



A RITMO DE TWEET: ARCHIVANDO ELECCIONES 2.0



Ciro Llueca, Daniel Cócera, Natalia Torres, Gerard Suades, Ricard De-la-Vega



Ciro Llueca es máster en documentación digital por la *Universitat Pompeu Fabra (UPF)*, licenciado en documentación por la *Universitat Oberta de Catalunya (UOC)* y diplomado en biblioteconomía y documentación por la *Universitat de Barcelona (UB)*. Coordinador de *Patrimoni Digital de Catalunya (Padicat)* desde su creación en la *Biblioteca de Catalunya* en 2005, es también profesor asociado en la *UB*. Fue vicepresidente (2003-2006) del *Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya (Cobdc)*.

Biblioteca de Catalunya
Hospital 56. 08001 Barcelona
clluca@bnc.cat

Gerard Suades es ingeniero en informática por la *Universitat Autònoma de Barcelona (UAB)* y trabaja como técnico de proyectos en el área de portales y repositorios del *Centre de Supercomputació de Catalunya (Cesca)*.

Centre de Supercomputació de Catalunya
Gran Capità 2-4 (Nexus). 08034 Barcelona
cesca.cat

Natalia Torres, ingeniera en informática por la *Universitat Politècnica de Catalunya (UPC)*, trabaja como técnica de proyectos en el área de portales y repositorios del *Centre de Supercomputació de Catalunya (Cesca)*.

Centre de Supercomputació de Catalunya
Gran Capità 2-4 (Nexus). 08034 Barcelona
ntorres@cesca.cat

Ricard De-la-Vega, ingeniero técnico en informática de gestión por la *Universitat Politècnica de Catalunya (UPC)*, coordina el área de portales y repositorios del *Centre de Supercomputació de Catalunya (Cesca)*.

Centre de Supercomputació de Catalunya
Gran Capità 2-4 (Nexus). 08034 Barcelona
rdelavega@cesca.cat

Daniel Cócera es licenciado en documentación y en historia por la *Universitat Autònoma de Barcelona (UAB)* y es el responsable de colecciones de *Padicat* en la *Biblioteca de Catalunya*.

Biblioteca de Catalunya
Hospital 56. 08001 Barcelona
padicat@bnc.cat

Resumen

Patrimoni Digital de Catalunya (Padicat) es el archivo web creado en 2005 por la *Biblioteca de Catalunya* con la misión de capturar, procesar y dar acceso permanente al patrimonio digital publicado en internet para el público de Catalunya. Su política de colección está basada en tres ejes: captura exhaustiva del dominio .cat; compilación selectiva de recursos digitales publicados por instituciones y empresas catalanas; y captura monográfica de acontecimientos de la vida pública catalana, tales como las campañas electorales en internet. Se describen las acciones de captura y preservación de las webs relativas a las últimas convocatorias electorales, con el objetivo de mostrar a las generaciones futuras cómo son las campañas electorales en internet a comienzos del siglo XXI.

Palabras clave

Repositorios digitales, Bibliotecas digitales, Archivos web, Preservación digital, Campañas electorales.

Title: Archiving elections 2.0 at the speed of tweet

Abstract

Patrimoni Digital de Catalunya (Padicat) is the web archive created in 2005 by the *Biblioteca de Catalunya (BC)*, with the aim of collecting, processing and providing permanent access to the digital heritage on internet for the people of Catalonia

Artículo recibido el 23-02-11
Aceptación definitiva: 26-04-11

(Spain). Its collection policy is based on three pillars: exhaustive harvesting of the .cat domain, selective compilation of digital resources published by Catalan institutions and companies, and monographic capture of social events, such as electoral campaigns on the Internet. The article describes the actions taken to capture and preserve web pages for the most recent electoral contests, in order to show future generations the internet electoral campaigns of the early 21st century.

Keywords

Digital repositories, Digital libraries, Web archives, Digital preservation, Electoral campaigns.

Llueca, Ciro; Cócera, Daniel; Torres, Natalia; Suades, Gerard; De-la-Vega, Ricard. "A ritmo de tweet: archivando elecciones 2.0". *El profesional de la información*, 2011, mayo-junio, v. 20, n. 3, pp. 309-314.

DOI: 10.3145/epi.2011.may.10

1. El archivo de internet

La *Biblioteca de Catalunya* (BC) inició en 2005 un proyecto para compilar y preservar la internet catalana: la creación del archivo web o depósito digital *Padicat*¹ (*Patrimoni Digital de Catalunya*). Pese a múltiples dificultades legislativas y técnicas, el archivo web de Catalunya se sumaba a otras iniciativas ya existentes² dirigidas a la preservación del patrimonio nacido digital más evidente: las páginas web que se publican en internet.

Padicat se planificó a partir del análisis de los proyectos existentes, basados en dos tipos de políticas de colección: el modelo integral o exhaustivo (mayoritario y característico especialmente de los países escandinavos), que persigue la integración automática de la Web a partir de determinados criterios infraestructurales (dominio, ubicación del servidor, lingüísticos, etc.); y el modelo selectivo (asimilado por Australia, el Reino Unido o Japón³, entre otros), dirigido a compilar la Web en base a una política determinada (sobre un espacio geográfico determinado, un tema de interés nacional, etc.). Estos dos modelos han dado paso en lo que es ya una tendencia generalizada, a modelos híbridos (cuyo caso más evidente es Dinamarca⁴) que complementan la captura periódica de la Web con acciones selectivas, ampliando esa cobertura a determinados acontecimientos de interés social (elecciones, competiciones deportivas, etc.). A este último modelo pertenece justamente la política de colección del depósito catalán, basada en tres ejes:

- compilación masiva de los recursos digitales publicados en abierto en internet, mediante la captura exhaustiva del dominio .cat, así como webs ubicadas en otros dominios;
- depósito sistemático de la producción web de las entidades catalanas, mediante la identificación y el convenio con entidades y empresas;
- promoción de líneas de investigación mediante la presentación temática de los recursos digitales de determinados acontecimientos de la vida pública catalana, como es el caso de las campañas electorales en internet, el fenómeno de la música en línea, o los museos en internet.

El archivo de internet es hoy razonablemente posible. *Padicat*, tras 6 años de existencia, contiene 118.039 capturas de 39.587 páginas web y está formado por 249 millones de ficheros, con un tamaño de 7,5 TB. Toda la colección se halla accesible en línea.

<http://www.padicat.cat>

Se han firmado acuerdos de colaboración con 450 instituciones y empresas. Y se han creado 7 monográficos dedicados a las campañas electorales en internet, música folk-rock, y museos catalanes. Con la actual infraestructura técnica, garantizada por el socio tecnológico del proyecto, el *Centre de Supercomputació de Catalunya* (*Cesca*), la previsión de crecimiento anual para 2011 se establece en 75.700 nuevas versiones de 32.000 webs, 50 nuevos convenios, y 2 nuevos monográficos. Se ha iniciado la elaboración de la hemeroteca digital de internet, con capturas diarias de 50 medios de comunicación digitales. Hay motivos sobrados para la satisfacción.

Pero también hay sombras que acompañan a estas luces. En primer lugar la infraestructura técnica del depósito debe poder crecer, y no está claro que en la actual coyuntura económica la administración pública catalana pueda mantener el ritmo de actualización de los recursos técnicos y su consolidación, lo que en el peor de los casos llevaría a la parálisis del archivo web catalán. En segundo lugar, como en otros archivos web del resto del mundo, se ha avanzado poco en el diseño de las estrategias de preservación digital de los ficheros que contiene el depósito, ya que se ha estado ocupado en alimentar la colección: crecer en recursos depositados. En tercer lugar, pese a la estandarización de los programas informáticos que usan *Padicat* y el resto de repositorios, aún no es posible un intercambio eficaz de ficheros y registros bibliográficos. El uso de pasarelas está todavía en una fase precaria, y los programas informáticos dedicados al archivo de sitios web están lejos de ser infalibles. Finalmente, para garantizar el retorno a la sociedad es imprescindible la cooperación entre archivos web de España, Iberoamérica, y el resto del mundo; y también entre repositorios digitales de las llamadas instituciones de la memoria: bibliotecas, archivos y museos. Como ha venido ocurriendo a lo largo de nuestra historia profesional, en pocas ocasiones conseguimos romper la inercia del trabajo aislado.

Estos retos de futuro no deben ocultar que el archivo de internet es actualmente posible, y que las instituciones públicas tienen el deber de garantizar el acceso y la preservación de los contenidos publicados en la Web: el patrimonio digital.

2. Elecciones en internet: nuevos escenarios de debate

En el sistema democrático español las administraciones públicas pueden renovar sus dirigentes políticos cada cuatro

años. La irrupción de las TIC en la política ha transformado la actividad de los cargos electos y representantes políticos y sociales⁵ porque el uso de las tecnologías por parte de los partidos políticos ha complementado sus estrategias comunicativas, basadas esencialmente en los impactos en medios de comunicación tradicionales⁶. La utilización de páginas web y blogs personales, la presencia generalizada en redes sociales, el uso de vídeos y aplicaciones multimedia, etc., ha llevado a una situación en la que una parte importante de las acciones de comunicación política durante una campaña electoral se dirimen en internet, exclusivamente en un formato digital: es parte de la filosofía 2.0 aplicada a la política.

El archivo de internet es hoy razonablemente posible

Esta circunstancia, el uso generalizado de la Red para la campaña política, y el hecho de que se focalice en un período de tiempo limitado de unos tres meses (precampaña, campaña, elecciones, postcampaña), llevó al archivo web australiano, *Pandora*, a iniciar en 1996 procesos monográficos de captura y preservación de campañas electorales en internet. Paulatinamente se sumaron a estas prácticas otros proyectos.

Padicat contempló en su política de colección la captura sistemática de las campañas electorales en internet, y en consecuencia ha venido realizado monográficos de las sucesivas convocatorias electorales con incidencia en Catalunya: al *Parlament de Catalunya* (2006), municipales (2007), al *Congreso* y el *Senado* (2008), al *Parlamento Europeo*⁷ (2009) y de nuevo al *Parlament* (2010), completando el ciclo iniciado cuatro años antes. En el momento de redactar el presente artículo se está llevando a cabo el monográfico para las elecciones municipales de mayo de 2011.

La acción de captura, proceso, preservación y difusión de las webs relativas a estas elecciones se ha centrado esencialmente en los recursos digitales publicados en internet por las candidaturas oficiales (partidos políticos con y sin representación en las administraciones); blogs y avatares sociales de los principales candidatos y candidatas (cabezas de lista y miembros destacados de las candidaturas); medios de comunicación; instituciones públicas afectadas por las elecciones (ayuntamientos, *Parlament*, *Moncloa*); fundaciones y centros de investigación política; y webs y blogs de activistas, articulistas y estrategias políticos. En la selección de recursos ha colaborado profesorado de ciencia y comunicación política de diversas universidades catalanas, así como la mayoría de partidos políticos que han participado en las campañas.

3. Para qué sirve un archivo web

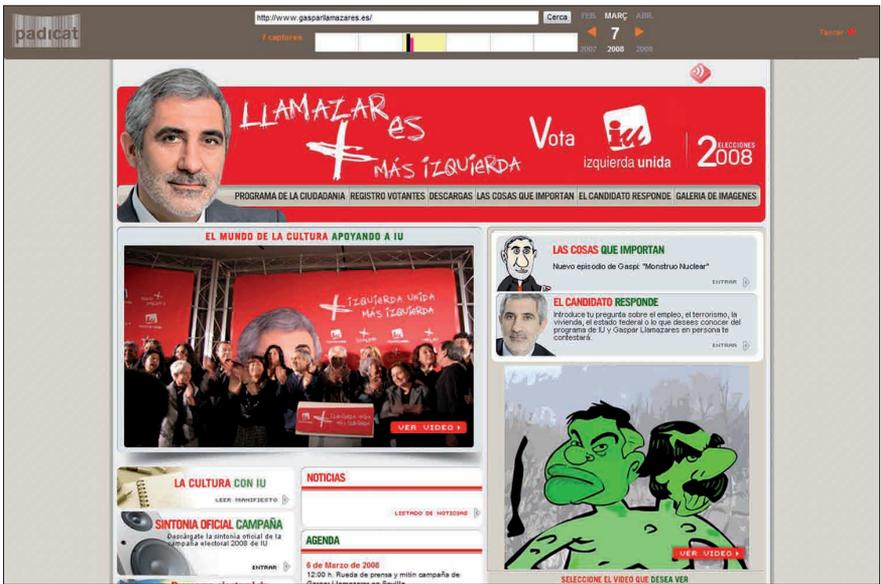
En el planteamiento inicial de creación de un depósito para preservar información publicada en internet convergen diversos factores, ninguno ajeno a la naturaleza misma de las instituciones patrimonialistas públicas. La *Unesco* sigue en sus directrices la lógica aplicable al resto del patrimonio de la humanidad: es de esperar que entre el público que utilice esos repositorios digitales se encuentren organizaciones culturales y de investigación como bibliotecas, archivos, museos, institutos de investigación, archivos de datos, editores, grupos comunitarios y demás interesados en consultar las fuentes de la historia, sean documentos impresos o digitales. El patrimonio digital corre peligro de perderse porque la evolución de la tecnología y la cultura digital ha sido tan rápida que los gobiernos e instituciones no han podido elaborar estrategias de preservación oportunas y bien fundamentadas⁸.

El reciente informe del *Comité de sabios* de la Unión Europea redonda en esta idea: no hay ambición mayor que perpetuar nuestro rico y diverso patrimonio cultural, siendo beneficios para las generaciones futuras el completo acceso y la democratización de la cultura y el conocimiento, especialmente para los sistemas educativos en escuelas y universidades. Complementariamente se persigue afianzar el enlace del pasado con el presente; proteger los intereses de los creadores europeos; nutrir creaciones futuras, incluyendo los esfuerzos de los no profesionales; contribuir a la educación; y estimular la innovación y el espíritu empresarial⁹.

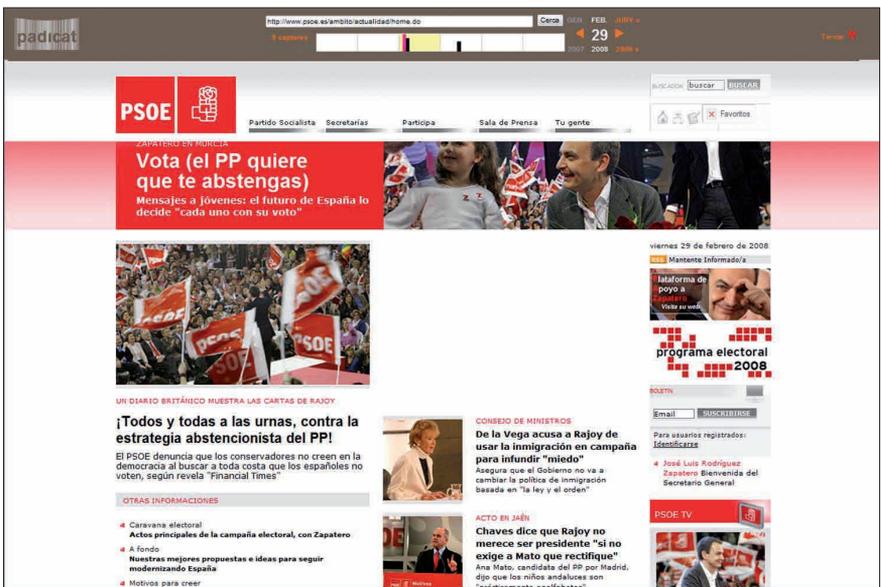
Las instituciones públicas tienen el deber de garantizar el acceso y la preservación de los contenidos publicados en la Web: el patrimonio digital

A una escala doméstica, exponemos tres ejemplos de experiencias recientes de uso de las colecciones de campañas

Web de Àngel Ros, candidato a la alcaldía de Lleida en 2007



Web de Gaspar Llamazares, candidato de IU al Congreso en 2008



Web del PSOE en la campaña electoral de 2008 al Congreso



Web de Mariano Rajoy, candidato del PP al Congreso en 2008

electorales de internet. El primero, para usos docentes y de investigación: una profesora universitaria imparte una asignatura de comunicación y marketing político. En sus clases acompaña a sus descripciones ejemplos reales de acciones comunicativas, concretamente sobre la campaña electoral al *Parlament de Catalunya* de 2006, comparando las consignas lanzadas por los principales partidos políticos en las campañas y el impacto de estos mensajes en el electorado. Por su antigüedad y la propia naturaleza de internet la mayoría de estas sedes web han desaparecido de la Red, siendo un ejemplo paradigmático la web institucional del presidente Pasqual Maragall, con sus declaraciones y discursos, imágenes y vídeos, que sólo existe ya en el depósito *Padicat*.

El segundo caso, fuente de información para los medios de comunicación: un diario híbrido prepara para las elecciones municipales de mayo un análisis sobre el uso de redes sociales y blogs por parte de los candidatos. A partir de los recursos depositados con motivo de las elecciones municipales de 2007 se realizó un informe sobre la presencia virtual de los partidos políticos en Catalunya, concretamente de los 278 candidatos a 48 ciudades catalanas. En total, 227 candidaturas tenían algún tipo de recurso digital al servicio de la campaña, básicamente parte de sedes web corporativas de cada partido. De estas, 165 candidaturas tenían web propia o blog de candidato. En el caso de los blogs se analizó el uso y, transcurridas las elecciones, numerosos casos de abandono.

El tercer caso, reutilización de contenidos: en octubre de 2010 la oficina de uno de los candidatos a la presidencia de la *Generalitat de Catalunya* contactó con *Padicat* para acceder a la propia biografía del candidato, publicada junto a fotos familiares digitalizadas en la anterior convocatoria electoral, de 2006. Como ocurre a menudo, la empresa que diseñó y creó la web del candidato hace cuatro años no conserva esos contenidos, pensados con finalidad efímera, y el equipo de campaña necesita recuperar esos textos e imágenes para agilizar sus procesos, reutilizar las partes oportunas y reducir costes. Tras la sorpresa inicial por una consulta tan poco previsible, se

Monográfico	Páginas web	Capturas	Ficheros	Volumen (GB)
Parlamento 2006	88	775	4.953.215	175
Municipales 2007	615	1.747	13.641.991	457
Generales 2008	147	896	3.117.638	135
Europeas 2009	170	613	5.404.291	233
Parlamento 2010	806	31.210	17.202.999	708
Municipales 2011 (previsión)	1.000	50.000	25.000.000	1.000

Tabla 1. Cifras relativas a las capturas realizadas por *Padicat* durante las últimas cinco convocatorias electorales en Catalunya.

La captura de elecciones en internet contiene webs de las candidaturas oficiales, blogs y avatares de los candidatos, webs de activistas y estrategias políticos...

2. Existen varios archivos web en funcionamiento, así como bibliografía que los ha detallado: **Lluca, Ciro**. "Webs siempre accesibles: las bibliotecas nacionales y los repositorios digitales nacionales". *BiD: textos universitaris de biblioteconomia i documentació*, 2005, diciembre, n. 15. http://www2.ub.edu/bid/consulta_articulos.php?fichero=15lluca1.htm

El patrimonio digital se pierde porque la evolución cultural digital ha sido tan rápida que las instituciones no han podido elaborar estrategias de preservación

Los más conocidos son los pioneros, creados en 1996: el sueco *Kulturarw3*, el australiano *Pandora*, e *Internet Archive*, de alcance internacional. Quince años más tarde podemos contar hasta 50 proyectos en diversas fases de implementación, siendo acciones consolidadas un tercio de esa

informó a la oficina de campaña que el acceso a esa información era libre, resolviendo fácilmente la petición formulada.

El apartado de capturas monográficas de campañas electorales es la fuente más consultada del depósito y el que más impactos mediáticos ha conseguido. Pese a la percepción generalizada de que es un buen producto, es pronto para evaluar el uso de este tipo de colecciones, como posiblemente también el resto del archivo web. En todo caso la realización sistemática de monográficos de elecciones es ya un eje estratégico de la política de la colección digital de la *Biblioteca de Catalunya* en *Padicat*, con el objetivo de mostrar a las generaciones futuras cómo son las campañas electorales en internet a comienzos del siglo XXI.



Twitter de CIU en la campaña electoral al *Parlament de Catalunya* en 2010

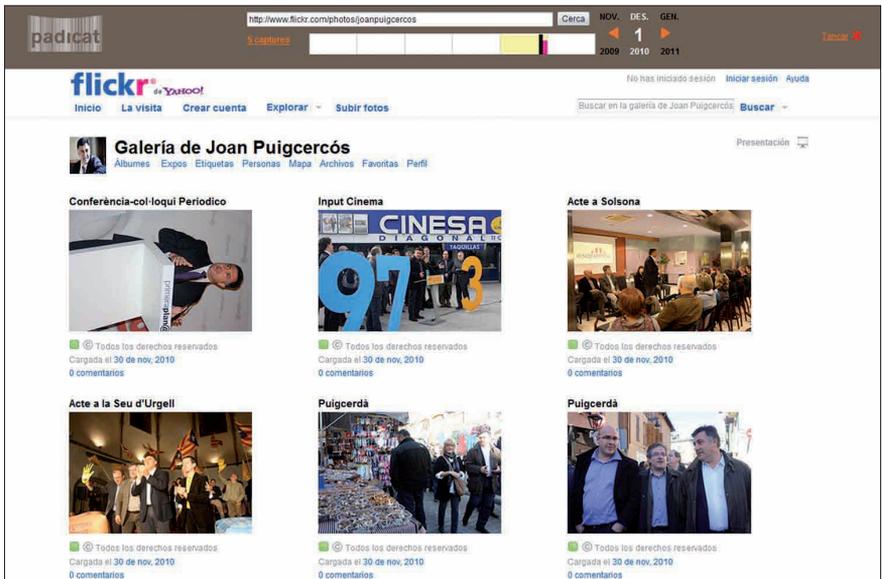
Notas

1. El proyecto se ha descrito a fondo anteriormente. En una fase inicial: **Lluca, Ciro**. "Archivando la Web, el proyecto *Padicat* (*Patrimonio Digital de Catalunya*)". *El profesional de la información*, 2006, noviembre-diciembre, v. 15, n. 6, pp. 473-478. http://eprints.rclis.org/archive/00007767/01/epi_padicat.pdf

Más recientemente: **Lluca, Ciro; Cócera, Daniel; Torres, Natalia; Suades, Gerard; De-la-Vega, Ricard**. "Padicat, el archivo de internet". En: *12as Jornadas Españolas de Documentación*, 2011 [inédito en el momento de presentar este artículo].



Facebook de José Montilla, candidato del *PSC* al *Parlament de Catalunya* en 2010



Flickr de Joan Puigercós, candidato de ERC al Parlament de Catalunya en 2010



Web de Alicia Sánchez-Camacho, candidata del PP al Parlament de Catalunya en 2010

4. El repositorio *Netarkivet.dk* inició su recorrido en 2005 a partir de la publicación de la emblemática ley danesa del Depósito legal de 22 de diciembre de 2004.

<http://www.netarchive.dk>

5. **Barreda, Mikel** (coord.). “La democracia electrónica”. *IDP: Revista de internet, derecho y política*, 2008, n. 6. <http://idp.uoc.edu/ojs/index.php/idp/issue/view/46>

6. Recomendamos el monográfico de *El profesional de la información* sobre “Medios de comunicación en internet”, 2010, noviembre-diciembre, v. 19, n. 6.

<http://www.elprofesionaldelainformacion.com/contenidos/2010/noviembre/index.html>

7. El monográfico dedicado a las elecciones de 2009 al *Parlamento Europeo* fue una acción conjunta de las bibliotecas nacionales de Alemania, Austria, Catalunya, Dinamarca, Francia, Polonia, Reino Unido, República checa y otros miembros del *International Internet Preservation Consortium (IIPC)*. <http://netpreserve.org>

8. *Directrices para la preservación del patrimonio digital*. Canberra: Unesco, 2003. <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>

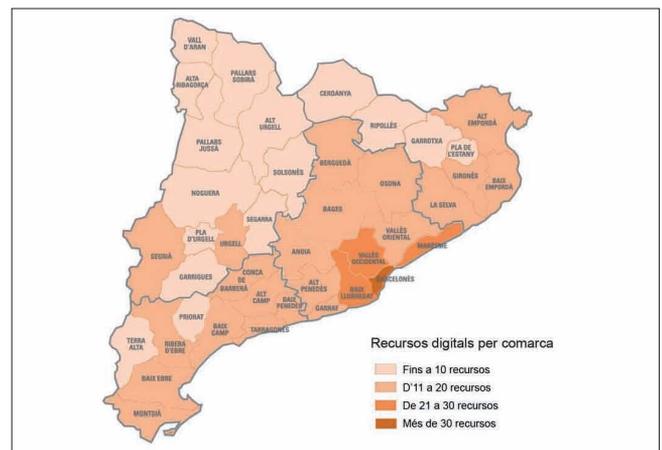
9. *The new renaissance: report of the Comité des Sages: reflection group on bringing Europe’s cultural heritage online*. Luxembourg: European Commission, Information Society and Media DG, 2010.

El apartado de capturas de campañas electorales es la fuente más consultada del depósito y el que más impactos mediáticos ha conseguido

http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/final_report_cds.pdf

cifra. En España, además de *Padicat*, el *Gobierno Vasco* y *EJIE (Sociedad Informática del Gobierno Vasco)* crearon en 2007 *Ondarenet, Archivo Electrónico del Patrimonio Digital Vasco*; desde 2009, la *Biblioteca Nacional de España* encarga capturas periódicas del dominio .es a *Internet Archive*. <http://www.ondarenet.kultura.ejgv.euskadi.net>

3. En Australia se creó en 1996 *Pandora*; en el Reino Unido en 2005 *UK Web Archive*; y en Japón en 2002 *WARP*. En los tres casos, fueron las respectivas bibliotecas nacionales sus principales impulsoras. <http://pandora.nla.gov.au> <http://www.webarchive.org.uk> <http://warp.ndl.go.jp>



Distribución comarcal de recursos digitales durante la campaña de las municipales de 2007 en Catalunya.