

JULIO/SEPTIEMBRE, 2023  
VOL. 37, NÚM. 96, MÉXICO  
ISSN 0187-358X

# Investigación Bibliotecológica

ARCHIVONOMÍA, BIBLIOTECOLOGÍA E INFORMACIÓN



# Investigación Bibliotecológica

ARCHIVONOMÍA, BIBLIOTECOLOGÍA E INFORMACIÓN

Vol. 37, núm. 96, julio/septiembre, 2023, México, ISSN: 2448-8321  
DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2023.96>



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

Instituto de Investigaciones Bibliotecológicas y de la Información

# Contenido

INVESTIGACIÓN BIBLIOTECOLÓGICA, VOL. 37, NÚM. 96, JULIO/SEPTIEMBRE, 2023, MÉXICO, ISSN: 2448-8321  
DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2023.96>

## ARTÍCULOS

- **Trayectoria y análisis del concepto de cultura escrita** [Trajectory and analysis of the concept of written culture] 13-26  
*Agustín Vivas Moreno y Carmen Solano Macías*
- **How Libraries in Indonesia Use TikTok: A Content Analysis of Library Accounts on the Platform** [Cómo las bibliotecas en Indonesia utilizan TikTok: Un análisis de contenido de las cuentas de bibliotecas en la plataforma] 27-44  
*Rahmat Fadhlil, Cecep Ibrahim, Miftahunnisa' Iqiriza and Bachrul Ilmi*
- **La gobernanza de los datos de investigación en el contexto de su organización y sistematización** [Governance of research data in the context of its organization and systematization] 45-64  
*Eder Ávila-Barrientos*
- **La valoración de ejemplares dedicados y firmados en bibliotecas y colecciones patrimoniales** [Valuation of inscribed and signed copies in libraries and heritage collections] 65-86  
*Manuel-José Pedraza-Gracia*
- **Las necesidades de información y comportamiento informativo de los usuarios de sistemas de información geográfica en comunidades de Facebook** [The information needs and information behavior of geographic information systems users in Facebook communities] 87-108  
*Eduardo Waybel Sánchez*
- **La ciencia abierta y su relación con la innovación: una revisión bibliométrica** [Open Science and Its Relationship to Innovation: A Bibliometric Review] 109-128  
*Felipe Guevara-Peoz*
- **Inconsistency in the registration of the Digital Object Identifier (DOI) of articles on Web of Science and Scopus** [Inconsistencia en el registro del Identificador de Objetos Digitales (DOI) de artículos en Web of Science y Scopus] 129-144  
*Erwin Krauskopf and Mauricio Salgado*

- **Veinticinco años de investigación en redes sociales: evolución de temas entre 1997 y 2021 empleando el algoritmo Asignación Latente de Dirichlet** [Twenty-five Years of Research in Social Networks: Evolution of Topics between 1997 and 2021 Based on Latent Dirichlet Allocation (LDA)] 145-177  
*Juan-Antonio Martínez-Comeche*
- **Productividad científica de investigadores de la Universidad Nacional de Asunción (UNA): comprobación del modelo de Lotka** [Scientific productivity of researchers at the National University of Asunción (UNA): Testing Lotka's model] 179-200  
*Emilce Sena Correa, María Luisa Lascurain, Mariana Cáceres Ruiz Díaz y Johana Raquel Pineda Alvarenga*
- **Arquivos contestados em contexto intranacional Em torno do caso da Região Autónoma da Madeira (Portugal)** [Subnational disputed archives Around the case of the Autonomous Region of Madeira (Portugal)] 201-218  
*L. S. Ascensão de Macedo*



Investigación Bibliotecológica: archivonomía, bibliotecología e información [en línea] / ed. por el Centro Universitario de Investigaciones Bibliotecológicas. -Vol. 1, No. 1 (ago. 1986) - . Ciudad de México: Universidad Nacional Autónoma de México, Centro Universitario de Investigaciones Bibliotecológicas, 1986 - V. Semestral, 1986 - 2007; a partir de vol. 22, no. 44 (ene. / abr. 2008) - , la periodicidad cambió a cuatrimestral.  
A partir de enero de 2018, la periodicidad cambió a trimestral.  
Resúmenes en español e inglés, a partir del vol. 3, no. 1 (jul / dic. 1987)  
Disponible también en idioma inglés a partir del vol. 28, no. 62 (ene. / abr. 2014)  
Publicado por la misma dependencia bajo su nombre actual: Universidad Nacional Autónoma de México, Instituto de Investigaciones Bibliotecológicas y de la Información vol. 26, no. 56 (ene. / abr. 2012) - .  
Disponible para su consulta en línea a partir del vol. 1, no. 1 (ago. 1986) - .  
Publicado en formato electrónico a partir del vol. 30, no. 70 (sep. / dic. 2016) - .  
Todos los artículos cuentan con DOI en forma individual.  
Disponible en: <http://rev-ib.unam.mx/ib>  
ISSN 0187-358X (impreso)  
ISSN 2448-8321 (en línea)



Investigación Bibliotecológica: archivonomía, bibliotecología e información, Vol. 37, No. 96, julio-septiembre 2023, es una publicación trimestral editada por la Universidad Nacional Autónoma de México, Ciudad Universitaria, Alcaldía Coyoacán, Ciudad de México, C.P. 04510, a través del Instituto de Investigaciones Bibliotecológicas y de la Información, Circuito Escolar s/n, Torre II de Humanidades, Piso 12, Ciudad Universitaria, Col. Copilco, Alcaldía Coyoacán, Ciudad de México, C.P. 04510, Tel. 55 5623 0349 y 55 5623 0337, <http://rev-ib.unam.mx/ib>, correos electrónicos: [revista@iibi.unam.mx](mailto:revista@iibi.unam.mx), [drevista@iibi.unam.mx](mailto:drevista@iibi.unam.mx). Editor responsable: Dr. Egbert John Sánchez Vanderkast, Reserva de Derechos al uso Exclusivo No. 04-2016-041813344600-203, ISSN: 2448-8321, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número, Dra. María de Jesús Madera Jaramillo, Instituto de Investigaciones Bibliotecológicas y de la Información, Circuito Escolar s/n, Torre II de Humanidades, Piso 12, Ciudad Universitaria, Col. Copilco, Alcaldía Coyoacán, Ciudad de México, C.P. 04510, Tel. 56230337. Fecha de la última modificación, septiembre 2023.

Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor de la publicación. Se autoriza la reproducción total o parcial de los textos aquí publicados siempre y cuando se cite la fuente completa y la dirección electrónica de la publicación.

DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2023.96>

#### REVISTA INDIZADA EN:

- |                                        |                  |
|----------------------------------------|------------------|
| •Clarivate Analytics<br>Web of Science | •Latindex        |
| •Clarivate Analytics JCR               | •DOAJ            |
| •Scopus                                | •Elsevier        |
| •SCImago Journal<br>& Rank             | Science Direct   |
| •Índice de Revistas<br>Mexicanas       | •LISA            |
| de Investigación                       | •LISTA Full Text |
| Científica y Tecnológica               | •LISS            |
| del CONACyT                            | •INFOBILA        |
| •SciELO Citation Index                 | •CSIC e-revist@  |
| •SciELO                                | •Dialnet         |
| •Google Academics                      | •CLASE           |
|                                        | •HAPI            |

Esta revista está disponible en texto completo y en acceso abierto en:

- Revista IIBI: <http://rev-ib.unam.mx/ib>
- SciELO: [http://www.scielo.org.mx/scielo.php?script=sci\\_serial&pid=0187-358X&lng=es&nrm=iso](http://www.scielo.org.mx/scielo.php?script=sci_serial&pid=0187-358X&lng=es&nrm=iso)
- Science Direct: <https://www.sciencedirect.com/journal/investigacion-bibliotecologica-archivonomia-bibliotecologia-e-informacion>

#### DIRECTOR DE LA REVISTA

DR. EGBERT JOHN SÁNCHEZ VANDERKAST

#### CONSEJO EDITORIAL

DRA. ESTELA MORALES CAMPOS  
Universidad Nacional Autónoma de México  
Ciudad de México, México

DR. CARLOS ADOLFO PIÑA GARCÍA  
Universidad Veracruzana  
Xalapa, México

PhD BIRGER HJØRLAND  
University of Copenhagen  
Copenhagen, Dinamarca

DRA. GLORIA PÉREZ SALMERÓN  
Stichting IFLA Foundation  
La Haya, Holanda

DRA. FERNANDA RIBEIRO  
University of Porto  
Porto, Portugal

DR. ELÍAS SANZ CASADO  
Universidad Carlos III  
Madrid, España



## COLABORADORES EN ESTE NÚMERO

INVESTIGACIÓN BIBLIOTECOLÓGICA, vol. 37, núm. 96, julio/septiembre, 2023, México,  
ISSN: 2448-8321

### *Agustín Vivas Moreno*

Grupo de Investigación ANDROPA  
Departamento de Información y Comunicación  
Universidad de Extremadura  
Plaza de Ib Marwán s/n 06071  
Badajoz, España  
Tel. +34 9 2428 9300  
aguvivas@unex.es

### *Bachrul Ilmi*

Department of Library Science  
Vocational School, Universitas Sebelas Maret  
Ir. Sutami Street 36 Kentingan, Jebres,  
Surakarta, Central Java, Indonesia  
Tel. +62 822 7872 6278  
bachrul.ildi@staff.uns.ac.id

### *Carmen Solano Macías*

Grupo de Investigación ANDROPA  
Departamento de Información y Comunicación  
Universidad de Extremadura  
Plaza de Ib Marwán s/n 06071  
Badajoz, España  
Tel. +34 9 2428 9300  
csolano@unex.es

### *Cecep Ibrahim*

Department of Library and Information Science  
Faculty of Social and Political Sciences  
Halu Oleo University  
Kambu, Kendari City, South East Sulawesi  
Indonesia  
Tel. +62 813 8356 5369  
cecep.ibrahim@uho.ac.id

### *Eder Ávila-Barrientos*

Instituto de Investigaciones Bibliotecológicas  
y de la Información  
Universidad Nacional Autónoma de México  
Torre II de Humanidades, piso 12, Circuito  
Escolar, Ciudad Universitaria, Coyoacán,  
C. P. 04510, Ciudad de México, México  
Tel. +52 55 5623 0380  
eder@iibi.unam.mx

### *Eduardo Waybel Sánchez*

Facultad de Filosofía y Letras  
Universidad Nacional Autónoma de México  
Colorines 8. Magdalena Contreras  
C. P. 16640, Ciudad de México, México  
Tel. +52 55 2497 4428  
eduardowaybel@gmail.com

### *Emilce Sena Correa*

Facultad Politécnica  
Universidad Nacional de Asunción  
Campus de la Universidad Nacional  
de Asunción, San Lorenzo, C. P. 111434,  
Departamento Central, Paraguay  
Tel. +59 52 1588 7000  
esena@pol.una.py

### *Erwin Krauskopf*

Vicerrectoría de Investigación,  
Universidad de las Américas  
Manuel Montt 948, Providencia,  
Santiago, Chile  
Tel. +56 2 2253 1634  
ekrauskopf@udla.cl

### *Felipe Guevara-Pezoa*

Dirección de Investigación y Postgrado,  
Vicerrectoría Académica,  
Universidad Central de Chile  
Toesca 1783, Santiago, Chile  
Tel. +56 2 2582 6189  
Felipe.guevara@ucentral.cl

### *Johana Raquel Pineda Alvarenga*

Facultad Politécnica  
Universidad Nacional de Asunción  
Campus de la Universidad Nacional  
de Asunción, San Lorenzo, C. P. 111434,  
Departamento Central, Paraguay  
Tel. +59 52 1588 7000  
jpineda@pol.una.py

*Juan-Antonio Martínez-Comeche*  
Departamento de Biblioteconomía y  
Documentación, Facultad de Ciencias  
de la Documentación, Universidad  
Complutense de Madrid, C/ Santísima Trinidad  
37, 28010, Madrid, España  
Tel. +34 9 1394 6665  
juaamart@ucm.es

*L. S. Ascensão de Macedo*  
Centros de Estudios Clásicos, Faculdade  
de Letras, Universidade de Lisboa,  
Alameda da Universidade, 1600-063,  
Lisboa, Portugal  
Tel. +35 12 1792 0005  
laureanomacedo@edu.ulisboa.pt

*Manuel-José Pedraza-Gracia*  
Departamento de Ciencias de la Documentación  
e Historia de la Ciencia, Instituto de Patrimonio  
y Humanidades, Universidad de Zaragoza  
Facultad de Filosofía y Letras. C/San Juan Bosco,  
7, 50009, Zaragoza, España  
pedraza@unizar.es

*María Luisa Lascurain*  
Facultad de Humanidades, Comunicación  
y Documentación, Universidad Carlos III  
de Madrid, C/Madrid, 126. 28903 Getafe,  
Madrid, España  
Tel. +34 9 1624 9249  
mlascura@bib.uc3m.es

*Mariana Cáceres Ruiz Díaz*  
Facultad Politécnica  
Universidad Nacional de Asunción  
Campus de la Universidad Nacional  
de Asunción, San Lorenzo, C. P. 111434,  
Departamento Central, Paraguay  
Tel. +59 52 1588 7000  
mcaceresrd@pol.una.py

*Miftahunnisa' Igriza*  
Department of Library Science  
Faculty of Law, Social and Political Sciences,  
Indonesia Open University  
Pd Cabe Raya Street, Ciputat, Tangerang  
Selatan, Banten, Indonesia  
Tel. +62 812 8869 7735  
Miftahunnisa.igriza@ecampus.ut.ac.id

*Mauricio Salgado*  
Centro de Estudios Públicos  
Monseñor Sótero Sanz 162,  
Providencia, Santiago, Chile  
Tel. +56 22 2328 2400  
msalgado@cepchile.cl

*Rahmat Fadhli*  
Research Group of Library Management,  
Department of Educational Management,  
Faculty of Educational Sciences and  
Psychology, Universitas Negeri Yogyakarta  
Colombo Street Yogyakarta No. 1, Karang  
Malang, Caturtunggal, Depok, Sleman  
Regency, Special Region of Yogyakarta,  
Indonesia  
Tel. +62 812 3440358  
rahmat.fadhli@uny.ac.id



## Trayectoria y análisis del concepto de cultura escrita

Agustín Vivas Moreno\*  
Carmen Solano Macías\*

*Artículo recibido:*  
2 de febrero de 2023  
*Artículo aceptado:*  
27 de abril de 2023

*Artículo de investigación*

### RESUMEN

La presente investigación tiene por objeto examinar las líneas evolutivas del concepto de cultura escrita. Se parte de las nociones de paleografía y alfabetización hasta llegar a la actual visión multidisciplinaria abierta a enfoques tan diversos como los que suscitan los cambios históricos en la definición del lector y las prácticas de lectura. Se observa en la trayectoria una ampliación del concepto de cultura escrita motivada fundamentalmente por la incorporación de principios pertenecientes a las disciplinas comunicativas, lingüísticas y sociológicas, lo que obliga a replanteamientos conceptuales en torno a tres principios: dialogismo, performatividad y activismo. El concepto de cultura escrita se inserta, en

\* Grupo de Investigación ARDOPA. Departamento de Información y Comunicación  
Universidad de Extremadura, España    aguvivas@unex.es    csolano@unex.es

consecuencia, en el marco de comunidades interpretativas conforme a los nuevos lenguajes multimodales que fomentan las prácticas culturales híbridas.

**Palabras clave:** Cultura escrita; Escritura; Lectura; Prácticas culturales

### Trajectory and analysis of the concept of written culture

*Agustín Vivas Moreno and Carmen Solano Macías*

#### ABSTRACT

This research aims to examine the evolutionary lines of the concept of written culture. It starts from the notions of paleography and literacy until reaching the current multidisciplinary vision open to approaches as diverse as those caused by historical changes in the definition of the reader and reading practices. An expansion of the concept of written culture is observed in the trajectory, motivated fundamentally by the incorporation of principles belonging to the communicative, linguistic and sociological disciplines, which forces conceptual rethinking around three principles: dialogism, performativity and activism. The concept of written culture is inserted, consequently, in the framework of interpretive communities according to the new multimodal languages that promote the hybrid cultural practices.

**Keywords:** Written culture; Writing; Reading; Cultural practices

#### INTRODUCCIÓN: LA CULTURA ESCRITA COMO ESTUDIO POLIÉDRICO

Intentar realizar una retrospectiva sobre el concepto de cultura escrita resulta ser una empresa extraordinariamente arriesgada. La amplitud temática y su profunda transversalidad son elementos que lo dificultan sobremanera. Sin embargo, entendemos que conviene hacer un ejercicio de carácter sistemático para determinar las líneas evolutivas del concepto a examinar. De esta forma, vislumbrar el trayecto recorrido y advertir las nuevas configuraciones serán los ejes en los que basaremos nuestro estudio.

No pretendemos hacer un ejercicio eminentemente histórico, pues toda retrospectiva deviene en análisis del presente. Así, cualquier examen del pasado conceptual de la cultura escrita resulta ser una interpretación de nuestro contexto cultural; y cualquier intento de comprensión de sus renovaciones conceptuales es resultado del análisis retrospectivo previamente confeccionado.

Los procesos de lectura y escritura resultan ser maneras esenciales de transmisión de la información. Naturalmente, de esta situación se deriva la ingente cantidad de estudios al respecto. Posiblemente, en ello intervenga el cambio de prototipo que se viene percibiendo gracias a las nuevas formas de comunicación.

En la actualidad nos es espinoso imaginar la historia de la humanidad sin la cultura escrita; no obstante, ello comenzó solamente hace cinco milenios y su desarrollo posterior, vinculado exclusivamente a las élites, suministrándoles poder (Viñao Frago, 1984). Sin embargo, en las páginas que siguen entenderemos que el concepto de cultura escrita debe sobrepasar diáfananamente el de las élites alfabetizadas.

Quizás debiéramos comenzar por una pregunta: ¿qué entendemos por cultura escrita? Si analizamos su atracción debiéramos retrotraernos al siglo ilustrado, cuando se desenvuelven la Paleografía y la Diplomática como disciplinas autónomas. Naturalmente, estos inicios vinculados a los caracteres formales poco tienen que ver con el proyecto transdisciplinar que se desarrolla hoy con múltiples matices y tonos.

Los estudios de cultura escrita se vieron desarrollados, en las últimas dos décadas del s. XX, con relaciones con otras disciplinas históricas, sociológicas, educativas o lingüísticas. Sin embargo, posiblemente sean las disciplinas documentales, desde la famosa obra de Otlet, las que han desplegado de manera ingente la noción de documento, identificándolo, en última instancia, con información mantenida en cualquier soporte. Ello, observaremos, tiene importantes consecuencias para la delimitación del concepto de cultura escrita.

Pero el desarrollo no se sostuvo ahí. El nuevo paradigma transdisciplinar ha convertido el estudio de lo escrito en un análisis complejo y sumamente heterogéneo, donde intervienen una variada amplitud de disciplinas científicas (Kallman, 2008). Las contribuciones de todas ellas son válidas, perfeccionando y ampliando el examen conceptual, hasta tal punto que deviene en ser una práctica cultural de enorme multiplicidad, con contribuciones de diferentes aplicaciones metodológicas.

De todas formas, para hacer un estudio de la temporalidad, conviene comenzar por el principio, y ello nos deviene en el análisis de dos materias: la Paleografía y el Alfabetismo (Castillo Gómez y Sáez Sánchez, 1994)

Seguiremos, pues, una metodología con un doble eje: por un lado, la selección y el análisis comparativo del *paradigma clásico* de la cultura escrita con el que

podríamos denominar *paradigma social con sus renovaciones historiográficas*; y, en consonancia con ello, la hermenéutica de dicho análisis mediante el examen de determinadas imbricaciones socio-culturales, con el objeto de percibir la compleja configuración del concepto y evolución de la cultura escrita.

## EVOLUCIÓN DEL CONCEPTO DE CULTURA ESCRITA

### *La Paleografía*

Para realizar cualquier análisis sobre la cultura escrita resulta imprescindible hacerlo, en primer lugar, sobre la Paleografía, dado que tradicionalmente ha sido percibida como una disciplina encargada del análisis de la escritura y sus signos (Petrucci, 1994). El fundamento de dicha disciplina consiste en entender los textos escritos, para lo cual resulta imprescindible cierta dosis de habilidad con el objetivo de conocer los sistemas de escritura, sus abreviaturas o las especificidades correspondientes para los diversos momentos históricos.

Como es sabido, el origen de la Paleografía como disciplina se encuentra vinculado con la Diplomática, siendo J. Mabillon (1681) quien desplegaría una metodología específica. La evolución posterior naturalmente desarrolló esta sistemática intentando dilucidar, para cada caso, cuándo surge lo escrito, cómo nace y dónde se engendra. No será sino hasta el s. XIX, centuria en la que se establecen institutos específicos relacionados con la Historia como disciplina, cuando la Paleografía adquiere una resonancia mayor. Surgen ahora colecciones documentales, el despliegue de la prensa histórica o el desarrollo de facsímiles fotográficos, lo cual ocasiona, como decimos, un avance cualitativo de la Paleografía de la mano de estudiosos como Ludwig Traube, Leopold Delisle, Edward Bond, Cesare Paoli o Bernhard Bischoff (García Tato, 2009).

Sin embargo, será en el s. XX cuando la Paleografía adquiera el carácter plenamente científico con autores como Giulio Battelli, Luigi Schiaparelli o Giorgio Cencetti. Éstos y otros dotan a la Paleografía de un cuerpo teórico con principios y métodos manifiestos y esclarecidos que son respetados en diferentes latitudes. En la década de los 30, sin embargo, tiene lugar un nuevo avance con M. Cohen, A. G. Gieysztor e István Hajnal. En ellos se observa una influencia marxista, lo cual induce a establecer una correlación entre la disciplina paleográfica y las características sociales en las que se desarrolla el texto escrito (Hajnal, 1952). En este orden, la escritura se configura como algo indisociable del entorno social en el que se ocasiona y desarrolla (Petrucci, 1979).

Más adelante, se crea el Comité Internacional de Paleografía. Sus inicios se dan en 1953, llegando hasta hoy mismo, si bien, desde 1985 se denomina Comité

Internacional de Paleografía Latina (CIPL). En la década de los años 60, influenciados por la Escuela de los *Annales*, se expresan nuevos recursos conceptuales derivados de la epistemología del nuevo concepto de *documento*. Quizás los máximos exponentes sean, entre otros, Lucien Febvre, Fernand Braudel, Jacques Le Goff y Roger Chartier.

Con todo, se observa que las disciplinas paleográficas no resultan apropiadas para el análisis del entorno social en el que se genera lo escrito, de tal forma que será A. Petrucci quien, junto con otros, se plantee novedosas interrogantes correspondientes esencialmente al último responsable de lo escrito, y a las causas y los fundamentos del mismo. La función de lo escrito adquiere, de esta forma, una notoriedad inusitada.

Naturalmente, todo ello conllevó modernos análisis, flamantes imbricaciones con el hecho social y, en última instancia, nuevos métodos hermenéuticos para el estudio del contexto social y sus vinculaciones con lo escrito y sus usos (Castillo Gómez y Sáez Sánchez, 1994). La denominación no es unívoca, si bien, lo más aceptado sea *Alfabetismo y cultura escrita* (Cassany, 2005). En última instancia, lo escrito transitó de ser algo estático y objeto firme de estudio de características meramente documentales, a ser algo dinámico y maleable, de propiedades vinculadas a la investigación social.

Como fruto de esta nueva tendencia, en 1977 tuvo lugar el congreso sobre *Alfabetismo e cultura scritta nella storia della società italiana*. Fue aquí donde, desde el contexto italiano, surgen nuevas propuestas y se amplían los soportes de la cultura escrita. El despliegue posterior vino acompañado por algunos seminarios permanentes, la prestigiosa revista *Scrittura e Civiltà* dirigida por Giuglielmo Cavallo, Alessandro Pratesi y el propio Petrucci, la inauguración de centros de investigación tales como el Archivio Diaristico Nazionale en Pieve Santo Stefano, en 1984; el Archivio della Scrittura Popolare en Trento, 1987; el Archivio Ligure della Scrittura Popolare en Génova, 1988, o el Archivio per la Memoria e la Scrittura delle Donne, Florencia, 1998. En España, por su parte, la Red de Archivos e Investigadores de la Escritura Popular, así como la Asociación Paleográfica Internacional: Cultura, Escritura, Sociedad (APICES). Y todo ello, aderezado con los estudios sobre alfabetismo como concepto historiográfico.

### ***El alfabetismo***

El concepto de *alfabetismo*, como intentaremos descifrar a continuación, ha tenido un despliegue parecido al de la Paleografía, pasando de ser una disciplina que examina las capacidades lecto-escritoras a algo más ambiguo, transversal y complejo, vinculándose a las imbricadas relaciones sociales y culturales.



Ciertamente, el interés por los estudios sobre alfabetización fue muy tardío en España, y naturalmente las nuevas tendencias no llegaron hasta bien entrada la década de los 70 del s. XX (Viñao Frago, 2009).

Con todo, el análisis sobre el hecho alfabetizador no es sencillo, sino todo lo contrario. La causa es clara: entre lo alfabetizado y lo analfabetizado hay amplias y sugerentes zonas de semialfabetización dificultosas de desentrañar. Sin descender a ello, cabe decir que los estudios sobre la firma (Soubeyroux, 1998) o los análisis acerca de la ausencia de escolarización y su vinculación con la falta de alfabetización (Luzuriaga, 1926; Olóriz, 1900), tan prolijos como controvertidos, son, a todas luces, insuficientes para evaluar los diferentes grados de alfabetización. Ello, como veremos en páginas posteriores, será bien contextualizado por los nuevos paradigmas de lo escrito.

En consecuencia, lo alfabetizado tiene un grave componente social que se presenta de forma voluble y gradual. No se trata, pues, de un cuerpo sólido de estudio, sino, podríamos decir, de un proceso en el que participan multitud de agentes, circunstancias sociales, intereses de poder, ideologías y, en última instancia, factores socio-culturales con abundantes imbricaciones.

Conviene quizás, para una mejor comprensión del devenir de la alfabetización, mencionar a Harvey J. Graff (1987), quien asevera que en el despliegue de los análisis sobre la alfabetización hay tres *generaciones* (Graff, 1987):

- La 1ª generación, correspondiente a la década de los 60 del s. XX, supuso el comienzo de los estudios sobre historia de la alfabetización. Se trata de una pléyade de autores que vislumbran que se encuentran ante una nueva disciplina científica (Stone, 1964; Schofield, 1968; Goody, 1963). De ahí que reconozcan y describan las fuentes documentales al respecto, realicen las primeras series cronológicas o confeccionen los pioneros análisis de carácter socio-cultural (Castillo Gómez y Sáez Sánchez, 1994).
- La 2ª generación, situada en los años 70 y los 80, tiene como rasgo identificativo la cada vez mayor vinculación entre los estudios de la alfabetización y determinados factores socio-culturales con los que se corresponden. De este modo, se pone en cuestión lo realizado en los años 60 y la disciplina se despliega entre nuevos parámetros como los diferentes modos y prácticas de lectura, con cardinales influencias de la Escuela de los *Annales*. Se trata de una generación abundantemente nutrida (Furet y Sachs, 1974; Clanchy, 1975; Cressy, 1977; Soltow y Stevens, 1977; Houston, 1982; Darnton, 1986; Johansson, 1987), fijándose como elemento clave la denominación de cultura escrita o la ampliación de los estudios, de forma transversal, a otras disciplinas.

- La 3ª generación, que fue trazada de manera muy sucinta por Graff, tiene su origen al finalizar la década de los 80 (Zemon Davis, 1993; Peredo Merlo, 1997). La perspectiva cultural, que origina multitud de estudios contextuales, y la interdisciplinariedad, son, posiblemente, las características que mejor describen este tiempo. Los efectos de lo escrito en el nuevo contexto social resultan ser intrincados por su versatilidad e inestabilidad. Por otra parte, los estudios de “literacidad” se hacen notorios, al vincularse con cierto énfasis, los conceptos de lectura e identidad social. Encontramos análisis que refieren las relaciones de los diferentes usos y prácticas de lo escrito para los diferentes conjuntos sociales (Street, 1984; Gee, 1990; Heath, 1991; Barton, D., 2012; Hamilton, M., Roz, 2000). Se trata, entendemos, de la base de la renovación historiográfica que veremos a continuación, donde las TICs y su imperativo tecnológico han abierto nuevos condicionamientos socio-culturales que se extienden al concepto de alfabetización.

## LA RENOVACIÓN HISTORIOGRÁFICA Y LOS NUEVOS ESTUDIOS

El *nuevo orden social* en el que nos encontramos también ha alcanzado, obviamente, a los estudios sobre el concepto de cultura escrita. Las nuevas apreciaciones del proceso alfabetizador han implicado el surgimiento de distintos enfoques que vienen determinando el polifacético devenir de su concepto (Cressy, 1980).

Intentemos simplificar algunas ideas, a riesgo de esquematizar lo que por esencia es múltiple.

### ***Los nuevos estudios de literacidad***

Tal y como lo hemos descrito, las investigaciones que se ocasionaron desde la década de los 70 admitieron implantar novedades en los análisis sobre alfabetización. Indicaremos dos: por un lado, la alfabetización como proceso, que permitió descubrir su riqueza conceptual, la variedad de modelos y la complejidad contextual; y, por el otro, los nuevos objetos de estudio, donde los análisis se multiplican sectorialmente, vinculando la historia del libro y de la lectura como elementos diferenciados, y donde los enfoques antropológicos y sociológicos cobran sentido.

En última instancia, el alfabetismo deviene en ser un estudio de los *procesos* alfabetizadores y sus múltiples y heterogéneas conexiones. Los estudios de literacidad, asimismo, acaban vinculándose con nuevos procesos que veremos a continuación.

### ***La nueva perspectiva socio-cultural***

Las contribuciones de la Escuela de los *Annales*, las denominadas “escrituras populares” (Lyons, 2007), y otros estudios culturales, se vieron renovados desde otros paradigmas diferentes. Quizás la orientación más amparada por los investigadores sea la literacidad de carácter sociocultural, en contra de otros enfoques que ponían mayor énfasis en aspectos lingüísticos y psicológicos. Este paradigma se basa en un análisis de la cultura escrita donde el lector existe en un contexto, está inserto en una comunidad con sus usos particulares de lectura y escritura, y, en consecuencia, percibiendo la literacidad como un fenómeno cultural concreto (Heath, 2000). Como resultado, nacen análisis denominados como *New Literacy Studies*, que conceptualizan la lectura como una práctica social contextualizada, muy diferentes de la concepción tradicional que no atiende a contextos ni pautas sociales.

Otra de las características de este nuevo paradigma es su enfoque integrador, de forma que la alfabetización se configura como un conjunto de prácticas sociales imbricadas entre sí. Los investigadores de esta corriente, B. Street (1984), J. P. Gee (1990), P. Conti, G. Franchini, Gibelli (2002), Lyons (2007), entienden que la literacidad no es una aptitud independiente, sino un conjunto de instrumentos comunicativos de una comunidad específica y en un contexto particular.

En definitiva, podemos indicar que no hay una única literacidad o una literacidad canónica, sino que ésta es plural y múltiple. Según investigadores de la Universidad de Ottawa, podemos encontrar tres “literacías” elementales: la escolar, la comunitaria y la personal (Cassany, 2006). Todas ellas alfabetizan de forma diferente a tenor del contexto. Así, los logros en la literacidad no devienen de resortes meramente académicos desvinculados del contexto ciudadano y familiar. Por otro lado, las inteligencias múltiples confirman estas múltiples “literacías”. Lo cierto es que, en el contexto digital, globalizado y líquido en el que nos encontramos inmersos, el concepto de “persona letrada” es cambiante, mudable y repleto de nuevas perspectivas.

### ***El nuevo enfoque socio-comunicativo***

Para comprender esta perspectiva, debemos tener en cuenta el concepto de “campo”, del sociólogo francés P. Bourdieu (1983), que parte de la similitud con la física. Así, reveló que los campos sociales son espacios de juego históricamente constituidos con leyes de funcionamiento propias. En este orden, tal y como sucede en un juego, subyacen reglas y confrontaciones, obteniéndose un sistema dinámico y activo.

De este modo, la cultura escrita se configura aquí no como un elemento estático y unívoco, sino repleto de multiplicidades que interactúan entre ellas, ya sean legitimadas por los grupos hegemónicos o por los grupos subalternos. El campo cultural será, en consecuencia, el espacio resultante de todas estas mediaciones. Así, por ejemplo, Martos y Vivas (2010) describieron la configuración de la cultura escrita en el *Quijote* según los criterios de Bourdieu, distinguiendo cuatro macrocampos o dominios: la cultura de creación o de innovación, la orientada a la instrucción-socialización, la cultura vinculada a la preservación-memoria y, finalmente, la cultura material, que trata de la producción o los soportes. Los cuatro dominios poseen sus oportunos agentes semiótico-materiales, si bien lo sustancial son las relaciones y sus múltiples combinaciones. Las diferentes prácticas diferencian la cultura principal o “hegemónica” de aquellas otras secundarias o “subalternas”.

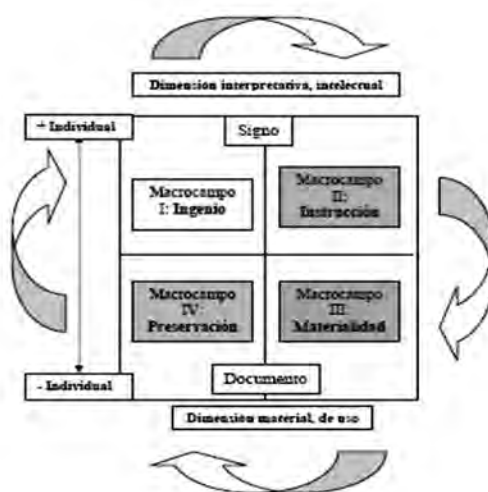


Figura 1: Macrocampos de la cultura escrita, según Bourdieu.

Fuente: Vivas Moreno y Martos García

En definitiva, en la teoría de campos no interesa tanto la figura o posición aislada, sino la red de correspondencias, concomitancias y la representación mental que acaba configurándose (Bourdieu, 1983; Aguirre Romero, 2005).

Este “dinamismo conflictivo” se encuentra vinculado con los nuevos estudios de literacidad. Según Barton y Hamilton (2000), en la cultura escrita hay unas prácticas más dominantes que otras. Así, la construcción de la identidad letrada es el vector resultante del “sistema de vectores fuerza”. La conducta comunicativa entre ellos determina el molde de la práctica letrada determinante, la que

ocasiona, por ejemplo, que el estudiante prime los apuntes y los manuales como artefactos casi únicos frente a la pluralidad de voces.

### ***Nuevas posibilidades de la cibercultura: un concepto extendido de la Cultura Escrita***

Como intentaremos describir a continuación, la cultura letrada tradicional descansa en un sistema jerárquico y hegemónico.

Ciertamente, “cibercultura” es un término que se utiliza para describir la cultura y las prácticas sociales que han surgido como resultado del uso generalizado de las tecnologías digitales, esto es, de cómo la tecnología ha influido en la forma en que las personas interactúan entre sí, cómo se comunican, trabajan, aprenden y se divierten. Naturalmente, la cibercultura abarca un amplio rango de actividades, desde la creación de comunidades en línea y las redes sociales, hasta el comercio electrónico, la ciberseguridad, la privacidad en línea, la educación a distancia, y la creación y distribución de contenido digital. También se relaciona con cuestiones más amplias de política y sociedad, como la regulación de internet, la propiedad intelectual y la libertad de expresión. Todo ello ha transformado las formas tradicionales de comunicación, trabajo y entretenimiento.

La cultura letrada y la cibercultura difieren en los medios de comunicación, la producción cultural, el acceso a la información, la comunicación, la velocidad de la información y la interacción social. Si la primera se basa en un sistema bien estratificado y los canales de comunicación resultan estar bien delimitados, en la cibercultura deambulamos hacia un modelo circular y rizomático (Deleuze y Guattari, 1980). La “ciudad letrada” no es simple, no tiene centros, no es lineal, no presenta jerarquías, sino que es compleja, descentralizada, dinámica y comunicativa.

En este orden, la cultura escrita resulta ser ahora un mosaico con ramificaciones heterogéneas. La cultura local, el enfoque performativo, la alfabetización informacional o la industria del entretenimiento son nuevos escenarios que surgen como elementos de las relaciones subyacentes que están teniendo lugar en este dinamismo que siempre es circunstancial y líquido.

Son destacables las obras de P. Lévy, filósofo y teórico de la información, conocido por sus trabajos en el campo de la cibercultura, la inteligencia colectiva y la filosofía de la tecnología (1997 y 2011). Dicho autor explora la relación entre la tecnología, la cultura y la sociedad, y argumenta que la tecnología digital está transformando la forma en que pensamos, creamos y compartimos conocimiento y donde la inteligencia colectiva resulta ser una fuerza transformadora en la sociedad.

Intentemos sistematizar el concepto de cultura escrita en la era digital:

1. La cultura escrita en el marco de las comunidades interpretativas: internet desarrolla una dinámica participativa. Los nuevos lectores se apropian de los clásicos conforme a los nuevos lenguajes multimodales, esto es, aquellos que utilizan múltiples representaciones para transmitir información o significado, lo que permite una comunicación más compleja.
2. El público homogéneo ha dejado paso a comunidades interpretativas estratificadas que comparten lecturas conformando múltiples subculturas. La red como “mapa del tesoro” donde cada uno traza sus itinerarios, búsquedas o cartografías. El mediador-profesor tiene un nuevo cometido: acompañar para interpretar las lecturas salvajes o enredadas a las que se refiere Chartier. Estas comunidades se nutren por analogías de identidad (género, edad, etc.).
3. Importancia de las prácticas de lectura y escritura, y sus contextos: la lectura virtual se torna en tarea constructiva, donde lo importante no es el soporte sino los usos y prácticas de lectura y escritura, y los contextos e interacciones que median entre ellas. Se trata, en definitiva, de una construcción interpersonal (Jenkins, 2008).
4. La percepción de la cibercultura como aliada de la cultura letrada. Se trata del problema descrito sobre cómo transitar de una “lectura salvaje” a una letrada. Daniel Link percibe que internet y la cultura letrada no son adversarias. El medio para establecer alianzas es fomentar prácticas culturales “anfíbias”, híbridas, eclécticas (Link, 1997).
5. Necesidad de replanteamientos básicos: el nuevo contexto nos obliga a repensar la noción de cultura escrita desde un punto de vista holístico en torno a tres principios: dialogismo, performatividad y activismo. El concepto de cultura letrada implica desarrollar múltiples voces y competencias. Las prácticas correspondientes a la cultura letrada no pertenecen ya al ámbito académico o formal, sino que cada vez se abren a entornos simulados y juegos de rol. Los usos antiguos quedan deslegitimados. Queda por ver dónde queda todo ello (Bajtin, 1974).

En conclusión, lo sustancial no es el objeto tecnológico sino los nuevos usos de lectura y escritura que devienen, y el novedoso carácter híbrido, extendido e integrador que se desprende. El concepto clásico de cultura escrita, según algunos autores, se ve beneficiado.

## CONCLUSIONES

Son tres las conclusiones a las que podemos llegar tras la retrospectiva realizada:

En primer lugar, el concepto de cultura escrita resulta ser polifacético, multiforme y complejo. En su transcurrir, transita de una estructura tradicional que descansaba en un sistema jerárquico y hegemónico vinculado a las élites, a otra de carácter circular, rizomática, descentralizada, dinámica y comunicativa. Se observa, pues, una evolución desde los conceptos de *alfabetización* y *paleografía* hasta llegar a la actual visión multidisciplinaria con perspectivas desiguales para delimitar al lector y sus formas y prácticas. Los elementos determinantes de esta evolución han sido los contextos de la alfabetización, la construcción de la textualidad, los usos sociales de la lengua escrita, las relaciones entre la oralidad y la escritura, los diferentes métodos descriptivos de representación, el determinismo tecnológico, los múltiples procesos de propagación, admisión y privación del texto, y, la más sustancial, la focalización actual del fenómeno lector como factor socio-cultural.

Por otra parte, actualmente nos hallamos en un nuevo paradigma repleto de incertidumbres. De este modo, en el análisis de la cultura escrita han formado parte, de forma transdisciplinar, cuestiones antropológicas, lingüísticas, literarias, sociales e históricas. Con todo ello, los estudios sobre cultura escrita han devenido en ser una disciplina propia y transdisciplinar tras un desarrollo de casi medio siglo. Durante estos últimos 30 años se ha pasado de una descripción del fenómeno del analfabetismo a la renovación historiográfica desde nuevas disciplinas que incorporaron nuevas variables, hasta llegar a la extensión del concepto con parámetros tocantes al universo de la comunicación y de la filosofía.

Finalmente, hemos de señalar que las nuevas corrientes suponen un alcance extendido del concepto de *cultura escrita*. Ésta se halla inserta en el marco de comunidades interpretativas conforme a los nuevos lenguajes multimodales. En este ámbito, la lectura se transforma en una tarea constructiva interpersonal, donde lo importante no es el soporte sino las prácticas de lectura y escritura, y los contextos e interacciones que median entre ellas. En consecuencia, la cibercultura no es adversaria de la cultura letrada sino un medio de fomentar prácticas culturales anfíbias e híbridas. Todo ello obliga a replanteamientos conceptuales en torno a tres principios: dialogismo, performatividad y activismo.

*Agradecimientos:*

Grupo de Investigación ARDOPA: Proyecto Ayuda para la realización de actividades de investigación y desarrollo tecnológico, de divulgación y de transferencia del conocimiento por los Grupos de Investigación de Extremadura, y se aprueba la primera convocatoria Junta de Extremadura, y Fondos FEDER

## REFERENCIAS

- Aguirre Romero, J. 2005. "La incidencia de las Redes de comunicación en el Sistema literario." *Espéculo*, n. 7.
- Barton, D. 2012. "Redefining vernacular literacies in the Age of Web 2.0". *Applied linguistics*, v. 33, n. 3, 282-298.
- Batjín, M. 1974. *La cultura popular en la Edad Media y el Renacimiento*. Barcelona: Barral Editores.
- Bourdieu, P. 1983. *Espacio social y campo de poder*. Barcelona: Anagrama.
- Cassany, D. 2005. "Investigaciones y propuestas sobre literacidad actual: multilateralidad, Internet y criticidad". Conferencia presentada en el *Congreso Nacional Cátedra UNESCO para la Lectura y la Escritura*, Universidad de Concepción.
- Castillo Gómez, A., coord. 2002. *Historia de la cultura escrita: del próximo Oriente Antiguo a la sociedad informatizada*. Gijón: Trea.
- Castillo Gómez, A. 1995. "De la Paleografía a la Historia: de las prácticas del escribir". *Historia a Debate*, v. II, 261-271.
- Castillo Gómez, A., y C. Sáez Sánchez. 1994. "Paleografía versus alfabetización: reflexiones sobre historia social de la cultura escrita". *Signo*, n. 1, 133-168.
- Chartier, R. 1985. *Pratiques de la lecture*. Marsella: Rivages.
- Chartier, R. 1988. *Cultural History Between Pratiques and Representations*. Cambridge: Cambridge University Press.
- Clanchy, M. T. 1975. "Modern in education and government in England". *Speculum*, v. 50, 671-688.
- Cressy, D. 1977. "Levels of illiteracy in England, 1530-1730". *The Historical Journal*, v. 20, n. 1, 1-23.
- Cressy, D. 2006. *Literacy and the social order: Reading and writing in Tudor and Stuart England*. Cambridge: Cambridge University Press.
- Darnton, R. 1986. "First Steps Towards a History of Reading". *Australian Journal of French Studies*, n. 23, 5-30.
- Darnton, R. 2006. *El negocio de la Ilustración: una historia editorial de la Enciclopedia, 1775-1800*. México: Fondo de Cultura Económica.
- Furet, F., y W. Sachs. 1974. "La croissance de l'alphabétisation en France, XVIIIe-XIXe siècle". *Annales: Histoire, Sciences Sociales*, v. 29, n. 3, 714-737.
- Gee, J. P. 1990. *Social Linguistics and Literacies: Ideology in Discourses*. Londres: Falmer Press.
- Goody, J. 1963. "The consequences of literacy". *Comparative studies in society and history*, v. 5, n. 3, 304-345.
- Graff, H. J. 1981. *Literacy in History: An Interdisciplinary Research Bibliography*. New York: Garland Publishing.
- Hajnal, I. 1952. *L'enseignement de l'écriture aux Universités Médiévales*. Budapest: Maison de l'Édition de l'Académie des Sciences de Hongrie.
- Houston, R. A. 1982. "The development of Literacy: Northern England, 1640-1750". *The Economic History Review*, v. 35, n. 2, 199-216.
- Jenkins, H. 2008. *Convergence Culture: La cultura de la convergencia en los medios de comunicación*. Barcelona: Paidós.
- Johansson, E. 1973. *Literacy and society in a historical perspective: a conference report*. Umea: University and Umea School of Education.



- Kalman, J. 2008. "Discusiones conceptuales en el campo de la cultura escrita". *Revista iberoamericana de educación*, n. 46, 107-134.
- Lévy, P. 1990. *Las tecnologías de la inteligencia. El futuro del pensamiento en la era de la informática*. Madrid: Cátedra,
- Lévy, P. 2011. *La esfera semántica*. Barcelona: Paidós.
- Link, D. 1997. "Literaturas comparadas, estudios culturales y análisis textual: por una pedagogía". *Filología*, n. 12, 514.
- Lyons, M. 2007. *Personal Narratives/Ordinary Writings. Towards a history of writing practices in the 19th and 20th century*. Berna: Peter Lang.
- Martos García, A., y A. Vivas Moreno. 2010. "Cultura escrita e historia de la cultura". *Álabe*, n. 2. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=3628346> (Consulta: 02/12/2021)
- Petrucchi, A. 1979. "Funzione della scrittura e terminologia paleografica". En: *Paleographica, Diplomatica et Archivistica: studi in onore di Giulio Battelli*. Roma: Edizioni di Storia e Letteratura.
- Schofield, R. S. 1968. "The measurement of literacy in pre-industrial England". En: *Literacy in traditional societies*, 311-325. London: Cambridge University Press.
- Solano Macías, C. s. f. "La cultura escrita en Extremadura a finales del s. XVIII a través del interrogatorio de la Real Audiencia: análisis y sistema de información histórica". Tesis doctoral. <https://dehesa.unex.es/handle/10662/5757>
- Soltow, L., y E Stevens. 1977. "Economic aspects of school participation in mid-nineteenth-century United States". *The Journal of Interdisciplinary History*, v. 8, n. 2, 221-243.
- Street, B. V. 1984. *Literacy in Theory and Practice*. New York: Cambridge University Press.
- Street, B. V., ed. 2005. *Literacy across educational contexts: mediating learning and teaching*. Philadelphia: Caslon.
- Viñao Frago, A. 1984. "Del analfabetismo a la alfabetización: análisis de una mutación antropológica e historiográfica, I". *Historia de la educación*, n. 3, 151-189.
- Vivas Moreno, A., y A. Martos García. 2010. "La cartografía conceptual y su utilidad para el estudio de la lectura como práctica histórico-cultural: El *Quijote* como ejemplo". *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, v. 24, n. 51, 95-124.
- Zemon Davis, N. 1993. *Sociedad y cultura en la Francia moderna*. Barcelona: Crítica.

### *Para citar este texto:*

Vivas Moreno, Agustín, y Carmen Solano Macías. 2023. "Trayectoria y análisis del concepto de cultura escrita". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 13-26.

<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58754>

# How Libraries in Indonesia Use TikTok: A Content Analysis of Library Accounts on the Platform

Rahmat Fadhli\*  
Cecep Ibrahim\*\*  
Miftahunnisa' Igiriza\*\*\*  
Bachrul Ilmi\*\*\*\*

*Artículo recibido:*  
23 de febrero de 2023  
*Artículo aceptado:*  
28 de abril de 2023

*Artículo de investigación*

## ABSTRACT

The objective of this study is to examine Indonesian libraries' TikTok accounts. The research was carried out on 12 TikTok accounts with a total of 688 posts to gain a better understanding of TikTok content, the type of information published, and follower engagement on posts. Data was collected by observing all TikTok accounts and posts. The result showed that most libraries began using TikTok

- \* Research Group of Library Management, Department of Educational Management, Faculty of Educational Sciences and Psychology, Indonesia  
rahmat.fadhli@uny.ac.id
- \*\* Department of Library and Information Science, Faculty of Social and Political Sciences, Indonesia  
cecep.ibrahim@uho.ac.id
- \*\*\* Department of Library Science, Faculty of Law, Social and Political Sciences, Indonesia  
Miftahunnisa.igiriza@ecampus.ut.ac.id
- \*\*\*\* Department of Library Science, Vocational School, Universitas Sebelas Maret, Indonesia  
bachrul.ildi@staff.uns.ac.id

in 2020, in response to the growing number of TikTok users and the shift during the pandemic. The most common types of published data are library collection information and other information. Announcements, services, and library facilities are the three types of information that receive the most engagement. This study has practical implications for libraries that use TikTok as a social media alternative. TikTok will allow them to reach younger audiences, boost creativity, spread joy, and save libraries' money. Based on the findings of this study, librarians can consider the format of published information. This study offers insights on the social media platform TikTok as a new opportunity to promote libraries. It emphasizes the enormous potential for using social media platforms such as this to gain more attention and bring libraries closer to patrons.

**Keywords:** Content analysis; Library; Social media

### **Cómo las bibliotecas en Indonesia utilizan TikTok: Un análisis de contenido de las cuentas de bibliotecas en la plataforma**

*Rahmat Fadbli, Cecep Ibrahim, Miftabunnisa' Igriza  
and Bachrul Ilmi*

El objetivo de este estudio es examinar las cuentas de TikTok de bibliotecas en Indonesia. La investigación se llevó a cabo en 12 cuentas de TikTok con un total de 688 publicaciones para obtener una mejor comprensión del contenido de TikTok, el tipo de información publicada y la interacción de los seguidores en las publicaciones. Los datos se recopilaban observando todas las cuentas y publicaciones de TikTok. Los resultados mostraron que la mayoría de las bibliotecas comenzaron a utilizar TikTok en 2020, en respuesta al creciente número de usuarios de TikTok y el cambio durante la pandemia. Los tipos más comunes de información publicada son información de la colección de la biblioteca y otra información. Los anuncios, servicios y espacio de la biblioteca son los tres tipos de información que reciben más interacción. Este estudio tiene implicaciones prácticas para las bibliotecas que utilizan TikTok como una alternativa de redes sociales. TikTok podrá llegar a una audiencia más joven, impulsar la creatividad, difundir la alegría y ahorrar dinero a las bibliotecas. Basándose en los hallazgos de este estudio, los bibliotecarios pueden considerar el formato de la información publicada. Este estudio ofrece perspectivas de

la plataforma de redes sociales TikTok como una nueva oportunidad para promocionar las bibliotecas. Se enfatiza el enorme potencial de utilizar plataformas de redes sociales como TikTok para llamar más la atención y acercar las bibliotecas a los usuarios.

**Palabras clave:** Análisis de contenido; Biblioteca; Redes sociales

## INTRODUCTION

In the modern digital age, many media and communication tools have been developed. Various positive impacts of the development of media and communication tools have begun to be adopted by organizations. This allows organizations or individuals to disseminate information promptly and connects people in remote locations. Social media is a platform where people connect and share information and/or feelings (Haenlein and Kaplan, 2010). Kapoor *et al.* (2018) defines social media as a variety of user-based platforms that facilitate the dissemination of exciting content, the creation of dialogue, and the user's communication to a larger audience. It is more broadly defined as a digital space created by people for people, providing a conducive environment for interaction and networking at a personal, professional, business, marketing, political, or social level.

The fast-paced nature of social media users encourages this medium to be more diverse. Social media has evolved over time from a tool for uploading and exchanging information and content to a platform for sharing information with the general public (Aichner *et al.*, 2021). This means that social media has been identified as a strategic tool for organizations to reach a large number of users. Many organizations use social media to disseminate information to patrons or customers because social interaction is becoming increasingly easier to do digitally. Social media is also widely used in government, health care, education, and even non-profit organizations. Libraries are also beginning to use social media as marketing tools. Taylor & Francis (no date) stated that more than 70% of libraries use social media tools for a variety of purposes, one of which is promotion.

For library promotion, the library has used various social media platforms such as Facebook, Twitter, and Instagram (Chan, Lam and Chiu, 2020; Rahmawati and Rahmi, 2021). According to the findings of these studies, social media is an effective tool for bringing libraries closer to their patrons. This means that the library can also increase patron engagement by collecting quick feedback from patrons of the services provided. As a result, social media is an appropriate and effective tool for connecting libraries with potential patrons. Besides,

the use of social media enhances the interactivity of library services, enables patrons to contribute to the creation of content, fosters a sense of community, and facilitates the sharing of information with the library's audience (Çakmak and Eroğlu, 2020).

The development of social media over time has also changed the pattern of social media usage in society. Libraries also need to adapt to meet the needs of their patrons. One of the growing forms of social media today is TikTok. China's video platform was launched in September 2016 to share short videos. TikTok is the social network of the moment. This app is based on the concept of short format video, and it provides patrons with a diverse selection of sounds and music, as well as the option to apply special effects and filters. The high demand for video sharing on online social networks has also led to a shift in organizational promotion through social media.

Tik Tok has also been used by libraries as a marketing tool. TikTok is a new platform that allows users to share information, interact with other users and libraries, and share flexible content with other accounts (Stephens, 2022). The value of TikTok lies in teaching, learning, and marketing library services and resources through short video content. This application can significantly help libraries use TikTok as an advocacy tool (Mashiyane, 2022). Furthermore, TikTok's existence can specifically target younger users. TikTok can also be used for outreach and marketing to introduce the library and its services (Anderson, 2020).

This study advances knowledge on how libraries use TikTok, a social media platform, for marketing and information tool purposes. Thus, this study proposes the following research questions:

- RQ 1. Which is the overview of the library's TikTok account?
- RQ 2. How is information disseminated by the library's TikTok account?
- RQ 3. How is the library's and users' response and engagement?

## METHODS

This study uses a descriptive analysis approach and content analysis to identify, describe, and analyze how libraries use TikTok as a medium of information and promotion. The object of the study is the libraries' TikTok accounts in Indonesia. A total of 486 posts by 12 library accounts in Indonesia were examined. The first step in conducting this study was to locate the libraries' TikTok accounts that would serve as the research sample. The following criteria were used in this study:

1. The library is a representative of the national, public, special, school, and academic libraries.
2. The minimum number of account followers is 100.

As a result, 14 library accounts in Indonesia were found that meet the criteria. However, two accounts were excluded from the analysis because, in addition to sharing library information, they were also used for personal purposes. The study's data collection is organized around the research question. In consequence, the data collected consists of a description of the general condition of the TikTok account, the disseminated information on each post, and the engagement of followers with the library.

The sample was described using statistical descriptive analysis in the second stage. The data described refers to the study's research question. Data is collected by identifying each post on each library's TikTok account. Posts are then coded and categorized using Microsoft Excel. The code is determined by reviewing the information or messages contained in them and then used to categorize the posts. The process of observing content is repeated several times in order to reduce inaccuracy. The data collection was conducted in January 2023, with a posting cutoff date of January 28th, 2023.

The clustering is based on groupings adapted from Al-Daihani and AlAwadhi (2015), Suharso and Muntiah (2020), and Rachman, Mutiarani and Putri (2018). The three previous studies categorized social media content on Twitter and Instagram. However, due to the increasing variety of information shared on TikTok, changes were made in this study. The following table summarizes the clusters identified in this study:

1. Announcements include content that disseminates important information about the library, for example, regulations or hours of operation.
2. News and events include content regarding documentation of activities carried out by the library or disseminating library activities to be carried out by the library or parent organization. For example: posters; library visits by students; seminars held in the library.
3. Entertainment includes content that aims to entertain users without any other purpose.
4. User greetings include content that greets or interacts with users.
5. Collection includes promoting or recommending the collection to the users.
6. Services include promoting and informing the public about the library's services.
7. Facilities encompass library tours and new facilities, as well as promotion and information about the library's spaces.
8. How-to videos involve content that instructs users on how to do something, such as borrowing materials, using the digital collection, or visiting the library during the pandemic.

9. Commemoration days, which include content acknowledging specific holidays.
10. Others may include content outside the context of the library, for example, reading tips, educationally themed content such as English lessons, or user testimonials.

Each piece of content is assigned to a single category. If a single piece of content includes information from multiple categories, the researcher groups them based on the piece's specific purpose. These twelve categories are used as a reference for clustering TikTok posts published by the library.

## RESULTS AND DISCUSSION

### *RQ. 1 Overview of the Library's TikTok account*

In the first stage, the researcher identified each chosen library's TikTok account between December 2022 and January 2023. The following is a general description of the libraries' TikTok accounts:

No	Account	Followers	Posts	Type	First Post
1	perpusnas_ri	12.500	220	National Lib	December 10, 2021
2	perpustakaan.adisutjipto	1002	63	Academic Lib	July 5, 2021
3	unairlibrary	499	62	Academic Lib	August 5, 2020
4	uinsaizulibrary	406	27	Academic Lib	November 27, 2020
5	uhwperbanaslib	264	54	Academic Lib	January 29, 2021
6	ui_library	217	10	Academic Lib	March 25, 2021
7	perpustakaan_umy	179	60	Academic Lib	July 28, 2021
8	perpus.balpem	413	4	Public Lib	November 18, 2020
9	balai_yampus	312	123	Public Lib	November 3, 2020
10	perpus.banjarkab	236	28	Public Lib	March 15, 2021
11	perpustakaanmansabdg	303	22	School Lib	June 11, 2021
12	perpustakaannetral	128	15	School Lib	October 5, 2022

Table 1. Overview of Library TikTok Account  
Data per January 28, 2023

According to *Table 1*, there is one National Library (NL) account, two school library accounts (16.67%), three public library accounts (PL) (25%), and five academic library (AL) accounts (50%). This figure contrasts sharply with the number of libraries in Indonesia, which includes 164,610 public, school, university, and special libraries. Only 0.0072% of Indonesian libraries have TikTok accounts, according to the study's criteria.

Using TikTok as a promotional medium for library information can present these institutions with a new challenge. Libraries can use the latest social media to get closer to potential and regular patrons, receive feedback more swiftly, and brand their libraries. Although the library already has social media, libraries must follow the trend of using social media based on generational characteristics. In the same way that many organizations and communities in the health sector provided health education during a pandemic, TikTok can also be used by libraries to connect with patrons and share library information.

The national library account had the most followers, with around 12.500 adherents. Meanwhile, other libraries continue to have numbers of followers ranging from 100 to 1000. According to Rachman, Mutiarani and Putri (2018), one of the most important aspects of social media use is the number of followers. With so many followers, the library has an outstanding opportunity to disseminate information in a way that enables patrons to comprehend its messages. Social media marketing has advantages over other forms of marketing.

Furthermore, the national library account has the most posts compared to other accounts. This is because the National Library of Indonesia has a service unit that focuses on community and social media. As a result, it outperforms other libraries in terms of content preparation. One of the challenges in the context of the library is the lack of social management personnel. According to Jones and Harvey (2019), one of the challenges for library staff is a lack of knowledge and willingness to invest time in managing social media. As a result, staff must be trained to use the tools to provide content tailored to the needs of patrons.

This study also identified the first post for each library's TikTok account. Since 2020, four libraries have started using their TikTok accounts, with the first being @unairlibrary on August 5, 2020 while the rest began in 2021 and 2022. These findings suggest that libraries used TikTok as one of the new informational and promotional media innovations during the pandemic. The existence of TikTok during the pandemic was translated in its being used for a variety of purposes, including disseminating health information, optimizing distance learning and deploying campaigns or facilitating education.



## RQ.2 Information types shared by the library's TikTok account

This study determined the number of posts across the libraries' TikTok accounts in response to the second research question. All posts published by each library's TikTok account are counted. The following is the outcome of grouping posts according to the researcher's clusters.

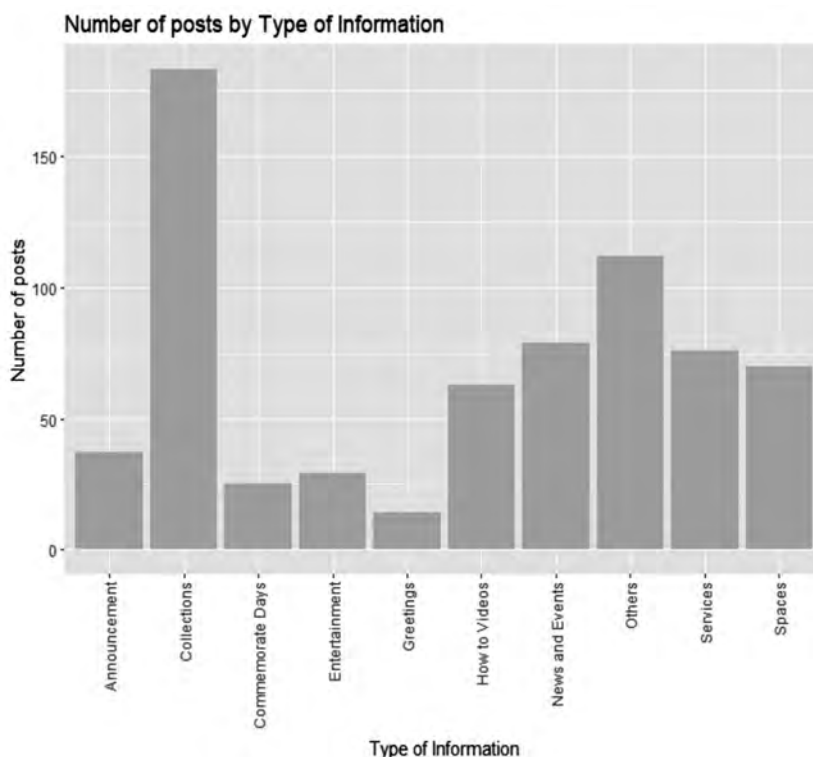


Figure 1. Type of information shared.

According to *Figure 1*, the most common type of information in the libraries' TikTok content is Collection information, which accounts for 183 posts (26.60%). The library promotes new books and reading recommendations to followers via short videos. In this study, content about collection information is presented in a variety of ways, among them displaying book covers, giving book reviews, promoting books that are currently being discussed, citing book quotes, and posting on specific anniversaries. These findings are consistent with those of Vaaler and Brantley (2016) and Wu and Yang (2021), who showed

how libraries can use social media to promote information sources owned by them. The distribution of information about the collections makes the user more aware of their existence. Furthermore, the findings of this study are consistent with the findings of a previous library TikTok study conducted by Alley and Hanshew (2022), who discovered that information about collections is one type of information frequently highlighted by library TikTok accounts in the United States.

Furthermore, the second-most shared category of information is other types of information, with a total of 112 posts (16.28%). The library provides educational content such as reading tips, tips on visiting the library during the pandemic, and other content outside the library context such as English lessons, tips for a healthy lifestyle, and environmental care campaigns. These findings emphasize that libraries can provide other types of information besides those strictly within the library context through social media platforms like TikTok, such as educational content, so that the information shared is also useful for library followers. However, libraries must also pay attention to the presentation of published information, including its quantity, because it is only secondary, not primary, information.

Libraries play a fundamental role in optimizing their social media platforms in order to disseminate information to their followers. Libraries use social media to spread information in order to increase their popularity and recognition (Quadri and Adebayo Idowu, 2016). As a result, information dissemination via library social media accounts must consider the balance of published information. According to Okike (2020), librarians should educate users by disseminating information through existing physical and digital media platforms. He goes on to say that, if the context of the information conveyed is outside the context of the library, librarians must establish working relationships with other organizations or agencies.

In this study, librarians needed to collaborate with related organizations or experts to package information for some of TikTok's content, notably educational content, especially on non-library topics. As a result, the information's validity and quality are preserved. The same is true for the subject of the library itself. Librarians must consult with related parties or service departments when disseminating information to ensure that it is not misleading. Therefore, each library is expected to be able to optimize its use of social media. This step is carried out by disseminating information on various topics.

### ***RQ 3. The library's and users' reactions and engagement***

Engagement with social media content is defined as a psychological state of mind experienced by consumers/audiences when perusing social media content which they absorb and enjoy. This condition causes interactive behavior such as

likes, shares, and comments in response to social media content (Syrdal and Briggs, 2018). In this context, patron engagement is classified as viewing/watching videos, liking, commenting on, and sharing posts.

No	Library	Average			
		View	Like	Comment	Share
1	perpusnas_ri	13,041.54	1,030.18	29.16	27.2
2	perpustakaan.adisutjipto	150.03	9.57	1.04	1.30
3	unairlibrary	965.03	41.22	1.22	3.38
4	uinsaizulibrary	146.77	8.77	0.55	0.96
5	uhwperbanaslib	156.81	3.67	0	0.29
6	ui_library	1,129.9	38.9	1.4	6
7	perpustakaan_umy	231.48	6.36	0.11	0.28
8	perpus.balpem	253.5	14	0	1
9	balai_yanpus	245.91	6.26	0.48	1.04
10	perpus.banjarkab	107.5	7.78	0	0.10
11	perpustakaanmansabdg	802	42	0.90	4.04
12	perpustakaannetral	755.86	42.06	1.13	1
<b>Average</b>		<b>1,498.86</b>	<b>104.23</b>	<b>3</b>	<b>3.88</b>

Table 2. Engagement based on library account.

According to *Table 3*, National Library accounts have the highest average engagement across all metrics: views, likes, comments, and shares. The National Library has the highest engagement in relation to its large number of followers (over 12,500). The average number of views on each account is 1,498.86. This figure is extremely high due to the large view numbers of the large national library account. If views on National Library accounts are excluded from the calculation, the average number of views per account is 449.52. The number of views indicates how many times the audience has watched a short video. As a result, this determines the extent to which content can reach both followers and non-followers.

This study also focuses on the patron engagement in terms of likes. The number of likes on National Library accounts contribute significantly to the average, which is 1,030.13 likes per post. If the National Library account is excluded from the calculation, the average of the eleven accounts is only 20.05 likes per post. This finding is also directly proportional to patron engagement in terms

of comments and shares: engagement at the National Library has a significant impact on the average comment and share. Other than the National Library, the level of engagement in libraries is very low, ranging from 0 to 1.4 comments per post and 0.1 to 4.04 shares per post. This study discovered that patrons/followers were more likely to follow or be exposed to information through the posts of the library's TikTok account, but did not respond significantly through commenting or sharing.

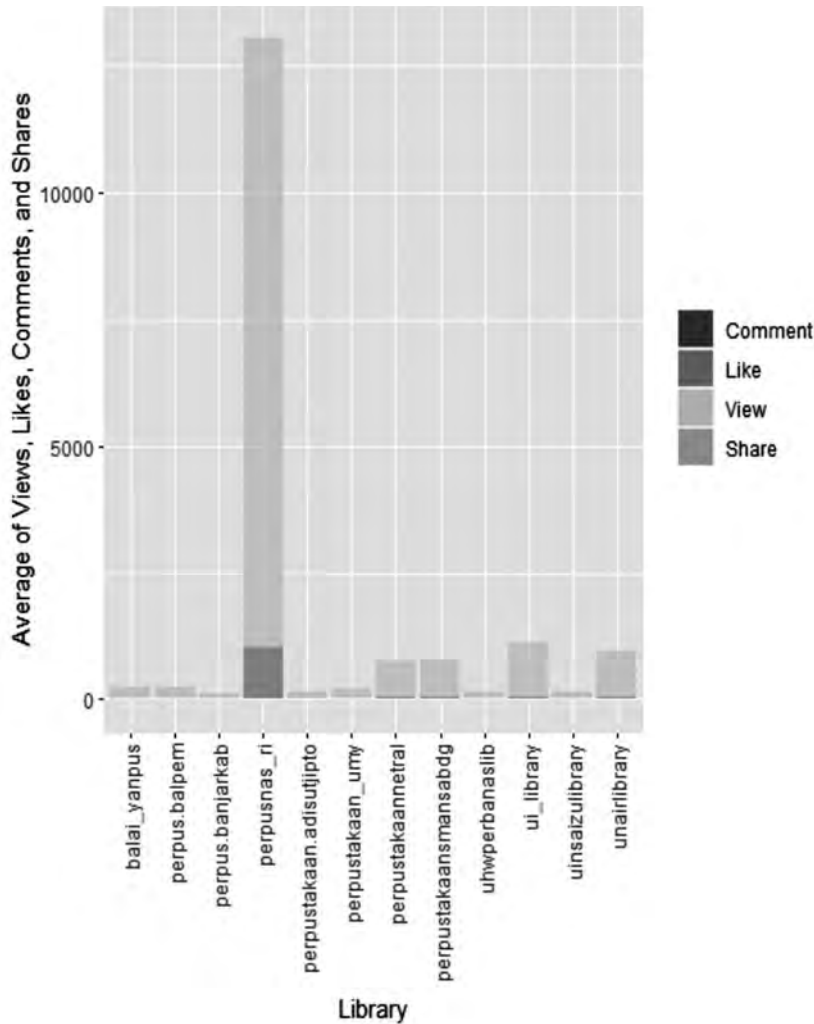


Figure 2. Engagement based on library account.

Social media allows patrons to participate and interact with their content in a variety of ways. Participation variations include likes, dislikes, comments, shares, and others. It is emphasized that it becomes a part of the patron's response and feedback to information shared through social media. Brands/organizations will benefit more from engagement, such as likes or reactions, because it indicates consumer affinity with the brand, thereby strengthening brand-consumer relationships. Likes, comments, and shares by social media followers/patrons are not the only ways to gauge the level of social media engagement. However, this must be a library concern in terms of disseminating information and developing relationships with followers.

The type of information in TikTok publications is then used to determine engagement. The outcome is as follows:

No	Type	Average			
		View	Like	Comment	Share
1	News and Events	694.92	27.58	0.84	1.75
2	Announcement	17,260.89	1,434.21	48.64	53
3	Collections	1,577.98	55.19	1.83	2.08
4	Spaces	1,3487.4	1,312.18	41.14	35.22
5	Services	4,938.28	253.38	7.94	7.71
6	How to Videos	2,905.01	172.38	4.9	8.78
7	Greetings	1,330.21	48	3.28	4
8	Commemoration Days	136.68	7.44	0.12	0.08
9	Entertainment	2,617.82	163.55	6.62	5.1
10	Others	4,076.53	363.29	4.07	3.04
Average		<b>4,902.57</b>	<b>383.72</b>	<b>11.93</b>	<b>12.07</b>

Table 3. Engagement based on type of information.

According to *Table 3*, posts with “announcement” type content receive the most engagement. The content types “spaces” and “services” come next, while the posts with the type of content that commemorates holidays are the lowest-engagement content. These results clearly show what type of content the library's social media followers prefer. Thus, libraries can use social media to disseminate information about announcements to followers, as well as to promote library services and spaces.

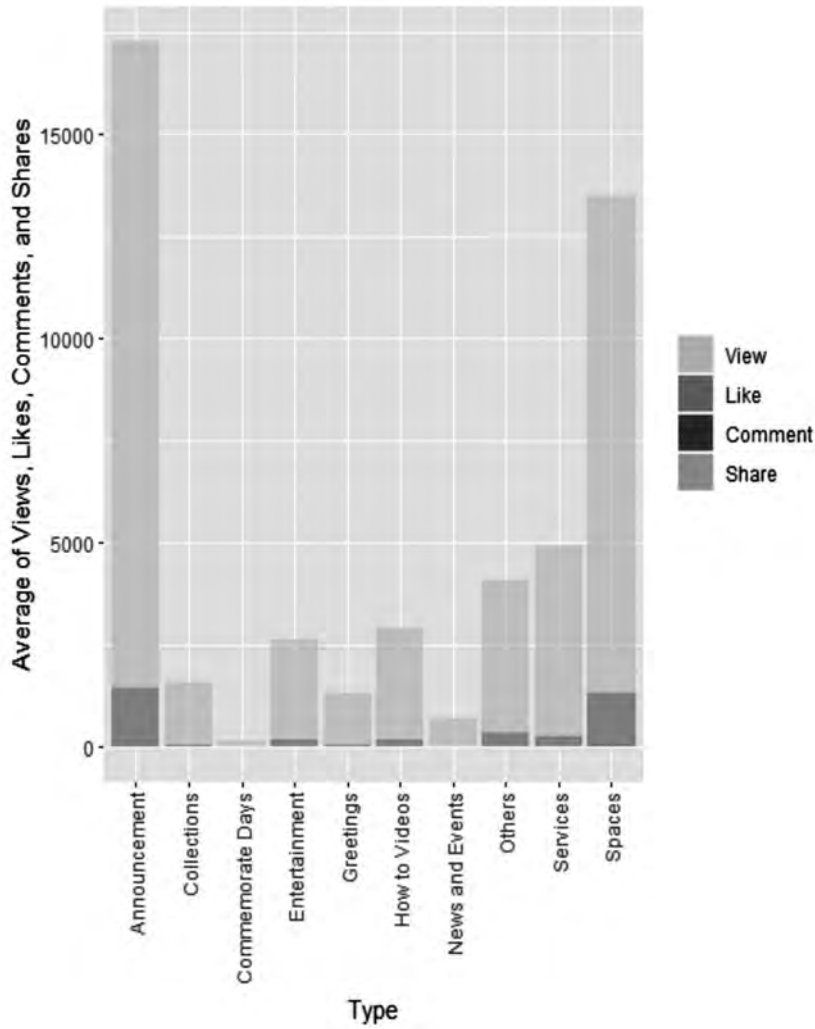


Figure 3. Engagement based on type of information.

Information should be packaged attractively in the form of a short video to make it easier for patrons to learn about the library. There were many changes and shifts in library services provided to patrons during the pandemic and post-pandemic. This, of course, must be quickly disseminated and published to followers. As a result, followers who are library patrons are aware of the most recent changes or information from the library via their social media accounts.

The size of the audience's contribution to the content indicates that social media users, particularly followers, were drawn to the library's content. That is, announcements, services, and library space are the types of information sought and responded to by library followers. However, follower engagement cannot be used as the sole criterion for assessing behavior. Syrdal and Briggs (2018) revealed that consumers who are engaged with content do not necessarily perform any measurable behavior such as sharing, commenting or even "liking".

Libraries that use social media should be concerned about patrons' interaction. This is done not just to draw followers' attention but also to gather feedback on the knowledge that the library disseminates. The number of followers does not necessarily translate into a high degree of engagement with the shared content. According to Idris (2018), there is a lack of interaction in campaign communication, particularly that which is focused on social media. Further explanation was provided, including the fact that organizations should be concerned about the high number of buzzers and phony followers while disseminating information on social media.

## CONCLUSION & FUTURE STUDIES

The data obtained in this study provides valuable insights into the use of TikTok by libraries in Indonesia. A total of 688 posts from 12 library TikTok accounts in Indonesia were subjected to content analysis. Content analysis may evaluate the extent to which different types of libraries use social media platforms such as TikTok as well as the efficient methods employed by libraries to effectively disseminate content. This research also provides a summary of follower interaction on each post and popular themes that libraries can post about.

According to this study, library TikTok accounts in Indonesia were only established in 2020, reflecting the application's growing popularity with a large number of users and short videos shared on the platform. Also, it suggests that libraries are adapting to the changing technological landscape and exploring new ways to engage with their audiences. The restrictions on library visits and access during a pandemic are thought to have encouraged libraries to develop new information-technology-based innovations by utilizing popular platforms.

The main library resources, such as collections, facilities, and services, are the types of information commonly published on TikTok. Outside of the library context, videos that are entertainment/fun or educational. The content analysis revealed that libraries on TikTok mostly publish information about their main resources, such as collections, space, and services. However, libraries also distribute videos that are educational or entertainment/fun, which suggests that they

are trying to reach a wider audience and engage with them in different ways. In this study, the type of announcement pertaining information receives the most engagement. Announcements are relevant because they provide important and timely information to the audience. This finding suggests that libraries should consider the type of information that they publish on social media platforms and focus on providing important and timely updates to their followers. This study suggests that libraries consider the type of information that will be published through social media.

Overall, this study highlights the potential of social media platforms like TikTok for libraries to reach out to their audiences and disseminate information. It also highlights the need for libraries to consider the type of content they publish on social media and to explore new ways to engage with their followers. However, it is important to note that this study only analyzed 12 library TikTok accounts in Indonesia and further research is needed to explore the use of TikTok by libraries in other countries and contexts. Further research is suggested to learn more about the demographics of the library account followers. It also necessitates further investigation into the reactions of followers, particularly comments on any information shared. Finally, it will be more beneficial for library managers to develop social media development strategies.

## REFERENCES

- Aichner, T. *et al.* (2021) 'Twenty-Five years of social media: A review of social media application and definitions from 1994 to 2019', *Cyberpsychology, Behavior and Social Networking*, 24(4), pp. 1–8.  
<https://doi.org/10.1089/cyber.2020.0134215>.
- Al-Daihani, S. M. and AlAwadhi, S. A. (2015) 'Exploring academic libraries' use of Twitter: A content analysis', *Electronic Library*, 33(6), pp. 1002–1015.  
<https://doi.org/10.1108/EL-05-2014-0084>.
- Alley, A. and Hanshew, J. (2022) 'A long article about short videos: A content analysis of U.S. academic libraries' use of TikTok', *Journal of Academic Librarianship*, 48(6).  
<https://doi.org/10.1016/j.acalib.2022.102611>.
- Anderson, K. E. (2020) 'Getting acquainted with social networks and apps: it is time to talk about TikTok', *Library Hi Tech News*, 37(4), pp. 7–12.  
<https://doi.org/10.1108/LHTN-01-2020-0001>.
- Çakmak, T. and Eroğlu, Ş. (2020) 'Public libraries on Facebook: content analysis of Turkish public libraries' posts', *Library Hi Tech*, 39(2), pp. 602–624.  
<https://doi.org/10.1108/LHT-04-2020-0082>.
- Chan, T. T. W., Lam, A. H. C. and Chiu, D. K. W. (2020) 'From Facebook to Instagram: Exploring user engagement in an academic library', *Journal of Academic Librarianship*, 46(6), pp. 1–8.  
<https://doi.org/10.1016/j.acalib.2020.102229>.



- Datareportal (2022) *Digital 2022 Indonesia*. Available at:  
<https://datareportal.com/>.
- Guarda, T. *et al.* (2021) 'The Impact of TikTok on Digital Marketing', in *Smart Innovation, Systems and Technologies*, pp. 35–44.  
[https://doi.org/10.1007/978-981-33-4183-8\\_4](https://doi.org/10.1007/978-981-33-4183-8_4).
- Haenlein, M. and Kaplan, A. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons*, 53(1), pp. 59–68. Available at:  
<http://search.proquest.com/docview/195360323?accountid=15920>.
- Heldman, A. B., Schindelar, J. and Weaver, J. B. (2013) 'Social media engagement and public health communication: Implications for public health organizations being truly "social"', *Public Health Reviews*, 35(1), pp. 1–18.  
<https://doi.org/10.1007/BF03391698>.
- Idris, I. K. (2018) 'Government social media in Indonesia: Just another information dissemination tool', *Jurnal Komunikasi: Malaysian Journal of Communication*, 34(4), pp. 337–356.  
<https://doi.org/10.17576/JKMJC-2018-3404-20>.
- Islam, M. M. and Habiba, U. (2015) 'Use of social media in marketing of library and information services in Bangladesh', *DESIDOC Journal of Library and Information Technology*, 35(4), pp. 299–303.  
<https://doi.org/10.14429/djlit.35.4.8455>.
- Jones, M. J. and Harvey, M. (2019) 'Library 2.0: The effectiveness of social media as a marketing tool for libraries in educational institutions', *Journal of Librarianship and Information Science*, 51(1), pp. 3–19.  
<https://doi.org/10.1177/0961000616668959>.
- Kapoor, K. K. *et al.* (2018) 'Advances in Social Media Research: Past, Present and Future', *Information Systems Frontiers*, 20(3), pp. 531–558.  
<https://doi.org/10.1007/s10796-017-9810-y>.
- Khan, M. L. (2017) 'Social media engagement: What motivates user participation and consumption on YouTube?', *Computers in Human Behavior*, 66, pp. 236–247.  
<https://doi.org/10.1016/j.chb.2016.09.024>.
- Li, Y. *et al.* (2021) 'Communicating COVID-19 information on TikTok: A content analysis of TikTok videos from official accounts featured in the COVID-19 information hub', *Health Education Research*, 36(3), pp. 261–271.  
<https://doi.org/10.1093/her/cyab010>.
- Li, Z. *et al.* (2012) 'SocialTube: P2P-assisted video sharing in online social networks', in *2012 Proceedings IEEE INFOCOM*. IEEE, pp. 2886–2890.  
<https://doi.org/10.1109/INFCOM.2012.6195721>.
- Mashiyane, D. M. (2022) 'Libraries breaking barriers through TikTok: enhancing access and visibility', *Library Hi Tech News*, 39(4), pp. 22–24.  
<https://doi.org/10.1108/LHTN-01-2022-0011>.
- Okike, B. I. (2020) 'Information dissemination in an era of a pandemic (COVID-19): librarians' role', *Library Hi Tech News*, 37(9), pp. 1–4.  
<https://doi.org/10.1108/LHTN-04-2020-0040>.
- Parveen, F., Jaafar, N. I. and Ainin, S. (2015) 'Social media usage and organizational performance: Reflections of Malaysian social media managers', *Telematics and Informatics*, 32(1), pp. 67–78.  
<https://doi.org/10.1016/j.tele.2014.03.001>.

- Quadri, G. O. and Adebayo Idowu, O. (2016) 'Social media use by librarians for information dissemination in three Federal University Libraries in Southwest Nigeria', *Journal of Library and Information Services in Distance Learning*, 10(1–2), pp. 30–40.  
<https://doi.org/10.1080/1533290X.2016.1156597>.
- Rachman, Y. B., Mutiarani, H. and Putri, D. A. (2018) 'Content analysis of Indonesian academic libraries' use of instagram', *Webology*, 15(2), pp. 27–37.
- Rahmawati, N. S. and Rahmi, R. (2021) 'Academic Libraries Promotion Through Social Media: A Systematic Literature Review', *Khazanah al-Hikmah: Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan*, 9(2), p. 157.  
<https://doi.org/10.24252/kah.v9i2a8>.
- Shutsko, A. (2020) 'User-generated short video content in social media. a case study of tiktok', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 108–125.  
[https://doi.org/10.1007/978-3-030-49576-3\\_8](https://doi.org/10.1007/978-3-030-49576-3_8).
- Stephens, R. (2022) 'Getting started with TikTok for library marketing', *Public Services Quarterly*, 18(1), pp. 59–64.  
<https://doi.org/10.1080/15228959.2021.2008286>.
- Suharso, P. and Muntiah, A. (2020) 'Pemanfaatan media sosial Instagram pada perpustakaan perguruan tinggi', *Edulib*, 10(1), pp. 1–14. Available at:  
<http://ejournal.upi.edu/index.php/edulib/index>.
- Swani, K. and Labrecque, L. I. (2020) 'Like, Comment, or Share? Self-presentation vs. brand relationships as drivers of social media engagement choices', *Marketing Letters*, 31(2–3), pp. 279–298.  
<https://doi.org/10.1007/s11002-020-09518-8>.
- Syrdal, H. A. and Briggs, E. (2018) 'Engagement with social media content: A qualitative exploration', *Journal of Marketing Theory and Practice*, 26(1–2), pp. 4–22.  
<https://doi.org/10.1080/10696679.2017.1389243>.
- Taylor & Francis (no date) *Use of social media by the library*, *Librarian Resources*. Available at:  
<https://librarianresources.taylorandfrancis.com/insights/library-advocacy/use-social-media-by-the-library/#> (Accessed: 4 April 2022).
- Unni, Z. and Weinstein, E. (2021) 'Shelter in Place, Connect Online: Trending TikTok Content During the Early Days of the U.S. COVID-19 Pandemic', *Journal of Adolescent Health*, 68(5), pp. 863–868.  
<https://doi.org/10.1016/j.jadohealth.2021.02.012>.
- Vaaler, A. and Brantley, S. (2016) 'Using a blog and social media promotion as a collaborative community building marketing tool for library resources', *Library Hi Tech News*, 33(5), pp. 13–15.  
<https://doi.org/10.1108/LHTN-04-2016-0017>.
- Wu, K. C. and Yang, T. Y. (2021) 'Library collections promotion for preadolescents using social media marketing strategies', *Library Hi Tech*.  
<https://doi.org/10.1108/LHT-03-2020-0073>.

*Para citar este texto:*

Fadhli, Rahmat, Cecep Ibrahim, Miftahunnisa' Igiriza and Bachrul Ilmi. 2023. "How Libraries in Indonesia Use TikTok: A Content Analysis of Library Accounts on the Platform". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 27-44.

<https://doi.org/10.22201/iibi.24488321xe.2023.96.58764>

# La gobernanza de los datos de investigación en el contexto de su organización y sistematización

Eder Ávila-Barrientos\*

*Artículo recibido:*  
22 de febrero de 2023  
*Artículo aceptado:*  
2 de mayo de 2023

*Artículo de investigación*

## RESUMEN

La *gobernanza* es un concepto aplicado al estudio de las ciencias políticas y sociales para explicar las regulaciones y buenas prácticas del Estado en temas de la ciudadanía y la administración pública. La gobernanza de los datos de investigación engloba el desarrollo de políticas para propiciar el efectivo manejo de los datos; como parte de este manejo, la sistematización y la organización son dos procesos que intervienen en la recuperación y accesibilidad a dichos datos. Por lo tanto, el objetivo de este trabajo consiste en identificar los atributos de la gobernanza de los datos de investigación, para comprender su interacción en el contexto de su organización y sistematización. A través de una metodología basada en dos etapas,

\* Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México  
eder@iibi.unam.mx

la primera de ellas, de carácter documental, apoyada en la hermenéutica del discurso, y la segunda en el método analítico-sintético para describir y representar a los datos de investigación, se presentan los hallazgos obtenidos en este estudio. Se estima que la creación y adaptación de un perfil de metadatos diseñado ex profeso para este tipo de datos puede contribuir a establecer un mecanismo interoperable para describir y representar a los datos de investigación en un sistema disponible en el contexto digital.

**Palabras clave:** Gobernanza; Datos de Investigación; Organización; Sistematización

### **Governance of research data in the context of its organization and systematization**

*Eder Ávila-Barrientos*

#### **ABSTRACT**

*Governance* is a concept applied to the study of political and social sciences to explain the regulations and good practices of the State in matters of citizenship and public administration. The governance of research data encompasses the development of policies to promote effective data management, as part of this management, systematization and organization are two processes involved in the recovery and accessibility of said data. Therefore, the objective of this paper is to identify the attributes of research data governance in order to understand their interaction in the context of their organization and systematization. Through a methodology based on two stages, the first of which is documentary based on the hermeneutics of discourse and the second on the analytic-synthetic method to describe and represent research data, the findings obtained in this study are presented. It is estimated that the creation and adaptation of a metadata profile designed specifically for this type of data can contribute to establishing an interoperable mechanism to describe and represent research data in a system available in the digital context.

**Keywords:** Governance; Research Data; Organization; Systematization

## INTRODUCCIÓN

Desde la década de los noventa del siglo pasado, la *gobernanza* es un concepto que se ha utilizado desde una perspectiva política para referirse a la eficacia, calidad y efectiva dirección de la participación del Estado en asuntos de la ciudadanía. Este concepto ha sido adaptado y configurado en un contexto informativo en donde los datos son de suma relevancia para la toma de decisiones en el ámbito gubernamental.

En el contexto de la investigación científica y académica que se desarrolla con fondos públicos, la *gobernanza de los datos de investigación* (GDI) se refiere a las políticas y directrices que permiten regular el manejo y la gestión de dichos datos. Esto conlleva propiciar su acceso y reutilización mediante el uso de sistemas digitales.

En este sentido, la organización y sistematización de los datos de investigación son procesos que involucran el uso de estándares internacionales, como es el caso de esquemas de metadatos desarrollados ex profeso para llevar a cabo la descripción y la representación de los datos bajo un perfil común.

Bajo esta premisa, el presente trabajo pretende abordar los siguientes cuestionamientos:

- ¿De qué manera pueden describirse y representarse los datos de investigación para propiciar su gobernanza?
- ¿Cuál es la interacción entre la gobernanza de los datos y su descripción y representación?

De acuerdo con Koltay (2016: 304) “la expresión gobernanza de datos (GD) podría referirse a los órganos organizativos; reglas, políticas, estándares; derechos de decisión; asignación de responsabilidades y métodos de ejecución que hacen posible la gestión de los datos”. De esta manera, la gobernanza de datos engloba a los procesos y técnicas para propiciar el acceso a los datos en un marco común.

De esta manera, el objetivo que persigue este trabajo es identificar los atributos de la gobernanza de los datos de investigación y comprender su interacción en el contexto de su organización y sistematización. De este modo, el estudio de los datos de investigación requiere de una perspectiva holística que permita comprender las aristas que forman parte de su fenomenología.

### ***Gobernanza de datos de investigación***

En años recientes, la gobernanza de los datos ha cobrado notable relevancia, debido a su dinamismo y a su aparición en esferas y contextos que versan sobre dominios como la administración pública y la privada, el gobierno abierto, la gestión organizacional, la transparencia y la rendición de cuentas. Sin embargo, su corpus teórico y conceptual aún presenta patrones de incertidumbre, al no reflejar con claridad su epistemología y razón de ser.

Desde la óptica de la gestión de datos, la GDI permite identificar el uso de políticas y lineamientos para hacer accesibles los datos que forman parte de la administración, pública y privada, de las instituciones gubernamentales y de la ciencia abierta. De acuerdo con Weber *et al.* (2009: 6), la GDI pone de manifiesto un marco común para ejercer los derechos de decisión y las responsabilidades para fomentar el comportamiento deseable en el uso de los datos. Si bien esta definición puede prestarse a confusiones, son de resaltarse las buenas prácticas que se deben llevar a cabo para utilizar los datos en diversos contextos.

Por otra parte, la Organización para la Cooperación y el Desarrollo Económico (OECD, por sus siglas en inglés) manifiesta que la gobernanza de los datos se refiere a diversas estrategias, que incluyen disposiciones técnicas, normativas, reglamentarias e institucionales, que pueden afectar a los datos y su creación, recopilación, almacenamiento, uso, protección, acceso, intercambio y eliminación, todo esto incluido en los dominios de las políticas organizacionales y gubernamentales (OECD, 2022: 13).

De esta manera, las políticas nacionales para el acceso a los datos forman parte de un conjunto de instrumentos normativos que tienen el propósito de regular el manejo y la accesibilidad de los datos en el contexto público y gubernamental; estas políticas tienen una interacción con el uso de los datos en diversas actividades y contextos productivos. Aunado a ello, “la gobernanza de los datos contempla a los procesos estándares y modelos que en un entorno común describen integralmente quién puede tomar una serie de acciones sobre los datos, esto de acuerdo con métodos predeterminados y fundamentados en los derechos de acceso” (Kuiler, 2022: 286).

Dentro de estos métodos pueden ubicarse aquellos relacionados con una estructura de metadatos para organizar, representar y hacer accesibles a los datos en sus diferentes formas y naturaleza. Pues la GDI debe contemplar su efectiva organización y representación. La complejidad de los datos y su dinamismo refleja su carácter interdisciplinario, en donde cada vez se hace más frecuente el uso de la inteligencia artificial y de algoritmos para captar datos de diversas

fuentes. Sin embargo, esto trae consigo una serie de retos y consideraciones que, desde la óptica de la gobernanza, han sido discutidos por diversos autores.

Por ejemplo, Jansen *et al.* (2020: 7) han recabado e interpretado los principios de la gobernanza de los datos desde su evaluación, la cual permite detectar sus irregularidades, hasta establecer un juicio imparcial al momento de describirlos y representarlos. De acuerdo con Fleckenstein y Fellows (2018: 64), la gobernanza de datos se trata, en gran medida, de colaboración, es decir, de conectar a productores y consumidores de datos y determinar quién tomará las decisiones fundadas en el uso de dichos datos con base en su accesibilidad. Desde este enfoque, la GDI remite al desarrollo de buenas prácticas para gestionar los datos en diferentes entornos, dejando ver un proceso que consta de etapas específicas para manejarlos desde un punto de vista organizativo y apegado a su sistematización.

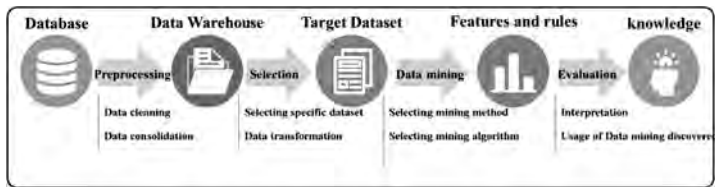


Figura 1. Four-steps data analytics process for data governance of NPO  
Fuente: Wang *et al.* (2019, 500)

Dentro de la literatura especializada en el tema de GDI, se puede identificar una amplia gama de procesos que tratan de explicar la manera en cómo los datos pueden ser manejados para obtener de ellos variables de estudio. Estos procesos son acompañados, la mayoría de las veces, de una metodología *ex profeso*. Por ejemplo, en la *Figura 1* puede observarse el proceso de análisis para la gobernanza de datos que toman parte en organizaciones sin ánimo de lucro. En este proceso destacan fases que engloban un proceso general para dar tratamiento a los datos que están depositados en una base de datos o un silo en particular.

Desde la limpieza de los datos, hasta su acomodo y transformación, se trata de fases que tienen el propósito de dirigir el manejo de los datos a un asunto en particular, tomando en consideración las fases de análisis a las cuales se exponen los datos. Técnicas tales como la minería de datos y su procesamiento algorítmico contribuyen a una mejor interpretación de los datos, con el objetivo de construir conocimientos sobre un tema, un fenómeno o una problemática a solucionar.

Con el auge de la ciencia abierta y la necesidad de contar con datos de investigación para reutilizarlos en la generación de nuevos conocimientos, la GDI tiene una interacción con la generación de políticas que permitan dar solidez y apertura a los datos, sobre todo en entornos académicos y científicos. En este sentido, “la ciencia abierta es el movimiento que busca la apertura de la investigación



científica (métodos, instrumentos, datos, etc.) para beneficio de toda la sociedad. La ciencia abierta se presenta como un medio capaz de articular y dinamizar las políticas de ciencia, tecnología e innovación” (AmeliCA, 2019: párr. 1).

De esta manera, la ciencia abierta y la GDI interactúan en un entorno de investigación científica y académica que busca eliminar las barreras de acceso a la producción intelectual que se genera en estos contextos. Esto mediante el desarrollo de políticas que regulen dicha apertura.

En este sentido, “el concepto de ciencia abierta supone la transparencia de la metodología de investigación y observación, así como el amplio acceso a las colecciones de conjuntos de datos científicos utilizados y adquiridos como resultado de experimentos, la libre accesibilidad y reutilización de dichos datos, la disponibilidad y transparencia de las comunicaciones científicas y el uso de plataformas virtuales de apoyo al flujo de trabajo científico colaborativo” (Karmansky et al., 2016: 400).

Bajo esta premisa, en la *Figura 2* puede observarse la interacción entre las políticas, la ciencia abierta y la GDI tomando en cuenta el contexto de la investigación científica y académica en donde se establece un marco común que motiva la necesidad de organizar y sistematizar a los datos que se generan en el contexto académico y científico.



*Figura 2.* Interacción entre las políticas, la ciencia abierta y la gobernanza de los datos de investigación en el contexto de la investigación científica y académica

Fuente: Elaboración propia

En este sentido, las políticas son un conjunto de actividades que se asocian con la toma de decisiones. Tratan de orientaciones o directrices que rigen la actuación de una persona o entidad en un asunto o campo determinado (RAE: 2023). Por lo tanto, las políticas pueden ser de diverso orden y naturaleza. En el contexto de la gobernanza de los datos de investigación, estas políticas permiten definir el manejo de los datos para su posterior accesibilidad a través del entorno digital.

De acuerdo con Solomonides (2019: 292), la gobernanza de los datos de investigación comprende los principios, políticas y estrategias adoptadas por las instituciones para proporcionar mecanismos operativos y de acceso a datos. Por lo tanto, las políticas para establecer la gobernanza de los datos conllevan el seguimiento de los principios que rigen a la recopilación, organización, transformación, el análisis, la recuperación y reutilización de los datos de investigación.

Con el propósito de ejercer una diferenciación entre las características de los conceptos de gobernanza, gobernanza de datos y gobernanza de datos de investigación, en la *Tabla 1* se ofrece una explicación de los componentes conceptuales de ambos elementos.

Concepto	Características
<b>Gobernanza</b>	<ul style="list-style-type: none"> <li>• Es un concepto que se desprende de la doctrina política y sociológica.</li> <li>• Explica la interacción entre actores involucrados en la toma de decisiones de asuntos públicos.</li> <li>• En su aplicación, figuran las políticas como instrumentos para ejecutar y fundamentar la toma de decisiones.</li> </ul>
<b>Gobernanza de datos (GD)</b>	<ul style="list-style-type: none"> <li>• En su corpus, están presentes elementos como la calidad y confiabilidad de los datos.</li> <li>• Es un concepto que expone el papel de las políticas y los procesos que permiten garantizar el acceso a los datos.</li> <li>• Se relaciona con el término de gobierno de datos, debido a su aplicación en diversas instituciones de índole pública y privada.</li> </ul>
<b>Gobernanza de datos de investigación (GDI)</b>	<ul style="list-style-type: none"> <li>• Es un concepto que remite a un proceso intelectual para organizar y sistematizar datos con el propósito de hacerlos accesibles y reutilizables en el contexto digital.</li> <li>• Refleja la importancia de la reutilización de los datos para propiciar el desarrollo de nuevos conocimientos.</li> <li>• Su aplicación permite definir planes para la gestión de los datos de investigación disponibles en las instituciones académicas y científicas.</li> </ul>

*Tabla 1.* Características de los conceptos de gobernanza, gobernanza de datos y gobernanza de datos de investigación

Elaboración propia. Fundamentada en las aportaciones de Cristina Zurbruggen (2011), T. Koltay (2016), y Rene Abraham, Johannes Schneider y Jan vom Brocke (2019)

De acuerdo con Abraham (2019: 426) “la gobernanza de los datos especifica un marco multifuncional para administrar datos como un activo organizacional estratégico. Al hacerlo, la gobernanza de los datos especifica los derechos de decisión y responsabilidades para la toma de decisiones de una organización sobre sus datos. Además, la gobernanza de datos formaliza las políticas, estándares y procedimientos aplicables a los datos supervisando su cumplimiento”.

Bajo esta premisa, la GD es un concepto genérico que, al momento de aplicarse en el contexto de la investigación científica y académica, remite a la presencia de los datos de investigación, en el sentido de su manejo con el propósito de fomentar su accesibilidad y reutilización. Por lo tanto, uno de los tópicos que la GDI aborda es la organización y la sistematización de los datos de investigación, pues estos procesos son claves para propiciar su accesibilidad y consulta en el contexto digital.

La GDI es un concepto holístico que no es exclusivo para el tratamiento de un solo tipo de datos de investigación. Pues la naturaleza de estos datos es ampliamente diversa y compleja, por lo que los procesos de organización y sistematización deben fundamentarse en el análisis de los atributos que presenten los datos de investigación.

### ***Organización y sistematización de los datos de investigación***

“Los datos de investigación se recopilan, observan, registran o crean para fines de análisis y producir resultados de investigación originales para un estudio” (Henderson, 2017: 2). El uso de metadatos para organizar a los datos de investigación es un tema que se ha planteado con anterioridad dentro de la literatura especializada. Si bien los metadatos son elementos descriptivos que permiten representar, recuperar y hacer accesibles los recursos y contenidos informativos disponibles en el entorno digital, cabe señalar que la concepción de la sistematización de los datos de investigación refleja un cambio sustancial en la manera de efectuar su organización.

Por ejemplo, Redkina (2019: 53) desarrolló un análisis de las políticas, directrices y lineamientos que son generados en diversos países cuyo propósito es dar mayor apertura a los datos de investigación mediante la aplicación de metadatos que forman parte de un plan para la gestión de datos de investigación. En este trabajo resalta la necesidad de contar con esquemas de metadatos específicos para organizar conjuntos de datos de investigación.

Por otra parte, Crystal-Ornelas *et al.* (2022: 1) llevaron a cabo un estudio de los datos que son generados en el contexto de las Ciencias de la Tierra y del Medio Ambiente, esto en el sentido de identificar la aplicación de los metadatos que utilizan para describir a los conjuntos de datos que son generados por

diversas comunidades científicas de ese campo disciplinario y hacerlos interoperables con los principios FAIR (Findability: *Localización*, Accessibility: *Accesibilidad*, Interoperability: *Interoperabilidad*, Reuse: *Reutilización*).

“Los principios FAIR hacen énfasis en mejorar la capacidad de las computadoras (es decir, la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos con una intervención humana mínima o nula) esto debido a que los humanos confían cada vez más en el soporte computacional para manejar los datos como resultado del aumento en el volumen, complejidad y velocidad de creación de los datos” (GoFAIR, 2023: párr. 1).

La capacidad de los metadatos para establecer un mecanismo de citación de datos otorga la posibilidad de contemplar los principios éticos relacionados con la certeza y calidad de los datos que son representados mediante los metadatos. Pues esto aumenta el nivel de confiabilidad y certeza al momento de respetar la propiedad intelectual y dar el debido reconocimiento a los creadores de los datos.

En este sentido, los datos, al momento de organizarse, “deben manejarse con cuidado durante su posesión y distribución. La propiedad y el control sobre los datos no son de naturaleza absoluta. Incluso el propietario principal, la fuente de los datos, sólo tiene derecho a utilizar los datos de forma ética. La propiedad de los datos se refiere tanto a la posesión como a la responsabilidad de la información” (Shukla et al., 2022: 69).

Desde esta perspectiva ética, la gobernanza de los datos de investigación puede ser útil para dotar de una mayor organización y sistematización a los datos; esto mediante la generación de políticas que regulen la sistematización de dichos datos.

Algunos ejemplos de estas regulaciones pueden observarse en proyectos como R3Data (<https://www.re3data.org/>), DataONE (<https://www.dataone.org/>) y Dryad (<https://datadryad.org/stash>), pues son proyectos que están basados en diversos lineamientos y políticas que, si bien son aplicables a diversas aristas de los datos de investigación, abordan diferentes temas respecto a su manejo y organización.

La sistematización de los datos de investigación pone de manifiesto una visión holística relacionada no sólo con la generación de repositorios de datos, sino también de instrumentos que regulen su apertura tomando en cuenta los factores técnicos, legales y económicos que forman parte del contexto académico y científico. En este sentido, la confiabilidad de un repositorio de datos debe enfocarse en:

- Adherirse a los estándares de metadatos y conservación, además de proporcionar la administración de los datos que almacena, por ejemplo, validación técnica, documentación, control de calidad, protección de la autenticidad y persistencia a largo plazo.

- Proporcionar servicios de datos, por ejemplo, interfaces de visualización, descarga de datos y procesamiento basado en servidor.
- Administrar los derechos de propiedad intelectual de los creadores de datos, la protección de los recursos de información sensible y la seguridad del sistema y su respectivo contenido. (Lin et al., 2020: 3)

En este sentido, la confiabilidad de la sistematización y organización de los datos dotará de una mayor certeza a los usuarios para poder utilizarlos, siendo este un principio relacionado con la gobernanza de los datos de investigación al momento de ser colocados en el entorno digital para su respectiva apertura, accesibilidad y re-utilidad.

## METODOLOGÍA

La metodología empleada en este trabajo se desarrolló en dos niveles. En el primer nivel, se llevó a cabo un proceso de revisión de la literatura, apoyado en la hermenéutica y el análisis del discurso, con la intención de analizar recursos de información vinculados al objeto de estudio planteado en la investigación. De esta manera el análisis argumentativo permitió interpretar los textos referentes al objeto de estudio, en función de la idea principal planteada, sus antecedentes y el propósito de su comunicación.

Para esto, se analizaron artículos de investigación, capítulos de libro y estudios de caso. Estos recursos de información fueron obtenidos en diversas bases de datos, como Elsevier, Springerlink, IEEE y Emerald; en repositorios institucionales, como el repositorio institucional de la UNAM y el repositorio de la CEPAL; y, además, en sitios web educativos e institucionales especializados en el objeto de estudio planteado en la investigación (Research Data Alliance, Research Data Management @Harvard). La selección de estas fuentes responde a su especificidad temática relacionada con el manejo y la gobernanza de los datos de investigación, pues son fuentes que tratan de manera particular el estudio de los datos desde diversas aristas, por ejemplo, procedimental, teórica y tecnológica.

Dentro de estas fuentes, se implementaron estrategias de búsqueda relacionadas con los siguientes términos: *gobernanza de datos*, *gobernanza de datos de investigación*, *organización de datos* y *sistematización de datos* y sus equivalentes en inglés. En la *Tabla 2*, se expone este proceso con el número de registros obtenidos.

<b>Términos de búsqueda</b>	<b>Springer-Link</b>	<b>Emerald Insight</b>	<b>Elsevier</b>	<b>IEEE Xplore</b>	<b>Repositorio UNAM</b>	<b>Repositorio CEPAL</b>
<i>Gobernanza de datos</i>	82	16	287	0	244	2,882
<i>Data governance</i>	237,411	62,000	136,628	3,431	21	11,409
<i>Gobernanza de datos de investigación</i>	55	11	224	0	76	2,774
<i>Research data governance</i>	214,728	60,000	127,901	1,447	6	9,640
<i>Organización de datos</i>	721	111	15,670	0	462	17,518
<i>Data organization</i>	1,837,646	209,000	1,830,628	45,141	159	11,133
<i>Sistematización de datos</i>	47	2	672	0	39	3,885
<i>Data systematization</i>	19,329	1,000	30,750	172	9	4,853

*Tabla 2.* Número de registros obtenidos a partir de las estrategias de búsqueda relacionadas con el objeto de estudio de la investigación  
Fuente: Elaboración propia

Cabe señalar que, aunque los términos empleados permitieron recuperar recursos que emplearan dichos términos dentro de su título, fue necesario revisar detalladamente cada recurso para identificar el enfoque disciplinario al que estaban dirigidos. Pues la noción conceptual de datos representa a múltiples variables de estudio, desde aquellas que versan sobre el rubro técnico y pragmático, hasta aquellas que permiten considerar sus nociones epistemológicas. En el caso de los sitios Research Data Alliance y Research Data Management @Harvard, no se realizaron búsquedas específicas de recursos, sino que se llevó a cabo un análisis de la documentación que contenían en sus páginas web que tratara los temas planteados en el objeto de estudio de esta investigación. De esta manera, la hermenéutica y el análisis del discurso permitió llevar a cabo la interpretación de los fenómenos particulares (gobernanza de datos de investigación) que acontecen en un contexto de tiempo y espacio definido (Cárcamo, 2005: 204).

Por lo tanto, se seleccionaron recursos publicados de 2015 a 2022, tomando en cuenta criterios que reflejasen una condicionante tecnológica y cronológica, pues el estudio de los datos se trata de un fenómeno que se encuentra en un constante dinamismo debido a la aparición de múltiples publicaciones que analizan su impacto

en diferentes contextos y disciplinas. Y, además, la mayoría de los recursos utilizados fueron publicados en idioma inglés.

La periodicidad de los años seleccionados responde a la constante actualización de los contenidos y los tópicos que se presenta en el objeto de estudio de los datos, pues en años recientes la literatura que versa sobre este tema ha crecido exponencialmente. Se optó por que los recursos seleccionados incluyeran, dentro de su espectro argumentativo, temas como *gobernanza de datos*, *gobernanza de datos de investigación*, *organización de datos* y *sistematización de datos*, para someterlos al análisis del discurso contenido dentro de su estructura argumentativa.

Los recursos de información publicados fuera del rango periódico mencionado fueron elegidos tomando en cuenta su grado de contribución y pertinencia respecto al análisis del objeto de estudio planteado en este trabajo; es decir, tomando en cuenta su relevancia para obtener una visión global del tema y que se relacionara con los tópicos de organización de la información y el conocimiento.

A su vez, tomando como sustento el método analítico-sintético, se seleccionó y descargó un conjunto de datos de investigación del directorio de repositorios de datos de investigación Dryad, disponible en <https://datadryad.org/stash>, con la finalidad de describirlo y representarlo mediante el uso de Datacite Metadata Schema, versión 4.4, esto con el propósito de ejemplificar las fases que intervienen en la organización y sistematización de los datos de investigación, mediante el uso de un esquema interoperable y estandarizado internacionalmente.

Los pasos a seguir para la tarea de organización y representación de estos datos se pueden consultar en la *Tabla 3*.

Fase	Descripción
1. <i>Selección del conjunto de datos de investigación</i>	Se seleccionaron conjuntos de datos con licencia abierta, que permitiera descargar y reutilizar a los datos de investigación o bien depositarlos en un repositorio respetando sus derechos de creación y propiedad intelectual.
2. <i>Análisis de los atributos del conjunto de datos</i>	Se identificó la naturaleza y temática de los datos descargados, mediante un análisis que permitiera obtener las propiedades representativas del conjunto.
3. <i>Descripción del conjunto de datos de investigación</i>	Tomando en cuenta los elementos descriptivos del esquema de metadatos, se describieron los atributos de los datos de investigación, siguiendo directrices señaladas en dicho esquema.
4. <i>Representación del conjunto de datos de investigación</i>	Mediante el diseño de un perfil descriptivo de metadatos, se representaron los datos de investigación, asignando puntos de acceso para su latente búsqueda y recuperación.

5. <i>Sistematización del conjunto de datos de investigación</i>	Las descripciones realizadas fueron almacenadas en un silo de datos para identificar las variables de la sistematización de los datos en un entorno digital.
------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 3. Metodología para la organización y sistematización de datos de investigación

Fuente: Elaboración propia

De esta manera, la elección de DataCite Metadata Schema, responde a su nivel de especificidad para describir y representar datos de investigación, al ser un esquema desarrollado ex profeso para este objetivo. Mediante el uso de este esquema pueden obtenerse metadatos para describir atributos de datos de investigación de diversa naturaleza, desde datos numéricos, textuales y alfa-numéricos.

Sin embargo, el esquema no puede concebirse como un modelo genérico para la descripción de datos de investigación, pues su sentido holístico permite implementarlo en diversos contextos disciplinarios acordes con la propia naturaleza de los datos; por esta razón, se sugiere primeramente analizar y conocer los atributos que caracterizan a los datos que serán organizados y sistematizados.

La metodología para el manejo de datos de investigación, desde la perspectiva de la GDI, aporta elementos para la elaboración de planes para la gestión de dichos datos, lo cual puede contribuir a visibilizar en mayor medida el trabajo de los investigadores. De acuerdo con Palsdottir (2021: 323), entre los beneficios de la gestión de datos está que facilita el intercambio de datos, lo que permite que otros investigadores verifiquen los resultados y reutilicen los datos.

## ANÁLISIS DE RESULTADOS

Al momento de descargar el conjunto de datos de investigación, fue preciso analizarlos y tener la posibilidad de consultarlos mediante el uso de *software* de visualización. Mediante el examen de las propiedades del conjunto, fue posible identificar los atributos de los datos y representarlos en un perfil de metadatos diseñado con DataCite Metadata Schema (véase la *Tabla 4*). De esta manera, fue posible aplicar una estructura de metadatos susceptible de fomentar la organización de los datos de investigación que fueron descritos como parte de este ejercicio.

El esquema de metadatos de DataCite, en su versión núcleo 4.4, está conformado por 20 elementos de metadatos, divididos en tres diferentes niveles de uso: Obligatorio, Recomendado y Opcional.

Mediante el uso del esquema base de metadatos fue necesario establecer un perfil de metadatos para describir y representar al conjunto de datos denominado “*Data from: Exploratory and confirmatory research in the open science era*”, tal y como se observa en la *Tabla 4*.



ID	Elemento	Descripción
1	Identificador	<a href="https://doi.org/10.5061/dryad.z612jm686">https://doi.org/10.5061/dryad.z612jm686</a>
2.1	Nombre del creador	Nilsen, Erlend B.
2.4. b	SchemeURI	<a href="https://orcid.org/0000-0002-5119-8331">https://orcid.org/0000-0002-5119-8331</a>
2.5	Afiliación	Norwegian Institute for Nature Research
2.1	Nombre del creador	Bowler, Diana.
2.5	Afiliación	German Center for Integrative Biodiversity Research
2.1	Nombre del creador	Linnell, John.
2.5	Afiliación	Norwegian Institute for Nature Research
3	Título	Data from: Exploratory and confirmatory research in the open science era
4	Editor	Dryad
5	Año de publicación	2020
6	Tema	Ciencia abierta
6	Tema	Revisión de la literatura
6	Tema	Investigación ecológica
8	Fecha	2020-02-03
8.a	Tipo de fecha	Publicación
9	Idioma	Inglés
10	Tipo de recurso	Conjunto de datos
12	Identificador relacionado	<a href="https://doi.org/10.1111/1365-2664.13571">https://doi.org/10.1111/1365-2664.13571</a>
12. a	Tipo de identificador relacionado	DOI
12. b	Tipo de relación	Es derivado de
13	Tamaño	70.3 kb
14	Formato	CSV
16	Derechos	Este trabajo tiene una licencia CC0 1.0 Universal (CC0 1.0) Public Domain Dedication .
16. a	URI de derechos	<a href="https://creativecommons.org/publicdomain/zero/1.0/">https://creativecommons.org/publicdomain/zero/1.0/</a>

17	Descripción	Applied ecological research is increasingly inspired by the Open Science movement. However, new challenges about how we define our science when biodiversity data is being shared and re-used are not solved. Among these challenges is the risk associated with blurring the distinction between research that mainly seeks to explore patterns with no a-priori articulated hypotheses (exploratory research), and research that explicitly tests a-priori formulated hypotheses (confirmatory research).
17 a.	Tipo de descripción	Resumen
19	Financiamiento	Norges Forskningsråd, Award: 251112

*Tabla 4.* Descripción y representación de un conjunto de datos de investigación mediante un perfil de metadatos diseñado con DataCite Metadata Schema  
Fuente: Elaboración propia, con datos de DataCite Metadata Schema, versión 4.4. 2021.

El registro de datos en cada uno de los elementos descriptivos del perfil fue llevado a cabo mediante el principio de “registro de observación”, es decir, de la fuente original del recurso, se tomaron los datos que representaban los atributos del *dataset* que fue examinado en el repositorio Dryad a través de su descarga. El perfil de metadatos emplea un nivel de ocurrencia que está relacionado con la repetibilidad de un determinado metadato. En el caso de los elementos 2.1 y 2.5, esta variable puede verse reflejada al momento de describir el nombre de los creadores y su respectiva afiliación.

Cabe señalar que el diseño de cada perfil de metadatos debe desarrollarse tomando en cuenta el análisis de los atributos que representan a cada recurso de información, en este caso a cada conjunto de datos de investigación que sea depositado y almacenado en un sistema. A su vez, el elemento 2.4. *b* refiere al Schema URI y permite apreciar las vinculaciones que pueden llevarse a cabo para identificar a los autores con sus respectivos identificadores digitales. En este caso el identificador empleado por el esquema particular de ORCID.

El nivel de exhaustividad al momento de llevar a cabo la descripción y representación del conjunto de datos de investigación permite identificar las propiedades que forman parte de la naturaleza dinámica del recurso, es decir, algunos de los datos registrados en el perfil son susceptibles de actualizarse con frecuencia dentro del propio sistema Dryad.

La gobernanza de los datos de investigación es un concepto que permite permitirse al uso de estándares de metadatos que faciliten la organización de los datos, mediante el análisis de sus atributos. Dicha gobernanza en su nivel más abstracto hace posible la interacción de diversos elementos normativos e informativos. Por ejemplo, la generación de políticas para regular y uniformar la representación de los datos es un escenario global.

A nivel de sistematización, la GDI supone un cambio en cuanto a la generación de sistemas para buscar y recuperar datos, por ejemplo, la generación de repositorios que, en un escenario global, están comenzando a desarrollarse en diversas instituciones académicas y de investigación a nivel mundial.

Si bien el uso de repositorios institucionales para fomentar el acceso abierto a las publicaciones científicas y académicas ha cobrado una notable madurez, el fenómeno de la generación vertiginosa de datos y su uso en la investigación está cobrando doble relevancia al momento de considerar la GDI como un mecanismo para propiciar la accesibilidad y la reutilización de los datos de investigación. Por lo tanto, los resultados obtenidos mediante el ejercicio planteado en esta investigación permiten obtener un acercamiento en cuanto a la descripción y representación de datos de investigación, mediante el uso de un esquema interoperable.

## DISCUSIÓN

En los resultados obtenidos, puede percibirse la aplicación de elementos de la gobernanza de los datos de investigación, los cuales son susceptibles de implementarse al momento de utilizar un esquema de metadatos formal para la descripción de datos de investigación. En este sentido, la GDI refleja un factor de especificidad mayor en cuanto al uso de principios jerárquicos para representar datos específicos. Por lo tanto, la GDI permite definir roles, procesos y puntos de acción para describir los datos de investigación.

A su vez, la GDI es un concepto que, al momento de implementarse pragmáticamente en la organización y sistematización de los datos, hace posible concebir la interrelación que debe existir entre las políticas y las directrices para describir y representar a los datos de una determinada institución. Este argumento fue adoptado como guía para el desarrollo de esta investigación, pues este tipo de gobernanza tiene un carácter holístico que permite contemplar la accesibilidad a los datos de investigación en un marco común, pero no totalitario ni genérico.

En el contexto de la ciencia abierta, la gobernanza de los datos de investigación tiene la particularidad de fomentar la interoperabilidad normativa y procedimental para construir un ecosistema de datos de investigación accesible para la sociedad y las comunidades que así lo requieran.

La propuesta que se presenta para la descripción y representación de los datos de investigación no es de carácter exhaustivo, pues el diseño del perfil de metadatos desarrollado con DataCite Metadata Schema tiene el propósito de ejemplificar los alcances para la organización y sistematización de dichos datos, ya que permite establecer un mecanismo interoperable para describir a los datos de investigación y su aplicación en un repositorio de datos, lo cual permite, a su vez,

dotar de mayor visibilidad a los datos generados en investigaciones científicas y académicas, además de propiciar su reutilización a través de un aparato de citación que propicie el respeto a la propiedad intelectual de dichos datos.

A su vez, la concepción de la gobernanza de los datos de investigación aún presenta limitaciones relacionadas con su propia epistemología, pues no queda claro su nivel de dominio dentro de las organizaciones académicas y científicas, y su importancia para propiciar una mayor comprensión de los problemas y fenómenos que caracterizan a la organización y la sistematización de los datos de investigación.

Al respecto, Mahanti (2021: 19) manifiesta que “existen factores, tanto internos como externos a una organización que pueden afectar el valor de los datos. Los datos pueden transformarse de un activo a un gran pasivo si no son aptos para su propósito, es decir, su calidad no está a la altura, por ejemplo, datos incorrectos o desactualizados, registros faltantes o datos que no son oportunos”.

Por lo tanto, la gobernanza de los datos de investigación se concentra en el manejo de los datos desde su perspectiva legal hasta la operativa, lo cual lleva a considerar su función dentro de los procesos de su organización y sistematización dentro de las instituciones científicas y académicas.

## CONSIDERACIONES FINALES

Como parte de la investigación realizada, se considera que la GDI es un concepto caracterizado por dos componentes, por un lado, su visión política, que refleja la obligación de transparentar el uso de los datos que emanan de la actividad pública y privada, así como de las instituciones de gobierno y, por el otro, su postura relativa al manejo y gestión de los datos en dichas instituciones. Es dentro de este rubro en donde los procesos de la organización y sistematización de los datos de investigación tienen una interacción conceptual.

Por otra parte, la gobernanza de los datos de investigación busca establecer estándares internos y externos en las instituciones de investigación científica y académica para dotar de una mayor calidad y confiabilidad a los datos con fines de su accesibilidad y reutilización. Esto conlleva el establecimiento de políticas para designar una infraestructura capaz de asignar roles y responsabilidades para manejar, organizar y sistematizar dichos datos.

A su vez, la organización y sistematización de los datos de investigación son dos procesos que interactúan en la gobernanza de los datos, pues forma parte del manejo y gestión que los datos deben tener al momento de pensar en su acceso y reutilización. En este sentido, un perfil de metadatos diseñado ex profeso para este tipo de datos puede contribuir a establecer un mecanismo interoperable para

describir y representar a los datos de investigación en un sistema disponible en el contexto digital.

De esta manera, la gobernanza de los datos de investigación es un tema con diversas aristas. En este trabajo se ha presentado aquella que tiene lugar en la organización y sistematización de los datos, mediante un esquema desarrollado ex profeso para llevar a cabo la descripción y la representación de los datos a través de un enfoque sistematizado, lo cual permite advertir que los datos de investigación no pueden ser tratados de una manera generalizadora, sino llevando a cabo el análisis de sus atributos para poder representarlos de acuerdo a los requerimientos de un determinado contexto disciplinario.

## REFERENCIAS

- Abraham, Rene, Johannes Schneider, y Jan vom Brocke. 2019. "Data Governance: A Conceptual Framework, Structured Review, and Research Agenda". *International Journal of Information Management* (49): 424-438.  
<https://doi.org/10.1016/j.ijinfomgt.2019.07.008>.
- AmeliCA. 2019. "Ciencia Abierta". Acceso 1 de febrero de 2023.  
<http://www.amelica.org/proyectos/ciencia.html>
- Cárcamo Vásquez, H. 2005. "Hermenéutica y análisis cualitativo". *Cinta de Moebio: Revista de Epistemología de Ciencias Sociales* 23 (1): 204-216.  
<https://www.moebio.uchile.cl/23/carcamo.html>
- Crystal-Ornelas, Robert, Charuleka Varadharajan, Dylan O’Ryan, Kathleen Beilsmith, Benjamin Bond-Lamberty, Kristin Boye, Madison Burrus, et al. 2022. "Enabling FAIR Data in Earth and Environmental Science with Community-Centric (Meta) Data Reporting Formats". *Scientific Data* 9 (1): 700.  
<https://doi.org/10.1038/s41597-022-01606-w>
- DataCite Metadata Working Group. 2021. "DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4". Application/pdf.  
<https://doi.org/10.14454/3W3Z-SA82>
- Fleckenstein, Mike, y Lorraine Fellows. 2018. "Data Governance". En *Modern Data Strategy*, editado por Mike Fleckenstein y Lorraine Fellows, 63-76. Cham: Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-68993-7\\_8](https://doi.org/10.1007/978-3-319-68993-7_8)
- Go FAIR. 2023. "FAIR Principles". Acceso 2 de febrero de 2023.  
<https://www.go-fair.org/fair-principles/>
- Henderson, Margaret E. 2017. "Data management: a practical guide for librarians". *Practical guides for librarians*, no. 28. Lanham, Maryland: Rowman & Littlefield.
- Jansen, M., Paul Brous, Elsa Estévez, Luis Barbosa, y Tomasz Janowski. 2020. "Data Governance: Organizing Data for Trustworthy Artificial Intelligence". *Government Information Quality* 37,  
<https://doi.org/10.1016/j.giq.2020.101493>

- Karmanovskiy, Nikolay, Dmitry Mouromtsev, Mikhail Navrotskiy, Dmitry Pavlov, e Iri-na Radchenko. 2016. "A Case Study of Open Science Concept: Linked Open Data in University". En *Digital Transformation and Global Society*, editado por Andrei V. Chugunov, Radomir Bolgov, Yuri Kabanov, George Kampis y Maria Wimmer, 400-403. Communications in Computer and Information Science. Cham: Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-49700-6\\_39](https://doi.org/10.1007/978-3-319-49700-6_39)
- Koltay, T. 2016. "Data governance, data literacy and the management of data quality". *IFLA Journal* 42 (4).  
<https://journals.sagepub.com/doi/full/10.1177/0340035216672238>
- Kuiler, Erik W. 2022. "Data Governance". En *Encyclopedia of Big Data*, editado por Laurie A. Schintler, y Connie L. McNeely, 286-290. Cham: Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-32010-6\\_306](https://doi.org/10.1007/978-3-319-32010-6_306)
- Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giarretta, Marisa De Giusti, et al. 2020. "The TRUST Principles for Digital Repositories". *Scientific Data* 7 (1): 144.  
<https://doi.org/10.1038/s41597-020-0486-7>
- Mahanti, Rupa. 2021. "Data and Its Governance". En *Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools*, editado por Rupa Mahanti, 5-82. Singapore: Springer.  
[https://doi.org/10.1007/978-981-16-3583-0\\_2](https://doi.org/10.1007/978-981-16-3583-0_2)
- OECD. 2022. "Going Digital to Advance Data Governance for Growth and Well-Being".  
<https://doi.org/10.1787/e3d783b0-en>
- Palsdottir, Agustá. 2021. "Data literacy and management of research data a prerequisite for the sharing of research data". *Aslib Journal of Information Management* 73 (2): 322-41.  
<https://doi.org/10.1108/AJIM-04-2020-0110>
- RAE. 2023. "Política". En *Diccionario de la lengua española*. Acceso 1 de febrero de 2023.  
<https://dle.rae.es/política>
- Redkina, N. S. 2019. "Current Trends in Research Data Management". *Scientific and Technical Information Processing* 46 (2): 53-58.  
<https://doi.org/10.3103/S0147688219020035>
- Shukla, Samiksha, Jossy P. George, Kapil Tiwari, y Joseph Varghese Kureethara. 2022. "Data Ethics". En *Data Ethics and Challenges*, editado por Samiksha Shukla, Jossy P. George, Kapil Tiwari y Joseph Varghese Kureethara, 61-72. Springer Briefs in Applied Sciences and Technology. Singapore: Springer.  
[https://doi.org/10.1007/978-981-19-0752-4\\_4](https://doi.org/10.1007/978-981-19-0752-4_4)
- Solomonides, A. 2019. "Research Data Governance, Roles, and Infrastructure". En *Clinical Research Informatics*, editado por Rachel L. Richesson y James E. Andrews, 291-310. Health Informatics. Cham: Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-98779-8\\_14](https://doi.org/10.1007/978-3-319-98779-8_14)
- Wang, Chen-Shu, Shiang-Lin Lin, Tung-Hsiang Chou, y Bo-Yi Li. 2019. "An Integrated Data Analytics Process to Optimize Data Governance of Non-Profit Organization". *Computers in Human Behavior* 101 (diciembre): 495-505.  
<https://doi.org/10.1016/j.chb.2018.10.015>
- Weber, Kristin, Boris Otto, y Hubert Österle. 2009. "One Size Does Not Fit All: A Contingency Approach to Data Governance". SSRN Scholarly Paper: Rochester, NY.  
<https://papers.ssrn.com/abstract=1728505>

Zurbriggen, Cristina. 2011. "Gobernanza: una mirada desde América Latina". *Perfiles latinoamericanos* 19 (38): 39-64.

*Para citar este texto:*

Ávila-Barrientos, Eder. 2023. "La gobernanza de los datos de investigación en el contexto de su organización y sistematización". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 45-64.

<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58763>

# La valoración de ejemplares dedicados y firmados en bibliotecas y colecciones patrimoniales

Manuel-José Pedraza-Gracia\*

*Artículo recibido:*  
28 de marzo de 2023  
*Artículo aceptado:*  
5 de mayo de 2023

*Artículo de investigación*

## RESUMEN

Se analiza el valor patrimonial que poseen los ejemplares contemporáneos con firma o dedicatoria de autor en bibliotecas y colecciones patrimoniales, y los aspectos y elementos imprescindibles para valorarlos. Estos ejemplares devienen en únicos, conformando “intradocumentos”: documentos soportados por otros documentos. Su importancia deriva de la información que aportan sobre el ejemplar, su autor o su poseedor. Se propone una gradación de las características para determinar su valor.

**Palabras clave:** Bibliotecas patrimoniales; Ejemplares dedicados; Ejemplares firmados; Valoración

\* Departamento de Ciencias de la Documentación e Historia de la Ciencia, Instituto de Patrimonio y Humanidades, Universidad de Zaragoza, España  
pedraza@unizar.es



## Valuation of inscribed and signed copies in libraries and heritage collections

Manuel-José Pedraza-Gracia

### ABSTRACT

It analyses the heritage value of signed or inscribed contemporary copies in libraries and heritage collections and the aspects and elements necessary to value them. These copies become unique, forming “intradocuments”: documents supported by other documents. Their importance is related to the information they provide about the copy, its author or its holder. A graduation of the characteristics is proposed to determine their value.

**Keywords:** Heritage libraries; Inscribed copies; Signed copies; Valuation

### INTRODUCCIÓN

En los libros se encuentran con asiduidad diferentes anotaciones manuscritas: *ex libris*, *ex donos*, dedicatorias, glosas, correcciones, *probationes penae*... que alteran el ejemplar en el que están incluidas, añadiendo o restándole méritos. Los ejemplares con estas adiciones pueden poseer distintos niveles de trascendencia patrimonial en función de la valoración que alcancen para un colectivo, una institución, un territorio, una nación o para el conjunto de la humanidad. Se entiende aquí por valoración la actuación de carácter técnico que pone de relieve las cualidades de un ejemplar, de una edición o de un fondo patrimonial. Al respecto, Julián Martín Abad (2004) dice: «La valoración es un proceso en el que mediante un examen minucioso, desde diversas perspectivas y con muy diferenciados saberes, se descubren los rasgos de una pieza o de un conjunto de piezas que nos permiten concretar su valor cultural». En consecuencia, no es posible llegar a confundir la valoración con la tasación, que consiste en hallar el precio del ejemplar o del fondo, ni con la valorización en la que intervienen todos los factores e investigaciones que dotan de valor(es) añadido(s) al ejemplar. *Cfr.* Manuel José Pedraza Gracia (2008, 2018).

Todas las incorporaciones manuscritas singularizan al ejemplar entre los de su edición, por lo que deviene en ejemplar único, informando sobre su génesis, su historia, sobre el poseedor o sobre los autores (Foster, 2017; Shammas,

2021). La calidad y cantidad de la información que incorporan los convierten en documentos dentro del documento –lo que se podría denominar «intradocumento»– y se convierten en una de las causas por las cuales algunos ejemplares contemporáneos son incorporados a los acervos patrimoniales conformando fondos específicos individualizados que ofrecen interés –y, consecuentemente, valor– bibliográfico, histórico, literario, científico... al margen de su antigüedad. Algunas bibliotecas, no exclusivamente patrimoniales, ofrecen los ejemplares que poseen de investigadores reconocidos para su dedicatoria, cuando éstos acuden a ellas como usuarios con el fin de incrementar este tipo de fondos. El problema que subyace en esta práctica es que, aunque el valor informativo de la adición se circunscribe a la indicación de la visita a la biblioteca, una vez firmados o dedicados, deben quedar exentos de préstamo, reservándose exclusivamente para ser consultados en sala.

Cuando se analizan las unidades bibliográficas que se integran en colecciones patrimoniales se citan seis grupos diferenciados (*Normas y directrices para bibliotecas universitarias y científicas*, 1999):

- 1) Los manuscritos (no necesariamente antiguos).
- 2) Las publicaciones antiguas, esto es, las anteriores a una fecha determinada que puede variar notablemente: 1820 (Universidad Complutense), 1901 (Universidad de Sevilla), 1958 (implantación del depósito legal)...
- 3) Los fondos de carácter unitario y valor histórico para la entidad.
- 4) Los ejemplares de publicaciones posteriores a la fecha establecida (en el apartado 2) que posean características singulares, tales como autógrafos o anotaciones de los autores, firmas, dedicatorias, encuadernaciones especiales, etc.
- 5) Los dibujos, grabados y mapas.
- 6) Cualquier ejemplar de publicación que se considere que debe ser objeto de una protección especial para su conservación.

Suelen incluirse a línea y renglón en los reglamentos de las universidades españolas, por ejemplo, los de la Universidad Complutense de Madrid o los de la Universidad de Zaragoza. Objeto de atención a la hora de la inclusión de un ejemplar en el fondo patrimonial es, como puede observarse, la condición de *unicum* y de rareza.

Esta atención especial a los ejemplares con autógrafos, anotaciones, firmados o dedicados (punto 4, aunque podrían tener cabida en el 1 o el 3) se fundamenta en que merecen ser investigados y conservados porque son únicos. Son de interés las anotaciones efectuadas por el autor después de difundida la edición en grado y con finalidades diferentes, y, dentro de este grupo de intervenciones, destacan

por su cantidad, calidad informativa, valores aportados y por su presencia en las colecciones patrimoniales públicas y privadas, las firmas y las dedicatorias. Este trabajo pretende, debido a la carencia de bibliografía científica al respecto, ofrecer una aproximación a la valoración patrimonial de ejemplares contemporáneos que tienen firmas y dedicatorias manuscritas de autor en los fondos de las bibliotecas públicas y privadas.

## FIRMAS Y DEDICATORIAS

El origen de la dedicatoria se ubica en tiempos remotos, cuando el autor ofrecía su obra al mecenas. Pero, ofrendándose la obra o la edición, este tipo de dedicatoria es general, pudiéndosela considerar como un género literario autónomo. Cuando lo que se pretende dedicar es el ejemplar, no la obra, es menester intervenir directamente en la unidad bibliográfica a la que afecta. Lo que procede entonces es que la intervención sea manuscrita. Las dedicatorias,

que varían mucho en la extensión y en la sinceridad, son muy importantes porque nos dan a conocer las relaciones y amistades de los autores, lo que puede contribuir a trazar la biografía de los escritores, especialmente la de aquellos que llegaron a la cumbre en el ámbito de la literatura... Como una derivación en esa costumbre existente en el mundo de las letras, ha surgido el coleccionismo de dedicatorias. La diferencia fundamental es que en el primer caso la iniciativa sale de los propios autores, mientras que en el segundo son los coleccionistas los que solicitan a los escritores que les escriban una dedicatoria, aunque sea ésa la única relación que existe entre ellos (Blázquez, 2009: 13).

Los primeros ejemplares contemporáneos dedicados y firmados conocidos aparecen en el siglo XIX. Estos primeros ejemplares no se dedicaban —sólo se firmaban— frecuentemente con el nombre propio o solamente indicando “el autor”. Pero hay que poner cuidado: un nombre o firma puede ser un *ex libris*. No obstante, no es difícil discernir entre una firma de autor —del texto, de la ilustración o de la edición— y una firma para indicar posesión. No todas las dedicatorias manuscritas tienen origen en el autor: traductores, editores literarios, ilustradores, etc. son posibles firmantes y dedicadores, y los valores que pueden aportar llegan a ser mayores que si las hubiese escrito el autor, ya que un ilustrador o un editor pueden ser más relevantes que el autor del texto. También es necesario diferenciar estas dedicatorias de las que pueda hacer una tercera persona que ha regalado el libro al poseedor. El interés y valor del ejemplar estará en función del dedicador y del dedicatario. Estas dedicatorias reciben en inglés el nombre de *gift inscriptions* que se podría traducir como “dedicatorias de regalo” o “dedicatorias de terceros”. Las

dedicatorias de regalo o de terceros adquieren valor cuando permiten ampliar el conocimiento sobre la red de relaciones de un personaje digno de estudio o las existentes entre dos personajes cuando se puede investigar la biblioteca de uno de ellos. En las bibliotecas patrimoniales hay cada vez más bibliotecas y archivos personales, que son el mejor objeto de estudio científico y biográfico sobre ese personaje. Salvo en estos casos, rara vez estas dedicatorias adquieren interés suficiente como para alcanzar nivel patrimonial. Desde la perspectiva bibliofílica tradicional esto se viene considerando un demérito del ejemplar.

Es necesario tratar sobre el valor de un ejemplar firmado o dedicado por el autor puesto que trasciende el estadio medio del ejemplar común de la edición en cuestión. Es imposible determinar el valor de una anotación de forma apriorística —sin un análisis ponderado— puesto que lo que hoy parece no tener valor podría llegar a poseerlo mañana y viceversa. El valor de un documento se modifica con el tiempo, más aún cuando interviene la evolución personal, profesional, histórica de los protagonistas de la dedicatoria. La estimación del valor siempre depende del observador, es subjetiva —no posee una unidad de medida— y variable. ¿Qué valor poseía un ejemplar con firma de Rafael Alberti el 28 de octubre de 1999?, ¿y el mismo ejemplar el 29, el día siguiente al de su fallecimiento? Podría responderse que los mismos valores en ambas fechas y, objetivamente, no se habían producido cambios en el ejemplar, sin embargo, la apreciación del valor, en cuanto subjetiva, ya no parece ser idéntica.

### EL INTERÉS DE LA FIRMA Y DEL LIBRO FIRMADO

En España e Iberoamérica, el coleccionismo de ejemplares firmados no posee la misma pujanza que en el mundo anglosajón y francés, en donde se presenta la disyuntiva en torno a la preferencia de la firma sobre la dedicatoria, fundamentada en el desinterés por ejemplares dedicados a otra persona. La facilidad para falsificar una firma y el número de falsificaciones existentes va cambiando esta percepción.

Una de las preguntas que me han hecho con más frecuencia en los últimos años es: “¿Qué es mejor, tener un libro firmado por el autor o dedicado?” En general, mi respuesta ha sido que cuanto más escriba el autor en un libro, mejor. Y especialmente he alentado a los coleccionistas a que, cuando firman sus libros, los dedique personalmente el autor... cuando compras un libro firmado estás comprando una firma, pero cuando compras un libro dedicado estás obteniendo una historia... cuanto más escriba el autor en un libro, mejor. (Lopez, 1997)

En los ejemplares con autógrafos, a la posible rareza de la edición se añade la de la dedicatoria. Desde una perspectiva bibliofílica, el atractivo de la firma está

sustentado en cierto componente fetichista: un individuo desea poseer algo tan particular como la firma de un personaje importante, mucho más cuando esa firma se ha escrito exclusivamente para dedicar el ejemplar al poseedor, reflejando el momento en el que el autor transmitió un sentimiento o un deseo hacia otra persona pretendiendo poseer ese momento. El colmo es que incluya cualquier expresión que manifieste proximidad o afecto con el poseedor del libro. En la biblioteca patrimonial el valor de una firma es relativo (número de firmas conservadas, importancia del personaje...), sin embargo, el valor de la “historia” del testimonio que aporta una dedicatoria añade a los de la firma los propios de la información que transmite. La investigación que realiza el bibliotecario, el bibliófilo, el librero... es la que revela esa historia para el resto de ciudadanos e investigadores.

Por lo que respecta a los ejemplares firmados (*signed books*, en inglés), se pueden presentar dos posibilidades:

- a) La firma del autor sin más aditamento (*Figura 1*), debida a diferentes causas, desde ejemplares firmados preparados para receptores no nominados, hasta el desinterés del autor... La importancia del ejemplar se circunscribe a la posesión de la firma autógrafa del autor.
- b) Puede tener un origen editorial evidente. La firma se incluye bien como mecanismo de control de tirada, bien como atractivo para el coleccionista. Estos ejemplares podrían denominarse “ejemplares con firma editorial” (*Figura 2a*). El único interés que poseen es conocer la firma del autor. Sin embargo, es posible encontrar ejemplares carentes de firma que adquieren, por ser más raros, un valor y un precio más elevado (*Figura 2b*).

Cuando los ejemplares están dedicados por el autor (*inscribed books*), la presencia de la dedicatoria está derivada de tres situaciones principales:

- a) La donación de un ejemplar por un autor a una persona o institución para dar a conocer su obra —como carta de presentación— o potenciar su imagen —como actividad mercadotécnica—, algo muy frecuente en una tipología documental actualmente desaparecida: la separata (*Figura 3*). Estos ejemplares se denominan ejemplares de presentación o *presentation copy*, en inglés. Es especial cuando el contenido del documento hace referencia al dedicatario (*Figura 3*).



Figura 1: Firma de Álvaro Mutis con bolígrafo negro, en la portada de *Diario de Lecumberri*. México: Utopía, 1976  
Fotografía: Librería La Social. Galería y Libros (Barcelona)

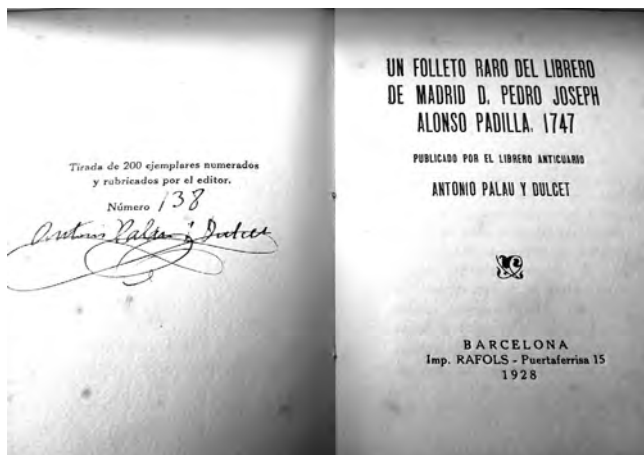


Figura 2a: Firma de Antonio Palau y Dulcet con pluma negra, en la anteportada del ejemplar y mención manuscrita del número del ejemplar (138) de: *Un folleto raro del librero de Madrid D. Pedro Joseph Alonso Padilla, 1747*. Barcelona: Imp. Rafols. 1928. Ejemplo de ejemplar con firma editorial.  
Fotografía del autor. Colección particular

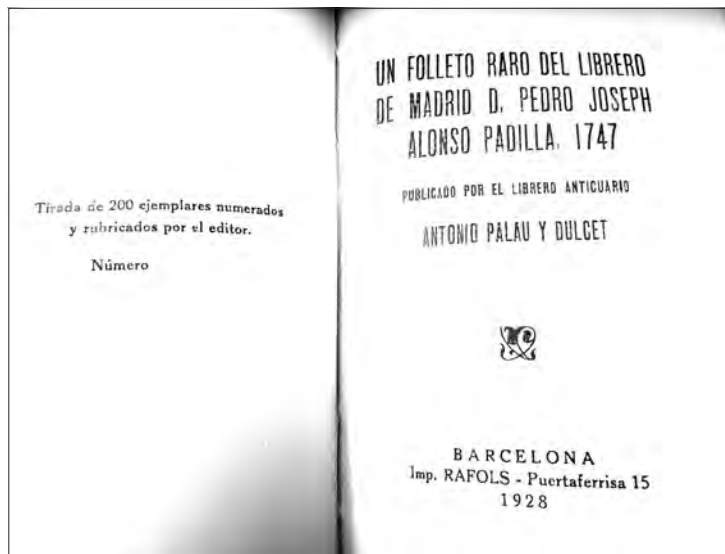


Figura 2b: Antonio Palau y Dulcet, *Un folleto raro del librero de Madrid D. Pedro Joseph Alonso Padilla, 1747*. Barcelona: Imp. Rafols. 1928. Ejemplar carente de firma editorial. Fotografía del autor. Colección particular

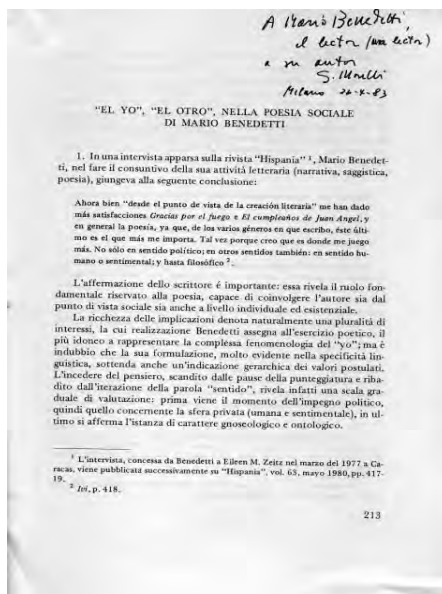


Figura 3: Dedicatoria datada y firmada por Gabrieli Morelli a Mario Benedetti con rotulador negro, en Milán el 26 de julio de 1983, en el encabezado de una separata de: «“El yo”, “el otro”, nella poesia sociale di Mario Benedetti», *Studi di letteratura ispano-americana*, 13-14 (1983). Fotografía: Centro de Estudios Literarios Iberoamericanos Mario Benedetti

- b) Cuando el ejemplar posee interés por el dedicator y el dedicatario se le denomina en inglés *association copy*, lo que en castellano podría traducirse como «ejemplar de vinculación» o «ejemplar de asociación», para mantener cierta simetría con la lengua inglesa (Figura 4). Los ejemplares dedicados como consecuencia de la relación entre dos colegas, o de parentesco, son también ejemplares vinculados o de asociación.

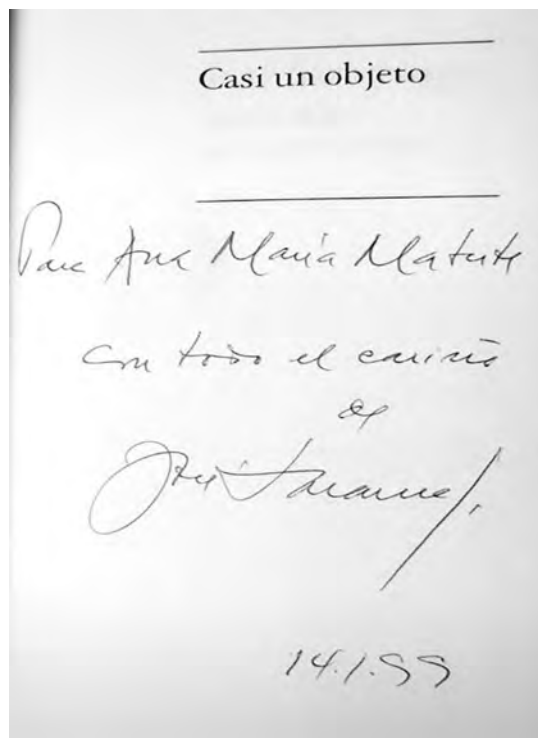


Figura 4: Dedicatoria datada y firmada por José Saramago a Ana María Matute con bolígrafo azul, el 14 de enero de 1999, en la anteportada de: *Casi un objeto*. Madrid: Alfaguara, 1998. Ejemplo de ejemplar de vinculación. Fotografía: Librería La Social. Galería y Libros (Barcelona)

- c) Finalmente, cuando la dedicatoria está destinada a quien se dedicó la edición, y así figura en ella, recibe el nombre de *dedication copy*, que en castellano podría denominarse «ejemplar de homenaje» (la designación «ejemplar de dedicatoria» podría dar lugar a una generalización). Son ejemplares que sólo pueden darse una vez por edición, haciéndolos extremadamente raros y suelen ser muy buscados y valorados.



- d) Frecuentemente un lector persigue la personalización de un ejemplar, lo que resulta favorecido por editoriales y librerías en ferias del libro. La dedicatoria deviene en una especie de certificado de exclusividad, que implica un vínculo (si bien fugaz) con el autor al conocerle en una feria, presentación o lectura pública, y cruzar unas palabras con él. Se han llegado a firmar más de 2 000 ejemplares en una sesión, como lo ilustra una nota periodística de Helena Evia (2008). Estos «ejemplares de promoción» (Figura 5), sin más valor que el de conocer la firma, pueden alcanzar precios notables.

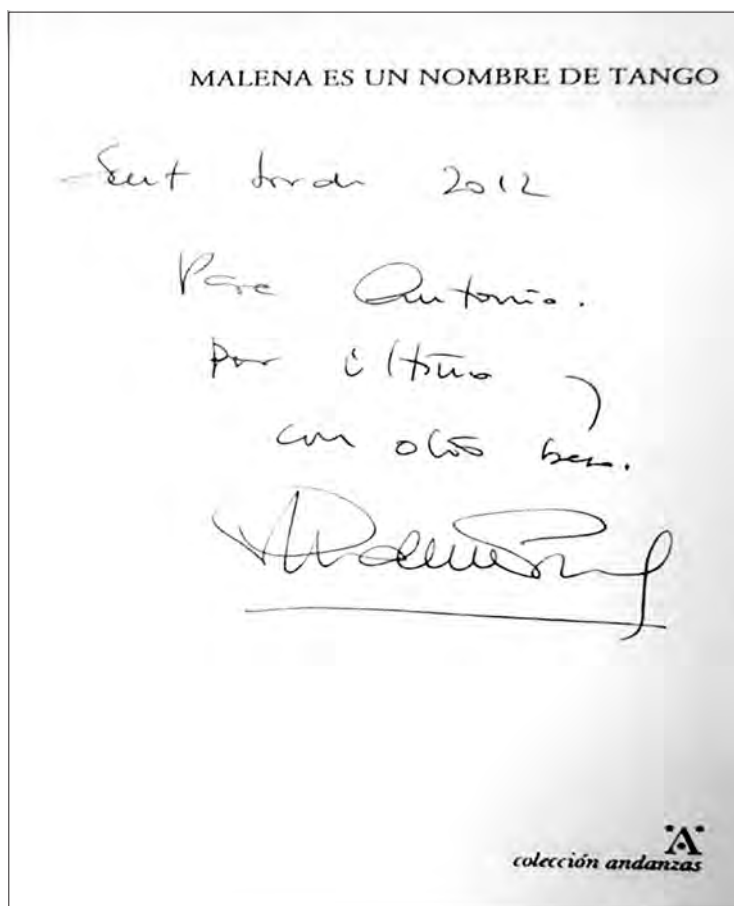


Figura 5: Dedicatoria datada y firmada de Almudena Grandes con bolígrafo azul, el "Día de San Jordi" de 2012, en la anteportada de: *Malena es nombre de tango*. Barcelona: Tusquets, 1994.

Ejemplo de ejemplar de promoción.

Fotografía: Librería La Social. Galería y Libros (Barcelona)

Esta clasificación resulta más evidente en la siguiente *Tabla*:

<b>Libro firmado</b> ( <i>Signed books</i> )	Sólo contiene la firma del autor	
	Ejemplar firmado	Sólo algunos ejemplares contienen la firma del autor
	Ej. con firma editorial	Todos los ejemplares poseen la firma del autor
<b>Libro dedicado</b> ( <i>Inscribed books</i> )	Contiene la dedicatoria (con o sin firma u otros elementos)	
	Ej. de presentación ( <i>Presentation copy</i> )	Dedicatoria para dar a conocer la obra entre otros autores e instituciones
	Ej. de vinculación ( <i>Association copy</i> )	Dedicatoria originada por la relación de amistad intelectual entre dos colegas
	Ej. de homenaje ( <i>Dedication copy</i> )	Dedicatoria que tiene como destinatario aquél a quien se dedicó la edición impresa
	Ej. de promoción	Dedicatoria ejecutada en una feria o presentación de la edición

*Tabla 1.* Somera clasificación de los ejemplares firmados y dedicados

No obstante, también suele denominarse a todos estos ejemplares dedicados como *association copies*, añadiendo aun aquellos que manifiestan alguna relación de cualquier tipo entre la posesión de un ejemplar con la edición, el autor o la propia obra.

John Carter y Nicholas Barker (2010: 27) indican: «Una *association copy* es un ejemplar que perteneció o fue anotada por el autor; o que en alguna ocasión perteneció a alguien relacionado con el autor o alguien reconocido por sí mismo; o, y quizás lo más interesante, por pertenecer a alguien especialmente relacionado con su contenido». Además, se incluye en este amplio ámbito «el libro dedicado por su autor a una persona famosa». La Independent Online Booksellers Association (ca. 2000) establece una clasificación y aclara:

En realidad, se aplica a cualquier ejemplar cuya relación con una persona relacionada con el libro esté plasmada en él. Se podría establecer la siguiente clasificación:

1. Un ejemplar propiedad del autor de la obra.
2. Un ejemplar propiedad de alguien involucrado en la producción de la edición; el ilustrador, el editor, etc.
3. Un ejemplar propiedad de alguien relacionado estrechamente con el autor del libro.
4. Un ejemplar propiedad de alguien relacionado directamente con el contenido del libro (la biografía que perteneció al biografiado, por ejemplo).

5. Un ejemplar propiedad de alguien conectado tangencialmente con el contenido del libro.
6. Un ejemplar regalado por el autor y firmado por éste.

Se puede resumir en dos grupos: aquellos que tienen que ver con la procedencia de los ejemplares, y los firmados y dedicados por el autor, que son el objeto de este trabajo.

## LOS EJEMPLARES DEDICADOS EN LA BIBLIOTECA PATRIMONIAL

La valoración de un ejemplar (no necesariamente antiguo) es una tarea compleja en la que se emplean conocimientos muy especializados para apreciar las cualidades de una obra, edición, variante, ejemplar, encuadernación, anotaciones, procedencias, y establecer su trascendencia. Los ejemplares contemporáneos dedicados o firmados se incluyen dentro de fondos patrimoniales porque aportan valores muchas veces únicos o en diferentes direcciones.

Una dedicatoria, una firma del autor, implica, aunque no siempre, la existencia de una relación entre dedicador y destinatario, acercando el personaje al investigador, el origen de sus posiciones, intereses, aspiraciones... Entonces, ¿qué se debe tener presente a la hora establecer los valores de un ejemplar dedicado o firmado?

El interés social, histórico, literario, político... del dedicador determina su valor. Una firma no siempre implica la existencia de una relación, algo que ocurre en los ejemplares de promoción obtenidos en firmas programadas de ejemplares. Lo único que se encuentra manuscrito en el ejemplar es la firma sin más aditamentos. Ésta se puede presentar de maneras diferentes: sólo el nombre formal, un diminutivo, el nombre y apellido(s), o el mote o alias, o incluyendo pequeños dibujos. El valor que aportan estos ejemplares para la biblioteca es discreto. La trascendencia del autor sería un factor determinante, resultando de interés los de autores verdaderamente relevantes, aunque se trate de ejemplares de promoción. En otros casos, tan sólo si se acompaña de dibujos o datos de relieve deberían tenerse en cuenta, lo que no es infrecuente y muy particular en algunos autores (*Figuras 10, 11 y 12*). Por el contrario, en el ámbito privado la firma adquiere trascendencia, siendo estos ejemplares buscados y coleccionados.

En especial hay aspectos morfológicos característicos de las firmas de determinados autores... que les dota de autenticidad y que facilitan su identificación (*Figuras 6 y 7*).

El segundo aspecto es analizar el destinatario. De la misma manera que con el dedicador, el interés de la persona o institución a quién se dedica debe ser considerado. No siempre es fácil identificar su nombre, por la cursividad de la escritura

o por aparecer oculto bajo apelativos personales o cariñosos. Por esta causa se requiere una indagación, puesto que el desconocimiento de la relación entre los referidos en la dedicatoria puede dar lugar a la pérdida de una información valiosa para el investigador.

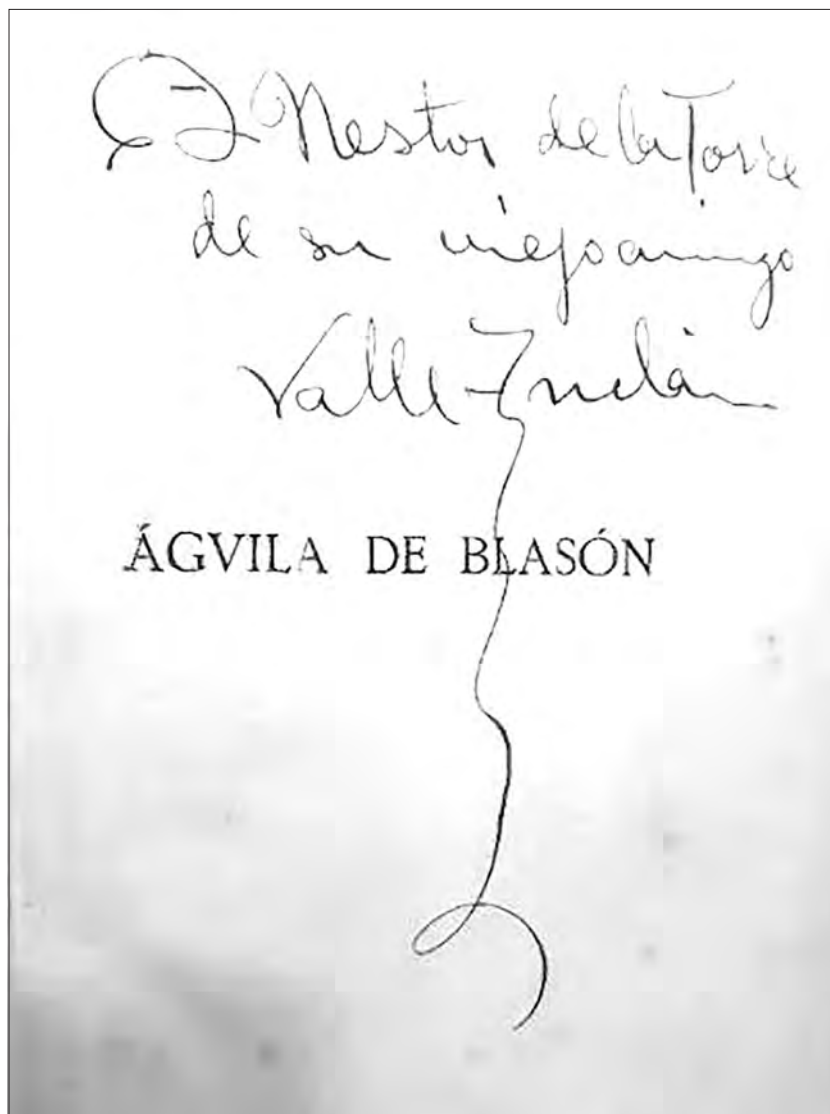


Figura 6: Dedicatoria firmada por Ramón del Valle-Inclán al pintor Néstor de la Torre con pluma negra, en anteportada de: *Águila de Blasón*. Madrid: Imprenta de Sáenz hermanos, 1922. Fotografía: Librería Lauso Books (Barcelona)

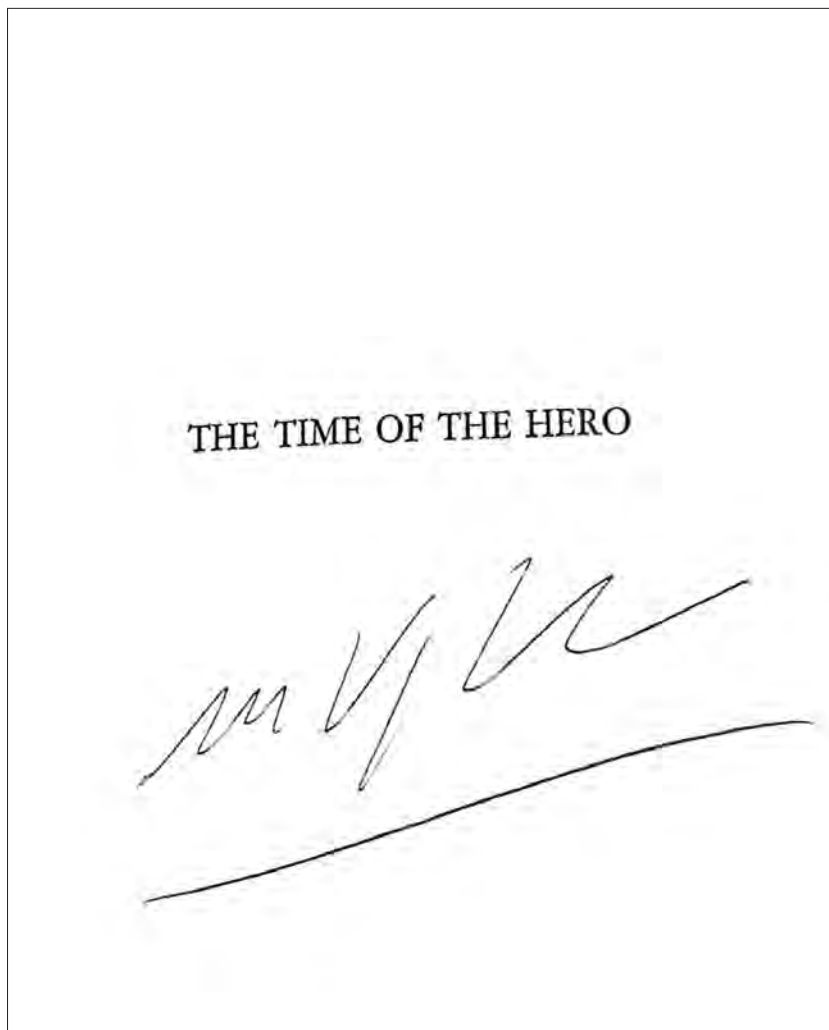


Figura 7: Firma de Mario Vargas Llosa con bolígrafo negro, en anteportada de:  
*The Time of Hero*. Brattleboro: Grove, 1966.  
Fotografía: Librería La Social. Galería y Libros (Barcelona)

El texto de la dedicatoria y cómo se hace son particularidades importantes. Una dedicatoria excesivamente formal o que, por el contrario, manifieste aprecio, una dedicatoria exigua o muy extensa, una genérica o concreta, son indicativos de una situación específica y del grado e intensidad de la relación entre los personajes referidos en ella. Ocasionalmente se incluyen informaciones de interés biográfico o histórico (*Figura 8*).

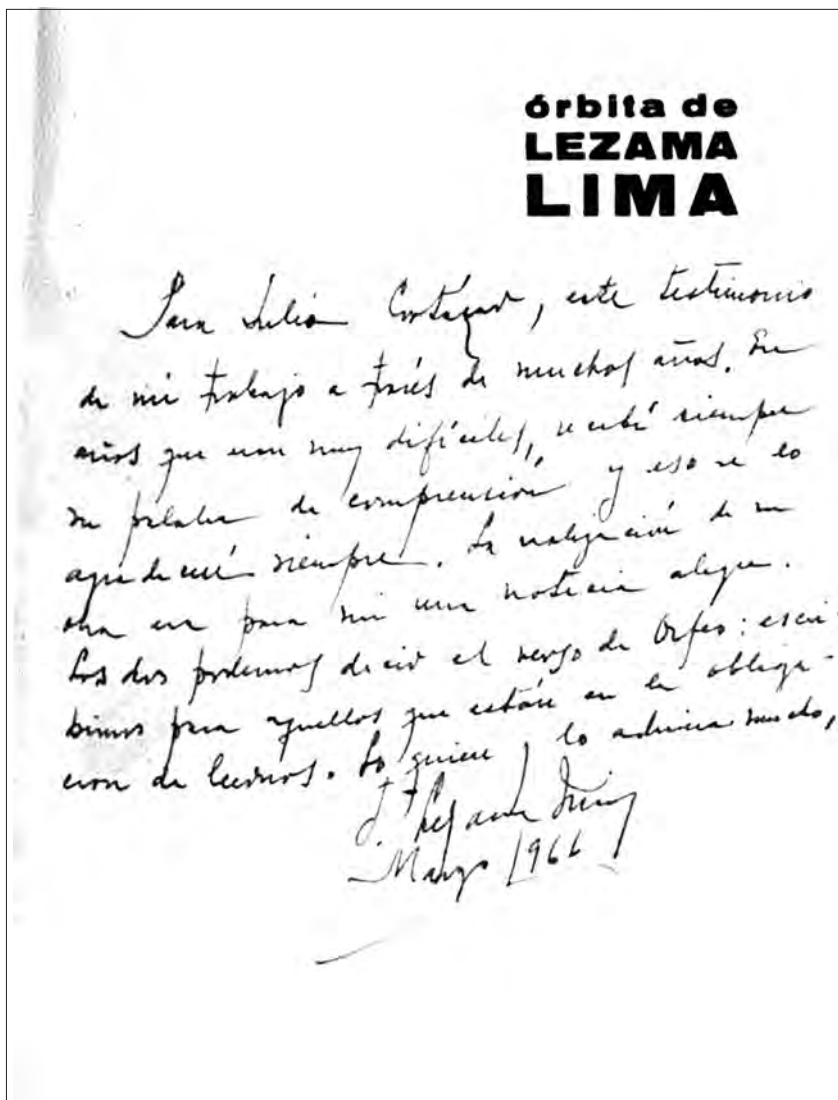


Figura 8: Dedicatoria datada y firmada de José Lezama Lima a Julio Cortázar, con bolígrafo azul, en mayo de 1966, en anteportada de: *Órbita de Lezama Lima*. La Habana: Unión de Escritores y Artistas de Cuba, 1966.

Fotografía: Fundación Juan March. Biblioteca [2000-...]

El lugar y la fecha de la firma o dedicatoria (*data tópica* y *data crónica*), cuando las hay, dan lugar a la obtención de información atinente de tipo biográfico de importancia para la investigación (Figura 9).



Figura 9: Dedicatoria datada y firmada de José Agustín Goytisolo, con pluma negra, en anteportada, en Barcelona, el 25 de junio de 1958, en: *Salmos al viento*. Barcelona: Instituto de Estudios Hispánicos, 1958.  
Fotografía: Llibreria Antiquària Farré

La presencia de dibujos, poemas... incorpora matices que no pueden ser obviados a la hora de valorar la dedicatoria y el ejemplar (Figura. 10, 11 y 12) que expresan una atención especial.

Por último, se ha de citar el lugar del ejemplar en el que se ha escrito la dedicatoria (portada, anteportada, cubierta...) y el método empleado (pluma, bolígrafo, lápiz... y el color de la tinta). Estas indicaciones no son intrascendentes, puesto que el uso de plumas, bolígrafos o rotuladores y determinadas tintas puede ayudar a la determinación de la autenticidad de las dedicatorias y firmas (Figura 10).



Figura 10: Dedicatoria datada y firmada de Rafael Alberti a Dolores Franco Manera con rotulador azul, en hoja de guarda, y representación marina con rotuladores verde, amarillo y morado en Roma, el 13 de diciembre de 1971, en: *Antología poética* (1924-1952), 5ª ed. aum. Buenos Aires: Losada, 1969. Fotografía del autor. Colección: Biblioteca Marqués de Valdecilla (Madrid)

Todos estos elementos, con sus diferentes particularidades, conforman lo que se ha denominado «intradocumento», un documento que se ha incluido en otro documento con el que se relaciona, que proporciona información relevante sobre la vida del ejemplar, la del autor o la del poseedor. De esta manera, este intradocumento es el que dota de «patrimonialidad» al ejemplar o, para ser más exactos: el documento que lo contiene se torna patrimonial por ser el soporte material del intradocumento al que se une, y no sólo por eso, sino por ser referente imprescindible para comprenderlo.





Figura 11: Dedicatoria y firma de Alejo Carpentier con una partitura, con bolígrafo azul, en anteportada de: La música en Cuba. México: Fondo de Cultura Económica, 1946. Fotografía: Librería La Social. Galería y Libros (Barcelona)

Estos ejemplares firmados han llegado a las bibliotecas de tres maneras diferentes: a) fueron enviados directamente por el autor, el dedicatario es la biblioteca o sólo llevan su firma; b) han llegado mediante legados o adquisiciones de bibliotecas en las que los ejemplares tienen un único dedicatario. Hay todavía una tercera vía de carácter más o menos artificial, que puede ser práctica frecuente de algunos bibliotecarios: c) la solicitud de la firma del autor en los ejemplares que ya posee de él la biblioteca con motivo de su visita como investigador.

Cada una de estas tres posibilidades ofrece informaciones diferentes. En el primero de los casos puede llegar a justificarse una relación del dedicante con

la biblioteca: bien sea por la pertenencia a la institución a la que está adscrita la biblioteca, o por una investigación prolongada en la misma..., en este caso, lo relevante es la manifestación de los documentos que han servido para la elaboración de la investigación dedicada. En el segundo, el reconocimiento de los dedicadores permite establecer la red de relaciones que mantuvo el primer propietario del fondo que posee la biblioteca y confluye en el mejor conocimiento del mismo. Jesús Marchamalo (2009: 49) lo refleja así: «Hace años se vendió parte de la biblioteca de Mercedes Fórmica. Afortunadamente, nadie eliminó su nombre, ni lo borró o alteró. Compré algunos de sus libros, y es curioso ver cómo era su relación con otros autores a través de sus dedicatorias». En el tercero de los casos solamente revela la presencia del dedicante en la biblioteca.

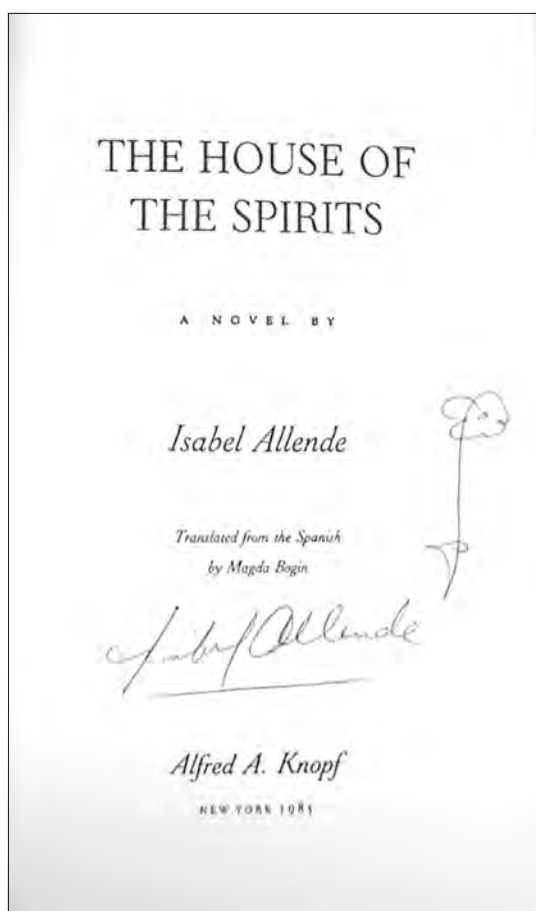


Figura 12: Firma de Isabel Allende con el dibujo de una flor, con bolígrafo negro, en la portada de: *The House of the Spirits*. New York: Alfred A. Knopf, 1985.  
Fotografía: La Social. Galería y Libros (Barcelona)

Son documentos únicos con informaciones de alcance desigual para la investigación y el conocimiento sobre la biblioteca o sobre los dedicatarios. Por esta causa, es necesario incorporar en el registro bibliográfico, entre los datos del ítem o ejemplar, la información completa referente a la firma o dedicatoria, con el objeto de que el usuario los conozca y pueda calibrar el interés que posee para su investigación. No resulta indiferente si el ítem sólo está firmado o dedicado y datado, si tiene el nombre del dedicatario y su posible relación con el autor, si incorpora un dibujo o un poema e, incluso, la descripción del procedimiento empleado para su escritura.

La evolución tecnológica de la difusión de la investigación ha hecho desaparecer las separatas, estados desgajados de la publicación ex profeso para ser firmadas y enviadas a los colegas o personas unidas afectivamente al autor, frecuentes hace sólo algunas décadas. Su valor es siempre relativo, puesto que el contenido está recogido en la publicación en la que aparecieron, no siendo raros en cuanto al contenido, aunque lo sean por su escasa tirada exenta.

## CONCLUSIÓN

El ejemplar que posee un autógrafo de su autor, ilustrador, traductor... con una dedicatoria ha entrado a formar parte de las bibliotecas patrimoniales en las que tiene su lugar debido a que estas anotaciones lo hacen devenir en único. El conjunto de valores que aportan tiene una relación unívoca con la información que transmiten.

Las dedicatorias manifiestan un vínculo entre dedicador y dedicatario. El conocimiento de estos vínculos proporciona una mejor percepción del autor, acercando al investigador al personaje, a la raíz de sus posiciones, intereses, empeños...

Es imprescindible manifestar en el registro bibliográfico quién firma, a quién va dirigido y el grado de relación reflejado en la dedicatoria: quién, a quién y qué, y, si es posible, cuándo y dónde, si incorpora expresiones adjuntas y el procedimiento usado para plasmarlo, porque todos ellos son necesarios para el investigador interesado en estos intradocumentos.

Estos ejemplares conforman fondos de especial valor en las bibliotecas patrimoniales que quedan independizados, por sus características, del resto de los fondos.

Para el bibliófilo, el encanto del autógrafo está apoyado en cierto componente fetichista, más exacerbado cuando la firma se ha plasmado para dedicar el ejemplar al poseedor.

## REFERENCIAS

- Blázquez, Guillermo. 2009. «Presentación», en *Dedicatorias: un siglo de libros dedicados: exposición formada en base a la colección de libros dedicados de José Luis Sánchez de Vívar*. Madrid, Gremio Madrileño de Libreros de Viejo.
- Carter, John; Nicholas Barker. 2010. *ABC for Book Collectors*. New Castle: Oak Knoll Books.
- Centro de Estudios Literarios Iberoamericanos Mario Benedetti. 2020. “*Querido Mario...: dedicatorias en la biblioteca madrileña de Mario Benedetti*”. Alicante, Centro de Estudios Literarios Iberoamericanos Mario Benedetti. [Consulta: 20/09/2022].  
<https://web.ua.es/es/centrobenedetti/querido-mario.-.-dedicatorias-en-la-biblioteca-madrilena-de-mario-benedetti-comisarias-eva-valero-juan-y-monica-ruiz-banuls.html>
- Consortium of European Research Libraries. 2010. *CERL Provenance Digital Archive*. London, CERL. [Consulta: 11/03/2022]  
<https://www.cerl.org/resources/provenance/main>
- Evia, Helena. 2008. «8 firmas por minuto: Ruiz Zafón, asediado por el público, estampó su rúbrica en 2.500 libros». *El Periódico* (24/04/2008). [Consulta: 28/09/2022].  
<https://www.elperiodico.com/es/actualidad/20080424/ocho-firmas-minuto-224332>
- Foster, Graham. 2022. *Burgess's Library of Inscribed Books*. Manchester: International Anthony Burgess Foundation. [Consulta: 25/04/2022].  
<https://www.anthonymburgess.org/object-of-the-week/object-week-burgesss-library-inscribed-books/>
- Fundación Juan March. Biblioteca. 2000. *Biblioteca digital de Julio Cortázar*. Madrid: Fundación Juan March. [Consulta: 11/03/2022].  
<https://www.march.es/es/coleccion/biblioteca-julio-cortazar/ficha/orbita-lezama-lima-3097>
- Lopez, Ken. 1997. *Signed vs. Inscribed*, Hadley, Ken Lopez Bookseller, Fine Books & Manuscripts. [Consulta: 11/03/2022].  
<https://lopezbooks.com/articles/signed/>
- Marchamalo, Jesús. 2009. «Formas de dedicar». En *Dedicatorias: un siglo de libros dedicados: exposición formada en base a la colección de libros dedicados de José Luis Sánchez de Vívar*, 17-57. Madrid: Gremio Madrileño de Libreros de Viejo.
- Martín Abad, Julián. 2004. «La valoración del libro: el punto de vista del bibliotecario de fondo antiguo». *Documentos de trabajo U.C.M. Biblioteca Histórica* 8: 1-25. [Consulta: 18/11/2021].  
<https://eprints.ucm.es/id/eprint/5698/>
- Normas y directrices para bibliotecas universitarias y científicas*. 1999. 2ª ed. aum. Madrid: Ministerio de Educación y Cultura-Dirección General del Libro, Archivos y Bibliotecas.
- Pedraza Gracia, Manuel José. 2018. «Representación, información, identificación y valoración del libro antiguo: bibliotecarios e investigadores ante un problema común». *Scire: Representación y organización del conocimiento* 24 (1): 23-34.
- Pedraza Gracia, Manuel José. 2008. «La valoración de los fondos antiguos en bibliotecas». *Ibersid: revista de sistemas de información y documentación - Journal of Information and Documentation Systems*, no. 2: 263-272.
- Pierce, Patricia. 2004. *The Great Shakespeare Fraud: The Strange, True Story of William-Henry Ireland*. Stroud: Sutton.
- Shammas, Victor Lund. 2021. *Michel Foucault's Library of Inscription Copies*. University of Agder. [Consulta: 25/04/2022].  
<https://www.victorshammas.com/blog/2021/4/30/michel-foucaults-library-of-inscription-copies>

- Universidad Complutense de Madrid. Biblioteca. 1992). *Reglamento para uso y conservación de los fondos antiguos y valiosos de la biblioteca de la Universidad Complutense de Madrid (aprobado por Junta Gobierno el día 18 de diciembre de 1992)*. Madrid: Universidad Complutense. [Consulta: 07/03/2022].  
<https://biblioteca.ucm.es/data/cont/docs/60-2014-03-25-reglamento280.pdf>
- Universidad Complutense de Madrid. Biblioteca Histórica Marqués de Valdecilla. [2010-...]. *Antiguos poseedores en las colecciones de la Biblioteca Complutense*. Madrid: Universidad Complutense. [Consulta: 11/03/2022].  
<https://biblioteca.ucm.es/historica/procedencias-1>
- Universidad de Salamanca. Biblioteca General. 2005. *Proyecto Exlibris*. Salamanca, Universidad de Salamanca. [Consulta: 11/03/2022].  
<http://bibliotecahistorica.usal.es/es/recursos/antiguos-poseedores>
- Universidad de Zaragoza. Biblioteca. 2010. *Normas de uso del fondo histórico*. Zaragoza: Universidad de Zaragoza. [Consulta: 07/03/2022].  
[http://biblioteca.unizar.es/sites/biblioteca.unizar.es/files/documentos/norma\\_uso\\_fondo\\_historico.pdf](http://biblioteca.unizar.es/sites/biblioteca.unizar.es/files/documentos/norma_uso_fondo_historico.pdf)
- Universitat de Barcelona-CRAI. 2005. *Antiguos poseedores*. Barcelona, Universitat de Barcelona. [Consulta: 11/03/2022].  
<https://marques.crai.ub.edu/es/impresores>

### *Para citar este texto:*

Pedraza-Gracia, Manuel-José. 2023. “La valoración de ejemplares dedicados y firmados en bibliotecas y colecciones patrimoniales”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 65-86.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58781>

# **Las necesidades de información y comportamiento informativo de los usuarios de sistemas de información geográfica en comunidades de Facebook**

Eduardo Waybel Sánchez\*

*Artículo recibido:*

*5 de marzo de 2023*

*Artículo aceptado:*

*29 de mayo de 2023*

*Artículo de investigación*

## **RESUMEN**

El objetivo principal de esta investigación es identificar las necesidades de información y el comportamiento informativo de usuarios de sistemas e información geográfica o geousuarios pertenecientes a comunidades de la red social Facebook. A partir de métodos tanto cualitativos como cuantitativos, se reconoce que este segmento de geousuarios cuenta con competencias y habilidades tecnológicas para la búsqueda, recuperación de información geográfica e interacción con los sistemas; sin embargo, se detecta que entre las necesidades primordiales están las que conciernen al conocimiento de fuentes de información geográfica y la capacitación en SIG, aspectos que abren la posibilidad de que las bibliotecas y otras

\* Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México, México  
eduardowaybel@gmail.com

unidades de información ofrezcan servicios potenciales relacionados con la geoinformación.

**Palabras clave:** Estudios de usuarios; Sistemas de información geográfica; Información geográfica; Geousuarios; Necesidades de información

### **The information needs and information behavior of geographic information systems users in Facebook communities**

*Eduardo Waybel Sánchez*

#### **ABSTRACT**

The main objective of this research is to identify the information needs and information behavior of users of geographic information systems or geousers belonging to communities of the social network Facebook. From both qualitative and quantitative methods, it is recognized that this segment of geousers have technological competences and skills for searching, retrieving geographic information and interacting with systems; however, it is detected that among the main needs are the knowledge of geographic information sources and GIS training, aspects that open the possibility for libraries and other information units to offer potential services related to geoinformation.

**Keywords:** User studies; Geographic information systems; Geographic Information; Geousers, Information needs

#### **INTRODUCCIÓN**

Los usuarios de información geográfica o geousuarios son comunidades que cuentan con competencias y habilidades tecnológicas, porque al hacer uso tanto de la información geográfica (IG, en adelante), como de los sistemas de información geográfica (SIG, en adelante) pueden tener acceso a fuentes de información geográfica para recuperar e importar archivos digitales con el objetivo de transformar la información y así satisfacer alguna necesidad.

Los geousuarios, al ser un segmento relacionado con el uso de las tecnologías, aprovechan el potencial de la *web* 2.0 para comunicarse con otros usuarios dentro de plataformas o redes sociales. Los grupos de Facebook funcionan como canales de interacción, donde se realizan actividades tales como responder y publicar preguntas relacionadas con la IG y los SIG, buscar recomendaciones, descargar herramientas, agilizar procesos, hacer uso de *software* o, sencillamente, compartir experiencias, resultando ser una red sólida de participación y retroalimentación entre los integrantes de la comunidad (Waybel, 2022: 95).

Los estudios de usuarios son un área dentro de la Bibliotecología y los Estudios de la Información, que tienen como objetivo determinar las necesidades de información, identificar el comportamiento informativo y el nivel de satisfacción de diferentes comunidades de usuarios, por lo tanto, los geousuarios son un segmento que debe ser estudiado debido a sus características particulares.

Para estudiar a los geousuarios de comunidades de grupos en Facebook, es importante conocer el contexto en el que se desarrollan, desde la perspectiva tecnológica y social; por ello, el objetivo principal del presente estudio sobre usuarios es identificar las necesidades de información y el comportamiento informativo de dicha comunidad.

Es importante mencionar que los geousuarios de grupos de Facebook no necesariamente son una comunidad ligada a las bibliotecas y otras unidades de información, sin embargo, se vislumbra que estas instituciones podrían tener cierta área de oportunidad para atender sus requerimientos informativos.

En este aspecto, si se identifican las necesidades de información y el comportamiento informativo de los geousuarios, se abre la posibilidad de ofrecer, dentro de las unidades de información, mejores resultados en procesos de toma de decisiones, colecciones, usuarios, proyectos y servicios bibliotecarios geográficos.

Cabe mencionar que los datos que se presentan como resultados de este estudio se derivan de la tesis de Waybel Sánchez (2022: 94-97) titulada “Los sistemas de información geográfica desde la perspectiva de los bienes comunes de información: posibles aportes de la geobiblioteca y del geobibliotecario”, publicada por la Universidad Nacional Autónoma de México.

## ANTECEDENTES

### ***Panorama nacional e internacional de la red social Facebook***

Debido al uso de las tecnologías y del internet, actualmente los usuarios conectados pueden crear y unirse a comunidades relacionadas por sus gustos, formas de vida, pensamientos, etcétera, en donde realizan diferentes actividades sociales,



siendo la difusión de información una de las actividades principales. Y esto, en tanto, gracias a la denominada *web 2.0*, la cual se caracteriza por la participación activa en la red. De acuerdo con Tomeo, en la *web 2.0* “el usuario intercambia contenidos, opina, forma grupos de referencia, ejerce amplio poder de influencia en los demás y genera nuevas relaciones interpersonales por medio de los vehículos y aplicaciones que facilita la propia *web 2.0*” (2014: 4).

La *web 2.0* es fundamental para el desarrollo y el uso de medios sociales, canales de comunicación y aplicaciones, entre otras plataformas. Una de las principales es Facebook, una red social que más de 2 700 millones de personas, en todo el mundo, utilizan a diario para compartir con su círculo social publicaciones de todo tipo: “la esencia de Facebook es muy sencilla: es una red que conecta personas con personas” (Berlanga, 2021).

Según el estudio “Digital 2021”, publicado por Hootsuite y We are Social (julio 2021), Facebook es la red social más usada a nivel mundial, ubicándose en la segunda posición como la plataforma favorita de los usuarios de internet en el rango de edad de 16 a 64 años (21.5%). El estudio también muestra las razones por las cuales los usuarios (del mismo rango de edad) usan las plataformas sociales, destacando las siguientes: compartir y discutir opiniones con otros (24.9%), para investigación y redes de trabajo (22.7%), encontrar comunidades afines o grupos de interés (22.0%).

Dentro del panorama nacional, Digital 2021 (México), muestra que Facebook es la segunda plataforma más utilizada por los usuarios entre 16 a 64 años (95.3%) con un rango de permanencia de 29.6 horas mensuales. Por otra parte, el “17°. Estudio sobre los Hábitos de los Usuarios de Internet en México” 2021 (mayo 2021), muestra que los internautas mexicanos utilizan internet para acceder a redes sociales (86.8%), leer / ver contenido relevante (20.2%), entre otros datos.

Facebook es una plataforma que permite la interacción entre sus comunidades de usuarios. Parte fundamental de esta red social son los denominados *grupos de Facebook*, los cuales “sirven para compartir intereses comunes entre los usuarios. Un *grupo* es un espacio en el que el usuario podrá dejar sus opiniones, enlaces, contenidos, comentarios... Además, aparte de participar en ellos, Facebook nos permite crearlos y gestionarlos” (Alcaraz, 2018).

Debido a su importancia, los grupos de Facebook y la red social en sí misma son un canal de comunicación donde los usuarios comparten sus opiniones, crean contenido e interactúan con personas afines a ellos.

### ***Los usuarios de información geográfica***

La información geográfica y los sistemas de información geográfica son herramientas que son utilizadas actualmente por diferentes sectores de usuarios, por

ejemplo, sectores sociales, empresariales y gubernamentales. En tanto, esto incluye la amplia diversidad y complejidad con la que los datos geográficos repercuten en áreas como la investigación o la de toma de decisiones.

Son pocas las investigaciones a nivel nacional e internacional en las que se estudia al usuario de información geográfica. Vale la pena mencionar la investigación que fue llevada a cabo en Estados Unidos (2014) por la *Asociación de Bibliotecas de Investigación* (ARL, por sus siglas en inglés), en la que participaron 115 bibliotecas académicas pertenecientes a dicha asociación. Tal estudio tenía el objetivo de comprender el uso de los datos geográficos y las tecnologías geoespaciales. Los resultados muestran que aproximadamente un 75% de los usuarios de los servicios geográficos corresponden a estudiantes de licenciatura y posgrado, 17% a profesores, 7% a personal (investigadores), 4% al público en general, y a ex alumnos el 1% (Holstein, 2015: 42).

En México, la literatura relacionada con los usuarios de la información geográfica y los SIG es escasa; sólo unos pocos investigadores han realizado estudios o proyectos relacionados con el tema. Uno de los más interesantes es el artículo “Niveles de usuarios de los sistemas de información geográfica”, publicado por la Dra. Antonia Santos, en donde menciona que “los usuarios de los SIG van desde especialistas técnicos, quienes diseñan y mantienen el sistema, hasta aquellos que los utilizan para realizar trabajos escolares, o simplemente para explorar el sistema” (2017: 174).

En conclusión, los usuarios de información geográfica y de sistemas de información geográfica son usuarios que fundamentalmente tienen las competencias y las habilidades tecnológicas necesarias para identificar, seleccionar, evaluar y recuperar información dentro de la web con el objetivo de incorporarla en algún sistema SIG, para procesarla, interpretarla y perfilarla para la satisfacción de alguna necesidad.

Por lo anterior, los geousuarios, al tener cierto nivel educativo y dichas habilidades tecnológicas, son una comunidad que puede ser tomada en cuenta a partir del modelo de comportamiento en la búsqueda de información propuesto por Saskia Brand-Gruwel (2009).

### ***El modelo de Saskia Brand-Gruwel en el contexto de los geousuarios y los sistemas de información geográfica***

Existe una gran variedad de modelos de comportamiento en la búsqueda de información. Entre los más importantes se encuentran los propuestos por Krikelas (1983), Ingwersen (1984), el modelo holístico de Kuhlthau (1991), el modelo de Marchionini (1995), etcétera. Sin embargo, Saskia Brand-Gruwel (2009) propone un modelo que describe una sociedad tecnológica cambiante, donde los

usuarios deben contar con competencias y habilidades para poder identificar necesidades de información, localizar las fuentes correspondientes, extraerla, organizarla y sintetizarla con el objetivo de resolver problemas complejos en internet.

IPS-I es el nombre de modelo propuesto por Brand-Gruwel, el cual describe habilidades y sub habilidades necesarias para resolver un problema de búsqueda de información en internet; tal modelo consta de cinco habilidades constitutivas:

1. Definir el problema de información: Esta habilidad se aplica al principio del proceso, para tener una visión clara del problema a resolver. Al definir el problema, se formulan las preguntas principales y secundarias.
2. Buscar información: Al realizar esta habilidad, se debe seleccionar una estrategia de búsqueda, especificar los términos y analizar los datos que aparezcan como resultados.
3. Escanear información: Tras una búsqueda de información, los resultados obtenidos se escanearán (*scan information*) para definir si el tipo de información es pertinente y relevante para el usuario.
4. Procesar información: Implica el análisis y la comprensión de la información con el objetivo de integrarla a los conocimientos previos que posee el usuario.
5. Organizar y presentar información: Las primeras cuatro habilidades corresponden a la parte de análisis, mientras que la fase de organizar y presentar información forma parte del proceso de síntesis. En ese sentido, al combinar dichas fases, se obtendrá como resultado un producto según lo requerido por el usuario, por ejemplo, un documento escrito. (Brand-Gruwel, 2009: 1208-1209).

Debido a que internet es una fuente muy extensa de información, el usuario, al momento de buscar información, lleva a cabo actividades de regulación, como la orientación, el seguimiento, la dirección y la evaluación, las cuales desempeñan un papel fundamental en la ejecución de la habilidad. Tanto las actividades como las habilidades deben llevarse a cabo en un nivel adecuado para resolver con éxito los problemas de información; para ello se distinguen las siguientes tres habilidades subyacentes a partir de los datos:

1. Habilidades de lectura: El usuario debe contar con habilidades de lectura en internet para identificar, por ejemplo, hipervínculos o enlaces a otras páginas web.
2. Habilidades de evaluación: Habilidades para que el usuario pueda identificar información pertinente y relevante a partir de títulos, resúmenes, bases de datos, etcétera, con el objetivo de evitar copiar y pegar la información.

3. Habilidades informáticas (las cuales pueden caracterizarse como condicionales): Las habilidades informáticas son importantes a la hora de buscar en la red de internet. Esto implica, entre otras cosas, habilidades de tecleo y de navegación (por ejemplo, teclear un término de búsqueda y seguir los enlaces) (Brand-Gruwel, 2009: 1216).

En tal contexto los geousuarios tienen la capacidad de realizar las habilidades constitutivas y las subyacentes, ya que para tener como resultado productos geográficos, esta comunidad debe ejecutar tareas de orden tecnológico y de análisis de información.

## MÉTODO

Para llevar a cabo esta investigación se empleó un enfoque mixto, por lo que se utilizó tanto el método cualitativo como el cuantitativo. El análisis se realizó a partir de los puntos que se exponen a continuación.

### *Identificación de usuarios*

Los usuarios de sistemas e información geográfica deben contar con competencias y habilidades necesarias para realizar diversas tareas al momento de su búsqueda de información e interacción con los sistemas. Por ejemplo, capacidad para identificar su necesidad, buscar, analizar y recuperar su información para integrarla al SIG. A partir de estas acciones se realizan procesos más complejos, como procesar, organizar y dar salida a nueva información en forma de mapa o plano.

De acuerdo con lo descrito, en el presente estudio se categoriza a los geousuarios en los siguientes tres grupos: principiantes, intermedios y expertos; cada uno corresponde a su nivel de habilidad tecnológica y de interacción con la información geográfica y los SIG.

La *web* 2.0 y las plataformas sociales son herramientas eficaces para la búsqueda y el intercambio de información. Para estas comunidades son importantes las páginas de Facebook orientadas a usuarios de información geográfica. Dentro de esta plataforma, destacan las siguientes páginas: Geomática (19 000 seguidores), Mundo GIS (17 000 seguidores), SIG Herramientas (12 000 seguidores), Aprende SIG (4 900 seguidores), Portal GIS (2 600 seguidores).

Debido al impacto y la relevancia que dichas páginas tienen para los geousuarios, se considera importante identificar cuáles son las necesidades de información y cómo es el comportamiento informativo al momento de su búsqueda de información geográfica.

### ***Tamaño de la muestra***

Para determinar el tamaño de la muestra en este estudio, se utilizó el método del muestreo estratificado, el cual “es una técnica de muestreo probabilístico en donde el investigador divide a toda la población en diferentes subgrupos o estratos. Luego, selecciona aleatoriamente a los sujetos finales de los diferentes estratos en forma proporcional” (*Explorable*, 2009).

Con base en el método del muestreo estratificado, se tomaron en cuenta cinco páginas de Facebook relacionadas con información y sistemas de información geográfica. Se identificó los números totales de seguidores y se calculó una población total de 43 300.

Posteriormente, se determinó el porcentaje que representa cada una de las cinco páginas y se calculó el tamaño de la muestra con un nivel de confianza del 95%, con margen de error del 5%, dando como resultado un total de 1 812 seguidores.

Debido a que se estimó que existe alta probabilidad de que cierto porcentaje de los geousuarios se encuentre en al menos dos o tres páginas de Facebook, se seleccionó aleatoriamente a Portal GIS como parte representativa del tamaño total de la muestra (*Tabla 1*).

<b>Grupos de Facebook</b>	<b>No. de seguidores</b>	<b>%</b>	<b>Nivel de confianza</b>	<b>Margen de error</b>	<b>Tamaño de la muestra</b>
QGIS3	13,700	31.64	95%	5	375
GIS México	10,100	23.33	95%	5	371
Portal GIS	2,600	6.00	95%	5	336
Aprende GIS	4,900	11.32	95%	5	357
GIS Herramientas	12,000	27.71	95%	5	373
<b>Total</b>	<b>43,300</b>	<b>100</b>			<b>1,812</b>

*Tabla 1.* Criterios para la determinación y cálculo de la muestra

### ***Instrumento y recogida de datos***

La recogida de datos consistió en la aplicación *online* de un cuestionario, entre el 1º y el 30 de noviembre de 2021, con Google Forms. Se obtuvo un total de 315 respuestas que corresponden al 93.7% de participación.

## Estructura

- *Sección 1 - Datos generales:* tiene el objetivo de identificar cuestiones generales de la comunidad SIG tales como edad, género, nivel de conocimiento en sistemas de información geográfica, etcétera.
- *Sección 2 - Necesidades de información:* identifica cuáles son las necesidades de información de la comunidad SIG (temáticas, *software*, etcétera).
- *Sección 3 - Comportamiento en la búsqueda:* identifica los recursos, criterios de búsqueda, métodos de evaluación y las fuentes principales de información de los geousuarios.
- *Sección 4 - Geousuarios y bibliotecas:* tiene como objetivo identificar cómo es la relación de los geousuarios con las bibliotecas y otras unidades de información.

## PRESENTACIÓN Y ANÁLISIS DE LOS RESULTADOS

### *Sección 1 - Datos generales*

Los geousuarios que participaron en el estudio son en su mayoría hombres (62.5%) y en minoría mujeres (37.1%). El rango de edad oscila entre los 18 a 39 años. El nivel académico en su mayoría es de licenciatura (56.2%), de posgrado (24.40%) y de técnico (4.0%).

En la *Tabla 2* se muestra el nivel de conocimiento en SIG, donde se observa que se identifican como usuarios de nivel básico, intermedio o experto.

Nivel	Porcentaje
Básico	98 (31.1%)
Intermedio	166 (52.7%)
Experto	51 (16.2%)

*Tabla 2.* Nivel de conocimiento en SIG

Dentro del ámbito de los sistemas de información geográfica, existen dos modalidades de uso de *software*: libre y comercial. En la *Tabla 3* se muestra la preferencia por el *software* libre. Es importante destacar que uno de los *softwares* libres más utilizados es QGIS, el cual es un sistema de información geográfica de Código Abierto licenciado bajo GNU (<https://www.qgis.org/es/site/>), y seguido por ArcSIG (*software* comercial).

Software	Porcentaje
Libre	233 (74%)
Comercial	82 (26%)

Tabla 3. Preferencias de software

Por otra parte, se identifica la razón por la que los geousuarios utilizan los sistemas de información geográfica. La *Figura 1* muestra que son utilizados primordialmente para resolver problemas del ámbito laboral y de desarrollo de proyectos, entre otros.

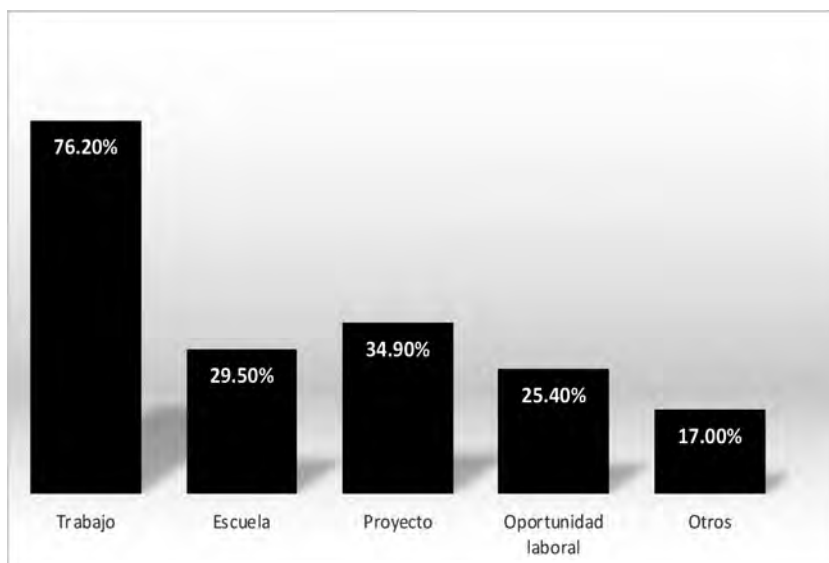


Figura 1. Uso de los SIG

## Sección 2 - Necesidades de información

La *Figura 2* muestra el porcentaje de uso de los grupos de Facebook, para resolver alguna necesidad. Se destaca que gran parte de los encuestados, sí logran su objetivo.

De acuerdo con la pregunta anterior, a los usuarios que respondieron “sí”, en la *Figura 3* se muestra el porcentaje de aquellos que utilizan las páginas de Facebook para satisfacer alguna necesidad de información. Destacan las fuentes de información geográfica, herramientas, procesos, procedimientos, manejo de *software* y capacitación en SIG.



Figura 2. Resolución de necesidades de información en grupos SIG de Facebook

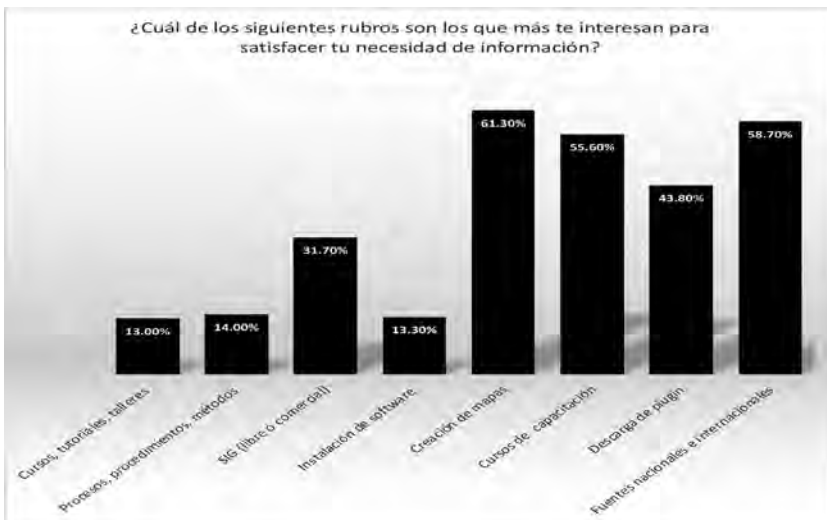
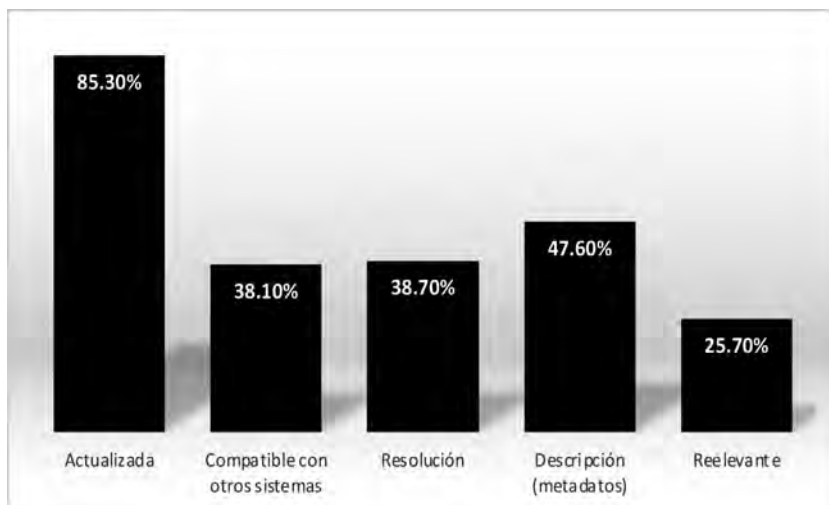


Figura 3. Necesidades de información de geousuarios en Facebook

La información geográfica que se encuentra disponible en internet se aloja en fuentes o bases de datos, en las que el usuario puede realizar diferentes tareas, tales como: buscarla, analizarla, descargarla e integrarla a sus sistemas. Por lo tanto, se identifica que los usuarios tienen necesidades específicas con base en la información que está disponible en dichas fuentes y en la *web*.



La *Figura 4*. muestra que los geousuarios necesitan información actualizada, organizada (metadatos), de calidad (resolución) y que sea compatible con sus sistemas.



*Figura 4.* Necesidades de información de geousuarios: fuentes

### *Sección 3 - Comportamiento en la búsqueda*

La *Tabla 4* muestra las fuentes donde los geousuarios, buscan y descargan su información geográfica (.SHP, Geotif, WFS, KML, CSV, etcétera).

Google (57.3%)	INEGI (72%)	CONABIO (45.5%)
ArcSIG Hub (12.4%)	Natural Earth 815%)	Open Street Map (36.3%)

*Tabla 4.* Principales fuentes de información de geousuarios

En la *Figura 5*, se muestran los criterios que los geousuarios toman en cuenta al momento de la búsqueda y recuperación de su información geográfica, resaltando la calidad y los atributos de la información.

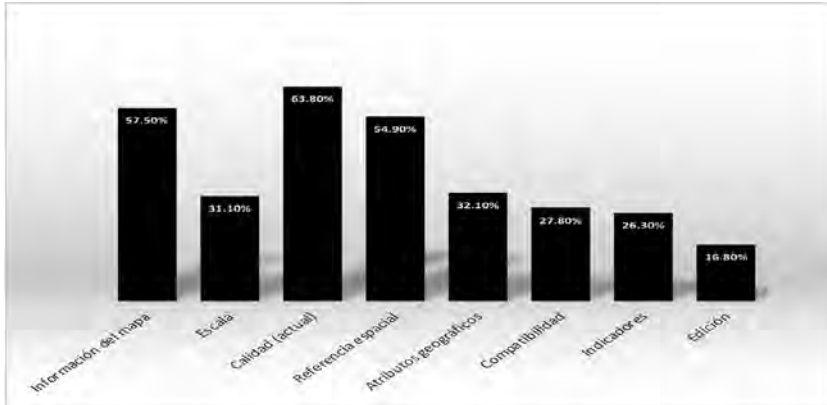


Figura 5. Criterios para la búsqueda y recuperación de información geográfica

Al momento del manejo e interacción con los sistemas de información geográfica, los geousuarios buscan características específicas para realizar dichas tareas. La *Tabla 5*, muestra las más importantes.

Características	Porcentaje
Diversidad de herramientas	82.8%
Intuitivo y fácil de usar	74.2%
Fácil de instalar	32.5%
Disponible en idioma español	28.7%

Tabla 5. Características de uso de SIG

Los geousuarios, al momento de gestionar, manejar y procesar la información dentro de los sistemas de información geográfica, prefieren presentar sus resultados de manera digital (*Tabla 6*).

Presentación de información	Porcentaje
Digital	71.3%
Compartida en alguna plataforma en la Nube	16.9%
Impresa	9.2%

Tabla 6. Preferencia de presentación de resultados en SIG

#### Sección 4 - Geousuarios y bibliotecas

Por último, se precisa como importante conocer qué servicios les gustaría recibir a los geousuarios desde las bibliotecas y otras unidades de información para resolver sus necesidades informativas (Tabla 7).

Servicios que requieren los geousuarios en bibliotecas	Porcentaje
Capacitación en SIG	32.8%
Manuales SIG	22.6%
Búsqueda y recuperación de información de archivos: SHP, Geotif, WFS, KML, CSV, etc.)	15%
Sección SIG dentro de la página web de la biblioteca	9.2%
Búsqueda y recuperación de información geográfica en geoportales	8.9%
Compartir proyectos en plataformas en la nube	6.4%
Descarga e instalación de software en computadora personal	3.2%
Todos los anteriores	1.9%

Tabla 7. Posibles servicios a ofrecer a geousuarios

## DISCUSIÓN

A partir de este estudio, se confirma que los geousuarios sí cuentan con habilidades y competencias tecnológicas, independientemente de su nivel de manejo de los sistemas de información geográfica (básico, intermedio o experto), porque estas comunidades tienen la capacidad de búsqueda y recuperación de información desde las bases de datos o fuentes disponibles que conocen en internet. Esta información, que es completamente digital, será analizada, integrada y manipulada dentro de los SIG para obtener un producto impreso, digital o en la Nube, todo ello para cumplir objetivos particulares; por lo tanto, ejecutan las habilidades y las habilidades subyacentes del modelo IPS-I, propuesto por Saskia Brand-Gruwel (2009).

Los sistemas de información geográfica pueden ser descargados a partir de dos modalidades: *software* comercial y *software* libre. El usuario SIG necesita que su sistema de información geográfica cuente con una interfaz intuitiva, fácil de usar, de instalar, en idioma español y, sobre todo, que ofrezca una adecuada

diversidad de herramientas, ya que la información geográfica presenta una infinidad de usos de acuerdo con el objetivo de cada usuario.

Se identifica que la comunidad SIG prefiere el *software* libre QGIS, porque este tipo de iniciativas resultan benéficas para las sociedades en red, destacando aspectos tales como el costo monetario, contrastante con los altos precios del *software* comercial; el derecho al acceso libre y universal a la información, y las libertades que permite el *software* libre.

Las páginas de Facebook enfocadas a comunidades de geousuarios son medios de comunicación en donde se realizan diferentes actividades, una de ellas es la resolución de alguna necesidad de obtención de información, mediante preguntas directas a la comunidad.

Independientemente de las competencias y habilidades tecnológicas de los geousuarios, se confirma que tienen ciertas limitaciones de conocimiento sobre fuentes de información nacionales e internacionales para la búsqueda y descarga de información geográfica. Por lo tanto, la falta de conocimiento de estas fuentes geográficas tiene una repercusión al momento del manejo y la integración de dichos documentos al sistema. Los geousuarios buscan conocer e interactuar con bases de datos en las que se ofrezca información geográfica actualizada, con descripción detallada (metadatos), de calidad (resolución, atributos geográficos, etcétera) y que sea compatible con otros sistemas.

Se detecta que la comunidad de geousuarios tiene la necesidad de capacitarse en cuestiones de sistemas de información geográfica, por ello, buscan tutoriales, guías de uso de *software*, etcétera. En este sentido, el conocimiento de actualizaciones o nuevas herramientas es otra de las necesidades primordiales de los geousuarios, ya que los sistemas se van renovando periódicamente.

Aunque los geousuarios pertenecientes a páginas de Facebook son comunidades que no necesariamente están relacionadas con bibliotecas y otras unidades de información, a partir de este estudio de usuarios se identifica que estas instituciones pueden tener un área de oportunidad para acercarse a estos segmentos de usuarios y resolver algunas de sus necesidades.

Respecto a bibliotecas y otras unidades de información, las temáticas que más les interesarían a los usuarios son: fuentes de información (bases de datos), uso de *software* (herramientas, creación de mapas, procesos), manejo, instalación de *software* y capacitación o tutoriales.

A partir de la biblioteca física o digital, se pueden diseñar y ofrecer productos y servicios geográficos para ayudar a estas comunidades a resolver sus necesidades. Los geousuarios de páginas de Facebook, en su mayoría, no consultan ni acuden a ninguna biblioteca, pero, de existir servicios especializados en SIG, estarían dispuestos a asistir a dichas instituciones.

## CONCLUSIONES Y RECOMENDACIONES

Con base en los resultados de este estudio sobre usuarios, se concluye que los geousuarios tanto de información geográfica como de SIG son comunidades complejas, debido a la diversidad de herramientas tecnológicas que utilizan, los objetivos particulares que tienen y las metodologías que emplean en su búsqueda e interpretación de información geográfica y en su interacción con los SIG.

Por lo tanto, es necesario que estos usuarios fortalezcan sus estrategias de búsqueda, recuperación e interpretación de la información; para lograrlo, es preciso que tengan conocimientos sobre estándares de metadatos, para que sus resultados sean pertinentes, relevantes, significativos y susceptibles de ser utilizados.

Tal como se describió en esta investigación, la calidad y la información detallada de la IG y de los mapas es una necesidad primordial para esta comunidad; por ende, sería fundamental que las instituciones creadoras de información geográfica, como, por ejemplo, el INEGI o el Conabio, homologuen y organicen la información geográfica de acuerdo a estándares de metadatos internacionales, todo ello con el objetivo de una mejor experiencia al momento de recuperar su información geográfica.

Adicionalmente se sugiere que las instituciones creadoras de datos geográficos desarrollen tácticas para la difusión de bases de datos o geoportales a nivel nacional e internacional, con el objetivo de ampliar su panorama de fuentes geográficas.

Es preciso recomendar a las bibliotecas y unidades de información que tomen a esta comunidad como referente para ofrecer servicios bibliotecarios de manera remota o presencial de acuerdo con las necesidades y requerimientos específicos de geoinformación. Como ejemplo, la creación de programas de formación de geousuarios, tanto sobre uso de fuentes de información geográfica como de uso y manejo de SIG; en tal contexto se abre la posibilidad de resolver varias necesidades informativas.

Para lograr este objetivo, la presencia del profesional de la información especializado en información geográfica, podría ser de suma importancia, ya que un profesional geobibliotecario que tenga las competencias, las habilidades y los conocimientos geográfico-tecnológicos, será capaz de orientar e instruir a los usuarios desde la unidad de información.

A partir de este trabajo se abre la posibilidad a nuevos estudios de geousuarios, porque la tecnología y la información geográfica es dinámica. Esto en tanto que el usuario de la información geográfica, debido a sus características, es susceptible de seguir siendo investigado.

## REFERENCIAS

- Arcarás, María. 2018. “¿Qué son y para qué sirven los grupos de Facebook?”. *Baetica*, abril 11, 2018.  
<https://baetica.com/que-son-grupos-de-facebook/>
- Asociación de Internet. 2021. “17º. Estudio sobre los Hábitos de los Usuarios de Internet en México 2021”.  
<https://www.asociaciondeinternet.mx/estudios/asociacion>
- Berlanga, Lucia. 2021. “Qué es Facebook, cómo funciona y qué te puede aportar esta red social”. Julio 30, 2022.  
<https://www.ciudadano2cero.com/que-es-facebook/>
- Brand-Gruwel, S., I. Wopereis, y A. Walraven. 2009. “A descriptive model of information problem solving while using internet”. *Computers & Education* 53.  
<https://www.sciencedirect.com/science/article/abs/pii/S0360131509001511>
- Brand-Gruwel, S., I. Wopereis, y Y. Vermetten. 2005. “Information problem solving by experts and novices: Analysis of a complex cognitive skill”. *Computers in Human*. 21 (3): 487-508.  
<https://research.ou.nl/en/publications/information-problem-solving-by-expert-and-novices-analysis-of-a->
- Explorable.com. 2009. “Método de muestreo estratificado”.  
<https://explorable.com/es/muestreo-estratificado>
- Holstein, A. 2015. “Geographic Information and Technologies in Academic Libraries: An ARL Survey of Services and Support”. *Information Technology and Libraries* 34 (1): 38-51.  
<https://ejournals.bc.edu/index.php/ital/article/view/5699>
- Hootsuite and We Are Social. 2021. “Digital 2021. Global Report (Q3 Update)”.  
<https://www.hootsuite.com/es/pages/digital-trends-2021>
- Hootsuite & We Are Social. 2021. “Digital 2021 México”  
<https://datareportal.com/reports/digital-2021-mexico>
- Santos Rosas, Antonia. 2017. “Niveles de usuarios de los sistemas de información geográfica”. En *Usuarios de la Información y Web 2.0*, editado por Juan José Calva, 169-180.  
<https://repositorio.unam.mx/contenidos/5002529>
- Tomeo, Fernando. 2014. *Redes sociales y tecnologías 2.0*, 2ª ed. Buenos Aires: Editorial Astrea.
- Waybel, Eduardo. 2022. “Los sistemas de información geográfica desde la perspectiva de los bienes comunes de información: posibles aportes de la geobiblioteca y del geobibliotecario”. Tesis de maestría. Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras.  
[https://tesiunam.dgb.unam.mx/F/AGK5DPYPQSQFLFN9P2IU2JQJIXVGH9QRQCM4XCY1YEQ3Q8CS3-16644?func=full-set-set&set\\_number=015583&set\\_enty=000001&format=999](https://tesiunam.dgb.unam.mx/F/AGK5DPYPQSQFLFN9P2IU2JQJIXVGH9QRQCM4XCY1YEQ3Q8CS3-16644?func=full-set-set&set_number=015583&set_enty=000001&format=999)

*Para citar este texto:*

Waybel Sánchez, Eduardo. 2023. "Las necesidades de información y comportamiento informativo de los usuarios de sistemas de información geográfica en comunidades de Facebook". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 87-108.

<http://dx.doi.org/10.22201/iibi.24488321xc.2023.96.58770>

## **Anexo 1**

### **Cuestionario**

Estimado(a) usuario(a) el objetivo de este cuestionario es identificar las necesidades y el comportamiento en la búsqueda de información por los geousuarios en grupos de Facebook. Le solicito, de la manera más atenta, indique las respuestas que usted considere que se acercan más a sus intereses. ¡Gracias!

#### **Sección 1 - Datos generales**

##### **Género**

Masculino

Femenino

Sin especificar

##### **Rango de edad**

18 – 29

30 – 39

40 – 49

50 + 60

61 +

##### **Nivel académico**

Licenciatura

Maestría

Posgrado

Técnico

Otro

**De acuerdo a tu nivel de conocimiento en información y sistemas de información geográfica, ¿cómo te consideras?**

Usuario novato

Usuario intermedio

Usuario experto

**¿Cuál de las siguientes modalidades de SIG prefieres?**

*Software* libre

*Software* comercial

Anota por favor el nombre de tu *software* preferido

-----

**¿Cuál es la razón por la que utilizas la información y los sistemas de información geográfica?**

Trabajo

Escuela

Investigación

Especialización

Proyecto

Oportunidad laboral

Otro (específica)

**Sección 2 - Necesidades de información**

**¿Consideras que los grupos de Facebook (SIG) te ayudan a resolver alguna necesidad de información?**

Sí

No

1. 1. ¿Consideras que los grupos de Facebook (SIG) te ayudan a resolver alguna necesidad de información? \_\_\_\_\_

2. Como usuario SIG (básico, intermedio o avanzado). ¿Cuál de los siguientes rubros, son los que más te interesan para satisfacer alguna necesidad?

Tipos de SIG (libre o comercial)

Instalación de *software*

Creación de mapas

Cursos de capacitación

Descarga y uso de plugins

Fuentes nacionales para descarga de archivos (tanto en fuentes nacionales como internacionales) SHP, Geotif, WFS, KML, CSV, etc.)

Fuentes internacionales para descarga de archivos

SHP, Geotif, WFS, KML, CSV, etc.)

Otros (específica)\_\_\_\_\_



**3. De la siguiente lista, indica ¿cuál o cuáles son los temas que más te interesarían conocer o desarrollar dentro de los sistemas de información geográfica para satisfacer alguna necesidad?**

Añadir capas  
Modelos vectoriales (Puntos, Líneas, Polígonos).  
Modelos ráster  
Simbología  
Etiquetado  
Importación y exportación de datos geográficos.  
Sistema de Coordenadas Geográficas  
Análisis espacial: Herramientas de selección y extracción de datos  
Publicación información geográfica en la web  
Capas virtuales  
Generación de atlas  
Georreferenciación  
Herramientas de geoposicionamiento  
Impresión y salida de imagen  
Otros (especifica) \_\_\_\_\_

**4. Cuando tienes una necesidad específica de información, ¿a dónde recurres para satisfacer dicha pregunta, duda, proceso, etc.?**

Busco en páginas de Facebook  
Busco en internet  
Pregunto a colegas  
Busco información en tutoriales  
Busco información en bibliotecas digitales o bases de datos  
Acudo a algún especialista  
Otro (especifica) \_\_\_\_\_

**5. ¿Qué es lo que más te interesa de la información geográfica disponible en internet?**

Que este actualizada  
Resolución  
Compatibilidad con otros sistemas  
Descripción detallada del archivo (metadatos)  
Relevante  
Otra (especifica) \_\_\_\_\_

### Sección 3 - Comportamiento en la búsqueda

**6. ¿Dónde buscas y descargas tu información geográfica?  
(descargas de archivos, .SHP, Geotif, WFS, KML, CSV, etc.)**

Google

INEGI

CONABIO

ArcSIG Hub

Natural Earth

OpenStreetMap

Otro (especifica) \_\_\_\_\_

**7- ¿Qué criterios consideras para la selección y descarga de tu información geográfica?**

Información básica del mapa o archivo

Escala

Calidad de la información

Atributos geográficos

Sistema de referencia espacial

Compatibilidad con otros sistemas

Indicadores

Edición

Otros (especifica) \_\_\_\_\_

**8. ¿Con base a sus características, que aspectos consideras relevantes al momento del manejo e interacción con tu sistema de información geográfica?**

Intuitivo y fácil de usar

Fácil de instalar

Diversidad de herramientas

Disponible en mi idioma

Otro (especifica) \_\_\_\_\_

**9. ¿Cómo presentas los resultados obtenidos a partir de la gestión y manejo de tu información con el sistema de información geográfica?**

Impresa

La comparto en alguna plataforma (Nube)

La conservo y la comparto en archivo digital

Otro (especifica)

#### Sección 4 - Geousuarios y bibliotecas

**10. ¿Acudes a alguna biblioteca para consultar bibliografía referente a información y sistema de información geográfica?**

Sí

No

En caso de responder “sí”, anota por favor el o los temas que más consultas \_\_\_\_\_

**11. Si la biblioteca de tu universidad o comunidad ofreciera un servicio SIG, por ejemplo donde un bibliotecario especializado en SIG te ayude y oriente para descargar información geográfica, uso de bases de datos, uso del *software*, etc., ¿asistirías a la biblioteca para utilizar dicho servicio?**

Sí asistiría

No asistiría

**12. Con base en la pregunta anterior ¿Cuáles son los temas o requerimientos que solicitarías como usuario SIG en dicha biblioteca? \_**  
\_\_\_\_\_

**13. De existir un servicio en SIG en la biblioteca de tu universidad o comunidad, ¿qué servicios te interesaría proponer o que ofreciera dicha biblioteca?**

Capacitación en SIG

Búsqueda y recuperación de información geográfica en geoportales

Búsqueda y recuperación de información de archivos: SHP, Geotif, WFS, KML, CSV, etc.)

Descarga e instalación de software en tu computadora personal

Compartir tus proyectos en plataformas en la Nube

Manuales SIG

Sección SIG dentro de la página web de la biblioteca

Oros (especifica) \_\_\_\_\_

**¡Gracias por tu colaboración!**

# La ciencia abierta y su relación con la innovación: una revisión bibliométrica

Felipe Guevara-Pezoa\*

*Artículo recibido:*

*15 de marzo de 2023*

*Artículo aceptado:*

*20 de junio de 2023*

*Artículo de investigación*

## RESUMEN

Trabajos recientes sugieren que la ciencia abierta puede mejorar diferentes disciplinas científicas, así como la difusión del conocimiento y la eficiencia del sistema de innovación. Sin embargo, las prácticas de ciencia abierta aún no se han adoptado de manera plena. El objetivo de este estudio fue explorar la relación entre la ciencia abierta y la innovación a través de una revisión bibliométrica sobre esta temática. Se observó un crecimiento de un 28,5 % anual y un predominio de países europeos en la productividad en la materia entre el año 2013 y 2022. Además, tras la evaluación de la productividad se advierte que el 91,8 % de los autores (1 401) participaron solo en una publicación. En suma, las temáticas con más desarrollo y relevancia las

\* Dirección de Investigación y Postgrado, Vicerrectoría Académica, Universidad Central de Chile, Chile  
felipe.guevara@ucentral.cl

representaron la interoperabilidad, el código y el hardware abiertos, el intercambio de datos, la propiedad intelectual y la política científica. También el trabajo muestra un creciente, pero aún incipiente interés en la relación entre ciencia abierta e innovación, con un foco en los recientes años hacia el estudio de temas relacionados con la educación (superior, innovación educativa), la investigación responsable, la innovación social, la co-creación y la inteligencia artificial.

**Palabras clave:** Ciencia abierta; Innovación; Análisis bibliométrico; Datos abiertos; Propiedad intelectual.

## **Open Science and Its Relationship to Innovation: A Bibliometric Review**

*Felipe Guevara-Peña*

### **ABSTRACT**

Recent work suggests that open science can improve different scientific disciplines, as well as the diffusion of knowledge and the efficiency of the innovation system. However, open science practices have not yet been fully adopted. The aim of this study was to explore the relationship between open science and innovation through a bibliometric review on this topic. A growth of 28,5 % per year and a predominance of European countries in productivity on the subject was observed between 2013 and 2022. Furthermore, after the evaluation of productivity, it was observed that 91,8 % of the authors (1 401) have participated in only one publication. Finally, the topics with the most development and relevance are interoperability, open source and open hardware, data sharing, intellectual property, and science policy. The present study shows a growing, but still incipient interest in the relationship between open science and innovation, with a focus in recent years on the study of topics related to education (higher education, educational innovation), responsible research, social innovation, co-creation and artificial intelligence.

**Keywords:** Open science; Innovation; Bibliometric analysis; Open data; Intellectual property.

## INTRODUCCIÓN

La ciencia abierta es un enfoque poderoso para comprender el papel de la comunicación y la difusión del conocimiento, sin embargo, la gran cantidad de literatura disponible muestra que el concepto de *ciencia abierta* es polisémico y, con frecuencia, carece de precisión (Invernizzi 2020). En este sentido, en una definición integrada, la ciencia abierta se define como el conocimiento transparente y accesible que se comparte y desarrolla a través de redes de colaboración (Vicente-Saez y Martínez-Fuentes 2018).

En la actualidad, las prácticas de ciencia abierta se están convirtiendo en un elemento clave para promover el intercambio de conocimientos mediante la difusión dinámica de los resultados de la investigación de una manera eficaz, temprana y lo más ampliamente accesible que se pueda (Vicente-Saez y Martínez-Fuentes 2018; Watson 2015). Además, se han identificado como herramienta poderosa para promover la colaboración, la transparencia y el acceso libre al conocimiento.

Dichas prácticas han demostrado que proporcionan importantes beneficios para los investigadores en comparación con las prácticas cerradas más tradicionales, tales como el aumento de la velocidad y la eficacia en el avance del conocimiento y la innovación (McKiernan *et al.* 2016). Sin embargo, a pesar de su creciente popularidad, la adopción de prácticas de ciencia abierta aún no se ha implementado de manera plena en la comunidad científica (Markowitz, Song y Taylor 2021; Song, Markowitz y Taylor 2021). Esto sugiere que aún existe un gran potencial para incrementar la adopción de este tipo de prácticas en la investigación científica y maximizar sus beneficios para la comunidad científica y la sociedad en general.

Los resultados sobre el efecto de la ciencia abierta en la calidad de la investigación son dispares. Por un lado, el uso de métricas cuantitativas en la evaluación de la investigación ejerce presión sobre el investigador, de modo que dedicar tiempo adicional a la publicación transparente de datos y resultados resulta difícil (Heuritsch 2020). Sin embargo, otros trabajos señalan que, en el largo plazo, las prácticas de ciencia abierta compensan los costes tanto para los investigadores individuales como para los del ámbito colectivo (LeBel, Campbell y Loving 2017).

Por otro, también hay quienes mencionan que las prácticas de publicación académica de acceso abierto basadas en APC disminuyen la calidad de los artículos de investigación (van Vlokhoven 2019). No obstante, por igual, existen posturas que declaran que la ciencia abierta puede mejorar diferentes disciplinas científicas, prácticas de investigación y la ciencia en general (Stracke 2020). Además, que contribuyen con aumentos de citas, atención mediática, oportunidades de empleo, de financiación y colaboradores potenciales (McKiernan *et al.* 2016).

El aumento de colaboradores potenciales tiene gran importancia en algunas etapas del proceso de investigación, como el desarrollo de la innovación, entendido este concepto como nuevas combinaciones de recientes o existentes conocimientos, recursos, equipos y otros factores (Schumpeter y Backhaus 2003). En un estudio de Woelfle, Olliaro y Todd (2011) se analiza cómo las prácticas de ciencia abierta actúan como aceleradores para la generación de asociaciones entre el mundo académico y la industria que conducen a desarrollos innovadores. Asimismo, otro estudio, de Frankenhuis y Nettle (2018), sostiene que las prácticas de ciencia abierta son liberadoras y fomentan la creatividad, sugiriendo que han hecho que el proceso de invención sea más creativo.

De acuerdo con este enfoque, los investigadores han determinado que las asociaciones público-privadas basadas en publicaciones de acceso abierto, datos y materiales abiertos y evitación de formas restrictivas de propiedad intelectual, fomentan la eficiencia del sistema de innovación (Gold 2021; Jong y Slavova 2014). Sumado a esto, se observa que las prácticas de colaboración conforman elementos sustanciales para la innovación y la ciencia abierta (Ramírez-Montoya y García-Peñalvo 2018). Esto resulta contrario a lo propuesto de manera tradicional, en cuanto a considerar los derechos de propiedad intelectual como un factor determinante de la innovación (Loukil 2020; Brüggemann *et al.* 2015).

Al respecto, un estudio exhaustivo de Murray y Stern (2007) descubrió que las restricciones a la apertura científica (como las creadas por la propiedad intelectual) limitan la diversidad y la experimentación de la propia investigación básica, lo que sugiere que las prácticas de ciencia abierta han hecho que el proceso de invención sea más diverso. Además, un documento de Fernández-Pinto (2020) sostiene que el movimiento de la ciencia abierta se ha centrado principalmente en la investigación financiada con fondos públicos, pero ha fomentado de manera activa los vínculos con el sector privado y también ha creado nuevas estrategias para comercializar la ciencia. De forma sorprendente, también se han encontrado pruebas de que los científicos que reciben financiación de la industria son más propensos a denegar a otros el acceso a los datos de investigación (Czarnitzki, Grimpe y Pellens 2014).

Se observa que existe una discrepancia entre el concepto de ciencia abierta y la realidad académica, con dos barreras principales: las individuales y las sistémicas (Scheliga y Friesike 2014). Con todo ello, algunos autores señalan que al analizar de modo serio las deficiencias de la ciencia moderna que el movimiento de ciencia abierta pretende remediar, las perspectivas de mejora resultan poco impresionantes (Mirowski 2018). Más aún, la mayoría de las pruebas de impacto de la ciencia abierta en las prácticas de investigación se basan en entrevistas, encuestas e inferencias (Fell 2019).

En un sentido estricto del aporte de las prácticas de ciencia abierta a la innovación se observan discrepancias. Mientras algunos estudios afirman que los ejercicios de innovación abierta tienen un fuerte impacto en el rendimiento de la innovación (Ebersberger *et al.* 2012; Inauen y Schenker-Wicki 2011), otros autores señalan que la relación entre prácticas abiertas e innovación depende del mercado y que, en las economías emergentes, puede tener incluso efectos negativos (Torres de Oliveira *et al.* 2020).

Teniendo en cuenta estos hallazgos contradictorios, se evidencia que existe una necesidad urgente de investigar más a fondo las contribuciones de la ciencia abierta al progreso científico y la innovación. Se requiere una revisión crítica y sistemática de la literatura existente para identificar de forma clara las fortalezas y debilidades de las prácticas de ciencia abierta y para comprender mejor cómo utilizarlas para maximizar el impacto en el rendimiento de la innovación.

En un esfuerzo por investigar la relación entre ciencia abierta e innovación, se llevó a cabo un análisis bibliométrico de las publicaciones indexadas en la base de datos Scopus entre los años 2013 y 2022. El enfoque metodológico adoptado implicó la recolección y el análisis estadístico de los datos obtenidos a partir de las publicaciones seleccionadas.

Los resultados indicaron que la investigación en esta área es todavía incipiente, con un número limitado de publicaciones y una escasa diversidad en cuanto a los temas abordados. Además, se detectó un bajo índice de colaboración entre investigadores, lo que apunta a que existe un gran potencial para el desarrollo de investigaciones colaborativas en este campo.

## METODOLOGÍA DEL ESTUDIO

El análisis presentado en este estudio fue realizado a través del uso del programa computacional Bibliometrix (Aria y Cuccurullo 2017).

El paquete de Bibliometrix está escrito en el lenguaje de programación R, el cual se caracteriza por un entorno de gran capacidad para computación estadística. R ha ganado popularidad a nivel mundial entre investigadores y analistas de datos de diversas áreas debido a su facilidad de uso, su extensa cantidad de paquetes disponibles y su capacidad para manejar grandes cantidades de datos (Aria y Cuccurullo 2017; Song *et al.* 2019). El uso de R en el análisis bibliométrico permite una alta flexibilidad y precisión en el estudio de datos.

El proceso de recopilación de datos se realizó a través de la utilización de la base de datos Scopus como fuente de información bibliográfica con la intención de conformar el conjunto final de artículos relevantes. La construcción de la cadena de búsqueda se efectuó mediante la concatenación de términos específicos



y el uso de operadores lógicos «OR» y «AND» para establecer las relaciones entre ellos y de esta manera asegurar la obtención de resultados precisos y relevantes.

El primer cuadro de búsqueda incluye las frases «open science», mientras que el segundo campo las palabras clave «innovation» y «intellectual property» combinadas con el operador «AND». De esta manera se logra la siguiente cadena de búsqueda:

(TITLE-ABS-KEY («open science») AND (TITLE-ABS-KEY («intellectual property») OR TITLE-ABS-KEY («innovation»)))).

La incorporación de la palabra clave «intellectual property» se justifica por la comprensión del concepto de propiedad intelectual como un factor determinante de la innovación, tal y como ha sido evidenciado por la literatura. Por lo tanto, cualquier artículo que incluya la noción de propiedad intelectual estará, por defecto, hablando de innovación.

La cadena de búsqueda arrojó un total de 397 artículos relevantes. Sin embargo, para garantizar la calidad y precisión de los datos, se aplicaron criterios de exclusión, limitando el rango de años entre 2013 y 2022 (siendo 2013 el primero donde se encontraron artículos relacionados con la temática), y restringiendo el idioma de los artículos a inglés y español. Al aplicar estos criterios, se obtuvo un listado final de 315 artículos.<sup>1</sup>

Posteriormente, estos datos fueron exportados y cargados en el paquete R del software Bibliometrix, a fin de convertirlos en un marco de datos bibliográficos para calcular las diferentes medidas presentadas en este artículo.

El marco de datos analizado contiene los metadatos necesarios para el número total de documentos, el de revistas, el de autores, el de citas y el de palabras claves utilizadas. Con estos datos fue posible calcular a su vez el número de artículos por autor, así como la tasa de crecimiento anual de la productividad analizada.

Junto con lo anterior, se evaluó la productividad académica en un marco geográfico, analizando la productividad académica en diferentes países. Además, se identificaron los autores más productivos, midiendo el número de publicaciones y citas recibidas. De igual forma, se estudiaron las fuentes más relevantes para el campo de estudio contenidas en cada revista mediante el número de publicaciones sobre el tema. También, se analizaron las palabras clave más recurrentes en las publicaciones considerando el primer y último año en que aparece cada palabra, así como el año con mayores ocurrencias. Esto permitió graficar una línea de tiempo de cada palabra, indicando el año con el mayor uso de cada una.

1 Los datos obtenidos a través de este proceso están disponibles en el repositorio en línea <https://figshare.com/s/d257acc1fddfffd18b4e>

Finalmente, se evaluó el grado de desarrollo y relevancia de las temáticas abordadas en las publicaciones mediante un mapa temático de co-ocurrencia de palabras clave (Aria y Cuccurullo 2017). En específico, el análisis temático tomó agrupaciones de palabras clave de los autores y sus interconexiones para organizar temas. Estos temas se caracterizaron por propiedades como la densidad y la centralidad, donde esta última representó el grado de correlación existente entre diferentes temas, mientras que la densidad conformó una medida de la cohesión entre los nodos. Los grados de ambas fueron cálculos proporcionados directamente por el software Bibliometrix, y determinaron si ciertos temas estuvieron bien desarrollados o no, y si son importantes o no.

Al graficar densidad y centralidad en un mapa bidimensional se posibilitó clasificar los temas en cuadrantes: (1) cuadrante superior derecho (alta centralidad y densidad): temas motores; (2) cuadrante inferior derecho: temas básicos (alta centralidad, baja densidad); (3) cuadrante inferior izquierdo: temas emergentes o en desaparición (baja densidad y centralidad); (4) cuadrante superior izquierdo: temas altamente especializados o de nicho (baja centralidad, alta densidad).

Esta metodología de estudio proporciona una visión general de la evolución de la productividad académica, además permite conocer las tendencias y patrones en el campo de estudio.

## RESULTADOS

La búsqueda bibliográfica se llevó a cabo en la base de datos Scopus con el objetivo de recolectar artículos relacionados con ciencia abierta e innovación. El rango temporal de búsqueda fue de 2013 a 2022. La selección arrojó como resultado 397 artículos, sin embargo, tras aplicar los criterios de exclusión establecidos en los materiales y métodos, solo 315 artículos fueron seleccionados para los análisis cuantitativos.

Los artículos elegidos presentaron una media de 11,72 citas por artículo, indicando un alto impacto en el campo académico. Los autores de estas publicaciones fueron 1 525, quienes escribieron una media de 0,21 artículos, cada uno en 231 fuentes diferentes.

Los datos presentados en la *tabla 1* muestran la distribución de los artículos seleccionados en función de las citas, los autores y las fuentes de publicación.

Fuentes (Journals, libros, etc.)	231
Documentos	315
Tasa de crecimiento anual %	28,95

Media de citas por documentos	11,72
Media de artículos por autor	0,21
Número de palabras claves	1 133
Número de autores	1 525

Tabla 1. Principales resultados del análisis bibliométrico para datos entre 2013 y 2022  
Fuente: Elaboración propia.

Al realizar un análisis de la evolución de la productividad académica en el área de estudio (relación entre ciencia abierta e innovación) se observó una tasa de crecimiento anual del 28,95 %. Ello refleja un aumento en el número de publicaciones sobre el tema, con un pico en el año 2017, en el que se publicaron 41 artículos. Sin embargo, también se analizó el número de citas de estas publicaciones. El estudio mostró que la media de citas totales alcanzó su máximo en 2014 con una media de 7,66 citas por año (MTCPY). Por su parte, el valor medio más bajo de 1,28 MTCPY se registró en 2022 (véase *figura 1*).

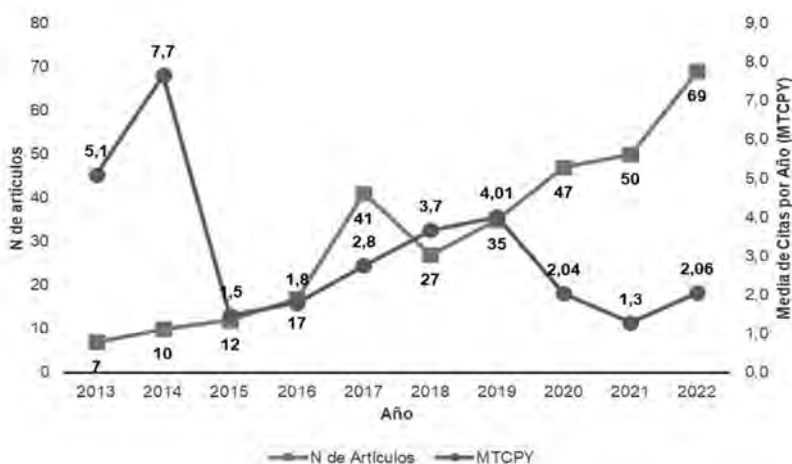


Figura 1. Evolución de la productividad académica.  
Fuente: Elaboración propia.

El análisis de la producción científica de diferentes países permitió determinar el número de artículos con autores correspondientes (*corresponding author*) de cada uno de ellos. Los resultados mostraron que el número de artículos oscilaba entre 1 y 21, siendo Estados Unidos con el mayor número de artículos publicados en el tema de estudio, con un total de 270 artículos. Le siguen el Reino Unido con 127 artículos y Canadá con 87 artículos (véase *tabla 2*).

En cuanto al número de citas, Estados Unidos también obtuvo el mayor número, con un total de 754 citas. Le siguen Alemania con 224 citas y el Reino Unido con 136. Por su parte, Dinamarca, a pesar de conformar el séptimo país en productividad (41 artículos), obtuvo solo dos citas (*tabla 2*).

<b>País</b>	<b>N artículos</b>	<b>Citas totales</b>	<b>N artículos autor correspondiente (<i>corresponding author</i>)</b>
Estados Unidos	270	754	21
Reino Unido	127	136	13
Canadá	87	61	10
España	72	32	5
Italia	66	116	9
Alemania	62	224	8
Dinamarca	41	2	2
Países Bajos	41	11	4
Francia	37	103	6
Suiza	35	99	7

*Tabla 2. Top 10 países más productivos*

Fuente: Elaboración propia.

La evaluación de la productividad institucional demostró que la Facultad de Medicina de Harvard (Estados Unidos) y la Universidad de Toronto (Canadá) ocuparon los primeros lugares en cuanto a producción científica relacionada con el tema de estudio. La Facultad de Medicina de Harvard publicó 14 artículos (4,5 % del total), mientras que la Universidad de Toronto 13 artículos (4,1 % del total). Le siguen la Ludwig Boltzmann Gesellschaft de Austria y la Université Laval de Canadá con 12 artículos publicados cada una (3,8 % del total).<sup>2</sup>

El análisis de las fuentes más productivas permitió determinar que «The ACM International Conference Proceeding Series (ICPS)» ocupó el primer lugar con nueve artículos publicados entre 2013 y 2022, lo que representó el 2,9 % del total. Le siguen «CEUR Workshop Proceedings», «F1000Research» y «Research Policy» con seis artículos publicados cada una.<sup>3</sup>

2 La tabla que ilustra esta información está disponible en <https://figshare.com/s/e0c29bacd-da46a4636b9>

3 Esta información se presenta en <https://figshare.com/s/d5bc6328aca09746b26d>

Además, se evaluó el ámbito de las revistas más relevantes, donde se encontró que el 55 % (n=11) de las mismas pertenecen al campo de la informática y/o las tecnologías de la información. Aunado a ello, siete de las fuentes más significativas forman parte de series de libros o congresos.<sup>4</sup>

De los 20 artículos más citados en el estudio, se encontró que 11 constituyen artículos originales, ocho artículos de revisión y uno corresponde a un editorial. El artículo más citado fue el de Mueller y Piper (2014), con 510 citas totales, seguido del de Himanen *et al.* (2019) con 234 citas totales, y el de Castellanos *et al.* (2013) con 214 citas totales.<sup>5</sup>

En cuanto a los autores más productivos, se encontró que Ramírez-Montoya representa el autor más productivo con cinco artículos (1,6 %), seguido de Gold E.R y Wang P., con cuatro documentos escritos cada uno (1,3 %). Curiosamente, la mayoría de los autores incluidos en la revisión publicaron solo un artículo, es decir, 1 401 autores que corresponden al 91,8 % del número total de ellos. Le siguen 11 autores que han escrito dos artículos, 10 con tres artículos y dos con cuatro artículos.<sup>6</sup>

El análisis de las palabras clave elegidas por los autores reveló que, una vez excluidos los términos incluidos en la definición de búsqueda de la base de datos (ciencia abierta, innovación, propiedad intelectual), las tres palabras clave más frecuentes fueron *innovación abierta*, *acceso y datos abiertos*, que aparecieron 34, 30 y 17 veces, respectivamente. Por su parte, al analizar la evolución de las palabras clave, se observó que los temas relacionados de manera directa con la ciencia abierta (datos abiertos, innovación abierta) abordados principalmente entre 2019 y 2020, han dado paso a temas relacionados con educación (educación superior, innovación educativa), investigación e innovación responsables, innovación social, co-creación e inteligencia artificial (véase *figura 2*).

Del análisis, se observa que los temas motores, es decir, aquellos con gran desarrollo y relevancia, están relacionados con la interoperabilidad, el código abierto y el hardware abierto, el intercambio de datos, la propiedad intelectual y la política científica.

El segundo cuadrante se refiere a los temas básicos. Se trata de aquellos poco desarrollados en la investigación, pero muy relevantes. Los grupos que caen en este cuadrante se caracterizan por bajos niveles de densidad y altos niveles de centralidad. En este cuadrante se encuentran temas relacionados con la comunicación académica y la publicación, la revisión abierta por pares, *big data*, la inteligencia artificial y el aprendizaje automático (*machine learning*) y, por último, la innovación abierta, el acceso abierto y la ciencia abierta.

4 Los datos detallados están disponibles en *Ibidem*.

5 *Ibid.*

6 *Ibid.*

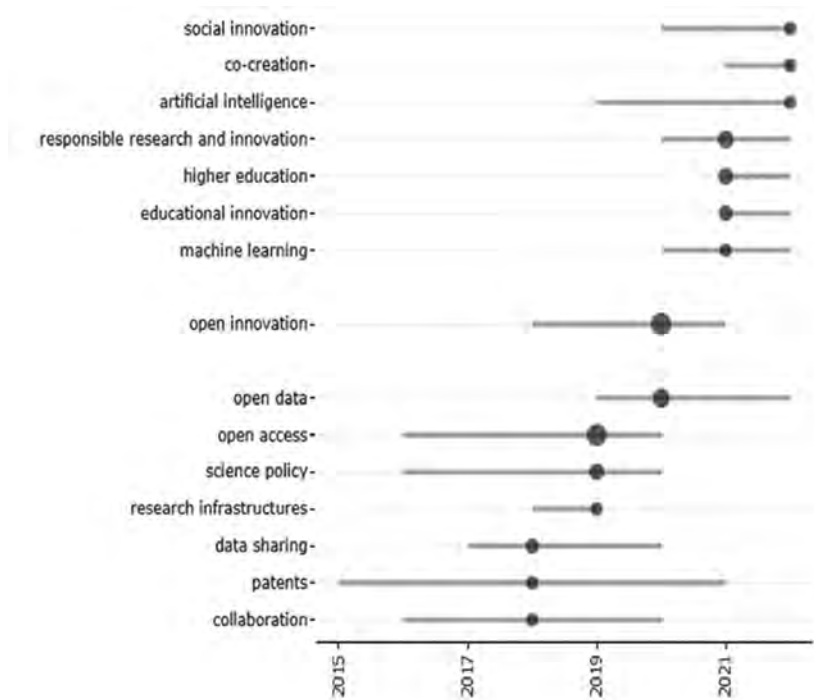


Figura 2. Evolución de las temáticas de tendencia de acuerdo con palabras claves.

Fuente: Elaboración propia.

El tercer cuadrante muestra temas emergentes o en declive. Aquí se hallan aquellos relacionados con la educación (educación superior, innovación educativa y pensamiento complejo), junto con el principio Fair (acrónimo usado en ciencia abierta para datos que cumplen con los principios de disponibilidad, accesibilidad, interoperabilidad y reutilización). Los grupos del cuadrante están formados por temas menos importantes y poco desarrollados en la investigación y esto ocurre en dos momentos de la vida de un tema, es decir, cuando emerge o declina.

En los límites divisorios de entre los temas emergentes o en declive y los básicos se encuentran las infraestructuras de investigación y las bibliotecas académicas, junto con Europa.

El cuarto cuadrante incluye los temas nicho, esto es, los muy desarrollados pero poco relevantes para la investigación. Aquí, como resultado del análisis, se observaron temas más específicos como repositorios e infraestructuras electrónicas, la nube europea de ciencia abierta, citas, *creative commons*, diseño de flujos, entornos colaborativos y diseño de chips (véase figura 3).

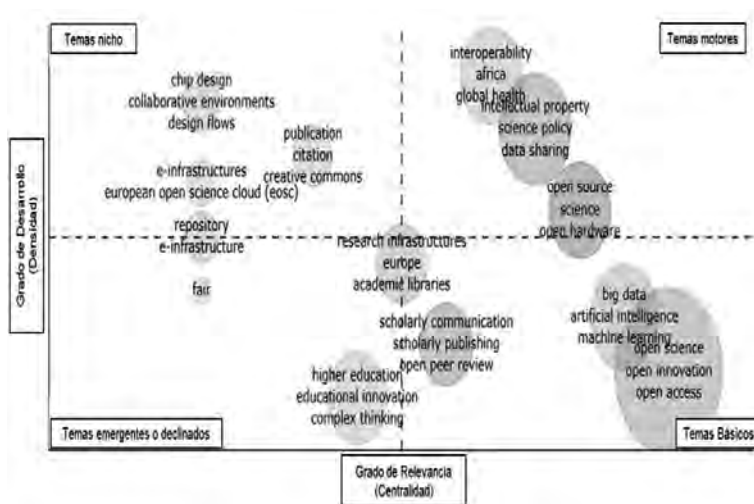


Figura 3. Agrupación de temáticas según su nivel de desarrollo y relevancia.  
Fuente: Elaboración propia.

## DISCUSIÓN

En este estudio se emplearon indicadores bibliométricos para describir la actividad científica global en relación con la ciencia abierta y la innovación. El objetivo principal fue brindar información útil para ayudar a los investigadores a comprender mejor la historia y el posible desarrollo futuro de la investigación sobre el tema. Según hasta donde se sabe, este es el primer estudio que busca evaluar cuantitativamente la tendencia evolutiva en la relación entre innovación y ciencia abierta, desde una perspectiva de productividad de países, instituciones, contribución de autores y revistas, así como de los principales temas abordados en las publicaciones.

Es importante considerar que la herramienta de análisis bibliométrico utilizada en este estudio tiene algunas limitaciones, como resultados falsos positivos y falsos negativos debido a la imposibilidad de generar una consulta de investigación perfecta y exhaustiva.

Además, sólo se incluyeron artículos indexados en Scopus, por lo que resulta imposible afirmar que la investigación fuera una revisión de toda la literatura relacionada con la ciencia abierta y la innovación. Sin embargo, Scopus cubre un gran número de artículos y proporciona un elevado número de registros en términos de citas (Heradio *et al.* 2016; Shen y Ho 2020). En consecuencia, se afirma que se obtuvieron datos suficientes para esbozar el panorama científico, los temas de investigación más destacados y otros análisis realizados en este estudio.

La presente indagación proporcionó una visión general sobre el estado actual de la investigación en el campo de la innovación y la ciencia abierta. Los resultados revelan que, aunque se observó un crecimiento en la producción científica relacionada con este tema en los últimos años, el nivel de desarrollo en este campo aún se encuentra en una etapa incipiente.

Una de las características que debe tener la innovación la constituye la globalidad, y la implicación de la ciencia abierta en su desarrollo debería ser un tema por estudiar en todo el mundo. Se plantea que la ciencia abierta puede facilitar la colaboración y el intercambio de conocimientos entre investigadores (Nicholas *et al.* 2020). Sin embargo, los datos obtenidos en esta sugieren que esto no se reflejó en la conformación de equipos permanentes de investigación sobre ciencia abierta e innovación. Lo cual se deduce al observar que el 91,8 % de los autores incluidos en esta revisión sólo publicaron un artículo relacionado con el tema objeto de estudio, y con tan solo 0,21 artículos por autor.

Esto resultó bastante sorprendente porque los avances tecnológicos y el rápido procesamiento y difusión de la información científica supone reducir las barreras de la distancia geográfica y ampliar la colaboración interdisciplinar, facilitando así una investigación más sofisticada. A su vez, un aumento de la productividad científica en la temática en estudio tendría un mayor impacto en la contratación, la promoción y la financiación académica para estudios en el área, lo que beneficiaría a todas las partes interesadas, consolidando equipos de investigación permanentes.

Al analizar la distribución geográfica de la producción científica, destacó la dominancia de países europeos, con siete de ellos situados entre los diez más productivos. Esto considera un fuerte compromiso y enfoque en la innovación y la ciencia abierta dentro de la comunidad científica europea, y es coherente con lo planteado en trabajos previos (Budroni, Claude-Burgelman y Schouppe 2019). Además, reforzó el análisis de los temas predominantes relacionados con ciencia abierta e innovación, donde Europa ocupa un lugar central en los temas obtenidos (*figura 3*), compartiendo el mismo nodo con infraestructuras de investigación.

Durante los últimos años se identifica como factor crucial de la infraestructura de investigación la creciente participación de los ciudadanos en proyectos científicos. De esta manera, la ciencia ciudadana se convierte en una de las ocho prioridades clave de la Agenda Europea de Ciencia Abierta (2018), en paralelo con establecimiento de la European Open Science Cloud (EOSC), que facilita la colaboración de infraestructuras de investigación multidisciplinares (Otsu y Masó 2023). Esta última iniciativa (EOSC) también apareció en el análisis como un tema con un alto grado de desarrollo, pero aún con poca relevancia en la literatura, siendo todavía un tema de nicho en la relación entre ciencia abierta e innovación (*figura 3*). La correspondencia de esta iniciativa con el tema de estudio



fue directa, pues la misma se define como un portal de acceso a datos de investigación, herramientas y servicios para la innovación y la educación.<sup>7</sup>

La restricción idiomática que se estableció en la metodología del estudio al utilizar solamente palabras claves en inglés pudo excluir fuertemente publicaciones de países de habla hispana. Sin embargo, se observó a España como el cuarto país en número de publicaciones (*tabla 2*). Si bien Latinoamérica cuenta con importantes avances en iniciativas que promueven la adopción del Acceso Abierto y la Ciencia Abierta en la región (Heredia 2022), aún se presenta un bajo desarrollo de la investigación en el campo de la innovación (Tello-Gamarra *et al.* 2018), por lo que resulta esperable que estudios sobre la relación de ambas áreas también sea incipiente en esta parte del globo.

De manera tradicional, los derechos de propiedad intelectual se consideran como factor determinante para la innovación (Loukil 2020; Brüggemann *et al.* 2015). Desde esta afirmación se desprende una compleja relación, puesto que la titularidad de la propiedad intelectual limita el intercambio de datos, mientras que las iniciativas de ciencia abierta promueven la apertura de la información como medio para hacer avanzar el conocimiento científico y acelerar el progreso de la investigación y la innovación (Woelfle, Olliaro y Todd 2011). Por ello, resultó llamativo que en el presente estudio se identificó como uno de los temas motores en la relación entre ciencia abierta e innovación, un nodo formado por: la propiedad intelectual, intercambio de datos y política científica. Es decir, el estudio de la interrelación de estos tres factores presenta un gran desarrollo y relevancia que refleja la importancia que está teniendo en la comunidad académica el estudiar la compleja relación entre los derechos de propiedad intelectual, el intercambio de datos y cómo esto se incorpora en las políticas científicas.

Lo anterior también se visualiza desde los análisis de las palabras clave, donde se obtuvo que los términos «innovación abierta», «acceso» y «datos abiertos» presentan la mayor frecuencia. Esto es interesante, ya que resultó acorde con lo planteado por otros estudios donde se demuestra que las prácticas de ciencia abierta actúan como un generador de asociaciones entre el mundo académico y la industria para fomentar la innovación (Woelfle, Olliaro y Todd 2011).

En línea con lo anterior, otro de los temas motores en el presente análisis fue el estudio del código abierto y del hardware abierto. El hallazgo resultó coherente con la tendencia creciente del código y hardware abiertos en la investigación científica que promueve la transparencia, la colaboración y la innovación. La observación se respaldó con otros resultados obtenidos en el análisis, como fue la gran presencia de revistas relacionadas con el área informática entre las más relevantes; 11 de las 20 revistas pertenecieron a la categoría de informática

y/o tecnologías de la información. Sumado a lo anterior, el artículo más citado en relación con la ciencia abierta e innovación correspondió al desarrollo de un software de código abierto para diseñar y ejecutar experimentos psicológicos (Muehler y Piper 2014).

El código abierto es transparente, fiable y ofrece un paradigma para el flujo de trabajo y la innovación (Seabold y Perktold 2010). Los proyectos de código abierto aportan prácticas de innovación novedosas que pueden emularse en muchos campos (von Krogh y von Hippel 2006), y también pueden derivar en proyectos de innovación rentables, como demuestran los modelos empresariales basados en la propiedad intelectual abierta (Erickson 2018).

Sin embargo, por igual resultó notable la presencia de revistas de diferentes áreas, como *Research Policy*, *Sustainability* o *Journal of Responsible Innovation*, reflejando que la investigación sobre innovación y ciencia abierta se está abarcando desde un enfoque multidisciplinar.

La ciencia abierta y la colaboración internacional son fundamentales para abordar los problemas sanitarios mundiales (Kittrie *et al.* 2017), cuestión que tomó relevancia con la pandemia de Covid-19 para generar soluciones innovadoras y eficaces con la finalidad de afrontar la situación de salud global. Razón por la que no extrañó el resultado que se obtuvo respecto a la interacción, como tema motor, entre la interoperabilidad y la salud global. Sin embargo, en este mismo nodo apareció África como tema relacionado. Si bien la literatura ha destacado que las prácticas de ciencia abierta tienen un fuerte impacto en el rendimiento de la innovación (Ebersberger *et al.* 2012; Inauen y Schenker-Wicki 2011), se debe considerar el mercado en el cual se ponen en marcha estas actividades, ya que en economías poco desarrolladas las prácticas de ciencia abierta podrían tener efectos negativos sobre la innovación (Torres de Oliveira *et al.* 2020). Es por esto que se tornan necesarios mayores estudios para abordar dicho resultado.

La ciencia abierta tiene una relación positiva con la educación. El contar con recursos educativos abiertos aumenta el potencial del aprendizaje al incrementar la oportunidad de que las personas accedan a los recursos educativos e interactúen con ellos (Brüggemann *et al.* 2015). Esto, además, quedó de manifiesto durante la pandemia Covid-19, donde prácticas de ciencia abierta se utilizaron para reducir el impacto de la pandemia en la educación (Stracke *et al.* 2022). Lo anterior se reflejó en los resultados de este estudio, donde se observó que los temas relacionados de manera directa con la ciencia abierta (datos abiertos, innovación abierta) abordados principalmente entre 2019 a 2020 (y que figuran ya como temas básicos de estudio), dieron paso a temas conectados con la educación, enfocados de forma primordial en educación superior, pero igual en innovación educativa, aunque aún en poco grado en tanto relevancia y desarrollo, lo que se explicaría por lo reciente de su investigación.

## CONCLUSIONES

En conclusión, este estudio bibliométrico proporciona una visión general del estado actual de la investigación en el campo de la ciencia abierta y la innovación. Como es común en cualquier investigación bibliométrica, existe la posibilidad de obtener resultados restringidos debido a la dificultad de generar una consulta de investigación perfecta y exhaustiva. La base de datos utilizada, Scopus, abarca una gran cantidad de artículos y facilita un amplio número de registros en términos de citas, por lo que se concluye que se han obtenido datos suficientes para proporcionar una visión general de la actividad científica en el campo de la relación entre ciencia abierta e innovación.

Cabe destacar que la restricción idiomática utilizada en la metodología del estudio pudo haber excluido publicaciones de países de habla hispana de manera significativa. A pesar de esto, se observa que España se ubica en un lugar importante dentro de los países más productivos. Sin embargo, la prevalencia de países europeos liderando la productividad sobre el tema en estudio sugieren la importancia de fomentar la investigación y el desarrollo de iniciativas relacionadas con la ciencia abierta y la innovación en los países de habla hispana y en Latinoamérica en general. Esto contribuiría a fortalecer el conocimiento científico en estas áreas y promover la colaboración y el intercambio de ideas a nivel global.

Se observó un crecimiento en la producción científica relacionada con el tema en estudio en los últimos años, no obstante, el nivel de desarrollo aún se encuentra en etapa incipiente. Como resultado, destaca la relevancia que se ha dado al estudio de la compleja conexión entre la propiedad intelectual, el intercambio de datos y las políticas científicas, así como el papel creciente del código abierto en relación con la innovación. Además, se resalta la trascendencia de la colaboración internacional y la ciencia abierta en la búsqueda de soluciones innovadoras para los desafíos globales, como se evidencia en el nodo de interoperabilidad y salud global. Aunque se identificaron áreas de fortaleza, como la presencia de revistas multidisciplinarias, se requieren más investigaciones para comprender plenamente el impacto de la ciencia abierta en el desarrollo de la innovación en diferentes contextos, como en economías menos desarrolladas y en el ámbito de la educación, donde esto último aún se observa como un tema emergente. En general, este estudio contribuye a la comprensión de la evolución y los temas clave en la investigación sobre ciencia abierta e innovación, proporcionando información valiosa para los investigadores interesados en este campo.

## REFERENCIAS

- Aria, Massimo, y Corrado Cuccurullo. 2017. «Bibliometrix : An R-Tool for Comprehensive Science Mapping Analysis». *Journal of Informetrics* 11 (4): 959-75.  
<https://doi.org/10.1016/j.joi.2017.08.007>.
- Bliss, Tj, T. Jared Robinson, John Hilton, y David A. Wiley. 2013. «An OER COUP: College Teacher and Student Perceptions of Open Educational Resources». *Journal of Interactive Media in Education* 2013 (1): 4.  
<https://doi.org/10.5334/2013-04>.
- Brüggemann, Julia, Paolo Crosetto, Lukas Meub, y Kilian Bizer. 2015. «Intellectual Property Rights Hinder Sequential Innovation -- Experimental Evidence». *SSRN Electronic Journal*.  
<https://doi.org/10.2139/ssrn.2545950>.
- Budroni, Paolo, Jean Claude-Burgelman, y Michel Schouppe. 2019. «Architectures of Knowledge: The European Open Science Cloud». *ABI Technik* 39 (2): 130-41.  
<https://doi.org/10.1515/abitech-2019-2006>.
- Castellanos, F. Xavier, Adriana Di Martino, R. Cameron Craddock, Ashesh D. Mehta, y Michael P. Milham. 2013. «Clinical Applications of the Functional Connectome». *NeuroImage* 80 (octubre): 527-40.  
<https://doi.org/10.1016/j.neuroimage.2013.04.083>.
- Czarnitzki, Dirk, Christoph Grimpe, y Maikel Pellens. 2014. «Access to Research Inputs: Open Science Versus the Entrepreneurial University». *SSRN Electronic Journal*.  
<https://doi.org/10.2139/ssrn.2411925>.
- Ebersberger, Bernd, Carter Bloch, Sverre J. Herstad, y Els Van De Velde. 2012. «OPEN INNOVATION PRACTICES AND THEIR EFFECT ON INNOVATION PERFORMANCE». *International Journal of Innovation and Technology Management* 09 (06): 1250040.  
<https://doi.org/10.1142/S021987701250040X>.
- Erickson, Kristofer. 2018. «Can Creative Firms Thrive without Copyright? Value Generation and Capture from Private-Collective Innovation». *Business Horizons* 61 (5): 699-709.  
<https://doi.org/10.1016/j.bushor.2018.04.005>.
- Fell, Michael J. 2019. «The Economic Impacts of Open Science: A Rapid Evidence Assessment». *Publications* 7 (3): 46.  
<https://doi.org/10.3390/publications7030046>.
- Fernández Pinto, Manuela. 2020. «Open Science for private Interests? How the Logic of Open Science Contributes to the Commercialization of Research». *Frontiers in Research Metrics and Analytics* 5 (noviembre): 588331.  
<https://doi.org/10.3389/frma.2020.588331>.
- Frankenhuis, Willem E., y Daniel Nettle. 2018. «Open Science Is Liberating and Can Foster Creativity». *Perspectives on Psychological Science* 13 (4): 439-47.  
<https://doi.org/10.1177/1745691618767878>.
- Gold, E. Richard. 2021. «The Fall of the Innovation Empire and Its Possible Rise through Open Science». *Research Policy* 50 (5): 104226.  
<https://doi.org/10.1016/j.respol.2021.104226>.
- Heradio, Ruben, Luis de la Torre, Daniel Galan, Francisco Javier Cabrerizo, Enrique Herrera-Viedma, y Sebastian Dormido. 2016. «Virtual and Remote Labs in Education: A Bibliometric Analysis». *Computers & Education* 98 (julio): 14-38.  
<https://doi.org/10.1016/j.compedu.2016.03.010>.

- Heredia, Ana. 2022. «A tradition of open, academy-owned, and non-profit research infrastructure in Latin America». Editado por Bonnie Lawlor. *Information Services & Use* 42 (3-4): 447-52.  
<https://doi.org/10.3233/ISU-220177>.
- Heuritsch, Julia. 2020. «Knowledge Utilization and Open Science Policies: Noble aims that ensure quality research or Ordering discoveries like a pizza?»  
<https://doi.org/10.48550/ARXIV.2005.14021>.
- Himanen, Lauri, Amber Geurts, Adam Stuart Foster, y Patrick Rinke. 2019. «Data-Driven Materials Science: Status, Challenges, and Perspectives». *Advanced Science* 6 (21): 1900808.  
<https://doi.org/10.1002/advs.201900808>.
- Inauen, Matthias, y Andrea Schenker –Wicki. 2011. «The Impact of Outside– in Open Innovation on Innovation Performance». Editado por Sven Carlsson. *European Journal of Innovation Management* 14 (4): 496-520.  
<https://doi.org/10.1108/14601061111174934>.
- Invernizzi, Noela. 2020. «Hacia una recomendación de la UNESCO sobre la ciencia abierta. Crear un consenso mundial sobre la ciencia abierta». Consulta Regional Virtual sobre la Recomendación de la UNESCO en Ciencia Abierta. UNESCO.  
[https://es.unesco.org/sites/default/files/comentarios\\_sobre\\_ciencia\\_abierta\\_noela\\_invernizzi.pdf](https://es.unesco.org/sites/default/files/comentarios_sobre_ciencia_abierta_noela_invernizzi.pdf).
- Jong, Simcha, y Kremena Slavova. 2014. «When Publications Lead to Products: The Open Science Conundrum in New Product Development». *Research Policy* 43 (4): 645-54.  
<https://doi.org/10.1016/j.respol.2013.12.009>.
- Kittrie, Elizabeth, Audie A. Atienza, Robert Kiley, David Carr, Aki MacFarlane, Vinay Pai, Jennifer Couch, *et al.* 2017. «Developing International Open Science Collaborations: Funder Reflections on the Open Science Prize». *PLOS Biology* 15 (8): e2002617.  
<https://doi.org/10.1371/journal.pbio.2002617>.
- Krogh, Georg von, y Eric von Hippel. 2006. «The Promise of Research on Open Source Software». *Management Science* 52 (7): 975-83.  
<https://doi.org/10.1287/mnsc.1060.0560>.
- LeBel, Etienne P., Lorne Campbell, y Timothy J. Loving. 2017. «Benefits of Open and High-Powered Research Outweigh Costs.» *Journal of Personality and Social Psychology* 113 (2): 230-43.  
<https://doi.org/10.1037/pspi0000049>.
- Loukil, Kamilia. 2020. «Intellectual Property Rights, Human Capital and Innovation in Emerging and Developing Countries». *Journal of Social Economics Research* 7 (1): 35-41.  
<https://doi.org/10.18488/journal.35.2020.71.35.41>.
- Markowitz, David Matthew, Hyunjin Song, y Samuel Hardman Taylor. 2021. «Tracing the Adoption and Effects of Open Science in Communication Research». Preprint. PsyArXiv.  
<https://doi.org/10.31234/osf.io/dsf67>.
- McKiernan, Erin C., Philip E Bourne, C Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, *et al.* 2016. «How Open Science Helps Researchers Succeed». *ELife* 5 (julio): e16800.  
<https://doi.org/10.7554/eLife.16800>.
- Mirowski, Philip. 2018. «The Future(s) of Open Science». *Social Studies of Science* 48 (2): 171-203.  
<https://doi.org/10.1177/0306312718772086>.

- Mueller, Shane T., y Brian J. Piper. 2014. «The Psychology Experiment Building Language (PEBL) and PEBL Test Battery». *Journal of Neuroscience Methods* 222 (enero): 250-59.  
<https://doi.org/10.1016/j.jneumeth.2013.10.024>.
- Murray, Fiona, y Scott Stern. 2007. «Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?» *Journal of Economic Behavior & Organization* 63 (4): 648-87.  
<https://doi.org/10.1016/j.jebo.2006.05.017>.
- Nicholas, David, Cherifa Boukacem-Zeghmouri, Abdullah Abrizah, Blanca Rodríguez-Bravo, Jie Xu, Marzena –wigo–, Anthony Watkinson, y Eti Herman. 2020. «Open science from the standpoint of the new wave of researchers: Views from the scholarly frontline1». Editado por Bonnie Lawlor. *Information Services & Use* 39 (4): 369-74.  
<https://doi.org/10.3233/ISU-190069>.
- Otsu, Kaori, y Joan Masó. 2023. «Open Science as the new normal, Citizen Science as the new component of research infrastructure». Other. oral.  
<https://doi.org/10.5194/egusphere-egu23-172>.
- Ramírez-Montoya, María-Soledad, y Francisco-José García-Peñalvo. 2018. «Co-creation and open innovation: Systematic literature review». *Comunicar* 26 (54): 09-18.  
<https://doi.org/10.3916/C54-2018-01>.
- Scheliga, Kaja, y Sascha Friesike. 2014. «Putting open science into practice: A social dilemma?» *First Monday*, 19 (9) agosto.  
<https://doi.org/10.5210/fm.v19i9.5381>.
- Schumpeter, Joseph, y Ursula Backhaus. 2003. «The Theory of Economic Development». En *Joseph Alois Schumpeter*, editado por Jürgen Backhaus, 1:61-116. The European Heritage in Economics and the Social Sciences. Boston: Kluwer Academic Publishers.  
[https://doi.org/10.1007/0-306-48082-4\\_3](https://doi.org/10.1007/0-306-48082-4_3).
- Seabold, Skipper, y Josef Perktold. 2010. «Statsmodels: Econometric and Statistical Modeling with Python». En 9th Python in Science Conference 92-96. Austin, Texas.  
<https://doi.org/10.25080/Majora-92bf1922-011>.
- Shen, Chien-wen, y Jung-tsung Ho. 2020. «Technology-Enhanced Learning in Higher Education: A Bibliometric Analysis with Latent Semantic Approach». *Computers in Human Behavior* 104 (marzo): 106177.  
<https://doi.org/10.1016/j.chb.2019.106177>.
- Song, Hyunjin, David Matthew Markowitz, y Samuel Hardman Taylor. 2021. «Trusting on the Shoulders of Open Giants? Open Science Increases Trust in Science for the Public and Academics». Preprint. Open Science Framework.  
<https://doi.org/10.31219/osf.io/g328c>.
- Song, Yu, Xieling Chen, Tianyong Hao, Zhinan Liu, y Zixin Lan. 2019. «Exploring Two Decades of Research on Classroom Dialogue by Using Bibliometric Analysis». *Computers & Education* 137 (agosto): 12-31.  
<https://doi.org/10.1016/j.compedu.2019.04.002>.
- Stracke, Christian M. 2020. «Open Science and Radical Solutions for Diversity, Equity and Quality in Research: A Literature Review of Different Research Schools, Philosophies and Frameworks and Their Potential Impact on Science and Education». En *Radical Solutions and Open Science*, editado por Daniel Burgos, 17-37. Lecture Notes in Educational Technology. Singapore: Springer Singapore.  
[https://doi.org/10.1007/978-981-15-4276-3\\_2](https://doi.org/10.1007/978-981-15-4276-3_2).

- Stracke, Christian M., Daniel Burgos, Gema Santos-Hermosa, Aras Bozkurt, Ramesh Chander Sharma, Cécile Swiatek Cassafieres, Andreia Inamorato dos Santos, *et al.* 2022. «Responding to the Initial Challenge of the COVID-19 Pandemic: Analysis of International Responses and Impact in School and Higher Education». *Sustainability* 14 (3): 1876.  
<https://doi.org/10.3390/su14031876>.
- Tello Gamarra, Jorge, Ricardo Machado Leo, Ariane Mello Silva Avila, y Jonatas Wendland. 2018. «Innovation Studies in Latin America: A Bibliometric Analysis». *Journal of Technology Management & Innovation* 13 (4): 24-36.  
<https://doi.org/10.4067/S0718-27242018000400024>.
- Torres de Oliveira, Rui, Martie-Louise Verreynne, Sandra Figueira, Marta Indulska, y John Steen. 2020. «How Do Institutional Innovation Systems Affect Open Innovation?» *Journal of Small Business Management* 60 (6): 1404-48.  
<https://doi.org/10.1080/00472778.2020.1775466>.
- Vicente-Saez, Ruben, y Clara Martinez-Fuentes. 2018. «Open Science Now: A Systematic Literature Review for an Integrated Definition». *Journal of Business Research* 88 (julio): 428-36.  
<https://doi.org/10.1016/j.jbusres.2017.12.043>.
- Vlokhoven, Has van. 2019. «The Effect of Open Access on Research Quality». *Journal of Informetrics* 13 (2): 751-56.  
<https://doi.org/10.1016/j.joi.2019.04.001>.
- Watson, Mick. 2015. «When Will ‘Open Science’ Become Simply ‘Science’?» *Genome Biology* 16 (1): 101.  
<https://doi.org/10.1186/s13059-015-0669-2>.
- Woelfle, Michael, Piero Olliaro, y Matthew H. Todd. 2011. «Open Science Is a Research Accelerator». *Nature Chemistry* 3 (10): 745-48.  
<https://doi.org/10.1038/nchem.1149>.

### *Para citar este texto:*

Guevara-Pezoa, Felipe. 2023. “La ciencia abierta y su relación con la innovación: una revisión bibliométrica”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 109-128.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58778>

# Inconsistency in the registration of the Digital Object Identifier (DOI) of articles on Web of Science and Scopus

Erwin Krauskopf\*  
Mauricio Salgado\*\*

*Artículo recibido:*  
19 de abril de 2023  
*Artículo aceptado:*  
1 de agosto de 2023

*Artículo de investigación*

## ABSTRACT

Almost 25 years ago, the Digital Object Identifier (DOI) was implemented with the purpose of providing a unique and persistent form of document identification. As DOIs can be assigned to any object, journals rapidly adopted their use as it eases the process of identifying a specific document, thus increasing its visibility. However, while studying the impact of papers published by Chilean researchers in disciplines related to Social Sciences, Arts, and Humanities, we noticed that some journals published a considerable proportion of documents without DOIs. Thus, we questioned if this was due to a lack of DOI adoption by these journals or if it was a database

- \* Vicerrectoría de Investigación, Universidad de las Américas, Chile  
ekrauskopf@udla.cl
- \*\* Centro de Estudios Públicos, Chile  
msalgado@cepchile.cl



processing error. Our findings indicate that while most journals have adopted the use of DOI for article-type documents, many of the Web of Science and Scopus records do not include this information. To overcome this issue, databases need to ensure the accuracy and consistency of their data, while the editorial management team of each journal ought to confirm that all the metadata from their articles has been properly registered by these databases. Undoubtedly, the use of DOI benefits all academics as it facilitates the discoverability and retrieval of the published articles.

**Keywords:** Digital Object Identifier; DOI; Web of Science; Scopus; Chile

### **Inconsistencia en el registro del Identificador de Objetos Digitales (DOI) de artículos en Web of Science y Scopus**

*Erwin Krauskopf y Mauricio Salgado*

#### **RESUMEN**

Hace aproximadamente 25 años, el Identificador de Objetos Digitales (DOI) fue implementado con el propósito de proveer una forma única y persistente de identificación de un documento. Dado que los DOIs puede ser asignados a cualquier objeto, las revistas comenzaron a adoptar su uso ya que facilita el proceso de identificación de un documento en particular, aumentando su visibilidad. Sin embargo, al analizar el impacto de artículos publicados por investigadores Chilenos en disciplinas relacionadas a las Ciencias Sociales, Artes y Humanidades, observamos que algunas revistas publicaban una proporción significativa de documentos sin DOIs. Por lo tanto, nos preguntamos si esto era consecuencia de una falta de adopción de DOI por las revistas o si era por un error de procesamiento de las bases de datos. Nuestros resultados confirman que, a pesar de que la mayoría de las revistas implementó el uso del DOI para sus artículos, varios de los registros de Web of Science y Scopus no incluyen esta información. Para resolver este problema, las bases de datos deben asegurar la precisión y consistencia de sus datos, mientras que los equipos editoriales de cada revista deben confirmar que todos los metadatos de los artículos han sido registrados apropiadamente por estas bases de

datos. Sin duda, el uso del DOI beneficia a todos los académicos ya que facilita la identificación y recuperación de los artículos publicados.

**Palabras Clave:** Identificador de Objetos Digitales; DOI; Web of Science; Scopus; Chile

## INTRODUCTION

“Publish or perish” a well-too-know axiom to academics involved in research throughout the world. In simple terms, researchers feel impelled to publish periodically to be recruited and to secure tenure or funding for their research. The fact that a high publication rate reflects academic success has led to a numbers game (Parnas, 2007) that has not only caused an increase in the quantity of worldwide publications but has also encouraged a series of misconducts (Fanelli, 2009; Hvistendahl, 2015). While in the year 2000 approximately 1.3 million publications were registered by Scopus, by the end of 2021 that number increased to approximately 3.7 million publications. Similarly, Web of Science registered approximately 1.2 million publications in the year 2000 versus 3.6 million in 2021. As the quantity of publications rises rapidly, there is a need to make one’s research visible to other academics.

Undoubtedly, the massification of the internet in the 1990s led to changes in the structure of scholarly communications. A turning point was the availability of the first graphical browser for the world-wide-web in 1994 (Haynes, 1999). Journals, which used to print issues to communicate research outcomes, began embracing the use of the World Wide Web to disseminate knowledge by establishing webpages which provided a venue to access the information that was being published. At first, access to these documents worked as expected. Nonetheless, after some years, users began encountering error messages (such as “404 not found”) when trying to access the journal’s content, which occurred when journals changed their URL (Germain, 2000). While this study established that, after a three-year period, 50% of URL citations from 31 journals could not be accessed (Germain, 2000), a later work reported that almost 17% of the URL references from all the articles published by PLoS Medicine between 2005-2007 were non-operational (Nagajara *et al.*, 2011). Due to the dynamics of the web, several other studies also found a high proportion of URL citations that were not accessible through the web browsers (Sampath Kumar and Vinay Kumar, 2013; Vinay Kumar and Sampath Kumar, 2017; Manjunatha *et al.*, 2020; Shanthakumari, 2021). As the URL refers to the home address of a journal, any changes in

its web architecture will affect the accessibility of a particular publication (Ansorge, 2023).

The digital object identifier (commonly known as DOI), which originated in 1998, offered a way to overcome this limitation. The DOI was conceived as a persistent and unique identifier of various types of physical and electronic objects (Paskin, 2005; Liu, 2021). It is managed by the International DOI Foundation, a non-profit organization that governs the federation of registration agencies (Wang *et al.*, 2018). The DOI name consists of a case-insensitive string made up of a prefix beginning with “10.” that represents a unique naming authority and a suffix, separated by a slash, that indicates the specific object (Paskin, 2005). Among its advantages are its capability to unambiguously identify an object, providing a simple and efficient system to track a journal article. This allows other web-based tools, such as Altmetric Explorer, to track journal articles and provide alternative metrics on the use and impact of the research at an article-level. While these metrics mainly focus on activity in online environments (such as Twitter, YouTube and Wikipedia), citation data is collected from the Dimensions database and patent data is collected from IFI Claims (<https://help.altmetric.com/support/solutions/articles/6000236721-patents>). The Altmetric Attention Score (AAS) is derived from these metrics, (<https://help.altmetric.com/support/solutions/articles/6000233311-how-is-the-altmetric-attention-score-calculated->). Though various studies have shown dissimilar results regarding a correlation between AAS and citations (Cintra *et al.*, 2018; Breuer *et al.*, 2022; Chen and Wang, 2022; Liu y Huang, 2022), Altmetrics still provides a means to assess online conversations surrounding a research output.

Limitations in the use of traditional bibliometrics to assess disciplines related to social sciences, arts, and humanities have been widely reported. Because some research topics are generally locally or regionally orientated, the corresponding articles are published in the language of the region where the research was developed (Frandsen y Nicolaisen, 2008; Hammarfelt and Haddow, 2018). Consequently, articles published in any other language besides English may not appear in search queries performed in either of the two main two databases –Web of Science and Scopus– since both index a significantly higher proportion of English-language journals (Hammarfelt and Haddow, 2018). Furthermore, some researchers tend to publish monographs or books to communicate their findings, and these outputs are usually not indexed by these databases (Larivière *et al.*, 2006; Marx and Bornmann, 2015). And contrary to what occurs in natural science, citation accumulation in social sciences, arts, and humanities tends to be slower over the years, causing a detrimental effect during curricular assessments of researchers (Pajic *et al.*, 2019). It is important to note that both databases tally citations based on the references registered by each database (Marx and Bornmann, 2015).

While studying the online attention generated by articles published by Chilean researchers in journals included in the Social Sciences Citation Index and the Arts and Humanities Citation Index from Web of Science, we noticed that not only some documents lacked a DOI, but some journals exhibited an irregular use of DOIs throughout the 5-year period assessed. To verify whether the lack of DOIs was due to database processing errors or to the absence of DOIs in the articles, we repeated the query using Scopus as the source of information and later verified the information by accessing the website of each journal. Therefore, this study aimed to establish whether the DOIs associated to every article are being registered by the two main citation databases, using data from 40 journals that exhibited the highest degree of irregularity in DOI usage according to the Web of Science output. To our knowledge, only one previous study reported issues with the registration of DOIs by a citation database for documents published by two journals between 2005-2014 (Gorraiz *et al.*, 2016). It is important to note that while more recent studies have reported that a proportion of the records downloaded from these databases lacked DOIs (Khurana *et al.*, 2022; Mugnaini *et al.*, 2021), none performed a DOI check in the journals' websites to determine the origin of the problem.

## MATERIALS AND METHODS

Bibliometric data for this study was retrieved from the Web of Science (WoS) database in December 2021. The WoS database was chosen for this study because the main Chilean research grant (Fondecyt) utilizes this database to assess all researchers as part of the evaluation process performed by the Chilean National Agency for Research and Development (ANID).

The search criteria (CU=Chile and PY=2015-2019) was used to extract publications registered by the Arts & Humanities Citation Index (A&HCI) and the Social Science Citation Index (SSCI) collections. This query resulted in 12,396 documents that registered at least one affiliation to a Chilean institution. The downloaded data was sorted and processed using Microsoft Excel. For each document we obtained the title, source, ISSN, publisher, document type and DOI.

The Scopus database was used to search for article-type documents for the 40 journals that exhibited the highest degree of irregularity in DOI usage according to the previous Web of Science query.

Furthermore, publisher names were normalized and classified into one of three categories (Krauskopf, 2021):

- Commercial Publisher: Defined as a profit-oriented firm not associated with universities.
- Non-university academic publisher: Refers to scientific societies or any other type of academic body not associated with universities.
- University Press: Any publisher belonging to a university.

## RESULTS

Among the 12,396 documents that were published between 2015-2019 by Chilean researchers, we observed that 3,085 (24.9%) were not associated to a DOI. These documents were published in 323 different journals. To assess the implementation of DOIs over the period 2015-2019, we selected all the journals that had published a minimum of 20 documents without DOIs (*Figure 1*). The purpose was to determine whether these journals were not employing DOIs or whether DOI usage was inconsistent throughout the selected period. This accounted for 40 journals (*Table 1*), which have been sorted by publisher type (commercial, non-university academic publisher or university publisher). Among the journals managed by commercial publishers, the Journal of Dental Education attracted attention because it was the only journal that appeared to not have implemented DOI usage during 2015 and 2016. As this seemed peculiar, we consulted the journal website (<https://onlinelibrary.wiley.com/loi/19307837>) to verify whether this was correct. To our surprise, we confirmed that all the papers published in 2015 and 2016 were associated with a DOI. Therefore, we questioned whether this inaccuracy extended to the other 39 journals. One way to solve this question was to download data from another database, Scopus, which indexed the same journals. However, as Web of Science and Scopus differ slightly in the type of documents registered, we focused our queries on article-type documents as these are recorded by both databases (Liu *et al.*, 2021).

*Table 2* shows the total number of article-type documents published by each journal between 2015 and 2019 and the proportion of these documents that registered a DOI, according to Web of Science and Scopus. The differences observed across data sources are concerning considering we have adjusted our queries to the same document type. Among 16 journals (Revista Argentina de Clínica Psicológica, RLA – Revista de Lingüística Teórica y Aplicada, ALPHA – Revista de Artes Letras y Filosofía, Arte Individuo y Sociedad, Convergencia – Revista de Ciencias Sociales, Estudios Filológicos, Historia Unisinos, Historia y Comunicación Social, Ideas y Valores, Pensamiento, Revista 180, Revista Chilena de Derecho, Revista Chilena de Literatura, Revista de Geografía Norte

Grande, Revista Musical Chilena, Teología y Vida) we established a difference of more than 20% in the number of article-type documents registered by both databases. Furthermore, in the case of the journals *Historia Unisinos* and *Historia y Comunicación Social* this difference is over 100%. It is important to note that Scopus's coverage of the journal *Revista de Filosofía* began in 2016, thus it was not included among these journals.

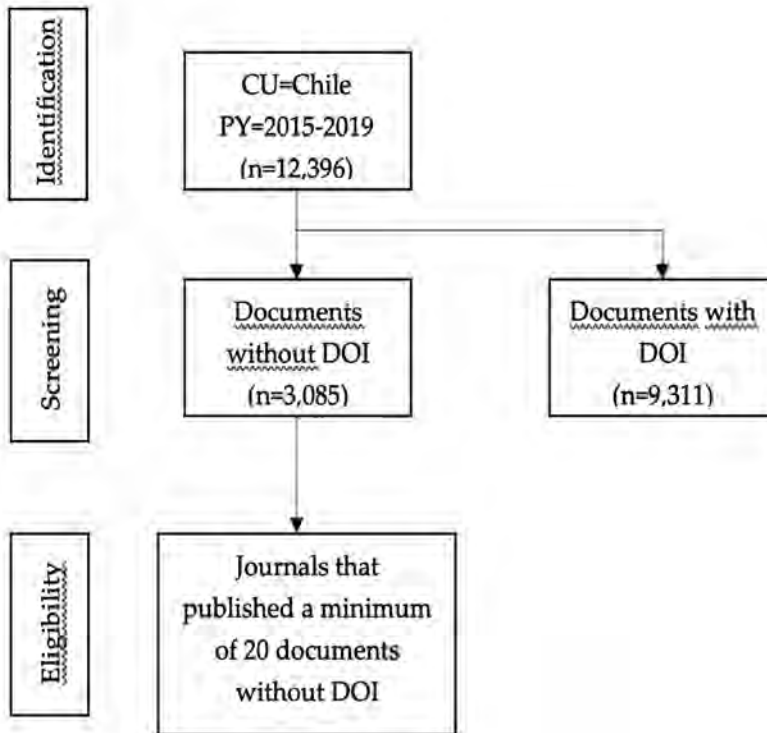


Figure 1. Overview of the search query and journal eligibility process

In addition, *Table 2* shows that Web of Science registered less DOIs associated to article-type documents in comparison to Scopus. In fact, while Web of Science registered 100% of DOIs for only three journals (*Economic Systems*, *Journal of the Royal Statistical Society Series*, and *Historia Unisinos*), Scopus reported 100% for 17 journals. Since we had previously established a database processing error with the *Journal of Dental Education*, we accessed the websites of the 23 journals that exhibited less than 100% of article-type documents associated to a DOI to confirm whether it was due to the absence of DOI in the document or another database processing error (*Table 3*).

An analysis of each article-type document published by the 23 journals revealed that the DOIs were under recorded by Web of Science and Scopus for 20 publications (*Table 3*). In fact, these 20 journals had a DOI associated to the totality of article-type documents. It is alarming that for a journal such as *Revista 180 Web of Science* only registered 21% of DOIs and Scopus 39%, even though the journal had a DOI associated to every article-type document. A similar situation was observed for another journal, *Revista Panamericana de Salud Publica*, for which less than 50% of DOIs were registered by both databases.

Finally, special attention needs to be drawn to the journal *Atenea*. While all the article-type documents published in the five-year period included a DOI (*Table 3*) only 26% were associated to a DOI (*Table 2*) according to Web of Science. A closer look revealed that, for some unknown reason, Web of Science did not register the DOIs for all the article-type documents published in 2016 and those published in 2019 lacked DOIs altogether. Contrarily, Scopus registered DOIs for 63% of article-type documents published by the journal on every year assessed.

Journal Title	Publisher <sup>1</sup>	2015	2016	2017	2018	2019
Economic Systems	CP	98%	100%	100%	100%	100%
Ergonomics	CP	93%	100%	100%	99%	100%
International Labour Review	CP	98%	100%	96%	100%	67%
Journal of Dental Education	CP	0%	0%	67%	98%	67%
Journal of the Royal Statistical Society Series A	CP	96%	94%	99%	97%	86%
Lancet Global Health	CP	100%	100%	98%	92%	100%
Acta Literaria	NUAP	32%	0%	37%	7%	27%
Revista Argentina de Clinica Psicológica	NUAP	0%	0%	89%	94%	100%
Revista Iberoamericana de Diagnostico y Evaluacion	NUAP	0%	47%	95%	95%	95%
Revista Médica de Chile	NUAP	86%	40%	71%	74%	47%
Revista Panamericana de Salud Publica	NUAP	1%	0%	38%	100%	97%
RLA – Revista de Lingüística Teórica y Aplicada	NUAP	44%	53%	50%	75%	29%
Terapia Psicológica	NUAP	96%	32%	31%	78%	42%
Acta Bioethica	UP	22%	19%	27%	27%	24%
ALPHA – Revista de Artes Letras y Filosofía	UP	46%	20%	42%	56%	93%

Andamios	UP	36%	38%	29%	77%	86%
ARQ	UP	31%	48%	22%	38%	44%
Arte Individuo y Sociedad	UP	88%	70%	98%	100%	100%
Atenea	UP	45%	0%	48%	27%	7%
Chungara – Revista de Antropología Chilena	UP	27%	93%	100%	100%	100%
Convergencia – Revista de Ciencias Sociales	UP	0%	0%	26%	100%	100%
Estudios Atacameños	UP	32%	9%	16%	100%	100%
Estudios Filológicos	UP	28%	30%	31%	43%	70%
EURE	UP	95%	86%	67%	100%	93%
Historia Unisinos	UP	97%	95%	90%	95%	96%
Historia y Comunicación Social	UP	81%	100%	86%	98%	98%
Ideas y Valores	UP	95%	87%	68%	95%	83%
Magallania	UP	76%	79%	71%	87%	29%
Papeles de Población	UP	0%	25%	92%	100%	100%
Pensamiento	UP	42%	8%	57%	38%	52%
Revista 180	UP	0%	0%	0%	5%	95%
Revista Chilena de Derecho	UP	17%	27%	29%	12%	20%
Revista Chilena de Literatura	UP	44%	20%	37%	32%	28%
Revista de Ciencias Política	UP	45%	45%	41%	86%	59%
Revista de Filosofía	UP	50%	55%	0%	28%	0%
Revista de Geografía Norte Grande	UP	27%	46%	46%	50%	42%
Revista Musical Chilena	UP	24%	29%	21%	24%	0%
Revista Signos	UP	100%	72%	100%	100%	100%
Teología y Vida	UP	67%	53%	77%	91%	36%
Universitas Psychologica	UP	26%	89%	95%	97%	100%

1 Publisher indicates whether it corresponds to a commercial publisher (CP), non-university academic publisher (NUAP) or university publisher (UP)

Table 1. Proportion of documents associated with a DOI published by each journal



Journal Title	Web of Science			Scopus		
	Total	DOI	% DOI	Total	DOI	% DOI
Economic Systems	209	208	100%	209	209	100%
Ergonomics	708	696	98%	692	692	100%
International Labour Review	137	125	91%	139	139	100%
Journal of Dental Education	613	316	52%	746	556	75%
Journal of the Royal Statistical Society Series A	284	284	100%	261	260	100%
Lancet Global Health	376	371	99%	414	414	100%
Acta Literaria	84	22	26%	80	80	100%
Revista Argentina de Clínica Psicológica	170	115	68%	213	154	72%
Revista Iberoamericana de Diagnóstico y Evaluación	200	165	83%	211	156	74%
Revista Médica de Chile	907	603	66%	977	969	99%
Revista Panamericana de Salud Publica	572	265	46%	481	232	48%
RLA - Revista de Lingüística Teórica y Aplicada	69	38	55%	54	54	100%
Terapia Psicológica	101	57	56%	105	104	99%
Acta Bioethica	148	46	31%	152	152	100%
ALPHA - Revista de Artes Letra y Filosofía	183	99	54%	147	146	99%
Andamios	191	115	60%	206	204	99%
ARQ	191	73	38%	222	204	92%
Arte Individuo y Sociedad	206	195	95%	160	160	100%
Atenea	129	33	26%	134	84	63%
Chungara - Revista de Antropología Chilena	211	174	82%	228	129	57%
Convergencia - Revista de Ciencias Sociales	136	63	46%	102	50	49%
Estudios Atacameños	160	96	60%	181	101	56%
Estudios Filológicos	126	59	47%	94	58	62%
EURE	184	163	89%	191	191	100%
Historia Unisinos	187	187	100%	89	89	100%

Historia y Comunicación Social	162	160	99%	56	55	98%
Ideas y Valores	234	228	97%	170	170	100%
Magallania	142	104	73%	156	156	100%
Papeles de Población	175	113	65%	185	114	62%
Pensamiento	317	172	54%	188	176	94%
Revista 180	98	21	21%	54	21	39%
Revista Chilena de Derecho	184	43	23%	103	98	95%
Revista Chilena de Literatura	163	52	32%	95	57	60%
Revista de Ciencias Política	138	72	52%	143	143	100%
Revista de Filosofía (*)	83	29	35%	46	46	100%
Revista de Geografía Norte Grande	155	68	44%	125	62	50%
Revista Musical Chilena	64	25	39%	91	90	99%
Revista Signos	132	124	94%	139	139	100%
Teología y Vida	98	62	63%	125	125	100%
Universitas Psychologica	543	436	80%	582	567	97%

"Total" indicates the total number of article-type documents published in the 5-year period "DOI" indicates the total number of article-type documents that had a DOI registered by the databases; "%" indicates the proportion of article-type documents that was associated to a DOI. (\*) Journal coverage began in 2016, thus the number of article-type documents associated with this journal is for the period 2016-2019

Table 2. Article-type documents published by each journal between 2015-2019

Journal Title	Website		
	Total	DOI	% DOI
Journal of Dental Education	746	746	100%
Revista Argentina de Clínica Psicológica	213	154	72%
Revista Iberoamericana de Diagnóstico y Evaluación	211	211	100%
Revista Médica de Chile	977	977	100%
Revista Panamericana de Salud Pública	481	481	100%
Terapia Psicológica	105	105	100%
ALPHA - Revista de Artes Letras y Filosofía	147	147	100%
Andamios	206	206	100%

ARQ	222	222	100%
Atenea	134	134	100%
Chungara – Revista de Antropología Chilena	228	228	100%
Convergencia – Revista de Ciencias Sociales	102	102	100%
Estudios Atacameños	181	181	100%
Estudios Filológicos	94	94	100%
Historia y Comunicación Social	56	56	100%
Papeles de Población	185	114	62%
Pensamiento	188	188	100%
Revista 180	54	54	100%
Revista Chilena de Derecho	103	103	100%
Revista Chilena de Literatura	95	95	100%
Revista de Geografía Norte Grande	125	125	100%
Revista Musical Chilena	91	90	100%
Universitas Psychologica	582	567	97%

Table 3. Proportion of article-type documents published by each journal that is associated to a DOI according to each journals' website

## DISCUSSION AND CONCLUSION

Databases play an important role during research assessment. Data extracted from Web of Science and Scopus, two high-cost subscription databases, is recurrently utilized to produce university rankings, evaluate tenure and promotion at universities, perform curricular assessments of researchers by grant funding agencies, etc. While these databases are perceived as accurate regarding the information registered, they are not exempt of errors (Krauskopf, 2019; Shuo *et al.*, 2019; Liu, 2020; Purnell, 2022; Savchenko y Kosyakov, 2022). Concerning errors in the DOI field, a few years ago Franceschini *et al.* (2015) established the existence of some DOIs that were shared between different publications indexed by Scopus, distorting the outcome of the bibliometric analysis. A later study from Zhu *et al.* (2019) reported 119 DOIs downloaded from Web of Science that were not found in the DOI System. Considering that the DOI was conceived as a unique and permanent identifier, the existence of such cases is perplexing.

In this study we report inconsistencies in DOI registration by Web of Science and Scopus. Even though most of the analyzed journals have associated a DOI to every article-type document (*Table 3*), both databases displayed problems registering this information. While Gorraiz *et al.* (2016) previously established that Web of Science had problems registering DOIs for documents published by two journals (out of 35), we determined this matter has not been solved for Latin American journals managed by non-university academic publishers and university presses in disciplines related to the Social Sciences, Arts, and Humanities. Likewise, Sanz-Casado *et al.* (2021) reported that DOI usage was higher (62%) for Norwegian, Finnish, and Spanish journals in the field of humanities, in comparison to journals managed by university presses (59%) and non-university academic publishers (29%). Besides the complications that may arise from citation tracking (Liu, 2021; Habermann, 2023), the missing DOIs in the downloaded data prevents users from utilizing these identifiers to track the online attention of these articles through Altmetrics.

In a recent study using data extracted from Web of Science, Khurana *et al.* (2022) established that 16.1% of the documents published by 1000 journals were without a DOI. It is important to note that this study did not discriminate between document types, therefore we expected to find a minor proportion of documents (such as news items or corrections) not associated to a DOI. Conversely, the study published by Mugnaini *et al.* (2021) which only considered three document types (articles, reviews, and proceedings papers) written by Brazilian researchers between 2012-2016, established that 68% of them included a DOI.

Another important issue that needs to be considered is the fact that to obtain a DOI publishers need to pay a fee, which may be cost-prohibited for all document types. A study on Ecuadorian journals suggested that this might be the reason why only 78% complied with the use of DOI (Santiago-Padilla y Ramos-Gil, 2022). Similarly, another study on Chinese journals reported that 80% used DOIs due to a lack of government support and lack of understanding about their importance. (Wang *et al.*, 2018).

While databases need to continually improve their procedure to annotate all the metadata provided by each journal to ensure the accuracy and consistency of the collected data, the editorial management team also share responsibility in this matter. By the time databases collect metadata from the journal's website, all articles should have DOIs assigned. A delay in the process of assigning DOIs will result in databases registries lacking DOIs. It is important to note that, for a researcher, choosing a journal to submit a manuscript is not trivial as everyone hopes that their work will be visible to others. Thus, the implementation of DOIs by journals will not only assist citation tracking but will increase their visibility.

Funding: This research was funded by ANID-FONDECYT, grant number 1211494.

## REFERENCES

- Ansorge, L. (2023) "Hidden limitations of analyses via alternative bibliometric services". *Scientometrics* 128(3): 2031-2033.  
<https://doi.org/10.1007/s11192-022-04626-w>
- Breuer, Timo., Philipp Schaer y Dirk Tunger. 2022. "Relevance assessments, bibliometrics, and altmetrics: a quantitative study on Pubmed and arXiv". *Scientometrics* 127(5): 2455-2478.  
<https://doi.org/10.1007/s11192-022-04319-4>
- Chen, Ming y Linzi Wang. 2022. "An Altmetrics and citation analysis of selected predatory journals in library and information science field". *Journal of Academic Librarianship* 48(6): 102618.  
<https://doi.org/10.1016/j.acalib.2022.102618>
- Cintra, Paulo Roberto, Ariadne Chloe Furnival y Douglas Henrique Milanez. 2018. "The impact of open access citation and social media on leading top information science journals". *Investigacion Bibliotecologica* 32(77): 117-132.  
<https://doi.org/10.22201/iibi.24488321xe.2018.77.57874>
- Fanelli, Daniele. 2009. "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data". *PLoS ONE* 4(5): e5738.  
<https://doi.org/10.1371/journal.pone.0005738>
- Franceschini, Fiorenzo, Domenico Maisano y Luca Mastrogiacomio. 2015. "Errors in DOI indexing by bibliometric databases". *Scientometrics* 102(3): 2181-2186.  
<https://doi.org/10.1007/s11192-014-1503-4>
- Frandsen, Tove Faber y Jeppe Nicolaisen. 2008. "Intradisciplinary differences in database coverage and the consequences for bibliometric research". *Journal of the American Society for Information Science and Technology* 59(10): 1570-1581.  
<https://doi.org/10.1002/asi.20817>
- Germain, Anne Carol. 2000. "URLs: Uniform resource locators or unreliable resource locators". *College & Research Libraries* 61(4): 359-365.  
<https://doi.org/10.5860/crl.61.4.35>
- Gorraiz, Juan, David Melero-Fuentes, Christian Gumpenberger y Juan Carlos Valderrama-Zurian. 2016. "Availability of digital object identifiers (DOIs) in Web of Science and Scopus". *Journal of Informetrics*. 10(1): 98-109.  
<https://doi.org/10.1016/j.joi.2015.11.008>
- Habermann, Ted. 2023. "Improving domain repository connectivity". *Data Intelligence* 5(1): 6-26.  
[https://doi.org/10.1162/dint\\_a\\_00120](https://doi.org/10.1162/dint_a_00120)
- Hammarfelt, Bjorn y Gaby Haddow. 2018. "Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators". *Journal of the Association for Information Science and Technology* 69(7): 924-935.  
<https://doi.org/10.1002/asi.24043>
- Haynes, J. 1999. "New Journal of Physics: a web-based and author-funded journal." *Learned Publishing*. Vol. 12: 265-269
- Hvistendahl, Mara. 2015. "China pursues fraudsters in science publishing: Measures may not be enough to stem the tide, some fear". *Science* 350(61264): 1015.  
<https://doi.org/10.1126/science.350.6264.1015>

- Khurana, Parul, Geetha Ganesan, Gulshan Kumar, y Kiran Sharma. 2022. "A Bibliometric Analysis to Unveil the Impact of Digital Object Identifiers (DOI) on Bibliometric Indicators". *Proceedings of Third International Conference on Computing, Communications, and Cyber-Security, Lecture Notes in Networks and Systems* 421: 859-869.  
[https://doi.org/10.1007/978-981-19-1142-2\\_67](https://doi.org/10.1007/978-981-19-1142-2_67)
- Krauskopf, Erwin. 2019. "Missing documents in Scopus: the case of the journal *Enfermeria Nefrologica*". *Scientometrics* 119(1): 543-547.  
<https://doi.org/10.1007/s11192-019-03040-z>
- Krauskopf, Erwin. 2021. "Article processing charge expenditure in Chile: The current situation". *Learned Publishing* 34(4): 637-646.  
<https://doi.org/10.1002/leap.1413>
- Larivière, Vincent, Éric Archambault, Yves Gingras y Étienne Vignola-Gagné. 2006. "The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities". *Journal of the American Society for Information Science and Technology* 57(8): 997-1004.  
<https://doi.org/10.1002/asi.20349>
- Liu, Jia. 2021. "Digital object identifier (DOI) and DOI services: An overview". *LIBRI* 71(4): 349-360.  
<https://doi.org/10.1515/libri-2020-0018>
- Liu, Chieh y Mu-Hsuan Huang. 2022. "Exploring the relationships between altmetric counts and citations of papers in different academic fields based on co-occurrence analysis". *Scientometrics* 127(8): 4939-4958.  
<https://doi.org/10.1007/s11192-022-04456-w>
- Liu, Weishu. 2020." Accuracy of funding information in Scopus: a comparative case study". *Scientometrics* 124(1): 803-811.  
<https://doi.org/10.1007/s11192-020-03458-w>
- Liu, Weishu, Meiting Huang y Haifeng Wang. 2021. "Same journal but different numbers of published records indexed in Scopus and Web of Science Core Collection: causes, consequences, and solutions". *Scientometrics* 126(5): 4541-4550.  
<https://doi.org/10.1007/s11192-021-03934-x>
- Marx, Werner y Lutz Bornmann. 2015. "On the causes of subject-specific citation rates in Web of Science". *Scientometrics* 102(2): 1823-1827.  
<https://doi.org/10.1007/s11192-014-1499-9>
- Mugnaini, Rogerio, Grischa Fraumann, Esteban F. Tuesta y Abel L. Packer. 2021. "Openness trends in Brazilian citation data: factors related to the use of DOIs". *Scientometrics* 126(3): 2523-2556.  
<https://doi.org/10.1007/s11192-020-03663-7>
- Manjunatha, G., B.T.S Kumar y H. Lakshmana. 2020. "Longevity of URL citations Cited in LIS journal articles: A Webometric Study". *Library Philosophy and Science* 2020: 3965
- Nagajara, Aragudije, Shine Joseph, S.A., Hyla Polen y Kevin Clauson. 2011. "Disappearing act: Persistence and attrition of uniform resource locators (URLs) in an open access medical journal". *Program Electronic Library and Information Systems* 45(1): 98-106.  
<https://doi.org/10.1108/00330331111107420>
- Pajic, Dejan, Tanja Jevremov and Marko Skoric. 2019. "Publication and citation patterns in the social sciences and humanities: A national perspective". *Canadian Journal of Sociology/Cahiers Canadiens de Sociologie* 44(1): 67-94.  
<https://doi.org/10.29173/cjs29214>

- Parnas, David Lorge. 2007. "Stop the numbers game". *Communications of the ACM* 50(11): 19-21.  
<https://doi.org/10.1145/1297797.1297815>
- Paskin, Norman. 2005. "Digital object identifiers for scientific data". *Data Science Journal* 4: 12-20.
- Purnell, Philip J. 2022. "The prevalence and impact of university affiliation discrepancies between four bibliographic databases—Scopus, Web of Science, Dimensions, and Microsoft Academic". *Quantitative Science Studies* 3(1): 99-121.  
[https://doi.org/10.1162/qss\\_a\\_00175](https://doi.org/10.1162/qss_a_00175)
- Sampath Kumar, B.T. y V. Vinay Kumar. 2013. "HTTP 404-page (not) found: Recovery of decayed URL citations". *Journal of Informetrics* 7(1):145–157.  
<https://doi.org/10.1016/j.joi.2012.09.007>
- Santiago-Padilla, Victor y Yalitza Therly Ramos-Gil. 2022. Indicators to achieve a better positioning of Ecuadorian scientific journals". *Revista Bionatura* 7(1): 2.  
<https://doi.org/10.21931/RB/2022.07.01.2>
- Sanz-Casado, Elias, Daniela De Filippo, Rafael Aleixandre Benavent, Vidar Roeggen y Janne Pölönen. 2021. "Impact and visibility of Norwegian, Finnish and Spanish journals in the fields of humanities". *Scientometrics* 126(11): 9031-9049.  
<https://doi.org/10.1007/s11192-021-044169-6>
- Savchenko, Igor y Denis Kosyakov. 2022. "Lost in affiliation: apatride publications in international databases". *Scientometrics* 127(6): 3471-3487.  
<https://doi.org/10.1007/s11192-022-04392-9>
- Shanthakumari, Keshava. 2021. "An analysis of persistence and obsolescence of web citations of pavement engineering literature". *Library and Philosophy Practice* 2021: 1-9
- Shuo, Xu, Liyuan Hao, Xin An, Dongsheng Zhai y Hongshen Pang. 2019. "Types of DOI errors of cited references in Web of Science with a cleaning method". *Scientometrics* 120(3): 1427-1437.  
<https://doi.org/10.1007/s11192-019-03162-4>
- Vinay Kumar, D., y V.T. Sampath Kumar. 2017. "Finding the unfound: Recovery of missing URLs through Internet Archive". *Annals of Library and Information Science*. 64(3): 165-171
- Wang, Weilang, Lvxiang Deng, Bin You, Ping Zhang y Yifeng Chen. 2018. "Digital object identifier and its use in core Chinese academic journals: A Chinese perspective". *Learned Publishing* 31: 19-154.  
<https://doi.org/10.1002/leap.1137>
- Zhu, Junwen; Guangyuan Hu y Weishu Liu. 2019. "DOI errors and possible solutions for Web of Science". *Scientometrics* 118(2): 709-718.  
<https://doi.org/10.1007/s11192-018-2980-7>

### Para citar este texto:

Krasuskopf, 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): 129-144.

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

# **Veinticinco años de investigación en redes sociales: evolución de temas entre 1997 y 2021 empleando el algoritmo Asignación Latente de Dirichlet**

Juan-Antonio Martínez-Comeche\*

*Artículo recibido:*  
15 de marzo de 2023  
*Artículo aceptado:*  
1 de junio de 2023

*Artículo de investigación*

## **RESUMEN**

El campo de las redes sociales ha sufrido importantes transformaciones en los últimos veinticinco años, en particular con la introducción de aplicaciones y plataformas digitales, así como la incorporación de estudios de otros campos del conocimiento que adoptan el enfoque de redes sociales en sus análisis. Este artículo ofrece una visión general de la evolución de los tópicos de investigación en este ámbito entre 1997 y 2021 a partir de la modelización de temas. El estudio parte de la producción académica que se recupera de la base de datos Scopus, considerando ventanas temporales de un año y utilizando el software Mallet. Se obtienen siete temas, cuya evolución en el tiempo se describe. Se concluye que los temas

\* Departamento de Biblioteconomía y Documentación, Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid, España  
juaamart@ucm.es



relacionados con los medios de comunicación social, así como las redes sociales en línea son estudiados con especial intensidad en los últimos años.

**Palabras clave:** Redes sociales; Evolución temática; SCOPUS; LDA; Medios de comunicación social

## **Twenty-five Years of Research in Social Networks: Evolution of Topics between 1997 and 2021 Based on Latent Dirichlet Allocation (LDA)**

*Juan-Antonio Martínez-Comeche*

### **ABSTRACT**

The social network field has suffered significant transformations in the last 25 years, particularly with the introduction of social networks online, as well as incorporated studies from many other knowledge fields that adopt the social network approach in their analyses. This paper offers an overview of the evolution of research topics in this field between 1997 and 2021 based on topic modeling. The methodology used draws from the Scopus database, considering time windows of a year and using the software, Mallet. Seven topics are obtained, whose evolution over time is described. It is concluded that the topics related to social media and social networks online have been studied with special intensity in the last years.

**Keywords:** Social networks; Topic evolution; SCOPUS; LDA; Social media

### **INTRODUCCIÓN**

Las Ciencias Sociales se interesan por el estudio de grupos sociales, concebidos en esencia como un conjunto de entidades que interrelacionan entre sí. En la década de 1930 Jennings y Moreno introducen los primeros sociogramas (Moreno y Jennings 1938), gráficos que muestran la red de relaciones existentes entre las unidades y la estructura social del grupo, como el de la *figura 1*.

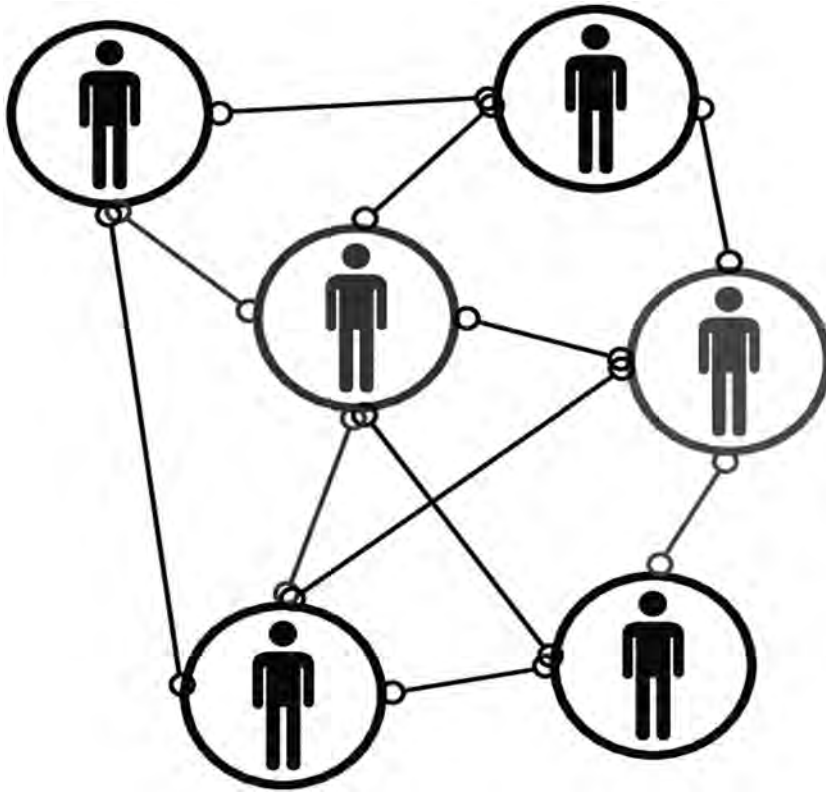


Figura 1: Estructura de relaciones de una red social  
Fuente: Tomado de Ricci (2018).

En esta figura se muestra la estructura de una red social compuesta por un grupo de unidades o nodos (personas en este caso) que se relacionan entre sí por compartir un cierto objetivo común. Los vínculos (representados por líneas en la figura) expresan en sus extremos los nodos o unidades que se relacionan directamente. Entre estos se establece un flujo de información que es unidireccional o bidireccional, como en el caso de la figura, indicando que el flujo informativo se establece en ambos sentidos.

Moreno constituye uno de los pioneros que aborda un enfoque cuantitativo para describir la organización de relaciones y los roles de los individuos en un grupo, aportando la perspectiva de la Sociometría, bajo la influencia de la teoría psicológica de la Gestalt (Moreno 1937). Por su parte, Heider (1946) desarrolla la teoría de los grupos dinámicos que, junto a la teoría de grafos desarrollada por Harary (1969), representan los pilares del denominado análisis de redes sociales (Wasserman y Faust 1994).

A finales de 1970 la búsqueda de patrones en las interacciones entre individuos empieza a reconocerse como una especialidad científica emergente basada en la teoría de grafos y en modelos estadísticos y computacionales, iniciándose así la unificación de este campo. Tres hechos contribuyen a unir las hasta entonces dispersas aproximaciones: la fundación del primer boletín (*Connections*) por Barry Wellman (Freeman 2004), la publicación de los primeros textos con una perspectiva estándar (Berkowitz 1982; Knoke y Kuklinski 1982), y el desarrollo de los primeros programas informáticos, SONIS y UCINET (Pappi y Stelck 1987; Freeman 1988).

Muchas áreas diversas entre sí de conocimiento emplean el enfoque de redes sociales en sus análisis: desde la zoología (Sueur y Pelé 2016; Zanardo *et al.* 2018) hasta la química (Chodera y Pande 2011; Gore *et al.* 2021), pasando por la informática (Wang y Robinson 2002) y la tecnología de las comunicaciones (Zhang *et al.* 2009). Las disciplinas que emplean este enfoque se encuentran en constante transformación, a modo de ejemplo: la economía, la medicina, la geografía, las ciencias de la información, la educación o las redes de transportes (Otte y Rousseau 2002). Las comunidades virtuales se incorporan a este campo como un dominio nuevo, concebido como grupos de individuos que comparten intereses e intercambian entre sí contenidos generados por ellos mismos a través de aplicaciones informáticas basadas en la web 2.0 (Ridings *et al.* 2002).

Con el objeto de evitar confusiones terminológicas, en este trabajo se emplea «medios de comunicación social» para designar cualquier plataforma o servicio en línea utilizados para crear comunidades virtuales, desde aplicaciones de mensajería instantánea (WhatsApp), proyectos colaborativos (Wikipedia) e intercambio de fotos o vídeos (Flickr, YouTube), hasta redes sociales (Facebook) o incluso mundos virtuales como Second Life, dada su amplia significación que posee en la actualidad (Aichner y Jacob 2015; Statista *et al.* 2022). Se utiliza «redes sociales en línea» para especificar servicios basados en la web que permiten a las personas estructurar y hacer públicas o semi públicas sus redes sociales (Boyd y Ellison 2008), y «aplicaciones de mensajería instantánea» para designar servicios que permiten a los usuarios intercambiar contenidos privados entre dos o más miembros específicos y conocidos de un grupo (Taipale y Farinosi 2018).

## ANTECEDENTES

La evolución temática dentro de un campo se puede abordar desde dos enfoques principales: modelos bibliométricos y modelización de temas (Song *et al.* 2014). A su vez, los modelos bibliométricos se basan en diversos elementos como las palabras, los autores, las revistas o las citas en un corpus de textos (Han 2020; Onyancha 2018; Chang *et al.* 2015).

Los autores de los textos son de gran utilidad para descubrir la evolución temática de un área de conocimiento, considerando un tema como el conjunto de intereses comunes de un grupo de autores (Jung y Yoon 2020). Por su parte, el análisis de las revistas donde se publican los artículos conforma una técnica ampliamente utilizada con esta finalidad, mediante el estudio detallado de contenido con respecto a los tópicos abordados en una selección de revistas del área (Tuomaala *et al.* 2014; Armann-Keown y Patterson 2020).

Cuando se opta por las palabras como los elementos de los documentos que permiten dar a conocer los temas presentes en un corpus y su evolución temporal, se emplean diversas técnicas, desde las palabras clave más frecuentes hasta su aparición simultánea o su distribución por temas o en ciertos ciclos de tiempo (Onyancha 2018; Peset *et al.* 2020).

Las citas, por su parte, se utilizan ampliamente para observar diversos aspectos como la estructura intelectual o la evolución temática, de manera habitual mediante análisis de co-citación (Chen 2006; Chabowski y Samiee 2023) o acoplamiento bibliográfico (Yanhui *et al.* 2021). En ocasiones se ha superpuesto el análisis de redes sociales a dichos métodos bibliométricos (Yang *et al.* 2012). Con ayuda de cualquiera de estos elementos se dan a conocer diversos indicadores de la actividad científica y su evolución temporal en relación con la productividad, la colaboración, la circulación y el impacto, entre otros aspectos.

Sin embargo, las técnicas bibliométricas no son las más apropiadas cuando el volumen de datos de entrada aparece elevado. En circunstancias semejantes a este estudio —donde el número de documentos analizados ronda los 150 000—, resulta más efectivo el empleo de técnicas de modelización de temas (Jeong y Min 2014; Yau *et al.* 2014; Jung y Yoon 2020), basadas en un análisis estadístico de palabras en el corpus, el cual permite reunir las en pocos grupos que constituyen los temas del área (Liu *et al.* 2020). No existe técnica única de modelado para la detección de temas. Unas se basan en las relaciones sintácticas entre las palabras del corpus (Ferrer *et al.* 2004), otras en las de semánticas entre las palabras (Arruda *et al.* 2016), e incluso algunas aportan el empleo de redes neuronales a estos modelos lingüísticos (Mikolov *et al.* 2013).

Sin embargo, las técnicas de modelado más utilizadas en el análisis de corpus grandes de carácter textual, como Asignación Latente de Dirichlet (LDA) (Blei *et al.* 2003), Análisis Probabilístico de Semántica Latente (PLSA) (Hofmann 1999) o Análisis de Semántica Latente (LSA) (Landauer *et al.* 2007), consideran los temas como grupos de palabras y se basan en procedimientos estadísticos de simple co-aparición de estas para su identificación. Aquí se empleó el modelo LDA porque ha mostrado ser un método efectivo en el descubrimiento de la estructura temática subyacente en corpus de temática dispersa en áreas de conocimiento distintas, como la de las redes sociales (Yau *et al.* 2014; Suominen y Toivanen 2015; Jeong y Min 2014; Banerjee y Basu 2007).

El algoritmo LDA aborda el descubrimiento de temas en un corpus a partir de las siguientes premisas: cada documento se representa mediante la co-ocurrencia de un conjunto de temas; cada tema se presenta a través de un conjunto de palabras co-ocurrentes. El propósito del algoritmo es generar un conjunto de palabras de alta probabilidad de ocurrencia para cada tema, expresión de una estructura semántica subyacente en el corpus más allá de palabras concretas empleadas en los documentos (Deerwester *et al.* 1990), de donde la inclusión de la palabra latente en la denominación del algoritmo (Wu *et al.* 2014).

Este procedimiento constituye un modelo de inferencia bayesiano en cuanto que genera esencialmente dos listados de probabilidades (Shen y Wang 2020): uno con las de aparición de todos los temas en cada documento del corpus y otro con las de aparición de todas las palabras en cada tema (Li y Lei 2021; Zhu *et al.* 2016; Kai *et al.* 2019). Además, constituye un algoritmo de aprendizaje no supervisado (Tdk Technologies 2020) en cuanto que la repetición de los cálculos de probabilidades de las palabras en grupos mejora los resultados en cada iteración sucesiva. Debe considerarse, sin embargo, la relación entre el tiempo invertido en un número alto de iteraciones y la mejora relativa en la precisión del modelo resultante.

De las varias implementaciones disponibles para ejecutar LDA, se optó por el paquete basado en Java Mallet (McCallum 2022a), conocido programa de código abierto para modelización de temas.

El análisis de la evolución de los temas se considera desde dos perspectivas: el desarrollo del contenido, que implica tanto cambios en las palabras distintivas de un tema como el aumento y disminución de su frecuencia con el tiempo; y el de la intensidad temática, que muestra cambios en la atención prestada por los investigadores al tema y la consiguiente mayor o menor presencia de este en los documentos del corpus (Shan y Li, 2010; Zhu *et al.* 2016). Ambos análisis se complementan y ofrecen un conocimiento profundo del desarrollo de los tópicos. En este estudio se aborda tanto la evolución del contenido como de la intensidad temática a partir del sometimiento del corpus de documentos al algoritmo LDA (Griffiths y Steyvers 2004).

## OBJETIVOS

El principal objetivo de este estudio radica en descubrir y analizar los principales temas de investigación en redes sociales y su evolución en contenidos e intensidad en los recientes 25 años (1997-2021). Así, la intensidad alude a la variación con el tiempo de la importancia relativa de los distintos temas en el corpus. En dicho análisis sobre las redes sociales se considera su naturaleza multidisciplinar

y la constante incorporación de nuevos dominios a este campo, entre los que destacan aquellas que se producen en línea, cuya primera aplicación (Six Degrees) aparece en 1997. Hasta donde se sabe, este tema no ha sido previamente abordado. La hipótesis parte de que el descubrimiento de los temas y su evolución en este período resulta posible mediante técnicas de modelización de temas, en concreto Latent Dirichlet Allocation (LDA) o Asignación Latente de Dirichlet, dado el amplio volumen de datos de entrada que dificulta la aplicación de métodos bibliométricos.

El segundo objetivo consiste en alcanzar una descripción detallada y al tiempo coherente desde el punto de vista temático. Existen técnicas de modelado de temas, como hLDA (Hierarchical Latent Dirichlet Allocation), que añade a la detección de estos su evolución temporal y la relación jerárquica entre ellos (Song *et al.* 2016). Sin embargo, se ha puesto de manifiesto que el resultado de estas técnicas, en principio más completas, suele ser de difícil interpretación (Ding y Chen 2014). Para superar el inconveniente, se opta por efectuar de manera independiente cada uno de los procesos necesarios, en vez de aplicar un solo programa o técnica, y poner en práctica la triangulación metodológica mediante el análisis textual de una muestra de los datos de entrada. Con ello se asegura la coherencia temática de los resultados, al tiempo que se detecta qué parámetros son decisivos para generar una descripción temáticamente congruente.

La originalidad del trabajo reside, además del campo objeto de estudio: las redes sociales en su conjunto —no únicamente las aplicaciones informáticas de éstas en línea—, también en dos aspectos metodológicos: uno, la aplicación de un enfoque cualitativo (análisis textual de 1 700 documentos) superpuesta a la técnica cuantitativa propia de la modelización de temas mediante LDA; dos, la aplicación de técnicas del Procesamiento del Lenguaje Natural (análisis morfológico, léxico y de reconocimiento de entidades) y de técnicas empleadas en la Recuperación de Información (ponderación TF.IDF o la similitud coseno) para mejorar los resultados de la modelización de temas mediante LDA.

## MÉTODOS Y MATERIALES

En este estudio se utilizó una metodología mixta, cuantitativa y cualitativa, para alcanzar el objetivo propuesto. Las técnicas cuantitativas incluyeron el desarrollo de diversos programas en Java y distintos scripts en lenguaje R para realizar cada una de las tareas de naturaleza estadística involucradas en el estudio, dado el alto volumen de documentación gestionada, así como el empleo del paquete Mallet para efectuar el proceso de modelización de temas, también de carácter estadístico. La técnica cualitativa involucró el análisis textual de una muestra de

entre 10 y 20 documentos más destacados de cada uno de los 155 temas generados de manera inicial por el programa Mallet, lo que supuso alrededor de 1 700 documentos, esto es, algo más del uno por ciento del total de documentación manejada en la investigación.

En este apartado se explican en detalle los pasos seguidos en el estudio y las técnicas empleadas en cada uno de ellos, cuyo esquema se muestra en la *figura 2*.

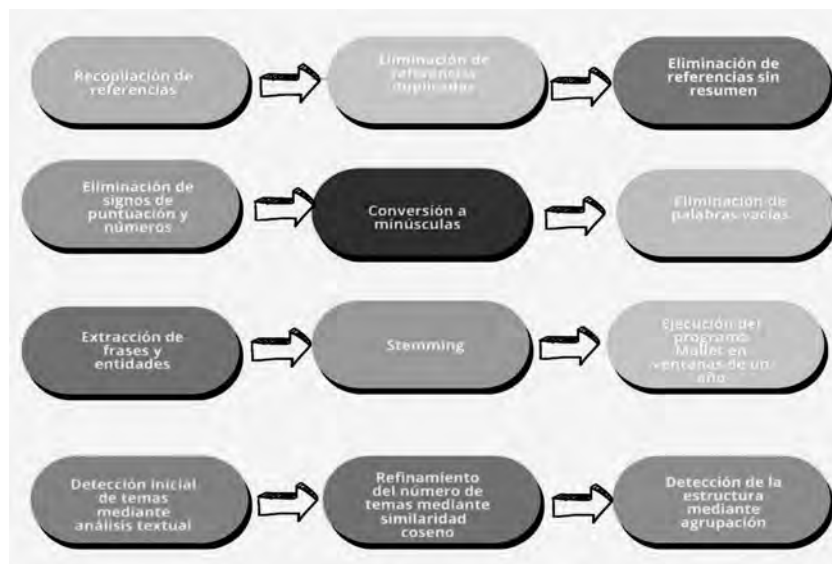


Figura 2: Esquema del proceso de análisis efectuado

Fuente: Elaboración propia.

Para la recopilación de los documentos sobre redes sociales de entre 1997 y 2021 se utilizó la base de datos Scopus, creada en 2004, publicada por Elsevier, siendo una de las bases con mayor volumen de documentación que abarca todas las ramas del conocimiento, empleando la consulta:

TITLE(«social network\*») OR KEY(«social network\*») OR TITLE(«instant messag\*») OR KEY(«instant messag\*») OR TITLE(«facebook») OR KEY(«facebook») OR TITLE(«myspace») OR KEY(«myspace») OR TITLE(«cyworld») OR KEY(«cyworld») OR TITLE(«twitter») OR KEY(«twitter») OR TITLE(«qzone») OR KEY(«qzone») OR TITLE(«instagram») OR KEY(«instagram») OR TITLE(«ask.fm») OR KEY(«ask.fm») OR TITLE(«tiktok») OR KEY(«tiktok») OR TITLE(«kuaishow») OR KEY(«kuaishow») OR TITLE(«tumblr») OR KEY(«tumblr») OR TITLE(«whatsapp») OR KEY(«whatsapp») OR TITLE(«msn messenger») OR KEY(«msn messenger») OR TITLE(«wechat») OR KEY(«wechat») OR TITLE(«telegram») OR KEY(«telegram») OR TITLE(«tencent qq»)

```
OR KEY(«tencent qq») OR TITLE(«snapchat») OR KEY(«snapchat») OR TITLE(«hangouts») OR KEY(«hangouts») OR TITLE(«skype») OR KEY(«skype») OR TITLE(«viber») OR KEY(«viber») OR TITLE(«kakaotalk») OR KEY(«kakaotalk») AND PUBYEAR AFT 1996 AND PUBYEAR BEF 2022 AND LANGUAGE(english) AND (DOCTYPE(ar) OR DOCTYPE(bk) OR DOCTYPE(ch) OR DOCTYPE(cp) OR DOCTYPE(re) OR DOCTYPE(sh))
```

En esta consulta se limitó la lengua de los documentos al inglés, pues la presencia de dos lenguas complicaría en exceso el desarrollo de los programas informáticos empleados en el proceso seguido, además el volumen que corresponde a los de inglés es superior al español, así como el tipo de documento a las vías principales de comunicación de la investigación científica, esencialmente: artículos, libros, capítulos de libro y ponencias. En la consulta no se incluyó la aparición de las palabras en el Abstract porque se captaron documentos que utilizaban las redes sociales como metodología de obtención de datos, pero no analizaban las redes sociales como tales. Se obtuvieron un total de 154 582 documentos.

Antes de aplicar el algoritmo de modelización de temas fue necesario preparar adecuadamente los textos a someter al proceso. Las sucesivas fases en que constó el preprocesamiento de documentos tras la recopilación de datos fueron: 1) Eliminación de las referencias duplicadas; 2) Así también, de las referencias que careciesen de resumen; 3) Por igual, de los signos de puntuación y los números, convirtiendo las mayúsculas a minúsculas, mediante el analizador SimpleAnalyzer incorporado a la biblioteca Java de código abierto Lucene Core (Apache Software Foundation 2022; McCandless *et al.* 2010); 4) De manera similar, de las palabras vacías que constasen de tres o más caracteres.

Además, para extraer la frecuencia de aparición de cada palabra en la muestra se utilizó el programa en Java desarrollado por Billy Joel Johnson (Johnson 2019). En la lista de palabras vacías se incluyeron solamente las 101 palabras con valor funcional o sintáctico que figuraban al menos en uno de los siguientes listados de palabras vacías en inglés: LingPipe (Alias-i 2016); la lista compilada por Kevin Bougé para el sistema de recuperación de información SMART (Brigadir 2022); la lista incorporada a la biblioteca NLTK (NLTK 2022); y la lista original de Snowball creada por Martin Porter (Snowball 2021). La eliminación de las palabras con 1 ó 2 caracteres se realizó posteriormente, pues el programa Mallet lleva a cabo esta tarea por defecto al preparar la colección antes de someterlo al proceso de detección de temas; 5) Extracción de frases y entidades. Para ello se emplearon todas las palabras clave del corpus compuestas de dos o más palabras (que incluyen también entidades geográficas, personales o temporales), convirtiéndolas en unitérminos (por ejemplo, *ecommerce*); 6) Stemming o reducción morfológica. Para lo cual se empleó el programa Snowball incorporado al proyecto OpenNLP, que



desarrolla el algoritmo Porter2 para el inglés (OpenNLP 2022); 7). Por último, importación de datos en Mallet, que por defecto elimina cualquier palabra de 1 ó 2 caracteres (Mallet 2022). El programa Mallet es un paquete basado en Java para el procesamiento estadístico del lenguaje natural, la clasificación de documentos, la agrupación o clustering y la extracción de información. Esta aplicación resultó de fácil instalación y manejo, siendo utilizada de modo amplio en el campo del aprendizaje máquina.

Dadas las peculiaridades de la colección durante el periodo que se analizó, se optó por considerar ventanas temporales de un año de duración, cada una de ellas acogiendo los documentos publicados el mismo año (Wu *et al.* 2014). Con ello se pretendió mejorar la adaptabilidad del algoritmo a los cambios que se han ido produciendo en estos años en un área tan dinámica como las redes sociales.

La aplicación del algoritmo LDA requirió la fijación previa de varios parámetros, además del número de periodos en que se dividió el corpus de entrada. El más importante fue el número de temas. Aunque requirió mucho más tiempo y carga de trabajo, no se basó en el cálculo de la perplejidad (Griffiths y Steyvers 2004), el consejo de expertos o en estudios previos para decidir un número aproximado que haya resultado de utilidad en áreas afines (McCallum *et al.* 2007), sino que se adoptaron los criterios que figuran en la propia documentación del programa Mallet (McCallum 2022b; Isasi 2022; Graham *et al.* 2021; Kumar 2022; Silge 2018). Así, tras aplicar el algoritmo LDA con varios números de temas, en una primera fase se seleccionaron los valores de temas con mejores rangos y valores medios de coherencia (*coherence*) y exclusividad (*exclusivity*); de entre ellos, en una segunda fase, los valores de temas con mejores rangos y valores medios del parámetro alpha; en una tercera y última fase, se eligieron para cada periodo anual el número de temas que presentó una mejor distribución de tales temas en la documentación correspondiente. En cuanto a otros parámetros, se fijaron 1 000 iteraciones para el desarrollo del modelo y 15 palabras destacadas por tema.

Una vez ejecutado el programa Mallet con cada ventana temporal de datos y obtenidos los temas correspondientes a cada año, el problema fue comprobar cómo se relacionaban los temas de años sucesivos. Este proceso de correlación de temas (Liu *et al.* 2020) se llevó a cabo en dos fases: primero se efectuó el análisis textual de las palabras destacadas, títulos y resúmenes de al menos los 10 documentos más relevantes señalados por Mallet de cada tema anual, llegando en algunos casos hasta los 20 documentos más destacados hasta discernir con claridad a qué tema adscribir cada tema anual. En caso de advertir cualquier peculiaridad temática, se optó por considerarlo de manera inicial un tema distinto. Ello supuso el análisis de unos 1 600 documentos, en torno al 1 % del volumen total de la colección considerada en el estudio. En esta primera fase se detectaron 20 temas.

A continuación, se procedió a refinar el número de temas y sus temas anuales componentes, mediante el cálculo de la similaridad coseno (Zhu *et al.* 2016). La similaridad coseno representó un procedimiento ampliamente utilizado en el área de la Recuperación de Información que calcula la similitud entre dos grupos de términos tomando en cuenta el número de términos comunes en ambos grupos y el peso o importancia de cada término para representar el grupo al que pertenece. La similaridad coseno más baja posible corresponde a cero, y la más alta a uno, en cuyo caso los grupos de términos fueron idénticos en número y representatividad.

Se partió de los resultados que proporciona Mallet para cada colección anual, en concreto, de la probabilidad de aparición de los términos de dicho año en cada uno de los temas generados en la mencionada anualidad. Considerando dicha probabilidad, como el peso de cada término en cada tema anual, y partiendo de los temas anuales del primer año 1997 como los temas iniciales, se procedió a calcular la similaridad entre cada uno de los temas anuales de los años sucesivos con todos los temas anteriores.

Para considerar un cierto tema anual, como el primero de un nuevo tema independiente, se impusieron las siguientes condiciones: en primer lugar, presentar una similaridad  $< 0,76$  con los últimos temas anuales que componen cualquier tema detectado de manera previa; en segundo, poseer una temática distintiva y claramente definida, ajena al resto de temas obtenidos hasta ese momento, mediante el análisis textual de los 20 primeros documentos más relevantes del tema candidato; en tercero, la imposibilidad de crear un tema compuesto de forma exclusiva por un tema único anual; por último, el segundo tema anual del nuevo debe presentar una similaridad  $\geq 0,76$  con el primer tema anual del nuevo tema. En esta segunda fase los 20 temas iniciales se redujeron a 10.

Con la finalidad de detectar posibles relaciones entre los 10 temas, mejorando la precisión del análisis de la evolución temática, se examinó su estructura temática sometiéndolos a agrupación jerárquica. El procedimiento consistió en imponer un peso a los términos que aparecen en cualquiera de los temas anuales componentes de un tema, empleando ponderación TF.IDF, que consiste en imponer a cada término un número representativo de su importancia en el grupo de términos al que pertenece. Este número, denominado peso del término, se calcula multiplicando dos cantidades: una corresponde a la frecuencia de aparición del término en el conjunto (TF del inglés *Term Frequency*), y otra el inverso del número de grupos en los que aparece el término, indicativo de su valor para discriminar si el término surge en muchos o en pocos grupos del corpus (IDF del inglés *Inverse Document Frequency*). El cálculo se realizó de la siguiente manera:

1. En cuanto a TF (*Term Frequency*), cada vez que un término aparece en un tema anual, se suma +1 a su frecuencia. Para paliar el sesgo producido por temas con muchos temas anuales frente a otros con muy pocos temas anuales, se adoptó la siguiente fórmula para el cálculo del TF (Singhal 2001):

$$TF(\text{term } t \text{ in great topic } d) = 1 + \log(1 + \log(tf_{t,d}))$$

2. En cuanto a IDF (*Inverse Document Frequency*), se adoptó la siguiente fórmula para su cálculo, que evita penalizar en exceso los términos que aparecen en bastantes temas anuales de la colección en su conjunto (Singhal 2001):

$$IDF(\text{term } t) = \log \frac{(N + 1)}{n}$$

donde N es el número total de temas anuales en la colección (155), y donde n es el número total de temas anuales en la colección en los que aparece el término t. A continuación, se calcula la similaridad coseno entre cada pareja de los 10 temas y la matriz resultante se somete al algoritmo hclust del lenguaje R (R Core Team 2022).

Una vez observada la evolución temporal del contenido de los temas, el panorama general de la investigación en redes sociales se complementó con el análisis de la importancia relativa concedida por los investigadores a cada tema frente a los demás, aspecto denominado intensidad temática (Zhu *et al.* 2016) o popularidad temática (Zou 2018).

Se han utilizado diversos procedimientos para comparar la relevancia de cada tema en un campo de investigación. En ocasiones se han empleado palabras determinantes que muestran tendencias (Ding *et al.* 2001; Kleinberg 2003). Otras alternativas consisten en comparar el número de citas recibidas por cada tema (Ha *et al.* 2014), o utilizar modelos de aprendizaje máquina (Xu *et al.* 2019).

Como el algoritmo LDA contempla múltiples temas en un mismo documento, para el cálculo de la popularidad o intensidad temática se encontró la probabilidad media de cada tema en todos los documentos de cada año, distinguiéndose entre temas cuya intensidad resultó creciente, estable o decreciente. El aumento o disminución de popularidad de un tema se evaluó mediante el test no paramétrico de tendencia Mann-Kendall (Pohlert 2022), con un valor de  $p < 0,001$  para considerar que fue estadísticamente significativo. Se incluyeron las gráficas de los temas con una intensidad creciente o decreciente.

## RESULTADOS Y DISCUSIÓN

De los 154 582 documentos iniciales obtenidos en Scopus, el preprocesamiento los redujo a 149 443 tratados finalmente por Mallet. Los cálculos para decidir el número de temas más adecuado en cada año supusieron el desarrollo de un total de 155 temas de alcance anual a lo largo de los 25 años del estudio. La *tabla 1* muestra el número final de documentos procesados y el número óptimo de temas para cada año.

Año	Número de temas	Número de documentos	Año	Número de temas	Número de documentos
1997	2	104	2010	7	5 226
1998	2	122	2011	7	6 487
1999	2	134	2012	7	7 997
2000	2	161	2013	8	9 520
2001	4	194	2014	8	11 740
2002	4	295	2015	8	13 841
2003	6	634	2016	8	12 086
2004	6	601	2017	8	12 088
2005	6	894	2018	8	13 404
2006	7	1 176	2019	8	14 525
2007	7	1 768	2020	8	15 345
2008	7	2 558	2021	8	14 825
2009	7	3 718	TOTAL	155	14 9443

*Tabla 1:* Número de documentos y temas desarrollados por Mallet para cada año  
Fuente: Elaboración propia.

Someter los 155 temas anuales a un proceso de agrupación mediante la similitud coseno para observar cómo se relacionan temáticamente entre ellos a lo largo del tiempo requiere la imposición de un valor límite. Si no se impone un tope que impida la unión de dos temas, todos los temas anuales terminarían uniéndose en uno solo que abarcaría todo el corpus. En el caso extremo, si se impone un límite de similitud muy alto para que dos temas anuales puedan unirse en un único tema, se tendría un número de temas cercano o igual a los 155 anuales de partida.

Para superar este inconveniente, se procedió en primer lugar a un análisis textual de los 155 temas anuales, con el objetivo de conocer el número máximo

de estos que deberían considerarse en el corpus y sus características. Luego se procedió a calcular, en sucesivos pasos, los valores de similitud coseno entre los temas anuales siguiendo el procedimiento expuesto en el apartado de Métodos y Materiales, poniendo especial cuidado ahora en no unir dos temas anuales cuando el análisis textual previo hubiese indicado que se trata de dos temas dispares que deben permanecer separados en grupos distintos.

De este proceso resultó la conveniencia de imponer un valor límite de 0,76, porque si se admitía la unión de temas con valores de similitud de 0,74 y 0,75, cuatro de ellos tendrían que unirse en dos (topic A y topic E, por una parte; topic J y topic T, por otra, respectivamente), presentando los cuatro una temática claramente independiente. Como resultado de esta segunda fase, los 20 temas que fijaba como máximo el análisis textual, quedaron reducidos a los 10 que figuran en la *tabla 2*.

Tema	Número de temas anuales	Similitud coseno media	Años de aparición	Años sin aparecer
A	25	0.913	1997-2021	--
B	7	0.747	1997-2004	2001
C	15	0.780	2001-2021	2005-2012 2016
D	29	0.749	2004-2021	--
E	23	0.865	2001-2021	--
I	10	0.868	2003-2012	--
J	16	0.878	2005-2020	--
K	20	0.836	2007-2021	--
M	3	0.511	2005-2008	2007
T	7	0.918	2016-2021	--

*Tabla 2:* Temas tras agrupación por similitud coseno y años de aparición  
Fuente: Elaboración propia.

Mallet proporciona una serie de palabras destacadas por cada uno de los temas anuales, a raíz de los cuales se deben escoger los descriptores adecuados para cada uno de ellos. El procedimiento seguido para seleccionar estos descriptores es el siguiente: primeramente se reúnen todas las palabras destacadas que proporciona Mallet de cada uno de los temas anuales que componen cada tema y se calcula su frecuencia; a continuación, se calcula el IDF de cada uno de dichos términos empleando la misma fórmula que se ha usado en el cálculo de la estructura temática, descrita en el apartado Métodos y Materiales; finalmente se calcula el

peso TF.IDF de cada uno de los términos. Los términos seleccionados finalmente para cada tema incluyen los que poseen un peso TF.IDF mayor, a los que se añaden aquellos considerados representativos del contenido del tema conforme al análisis textual efectuado, pero que se hallan en un rango de frecuencias media o baja. En la *tabla 3* se resumen los términos seleccionados para cada tema de entre los obtenidos por Mallet, junto con una breve descripción en inglés de su contenido con base al análisis textual efectuado.

Temas	Términos	Descripción textual
A	health, social network, age, women, relation, social support, patient, family, friend, network, hiv, sexual, behavior, parent, contact	Documents about health, focusing on the social network (e.g., family, friends) of the patients. Factors as age, gender or race are usually considered.  [Documentos sobre salud, centrándose en las redes sociales –familia, amigos– de los pacientes. Factores como la edad, el género o la raza son considerados habitualmente.]
B	develop, online, network, community, inform, support, structure, work, migrate, instant messaging, converse, email, collaborate, economics, learn	Documents about diverse online networks (e.g., students, workers) and physical communities (for political, religious or migration support), discussing their structure or information diffusion, and announcing the development of the first instant messaging and social network online applications.  [Documentos sobre diversas redes en línea (por ejemplo, estudiantes, trabajadores) y comunidades físicas (de apoyo político, religioso o a la emigración), analizando su estructura o la difusión de información en ellas, y anunciando el desarrollo de las primeras aplicaciones de mensajería instantánea y de redes sociales en línea.]
C	web, system, design, application, data, privacy, security, social network, algorithm, process, big data, social media, semantic web, interface, mobile	Documents on technical aspects of the design of web-based applications, algorithms related to privacy, security, and integrity of information, efficient techniques for data storage, processing, and sharing, software for mobile communication systems, and their use in social media and online social networks.  [Documentos sobre aspectos técnicos del diseño de aplicaciones basadas en la web, algoritmos relativos a la privacidad, seguridad e integridad de la información, técnicas eficientes para el almacenamiento, procesamiento y compartición de datos, programas para sistemas móviles de comunicación y su empleo en medios de comunicación social y en redes sociales en línea.]

D	tweet, Twitter, node, algorithm, graph, sentiment, event, detect, recommend, trust, network analysis, privacy, semantic web, Facebook, evolution	<p>Documents on graphs, semantic web, social networks, and social media, addressing issues related to the structure of networks, their attributes (e.g., influential nodes, trust relationships), their evolution over time, and multiple aspects related to social media (e.g., user interests, recommendation algorithms, trends, sentiment analysis, fake news detection or information extraction).</p> <p>[Documentos sobre grafos, web semántica, redes sociales y medios de comunicación social, abordando aspectos relativos a la estructura de las redes, sus atributos (el concepto de influencia y relaciones de confianza, por ejemplo) o su evolución temporal, además de múltiples aspectos relacionados con los medios de comunicación social (intereses de los usuarios, algoritmos de recomendación, tendencias, análisis de sentimientos, detección de bulos o extracción de información.)]</p>
E	network, social, social network, community, group, politics, social network analysis, business, culture, economics, govern, immigrate, policy, social capital, science	<p>Documents on economy, social capital, market, communities, and groups (e.g., immigrants, scientists), governance and policies, social, political, and cultural movements, all aspects analyzed from a social network perspective.</p> <p>[Documentos sobre economía, capital social, mercado, comunidades y grupos (emigrantes, científicos), gobernanza y políticas, movimientos sociales, políticos y culturales, analizados desde la perspectiva de las redes sociales.]</p>
I	instant messaging, implement, application, base, service, platform, message, communicate, mobile, network, protocol, architecture, server, distribute, standard	<p>Documents on text, voice, video and image processing and communication, protocols and mechanisms related to encryption, authentication, and access for mobile or wireless networks, and implementation of instant messaging and chat services.</p> <p>[Documentos sobre procesamiento y comunicación de texto, voz, vídeos e imágenes, protocolos y mecanismos de encriptación, autenticación y acceso en redes de móviles o sin cable, así como implementación de servicios de mensajería instantánea y de chat.]</p>
J	online, course, learn, student, education, e-learn, social media, technology, virtual, teacher, academic, computer, skill, web, chat	<p>Documents on e-learning, collaborative learning, and technology use in education, with special attention to virtual learning communities and the use of social media and social networks online in learning.</p> <p>[Documentos sobre aprendizaje en línea o virtual, aprendizaje colaborativo y la utilización de la tecnología en la educación, prestando especial atención a las comunidades virtuales de aprendizaje y al empleo de medios de comunicación social y de redes sociales en línea en la enseñanza.]</p>

K	social network, online, social media, consume, use, privacy, behavior, factor, market, tag, weblog, tweet, game, student, public	<p>Documents about social networks online (e.g., Facebook, Twitter) and social media (e.g., YouTube, Digg, Flickr), the user behavior and tagging patterns, with special emphasis on privacy issues and factors affecting consume and purchase decision related to brands accounts.</p> <p>[Documentos sobre redes sociales en línea -Facebook, Twitter, por ejemplo- y medios de comunicación social -como YouTube, Digg, Flickr-, el comportamiento del usuario y patrones de etiquetado, con énfasis en la privacidad y en los factores que afectan al consumo y decisión de compra de marcas comerciales.]</p>
M	immigrate, migrate, culture, politics, local, community, network, economics, social, policy, Irish, refugee, China, ethnic, transnational	<p>Documents about networks and communities of various kinds (immigrants, refugees, ethnic groups, organized crime or students), showing local cultural, political, economic or social aspects (in the host country) and those of the country of origin.</p> <p>[Documentos sobre redes y comunidades diversas (emigrantes, refugiados, grupos étnicos, crimen organizado o estudiantes), mostrando aspectos locales de la cultura, la política, la economía y de carácter social en el país de acogida frente a estos aspectos en el país de origen.]</p>
T	Twitter, social media, discuss, politics, public, communicate, social, inform, engage, active, covid pandemic, pandemic, vaccine, practice, message	<p>Documents on the use of social networks online and social media to disseminate political and religious propaganda, public information campaigns (e.g., gun control or health advice), and as a stage for protest movements (e.g., sexual harassment) and discussions on relevant news.</p> <p>[Documentos sobre la utilización de redes sociales en línea y medios de comunicación social para la propaganda política, religiosa o las campañas informativas de carácter público (control de armas o avisos sanitarios, por ejemplo), así como su empleo como plataformas para movimientos de protesta (sobre acoso sexual, por ejemplo) y como foros de discusión sobre noticias relevantes.]</p>

Tabla 3: Términos obtenidos por Mallet y descripción textual de los tema

Fuente: Elaboración propia.

A fin de disponer de más información sobre las características de los temas, en especial para detectar posibles relaciones entre ellos, de ayuda en el análisis de la evolución temática, se examinó su estructura sometiendo a agrupación jerárquica. Siguiendo el procedimiento descrito en la sección anterior, se obtiene la estructura temática que muestra la *figura 3*.



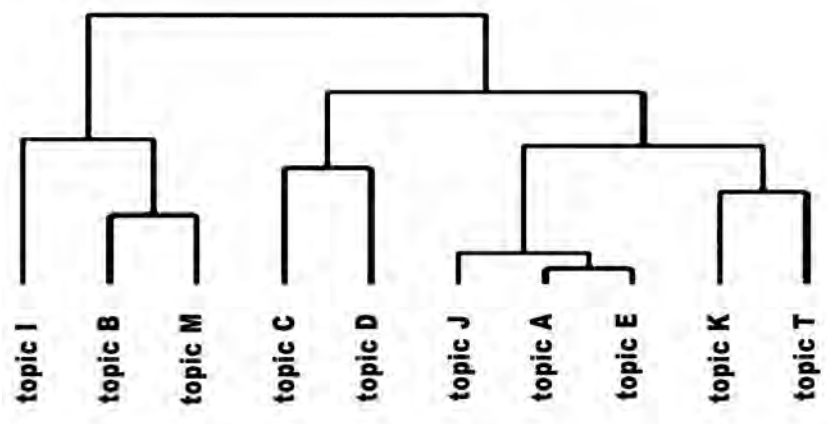


Figura 3. Estructura temática de los temas mediante agrupación jerárquica  
Fuente: Elaboración propia.

En esta gráfica se observa una clara separación entre los temas B, M e I frente al resto. Si se advierte en el grupo compuesto por los de B y M, se ve que comparten varias características: son los temas con un menor número de temas anuales; a excepción del de T, los temas con una menor duración temporal; además, los temas con un menor valor de similaridad media. Como entre 1997 y 2000 —los cuatro primeros años— solo hay dos temas, y el de A tiene un marcado carácter sanitario, necesariamente el tema B aún todos los aspectos restantes, incluyendo los primeros artículos sobre las primeras aplicaciones tanto de mensajería instantánea como de plataformas de redes sociales en línea. El *topic* M comparte este mismo carácter genérico, al acoger redes y comunidades de diversa naturaleza, desde bandas de crimen organizado hasta refugiados. En consecuencia, los temas B y M pueden considerarse caracterizados por reunir aquellos aspectos temáticos vinculados con redes sociales o comunidades relegados por el resto de los temas en las anualidades donde aparecen.

En cuanto al tema I, comparte con los de B y M esa misma amplitud temática, aunque no centrada en grupos o redes sociales, sino en aspectos técnicos relacionados con el tratamiento de cualquier tipo de dato (texto, voz o imagen) de cara a su transmisión en redes móviles o inalámbricas. En resumen, los temas B, M e I se distinguen por su gran amplitud temática, y cuya aparición se justifica durante aquellos años de transición en que el resto de los temas no pueden acoger ese tipo de contenidos.

Con relación a la popularidad temática, el cálculo de la probabilidad media de cada tema en todos los documentos de cada año muestra —en general— una gran variabilidad en la intensidad o popularidad relativa de todos los temas, indicativo

de un área sometida a frecuentes cambios por la aparición de ramificaciones y facetas temáticas novedosas. El test no paramétrico de tendencia Mann-Kendall muestra que los de A y E tienen una tendencia decreciente a lo largo de los 25 años del estudio, aunque solo la variación del tema A es estadísticamente significativa. Del mismo modo, los de C, D, J, K y T presentan una tendencia general creciente en popularidad a lo largo de los años, pero solamente los temas K y T con significación estadística.

Por tanto, todos los aspectos relativos a los medios de comunicación social y de redes sociales en línea, en particular el comportamiento del usuario, o su utilización en las esferas política, económica, sanitaria, religiosa o social como escenario para la discusión pública, han sido investigados con especial intensidad en los últimos años. En las *figuras 4, 5 y 6* se incluyen las gráficas de los temas con intensidad estadísticamente significativa de carácter decreciente o creciente.

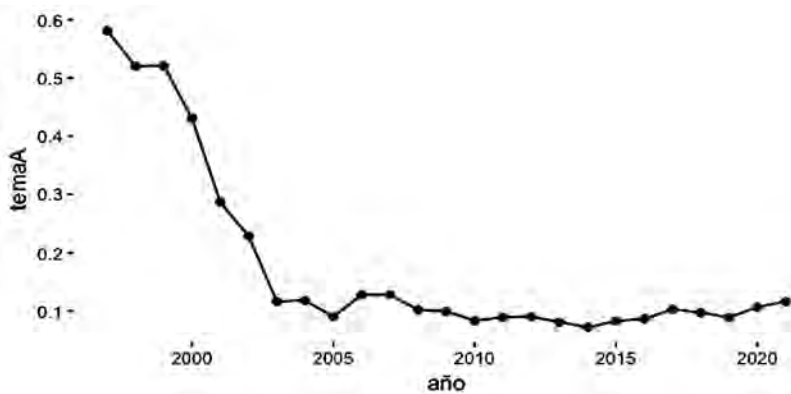


Figura 4. Evolución de la probabilidad media del tema A con intensidad decreciente

Fuente: Elaboración propia.

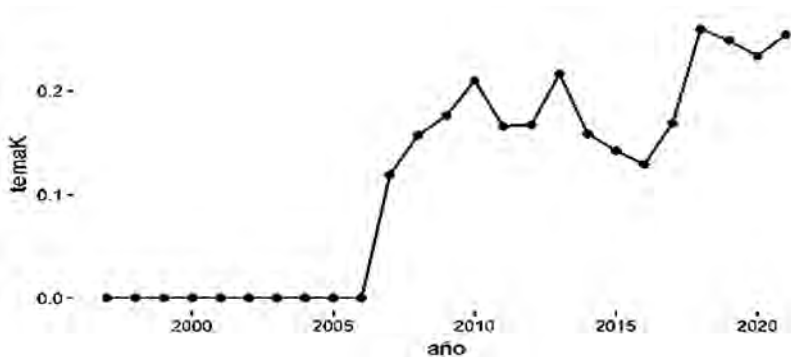


Figura 5. Evolución de la probabilidad media del tema K con intensidad creciente

Fuente: Elaboración propia.

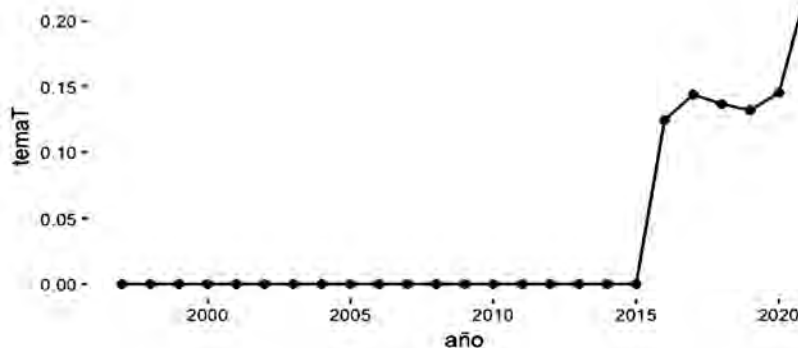


Figura 6. Evolución de la probabilidad media del tema T con intensidad creciente

Fuente: Elaboración propia.

El análisis sugiere la distinción de tres periodos o etapas en la investigación sobre redes sociales entre 1997 y 2021. En el primero (1997-2004), el estudio de las redes sociales como estructuras sociales basadas en el establecimiento de relaciones entre sus componentes (perspectiva sociológica) es la predominante hasta 2001 a través del tema A. Entre 2002 y 2004 se produce el progresivo ascenso y finalmente la preponderancia de los temas B y C, que incorporan nuevos enfoques de carácter más tecnológico, incluyendo las primeras aplicaciones de redes sociales en línea y medios de comunicación social. El tema A incide a lo largo de esta etapa inicial sobre todo en enfermedades psiquiátricas y de transmisión sexual, y en la labor que llevan a cabo las redes sociales compuestas por familia, amistades, compañeros o compañeras y parejas. El tema B aporta el estudio de comunidades virtuales de estudiantes así como de usuarios y de usuarias de sistemas y aplicaciones basados en la web, junto con la descripción de las primeras aplicaciones de mensajería instantánea (Keiretsu o Hubbub) y de medios de comunicación social (Twitter o Swisshouse). El tema C añade, en esta etapa, estudios técnicos sobre programas informáticos relacionados con sistemas móviles o interactivos (protocolos de Internet para comunicación móvil o sistemas interactivos basados en la web para el aprendizaje de idiomas).

En el segundo (2005-2012), desde la perspectiva sociológica, el tema E es el más abordado por los investigadores. Desde el enfoque que incorpora aspectos tecnológicos a los sociológicos, el tema D es el preponderante tras irrumpir con fuerza en esta etapa, con una popularidad en aumento hasta convertirse en el tema más popular desde 2010. En esta etapa el tema E analiza las redes sociales en relación con la economía en general (en la industria y en las empresas en particular), abordando también comunidades y grupos diversos (desde los étnicos

o criminales hasta los de trabajo o comunidades religiosas). Entre los aspectos que aborda el tema D en esta etapa destaca la aplicación del Análisis de Redes Sociales (SNA, por el término en inglés) a redes sociales de masas, la detección automática de relaciones y atributos en redes sociales (confianza o liderazgo, entre otros), así como la descripción y aspectos generales (comportamiento del usuario, etiquetado, detección de opiniones o temas) en medios de comunicación social y en redes sociales en línea (Last.fm, EventBrite, Facebook, Twitter, YouTube o Flickr, entre otros).

En el tercero (2013-2021) se consolida la disminución en la importancia relativa de los temas de enfoque sociológico frente a los de corte más tecnológico, tendencia que empezó a observarse a finales de la segunda etapa. El tema D se mantiene como el tema más popular hasta 2018, año en que cede esta primera posición a favor del tema K, posición que ostentará hasta el final de esta tercera fase. El tema K aborda principalmente el empleo de medios de comunicación social y de redes sociales en línea en diversos ámbitos como los medios de comunicación de masas (periódicos, televisión), los servicios públicos (bomberos, sanidad, administración local), la mercadotecnia y el consumo (imagen de marca, publicidad, persuasión del consumidor, intención de compra, comportamientos relacionados con la moda). Con menor frecuencia surgen aspectos generales relativos a las aplicaciones de mensajería instantánea (intención de continuación con el empleo de la aplicación, intención de abandono de esta, comportamiento al compartir información) y aspectos vinculados a la privacidad o la seguridad (suplantación de identidad, políticas de seguridad, malware). El tema T se especializa en la relación entre los medios de comunicación social o de las redes sociales en línea y el ámbito político (control ideológico, campañas electorales, identidad nacional), religioso (radicalización religiosa, actos de defensa del Islam), sanitario (desinformación durante la pandemia de Covid-19, campañas sanitarias), periodístico (difusión de noticias falsas, noticias sobre temas de actualidad) o policial (terrorismo, persecución de delitos). Con menor frecuencia también se abordan en este tema la utilización de medios de comunicación social y de redes sociales en línea en debates polémicos (anti y pro refugiados o migrantes), en movimientos de protesta o por el activismo (huelgas estudiantiles, violencia de género, discursos de minorías o activismo feminista).

## DISCUSIÓN Y CONCLUSIONES

La literatura acerca de modelado de temas destaca, entre las múltiples herramientas disponibles, tres que incluyen tanto la detección de temas como su evolución temporal: Hierarchical Topic Evolution Model (HTEM), que organiza de manera

jerárquica los temas obtenidos y ofrece su evolución temporal (Song *et al.* 2016); Topic Over Time (TOT), que añade el factor tiempo al descubrimiento de los temas en un corpus (Wang y McCallum 2006); y Dynamic Topic Model (DTM), que muestra la evolución del contenido de los temas extraídos del análisis (Blei y Lafferty 2006). La revisión de estas técnicas destaca dos inconvenientes: se puede observar la evolución de cada tema a lo largo del tiempo, lo que mejora la percepción de sus características, pero no es posible descubrir la aparición o desaparición de nuevos temas a los inicialmente obtenidos (Blei y Lafferty 2006); además, ignoran las relaciones existentes entre los temas y la modificación temporal de tales vinculaciones entre temas, como la unión o la división de estos a lo largo del tiempo (Song *et al.* 2016).

Lo anterior ha provocado que el modelo Asignación Latente de Dirichlet (LDA) atraiga cada vez más adeptos y sea la técnica utilizada más ampliamente en la actualidad, aunque no incorpora el análisis temporal (Ballester y Penner 2022). Por estos motivos, además de por su facilidad de empleo y por el hecho de aportar resultados más interpretables humanamente que otros modelos (permitiendo el etiquetado de los temas, como se ha hecho aquí), se ha decidido su utilización en el presente estudio, tratando al tiempo de que sirva para evaluar la pertinencia o utilidad de estas herramientas en determinadas circunstancias en análisis propios de nuestra disciplina.

El descubrimiento de las relaciones entre los temas se basa en técnicas automáticas de agrupamiento (*clustering* en inglés), empleando para ello medidas matemáticas de similitud. Entre estas destacan dos: symmetric Kullback-Leibler divergence, la más frecuente entre los programas de modelado de temas que incorporan esta faceta (Zhou *et al.* 2017); y la similitud coseno, habitual en estudios bibliométricos, en estudios basados de programas de modelado como el presente que no incluyen el agrupamiento, y en el campo de la recuperación de información (Zhang *et al.* 2018; Leydesdorff 2008; Cacheda Seijo *et al.* 2011).

En este estudio no se emplea, sin embargo, la fórmula clásica de TF e IDF para calcular los pesos que figuran en la fórmula de la similitud coseno. Se introduce una variante novedosa, apropiada cuando los documentos tratados (los resúmenes de los textos) presentan longitudes muy distintas, como sucede en la colección utilizada aquí. Con ello se evitan sesgos debido a que los documentos más largos tienden a contener mayor número de términos y con mayor frecuencia que los textos cortos, alterando al alza los valores de similitud que presentan con otros documentos.

Por tanto, en el estudio se sigue una de las aproximaciones más recientes mediante el empleo de LDA y la detección posterior de los temas a raíz del agrupamiento de los temas detectados en las ventanas temporales seleccionadas (temas anuales en este caso). Una diferencia, sin embargo, se ha introducido en

este análisis. El procedimiento habitual consiste en efectuar LDA con todo el corpus en su integridad, y mediante similaridad coseno descubrir la modificación, unión o división de los temas globales (correspondientes a la colección completa de documentos) mediante el incremento o disminución del número de temas anuales conectados al mismo tema global (Chen *et al.* 2017).

En el presente estudio se introduce un procedimiento novedoso, que no exige la selección previa del número de temas considerados. Se calcula el número óptimo de temas en cada periodo de un año, y mediante similaridad coseno entre los temas anuales se descubre el número de temas que deben considerarse en la totalidad del periodo analizado, además de su evolución temporal. Sea cual sea el procedimiento seguido, la existencia de una correlación fuerte entre los temas involucrados se detecta cuando la similaridad entre temas traspasa un valor umbral (Gaul y Vincent 2017), dependiente en primera instancia del área de conocimiento analizado. En el caso de las redes sociales se obtiene un valor límite o umbral de 0,76, muy semejante al valor 0,75 obtenido en el campo de la recuperación de información (Chen *et al.* 2017), pero distante del valor 0,85 detectado al analizar modelos mentales (Ma *et al.* 2023).

Otro aspecto novedoso del presente estudio radica en el procedimiento de selección del número de temas en cada año de manera independiente. La *perplexity* consituye la métrica más utilizada para la elección del número de temas en un corpus, que indica el grado de incertidumbre sobre la pertenencia de un documento a un tema. Se pone de manifiesto, sin embargo, que esta medida es adecuada siempre que el número de temas se sitúe entre 80 y 100 (Chen *et al.* 2017). Este resulta un número demasiado grande en nuestro caso, porque se pretende en principio descubrir amplios temas de carácter general. Por este motivo se opta por aplicar los criterios y medidas sugeridos por el programa Mallet empleado en el estudio: la coherencia (*coherence*), la exclusividad (*exclusivity*), el parámetro ‘alpha’ y la distribución lo más homogénea posible del tema en el corpus correspondiente. Este modo de proceder es mucho más laborioso, pero permite detectar un número de temas variable por año, el más adecuado en cada una de estas ventanas temporales. De este modo se facilita el análisis de la evolución temática detallada en un área con grandes cambios en el periodo considerado.

En relación con el preprocesamiento de los documentos, se siguen aquí las fases habituales, con una única excepción relativa a la extracción de bigramas o trigramas. Suele realizarse habitualmente la detección de bigramas o trigramas en todos los documentos del corpus (Buehling 2021), mientras que en el presente estudio se opta por emplear las palabras clave compuestas de dos o más palabras seleccionadas por los autores en la descripción de los documentos de la colección. De esta manera se consigue extraer no solo términos relevantes de naturaleza temática, sino también entidades geográficas, personales o temporales de interés.

En este trabajo se analizan las investigaciones llevadas a cabo sobre cualquier tipo de relación establecida entre los miembros de un grupo, sean cuales sean sus objetivos y su naturaleza (humana o no), entre 1997 y 2021. En consecuencia, se afronta el estudio del área de conocimiento de las redes sociales en su integridad, no solo el de las plataformas de redes sociales en línea como Twitter o Facebook, temática no abordada previamente –hasta donde conocemos–. Se han localizado escasos estudios generales sobre la historia de las redes sociales en línea (Boyd y Ellison 2008), así como diversos análisis de la evolución del empleo de las redes sociales en línea en ámbitos concretos como el sanitario (Cho *et al.* 2020), el histórico (Bunnenberg *et al.* 2021), el económico (Agarwal *et al.* 2021) o el educativo (Yu *et al.* 2023), pero ningún análisis comparable al que se realiza aquí.

Las contribuciones de este enfoque se resumen en los siguientes aspectos principales: 1) Incardinar la investigación en redes sociales en línea en su área de conocimiento originario, lo que aporta al área de las Ciencias de la Información una perspectiva teórica más amplia que enriquece los métodos empleados; 2) Evaluar la utilidad –cuando las características del tema lo demanda– de ciertas herramientas informáticas en estudios propios de la disciplina; y 3) Aportar elementos de juicio más ajustados a la realidad sobre el verdadero alcance e importancia relativa del estudio de las redes sociales en línea frente a cualquier otra temática relacionada con las redes sociales en general, desde su aparición en 1997 hasta la actualidad.

Los resultados muestran la pertinencia de distinguir tres fases en la investigación sobre las redes sociales en su integridad: 1997-2004, 2005-2012 y 2013-2021. Los estudios consultados sobre la evolución de las plataformas de redes sociales y medios de comunicación social refuerzan esta distinción, al considerar que a partir de 2003-2004 se inicia una nueva fase en la evolución de las redes sociales en línea (Boyd y Ellison 2008), que a partir de 2013 se distingue una nueva etapa en la literatura sobre medios de comunicación social (Shibuya *et al.* 2022), o que en 2012 acaba un periodo y en 2013 se inicia otro nuevo en la investigación sobre medios de comunicación social (Gálvez 2019).

En relación con el primer periodo (1997-2004), Boyd y Ellison (2008) constatan que entre 1997 y 2001 surgen las primeras aplicaciones de redes sociales en línea de carácter personal, profesional y social; y que en 2001 se inicia la preponderancia de redes profesionales y el carácter económico de las mismas. Estas características coinciden en buena medida con los términos extraídos por el programa Mallet para el tema B (*develop, community, network, work, economics*) y con la descripción textual a partir del análisis cualitativo del corpus («Documentos sobre diversas redes en línea –por ejemplo estudiantes, trabajadores–) [...] anunciando el desarrollo de las primeras aplicaciones de mensajería instantánea y de redes sociales en línea»). Al tiempo, el análisis desde la perspectiva de las redes sociales

en general indica que los aspectos sanitarios son los más populares en esta etapa, y que los aspectos tecnológicos —entre los que se incluyen las primeras aplicaciones de redes sociales en línea, pero que también abarca la investigación sobre sistemas móviles, por ejemplo— presentan un progresivo ascenso en esta fase, mostrando que las redes sociales en línea representa una temática en desarrollo que necesita de un paralelo avance en las herramientas informáticas que las hacen posible.

En el segundo (2005-2012) se destaca la irrupción de las tecnologías y de las aplicaciones de redes sociales en línea (Gálvez 2019). Estas características se reflejan en el surgimiento del tema I entre 2003 y 2012, que desde un enfoque general engloba los aspectos más técnicos del procesamiento de texto, voz e imágenes en móviles y la implementación de servicios de mensajería instantánea y aplicaciones de chat. Por su parte, estudiando la evolución de la mercadotecnia, Sharma y Verma (2018) destacan entre 2001 y 2013 el auge de investigaciones relativas a los usuarios de medios de comunicación social como herramienta de mejora de la relación entre las marcas y sus consumidores. Este aspecto forma parte de los contenidos del tema K, que pone el acento en los factores que afectan al consumo y la decisión de compra de marcas comerciales por los usuarios de redes sociales en línea. El enfoque general que se adopta aquí enriquece el alcance de estas características al mostrar que en esta etapa el tema E (que aborda las redes sociales en relación con la economía en su integridad) resulta más popular, y que el tema K presenta un progresivo auge durante estos años, incorporando el estudio de aspectos como los estados emocionales de los consumidores o la credibilidad de las valoraciones. Incluso desde el punto de vista tecnológico el tema I no es el preponderante en la investigación en estos años, sino el tema D, que pone el acento en la detección automática de relaciones y atributos en redes sociales, así como la descripción del comportamiento del usuario —etiquetado, opiniones— en medios de comunicación social y en redes sociales en línea.

En relación con el tercero (2013-2021), se advierte entre 2013 y 2017 un incremento notable de estudios sobre la conversión de las redes sociales en línea en herramientas influyentes en la política y en instrumentos de movimientos sociales (Gálvez 2019). Estos aspectos están incorporados al tema T, pues en él se aborda la utilización de los medios de comunicación social en la propaganda política, en movimientos de protesta y en foros de discusión pública. También se destaca que entre 2013 y 2016 surgen con fuerza los estudios sobre aspectos económicos relacionados con las redes sociales (Shibuya *et al.* 2022). El enfoque general que se adopta aporta matices de interés, pues el tema E en estos años se interesa —entre otros aspectos— por las características de las plataformas de redes sociales que son de utilidad económica (como la exploración de información social que facilite las recomendaciones del sistema a los usuarios o la reducción de



rumores negativos). En cuanto al tema T, durante estos años también aborda el empleo de las redes sociales en línea en los ámbitos religioso, sanitario, periodístico o policial.

En resumen, aspectos reseñados en estudios anteriores tienen su correlato en los temas que se analizan aquí, aunque en el presente trabajo se incluyen otros temas no reseñados con anterioridad, además de aspectos no descritos en los temas involucrados, lo que enriquece el análisis y aporta una visión más detallada y completa de la investigación durante estos años.

En conclusión, el estudio muestra que es posible abordar la evolución de la investigación en un campo en permanente transformación durante un amplio lapso y con un volumen elevado de datos, como es el caso de las redes sociales, mediante modelización de temas. Los problemas derivados en cuanto a precisión y coherencia de los resultados se superan mediante un procesamiento más extenso de los datos de entrada. Así, para alcanzar una descripción detallada, se disminuye la amplitud de las ventanas temporales, reducidas a un año en nuestro caso, seleccionando el número de temas óptimo para cada una de dichas ventanas. Ello conlleva, a su vez, la necesidad de someter estos datos fragmentarios a procesamientos adicionales para descubrir las conexiones entre dichas temáticas parciales, aportando aspectos poco frecuentes en el tradicional análisis de modelización de temas.

El análisis muestra la existencia de siete grandes temas en el campo de las redes sociales, que se describen sucintamente como los aspectos: sanitarios, económicos, educativos, técnicos (seguridad, almacenamiento, comunicación o procesamiento de la información), sociales (influencia, relaciones entre los componentes de la red, intereses de los miembros), redes sociales en línea (comportamiento de los usuarios, privacidad, patrones de etiquetado) y la utilización de las redes sociales en línea como medio de comunicación social de ideas políticas o religiosas. En cuanto a las fases que pueden discernirse en estos veinticinco años, el análisis sugiere la distinción de tres periodos o etapas en la investigación sobre redes sociales entre 1997 y 2021, cuyas características se describen. Cabe enfatizar que en el tercer periodo (2013-2021) el estudio de las redes sociales en línea (tema K) y su utilización en distintos ámbitos como el político o el religioso (tema T) representan los aspectos investigados con especial intensidad.

## REFERENCIAS

- Agarwal, T., Arya, S. y Bhasin, K. (2021). The evolution of internal employer branding and employee engagement: The temporal role of internal social media usage. *Journal of Information and Knowledge Management*, 20(1), 2150012. <https://doi.org/10.1142/S021964922150012X>
- Aichner, T. y Jacob, F. H. (2015). Measuring the degree of corporate social media use. *International Journal of Market Research*, 57(2), 257-275. <https://doi.org/10.2501/IJMR-2015-018>

- Alias-I (2016). *LingPipe Home*. Alias-i.  
<http://www.alias-i.com/lingpipe/>
- Apache Software Foundation (2022, 25 de octubre). *Welcome to Apache Lucene*. ASF.  
<https://lucene.apache.org>
- Armstrong-Keown, V. Y Patterson, L. (2020). Content analysis in library and information research: An analysis of trends. *Library & Information Science Research*, 42(4), art. 101048.  
<https://doi.org/10.1016/j.lisr.2020.101048>
- Arruda, H. F., Costa, L. F. y Amancio, D. R. (2016). Topic segmentation via community detection in complex networks. *Chaos (Woodbury, N.Y.)*, 26(6), 063120.  
<http://dx.doi.org/10.1063/1.4954215>
- Ballester, O. y Penner, O. (2022). Robustness, replicability and scalability in topic modelling. *Journal of Informetrics*, 16 (1), 101224.  
<https://doi.org/10.1016/j.joi.2021.101224>
- Banerjee, A. y Basu, S. (2007). Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning. En *Proceedings of the seventh SIAM international conference on Data Mining* (pp. 431-436).  
<https://doi.org/10.1137/1.9781611972771.40>
- Berkowitz, S. D. (1982). *An introduction to structural analysis: The network approach to social research*. Butterworths.
- Blei, D. M. y Lafferty, J. D. (2006). Dynamic topic models. En *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 113-120). ACM Press.  
<https://doi.org/10.1145/1143844.1143859>
- Blei, D. M., Ng, A. Y. y Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4-5), 993-1022.  
<https://jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Boyd, D. M. y Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.  
<https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brigadir, I. (2022, 25 de octubre). *Default English stopwords lists from many different sources*. Github.  
<https://github.com/igorbrigadir/stopwords>
- Buehling, K. (2021). Changing research topic trends as an effect of publication ranking: The case of German economists and the Handelsblatt Ranking. *Journal of Informetrics*, 15(3), 101199.  
<https://doi.org/10.1016/j.joi.2021.101199>
- Bunnenberg, C., Logge, T. y Steffen, N. (2021). Social Media History. *Historische Anthropologie*, 29(2), 267-283.  
<https://doi.org/10.7788/hian.2021.29.2.267>
- Cacheda Seijo, F., Fernández Luna, J. M. y Huete Guadix, J. F. (coords.) (2011). *Recuperación de información: un enfoque práctico y multidisciplinar*. Ra-Ma.
- Chabowski, B. R. y Samiec, S. (2023). A bibliometric examination of the literature on emerging market MNEs as the basis for future research. *Journal of Business Research*, 155, art. 113263.  
<https://doi.org/10.1016/j.jbusres.2022.08.027>
- Chang, Y.-W., Huang, M.-H. y Lin, C.-W. (2015). Evolution of research subjects in Library and Information Science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*, 105(3), 2071-2087.  
<https://doi.org/10.1007/s11192-015-1762-8>

- Chen, B., Tsutsui, S., Ding, Y. y Ma, F. (2017). Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval. *Journal of Informetrics*, 11(4), 1175-1189.  
<https://doi.org/10.1016/j.joi.2017.10.003>
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359-377.  
<https://doi.org/10.1002/asi.20317>
- Cho, S. M., Park, C. y Song, M. (2020). The evolution of social health research topics: A data-driven analysis. *Social Science & Medicine*, 265, 113299.  
<https://doi.org/10.1016/j.socscimed.2020.113299>
- Chodera, J. D. y Pande, V. S. (2011). The social network (of protein conformations). *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), 12969-12970.  
<https://doi.org/10.1073/pnas.1109571108>
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. y Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.  
[https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
- Ding, W. y Chen, C. (2014). Dynamic topic detection and tracking: A comparison of HDP, C-word, and cocitation methods. *Journal of the Association for Information Science and Technology*, 65(10), 2084-2097.  
<https://doi.org/10.1002/asi.23134>
- Ding, Y., Chowdhury, G. y Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing y Management*, 37(6), 817-842.  
[https://doi.org/10.1016/S0306-4573\(00\)00051-0](https://doi.org/10.1016/S0306-4573(00)00051-0)
- Ferrer, R., Solé, R. V. y Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, 69(5), 051915.  
<http://dx.doi.org/10.1103/PhysRevE.69.051915>
- Freeman, L. C. (1988). Computer programs in social network analysis. *Connections*, 11(2), 26-31.  
[https://www.researchgate.net/publication/239060164\\_Computer\\_Programs\\_and\\_Social\\_Network\\_Analysis](https://www.researchgate.net/publication/239060164_Computer_Programs_and_Social_Network_Analysis)
- Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science*. Empirical Press.  
[https://www.researchgate.net/publication/238341375\\_The\\_Development\\_of\\_Social\\_Network\\_Analysis\\_A\\_Study\\_in\\_the\\_Sociology\\_of\\_Science](https://www.researchgate.net/publication/238341375_The_Development_of_Social_Network_Analysis_A_Study_in_the_Sociology_of_Science)
- Gálvez, C. (2019). Evolución del campo de investigación de los social media mediante mapas de la ciencia (2008-2017). *Communication & Society*, 32(2), 61-76.  
<https://doi.org/10.15581/003.32.2.61-76>
- Gaul, W. y Vincent, D. (2017). Evaluation of the evolution of relationships between topics over time. *Advances in Data Analysis and Classification*, 11, 159-178.  
<https://doi.org/10.1007/s11634-016-0241-2>
- Gore, D. J., Schueler, K., Ramani, S., Uvin, A., Phillips, G., McNulty, M., Fujimoto, K. y Schneider, J. (2021). HIV response interventions that integrate HIV molecular cluster and social network analysis: A systematic review. *AIDS and Behavior*, 26(6), 1750-1792.  
<https://doi.org/10.1007/s10461-021-03525-0>

- Graham, S., Weingart, S. y Milligan, I. (2021, 3 de septiembre). *Getting started with topic modeling and Mallet*.  
<https://programminghistorian.org/en/lessons/topic-modeling-and-mallet>
- Griffiths, T., Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(suppl. 1), 5228-5235.  
<https://doi.org/10.1073/pnas.0307752101>
- Ha, I., Park, H. y Kim, C. (2014). Analysis of Twitter research trends based on SLR. En *16th International Conference on Advanced Communication Technology* (pp. 774-778). IEEE.  
<https://doi.org/10.1109/ICACT.2014.6779067>
- Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics*, 125(3), 2561-2595.  
<https://doi.org/10.1007/s11192-020-03721-0>
- Harary, F. (1969). *The Graph Theory*. Addison-Wesley Publishing Company.
- Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, 21, 107-112.  
<https://doi.org/10.1080/00223980.1946.9917275>
- Hofmann, T. (1999). Probabilistic latent semantic indexing. En *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50-57). Association for Computing Machinery.  
<https://doi.org/10.1145/312624.312649>
- Isasi, J. (2022, 15 de noviembre). *Modelado de temas con Mallet*.  
<https://repositories.lib.utexas.edu/handle/2152/72737>
- Jeong, D. H. y Min, S. (2014). Time gap analysis by the topic model-based temporal technique. *Journal of Informetrics*, 8(3), 776-790.  
<https://doi.org/10.1016/j.joi.2014.07.005>
- Johnson, B. J. (2019, 23 de febrero). *Contar todas las palabras diferentes en un archivo de texto*. LWP, lawebdelprogramador.  
<https://www.lawebdelprogramador.com/foros/Java/1685229-Contar-todas-las-palabras-diferentes-en-un-archivo-de-texto.html>
- Jung, S. y Yoon, W. C. (2020). An alternative topic model based on Common Interest Authors for topic evolution analysis. *Journal of Informetrics*, 14(3), 101040.  
<https://doi.org/10.1016/j.joi.2020.101040>
- Kai, H., Qing, L., Kunlun, Q., Siluo, Y., Jin, M., Xiaokang, F., Jie, Z., Huayi, W., Ya, G. y Qibing, Z. (2019). Understanding the topic evolution of scientific literatures like an evolving city: Using Google Word2Vec model and spatial autocorrelation analysis. *Information Processing and Management*, 56(4), 1185-1203.  
<https://doi.org/10.1016/j.ipm.2019.02.014>
- Kleinberg, J. (2002). Bursty and hierarchical structure in streams. *Data Mining y Knowledge Discovery*, 7(4), 373-397.  
<https://doi.org/10.1145/775047.775061>
- Knoke, D. y Kuklinski, J. H. (1982). *Network analysis*. Sage.
- Kumar, K. (2018, 3 de mayo). *Evaluation of topic modeling: Topic coherence*.  
<https://datascienceplus.com/evaluation-of-topic-modeling-topic-coherence/>
- Landauer, T. K., McNamara, D. S., Dennis, S. y Kintsch, W. (2007). *Handbook of latent semantic analysis*. Taylor y Francis Group.  
<https://doi.org/10.4324/9780203936399>

- Leydesdorff, L. (2007). On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, 59(1), 77-85.  
<https://doi.org/10.1002/asi.20732>
- Li, X. y Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000-2017). *Journal of Information Science*, 47(2), 161-175.  
<https://doi.org/10.1177/0165551519877049>
- Li, Y., Wu, Y. y Chen, Y. (2021). A review of enterprise social media: visualization of landscape and evolution. *Internet Research*, 31(4), 1203-1235.  
<https://doi.org/10.1108/INTR-07-2020-0389>
- Liu, H., Chen, Z., Tang, J., Zhou, Y. y Liu, S. (2020). Mapping the technology evolution path: a novel model for dynamic topic detection and tracking. *Scientometrics*, 125(3), 2043-2090.  
<https://doi.org/10.1007/s11192-020-03700-5>
- Ma, J., Wang, L., Zhang, Y.-R., Yuan, W. y Guo, W. (2023). An integrated latent Dirichlet allocation and Word2vec method for generating the topic evolution of mental models from global to local. *Expert Systems With Applications*, 212, 118695.  
<https://doi.org/10.1016/j.eswa.2022.118695>
- Mallet (2022, 12 de julio). *Importing data Mallet*. Mallet.  
<https://mimno.github.io/Mallet/import.html>
- McCallum, A. K. (2022a, 15 de noviembre). *MALLET: A Machine Learning for Language Toolkit*.  
[https://mallet.cs.umass.edu/index.php/Main\\_Page](https://mallet.cs.umass.edu/index.php/Main_Page)
- McCallum, A. K. (2022b, 15 de noviembre). *Topic model diagnostics*.  
<https://mallet.cs.umass.edu/diagnostics.php>
- McCallum, A., Wang, X. y Corrada-Emmanuel, A. (2007). Topic and role discovery in social networks with experiments on Enron and academic email. *Journal of Artificial Intelligence Research*, 30, 249-272.
- McCandless, M., Hatcher, E. y Gospodnetic, O. (2010). *Lucene in action*. Manning.
- Mikolov, T., Chen, K., Corrado, G. y Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*, 1301.3781 [cs.CL].  
<https://doi.org/10.48550/arXiv.1301.3781>
- Moreno, J. L. (1937). Inter-personal therapy and the psychopathology of inter-personal relations. *Sociometry*, 1(1-2), 9-76.  
<https://doi.org/10.2307/2785258>
- Moreno, J. L., Jennings, H. H. (1938). Statistics of social configurations. *Sociometry*, 1(3-4), 342-373.  
<https://doi.org/10.2307/2785588>
- NLTK (2022, 25 de octubre). *Natural Language Toolkit*. NLTK.  
<https://www.nltk.org>
- Onyancha, O. B. (2018). Forty-five years of LIS research evolution, 1971-2015: An informetrics study of the author-supplied keywords. *Publishing Research Quarterly*, 34(3), 456-470.  
<https://doi.org/10.1007/s12109-018-9590-3>
- OpenNLP (2022, 5 de junio). *SnowballStemmer (Apache OpenNLP Tools 1.8.0 API)*. OpenNLP Tools.  
<https://opennlp.apache.org/docs/1.8.0/apidocs/opennlp-tools/opennlp/tools/stemmer/snowball/SnowballStemmer.html>

- Otte, E. y Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.  
<https://doi.org/10.1177/016555150202800601>
- Pappi, F. U. y Stelck, K. (1987). Ein Databanksystem zur Netzwerkanalyse. En Pappi, F. U. (ed.), *Methoden Netzwerkanalyse* (1st ed., pp. 253-265). Oldenberg.
- Peset, F., Garzón-Farinos, F., González, L. M. et al. (2020). Survival analysis of author keywords: An application to the library and information sciences area. *Journal of the Association for Information Science and Technology*, 71(4), 462-473.  
<https://doi.org/10.1002/asi.24248>
- Pohlert, T. (2022, 26 de marzo). *Non-parametric trend tests and change-point detection*. R project.  
<https://cran.r-project.org/web/packages/trend/vignettes/trend.pdf>
- R Core Team (2022, 15 de noviembre). *Hclust function: Hierarchical Clustering*.  
<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust>
- Ricci, R. (2018). Movimentos e mobilizações sociais no Brasil: de 2013 aos dias atuais. *Saúde em Debate*, 42, 90-107.  
<https://doi.org/10.1590/0103-11042018S308>
- Ridings, C. M., Gefen, D. y Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *The Journal of Strategic Information Systems*, 11(3-4), 271-295.  
[https://doi.org/10.1016/s0963-8687\(02\)00021-5](https://doi.org/10.1016/s0963-8687(02)00021-5)
- Shan, B. y Li, F. (2010). A survey of topic evolution based on LDA. *Journal of Chinese Information Processing*, 24(6), 43-50.
- Sharma, S. y Verma, H. V. (2018). Social media marketing: Evolution and change. En G. Hegde y G. Shainesh (eds.). *Social Media Marketing: Emerging Concepts and Applications*, pp. 19-36. Springer.
- Shen, X. y Wang, L. (2020). Topic evolution and emerging topic analysis based on open source software. *Journal of Data and Information Science*, 5(4), 126-136.  
<https://doi.org/10.2478/jdis-2020-0033>
- Shibuya, Y., Hamm, A. y Pargman, T. C. (2022). Mapping HCI research methods for studying social media interaction: A systematic literature review. *Computers in Human Behavior*, 129, 107131.  
<https://doi.org/10.1016/j.chb.2021.107131>
- Silge, J. (2018, 8 de septiembre). *Training, evaluating, and interpreting topic models*.  
<https://juliasilge.com/blog/evaluating-stm/>
- Singhal, A. (2001). Modern information retrieval: a brief overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 24(4), 35-43.  
<http://www1.cs.columbia.edu/~gravano/Qual/Papers/singhal.pdf>
- Snowball (2021, 25 de octubre). *Snowball*. Snowball.  
<https://snowballstem.org>
- Song, M., Heo, G. E. y Kim, S. Y. (2014). Analyzing topic evolution in bioinformatics: Investigation of dynamics of the field with conference data in DBLP. *Scientometrics*, 101(1), 397-428.  
<https://doi.org/10.1007/s11192-014-1246-2>
- Song, J., Huang, Y., Qi, Y., Li, Y., Li, F., Fu, K. y Huang, T. (2016). Discovering Hierarchical Topic Evolution in Time-Stamped Documents. *Journal of the Association for Information Science and Technology*, 67(4), 915-927.  
<https://doi.org/10.1002/asi.23439>

- Statista, We Are Social, Hootsuite y DataReportal (2022, 26 de enero). *Most popular social networks worldwide as of January 2022, ranked by number of monthly active users (in millions)*. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Sueur, C. y Pelé, M. (2016). Social network and decision-making in primates: A report on Franco-Japanese research collaborations. *Primates*, 57(3), 327-332. <https://doi.org/10.1007/s10329-015-0505-z>
- Suominen, A. y Toivanen, H. (2015). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology*, 67(10), 2464-2476. <https://doi.org/10.1002/asi.23596>
- Taipale, S. y Farinosi, M. (2018). The big meaning of small messages: The use of WhatsApp in intergenerational family communication. En J. Zhou, J. y Salvendy, G. (eds.), *Human aspects of IT for the aged population. Acceptance, Communication and Participation* (pp. 532-546). Springer. [https://doi.org/10.1007/978-3-319-92034-4\\_40](https://doi.org/10.1007/978-3-319-92034-4_40)
- Tdk Technologies (2020, 12 de noviembre). *Topic modeling explained: LDA to Bayesian Inference*. <https://www.tdktech.com/tech-talks/topic-modeling-explained-lda-to-bayesian-inference/>
- Tuomaala, O., Järvelin, K. y Vakkari, P. (2014). Evolution of library and information science, 1965-2005: Content analysis of journal articles. *Journal of the Association for Information Science and Technology*, 65(7), 1446-1462. <https://doi.org/10.1002/asi.23034>
- Wang, G. y Robinson, R. (2002). An architecture for web-enabled engineering applications based on lightweight high-performance CORBA. En Williams, A. D. (ed.), *Proceedings of the 6th International Enterprise Distributed Object Computing Conference* (pp. 249-257). IEEE Computer Society. <https://doi.org/10.1109/EDOC.2002.1137714>
- Wang, X. y McCallum, A. (2006). Topics over time: A non-Markov continuous-time model of topical trends. En *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*, pp. 424-433). ACM Press <https://doi.org/10.1145/1150402.1150450>
- Wasserman, S. y Faust, K. (1994). Social network analysis in the social and behavioral sciences. En *Social network analysis: Methods and applications* (Structural Analysis in the Social Sciences, pp. 3-27). Cambridge University Press. <https://doi.org/10.1017/CB09780511815487.002>
- Wu, Q., Zhang, C., Hong, Q. y Chen, L. (2014). Topic evolution based on LDA and HMM and its application in stem cell research. *Journal of Information Science*, 40(5), 611-620. <https://doi.org/10.1177/0165551514540565>
- Xu, S., Hao, L., An, X., Yang, G. y Wang, F. (2019). Emerging research topics detection with multiple machine learning models. *Journal of Informetrics*, 13(4), 100983. <https://doi.org/10.1016/j.joi.2019.100983>
- Yang, C., Tang, X., Kim, S. Y. y Song, M. (2012). A trend analysis of domain-specific literatures with content and co-author network similarity. En Chen, H. H. y Chowdhury, G. (eds.), *The 14th International Conference on Asia-Pacific Digital Libraries (ICADL 2012)*, pp. 73-76). Springer. [https://doi.org/10.1007/978-3-642-34752-8\\_10](https://doi.org/10.1007/978-3-642-34752-8_10)



- Yanhui, S., Lijuan, W. y Junping, Q. (2021). A comparative study of first and all-author bibliographic coupling analysis based on Scientometrics. *Scientometrics*, 126(2), 1125-1147. <https://doi.org/10.1007/s11192-020-03798-7>
- Yau, C. K., Porter, A., Newman, N. y Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, 100(3), 767-786. <https://doi.org/10.1007/s11192-014-1321-8>
- Yu, Z., Sukjairungwattana, P. y Xu, W. (2023). Bibliometric analyses of social media for educational purposes over four decades. *Frontiers in Psychology*, 13, 1061989. <https://doi.org/10.3389/fpsyg.2022.1061989>
- Zanardo, N., Parra, G. J., Diaz-Aguirre, F., Pratt, E. A. L. y Möller, L. M. (2018). Social cohesion and intra-population community structure in southern Australian bottlenose dolphins. *Behavioral Ecology and Sociobiology*, 72(9), 1-13. <https://doi.org/10.1007/s00265-018-2557-8>
- Zhang, J., Chen, H., Chan, H. C. B. y Leung, V. C. M. (2009). PUCS: Personal unified communications over heterogeneous wireless networks. En Ramasubramanian, S. y Aracil-Rico, J. (eds.), *Proceedings of the 2009 6th International Conference on Broadband Communications, Networks and Systems, BROADNETS 2009* (article number 5336353). <https://doi.org/10.4108/ICST.BROADNETS2009.7851>
- Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H. y Zhang, G. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics*, 12(4), 1099-1117. <https://doi.org/10.1016/j.joi.2018.09.004>
- Zhou, H., Yu, H. y Hu, R. (2017). Topic evolution based on the probabilistic topic model: a review. *Frontiers of Computer Science*, 11(5), 786-802. <https://doi.org/10.1007/s11704-016-5442-5>
- Zhu, M., Zhang, X. y Wang, H. (2016). A LDA based model for topic evolution: Evidence from Information Science journals. En *Proceedings of 2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications* (pp. 49-54). Atlantis Press. <https://doi.org/10.2991/msota-16.2016.12>
- Zou, C. (2018). Analyzing research trends on drug safety using topic modeling. *Expert opinion on drug safety*, 17(6), 629-636. <https://doi.org/10.1080/14740338.2018.1458838>

### Para citar este texto:

Martínez-Comeche, Juan-Antonio. 2023. “Veinticinco años de investigación en redes sociales: evolución de temas entre 1997 y 2021 empleando el algoritmo Asignación Latente de Dirichlet”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 145-177.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58777>



# Productividad científica de investigadores de la Universidad Nacional de Asunción (UNA): comprobación del modelo de Lotka

Emilce Sena Correa\*

María Luisa Lascurain\*\*

Mariana Cáceres Ruiz Díaz\*

Johana Raquel Pineda Alvarenga\*

*Artículo recibido:*

*12 de mayo de 2023*

*Artículo aceptado:*

*16 de agosto de 2023*

*Artículo de investigación*

## RESUMEN

Este trabajo plantea la comprobación del modelo sobre productividad de los autores propuesto por Lotka. En su forma generalizada, para los investigadores de la Universidad Nacional de Asunción que publicaron entre los años 2015 y 2020 en revistas paraguayas recogidas en SciELO particularmente en las áreas de Ciencias Médicas y de la Salud, Ciencias Sociales, Ciencias Veterinarias y Agrícolas. Se recuperaron un total de 627 documentos incluidos en 15 títulos de revista. En las cuatro áreas de estudio se comprueba el ajuste de la producción de los autores al postulado de Lotka con una importante proporción de autores esporádicos. El análisis demuestra la posibilidad de

\* Facultad Politécnica, Universidad Nacional de Asunción, Paraguay

esena@pol.una.py mcaceresrd@pol.una.py jpineda@pol.una.py

\*\* Facultad de Humanidades, Comunicación y Documentación, Universidad Carlos III de Madrid, España mlascura@bib.uc3m.es

comprobación del modelo en distintas áreas científicas en un periodo inferior a los 10 años.

**Palabras clave:** Paraguay; Revistas científicas; Modelo de Lotka; Productividad de autores paraguayos; Producción científica; Universidad Nacional de Asunción; Investigadores.

### Scientific productivity of researchers at the National University of Asuncion (UNA): Testing Lotka's model

*Emilce Sena Correa, María Luisa Lascrain, Mariana Cáceres Ruiz Díaz and Johana Raquel Pineda Alvarenga*

#### ABSTRACT

This paper proposes the verification of the model about author productivity proposed by Lotka, in its generalized form, for researchers of the National University of Asuncion who published between 2015 and 2020 in Paraguayan journals, collected in SciELO, in the areas of Medical and Health Sciences, Social Sciences, Veterinary and Agricultural Sciences. A total of 627 documents included in 15 journal titles were collected. In the four areas studied, the adjustment of the authors' production to Lotka's postulate was verified, with a significant proportion of sporadic authors. The study demonstrates the possibility of testing the model in different scientific areas in a period of less than 10 years.

**Keywords:** Paraguay; Scientific journals; Lotka Model; Productivity of Paraguayan authors; Scientific production; National University of Asuncion; Researchers.

#### INTRODUCCIÓN

En su artículo publicado en *Journal of the Washington Academy of Science*, Lotka (1926) desarrolló el modelo matemático que serviría para determinar la contribución de los autores de distintas categorías al progreso de la ciencia. A partir de ese momento, su ley epónima del cuadrado inverso estableció que, en todas las comunidades científicas, el número de autores ( $A$ ) que han publicado

un determinado número de trabajos ( $n$ ) en el curso de varios años de actividad ( $An$ ), es igual a la cantidad de autores que han publicado un único trabajo ( $A1$ ) dividida entre el cuadrado de  $n$  ( $An=A1/n^2$ ). Se trata, como señala Gorbea Portal (2005, 67), de un modelo matemático que sirve para determinar «la presencia de cierta regularidad estadística en el comportamiento cuantitativo de la productividad de autores científicos dentro de una disciplina específica».

Como recogen varios autores (Potter 1981; Urbizagástegui Alvarado 1999; 2005; Gorbea Portal 2005; Pulgarín 2012; Egghe 2012), el modelo de Lotka ha tenido numerosas críticas tanto en sus aspectos metodológicos y matemáticos como por su dispar eficacia al ser aplicado en distintas áreas temáticas.

En esta investigación, el propósito radica en comprobar el ajuste al modelo de Lotka del comportamiento de la producción científica de los investigadores de la UNA en revistas paraguayas indexadas en SciELO en el periodo 2015-2020.

Se pretende validar el ajuste de la productividad estudiada en cuatro áreas científicas y realizar comparaciones en los patrones de comportamiento, a fin de establecer una línea de base para estudios posteriores respecto a la productividad de autores, según campos de la ciencia.

La investigación parte de una perspectiva bibliométrica para abordar la evaluación de los resultados de la actividad científica como es el análisis de la productividad. Dado que las publicaciones científicas dan una mayor visibilidad y prestigio tanto a la Universidad como a sus investigadores, se destaca la relevancia de este trabajo para la UNA.

## METODOLOGÍA

Se planteó un estudio con enfoque cuantitativo y alcance correlacional, ya que se realizó la comprobación de ajustar los datos que se observaron a los esperados según el modelo de Lotka, para lo cual se aplicó una prueba de hipótesis. El diseño fue no experimental y de corte transversal.

En el contexto de este estudio, se denominó investigador de la UNA a todo autor que declaró esta filiación institucional en su publicación. En lo que respecta a los casos de filiación múltiple de los autores, se consideró la de la UNA independientemente del orden de presentación de ésta.

Se examinaron las publicaciones indexadas en la base de datos bibliográfica SciELO (<https://scielo.org/>), correspondientes al periodo 2015-2020. Entre los meses de enero y febrero del año 2021, se procedió a exportar de SciELO los siguientes metadatos: número de identificación de la publicación, título de la publicación, nombres y apellidos de la totalidad de autores, nombre de la revista, idioma, año de publicación, dirección electrónica de la publicación.

Los datos fueron alojados y procesados en hojas de cálculo de Microsoft Excel. Con estos datos se conformó la primera base, en la cual se identificaron las publicaciones que contaban con al menos un investigador de la UNA como autor. Para ello se examinaron las publicaciones utilizando la dirección electrónica correspondiente, con el objeto de verificar la filiación institucional de los autores.

A continuación, se conformó una segunda base de datos, que contempló únicamente las publicaciones con al menos un investigador de la UNA como autor. A esta segunda base se añadieron las variables siguientes:

- Institución de filiación
- Orden de firmas
- Tipo de documento: artículo de revisión, artículo original, caso clínico, comunicación breve, reporte de caso, revisión bibliográfica, tema de actualidad, carta al editor, editorial, nota al editor, proyecto educativo, reflexión, resumen de congreso.

Por último, la base de datos definitiva se constituyó con las variables que se citan a continuación:

- Número de identificación de la publicación
- Título de la publicación
- Nombres y apellidos del autor
- Nombre de la revista
- Idioma de la publicación
- Año de publicación
- Dirección electrónica de la publicación
- Institución de filiación del autor
- Orden de firmas
- Área de la ciencia
- Tipo de documento

Las categorías de las variables fueron normalizadas y codificadas.

En cuanto a los nombres y apellidos de los autores, éstos fueron depurados por medio del programa OpenRefine (<https://openrefine.org/>). Esta herramienta de código abierto libre permite ordenar y depurar los datos para disponerlos de forma utilizable (Centro Universitario de Liderazgo y Tecnología Avanzada 2016). Con ello, se logró la normalización de los nombres y apellidos de los autores de forma semiautomática, realizándose la corrección de diferencias de ortografía, espacios en blanco, entre otros, mediante el uso de filtros y

agrupamiento de los datos. Este proceso posibilitó después determinar la cantidad de publicaciones por autor y superar el sesgo que en este tipo de análisis supone las firmas incorrectas.

Cabe destacar que en muchos casos la visibilidad nacional e internacional de los trabajos científicos se ve afectada por la falta de normalización de la firma del investigador, tal como lo mencionan Barceló *et al.* (2009, 154): «La ambigüedad que podemos encontrar en nombres de autores dedicados a la investigación, la literatura y las artes, dificulta la adjudicación acertada de las obras concebidas», un factor que incide en la realización de análisis métricos de la producción científica y de la visibilidad e impacto de la misma.

En este sentido, se presentaron diversas limitaciones y obstáculos durante el desarrollo del trabajo.

Asimismo, es significativo apuntar que se escogió SciELO para la realización de este estudio por ser una de las importantes bases de datos bibliográficas que da visibilidad a la productividad científica generada en América Latina.

En cuanto a los criterios establecidos para la selección:

- Se consideraron únicamente aquellas revistas científicas pertenecientes a Paraguay, con el objeto de aproximar el aporte de los investigadores de la UNA en las revistas nacionales.
- Para la elección de las principales áreas de la ciencia se tuvo en consideración, por un lado, las que concentraban la mayor cantidad de publicaciones, y por otro, la relevancia de las temáticas abordadas. Las denominaciones de las áreas corresponden al Manual de Frascati, edición del año 2015, con la salvedad de que, para este estudio, Ciencias Agrícolas y Ciencias Veterinarias fueron presentadas de manera independiente, conforme los requerimientos del presente trabajo.
- Seguidamente, a fin de determinar las revistas correspondientes a cada área se llevó a cabo la revisión de los aspectos temáticos y finalidad de estas.
- Por último, se seleccionó la tipología documental que atañe a los tipos de documentos generalmente considerados en los estudios métricos: artículo de revisión, artículo original, reporte/revisión de caso y caso clínico.

Conforme a los criterios establecidos, la aplicación del modelo de Lotka se llevó a cabo en la producción científica de los investigadores de la UNA de las áreas: Ciencias Médicas y de la Salud, Ciencias Sociales, Ciencias Veterinarias y Ciencias Agrícolas, en 15 de 18 revistas paraguayas que se encontraban indexadas en SciELO, en el periodo seleccionado. Fueron exceptuadas tres revistas que congregaron alrededor de 9 % del total de documentos que presentaron temática multidisciplinaria, las cuales no fueron consideradas debido a los requerimientos del modelo objeto de este análisis.

En lo que hace propiamente al modelo de Lotka cabe mencionar que para el conteo de las contribuciones se adoptó el método completo, según el cual la contribución es atribuida tanto al primer autor como a los demás autores (Urbizagástegui Alvarado 1999).

Se propuso la comprobación de la forma general del modelo de Lotka que se conoce como el de «poder inverso generalizado», cuya expresión es (Urbizagástegui Alvarado 2005, 55):

$$y_x = \frac{C}{x^n}$$

Donde:

$y_x$ : Probabilidad de que un autor realice "x" contribuciones

C: Parámetro

n: Parámetro

Para la comprobación del modelo se siguieron los lineamientos dados por Urbizagástegui Alvarado (2005), siendo éstos:

- Recopilación y tabulación de datos

La recopilación de los datos se realizó con base en las contribuciones de los autores en un ámbito determinado de la ciencia, acreditándose la contribución a la totalidad de autores presentes en la obra.

Los datos se tabularon de tal manera que en la primera columna se dispuso en forma ascendente la cantidad de contribuciones (columna x) y, en la segunda, la cantidad de autores que realizan x contribuciones (columna y).

- Cálculo del valor de n

Se utilizó el método de mínimos cuadrados, el cual permite estimar los parámetros en una ecuación mediante la minimización de la suma de los cuadrados de las diferencias entre los valores observados y los correspondientes a la ecuación de predicción (Canavos 2001), para llevar a cabo la operación del valor de n ajustado a los datos observados, aplicando la fórmula siguiente:

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

Donde:

X: Logaritmo de base 10 de x

Y: Logaritmo de base 10 de y

N: Cantidad de pares de datos en la muestra

- Cálculo del valor de C

Para calcular el valor del parámetro C se utilizó una fórmula de aproximación exacta que se expresa como:

$$C = \frac{1}{\sum_{x=1}^{P-1} \left( \frac{1}{x^n} \right) + \frac{1}{(n-1)P^{(n-1)}} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{(n+1)}}$$

Donde:

P: Cantidad de pares de datos observados xy

- Prueba de ajuste

El estadístico de prueba que se utilizó para la comprobación estadística del modelo fue el de Kolmogorov-Smirnov (K-S), a un nivel de significación de 1 %.

Cabe señalar que la regla de decisión aplicada fue: Si la máxima desviación absoluta es menor al valor crítico ( $D_{máx} < \text{Valor crítico}$ ) la hipótesis nula no puede ser rechazada. Por consiguiente, la productividad científica de los investigadores observados se ajusta a la forma generalizada del modelo matemático de Lotka.

## PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

Se recuperó un total de 693 documentos con al menos un investigador con filiación a la UNA, correspondientes al periodo 2015-2020, presentes en 18 revistas paraguayas indexadas en SciELO. Se observó que la mitad de estas fueron del área de Ciencias Médicas y de la Salud y agruparon aproximadamente siete de cada 10 documentos, seguida por el área de Ciencias Sociales que concentró alrededor de 22 % de las revistas y cerca de 8 % de los documentos. Por su parte, las áreas de Ciencias Veterinarias y Ciencias Agrícolas presentaron cada una de ellas una revista única, agrupando 5,5 % y 5,2 % de los documentos respectivamente (véase *tabla 1*). Conforme a estos hallazgos las áreas de la ciencia seleccionadas para la comprobación del modelo de Lotka fueron las cuatro mencionadas con anterioridad. En estas se recuperaron 627 documentos en total, correspondientes a 15 revistas entre los 18 títulos del país recogidos en SciELO.

La cuantificación de los investigadores mostró que el campo de Ciencias Médicas y de la Salud concentró a la mayor parte de ellos con 78 %, posicionándose a continuación las Ciencias Veterinarias con cerca de 10 %, seguida por las Ciencias Sociales y las Ciencias Agrícolas con aproximadamente 7 % y 6 %, respectivamente (véase *tabla 2*).

Así, se observó un comportamiento de correspondencia entre la concentración de publicaciones e investigadores para el área de Ciencias Médicas y de la Salud, con mayor cantidad de publicaciones e investigadores.

Área de la ciencia	Cantidad de revistas	Cantidad de documentos
Ciencias Médicas y de la Salud	9	500
Ciencias Sociales	4	53
Ciencias Veterinarias	1	38
Ciencias Agrícolas	1	36
Otras áreas	3	66
Total	18	693

*Tabla 1.* Distribución de cantidad de revistas y publicaciones con al menos un investigador con filiación a la UNA, según áreas de la ciencia. Periodo 2015-2020  
Fuente: Elaboración propia a partir de datos relevados en el estudio.

Área de la ciencia	Cantidad de investigadores
Ciencias Médicas y de la Salud	992
Ciencias Sociales	83
Ciencias Veterinarias	120
Ciencias Agrícolas	70
Total	1 265

*Tabla 2.* Distribución de investigadores con filiación a la UNA que poseen publicaciones en revistas paraguayas indexadas en SciELO, según áreas de la ciencia seleccionadas. Periodo 2015-2020  
Fuente: Elaboración propia a partir de datos relevados en el estudio.

Posteriormente, se presentan los hallazgos según el área de la ciencia, en lo que respecta a la comprobación del modelo de Lotka en su forma generalizada.

### ***Ciencias Médicas y de la Salud***

En el área de Ciencias Médicas y de la Salud se registró que un total de 992 investigadores produjeron 500 publicaciones, contando con una sola publicación cerca de 67 % y con más de cinco aproximadamente 5 % de ellos (véase *tabla 3*). El porcentaje de autores con una única contribución resulta mayor al encontrado por Cáceres Ruiz Díaz (2019) en una muestra de investigadores de Ciencias Médicas categorizados en el PRONII de Paraguay, con un valor aproximado de 38 % para investigadores con una sola publicación.



Distribución observada					Distribución esperada según el modelo de Lotka			Prueba de ajuste según K-S	
Cantidad de contribuciones $x$	Cantidad de investigadores observados y	Cantidad de investigadores observados acumulados	Cantidad de investigadores observados relativo $s(x)$	Cantidad de investigadores esperados	Cantidad de investigadores esperados acumulados	Cantidad de investigadores esperados relativo $f(x)$	$f(x) - s(x)$	$ f(x) - s(x) $	
1	669	669	0,67440	637	637	0,65664	-0,01776	0,01776	
2	154	823	0,82964	149	787	0,81070	-0,01894	0,01894	
3	67	890	0,89718	64	851	0,87667	-0,02050	0,02050	
4	31	921	0,92843	35	886	0,91282	-0,01561	0,01561	
5	25	946	0,95363	22	908	0,93548	-0,01814	0,01814	
6	11	957	0,96472	15	923	0,95096	-0,01375	0,01375	
7	11	968	0,97581	11	933	0,96218	-0,01363	0,01363	
8	7	975	0,98286	8	942	0,97066	-0,01221	0,01221	
9	2	977	0,98488	6	948	0,97729	-0,00759	0,00759	
10	4	981	0,98891	5	953	0,98260	-0,00631	0,00631	
11	1	982	0,98992	4	958	0,98696	-0,00296	0,00296	
13	2	984	0,99194	3	961	0,99003	-0,00190	0,00190	
15	2	986	0,99395	2	963	0,99231	-0,00164	0,00164	

16	1	987	0,99496	2	965	0,99430	-0,00066	0,00066
17	1	988	0,99597	2	966	0,99605	0,00008	0,00008
18	1	989	0,99698	2	968	0,99761	0,00063	0,00063
19	1	990	0,99798	1	969	0,99900	0,00101	0,00101
28	1	991	0,99899	1	970	0,99961	0,00062	0,00062
35	1	992	1,00000	0	970	1,00000	0,00000	0,00000

Tabla 3. Distribución de contribuciones por investigador y comprobación estadística del modelo matemático de Lotka en el área de Ciencias Médicas y de la Salud. Periodo 2015-2020  
Fuente: Elaboración propia a partir de datos relevados en este estudio.

La comprobación del modelo de Lotka en esta área arrojó un valor de  $n=-2,0916$  (ecuación 1) y un valor del parámetro  $C=0,6422$  (ecuación 2). Se observó un coeficiente de determinación de 92,67 % entre la cantidad de publicaciones e investigadores (figura 1). Como la máxima desviación absoluta 0,02050 fue menor al valor crítico 0,05175 la hipótesis nula no fue rechazada (tabla 3), por tanto, el modelo de Lotka se cumplió para el área de Ciencias Médicas y de la Salud.

Ecuación 1

$$n = \frac{19(7,44) - (17,85)(14,16)}{19(19,58) - (17,85)^2} = -2,0916$$

Ecuación 2

$$C = \frac{1}{\sum_{x=1}^{18} \left( \frac{1}{x^{2,0916}} \right) + \frac{1}{(2,0916-1)(19)(2,0916-1)} + \frac{1}{2(19)2,0916} + \frac{2,0916}{24(19-1)(2,0916+1)}} = \frac{1}{1,5602} = 0,6422$$

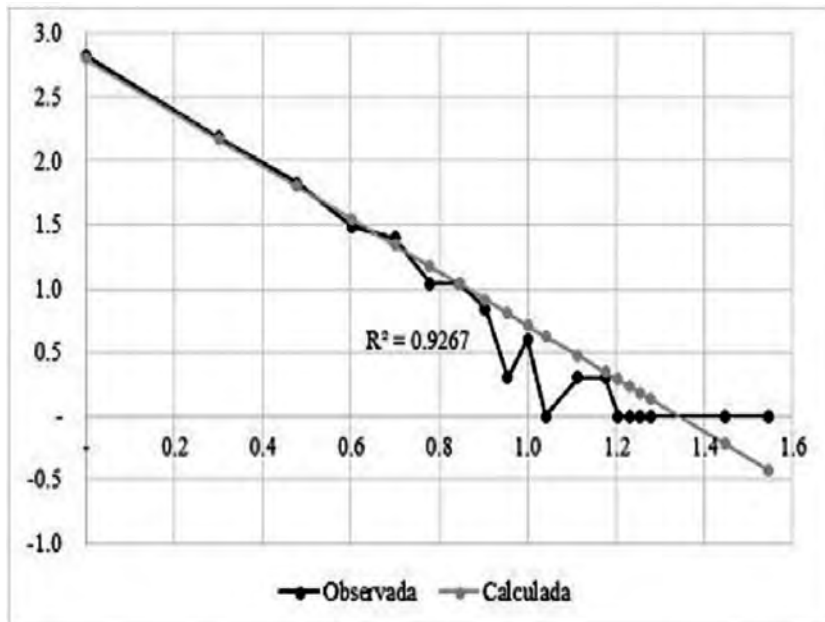


Figura 1. Distribución de productividad de investigadores observados y esperados en el área de Ciencias Médicas y de la Salud (escala logarítmica)

Fuente: Elaboración propia a partir de datos relevados en este estudio.

## Ciencias Sociales

En Ciencias Sociales 83 investigadores generaron 53 publicaciones. Se observó que en torno a 93 % de los mismos produjeron una única publicación, evidenciándose por ende una elevada presencia de pequeños productores (véase *tabla 4*).

La comprobación del modelo de Lotka para el área de Ciencias Sociales mostró un valor de  $n=-2,9075$  (*ecuación 3*) y un valor del parámetro  $C=0,8183$  (*ecuación 4*). El coeficiente de determinación entre la cantidad de publicaciones e investigadores tuvo un valor de 82,24 % (*figura 2*). Resultó que la máxima desviación absoluta 0,08881 fue menor al valor crítico 0,17892, por lo cual se rechazó la hipótesis nula (véase *tabla 4*), razón por la que se comprobó que se cumplió el modelo de Lotka para el área de Ciencias Sociales.

### Ecuación 3

$$n = \frac{4(0,32) - (1,38)(2,66)}{4(0,68) - (1,38)^2} = -2,9075$$

### Ecuación 4

$$C = \frac{1}{\sum_{x=1}^3 \left( \frac{1}{x^{2,9075}} \right) + \frac{1}{(2,9075-1)(4)} + \frac{1}{2(4)^{2,9075}} + \frac{2,9075}{24(4-1)(2,9075+1)}} = \frac{1}{1,2221} = 0,8183$$

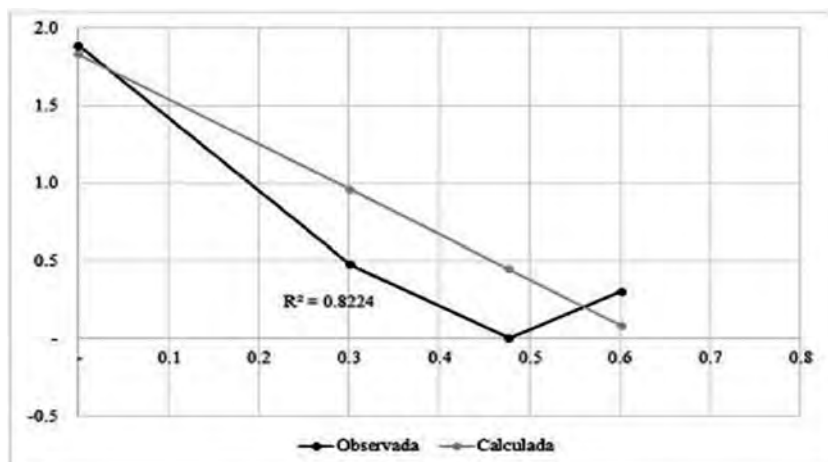


Figura 2. Distribución de productividad de investigadores observado y esperados en el área de Ciencias Sociales (escala logarítmica)  
Fuente: Elaboración propia a partir de datos relevados en este estudio.

Cantidad de contribuciones x	Distribución observada					Distribución esperada según el model de Lotk		Prueba de ajuste según K-S	
	Cantidad de investigadores observados y	Cantidad de investigadores observados acumulado	Cantidad de investigadores observados acumulado relativo $s(x)$	Cantidad de investigadores esperados	Cantidad de investigadores esperados acumulado	Cantidad de investigadores esperados acumulado relativo $f(x)$	$f(x) - s(x)$	$ f(x) - s(x) $	
1	77	77	0,92771	68	68	0,83890	-0,08881	0,08881	
2	3	80	0,96386	9	77	0,95071	-0,01315	0,01315	
3	1	81	0,97590	3	80	0,98510	0,00920	0,00920	
4	2	83	1,00000	1	81	1,00000	0,00000	0,00000	

Tabla 4. Distribución de contribuciones por investigador y comprobación estadística del modelo matemático de Lotka en el área de Ciencias Sociales. Período 2015-2020.  
Fuente: Elaboración propia a partir de datos relevados en este estudio.

## Ciencias Veterinarias

En el área de Ciencias Veterinarias se registraron 120 investigadores que generaron 38 publicaciones, en la cual los pequeños productores representan el 77,5 %. Por su parte, un único investigador presentó cinco publicaciones, que correspondió al número más elevado de contribuciones (véase *tabla 5*).

La comprobación del modelo de Lotka para el área de Ciencias Veterinarias produjo un valor de  $n=-2,5917$  (*ecuación 5*), un parámetro  $C=0,7643$  (*ecuación 6*) y un coeficiente de determinación de 96,63 % (*figura 3*). En vista de que la máxima desviación absoluta 0,02031 resultó menor al valor crítico 0,14880 la hipótesis nula no fue rechazada (véase *tabla 5*). Por consiguiente, se comprobó el cumplimiento del modelo de Lotka en esta área.

### Ecuación 5

$$n = \frac{5(1,12) - (2,08)(4,59)}{5(1,17) - (2,08)^2} = -2,5917$$

### Ecuación 6

$$C = \frac{1}{\sum_{x=1}^4 \left( \frac{1}{x^{2,5917}} \right) + \frac{1}{(2,5917-1)(5)} + \frac{1}{2(5)(2,5917-1)} + \frac{1}{24(5-1)(2,5917+1)}} = \frac{1}{1,3084} = 0,7643$$

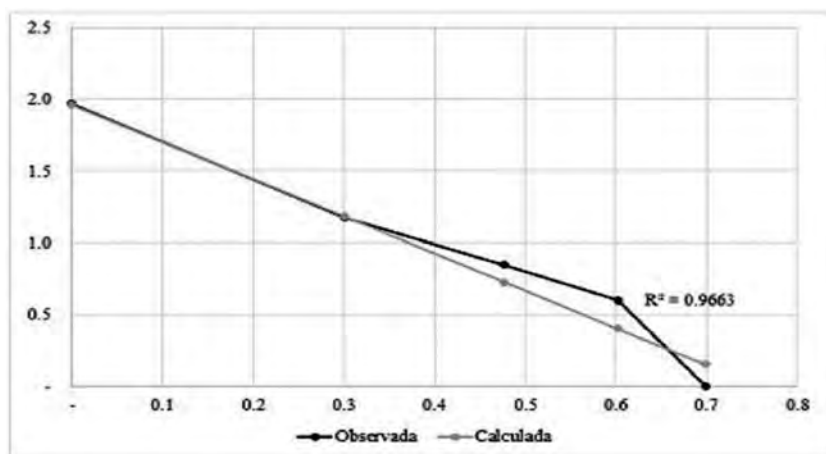


Figura 3. Distribución de productividad de investigadores observados y esperados en el área de Ciencias Veterinarias (escala logarítmica)  
Fuente: Elaboración propia a partir de datos relevados en este estudio

Cantidad de contribuciones $x$	Distribución observada			Distribución esperada según el modelo de Lotka			Prueba de ajuste según K-S	$f(x) \cdot s(x)$
	Cantidad de investigadores observados $y$	Cantida de investigadores observados acumulados	Cantidad de investigadores observados acumulados relativo $s(x)$	Cantidad de investigadores esperados	Cantidad de investigadores esperados acumulados	Cantidad de investigadores esperados acumulados relativo $f(x)$		
1	93	93	0,77500	92	92	0,78936	0,01436	0,01436
2	15	108	0,90000	15	107	0,92031	0,02031	0,02031
3	7	115	0,95833	5	112	0,96609	0,00776	0,00776
4	4	119	0,99167	3	115	0,98782	-0,00385	0,00385
5	1	120	1,00000	1	116	1,00000	0,00000	0,00000

Tabla 5. Distribución de contribuciones por investigador y comprobación estadística del modelo matemático de Lotka en el área de Ciencias Veterinarias. Período 2015-2020

Fuente: Elaboración propia a partir de datos relevados en este estudio.

## Ciencias Agrícolas

En las Ciencias Agrícolas se encontró que 70 investigadores generaron 36 publicaciones, donde los pequeños productores resultaron aproximadamente 77 %, un valor muy cercano al obtenido en este estudio para el área de Ciencias Veterinarias. Por su parte, solo dos investigadores poseían la cantidad máxima de contribuciones, que ascendió a un valor de cuatro (véase *tabla 6*).

Se obtuvo un valor de  $n=-2,4609$  (*ecuación 7*) y un valor del parámetro  $C=0,7365$  (*ecuación 8*). El coeficiente de determinación entre la cantidad de publicaciones e investigadores fue de 99,08 % (*figura 4*). Considerando que la máxima desviación absoluta 0,00885, que en comparación con el valor crítico 0,19482 fue menor, la hipótesis nula no fue rechazada (véase *tabla 6*). De ahí que, se comprobó que el modelo de Lotka se cumplió para el área de Ciencias Agrícolas.

### Ecuación 7

$$n = \frac{4(0,72) - (1,38)(3,55)}{4(0,68) - (1,38)^2} = -2,4609$$

### Ecuación 8

$$C = \frac{\sum_{x=1}^3 \left( \frac{1}{x^2,4609} \right) + \frac{1}{(2,4609-1)(4)} + \frac{1}{2(4)^2,4609} + \frac{2,4609}{24(4-1)(2,4609+1)}}{1} = \frac{1}{1,3577} = 0,7365$$

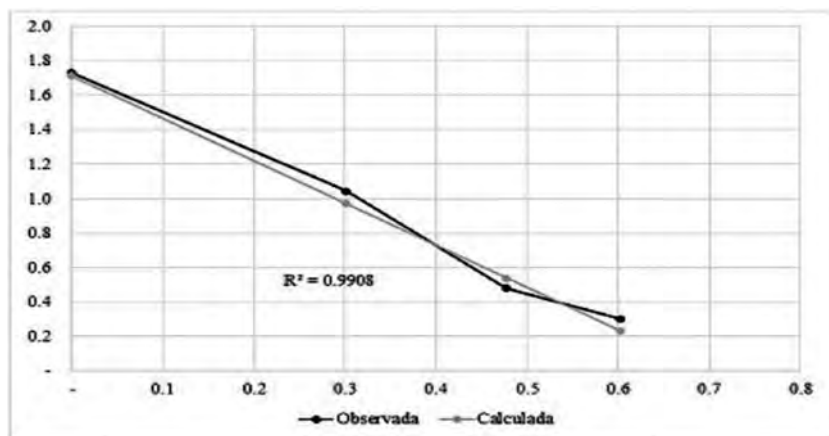


Figura 4. Distribución de productividad de investigadores observados y esperados en el área de Ciencias Agrícolas (escala logarítmica)  
Fuente: Elaboración propia a partir de datos relevados en este estudio.



Cantidad de contribuciones x	Distribución observada			Distribución esperada según el modelo de Lotka			Prueba de ajuste según K-S	
	Cantidad de investigadores observados y	Cantidad de investigadores observados acumulado	Cantidad de investigadores observados relativo s(x)	Cantidad de investigadores esperados	Cantidad de investigadores esperados acumulado	Cantidad de investigadores esperados relativo f(x)	$f(x) - s(x)$	$f(x) \cdot s(x)$
1	54	54	0,77143	52	52	0,78028	0,00885	0,00885
2	11	65	0,92857	9	61	0,92201	-0,00657	0,00657
3	3	68	0,97143	3	64	0,97426	0,00283	0,00283
4	2	70	1,00000	2	66	1,00000	0,00000	0,00000

Tabla 6. Distribución de contribuciones por investigador y comprobación estadística del modelo matemático de Lotka en el área de Ciencias Agrícolas. Periodo 2015-2020  
Fuente: Elaboración propia a partir de datos relevados en este estudio.

A continuación, se presenta un cuadro comparativo del modelo de Lotka por áreas de la ciencia.

Áreas de la ciencia	Coficiente de variación ( $R^2$ )	Valor de $n$	Valor de $C$	Valor crítico Prueba K-S	Desviación máxima	Ajuste al modelo de Lotka
Ciencias Médicas y de la Salud	0,9267	-2,0916	0,6422	0,05175	0,02050	Se ajusta
Ciencias Sociales	0,8224	-2,9075	0,8183	0,17892	0,08881	Se ajusta
Ciencias Veterinarias	0,9663	-2,5917	0,7643	0,14880	0,02031	Se ajusta
Ciencias Agrícolas	0,9908	-2,4609	0,7365	0,19482	0,00885	Se ajusta

*Tabla 7.* Comparación del modelo de Lotka según áreas de la ciencia  
Fuente: Elaboración propia a partir de datos relevados en este estudio.

## DISCUSIÓN

La cantidad de investigadores con filiación a la UNA según área de la ciencia evidenció un comportamiento diferenciado en comparación a la distribución de investigadores (personas físicas) de Paraguay para el año 2020, según datos preliminares, la mayoría de los investigadores declarados en Actividades Científicas y Tecnológicas (ACT) pertenecen al área de Ciencias Sociales, aproximadamente 24 %, y en segundo lugar el área de Ciencias Médicas y de la Salud, con alrededor de 21 % (Consejo Nacional de Ciencia y Tecnología 2020). De hecho, estas cifras no son estrictamente comparables en consideración a la metodología que implica la concepción de las áreas de la ciencia, pero permite dar cuenta que a nivel nacional las dos principales áreas de incursión de los investigadores son las Ciencias Médicas y de la Salud y las Ciencias Sociales.

En cuanto a la distribución de las publicaciones por área de la ciencia los resultados coinciden con los reportados por la Red de Indicadores de Ciencia y Tecnología Iberoamericana e Interamericana (2019) en el periodo 2015-2019, conforme a la base de datos bibliográfica Scopus, con un valor aproximado de 40 %, cuya área principal fue la de Ciencias de la Salud. Asimismo, un estudio cuantitativo realizado sobre los resultados de investigación de la Universidad Rashttrasant Tukadoji Maharaj Nagpur reveló una escasa producción investigativa en las Ciencias Sociales, así como en Humanidades, en comparación con demás áreas temáticas (Bapte y Kherde 2019), lo cual se asemeja a lo obtenido en este estudio.

En lo referido a la comprobación del modelo de Lotka en el área de Ciencias Médicas y de la Salud, al igual que en los estudios de Cáceres Ruiz Díaz (2019, 51) y Urbizagástegui Alvarado (2002, 18), se comprobó el cumplimiento del modelo de Lotka en su forma general, pero con un valor de  $n$  mucho mayor en comparación con el primer estudio, que mostró  $n=-1,2058$ , y más cercano al segundo, que reportó un valor de  $n=-2,76$  para el área de Medicina. Además, en un estudio sobre la aplicabilidad de la ley de Lotka en la rama de la Parasitología de la India, se comprobó el ajuste de la literatura en este campo a dicha ley (Naheem, Sivaraman y Saravanan 2019).

Por otra parte, en la comprobación del modelo en el área de las Ciencias Sociales, resulta significativo traer a colación un estudio sobre productividad realizado dentro de un campo particular de esta área, el de la Bibliotecología y Ciencias de la Información, el cual mostró que la distribución de la productividad de autores se ajusta a ley de Lotka para la literatura correspondiente a India (Bisaria 2020). De manera análoga, en una investigación sobre publicaciones realizadas en una revista de documentación se comprobó la aplicabilidad de la ley de Lotka (Kumar 2017). Asimismo, en un trabajo hecho sobre patrones de autoría se concluyó que la distribución de Lotka es válida para la literatura de negocios internacionales (Naqvi y Fatima 2017). Lo expuesto ofrece indicios de la aplicabilidad de la ley de Lotka en el área de las Ciencias Sociales, de forma coincidente con los hallazgos del presente estudio.

En contrapartida, la comprobación del cumplimiento del modelo de Lotka en el área de Ciencias Veterinarias, se contrapone al hallazgo de Urbizagástegui Alvarado (2002), que comprobó que las publicaciones de los profesores de la Escola de Veterinária de la UFMG (Brasil) no se ajusta al cuadrado inverso del modelo de Lotka en el área de Veterinaria.

En tanto que, el cumplimiento del modelo de Lotka en el área de Ciencias Agrícolas, se da de manera semejante a lo reportado en un trabajo realizado por Urbizagástegui Alvarado (2005), en el cual se aplicó la Ley de Lotka por el método del «poder inverso generalizado» a datos de un estudio realizado por Oliveira (1983), éste halló un ajuste de la literatura estudiada al modelo de Lotka. Cabe señalar que Oliveira estudió el comportamiento de la literatura sobre jaca, a partir de resúmenes del ámbito agrícola.

Considerando que la ciencia posee productividad que se encuentra sesgada, es de vital relevancia que los recursos disponibles se encuentren orientados a los investigadores que conforman el grupo élite en un campo de la ciencia (González de Dios, Moya y Mateos Hernández 1997).

La relevancia del modelo de Lotka, según lo plantea Rau (2011, 214), radica en «una aplicación práctica del cumplimiento o no de esta ley, podría traducirse en la propuesta de fondos gubernamentales para el desarrollo de un área científica».

## CONCLUSIONES

La comprobación del modelo de Lotka en su forma generalizada, con el correspondiente cálculo del valor de  $n$  por el método de los mínimos cuadrados y el valor de  $C$ , mediante la aplicación de la prueba Kolmogorov-Smirnov evidencia un ajuste de la producción científica de los investigadores con filiación a la UNA, en revistas paraguayas indexadas en SciELO, en las áreas estudiadas, con respecto al postulado teórico del modelo, a un nivel de 1 % de significación.

En las cuatro áreas de estudio se observa una importante proporción de autores esporádicos, autores que publican en una sola ocasión, cuyos valores superan al predicho por Lotka (1926) que en su estudio presenta un valor que ronda en 60 %. Esto implica que predominan los pequeños productores, es decir, la mayor parte manifiesta baja productividad, en contrapartida, representa una pequeña cantidad de éstos la que agrupa la mayor proporción de las publicaciones.

Aunque la metodología para la comprobación del modelo de Lotka debería preferentemente realizarse en una temática específica y en un lapso de tiempo considerable, al menos 10 años, se ha demostrado en este estudio que también resulta plausible la comprobación del modelo considerando áreas de la ciencia y un rango temporal no muy amplio, en este caso seis años.

## REFERENCIAS

- Bapte, Vishal Dattatray y Mohan R. Kherde. 2019. An Institutional Collaboration and Application of Lotka's Law to the Research Output of Rashtrasant Tukadoji Maharaj Nagpur. *Library Philosophy and Practice (e-Journal)*.
- Barceló, Grettel, Eduardo Cendejas, Igor Bolshakov y Grigori Sidorov. 2009. Ambigüedad en nombres hispanos. *Revista Signos* 42(70): 153-169.  
<https://www.redalyc.org/articulo.oa?id=157013778001>
- Bisaria, Garima. 2020. Lotka's Law and Authorship Trends in Library and Information Science: A Study Based on Select Journals of India, US and UK. *Library Philosophy and Practice (e-journal)*: 1-22.
- Cáceres Ruiz Díaz, Mariana. 2019. Productividad científica de los investigadores categorizados del área de Ciencias Médicas y de la Salud: aplicación del modelo matemático de Lotka. *Memorias del Instituto de Investigaciones en Ciencias de la Salud* 17(2): 44-55.  
<http://scielo.iics.una.py/pdf/iics/v17n2/1812-9528-iics-17-02-44.pdf>
- Canavos, George C. 2001. *Probabilidad y Estadística*. México: McGraw-Hill.
- Centro Universitario de Liderazgo y Tecnología Avanzada. 2016. Metodología de limpieza de datos con la herramienta de OpenRefine.  
<http://www.visualix.mx/files/5eb360388714341c277936ecb036e370.pdf>

- Consejo Nacional de Ciencia y Tecnología. 2020. *Indicadores de Ciencia y Tecnología de Paraguay* 2020.  
[https://www.conacyt.gov.py/sites/default/files/upload\\_editores/u489/Resultados-ACT2020-Indicadores-CyT\\_Paraguay-2020.pdf](https://www.conacyt.gov.py/sites/default/files/upload_editores/u489/Resultados-ACT2020-Indicadores-CyT_Paraguay-2020.pdf)
- Egghe, L. 2012. Theoretical evidence for empirical findings of A. Pulgarin on Lotka's law. *Malaysian Journal of Library & Information Science* 17(3): 1-15.  
<https://ajba.um.edu.my/index.php/MJLIS/article/view/6637>
- González de Dios, J., M. Moya y M. A. Mateos Hernández. 1997. Indicadores bibliométricos: características y limitaciones en el análisis de la actividad científica. *Anales Españoles de Pediatría* 47(3): 235-244.  
<https://www.aeped.es/sites/default/files/anales/47-3-3.pdf>
- Gorbea Portal, Salvador. 2005. *El modelo matemático de Lotka: su aplicación a la producción científica latinoamericana en ciencias bibliotecológica y de la información*. México: Universidad Nacional Autónoma de México, Centro Universitario de Investigaciones Bibliotecológicas.  
[http://ru.iibi.unam.mx/jspui/bitstream/IIBI\\_UNAM/L82/1/modelo\\_matematico\\_lotka.pdf](http://ru.iibi.unam.mx/jspui/bitstream/IIBI_UNAM/L82/1/modelo_matematico_lotka.pdf)
- Kumar, Suresh. 2017. Author Productivity and the Application of Lotka's Law in LIS Publications. *Annals of Library and Information Studies* 64(4): 234-241.
- Lotka, Alfred. J. 1926. The Frequency Distribution of Scientific Productivity. *Journal of the Washington Academy of Sciences* 16(12): 317-323.
- Naheem, K. T., P. Sivaraman y G. Saravanan. 2019. Applicability of Lotka's Law in Parasitology Research Output of India. *Library Philosophy and Practice* (e-journal).
- Naqvi, S. H. y N. Fatima. 2017. Authorship Patterns in International Business Literature: Applicability of Lotka's Law. *Annals of Library and Information Studies* 64(4): 253-259.
- OCDE. 2018. Manual de Frascati 2015: *Guía para la recopilación y presentación de información sobre la investigación y el desarrollo experimental*. Madrid: FECYT.  
<https://doi.org/10.1787/9789264310681-es>
- Oliveira, Silas Marques. 1983. Aplicação da lei de produtividade de autores de Lotka á literatura de Jaca. *Revista de Biblioteconomia de Brasília* 11(1): 125-130.  
<https://1library.org/document/yrwg8lpz-aplicacao-da-lei-produtividade-autores-lotka-literatura-jaca.html>
- Potter, W. G. 1981. Lotka's Law revisited. *Library Trends* 31: 21-29.
- Pulgarín, A. 2012. Dependence of Lotka's law parameters on the scientific area. *Malaysian Journal of Library & Information Science* 17(1): 41-50.  
<https://mjlis.um.edu.my/index.php/MJLIS/article/view/1881>
- Rau, Jaime R. 2011. ¿Sigue la producción de artículos ISI de los ecólogos chilenos (sensu lato) la ley de Lotka (1926)? *Revista Chilena de Historia Natural* 84: 213-216.  
<http://dx.doi.org/10.4067/S0716-078X2011000200007>
- Red de Indicadores de Ciencia y Tecnología Iberoamericana e Interamericana. 2019. Publicaciones. Acceso el 26 de enero de 2022.  
[http://app.riicyt.org/ui/v3/bycountry.html?country=PY&subfamily=CTI\\_BIB&start\\_year=2010&end\\_year=2019](http://app.riicyt.org/ui/v3/bycountry.html?country=PY&subfamily=CTI_BIB&start_year=2010&end_year=2019)
- Urbizagástegui Alvarado, Rubén. 1999. La ley de Lotka y la literatura de bibliometría. *Investigación Bibliotecológica* 13(27): 125-141.  
<http://dx.doi.org/10.22201/iibi.0187358xp.1999.27.3913>

- \_\_\_\_\_. 2002. A Lei de Lotka na bibliometria brasileira. *Ciência da Informação* 31(2): 14-20.  
<https://doi.org/10.1590/S0100-19652002000200002>
- \_\_\_\_\_. 2005. La productividad científica de los autores. Un modelo de aplicación de la Ley Lotka por el método del poder inverso generalizado. *Información, cultura y sociedad* 12: 51-73.  
<http://repositorio.filo.uba.ar/bitstream/handle/filodigital/8017/n12a04.pdf?sequence=1&isAllowed=y>

*Para citar este texto:*

Sena Correa, Emilce, María Luisa Lascurain, Mariana Cáceres Ruiz Díaz y Johana Raquel Pineda Alvarenga 2023. “Productividad científica de investigadores de la Universidad Nacional de Asunción (UNA): comprobación del modelo de Lotka”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 170-200.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58793>

# **Arquivos contestados em contexto intranacional Em torno do caso da Região Autónoma da Madeira (Portugal)**

L. S. Ascensão de Macedo\*

*Artículo recibido:*  
9 de mayo de 2023  
*Artículo aceptado:*  
1 de agosto de 2023

*Artículo de investigación*

## **RESUMO**

O artigo discute o fenómeno dos «arquivos deslocados», um dos problemas mais complexos da arquivística contemporânea. Embora os casos internacionais tenham sido bem estudados, o objetivo deste artigo consiste em analisar um caso subnacional, menos abordado na produção científica a partir da representação destes arquivos em instrumentos de acesso à informação. O artigo concentra-se num caso específico de disputa de custódia de arquivos entre a Região Autónoma da Madeira (ABM) e o Arquivo Nacional da Torre do Tombo (ANTT) em Portugal. Metodologicamente, este artigo caracteriza-se por uma investigação qualitativa. A pesquisa utilizou fontes documentais e arquivísticas para explorar estratégias de representação de conjuntos documentais em disputa,

\* Centro de estudios Clásicos, Facultad de letras, Universidade de Lisboa, Portugal  
laureanomacedo@edu.ulisboa.pt

com base em dados sobre a representação do conteúdo em instrumentos de acesso à informação publicados pelo ANTT e pelo ABM. Os resultados apontam para o facto de as entidades custodiantes não terem envidado esforços para alcançar uma representação da informação técnica e cientificamente reunificada. O estudo recomenda a concessão de novas garantias literárias para a descrição de conjuntos documentais deslocados que potenciem a sua reunificação.

**Palavras-chave:** Representação da informação; Arquivos deslocados; Disputas subnacionais; Instrumentos de acesso à informação.

### **Archivos disputados en contexto intranacional Alrededor del caso de la Región Autónoma de Madeira (Portugal)**

*L. S. Ascensão de Macedo*

#### **RESUMEN**

El artículo discute el fenómeno de los «archivos desplazados», uno de los problemas más complejos en la archivística contemporánea. Si bien los casos internacionales han sido bien estudiados, el objetivo de este artículo es analizar un caso intranacional, menos abordado en la producción científica, desde la perspectiva de la representación de estos archivos en instrumentos de descripción. El artículo se centra en un caso específico de disputa de custodia de archivos entre la Región Autónoma de Madeira (ABM) y el Archivo Nacional «Torre do Tombo» (ANTT) en Portugal. Metodológicamente, este artículo se caracteriza por una investigación cualitativa. La investigación utilizó fuentes documentales y archivísticas para explorar estrategias de representación de fondos documentales en disputa, basadas en datos sobre la representación de contenido en instrumentos de descripción archivística publicados por ANTT y ABM. Los resultados señalan que las entidades custodias no han hecho esfuerzos para lograr una representación técnicamente y científicamente reunificada de la información. El estudio concluye que es importante desarrollar garantías literarias para la descripción de archivos desplazados.

**Palabras clave:** Representación de la información; Archivos desplazados; Reclamaciones subnacionales; Instrumentos de búsqueda.



# Subnational disputed archives Around the case of the Autonomous Region of Madeira (Portugal)

L. S. Ascensão de Macedo

## ABSTRACT

The article discusses the phenomenon of «displaced archives», one of the most complex problems in contemporary Archival Science. While international cases have been well studied, the aim of this article is to analyze a subnational case, which is less addressed in scientific production, from the perspective of representation of these archives in finding aids. The article focuses on a specific case of a custody dispute over archives between the Autonomous Region of Madeira (ABM) and the National Archives «Torre do Tombo» (ANTT) in Portugal. Methodologically, this article is characterized by unobtrusive qualitative research techniques. The research used documentary and archival sources to explore strategies for representing fonds or collections in dispute, based on data on content representation in finding aids published by ANTT and ABM. The results point to the fact that custodial entities have not made efforts to achieve a technically and scientifically reunified representation of information. The study concludes that it is important to develop literary warrants for the description of displaced archives.

**Keywords:** Information representation; Displaced archives; Subnational claims; Finding aids.

## INTRODUÇÃO

Em 2020, o Grupo de Peritos para o Património Arquivístico Partilhado do Conselho Internacional dos Arquivos emanou um relatório elaborado por James Lowry (doravante, *Relatório* de Lowry) (International Council on Archives 2020) que atualiza o ponto da situação do contencioso arquivístico internacional. A remoção dos arquivos dos seus contextos sociais e territoriais originários, independentemente dos seus factores incoativos, e as reivindicações pela sua restituição constituem um dos problemas mais complexos da arquivística contemporânea. O fenómeno tem sido denominado pela comunidade arquivística internacional como arquivos deslocados –apesar de coexistirem outras denominações sinónimas, com definições nem sempre consistentes– entendidos como

«arquivos removidos do local da sua criação, onde a posse e propriedade dos arquivos são disputadas por duas ou mais partes» (International Council on Archives 2020, 5).

Durante muito tempo, o fenómeno subjacente aos arquivos deslocados permaneceu associado a casos internacionais, por força dos estudos produzidos por Kecskeméti (1977) e por Auer (1998). Segundo Lowry (2017), a via jurídica revelou-se, na maior parte das vezes, ineficaz na resolução destes conflitos, considerando os poucos casos resolvidos face à emergência de novos pedidos de restituição, desta vez com contornos muito distintos. Embora a natureza deste conflito possa estar associada a fenómenos derivados do conflito armado, da secessão/irredentismo, do colonialismo, das migrações forçadas/diáspora, do tráfico ilícito de bens culturais, estes eventos também podem ocorrer em contexto subnacional, com a agravante de as remoções poderem também ocorrer *ope legis* pela ação das administrações centrais (Macedo 2019).

Entre os diversos casos emergentes destacam-se os conflitos pela custódia de arquivos em contexto subnacional ou intranacional, ou seja, as reivindicações pela restituição de arquivos dentro de um mesmo país. De acordo com o *Relatório* de Lowry, emergiram dois casos: a Região Autónoma da Gronelândia *versus* Reino da Dinamarca e a Região Autónoma da Madeira *versus* Portugal. Embora os casos apresentados pela Região Autónoma da Gronelândia se apresentem como remoções derivadas da postura colonialista do governo dinamarquês em relação àquela comunidade, a Região Autónoma da Madeira denomina o fenómeno como um caso intranacional. Não obstante existirem diversos conflitos pela custódia de arquivos em contexto subnacional ou intranacional no mundo, como o caso dos «Papeles de Salamanca» em Espanha ou a disputa pela domiciliação legal de arquivos entre o Arquivo Público Municipal de Ouro Preto e o Arquivo Público Mineiro, no Brasil, estes não se encontram reportados no *Relatório* de Lowry, contribuindo, em certa medida, para a invisibilidade do fenómeno em apreço ao nível internacional.

Este artigo versa-se sobre um caso concreto referente aos pedidos de restituição de arquivos reivindicados pela Região Autónoma da Madeira (RAM) ao Arquivo Nacional da Torre do Tombo (ANTT). Reconhecendo que os conflitos pela custódia de arquivos em contexto subnacional constituem uma dimensão pouco abordada na produção científica, a nossa análise incide especialmente sobre as estratégias de representação da informação dos conjuntos documentais reivindicados pela RAM em instrumentos de acesso à informação (doravante, IAI), de qualquer materialização genológica. Se para Winn (2015) a inexistência de IAI constitui um dos factores limitadores para a identificação de casos de arquivos na condição de deslocados, é importante, no caso de estes *instrumenta* existirem, conhecer como estes conjuntos documentais têm sido representados

pelas entidades custodiantes, atendendo que as disputas pela custódia de arquivos não se restringem apenas a questões jurídicas sobre posse ou propriedade de bens arquivísticos de conservação permanente, mas também aos critérios de representação de informação adotados pelas entidades custodiantes contestadas.

Assim, partindo do caso escolhido, pretende-se neste artigo responder a duas perguntas de investigação:

- 1) *Quais são os fundos documentais reivindicados pelas autoridades madeirenses ao ANTT?*
- 2) *Como é que estes conjuntos documentais se encontram representados nos IAI do ANTT e da RAM, especialmente no que respeita à forma de aquisição?*

A principal contribuição deste estudo consiste não só em tornar visível a complexidade subjacente ao fenómeno dos arquivos deslocados em contexto subnacional ou intranacional, tomando o caso português como ponto de partida e reconhecendo a escassez de casos congêneres na literatura, mas também em explorar abordagens metodológicas alternativas que permitam conhecer a causalidade do fenómeno, identificando as suas características diferenciadoras.

## METODOLOGIA

Para responder às questões de investigação supramencionadas, realizou-se uma investigação de natureza qualitativa de tipo documental e arquivística (C. G. da Silva 2021). Esta estratégia possibilita combinar dados sobre a representação do conteúdo de conjuntos documentais em IAI publicados pelo ANTT e pelo ABM com o grau de dispersão das unidades documentais.

Para a identificação dos arquivos reivindicados pela Resolução da Assembleia Legislativa da Madeira n.º 3/2017/M, de 12 de janeiro, a análise de IAI possibilita-nos explorar estratégias de representação dos conjuntos documentais em disputa. Para Ribeiro (2013), a representação da informação arquivística materializada em IAI (genologicamente materializados em catálogos, inventários, guias, índices ou bases de dados) continua a ser aplicada pela comunidade profissional para servir uma dupla função, por um lado, para «organizar/arrumar (fisicamente) documentos» e, por outro, para «representar/recuperar informação». Neste caso específico, para a questão (1) a nossa análise incide na extração de dados referentes à zona de identificação (metadados *Código(s) de referência, Título, Data(s), Nível de descrição, Dimensão e suporte*) da norma internacional ISAD(G) (Conselho Internacional dos Arquivos 2002). Do ponto de vista do nível de descrição, cingir-nos-emos ao fundo. Para a contextualização dos processos de aquisição,

interessa-nos particularmente o metadado «fonte imediata de aquisição ou transferência» da zona de contexto da norma referenciada para responder à questão de investigação (2). Convém reconhecer que os IAI têm evoluído consideravelmente, tanto em suporte analógico como em suporte digital (MacNeil 2012), de modo que os dados disponibilizados nos portais institucionais pelas entidades custodiantes podem ser objeto de modificação.

A pesquisa decorreu através de uma recolha em bases de dados ou listagens disponíveis no *Portal Português de Arquivos* (Direção-Geral do Livro dos Arquivos e das Bibliotecas 2017) e no portal da Direção Regional do Arquivo e Biblioteca da Madeira (Arquivo Regional e Biblioteca Pública da Madeira 2017).

Nos motores de pesquisa do portal institucional *web do Portal Português de Arquivos* utilizamos descritores para recuperar informação em título de fundo, utilizando a denominação dos fundos enumerados na Resolução supra exarada, combinados com «Madeira» e «Funchal». Os resultados foram extraídos e processados em .xlsx (Microsoft Excel). Para suporte à análise de conteúdo, baseamos-nos num conjunto de garantias literárias emanadas pelo Conselho Internacional de Arquivos e pelo órgão nacional dos arquivos (Direção-Geral dos Arquivos 2011), atendendo às recomendações para a descrição dos metadados supra indicados. Uma vez extraídos os dados, complementa-se com investigação documental, com recurso à legislação portuguesa (Assembleia da República 2010; Imprensa Nacional-Casa da Moeda 2016) e à investigação arquivística dos fundos dos «arquivos dos arquivos» (Arquivo Nacional da Torre do Tombo 2011; Governo Regional da Madeira 2021).

## RESULTADOS

### ***Quais são os fundos reivindicados pelas autoridades madeirenses ao ANTT?***

A Resolução n.º 3/2017/M, de 12 de janeiro, enumerou cinco fundos a *transferir* para a RAM: «Cabido da Sé do Funchal, do Convento de Santa Clara, do Convento da Encarnação, da Provedoria da Real Fazenda e da Alfândega do Funchal» (Assembleia Legislativa da Região Autónoma da Madeira 2017, 409). Contudo, os resultados da pesquisa efetuada no portal institucional devolveram onze fundos madeirenses na custódia do ANTT, a saber: *Convento de Santa Clara do Funchal* (CSCF); *Convento de Nossa Senhora da Encarnação do Funchal* (CNSEF); *Convento de São Francisco do Funchal* (CSFF); *Cabido da Sé do Funchal* (CSF); *Convento de Nossa Senhora da Piedade de Santa Cruz* (CNSPSC); *Convento de São Bernardino de Câmara de Lobos* (CSB-CL); *Convento de Nossa Senhora da Porciúncula da Ribeira Brava* (CNSPRB); *Convento de*

*São Sebastião da Calheta* (CSSC); *Provedoria e Junta da Real Fazenda do Funchal* (PJRFF); *Comissão da Fazenda do Distrito da Madeira e Porto Santo* (CFMPS) e *Alfândega do Funchal* (ALF). Foram devolvidos subfundos *Confraria das Escravas de Nossa Senhora do Monte* (CSCF), *Irmadade da Ordem Terceira de São Francisco* (CSSC) e *Confraria de Nossa Senhora Mãe dos Homens e do Patriarca São José* (CSFF). Foi igualmente devolvido pelos motores de pesquisa do ANTT sub-fundos como *Delegação do Funchal da PIDE-DGS* e *Comissão Distrital do Funchal da União Nacional*, que se excluiu, não só por não estar mencionada na citada Resolução.

Numa primeira análise, o facto de a Resolução apenas enumerar cinco fundos só pode ter justificação por motivo de o legislador ter recorrido a um IAI anterior ao ano de 2000. Esta discrepância de número de fundos deriva do desconhecimento da intervenção feita pelo ANTT, em 2002, quando se empreendeu uma profunda reorganização dos fundos conventuais, cujo inventário (Instituto dos Arquivos Nacionais/Torre do Tombo 2002) foi coordenado pela arquivista madeirense Maria do Carmo Jasmins Dias Farinha (1937-2014), sob a direção José Mattoso. Basicamente, foram individualizados seis novos fundos que permaneceram indistintos no fundo da PJRFF desde a sua transferência para Lisboa em finais do século XIX, mais concretamente os conventos masculinos (CFMPS, CNSPRB, CNSPSC, CSBCL, CSFF e CSSC) (Arquivo Nacional da Torre do Tombo 2002).

Importa referir que, desde 1937, o Arquivo Distrital do Funchal (entidade antecessora da actual Direção Regional do Arquivo e Biblioteca da Madeira) procurou inventariar no ANTT a documentação reivindicada, cujos resultados foram publicados no boletim *Arquivo Histórico da Madeira* (1939d; 1939c; 1939e; 1939b; 1939a). Os fundos ALF, PJRFF, CNSEF e CSCF foram microfilmados e parcialmente digitalizados a partir do microfilme, projeto conduzido pelo Centro de Estudos de História do Atlântico, organismo do Governo Regional da Madeira, cuja base de dados *Nesos* (Governo Regional da Madeira 1998) se encontra atualmente descontinuada.

O portal institucional da Direção Regional do Arquivo e Biblioteca da Madeira (Arquivo Regional e Biblioteca Pública da Madeira 2017), por seu turno, apenas disponibiliza um inventário digitalizado e não dispõe de qualquer representação digital sobre os conjuntos documentais que custodiam nas bases de dados institucionais. De igual forma, não dispõe de um IAI atualizado há várias décadas. A tabela seguinte enumera os conjuntos documentais constantes desta entidade, partindo da informação unicamente providenciada através do guia impresso (Arquivo Regional da Madeira 1997).

Conforme se pode verificar através do *Gráfico 1*, os conjuntos documentais foram incorporados no Arquivo Distrital do Funchal (entidade da administração central desconcentrada pré-autonómica) desde 1937 (CSFF, CNSEF, CSBCL, CSCF, CNSPSC). Sobre a PJRFF, não foi documentado qualquer auto de

incorporação anterior a 1957 e, no que respeita ao fundo ALF, a documentação foi incorporada ao longo dos anos entre 1951 até 2012 (Arquivo Regional da Madeira 1997). Apesar da amplitude cronológica, alguns de vários séculos, houve perdas documentais significativas ocorridas no passado devido a factores humanos e naturais, que se pode depreender através de lapsos cronológicos. Os IAI disponibilizados pelo ABM constituem simplesmente listagens de unidades de instalação, que não cumprem com as normas internacionais de descrição arquivística, sem representações digitais destas unidades nas bases de dados desta entidade. Assim, podemos concluir, através desta breve análise, que o problema presente é o da reunificação de fundos, como o gráfico seguinte elucida:

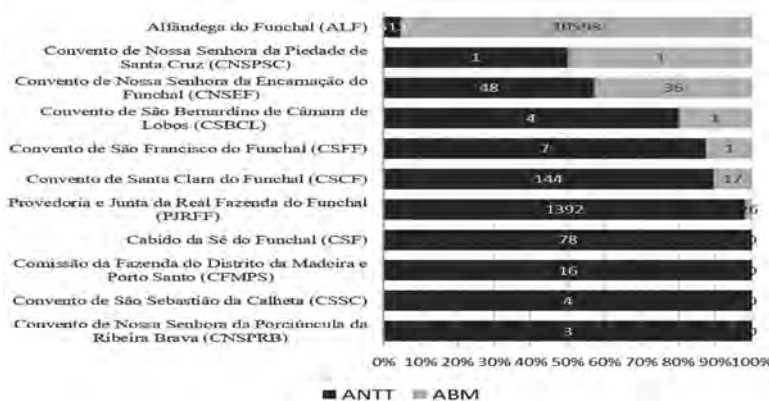


Gráfico 1. População de fundos custodiados no ANTT e no ABM ( $n= 12890$  u. i.)  
 Fonte: elaborado a partir do portal do ABM (Arquivo Regional e Biblioteca Pública da Madeira 2017) e Portal Português de Arquivos (Direção-Geral do Livro dos Arquivos e das Bibliotecas 2017).

Através do Gráfico 1, é possível visualizar a existência de dois grupos distintos, a saber:

- (1) Grupo 1 (G1): sete fundos repartidos entre o ABM e ANTT ( $n= 12789$ ): ALF (ABM [1650/2000],  $n= 10598$  u. i. *versus* ANTT [1620/1834],  $n= 513$  u. i.); PJRF (ABM [1649/1833],  $n= 26$  u. i. *versus* ANTT [1569/1834],  $n= 1392$  u. i.); CSCF (ABM [1634/1897],  $n= 17$  u. i. *versus* ANTT [1447/1900],  $n= 144$  u. i.); CNSEF (ABM [1645/1895],  $n= 36$  u. i. *versus* ANTT [1660/1890],  $n= 48$  u. i.); CSFF (ABM [1710/1730],  $n= 1$  u. i. *versus* ANTT [1732/1832],  $n= 7$  u. i.); CSBCL (ABM [1783/1824],  $n= 1$  u. i. *versus* ANTT [1792/1825],  $n= 4$  u. i.); CNSPSC (ABM [1705/1710],  $n= 1$  u. i. *versus* ANTT [1772/1776],  $n= 1$  u. i.).
- (2) Grupo 2 (G2): quatro fundos que se encontram custodiados na totalidade pelo ANTT ( $n= 101$  u. i.), a saber: CSF ( $n= 78$  u. i., 1478/1861); CFMPS ( $n= 16$  u. i., 1834/1851); CSSC ( $n= 4$  u. i., 1674/1811); CNSPRB ( $n= 3$  u. i., 1736/1809).

Com base no exposto, a proporção de fundos entre o ANTT e o ABM reparte-se, respetivamente, entre 16,5 % e 83,5 % face ao total de unidades de instalação identificadas ( $n=12890$ ). O fundo mais extenso é o ALF, com cerca de 86,2 % do volume total de unidades de instalação, e o menos extenso é o CNSPSC (0,01%), ambos do G1.

Independentemente do volume documental de cada fundo, alguns dos quais aparentam ser muito fragmentários (com  $> 10$  u. i.), verifica-se que os conjuntos documentais mais antigos se encontram na custódia do ANTT, especialmente o CSCF e o CSF, que remontam ao século XV e que se estendem por vários séculos. Acresce a isto o facto de, não obstante, os fundos do G1 apresentarem datas que indiciam lapsos temporais em determinadas séries documentais. Embora esta análise se cinja ao nível do fundo, este aspeto pode ser importante, pois sugere que uma representação hipoteticamente reunificada destes conjuntos documentais depende da forma como cada unidade de instalação foi descrita por ambas as entidades custodiantes, assim como as classes intermédias (secção/subsecção e série/subsérie). Evidências empíricas destas ações podem ser encontradas nos IAI emanados pelo ANTT, designadamente na forma como esta entidade custodiante optou não só por desagregar unidades de um fundo para gerar novos fundos –como é o caso dos conventos masculinos (CSFF, CSBCL, CSSC, CNSPSC e CNSPRB) e entidades como a CFMPJ, ambos inicialmente integrados na PJRFF– como também por reconduzir unidades de instalação entre fundos (*u. g.*, PJRFF e ALF e CSCF e CNSEF). Oscilações no cômputo global das unidades de instalação de cada um desses fundos podem limitar a possibilidade de alcançar-se uma representação orgânica e cientificamente reunificada, por motivo da ausência de intervenção por parte do ABM na descrição dos conjuntos documentais que se encontram sob custódia.

Com base no exposto, a identificação de fundos disputados não constitui um processo linear, atendendo que nada indica que a reivindicação feita pelas autoridades insulares tenha partido do conhecimento direto dos conjuntos documentais e dos seus IAI. Além disto, é indispensável analisar genologicamente os IAI emanados pelas entidades custodiantes, sendo perceptível que diversas representações destes fundos disputados resultaram numa erosão causada pela mediação arquivística, especialmente ao fazer prevalecer a perspetiva da entidade custodiante em detrimento de uma representação técnica e cientificamente reunificada com a parte remanescente na Região Autónoma da Madeira. Por exemplo, os IAI que descrevem o CNSEF na custódia do ANTT não mencionam em qualquer um dos metadados a existência da outra parte, também CNSEF, na custódia do ABM. Não obstante, em última análise, o recurso a garantias literárias para a descrição arquivística pode não proporcionar um cenário de representação reunificada (independentemente do *medium*) de fundos dispersos, se na base da sua conceção não estiver o propósito de alcançar a sua reunificação.

***Como é que estes conjuntos documentais se encontram representados nos IAI do ANTT e da RAM, especialmente no que respeita à forma de aquisição?***

O elemento metainformativo «Fonte imediata de aquisição» (FIA) tem como objetivo «identificar a fonte imediata de aquisição ou transferência da unidade de descrição» (Conselho Internacional dos Arquivos 2002, 28). Visa informar «formas de aquisição, ou seja, de obtenção da propriedade e/ou custódia de documentos de arquivo» (Direção-Geral dos Arquivos 2011, 78). Sendo um elemento obrigatório na descrição de nível de fundo em Portugal, mas facultativo para os níveis de descrição inferiores, a informação disponibilizada pelo ANTT sobre os onze fundos reivindicados encontra-se exarada no seguinte *Quadro*:

<b>Fundo</b>	<b>FIA</b>	<b>Fonte</b>
ALF	Incorporação ao abrigo de Portaria do Ministério do Reino.	<a href="https://digitarq.arquivos.pt/details?id=3886618">https://digitarq.arquivos.pt/details?id=3886618</a>
CFMPS	Incorporação ao abrigo de Portaria do Ministério do Reino.	<a href="https://digitarq.arquivos.pt/details?id=3909719">https://digitarq.arquivos.pt/details?id=3909719</a>
CNSPRB	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1375741">https://digitarq.arquivos.pt/details?id=1375741</a>
CNSPSC	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1375743">https://digitarq.arquivos.pt/details?id=1375743</a>
CSBCL	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1379034">https://digitarq.arquivos.pt/details?id=1379034</a>
CSCF	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1379435">https://digitarq.arquivos.pt/details?id=1379435</a>
CSFF	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1379962">https://digitarq.arquivos.pt/details?id=1379962</a>
CSSC	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1382433">https://digitarq.arquivos.pt/details?id=1382433</a>
PJRFF	Incorporação, em 1886, ao abrigo de Portaria do Ministério do Reino.	<a href="https://digitarq.arquivos.pt/details?id=4310201">https://digitarq.arquivos.pt/details?id=4310201</a>
CSF	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1379940">https://digitarq.arquivos.pt/details?id=1379940</a>
CNSEF	[Omite]	<a href="https://digitarq.arquivos.pt/details?id=1375671">https://digitarq.arquivos.pt/details?id=1375671</a>

*Quadro 1* – Listagem dos fundos custodiados pelo ANTT (FIA)

Fonte: Portal Português de Arquivos (Direção-Geral do Livro dos Arquivos e das Bibliotecas 2017).

Conforme se pode verificar no *Quadro 1*, apenas três fundos mencionam explicitamente o critério de aquisição por *incorporação* por diploma, a Portaria do Ministério do Reino de 9 de junho de 1886, enquanto os fundos conventuais não mencionam qualquer modo de aquisição, não se cumprindo, assim, o requisito estipulado pelo órgão de gestão dos arquivos em termos de descrição arquivística.



No entanto, as informações sobre os processos de aquisição destes bens pelo ANTT encontram-se no metadado «História Custodial e Arquivística». Por um lado, o ANTT justifica a aquisição ao abrigo do Decreto de 2 de outubro de 1862 (Ministério da Justiça 1862). Apesar de os restantes fundos integrantes da população terem em comum o facto de partilharem o fatum provocado por esse diploma e pela Portaria de 9 de junho de 1886 (Ministério do Reino), apenas o fundo CNSEF omite as bases legais que determinaram a sua transferência, concretizada a 21 de junho de 1894 (Arquivo Nacional da Torre do Tombo 1894), para o ANTT. Atendendo que este fundo procede de um cenóbio feminino, a Lei de 4 de abril de 1861 (Ministério dos Negócios da Fazenda 1862), que regulou a extinção dos conventos femininos, executando os bens com o falecimento da última religiosa professa, só foi efetivamente executada a 24 de outubro de 1890, passando os conjuntos documentais imediatamente para a custódia da Repartição da Fazenda do Funchal, nos termos do Decreto e Instruções de 31 de Maio de 1862 (Ministério e Secretaria de Estado dos Negócios da Fazenda 1871), e transferido para o ANTT, por força do Decreto de 29 de Dezembro de 1887 (Ministério do Reino 1887), que institucionalizou a Inspeção Superior das Bibliotecas e Arquivos.

O diploma mais problemático é a Portaria de 9 de junho de 1886, do Ministério do Reino, mencionada pelos diversos IAI do ANTT (e replicado pelo ABM). As tentativas de identificação deste diploma remetem-nos para o facto de não existirem evidências documentais que atestem para a sua existência, apesar da sua obrigatoriedade nos termos do artigo 3.º do Decreto de 2 de outubro de 1862. A sua existência não consta nem nos índices dos boletins oficiais do *Diário de Governo* do ano de 1886, nem no Arquivo Histórico da Assembleia da República, muito menos no «arquivo do arquivo» do ANTT. A primeira menção desta portaria surge com a *Relação* (Arquivo Nacional da Torre do Tombo 1886) elaborada pelo oficial de origem madeirense (e posteriormente diretor do ANTT), Roberto Augusto da Costa Campos, que veio à ilha da Madeira por três vezes para recolher os fundos documentais para o ANTT (precisamente nos anos de 1886, 1887 e 1894). Depois da primeira reivindicação feita pelos autores do *Elucidário Madeirense* (1921), somente em 1934, Machado classificou este diploma como «portaria surda» (1935, 171). Presumindo que este oficial estivesse munido de uma autorização ao mais alto nível para executar a remoção dos arquivos para o ANTT, por que motivo a Portaria de 9 de junho de 1886, do Ministério do Reino, não foi objeto de publicação nos boletins oficiais governamentais? Apesar de o artigo 3.º do Decreto de 2 de outubro de 1862 estabelecer como procedimento de aquisição de arquivos a publicação de «uma portaria expedida ao competente prelado diocesano, pela direcção geral dos negócios eclesiásticos» (Ministério da Justiça 1862, 304), a ausência de elementos factuais que atestem a existência da denominada Portaria do Ministério do Reino de 9 de junho de 1886, não deixou de

ser um factor causal relevante para produzir um determinado efeito, que o teve, ao menos na forma como este diploma tem sido referenciado pelos IAI sucessores, sem indagação por parte das entidades custodiantes acerca do seu teor.

Sobre a segunda questão de investigação, em última análise, é possível verificar que as entidades custodiantes (e mais particularmente o ANTT) não escrutinaram suficientemente sobre o teor da Portaria de 9 de junho de 1886, do Ministério do Reino, sobre o qual não se possui qualquer registo publicado. Apesar de o caso da Madeira não ser único, podendo ser conferido, por exemplo, em relação aos fundos documentais removidos da região de Coimbra para o ANTT ao abrigo da Portaria de 9 de julho de 1863, do Ministério do Reino, a descrição arquivística no metadado FIA poderia sustentar-se na informação, pelo menos do «arquivo do arquivo» do ANTT, de modo a apresentar uma justificação minimamente plausível sobre o processo de aquisição dos fundos madeirenses, aspeto que não se verifica em nenhum dos IAI emanados pelas entidades custodiantes.

## DISCUSSÃO

De acordo com Ribeiro, como se denominam os arquivos que foram «desalojados do seu *habitat* original» (1998, 522), quando este ocorre dentro de um mesmo país? Este artigo partiu do caso português veiculado pelo *Relatório* de Lowry, mais especificamente a reivindicação pela restituição de arquivos removidos do arquipélago da Madeira para o ANTT em finais do século XIX. O principal contributo deste artigo consiste em trazer um fenómeno pouco conhecido na literatura científica, atendendo que o conflito pela custódia de arquivos tem sido entendido como um fenómeno que ocorre predominantemente em contexto internacional.

O enfoque deste artigo consiste em analisar, não do ponto de vista jurídico-normativo, mas no que diz respeito à representação da informação dos conjuntos documentais objeto de disputa em IAI elaborados por cada uma destas entidades custodiantes. É curioso verificar que os casos subnacionais (ainda que outros autores atribuam causas diversas para o mesmo fenómeno) ocorrem com maior incidência em regiões insulares, que, para além da ilha da Madeira (Macedo, Silva y Freitas 2022), podem ser presenciados noutros casos estudados sobre os arquivos removidos das Ilhas Virgens para o Reino da Dinamarca e para os Estados Unidos da América por Bastian (2001), sobre a política extrativista de arquivos por parte do Reino Unido sobre as suas antigas possessões insulares em Rawlings (2015) ou da importância da sua restituição às comunidades originárias não só como forma de reparação de injustiças históricas ocorridas no passado mas também na revitalização de práticas culturais em situação de

vulnerabilidade (Lancefield 1998). Contudo, fenómenos subnacionais também existem em contexto continental e entre diversas jurisdições territoriais dentro de um mesmo país, como é o caso dos Papeles de Salamanca (Cruanyes 2003), em Espanha.

O fenómeno de remoção dos arquivos da ilha da Madeira para o ANTT, que se ampara na legitimação do seu cânone arquivístico nacionalista, decorreu de uma forma, senão tácita, bastante acidentada no passado. O resultado da dispersão de fundos entre entidades custodiantes tornou a hipótese de reunificação de fundos uma tarefa complexa, onde as tecnologias de informação e de reprodução são instrumentalizadas, ao mesmo tempo, como uma solução fungível à restituição física e uma relação de poder sobre a comunidade reivindicante.

Neste sentido, até que ponto as garantias literárias para a descrição arquivística proporcionam um cenário de representação reunificada de fundos dispersos do mesmo produtor entre diversas entidades custodiantes? Não se pretende com isto sugerir a ideia de falência dos modelos de representação da informação em IAI construídos com o suporte das garantias literárias para a descrição arquivística. Foi possível perceber que os IAI, no caso particular de Portugal, se limitaram a representar conjuntos documentais como meras listas de ativos patrimoniais, transmitindo, em certa medida, uma falsa perceção de completude, de integridade e de organicidade ao utilizador, designadamente na estratégia de as entidades custodiantes tendencialmente não representarem as lacunas ou os erros nos IAI. No caso dos arquivos removidos das Ilhas Virgens, Bastian criticou a forma como «each custodian can only provide access to a portion of the records without any clear idea of what else exists» (2001, 111). De igual forma, Ribeiro criticou a forma como fundos documentais beneditinos dispersos entre diversas entidades custodiantes em Portugal eram representados nos IAI, que, em várias situações, «ocasionou uma perda de inteligibilidade do todo e o tratamento técnico a que cada uma das parcelas foi sujeita descurou a visão global» (2006, 311). No presente caso analisado, umas das hipóteses consiste no facto de haver representações idiossincrásicas em IAI (tanto analógico como digital) que tendem a reproduzir *silos* informacionais, em vez de alcançar uma representação descritiva dos fundos documentais técnica e cientificamente reunificada.

Não obstante a reivindicação subjacente à supracitada *Resolução*, existem diversos conjuntos documentais removidos do arquipélago da Madeira que saíram ilegalmente do país, nos termos do Decreto Regional n.º 14/78/M, de 10 de março (Assembleia Regional da Região Autónoma da Madeira 1978). Nesta condição estão, a título de exemplo, os arquivos privados de natureza empresarial, como *Vera Way Marghab Papers* (South Dakota State University 2000) e *Cossart, Gordon and Co.* (London Metropolitan Archives 2011). Até ao momento,

não se dispõe de uma listagem pública dos conjuntos documentais removidos do arquipélago.

Ainda assim, posto que os estudos sobre arquivos deslocados em contexto subnacional se encontram numa condição incipiente, muitos casos permaneceram obscurecidos, sendo que a pesquisa documental e arquivística nos «arquivos dos arquivos» (Henry 2009) pode fazer emergir diversos casos silenciados. Como Lowry refere, «the catalogue is the key» (2017, 8), não só para identificar casos, mas também para fazer emergir através da análise genológica, retórica ou discursiva processos subjacentes à representação de arquivos que se encontram na condição de deslocados.

### CONSIDERAÇÕES FINAIS

Este artigo abordou o caso específico dos pedidos de restituição de arquivos reivindicados pela Região Autónoma da Madeira (RAM) ao Arquivo Nacional da Torre do Tombo (ANTT). O estudo teve como principal objetivo tornar visível a complexidade dos arquivos deslocados em contexto subnacional ou intranacional, utilizando o caso português como ponto de partida. Também explorou abordagens metodológicas alternativas para compreender a causalidade desse fenómeno e identificar as suas características diferenciadoras.

A investigação realizada adotou uma abordagem qualitativa, utilizando análise documental e arquivística. Os dados foram recolhidos a partir de IAI publicados pelo ANTT e pelo Arquivo Regional e Biblioteca Pública da Madeira. A análise concentrou-se na representação da informação dos conjuntos documentais em disputa nos IAI, com ênfase na forma de aquisição.

Os resultados revelaram que a Resolução n.º 3/2017/M, de 12 de janeiro, listou cinco fundos a serem transferidos para a RAM. No entanto, a pesquisa realizada no portal institucional do ANTT identificou onze fundos madeirenses sob a custódia do ANTT. Essa discrepância pode ser atribuída ao facto de que a Resolução pode ter utilizado um IAI anterior a 2000.

O estudo destacou a importância de compreender não apenas as questões jurídicas de posse e propriedade dos arquivos, mas também os critérios de representação de informação adotados pelas instituições custodiantes contestadas. Nesse sentido, a análise dos IAI permitiu identificar as estratégias de representação dos conjuntos documentais em disputa e examinar a forma como foram organizados e descritos.

Em suma, este estudo procurou dar um contributo para uma melhor compreensão da complexidade dos arquivos deslocados em contexto subnacional ou intranacional, utilizando o caso da Região Autónoma da Madeira como exemplo.

Também sublinhou a necessidade de abordagens metodológicas alternativas para estudar este fenómeno e identificar as suas características específicas. Ao tornar visível esta problemática, espera-se encontrar soluções mais adequadas para lidar com os arquivos deslocados e as suas reivindicações de restituição.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Anón. 1939a. A Madeira nos arquivos nacionais: Alfândega do Funchal. Parte 1 de 8. *Arquivo Histórico da Madeira* 6: 129–32.
- . 1939b. A Madeira nos arquivos nacionais: Cabido da Sé Catedral. Parte 2 de 8. *Arquivo Histórico da Madeira* 6: 129–32.
- . 1939c. A Madeira nos arquivos nacionais: Convento da Encarnação. Parte 3 de 8. *Arquivo Histórico da Madeira* 6: 129–32.
- . 1939d. A Madeira nos arquivos nacionais: Convento de Santa Clara. Parte 4 de 8. *Arquivo Histórico da Madeira* 6: 129–32.
- . 1939e. A Madeira nos arquivos nacionais: Livros da Repartição da Fazenda do Funchal. Parte 7. *Arquivo Histórico da Madeira* 6: 160–68.
- Arquivo Nacional da Torre do Tombo. 1886. Relação dos livros pertencentes ao arquivo da Alfândega do Funchal, transferidos para o Arquivo Nacional da Torre do Tombo (L. 266). Lisboa.
- . 1894. Relação de todos os livros encontrados nos archivos dos suprimidos Conventos de Santa Clara e de Nossa Senhora da Encarnação, e de que n'esta data se faz entrega pela repartição de fazenda do districto do Funchal. Funchal.
- . 2002. Provedoria e Junta da Real Fazenda do Funchal. En *Guia Geral dos Fundos da Torre do Tombo, Vol.IV: Administração Periférica. Domínios Casa Real e Casa Anexas*, editado por Maria do Carmo Jasmins Dias Farinha y António Frazão, 318–72. Lisboa: Arquivo Nacional da Torre do Tombo.
- . 2011. Arquivo do Arquivo (PT/TT/AA). DigitArq. 2011. <https://digitarq.arquivos.pt/details?id=4201111>.
- Arquivo Regional da Madeira. 1997. Guia do Arquivo Regional da Madeira. *Arquivo Histórico da Madeira: Boletim do Arquivo Regional da Madeira* 20: 11–204.
- Arquivo Regional e Biblioteca Pública da Madeira. 2017. “Arquivo Regional e Biblioteca Pública da Madeira [portal institucional]”. Página institucional. 2017.
- Assembleia da República. 2010. Legislação Régia. Assembleia da República. 2010. <http://legislacaoregia.parlamento.pt>.
- Assembleia Legislativa da Região Autónoma da Madeira. 2017. Resolução da Assembleia Legislativa da Região Autónoma da Madeira n.º 3/2017/M. En *Diário da República n.º 9/2017, Série I de 2017-01-12*.
- Assembleia Regional da Região Autónoma da Madeira. 1978. Decreto Regional n.º 14/78/M. En *Diário da República n.º 58/1978, Série I de 1978-03-10*.
- Auer, Leopold. 1998. *Disputed archival claims: analysis of an international survey: a RAMP study*. UNESCO. París: Unesco. <https://unesdoc.unesco.org/ark:/48223/pf0000113472>.
- Bastian, Jeannette Allis. 2001. A question of custody: The colonial archives of the United States Virgin Islands. *The American Archivist* 64 (1): 96–114. <https://doi.org/https://doi.org/10.17723/aarc.64.1.h6k872252u2gr377>.

- Conselho Internacional dos Arquivos. 2002. *ISAD(G) : Norma Geral Internacional de Descrição Arquivística : adoptada pelo Comité de Normas de Descrição, Estocolmo, Suécia, 19-22 de Setembro de 1999*. 2.a. Lisboa: Instituto dos Arquivos Nacionais/Torre do Tombo.
- Cruanyes, Josep. 2003. *Els papers de Salamanca : l'espoliació del patrimoni documental de Catalunya (1938-1939)*. Barcelona: Edicions 62.
- Direção-Geral do Livro dos Arquivos e das Bibliotecas. 2017. Rede Portuguesa de Arquivos: Portal Português de Arquivos. 2017.  
<https://portal.arquivos.pt>.
- Direção-Geral dos Arquivos. 2011. *Orientações para a Descrição Arquivística*. 3.a versão. Lisboa: Direção-Geral dos Arquivos.  
<https://act.fct.pt/wp-content/uploads/2014/05/ODA-3a-versão.pdf>.
- Governo Regional da Madeira. 1998. NESOS: Base de Dados de História das Ilhas Atlânticas. Centro de Estudos de História do Atlântico. 1998.  
<http://nesos.madeira-edu.pt/>.
- . 2021. Arquivo Distrital do Funchal (PT/ABM/ADFUN). Archeevo. 2021.  
<https://arquivo-abm.madeira.gov.pt/details?id=1387943>.
- Henry, Christy Morgan. 2009. Toward the Archives of Archives: The New Archival History, Accountability and the Documentation of Archival Appraisal. University of Manitoba/University of Winnipeg.  
<http://hdl.handle.net/1993/21532>.
- Imprensa Nacional-Casa da Moeda. 2016. DRE Diário da República Eletrónico. Imprensa Nacional-Casa da Moeda. 2016.  
<https://dre.pt>.
- Instituto dos Arquivos Nacionais/Torre do Tombo. 2002. *Ordens monástico-conventuais: inventário: Ordem de São Bento, Ordem do Carmo, Ordem dos Carmelitas Descalços, Ordem dos Frades Menores, Ordem da Conceição de Maria*. Editado por José Mattoso y Maria do Carmo Jasmins Dias Farinha. Lisboa: Instituto dos Arquivos Nacionais/Torre do Tombo.
- International Council on Archives. 2020. Disputed Archival Claims: An International Survey (2018/2019): Report to the International Council on Archives' Expert Group on Shared Archival Heritage. Paris.  
[https://www.ica.org/sites/default/files/disputed\\_archival\\_claims\\_an\\_international\\_survey\\_2018\\_2019.pdf](https://www.ica.org/sites/default/files/disputed_archival_claims_an_international_survey_2018_2019.pdf).
- Kecskeméti, Charles. 1977. Archival claims; preliminary study on the principles and criteria to be applied in negotiations. PGI-77/WS/1. París.  
[https://unesdoc.unesco.org/ark:/48223/pf0000029879\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000029879_fre).
- Lancefield, Robert C. 1998. Musical 'Traces' Retractable Paths: The Repatriation of Recorded Sound. *Journal of Folklore Research* 35(1): 47–68.  
<https://doi.org/10.2307/3814785>.
- London Metropolitan Archives. 2011. Cossart Gordon Company. *AIM25 Archives in London and the M25 area*.  
[https://aim25.com/cgi-bin/vcdf/detail?coll\\_id=16850&inst\\_id=118&nv1=-browse&nv2=sub&fbclid=IwAR1a2PahjtNfgbbb\\_BhG6-WrmlACmtlWSxteUIX-5HM2w\\_ECqi3oqtm\\_J7Tl](https://aim25.com/cgi-bin/vcdf/detail?coll_id=16850&inst_id=118&nv1=-browse&nv2=sub&fbclid=IwAR1a2PahjtNfgbbb_BhG6-WrmlACmtlWSxteUIX-5HM2w_ECqi3oqtm_J7Tl).
- Lowry, James. 2017. Introduction: *displaced archives*. En *Displaced archives*, editado por James Lowry, 1–10. London ; New York: Routledge.
- Macedo, L. S. Ascensão de. 2019. Arquivos deslocados: mapeamento de literatura. *Brazilian Journal of Information Science* 13 (4): 5–34.  
<https://doi.org/10.36311/1981-1640.2019.v13n4.02.p5>.

- Macedo, L S Ascensão de, Carlos Guardado da Silva, y Maria Cristina Vieira de Freitas. 2022. Below the Nation State: Power Asymmetry and Jurisdictional Boundaries around the Archives of Madeira Archipelago. En *Contested Archival Heritage*, editado por James Lowry, 137–70. London: Routledge.
- Machado, João Franco. 1935. Alguns Documentos do Mosteiro de Santa Clara do Funchal. *Arquivo Histórico da Madeira* 4: 171–72.
- MacNeil, Heather. 2012. What finding aids do: archival description as rhetorical genre in traditional and web-based environments. *Archival Science* 12(4): 485–500. <https://doi.org/10.1007/s10502-012-9175-4>.
- Ministério da Justiça. 1862. Decreto de 2 de outubro de 1862. En *Collecção Official da Legislação Portuguesa: anno de 1862*, 303–6. Lisboa: Imprensa Nacional.
- Ministério do Reino. 1887. Decreto de 29 de dezembro de 1887. En *Collecção de Legislação Portuguesa: anno de 1887*, 688–93. Lisboa: Imprensa Nacional.
- Ministério dos Negócios da Fazenda. 1862. Lei de 4 de abril de 186. En *Collecção official da Legislação Portuguesa: anno de 1861*, 155–57. Lisboa: Imprensa Nacional.
- Ministério e Secretaria de Estado dos Negócios da Fazenda. 1871. Decreto e Instruções de 31 de maio de 1862. En *Collecção official da legislação Portuguesa: anno de 1870*, 33–34. Lisboa: Imprensa Nacional.
- Rawlings, Gregory. 2015. Lost Files, Forgotten Papers and Colonial Disclosures: The ‘Migrated Archives’ and the Pacific, 1963–2013’. *The Journal of Pacific History* 50(2): 189–212. <https://doi.org/10.1080/00223344.2015.1048585>.
- Ribeiro, Fernanda. 1998. O acesso à informação nos arquivos. Universidade do Porto. <https://repositorio-aberto.up.pt/handle/10216/7058>.
- . 2006. Os Instrumentos de acesso à informação das instituições monásticas beneditinas: uma abordagem crítica. En *Estudos em homenagem a José Amadeu Coelho Dias*, v. 1, 307–20. Porto: Faculdade de Letras da Universidade do Porto.
- . 2013. O Uso da classificação nos arquivos como instrumento de organização, representação e recuperação da informação. En *I Congresso ISKO Espanha e Portugal Porto, 7 a 9 de novembro de 2013*, 528–39. <https://repositorio-aberto.up.pt/handle/10216/69659>.
- Silva, Carlos Guardado da. 2021. Investigação documental. En *Manual de investigação qualitativa: Conceção, análise e aplicações*, editado por Sónia Gonçalves, Joaquim P Gonçalves, y Célio Gonçalo Marques, 103–23. Lisboa: Pactor.
- Silva, Fernando Augusto da, y Carlos Azevedo de Menezes. 1921. *Elucidario madeirense*. Funchal: Tipografia Esperança.
- South Dakota State University. 2000. Vera Way Marghab Papers. South Dakota State University. 2000. <https://www.sdstate.edu/sdsu-archives-and-special-collections/vera-way-marghab-papers>.
- Winn, Samantha R. 2015. Ethics of Access in Displaced Archives. *Provenance, Journal of the Society of Georgia Archivists* 33 (1): 6–13. <https://digitalcommons.kennesaw.edu/provenance/vol33/iss1/5>.

*Para citar este texto:*

Macedo, L.S. Ascensão de 2023. “Arquivos contestados em contexto intranacional Em torno do caso da Região Autónoma da Madeira (Portugal)”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (96): 201-218.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2023.96.58791>