

La fiabilidad de *Web of Science* y *Scopus* y su repercusión en el ecosistema científico

Inmaculada Bravo-García

Universidad de Salamanca, España

<https://ror.org/02f40zc51>

<https://orcid.org/0000-0003-3388-0545>

inma@usal.es

Valeria Montoya-Roncancio

Universidad de Salamanca, España

<https://ror.org/02f40zc51>

<https://orcid.org/0000-0001-7357-3326>

valeriamontoya@usal.es

Bravo-García, Inmaculada; Montoya-Roncancio, Valeria (2026). “La fiabilidad de *Web of Science* y *Scopus* y su repercusión en el ecosistema científico”. En: Orduña-Malea, Enrique; Olea, Isabel (coords). *El valor de las revistas científicas* (pp. 30-36). Ediciones Profesionales de la Información.

<https://doi.org/10.3145/vrc.03>

Resumen

A pesar de la relevancia de *Web of Science* (WoS) y *Scopus* como principales bases de datos bibliográficas, se evidencia que contienen errores en los metadatos que afectan la visibilidad y las métricas de los investigadores. Este estudio examina la fiabilidad de estas bases de datos y su impacto en la precisión de los perfiles de autor y en la evaluación científica. Para ello, se desarrolla un modelo replicable denominado “Detector de Inconsistencias en Perfiles Científicos”, aplicado en cinco universidades españolas. La metodología comprende cuatro fases: recopilación de datos institucionales, consulta automatizada a las API de WoS y Scopus, comparación de metadatos con análisis automatizado y revisión manual de los casos conflictivos. Los resultados permiten clasificar los registros como correctos o con distintos tipos de incidencias, generando informes que facilitan la corrección manual. Las universidades participantes confirman la utilidad del modelo para depurar y mejorar la calidad de sus datos científicos.

Se concluye que la fiabilidad de *WoS* y *Scopus* tiene implicaciones directas en la evaluación del impacto académico, pues los errores reducen la precisión de las métricas y rankings. Se propone institucionalizar la revisión periódica de perfiles y fortalecer la integración con identificadores persistentes como Orcid, garantizando datos más fiables y una representación más justa del desempeño investigador.

Palabras clave

Falibilidad; impacto; Perfiles de autor; *Web of Science*; *Scopus*; Bases de datos bibliográficas.

Abstract

Despite the relevance of *Web of Science* (*WoS*) and *Scopus* as the main bibliographic databases, evidence shows that they contain metadata errors that affect researchers' visibility and metrics. This study examines the fallibility of these databases and its impact on the accuracy of author profiles and research assessment. To this end, a replicable model named "Detector of Inconsistencies in Scientific Profiles" was developed and applied across five Spanish universities. The methodology comprises four phases: collection of institutional data, automated querying of the *WoS* and *Scopus* APIs, automated metadata comparison, and manual review of conflicting cases. The results allow records to be classified as correct or affected by different types of inconsistencies, generating reports that facilitate manual correction. Participating universities confirm the model's usefulness in cleansing and improving the quality of their scientific data. The study concludes that the fallibility of *WoS* and *Scopus* has direct implications for academic impact evaluation, as such errors reduce the accuracy of metrics and rankings. It therefore proposes institutionalising the periodic review of author profiles and strengthening integration with persistent identifiers such as Orcid, ensuring more reliable data and a fairer representation of researchers' performance.

Keywords

Fallibility; Impact; Author profiles; *Web of Science*; *Scopus*; Bibliographic databases.

1 Introducción

En la actualidad, *Web of Science* (*WoS*) y *Scopus* son consideradas como las bases de datos bibliográficas más influyentes del ecosistema científico. Su capacidad de indexación con más de 91 millones de documentos y de 22.000 revistas las convierten no sólo en herramientas de búsquedas relevantes para descubrir información, sino también, como el método para identificar, valorar y evaluar la calidad de la producción científica que desarrollan los investigadores.

La falibilidad de *WoS* y *Scopus* es un concepto que ha sido poco estudiado, aunque existen preocupaciones significativas sobre la precisión de los metadatos relacionados con el mantenimiento y creación de perfiles de autor que suministran estas bases bibliográficas (Olivas-Ávila; Musi-Lechuga, 2014; Eck; Waltman, 2019; Bravo-García et al., 2024). Rechazar la premisa de la infalibilidad, permite construir una red de información que considera que el error es inevitable (Harari, 2024).

La presente investigación pretende identificar las problemáticas comunes en la duplicación de perfiles de autores, asignación de autorías incorrectas, creación de un perfil con varios autores mezclándose la producción de todos ellos, autores en diferente orden (Chinchilla-Rodríguez et al., 2024) y un mismo DOI para diferentes artículos (Krauskopf; Salgado, 2023). Errores que evidencian que hay metadatos que pueden faltar, fallar y convertir la información suministrada en inexacta, incierta y, por tanto, falible; tanto en *WoS* como en *Scopus*.

El objetivo principal de esta investigación es identificar las inconsistencias en los metadatos de *Web of Science* y *Scopus* en los perfiles de los investigadores analizados. Para ello se plantean los siguientes objetivos específicos:

- Desarrollar un modelo replicable: “Detector de Inconsistencias en Perfiles Científicos”.
- Generar informes que faciliten la corrección manual de las inconsistencias.
- Detectar perfiles de autor/autora duplicados, errores de atribución de publicaciones y otras incidencias menores.
- Colaborar con universidades que validen el modelo propuesto.

2 Metodología y resultados

Esta investigación se llevó a cabo a través de cuatro fases que se describen a continuación.

La primera fase consistió en obtener de la universidad de origen datos relacionados con sus publicaciones con DOI, así como de los autores de las mismas (figura 1).

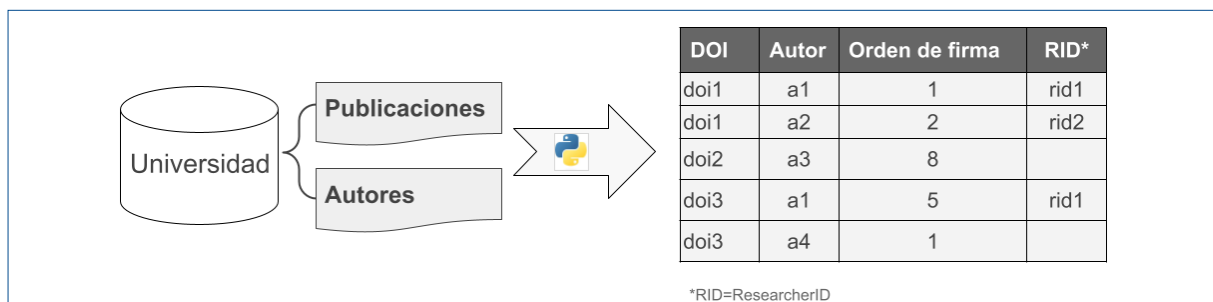


Figura 1. Obtención de datos de origen
Fuente: elaboración propia (2025)

En la segunda fase, se consultó las API de *WoS* y *Scopus* programando una serie de scripts en *Python*, utilizando *Spyder* como entorno de desarrollo y *SQLite* como base de datos. Se realizó posteriormente una búsqueda sistemática de todos los autores y publicaciones que formaban parte de una universidad en ambas plataformas (figura 2).

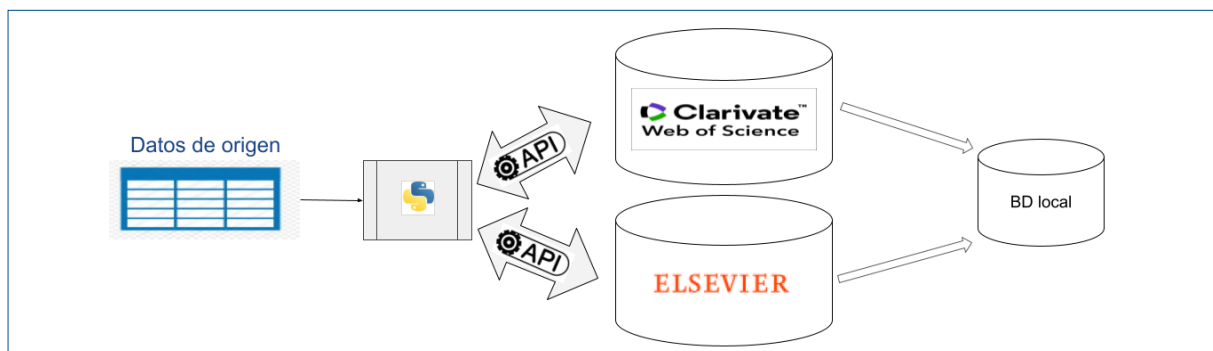


Figura 2. Consulta de API de WoS y Scopus
Fuente: elaboración propia

En la tercera fase, una vez se recolectaron los datos, se realizó un primer análisis automatizado por cada base de datos, comparando los datos de origen, es decir, los que muestra la universidad en el portal de producción científica, y los obtenidos al consultar la API de *WoS* y la API de *Scopus* por cada registro. Tras ello se generan informes en formato hoja de cálculo como se muestra en la figura 3.

PORTAL_ID	FIRMA	ORCID	SCOPUS_ID	S_SCOPUS_ID	S_FIRMA	Clasificación origen/API-BD	PUBLICACIONES
57689	Héctor del Estal Sánchez	0000-0003-0827-4076		57201878579	Sanchez H.D.E.	Incendencia_ID	7
157391	Hernández Mateos, Alberto	0000-0003-0886-5781		41861501300	Hernandez Mateos A.	Incendencia_ID	3
157391	Alberto Hernández Mateos	0000-0003-0886-5781		57376949600	Mateos A.H.	Incendencia_ID	2
57717	Mateos, M.-V.	0000-0003-2390-1218	35405592800			Orden_excedido	1
57717	Mateos, M.-V.	0000-0003-2390-1218	35405592800	35405592800	Mateos, Maria	En_diferente_orden	2
57717	Mateos, M.V.	0000-0003-2390-1218	35405592800	35405592800	Mateos, M. V.	Correcto	29
57717	Mateos, M.-V.	0000-0003-2390-1218	35405592800	57223876573	Mateos, Maria Victoria	Incendencia_ID	482
57717	Mateos, V.	0000-0003-2390-1218	35405592800	9235544900	Mateos, Victoria	Incendencia_ID	1

<https://produccioncientifica.usal.es/investigadores/57717/publicaciones>

<https://orcid.org/0000-0003-2390-1218>

<https://www.scopus.com/authid/detail.uri?authorid=9235544900>

<https://produccioncientifica.usal.es/documentos/5e723e1e2999525b1d0d1744>
<https://www.scopus.com/record/display.uri?eid=2-s2.0-27744549869&origin=resultlist>

Figura 3. Informe de autores con incidencias en *Scopus*

Fuente: elaboración propia

En la figura 3, se observa un informe de *Scopus*, si bien se genera otro igual para *WoS*. Las primeras columnas hacen referencia a los datos recolectados desde el portal de producción científica de la universidad, seguidos de los obtenidos en la API. La columna “Clasificación origen/API-BD” se refiere a la clasificación automatizada de la comparación entre ambas:

1. Correcto: la información en origen y en la consulta a la API es la misma y por tanto, todo es correcto.
2. Incendencia_ID: se detectan inconsistencias entre los identificadores de autor en los datos de origen y los hallados en la API.
3. Otras incidencias menores:
 - Orden_excedido: el orden del autor en origen, excede el número de autores registrados en *WoS* o *Scopus*. Por lo cual, la publicación no se muestra en el perfil del investigador.
 - En_diferente_orden: la obra está bien atribuida al autor, porque se ha detectado el mismo como uno de los autores firmantes, pero en diferente orden de firma que en origen.
 - ERROR_DOI_más_de_uno: el mismo DOI asignado a diferentes artículos. Este error sólo se ha detectado en *WoS*.

La última columna “PUBLICACIONES”, corresponde con el número de publicaciones de un autor que tienen la misma “Clasificación origen/API-BD” y los mismos metadatos.

Para facilitar el análisis manual, se incluyeron enlaces en cada registro a los perfiles del autor y a las publicaciones tanto en el portal de origen como en *WoS*, en *Scopus* y en *Orcid*, con el fin de tener toda la información útil para tomar una decisión y ejecutar o solicitar la acción correctora necesaria.

La cuarta fase del proyecto se focalizó en los registros cuyo campo “Clasificación origen/API-BD” es “Incidencia_ID”. En estos casos, fue necesario un análisis manual, caso a caso, por personal técnico (figura 4).



Figura 4. Análisis manual de incidencias: “Incidencia_ID”
Fuente: elaboración propia

Ante cada una de estas “Incidencia_ID” (es decir, cuando son diferentes los identificadores del autor en los datos de origen y los hallados en la API), el identificador de autor obtenido en *WoS* o *Scopus* se tipificó en alguno de los siguientes estados:

- Perfil principal del autor obtenido en la API, con inconsistencias en el portal de investigación de la universidad (no está definido o es diferente).
- Perfil duplicado del autor en *WoS* o *Scopus*.
- De otro investigador, la obra está mal atribuida.
- Es un error de orden, no es una incidencia de identificadores.
- Perfil con varios autores, mezclándose la producción de todos ellos en un perfil. Este caso sólo se ha detectado en *WoS*.

3. Discusión y conclusiones

La implementación del modelo “Detector de Inconsistencias en Perfiles Científicos” ha permitido establecer un sistema automatizado y replicable para la identificación de discrepancias entre los datos académicos registrados en las plataformas institucionales de investigación y aquellos disponibles en las bases de datos bibliográficas, consultadas mediante API (*Web of Science* y *Scopus*).

Como parte del desarrollo metodológico, se diseñó una clasificación estructurada de las inconsistencias detectadas, lo que ha permitido diferenciar entre registros correctos y diversos tipos de anomalías. Adicionalmente, se estableció una tipificación de los estados de las Incidencias_ID, con el fin de facilitar su interpretación y resolución según su origen, naturaleza y complejidad.

El modelo ha sido aplicado a los perfiles científicos de cinco universidades públicas españolas, generando para cada institución un informe detallado con los resultados obtenidos en ambas plataformas, encontrándose en el momento de redacción de este trabajo en la cuarta fase de tipificación manual.

Uno de los principales aspectos de este trabajo es poner el foco en la realidad de que *WoS* y *Scopus* son falibles. Estas bases de datos se utilizan como fuentes de información para recolectar publicaciones, citas y otros indicadores, sin tener en cuenta los posibles errores de perfiles de autor duplicados o atribuciones incorrectas de publicaciones.

La recopilación de información incompleta, imprecisa, errónea y sesgada, influye no sólo en la visibilidad del investigador, sino también, en la evaluación de su impacto académico, ya que los errores en la atribución y duplicación de perfiles pueden disminuir el conteo de citas y afectar métricas claves utilizadas en los rankings de universidades (Vernon et al., 2018). Así mismo, puede repercutir en las decisiones de contratación, la promoción académica de los profesores y la categorización de las revistas científicas (Merlo-Vega; Montoya-Roncancio, 2023).

Con esta investigación se busca contribuir a la mejora de mecanismos de detección y resolución de incidencias que permitan garantizar la precisión de los perfiles de autor. Al conseguir que la identidad digital sea más precisa, se incrementará su reputación en el ecosistema científico (Cruz-Cruz, 2019).

La retroalimentación obtenida de las universidades sobre la implementación del modelo es bastante uniforme, les permite abordar la depuración y mejorar la calidad de los datos de sus portales científicos. Anteriormente, se detectaban las incidencias o bien por comunicación del personal investigador, o las detectaban en sus comprobaciones aleatorias, pero no partían de una información exhaustiva y pertinente. Aunque existen una gran cantidad de incidencias a resolver, los enlaces que se proporcionan a los perfiles de autor y a los documentos en las distintas bases de datos, les facilita el trabajo. Las universidades proponen realizar el estudio periódicamente para mantener la calidad de los datos.

El motivo fundamental de las incidencias en la autoría de las publicaciones es la ambigüedad con los nombres de los autores. Para solucionar estos problemas y conseguir la desambiguación surgió el proyecto Orcid (Marín-Arraiza; Mejías, 2020). A pesar del esfuerzo por los investigadores y revistas de aprovisionar este dato del Orcid en todos sus artículos, parece que los algoritmos que utilizan estas bases de datos continúan sin resolver ciertas inconsistencias al recopilar los metadatos de los autores.

Por ello, este estudio se centra en detectar los errores, tipificarlos y proponer una cadena de actuaciones para mitigarlos por parte de la universidad o del propio investigador, de la siguiente forma:

- Detectar de manera automática inconsistencias en los perfiles de autor de *WoS* y *Scopus*.
- Informar a la Institución.
- Los bibliotecarios revisan y tipifican manualmente las incidencias detectadas e informan a los investigadores.
- Los investigadores o los bibliotecarios solicitan su corrección a *WoS* y *Scopus*.

Una vez resueltas las incidencias, la calidad de los datos de sus portales científicos será veraz y completa, mejorando el prestigio, el impacto y la visibilidad tanto de los investigadores como de la institución, no sólo en *WoS* y en *Scopus*, sino en todo el ecosistema científico.

4. Referencias

Bravo-García, Inmaculada; Vega, Juan A.; Montoya-Roncancio, Valeria (2024). "Precisión de los perfiles de autor en portales institucionales de producción científica". *Investigación bibliotecológica: Archivonomía, Bibliotecología e Información*, v. 38, n. 101.
<https://doi.org/10.22201/iibi.24488321xe.2024.101.58897>

Chinchilla-Rodríguez, Zaida; Costas, Rodrigo; Robinson-García, Nicolás; Larivière, Vincent (2024). "Examining the quality of the corresponding authorship field in *Web of Science* and *Scopus*". *Quantitative science studies*, v. 5, n. 1, pp. 76-97.

https://doi.org/10.1162/qss_a_00288

Cruz-Cruz, Elso M. (2019). "Identidad digital y reputación científica de investigadores e instituciones". *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, v. 44, n. 2.

<https://revzoiolomarinello.sld.cu/index.php/zmv/article/view/1760>

Harari, Y. N. (2024). *Nexus: Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. Debate.

Krauskopf, Erwin; Salgado, Mauricio (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*, v. 37, n. 96.

<https://doi.org/10.22201/iibi.24488321xe.2023.96.58784>

Merlo Vega, Juan A.; Montoya-Roncancio, Valeria (2023). "Criterios de evaluación de revistas científicas". *Revista estudios de la información*, v. 1, n. 1, pp. 71-89.

<https://doi.org/10.54167/rei.v1i1.1223>

Olivas-Ávila, Jose A.; Musi-Lechuga, Bertha (2014). "Validez y fiabilidad del Researcher ID y de 'Web of Science Production of Spanish Psychology'". *International Journal of Clinical and Health Psychology*, v. 14, n. 1, pp. 58-66.

[https://doi.org/10.1016/S1697-2600\(14\)70037-7](https://doi.org/10.1016/S1697-2600(14)70037-7)

Van-Eck, Nees-Jan; Waltman, Ludo (2019). "Accuracy of citation data in *Web of Science* and *Scopus*". ArXiv [preprint].

<https://doi.org/10.48550/arXiv.1906.07011>

Vernon, Marlo M.; Balas, E. Andrew; Momani, Shaher (2018). "Are university rankings useful to improve research? A systematic review". *Plos one*, v. 13, n. 3.

<https://doi.org/10.1371/journal.pone.0193762>