

# ¿Cuáles de los metadatos relevantes para bibliometría son iguales y cuáles diferentes al pasar de *Microsoft Academic Graph* a *OpenAlex*?

Which of the metadata with relevance for bibliometrics are the same and which are different when switching from *Microsoft Academic Graph* to *OpenAlex*?

Thomas Scheidsteger; Robin Haunschild

**Note:** This article can be read in its English original version on:  
<https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87295>

Cómo citar este artículo.

Este artículo es una traducción. Por favor cite el original inglés:

**Scheidsteger, Thomas; Haunschild, Robin** (2023). "Which of the metadata with relevance for bibliometrics are the same and which are different when switching from *Microsoft Academic Graph* to *OpenAlex*?". *Profesional de la información*, v. 32, n. 2, e320209.

<https://doi.org/10.3145/epi.2023.mar.09>

Artículo recibido el 18-01-2023  
Aprobado el 08-02-2023



**Thomas Scheidsteger** ✉  
<https://orcid.org/0000-0001-8351-2498>  
Max Planck Institute for Solid State  
Research  
IVS-CPT  
Heisenbergstr. 1  
70569 Stuttgart, Alemania  
[T.Scheidsteger@fkf.mpg.de](mailto:T.Scheidsteger@fkf.mpg.de)



**Robin Haunschild**  
<https://orcid.org/0000-0001-7025-7256>  
Max Planck Institute for Solid State  
Research  
IVS-CPT  
Heisenbergstr. 1  
70569 Stuttgart, Alemania  
[R.Haunschild@fkf.mpg.de](mailto:R.Haunschild@fkf.mpg.de)

## Resumen

Con el anuncio del cierre de *Microsoft Academic Graph* (MAG), la organización sin fines de lucro *OurResearch* anunció que proporcionaría un recurso similar bajo el nombre de *OpenAlex*. Por lo tanto, comparamos los metadatos útiles para los análisis bibliométricos de la última versión de MAG con una versión anterior de *OpenAlex*. Prácticamente todos los trabajos de MAG fueron transferidos a *OpenAlex* conservando sus datos bibliográficos, año de publicación, volumen, primera y última página, y DOI, así como el número de referencias, que son ingredientes importantes del análisis de citas. Más del 90% de los documentos MAG tienen tipos de documentos equivalentes en *OpenAlex*. De los restantes, especialmente las reclasificaciones de los tipos de documento de revista-artículo y libro-capítulo de *OpenAlex* parecen ser correctas y ascienden a más del 7%, por lo que las especificaciones de tipo de documento han mejorado significativamente de MAG a *OpenAlex*. Como otro elemento de metadatos bibliométricos relevante, observamos la clasificación de materias en MAG y en *OpenAlex*. Encontramos más documentos con asignación de clasificación de materias en *OpenAlex* que en MAG. En el primer y segundo nivel, la estructura de clasificación es casi idéntica. Presentamos datos sobre las reclasificaciones de materias en ambos niveles en forma tabular y gráfica. La evaluación de las consecuencias de las abundantes reclasificaciones de materias en las evaluaciones bibliométricas de campos normalizados no está dentro del alcance del presente artículo. Aparte de esto, *OpenAlex* parece ser en general al menos tan adecuado para los análisis bibliométricos como MAG para las publicaciones de los años anteriores a 2021 o incluso mejor debido a la cobertura más amplia de las asignaciones de tipos de documentos.

## Palabras clave

Clasificación de materias; Campos de estudio; Conceptos; Datos bibliográficos; Metadatos; Tipos de documentos; Análisis de citas; Bibliometría; *Microsoft Academic Graph*; MAG; *OpenAlex*.



## Abstract

With the announcement of the retirement of *Microsoft Academic Graph (MAG)*, the non-profit organization *OurResearch* announced that they would provide a similar resource under the name *OpenAlex*. Thus, we compare the metadata with relevance to bibliometric analyses of the latest *MAG* snapshot with an early *OpenAlex* snapshot. Practically all works from *MAG* were transferred to *OpenAlex* preserving their bibliographic data publication year, volume, first and last page, DOI as well as the number of references that are important ingredients of citation analysis. More than 90% of the *MAG* documents have equivalent document types in *OpenAlex*. Of the remaining ones, especially reclassifications to the *OpenAlex* document types journal-article and book-chapter seem to be correct and amount to more than 7%, so that the document type specifications have improved significantly from *MAG* to *OpenAlex*. As another item of bibliometric relevant metadata, we looked at the paper-based subject classification in *MAG* and in *OpenAlex*. We found significantly more documents with a subject classification assignment in *OpenAlex* than in *MAG*. On the first and second level, the classification structure is nearly identical. We present data on the subject reclassifications on both levels in tabular and graphical form. The assessment of the consequences of the abundant subject reclassifications on field-normalized bibliometric evaluations is not in the scope of the present paper. Apart from this open question, *OpenAlex* seems to be overall at least as suited for bibliometric analyses as *MAG* for publication years before 2021 or maybe even better because of the broader coverage of document type assignments.

## Keywords

Subject classification; Fields of study; Concepts; Bibliographic data; Metadata; Document types; Citation analysis; Bibliometrics; *Microsoft Academic Graph*; *MAG*; *OpenAlex*.

### Agradecimientos

Agradecemos a Jason Priem sus útiles comentarios sobre un borrador anterior. El presente estudio se basa en la comunicación presentada en la conferencia *STI2022* de Granada, España (Scheidsteger; Haunschild, 2022), pero se complementa especialmente con una investigación mucho más detallada de las clasificaciones de materias en las dos bases de datos estudiadas.

## 1. Introducción

Desde su lanzamiento en 2015, *Microsoft Academic Graph (MAG)* (Sinha et al., 2015) había sido una fuente de datos nueva y prometedora para los análisis bibliométricos debido a su gran cobertura y al conjunto de metadatos disponibles (Harzing; Alakangas, 2017). Por lo tanto, *MAG* ha sido objeto de muchos estudios, en particular comparaciones con otras bases de datos bibliográficas importantes. En uno de los últimos y más amplios estudios, Visser, Van-Eck y Waltman (2021) compararon *MAG* con *Web of Science*, *Scopus*, *Dimensions* y *Crossref*.

*Microsoft Blog* (2021) anunció que la web *Microsoft Academic*, las APIs y los *snapshots* se retirarían el 31 de diciembre de 2021. Poco después, la organización sin fines de lucro *OurResearch*, con el objetivo de proporcionar “una catálogo completamente abierto del sistema de investigación global” (*OurResearch*, 2021), anunció que conservaría y ampliaría el último corpus completo de *MAG*, excluyendo solo los datos de patentes, y que esperaba mejorarlo. Otra fuente principal de datos es *Crossref*. En enero de 2022 se lanzó *OpenAlex* (<http://docs.openalex.org>) y proporcionó acceso API a sus servicios, así como volcados de datos para cualquier propósito. La *Curtin University's Open Knowledge Initiative (COKI)* ya ha comenzado a monitorear el desarrollo de *OpenAlex*, en particular evaluando y comparando el valor añadido por *OpenAlex* a *MAG* y *Crossref*, tanto en la cobertura de publicaciones como en otros resultados de investigación (Kramer, 2022).

Scheidsteger et al. (2018) estudiaron la posibilidad de utilizar datos *MAG* para el cálculo de puntajes de citación normalizada por campo y fecha. Compararon las puntuaciones derivadas de las clasificaciones de las materias y cobertura en *MAG* a las obtenidas de categorías temáticas y cobertura en la *Web of Science (WoS)*. En el presente estudio, vamos a comparar metadatos que son relevantes para los análisis bibliométricos (en particular, la normalización de fecha y de campo de las citas) de *MAG* y *OpenAlex*:

- cobertura de documentos a lo largo de los años,
- concordancia de los datos bibliográficos,
- número de referencias de cada documento,
- clase y distribución de los tipos de documentos,
- distribución y relación entre clasificaciones de materias.

## 2. Datos y métodos

### *Microsoft Academic Graph (MAG)*

Descargamos el conjunto de datos de *Microsoft Academic Graph (MAG)* a través del portal *Microsoft Azure* a fines de diciembre de 2021 y obtuvimos datos con marca de fecha del 6 de diciembre de 2021 (Sinha et al., 2015). Ver: <https://www.microsoft.com/en-us/research/project/academic>

No pudimos obtener datos más nuevos a principios de 2022 después de la fecha del vencimiento oficial del servicio *MAG*. Según la *OpenAlex migration guide* (OpenAlex, 2021), no se han transferido patentes de *MAG* a *OpenAlex*. Por lo tanto, excluimos de la comparación todos los artículos con tipo de documento *Patent*. Para facilitar la distinción entre las dos bases de datos, mantenemos los nombres de tipos de documentos tal como se utilizan en ambas bases de datos. En particular, los tipos de *MAG* se escriben con mayúscula. Debido a que los datos de *MAG* no contienen el año 2021 completo, restringimos nuestros análisis a los años de publicación anteriores a 2021. Por lo tanto, consideramos 197.445.041 artículos en *MAG*, de los cuales 95.160.734 poseen un DOI.

### OpenAlex

El volcado de datos de *OpenAlex* se realizó el 9 de febrero de 2022, con una actualización con fecha del 31 de enero de 2022 en la tabla principal (*works*). Ambos conjuntos de datos se importaron y procesaron en nuestra base de datos *PostgreSQL* mantenida localmente en el *Max Planck Institute for Solid State Research* (Stuttgart, Alemania). Antes del año de publicación 2021, tenemos un total de 198.606.165 documentos en *OpenAlex*, de los cuales 96.268.256 poseen DOI.

Los documentos en *MAG* y *OpenAlex* se pueden vincular a través de una identificación única. Tanto *OpenAlex* como *MAG* solo contienen referencias vinculadas. Para la mayoría de los trabajos hay disponibles “Fields of Study” (campos de estudio), llamados “concepts” en *OpenAlex* y (solo allí) todos vinculados al *Wikidata* ID respectivo a través de la tabla *concepts*. Para obtener más detalles sobre el enfoque y la estructura de *OpenAlex*, consultar **Priem, Piwowar y Orr (2022)**.

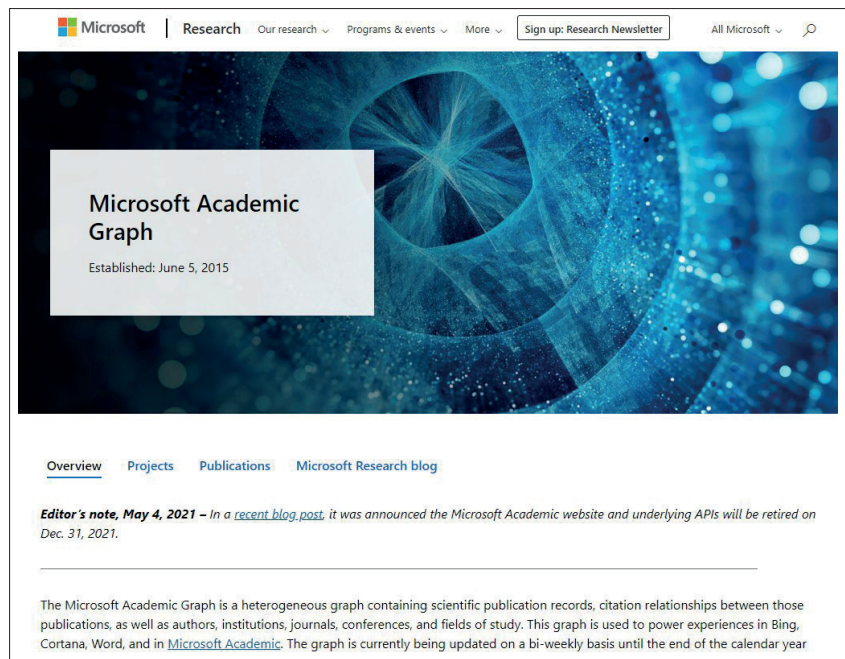
### Software

Las evaluaciones estadísticas se han realizado utilizando *R* (*R Core Team*, 2020), la presentación gráfica en la figura 1 usando el paquete *R ggplot2* (Wickham, 2016), y los gráficos aluviales de las figuras 2, 3, 4 y 5 utilizando el *R package alluvial* (Bojanowski; Edwards, 2016).

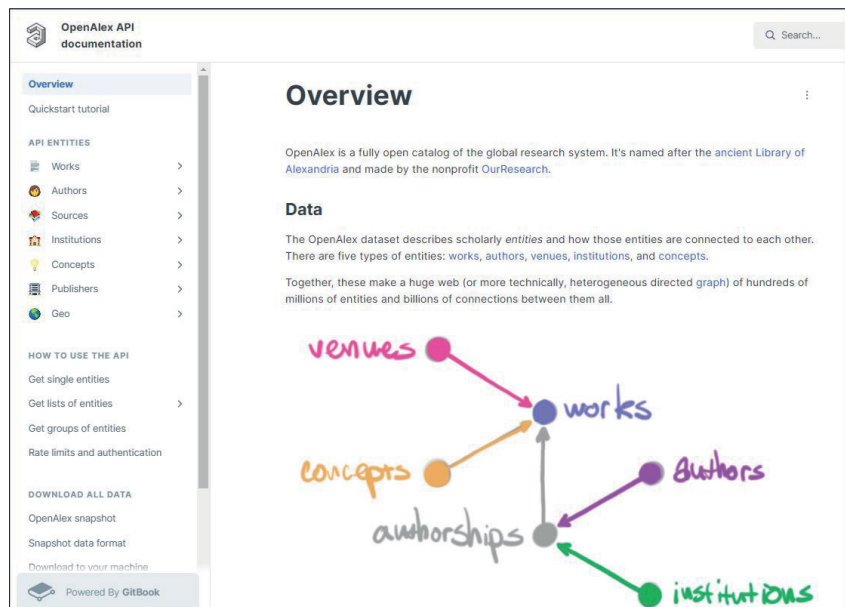
## 3. Resultados

### Cobertura de los años de publicación

Solo 777 IDs de *MAG* no están incorporados en *OpenAlex*, comenzando con un ítem en 1952 y alcanzando un máximo de 201 en 2020. Los tipos de documentos en *MAG* de estos ítems que faltan son alrededor del 40% *Journal* y *None*, cada uno, y alrededor del 15% *BookChapter*. Durante todo el período desde 1952, de las 777 *MAG IDs*, 654 tienen DOI, solo dos de ellas no se pudieron encontrar en *Crossref*. 347 de estos DOI contienen el prefijo ISBN Bookland “978” o “979” y por lo tanto apuntan a libros o capítulos de libros, pero solo un tercio de ellos está asignado a los tipos *Book* o *Book-*



<https://www.microsoft.com/en-us/research/project/microsoft-academic-graph>



<http://docs.openalex.org>

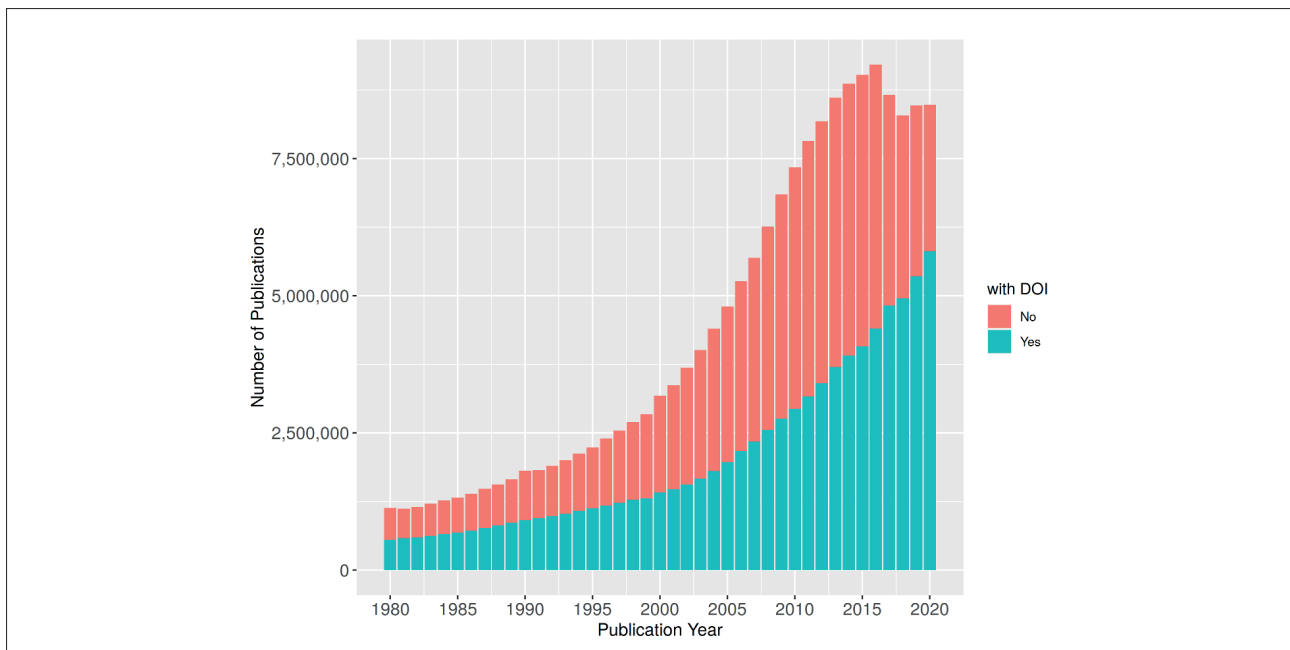


Figura 1. Número de documentos *OpenAlex* - *MAG* comunes entre los años 1980 y 2020

*Chapter* en *MAG*. Como era de esperar, el número 777 de IDs de *MAG* que faltan coincide exactamente con la diferencia entre el número total de artículos de *MAG* y 197.444.264 trabajos de *OpenAlex* que tienen asociado un ID *MAG*. De los 654 DOIs, 23 habían sido asociados a más de un ID *MAG* (dos de ellos con tres, los otros con dos) y –salvo uno (*10.1016/j.physrep.2013.03.005*)– todos se encontraban en *OpenAlex* y cada uno tuvo una aparición menos. En 16 casos la respectiva entrada de preprint se había descartado a favor del artículo de la revista, y en dos casos la entrada como artículo de revista había sido descartada a favor de una entrada como capítulo de libro.

Hay 1.161.901 trabajos indexados en *OpenAlex* que no tienen el correspondiente registro en *MAG*, 1.108.176 de ellos tienen un DOI en *OpenAlex*, en particular 1.877 documentos anteriores a 1800, el primer año de publicación en *MAG*. A continuación, solo se investigarán los documentos que ambas bases de datos tienen en común.

La figura 1 muestra el número anual de documentos comunes con y sin DOI entre los años 1980 y 2020. La disminución inesperada del número total a partir de 2017 se debe a la reducción del número de documentos sin DOI que, a su vez, está dominada por el número de documentos *sin* tipo de documento asignado.

### Comparación de datos bibliográficos

Para los 197.444.264 documentos en *OpenAlex* con una ID en *MAG*, primero verificamos si los datos bibliográficos de *MAG*, como el volumen, el número, la primera página, la última página y el DOI, se conservan después de la transferencia a *OpenAlex*. Cuando el volumen o el número estaban disponibles en *MAG*, estos datos se transfirieron por completo a *OpenAlex*. Este también es el caso de las primeras y últimas páginas y los DOIs. Durante nuestra investigación encontramos algunos problemas con la calidad de los datos (del *MAG* original):

- (i) En más de 28.800 casos, los campos “primera página” y “última página” no contenían un solo número sino el mismo rango de números, por ejemplo, “35-46”.
- (ii) Más de 810.028 DOIs aparecen *más de una vez* en el conjunto de datos, 7.626 de ellos al menos diez veces y 235 al menos 100 veces. De los 100 DOIs más frecuentes, solo se pudieron resolver 29.
- (iii) Más de 6.000 DOIs contienen caracteres no latinos. De ellos solo podrían resolverse menos de 200.

En segundo lugar, con respecto al número de referencias (vinculadas) para un documento, contamos las entradas en la tabla de referencias de *OpenAlex* para cada obra (*works\_referenced\_works*) y no encontramos ninguna desviación de los valores respectivos (*nref*) en la tabla correspondiente de *MAG*. Pero *nref* se había calculado incluyendo referencias de patentes. Evidentemente, en *OpenAlex* se habían mantenido las referencias a las patentes pero no los documentos de las patentes en sí. Esto debe tenerse en cuenta al realizar análisis de citas.

“Inesperadamente, la cantidad total de documentos comienza a disminuir en 2017. Esto se debe a la disminución de la cantidad de documentos sin DOI que, a su vez, está dominada por la cantidad de documentos *sin* tipo de documento asignado”

Tabla 1. Número y porcentajes de tipos de documentos en MAG

Tipos de documentos en MAG	Número de ítems	Porcentaje de ítems
Journal	87.430.385	44,28
None	85.844.335	43,48
Thesis	5,925,439	3,00
Conference	5,053,232	2,56
Repository	4,779,269	2,42
Book	4,588,285	2,32
BookChapter	3,691,552	1,87
Dataset	132,544	0,07
Suma	197.445.041	100

Tabla 2. Número y porcentajes de tipos de documentos en OpenAlex

Tipos de documentos en OpenAlex	Número de ítems	Porcentaje de ítems
journal-article	96.547.138	48,61
none	70.155.602	35,32
book-chapter	9.588.895	4,83
proceedings-article	7.051.207	3,55
dissertation	6.126.640	3,08
book	4.522.989	2,28
posted-content	3.093.874	1,56
report	464.164	0,23
dataset	276.311	0,14
monograph	212.401	0,11
otros tipos	566.944	0,29
Suma	198.606.165	100

### Tipos de documentos

En *MAG* tratamos con siete tipos de documentos: *Book*, *BookChapter*, *Conference*, *Dataset*, *Journal*, *Repository* y *Thesis*. La tabla 1 enumera los números y porcentajes de sus ocurrencias. Casi el 45% de los documentos se clasifican como *Journal*, pero casi la misma cantidad de documentos *no* tiene ningún tipo de documento asignado (*None*).

En *OpenAlex* hay 26 tipos de documentos que heredan su definición de otra fuente de datos importante, *Crossref*, como se documenta en la *Crossref's content type markup guide* (*Crossref*, 2021). Obviamente, todos los trabajos en *OpenAlex* con un DOI de *Crossref* reciben su tipo de documento desde allí. Los tipos de documentos con una presencia de más del 0,1% del total se enumeran en la tabla 2. Hay nueve millones de ítems adicionales en *OpenAlex* asignados al tipo *journal-article* en comparación con *Journal* de *MAG*. Los ítems de *OpenAlex* del tipo *journal-article* cubren casi la mitad de todos los documentos, pero los ítems sin tipo de documento (*none*) siguen siendo más de un tercio del total. Sin embargo, los tipos de documento *Journal* y *None* aparecen con la misma frecuencia en *MAG*. El aumento del número de artículos de revista, actas de congreso y capítulos de libro es especialmente interesante desde el punto de vista bibliométrico.

Como se muestra en la tabla 3, alrededor del 90,1% de todos los ítems tienen tipos de documentos obviamente equivalentes en ambas bases de datos.

Tabla 3. Proporciones de transferencias de tipos de documentos equivalentes entre *MAG* y *OpenAlex*

MAG	OpenAlex	Número de ítems	Porcentaje de ítems
Journal	journal-article	86.395.430	43,76
None	none	70.154.418	35,53
Thesis	dissertation	5.917.802	3,00
Book	book	4.421.867	2,24
Conference	proceedings-article	4.285.360	2,17
BookChapter	book-chapter	3.662.705	1,86
Repository	posted-content	3.018.186	1,53
Dataset	dataset	132.421	0,07
Suma		177.988.189	90,15

Los casos más interesantes son las reclasificaciones. En la figura 2 mostramos un diagrama aluvial de los tipos de documentos correspondientes en ambas bases de datos, *excluyendo* las transferencias de la tabla 3.

Esas reclasificaciones que ocurren en números importantes (suman casi el 9,3% del total de documentos), se enumeran en la tabla 4. Para tener una idea de la calidad de estas reclasificaciones, hemos añadido algunas características de muestras aleatorias respectivas de diez documentos cada una. Todos tenían un DOI, como era de esperar debido a que *Crossref* es la principal fuente de información del tipo de documento. De hecho, se han reclasificado menos de 10.000 documentos *sin* DOI, es decir, alrededor del 0,05% de los 20 millones de reclasificaciones.

“ Cuando el volumen o el número estaban disponibles en *Microsoft Academic Graph*, estos datos se transfirieron por completo a *OpenAlex* ”

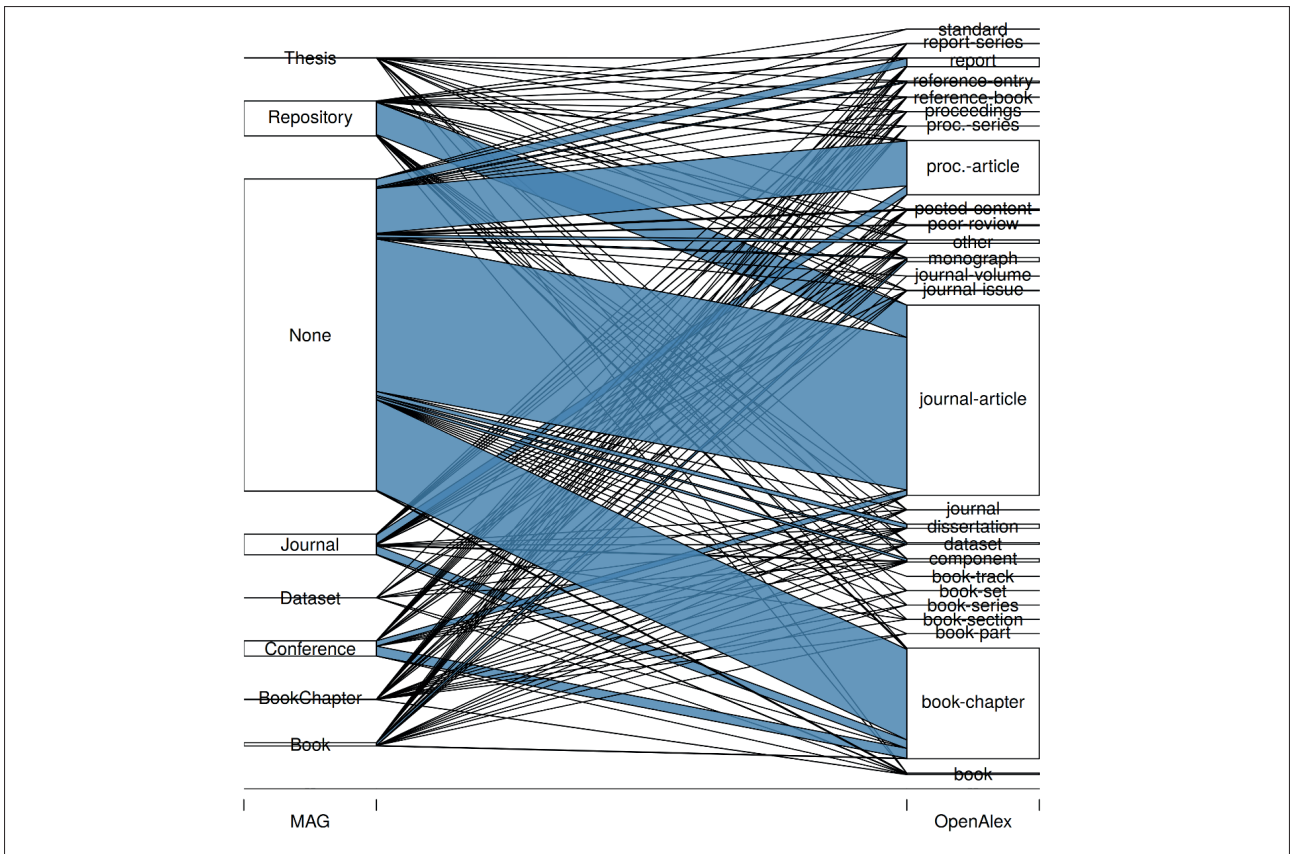


Figura 2. Diagrama aluvial de reclasificaciones de tipos de documentos de MAG a OpenAlex

Tabla 4. Proporciones de reclasificaciones de tipos de documentos de MAG a OpenAlex junto con algunas características de las muestras aleatorias correspondientes de diez documentos. Se muestran las que al menos representan un 0,1%.

Tipos de documentos			Muestras aleatorias de diez documentos	
MAG	OpenAlex	Porcentaje de todos los documentos	Años de publicación	Otras características
None	journal-article	3,88%	1928 - 2018	Ocho títulos con caracteres cirílicos, orientales o árabes; un documento holandés con título en inglés
None	book-chapter	2,30%	1984 - 2020	Todos los DOIs que contienen el prefijo Bookland "978"; un título alemán
None	proceedings-article	1,14%	1971 - 2019	Siete títulos cirílicos o árabes; solo dos documentos de conferencia identificables
Repository	journal-article	0,82%	1988 - 2016	Cuatro ChemInform Abstracts; cinco artículos de arXiv; todos los DOIs apuntan a artículos publicados; un preprint de SSRN de 2012, publicado en 2016 en una revista
Conference	book-chapter	0,25%	2001 - 2020	Todos publicados en actas de conferencias por Springer como parte de una serie de libros; ocho DOIs contienen el prefijo Bookland "978"; siete documentos de LNCS; solo un documento señalado por Springer como capítulo, los demás como documentos de conferencia
Journal	proceedings-article	0,23%	2010 - 2017	Tres resúmenes de presentación de póster en el suplemento de una revista; cuatro documentos de Proceedings of SPIE; dos documentos en las actas de una conferencia médica como suplemento a una revista.
Journal	book-chapter	0,22%	1965 - 2017	Todos los capítulos de libros; ocho DOIs contienen el prefijo Bookland "978"
Conference	journal-article	0,14%	1987 - 2014	No hay papers de conferencia; cuatro editores incorrectos en MAG
None	report	0,20%	1964 - 2019	Siete informes técnicos o datos de estudios geológicos de los gobiernos de EUA y Canadá
None	dissertation	0,10%	1973 - 2018	Tesis y disertaciones en repositorios institucionales (cinco estadounidenses, cuatro brasileños, uno griego)

La reclasificación a tipo *book-chapter* en *OpenAlex* parece funcionar bastante bien. Este es también el caso de *journal-article*. En particular, muchos documentos que usan conjuntos de caracteres no latinos ahora están clasificados, y una cantidad considerable de ítems con DOIs que *Microsoft Academic Graph* había etiquetado como preprints de *arXiv* con tipo de documento *Repository* se reconocen correctamente como *journal-article*. Por otro lado, la asignación de abstracts de *ChemInform* a este tipo de documento es discutible, pero definitivamente no son preprints. Las comunicaciones de conferencias parecen ser un caso especial: las asignadas incorrectamente a *Journal* se corrigen a *proceedings-article*, pero para los documentos sin un tipo de documento en *MAG*, la asignación de *proceedings-article* no es tan precisa o al menos difícil de verificar. En el caso del tipo *Conference* de *MAG*, la reclasificación a *journal-article* parece ser correcta en general, mientras que la reclasificación de las contribuciones de *Lecture notes in computer science (LNCS)* a *book-chapter* parece ser el resultado de su aparición como parte de una serie de libros y del formato de sus DOIs que contienen el prefijo Bookland "978" (DOI.org, 2019). Este hecho debe tenerse en cuenta para los estudios bibliométricos en ciencias de la computación, que probablemente también deberían incluir capítulos de libros.

La reclasificación a tipo *book-chapter* en *OpenAlex* parece funcionar bastante bien. Este es también el caso de *journal-article*. En particular, muchos documentos que usan conjuntos de caracteres no latinos están ahora clasificados, y una cantidad considerable de ítems con DOIs y el tipo de documento *Repository* en *Microsoft Academic Graph* se reconocen correctamente como *journal-article*

### Clasificaciones de materias

*OpenAlex* afirma en su guía de migración (*OpenAlex*, 2021) que utilizan la misma taxonomía que *MAG* pero han reducido la cantidad de campos de estudio ("Fields of Study", FoS) al eliminar los que tienen menos de 500 documentos asociados. Además, aplicaron un algoritmo diferente, es decir, el modelo *V1*, que usaba títulos de artículos y algunas otras características, pero no datos de los abstracts. Estos últimos solo se usaron más adelante en 2022 con la implementación del modelo *V2* de su software de código abierto (*Priem; Piwowar*, 2022).

Una mirada rápida revela la persistencia de los 19 FoS de nivel superior (nivel = 0) de *MAG*, así como de 284 de los 292 FoS del siguiente nivel (nivel = 1). La tabla 5 enumera la distribución de todos los niveles de FoS del 0 al 5 en ambas bases de datos. La reducción más fuerte de los números de FoS ocurre en los niveles 3 a 5 donde permanece menos del 10%. Sin embargo, en ambas bases de datos, *MAG* y *OpenAlex*, la granularidad no necesariamente aumenta con el nivel de FoS. El nivel 3 tiene el mayor número de FoS. El número total de FoS en todos los niveles es de 714.971 en *MAG* y solo de 65.073 en *OpenAlex*, lo que supone una reducción del 9,1%. Curiosamente, en los niveles 2 a 5, un número sustancial de FoS tienen menos de 500 trabajos asignados, por ejemplo, más de 4000 FoS en los niveles 2 y 3, respectivamente.

Tabla 5. Distribución de FoS en *MAG* y *OpenAlex*

Nivel	#MAG	#OpenAlex	Diferencia (#MAG - #OpenAlex)	Porcentaje (#OpenAlex / #MAG *100)
0	19	19	0	100,00
1	292	284	8	97,26
2	137.415	21.460	115.955	15,62
3	330.275	24.768	305.507	7,50
4	134.843	12.406	122.437	9,20
5	112.127	6.136	105.991	5,47
Todos los niveles	714.971	65.073	649.898	9,10

### Campos de estudio de nivel superior

Incluso si los FoS de nivel superior persisten, están asociados de manera muy diferente a los documentos. Por ejemplo, el paper: <https://api.openalex.org/works/W2178938397>

(consultado el 26 de abril de 2022) tenía un FoS de nivel superior y un FoS de nivel 1 en *MAG*, pero tiene seis FoS de nivel superior adicionales y un FoS de nivel 1 adicional en *OpenAlex*.

La tabla 6 muestra algunas medidas estadísticas del conjunto de publicaciones comunes con respecto al número de artículos con un FoS asignado. El número total de documentos con un FoS aumenta significativamente: 30,5 de 48,9 millones de documentos sin ningún FoS en *MAG* tienen al menos un FoS en *OpenAlex* –28,8 millones tienen un FoS de nivel superior en *OpenAlex*– por lo que la cobertura aumenta del 74,6% al 86,6%. El número de asignaciones por artículo a un FoS de nivel superior aumenta drásticamente en *OpenAlex*: alrededor de 147 millones de artículos en *MAG* y alrededor de 171 millones de artículos en *OpenAlex* tienen asignado al menos un FoS de nivel superior. De esos documentos, 65.000 en *MAG* y 53 millones en *OpenAlex* tienen múltiples FoS de nivel superior (hasta siete) asignados. Por lo tanto, al aplicar el algoritmo de *OpenAlex*, proliferó la asignación múltiple a FoS de nivel superior.

Tabla 6. Números y porcentajes de documentos con FoS asignado

Medidas estadísticas del conjunto de publicaciones comunes	MAG	OpenAlex
Número de documentos	197.444.264	197.444.264
Número de asignaciones a algún FoS	1.092.748.572	1.095.801.888
Número de documentos con algún FoS	148.518.539	175.993.558
Número de documentos sin FoS	48.925.725	21.450.706
Cobertura de documentos con algún FoS	75,22%	89,14%
Asignaciones medias a algún FoS por documento	7,36	6,23
Número de asignaciones a un FoS de nivel superior	147.426.219	229.560.450
Número de documentos con un FoS de nivel superior	147.360.860	170.900.225
Cobertura de documentos con un FoS de nivel superior	74,63%	86,56%
Número de documentos con múltiples asignaciones a FoS de nivel superior	65.359	52.966.153
Porcentaje de documentos con múltiples asignaciones a FoS de nivel superior de todos los documentos con asignaciones de FoS	0,04%	30,99%
Asignaciones medias a un FoS de nivel superior por documento	1,000444	1,343243

La tabla 7 compara el número de asignaciones a FoS de nivel superior en ambas bases de datos. El mayor incremento relativo se da en Art y el mayor descenso en Engineering.

Tabla 7. Comparación del número de asignaciones de FoS de nivel superior en MAG y OpenAlex

FoS	#MAG	% MAG	#OpenAlex	% OpenAlex	% Open Alex / % MAG
Art	3.717.975	2,52	12.873.508	5,61	2,22
Biology	13.169.649	8,93	14.242.938	6,20	0,69
Business	5.174.422	3,51	12.010.241	5,23	1,49
Chemistry	14.191.693	9,63	20.029.716	8,73	0,91
Computer science	12.312.525	8,35	25.678.965	11,19	1,34
Economics	3.130.346	2,12	4.064.798	1,77	0,83
Engineering	8.472.749	5,75	3.117.013	1,36	0,24
Environmental science	3.533.640	2,40	6.702.031	2,92	1,22
Geography	4.447.923	3,02	7.053.608	3,07	1,02
Geology	3.061.102	2,08	3.621.249	1,58	0,76
History	3.059.007	2,07	4.454.112	1,94	0,94
Materials science	11.063.791	7,50	17.437.416	7,60	1,01
Mathematics	6.021.856	4,08	6.101.501	2,66	0,65
Medicine	27.897.600	18,92	36.085.634	15,72	0,83
Philosophy	2.010.846	1,36	5.916.152	2,58	1,89
Physics	6.873.294	4,66	9.938.209	4,33	0,93
Political science	6.775.718	4,60	17.760.011	7,74	1,68
Psychology	8.063.945	5,47	13.966.595	6,08	1,11
Sociology	4.448.138	3,02	8.506.753	3,71	1,23
Todas las asignaciones	147.426.219		229.560.450		

De los casi 49 millones de documentos sin ningún FoS en MAG, casi 29 millones recibieron al menos un FoS de nivel superior en OpenAlex. La tabla 8 muestra la distribución de tipos de documentos en MAG en ambos conjuntos. Alrededor del 80% de los documentos sin ningún FoS no tienen ningún tipo de documento y la mitad de ellos reciben un FoS de nivel superior en OpenAlex que asciende a dos tercios de todos estos documentos. Teniendo en cuenta los tipos de documentos bibliométricamente más interesantes *Journal*, *BookChapter*, *Book* y *Conference*, aproximadamente el 90% tiene un FoS de nivel superior en OpenAlex, que representa una cuarta parte de todos esos documentos.

“ En OpenAlex, el número total de trabajos con un campo de estudio aumenta significativamente. Igualmente, también aumenta mucho el número de asignaciones por trabajo a un campo de estudio de nivel superior ”



Tabla 8. Distribución de tipos de documentos en MAG entre los documentos sin ningún FoS en MAG y proporción de documentos con algún FoS de nivel superior en *OpenAlex*

Tipo de documento en MAG	#MAG sin FoS	#MAG sin FoS, pero FoS de nivel superior en <i>OpenAlex</i>	Porcentaje de trabajos con FoS de nivel superior en <i>OpenAlex</i> pero sin FoS en MAG
None	38.273.234	19.266.767	50,34
Journal	5.102.461	4.566.743	89,50
Thesis	2.453.810	2.249.289	91,67
BookChapter	1.416.913	1.235.266	87,18
Book	1.356.096	1.241.835	91,57
Repository	251.384	228.099	90,74
Conference	58.961	54.332	92,15
Dataset	12.866	9.561	74,31
Suma resp. promedio	48.925.725	28.851.892	58,97

Alrededor del 77,2% de todas las asignaciones de nivel superior en MAG persisten en *OpenAlex*, pero esta proporción varía significativamente entre los 19 FoS de nivel superior, como muestra claramente la tabla 9: desde menos de una cuarta parte para Ingeniería hasta más del 90 % para Ciencias de los materiales y Medicina.

Tabla 9. Distribución de FoS de nivel superior en MAG, y número y porcentaje de FoS de nivel superior persistentes en *OpenAlex*

FoS	#MAG	#OpenAlex	% persistencia
Art	3.717.975	2.620.365	70,48
Biology	13.169.649	10.411.044	79,05
Business	5.174.422	4.200.803	81,18
Chemistry	14.191.693	12.194.451	85,93
Computer science	12.312.525	10.878.013	88,35
Economics	3.130.346	2.131.877	68,10
Engineering	8.472.749	2.023.815	23,89
Environmental science	3.533.640	2.712.884	76,77
Geography	4.447.923	2.366.289	53,20
Geology	3.061.102	2.302.537	75,22
History	3.059.007	1.650.999	53,97
Materials science	11.063.791	10.010.937	90,48
Mathematics	6.021.856	4.028.415	66,90
Medicine	27.897.600	25.953.084	93,03
Philosophy	2.010.846	1.240.834	61,71
Physics	6.873.294	5.517.376	80,27
Political science	6.775.718	4.899.049	72,30
Psychology	8.063.945	6.198.019	76,86
Sociology	4.448.138	2.520.233	56,66
Todas las asignaciones	147.426.219	113.861.024	77,23

La figura 3 muestra un gráfico aluvial de la transferencia de clasificaciones de materias *sin* las asignaciones persistentes de FoS de la tabla 9 para que las reclasificaciones restantes se vuelvan más visibles. Dado el hecho de que *todas* las 342 posibles reclasificaciones efectivamente ocurren en nuestro conjunto de publicaciones, solo se muestran las 94 conexiones con al menos 200.000 ocurrencias. Varias reclasificaciones ocurren en medidas comparables en ambas direcciones, por ejemplo, en los pares Sociology y Psychology, Sociology y Political Science, o Psychology y Medicine. Otros muestran una transferencia significativa principalmente en una dirección, como de Engineering a Computer Science, de Mathematics a Computer Science, de Biology a Chemistry, o de Chemistry a Materials Science.

Como ejemplo, la distribución de las asignaciones de los documentos con FoS Engineering de nivel superior en MAG a FoS de nivel superior en *OpenAlex* se muestra en la tabla 10. Las dos columnas de la derecha indican la proporción de las asignaciones solo al respectivo FoS.

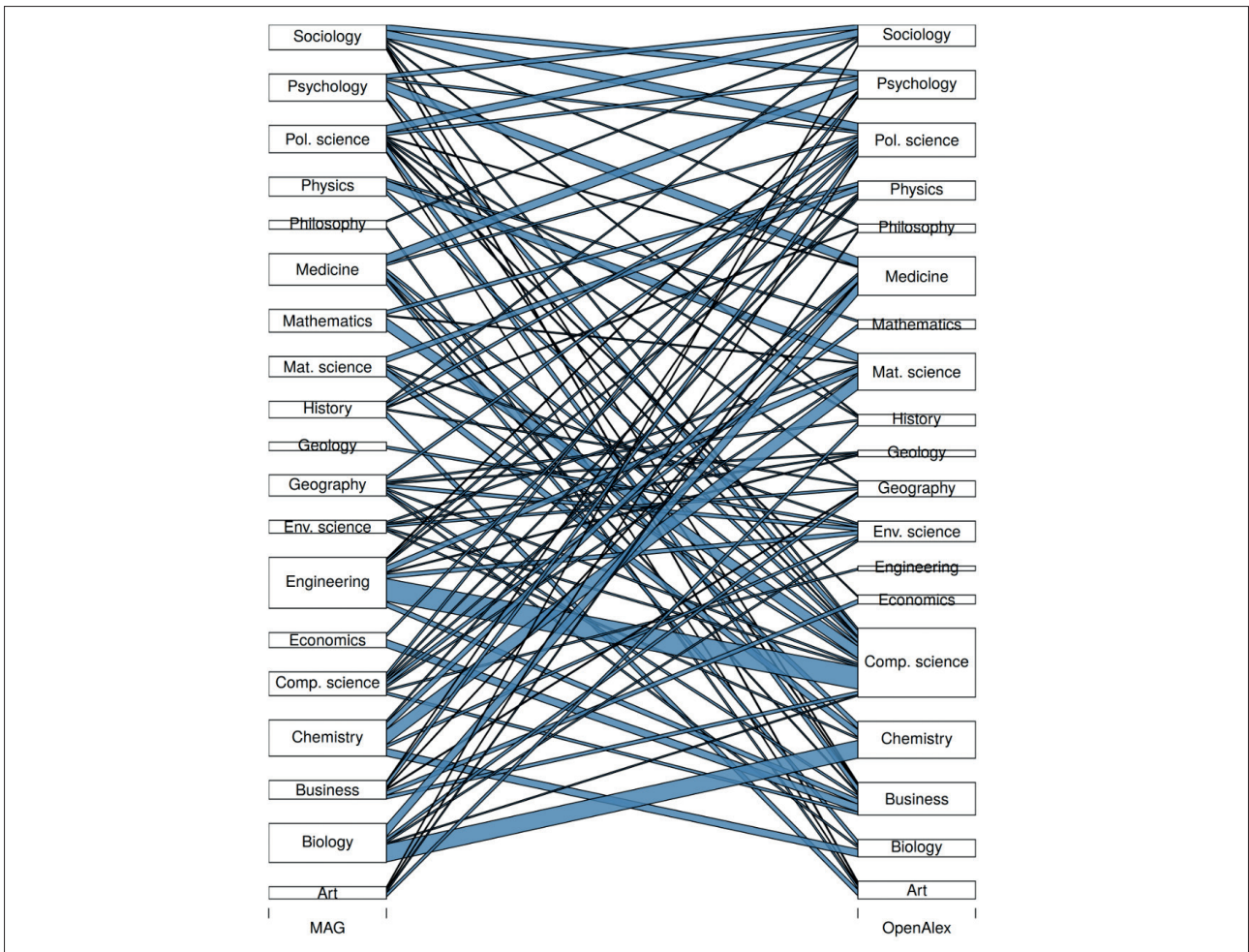


Figura 3. Diagrama aluvial para las reclasificaciones de FoS de nivel superior de *MAG* a *OpenAlex*, que muestra solo las reclasificaciones que ocurren al menos 200.000 veces

Tabla 10. Distribución de asignaciones de los 8.472.749 documentos con FoS Engineering de nivel superior en *MAG* y sus FoS de nivel superior en *OpenAlex*

FoS en <i>OpenAlex</i>	Número	Porcentaje	Número de trabajos que se asignan solo a este FoS	Porcentaje de trabajos que se asignan solo a este FoS
Art	155.994	1,84	83.922	53,80
Biology	37.572	0,44	21.505	57,24
Business	890.319	10,51	371.279	41,70
Chemistry	117.583	1,39	47.157	40,11
Computer science	3.880.097	45,80	2.199.873	56,70
Economics	52.304	0,62	10.490	20,06
Engineering	2.023.815	23,89	649.984	32,12
Environmental science	715.720	8,45	277.495	38,77
Geography	121.729	1,44	46.386	38,11
Geology	222.817	2,63	98.828	44,35
History	66.291	0,78	16.244	24,50
Materials science	1.028.047	12,13	524.787	51,05
Mathematics	122.422	1,44	20.450	16,70
Medicine	175.371	2,07	86.643	49,41
Philosophy	22.304	0,26	7.247	32,49
Physics	310.876	3,67	90.731	29,19
Political science	412.799	4,87	155.831	37,75
Psychology	245.947	2,90	99.949	40,64
Sociology	166.826	1,97	44.179	26,48
Suma/Promedio	10.768.833	127,10	4.852.980	45,07

Solo casi una cuarta parte de los 8.472.749 documentos que se asignan a Engineering en MAG se asignan al mismo FoS en OpenAlex, pero casi la mitad de ellos se asignan al FoS Computer science. Hay una cantidad significativa de asignaciones múltiples en OpenAlex, por lo que el número total de asignaciones de FoS aumenta en más de una cuarta parte a 10,768,833.

La tabla 11 muestra las 10 listas más frecuentes de asignaciones de FoS. Los FoS individuales están separadas por un punto y coma. Curiosamente, la combinación Computer science; Engineering ocupa el segundo lugar y, junto con Engineering en el tercer lugar, representa más del 75 % de las asignaciones a Engineering en OpenAlex.

Tabla 11. Las 10 listas más frecuentes de asignaciones de FoS de nivel superior en *OpenAlex* para los 8.472.749 documentos asignados a Engineering en MAG

Asignaciones de FoS de nivel superior más frecuentes en <i>OpenAlex</i>	Número de trabajos con asignaciones FoS de nivel superior más frecuentes en <i>OpenAlex</i>	Porcentaje de trabajos con asignaciones FoS de nivel superior más frecuentes en <i>OpenAlex</i>
Computer science	2.199.873	25,16
Computer science; Engineering	827.861	9,47
Engineering	649.984	7,43
Materials science	524.787	6,00
Business	371.279	4,25
Environmental science	277.495	3,17
Political science	155.831	1,78
Business; Computer science	151.568	1,73
Computer science; Materials science	149.538	1,71
Engineering; Environmental science	101.786	1,16

Para tener una idea de la validez de las reclasificaciones (automáticas) de materias de MAG a *OpenAlex*, observamos los casos extremos relacionados con la proliferación de FoS de nivel superior, como se indica en la tabla 7. Tomamos muestras aleatorias de esos documentos en el conjunto de datos comunes que tienen un FoS Art resp. Engineering en MAG. Las restricciones adicionales fueron el año de publicación 2020, el tipo de documento *Journal* y la disponibilidad de un DOI. Finalmente, solo elegimos documentos que recibieron un FoS único de nivel superior en *OpenAlex*. Recuperamos los documentos e intentamos evaluar la idoneidad de sus asignaciones de FoS en ambas bases de datos, indicadas por una puntuación numérica de 1 (correcto), -1 (no correcto) y 0 (posiblemente plausible). Las tablas 12 y 13 muestran los resultados respectivos, incluida la información de contenido que condujo a nuestra evaluación y un puntaje de idoneidad resumido para cada base de datos. Por supuesto, el método aplicado solo puede producir resultados preliminares con poca precisión, pero podrían dar una pista para futuras investigaciones.

Tabla 12. Evaluación de la idoneidad de asignaciones únicas de FoS de nivel superior para una muestra aleatoria de los 3.717.975 documentos asignados a Art en MAG

FoS en <i>OpenAlex</i>	DOI del documento de muestra	Información de contenido	Puntuación de idoneidad de los FoS de MAG	Puntuación de idoneidad de los FoS de <i>OpenAlex</i>
Art	10.25162/afmw-2020-0005	Musicología	1	1
Biology	10.25223/brad.n38.2020.a10	Botánica	-1	1
Business	10.47287/cen-09844-newscrits	Noticias sobre química e ingeniería: productos de consumo; Guía de regalos navideños 2020	-1	1
Chemistry	10.4312/keria.22.1.143-202	Poesía latina	1	-1
Computer science	10.1016/s1634-7358(20)44295-0	Endocrinología (Medicina)	-1	-1
Economics	10.33008/ijcmr.202019	Instalación performativa creativa para una película.	1	-1
Engineering	10.1061/(asce)gm.1943-5622.0001816	Editorial en <i>Journal of Geomechanics</i>	-1	1
Environmental science	10.1002/awwa.1472	Obituario en <i>Journal American Water Works Association</i>	0	1
Geography	10.4000/geomorphologie.14981	Arqueología en número especial sobre "geomorphologie et environnements karstiques"	0	1
Geology	10.1180/mgm.2020.16	Obituario en <i>Mineralogical Magazine</i>	1	1
History	10.18223/hiscult.v9i1.3154	Historia y cultura	0	1

FoS en OpenAlex	DOI del documento de muestra	Información de contenido	Puntuación de idoneidad de los FoS de MAG	Puntuación de idoneidad de los FoS de OpenAlex
Materials science	10.33383/2020-006	"Recomendaciones para la restauración de revestimientos transparentes históricos en el Museo Pushkin"	-1	1
Mathematics	10.51481/amc.v56i1.823	Obituario en <i>Acta Médica Costarricense</i> (Biomedicina)	1	-1
Medicine	10.1136/bmj.m3665	Obituario en <i>BMJ</i>	1	1
Philosophy	10.13125/medea-4529	Poesía Italiana (Estudios Culturales)	1	1
Physics	10.1038/d41586-020-02136-4	Foto del sol (Noticias en <i>Nature</i> )	-1	1
Political science	10.1080/10402659.2020.1823575	<i>Peace review: A journal of social justice</i>	1	1
Psychology	10.1037/amp0000602	Obituario de un psicólogo	1	1
Sociology	10.35293/srsa.v37i1.229	Reseña de libro sobre la historia reciente de Sudáfrica con dimensión socioeconómica	-1	0
Suma de puntajes de idoneidad			2	10

Tabla 13. Evaluación de la idoneidad de las asignaciones únicas de FoS de nivel superior para una muestra aleatoria de los 8.472.749 documentos asignados a Ingeniería en MAG

FoS en OpenAlex	DOI del documento de muestra	información de contenido	de MAG FoS	Puntuación de idoneidad de OpenAlex FoS
Art	10.22452/sare.vol57no2.10	Poesía y ficción	-1	1
Biology	10.1182/blood.2020008691	Células en sangre	-1	1
Business	10.4028/www.scientific.net/amm.896.371	Régimen fiscal de las edificaciones en <i>Applied Mechanics and Materials</i>	1	1
Chemistry	10.1021/cen-09813-scicon10	C&EN (sección Materiales) sobre micro y nanoestructuras diseñadas que imitan a las arañas	1	1
Computer science	10.17577/ijertv8is120305	Estudio comparativo de sistemas de calificación de edificios	1	-1
Engineering	10.1007/s35658-020-0295-y	Autobuses lanzadera automatizados	1	1
Environmental science	10.3130/aije.85.19	Estudio del comportamiento higrótérmico del montaje de muros... impacto de la penetración de lluvia y absorción de agua	1	1
Geography	10.5632/jila.83.539	Desarrollo de 'negocios-utilizando parques urbanos'	-1	1
History	10.1146/annurev-anchem-091119-120456	Evolución de las ciencias analíticas en los Estados Unidos: un relato histórico	-1	1
Materials science	10.1063/1.5145201	Desarrollo de microLED en <i>Appl.phys.lett</i>	1	1
Mathematics	10.1007/s11071-020-05950-7	Obituario en <i>Nonlinear dynamics</i>	1	1
Medicine	10.1055/a-1309-0141	Tecnología dental	1	1
Philosophy	10.1007/s40544-020-0360-9	Obituario en <i>Mechanical engineering</i>	1	-1
Physics	10.1007/s35658-019-0154-x	Evaluación de riesgos sobre propagación térmica	1	1
Political science	10.1038/s41587-020-0702-1	Corrección en <i>Nature biotechnology</i>	1	-1
Psychology	10.12775/jehs.2020.10.05.035	Futuros licenciados en transporte motorizado	1	-1
Sociology	10.4079/gbl.v20.1	Artículo publicado por <i>Global business languages</i>	-1	-1
Suma de puntajes de idoneidad			7	7

Tal vez, los obituarios y editoriales en la muestra de la tabla 12 tienden a clasificarse como Art en *MAG*, pero en *OpenAlex*, el tema científico parece ser más importante. Según el puntaje de idoneidad, para esta muestra las clasificaciones de materias de *OpenAlex* son definitivamente más adecuadas. En el caso de Engineering (consultar la tabla 13), hay muchos más casos que podrían asignarse plausiblemente a dos o más FoS, lo que también se expresa por los puntajes de idoneidad uniformes y relativamente altos para ambas bases de datos.

### Campos de estudio de nivel 1

Los FoS de nivel 1 son aún más interesantes para las evaluaciones y comparaciones bibliométricas porque su número es similar al número de categorías temáticas en *WoS* y *Scopus*, lo que permite una granularidad similar para las normalizaciones por campo. En nuestro conjunto de datos se produjeron 74.454 reclasificaciones, es decir, en promedio alrededor de 255 para cada uno de los 292 FoS en *MAG*. Debido a su número ya no es factible presentar las reclasificaciones en forma de tabla. En la figura 4 se muestran las reclasificaciones más frecuentes de FoS de nivel 1 de *MAG* a *OpenAlex*. En

Tabla 14. Número de documentos en *MAG* de los ocho FoS de nivel 1 que faltan en OpenAlex

Falta FoS de nivel 1 en OpenAlex	#Documentos en MAG
Algebra	398.751
Analytical chemistry	1.989.709
Calculus	205.406
Ceramic materials	725
Control theory	967.322
Hydrology	524.127
Pattern recognition	1.417.305
Topology	277.926

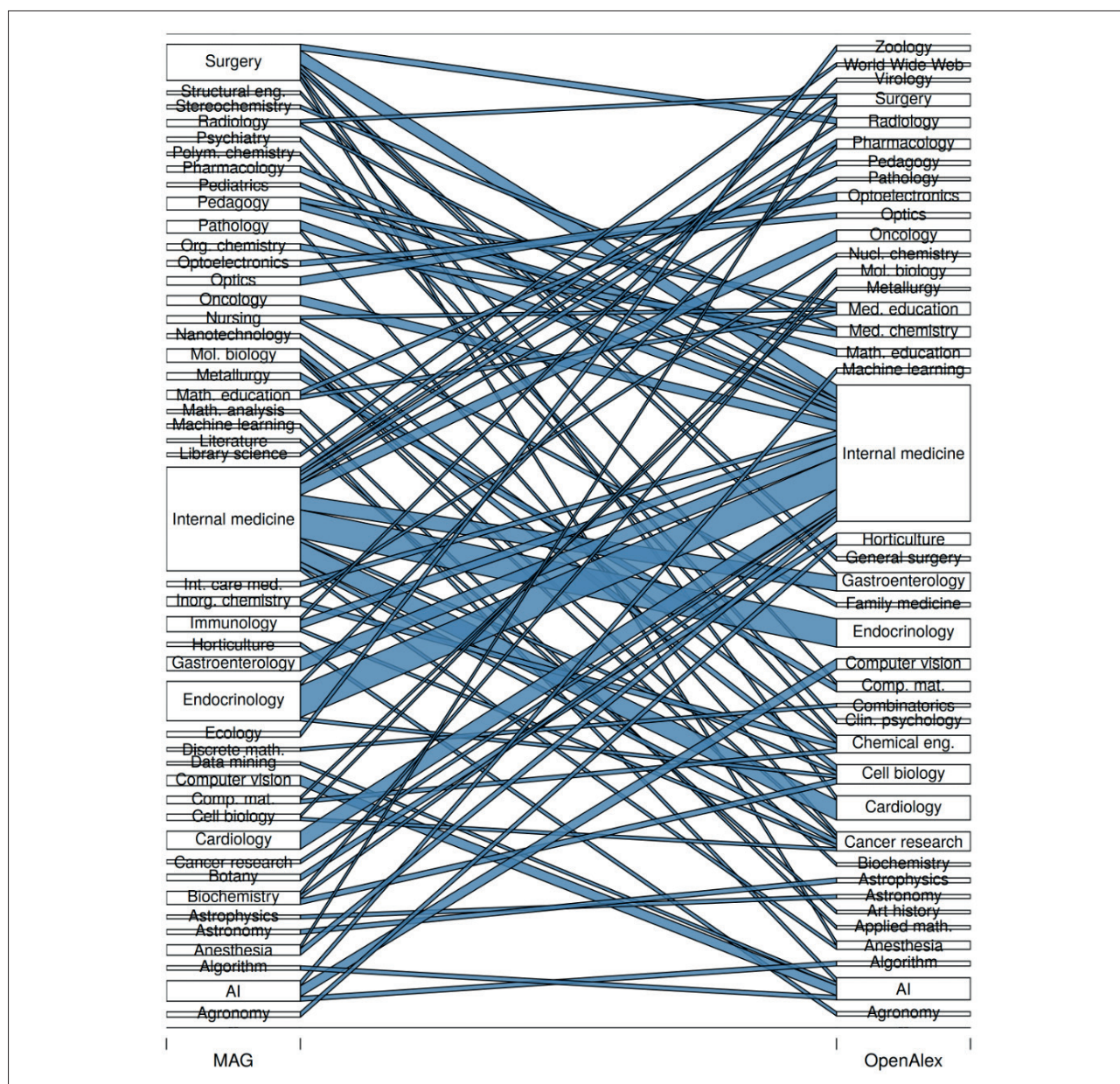


Figura 4. Diagrama aluvial para las reclasificaciones de FoS de nivel 1 de *MAG* a *OpenAlex*, que muestra solo las reclasificaciones que ocurren al menos 250.000 veces. Algunos de los nombres originales de FoS eran bastante largos, por lo que en este gráfico los abreviamos. Los nombres abreviados de FoS con sus nombres originales se pueden encontrar en la tabla A1 en el Apéndice.

este nivel se puede detectar una buena cantidad de simetría. Por ejemplo, en el caso de Internal medicine, las mayores transferencias a otros FoS parecen ocurrir en ambas direcciones. Esta impresión se mantiene incluso si se va a un valor de umbral mucho más pequeño como, por ejemplo, 50.000 reclasificaciones.

Como se mencionó anteriormente, ocho de los 292 FoS de nivel 1 presentes en *MAG* faltan en *OpenAlex*. Sus números de documentos en *MAG* se ofrecen en la tabla 14. Especialmente, la eliminación de los FoS más poblados Analytical chemistry y Pattern recognition con casi 2 resp. 1,5 millones de trabajos en *MAG* es sorprendente.

La figura 5 muestra los números de asignaciones de FoS de nivel 1 en *OpenAlex* que ocurren al menos 30.000 veces. Por lo tanto, Ceramic materials no se muestra. Un primer vistazo revela principalmente reclasificaciones a FoS estrechamente relacionados.

#### 4. Discusión y conclusiones

*OpenAlex* ha transferido prácticamente todos los trabajos de *MAG* conservando sus datos bibliográficos: año de publicación, volumen, primera y última página, DOI y el número de referencias que son ingredientes importantes para el análisis de citas.

Más del 90% de los documentos *MAG* tienen tipos de documentos equivalentes en *OpenAlex*. De los restantes, especialmente las reclasificaciones de los tipos *journal-article* y *book-chapter* de *OpenAlex* parecen ser correctas y ascienden a más del 7%, por lo que las especificaciones de tipo de documento han mejorado significativamente de *MAG* a *OpenAlex*.

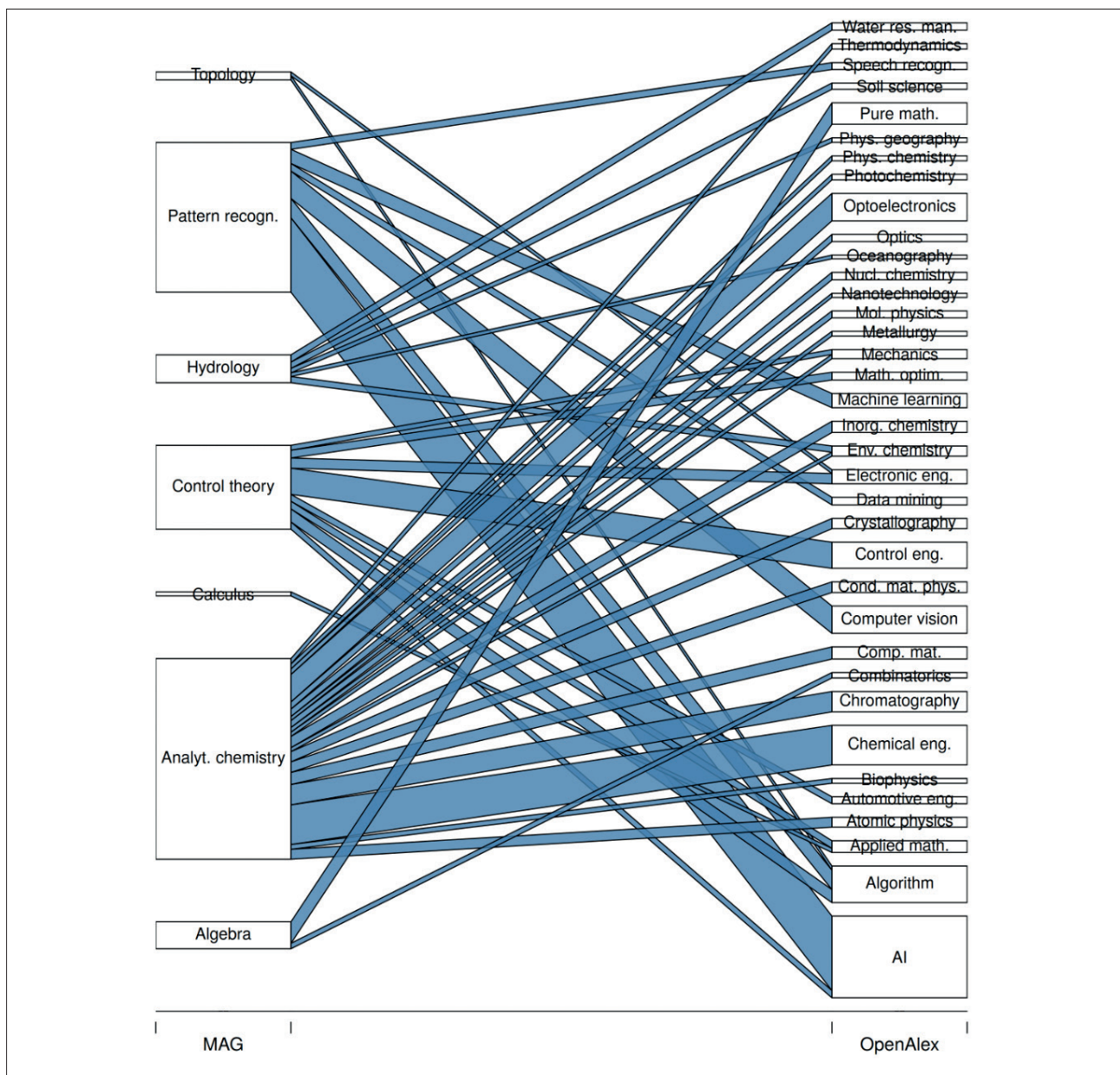


Figura 5. Diagrama aluvial para las reclasificaciones de los ocho FoS de nivel 1 que faltan en *OpenAlex* de *MAG* a *OpenAlex*. Se muestran solo las reclasificaciones que ocurren al menos 30.000 veces. Algunos de los nombres originales de FoS eran bastante largos, por lo que se han abreviado. Los nombres abreviados de FoS con sus nombres originales se pueden encontrar en la tabla A1 en el Apéndice.

Hasta ahora, *OpenAlex* parece ser más adecuado para análisis bibliométricos que *MAG*.

Como último elemento importante de metadatos bibliométricos, observamos la clasificación de materias a campos de estudio ("Fields of Study", FoS) en *MAG* y en *OpenAlex*. Encontramos significativamente más documentos con una asignación de FoS en *OpenAlex* que en *MAG*. En el nivel superior y en el nivel 1, la estructura de FoS es idéntica o casi idéntica, pero en los niveles más profundos, el número de FoS disponibles se reduce drásticamente a alrededor del 10%. Una característica llamativa en el nivel superior es la proliferación y abundante reclasificación de los 19 FoS, muy desigualmente distribuidos entre ellos. Este también es el caso de una muestra aleatoria utilizada para evaluar la idoneidad de las asignaciones de FoS en ambas bases de datos. En el nivel 1, las reclasificaciones de FoS parecen ser mucho más simétricas y los ocho FoS que faltan se distribuyen a otros estrechamente relacionados, de modo que el efecto neto sobre las conclusiones extraídas de análisis bibliométricos anteriores utilizando FoS de nivel 1, como por **Scheidsteger et al.** (2018), podría ser pequeño. Pero eso aún debe investigarse, así como las consecuencias del cambio de *OurResearch* al modelo V2, un algoritmo de clasificación de materias diferente, durante el año 2022, que prometía traer una mejora sustancial (**Priem; Piwowar**, 2022).

Ocho campos de estudio (Algebra, Analytical chemistry, Calculus, Ceramic materials, Control theory, Hydrology, Pattern recognition, Topology) de los 292 presentes en *Microsoft Academic Graph* faltan en *OpenAlex*

En general, *OpenAlex* parece ser al menos tan adecuado para los análisis bibliométricos como *MAG* para las publicaciones de antes de 2021. Sin embargo, esta primera impresión debe verificarse mediante estudios más detallados.

## 5. Referencias

**Bojanowski, Michał; Edwards, Robin** (2016). *Alluvial: R package for creating alluvial diagrams*. R package version: 0.1-2. <https://github.com/mbojan/alluvial>

**Crossref** (2021). *Content type markup guide*.

<https://www.crossref.org/documentation/content-registration/content-type-markup-guide>

**DOI.org** (2019). *DOI System and the ISBN System*.

<https://www.doi.org/factsheets/ISBN-A.html>

**Harzing, Anne-Wil; Alakangas, Satu** (2017). "Microsoft Academic: is the phoenix getting wings?". *Scientometrics*, v. 110, n. 1, pp. 371-383.

<https://doi.org/10.1007/s11192-016-2185-x>

**Kramer, Bianca** (2022). *COKI Open metadata report* (Update March 25, 2022).

<https://github.com/Curtin-Open-Knowledge-Initiative/open-metadata-report>

**Microsoft Blog** (2021). *Microsoft Academic*.

<https://www.microsoft.com/en-us/research/project/academic>

**OpenAlex** (2021). *Migration guide*.

<https://docs.openalex.org/download-snapshot/mag-format/mag-migration-guide>

**OurResearch** (2021). *We're building a replacement for Microsoft Academic Graph*.

<https://blog.ourresearch.org/were-building-a-replacement-for-microsoft-academic-graph>

**Priem, Jason; Piwowar, Heather** (2022). *OpenAlex-concept-tagging*.

<https://github.com/ourresearch/openalex-concept-tagging>

**Priem, Jason; Piwowar, Heather; Orr, Richard** (2022). "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts". In: 26<sup>th</sup> *International conference on science, technology and innovation indicators (STI 2022)*, Granada, Spain.

<https://doi.org/10.5281/zenodo.6936227>

**R Core Team** (2020). *R: A language and environment for statistical computing*.

<https://www.R-project.org>

**Scheidsteger, Thomas; Haunschild, Robin** (2022). "Comparison of metadata with relevance for bibliometrics between Microsoft Academic Graph and OpenAlex until 2020". In: 26<sup>th</sup> *International conference on science, technology and innovation indicators (STI 2022)*, Granada, Spain.

<https://doi.org/10.5281/zenodo.6975102>

**Scheidsteger, Thomas; Haunschild, Robin; Hug, Sven E.; Bornmann, Lutz** (2018). "The concordance of field-normalized scores based on Web of Science and Microsoft Academic data: A case study in computer sciences". In: 23<sup>rd</sup> *International conference on science, technology and innovation indicators (STI 2018)*, Leiden, The Netherlands.

<https://hdl.handle.net/1887/65358>

**Sinha, Arnab; Shen, Zhihong; Song, Yang; Ma, Hao; Eide, Darrin; Hsu, Bo-June-Paul; Wang, Kuansan** (2015). "An overview of Microsoft Academic Service (MAS) and applications". In: *24<sup>th</sup> International conference on World Wide Web (WWW'15 Companion)*, Florence, Italy.

<https://doi.org/10.1145/2740908.2742839>

**Visser, Martijn; Van-Eck, Nees-Jan; Waltman, Ludo** (2021). "Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic". *Quantitative science studies*, v. 2, n. 1, pp. 20-41.

[https://www.doi.org/10.1162/qss\\_a\\_00112](https://www.doi.org/10.1162/qss_a_00112)

**Wickham, Hadley** (2016). *ggplot2: Elegant graphics for data analysis*: New York: Springer-Verlag. ISBN: 978 3 319 24277 4

## 6. Apéndice

**Tabla A1. Lista de abreviaturas de los nombres de campos de estudio ("Fields of Study", FoS) usados en los diagramas aluviales de las figuras 4 y 5**

Nombre abreviado del FoS	Nombre original del FoS
AI	Artificial Intelligence
Analyt. chemistry	Analytical chemistry
Applied math.	Applied mathematics
Automotive eng.	Automotive engineering
Chemical eng.	Chemical engineering
Clin. psychology	Clinical psychology
Comp. mat.	Composite material
Cond. mat. phys.	Condensed matter physics
Control eng.	Control engineering
Discrete math.	Discrete mathematics
Electronic eng.	Electronic engineering
Env. chemistry	Environmental chemistry
Inorg. chemistry	Inorganic chemistry
Int. care med.	Intensive care medicine
Math. analysis	Mathematical analysis
Math. education	Mathematical education
Math. optim.	Mathematical optimization
Med. chemistry	Medicinal chemistry
Med. education	Medical education
Mol. biology	Molecular biology
Mol. physics	Molecular physics
Nucl. chemistry	Nuclear chemistry
Org. chemistry	Organic chemistry
Pattern recogn.	Pattern recognition
Phys. chemistry	Physical chemistry
Phys. geography	Physical geography
Polym. chemistry	Polymer chemistry
Pure math.	Pure mathematics
Speech recogn.	Speech recognition
Structural eng.	Structural engineering
Water res. man.	Water resource management