

OpenAlex. Grandeza y cicatrices de un Frankenstein bibliográfico

OpenAlex. The grandeur and scars of a bibliographic Frankenstein

Daniel Torres-Salinas y Wenceslao Arroyo-Machado

Torres-Salinas, D., & Arroyo-Machado, W. (2026). "OpenAlex. Grandeza y cicatrices de un Frankenstein bibliográfico". *Anuario ThinkEPI*, 20, e20a12. <https://doi.org/10.3145/thinkepi.2026.e20a12>

Publicado en *IweTel* el 14 de mayo de 2026

Daniel Torres-Salinas

<https://www.directorioexit.info/ficha772>

<https://orcid.org/0000-0001-8790-3314>

Universidad de Granada

torressalinas@go.ugr.es

Wenceslao Arroyo-Machado

<https://www.directorioexit.info/ficha4831>

<https://orcid.org/0000-0001-9437-8757>

INGENIO (CSIC-UPV)

warrmac@upvnet.upv.es



Resumen: *OpenAlex* se ha consolidado como la gran alternativa abierta y gratuita frente a *Web of Science* y *Scopus*. Sin embargo, su enorme cobertura contrasta con la fragilidad de sus cimientos. Por un lado, supera los 450 millones de registros y aporta mayor diversidad lingüística, disciplinar y geográfica. Por otro, su agregación automatizada y sin curación arrastra metadatos deficientes y miles de revistas potencialmente predatorias. A partir de ese diagnóstico, se ofrecen recomendaciones sobre cuándo conviene usar la base y con qué cautelas.

Palabras clave: *OpenAlex*; Bibliometría; Bases de datos bibliográficas; Metadatos; Acceso abierto.

Abstract: *OpenAlex* has established itself as the major open and free alternative to *Web of Science* and *Scopus*. Its vast coverage, however, contrasts with the fragility of its foundations. On the one hand, it exceeds 450 million records and offers greater linguistic, disciplinary, and geographical diversity. On the other, its automated and uncurated aggregation carries over poor metadata and thousands of potentially predatory journals. Drawing on this diagnosis, the article offers recommendations on when the database should be used and with what cautions.

Keywords: *OpenAlex*; Bibliometrics; Bibliographic databases; Metadata; Open access

1. Introducción

La irrupción de *OpenAlex* en el ecosistema de la información científica representa uno de los cambios más significativos de la última década. Lanzada en 2022 por *OurResearch* como sucesora de *Microsoft Academic Graph*, se ha presentado como la alternativa abierta y gratuita a *Web of Science* y *Scopus*. Sin embargo, la velocidad y el impacto de su adopción contrastan con la fragilidad de sus cimientos técnicos. Conviene analizar con datos en la mano qué estamos ganando exactamente y qué estamos, quizás, decidiendo ignorar.

Los argumentos se apoyan en dos trabajos recientes de los propios autores (Torres-Salinas y Arroyo-Machado, 2026; Forchino y Torres-Salinas, 2026) que, leídos en paralelo, permiten construir un retrato más nítido, y también más incómodo, de lo que esta base de datos es hoy realmente. En la primera parte de este *ThinkEPI* ofrecemos una comparativa básica de la cobertura de las tres principales bases de datos del momento. En la segunda ofrecemos una mirada crítica sobre la calidad de la información en *OpenAlex*. Finalmente nos atrevemos a cerrar con una serie de recomendaciones a tener en cuenta antes de usar esta base de datos.

2. La grandeza de indexarlo (casi) todo

Antes de entrar en los números conviene entender qué hay detrás de cada una de estas tres plataformas, porque sus diferencias no son solo técnicas sino filosóficas. Tienen que ver sobre todo con qué se incluye y qué se ignora. *Web of Science* aplica 28 criterios de evaluación de revistas (24 de calidad y 4 de impacto) y, eventualmente, revisa y sanea estas listas. En 2024 excluyó 17 revistas por manipulación de citas, sumando 67 expulsadas en dos años. *Scopus* recibe cerca de 3.500 solicitudes anuales, pero solo el 34% supera los requisitos mínimos para ser evaluado y, de ese grupo, apenas la mitad es finalmente aceptada. Ambas, son bases de datos que se ganan a pulso cada registro que incluyen. Modelos selectivos, costosos, imperfectos sin duda, pero con criterios conocidos, públicos y, la mayoría de las veces, auditables.

OpenAlex, en cambio, indexa 50.000 registros diarios de forma automatizada desde fuentes como *Crossref*, *PubMed*, *ORCID* y repositorios, sin filtros previos ni comités editoriales, sin intervención humana. Ese modelo de agregación voraz es lo que convierte a *OpenAlex* en una suerte de Frankenstein bibliográfico, ensamblado con retazos de múltiples fuentes abiertas con no pocas cicatrices. El resultado final es apabullante. Según su propio sitio web, *OpenAlex* supera ya los 450 millones de registros totales incluyendo más de 2.000 millones de enlaces de citas, mientras *Scopus* ha alcanzado los 100 millones y *Web of Science* ronda los 97 millones. Si hablamos solo de artículos científicos entre 2015-2024, *Scopus* reúne 27,3 millones, 23,9 millones de *Web of Science* y *OpenAlex* alcanza los 54,1 (poco más de la mitad si se limita la fuente a revistas).

El volumen es, sin duda, su grandeza, pero si vamos más allá del total de registros, los datos revelan diferencias estructurales si consideramos el perfil temático y la cobertura lingüística.

En cuanto al perfil temático, las bases comerciales siguen manteniendo su sesgo hacia las STEM, con especial inclinación hacia Biomedicina y Ciencias de la Salud. *Web of Science* dedica el 43,3% de su cobertura a esta área, *Scopus* el 37,3% y *OpenAlex* el 34%, una diferencia real pero no abismal. Donde sí encontramos una divergencia más marcada es en Ciencias Sociales y Humanidades. *OpenAlex* destina el 20,4% de su cobertura a estas disciplinas, frente al 9,0% de *Web of Science* y el 12,1% de *Scopus*. El perfil lingüístico apunta en la misma dirección más representativa. El 77,5% de los artículos de *OpenAlex* están en inglés, frente al 96,2% de *Web of Science* y el 90,8% de *Scopus* lo que configura un perfil más plural. El español se convierte en la segunda lengua por volumen con un 2,8% del total, frente al 1,1% de *Web of Science* y el 1,0% de *Scopus*. Tampoco está la cosa para tirar cohetes. *OpenAlex* ha hecho más grande el mapa, pero no ha cambiado quién vive en el centro.

3. Las cicatrices del jovencito Frankenstein

Antes de señalar los problemas conviene reconocer lo que *OpenAlex* ha conseguido. Indexa el 89% de las revistas del DOAJ, el 95% de las de acceso abierto dorado y el 71% de las activas en *Open Journal Systems*, con una representación del Sur Global que ninguna base comercial puede igualar. Brasil, Argentina y Colombia lideran una adopción temprana que no es casual, ya que sus comunidades científicas han dependido durante décadas de herramientas prohibitivas en términos de coste.

Su API robusta ha permitido que investigadores sin licencias comerciales realicen análisis de gran escala que antes eran imposibles. Sin embargo, la ausencia de cualquier proceso de curación tiene consecuencias que no se pueden ignorar.

En *OpenAlex* conviven registros perfectamente válidos con otros de dudosa fiabilidad y metadatos mínimos, sin que el usuario reciba ningún aviso al respecto. El 61,5% de sus registros carecen de afiliación institucional, y esa cobertura cae al 55% precisamente para los registros exclusivos, es decir, aquellos donde la base presume de aportar cobertura única. La paradoja se enuncia sola, ya que cuanto más original es la cobertura de *OpenAlex*, menos fiables son sus metadatos. La clasificación documental añade otro frente problemático. Antes de mayo de 2024 *OpenAlex* etiquetaba como *article* el 99,42% de sus ítems. Tras una posterior revisión la cifra bajó al 89,59%.

Dos problemas adicionales merecen atención. El primero afecta a la cobertura de campos clave como el resumen. *Elsevier*, alegando derechos de *copyright*, ha retirado sus resúmenes de *OpenAlex* al dejar de compartirlos en abierto, con una pérdida de hasta 11,5 millones de *abstracts* de un plumazo. Un recordatorio de que la gratuidad de la base depende, en parte, de la buena voluntad de los mismos actores comerciales a los que pretende sustituir. El segundo es más inquietante. Un análisis de 400 revistas predatorias del catálogo *Cabells* encontró que el 36,5% estaban presentes en *OpenAlex*; extrapolando, entre 6.300 y 8.200 revistas potencialmente depredadoras habitan la base. Entre ellas, el editor *OMICS*, multado con 50 millones de dólares por la FTC estadounidense, que acumula más de 215.000 registros.

Para *OurResearch* esto no es un problema de su incumbencia ya que el filtrado se delega en el usuario. Pero no siempre el usuario final conoce estas cosas. Cuando las decisiones que se toman con esos datos afectan a carreras académicas y asignación de fondos, la filosofía del *caveat emptor* se convierte en un problema ético, no solo técnico.

4. OpenAlex: ¿sí o no?

OpenAlex encarna una promesa política genuina, la de una infraestructura científica que no discrimine por idioma, región ni capacidad económica institucional. Que la Sorbona haya cancelado sus contratos con *Clarivate* para adoptar soluciones abiertas, o que el *Leiden Ranking* publique su *Open Edition* basado exclusivamente en *OpenAlex*, son señales de una dirección que el campo ha tomado de forma deliberada. Pero la velocidad de adopción no puede ser la coartada que exima a una herramienta del escrutinio técnico que aplicamos a cualquier otra fuente bibliométrica. A nadie se le ocurriría usar un instrumento de medición sin calibrar para un análisis con consecuencias reales. Proponemos tres consideraciones a la hora de decidir sobre su uso:

- Naturaleza del análisis. Si el estudio tiene consecuencias evaluativas directas (acreditaciones, rankings institucionales, informes de agencia), *OpenAlex* no debería emplearse como fuente única ni principal, dada la fragilidad de sus metadatos de afiliación y clasificación documental.
- Objeto de estudio. Si el análisis se orienta hacia cobertura lingüística diversa, acceso abierto diamante, producción iberoamericana o disciplinas subrepresentadas, *OpenAlex* añade un valor genuino e insustituible, siempre con validación cruzada.
- Cultura de la depuración. Incorporar *OpenAlex* obliga a destinar recursos específicos a la limpieza de datos: duplicados, afiliaciones, tipos documentales, fuentes predatorias. Quien no planifique ese coste adicional está construyendo sobre arena.

El futuro de la evaluación responsable pasa por las bases de datos abiertas. Pero ese futuro exige construir esas infraestructuras con la misma exigencia que durante décadas hemos aplicado a las comerciales. *OpenAlex* tiene la arquitectura y la ambición para ser el recurso del futuro. Tiene, y en

abundancia, lo que más entusiasma a quienes trabajan con datos: una API potente, capacidades de descarga masiva y un ecosistema interesante de herramientas de análisis. Esos aspectos técnicos, los flujos de trabajo recomendados y los procedimientos de validación que se deberían implementar, de forma crítica, antes de utilizar *OpenAlex* para un análisis bibliométrico serán el objeto de nuestro próximo *ThinkEPI*.

5. Declaración de uso de IA

Los autores, una vez finalizado el texto, le pidieron a *Perplexity* (LLM=Claude Sonnet 4.6. Thinking) una revisión mediante el siguiente *prompt*: “*Revisa ortografía y gramática del texto adjunto*”.

6. Bibliografía

Torres-Salinas, D., & Arroyo-Machado, W. (2026). *Las Big Three de la información científica: revisión bibliométrica comparativa de Web of Science, Scopus y OpenAlex* (Informe PLUGG01XX26, Versión 1.2). EC3metrics / InluScience Ediciones. <https://doi.org/10.5281/zenodo.18336510>

Forchino, M. V., & Torres-Salinas, D. (2026). The OpenAlex database in review: Evaluating its applications, capabilities, and limitations. *Journal of Informetrics*, 20, 101800. <https://doi.org/10.1016/j.joi.2026.101800>