

Precisión de los perfiles de autor en portales institucionales de producción científica

Inmaculada Bravo-García*
José Antonio Merlo Vega**
Valeria Montoya-Roncancio***

Artículo recibido:
5 de marzo de 2024
Artículo aceptado:
13 de agosto de 2024

Artículo de investigación

RESUMEN

Las universidades difunden su producción científica a través de portales de investigación donde integran las publicaciones de sus investigadores, las cuales proceden de bases de datos bibliográficas como Web of Science. Este estudio examinó la correspondencia entre la información de los investigadores de la Universidad de Salamanca en Web of Science y en el Portal de la Investigación de la USAL. La metodología empleada fue mixta e incluyó técnicas de análisis automatizados y manuales para el tratamiento de datos de Web of Science que per-

- * Servicios Informáticos (Centro de Proceso de Datos), Universidad de Salamanca, España inma@usal.es
- ** Departamento de Biblioteconomía y Documentación, Universidad de Salamanca, España merlo@usal.es
- *** Programa de Doctorado Formación en la Sociedad del Conocimiento, Universidad de Salamanca, España valeriamontoya@usal.es

mitieron identificar las inconsistencias en los datos de publicaciones y autores. Se desarrolló una metodología en cinco fases que apoyó en la extracción semiautomática de identificadores ResearcherID de Web of Science de los investigadores que carecían de dicho identificador en el Portal USAL, así como en la validación de la corrección de los identificadores existentes. El método utilizado en el análisis y en la tipificación de incidencias son replicables en cualquier entidad de investigación. Las conclusiones determinan la consistencia de la información procedente de Web of Science, abogan por la autonomía de las plataformas institucionales de las fuentes externas y ofrecen soluciones para alcanzar la mejor calidad en los datos de una universidad en las plataformas comerciales de producción científica.

Palabras clave: Producción científica; Perfiles de autor; ResearcherID; Portales de investigación

Accuracy of Author Profiles in Institutional Portals of Scientific Production

Inmaculada Bravo-García, José Antonio Merlo Vega and Valeria Montoya-Roncancio

ABSTRACT

Universities disseminate their scientific production using research portals integrating their researchers' publications from bibliographic databases such as Web of Science. This study examined the correspondence between the available information of the University of Salamanca researchers in Web of Science and the USAL Research Portal. The methodology employed was mixed and included automated and manual analysis techniques for Web of Science data processing, which made it possible to identify inconsistencies in the publication and author data. We developed a five-phase methodology; it provided a semi-automatic extraction of ResearcherID identifiers from Web of Science of researchers who did not have this identifier in the USAL Portal and helped validate the correctness of the existing identifiers. The method used in the analysis and the typification of incidences are replicable in any research entity. The conclusions determine the consistency of the information provided by Web of Science, advocate for the autonomy of institutional

platforms from external sources, and offer solutions to achieve the highest quality in a university's data in scientific production commercial platforms.

Keywords: Scientific Production; Author Profiles; ResearcherID; Research Portals

INTRODUCCIÓN

Planteamiento

Las universidades divulgan su producción científica a través de portales institucionales donde presentan de forma visual los datos de las publicaciones de sus investigadores, a menudo aportando indicadores de impacto y de acceso abierto. Los portales de investigación son las plataformas mediante las cuales las universidades difunden su producción científica. Las publicaciones de las que sus autores los investigadores de una institución se denominan producción científica. Los datos bibliográficos y los perfiles de autor de los portales proceden de fuentes internas y externas. La información se integra de forma automatizada y manual y se revisa por los servicios de apoyo a la investigación encargados de la gestión de los portales que, con frecuencia, son los sistemas bibliotecarios.

La integración de datos se realiza partiendo de los sistemas de gestión curricular de la institución (Current Research Information Systems, CRIS) y de las bases de datos bibliográficas que disponen de información sobre producción científica y de los perfiles de autor, como Web of Science, Scopus y Dialnet. Los perfiles de autor son los datos de identificación, producción científica y citas correspondientes a un investigador, disponibles en un portal o plataforma especializada.

Los identificadores persistentes son códigos normalizados asignados de forma única a una persona, entidad o publicación. Las bases de datos emplean identificadores propios de autores, como el identificador ResearcherID de Web of Science; además de identificadores abiertos de autor, como Open Researcher and Contributor ID (ORCID); e identificadores únicos de publicaciones, especialmente Digital Object Identifier (DOI).

Los portales institucionales de producción científica difunden la investigación publicada realizada en una universidad (Calderón-Rehecho, 2017). Estos no sustituyen a los CRIS ni a los repositorios institucionales, sino que son un canal desde donde se promocionan las publicaciones generadas por el personal investigador de la institución. Los portales asumen una función de transparencia en la

divulgación de la producción científica, por lo que deben asegurar que contienen toda la investigación de la universidad y que sus datos son fiables.

La integración de datos externos puede realizarse mediante una Interfaz de Programación de Aplicaciones (por sus siglas en inglés, API), que es una tecnología que facilita la incorporación de la información bibliográfica desde una fuente externa. Web of Science (en adelante, WoS) y Scopus disponen de sus propias API que permiten extraer los datos particulares de investigadores y publicaciones. La importación requiere de identificadores como ResearcherID en WoS o AuthorID en Scopus. En el caso de que las fuentes externas aporten datos erróneos o ignoren información publicada, estas carencias se reflejarán en los portales institucionales si no se realizan revisiones constantes.

La Universidad de Salamanca (USAL), como muchas otras, divulga las publicaciones de sus miembros a través del Portal de la Investigación, disponible en <<https://produccioncientifica.usal.es/>>, que agrega la información de las publicaciones desde bases de datos como Dialnet, Scopus y WoS. Esta investigación ha analizado la consistencia de los datos de los investigadores de la USAL disponibles en WoS. La metodología ha permitido incrementar la visibilidad de la producción científica realmente generada en la universidad, ya que se detectaron inconsistencias, se eliminaron ambigüedades y se asignaron las autorías correctas a publicaciones mal identificadas.

Objetivos

El objetivo general de la investigación fue evaluar la consistencia de la información que se extrae de las bases de datos bibliográficas que funcionan como las fuentes de recolección de información para los portales institucionales. El objetivo general se concretó en tres específicos, todos ellos centrados en WoS como fuente de datos de producción científica para su integración en el Portal de la Investigación de la Universidad de Salamanca (Portal USAL):

Objetivo específico 1: examinar la correspondencia entre las publicaciones de investigadores de la Universidad de Salamanca que deberían ofrecerse en WoS y las que realmente se les asignan.

Objetivo específico 2: identificar las discrepancias de WoS en la descripción y la asignación de publicaciones de investigadores de la Universidad de Salamanca.

Objetivo específico 3: desarrollar un procedimiento autónomo para la extracción de datos desde WoS que enriquezca los contenidos del Portal USAL.

La hipótesis de la que se partió consistió en la necesidad de revisar la cantidad y la calidad de la información que se obtiene desde WoS (objetivo 1 y objetivo 2), para aprovechar su valor como fuente de datos para los portales institucionales (objetivo 3).

Revisión bibliográfica

El tema principal de la investigación fue la calidad de los datos bibliográficos procedentes de bases de datos como Web of Science como fuentes para la importación automática en portales institucionales de producción científica (Birkle *et al.*, 2020). Aunque la literatura académica ha abordado tangencialmente este tema y otros aspectos relacionados, como el uso de las API, las tecnologías para la extracción de datos, la utilidad de los identificadores persistentes y el análisis de varios tipos de inconsistencias en las grandes bases de datos bibliográficas, aún hay una escasez de estudios sobre la calidad de los datos de WoS en lo que respecta a la producción científica de una universidad específica, por lo que este artículo contribuirá a la investigación sobre este tema. En el caso de la Universidad de Salamanca, ya se ha analizado la producción científica indexada en Scopus con un enfoque descriptivo y cuantitativo (Medina *et al.*, 2024).

Las principales bases de datos que contienen información bibliográfica y de autorías ofrecen una API para recuperar publicaciones. Velez-Estevez *et al.* (2023) realizaron un exhaustivo estudio sobre las tendencias de las API en bibliometría y establecieron una taxonomía y un análisis comparativo de las API bibliográficas, con cada vez más presencia en los contextos bibliométricos, como señalan Torres-Salinas y Arroyo-Machado (2022), quienes también aportaron una completa relación de API de fuentes bibliográficas. Por su parte, Moral-Muñoz *et al.* (2020) seleccionaron y describieron programas que habilitaron la extracción automática de datos de plataformas como WoS, Scopus y Google Scholar.

La fiabilidad de los datos de WoS ha sido parcialmente estudiada por varios autores, por ejemplo, Olivas-Ávila y Musi-Lechuga (2014) cuestionaron la validez de la información registrada en WoS en autores españoles del ámbito de la psicología. Por otra parte, Zhu, Hu, y Liu (2019) detectaron la asignación incorrecta del mismo DOI en diferentes trabajos tanto en Scopus como en WoS y propusieron soluciones. Krauskopf y Salgado (2023) analizaron la omisión del DOI en WoS en múltiples artículos de investigadores chilenos. Donner (2017) identificó la calidad de la asignación de los tipos de documentos en WoS, mientras que Wang y Waltman (2016) centraron su investigación en los sistemas de clasificación de revistas de WoS y Scopus. Lazonder y Janssen (2022), estudiaron la calidad de las citas en WoS, mientras que Penteado Filho y Fonseca Júnior (2017) examinaron errores en las filiaciones.

Adicionalmente, Franceschini, Maisano y Mastrogiacomo (2016) realizaron un amplio estudio que permitió tipificar los errores contenidos en Scopus y en WoS, aportando una comparación detallada de ambas bases de datos. Boudry y Durand-Barthez (2020) analizaron el uso de los identificadores ORCID y ResearcherID de autores de su institución en redes académicas.

Los identificadores únicos persistentes son materia de estudio en múltiples investigaciones (Cujba, 2019; Santos y Maimone, 2023). La variedad de identificadores de autor está ampliamente desarrollada en la bibliografía especializada (Borrego, 2014; Mering, 2017; Craft, 2020), aunque solo un número limitado de estudios se ha posicionado en torno al control de la producción científica a través de identificadores como ResearcherID de WoS, Author ID de Scopus y, sobre todo, ORCID. Mallery (2016) ha indicado que el código de autor ORCID es de uso generalizado por su carácter universal e independiente. Harrison y Harrison (2016) señalaron la obligación de los investigadores de crear y mantener múltiples perfiles e identificadores de autor para estar presentes en las plataformas de impacto y difusión de la producción científica.

La función de las bibliotecas como centros de apoyo a la investigación y a la difusión de la producción científica de las universidades ha sido ampliamente desarrollada en la bibliografía profesional. El contexto teórico se presenta en estudios como los de Jubb (2016), Hoffman (2016) y Tran y Chan (2020). Asimismo, la evolución de los perfiles profesionales ha sido analizada por Davis y Saunders (2020), Iribarren-Maestro (2018) y otros autores que han defendido la actuación de las bibliotecas en el control de los identificadores de autores de una institución (Sandberg y Jin, 2016; Craft, 2020; Silva y Guimarães, 2023).

METODOLOGÍA

Diseño metodológico

La metodología se define como mixta, en términos de Hernández-Sampieri y Mendoza Torres (2018), ya que se obtuvieron y analizaron datos procedentes de diferentes técnicas de investigación que se interpretaron de forma conjunta. Se trató de una investigación cuantitativa con diseño no experimental transversal correlacional, ya que se ha trabajado con un análisis de datos tratados estadísticamente, recolectados en un único momento y que sirvieron para establecer relaciones. Igualmente, se ha realizado un estudio de caso cualitativo. La combinación de técnicas cuantitativas y cualitativas determina la metodología empleada como mixta.

El estudio de caso cualitativo analizó los datos de WoS correspondientes a la Universidad de Salamanca. El estudio correlacional se diseñó para conocer la relación entre los datos de WoS y los del portal de la producción científica de la Universidad de Salamanca. La investigación no experimental transversal de tipo correlacional o causal consistió en la recolección de datos en una fecha determinada y en el posterior análisis de las relaciones entre los datos estudiados. La fecha de extracción de datos fue el 23 de julio de 2023. La tabulación y el análisis de datos se realizó entre julio y octubre de 2023. El informe de resultados se obtuvo en noviembre de 2023. La aplicación práctica de la metodología, en cuanto a la corrección de errores detectados, se inició en febrero de 2024.

Desarrollo de la investigación

Se diseñó un procedimiento semiautomatizado para la obtención y el análisis de los datos que se estructuró en cinco fases: 1. Obtención de los datos del Portal USAL; 2. Obtención de los datos de WoS; 3. Análisis de los datos y generación de archivos Excel; 4. Análisis manual de una muestra con incidencias; y 5. Nueva consulta a la API de WoS con los identificadores de la muestra seleccionada con incidencias específicas. Se partió de los datos disponibles en el Portal USAL, de los que se seleccionaron únicamente las publicaciones con DOI y con autores de la USAL. Los datos brutos se conservan para posteriores verificaciones y la posibilidad de difusión en acceso abierto en repositorios de datos. La Figura 1 esquemiza las fases y los procesos realizados.

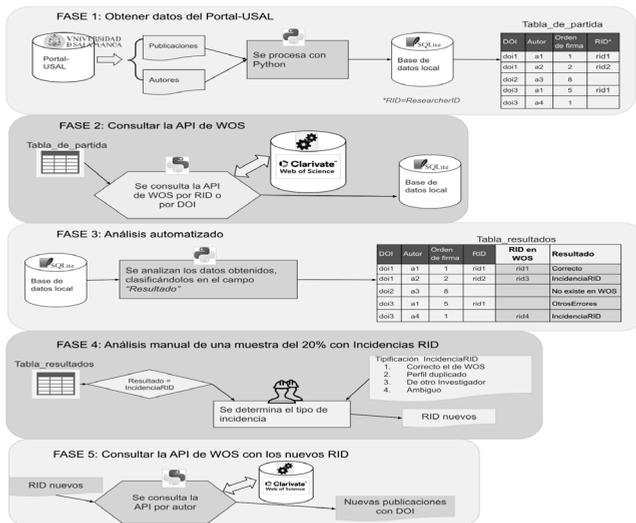


Figura 1. Fases de la metodología de la investigación
 Fuente: elaboración de los autores (2024)

Los datos fueron tabulados tipificando los resultados y comparando los datos de cada registro del Portal USAL con los obtenidos en la consulta a la API de WoS. Se organizaron en cuatro categorías: a) DOI no existe, b) DOI y ResearcherID (RID) coinciden, c) incidencia por identificadores no coincidentes o ausentes, y d) errores menores. El archivo principal con los datos obtenidos mostró a los autores con alguna información que requería revisión (*Tabla 1*).

El análisis manual se aplicó a una muestra de los autores con incidencias RID obtenidas automáticamente en las fases anteriores. Los autores se dividieron en dos grupos dependiendo de si disponían o no de identificador RID definido en el Portal USAL. Cada grupo se subdividió en otros dos, uno con los autores con un único RID en WoS en todas sus publicaciones y otro con los autores que tenían más de uno. Se seleccionó un 20 % de cada uno de los grupos descritos, descartando a los autores con menos de cinco publicaciones. Las incidencias en las que el RID hallado en WoS no coincidía con el del Portal USAL se clasificaron en cuatro posibles estados:

1. ResearcherID correcto en WoS con inconsistencias en el Portal USAL (no definido o diferente).
2. ResearcherID de perfil duplicado del autor en WoS.
3. ResearcherID de otro investigador.
4. Ambigüedad en la identificación del perfil del autor.

Posteriormente, se realizó una nueva consulta a la API de WoS con los identificadores ResearcherID de la muestra seleccionada cuyas incidencias fueron clasificadas dentro de los estados 1 o 2, ya que, en ambos casos, eran ResearcherID de investigadores de la USAL, aunque en el estado 2 su producción apareciera dispersa en WoS. De esta forma, se obtuvieron de WoS las publicaciones con DOI y se comprobó si existían o no en el Portal USAL.

PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

La metodología habilitó la obtención de identificadores ResearcherID que no estaban definidos en el Portal USAL, así como la verificación de los existentes. También se hallaron investigadores con múltiples perfiles en WoS, de los cuales bien eran perfiles duplicados o de otro autor. Se localizaron publicaciones diferentes identificadas con el mismo DOI, las cuales fueron excluidas del análisis. Los resultados afectaban al 59 % (2631) del personal investigador de la USAL y al 41 % (38 773) de su producción científica. De los investigadores seleccionados, el 45 % (1 196) tenía un identificador ResearcherID en el Portal USAL (*Tabla 2*).

PORTAL_ID	ALIASFIRMA	RID	O_CLAIM	W_DOCS	W_DOIS	W_RID	W_CLAIM	W_FIRMA	RESULTADO	PUBLICACIONES
107673	Lozano, E.						TRUE	Lozano, Elisa	Incidencia_RID	34
57640	Franco, M.	D-1752-2011	TRUE	157	113				Orden_excedido	2
57640	Franco, M.	D-1752-2011	TRUE	157	113	D-1752-2011	TRUE	Martin-Carrasco, Manuel	En_diferente_orden	4
57640	Franco-Martín, M.	D-1752-2011	TRUE	157	113	D-1752-2011	TRUE	Franco-Martín, Manuel	Correcto	82
57640	Franco-Martín, M.A.	D-1752-2011	TRUE	157	113	IKV-9682-2023	FALSE	Franco-Martín, Manuel A.	Incidencia_RID	3
57642	Orfao, A.	B-5801-2017	TRUE	963	607				Orden_excedido	6
57642	Orfao, A.	B-5801-2017	TRUE	963	607	B-5801-2017	TRUE	Orfao, Alberto	En_diferente_orden	6
57642	Orfao, A.	B-5801-2017	TRUE	963	607	B-5801-2017	TRUE	Orfao, Alberto	Correcto	539
57642	Orfao, J.A.	B-5801-2017	TRUE	963	607		FALSE	Alberto Orfao, Jose	Incidencia_RID	1
57642	Orfao, A.	B-5801-2017	TRUE	963	607		FALSE	Orfao, Alberto	Incidencia_RID	137

Tabla 1. Ejemplo del archivo de autores con algún dato que corregir
Fuente: elaboración de los autores (2024)

Portal de la Investigación USAL		Totales	Seleccionados
Publicaciones	Totales	94 717	38 773
	Con DOI repetido	466	0
Investigadores	Totales	4 446	2 631
	Con ResearcherID	1 246	1 196

Tabla 2. Datos del Portal de la Investigación USAL
Fuente: elaboración de los autores (2024)

Localización de publicaciones

El estudio mostró que el 20 % de las publicaciones no existía en WoS por pertenecer a revistas no indexadas en esta base de datos. El 35 % tenía algún autor sin RID en el Portal USAL. El 45 % eran publicaciones con todos sus autores USAL con RID definido en el Portal USAL, de ellas el 33 % presentó incidencias RID debido a alguna de las siguientes causas: el RID en el Portal USAL era obsoleto, el autor en WoS estaba duplicado, la publicación estaba mal imputada o porque había discrepancias en el orden de las firmas (*Figura 2*).

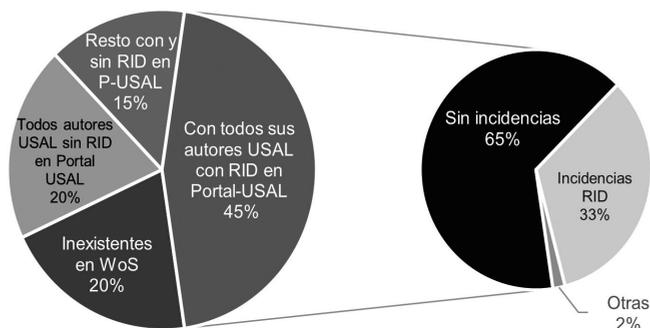


Figura 2. Publicaciones con DOI analizadas
Fuente: elaboración de los autores (2024)

Uso de identificadores

El 13 % de los investigadores no disponía de publicaciones en WoS. Entre los investigadores sin el identificador ResearcherID en el Portal USAL, el 52 % tenía todas sus publicaciones en WoS con el mismo identificador, por tanto, podrían extraerse automáticamente para ser incorporados en el Portal USAL. De los investigadores con el identificador ResearcherID en el Portal USAL únicamente el

28 % no presentaba incidencias, en el otro 72 % había discrepancias entre los datos externos de WoS y los internos del Portal USAL (*Tabla 3*).

Investigadores		%
Sin publicaciones en WoS	344	13 %
Con publicaciones en WoS. Con RID en Portal USAL	1 115	42 %
Con publicaciones en WoS. Sin RID en Portal USAL	1 172	45 %
Total	2 631	100
Investigadores. Con publicaciones en WoS. Con RID en Portal USAL		
		%
Sin incidencias	316	28 %
Con incidencias RID. Solo un RID en WoS	351	31 %
Con incidencias RID. Varios RID en WoS	407	37 %
Otras incidencias	41	4 %
Total	1 115	100
Investigadores. Con publicaciones en WoS. Sin RID en Portal USAL		
		%
Con incidencias RID. Solo un RID en WoS	610	52 %
Con incidencias RID. Varios RID en WoS	554	47 %
Otras incidencias	8	1 %
Total	1 172	100

Tabla 3. Datos de investigadores del Portal USAL analizados
Fuente: elaboración de los autores (2024)

En consonancia con el objetivo específico 2 de esta investigación, el estudio mostró que en el 72 % de los investigadores del Portal USAL con publicaciones en WoS e identificador RID en el Portal USAL había discrepancias entre ambas plataformas, ya fuera por identificación incorrecta de autorías, por duplicación de identificadores de un mismo autor o por la asignación de publicaciones a investigadores que no habían participado en estas.

Los resultados del análisis de los datos de los investigadores pueden desagregarse para mostrar las diferencias entre las incidencias de investigadores con o sin RID en el Portal USAL (*Figura 3*).

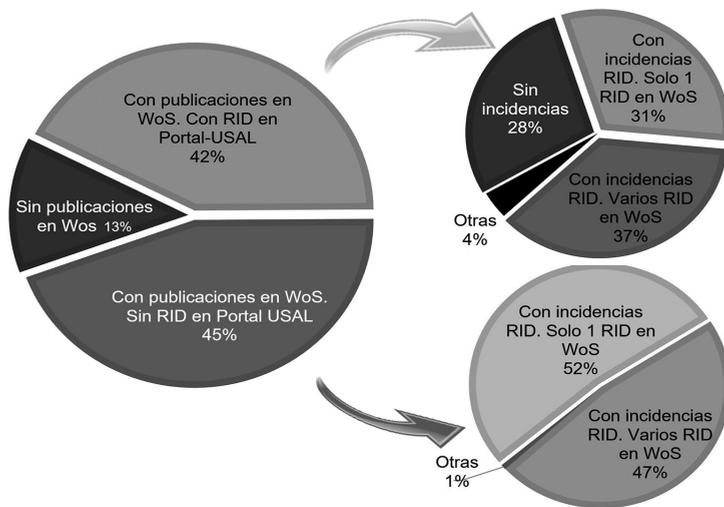


Figura 3. Investigadores analizados
Fuente: elaboración de los autores (2024)

Nuevas publicaciones localizadas

Un resultado significativo de la investigación fue la detección de la producción científica de la USAL en WoS que todavía no estaba incorporada en el portal institucional. Esto incrementaría en 11.1 % las publicaciones en el Portal USAL procedentes de WoS (Tabla 4).

Publicaciones con DOI halladas en WoS	Total	%
Documentos con DOI de investigadores USAL hallados en WoS	34 757	100 %
Presentes en Portal USAL	30 900	88.9 %
No presentes en Portal USAL	3 857	11.1 %

Tabla 4. Publicaciones USAL halladas en WoS
Fuente: elaboración de los autores (2024)

El estudio localizó publicaciones en WoS que aún no estaban incorporadas al Portal USAL e, igualmente, identificó publicaciones que no aparecían correctamente ligadas a la Universidad de Salamanca.

Tipos de incidencias

El estudio manual singularizado realizado en la cuarta fase de la investigación clasificó los identificadores ResearcherID de los investigadores de la USAL encontrados en WoS que presentaban incidencias. El número de autores con incidencias RID fue de 1 922 (*Tabla 3*), de los cuales se revisaron 385, correspondiente al 20% del total. Se hallaron 761 identificadores RID –que es un número superior al de investigadores de la muestra– porque un mismo investigador puede disponer de varios ResearcherID. Los datos de la muestra pueden extrapolarse al total de los investigadores USAL con publicaciones en WoS (2 287) (*Tabla 5*).

Incidencias en la muestra	RID	Investigadores de la muestra afectados		Extrapolación de los investigadores con publicaciones en WoS
Totales	761	385	%	2287
1. RID correcto en WoS. No informado o erróneo en el Portal USAL	262	262	68.05 %	57.19 %
2. RID de perfil duplicado en WoS	312	208	54.03 %	45.40 %
3. RID de otro investigador	161	87	22.60 %	18.99 %
4. Ambigüedad en la identificación	26	9	2.34 %	1.96 %

Tabla 5. Incidencias en ResearcherID en la muestra
Fuente: elaboración de los autores (2024)

El estudio habilitó la identificación de diferencias en la fuente de datos y la estimación de las incidencias, relacionadas al objetivo específico 1 de la investigación, entre las que destacan la creación de nuevos ResearcherID (45.40 %), aunque ya existieran perfiles anteriores verificados por los investigadores, así como la asignación incorrecta de documentos a autores a partir de sus apellidos o sus filiaciones institucionales (18.99 %).

Incremento de publicaciones en el portal institucional

El análisis de los identificadores RID de los 385 investigadores de la muestra aportó 574 nuevos RID que no estaban disponibles en el Portal USAL. Una segunda consulta a la API de WoS con los nuevos RID posibilitó localizar 6 301 publicaciones con DOI, de las cuales 1 360 no estaban en el Portal USAL. La extrapolación de estos datos permitió deducir que el volumen de publicaciones con DOI existentes se incrementaría en torno al 17.51 %. Los datos desagregados

de los tipos de incidencias ofrecen una comparación entre las inconsistencias correspondientes a los autores con o sin RID en el Portal USAL, lo cual desarrolla el objetivo específico 3 de esta investigación (Figura 4).

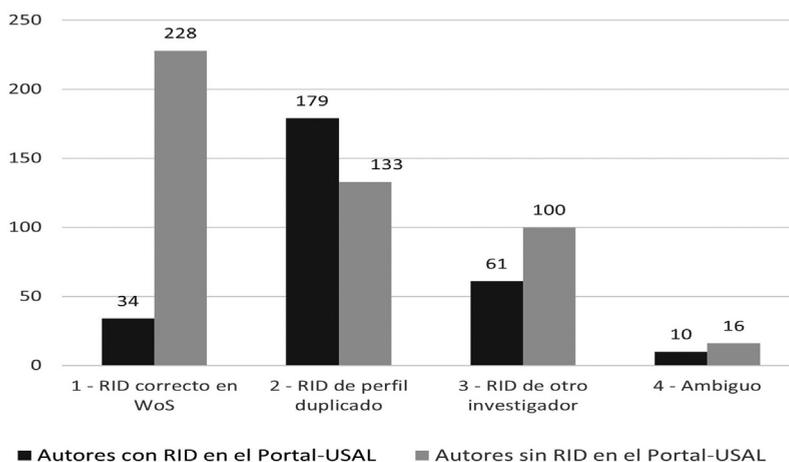


Figura 4. Tipos de incidencias RID en la muestra
Fuente: elaboración de los autores (2024)

DISCUSIÓN

La investigación encontró que existen investigaciones de autores de la Universidad de Salamanca publicadas en revistas indexadas en WoS que no están asignadas a la USAL. El número de publicaciones procedentes de WoS podría incrementarse en un 17.51 %, con esto podemos afirmar que una revisión de los datos de origen es necesaria. Los hallazgos de esta investigación detectaron problemas similares a los señalados por Olivas-Ávila y Musi-Lechuga (2014) en torno a la necesidad de implementar medidas que garanticen la precisión de ResearcherID, así como a los percibidos por Boudry y Durand-Barthez (2020) en relación con la dificultad para encontrar información de la producción científica de los investigadores debido a los cambios, las similitudes y las diferencias en la asignación de los nombres y afiliación de los autores.

El estudio determina que hay un número significativo de diferencias en relación con los datos propios de la USAL que requieren de un proceso de revisión por parte de los investigadores o por los gestores de los portales de producción científica. La investigación se ha realizado con una intención eminentemente práctica para diseñar un método que localizara errores en WoS y en el portal institucional para protocolizar el control de calidad con el cual corregir

las inconsistencias encontradas. Esta investigación es replicable en cualquier universidad que mantenga un portal propio de producción científica. Este estudio ha mostrado la necesidad de verificar la consistencia de la información procedente de plataformas externas.

A su vez, una importante aportación de esta investigación ha sido la programación de la extracción de los datos, con *scripts* en Python para la descarga, la integración y el análisis de datos, útiles para la importación y la exportación de información bibliográfica. El procedimiento puede emplearse con cualquier base de datos bibliográfica que use identificadores persistentes de autores y de documentos y que, además, proporcione una API para consultar sus datos.

Web of Science es un recurso útil para la ingesta de datos en los portales institucionales que requiere de un procedimiento como el utilizado en esta investigación para garantizar la consistencia de los datos de autores y publicaciones. Por tanto, sigue siendo necesario que plataformas como WoS supervisen la calidad de la información que ofrecen, ya que se ha demostrado que el control de los sistemas de identificación facilitará las actividades de monitorización e integración de los resultados de investigación en portales institucionales. Consideramos que, igualmente, la revisión de la calidad de los datos de la producción científica de las universidades es responsabilidad de las bibliotecas o departamentos que dan mantenimiento a los portales institucionales de producción científica. La investigación también corrobora la apreciación de Krauskopf y Salgado (2023) sobre la responsabilidad compartida de los gestores de revistas en lo que respecta a la exactitud y coherencia de los datos publicados.

CONCLUSIONES Y RECOMENDACIONES

El objetivo general del estudio fue evaluar la consistencia de fuentes de datos externas, tomando el caso de WoS, para integrar sus contenidos en portales institucionales, concretamente en el Portal de la Investigación de la Universidad de Salamanca. La investigación evidencia que el Portal USAL incrementaría sus contenidos en un 17.51 % con la detección de nuevos identificadores ResearcherID de WoS y resolviendo las incidencias halladas en ambas plataformas, tanto en WoS como en el Portal USAL. La conclusión general es que, para garantizar que los perfiles de investigadores en portales institucionales de producción científica sean correctos, se requiere la revisión y resolución de inconsistencias procedentes de las fuentes externas integradas.

La investigación permite afirmar que es necesario administrar los portales de forma autónoma, revisando los datos cosechados mediante API o exportaciones. Se propone que una vez localizada la información que debe modificarse, las

universidades establezcan canales de comunicación entre sus investigadores y los gestores de identificadores persistentes, como ResearcherID, y que dispongan de procedimientos de revisión para conseguir que toda la producción científica esté identificada y correctamente asignada, que todos los autores tengan un único identificador y que su producción esté reunida bajo el mismo perfil de autor. De esta manera, la presencia y el impacto de los investigadores en el ecosistema científico será mayor y su producción no estará dispersa a causa de inconsistencias y ambigüedades en la identificación de sus publicaciones.

Agradecimientos

Esta investigación se llevó a cabo gracias al apoyo brindado por el Servicio de Archivos y Bibliotecas y por los Servicios Informáticos de la Universidad de Salamanca.

REFERENCIAS

- Birkle, Caroline, David Pendlebury, Joshua Schnell y Jonathan Adams. 2020. "Web of Science as a Data Source for Research on Scientific and Scholarly Activity". *Quantitative Science Studies* 1 (1): 363-76.
https://doi.org/10.1162/qss_a_00018
- Borrego, Ángel. 2014. *Sistemas de identificación unívoca de investigadores*. Madrid: Conferencia de Rectores de las Universidades Españolas (CRUE); Red de Bibliotecas Universitarias Españolas (Rebiun).
<http://hdl.handle.net/20.500.11967/43>
- Boudry, Christophe, y Manuel Durand-Barthez. 2020. "Use of Author Identifier Services (ORCID, ResearcherID) and Academic Social Networks (Academia.edu, ResearchGate) by the Researchers of the University of Caen Normandy (France): A Case Study". *PLoS One* 15 (9), e0238583.
<https://doi.org/10.1371/journal.pone.0238583>
- Calderón-Rehecho, Andoni. 2017. "Portales de producción científica". *Anuario ThinkEPI* 11: 247-55.
<https://doi.org/10.3145/thinkepi.2017.45>
- Craft, Anna. 2020. "Managing Researcher Identity: Tools for Researchers and Librarians". *Serials Review* 46 (1): 44-49.
<https://doi.org/10.1080/00987913.2020.1720897>
- Cujba, Rodica. 2019. "The Role of Persistent Identifiers in E-Science". *Journal of Social Sciences* 2 (4): 40-46.
<https://doi.org/10.5281/zenodo.3550705>
- Davis, Rebecca, y Laura Saunders. 2020. "Essential Skills for Corporate and Special Librarians". *Journal of Library Administration* 60 (7): 762-83.
<https://doi.org/10.1080/01930826.2020.1786984>
- Donner, Paul. 2017. "Document Type Assignment Accuracy in the Journal Citation Index Data of Web of Science". *Scientometrics* 113 (1): 219-36.
<https://doi.org/10.1007/s11192-017-2483-y>

- Franceschini, Fiorenzo, Domenico Maisano y Luca Mastrogiacomio. 2016. "Empirical Analysis and Classification of Database Errors in Scopus and Web of Science". *Journal of Informetrics* 10 (4): 933-53.
<https://doi.org/10.1016/j.joi.2016.07.003>
- Harrison, Andrew Marc, y Anthony Mark Harrison. 2016. "Necessary but Not Sufficient: Unique Author Identifiers". *BMJ Innovations* 2 (4): 141-43.
<https://doi.org/10.1136/bmjinnov-2016-000135>
- Hernández-Sampieri, Roberto, y Christian Paulina Mendoza Torres. 2018. *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. Ciudad de México: McGraw Hill Education.
- Hoffman, Starr, ed. 2016. *Dynamic Research Support for Academic Libraries*. Londres: Facet Publishing.
<https://doi.org/10.29085/9781783301225>
- Iribarren-Maestro, Isabel. 2018. "Bibliometría y bibliotecas universitarias: ¿matizando el perfil profesional?". *Anuario ThinkEPI* 12: 142-51.
<https://doi.org/10.3145/thinkepi.2018.15>
- Jubb, Michael. 2016. "Libraries and the Support of University Research". En *Quality and the Academic Library: Reviewing, Assessing and Enhancing Service Provision*, editado por Jeremy Atkinson, 143-56. Cambridge, Massachusetts: Chandos Publishing.
<https://doi.org/10.1016/B978-0-12-802105-7.00014-2>
- Krauskopf, Erwin, y Mauricio Salgado. 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): 129-44.
<https://doi.org/10.22201/iibi.24488321xe.2023.96.58784>
- Lazonder, Ard, y Noortje Janssen. 2022. "Quotation Accuracy in Educational Research Articles". *Educational Research Review* 35, e100430.
<https://doi.org/10.1016/j.edurev.2021.100430>
- Mallery, Mary. 2016. "Scholarly Identification Systems in a Global Market: The ORCID Solution". *International Information and Library Review* 48 (4): 269-73
<https://doi.org/10.1080/10572317.2016.1243962>
- Medina-González, Alejandro, Carlos Díaz-Redondo, Blanca Rodríguez-Bravo y José Antonio Frías. 2024. "Análisis de la producción científica de la Universidad de Salamanca indexada en Scopus (2010-2015)". *Información, Cultura y Sociedad* (50): 49-67.
<https://doi.org/10.34096/ics.i50.13697>
- Mering, Margaret. 2017. "Correctly Linking Researchers to Their Journal Articles: An Overview of Unique Author Identifiers". *Serials Review* 43 (3-4): 265-67.
<https://doi.org/10.1080/00987913.2017.1386056>
- Moral-Muñoz, José Antonio, Enrique Herrera-Viedma, Antonio Santisteban-Espejo y Manuel Jesús Cobo. 2020. "Software Tools for Conducting Bibliometric Analysis in Science: An Up-to-Date Review". *El Profesional de la Información* 29 (1), e290103.
<https://doi.org/10.3145/epi.2020.ene.03>
- Olivas-Ávila, José Alonso, y Bertha Musi-Lechuga. 2014. "Validez y fiabilidad del Researcher ID y de 'Web of Science Production of Spanish Psychology'". *International Journal of Clinical and Health Psychology* 14 (1): 58-66.
[https://doi.org/10.1016/S1697-2600\(14\)70037-7](https://doi.org/10.1016/S1697-2600(14)70037-7)
- Penteado Filho, Roberto de Camargo, y Wilson Corrêa da Fonseca Júnior. 2017. "O problema da padronização das afiliações de autores na base de dados Web of Science: o caso Embrapa e sua solução". Número especial, *Em Questão* 23: 74-93.
<https://doi.org/10.19132/1808-5245230.74-93>

- Sandberg, Jane, y Qiang Jin. 2016. "How Should Catalogers Provide Authority Control for Journal Article Authors? Name Identifiers in the Linked Data World". *Cataloging and Classification Quarterly* 54 (8): 537-52.
<https://doi.org/10.1080/01639374.2016.1238429>
- Santos, Thamyres Vieira dos, y Giovana Deliberali Maimone. 2023. "Recuperação da informação e visibilidade científica sob a perspectiva de identificadores persistentes". *Revista Ibero-Americana de Ciência da Informação* 16 (3): 518-42.
<https://doi.org/10.26512/rici.v16.n3.2023.44053>
- Silva, Eduardo Graziosi, y José Augusto Chaves Guimarães. 2023. "O escritório de comunicação científica como perspectiva de atuação para bibliotecas universitárias". *Ibersid: Revista de Sistemas de Información y Documentación* 17 (1): 47-58.
<https://doi.org/10.54886/ibersid.v17i1.4917>
- Torres-Salinas, Daniel, y Wenceslao Arroyo-Machado. 2022. "APIs en contextos bibliométricos: introducción básica y corpus exhaustivo". *Anuario ThinkEPI* 16, e16a09.
<https://doi.org/10.3145/thinkepi.2022.e16a09>
- Tran, Ngoc-Yen, y Emily Chan. 2020. *Supporting Scholarly Research: Current and New Opportunities for Academic Libraries*. Choice, Asociación de Bibliotecas Universitarias y de Investigación.
<https://www.choice360.org/research/supporting-scholarly-research-current-and-new-opportunities-for-academic-libraries/>
- Velez-Estevez, Antonio, Ignacio Javier Perez, Pablo García-Sánchez, José Moral-Munoz y Manuel Jesús Cobo. 2023. "New Trends in Bibliometric APIs: A Comparative Analysis". *Information Processing and Management* 60 (4), e103385.
<https://doi.org/10.1016/j.ipm.2023.103385>
- Wang, Qi, y Ludo Waltman. 2016. "Large-Scale Analysis of the Accuracy of the Journal Classification Systems of Web of Science and Scopus". *Journal of Informetrics* 10 (2): 347-64.
<https://doi.org/10.1016/j.joi.2016.02.003>
- Zhu, Junwen, Inmaculada, Guangyuan Hu y Weishu Liu. 2019. "DOI Errors and Possible Solutions for Web of Science". *Scientometrics* 118 (2): 709-18.
<https://doi.org/10.1007/s11192-018-2980-7>

Para citar este texto:

- Bravo-García, Inmaculada, José Antonio Merlo Vega y Valeria Montoya-Roncancio. 2024. "Precisión de los perfiles de autor en portales institucionales de producción científica". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 38 (101): 145-162.
<http://dx.doi.org/10.22201/iibi.24488321xe.2024.101.58897>