at creating a strategic plan as the way to build and express the goals of the library and the means of achieving them. It is hard to do strategic planning quickly and without in involving lots of people. This has the advantage on engaging the library's many stakeholders, but it tends to consume the organization for one or more years. I believe that a strategic strategy, like the one laid out in my book, this article and others, is important (**Lewis**, 2007; 2016), but have come to see strategy planning as not the best use of an organizations time and energy. Planning is useful in some areas as I have suggested above, for example collections and personnel, where there are long-term commitments and large sums of money at stake. But for many areas the particulars of how best to move forward are unclear and the better approach is to experiment is small ways and learn from success and failures. Within the strategic strategy staff can use their own initiative in ways that tend to be difficult when strategies and tactics are set by a strategic plan.

It is a bit of a cliché, but I believe it is true. If you hire good people, point them in the right direction, trust their ability to do good, and accept and learn from failures, then the organization will get to where it needs to go. Stop periodically to assess how you are doing and celebrate your successes. If your organization can do this then it has the right culture.

3.10. Nine (in the book item ten): Sell the change

There are three constituencies who need to be sold on the changes the library will need to make: students, faculty, and administrators.

Students are the easiest. Given them good space and the library will be rated by students as one of the most values services on campus. Provide a good instruction program and by the time they graduate many students will come value what librarians can contribute to their educations.

The basic truth about faculty is that they have achieved their positions because they have mastered the scholarly system as it was and how it is. They live very busy lives and changing the way they do their scholarship and how they publish takes time and effort that they don't feel they have. The *Ithaka S+R 2015 Faculty survey* documents this clearly. Among the major finding of the survey were:

- "There is no observable trend towards a format transition for monographs. Faculty members' preference for using scholarly monographs in various ways in print format rather than digital format has, if anything, increased since the previous cycle of the survey.
- Faculty members prefer to be self-reliant in their data management and preservation processes… Nearly 90% of respondents organize these data on their own computer.
- Traditional scholarly incentives continue to motivate behaviors around research and its dissemination. Respondents generally believe that more recognition should be awarded for traditional research publications, such as journal articles and books… [and] are most interested in reaching scholars in their specific subdiscipline or field of research… consistent with findings with the previous cycle of the survey" (**Wolff-Eisenberg**; **Rod**; **Schonfeld**, 2015, pp. 6-7).

Jingfeng Xia conducted a meta-analysis of a several dozen studies of faculty attitudes toward open access publishing going back 20 years. He found,

> "that although an increase in the publishing and awareness rates of scholars with regard to OA journals has been observed, scholars have been consistently concerned with the low prestige of such journals and their lack of peer review, which is not the case in practice" (**Xia**, 2010).

A recent study of non-Western and non-English-specking academic authors found that 45% of the respondence to the survey had published in an open access journal, but that 9% did know what open access was (*Editage Insights*, 2018). Another recent study of contributors to Spanish scholarly journals demonstrated a general understanding of and support for open access, though there was skepticism about open peer review and altmetrics (**Segado-Boj**; **Martín-Quevedo**; **Prieto-Gutiérrez**, 2018).

I think it is fair to use an understanding of and willingness to publish in OA venues as a proxy for faculty willingness to support the changing library mission. If this is the case, we can see a slow growth of understanding and probably support, but it is far from universal. Changing faculty and attitudes and practice will be a long and slow process.

Adam Murray and Ashley Ireland surveyed U.S. provosts and chief academic officers to establish their view on the value libraries provide and the best ways for libraries to communicate that value. They concluded that,

> "Overall, provosts or chief academic officers indicate that academic libraries are involved with important campus initiatives such as student retention and success, faculty research productivity, and accreditation. They also tend to note that the biggest barrier for academic libraries to be involved with campus initiatives is that the campus overall doesn't recognize that potential."

They recommended a communications approach of library communicating,

> "evidence with provosts and other institutional administrators through means that are timely and relevant, such as a formal annual report or a dedicated budget meeting. Library administrators must strategically use endorsements from deans, directors, or other administrators, as well as user satisfaction data" (**Murray**; **Ireland**, 2018, p. 359).

Mark Robertson conducted a smaller scale study of Canadian provosts and found that they,

> "perceive libraries making significant contributions to research and student learning, particularly through the provision of access to information and the evolving role of library as place respectively. Other areas of library expertise, such as scholarly communication, appear somewhat less familiar to provosts" (**Robertson**, 2015, p. 490).

It seems that both faculty and administrators recognize that the library makes contributions to the institutions and their priorities, but that the library which may still rhetorically the "heart of the university," is really not a top priority. As noted above, faculty lead busy lives and administrators have lots of difficult problems to contend with. The important thing is that if you are going to sell the change you have to be making the change. If you have an instruction program that is well structured and you are measuring its impacts, you don't have to connect all of the dots from a library instruction session to a student's GPA (grade point average) to demonstrate the library's contribution to student success. But you will need some evidence to make the case. If you are altering your collections strategies to on-demand purchasing, the systems

you use to acquire books and journals need to work well and be fast. You can measure the citation impact of faculty authors who deposit in your institutional repository. It is also important to state the obvious, that the current system of scholarly communication, with exploitive commercial publishers extracting monopoly profits, is not sustainable. The best way to sell the change is to do it and show the impact. It will take time and there will be bumps in the road, but in the end, there are really no other choices.

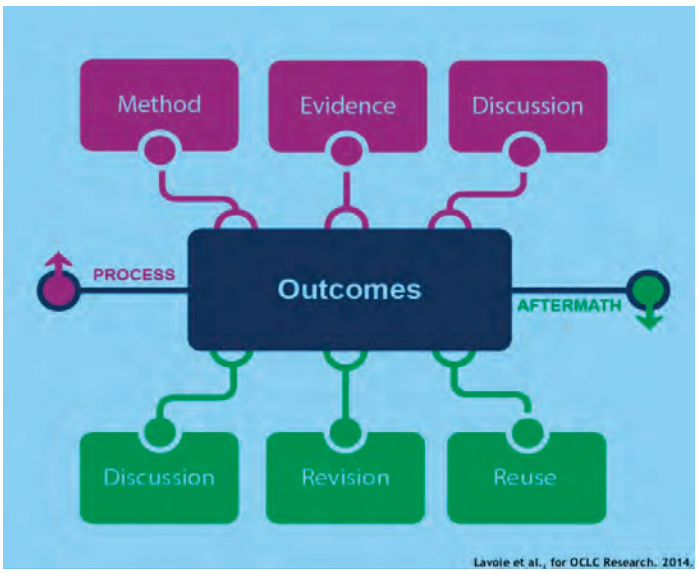


Figure 5. Evolving scholarly record (**Lavoie** *et al.*, 2014, p. 8)
https://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-evolving-scholarly-record-2014-a4.pdf

4. The most important thing to do next: Support the development and sustainability of network-level tools and services

In the time since my book was published, the thing that has become clearest to me is that while there are many challenges that individual libraries face, the biggest and most important challenge is one that we all face together as a community of libraries. To go back to Jean-Claude Guédon' description of the open scholarly commons and the what it will take to get to it. He sets out a second function after describing the library's role in curating content created on campus. He says,

> "the libraries, with their sister institutions, are involved in the task of ensuring a vibrant knowledge-nurturing life for their documents: they will circulate, be discoverable, be interoperable, be evaluated, etc., with the second function, the libraries connect to bring the knowledge infrastructure that we all really need" (**Guédon**, 2017).

Accomplishing this will be difficult because building infrastructure is difficult, and because there are large commercial entities that are attempting to capture the infrastructure of scholarly communications as they did scientific journals and use this position to extract monopoly rents from the academy, and finally because collective action is difficult.

We will begin our consideration of the infrastructure defining what outcome would be in the best interest of libraries, and more importantly science and scholarship. We will then look at the three problems listed above.

4.1. Integrated community controlled open source infrastructure

The infrastructure we need must accommodate scholarship in the digital world. In a 2014 *OCLC* Brian Lavoie and his colleagues attempted to document the evolution of the scholarly record (**Lavoie** *et al.*, 2014, p. 8). They produced the diagram shown in Figure 5.

The important point that this diagram makes is that the scholarly records, now firmly situated in the digital world, is much more complex than it was in the print-based past. The blue section in the center of the diagram represents what had been, in the paper world, the scholarly record, that part of the scholarly practice that was captured and preserved. It encompassed mostly books and articles. In the digital world it is now possible, and, Lavoie and his colleagues argue, desirable, to capture aspects of the scholarly process that take place before the traditional outcomes are produced and aspects that come after the traditional outcomes. What Lavoie and his colleagues argues is that the whole

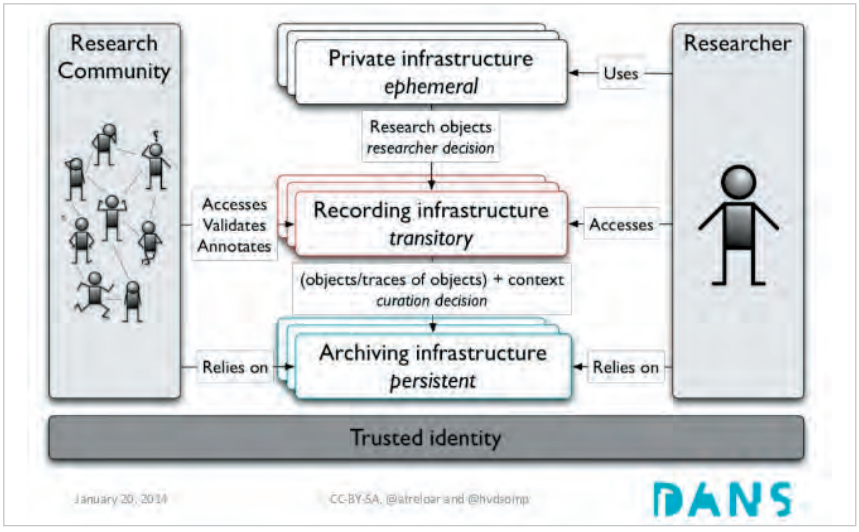


Figure 6. Infrastructure that assures reliable future access
http://www.ncdd.nl/on-line-scholarly-communications-vd-sompel-and-treloar-sketch-the-future-playing-field-of-digital-archives

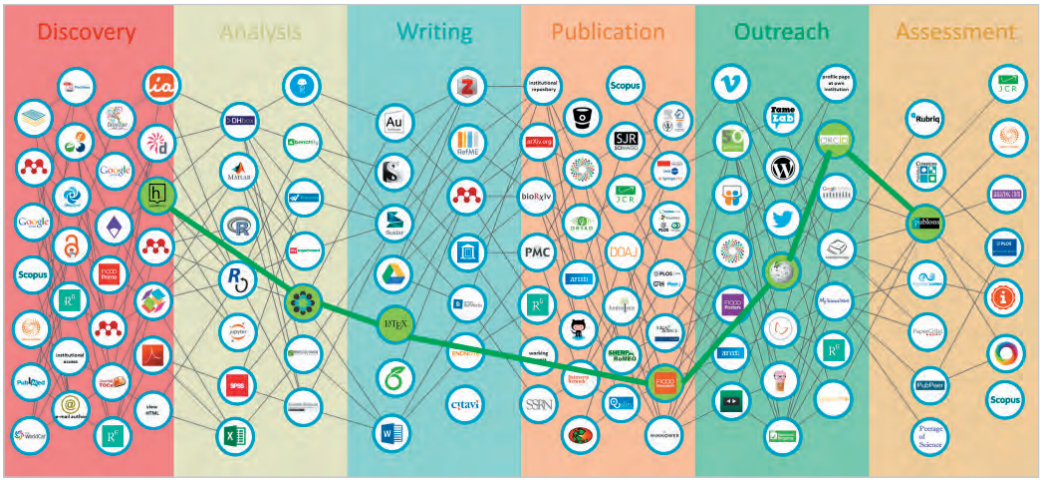
of the scholarly workflow needs to be considered part of the scholarly record.

Herbert Van de Sompel looks at the digital scholarly record and makes the important distinction between three different types of infrastructure (**Van-de-Sompel**, 2014):

- Private infrastructure. This is used by the researcher to create research objects. This infrastructure and the objects created by it are ephemeral.
- Recording Infrastructure. The researcher makes the decision to move the research object into the recording infrastructure. This makes the work public and as such discoverable and accessible and through review validates the work. It may also put the work in the context of other work. This infrastructure would include roles traditionally played by publishers and libraries.
- Archiving infrastructure. Works enter the archiving infrastructure as the result of curation decisions made by publishers and/or libraries. This is the infrastructure that assures reliable future access.

Figure 6 shows a diagram illustrating this structure.

The infrastructure needed to support this more complex scholarly/scientific workflow can be shown as a linear or a circular process. Bianca Kramer and Jeroen Bosman of the *Utrecht University Library* do both from the perspective of the researcher (**Kramer**; **Bosman**, 2018). Both are shown in Figures 7 and 8.



Figures 7 and 8. Infrastructure needed to support a more complex scholarly/scientific workflow (**Kramer**; **Bosman**, 2018)
https://101innovations.wordpress.com/workflows/#MVP
https://figshare.com/articles/Views_on_innovation_The_scholarly_communication_landscape_and_changing_research_workflows_/3185293/1

A somewhat different approach is taken by Alejandro Posada and George Chen. Note that Posada and Chen include entities as well as functions, for example, open repositories and libraries, and maybe journal editorial offices (**Posada**; **Chen**, 2018, p. 6) (Figure 9).

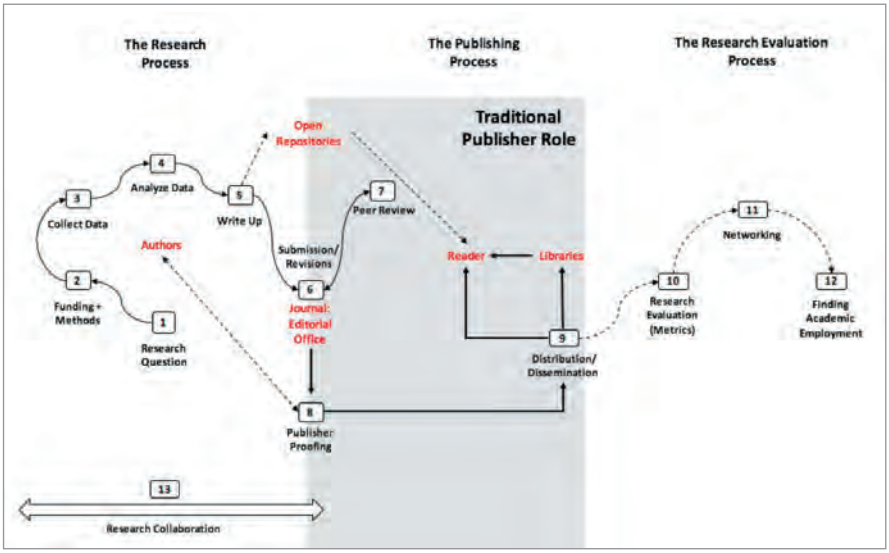


Figure 9. Academic knowledge production process (**Posada**; **Chen**, 2018, p. 6)
http://knowledgegap.org/index.php/sub-projects/rent-seeking-and-financialization-of-the-academic-publishing-industry/preliminary-findings

I would suggest that the required infrastructure, at least that part that is in the domain of academic libraries should include the functions that encompass Van de Sompel's recording and preservation infrastructure. The tools that Van de Sompel defines as private infrastructure, those use to create knowledge and that support what Lavoie and his colleagues refer to as process, are important, but are, I would argue, in the domain of academic departments, research groups and labs. It should be noted that the most significant functions that are part of the scholarly record that we have eliminated can be managed by tools that make up what Van de Sompel terms the recording infrastructure. For example, tools that facilitate discussion after publication can be used just as well prior to publication.

If we use the maps of the scholarly work flow provided by Kramer and Bosman and by Posada and Chen we get the more detail to the infrastructure tools list below. I have provided a single example in each case, for most functions there are multiple systems.

1. Research outcomes
   1.1. Repositories
      - Institutional (*DSpace*)
      - Disciplinary/Pre-print servers (*ArXiv*)
      - Content repositories (*Internet Archive*)
   1.2. Formal publishing
      - Submission and peer review systems (*xPub*)
      - Journal publishing systems (*Open Journal Systems*)
      - Book publishing systems (*Fulcrum*)
   1.3. Systems for sharing non-published items (*Figshare*)

2. Reuse and response
   2.1. Annotation systems (*Hypothes.is*)
   2.2. Discovery
      - Search systems (*DOAJ*)
      - Authority control systems (*Orcid*)
      - Personal information management (*Zotero*)
   2.3. Evaluation and assessment
      - Research profiles (*Vivo*)
      - Ranking systems (*Impact Story*)

3. Preservation systems (*Lockss*)

Having defined the functions that are necessary it is important to look at the characteristics we want in the tools and services and in the organizations that build and maintain then. Geoffrey Bilder, Jennifer Lin, and Cameron Neylon have laid out principles for defining openness of scholarly infrastructure (**Bilder**; **Lin**; **Neylon**, 2015). These principle cover governance, sustainability, and insurance and are summarized below.

- *Governance*. If an infrastructure is successful and becomes critical to the community, we need to ensure it is not co-opted by particular interest groups. Similarly, we need to ensure that any organization does not confuse serving itself with serving its stakeholders. It should be **stakeholder governed**, have n**on-discriminatory membership**, and t**ransparent operations**.
- *Sustainability*. Financial sustainability is a key element of creating trust. To be a financially stable the organization should use time limited funds in the time allotted, have a goal to generate a surplus so as to have a contingency fund

to support at least a year of operations, revenue should be mission driven and based on services, not data.

- *Insurance*. Even with the best possible governance structures, critical infrastructure can still be co-opted by a subset of stakeholders or simply drift away from the needs of the community. Long term trust requires the community to believe it retains control. Systems should be open source. Constraint with privacy laws data should be open. No patent writes should be asserted.

Bilder, Lin, and Neylon's insurance principles are similar to those proposed by Paul Peters. Peters asks how can for-profit organizations be trusted to appropriately contribute to open science. He argues that they can do so if they follow the following principles of openness:

- Open source
- Open data
- Open integrations
- Open contracts

	Community Control	Commercial
Open source and Open systems	*DSpace*	*Ubiquity*
Proprietary code and closed systems	*Contentdm*	*Bepress*

Figure 10. Openness options considering community control versus commercial, and open versus closed

The options we face can be summarized in the Figure 10. There are two dimensions: 1) community control versus commercial, and 2) open versus closed. I have put examples of repository systems in each cell.

Community control/open is clearly best. Commercial/open and community control/proprietary both have disadvantages, but are better than the final option commercial/closed. The goal of the community should be to move, to the extent possible to the upper left cell for all of the systems need to support the scholarly workflows define above.

Finally, the infrastructure needs to be integrated. Today there are many open community-controlled projects, but there is not an integrated system of tools that work well together. The work of integrating these many projects is important and there have been beginnings, especially the *Joint roadmap for open science tools* (*Jrost*): *https://jrost.org*

4.2. The nature of infrastructure and why it is so difficult

Brett M. Frishmann, in his book *Infrastructure: The social value of shared resources* defines infrastructure as follows,

> "Infrastructure resources are intermediate capital resources that serve as critical foundations for productive behavior within economic and social systems. Infrastructure resources effectively structure in-system behavior at the micro-level by providing and shaping the available opportunities of many actors."

Importantly infrastructure facilitates behaviors that affect third parties. As he puts it,

> "the *social returns* on infrastructure investment and use may exceed the *private returns* because society realizes benefits above and beyond those realized by producers and user. [italics in the original]" (**Frischmann**, 2012, pp. 11-21).

Frischmann makes the obvious but important observation that, both governments and markets struggle to adequately supply the public with the infrastructures that it needs. There are good reasons for this. First, infrastructure is often expensive to build and maintain, and raising the necessary capital can be difficult. It is also the case that the nature of the benefits that infrastructure makes possible are not always easily identified and often the impact on any one individual is small. This makes the social value of infrastructure difficult to fully appreciated. Without this concrete sense of the return, investment is often difficult to justify.

Governments are often involved in the provision of infrastructure because governments are able to assess taxes and use them to fund infrastructure without being dependent on the reliably expected returns that the market requires. If is often not this simple, but taxing for the public good is often the best place to start in funding infrastructure.

Frischmann identifies three criteria that infrastructure must meet from a system perspective.

- The resources may be consumed nonrivalrously for some appreciable range of demand.
- Social demand for the resource is driven primarily by downstream productive activity that requires the resource as an input.
- The resource may be used as an input into a wide range of goods and services, which may include private goods, public goods, and social goods. (**Frischmann**, 2012, p. xiv).

The first criterion requires nonrivalrous consumption, though not over all levels of demand. Nonrivalrous use means that one person's use does not interfere with the use of another person. Ideas are nonrivalrous, so is air or a road or library, though a library or a road are subject to congestion at high level of use and air is subject to uses by some that degrade its usability for others.

The second criterion makes clear that infrastructure is a means to an end not the end itself. The value of infrastructure is in what it makes possible. Often what it makes possible happens indirectly. This is why the value of infrastructure is not always fully appreciated. Its full value is hard to observe.

Finally, the use of infrastructure can provide benefits across a variety of activities. Infrastructure is a shared means to many ends. Some of those benefits accrue to society at large and some are captured by individuals. Either is acceptable. As Frischmann says,

> "The key insights from this analysis are that infrastructure resources are basic inputs into a wide variety of productive activities and infrastructure users who choose to engage in such activities often produce public and social goods that generate spillovers that benefit society as a whole" (**Frischmann**, 2012, p. xii).

Because infrastructure enables nonrivalrous use of a resource, the widest possible access to the resource with the fewest impediments to use will maximize the value of the infrastructure resource to society. Libraries and most other involved with scholarship espouse this goal, but the structures that have evolved as scholarship has become digital often restrict access rather than encouraging unfettered access, thus limiting the value of the work.

From an economic perspective infrastructure is most often either a natural monopoly or a public good. Except when congestion occurs at high levels of demand the more an infrastructure is used the better, as this creates the most benefit. This is typical done by charging users the marginal cost. For both natural monopolies and public goods, the average costs decrease as output increase and marginal costs are below average costs. Efficient price thus precludes the investors recouping all of the costs. Because infrastructure often requires substantial investments the difference between average and marginal costs can be significant. This is a problem. It is generally solved in one of four ways: government provision; government subsidy; nonprofit or community-based provision, or market provisions that allow providers to charge more than the marginal cost.

So, the reasons that infrastructure in general is difficult are:

- There is often a very high initial cost.
- Many of the benefits of infrastructure accrue to individuals, firms, and to society at large, making the measurement of impact difficult. They often also accrue indirectly and long after the actual use. The value of infrastructure is thus had to assess, and is therefore not always appreciated as it should be. This makes raising the funds required to build and maintain infrastructure difficult.
- The economics of infrastructure, as a natural monopoly or a public good, means that the most benefit is generated when marginal cost is charged for use. Returns from margin cost pricing will not often cover the cost of the initial investment. This generally means some form of tax or philanthropy is required to fund the start-up costs for infrastructure and often its maintenance.

All of this makes infrastructure in general difficult. All of these problems apply as we try to create the infrastructure we need to support the open scholarly commons.

4.3. The commercial competition

We already have a clear understand that the scholarly journal market is an oligopoly. Vincent Larivière, Stefanie Haustein, and Philippe Mongeon document the extent of the consolidation in scholarly journal publishing (**Larivière**; **Haustein**; **Mongeon**, 2015). They look at the Natural and Medical Sciences (NMS) and Social Sciences and Humanities (SSH). In NMS the top publishers were *Reed-Elsevier*, *Wiley-Blackwell*, *Springer*, *Taylor & Francis* and the *American Chemical Society*. In SSH the top five were *Reed-Elsevier*, *Wiley-Blackwell*, *Springer*, *Taylor & Francis*, and *Sage*. In NMS the top five publishers accounted for a little more than 20% of the papers in 1973. Their share increased to 30% in 1996; passed 50% in 2006, and was 53% in 2013. *Reed-Elsevier*, *Wiley-Blackwell*, and *Springer* accounted for 47% of all NMS papers in 2013. In SSH the top five publishers accounted for about 10% of the papers between 1973 and 1990, but by 2013 their share of papers had increased to 51%. This consolidation fortified the "big deal" strategy of the publishers and made it more difficult for libraries to withdraw from them. This in turn has led to excessive price increases and profits for these publishers that are widely reported as being between 35% and 40%. This is well above profit margins in most other industries.

As Jonathan Tennant and Björn Brem put it in the conclusion of the complaint against *RELX* (*Elsevier*'s parent company) to the *EU Competition Authority*:

> "We believe that the present scholarly communication market is clearly not functioning well due to a number of related reasons. High subscription charges still reign, publishers still offer limited access to research to the wider public, many continue to reap excessively high profits, and many financial elements of the process is shrouded in secrecy. Much of these peculiarities exist due to a combination of content aggregation and concentration by a few large players, chief among which is *Elsevier*, that each individual research article acts as a mini-monopoly meaning that consumers have no buyer power over content, and the use of non-disclosure clauses over licensing agreements which restricts any sort of competitive consumer power" (**Tennant**; **Brems**, 2018).

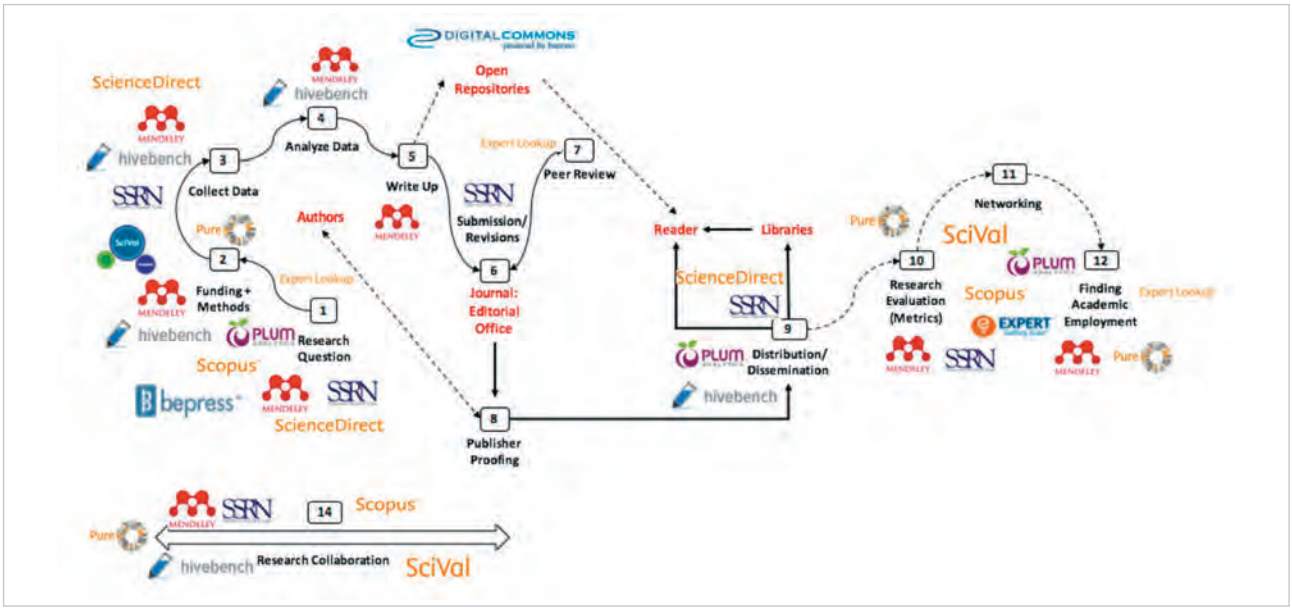


Figure 11. *Elsevier* acquisitions on a version of the scholarly workflow (**Posada**; **Chen**, 2018, p. 6)

What is especially concerning is that these powerful players are now looking to assert control of academic workflows and educational processes beyond journal content. The purchase of *Bepress* by *Elsevier* in the summer of 2017 was a stark reminder to the library community that its infrastructure was vulnerable to acquisition. This should not have been a surprise. As Posada and Chen have documented, *Elsevier* has pursued a clear strategy of acquiring infrastructure components across the scholarly workflow. The *Elsevier* strategy is clearly seen in Figure 11 that overlays *Elsevier* acquisitions on a version of the scholarly workflow (**Posada**; **Chen**, 2018, p. 6).

Posada and Chen summarize the situation we find ourselves in:

> "As major academic publishers' have redirected their business strategies to open access and alternative paying structures, it could be argued that this represents a move towards more democratic access to knowledge. However, this paper problematizes this claim by documenting and examining what has been a simultaneous redirection of big publishers' business strategy towards the acquisition and integration of scholarly infrastructure, the tools and services that underpin the scholarly research life cycle, many of which are geared towards data analytics. We argue that moves toward openness and increased control of scholarly infrastructure are simultaneous processes of rent-seeking which could further entrench publishers' power and exacerbate the vulnerability of already marginalized researchers and institutions" (**Posada**; **Chen**, 2018, p. 1).

They go on to argues that *Elsevier* is pursuing the research work flow beginning to end and that *Wiley* is attempting to lockdown online/digital education. In both cases the aim is not just to control the infrastructure, but to harvest and have monopoly control of the data that the infrastructure generates. Should *Elsevier* or *Wiley* or one of the other large companies in this space —*Springer/Nature*, *Clarivate Analytics*, *Academic Analytics*, *Pearson*, *McGraw-Hill*, or *Cengage*- come to control a large segment of either the research workflow or online/digital education, they would know more about the university's business than the university. Think for a minute about the research workflow. If a single business controls a large enough segment of this workflow across a large enough number of institutions, it has a unique, privately held, understanding of where science is going, what discoveries are being made, and where money might be made as a result.

Tim Wu examines information industries in his book *The master switch* (**Wu**, 2010). He argues that information industries, as is the case with scholarly communication, evolve toward oligopoly and monopoly. That once they gain control of a market, they lock in the revenue streams, resist change, and build barriers to new entrants in the market place. Wu documents this evolution with the telegraph, radio, telephones, television, and the Internet. He concludes that only regulatory action by governments alter the situation. The action can be either regulatory control, as was the case with *AT&T* in the United States in the first half of the 20th century, or breaking up the monopoly enterprise, as was the case with *AT&T* in the 1980s.

Despite some modest attempts to make publicly funded research openly available, for example the *U.S. National Institutes of Health* public access policy, the large publishers have been effective in lobbying governments on both sides of the Atlantic to stop any alteration of their practices. *Plan S* can be seen as the most robust attempt to date to regulate the commercial publisher's monopolistic behavior. It will be interesting to see how successful *Plan S* will be given the inevitable push back from the publishers and their lobbyists. We can hope that because most to the money that funds the commercial publisher's excess profits comes from governments, that these government will at some point respond to this exploitation, but we will have to wait and see.
https://www.scienceeurope.org/making-open-access-a-reality-by-2020

4.4. The collective action problem

It is not enough to wait for government intervention on our behalf, libraries need to build our own community-controlled infrastructure so that here is an alternative to what the large commercial players provide. As we noted at the beginning, there are forces that are working in our favor, the power of technology is increasing and the cost is declining, the nature of digital content works in our favor, and much of the money in the system passes through our hands. However, to take advantage of these forces we need to solve a very difficult problem. We need to overcome the collective action problem. We need to develop the capacity to act collectively to create and nurture a public good.

In 1965 Mancur Olson published *The logic of collective action* in which he laid out what has become known as the collective action problem (**Olson**, 1965). Olson argued that groups have a difficult time providing themselves with collective goods, even when the provision of such goods would greatly advantage the group. He argues that in some cases small groups will be able to do so, but that large groups can almost never do so without some form of compulsion or the ability to provide individual members of the group some unrelated benefit. This is because the cost of coordinating collective action and, with large groups, a single individual's contribution is small and individuals can generally benefit whether or not they contribute.

As Clay Shirky, in his book *Here comes everybody* (**Shirky**, 2008), and Yochai Benkler, in his book *The wealth of networks* (**Benkler**, 2006), powerfully demonstrate the Internet has radically reduced coordination costs to the extent that, for example, a group can collectively create an encyclopedia. None-the-less, the collective action problem remains as the incentives for members of large groups to contribute remain weak. The inability of academic libraries to make the collective investment required to create the infrastructure we need is a key problem that we need to face and to solve.

John Wenzler examines the collective action problem academic libraries face and summarizes the situation well:

> "Economists and political scientists long have argued that it is extremely difficult to replace the unconscious coordination made possible by market transactions with the conscious coordination required for collective action. Even when everyone involved understands and desires the benefits of cooperation, it is often impossible for them to work together to achieve it... Because everyone benefits from the creation of public good regardless of whether or not they help to produce it, individuals are tempted to become "free riders" who exploit the efforts of others. Even those who have no intention of free-riding for themselves are reluctant to invest in a public good because they worry that their effort will be wasted if too many other people chose to ride for free. In small communities, it often is possible to build bonds of reciprocal trust that allow groups to achieve collective ends, but it is more challenging for larger and more distributed groups to do so" (**Wenzler**, 2017, pp. 184-185).

Wenzler argues that it is this collective action problem that has made it impossible for the academic library community to benefit from digital technologies that should make scholarly communication, as was described above, cheaper and more easily available to all of the people who wish to use it. Because we have not overcome the collective action problem, we face continued steep price increases from for-profit publishers who maintain monopolistic control over much of the scholarly communications system. Wenzler is frank about our prospects:

> "Although it is likely that university libraries could develop a more efficient system of scholarly communication if they were to redeploy their collective subscription budgets, each individual library —when it decides how to spend its own little piece of that huge pie- has little incentive to redirect its own expenditures... Unfortunately, if every librarian waits for every other librarian to make the investments necessary to develop a sustainable system of Gold OA publishing, it may never happen" (**Wenzler**, 2017, p. 192).

Wenzler's critique clearly defines the problem we face, but it is incomplete. Cameron Neylon also uses Olson's work to consider the collective action problem faced by the scholarly community (2017). He notes that Olson argues there are three ways to solve the provisioning problem presented by public goods. They are:

- Mechanisms that require contributions from the whole community. Taxation is the clearest example. For scholarly infrastructure, such mechanisms include indirect cost taken by institutions and top-slicing of funder budgets to provide infrastructure and services.
- Mechanisms that provide additional private good or services to contributors. This approach is generally referred to as a club good. It requires the creation of a good or service that is non-rivalrous but excludable. Neylon cites *Crossref* as an example. Contributors gain the exclusive right to mint DOIs and the broader community gain an open database of scholarly content. For much of scholarly infrastructure creating club goods can be problematic. As Neylon says,

  > "Digital resources are not natively excludable; a technical barrier has to be put in place" (**Neylon**, 2017, p. 2).

  This is both a technical and a philosophical problem.
- Reducing the size of the group to the extent that all of the members can agree that contributions are in their interest. Such oligopolies for scholarly infrastructure are likely to be either groups of large funders or governments. Neylon cites *Europe PubMed Central* as an example. Importantly, Neylon notes that oligopoly among funders can look like taxation to the individual researcher. Thus we see the "academic freedom" response to *Plan S*.

The effectiveness of the various strategies depends on the size of the group attempting to create the public good. Neylon makes the critical point,

> "The difficult truth that Olson articulates is that there is no mechanism that will lead directly to a large community supporting the provision of a large-scale public-good infrastructure. Any successful sustainability model will depend on some mixture of these three approaches for resourcing."

He goes on to state,

> "If our challenge in delivering on the openness and transparency agenda is how to support the conversion of successful medium-scale club-like infrastructures into open systems that provide collective goods, then we need to solve the political and economic problems of transitioning from the club state to a model that successfully provides a mix of these models" (**Neylon**, 2017, p. 3).

To the extent that academic libraries act in isolation, they will inevitably act in their own narrow interest and we will not be able to create the infrastructure that will be required to support the open scholarly commons. This is Wenzler's point, and it is important. Neylon provides a more nuanced view that argues for the need to look at the political economy of scholarly infrastructure and consider governance and community building as well as financial sustainability if we are to successful create the institutional structures required to support scholarly infrastructure, Neylon argues, we may be able to find the necessary resources.

Elinor Ostrom won the Nobel Prize in Economics for her work on commons governance. Her book, *Governing the commons: The evolution of institutions for collective action*, examines situations and institutions where commons are successful (1990). Her work demonstrates that the collective action problem though real, can be overcome given the right circumstances, incentives, and motivations. Her work can provide both hope and guidance.

Ostrom looks at a variety of common pool resources (CPRs) and develops design principles form this study. Her work was with physical systems such as forests, fisheries, or irrigation systems. These systems require mechanisms both provisioning (the work require to create and maintain the systems) and appropriation (managing how the resource is used).

The commons we are endeavoring to create is digital and networked and as we have observed, digital content can be distributed for all practical purposes at no cost. This means that the open scholarly commons do not need to be concerned with appropriation, as use does not diminish the resource. Once provisioning is accomplished, that is once the resource is created and there is a mechanism for ongoing maintenance and enhancement, the commons can function. So, provisioning the open scholarly commons is what we have to accomplish. This simplifies the task, but it is still significant. Ostrom states the challenge:

> "Designing and adopting new institutions to solve CPR problems are difficult tasks, no matter how homogeneous the group, how well informed the members are about the conditions of their CPR, and how deeply ingrained are generalized norms of reciprocity. Given the strong temptation to shirk, free-ride, and generally act opportunistically that usually are present when individuals face CPR problems, overcoming such problems can never be assured. No strong external pressures drive individuals toward positive solutions to such problems" (**Ostrom**, 1990, pp. 210-211).

Ostrom suggests that success in creating a CPR is best when the follow conditions exist:

- Parties share a common judgement that they will be harmed if they do not adopt an alternative arrangement.
- Parties will be affected in similar ways by the proposed arrangement.
- Parties highly value the activities from the common pool resources (CPR), they have low discount rates.
- Parties face relatively low information, transformation, and enforcement costs.
- Parties share norms of reciprocity and trust that can be used as initial social capital.
- The group is relatively small and stable (**Ostrom**, 1990, p. 211).

When we look at the current state of academic libraries and scholarly communication in light of Ostrom's work we see that there are some positives and some negatives.

1. Parties share a common judgement that they will be harmed if they do not adopt an alternative arrangement. This is a big plus. We understand the system's problems and the harm that is being done. We know we should act.

> "Libraries have to reshape themselves so that they are nodes on the global network that will make up the open scholarly commons"

2. Parties will be affected in similar ways by the proposed arrangement. This is a minus. At the present time, there is no shared understanding of what the system we are trying to create should be. There is no roadmap. There is probably general agreement, but agreement on strategies and tactics to construct the system are not widely understood or accepted. Nor do we know what resources individual libraries are now committing to activities that support the creation of the open scholarly commons. Because we don't know what others in our community are doing there are no norms to guide behavior. Creating a general understanding on levels of contribution and developing norms is the aim of the 2.5% commitment initiative, but this work is in its infancy (**Lewis**, 2017c).

3. Parties highly value the activities from the CPR, they have low discount rates. This is a plus. The academic library community's default is to share. Having a low discount rate means that the future value of the resource declines slowly. Libraries by their nature have long time horizons and so our discount rates are general low.

4. Parties face relatively low information, transformation, and enforcement costs. It is hard to know what the information costs will be, but it is likely that they will be relatively low as there are many existing structures to expedite them. The academic library community tends to be overly deliberative, which increases coordination costs.

> "Libraries have to create and sustain the infrastructure that will support the open scholarly commons"

5. Parties share norms of reciprocity and trust that can be used as initial social capital. Probably a plus. In general, academic libraries share norms of reciprocity and trust that can be used as initial social capital. Although trust is highest in small groups and in most cases the level of investment is small so the risks of trust are small. To generate trust across the system, it will be necessary for organizations providing systems and services to be exceptional transparent in the ways proposed by **Bilder**, **Lin**, and **Neylon** (2015) and use open contracts as proposed by **Peters** (2017).

6. The group is relatively small and stable. This is a minus. Academic libraries as a whole are a very large and diverse group, and the system we are attempting to create is global.

When we look at the conditions Ostrom sets for the successful creation of a CPR, our problems are clear. However, it may be that the scale of the threat and the need response will motivate the library community. Ostrom's work suggests the need leverage small scale collaborations and build hierarchical structures on them to support global scale work. It argues for the necessity of transparent governance and finances in order to build trust, and the development of shared values, in my view, especially around appropriate resource contributions to the commons. We also need to make the landscape and the roadmap forward clear as a means of developing a shared understanding of what needs to be accomplished and how it will be done. If we can do these things, Ostrom's work provides a reason for hope.

5. Conclusion

So, there is much work to do. Some of it will be in reshaping our individual libraries so that they are nodes on the global network that will make up the open scholarly commons. I believe this is the easier half of what needs to be done. We will need new skills and expertise, but the work is an extension of what libraries have done for a long time. It will be both challenging and exciting. The second half of our work is to create and sustain the infrastructure that will support the open scholarly commons. This infrastructure will need to be integrated and community controlled. It will require ongoing financial contributions from universities and governments. A sizeable portion of this funding will come the reallocation of library budgets.

> "If we do not build and maintain infrastructures as a public good, the private sector will build it and use it to extort excessive profits fron our institution and the inequities of access to the scholarly record will continue"

If we do not build and maintain this infrastructure as a public good, the private sector will build it and use it to extort excessive profits from our institutions and the inequities of access to the scholarly record will continue. These inequities are of two types. First those who could benefit from the knowledge and cannot afford to acquire it. Second, those, largely from the Global South, who do not have access to the venues that would bring their research to the world.

The challenge is clear. It will not be easy. But I am optimistic. We get to create something new, something that was not possible before. Because nature of the network and digital information on it, we have the tools to create the open scholarly commons where the world's knowledge openly and easily available to everyone. What could be more important?

Note

1. *Association of Research Libraries* (*ARL*)

6. References

ACRL (2016). *Framework for information literacy for higher education*. Association of College & Research Libraries. *http://www.ala.org/acrl/standards/ilframework*

Agrawal, Ajay (2018). "The economics of artificial intelligence". *McKinsey quarterly*. *https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-economics-of-artificial-intelligence*

Albertson, Dan; **Spetka, Kaitlyn**; **Snow, Kristen** (eds.) (2015). *Library and information science education statistical report 2015*. Seattle WA: Association for Library and Information Science Education (Alise). *https://ali.memberclicks.net/assets/documents/statistical_reports/2015/alise_2015_statistical_report.pdf*

Almanac 2018 (2018). "Library spending by sector 2015-16". *The chronicle of higher education*, August 19.
https://www.chronicle.com/article/Library-Spending-by-Sector/244160

Alperin, Juan-Pablo; **Muñoz-Nieves, Carol**; **Schimanski, Lesley**; **Fischman, Gustavo E.**; **Niles, Meredith T.**; **McKiernan, Erin C.** (2018). "How significant are the public dimensions of faculty work in review, promotion, and tenure documents?". *Humanities commons* [preprint].
https://doi.org/10.17613/M6W950N35

Applegate, Rachel (2009). "The library is for studying: Student preferences for study space". *Journal of academic librarianship*, v. 35, n. 4, pp. 341-346.
https://doi.org/10.1016/j.acalib.2009.04.004

Aries Marketing (2016). "Artificial intelligence integration allows publishers a first look at meta bibliometric intelligence". *Aries Systems*, October 17.
https://www.ariessys.com/views-press/press-releases/artificial-intelligence-integration-allows-publishers-first-look-meta-bibliometric-intelligence/

ARL (2010). *Principles to guide vendor/publisher relations in large-scale digitization projects of special collections materials*. Association of Research Libraries.
http://www.arl.org/storage/documents/publications/principles_large_scale_digitization.pdf

Bedi, Param; **Snyder, Jason** (2015). "Making a difference: Moving your organization from transactional to transformational". *Educause review*, March 16.
https://er.educause.edu/articles/2015/3/making-a-difference-moving-your-organization-from-transactional-to-transformational

Benkler, Yochai (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press. ISBN: 978 0 300110562
http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf

Bergstrom, Ted (2018). *Do download reports reliably measure journal usage? Trusting the fox to count your hens?*. UC Santa Barbara: Department of Economics, UCSB. Ted Bergstrom papers.
https://escholarship.org/uc/item/2cd2h7vt

Bergstrom, Ted; **Uhrig, Richard**; **Antelman, Kristin** (2018). *Looking under the counter for overcounted downloads*. UC Santa Barbara: Department of Economics, UCSB. Departmental working papers.
https://escholarship.org/uc/item/0vf2k2p0

Bilder, Geoffrey; **Lin, Jennifer**; **Neylon, Cameron** (2015). "Principles for Open Scholarly Infrastructure". *Science in the open*, February 23.
https://doi.org/10.6084/m9.figshare.1314859

Chang, Hui-Fen (2013). "Ethnic and racial diversity in academic and research libraries: Past, present, and future". In: *Proceedings of the 2013 ACRL Conference*, Indianapolis, IN, April 10-13.
http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/Chang_Ethnic.pdf

Courant, Paul N.; **Nielsen, Matthew "Buzzy"** (2010). "On the cost of keeping a book". In: *CLIR*. *The idea of order: Transforming research collections for 21st Century scholarship*. Washington, DC: Council on Library and Information Resources, pp. 81-105. ISBN: 978 1 932326352
http://www.clir.org/wp-content/uploads/sites/6/pub147.pdf

Dempsey, Lorcan (2016). "Library collections in the life of the user: Two directions". *Liber quarterly*, v. 26, n. 4, pp. 338-359.
http://doi.org/10.18352/lq.10170

Dempsey, Lorcan; **Lavoie, Brian**; **Malpas, Constance** (2013). *Understanding the collective collection: Towards a system-wide perspective on library print collections*. Dublin, OH: OCLC Research. ISBN: 978 1 556534614
https://www.oclc.org/content/dam/research/publications/library/2013/2013-09.pdf

Editage Insights (2018). *Author perspectives on academic publishing: Global survey report 2018*.
https://www.editage.com/files/Editage-Global-Author-Survey.pdf

Frazier, Kenneth (2001). "The librarians' dilemma: Contemplating the costs of the 'big deal'". *D-Lib magazine*, v. 7, n. 3.
http://www.dlib.org/dlib/march01/frazier/03frazier.html

Frischmann, Brett M. (2012). *Infrastructure: The social value of shared resources*. New York, NY: Oxford University Press. ISBN: 978 0 199975501

Fussler, Herman H.; **Simon, Julian** (1961). *Patterns in the use of books in large research libraries*. Chicago, IL: University of Chicago Press. ISBN: 978 0 226275567
https://catalog.hathitrust.org/Record/003908643

Gagnon, Stéphanie (2017). *Journal publishers' big deals: Are they worth it?*. Les bibliothèques. Université de Montréal.
https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18507/Gagnon_Stephanie_2017_article.pdf

Gensler Research (2015). *How do students envision the present and future academic library?* Research project: Libraries are for Studying
https://www.gensler.com/research-insight/gensler-research-institute/students-on-libraries

Gore, Daniel (1976). "Farewell to Alexandria: The theory of the no-growth, high-performance library". In: Gore, Daniel (ed.). *Farewell to Alexandria*. Westport, CT: Greenwood Press. ISBN: 978 0 837185873

Guédon, Jean-Claude (2017). "Open access: Toward the internet of the mind". *Budapest open access initiative*, February 23.
http://www.budapestopenaccessinitiative.org/open-access-toward-the-internet-of-the-mind

Harvard Library (2014). *Harvard's School of Engineering and Applied Sciences recommends open-access deposit for faculty review process*. October 22.
http://officeforscholarlycommunicationharvardlibrary.createsend.com/t/ViewEmailArchive/t/0B74E30D40EFCA3A/C67FD2F38AC4859C

Hirtle, Peter B.; **Kenney, Anne R.**; **Ruttenberg, Judy** (2012). "Digitization of special collections and archives: Legal and contractual issues". *Research library issues*, n. 279, pp. 2-4.
https://publications.arl.org/rli279

Jackson, Heather-Lea; **Hahn, Trudi-Bellardo** (2011). "Serving higher education's highest goals: Assessment of the academic library as place". *College & research libraries*, v. 72, n. 5, pp. 428-442.
https://doi.org/10.5860/crl-123

JISC (2014). *Developing digital literacies*.
https://www.jisc.ac.uk/guides/developing-digital-literacies

Julien, Heidi; **Gross, Melissa**; **Latham, Don** (2018). "Survey of information literacy instructional practices in the U.S. academic libraries". *College & research libraries*, v. 79, n. 2, pp. 179-199.
https://doi.org/10.5860/crl.79.2.179

Kent, Allen (1979). *Use of library materials: The University of Pittsburgh study*. New York, NY: Dekker. ISBN: 978 0 824768072

Kramer, Bianca; **Bosman, Jeroen** (2018). "Mapping the open science ecosystem: Looking at tools and platforms". In: *JROST Workshop*, Berkeley, CA August 27.
http://tinyurl.com/JROSTmap

Kurzweil, Raymond (2010). "Exponential growth of computing for 110 years". *Kurzweil accelerating intelligence*, April 9.
http://www.kurzweilai.net/exponential-growth-of-computing#!prettyPhoto

Larivière, Vincent; **Haustein, Stefanie**; **Mongeon, Philippe** (2015). "The oligopoly of academic publishers in the digital era". *PLoS one*, v. 10, n. 6, e0127502.
https://doi.org/10.1371/journal.pone.0127502

Lavoie, Brian F.; **Childress, Eric**; **Erway, Ricky**; **Faniel, Ixchel**; **Malpas, Constance**; **Schaffner, Jennifer**; **Van-der-Werf, Titia** (2014). *The evolving scholarly record*. Dublin, Ohio: OCLC Research. ISBN: 978 1 556534768
https://doi.org/10.25333/C3763V

Lavoie, Brian F.; **Schonfeld, Roger C.** (2006). "Books without boundaries: A brief tour of the system-wide print book collection". *Journal of electronic publishing*, v. 9, n. 2.
https://doi.org/10.3998/3336451.0009.208

Levenson, Helen N. (2015). "Michigan shared print initiative and GreenGlass for groups for data analysis in developing a collaborative collective collection". *Journal of interlibrary loan, document delivery & electronic reserve*, v. 25, n. 3-5, pp. 89-105.
https://doi.org/10.1080/1072303X.2016.1254701

Lewis, David W. (2007). "A strategy for academic libraries in the first quarter of the 21st Century". *College & research libraries*, v. 68, n. 5, pp. 418-434.
https://doi.org/10.5860/crl.68.5.418

Translated into Catalan as: "Una estratègia per a les biblioteques acadèmiques del primer quart del segle XXI". Traduccions del *Consorci de Biblioteques Universitàries de Catalunya*, n. 39, Setembre 2008.
https://www.recercat.cat/handle/2072/10361

Lewis, David W. (2010). "The user-driven purchase give away library". *Educause review*, v. 45, n. 5, pp. 10-11.
https://er.educause.edu/articles/2010/10/the-userdriven-purchase-giveaway-library

Lewis, David W. (2012). "The inevitability of open access". *College & research libraries*, v. 73, n. 5, pp. 493-506.
https://doi.org/10.5860/crl-299

Lewis, David W. (2013). "From stacks to the Web: The transformation of academic library collecting". *College & research libraries*, v. 74, n. 2, pp. 159-176.
https://doi.org/10.5860/crl-309

Translated into Catalan as: "Del prestatge a la web: la transformació de la formació de les col·leccions a les biblioteques universitàries". Traduccions del *Consorci de Biblioteques Universitàries de Catalunya*, n. 53, October 2013.
http://www.recercat.cat/handle/2072/219163

Lewis, David W. (2016). *Reimagining the academic library*. Lanham, MD: Rowman & Littlefield. ISBN: 978 1 442238589

Lewis, David W. (2017a). "Library as place". In: Gilman, Todd (ed.). *Academic librarianship today*. Lanham, MD: Rowman & Littlefield, pp. 161-170. ISBN: 978 1 442278752

Lewis, David W. (2017b). "Measures of change in academic library behavior". *Library leadership & management*, v. 32, n. 1.
https://journals.tdl.org/llm/index.php/llm/issue/view/382

Lewis, David W. (2017c). *The 2.5% commitment*.
https://doi.org/10.7912/C2JD29

Lewis, David W.; **Orr, Kindra** (2018). "The age demographics of librarians and the organizational challenge facing academic libraries". *Library leadership & management*, v. 32, n. 3.
https://journals.tdl.org/llm/index.php/llm/article/view/7282

Lynch, Clifford (2003). "Institutional repositories: Essential infrastructure for scholarship in the digital age". *ARL: A bimonthly report*, n. 226.
http://old.arl.org/resources/pubs/br/br226/br226ir~print.shtml

Lynch, Clifford (2017). "Updating the agenda for academic libraries and scholarly communications". *College & research libraries*, v. 78, n. 2.
https://doi.org/10.5860/crl.78.2.16577

Maron, Nancy L.; **Pickle, Sarah** (2013). *Searching for sustainability: Strategies from eight digitized special collections*. Ithaka S+R; Association of Research Libraries.
http://sr.ithaka.org/wp-content/uploads/2015/08/SR_IMLS_20140129.pdf

McAfee, Andrew; **Brynjolfsson, Erik** (2017). *Machine, platform, crowd: Harnessing our digital future*. New York, NY: W. W. Norton & Company. ISBN: 978 0 393254297

Murray, Adam; **Ireland, Ashley** (2018). "Provosts' perceptions of academic library value and preferences for communication: A national study". *College & research libraries*, v. 79, n. 3, pp. 336-365.
https://doi.org/10.5860/crl.79.3.336

Neylon, Cameron (2017). "Sustaining scholarly infrastructures through collective action: The lessons that Olson can teach us". *KULA: Knowledge creation, dissemination, and preservation studies*, v. 1, n. 1, p. 3.
https://doi.org/10.5334/kula.7

Odell, Jere (2018). *Iupui open access policy: Annual report for 2017*. *Iupui University Library*.
https://scholarworks.iupui.edu/handle/1805/16208

Odell, Jere; **Coates, Heather**; **Palmer, Kristi** (2016). "Rewarding open access scholarship in promotion and tenure: Driving institutional change". *College & research library news*, v. 77, n. 7, pp. 322-325.
https://doi.org/10.5860/crln.77.7.9518

O'Gara, Genya; **Osterman, Anne** (2015). "Curating collective collections— collaborating to build: Using consortial collection analysis to inform collection development". *Against the grain*, v. 27, n. 3, pp. 67-69.
https://doi.org/10.7771/2380-176X.7113

Oldenburg, Ray (1989). *The great good place: Cafés, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day*. New York, NY: Paragon House. ISBN: 978 1 557781109

Olson, Mancur (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press. ISBN: 978 0 674537514

Ostrom, Elinor (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press. ISBN: 978 0 5214005997

Payne, Lizanne (2007). *Library storage facilities and the future of print collections in North America*. Dublin, OH: OCLC Programs & research. ISBN: 978 1 556533716
https://www.oclc.org/content/dam/research/publications/library/2007/2007-01.pdf

Peters, Paul (2017). "A radically open approach to developing infrastructure for open science". *Hindawi*, October 23.
https://about.hindawi.com/blog/a-radically-open-approach-to-developing-infrastructure-for-open-science

Piwowar, Heather; **Priem, Jason**; **Larivière, Vincent**; **Alperin, Juan-Pablo**; **Matthias, Lisa**; **Norlander, Bree**; **Farley, Ashley**; **West, Jevin**; **Haustein, Stefanie** (2018). "The state of OA: A large-scale analysis of the prevalence and impact of open access articles". *PeerJ* 6:e4375.
https://doi.org/10.7717/peerj.4375

Posada, Alejandro; **Chen, George** (2018). "Inequality in knowledge production: The integration of academic infrastructure by big publishers". In: *Elpub 2018*, Toronto, Canada.
https://doi.org/10.4000/proceedings.elpub.2018.30

Regalado, Mariana; **Smale, Maura A.** (2015). "'I am more productive in the library because it's quite': Commuter students in the college library". *College & research libraries*, v. 76, n. 7, pp. 899-913.
https://doi.org/10.5860/crl.76.7.899

Rentier, Bernard (2015). "Institutional repositories: It's a matter of sticks and carrots". *Research Europe*, September 10.
https://orbi.uliege.be/handle/2268/185861

Rieger, Oya Y. (2018). *The state of digital preservation in 2018: A snapshot of challenges and gaps*. New York, NY: Ithaka S+R.
https://doi.org/10.18665/sr.310626

Robertson, Mark, (2015). "Perceptions of Canadian provosts on the institutional role of academic libraries". *College & research libraries*, v. 76, n. 4, pp. 490-511.
https://doi.org/10.5860/crl.76.4.490

Schonfeld, Roger (2018). "The new Clockss succession plan: An interview with executive director Craig Van Dyck". *The scholarly kitchen*, November 6.
https://scholarlykitchen.sspnet.org/2018/11/06/interview-clockss-craig-van-dyck/?utm_source=dlvr.it&utm_medium=twitter

Schonfeld, Roger; **Sweeney, Liam** (2017). *Inclusion, diversity, and equity: Members of the Association of Research Libraries: Employee demographics and director perspectives*. New York, NY: Andrew W. Mellon Foundation and Ithaka S+R.
https://doi.org/10.18665/sr.304524

Segado-Boj, Francisco; **Martín-Quevedo, Juan**; **Prieto-Gutiérrez, Juan-José** (2018). "Attitudes toward open access, open peer review, and altmetrics among contributors to Spanish scholarly journals". *Journal of scholarly publishing*, v. 50, n. 1, pp. 48-70.
https://doi.org/10.3138/jsp.50.1.08

Shirky, Clay (2008). *Here comes everybody: The power of organizing without organizations*. New York, NY: Penguin Press. ISBN: 978 0 713999891

Shirky, Clay (2009). "Newspapers and thinking the unthinkable". *Clay Shirky*.
http://www.shirky.com/weblog/2009/03/newspapers-and-thinking-the-unthinkable

Sparc (sf). "Big deal cancellation tacking". *Sparc.*
https://sparcopen.org/our-work/big-deal-cancellation-tracking

Stemmer, John K.; **Mahan, David M.** (2016). "Investigating the relationship of library usage to student outcomes". *College & research libraries*, v. 77, n. 3, pp. 359-375.
https://doi.org/10.5860/crl.77.3.359

Tennant, Jonathan; **Brembs, Björn** (2018). *RELX referral to the EU competition authority*, October 26.
https://zenodo.org/record/1472045

Tillman, Ruth-Kitchin (2017). "Where are we now? Survey on rates of faculty self-deposit in institutional repositories". *Journal of librarianship and scholarly communication*, v. 5, n. 1, p.eP2203.
https://doi.org/10.7710/2162-3309.2203

University of California Libraries (2018). *Pathways to open access*.
https://libraries.universityofcalifornia.edu/groups/files/about/docs/UC-Libraries-Pathways%20to%20OA-Report.pdf

Van-de-Sompel, Herbert (2014). "On-line scholarly communications: vd Sompel and Treloar sketch the future playing field of digital archives". *NCDD blog*, January 23.
http://www.ncdd.nl/on-line-scholarly-communications-vd-sompel-and-treloar-sketch-the-future-playing-field-of-digital-archives

Wenzler, John (2017). "Scholarly communication and the dilemma of collective action: Why academic journals cost too much". *College & research libraries*, v. 78, n. 2, pp. 183-200.
https://doi.org/10.5860/crl.78.2.16581

Wetterstrand, Kris A. (2018). *DNA sequencing costs 2017: Data from the Nhgri Genome Sequencing Program (GSP).* National Human Genome Research Institute.
https://www.genome.gov/27541954/dna-sequencing-costs-data

Wolff-Eisenberg, Christine; **Rod, Alisa B.**; **Schonfeld, Roger C.** (2016). *Ithaka S+R US faculty survey 2015*. New York, NY: Ithaka S+R.
https://doi.org/10.18665/sr.277685

Wu, Tim (2010). *The master switch: The rise and fall of information empires*. New York, NY: Vintage Books. ISBN: 978 0 307390998

Xia, Jingfeng (2010). "A longitudinal study of scholars attitudes and behaviors toward open-access journal publishing". *Journal of the American Society for Information Science & Technology, v.* 61, n. 3, pp. 615-624.
https://doi.org/10.1002/asi.21283

Yang, Liu; **Vembu, Shankar**; **Adawi, Amr**; **Shai, Ofer** (2016). *Enabling editors through machine learning*. Meta. Bibliometric intelligence.
https://medium.com/@meta_6493/enabling-editors-through-machine-learning-81b528b496ce

Adopción de redes sociales por revistas científicas de ciencias sociales

Adoption of social media by social sciences scientific journals

Carlos Arcila-Calderón; Mabel Calderín-Cruz; Patricia Sánchez-Holgado

Cómo citar este artículo:

Arcila-Calderón, Carlos; **Calderín-Cruz, Mabel**; **Sánchez-Holgado, Patricia** (2019). "Adopción de redes sociales por revistas científicas de ciencias sociales". *El profesional de la información*, v. 28, n. 1, e280105.

https://doi.org//10.3145/epi.2019.ene.05

Artículo recibido el 11-10-2018
Aceptación definitiva: 14-12-2018

Carlos Arcila-Calderón
http://orcid.org/0000-0002-2636-2849

Universidad de Salamanca
Facultad de Ciencias Sociales
Campus Miguel de Unamuno, Edificio FES
Av. Francisco Tomás y Valiente, s/n.
37071 Salamanca, España
carcila@usal.es

Mabel Calderín-Cruz
http://orcid.org/0000-0002-7750-4608

Centro Cultural Padre Carlos Guillermo Plaza
Universidad Católica Andrés Bello.
Moltabán, Caracas, Venezuela
mcalderi@ucab.edu.ve

Patricia Sánchez-Holgado ✉
http://orcid.org/0000-0002-6253-7087

Universidad de Salamanca
Facultad de Ciencias Sociales
Campus Miguel de Unamuno, Edificio FES
Av. Francisco Tomás y Valiente, s/n.
37071 Salamanca, España
patriciasanc@usal.es

Resumen

La expansión de las redes sociales obliga a la comunicación de la ciencia y a las revistas científicas a adaptarse al nuevo entorno. Estudios anteriores y modelos teóricos como Utaut y Utaut2, han probado que existe una relación entre la percepción de desempeño futuro de la tecnología (Expectativa de rendimiento), la dificultad de uso (Expectativa de esfuerzo) y la Influencia social a la que el individuo está sometido, sobre el uso real de dicha tecnología, pero aún existe poca investigación acerca de los mecanismos que causan dicho efecto. Partiendo de Utaut, la hipótesis planteada es que existe un efecto condicional indirecto de la Expectativa de rendimiento, la Expectativa de esfuerzo y la Influencia social, sobre el uso real de redes sociales en las revistas científicas, donde dicho efecto es mediado por la intención de uso y moderado por el género y la edad. Se realizó una encuesta a 300 editores de ciencias sociales del ranking *Google Scholar*. Confirmando parcialmente el modelo Utaut, los resultados de la mediación moderada evidencian que la Expectativa de rendimiento y la Influencia social en los editores inciden en el uso real de *Facebook* y *Twitter*, a través de la intención de uso, pero no sucede así en el caso de *ResearchGate*, *Academia* y *LinkedIn*. No hay evidencia clara de que la edad y el género moderen los predictores estudiados.

Financiación

Centro Cultural Padre Carlos Guillermo Plaza SJ de la *Universidad Católica Andrés Bello*, Venezuela.

Agradecimiento

A Andrea Urbina por su participación en la recogida de datos.

Palabras clave

Adopción de tecnologías; Uso de tecnologías; *User acceptance and use of technology*; Utaut; Redes sociales; Medios sociales; Editores; Revistas científicas; Ciencias sociales; *Twitter*; *Facebook*; *LinkedIn*; *ResearchGate*; *Academia.edu*.

Abstract

The expansion of social media forces the communication of science and scientific journals to adapt to this new environment. Previous studies and theoretical models such as Utaut and Utaut2 have proven that there is a relationship between the perception of future performance of technology (performance expectancy), the difficulty of use (effort expectancy) and the social influence to which the individual is subject, on the real use of technology, but there is still scarce research about the mechanisms that cause this effect. Starting with Utaut, the hypothesis is that there is an indirect conditional effect of the performance expectancy, the effort expectancy and social influence, over the actual use of social media in scientific journals, where the effect is driven through the intention of use and the size of the effect is moderated by gender and age. A survey was carried out on 300 editors of the top 20 in Social Sciences of the *Google Scholar* ranking. Confirming partially Utaut, the results of moderated mediation show that performance expectancy and social influence of editors affect the use of *Facebook* and *Twitter*, through the intention to use, but it is not the case of *ResearchGate*, *Academia* and *LinkedIn*. There is not enough evidence about the moderation of age and gender.

Keywords

Adoption of technologies; Use of technologies; User acceptance and use of technology; Utaut; Social media; Editors; Scientific journals; Social Sciences; *Twitter*, *Facebook*; *LinkedIn*; *ResearchGate*; *Academia.edu*.

1. Introducción

El uso de las nuevas tecnologías de la información y la comunicación (TIC) ha cambiado las prácticas de comunicación científica (**Gibbons** *et al.*, 1994) y las revistas científicas dentro del paradigma de la llamada ciencia 2.0 (**Waldrop**, 2008), o ciencia abierta (**Neylon**; **Wu**, 2008; **García-Peñalvo**, 2017).

De hecho, la ciencia abierta ha modificado la forma en que investigadores, editores y revistas hacen su trabajo en ciencias sociales (**Borgman**, 2007; **Dutton**, 2010), incluso aunque la adopción de las TIC pueda parecer lenta, comparada con otros campos como la Física (**Gentil-Beccot** *et al.*, 2009).

Las revistas científicas son consideradas un instrumento principal en la comunicación de la ciencia (**Castillo-Esparcia**, 2012). En este contexto, aunque sigue habiendo resistencia al movimiento de acceso abierto (**Schiermeier**, 2017), algunos editores están dispuestos a adoptar redes sociales como medio de compartir la ciencia y sus resultados entre investigadores y públicos no especializados, migrando publicaciones tradicionales hacia interactivas. El nacimiento de este tipo de publicaciones académicas en ciencias sociales –debido a la adopción de las redes sociales– representa una clara innovación en la forma en que las revistas interactúan con sus autores y lectores.

> Los investigadores han incorporado las redes sociales como *Facebook* o *Twitter* a sus prácticas habituales de comunicación

Por otra parte, la difusión del conocimiento científico a través de las redes sociales acelera el alcance obtenido por los resultados de la investigación, dando más visibilidad a la ciencia y promoviendo el aprendizaje abierto. Los editores son actores clave en este proceso, debido a su gran influencia sobre las editoriales y asociaciones científicas. Como expertos locales, los editores son vitales en la domesticación de las TIC (**Stewart**, 2007), lo que hace relevante la comprensión de los factores que impulsan a estos agentes a adoptar los medios sociales en sus revistas y la forma en que estos factores directa o indirectamente afectan el uso real de las redes sociales en interacción con atributos personales como el sexo o la edad.

Algunos estudios anteriores han abordado las razones por las que los investigadores adoptan las redes sociales para comunicar sus resultados (**Santana-Arroyo**, 2011; **Arcila-Calderón**; **Calderín-Cruz**; **Aguaded**, 2015), sin embargo no existen estudios precedentes sobre los editores de revistas científicas desde una perspectiva teórica de adopción de innovaciones tecnológicas.

Este trabajo pretende llenar ese vacío analizando los factores que impulsan a los editores de revistas de ciencias sociales a adoptar los medios sociales. Sobre todo, busca comprender los mecanismos y condiciones bajo los cuales operan estas variables, es decir, valorar las relaciones de mediación para conocer los efectos indirectos de una variable sobre otra a través de una tercera y las relaciones de moderación. En las siguientes secciones ofrecemos un panorama de la investigación realizada y el marco teórico que apoya nuestras hipótesis. A continuación, describimos el método de esta investigación, un informe de las conclusiones, y finalmente presentamos una discusión.

2. Medios sociales y publicaciones científicas

En los últimos años las revistas científicas han ido adoptando progresivamente tanto redes sociales generalistas (*Twitter* y *Facebook*), como académicas (*ResearchGate* y *Academia*); y profesionales (*LinkedIn*) (**Procter** *et al.*, 2010; **Ponte**; **Simon**, 2011; **Campos-Freire**; **Valencia**, 2015; **Campos-Freire**; **Rúas-Araújo**, 2016).

Este hecho se explica por las múltiples interconexiones que los medios sociales producen entre los estudiosos, lo que permite compartir fácilmente datos y acelerar la colaboración científica (**Murthy**; **Lewis**, 2015). En este sentido, la adopción de estas innovaciones podría estar relacionada con la tasa de colaboración científica (**Hara** *et al.*, 2003; **Cummings**; **Kiesler**, 2005; **Rigby**; **Edler**, 2005) o incluso el factor de impacto individual (**Liao**, 2011; **Direito-Rebollal**; **Campos-Freire**, 2016). Existe una extensa bibliografía sobre cómo los investigadores se han adaptado a las e-infraestructuras emergentes y a las TIC en ciencias sociales y humanidades en Estados Unidos (**Barjak** *et al.*, 2010), Europa (**Voss** *et al.*, 2010; **Procter**; **Voss**; **Asgari-Targhi**, 2013) y América Latina (**Briceño**; **Arcila-Calderón**; **Said-Hung**, 2012; **Arcila-Calderón**; **Piñuel-Raigada**; **Calderín-Cruz**, 2013; **Briceño**, 2014). Hay pruebas sólidas de que las actitudes hacia la *e-social research* son positivas (**Meyer**; **Dutton**, 2009) pero con un apoyo inicial limitado (**Dutton**; **Meyer**, 2009).

Específicamente, compartir los trabajos académicos en la web 2.0 ha sido considerado un reto (**Acord**; **Harley**, 2013), porque el uso de sitios como *Twitter* aumenta significativamente la difusión y el impacto de la investigación original (**Bruns**; **Liang**, 2012). Estudios similares también han abordado esta preocupación en el campo de la publicación académica examinando el papel de los repositorios abiertos que usan software *DSpace* o *Eprints*, el intercambio de documentos peer-to-peer (email, *ResearchGate*, etc.), o piratería (*Sci-Hub*, *LibGen*).

En países como España, estudios anteriores encontraron que las herramientas de la web 2.0 apenas fueron usadas por revistas de ciencias sociales (**Oller-Alonso**; **Segarra-Saavedra**; **Plaza-Nogueira**, 2012), incluso cuando esta innovación podría impactar positivamente en la comunicación científica en términos de marketing viral (**Bellón-Rodríguez**; **Sixto-García**, 2011). Sin embargo, no existen estudios comparativos o internacionales que expliquen la adopción de los medios sociales por revistas en ciencias sociales, por lo que nos preguntamos:

> RQ1. ¿Hasta qué punto redes las sociales generalistas (*Twitter* o *Facebook*), académicas (*ResearchGate* o *Academia*), y profesionales, (*LinkedIn*) son adoptadas por revistas científicas en ciencias sociales?

3. Expectativa de rendimiento, Expectativa de esfuerzo e Influencia social

El marco de difusión de la innovación se ha aplicado a las TIC en diferentes contextos como la adopción de medios sociales por los políticos (**Quinlan** *et al.*, 2017) o por científicos (**Arcila-Calderón**; **Calderín-Cruz**; **Aguaded**, 2015). Ahora buscamos los factores que conforman la adopción de redes sociales en publicaciones académicas (utilizando sus editores como una fuente de confianza) y explorando los mecanismos y condiciones bajo los cuales operan. Existen varios modelos de aceptación de la tecnología que derivan del paradigma de difusión de innovaciones. Entre ellos, Utaut (*User acceptance and use of technology* o *Teoría unificada de aceptación y uso de tecnología*), propuesta por **Venkatesh** *et al*. (2003), que maneja variables como la expectativa de rendimiento, expectativa de esfuerzo, influencia social, condiciones facilitadas, entre otras; y el modelo Utaut2 (**Venkatesh**; **Thong**; **Xu**, 2012), que incorpora las variables de motivación hedónica, valor de precio, experiencia y hábito.

En este estudio seguimos un modelo simplificado de tres factores principales basado en Utaut (figura 1):

1) Expectativa de rendimiento: grado en que el uso de una tecnología proporcionará algún beneficio al usuario al realizar ciertas actividades;

2) Expectativa de esfuerzo: facilidad de utilización de una tecnología percibida por parte del usuario;

3) Influencia social: medida en que los usuarios perciben que personas cercanas consideran apropiado el uso de una tecnología.

El modelo establece que hay un efecto condicional indirecto de estas tres variables sobre el uso real de las redes sociales. Dicho efecto es mediado por la intención de uso y moderado por el género y la edad. En línea con Utaut y para probar estas afirmaciones, planteamos que:

> H1. La Expectativa de rendimiento influye de manera positiva en la intención de uso de redes sociales por los

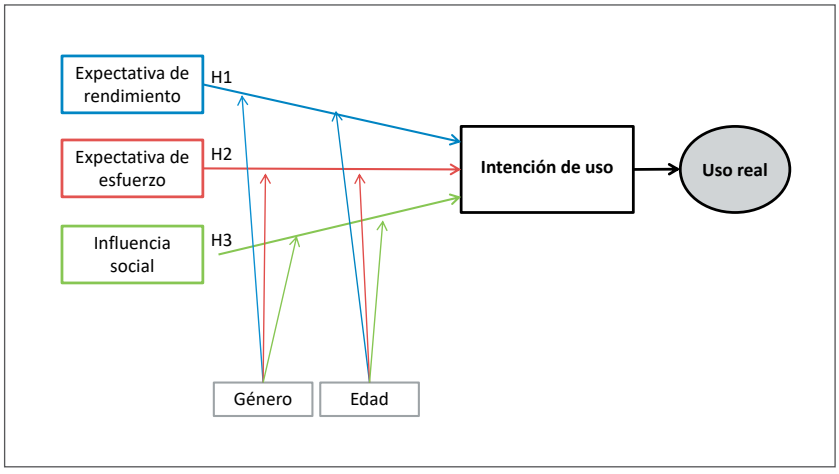


Figura 1. Modelo simplificado de tres factores principales de Utaut.

editores, generando un efecto indirecto positivo en su uso real. Dicho efecto indirecto está moderado por el género y la edad, de manera que será más fuerte en los hombres y en los jóvenes.

H2. La Expectativa de esfuerzo influye negativamente en la intención de uso de redes sociales por los editores, generando un efecto indirecto negativo en su uso real. Este efecto indirecto está moderado por el género y la edad, siendo más fuerte en las mujeres y en las más mayores.

H3. La Influencia social de los editores influye positivamente en la intención de uso de redes sociales, lo que a su vez genera una influencia positiva en su uso real. El efecto indirecto es moderado significativamente por el género y la edad, de modo que será más fuerte en las mujeres y en las personas de más edad.

4. Método

4.1. Participantes y procedimientos

> Los editores son actores clave en los procesos de comunicación de la ciencia, lo que hace relevante comprender los factores que les impulsan a adoptar las redes sociales en sus revistas

Se calculó el tamaño muestral mínimo requerido para contrastar las hipótesis formuladas teniendo en cuenta que se realizaría un análisis de regresión lineal con tres predictores basado en mínimos cuadrados ordinarios. La estimación se hizo en *G*Power 3.1* (**Faul** *et al.*, 2009), utilizando como parámetros los tamaños de los efectos ya conocidos por los predictores de Utaut (**Venkatesh**; **Thong**; **Xu**, 2012; **Taiwo**; **Downe**, 2013), un error tipo 1, α, de 0,05 y una potencia estadística (1-β) de 0,80. El meta-análisis de **Taiwo** y **Downe** (2013) encontró correlaciones (Zr) *pequeñas* o *medianas* entre la Expectativa de rendimiento y la intención de uso (Zr=0,54), la Expectativa de esfuerzo y la intención de uso (Zr=0,44) y la Influencia social y la intención de uso (Zr=0,42). Usando esta referencia y para detectar efectos *medianos* (f^2=0,15) (**Cohen**, 1992; **Faul** *et al.*, 2007) en nuestra regresión lineal múltiple con tres predictores, obtuvimos que la muestra mínima necesaria es de 43 casos.

Para obtener esta muestra se invitó a 300 editores de las 20 principales revistas de ciencias sociales en *Google Scholar Metrics*, que forman parte de las siguientes áreas:

- Antropología;
- Negocios, economía y administración;
- Comunicación;
- Ciencias sociales;
- Educación;
- Geografía y estudios urbanos;
- Derecho;
- Biblioteca y ciencias de la información;
- Ciencia política, administración y relaciones internacionales;
- Psicología;
- Sociología;
- Obras sociales y servicios;
- Pruebas académicas y psicológicas;
- Estudios africanos e historia;
- multidisciplinar.

Después de la invitación en octubre de 2016, 56 editores respondieron correctamente al cuestionario, una tasa de respuesta del 20%.

4.2. Medidas utilizadas

Los indicadores utilizados se componían de varios items para cada variable:

- Expectativa de rendimiento (4 items, α_c=0,87);
- Expectativa de esfuerzo[1] (4 items, α_c=0,87) (medida del 1 al 5, donde el 5 es más facilidad);
- Influencia social (3 items, α_c=0,83).

Para realizar la medición se utilizó una escala de Likert de 5 puntos, desde el 1 (completamente en desacuerdo) hasta el 5 (completamente de acuerdo) (tabla 1). Las mediciones de la intención de uso diario, frecuente y futuro, se hicieron para cada red social, por el mismo sistema, encontrando en cada caso una fiabilidad de:

- *Twitter*: α_c=0,94
- *Facebook*: α_c=0,92
- *ResearchGate*: α_c=0,75
- *Academia*: α_c=0,58
- *LinkedIn*: α_c=0,72.

La frecuencia de uso se midió con la misma escala, desde nunca (1) hasta siempre (5). El uso real se midió con respuesta dicotómica (0, No adoptado; 1, Sí adoptado). Además se consideraron variables sociodemográficas:

- edad;
- género (1=hombre, 2=mujer);
- nivel académico;
- estatus laboral;
- país de trabajo;
- temática editorial.

> No hay estudios precedentes que exploren a los editores de revistas científicas desde una perspectiva teórica de adopción de innovaciones tecnológicas

El estadístico Alpha de Cronbach (α_c) se utilizó para determinar la fiabilidad por la consistencia interna, garantizando el mínimo requerido 0,70 (**Cronbach**, 1951).

4.3. Análisis

Se comenzó el análisis realizando una revisión de los datos después de codificarlos para detectar posibles errores. Los valores perdidos fueron estimados (**Myers**, 2011) usando regresiones ordinarias de mínimos cuadrados con 5 imputaciones. A continuación, con los datos depurados, se realizó un análisis estadístico inductivo-exploratorio que aportaba información del uso de las redes sociales. Para finalizar se llevó a cabo un análisis de mediación moderada para comprobar las hipótesis por medio del programa *SPSS* y la macro *Process* (Modelo 9) (**Hayes**, 2013), con 10.000 muestras de *bootstrap* (95% e intervalos de confianza corregidos por sesgo). Este método está basado en el análisis de regresión lineal múltiple estimado por mínimos cuadrados ordinarios (OLS).

5. Resultados

5.1. Análisis exploratorio de los datos

Los resultados exploratorios obtenidos sobre la tasa de adopción de los medios sociales estudiados (RQ1), indican que la tasa promedio es del 38,21% (M=1,88 DE=0,63), lo que significa que la adopción es baja si se tienen en cuenta todas las redes en su conjunto. Pero el uso real de cada red a nivel individual varía, siendo las tasas de adopción:

- *Twitter*: 69,60% (M=0,69 DE=0,46)
- *Facebook*: 57,14% (M=0,57 DE=0,49)
- *LinkedIn*: 26,80% (M=0,26 DE=0,44)
- *ResearchGate*: 21,40% (M=0,21 DE=0,41)
- *Academia*: 16,10% (M=0,16 DE=0,37).

Las redes con mayor frecuencia de acceso en la escala de 1-5 son, de mayor a menor:

- *Twitter*: (M=2,66 DE=1,37)
- *Facebook*: (M=2,24 DE=1,30)
- *ResearchGate*: (M=1,63, DE=1,04)
- *LinkedIn*:(M=1,54 DE=0,83)
- *Academia*: (M=1,34 DE=0,63).

Tabla 1. Expectativa de rendimiento, Expectativa de esfuerzo e Influencia social.

Expectativa de rendimiento	**M**	**DE**
Usar las redes sociales aumenta el impacto de mi revista	3,52	1,10
El uso de las redes sociales aumenta las posibilidades de mi revista de lograr cosas importantes para ella	3,50	1,10
Encuentro los medios sociales útiles en mi revista	3,30	1,18
El uso de las redes sociales ayuda a mi revista a lograr cosas más rápidamente	2,97	1,09
Total	3,32	0,95
Expectativa de esfuerzo	M	DE
El personal de mi revista encuentra que las redes sociales son fáciles de usar	3,54	1,01
Es fácil para el personal de mi revista llegar a ser hábil en las redes sociales	3,51	1,03
Aprender a usar las redes sociales es fácil para el personal de mi revista	3,48	1,00
La interacción del personal de mi revista con las redes sociales es clara y comprensible	3,12	1,05
Total	3,41	0,87
Influencia social	**M**	**DE**
Las personas que influyen en el comportamiento de mi revista (editor, patrocinadores) piensan que debería usar las redes sociales	3,56	1,22
Las personas que son importantes para mi revista, piensan que debería usar las redes sociales	3,34	1,11
Las personas cuyas opiniones son valoradas por mi revista (autores, lectores) prefieren que use las redes sociales	3,23	1,16
Total	3,38	1,01

Twitter y *Facebook* son por tanto las más usadas y las de acceso más frecuente.

La intención de uso (en una escala de 1 a 5) muestra también que *Twitter* (M=3,38 DE=1,50) y *Facebook* (M=2,82 DE=1,55) son las preferentes, seguidas de *ResearchGate* (M=1,77 DE=1,33), *Academia* (M=1,47 DE=0,85) y por último *LinkedIn* (M=1,52 DE=1,13). Tomando estos valores, la media se sitúa en 2,18 (DE=0,66).

Los datos obtenidos muestran que los editores tenían una Expectativa de rendimiento (EdR) en una escala de 1 a 5 hacia las redes sociales, con valores de 2,97 a 3,52, lo que implica que su percepción es positiva, siendo la media de 3,32 (M=3,32 DE=0,95).

En cuanto a la Expectativa de esfuerzo (EdE) en la escala de 1 a 5, los valores van de 3,12 a 3,54, obteniendo una media de 3,41, lo que implica que su percepción es baja.

La Influencia social ha obtenido unos valores que van de 3,23 a 3,56, obteniendo una media de 3,38, lo que implica que su percepción es alta (tabla 1).

5.2. Análisis de mediación moderada

5.2.1. Expectativa de rendimiento

Para la variable independiente Expectativa de rendimiento (EdR) se ha realizado un análisis de mediación moderada en cada red social (figura 2).

Para *Facebook* obtenemos que el efecto directo de la Expectativa de rendimiento sobre el uso real no se ha mostrado significativo (B=-1,07, p=0,07, IC=de -2,24 a 0,09). A pesar de ello existe un efecto condicional indirecto que delata esa relación (tabla 2) a través de la intención de uso (el efecto, con valores de 0,86 a 1,85, fue significativo para los 3 niveles de género moderador masculino y no significativo en la totalidad del género femenino). Este resultado sostiene H1, que postulaba que la Expectativa de rendimiento influencia la intención de uso y ésta, a su vez, influye de modo positivo en el uso real de las redes sociales, con su efecto moderado por la edad y el género, siendo más fuerte en los hombres jóvenes.

> La Expectativa de rendimiento es el factor predictivo más fuerte en la adopción de una tecnología, pero en este estudio no se confirma totalmente el modelo Utaut

Para *Twitter*, el efecto directo de la Expectativa de rendimiento sobre el uso real no se mostró significativo (B=0,22, p=0,71, IC=-0,94 a 1,38). Pero sí se observó esa relación a través de un efecto condicional indirecto (el efecto resultó significativo para los 4 niveles de la variable edad moderadora más altos en ambos géneros, con variación de 1,49 a 1,64), lo que apoya H1 y fue más fuerte en las mujeres de menor edad, lo que nos indica que existe moderación de la edad y el género, pero contrario a lo planteado en H1.

En *ResearchGate* no se aprecia que sea significativo el efecto directo de la Expectativa de rendimiento sobre el uso real (B=-0,29; p=0,54; IC=-1,22 a 0,65). Tampoco se produce relación de modo indirecto a través de la intención de uso (el efecto no se mostró significativo en edad ni género, variando de -0,52 a 0,93, siendo más fuerte en las mujeres mayores, en la dirección opuesta a lo planteado en H1).

Para *Academia*, no fue significativo el efecto directo que ejerce la Expectativa de rendimiento sobre el uso real (B=-0,33, p=0,60, IC=-1,59 a 0,93). La relación tampoco se ha revelado a través de un efecto condicional indirecto de la intención de uso (los valores del efecto variaron de -0,37 a -0,47, no siendo sig-

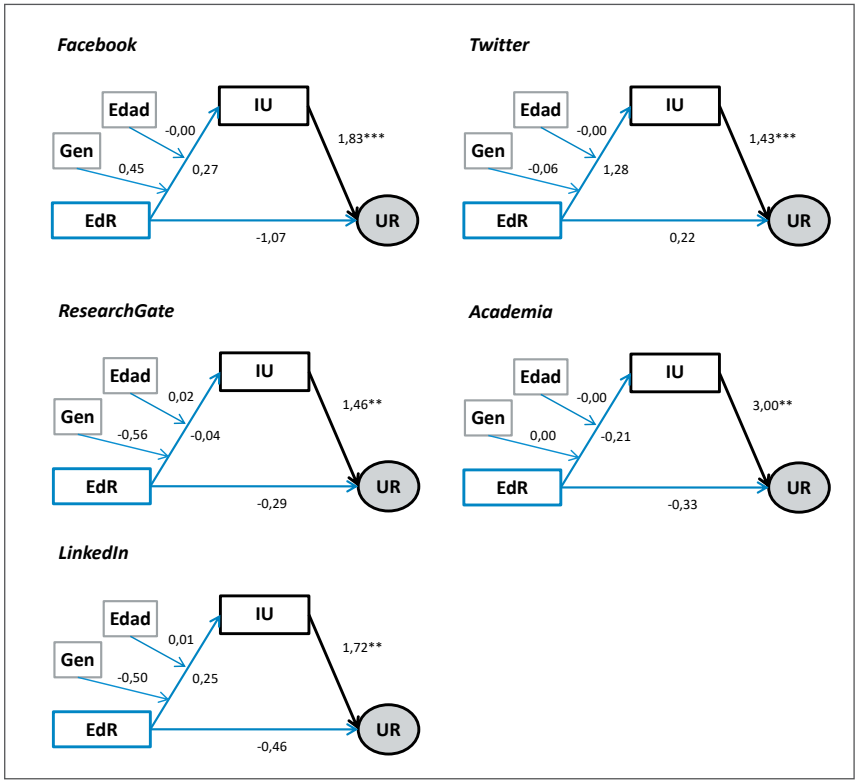


Figura 2. Expectativa de rendimiento. Coeficientes no normalizados de regresión (B) para cada modelo. (* = P < 0,05; ** = P < 0,01; *** = P < 0,001)

nificativo para edad ni género, a pesar de ser más fuerte en hombres y mujeres en los tramos de mayor edad, de modo opuesto a lo reflejado en H1).

Para *LinkedIn* tampoco fue significativo el efecto directo sobre el uso real de la Expectativa de rendimiento (B=-0,46, p=0,27, IC=-1,29 a 0,36). Igualmente, no se produce relación mediando la intención de uso (la variación del efecto fue de -0,53 a 0,74, no siendo significativo para el moderador edad ni género, y mostrándose más fuerte en mujeres de los 3 niveles de edad, opuesto a H1) (tabla 2).

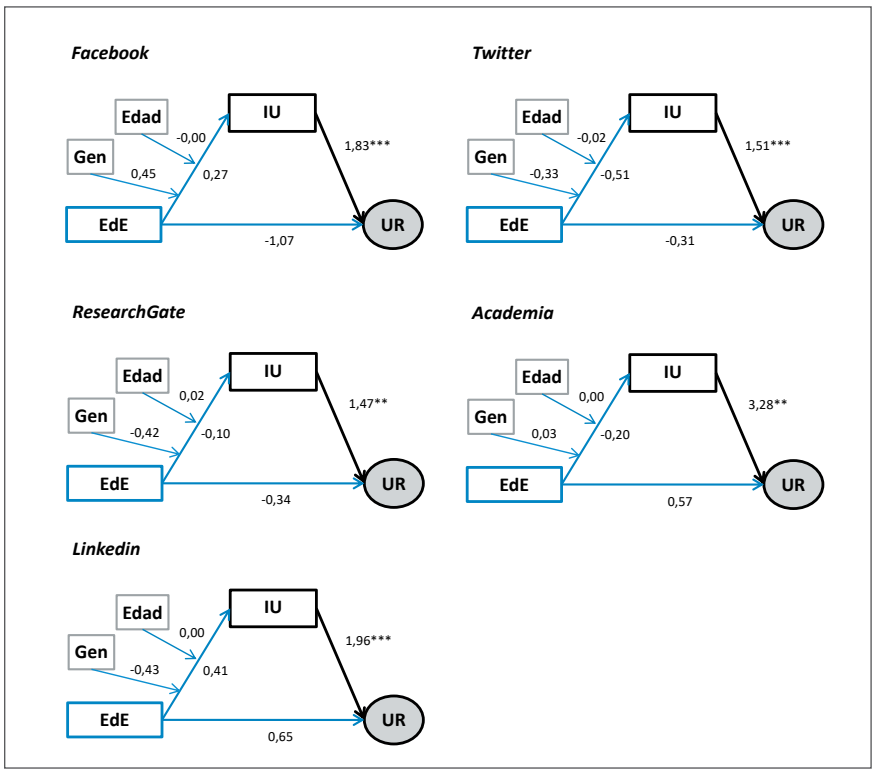


Figura 3. Expectativa de esfuerzo. Coeficientes no normalizados de regresión (B) para cada modelo. (* = P < 0,05; ** = P < 0,01; *** = P < 0,001)

5.2.2. Expectativa de esfuerzo

Para la variable independiente Expectativa de esfuerzo (EdE), se ha realizado un análisis de mediación moderada para una de las redes sociales (figura 3).

Comenzando con *Facebook*, el efecto directo de la Expectativa de esfuerzo sobre el uso real no fue significativo (B=-0,44, p=0,32, IC=de -1,30 a 0,43), ni hubo relación a través de la intención de uso mediante un efecto condicional indirecto (sus valores variaron de -0,02 a 0,48, no siendo significativo para el moderador edad ni género, y mostrándose más fuerte en los hombres de más edad, en el sentido opuesto de género al planteado en H2).

Atendiendo a *Twitter*, no se mostró significativo el efecto directo de la Expectativa de esfuerzo sobre el uso real (B=-0,314, p=0,55, IC=de -1,35 a 0,72). Esta relación tampoco se produjo a través de la intención de uso mediante un efecto condicional indirecto (los valores variaron de -0,24 a 1,10, no siendo significativo para el moderador edad ni género, y mostrándose más fuerte en las mujeres más mayores, lo que apoya H2).

Para *ResearchGate* el efecto directo de la Expectativa de esfuerzo sobre el uso real no fue significativo (B=-0,34, p=0,50, IC=-1,33 a 0,65). No se encontró otra vía de relación mediante la intención de uso con un efecto no directo (los valores variaron de -0,20 a 1,08, no siendo significativo para el moderador edad ni género, pero mostrándose con más intensidad en mujeres más mayores, en línea con H2).

> La tasa de adopción de cada red individual varía notablemente, siendo *Twitter* y *Facebook* las más utilizadas

En *Academia* no se ha demostrado un efecto directo de la Expectativa de esfuerzo sobre el uso real (B=-0,57, p=0,39, IC=-0,72 a 1,86). De forma similar, no se ha producido un vínculo a través de la intención de uso mediante un efecto condicional indirecto (los valores variaban de -0,20 a 0,10, de manera que no fue significativo para el moderador edad ni género, siendo más fuerte en los hombres más mayores, lo que indica un sentido opuesto de género al planteado en H2 pero coincidiendo en edad).

Por último, en *LinkedIn* no fue significativo un efecto directo de la Expectativa de esfuerzo sobre el uso real (B=0,65, p=0,20, IC=-0,35 a 1,66). No se ha revelado tampoco conexión indirecta mediante la intención de uso (los valores del efecto variaron de -0,73 a 0,20, y no fue significativo en el moderador edad ni género –excepto en hombres de 54 años, pero siendo más intenso en las mujeres de más edad, sosteniendo lo propuesto en H2).

5.2.3. Influencia social

Para la variable independiente Influencia social (IS), se ha realizado un análisis de mediación moderada en cada uno de los medios sociales (figura 4).

En el caso de *Facebook*, el efecto directo de la Influencia social sobre el uso real no fue significativo (B=-0,47, p=0,27, IC=-1,30 a 0,36). A pesar de ello, a través de la intención de uso se genera un efecto condicional indirecto (los valores variaron de 0,75 a 0,98 y sólo fueron significativos para el nivel intermedio de edad en hombres), siendo más fuerte en los hombres más jóvenes, lo que muestra un sentido opuesto al planteado en H3 (que indicaba que el efecto de la Influencia

social sobre la intención de uso es moderado significativamente por el género y la edad, de tal manera que el efecto será más fuerte en las mujeres y en las personas de más edad).

En *Twitter*, de un modo directo no fue significativo el efecto de la Influencia social sobre el uso real (B=-0,59, p=0,23, IC=-0,37 a 1,56). La relación en cambio, sí se aprecia a través de la intención de uso, con un efecto condicional indirecto (los valores variaron de 0,50 a 1,23 y sólo fueron significativos para el nivel intermedio de edad en hombres), siendo más fuerte en los hombres y en los más jóvenes, en sentido opuesto a H3.

ResearchGate no muestra un efecto directo significativo de la Influencia social sobre el uso real (B=-0,04, p=0,92, IC=-0,87 a 0,79). Además, no se aprecia que la intención de uso tenga un efecto condicional indirecto (varió de -0,76 a 0,27, no fue significativo para edad ni género, aunque mostraba más intensidad en los hombres y en los más jóvenes, en sentido opuesto a H3).

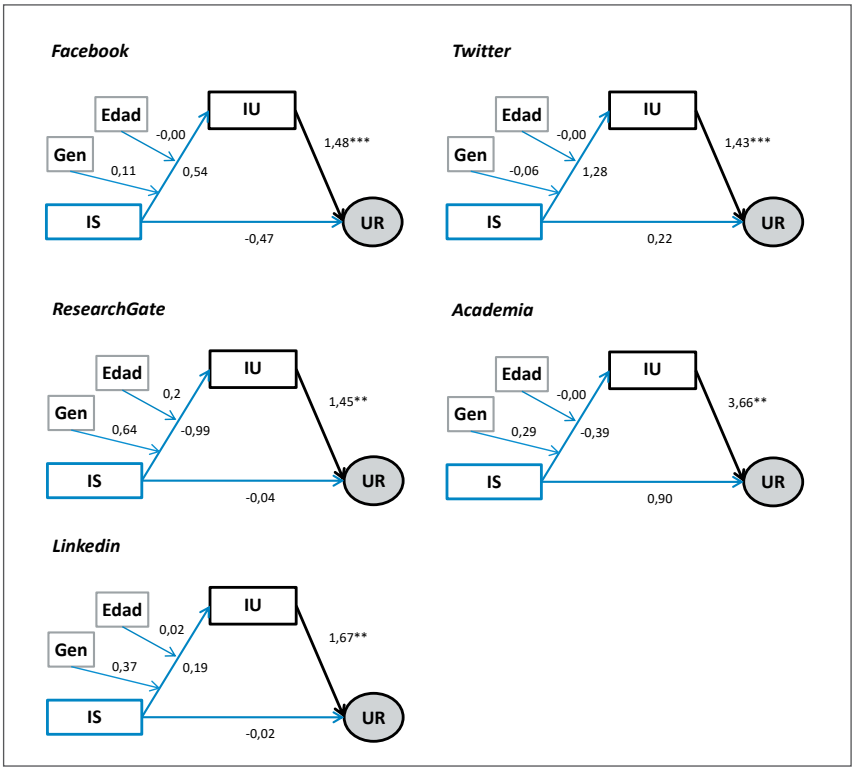


Figura 4. Influencia social. Coeficientes no normalizados de regresión (B) para cada modelo. (* = P < 0,05; ** = P < 0,01; *** = P < 0,001)

En *Academia* no se muestra significancia en el efecto directo de la Influencia social sobre el uso real (B=0,90, p=0,14, IC=-0,29 a 2,10). De igual forma, no hubo relación mediante un efecto condicional indirecto a través de la intención de uso (el valor variaba de -1,25 a 0,13, no fue significativo para ninguno de los niveles de edad ni género, pero era más intenso en los hombres y en los más jóvenes, en sentido opuesto a H3).

En *LinkedIn* no fue significativo el efecto directo de la Influencia social sobre el uso real (B=-0,02, p=0,96, IC=-0,82 a 0,78). De igual modo no se aprecia un efecto condicional indirecto reseñable a través de la intención de uso (varió de -1,09 a 0,25, pero no fue significativo para ninguno de los niveles de edad ni género, mostrando más intensidad en los hombres y en los más jóvenes, lo que es opuesto a H3).

6. Discusión y conclusión

Hay que considerar en primer lugar que los tipos de redes sociales estudiadas son diversos, puesto que se han incluido redes sociales generalistas (*Twitter* y *Facebook*), académicas (*ResearchGate* y *Academia*), y profesionales (*LinkedIn*), para estudiar el uso que hacen de ellas los editores de revistas de ciencias sociales específicamente. A pesar de que las dos redes generalistas son las más extendidas, su uso por investigadores o con fines académicos genera escepticismo (**Rúas-Araújo**; **Campos-Freire**; **Puentes-Rivera**, 2016). En este caso se trata la utilización de estas de redes sociales por revistas de ciencias sociales, cuya finalidad es la difusión de su contenido y trabajo diario.

En atención a los datos obtenidos, la tasa de adopción real de las redes sociales (RQ1) es baja (38,21%) cuando revisamos el conjunto de las redes analizadas (*Facebook*, *Twitter*, *ResearchGate*, *Academia* y *LinkedIn*).

Pero si analizamos la tasa de adopción individual apreciamos que varía, siendo la de mayor adopción *Twitter* (69,60%), seguida de *Facebook* (57,14%), *LinkedIn* (26,80%), *ResearchGate* (21,40%) y *Academia* (16,10%). En la frecuencia de acceso se colocan en primer lugar *Twitter* y *Facebook*, seguidas de *ResearchGate*, *LinkedIn* y *Academia*. La intención de uso medio es mayor en *Twitter* y *Facebook*, seguidas de *ResearchGate*, *LinkedIn* y *Academia*. No se aprecia en ninguno de los casos que exista un efecto directo de la Expectativa de rendimiento sobre el uso real, sino que en alguna de las relaciones el efecto es *mediado* por la intención de uso. A su vez, el uso real recibe un efecto indirecto de la Expectativa de rendimiento que está moderado en algunos casos por el género y la edad, según propone el modelo Utaut (**Venkatesh** *et al.*, 2003; **Venkatesh**; **Thong**; **Xu**, 2012), generando un efecto condicional indirecto.

> Los efectos de la Expectativa de rendimiento sobre el uso real a través de la intención de uso de medios sociales por los editores de las revistas sólo son significativos en el caso de *Facebook* y *Twitter*

Tabla 2. Efecto condicional indirecto de la EdR, EdE e IS en el uso real a través de la intención de uso, moderada por edad y género.

			Expectativa de rendimiento				**Expectativa de esfuerzo**				**Influencia social**			
Redes sociales	Edad	Género	Efecto	BOOT SE	BOOT LLCI	BOOT ULCI	Efecto	BOOT SE	BOOT LLCI	BOOT ULCI	Efecto	BOOT SE	BOOT LLCI	BOOT ULCI
FB	42	F	1,02	1,44	-1,02	3,89	-0,02	0,90	-1,55	2,03	0,83	0,98	-0,92	2,85
	42	M	1,85	1,28	**0,32**	**3,34**	0,31	0,97	-1,74	2,08	0,98	0,89	-0,38	2,88
	54	F	0,94	1,47	-1,89	3,49	0,06	0,79	-1,40	1,79	0,79	1,07	-1,36	2,73
	54	M	1,77	1,05	**0,75**	**3,14**	0,40	0,64	-0,70	1,77	0,95	0,67	**0,06**	**2,42**
	65	F	0,86	1,69	-2,65	3,76	0,15	0,95	-1,85	1,99	0,75	1,29	-1,98	3,07
	65	M	1,69	1,08	**0,44**	**3,89**	0,48	0,61	-0,48	1,98	0,91	0,67	-0,15	2,23
TW	42	F	1,64	1,88	-0,22	3,89	0,26	1,24	-1,30	2,86	0,96	1,38	-0,57	4,24
	42	M	1,55	1,56	-0,15	3,39	-0,25	1,47	-4,48	1,37	1,23	1,48	-0,24	3,96
	54	F	1,61	1,82	0,00	4,21	0,68	1,20	-1,41	2,47	0,73	1,29	-0,84	4,18
	54	M	1,52	1,43	**0,07**	**3,45**	0,17	1,05	-1,90	1,64	1,00	1,16	**0,00**	**3,44**
	65	F	1,58	1,85	**0,05**	**5,31**	1,10	1,74	-1,39	4,46	0,50	1,47	-1,72	3,78
	65	M	1,49	1,41	**0,41**	**5,43**	0,59	1,31	-0,71	4,23	0,77	1,11	-0,43	3,41
RG	42	F	0,29	1,54	-1,24	4,39	0,42	1,14	-1,16	2,81	-0,67	2,06	-3,52	0,30
	42	M	-0,52	1,41	-3,27	0,24	-0,20	1,14	-2,74	0,87	0,27	0,94	-0,49	2,78
	54	F	0,61	2,03	-0,98	5,27	0,75	1,56	-0,90	4,00	-0,72	2,21	-4,14	0,44
	54	M	-0,20	0,75	-1,61	0,45	0,13	0,75	-0,97	1,65	0,23	0,72	-0,24	2,22
	65	F	0,93	2,69	-1,08	6,65	1,08	2,28	-0,97	5,65	-0,76	2,52	-5,53	0,84
	65	M	0,12	0,69	-0,71	2,04	0,46	1,21	-0,82	2,85	0,18	1,01	-0,99	2,36
AC	42	F	-0,47	1,51	-2,50	2,36	-0,20	1,66	-2,13	4,51	-0,93	3,05	-7,89	1,87
	42	M	-0,47	1,59	-3,79	0,73	-0,09	1,72	-3,70	2,45	0,13	1,27	-3,20	2,36
	54	F	-0,42	1,55	-3,09	2,13	-0,10	1,75	-2,38	3,62	-1,09	3,30	-9,10	1,78
	54	M	-0,41	1,02	-2,23	0,50	0,01	1,33	-2,07	2,89	-0,03	0,97	-1,97	2,20
	65	F	-0,37	1,98	-4,10	3,04	0,00	2,43	-4,05	3,65	-1,25	3,80	-10,84	2,64
	65	M	-0,37	0,94	-1,81	1,45	0,11	1,77	-2,68	3,97	-0,19	1,48	-2,78	4,05
LK	42	F	0,33	0,74	-0,95	1,78	0,11	0,71	-1,14	1,66	-0,37	1,45	-3,86	0,84
	42	M	-0,53	0,77	-2,79	0,09	-0,74	0,77	-2,86	0,32	0,25	0,59	-0,59	1,26
	54	F	0,53	0,86	-1,20	2,00	0,16	0,80	-1,58	1,44	-0,73	1,83	-5,10	0,56
	54	M	-0,32	0,53	-1,63	0,18	-0,69	0,55	-2,05	-0,05	-0,11	0,30	-0,68	0,43
	65	F	0,74	1,10	-1,45	2,74	0,20	1,06	-2,38	1,55	-1,09	2,24	-6,31	0,49
	65	M	-0,12	0,53	-0,99	0,92	-0,65	0,59	-1,68	0,29	-0,47	0,50	-1,38	0,25

Los resultados obtenidos sobre este efecto condicional indirecto ejercido por la variable Expectativa de rendimiento, a través de la intención de uso, sobre el uso real de redes sociales por los editores de las revistas (H1), sólo son significativos en el caso de *Facebook* y *Twitter*. Esto supone que si un editor considera que la adopción de esos medios sociales puede mejorar el rendimiento de su revista en el futuro, aumenta su intención de usarlo y por lo tanto la probabilidad de uso real.

Facebook muestra un efecto indirecto condicional más fuerte en los hombres jóvenes, apoyando H1. *Twitter* muestra un efecto indirecto más fuerte en las mujeres jóvenes, en sentido opuesto a H1. Los resultados de la mediación no son significativos en *ResearchGate* y *LinkedIn*, donde el efecto indirecto fue más intenso en mujeres de mayor edad, mientras que en *Academia* no se aprecia diferencia de género y el efecto aumenta con la edad.

No hay suficiente evidencia empírica de que la variable Expectativa de rendimiento tenga influencia en el uso real de estas tres redes sociales a través de la variable intención de uso, ni que la edad o el género moderen el tamaño en ningún sentido.

Por lo tanto, podemos indicar que la Expectativa de rendimiento es un importante factor de predicción de uso en determinadas redes sociales con medidas de coeficientes no normalizados de hasta 3,00. También conocida como utilidad percibida (**Davis**, 1989; **Davis**; **Bagozzi**; **Warshaw**, 1992) es determinante en la adopción de tecnologías, indicando que la decisión de adoptar o no una tecnología está condicionada por el convencimiento individual.

Revisando H2, los resultados del análisis del efecto indirecto negativo ejercido por la Expectativa de esfuerzo a través de la intención de uso, sobre el uso real de redes sociales por los editores de las revistas, no son significativos. Si los editores valoran el esfuerzo que supone la adopción de determinadas redes sociales baja la probabilidad de su uso real. No se aprecia en ninguno de los casos un efecto directo o indirecto de la Expectativa de esfuerzo sobre el uso real, según propone Utaut (**Venkatesh** *et al.*, 2003; **Venkatesh**; **Thong**; **Xu**, 2012). En *Twitter*, *ResearchGate* y *LinkedIn* el efecto de interacción fue más fuerte en las mujeres mayores. En *Facebook* y *Academia* el efecto es más intenso en hombres mayores. Por tanto, la Expectativa de esfuerzo no se puede considerar un predictor de uso negativo de los medios sociales con medidas de coeficientes no normalizados hasta 3,28, y no podemos demostrar empíricamente el efecto condicional indirecto, ni una moderación de género o edad con una dirección clara y aplicable a todas las redes sociales estudiadas.

Por último, valorando H3, los resultados confirman un efecto condicional indirecto positivo ejercido por la Influencia social a través de la intención de uso, sobre el uso real de redes sociales por los editores, pero sólo en caso de *Facebook* y *Twitter*. Esto supone que cuando los editores se sienten influenciados positivamente en su entorno social, colaboradores o amigos, aumenta la probabilidad de que adopten dichos medios sociales. No apreciamos en ningún caso un efecto directo significativo de la Influencia social sobre el uso real. El efecto de la Influencia social sobre el uso real, se modera con las variables de género y edad según propone el modelo Utaut (**Venkatesh** *et al.*, 2003; **Venkatesh**; **Thong**; **Xu**, 2012), y muestra un efecto más intenso en mujeres jóvenes, opuesto a H3. Podemos considerar que la Influencia social es predictor del uso de *Facebook* y *Twitter*, con medidas de coeficientes no normalizados hasta 3,66, y que el género y la edad pueden ser moderadores más fuertes en mujeres jóvenes.

Podemos valorar las implicaciones teóricas o prácticas de estos resultados. Para empezar, en línea con los estudios previos existentes, se confirma que el factor predictivo más fuerte en la adopción de una tecnología es la Expectativa de rendimiento. A pesar de ello, en esta investigación concreta se ha demostrado un desacuerdo con el modelo Utaut y Utaut2 en cuanto a la influencia que las variables género y edad ejercen en el uso real de una tecnología, porque no son en todos los casos moderadoras de la relación entre los factores explicativos y el uso real. Tampoco apreciamos que la variable Expectativa de esfuerzo sea un predictor significativo en ninguno de los casos analizados. Por último, atendiendo a las implicaciones prácticas, este estudio puede servir de base para la generación de políticas de adopción de TIC por investigadores, considerando la utilidad que se percibe y los beneficios obtenidos. Hay algunas limitaciones que también hay que reconocer. Estamos considerando un uso profesional de los medios sociales aplicados en revistas científicas de ciencias sociales, por lo que investigaciones futuras se pueden encaminar a incluir los condicionantes existentes en este entorno como son la formación y experiencia previa, las condiciones facilitadas y las competencias tecnológicas individuales. También se pueden estudiar los usos concretos que las revistas aplican en cada medio social teniendo en cuenta los perfiles de uso y consumo concreto de cada tipo de red y ampliar la investigación a otros medios sociales según el objetivo de comunicación de la propia revista.

> “La Influencia social es predictor del uso de *Facebook* y *Twitter*, y el género y la edad pueden ser moderadores más fuertes en mujeres jóvenes”

7. Nota

1. La Expectativa de esfuerzo se mide como el nivel de facilidad percibida, de modo contrario a la teoría del modelo UTAUT, por lo que la correlación esperada es positiva. Números altos suponen baja Expectativa de esfuerzo y número bajos suponen alta Expectativa de esfuerzo.

8. Referencias

Acord, Sophia-Krzys; **Harley, Diane** (2013). “Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0”. *New media & society*, v. 15, n. 3, pp. 379-397.
https://doi.org/10.1177/1461444812465140

Arcila-Calderón, Carlos; **Calderín-Cruz, Mabel**; **Aguaded, Ignacio** (2015). “Adoption of ICTs by communication researchers for scientific diffusion and data analysis”. *El profesional de la información*, v. 24, n. 5, p. 526-536.
https://doi.org/10.3145/epi.2015.sep.03

Arcila-Calderón, Carlos; **Piñuel-Raigada, José-Luis**; **Calderín-Cruz, Mabel** (2013). “The e-research on media & communications: Attitudes, tools and practices in Latin America researchers”. *Comunicar*, v. 20, n. 40, pp. 111-118.
https://doi.org/10.3916/C40-2013-03-01

Barjak, Franz; **Lane, Julia**; **Poschen, Meik**; **Procter, Rob**; **Robinson, Simon**; **Wiegand, Gordon** (2010). “E-infraestructure adoption in the social sciences and humanities”. *Information, communication & society*, v. 13, n. 5, pp. 635-651.
https://doi.org/10.1080/13691180903095849

Bellón-Rodríguez, Ana; **Sixto-García, José** (2011). “Aplicación y uso de la web 2.0 y de las redes sociales en la comunicación científica especializada: del marketing viral al usuario activo”. *Anagramas. Rumbos y sentidos de la comunicación*, v. 9, n. 18, pp. 61-70.
https://doi.org/10.22395/angr.v9n18a4

Borgman, Christine L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. The MIT Press, ISBN: 978 0 262026192

Briceño, Ysabel (2014). "Saber y medios: hacia un modo emergente de la comunicación de la ciencia". *Bitácora-e. Revista electrónica latinoamericana de estudios sociales, históricos y culturales de la ciencia y la tecnología*, v. 1.
http://www.saber.ula.ve/handle/123456789/38746

Briceño, Ysabel; **Arcila-Calderón, Carlos**; **Said-Hung, Elías** (2012). "Colaboración y comunicación científica en la comunidad latinoamericana de físicos de altas energías". *E-colabora. Revista de ciencia, educación, innovación y cultura apoyadas por redes de tecnología avanzada*, v. 2, n. 4, pp. 131-144.

Bruns, Axel; **Liang, Yuxian-Eugene** (2012). "Tools and methods for capturing Twitter data during natural disasters". *First Monday*, v. 17, n. 4.
https://doi.org/10.5210/fm.v17i4.3937

Campos-Freire, Francisco; **Rúas-Araújo, José** (2016). "Uso de las redes sociales digitales profesionales y científicas: el caso de las 3 universidades gallegas". *El profesional de la información*, v. 25, n. 3, p. 431-440.
https://doi.org/10.3145/epi.2016.may.13

Campos-Freire, Francisco; **Valencia, Andrea** (2015). "Managing academic profiles on scientific social networks". En: Rocha, Álvaro; Correia, Ana-María; Costanzo, Sandra; Reis, Luis-Paulo (eds.). *New contributions in information sytems and technologies*, v. 353 , pp. 265-273. ISBN: 978 3 319 16485 4
https://doi.org/10.1007/978-3-319-16486-1_27

Castillo-Esparcia, Antonio (2012). "Investigación e investigadores. Las revistas científicas como instrumento de comunicación". *Vivat Academia. Revista de comunicación*, v. 14, n. 117, pp. 1002-1017.
https://doi.org/10.15178/va.2011.117E.1002-1017

Cohen, Jacob (1992). "Statistical power analysis". *Current directions in psychological science*, v. 1, n. 3, pp. 98-101.
https://doi.org/10.1111/1467-8721.ep10768783

Cronbach, Lee J. (1951). "Coefficient alpha and the internal structure of tests". *Psychometrika*, v. 16, n. 3, pp. 297-334.
https://doi.org/10.1007/BF02310555

Cummings, Jonathon N.; **Kiesler, Sara** (2005). "Collaborative research across disciplinary and organizational boundaries". *Social studies of science*, v. 35, n. 5, pp. 703-722.
https://doi.org/10.1177/0306312705055535

Davis, Fred D. (1989). "Perceived usefulness, perceived ease of use, and user acceptance of information technology". *MIS quarterly*, v. 13, n. 3, p. 319-340.
https://doi.org/10.2307/249008

Davis, Fred D.; **Bagozzi, Richard P.**; **Warshaw, Paul R.** (1992). "Extrinsic and intrinsic motivation to use computers in the workplace". *Journal of applied social psychology*, v. 22, n. 14, pp. 1111-1132.
https://doi.org/10.1111/j.1559-1816.1992.tb00945.x

Direito-Rebollal, Sabela; **Campos-Freire, Francisco** (2016). "Altmetrics: A measure of scientific impact on social networks". En: *11th Iberian conf on information systems and technologies (Cisti)*, pp. 1-5.
https://doi.org/10.1109/CISTI.2016.7521616

Dutton, William H. (2010). "Reconfiguring access in research: information, expertise and experience". En: Dutton, William H.; Jeffreys, Paul W. *World wide research: reshaping the sciences and humanities*, pp. 21-39. MIT Press Scholarship Online. ISBN: 978 0 262014397

Dutton, William H.; **Meyer, Eric T.** (2009). "Experience with new tools and infrastructures of research: An exploratory study of distance from, and attitudes toward, e-research". *Prometheus*, v. 27, n. 3, pp. 223-238.
https://doi.org/10.1080/08109020903127802

Faul, Franz; **Erdfelder, Edgar**; **Buchner, Axel**; **Lang, Albert-Georg** (2009). "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses". *Behavior research methods*, v. 41, n. 4, pp. 1149-1160.
https://doi.org/10.3758/BRM.41.4.1149

Faul, Franz; **Erdfelder, Edgar**; **Lang, Albert-Georg**; **Buchner, Axel** (2007). "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences". *Behavior research methods*, v. 39, n. 2, pp. 175-191.
https://doi.org/10.3758/BF03193146

García-Peñalvo, Francisco-José (2017). *Marco para la ciencia abierta*. Seminario.
https://doi.org/10.5281/zenodo.1069582

Gentil-Beccot, Anne; **Mele, Salvatore**; **Holtkamp, Annette**; **O'Connell, Heath B.**; **Brooks, Travis C.** (2009). "Information

resources in high-energy physics: Surveying the present landscape and charting the future course". *Journal of the Association for Information Science and Technology*, v. 60, n. 1, pp. 150-160.
https://doi.org/10.1002/asi.20944

Gibbons, Michael; **Limoges, Camille**; **Nowotny, Helga**; **Schwartzman, Simon**; **Scott, Peter**; **Trow, Martin** (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE Publications. ISBN: 978 0 803977945

Hara, Noriko; **Solomon, Paul**; **Kim, Seung-Lye**; **Sonnenwald, Diane H.** (2003). "An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration". *Journal of the Association for Information Science and Technology*, v. 54, n. 10, pp. 952-965.
https://doi.org/10.1002/asi.10291

Hayes, Andrew F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford. ISBN: 978 1 609182304

Liao, Chien-Hsiang (2011). "How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks". *Scientometrics*, v. 86, n. 3, pp. 747-761.
https://doi.org/10.1007/s11192-010-0309-2

Meyer, Eric T.; **Dutton, William H.** (2009). "Top-down e-infrastructure meets bottom-up research innovation: The social shaping of e-research". *Prometheus*, v. 27, n. 3, pp. 239-250.
https://doi.org/10.1080/08109020903127810

Murthy, Dhiraj; **Lewis, Jeremiah P.** (2015). "Social media, collaboration, and scientific organizations". *American behavioral scientist*, v. 59, n. 1, pp. 149-171.
https://doi.org/10.1177/0002764214540504

Myers, Teresa A. (2011). "Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data". *Communication methods and measures*, v. 5, n. 4, pp. 297-310.
https://doi.org/10.1080/19312458.2011.624490

Neylon, Cameron; **Wu, Shirley** (2008). "Open science: Tools, approaches and implications". En: *Biocomputing 2009*, pp. 540-544.
https://doi.org/10.1142/9789812836939_0051

Oller-Alonso, Martín; **Segarra-Saavedra, Jesús**; **Plaza-Nogueira, Alberto** (2012). "La presencia de las revistas científicas de ciencias sociales en los social media: de la Web 1.0 a la 2.0". *Index.comunicación: Revista científica en el ámbito de la comunicación aplicada*, v. 2, n. 1, pp. 49-68.
http://journals.sfu.ca/indexcomunicacion/index.php/indexcomunicacion/article/view/27/367

Ponte, Diego; **Simon, Judith** (2011). "Scholarly communication 2.0: Exploring researchers' opinions on web 2.0 for scientific knowledge creation, evaluation and dissemination". *Serials review*, v. 37, n. 3, pp. 149-156.
https://doi.org/10.1080/00987913.2011.10765376

Procter, Rob; **Voss, Alex**; **Asgari-Targhi, Marzieh** (2013). "Fostering the human infrastructure of e-research". *Information, communication & society*, v. 16, n. 10, pp. 1668-1691.
https://doi.org/10.1080/1369118X.2012.715667

Procter, Rob; **Williams, Robin**; **Stewart, James**; **Poschen, Meik**; **Snee, Helene**; **Voss, Alex**; **Asgari-Targhi, Marzieh** (2010). "Adoption and use of Web 2.0 in scholarly communications". *Philosophical transactions of the Royal Society A: Mathematical, physical, and engineering sciences*, v. 368, n. 1926, pp. 4039-56.
https://doi.org/10.1098/rsta.2010.0155

Quinlan, Stephen; **Gummer, Tobias**; **Roßmann, Joss**; **Wolf, Christof** (2017). "'Show me the money and the party!' – variation in Facebook and Twitter adoption by politicians". *Information, communication & society*, v. 21, n. 8, pp. 1-19.
https://doi.org/10.1080/1369118X.2017.1301521

Rigby, John; **Edler, Jakob** (2005). "Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality". *Research policy*, v. 34, n. 6, pp. 784-794.
https://doi.org/10.1016/j.respol.2005.02.004

Rúas-Araújo, José; **Campos-Freire, Francisco**; **Puentes-Rivera, Iván** (2016). "Utilización y valoración de las redes sociales generalistas y buscadores bibliográficos en las universidades gallegas". *Revista latina de comunicación social*, n. 71, pp. 1187-1207.
https://doi.org/10.4185/RLCS-2016-1141

Santana-Arroyo, Sonia (2011). "Redes de intercambio de información científica y académica entre los profesionales, en el contexto de la Web 2.0". *Revista cubana de información en ciencias de la salud*, v. 21, n. 3.
http://scielo.sld.cu/scielo.php?pid=S1024-94352010000300006&script=sci_abstract

Schiermeier, Quirin (2017). "Science publishers try new tack to combat unauthorized paper sharing". *Nature*, v. 545, n. 7653, pp. 145-146.
https://doi.org/10.1038/545145a

Stewart, James (2007). "Local experts in the domestication of information and communication technologies". *Information, communication & society*, v. 10, n. 4, pp. 547-569.
https://doi.org/10.1080/13691180701560093

Taiwo, Ayankunle; **Downe, Alan** (2013). "The theory of user acceptance and use of technology (Utaut): A meta-analytic review of empirical findings". *Journal of theoretical and applied information technology*, v. 49, n. 1, pp. 48-58.
http://www.jatit.org/volumes/Vol49No1/7Vol49No1.pdf

Venkatesh, Viswanath; **Morris, Michael G.**; **Davis, Gordon B.**; **Davis, Fred D.** (2003). "User acceptance of information technology: Toward a unified view". *MIS quarterly*, v. 27, n. 3, pp. 425-478.
https://doi.org/10.2307/30036540

Venkatesh, Viswanath; **Thong, James Y. L.**; **Xu, Xin** (2012). "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use technology". *MIS quarterly*, v. 36, n. 1, pp. 157-172.

Voss, Alex; **Asgari-Targhi, Marzieh**; **Procter, Rob**; **Fergusson, David** (2010). "Adoption of e-infrastructure services: configurations of practice". *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, v. 368, n. 1926, pp. 4161-4176.
https://doi.org/10.1098/rsta.2010.0162

Waldrop, Mitchell (2008). "Science 2.0. Is open access science the future? Is posting raw results online, for all to see, a great tool or a great risk?". *Scientific American*, v. 298, n. 5, pp. 68-73.

Participación digital del público en la ciencia de excelencia española: análisis de los proyectos financiados por el *European Research Council*

Digital public engagement in excellent Spanish science: Analysis of research projects funded by *European Research Council*

Lourdes López-Pérez; María-Dolores Olvera-Lobo

Cómo citar este artículo:

López-Pérez, Lourdes; **Olvera-Lobo, María-Dolores** (2019). "Participación digital del público en la ciencia de excelencia española: análisis de los proyectos financiados por el *European Research Council*". *El profesional de la información*, v. 28, n. 1, e280106.

https://doi.org//10.3145/epi.2019.ene.06

Artículo recibido el 17-08-2018
Aceptación definitiva: 04-12-2018

Lourdes López-López ✉
https://orcid.org/0000-0002-5787-1531

Universidad de Granada
Facultad de Comunicación y Documentación
Colegio Máximo de Cartuja.
Campus Universitario de Cartuja.
18071 Granada, España
lourdes.lpez@gmail.com

María-Dolores Olvera-Lobo
https://orcid.org/0000-0002-0489-7674

Universidad de Granada
Facultad de Comunicación y Documentación
Colegio Máximo de Cartuja.
Campus Universitario de Cartuja.
18071 Granada, España
molvera@ugr.es

Resumen

Se analiza cómo la ciencia excelente usa las aplicaciones web 2.0 para fomentar la participación del público en el proceso de investigación. Se entiende participación desde la perspectiva de la investigación e innovación responsables, no sólo como la comunicación de los resultados científicos, sino también como la implicación de la sociedad en el desarrollo científico. Para el análisis se ha diseñado y validado una encuesta *ad hoc* mediante el método Delphi. La muestra de estudio está constituida por los proyectos españoles financiados por el *European Research Council* en 2015. Entre los resultados obtenidos destaca el reducido uso de las herramientas digitales. De hecho, sólo el 23,9% de los proyectos tiene web y menos de un 15% perfiles sociales. En cuanto a los mecanismos de participación, un limitado 3% utiliza la web 2.0 para implicar a los ciudadanos. Sobre las características de la comunicación se puede concluir que es unidireccional desde la institución-responsable del proyecto hacia la sociedad –97% de los casos– y presenta un alto nivel de especialización –en más del 80% de los casos se utiliza un lenguaje técnico y los mensajes publicados están dirigidos a expertos–.

Palabras clave

Comunicación científica; Comunicación pública de la ciencia; Participación; Medios sociales; Redes sociales; Internet; Investigación e innovación responsables; RRI; Proyectos de investigación; Método Delphi.

Abstract

How much the excellent science uses web 2.0 tools to encourage public engagement in the research process is analyzed. Understanding participation, from the perspective of responsible research and innovation, not only as the communication of scientific results, but also the involvement of society in scientific development. For the analysis, an ad-hoc questionnaire developed using the Delphi method was designed and validated. The study sample is constituted by Spanish projects financed by the *European Research Council* in 2015. Among the results obtained, the reduced use of digital applications stands out. In fact, only 23.9% of projects have a website and less than 15% have a social profile. Regarding participation mechanisms, a limited 3% uses web 2.0 to involve citizens. With reference to the characteristics of the

communication, it can be concluded that it is unidirectional from the institution responsible for the project to society -97% of the cases- and presents a high level of specialization -in more than 80% of the cases a technical language is used and the messages posted are aimed solely at experts-.

Keywords

Science communication; Public communication in science; Participation; Social networks; Social media; Internet; Responsible research and innovation; RRI; Research projects; Delphi method.

1. Introducción

La transformación de la relación entre ciencia y sociedad se ha reflejado en la bibliografía académica como un cambio de objeto de estudio, influido por el acercamiento de ambas en las dos últimas décadas. Se ha trasladado el foco de interés de la alfabetización científica del público a su implicación en el proceso científico; de la comunicación al diálogo; de la mera unión de ciencia y sociedad a una ciencia hecha con y para la sociedad; y del modelo de déficit cognitivo a la implicación de la sociedad en el proceso científico (**Marschalek**, 2017). Así hasta llegar al actual modelo de investigación e innovación responsable propiciado por la *Comisión Europea* (2014), que defiende que los procesos y los resultados científicos estén alineados con los valores, necesidades y expectativas de la sociedad.

Este cambio se ha producido en gran medida por la irrupción de internet y las herramientas de la web 2.0, que han facilitado como ningún otro medio la comunicación multidireccional (**Weitze**, 2017), construyendo un espacio conversacional (**Grand** *et al.*, 2016; **Brown**, 2016) que estimula una ciudadanía más activa (**Coleman**, 2001).

Los trabajos que estudian el impacto de internet y los instrumentos digitales en la comunicación pública de la ciencia son todavía exiguos y la bibliografía científica sobre el uso y la eficacia de este nuevo canal para implicar a los ciudadanos en el proceso científico es casi inexistente. Esto se debe a la juventud del modelo de participación del público en la ciencia, así como a la propia naturaleza cambiante de la Red.

Este trabajo pretende contribuir a generar una línea académica en este sentido. Para ello se ha analizado el uso que los proyectos de investigación españoles financiados por el *European Research Council* (*ERC*) hacen de los medios digitales para implicar a la sociedad en el proceso científico, así como las características y efectividad de ese uso. La muestra está constituida por los proyectos españoles que consiguieron financiación del *ERC* en 2015.

Seguidamente se presenta el estado de la cuestión, así como la metodología aplicada, que se sustenta en un cuestionario diseñado *ad hoc*, elaborado y validado mediante el método Delphi. A continuación se detallan los resultados obtenidos, que apuntan al escaso uso de las herramientas digitales y a la existencia de una comunicación unidireccional y dirigida fundamentalmente a expertos, y las conclusiones que de éstos se derivan.

2. Modelo de participación del público en la ciencia desde la dimensión de la investigación e innovación responsables

> “La comunicación que los responsables de los proyectos realizan a la sociedad es unidireccional y utiliza un lenguaje técnico”

Para comprender el modelo de participación del público en la ciencia desde la dimensión de la investigación e innovación responsables (*Responsible Research and Innovation*, RRI) es necesario contextualizar las implicaciones de este sistema de I+D+i. Aunque la bibliografía científica ha planteado múltiples aproximaciones a este concepto, la que propone la *Comisión Europea* (2014) sintetiza los aspectos anteriormente propuestos.

Desde esta perspectiva la investigación e innovación responsable es aquella en la que los actores sociales trabajan juntos durante el curso de la investigación con el objetivo de alinear los procesos y sus resultados con los valores, necesidades y expectativas de la sociedad. Se trata de un término amplio e integrado por seis dimensiones: participación del público en la ciencia, acceso abierto, igualdad de género, educación científica, ética y gobernanza (*Comisión Europea*, 2015).

Académicos de la comunicación pública de la ciencia han contribuido a definir el modelo de la participación del público en la ciencia desde diferentes perspectivas. **Rowe** y **Frewer** (2005) consideran la implicación del público en el proceso científico como una combinación entre la comunicación, la consulta y la participación del público en el marco de la investigación y la innovación. Por su parte, **Ravn** y **Mejlgaard** (2014) parten de la categorización de **Rowe** y **Frewer** (2005) y plantean una clasificación que engloba las iniciativas de participación del público como: comunicación pública, activismo público, consulta pública y deliberación pública.

Otros académicos como **Bonney** *et al.* (2009) definen la participación del público ciñéndola a las etapas del proceso de investigación e innovación en las que los ciudadanos pueden ser partícipes.

Este concepto también se ha asociado con la implicación de los investigadores en la comunicación de los resultados científicos (**Bauer**; **Jensen**, 2011). Conferencias públicas, entrevistas en medios de comunicación, redacción de libros de

divulgación, participación en debates públicos, o colaboración con organizaciones no gubernamentales, son algunas de las actividades integradas en esta definición.

3. Internet como canal de interacción entre ciencia y sociedad

El desarrollo del modelo de participación pública de la ciencia ha transcurrido de forma paralela a la consolidación de internet como uno de los principales canales de comunicación de la sociedad contemporánea (**Acord**; **Harley**, 2013; **Castells**, 2001; 2011; **Middaugh**; **Kahne**, 2013; **Papacharissi**, 2002). En la propia naturaleza de la web 2.0 están implícitos los conceptos de interacción y conversación (**Grand** *et al.*, 2016; **Brown**, 2016; **Könneker**; **Lugger**, 2013; **Weilgod**; **Treise**, 2004; **Flores-Vivar**, 2009; **Castells**, 2001) indispensables en la implementación de investigación e innovación responsable.

Las ventajas evidenciadas del entorno digital son coincidentes en la bibliografía científica publicada. Varios autores han establecido como principales beneficios de la web 2.0 en la comunicación pública de la ciencia:

- creación de un espacio conversacional (**Grand** *et al*., 2016; **Brown**, 2016; **Weilgod**; **Treise**, 2004);
- visibilización de la producción científica (**López-Goñi**; **Sánchez-Agudo**, 2018; **Wolf**, 2017; **Shuai**; **Pepe**; **Bolen**, 2012);
- fomento de debate y discusión sobre controversias científicas (**Ke**; **Ahn**; **Sugimoto**, 2017);
- transformación de la comunicación unidireccional a la bidireccional (**Weitze**, 2017; **Brossard**; **Schefeule**, 2013; **Weilgod**; **Treise**, 2004).

El interés académico en la dimensión digital de la comunicación pública de la ciencia como objeto de estudio es incipiente y aún es más exiguo en lo que respecta a la participación del público en la ciencia. Los trabajos académicos que han abordado la evaluación de la implicación del público en la ciencia se han centrado en la faceta convencional a través del diseño de sistemas de evaluación (*Comisión Europea*, 2015; 2018; **Meijer** *et al*., 2016) que no integran la esfera digital como objeto de análisis. Sólo un número limitado de estudios (**López-Pérez**; **Olvera-Lobo**, 2018a; 2018b; **Neresini**; **Bucchi**, 2011) han abordado la validación científica de indicadores que permitan estudiar el papel de las redes sociales y de la Web en la participación del público en la ciencia.

Según lo expuesto, a pesar de las cualidades de internet como canal para favorecer la participación ciudadana en la ciencia, no hay resultados empíricos que demuestren si este nuevo canal ha reducido la brecha entre científicos y sociedad, una pregunta de investigación a cuya respuesta pretende contribuir el presente trabajo.

4. Metodología

Con el fin de obtener criterios e indicadores validados consensuadamente para analizar la implicación digital del público en la ciencia de excelencia, se ha aplicado el método Delphi (**Osborne** *et al*., 2003; **Clayton**, 1997; **Murry**; **Hammons**, 1995). Éste ha sido utilizado anteriormente para el diseño de propuestas metodológicas de análisis de la comunicación de la ciencia (**Ouarichi**; **Olvera-Lobo**; **Pérez-Gutiérrez**, 2017; **Seakins**; **Dillon**, 2013) y de la educación científica (**Smith**; **Simpson**, 1995; **Blair**; **Uhl**, 1993). Se trata de un proceso sistemático, interactivo y grupal encaminado a la obtención de opiniones y consenso, a partir de las experiencias y juicios subjetivos de un grupo de expertos (**Scapolo**; **Miles**, 2006; **Osborne** *et al.*, 2003).

Para la aplicación del Delphi en este trabajo, la selección de expertos se ha realizado teniendo en cuenta cuatro aspectos:

- área de conocimiento afín a la materia abordada que los acredita para validar el cuestionario;
- relevancia en el área;
- igualdad de género;
- diversidad geográfica.

De este modo, las áreas de conocimiento han sido:

- comunicación pública de la ciencia;
- educación científica;
- investigación e innovación responsables;
- científicos blogueros;
- redes sociales.

Los criterios para medir la relevancia en el área han sido:

- calidad y cantidad de publicaciones científicas;
- experiencia profesional y académica en el área;
- impacto social (este ítem se ha fijado principalmente en el caso de los científicos blogueros y es relativo al número de seguidores);
- formación;
- coordinación y organización de proyectos internacionales que implican la participación del público en el proceso de investigación o que están vinculados con la investigación e innovación responsables.

También se han seguido valores como la igualdad de género y de los 14 participantes, 8 fueron mujeres y 6 hombres.

Asimismo, se consideró relevante la procedencia geográfica haciendo especial hincapié en aquellos países, además de España, en los que se están realizando importantes esfuerzos para la implementación de la RRI, como Reino Unido, Alemania y Dinamarca, y que por tanto poseen un mayor conocimiento de las ventajas y desventajas de este modelo.

Tabla 1. Expertos método Delphi

Área	Experto	Origen
Comunicación pública de la ciencia	Profesor de comunicación científica	Reino Unido
	Técnica en participación pública de la ciencia	Reino Unido
	Periodista científico	España
Educación científica	Profesora de educación científica	Alemania
	Educadora científica en un museo de ciencia	Alemania
	Educadora científica en un museo de ciencia	España
Investigación e innovación responsables (RRI)	Experta en gestión de proyectos RRI	España
	Experta en humanidades digitales y RRI	España
Blogueros especializados en ciencia	Bloguero especializado en ciencia	España
	Bloguero especializado en ciencia	España
Comunicación pública de la ciencia en redes sociales	Profesora de medios interactivos	Dinamarca
	Profesor de comunicación en social media	España
	Experto en comunicación pública de la ciencia a través de los social media	España
	Experta en comunicación pública de la ciencia a través de los social media	España

El aspecto clave en el desarrollo de la metodología de trabajo ha sido conseguir el consenso del grupo, pero con la máxima autonomía por parte de los participantes. Para ello se han realizado tres rondas de consultas en un proceso interactivo y anónimo que ha permitido a los participantes opinar, recibir las conclusiones del resto del grupo en cada una de las rondas y, finalmente, reconsiderar sus opiniones en una última fase. Los criterios e indicadores de evaluación establecidos por consenso son los que se muestran en la tabla 2.

Tabla 2. Criterios de evaluación validados por el método Delphi

Criterios	Indicadores
Identificación	Nombre del proyecto de investigación
	Institución responsable
	Fecha de comienzo y cierre del proyecto
	Transdisciplinariedad
	Países implicados en el proyecto
Tipos de herramientas online	Webs. Existencia de una web específica para los proyectos de investigación
	Redes sociales (*Facebook*, *Twitter*, *YouTube* e *Instagram*)
	Blogs. Uso de blogs para la comunicación de los resultados científicos o para promover la participación ciudadana
	Apps. Uso de aplicaciones para recopilar información u opiniones del público
Categoría de participación	Comunicación
	Consulta
	Participación
	Co-creación
Mecanismos de participación	Encuestas
	Materiales educativos
	Consultas. Recopilar opiniones a favor o en contra en alguna de las fases del proceso de investigación
Características de la participación	Facilidad de acceso a la web y perfiles sociales
	Nivel de comunicación (uni o bidireccional)
	Dirección de la comunicación (científico-público/público-científico/científico-científico/público-público)
	Lenguaje utilizado en la comunicación
	Tipo de público
	Expertos implicados
	Presentación de suficiente información
	Comunicación de los resultados de la participación
	Dirección de correo electrónico para contacto del público
	Tipo de información presentada al público
	Acceso abierto
	Tema de discusión
Intensidad de la participación	Número de seguidores en redes sociales
	Número de personas implicadas en cada categoría de la participación
	Cuantificación de la participación
	Demanda de la información por parte del público

En la primera ronda los expertos evaluaron en una escala Likert (1 = importancia baja, 2 = importancia media, o 3 = importancia alta) un conjunto de criterios e indicadores basados en las definiciones y aproximaciones al concepto de participación del público en la ciencia planteadas por la bibliografía científica, y se añadieron tres preguntas abiertas que perseguían:

I) evaluar la idoneidad del objeto de estudio –¿considera que internet y las herramientas online son un buen canal para promover la participación del público en el desarrollo científico y tecnológico?-, o

II) mejorar y ampliar los criterios e indicadores propuestos por el grupo coordinador –¿qué criterios y/o indicadores añadiría para la mejora de un sistema de evaluación de la participación del público a través de herramientas online?, ¿qué criterios y/o indicadores eliminaría para la mejora de un sistema de participación del público a través de las herramientas online?–.

Con el fin de incluir las aportaciones de los expertos a través de las preguntas abiertas y someterlas al consenso del grupo, el cuestionario enviado en la segunda ronda incluyó los indicadores propuestos e integró aquellos que en la primera ronda no habían alcanzado una valoración promedio de 2 puntos sobre 3, que fue lo establecido como consenso para su inclusión. El objetivo de incluir estos indicadores fue someter a una segunda reflexión la consideración de los mismos, tal y como marca el método Delphi, antes de eliminarlos definitivamente del cuestionario.

A partir de las respuestas obtenidas, se diseñó la tercera versión del cuestionario, en el que se incluyeron los indicadores que habían logrado consenso en la primera y segunda ronda. En este caso el objetivo se centró en comprobar la estabilidad en las respuestas entre los cuestionarios 1, 2 y 3, e integrar de forma definitiva aquellos indicadores que hubiesen alcanzado un consenso superior a una valoración promedio de 2 puntos.

El proceso, que se ha descrito en trabajos anteriores (**López-Pérez**; **Olvera-Lobo**, 2018a; 2018b) concluyó al alcanzarse el criterio de saturación establecido por el consenso y la estabilidad de las valoraciones de los expertos de los indicadores incluidos en el cuestionario.

5. Muestra de estudio

La muestra de estudio está constituida por los proyectos financiados por el *European Research Council* a través de sus convocatorias *Starting grant*; *Consolidator grant*; *Advanced grant*; *Proof of concept* y *Synergy grant*. En las cuatro primeras se han seleccionado los proyectos españoles que recibieron financiación en 2015 y en el caso de las *Synergy grants* se han integrado los beneficiarios de 2013 por ser la última convocatoria publicada de este programa. Suman un total de 67 proyectos.

La elección de los proyectos financiados por el *ERC* responde a tres criterios:

- excelencia: el organismo europeo se rige por este valor para destinar sus recursos;
- temporalidad: se han seleccionado los financiados en 2015 para no sesgar los resultados de participación y comunicación en proyectos que hayan sido financiados recientemente y se encuentren en su fase inicial. En los de 2015 las investigaciones se encuentran en una fase intermedia que ya ha permitido el desarrollo de estrategias de difusión e implicación del público en el proceso científico. La recogida de datos se llevó a cabo durante el mes de abril de 2018;
- coherencia: se refiere al hecho de que la Unión Europea ha sido la impulsora de la implementación de la investigación e innovación responsables (RRI) dentro del programa *Horizon 2020* y, por coherencia, los proyectos que reciben su financiación deberían responder a los valores de la RRI, entre los que se integra la participación del público en el proceso de investigación.

6. Resultados

Los resultados se presentan según los diferentes criterios considerados:

6.1. Identificación

> Sólo un 23,9% de los proyectos analizados tiene web propia

El hecho de que sólo un 23,9% de los proyectos de investigación analizados tenga web ha propiciado que el análisis del criterio de identificación se extienda más allá de los medios de comunicación específicamente diseñados para la difusión de los proyectos. La información más detallada de los proyectos aparece publicada en la web del *ERC*, en la que en el 100% de los casos se publica el nombre en inglés del proyecto y se identifica la entidad beneficiaria, así como el investigador principal.

Las fechas de inicio y conclusión de los proyectos también se detallan en la web del *ERC*. El 76,1% de los proyectos comenzó en 2016, un 13,7% en 2017, un 11,5% en 2015 y un 5,8% en 2014. En cuanto a su conclusión, un 28,2% ha finalizado antes de 2018, un 4,4% lo hará entre 2019 y 2020 y un 67,4% en 2021. El hecho de que la mayoría de los proyectos esté en el estadio intermedio de ejecución incita a la repetición del análisis a su conclusión para determinar la evolución del uso de medios de participación del público en las distintas fases de los procesos de investigación.

La transdisciplinariedad se reseña en el 19,4 % de los casos. Por último, el indicador relativo a países implicados, en todos los proyectos se indica únicamente el país de la institución beneficiaria, en este caso España.

6.2. Tipos de medios online

Gráfico 1. Tipos de medios online

Los proyectos de investigación analizados hacen un escaso uso de los sistemas online en las dos dimensiones de la participación pública en la ciencia, es decir, tanto en lo que se refiere a la comunicación como en lo relacionado con la interacción-implicación de los ciudadanos. Esto es algo que se refleja en los reducidos porcentajes obtenidos del uso de los tipos de herramientas. Como muestra el gráfico 1, un 23,9% de los proyectos analizados cuentan con una web, el medio más usado, seguido de *Twitter*: un 13,4% tiene cuenta en esta red de microblogging. El uso de *YouTube*, *Facebook* y de blogs es aún más escaso con un 9%, un 7,5% y un 6% respectivamente. Las herramientas más minoritarias son *Instagram* y otras apps, que comparten un 1,5%.

Los resultados evidencian además una preponderancia de los canales de comunicación unidireccionales sobre los que permiten interacción y diálogo. Este escenario sugiere que el papel del público en la dimensión digital de la ciencia sigue siendo de usuario final y no de participante activo.

6.3. Categoría de participación

> La implicación de los ciudadanos en el proceso del proceso de investigación a través de la web 2.0 es casi inexistente

Los proyectos analizados no fomentan el uso de internet y de las herramientas web 2.0 como canales de interacción con la sociedad. Un 12,1% de los proyectos emplea estrategias de comunicación unidireccional (difusión de información sin diálogo o interacción) y, dentro de éstos, solo el 9% cuenta con un canal de noticias para divulgar sus principales resultados.

En cuanto a la implicación activa del público en el proceso de investigación, en las categorías de consulta y cocreación esta implicación es inexistente. Sólo un 3% de los proyectos analizados emplea alguna de las herramientas digitales analizadas para hacer posible la participación de los ciudadanos en la recogida de datos para el proyecto o a través de la recopilación de ideas sobre el mismo.

6.4. Mecanismos de participación

Un 9% de los proyectos utiliza materiales educativos y un 3% mecanismos de consulta. El resto de items contemplados en el cuestionario de evaluación no son utilizados en ningún caso. Los recursos educativos más utilizados son la publicación de conferencias, documentales o vídeos online para visibilizar sus logros y líneas de investigación.

6.5. Características de la participación

El fácil acceso a su web o a los perfiles sociales creados para el proyecto indudablemente facilitan su visibilización y difusión. La facilidad de acceso se ha evaluado con una puntuación de 0 a 5, siendo 0 muy difícil y 5 muy fácil. Como se observa en el gráfico 2, el estudio muestra que el 53,7% de las páginas web son de muy difícil acceso, es decir, no están ubicadas en sitios visibles como las páginas principales de las instituciones beneficiarias, el 19,4% son de difícil acceso y tan sólo un 26,9% están publicitadas en espacios fácilmente accesibles (gráfico 2). En cuanto a los perfiles sociales, en el 82,8% de los casos están publicados en espacios no visibles o de difícil acceso.

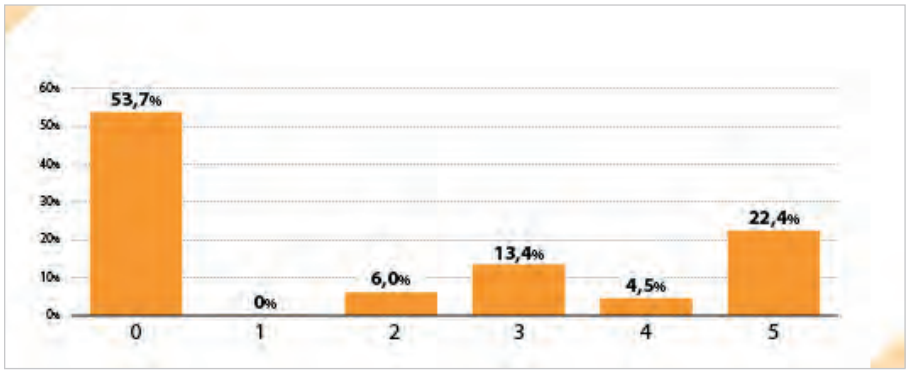


Gráfico 2. Facilidad de acceso a la web

Aunque la web 2.0 se caracteriza por ser un espacio conversacional que facilita el diálogo y la interacción, el 97% de la comunicación que se realiza en las redes sociales analizadas es unidireccional, y se produce de arriba abajo, es decir desde la institución-responsable del proyecto hacia la sociedad.

El lenguaje utilizado en las comunicaciones es técnico en el 80,6% de los casos, lo que probablemente no contribuye a potenciar el acercamiento al ciudadano medio. Esta especialización en el contenido hace que la dirección de la comunicación sea en un 83,6% de los casos de científico a científico y el 16,4% científico-público (gráfico 3). En cuanto al tipo de

público no es posible definirlo al no existir comentarios de usuarios en las redes analizadas que nos permitan inferir sus características.

El 93,8% de los proyectos que sí utilizan las herramientas de la web 2.0 para interaccionar con el público, realiza una comunicación institucional frente a la personal elegida por un 6,2% de los proyectos (gráfico 4).

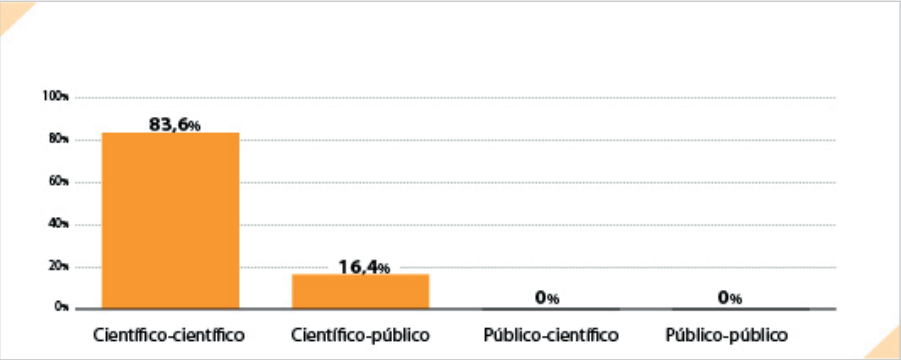


Gráfico 3. Dirección de la comunicación

En lo que respecta a la información sobre el proyecto como sus objetivos o líneas de investigación, el 76,1% publica una información insuficiente (entendida la información del proyecto como aquella que incluye el título, objetivos, metodología, período de ejecución, líneas de investigación a desarrollar y presupuesto) frente al 23,9% que sí ofrece una información completa.

La publicación de un correo electrónico de contacto se da en un 22,4% de los casos, frente al 77,6% que no integra esta información.

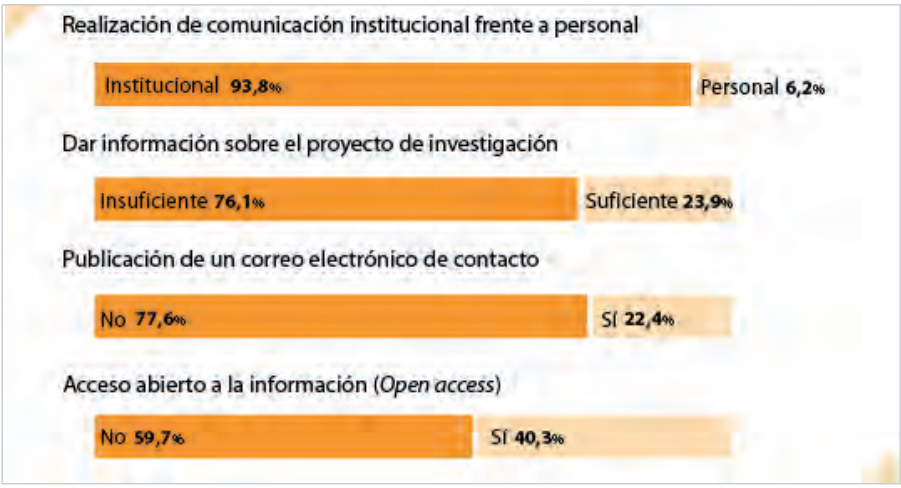


Gráfico 4. Características de la participación

Con respecto a la publicación de resultados científicos en abierto, tal y como puede observarse en el gráfico 4, el 40,3% de los proyectos con website publica en abierto sus resultados, frente al 59,7% que no lo hace (gráfico 4).

Finalmente, en referencia a la información sobre los proyectos que los beneficiarios publican online, en el 65,6% de los casos se publica un resumen del proyecto. El 57,1% de los proyectos se publican en la web del grupo de investigación, el 40,5% en la sección de investigación de la web de la institución beneficiaria de la financiación del *ERC* y sólo en un 2,4% la información es publicada en los portales de transparencia de estas mismas instituciones (gráfico 5).

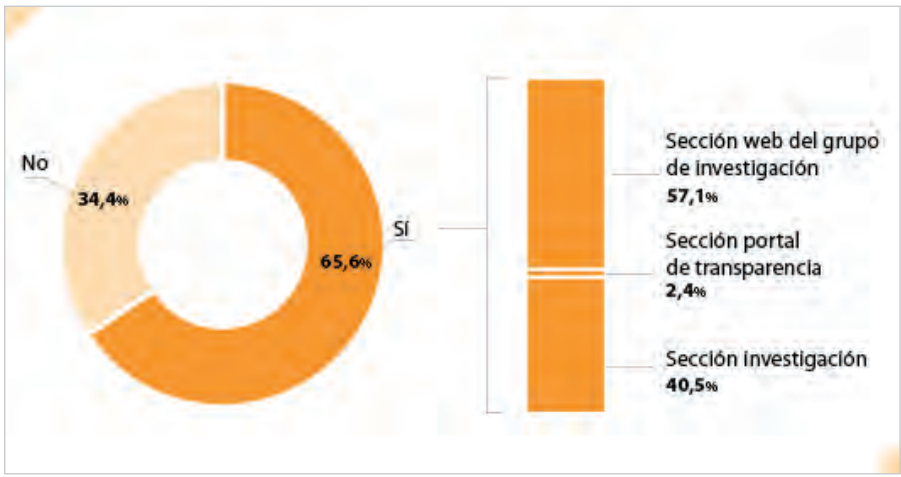


Gráfico 5. ¿Publican la información del proyecto en la web de la institución?

6.6. Intensidad de la participación

La red social que presenta mayor intensidad de participación en términos de seguidores y posts publicados es *Twitter*: el 9,1% de perfiles supera los 5.000 seguidores, otro 9,1% tiene entre 1.000 y 5.000 seguidores y el 81,8% restante tiene menos de mil. En el caso de *Facebook* las cifras son similares, el 80% cuenta con menos de 1.000 seguidores mientras que el 20% tiene más de 5.000 y ninguno entre 1.000 y 5.000.

En lo que respecta a la publicación y actualización de contenidos, vuelve a ser *Twitter* la red con mayor actividad: un 22,2% de los proyectos con perfil en *Twitter* publicó más de 50 tweets en un mes mientras que en el 77,8% de los casos

no se superan los 20 tweets mensuales (no hay perfiles con publicaciones entre 21 y 49 tweets). En el caso de *Facebook* el 50% de los proyectos publicó menos de 10 posts y el otro 50% entre 10 y 24. En *YouTube*, el 60% publicó 3 vídeos en el período analizado, un 20% 1 vídeo y el 20% restante ninguno. El único proyecto con presencia en *Instagram* no publicó ningún post en el período analizado.

> Casi el 60% de los proyectos de investigación analizados no tiene información publicada en abierto

7. Conclusiones

La ciencia de excelencia española aún no ha encontrado en internet un canal útil de interacción con los ciudadanos. Esto es lo que sugieren los resultados obtenidos sobre el uso de la web 2.0. Sólo un 23,9% cuenta con una web, y menos de un 15% ha creado perfiles en redes como *Facebook*, *Twitter* o *YouTube*. Además, cuando tienen un perfil la repercusión de las informaciones publicadas en el mismo tienen muy poco impacto por el escaso número de seguidores, y es que el 80% de los perfiles tiene menos de 5.000 usuarios. Por otro lado destaca la escasa publicación de contenidos en estas redes –el 50% de los proyectos publica menos de 10 posts mensuales–.

Un aspecto que podría explicar el reducido número de seguidores es el difícil acceso tanto a sus páginas web como a los perfiles sociales. Es destacable el hecho de que sólo el 40,5% de las instituciones beneficiarias publique información del proyecto y de su financiación en sus páginas principales, y que un reducido 2,4% integre esta información en sus portales de transparencia. La información más completa se publicó en más de un 50% de los casos en las páginas propias de los grupos de investigación, un resultado que confirma las conclusiones ya apuntadas por **Pacios**, **Vianello-Osti** y **Rodríguez-Bravo** (2016) en un trabajo sobre transparencia y acceso a la información de los proyectos de investigación en las universidades públicas españolas.

Pese a que entre las principales características de la web 2.0 destacan la interactividad, el diálogo y la participación, sólo el 3% de los proyectos utiliza los canales digitales con este objetivo, y en un 77,6% de las webs no se incluye siquiera una dirección de correo electrónico para contactar con los miembros del equipo investigador. Esta circunstancia demuestra que la implementación de la dimensión digital de la participación pública dentro la investigación e innovación responsable en España está aún en una fase incipiente y que no está asumida por los investigadores. El 97% de la comunicación que se realiza en las redes sociales analizadas es unidireccional, y se produce de arriba abajo, es decir desde la institución-responsable del proyecto a la sociedad.

Otra dimensión de la RRI que aún no está consolidada, según los resultados obtenidos, es el acceso abierto a la investigación y sus resultados. El 60% de los proyectos no publica información en abierto sobre sus resultados. Aunque este dato habría que contrastarlo una vez finalicen los proyectos (el 67,4% concluirá en 2021) ya que en muchos de ellos los resultados relevantes posiblemente se obtengan en la última fase del proceso.

Según se observa, en el entorno digital los equipos de investigación están más preocupados por la difusión de la ciencia entre científicos que por la comunicación con el público. En más del 80% de los casos utilizan lenguaje técnico y dirigen sus mensajes a los expertos. La comunicación es fundamentalmente institucional y hay escasa personalización de los mensajes, a pesar de que existen evidencias científicas de que en el caso de los medios sociales la autenticidad y la conexión personal se han establecido como criterios adicionales para el fomento de la credibilidad de la información (**Betsch** *et al*., 2010).

Por otro lado, el análisis realizado destaca la importancia marginal que la participación del público tiene en la ciencia de excelencia en España y la necesidad de trabajar en este sentido, tanto para su implementación como para su evaluación. Por esto, aunque este estudio se ha concentrado principalmente en determinar los esfuerzos que lleva a cabo la ciencia excelente por involucrar a los ciudadanos más que en el estudio de la calidad de esos esfuerzos, los indicadores validados pueden contribuir al desarrollo de análisis centrados en evaluar la efectividad de la participación digital del público en la ciencia en nuestro país.

8. Referencias

Acord, Sophia-Krzys; **Harley, Diane** (2013). "Credit, time and personality: The human challenges to sharing scholarly". *New media & society*, v. 15, n. 3, pp. 379-397.
https://doi.org/10.1177/1461444812465140

Bauer, Martin W.; **Jensen, Pablo** (2011). "The mobilization of scientists for public engagement". *Public understanding of science*, v. 20, n. 1, pp. 3-11
https://doi.org/10.1177/0963662510394457

Betsch, Cornelia; **Renkewitz, Frank**; **Betsch, Tilmann**; **Ulshöfer, Corina** (2010). "The influence of vaccine-critical websites on perceiving vaccination risks". *Journal of health psicology*, n. 15, v. 3, pp. 446-455.
https://doi.org/10.1177/1359105309353647

Blair, Shirley; **Uhl, Norman P.** (1993). "Using the Delphi method to improve the curriculum". *The Canadian journal of higher education*, v. 23, n. 3, pp. 107-128.

http://journals.sfu.ca/cjhe/index.php/cjhe/article/view/183175

Bonney, Rick; **Ballard, Heidi**; **Jordan, Rebecca**; **McCallie, Ellen**; **Phillips, Tina**; **Shirk, Jennifer**; **Wilderman, Candie C.** (2009). *Public participation in scientific research: Defining the field and assessing its potential for informal science education.* A Caise Inquiry Group Report. Washington, D.C., EUA: Center for Advancement of Informal Science Education (Caise).
http://www.birds.cornell.edu/citscitoolkit/publications/CAISE-PPSR-report-2009.pdf

Brossard, Dominique; **Scheufele, Dietram** (2013). "Science, new media, and the public". *Science*, v. 339, n. 6115, pp. 40-41.
https://doi.org/10.1126/science.1232329

Brown, David J. (2016). *Access to scientific research. Challenges facing communications in STM*. Berlin: Walter de Gruyter. ISBN: 978 3 110376 16 9
https://doi.org/10.1002/asi.23764

Castells, Manuel (2001). *La galaxia de internet*. Barcelona: Areté. ISBN: 978 8 401386 06 0

Castells, Manuel (2011). *Comunicación y poder*. Madrid: Alianza Editorial. ISBN: 978 84 20684 99 4

Clayton, Mark (1997). "Delphi: a technique to harness expert opinion for critical decision-making tasks in education". *Educational psychology*, n. 17, v. 4, pp. 373-386.
https://doi.org/10.1080/0144341970170401

Coleman, Stephen (2001). "The transformation of citizenship?" In: Axford, Barrie; Huggins, Richard (eds.). *New media and politics*. London: SAGE, pp. 109-126. ISBN: 978 0 761962007
https://doi.org/10.4135/9781446218846.n5

Comisión Europea (2014). *Responsible research and innovation — Europe's ability to respond to societal challenges*.
https://ec.europa.eu/research/swafs/pdf/pub_rri/KI0214595ENC.pdf

Comisión Europea (2015). *Indicators for promoting and monitoring responsible research and innovation*.
http://ec.europa.eu/research/swafs/pdf/pub_rri/rri_indicators_final_version.pdf

Comisión Europea (2018). *Monitoring the evolution and benefits of responsible research and innovation*.
https://www.rri-tools.eu/-/morri-monitoring-the-evolution-and-benefits-of-responsible-research-and-innovation

Flores-Vivar, Jesús-Miguel (2009). "Nuevos modelos de comunicación, perfiles y tendencias en las redes sociales". *Comunicar*, v. 17, n. 33, pp. 73-81.
https://doi.org/10.3916/c33-2009-02-007

Grand, Ann; **Holliman, Richard**; **Collins, Trevor**; **Adams, Anne** (2016). "'We muddle our way through': Shared and distributed expertise in digital engagement with research". *JCOM*, v. 15, n. 4, A05.
https://doi.org/10.22323/2.15040205

Ke, Quing; **Ahn, Yong-Yeol**; **Sugimoto, Cassidy R.** (2017). "A systematic identification and analysis of scientists on Twitter". *PLoS one*, v. 12, n. 4.
https://doi.org/10.1371/journal.pone.0175368

Könneker, Carsten; **Lugger, Beatrice** (2013). "Public science 2.0 - back to the future". *Science*, v. 342, n. 6154, pp. 49-50.
https://doi.org/10.1126/science.1245848

López-Goñi, Ignacio; **Sánchez-Angulo, Manuel** (2018). "Social networks as a tool for science communication and public engagement: focus on Twitter". *FEMS Microbiology letters*, v. 365, n. 2.
https://doi.org/10.1093/femsle/fnx246

López-Pérez, Lourdes; **Olvera-Lobo, María-Dolores** (2018a). "Public engagement in science via web 2.0 technologies. Evaluation criteria validated using the Delphi method". *JCOM*, v. 17, n. 2.
https://doi.org/10.22323/2.17020208

López-Pérez, Lourdes; **Olvera-Lobo, María-Dolores** (2018b). "Criterios para la evaluación de la implicación del público en la ciencia a través de la web 2.0". En: Romero-Frías, Esteban; Bocanegra-Barbecho, Lidia (eds.). *Ciencias sociales y humanidades digitales aplicadas. Casos de estudio y perspectivas críticas*. Granada: Universidad de Granada; Downhill Publishing (NY), 2018. ISBN: 978 84 338 6318 8
http://digibug.ugr.es/handle/10481/54383

Marschalek, Ilse (2017). *Public engagement in responsible research and innovation. A critical reflection from the practitioner's point of view*. Wien: University of Wien.
https://www.zsi.at/object/publication/4498/attach/Marschalek_Public_Engagement_in_RRI.pdf

Meijer, Ingeborg; **Mejlgaard, Niels**; **Woolley, Richard**; **Rafols, Ismael**; **Wroblewski, Angela** (2016). "Monitoring the evolution and benefits of responsible research and innovation (MoRRI) – a preliminary framework for RRI dimensions &

indicators". In: *21st Intl conf on science and technology indicators: Peripheries, frontiers and beyond*.
http://hdl.handle.net/10261/161892

Middaugh, Ellen; **Kahne, Joseph** (2013). "Nuevos medios como herramienta para el aprendizaje cívico". *Comunicar*, v. 20, n. 40, pp. 99-108.
https://doi.org/10.3916/c40-2013-02-10

Murry, John W.; **Hammons, James O.** (1995). "Delphi: A versatile methodology for conducting qualitative research". *The review of higher education*, v. 18, n. 4, pp. 423-436.
https://doi.org/10.1353/rhe.1995.0008

Neresini, Federico; **Bucchi, Massimiano** (2011). "Which indicators for the new public engagement activities? An exploratory study of European research institutions". *Public understanding of science*, v. 20, n. 1, pp 64-79.
https://doi.org/10.1177/0963662510388363

Osborne, Jonathan; **Collins, Sue**; **Ratcliffe, Mary**; **Milar, Robin**; **Duschl, Rick** (2003). "What 'ideas-about-science' should be taught in school science? A Delphi study of the expert community". *Journal of research in science teaching*, v. 40, n. 7, pp. 692-720.
https://doi.org/10.1002/tea.10105

Ouariachi, Tania; **Olvera-Lobo, María-Dolores**; **Gutiérrez-Pérez, José** (2017). "Analyzing climate change communication through online games". *Science communication*, v. 39, n. 1, pp. 10-44
https://doi.org/10.1177/1075547016687998

Pacios, Ana-Reyes; **Vianello-Osti, Marina**; **Rodríguez-Bravo, Blanca** (2016). "Transparencia y acceso a la información sobre proyectos de investigación en las universidades públicas españolas". *El profesional de la información*, v. 25, n. 5, pp. 721-729.
https://doi.org/10.3145/epi.2016.sep.02

Papacharisi, Zizi (2002). "The virtual sphere: The internet as a public sphere". *New media & society*, n. 4, v. 1, pp. 9-27.
https://doi.org/10.1177/14614440222226244

Ravn, Tine; **Mejlgaard, Niels** (eds.) (2014). *Public engagement innovation for Horizon 2020. Catalogue of PE initiatives*. University of Helsinky; University of Aarhus; Vilnius University International Business School; Laboratorio di Scienze della Cittadinanza.
http://www.vm.vu.lt/uploads/pdf/Public_Engagement_Innovations_H2020-2.pdf

Rowe, Gene; **Frewer, Lynn J.** (2005). "A typology of public engagement mechanisms". *Science, technology & human values*, v. 30, n. 2, pp. 251-290.
https://doi.org/10.1177/0162243904271724

Scapolo, Fabiana; **Miles, Ian** (2006). "Eliciting experts' knowledge: A comparison of two methods". *Technological forecasting and social change*, n. 73, v. 6 pp. 679-704.
https://doi.org/10.1016/j.techfore.2006.03.001

Seakins, Amy; **Dillon, Justin** (2013). "Exploring research themes in public engagement within a Natural History Museum: A modified Delphi approach". *International journal of science education*, v. 3, n. 1, pp. 52-76.
https://doi.org/10.1080/21548455.2012.753168

Shuai, Xin; **Pepe, Alberto**; **Bolen, Johan** (2012). "How the scientific community reacts to newly submitted preprints: Article downloads, Twitter mentions, and citation". *PLoS one*, v. 7, n. 11.
https://doi.org/10.1371/journal.pone.0047523

Smith, Kathleen S.; **Simpson, Ronald D.** (1995). "Validating teaching competencies for faculty members in higher education: A national study using the Delphi method". *Innovative higher education*, v. 19, n. 3, pp. 223-234.
https://doi.org/10.1007/bf01191221

Weigold, Michael F.; **Treise, Debbie** (2004). "Attracting teen surfers to science web sites". *Public understanding of science*, v. 13, n. 3, pp. 229-248.
https://doi.org/10.1177/0963662504045504

Weitze, Marc-Denis (2017). *Social media and digital science communication. Analysis and recommendations for dealing with risks and opportunities in a democracy*. Berlin: National Academy of Science and Engineering, German National Academy of Sciences Leopoldina and Union of the German Academies of Sciences and Humanities. ISBN: 978 3 804736320
https://www.acatech.de/wp-content/uploads/2018/03/WOM2_EN_web_final.pdf

Wolf, Julie M. (2017). "The multipurpose tool of social media applications for scientists, science communicators and educators". *Clinical microbiology newsletter*, v. 39, n. 10, pp. 75-79.
https://doi.org/10.1016/j.clinmicnews.2017.04.003

Off-page SEO and link building: General strategies and authority transfer in the digital news media

Carlos Lopezosa; Lluís Codina; Carlos Gonzalo-Penela

Nota: Este artículo se puede leer en español en:
http://www.elprofesionaldelainformacion.com/contenidos/2019/ene/09_es.pdf

How to cite this article:

Lopezosa, Carlos; **Codina, Lluís**; **Gonzalo-Penela, Carlos** (2018). "Off-page SEO and link building: General strategies and authority transfer in the digital news media". *El profesional de la información*, v. 28, n. 1, e280107.
https://doi.org//10.3145/epi.2019.ene.07

Article received on June 15th, 2018
Approved on December 12th, 2018

Carlos Lopezosa ✉
http://orcid.org/0000-0001-8619-2194

Pompeu Fabra University
Department of Communication
Roc Boronat, 138. 08018 Barcelona, Spain
carlos.lopezosa@upf.edu

Lluís Codina
http://orcid.org/0000-0001-7020-1631

Pompeu Fabra University
Department of Communication
Roc Boronat, 138. 08018 Barcelona, Spain
lluis.codina@upf.edu

Carlos Gonzalo-Penela
https://orcid.org/0000-0002-3380-6823

Pompeu Fabra University
Department of Communication
Roc Boronat, 138. 08018 Barcelona, Spain
carlos.gonzalo@upf.edu

Abstract

In recent years, a number of digital news media outlets have begun to include paid links in their content. This study seeks to identify and analyse this content whose sole purpose is to improve the website authority of the advertisers and their search engine rankings. To do so, it employs two basic methodologies: first, it undertakes a systematic review of off-page SEO practices, the digital press and native advertising; and, second, it reports a case study based on the identification and analysis of 150 news items that contain specially commissioned links resulting from a commercial transaction. The study provides evidence of a new revenue stream for the digital news media, one that is not clearly disclosed and which is based on the sale of links. The article includes a discussion of the case study findings, and presents future guidelines for the use of paid links based on the emerging concept of 'native advertising'.

Keywords

Digital news media; Online journalism; Digital journalism; SEO; Off-page SEO; Web positioning; Link building; Native advertising; Journalism ethics.

1. Introduction and study goals

This study examines a new activity being practiced by the digital news media involving the sale of links and aimed at improving the web positioning of the sites that receive them. These links are embedded in articles that are written solely with this purpose in mind. In some instances, the content of the articles is actually provided by the buyer of the links. It is, therefore, a form of advertising which, as it is not clearly disclosed, can easily be confused with editorial content.

Such content can give rise to three types of problem:

- as it is produced with the sole purpose of serving as a link vector, its quality is of secondary importance and, moreover, it responds not to journalistic criteria but rather to those of advertising;
- as it is not clearly identified as sponsored content, it threatens to undermine professional ethics, which requires the unambiguous separation of this type of content from editorial content.
- as it is presented in a non-transparent fashion, not only is the public unaware of such practices, but many studies (and experts) of the digital media are also unaware of it.

If, in the current digital information ecosystem, this is one of the engines of content creation, then experts undisputedly have an incomplete picture of this ecosystem if they are unfamiliar with this practice.

> We are dealing with a form of advertising that, as it is not clearly disclosed, can easily be confused with editorial content

To understand the emergence of this activity, in this study we examine the SEO strategy that underpins it, based on what is known as 'link building'. More specifically, the goals of this study can be stated as follows:

- To analyse and characterise a new line of activity in the digital news media centred on link buying/selling and to identify the actors involved.
- To classify the content published as a result of this activity and to examine its implications for off-page SEO strategies.
- To provide guidelines for the possible improvement of this activity by developing a set of best practices modelled on so-called native advertising.

In keeping with these objectives, we address the following research questions:

- What are the main characteristics of this new line of activity centred on link buying/selling in the digital news media and who are its main actors?
- What are the main characteristics of the content published as link vectors?
- Is it possible to develop a set of best practices for this activity based on native advertising in order to improve it?

1.1. Methodology

The following two basic methodologies were employed in conducting this study:

> As it is not identified as sponsored content, it threatens to undermine professional ethics, which requires the unambiguous separation of this type of content from editorial content

- Systematic literature review (**Hart**, 2008; **Booth**; **Papaionnou**; **Sutton**, 2012) of the articles listed in the bibliography of this paper, based on a prior consultation of the *Scopus*, *Web of Science*, *Lista* and *Communication Source* databases. We also consulted the most authoritative professional sources in relation to SEO –including *Search Engine Journal* and *Search Engine Land*- and native advertising –including the *Native Advertising Institute* and the *Nieman Reports*-.
- Case study research (**Yin**, 2014) involving the identification, selection and classification of a set of 150 news items, published by three digital news media outlets and originating from link buying/selling.

The results of the systematic literature review are presented first followed by the case study findings. The former enable us to outline the main characteristics of off-page SEO strategies, which are the origin of this practice; the latter allow us to classify the products of this activity.

2. Off-page SEO

It is worth recalling that *Google* was the first search engine to apply a technique based on hyperlink (i.e. the links between web pages) analysis to determine the relative importance of all pages on the World Wide Web.

For analyses of this type, the inventors of *Google* based their work on citation analysis in the academic world and its corresponding impact factor. In this way, they designed a metric –*PageRank*- that serves to express the results of such an analysis (**Brin**; **Page**, 2000).

Given its enormous efficacy, *Google* has had an enduring influence on the way in which search engines display their results pages, with all of them adopting the same basic idea (**Kleinberg**, 1998; **Lewandowski**, 2012; **Giomelakis**; **Veglis**, 2015). The reason for its widespread adoption is that it provided the first genuinely efficient response to all the challenges posed by Internet searches (**Gonzalo-Penela**; **Codina**; **Rovira**, 2015), although initially no firm in the search engine sector seemed to realise.

More specifically, the new idea developed by *Google* was the following: instead of calculating the relevance of each page exclusively in terms of its intrinsic characteristics –including, for example, the number of times the keyword appears-, it also took into account its extrinsic characteristics, most notably, the number and quality of links it receives (**Harry**, 2013).

What was the underpinning rationale? In broad terms, given two pages addressing the same theme, the more important of the two is considered to be the one that receives the greater number of backlinks from websites which, in turn, are highly linked (**Brin**; **Page**, 2000; **Thelwall**, 2004; **Gonzalo-Penela**, 2006).

Here, the key point is that part of a page's *PageRank* can be transferred to other pages if they are linked to it. *PageRank* is also a measure of a page's authority in the same sense that a journal's impact factor is a measure of its authority.

In this way, the net effect of these links –indistinctly known as backlinks, inbound links or external links- is to transfer authority from the page that points to the linked page, improving its visibility in the search engines (**Crowe**, 2017; **Giomelakis**; **Veglis**, 2016).

> On-page SEO: actions to optimize web page content.
> Off-page SEO: link building actions (to obtain backlinks)

Consequently, the number and quality of the links that link to a website are an indicator of its essential relevance, as well as being one of the most influential positioning factors (**Fishkin**, 2016; **García-Carretero** *et al*., 2016). It is not surprising, given these circumstances, that firms' SEO managers seek to implement link building strategies (**Gonzalo-Penela**, 2006; **Serrano-Cobos**, 2015). This, in turn, has led to two major branches of SEO:

- On-page SEO: actions to optimize web page content.
- Off-page SEO: actions to obtain backlinks, that is link building.

Several link building procedures have been developed (**Monterde**, 2016; *Publisuites*, 2018), among which two stand out:

- Natural or editorial link building: this is based on a similar logic to that of the impact factor of academic articles, whereby a high quality article is one that will be highly cited, thus establishing itself as an article of great authority. In the case of the web, this type of link building is achieved by creating high quality content.
- Strategic link building: this is a proactive practice that requires direct contact between the website manager and the author of another site to which a link is requested. If performed on a massive scale, *Google*, *Bing*, *Yahoo*, *Yandex*, etc. are able to identify patterns of unnatural links, and if so, penalize those web sites by pushing them down the search result rankings, or even excluding them from their indexes

The main goals, therefore, of off-page SEO professionals are (**Cámaras-León**, 2018; **Rowe**, 2018):

- to search for and obtain a large number of backlinks;
- to multiply the strength of backlinks by ensuring that the sites from which the links originate are in turn highly linked.

> *Dofollow* links link related themes. Given their editorial nature, they transfer authority to the linked website

There exist various websites where it is possible to obtain free backlinks. Primarily they can be obtained from web profiles, forums, social networks, blogs 2.0, comments on websites/blogs, wikis, content aggregators, directories, newspapers, third-party websites, etc., and of course from other websites (**Cooper**, 2012).

From a technical point of view, but with far-reaching implications for the matter in hand, there are two types of backlink: *dofollow* links (also known as follow), and *nofollow* links (**Dean**, 2018).

Both types of link are identified by means of the corresponding labelling of the source code (not visible on the page). They can be explained as follows:

- *dofollow* links fulfil the original function of hyperlinks, that is, they link related themes. Due to their editorial nature, *Google* considers them a way of transferring authority to the linked website, and the amount of authority or of PageRank transferred depends on the quality or authority of the page that creates the link. *Dofollow* means that *Google* will follow the link and attribute PageRank to the page that receives it. In theory, *dofollow* links are limited to editorial links. *Dofollow* links do not have a brand. In other words, a standard link, without any additional brand, is a *dofollow* link.
- *nofollow* links, on the other hand, include a source code tag that tells searchers that this link cannot be used for PageRank. It is a code that informs search engine robots not to follow the link (hence its name). Since they correspond to advertising links, the transmission of authority in this case is zero.

2.1. Anchor text

> *Nofollow* links include a label that tells the search engines they cannot be used to transfer authority because they are advertisements

The links are made up not only of the corresponding URL, but also of a text known as the anchor text (**González-Villa**, 2017). This is the portion of the text that activates the link on the web page from which it originates.

For *Google*, the anchor text forms part of the content of the linked site, and it is used to determine whether that site is relevant for the keyword contained in the anchor text (Figure 1).

In short, we should stress the following: the authority of the site from which a link originates, the link's anchor text and the context in which that link is included are the most important elements of link building.

Figures 1 and 2 illustrate the main concepts associated with links, as presented above.

Figure 1 shows the structure of a link using the source code. It can be seen that:

- the link's destination, that is, the page that will open in the browser if the user clicks on it is https://es.unesco.org
- the anchor text is Unesco.

Figure 1. Source code of a *dofollow* type link

This is a *dofollow* link because it does not include any additional coding (see Figure 2). For this reason, this link transfers PageRank or authority to the Unesco page. If the page containing this link belongs to a leading digital newspaper, such as *The New York Times*, the authority transferred will be very high. Moreover, *Google* will understand that the Unesco keyword in the anchor text is part of the content of the destination page.

Figure 2 shows the structure of a *nofollow* link, since it incorporates the 'rel' attribute, with the *nofollow* value. Due to this attribute, the link does not transfer authority to the (fictitious) destination page *Store*. In this case, the authority of the page containing the link is of no significance. Moreover, because of this attribute, *Google* will not follow the link and will not transfer any value.

Figure 2. Source code of a *nofollow* type link

3. Hyperlink buying/selling sector

Any human activity that can give rise to a supply and demand relationship will eventually generate a market. Here, news media websites have something (offer) that the SEO managers of other sites need (demand), namely, backlinks of great authority.

Hence, it was only a matter of time before SEO managers began to explore the possibility of buying links in the digital news media. To mediate in this relationship, a number of intermediaries have emerged to act as go-betweens for the website managers that need backlinks and the online news media.

> The authority of the site from which a link originates, the link's anchor text and the context in which that link is included are the most important elements of link building

Given this change in the digital information ecosystem, one of the objectives of this study was to identify the leading firms operating in this sector. Although it is impossible to determine exactly how many companies are operating in Spain, based on our monitoring of professional forums, social networks and other sources regularly used by the sector's professionals, we believe that those shown in Tables 1 to 5 are, by far, the most important.

To outline the main characteristics of these intermediaries, we present their principal features in the following data files. All data are derived from their respective websites (Tables 1 to 5).

Table 1. *PrensaRank*

Name	***PrensaRank***
URL	*https://prensarank.com*
Description	Its website claims they have 3,660 customers, 408 newspapers from which they can obtain links, and (as of February 2018) they had sold 30,024 articles to the news media.
Digital news media	On registration the user obtains a link purchase interface. We identify 305 newspapers, distributed geographically as follows: Andorra: 1 newspaper; Saudi Arabia: 1; Argentina: 9; Chile: 4; Spain: 236; USA: 2; Mexico: 45; Nicaragua: 1; Peru: 1; Portugal: 1; United Kingdom: 2; Venezuela: 2.
Themes	Current affairs; Love, weddings, relationships, and couples; Betting and casinos; Art, decoration and design; Film and television; Cooking and gastronomy; Dating; Sports; Economics and politics; Education and culture; Company (advertising); Home, decoration and DIY; Humour and leisure; Computers and technology; Games and video consoles; Marketing and SEO; Pets and nature; Music and shows; Fashion and beauty; Cars; Women, babies, and children; Others; Religion, mysticism and esotericism; Health; Estate agency services; Sex shops; Sexuality; Tarot; Travel, hotels and tourism.
Maximum price for link	€ 950

Table 2. *Unancor*

Name	***Unancor***
URL	*https://www.unancor.com*
Description	Its website claims (as of February 2018) they have 6,000 customers and 500 newspapers from which they can obtain links.
Digital media	On registration the user obtains a link purchase interface. We identify 431 newspapers, distributed geographically as follows: Germany: 53 newspapers; Argentina: 39; Canada: 1; Chile: 13; Colombia: 6; Costa Rica: 1; El Salvador: 1; Arab Emirates: 1; Spain: 213; USA: 18; France: 17; Italy: 1; Morocco: 1; Mexico: 44; Monaco: 1; Nicaragua: 1; Panama: 1; Peru: 2; Uruguay: 3; Venezuela: 5
Themes	All these newspapers are associated with one or more of the following topics: Art and culture; Health and sport; Economics and business; Education; Home, decoration and DIY; Cooking and recipes, gastronomy; Computers, technology, mobiles and apps; Marketing (offline and online); Nature (animals and plants); Cars and motorcycles; Cinema, TV and music; News and politics; Travel and tourism; Others; Fashion and beauty; Erotica; Love, relationships, couples; Services (locksmiths, home improvements, plumbers, etc.); Legal; Children; Tarot
Maximum price for link	€ 10,000

Table 3. *Publisuites*

Name	***Publisuites***
URL	*https://www.publisuites.com/es*
Description	Its website claims they have 54,967 users and 478 newspapers from which they can obtain links. As of February 2018, they had sold 39,334 articles to the news media and blogs.
Digital media	On registration the user obtains a link purchase interface. We identify 478 newspapers, distributed geographically as follows: Argentina: 19 newspapers; Australia: 1; Bolivia: 1; Brazil: 4; Chile: 6; Colombia: 4; El Salvador: 1; Spain: 304; USA: 3; France: 26; Honduras: 1; Italy: 70; Jersey: 1; Mexico: 14; Nicaragua: 1; New Zealand: 1; Panama: 1; Paraguay: 1; Peru: 6; Portugal: 3; United Kingdom: 1; Dominican Republic: 1; Senegal: 1; South Africa: 1; Venezuela: 6.
Themes	All these newspapers are associated with one or more of the following topics: Betting, casinos and lotteries; Celebrities; Cooking, recipes and gastronomy; Trivia; Sports; Economy; Education and training; Entrepreneurs and SMEs; Computers and programming; Literature and culture; Music and radio; Marketing, SEO and social platforms; Miscellaneous; Fashion and accessories; Cars and motorcycles; Nature and ecology; News; Leisure and free time; Politics; Health; Technology; Mobile telephones and apps; Travel and tourism.
Maximum price for link	€ 1,943

Table 4. *RT Gopress*

Name	***RT Gopress***
URL	*https://rtgopress.com*
Description	Its website claims it is the most economically competitive Seo MarketPlace, Social Media and Growth Hacking firm in the market. They do not indicate how many newspapers or customers they have.
Digital media	On registration the user obtains a link purchase interface. We identify 155 newspapers distributed geographically as follows: Argentina: 3; Mexico: 26; Spain: 126.
Themes	All these newspapers are associated with one or more of the following topics: Current affairs; Stock market; Sports; Economics; Gastronomy; Marketing; Cars; Tourism; News; Technology; Health; Video games.
Maximum price for link	The website operates a price filter, but it appears not to be operative.

Table 5. *Dofollow*

Name	***Dofollow.es***
URL	*http://dofollow.es*
Description	While offering a similar service to the above firms, it operates differently. Thus, they offer what they call the *dofollow pack*. The customer writes a press release including two links to its website (1 for a brand and the other a keyword) and they undertake to publish the press release in four digital newspapers.
Digital media	Its website includes general, regional, and specialized newspapers of all types. They state that these media may vary depending on availability.
Themes	Unknown.
Maximum price for link	€ 339 for its most complete package.

4. Native advertising

The general absence of studies examining the link buying/selling sector in the digital news media and, hence, the development of any guidelines for its self-regulation, leads us here to consider the possibility of applying best practices in the so-called native advertising industry.

The *Native Advertising Institute* defines native advertising as the use of paid ads that match the look, feel and function of the content of the platform in which they appear (**Schauster**; **Ferrucci**; **Neill**, 2016; **Pollitt**, 2018).

Native advertising consists of news items, reports and, in general, quality content, its effectiveness being based on credibility. These characteristics can be used to provide quality content to publications (**Sweetser** *et al.*, 2016; **Carlson**, 2016). An essential point is that native advertising must present a branded message that allows readers to recognize not only the fact that it is sponsored content (**Ferrer-Conill**, 2016; **Amazeen**; **Muddiman**, 2017; **Amazeen**; **Wojdynski**, 2018), but also the logical intent of the advertisement to persuade and sell (**Mathiasen**, 2018).

> Native advertising must present a branded message that allows readers to recognize not only the fact that it is sponsored content, but also the logical intent of the advertisement to persuade and sell

As such, the idea is that the digital press should have a model for including sponsored content that allows it to be differentiated from their editorial content, and which, moreover, ensures it can be integrated naturally in the publication, maintaining a level of quality similar to that of the platform that hosts it (**Cramer**, 2016; **Li**, 2017; **Batsell**, 2018).

5. Case study

Having identified the key components and actors operating in the link buying/selling industry, we present our case study, which consists of a comparative analysis. We examined 150 news items that have been published as a direct result of the buying/selling of links. As such, we are dealing with content specially commissioned with the aim of including links to improve the website authority of the customers who purchase them.

To shed greater light on this procedure, we first explain how the whole process works. First, the customer contacts one of the link building firms described in the section above to purchase backlinks to its website from the digital news media. This news outlet then publishes content that includes links to the customer's website. In so doing so the process is terminated, following payment by the buyer at the price stipulated for receipt of backlinks. It should be stressed that what is purchased is the link or backlink and that the content is merely the vehicle in which it is included, which generally results in content unrelated to the newspapers normal editorial line.

To explore this market, we conducted an analysis whose object of study was three digital news media of medium to high importance.

Given the nature of this analysis, we do not explicitly identify the name of each news outlet, but rather describe them as accurately as possible using a series of data files (Tables 6 to 8). In these files we specifically incorporate the data provided by *Alexa Rank*, a ranking developed by *Amazon*, based on web traffic.

In addition, to lend greater credibility to the ranking of these three digital news media companies, we incorporate daily unique user data for each of the three websites. To do so, we used *Site Worth Traffic*, which measures website traffic providing unique and total user data, social network performance metrics, and a complete analysis of the site's evolution.

To select the news stories from the three media companies, we purchased three items from the *Prensarank* website (one item for each news media). Then, having

> News media websites can offer backlinks of great authority

Table 6. Digital news media company 1

Media company 1 (MC1)	
Media type	Generalist
Country	Spain
Alexa ranking	Ranked 252nd in Spain (June 2018)
Daily unique users	69,782

Table 7. Digital news media company 2

Media company 2 (MC2)	
Media type	Generalist
Country	Spain
Alexa ranking	Ranked 702nd in Spain (June 2018)
Daily unique users	20,864

Table 8. Digital news media company 3

Media company 3 (MC3)	
Media type	Generalist
Country	Spain
Alexa ranking	Ranked 4,763rd in Spain (June 2018)
Daily unique users	5,979

examined these news stories, we were able to identify a search pattern for each item, and with this to create what is known as its 'footprint': that is, a type of advanced search (*Google*, 2018) that allows the highly precise selection of well-characterized web page types.

In this way we identified three footprints that allowed us to locate 50 news items purchased from each of the three media companies. Each of these footprints, in the form of an advanced search equation, is constructed as follows (using the site search operator):

- site: MC1(media company 1) + name of link buying/selling company.
- site: MC2(media company 2) + the word "*remitido*" (or "press release").
- site: MC3(media company 3) + name of a news item contributor.

Having obtained the 150 news items (50 for each media company) by applying the respective equations, we were then able to isolate the following elements by responding to the six questions below, based on recommendations made by the *Native Advertising Institute* and the *Nieman Reports*:

- Is the news item specifically identified as sponsored content?
- Is the story reported newsworthy, that is, is the item directly linked to a breaking news story or current affairs?
- How many hyperlinks are included in each news item?
- Are the hyperlinks coherent with the content of the news item?
- Do the hyperlinks point to an authoritative website providing users with complementary quality information?
- What themes are the commissioned news items included in?

6. Results

Below, we first present our main findings. Next, we review our research objectives and questions in order to present our conclusions, and we finish with proposals for the development of new lines of research.

> "A number of intermediaries have emerged to act as go-betweens for the website managers that need backlinks and the online news media"

6.1. Main findings

From our study of the 150 news items commissioned in the three digital news media companies, the following results can be highlighted:

- News originating from the purchasing of a link is not clearly identified as sponsored content or advertising.
- The content does not describe or narrate a breaking news story or current affairs, that is, it is not a typical news story, but rather the content is timeless, generally involving recommendations and advice.
- The need to include the literal anchor text (the text that activates the link) as commissioned by the customer leads to errors of grammar and syntax in the writing of the content. The reason for this is that the authors opt to respect the keyword or phrase commissioned by the customer even if it does not fit with the syntax or phrase in which it is embedded.
- When a news item contains more than one link, the need to maintain two or more links in the same item for sites of distinct natures results in a lack of coherence between the links and the content of the news story.

Tables 9 to 11 show the results for each of the three digital news media companies in greater detail.

Table 9. Results for MC 1

Are news items identified as sponsored content?	Identification somewhat ambiguous. Items are identified as a *Communicado* (or news release). The headline is displayed in the following format: "News release: title of the story".
Is the content newsworthy?	No. It is timeless involving recommendations and offering tips.
How many hyperlinks are included in each item?	Of the 50 items analysed: 1 item includes one hyperlink 8 items include two hyperlinks 4 items include three hyperlinks 37 items include four hyperlinks
Are the hyperlinks coherent with the content?	No. Most are shoehorned into the item; others use a syntactically incorrect generic anchor text. In some of the items with more than one hyperlink, there is no thematic link between them.
Do the hyperlinks point to an authoritative website providing users with complementary quality information?	No. In general, hyperlinks point to websites that are not authoritative and, therefore, do not provide noticeable added value for the user-reader.
What themes are the commissioned news items included under?	The main themes are business, the home, beauty, tourism, productivity, cars, weddings, fashion, health, and decoration.

Table 10. Results for MC2

Are news items identified as sponsored content?	Ambiguous. Items are identified with a tag that reads *Remitido* (or news/press release) followed by the headline.
Is the content newsworthy?	No. It is timeless involving recommendations and offering tips.
How many hyperlinks are included in each item?	Of the 50 items analysed: 10 items include one hyperlink 13 items include two hyperlinks 6 items include three hyperlinks 21 items include four hyperlinks
Are the hyperlinks coherent with the content?	No. Most are shoehorned into the item; others use a syntactically incorrect generic anchor text. In some of the items with more than one hyperlink, there is no thematic link between them.
Do the hyperlinks point to an authoritative website providing users with complementary quality information?	No. In general, hyperlinks point to websites that are not authoritative and, therefore, do not provide noticeable added value for the user-reader.
What themes are the commissioned news items included under?	The main themes are work, recipes and gastronomy, business, cars, healthy living, gadgets, cooking, fashion trends and styles, and fortunes and tarot.

Table 11. Results for MC3

Are news items identified as sponsored content?	No. Items are presented as another news story, that is, as editorial content.
Is the content newsworthy?	No. It tends to be timeless involving recommendations and offering tips.
How many hyperlinks are included in each item?	Of the 50 items analysed: 17 items include one hyperlink 11 items include two hyperlinks 2 items include three hyperlinks 20 items include four hyperlinks
Are the hyperlinks coherent with the content?	No. Most are shoehorned into the item; others use a syntactically incorrect generic anchor text. In some of the items with more than one hyperlink, there is no thematic link between them
Do the hyperlinks point to an authoritative website providing users with complementary quality information?	No. In general, hyperlinks point to websites that are not authoritative and, therefore, do not provide noticeable added value for the user-reader.
What themes are the commissioned news items included under?	The main themes are business, virtual spaces, tourism, music, health, cars, investments, holidays, problem pages, and travel.

7. Discussion and conclusions

7.1. Discussion

Link building strategies and link buying can greatly benefit both SEO companies in their bid to provide their clients' websites with greater authority, as well as news media companies as they seek to grow their revenue. However, without adequate regulation, users stand to suffer, being presented with poor quality content and information that do not respond to criteria of journalistic or editorial feeds, but rather to those of advertising.

Specifically, we have seen that in two of the three Spanish news media companies analysed some attempt is made to signal a divide between sponsored and editorial content, but such attempts are ultimately ambiguous.

> "*Google* has had a major influence on the way in which search engines display their results pages"

Instead of identifying the content with an unmistakable tag indicating sponsorship or advertising, other labels are employed, such as *comunicado* (news release) or *remitido* (news/press release). This is better than nothing, but it remains nevertheless ambiguous. *Comunicado* or *remitido* are usual journalistic terms for referring to press releases that serve as the basis for perfectly valid editorial content, which is why these tags must be considered inadequate, albeit that they do represent some attempt on the part of the publication to indicate their actual content.

In contrast, one of the news media companies does not seek to make any distinction in content, which is a more serious matter.

In all three cases, readers may well think they are reading editorial content and, therefore, believe that the linked sites have been selected for their quality when in fact what they are reading is advertising or sponsored content.

Moreover, as their origin is not strictly editorial, the content tends to be largely superficial and to have little or no relationship with the linked sites.

These two closely related factors have a somewhat negative impact on the quality of the content of the news media companies. However, it is apparent that if we adhere to the Native Advertising model, the interests of all parties can be reconciled: Advertisers can obtain authoritative links, the content can be genuinely interesting – while at the same time being clearly identified as sponsorship – and the media can have a new model of sustainability.

7.2. Conclusions

To present the conclusions, we first go back to the study's initial objectives to consider how far they have been fulfilled. Then we do the same with the research questions.

Objectives

Objective 1. To analyse and characterise a new line of activity in the digital news media centred on link buying/selling and to identify the actors involved.

We have shown that a new model of economic activity has emerged based on link buying/selling and that this activity is becoming increasingly more commonplace, as demonstrated by our close monitoring of the sector over the last two years. As a result, the number of news media companies now included on the websites studied here (*Prensarank*, *Unancor*, *Publisuite* and *RT Gopress*) has experienced constant growth.

We have shown that this line of activity adds value to each party involved –the digital news media, the customers that buy links and the firms that act as intermediaries in the sales transaction- as it seeks to fulfil three main objectives:

- Providing a new revenue stream, albeit that for the time being it remains a fairly marginal stream for news media companies.
- Obtaining greater website authority and improving the visibility of the websites that buy backlinks.
- Generating revenue in the form of commissions to the intermediary firms dealing in hyperlinks.

Objective 2. To classify the content published as a result of this activity and its implications for off-page SEO strategies.

We have shown that the sector does not operate a system of self-regulation, since each of the three news media companies analysed applies different criteria. Furthermore, contrary to native advertising, the sponsored content does not conform to the look, feel and function of the content of the platform on which they appear.

Different degrees of ethical awareness can also be identified, since news media companies 1 and 2 at least go some way to specifically identifying this content (by labelling items as *comunicados* or *remitidos*), while company 3 avoids drawing any distinction between editorial and sponsored content.

Objective 3. To provide guidelines for the possible improvement of this activity by developing a set of best practices modelled on so-called native advertising.

Based on native advertising regulations, an initial proposal of best practices for the writing of news items for link selling should consider the following guidelines:

- There should be a clear indication that the news story published is sponsored content or advertising – the distinction being that the latter is provided by the advertiser, the former by the news media company itself.
- The news item should match the look, feel and function of the content of the platform on which it appears.
- The information included in the commissioned news item should be newsworthy or, at least, useful for the reader, and should be based on current news stories. The news story ought to be written with the user in mind and should not be motivated solely by the hyperlink that has been purchased. Its features should serve not only the needs of the link buyers but also those of the readers.
- More than one link can be included in a news story provided there is a thematic connection between them that does not affect the story's overall coherence.
- The hyperlinks and their anchor texts must be orthographically and syntactically coherent with the text of the news item.
- As a rule, if the hyperlinks do not lead to an authoritative website that provides useful, complementary information to readers, then this link should not be added to editorial content. Instead, these links should be published in a section dedicated exclusively to sponsored content or advertising and separated from the newspapers' usual sections.

Research questions

Next, we return to the research questions posed at the outset to examine the responses obtained from the case study reported above.

Question 1. What are the main characteristics of this new line of activity centred on link buying/selling in the digital news media and who are its main actors?

We have shown that it is possible to both clearly identify and determine the characteristics of this line of activity in the news media centred on the acquisition of links and content that act as vectors for these links and content.

It is a business model in which the three main actors, i.e. the digital news media, their customers, and link buying intermediaries, all benefit. The news media and the intermediaries obtain an economic return, while the clients obtain greater web site authority and visibility. The loser in the activity is, however, journalistic quality and, with it, the readers of the news media.

Question 2. What are the main characteristics of the content published as link vectors?

The analysis shows that the news items identified in this case study present the following characteristics:

- They do not carry clear labels identifying their content as advertising or sponsored.
- They are timeless, focusing primarily on providing advice and basic recommendations on a huge variety of topics ranging from tourism, cooking, and cars, to investments, beauty, and technology, and many others.
- They can include up to four backlinks. These links are often shoehorned into the content, not only because they are poorly constructed in terms of their semantics but also because they link to websites that do not provide complementary quality information for their readers.

Question 3. Is it possible to develop a set of best practices for this activity based on native advertising in order to improve sector practices?

Here, we have taken the concept of native advertising as our reference because it can be considered to provide interesting precedents and, as such, to be a model for future regulations governing paid links in the digital press.

Broadly speaking, the news items in our sample point clearly to the need to develop a set of best practices, preferably so that the media companies can self-regulate themselves, rather than depend on an external regulator.

Digital news media readers deserve the highest degree of quality and transparency, characteristics that ultimately benefit the news media themselves, especially if we consider the acute crisis they are currently experiencing. It is important that the media generate additional revenue streams, which is why this line of business should be understood as being both necessary and timely.

However, the sector's legitimacy calls for a highly transparent and stringent system of self-regulation and, here, we have identified some of the essential elements that need to be taken into consideration in developing such a system. The

key idea in the process is that the transfer of authority effected by link buying/selling should not negatively impact the content quality or the reading experience of the news media that participate in this business model. Additionally, maximum transparency must be guaranteed at all times.

> The transfer of authority effected by the link buying/selling should not negatively impact the quality or the reading experience of the news media involved in this business model

8. Future research

More ethical studies need to be undertaken within the digital news media to determine best practices for the selling of commissioned news items and hyperlinks. In this way it should be possible to reconcile the sector's legitimate interest for sponsorship or advertising revenue with the interests of their users who consume news and with their right to receive quality content which, even if sponsored, should be in line with the general orientation of the news outlet.

Within the field of SEO, analyses could be undertaken of the actual impact of links of this type in terms of improving the ranking of the websites that receive them. To do this, analytical frameworks need to be designed and employed in conjunction with such SEO tools as *Sistrix*, *SEMrush*, *Ahrefs*, or *Majestic*, among others.

9. References

Amazeen, Michelle A.; **Muddiman, Ashley R.** (2017). "Saving media or trading on trust? The effects of native advertising on audience perceptions of legacy and online news publishers". *Digital journalism*, v. 6, n. 2, pp. 176-195.
https://open.bu.edu/bitstream/handle/2144/27151/Amazeen_Muddiman_2017.pdf?sequence=1
https://doi.org/10.1080/21670811.2017.1293488

Amazeen, Michelle A.; **Wojdynski, Bartosz W.** (2018). "The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers". *Journalism*, pp. 1-20.
https://open.bu.edu/handle/2144/27308
https://doi.org/10.1177/1464884918754829

Batsell, Jake (2018). "4 steps to bring ethical clarity to native advertising". *Neiman report*, September 23rd.
https://niemanreports.org/articles/4-steps-to-bring-ethical-clarity-to-native-advertising

Booth, Andrew; **Papaionnou, Diana**; **Sutton, Anthea** (2012). *Systematic approaches to a successful literature review*. London: Sage. ISBN: 978 0 857021359

Brin, Sergey; **Page, Lawrence** (2000). "The anatomy of a large-scale hypertextual web search engine". Stanford University.
http://infolab.stanford.edu/~backrub/google.html

Cámaras-León, Nuria (2018). "Linkbuilding 2018, guía de enlazado perfecto (+12 predicciones expertos)". *Unancor*, 11th January.
https://www.unancor.com/blog/guia-linkbuilding

Carlson, Matt (2014). "When news sites go native: Redefining the advertising – editorial divide in response to native advertising". *Journalism*, v. 16, n. 7, pp. 849-865.
https://doi.org/10.1177/1464884914545441

Cooper, Jon (2012). "Link building tactics. The complete list". *Point Blank SEO*, April 1st.
http://pointblankseo.com/link-building-strategies

Cramer, Theresa (2016). "The deal with disclosure and the ethics of native advertising". *Digital content text*, Sept. 23rd.
https://digitalcontentnext.org/blog/2016/09/06/the-deal-with-disclosure-and-the-ethics-of-native-advertising

Crowe, Anna L. (2017). "Illustrated guide to link building". *Search engine journal*.
https://www.searchenginejournal.com/link-building-guide

Dean, Brian (2018). "The definitive guide (2018 update)". *Backlinko*, March 11th.
https://backlinko.com/link-building

Ferrer-Conill, Raul (2016). "Camouflaging church as state". *Journalism studies*, v. 17, n. 7, pp. 904-914.
https://doi.org/10.1080/1461670X.2016.1165138

Fishkin, Rand (2016). "Targeted link building in 2016 - Whiteboard Friday". *Moz*, Jan 29th.
https://moz.com/blog/targeted-link-building-in-2016

García-Carretero, Lucía; **Codina, Lluís**; **Díaz-Noci, Javier**; **Iglesias-García, Mar** (2016). "Herramientas e indicadores SEO: características y aplicación para análisis de cibermedios". *El profesional de la información*, v. 25, n. 3, pp. 497-504.
https://doi.org/10.3145/epi.2016.may.19

Giomelakis, Dimitrios; **Veglis, Andreas** (2015). "Employing search engine optimization techniques in online news articles". *Studies in media and communication*, v. 3, n. 1, pp. 22-33.
https://doi.org/10.11114/smc.v3i1.683

Giomelakis, Dimitrios; **Veglis, Andreas** (2016). "Investigating search engine optimization factors in media websites. The case of Greece". *Digital journalism*, v. 4, n. 3, pp. 379-400.
https://doi.org/10.1080/21670811.2015.1046992

González-Villa, Juan (2017). "Cómo hacer link building: estrategias y ejemplos prácticos. *Useo*, 30th March.
https://useo.es/como-hacer-link-building

Gonzalo-Penela, Carlos (2006). "Tipología y análisis de enlaces web: aplicación al estudio de los enlaces fraudulentos y de las granjas de enlaces". *BiD: textos universitaris de biblioteconomia i documentació*n, n. 16.
http://bid.ub.edu/16gonza2.htm

Gonzalo-Penela, Carlos; **Codina, Lluís**; **Rovira, Cristòfol** (2015). "Recuperación de información centrada en el usuario y SEO: categorización y determinación de las intenciones de búsqueda en la Web". *Index comunicación*, v. 5, n. 3, pp. 19-27.
http://journals.sfu.ca/indexcomunicacion/index.php/indexcomunicacion/article/view/197/175

Google (2018). *Google guide making searches even easier. Search operators*.
http://www.googleguide.com/advanced_operators_reference.html

Harry, David (2013). "How search engines rank web pages". *Search engine watch*, Sept. 23rd.
https://searchenginewatch.com/sew/news/2064539/how-search-engines-rank-web-pages

Hart, Chris (2008). *Doing a literature review: Releasing the social science research imagination*. London: Sage. ISBN: 978 0 761959755

Kleinberg, Jon M. (1998). "Authoritative sources in a hyperlinked environment". In: *Procs. of the ACM-SIAM Symposium on discrete algorithms*, pp. 1-33.
https://www.cs.cornell.edu/home/kleinber/auth.pdf

Lewandowski, Dirk (2012). "A framework for evaluating the retrieval effectiveness of search engines". In: Jouis, Christophe; Biskri, Ismail; Ganascia, Jean-Gabriel; Roux, Magali. *Next generation search engine: Advanced models for information retrieval.* Hershey, PA: IGI Global, pp. 456-479. ISBN: 978 1 466603318
https://arxiv.org/pdf/1511.05817.pdf

Li, You (2017). "Contest over authority". *Journalism studies*, pp. 1-19.
https://doi.org/10.1080/1461670X.2017.1397531

Mathiasen, Stine F. (2018). "10 quick takeaways from native advertising days 2018". *Native Advertising Institute*, Sept. 23th.
https://nativeadvertisinginstitute.com/blog/takeaways-native-advertising-days-2018

Monterde, Nacho (2016). "Introducción al link building". *SEO azul*, March 4th.
https://www.seoazul.com/introduccion-al-link-building

Pollitt, Chad (2018). *The global guide to technology 2018. A resource for marketers, advertisers, media buyers, communicators, publishers and ad tech professionals*.
https://nativeadvertisinginstitute.com

Publisuites (2018). "Estudio del uso de linkbuilding". *Publisuite*, 15th March.
https://www.publisuites.com/blog/estudio-de-linkbuilding-publisuites

Rowe, Kevin (2018). "How link building will change in 2018". *Search engine journal*, Feb. 2nd.
https://www.searchenginejournal.com/how-link-building-will-change/231707

Schauster, Erin E.; **Ferrucci, Patrick**; **Neill, Marlene S.** (2016). "Native advertising is the new journalism: How deception affects social responsibility". *American behavioral scientist*, v. 60, n. 12, pp. 1408-1424.
https://doi.org/10.1177/0002764216660135

Serrano-Cobos, Jorge (2015). "SEO: Introducción a la disciplina del posicionamiento en buscadores". *Colección EPI Scholar.* Barcelona: Editorial UOC. ISBN: 978 84 9064 956 5

Sweetser, Kaye D.; **Joo, Sun**; **Golan, Guy J.**; **Hochman, Asaf** (2016). "Native advertising as a new public relations tactic". *American behavioral scientist*, v. 60, n. 12, pp. 1442-1457.
https://doi.org/10.1177/0002764216660138

Thelwall, Mike (2004). *Link analysis: An information science approach*. Amsterdam: Elsevier. ISBN: 978 0 12 088553 4

Yin, Rober K. (2014). *Case study research. Design and methods*. Canada: SAGE. ISBN: 978 1 452242569

Mi Biblioteca

La revista del mundo bibliotecario

reportajes

opinión

Redes sociales

actualidad

bibliobuses

públicas

entrevistas

escolares

universitarias

Suscríbete

952 23 54 05

www.alonsoquijano.org

belen@alonsoquijano.org

Construcción social de la imagen de una ciudad a través de *Instagram*: el caso de Granada

The social construction of the image of a city through *Instagram*: The case of Granada

Francisco-Javier Cantón-Correa; Jordi Alberich-Pascual

Cómo citar este artículo:

Cantón-Correa, Francisco-Javier; **Alberich-Pascual, Jordi** (2019). "Construcción social de la imagen de una ciudad a través de *Instagram*: el caso de Granada". *El profesional de la información*, v. 28, n. 1, e280108.

https://doi.org/10.3145/epi.2019.ene.08

Artículo recibido el 29-09-2018
Aceptación definitiva: 13-12-2018

Francisco-Javier Cantón-Correa

https://orcid.org/0000-0002-8466-1679

Universidad de Granada
Facultad de Comunicación y Documentación
Campus Universitario de Cartuja.
18071 Granada, España
javicanton@ugr.es

Jordi Alberich-Pascual ✉

https://orcid.org/0000-0001-6871-4614

Universidad de Granada
Facultad de Comunicación y Documentación
Campus Universitario de Cartuja.
18071 Granada, España
jalberich@ugr.es

Resumen

Nuestro mundo digital es cada vez más visual. Aplicaciones móviles centradas en la fotografía digital como *Instagram* son vehículos para la creación, manipulación y difusión instantánea de imágenes. *Instagram* supone por tanto una ventana abierta a la investigación en Ciencias Sociales y Humanidades Digitales, y una oportunidad de investigar cómo los jóvenes usuarios de esta aplicación desarrollan la cultura visual en sus entornos locales a través de lenguajes visuales globales. El presente trabajo analiza la producción de *Instagram* en Granada a partir de una muestra de 955.564 publicaciones y 375.758 imágenes publicadas y geolocalizadas recogidas a lo largo de un año (entre abril de 2017 y abril de 2018), con el objetivo de mostrar cómo se construye socialmente la imagen de una ciudad. El análisis geográfico, de redes sociales y de contenido de la muestra del estudio ha permitido establecer cuáles son los principales focos de atracción visual de la ciudad, así como los barrios y distritos más activos en *Instagram*; los miembros más relevantes de una red de usuarios autoconstruida de manera heterogénea y abierta, con fuertes nodos intermediarios de estilo "instagramista" y que conectan subcomunidades; así como el uso extendido de la aplicación tanto para la promoción empresarial como para el desarrollo de la creatividad individual. Todo ello indica cómo las publicaciones de *Instagram* en Granada visibilizan la importancia e incidencia del turismo en la ciudad.

Palabras clave

Medios sociales; Redes sociales; Visualización; Visualización de datos; Información geográfica; Geolocalización; Cultura visual; Imagen; Fotografía, Ciudades; Urbana; Digitalización; *Instagram*.

Financiación

Este artículo se enmarca dentro del proyecto *Transmedialización y crowdsourcing en las narrativas de ficción y no ficción audiovisuales, periodísticas, dramáticas y literarias* (Referencia CSO2017-85965-P), dirigido por los profesores Domingo Sánchez-Mesa Martínez y Jordi Alberich-Pascual. *Ministerio de Economía, Industria y Competitividad*, Gobierno de España

Abstract

Our digital world is increasingly visual. Mobile applications focused on digital photography such as *Instagram* are vehicles for the creation, manipulation and instant diffusion of images. *Instagram* is therefore an open window to research in Social Sciences and Digital Humanities, and an opportunity to investigate how young users of this application develop visual culture in their local environments through global visual languages. The present work analyzes the production of *Instagram* in Granada from a sample of 955.564 publications and 375.758 images published and geolocated over a year (between April 2017 and April 2018), with the aim of showing how the image of a city is socially constructed. The geographical, social network, and content analysis of the study sample has allowed to establish which are the main focal points of visual attraction of the city, as well as the most active neighborhoods and districts; the most relevant members of a user network that is self-constructed in a heterogeneous and open manner, with strong intermediary nodes of an "instagramist" style that connect subcommunities; as well as the widespread use of the application both for business promotion and for the development of individual creativity. All this indicates how the *Instagram* publications in Granada make visible the importance and incidence of tourism in the city.

Keywords

Social media; Social networks; Visualization; Data visualization; Geographic information; Geolocation; Visual culture; Image; Photography; Cities; Urban; Digitization; *Instagram*.

1. Introducción

La irrupción de los *smartphones* y su conexión permanente a internet ha permitido la creación, modificación y publicación de imágenes a un ritmo vertiginoso. Además, el aumento en el ancho de banda ha sido clave para pasar del texto a las imágenes fijas y en movimiento. El nuevo lenguaje transmedia e intertextual está transformando nuestra cultura. En un mundo en el que consumo y producción cultural van de la mano y los medios convergen entre sí (**Jenkins**, 2008) en un "metamedio", nuestra cultura visual se desarrolla a través de la multiplicación icónica que han propiciado las cámaras digitales incluidas en los *smartphones*: todos producimos imágenes al mismo tiempo que las consumimos.

Joan Fontcuberta, acuñando un nuevo término para definir este cambio en el mundo de la imagen, afirma que vivimos en la era de la postfotografía, donde la cantidad e inmediatez de la imagen se impone a su calidad y cualidades mismas, y que necesitamos de una "ecología de lo visual" para sobrevivir a la "polución icónica" que nos invade:

> "vivimos en la imagen, y la imagen nos vive y nos hace vivir" (**Fontcuberta**, 2011).

La imagen se convierte no sólo en un medio de expresión y de prescripción de sentidos, sino en un lenguaje universal, una forma natural y espontánea de relacionarnos con los demás. Esta revalorización de la imagen también nos obliga a reflexionar sobre la importancia de:

- la construcción social de lo visual;
- comprender el acto de ver y los mecanismos de nuestro pensamiento visual;
- el desarrollo de nuestra cultura visual, de cómo aprendemos a ver, a visibilizar y a invisibilizar (**Mirzoeff**, 2016).

La hegemonía de la imagen está ya muy presente: los más de 1.000 millones de usuarios de *YouTube* consumen cada día 200 millones de horas de vídeos y producen y comparten 400 horas de contenido nuevo cada minuto (*Google*, 2016). En esta revolución icónica los jóvenes son los que están adoptando este tipo de herramientas para construir cultura visual con mayor profusión y profundidad. Primero, porque según los datos del *INE*, un 70% de los jóvenes entre 10 y 15 años ya cuentan con móvil, porcentaje que crece a más del 95% en la franja de edad entre 16 y 24 años (*INE*, 2017). Segundo, porque es una tendencia que continúa creciente: el porcentaje sigue subiendo en los últimos años. A lo que se le añade, según informes de *Pew Research Center* (**Anderson**; **Jiang**, 2018), que las plataformas sociales online más populares son aquellas con un alto componente visual en sus contenidos, como ocurre con *YouTube* (85%), *Instagram* (72%) o *Snapchat* (69%), mientras otras redes sociales enfocadas principalmente en el texto como *Twitter* o *Facebook*, más extendidas en el resto de franjas de edad, no reciben la misma atención por parte de adolescentes y jóvenes.

Por eso han surgido numerosas aplicaciones centradas en este poder de la imagen y que simplifican el proceso de producción y edición fotográfica. De ellas, *Instagram* es la que ha experimentado un mayor auge en menos tiempo y la más extendida de su clase. Además es una aplicación móvil muy enfocada a la conectividad social, por lo que *Instagram* supone para las ciencias sociales y la investigación social una ventana abierta a la observación de cómo se construye socialmente la cultura visual a nivel global desde entornos locales.

Instagram nació un 6 de octubre de 2010 para hacer más accesible la creación, edición y publicación web de fotografías mediante los *smartphones* dentro de una red social propia, como una especie de "*Twitter* con imágenes", pero con filtros prefabricados para editar rápidamente las imágenes. En junio de 2018 la red contaba con 1.000 millones de usuarios activos mensuales, de los que la mitad usan la aplicación diariamente. Su ritmo de crecimiento ha sido espectacular para tan corta vida: cada día se registran más de 730.000 nuevos usuarios. Es una red donde se comparten diariamente más de 80 millones de fotos y vídeos y con una alta presencia empresarial (*Instagram Press*, 2017). Las claves del éxito de

Instagram son su conectividad social y la rapidez en la edición y publicación de las imágenes. Su propio nombre es una fusión de *instant camera* y *telegram* y su icono original, un recuerdo de las antiguas *Polaroid*, de las que también copia la forma cuadrada de las imágenes. El algoritmo que ordena las imágenes de *Instagram* para cada usuario forma parte de la experiencia, pues ordena el *feed* o *timeline* de cada usuario en función de la afinidad del usuario, la ponderación y la antigüedad del contenido. Sacar y embellecer creativamente las fotos, su función original, es hoy una parte más del flujo comunicacional que permite la aplicación. Por sus características y evolución, algunos autores como Lev Manovich afirman que *Instagram* se ha convertido hoy en una parte fundamental de la historia de la imagen mundial y la cultura visual contemporánea (**Manovich**, 2017).

Esta investigación es una aplicación de las investigaciones y metodologías creadas por Manovich y su equipo *Software Studies Initiative* para grandes ciudades en una ciudad mediana como Granada, aprovechando el conocimiento del entorno local para contextualizar los datos recabados. Es el primer estudio en España que ha usado *Instagram* como una herramienta científica para la observación de la construcción de la cultura visual y la imagen pública de una ciudad española de tamaño medio, pues otros trabajos centrados en lo urbano han analizado grandes urbes españolas como Madrid (**Sued**, 2018), pero lo han hecho de manera comparativa con otras megalópolis, sin dotar de entidad propia a la ciudad objeto de estudio. Ofrece además una metodología sencilla y replicable que garantiza su aplicación para el estudio de cualquier otra ciudad a partir de los datos generados.

> “Detrás de una foto hay mucho más que la propia imagen. El usuario de *Instagram* está construyendo la imagen pública de su entorno”

2. Materiales y métodos

A través de este trabajo se pretende analizar la visualidad de una ciudad española concreta, Granada, mostrando simultáneamente algunas de las posibilidades que existen para el análisis de datos de *Instagram*, con el fin de descubrir la configuración y pautas de uso de la aplicación en esta ciudad. El objetivo principal es observar cómo se construye socialmente la imagen de una ciudad en *Instagram* y, por ende, la cultura visual en un entorno local. Para ello se realizan, sobre una muestra de publicaciones de *Instagram* en Granada, análisis cuantitativos de tres tipos:

- información geográfica;
- análisis de redes sociales;
- contenido visual.

Se ha tomado en especial consideración que, aunque el uso de la fotografía en investigación social está ampliamente desarrollado, el mundo digital supone un nuevo campo de aplicación de las metodologías visuales. El margen de actuación trasciende a la propia imagen, que ahora va acompañada de metadatos que completan la información disponible: disponemos del contexto de la imagen dentro de la propia imagen. Sus implicaciones tienen un alcance revolucionario para las ciencias sociales y las humanidades digitales. En su estudio de las metodologías visuales, **Gillian Rose** (2016) resumía en tres los lugares en los que podemos encontrar el significado de las imágenes:

- en la propia imagen;
- en cómo se ha producido;
- en hacia quién va dirigida.

Y para cada caso, existen tres modalidades de investigación (tecnológica, composicional y social), de tal forma que en función de la elección metodológica, hay diferentes técnicas disponibles procedentes de la semiología, el análisis de discurso o el de contenido, la interpretación composicional, el enfoque antropológico o los estudios de audiencia. *Instagram* es una herramienta única porque proporciona suficientes datos desde cualquiera de estos enfoques, y también porque reúne las suficientes características de red social para que pueda ser aprovechada para la investigación científica (**Boyd**, 2010):

- persistencia de las imágenes publicadas;
- visibilidad del contenido;
- buscabilidad permitida por la aplicación;
- alto grado de interpretabilidad de los datos generados.

No es extraño por tanto que la aplicación haya suscitado el interés académico, especialmente en los últimos años: sólo dos años después de su aparición comenzaron a aparecer trabajos de importancia que reflexionaban sobre su papel en la redefinición de la fotografía (**Champion**, 2012) o sus posibilidades para trazar ritmos visuales culturales (**Hochman**; **Schwartz**, 2012). En los siguientes años fue aumentando paulatinamente el número de trabajos publicados con *Instagram* como objeto de estudio, sobre todo artículos de investigación. Pero fue el año pasado (2017), según nuestra revisión bibliográfica, cuando se produjo una auténtica explosión investigadora sobre *Instagram*: del total de 482 trabajos revisados, 196 estudios de relevancia (más del 40%) se publicaron en 2017, ya sea como artículos (115) en revistas de impacto o de divulgación (12), capítulos (12), tesis (26), presentaciones (18) o libros (3). En 2018, aún inacabado, el número permanece alto: un total de 166 trabajos publicados.

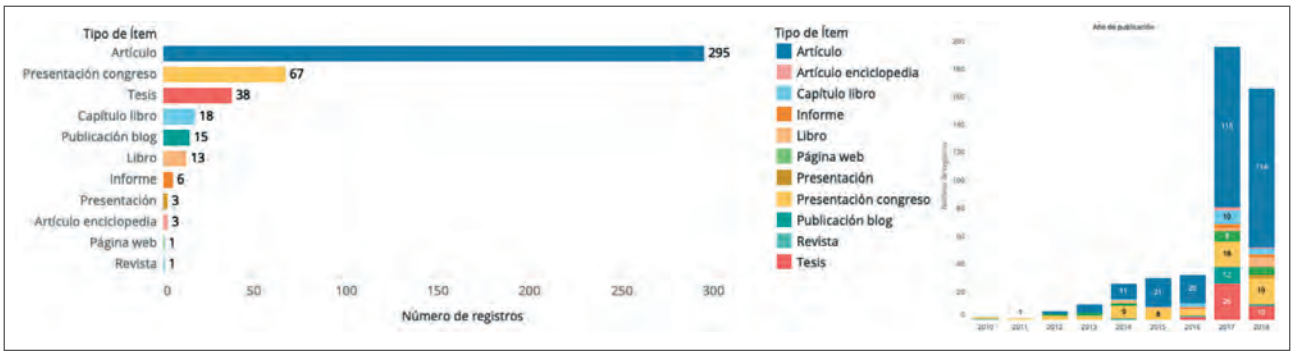


Gráfico 1. Evolución del número de trabajos publicados sobre *Instagram*, por año de publicación y tipo de ítem (2010-2018).

Durante este auge investigador, **Laestadius** (2017) propuso una clasificación de las investigaciones realizadas hasta ese momento sobre *Instagram* en cuatro áreas metodológicas:

1) aproximaciones cuantitativas desde técnicas de *big data*;

2) aproximaciones desde las humanidades digitales como el análisis cultural, que buscan aprovechar mejor la riqueza de los datos de *Instagram*;

3) pequeños conjuntos de datos analizados mediante enfoques cualitativos como el análisis de contenido;

4) el estudio directo de los usuarios de *Instagram* a través de entrevistas y trabajo etnográfico.

De estos cuatro campos, este trabajo es heredero del enfoque del análisis cultural, cuyas ideas y conceptos aplicamos aquí. Definido por Lev Manovich, el análisis cultural y los estudios de software son una forma de

> "analizar grandes conjuntos y flujos de datos de tipo cultural, usando técnicas de visualización y computacionales", centrando la atención en el "uso del procesamiento de imágenes digitales y la visualización para el análisis exploratorio de grandes colecciones de imagen y vídeo" (*Software Studies Initiative*, 2015).

Para ello, combina técnicas de *big data* con técnicas de procesamiento de imagen y visión computerizada para crear visualizaciones que usan como datos principales las propias imágenes, con la intención de

> "revelar los patrones que yacen debajo de las capacidades autónomas de la percepción humana y la cognición" (**Manovich**, 2009; 2012a; 2012b).

Estos principios han sido la base para tres proyectos que muestran la aplicación de este enfoque y han resultado clave para el desarrollo de este trabajo:

- *Phototrails*: un análisis de 2,3 millones de imágenes;
- *Selfiecity*: que exploraba los usos del selfi en cinco ciudades (**Manovich** *et al.*, 2014);
- *On Broadway*: que examina la publicación de imágenes a lo largo de la calle Broadway de Nueva York (**Manovich** *et al.*, 2015).

Estos trabajos se sirven de la ciudad como ente vivo y generador de datos, ya que más de la mitad de la población del mundo vive en áreas urbanas, en ciudades cada vez más densificadas, lo que genera un interés científico por estudiarlas, como hacen los fundamentales trabajos de **Hochman** y **Manovich** (2013); **Hochman** y **Schwartz** (2012); y **Boy** y **Uitermark** (2016; 2017).

Otros estudios centrados en lo urbano también constituyen una inspiración y un precedente relacionado con este artículo, como los centrados en estudiar:

- las propias dinámicas urbanas y el ritmo y patrones espacio-temporales y visuales de las ciudades (**Rodríguez-Domínguez** *et al.*, 2017; **Xia** *et al.*, 2014);
- desigualdades sociales y digitales (**Indaco**; **Manovich**, 2016; **Shelton**; **Poorthuis**; **Zook**, 2015);
- puntos turísticos de interés (**Li** *et al.*, 2015; **Mukhina**; **Rakitin**; **Visheratin**, 2017);
- el cambio urbano (**Lazzarini**; **López-Baeza**, 2017) y los procesos de gentrificación (**Huang**, 2017);
- estudio de eventos locales concretos (**Giridhar**; **Abdelzaher**, 2017; **Liu**; **Jansson**, 2017);
- mapeado del *street art* y el graffiti urbano (**Honig**; **MacDowall**, 2017; **MacDowall**; **De-Souza**, 2017).

Debido a las actuales dificultades de acceso a la API (interfaz de programación de aplicaciones) de *Instagram*, ha sido necesario recurrir a *Netlytic* (**Gruzd**, 2016), un servicio mantenido por el *Social Media Lab* de la *Universidad de Ryerson*, para la recolección de datos a partir de unas coordenadas geográficas y durante un período de un año (abril de 2017-abril de 2018).

El conjunto de datos resultante contiene un total de 955.564 observaciones (560.716 publicaciones y 394.848 comentarios) con distintas variables:

- coordenadas de geolocalización;
- urls de la publicación y la imagen;
- fecha;
- usuario-autor;
- usuarios que han interactuado con la publicación;
- texto de la publicación;
- número de *likes*;
- nombre del filtro usado.

Estas variables pueden usarse para analizar diversos aspectos de las publicaciones, y fue lo que hicimos tratando los datos a través del lenguaje *R* (*R Development Core Team*, 2008).

Para el análisis y visualización de datos de manera general se usó también el software *Tableau Desktop*.

Para el análisis geográfico se dividió la variable de coordenadas en las de latitud y longitud, y se trataron los datos con el software *QGIS 3.2* (*QGIS Development Team*, 2018), de tal forma que fusionando estos datos con los archivos de polígonos que delimitan barrios y distritos administrativos de la ciudad, se crearon dos nuevas columnas para cada dato que incluía el barrio y distrito al que pertenecía cada publicación.

Para el análisis de redes sociales y su visualización se creó un subconjunto de datos con las variables de autor-usuario y la de usuarios que interactuaron con cada publicación, de tal forma que pudo importarse al software *Gephi* (*Gephi Consortium*, 2018), que requiere una tabla simple con ambas variables (a modo de origen y destino de la aplicación).

Para el análisis del *corpus* textual se usaron *Meaning Cloud* (un *add-on* para *Excel* de *Windows*) y una de las herramientas de *Google Spreadsheets*, que identifica el lenguaje usado en cada publicación.

Para el análisis visual se descargaron 355.797 imágenes mediante un *script* de *Python*, que han sido analizadas mediante el software *ImageJ* (**Rueden** *et al.*, 2017; **Schneider**; **Rasband**; **Eliceiri**, 2012), procesándose una muestra de ellas, de unas 10.000 imágenes a través de *Google Cloud Vision*.

3. Resultados

3.1. Análisis geográfico

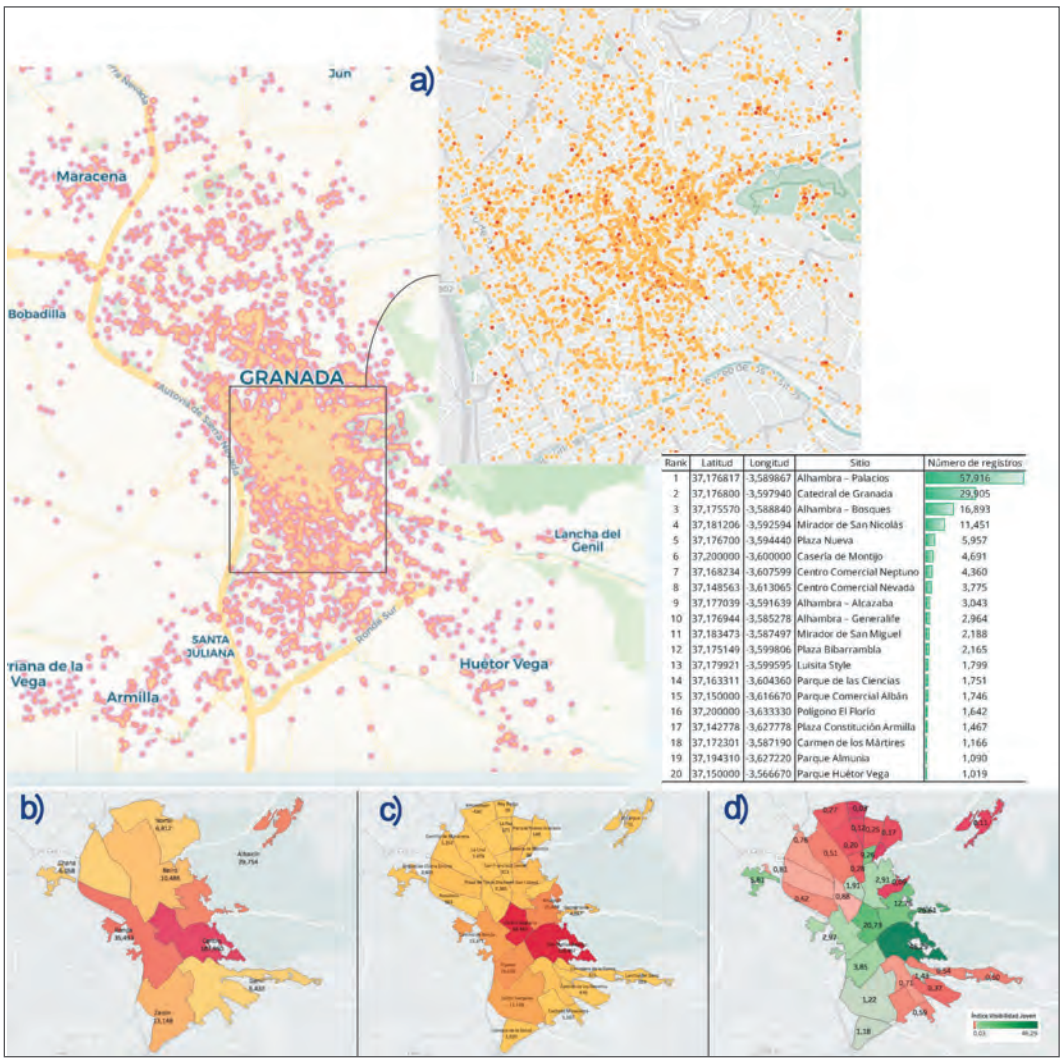
El primer análisis fue para localizar los principales puntos visuales de interés en Granada, mostrados en el gráfico 2. Para ello se escogieron los sitios que concentraban más de 500 imágenes de la muestra, resultando un *ranking* que comprende negocios, parques comerciales, parques públicos, lugares emblemáticos y espacios deportivos y culturales. Los casi 40 puntos visuales del ranking concentran más de la mitad (51,63%) de las imágenes contenidas en la muestra. No es sorprendente que la Alhambra, el monumento más visitado de España, sea el punto que atraiga las miradas de los *instagramers* en una ciudad que define su identidad a partir de dicho complejo monumental. Un 31,4% de las imágenes se realizaron dentro de, o a lugares del conjunto nazarí, como los palacios (17,67%), los bosques de la Alhambra (5,16%), la Alcazaba, el Generalife, o sitios que están muy relacionados visualmente con el monumento, como la Plaza Nueva, los miradores de San Nicolás o San Miguel, el Carmen de los Mártires, el Hotel Alhambra Palace o el Paseo de los Tristes.

> El análisis de las publicaciones de *Instagram* en Granada es útil para visibilizar el impacto del turismo: una de cada cinco imágenes (22,71%) se realizan desde la Alhambra o zonas aledañas

El siguiente monumento en importancia es la Catedral de Granada (9,13%), que también comprende la Plaza de las Pasiegas y la Capilla Real. Lugares de ocio y compras como el Parque Comercial Albán y los Centros Comerciales Neptuno y Nevada concentran un 3% de las fotos. Otros lugares de interés para el usuario de *Instagram* son algunos espacios públicos de encuentro como la Plaza Bibarrambla, cerca de la Catedral y en pleno centro (0,66%), el Parque Almunia (0,33%) o recintos como el Parque de las Ciencias (0,53%).

En total, los lugares culturales concentran en torno al 30% de la muestra, mientras los lugares de interés paisajístico suponen un 15% y los relacionados con negocios de todo tipo con fuerte presencia visual, un 5,37%.

Al segmentar los datos por barrios y distritos administrativos queda clara la concentración zonal de las publicaciones: eliminando las 227.340 publicaciones que se han geolocalizado de manera genérica en Granada (esto es, no han especificado unas coordenadas concretas), el distrito Centro concentra el 56,38% de las publicaciones, seguido muy de lejos de los distritos limítrofes: Ronda (10,65%) y Albaicín (8,93%). El resto de distritos concentra el restante 12,5% del total de publicaciones, repartido entre los distritos del Zaidín (3,94%), Beiro (3,15%), Norte (2,04%), Chana (1,82%) y Genil (1,63%). En el distrito Otros se ha concentrado el resto de publicaciones (11,47%), geolocalizadas en las zonas que conforman el denominado cinturón metropolitano de la ciudad, y al que pertenecen municipios como Armilla, Maracena o Huétor Vega.



Rank	Latitud	Longitud	Sitio	Número de registros
1	37,176817	-3,589867	Alhambra - Palacios	57,916
2	37,176800	-3,597940	Catedral de Granada	29,905
3	37,175570	-3,588840	Alhambra - Bosques	16,893
4	37,181206	-3,592594	Mirador de San Nicolás	11,451
5	37,176700	-3,594440	Plaza Nueva	5,957
6	37,200000	-3,600000	Casería de Montijo	4,691
7	37,168234	-3,607599	Centro Comercial Neptuno	4,360
8	37,148563	-3,613065	Centro Comercial Nevadá	3,775
9	37,177039	-3,591639	Alhambra - Alcazaba	3,043
10	37,176944	-3,585278	Alhambra - Generalife	2,964
11	37,183473	-3,587497	Mirador de San Miguel	2,188
12	37,175149	-3,599806	Plaza Bibarrambla	2,165
13	37,179921	-3,599595	Luisita Style	1,799
14	37,163311	-3,604360	Parque de las Ciencias	1,751
15	37,150000	-3,616670	Parque Comercial Albán	1,746
16	37,200000	-3,633330	Polígono El Florio	1,642
17	37,142778	-3,627778	Plaza Constitución Armilla	1,467
18	37,172301	-3,587190	Carmen de los Mártires	1,166
19	37,194310	-3,627220	Parque Almunia	1,090
20	37,150000	-3,566670	Parque Huétor Vega	1,019

Gráfico 2. Principales puntos visuales de Granada según los usuarios de *Instagram*. (a) Mapa de calor general; (b y c) mapas coropléticos de frecuencias y (d) de índice de visibilidad, por barrios y distritos y ranking de puntos.

Los barrios con mayor número de publicaciones son los dos pertenecientes al distrito Centro:

- San Matías Realejo (al que pertenece la Alhambra), con el 35,84% de las publicaciones;
- Centro Sagrario, donde su ubica el casco histórico alrededor de la Catedral, con el 20,54%.

El tercero en importancia es el barrio del Albaicín, que forma parte anexa del conjunto monumental de Alhambra y Generalife, con un 7,63%, seguido por el barrio Fígares (5,86%), donde se ubican puntos relevantes de la ciudad como el Parque de las Ciencias, o el Museo Caja Granada Memoria de Andalucía, y el barrio Camino de Ronda (4,61%), que recientemente ha experimentado una renovación del tejido comercial, ambos del distrito Ronda. Es destacable también la importancia del histórico barrio de Zaidín Vergeles (3,94%), debido en parte a la revitalización por la creación del Parque Tecnológico de la Salud.

Usando los datos del *Padrón Municipal de Habitantes* de 2016, podemos relacionar el número de publicaciones con el número de empadronados, creando un índice de visibilidad que muestra si el barrio está sobrerrepresentado o infrarrepresentado visualmente. Por ejemplo, el barrio del Sacromonte, dentro del Albaicín, con un índice de visibilidad de 26,6, resulta bastante visible en *Instagram* por sus características turísticas en comparación con la poca gente que está empadronada allí. Los barrios de Campo Verde en distrito Norte, con un índice de 9,26, o el de Bobadilla en distrito Chana, con 5,8 son barrios poco poblados pero muy representados visualmente, tal vez por su población más joven. Todo lo contrario que los barrios de Fígares y Camino de Ronda, barrios bastante envejecidos demográficamente, pero con un alto índice de visibilidad por su carácter céntrico.

3.2. Análisis de redes sociales (SNA)

Usando dos variables del *dataset* generado podemos también analizar y visualizar las redes generadas entre los usuarios de *Instagram* en Granada. En cuestión de número de publicaciones, los usuarios de *Instagram* más activos son:

- Michael Goatherd (*S&M Escuela de Modelos* de Granada);
- Josep Dely (fotógrafo especializado en temática gay);

- *El Capricho* (complejo hotelero y gastronómico);
- Juant Cyan (particular);
- *Artesanía del Árbol* (productos de taracea);
- *Granada Tatto Shop* (tatuajes);
- *Luna Lorquiana* (reportajes de viajes).

El ranking cambia si se tienen en cuenta los comentarios a otras publicaciones, lo que indica que es posible posicionarse en la red sin necesidad de publicar imágenes.

Según las métricas propias del análisis de redes sociales (**Pérez-Solà**; **Casas-Roma**, 2016), y tratada como una red no dirigida con 344.433 nodos (usuarios), conectados a través de 394.848 aristas (conexiones entre usuarios), la red de usuarios de *Instagram* en Granada está configurada como una red con un diámetro medio de 10 y una distancia media de camino de 4,499, es decir, una red muy dispersa pero cuyos usuarios no están muy separados entre sí, además de que es poco densa (0,075). Unido a un coeficiente de *clustering* de 0,455, podríamos hablar de que la red de usuarios de *Instagram* en Granada tiene características de "mundo pequeño".

Hay varias métricas para analizar la centralidad de nodos de la red y su relación con los clusters de usuarios, visibles en las redes del gráfico 3. Según la de vector propio (*Eigenvector*) el nodo central sería Mauro Tittarelli, pero esta centralidad se debe a su número de conexiones a través de sus interacciones, no a sus publicaciones, por lo que no resulta una métrica fiable para nuestros intereses al no tener en cuenta las "corredurías" (*brokerage*). Para evitar esta distorsión y detectar los nodos verdaderamente relevantes de la red hemos analizado también*:*

- la intermediación (usuarios que conectan subcomunidades)*;*
- la proximidad o cercanía (mejores situaciones con respecto al resto).

El nodo con más intermediación es @granadademoda, una cuenta que promociona diversos negocios relacionados con moda, salud y belleza. La moda se configura como un tema importante en la red granadina, ya que en torno a ella surgen otros nodos importantes, como @luisitastyle, bloguera de moda que concentra en torno a ella a otros usuarios como @tetecaballero, @anasaez___, @mirrorofstyle, @entretelasyretales o @marikillaruiz que se posicionan a su vez como nodos con altos valores de intermediación. Otros usuarios *gatekeepers* son*:* @srtahachi, @yamilaweed, @antonyzetta, @joselop44, @iharvera y @jcarlosisla, poseedores de un estilo propio, estéticamente atractivo y que conecta bien con el resto de usuarios. Serían el prototipo de *instagramers* que podrían considerarse dentro del estilo del *instagramismo* (**Manovich**, 2017).

De los usuarios con altos índices en proximidad encontramos a @vinilavonbismarck (cantante y artista granadina), @vigorexito (modelo), @pirugozadera (locutor de radio), @valen_viajescoquitos (blogueros de viajes), @pinkrunnersmagazine (*running*) o @paula_videostar (*youtuber*). Todos ellos tienen en común que realmente no están conectados directamente con Granada (excepto la primera mencionada, Vinila) y tienen un buen número de seguidores (presumiblemente muchos de ellos en Granada). Excepciones son @discosmarcapasos, una histórica tienda de vinilos de Granada y @mariaamoraleess, una joven que con sólo dos publicaciones y más de 1.800 seguidores se posiciona como nodo cercano en la red analizada.

3.3. Análisis de contenido (textual y visual)

Una de las características distintivas de *Instagram* son los filtros. Sin embargo, el 70% de los usuarios de *Instagram* en Granada no usó ningún filtro en absoluto (o bien ha editado la foto en otra aplicación o programa externo). De las publi-

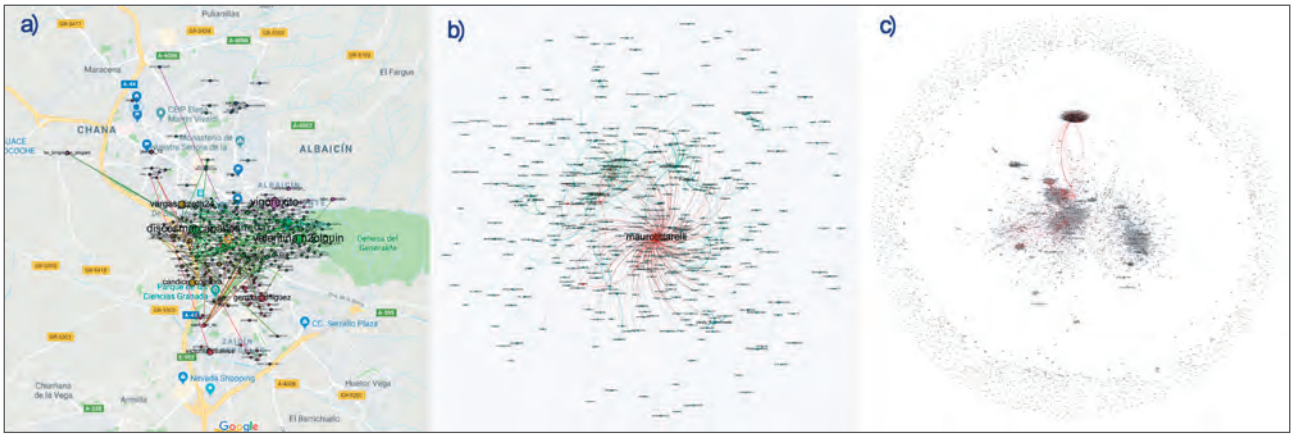


Gráfico 3. Distintas configuraciones de la red de usuarios de *Instagram* filtradas por centralidad de vector propio: (a) red simple (320 usuarios) visualizada por medida central de proximidad y geolocalización; (b) red simple (320 usuarios), por intermediación y grado mediante distribución *Yifan Hu Proporcional*; (c) red compleja (5.861 usuarios), por intermediación y proximidad mediante distribución *Force Atlas 2*. Versión completa en alta resolución para descargar y explorar en: *http://bit.ly/redinstagramGRX*
Fuente: elaborado en *Gephi*.

caciones que sí lo usan, un 22,21% usó el filtro *Clarendon*, seguido de los filtros *Lark* (8,72%) y *Juno* (8,07%). Otros muy utilizados son *Ludwig* (6,31%), *Gingham* (6,15%), *Lo-fi* (4,40%) y *Valencia* (4,10%). El uso de filtros por barrios y distritos es relativamente similar.

En cuanto a las palabras y "hashtags" más usados aparecen, en primer lugar, #granada, #spain y #alhambra, seguidos de #travel, #andalucia y #españa. Los hashtags más usados a nivel mundial aparecen también entre los más frecuentes en Granada: #instagood (el segundo más usado en todo el mundo), #love (el primero), #photooftheday (el tercero), #beautiful (el quinto) y #fashion (el cuarto). También son destacables #picoftheday (que es el undécimo más usado mundialmente) o #photography (el trigésimo séptimo). Más propios de la ciudad granadina parecen #travelgram, #travelphotography, #architecture, #instatravel o #wanderlust (pasión por viajar).

Una detección realizada sobre las primeras 3.000 publicaciones más puntuadas con *likes* revela que la mayoría del texto (46%) está en español, seguido del inglés (23%). El resto de idiomas usados son mucho más minoritarios: italiano (3,23%), catalán (2,78%), portugués (2,65%), francés (2,31%), ruso (1,26%) o coreano (1,18%).

Analizando el contenido de las imágenes, a través de una muestra de 10.000 imágenes procesadas mediante *Google Cloud Vision*, podemos establecer también un ranking de los contenidos más compartidos en *Instagram* en Granada. En la tabla 1 pueden verse las 20 etiquetas más frecuentes en el contenido detectado, por cómputo total y por la primera etiqueta detectada (algunas imágenes recibían hasta 10 etiquetas de contenido, mientras otras podían recibir una o dos). Las etiquetas "árbol" y "cielo" son de las más repetidas en ambas categorías, así como las que tienen que ver con elementos arquitectónicos ("calle", "ventana", "fachada", "edificio", "muro", "propiedad") y paisajes urbanos ("ciudad", "municipio", "lugar histórico" o "lugar conocido").

Se refuerza la importancia del turismo en las imágenes compartidas, al aparecer repetidamente etiquetas como "turismo", "vacaciones", "diversión" o "viaje". También son frecuentes las imágenes donde aparece gente, como ocurre con las etiquetas "gafas", "ropa" o "pelo" (elementos personales).

A través del *software* creado por Manovich y su equipo, *ImageJ* (**Rueden** *et al.*, 2017; **Schneider** *et al.*, 2012) pueden generarse *imagengramas* como los mostrados en el gráfico 4, que visualizan 100.000 imágenes de la muestra ordenadas, radial y linealmente, por sus valores medianos (menos sensibles que el promedio a las variaciones de *outliers*) de brillo y matiz. Al pasar dicha visualización por la plataforma web *Adobe Color*, obtenemos las tonalidades predominantes, la paleta cromática característica de Granada, en cinco variaciones: colores plenos, colores claros, apagados, intensos y oscuros.

Tabla 1. Rankings de usuarios, hashtags, filtros más usados y etiquetas extraídas de las publicaciones de *Instagram* en Granada

	Usuarios	Hashtags	Filtros	Etiquetas (todas)		Etiqueta (1er lugar)	
1	michaelgoatherd	#granada	Clarendon	Árbol	3.739	Cielo	1.404
2	josepdely	#spain	Lark	Cielo	3.625	Fotografía	561
3	elcapricho	#alhambra	Juno	Chica	3.009	Árbol	418
4	juantcyan	#travel	Ludwig	Diversión	2.312	Municipio	404
5	artesania_del_arbol	#andalucia	Gingham	Edificio	2.136	Arquitectura	351
6	granadatattooshop	#españa	Lo-fi	Planta	2.113	Pelo	349
7	lunalorquiana	#instagood	Valencia	Ciudad	2.067	Planta	337
8	ikkaaccesorios	#picoftheday	Mayfair	Turismo	2.021	Lugar conocido	297
9	avefenix16177	#love	X-Pro II	Vacaciones	1.619	Azul	270
10	sietepecadostatoogranada	#photography	Hefe	Producto	1.480	Lugar histórico	234
11	et_nuit	#photooftheday	Aden	Entretenimiento	1,320	Gafas de sol	227
12	rg_life.style	#travelgram	Slumber	Fachada	1.319	Flor	225
13	garciaillana	#architecture	Crema	Municipio	1.262	Edificio	220
14	ingriddelacruz_oficial	#beautiful	Rise	Ventana	1.166	Ropa	204
15	discosmarcapasos	#instatravel	Amaro	Imagen	1.151	Gafas	200
16	ransett	#travelphotography	Inkwell	Lugar histórico	1.140	Blanco	174
17	borealrelax	#wanderlust	Sierra	Calle	1.053	Texto	173
18	sony_liz23	#art	Moon	Arquitectura	1.032	Muro	172
19	mayte_lovenevada	#europe	Reyes	Fotografía	961	Noche	161
20	maribelcaffeina	#photo	Hudson	Viaje	927	Ciudad	158

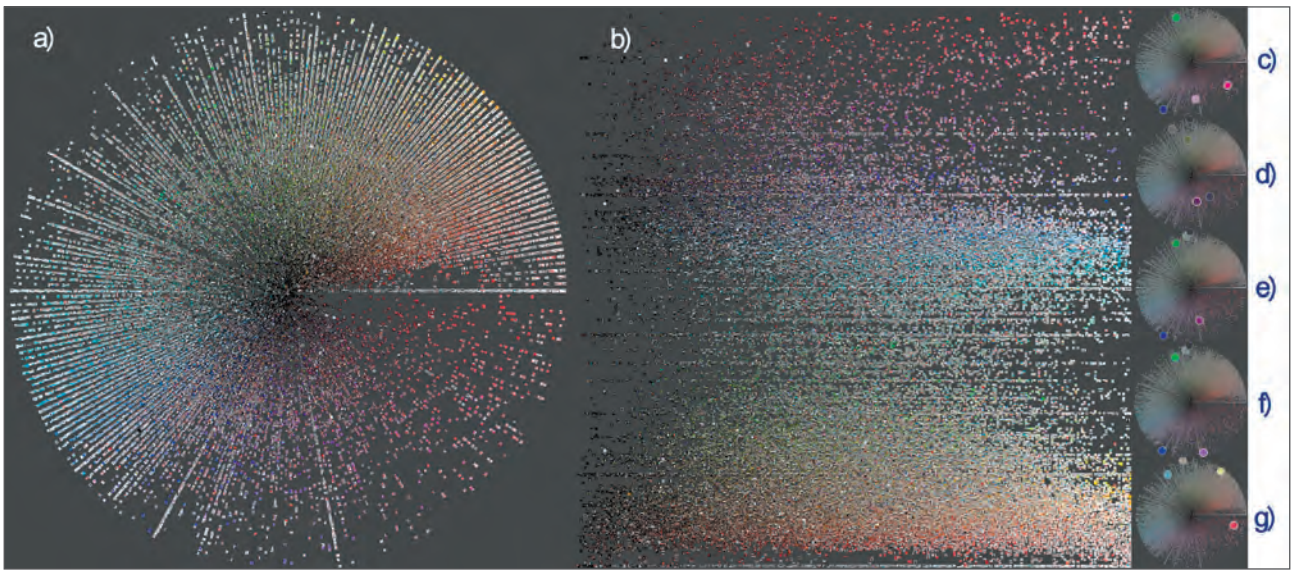


Gráfico 4. *Imagengramas* de Granada, radial (a) y lineal (b), realizadas a partir de 100.000 imágenes por sus valores de brillo y tono; y extracción de su paleta cromática (c, d, e, f y g) en variaciones de colores plenos, colores claros, apagados, intensos y oscuros.
Disponible versión en alta resolución en: *http://bit.ly/ig_grx_01*

4. Discusión y conclusiones

Este análisis de las publicaciones de *Instagram* en Granada es útil para visibilizar el impacto e incidencia del turismo en la ciudad. El atractivo visual de la Alhambra es sobradamente conocido, y numerosos puntos se sirven de ella como escaparate: una de cada cinco imágenes (22,71%) que se publican en *Instagram* desde Granada se realizan desde la Alhambra o alguna de sus zonas aledañas, como el Mirador de San Nicolás, desde el que se obtiene una de las mejores vistas de la ciudad palatina andalusí. Junto con la Catedral de Granada y el centro histórico a su alrededor, que concentra también numerosas imágenes, son los dos polos sobre los que se construye la imagen de la ciudad en *Instagram*. De ahí que los barrios con mayor número de imágenes geolocalizadas en sus demarcaciones sean los del distrito Centro, en el que están situados ambos monumentos. Los barrios limítrofes, Ronda y Albaicín, se benefician de esta atracción, captando también la atención de los usuarios de *Instagram*, aunque en menor medida. El número de publicaciones en el resto de barrios va disminuyendo por lejanía hacia estos focos, lo que nos habla de un modelo de centro-periferia en la atracción y construcción visual de la imagen de Granada en *Instagram*. Esta importancia del turismo también es visible en los *hashtags* usados en las publicaciones, que incluyen, no sólo el nombre de la propia ciudad en el primer puesto, o el de la Alhambra en el tercero, sino otras referencias a viajes, turismo y arte. Otros datos geográficos también hablan de la revitalización visual de ciertas zonas y barrios, con población mayoritariamente envejecida, pero que está muy visibilizada, como el barrio de Fígares o Camino de Ronda. La desigualdad visual de Granada por barrios y distritos queda patente.

En cuanto a la red de usuarios de *Instagram*, hemos podido ver que es una red extensa, de gran diámetro, pero cuya poca distancia y mediano coeficiente de *clustering* la configuran como del tipo "pequeño mundo"; es decir, una red cuyos usuarios no están muy separados entre sí, aunque sea poco densa, y que muestra algunos clusters en torno a moda, deporte y salud. Se ha visto cómo la importancia de los usuarios para la red no radicaba en su centralidad de grado o vector, es decir, en su número de conexiones con otros usuarios, sino que resultaban más esenciales para el flujo de conexiones aquellos usuarios que actuaban como intermediarios entre los diversos clusters y que encajaban mejor en los ideales del *instagramer*, con perfiles estéticamente muy cuidados, cercanos al estilo *instagramista*, y muy atractivos visualmente, capaces de aprovechar todo el potencial de la aplicación. Otros usuarios, con valores altos de cercanía, también se constituyen en nodos de conexión con el resto de usuarios por su prestigio dentro de la red social de *Instagram*.

Si interpretamos el uso del filtro *Normal* como la ausencia de filtro (o que forme parte de la tendencia generalizada de no usar filtro alguno intencionadamente), podríamos afirmar entonces que tan sólo un 30% de los usuarios decide ser creativo con sus imágenes.

El amplio uso del idioma inglés también habla de la internacionalización de las publicaciones de *Instagram* en Granada, y de nuevo, el peso del turismo en la ciudad, algo que también resulta visible al analizar el contenido de las imágenes compartidas. El contenido compartido igualmente habla de la importancia de la fotografía paisajística y arquitectónica, por encima de la personal.

La colorimetría aplicada a las imágenes muestra también cómo los colores predominantes, azules oscuros intensos y verdes naturales, podrían estar asociados a imágenes compuestas de elementos naturales como árboles y cielo (que también aparecen en las etiquetas de contenido), así como los colores más rosados y anaranjados pueden asociar-

se, principalmente, a imágenes con contenido de rostros y espacios arquitectónicos de la Alhambra.

La investigación que se ha mostrado aquí es un avance de los resultados que se están consiguiendo sobre la base del *dataset* principal mencionado al principio del artículo. Al mismo tiempo que se recopilaban estos datos, se han recolectado simultáneamente otros *datasets* de otras ciudades (Málaga, Madrid y Barcelona), con los que se pretende aplicar la misma metodología en ciudades de tamaño superior. El mapa comparativo resultante puede verse en el gráfico 5, en el que se muestran los principales puntos visuales de estas cuatro ciudades.

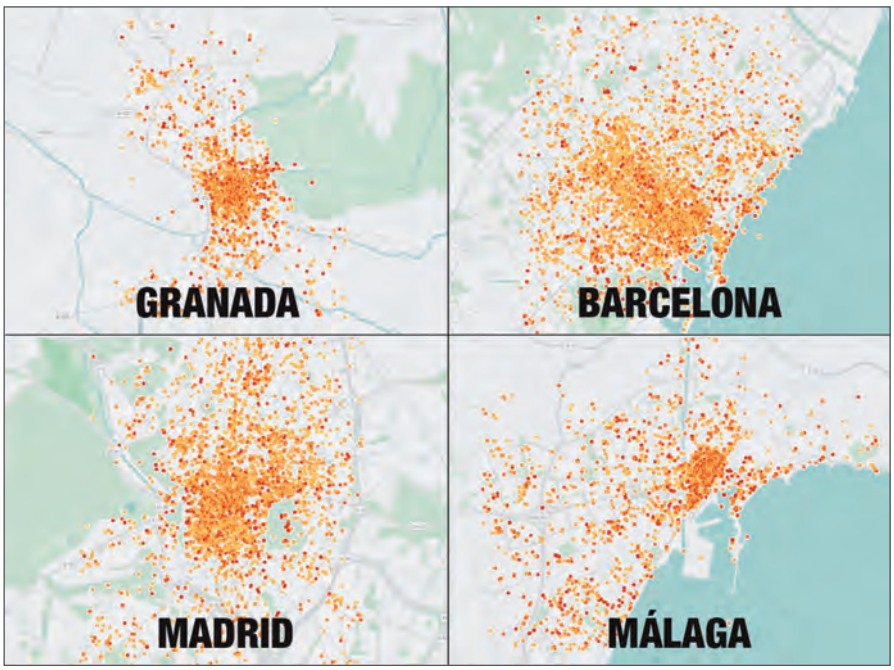


Gráfico 5. Principales puntos visuales de Granada, Barcelona, Madrid y Málaga según los usuarios de *Instagram*

Como puede verse, la distribución de puntos en cada ciudad es muy diferente. Si en el caso de Granada, como hemos visto, hay una alta concentración de puntos en torno a la Alhambra y el centro histórico de la ciudad, en el caso de Málaga, los primeros análisis muestran que la mayoría de puntos se concentra en torno a una zona recientemente revitalizada de la ciudad: el puerto, cuyo atractivo visual y urbano concentra, al igual que la Alhambra, una de cada tres publicaciones realizadas. Por otro lado, en el caso de grandes urbes como Madrid y Barcelona, los puntos visuales de interés son monumentos como la Sagrada Familia, el Palacio Real o el Parque Güell, pero también lugares emblemáticos como la Puerta del Sol o la Plaza Catalunya, o especialmente, el Santiago Bernabeu, estadio del Real Madrid, que tiene la cuenta española en *Instagram* con más seguidores del mundo: 65 millones.

Queda claro pues que un monumento de la importancia visual de la Alhambra determina fuertemente la construcción de la visualidad mediante medios sociales en una ciudad como Granada, pero esta distorsión no debería ser tan fuerte en otras ciudades cuyos monumentos no tengan ese atractivo o cuyo interés esté más repartido por las distintas zonas urbanas, de tal forma que la foto que pueda establecerse de esta ciudad sea más fiel al interés visual de los ciudadanos que la habitan y los turistas que la visitan.

El acto de compartir una imagen no es trivial, puede parecer simple interés por recordar algo bello, capturar algo espontáneo o comunicar algo a los demás, pero hay algo más: el usuario de *Instagram* construye su identidad mediante las imágenes que comparte y también está construyendo la imagen pública de su entorno, visualizando públicamente un objeto, momento, lugar o persona, dándole visibilidad ante los demás y con ello, invisibilizando a la vez otros elementos. Por tanto está poniendo su granito de arena local en el inacabado proyecto de construcción de la cultura visual global y urbana. Detrás de una foto hay mucho más que la propia imagen. Los análisis realizados son útiles para ver precisamente cómo se construye socialmente la imagen de una ciudad como Granada cuando se comparte en aplicaciones como *Instagram*, por lo que este trabajo puede ser útil para el desarrollo de futuras líneas de estudio que amplíen esta investigación, ya sea con el uso complementario de otras redes como *Twitter* o *Facebook*, o con un mayor aprovechamiento de los propios datos de *Instagram*, por ejemplo, los datos temporales o la identificación de perfiles mediante técnicas de *machine learning*. Creemos en definitiva que este trabajo muestra el potencial de los datos generados por los medios sociales para analizar cómo se construye la imagen y la visualidad de una ciudad.

5. Referencias

Anderson, Monica; **Jiang, Jingjing** (2018). *Teens, social media & technology 2018*. Pew Research Center.
http://assets.pewresearch.org/wp-content/uploads/sites/14/2018/05/31102617/PI_2018.05.31_TeensTech_FINAL.pdf

Boy, John D.; **Uitermark, Justus** (2016). "How to study the city on Instagram". *PLoS one*, v. 11, n. 6.
https://doi.org/10.1371/journal.pone.0158161

Boy, John D.; **Uitermark, Justus** (2017). "Reassembling the city through Instagram". *Transactions of the Institute of British Geographers*, v. 42, n. 4, pp. 617-624.
https://doi.org/10.1111/tran.12185

Boyd, Danah (2010). "Social network sites as networked publics: Affordances, dynamics, and implications". In: Papacharissi, Zizi (ed.). *A networked self: Identify, community, and culture on social network sites*. Routledge, pp. 47-66. ISBN: 978 0 415801812
https://www.danah.org/papers/2010/SNSasNetworkedPublics.pdf

Champion, Charlotte (2012). "Instagram: Je-suis-là?". *Philosophy of photography*, v. 3, n. 1, pp. 83-88.
http://cenestpas.com/wp-content/uploads/2013/08/Charlotte-Champion-Instagram-je-sui-la.pdf
https://doi.org/10.1386/pop.3.1.83_7

Fontcuberta, Joan (2011). "Por un manifiesto posfotográfico". *Suplemento cultural de La Vanguardia*.
http://www.lavanguardia.com/cultura/20110511/54152218372/por-un-manifiesto-posfotografico.html

Gephi Consortium (2018). *Gephi*.
https://gephi.org

Giridhar, Prasanna; **Abdelzaher, Tarek** (2017). "Visualization of events using Twitter and Instagram". In: *2017 IEEE Intl conf on pervasive computing and communications workshops (PerCom workshops)*, 13-17 March, pp. 82-84.
https://doi.org/10.1109/PERCOMW.2017.7917530

Google (2016). "La revolución del vídeo en la era del contenido audiovisual". *Think with Google*, mayo.
http://bit.ly/2Isl6mI

Gruzd, Anatoliy (2016). *Netlytic: Software for automated text and social network analysis*.
https://netlytic.org

Hochman, Nadav; **Manovich, Lev** (2013). "Zooming into an Instagram city: Reading the local through social media". *First Monday*, v. 18, n. 7.
https://doi.org/10.5210/fm.v18i7.4711

Hochman, Nadav; **Schwartz, Raz** (2012). "Visualizing Instagram: Tracing cultural visual rhythms". In: *Procs of the Workshop on social media visualization (SocMedVis) in conjunction with the Sixth intl AAAI conf on weblogs and social media (ICWSM–12)* (pp. 6-9).
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.851.377&rep=rep1&type=pdf

Honig, Christopher D. F.; **MacDowall, Lachlan** (2017). "Spatio-temporal mapping of street art using Instagram". *First Monday*, v. 22, n. 3.
https://doi.org/10.5210/fm.v22i3.7072

Huang, Cheng-Chia (2017). *Developing a data mining framework to identify a sense of gentrification through social media data: A case study using Instagram posts in Salt Lake City, Utah*. Thesis presented to the Faculty of San Diego State University. Ann Arbor, United States.
https://search.proquest.com/docview/1987947655/abstract/68F7D62F1CCB4318PQ/1

Indaco, Agustin; **Manovich, Lev** (2016). "Urban social media inequality: definition, measurements, and application". *arXiv preprint arXiv:1607.01845*.
https://arxiv.org/abs/1607.01845

INE (2017). *Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares 2017*.
https://www.ine.es/prensa/tich_2017.pdf

Instagram Press (2017). *Our story.*
https://instagram-press.com/our-story

Jenkins, Henry (2008). *Convergence culture. La cultura de la convergencia de los medios de comunicación.* Barcelona: Paidós. ISBN: 978 84 493 2153 5
https://stbngtrrz.files.wordpress.com/2012/10/jenkins-henry-convergence-culture.pdf

Laestadius, Linnea (2017). "Instagram". In: Sloan, Luke; Quan-Haase, Anabel (eds.). *The SAGE handbook of social media mesearch methods*. SAGE, pp. 573-592. ISBN: 978 1 4739 1632 6

Lazzarini, Luca; **López-Baeza, Jesús** (2017). "La strada dei funghi: studiare la percezione delle trasformazioni urbane attraverso Instagram". *Urbanistica*, n. 157, pp. 50-54.
https://bit.ly/2AggoW6

Li, Xutao; **Pham, Tuan-Anh-Nguyen**; **Cong, Gao**; **Yuan, Quan**; **Li, Xiao-Li**; **Krishnaswamy, Shonali** (2015). "Where you Instagram?: Associating your Instagram photos with points of interest". In: *Proceedings of the 24th ACM Intl on conf on information and knowledge management*. New York, NY, USA: ACM, pp. 1231-1240. ISBN: 978 1 4503 3794 6
https://doi.org/10.1145/2806416.2806463

Liu, Shuhua; **Jansson, Patrick** (2017). *City event identification from Instagram data using word embedding and topic.* Arcada working papers, n. 7. ISBN: 978 952 5260 85 4
https://www.theseus.fi/bitstream/handle/10024/140582/AWP_7-2017_Liu_Jansson.pdf?sequence=1&isAllowed=y

MacDowall, Lachlan-John; **De-Souza, Poppy** (2017). "'I'd double tap that!!': street art, graffiti, and Instagram research". *Media, culture & society*, v. 40, n. 1, pp. 3-22.
https://doi.org/10.1177/0163443717703793

Manovich, Lev (2009). "How to follow global digital cultures, or cultural analytics for beginners". In: Becker, Konrad; Stalder, Felix. *Deep search: The politics of search beyond Google. Wien: Studien Verlag*, pp. 198-211.
http://manovich.net/index.php/projects/how-to-follow-global-digital-cultures

Manovich, Lev (2012a). "How to compare one million images?". In: Berry, David M. *Understanding digital humanities*. London: Palgrave Macmillan, pp. 249-278. ISBN: 978 0 230 29265 9
https://doi.org/10.1057/9780230371934_14

Manovich, Lev (2012b). "Media visualization: Visual techniques for exploring large media collections". *Media studies futures*, pp. 1-21.
http://manovich.net/index.php/projects/media-visualization-visual-techniques-for-exploring-large-media-collections
https://doi.org/10.1002/9781444361506.wbi

Manovich, Lev (2017). *Instagram and contemporary image*.
http://manovich.net/index.php/projects/instagram-and-contemporary-image

Manovich, Lev; **Goddemeyer, Daniel**; **Stefaner, Moritz**; **Baur, Dominikus** (2015). *On Broadway*.
http://on-broadway.nyc

Manovich, Lev; **Stefaner, Moritz**; **Yazdani, Mehrdad**; **Baur, Dominikus**; **Goddemeyer, Daniel**; **Tifentale, Alise** (2014). *Selfiecity*.
http://selfiecity.net

Mirzoeff, Nicholas (2016). *Cómo ver el mundo: una nueva introducción a la cultura visual*. Madrid: Paidos. ISBN: 978 84 493 3216 6

Mukhina, Ksenia D.; **Rakitin, Stepan V.**; **Visheratin, Alexander A.** (2017). "Detection of tourists attraction points using Instagram profiles". *Procedia computer science*, v. 108, pp. 2378-2382.
https://doi.org/10.1016/j.procs.2017.05.131

Pérez-Solà, Cristina; **Casas-Roma, Jordi** (2016). *Análisis de datos de redes sociales*. Editorial UOC. ISBN: 978 84 91163664

QGIS Development Team (2018). *QGIS Geographic Information System. Open Source Geospatial Foundation Project.*
https://qgis.org

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
http://www.R-project.org

Rodríguez-Domínguez, Daniel; **Díaz-Redondo, Rebeca P.**; **Fernández-Vilas, Ana**; **Ben-Khalifa, Mohamed** (2017). "Sensing the city with Instagram: Clustering geolocated data for outlier detection". *Expert systems with applications*, v. 78, pp. 319-333.
https://doi.org/10.1016/j.eswa.2017.02.018

Rose, Gillian (2016). *Visual methodologies: An introduction to researching with visual materials*. Sage. ISBN: 978 1 473948907

Rueden, Curtis T.; **Schindelin, Johannes**; **Hiner, Mark C.**; **DeZonia, Barry E.**; **Walter, Alison E.**; **Arena, Ellen T.**; **Eliceiri, Kevin W.** (2017). "ImageJ2: ImageJ for the next generation of scientific image data". *BMC Bioinformatics*, v. 18, pp. 529-555.
https://doi.org/10.1186/s12859-017-1934-z

Schneider, Caroline A.; **Rasband, Wayne S.**; **Eliceiri, Kevin W.** (2012). "NIH Image to ImageJ: 25 years of image analysis". *Nature methods*, v. 9, n. 7, pp. 671-675.
https://doi.org/10.1038/nmeth.2089

Shelton, Taylor; **Poorthuis, Ate**; **Zook, Matthew** (2015). "Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information". *Landscape and urban planning*, v. 142, pp. 198-211.
https://doi.org/10.1016/j.landurbplan.2015.02.020

Software Studies Initiative (2015). *Cultural analytics*.
http://lab.softwarestudies.com/p/overview-slides-and-video-articles-why.html

Sued, Gabriela (2018). "Métodos digitales para el estudio de la fotografía compartida. Una aproximación distante a tres ciudades iberoamericanas en Instagram". *Empiria. Revista de metodología de ciencias sociales*, n. 40, pp. 15-39.
https://doi.org/10.5944/empiria.40.2018.22009

Xia, Chaolun; **Schwartz, Raz**; **Xie, Ke**; **Krebs, Adam**; **Langdon, Andrew**; **Ting, Jeremy**; **Naaman, Mor** (2014). "CityBeat: real-time social media visualization of hyper-local city data". In: *Procs of the 23rd Intl conf on world wide web*. ACM, pp. 167-170.
http://dl.acm.org/citation.cfm?id=2577020

Del flujo al *stock*: el programador radiofónico ante la gestión del catálogo digital

From flow to stock: The radio programmer facing the management of the digital catalog

Montse Bonet; **Toni Sellas**

Cómo citar este artículo:

Bonet, Montse; **Sellas, Toni** (2019). "Del flujo al *stock*: el programador radiofónico ante la gestión del catálogo digital". *El profesional de la información*, v. 28, n. 1, e280109.

https://doi.org//10.3145/epi.2019.ene.09

Artículo recibido el 04-07-2018
Aceptación definitiva: 04-12-2018

Montse Bonet ✉
https://orcid.org/0000-0001-8165-4898

Universitat Autònoma de Barcelona
Facultat de Ciències de la Comunicació
Departament de Comunicació Audiovisual i Publicitat
Carrer de la Vinya, s/n. Edifici I. 08193
Bellaterra (Barcelona), España
montserrat.bonet@uab.cat

Toni Sellas
https://orcid.org/0000-0003-2653-0547

Universitat de Vic – Universitat Central de Catalunya
Facultat d'Empresa i Comunicació
Departament de Comunicació
Sagrada Família, 7. 08500 Vic (Barcelona), España
toni.sellas@uvic.cat

Resumen

Más allá del indudable resurgir del podcast, su presencia en la oferta de algunas emisoras españolas merece un análisis detallado. Este estudio presenta el estudio de dicha oferta, así como una aproximación a los retos e interrogantes que hay tras la producción de podcasts nativos, de consumo exclusivo vía web. Los últimos responsables de la programación radiofónica deben lidiar con la tradicional parrilla de programas en flujo y añadir ahora la gestión de un catálogo de podcasts. El análisis cuantitativo se combina con el cualitativo a través de entrevistas a algunos de los máximos responsables de decidir cómo deben convivir estas dos ofertas radiofónicas complementarias. Los resultados muestran que de momento el podcast podría quedar definido como un producto cultural más que como un producto de una industria cultural.

Palabras clave

Radio; Emisoras de radio; Podcasts; Catálogos; Archivos de radio; Flujos; Programación; Programadores; *Stock*.

Abstract

Beyond the undoubted revival of the podcast, its presence in the offer of some Spanish radio stations deserves a detailed analysis. This study presents the analysis of this offer, as well as an approach to the challenges and questions behind the production of native podcasts, exclusively for web consumption. People ultimately responsible for radio programming has to deal with the traditional program schedule, in flow, and now has to add the management of a catalog of podcasts. The quantitative analysis is combined with the qualitative analysis through interviews with some of the top managers responsible for deciding how these two complementary radio offers should coexist. Results show that so far podcasts could be defined as a cultural product rather than a product of a cultural industry.

Keywords

Radio; Radio stations; Podcasts; Catalogs; Radio archives; Flows; Program schedule; Programmers; Stock.

1. Introducción

La digitalización supuso un reto intelectual para la economía política de las industrias culturales, perspectiva teórica que tanto ha contribuido a entender el funcionamiento de estas industrias entre las que se incluye la radio. Las clásicas tipologías sobre productos culturales o industrias culturales de Flichy, Zallo, Miège, Lacroix y Tremblay (**Bonet**, 2017) tuvieron que adaptarse a un código binario que ya en los años 90 del siglo pasado empezó a dar sus primeros pasos en distintas etapas de la cadena de valor de la industria radiofónica.

Desde esta perspectiva tradicional, la radio quedaría definida como una industria cultural de difusión continua, representante de la *culture du flot* o lógica de producción de flujo, cuyo contenido es efímero, difunde prototipos, controla el flujo estructurándolo en una parrilla programática y funciona con un sistema indirecto de financiación, ya sea publicidad o subvenciones públicas. La idea de una "programación de flujo" continua, de 24 horas al día, cuyas representantes tipo son la radio y la televisión, se contrapone a la "programación de *stock*", también conocida como de edición discontinua, representada por la industria discográfica o la cinematográfica, en las que la gestión de un catálogo (de discos, de películas) juega un papel fundamental en la cadena de valor.

La digitalización empezó a desdibujar la línea que separaba estos dos grandes tipos de productos culturales, de manera que algunas características del proceso de producción y difusión fueron entremezclándose. Uno de los primeros ejemplos fue la televisión de pago (*pay-per-view*, *pay-per-channel* o *pay-per-program*) o, con menor éxito, la radio de pago (*pay-per-listen*), cuando se rompió el flujo y pasó a ofrecerse un programa televisivo fuera de parrilla y horario prefijado, pudiendo ser visionado de forma individual, y adquirido como se adquiere un libro o un DVD.

Digitalizar supone trasladar el foco y centrarlo en las matrices (vídeo, audio, texto, imagen fija)

> "que, una vez convertidas en ceros y unos, quedan uniformizadas y pueden ser distribuidas sin atender demasiado a qué las hacía diferentes antes como medios: radio, televisión, prensa" (**Bonet**, 2012, p. 195).

Este es uno de los aspectos que influye en el proceso de producción de la radio. Del estudio de los pocos modelos que encontramos de cadena de valor del medio radio se pueden apuntar algunos cambios, consolidados o en fase de desarrollo (**Sellas**; **Bonet**, 2016):

- Contenidos: gracias a la universalización tecnológica, todos podemos ser productores y difusores. Es decir, se añaden más fuentes e intermediarios a la cadena.
- Agregadores: los radiodifusores tradicionales no son el único actor de la cadena que empaqueta y/o agrega contenidos de audio, pero la distancia es cada vez menor con nuevas empresas que gestionan el audio como parte de su oferta.
- Consumidores: se multiplican o multiplican su consumo gracias a la aparición de nuevos dispositivos, en lo que **Kischinhevsky** (2017) llama "radio expandida".
- Distribuidor: la lucha entre el *broadcast* y *broadband* y la existencia de redes sociales han sembrado cierta inquietud entre los radiodifusores. Nadie duda de que es necesario tener presencia en las redes, pero la cuestión es hasta qué punto se pierde el control sobre un contenido que ayuda a la Red a enriquecerse pero que la empresa radiofónica no sabe o no puede monetizar o rentabilizar en beneficio propio.

2. Análisis de la oferta: flujo de antena y catálogo digital

El objetivo básico de este artículo es el análisis de la oferta de contenidos mediante podcasting y las estrategias programáticas que guían esta oferta. Aunque es de indudable interés, no son objetivo de este análisis las herramientas o los modelos de gestión de contenidos, que ya han sido tratados con profusión en otros artículos (por ejemplo, en **Bonet**; **Fernández-Quijada**, 2006; **Fernández-Sande**; **Rodríguez-Barba**; **Rodríguez-Pallares**, 2013; **Rodríguez Pallares**, 2016; **Sellas**, 2012b).

Para ello se sigue una metodología tanto cuantitativa como cualitativa, que incluye una ficha de análisis y la interpretación de los datos recogidos, y entrevistas semiestructuradas en profundidad a responsables de programación de diversas emisoras públicas y privadas, realizadas entre el 25 de mayo y el 25 de junio de 2018. Dichas entrevistas complementan las que se realizaron para una investigación anterior de la que este artículo es deudor (**Sellas**; **Bonet**, 2016). El cuestionario planteado a los entrevistados estaba dividido en grandes bloques y adaptado en alguna pregunta a las características propias de cada empresa o proyecto. Estos bloques fueron:

a) perfil del programador radiofónico hoy en día: evolución, necesidades, formación;

b) producción no lineal (cuando el producto que alimenta a la web no procede de la antena): rutinas productivas, personal implicado, principales decisiones;

c) audiencia y cultura de consumo online;

d) comercialización de la producción no lineal;

e) la importancia del archivo (la nostalgia vende).

En la observación de la oferta, distinguimos entre los podcasts radiofónicos, que derivan de la antena, y los podcasts nativos digitales, no vinculados a la emisión. Esta dualidad es fruto de la propia evolución del podcasting, situado en una

fase de madurez (**Berry**, 2016; **Bonini**, 2015), pero también refleja un cambio de perspectiva en la industria radiofónica. En un inicio los podcasts fueron básicamente una extensión de la parrilla (**Gallego-Pérez**, 2012; **Sellas**, 2012a), pero en los últimos años la radio ha empezado a producir contenidos concebidos y distribuidos directamente como podcasts.

> Deporte y ficción concentran buena parte de los podcasts exclusivos online de las emisoras

La tabla 1 expone la oferta de podcasts de las emisoras analizadas, con fecha 2 de julio de 2018. Las emisoras son las 6 primeras del *ranking* del *EGM* (*Estudio General de Medios*) en la categoría de radio generalista. Este artículo se centra en las emisoras generalistas, a pesar de contener comentarios y respuestas sobre la programación musical, dado que de momento son las radios que están ofertando mayor variedad, no sólo de contenidos subidos a la Red desde la antena sino también de productos elaborados exclusivamente para ser consumidos online.

Para realizar esta observación, hemos recogido los datos cuantitativos y cualitativos de la web de la emisora, consultando la pestaña identificativa ya sea de los podcasts o de la sindicación RSS. Como se puede observar en la columna "Identificación y ubicación", no todas las emisoras utilizan la palabra podcast para denominar dicho espacio ni el url de la página web donde se encuentran los podcasts la incluye siempre.

La cuantificación del número ofertado toma como unidad el podcast de cada programa, sección o parte de programa. Es decir, se considera como podcast el espacio en cuestión, no cada una de sus ediciones. Las 6 emisoras analizadas suman 684 podcasts y la media es de 114. Las dos emisoras catalanas, la privada *RAC1* y la pública *Catalunya Ràdio*, son las que más ofrecen: 234 y 151. Les sigue la *Cadena SER* con 117. Es menor la oferta en *Onda Cero Radio* (*OCR*) (93), *RNE* (67) y especialmente en *COPE* (22).

La descripción que sigue (tabla 1) expone el tipo de podcast ofrecido. Es una tipología de elaboración propia, a partir del análisis y comparación del elemento de la programación que se ofrece como tal. La mayoría ofrece el podcast del programa completo o por tramos. *RAC1* y *Catalunya Ràdio* destacan por el nivel de detalle, dado que permiten el acceso al podcast de todas o casi todas las secciones de los programas. El resto también ofrece podcasts de secciones, pero menos.

Tabla 1. Podcasts de las emisoras

Emisor (grupo)	Podcasts	Descripción	Estructura	Contenidos	Identificación y ubicación	Distribución
SER (*Prisa*)	117	Programas: por tramos, exclusivos online, secciones, firmas/comentarios, emisoras, archivo	Por programas Temática Geográfica	Actualidad, Bienestar, Consumo, Cultura, Deporte, De nuestras emisoras, Entretenimiento, Firmas, Gastronomía, Historia, Humor, Misterio, Música, Noche, Otros, Periodismo, Televisión, Tendencias, Toros	*Cadena SER Podcasts* ***http://cadenaser.com/ser/podcasts***	RSS *iTunes* *iVoox*
COPE (*COPE*)	22	Programas: completos, por tramos, audios destacados, archivo	Por programas	Podcasts de *COPE* Podcasts de otros programas	RSS de *COPE* ***http://www.cope.es/pagina/rss***	RSS *iTunes* *iVoox* *Feedly*
Onda Cero Radio (*Atresmedia*)	93	Programas: completos, exclusivos online, secciones, firmas/comentarios, emisoras locales, emisoras regionales	Por programas Temática Geográfica	De lunes a viernes, Fin de semana, Ciencia, Cultura, Deportes, Economía, Humor, Historia, Informativos, Local, Música, Opinión, Regional, Tecnología	*Podcasts de Onda Cero* ***http://www.ondacero.es/podcast***	RSS *iTunes*
RNE (*RTVE*)	67	Programas: completos, por tramos, secciones, audios destacados	Temática. Géneros	Ciencia y Tecnología, Cultura, Documentales, Deporte, Entretenimiento, Informativos, Informativos territoriales, Magazín, Religiosos, Otros programas	*RNE a la carta* ***http://www.rtve.es/radio/podcast*** ***http://www.rtve.es/alacarta/programas/radio-nacional/todos/1***	RSS *iTunes*
RAC1 (*Grupo Godó*)	234	Programas: completos, por tramos, secciones	Por programas	Programas y secciones (no se establecen categorías)	RSS Àudios ***https://www.rac1.cat/rss***	RSS *iTunes*
Catalunya Ràdio (*CCMA*)	151	Programas: completos, exclusivos online, secciones	Por programas	Programas y secciones (orden alfabético)	*Podcasts* ***http://www.ccma.cat/catradio/podcast*** ***http://www.ccma.cat/catradio/podcast/catalunya-radio***	RSS *iTunes* *iVoox*

Fuente: elaborado a partir del análisis de las webs de las emisoras de la muestra.

> *+CatRàdio*, *Playzcast* y especialmente *Podium Podcast* responden a estrategias diversas de construcción de un catálogo nativo digital

Cadena SER y *Onda Cero* incluyen los de sus emisoras regionales o locales (el feed RSS permite suscribirse a lo más destacado de la emisora regional/local). *SER* y *COPE* ofrecen el feed de algún programa de archivo, mientras que las emisoras con un catálogo digital (programas emitidos sólo vía web, no en antena) incluyen también el podcast de esos espacios. Por último, *SER*, *COPE*, *Onda Cero* y *RNE* utilizan el podcasting para ofrecer a sus oyentes "audios destacados", como los comentarios de opinión de algunos de sus profesionales o colaboradores, declaraciones de personajes de actualidad o incluso determinados fragmentos de boletines horarios o espacios informativos.

En cuanto a la estructura de la oferta, en las públicas *Catalunya Ràdio* y *RNE* existe un estadio previo que es la oferta global de podcasts de todas las emisoras del grupo, que permite seleccionar cualquiera de ellas. Las webs analizadas presentan una estructura ordenada por programas. En algunos casos, esta organización se fusiona también con otros dos criterios: el temático y el geográfico (regional/local). A la hora de mostrar los podcasts por programas, observamos varios modos de ordenación:

- alfabética, utilizada por *Catalunya Ràdio*;
- modelo que prioriza los principales programas de la parrilla de programación, ya sea mediante un listado (*RAC1*, *COPE*) o con las imágenes de los profesionales que encabezan el programa (es decir, las "estrellas"; este es el caso de *SER* y *Onda Cero*);
- un modo singular es el de *RNE*, que compagina los podcasts por programas (reconocidos con el nombre del espacio) con los identificados por tipo (informativos, informativos territoriales, documentales, magazín);
- una estructura habitual es la que combina una muestra destacada de los podcasts de los principales programas de la emisora (incluso gráficamente) y una lista temática del resto o de su totalidad.

Sobre el tipo de contenidos, *SER*, *Onda Cero* y *RNE* establecen los temas. Por el contrario, *COPE* distingue entre los programas más destacados y el resto ("Podcasts de otros programas"), mientras que *Catalunya Ràdio* y *RAC1* presentan una lista (alfabética y jerárquica respectivamente). Dado que la tipología deriva de los podcasts ofertados, la clasificación que encontramos en las emisoras es la propia de una programación de modelo mosaico, protagonizada por los magazines de mañana y tarde y los programas deportivos: actualidad, deportes, entretenimiento, opinión. Junto a ello, vemos otras muchas categorías que resultan de secciones de temática específica o de programas especializados (cultura, humor, historia o economía, entre otras). Finalmente, también caben categorías por géneros programáticos (informativos, documentales, magazín, ficción) y otras que muestran la dimensión territorial (local, regional).

Por último, observamos la distribución. Más allá de la escucha o descarga en la propia web y del feed RSS del podcast, la mayoría utiliza *iTunes* e *iVoox*. En tal caso se incluye el icono correspondiente como elemento visual identificativo y su enlace. La excepción más destacada es *RAC1*. Esta emisora del *Grupo Godó* mantiene un acuerdo con *iTunes* por el que algunos de sus podcasts están presentes en la plataforma de *Apple*, pero no en *iVoox*. Sin embargo, el listado de podcasts de la emisora sólo incluye el feed RSS y no utiliza ningún símbolo que identifique a *iTunes*. Para suscribirse el usuario debe realizar una búsqueda en la plataforma.

La tabla 2 expone el catálogo digital de las emisoras generalistas analizadas, es decir, la oferta de contenidos exclusivos online incluyendo el análisis de *Podium Podcast* como plataforma nativa de podcasting del grupo *Prisa*. De este modo podemos observar la oferta no sólo en la web de las emisoras sino también en nuevos espacios creados para esta finalidad.

Para llevar a cabo el análisis hemos identificado los contenidos que los emisores especifican como programas o podcasts diferenciados de la programación convencional y de difusión exclusiva por internet. En el período de observación, suman 120 programas y podcasts, de los que 55 corresponden a *Podium Podcast* (cuantificados como unidad por programa, sin considerar cada uno de los episodios que lo integran). La plataforma de podcasts nativos digitales de *Prisa* prácticamente iguala en número a la suma de los contenidos exclusivos online del resto de emisores (incluida la *Cadena SER*, del mismo grupo). La oferta de los demás oscila entre los 8 de *RNE* y los 22 de *COPE*.

Si observamos la identificación y ubicación de los contenidos online, podemos distinguir tres modos de oferta:

- Las emisoras que los sitúan en su web y con poca distinción: *COPE* y *Onda Cero* coinciden en la expresión "sólo en…" seguida del nombre web de la emisora, mientras que la *Cadena SER* no utiliza una denominación específica. Estas emisoras mantienen los contenidos exclusivos online en la misma página que el conjunto de programas.
- Las emisoras que utilizan una nueva marca o submarca para identificar los contenidos nativos digitales, pero los ubican en espacios que dependen estructuralmente del sitio web principal. *Catalunya Ràdio* agrupa su catálogo bajo la marca *+CatRàdio* y en una página específica, situada dentro del sitio web de la emisora. Por su parte, *COPE* distingue el proyecto *Sin mi identidad* mediante un subdominio propio, a diferencia del resto de sus podcasts online.
  En *RNE* los podcasts nativos se enmarcan en una estrategia de desarrollo de nuevos contenidos basada en la ficción binaural y las narrativas transmedia. Están ubicados en *Playz*, espacio con un dominio propio pero ubicado dentro del sitio web de *RTVE*. También pueden encontrarse en la página de "Ficción sonora" de *RNE*, donde la emisora pública agrupa los contenidos de ficción que ha producido y emitido en los últimos años, ya sea en antena o en internet.

Tabla 2. Catálogo digital

Emisor (grupo)	Catálogo digital	Descripción	Estructura	Contenidos	Identificación y ubicación	Distribución
SER (*Prisa*)	14	Programas temáticos	Catálogo	Bienestar, Música, Deporte	*Cadena SER Programas* *http://cadenaser.com/ser/programas*	RSS *iTunes* *iVoox*
COPE (*COPE*)	22	Programas temáticos, Ficción sonora, Proyectos transmedia	Listado Catálogo	Cine, Deporte, Documental, Educación, Ficción, Gastronomía, Música, Tradiciones, Social	*Sólo en cope.es* *http://www.cope.es* *Sin mi identidad. Un podcast de COPE* *http://sinmiidentidad.cope.es*	RSS *iTunes* *iVoox*
Onda Cero Radio (*Atresmedia*)	9	Programas temáticos, Ficción sonora	Catálogo	Actualidad internacional, Cultura, Deporte, Ficción, Música, Social	*Sólo en ondacero.es* *http://www.ondacero.es/solo-ondaceroes*	RSS *iTunes* *iVoox*
RNE (*RTVE*)	8	Ficción sonora, Proyectos transmedia	Catálogo	Ficción	*Playz (Playzcast)* *http://www.rtve.es/playz*	RSS *iTunes*
RAC1 (*Grupo Godó*)	-	-	-	-	-	-
Catalunya Ràdio (*CCMA*)	12	Programas temáticos, Serie de no ficción	Catálogo	Arquitectura, Bienestar, Cultura, Deporte, Economía, Educación, Familia, Gastronomía, Meteorología, Música, Periodismo de investigación, Sexualidad, Tecnología	*+CatRàdio* *http://www.ccma.cat/catradio/mes-catradio*	RSS *iTunes* *iVoox*
Podium Podcast (*Prisa*)	55	Podcasts por temporadas y episodios	Catálogo por temáticas	Ficción, Esenciales, Entretenimiento, Periodismo	*Podium Podcast* *http://www.podiumpodcast.com*	RSS *iTunes* *iVoox*

Fuente: elaborado a partir del análisis de las webs de las emisoras de la muestra.

- La estrategia más singular es la de *Prisa* con *Podium Podcast*, una nueva marca y plataforma, con su propio espacio autónomo en internet.

En cuanto a la distribución, más allá de su web, vemos que todos los emisores utilizan *iTunes* e *iVoox* como plataformas complementarias para difundir los contenidos.

En el análisis observamos también el tipo de oferta, la estructura y los contenidos. Todos los emisores presentan su propuesta exclusiva online por programas, como una oferta temática y especializada, mediante una muestra de tipo catálogo. No obstante encontramos algunas singularidades: En *RNE* los podcasts nativos se exponen como ficciones sonoras insertadas en proyectos transmedia o como ficciones binaurales. En *+CatRàdio* se utiliza la descripción temática, aunque en el caso de *Tor, tretze cases i tres morts* se hace referencia al género (serie de no ficción). La plataforma *Podium Podcast* estructura su más de medio centenar de productos en cuatro bloques temáticos: entretenimiento, esenciales (archivo), ficción y periodismo. En este caso la estructura incluye las temporadas de cada podcast.

Más allá de la estructura de la oferta y en cuanto al tipo de contenidos, en las tres principales radios generalistas privadas el deporte tiene una presencia destacada: 7 de 14 podcasts en la *SER*, 11 de 22 en *COPE*, y 4 de 9 en *Onda Cero*. Se trata de programas temáticos, dedicados a un deporte o competición específico, mediante los que las emisoras pueden atender a audiencias muy de nicho y aprovechar el conocimiento especializado de sus profesionales. En estas emisoras, la música también tiene bastante presencia en la oferta exclusiva online (programas temáticos, sobre géneros musicales o sobre la música de un determinado período). Y junto a ello encontramos podcasts dedicados a otros temas especializados y para audiencias segmentadas, como cine, gastronomía, educación, actualidad internacional o bienestar, entre otros. Estas categorías forman parte también del catálogo de *+CatRàdio*, en el que se incluyen otros temas específicos (arquitectura, meteorología, sexualidad) o vinculados a la cultura popular (sardanas y música de cobla). La radio pública catalana destaca también una serie de no ficción basada en el periodismo de investigación.

> Podríamos definir el podcast como un producto cultural más que como un producto de industria cultural

La ficción sonora es otra apuesta de estas emisoras. En *RNE* con proyectos transmedia y ficciones binaurales o bien algunas producciones de ficción esporádicas que *Onda Cero* incluye en su catálogo digital y que suponen una vía de experimentación de nuevos contenidos. El caso más reciente (julio de 2018) es *Sin mi identidad*, de *COPE*: una serie de ficción de ocho capítulos en el marco de un proyecto transmedia que incluye otros dos podcasts documentales, vídeos, blog y banda sonora propia.

La ficción es uno de los ejes de *Podium Podcast*. En sólo dos años, la plataforma ha llevado a cabo una estrategia de producciones de ficción que incluye:

- series originales (*El gran apagón*, *Bienvenido a la vida peligrosa*);
- ficciones concebidas como *branded content* para marcas (*Aerolínea Momentos*, coproducida con *Iberia*);
- cronoficciones y recreaciones de casos reales (en *Negra y criminal*).

Además, en la categoría de "Esenciales" ha puesto en valor el archivo histórico de la *Cadena SER* mediante la recuperación de algunos programas históricos de ficción, como *Historias de medianoche*, el radioteatro de los años cincuenta y sesenta en *Teatros del aire* o la radionovela de los setenta *Tentación a medianoche*.

3. Estrategias programáticas, programación del flujo y gestión del *stock*

La programación es el producto que la radio ofrece a sus oyentes, los cuales a su vez son lo que la empresa ofrece a sus anunciantes. A lo largo de su historia, como recuerda **Martí** (2000), la radio ha ido perfeccionando su programación, ajustándola más a las características propias del medio, llenando más horas al día, haciendo que los programas duraran más o menos hasta llegar a configurar unos modelos programáticos ya perfectamente delimitados por una tipología. El mismo autor nos recuerda que la programación es un arte que debe garantizar tres cosas:

- continuidad (el flujo no puede parar);
- diversidad o especialización;
- homogeneidad,

pues al fin y al cabo la parrilla de programación de una emisora es un conjunto de programas, ordenados por franjas horarias y es al mismo tiempo su sello, su marca.

> "En la radio no lineal, el catálogo es *stock*, nicho, experimentación y heterogeneidad"

Durante años el programador radiofónico ha ido tomando decisiones sobre un producto que quedaba definido y reconocido por su continuidad, por su fluir. Sin embargo, las estrategias aplicadas y las decisiones tomadas empezaron a cambiar el día en que el flujo programático pudo convertirse en producto cultural de *stock* y empezar a configurar un catálogo. De momento fue posible solamente con el "despiece" de la parrilla programática ya emitida en antena y, por consiguiente, las decisiones debían girar alrededor de qué programas se subían a la Red, si se subían enteros o troceados, etc. (**Sellas**, 2012a).

A decir de los entrevistados para este estudio, de momento la figura del programador no ha variado tanto, lo que sí cambia es:

- el flujo de información que le llega (Celso Del Río, *OCR*);
- el nombre del cargo, la perspectiva o mentalidad desde la que se programa (Vicent Argudo, *Prisa Radio*);
- cambia el nombre, que pasa de jefe de programas a director de contenidos (Santi Faro, *Catalunya Ràdio*).

A este respecto añade Argudo que hay una parte del perfil de programador que no se transforma porque tampoco lo hace la radio en antena, pero lo que sí debe cambiar es la parte de la radio no lineal, sobre la que llegan muchos más datos y de manera más rápida, y no es fácil de gestionar. Por ello esta parte suele llevarla más el departamento de marketing, porque hay que buscar oyentes, nichos a los que ofrecer algo que pueda interesarles. En internet la radio ya no emite, no se lanza simplemente un mensaje, en el entorno digital

> "la radio tiene que escuchar más que hablar y a eso no estamos acostumbrados" (Vicent Argudo).

Faro apostilla que el programador debe ser

> "un profesional atento a los cambios del entorno de la comunicación",

algo que concuerda con esta idea de una toma de decisiones rápida, dependiente de un flujo constante de miles de datos y búsqueda de públicos.

La distinción entre radio lineal y no lineal no es meramente semántica ya que, como bien explican los entrevistados, la radio en antena forma parte de un sistema cuyas reglas de funcionamiento fueron aceptadas por diversas partes y las programaciones en flujo dependen de cómo les mide el *EGM*.

En el "dial digital" (Santi Faro, *Catalunya Ràdio*), por el contrario, no existe todavía un sistema de medición reconocido y que inspire suficiente confianza a los anunciantes (Mónica Muñoz, *OCR*). Y este es un detalle fundamental para dar empuje a los intentos de monetizar los contenidos digitales. A excepción de *Podium Podcast* (M. Jesús Espinosa de los Monteros) que apunta cierto optimismo, el resto de entrevistados reconoce que todavía es más un generador de gastos que de beneficios, motivo por el cual en *Catalunya Ràdio* se habla de la "producción corresponsable", para referirse a los proyectos que llegan de fuera de la emisora y que deben presentarse junto con un patrocinador.

También coinciden en que, online o no, lo que se busca es que haya un sello, una marca fácilmente identificable con la de la radio ya conocida por el público. Por ello no tendría mucho sentido que se tomaran las decisiones sobre los nuevos productos online por separado y no hubiera "continuidad entre antena e internet" (Francisco J. Pomares, *OCR*). Juega un

papel importante la experiencia acumulada. Ninguna de las emisoras de la muestra es nueva en el entorno digital, pero reconocen, usando para ello la misma expresión, que hasta hace poco la web era un poco un "cajón de sastre" y ahora se cuidan más los criterios. Insiste en ello Mónica Muñoz:

> "que quien escuche programas exclusivos de la web sepa igualmente que eso es el sello de *Onda Cero*".

En este nuevo contexto, puntualiza Argudo (*Prisa Radio*), hay que trabajar las marcas desde otra perspectiva, dado que "en la Red no vas a vender marcas, vas a vender audiencias" y por ello hacen falta muchos productos distintos que satisfagan a *targets* variados. De hecho, *Podium Podcast* es un proyecto reciente que dispone de la fuerza de un grupo, pero para el que se eligió un nombre distinto. M. Jesús Espinosa apunta la necesidad de distinguirlo precisamente como un nuevo proyecto del grupo *Prisa* que tiene planeado dar el salto a Latinoamérica, es decir, ofrecer un catálogo de podcasts (ella es una de las pocas personas entrevistadas que habla directamente de "catálogo") que incluya producciones realizadas desde las emisoras latinoamericanas, y eso solamente puede hacerse si no lleva la marca *SER*.

> "Quien programa no sólo debe atender la lógica de la programación en flujo sino también saber gestionar un catálogo de contenidos en audio"

Existe consenso al hablar de la vertiente digital como un espacio de experimentación, mayor libertad, ensayo (de voces y de contenidos), pero también un lugar donde dar una segunda oportunidad a programas que en su versión original no tuvieron éxito, no acabaron de cuajar o se emitían en horarios residuales; o bien lo contrario, un lugar donde probar contenidos que después dan el salto a la antena. Por ello en *Catalunya Ràdio*, Santi Faro insiste en la necesidad de pensar productos que no estén "radiofónicamente estructurados", sin duración estándar, cambiando el tono en función de una franja horaria, incluso "sin cola musical para cuadrarlo con los boletines horarios". Los contenidos nativos digitales no forman parte de ningún flujo que los condicione.

Finalmente el archivo de las emisoras juega un papel clave alimentando el entorno digital: *Onda Cero Radio* está en proceso de digitalizarlo por entero, *Podium Podcast* y *Catalunya Ràdio* se alimentan de él para proyectos actuales y futuros. La historia de España, como recuerda Francisco J. Pomares (*OCR*), está muy ligada a la radio y el archivo es una pieza fundamental. La última apuesta programática ha sido crear podcasts con programas que giran alrededor del uso y aprovechamiento del archivo sonoro de la emisora. La nostalgia vende y, sin ser nada nuevo, ahora se recurre al archivo no solamente para elaborar un programa especial de celebración, aniversario o similar, sino también para ser el sustento del programa, con una serialización que permite explotar mejor sus posibilidades de fidelización de audiencias y potencial de monetización.

4. Conclusiones

La radio en internet, especialmente pero no sólo en su vertiente de formato podcast, está suponiendo un reto para las personas responsables de la programación. La variedad de plataformas entre las que distribuir contenidos de audio se ha ampliado considerablemente. Si hace unas décadas se podía elegir entre FM y OM, después se añadió la decisión de qué contenidos replicar en la TDT (Televisión Digital Terrestre), hasta llegar a la actualidad.

La "radio extendida" (**Kischinhevsky**, 2017), de "dial digital" (según expresión usada en *Catalunya Ràdio*) no sustituye, no mata a la radio lineal, pero exige un esfuerzo complementario. En la radio lineal la programación queda definida por la continuidad, la diversidad o la especialización y la homogeneidad (en términos de estrategias y programación, más que de contenidos). En la radio no lineal, el catálogo es *stock*, nicho, experimentación y heterogeneidad. De momento parecen bastante incompatibles y, sin embargo, conviven y esta convivencia obliga a los programadores radiofónicos a encontrar la manera de o bien adaptar la cadena de valor tradicional, o bien crear otra en paralelo, para un producto de audio realizado, ofrecido y consumido de una forma que hasta el momento no le era habitual.

Curiosamente el catálogo digital dispone de datos más inmediatos y exactos (aunque todavía insuficientes) sobre los oyentes, que permiten la toma de decisiones con una rapidez y capacidad de reacción que no tiene la radio de flujo. Ello se debe en buena parte al sistema industrial tejido durante años en el que los emisores y anunciantes convinieron aceptar un sistema de medición determinado (*EGM*). La radio no lineal carece todavía de un sistema pactado y aceptado que inspire suficiente confianza. Por ello actualmente podríamos definir el podcast como un producto cultural más que como un producto de industria cultural. De hecho podríamos incluso apuntar que *Podium Podcast* es un primer intento de convertir el podcast en un producto de industria cultural, dado su interés no sólo en explorarlo sino también en comercializarlo y posicionarlo en el mercado, fomentando su consumo.

Transformación, cambio, mutación, son las palabras más utilizadas al intentar definir un perfil profesional que no siempre recae en una sola persona. Por ello no existe todavía una estrategia clara de programación, más bien experimentación en la gestión de un catálogo que se ofrece a los oyentes, nuevos o no, al tiempo que se sigue alimentando el flujo de unas emisoras que ya llevan muchos años de experiencia a sus espaldas. Fruto de esa experimentación, la oferta online (nativa o no) es heterogénea y todavía le falta encontrar su propio orden y organización, al estilo de los programas distribuidos en una parrilla.

Quizá uno de los detalles más destacables es constatar que finalmente se extiende entre los trabajadores de la radio la conciencia de que las emisoras son gestoras de contenidos de audio, no simplemente productoras de contenido radio-

fónico (aunque todavía cueste cambiar algunas mentalidades). Ello les abre un abanico de oportunidades para expandir su oficio, que no es otro que ofrecer buenos productos elaborados con la matriz audio.

5. Entrevistas

Vicent Argudo (*Prisa Radio*). Director de Cadenas Musicales.

Celso Del Rio (*Onda Cero Radio*). Subdirector de Programación.

M. Jesús Espinosa de los Monteros. Jefa de Proyecto de *Podium Podcast*.

Santi Faro (*Catalunya Ràdio*). Director de Contenidos.

Mónica Muñoz (*Onda Cero Radio*). Editora de contenido de *ondacero.es*.

Francisco J. Pomares (*Atresmedia Radio*). Jefe de prensa.

Así como las entrevistas realizadas previamente en el proyecto de investigación de **Sellas** y **Bonet** (2016).

6. Referencias

Berry, Richard (2016). "Podcasting: Considering the evolution of the medium and its association with the word 'radio'". *Radio journal: International studies in broadcasting & audio media*, v. 14, n. 1, pp. 7-22.
https://doi.org/10.1386/rjao.14.1.7_1

Bonet, Montse (2012). "Distribución: analógica o digital, lo importante es llegar al oyente". En: Gallego-Pérez, Juan-Ignacio; García-Leiva María-Trinidad (coords.). *Sintonizando el futuro: radio y producción sonora en el siglo XXI.* Madrid: Instituto RTVE, pp. 189-207. ISBN: 978 84 88788856

Bonet, Montse (2017). "The Spanish contribution to the study of cultural industries. The first steps". En: Tosoni, Simone; Carpentier, Nico; Murru, Maria-Francesca; Kilborn, Richard; Kramp, Leif; Kunelius, Risto; McNicholas, Anthony; Olsson, Tobias; Pruulmann-Vengerfeldt, Pille (eds.). *Present scenarios of media production and engagement*. Bremen: Edition Lumière, pp. 85-94. ISBN: 978 3 943245721
http://www.researchingcommunication.eu/SuSobook2016.pdf

Bonet, Montse; **Fernández-Quijada, David** (2006). "El reto de la digitalización del archivo sonoro en los servicios públicos de radiodifusión. El caso de *Catalunya Ràdio*". *El profesional de la información*, v. 15, n. 5, pp. 390-396.
http://elprofesionaldelainformacion.com/contenidos/2006/septiembre/8.pdf

Bonini, Tiziano (2015). "La 'segona era' del podcàsting: el podcàsting com a nou mitjà de comunicación de masses digital". *Quaderns del CAC*, v. 41, n. 18, pp. 23-33.
https://dialnet.unirioja.es/servlet/articulo?codigo=5168807

Fernández-Sande, Manuel; **Rodríguez-Barba, Dolores**; **Rodríguez-Pallares, Miriam** (2013). "La gestión de contenidos como actividad estratégica en empresas de radiodifusión. Estudio de casos en la radio comercial española". *El profesional de la información*, v. 22, n. 5, pp. 392-398.
https://doi.org/10.3145/epi.2013.sep.03

Gallego-Pérez, Juan-Ignacio (2012). "Podcasting in Spain: A new business model or a reflection of traditional radio?". *Radio journal: International studies in broadcast & audio media*, v. 10, n. 1, pp. 23-34.
https://doi.org/10.1386/rjao.10.1.23_1

Kischinhevsky, Marcelo (2017). *Radio y medios sociales. Mediaciones e interacciones radiofónicas digitales*. Barcelona: Editorial UOC. ISBN: 978 84 9180 011 8

Martí, Josep M. (2000). *De la idea a l'antena. Tècniques de programació radiofónica*. Barcelona: Pòrtic. Eines de Periodistes, n. 5. ISBN: 978 84 73066105

Rodríguez-Pallares, Miriam (2016). "Propuesta conceptual de un modelo de gestión de contenidos y del conocimiento en la empresa radiofónica". *Revista española de documentación científica*, v. 39, n. 2, pp. 1-17.
https://doi.org/10.3989/redc.2016.2.1271

Sellas, Toni (2012a). "A two-dimensional approach to the study of podcasting in Spanish talk radio stations". *Radio journal: International studies in broadcast & audio media*, v. 10, n. 1, pp. 7-22.
https://doi.org/10.1386/rjao.10.1.7_1

Sellas, Toni (2012b). "Repositorios sonoros y prescripción de contenidos. El caso iVoox". *El profesional de la información*, v. 21, n. 2, pp. 206-209.
https://doi.org/10.3145/epi.2012.mar.13

Sellas, Toni; **Bonet, Montse** (2016). *L'àudio a internet: de la ràdio en línia als projectes nadius*. Barcelona: Consell de l'Audiovisual de Catalunya.
https://www.cac.cat/sites/default/files/migrate/study/IX_Ajuts_Audio_Internet_Sellas_Bonet.pdf

Producción de ficción televisiva española a partir de la desregulación: entre la atomización de las empresas y la concentración vertical

Spanish television fiction since desregulation: Between atomization and vertical concentration of production

Charo Lacalle; Mariluz Sánchez-Ares

Cómo citar este artículo:

Lacalle, Charo; **Sánchez-Ares, Mariluz** (2019). "Producción de ficción televisiva española a partir de la desregulación: entre la atomización de las empresas y la concentración vertical". *El profesional de la información*, v. 28, n. 1, e280110.

https://doi.org/10.3145/epi.2019.ene.10

Artículo recibido el 19-07-2018
Aceptación definitiva: 12-11-2018

Charo Lacalle ✉
https://orcid.org/0000-0002-0024-6591

Universitat Autònoma de Barcelona
Facultat de Ciències de la Comunicació
08193 Bellaterra, Barcelona, España
rosario.lacalle@uab.es

Mariluz Sánchez-Ares
https://orcid.org/0000-0002-0076-7522

Jungle Media
33 Jefferson Ave., Toronto
ON M6K 1Y4, Canadá
mariluz.sanchez.ares@gmail.com

Resumen

Este artículo intenta paliar la ausencia de datos y de investigaciones empíricas sobre las diferentes modalidades de producción de ficción televisiva española, desde el predominio inicial de la producción interna hasta la casi completa externalización actual. La muestra incluye las 490 producciones de ficción propia, estrenadas en las cadenas estatales entre 1990 y 2017, en las que participaron 234 empresas. La investigación revela que más del 60% de las productoras independientes únicamente han participado en una sola ficción, mientras que 10 empresas (los grupos *RTVE*, *Atresmedia* y *Mediaset*, y siete productoras independientes) son responsables de más de la mitad de los estrenos. El análisis pone de manifiesto el carácter paradójico de un sector cuya creciente especialización y concentración vertical convive con la atomización de las productoras independientes.

Palabras clave

Producción audiovisual; Televisión; Ficción; Series de televisión; Industrias culturales; España.

Financiación

Este artículo ha sido elaborado en el marco del proyecto *Historia de la programación y programas de ficción televisiva en España (cadenas de ámbito estatal): de la desregulación al apagón analógico, 1990-2010* (CSO2015-66260-C4-4-P), financiado por el *Ministerio de Economía y Competitividad.* Los datos de la producción de ficción española entre 2008 y 2017 proceden de los informes elaborados por el equipo español de *Obitel* (*Observatorio Iberoamericano de Ficción Televisiva*), que dirige Charo Lacalle.

Abstract

This article attempts to mitigate the absence of data and empirical research on the different types of production of Spanish television fiction, from the initial predominance of in-house production to the near complete industry wide trend of outsourcing. The sample includes 490 productions aired on national networks between 1990 and 2017 across 234 production houses. The research reveals that more than 60% of independent producers have partaken in only one production, while 10 studios (including the three major national networks) are responsible for more than half of the productions. The analysis highlights the paradoxical nature of a sector whose growing specialization and vertical concentration coexists with the atomization of independent producers.

Keywords

Audiovisual production; Television; Fiction; TV series; Cultural industries; Spain.

1. Introducción

Las industrias culturales y creativas (ICC) constituyen uno de los puntales de la economía española, con el 6% del PIB y el 5% del empleo (*Deloitte*, 2017). Las empresas audiovisuales, el grupo económico más relevante de las ICC, facturaron en 2016 el 68,0% del total, el 38,2% del cual correspondía a la televisión y la radio, y el 29,8% restante al cine y el vídeo (*Ontsi*, 2017). A pesar de su indudable relevancia económica y social, la ausencia de estadísticas oficiales sobre géneros y categorías de la programación impiden a los investigadores determinar el incremento de la producción y diferenciar de manera precisa la producción independiente de la producción interna de las cadenas, una situación que Bustamante y Zallo ya evidenciaban en 1988 (**Bustamante**; **Zallo**, 1988).

Este artículo analiza la producción de ficción televisiva española emitida por las cadenas estatales entre 1990 y 2017, con el objetivo de determinar tanto la evolución del sector como su configuración actual y colmar la escasez y la dispersión de las informaciones sobre el tema. La muestra de análisis está integrada por las 234 empresas que participaron en régimen de producción interna, externa (asociada o financiada) o coproducción con otras cadenas[1], en los 490 programas de ficción estrenados en el período de referencia (series, seriales, miniseries, *TVmovies* y *sketches*). El resumen de los resultados obtenidos está precedido por una panorámica sobre la evolución de la producción de ficción televisiva en España, destinada a contextualizar el análisis.

La investigación asume como punto de partida que la "marketización" originada por la desregulación a partir de 1990

> "ayudó a crear un contexto en el que las industrias culturales eran percibidas como un buen sector para las inversiones" (**Hesmondhalgh**, 2002, p. 257).

Esto habría propiciado entre la segunda mitad de los noventa y los primeros años de 2000 un período de expansión de la producción audiovisual particularmente prolífico a nivel mundial (*United Nations*, 2008). Sin embargo a diferencia de otros países europeos, el despegue tardío de la producción independiente en España por efecto del largo monopolio televisivo determinó en un primer momento la implantación de un sistema de producción interna, que fue substituido en un segundo período por otra modalidad basada principalmente en la producción externa asociada y, sucesivamente, en la producción externa financiada.

2. Evolución de la producción de ficción televisiva en España

La historia de la producción de ficción televisiva española se puede dividir a grandes rasgos en cinco períodos, desde sus comienzos a finales de los años cincuenta hasta la actualidad, asociados de manera indisoluble a la evolución de la producción audiovisual.

2.1. Primeros pasos: políticas públicas de impulso a la producción

TVE inauguraba sus emisiones de ficción propia en 1958, con un espacio dirigido por Jaime de Armiñan (Érase una vez) que acogía escenificaciones de cuentos populares. Iniciaba así la primera etapa de la historia de la ficción televisiva española, centrada en la grabación de teleteatro, series cómicas y adaptaciones por capítulos de clásicos literarios, que concluiría con un acercamiento al cine a comienzos de los setenta, cuyo corolario sería el incremento de la producción de series dramáticas y miniseries.

Con el inicio de la Transición en 1975 arrancaba una segunda etapa, en la que la comedia costumbrista tan popular en los años sesenta y setenta iría cediendo terreno a las diferentes variantes del drama (sentimental, de acción, de aventuras, etc.), así como a las adaptaciones de grandes obras literarias (**Palacio**, 2001; **García-de-Castro**, 2002; **Lacalle**, 2018). En 1979 el primer gobierno de *UCD* concedió a *TVE* un crédito extraordinario de 1.300 millones de pesetas (equivalentes a 37 millones de euros de 2018)[2] con el objetivo de promover la producción de series y películas para televisión mediante concurso público. En 1985 el primer gobierno de Felipe González creó el *Instituto de la Cinematografía y las Artes Audiovisuales* (*ICAA*), destinado a incrementar la producción y a mejorar el grado de desarrollo de las empresas. El fomento de la producción constituía también uno de los principales objetivos del fondo *Euroimages*, instituido por el *Consejo de*

Europa en 1988; un período en el que las industrias culturales eran percibidas por las políticas públicas europeas como un sector de gran interés (**Hesmondhalgh**; **Pratt**, 2005). En 1991 la UE lanzaba el programa *Media* para impulsar la transformación audiovisual y estimular la colaboración entre los países miembros (convertido en 2014 en subprograma de *Europa Creativa* para el fomento de la industria audiovisual europea). La *Ley 25/1994, de 12 de julio*, incorporaba al ordenamiento jurídico español la denominada *Directiva sobre Televisión sin Fronteras* (*Directiva 89/552/CEE, de 3 de octubre*), que obligaba a los operadores de televisión a destinar un mínimo del 5% de los ingresos de explotación a la financiación de largometrajes y películas para televisión europeas.

2.2. El efecto *Farmacia de guardia*

El proceso de desregulación, que autorizaba la creación de las cadenas privadas, introdujo a partir de 1990 un nuevo régimen de competencia caracterizado por una ofensiva de estas últimas a *TVE*, que copiaba las propias estrategias de la cadena pública (**Artero**; **Herrero**; **Sánchez-Tabernero**, 2005, p. 87). La incorporación de *Antena3* y *Tele5* a la producción de ficción en 1991 abría una tercera etapa que impulsaría la transformación del sistema tras el dominio durante el decenio anterior de un modelo dependiente de la industria cinematográfica.

Las perspectivas de expansión indujeron a los productores audiovisuales a aunar esfuerzos para poder optimizar las ayudas públicas y la difusión de sus productos. Nacía así la *Federación de Asociaciones de Productores Audiovisuales* (*Fapae*), que en 1999 ya reunía 323 empresas, la práctica totalidad de las productoras existentes en aquel momento (*Fapae*, 2000, p. 14). En 2010 el número de asociadas ascendía a 400 (*Fapae*, 2010, p. 7)[3].

> Internet se ha convertido en el mejor aliado de la ficción televisiva, que amplifica su alcance sin fragmentar el producto

El éxito de *Farmacia de guardia* (*Antena3*, 1991-1995) estimuló la producción de ficción televisiva e indujo a las cadenas privadas a externalizarla. Una tendencia a la que también se sumaría progresivamente *TVE* tras la "interrupción en la producción propia" que, según Antonio Mercero, se habría producido en los años ochenta por la alianza de la televisión pública con el cine (Mercero, en **Veiga**; **Ibáñez**, 2006, p. 114). El alcance de *Farmacia de guardia*, cuyos 169 episodios reportaron a *Antena3* un *share* medio del 48%, sobrepasó su propio éxito de audiencia y afianzó la creencia de que la ficción española podía constituir una verdadera alternativa a la ficción estadounidense y a las telenovelas latinoamericanas, además de fijar el canon "para muchas telecomedias posteriores" (**Palacio**, 2002, p. 181).

2.3. Consolidación del sistema

Médico de familia (*Tele5*, 1995 y 1999), otra de las series españolas más populares, canalizaba el tirón de su predecesora y abría una cuarta etapa con un sector audiovisual en expansión, integrado por una multitud de pequeñas empresas que iban surgiendo de la diferenciación de las competencias requeridas en cada fase de la producción (**Azpillaga**; **De-Miguel**; **Zallo**, 1998). La serie de *Globomedia*, que había adoptado las técnicas productivas de la ficción norteamericana (*GECA*, 1996; **García-de-Castro**, 2008; **Medina-Laverón**, 2008), impulsó un movimiento renovador de la ficción española que atribuía a las pequeñas productoras independientes

> "un sustancial protagonismo en la dinamización del sector" (**García-de-Castro**, 2008, p. 151).

En este cuarto período las ayudas al audiovisual recibieron un notable impulso con la apertura en 1997 de la oficina *Media España*, destinada a asesorar e informar a los profesionales. Ese mismo año surgía el *Programa de Desarrollo Audiovisual Ibermedia* con el fin de estimular las coproducciones iberoamericanas. Sin embargo algunos investigadores responsabilizan a las políticas públicas de los años noventa de fomentar la creación de empresas inoperantes,

> "anquilosadas en el continuo recurso a la protección pública" (**Pablo-Martí**; **Muñoz-Yebra**, 2001, pp. 132-133).

Otros autores señalan que a pesar de la relevancia de las diferentes medidas de apoyo al audiovisual en la expansión de la producción televisiva, el sector presentaba aún grandes deficiencias a comienzos de 2000 a consecuencia tanto de la atomización como de "la mezcla en un mismo saco de las condiciones legales de producción" para los diferentes tipos de empresas (**Fernández-Beaumont**, 2005, p. 2). De hecho, además de las cadenas y de las productoras independientes de nueva creación, muchas productoras de televisión habían surgido desde el entono cinematográfico, animadas en una clara apuesta por la diversificación de un negocio que se resentía particularmente de las crisis cíclicas (**Fernández-Quijada**, 2008).

La atomización del sector llevó a las cadenas a apostar por las productoras responsables de los programas de éxito, lo que favorecía la expansión de unas pocas empresas que controlaban la totalidad del mercado, mientras que la mayor parte tan sólo llegaba a producir una única ficción (**Bustamante**; **Álvarez-Monzoncillo**, 1999; **Guerrero**, 2010). Por otra parte la formación de conglomerados nacionales e internacionales en el mercado televisivo español, a consecuencia de la transformación de las empresas de difusión en empresas multimedia, avanzaba desde principios de 2000 hacia una migración digital tan difícil como imparable, que apuntaba asimismo a una concentración vertical sin otros mecanismos de contrapeso (**Álvarez-Monzoncillo**; **Zallo**, 2002).

La apuesta por las productoras de éxito y la formación de conglomerados propiciaron inicialmente un proceso de integración horizontal, que a su vez impulsó la especialización y la expansión internacional tras más de tres décadas de penetración estadounidense y de consumo interior de los productos locales. Con todo, *TVE*, *Antena3* y *Tele5* seguían siendo responsables de la mayor parte de la producción independiente española, que entre 1999 y 2001 ascendía aproximadamente al 45% del total de emisiones (**Pardo**, 2001, p. 45). De ahí que en 1999 no se pudiera hablar en España de producción independiente en sentido estricto, sino más bien de una producción delegada o financiada por las cadenas (**Álvarez-Monzoncillo**; **López-Villanueva**, 1999).

> Las televisiones estatales estrenaron 23.684 episodios/capítulos de ficción propia entre 1990 y 2017

La salida a bolsa de las empresas televisivas entre 2000 y 2004 ratificaba la consolidación del sistema audiovisual español, cuyas inversiones se multiplicaban debido a

> "un mercado publicitario en crecimiento sostenido y a la mejora de la gestión de costes" (**Artero**; **Herrero**; **Sánchez-Tabernero**, 2005, p. 89).

La creación de *Cuatro* y de *LaSexta* en 2005, en plena bonanza económica, y la creciente demanda de los canales autonómicos, contribuyeron a intensificar estas tendencias, con una aportación cercana al 10% de la producción independiente española entre las temporadas 2004-2005 y 2005-2006 (**Fernández-Quijada**, 2008, p. 84).

2.4. Internacionalización y concentración vertical

El apagón analógico de 2010 inauguraba el período actual y aceleraba un proceso de internacionalización que en 2014 situaba a España entre los principales proveedores mundiales de formatos de ficción televisiva (*The Wit*, 2014). La renuncia de *TVE* a emitir publicidad, la absorción de *Cuatro* por *Tele5* y de *LaSexta* por *Antena 3*, y la entrada de *Mediaset* y *Telefónica* en *Digital+*, han impulsado una concentración del sector audiovisual, auspiciada a su vez por el

> "giro copernicano de política comunicativa en un sentido neoliberal"

que se produjo en el segundo gobierno de José Luis Rodríguez Zapatero (**Zallo**, 2010, p. 12).

A diferencia de las empresas de producción de contenidos periodísticos, cuyos contenidos son mezclados con información sesgada y compiten en la Red con las fuentes de información en tiempo real del "periodismo ciudadano" y de las redes sociales (**Álvarez-Monzoncillo**; **Suárez-Bilbao**; **De-Haro**, 2016, p. 268), internet se ha convertido en el mejor aliado de la ficción televisiva, que amplifica su alcance sin fragmentar el producto. En paralelo las ventanas de oportunidades a la explotación se incrementan y están en constante evolución por efecto de las transformaciones en la manera en que la televisión es distribuida y consumida (**Doyle**, 2016). Pero a pesar de que las numerosas oportunidades de comercialización de los contenidos incrementan el protagonismo de los productores, la gran inversión que requiere la producción audiovisual aboca a las empresas independientes a integrase en los grandes grupos multimedia controlados por las cadenas de televisión o a aliarse con las grandes plataformas internacionales (**Álvarez-Monzoncillo**; **López-Villanueva**, 2014).

Las operaciones corporativas y la legislación han terminado por construir una especie de "oligopolio imperfecto" (**García-Santamaría**; **Fernández-Beaumont**, 2012), en una situación

> "propia de un mercado que, como el de la producción independiente, presenta un número reducido de clientes potenciales" (**Fernández-Quijada**, 2008, p. 65).

3. Método

Este artículo analiza la producción de ficción televisiva española estrenada entre 1990 y 2017. La muestra de análisis comprende 490 programas (series, seriales, miniseries, *TVmovies* y *sketches*) en cuya elaboración participaron en diferentes regímenes de producción 234 empresas audiovisuales.

La base de datos de tipo relacional, utilizada para almacenar, organizar y explotar la información a través de consultas SQL, incluye atributos de tipo cadena (productoras, descripción de la ficción y comentarios), de carácter numérico (datos de audiencia) y fechas (día, mes y año de emisión).

> La creciente especialización y concentración vertical convive con la atomización de las productoras independientes

La amplitud de la muestra y el extraordinario dinamismo del contexto audiovisual español e internacional han determinado, a lo largo de los 28 años que constituyen la referencia del estudio, la creación, desaparición, fusión o integración en diferentes conglomerados de numerosas productoras audiovisuales. Con el fin de mantener la coherencia de la investigación, se ha optado por conservar la identidad de las empresas en el momento en que realizaron la producción, incluso en aquellos casos en los que sucesivamente ha variado la composición de su accionariado.

La panorámica sobre la evolución de la producción de ficción televisiva en España, presentada en el primer apartado del

Tabla 1. Producción de ficción televisiva española 1990-2017

Décadas	Canales	Producciones	Episodios/ capítulos	Total horas	*Share*
1990-1999	4	131	6.533	4.180	21
2000-2009	6	180	9.373	7.606	15
2010-2017	18	179	7.778	6.492	12
Total		490	23.684	18.279	15

Tabla 2. Productoras agrupadas por número de títulos estrenados

Producciones	Productoras	% del total
1 título	144	62
2-9 títulos	78	33
10-19 títulos	7	3
20+ títulos	5	2
Total	234	100

artículo, nos ha permitido identificar tres etapas diferenciadas en el período comprendido entre la desregulación y la actualidad: 1990-1999, 2000-2010 y 2011-1017, utilizadas como referencia para estructurar el análisis. La distribución de los programas por etapas se ha realizado tomando como referencia el año de estreno, independientemente del número de temporadas que hubieran podido permanecer en antena.

4. Resultados

Las televisiones estatales estrenaron 23.684 episodios/capítulos de ficción propia entre 1990 y 2017, correspondientes a 18.279 horas de emisión, que reportaron a las cadenas una media de 2.318.363 de espectadores (15% de *share*[4]). Por etapas se observa un incremento progresivo de los estrenos, con 131 entre 1990-1999, 180 entre 2000 y 2009, y 179 entre 2010 y 2017.

La comedia y el drama son los dos géneros que cuentan con una mayor presencia en el período analizado, con 182 y 174 estrenos respectivamente. Sin embargo, mientras que el número de producciones dramáticas se incrementa de manera paulatina a lo largo de las tres etapas estudiadas (38, 56 y 80 estrenos respectivamente), la comedia parece haberse estancado, con 69 estrenos en cada una las dos primeras décadas y 44 en los ocho años que comprende la etapa actual. Los géneros de intriga (policíaco y thriller) también experimentan un crecimiento significativo, con 15 estrenos en la primera década, 46 en la segunda y 35 en el período actual.

La serie es el formato predominante de la ficción española, con 293 estrenos (60% del total), seguida por la miniserie (84 estrenos y 17%) y el serial (50 estrenos y 10%), aunque el *share* de la miniserie es el más elevado del conjunto y oscila menos que el del resto de los formatos (21% en la primera década, 19% en la segunda y 14% entre 2010 y 2017). A partir de 2010 se consolida el auge de la miniserie iniciado a finales de la década anterior y se estabiliza el estándar generalizado de 13 episodios por temporada/año en la serie, al tiempo que el número de seriales disminuye. En consecuencia se incrementan las producciones (179 entre los ocho años de la década actual incluidos en el análisis), al tiempo que el número de episodios/capítulos se reduce en un 17% y el tiempo de emisión en un 15% respecto de las dos décadas precedentes.

El número de productoras en activo se mantuvo estable entre 1990-1999 y 2000-2009, con 95 y 97 empresas respectivamente, mientras que el período comprendido entre 2010 y 2017 registra un aumento del 9% respecto al anterior. El incremento es significativamente mayor en las empresas que han participado en la producción de miniseries y *TVmovies* que en el resto, con 22 productoras en la primera década, 35 en la segunda y 54 en el tercer período. Sin embargo tan sólo el 2% de las 234 empresas audiovisuales activas entre 1990 y 2017 han producido más de 20 títulos (*RTVE*, *Mediaset*, *Atresmedia*, *Globomedia* y *BocaBoca*), mientras que el 62% (144 empresas) han participado en la producción de una única ficción. También cabe destacar que sólo nueve empresas han producido ficción televisiva a lo largo de los tres períodos de referencia: los tres grandes grupos estatales (*RTVE*, grupo *Antena3* y grupo *Mediaset*), cinco productoras españolas independientes (*Alba Adriática*, *BocaBoca*, *Diagonal TV*, *Globomedia* y *José Frade Producciones Cinematográficas*) y una empresa alemana (*Beta Film*).

La mayor parte de la ficción española ha sido realizada mediante la modalidad de producción externa financiada (331 títulos, 68%), seguida a considerable distancia por la producción externa asociada (83 títulos, 17%), la producción interna (49 títulos, 10%) y las coproducciones internacionales (27 títulos, 5%). Se observa además que el incremento de la producción externa financiada mantiene un ritmo sostenido década tras década, en detrimento de las modalidades de producción interna y asociada. La producción interna pasa del 32% de las fic-

Tabla 3. Tipología de productoras y actividad por períodos

Tipo de productora	1990-1999		2000-2009		2010-2017		Total	
	Títulos		Títulos		Títulos		Títulos	
	n	%	n	%	n	%	n	%
Interna	42	32	4	2	3	2	49	10
Externa financiada	47	36	140	78	144	80	331	68
Externa asociada	29	22	30	17	24	13	83	17
Coproducción	13	10	6	3	8	4	27	6
Total	131	100	180	100	179	100	490	100

ciones estrenadas entre 1990 y 1999 al 2% de cada uno de los dos períodos siguientes.

Los tres grandes grupos de televisión en abierto (*RTVE*, *Mediaset*, *Atresmedia*) y la productora independiente *Globomedia* encabezan el número de estrenos. De manera acorde con la progresiva externalización de la producción de ficción televisiva, la actividad de *Globomedia* se incrementó a partir del año 2000, con el 87% de los 40 estrenos producidos por esta empresa entre 2000-2009 y 2010-2017. Por el contrario la mayor parte de la producción de *RTVE* se remonta a los años 90, con 52 títulos equivalentes al 72% del total de sus 74 estrenos.

Tabla 4. Ranking de las empresas por número de producciones

Productoras	1990-1999	2000-2009	2010-2017	Total
RTVE	53	13	8	74
Globomedia	6	27	13	46
Mediaset	14	17	8	39
Atresmedia	13	8	7	28
BocaBoca	4	9	7	20
Diagonal TV	2	7	10	19
Bambu	-	1	15	16
Cartel	6	9	-	15
Ida y Vuelta	-	8	3	11
Videomedia	-	8	3	11
Total	98	107	74	279

5. Discusión y conclusiones

La producción propia de ficción televisiva se ha ido afianzado en el panorama audiovisual español a lo largo de los 28 años estudiados, con un incremento progresivo de los estrenos, una mayor diversificación de géneros, formatos y contenidos, y una reducción significativa del número de episodios/capítulos, que reflejan la madurez y versatilidad del sistema. En línea con las tendencias internacionales, la reducción del número de episodios de las series y la preferencia por las tramas de continuidad manifiestan su creciente hibridación con la miniserie, acorde con el auge de este último formato desde 2010.

El abandono progresivo de las modalidades de producción interna y externa asociada a favor de la producción externa financiada ha contribuido a perpetuar la tradicional atomización de un sector en el que más del 60% de las productoras independientes han participado en un único proyecto entre 1990 y 2017. Al mismo tiempo la tendencia a la especialización de los contenidos en un contexto de creciente competitividad y la dependencia endémica de las grandes cadenas por parte de las productoras televisivas, han impulsado la concentración vertical del sector audiovisual. En consecuencia algunas de las mayores productoras españolas aparecen asociadas a determinados géneros o subgéneros cada vez con mayor frecuencia, al tiempo que 9 de las 10 empresas que han participado en la producción del 51% de las 490 ficciones estrenadas en el período de referencia[5] se han integrado o están participadas por grandes grupos empresariales o de televisión. Estos datos concuerdan con el status de las productoras de televisión generalista, según un estudio de la consultora *Dos30* que atribuía a 10 empresas la producción del 71% de los programas emitidos en 2016 (**Espinel**, 2017).

> La mayor parte de la ficción española ha sido realizada mediante la modalidad de producción externa financiada

Al igual que ocurre en otros sectores de las ICC, la producción de ficción televisiva se configura como una "larga cola" (**Anderson**, 2006), con una gran cabeza integrada por el 5% de productoras (12 empresas) que han participado en más de 10 estrenos y una "larga cola" constituida por el 95% restante.

La verificación empírica realizada confirma la situación de oligopolio que caracteriza la estructura de la producción audiovisual española, donde el difícil equilibrio entre la atomización del sistema y la tendencia a la concentración vertical se mantiene a expensas de la dependencia de las productoras independientes respecto de las grandes cadenas (**Álvarez-Monzoncillo**; **Baraybar-Fernández**; **López-Villanueva**, 2015). Una dependencia endémica, acentuada por la obligatoriedad de invertir el 5% de sus ingresos en obras cinematográficas por parte de los operadores de televisión en abierto y de telecomunicaciones (tras la entrada en vigor de la *Ley del Audiovisual* de 2010), que ha impedido afianzar la producción independiente,

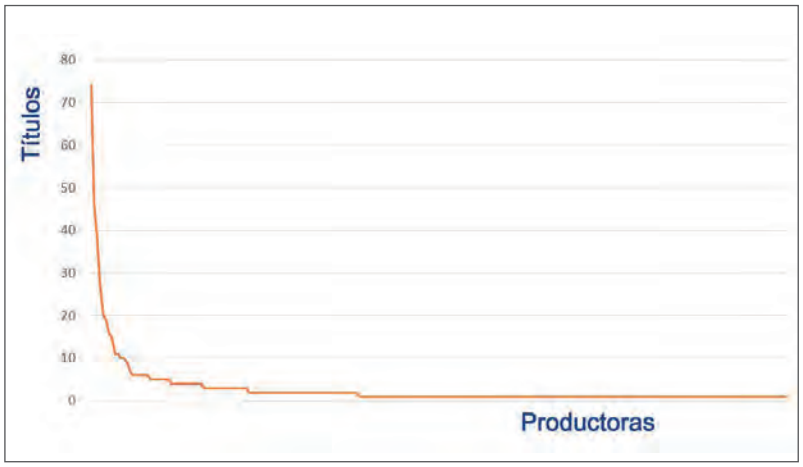


Gráfico 1. Larga cola de la producción de ficción española

"desplazada por los procesos de integración de las grandes productoras con los operadores y subordinada a la externalización de riesgos de los difusores" (**Álvarez-Monzoncillo**; **Zallo**, 2002, p. 5).

La incursión en la producción propia de

los distribuidores de ficción en *streaming* (*Netflix*, *Amazon Prime Video*, etc.) y de los teleoperadores (*Movistar+*), con una clara apuesta por la televisión de nicho, prefiguran una reconfiguración cuyo alcance a medio y largo plazo parece indisolublemente ligado al futuro, difícil de vislumbrar por el momento, de la televisión generalista en abierto.

> Más del 60% de las productoras independientes únicamente han participado en una sola ficción, mientras que 10 empresas son responsables de más de la mitad de los estrenos

Notas

1. Nos basamos en la tipología de **Bustamante** (1999) para diferenciar las modalidades de producción identificadas en este estudio. La serie *Sé quién eres* (*Tele5*, 2017) constituye un caso especial, pues se trata de una producción de *Tele5* realizada a partir de un guion elaborado por Pau Freixas y vendido después a la cadena (**Onieva**, 2017).

2. *http://www.ine.es/varipc*

3. En el momento de escribir estas líneas, *EGEA* afronta un posible concurso de acreedores que culminaría una crisis abierta hace tres años, agravada en 2017 con la salida de una buena parte de sus asociados (**Belinchón**, 2018).

4. Los datos de audiencia proceden del 86% de las 490 ficciones incluidas en la muestra de análisis. No se dispone de datos relativos al 14% restante, emitidas en su mayor parte en los primeros años noventa.

5. La décima es *Cartel Producciones*, inactiva en el momento de escribir este artículo.

7. Referencias

Álvarez-Monzoncillo, José-María; **Baraybar-Fernández, Antonio**; **López-Villanueva, Javier** (2015). "Audiovisual production in Spain. Fewer resources, same problems, new challenges". *Economia della cultura*, v. 2, pp. 211-222.
http://www.rivisteweb.it/doi/10.1446/81241
https://doi.org/10.1446/81241

Álvarez-Monzoncillo, José-María; **López-Jiménez, Javier** (1999). "La producción de ficción en España: un cambio de ciclo". *ZER*, v. 4, n. 7, pp. 65-88.
http://www.ehu.eus/ojs/index.php/Zer/article/view/17398

Álvarez-Monzoncillo, José-María; **López-Villanueva, Javier** (2014). "El audiovisual español: evolución en curso". En: Bustamante, Enrique; Rueda, Fernando (eds.). *Informe sobre el estado de la cultura en España*. Madrid: Lúa Ediciones 3.0, pp. 65-74.
http://www.fundacionalternativas.org/public/storage/publicaciones_archivos/74a26d1a8b4c8f2bc9c5e92b2cc915fb.pdf

Álvarez-Monzoncillo, José-María; **Suárez-Bilbao, Fernando**; **De-Haro, Guillermo** (2016). "Challenges and considerations of the new labor market in the media industry". *El profesional de la información*, v. 25, n. 2, pp. 262-271.
https://doi.org/10.3145/epi.2016.mar.13

Álvarez-Monzoncillo, José-María; **Zallo, Ramón** (2002). "Las políticas culturales y la comunicación para el desarrollo de los mercados digitales: un debate necesario". *Zer*, v. 7, n. 13, pp. 1-9.
http://www.ehu.eus/ojs/index.php/Zer/article/view/6022

Anderson, Chris (2006). *La economía long tail. De los mercados de masas al triunfo de lo minoritario*. Barcelona: Tendencias Editores. ISBN: 978 84 92452316

Artero, Juan-Pablo; **Herrero, Mónica**; **Sánchez-Tabernero, Alfonso** (2005). "Monopolio, oligopolio y competencia en los últimos quince años". *Sphera publica*, n. 5, pp. 83-98.
http://sphera.ucam.edu/index.php/sphera-01/article/view/24

Azpillaga, Patxi; **De-Miguel, Juan-Carlos**; **Zallo, Ramón** (1998). "Las industrias culturales en la economía internacional". *ZER*, v. 3, n. 5, pp. 53-74.
http://www.ehu.eus/ojs/index.php/Zer/article/view/17360

Belinchón, Gregorio (2018). "Los productores de cine cierran su histórica asociación por quiebra". *El país*, 15 febrero.
https://elpais.com/cultura/2018/02/14/actualidad/1518629258_075425.html

Bustamante, Enrique (1999). *La televisión económica. Financiación, estrategias y mercados*. Barcelona: Gedisa. ISBN: 84 7432 745 8

Bustamante, Enrique; **Álvarez-Monzoncillo, José-María** (1999). "España: la producción audiovisual en el umbral digital". *Zer*, v. 4, n. 7, pp. 45-64.
http://www.ehu.eus/ojs/index.php/Zer/article/view/17397

Bustamante, Enrique; **Zallo, Ramón** (coords.) (1988). *Las industrias culturales en España. Grupos multimedia y transnacionales.* Madrid: Akal. ISBN: 978 84 76003398

Deloitte (2017). *Televisión en abierto. Contribución a la sociedad española*.
https://goo.gl/gcAaPB

Doyle, Gillian (2016). "Digitization and changing windowing strategies in the television industry: Negotiating new windows on the world". *Television and new media*, v. 17, n. 7, pp. 629-645.
http://eprints.gla.ac.uk/119595/
https://doi.org/10.1177/1527476416641194

Espinel, Rodrigo (2017). "¿Cuáles son las 45 productoras que mandan en la televisión española?". *Produccion audiovisual*, 14 marzo.
https://produccionaudiovisual.com/produccion-tv/45-productoras-mandan-television-espanola

Fapae (2000). *Informe anual*. Federación de Asociaciones de Productores Audiovisuales Españoles.

Fapae (2010). *Informe anual*. Federación de Asociaciones de Productores Audiovisuales Españoles.

Fernández-Beaumont, José (2005). "Movimientos de concentración en un sector muy atomizado". *Telos*, n. 65, pp. 1-5.
https://telos.fundaciontelefonica.com/archivo/numero065/movimientos-de-concentracion-en-un-sector-muy-atomizado

Fernandez-Quijada, David (2008). "El mercado de producción independiente en España ante la aparición de Cuatro y LaSexta". *Comunicación y sociedad*, v. 22, n. 1, pp. 59-88.
https://www.unav.es/fcom/communication-society/es/articulo.php?art_id=21

García-de-Castro, Mario (2002). *La ficción televisiva popular. Una evolución de las series de televisión en España*. Barcelona: Gedisa. ISBN: 84 7432 981 7

García-de-Castro, Mario (2008). "Los movimientos de renovación en las series televisivas españolas". *Comunicar*, v. 30, pp. 147-153.
https://doi.org/10.3916/c30-2008-02-008

García-Santamaría, José-Vicente; **Fernández-Beaumont, José** (2012). "Concentración y pluralismo en el nuevo escenario audiovisual español. Análisis de la absorción de la cadena Cuatro por Telecinco". *Anàlisi. Quaderns de comunicació i cultura*, n. 47, pp. 57-74.
https://www.raco.cat/index.php/Analisi/article/view/304839

GECA (1996). *Cómo se hizo Médico de familia*. Madrid: Ediciones GECA. ISBN: 978 84 605 5331 1

Guerrero, Enrique (2010). "Las productoras de contenidos: protagonistas de una industria multimedia. El caso de España". *Revista de comunicación*, v. 9, pp. 150-173.

Hesmondhalgh, David (2002). *The cultural industries*. London: Sage. ISBN: 978 1 446209264

Hesmondhalgh, David; **Pratt, Andy C.** (2005). "Cultural industries and cultural policy". *International journal of cultural policy*, v. 11, n. 1, pp. 1-14.
http://individual.utoronto.ca/diep/c/hesmondhalgh.pdf
https://doi.org/10.1080/10286630500067598

Lacalle, Charo (2018). "Memoria, historia y democracia: la ficción en el período socialista (1982-1999)". En: Montero-Díaz, Julio (dir.). *Una televisión con dos cadenas. Programas y programación (1956-1990)*. Madrid: Cátedra, pp. 611-636. ISBN: 978 84 37638409

Medina-Laverón, Mercedes (2008). *Series de televisión. El caso de "Médico de familia", "Cuéntame cómo pasó" y "Los Serrano"*. Madrid: Ediciones Internacionales Universitarias. ISBN: 978 84 84692423

Onieva, Álvaro (2017). "Los secretos de 'Sé quién eres' según Pau Freixas". *Fotogramas*, 3 junio.
http://www.fotogramas.es/series-television/Se-quien-eres-serie-Telecinco-secretos-Pau-Freixas

Ontsi (2017). *Informe anual del sector de los contenidos digitales en España*. Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información.
https://bit.ly/2CbJok1

Pablo-Martí, Federico; **Muñoz-Yebra, Carlos** (2001). "Economía del cine y del sector audiovisual en España". *Información comercial española ICE*, v. 792, pp. 124-138.

Palacio, Manuel (2001). *Historia de la televisión en España*. Barcelona: Gedisa. ISBN: 978 84 74328066

Palacio, Manuel (2002). "Enseñar deleitando. Las adaptaciones literarias en televisión". *Cuadernos de la Academia*, n. 11-12, pp. 519-537.

Pardo, Alejandro (2001). "La producción televisiva: un sector en alza". En: García-Matilla, Agustín; Arnanz, Carlos (eds.). *La nueva era de la televisión*. Pozuelo de Alarcón: Academia de las Ciencias y las Artes de Televisión de España, pp. 44-51. ISBN: 84 607 3469 2
http://dadun.unav.edu/handle/10171/38223

The Wit (2014). *The Wit guide to scripted formats 2014*.
http://www.miptrends.com/wp-content/uploads/2016/11/miptv-mipcom-thewit-wp-scripted-formats-2014.pdf

United Nations (2008). *Creative economy. Report 2008. The Challenge of Assessing the Creative Economy: towards informed policy-making.* ISBN: 978 0 9816619 0 2
https://unctad.org/en/docs/ditc20082cer_en.pdf

Veiga, Yolanda; **Ibáñez, Isabel** (2006). *Religión catódica. 50 años de televisión en España*. Madrid: Rama Lama Music. ISBN: 978 84 93430764

Zallo, Ramón (2010). "La política de comunicación audiovisual del gobierno socialista (2004-2009): un giro neoliberal". *Revista latina de comunicación social*, v. 65, pp. 14-29.
https://doi.org/10.4185/RLCS-65-2010-880-014-029

Fiscalidad indirecta del libro en Europa: implicaciones y políticas

Indirect taxation of the book in Europe: Implications and policies

Marta Magadán-Díaz; **Jesús Rivas-García**

Cómo citar este artículo:

Magadán-Díaz, Marta; **Rivas-García, Jesús** (2019). "Fiscalidad indirecta del libro en Europa: implicaciones y políticas". *El profesional de la información*, v. 28, n. 1, e280111.

https://doi.org//10.3145/epi.2019.ene.11

Artículo recibido el 07-07-2018
Aceptación definitiva: 13-11-2018

Marta Magadán-Díaz ✉
https://orcid.org/0000-0003-3178-3215

Universidad Internacional de La Rioja
Facultad de Empresa y Comunicación
Avda. de la Paz, 137. 26006 Logroño, España
marta.magadan@unir.net

Jesús Rivas-García
https://orcid.org/0000-0003-0576-5961

Universidad Internacional de La Rioja
Facultad de Empresa y Comunicación
Avda. de la Paz, 137. 26006 Logroño, España
jesus.rivas@unir.net

Resumen

El objetivo de este estudio es determinar el papel de la tributación indirecta en la industria editorial española. En el trabajo se realiza: a) una comparación del tratamiento fiscal del libro en España frente a los países europeos de su entorno, b) una evaluación de las opciones de exención o reducción del impuesto sobre el valor añadido (IVA) aplicado a los libros y c) un análisis de la política de precio fijo del libro en Europa. Las conclusiones del trabajo apuntan, en primer lugar, a la necesidad de armonizar el IVA en el espacio de la UE; en segundo lugar, a reducir la discriminación fiscal entre formatos; y en tercer lugar, a subrayar que la tributación indirecta podría convertirse, si no lo es ya, en un freno al desarrollo de la digitalización de la industria editorial española.

Palabras clave

Libro; Industria editorial; Fiscalidad; Impuesto sobre el valor añadido; IVA; Precio fijo del libro; Discriminación fiscal; Digitalización; Europa; España.

Abstract

The main goal of this study is to determine the role of indirect taxation in the Spanish publishing industry. The work includes: a) a tax treatment comparison of books in Spain with respect to the rest of European countries; b) an evaluation of the options for exemption or reduction of the value added tax (VAT) applied to books; and c) an analysis of book fixed-price policy applied in Europe. The conclusions of this work point, firstly, to the need to harmonize VAT in the EU space; secondly, to reduce tax discrimination between formats and, thirdly, to emphasize that indirect taxation could be converted, if it is not already, in a brake for the development of the digitization of the Spanish publishing industry.

Keywords

Book; Publishing industry; Taxation; Value added tax; VAT; Book's fixed price; Tax discrimination; Digitization; Europe; Spain.

1. Introducción

El libro, al igual que otros bienes y servicios culturales, está experimentando un proceso crucial de cambio tecnológico (**Carreiro**, 2010; **Donoughue**, 2010; **Lichtenberg**, 2011) que le hace transitar del formato papel al formato digital.

En la última década el impacto transversal del sector tecnológico sobre el editorial ha provocado la mayor transformación de la industria desde Gutenberg y la aparición de la imprenta (**Carreiro**, 2010), incrementado significativamente la complejidad del mercado (**Magadán**; **Rivas**, 2018a; 2018b). La imposición indirecta tampoco permanece ajena al nuevo contexto, marcando las diferencias entre ambos soportes: el papel y el digital (**Colbjørnsen**, 2015).

La rápida difusión y mejora de las herramientas informáticas y de internet ha dado lugar a que los bienes y servicios culturales se produzcan y distribuyan de una manera diferente a la empleada para sus soportes físicos, exigiendo a la industria una adaptación de sus procesos productivos a los nuevos continentes y a nuevas regulaciones públicas (**Terzi**, 2011; **Martin**; **Tian**, 2016), promoviendo al tiempo un cambio notable en los hábitos de consumo del público (**Hibbert**, 1999; **Lichtenberg**, 2011). Por ejemplo, la impresión digital y las librerías virtuales cubren nuevas necesidades: han aumentado las posibilidades de visibilidad y de difusión de la producción editorial y han eliminado barreras geográficas y físicas (**Tian**; **Martin**, 2010; **Magadán**, 2017).

Internet se ha convertido en una ventana que permite a los creadores poner su trabajo a disposición de los ciudadanos sin recurrir exclusivamente a los canales minoristas convencionales (**Zhang**; **Zhu**; **Ye**, 2016), y los consumidores por su parte ya no necesitan moverse de su entorno para adquirir y disfrutar de bienes y servicios. La posibilidad de interacción entre proveedores, oferta y demanda sin la necesidad de un mercado físico, implica cambios sustanciales en los elementos esenciales del intercambio: desde precios hasta el medio de difusión y el rango de este último, por ejemplo.

La creciente accesibilidad de la Web ha facilitado su uso fraudulento, generando uno de los problemas más graves en el sector cultural actual, la posibilidad de copia y distribución ilegal de dichos bienes (**Jain**, 2008). El aumento del ancho de banda y la mejora de la tecnología de compresión de archivos, mejora la distribución de los productos y servicios, pero también facilita el pirateo de libros, películas y programas de televisión (**Allen-Robertson**, 2013). Una vez que la información se transforma al formato digital, puede copiarse y distribuirse a un coste marginal cercano a cero (**Waldfogel**, 2017).

Los Estados han reaccionado modificando y reforzando la legislación sobre protección de derechos de autor, penalizando el acceso y las descargas ilegales. Sin embargo, el comportamiento oportunista del consumidor no sólo afecta al autor o la industria cultural, sino que también tiene consecuencias para la Hacienda Pública, ya que significa una pérdida importante de recursos.

Para evitar la piratería, las editoriales han tratado de mantener su posición de dominio, reclamando un reforzamiento de las leyes que regulan y protegen los derechos de autor, optado por el DRM (*digital rights management*) de *Adobe* como solución que les permita la protección de los derechos de las obras, controlar la duplicación, distribución y reproducción de los contenidos (**Boyle**, 2008). No obstante, algunos autores consideran que las descargas ilegales son una forma de publicidad que favorece la visibilidad de las obras (**Conner**; **Rumelt**, 1991; **Jiang**; **Sarkar**, 2003) y que en todo caso no impide que los lectores interesados dejen de acudir a los sitios web de editoriales y librerías online para adquirir una versión autorizada del libro (**Czerwinski**; **Fromm**; **Hodes**, 2007; **Van-Eijk**; **Poort**; **Rutten**, 2010; **Cordón-García**; **Gómez-Díaz**; **Alonso-Arévalo**, 2011). Cierto grado de piratería puede ser beneficioso e incluso puede favorecer procesos de innovación novedosos (**Conner**; **Rumelt**, 1991; **Jiang**; **Sarkar**, 2003; **Czerwinski**; **Fromm**; **Hodes**, 2007).

Esta investigación tiene como objetivo analizar a través del uso del método comparativo, el papel de la tributación indirecta en el negocio de las empresas editoriales desde una doble perspectiva:

a) la comparación del tratamiento fiscal del libro en España en relación con los países europeos de su entorno, situando el foco en la posibilidad de exención o reducción del impuesto sobre el valor añadido (en adelante, IVA) aplicado, de cara a promover la reflexión en las futuras agendas de investigación de políticas culturales y de la industria editorial española;

b) la consideración de la tributación indirecta como un freno al desarrollo del nuevo negocio digital emergente que facilite el tránsito de las empresas a una nueva realidad definida por el cambio tecnológico.

Para abordar esta doble perspectiva, se mostrarán las diferencias tributarias presentes en el IVA del libro tanto en papel como digital, en el contexto de la Unión Europea (en adelante, UE). Esta sección descriptiva es indispensable, tanto por la evidencia que subraya (desigualdad tributaria en el consumo de libros impresos y electrónicos) como por la generalización de las implicaciones que supone, siendo el IVA un impuesto de naturaleza europea.

En la práctica algunos Estados europeos han optado por reducir el IVA de los libros con el fin de estimular la lectura y la alfabetización, como por ejemplo los Países Bajos que bajaron el IVA que grava los libros 21 al 6%, o Irlanda del 21 al 0%, o Suecia del 25 al 6%. Francia y Luxemburgo introdujeron también un tipo reducido de IVA para libros electrónicos a partir de 2012, en 5,5 y 3%, respectivamente. Sin embargo, se hizo sin la aprobación de la *Comisión Europea* y como resultado en 2015 el *Tribunal de Justicia* de la UE dictaminó que los libros electrónicos suministrados por descarga o transmisión son un servicio y por lo tanto quedan fuera de la exención del IVA. La tasa del IVA se establece en función del operador y no del contenido, lo que resulta un incentivo negativo para el desarrollo tecnológico en Europa. Esto aparece

en marcado contraste con el apoyo sustancial y la financiación para mejorar y promover el patrimonio cultural digital de Europa (**Borowiecki**; **Navarrete**, 2018).

Este trabajo está estructurado de la siguiente manera: la sección 2 hace una breve exposición del marco teórico de la incidencia fiscal en equilibrio parcial para avanzar sobre el análisis de la no discriminación fiscal por formato de libro; la sección 3 proporciona una visión de la metodología utilizada; la sección 4 presenta el estudio comparativo sobre el tratamiento fiscal del libro en España y en los países europeos de su entorno; finalmente la última sección condensa las conclusiones del estudio.

2. Marco teórico de un cambio en el IVA sobre los precios y el bienestar

La lectura tiene un efecto directo en las habilidades del individuo (**Krashen**; **Lee**; **McQuillan**, 2012), como la comprensión lectora y la ortografía, que a su vez están asociadas con un incremento en rendimiento general (**Mol**; **Bus**, 2011), una mejora en la alfabetización, y también tiene efectos positivos sobre la economía local (**Dent**, 2007). El acceso a los libros también tiene beneficios a largo plazo, como el aumento del vocabulario, el conocimiento básico y las habilidades de comprensión (**Neuman**; **Moland**, 2016). La lectura y la alfabetización permiten procesar información que conducirá a la toma inteligente de decisiones, vital en nuestra sociedad, mediante la identificación de fuentes de información, la evaluación y organización de la información y su uso en el pensamiento crítico y la resolución de problemas (**Hughes-Hassell**; **Rodge**, 2007). Los gobiernos reconocen los beneficios de leer y apoyar las políticas que facilitan el acceso a los libros (**Colbjørnsen**, 2015). Reducir el impuesto del IVA sobre los libros es una de esas políticas que estimulan la lectura y la alfabetización (**Borowiecki**; **Navarrete**, 2018).

> Se deben reducir las diferencias de trato fiscal entre el libro en papel y el libro electrónico

La mayoría de los impuestos crean distorsiones en el sistema de precios de mercado. Estas distorsiones generan asignaciones ineficientes de recursos. Los efectos distributivos de los impuestos se estiman tomando como referencia su impacto formal o legal: los impuestos inciden sobre aquellos individuos a quienes la legislación les impone la obligación de pago. A través del análisis de la incidencia impositiva se puede demostrar que no necesariamente la carga de los impuestos es soportada por los individuos que legalmente están obligados a pagarlo. La razón es que la modificación de un impuesto altera el equilibrio de la economía. Por tanto, los consumidores y los productores reajustarán sus decisiones, respondiendo a los nuevos costes y precios relativos.

Aplicamos el análisis de la incidencia impositiva en equilibrio parcial. Supongamos que un mercado perfectamente competitivo, en ausencia de impuestos, la intersección de las funciones de oferta y demanda proporciona la cantidad de equilibrio eficiente del bien consumido.

$$X_D(p) = X_S(p),\ p = p_0 \qquad [1]$$

Donde $X_D(p)$ y $X_S(p)$ son las funciones de demanda y oferta, y p el precio del bien X. En este caso la demanda no depende de la renta ni de los precios de los demás bienes, p_0 es el precio de equilibrio.

Considérese la introducción de una tasa de IVA estándar sobre X al tipo t. El equilibrio asociado a la introducción de una tasa de IVA sería:

$$X_D(p) = X_S(p - t),\ p = p_1 \qquad [2]$$

donde p es el precio pagado por el consumidor. Diferenciando con respecto a t y operando se puede obtener:

$$\frac{dp}{dt} = \frac{e_X^S}{e_X^S - e_X^D} \qquad [3]$$

donde e_X^D y e_X^S son las elasticidades de demanda y oferta. Dado que el precio del productor es igual al precio pagado por el consumidor menos el impuesto, $p - t$, el cambio en el precio del productor será:

$$\frac{dp}{dt} - 1 = \frac{e_X^D}{e_X^S - e_X^D} \qquad [4]$$

La imposición de una tasa de IVA estándar sobre un bien o servicio cambia la curva de oferta que la desplaza hacia arriba, aumentando el precio del bien y disminuyendo la cantidad de bien intercambiado, lo que provoca una pérdida de bienestar originada por la distorsión de los precios creada por la carga fiscal.

El análisis de equilibro parcial, aunque presenta ciertas limitaciones, sirve para realizar una primera aproximación de los cambios en los tipos del IVA en los libros. Si se reduce el tipo impositivo del IVA aplicado sobre los libros, con la aplicación

de una tasa reducida (*tr*), entonces el nuevo equilibrio asociado a la reducción del IVA sería:

$$X_D(p) = X_S(p - tr),\ p = p_2 \qquad [5]$$

Al reducir la tasa de impuesto estándar (*tr*) a la tasa reducida (*tr*) desplazará la curva de oferta hacia abajo hacia el equilibrio eficiente. Una reducción fiscal da como resultado una disminución en el precio y un aumento en la cantidad; por tanto, la aplicación de una tasa reducida conduce a una menor pérdida de bienestar para la sociedad; y si el tipo impositivo que se aplica es cero, entonces se vuelve al equilibrio eficiente y no se produce ninguna distorsión.

3. Método

El método aplicado en el estudio empírico es análisis comparativo, de amplio y frecuente uso en las ciencias sociales (**Lijphart**, 1971; **Ragin**, 1987; **Smith**, 1991) y empleado en el caso que nos ocupa como una herramienta útil no sólo para describir desde una perspectiva sincrónica los aspectos clave de las comparaciones consideradas, sino para servir de fundamento a la interpretación de los hechos objeto de estudio (**Ragin**, 2006).

Este trabajo se mueve en el terreno de la política económica fiscal, donde las coordenadas espaciales y temporales definen un número reducido de casos para abordar el tratamiento fiscal que cada país considerado en el análisis aplica al libro en papel y el ebook. Todo ello hace que el análisis comparativo se muestre como una vía eficaz para los propósitos de esta investigación.

4. Análisis de los datos

Esta sección presenta el estudio comparativo sobre el tratamiento fiscal del libro en España y en los países europeos de su entorno.

4.1. Tipos de IVA y política fiscal europea

Antes de entrar en un análisis sobre el papel del IVA como herramienta de política cultural, es necesario hacer una aproximación a la figura de este tributo para posteriormente ver cómo se aplica en Europa, poniendo de relieve la disparidad de tratamientos fiscales entre los libros en papel y electrónicos según el país europeo considerado.

Aunque la UE cuenta con normas generales sobre el IVA, su aplicación puede variar –y de hecho lo hace– en cada país. Normalmente el IVA se paga sobre todos los bienes y servicios –salvo las exenciones establecidas en cada normativa nacional–, orientados tanto al consumidor final como a todos aquellos que se puedan dirigir a consumos intermedios dentro de las fases del proceso de producción de las organizaciones: de ahí la necesidad de determinar la cuota del IVA por diferencia entre el IVA repercutido (a los consumidores) y el IVA soportado (por la empresa en la adquisición y/o contratación de bienes y servicios necesarios para su producción).

El IVA es un impuesto indirecto que, si bien grava el valor añadido generado en todas las fases del proceso productivo/operacional de la empresa (sujeto pasivo), este acaba por trasladarse vía precios al consumidor final (contribuyente).

No todos los bienes y servicios están sujetos al mismo tipo im-

Tabla 1. Tipos de IVA aplicados en los Estados miembros

Estados miembros	Código	Tipo súper reducido	Tipo reducido	Tipo general
Alemania	DE	-	7	19
Austria	AT	-	10 / 13	20
Bélgica	BE	-	6 / 12	21
Bulgaria	BG	-	9	20
Chipre	CY	-	5 / 9	19
Croacia	HR	-	5 / 13	25
Dinamarca	DK	-	-	25
Eslovaquia	SK	-	10	20
Eslovenia	SI	-	9,5	22
España	ES	4,0	10	21
Estonia	EE	-	9	20
Finlandia	FI	-	10 /14	24
Francia	FR	2,1	5,5 / 10	20
Grecia	EL	-	6 / 13	24
Holanda	NL	-	6	21
Hungría	HU	-	5 / 18	27
Irlanda	IE	4,8	9 / 13,5	23
Italia	IT	4,0	5 / 10	22
Letonia	LV	-	12	21
Lituania	LT	-	5 / 9	21
Luxemburgo	LU	3,0	8	17
Malta	MT	-	5 / 7	18
Polonia	PL	-	5 / 8	23
Portugal	PT	-	6 / 13	23
Reino Unido	UK	-	5	20
República Checa	CZ	-	10 / 15	21
Rumanía	RO	-	5 / 9	19
Suecia	SE	-	6 / 12	25

Fuente: *European Commission*, 2018.
https://ec.europa.eu/taxation_customs/sites/taxation/files/resources/documents/taxation/vat/how_vat_works/rates/vat_rates_en.pdf

positivo y se pueden encontrar diferencias significativas, en tipos y/o exenciones para determinadas operaciones de compraventa, entre las cuales se incluyen las transacciones de naturaleza cultural (**Alino**; **Schneider**, 2012; **Álamo-Cerrillo**; **Lagos-Rodríguez**, 2016; **Stevenson**; **Balling**; **Kann-Rasmussen**, 2017; **Borowiecki**; **Navarrete**, 2018). Sirva como ejemplo la lista de tipos de IVA aplicados en la UE, recogidos en la tabla 1.

La utilización de tipos reducidos para bienes y servicios de naturaleza cultural, en principio obedece a la idea de que los ciudadanos accedan a estos bienes de un modo más económico y conseguir que se incentiven las ventas en el sector cultural (**Colbjørnsen**, 2015). La justificación tradicional para otorgar un tratamiento fiscal favorable a los bienes culturales se ha apoyado en la posible existencia de economías externas y de asimetrías de información (**Sanz-Sanz**; **Romero-Jordán**; **Prieto-Rodríguez**, 2003).

Los estudios empíricos sobre el impacto de la variación del IVA sobre el precio del libro muestran los siguientes resultados: en Suecia la reducción del IVA del 25% al 6% incrementó las ventas pero no el hábito de lectura; en Letonia la subida del IVA del 5% al 21% redujo las ventas y la producción editorial sin afectar tampoco al hábito de lectura (**Towse**, 2010; **Colbjørnsen**, 2015; **Borowiecki** y **Navarrete**, 2018); en Francia, donde el precio de los libros es fijo, la reducción del IVA en los libros en 1989 (del 5,5% al 2,1%) no se trasladó a los precios por los editores, quienes se apropiaron del espacio dejado por el impuesto para incrementar sus márgenes (**Perona**; **Pouyet**, 2010).

Lo que se deriva de todos estos casos, el sueco, el letón y el francés, es que en primer lugar no siempre mercado y hábitos de lectura van de la mano, y en segundo lugar las reducciones en la imposición indirecta pueden no llegar al consumidor final.

> La reducción del IVA es un factor positivo para las empresas implicadas en la industria del libro y para los consumidores

En definitiva, ni las políticas de fomento de la lectura necesariamente impactan en la facturación editorial, ni las reducciones de la fiscalidad indirecta tienen un impacto evidente en el aumento del hábito de lectura. Donde sí hay un impacto evidente de la variación de tipos de IVA es en las ventas y la actividad productiva de las empresas editoriales.

4.2. Reducción o exención del IVA

El consumo de libros está asociado a externalidades positivas. Los que leen libros no sólo aumentan su propia utilidad, sino que también generan efectos positivos para la sociedad; por ejemplo, en la forma de una mejor educación y alfabetización (**Krashen**; **Lee**; **McQuillan**, 2012).

La teoría del bienestar nos indica la existencia de peso muerto si se consumen bienes y servicios que incorporan externalidades. Esto es causado por el hecho de que los mercados no internalizan las externalidades y por lo tanto el comercio no ocurre en cantidades eficientes. Por lo tanto se requiere la acción de los legisladores para que se suministren y consuman las cantidades correctas.

Mientras que el IVA es un medio para generar ingresos para el Estado, la reducción o exención de dichos impuestos puede servir para propósitos políticos específicos para el Gobierno. Una característica común, tanto en la UE como en otros lugares, es la reducción de impuestos o la exención para productos de medios tales como periódicos, publicaciones periódicas y libros (**Colbjørnsen**, 2015).

A diferencia de los subsidios directos de producción cultural o negocios editoriales, la reducción / exención del IVA se considera una forma indirecta de subsidio (**Colbjørnsen**, 2015), dirigida a alterar las estructuras institucionales y de asignación de recursos (**Mitchell**, 2003). Conceder tasas reducidas o exención del IVA sería una forma de:

a) apoyar la producción de bienes que se consideran particularmente vitales o beneficiosos para la población (**Ahearne**, 2009) y

b) corregir las deficiencias y distorsiones del mercado para la promoción de un bien cultural (**Hesmondhalgh**, 2007).

La distinción entre el tipo cero y la exención es que el primero permite a las empresas editoriales recuperar el IVA soportado: así, el precio final del libro con tipo cero no incorpora absolutamente nada del impuesto. En cambio, si los libros estuviesen exentos el editor no podría recuperar el IVA soportado y, por consiguiente, el precio de venta incorporaría un componente impositivo asociado al consumo de bienes y servicios empleados el proceso productivo, haciendo de dicho componente un coste más de producción, que será trasladado vía precios y conducirá a un precio final del libro superior al que tendría con un tipo cero.

> La reducción del IVA no implicaría una merma significativa de recaudación para el Estado

Por lo tanto, sobre los principios teóricos expuestos es preferible una tasa cero de IVA que una exención. En definitiva, una tasa cero de IVA significa la creación de empleos y el desarrollo de una economía del conocimiento. Este razonamiento también se aplicaría al tratamiento fiscal de los libros electrónicos.

Sin embargo, en la UE mientras que el 92,6% de los países aplican un tratamiento especial de IVA a los libros impresos (en la mayoría de los casos una tasa de IVA reducida), sólo el 7,4% aplica una tasa cero de IVA (Reino Unido e Irlanda).

4.3. Precio fijo del libro

Bajo un sistema de precio fijo es el editor quien fija el precio final del libro, que debe ser respetado por todos los intermediarios del canal de distribución, quienes fijan su ganancia sobre una proporción de dicho precio final.

Los países con una importante industria editorial, en algún momento de su historia han introducido una ley o un acuerdo para fijar el precio de venta al público. En algunos países los sistemas de precios fijos del libro se establecen y se rigen por ley (por ejemplo, Alemania, España y Francia, entre otros) de tal manera que es el editor quien fija el precio de venta final al público y no los otros agentes que operan en el canal de distribución (distribuidoras y librerías, por ejemplo) para quienes su margen de ganancia se determina por una participación en ese precio y no por un incremento del mismo como puede ocurrir en otros sectores. En total 12 Estados miembros de la UE tienen regulaciones para fijar el precio de los libros impresos y al menos 8 de éstos ya han extendido dicha regulación a los libros electrónicos (**Benhamou**, 2015). En el resto de países de la UE, sin precio fijo, los libreros y los editores pueden acordar contractualmente el precio final de venta (ver tabla 2).

El argumento principal para establecer sistemas de precios fijos del libro que otorgan el poder de fijar el precio de venta final al editor sería el de proteger los valores culturales del mercado del libro (**Appelman**, 2003; **Van-der-Ploeg**, 2004; **Canoy**; **Van-Ours**; **Van-Der-Ploeg**, 2006). Sin embargo es difícil encontrar evidencia convincente, bien teórica o empírica, de que la política de precios fijos en los libros es la mejor opción (**Ringstad**, 2004; **Poort**; **Van-Eijk**, 2017).

Quienes se oponen a la fijación de precios fijos señalan que la industria funciona mejor cuando opera en condiciones de libre mercado y que el precio fijo aumenta artificialmente los precios que las personas pagan por los libros (**Løyland**; **Ringstad**, 2012), es decir, los minoristas deberían poder determinar libremente los precios de sus productos, en función de la demanda real (**Poort**; **Van-Eijk**, 2017).

Los oponentes a la fijación de precios fijos en los libros creen que los gobiernos pueden apoyar mejor a la industria del libro mediante políticas que respalden la alfabetización y el desarrollo de las industrias creativas. En la tabla 2 se muestran las medidas que utilizan los Estados para ayudar a la industria del libro.

Tanto en los 12 países miembros de la UE que tienen implantado el precio fijo del libro como en los 16 que no, sólo el 50% -en cada caso- articulan mecanismos de financiación pública directa para favorecer el desarrollo de la producción editorial. De ello se puede concluir que no hay una correlación clara entre las medidas públicas de fomento de la edición y el sistema de fijación de precios en la industria editorial.

Tabla 2. Comparativa de medidas y aplicación del IVA en los Estados miembros de la UE

		IVA			Medidas para respaldar los mercados de libros	
Estados miembros	**Código**	**Libro en papel**	**Libro electrónico**	**Tipo general**	**Precio fijo del libro**	**Financiación pública directa para editores de libros**
Alemania	DE	7	19	19	Sí	No
Austria	AT	10	20	20	Sí	Sí
Bélgica	BE	6	21	21	No	No
Bulgaria	BG	20	20	20	Sí	Sí
Chipre	CY	5	19	19	No	No
Croacia	HR	5	25	25	Sí	Sí
Dinamarca	DK	25	25	25	Sí	Sí
Eslovaquia	SK	10	20	20	No	Sí
Eslovenia	SI	9,5	22	22	No	Sí
España	ES	4	21	21	Sí	Sí
Estonia	EE	9	20	20	No	Sí
Finlandia	FI	10	24	24	No	No
Francia	FR	5,5	20	20	Sí	No
Grecia	EL	6	24	24	Sí	Sí
Holanda	NL	6	21	21	Sí	No
Hungría	HU	5	27	27	Sí	No
Irlanda	IE	0	23	23	No	Sí
Italia	IT	4	22	22	Sí	No
Letonia	LV	12	21	21	No	Sí
Lituania	LT	9	21	21	No	Sí
Luxemburgo	LU	3	17	17	No	No
Malta	MT	5	18	18	No	No
Polonia	PL	8	23	23	No	No
Portugal	PT	6	23	23	Sí	No
Reino Unido	UK	0	20	20	No	No
República Checa	CZ	15	21	21	No	No
Rumanía	RO	5	19	19	No	Sí
Suecia	SE	6	25	25	No	Sí

Fuente: elaborado a partir de *European Commission*, 2018.
https://ec.europa.eu/taxation_customs/sites/taxation/files/resources/documents/taxation/vat/how_vat_works/rates/vat_rates_en.pdf

Por otra parte, de los 14 países donde existe financiación pública directa para editores de libros, 9 de ellos presentan unos tipos de IVA aplicados en el libro en papel inferiores al 10% frente a los otros 6 cuyos tipos de IVA son iguales o superiores al 10%. De esta evidencia se puede deducir que las ayudas públicas directas se orientan claramente más a la producción en papel -donde se percibe esta diferencia significativa de tratamiento fiscal- que al libro electrónico y que reflejan una actitud claramente proactiva y de compromiso por parte de aquellos Estados que no sólo discriminan fiscalmente en positivo al libro en papel, beneficiando así a toda la cadena de distribución, sino que actúan como promotores públicos de la actividad editorial.

La *Ley 10/2007 de 22 de junio de la Lectura, del Libro y de las Bibliotecas* española tiene como finalidad ensalzar el valor del libro, el proceso lector e instaurar principios que regulen el comercio del libro. El Artículo 9 desarrolla la regulación respecto al precio del libro, que está sujeto a un precio fijo de venta al público que debe acatar toda la cadena de valor: el editor, el distribuidor y la librería.

Más allá de los posicionamientos a favor o en contra del precio fijo del libro, lo que resulta técnicamente incuestionable es que una reducción del IVA parcial o total (tipo cero) llegará al cliente final; sin embargo, con un sistema de precios pactados –no fijos- esa misma reducción no tiene porqué llegar al cliente final y podría acabar siendo absorbida por el margen de ganancia de cualquiera de los restantes miembros del canal de distribución.

4.4. Libros y tipos de IVA

Las diferencias entre los tipos impositivos en la provisión de bienes y servicios de naturaleza cultural son muy acusadas entre los Estados miembros. En Suecia por ejemplo, el tipo del IVA en libros impresos se redujo en 2002 del 25 al 6%, armonizando así la línea de la política de medios sobre libros (impresos), periódicos y publicaciones periódicas. Posteriormente se comprobó que la reducción del IVA había tenido un efecto positivo en la rentabilidad y la facturación de los editores de libros y revistas (**Colbjørnsen**, 2015) y que las ventas de libros aumentaron un 20% después de la reducción (**Towse**, 2010; **Borowiecki**; **Navarrete**, 2018).

En el libro electrónico también se observan importantes diferencias de tratamiento fiscal, no sólo entre países de la UE sino en comparación con su versión en papel dentro de cada país (ver tabla 2) a pesar de haber existido intentos (Francia y Luxemburgo) infructuosos de igualar tipos para ambos soportes.

Las diferencias en las tasas impositivas entre libro en papel y electrónico son muy importantes. En el caso de Irlanda y el Reino Unido los libros están gravados a tasa cero, siguiendo lo indicado en el Título VIII, Capítulo 4 de la *Directiva IVA 2006/112 / CE*. Precisamente el estado en el que encontramos la mayor diferencia es Irlanda, donde los libros impresos tienen una tasa de impuesto cero, mientras que los libros electrónicos pagan un 23% de impuestos. También vale la pena señalar el caso de Luxemburgo, donde ambos tipos de libros están sometidos al 3%. En general, en todos los Estados miembros existe una diferencia importante entre la imposición de los libros electrónicos y los libros tradicionales, salvo en Dinamarca donde ambos formatos están sujetos al 25%.

De los 28 Estados miembros de la UE analizados, 26 permiten tipos de IVA reducidos o nulos en libros impresos, excepto Bulgaria y Dinamarca, de conformidad con la *Directiva del IVA*, en reconocimiento del valor cultural, social y económico del libro y la lectura. Sin embargo, persisten diferencias en el tratamiento fiscal de los libros con respecto a su formato (impreso o electrónico) y los medios de entrega (online o fuera de línea).

Aplicar una política fiscal diferenciada para ambos formatos sin pensar en las consecuencias para el desarrollo digital de las empresas editoriales o en el impacto sobre los patrones de consumo, quizá resulte poco prudente (**Gesko**, 2013). El libro electrónico y el libro físico son dos productos que difieren sólo en su método de producción y distribución (**Guillon**; **Thierry**, 2013), por lo que aplicarles diferentes tipos impositivos violaría el principio de neutralidad de las tasas del IVA (**Barraud**, 2013). Este argumento se apoya en el principio de neutralidad fiscal, creado por el *Tribunal de Justicia* de la UE, según el cual una medida fiscal no debe falsear la competencia por el mismo producto en el mismo mercado, independientemente del medio o forma de distribución (**Respingue**-**Perrin**, 2013).

A finales de 2011 el *Parlamento Europeo* en una resolución sobre la modernización del IVA recomendó la reducción de las tasas de los productos culturales online evitando la discriminación de tipos según el soporte. Sin embargo, la *Comisión* emitió una comunicación en la que no apoyaba el uso de tipos reducidos de IVA en los bienes culturales online, como es el caso de los libros electrónicos. La *CE* sostuvo que los libros electrónicos y los libros impresos no son sustitutos perfectos y por lo tanto justificó su resistencia a que se apliquen tarifas reducidas a los libros electrónicos. Desde enero de 2012 Francia y Luxemburgo comenzaron a aplicar tipos de IVA reducidos a todo tipo de libros, independientemente del soporte. En julio de 2012 la *CE* inició un procedimiento de infracción contra ellos y en 2013 se remitió al *Tribunal de Justicia* de la UE para resolver el problema. El *TJUE* falló contra los dos Estados miembros el 5 de marzo de 2015. Mientras tanto, a partir de enero de 2015 Italia comenzó a aplicar también tipos reducidos de IVA a los libros electrónicos.

En julio de 2016 la *Comisión* lanzó una consulta pública sobre tipos de IVA reducidos para las publicaciones suministradas electrónicamente, que se clausuró en septiembre, en preparación de la propuesta. El 1 de diciembre de 2016 la *Comisión* publicó su propuesta. Su objetivo es permitir que los Estados miembros apliquen los mismos tipos de IVA reducido o superreducido que aplican a los libros impresos, también a todos los demás tipos de libros. El 7 de marzo de 2017 el *TJCE*

dictaminó que la discriminación en el IVA estaba justificada en el marco legislativo actual. En este contexto concluye que la exclusión de la aplicación de un tipo reducido de IVA al suministro de libros digitales por vía electrónica es consecuencia del régimen especial de IVA aplicable al comercio electrónico. Es decir, que los servicios prestados por vía electrónica no pueden tributar con tipos reducidos de IVA.

Aunque la idea de reducir el IVA de los libros electrónicos recibió el visto bueno del *Parlamento*, la reducción debe ser aprobada por los países de la UE por unanimidad. Sin embargo no se puede por el bloqueo que ejerce la República Checa contra esta medida. La normativa permitiría a los países, una vez que sea aprobada, aplicar un IVA reducido a ebooks y demás publicaciones digitales.

5. Conclusiones

Del recorrido efectuado por este análisis sobre la fiscalidad de los libros en la UE se extraen varias conclusiones. La primera es sin duda la necesidad de un tratamiento fiscal armonizado dentro de la UE que promueva el desarrollo de la industria cultural, columna vertebral de la sociedad del conocimiento. La segunda apunta a reducir las diferencias de trato fiscal entre el libro en papel y el libro electrónico.

Un mercado con el potencial de crecimiento del libro electrónico no puede verse desacelerado por una política fiscal que ponga límites a su expansión. Dicha expansión es positiva porque:

a) promueve el tránsito digital de las empresas editoriales;

b) difunde la cultura digital en la sociedad, elevando su alfabetización tecnológica; y

c) potencia la inversión en I+D+i dentro y fuera del propio sector editorial.

La inversión pública en el hábito lector queda muy lejos de estas otras aspiraciones colaterales del sector. El hábito de leer es una condición necesaria pero no suficiente para consolidar una industria que pasa por unos momentos delicados como consecuencia del estancamiento de la facturación del libro en papel y su desubicación ante la disrupción tecnológica proveniente de sectores ajenos al editorial.

Del análisis realizado se observa por un lado que es posible identificar una correlación clara entre las medidas públicas de fomento de la edición y el sistema de fijación de precios en la industria editorial. Por otro lado se aprecia que las ayudas públicas directas se orientan claramente más a la producción en papel que al libro electrónico, mostrando un mayor compromiso con la industria editorial por parte de aquellos Estados que discriminan fiscalmente en positivo al libro en papel, beneficiando así a toda la cadena de distribución.

De la experiencia internacional recogida en los estudios empíricos, se puede observar la sensibilidad de las ventas de ejemplares ante variaciones del IVA y la insensibilidad del hábito de lectura ante similares cambios (**Towse**, 2010; **Colbjørnsen**, 2015; **Borowiecki**; **Navarrete**, 2018). En consecuencia, una reducción fiscal en un contexto de precio fijo no sólo podría resultar un factor positivo para las empresas implicadas en la industria del libro (en términos de facturación), sino que el consumidor final, independientemente de su hábito de lectura, también se vería plenamente favorecido por acceder a contenidos –digitales o en papel- de un modo más económico.

Finalmente, el peso que sobre la recaudación total del IVA tendría una reducción del mismo (incluso a tipo cero) sobre un sector que apenas supone un 1,1% del PIB, no implicaría una merma significativa sobre las arcas del Estado. Tampoco la pérdida recaudatoria y su posible impacto negativo sobre la financiación de los servicios públicos sería un argumento razonable para seguir manteniendo el 21% de IVA en el libro electrónico y el 4% en el libro en papel.

Sería por tanto necesaria una reflexión sosegada sobre el sector editorial, su valor estratégico como espina dorsal del desarrollo cultural y cimiento de esa sociedad del conocimiento en la que la UE debería posicionarse globalmente como líder.

6. Referencias

Ahearne, Jeremy (2009). "Cultural policy explicit and implicit: A distinction and some uses". *International journal of cultural policy*, v. 15, n. 2, pp. 141-153.
http://wrap.warwick.ac.uk/36674
https://doi.org/10.1080/10286630902746245

Álamo-Cerrillo, Raquel; **Lagos-Rodríguez, María-Gabriela** (2016). "Tax implications of selling electronic books in the European Union". *Journal of theoretical and applied electronic commerce research*, v. 11, n. 2, pp. 28-40.
https://doi.org/10.4067/S0718-18762016000200004

Alino, Nelson U.; **Schneider, Gary P.** (2012). "Consumption taxes on digital products in the European Union". *Journal of legal, ethical and regulatory issues*, v. 15, n. 2, pp. 1-8.
https://www.abacademies.org/articles/jlerivol15no22012.pdf

Allen-Robertson, James (2013). *Digital culture industry: a history of digital distribution.* UK: Palgrave Macmillan. ISBN: 978 1 137 03347 5

Appelman, Marja (2003). "Fixed book price". In: Towse, R. *A handbook of cultural economics*. Cheltenham, UK: Edward Elgar Publishing Limited, pp. 237-242.
https://core.ac.uk/download/pdf/18507444.pdf

Barraud, Boris (2013). "Le régime fiscal du livre numérique". *Legicom*, v. 3, pp. 61-71.
https://www.cairn.info/revue-legicom-2013-3-p-61.htm

Benhamou, Françoise (2015). "Fair use and fair competition for digitized cultural goods: The case of ebooks". *Journal of cultural economics*, v. 39, n. 2, pp. 123-131.
https://doi.org/10.1007/s10824-015-9241-x

Borowiecki, Karol; **Navarrete, Trilce** (2018). "Fiscal and economic aspects of book consumption in the European Union". *Journal of cultural economics*, v. 42, n. 2, pp. 309-339.
https://doi.org/10.1007/s10824-017-9307-z

Boyle, James (2008). *The public domain: enclosing the commons of the mind.* New Haven, CT: Yale University Press. ISBN: 978 0 300 13740 8
http://thepublicdomain.org/thepublicdomain1.pdf

Canoy, Marcel; **Van-Ours, Jan**; **Van-der-Ploeg, Frederick** (2006). "The economics of books". *Handbook of the economics of art and culture*, v. 1, pp. 721-761.
https://doi.org/10.1016/S1574-0676(06)01021-0

Carreiro, Erin (2010). "Electronic books: How digital devices and supplementary new technologies are changing the face of the publishing industry". *Publishing research quarterly*, v. 26, pp. 219-235.
https://doi.org/10.1007/s12109-010-9178-z

Colbjørnsen, Terje (2015). "What is the VAT? The policies and practices of value added tax on ebooks in Europe". *International journal of cultural policy*, v. 21, n. 3, pp. 326-343.
https://dx.doi.org/10.1080/10286632.2014.904298

Conner, Kathleen; **Rumelt, Richard** (1991). Software piracy: an analysis of protection strategies. *Management science*, v. 37, n. 2, pp. 125-139.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.611.9166&rep=rep1&type=pdf
https://doi.org/10.1287/mnsc.37.2.125

Cordón-García, José-Antonio; **Gómez-Díaz, Raquel**; **Alonso-Arévalo, Julio** (2011). "Libros electrónicos: oferta comercial y redes". *El profesional de la información*, v. 20, n. 2, pp. 149-158.
https://doi.org/10.3145/epi.2011.mar.04

Czerwinski, Steve; **Fromm, Richard**; **Hodes, Todd** (2007). *Digital music distribution and audio watermarking.* Berkeley: Computer Science Division, University of California.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.1330&rep=rep1&type=pdf

Dent, Valeda (2007). "Local economic development in Uganda and the connection to rural community libraries and literacy". *New library world*, v. 108, n. 5-6, pp. 203-217.
https://rucore.libraries.rutgers.edu/rutgers-lib/21707/PDF/1/play
https://doi.org/10.1108/03074800710748777

Donoughue, Peter (2010). "Beyond the fear of cannibalisation will the book publishing industry survive the digital revolution?". *Logos*, v. 21, n. 3, pp. 167-178.
https://doi.org/10.1163/095796511X560060

European Commission (2018). *VAT rates applied in the member states of the European Union.* European Commission.
http://ec.europa.eu/taxation_customs/resources/documents/taxation/vat/how_vat_works/rates/vat_rates_en.pdf

Gesko, Martin (2013). "Factors influencing the microeconomic and fiscal effects of the reduced VAT on books". In: *European financial systems 2013. Proceedings of the 10th international scientific conference*. Brno: Masaryk University, pp. 99-106. ISBN: 978 80 210 6294 8
http://is.muni.cz/do/econ/soubory/konference/efs/Sbornik_2013.pdf

Guillon, Olivia; **Thierry, Clémence** (2013). "Is ebook pricing structured to mirror paper book prices? The cases of France and the USA in 2011". *Canadian journal of information and library science*, v. 37, n. 3, pp. 207-224.
https://doi.org/10.1353/ils.2013.0017

Hesmondhalgh, David (2007). *The cultural industries.* London: Sage. ISBN: 978 1 412908085

Hibbert, Bill (1999). "Publishing and the media industries in the digital age". *Journal of policy, regulation and strategy for telecommunications, information and media*, v. 1, n. 5, pp. 393-403.
https://doi.org/10.1108/14636699910801151

Hughes-Hassell, Sandra; **Rodge, Pradnya** (2007). "The leisure reading habits of urban adolescents". *Journal of adolescent y adult literacy*, v. 51, n. 1, pp. 22-33.
https://doi.org/10.1598/JAAL.51.1.3

Jain, Sanjay (2008). "Digital piracy: a competitive analysis". *Marketing science*, v. 27, n. 4, pp. 610-626.
http://doi.org/10.1287/mksc.1070.0313

Jiang, Zhengrui; **Sarkar, Sumit** (2003). "Free software offer and software diffusion: The monopolist case". In: *Procs of the 24th Intl conf on information systems.* Seattle, WA.
https://works.bepress.com/zhengrui-jiang/4

Krashen, Stephen; **Lee, Syying**; **McQuillan, Jeff** (2012). "Is the library important? Multivariate studies at the national and international level". *Journal of language and literacy education*, v. 8, n. 1, pp. 26-38.
http://jolle.coe.uga.edu/wp-content/uploads/2012/06/Is-the-Library-Important.pdf

Lichtenberg, James (2011). "In from the edge: The progressive evolution of publishing in the age of digital abundance". *Publishing research quarterly*, v. 27, n. 2, pp. 101-112.
https://doi.org/10.1007/s12109-011-9212-9

Lijphart, Arend (1971). *Comparative politics and the comparative method*. *American political science review*, v. 65, n. 3, pp. 682-693.
https://doi.org/10.2307/1955513

Løyland, Knut; **Ringstad, Vidar** (2012). "Fixed or free book prices: Is a hybrid system superior?". *International journal of cultural policy*, v. 18, n. 2, pp. 238-254.
https://doi.org/10.1080/10286632.2011.573850

Magadán-Díaz, Marta (2017). *Cambio tecnológico e innovación en el sector editorial español: efectos organizativos.* Oviedo: Tesis doctoral, Universidad de Oviedo.

Magadán-Díaz, Marta; **Rivas-García, Jesús** (2018a). "Digitization and business models in the Spanish publishing industry". *Publishing research quarterly*, v. 34, n. 3, pp. 333-346.
https://doi.org/10.1007/s12109-018-9593-0

Magadán-Díaz, Marta; **Rivas-García, Jesús** (2018b). "Variables conductoras de la industria editorial española". E*l profesional de la información*, v. 27, n. 6, pp. 1335-1345.
https://doi.org/10.3145/epi.2018.nov.16

Martin, Bill; **Tian, Xuemei** (2016). *Books, bytes and business: the promise of digital publishing.* New York: Routledge. ISBN: 978 0 754678373

Mitchell, Richard (2003). "Nordic and European cultural policies". In: Duelund, Peter. *The Nordic cultural model*. Copenhagen: Nordic Cultural Institute, pp. 437-478. ISBN: 978 87 98622550

Mol, Suzanne; **Bus, Adriana** (2011). "To read or not to read: A meta-analysis of print exposure from infancy to early adulthood". *Psychological bulletin*, v. 137, n. 2, pp. 267-296.
https://goo.gl/spDUpx
https://doi.org/10.1037/a0021890

Neuman, Susan; **Moland, Naomi** (2016). "Book deserts: The consequences of income segregation on children's access to print". *Urban education*, v. 1, pp. 1-22.
http://detroitlittlelibraries.org/wp-content/uploads/2016/07/BookDesertsStudy.pdf
https://doi.org/10.1177/0042085916654525

Perona, Mathieu; **Pouyet, Jérôme** (2010). *Le prix unique du livre à l'heure du numérique*. Paris: Éditions Rue d'Ulm. ISBN: 978 2 7288 0439 9
http://www.cepremap.fr/depot/opus/OPUS19.pdf

Poort, Joost; **Van-Eijk, Nico** (2017). "Digital fixation: The law and economics of a fixed e-book price. *International journal of cultural policy*, v. 23, n. 4, pp. 464-481.
https://doi.org/10.1080/10286632.2015.1061516

Ragin, Charles (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press. ISBN: 978 0 520280038

Ragin, Charles (2006). "Set relations in social research: Evaluating their consistency and coverage". *Political analysis*, v. 14, n. 3, pp. 291-310.
https://doi.org/10.1093/pan/mpj019

Respingue-Perrin, Sébastien (2013). "Too early, too fast? The regulation of the ebook market in France and its possible

effects on EU libraries". *Liber quarterly*, v. 23, n. 2, pp. 81-109.
https://doi.org/10.18352/lq.8539

Ringstad, Vidar (2004). "On the cultural blessings of fixed book prices: facts or fiction?". *International journal of cultural policy*, v. 10, n. 3, pp. 351-365.
https://doi.org/10.1080/1028663042000312570

Sanz-Sanz, José-Félix; **Romero-Jordán, Desiderio**; **Prieto-Rodríguez, Juan** (2003). "¿Qué puede esperarse de una reducción de la imposición indirecta que recae sobre el consumo cultural?: un análisis a partir de las técnicas de microsimulación". *Papeles de trabajo del Instituto de Estudios Fiscales. Serie Economía*, v. 14, pp. 7-27.
https://dialnet.unirioja.es/servlet/articulo?codigo=650704

Smith, Dennis (1991). *The rise of historical sociology*. Philadelphia: Temple University Press. ISBN: 978 0 877229209

Stevenson, David; **Balling, Gitte**; **Kann-Rasmussen, Nanna** (2017). "Cultural participation in Europe: Shared problem or shared problematisation?". *International journal of cultural policy*, v. 23, n. 1, pp. 89-106.
https://doi.org/10.1080/10286632.2015.1043290

Terzi, Nuray (2011). "The impact of e-commerce on international trade and employment". *Procedia-social and behavioral sciences*, v. 24, pp. 745-753.
https://doi.org/10.1016/j.sbspro.2011.09.010

Tian, Xuemei; **Martin, Bill** (2010). "Digital technologies for book publishing". *Publishing research quarterly*, v. 26, n. 3, pp. 151-167.
https://doi.org/10.1007/s12109-010-9164-5

Towse, Ruht (2010). *A textbook of cultural economics.* Cambridge University Press. ISBN: 978 0 521717021

Van-der-Ploeg, Frederick (2004). "Beyond the dogma of the fixed book price agreement". *Journal of cultural economics*, v. 28, n. 1, pp. 1-20.
https://goo.gl/qhahy4
https://doi.org/10.1023/B:JCEC.0000009817.61107.f3

Van-Eijk, Nico; **Poort, Joost**; **Rutten, Paul** (2010). "Legal, economic and cultural aspects of file sharing". *Communication and strategies*, v. 1, pp. 35-54.
https://ssrn.com/abstract=1609847

Waldfogel, Joel (2017). "How digitization has created a golden age of music, movies, books, and television". *Journal of economic perspectives*, v. 31, n. 3, pp. 195-214.
https://doi.org/10.1257/jep.31.3.195

Zhang, Dan; **Zhu, Pengyu**; **Ye, Yanmei** (2016). "The effects of e-commerce on the demand for commercial real estate". *Cities*, v. 51, pp. 106-120.
https://doi.org/10.1016/j.cities.2015.11.012

Sci-Hub, a challenge for academic and research libraries

Llarina González-Solar; Viviana Fernández-Marcial

How to cite this article:

González-Solar, Llarina; **Fernández-Marcial, Viviana** (2019). "*Sci-Hub*, a challenge for academic and research libraries". *El profesional de la información*, v. 28, n. 1, e280112.

https://doi.org/10.3145/epi.2019.ene.12

Article received on 26-09-2018
Approved on 24-11-2018

Llarina González-Solar ✉
https://orcid.org/0000-0003-4443-4102

Universidade da Coruña (Spain)
Grupo de Investigación Historia, Arqueología, Documentación y Cultura (*Hadoc*)
llarina.gonzalez@udc.es

Viviana Fernández-Marcial
https://orcid.org/0000-0002-9277-266X

Universidade da Coruña (Spain)
Facultad de Humanidades y Documentación
viviana.fernandez@udc.es

Abstract

Sci-Hub emerged into the field of scientific communication in 2011 as a platform for free access to scientific papers. It is the most popular of the so-called shadow libraries, systems that overcome the limits of legal access to scientific publications, standing apart from the open access movement. Besides from the media coverage that has served to boost its popularity, several studies reveal the impact of *Sci-Hub* among researchers, who have embraced this initiative. *Sci-Hub* has revealed new forms of access to scientific information, affecting academic and research libraries that cannot remain on the sidelines. This study addresses the *Sci-Hub* phenomenon and its implications for academic and research libraries from different points of view, through a bibliographic review and an analysis of examples of action.

Keywords

Academic libraries; University libraries; Shadow libraries; Information behaviour; *Sci-Hub*; Scientific communication; Ethics; Legality; Legal issues; Information access; Open access; Free information; Information usage habits.

1. Introduction

In recent years, access to scientific publications has become a battleground, with recurrent reports of scientific and academic entities announcing the cancellation or modification of subscriptions to important publishers. The fundamental cause of this situation can be found in difficulties faced by libraries in meeting ever-growing subscription costs (**Piwowar** *et al.*, 2018).

Access to scientific content is uneven. Thus, only a small proportion of the world's population has access to most publications, with small institutions in developing countries in the least favourable position (**Kirsop**; **Chan**, 2005; **Meadows**, 2015; **Bendezú-Quispe** *et al.*, 2016). The open access (OA) movement was born precisely to guarantee access to scientific publications for all, including those who could not afford to pay for them. After years of work, significant advances have been achieved, but the most recent calculations estimate that only a quarter of scientific literature is free of paywalls, in open access (**Khabsa**; **Giles**, 2014; **Piwowar** *et al.*, 2018; **Bosman**; **Kramer**, 2018), although the proportion is higher when referring to publications in recent years (**Himmelstein** *et al.*, 2018; **Piwowar** *et al.*, 2018).

It is in this context that pirate websites such as *Aaaaarg* and *Library Genesis* appear, ignoring or circumventing intellectual property restrictions (**Lawson**, 2017). *Sci-Hub* is the best known of all these shadow libraries (**Gardner**; **McLaughlin**; **Asher**, 2017) and probably the most widely used, not only because of the volume of publications to which it gives access, but also because it has received unprecedented attention from both the mass media and academia.

Academic and research libraries are the traditional providers of scientific information to users which once went, in person, to consult journals (**Sathe**; **Grady**; **Giuse**, 2002). Today, most have chosen to access electronic publications and users have moved to this system (**Tenopir** *et al.*, 2009). Over the past decades the price of subscriptions has been continuously increasing (**Larivière**; **Haustein**; **Mongeon**, 2015), and the contract models imposed since the 1990s, the big deals, have proved to be excessively rigid and costly in the long term (*Sparc*, 2018). As a result, research libraries are in a complex position in relation to pirate websites, as the services with which they compete, such as interlibrary loan, have greatly improved, but they cannot provide the level of immediacy of the former. Moreover, experience shows that libraries continue to be where users turn to in order to seek advice and guidance on sources of information, as demonstrated by the development of research support services in many universities (**Fernández-Marcial**; **Costa**; **González-Solar**, 2016), and they must be prepared to face information queries on these illegal systems.

“Since *Sci-Hub* has entered the panorama of access to scientific information, it has generated changes in the consumption of scientific information and has been a turning point for libraries”

Since *Sci-Hub* has entered the panorama of access to scientific information, it has generated changes in the consumption of scientific information and has been a turning point for libraries. **Crissinger** (2017) considers that there is evidence that an analysis of *Sci-Hub*, from an ethical, technological, intellectual property and inequality point of view, is now a mature topic for debate. The objective of this paper is to describe the various facets of *Sci-Hub* in order to give as complete a picture as possible of this phenomenon and to assess its impact on university libraries.

2. *Sci-Hub*. The new *Napster*?

It is not uncommon to find comparisons between *Sci-Hub*[1] and the P2P phenomenon that radically changed the music industry at the beginning of the 21st century and which, beyond its own history of growth, popularity and decline, helped to shape the new path that record companies are following today with streaming services as major protagonists. A brief analysis of the history and characteristics of *Sci-Hub* allows us to understand if its possible final effect on the publishing system will have any parallelism with *Napster*'s.

Sci-Hub's story has its own mythology. Neuroscientist Alexandra Elbakyan, at the tender age of 22, created *Sci-Hub*, which was released on September 5, 2011. This was her response to difficulties in accessing scientific publications after her return to Kazakhstan. At that time, the options available to circumvent the paywalls were to write directly to the authors of the publication or to request the document via *Twitter* with the hashtag #IcanhazPDF, in the hope that someone with access could send a copy (**Bonhanon**, 2016a; 2016b; **Himmelstein** *et al.*, 2018).

Its implementation has been growing over the past seven years, as has the media attention it has received, with Elbakyan being named among “*Nature's 10*: Ten people who mattered this year” in December 2016 as milestone. Both legal battles and articles written about *Sci-Hub* or Elbakyan (and their consideration among “*Nature's 10*”) make them trending on *Google* searches (**Himmelstein** *et al.*, 2018). In other words, every time *Sci-Hub* and/or its creator are the object of media attention, even if it is from a critical position or a company demands it, a significant amount of free publicity is generated.

But what is *Sci-Hub*, and how does it work? To start with, it can be classified in a number of ways: as an online search engine, a web page, or as a collection of academic articles. **Elbakyan** (2017) describes it as follows:

> “The core of *Sci-Hub* is a script that downloads html and pdf pages from the Web. In that sense, *Sci-Hub* is technically more similar to a web scraper [...] *Sci-Hub* technically is by itself a repository, or a library if you like, and not a search engine for some other repository. But of course, the most important part in *Sci-Hub* is not a repository, but the script that can download papers closed behind paywalls”.

The creator of *Sci-Hub* does not offer explanations on how it obtains the papers hosted in the repository. **Bonhanon** (2016b) indicates that this harvesting of documents involves collecting user data to overcome payment barriers; access credentials can be obtained by voluntary contributions from researchers or through unethical methods, such as phishing emails, as stated by publishers. **Elbakyan** (2018) does not reveal the origin of the accounts it uses but indicates that they come from various sources, including some illicit ones.

In order to maintain access to this information, it uses a succession of mirror sites, as do other pirate information systems such as *The Pirate Bay* (**Penn**, 2018) using a decentralized scheme that is difficult to control and has a wide international scope. It has a bot for *Telegram* users (@scihubot) although the most frequent way to access *Sci-Hub* is through its web page which, due to lawsuits, has been deftly and repeatedly changing its domain name, and has been hosted in various countries (Laos, Hong Kong, Mongolia, Taiwan, Russia..., among others) and in *Tor* (*The Onion Router*).

Over the years multiple attempts have been made to quantify the number of documents collected by *Sci-Hub* and their coverage. It is estimated that it contains 69% of the scientific articles with DOI, of which 54.5 million articles are for payment and account for 85.1% of the total number of articles behind paywall (**Himmelstein** *et al.*, 2018). These data correspond to the documents that have already been downloaded to its repository, not to those that could be accessed on demand whose

number would be substantially higher. **Greshake** (2017) estimates that 95% of these publications are subsequent to 1950 despite having some as old as the 1619 edition of the book *Descriptio cometæ*. Its size would represent a real threat to the current model of scientific publication and, in particular, to payment by subscription (**McKenzie**, 2017).

In terms of coverage, **Houle** (2017) detects variances between disciplines, with less presence of law, music and some areas of economics; and **Himmelstein** *et al.* (2018) highlight the strong presence of chemistry and health sciences. Most of the documents contained are journal articles for which it represents 77.8% of what is present in *Crossref*. This high coverage rate is also shown in contributions to congresses (79.7%) and is very limited in other types of documents such as book chapters (14.2%) or standards (1.5%) (**Himmelstein** *et al.*, 2018).

3. Who uses *Sci-Hub* and how?

The above data demonstrate the magnitude of *Sci-Hub* as an information resource. The key question then arises, who uses it? Bearing in mind that its objective would be to facilitate access to scientific literature for those who cannot afford it because of its high cost.

Regarding the intensity of *Sci-Hub* use, **Van-Noorden** (2016) reports a number of downloads exceeding 75 million in 2016 (surpassing 42 million in 2015), which represents 3% of all downloads of this type of publication worldwide. Documents after 1985 account for 95% of downloads, and 35% are less than two years old at the time of download (**Greshake**, 2017). Their growth appears to be soaring, reaching 88% per year (**Himmelstein** *et al.*, 2018).

As to *Sci-Hub* utilization data (2015-2016), **Bonhanon** (2016b) estimates, not without surprise, that a quarter of the requests for articles come from rich countries, presumably without problems of access from their institutions,

> "in fact, some of the most intense use of *Sci-Hub* appears to be happening on the campuses of U.S. and European universities" (p. 510).

However, his data show that the countries with the highest volume of downloads to be, in the following order: Iran, China, India and Russia. **Greshake** (2016) detects an intensive use in Portugal, Iran, Tunisia and Greece; **Himmelstein** *et al.* (2018) affirm that preferential use is occurring in countries with lower access capacity through institutional channels.

The reason for these variations and the impact on specific campuses could be related to the growth of text mining techniques, which require a large volume of text to work with and which, unlike automatic downloads at *Sci-Hub*, are very costly to obtain through the usual services in terms of time spent. Some countries such as Iran (where national legislation permits) appear to be making national mirrors (**Bonhanon**, 2016b).

In early 2018, *Sci-Hub* posted full data for 2017[2] on *Twitter*. An analysis of these using the *OpenRefine* tool allows us to affirm that many of these trends are maintained today. It is confirmed that among the countries with the highest downloads we have emerging powers such as China, India or Brazil, atypical cases such as Iran and also countries with highly developed economies and research structures such as the USA or France (Figure 1). These data should be studied in the future taking into account correction and weighting factors such as the number of researchers, research centres, investment in science, etc.

Raw usage data for the months of September 2015 to February 2016[3] show that Spain accounted for an average of 1.5% of total *Sci-Hub* downloads worldwide, with some concentration of IP in the main research poles. The data corresponding to 2017 show a certain increase in downloads, so that Spain represents 2.04% (a total of 3,089,349), occupying position 11 in this peculiar ranking. There is some correspondence between the areas with the highest downloads in the

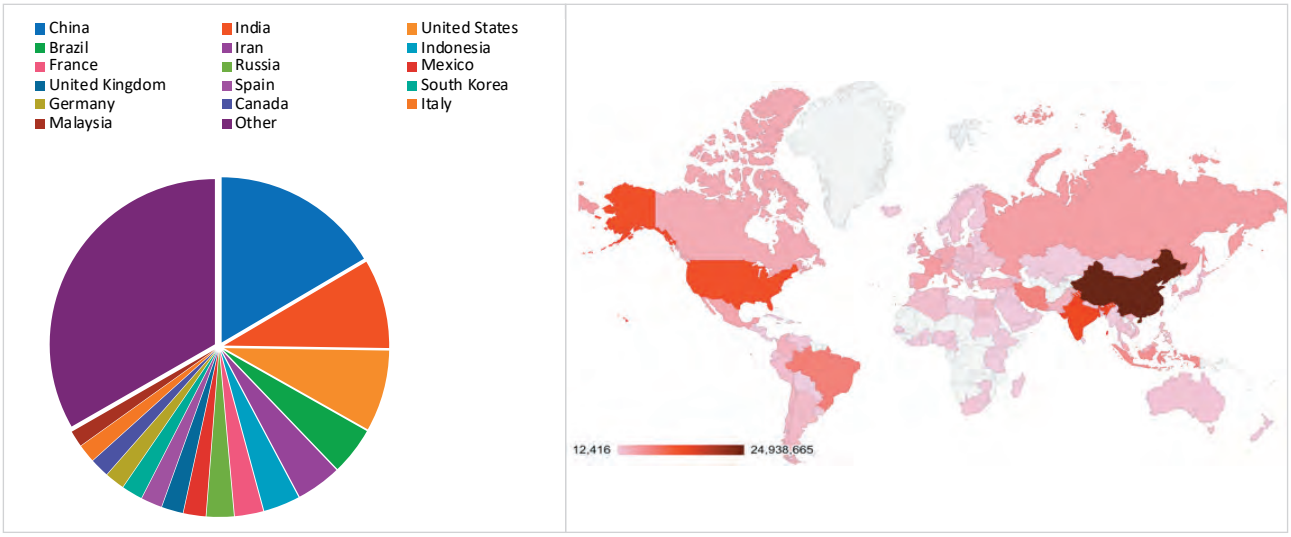


Figure 1. Distribution by country of *Sci-Hub* downloads and details of countries with more than 1.5% of the total

data for 2015-2016 and those for 2017 (Table 1), with large cities with consolidated university campuses having the highest incidence, and Madrid having nearly twice as many downloads by 2017 as Barcelona. The articles with the highest number of downloads in Spain show a preference for the areas of health sciences, sciences and engineering and with a greater presence of articles subsequent to the year 2000.

Faced with this ubiquity of downloads, it is worth asking the reasons for them. In those countries where access is limited due to the scarcity of economic resources, the answer is simple: in order to do science, access to scientific contents is needed and this is the only possible way for articles that are not in open access, avoiding any payment.

Table 1. Spanish cities with the highest number of *Sci-Hub* downloads

Position	2015-2106[4]		2017	
	City	Downloads	City	Downloads
1	Madrid	98.143	Madrid	868.322
2	Barcelona	78.535	Barcelona	488.101
3	Valencia	26.634	Valencia	215.690
4	Bilbao	12.622	Sevilla	72.613
5	Zaragoza	10.795	Bilbao	67.899
6	Santander	10.377	Zaragoza	61.313
7	Murcia	8.819	A Coruña	60.267
8	A Coruña	8.360	Murcia	54.065
9	Sevilla	7.432	Valladolid	40.853
10	Oviedo	7.406	Alacant	40.040

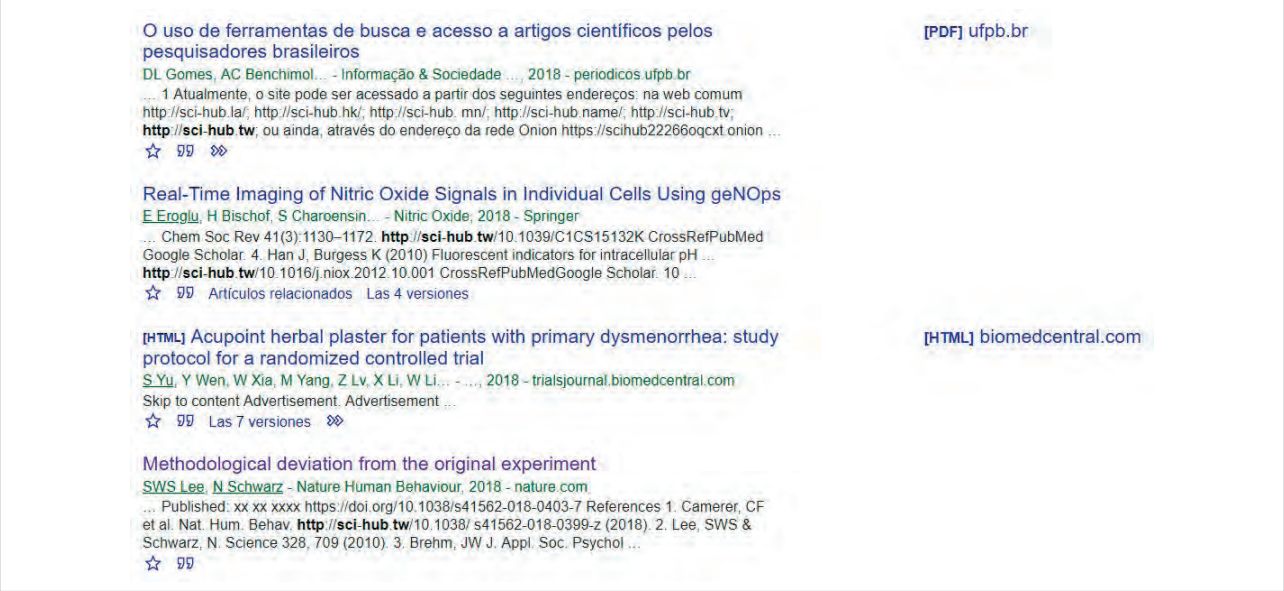


Figure 2. *Google Scholar* search that exemplifies the lack of knowledge about *Sci-Hub*, used when referencing bibliographic citations. Source: *Google Scholar*

This justification is, however, insufficient as it cannot be applied to the United States or Europe. A survey published in *Science* attempts to discover the underlying reasons: researchers claim lack of access to documents (50%), disagreement with editing and publication models (23%) and convenience (17%) (**Travis**, 2016). For countries with good access **Bonhanon** (2016b) focuses on two possible reasons: confusion about what *Sci-Hub* is (Figure 2) and convenience.

4. The debate of morality

Elbakyan has justified the creation and maintenance of *Sci-Hub* as illegal but ethical, arguing that the concept of property and intellectual property are unnecessary, communism or theft as a fundamental element for the advancement of science... and comparing its creation with the acts of characters such as Robin Hood or Hermes (Figure 3).

Figure 3. Example of Elbakyan's (2016) discourse developed in her presentations.

In this debate, which **Triggle** and **Triggle** (2016, p. 6) summarize by asking "Is *Sci-Hub* altruism or copyright theft?" One can see the opinions of different groups, which must be confronted. Obviously on one side are the publishers and, at the opposite extreme, Elbakyan. In addition, the opinion of researchers should be valued. They use *Sci-Hub* extensively (60%) and think that using it is not incorrect (88%), according to the *Science* survey (**Travis**, 2016). Among people who study SciComm or scientific publication as a process, the debate has been very intense and has spread in recent years over news, forums, blogs and scientific literature (**Machin-Mastromatteo**; **Uribe-Tirado**; **Romero-Ortiz**, 2016; **McNutt**, 2016; **Priego**, 2016; **Woolston**, 2016).

The central argument of support for *Sci-Hub* lies in the inequality of the scientific publishing system and in considering that the fundamental objective of changing it and making it fairer requires any means. In this line of thought *Sci-Hub* may not be the solution but it is a wake-up call as to the need to build a common scientific and academic heritage, beyond the limits imposed by intellectual property (**Lawson**, 2017). *Sci-Hub* is a symptom, a reaction to a problem that has to be solved no matter what.

The signatories of the open letter "In solidarity with *Library Genesis* and *Sci-Hub*" are signified in the group of people with the greatest affinity for the platform (*Custodians Online Campaign*, 2015). The largest number of *Sci-Hub* supporters seem to be in the area of medicine, not only among researchers, but among physicians seeking better alternatives for their patients (**Bendezú-Quispe** *et al.*, 2016; **Faust**, 2016; **Triggle**; **Triggle**, 2016). This idea is clearly evidenced in Latin America with the positive perception of *Sci-Hub* among medical students (**Mejia** *et al.*, 2017) and high download volume data, to which **Machin-Mastromatteo** *et al.* (2016) propose to improve access to medical scientific literature especially through the promotion of open access.

Those who look at the phenomenon from a critical perspective point out that the platform infringes laws, being "copyright-breaking on a grand scale" (**Van-Noorden**, 2016), which is neither admissible nor justifiable, insisting that there are legal alternatives such as open access (**Greco**, 2017). **McNutt** (2016) emphasizes other collateral factors such as the loss of usage data and its implication in the management of subscriptions, the repercussions for publishing house workers, the impact on smaller publishing houses such as non-profit scientific societies... We can add to all these the possible and unpredictable consequences for authors.

The relationship between *Sci-Hub* and OA is controversial. *Sci-Hub* is generally not considered as part of the OA movement (**Piwowar** *et al.*, 2018) although in some cases it has been labeled as a subtype of it, following the ideas put forward by the "*Guerilla Open Access Manifesto*" (**Swartz**, 2008). **Björk** (2017) and **Penn** (2018) call it black open access (together with *ResearchGate*, *Academia.edu*, *Google*, the tag *#icanhazpdf*...) considering it one of the greatest challenges to the traditional model of academic publication.

Peter Suber warns that being confused with piracy can have a "strategic cost" for the OA movement (**Bonhanon**, 2016b) in terms of distorting its purpose and even replacing one paywall with another by reinforcing the golden route (**Novo**; **Onishi**, 2017). However, from the ranks of OA there has been explicit support for *Sci-Hub* (**Cochran**, 2016) and tweets are frequent in providing their new domains. Heather Piwowar gives it an instrumental role by scaring editors and pushing them to "do the right thing" and bet on open access (**Van-Noorden**, 2016). There is some consensus that *Sci-Hub* has generated debate and attracted media attention (**Emery**, 2016; **Van-Noorden**, 2016; **Novo**; **Onishi**, 2017; **Piwowar** *et al.*, 2018). **De-Castro** (2016) even compares Alexandra Elbakyan with the girl in Andersen's tale who pointed out that the emperor has no clothes.

Sci-Hub is attributed the quality of "disruptive" (**Bonhanon**, 2016b; **Emery**, 2016; **Machin-Mastromatteo** *et al.*, 2016; **Steel**, 2016; **Travis**, 2016; **Himmelstein** *et al.*, 2018; **Marple**, 2018, **Nicholas** *et al.*, 2018). However, for **Priego** (2016) the platform does not represent a substantial cultural change, since gratuity does not alter the basis of the current system of scientific communication. To make this assessment, the judgment of time will indicate whether or not the existence of *Sci-Hub* entails a real change in the model of information consumption and scientific communication.

It is necessary to raise the risks of dependence on a system that is managed by one person. Those who choose to rely on *Sci-Hub* for their access to science (for whatever reason) are subject to the whims of Elbakyan. A telling example occurred in September 2017 when she brought down the system in Russia for several personal offenses (**Travis**, 2017).

5. *Sci-Hub*: a turning point for libraries

The emergence of *Sci-Hub* has signified a turning point for academic and research libraries, representing a challenge for the information retrieval systems offered through them. *Sci-Hub*'s interface is simple and friendly, similar to *Google*. Users access the content of the document using only the DOI or text title. **Faust** (2016) relates his experience, highlighting the simplicity of a search for a papers noting that the

> *Sci-Hub* has been a challenge for academic and research libraries and for the information retrieval systems offered through them

> "'click burden' using *Sci-Hub* was substantially lower than going through my hospital's online library, and it saved me many seconds".

Oakely (2016), who compares a paper search in the *Georgetown University Library* and *Sci-Hub* (Figure 4), points in the same direction.

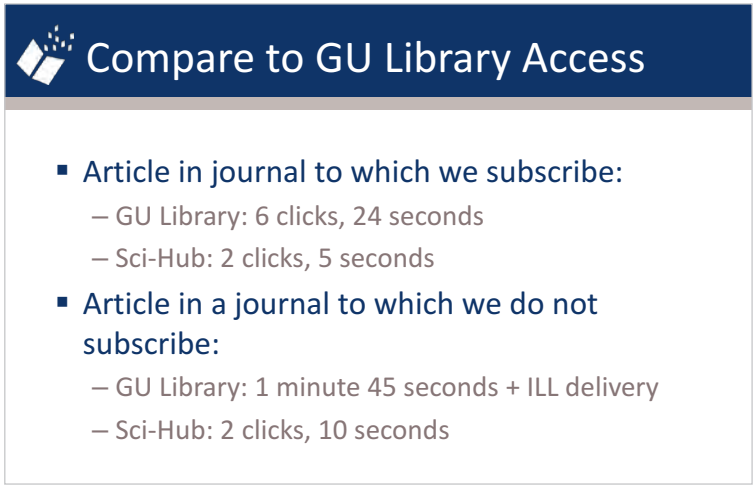


Figure 4. Comparison of **Oakley** (2016) between the process of locating an article in the *Georgetown University Library* and *Sci-Hub*.

This simplicity is one of its strengths against library models, so it could be emerging as the go-to resource even for journals acquired by libraries (**Bonhanon**, 2016b). Some publishers even point the finger at libraries for using unfriendly retrieval systems or for not adequately "educating" researchers in them (**Bonhanon**, 2016b).

Faust (2016) highlights the reliability of the recovery as an aspect in *Sci-Hub*'s favour:

> "*Sci-Hub*'s appeal does not rest on speed alone but rather its reliability. Some have observed that finding an article through a hospital or university library doesn't guarantee that the article will actually be available for download. Surely we have all encountered the dreaded "request access" hyperlink for a sought article found in a library's database, a phrase that might as well be synonymous with "fuhgeddaboudit." In contrast, when *Sci-Hub* finds an article, you're always 1 click away from the pdf file. The appeal quickly becomes clear. Alternatively, of course, there's always *Google*. By Googling my article, I found that I could rent it for $6 per day or buy it for $38. (p. 15A).

Furthermore, *Sci-Hub*'s existence affects the contracting of resources. **Steel** (2016) points out that libraries do not take decisions such as canceling their licenses due to the availability of articles in *Sci-Hub*; but it does appear that the existence of these systems, which are still resources that users are already using, is influencing the development of negotiations for the subscription of journals, specifically when it comes to big deals (**Esposito**, 2017). Other authors such as **Strielkowski** (2017) consider that the impact of *Sci-Hub* can lead to a change of business model based on more reasonable prices, although subscriptions would prevail. **Himmelstein** *et al.* (2018) study the evolution of contracting, from the hegemony of big deals to their gradual replacement by "à la carte" selections and the recent cancellations of important contracts with *Elsevier*. The authors observe that this publisher's lawyers already anticipated the possibility that *Sci-Hub* and *LibGen* would reach a level at which they could

> "serve as a functionally equivalent, although patently illegal, replacement for *ScienceDirect*" (p. 13).

The possibility has been raised that the existence of *Sci-Hub* or other academic publication exchange systems may be affecting the use of library services, specifically interlibrary loan (**Gardner**; **Gardner**, 2015), however this fact has yet to be demonstrated as there are no reliable data.

Cochran (2016) called into question, from the publishers' perspective, the role of libraries and universities as responsible for the illegitimate use of credentials, breaking licensing agreements and allowing illegal downloading of articles. Watermarks make it possible to know which university each article comes from, which is why he considers universities obliged to prevent this from happening and to reflect their position in codes of conduct. There is some debate about whether libraries can monitor patterns of credential use regularly enough to avoid these deviations, and even whether it may be contrary to their very nature to do so (**Ruff**, 2016; **Russell**; Sánchez, 2016; **Hoy**, 2017).

"The ethical and legal aspects surrounding Sci-Hub also place libraries in a new landscape. Librarians debate between defending access to scientific information and the legal implications of its use"

The ethical and legal aspects surrounding *Sci-Hub* also place libraries in a new landscape. Librarians debate between defending access to scientific information and the legal implications of its use. There is some latent support among librarians, although few would recommend it openly since, even though they may agree with its objectives or motivations, few would agree with the means it uses (**Ruff**, 2016; **Steel**, 2016; **Penn**, 2018). A good example of this is the diatribe between Gabriel J. Gardner (librarian at *California State University*) and Thomas H. Allen (president of the *Association of American Publishers*) about the possibility that the former would have recommended the use of *Sci-Hub* during a talk with librarians (**Reichman**, 2016).

6. Communication about *Sci-Hub* in libraries

Beyond the subjective aspects of support or disapproval of *Sci-Hub*, it is interesting to analyze the real way in which libraries communicate this phenomenon or not. There does not seem to be a homogeneous model of communication about shadow libraries. An exploration of Spanish academic libraries shows that this information is not common on the

websites themselves, although it can be found on the blogs they use as an informal communication system. It appears as one more piece of news of one more system, gathering the content of an article or report that refers to it or as a complement to information on topics such as open access, as we see, for example, in:

> There does not seem to be a homogeneous model of communication about shadow libraries

- Blog *Universo Abierto* of the *Universidad de Salamanca* (Figure 5): *https://universoabierto.org/2018/04/02/sci-hub-proporciona-acceso-a-mas-de-dos-tercios-de-todos-los-articulos-cientificos-publicados*
- Blog *Acceso abierto* of the *Universidad de las Palmas de Gran Canaria*: *http://bibwp.ulpgc.es/accesoabierto/2017/02/10/sci-hub-tiene-ya-en-torno-a-60-millones-de-articulos-cientificos-en-abierto*
- *BujaBlog* of the *Universidad de Jaén*: *https://blogs.ujaen.es/biblio/?tag=sci-hub*
- Blog *Digitum* of the *Universidad de Murcia*: *http://digitum-um.blogspot.com/2016/11/posibles-efectos-de-sci-hub-en-las.html*
- *Boletín DIB* of the *Universidad de Extremadura*: *https://deinteresparaelbibliotecario.wordpress.com*
- *Canal Biblog* of the *Universidad Autónoma de Madrid*: *http://canalbiblos.blogspot.com/2017/06/sci-hub-pierde-la-batalla-frente.html*
- Blog *Fonseca* of the *Universidade de Santiago de Compostela*: *https://busc.wordpress.com/tag/sci-hub*

Figure 5. The library blog of the *Universidad de Salamanca* collects information from an article on *Sci-Hub* coverage.

In the English-speaking university libraries, however, a greater presence of this information is detected. For instance, *Sci-Hub* appears in the subject guides of the libraries of the:

- *University of Wisconsin-Milwaukee* *http://guides.library.uwm.edu/scihub*

- *University of Delaware*
  https://guides.lib.udel.edu/sci-hub
- *Kansas State University*
  https://guides.lib.k-state.edu/c.php?g=645013&p=4520198
- *Washington University in St. Louis*
  https://library.wustl.edu/thoughts-sci-hub-easy-access
- *Bond University*
  https://library.bond.edu.au/news/48268/antoinette-cass-manager-scholarly-publications-and-copyright
- *University of Windsor*
  http://leddy.uwindsor.ca/sci-hub-problems-and-questions

And frequently asked questions such as in the *Australian Curtin University*:
https://answers.library.curtin.edu.au/faq/204046

Events such as the symposium "Online piracy: why *Sci-Hub* is disrupting scholarly publishing" of the *Georgetown University Library* in 2016[5] or the one entitled "*Sci-Hub* and *LibGen* in perspective" at the *University of Texas at Austin* library in early 2018[6] have also been raised.

Some of this information shows a clear position against *Sci-Hub*. For example, the *Kansas State University* library guide states

> "This page explains why librarians disagree with using resources accessible from *Sci-Hub*".

The *Bond University Library* avoids a clear positioning but reminds that using *Sci-Hub* or other similar sites is illegal and that its users must abide by the university regulations that also prohibit it.

In several cases, it is insisted from the beginning of the guide that users should not offer their credentials to *Sci-Hub*, since publishers can act by terminating contracts with the library (Figure 6), this information can be supplemented with how to act in case of stolen credentials. Likewise, these guides strive to focus attention on the alternatives offered by the library for obtaining academic information. Less commonly, some libraries such as *Cornell University*[7] report how to use *Sci-Hub* (Figure 6).

Some libraries have offered to the creator of *Sci-Hub* a space to communicate her points of view, for example in the framework of the *Open access symposium 2016* organized by the *University of North Texas Libraries* with a presentation titled "Why science is better with communism? The case of *Sci-Hub*" (**Elbakyan**, 2016) and available on *YouTube*;
https://youtu.be/hr7v5FF5c8M

Another example is her interven-

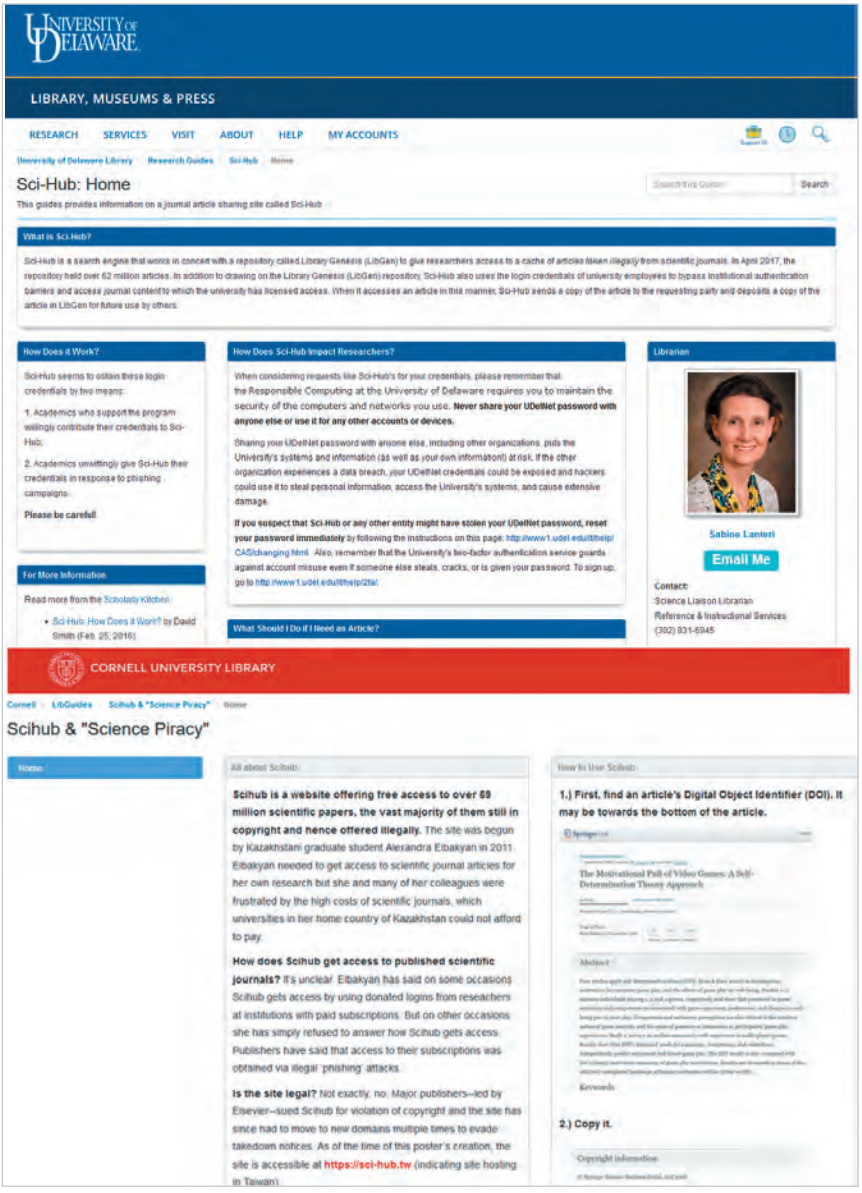


Figure 6. Two different approaches to *Sci-Hub* from libraries. In the image above the *University of Delaware Library Guide* warns of the consequences of donating credentials to *Sci-Hub* and how to act if they have been obtained without authorization. On the bottom, *Cornell University*'s guide explains the use of *Sci-Hub* without openly positioning for or against it.

tion in the *Workshop for the creation of a network of research libraries* held in 2017, a network that belongs to the *Ministerio Coordinador de Conocimiento y Talento Humano de Ecuador*.
https://youtu.be/w_jc7wNET1Q

7. Conclusions

If an anonymous survey or study were carried out among librarians in relation to *Sci-Hub*, what would be the widespread opinion about its impact and use in libraries? It is evident that the library is a service that, in this context, can have a dual and contradictory vision. On the one hand, it is faced with user demands that can be satisfied with this accessible and unlimited resource and, on the other, with the defense of current hiring models and respect for the ethical use of information, especially from the institutional perspective as a service of research institutions.

> One of the issues that *Sci-Hub* has raised is that users, or part of them, prioritize access to scientific content regardless of the legal or ethical connotations that this may imply

Beyond this acceptance or rejection, shadow libraries have reached the field of scientific communication to mark a before and an after, in an environment that had already been strongly altered by a search engine like *Google*. As evidence of the impact of the aforementioned search engine on academic activity, we must observe how this company has been creating and generating products that are clearly oriented towards the researcher as a target audience, with *Google Scholar* being its best exponent. Could the scientific community and information professionals glimpse that citations were going to have a key reference *Google Citations*?

One of the issues that *Sci-Hub* has raised is that users, or part of them, prioritize access to scientific content regardless of the legal or ethical connotations that this may imply. This is evident not only in access but also in sharing or disseminating research results. This is a significant change for academic and research libraries because, once again, they face new forms of competition.

The shadow libraries seem to have arrived, perhaps to stay, perhaps to definitively alter the context of scientific communication as *Napster* did with the distribution of music. *Sci-Hub* has been the object of this study due to its impact, but there are others, and in the future, new initiatives may be added and, at the moment, we do not know what these may lead to. Librarians can position themselves for or against shadow libraries, from a personal or institutional point of view, but beyond that the truth is that the informational behaviour of researchers can no longer be classified or delimited in a space. Users are the raison d'être of libraries. If they change then....

8. Notes

1. For example

- *The Washington Post* article "Russia is building a new Napster — but for academic research". *https://wapo.st/2utfi5O?tid=ss_tw-bottom&utm_term=.b206f3005e50*
- Rick Anderson post "Napster vs. Record Labels, Sci-Hub vs. Publishers" in the blog: *https://scholarlykitchen.sspnet.org/2018/01/03/napster-vs-record-labels-sci-hub-vs-publishers-part-1-parallels*
- Or the papers **Crissinger** (2017) and **Nicholas** *et al.* (2018)

2. Available at *https://t.co/qdp7oNu2ay*

3. Available at *Dryad*:
https://doi.org/10.5061/dryad.q447c

4. For this table it is necessary to remember that these data refer only to the period between September 2015 and February 2016, while those corresponding to 2017 refer to the whole year.

5. Whose session can be followed in:
https://youtu.be/Qc6PqKlpfkw

6. Information about this event can be found at:
https://lib.utexas.edu/events/87

7. This guide, hosted at:
https://guides.library.cornell.edu/c.php?g=862841&p=6185832
was active until mid-September 2018 when it was withdrawn.

9. References

Bendezú-Quispe, Guido; **Nieto-Gutiérrez, Wendy**; **Pacheco-Mendoza, Josmel**; **Taype-Rondan, Alvaro** (2016). "Sci-Hub and medical practice: an ethical dilemma in Peru". *The lancet global health*, v. 4, n. 9, e608.
https://doi.org/10.1016/s2214-109x(16)30188-7

Björk, Bo-Christer (2017). "Gold, green, and black open access". *Learned publishing*, v. 30, n. 2, pp. 173-175.
https://doi.org/10.1002/leap.1096

Bohannon, John (2016a). "The frustrated science student behind Sci-Hub". *Science*, v. 352, n. 6285, p. 51.
https://doi.org/10.1126/science.352.6285.511

Bohannon, John (2016b). "Who's downloading pirated papers? Everyone". *Science*, v. 352, n. 6285, pp. 508-512.
https://doi.org/10.1126/science.352.6285.508

Bosman, Jeroen; **Kramer, Bianca** (2018). "Open access levels: A quantitative exploration using Web of Science and oaDOI data". *PeerJ preprints*, 6, e3520v1.
https://doi.org/10.7287/peerj.preprints.3520v1

Cochran, Angela (2016). "A funny thing happened on the way to OA". *The scholarly kitchen*, 25 February.
https://scholarlykitchen.sspnet.org/2016/02/25/a-funny-thing-happened-on-the-way-to-oa

Crissinger, Sarah (2017). "Access to research and Sci-Hub: Creating opportunities for campus conversations on open access and ethics". *College & research libraries news*, v. 78, n. 2, pp. 86-95.
https://doi.org/10.5860/crln.78.2.9621

Custodians Online Campaign (2015). *In solidarity with Library Genesis and Sci-Hub*.
http://custodians.online

De-Castro, Pablo (2016). "Breves notas sobre la conferencia anual 2016 de *Liber* en Helsinki". *Blog EPI*, 12 July.
http://www.elprofesionaldelainformacion.com/notas/breves-notas-sobre-la-conferencia-anual-2016-de-liber-en-helsinki-pablo-de-castro-liber-open-access-project-officer-miembro-del-grupo-thinkepi

Elbakyan, Alexandra (2016). *Why science is better with Communism? The case of Sci-Hub.* Presentation.
https://digital.library.unt.edu/ark:/67531/metadc850001

Elbakyan, Alexandra (2017). "Some facts on Sci-Hub that Wikipedia gets wrong". *Engineuring*, 2 July.
https://engineuring.wordpress.com/2017/07/02/some-facts-on-sci-hub-that-wikipedia-gets-wrong

Elbakyan, Alexandra (2018). "Corrections to The Verge article about Sci-Hub: part 1". *Engineuring*, 16 February.
https://engineuring.wordpress.com/2018/02/16/corrections-to-the-verge-article-about-sci-hub-part-1

Emery, Jill (2016). "Heard on the net: Open access rhapsody". *The Charleston advisor*, v. 18, n. 1, pp. 41-42.
https://doi.org/10.5260/chara.18.1.41

Esposito, Joseph (2017). "Sci-Hub moves to the center of the ecosystem". *The scholarly kitchen*, 5 September.
https://scholarlykitchen.sspnet.org/2017/09/05/sci-hub-moves-center-ecosystem

Faust, Jeremy S. (2016) "Sci-Hub. A solution to the problem of paywalls, or merely a diagnosis of a broken system?". *Annals of emergency medicine*, v. 68, n. 1, pp. A15-A17.
https://doi.org/10.1016/j.annemergmed.2016.05.010

Fernández-Marcial, Viviana; **Costa, Luis-Miguel**; **González-Solar, Llarina** (2016). "Top universities, top libraries: Do research services in academic libraries contribute to university output?". In: *Library Leadership in a Sea of Change*, *37th Iatul Conference*, Halifax (Canada).
http://docs.lib.purdue.edu/iatul/2016/spaces/2

Gardner, Carolyn; **Gardner, Gabriel** (2015). "Bypassing interlibrary loan via Twitter: An exploration of #icanhazpdf requests". In: *ACRL 2015*.
http://hdl.handle.net/10760/24847

Gardner, Gabriel J.; **McLaughlin, Stephen R.**; **Asher, Andrew D.** (2017). "Shadow libraries and you: Sci-Hub usage and the future of ILL". In: *ACRL 2017*.
http://hdl.handle.net/10760/30981

Greco, Albert N. (2017). "The Kirtsaeng and Sci-Hub cases: The major U.S. copyright cases in the twenty-first century". *Publishing research quarterly*, v. 33, n. 3, pp. 238-253.
https://doi.org/10.1007/s12109-017-9522-7

Greshake-Tzovaras, Bastian (2016). "Correlating the Sci-Hub data with World Bank indicators and identifying academic use". *The winnower*, v. 5, e146485.57797.
https://doi.org/10.15200/winn.146485.57797

Greshake-Tzovaras, Bastian (2017). "Looking into Pandora's box: The content of Sci-Hub and its usage". *F1000Research*, v. 6, n. 541.
https://doi.org/10.12688/f1000research.11366.1

Himmelstein, Daniel S.; **Rodriguez-Romero, Ariel**; **Levernier, Jacob G.**; **Munro, Thomas-Anthony**; **McLaughlin, Stephen-Reid**; **Greshake-Tzovaras, Bastian**; **Greene, Casey S.** (2018). "Sci-Hub provides access to nearly all scholarly literature". *eLife*, v. 7, e32822.
https://elifesciences.org/articles/32822

Houle, Louis (2017). "Sci-Hub and LibGen: what if… why not?" In: *IFLA WLIC 2017 Libraries. Solidarity. Society*, 16-17 August, Gdansk (Poland).
http://library.ifla.org/id/eprint/1892

Hoy, Matthew B. (2017). "Sci-Hub: What librarians should know and do about article piracy". *Medical reference services quarterly*, v. 36, n. 1, pp. 73-78.
https://doi.org/10.1080/02763869.2017.1259918

Khabsa, Madian; **Giles, C. Lee** (2014). "The number of scholarly documents on the public web". *PLoS one*, v. 9, n. 5, e93949.
https://doi.org/10.1371/journal.pone.0093949

Kirsop, Barbara; **Chan, Leslie** (2005). "Transforming access to research literature for developing countries". *Serials review*, v. 31, n. 4, pp. 246-255.
https://doi.org/10.1080/00987913.2005.10764998

Larivière, Vincent; **Haustein, Stefanie**; **Mongeon, Philippe** (2015). "The oligopoly of academic publishers in the digital era". *PLoS one*, v. 10, n. 6, e0127502.
https://doi.org/10.1371/journal.pone.0127502

Lawson, Stuart (2017). "Access, ethics and piracy". *Insights. The UKSG journal*, v. 30, n. 1, pp. 25-30.
https://doi.org/10.1629/uksg.333

Machin-Mastromatteo, Juan D.; **Uribe-Tirado, Alejandro**; **Romero-Ortiz, Maria E.** (2016). "Piracy of scientific papers in Latin America". *Information development*, v. 32, n. 55, pp. 1806-1814.
https://doi.org/10.1177/0266666916671080

Marple, Holly (2018). "Parasite, pirate, and Robin Hood: Sci-Hub is disrupting the world of academic publishing". *The journal*, v. 3, n. 2.
https://theijournal.ca/index.php/ijournal/article/view/29480

McKenzie, Lindsay (2017). "Sci-Hub's cache of pirated papers is so big, subscription journals are doomed, data analyst suggests". *Science news*, 27 July.
https://doi.org/10.1126/science.aan7164

McNutt, Marcia (2016). "My love-hate of Sci-Hub". *Science*, v. 352, n. 6285, pp. 497-497.
https://doi.org/10.1126/science.aaf9419

Meadows, Alice (2015). "Beyond open: Expanding access to scholarly content". *The journal of electronic publishing*, v. 18, n. 3.
https://doi.org/10.3998/3336451.0018.301

Mejia, Christian R.; **Valladares-Garrido, Mario J.**; **Miñan-Tapia, Armando**; **Serrano, Felipe T.**; **Tobler-Gómez, Liz E.**; **Pereda-Castro, William**; **Mendoza-Flores, Cynthia R.**; **Mundaca-Manay, Maria Y.**; **Valladares-Garrido, Danai** (2017). "Use, knowledge, and perception of the scientific contribution of Sci-Hub in medical students: Study in six countries in Latin America". *PLoS one*, v. 12, n. 10, e0185673.
https://doi.org/10.1371/journal.pone.0185673

Nicholas, David; **Boukacem-Zeghmouri, Chérifa**; **Xu, Jie**; **Herman, Eti**; **Clark, David**; **Abrizah, Abdullah**; **Rodríguez-Bravo, Blanca**; **Świgoń, Marzena** (2018). "Sci-Hub: The new and ultimate disruptor? View from the front". *Learned publishing*, 1 November.
https://doi.org/10.1002/leap.1206

Novo, Lluís A. B.; **Onishi, Viviani C.** (2017). "Could Sci-Hub become a quicksand for authors?". *Information development*, v. 33, n. 3, pp.324-325.
https://doi.org/10.1177/0266666917703638

Oakley, Meg (2016). *Online piracy: Why Sci-Hub is disrupting scholarly publishing*. Georgetown University Library.
https://www.library.georgetown.edu/sites/default/files/sci-hub-intro.pdf

Penn, Louise (2018). "Alternative ways of obtaining scholarly articles and the impact on traditional publishing models from a UK/European perspective". *Serials review*, v. 44, n. 1, pp. 40-50.
https://doi.org/10.1080/00987913.2018.1433906

Piwowar, Heather; **Priem, Jason**; **Larivière, Vincent**; **Alperin, Juan-Pablo**; **Matthias, Lisa**; **Norlander, Bree**; **Farley, Ashley**; **West, Jevin**; **Haustein, Stefanie** (2018). "The state of OA: A large-scale analysis of the prevalence and impact of open access articles". *PeerJ*, v. 6, e4375.
https://doi.org/10.7717/peerj.4375

Priego, Ernesto (2016). "Signal, not solution: Notes on why Sci-Hub is not opening access". *The winnower*, v. 5, e145624.49417.
https://doi.org/10.15200/winn.145624.49417

Reichman, Hank (2016). "News: Is it legal?". *Journal of intellectual freedom and privacy*, v. 1, n. 2-3, p. 50.
https://doi.org/10.5860/jifp.v1i2-3.6173

Ruff, Corinne (2016). "Librarians find themselves caught between journal pirates and publishers". *The chronicle of higher education*, 18 February.
https://www.chronicle.com/article/Librarians-Find-Themselves/235353

Russell, Carrie; **Sanchez, Ed** (2016). "Sci-Hub unmasked: Piracy, information policy, and your library". *College & research libraries news*, v. 77, n. 3, pp. 122-125.
https://doi.org/10.5860/crln.77.3.9457

Sathe, Nila A.; **Grady, Jenifer L.**; **Giuse, Nunzia B.** (2002). "Print *versus* electronic journals: A preliminary investigation into the effect of journal format on research processes". *Journal of the Medical Library Association*, v. 90, n. 2, pp. 235-243.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC100770

Sparc (2018). *Big deal cancellation tracking*.
https://sparcopen.org/our-work/big-deal-cancellation-tracking

Steel, Graham (2016). "Wallets with a serious case of Stockholm syndrome: Sci-Hub and the future of scholarly communication". *The winnower*, v. 5, e145677.73287.
https://doi.org/10.15200/winn.145677.73287

Strielkowski, Wadim (2017). "Will the rise of Sci-Hub pave the road for the subscription-based access to publishing databases?". *Information development*, v. 33, n. 5, pp. 540-542.
https://doi.org/10.1177/0266666917728674

Swartz, Aaron (2008). *Guerilla open access manifesto*.
https://archive.org/details/GuerillaOpenAccessManifesto

Tenopir, Carol; **King, Donald W.**; **Edwards, Sheri**; **Wu, Lei** (2009). "Electronic journals and changes in scholarly article seeking and reading patterns". *Aslib proceedings*, v. 61, n. 1, pp. 5-32.
https://doi.org/10.1108/00012530910932267

Travis, John (2016). "In survey, most give thumbs-up to pirated papers". *Science news*, May 6th.
https://doi.org/10.1126/science.aaf5704

Triggle, Chris R.; **Triggle, David J.** (2016). "From Gutenberg to open science: An unfulfilled odyssey". *Drug development research*, v. 78, n. 1, pp. 3-23.
https://doi.org/10.1002/ddr.21369

Van-Noorden, Richard (2016) "Alexandra Elbakyan: Paper pirate". *Nature*, v. 540, n. 7634.
https://doi.org/10.1038/540507a

Woolston, Chris (2016). "Paper piracy sparks online debate". *Nature*, 2 May.
https://doi.org/10.1038/nature.2016.19841

PUBLICACIONES *EL PROFESIONAL DE LA INFORMACIÓN*

REVISTA *EL PROFESIONAL DE LA INFORMACIÓN*

Desde 2015 sólo online

- ☐ Suscripción anual institucional ..... 170 €
- ☐ Suscripción anual individual ..... 115 €

http://recyt.fecyt.es/index.php/EPI/index

ANUARIO THINKEPI

Desde 2019 no es necesaria suscripción. Pasa a ser una publicación OA.

COLECCIÓN LIBROS *EPI SCHOLAR* (Editorial *UOC*)

- ☐ 1. Cibermetría. Midiendo el espacio red ..... 20,00€
  Enrique Orduña-Malea e Isidro F. Aguillo
- ☐ 2. La web social como nuevo medio de comunicación y evaluación científica ..... 21,00€
  Amalia Mas-Bleda e Isidro F. Aguillo
- ☐ 3. SEO. Introducción a la disciplina del posicionamiento en buscadores ..... 20,00€
  Jorge Serrano Cobos
- ☐ 4. Métricas de la web social para bibliotecas ..... 26,00€
  Nieves González-Fernández-Villavicencio
- ☐ 5. Calidad en sitios web. Métodos de análisis general, e-comerce, Imágenes, hemerotecas y turismo ..... 19,50€
  Rafael Pedraza-Jiménez, Lluís Codina, Javier Guallar
- ☐ 6. Gestión de datos de investigación ..... 17,00€
  Fabiano Couto Corrêa
- ☐ 7. Desafío a la investigación estándar en comunicación. Crítica y altermativas ..... 20,00€
  Manuel Goyanes
- ☐ 8. Bibliotecas integradas ..... 15,00€
  Pablo Parra-Valero
- ☐ 9. Gestión de información personal. Integración mediante el correo electrónico ..... 19,00€
  Jorge Franganillo
- ☐ 10. Datos abiertos enlazados de archivos, bibliotecas y museos en la web ..... 15,00€
  Carlos Henrique Marcondes
- ☐ 11. Turismo rural online. Páginas web y redes sociales ..... 18,00€
  Gersón Beltrán

http://www.elprofesionaldelainformacion.com/librosEPIScholar.html

El profesional de la información

ANUARIO Think EPI

En el caso de las publicaciones en papel hay que añadir los gastos de envío

Información y pedidos:

Isabel Olea
epi.iolea@gmail.com
Tel.: +34-608 491 521

LIBROS *EL PROFESIONAL DE LA INFORMACIÓN*

(Editorial *UOC*)

- ☐ 47. Makerspaces y bibliotecas ..... 15,00 €
  Julio Alonso-Arévalo
- ☐ 46. Bibliotecas ciudadanas. Espacios de desarrollo y participación ..... 12,00 €
  João De-Sousa-Guerreiro y María-Ramona Domínguez-Sanjurjo
- ☐ 45. Gamificación en bibliotecas. El juego como inspiración ..... 15,00 €
  Ana Ordás
- ☐ 44. Biblioteca: el servicio público con corazón ..... 14,00 €
  Magdalena Ortiz-Macías
- ☐ 43. La Hora del Cuento en las bibliotecas. Claves para su organización... 11,50 €
  Joan de Boer
- ☐ 42. Curación de contenidos para bibliotecas..... 11,50 €
  Evelio Martínez-Cañadas
- ☐ 41. Plan de contenidos para medios sociales ..... 11,50 €
  Eva Sanagustín
- ☐ 40. Legislación de archivos ..... 12,50 €
  Vicente Giménez-Chornet
- ☐ 39. Gestión de proyectos paso a paso ..... 14,50 €
  Candela Ollé y Berta Cerezuela
- ☐ 38. Fuentes de información. Guía básica y nueva clasificación ..... 13,00 €
  Rosana López-Carreño
- ☐ 37. La medición en plataformas sociales ..... 11,00 €
  Adrián Segovia
- ☐ 36. Big data. Gestión y explotación de grandes volúmenes de datos .... 15,00 €
  Montserrat García-Alsina
- ☐ 35. Estudios de usuarios de información..... 12,00 €
  Isabel Villaseñor-Rodríguez
- ☐ 34. Autopublicar. Los nuevos circuitos para autores e investigadores.. 12,00 €
  Almudena Mangas-Vega
- ☐ 33. Lectura digital infantil. Dispositivos, aplicaciones y contenidos..... 13,00 €
  Araceli García-Rodríguez y Raquel Gómez Díaz
- ☐ 32. Geolocalización online. La importancia del dónde ..... 12,00 €
  Gersón Beltrán
- ☐ 31. Biblioteca pública. Mientras llega el futuro ..... 14,50 €
  Fernando Juárez-Urquijo
- ☐ 30. Geobibliotecas ..... 11,50 €
  Estefanía Aguilar-Moreno y Carlos Granell-Canut
- ☐ 29. Los sexenios de investigación ..... 12,00 €
  Álvaro Cabezas-Clavijo y Daniel Torres-Salinas
- ☐ 28. Los impresos antiguos ..... 12,50 €
  Jon Zabala
- ☐ 27. La intranet social ..... 13,50 €
  Ana Carrillo Pozas
- ☐ 26. Innovación en bibliotecas ..... 13,50 €
  Isabel Riaza
- ☐ 25. Gestión de la calidad en la biblioteca ..... 13,50 €
  Nuria Balagué y Jarmo Saarti
- ☐ 24. El content curator ..... 14,00 €
  Javier Guallar y Javier Leiva-Aguilera
- ☐ 23. Marca y comunicación empresarial ..... 11,00 €
  Pablo Medina-Aguerrebere
- ☐ 22. Bibliotecas escolares ..... 10,00 €
  Concepción Mª Jiménez-Fernández y Raúl Cremades-García
- ☐ 21. Inteligencia en redes sociales ..... 11,00 €
  Eva Moya
- ☐ 20. Archivos ..... 14,00 €
  Ramón Alberch-Fugueras
- ☐ 19. Documentación cinematográfica ..... 12,00 €
  Elena De la Cuadra
- ☐ 18. Documentación fotográfica ..... 14,00 €
  Juan-Miguel Sánchez-Vigil y Antonia Salvador-Benítez
- ☐ 17. Gestión de contenidos ..... 13,00 €
  Ricardo Eíto-Brun

BOLETÍN DE SUSCRIPCIÓN

El profesional de la información

Deseo suscribirme a la revista EPI a partir del mes de enero del año 20 ☐ Las suscripciones van por años naturales, de enero a diciembre

Suscripción ☐ Institucional ☐ Personal

Nombre _____ Institución _____

(Los suscriptores individuales no han de escribir ningún nombre de institución, sólo indicar la dirección particular)

Departamento _____ NIF institucional _____

Dirección _____ Código postal _____ Ciudad _____ País _____

Teléfono _____ Fax _____ Correo-e _____

Método de pago

☐ Tarjeta de crédito ☐ VISA ☐ Master Card

Titular de la tarjeta _____ Código de seguridad CVC2 _____

Número de tarjeta _____ Caducidad (mm/aaaa) _____

☐ Cheque nominativo en euros a nombre de *El profesional de la información*

☐ Transferencia bancaria a la cuenta de La Caixa **ES95 2100 0818 93 0200745544** Enviar, fotocopiado o escaneado, el resguardo de la transferencia.

☐ Giro postal al apartado de correos 13 de León Enviar, fotocopiado o escaneado, el resguardo del giro.

☐ Domiciliación en cuenta bancaria

Entidad _____ Oficina _____ DC _____ Núm _____

Titular de la cuenta _____

☐ PayPal a la cuenta EPISCP@gmail.com

Enviar el boletín cumplimentado por correo electrónico, o postal al APARTADO 13 - 24080 LEÓN - ESPAÑA.
Consultas: suscripciones@elprofesionaldelainformacion.com o +34 608 491 521

Open choice. Los autores pueden liberar su artículo en open access en las webs de *EPI* mediante el pago de 650€ + 21%IVA

Precios 2019

Importante: Desde enero de 2015 EPI sólo se publica online

Suscripción institucional EPI: 140,50 € + 21% IVA = 110 €

Suscripción personal: 95,05 € + 21% IVA = 115 €

SNúmero suelto: 30,58 € + IVA = 37 €

NORMAS PARA LOS AUTORES

El profesional de la información tiene dos secciones principales:
ARTÍCULOS: Trabajos de investigación y temas analizados en profundidad.
ANÁLISIS: Experiencias, estudios de casos, análisis de productos, etc.

Las contribuciones han de ser originales e inéditas, no pueden haberse publicado previamente en soporte papel o electrónico.

El texto ha de enviarse en Word, rtf u odt. Las tablas deberán ir pegadas en el mismo documento. Todos los materiales gráficos (diagramas, fotografías, capturas de pantalla, etc.) deben pegarse en el Word y además enviarse en ficheros independientes (en formatos xls, jpg, pdf, etc.). Las imágenes jpg deben tener una resolución de al menos 300 pp (unos 200 KB cada una).

El texto debe presentarse completamente plano, sin autoformatos ni automatismos de Word (subsecciones, viñetas, citas enlazadas, pies de página, sangrías, tabulaciones, colores, etc.), pero debe seguir el estilo de EPI en cuanto a **negritas** (nombres de los autores citados), *cursivas* (instituciones, títulos de revista, marcas) y mayúsculas. Los urls deben estar sin hipervínculo.

Las citas bibliográficas en el texto se realizarán de la forma: (**Apellido**, año) o (**ApellidoAutor1**; **ApellidoAutor2**, año).

La redacción debe ser concisa y precisa, evitando la retórica.

Los trabajos deben incluir: a) título, b) resumen de 100-150 palabras, c) 5-10 palabras clave, d) title, e) abstract de 100-150 palabras, f) 5-10 keywords.

Aparte se incluirá el nombre de los autores, su lugar de trabajo y dirección (postal y electrónica), su foto tipo carnet superior a 100 KB en jpg, un currículum de unas 70 palabras, y su orcid

EVALUACIÓN

Los trabajos son revisados según el sistema tradicional "peer review" en doble ciego por al menos dos expertos en el tema, del Consejo Asesor de la revista y/o externos. La revista se compromete a informar del resultado a los autores.

REFERENCIAS BIBLIOGRÁFICAS

Ordenadas alfabéticamente por autor, se limitarán a las obras citadas en el artículo. No se acepta bibliografía de relleno.

Artículos de una publicación periódica:
Apellido, Nombre; **Apellido2, Nombre2**. "Título del artículo".
Título de la publicación periódica, año, v., n., pp. xx-yy.
Dirección url iniciada en nueva línea sin barra y sin punto finales
http://dx.doi.org/10.xxxx/doi

Ponencia presentada en un congreso:
Apellido, Nombre; **Apellido2, Nombre2**. "Título de ponencia".
En: *Nombre del congreso*, año, pp. xx-yy.
Dirección url iniciada en nueva línea sin barra y sin punto finales
http://dx.doi.org/10.xxxx/doi

Monografías e informes:
Apellido, Nombre; **Apellido2, Nombre2**. *Título del trabajo*.
Lugar de publicación: editor, fecha, ISBN: espacios, sin guiones
Dirección url iniciada en nueva línea sin barra y sin punto finales
http://dx.doi.org/10.xxxx/doi

Capítulo de una monografía:
Apellido, Nombre; **Apellido2, Nombre2**. "Título del capítulo".
En: Apellido, Nombre; Apellido2, Nombre2. *Título de la monografía*. Lugar de publicación: editor, fecha, pp. xx-yy. ISBN: espacios, sin guiones
Dirección url iniciada en nueva línea sin barra y sin punto finales

Recurso online:
Apellido, Nombre; **Apellido2, Nombre2**. *Título del recurso (sin fecha de la consulta)*.
Dirección url iniciada en nueva línea sin barra y sin punto finales

Todas las contribuciones se tienen que enviar a la sección *EPI* de la plataforma *OJS* del *Repositorio Español de Ciencia y Tecnología* (*Recyt*) de la *Fecyt*:
http://recyt.fecyt.es/index.php/EPI/index

Previamente los autores deben registrarse en:
http://recyt.fecyt.es/index.php/EPI/user/registerJournal