





# ***SCIRE***

Representación y organización del conocimiento



# SCIRE

Representación y organización del conocimiento

Vol. 25, n.º 1, enero-junio 2019

ISSN 1135-3716

Scire:

knowledge representation and organization

Vol. 25, n. 1, January-June 2019

Ibersid:

Red de Investigación  
en Sistemas de Información  
y Documentación

© 2019 Francisco Javier García Marco

Reservados todos los derechos. Prohibida la reproducción por cualquier medio de la totalidad o parte de la presente edición con fines comerciales sin permiso escrito de los titulares del copyright. Queda, sin embargo, autorizada expresamente la reproducción de los resúmenes y palabras clave de los artículos en inglés y español. También se permite la reproducción de sus textos con objetivos exclusivamente docentes.

Los artículos publicados están sometidos a evaluación científica y editorial en beneficio de la comunidad científica y los autores, pero dicha evaluación se realiza dentro de los límites del desarrollo actual de la disciplina, las limitaciones humanas de editores y revisores, y con una perspectiva ideológica, teórica y metodológica abierta, por lo que la dirección no siempre comparte las opiniones, métodos y procedimientos reflejados en la revista. Los autores se responsabilizan personalmente del cumplimiento de los códigos éticos y la legislación vigente, así como de la obtención de los permisos de copia de materiales ajenos.

Se permite expresamente el autoarchivo de la versión electrónica por parte de los autores en sus repositorios personales e institucionales. Los autores pueden reproducir sus textos y figuras sin permiso expreso del editor indicando la fuente de publicación, y conservan los derechos de su original con la obligación de citar la presente edición.

ISSN: 1135-3716 = Scire (Zaragoza)  
Depósito legal: Z. 1.790 — 1995

Edita: Ibersid® con la colaboración de Prensas de la Universidad de Zaragoza

Foto de cubierta: Iuri Rocio Franco Rizzi

Imprime:  
Servicio de Publicaciones. Universidad de Zaragoza.  
Edificio de Ciencias Geológicas, C/ Pedro Cerbuna, 12.  
50009 Zaragoza, España. Tel.: 976 761 330. Fax: 976 761 063.

---

# Scire

representación y organización  
del conocimiento

---

## Alcance y objetivos

---

*Scire: representación y Organización del Conocimiento* es una publicación semestral de carácter interdisciplinar sobre la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

## Difusión

---

*Scire* tiene difusión internacional. Agradecemos la inclusión en los siguientes servicios de referencia: Scopus, ESCI, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC y Catálogo Latindex.

## Agradecimientos

---

*Scire* publica en este volumen ponencias y comunicaciones arbitradas presentadas en la edición de 2018 del congreso Ibersid, que contó con una ayuda del Departamento de Investigación, Innovación y Universidad del Gobierno de Aragón (BBC 11\_2018).

## Redacción, distribución y canje

---

Revista Scire  
Departamento de Ciencias de la Documentación e Historia de la Ciencia  
Facultad de Filosofía y Letras  
Universidad de Zaragoza  
C/ Pedro Cerbuna 12,  
E-50.009 Zaragoza (Spain)

Tfno: int + 34 976 762239. Fax: 34 976761506.  
E-mail: <mailto:scire@unizar.es>

## Suscripciones y números sueltos

---

Suscripción anual / Annual subscription: 30 €.  
Número suelto: 20 €.  
(IVA incluido)

---

# Scire

knowledge organization  
and representation

---

## Scope and aims

---

*Scire: Representación y Organización del Conocimiento* is an interdisciplinary journal published twice a year that is devoted to the representation, standardization, treatment, retrieval and communication of information and knowledge.

## Dissemination

---

*Scire* has international distribution. We acknowledge its inclusion in the following reference services: Scopus, ESCI, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC and Catálogo Latindex.

## Acknowledgments

---

In this volume, *Scire* publishes referred invited and contributed papers to the congress Ibersid 2018, which received a grant from the Department of Research, Innovation and Universities of the Government of Aragón (BBC 11\_2018).

## Contact address

---

Revista Scire  
Departamento de Ciencias de la Documentación e Historia de la Ciencia  
Facultad de Filosofía y Letras  
Universidad de Zaragoza  
C/ Pedro Cerbuna 12,  
E-50.009 Zaragoza (Spain)

Tel.: int + 34 976 762239. Fax: 34 976761506.  
E-mail: [scire@unizar.es](mailto:scire@unizar.es)

## Subscriptions

---

Annual subscription: 30 €.  
Issue: 20 €.  
(VAT included)

## Editor

Francisco Javier García Marco, Univ. de Zaragoza. E-mail: jgarcia@unizar.es

### Consejo de redacción / Editorial council

Mario Guido Barité Roqueta, Universidad de Montevideo, Uruguay	João Batista Ernesto Moraes, Universidade Estadual Paulista, Brasil	Francisco Javier Martínez Mendez, Universidad de Murcia, España
José Augusto Chaves Guimarães, Universidade Estadual Paulista, Brasil	Francisco Javier García Marco, Universidad de Zaragoza, España	Álvaro Quijano Solís, Colegio de México, México

### Consejo científico / Scientific council

Tomás Baiget, EPI S. A., España	José María Izquierdo Arroyo, Universidad de Murcia, España	Manuel José Pedraza Gracia, Universidad de Zaragoza, España
José Luis Bonal Zazo, Univ. de Extremadura, España	María Pilar Lasala Calleja, Universidad de Zaragoza, España	María Pinto Molina, Universidad de Granada, España
Mercedes Caridad Sebastián, Universidad Carlos III de Madrid, España	Alfonso López Yepes, Universidad Complutense de Madrid, España	Gloria Ponjuán Dante, Universidad de La Habana, Cuba
Alberto Carreras Gargallo, Universidad de Zaragoza, España	José López Yepes, Universidad Complutense de Madrid, España	José Vicente Rodríguez Muñoz, Universidad de Murcia, España
Emilia Currás Puente, Universidad Autónoma de Madrid, España	Pedro Marijuán Fernández, Universidad de Zaragoza, España	Adelaida Román Román, CINDOC (Madrid), España
Constança Espelt Busquets, Universidad de Barcelona, España	Bonifacio Martín del Brío, Universidad de Zaragoza, España	Juan Ros García, Universidad de Murcia, España
María Eulalia Fuentes Pujol, Universidad Autónoma de Barcelona, España	José Antonio Moreira González, Universidad Carlos III de Madrid, España	Francisco José Ruiz de Mendoza Ibáñez, Universidad de La Rioja, España
Fernando Galindo Ayuda, Universidad de Zaragoza, España	Purificación Moscoso Castro, Universidad de Alcalá, España	Félix Sagredo Fernández, Universidad Complutense de Madrid, España
Blanca Gil Urdicián, Universidad Complutense de Madrid, España	Félix Moya Anegón, Universidad de Granada, España	Elías Sanz Casado, Universidad Carlos III de Madrid, España
Alan Gilchrist, Cura Consortium, Reino Unido	María del Carmen Negrete Gutiérrez, Universidad Autónoma de México	Carlos Serrano Cinca, Universidad de Zaragoza, España
Vicente Pablo Guerrero Bote, Univ. de Extremadura, España	José Luis Otañal, Universidad Jaime I de Castellón, España	Emilio Setién Quesada, Asociación Cubana de Bibliotecarios, Cuba

### Revisores externos del número / External reviewers in this issue

Agradecemos la colaboración altruista y desinteresada de los revisores externos del presente número: Ernest Abadal Falgueras, María Luisa Alvíte Díez, Antonio García Jiménez, Javier Guallar, Luis Miguel Moreno Fernández, Javier Nogueras Iso, Blanca Rodríguez Bravo, Maria Graça de Melo Simões y Roberta Dal' Evedove Tartarotti.

#### Candidaturas al consejo científico

Se aceptan candidaturas al consejo científico de especialistas del área de Biblioteconomía y Documentación y de otras disciplinas relacionadas (Informática, Ciencias Sociales, Lingüística, Filosofía, Psicología, etc.) con experiencia profesional e investigadora demostrada. En el sistema público de investigación español, suele ser equivalente al doctorado y dos sexenios de investigación o méritos equivalentes.

#### Scientific council membership policy

Candidatures of researchers from LIS and other related disciplines (Computer Science, Social Sciences, Linguistics, Philosophy, Psychology, etc.) with demonstrated professional and research experience are welcomed. In the Spanish public research system, for example, this usually means having a doctorate and two scientific productivity sexennia or equivalent outputs.

---

# Tabla de contenidos en español

*Table of contents in Spanish*

---

*Tabla de contenidos en español* ..... 9

*Tabla de contenidos en inglés* ..... 11

## **Artículos**

*Correspondencias entre la norma aragonesa para la descripción de autoridades de archivos (ARANOR) y MARC21: hacia la web semántica*

Miguel Ángel del PRADO MARTÍNEZ..... 13

*La dimensión cultural de la organización del conocimiento: un análisis en el contexto internacional de la ciencia de la información*

José Augusto Chaves GUIMARÃES

Isadora Victorino EVANGELISTA

Gabriele de Araújo Medeiros LUZ

Henrique Fiamengue OSAWA ..... 25

*Presente y futuro de las revistas científicas*

Nicoleta-Roxana DINU

Tomàs BAIGET ..... 37

*Sistemas y procesos de organización del conocimiento: un análisis conceptual de la literatura brasileña de ciencia de la información*

Walter MOREIRA

Fernanda Carolina PEGORARO NOVAES

Isabela SANTANA DE MORAES ..... 47

*Estudio sobre la indización/etiquetado y los lenguajes documentales*

*en cinco diarios españoles* ..... 55

Antonio GARCÍA-JIMÉNEZ

David RODRÍGUEZ-MATEOS

Beatriz CATALINA-GARCÍA ..... 55

*Evaluación de la coincidencia de los descriptores de las revistas españolas de PubMed*

Miguel Ángel PELLÉS GARCÍA

José Antonio SALVADOR OLIVÁN..... 65

*Procedimiento de evaluación*..... 73

*Instrucciones para la presentación de trabajos*..... 75

*Índice de autores* ..... 78

*Índice de materias en español*..... 78

*Índice de materias en inglés*..... 79



---

# Tabla de contenidos en inglés

*Table of contents in English*

---

*Table of contents in Spanish* ..... 9  
*Table of contents in English* ..... 11

## Articles

*Correspondences between the Aragonese  
Standard Archival Authority Records  
(DARANOR) and MARC21:  
towards the semantic web*  
Miguel Ángel del PRADO MARTÍNEZ..... 13

*The cultural dimension of Knowledge  
Organization: an analysis in the Information  
Science international context*  
José Augusto Chaves GUIMARÃES  
Isadora Victorino EVANGELISTA  
Gabriele de Araújo Medeiros LUZ  
Henrique Fiamengue OSAWA ..... 25

*Present and future of scientific journals*  
Nicoleta-Roxana DINU  
Tomàs BAIGET ..... 37

*Systems and processes of knowledge  
organization: a conceptual analysis  
of the Brazilian Information Science literature*  
Walter MOREIRA, Fernanda Carolina PEGORARO  
NOVAES, Isabela SANTANA DE MORAES..... 47

*Study on indexing/tagging and controlled  
vocabularies in five Spanish newspapers*  
Antonio GARCÍA-JIMÉNEZ  
David RODRÍGUEZ-MATEOS  
Beatriz CATALINA-GARCÍA ..... 55

*An assessment of descriptor coincidence  
in the Spanish journals of Pub-Med*  
Miguel Ángel PELLÉS GARCÍA  
José Antonio SALVADOR OLIVÁN..... 65

*Evaluation process*..... 73  
*Instructions for authors*..... 75  
*Author index* ..... 78  
*Subject index in Spanish*..... 78  
*Subject index in English* ..... 79



---

# Correspondencias entre la norma aragonesa para la descripción de autoridades de archivos (ARANOR) y MARC21: hacia la web semántica

*Correspondences between the Aragonese Standard Archival Authority Records (DARANOR) and MARC21: towards the semantic web*

---

**Miguel Ángel del PRADO MARTÍNEZ**

Archivo Histórico Provincial de Zaragoza, Zaragoza, España, [madelprado@aragon.es](mailto:madelprado@aragon.es)

## Resumen

Se presentan las correspondencias establecidas entre la Norma Aragonesa para la Descripción de Autoridades de Archivos (ARANOR) y MARC21 para datos de autoridad, analizando las repercusiones de su aplicación para el sistema de información de archivos DARA (Documentos y Archivos de Aragón).

**Palabras clave:** Descripción archivística. Registros de autoridad. Control de autoridades. Normas de descripción archivística. Sistemas de información. ARANOR. MARC21. DARA. Aragón. Web semántica.

## 1. Introducción

Los archiveros aragoneses han demostrado en las dos últimas décadas su interés por el control y normalización de las autoridades archivísticas. Buen ejemplo de ello ha sido la elaboración por el Grupo de Trabajo de Autoridades de Aragón de unas pioneras y modélicas normas para la descripción de autoridades de archivos (ARANOR) y para la creación de descriptores geográficos (GEODARA), sin olvidar el impulso para la creación de un fichero de autoridades de instituciones, personas y familias de los archivos aragoneses. Sin embargo todavía no se había abordado la codificación informática de las descripciones completas de las autoridades archivísticas y puntos de acceso (personas, familias, instituciones, lugares y temas), una tarea imprescindible para posibilitar su redacción, comunicación e intercambio a través de los sistemas informáticos y muy especialmente a través de la web. Con este objetivo, y con la vista puesta en la redacción de una nueva norma para la estructuración de datos para registros de autoridad en los archivos de Aragón, que recibirá el nombre de DARANOR, se está actualmente trabajando. El núcleo de este estándar lo constituyen las equivalencias establecidas entre la norma de contenido que rige las descripciones de autoridades en Aragón (ARANOR) y MARC21 para datos de autoridad, el formato informático utilizado por el sistema de información de los archivos de Aragón (DARA) para

## Abstract

A proposal of correspondences between the Aragonese Standard Archival Authority Records (ARANOR) and the MARC21 format for authority data is presented, analyzing their impact on DARA (Documents and Archives of Aragón).

**Keywords:** Archival description. Authority records. Authority control. Standards for archival description. Information systems. ARANOR. MARC21. DARA. Aragón. Semantic web.

la representación y codificación de dichas descripciones.

Con este artículo se pretende presentar las correspondencias establecidas entre ARANOR y MARC21, deteniéndose en el análisis de las ventajas que su aplicación práctica suponen para la mejora tanto del funcionamiento interno de DARA como de sus relaciones con otros sistemas de información externos. Así mismo se quiere presentar el contexto en el que se desarrolla este trabajo, marcado por el acercamiento entre las instituciones de diferentes tipologías responsables de la custodia y difusión del patrimonio cultural (bibliotecas, archivos, museos...), y condicionado por el desarrollo tecnológico y muy especialmente por la web semántica y los nuevos modelos de datos abiertos y enlazados.

## 2. La descripción normalizada de autoridades

### 2.1. La descripción de las autoridades bibliográficas

El concepto de autoridad surge en el ámbito bibliotecario entendido como un término normalizado que permite acceder de forma única a los registros bibliográficos que lo contienen. Los nombres de personas, familias y entidades corporativas, los títulos uniformes, las materias y lugares geográficos o los títulos de las series, una

vez normalizados y establecidos como puntos de acceso autorizados, constituyen los diferentes tipos de autoridades. El proceso por el que se establecen estas formas únicas autorizadas y se señalan las relaciones entre ellas y todas sus formas variantes se conoce como control de autoridades. Este proceso tiene por finalidad facilitar la identificación y recuperación de la información bibliográfica, eliminando la ambigüedad en las denominaciones de personas, entidades o temas que sean sinónimos, homónimos o presenten variedad de nombres.

El resultado del control de autoridades queda plasmado en los registros de autoridad que recogen la información sobre los puntos de acceso normalizados, incluyendo la forma autorizada, las formas no autorizadas, las fuentes consultadas que justifican cada variante seleccionada y las posibles relaciones con otros puntos de acceso. Además, pueden aparecer notas explicativas para el usuario, instrucciones para el catalogador, y datos de tipo interno como el número de registro, códigos de validación, fecha de creación y modificación, etc. Complemento necesario de los registros de autoridad son los registros de referencias que contiene un punto de acceso no autorizado y que dirigen al usuario desde una variante no admitida a la forma autorizada. Un tipo especial de registros de referencias son los registros explicativos generales que desde una forma no autorizada sugieren la consulta de una serie de registros de autoridad. El conjunto estructurado y ordenado de registros de autoridad forman el fichero o catálogo de autoridades, un producto derivado y auxiliar del catálogo bibliográfico.

De acuerdo con la *Declaración Profesional de la IFLA sobre el Control Bibliográfico Universal* (IFLA, 2012) las agencias bibliográficas nacionales son las responsables de la elaboración de los registros de autoridad de las obras relacionadas con su país, debiendo permitir el acceso a sus ficheros y compartir sus autoridades con otras agencias, bibliotecas y otros colectivos como archivos y museos. Las limitaciones de las agencias nacionales se ven compensadas por el trabajo normalizador realizado por diferentes instituciones bibliotecarias que crean y dan acceso abierto a sus propios ficheros de autoridades. Así mismo se están desarrollando proyectos cooperativos como el Fichero de Autoridades Virtual Internacional (VIAF) que tiene por objetivo crear una base de datos disponible a través de Internet que combine de manera virtual los registros de autoridad de las principales instituciones bibliotecarias internacionales.

Las bibliotecas se han dotado de un conjunto de principios y normas que posibilitan la elaboración homogénea de los registros de autoridades. Los

criterios a seguir en el trabajo y a la hora de toma de decisiones están contenidos en la *Declaración de principios internacionales de catalogación* (IFLA, 2017). El modelo conceptual de descripción vigente se ha definido en *Requisitos Funcionales de los Datos de Autoridad* (FRAD) (IFLA, 2009), donde se especifican las entidades que se utilizan como base para la construcción de puntos de acceso controlados (persona, familia, entidad corporativa, obra, expresión, manifestación, ejemplar, concepto, objeto, acontecimiento y lugar), los atributos de cada una de las entidades y los tipos de relaciones que se establecen entre los diversos tipos de entidades. El objetivo de este modelo es compartir datos con sectores como archivos, museos y otras organizaciones que gestionan derechos de autor, trascendiendo por lo tanto el ámbito bibliotecario. Como norma de estructura está la *Directrices para Registros de Autoridad y Referencia* (GARR) (IFLA, 2004) en la que se definen los elementos que han de incluirse en el asiento o registro de autoridad bibliográfica, asignan un orden a los elementos, y especifican un sistema de puntuación.

A falta de un código internacional, las reglas de catalogación vigentes en la zona donde se realiza la catalogación y diferentes listas de autoridades se utilizan a modo de normas de contenido para definir las formas definitivas de las autoridades.

Para la representación y codificación informática, la mayoría de los sistemas bibliotecarios automatizados utilizan el formato *MARC21 para datos de autoridad* (<http://www.loc.gov/marc/authority/>). Este formato diseñado para servir como portador de la información relativa a las formas autorizadas usadas como puntos de acceso en los registros MARC de datos bibliográficos, soporta la descripción de personas, instituciones, reuniones y materias cronológicas, temáticas o geográficas. La estructura de los registros, de acuerdo con la norma ISO 2709, está constituida por una colección de campos con sus correspondientes etiquetas identificativas, una etiqueta de registro, un directorio y los datos. Gracias al esquema MARCXML, los registros MARC pueden representarse en XML, el estándar para el acceso e intercambio de información en la web. De este modo los registros de autoridad se pueden transformar en recursos digitales compartidos y reutilizables a través de la web por diferentes sistemas no exclusivamente bibliotecarios. Un conjunto de programas informáticos y de hojas de estilo desarrollados por la Library of Congress' Network Development and MARC Standards Office permiten la conversión sin pérdida de datos desde MARC21 a MARCXML, y desde éste a otros sistemas de metadatos. Sirviéndose del esquema MADS, MARC21 para datos de autoridad

también puede adaptarse al modelo de la web semántica. MADS es el instrumento terminológico específico con el que representar un vocabulario de autoridad bibliográfica en su contextualización semántica. Desarrolla un esquema de codificación de datos en formato MARC21 a *Resource Description Framework* (RDF) para su uso en aplicaciones de la web semántica (López, 2012; Garzón y Peset, 2018b).

## 2.2. La descripción de autoridades archivísticas

Tanto el marco conceptual como el modelo regulador de las autoridades bibliotecarias han sido replicados por la comunidad archivística, adaptándolos a los objetivos y a la naturaleza de sus documentos e informaciones. Tampoco ha sido extraño el uso de los mismos formatos de codificación por ambas comunidades.

Así el Consejo Internacional de Archivos (ICA) (1992, 2004) ha definido los datos de autoridad como formas normalizadas de términos que incluyen nombres (personales, corporativos o geográficos) utilizados como puntos de acceso, junto con otra información relevante asociada con los términos, como la fuente de un nombre, la relación con otros términos, etc. Por control de autoridades el ICA entiende al control de las formas normalizadas de términos utilizados como puntos de acceso. Se trata de un concepto relativamente reciente, adoptado plenamente por la comunidad archivística internacional con la publicación de las ISAAR-CPF en 1996, aunque ha costado su plena aceptación (Lafuente, 2018). Los registros de autoridad archivísticos se conciben como la forma autorizada del nombre, que combinada con otros elementos informativos permite identificar y describir la entidad mencionada y también remite a otros registros de autoridad relacionados. Estos registros se pueden utilizar para describir una institución, persona o familia como unidades dentro de un sistema de descripción archivístico; para controlar la creación y utilización de los puntos de acceso en las descripciones archivísticas; y para documentar las relaciones entre diferentes productores de documentos, los documentos creados por ellas, y otras entidades o recursos que les conciernen. Por lo tanto, aun reconociéndose de manera explícita su similitud con los registros de bibliotecas, los registros de autoridad de archivos deben responder a un número mayor de exigencias por lo que contienen mucha más información.

Los distintos sistemas archivísticos se responsabilizan de la creación y mantenimiento de sus propios ficheros de autoridades. En España, en el Sistema Español de Archivos, el Centro de Información y Documentación de Archivos (CIDA)

funciona como agencia normalizadora responsable del mantenimiento de las autoridades archivísticas (Lafuente y Muriel, 2014; Lafuente, 2018). En Aragón, el Grupo de Trabajo de Autoridades de Aragón (GTAA) asumió como objetivo el desarrollo de los instrumentos necesarios para construir el catálogo de autoridades de instituciones, personas y familias de los archivos aragoneses (GTAA, 2008). Los ficheros de autoridades en los archivos deben asegurar el control de los puntos de acceso normalizado; actuar como sistemas de referencia e información con la finalidad de proporcionar a los archiveros y a los usuarios de los archivos, la información básica de carácter histórico o descriptivo sobre la persona, entidad, familia o lugar tratado; representar las relaciones entre entidades; y facilitar el intercambio de información archivística, bibliográfica y documental (Bonal, Generele y Travesí, 2006).

El entramado normativo para la elaboración de los registros de autoridades archivísticas se ciementa en la *Declaración de Principios sobre la Descripción de Archivos* del (ICA, 1992) donde se estipula que el propósito de la descripción archivística es "identificar y explicar el contexto y el contenido del material archivístico para promover su accesibilidad".

También se dispone de modelos conceptuales propios que desde una aproximación multientidad establecen los tipos de entidades que deben ser objeto de la descripción archivística, sus relaciones y atributos. En el ámbito nacional el modelo conceptual de la Comisión de Normas Españolas de Descripción Archivística (2017) identifica seis tipos de entidades principales: documentos; agentes (instituciones, familias y personas); funciones; normas; conceptos, objetos o acontecimientos; y lugares. También especifica el conjunto de relaciones o vinculaciones que se establecen entre las entidades, así como los atributos o propiedades de las entidades y de las relaciones relevantes para un sistema de información descriptiva. Con una perspectiva internacional, el ICA (2016) ha elaborado *Records in Context Conceptual Model (RiC-CM)* un modelo conceptual que integra e interrelaciona las normas de descripción sobre documentos, autoridades, funciones e instituciones en sistemas de información archivística completos.

Un conjunto de estándares internacionales, nacionales y regionales permiten la creación de registros de autoridades homogéneos y facilitan el intercambio de información archivística. La *Norma Internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias: ISAAR-CPF* (ICA, 2004) sienta las reglas generales para la normalización de las descripciones de los productores de documentos

de archivo y de su contexto de producción. Es un estándar de estructura en el que se identifican los elementos de información necesarios para la descripción de las autoridades archivísticas, agrupados en cuatro áreas: identificación, descripción, relaciones y control. Además proporciona orientaciones para vincular los registros de autoridad con las descripciones de documentos y con otros recursos y sistemas descriptivos más allá del ámbito archivístico.

El contenido y la formalización de los diferentes registros de autoridad están regulados por normas nacionales y regionales que responden a las características propias de sus contextos archivísticos. La *Norma Aragonesa para la Descripción de Autoridades de Archivos (ARANOR)* es el documento de referencia para la elaboración de registros de autoridad en los archivos aragoneses (GTAA, 2014). ARANOR proporciona las directrices para normalizar el contenido de cada uno de los elementos de ISAAR-CPF, establece las subdivisiones necesarias en cada elemento, con vistas a un etiquetado e intercambio de la información; establece la formalización de cada elemento: puntuación, orden de los elementos, lengua, etc.; normaliza la forma de integrar los registros en un sistema descriptivo controlado (elección de la forma autorizada, una sola entrada para cada autoridad, etc.). Además indica las fuentes adecuadas para cada elemento así como las normas especiales para algunas autoridades. Aragón también cuenta con una norma de contenido para el tipo de entidad “lugar”, se trata de *Geodara: Norma aragonesa para la creación de descriptores geográficos* (Grupo de Trabajo de Descriptores Geográficos de Aragón, 2015a). Su objeto es regular la normalización de los términos que designan lugares, con el fin de crear descriptores geográficos autorizados para los archivos y sistemas de información aragoneses. Además se dispone de una lista de valores normalizados para la indización de lugares: la *Lista de descriptores geográficos de Aragón: localidades, municipios, comarcas y provincias actuales y desaparecidos* (Grupo de Trabajo de Descriptores Geográficos de Aragón, 2015b). Este Grupo de Trabajo ha generado una *Lista de descriptores geográficos de Aragón. Construcciones* (v. 2, noviembre 2017) como aplicación práctica de la Norma Geodara a un caso concreto, los elementos del patrimonio cultural aragonés, que son referencias comunes a las bases de datos de información sobre el patrimonio.

Para posibilitar la comunicación e intercambio de las descripciones archivísticas en entornos electrónicos, es necesario codificar esta información de modo que pueda ser procesada informáticamente. El éxito del intercambio automatizado de

la información de autoridades a través de redes informáticas, depende de la adopción de un formato de comunicación adecuado para todos los archivos que participen en dicho intercambio, tal y como se recoge en la ISAAR-CPF (ICA, 2004, p. 15). Esta norma señala al *Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF)* como uno de los formatos de comunicación que permite el intercambio de datos de autoridad de archivos a través de Internet. Pero no es el único formato posible, como lo demuestra la utilización en algunos sistemas archivísticos del formato bibliotecario MARC21 para datos de autoridad.

EAC-CPF es un estándar de comunicación para el intercambio de descripciones de agentes (instituciones, familias y personas) conformes con la segunda edición de la ISAAR-CPF, que regula de manera detallada la estructura y codificación XML de estas representaciones. El modelo está expresado como un lenguaje XML para favorecer la independencia de la plataforma tecnológica que lo use y la portabilidad de la información. EAC-CPF define los elementos utilizados para describir los agentes y la estructura de interrelaciones entre dichos elementos. Estos elementos y estructura permiten el descubrimiento, navegación y presentación de la información contextual sobre los agentes y la vinculación de esta información con descripciones de documentos de archivo o con otras entidades contextuales, especialmente aquellas codificadas con EAD —formato utilizado para crear, codificar e intercambiar descripciones de documentos de archivo—, MARC y otros estándares similares. Además permite la vinculación de descripciones de entidades contextuales con representaciones digitales u otros sustitutos de dichas entidades (*Encoded Archival Context Working Group*, 2012). Se han realizado proyectos para transformar el formato EAC-CPF en una ontología para convertir los registros de autoridad contenidos en este formato en datos abiertos enlazados (Puente, 2018).

### 2.3. La descripción de autoridades en la comunidad GLAM

Cada vez se hace más patente la importancia de la colaboración entre archivos y bibliotecas para mejorar el acceso y difusión de sus colecciones digitales. En este sentido se ha generalizado el uso del acrónimo *GLAM* referido a las instituciones que recopilan materiales del patrimonio cultural (galerías de arte, bibliotecas, archivos y museos) englobando a las actividades que llevan a cabo para la difusión en el mundo digital de sus acervos culturales e históricos. Y es que tal y como señala la *Coalition for Networked Informa-*

tion (2017) la sinergia entre las diferentes colecciones GLAM tiene un valor añadido que nunca se podría conseguir utilizando estas colecciones por separado.

Sin embargo, según el *Informe Final del Grupo Incubador de Datos Vinculados de Bibliotecas* (2012), los actuales datos de bibliotecas —y añadimos, de archivos— residen mayoritariamente en bases de datos que, aunque disponen de interfaces de búsqueda web, no están integrados entre ellos ni con otras fuentes de datos de la web. Hay una considerable cantidad de datos bibliográficos, archivísticos y de otro tipo que tienen puntos en común con otros recursos de la web como las fechas, la información geográfica, las personas, las organizaciones, los acontecimientos... y que sin embargo están escasa o nula-mente relacionados. Es por ello que se hace necesario superar las formas en que las bibliotecas, archivos y museos difunden sus contenidos, desarrollando nuevos sistemas que permitan tanto el acceso conjunto a sus diferentes colecciones, como la vinculación con otros datos de la web. La tecnología de la web semántica y la creación de modelos o estándares para metadatos y datos enlazados se postulan como la solución para lograr esta necesaria confluencia. Pero de manera previa es necesario lograr la aproximación entre las distintas herramientas desarrolladas por las instituciones GLAM con la tecnología de la web semántica y los datos abiertos. Inexcusablemente, los estándares especialmente en el área del control de autoridades se deben armonizar.

Los colectivos bibliotecarios y archivísticos han desarrollado unos marcos teóricos, unos modelos conceptuales y unas normas propias, si bien cada vez más próximas y con objetivos confluyentes. Recientemente Garzón y Peset (2018a) han realizado las equivalencias entre las diferentes normativas sobre autoridades de persona (FRAD, ISAAR-CPF, EAC-CPF, GARR, MARC21, MADS, Dublin Core RDF, Marc XMLy FOAF), demostrando que es posible que el mundo bibliotecario y archivístico confluyan con relativa facilidad entre ellos y con la web semántica. Estas mismas autoras (2018b) han propuesto un modelo de formatos de exportación que partiendo del formato MARC21 conciso para datos de autoridad, facilitan el intercambio de datos de autoridades de personas entre sistemas de información con características similares, acordes con algunas iniciativas de la web semántica. En este modelo, establecidas las equivalencias entre formatos y haciendo uso de una serie de campos de MARC 21, se consigue la vinculación de los registros de autoridad con fuentes de información externa y su consiguiente enriquecimiento semántico. Y es precisamente el enriquecimiento semántico de entidades

contextuales, representadas a través de registros de autoridad, lo que permite mejorar la calidad de los datos y aumentar la relevancia de las colecciones digitales de archivos, bibliotecas y museos dentro de la web. Es por ello que se trata de una técnica que debe incorporarse a las tareas de control de autoridades (Agenjo y Hernández, 2018).

En este entorno, la pauta de actuación la está marcando Europeana, el portal de patrimonio cultural europeo que agrega, procesa y enriquece los recursos digitales de bibliotecas, archivos y museos de toda Europa. Para lograrlo ha desarrollado el *Europeana Data Model (EDM)*, un marco normalizador a través del cual se pueden reunir descripciones bibliográficas, archivísticas o de museos y establecer las líneas de actuación sobre las que realizar un enriquecimiento semántico de las colecciones documentales. EDM define cinco clases de entidades contextuales objeto de descripción: agentes (personas y organizaciones), eventos, lugares, fechas y materias con el objetivo de responder a las preguntas ¿quién? ¿qué? ¿cómo? ¿cuándo? y ¿dónde? Gracias al uso de este formato es posible interop-erar entre recursos de diferentes instituciones y relacionar los datos de las colecciones con otros datos disponibles en la web basándose en los principios generales de la web semántica (Juanes, 2015a, 2015b).

### 3. La codificación de las descripciones de autoridades archivísticas en DARA

El sistema de información de los archivos de Aragón se conoce como DARA, acrónimo de Documentos y Archivos de Aragón (<http://www.dara-aragon.es>). Se trata de una plataforma tecnológica que actúa como un repositorio digital común a todos los archivos integrados en el sistema, permitiendo el acceso a través de Internet al patrimonio documental aragonés. Coordinado por el Gobierno de Aragón, está abierto a todas las instituciones que quieran incorporarse. Actualmente 70 archivos de entidades de diferente tipología difunden sus fondos documentales mediante DARA. A fecha de abril de 2019 es posible acceder a 775.831 registros descriptivos, incluyendo 235.797 documentos digitalizados.

Desde hace una década los archivos de distintas administraciones públicas aragonesas, entre ellos los tres históricos provinciales, tienen implantada una aplicación informática que utiliza los formatos *MARC21* para la representación y codificación de las descripciones archivísticas. Concretamente *MARC21 para datos bibliográficos* como formato de las descripciones de documen-

tos en sus diferentes niveles y *MARC21 para datos de autoridad* como formato para la descripción y control de personas, familias, instituciones, lugares y temas. Estos mismos formatos sirven de base para la normalización en DARA, utilizándose como estructura común para la comunicación e intercambio de las descripciones, soportando tanto la importación e integración de los recursos provenientes de los diferentes archivos del sistema como su difusión a través de la web.

Para lograr la integración en DARA de las descripciones archivísticas normalizadas y posibilitar posteriormente su eficaz búsqueda, recuperación y difusión, se requiere que las estructuras de las normas de descripción usadas por los archivos del sistema estén alineadas con los formatos MARC, unos formatos diseñados especialmente para los sistemas bibliotecarios. Así mismo son necesarios unos criterios homogéneos que garanticen la aplicación uniforme de las descripciones por todos los archivos que comparten sus registros a través de DARA.

Con este empeño, en el año 2010 se publicaba *EDARA: Estructura de Datos de los Archivos de Aragón*, creada para codificar la información archivística de materiales textuales impresos y manuscritos, archivos de ordenador, mapas, música, materiales visuales y materiales mixtos, basándose en el formato de codificación *MARC21 para datos bibliográficos* y en la norma ISAD(G). Gracias a la aplicación de EDARA, se ha logrado integrar en un mismo buscador las descripciones de los documentos procedentes de plataformas y aplicaciones diferentes, y que a su vez los archivos integrados en DARA no dependan de programas específicos para almacenar y compartir su información. Por su parte, en virtud a la alineación de MARC21 con ISAD (G), es posible la conversión y exportación de las descripciones documentales de DARA a formatos de codificación archivísticos como EAD y a formatos de propósito más generalista como Dublin Core o EDM.

El Grupo de Trabajo de Autoridades de Aragón (GTAA) ha trabajado intensamente desde el año 2001 en la normalización de la descripción de autoridades, cuyo resultado más sobresaliente ha sido la publicación de la *Norma Aragonesa para la Descripción de Autoridades de Archivos (ARANOR)*. No obstante, y a pesar de los tempranos trabajos realizados para la elaboración de un fichero de autoridades de personas y familias de Aragón, seguía siendo una tarea pendiente la codificación informática de las descripciones completas de las autoridades archivísticas y puntos de acceso (personas, familias, instituciones, lugares y materias) para su integración en DARA. Con este objetivo un equipo formado por representantes de los archivos históricos provinciales

aragoneses, del archivo de la Universidad de Zaragoza, y de la empresa Keensoft iniciaron los trabajos para la elaboración de una nueva norma: *DARANOR Estructura de datos para registros de autoridad en archivos de Aragón*. Esta norma incluirá:

1. La definición de una estructura de datos para la descripción de autoridades de archivo válida para DARA, que alinee la estructura de la segunda edición de la Norma Aragonesa para la Descripción de Autoridades de Archivos (ARANOR) con el formato MARC21 para datos de autoridad, según su última versión de noviembre de 2018. La estructura de datos definida debe permitir:

- La redacción de registros completos e independientes, de los diferentes tipos de autoridades archivísticas, especialmente de personas, familias e instituciones, pudiéndose extender a lugares y temas.
- Establecer las relaciones necesarias entre los registros de autoridad y otros tipos de entidades archivísticas.
- El enriquecimiento semántico de los registros de autoridad mediante el vínculo con otras entidades relacionadas y con fuentes de información externa disponibles en la web.
- La interoperabilidad de DARA con otros sistemas de información, garantizando el intercambio de información y la utilización de la información intercambiada.

2. Los componentes y los contenidos de los registros de autoridad. El núcleo de *DARANOR* lo constituyen las equivalencias establecidas entre los elementos de *ARANOR* / *ISAAR-CPF* y los campos y subcampos de *MARC21* para datos de autoridad, y que se ofrece en la Tabla I. En su elaboración se ha tenido en cuenta las descripciones de autoridades realizadas por la Biblioteca del Congreso de Estados Unidos, la Biblioteca Nacional de España y la Biblioteca Virtual Ignacio Larrañendi de Polígrafos, así mismo se han analizado las paridades identificadas por Garzón y Paset (2018a).

Elementos ISAAR - ARANOR	Etiqueta MARC	Descripción
1. Área de identificación		
1.1. Tipo de entidad	075\$a	Tipo de entidad (R)
1.2. Forma(s) autorizada(s) del nombre		
1.2. Nombre de persona	100 1\$a	Nombre personal (NR)
1.2. Fecha de persona como calificador	100 1\$d	Fechas asociadas con el nombre (NR)

1.2. Información personal como calificador	100 1\$g	Información miscelánea (R)
1.2. Atributos	100 1\$j	Calificador de atribución (R)
1.2. Nombre de familia	100 3\$a	Nombre de familia
1.2. Nombre de institución	110 2\$a	Nombre corporativo o de jurisdicción (NR)
1.2. Nombre de institución subordinada	110 2\$b	Unidad subordinada [R]
1.2. Fechas de institución	110 2\$d	Fecha de la reunión o de la firma de un tratado [R]
1.2. Información institucional como calificador	110 2\$g	Información miscelánea [NR]
1.2. Nombre de reunión	111 2\$a	Nombre de reunión o de jurisdicción [NR]
1.2. Lugar de reunión	111 2\$c	Lugar de la reunión [NR]
1.2. Fecha de reunión	111 2\$d	Fecha de la reunión [NR]
1.2. Información de la reunión como calificador	111 2\$g	Información miscelánea [NR]
1.3. Formas paralelas del nombre		
1.3. Formas paralelas nombre persona	400 1\$a\$!	Idioma de una obra [NR]
1.3. Formas paralelas nombre familia	400 3\$a\$!	Idioma de una obra [NR]
1.3. Formas paralelas nombre institución	410 2\$a\$!	Idioma de una obra [NR]
1.3. Formas paralelas nombre reunión	411 2\$a\$!	Idioma de una obra [NR]
1.4. Formas normalizadas del nombre según otras reglas		
1.4. Formas normalizadas del nombre de persona según otras reglas	700 1\$a\$4	Asiento ligado al encabezamiento – nombre personal [R] Nombre personal [NR]
1.4. Formas normalizadas del nombre de familia según otras reglas	700 3\$a\$4	Asiento ligado al encabezamiento – nombre personal [R] Nombre de familia
1.4. Formas normalizadas del nombre de institución según otras reglas	710 2\$a\$4	Asiento ligado al encabezamiento – nombre corporativo [R] Nombre corporativo o de jurisdicción [NR]
1.4. Formas normalizadas del nombre de reunión según otras reglas	711 2\$a\$4	Asiento ligado al encabezamiento – nombre reunion [R] Nombre de reunión o de jurisdicción [NR]
1.5. Otras formas del nombre		
1.5. Otras formas del nombre persona	400 1\$a	Referencia de envío "vease" - nombre personal [R] Nombre personal [NR]
1.5. Otras formas del nombre familia	400 3\$a	Referencia de envío "vease" - nombre personal [R] Nombre de familia [NR]
1.5. Otras formas del nombre institución	410 2\$a	Referencia de envío "vease" - nombre corporativo [R] Nombre corporativo o de jurisdicción [NR]
1.5. Otras formas del nombre reunión	411 2\$a	Referencia de envío "vease" - nombre de reunion [R] Nombre de reunión o de jurisdicción [NR]
1.6. Identificadores para instituciones	024\$a\$2	Otro identificador normalizado [R] Fuente del número o código [NR]
<b>2. Área de descripción</b>		
2.1. Fechas de existencia		
2.1. Persona: Fechas de nacimiento	046\$f	Fecha nacimiento (NR)

2.1. Persona: Fechas de muerte	046\$g	Fecha muerte (NR)
2.1. Familia: Fechas inicio	046\$s	Comienzo de periodo (NR)
2.1. Familia: Fechas fin institución	046\$t	Fin de periodo (NR)
2.1. Institución: fecha de establecimiento	046\$q	Fecha de establecimiento (NR)
2.1. Institución: fecha de finalización	046\$r	Fecha de finalización (NR)
Género	375\$a	Género
2.2. Historia		
2.2. Biografía	678 0\$a	Datos biográficos o históricos [R]
2.2. Historia administrativa	678 1\$a	Historia administrativa
	678\$u	Identificador Uniforme del Recurso [R]
2.3. Lugares		
2.3. Lugar de nacimiento	370\$a	Lugar de nacimiento (NR)
2.3. Lugar de muerte	370\$b	Lugar de muerte (NR)
2.3. Lugar de residencia	370\$e	Lugar de residencia / sede (R)
2.3. País	370\$c	País asociado (R)
2.3. Inicio periodo	370\$s	Inicio periodo (NR)
2.3. Fin periodo	370\$t	Fin periodo (NR)
	370\$2	Fuente del término (NR)
	370\$6	Enlace (NR)
2.4. Estatuto jurídico		
2.4. Naturaleza jurídica	368\$a	Tipo de entidad corporativa (R)
2.4. Tipo de institución	368\$b	Tipo de jurisdicción (R)
2.4. Fecha inicio estatuto jurídico	368\$s	Inicio periodo (NR)
2.4. Fecha fin estatuto jurídico	368\$t	Fin periodo (NR)
2.5. Funciones, ocupaciones y actividades		
2.5. Actividades	372\$a	Campo de actividad (R)
2.5. Actividades	372\$s	Inicio periodo (NR)
2.5. Actividades	372\$t	Fin periodo (NR)
2.5. Ocupaciones	374\$a	Ocupación (R)
2.5. Ocupaciones	374\$s	Inicio periodo (NR)
2.5. Ocupaciones	374\$t	Fin periodo (NR)
2.6. Atribución(es) / Fuente(s) legal(es)	670\$a\$i	Fuente de la cita (NR)
2.7. Estructura(s) interna(s)/Genealogía		
Genealogía familiar	678 0\$b	Biografía o datos históricos (R). Expansión (NR) (Ampliación de los datos biográficos dados en el subcampo \$ a)
Estructura institución	678 1\$b	Biografía o datos históricos (R). Expansión (NR). (Ampliación de los datos históricos dados en el subcampo \$ a)
2.8. Contexto general	680\$i	Nota general pública. Texto explicativo (R)
<b>3. Área de relaciones</b>		
3.1. Nombre(s)/ Identificadores de las instituciones, personas o familias relacionadas		
Persona	500 1\$a	Referencia de envío "vease ademas" - nombre personal

Familia	500 3\$a	Referencia de envío "vease ademas" - nombre familia
Institución	510 2\$a\$b	Referencia de envío "vease ademas" - nombre corporativo
Reunión	511 2\$a\$c	Referencia de envío "vease además" - nombre de Reunion
3.2. Naturaleza de la relación		
Persona	500 1\$e	Término de relación [R]
Familia	500 3\$e	Término de relación [R]
Institución	510 2\$e	Término de relación [R]
Reunión	511 2\$e	Término de relación [R]
3.3. Descripción de la relación		
Persona	500 1\$w\$r\$i	Subcampo de control [NR] Frase de instrucción [NR]
Familia	500 3\$w\$r\$i	Subcampo de control [NR] Frase de instrucción [NR]
Institución	510 2\$w\$r\$i	Subcampo de control [NR] Frase de instrucción [NR]
Reunión	511 2\$w\$r\$i	Subcampo de control [NR] Frase de instrucción [NR]
3.4. Fechas de la relación		
Persona	500 1\$d	Fechas asociadas con el nombre [NR]
Familia	500 3\$d	Fechas asociadas con el nombre [NR]
Institución	510 2\$b	Fechas asociadas con el nombre [NR]
Reunión	511 2\$d	Fechas asociadas con el nombre [NR]
4. Área de control		
4.1. Identificador del registro de autoridad		
	001	Número de Control (NR)
	035\$a	Número de control del sistema (R)
4.2. Identificadores de la institución		
	040\$a	Agencia de catalogación original (NR)
	040\$c	Agencia de transcripción (NR)
4.3. Reglas y/o convenciones		
	040\$e	Convenciones de descripción (R)
4.4. Estado de elaboración		
	042\$a	Código de autenticación (R)
4.5. Nivel de detalle		
	?	
4.6. Fechas de creación, revisión o eliminación		
	005	Fecha y hora de la última transacción (NR)
4.7. Lengua(s) y escritura(s)		
	040\$b	Idioma de catalogación (NR)
4.8. Fuentes		
	670\$a\$i	Fuente de la cita (NR)
4.9. Notas de mantenimiento		
	688\$a	Nota de historial de aplicación (NR)

*Tabla 1. Equivalencias entre ARANOR y MARC21 para datos de autoridad*

Una de las principales dificultades encontradas a la hora de establecer las equivalencias se ha debido a la mayor información que ofrecen los registros de autoridades archivísticas con respecto a los bibliográficos, por lo que no siempre ha resultado sencillo encontrar los campos y subcampos de MARC21 en los que incluir los datos de los elementos, atributos y calificadores de ARANOR e

ISAAR-CPF. Así, especialmente problemático ha sido identificar un campo de MARC específico para el elemento "2.6. Atribución(es)/Fuente(s) legal(es)", por lo que se ha optado por utilizar el campo "670. Fuente en la que se localizaron datos" compartido con el elemento "4.8 Fuentes". Esta solución se inspira en las descripciones de autoridades realizadas por las bibliotecas y en las que se utiliza el campo 670 para las fuentes bibliográficas y legislativas. Por su parte, para el elemento "4.5. Nivel de detalle" no hemos podido asignar campo alguno, por lo que queda pendiente su resolución. Por otra parte MARC21 de autoridades nos ofrece la posibilidad de enriquecer los registros con datos no contemplados o no individualizados por ARANOR e ISAAR-CPF como el campo "375. Género".

#### 4. Beneficios de la codificación basada en las equivalencias definidas

La aplicación de la norma DARANOR y más concretamente la correspondencia establecida entre ARANOR y MARC21 van a mejorar el funcionamiento interno del sistema de archivos aragoneses a la par que va a abrir nuevas vías para explotar y compartir a través de la web los datos de autoridades de DARA y los de otros sistemas de información, ya sean archivísticos, bibliotecarios o de cualquier otro tipo.

##### 4.1. Mejoras en el funcionamiento interno de DARA

En primer lugar el trabajo de equivalencias realizado ha supuesto una fase previa e imprescindible para la efectiva implantación de un módulo para la creación y gestión de registros de autoridades, conforme a las normas descriptivas aragonesas, en la plataforma tecnológica de DARA y en las aplicaciones informáticas de los archivos de Aragón que utiliza los formatos MARC21.

Este módulo permite elaborar registros de autoridad de personas, familias e instituciones completos, conteniendo todos los elementos y subelementos estipulados en la segunda edición de la norma ARANOR, incluidos calificadores para identificar y distinguir autoridades con el mismo nombre. Gracias a los nuevos campos introducidos por las últimas actualizaciones de MARC21 (Library of Congress, 2014) en su acercamiento a los archivos, museos y otras organizaciones de gestión de la información, los registros de autoridad además de plasmar de manera estructurada y diferenciada los elementos y calificadores acordados en ARANOR, pueden comprender otros datos de interés, como el género de las personas. Por otra parte, y dado que los campos de MARC21 para datos de autoridad dan cabida a la

representación de descripciones de lugares y temas, es factible extender el alcance del módulo de autoridades a la descripción de estas entidades. De este modo el sistema de archivos aragonés puede llevar a la práctica un modelo de descripción multientidad que además de los documentos abarque a los agentes implicados en la producción y gestión de la documentación (personas, familias o instituciones) así como los lugares y temas relacionados. La elaboración de registros de autoridad independientes de las descripciones documentales posibilita que su información pueda asociarse a diferentes entidades documentales sin tener que repetirla. Así mismo gracias a las relaciones definidas en ARANOR y codificadas por MARC21 se pueden conectar todos los registros de autoridad creados.

Otro de los beneficios reportados por DARANOR es la disposición de una estructura común de datos de autoridad que permite integrar en DARA las descripciones de los diferentes archivos del sistema. De este modo, y contando con las aportaciones de los miembros del sistema se va a poder construir de manera cooperativa el fichero de autoridades de instituciones, personas y familias de los archivos aragoneses. Un fichero que contendrá los nombres normalizados de personas, familias e instituciones que deben ser utilizados como formas preferentes por los archivos del sistema aragonés en sus descripciones, así como sus formas alternativas u otras formas del nombre. El fichero se debe convertir por lo tanto, en un instrumento de referencia de uso habitual por los catalogadores de los archivos aragoneses. Por otra parte al compartirse de manera abierta todo el fichero, se evitarán trabajos redundantes como la repetición de descripciones de autoridades ya registradas. Este trabajo colaborativo realizado de manera coordinada y controlada mediante el rol de Gestor de Autoridades, proporcionará consistencia al buscador de DARA facilitando a sus usuarios la identificación y recuperación de información de manera más pertinente. Por su parte, este fichero va a suponer un incremento notable en la calidad y cantidad informativa ofertada por DARA, al añadir los datos de miles de personas, familias e instituciones que crearon, gestionaron, usaron o están citados en los documentos custodiados por los archivos aragoneses y sus relaciones.

#### 4.2. Nuevas relaciones con sistemas de información externos

La estructura de datos definida por DARANOR permite a DARA relacionarse e interactuar con otros sistemas de información o fuentes de datos externos, obteniendo beneficios variados de estos contactos.

En primer lugar podemos señalar la posibilidad de enlazar los registros de autoridad con datos aportados por fuentes externas –ajenas o propias- disponibles en la web. Estas fuentes podrán ser archivísticas, bibliográficas, museísticas o de cualquier otro tipo, así como expresadas en diferentes lenguas o escrituras. Con ello se consigue el enriquecimiento semántico de los registros, lo que mejora la calidad de los datos y aumenta la relevancia informativa de DARA. Este procedimiento consiste en incluir en cada una de las entidades enriquecidas el identificador uniforme de recursos (URI) del dato o conjunto de datos a vincular. El ejemplo más representativo de enriquecimiento semántico lo ofrece el uso del campo 024 (otro identificador normalizado) de MARC21 en equivalencia con el elemento 1.6 (identificadores para instituciones) de ARANOR. De este modo se puede establecer la vinculación con los registros de una pluralidad de conjuntos de datos y vocabularios como VIAF, Geonames, DBpedia, Wikidata, el servicio Linked Data de la *Library of Congress* (<http://id.loc.gov/>), [datos.bne.es](http://datos.bne.es)... Otro recurso para integrar información externa es el campo 856 (localización y acceso electrónico) que provee información suplementaria, disponible en forma electrónica, sobre la entidad para la que se crea el registro de autoridad (Juanes 2015; Agenjo y Hernández, 2018; Garzón y Peset 2018b). Así mismo MARC21 nos ofrece unos subcampos que permiten enriquecer elementos concretos de nuestros registros, como el \$u (identificador uniforme del recurso) o el \$1 (URI de objetos del mundo real).

```

001 XX950223
003 SpMaBN
005 20140218
008 901120nn ajjnnaabn a aaa
016 $aBNE19900080491
024 7 $ahhttp://viaf.org/viaf/76310334$2viaf
024 7 $ahhttp://www.isni.org/isni/0000000108150650$2isni
024 7 $ahhttps://www.wikidata.org/wiki/Q360512$2wikidata
040 $aSpMaBN$bspa$cSpMaBN$erdc$fembne
046 $f1601$g1658
100 1 $aGracián, Baltasar$d1601-1658
372 $aFilósofos$vWikidata
374 $aFilósofos$aEscritores$vWikidata
375 $aMasculino
377 $!Español
370 $aBelmonte de Gracián, Zaragoza, España$v(Wikidata)$0
400 1 $aGracián, Baltasar de$d1601-1658
400 1 $aGracián, Lorenzo$d1601-1658
400 1 $aGracián y Morales, Baltasar$d1601-1658
400 1 $aGelaxi'an, Ba'ertasa'er$d1601-1658
400 1 $a格拉西安, 巴尔塔萨尔$d1601-1658
670 $aEl arte de la prudencia, 1993$bport. (Baltasar Gracián) p. VI
670 $aZhi hui shu, 2003$bport. (巴尔塔萨尔-格拉西安 = Ba'ertasa'er-
670 $aObras de Lorenzo Gracián, 1700

```

Figura 1. Fichero en formato MARC21 de Baltasar Gracián obtenido de la Biblioteca Nacional de España (<http://catalogo.bne.es>)

La estructura de datos desarrollada permite a DARA intercambiar información con diferentes agencias, servicios, sistemas o unidades productoras y/o consumidoras de datos de autoridad. Así, el uso del estándar bibliotecario MARC21 de autoridades garantiza la interoperabilidad con numerosos sistemas bibliotecarios, hasta ahora las principales fuentes generadoras de datos de autoridad. De este modo se pueden obtener registros para la importación y posterior uso por DARA, pero teniendo presente que generalmente se precisará una información más extensa que la aportada por los ficheros bibliotecarios (Figura 1, en la página anterior). El fichero de autoridades de DARA también se podrá asociar con los desarrollados por las bibliotecas del Gobierno de Aragón y de otras instituciones públicas consiguiendo un punto de acceso centralizado al patrimonio documental y bibliográfico aragonés.

Pero como ya hemos visto, el sector bibliotecario no es el único sector que crea y utiliza datos de autoridad. Archivos, museos, entidades de gestión de derechos y otros, crean ficheros para servir a sus propios fines. La conversión y exportación de la información codificada de acuerdo a MARC21 a los formatos de intercambio utilizados por estas diferentes comunidades posibilita la integración de DARA en proyectos cooperativos que incrementen la difusión, conocimiento y uso del patrimonio documental aragonés. Mediante el esquema MARCXML, los registros de DARA se pueden representar en XML, transformándose de este modo en recursos digitales compartidos y reutilizables a través de la web por diferentes sistemas no exclusivamente bibliotecarios. Los ficheros XML a su vez se pueden convertir a otros sistemas de metadatos como *Europeana Data Model*. La alineación realizada entre MARC21 de autoridades y ARANOR también permiten su conversión a EAC-CPF, el estándar de codificación e intercambio de autoridades archivísticas.

De este modo quedan abiertas las puertas a la inclusión de los registros de autoridad de DARA en iniciativas internacionales orientadas a la difusión de recursos archivísticos como el Archives Portal Europe (s. d.) que ha desarrollado el perfil apeEAC-CPF para la integración de los datos de autoridad. E igualmente puede aportar sus registros a Europeana, el punto de acceso único a millones de recursos digitales de archivos, museos, bibliotecas y colecciones audiovisuales europeas y que ofrece un escaparate internacional que mejorará la visibilidad y el uso de DARA.

#### 4.2. Proyección hacia la web semántica

El fichero de autoridades de DARA también se puede proyectar hacia una web semántica de datos abiertos y enlazados. Ello implica que el centro de atención del tratamiento y gestión pase de los registros de autoridad como unidades cerradas contendedoras de información, a los propios datos que los conforman considerados de manera individualizada. De este modo, ya no se tratarán los registros completos de personas o instituciones, sino los datos específicos que se disponen de los mismos: fechas de nacimiento, de muerte, lugares, actividades, ocupaciones, biografías, historias institucionales... Para ello todos estos datos individualizados deben estar identificados mediante URIs. Su descripción se guiará por un estándar semántico como RDF (*Resource Description Framework*) que los representará en forma de tripletes del tipo sujeto-predicado-objeto. Además para la definición en formato semántico será necesario disponer de un vocabulario u ontología que recoja de manera procesable automáticamente la información sobre los tipos de recursos a los que se refiere el conjunto de datos (clases) y las características que definen a cada uno de estos recursos (propiedades).

LC control no.: n 50047296  
 LCCN Permalink: <https://lccn.loc.gov/n50047296>  
 HEADING: Gracián y Morales, Baltasar, 1601-1658  
 000 01170cz a2200313n 450  
 001 2800609  
 005 20120904074916.0  
 008 800812n| azannaabn |a aaa  
 010 \_\_ |a n 50047296  
 035 \_\_ |a (OCoLC)oca00082382  
 035 \_\_ |a (Uk)000168107  
 040 \_\_ |a DLC |b eng |c DLC |e rda |d DLC |d UPB |d OCoLC |d Uk |d DL  
 046 \_\_ |f 1601 |g 16581206  
 053 \_\_ |a PQ6398.G3 |b PQ6398.G4  
 100 1\_ |a Gracián y Morales, Baltasar, |d 1601-1658  
 377 \_\_ |a spa  
 400 1\_ |a Gracjan, Baltazar, |d 1601-1658  
 400 1\_ |a Gratian, Baltasar, |d 1601-1658  
 400 1\_ |a Gratsian, Baltazar, |d 1601-1658  
 400 1\_ |a Morales, Baltasar Gracián y, |d 1601-1658  
 400 1\_ |a Gracián, Lorenzo, |d 1601-1658.  
 400 1\_ |a Gracián, Baltasar, |d 1601-1658  
 400 1\_ |a Gracian, Balthasar, |d 1601-1658  
 400 1\_ |a Gracián, Baltazar, |d 1601-1658  
 670 \_\_ |a His The art of worldly wisdom, 1993: |b t.p. (Balthasar Gracian)  
 670 \_\_ |a His palacio sin puertas, 1991: |b t.p. (Baltasar Gracián)  
 670 \_\_ |a Agudeza y arte de ingenio, 2004: |b t.p. (Baltasar Gracián) p. [xii  
 952 \_\_ |a RETRO  
 953 \_\_ |a xx00 |b be04

Figura 2. Registro de Baltasar Gracián: formato Marc21 de autoridades (<https://bit.ly/2Gs50sy>)

La adaptación de los registros de autoridad de DARA al modelo de la web semántica se puede realizar a través de MADS, el esquema para la codificación de datos en formato MARC21 a RDF, y del esquema RDF para la descripción de metadatos de autoridades MADS/RDF (Library of Congress, 2015) diseñado para su uso con valo-

res controlados para nombres (personal, corporativo, geográfico, etc.), tesauros, taxonomías, sistemas de encabezado de materia y otras listas de valores controlados (Figuras 2, 3 y 4). De esta manera se conseguirá la transformación de toda la información contextual de DARA en un conjunto de datos abiertos y reutilizables.

```
<rdf:RDF>
  <madsrdf:PersonalName rdf:about="http://id.loc.gov/authorities/names/n50047296">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Authority"/>
    <madsrdf:authoritativeLabel xml:lang="en">Gracián y Morales, Baltasar, 1601-1658</madsrdf:authoritativeLabel>
  </madsrdf:elementList rdf:parseType="Collection">
    <madsrdf:FullNameElement>
      <madsrdf:elementValue xml:lang="en">Gracián y Morales, Baltasar</madsrdf:elementValue>
    </madsrdf:FullNameElement>
    <madsrdf:DateNameElement>
      <madsrdf:elementValue xml:lang="en">1601-1658</madsrdf:elementValue>
    </madsrdf:DateNameElement>
    <madsrdf:elementList>
      <madsrdf:hasVariant>
        <madsrdf:PersonalName>
```

Figura 3. Registro de Baltasar Gracián: extracto del fichero RDF generado con MADS/RDF (<http://id.loc.gov/authorities/names/n50047296.rdf>)

```
<http://id.loc.gov/authorities/names/n50047296>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.loc.gov/mads/rdf/v1#PersonalName> .

<http://id.loc.gov/authorities/names/n50047296>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.loc.gov/mads/rdf/v1#Authority> .

<http://id.loc.gov/authorities/names/n50047296>
<http://www.loc.gov/mads/rdf/v1#authoritativeLabel>
"Gracián y Morales, Baltasar, 1601-1658"@en .
```

Figura 4. Registro de Baltasar Gracián: tripletas RDF (MADS and SKOS) (<http://id.loc.gov/authorities/names/n50047296.nt>)

El fichero de autoridades de DARA en forma de datos abiertos se podrá publicar para su explotación, junto a otros *datasets*, en Aragón Open Data, el portal de datos abiertos del Gobierno de Aragón (<https://opendata.aragon.es/>). Además se podrá crear un *triplestore* o base de datos para el almacenamiento y recuperación de las tripletas RDF de DARA, así como un punto SPARQL que permita realizar consultas sobre las descripciones semánticas de los datos de autoridad del sistema de archivos de Aragón y cuyos resultados se obtendrán en formatos procesables automáticamente para que puedan ser reutilizados directamente.

#### 4. Conclusiones

El sistema de información de los archivos de Aragón (DARA) necesita avanzar en la normalización, control y difusión de las autoridades archi-

vísticas. La elaboración de una norma para la estructuración de datos para registros de autoridad en los archivos de Aragón es una prioridad inexcusable para el desarrollo del sistema y garantía de su integración con otros sistemas de información. En este proceso debe tenerse en consideración el acercamiento a otras instituciones culturales como bibliotecas o museos, así como las potencialidades que ofrece la web semántica y los modelos de datos abiertos y vinculados.

La aplicación de esta norma va a permitir que DARA —al igual que están haciendo otros grandes sistemas archivísticos como el Portal de Archivos Españoles PARES— ponga en práctica un modelo de descripción multientidad que abarque a los documentos en sus diferentes niveles, a los agentes implicados en la producción y gestión de la documentación (personas, familias o instituciones), pudiendo además dar cabida a lugares y temas relacionados. Uno de los resultados más positivos será la confección de manera colaborativa del fichero de autoridades de instituciones, personas y familias de los archivos aragoneses. Pero además, la estructura definida por DARANOR va a permitir que DARA se relacione e interactúe con otros sistemas de información o fuentes de datos externos, pudiendo conseguir desde el enriquecimiento semántico de sus registros hasta su integración y participación en proyectos cooperativos regionales, nacionales o internacionales, tanto si están orientados a la difusión de recursos archivísticos como si además incluyen los bibliotecarios, museísticos o los de otras instituciones culturales. Además si se avanza hacia la conversión de sus registros de autoridad en datos abiertos, enlazados y procesables semánticamente podrá participar de iniciativas como Aragón Open Data o incluso desarrollar su propia infraestructura para la explotación y reutilización de estos datos.

#### Referencias

- Archives Portal Europe (s. d.). apeEAC-CPF. <http://wiki.archives-portaleurope.net/index.php/apeEAC-CPF> (2019-06-15).
- Agenjo-Bullón, Xavier; Hernández-Carrascal, Francisca (2018). Registros de autoridades, enriquecimiento semántico y Wikidata. // Anuario ThinkEPI. 12 (2018) 361-372.
- Bonal Zazo, José Luis; Generelo Lanaspa, Juan José; Travesí de Diego, Carlos (2006). Manual de Descripción Multinivel: Propuesta de adaptación de las normas internacionales de descripción archivística. 2º ed. Rev. [Valladolid]: Junta de Castilla y León, 2006.
- Coalition for Networked Information (2017). Library and IT Partnerships with Campus Museums and Archives. Report of a CNI Executive Roundtable. [Washington]: Coalition for Networked Information, 2017.
- Comisión de Normas Españolas de Descripción Archivística (2017). NEDA-MC. Modelo conceptual de descripción archivística: Entidades, Relaciones y Atributos. Madrid: Ministerio de Educación, Cultura y Deporte, 2017.
- Encoded Archival Context Working Group de la Society of American Archivists. Contexto Archivístico Codificado –

- Instituciones, Personas y Familias (EAC-CPF). Repertorio de Etiquetas. Versión 2010 (Edición Inicial). Traducción de la versión original en inglés al español. Madrid: Subdirección General de los Archivos Estatales, 2012.
- Garzón Farinós, M. F. (2015). El registro de autoridades personales tras la aparición del web. Valencia: Editorial Universitat Politècnica de València, 2015. Tesis doctoral.
- Garzón Farinós, M. F. y Peset, F. (2018a) Normativas interoperables sobre control de autoridades para bibliotecas digitales semánticas. // Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información. 77 (2018) 167-188.
- Garzón Farinós, M. F.; Peset, F. (2018b) Los registros de autoridades personales de la comunidad bibliotecaria revisitados en la web semántica. // Anales de Documentación. 21:1 (2018) 1-15.
- Grupo de Estructura de Datos de los Archivos Aragoneses (2010). EDARA: Estructura de Datos de los Archivos de Aragón. Zaragoza: Gobierno de Aragón, Departamento de Educación, Cultura y Deporte, 2010.
- Grupo de Trabajo de Autoridades de Aragón (2014). Norma Aragonesa para la Descripción de Autoridades de Archivos (ARANOR). 2ª ed. [Zaragoza]: Departamento de Educación, Universidad, Cultura y Deporte, 2014.
- Grupo de Trabajo de Autoridades de Aragón (2004). Compartir autoridades: la experiencia aragonesa. // Tabula. (2004), 89-105.
- Grupo de Trabajo de Autoridades de Aragón (2004). La normalización y el control de los puntos de acceso en la descripción archivística: fórmulas de colaboración para el control de autoridades. XV Jornadas de Archivos Municipales. Madrid: Comunidad de Madrid, Dirección General de Archivos, Museos y Bibliotecas, 2004, 283-311.
- Grupo de Trabajo de Autoridades de Aragón (2008). La norma de autoridades de Aragón (ARANOR): Aplicación práctica. Compartir archivos: Actas de las VIII Jornadas de Archivos Aragoneses. Huesca: Gobierno de Aragón; Diputación Provincial, 2008, t. I, 97-102.
- Grupo de Trabajo de Descriptores Geográficos de Aragón (2015a). Geodara: Norma aragonesa para la creación de descriptores geográficos. Zaragoza : Gobierno de Aragón, Departamento de Educación, Cultura y Deporte, 2015.
- Grupo de Trabajo de Descriptores Geográficos de Aragón (2015b). Lista de descriptores geográficos de Aragón: localidades, municipios, comarcas y provincias actuales y desaparecidos. Zaragoza : Gobierno de Aragón, Departamento de Educación, Cultura y Deporte, 2015.
- Grupo de Trabajo de Descriptores Geográficos de Aragón (2017). Lista de descriptores geográficos de Aragón. Construcciones. Zaragoza : Gobierno de Aragón, Departamento de Educación, Cultura y Deporte, 2017.
- Grupo Incubador de Datos Vinculados (2012). Informe Final del Grupo Incubador de Datos Vinculados de Bibliotecas. Madrid: Fundación Ignacio Larramendi, 2012.
- Consejo Internacional de Archivos (1992). Declaración de Principios sobre la Descripción de Archivos.
- Consejo Internacional de Archivos (2004). Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias. 2ª ed.
- Consejo Internacional de Archivos (2016). Records in Context Conceptual Model (RiC-CM).
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) (2004). Directrices para Registros de Autoridad y Referencia (GARR).
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) (2009). Requisitos Funcionales de los Datos de Autoridad (FRAD).
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) (2012). Declaración Profesional de la IFLA sobre el Control Bibliográfico Universal.
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) (2017). Declaración de principios internacionales de catalogación.
- Juanes Hernández, César (2015a). Enriquecimiento semántico de colecciones digitales a través del Europeana Data Model. Digibis, 2015.
- Juanes Hernández, César (2015b). El Europeana Data Model: su papel normalizador en los procesos de enriquecimiento semántico a través de tecnología Linked Open Data. // Actas Fesabid, 2015. 147-160.
- Lafuente Urien, Atrantxa (2016). PARES hacia la web semántica: autoridades y tesauros en los archivos estatales. El lenguaje sobre el patrimonio: Estándares documentales para la descripción y gestión de colecciones. - [Madrid]: Ministerio de Educación, Cultura y Deporte, Subdirección General de Documentación y Publicaciones, 2016. 181-194.
- Lafuente Urien, Atrantxa (2018). En torno al nombre del productor: concepto de autoridad archivística, de las ISAAR-CPF al RIC. // Boletín ANABAD. 68:3 (jul-dic. 2018) 134-147.
- Lafuente Urien, Atrantxa; Muriel Hernández, Santiago (2014). Control de autoridades y puntos de acceso en el Portal de Archivos Españoles (PARES). Madrid: Ministerio de Educación Cultura y Deporte, 2014.
- Library of Congress' Network Development and MARC Standards Office. MADS. <http://www.loc.gov/standards/mads/> (2019-06-12).
- Library of Congress' Network Development and MARC Standards Office. MARCXML. <http://www.loc.gov/standards/marcxml/> (2019-06-12).
- Library of Congress (2014). RDA in MARC. <http://www.loc.gov/marc/RDAinMARC.html> (2019-06-18).
- Library of Congress (2015). MADS/RDF: Metadata Authority Description Schema in RDF. <http://id.loc.gov/ontologies/madsrdf/v1.html#RWO> (2019-06-18).
- López de Guillamón, Ignacio (2012). Metadatos para la codificación y transferencia de datos bibliográficos. // Boletín ANABAD. 62:2, (abr.-jun. 2012), 219-235.
- Llanes Padrón, Dunia; Suárez Maresma, Isabel. La descripción del productor y los registros de autoridades en los contextos archivísticos. // Revista Interamericana de Bibliotecología. 39:3 (2016) 301-313.
- Navarro Santolaria, Santiago (2008). Sistema de archivos y normas de descripción: hacia un manual de estructura de datos del Gobierno de Aragón. Compartir archivos: Actas de las VIII Jornadas de Archivos Aragoneses. Huesca: Gobierno de Aragón; Diputación Provincial, 2008. T. I, 71-80.
- Puente, Marcelo de la (2018). EAD y EAC: Aplicación de metadatos en la descripción de colecciones archivísticas. Consultora de Ciencias de la Información. Serie Documentos de Trabajo, 68 (en. 2018).

---

Enviado: 2019-05-05. Segunda versión: 2019-06-12.  
Aceptado: 2019-06-14.

---

---

# A dimensão cultural da organização do conhecimento: uma análise no contexto internacional da Ciência da Informação

*La dimensión cultural de la organización del conocimiento: un análisis en el contexto internacional de la ciencia de la información*

*The cultural dimension of Knowledge Organization: an analysis in the Information Science international context*

---

**José Augusto Chaves GUIMARÃES, Isadora Victorino EVANGELISTA,  
Gabriele de Araújo Medeiros LUZ, Henrique Fiamengue OSAWA**

Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Avenida Hygino Muzzi Filho, 7373, Caixa Postal 420, 17525-900 Marília (Brasil); [chaves.guimaraes@unesp.br](mailto:chaves.guimaraes@unesp.br), [isadora.evangelista@unesp.br](mailto:isadora.evangelista@unesp.br), [gpfapoi@gmail.com](mailto:gpfapoi@gmail.com), [gpfapoi@gmail.com](mailto:gpfapoi@gmail.com)

## Resumen

Los estudios acerca de la dimensión cultural de la organización del conocimiento vienen ganando un papel notorio en las comunicaciones científicas, revelando su importancia en el escenario académico actual. Las investigaciones en comunidades epistémicas pueden ayudar en el estudio de esta temática, identificando a sus principales teóricos y fuentes de información, de modo a auxiliar el desarrollo de futuras investigaciones. Fue posible identificar una fuerte contribución teórica de investigadores actuantes en Brasil, además de fuentes de investigación prominentes en el contexto de ISKO y revistas como *Scire*. En conclusión, fue posible identificar una comunidad que contribuye de manera provechosa al área y que puede servir de base para la construcción de nuevos conocimientos sobre los aspectos culturales de la organización del conocimiento.

**Palabras clave:** Organización del conocimiento. Dimensión cultural. Comunidades epistémicas.

## 1. Introdução

A dimensão cultural é inerente à organização do conhecimento e vem ganhando espaço de pesquisa nesse campo de conhecimento, mais especialmente no universo da ISKO, mais significativo espaço catalizador da pesquisa na área. Essa abordagem centra-se na impossibilidade de “neutralidade” na representação do conhecimento uma vez que essa pode levar a discriminações, a exclusões e a dominações culturais. No entanto, ainda não se tem, de forma sistematizada e analítica, um panorama que permita caracterizar essa comunidade de pesquisadores sobre o tema, em especial quanto a concepções específicas, interrelações, dialogicidades e desdobramentos que revelem influências teóricas.

Desse modo, objetiva-se analisar a composição e o comportamento científico da comunidade epistêmica ligada à dimensão cultural da organização do conhecimento (Meyer, Molineux-Hodgson,

## Abstract

The studies on the cultural dimension of the knowledge organization have been getting a prominent role in scientific communications, revealing their importance in the current academic context. The investigations in epistemic communities can help in the study of this topic in evidence, identifying its main theoreticians and sources of information, in order to help the development of future researches. It was possible to identify a strong theoretical contribution in researchers working in Brazil, as well as prominent research sources in the ISKO context and periodicals such as *Scire*. In conclusion, it was possible to identify a community that contributes in a profitable way to the area and that can serve as a basis for the construction of new knowledge about the cultural aspects of knowledge organization.

**Keywords:** Knowledge Organization. Cultural dimension. Epistemic communities.

2010), mais especialmente no contexto da ISKO, assim como o seu poder de influência teórica no campo, de modo a verificar o grau de presença das questões culturais na literatura científica internacional de organização do conhecimento, a evolução e as tendências conceituais dessa temática, na área; e o “power to influence” dessa comunidade em outros pesquisadores (Wolfram, 2016), sistematizando as redes se estabelecem entre tais referentes, seu grau de dialogicidade e se essas redes poderiam evidenciar correntes teóricas na área.

Tendo como campo de pesquisa a literatura científica da International Society for Knowledge Organization, parte-se de um conjunto de dez autores mais representativos internacionalmente, conforme identificado por Guimarães (2016), com busca de sua produção completa sobre o tema (base KO Literature da ISKO, Web of Science e Scielo), verificação de seu grau de interlocução e de influência relativamente a novas gerações de

pesquisadores a partir das citações por eles recebidas. Com isso será possível traçar um panorama da pesquisa internacional nessa área temática, relativamente ao seu núcleo de autores e instituições, suas interlocuções, influências teóricas geradas e veículos de socialização do conhecimento, em uma perspectiva diacrônica.

## 2. A abordagem sócio-cognitiva da Organização do Conhecimento

Partindo do pressuposto de que o objeto de estudo da Organização do Conhecimento reside, como destaca Barité (2001), em um conhecimento socializado, o referido autor destaca que esse conhecimento, na condição de produto, necessidade e dinamismo social, produz-se a partir da informação e nela novamente se transforma quando registrado em documentos, por meio de conceitos oriundos de sistemas de conceitos, (com fins funcionais, científicos ou documentários) valendo-se de um sistema aberto que evidencia sua estrutura e seu processo de comunicação. A organização de se conhecimento, por sua vez, busca um melhor aproveitamento individual e social e se faz de “n” formas possíveis, em um processo artificial, provisório e determinista.

Desse modo, tem-se um conhecimento gerado pela sociedade que, uma vez socializado por conta de sua materialização ou registro, constitui-se naquilo que Michael Buckland denomina “informação como coisa” (Buckland, 1991). No entanto, para que esse conhecimento socialmente produzido possa ter um uso social, necessário se torna um processo mediador, de organização, em que se estabelecem “substitutos do conhecimento” (*surrogates of knowledge*) para que os contextos de produção e de uso possam ser colocados em diálogo e para que se propicie a realização de um processo helicoidal em que um conhecimento produzido e socializado é organizado para que possa ser acessado e apropriado socialmente para fins de geração de um novo conhecimento que, por sua vez, será organizado e acessado, em um processo contínuo (Guimarães, 2008).

A questão da configuração desse campo de estudos merece ser abordada no seu âmbito científico mais abalizado – a *International Society for Knowledge Organization* – como um nítido representante da dimensão interna social revelada nos quadrantes metateóricos de Ritzer (1991). Nesse contexto, e como destaca Arboit (2014, p. 236), ao analisar a configuração cognitiva do campo de OC a partir da literatura da ISKO, destaca que:

No caso da ISKO assim como no referencial teórico sobre o processo constitutivo do domínio de OC, ve-

rificou-se que no início os interesses eram majoritariamente permeados por perspectivas técnicas, idealistas e cognitivas do conhecimento, como resultado do acolhimento das ideias defendidas principalmente por Dahlberg e Ranganathan. Depois uma mudança de orientação para enfoques genealógicos, pragmáticos, culturais e éticos, representados principalmente por autores como Hjørland, Frohmann, Olson, Beghtol, Guimarães e García Gutiérrez, mas não se abandonando completamente o idealismo em virtude da defesa do naturalismo e do enfoque ontológico defendido por Gnoli, conforme respaldado no referencial teórico.

Esse *turning point* reflete aquilo que Mai (2013, p. 251) considera como um “movimento que se dá a partir de uma concepção essencialista de universo do conhecimento em direção a uma visão mais pragmática e pluralista dos domínios de conhecimento”.

Em suma, pode-se dizer que, em um primeiro momento do estatuto científico da organização do conhecimento, no entorno da ISKO, tem-se uma perspectiva mais pautada na natureza do ser, da sua existência em um contexto de realidade refletidas em categorias básicas de conhecimento que levam a um conjunto estabelecido de conceitos, tal como evidenciado no *Classification Scheme for Knowledge Organization Literature*, cuja sequência estrutural é representada por uma matriz 3x3 de classes principais chamada de “*Systematifier*” (Dahlberg, 1993). Passada cerca de uma década, ou seja, no limiar do século XXI, emerge uma perspectiva, na área, mais voltada para “a natureza, as etapas e os limites do conhecimento humano, especialmente no que tange às relações estabelecidas entre o sujeito e o objeto do conhecimento” (Guimarães, 2017, p. 90), entendendo esse conhecimento como algo dinâmico, culturalmente delimitado.

Como se pode observar, a historicidade da ISKO, a partir de seus objetos de investigação ao longo do tempo, gravita entre uma concepção inicial, de natureza mais ontológica, notadamente pautada nas ideias de Ranganathan e tendo como maior representante Ingetraut Dahlberg, com a sua teoria do conceito (Dahlberg, 1978a,b; 1993; 1995; 2006; 2008) (1) e uma concepção de natureza mais epistemológica, com uma abordagem sociocognitiva, que incorpora os usos, o usuário e o contexto desse conhecimento, e que tem em Birger Hjørland seu maior expoente (Hjørland e Albrechtsen, 1995; Hjørland, 2002, 2003, 2004, 2008).

A essa preocupação epistemológica se alia, ainda, uma abordagem sócia cognitiva, em que se destacam os trabalhos de Hjørland e de Frohman, no sentido de se contextualizarem os processos, produtos e instrumentos de organização

do conhecimento a partir das comunidades discursivas envolvidas. Nesse sentido, as concepções que por séculos permearam a área, relativamente a sistemas e abordagens universais dão lugar a abordagens locais que possam dialogar entre si, abrindo espaço para a denominada interoperabilidade cultural.

Contemplando aspectos da sociologia da ciência, notadamente no que tange à constituição e à dinâmica, intrínsecos a um dado contexto científico, busca identificar como se comporta e se articula uma dada comunidade discursiva, nela identificando aspectos que a possam caracterizá-la como tal: terminologia, referentes teóricos, colégios invisíveis, escolas de pensamento, produção científica, etc.

Nesse âmbito, autores como Birger Hjørland, da University of Copenhagen, Joseph Tennis, da University of Washington, e Richard Smiraglia, da University of Wisconsin-Milwaukee, entre outros, vêm contribuindo significativamente para a consolidação teórica e metodológica da análise de domínio, em especial no que tange à organização do conhecimento.

Especial menção merece, na literatura especializada da ISKO, o fascículo seminal publicado em 2008 (v.35, n.2/3), sobre a base epistemológica da área, tendo como ponto de partida a questão: "O que é organização do conhecimento?" (Mcillwaine e Mitchell, 2008, p. 79-85; Hjørland, 2008, p. 86-101), em cujo âmbito foram discutidas, entre outras questões, sobre a configuração epistemológica da área (Tennis, 2008, p. 102-112), e as questões investigativas que lhe são mais pungentes na atualidade (López-Huertas, 2008, p. 113-136; Gnoli, 2008, p. 137-149).

A abordagem sócio-cognitiva da organização do conhecimento defendida por Hjørland assenta-se, como destacado por Guimarães (2017) em uma tríade de natureza dialógica e complementar: as perspectivas de análise de domínio, semiótica e cultural.

A análise de domínio, ao permitir identificar as condições pelas quais o conhecimento científico se constrói e se socializa, possibilita que se verifique o que é efetivamente importante ou significativo em um dado campo, de tal modo que aspectos como tendências, padrões, processos, agentes e seus relacionamentos possam ser identificados e analisados (Hjørland e Albrechtsen, 1995; Moya-Anegón e Herrero-Solana, 2001; Hjørland, 2002, 2004; Tennis, 2003; Danuello, 2007; Smiraglia, 2011, 2012). Essa perspectiva, decorrente de comunidades discursivas oriundas da divisão social do trabalho, considera o conhecimento como algo que somente pode ser compreendido a partir de um dado contexto e de uma

dada comunidade. Assim, a partir de uma abordagem mais centrada no contexto (perspectiva sociológica) que no indivíduo (perspectiva cognitivista), a análise de domínio, valendo-se de processo de categorização, tem por objeto grupos que revelam uma ontologia coerente e uma mesma epistemologia (que lhe confere as fronteiras intelectuais), a partir de elementos intelectuais comuns e de um discurso efetivo que ocorrem em uma unidade socialmente estruturada (Smiraglia, 2013) (2).

Hjørland (2002) identifica um conjunto de onze abordagens para a análise de domínio na Ciência da Informação: Produção de obras de referência, Construção de linguagens de indexação, Indexação e recuperação da informação, Estudo de usuários, Estudos bibliométricos, Estudos históricos, Estudos de gêneros/tipologias documentais, Estudos epistemológicos e críticos, Estudos terminológicos, Comunicação científica, Cognição científica, conhecimento especializado e inteligência artificial. A essas abordagens soma-se ainda a proveniência arquivística, como defendido por Guimarães e Tognoli (2015) e a análise do discurso, ressaltada por Smiraglia (2015).

A perspectiva semiótica da organização do conhecimento encontra-se mais diretamente relacionada às questões de recepção e de significação, em que se inserem aspectos como os de interpretação de um documento por ocasião de sua análise, o que envolve não apenas a forma como o indexador compreende esse conteúdo como também a maneira que ele o representará de modo a garantir que seja compreendido e acessado pela comunidade usuária. Mai (2001) destaca esse processo de representação como um conjunto de múltiplas interpretações que evidenciam uma semiose ilimitada na concepção Peirceana, em que se conectam signos ou se tem um signo dando origem a outro.

Thellefsen Thelklkefsen e Sorensen (2013), nesse contexto, destacam a importância de que se reconheça a existência de um processo comunicativo, processo esse constantemente modificado ou reinterpretado em virtude de diferentes situações. Tal processo apresenta dinâmica própria, "o que sugere a criação de interpretantes mutuamente relacionados em um processo contínuo de significação (Almeida, Fujita e Reis, 2013, p. 239).

Objeto específico da presente investigação, e em decorrência direta das duas perspectivas anteriores, tem-se a perspectiva cultural da organização do conhecimento, que considera o contexto em sua historicidade e em suas idiossincrasias como determinante para o modo como se desenvolvem os processos, os instrumentos e os produtos da

organização do conhecimento. Como destacado por Guimarães (2017), essa abordagem pressupõe a impossibilidade de neutralidade, na medida em que a organização do conhecimento constitui “*a slanted field*”, em grande parte modulado pelas dimensões de tempo e de espaço.

Essa abordagem pressupõe que tanto o autor, como o indexador, quanto o usuário são oriundos de comunidades específicas e refletem os valores dela, de tal modo que o processo comunicacional que entre eles se estabelece deve ser mediado a ponto de que se chegue a elementos mínimos comuns que propiciem uma mútua compreensão. A isso se alia, por exemplo, a própria diversidade cultural inerente às múltiplas comunidades usuárias que poderão ter acesso àquela informação, o que relativiza e contextualiza a missão do profissional da informação, sintetizada por Richard Mason (1990) como a de fornecer a informação certa, da fonte certa, no tempo certo, no local certo e ao usuário certo.

O reconhecimento da não neutralidade nos processos, produtos e instrumento desenvolvidos assim como da diversidade cultural que permeia os contextos do autor, do indexador e dos usuários leva a uma dimensão ética da organização do conhecimento, tendo por pressuposto que as atitudes éticas encontram fulcro em valores morais de um dado grupo social e tais valores inspiram nossas ações e servem de referência para avaliar a aceitabilidade ética das mesmas.

Relativamente às questões éticas em organização do conhecimento, já afirmavam Fernandez Molina e Guimarães (2002) uma menor incidência de literatura sobre essa temática face a aspectos como disseminação da informação ou a administração de unidades de informação (como por exemplo, na formação e desenvolvimento de coleções), o que decorria, por sua vez, de uma concepção de organização do conhecimento de viés mais pragmático, mais voltada para a utilização de instrumentos ou a geração de produtos. No entanto, com o impacto das tecnologias de informação e de comunicação, aspectos como aqueles relativos a propriedade intelectual, privacidade e liberdade intelectual forçaram uma reflexão ética mais aprofundada nesse campo de estudos.

Especificamente no que tange aos valores envolvidos, Arot (2000) refere-se a valores de *inter-relacionamento* e de *ordenação* como algo inerente às atividades de organização e representação do conhecimento, indo além das práticas propriamente ditas. Isso reflete o que Leblond (1999) já denominava como compromisso ético do profissional com o *espírito do sistema*, aqui envolvida a instituição, o usuário e os critérios de organização

documental existentes. Pela mesma época, Guimarães (2000) delineava como compromissos éticos fundamentais do profissional nesse campo: com o usuário, com a organização, com a informação, com a profissão e consigo mesmo enquanto cidadão.

Historicamente falando, a abordagem da dimensão ética como um componente cultural da organização do conhecimento remonta aos estudos de Sanford Berman, na década de 70, que analisou a presença de preconceitos e discriminações na estrutura de sistemas de organização do conhecimento (no caso, os *Library of Congress Subject Headings*). Na ocasião, já destacava o referido autor um conjunto de temáticas ligadas a raça, religião, sexo e outras cujos cabeçalhos de assunto evidenciavam sua inaplicabilidade (e, muitas vezes, um confronto) a distintos padrões culturais, especialmente em um momento que a sociedade caminhava a passos largos para uma efetiva pluralidade cultural (Berman, 1971). Como destaca Knowlton (2005, p.124), as críticas de Berman foram além da inadequação dos termos por si sós, mais voltados para um usuário padrão “*American/Western European, Christian, white, heterosexual, and male*” como também no que tange às estruturas sindéticas de subordinação e de relacionamento entre termos.

As críticas de Berman, por sua vez, traziam subjacentes as preocupações com o estabelecimento de uma relação comunicativa efetiva e multidirecional entre a informação, o indexador e o(s) usuário(s), o que levou Antonio García Gutiérrez a contemplar essa dimensão como elemento norteador da construção de um sistema de organização do conhecimento, no caso, relativamente ao patrimônio histórico da Andaluzia (García Gutiérrez, 1998). Tal aspecto, por sua vez, levou o autor a formalizar a ideia de uma “epistemologia interativa” que considerasse a diversidade das culturas envolvidas e promovesse, assim, uma “ética transcultural de mediação” de modo a que os processos, produtos e instrumentos de organização do conhecimento estejam efetivamente comprometidos com a promoção o diálogo intercultural (García Gutiérrez, 2002).

Criticando o caráter dominador das classificações utilizadas atualmente, García-Gutiérrez (2014) propõe como alternativa o que ele denomina como “desclassificação”. Para o autor, os processos classificatórios abordam uma ideologia dominante, revelando estruturas totalitárias e que buscam o objetivismo. A desclassificação poderia ser compreendida então, como uma operação hermenêutica, agregadora e aberta ao pluralismo.

Para tanto, o autor propõe três estratégias que devem ser utilizadas nesse processo: as instâncias devem ser admitidas como poli-hierárquicas, não correspondendo necessariamente a um único correspondente hierárquico; reconhece-se a possibilidade do “não ser”, evitando-se os reducionismos; e ainda, permite que as instâncias sejam vistas além de dicotomias básicas, em uma sobreposição dissolutiva (García-Gutiérrez, 2014) (3).

Essa questão cultural perpassa, necessariamente, a dimensão da linguagem, como destaca Michéle Hudon (1997, 1999), ao referir-se aos tesouros multilíngues, de modo a que se garanta igual tratamento a todas as línguas envolvidas, em especial pelo fato de que as línguas, mais do que um conjunto de palavras e regras sintáticas colocadas lado a lado, refletem os valores culturais de seus falantes.

A questão cultural na organização do conhecimento se manifesta ainda por meio de dois conceitos trabalhados por Clare Beghtol (2002, 2005): a *garantia cultural* dos processos, produtos e instrumentos de organização do conhecimento de modo a que possam refletir os valores das culturas envolvidas, e a *hospitalidade cultural*, que se evidencia pela capacidade de os sistemas de organização do conhecimento estarem abertos para inclusões, exclusões ou alterações de notações em função das mudanças culturais ocorridas no tempo e no espaço.

Os fenômenos do inter e do multiculturalismo são evidenciados por Maria José López-Huertas (2008) como uma realidade social e, como decorrência, uma necessidade de pesquisa no campo da organização do conhecimento, tangenciando aspectos relativos à diversidade sócio-cultural, à tensão entre as abordagens globais e as abordagens locais, e à forma pela qual os sistemas de organização do conhecimento contemplam a realidade das minorias.

No bojo da temática das minorias, tem-se o exemplo do universo da homossexualidade masculina em que se discutem as deletérias consequências de sistemas que fazem “tabula rasa” dessas realidades e ainda das tensões que podem ocorrer a partir dos diferentes significados que um cabeçalho pode ter (muitos dos quais evidenciadores de visões preconceituosas) nesse âmbito (Campbell, 2000; Pinho e Guimarães, 2012; Campbell, Guimarães, Pinho, Martínez-Avila e Nascimento, 2017).

Tais aspectos evidenciam uma dimensão eminentemente ética visto envolverem valores. Assim, a atividade de representação de conteúdos reveste-se de um “poder de nomear”, permeado pelos valores e idiosincrasias do document, do

sistema e do indexador, evocando uma dimensão situacional e política da organização do conhecimento, especialmente no que tange a dar voz à diversidade cultural das comunidades usuárias envolvidas. Nessa perspectiva, fundamental se torna o reconhecimento da diversidade como inerente à organização do conhecimento, do impacto que os fatores humanos e sociais trazem a esse campo, de sua concepção como produto social, por esta sofrido, de sua natureza antes sugestiva que prescritiva, de seu compromisso com a promoção de um diálogo global e com a interoperabilidade de sistemas (Olson, 2002; Guimarães, Pinho e Milani, 2016).

### 3. Comunidades Epistêmicas

As comunidades epistêmicas, como redes de *experts* em um dado tema ou campo de pesquisa, influenciam nos produtores e usuários de informação desse contexto, sendo em grande parte responsáveis pela literatura produzida. Pelo fato de contar com especialistas com efetiva competência científica, as comunidades epistêmicas apresentam autoridade e relevância política no campo a que se referem. Isso pressupõe que seus atores compartilhem, como elemento comum, um conjunto de crenças, práticas e competências profissionais (Haas, 1992). Hakanson (1992), no mesmo sentido, refere-se a um conjunto de conhecimento e práticas em comum que têm a capacidade de afetar como esse conhecimento é compartilhado.

Trazendo essa questão para o universo científico, em especial no que tange à atividade acadêmica e à conseqüente produção científica, tem-se a ciência como uma arte coletiva cuja construção se dá a partir de publicações, eventos, departamentos de universidades, pareceristas, listas de discussões e ainda pela formação de novos coletivos. Sob essa perspectiva, afirmam que, além de produzir novos conhecimentos, constantemente contribuindo para a evolução de áreas científicas, de forma a produzir soluções para problemas, essas comunidades epistêmicas afetam os conhecimentos produzidos, influenciando-os (Meyer e Molyneux-Hodgson, 2010).

Como destacam as referidas autoras, pode-se ainda relacionar o conceito de comunidade epistêmica, como o conceito de culturas epistêmicas de Knorr-Cetina (1999), uma vez que a ênfase vai além da produção de conhecimentos, atingindo os mecanismos culturais que levam a essa produção, os quais revelam a fragmentação das ciências, a arquitetura dos conhecimentos epistêmicos das áreas, a construção de referentes teóricos e, ainda, a diversidade da produção de instrumentos.

Outro conceito que pode ser relacionado às comunidades epistêmicas é o de comunidades de prática, baseado no estudo das práticas que grupos de especialistas possuem, uma vez que as sociedades sempre formaram comunidades que dividiam práticas culturais comuns, com base no conhecimento coletivo, resultando na competência sobre determinado assunto, e contribuindo para a transmissão dessas práticas comuns aos novos integrantes da comunidade (Wenger, 2000).

O conceito de paradigmas da ciência definido por Kuhn (1998) também pode contribuir e se relacionar com o conceito de comunidades epistêmicas uma vez que, segundo o autor, a ciência vive de paradigmas, que se manifesta cada vez que uma teoria é reformulada com uma contribuição. Para que esses paradigmas aconteçam, é necessário que a área seja estudada e que comunidades científicas saibam dessas pesquisas por meio de divulgações científicas, como as publicações. A questão dos paradigmas, exposta por Kuhn, trouxe efetiva contribuição para os estudos da Sociologia da Ciência, pois apresenta uma nova visão a respeito do funcionamento e desenvolvimento da ciência (Marcelo e Hayashi, 2013).

Segundo Barjak (2006), torna-se necessário diferenciar os conceitos de sociologia da ciência e sociologia do conhecimento, uma vez que o primeiro campo parte do pressuposto que a comunicação científica, enquanto difusora de conhecimentos produzidos e de pesquisas realizadas em uma área, leva a socialização desses estudos. E é da consequência dessas ações que se vale a sociologia do conhecimento, ao voltar-se para o estudo dessas comunicações como elemento para o reconhecimento dos autores e daqueles que se valem de seus trabalhos como fonte para futuras pesquisas.

Referindo-se especificamente à Sociologia da Ciência, Hjørland (2015) destaca como objeto os estudos empíricos nas atividades científicas (tanto em relação à própria Ciência quanto em relação à sociedade, aqui englobadas relações de poder, de natureza econômica, política, etc.). Assim, e em estreito diálogo com as Ciências Métricas, encontra-se envolvida com questões sobre os estudos culturais da Ciência, em questões como as de gênero e étnicas, assim como as que envolvem os motivos que levam os cientistas a agirem sistematicamente de uma maneira.

Nesse contexto, avaliar a informação científica produzida passa ser a uma necessidade ética, uma vez que a falta de integridade na pesquisa traz consequências deletérias uma vez que a pesquisa se desenvolve e intervêm em ambientes complexos, permeados por uma diversidade

de valores, pontos de vista, teorias e conceitos, em cujo âmbito se torna necessário considerar o conhecimento em rede e suas formas sociais e epistêmicas. Nessa tônica, destacam-se as citações como importantes indicadores de validação da ciência, pois não basta saber os resultados a que se uma pesquisa chegou, mas também, o que está sendo feito em determinada temática e por quem esse trabalho é realizado. Indo além, tem-se que os dispositivos de representação da ciência devem levar em conta os contextos que normatizam e validam as práticas de produção científica (González de Gómez, 2013, 2015).

Esse processo de qualificação da produção científica resulta em uma estrutura de redes enquanto grupos de relações significativas de modo a demonstrar a interlocução entre assuntos e questionamentos em um dito estado do conhecimento. Ou seja, a relevância de uma determinada investigação se explicita após o surgimento de novas pesquisas e intervenções epistemológicas e empíricas (González de Gómez, 2012).

Outro conceito próximo do conceito de comunidades epistêmicas é o de comunidades discursivas, evidenciadas por meio de redes de autores de um dado campo “[...] consistindo em instrumento essencial para a compreensão do estado da arte em temas específicos, bem como identificar graus de relações interdisciplinares.” (Lucas, 2014, p. 37).

Tem-se, pois que, a abordagem e a análise de comunidades epistêmicas em áreas em processo de desenvolvimento e consolidação, tal como ocorre com a organização do conhecimento, pode ser uma forma de medir seu impacto na sociedade e na academia mais especialmente, pois quanto maior o impacto de relevância de um campo, maior será sua capacidade de gerar conhecimento. Assim, a importância dos estudos de comunidades epistêmicas nesse campo evidencia-se mais especialmente quando Hjørland (2017) a elas se refere como conceito de importante discussão dentro da análise de domínio, fazendo referência aos estudos em comunidades epistêmicas realizados (Guimarães, Martinez-Avila e Alves, 2015; Martinez-Avila, Guimarães e Evangelista, 2017).

#### 4. Análise e discussão dos dados

A partir da coleta realizada deu-se início a um estudo bibliométrico desse corpus.

A Bibliometria (e, por decorrência, a Cientometria) origina-se na Ciência da Informação (CI), sendo inicialmente utilizada como forma de análise da ciência enquanto um processo informativo e em grande parte motivada pela necessidade de

se avaliarem publicações para fins de aquisição em bibliotecas universitárias e especializadas. A isso se aliou a necessidade de a CI valer-se de métodos estatísticos tanto para mensurar ou modelar fenômenos específicos da comunicação científica quanto para melhor compreendê-los. Pode-se dizer que enquanto a Bibliometria tem por objeto basicamente a produção bibliográfica, e mais especificamente a produção científica, a Cientometria centra-se no estudo do comportamento da Ciência, incorporando elementos da Sociologia da Ciência.

Nesse contexto destacam-se os estudos seminais levados a cabo, entre outros, por Eugene Garfield, linguista norte-americano, no Institute for Scientific Information. Especial referência merece o trabalho de Garfield (2006, p. 90) na criação do Science Citation Index, em 1964, precursor da atual Web of Science, e do conceito de Fator Impacto. Garfield colaborou, ainda, na criação do Current Contents, do Journal Citation Reports (JCR), do Index Chemicus e da revista científica *The Scientist* (Garfield, Cronin e Atkins, 2000). (4)

Hoje esse campo de estudos fornece base metodológica para a comunicação científica (que, por sua vez, vai além da comunicação acadêmica), sendo usado para a avaliação da pesquisa e agregando aspectos quantitativos e qualitativos. Alguns desafios se colocam, ligados às características da informação científica da atualidade: crescimento exponencial, superposição (overload), complexidade, acesso competitivo a recursos humanos e financeiros, crise no sistema de avaliação por pares e mudança de uma *little science* para uma *big science*, como já havia previsto Derek de Solla Price há mais de meio século (Price, 1963). A base Scopus tende a abranger maior diversidade de idiomas e melhor cobertura em ciências sociais, enquanto a base de dados Web of Science é mais voltada para as ciências exatas, em trabalhos primordialmente publicados em inglês.

Consolidada como área – ou campo – de estudos, a Bibliometria e a Cientometria revelam uma crescente produção científica que se manifesta, entre outros espaços, nas conferências internacionais da International Society for Scientometrics and Informetrics – ISSI e nas conferências nacionais no âmbito dos Encontros Nacionais de Bibliometria e Cientometria, nos quais, como destaca Leta (2013), temas como produtividade científica, indicadores bibliométricos, webometria, colaboração científica e políticas científicas ocupam lugar de destaque.

Essa área de estudos possibilita, entre outros aspectos: a) identificar o que é mais significativo

(hot) em uma área de conhecimento e as redes que nela se estabelecem; b) realizar estudos de mapeamento e visualização sobre a dinâmica da ciência; c) identificar novas tendências e campos investigativos; e d) subsidiar a avaliação e o financiamento da pesquisa por meio de métodos para medir produtos e performances de pesquisa, fórmulas de financiamento e estudos comparativos.

Para tanto, um conjunto de etapas ou passos de natureza metodológica são previstos nesse campo: a) Quantificação (extração, redução, limpeza e desambiguação de dados); b) Mensuração (definindo-se o que pode e o que deve ser medido e quais as medidas adequadas para o objetivo almejado); c) Benchmarking (contextualização dos dados e realização de comparações entre dimensões comparáveis); e Apresentação e Interpretação (preparação de tabelas e mapas, interpretação das observações e relacionamento de indicadores bibliométricos com outros métodos). Um aspecto métrico especialmente importante nesse campo de estudos reside nos denominados indicadores bibliométricos de produção científica, entendidos como resultados de quantificação de resultados científicos atribuíveis tanto a agentes determinados quanto a contextos acadêmicos como um todo (Maltrás Barba, 2003).

Calon, Courtial e Penan (1995) destacam que os indicadores bibliométricos podem ser classificados em: a) indicadores de atividades (relativos à produção científica e de patentes); b) indicadores relacionais de primeira geração (coautorias, redes de citações e de cocitações); e c) indicadores relacionais de segunda geração (co-ocorrência de termos, estudos de centralidade, estruturação de campos, atores científicos e sua utilização estratégica, e ciclos de vida, seja na ciência, por meio das publicações científicas, seja em tecnologia, nas patentes).

A atuação da Bibliometria / Cientometria pode dar-se nos níveis macro (analisando países, regiões ou campos de pesquisa), meso (analisando instituições, revistas ou sub-campos), micro (analisando pesquisadores ou grupos de pesquisa) ou nano (analisando um dado paper ou um dado pesquisador). Cumpre destacar que cada nível de análise demanda padrões específicos e quanto maior a amplitude ou o nível de agregação desses estudos, mais abrangentes deverão ser os métodos empregados. Por outro lado, os estudos de menor nível de agregação demandam análises bastante verticalizadas e contextualizadas, e de maior espectro qualitativo.

Esse campo de estudos vale-se de métodos que têm por requisitos a validade (serem capazes de medir efetivamente o que se pretende medir), re-

levância (seus resultados devem ser significativos ao contexto), confiabilidade (precisão), replicabilidade (igualmente aplicáveis a situações congêneres) e robustez (representativos de um conjunto considerável de dados). Mais especificamente no que se refere à robustez, recordam-se situações em que um único paper altamente citado altera a posição de uma universidade em um ranking, constituindo um desvio pois deixa de ser um dado representativo.

As fontes de dados da Bibliometria residem nas bibliografias, bases de dados, índices de citações e de abstracts, bem como em bases de dados de texto integral, repositórios de dados e outros.

Esse campo tem levado profissionais oriundos de diversas áreas (Ciência da Informação, Ciência da Computação, Sociologia, Economia, Ciências Biológicas, Estatística, Matemática, etc) a se voltarem para estudos acerca do desenvolvimento da ciência e de seus produtos e veículos, com campos de excelência investigativa em distintas instituições do mundo, como é o caso da Universidade de Leuven (Bélgica), Universidade de Leiden (Holanda), Universidade de Viena (Áustria), Universidade Humboldt de Berlim (Alemanha), Universidade Carlos III de Madrid (Espanha) e Universidade de Granada (Espanha), entre outros. No Brasil, grupos de pesquisa na UNESP, na Universidade Federal do Rio Grande do Sul, na Universidade Federal do Rio de Janeiro e na Universidade de Pernambuco, entre outros, vêm se destacando na pesquisa nessa área, ressaltando-se, ainda, o especial apoio fornecido por agências de fomento à pesquisa, tais como a Fapesp e o CNPq, em estudos dessa natureza.

Esse contexto de pesquisa vem se revelando, inclusive, um promissor universo profissional, como se verifica na atuação do Bibliometrics and Publication Strategies Department da Biblioteca da Universidade de Viena, que, sob a coordenação de Juan Gorraiz, possui, entre outras atribuições, a elaboração do perfil bibliométrico dos pesquisadores daquela instituição, importante subsídio à avaliação e ao delineamento de estratégias e metas (Gumpenberger, Wieland e Gorraiz, 2012).

Em termos de perspectivas e desafios para essa área na atualidade, destaca-se ao fato de, por trabalhar com toda uma ampla variedade de dados (que, por sua vez, necessitam estar devidamente limpos, sem ambiguidades e corretamente expressos), tais como publicações, citações (trabalhos citantes), referências (trabalhos citados), afiliações institucionais e tipos documentais, a Bibliometria hoje se depara com importantes desafios, tais como: a tensão entre a completeza, a representatividade e a relevância dos dados; a

variação de cobertura de dados entre as diferentes bases de dados (p.ex. Web of Science e Scopus); o desequilíbrio na cobertura dos diferentes campos de pesquisa, com privilegiamento de alguns (p.ex. C. Biológicas) em detrimento de outros (p.ex. C. Humanas e Sociais); e os problemas relativos a autoria, afiliação institucional e fonte, em muitos casos devidos a questões ligadas a homonímia, fusões, reorganizações e variações de uso em denominações institucionais.

A isso se alia, ainda, o fato de as grandes bases dedicaram-se predominantemente à produção científica em inglês, o que traz problemas a campos como as Ciências Humanas e Sociais em que as línguas mais importantes de comunicação científica podem ser o italiano em Direito, o alemão em Filosofia, o francês em História, etc.

Nos dias de hoje, novas perspectivas se abrem para esse campo, tais como: mais ampla aplicação a áreas distintas das Ciências Básicas, possibilidade de estabelecimento de relação entre literatura científica e patente como indicador de produção e de transferência de conhecimento, e análise da visibilidade científica a partir da web e as estatísticas de uso como um componente importante dessa visibilidade (downloads, acessos, visualizações, visitas, etc), como demonstram os estudos de Altmetria, baseados em mídias sociais (Mendeley, CiteUlike, etc) e redes sociais (Researchgate, Scholarmetrics, Academia, Facebook, Twitter, etc).

Cumprir destacar que os estudos bibliométricos constituem, tradicionalmente, uma das abordagens mais utilizadas de análise de domínio, visto fornecer ferramentas para que melhor se compreenda o comportamento da ciência em um dado, por meio de sua produção científica.

Considerando que essa análise bibliométrica do corpus ainda se encontra em andamento, apresentam-se os resultados até então obtidos.

Para recuperação do corpus, utilizou-se as bases de dados Scopus, Web of Science e Library and Information Science Abstracts – LISA, por meio da busca pelos termos “Knowledge organization” and “Cultur\*”; “Knowledge organization” and “Semiotic\*”; “Knowledge organization” and “Society”; “Knowledge organization” and “Ethic\*”; “Knowledge organization” and “Sociocognitiv\*”.

As buscas foram realizadas no período de setembro a novembro de 2017, estabelecendo como intervalo temporal os artigos publicados a partir do ano de 2001, compreendendo assim as publicações do atual século sobre o tema.

No que tange aos autores dos trabalhos, verificou-se um conjunto de 242 autores. A esse grupo foi aplicada a Lei de elitismo de Price (Price,

1963) segundo a qual a raiz quadrada de um dado conjunto de autores constitui a elite daquele campo em termos de produtividade. Aplicando-se, pois, essa lei, chegou-se a cerca de 15 autores, número que foi elevado a um grupo de 21 autores que publicaram ao menos 3 artigos cada. É interessante observar, ainda, que 207 autores, correspondendo a cerca de 86% dessa produção, publicaram apenas um artigo cada, o que revela, por um lado, que essa temática ainda se encontra bastante dispersa na literatura da área, mas, por outro, que essa elite se confirma. Esse conjunto de autores pulcou 178 artigos o que perfaz uma média de 1,2 autores por artigo, caracterizando uma tradição ainda bastante voltada para a autoria individual.

Autor	País	Instituição	Artigos
Guimarães, J.A.C.	Brasil	UNESP	12
Neelameghan, A.	Índia	Indiana Statistical Institute/ Sarada Ranganathan Endowment for Library Science	08
Martínez-Avila, D.	Brasil	UNESP/UWM/UC3M	06
Narayana, G.	Índia	Aeronautical Development Establishment	05
Beghtol, C.	Canadá	U. de Toronto	04
Garcia Gutiérrez, A.	Espanha	U. de Sevilla	04
Sales, R.	Brasil	UFF	04
Smiraglia, R. P.	EUA	UWM	04
Almeida, C.C.	Brasil	UNESP	04
Pinho, F.A.	Brasil	UFPE	03
Williamson, N.J.	Canadá	U. Toronto	03
Andersen, J.	Dinamarca	U. of Copenhagen	03
Thellefsen, M.	Dinamarca	U. of Copenhagen	03
Ohly, H.P.	Alemanha	Social Science Information Center	03
Farias, M.C.Q. S.	Brasil	UNESP	03
Milani, S.O.	Brasil	UFF	03
Murguia, E. I.	Brasil	UFF	03
Raghavan, K.S.	Índia	Indian Statistical Institute	03
Monteiro, S.D.	Brasil	UEL	03
Chen, H.-H.	China	U. Nacional de Taiwan	03
López-Huertas, M.	Espanha	U. de Granada	03
		Total	87

Tabela I. Autores mais produtivos

Observa-se, nesse grupo, um conjunto de 21 autores provenientes de 8 países distintos: Alemanha, Brasil, Canadá, China Dinamarca, Espanha, Estados Unidos e Índia. Nesse contexto, o Brasil

ocupa posição de destaque, respondendo por 41 dos 87 trabalhos recuperados. Isso se explica pelo fato de a pesquisa em organização do conhecimento, no Brasil, haver fortemente abraçado a questão cultural, muito em função da diversidade existente no próprio país. Exemplo disso reside no fato de os quatro congressos da seção brasileira da ISKO, realizados bianualmente a partir de 2011, sempre contarem com essa abordagem como uma seção temática especial.

A produção da área, por sua vez, encontra-se disseminada em um conjunto de 75 revistas, predominantemente em língua inglesa, como é o padrão tradicional da área de Ciência da Informação. Aplicando-se novamente a lei de elitismo de Price, chega-se ao valor inteiro aproximado de 8 revistas, correspondendo àquelas que publicaram ao menos três artigos cada. Esse grupo responde por 101 dos 178 artigos publicados, ou seja, cerca de 57% dessa produção. Observe-se que, do corpus de 185 publicações teve-se um total de 178 artigos e 7 publicações de outra natureza, mais especialmente capítulos de livros.

Revista	Artigos
Knowledge Organization	38
Advances in Knowledge Organization <sup>1</sup>	26
Scire	16
Ciência da Informação	06
Information Studies	05
Cataloging and Classification Quarterly	04
Journal of Documentation	03
Lecture Notes in Computer Science	03
Total	101

Tabela II. Revistas mais produtivas

É interessante observar, nesse contexto, o papel preponderante da International Society for Knowledge Organization – ISKO, cujas publicações Knowledge Organization e Advances in Knowledge Organization respondem, juntas, com 64 artigos, correspondendo a 36% do total. Destaca-se, ainda, a revista espanhola Scire, igualmente voltada para questões de organização do conhecimento, com 16 publicações. Os periódicos respondem juntos, por 79,2% dessa produção, o que faz sentido, pois constituem tradicionalmente os três veículos científicos de maior representatividade nas temáticas de organização do conhecimento em âmbito internacional.

De uma análise temática geral e ainda preliminar da literatura que compõe o corpus, pôde-se ob-

servar dois temas fortemente prevalentes: a Cultura propriamente dita (envolvendo aspectos como mudanças, configurações, impactos, variações, interculturalidade, etc) e a Ética, cada qual com 30 incidências. Em seguida, observam-se temas ligados às tensões entre as abordagens globais e locais – Glocal KO (15 incidências), à Semiótica (18 incidências) e ao entorno digital (9 incidências) que, juntos, respondem por cerca de 50% do corpus analisado.

## 5. Conclusões

Os estudos relacionados à dimensão cultural demonstram cada vez mais sua evidência e importância no universo da Organização do Conhecimento, revelando por meio das publicações a constante evolução dessa temática. As comunidades epistêmicas são uma importante forma de visualização da ciência, permitindo analisar quais os maiores produtores de conhecimento em determinado tema e como estes influenciam a produção de novos conhecimentos. A partir da identificação da comunidade epistêmica da dimensão cultural em organização do conhecimento, é possível discernir os principais teóricos da área, além das revistas que mais se preocupam em publicar sobre.

Por meio dos resultados, pudemos identificar uma forte liderança do Brasil em estudos nesses aspectos, justificada possivelmente pelo fato da diversidade cultural existente no país. Essa posição de destaque revela que o país pode ser considerado um importante centro de pesquisas nesse sentido, sendo inclusive indicado para colaborações internacionais. Como segundo país mais produtivo tem-se a Índia, o que também pode ser justificado pela forte pluralidade da cultura local.

Como revistas mais produtivas, os periódicos *Knowledge Organization* e *Scire* e a série *Advances in Knowledge Organization*, mostraram sua relevância, correspondendo a quase 80% de toda a publicação nesse tema. Essa proeminência demonstra que essas fontes certamente devem ser as primeiras a serem buscadas por investigadores que pretendem se aproximarem de temas nesse sentido.

O presente trabalho pretendeu explicar sobre os conceitos de comunidades epistêmicas e de como eles podem auxiliar na análise da ciência, principalmente em temáticas emergentes, como os aspectos culturais de organização do conhecimento. Dessa forma, buscou-se deixar como principal contribuição a indicação de um grupo de teóricos e fontes informacionais em evidência nessa temática, de maneira a auxiliar o desenvolvimento de futuras pesquisas na área.

## Notas

- (1) Veja-se, para tanto, a definição de conhecimento em OC proposta por Dahlberg (1995, p.11): "certeza subjetiva ou objetivamente conclusiva da existência de um fato ou do estado de um caso, não sendo transferível e somente podendo ser adquirido por meio da reflexão".
- (2) Como destaca o autor (Smiraglia, 2012), tem-se como elementos caracterizadores de um domínio: base ontológica, finalidade subjacente, hipóteses comuns, consenso epistemológico sobre abordagens metodológicas e semântica social. Em outros trabalhos posteriores, o autor (García Gutiérrez, 2002, 2003, 2004, 2014) destaca como os sistemas de organização do conhecimento, como instrumentos de representação servem, no mais das vezes, a uma realidade *mainstream*, ignorando ou mesmo abafando valores oriundos de outras realidades culturais.
- (3) Como destaca Glänzel (2014), esses dois campos de conhecimento em muito se assemelham e se sobrepõem no que tange a seus objetos e a suas formas de análise, o que lhes confere fronteiras tênues, aparecendo na literatura, muitas vezes, como sinônimos. No entanto, pode-se dizer que o objeto da Cientometria é um tanto mais amplo que o da Bibliometria, por preocupar-se com o comportamento da ciência, indo além da produção científica em si.
- (4) Mugnaini (2013) traça um importante panorama histórico da Bibliometria e da Cientometria em âmbito nacional, e seus consequentes impactos no Brasil.

## Agradecimentos

A autora Isadora Victorino Evangelista agradece o financiamento por meio do processo nº 2017/02327-8, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Os autores José Augusto Chaves Guimarães, Gabriele Luz e Henrique Osawa agradecem o financiamento por meio do processo nº 2017/00584-3, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

## Referências

- Almeida, C.C.; Fujita, M.S.L.; Reis, D.M. (2013). Peircean semiotics and subject indexing: contributions of speculative grammar and pure logic. // *Knowledge Organization*. 40:4, 225-241.
- Arboit, A. E. (2014). O processo de institucionalização socio-cognitiva do domínio de Organização do Conhecimento a partir dos trabalhos científicos dos congressos da ISKO. (Doutorado em Ciência da Informação). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.
- Arot, D. (2000). Les valeurs professionnels du bibliothécaire. // *Bulletin des Bibliothèques de France*. 45, 33-41.
- Barité, M. (2001). Organización del conocimiento: un nuevo marco teórico conceptual em Bibliotecología y Documentación. // Carrara, K. (org.) Educação, universidade e pesquisa. Marília: Unesp Publicações; São Paulo: FAPESP, 35-60.
- Barjak, F. (2006). The role of the internet in informal scholarly communication. // *Journal of the American Society for Information Science and Technology*. 57:10, 1350-1367.
- Beghtol, C. (2002). A proposed ethical warrant for global knowledge representation and organization systems. // *Journal of Documentation*. 58:5, 507-532.
- Beghtol, C. (2005). Ethical decision-making for knowledge representation and organization systems for global use. // *Journal of the American Society for Information Science and Technology*. 56:9, 903-912.

- Berman, S. (1971). *Prejudices and antipathies: a tract of Library of Congress Subjects Headings concerning people*. Metuchen: Scarecrow Press, 1971.
- Buckland, M.K. (1991). Information as thing // *Journal of the American Society for Information Science*. 42:5 (June 1991) 351-360.
- Dahlberg, I. (1978a). Fundamentos teóricos-conceituais da classificação // *Revista de Biblioteconomia de Brasília*, 6:1 (Jun. 1978) 9-21.
- Dahlberg, I. (1978b). Teoria do conceito // *Ciência da Informação*, 7:2 (1978b) 101-107.
- Dahlberg, I. (1993). Knowledge organization: its scope and possibilities. // *Knowledge Organization*, 20:4 (1993) 211-222.
- Dahlberg, I. (1995). Current trends in knowledge organization. // Garcia Marco, F. J. (ed.). *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: Librería General, 1995, 7-26.
- Dahlberg, I. (2008). Interview with Ingetraut Dahlberg // *Knowledge Organization*, 35:2/3 (2008) 82-85.
- Dahlberg, I. (2006). Knowledge organization: a new science? // *Knowledge Organization*, 33:1 (2006) 11-19.
- Danuello, J. C. (2007). *Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise de domínio*, Marília: UNESP, 2007.
- Fernandez-Molina, J. C.; Guimaraes, J. A. C. (2002). Ethical aspects of knowledge organization and representation in the digital environment: their articulation in professional codes of ethics // Lopez-Huertas, M. J. (Org.). *Challenges in knowledge representations and organization for the 21st century: integration of knowledge accross boundaries*; Würzburg: Ergon Verlag. 487-492.
- García Gutiérrez, A. (1998). *Principios de lenguaje epistemográfico: la representación del conocimiento sobre Patrimonio Histórico Andaluz*. Sevilla: Instituto Andaluz del Patrimonio Histórico, 1998.
- García Gutiérrez, A. (2002). Knowledge organization from a "culture of the border": towards a transcultural ethics of mediation. // López-Huertas, M. J. (Ed.) *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Würzburg: Ergon-Verlag, 516-522.
- García Gutiérrez, A. (2003) *Proyectar la memoria: del ordo nacional a la reapropiación crítica* // *Transinformação*. Campinas, 15(1), p. 7-13, (2003).
- García Gutiérrez, A. (2004). *Otra memoria es posible: estrategias descolonizadoras del archivo mundial*. Sevilla: Universidad de Sevilla 2004.
- García Gutiérrez, A. (2014). Declassifying Knowledge Organization. // *Knowl. Org.* 41:5, 393-409.
- Garfield, E.; Cronin, B.; Atkins, H. B. (2000). *The web of knowledge: a festschrift in honor of Eugene Garfield*. Medford, N.J.; Information Today, 2000.
- Gnoli, C. (2008). Tem long-term research questions in Knowledge Organization // *Knowledge Organization*. 35:2/3 (2008) 137-149.
- González de Gómez, M. N. (2013). Sobre normas e algoritmos: os critérios de validade da ciência e da informação // *Encontro nacional de pesquisa em ciência da informação – ENANCIB*, 14, 2013, Florianópolis. Anais eletrônicos... Florianópolis, 2013.
- González de Gómez, M. N. (2015). Validade científica: da epistemologia à política e à ética. // *Liinc em Revista*, Rio de Janeiro. 11:2 (2015) 339-359.
- Guimarães, J.A.C. (2000). O profissional da informação sob o prisma de sua formação // Valentim, M.L. P. (Ed.) *Profissionais da Informação: formação, perfil e atuação profissional*, São Paulo: Polis. 53-70.
- Guimarães, J. A. C. (2017). Slanted knowledge organization as a new ethical perspective // Andersen, Jack; Skouvig, Laura. (Org.) *The organization of knowledge: caught between global structures and local meaning*, Bingley: Emerald Publishing Limited, 12:1, 87-102.
- Guimarães, J. A. C.; Tognoli, N. B. (2015). Provenance as a domain analysis approach in archival knowledge organization. // *Knowledge organization*. 42:8 (2015) 562-569.
- Guimarães, J. A. C.; Pinho, F. A.; Milani, S. de O. (2016). Theoretical dialogs about ethical issues in knowledge organization: García Gutiérrez, Hudon, Beghtol, and Olson // *Knowledge Organization*. 43:1 (2016) 338-350.
- Guimarães, J. A. C. (2008). A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). // *Revista Ibero-Americana de Ciência da Informação*, Brasília. 1:1 (2008) 77-99.
- Gumpenberger, C.; Wieland, M.; Gorraiz, J. (2012). Bibliometric practices and activities at the University of Vienna. // *Library Management*. 33:3 (2012) 174-183.
- Haas, P. M. (1992). Epistemic communities and international policy coordination. // *International Organization*, Cambridge. 46:1 (1992) 01-35.
- Hjørland, B. (2004). Domain analysis: a socio-cognitive orientation for Information Science research. // *Bulletin of the American Society for Information Science and Technology*. 3:3.
- Hjørland, B. (2003). Arguments for epistemology in information science. // *Journal of the American Society for Information Science and Technology*. 54:8 (2003) 805-806.
- Hjørland, B. (2003). Deliberate bias in knowledge organization? // Arsenault, C.; Tennis, J.T. (Ed.) *Culture and identity in knowledge organization*, 10th, 2008, Canada, Proceedings of the Tenth International ISKO Conference, Würzburg: Ergon, 256-261.
- Hjørland, B. (2002). Domain analysis in information science: eleven approaches-traditional as well as innovative. // *Journal of Documentation*. 58:4 (2002) 422-462.
- Hjørland, B. (2008). What is Knowledge Organization // *Knowledge Organization*. 35:2/3 (2008) 86-101.
- Hjørland, B.; Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis // *Journal of the American Society for Information Science*. 46:6 (1995) 400-425.
- Hjørland, B. (2017) Domain analysis // *ISKO Encyclopedia of Knowledge Organization*. (2017). [http://www.isko.org/cyclo/domain\\_analysis#ref](http://www.isko.org/cyclo/domain_analysis#ref) (26 maio 2017).
- Hjørland, B. (2015). Informetrics needs a foundation in the theory of science. // Sugimoto, C. (ed.) *Theories of Informetrics and Scholarly Communication*, Berlin: Walter de Gruyter, 2015.
- Hudon, M. (1997). Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts // *Knowledge Organization*. Würzburg, 24:2 (1997) 84-91.
- Knorr-Cetina, K (1999). *Epistemic cultures: how the sciences make knowledge*. Cambridge: Harvard University Press, 1999.
- Knowlton, S. A. (2005). Three decades since *Prejudices and antipathies*: a study of changes in the LCSH. // *Cataloging & Classification Quarterly*. 4:2 (2005) 123-145.
- Kuhn, T. (1998). *A estrutura das revoluções científicas*. São Paulo: Perspectiva, 1998.

- Leblond, F. (1999). *Étique et légitimité du professeur documentaliste*. Paris: Hachette Éducation, 1999.
- Leta, J. (org) (2013). *Bibliometria e cientometria: reflexões teóricas e interfaces*. São Carlos: Pedro & João Editores, 2013.
- López-Huertas, M. J. (2008). Some current research questions in the field of Knowledge Organization. // *Knowledge Organization*. 35:2/3 (2008) 113-136.
- López-Huertas, M. J. (2008). Cultural impact on Knowledge Representation and Organization in a Subject Domain. // Arsenault, C.; Tennis, J. T. (Ed.). *Cultural and Identity in Knowledge Organization*. Würzburg: ERGON Verlag, (2008) 340-346.
- Lucas, E. R. O. (2014). *Capital social e capital científico na produção científica sobre linguagens documentárias e sistemas de organização do conhecimento no campo da knowledge organization (KO) nos idiomas espanhol, francês e português*. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo. São Paulo 2014.
- Mai, J.-E. (2001). Semiotics and indexing: an analysis of the subject indexing process. // *Journal of Documentation* 57:5 (2001) 591–622.
- Mai, J.-E. (2013). Ethics, values and morality in contemporary Library Classifications. // *Knowledge Organization* 40:4 (2013) 242-253.
- Marcelo, J. F.; Hayashi, M. C. P. I. (2013) Estudo bibliométrico sobre a produção científica no campo da sociologia da ciência. // *Informação & Informação*, Londrina 18:3 (2013) 138–153.
- Martínez-Ávila, D.; Alves, B. H.; Guimarães, J.A.C. (2015) Epistemic communities in knowledge organization: an analysis of research trends in the Knowledge Organization journal. // *ISKO*, London. Promoting the theory and practice of organizing knowledge and information. London: (2015) 1-22.
- Martínez-Ávila, D.; Guimarães, J. A. C.; Evangelista, I. V. [no prelo]. Epistemic communities in Knowledge Organization: An analysis of the NASKO meetings proceedings. // *Visualizing Knowledge Organization: Bringing Focus to Abstract Realities - NASKO 2017*, University of Illinois at Urbana-Champaign, jun. 15–16.
- Mcillwaine; I. Mitchell, J. What is Knowledge Organization? // *Knowledge Organization*. 35:2/3 (2008) 79-81.
- Meyer, M.; Molyneux-Hodgson, S. The dynamics of epistemic communities. // *Sociological Research Online*. 15:2.
- Moya-Anegón, F.; Herrero-Solana, V. (2001). Análisis de dominio de la revista mexicana de investigación bibliotecológica. // *Información, cultura y sociedad*. 5:1 (2001) 10-28
- Mugnaini, R. (2013). 40 anos de bibliometria no Brasil: da bibliografia estatística à avaliação da produção científica nacional. // Hayashi, Maria Cristina Piumbato Innocentini; Leta, Jacqueline. (Org.). *Bibliometria e cientometria: reflexões teóricas e interfaces*. São Carlos: Pedro e João Editores. 37-58.
- Olson, H. (2002). *The power to name: locating the limits or subject representation in libraries*. Dordrecht: Kluwer Academic Publisher, 2002.
- Smiraglia, R. (2015). *Domain analysis for knowledge organization: tools for ontology extraction*. Oxford: Chandos, 2015.
- Smiraglia, R. P. (2011). Domain coherence within Knowledge Organization: people, interacting theoretically, across geopolitical and cultural boundaries. // McKenzie, P.; Johnson, K.; Stevens, S. (ed.). *Exploring interactions of people, places and information*, Proceedings of the 39th Annual CAIS/ACSI Conference, University of New Brunswick, Canada, Jun. 2011. <http://www.caais-acsi.ca/conferences.htm>.
- Smiraglia, R. P. (2012). Epistemology of Domain Analysis. In: Lee, H-L.; Smiraglia, R. P. (ed.). // *Cultural frames of knowledge*. Würzburg: Ergon (2012) 111-124.
- Solla Price, D. (1963). *Little science, big science*. New York: Columbia University Press, 1963.
- Tennis, J. T. (2008) Epistemology, Theory, and Methodology in Knowledge Organization: toward a classification, meta-theory, and research framework. // *Knowledge Organization*. 35:2/3 (2008) 102-112.
- Tennis, Joseph T. (2003). Two Axes of Domains for Domain Analysis. // *Knowledge Organization*. 30: 3/4 (2003) 191-195.
- Thellefsen, M.; Thellefsen, T.; Sørensen, B. (2013). A pragmatic semeiotic perspective on the concept of information need and its relevance for knowledge organization. // *Knowledge Organization*. 40:4 (2013) 213-224.
- Wenger, F. (2000). Communities of practice and social learning systems. // *Organization Stanford*. 7:2 (2000) 225-246.

---

Enviado: 2019-04-03. Segunda versão: 2019-06-16.  
 Aceptado: 2019-06-17.

---

---

# Presente y futuro de las revistas científicas

*Present and future of scientific journals*

---

Nicoleta-Roxana DINU (1), Tomàs BAIGET (2)

(1) Biblioteca Nacional de Rumanía, Bucarest, Rumanía, Bulevardul Unirii, 22. București 030833, Romania; roxana.dinu@bibnat.ro.  
(2) Ediciones Profesionales de la Información SL, Barcelona, España; baiget@gmail.com

## Resumen

Evolución de las revistas científicas en los últimos 40 años y análisis de su situación actual considerando los 4 factores que están condicionando su evolución y futuro: 1) movimiento hacia el acceso abierto, que avanza inexorablemente, 2) gran competencia con otros canales de información (repositorios, blogs, redes sociales), así como entre las propias revistas, y en especial las megarevistas, 3) piratería (revistas depredadoras y Sci-Hub), y 4) calidad. Dependiendo de lo "imprescindibles" que sean algunos títulos, la gran mayoría de revistas que se financian mediante suscripciones deberán pasar a open access y financiarse cobrando a los autores. Las revistas que no han conseguido estar indexadas por las dos bases de datos que las agencias de acreditación toman como referencia y ranking de calidad (Web of Science y Scopus) no serán suficientemente atractivas para los autores, por lo que tienen un futuro incierto, a menos que estén subvencionadas por una universidad o institución. En este caso representarán un papel secundario —aunque no menos importante— de divulgación de la ciencia y de pedagogía, pero probablemente no publicarán investigaciones originales.

**Palabras clave:** Revistas científicas. Revistas académicas. Evolución. Perspectiva. Futuro. Acceso abierto. Revistas piratas. Calidad. Mega-revistas. Suscripciones.

## 1. Introducción

A lo largo de los últimos 30 años, una serie de factores más o menos interrelacionados entre sí como Internet, globalización, creciente predominio del idioma inglés, mayor transparencia, crisis económicas, sistemas de acreditación de los investigadores, programas informáticos..., han dado un vuelco a la publicación científica. A pesar de que todos los cambios producen algún malestar al tener que adaptar las rutinas de trabajo (Harzing y Adler, 2015; Schimmer y Geschuhn; Vogler, 2015; Segado-Boj *et al.*, 2018), podríamos decir que en conjunto la situación actual ha ido mejorando en favor de la ciencia: tenemos abundancia de información (Anglada, 2019; Dinu, 2018), mejor acceso a la misma, se trabaja con más ética (Baiget, 2010) y más transparencia (Wicherts, 2016).

## Abstract

Evolution of scientific journals in the last 40 years and analysis of their current situation considering the 4 factors that are conditioning their evolution and future: 1) movement towards open access, that advances inexorably, 2) great competition with other information channels (repositories, blogs, social networks), as well as between the journals themselves, and especially the mega-journals, 3) piracy (predatory journals and Sci-Hub), and 4) quality. Depending on how "essential" certain titles are, the vast majority of journals that currently are financed by subscriptions will need to be open access and financed by charging the authors. The journals that have not been indexed by the two databases that the accreditation agencies take as reference and quality rankings (*Web of Science* and *Scopus*) will not be sufficiently attractive for the authors, so they have an uncertain future, unless they are subsidized by a university or other type of institution. In this case, they will play a secondary (though not less important) role in the dissemination of science and pedagogy, but they will probably not publish original research.

**Keywords:** Journals. Evolution. Future. Open access. OA. Predatory journals. Piracy. Quality. Mega-journals. Subscriptions.

De todos los tipos de publicaciones que sirven para difundir la ciencia: congresos, libros, patentes, revistas..., son éstas las que han experimentado mayor transformación, habiendo perdido al mismo tiempo una parte de su protagonismo.

En este artículo se revisa la situación actual y la evolución de las revistas científicas, analizando las tendencias que las están afectando y que pueden configurar su futuro. El objetivo es que viendo la trayectoria seguida a lo largo de los últimos 40 años se entienda mejor la compleja situación actual, y se pueda imaginar cómo será en los próximos años.

Como metodología se ha hecho una revisión bibliográfica, y se han utilizado datos y estadísticas ya existentes, excepto los datos de suscriptores de una revista que han sido suministrados por su editor para este artículo.

## 2. Acceso abierto, un movimiento imparable

### 2.1. Antecedentes

El movimiento de acceso abierto, que propugna que los artículos resultado de la investigación científica —en especial la financiada con dinero público— estén en la Web accesibles para todos, tiene sus raíces en la llamada *espiral de las revistas* que se desarrolló entre los años 70 y 90: En esos años las universidades fueron centralizando las bibliotecas, unificando las pequeñas bibliotecas de los departamentos en una biblioteca de facultad. La integración continuó unificando las adquisiciones de libros y las suscripciones de revistas, que pasaron poco a poco a realizarse desde la biblioteca central de la universidad. Así, se cancelaron las suscripciones independientes de las facultades y sólo quedó una única suscripción compartida por toda la universidad.

En los años 80 las bibliotecas universitarias se fueron informatizando. Como al principio la informática era cara se crearon cooperativas que compartieron unos servicios centrales, un centro de cálculo. Poco a poco se implementaron catálogos colectivos de la cooperativa, lo que representó un paso más hacia la cancelación de suscripciones de algunas revistas, entonces compartidas ya por varias universidades.

Empezaron a funcionar de forma eficiente los servicios de obtención de documentos (SOD), primero a través de correo postal, luego por fax y finalmente por correo electrónico.

Las caras fotocopias de los 70 se abarataron drásticamente en los 80, y en los 90 aparecieron los primeros escaners a precios asequibles a cualquier entidad, lo que permitió digitalizar los artículos y enviarlos por mail. Así se acabó de implantar la filosofía *just in time* (pedir copias de los artículos en el momento que alguien los necesita, puesto que los avances técnicos permitían obtener la copia del artículo en pocos días u horas) en contraposición a la filosofía *just in case* (estar suscritos a la revista por si algún día alguien necesita un artículo). Los investigadores iban teniendo cada vez mejor acceso a la información bibliográfica, tanto por el acceso a las revistas de abstracts, como sobre todo a partir de los 80 a las bases de datos online, que ofrecían información de muchas más revistas que las suscritas por su biblioteca.

Las propias editoriales de revistas se echaron tierra encima al ofrecer el servicio *pay per view* o de pago por leer: no era necesario estar suscrito a la revista, se podían comprar artículos sueltos.

Esta opción era más barata si anualmente se necesitaban pocos artículos de determinado título.

La cooperación bibliotecaria se amplió con la creación de consorcios de bibliotecas universitarias, como en España, en 1996, el *Consorti de Biblioteques Universitàries de Catalunya (CBUC)*, o en 2000 el *Consortio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria (Madroño)*; etc. (Térmens-Graells, 2005; Anglada, 2015).

También se crearon redes exclusivamente para el intercambio de fotocopias de artículos como la *Coordinadora de Documentació Biomèdica (CDB)*, que usaba un catálogo colectivo propio, antecedente del *C17 (Catálogo de Publicaciones Periódicas en Bibliotecas de Ciencias de la Salud Españolas)* (Amat y Serra, 1991; San-José-Montano *et al.*, 2003). En años posteriores se instaló en casi todas las bibliotecas de las universidades españolas el programa *GTBib-SOD*, de la empresa española *Kronosdoc*, que facilitó aún más el intercambio de fotocopias (Rodríguez-Gairín, 2012, 2015).

Entre los antecedentes (y realidades actuales) del movimiento OA también hay que tener en cuenta los repositorios, tanto temáticos como institucionales, que volveremos a citar en el punto 3.1. El primer repositorio fue *ArXiv*, puesto en marcha en 1991 por Paul Ginsparg, en *Los Alamos Laboratory*, que pasó en 2001 a la *Cornell University*.

Finalmente, un hecho colateral digno de señalar es el proyecto *Tulip (The University Licencing Program)* que se llevó a cabo en los años 1993-1995 entre *Elsevier* y 8 universidades norteamericanas y que marcó el principio del paso de la “posesión” de las revistas al “alquiler” bajo una licencia (Keefer, 1996).

Todos los hechos descritos hasta aquí fueron causa de cancelaciones de suscripciones de revistas, a lo cual las editoriales reaccionaron subiendo los precios mucho más que el índice de precios al consumo: hemos descrito la *espiral o círculo vicioso de las revistas*.

### 2.2. Movimiento acceso abierto (OA)

Propiamente, el movimiento acceso abierto empezó con las declaraciones de Budapest (2002), Berlín (2003) y Bethesda (2003) (Melero, 2005) en las que diversos investigadores, como reacción al gran encarecimiento que habían experimentado las suscripciones de las revistas, y habiendo entrado ya plenamente en la época de la información electrónica, hicieron los llamamientos liderados por Peter D. Suber (*Berkman Klein Center for Internet & Society*, director del *Harvard Open Access Project*) y Stevan Harnad (profesor

de la *Université du Québec à Montréal*, y de la *University of Southampton*, UK).

En estos 17 años ha habido muchas llamadas de la *Comisión Europea (CE)* y otras instituciones para avanzar hacia el OA. Algunos hitos han sido:

2012, julio. Primera recomendación de la CE.

2014. *Charity Open Access Fund (COAF)*, integrado por *Wellcome Trust* y otras 5 instituciones, dan dinero a universidades del Reino Unido para pagar las tasas por publicar (*article processing charges* o APCs) en revistas de acceso abierto.

2015. La CE publica su proyecto *European Open Science Cloud (EOSC)*.

2016. La CE publica el informe *Realising the European Open Science Cloud*.

2018, 4 sept. La organización *cOAlition S*, parte de *Science Europe* (fundada en 2011 e integrada por instituciones financiadoras de investigación de 27 países, con el CSIC por parte de España), dio a conocer su llamado *Plan S (2018)*: "Making open access a reality by 2020". Según este *Plan*, desde el 1 de enero de 2020 los artículos resultado de investigaciones financiadas por las instituciones miembro de *Science Europe*: a) no deben sufrir ningún embargo después de su publicación; b) no deben publicarse en revistas híbridas (revistas con acceso mediante suscripción pero que aceptan la liberación de artículos en OA pagando una APC); c) los autores deben retener el copyright con el fin de poder difundir su artículo sin limitación; y d) las APCs deben tener un límite superior (del orden de 1.000 €)

2018, 5 nov. Wellcome Trust y Bill and Melinda Gates Foundation manifiestan su apoyo al Plan S.

2018, 21 nov. La CE publica el informe *Prompting an EOSC in practice*.

2019, 31 mayo. Revisión del *Plan S*. Con el fin de proporcionar más tiempo para que los investigadores y editores se adapten a los cambios, el plazo fue extendido por un año hasta 2021. Entre otros detalles, los acuerdos para transformar revistas de suscripción a OA serán apoyados hasta 2024.

Una importante contribución al acceso abierto procede de los mandatos de las universidades (Meleró et al., 2018) y de las agencias de financiación para que los autores depositen en repositorios sus artículos inmediatamente después de ser publicados, o pasados unos pocos meses de embargo o *time wall*, período que depende de las condiciones impuestas por las editoriales de las revistas.

Diversos estudios han ido evaluando el porcentaje de artículos científicos disponibles en OA. La cifra actual, principios de 2019, teniendo en cuenta tanto los publicados directamente en OA (la llamada *vía dorada* al acceso abierto) como los depositados en repositorios (la *vía verde* al acceso abierto) o puestos en abierto por algunas editoriales transcurrido un período de embargo, se sitúa entre el 50% y el 60% de todos los existentes (Piwowar, 2018). Es de prever que este porcentaje siga aumentando en los próximos años y llegue a estar muy próximo al 100%. En junio de 2019, en el *Directory of Open Access Journals (DOAJ)* se listan 13.405 revistas que dan acceso libre y gratuito a más de 4 millones de artículos.

Abadal (2017, p. 182) considera las revistas de acceso abierto (la *vía dorada*) uno de los pilares fundamentales para extender el nuevo modelo de comunicación científica, pero deben afrontar tres grandes retos:

- a) Incrementar su número y su proporción para conseguir la hegemonía en el sistema.
- b) Superar el debate sobre su calidad (y evitar la "contaminación" por parte de las denominadas revistas 'depredadoras').
- e) Consolidar las vías de financiación para su sostenimiento económico.

El éxito del movimiento Open Access para acceder libremente a los artículos de revista ha alentado su ampliación a un concepto más general: la llamada Open Science, con una serie de iniciativas adjetivadas con "open": O. Data, O. Peer review, O. Research, etc. Estas iniciativas afectan también a las revistas:

- Open Data busca que los investigadores pongan accesibles y en abierto los datos usados para su investigación. Éstos pueden ser hojas de cálculo, imágenes, cuestionarios, estadísticas, etc. Algunas revistas han dispuesto espacio para esos objetos digitales, pero lo más frecuente es que los autores los depositen en los repositorios especializados como *Figshare*, *Zenodo*, *Dryad*, o *Mendeley Data*. La finalidad de Open Data es doble: poder reutilizar los datos y poder comprobar si la investigación se hizo correctamente.
- Open peer review propone que sean públicas las evaluaciones de los manuscritos, así como los nombres de los evaluadores. Esto sería muy deseable para la transparencia del proceso, sin embargo en la práctica choca con los conflictos humanos que se pueden producir entre las personas, sobre todo si éstas se conocen. En nuestra opinión, y a pesar

de los posibles defectos, seguirá practicándose la evaluación en simple o doble anonimato (esta última claramente preferida por los autores), especialmente en revistas de alcance básicamente nacional.

- Depositar en un repositorio los artículos en versión *preprint* (= manuscrito del autor antes de pasar la evaluación por pares) para obviar el retraso que se produce hasta que son aprobados y publicados. Desde hace años ésta es una práctica habitual en Física y disciplinas relacionadas, pero cuesta que sea aceptada por las revistas de otras temáticas. Si las revistas tienen retraso en diagramar sus artículos ya aprobados, se les recomienda que mientras publiquen versiones provisionales llamadas *postprints* (= artículo que ha superado la evaluación por pares).

### 3. Gran competición

#### 3.1. Otros canales

Antes de internet las revistas eran los únicos canales de comunicación —especialmente en las ciencias experimentales—, junto con los libros y las conferencias. Con internet han aparecido otros medios y canales que compiten con las revistas:

- Repositorios, de los cuales según *OpenDoar* existen 3.520 en el mundo, tanto institucionales como temáticos, y de ellos 133 en España.
- Redes sociales (*ResearchGate*, *Academia.edu*, *Mendeley*, *Zotero*...) que permiten difundir cualquier tipo de documento subido por los usuarios (Rodríguez-Fernández et al., 2018).
- Blogs, webs personales...

La información distribuida por estos canales en general no está evaluada por pares —que es el gran activo de las revistas—, pero puede tener el aval de autores prestigiosos o de la institución donde está disponible, por lo que absorbe una parte de la atención de los autores que de otro modo hubiera ido íntegramente a las revistas. Y si una revista se consulta poco, posiblemente la biblioteca cancele la suscripción.

#### 3.2. Muchas revistas

El número de revistas sigue aumentando al ritmo de entre 3 y 13% anual, según las disciplinas (Larsen y Von-Ins, 2010). Según Ware y Mabe (2015) a finales de 2014 existían 34.550 revistas revisadas por pares (81% de ellas en inglés). Es muy fácil poner en marcha una revista online, especialmente si en la universidad o institución ya

funciona un programa de gestión editorial (como por ejemplo *Open Journal Systems*, *OJS*).

Las llamadas megarevistas son un fenómeno iniciado en 2006 con *PLoS One*, publicada por la empresa *Public Library of Science*. Extrañamente, ya en 2007 dicha revista fue indexada por la *Web of Science* y en 2009 contó con un factor de impacto de 4,351 (*Clarivate Analytics* no ha dado ninguna explicación a esa rapidez, aunque no cuesta imaginar a qué se debió tratándose de empresas). Muchos investigadores se lanzaron a publicar en ella dada su ligera revisión de los manuscritos: no se realiza una evaluación por pares, sino sólo una revisión superficial formal, dejando el artículo accesible online a merced de un aleatorio *open peer review* por parte de cualquier lector. Esta revisión tan superficial ha generado mucha desconfianza entre los editores tradicionales (Buriak, 2015). La revista fue aumentando rápidamente el número de artículos publicados, alcanzando un máximo de 31.500 en 2013. En los últimos años el número de artículos va bajando (22.000 en 2016) al mismo tiempo que baja su JIF (2,766 en 2017). Parecía que la causa del bajón era que los investigadores prefieren publicar en revistas especializadas en su disciplina, pero sin embargo otra megarevista generalista está teniendo mucho éxito (Villatoro, 2018). Se trata de *Scientific Reports*, publicada por *Springer Nature*, que en septiembre de 2016 superó a *PLoS One* en número de artículos (gráfico 1). Su JIF fue 4,122 en 2017. Otras editoriales han puesto en marcha megarevistas (*Sage Open*, *Real Society Open Science*, *BMJ Open*...) pero están teniendo mucho menos éxito (Björk, 2015). Según Ware y Mabe (2015) y Spezi et al. (2017) existen más de 50 megarevistas.

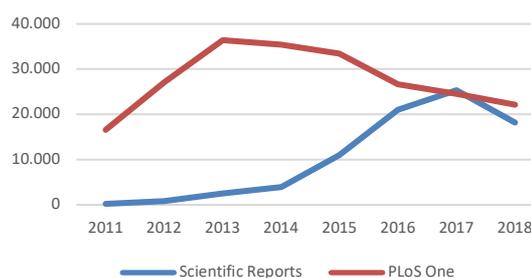


Gráfico 1. Artículos publicados por las megarevistas *PLoS One* y *Scientific Reports*, 2011-2018.

### 4. Los piratas

#### 4.1. Revistas depredadoras

La existencia de editoriales que publican revistas en acceso abierto cobrando una cantidad a los autores por artículo publicado ha sido aprovechado

por personas sin escrúpulos que han puesto en marcha “editoriales” que cobran pero no realizan ninguna evaluación ni rechazan los artículos de mala calidad, pues cuanto más publican más cobran. Tales revistas hacen un gran daño a la ciencia, contaminándola con toda clase de textos que pueden contener las mayores falsedades.

En 2008, Jeffrey Beall, bibliotecario de la *University of Colorado Denver*, puso en marcha una lista denunciando a estas editoriales y revistas, a las que llamó “depredadoras” (Beall, 2013). La lista causó muchos problemas a Beall (demandas judiciales por difamación de las editoriales, y malestar internamente en su propia universidad) por lo que en enero de 2017 cerró su web. Existe una copia de la lista, sin actualizar, en:

<https://beallist.weebly.com>

La empresa *Cabells* se hizo heredera de la lista de Beall, aunque con ella ahora el acceso es de pago. Según *Cabells*, su lista negra de revistas tiene más de 8.000 títulos.

<https://www2.cabells.com/blacklist>

Otros estiman que puede haber 30.000 títulos. Todas viven de los autores incautos que desconocen que publicar en ellas no les va a dar ningún rédito académico, pues son revistas que en general no están indexadas en ninguna base de datos. Es difícil distinguir entre las *buenas* y las *malas* revistas, pues la frontera entre ambas es difusa. Puede haber editoriales que no están en la lista negra, pero que no hacen evaluaciones correctas y publican investigaciones erróneas o falsas. Somoza-Fernández, Rodríguez-Gairín y Urbano (2016) analizaron las revistas depredadoras listadas por Beall que habían conseguido ser indexadas por las bases de datos. Las editoriales depredadoras se caracterizan por:

- Localizar a autores concretos invitándoles a publicar en su revista mediante un mensaje adulador. Ocultan que al final del proceso cobrarán. Ninguna editorial seria se dirige directamente a los autores, a menos que se trate de encargar artículos a VIPs de la disciplina y en este caso tal vez pague en vez de cobrar.
- Ponerse nombres muy generales para poder abarcar cualquier tema, por ej.: *Journal of Academic Science*, o *American Journal of Scholarly Research*...
- Decir que están indexadas en bases de datos y citar indicadores con nombres parecidos, como ICI, JSR, Índice de Impacto, *Web of Scientific*s...
- Publicar los artículos a los pocos días de recibirlos.

Si un investigador recibe una invitación sospechosa debe consultar si la revista figura en la lista negra de Beall, o de *Cabells*; está indexada en las bases de datos donde dicen estar; y figura en la lista blanca *Directory of Open Access Journals (DOAJ)*: <https://doaj.org>.

## 4.2. Sci-Hub

*Sci-Hub* es una web que da acceso gratuito a millones de artículos y libros de investigación, tanto de pago como de acceso abierto. Obtiene los artículos gracias a colaboradores que tienen acceso a las colecciones digitales de las universidades y le ceden sus contraseñas. *Sci-Hub* los descarga y almacena en sus propios servidores.

*Sci-Hub* fue fundado por Alexandra Elbakyan en 2011 en Kazajistán, como una reacción al alto costo de los documentos de investigación (González-Solar; Fernández-Marcial, 2019). Se calcula que en 2017 tenía el 86% de todos los pdfs de pago, y servía 6 millones de descargas al mes, según Himmelstein et al. (2018) y Tzovaras (2018). El sitio es utilizado sobre todo desde los países en desarrollo, pero según los citados autores también tiene muchos usuarios en países avanzados, lo cual representa una sangría en el uso de las revistas suscritas legalmente. A corto plazo es posible que las bibliotecas de los países avanzados sigan suscritas a las revistas como hasta ahora, pero es probable que si algunas revistas se usan menos al final se cancele su suscripción, como ya apuntamos unas líneas más arriba.

## 5. Futuro de las revistas

### 5.1. Crisis de 2008

Los factores descritos en los apartados anteriores causaron y aún están causando importantes cambios en el mundo de las revistas, a los que se unió la crisis económica iniciada en 2008 que hizo que muchas bibliotecas cancelaran las suscripciones por falta de fondos. Quienes más se resintieron fueron las editoriales pequeñas, pues para gastar su reducido presupuesto las bibliotecas dieron prioridad a los *big deals* (compras de paquetes cerrados de cientos de títulos, entre los que están algunas revistas “cebo”: las que los investigadores consideran “imprescindibles”) vendidos por las grandes editoriales. “El pequeño editor encuentra cada vez más dificultades para mantenerse en el circuito comercial y muchas más para obtener beneficios” (Rodríguez-Yunta, 2016). Una vez canceladas algunas suscripciones, cuesta mucho volver a darlas de alta: bibliotecas e investigadores han vivido los peores años de la crisis sin ellas, y han aprendido a pasar sin —aparte de que el sector ha cambiado y

han aparecido otras fuentes de información que compiten—. Aunque secundaria, otra circunstancia paralela que ha favorecido las cancelaciones es el paulatino abandono de las versiones en papel de las revistas, que ha hecho desaparecer un poco el “sentido de tener colecciones completas” de los bibliotecarios.

## 5.2. Falta de reconocimiento del trabajo editorial

Otro factor que no ayuda a las revistas es la falta de reconocimiento de su función, y del trabajo de su equipo de redacción. En general se tiene la idea de que los editores aportan poca cosa al proceso de publicación. Se ignora su gran valor añadido detectando posibles plagios, gestionando la evaluación por pares, revisando y mejorando los textos, tablas y gráficos, aplicando un estilo homogéneo a los manuscritos, detectando errores, resolviendo dudas, maquetando, subiendo los pdfs a la web, manteniendo la correspondencia con autores y evaluadores, extendiendo certificados, promoviendo los artículos por las redes sociales, haciendo marketing de la revista en congresos...

La ignorancia del trabajo editorial se manifiesta con frecuencia en los ataques a las grandes editoriales multinacionales que leemos en la prensa. No vamos a entrar en el debate de si obtienen muchos o pocos beneficios, pero no pueden aceptarse términos como que las editoriales “son un obstáculo para el avance de la ciencia”, “ahora que ya no se edita en papel el coste de las revistas debería ser prácticamente cero”, “las editoriales son casi el 100% parásitas” o “las universidades pagan 2 veces por los artículos, primero su producción y luego su lectura” (debería ser evidente que alguien tiene que pagar por el trabajo de edición y publicación).

Como consecuencia, los editores que trabajan en universidades, por ejemplo, se encuentran solos y poco valorados, con unos recursos mínimos tanto económicos como de tiempo (no reciben una compensación tal como reducción de clases, de actividad en comités, etc.). Aunque se refieren a las monografías, Abadal, Ollé y Redondo (2018) dicen:

[...] se necesita el apoyo directo de los rectores para que se confirme el apoyo económico de la universidad, sin el cual no se puede llevar a cabo una política de acceso abierto. [...] queda clara y patente la contradicción existente entre las declaraciones teóricas y el apoyo financiero.

Contar con medios económicos permite profesionalizar los procesos, lo que se traduce en importantes mejoras en la calidad de las revistas: pagar a los revisores y poder exigir rapidez; pagar traducciones profesionales al inglés, tanto de los

abstracts como de los textos completos; pagar a colaboradores y exigir calidad (sin dinero también se pueden tener colaboradores voluntarios que ayuden en las labores editoriales, pero a condición de aceptar lo que “buenamente” pueden hacer, así como sus retrasos); poder encargar artículos a autores importantes que ayuden a dar realce a la revista y aumente su factor de impacto; mejorar la web (diseño atractivo y más funcional, incorporar altmétricas de los artículos, búsqueda de artículos similares, navegación entre autores y entre palabras clave, posibilidad de subir los datos usados en las investigaciones de los artículos, asegurar la preservación en *Lockss*, *DuraCloud*, *Preservica*, *Arkivum*...); realizar más marketing, una actividad absolutamente necesaria debido a la gran competitividad entre revistas de la que hablamos en el apartado 3. Disponer de un folleto bien diseñado y actualizado (imprescindible para promover la revista si ésta sólo se publica online), estar presentes en las redes sociales, enviar los pdfs a los autores citados, presentarla en congresos...

## 5.3. Disminución de las suscripciones

Como resultado de los factores reseñados anteriormente: más información en acceso abierto, competencia, piratería y crisis económica iniciada en 2008 y todavía no superada, las revistas que se financian mediante suscripciones ven como de año en año va disminuyendo el número de suscriptores. Véase por ejemplo el gráfico 2 correspondiente a la revista *El profesional de la información (EPI)*.

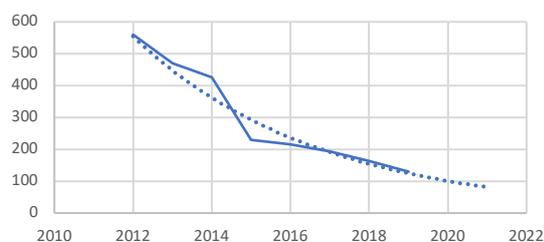


Gráfico 2. Evolución de las suscripciones de la revista *El profesional de la información* (La línea de puntos señala la tendencia exponencial)

La disminución de suscripciones de *EPI* la hizo económicamente inviable a partir de 2015, por lo que se decidió empezar a cobrar a los autores, pasando así a ser una revista de las llamadas *híbridas*: dos vías de financiamiento: mediante suscripciones y mediante pagos por publicar (*article processing charges* o APCs). En el caso de *EPI* hay dos tipos de APC, 1) pago sólo por publicar, y 2) pago por publicar y además liberar el artículo, para que esté en acceso abierto en las dos webs

de la revista. Los autores están aceptando las tarifas de *EPI* gracias a que está indexada en *Scopus* (situada en Q1) y en la *Web of Science* (Q2).

Estar en modo híbrido ha sido considerado en general como un paso intermedio hacia el acceso abierto completo, eliminando las suscripciones, y así está sucediendo con *EPI*, pero ocurre que para la mayoría de revistas híbridas no es así: se quedan en esta situación sin evolucionar (Björk, 2012). Es por esto que en el citado *Plan S de Science Europe*, las instituciones financiadoras a partir de 2020 no aceptarán la publicación en revistas híbridas (a menos que éstas demuestren estar avanzando hacia el pleno acceso abierto).

Las revistas que actualmente se financian mediante suscripciones deberán ir pensando en financiarse cobrando a los autores una APC, pero éstos sólo querrán publicar en ellas si están indexadas en alguna de las dos bases de datos citadas, *Scopus* o *Web of Science*, que las agencias de acreditación toman como referentes de calidad. Ha habido diversas declaraciones (*Dora*, *Manifiesto de Leiden*) para que en vez de tomar el factor de impacto de las revistas como proxy de la calidad de los artículos que publican, se evalúe cada artículo, independientemente de la revista. Incluso el *Tribunal Supremo* de España dio la razón a una profesora que reclamó que no se había valorado su artículo publicado en una revista no indexada en *Scopus* ni *WoS* (Aunión, 2018). Sin embargo esto representaría un trabajo ingente para las agencias de acreditación, y por ahora es inasumible. “La producción científica es tan amplia que no es posible evaluar cada artículo de forma individual” (Rodríguez-Yunta, 2016). Es previsible que el sistema de evaluación actual siga igual en los próximos años.

Las revistas comerciales con un alto factor de impacto y muchos suscriptores (como por ejemplo *New England journal of medicine* (JIF 2017 = 79,260; *Lancet*, JIF = 53,254; *Nature*, JIF 2017 = 41,577...) con toda seguridad seguirán siendo de suscripción varios años más, mientras sigan gozando de su privilegiada posición. Si sus editoriales quisieran obtener los mismos beneficios cobrando APCs tendrían que poner unos precios tan altos que pocos investigadores los podrían pagar.

En esta web se informa sobre cancelaciones de suscripciones:

<http://tagteam.harvard.edu/hubs/oatp/tag/oa.cancellations>

#### 5.4. Revistas Q5

Según el *Directorio Latindex* en 2018 en España se publican 4.474 revistas. De éstas, las 2.298

que cumplen unos mínimos formales figuran también en el *Catálogo Latindex v. 1.0*. Las revistas que no cumplen tales requisitos mínimos no significa que no puedan publicar contenidos de calidad, pero dada la relación existente en general entre presentación y contenido, las que no están en el *Catálogo* no ofrecen mucha garantía de que trabajen bien. En información científica los protocolos formales son importantes.

En 2015 Abejón y Rodríguez-Yunta estimaban que el 13% de las revistas listadas en *Latindex* dejaron de publicarse o estaban en una situación desconocida, por lo que de conservarse el porcentaje en 2018, en España quedarían unas 2000 revistas vivas.

De éstas 2000, sólo está indexadas en *Scopus* 557 (23 de ellas en el año 2018 no hace los 4 años necesarios y todavía no tienen el índice de calidad SJR calculado).

Con los actuales criterios de acreditación de los profesores universitarios de la *ANECA* y la *CNEAI*, valorando sólo los artículos publicados en revistas indexadas en la *Web of Science* o *Scopus*, podemos preguntarnos cuál es el futuro de las 1.443 revistas no indexadas, a las que llamamos “revistas Q5”, cosa que trataremos en la sección siguiente.

Muchas de las revistas españolas se publican en acceso abierto, subvencionadas por las instituciones, principalmente las universidades. Algunas tienen un buen nivel científico (Ruiz-Pérez et al. 2015): así, de las 534 revistas españolas indexadas en *Scopus* y con SJR, 144 (27%) las publican universidades (gráfico 3).

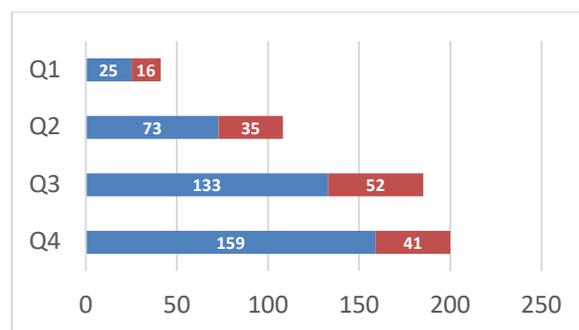


Gráfico 3. Revistas españolas indexadas en la base de datos *Scopus*, por cuartiles (En color naranja se muestra el número de revistas publicadas por las universidades)

En el directorio *Dulcinea* se listan 1.846 revistas, de las cuales 985 (53%) están publicadas por el CSIC y las universidades, y de éstas, 784 (80%) son gratuitas. Siguiendo con los datos proporcionados por *Dulcinea*, las universidades más el CSIC publican 323 revistas de humanidades y

525 de ciencias sociales. O sea, de las 985 revistas publicadas por las universidades más el CSIC, 848 (86%) son de ciencias sociales o humanidades.

Rodríguez-Yunta y Giménez-Toledo (2013) analizaron las revistas de estas temáticas y hallaron que:

- muchas son unipersonales y por ello, con la falta de recursos y de dedicación, están producidas con poca profesionalidad;
- reciben pocos originales y tienen poca producción (el 68% de revistas publica menos de 20 artículos al año) —esto es importante de cara a *Google Scholar Metrics*, que no tiene en cuenta las revistas que no publican al menos 100 artículos en los últimos 5 años—;
- reciben pocas citas debido a que publican artículos sin interés o que aportan poca novedad a la ciencia... “de las 41 revistas presentes en *Resh* —un ranking de calidad realizado por el CSIC que dejó de actualizarse en 2009— para la categoría Historia moderna y contemporánea, 13 no han recibido ninguna cita en el período 2004-2008. Lo mismo ocurre con 90 de las 188 revistas de Educación editadas en España”.
- se retrasan respecto a su calendario programado (24% publica con retrasos superiores a 1 año);
- llevan a cabo una evaluación poco rigurosa, pues no pueden quedarse sin artículos, y publican lo que les llega, aunque no sea de la mejor calidad;
- tienen endogamia (sólo un 45% publica más de 75% de artículos de autores externos).

Rodríguez-Yunta y Giménez-Toledo (2013) abogan por la fusión o la coedición de revistas, de lo cual existen algunas experiencias con buenos resultados (Agustí *et al.*, 2015). “La dispersión de títulos dificulta tanto la visibilidad de la producción como el rigor de la revisión en el proceso editorial” (Rodríguez-Yunta, 2016). Con la fusión se puede obtener una revista más competitiva, evitar la multiplicación de esfuerzos, no malgastar fondos y recursos, y fomentar la profesionalización (Rodríguez-Yunta; Tejada-Artigas, 2013; Expósito-García *et al.*, 2018). Pero llegar a un acuerdo no es fácil por las pérdidas de autonomía y protagonismo, sobre todo teniendo en cuenta que uno de los dos títulos tiene que desaparecer. Un nuevo título consensuado debería llevar un nuevo ISSN y eso es casi como comenzar la revista desde cero. Pero “...los editores deben ser conscientes de que lo importante no es 'estar' en estos sistemas, sino 'cómo se está’”.

Las revistas que pretenden tener el estatus de publicación de ciencia original y que no consigan estar indexadas en las bases de datos de referencia deberán plantearse fusionarse o desaparecer. No tiene mucho sentido que sigan manteniendo una vida lánguida y anémica, sin impactar nada a la disciplina científica a la que sirven. Otra cosa es que cambien su misión y se conviertan claramente en revistas de divulgación, con un papel en la enseñanza y en la formación de futuros investigadores.

## 6. Discusión y conclusiones

Se han presentado brevemente los fenómenos y causas que están determinando una nueva situación de las revistas en todo el mundo, gestada ya en el siglo pasado pero acelerada en los últimos diez años: abandono del papel, crisis económica, acceso abierto, excesivo número de revistas, megarevistas, competencia de otros recursos de información, revistas depredadoras, *Sci-Hub*...

La sobreabundancia de información, con bastantes redundancias, hace replantear la misión de las revistas, en especial las institucionales que no consiguen estar indexadas en las bases de datos *Scopus* o *WoS*, tomadas por las agencias de acreditación como referencias de calidad, y por defecto, por la calidad de los artículos publicados en ellas y a su vez por el nivel científico de sus autores. Tales revistas que hemos llamado Q5 deberían fusionarse. Por lo mismo, si un promotor no cuenta con un equipo importante de colaboradores, no debería crear una nueva revista sino ver de integrarse y mejorar una revista ya existente.

El modelo de financiación para la gran mayoría de revistas comerciales basado en las suscripciones que pagan las bibliotecas se va viendo obligado a cambiar al modelo de pago de APCs (*article processing charges*) para publicar por parte de los autores, pasando en algunos casos por un período de transición híbrido, hasta conseguir publicarse completamente en abierto. La publicación en abierto es el estado deseable para cualquier revista, con el fin de tener más visibilidad (pudiendo también los autores depositar libremente sus pdfs en los repositorios), más lecturas y recibir más citas, que son la medida de su calidad e impacto, y de su éxito.

Un problema importante sufrido por las revistas es la invisibilidad y la falta de reconocimiento del trabajo de sus redacciones. La idea imperante es que los editores no hacen más que recibir los manuscritos y subirlos a una web: no se conoce su trabajo. El hecho de tener que pagar APCs para publicar puede tener el efecto secundario beneficioso de ayudar a entender el gran valor añadido

que aportan las revistas. Gran parte de las críticas proceden del entorno universitario, con comentarios que claman que las universidades se ven obligadas a pagar dos veces: una por el investigador y su trabajo y otra por la publicación. La realidad es que —salvo unas pocas excepciones— las universidades no han sabido crear servicios eficientes y de calidad —no endogámicos— de publicación de revistas (Expósito-García et al., 2018). A primera vista los precios de las APCs (se considera que entre 1.000 y 1.500 USD es una cifra correcta) sorprenden a las personas que desconocen dicho trabajo, pero quizá poco a poco vaya permeando la idea de que se trata de precios justos para compensar las horas de una labor altamente especializada para la que se requiere una larga experiencia.

Finalmente, es necesario que bibliotecarios y profesores realicen mucho trabajo pedagógico para evitar que los autores, sobre todo los novales, caigan en manos de las revistas depredadoras, que son una plaga para el conocimiento científico y que por desgracia van a seguir engañando y emponzoñando la ciencia con artículos no evaluados en los próximos años.

## Referencias

- Abadal, Ernest (ed.) (2017). *Revistas científicas. Situación actual y retos de futuro*. // Barcelona: Universitat de Barcelona. ISBN: 978 84 9168 038 3.
- Abadal, Ernest; Ollé, Candela; Redondo, Sílvia (2018). Publicación de monografías en acceso abierto por editoriales universitarias españolas. // *El profesional de la información*. 27:2, 300-311. <https://doi.org/10.3145/epi.2018.mar.08>
- Abejón-Peña, Teresa; Rodríguez-Yunta, Luis (2015). La participación española en Latindex: valoración de resultados e impacto sobre la calidad y evaluación de las publicaciones científicas. // *Ciencia da Informação*. 44:2, 258-275. <http://revista.ibict.br/ciinf/article/view/1795>
- Agustí, Lluís; Ardanuy, Jordi; López-Borrull, Alexandre; Rodríguez-Gairín, Josep-Manuel; Vall, Aurora; Vázquez, Mercè; Pérez-Montoro, Mario; Ollé, Candela (2015). // *Revista BiD: el valor de colaborar, experiencias a compartir*. // Conferencia intl sobre revistas de ciencias sociales y humanidades (CRECS), Murcia (Spain), 7 May 2015. <http://eprints.rclis.org/28213>
- Amat, Lluís; Serra, Gemma (1991). La CDB: un projecte de biblioteques biomèdiques de Catalunya. // *ITEM*:9, 68-75. <http://www.raco.cat/index.php/Item/article/download/22390/22224>
- Anglada, Lluís M. (2015). Compras consorciadas, big deals, concentración editorial, acceso abierto y el circuito de la comunicación científica. // *Anuario ThinkEPI*. 9, 41-48. <http://dx.doi.org/10.3145/thinkepi.2015.08>
- Anglada, Lluís M. (2019). Muchos cambios y algunas certezas para las bibliotecas de investigación, especializadas y centros de documentación. // *El profesional de la información*. 28:1, e280113. <https://doi.org/10.3145/epi.2019.ene.13>
- Aunión, Juan-Antonio (2018). La justicia obliga a leer los trabajos para acreditar los méritos de investigación en la Universidad. // *El País*, 18 de septiembre. [https://elpais.com/sociedad/2018/09/18/actualidad/1537270842\\_896963.html](https://elpais.com/sociedad/2018/09/18/actualidad/1537270842_896963.html)
- Baiget, Tomàs (2010). Ética en revistas científicas. // *Ibersid*. 4, 59-65. <https://ibersid.eu/ojs/index.php/ibersid/article/view/3873>
- Beall, Jeffrey (2012). Predatory publishers are corrupting open access. // *Nature*. 489, p. 179, September 12th. <https://doi.org/10.1038/489179a>
- Björk, Bo-Christer (2012). The hybrid model for open access publication of scholarly articles—a failed experiment? // *Journal of the American Society for Information Sciences and Technology*. 63:8, 1496-1504. <https://doi.org/10.1002/asi.22709>
- Björk, Bo-Christer (2015). Have the 'mega-journals' reached the limits to growth? // *PeerJ*. 3, e981. <https://doi.org/10.7717/peerj.981>
- Buriak, Jillian M. (2015). Mega-journals and peer review: can quality and standards survive? Editorial. // *Chemistry of materials*. 27:7, p. 2243. <https://doi.org/10.1021/acs.chemmater.5b01142>
- Dinu, Nicoleta-Roxana (2018). Tendiște la nivel mondial în privința revistelor științifice. // *Biblioteca*:12, 355-357. <http://eprints.rclis.org/33903>
- Expósito-García, Alfonso; Velasco-Morente, Francisco (2018). How efficient are universities at publishing research? A data envelopment analysis of Spanish state universities. // *El profesional de la información*. 27:5, 1108-1115. <https://doi.org/10.3145/epi.2018.sep.14>
- González-Solar, Llarina; Fernández-Marcial, Viviana (2019). Sci-Hub, a challenge for academic and research libraries. // *El profesional de la información*. 28:1, e280112. <https://doi.org/10.3145/epi.2019.ene.12>
- Harzing, Anne-Wil; Adler, Nancy J. (2015). Essays, dialogues & interviews disseminating knowledge: From potential to reality—New open-access journals collide with convention. // *Academy of management learning & education*. 15:1. <https://doi.org/10.5465/amle.2013.0373>
- Himmelstein, Daniel S. et al. (2018). Sci-Hub provides access to nearly all scholarly literature. // *eLife*. 7, e32822, PMID: 29424689, PMC5832410. <https://dx.doi.org/10.7554/eLife.32822>
- Keefer, Alice (1996). Proyecto Tulip: análisis crítico. // *Información world en español*:42, 24-25. [http://www.profesionaldelainformacion.com/contenidos/1996/marzo/proyecto\\_tulip\\_analisis\\_critico.html](http://www.profesionaldelainformacion.com/contenidos/1996/marzo/proyecto_tulip_analisis_critico.html)
- Larsen, Peder-Olesen; Von-Ins, Markus (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. // *Scientometrics*:84, 575–603. <https://doi.org/10.1007/s11192-010-0202-z>
- Melero, Remedios (2005). Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. // *El profesional de la información*. 14:4, 255-266. <http://eprints.rclis.org/65711/EPI-rmelero.pdf>
- Melero, Remedios; Melero-Fuentes, David; Rodríguez-Gairín, Josep-Manuel (2018). Monitoring compliance with governmental and institutional open access policies across Spanish universities. // *El profesional de la información*. 27:4, 858-878. <https://doi.org/10.3145/epi.2018.jul.15>
- Piowar, Heather; Priem, Jason; Larivière, Vincent; Alperin, Juan-Pablo; Matthias, Lisa; Norlander, Bree; Farley, Ashley; West, Jevin; Haustein, Stefanie (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. // *PeerJ*, 6, e4375. PubMed 29456894. <https://doi.org/10.7717/peerj.4375>
- Plan S (2018). Science Europe. <https://www.scienceurope.org/wp-content/uploads/2018/09/cOAlitionS.pdf>
- Rodríguez-Fernández, María-Magdalena; Sánchez-Amboage, Eva; Martínez-Fernández, Valentín-Alejandro (2018). Utilización, conocimiento y valoración de redes

- sociales digitales científicas en las universidades gallegas. // *El profesional de la información*. 27:5, 1097-1107. <https://doi.org/10.3145/epi.2018.sep.13>
- Rodríguez-Gairín, Josep-Manuel (2012). 20 años de automatización de préstamo interbibliotecario en España (1992-2012). // *El profesional de la información*. 21:6, 557-566. <http://dx.doi.org/10.3145/epi.2012.nov.02>
- Rodríguez-Gairín, Josep-Manuel (2015). Préstamo interbibliotecario e Internet: análisis de un modelo para bibliotecas universitarias españolas basado en agentes. Tesis doctoral. // Doctorado en información y documentación en la sociedad del conocimiento, Barcelona. <https://core.ac.uk/download/pdf/43551585.pdf>
- Rodríguez-Yunta, Luis (2015). Evaluación de las revistas españolas: pasos hacia un sistema que mida la apertura y la internacionalidad. // *Anuario ThinkEPI* 2015. 9, 189-195. <http://hdl.handle.net/10261/125672>. <https://doi.org/10.3145/thinkepi.2015.44>
- Rodríguez-Yunta, Luis (2016). Difusión y evaluación de la investigación histórica en la era digital: revistas españolas y bases de datos. // *Estudios Humanísticos*:15, 205-237. <http://revpubli.unileon.es/index.php/EEHHHistoria/article/view/5048>
- Rodríguez-Yunta, Luis (2017). Evaluación de publicaciones en humanidades: cambios necesarios frente a indicadores inadecuados. // *Anuario ThinkEPI* 2017, vol. 11. <https://doi.org/10.3145/thinkepi.2017.43>
- Rodríguez-Yunta, Luis; Giménez-Toledo, Elea (2013). Fusión, coedición o reestructuración de revistas científicas en humanidades y ciencias sociales. // *El profesional de la información*, enero-febrero. 22:1, 36-45. <http://dx.doi.org/10.3145/epi.2013.ene.05>
- Rodríguez-Yunta, Luis; Tejada-Artigas, Carlos-Miguel (2013). El editor técnico: un perfil necesario para la profesionalización de la edición de revistas científicas en el entorno digital. // *Anales de documentación*. 16:2. <http://revistas.um.es/analesdoc/article/view/176391>.
- Rodríguez-Yunta, Luis; Tejada-Artigas, Carlos-Miguel. Recomendaciones a los editores para fortalecer la revisión por pares en su revista científica. // *Anuario ThinkEPI* 2016, vol. 10. <https://doi.org/10.3145/thinkepi.2016.44>
- Ruiz-Pérez, Rafael; Martín-Martín, Alberto; Delgado-López-Cózar, Emilio (2015). Las revistas universitarias en el marco de los criterios de evaluación de la actividad investigadora en España. // *Revista española de documentación científica*. 38:2, p. e081. <http://dx.doi.org/10.3989/redc.2015.2.1191>
- San-José-Montano, Blanca; Espantaleón-Agreda, Manuel; De-Santiago-Puchol, Ana; Díaz-Galcerán, Victoria (2003). El Servicio de Obtención de documentos en las bibliotecas virtuales: Nuevas oportunidades en la gestión de Documentos = Conocimiento. // *IX Jornadas de información y documentación en ciencias de la salud. Gestión del conocimiento y bibliotecas de salud, Málaga (España)*, 13 de noviembre, 611-617. <http://eprints.rclis.org/25543>
- Segado-Boj, Francisco; Martín-Quevedo, Juan; Prieto, Juan-José (2018). Percepción de las revistas científicas españolas hacia el acceso abierto, open peer review y altmetrics. // *Ibersid*. 12:1, 27-32. <https://www.ibernid.eu/ojs/index.php/ibernid/article/view/4407>
- Somoza-Fernández, Marta; Rodríguez-Gairín, Josep-Manuel; Urbano, Cristóbal (2016). Presence of alleged predatory journals in bibliographic databases: analysis of Beall's list. // *El profesional de la información*. 25:5, 730-737. <https://doi.org/10.3145/epi.2016.sep.03>
- Spezi, Valerie; Wakeling, Simon; Pinfield, Stephen; Creaser, Claire; Fry, Jenny; Willett, Peter (2017). Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review. // *Journal of documentation*. 73:2, 263-283. <http://dx.doi.org/10.1108/JD-06-2016-0082>
- Térmens-Graells, Miquel (2005). Los consorcios, una nueva etapa de la cooperación bibliotecaria. // *El profesional de la información*. 14:3, 166-173. <http://www.elprofesionaldelainformacion.com/contenidos/2005/mayo/2.pdf>
- Tzovaras, Bastian-Greshake (2018). Sci-Hub download log of 2017. <https://zenodo.org/record/1158301>
- Villatoro, Francisco R. (2018). Scientific Reports ya es la mayor megarrevista científica superando a PLoS ONE. // *La ciencia de la Mula Francis*. 2018-02-19. <https://francis.naukas.com/2018/02/19/scientific-reports-supera-a-plos-one-en-2017-siendo-la-mayor-megarrevista-cientifica>
- Ware, Mark; Mabe, Michael (2015). The STM report: An overview of scientific and scholarly journal publishing. // *International Association of Scientific, Technical and Medical Publishers (STM), The Hague, The Netherlands*. <https://digitalcommons.unl.edu/scholcom/9>
- Wicherts, Jelte M. (2016). Peer review quality and transparency of the peer-review process in open access and subscription journals. // *PLoS One*. 1:1, e0147913. <https://doi.org/10.1371/journal.pone.0147913>.

---

Enviado: 2019-02-26. Segunda versión: 2019-06-14.  
Aceptado: 2019-06-17.

---

---

# Sistemas e processos de organização do conhecimento: uma análise conceitual da literatura brasileira de ciência da informação

*Sistemas y procesos de organización del conocimiento:  
un análisis conceptual de la literatura brasileña de ciencia de la información*

*Systems and processes of knowledge organization: a conceptual analysis of the Brazilian Information Science literature*

---

**Walter MOREIRA, Fernanda Carolina PEGORARO NOVAES, Isabela SANTANA DE MORAES**

Universidade Estadual Paulista Júlio de Mesquita Filho, Avenida Hygino Muzzi Filho, 737, Marília, SP, Brasil;  
walter.moreira@unesp.br, cttfernanda@gmail.com, isabela.santanademoraes@hotmail.com

## Resumen

En la terminología de la intersección entre las Ciencias de la Información y la Organización del Conocimiento, algunos términos han ganado importancia sin que sus ámbitos sean lo suficientemente claros. Teniendo en cuenta este marco, establecimos como objetivos analizar y sistematizar el uso de los términos “sistemas de organización del conocimiento” (KOS) y “procesos de organización del conocimiento” (KOP) en las revistas brasileñas de Ciencias de la Información. Para ello, utilizamos la metodología de análisis de contenido y utilizamos como corpus un conjunto de artículos publicados en revistas brasileñas clasificadas con “Qualis / Periódicos” “A1” o “A2” que discute los términos KOS y KOP. Los conceptos de análisis se discutieron a partir de las siguientes categorías definidas a posteriori: comparación con otros instrumentos, tipologías, bases teóricas y aplicaciones y usos. Se verificó que todavía no hay consenso sobre el alcance del concepto KOS y que la expresión KOP aún no se utiliza ampliamente en la literatura brasileña. Las tipologías aún no están declaradas formalmente. El tesaurus, los sistemas de clasificación y las ontologías se indican como los tipos más comunes de KOS; y como tipos comunes de KOP la clasificación y la indización, todos orientados a la organización, representación y recuperación de información. Como desafío para trabajos futuros, creemos que es posible enriquecer el análisis conceptual mediante la aplicación del eje onomasiológico.

**Palabras clave:** Sistemas de organización del conocimiento. Procesos de organización del conocimiento. Organización del conocimiento. Brasil.

## 1. Introdução

A abrangência e o alcance de alguns conceitos já consensuais e relativamente estáveis na biblioteconomia e na ciência da informação, tais como “linguagens documentárias”, “sistemas de classificação”, “instrumentos de organização e representação do conhecimento” e “vocabulários controlados”, estão sendo novamente discutidos em função de uma possível revisão terminológica.

## Abstract

In the terminology interface between Library and Information Science and Knowledge Organization, some terms have gained prominence without their scopes being however sufficiently clear. Considering this framework, we established as objectives to map and systematize the use of the terms knowledge organization systems (KOS) and knowledge organization processes (KOP) in Brazilian journals of Library and Information Science. To do so, we used the methodology of content analysis and used as corpus a set of papers published in Brazilian journals classified with “Qualis/Periódicos” “A1” or “A2” that discusses the terms KOS and KOP. The concepts under analysis were discussed based on the following categories defined a posteriori: comparison with other instruments, typologies, theoretical bases, and applications and uses. It was verified that there is still no consensus on the scope of the concept KOS; and that the expression KOP is not yet widely used in the Brazilian literature. The typologies are not yet formally stated. Thesaurus, classification systems, and ontologies are indicated as more common types of KOS, and as common types of KOP classification and indexing, all oriented towards information organization, representation, and retrieval. As a challenge for future works, we think it's possible to enrich the conceptual analysis by the application of onomasiology axis.

**Keywords:** Knowledge organization systems. Knowledge organization processes. Knowledge organization. Brasil.

O conceito de “linguagens documentárias” inclui, numa espécie de definição extensional, os sistemas de classificação bibliográfica, as listas de cabeçalhos de assunto e os tesausos. Em função da semelhança de algumas propostas no que tange às suas aplicações, às suas estruturas de base relacional e a alguns subsídios teórico-metodológicos compartilhados, tornou-se comum, principalmente na literatura da *International So-*

ciety of Knowledge Organization (ISKO), a utilização da expressão *knowledge organization systems* (KOS) – em português, *sistemas de organização do conhecimento* (SOC) – para se referir aos instrumentos utilizados para a organização e representação do conhecimento (Zeng, 2008; Barité Roqueta, 2011; Souza, Thudope e Almeida; 2012; Hjørland, 2015; Hjørland, 2016; Mazzocchi, 2017a, 2017b). Verifica-se um movimento semelhante em algumas publicações brasileiras (Carlan, 2010; Brascher e Carlan, 2010; Carlan e Medeiros, 2011; Lara, 2015).

Esta investigação tem como objetivos: a) mapear a utilização dos termos “sistemas de organização conhecimento” e “processos de organização do conhecimento” em periódicos representativos da literatura brasileira de ciência da informação e b) sistematizar e analisar as definições encontradas para os dois termos.

Na Teoria do Conceito (Dahlberg, 1978), o conceito constitui-se como uma unidade de conhecimento. Nesse sentido, sua compreensão orienta e, quando tomada literalmente, condiciona o acesso ao conhecimento. Barité *et al.* (2013, p. 43) apresentam a seguinte definição para *conceito* em seu *Dicionário de organización del conocimiento*:

Abstracción o noción que refiere a una unidad de conocimiento, independiente de su expresión lingüística, y comprende el conjunto de sus rasgos esenciales. [...] es el elemento indivisible que permite representar el conocimiento contenido en los documentos y organizar los enunciados correspondientes a la idea que se tiene de cualquier cosa.

Na relação entre o conceito e sua designação verbal, ambos unidades separadas apenas virtualmente e unificados pelo *termo*, o conceito tem prioridade na terminologia, isto é, relação de significação é organizada pelo percurso onomasiológico. O termo, contrariamente à palavra, é uma unidade contextualizada e possui um referente de interpretação. Assim, conforme Lara (2004), “a palavra no discurso – o termo – associa-se a uma classe de objetos, às coisas do mundo real, tendo, dessa forma, uma extensão”.

Os SOC são, essencialmente, compostos por termos, conceitos e relações semânticas. Assim, as discussões identitárias a respeito de termos e conceitos, bem como sobre as relações triádicas de significação em que se inserem, constituem-se como foco dos estudos que envolvem a compreensão, o desenvolvimento e aplicação de SOC. Dahlberg (2009) alçou essa discussão sobre conceitos e termos ao patamar de “maiores desafios da ISKO”.

Para corroborar a concepção de que os conceitos formam as unidades de conhecimento e que a

compreensão de uma área passa pela compreensão e pela apreensão de sua linguagem, destacam-se o pensamento de Benveniste (1989) e de Sager (1993).

Para o primeiro, a terminologia coloca-se como a própria condição da existência de uma determinada ciência. Não há outro meio para delinear seu objeto e legitimar-se a não ser por meio da construção de uma linguagem própria que se lhe afigure como identitária. Na expressão direta do pensamento de Benveniste (1989, p. 52, tradução livre) pode-se ler:

[...] Uma ciência só começa a existir ou consegue se impor na medida em faz existir e em que impõe seus conceitos, através de sua denominação. Ela não tem outro meio de estabelecer sua legitimidade senão por especificar seu objeto denominando-o, podendo este constituir uma ordem de fenômenos, um domínio novo ou um modo novo de relação entre certos dados [...]. Denominar, isso é, criar um conceito, é, ao mesmo tempo, a primeira e última operação de uma ciência.

Para o segundo, “conocer una materia equivale a tener un dominio de parte de los lenguajes de esa materia; dominar los lenguajes de una materia equivale a tener cierta comprensión de la materia” (Sager, 1993, p. 15).

As terminologias científicas a que se referem os dois autores citados são recolhidas e organizadas a partir das linguagens de especialidade. Os SOC, na condição de instrumentos semânticos logicamente organizados, utilizados nas fases de produção, organização e uso da informação, também constroem, de modo solidário, as terminologias de cada área.

## 2. Novos conceitos = novos modelos de entendimento

O termo SOC, segundo a definição de Hodge (2000, p. 1, tradução livre), um dos primeiros autores a utilizar a expressão,

[...] pretende abranger todos os tipos de esquemas para organizar a informação e promover a gestão do conhecimento. [...] inclui esquemas de classificação e categorização que organizam materiais em um nível geral, cabeçalhos de assunto que fornecem acesso mais detalhado e arquivos de autoridade que controlam versões variantes de informações-chave, tais como nomes geográficos e nomes pessoais. [...] também inclui vocabulários altamente estruturados, como tesouros, e esquemas menos tradicionais, tais como redes semânticas e ontologias. Como os sistemas de organização do conhecimento são mecanismos para organizar a informação, eles estão no coração de cada biblioteca, museu e arquivo.

Ater-se, contudo, apenas ao sentido que Hodge (2000) empresta ao termo, significa obliterar a compreensão de que tais SOC são organizados

a partir de concepções historicamente construídas dos fundamentos teórico-epistemológicos que lhes sustentam, isso é, que são artefatos contextualmente inseridos em domínios específicos, mesmo aqueles que demonstram vocação enciclopédica.

Desse modo, é preciso ampliar a discussão sobre os SOC e incluir, para além das análises sobre os instrumentos, em si, as concepções mais amplas que contemplem o modo como o conhecimento é socialmente organizado. Para tanto, toma-se como princípio taxonômico a divisão apresentada em Broughton *et al.* (2005) e corroborada por Hjørland (2008).

O campo da organização do conhecimento pode ser compreendido em sentido amplo e em sentido estrito. Em sentido amplo, diz respeito à divisão social do trabalho intelectual, isto é, à organização do conhecimento nas universidades e em outras instituições de ensino superior e de pesquisa e, por extensão, ao modo como as disciplinas e as profissões são estruturadas. Refere-se, portanto, ao modo como o conhecimento é organizado em diferentes domínios (Hjørland, 2008).

Em sentido estrito a organização do conhecimento preocupa-se com a construção e utilização de sistemas de organização do conhecimento, tais como os sistemas de classificação, os tesouros e as ontologias, e com a natureza e a qualidade dos processos de organização do conhecimento, tais como a classificação e a indexação (Broughton *et al.*, 2005).

Tem sido utilizada, ainda timidamente, na literatura da ISKO a sigla KOP (POC, em tradução livre para o português) para representar os processos de organização do conhecimento (*knowledge organization processes*) (Hjørland, 2016; Mazzocchi, 2017).

Os SOC, como se pode depreender das citações anteriores, apresentam duas facetas interdependentes as quais envolvem, quer seja no espectro amplo, quer seja no espectro estrito, discussões de caráter sintático, semântico e pragmático relativamente à estrutura das relações conceituais que lhes asseguram o caráter de sistemas conceituais.

Considerando-se, portanto, que os SOC, observados em sentido amplo, estabelecem relações que poder-se-ia chamar de dialéticas com os processos de organização do conhecimento, bem como com o sentido estrito de SOC, e que tanto uns quanto outros são operacionalizados a partir da definição de sistemas conceituais estruturados apresentam-se os procedimentos metodológicos e os resultados da pesquisa.

### 3. Procedimentos metodológicos

Os procedimentos metodológicos da pesquisa envolvem a adoção de dois tipos de *corpora*, sendo eles o *corpus documental* e o *corpus* de análise.

O material de análise utilizado na pesquisa foi constituído pelo conjunto dos artigos publicados em periódicos brasileiros de biblioteconomia e ciência da informação classificados com Qualis A1 ou A2 no quadriênio 2013-2017. A coleta de dados para a pesquisa cobriu o período 2000-2017.

O Qualis-Periódicos, disponível para acesso *online* (Plataforma Sucupira), é (Brasil, 2018)

um sistema usado para classificar a produção científica dos programas de pós-graduação no que se refere aos artigos publicados em periódicos científicos. [...] Como resultado, disponibiliza uma lista com a classificação dos veículos utilizados pelos programas de pós-graduação para a divulgação da sua produção.

No processo de avaliação *Qualis*, os periódicos são categorizados em conformidade com os parâmetros das áreas em que se inserem e são distribuídos nos seguintes estratos indicadores de qualidade, em ordem decrescente: A1, A2, B1, B2, B3, B4, B5 e C.

Considerando-se os critérios definidos para a pesquisa, foram selecionados seis periódicos, nomeadamente, (Qualis A1): Informação & Sociedade: estudos, Perspectivas em Ciência da Informação e Transinformação; (Qualis A2): Em Questão, Encontros Biblio e Informação & informação.

Para a seleção dos artigos que compuseram o *corpus documental*, utilizou-se como estratégia de busca, aplicada diretamente nas janelas de busca dos periódicos selecionados, as expressões “sistemas de organização do conhecimento” e “processos de organização do conhecimento” com suas variações de número (plural e singular), de idioma (inglês e espanhol) e de uso de siglas (SOC, KOS, POC, KOP).

Desse modo, 27 artigos foram selecionados para a composição do *corpus* de análise, de um total de 2.766. Os artigos selecionados distribuíram-se do seguinte modo: *Informação & Sociedade* (3), *Perspectivas em Ciência da Informação* (5), *Transinformação* (5), *Em Questão* (1), *Encontros Biblio* (6), *Informação & Informação* (7). Cumpre destacar que no periódico *Informação & Sociedade*, Qualis A1, relativamente aos anos 2001 a 2004, as recuperações foram parciais, pois não foi possível abrir alguns artigos por conta do mau funcionamento dos respectivos *links*. Registra-se, ainda, que a página do periódico *Transinformação*, também Qualis A1, não apresentava publicações relativas ao ano de 2017.

A análise foi realizada com recurso da metodologia de análise de conteúdo (Bardin, 2003). Esse tipo de análise organiza-se em três fases: a) pré-análise, b) exploração do material, c) tratamento dos resultados, inferência e interpretação. A definição das categorias de análise foi feita *a posteriori*, ou seja, não foi realizada nenhuma projeção teórica anteriormente à leitura analítica dos documentos componentes do *corpus* de análise. Desse modo, as categorias de análise utilizadas como base para as categorizações dos artigos analisados foram as seguintes: comparação, tipologia, bases teóricas, aplicações e usos.

#### 4. Resultados e discussão

A expressão SOC foi localizada em 26 artigos, aparecendo pela primeira vez em Furgeri (2006). Trata-se de um trabalho cuja base teórica é oriunda da ciência da computação e que aborda basicamente os tesouros como modelos SOC e alguns elementos para implementação do SKOS.

O termo POC, isoladamente, foi localizado em apenas um artigo, e, em co-ocorrência com SOC em outros quatro, totalizando cinco. A primeira ocorrência do termo POC no *corpus* de análise aparece em Brandt e Medeiros (2010), já alinhado ao modo como a organização do conhecimento é compreendida na literatura da ISKO (Hjørland, 2017; Mazzocchi, 2017), tomando-se os SOC e os POC como vertentes da organização do conhecimento.

Os textos que mencionam os POC (Brandt; Medeiros, 2010; Pontes; Lima, 2012; Vignoli; Souto; Cervantes, 2013; Guimarães; Milani; Evangelista, 2015; Smiraglia, 2015) não o conceituam. Apresentam apenas informações para o que poderia compor uma definição extensional, por meio da enumeração da seguinte tipologia: folksonomia, catalogação, classificação, indexação e análise de assunto.

##### 4.1. Comparação de SOC com outros instrumentos

Observou-se uma relativa confusão terminológica em relação ao conceito coberto pela designação SOC. Há, inclusive, acepções que se desviam da compreensão mais moderna sobre esse conceito. Assim, afirmar que os SOC *existem desde tempos remotos* (Monteiro; Giralde, 2008) revela inadequação ou apropriação indevida do conceito para aplicação em contexto no qual ele, na verdade, não ocorria. Tesouros e ontologias, apontados como exemplos típicos de SOC neste trabalho, por exemplo, surgem apenas no século XX, ganhando projeção mais exatamente na segunda metade desse século.

Em alguns trabalhos (Monteiro; Giralde, 2008; Boccato; Fujita, 2010; Bufrem; Gabriel Junior, 2011; Maculan; Lima, 2014; Aganette; Almeida, 2015; Souza; Bezerra, 2016) os SOC foram tomados como equivalentes de: sistemas de informação, linguagens documentárias, obras terminológicas, ontologias, catálogos ou de repositórios digitais. A variedade de comparações evidencia a relativa falta de clareza a respeito do conceito na literatura analisada.

Há, contudo, relações de equivalência que, ainda que não possam ser tomadas de modo absoluto, referem-se a conceitos mais próximos do conceito de SOC, tais como as que relacionam SOC a “sistemas de organização de conceitos” (Kobashi; Francelin, 2011) ou que destacam seu caráter eminentemente hierárquico (Barité *et al.*, 2010).

Afora as designações que poderiam, a partir de uma análise mais criteriosa, ser tomadas como equivalentes de SOC, há algumas outras que seriam mais adequadamente aplicadas como termos associados, tais como *esquemas de representação do conhecimento* (Brandt e Medeiros, 2010).

##### 4.2. Tipologia de SOC e POC

A Tabela I apresenta os tipos mais comuns de SOC enumerados na literatura analisada. Foram encontrados 49 tipos de SOC e seis de POC, números que revelam a clareza na distinção entre os dois conceitos, pois apenas *folksonomia* ocorre nos dois campos, e, ao mesmo tempo evidenciam a amplitude do conceito de SOC. Em “outras designações”, foram agrupados os tipos de SOC (*terminologia, registros bibliográficos, anéis de sinônimos, dicionário e glossário*) e POC (*catalogação, indexação, análise de assunto*) que ocorreram apenas uma vez.

Tipologia	SOC	POC
Tesouro	12	-
Sistema de classificação	8	-
Ontologia	8	-
Taxonomia	7	-
Lista de cabeçalho de assunto	4	-
Redes semânticas	2	-
Listas de autoridade	2	-
Classificação	-	2
Folksonomia	1	1
Outras designações	5	3

Tabela I. Tipologia de SOC e POC

### 4.3. Bases teóricas

Utilizou-se também como categoria de análise do *corpus* a identificação das bases teóricas dos artigos. Há uma identificação quase natural dos SOC ou POC com alguns recursos computacionais, isso é, tornou-se quase impossível falar deles sem discutir as tecnologias que lhes são associadas. Ainda assim, apenas um trabalho evidenciou essa relação (Furgeri, 2006), exatamente o primeiro artigo em que o termo SOC ocorreu.

Kobashi e Francelin (2011) destacam a ausência de explicitação das bases teóricas com que os SOC operam. Na literatura sobre SOC, entretanto, as bases teóricas são formalmente declaradas.

Assim, há conexões quase naturais entre SOC e POC com campos mais amplos e mais tradicionais, como a arquivologia, a biblioteconomia e a documentação (Barité *et al.*, 2010; Vignoli, Souto e Cervantes, 2013).

Uma das críticas mais recorrentes aos SOC mais tradicionais, tais como os sistemas de classificação mais conhecidos, relaciona-se ao seu caráter enciclopédico e universalista. A análise de domínio aparece em alguns trabalhos como fundamento teórico, o que parece indicar preocupações com a construção de SOC mais precisos em função da definição de campos conceituais mais delimitados (Bufrem e Gabriel Junior, 2011; Vignoli, Souto e Cervantes, 2013; Café, Barros e Fonseca, 2014).

Naturalmente, há variações em relação aos elementos constitutivos de cada tipo de SOC do mesmo modo como variam suas estruturas e finalidades. Como aspecto comum, contudo, os SOC apresentam um conjunto de termos e uma rede de relações semânticas e, desse modo, requerem aportes da teoria do conceito e da análise conceitual, conforme evidenciado em alguns trabalhos (Bufrem e Gabriel Junior, 2011; Vignoli, Souto e Cervantes, 2013; Café, Barros e Fonseca, 2014) e também da teoria da terminologia (Bufrem e Gabriel Junior, 2011; Café, Barros e Fonseca, 2014; Aganette e Almeida, 2015).

SOC refletem um “pensar sobre o mundo” (Barité *et al.*, 2010) e possibilitam a “organização da cultura” (Oliveira e Araújo, 2012). A preocupação com os aspectos éticos que envolvem e permeiam os POC e, por extensão, os SOC, estes tomados como instrumentos daqueles (Vignoli, Souto e Cervantes, 2013; Café, Barros e Fonseca, 2014), aparecem nos seguintes estudos: Guimarães, Milani e Evangelista (2015), Smiraglia (2015) e Quintslr *et al.* (2017).

Observando-se, principalmente, o caráter inclusivo que devem assumir, podem-se também conectar aos estudos que discutem a aplicação dos princípios éticos aos sistemas e processos de organização do conhecimento os artigos que os associam às garantias cultural e literária (Dias, 2015; Smiraglia, 2015; Barité *et al.*, 2010).

### 4.4. Aplicações e usos

Conforme era esperado e ainda de acordo com a compreensão geral a respeito dos SOC e POC como vertentes da organização do conhecimento, suas principais aplicações referem-se à organização da informação e do conhecimento visando a recuperação da informação (Monteiro; Giralde, 2008; Vignoli; Souto; Cervantes, 2013; Maculan; Lima, 2014, Berti Júnior *et al.*, 2017).

De modo geral, os conceitos de organização e recuperação da informação ou do conhecimento são tratados genericamente pelas características que os aproximam, sem especificações das peculiaridades que os distinguem formalmente. As exceções, no *corpus*, são os textos de Café e Brascher (2011) e de Brandt e Medeiros (2010), que discutem tais conceitos com base em Brascher e Café (2008).

As interpenetrações entre os espectros semânticos do conceito geral de SOC e de POC e o conceito, também genérico, de linguagem documental, evidenciam-se em alguns trabalhos (Barité *et al.*, 2010; Boccato e Fujita, 2010; Café, e Brascher, 2011; Maculan e Lima, 2014; Ribeiro, Decourt e Almeida, 2017).

Outras aplicações de SOC apontadas, que ganham destaque relativo por conta do número baixo de ocorrências, são disseminação da informação (Monteiro e Giralde, 2008), *web* semântica (Vignoli, Souto e Cervantes, 2013; Souza e Bezerra, 2016) e organização de acervos eletrônicos (Ribeiro, Decourt e Almeida, 2017).

A dispersão conceitual revelada na possibilidade de utilização dos SOC ou POC para a organização de acervos que tomam por base o suporte da informação, ou mesmo sua possível utilização para a disseminação da informação, não prejudicam a compreensão geral a respeito dos contextos de aplicação dos SOC e da sua ação instrumentalizadora em relação aos POC revelados na literatura.

De modo geral, aliás, os dados analisados demonstram que a questão sobre como/onde aplicar os SOC parece mais resolvida do que o que são exatamente os SOC.

## 5. Considerações finais

Considerando-se que a expressão sistemas de organização do conhecimento é utilizada pelo menos desde 2000, data da publicação do trabalho de Hodge, sua utilização se dá de modo bastante tardio na amostra que representa neste estudo a literatura periódica brasileira de ciência da informação, aparecendo seis anos depois no trabalho de Furgeri (2006).

Ainda não há consenso sobre a abrangência do conceito de SOC e a expressão POC ainda não é, *grosso modo*, utilizada na literatura brasileira. Na literatura internacional, contudo, principalmente no âmbito das publicações relacionadas à ISKO, tais termos (KOS e KOP) vêm ganhando popularidade.

As tipologias, que subsidiam as definições extensionais e auxiliam na compreensão dos conceitos, ainda não estão formalmente declaradas. De modo mais comum, quando apresentam tipologias, os trabalhos enumeram exemplos de SOC ou de POC, sendo os tesouros, os sistemas de classificação e as ontologias os mais frequentes. Neste trabalho buscou-se compilar o rol de exemplos apresentados visando a facilitar a organização de uma tipologia.

A análise dos conceitos em discussão neste trabalho tomou como parâmetro para a observação da literatura brasileira a literatura internacional da ISKO, a qual, contudo, não deve ser e não foi tomada como parâmetro de ordem normativa. Tendo-se em mente, contudo, a relevância da instituição no cenário internacional da OC, acredita-se que tal literatura apresenta plenas condições de ser observada como orientadora e, ao mesmo, tempo como catalisadora dos discursos sobre OC, SOC e POC.

Por fim, cabe ressaltar que a análise realizada traça um percurso semasiológico, isso é, parte do termo para a identificação do conceito e é clivada por essa restrição. Considerando-se a imprecisão terminológica dos conceitos de SOC e, mais acentadamente, de POC, como se defendeu neste artigo, entende-se como perfeitamente possível a identificação do conceitos (conteúdos) analisados designados por outros termos (expressões). Resulta, portanto, como proposta para trabalhos futuros o aprofundamento da análise conceitual neste sentido.

## Referências

Aganette, E. C.; Almeida, M. B. (2015). Avanços teóricos na comparação entre Teoria da Ontologia Aplicada e Teoria da terminologia. // *Encontros Bibli.* 20:44 (Setembro/Dezembro 2015) 03-24.

Bardin, L. (2003). *Análise de conteúdo*. Lisboa: Edições 70, 2003.

Barité Roqueta, M. G. (2011). Sistemas de organización del conocimiento: una tipología actualizada. // *Información & Información.* 16:esp (Janeiro/Junho 2011) 122-139.

Barité, M. G.; et al. (2010). Garantia literária elementos para uma revisão crítica após um século. // *TransInformação.* 22:2 (Maio/Agosto 2010) 123-138.

Barité, M. G.; et al. (2013). *Diccionario de organización del conocimiento: clasificación, indización, terminología*. Montevideo: PRODIC, 2013.

Benveniste, É. (1989). *Problemas de linguística geral II*. Campinas: Pontes, 1989.

Berti Junior, D. W. et al. (2017). Semiautomação de relações em tesouros: uma proposta para o refinamento de relacionamentos semânticos a partir do tesouro Agrovoc. // *Informação & Informação.* 22:3 (December 2017) 377-404.

Bocato, V. R. C.; Fujita, M. S. L. (2010). O uso de linguagem documental em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal. // *Perspectivas em Ciência da Informação.* 15:3 (Setembro/Dezembro 2010) 23-51.

Brandt, M.; Medeiros, M. B. B. (2010). Folksonomia: esquema de Representação do Conhecimento?. // *TransInformação.* 22:2 (Maio/Agosto 2010) 111-121.

Brascher, M.; Café L. (2008). *Organização da informação ou organização do conhecimento*. // *Encontro Nacional da pesquisa em Ciência da Informação (2008)*. São Paulo: USP, 2008. 01-14.

Brascher, M.; Carlan, E. (2010). *Sistemas de organização do conhecimento: antigas e novas linguagens*. // Robredo, J.; Brascher, M (2010). *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: IBICT; Brasília: IBICT, 2010. ISBN: 978-85-7013-072-3. 147-176.

Brasil (2018). Ministério da Educação. Fundação Capes. *Classificação da produção intelectual*. 2018. <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/classificacao-da-producao-intelectual> (2008-04-03).

Broughton, V.; et al. (2005). *Knowledge organization*. // Kajberg, L.; Loring, L. (2005). *European curriculum reflections on library and information science education*. Copenhagen: The Royal School of Library and Information Science; Copenhagen: The Royal School of Library and Information Science, 2005. 133-148.

Buřem, L. S.; Gabriel Junior, R. F. (2011). A apropriação do Conceito como objeto na literatura periódica científica em Ciência da Informação. // *Informação & Informação.* 16:3 (Janeiro/Junho 2011) 52-91.

Café, L. M. A.; Barros, C. M.; Fonseca, M. S. (2014). *Organização do Conhecimento: uma análise conceitual nos anais do ENANCIB*. // *Em Questão.* 20:1 (Janeiro/Junho) 86-112.

Carlan, E. (2010). *Sistemas de organização do conhecimento: uma reflexão no contexto da Ciência da Informação*. Brasília: UnB, 2010. *Dissertação (Mestrado em Ciência da Informação)*.

Carlan, E.; Medeiros, M. B. B. (2011). *Sistemas de organização do conhecimento na visão da ciência da informação*. // *Revista Ibero-americana de Ciência da Informação.* 4:2 (Agosto) 53-73.

Dahlberg, I. (2009). Brief communication: concepts and terms: ISKO's major challenge. // *Knowledge Organization.* 36:2/3 (2009) 169-177.

Dahlberg, I. (1978). *Teoria do conceito*. Ciência da informação. 7:2, 101-107.

Dias, C. C. A (2015). *análise de domínio, as comunidades discursivas, a garantia de literatura e outras garantias*. // *Informação & Sociedade: estudos.* 25:2, (Maio/Agosto 2015) 6-17, 2015.

- Furgeri, S. (2006). O papel das linguagens de marcação para a Ciência da Informação. // *TransInformação*. 18:3 (Setembro/ Dezembro 2006) 225-239.
- Guimarães, J. A. C.; Milani, S. O.; Evangelista, I. V. (2015). Valores éticos na organização e representação do conhecimento. // *Encontros Bibli*. 20:esp.1 (Janeiro 2015) 19-32.
- Hjørland, B. (2008). What is knowledge organization (KO)?. // *Knowledge Organization*. 35:2/3 (2008) 86-101.
- Hjørland, B. (2015). Theories are knowledge organizing systems (KOS). // *Knowledge Organization*. 42:2 (2015) 113-128.
- Hjørland, B. (2016). Knowledge organization. // *Knowledge Organization*. 43:6 (2016) 475-484.
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: Council on Library and Information Resources, 2000.
- Kobashi, N. Y.; Francelin, M. M. (2011). Conceitos, Categorias e Organização do Conhecimento. // *Informação & Informação*. 16:2 (Dezembro 2011) 01-24.
- Lara, M. L. G. (2004). Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documental. // *Ciência da Informação*, 33:2, (Maio/Agosto 2004) 91-96.
- Lara, M. L. G. (2015). Propostas de tipologias de KOS: uma análise das referências de formas dominantes na organização do conhecimento. // *Encontros Bibli*. 20: esp. 1 (Fevereiro 2015) 89-107.
- Maculan, B. C.; Lima, G. A. B. O. (2014). Relacionamentos em tesouros: o valor semântico dos verbos. // *Perspectivas em Ciência da Informação*. 19:4 (Dezembro 2014) 182-201.
- Mazzocchi, F. (2017a). Knowledge organization systems (KOS). // Hjørland, B.; Gnoli, C. (2017). *Encyclopedia of knowledge organization*. (2017a). <http://www.isko.org/cyclo/kos> (2018-03-26).
- Mazzocchi, F. (2017b). Relations in KOS: is it possible to couple a common nature with different roles?. // *Journal of Documentation*. 73:2 (2017b) 368-383.
- Monteiro, S. D.; Giraldes, M. J. C. (2008). Aspectos lógico-filosóficos da Organização do Conhecimento na esfera da Ciência da Informação. // *Informação & Sociedade*. 18:3 (Setembro/Dezembro 2008) 13-27.
- Oliveira, D. A.; Araújo, R. F. (2012). Construção de linguagens documentárias em sistemas de recuperação da informação: a importância da garantia do usuário. // *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*. 17:34 (2012) 17-30.
- Plataforma Sucupira. <https://sucupira.capes.gov.br/sucupira/#> (07 abr. 2018).
- Pontes, F. V.; Lima, G. A. B. O. (2012). A organização do conhecimento em ambientes digitais: aplicação da teoria da classificação facetada. // *Perspectivas em Ciência da Informação*. 17:4 (Dezembro 2012) 18-40.
- Quintslr, M. M. M.; et al. (2017). Visibilidade social de indivíduos transgênero e sistemas de organização do conhecimento. // *Informação & Informação*. 22:2 (Outubro 2017) 235-264.
- Ribeiro, A. R. P.; Decourt, B.; Almeida, T. (2017). A representação do domínio "gênero" no âmbito das linguagens documentárias: um mapeamento conceitual em instrumentos terminológicos. // *Informação & Informação*. 22:2 (Outubro 2017) 208-234.
- Sager, J. C. (1993). *Prólogo: la terminología: puente entre varios mundos*. // Cabré, M. T. (2013). *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida/Empúries, 1993.
- Smiraglia, R. P. (2015). Ética na Organização do Conhecimento: aplicação em duas conferências para apontar um novo núcleo de Domínio. // *Encontros Bibli*. 20:esp.1 (Fevereiro 2015) 01-18.
- Souza, E. G.; Bezerra, D. A. (2016). Os Functional Requirements for Bibliographic Records no contexto da Web Semântica: as contribuições de Paul Otlet. // *Transinformação*. 28:2 (Maio/Agosto 2016) 143-157.
- Souza, R. R.; Tudhope, D.; Almeida, M. B. (2012). Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. // *Knowledge Organization*. 39:3 (2012) 179-192.
- Vignoli, R. G.; Souto, D. V. B.; Cervantes, B. M. N. (2013). // *Informação & Sociedade*. 23:2 (Maio/Agosto 2013) 53-72.
- Zeng, M. L. (2008). Knowledge organization systems (KOS). // *Knowledge Organization*. 35:2/3 (2008) 160-182.

---

Enviado: 2018-07-28. Segunda versão: 2015-0-  
Aceptado: 2013-0-

---



---

# Estudio sobre la indización/etiquetado y los lenguajes documentales en cinco diarios españoles

*Study on indexing/tagging and controlled vocabularies in five Spanish newspapers*

---

**Antonio GARCÍA-JIMÉNEZ (1) David RODRÍGUEZ-MATEOS (2) Beatriz CATALINA-GARCÍA (3)**

(1) Facultad de Ciencias de la Comunicación, Universidad Rey Juan Carlos, c/ Camino del Molino, s/n, Campus de Fuenlabrada, 28943; antonio.garcia@urjc.es (2) Departamento de Periodismo y Comunicación Audiovisual, Universidad Carlos III de Madrid, c/Madrid, 128, 28903 Getafe (Madrid), david.rodriguez@uc3m.es (3) Facultad de Ciencias de la Comunicación, Universidad Rey Juan Carlos; beatriz.catalina@urjc.es

## Resumen

Esta investigación parte de la preocupación sobre el modo en el que se organiza el conocimiento en los diarios españoles. Tiene un doble objetivo: por un lado, determinar cómo se desarrolla el análisis, indización y etiquetado de los contenidos periodísticos. El segundo es la delimitación de la existencia, características y aplicación operativa de vocabularios/lenguajes documentales. Se ha remitido un cuestionario a un profesional especializado de los siguientes medios: El País, El Mundo, ABC, La Vanguardia y Heraldo de Aragón. Se observa divergencias en las tareas de indización y etiquetado, la conexión con técnicas SEO, así como la continuación en el empleo de lenguajes documentales, si bien no en todos los casos.

**Palabras clave:** Documentación informativa. Etiquetado. Indización. Periodismo. Medios. Lenguajes documentales. España.

## Abstract

This research comes from a major issue, knowledge organization at Spanish newspapers, with two main aims. First, the way its contents are analysed, indexed and tagged. Second, the use of controlled vocabularies, its presence, main features and implementation is studied. A survey was conducted among some news librarians from El País, El Mundo, ABC, La Vanguardia and Heraldo de Aragón. Some indexing and tagging variations were found, and there is a relation to SEO techniques. Controlled vocabularies are also applied, although there are some differences among those media.

**Keywords:** Media libraries. Tagging. Indexing. Journalism. Media. Controlled vocabularies. Spain.

## 1. Introducción y objetivos

En el contexto de los cambios vertiginosos y de calado que se están produciendo en el ámbito mediático-periodístico y, por ende, en lo que tiene que ver con la documentación informativa, esta investigación se preocupa por los procedimientos e instrumentos con los que se describen los contenidos periodísticos para su posterior enlazado y/o recuperación al objeto de contribuir a las tareas informativas (García Jiménez, 2016).

Dos son los objetivos específicos. El primero es determinar cómo se aplica en los diarios digitales españoles los procedimientos de análisis, indización y etiquetado de los contenidos periodísticos. El segundo es la delimitación de la existencia, características y aplicación operativa de vocabularios/lenguajes documentales. Conectado a estos objetivos, también son de interés el marco profesional de las funciones documentales de los diarios españoles. Este es un trabajo (1) de naturaleza aplicada y a medio camino entre el ámbito

de la Documentación Informativa y el de la Organización del Conocimiento. Aunque entendemos que sería pertinente en este caso, no se planteará ninguna reflexión conceptual sobre el significado, y su correspondiente transformación (lo que daría muestra de un cambio de paradigma) de algunos términos. El ejemplo más claro lo tenemos en las interacciones, solapamientos y fugas existentes entre “análisis documental” y “etiquetado”. Tampoco se hará una valoración crítica de las consecuencias de naturaleza epistemológica e ideológica de las actividades convertidas aquí en objeto de estudio.

## 2. Estado de la cuestión

### 2.1. Situación del periodismo

Es comúnmente aceptado que el periodismo vive un momento convulso. Y esto se debe a la casi absoluta digitalización de sus diferentes procesos. Un dominio en continua interacción con los

medios o redes sociales (Fondevila, 2017), donde los procedimientos y técnicas SEO tienen un peso relevante (Codina et al., 2017), lo que influye en la actividad periodística, y por ende, documental. Lo que también se trasluce en los términos-etiquetas y lenguajes con los que se organiza el conocimiento.

Otra de las características del periodismo moderno tiene que ver con los crecientes procesos de automatización y “algoritmización” (Saad y Bertocchi, 2012; Jung et al., 2017). De hecho, es cada vez mayor la aportación del Big Data (Stone, 2014; Renó y Renó, 2017), por ejemplo, en términos de análisis automatizado de titulares, adaptación a las audiencias, segmentación y personalización de contenidos, seguimiento de noticias en tiempo real o generación de relatos. Unido a este fenómeno, se añade la mayor presencia de una corriente periodística que tiene una gran conexión con la Documentación, concretamente el periodismo de datos (López-García et al., 2016).

## 2.2. Documentación Informativa

El ámbito primario donde se ubica este trabajo es el de la Documentación Informativa. En este sentido, son varias las tendencias que se advierten en la literatura especializada, con datos complementarios y en ocasiones, contradictorios. En el trabajo centrado en Andalucía de Meléndez-Malavé y Hirschfeld-Suárez (2016), se observa que no todos los medios gozan de un centro de documentación. Se podría hablar incluso de una lenta, aunque progresiva, desaparición de los mismos. Asimismo, es destacable el descenso del número de documentalistas que trabajan en los medios, lo que podría vincularse a la doble crisis tanto económica como la relativa al propio del negocio de la prensa. De acuerdo con este estudio, la información más requerida por los periodistas se basa en noticias y fotos de archivo, especialmente vinculadas al ámbito local.

Por su parte, en la investigación de Marcos-Recio y Edo (2015) se afirma que un 33% de los periodistas españoles ven el servicio de documentación como imprescindible y el 50% como necesario. Al mismo tiempo, se detectan porcentajes relevantes de periodistas que usan de forma intensiva el centro de documentación, si bien son los recursos electrónicos los que se constituyen como las fuentes con un mayor índice de consulta. En este sentido, el estudio de Pintado (2013) confirma que habitualmente los periodistas solo se documentan cuando tienen dudas sobre un determinado aspecto de la información que están elaborando, y no aluden a esta actividad como un modo genérico de comportamiento. Destaca el hecho de que el uso de fuentes, que

en cualquier caso son mayoritariamente institucionales y electrónicas, se diferencia por el género del texto y por la temática tratada.

## 2.3. Etiquetado y lenguaje

Los continuos cambios que tienen lugar en los modos de producción y difusión de la información periodística tienen sus consecuencias en la organización del conocimiento. Influyen en la asignación terminológica correspondiente y en los instrumentos que la gestionan, y que vinculamos al subcampo teórico que se ocupa de herramientas como las clasificaciones, los tesauros, las folksonomías y, más tangencialmente, las ontologías.

### 2.3.1. Vocabularios y lenguajes documentales

Uno de los primeros aspectos que se tendrá en cuenta en esta investigación es si, por un lado, el medio analizado emplea un lenguaje/vocabulario, así como su naturaleza y sus características. En primer lugar, nos parece útil, de acuerdo con Soler y Gil (2010), y conforme a unas categorías válidas para nuestra investigación, la determinación del nivel de complejidad de los lenguajes a los que aquí hacemos referencia: de menor a mayor, las folksonomías, las taxonomías, los tesauros y las ontologías.

En este sentido, resulta relevante la defensa por parte de autores como Yedid (2013) de las folksonomías. Desde su punto de vista, no existen grandes diferencias en relación a los términos-etiquetas de otros vocabularios controlados, a pesar del mayor grado de inconsistencias existentes en el uso del plural-singular. Asimismo, desde su perspectiva, los términos en uso están más relacionados con el vocabulario que se emplea realmente por parte de los usuarios, lo que sugeriría una mejora en la recuperación de información, puesto que se deriva de un proceso de selección-elaboración más intuitivo. Junto al menor costo que supone, hay quien aduce, a su vez, una mayor democratización en el sentido de uso de un conocimiento compartido y colaborativo, frente a una visión más jerárquica. En cualquier caso, el debate sigue abierto (du Preez, 2015).

Inicialmente, desechamos el uso real y operativo de instrumentos de gran complejidad. Es el caso de las ontologías. Frente a estas, y en un plano puramente documental, diversos autores (García Jiménez, 2004; Codina y Pedraza, 2011; Mendes et al., 2015) apuntan la validez de los tesauros, demostrada en una trayectoria ya larga. Precisamente, desde una mirada vinculada a las clasificaciones documentales, Szostak (2014) hace hincapié en las dificultades que encuentran las ontologías para su aplicación documental con-

creta: por un lado, por la pobre concepción terminológica subyacente, y por otro, sus diferentes planteamientos y bases, lo que complica en exceso sus opciones de interoperabilidad.

Además, en gran medida gracias a SKOS (Simple Knowledge Organization System), el tesoro ya goza de opciones claras de formalización, aunque condicione su estructura relacional en el contexto de la web semántica (Sánchez, Colmenero y Moreiro, 2012). De hecho, tal y como señalan Pastor et al. (2012), al analizar el funcionamiento de diversos lenguajes controlados que incorporan SKOS, los tesauros, junto a las clasificaciones, obtienen buenos resultados, salvo en lo que se refiere a la interoperabilidad y la interconexión entre ellos mismos.

Si bien hasta el momento no se han descrito procesos de automatización en los instrumentos empleados en prensa, en cualquier caso, en una revisión de este tipo, se tendrá que tomar en consideración los gestores de tesauros. Precisamente, Martínez y Alvite (2014) aplican un método para la evaluación de este tipo de instrumentos. Algunos de los puntos de análisis son: el propósito, la disposición en términos de integridad y coherencia, así como los elementos asociados a la interoperabilidad, integración y a la compatibilidad con los estándares propios de la web semántica (RDF/SKOS).

### 2.3.2. Análisis, etiquetado y formalización

Son escasas las investigaciones centradas en el análisis-etiquetado humano. En este sentido, podemos acudir a Pérez et al. (2014) cuando hace mención al etiquetado social, que hace posible que los términos sean más concretos y están más pegados a los intereses de los usuarios. Al mismo tiempo, se adapta más a los ámbitos colaborativos. Incluso se vinculan con comunidades de usuarios concretas. Por su parte, sí encontramos algunos textos relevantes sobre la formalización del análisis y marcado semántico. Así, la investigación de Pastor (2013) se centra en la situación y aplicabilidad de Schema.org, los datos estructurados en Xhtml, los microdatos, microformatos o el RDFa.

## 2.4. Etiquetado, indización y lenguajes en el periodismo

En primer lugar, acudimos al trabajo realizado por Rubio Lacoba (2012) que analiza lo que podríamos definir como folksonomía controlada del diario *El País*. En efecto, se trata de una elaboración que tiene como fuente fundamental la indexación social realizada sobre los productos periodísticos. Una tarea iniciada por los periodistas

cuando incorporan las etiquetas, que son filtradas posteriormente por los documentalistas con vistas a evitar la sinonimia y la polisemia, y para mejorar las relaciones entre las citadas etiquetas. En cualquier caso, son procesos que se realizan tomando en consideración las técnicas SEO.

Se trata de un lenguaje que se ha alimentado de las bases de datos propias y de tesauros especializados. Su estructura radica en diferentes áreas: temas, personajes, organizaciones, lugares y eventos. También hay que contar con la conexión automatizada de las noticias con las etiquetas o temas. Finalmente, este lenguaje está vinculado a un editor que gestiona: la solicitud e incorporación de etiquetas, las acciones de desambiguación alrededor de fechas, cargos y siglas, así como el establecimiento de etiquetas, que no se situarán en el lenguaje principal pero que pertenecerán a lo que denominan "conceptos editoriales".

En un plano más teórico, destacan varias líneas de trabajo vinculadas con el periodismo. Así, Baños (2013) aborda las posibilidades que ofrecen las noticias de divulgación científica en lo que se refiere a la actualización de los tesauros. Aunque muestran algunas limitaciones, sí se sugiere su capacidad para colaborar con la actualización de estas herramientas, fundamentalmente como "yacimientos" de términos. Dentro también del ámbito periodístico, el estudio de Baños et al. (2015) revisa otro espacio relevante: los modelos de representación de los metadatos. Fundamentalmente destaca la ausencia de "uniformidad" lo que dificulta que se aplique realmente la interoperabilidad. En este trabajo se advierte que, frente a la mayor dedicación en la literatura especializada sobre NewsML y NITF, en el campo profesional predomina en los códigos fuente schema.org junto a los esquemas, Twitter Cards y Open Graph Protocol (en Facebook).

Con un planteamiento más especulativo, aunque con orientación operativa mantenida en el tiempo, García Gutiérrez (2011) propone las exomemorias. Con una base crítica, transcultural y participativa, intenta superar la investigación descriptiva y desprovista de una perspectiva positiva y estereotipada. A su vez, (García Gutiérrez, 2014) desarrolla un modelo de análisis de textos periodísticos, para su posterior recuperación, que se centra en los siguientes puntos: acción (a través de una macroproposición global), sujetos, objetos, elementos asociativos, situación, causa, finalidad, consecuencia, modo, instrumento, lugar y tiempo. No deja de sorprender que sea de los pocos investigadores preocupados por los "matices" ideológicos y el "impacto" en la sociedad de este tipo de tareas, más si cabe cuando nos referimos al dominio periodístico.

Finalmente, traemos a colación un estudio (Søbak y Pharo, 2017) que, aunque centrado en televisión, aborda el etiquetado-indización en una cadena, en este caso la Norwegian Broadcasting Corporation (NRK). En su dinámica de trabajo, y aunque no se cuenta con un vocabulario controlado, los indizadores tienen acceso a las etiquetas que han sido previamente usadas. Uno de los hallazgos más relevantes es el referido a la influencia de la indización descentralizada, tanto en los porcentajes de temas y nombres de personas o instituciones representados como en la variación de los niveles de exhaustividad, o la presencia de un bajo nivel de cobertura, en gran medida a partir de las diferencias de prácticas profesionales (con diferentes niveles de formación) de etiquetado encontradas.

### 3. Metodología

Esta investigación presenta un análisis preliminar, en el marco de la actividad documental en algunos de los principales medios periodísticos españoles en versión digital (tanto si tienen además versión impresa como si no), el uso de vocabularios/lenguajes y del análisis/descripción documental o etiquetado en relación a las noticias que publican. Por una parte, se pretende delimitar si el etiquetado es realizado con fines documentales, esto es, si ha sido empleado para describir cada contenido periodístico (noticia, reportaje, artículo de opinión, etc.) con el fin de facilitar la relación y asociación entre distintos contenidos, así como su recuperación en un futuro. Ello supone previamente averiguar si en estos medios existen profesionales de la documentación y, a continuación, saber si estos profesionales participan en la definición de los textos empleados como etiquetas. Es también posible que ese etiquetado sea realizado por otro tipo de profesionales, ya sean los mismos periodistas que elaboran los contenidos, o bien especialistas en la creación de términos que sirvan para destacar la presencia de estos medios, a corto plazo, en otros recursos de Internet, tanto en buscadores como en distintas redes sociales, mediante técnicas como el search engine optimization, o SEO (Carroll, 2010).

Se ha pretendido revisar, de acuerdo con la información ofrecida por estos medios, quién y cómo define esta descripción mediante etiquetas: si se plantea la generación de esos descriptores basados en lenguajes documentales, y de qué tipos, o bien, si se realiza mediante lenguaje libre; cómo se evalúa la calidad de ese etiquetado y su efectividad; y además, qué profesionales generan, de hecho, esos contenidos.

En relación a los vocabularios/lenguajes documentales empleados, se analizan los siguientes

aspectos: su existencia, su nivel de control terminológico, el número de términos existente, su estructura, las fuentes que lo alimentan, las interacciones y conexiones con otros instrumentos, su base tecnológica, los modos de integración en el plano documental y en la redacción, su impacto en la recuperación de información, los profesionales que se ocupan de ellos, los procesos de actualización, y el papel de los protocolos de evaluación y determinación de su calidad.

Para cumplir con los objetivos previamente marcados, se ha empleado una doble estrategia: inicialmente, se ha tratado de localizar y contactar, para cada medio analizado, a algún miembro del mismo que pudiera estar vinculado o participar en el etiquetado de sus contenidos. El contacto se ha realizado vía telefónica, a través del correo electrónico, o mediante redes sociales (como LinkedIn). Se ha pretendido obtener a algún profesional de la documentación dentro del medio o, secundariamente, a algún otro posible implicado en la generación de etiquetas (sea un responsable de la redacción, un experto en SEO, etc.). En aquellos casos en los que se ha conseguido ese contacto, se ha realizado una encuesta detallada, mediante Google Docs, empleando una combinación de preguntas tanto cerradas como abiertas, con el fin de amplificar las respuestas. La extracción de datos de ese cuestionario se presentan en el apartado de resultados.

La base de trabajo ha estado compuesta por los 15 periódicos españoles de mayor audiencia, medida en el número de visitas a su versión digital, de acuerdo con los baremos oficiales aceptados por los propios medios en los meses de julio, agosto y septiembre de 2017, de acuerdo con la empresa Comscore. Dado que esta compañía solo hace públicos sus datos a sus asociados, se han consultado las referencias publicadas en uno de los periódicos analizados (OK Diario 2017a, 2017b, 2017c). Se han tomado datos de tres meses, dado que algunos de los medios no aparecían en todas las mediciones, por lo que se han dado por válidos cuando aparecían al menos en dos de los meses analizados. Durante la investigación, no ha sido posible contar con respuestas concretas de muchos de estos medios. Sí que han respondido a esta cuestión cuatro de los diarios con mayor difusión digital: *El Mundo*, *El País*, *La Vanguardia* y *ABC*, en los cuales existe, aún, un servicio de documentación como tal. En cuanto al quinto medio que respondió, 20 Minutos, no consta que mantenga ya dicho servicio, aunque sí lo hace el que es ahora el diario matriz de su grupo, *Heraldo de Aragón*, a quien se le ha aplicado también la encuesta.

	<i>El País</i>	<i>ABC</i>	<i>El Mundo</i>	<i>La Vanguardia</i>	<i>Heraldo de Aragón</i>
Utilización	Sí	Sí	No	No	Sí
Control de términos	Sí	Sí			Sí
Estructura	Colaborario	Tesoro		Taxonomía	Tesoro
Fuentes	Antiguas del grupo Prisa. Estándar IPTC. Tesauros especializados: CSIC y CINDOC.	Clasificación propia. Tesauros externos referenciales		www.classora.com	Tesauros UNESCO
Nº términos	Más de 6000	Más de 6000		Menos de 1000	Más de 6000
Impacto en la recuperación 1-5	5	4		3	4
Fórmulas para desambiguar	Notas aclaratorias Nombres normalizados (URI)	Relaciones tipo		No se emplean	Notas aclaratorias
Relaciones semánticas entre términos	Específicos y generales	Específicos y generales		Específicos y generales	No hay relaciones
Tipo de categorización	Personas Lugares Organizaciones Eventos Temas	Por asignación Basado en metadatos		Ciudades Ubicaciones Equipos Nombres propios, etc	Facetada
Profesionales encargados	Documentalistas	Documentalistas		Documentalistas y programadores informáticos	Documentalistas y programadores informáticos
Frecuencia de actualización	Menos de un mes	Menos de un mes		Menos de un mes	Mensualmente
Conexión con otros lenguajes	RR. internas con Enciclopedia (PRISA Radio) Futuro: Linked Data	Etiquetado de contenidos de página web		No hay conexión	No hay conexión
Modelo-Base para el lenguaje	Procedente del Colaborario con metas DC y Schema	UNE 50106:1900		No hay modelo	No hay modelo
Evaluación funcionamiento y calidad	Humana Automática	Humana		No se evalúa	No se evalúa
Coste		10% capital humano		No disponible	Actualmente mínimo
Integración en actividad cotidiana	Publicación diaria de contenidos. Arquitectura de información Propagación en RR.SS. Estrategia para negocio online.	Catalogación de registros		Creación de temas para mejorar el enlazado SEO	Imprescindible para búsquedas de envergadura
Elementos para marcado semántico	Estrategia aplicada por equipo SEO	No registrado	Palabras claves en el resumen	No registrado	Resumen documental y descriptores (papel). Taxonomía SEO (web)
Contenidos etiquetados	Los publicados por el medio en su web	Los publicados por el medio en su web	Contenidos en papel	Los publicados por el medio en su web	Los publicados por el medio en su web
Etiquetas en medios y redes sociales	No emplean redes	No emplean redes	No emplean redes	No emplean redes	No emplean redes
Responsables de etiquetado	Periodistas Especialistas en medios sociales	Periodistas Documentalistas	Documentalistas	Herramienta de Classora	Periodistas Documentalistas
Relación etiquetado-lenguaje documental	El Colaborario genera una página optimizada por equipo SEO. El periodista tiene posibilidad de enlazar contenidos relacionados- Algoritmos para establecer potenciales intereses del lector	El etiquetado se desarrolla a partir del lenguaje	Ninguna	Ninguna	Ninguna
Modo de evaluar el etiquetado	Estadísticas internas Herramientas propias del equipo SEO	Indicadores de posicionamiento web	No	Prevalecen términos genéricos para que no dupliquen contenidos clasificados y carezcan de valor SEO	No

*Tabla I. Resultados de la encuesta*

## 4. Resultados

### 4.1. Los centros de documentación

En todos los diarios analizados se mantiene la figura del documentalista que, según los casos, desempeña su labor junto a profesionales de otros ámbitos; principalmente programadores y/o informáticos que están presentes en tres de los centros; además de un periodista, un diseñador gráfico, un especialista en medios sociales y, finalmente, un responsable de métrica web.

En la mayoría de los casos, no son más de tres personas las encargadas de este cometido; solo en uno de los centros (*El País*) trabajan más de 5 profesionales que, atendiendo al volumen de información generada y recibida por los periódicos analizados, suponen una cantidad adecuada para la gestión documental en un medio de comunicación. La duda que surge en este punto es la relación entre número de profesionales y el desempeño de las funciones correspondientes.

La principal tarea encomendada se relaciona, en todos los casos, con la gestión y actualización del fondo documental. Y en cuatro de los medios el trabajo documental también se orienta a la búsqueda de información y el etiquetado de documentos. El asesoramiento en la recuperación de información a periodistas, el diseño y mantenimiento de lenguajes y la elaboración de productos documentales y/o periodísticos son tareas que se desempeñan en tres de estos centros. Solo uno de ellos dedica parte de su trabajo a la digitalización y conservación del fondo antiguo. Este último resultado sugiere que, o bien ya tienen todo el material analógico reproducido en los nuevos formatos, o bien que estos centros carecen de un archivo antiguo porque su creación se produjo en la era digital.

Cuatro de los cinco periódicos encuestados afirman emplear una base de datos para el desarrollo de su trabajo cotidiano. Dos de ellos (*El País* y *El Heraldo de Aragón*) mantienen un sistema de gestión propio, aunque el utilizado por el periódico nacional es compartido con todos los medios que conforman el Grupo PRISA. En *ABC* y *El Mundo* se utiliza el archivo digital multimedia *Quay*, y en este último diario se combina con otras bases como la plataforma diseñada por la Agencia EFE (*EFE Data*), *Informa* y la estadounidense *Factiva*.

### 4.2. Vocabularios/lenguajes documentales

Tanto *El Mundo* como *La Vanguardia* reconocen en el cuestionario no disponer de un tipo de lenguaje documental establecido. No obstante, por las respuestas recibidas de este último medio, entendemos que emplean un lenguaje que no es

propio, o bien que no tiene un control documental real, y que puede basarse en una mera clasificación temática que reutilizan en algunas ocasiones. En principio, se han aceptado las respuestas en este sentido, si bien este punto necesitaría de una ulterior revisión. El resto, *El País* y *ABC*, en el ámbito nacional, y *El Heraldo de Aragón* (aquí y en adelante), en el regional, aplican en todos ellos un lenguaje controlado.

Se utilizan diferentes tipos de lenguajes: taxonomía en *La Vanguardia*, tesauros aplicados por *ABC* y *Heraldo de Aragón*, y en *El País* se maneja lo que en la actualidad denominan Colabulario, similar a los tesauros pero con una adaptación mayor a los entornos digitales.

La diversidad que caracteriza a los periódicos estudiados se manifiesta también en las fuentes que utilizan para el enriquecimiento de su lenguaje documental. *El País* y *ABC* recurren a una mayor diferenciación de procedencias, mientras que los diarios de ámbito regional se ciñen respectivamente a una sola fuente: el catalán se sustenta en la base digital y abierta de Classora (<http://www.classora.com/>), y el aragonés se apoya en tesauros de la UNESCO. Igualmente se encuentran diferencias en función del ámbito que cubren los periódicos en lo que respecta a los profesionales encargados de estos lenguajes: si bien en todos los diarios esta labor es ejercida por documentalistas, en el caso de *La Vanguardia* y *El Heraldo de Aragón* se incorporan también a este cometido programadores informáticos.

Se observan divergencias también entre las distintas coberturas geográficas: en los regionales no se establece conexión con otros lenguajes, no existe un modelo que sirva como base y tampoco se aplican protocolos de evaluación y calidad del lenguaje empleado. En el caso de *El País* se recurre nuevamente al Colabulario para establecer un modelo del lenguaje, y para las conexiones lingüísticas se desarrollan relaciones internas con la Enciclopedia (procedente de PRISA Radio). *ABC* conecta con el etiquetado de contenidos en la página web, y el modelo-base que utiliza es el establecido por la Asociación Española de Normalización, UNE. En ambos periódicos de ámbito nacional son los responsables del centro quienes evalúan y supervisan el funcionamiento de los lenguajes, aunque en *El País* se apoyan también en sistemas automáticos.

Otras cuestiones puntuales que marcan la diferencia entre unos periódicos y otros se derivan del número de términos empleados. Todos manejan una cantidad superior a los 6.000 a excepción de *La Vanguardia*, cuyo número no alcanza el millar (y tampoco este periódico aplica fórmulas para eliminar ambigüedades terminológicas). Por su parte,

*El Heraldo de Aragón* no marca relaciones semánticas entre términos, mientras que los otros tres establecen tanto conexiones específicas como generales. En todos ellos se pretende organizar los términos alrededor de categorías, aunque no coincidan entre sí: por ejemplo, “personas”, “ciudades”, etc. Igualmente, es el periódico aragonés el que registra diferencias en cuanto a la actualización del lenguaje, que desarrolla este cometido una vez al mes, mientras que el resto de periódicos lo hace con una mayor frecuencia.

Por otra parte, los profesionales que cumplieron el cuestionario perciben un importante impacto de los vocabularios/lenguajes en la recuperación de información del periódico. En una escala de grado creciente entre el 1 y 5, el único que presenta una visión más modesta (3) es *La Vanguardia*; mientras que en *El Heraldo de Aragón* y *ABC* entienden que el impacto es notable (4) y muy grande (5) en *El País*. Estos índices se relacionan también con la integración en la actividad cotidiana de cada periódico analizado. Al respecto, el centro de *Prisa* utiliza su lenguaje documental para múltiples cometidos: la publicación diaria de contenidos, el diseño de la arquitectura de la información, la difusión en redes digitales e incluso como un elemento más de la estrategia global para el negocio online. En el resto de medios se reconoce también un cierto nivel de integración, aunque en menor medida y en aspectos más puntuales: en *ABC* para la catalogación de registros; como vía para crear temas que mejoren el enlazado SEO dentro de *La Vanguardia*; y, finalmente, la incorporación de lenguajes es clave en *El Heraldo de Aragón* para búsquedas con cierta envergadura.

#### 4.3. Procedimientos de análisis y etiquetado

La única coincidencia detectada en todos los periódicos es la ausencia generalizada de control documental sobre el etiquetado de sus contenidos presentados en medios y redes sociales. En el resto de ítems establecidos se observan diferencias entre todos los centros analizados. Igualmente, en función de la cobertura geográfica, no existe una relación significativa en ningún aspecto. Como punto preliminar, el marcado semántico de *El País* y *El Heraldo de Aragón* es aplicado en el entorno de la estrategia del equipo SEO, aunque en el formato papel, y el periódico aragonés coincide con *El Mundo* al utilizar los descriptores incluidos en los resúmenes. Este diario se diferencia del resto en la selección de contenidos que son etiquetados, correspondientes a los incluidos en la versión impresa, mientras que los otros diarios se centran en los publicados en su web.

Los responsables de este cometido son documentalistas en el caso de *El Mundo*, *ABC* y *El Heraldo de Aragón*; aunque en estos dos últimos también participan periodistas. Precisamente son estos profesionales, junto a los especialistas en medios sociales, los encargados de esta tarea en *El País*. En *La Vanguardia*, esta responsabilidad se delega en la herramienta aportada por Classora.

Dos de los diarios nacionales, *ABC* y especialmente *El País*, establecen relaciones entre el marcado semántico y el lenguaje documental. Este último aplica las vinculaciones mediante tres sistemas: a través del Colabulario que se proyecta en la web con la optimización del SEO, los enlaces aportados por los periodistas y, finalmente, los algoritmos que se aplican en función de los intereses detectados en los lectores. En *ABC* el etiquetado se desarrolla a partir del lenguaje documental; mientras que en los otros tres periódicos no se establece relación alguna.

Finalmente, la evaluación del etiquetado no se desarrolla en todos los periódicos y los que lo hacen, lo establecen de forma diferente. A partir de las estadísticas internas y de herramientas SEO es la línea que *El País* aborda para esta cuestión; *ABC* consulta indicadores de posicionamiento web, mientras que *La Vanguardia* se rige por la prevalencia de términos generales con el fin de evitar duplicaciones en contenidos ya clasificados. Tanto *El Mundo* como *El Heraldo de Aragón* manifiestan no aplicar ningún sistema de evaluación.

## 5. Discusión y conclusiones

En buena parte de los periódicos analizados, se mantiene la presencia documental sobre los contenidos. Los responsables documentales continúan siendo conscientes de su importancia para la recuperación de información, y se mantiene el uso de lenguajes documentales. En este sentido, a la diversidad tipológica detectada se añade la que se observa en lo tocante a sus fuentes de alimentación. Nos encontramos ante instrumentos que suelen tener un número alto de términos, y que suelen contar con relaciones semánticas, aunque no en todas las opciones estudiadas. Cuentan con diferenciados objetivos y en su elaboración y mantenimiento intervienen documentalistas y, en ocasiones, otros profesionales. Con una actualización, por regla general, en un tiempo no superior al mes, no emplean ningún protocolo concreto de evaluación ni de determinación de su calidad. En cualquier caso, sí ofrecen un claro impacto en la recuperación de información.

No obstante, esa actividad documental parece tener un reflejo cada vez menos perceptible en el etiquetado real de los contenidos en la red. Las

etiquetas son aplicadas por documentalistas en la versión web, pero no son empleadas en las versiones de esos contenidos en medios o redes sociales. En tres de los cinco casos analizados, ese etiquetado no está realmente relacionado con un lenguaje documental, tampoco se corresponde con un marcado semántico en detalle de los contenidos y, cuando ocurre, está enfocado más bien a la actividad del medio en SEO (es decir, a mejorar su difusión a corto plazo en la red). Principalmente, se emplea para una descripción de personas, lugares y de ciertos temas. Tampoco se emplean, en su mayoría, estrategias de evaluación y de control, desde el punto de vista documental sobre el etiquetado realizado.

Otro hallazgo de esta investigación es que en solo dos de los cinco medios, la actividad documental tiene algún influjo real en la actividad digital del medio. Una de las principales motivaciones de estos resultados radica en el escaso número de profesionales, con responsabilidades sobre la documentación, existentes en los medios analizados.

Como conclusión general, se puede afirmar que los servicios de documentación que permanecen en los periódicos analizados están centrados, fundamentalmente, en tareas de recuperación de información que en la actividad digital del medio, y apenas participan en el etiquetado de contenidos, al menos, de cara al exterior. No ha sido posible encontrar referencias sobre servicios de documentación en periódicos nativos digitales, aunque este extremo requiere de una investigación más en profundidad: por un lado, para confirmar si esta ausencia es real y masiva; y por otro, para confirmar hasta qué punto la descripción y recuperación de información en estos medios, tanto interna (en la generación de contenidos) como externamente (en el etiquetado de los contenidos que se añade a las noticias para sus lectores), se ha visto afectada por esta situación.

Finalmente, esta investigación presenta diferentes limitaciones. En primer lugar, los datos expuestos proceden de 5 medios de comunicación que coinciden en una misma trayectoria: disponen de edición digital y de edición impresa, por lo que su servicio documental ya existía antes de su presencia en la red y se ha adaptado, en lo posible, al entorno digital. En segundo lugar, se ha aplicado un cuestionario, que es una técnica útil para los objetivos propuestos, si bien para una profundización en algunos aspectos sería útil acudir a otras aproximaciones metodológicas para completar algunos de los resultados y para despejar algunas dudas. A las dificultades aquí resaltadas conviene añadir que nos situamos en un momento de evidente cambio en la realidad documental.

## Agradecimientos

Los autores agradecen la colaboración de los diarios analizados y especialmente a los profesionales que han colaborado.

## Apéndice: Cuestionario empleado

### Datos generales

Medio:

Nombre de la persona que responde este cuestionario:

Puesto que ocupa en el medio:

### Sección 1: Preguntas genéricas para todos los medios

1. Si su medio dispone de versiones en papel y digital, ¿las redacciones digital e impresa están integradas? Sí / No
2. ¿Qué perfiles profesionales de su medio están vinculados a la actividad documental? Periodistas / Documentalistas / Programadores/informáticos / Gestores de la información digital / Especialistas en medios sociales (SEO, SEM, etc.) / Gestores de medios/redes sociales / Diseñadores gráficos/responsables de visualización / Responsables de métrica web / Otra...
3. ¿Su medio dispone de un servicio de documentación? Sí / No
4. En caso de respuesta afirmativa a la anterior cuestión, ¿cuáles son sus tareas? Asesoramiento en la recuperación de información a periodistas / Gestión y actualización del fondo documental / Elaboración y mantenimiento de lenguajes / vocabularios documentales / Etiquetado de documentos / Búsqueda de información / Elaboración de productos documentales / Elaboración de productos periodísticos / Otra...
5. ¿Cuántos documentalistas componen el servicio de documentación, si existe? 1 persona / Entre 1 y 3 personas / Entre 3 y 5 personas / Más de 5
6. ¿Qué bases de datos se emplean en vuestro trabajo cotidiano? (Respuesta abierta)
7. ¿Qué tipo de fondo documental se genera y se mantiene en vuestro medio? (Respuesta abierta)

### Sección 2: Cuestiones sobre el lenguaje / vocabulario documental

8. ¿Su servicio documental dispone de algún tipo de lenguaje documental? Sí / No

(El resto de esta sección, preguntas 9 a 26, se responderán solo en caso afirmativo a la pregunta anterior)

9. ¿De qué tipo de lenguaje se trata, teniendo en cuenta el control de los términos empleados? Lenguaje libre / Lenguaje controlado
10. ¿De qué tipo es ese lenguaje documental, teniendo en cuenta su estructura? Taxonomía / lenguaje de clasificación / Folksonomía / Tesauro / Ontología / Otra...
11. ¿Cuáles han sido las fuentes iniciales para la construcción del lenguaje documental? (Respuesta abierta)
12. ¿Cuál es el número de términos (aproximado) del lenguaje documental? Por debajo de 1000 términos / Entre 1001 y 3000 términos / Entre 3001 y 6000 términos / Más de 6000 términos
13. ¿En qué medida el lenguaje documental tiene impacto en la recuperación de información de los contenidos del medio? (1 = Nada / 5 = En gran medida)
14. ¿Qué fórmulas se emplean para la desambiguación de términos? No se emplean / Mediante notas aclaratorias / Mediante relaciones del tipo "usado por/use" / Otra...
15. El lenguaje documental ¿plantea relaciones semánticas entre los términos? Sí / No
16. En caso afirmativo a la pregunta anterior, ¿qué relaciones semánticas se emplean? Términos relacionados / Términos específicos/generales / Otra...
17. Ese lenguaje ¿tiene formas de categorización? Sí / No
18. En caso afirmativo, ¿qué tipo de categorización se emplea? (Pregunta abierta)
19. ¿Qué profesionales son los encargados de la elaboración y mantenimiento del lenguaje documental? Periodistas /

- Documentalistas / Programadores/informáticos / Gestores de la información digital / Especialistas en medios sociales (SEO, SEM, etc.) / Gestores de medios/redes sociales / Diseñadores gráficos/Responsables de visualización / Responsable de métrica web / Otros (Si es así, ¿cuales?)
20. ¿Con qué frecuencia se actualiza el lenguaje documental? En un lapso menor a un mes / Mensualmente / Trimestralmente / Anualmente / Más de un año / El lenguaje no se ha actualizado desde su creación
  21. ¿Ese lenguaje tiene conexión con otro tipo de lenguajes? Sí / No
  22. En caso afirmativo a la pregunta anterior, ¿con qué otros lenguajes? (Respuesta abierta)
  23. ¿El lenguaje está basado en algún modelo general de lenguajes de marcado semántico como: Dublin Core / OWL / Otra...
  24. ¿Se evalúa de alguna forma el funcionamiento y la calidad del lenguaje documental? Sí / No
  - 24B. En caso de respuesta afirmativa a la anterior, ¿cómo se evalúa? (Respuesta abierta)
  25. ¿Cuál es el coste de la elaboración y mantenimiento del lenguaje documental? (Respuesta abierta)
  26. ¿De qué forma se integra este lenguaje en la actividad cotidiana del medio? (Respuesta abierta)
- Sección 3: Preguntas sobre el marcado semántico de los contenidos del medio*
27. ¿Qué elementos y categorías de descripción semántica se emplean para realizar el marcado semántico de cada documento? (Respuesta abierta)
  28. ¿Qué contenidos del medio son etiquetados? Contenidos publicados por el medio en su sitio web / Contenidos publicados por el medio en medios sociales/redes sociales / Otra...
  29. Si se etiquetan contenidos publicados en medios o redes sociales, ¿qué redes o medios se emplean? Facebook / Twitter / YouTube / Instagram / Google+ / Snapchat / Tumblr / Pinterest / Otra u otras [Cuál o cuáles]
  30. ¿Quiénes son los encargados de etiquetar los contenidos del medio? Periodistas / Documentalistas / Programadores/informáticos / Gestores de información digital / Especialistas en medios sociales (SEO, SEM, etc.) / Diseñadores gráficos y responsables de visualización / Responsables de métrica web / Otra...
  31. ¿Qué relación existe entre el etiquetado o marcado semántico y el lenguaje documental que usa el medio? (Respuesta abierta)
  32. ¿Existe alguna forma de evaluar la efectividad del etiquetado? Sí / No
  32. En caso afirmativo, ¿cuál es el procedimiento para medir esa efectividad? (Respuesta abierta)

## Referencias

- Baños Moreno, María José (2013). Fuentes para la actualización de macrotesoros: noticias de divulgación científica. // Cuadernos de Gestión de Información. 3 (2013) 13-24.
- Baños Moreno, María José; Felipe, Eduardo R.; Pastor Sánchez, Juan Antonio; Martínez Béjar, Rodrigo y Lima, Germina (2015). Metadatos en noticias: un análisis internacional para la representación de contenidos en periódicos. // XII Congreso ISKO España y II Congreso ISKO España-Portugal, 19-20 de noviembre, 2015.
- Carroll, Nicholas (2010): Search Engine Optimization. // Bates, Marcia J. (coord). Encyclopedia of library and information sciences. Vol. 6, 4613-4629.
- Codina, Lluís; Gonzalo-Penela, Carlos; Pedraza-Jiménez, Rafael; Rovira, Cristófol (2017). Posicionamiento Web y Medios de Comunicación: Ciclo de Vida de una Campaña y Factores SEO. Barcelona: Departamento de Comunicación. Serie Editorial DigiDoc, 2017.

- Codina, Lluís; Pedraza Jiménez, Rafael (2011). Tesoros y ontologías en sistemas de información documental. // El profesional de la información, 20:5 (sept-oct, 2011) 555-563.
- du Preez, Madely (2015). Taxonomies, folksonomies, ontologies: what are they and how do they support information retrieval?. // The Indexer, 33:1, 29-37.
- Fondevila Gascón, Joan Francesc (2017). Algoritmos sobre el impacto de los medios de comunicación en medios sociales: estado de la cuestión. // Icono14. 15:1 (2017) 21-41.
- García Gutiérrez, Antonio (2011). Epistemología de la Documentación. Barcelona: Stonberg, 2011.
- García Gutiérrez, Antonio (2014). Análisis documental de noticias de prensa en sistemas de información factual. // Revista Española de Documentación Científica. 37:2 (abril-junio, 2014) e046.
- García Jiménez, Antonio (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. // Anales de Documentación. 7 (2004) 79-95. <http://revistas.um.es/analesdoc/article/view/1691/1741>
- García Jiménez, Antonio (2016). Organización del conocimiento para la documentación en periodismo: situación y prospectiva. // Scire. 22:2 (jul.-dic. 2016) 21-28.
- Jung, Jaemin; Song, Haeyeop; Kim, Youngju; Im, Hyunsuk; Oh, Sewook (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. // Computers in Human Behavior. 71 (2017) 291-298.
- López-García, Xosé; Toural-Bran, Carlos; Rodríguez-Vázquez, Ana Isabel (2016). Software, estadística y gestión de bases de datos en el perfil del periodista de datos. // El profesional de la información (EPI). 25:2 (2016) 286-294.
- Marcos-Recio, Juan Carlos; Edo, Concha (2015). Análisis de la nueva perspectiva de la documentación periodística en los medios de comunicación españoles. // Revista general de información y documentación. 25:2 (2015) 389-423
- Martínez González, M. Mercedes; Alvite Díez, M. Luisa (2014). Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. // Anales de Documentación. 17:1 (2014).
- Meléndez-Malavé, Natalia; Hirschfeld-Suárez, Rocío (2016). Situación de los centros de documentación escritos andaluces. // El profesional de la información. 25:4 (julio-agosto 2016) 606-615
- Mendes, Paula Raphisa; Martins dos Reis, Raquel; Coura Moreira dos Santos Maculan, Benildes (2015). Tesoros no acesso à informação: uma retrospectiva. // Revista ACB: Biblioteconomia em Santa Catarina. 20:1 (2015) 49-66.
- OK Diario (2017a). Nuevo récord de visitas de OKDIARIO: 33.089.131. // OK Diario. 19/8/2017. <https://okdiario.com/audiencia/2017/08/19/comscore-julio-2017-1251511>
- OK Diario (2017b). Nuevo récord de visitas: 37.545.412 y de usuarios únicos: 9.160.000. // OK Diario. 22/09/2017. <https://okdiario.com/audiencia/2017/09/22/comscore-agosto-2017-1343824>
- OK Diario (2017c). Nuevo récord de visitas de OKDIARIO: 39.380.408 y ya es sexto en el Top 10 de periódicos generalistas. // OK Diario. 20/10/2017. <https://okdiario.com/audiencia/2017/10/20/comscore-septiembre-2017-1433604>
- Pastor Sánchez, Juan Antonio (2013). Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. // Scire. 19:2 (2013) 55-68.
- Pastor Sánchez, Juan Antonio; Martínez Méndez, Francisco Javier; Rodríguez Muñoz, José Vicente (2012). Aplicación de SKOS para la interoperabilidad de vocabularios

- controlados en el entorno de linked open data. // *El profesional de la información*. 21:3 (may-jun, 2012) 245-253.
- Pérez Sanchidrián, Elaine; Campos Posada, Raúl; Campos Posada, Gloria Elisa (2014). Etiquetado social: un modelo de representación de la información en la blogosfera. // *Biblios*. 56 (2014) 19-27.
- Pintado Navarro, Rocío (2013). El comportamiento informacional de los periodistas en la Región de Murcia. // *Cuadernos de Gestión de Información*. 3 (2013) 25-51.
- Renó, Denis; Renó, Luciana (2017). Algoritmo y noticia de datos como el futuro del periodismo transmedia imágético. // *Revista Latina de Comunicación Social*. 72, (2017) 1.468-1.482. <http://www.revistalatinacs.org/072paper/1229/79es.html>
- Rubio Lacoba, María (2012). Nuevas destrezas documentales para periodistas: el vocabulario colaborativo del diario El País. // *Trípodos*. 31 (2012) 65-78.
- Saad Corrêa, Elizabeth; Bertocchi, Daniela (2012). A cena intercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. // *MATRIZES*. 5:2 (ene-jun, 2012) 123-144
- Sánchez Cuadrado, Sonia; Colmenero Ruiz, María Jesús; Moreiro, José Antonio (2012). Tesoros: estándares y recomendaciones. // *El profesional de la información*. 21:3 (may-jun, 2012) 229-235.
- Søbak, V.; Pharo, N. (2017). Decentralized subject indexing of television programs: The effects of using a semicontrolled indexing language. // *Journal of the Association for Information Science and Technology*. 68:3 (2017) 739-749.
- Soler Monreal, Concha; Gil Leiva, Isidoro (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. // *Revista Interamericana de Bibliotecología*. 33:2 (jul-dic, 2010) 361-377.
- Stone, Martha L. (2014). Big data for Media. Reuters Institute for the Study of Journalism. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media\\_0.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media_0.pdf)
- Szostak, Rick (2014). Advances in Classification Research Online 2013. Classification, Ontology, and the Semantic Web. // *Advances In Classification Research Online*. 24:1 (2014) 30-37.
- Yedid, Nadina (2013). Introducción a las folksonomías: definición, características y diferencias con los modelos tradicionales de indización. // *Información, Cultura y Sociedad*. 29 (dic. 2013) 13-26.

---

Enviado: 2018-04-12. Segunda versión: 2018-10-14.  
Aceptado: 2018-10-31.

---

---

# Evaluación de la coincidencia de los descriptores de las revistas españolas de PubMed

*An assessment of descriptor coincidence in the Spanish journals of Pub-Med*

---

Miguel Ángel PELLÉS GARCÍA (1, 2), José Antonio SALVADOR OLIVÁN (2)

(1) Biblioteca de Aragón, Zaragoza (España) (2) Departamento de Ciencias de la Documentación e Historia de la Ciencia. Universidad de Zaragoza. C/ Pedro Cerbuna, 12, 50009 Zaragoza (España); mapelles@unizar.es, jaso@unizar.es

## Resumen

Estudio que analiza la coincidencia en la indización que se da en revistas españolas en PubMed entre autores, indizadores profesionales y *MeSH on Demand*. El objetivo es determinar si la herramienta de indización automática es eficaz para la recomendación de términos de indización. Se extrajo una muestra representativa de 501 artículos sobre la población de revistas españolas ingresadas en PubMed en 2015 y se compararon artículo por artículo los descriptores asignados en cada caso por cada uno de los sujetos del estudio. Los resultados obtenidos demuestran que la coincidencia entre indizadores profesionales y *MeSH on Demand* es más alta que la obtenida por los autores. Así pues, se recomienda a las revistas científicas que animen a los autores a utilizar la herramienta de indización automática para la asignación de palabras clave con objeto de facilitar el acceso a la información.

**Palabras clave:** Indización. Grado de coincidencia en la indización. Indización automática. Métodos de indización. Medical Subject Headings (MeSH). MeSH on Demand. España.

## 1. Introducción

Actualmente tanto los productores de información científica, cómo las bases de datos más importantes de las distintas disciplinas destinan importantes recursos económicos y humanos a las tareas de indización de los documentos que componen sus colecciones. En esta línea y a modo de ejemplo, MEDLINE a comienzos de los 90 ya destinaba más de dos millones de dólares a estos menesteres (Hersh, Haynes y McKibbon, 1992). Una alternativa en la reducción de esos elevados costes de la indización humana viene de la mano de la indización automática. Esta variante al sistema de indización tradicional o humana viene siendo utilizada desde la década de los 60 por estos grandes productores de información científica (Gil Leiva y Rodríguez Muñoz, 1996).

Aunque no sin una cierta oposición al uso de la máquina para la realización de este tipo de tareas que desde tiempo remoto venían siendo desarrolladas por los profesionales de la información, se puede decir que hace ya tiempo, gracias al trabajo

## Abstract

The study aims to compare the indexing consistency between authors, professional catalogers and *MeSH on Demand* in the Pubmed database. The objective is to identify if the automatic indexing tool is effective in recommending indexing terms. A representative sample of 501 papers published in Spanish journals in 2015 was extracted, and the three set of assigned descriptors were compared for each individual paper. The major findings of the study indicate that the consistency value between professional indexers and *MeSH on Demand* is higher than the value obtained by the authors. As a final conclusion, there is a need for scientific journals to recommend authors to use the automatic indexing tool when assigning keywords to their papers in order to facilitate access to information.

**Keywords:** Indexing. Degree of coincidence in indexing. Automatic indexing. Indexing methods. MeSH terms. MeSH on Demand. Spain.

de grupos multidisciplinares que tratan con las técnicas de procesamiento del lenguaje natural (PLN), se están desarrollando herramientas automáticas cada vez más complejas y precisas que sirven para descargar de trabajo a estas secciones “humanas” de indización, o, al menos facilitar el trabajo tanto de los autores de textos científicos, como de los indizadores profesionales. En este sentido, en la US National Library of Medicine (NLM) se dan estos dos parámetros para desarrollar las tareas de indización de documentos: de un lado cuenta con una importante sección humana destinada estas tareas de indización de documentos; y de otro, dispone de una herramienta de indización automática que permite procesar para el indexado, los textos de los artículos biomédicos que ingresan en Medline, ofreciendo a autores o profesionales de PubMed, recomendaciones de términos para la indización de dichos textos.

A esta herramienta se le conoce como MeSH on Demand (Cho, 2014) y está accesible al público desde 2014. Su programa de procesamiento de textos se denomina Medical Text Indexer (MTI)

(Aronson, 2004), está operativo desde 2002 y su uso se hace más intensivo en tareas de indización a partir de 2011. Dicha herramienta tiene su antecedente en el Proyecto Sapphire (Hersh y Greenes, 1990), proyecto experimental para la indización automática y la recuperación de la información en Medline que se puso en funcionamiento en la década de los 90.

Son varios los estudios realizados sobre la coincidencia en la indización en la asignación de palabras clave, materias o descriptores a un documento (Abad García, Abad Pérez y Aleixandre Benavent, 1998; Moreno Fernández, Izquierdo Alonso y Vázquez Montávez, 2006; Evedove Tratarotti, Dal'Evedore y Lopes Fujita, 2017), pudiendo ser ésta realizada en periodos distintos o entre indizadores profesionales (Gil Leiva, 2002). La primera de ellas es denominada intraindización, y la segunda interindización; ambas se inscriben en la investigación de las diferencias individuales en la recuperación de la información (Saracevic, 1987) o, en otras palabras y fundamentalmente en el caso de la interindización, es una medida de la similitud de las reacciones de seres humanos diferentes cuando procesan la misma información (Zunde y Dexter, 1966).

Por tanto, este trabajo, aunque en cierta forma sigue la línea iniciada por Marcetich y Schuyler (1981), no pretende realizar un análisis de la coincidencia de la indización en sentido estricto. La finalidad del presente estudio es determinar el grado de coincidencia de los descriptores que asignan sujetos radicalmente diferentes (médicos, indizadores profesionales y una herramienta de indización automática); utilizando para ello herramientas de indización también diferentes; e incluso tomándose como base para la indización elementos distintos, ya que autor e indizadores de PubMed utilizan el texto completo; y para *MeSH on Demand* se ha utilizado el resumen elaborado por los propios autores.

Con este estudio se tratará de comprobar si la mencionada herramienta de indización automática permite, por un lado aconsejar de forma precisa a los autores de los trabajos médicos a la hora de elegir las palabras clave que acompañarán a sus artículos, y, por otro lado, ofrecer una lista adecuada de términos de cada artículo a los indizadores profesionales con el fin de poder ahorrar tiempo, y recursos a la NLM en estas tareas relativas a la indización.

## 2. Método

La población está constituida por los artículos publicados en revistas españolas indizados en Medline/PubMed en el año 2015. La estrategia

de búsqueda utilizada para poder seleccionar posteriormente la muestra ha sido:

Spain [PL] AND 2015 [DP] AND 2015/01:2015/12/31 [EDAD] AND medline [sb].

Los resultados obtenidos se filtraron por tipo de artículo, seleccionando de entre todos los disponibles los tres siguientes: *Journal Article*, *Review* y *Systematic Reviews*. El objetivo de este proceso de filtrado es dejar exclusivamente aquellos documentos que presentan una mayor posibilidad de inclusión de los términos de indización por parte de los profesionales de PubMed.

El resultado final de la búsqueda con su correspondiente filtrado fue de un total de 3009 artículos. Debido al número elevado de registros recuperados, se exportaron a 12 ficheros (uno por mes) en Refworks con el formato completo: título, fuente, autores, resumen y términos MeSH. Posteriormente, se exportaron a un único fichero Excel y de allí al programa estadístico SPSS, v. 20 para seleccionar la muestra. Se calculó el tamaño de la muestra que para un nivel de confianza del 95%, un error de estimación del 4% y un valor desconocido de la  $p$  y  $q$ , resultó en un total de 501 ítems. Los ítems se seleccionaron con el procedimiento de muestra aleatoria de casos del programa SPSS.

Las variables estudiadas, además de los datos obtenidos de los registros de Medline/PubMed fueron las siguientes: número de descriptores MeSH; número de descriptores asignados por los autores; número de descriptores que proporciona *MeSH on Demand*; coincidencia entre los términos MeSH y los propuestos por los autores; coincidencia entre los términos MeSH y los asignados por *MeSH on Demand*; y coincidencia entre los términos propuestos por los autores y los asignados por *MeSH on Demand*.

El trabajo de campo consistente en comparar los términos de indización asignados por autores de los artículos, indizadores profesionales de PubMed y la herramienta de indización automática *MeSH on Demand* se llevó a cabo según la metodología que se describe a continuación.

En primer lugar, se buscó el documento en PubMed a través del código PMID que tiene cada uno de los artículos ingresados en esta base de datos. Con el documento en pantalla se obtuvieron los siguientes datos: a) los términos MeSH asignados por los indizadores profesionales; b) el resumen del artículo incluido en la base de datos; y c) las palabras clave del artículo incluido en la base de datos. Se copió el resumen del documento, se pegó en la herramienta de indización automática *MeSH on Demand* y se procesó con objeto de obtener los descriptores generados a

partir del proceso de indización automática. Finalmente, una vez obtenidos todos los parámetros se efectuó el análisis comparativo de los mismos teniendo en cuenta las siguientes circunstancias “especiales” y la solución determinada en cada caso.

El autor principal comprobó si los descriptores coincidían entre los distintos procedimientos de indización, y descartó los registros problemáticos.

En primer lugar, no en todos los ítems de la muestra se puede llevar a cabo el contraste entre descriptores. El motivo es que, o bien faltan las palabras clave en el original, o bien falta el resumen; e incluso ambos en alguno de los ítems. El total de ítems incompletos en alguno de los citados elementos es de 89 sobre un total de 501 registros, lo que hace que casi un 17,8% de la muestra quede fuera del estudio.

Por otra parte, hay autores que usan descriptores que coinciden en su contenido, pero los equivocan en su forma, afectándolos en aspectos como el orden de los términos y sus correspondientes inversiones cuando se trata de descriptores compuestos, las palabras incluidas en el descriptor, etc. También hay problemas en el número gramatical de los descriptores; por ejemplo, el autor emplea un singular cuando el descriptor va en plural o a la inversa. En estos casos, se decidió no considerarlos como coincidentes.

Además, hay que mencionar también algunos cambios producidos en la nueva versión (julio de 2017) de *MeSH on Demand* (*MeSH on Demand Update*, 2017) que afectan a las pautas en el trabajo de campo. En primer lugar, la nueva versión no informa del número de caracteres del texto ingresado, aspecto que se estaba recogiendo hasta el momento en el estudio con objeto de comprobar si el tamaño del texto ingresado es directamente proporcional al número de términos extraídos. El problema se solventó pasando el texto por un contador de caracteres y palabras gratuito y así se recupera dicho dato —concretamente <http://www.contadordecaracteres.com/>—. En segundo lugar, se ha variado el orden el que presentan los descriptores obtenidos. En la versión anterior se presentaban en orden alfabético y ahora en función de la relevancia del término. Esto afecta a la rapidez de contraste entre los descriptores que dan los indizadores de PubMed (expuestos en orden alfabético) y los que ofrece la herramienta. Finalmente, otro cambio relevante afecta a los descriptores más “básicos” destinados a delimitar lo que se puede denominar como “objeto de estudio”, que en general son “menos importantes”, como ocurre con los identificadores del tipo de grupo al que va dirigido el estudio *Humans, Animals, Male, Female*, etc. son

ahora calificados como *Additional Terms*. Como hasta ahora no se hacía esta distinción se decide computarlos como si estuvieran integrados dentro de la sección de términos MeSH. Tanto en lo que atañe en el cómputo global de términos que ofrece dicha herramienta, como a la hora de contrastar la información dada con los términos MeSH que dan los indizadores profesionales.

En cuanto al proceso y análisis de los datos, el análisis estadístico se realizó con el programa SPSS. Se obtuvo la estadística descriptiva de cada variable y se analizó la correlación mediante el coeficiente de Pearson. Para las comparaciones se utilizó el análisis de la varianza para medidas repetidas.

### 3. Resultados

El número total de artículos estudiados ha sido de 501 distribuidos en 54 revistas. Las revistas con mayor número de artículos en la muestra seleccionada para el trabajo son: *Nutricion hospitalaria* con un total de 94 artículos analizados, *Cirurgia y cirujanos* con un total de 24 artículos analizados, *Revista de neurología* con un total de 23 artículos analizados y *Histology and histopathology* con un total de 21 artículos analizados. Estas cuatro revistas comprenden el 32.3% de los artículos estudiados, frente al 67.7% restante distribuido entre las 50 revistas restantes.

#### 3.1. Número de descriptores

El número de términos MeSH por artículo asignados por los profesionales que indizan artículos en PubMed y los asignados por la herramienta *MeSH on Demand*, es muy similar, obteniendo en ambos casos una media de 12,6 descriptores. Sin embargo, los autores de los artículos incluyen una media de 4,5 descriptores, un dato muy inferior.

	Media	Moda	Rango
Profesionales	12,6	12	2-33
MeSH on Demand	12,6	12	3-38
Autores	4,5	4	1-13

Tabla 1. Estadística descriptiva de descriptores

La moda obtenida también es coincidente en el trabajo realizado por los indizadores profesionales y por la herramienta de indización automática, frente a los autores de los textos. En este sentido el número de descriptores asignados a cada uno de los artículos por cada uno de los tres elementos estudiados, muestra que las cifras más repetidas de términos incluidos en un artículo son 12

términos para indizadores profesionales y *MeSH on Demand* y 4 términos para los autores.

En cuanto al número máximo y mínimo de descriptores asignados a cada artículo por parte de las diferentes variables estudiadas decir que el rango entre indizadores profesionales y herramienta es prácticamente coincidente, se está en ambos casos ante una horquilla muy amplia y similar, mientras que dicho rango es mucho menor cuando se trata de los autores.

En la Tabla II se presentan las diferencias entre los tres elementos estudiados. El número de descriptores utilizado por los profesionales es prácticamente el mismo que ofrece la herramienta *MeSH on Demand*; sin embargo, sí hay diferencias significativas cuando se comparan con el número de palabras clave empleadas por los autores, con un promedio de 8 descriptores más.

	Dif. medias	IC 95% diferencia	p <sup>a</sup>
Profesionales - <i>MeSH on Demand</i>	-0,29	-0,77 – 0,71	1,00
Profesionales - Autores	8,05	7,44 – 8,66	<0,001
<i>MeSH on Demand</i> - Autores	8,08	7,51 – 8,66	<0,001

Tabla II. Comparación de descriptores en los tres casos

(Dif. Medias: Diferencia de medias. IC: Intervalos de confianza. <sup>a</sup>Valor de la p para el test de MANOVA para medidas repetidas)

Hay evidencia también de una relación lineal positiva moderada entre el número de descriptores asignados por los profesionales y el de *MeSH on Demand* ( $r=0,161$ ,  $p=0,001$ ), asociándose en ambos casos un mayor número de términos con una mayor longitud del resumen, aunque la intensidad de esta relación es mayor con *MeSH on Demand* ( $r=0,474$ ,  $p=0,001$  vs  $r=0,216$ ,  $p=0,001$ ).

### 3.2. Análisis de coincidencia de descriptores

En este apartado se analiza el grado de coincidencia entre los descriptores asignados a los artículos por los profesionales de la indización, autores y la herramienta de indización automática.

El grado de coincidencia entre los términos dados por los autores de los artículos y los términos dados por los indizadores profesionales de PubMed es de 21,1%. Solo hay 8 artículos que presentan una coincidencia del 100%, lo que supone el 1,9% del total, mientras que la coincidencia es nula (0%) en 162 artículos (39,3% de la muestra).

El grado de coincidencia entre los términos dados por los autores de los artículos y la herramienta de indización automática es muy bajo, del 8,1%, siendo nula en 166 artículos y no hay ningún artículo donde sea total.

	Media	IC 95% diferencia
Autores - Profesionales	21,1%	18,9%-23,2%
Autores – <i>MeSH on Demand</i>	8,1%	7,2%-9,0%
<i>MeSH on Demand</i> – Profesionales	45,9%	44,0%-47,7%

Tabla III. Comparación de descriptores entre los tres elementos

El grado de coincidencia entre los términos dados por los indizadores profesionales de PubMed y los términos dados por *MeSH on Demand* es del 45,9%. Solo hay 2 artículos en los que la coincidencia es completa, y otros 2 en los que no coincide ningún término.

Cabe destacar que el grado de coincidencia en los términos asignados para cada artículo entre profesionales de la indización y la herramienta de indización automática es muy superior a cualquiera de las otras dos situaciones, quedando como menos coincidentes los términos asignados entre los autores de los artículos y la herramienta de indización automática. Hay que resaltar que esta alta tasa de coincidencia entre los indizadores profesionales y la herramienta de indización automática se produce en términos de indización básicos, es decir, aquellos que ayudan a delimitar el tema objeto de estudio del artículo.

Algunas cuestiones significativas que afectan a la asignación de encabezamientos básicos son los siguientes: a) la identificación del fin o el objeto del estudio: si está dirigido a humanos o animales, o por diferencias de género entre hombres y mujeres, etc.; y b) aludiendo a estudios concretos vinculados a una forma específica o a un tipo de estudio específico, como por ejemplo si son encuestas o cuestionarios, estudios prospectivos o estudios retrospectivos, etc.

Sin embargo y en la línea propuesta en el punto anterior relativa a la coincidencia de términos básicos, hay que decir que también se observan diferencias en la asignación de este tipo de descriptores entre los indizadores expertos de PubMed y la herramienta de indización automática. Esta circunstancia se ve por ejemplo en el caso de la asignación de descriptores referidos al grupo de edad al que afecta el estudio realizado en el artículo, es decir, si se trata de un estudio

cuyo sujeto son niños, jóvenes, adultos o personas mayores. Mientras que los indizadores profesionales sí que proporcionan esta información detallando ese grupo de edad sobre el que se efectúa el estudio, de forma general se puede decir que *MeSH on Demand* obvia este tipo de información y no asigna este tipo de descriptores o lo hace en muy pocas ocasiones.

Otro aspecto a destacar es que hay gran coincidencia entre los indizadores profesionales y la herramienta automática cuando en el artículo se trata una enfermedad específica y concreta, como un síndrome con un nombre concreto (por ej. *Metabolic Syndrome X*) o similares. En general *MeSH on Demand* identifica bien este tipo de situaciones y tiene una tasa de coincidencia elevada con los indizadores profesionales.

En cuanto a los aspectos negativos de la herramienta de indización automática, se observa una circunstancia que tiene que ver con la incoherencia en la selección de términos por parte de *MeSH on Demand*. Esta incoherencia se ve, por ejemplo, en artículos vinculados al tipo de sujeto sobre el que versa el estudio. Por ejemplo, a un estudio hecho sobre ratas (es decir, dentro de la categoría de 'Animales'), *MeSH on Demand* le asigna también el descriptor 'Humanos', es decir, en un estudio adscrito a la categoría de animales, la herramienta asigna los dos descriptores, antagónicos entre sí. Esto no ocurre prácticamente nunca en los descriptores asignados por los profesionales, en los que se mantiene la coherencia en todos los casos. Un ejemplo de esta circunstancia se ve en los artículos con PMID 26274892 o y 26238592.

A continuación, se presentan las revistas españolas en las que existe mayor y menor coincidencia en la asignación de descriptores:

- Coincidencia mayor del 90% entre autores e indizadores profesionales de PubMed. Tras el estudio se encuentran un total de 8 artículos distribuidos en 6 títulos de revista en los que la coincidencia entre los términos dados por los autores de los artículos y los términos dados por los indizadores profesionales es mayor del 90%. Siendo la revista *Nutrición hospitalaria* con 3 artículos la que mayor concordancia ha alcanzado en esta comparativa.
- Coincidencia mayor del 90% entre autores y *MeSH on Demand*. No se obtienen resultados en esta faceta, es decir, en ningún artículo se alcanza una tasa de coincidencia superior al 90% entre los términos asignados por los autores de los artículos y la herramienta de indización automática.
- Coincidencia mayor del 90% entre indizadores profesionales de PubMed y *MeSH on Demand*. Tras el estudio, se encuentran un total de 3 artículos distribuidos en 3 títulos de revista, en los que la coincidencia entre los términos dados por los indizadores profesionales y los términos dados por la herramienta de indización automática es mayor del 90%. Las revistas que mayor concordancia ha alcanzado en esta comparativa son *Psicothema*, *Reumatología clínica* y la *Revista Española de Sanidad Penitenciaria*.
- Coincidencia menor del 10% entre autores e indizadores profesionales de PubMed. El estudio detalla que hay un total de 164 artículos distribuidos en 42 títulos de revista en los que la coincidencia entre los términos dados por los autores de los artículos y los términos dados por los indizadores profesionales es menor del 10%. Los títulos de revista con mayor número de artículos con esta escasa coincidencia entre las palabras clave asignadas por los dos entes objeto de estudio para este caso son *Nutrición hospitalaria* (con 23 artículos); *Revista de neurología* (con 12 artículos) y; *Cirugía y cirujanos* (con 11 artículos).
- Coincidencia menor del 10% entre autores y *MeSH on Demand*. Hay un total de 265 artículos distribuidos en 49 títulos de revista en los que la coincidencia entre los términos dados por los autores de los artículos y los términos asignados por la herramienta de indización automática es menor del 10%. Los títulos de revista con mayor número de artículos con esta escasa coincidencia entre las palabras clave asignadas por los dos entes objeto de estudio para este caso son *Nutrición hospitalaria* (con 56 artículos); *Cirugía y cirujanos* (con 17 artículos); *Revista de neurología* (con 15 artículos); *Archivos Españoles de Urología* (con 12 artículos) y; *Journal of Physiology and Biochemistri* (con 10 artículos).
- Coincidencia menor del 10% entre indizadores profesionales de PubMed y *MeSH on Demand*. El estudio revela que hay un total de 5 artículos distribuidos en 5 títulos de revista en los que la coincidencia entre los términos dados por los indizadores profesionales de PubMed y los términos asignados por la herramienta de indización automática es menor del 10%.

#### 4. Conclusiones

Tras el trabajo de campo y el análisis de los resultados obtenidos en él se extraen las siguientes conclusiones:

1. El número medio de descriptores asignados por artículo es muy bajo cuando se trata de las palabras clave que dan los autores para su artículo. Con una media de 4 descriptores por artículo, frente a la media de 12 que dan tanto indizadores profesionales como *MeSH on Demand*, se determina como una seria traba de cara a la recuperación de la información, pues ese menor número de descriptores asignados por autores a los artículos reduce las posibilidades a la hora de recuperar los artículos a través de su contenido. A tenor de los resultados, si los autores emplearan la herramienta de indización, ésta les recomendaría un mayor número de descriptores y por tanto se incrementarían las posibilidades de recuperar la información.
2. El grado medio de coincidencia alcanzada entre indizadores profesionales y *MeSH on Demand* es más del doble que la media de cualquiera de las combinaciones en las que participan los autores como sujeto de comparación en dicha coincidencia en la indización. Si se da como premisa que la indización óptima es la efectuada por los indizadores profesionales de PubMed, es decir, se considera que los profesionales asignan los términos con mayor precisión en relación con el contenido del documento, es altamente recomendable que los autores utilicen la herramienta de indización automática, con lo que mejoraría la calidad de la indización y redundaría en una más eficaz recuperación de la información (Gil Leiva, 2008).
3. La herramienta de indización automática evita en la mayoría de las ocasiones los errores de número o forma en los descriptores elegidos por parte de los autores. Errores que tienen que ver con la asignación de palabras clave en las que equivocan el número, dando un plural en vez de un singular o a la inversa; y errores de forma que tienen que ver con la inversión de términos, la elección de una sigla en vez de la forma desarrollada o la inversa, etc. La herramienta de indización automática no falla nunca en este tipo de cuestiones al recomendar los términos de indización. El acertar en estos términos, en ocasiones muy relevantes por ser la materia principal sobre la que versa el artículo, elevaría el grado de coincidencia en la indización favoreciendo de nuevo la recuperación de la información.

A tenor de lo expuesto, decir que *MeSH on Demand* es una herramienta de indización automática que, a pesar de las constantes mejoras que ha recibido en los últimos años, sigue presentando problemas en las recomendaciones de descriptores MeSH que hace a sus usuarios.

No obstante, tasas como la más que aceptable coincidencia en la indización en la pareja conformada por indizadores profesionales y la propia herramienta, demuestran que es una herramienta bastante eficaz en la tarea de indización, que además de rapidez en este trabajo, mejora con mucho el trabajo realizado por parte de los autores en la asignación de las palabras clave que acompañan a sus artículos.

Las conclusiones a las que se llega tras este estudio confluyen en dos sugerencias básicas. La primera de ellas es que las editoriales o revistas científicas españolas responsables de los artículos albergados en PubMed recomienden a sus autores extraer las palabras clave del resumen que acompaña a su artículo, filtrando dicho artículo a través de la herramienta de indización automática *MeSH on Demand*.

El uso de *MeSH on Demand* en el proceso de asignación de palabras clave no solo facilita y mejora la calidad de los descriptores, sino que también asigna de forma muy sencilla información relevante sobre el propio artículo que los autores no suelen incluir a través de sus palabras clave: sobre el fin o el objeto de estudio, del tipo de estudio que es, a qué grupo de personas va dirigido, etc. La herramienta de indización automática sí que asigna términos en esta línea permitiendo a un usuario final, conocer el tipo de artículo ante el que se encuentra o cuestiones particulares sobre las que versa dicho artículo.

La segunda de las sugerencias va dirigida al equipo de indizadores profesionales de Medline/Pubmed, invitándoles a depositar parte de su confianza en los términos MeSH ofrecidos de forma automática por *MeSH on Demand*, puesto que a partir del resumen del artículo y en función de su extensión y de la corrección del mismo, se permite ver de forma muy rápida cuál es el tema central sobre el que versa el documento y cuáles son algunos de los términos de delimitación más importantes del objeto de estudio. Como ya se ha dicho anteriormente, dichos términos son solo recomendaciones y quedarían a expensas de un posterior tratamiento humano con el fin de subsanar posibles errores de la herramienta de indización automática. No obstante como se puede ver tras los datos obtenidos en este estudio, el uso de *MeSH on Demand* permitiría al menos orientar a los profesionales de la indización de Medline/Pubmed en su trabajo diario.

## Referencias

- Abad García, M. F.; Abad Pérez, I.; Aleixandre Benavent, R. (1998). Evaluación de la consistencia en la indización del repertorio Documentación Médica Española. // Revista Española de Documentación Científica. 21:4, 389-401.

- Aronson, A. R.; y otros (2004). The NLM Indexing Initiative's Medical Text Indexer // Medinfo (sept.). 268-272.
- Cho, D. (2014). MeSH on Demand Tool: An Easy Way to Identify Relevant MeSH Terms. // NLM Technical Bulletin. 398. [https://www.nlm.nih.gov/pubs/techbull/mj14/mj14\\_MeSH\\_on\\_demand.html](https://www.nlm.nih.gov/pubs/techbull/mj14/mj14_MeSH_on_demand.html) (may-june) (2018-01-12).
- Dal' Evedove Tartarotti, R. C.; Dal' Evedore, P.R.; Spotti Lopes, M. (2017). Avaliação Da Consistência Da Indexação Em Biblio-tecas Universitárias Federais Da Região Nordeste do Brasil. // Anales De Documentación. 20:1, 1-19. <https://doi.org/10.6018/analesdoc.20.1.253201> (2018-01-12).
- Hersh, W. R.; Greenes, R. A. (1990). Sapphire: An information retrieval system featuring concept matching, automatic indexing, probabilistic retrieval, and hierarchical relationships. // Computers and Biomedical Research. 23:5, 410-425.
- Hersh, W.R.; Haynes, R. B.; McKibbon, K. A. (1992). Evaluation of SAPHIRE: an automated approach to indexing and retrieving medical literature. // Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care: A Conference of the American Medical Informatics Association. 808-12.
- Gil Leiva, I. (2002). Consistencia en la indización de documentos entre indizadores noveles. // Anales de documentación. 5, 99-111.
- Gil Leiva, I. (2008). Manual de indización: Teoría y práctica. Gijón: Trea.
- Gil Leiva, I.; Rodríguez Muñoz, J. V. (1996). Tendencias en los sistemas de indización automática: Estudio evolutivo. // Revista Española de Documentación Científica. 19:3, 273-291.
- Marcetich, J.; Schuyler, P. (1981). The use of AID to promote indexing consistency at the National Library of Medicine. // Eight-first Annual Meeting of the Medical Library Association. Montreal: Medical Library Association.
- MeSH on Demand Update (2017). // NLM Technical Bulletin. 416 (may-june). [https://www.nlm.nih.gov/pubs/techbull/mj17/mj17\\_MeSH\\_on\\_demand\\_update.html](https://www.nlm.nih.gov/pubs/techbull/mj17/mj17_MeSH_on_demand_update.html) (2018-01-12).
- Moreno Fernández, L. M.; Izquierdo Alonso, M.; Vázquez Montávez, A. J. (2006). La consistencia como indicador de la fiabilidad (corrección) de la indización: los casos de LISA, PASCAL y VOCED. // Scire. 12:1 (en.-jun. 2006) 85-98.
- Mork, J.G.; Jimeno Yepes, A.J.; Aronson, A.R. (2013). The NLM medical text indexer system for indexing biomedical literature. // BioASQ Workshop. Valencia. <http://bioasq.org/sites/default/files/Mork.pdf> (2018-01-12).
- Saracevic, T. (1991). Individual differences in organizing, searching and retrieving information. // Proceedings of the 54TH Annual Meeting of the American Society of Information Science. 82-86.
- Stevens, M.E. (1965). Automatic indexing: a state of the art report, Monograph 91. National Bureau of Standards, Washington.
- Zunde, P.; Dexter, M. (1969). Indexing consistency and quality. // American Documentation. 20:3. 259-267.

---

Enviado: 2018-03-20. Segunda versión: 2019-05-02.

Aceptado: 2019-05-02.

---



---

# Procedimiento de evaluación

*Evaluation process*

---

## 1. Introducción

El procedimiento de evaluación en Scire es el de revisión por pares mediante el sistema de doble ciego —los revisores no conocen el nombre de los autores, ni viceversa— semejante al de otras revistas científicas, aunque tiene algunas peculiaridades que permiten ajustarlo a las condiciones específicas de su campo científico y de sus lectores.

## 2. Objetivos

El objetivo del procedimiento de evaluación es que se puedan admitir con rapidez los trabajos buenos, mejorar los que pueden ser susceptibles de aceptación —esto es, que, siendo interesantes, contienen errores o lagunas que deben ser corregidas— y rechazar los que no son adecuados para la revista por su tema o calidad, redirigiéndolos hacia otras publicaciones adecuadas o haciendo sugerencias al autor para la posible reutilización y reorientación de su trabajo.

El procedimiento de revisión tiene que tener en cuenta a los autores de los artículos que revisa, con los que el revisor debe intentar ser justo, pero también debe responder ante los lectores y los demás autores que compiten por el espacio de la revista, así como, de forma más general, con el proyecto editorial, la propia comunidad científica y el conjunto de la sociedad.

## 3. Procedimiento

Cada artículo requiere al menos la opinión favorable de al menos dos revisores y cada nota breve de al menos uno. Si el director de la revista estima que el artículo entra dentro de los objetivos de Scire —que se encuentran públicamente disponibles en cada número y en la página web de la revista—, lo envía a dos revisores —o, si es necesario, a más— elegidos por el director. Los evaluadores pueden ser miembros del consejo científico, revisores propuestos por los miembros del consejo científico con anterioridad o con motivo del artículo en cuestión, o científicos de reconocido prestigio propuestos por el autor o autores del artículo. La elección se hace según su conocimiento del tema y el número de trabajos que han evaluado hasta el momento (para equilibrar la carga de trabajo entre ellos).

Si el director estima sin lugar a dudas que el tema del artículo no se adecua al de la revista, se lo comunica al autor con una explicación suficiente, y, si le es posible, le sugiere otra revista que pueda ser más adecuada. De esta forma, se ahorra tiempo del autor y de los revisores.

La agilidad del procedimiento de revisión es un aspecto fundamental, por lo que el director solicita a los revisores que envíen su valoración en dos semanas, y que, en caso contrario, renuncien expresamente al proceso de revisión. Sin embargo, si el revisor no realiza ninguna de estas acciones, el proceso se retrasa sin remedio. En este sentido, el autor debe tener en cuenta que el proceso de revisión tiene un carácter voluntario. La decisión de aceptar el trabajo depende, salvo en casos excepcionales, de la opinión de los revisores. En los casos en que no están de acuerdo, la decisión es tomada por el director de la revista, que, excepcionalmente, puede decidir enviar el artículo a un nuevo revisor.

## 4. Criterios de evaluación

Se solicita a cada revisor que indique su valoración sobre los siguientes aspectos en una escala del 1 al 5 (1 muy cuestionable, 2 marginal, 3 suficiente, 4 sólido, 5 excelente):

1. *Tipo de contribución*: Se indica si se trata de una contribución de tipo teórico o metodológico, un estado de la cuestión, una comunicación de resultados de investigación, una experiencia específica, una exposición con valor introductorio o docente sobre el tema o una noticia de interés. Si es una mezcla de varios tipos, es posible marcar más de uno.

2. *Adecuación del tema de la revista*: Aunque al enviar el artículo al revisor el director de la revista ya ha realizado un juicio sobre su adecuación al enfoque de la revista Scire, el revisor puede también valorar este aspecto, comparando el tema del artículo con la declaración de objetivos y alcance de la revista, disponible en cada número y en la página web.

3. *Interés e importancia del tema*: Grado en el que el artículo aborda una cuestión de gran interés para el área científica de Scire, ya sea tradicionalmente, en el momento actual o en el futuro previsible.

4. *Actualidad del tema*: Grado en el que el tema se sitúa en los frentes actuales de investigación. No es un aspecto imprescindible, pero sí se considera informativo.

5. *Novedad y originalidad teórica de la aportación*: Grado en el que el artículo aporta una novedad teórica a su campo de investigación; esto es, un nuevo problema, un nuevo enfoque, un nuevo modelo, una nueva relación o ley, una nueva hipótesis, un nuevo concepto o una interpretación o modificación original de otros anteriores. Debe tratarse lógicamente de una aportación relevante y fundamentada.

5. *Novedad y originalidad metodológica de la aportación*: Grado en el que el artículo aporta una novedad metodológica a su campo de investigación; esto es, un nuevo procedimiento de estudio, de análisis o de prueba, aunque sea a un problema conocido.

7. *Aplicabilidad práctica*: Grado en el que las ideas del artículo pueden servir para la mejora de procedimientos y actividades prácticas y, en definitiva, para el desarrollo y la innovación.

8. *Rigor metodológico y validez de los resultados*: Grado de coherencia, exactitud, precisión y cuidado con el que se aborda el tema de estudio; validez y actualidad de la metodología utilizada; rigor en la selección de las muestras y en la aplicación de la metodología a las mismas; grado de replicabilidad del estudio.

9. *Rigor en el estado de la cuestión y el reconocimiento del trabajo previo*: Hasta qué punto se ha realizado un estado de la cuestión suficiente y actualizado y se han manejado las referencias pertinentes para el problema abordado.

10. *Claridad expresiva, orden y facilidad de lectura*: El artículo debe estar redactado de forma clara y bien ordenada, sin redundancias ni ideas fuera de contexto; y debe ser fácilmente entendible por cualquier lector potencial de Scire. No debe faltar información necesaria ni deben existir lagunas sobre el problema abordado, el método, los resultados, la discusión, las conclusiones o las recomendaciones. Si el artículo contiene aparato matemático, debe ser comprensible por un lector no especialista.

11. *Adecuación de gráficos, ilustraciones, tablas y apéndices*: Los gráficos, ilustraciones, tablas y apéndices deben ser necesarios y relevantes, estar bien resueltos técnicamente y disponer de una leyenda suficientemente explicativa.

13. *Conformidad de las referencias a las normas de la revista*: Grado en el cuál las referencias del artículo son suficientes y se conforman a las normas de la revista Scire.

## 5. Familiaridad del revisor con el tema

Aunque el director de la revista envía los artículos para su revisión intentando que coincidan con el campo de especialidad del revisor, pueden producirse desajustes. El revisor puede rechazar la revisión sugiriendo o no un revisor alternativo, o puede efectuar la revisión indicando de 1 (mínimo) a 5 (máximo) su grado de familiaridad con el tema del artículo evaluado. Si no rellena esta información, se entenderá que su grado de familiaridad es alto (4) o muy alto (5).

Esta información sirve para que el director de la revista pueda estimar la necesidad de enviar el artículo a un nuevo revisor en caso de conflicto de pareceres entre los revisores.

## 6. Valoración final

El revisor puede realizar cinco tipos de propuestas finales:

1. Aceptar el artículo para su publicación sin restricciones ni cambios, porque es un trabajo de alta calidad.

2. Aceptar el artículo para su publicación una vez realizadas las correcciones indicadas, bien sea incondicionalmente, al ser un trabajo de calidad, o bien de forma condicionada a que haya espacio disponible en el número, por ser un trabajo de calidad marginal. Las correcciones solicitadas se indican en las instrucciones para el autor, son de carácter menor y su comprobación puede ser delegada en el comité editorial.

3. Someter a reevaluación el artículo una vez realizadas las correcciones indicadas en las instrucciones para el autor, porque son de importancia y amplitud, y deben ser comprobadas por el revisor original u otro revisor.

4. Rechazar el artículo por las razones indicadas en las instrucciones para el autor, donde el revisor especificará con claridad y amabilidad las razones por las que su trabajo debe ser rechazado y, si le resulta posible, realizará sugerencias sobre su mejora o lugares alternativos de publicación.

5. El revisor rechaza evaluar el artículo por su falta de familiaridad con el tema o por razones que especifica en los Comentarios confidenciales para el equipo directivo.

La labor de los revisores es fundamental para que una publicación científica realice su labor de manera adecuada. Scire agradece de forma muy especial el tiempo y esfuerzo que dedican, y reconoce su trabajo publicando periódicamente sus nombres.

---

# Instrucciones para la presentación de trabajos

*Instructions for authors*

---

## 1. Cuestiones generales

---

Scire admite artículos con estados de la cuestión, discusiones de carácter teórico y metodológico, e informes de investigación sobre proyectos terminados, así como noticias sobre proyectos en curso, experiencias, notas y reseñas, que traten de la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

Las lenguas de publicación de Scire son el español, el inglés y el portugués.

Los artículos son sometidos a proceso de revisión por pares según el sistema de doble ciego, aunque los autores pueden proponer revisores libremente. El procedimiento es público, y puede ser leído en el apartado anterior a las instrucciones.

Los autores se comprometen a presentar y discutir públicamente y presencialmente sus trabajos en el congreso Scire correspondiente al año en curso.

Los trabajos deben ser originales y tendrán una extensión máxima de 20.000 caracteres en el caso de los artículos; 12.000 caracteres para los estudios de caso; y 6.000 caracteres para las noticias, proyectos en curso, experiencias, notas y reseñas. Excepcionalmente, estas cifras podrán ajustarse en virtud del interés del trabajo, aunque siempre con el acuerdo previo del editor.

Los autores deben presentar en una hoja aparte sus datos personales completos: nombre, cargo, dirección, teléfono, fax y correo electrónico.

Los artículos han de presentarse en un documento Word o RTF debidamente maquetados según la plantilla disponible en [www.iversid.org](http://www.iversid.org) con sus gráficos incluidos. Irán precedidos de sus títulos, resúmenes informativos y palabras clave, todos ellos en inglés y español. Se incluirán en ficheros aparte los gráficos e imágenes en su archivo fuente para su eventual tratamiento, y, en cualquier caso, con resolución de impresión.

Todos los apartados, incluidos las notas, referencias y apéndices irán numerados secuencialmente. Los subapartados —cuyos títulos serán claros y sintéticos— deberán tener la forma “1.1.”, “1.2.”, “1.2.1.” y así sucesivamente. No numere los resúmenes en español y en inglés, ni

los apartados de notas, agradecimientos y referencias. Evite un rango de encabezamientos superior a tres.

No se deben relatar datos analíticos repetitivos ni elencos de recursos en el texto principal. Por favor, preséntelos en forma de tablas o, si fuera imposible, en forma de apéndice.

El editor podrá realizar cambios de estilo para adecuar el trabajo a los requisitos de la revista.

Se recuerda al autor que la violación de derechos intelectuales y códigos éticos es un comportamiento inaceptable y puede constituir un grave delito.

## 2. Formatos de texto

---

No utilice subrayados, negritas ni versales.

Utilice mayúsculas sólo para acrónimos. No ponga puntos y espacios entre las letras que los forman.

Las cursivas se deben utilizar tan sólo para resaltar términos en otras lenguas, títulos de obras, palabras clave muy importantes que podrían servir a modo de títulos de un párrafo o conjunto de párrafos, y, con parquedad, para enfatizar los demás términos y frases a gusto del autor.

No utilice listas con salto de línea o de párrafo. Redacte párrafos completos según las reglas y los recursos gramaticales disponibles en su lengua de redacción. Puede indicar orden en las enumeraciones por medio de números o letras entre paréntesis, como, por ejemplo, (1), (2)...

## 3. Figuras y gráficos

---

Las figuras deben presentarse también en ficheros aparte en formato EPS o PDF con resolución de impresión de alta calidad (600 ppp) y, además, en el formato de la aplicación original con la que se creó. Tenga cuidado si genera imágenes a partir de la pantalla de su ordenador. Si recurre a imprimir pantalla en Windows obtendrá solo una resolución de 72 ppp. En ese caso, utilice una pantalla lo más grande posible, amplíe la ventana al máximo, copie con la tecla “impr pant”, pegue en el programa “Paint” y guarde el fichero; compruebe luego los resultados. Existen programas que permiten obtener imágenes de pantalla a una resolución aceptable. En el cuerpo

del artículo las imágenes portarán epígrafes indicativos de su contenido precedidos de la palabra “Figura” y un número correlativo (por ejemplo, “Figura 1.”).

#### 4. Tablas

Deben realizarse con el mismo procesador de textos que el artículo. En texto portarán epígrafes indicativos de su contenido precedidos de la palabra “Tabla” y un numeral romano correlativo (por ejemplo, “Tabla I.”).

#### 5. Notas

Las notas explicativas deben usarse excepcionalmente. No utilice el sistema automático de su procesador de textos. Refiéralas en texto tecleando su número secuencial entre paréntesis, e inclúyalas al final del trabajo en un apartado titulado “Notas” situado inmediatamente antes del dedicado a las “Referencias”, en orden secuencial y precedida cada una de ellas de su correspondiente número entre paréntesis seguido de tabulador. Las citas bibliográficas se realizarán como en el resto del texto; y la referencia completa se incluirá en el apartado de “Referencias”. Las páginas web se consideran referencias bibliográficas y deben tratarse como tales, evitando su cita como nota.

#### 6. Citas bibliográficas en texto

Las referencias bibliográficas en texto se denotarán por el apellido del autor, el año del trabajo y, si es necesario, el número de página. Si conviene, el nombre y el apellido del autor podrán quedar fuera del paréntesis. Si dos autores de la lista de referencias poseen el mismo primer apellido, se referirán por sus dos apellidos y, si tienen ambos apellidos iguales, por el nombre también. Si dos o más trabajos del mismo autor son del mismo año, se distinguirán mediante letras del alfabeto secuenciales en minúsculas puestas a la fecha sin espacio. Si hay varias referencias dentro del paréntesis, se separarán por punto y coma si son de diferentes autores, y coma si son del mismo.

Así, por ejemplo, “Es una teoría sustentada por varios autores (Gallego, 1975; Fernández y Alonso, 1993) [...]”. “Según Jaime Pérez (1993, p. 24) se establecen [...]”. “El mismo autor (Pérez, p. 27) hace constar la conveniencia de [...]”. “Se han detectado diferencias en la replicación de la experiencia (Menéndez, 1994a, 1994b; Menéndez y Alonso, 1997)”.

#### 7. Formato de las referencias bibliográficas

Las referencias bibliográficas se incluirán por orden alfabético al final del artículo bajo el título “Referencias”. Deberán seguir la norma UNE 50104:1996/ISO 690:1987, y, en cuanto a la puntuación, aproximarse lo más posible a las ISBD abreviadas, aunque sin espacio delante de punto, coma y punto y coma. Las referencias a páginas web deben incluirse también en este apartado, no en el texto ni en nota. Refiera adecuadamente en el texto del artículo toda la bibliografía utilizada; no incluya en el apartado referencias documentos que no haya incluido en el texto. Se presenta a continuación el formato para los principales tipos de materiales, con los elementos que pueden darse o no en cursivas, y seguidos de un ejemplo.

##### 7.1. Monografías, congresos, informes, normas y tesis

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. Edición. Lugar de publicación: editor; lugar de publicación: editor. ISBN número. Tesis doctoral.

Delclaux, Isidoro; Seoane, Julio (1982). *Psicología cognitiva y procesamiento de la información: teoría, investigación y aplicaciones*. Madrid: Ediciones Pirámide.

##### 7.2. Artículos de publicaciones periódicas

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. // Título de la publicación periódica. ISSN número. Volumen: número, primera página-última página.

Ellis, David (1992a). The physical and cognitive paradigms in Information Retrieval Research. // *Journal of Documentation*. 48:1, 45-46.

##### 7.3. Capítulos de una monografía, de un informe o contribuciones a un congreso

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. // Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. Edición. Lugar de publicación: editor; lugar de publicación: editor. ISBN número. Primera página-última página.

Markey, Karen (1990). Keyword searching in an online catalog enhanced with a library classification. // Bengtson, Betty G.; Hill, Janet Swan (eds.). *Classification of library materials: current and future potential for providing access*. New York: Neal-Shuman Publishers. 99-125.

Smith, Ph. J.; Beghtol, C.; Fidel, R.; Kwasnik, B. H. (eds.) (1993). *Proceedings of the 4th ASIS SIG/CR Classification Research Workshop: Columbus, OH, Oct.24, 1993*. Silver Spring, MD.: American Society for Information Science.

#### 7.4. Páginas web y recursos en línea

Se tratarán como monografías, artículos o capítulos según corresponda y se añadirá al final detrás de punto el URL y la fecha de consulta entre paréntesis:

Referencia. URL (Fecha de consulta).

Sagredo Fernández, Félix; Espinosa Temiño, María Blanca (2000). *Del libro, al libro electrónico-digital. // Cuadernos de Documentación Multimedia*. 9 (2000). <http://www.ucm.es/info/multi doc/multidoc/revista/num9/cine/sagredo.htm> (2001-01-22).

### 8. Apéndice I. El resumen

Un resumen proporciona información sobre objetivos, alcance, metodología, resultados, conclusiones y recomendaciones. Por alcance se entiende la determinación de hasta qué punto se han tratado los objetivos. Los resultados, conclusiones y recomendaciones se distinguen unos de otros así: los resultados son las informaciones o datos específicos y concretos que se obtienen de la aplicación de la metodología; las conclusiones se derivan de los resultados bien al interpretarlos en el marco de conocimientos teóricos previos o bien al crear un marco de interpretación nuevo a partir de ellos a través de un proceso de abstracción; finalmente, las recomendaciones se refieren a los nuevos problemas y vías de estudio que plantean a la vista de los resultados y conclusiones.

El resumen se debe redactar de forma clara y concisa, sin redundancias ni perífrasis ni información difícil de interpretar, como neologismos o abreviaturas que no se desarrollan.

Es muy conveniente que su primera frase contenga en primer lugar la mención del tipo de documento. Por ejemplo, “Estado de la cuestión sobre la aplicación del modelo europeo de calidad a las bibliotecas universitarias”, “Proyecto de aplicación del MARC21 a la catalogación de páginas web institucionales”, etc.

### 9. Apéndice II. La redacción del artículo científico: algunos aspectos obvios frecuentemente descuidados

#### 9.1. Sistemática

El artículo científico debe redactarse de forma muy organizada, con una estructura textual firme. En los informes de resultados de proyectos, se suele utilizar el esquema objetivos-metodología-resultados- conclusiones- recomendaciones. En los estados de la cuestión se aborda el asunto en orden cronológico o bien sistemático, según la ontología propia del área científica en cuestión; sigue la discusión, la síntesis, las conclusiones y las recomendaciones. Los artículos críticos suelen seguir el esquema tesis-síntesis-antítesis. Otros tipos documentales pueden exigir otras estructuras textuales, pero, en cualquier caso, dicha estructura debe ser explícita y aparente.

#### 9.2. Claridad y concisión

El lenguaje del artículo científico debe ser claro y conciso. Excepcionalmente, en los trabajos de tipo ensayístico las personas con un alto dominio del lenguaje escrito pueden permitirse la utilización de recursos literarios.

#### 9.3. Pertinencia y parsimonia

Debe evitarse proporcionar información ya conocida, salvo si constituye el punto central de la discusión. A ella cabe referirse por medio de las citas o, si estas no son necesarias, suponerla patrimonio de la comunidad profesional y científica. Sistematizar conocimiento es una función que se realiza solo en un tipo de artículo científico —la revisión o estado de la cuestión—; en los demás casos es contraproducente, aunque se deben citar las fuentes y las obras clave en la introducción del trabajo.

#### 9.4. Civildad

Es necesario citar a otros autores que han trabajado sobre el tema, demostrando que se conoce la literatura previa y que el trabajo se inserta civilizadamente en un determinado contexto científico.

---

# Índice de autores

*Author index*

---

BAIGET, Tomàs, 37  
CATALINA-GARCÍA, Beatriz, 55  
DINU, Nicoleta-Roxana, 37  
EVANGELISTA, Isadora Victorino, 25  
GARCÍA-JIMÉNEZ, Antonio, 55  
GUIMARÃES, José Augusto Chaves, 25  
LUZ, Gabriele de Araújo Medeiros, 25  
MOREIRA, Walter, 47

OSAWA, Henrique Fiamengue, 25  
PEGORARO NOVAES, Fernanda Carolina, 47  
PELLÉS GARCÍA, Miguel Ángel, 65  
PRADO MARTÍNEZ, Miguel Ángel del, 13  
RODRÍGUEZ-MATEOS, David, 55  
SALVADOR OLIVÁN, José Antonio, 65  
SANTANA de Moraes, Isabela, 47

---

# Índice de materias en español

*Subject index in Spanish*

---

Acceso abierto, 37  
Aragón, 13  
ARANOR, 13  
Brasil, 47  
Calidad, 37  
Comunidades epistémicas, 25  
Control de autoridades, 13  
DARA, 13  
Descripción archivística, 13  
Dimensión cultural, 25  
Documentación informativa, 55  
España, 55, 65  
Etiquetado, 55  
Evolución, 37  
Futuro, 37  
Grado de coincidencia en la indización, 65  
Indización, 55, 65  
Indización automática, 65  
Lenguajes documentales, 55

MARC21, 13  
Medical Subject Headings (MeSH), 65  
Medios, 55  
Mega-revistas, 37  
MeSH on Demand, 65  
Métodos de indización, 65  
Normas de descripción archivística, 13  
Organización del conocimiento, 25, 47  
Periodismo, 55  
Perspectiva, 37  
Procesos de organización del conocimiento, 47  
Registros de autoridad, 13  
Revistas académicas, 37  
Revistas científicas, 37  
Revistas piratas, 37  
Sistemas de información, 13  
Sistemas de organización del conocimiento, 47  
Suscripciones, 37  
Web semántica, 13

---

# Índice de materias en inglés

*Subject index in English*

---

Aragon, 13  
ARANOR, 13  
Archival description, 13  
Authority control, 13  
Authority records, 13  
Automatic indexing, 71  
Brasil, 51  
Controlled vocabularies, 59  
Cultural dimension, 27  
DARA, 13  
Degree of coincidence in indexing, 71  
Epistemic communities, 27  
Evolution, 41  
Future, 41  
Indexing, 59, 71  
Indexing methods, 71  
Information systems, 13  
Journalism, 59  
Journals, 41  
Knowledge Organization, 27, 51  
Knowledge organization processes, 51  
Knowledge organization systems, 51  
Media, 59  
Media libraries, 59  
Mega-journals, 41  
MeSH on Demand, 71  
MeSH terms, 71  
OA, 41  
Open access, 41  
Piracy, 41  
Quality, 41  
Semantic web, 13  
Spain, 59, 71  
Standards for archival description, 13  
Subscriptions, 41  
Tagging, 59