

El problema de las bases de datos en la evaluación de la producción científica en América Latina. Una exploración del potencial de OpenAlex para las universidades argentinas¹²³

Recibido: 07/01/2024

Aprobado: 05/06/2024

Oswaldo Gallardo

CONICET, Centro de Estudios de la Circulación del Conocimiento—Universidad Nacional de Cuyo, Argentina

<https://orcid.org/0000-0003-0662-2196>

ogallardo@conicet.gov.ar

Manuel Bruccoleri Ochoa

Centro de Estudios de la Circulación del Conocimiento—Universidad Nacional de Cuyo, Argentina

manubrullo97@gmail.com

<https://orcid.org/0009-0006-7708-7943>

Resumen:

En este artículo se analiza la cobertura de OpenAlex y Scopus de las publicaciones científicas de autores/as de Argentina, tanto para el país completo como para un conjunto de 10 universidades nacionales seleccionadas, en el período 2019-2023. El objetivo es comparar el desempeño de ambas bases de datos en cuanto a volumen global de publicaciones indexadas, idiomas y áreas disciplinares. Los resultados arrojan que la cobertura de OpenAlex es más amplia y diversa, en particular para las publicaciones en español, en ciencias sociales y humanidades y con un único autor o autora. También se observa que la cobertura de OpenAlex es particularmente relevante para las universidades

¹ El origen del presente artículo fue el seminario “Presentes y futuros de la evaluación académica y la ciencia abierta: debates y perspectivas desde el Perú”, realizado el 19 y 20 de septiembre del 2023 en la Universidad Nacional Mayor de San Marcos y coorganizado por CLACSO-FOLEC. Allí se discutieron algunos de los puntos centrales que derivaron en este texto.

² Una versión avanzada del texto se discutió en el Seminario “Open Access and publishing asymmetries: Studies on APC costs and diamond publishing”, realizado el 27 y 28 de marzo del 2024 en el Global Research Institute of Paris, Francia. Agradecemos a las y los participantes del evento por los valiosos comentarios, en particular a Rodrigo Costas (Universidad de Leiden).

³ Para esta investigación se ha contado financiamiento de los proyectos “Mapeando la heterogeneidad estructural y capacidades diferenciales de producción y circulación de conocimientos en Argentina (2022-2024): indicadores de espacialidad académica para la formulación de políticas de redistribución y fortalecimiento CTI, con enfoque federal, multidisciplinar y de género” (director Víctor Algañaraz, CONICET-Universidad Nacional de San Juan), financiado por el programa PISAC del (ex) Ministerio de Ciencia, Tecnología e Innovación de Argentina, y el proyecto “Open science in Latin America and Africa” (directora Fernanda Beigel, CONICET-Universidad Nacional de Cuyo), financiado por el Global Research Institute of Paris-Université Paris-Cité.

de menor tamaño y para los artículos con colaboración nacional o de autoría individual, modalidades tradicionalmente poco representadas en las bases de datos bibliométricos tradicionales. Estos resultados son particularmente relevantes en un contexto de ataque a la ciencia producida en Argentina de parte del Gobierno federal electo en el 2023 y del corte al acceso a Scopus como fuente para la investigación científica.

Palabras clave: Argentina, OpenAlex, Scopus, evaluación, rankings universitarios.

Databases in the evaluation of scientific production in Latin America. Exploring OpenAlex's potential for Argentine universities

Abstract:

This article examines the coverage of OpenAlex and Scopus for scientific publications by Argentine authors, both nationwide and for a selection of 10 national universities, over the period 2019-2023. The objective is to compare the performance of both databases in terms of the overall volume of indexed publications, languages, and disciplinary areas. The results show that OpenAlex has a broader and more diverse coverage, in particular for publications in Spanish, in social sciences and humanities, and authored by a single individual. Additionally, it is observed that OpenAlex's coverage is especially relevant for smaller universities and articles with national collaborations or single authorship, which are often underrepresented in traditional bibliometric databases. These findings hold special significance amidst current challenges to science produced in Argentina made by the federal government elected in 2023 and the restriction of access to Scopus as a source for scientific research.

Keywords: Argentina, OpenAlex, Scopus, evaluation, university rankings.

1. Introducción

A principios del 2024 se publicó una nueva versión del ranking de instituciones científicas realizado por Scimago. En Argentina se le dio particular difusión, dada la embestida frontal del Gobierno nacional contra las instituciones científicas y universitarias, entre otras. Pero también la difusión fue amplia por el buen desempeño de las instituciones

argentinas, en particular del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).⁴

Los rankings como los de Scimago sin duda evalúan la producción científica —a veces otro tipo de producciones— de las universidades y de las instituciones científicas alrededor del mundo. Sin embargo, lo que a menudo se olvida es que no evalúan toda la producción científica o de otro tipo que se genera en el seno de esas instituciones. Estas, en particular las universidades latinoamericanas, producen distintos tipos de conocimiento que circula por múltiples vías, las cuales no pueden medirse con una misma metodología (Beigel, 2018; 2021).

El núcleo epistemológico de la producción de los rankings descansa sobre el volumen de artículos y patentes y las citas que estos reciben. Aun cuando en años recientes la metodología de Scimago ha ido incorporando nuevas dimensiones e indicadores, es claro que el peso del dispositivo artículos/citas/patentes es decisivo para “rankear” a las instituciones. Esto es claro en el factor Investigación, que representa el 50 % de la puntuación, pero también en Innovación (30 %), donde las tres subcategorías están determinadas por las citas.⁵

Hay, además, un problema específico para las instituciones latinoamericanas. Las fuentes que indexan las revistas científicas y que fueron creadas en los centros del sistema académico mundial tradicionalmente no han incluido revistas de las periferias, al tiempo que se han concentrado casi exclusivamente en el inglés (Ortiz, 2009). Si bien en los últimos años —y producto de una competencia puramente empresarial— varias de estas fuentes han ampliado su alcance, la producción latinoamericana sigue claramente subrepresentada, máxime si se analizan las ciencias sociales y humanidades, los formatos distintos del *paper* y los idiomas distintos del inglés (Sivertsen, 2021).

Los desafíos que se derivan de este dilema son múltiples. En este artículo nos centraremos en uno específico, el de la medición del alcance de una fuente emergente y, en principio, más amplia que las tradicionales. Para ello, exploraremos y compararemos la cobertura de la producción científica de distintas universidades argentinas en una base

⁴ Véase <https://n9.cl/ovzc2>

⁵ “Conocimiento Innovador” contabiliza la producción de publicaciones científicas de una institución citadas en patentes; “Impacto Tecnológico” es el porcentaje de la producción de publicaciones científicas citadas en patentes; y “Patentes” el número de solicitudes de patentes según el conteo de la Oficina Europea de Patentes. Véase <https://n9.cl/n2ko37>.

de datos que podemos considerar tradicional y de acceso restringido (Scopus) y una base de datos emergente y de acceso abierto (OpenAlex).

La elección de Scopus se justifica no solamente por el particular peso del ranking Scimago en la opinión pública argentina (el ranking se elabora con datos de Scopus), sino también porque es, para los países latinoamericanos, una fuente con mayor cobertura que Web of Science. También porque, al momento de recogida de los datos que aquí analizamos, todavía era accesible a través de la suscripción del Ministerio de Ciencia, Tecnología e Innovación de Argentina, degradado por el actual Gobierno.⁶ Web of Science, en cambio, no ha sido accesible desde el país en los últimos años.

OpenAlex es, por el contrario, una fuente de acceso abierto y con alcance notoriamente superior al de Web of Science, Scopus y otras bases de datos de acceso restringido. Esto ha sido claramente establecido en numerosos estudios recientes (Priem, Piwowar y Orr 2022; Simard *et al.*, 2024; Culbert *et al.*, 2024).

Sin embargo, así como son claras las limitaciones de las fuentes bibliométricas tradicionales para orientar la evaluación de instituciones e individuos, no es tan claro que las fuentes emergentes representen una alternativa fehaciente y sostenible para ser utilizadas como insumos en la evaluación. Nuestra intención es, por lo tanto, explorar qué cobertura diferencial ofrecen Scopus y OpenAlex de la producción científica de los investigadores e investigadoras de las instituciones argentinas.

En este trabajo, de carácter exploratorio, optamos por comparar un conjunto específico de diez universidades, en lugar de la Argentina como conjunto o de todas las universidades que hay en el país. Usar solo diez universidades permite relacionar sus indicadores bibliométricos con otras características como su antigüedad, su peso en el campo científico-universitario, su prestigio, y su ubicación, sin tener que realizar este análisis para el conjunto del campo, objetivo que excede a este trabajo. Se trata, no obstante, de una selección relevante de universidades. Según datos correspondientes al 2022, estas diez instituciones reúnen 708 553 estudiantes de grado (36.2 % del total del sistema público). Como institución de afiliación, agrupan el 32 % de las publicaciones indexadas por OpenAlex en el período 2019-2023, y el 36.9 % de las registradas por Scopus.

⁶ Al 28 de junio del 2024, Scopus no es accesible desde la Argentina. Desconocemos si se trata de una situación puntual o permanente.

A diferencia de otros contextos, en Argentina es muy habitual la doble afiliación institucional. El CONICET es la principal institución científica, sus investigadores e investigadoras tienen puestos de tiempo completo. La mayoría de ellos/as, al mismo tiempo, tienen cargos de profesores/as en las universidades (Beigel, Gallardo y Bekerman, 2018). Esto supone dos problemáticas que no pueden ser pasadas por alto en este tipo de análisis.

En primer lugar, Scimago contabiliza la doble afiliación (*multiple affiliations*) a partir de un conteo fraccional, distribuyendo el peso de cada documento entre la cantidad de afiliaciones de su autor para aquellos casos en que el este posea más de una afiliación institucional. Esto resulta problemático para el caso argentino, donde la doble afiliación es una práctica que forma parte de una legítima tradición científica nacional.

El otro inconveniente que observamos es que no parecen estar del todo bien contabilizadas las afiliaciones múltiples en Scimago para el caso argentino. Un 7.2 % de los artículos de autores/as del país aparece con más de una afiliación,⁷ pero no figura el CONICET como institución de pertenencia. Siendo este el principal organismo científico del país, hay por lo tanto dudas importantes respecto de la metodología seguida por Scimago.

Por otra parte, no está claro todavía cómo se administra la pertenencia institucional a la hora de firmar como autores/as en la academia argentina. Pero sí es evidente que no hay un criterio homogéneo entre los/as autores/as. A ello se suma las pérdidas de información que pueden producirse dentro de los procesos editoriales de las revistas, así como en el proceso de indexación por las bases de datos.

Por ello, comparar la cobertura de OpenAlex y Scopus a partir de cómo estas bases recuperan la afiliación institucional puede ser un primer paso para mensurar su potencial para la evaluación individual o institucional. Pero es claro que un segundo paso necesario es identificar planteles completos de investigadores/as a partir de otro tipo de fuentes y luego buscar las publicaciones de ellos/as en las distintas bases de datos, incluidos los repositorios institucionales y las iniciativas del estilo CRIS (Sivertsen, 2019).

⁷ Véase <https://n9.cl/u6449>.

2. Estado del arte

En trabajos previos, hemos analizado *corpus* de publicaciones de planteles completos de investigación, como parte de una línea de investigación que apunta a ir más allá de las mencionadas limitaciones de las bases de datos tradicionalmente utilizadas para caracterizar la producción científica en los países periféricos. Para el caso del CONICET, fue posible establecer la relevancia de las publicaciones en español y en libros, aun cuando el *paper* en inglés es claramente dominante; lo más relevante, sin embargo, es que esta bibliodiversidad no está necesariamente restringida a las ciencias sociales y humanidades (Beigel y Gallardo, 2021; Beigel *et al.*, 2023).

Scopus y Web of Science han sido las fuentes tradicionalmente utilizadas para los estudios bibliométricos sobre instituciones argentinas, o que incluyen al país en un panorama más amplio. Pero no hay demasiados estudios respecto de la validez de estas fuentes para caracterizar la producción completa de esas instituciones. Sobre una muestra de revistas de ciencias sociales, por ejemplo, se ha concluido que los “datos de Scopus no son indicadores adecuados ni suficientes para comprender su uso, alcance e impacto” (Rozemblum, Alperin y Unzurrunzaga, 2021, p. 16). En cualquier caso, los sesgos y limitaciones del circuito *mainstream* para con las revistas latinoamericanas son claros (Salatino, 2018; Beigel *et al.*, 2024). Respecto de OpenAlex, el número de investigaciones que la utilizan como fuente está creciendo,⁸ pero no hemos encontrado todavía un estudio específico para Argentina.

El país sí aparece analizado, entre muchos otros, en algunas publicaciones que han abordado el asunto de la múltiple afiliación institucional. Esta ha crecido en los últimos años, es decir, ha aumentado la proporción de autores/as que declaran vínculos con más de una institución, una de ellas argentina. Esta es claramente una tendencia global y Argentina parece mostrar un crecimiento menor al de países vecinos (Halevi *et al.*, 2023) y, al mismo tiempo, un mayor peso de la múltiple afiliación dentro del propio país antes que la internacional (Hottenrott, Rose y Lawson, 2021).

Finalmente, se han detectado importantes faltantes de información en el campo de afiliación institucional en OpenAlex. Lejos de ser un problema puntual, es una de las dificultades que afronta cualquier base que agregue metadatos de publicaciones. Para el

⁸ La búsqueda en la misma base de datos del término “openalex” en títulos y *abstracts* arroja 4 publicaciones en el 2021, 30 en el 2022, 112 en el 2023 y 28 el en 2024 (al 9 de abril). Búsqueda: <https://n9.cl/d24029>.

caso particular de OpenAlex, se ha cuantificado esta pérdida de información, si bien no hay una distribución uniforme a lo largo de todos los países y no parece ser un problema particularmente acentuado para América Latina (Zhang *et al.*, 2024).

3. Metodología

Analizamos los documentos indexados por Scopus y OpenAlex con fecha de publicación entre el 2019 y el 2023, ya que la cobertura de las bases de datos extrarregionales es más incierta en años alejados, pero más confiable en los recientes (Gallardo, 2022). Se apuntó a todos los tipos de documentos a fin de analizar la bibliodiversidad de los registros, si bien en el apartado de Resultados se analizan también los artículos por separado. La primera parte de esta sección analiza el conjunto de los documentos presentes en cada una de las bases de datos de manera agregada.

Luego, para la segunda parte, se identificó el desempeño de 10 universidades nacionales argentinas a partir del campo “afiliación institucional” de los/as autores/as. Esta metodología para un país como Argentina —en donde es muy común la doble afiliación— suma desafíos a los que están implícitos para este tipo de enfoque, en el que se caracteriza una institución a partir de la información recogida en ese campo por las bases de datos bibliométricas y que es problemático para cualquier base de datos, incluida OpenAlex (Zhang *et al.*, 2024), como mencionamos. Sin embargo, no hay por el momento una alternativa factible para caracterizar la producción científica indexada de los/as profesores/as e investigadores/as de las universidades argentinas.

OpenAlex ya tenía identificadas a todas las universidades seleccionadas, con la denominación normalizada, por lo que no se encontraron resultados por fuera de los artículos ya vinculados a cada una de estas universidades. No fue igual para Scopus, por lo que fue necesario buscar las instituciones con distintas denominaciones, como Universidad Nacional de Cuyo, National University of Cuyo, Universidad de Cuyo, etc. Luego de una primera identificación, a través de R, se seleccionó una muestra al azar para descartar falsos positivos y, al no encontrar ninguno, se consideró que el método resultaba válido.

La descarga de los datos se realizó durante febrero del 2024 y apuntó a los documentos con al menos un autor con afiliación a instituciones argentinas. Para Scopus se utilizó su sitio web a través del acceso que todavía proveía el desaparecido Ministerio

de Ciencia, Tecnología e Innovación. Para la descarga de OpenAlex se utilizó el API (Application Programming Interface) de la misma a través del paquete openalexR (Aria y Le, 2024) del software libre R (R Core Team, 2023). El análisis se realizó a través de Microsoft Excel y de R, en particular los paquetes openalexR y bibliometrix (Aria y Cuccurullo, 2017).

La selección de las 10 universidades se hizo de manera intencional, ya que no se intenta una muestra representativa sino una primera aproximación. Se buscó de todas maneras que estas conformaran un grupo variado tanto en antigüedad y tamaño como en su distribución geográfica. El listado se presenta en la tabla 1, en la que puede observarse la importante diversidad en las tres variables consideradas.

Tabla 1

Características de las universidades seleccionadas para el estudio empírico

Universidad	Año de creación	Estudiantes de grado (2022)	Región de ubicación
Buenos Aires	1821	347.280	Metropolitana
Córdoba	1613	169.395	Centro
Cuyo	1939	36.113	Cuyo
Centro	1974	18.957	Bonaerense
San Juan	1973	28.075	Cuyo
Río Negro	2007	13.274	Sur
Misiones	1973	28.554	Noreste
Santiago del Estero	1973	18.066	Noroeste
Patagonia Austral	1994	11.406	Sur
Hurlingham	2014	37.433	Metropolitana

Fuente: elaboración propia con base en anuarios estadísticos del ex-Ministerio de Educación de la nación⁹

La hipótesis de trabajo que subyace a esta selección es que la cobertura puede resultar diferencial entre las distintas bases de datos y universidades, en tanto no todas tienen el mismo perfil de producción y circulación de conocimiento. Universidades como las de Córdoba o Buenos Aires sin duda tienen una fuerte presencia en cualquier indicador, no solo en virtud de su tamaño comparativamente muy grande sino también por la acumulación de recursos y prestigio. Otras universidades, en cambio, son de mucha más

⁹ Véase <https://n9.cl/i7y36b>.

reciente fecha de creación y no acumulan prestigio ni recursos en la misma medida que las otras.

4. Resultados

Aplicando los criterios señalados, de Scopus se obtuvo un total de 79 715 registros (sin distinción de tipo) y de OpenAlex un número sensiblemente mayor, 134 067, en ambos casos para el período 2019-2023. Una primera mirada sobre estos datos es la determinación de en qué medida los registros son los mismos o no. El solapamiento entre ambas bases de datos se calculó únicamente a partir del DOI (Digital Object Identifier) de las publicaciones, aun cuando es un instrumento no exento de problemas para los países periféricos (Krauskopf y Salgado, 2023). En nuestro caso, los registros de OpenAlex cuentan con DOI en una proporción menor: 87.7 % frente al 93.3 % de los registros de Scopus, es decir, porcentajes altos.

La combinación de las dos bases de datos arroja un total de 150 084 registros de publicaciones con autores/as de Argentina en el período 2019-2023. A futuro, es preciso realizar el cruzamiento con un método de mayor calidad para buscar registros compartidos, aunque estos no tengan DOI. Sin embargo, los resultados obtenidos hasta el momento parecen suficientes para ilustrar con una certeza razonable la cobertura de cada base de datos para un país como Argentina.

En la tabla 2 se puede observar que el porcentaje de registros cubiertos en ambas bases de datos es del 42.4 %. Únicamente en OpenAlex se encuentra el 35.9 % (con DOI) además del 11 % (sin DOI). Estos porcentajes son sensiblemente menores para Scopus: 7.1 % (con DOI) y 3.5 % (sin DOI). Así, resulta claro que Scopus está en buena medida contenido en OpenAlex y, a la vez, que esta última tiene una importante proporción de registros sin cobertura por parte de Scopus.

Tabla 2

Cobertura entre OpenAlex y Scopus (vía DOI) de documentos con autores/as afiliados/as a instituciones de Argentina, 2019-2023

	DOI-Base de datos	Registros	Registros %
Con DOI	Con DOI-Solo Scopus	10 714	7.14 %
	Con DOI-Solo OpenAlex	53 816	35.86 %
	Con DOI-Scopus y OpenAlex	63 698	42.44 %
Sin DOI	Sin DOI-Scopus	5303	3.53 %
	Sin DOI-OpenAlex	16 553	11.03 %
Total		150 084	100 %

N=150.084

Fuente: elaboración propia

Lo relevante de este diferencial en la cobertura se puede observar en la diversidad disciplinar de cada base de datos, si bien no es posible una comparación totalmente directa debido a la diferente forma de clasificar los registros de cada una de las bases de datos. OpenAlex aplica distintos métodos automatizados para clasificar cada registro en un gran dominio, campos y subcampos, combinando el análisis textual con las redes de citación y otras variables. Los registros de Scopus, en cambio, pueden analizarse en función de la clasificación disciplinar de las revistas, que puede consultarse en Scimago.

De todas maneras, puede observarse el distinto peso de las ciencias sociales, punto particularmente relevante para un país como Argentina, donde la producción de esta área tiende a estar sobre todo fuera del circuito *mainstream*. En el caso de OpenAlex, 44 986 registros son clasificados en el dominio ciencias sociales (33.6 % del total registrado para el período 2019-2023). En el caso de Scopus, 11 454 documentos han sido publicados en revistas que Scimago identifica como de ciencias sociales o artes y humanidades (se trata del 14.3 %).

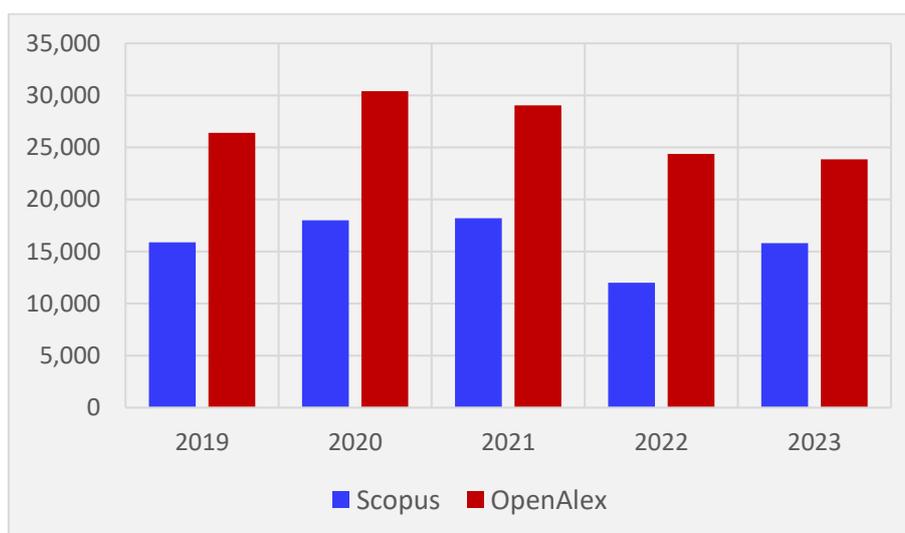
Es evidente que la cobertura entre las bases es distinta en función de las áreas disciplinares. Por ejemplo, si analizamos las cuatro grandes áreas en que se agrupan los registros obtenidos de OpenAlex, globalmente están cubiertos por Scopus el 54.2 % de los 117 514 registros de OpenAlex con DOI. Pero si segmentamos por área disciplinar, la cobertura es solo del 25.7 % en el caso de ciencias sociales y del 57.1 % en el de los

registros catalogados en ciencias de la salud. En cambio, es claramente mayor para ciencias de la vida (70.1 %) y ciencias físicas (70.7 %).

La evolución por año muestra un patrón similar para ambas fuentes, con una caída en el 2022 y, para Scopus, un repunte en el 2023 (ver gráfico 1). Es un patrón común para la región cuando se analiza la evolución de la cantidad de publicaciones en los años recientes (Gallardo, 2022; Beigel *et al.*, 2024), algo que presumiblemente esté vinculado con las dificultades en los procesos editoriales. No hay que olvidar, por otro lado, que los años considerados incluyen el desarrollo de la pandemia por COVID-19, lo que puede haber impactado en la productividad. En cualquier caso, es evidente que la diferencia entre OpenAlex y Scopus se verifica a lo largo de todo el período.

Gráfico 1

Documentos de autores/as afiliados/as a instituciones de Argentina en Scopus y OpenAlex por año, 2019-2023



Fuente: elaboración propia

Otro indicador relevante para la Argentina es el idioma de publicación de los documentos, el cual también muestra una diferencia muy importante (ver tabla 3). En Scopus, 12 835 documentos fueron publicados en español (apenas el 16.1 %). En OpenAlex, en cambio, son 45 033 (33.6 %); también otorga un lugar más relevante al portugués, al francés y a otros idiomas. Es una diferencia muy significativa, ya que uno de los principales sesgos que tienen las bases de datos tradicionales es su concentración en el inglés, producto de

unas condiciones específicas de origen que luego se universalizaron como medidas de la ciencia legítima (Ràfols, 2019; Gingras, 2016; Ortiz, 2009).

Tabla 3

Documentos de autores/as afiliados a instituciones de Argentina en Scopus y OpenAlex según idioma de publicación, 2019-2023

Idioma	Scopus	Scopus %	OpenAlex	OpenAlex %
Inglés	66.727	83,5%	85.167	63,5%
Español	12.835	16,1%	45.033	33,6%
Portugués	174	0,2%	1.811	1,4%
Francés	88	0,1%	579	0,4%
Otros	80	0,1%	1.477	1,1%
Total	79.904	100%	134.067	100%

Fuente: elaboración propia

En cuanto a tipo de documentos, hay mayor diversidad en Scopus (ver tabla 4). OpenAlex no tiene ningún registro consignado como reseña (es posible que algunas estén contadas como artículos). En la tabla siguiente se utilizó el campo de tipo de artículo según CrossRef para OpenAlex, ya que resulta más comparable con Scopus (la clasificación de la propia base de datos es distinta).

Tabla 4

Documentos de autores/as afiliados/as a instituciones de Argentina en Scopus y OpenAlex por tipo de documento, 2019-2023

Tipo	Scopus	Scopus %	OpenAlex	OpenAlex %
Artículo	63.756	79,8 %	118.651	88,5 %
Reseña	5.758	7,2 %	0	0 %
Paper de congreso	3.259	4,1 %	3.067	2,3 %
Capítulo de libro	2.939	3,7 %	5.607	4,2 %
Otros	4.192	5,3 %	6.742	5 %
Total	79.904	100 %	134.067	100 %

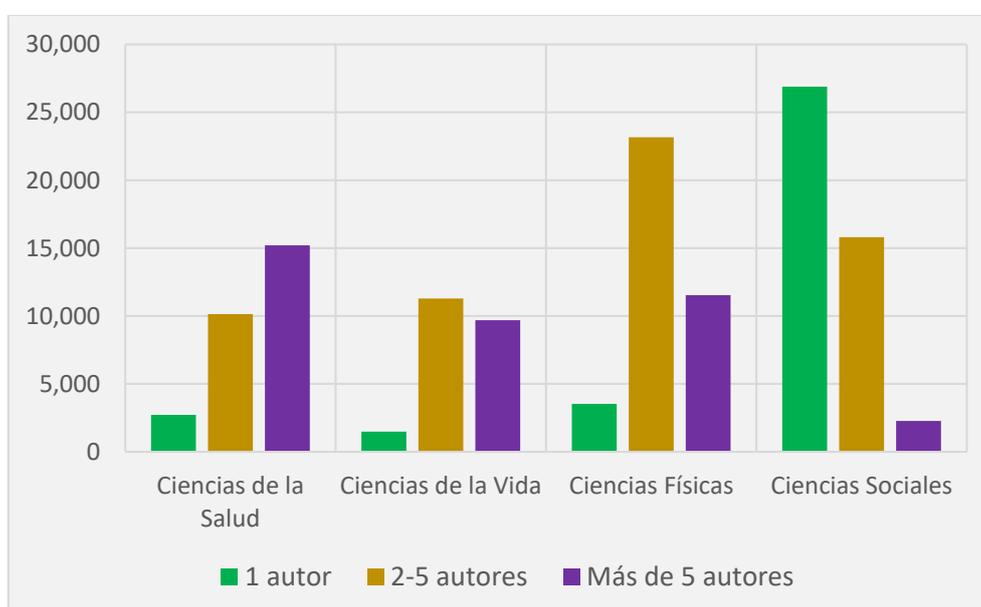
Fuente: elaboración propia

Otro tipo de publicación que generalmente no está bien representado en las bases de datos tradicionales es la publicación individual, generalmente más asociada con la publicación

en español en Argentina o América Latina y, mayoritaria pero no exclusivamente, con las ciencias sociales y humanidades (Beigel y Gallardo, 2021). En el gráfico 2 se muestran los registros de OpenAlex según los cuatro grandes campos disciplinares de esta base de datos y la cantidad de autores de cada publicación. Es llamativo que la categoría “Más de 5 autores” solo sea mayoritaria en ciencias de la salud, y que se ubique en segundo lugar en ciencias de la vida y físicas. En ciencias sociales el resultado sí es más cercano al esperado *a priori*, con un claro predominio de las publicaciones individuales.

Gráfico 2

Documentos de autores/as afiliados/as a instituciones de Argentina en OpenAlex por área disciplinar, 2019-2023



Fuente: elaboración propia

En cuanto a citas recibidas contabilizadas por OpenAlex, el promedio general para los registros analizados es de 4.5. Los registros con al menos una cita son 58 505 (43.6 % del total) y en este caso el promedio de citas es de 10.2. En todos los casos, el mayor promedio corresponde a los artículos etiquetados en ciencias de la salud y el menor, a los de ciencias sociales.

También es relevante para un sistema científico como el argentino el país de publicación de los registros. Para realizar un análisis exploratorio, se analizó el país de las fuentes —revistas u otras— según las identifica OpenAlex. El 13.8 % de los registros aparece publicado en revistas que han sido catalogadas como editadas en Argentina. En

sí mismo esto es un resultado interesante, ya que es notoriamente mayor al casi nulo margen de publicaciones que pueden identificarse en el país en las bases de datos del circuito *mainstream*.

Por otro lado, es claro que este grupo de publicaciones corresponde mayoritariamente a las ciencias sociales, lo que resulta congruente con las anticipaciones de sentido que se pueden postular. El mencionado 13.8 % de publicaciones en Argentina se descompone en un 8.6 % registrado como correspondiente a ciencias sociales, un 2.2 % a ciencias de la salud, un 1.7 % a ciencias físicas y un 1.2 % a ciencias de la vida.

La caracterización sencilla que se ha planteado para Argentina a partir de OpenAlex resulta mucho más cercana a la que emergería del análisis de la totalidad de la producción científica del país que la que es posible construir con Scopus o Web of Science, al menos en esta primera aproximación. El peso de las ciencias sociales y del español son fuertes indicadores de que OpenAlex está menos sesgada que las bases de datos *mainstream*.

Sin embargo, aquí no abordamos tres problemas fundamentales. Por un lado, no hemos evaluado en profundidad la calidad de los metadatos obtenidos de OpenAlex. Se aplicó solamente un *data cleaning* inicial (eliminación de registros repetidos o con errores en los campos). En segundo lugar, no existen bases de datos que permitan estimar las principales tendencias de la producción científica argentina. Los estudios disponibles usualmente se realizan a partir de las bases de datos del circuito *mainstream* o de instituciones específicas. En nuestros trabajos previos y en curso nos hemos concentrado en CONICET, dada la disponibilidad de datos, pero no en las universidades nacionales. No existen en Argentina iniciativas del tipo CRIS ni hay bases de datos de antecedentes o CV que puedan ser utilizadas para ese fin. Por último, los metadatos sobre afiliación institucional necesitan un tratamiento particularmente cuidadoso para un país como Argentina, donde es muy frecuente la doble afiliación institucional, como ya mencionamos. No se ha medido todavía qué márgenes de error se producen al utilizar los metadatos de afiliación institucional de las bases de datos bibliométricas. Esto se dificulta, también, por lo mencionado, en tanto no existen listados de profesores/as o investigadores/as por institución.

Un punto de partida posible para iniciar el análisis con mayor profundidad es la Universidad de Buenos Aires (UBA), la universidad de mayor peso dentro del sistema científico argentino. Por su tamaño y por su impacto, su cobertura disciplinar es más

amplia que, presumiblemente, cualquier otra universidad. Alberga en su seno autores/as con múltiples modalidades de circulación, desde campos fuertemente internacionalizados e insertos en el circuito *mainstream* en una posición relativamente relevante; hasta múltiples circuitos locales, publicados mayoritariamente en español y que no son necesariamente dominados por el artículo como modo privilegiado de comunicación del conocimiento.

Los registros que fue posible identificar en OpenAlex de autores/as vinculados/as con la UBA en el período 2019-2023 son 24 604. La gran mayoría de ellos son artículos (91.7 %), seguidos por capítulos de libro (1075 registros, el 4.4 %). Los registros de libros son apenas 33 (0.1 %). Como se mencionó, es difícil estimar cuál puede ser la producción publicada en formato de libros, pero para una universidad del tamaño de la UBA, es muy posible que la cantidad real sea sustancialmente mayor.

Los artículos en autoría individual son el 29 % (ver tabla 5) y una gran mayoría de ellos (casi 9 de cada 10) es clasificado por OpenAlex dentro de las ciencias sociales. Es marginal el peso que tiene la autoría individual dentro de las otras áreas. En los artículos firmados en coautoría, aquellos únicamente con autores/as vinculados a instituciones argentinas representan un 34 % del total. Por su parte, la colaboración internacional se produce en el 37 % de los artículos.

Tabla 5

Artículos según tipo de autoría y colaboración. Autores/as con afiliación a la Universidad de Buenos Aires, 2019-2023

Tipo de autoría y colaboración	Registros	Registros %
Autoría individual	6542	29 %
Coautoría con colaboración nacional	7661	34 %
Coautoría con colaboración internacional	8355	37 %
Total	22 558	100 %

N=22.258

Fuente: elaboración propia

Este es un punto interesante, pues en las bases de datos *mainstream* no solo está sobrerrepresentado el inglés y el *paper*, sino también la publicación en coautoría y en colaboración internacional. La colaboración nacional, para los países periféricos, está más enfocada en la publicación dentro del país o la región, sin que esto deba ser entendido

como un rasgo, negativo, de la llamada endogamia (Beigel *et al.*, 2024). En el caso de la UBA, hay una aproximada división en tercios entre las tres modalidades representadas en la tabla 5, lo que sin dudas es un aspecto novedoso sobre la diversidad de la producción científica contenida en OpenAlex, aun cuando esté mayoritariamente concentrada en artículos en revistas.

Establecido que la comparativa a nivel de universidades parece ser un camino propicio, el paso siguiente fue ampliar la misma a otras 9 universidades (ver tabla 6). Como mencionamos, se trata de instituciones seleccionadas de manera intencional pero diversas en cuanto a su ubicación geográfica, su antigüedad y su peso en el campo científico-universitario argentino. La comparativa se presenta a nivel de documentos totales y de artículos que es posible encontrar en OpenAlex y Scopus. En el caso de la primera, las 10 universidades están ya identificadas y han seguido el proceso de desambiguación. Esto no es así en el caso de Scopus, donde fue preciso buscar distintas formas de los nombres y siglas de las universidades para identificar los artículos firmados por sus miembros. A modo de validación, se aplicó el mismo método para OpenAlex, con resultados negativos, aunque cabe aclarar que se tomó como punto de partida el conjunto de artículos con autores/as de Argentina, no la base de datos completa.

Tabla 6

Documentos y artículos en OpenAlex y Scopus según universidad de afiliación de autores/as y diferencia porcentual, 2019-2023

Universidad Nacional	Documentos			Artículos		
	OpenAlex	Scopus	Diferencia	OpenAlex	Scopus	Diferencia
Buenos Aires	24 604	18.125	36 %	22 558	14 500	56 %
Córdoba	8917	5.588	60 %	8517	4843	76 %
Cuyo	2982	2.370	26 %	2839	1968	44 %
Centro	1675	705	138 %	1586	618	157 %
San Juan	1278	946	35 %	1202	788	53 %
Río Negro	1277	698	83 %	1208	623	94 %
Misiones	1204	532	126 %	1173	479	145 %
Santiago del Estero	514	250	106 %	472	205	130 %
Patagonia Austral	466	239	9%	445	208	114 %
Hurlingham	n. d.	86	-	n. d.	73	-

Fuente: elaboración propia. La diferencia porcentual se calculó como (registros OpenAlex – registros Scopus)/registros Scopus.

De la tabla precedente se pueden extraer varias conclusiones. Primero, el diferencial en cantidad de registros entre OpenAlex y Scopus es mayor para los artículos que para el conjunto de los documentos. Esto va en línea con lo ya mencionado respecto de una mayor diversidad en los tipos de documentos en Scopus. Se refuerza entonces la mejor cobertura de OpenAlex para las universidades argentinas cuando se restringe el análisis a los artículos.

Segundo, la diferencia en los artículos registrados por cada base de datos tiende a ser menor en las universidades colocadas al principio de la tabla y a aumentar conforme se analizan las universidades ubicadas al final. Las instituciones están ordenadas en función de los documentos en OpenAlex y el orden resultante refleja apropiadamente las diferencias de profesores/as, estudiantes e incluso antigüedad (ver tabla 1). Es decir, OpenAlex parece ser particularmente útil, o más representativa, para las universidades argentinas de menor tamaño (esto es, de menor impacto en las bases de datos bibliométricas, pero también de menor tamaño en general). En la muestra de universidades seleccionada, la excepción la constituye la Universidad Nacional de Hurlingham, de muy reciente creación. Esta universidad está identificada por OpenAlex, pero no había, al momento de extracción de los datos, ningún documento asociado con ella.

Otra comparación que puede hacerse para este conjunto de universidades es la colaboración internacional. La tabla 7 muestra el porcentaje de documentos que tienen este tipo de colaboración en la autoría en cada base de datos. Lo más relevante es que este porcentaje es mayor para Scopus que para OpenAlex en casi todos los casos. Si bien las diferencias varían entre las distintas universidades, se trata de un indicio claro de que OpenAlex puede contener una mayor proporción de documentos con colaboración exclusivamente nacional. Estos, presumiblemente, corresponderán en una medida considerable a publicaciones realizadas en Argentina y en español. Evidentemente, es necesario un análisis más profundo para evaluar correctamente esta tendencia. Sobre todo, sería preciso cuantificar, y eventualmente excluir del análisis, los registros con información faltante sobre el país de los autores o autoras.

Tabla 7

Documentos en colaboración internacional y principales países de colaboración en OpenAlex y Scopus según universidad de afiliación de autores/as, 2019-2023

Universidad Nacional	% en colaboración internacional		Principales países de colaboración	
	OpenAlex	Scopus	OpenAlex	Scopus
Buenos Aires	38.9 %	45.0 %	Estados Unidos, Brasil, España, Chile, Alemania	Estados Unidos, España, Brasil, Alemania, Francia
Córdoba	39.3 %	45.3 %	Estados Unidos, Brasil, España, Chile, Italia	Estados Unidos, España, Brasil, Chile, Reino Unido
Cuyo	41.6 %	47.9 %	Estados Unidos, España, Brasil, Chile, Alemania	España, Estados Unidos, Chile, Brasil, Alemania
Centro	29.6 %	31.5 %	México, España, Estados Unidos, Colombia, Uruguay	España, México, Estados Unidos, Colombia, Uruguay
San Juan	40.1 %	41.1 %	Brasil, Ecuador, Estados Unidos, Colombia, Rusia	Brasil, Estados Unidos, Colombia, Ecuador, Chile
Río Negro	49.7 %	46.4 %	Canadá, España, Brasil, Estados Unidos, Alemania	España, Brasil, Canadá, Estados Unidos, Italia
Misiones	43.4 %	45.5 %	Brasil, Estados Unidos, Cuba, Paraguay, Australia	Brasil, Estados Unidos, España, Australia, Alemania
Santiago del Estero	32.9 %	33.6 %	España, India, México, Brasil, Chile	España, Brasil, Estados Unidos, Chile, México
Patagonia Austral	34.1 %	40.6 %	Alemania, Brasil, España, Estados Unidos, Australia	España, Estados Unidos, Alemania, Suiza, Canadá
Hurlingham	Sin registros	37.21 %	-	Estados Unidos, España, Reino Unido, Italia, Chile

Fuente: elaboración propia

En cuanto a los países con los que se produce mayor colaboración, no aparecen en este vistazo rápido grandes diferencias. En casi todos los casos, los principales países son los Estados Unidos, Brasil y España. OpenAlex, sin embargo, ofrece algunas particularidades interesantes. Por ejemplo, para la Universidad Nacional de Misiones aparecen países importantes en cuanto a colaboraciones autores/as de Cuba, Paraguay y Australia. También aparece la India como un país relevante de colaboración para la Universidad Nacional de Santiago del Estero, una particularidad que podría ser explorada a futuro, sea que se trate de un rasgo distintivo o de un error de la base de datos.

Sobre esta última posibilidad, no hemos evaluado la calidad de los datos de afiliación institucional por fuera de las universidades seleccionadas. Pero, en principio, no parece haber una gran pérdida de información en ese campo. Del total de registros de documentos en OpenAlex de las diez universidades, apenas el 10 % de los autores/as totales de tales documentos aparece sin ninguna información sobre afiliación institucional.

Conclusiones

En este trabajo nos propusimos responder a la pregunta sobre la representatividad y utilidad de las bases de datos bibliométricos emergentes como fuentes para la evaluación del desempeño de individuos y universidades en Argentina, un centro periférico del sistema académico mundial. Los datos analizados nos permiten dar un primer paso en la construcción de esa respuesta, así como nos permiten nutrir las preguntas de investigación que alimenten nuestra indagación en el futuro próximo.

En las variables que analizamos en la comparación general entre OpenAlex y Scopus, la primera se reveló como más amplia y representativa que la segunda. Esto sucede en todos los casos, excepto en la diversidad documental, dada la inclusión en Scopus de reseñas, algo que OpenAlex parece no hacer.

Es claro que OpenAlex no solo incluye más documentos cuya autoría corresponde a investigadores/as afiliados/as a instituciones argentinas, sino que es más diverso el panorama que representa. Este es un punto fundamental, en tanto se trata de una base de datos global que, *a priori*, no está tan sesgada contra el español, la autoría individual y las ciencias sociales y humanidades. Es necesario, sin duda, profundizar en el análisis de estos datos, pero parece un escenario promisorio. Sobre todo, es interesante que haya un lugar más relevante para la autoría individual y la colaboración nacional, modalidades con una muy baja representación en las bases de datos tradicionales para un país como Argentina.

No obstante, el aporte de este trabajo parece ser el diferencial encontrado en la cobertura, siendo más relevante OpenAlex para las universidades pequeñas que para las más grandes (ver tabla 6). En varias de las instituciones de menos peso cuantitativo en el campo científico-universitario argentino, OpenAlex registra más del doble de artículos respecto de los que contabiliza Scopus. Esta ampliación de la base empírica sobre la que

se pueden calcular tendencias en la colaboración, redes de citación y coautoría, impacto de las revistas con APC, etc., representa una muy interesante perspectiva para caracterizar a todas las instituciones científicas y no solo a aquellas que tienen más peso en Web of Science o Scopus.

En el futuro próximo esperamos profundizar el análisis a partir de este primer paso en dos direcciones fundamentales. La primera es una estimación de la fiabilidad y validez de los datos recopilados y procesados por OpenAlex. Web of Science y Scopus no están exentas de problemas en cuanto a la calidad de sus datos y OpenAlex necesita ser analizada en profundidad sistemáticamente. Para el caso argentino, parece particularmente relevante analizar la confiabilidad del campo de afiliación institucional

Aun así, la segunda dirección apunta a intentar superar las limitaciones intrínsecas de tales campos de afiliación institucional en OpenAlex y cualquier base de datos, dificultades particularmente relevantes para el contexto argentino. El camino más prometedor parece ser el de los estudios de abajo hacia arriba, esto es, que parten de los planteles de investigadores/as y profesores/as y, desde allí, van hacia los datos bibliométricos (Beigel *et al.*, 2023). Metodológicamente esto incluye una diversidad de desafíos considerables cuando estos planteles son muy grandes o directamente no hay certeza sobre los mismos. Pero no parece posible comenzar a superar los sesgos, usos y abusos (Gingras, 2016) de cualquier base de datos bibliométricos si no se restituye a las personas que ocupan una posición concreta en el campo científico, y no solo a los autores/as detrás de las publicaciones.

Referencias bibliográficas

- Aria, M, y Cuccurullo, C. (2017). Bibliometrix: An R–Tool for Comprehensive Science Mapping Analysis. *Journal of Informetrics*, 11(4), 959-975.
- Aria, M. y Le, T. (2024). *OpenalexR: Getting Bibliographic Records from OpenAlex*. <https://n9.cl/isdmtd>
- Beigel, F., Packer, A. P., Gallardo, O. y Salatino, M. (2024). OLIVA: la producción científica indexada en américa latina. diversidad disciplinar, colaboración institucional y multilingüismo en SciELO y Redalyc (1995-2018). *Dados*, 67(1), e20210174. <https://doi.org/10.1590/dados.2024.67.1.307>

- Beigel, F. (2021). A Multi-Scale Perspective for Assessing Publishing Circuits in Non-Hegemonic Countries. *Tapuya: Latin American Science, Technology and Society* 4(1), 1845923. <https://doi.org/10.1080/25729861.2020.1845923>
- Beigel, F., Almeida, A. M., Gallardo, O., Digiampietri, L., Gómez, S., Rangel Candido, M., Ciriza, A., Rossomando, P., Moschkovich, M. y Pecheny, M. (2023). Scientific Production and Gender Inequalities in Two Academic Elites: Brazil and Argentina. *Revue d'histoire Des Sciences Humaines*, 42, 255-280.
- Beigel, F. (2018). Las relaciones de poder en la ciencia mundial. Un anti-ranking para conocer la ciencia producida en la periferia. *Nueva Sociedad. Democracia y política en América Latina*. <https://n9.cl/1jl05>
- Beigel, F., Gallardo, O. y Bekerman, F. (2018). Institutional Expansion and Scientific Development in the Periphery: The Structural Heterogeneity of Argentina's Academic Field. *Minerva*, 56(3), 305-331. <https://doi.org/10.1007/s11024-017-9340-2>
- Beigel, F. y Gallardo, O. (2021). Productividad, bibliodiversidad y bilingüismo en un corpus completo de producciones científicas. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad-CTS*, 16(46). <https://n9.cl/55m1qm>
- Culbert, J., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P. y Mayr, P. (2024). Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus. *arXiv*. <https://doi.org/10.48550/arXiv.2401.16359>
- Gallardo, O. (2022). La evolución de la producción científica en revistas indexadas en Iberoamérica. Principales tendencias y diferencias con el circuito mainstream. *Papeles del Observatorio*, 23, 5-41. <https://n9.cl/xgbv77>
- Gingras, Y. (2016). *Bibliometrics and Research Evaluation: Uses and Abuses*. <https://doi.org/10.7551/mitpress/10719.001.0001>
- Halevi, G., Rogers, G., Guerrero-Bote, V. P. y De-Moya-Anegón, F. (2023). Multi-affiliation: a growing problem of scientific integrity. *El Profesional de la información*, e320401. <https://doi.org/10.3145/epi.2023.jul.01>
- Hottenrott, H., Rose, M. E. y Lawson, C. (2021). The Rise of Multiple Institutional Affiliations in Academia. *Journal of the Association for Information Science and Technology*, 72(8), 1039-1058. <https://doi.org/10.1002/asi.24472>

- Krauskopf, E. y Salgado, M. (2023). Inconsistency in the registration of the Digital Object Identifier (DOI) of articles on Web of Science and Scopus. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 37(96), 129-144. <https://doi.org/10.22201/iibi.24488321xe.2023.96.58784>
- Ortiz, R. (2009). *La supremacía del inglés en las ciencias sociales*. Siglo Veintiuno Editores.
- Priem, J., Piwowar, H. y Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*. <https://doi.org/10.48550/arXiv.2205.01833>
- R Core Team (2023). *R: A language and environment for statistical computing*. <https://www.R-project.org>
- Ràfols, I. (2019). S&T Indicators in the Wild: Contextualization and Participation for Responsible Metrics. *Research Evaluation*, 28(1), 7-22. <https://doi.org/10.1093/reseval/rvy030>
- Rozemblum, C., Alperin, J. P. y Unzurrunzaga, C. (2021). Las limitaciones de Scopus como fuente de indicadores: Buscando una visibilidad integral para revistas argentinas en ciencias sociales. *e-Ciencias de la Información*. <https://doi.org/10.15517/eci.v11i2.44300>
- Salatino, M. (2018). Más allá de la indexación: circuitos de publicación de ciencias sociales en Argentina y Brasil. *Dados*, 61, 255-287. <https://doi.org/10.1590/001152582018152>
- Simard, M.-A., Basson, I., Hare, M., Larivière, V. y Mongeon, P. (2024). The open access coverage of OpenAlex, Scopus and Web of Science. *arXiv*. <https://doi.org/10.48550/arXiv.2404.01985>
- Sivertsen, G. (2019). Developing Current Research Information Systems (CRIS) as Data Sources for Studies of Research. En W. Glänzel, H. F. Moed, U. Schmoch y M. Thelwall (Eds.). *Springer Handbook of Science and Technology Indicators* (pp. 667-683). Springer Handbooks. https://doi.org/10.1007/978-3-030-02511-3_25
- Sivertsen, G. (2022). Publishing in the Social Sciences and Its Representation in Research Evaluation and Funding Systems. En T. Engels y E. Kulczycki. *Handbook on Research Assessment in the Social Sciences* (pp. 238-261). Edward Elgar Publishing. <https://n9.cl/yqjvg>

Zhang, L., Cao, Z., Shang, Y., Sivertsen, G. y Huang, Y. (2024). Missing Institutions in OpenAlex: Possible Reasons, Implications, and Solutions. *Scientometrics*.
<https://doi.org/10.1007/s11192-023-04923-y>