

La presente obra está bajo una licencia de:
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>



Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

Este es un resumen legible por humanos (y no un sustituto) de la [licencia](#). [Advertencia](#).

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

Adaptar — remezclar, transformar y construir a partir del material

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



CompartirIgual — Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la [misma licencia](#) del original.

Organización del conocimiento
Bibliotecología, Terminología e Información

COLECCIÓN
SISTEMATIZACIÓN DE LA INFORMACIÓN DOCUMENTAL
Instituto de Investigaciones Bibliotecológicas y de la Información

Organización del conocimiento
Bibliotecología, Terminología e Información

Coordinadoras
Catalina Naumis Peña
Adriana Suárez Sánchez



Universidad Nacional Autónoma de México
2025

Z1006
O743

Organización del conocimiento : bibliotecología, terminología e información /
coordinadoras Catalina Naumis Peña, Adriana Suárez Sánchez. –Primera edición.
– Ciudad de México : Universidad Nacional Autónoma de México, Instituto de
Investigaciones Bibliotecológicas y de la Información, 2025.
xi, 240 páginas. – (Sistematización de la información documental)

Este libro fue producido en el marco del Proyecto PAPIIT IT 400122 Sistema
de información terminológica sobre bibliotecología, archivología e información.

ISBN: 978-607-587-131-8

1. Organización del conocimiento. 2. Terminología bibliotecológica.
I. Naumis Peña, Catalina, editor. II. Suárez Sánchez, Adriana, editor. III. serie.

Este libro fue producido en el marco del Proyecto PAPIIT IT 400122 *Sistema de
información terminológica sobre bibliotecología, archivología e información.*

Diseño de portada: Carlos Papaqui Landeros

Primera edición: febrero de 2025

D. R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Instituto de Investigaciones Bibliotecológicas y de la Información
Circuito Interior s/n, Torre II de Humanidades,
pisos 11, 12 y 13, Ciudad Universitaria, C. P. 04510,
Alcaldía Coyoacán, Ciudad de México

ISBN: 978-607-587-131-8

Esta edición y sus características son propiedad de la Universidad Nacional Autónoma
de México. Prohibida la reproducción total o parcial por cualquier medio sin la
autorización escrita del titular de los derechos patrimoniales.

Publicación dictaminada
Hecho en México

Contenido

Presentación	vii
<i>Birger Hjørland</i>	
Prefacio	ix
<i>Catalina Naumis Peña</i>	

I.

PERSPECTIVAS DE EXPERTOS

La transformación conceptual y terminológica sobre catalogación en el siglo XXI	3
<i>Catalina Naumis Peña</i>	
Distribución de relaciones léxicas como indicador de propiedades estructurales en vocabularios controlados: el caso de la sinonimia	21
<i>Diego Ferreyra</i>	
La enseñanza de la construcción de tesauros: experiencias y reflexiones	39
<i>Francisco Javier García Marco</i>	

II.

APLICACIONES PRÁCTICAS

La representación temática facetada: análisis y propuesta de conversión y representación de encabezamientos de materia a FAST	63
<i>Noé Ángeles Escobar</i>	
Los sistemas de anotación como instrumento de organización del conocimiento	87
<i>Jesús Tramullas</i>	

Hacia una sistematización de la terminología de las ediciones: una contribución desde la organización del conocimiento a la industria editorial. 103

Mario Barité

Carolina Saravia Rebollo

Consolidando un ecosistema de documentación del patrimonio histórico y cultural de la UNAM: propuesta metodológica *ad hoc* para el desarrollo de un microtesauro. 121

Claudio Molina Salinas

III.

PERSPECTIVAS INNOVADORAS

Gestión de datos digitales: desafíos de la organización de información. 141

Lourdes Feria Basurto

Elizabeth Nashielli Márquez Sánchez

Repositorios institucionales, redes y organización del conocimiento: un estado de la cuestión en Hispanoamérica. 163

Gema Bueno-de-la-Fuente

María del Carmen Agustín-Lacruz

IV.

ORGANIZACIÓN DEL CONOCIMIENTO DIGITAL

Mapas tópicos y ontologías terminológicas: características y tendencias en la organización temática de colecciones digitales. 195

Adriana Suárez Sánchez

Las relaciones terminológicas entre los sistemas de información y los sistemas informáticos. 211

Hugo Alberto Guadarrama Sánchez

La organización de la información en la era digital. 227

Blanca Estela Sánchez Luna

Presentación

Dear Reader!

Welcome to an exploration of one of the most fundamental and fascinating aspects of librarianship: the organization of knowledge. This book presents some of the intricate systems and principles that underpin how information is classified, stored, and retrieved, making it an indispensable resource for any aspiring librarian.

Any information system needs a structured catalog system. Knowing the resources that a system contains and having the possibility of understanding the contributions that can lead us towards useful work in our work is a product of the organization of the knowledge that a system has. Knowledge organization transforms this chaotic scenario into a navigable, user-friendly environment where information is easily accessible.

CONTENT HIGHLIGHT:

EXPERT-INSIGHTS

- The conceptual and terminological transformation on cataloging in the 21st Century
- Distribution of lexical relations as an indicator of structural properties in controlled vocabularies: the case of synonymy
- Teaching thesaurus construction: experiences and reflections

PRACTICAL APPLICATIONS

- Faceted thematic representation: analysis and proposal for conversion and representation of subject headings to FAST
- Annotation systems as an instrument for knowledge organization
- Towards a systematization of the terminology of editions: a contribution from the organization of knowledge to the publishing industry
- Consolidating an ecosystem of documentation of the historical and cultural heritage of the UNAM: ad hoc methodological proposal for the development of a microthesaurus

INNOVATIVE PERSPECTIVES

- Digital data management: challenges of information organization
- Institutional repositories, networks and knowledge organization: a state of the art in Spanic America

DIGITAL KNOWLEDGE ORGANIZATION

- Topical maps and terminological ontologies: characteristics and trends in the organization of digital collections
- The terminological relationships between information systems and computer systems
- The organization of information in the digital age

I invite you to a journey through the contents of this book in Spanish on the great topic of the Organization of knowledge. Whether you are a student beginning your career in library science or a professional looking to update your skills, this book can help you learn about the topics related to the specialty and understand the path to producing your own contributions to the topic of organization of knowledge that is the backbone of library science.

Birger Hjørland
*University of Copenhagen,
Department of Communication
July 16, 2024*

Prefacio

Desde hace tres años buena parte del grupo que conforma el Seminario de Organización del Conocimiento del Instituto de Investigaciones Bibliotecológicas y de la Información (IIBI) desarrolla el proyecto *Sistema de información terminológica sobre bibliotecología, archivología e información*, apoyado por la Dirección General de Asuntos del Personal Académico. En el marco del seminario se han discutido muchos de los temas relacionados con el proyecto, pero también otros relacionados con la representación de significados contenidos en diferentes proyectos que desarrollan otros colegas. De ello surge esta obra que recoge las ideas, trabajos y plasma aportaciones muy interesantes para conectar los sentidos de las palabras en la organización del conocimiento.

El Sistema de Información Terminológica que se construye en el IIBI es una plataforma diseñada para gestionar, almacenar y facilitar el acceso a la terminología específica del dominio de conocimiento que se estudia en la institución y que además cuenta con un repositorio y bases de datos sobre el tema. Este sistema recopila y organiza términos y sus definiciones, así como sus relaciones y contextos de uso para agrupar datos, información y conocimiento con la finalidad de facilitar la recuperación y la interoperabilidad con otros sistemas.

Los sistemas de información terminológica son esenciales para garantizar la precisión, consistencia y eficiencia en la gestión y el uso de terminología especializada, apoyando diversas actividades profesionales y académicas y es uno de los grandes temas de la organización del conocimiento.

La organización del conocimiento ha sido una parte sustancial de los catálogos de bibliotecas y los archivos durante siglos. Sin embargo, a partir de la segunda mitad del siglo xx y en las primeras dos décadas que llevamos del siglo xxi, hemos presenciado una transformación significativa: en los recursos

documentales sobre los cuales se estudian y analizan los contenidos; en los métodos para realizar las operaciones que promueven el conocimiento de los contenidos documentales y, como consecuencia de ello, la inclusión de nueva terminología para designar muchas de las actividades inducidas en gran medida por los avances tecnológicos y el cambio en las necesidades de los usuarios.

La ciencia no existe aislada o fuera de la sociedad, sino que persigue reflejar los cambios, convirtiéndose al mismo tiempo en objeto y sujeto de las transformaciones que ocurren en la vida cotidiana. Es sujeto en el proceso de participar o crear condiciones y objeto cuando es receptáculo de interferencias sociales.

Este libro nace de la observación de estas transformaciones y de la necesidad de compartir estudios que se realizan en la especialidad, para motivar a los profesionales de la información y que, a su vez, operen como sugerencias de los temas que pueden ser abordados en las tesis de posgrado en Bibliotecología y Estudios de la Información.

El proceso de organizar el conocimiento se basa en estructurar, categorizar y sistematizar la información y los datos generados en las actividades y estudios para facilitar su acceso, comprensión y uso. Entre las prácticas que se realizan para llegar al conocimiento de los contenidos se incluyen la indización, la clasificación, la catalogación y la creación de sistemas de información como los repositorios.

Esas prácticas que sustentan los conocimientos de la especialidad son útiles en la conformación de bases de datos científicas, la creación de sistemas de gestión de información en las empresas o archivos privados o los procesos de catalogación, clasificación y ordenamiento de colecciones en bibliotecas y archivos.

Agradecemos a Birger Hjørland por su interés en los trabajos que se presentan al interior de la obra y por ser él mismo un promotor y el mayor experto sobre el tema a nivel internacional. En la actualidad trabaja en términos que se presentan en la *Encyclopedia of Knowledge Organization* y su prolífica obra sobre el mismo tema demuestra su competencia a nivel internacional. Entre otros, nos interesa destacar de su obra los estudios analíticos de dominio basados en estudios epistemológicos del conocimiento, que intensifican la cooperación entre terminología y organización del conocimiento.

Otro de los aspectos a destacar de la obra de Hjørland es el énfasis en que los sistemas de organización del conocimiento no son neutrales, sino que reflejan paradigmas teóricos que subyacen y es necesario estudiar para realizar clasificaciones más eficaces sustentadas en sistemas de organización del conocimiento más sólidos y académicamente más rigurosos. Su obra es muy extensa y su

contribución puede ser conocida a través de varios de sus trabajos publicados en las revistas *Knowledge Organization* y *Journal of Documentation*.

El hilo conductor de los trabajos presentados es dar cuenta de los estudios sobre el tema que se realizan en el idioma español, para que su comprensión no se vea sesgada por traducciones que a veces son influenciadas por la dificultad de encontrar la expresión más común y significativa utilizada en la comunidad hispano hablante a la que se encuentra destinada esta obra.

Los autores que comparten las investigaciones realizadas son expertos reconocidos en sus comunidades como investigadores o profesores en universidades de prestigio y con sólidos conocimientos en el tema de la obra.

Los autores reconocidos en sus respectivos países y a nivel internacional provienen de países como Argentina, Uruguay, España y México.

Esperamos que este libro sirva como una herramienta útil y práctica para todos aquellos interesados en la organización del conocimiento. La estructura del libro está diseñada para proporcionar tanto teoría como aplicaciones prácticas, facilitando así la comprensión en diferentes niveles de lectores que lo consulten.

Les invitamos a embarcarse en este viaje de transformación y descubrimiento, esperando que encuentren en estas páginas el conocimiento y la inspiración necesarios para adaptarse y sobresalir en el dinámico campo de la organización del conocimiento en el siglo XXI.

Catalina Naumis Peña
Ciudad Universitaria, a 22 de julio de 2024

I.
PERSPECTIVAS DE EXPERTOS

La transformación conceptual y terminológica sobre catalogación en el siglo XXI

The conceptual and terminological transformation on cataloging in the 21 century

Catalina Naumis Peña
*Instituto de Investigaciones Bibliotecológicas y de la Información
Universidad Nacional Autónoma de México*

INTRODUCCIÓN

En la década de 1990 la catalogación comienza una transformación conceptual con la consecuente aparición de nuevos términos. El objetivo de este trabajo es identificar la terminología de la catalogación en el presente y compararla con los términos que contenía el Tesauro Latinoamericano en Ciencia Bibliotecológica y de la Información (TELACIBIN) publicado en el año 1999¹ y que recogía la terminología usada en América Latina sobre el tema para analizar las variaciones observadas en la actualidad, en la cual, los metadatos y las estructuras de metadatos son el mecanismo indiscutible que sustenta la catalogación. En el presente año se está concluyendo un proyecto para actualizar el TELACIBIN y es una tarea obligada el análisis de cada temática de la bibliotecología, la archivología y la información, entre las que se encuentra la catalogación, como uno de los temas del área de Organización de la información y el conocimiento.

El objetivo es analizar los cambios conceptuales y de paradigmas en la catalogación en los últimos años mediante la comparación de algunos de los términos que incluía un tesauro elaborado a fines del siglo pasado en la disciplina bibliotecológica y los que se necesitan incorporar en la actualidad para reflejar la construcción del significado en la temática.

1 Catalina Naumis Peña *et al.*, *Tesauro latinoamericano de ciencia bibliotecológica y de la información* (México: Centro Universitario de Investigaciones Bibliotecológicas, 1999).

EL COMIENZO DEL CAMBIO CONCEPTUAL EN LA CATALOGACIÓN

Cuando se publica el TELACIBIN la catalogación que se usaba en la generalidad de los sistemas de información era automatizada y basada en sistemas comerciales, sobre los cuales se discutían y hacían públicos pocos aspectos, uno de los cuales era el formato MARC (Machine Readable Cataloging) estructurado para recoger los datos de los recursos bibliográficos en las diferentes áreas en que los distribuyen las normas RCAA2 (Reglas de catalogación angloamericanas, segunda edición) y que reproducía la presentación de los registros catalográficos manuales. El formato MARC es la estructura de registros, que define las designaciones de campos dentro de cada registro, y el propio contenido del registro que usan los catalogadores para conformar una base de datos, cada una elaborada con su propuesta particular de *software* y que además fue evolucionando como formato a lo largo de los años y de acuerdo con las nuevas necesidades.

El MARC comenzó a desarrollarse en los años sesenta, las bibliotecas fueron incorporándolo en las siguientes décadas para sustentar la catalogación y llegó a su máximo uso en los noventa. El surgimiento de los esquemas de metadatos que se desarrollan para facilitar la interoperabilidad de los recursos digitales trae consigo el esquema Dublin Core en 1995, que a través de un conjunto de 15 elementos de metadatos permite describir recursos documentales. Fue elaborado y auspiciado por la DCMI (Dublin Core Metadata Initiative), una organización dedicada a fomentar la adopción extensa de los estándares de metadatos y que a lo largo de los años continuó actualizando la tecnología necesaria para su inclusión en la web (World Wide Web).

Para despedir el siglo, en 1998 la IFLA (International Federation of Library Associations and Institutions) presenta el modelo conceptual FRBR (Functional Requirements for Bibliographic Records) acompañado del planteamiento sobre el modelo entidad-relación que irrumpe como una manera de establecer vínculos explícitos entre datos de los contenidos documentales.

A partir de esa fecha, a través de los años fue madurando la idea con mayores elementos que explican la importancia de las relaciones entre datos y se presentan aplicaciones que las posibilitan hasta convertirse en la actualidad en el medio de navegación en la web. Sin duda, el acceso a los contenidos documentales a través de las relaciones entre los datos que los definen ha cambiado las comunicaciones entre la información y los usuarios, como una nueva tecnología de acceso, sobre la cual Arlene G. Taylor resaltaba la tradición de la actividad de organización del conocimiento registrado en las bibliotecas

que explica ahora la arquitectura propuesta y que no difiere demasiado de lo practicado por muchos años en estas instituciones.²

El desarrollo de los modelos conceptuales de la familia FRBR prosiguió varios años y su incorporación para entender los diferentes aspectos que se deben cubrir fue madurando, y dichos aspectos fueron discutidos ampliamente en la literatura sobre catalogación. Un trabajo que resume los diferentes aspectos de esta peculiar familia y donde se puede profundizar al respecto es un artículo publicado en 2016 por Merčun, Žumer y Aalberg.³

EVOLUCIÓN CONCEPTUAL Y PRÁCTICA DE LA CATALOGACIÓN EN EL SIGLO XXI

El MARC 21 fue creado en 1999 como resultado de la armonización de los formatos MARC de Estados Unidos y Canadá, y UNIMARC, de uso amplio en Europa. Hoy en día, además de los formatos para registros bibliográficos, en la familia de estándares existen formatos MARC 21 para diferentes tipos de registros.⁴ En el 2002, la Biblioteca del Congreso de los Estados Unidos de América desarrolló el esquema MARC-XML como una estructura alternativa que permite que los registros MARC sean representados con una nueva tecnología, propia del entorno web.

Es así como a partir del surgimiento del XML (eXtensible Markup Language) se comienza a utilizar para intercambiar datos.⁵ Más tarde se agrega el RDF (Resource Description Framework) y los esquemas de metadatos comenzaron a desarrollarse con estos estándares diseñados para describir e interconectar datos en la web, facilitando la interoperabilidad entre diferentes sistemas y aplicaciones. MARC2RDF es otra propuesta para convertir registros en formato ISO 2709 en registros que incorporan el uso de las tripletas RDF. Sobre el particular como una nueva alternativa, la Biblioteca del Congreso de Estados Unidos de América trabaja en incorporar el BIBFRAME (Bibliographic Framework) para sustituir al formato MARC y otorgar un mejor acceso a los modelos RDF.

2 Arlene G. Taylor, *The organization of information* (2.^a ed.) (New York: Libraries Unlimited, 2004), 18.

3 Tanja Merčun, Maja Žumer, y Trond Aalberg, “Presenting bibliographic families: Designing an FRBR-based prototype using information visualization” *The Journal of documentation* 72, n.º 3 (2016), 490-526, doi: 10.1108/JD-01-2015-0001

4 “Conociendo MARC Bibliográfico: Catalogación Legible por Máquina”, acceso 29 de mayo de 2024, <https://www.loc.gov/marc/umbspa/>

5 “MARCXML MARC 21 XML schema official Web Site”, acceso 29 de mayo de 2024, <https://www.loc.gov/standards/marcxml/>

El modelo BIBFRAME es una iniciativa para desarrollar los estándares de descripción bibliográfica hacia un modelo de datos enlazados, con el fin de que la información bibliográfica sea más útil tanto dentro como fuera de la comunidad bibliotecaria. Cuando se cataloga un recurso (por ejemplo, un libro), la descripción resultante incluye elementos de información como el autor, el tema del libro, diversos formatos publicados e información sobre copias del libro. BIBFRAME organiza esta información en tres niveles básicos de abstracción: Obra (Work en inglés), Ejemplar (Instance) o Ítem (Item); con tres clases adicionales (agente, materia, evento), las cuales se relacionan con las categorías principales.⁶

Con base en los nuevos términos que están definiendo la actividad tradicional de catalogación documental se debe considerar que el registro realizado a partir de esquemas de metadatos está basado en la tecnología de la web en la cual conviven con las descripciones de objetos del mundo real. Como lo expresa Guerrini:

*For those already familiar with cataloguing, the new digital context requires a further paradigm shift: it requires metanoia, 'a change of mind', an open intellectual disposition; a conscious disposition and one not conditioned by conventions, which are always provisional.*⁷

Las transformaciones y actualizaciones de los modelos conceptuales, los cambios en los códigos de catalogación y los estándares de metadatos son constantes para integrarse a la web semántica:

*Models of the librarianship domain underlie the compilation of codes of cataloguing rules and metadata standards. Creating a model is affected by transformations due to constant acquisitions and innovations. The impact of technological evolution is evident when comparing the models of the FR family with IFLA LRM; the latter is specifically conceived for the Semantic Web context.*⁸

6 Overview of the BIBFRAME 2.0 Model (BIBFRAME-Bibliographic Framework Initiative, Library of Congress), acceso 29 de mayo de 2024, <https://www.loc.gov/bibframe/docs/bibframe2-model.html>

7 Mauro Guerrini, *From Cataloguing to Metadata Creation: A Cultural and Methodological Introduction* (London: Facet Publishing, 2023), 6.

8 Jan Pisanski and Marcia Žumer, "Mental Models of the Bibliographic Universe". Part 1: "Mental Models of Descriptions"; Part 2: "Comparison Task and Conclusions", *Journal of Documentation* 66, n.º 5 (2010), 668.

El Metadata Object Description Schema (MODS) fue desarrollado por la Biblioteca del Congreso de Estados Unidos y su primera versión se publicó en diciembre de 2002. MODS se creó como un esquema XML para describir recursos bibliográficos y fue diseñado para ser más detallado que Dublin Core, pero menos complejo que MARC.

A partir de la década del dos mil el uso frecuente de la web semántica trae consigo la creación y el uso de esquemas de metadatos más complejos, como OWL (Web Ontology Language) y SKOS (Simple Knowledge Organization System), para mejorar la interoperabilidad y la vinculación de datos temáticos en la web. A ello se agregaron esquemas de metadatos específicos para diversos dominios, como METS (Metadata Encoding and Transmission Standard) para la preservación digital, y TEI (Text Encoding Initiative) para textos literarios y lingüísticos.⁹

Las Resource Description and Access (RDA) surgieron por primera vez el 22 de junio de 2010. Fueron desarrolladas para reemplazar las Anglo-American Cataloguing Rules, Second Edition (AACR2), ya que son compatibles con estándares internacionales de metadatos para integrarse en el entorno digital. Las actualizaciones son realizadas de manera continua por el Joint Steering Committee for Development of RDA (JSC), actualmente conocido como el RDA, Steering Committee (RSC) y se consultan en el sitio web del RDA Toolkit.

Desde 2016, siguiendo instrucciones de la Sección de Catalogación de la IFLA, la ISBD (en español, Descripción Bibliográfica Internacional Normalizada) como estándar y los modelos bibliográficos (FRBR y ahora IFLA Library Reference Model (LRM)) son responsabilidad del Comité de Normas de la IFLA. Al mismo tiempo el Grupo de Estudio de Datos Vinculados de ISBD publicó las directrices para utilizar las ISBD como datos vinculados. La utilización de los datos vinculados significa la inserción en la semántica de la web.

En 2021 se publica con el título ISBD: International Standard Bibliographic Description Actualización de 2021 a la edición consolidada de 2011, en la que se señala que:

Los términos de referencia que el Grupo de Revisión de ISBD propuso al Grupo de Trabajo de Actualización de Contenido de ISBD encargado de esta revisión, fueron: 1. proporcionar una actualización del contenido de ISBD de acuerdo con los comentarios y propuestas recibidos por la comunidad de interés y usuarios del estándar ISBD; 2. ampliar la cobertura de ISBD para incluir una mayor variedad de recursos, especialmente los recursos no publicados. La relación con el Modelo de Referencia

9 Pastor Sánchez, “Marcado semántico”, 56.

Bibliotecaria de la IFLA (IFLA LRM) está en proceso. El mapeo de ISBD a LRM ya fue publicado, pero el alineamiento con LRM, que implicaría darle al estándar ISBD una nueva forma basada en la visión de LRM del universo bibliográfico, es el tema central de la Revisión de ISBD que actualmente está en desarrollo por el Grupo de Trabajo de la Manifestación dependiente del Grupo de Revisión de ISBD.¹⁰

NUEVOS ENTORNOS DE CATALOGACIÓN ANALÍTICA

En párrafos anteriores se alude a metadatos utilizados en la indización temática de los registros documentales que se obtienen de la catalogación analítica, con la finalidad de explicitar los términos que califican los contenidos.

Además de los metadatos de la catalogación analítica se usa el etiquetado para indizar, también conocido como etiquetado semántico o etiquetado de contenido, que es un proceso utilizado para asignar etiquetas a diferentes elementos de información, como documentos, imágenes, videos, textos de plataformas, etcétera. Estas etiquetas están diseñadas para describir el contenido de manera precisa y relevante, lo que facilita la búsqueda y recuperación de la información en sistemas de gestión de contenidos.¹¹

Otra forma de indizar contenidos es a través del marcado de información, que es una técnica crucial para estructurar y anotar datos de manera que sean fácilmente comprensibles y utilizables tanto por humanos como por máquinas, mejorando la organización, búsqueda y recuperación de información en diversos contextos. El marcado de información a diferencia del etiquetado debe utilizar lenguajes de marcado como: HTML (HyperText Markup Language) utilizado para estructurar y presentar contenido en la web; XML (eXtensible Markup Language) lenguaje flexible para definir estructuras de datos personalizadas, comúnmente usado para el intercambio de datos entre siste-

10 IFLA. ISBD Descripción Bibliográfica Internacional Normalizada: Actualización de 2021 a la edición consolidada de 2011, aprobada por el Grupo de Revisión de ISBD de IFLA, aprobada por el Comité Consultivo de Estándares de IFLA, febrero 2022, <https://repository.ifla.org/bitstream/123456789/2491/1/ISBD%20consolidada%20Espa%C3%B1ol%202021-%20Act2021.pdf>

11 Gonzalo Mochón-Bezares, Eva Méndez-Rodríguez y Ángela Sorli-Rojo, “Etiquetado social y blog-scraping como alternativa para la actualización de vocabularios controlados: Aplicación práctica a un tesoro de Biblioteconomía y Documentación”, *Información, cultura y sociedad* 37 (diciembre 2017), 13, <http://repositorio.filo.uba.ar/handle/filodigital/11218>

mas; y Markdown, lenguaje ligero de marcado para formatear texto de manera sencilla, comúnmente usado en documentación y foros.¹²

Los catálogos bibliotecarios manuales relacionaban datos como títulos de las obras, autores y temas y ayudaban al acceso a la información, ahora ese acceso se ha multiplicado y potenciado. La adopción de modelos conceptuales en el medio bibliotecario internacional es una respuesta a las demandas de organización, interoperabilidad, consistencia y adaptación tecnológica que impone la era digital. Estos modelos no solo optimizan la gestión de la información, sino que también significan mejor acceso y uso de los recursos documentales.

En un estudio realizado sobre la evolución temática de la investigación sobre Información y Documentación en español en la base de datos LISA en el 2020, se relaciona la catalogación con la cooperación bibliotecaria, marcando un punto de vista diferente del tradicional:

“Catalogación y cooperación bibliotecaria” da título a un conjunto de documentos que tratan sobre la catalogación bibliográfica y todo el contexto que la rodea. Aunque el conjunto pueda parecer a primera vista algo artificial, tiene todo el sentido si se piensa que durante gran parte de la historia contemporánea de las bibliotecas, la catalogación bibliográfica y el desarrollo de normas y formatos de catalogación e intercambio han sido el centro de las actividades de cooperación entre las bibliotecas y han constituido la base para otros servicios como el intercambio de documentos y la creación de catálogos colectivos, por lo que han sido centrales en el origen de organizaciones como la IFLA o la OCLC.¹³

La diferencia que conlleva la catalogación en la actualidad es que el catalogador debe conocer esquemas y técnicas para registrar la información y que también debe conocer la forma en que pueden hacer la información accesible al usuario.

Como se puede observar, un fenómeno generalizado es el uso de las abreviaturas en los nombres de las tecnologías, los esquemas y los estándares para reducirlos, situación que se aplica también en la catalogación. Muchas de estas tecnologías fueron lentamente agregadas en los sistemas bibliotecarios, sobre

12 Jesús Tramullas, “Desarrollos en elaboración de documentación técnica: los lenguajes de marcado ligero”. *Anuario ThinkEPI* 13 (2019), 1, doi.org/10.3145/thinkepi.2019.e13f03

13 Francisco Javier García-Marco, Carlos G. Figuerola y María Pinto, “Análisis de la evolución temática de la investigación sobre Información y Documentación en español en la base de datos LISA mediante modelado temático (1978-2019)”. *Profesional de la información* 29, n.º 4 (2020), 9, https://doi.org/10.3145/epi.2020.jul.27

todo en Latinoamérica, pero la discusión acerca de la necesidad de incorporar la tecnología para catalogar está presente en la literatura bibliotecológica y, como consecuencia, la cantidad de abreviaturas que serán incorporadas a la nueva propuesta de tesaurus.

Un trabajo terminológico debe estar basado en la garantía literaria y la garantía cultural de la terminología de una especialidad. La garantía literaria se refiere al uso lingüístico plasmado en la literatura sobre una especialidad que debe ser analizada desde la perspectiva de la frecuencia de uso y la aceptación o el reconocimiento como formas de nombre válidas.¹⁴

Cuando se dice que la garantía literaria es el respaldo con que se cuenta para justificar la inclusión de un término en un vocabulario controlado o su utilización como palabra clave, se está diciendo también que se requiere la existencia probada de documentación sobre ese tema para que pueda ser considerado como herramienta de almacenamiento, recuperación, comunicación e intercambio de información.¹⁵

La garantía cultural supone la comprensión del lenguaje reflejado en el contexto cultural estudiado y que refleja la valoración realizada por los hablantes del ámbito de estudio. Beghtol, una teórica de la clasificación, analizó las garantías para validar terminología desde el año 1986, propuso, además, incorporar el concepto de garantía cultural a considerar en los estudios terminológicos:

cultural warrant posits that every classification system is based on the assumptions and preoccupations of a certain culture, whether the culture is that of a country, or of some smaller or larger social unit (e.g. ethnic group, academic discipline, arts domain, political party, religion and/or language).¹⁶

Tanto la garantía literaria como la cultural son útiles para seleccionar los términos representativos de un dominio temático y que posteriormente pasarán a la etapa de definición del descriptor, o sea el término que mejor represente una temática de acuerdo con la frecuencia del uso, el valor de quienes lo utilizaron como término en su discurso y además la adopción del término por la comunidad a la que irá destinada el vocabulario controlado que se construye.

14 Mario G. Barité, “Garantía literaria y normas para construcción de vocabularios controlados: aspectos epistemológicos y metodológicos”. *Scire: representación y organización del conocimiento* 15, n.º 2 (2009), 16, <https://doi.org/10.54886/scire.v15i2.3709>

15 Barité, “Garantía literaria”, 15.

16 Clare Beghtol, “Universal concepts, cultural warrant and cultural hospitality”. En *Proceedings of the 7th International ISKO Conference*/ López-Huertas, M.J., editor (Granada: Würzburg: Ergon Verlag, 2002), 45.

ANÁLISIS COMPARATIVO DE LA TERMINOLOGÍA SOBRE CATALOGACIÓN

En virtud de la transformación de la terminología en el tema de catalogación sobre la cual ya se ha visto algo de lo que ocurre, se hará un estudio comparativo entre la terminología reconocida a través de un tesaurus y la terminología que se rescata de la literatura bibliotecológica sobre el tema en los años transcurridos desde la publicación del tesaurus.

Sobre los estudios comparados, Fuentes Romero y Fernández Rodríguez establecen:

Algunos autores afirman que la comparación es inherente a cualquier procedimiento científico, es decir, que el método científico es inevitablemente comparativo, puesto que para verificar una hipótesis es necesario comparar los resultados obtenidos después de manipular determinadas variables y observar los resultados. Pero esto no supone que siempre que se compare se haga siguiendo unos procedimientos científicos. Para que esto ocurra se tienen que dar los siguientes pasos, presentados de una manera breve y general: determinar los objetos a comparar; en qué aspectos son comparables y, además, seguir unas estrategias de análisis para llegar a unas conclusiones.¹⁷

En el año 1999 se publicó el TELACIBIN, de donde se recogieron para comentar algunos de los términos que representaban tanto la acción de la catalogación como de los productos de la actividad y que ahora han sufrido cambios muy drásticos, a veces en sus relaciones, otras en la erradicación de su uso.

Por mencionar algunos casos, entre los diferentes *catálogos* se cuentan *catálogos de autoridad*, *catálogo de autoridad de autor* y *catálogo de autoridad de materia*. Estos catálogos estaban relacionados con la actividad control de autoridad. En la actualidad, de acuerdo con la nueva normatividad, un término utilizado es *catálogo de puntos de acceso*, no usado en el medio bibliotecario mexicano. Actualmente el proceso de normalización de un punto de acceso en la literatura es denominado *control de puntos de acceso*, antes del desarrollo de RDA fue conocido sin discusión, como control de autoridades.¹⁸ Ahora la nueva normatividad de registro documental para el mundo digital en que estamos

17 Juan José Fuentes Romero y Vanesa Fernández Rodríguez, “Una revisión bibliográfica de los estudios comparativos. Su evolución y aplicación a la ciencia de las bibliotecas”, *Revista Interamericana de Bibliotecología* 32, n.º 2 (2009), 413.

18 Barbara Tillett, *RDA: antecedentes y aspectos de su implementación: manual para instructor*. Conferencia presentada en febrero de 2009, acceso 29 de mayo de 2024, https://www.loc.gov/catdir/cpso/RDA/RDAantecedentes_instructor.pdf

insertos recomienda control de acceso. Quizás los sistemas de información mantengan sus catálogos de autoridad todavía por un buen tiempo o quizás se comiencen a construir de otra manera.

El *control de acceso* es un término discutido por Jiménez y García porque al autorizar muchos posibles puntos de acceso en el registro documental, como se esboza en la normatividad que se está proponiendo, en donde se aspira a vincular todas las variantes encontradas en los catálogos y que designan a una misma persona, entidad, título o materia. Es decir, no hay una forma que prevalezca con respecto a la otra.¹⁹

En la actualización del tesoro se deberán mantener el término *catálogo de autoridades* y la *actividad control de autoridades* porque siguen siendo las formas predominantes en la literatura sobre la materia, pero usándolos como términos equivalentes: de la actividad *control de puntos de acceso*, *control de acceso*, *gestión de autoridades*, *gestión y mantenimiento de autoridades*.²⁰

Otro núcleo de catálogos integrados entre los descriptores del TELACIBIN que son dignos de mencionar están: *catálogos diccionarios*, *catálogos divididos* y *catálogos en fichas*. Este tipo de catálogos son casi de tipo histórico porque existen pocos catálogos manuales y, quizás, sea conveniente crear esa categoría de catálogos históricos.

En el tesoro comentado aparece *metadatos* y, por supuesto, se incorpora formato MARC, formato MARCAL, formato UNISIST, que eran los habituales y usados en 1999. A partir de los cuales se promovió el MARC 21, que es el formato actual utilizado en los sistemas bibliotecarios. Los formatos están integrados como parte del área de conocimiento titulada en el TELACIBIN *Transferencia de información documental* y en el área de *Análisis y sistematización de la información* aparecen como términos relacionados de *catalogación automatizada*. No existe ninguna referencia a los principios de catalogación y tampoco se incorporó el Modelo conceptual FRBR, que ya se venía perfilando desde principios de los noventa.

La *descripción bibliográfica* es un descriptor del TELACIBIN entre el área de *Análisis y sistematización de la información* y el descriptor específico *catalogación* y al respecto la concepción que se perfila es incorporar *universo bibliográfico* como término genérico de *catalogación descriptiva* y *catalogación analítica*, la primera desprendiéndose de la rama *organización de la información*, y la segunda de la *organización del conocimiento*. Como términos específicos se presenta la familia de las ISBD, donde se aceptan las abreviaturas con el correspondiente *usado por* donde se desglosa el significado.

19 Jesús Jiménez Pelayo y Rosa García Blanco, *El catálogo de autoridades: creación y gestión en unidades documentales* (Gijón: Ediciones Trea, 2002), 71.

20 Jiménez Pelayo y García Blanco, *El catálogo de autoridades*, 28.

De *catalogación descriptiva* se desprende la *descripción física*, y de la *catalogación analítica* el *análisis semántico* y la *indización*. Sin duda, tanto la *catalogación descriptiva* como la *catalogación analítica* tienen como término asociado los *datos enlazados*.

En la terminología y descriptores de 1999 se observa la separación entre *catalogación* y *metadatos*, que en la actualidad son inseparables porque es la tecnología que permite la anotación de los campos de datos de la información. Los estándares relacionados con metadatos son documentos formales que establecen criterios uniformes, métodos, procesos y prácticas. Ellos han sido creados para propósitos específicos: guiar el diseño, la creación e implementación de la estructura de datos, los valores de los datos, los contenidos de los datos y el intercambio de datos de manera eficiente y consistente.²¹

Incorporar al nuevo tesoro los estándares de metadatos es un reto difícil. Marcia L. Zeng hace una propuesta para agruparlos que puede constituir una solución, la autora divide los tipos de estándares de metadatos en cuatro: 9 y + estándares para estructura de datos; 3 y + estándares para datos de contenido; 8 y + estándares para intercambio de datos y por último 7 y + estándares para valores de datos. Por supuesto, el diagrama de Venn da por entendido que todos ellos están interrelacionados en el proceso de catalogación.

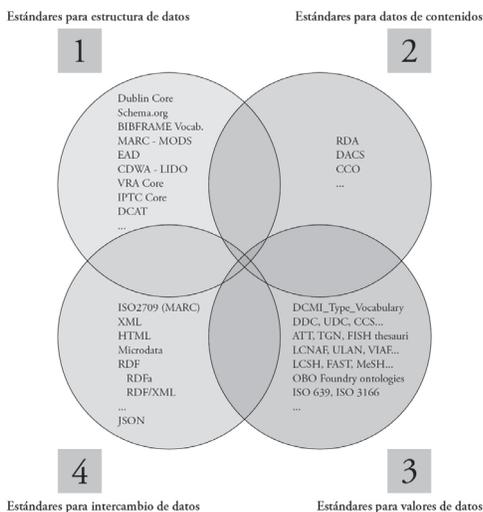


Figura 1. Adaptado de Tipos de metadatos de Marcia Lei Zeng and Jian Qin, 24.

21 Marcia Lei Zeng and Jian Qin, *Metadata*, 3.^a ed. (Chicago: ALA Neal Schuman, 2022), 23.

Como se puede observar en la figura, los metadatos usados en la conformación de los registros documentales son bastantes y tienen la particularidad que pierden actualidad y es necesario estar pendientes para observar los que están en uso. La siguiente observación sobre los metadatos es que están interrelacionados entre los que se emplean en el proceso de catalogación, por ejemplo: se aplican las RDA para diseñar el contenido, pero esos datos pueden estar estructurados con MARC-MODS; mediante el estándar RDF se preparan los datos para relacionarlos y permitir su interconexión entre los diferentes campos utilizados en registros de películas, obras escritas, cuadros, etcétera; en el campo temático, los términos que se insertan se obtienen en el MESH, cuando pertenecen al ámbito de la salud para darle valor al dato temático, o se usa el DCMI_Type_Vocabulary para explicitar el medio físico en que está presentado, ya sea papel o DVD.

Otro grupo de términos nuevos en el universo bibliográfico son los *datos enlazados*, mediante el servicio que brinda el estándar RDF. Si bien en el TELACIBIN se hallan varios términos y descriptores con *datos*, no se incluye el mencionado.

Los datos enlazados son un imprescindible en la catalogación actual y los grandes sistemas bibliotecarios se están plegando al movimiento porque es una nueva concepción del registro de información que relaciona los datos de manera que se comprenda la dimensión de lo publicado por un autor, las variedades de soportes en que puede aparecer una obra, las versiones diferentes de una obra y mucho más. La propuesta de este núcleo terminológico nuevo sería la siguiente:

Datos enlazados

- UP: Datos vinculados
- UP: Datos ligados
- TG: Catalogación
- TE: *Datos abiertos enlazados*
- TR: URIS
- HTTP
- RDF
- OWL
- SPARQL

La terminología comentada en este apartado no es la totalidad del análisis sobre la catalogación, sólo una muestra general sobre descriptores y términos

existentes en el TELACIBIN con el agregado, además, de algunos de los términos que se perfilan para incorporar al tesoro que se propone.

Entre los términos nuevos que se manejan en los modelos de datos está entidad y recurso documental, que en realidad no pertenecen al campo de la catalogación, pero están relacionados y tendrán que ser vinculados con el universo bibliográfico. Al respecto de recurso y entidad, Ávila Barrientos dice:

Cualquier cosa puede ser un recurso, incluidos los objetos físicos, documentos, conceptos abstractos, números y cadenas de texto; el término recurso es sinónimo de “entidad”, tal y como se usa en la especificación RDF Semantics RDF I I-MT.²²

La opinión vertida por este autor supone revisar el término recurso documental para definirlo probablemente como un término específico de recurso de información, que refleja una mayor apertura.

DISCUSIÓN DE LOS RESULTADOS

En el nuevo entorno de la web, la *asignación de metadatos*, el *etiquetado de contenidos* o el *marcado de los recursos de información* es una tarea crucial para organizar, recuperar y gestionar información, con la consiguiente incrustación de nuevos términos en la tarea de catalogación.

La tarea de catalogación y, por lo tanto, su mantenimiento como uno de los grandes temas de la bibliotecología es útil en el entorno digital actual porque las relaciones dentro de una red de significados siguen siendo el atributo humano que apunta a lograr la garantía de mejorar la precisión y la confiabilidad de la indización. Se agrega que, al utilizar una red de significados común, se facilita la interoperabilidad entre diferentes sistemas de información, porque los datos registrados pueden ser comprendidos entre sistemas o plataformas de otros entornos diferentes al que fueron creados. La red de significados ayuda a captar el contexto y las relaciones semánticas entre los datos asentados para representar a los recursos de información. Una misma estructura semántica usada por diferentes sistemas de información ayuda a los usuarios a entender mejor cómo funcionan las etiquetas en relación con los contenidos, para obtener una mejor recuperación de recursos de información.

22 Eder Ávila Barrientos, *Recuperación de información con datos abiertos enlazados* (México: UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2022), 21.

...ni la catalogación, ni el control de vocabulario, ni las bibliotecas han muerto, como algunos pronosticaban, aunque haya cambiado el lenguaje y ahora se hable de etiquetado, metadatos, esquemas de metadatos, ontologías, repositorios, bibliotecas y curadores digitales. En realidad, la web semántica puede suponer a medio plazo –junto con la integración entre datos y documentos y el desarrollo de agentes automáticos capaces de explotar los nuevos formatos– una inmensa “re-bibliotecarización” de la internet –si se me permite el neologismo–, aunque bajo bases y protagonistas nuevos, en un entorno en el debemos ganarnos el sitio en colaboración y competencia con otros profesionales.²³

El tema de la catalogación y el acercamiento a ella de los buscadores, para solventar muchas de las situaciones con la finalidad de lograr efectividad en la comunicación documental, que no logran sólo usando los algoritmos para obtener las ganancias esperadas, es de gran preocupación para García Marco e insiste en otro trabajo:

...algo ha cambiado: la labor no se realizará centralizadamente por ahora en los grandes servicios de búsqueda de internet. Se trata ahora de que cataloguen los productores, no los agregadores. El que está en la parte de abajo de la pirámide trabaja; el que está en la parte de arriba, dirige y recoge. O, dicho de forma más neutra, que cada uno se centre en lo suyo. ¡Bienvenidos al *cataloguing in publication* (cip) de internet!²⁴

La catalogación y, por lo tanto, la asignación de metadatos es mediante el uso y aplicación de formatos que lo permitan, pero también se necesita un llenado de contenidos, teniendo en cuenta criterios para organizar los datos pensando, por ejemplo, en los temas tratados, las relaciones jerárquicas y asociativas que ofrecen esquemas conceptuales como los tesauros o las ontologías.

Uno de los grandes problemas a resolver es entender cuáles tecnologías son útiles en la catalogación para integrarlas en el tesoro que se construye, esperando que se mantengan un buen tiempo los mecanismos que se están tratando de imponer en la actualidad. La agrupación propuesta por Marcia L. Zeng es interesante en el modo que divide los estándares de metadatos y hace más

23 Francisco Javier García Marco, “Educación y aprendizaje de la información y la documentación: raíces, desafíos y líneas de acción”, *El profesional de la información*, n.º 6, 22 (noviembre-diciembre 2013), 492, <http://dx.doi.org/10.3145/epi.2013.nov.01>

24 Francisco Javier García Marco, “Schema org: la catalogación revisitada”, *Anuario ThinkEPI* 7 (2013), 172.

fácil la revisión constante de sus nombres, ya que unas tecnologías surgen y otras se desvanecen.

La cantidad de abreviaturas de los estándares de metadatos que se manejan hace difícil hallar todas las útiles en la catalogación, pero no se pueden soslayar para ayudar a los catalogadores a tenerlos presentes en los contenidos reflejados en los sistemas de información.

CONCLUSIONES

El registro documental de la información que genera la humanidad, en cualquier formato y sobre cualquier soporte, sigue siendo fundamental en un mundo lleno de datos que la tecnología pretende manipular a través de las computadoras, pero estas necesitan datos verificados por el ser humano para funcionar, por lo que se ha observado hasta el presente en los sistemas de información.

La catalogación a partir de la publicación de un tesoro en 1999 es una de las áreas de conocimiento en la Bibliotecología y los Estudios de la Información que ha cambiado en su quehacer y, como consecuencia, ha incrementado su terminología y, por lo tanto, sus descriptores, pero que también ha disminuido otros términos y otros descriptores.

En el tesoro de 1999 se observaba la presencia de muchísimas abreviaturas usadas para nombrar las tecnologías aplicables en el tema, pero no llegaban a la cantidad actual que es desmesurada. Es un tema difícil, porque así como aparecen nuevas tecnologías otras caen en desuso, y es un tema que debe ser revisado con frecuencia para no confundir a los usuarios del lenguaje bibliotecológico.

No menos difíciles son los cambios en la estructura cognoscitiva, que suponen una responsabilidad muy especial porque desacomodan una concepción que había sido usada, pero que ya no es funcional. Sin embargo, la masa de usuarios que abordan las diferentes áreas de la disciplina se desconcierta y es necesario ser cuidadosos con las nuevas propuestas.

El avance de la tecnología en la web semántica para lograr su funcionamiento ha sido constante y seguramente lo será más. La inserción de la inteligencia artificial con el uso de los transformers comenzó hace relativamente poco tiempo y continuamente recibimos noticias sobre su avance. Esto se refleja en el vocabulario de la disciplina insertado en el ámbito de la información, los datos y la transferencia de conocimientos documentados que son los elementos que definen las relaciones de la sociedad actual.

REFERENCIAS BIBLIOGRÁFICAS

- Ávila Barrientos, Eder, *Recuperación de información con datos abiertos enlazados*, México: UNAM - Instituto de Investigaciones Bibliotecológicas y de la Información, 2022.
- Barité, Mario G., “Garantía literaria y normas para construcción de vocabularios controlados: aspectos epistemológicos y metodológicos”, *Scire: representación y organización del conocimiento* 15, n.º 2 (2009), 13-24, <https://doi.org/10.54886/scire.v15i2.3709>
- Beghtol, Clare, “Universal concepts, cultural warrant and cultural hospitality”. En *Proceedings of the 7th International isko Conference*, López-Huertas, M.J., editor, Granada: Würzburg: Ergon Verlag, 2002, 45-49.
- “Conociendo MARC Bibliográfico: Catalogación Legible por Máquina”, acceso 29 de mayo de 2024, <https://www.loc.gov/marc/umbpsal/>
- Fuentes Romero, Juan José y Vanesa Fernández Rodríguez, “Una revisión bibliográfica de los estudios comparativos. Su evolución y aplicación a la ciencia de las bibliotecas”, *Revista Interamericana de Bibliotecología* 32, n.º 2 (2009), 411-433.
- García-Marco, Francisco-Javier, “Educación y aprendizaje de la información y la documentación: raíces, desafíos y líneas de acción”, *El profesional de la información* 22, n.º 6 (noviembre-diciembre 2013), 489-504, <http://dx.doi.org/10.3145/epi.2013.nov.01>
- García-Marco, Francisco Javier, “Schema org: la catalogación revisitada”, *Anuario ThinkEPI* 7 (2013), 169-172.
- García-Marco, Francisco Javier, Carlos G. Figuerola y María Pinto, “Análisis de la evolución temática de la investigación sobre Información y Documentación en español en la base de datos LISA mediante modelado temático (1978-2019)”, *Profesional de la información* 29, n.º 4 (2020), 1-23, <https://doi.org/10.3145/epi.2020.jul.27>
- Guerrini, Mauro, *From Cataloguing to Metadata Creation: A Cultural and Methodological Introduction*, London: Facet Publishing, 2023, <https://ebookcentral.proquest.com/lib/unam/detail.action?docID=7109478>

- IFLA, ISBD, Descripción Bibliográfica Internacional Normalizada: Actualización de 2021 a la edición consolidada de 2011, aprobada por el Grupo de Revisión de ISBD de IFLA, aprobada por el Comité Consultivo de Estándares de IFLA, febrero 2022, <https://repository.ifla.org/bitstream/123456789/2491/1/ISBD%20consolidada%20Espa%C3%B1ol%202011-%20Act2021.pdf>
- Jiménez Pelayo, Jesús y Rosa García Blanco, *El catálogo de autoridades: creación y gestión en unidades documentales*, Gijón: Ediciones Trea, 2002.
- “MARCXML MARC 21 XML esquema oficial Web Site”, acceso 29 de mayo de 2024, <https://www.loc.gov/standards/marcxml/>
- Merčun, Tanja, Maja Žumer y Trond Aalberg, “Presenting bibliographic families: Designing an FRBR-based prototype using information visualization”, *The Journal of documentation* 72, n.º 3 (2016), 490-526, doi: 10.1108/JD-01-2015-0001
- Mochón-Bezares, Gonzalo, Eva Méndez-Rodríguez y Ángela Sorli-Rojo, “Etiquetado social y blog-scraping como alternativa para la actualización de vocabularios controlados: Aplicación práctica a un tesoro de Biblioteconomía y Documentación”, *Información, cultura y sociedad* 37 (diciembre 2017), 13-26, <http://repositorio.filo.uba.ar/handle/filodigital/11218>
- Naumis Peña, Catalina, María Texia Iglesias Maturana, Ana Cecilia Osuna Dumont, Lucy Espinosa Ricardo, *Tesoro latinoamericano de ciencia bibliotecológica y de la Información*, México: Centro Universitario de Investigaciones Bibliotecológicas, 1999.
- “Overview of the BIBFRAME 2.0 Model (BIBFRAME-Bibliographic Framework Initiative, Library of Congress)”, acceso el 29 de mayo de 2024, <https://www.loc.gov/bibframe/docs/bibframe2-model.html>
- Pastor Sánchez, Juan Antonio, “Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data”, *Scire* 19, n.º 2 (jul.-dic. 2013), 55-68, doi:10.54886/scire.v19i2.4112

- Pisanski, Jan y Marcia Žumer, “Mental Models of the Bibliographic Universe. Part 1: Mental Models of Descriptions; Part 2: Comparison Task and Conclusions”, *Journal of Documentation* 66, n.º 5 (2010), 643-667, 668-680.
- Rodríguez-Bravo, Blanca, Maria-da-Graça Simões, Maria-Cristina Vieira-de-Freitas, José-Antonio Frías, “Descubrimiento de información científica: ¿todavía misión y visión de la biblioteca académica?”, *El profesional de la información* 26, n.º 3, 464-479, <https://doi.org/10.3145/epi.2017.may.13>
- Taylor, Arlene G., *The organization of information*, 2.ª ed., Westport, Conn.: Libraries Unlimited, 2004.
- Tillett, Barbara, *RDA: antecedentes y aspectos de su implementación: manual para instructor*, Conferencia presentada en febrero de 2009, acceso 29 de mayo de 2024, https://www.loc.gov/catdir/cpsd/RDA/RDAantecedentes_instructor.pdf
- Tramullas, Jesús, “Desarrollos en elaboración de documentación técnica: los lenguajes de marcado ligero”, *Anuario ThinkEPI* 13 (2019), 1-5, doi.org/10.3145/thinkepi.2019.e13f03
- Zeng, Marcia Lei y Jian Qin, *Metadata*, 3.ª ed., Chicago: ALA Neal Schuman, 2022.

Distribución de relaciones léxicas como indicador de propiedades estructurales en vocabularios controlados: el caso de la sinonimia

Distribution of lexical relations as an indicator of structural properties in controlled vocabularies: the case of synonymy

Diego Ferreyra
Facultad de Filosofía y Letras
Universidad de Buenos Aires

Los vocabularios controlados constituyen un tipo específico de sistema de organización del conocimiento (soc). Delimitan y formalizan un dominio estableciendo los términos que lo conforman, el alcance conceptual de cada uno de ellos y las relaciones que los articulan. Se trata de dispositivos de mediación social que basan su productividad y agencia en su capacidad para designar de manera no ambigua un conjunto de referentes en el contexto de un dominio y articularlos a través de una red coherente de relaciones léxico-semánticas. Entre las relaciones léxicas contempladas se encuentra la relación de sinonimia, entendida como la presencia en el lexicón de diversas formas lingüísticas con un significado similar o equivalente en el contexto de un dominio específico. Este artículo presenta un modelo de análisis y evaluación que permite definir una proporción óptima entre términos alternativos y preferentes en el contexto global de un vocabulario controlado.

Estudiaremos modelos que abordan el léxico como una red autoorganizada y se presentará una serie de modelos aplicables al contexto de vocabularios controlados. Finalmente, explicaremos la implementación de estos criterios en la herramienta de gestión de vocabularios controlados TemaTres.

Nocetti y Figueiredo (1978) relevan los puntos de contacto entre los vocabularios controlados y los lenguajes naturales. En su trabajo identifican, entre otros aspectos, el caso de relaciones léxicas entre términos como uno de los fenómenos de estudio compartidos. La sinonimia –y la polisemia– constituyen un aspecto estructural en los lenguajes naturales y un componente funcional

de los vocabularios controlados. En los lenguajes naturales estos procesos forman parte de las dinámicas que proveen elasticidad, economía y polivalencia como dispositivo sociocomunicacional. En el caso de los vocabularios controlados, la sinonimia forma parte de los fenómenos que pretenden corregir y remediar, la diversidad de términos referidos a un mismo concepto es abordada como una falla que debe ser subsanada a través del control (ANSI/NISO Z39.19:2005 [R2010]; ISO 25964-1:2011). Observando este particular punto de articulación, creo que es posible realizar un acercamiento, definir aspectos distintivos, formular algunas preguntas e intentar explorar herramientas para el análisis y la elaboración de sistemas de organización del conocimiento basados en vocabularios controlados.

LA SINONIMIA COMO RELACIÓN LÉXICA

La sinonimia es una relación entre unidades léxicas caracterizada por la verificación de la convergencia de significados entre palabras. Los procesos de sinonimia establecen una relación de equivalencia entre los significados léxicos de dos o más formas léxicas. Por tratarse de un proceso omnipresente en todos los actos de habla, todos los hablantes disponen de una aproximación nocional de la sinonimia.

Existen diversos enfoques para determinar la sinonimia como relación léxica. Desde la perspectiva de la semántica composicional, la sinonimia puede ser explicada como la coincidencia entre los rasgos de sentido composicionales de dos unidades léxicas. Desde una perspectiva funcional, la sinonimia se verifica y estudia en los usos y la posibilidad de sustitución entre ítems léxicos en contextos oracionales. Desde una perspectiva contextualista, las relaciones de sinonimia se analizan con base en la recurrencia de ítems léxicos en corpus textuales. Más allá de los marcos teóricos, es posible establecer un acuerdo en la consideración de la sinonimia como una relación caracterizada por la convergencia de diversas formas verbales sobre un significado. Este punto resulta central en la presente indagación: entender la sinonimia como la red de formas o unidades léxicas que se organizan alrededor de un mismo significado o concepto. Es pues una relación que agrupa una diversidad de entidades que en el contexto del lenguaje conforman la dimensión más concreta y material posible –como lo son los significantes o las formas verbales– en vinculación con un significado o concepto. Es quizás la dimensión más material de los términos al estar basada en relaciones entre diversos significantes y un significado y no estrictamente en las relaciones entre significados (Casas Gómez 1999; García Jurado 2003), como es el caso de otras relaciones léxicas (hiperonimia-

hiponimia, meronimia, términos asociados). Si la polisemia opera a través de la acumulación de significados en un mismo significante, la sinonimia, como proceso, acumula significantes en torno a un significado.

La onomasiología es el estudio del significado con respecto a los distintos significantes que pueden designarlo (García Jurado 2003). En términos de actividad terminológica, podemos considerarla como un derrotero que parte de un significado establecido en un dominio específico y que busca, selecciona y asigna un significante puntual. Si la sinonimia explora la posibilidad de convergencia entre diferentes términos y un único significado, es pues la relación onomasiológica por excelencia.

Nos interesa concebir la relación sinonímica como un proceso, es decir, como una relación dinámica en el tiempo y en el contexto de los conceptos y significados que organizan un dominio o una situación comunicativa.

En términos dialógicos, es decir, a través del tiempo, tal como Casas Gómez (1999), se analizan diferentes estados sinonímicos en una determinada lengua siguiendo diversos procesos (semantización/desemantización, irradiación o derivación semántica, etcétera) e incluso es posible identificar momentos en los que una palabra es utilizada para expresar significados opuestos (antónimos) o para reunir diferentes hipónimos (es decir, términos más específicos) bajo un mismo término. Es así que en determinados momentos una serie de palabras comparten un mismo significado y, eventualmente, a lo largo del tiempo pueden ir diferenciándose hasta incluso conformar series con significados contrarios (Abel 1988).

La cantidad de sinónimos posibles para un mismo concepto o significado varía según sea la lengua, el dominio o la situación comunicativa (Casas Gómez 1999; Strori, Bombaci y Bingol 2007; Makaruk y Owczarek 2008; ISO 25964-1:2011).

Desde una perspectiva psicolingüística, la sinonimia se explica como una relación gradual y continua entre significados y definida en términos de semejanzas en el contexto de una red de ítems léxicos. Esto implica que se verifica experimentalmente (Horno-Chéliz, Timor y Sarasa 2017) que el lexicón mental se encuentra organizado como una red en la que todas las piezas léxicas difieren –ya sea a través de matices, contextos de aplicación, registro de uso u otros aspectos– y donde la relación de sinonimia es de naturaleza gradual en términos de gradientes de semejanza y no de identidad. Eventualmente estas aproximaciones establecen que un término en el contexto del lenguaje natural no opera como una etiqueta léxica que se adosa a un concepto, sino que se desprende de la red de relaciones que organiza un léxico o vocabulario (Zapico y Vivas 2014). Esto presenta un conjunto de problemas y promesas para el campo de los vocabularios controlados y los sistemas de organización

del conocimiento, problemas que no son soslayados pero que serán abordados en futuros trabajos.

Se observa, pues, que la sinonimia se presenta como un proceso fluido, dinámico y situado ya sea que sea abordado como un fenómeno a través del tiempo, como un aspecto de la lengua en tanto sistema o desde un enfoque psicolingüístico. Corresponde analizar la función de esta relación léxica y su operacionalización en el contexto de los vocabularios controlados como componentes de un sistema de organización del conocimiento (soc).

Los soc son dispositivos de mediación social singularizados por su carácter formal, sistemático y orientado a objetivos específicos (Ferreya 2021; Glushko 2014). En ese contexto, los vocabularios controlados constituyen un componente articular que organiza un dominio o espacio semántico a través de una red de relaciones léxico-semánticas coherente, es decir, relaciones no contradictorias entre sí según un régimen conceptual definido. En ese contexto, las relaciones de equivalencia léxica –como la sinonimia– constituyen un recurso central que permite diversificar el espectro de ítems léxicos que refieren un mismo significado, aportando mayor asertividad léxica y, por lo tanto, disminuyendo el nivel de ambigüedad y mejorando, por lo tanto, las capacidades para operar en un dominio dado.

La ambigüedad es un problema en cualquier contexto comunicativo. El lenguaje natural recurre a la redundancia –en registros diversos y concurrentes– para resolver ambigüedades y proveer eficacia comunicativa. Los vocabularios controlados, en cambio, se caracterizan por recurrir a un esquema de coherencia estructural en la red de relaciones léxicas. Esta coherencia estructural permite definir y diferenciar cada valor léxico de manera interdependiente. Los lenguajes artificiales en general tienden a un ideal modélico en el que cada término delimita con otros, conformando conjuntos diferenciados e interdependientes (Geckeler 1976). Sin embargo, es importante remarcar que se trata de una orientación, un modelo ideal que como tal se encuentra en permanente tensión y torsión con los procesos comunicativos basados en el lenguaje.

Aun así, esta es una de las fortalezas distintivas de los vocabularios controlados y los léxicos en general: la portabilidad operativa que le permite intervenir y asignar significados con relativa autonomía con respecto a diversos contextos. Un vocabulario controlado no varía en cada situación comunicativa, en cada registro o según los interlocutores y –en los casos óptimos– resuelve procesos de asignación de significados con eficacia. Es aquí donde las relaciones de equivalencia tienen un rol central en la consolidación de la eficacia de un vocabulario controlado o léxico. La relación entre términos preferentes y términos alternativos define el territorio total del campo semántico entendido como inventario léxico del dominio. A través de las relaciones de equivalencia

léxica se establece y delimita la totalidad de lo que puede ser denominado en el contexto del dominio. Las otras relaciones léxicas (relaciones jerárquicas de inclusión, relaciones asociativas) estructuran el léxico, organizan el territorio pero no amplían el contorno lexical. Aunque parece ser una relación léxica desjerarquizada y considerada de menor complejidad, la relación léxica de sinonimia define el inventario total de material léxico. Las relaciones de sinonimia resultan en ocasiones subvaluadas como instrumentos menores, considerados estamentos primarios en la jerarquía de complejidad de los vocabularios controlados (ANSI/NISO Z39.19:2005 [R2010]; Zeng 2008) y, sin embargo, tienen un rol preeminente y son el común denominador de todas las tipologías de léxicos y vocabularios controlados.

Consideremos que el mundo en general o un dominio en particular como destino referencial constituyen entidades infinitas, mientras que el repertorio terminológico consolidado en un vocabulario controlado es un fenómeno finito. En tal sentido, la capacidad de un vocabulario controlado para operativizar un dominio o proyecto comunicacional es justamente establecer un espacio finito de significados a través de un léxico puntual y discreto.

El fenómeno del proceso de sinonimia expresa como ningún otro la materialidad del lenguaje. Diversas formas pueden expresar un mismo significado y eventualmente denotar un mismo referente, pudiendo variar según el dominio, el lenguaje o el contexto comunicacional.

Se nos presenta entonces una pregunta importante para el campo de los vocabularios controlados y los SOC: ¿existe una distribución de sinónimos característica en el contexto de un léxico? Si todos los términos pueden ser expresados por uno o más sinónimos, ¿sería deseable o correcto que todo término disponga de un sinónimo?, ¿cada término debería disponer de más de un sinónimo?, ¿cuántos? Estas preguntas giran en torno de una pregunta subyacente: ¿existe un patrón o modelo ideal de distribución de sinónimos en el léxico? Y si existiera, ¿qué tipo de indicador es y cómo debería interpretarse?

Los sinónimos no se distribuyen de manera uniforme en el léxico, este fenómeno se verifica tanto en el lenguaje natural como en los lenguajes controlados (Casas Gómez 1999; Strori, Bombaci y Bingol 2007; Makaruk y Owczarek 2008). Quizás la noción misma de clasificación como actividad pueda ofrecer un modelo explicativo de este fenómeno. La clasificación es, en los términos más generales posibles, el establecimiento de una discontinuidad sobre el continuo de la realidad, implica obviar ciertas diferencias para priorizar determinados rasgos salientes que nos permitan establecer divisiones (Bowker y Star 2000; Bauman 1996). Los procesos de sinonimia implican establecer una indiferenciación entre significantes para un determinado contexto comunicacional o dominio, crear una continuidad material entre varios

ítems léxicos en relación con un concepto. El proceso sinonímico generaliza y obvia diferencias entre formas léxicas en torno a un referente o concepto. Articula pues dos conjuntos no congruentes: agrupa una diversidad de significados como una unidad bajo una variedad de ítems léxicos.

Es así que la sinonimia expresa un aspecto central del conocimiento y del lenguaje en un dominio o contexto comunicacional: identifica los espacios en los que un dominio prefiere ampliar el rango de amplitud léxica para un significado que no debería ser segmentado. Es un indicador de la necesidad de una comunidad de prácticas de diversificar y extender el material verbal disponible para abarcar un campo semántico puntual. La presencia de sinónimos resulta ser un indicador útil para detectar la densificación del material léxico en una determinada zona del conocimiento de una comunidad o proyecto. Es un analizador que permite reconocer las zonas en las que una comunidad de prácticas o dominio opone una continuidad semántica frente a la evidencia de discontinuidades léxicas en las prácticas sociales en el seno de un dominio dado.

Hemos podido observar la importancia de las relaciones de sinonimia en la conformación del universo material y léxico de un vocabulario controlado. Reconocer los componentes de la dimensión material del léxico en nuestro campo resulta central, ya que constituyen el cuerpo y la geografía sobre las cuales habremos de diseñar y elaborar las operaciones de identificación, representación, búsqueda y organización de los artefactos culturales que conforman nuestros entornos de trabajo. Hemos observado también que los procesos de sinonimia exponen un campo de tensiones en el que es posible reconocer tanto concentraciones como vacíos. Ha llegado el momento de explorar la estructura de estas afinidades.

REDES DE RELACIONES LÉXICAS

Es posible analizar el léxico como una red de relaciones en la que cada palabra es un nodo y cada relación léxica un arco que la vincula a través de una relación específica con otra palabra (Cancho y Solé 2001; Sigman y Cecchi 2001; Zapico y Vivas 2014). Si en el lenguaje natural estas relaciones pueden analizarse a partir de la interacción recurrente entre palabras, en el caso de los léxicos y vocabularios controlados el análisis resulta más sencillo. En nuestro caso los términos –no las palabras– son los nodos y sus relaciones los vínculos que los conectan. Es así que los vocabularios controlados pueden analizarse como un espacio de nodos y conexiones: cada término/nodo puede establecer más de tipo de relación léxica y cada relación dispone de una direccionalidad (Holanda, Torres Pisa, Kinouchi, Souto Martínez y Seron Ruiz 2004).

Para avanzar en este análisis resulta necesario recurrir a ciertos conceptos propios de la teoría de grafos. Esta teoría parte de la consideración de que un grafo es una representación que se compone de una serie de nodos –también llamados vértices– y vínculos –también llamados arcos– que relacionan pares de nodos. Si las relaciones entre nodos expresan una direccionalidad, será un grafo dirigido, caso contrario será un grafo no dirigido. Con estos elementos es posible caracterizar una red según aspectos como ser el promedio de pasos entre cualesquiera dos nodos (distancia promedio), el diámetro máximo de la red (la distancia máxima entre dos nodos) o coeficiente de agrupamiento –puede ser entendido como el promedio de relaciones de los nodos de una red (Steyvers y Tenenbaum, 2005).

Los procesos de sinonimia en lenguajes naturales se analizan como un grafo no dirigido en el que un conjunto de palabras forman parte de una red vinculada por un significado similar. Es así que reciben a veces la denominación de synset o de anillos de sinónimos (ISO 25964-1:2011; ANSI/NISO Z39.19:2005 [R2010]). En una red de este tipo, los nodos establecen relaciones directas y simétricas entre sí (Strori, Bombaci y Bingol 2007). Un grafo no dirigido de sinónimos podría graficarse de la siguiente manera (Figura 1):

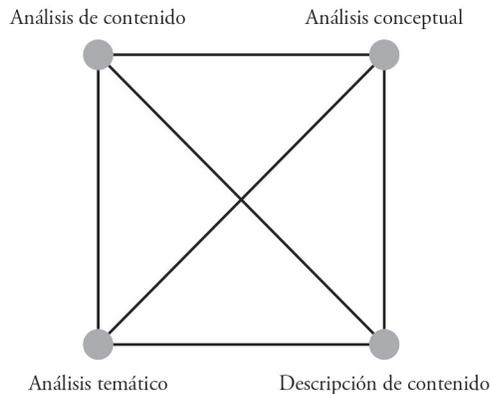


Figura 1.

Por el contrario, en vocabularios controlados orientados a procesos de indicación, se trata de un grafo dirigido, ya que se establece un término preferente y un conjunto de posibles términos equivalentes alternativos. En una red de este tipo un nodo –el término preferente– establece relaciones directas y asi-

métricas hacia otros nodos –los términos alternativos. Un grafo dirigido de sinónimos podría graficarse de la siguiente manera (Figura 2):

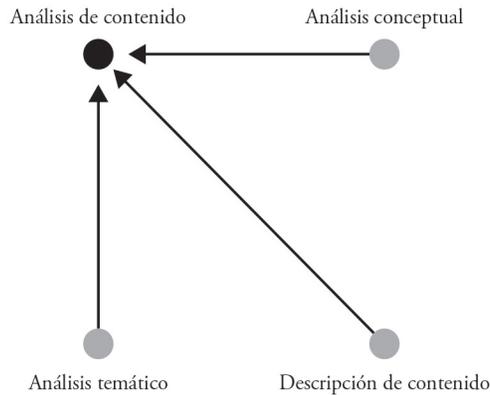


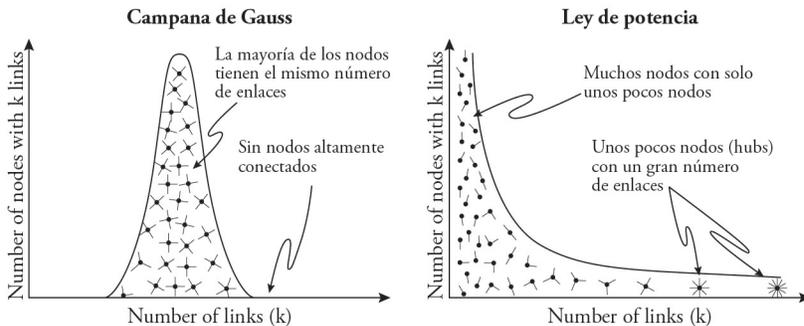
Figura 2.

En este sentido, un vocabulario controlado puede ser visto como la documentación formal de la red léxica que caracteriza un dominio.

Más allá de variaciones en los métodos, corpus o lenguas, se verifica en las investigaciones relevadas que las redes de relaciones léxicas en el lenguaje presentan propiedades compatibles con el modelo de redes libres de escala (Budel *et al.* 2023; Cancho y Solé 2001; Csermely 2006; Fóris 2013; Makaruk y Owczarek 2008; Sigman y Cecchi 2001; Steyvers y Tenenbaum 2005; Stori, Bombaci y Bingol 2007; Zapico y Vivas 2014). Las redes libres de escala formalizan un modelo presentado inicialmente por Barabási y Albert (1999) en el que no existe un patrón homogéneo en la distribución de enlaces por nodo, sino que un número acotado de nodos concentra proporcionalmente la mayor cantidad de vínculos, en tanto la mayoría de los nodos establecen pocos vínculos. Es un modelo representado gráficamente como una larga cola donde no es posible establecer un número característico o predominante de nodos según cantidad de relaciones. Este patrón, en términos de representación gráfica, resulta compatible con la distribución resultante de una relación de potencia entre 2 magnitudes (ley de potencia) o con la distribución de Pareto (Newman 2005). Este tipo de redes se verifica en un sinnúmero de fenómenos y actividades, como ser las citas científicas, la frecuencia de palabras, frecuencia y duración de llamadas telefónicas, diámetros de cráteres o el tamaño de las ciudades (Barabási 2003; Newman 2005; Solé 2009). Y en

el análisis de esta recurrencia se identifican una serie de aspectos característicos como una singularidad de las redes libres de escala (Figura 3):

- Conforman redes caracterizadas por la robustez como capacidad de tolerancia frente errores o variaciones en la cantidad de nodos o las conexiones (Fóris 2003).
- Se presenta siempre el fenómeno de *mundos pequeños*, esto significa que es posible acceder en pocos pasos entre nodos aparentemente distantes entre sí y entre nodos poco conectados; pese a que la mayoría de los nodos dispone de pocos enlaces, la existencia de hubs o concentradores permite disminuir drásticamente la distancia media entre dos nodos cualesquiera en la red (Watts y Strogatz 1998).
- Se autoensamblan y evolucionan nodo a nodo según el principio de conexión preferencial (*preferential attachment*); este principio considera que cada nodo que se suma a la red tiende a seleccionar nodos más densamente conectados; resulta un principio relevante por su capacidad explicativa de fenómenos evolutivos en grandes estructuras no coordinadas hasta la emergencia de patrones organizados robustos y sistémicamente estables (Klemm y Eguiluz 2001; Barabási 2003; Solé 2009; Fóris 2013).
- Permite describir globalmente una estructura compleja con una distribución proporcional consistente con la denominada ley de Pareto –también conocida como principio 80-20– según la cual el 80% de los recursos o eventos de algún tipo se concentra en el 20% de actores o factores (Barabási 2003; Newman 2005; Makaruk y Owczarek 2008).



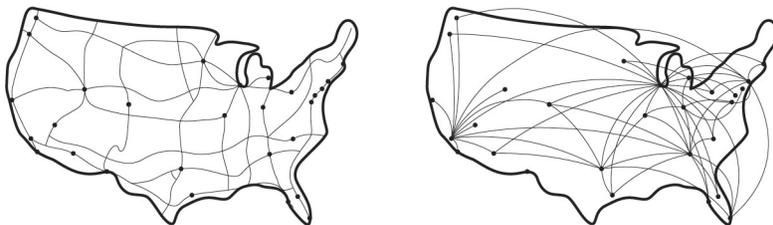


Figura 3. La distribución de enlaces en redes aleatorias se refleja como una curva de campana donde la mayor parte de los nodos tienen similar cantidad de enlaces. Las redes libres de escala –con una distribución acorde con la ley de potencia– concentran en pocos nodos la mayor cantidad de enlaces. Fuente: Barabási 2003.

Reconocer la conformación de redes compatibles con el modelo de redes sin escala en el lenguaje es relevante, especialmente por los trabajos que verifican este fenómeno en particular en el caso de las relaciones de sinonimia (Strori, Bombaci y Bingol 2007; Makaruk y Owczarek 2008).

Determinar distribuciones de relaciones léxicas que se verifican en cualquier lengua natural permite suponer que se trata de rasgos sistémicos o adaptativos que, a través de un sinnúmero de reiteraciones y variaciones, ofrecen condiciones de tolerancia, robustez y eficiencia. Si se pudieran identificar los aspectos que caracterizan la red de relaciones léxicas resultante, sería eventualmente posible aplicar los mismos patrones estructurales en la elaboración de vocabularios controlados y léxicos en general. Se espera que, a diferencia de otros aspectos del lenguaje en general y de los vocabularios controlados en particular, estas características resulten independientes de la escala o magnitud, el dominio, la lengua o el contexto comunicacional. Esto permite especular con la posibilidad de establecer un indicador regular e invariante en todos los escenarios y casos (Figuras 4 y 5).

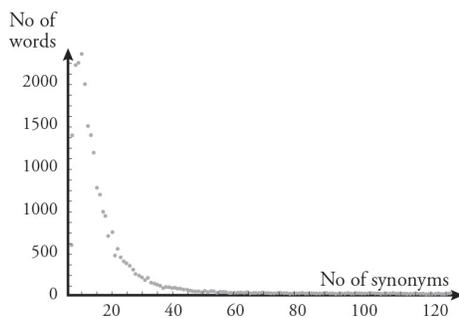


Figura 4. Número de sinónimos frente al número de palabras en polaco. Fuente: Makaruk y Owczarek, 2008.

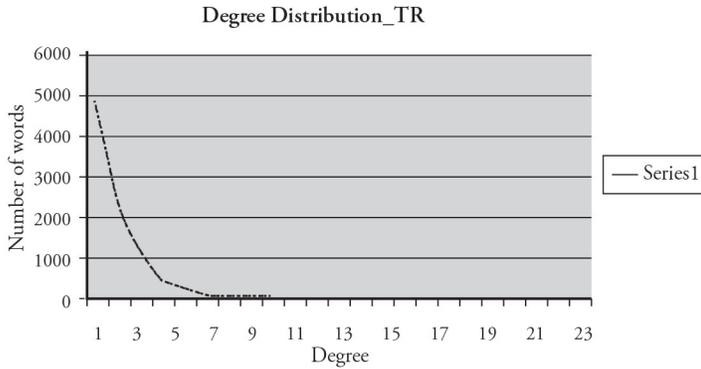


Figura 5. Número de sinónimos frente al número de palabras en turco. Fuente: Strori, Bombaci y Bingol, 2007.

INDICADORES DE DISTRIBUCIÓN DE RELACIONES LÉXICAS INCORPORADOS A TEMATRES

TemaTres es una aplicación web de gestión de vocabularios controlados.¹ Existen cientos de implementaciones, una familia de herramientas web asociadas a la explotación y reúso de servicios terminológicos² y dispone de una agenda de desarrollo activa y participativa. Tal como se señaló anteriormente, el interés de la presente exploración es proponer la consideración de indicadores asociados a la gestión de vocabularios controlados y léxicos que sean capaces de dar cuenta de propiedades estructurales y globales con base en las relaciones léxicas formalizadas en los mismos. Como se ha podido comprobar a lo largo de nuestra indagación, en el caso de las relaciones léxicas de sinonimia es posible verificar la emergencia sistémica de patrones de distribución compatible con el modelo de redes libres de escala. Es así que en la última versión disponible de TemaTres se han incorporado tres componentes informativos orientados a caracterizar la distribución global de relaciones léxicas de sinonimia y ofrecer en tal sentido indicadores que permitan evaluar el grado de adecuación de cada vocabulario, con una distribución compatible de las redes libres de escala, a saber:

1 <https://github.com/tematres/TemaTres-Vocabulary-Server/releases>

2 <https://github.com/tematres>

1. Un gráfico que describe la relación de términos preferentes según la cantidad de términos alternativos.
2. Un gráfico que describe la relación porcentual sobre el total de términos de los términos preferentes y los términos alternativos.
3. El coeficiente de agrupamiento (*clustering*) correspondiente a la red de relaciones léxicas resultante entre términos preferentes y los términos alternativos.

El primero de los indicadores permite establecer si la red de relaciones léxicas resultante describe una función compatible con la ley de potencia y, por lo tanto, podría caracterizarse como una red libre de escala. Debería verificarse que un conjunto minoritario de términos concentra la mayor cantidad de términos alternativos mientras que la mayoría de los términos dispone de un número drásticamente menor de términos alternativos.

El segundo gráfico indicador permite establecer si la distribución de relaciones de equivalencia léxica es compatible con la ley de Pareto. Dicha ley establece que el universo de términos se compone de un 80% de términos preferentes y un 20% de términos alternativos. Este gráfico resulta en realidad un orientador intuitivo, un indicador riguroso debería establecer la concentración del 80% de las relaciones de equivalencia en el 20% de los términos preferentes. Siendo que esa proporción se encuentra informada detalladamente en el indicador gráfico anterior, el presente gráfico espera orientar un criterio aproximado relativo a la necesidad de incrementar o decrementar la cantidad global de términos equivalentes.

Finalmente, el tercer indicador ofrece un coeficiente de agrupamiento, es el promedio de relaciones léxicas de equivalencia establecidas en el vocabulario. Estima en qué medida los términos del vocabulario, entendidos como una red léxica, están agrupados o conectados entre sí. El coeficiente permite establecer un valor numérico que describa la tendencia a formar grupos densamente interconectados a través de relaciones de equivalencia léxica entre términos en el contexto del vocabulario. Es un indicador cuya mayor utilidad se encuentra orientada a establecer métricas comparativas entre vocabularios controlados o indicadores evolutivos en el contexto de un mismo vocabulario controlado a través del tiempo.

Se presentan a modo de ejemplo los indicadores resultantes para el vocabulario el Tesouro Spines (Figura 6), el Tesouro latinoamericano en ciencia bi-

blotecológica y de la información (Figura 7) y el Tesauro brasileiro de ciência da informação (Figura 8).

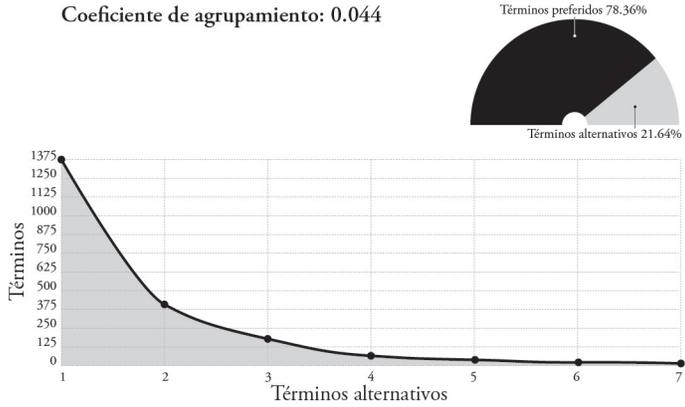


Figura 6. Términos alternativos según término. Tesauro SPINES, <https://vocabularyserver.com/spines/sobre.php>

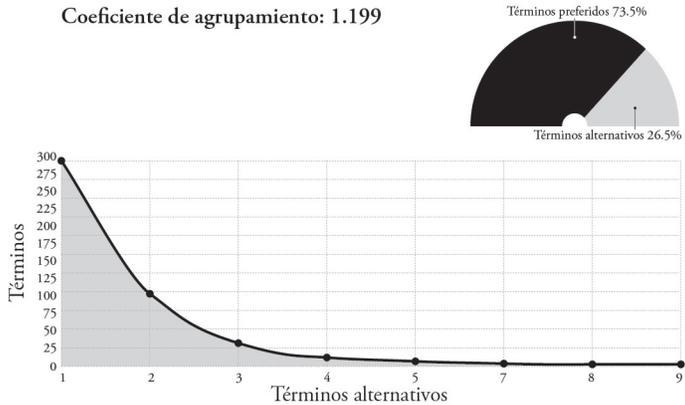


Figura 7. Términos alternativos según término. Tesauro latinoamericano en ciencia bibliotecológica y de la información, <https://vocabularyserver.com/telacibin/sobre.php>

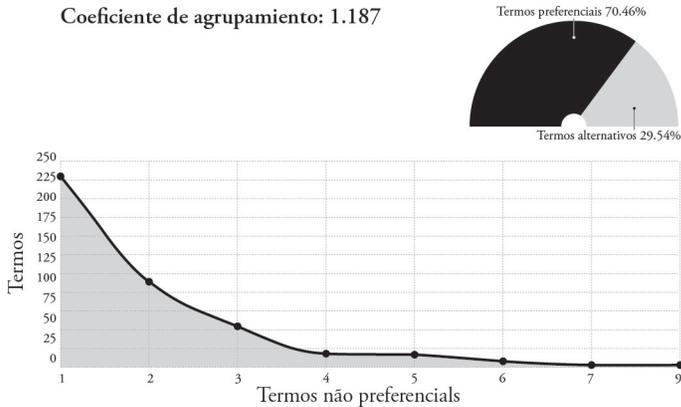


Figura 8. Términos alternativos según término. Tesauro brasileiro de ciência da informaçã, <https://vocabularyserver.com/tbci/sobre.php>

CONCLUSIONES

La construcción de una red coherente y robusta de relaciones léxicas promueve la persistencia estructural de un paradigma de sentido. En este trabajo se ha observado que las relaciones de sinonimia no se distribuyen de manera aleatoria ni homogénea. La conformación de una red de relaciones léxicas compatible con el modelo de redes libres de escala parece ofrecer mejores posibilidades para sobrellevar variaciones de todo tipo. En el caso de la sinonimia, fortalece las capacidades de un léxico para absorber las variaciones léxicas que se concentran en determinadas zonas de un dominio. El presente análisis sostiene que el análisis de las redes de relaciones léxicas de sinonimia permite identificar los *centros de atracción* en el contexto de un dominio. Establece un espectro de indiferencias entre variantes léxicas asociadas a un mismo concepto, es una forma de ampliar el alcance asertivo de un conjunto de ítems léxicos sobre un mismo vértice conceptual. Analizar la distribución de las relaciones de sinonimia permite evaluar la robustez de la red de ítems léxicos en su estamento más material posible y menos abstracto. En segundo término, facilita la diferenciación de espacios concentrados de significado en el contexto de un dominio al identificar el conjunto minoritario de términos que reúne la mayor cantidad de variaciones lexicales. En tercer lugar, permite establecer comparativas entre vocabularios controlados. Por último, considero que la disponibilidad de indicadores capaces de caracterizar vocabularios controlados como redes de relaciones léxicas abre importantes posibilidades

para nuevos interrogantes: ¿es posible establecer un análisis similar en otras relaciones léxicas adoptadas en los vocabularios controlados (relaciones de inclusión, relaciones asociativas)?, ¿existen efectos o factores articulares entre los modelos de red de cada tipo de relación léxica?, ¿la verificación de patrones de distribución de relaciones de sinonimia en diversos dominios, escalas y lenguas permitiría considerar la existencia de universalidades estructurales y variaciones contextuales según sean las zonas en las que se concentran los casos de sinonimia? El modelo permite comparar variaciones diferenciales según cada lengua en la composición de las redes de sinonimia. En el caso de los vocabularios multilingües ¿qué implicancias conlleva para el enfoque centrado en conceptos (web semántica, ISO 25964-1:2011) las diferencias entre las redes de sinonimia –que varían según cada lengua– en el contexto de un mismo vocabulario multilingüe? (Makaruk y Owczarek 2008).

REFERENCIAS BIBLIOGRÁFICAS

- Karl Abel, "Acerca del sentido antitético de las palabras primitivas", Karl Abel, Wilhelm Von Humboldt y Hans Sperber, *El psicoanálisis y las teorías del lenguaje*, Catálogos, 1998.
- Baldinger, Kurt, *Teoría Semántica: hacia una semántica moderna*, Madrid: Alcalá, 1977.
- Barabási, Albert-László y Réka Albert, "Emergence of Scaling in Random Networks", *Science* 286 (5439), 509-12, <https://doi.org/10.1126/science.286.5439.509>
- Barabási, Albert-László, *Linked: The New Science of Networks*, Cambridge: Perseus Publishing, 2003.
- Bauman, Zigmunt, "Modernidad y ambivalencia", Josetxo Beriain, comp., *Las consecuencias perversas de la modernidad: modernidad, contingencia y riesgo*, Anthropos Editorial, 1996.
- Bowker, Geoffrey C., y Susan Leigh Star, *Sorting Things Out*, Cambridge: The MIT Press, 1999, <https://doi.org/10.7551/mitpress/6352.001.0001>
- Budel, Gabriel, Ying Jin, Piet Van Mieghem, y Maksim Kitsak, "Topological Properties and Organizing Principles of Semantic Networks", *Scientific Reports* 13 (1), <https://doi.org/10.1038/s41598-023-37294-8>
- Cancho, Ramon Ferrer I. y Ricardo Vicente Solé, 2001, "The Small World of Human Language", *Proceedings of the Royal Society B. Biological Sciences* 268 (1482), 2261-65, <https://doi.org/10.1098/rspb.2001.1800>
- Casas Gómez, Miguel, *Las relaciones léxicas*, Walter de Gruyter, 2015.
- Csermely, Peter, *Weak Links: The Universal Key to the Stability of Networks and Complex Systems*. Springer, 2009.
- Ferreya, Diego, "Sistemas de organización del conocimiento: propuesta de un modelo unificado de definición", *Palabra Clave* 11 (1), e141, <https://doi.org/10.24215/18539912e141>
- Geckeler, Horst, *Semántica estructural y teoría del campo léxico*, Barcelona: Gredos, 1976.
- Glushko, Robert J. *The Discipline of Organizing: Professional Edition*, O'Reilly Media, Inc., 2016.
- Holanda, Adriano de Jesus, Ivan Torres Pisa, Osame Kinouchi, Alexandre Souto Martinez y Evandro Eduardo

- Seron Ruiz, "Thesaurus as a Complex Network", *Physica*, A 344 (3-4), 530-36, <https://doi.org/10.1016/j.physa.2004.06.025>
- Horno-Chéliz, María del Carmen, Raquel Timor y Antonio Sarasa Cabezuelo, "¿Qué ocurre cuando comparamos dos unidades léxicas sinónimas? Un estudio psicolingüístico sobre la naturaleza de la sinonimia", *RLA, Revista de lingüística teórica y aplicada* 55 (1), 149-68, <https://doi.org/10.4067/s0718-48832017000100149>
- ISO, ISO 25964-1:2011 Information and documentation - *Thesauri and interoperability with other vocabularies* - Part 1: Thesauri for information retrieval, 2011.
- Jurado, Francisco García, *Introducción a la semántica latina: de la semántica tradicional al cognitivismo*, Madrid: Universidad Complutense de Madrid, 2003.
- Klemm, Konstantin y Víctor M. Eguíluz, "Highly Clustered Scale-free Networks", *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 65 (3), <https://doi.org/10.1103/physreve.65.036123>
- Makaruk, Hanna E. y Robert Owczarek, "Hubs in Languages: Scale Free Networks of Synonyms", arXiv.Org, February 28, 2008, <http://arxiv.org/abs/0802.4112>
- National Information Standards Organization, ANSI/NISO Z39.19:2005 [R2010], *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, 2005.
- Newman, M. E. J., "Power Laws, Pareto Distributions and Zipf's Law", *Contemporary Physics* 46 (5), 323-51, <https://doi.org/10.1080/00107510500052444>
- Nocetti, Milton A. y Regina Célia Figueredo, "Línguas naturais e linguagens documentárias: traços inerentes e ocorrências de interação", *R. Bibliotecon, Brasília* 6 (1); 23-27, jan./jun.1978.
- Sigman, Mariano y Guillermo A. Cecchi, "Global Organization of the Wordnet Lexicon", *Proceedings of the National Academy of Sciences of the United States of America* 99 (3), 1742-47, <https://doi.org/10.1073/pnas.022341799>
- Solé, Ricard, *Redes complejas: del genoma a Internet*, Tusquets Editores, 2016.

- Steyvers, Mark y Joshua B. Tenenbaum, “The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth”, *Cognitive Science* 29 (1), 41-78, https://doi.org/10.1207/s15516709cog2901_3
- Strori, Dorina, Ahmet Bombaci y Haluk Bingol, “Cross Comparison of Synonym Graphs in a Multi Linguistic Context”, arXiv.Org, September 16, 2007, <https://arxiv.org/abs/0709.2476>
- Watts, Duncan J. y Steven H. Strogatz, “Collective Dynamics of ‘Small-world’ Networks”, *Nature* 393 (6684), 440-42, <https://doi.org/10.1038/30918>
- Zapico, Martín y Jorge Vivas, “La sinonimia como caso particular de distancia semántica”, *Encontros Bibli* 19 (40), 253, <https://doi.org/10.5007/1518-2924.2014v19n40p253>
- Zeng, Marcia Lei, “Knowledge Organization Systems (kos)”, *Knowledge Organization* 35 (2-3), 160-82, <https://doi.org/10.5771/0943-7444-2008-2-3-160>

La enseñanza de la construcción de tesauros: experiencias y reflexiones

Teaching thesaurus construction: experiences and reflections

Francisco Javier García Marco
*Universidad de Zaragoza,
Área de Biblioteconomía y Documentación*

1. INTRODUCCIÓN

1.1. Objetivo, alcance y metodología

En este capítulo se van a presentar y discutir diferentes posibles metodologías docentes para enseñar la construcción, si se quiere diseño,¹ de tesauros. El objetivo es analizar sus posibles ventajas e inconvenientes y facilitar así la decisión de los profesores o, en el caso del autoaprendizaje y el aprendizaje por pares, de los propios estudiantes. El enfoque que se recomienda para la enseñanza de grado es el descendente, aunque puede complementarse con otras estrategias ascendentes para consolidar y precisar los conceptos clave.

Se va a tratar tan sólo la construcción de tesauros, no su utilización. La indicación con tesauros ya existentes es un tema que requiere enfoques y metodologías diferentes. Sin embargo, se asume que para tener garantías de éxito en la construcción de tesauros nuevos es necesario primero saber indizar con diferentes modelos de tesauros con cierta soltura y amplitud de técnicas, aunque sólo sea para haber alcanzado con éxito el arco “conocimientos-habilidades-actitudes” en la comprensión y uso de sus conceptos fundamentales (conceptos, términos, relaciones, disposiciones, postcoordinación, precoordinación, etcétera).

1 Aunque este término creemos que es mejor dejarlo para una de las etapas del proceso.

La metodología adoptada en este capítulo ha sido la de un estudio de caso. Se presenta la experiencia de una década de enseñanza de la asignatura “Construcción y evaluación de tesauros” del grado en Información y Documentación de la Universidad de Zaragoza (2014-2024), se compara con otras experiencias documentadas, se señalan las lecciones aprendidas, se reflexiona sobre la trayectoria seguida en relación con experiencias docentes anteriores, y se exploran algunos retos que se vislumbran en el futuro de la enseñanza de la organización del conocimiento y los lenguajes documentales. Anteriormente se había enseñado también la asignatura en el marco de asignaturas más amplias de lenguajes documentales, principios de la clasificación, y tratamiento y recuperación de la información, por lo que se conocían los problemas a los que se enfrentan los estudiantes al abordar la construcción de tesauros con éxito y, especialmente, para conseguir un sentimiento capacitante de autocompetencia que les permita abordar este tipo de proyectos en centros medianos y pequeños sin supervisión o apoyo en el futuro con seguridad e independencia.

1.2. Por qué se ha enseñado a construir tesauros y por qué puede seguir siendo útil

Los tesauros son un sistema de organización del conocimiento (kos en inglés, de *Knowledge Organization Systems*, se utilizará la abreviatura inglesa) que surgió en un contexto documental, profesional y tecnológico muy diferente a sus antecesores —las clasificaciones bibliográficas y las listas de autoridades y materias—, lo que ha provocado que durante décadas se hayan enseñado frecuentemente en asignaturas diferentes e incluso desconectadas. Esto ha contribuido a transmitir a los estudiantes una visión fraccionada de los kos, con frecuentes conflictos terminológicos y perspectivas muy diferentes sin integrar. Por ello, es importante presentar los tesauros a los estudiantes en relación con otros kos, tanto históricamente como, sobre todo, funcionalmente; y este objetivo debe ser una parte importante de cualquier curso, especialmente dados los requisitos de interoperabilidad que se presentan cada vez más en el entorno de la internet.

Las clasificaciones bibliográficas y lenguajes de autoridades fueron (y son) lenguajes precoordinados desarrollados por bibliotecarios para buscar libros y publicaciones seriadas en secuencias sistemáticas (libre acceso, catálogo sistemático) y listas alfabéticas (catálogo alfabético) en las últimas décadas del siglo XIX y primeras del XX. Por el contrario, los tesauros son lenguajes postcoordinados que fueron desarrollados posteriormente por documentalistas con formación fundamentalmente científica y técnica para abordar el ingente volumen de artículos, informes y documentos no librarios que habían resultado de

la explosión de la información científica en el periodo de las guerras mundiales y la posguerra,² y cuya recuperación temática resultaba difícil de abordar con los métodos anteriores. No se pensaron inicialmente para buscar en listas, sino para aprovechar la invención de máquinas de ordenación y cálculo con funciones booleanas cada vez más avanzadas, que funcionaban extrayendo fichas y luego registros lógicos. En el fondo se trataba de aplicar la potencia de la recuperación booleana a los tradicionales índices analíticos, ya usados para indizar analíticamente las publicaciones seriadas.

Significativamente, los tesauros no surgieron inmediatamente,³ sino como respuesta a las limitaciones de las soluciones más sencillas que los ingenieros de sistemas de información y documentación iban adoptando.⁴ Primero, se vio que los unitérminos sólo permitían búsquedas precisas si eran muy técnicos. Luego, se constató que las palabras clave dejaban a los sinónimos en el limbo, produciendo más o menos silencio según la pericia del recuperador, por lo menos hasta que el texto libre amplió la disponibilidad de términos de búsqueda. Finalmente, se comprobó que las listas de descriptores⁵ sólo permitían buscar alfabéticamente y que, cuando crecían en tamaño, “los árboles” (conceptos) “no dejaban ver el bosque” (dominio de conocimiento), lo que dificultaba la selección de los términos de búsqueda y conducía a resultados con más ruido y menos precisión. Así que los postcoordinacionistas fueron humildemente incorporando las lecciones que ya habían tenido que aprender los precoordinacionistas en el mundo de las bibliotecas.

Su forma actual cuajó a finales de la década de los cincuenta, fueron recibidos con entusiasmo en el ámbito de los centros de documentación y se consolidaron con la primera generación de normas de tesauros en los ochenta.⁶ Enseguida convergieron con el otro gran movimiento de reforma de los kos, las clasificaciones facetadas promovidas por Bliss y Ranganathan, para

2 Slype, Georges van, *Les langages d'indexation: conception, construction et utilisation dans les systèmes documentaires*. París: les éditions d'organisation, 1987.

3 Aitchison, J., and S. D. Clarke, “The thesaurus: a historical viewpoint, with a look to the future,” in Roe, S. K., and A. R. Thomas. 2004. *The Thesaurus: Review, Renaissance, and Revision*, 5-21 (New York: Haworth Press).

4 García Marco, Francisco Javier, “The evolution of thesauri and the history of knowledge organization: between the sword of mapping knowledge and the wall of keeping it simple,” *Brazilian Journal of Information Studies: Research Trends*, 10, n.º 1 (2016), 1-11, <http://www.bjis.unesp.br/revistas/index.php/bjis/article/view/5786>

5 Varias de esas listas ya fueron denominadas tesauros por analogía con el del lexicógrafo británico Peter Roget.

6 La historia se resume en García Marco, Francisco Javier, “Las normas de tesauros se ponen al día: vocabularios estructurados para la recuperación de información en el entorno digital”, *Anuario ThinkEPI 2008: Análisis de tendencias en información y documentación*, 2008, 57-62.

transmutarse en tesauros facetados, ganando en una metodología clara para su construcción.⁷ Por otra parte, los términos de las listas de autoridades y materias terminaron gestionándose como tesauros en los principales sistemas bibliotecarios del mundo. Como resultado del proceso se llegó a un kos que era próximo a las listas de materias por su lenguaje natural y riguroso control del vocabulario, y a las clasificaciones por su estructuración, pero por otro lado, mantenía la suficiente flexibilidad para adaptarse a las nuevas necesidades –lo que no era posible en las clasificaciones bibliográficas tradicionales por su carácter consensual entre organizaciones y países y su complicada gobernanza. En definitiva, triunfaron por su gran capacidad de adaptación e integración de soluciones en el campo de la recuperación documental.

Los tesauros se incorporaron muy pronto a los programas de formación en biblioteconomía y documentación, como herramientas punteras para mejorar la recuperación de la información y la organización del conocimiento (oc), especialmente en la rama de Documentación Científica. Posteriormente fueron evolucionando en España⁸ dentro de las asignaturas de lenguajes documentales de la Diplomatura de Biblioteconomía y Documentación de la posterior Licenciatura en Documentación, y en este siglo del grado en Información y Documentación. En su inicio, por lo general, los lenguajes precoordinados (autoridades, materias y clasificaciones bibliográficas) y postcoordinados (indización) se enseñaban en asignaturas separadas, a veces con alguna más general de introducción. Con el tiempo, especialmente tras la autorización de la licenciatura en Documentación, se distinguió entre la formación sobre su uso y su construcción y, con mayor o menor éxito, hubo un esfuerzo por constituir asignaturas compactas que abordaran los distintos tipos de sistemas. En general, los docentes y la academia han visto el problema de la falta de integración, pero han tenido que sucumbir en mayor o menor medida al principio de la realidad: la coexistencia en la actividad práctica de numerosos sistemas en diferentes áreas de aplicación, cada uno con sus ventajas e inconvenientes. Sin embargo, en este aparente caos, su carácter convergente entre diferentes sistemas convertía a los tesauros en herramientas excelentes para formar a los estudiantes en las técnicas básicas de construcción de lenguajes documentales, que luego podrían aprovechar más fácilmente con los demás kos.

Con la expansión de la World Wide Web, los avances en el análisis semántico latente (modelos probabilísticos y vectoriales, ranking de relevancia...) robaron protagonismo a los tesauros. Sin embargo, al tratarse de herramientas clásicas y ampliamente utilizadas en la organización del conocimiento, han

7 Broughton, Vanda, *Essential thesaurus construction* (Facet, London, 2006), 10.

8 Gil Urdiciáin, Blanca, “Origen y evolución de los tesauros en España”, *Revista General de Información y Documentación* 8, n.º 1, 1998, 63-110.

seguido formando parte de los programas de enseñanza de los estudios de biblioteconomía y documentación –ahora estudios de información– como asignaturas específicas, o al menos como módulos dentro de ellas. De todas formas, su posición quedaba amenazada, y un sector de los profesionales, investigadores y docentes las han llegado a ver como casi obsoletas ante el empuje del análisis semántico latente, cuanto más recientemente con el despliegue de la inteligencia artificial generativa y los grandes modelos lingüísticos. En este estado de opinión puede resultar difícil percibir la dialéctica micro-macro que se da en estos sistemas, y que requiere puertas de incorporación de la organización del conocimiento de dominios específicos para lo que los tesauros pueden constituir un modelo adecuado y práctico, ni demasiado prolijo, ni demasiado simple.⁹

Efectivamente, en la internet se ha visto un proceso de “redescubrimiento de la rueda” semejante al que se dio con los propios tesauros. Al crecer vertiginosamente para incorporar todos los antiguos medios, y cada vez más todos los repositorios existentes, sus ingenieros se han tenido que enfrentar a las limitaciones del análisis basado en el significante –que disminuían la precisión de la recuperación y provocaban ruido–, y desde principios de siglo está inmersa en procesos de “semantización”, basados parcialmente en el apoyo de grafos de conocimiento en cuya ayuda han venido, en los últimos años, los grandes modelos lingüísticos vectoriales como puente entre las dos aproximaciones (análisis latente y formulación explícita). En ese marco, los lenguajes estructurados y su probable sucesor emergente, las ontologías “ligeras”, han recibido una creciente atención. En el tránsito han visto la luz una nueva generación de normas de tesauros que intentan simplificar las normas anteriores –asumiendo el multilingüismo como una necesidad– y abrirlos a la interoperabilidad semántica en la internet.¹⁰

Y es que la revolución de internet ha convertido cualquier campo en un dominio informacional, desde una tienda en línea a un archivo histórico, pasando por un periódico. Por ello, debido al crecimiento cada vez mayor de nuevos campos especializados de información y documentación y a la continua incorporación de nuevos sistemas a la internet apoyados en tecnologías de la web semántica, vuelve a ser importante que los estudiantes de grado sean capaces de desarrollar nuevos KOS pequeños o medianos, adaptados a comunidades de usuarios emergentes o muy específicas, y que contribuyan al mantenimiento de los ya existentes, añadiéndoles mejoras o trabajando en su interoperabili-

9 García Marco, Francisco Javier. “The evolution of thesauri”.

10 Dextre Clarke, S. y M. L. Zeng, “From iso 2788 to iso 25964: the evolution of thesaurus standards towards interoperability and data modeling (2012),” *Information standards quarterly* 24, n.º 1 (2012), 20-26.

dad. Así pues, creemos que la necesidad de instruir a los estudiantes de grado en el diseño de KOS vuelve a convertirse en una prioridad. Además, los tesauros son una excelente opción porque a partir de ellos es fácil derivar los KOS más usados en internet –las taxonomías, folksonomías curadas y ontologías ligeras– y organizar su interoperabilidad.

En conclusión, los tesauros constituyen una gran plataforma para la enseñanza del diseño de KOS porque sintetizan el enfoque alfabético, terminológico y el sistemático, permitiendo de forma flexible las diferentes opciones de organización jerárquica –campos terminológicos, categorías, facetas y disciplinas...– y pueden proporcionar un modelo adecuado para los retos que presenta la internet a la organización del conocimiento.

En este artículo se discuten enfoques alternativos para la formación en la construcción de tesauros, haciendo hincapié en las dificultades experimentadas por los profesores y los estudiantes universitarios de los programas de Biblioteconomía y Documentación y de Estudios de la Información.

2. ENFOQUES EN LA ENSEÑANZA DE LA CONSTRUCCIÓN DE TESAURUS

La enseñanza de la organización del conocimiento en los programas de grado suele centrarse en la comprensión de los conceptos fundamentales de los KOS y en el uso adecuado de estas herramientas a la hora de catalogar e indizar. Se trata de un objetivo clave, ya que se supone que los licenciados en Biblioteconomía y Documentación deben conocer en profundidad cómo se construyen los catálogos, bibliografías y otras herramientas de referencia, de modo que estén capacitados para explotarlos para la búsqueda y referencia en beneficio de sus usuarios, y también para mantenerlos, incorporando nuevos conceptos, términos y relaciones.

En consecuencia, la enseñanza de la construcción de tesauros suele estar muy relacionada con la enseñanza de la indización. Frecuentemente, ambos temas son parte del currículo de la misma asignatura, y la construcción de tesauros se enseña al mismo tiempo que la indización con un vocabulario controlado, normalmente un tesoro concreto, aunque los temas más avanzados –como su construcción y mantenimiento– se suelen explicar al final.

En otras ocasiones, la indización y la construcción de tesauros se enseñan en asignaturas diferentes y secuenciales, pero el profesor opta también por un enfoque ascendente para enseñar el desarrollo de tesauros, muy en conexión con los procesos de indización. Desde este punto de vista, los tesauros se explican y construyen principalmente como herramientas de control del vocabulario, y los alumnos aprenden a interconectar términos y conceptos para construir

la lista controlada que, poco a poco, se va transformando en una red de conceptos. Esta aproximación es perfecta para futuros profesionales que se incorporarán a bibliotecas, archivos y centros de documentación que ya trabajan con sus propios lenguajes documentales ya desarrollados, y para los que, como mucho, propondrán conceptos, términos, clases y relaciones candidatas que luego serán supervisados por los responsables de dichos lenguajes.

Sin embargo, como se ha presentado en la sección anterior, existe una necesidad creciente de desarrollar KOS para equipos de trabajo específicos, pequeñas unidades de información y sitios web que, a pesar de su menor tamaño, necesitan también una herramienta potente que sea capaz de organizar su dominio del conocimiento, proporcionando una visión sistemática del mismo. En este sentido, es necesario que los alumnos se pongan realmente manos a la obra con nuevos tesauros desde cero, para que adquieran confianza a la hora de diseñar estas herramientas en las que suelen trabajar como usuarios –siquiera cualificados, pero no como desarrolladores.

Precisamente, una de las mayores dificultades que encuentran los estudiantes de grado a la hora de construir un tesoro es precisamente aprender a seleccionar y organizar sus conceptos en disciplinas, categorías o facetas, es decir, a construir su arquitectura general. Con frecuencia, en esta parte se bloquean, necesitan mucha ayuda de los profesores y, tras terminar su proyecto, no suelen sentirse lo suficientemente competentes en la organización de los conceptos de un dominio específico. Al contrario, llegan a la conclusión de que diseñar la arquitectura global de un tesoro es una tarea difícil que deben dejar en manos de especialistas más avanzados. Esto puede ser útil para defender la existencia de especialistas en tesauros, pero al final nos condena a ser una minoría en un contexto en el que podríamos estar creciendo con fuerza, al no ser suficientes para aprovechar la oportunidad.

En consecuencia, en este capítulo se presenta y discute un enfoque alternativo al diseño de tesauros que hace hincapié en un enfoque descendente (*top-down*), de modo que los estudiantes puedan terminar el curso con una fuerte sensación de autocompetencia (*self-competence*), pero sin descuidar la necesidad de prestar atención a los detalles, refiriendo cuidadosamente su trabajo a los estándares actuales,¹¹ que, por cierto, son excelentes y sometidos a mantenimiento constante. De hecho, la Sociedad Internacional de Normalización

11 Asociación Española de Normalización y Certificación, *UNE 25964-1:2014: Información y documentación. Tesauros y su interoperabilidad con otros vocabularios. Parte 1: Tesauros para la recuperación de información* (Madrid: AENOR, D.L. 2014-11-26), 47, equivalente a ISO 25964-1; Asociación Española de Normalización y Certificación. *UNE-ISO 25964-2:2016: Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 2: Interoperabilidad con otros vocabularios* (Madrid: Aenor, 2016-12-14), 116, equivalente a ISO 25964-2.

(ISO) está ya trabajando en su revisión sistemática más adaptada todavía a las necesidades de interoperabilidad semántica.¹²

3. RESULTADOS

Los resultados del proyecto de desarrollo del curso pueden agruparse en dos grandes apartados: las lecciones obtenidas sobre la evolución, ventajas e inconvenientes de los enfoques ascendente y descendente en el diseño, construcción y evaluación de tesauros; y las características concretas del curso real que se diseñó.

3.1. Enseñanza de tesauros: evolución, ventajas e inconvenientes de los enfoques ascendente y descendente

En 1972, una década después de que se publicaran los primeros tesauros propiamente dichos, aparecieron tanto el primer manual sobre la utilización, construcción y mantenimiento de tesauros de Aitchison y Gilchrist¹³ como el más general Lancaster,¹⁴ que también los abordaba en detalle. Les siguieron otros excelentes cursos sobre construcción y desarrollo de tesauros tanto monográficos como incluidos en manuales de temática más amplia, a los que seguirían en la década de los ochenta las recomendaciones de los organismos de normalización, que también tienen forma de manual. Fueron monográficos las numerosas y exitosas ediciones y trabajos de colaboración posteriores del antedicho manual de Aitchison y Gilchrist.¹⁵ Entre ellos el de Currás, Aitchison y Gilchrist de 1991¹⁶ en español, que sería seguido de numerosas ediciones ampliadas o adaptadas, traducciones de Currás,¹⁷ y que ha tenido

12 ISO TC46 SC9, resoluciones internas de 2023.

13 Aitchison, J. y A. Gilchrist, *Thesaurus construction: a practical manual* (London: Aslib, 1972).

14 Lancaster, F. W. *Vocabulary control for information retrieval* (Washington: Information Resources Press, 1972).

15 Aitchison, J. y A. Gilchrist, *Thesaurus construction*.

16 Currás, Emilia, Jane Aitchison y Alan Gilchrist, *Thesaurus: lenguajes terminológicos* (Madrid: Paraninfo, 1991).

17 Currás Puente, Emilia, Tesauros: Lenguajes Terminológicos (Brasilia: CNPq-IBICT, 1995), que es traducción de: *Tesauros: lenguajes terminológicos* (Madrid: Paraninfo, 1991); Currás Puente, Emilia, *Tesauros: manual de construcción y uso* (Madrid: Kaher II, D.L. 1998); Currás Puente, Emilia, *Ontologías, taxonomía y tesauros: manual de construcción y uso*, 3.^a ed., act. y amp (Gijón: Trea, 2005).

una gran influencia en Hispanoamérica—; o el de Lancaster¹⁸ de 1985 para la UNESCO. Otros se incluyen dentro de manuales de mayor alcance, como el ya citado de Lancaster de 1973¹⁹ y sus sucesivas ediciones y traducciones, o el muy usado en España de van Slype²⁰ de 1987. Muchos de ellos estaban orientados a profesionales que tienen un perfil muy diferente al de los estudiantes de grado en cuanto a sus motivaciones, su formación teórica y práctica, y la extensión y calidad de sus conocimientos enciclopédicos; pero se han utilizado con éxito como libros de texto en cursos de grado y posgrado.

Ya en este siglo se han ido añadiendo aportaciones muy interesantes. En 2004 Nielsen²¹ ofreció una excelente revisión de la bibliografía que había hasta la fecha sobre la enseñanza de tesauros; y también Thomas²² referencias y consejos muy útiles. Al año siguiente destaca el artículo de Ezzo,²³ que ofrece de forma desenfadada, pero incisiva, sus lecciones aprendidas en la construcción de tesauros, muy valiosas para comprender mejor la perspectiva de nuestros estudiantes. También en 2005 Zeng²⁴ abordó la cuestión de la utilización de aplicaciones de tesauros en la docencia, para la que poco después, con otro enfoque, se dispuso de los manuales elaborados por Gazan²⁵ (2006, revisado en 2013) para el *Catalogers Learning Workshop* de la *Library of Congress*, que destacan por orientarse a la formación para un entorno de catalogación e indexación en red.

En general, estos cursos siguen un enfoque inductivo y ascendente. Por ejemplo, el manual creado por F. W. Lancaster²⁶ para los instructores del Pro-

18 Lancaster, F. W., *Thesaurus Construction and Use: A Condensed Course* (Paris: United Nations Educational, Scientific and Cultural Organization, General Information Programme, 1985). <http://unesdoc.unesco.org/images/0007/000703/070359EB.pdf>

19 Lancaster, F. W., *Thesaurus*.

20 Slype, Georges van, *Les langages d'indexation: conception, construction et utilisation dans les systèmes documentaires* (Paris: les éditions d'organisation, 1987). Traducido al español por la Fundación Germán Sánchez Ruipérez.

21 Nielsen, M. L., "Thesaurus Construction: Key Issues and Selected Readings", *Cataloging y Classification Quarterly* 37, n.º 3-4 (2004), 57-74, doi:10.1300/J104v37n03_05

22 Thomas, A. R., "Teach Yourself Thesaurus: Exercises, Readings, Resources," *Cataloging & Classification Quarterly* 37, n.º 3-4 (2004), 23-34, doi:10.1300/J104v37n03_03.

23 Ezzo, Jaenette, "Bella and Yakov and Tillie's Panties: What I learned in "construction and maintenance of indexing languages and thesauri", *Bulletin of the American Society for Information Science and Technology* 31, n.º 4 (2005), 23-26, <https://www.doi.org/10.1002/bult.1720310408>

24 Zeng, Marcia Lei, "Using software to teach thesaurus development and indexing in graduate programs of LIS and IAKM," *Bulletin of the ASIS&T* 31, n.º 6 (2005), 11-13.

25 Gazan, Rich, *Controlled vocabulary & thesaurus design* (Association for Library Collections & Technical Services, [2013]), <https://www.loc.gov/catworkshop/courses/thesaurus/index.html>

26 Lancaster, F. W., *Thesaurus*.

grama de Información General de la UNESCO en 1985 –y que está basado en un curso impartido en Buenos Aires en 1978– está organizado en 14 capítulos: (1) objetivos del control del vocabulario; (2) principales componentes de un vocabulario controlado; (3) recopilación de los términos; (4) organización de los términos; (5) la relación jerárquica; (6) la relación asociativa; (7) características de los descriptores; (8) el vocabulario de entrada; (9) notas de alcance e identificadores; (10) formato y visualización del tesoro; (11) crecimiento y actualización; (12) uso informatizado; (13) factores del vocabulario que afectan al rendimiento de los sistemas de información; y (14) sistemas de lenguaje natural. Así, tras una presentación global de los tesauros en el marco del control del vocabulario, la organización de los términos ocupa el cuarto lugar después de la recopilación de términos (tercer capítulo).

Este curso sigue, por tanto, un enfoque inductivo y ascendente. Y este es también el caso de las principales normas sobre construcción de tesauros y de algunas publicaciones más recientes sobre la enseñanza de la construcción de tesauros, como las muy prácticas y útiles de Shearer²⁷ (2004), Naumis²⁸ (2005, 2007) –etapas de recolección, conceptual y de evaluación– o la anterior de Cabero y Castro²⁹ (1997).

Este enfoque es coherente con el origen y la evolución del uso de los tesauros para la recuperación de información. Los tesauros para la recuperación de información evolucionaron a partir de la indización postcoordinada, como una forma de controlar conceptos y términos, y de agruparlos para permitir una selección de términos más fácil por parte de los indizadores y para mejorar la expansión y el refinamiento de las búsquedas. Sólo después de algunos años se introdujo la presentación sistemática, primero jerárquica (principalmente en disciplinas o categorías más específicas, dependiendo del ámbito del tesoro), y después facetada.³⁰

Esta estructura tiene también la ventaja de ir desde las unidades más elementales –los conceptos, términos y relaciones– hacia los temas más com-

27 Shearer, J. R., “A Practical Exercise in Building a Thesaurus,” *Cataloging & Classification Quarterly* 37, n.º 3-4 (2004), 35-56, doi:10.1300/J104v37n03_04

28 Naumis Peña, Catalina, “Criterios teóricos y metodológicos para la construcción de un tesoro,” *Boletín de la ANABAD* 55, n.º 1-2 (2005), 97-106; Naumis Peña, Catalina, *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia* (México: UNAM, CUIB, 2007).

29 Moro Cabero, Manuela y Carmen Caro Castro, “Propuesta metodológica para la enseñanza de la utilización y elaboración de tesauros”, *Organización del conocimiento en Sistemas de Información y Documentación*, n.º 2 (1997), 159-167.

30 Aitchison, J. y S. D. Clarke, “The thesaurus”; García Marco, Francisco Javier, “The evolution of thesauri”.

plejos, como la organización del tesoro, su edición para publicación y su mantenimiento.

Por último, la aproximación ascendente es normalmente muy adecuada para la formación de profesionales que se van a integrar en un grupo de trabajo o en el equipo de desarrollo y mantenimiento de un tesoro o una lista de autoridades tesaurizada ya existente. Normalmente, tendrán que ocuparse de conceptos, términos y relaciones concretas, mientras que el diseño global quedará en manos de un grupo reducido, de dirección.

Pero también tiene algunos inconvenientes graves si tenemos en cuenta el contexto pedagógico de los programas de grado. El tema no suele tener muchas horas asignadas y, como el procedimiento inductivo lleva mucho tiempo, los estudiantes de licenciatura acaban con un sentimiento de incompetencia a la hora de enfrentarse a la tarea más abstracta y difícil de organizar un dominio de conocimiento. Como consecuencia, se pierde uno de los principales objetivos del curso: fomentar la autocompetencia.

Debido a una larga experiencia con este relativo fracaso en cursos de lenguajes documentales más generales y teniendo el encargo de organizar una asignatura específica sobre construcción de tesauros, se puso en marcha un proyecto educativo innovador para investigar si un enfoque descendente del aprendizaje y la enseñanza de los tesauros, inspirado en el concepto de Gestalt o cierre cognitivo, podía servir para promover la autocompetencia de los estudiantes de un modo más eficaz. Como resultado, se programó desde cero un horario y un conjunto de actividades para la asignatura “Construcción y evaluación de tesauros” que actualmente se imparte en el Programa de Posgrado en Información y Documentación de la Universidad de Zaragoza (España), que tiene una carga lectiva de 6 ECTS.³¹ Se trata de una asignatura optativa que los alumnos pueden elegir en sus cursos tercero o cuarto,³² por lo que se trata de estudiantes de grado avanzados, con conocimientos previos de indización, catalogación y clasificación.

31 ECTS es el Sistema Europeo de Transferencia y Acumulación de Créditos que se utiliza para facilitar la convalidación de títulos y el intercambio de estudiantes en la Unión Europea. En el sistema universitario español, cada crédito suele equivaler a al menos 25 horas de carga lectiva por crédito, y todas las materias y sus asignaturas deben establecerse en ECTS. En ellas se incluyen las clases teóricas; prácticas en clase, laboratorio o en entornos de aprendizaje externo; trabajos dirigidos; estudio; y evaluación. La proporción es variable según las necesidades de la titulación y la asignatura.

32 En la actualidad, los programas de grado en España (2023) tienen una duración de cuatro años, con excepción de Medicina y Cirugía. Su carga lectiva total es de 240 ECTS. Cada curso anual suele ser de 60 ECTS.

3.2. Diseño del curso

La asignatura se ha diseñado de acuerdo con la filosofía y requisitos del Espacio Europeo de Educación Superior: contabilizar la carga total de trabajo del alumno en créditos de 25 horas, establecer un conjunto de competencias genéricas y específicas como objetivos educativos finales, determinar los correspondientes resultados de aprendizaje como variables operativas y medibles, y fijar una serie de actividades instrumentales para alcanzar los resultados de aprendizaje, así como un sistema de evaluación adecuado a los mismos. El concepto de competencia es central, pues supone intentar que el estudiante sea “competente” en una serie de desempeños, lo cual supone garantizar que va a tener una evaluación positiva por parte de sus responsables en sus futuros contextos laborales, sean profesionales, o en su caso y más específicamente, de investigación o docentes. En la actualidad, el foco de atención se está desplazando de las competencias a los resultados de aprendizaje porque son más controlables en el ámbito estricto de la docencia –es difícil y caro evaluar el desempeño en los trabajos– y están conectados directamente con la evaluación. No obstante, se trata de niveles de abstracción en la programación educativa diferentes, y ambos siguen siendo importantes.

Según la configuración de la asignatura en el plan de estudios de grado, los alumnos deben dedicar un total de 150 horas, aunque diferentes relatos han demostrado que el alumno medio no alcanza este nivel de dedicación, sólo los mejores. La actual guía docente fue elaborada por el autor en marzo 2014 a solicitud del coordinador del grado, el doctor José Antonio Salvador Oliván, cuando la asignatura comenzó a impartirse por primera vez para el curso 2014-2015.

Las competencias propuestas para la asignatura son un subconjunto de las definidas para todo el programa de grado.³³ Cuatro competencias son genéricas: el desarrollo del aprendizaje autónomo, la orientación hacia la mejora continua y la innovación, la mejora de las capacidades personales de organización y planificación, y la promoción de un compromiso ético con los usuarios y su entorno de trabajo. Y dos competencias son específicas: análisis y

33 La mayoría de los programas de grado españoles en Biblioteconomía y Documentación han estado siguiendo un Libro Blanco elaborado por consenso entre los departamentos y centros que lo enseñan en 2003-4 bajo la dirección de la Agencia Nacional de Evaluación de la Calidad y Acreditación en 2005. El Libro Blanco sigue el marco de competencias elaborado en el proyecto DECIDOC, desarrollado en el marco del programa Leonardo da Vinci de la Unión Europea (“Euroguide”). La crisis de ingreso en las titulaciones ha llevado a que varias universidades se hayan apartado de este modelo para ofrecer estudios en Gestión de la información y contenidos digitales (Barcelona, Carlos III, Murcia y próximamente Zaragoza).

representación de la información, y organización y almacenamiento de la información. Como puede verse, la construcción de tesauros fue seleccionada al diseñar el programa como una de las materias en las que mejor se pueden enseñar la creatividad, la innovación y el compromiso ético. Este reto específico se asumió con fuerza a la hora de diseñar el proceso de aprendizaje-enseñanza de la asignatura.

Las competencias se despliegan en doce resultados de aprendizaje. De acuerdo con ellos, el estudiante competente:

- 1) Identifica, analiza y describe los objetivos de un tesoro, sus componentes, su estructura y los mecanismos para su creación, mantenimiento, difusión y utilización
- 2) Organiza el conocimiento de un dominio para facilitar su recuperación
- 3) Detecta y argumenta las implicaciones de la selección de términos para la búsqueda y recuperación de la información
- 4) Incorpora al diseño de tesauros las necesidades de difusión de la información
- 5) Planifica y gestiona la construcción de un tesoro como un proyecto
- 6) Evalúa tesauros
- 7) Comprende y utiliza la norma ISO- UNE 25964
- 8) Construye tesauros especializados utilizando aplicaciones informáticas
- 9) Organiza su agenda de trabajo
- 10) Analiza las implicaciones éticas de sus decisiones
- 11) Planifica y ejecuta su trabajo de forma autónoma
- 12) Mejora su tesoro tomando decisiones innovadoras

3.3. Infraestructura

Se utiliza la norma ISO 25964-1³⁴ como contexto de enfoque y referencia.

Como aplicación de ayuda se seleccionó TemaTres de Diego Ferreyra (TemaTres, 2006-; Ferreyra, 2016), un *software* PHP de código abierto que puede utilizarse en entornos en red y permite muchos formatos de exportación: Skos-Core, Zthes, TopicMap, Dublin Core, MADS, BS8723-5, RSS, SiteMap, txt, SQL. Existen por supuesto un buen número de alternativas como MultiThes, ThManager, VocBench, IQVOC o SKOSed para Protégé.

El *software* se implantó inicialmente en un OS X Server 10.11, con MySQL 5.6.21 y PHP 7.0.6, y en los últimos años se trasladó a un entorno gestionado

34 Asociación Española de Normalización y Certificación, *UNE 25964-1:2014*.

en la nube. Cada alumno dispone de una instalación completa de TemaTres y puede invitar a otros alumnos a colaborar con él.

3.4. Actividades de aprendizaje

Los estudiantes deben trabajar con referencias seleccionadas para construir una cronología categorizada de la evolución de los tesauros. De este modo, adquieren una visión general de su lugar en la ecología de los kos. Las referencias sugeridas cambian parcialmente de un año a otro para evitar copiar y pegar de tareas de alumnos anteriores. Después de completar esta tarea se les proporcionan referencias adicionales para el caso de que deseen ampliar su estado de la cuestión.³⁵

Posteriormente, se les presentan las decisiones previas que deben tomar antes de iniciar un proyecto de tesoro, así como las principales alternativas disponibles. Mientras tanto, deben encontrar varios temas potenciales que puedan convertirse en el objeto de su proyecto, y elegir uno entre ellos. Para ello, utilizan la perspectiva vocacional común de encontrar intersecciones entre sus intereses y capacidades personales, y las necesidades de las personas con las que se relacionan (“clientes”). A continuación, deben hacer una evaluación de los posibles problemas y dificultades, para poder elegir un proyecto que se ajuste a la duración del curso y a sus conocimientos previos. Como la construcción de tesauros puede ser una actividad muy técnica y laboriosa, se intenta enérgicamente que el tema que elijan refuerce su motivación en lugar de convertirse en un obstáculo, problema habitual cuando se les proporciona una lista de temas para elegir, seleccionados por el profesor. Por otra parte, necesitan conocer el tema para no perder demasiado tiempo en ese aspecto y centrarse en los aspectos técnicos de la construcción de un tesoro, que es el objetivo de la asignatura. Ezzo³⁶ señala con agudeza que: “*Creating a thesaurus without knowing the discipline well is like trying to navigate in a foreign country without speaking the language-you can quickly drown in a sea of words*”.

35 García Marco, Francisco Javier, “Normas y estándares para la elaboración de tesauros de patrimonio cultural”, en *El lenguaje sobre el patrimonio: estándares documentales para la descripción y gestión de colecciones*, Secretaría General Técnica, Centro de Publicaciones, Ministerio de Educación, Cultura y Deporte (ed.) (Madrid: Ministerio de Educación, Cultura y Deporte, 2016), 29-46.

36 Ezzo, Jaenette, “Bella and Yakov and Tillie’s Panties: What I learned in “construction and maintenance of indexing languages and thesauri”, *Bulletin of the American Society for Information Science and Technology* 31, n.º 4 (2005), 23-26, <https://www.doi.org/10.1002/bult.1720310408>, 23

Un beneficio colateral de que los estudiantes escojan sus temas de trabajo es la extraordinaria variedad que se produce si el profesor está dispuesto a ser flexible.³⁷ Unos eligen temas más canónicos y disciplinares y otros más relacionados con la vida común –por ejemplo, cocina o distintos *hobbies*–, con negocios concretos o más rupturistas –como videojuegos o productos audiovisuales concretos. En las puestas en común los estudiantes tienen la oportunidad de ver cómo sus compañeros han abordado dominios muy diferentes, y se verán expuestos a esquemas generales de organización o adaptaciones específicas y variadas, contribuyendo a adquirir una visión mucho más amplia del desarrollo de tesauros y de su utilidad.

Se les sugiere que elijan temas lo más concretos posibles para no quedar desbordados y también para que sea más difícil encontrar un modelo que copiar, ya que entonces la actividad dejaría de ser creativa para ser meramente productiva. Además, tienen que identificar claramente quiénes serán los usuarios, qué tipos de documentos se indizarán y en que centros o unidades de información se van a utilizar (biblioteca, sitio web, etcétera), de manera que tengan referencias muy concretas para sus decisiones. Es muy importante que perciban su tesoro como un proyecto nuevo, un reto y un problema que abordar creativamente, pero con anclajes firmes en la realidad.

Por supuesto, tienen que realizar un estado de la cuestión sobre los KOS existentes en su campo, que tienen que considerar, rechazar, aceptar parcialmente o adaptar. Si ya existe un sistema muy desarrollado se acepta que no conviene seguir adelante y que tienen que cambiar de tema, aunque en ese caso se les apoya para que su carga de trabajo no supere los créditos previstos. El tesoro tiene que ser multilingüe, al menos en español e inglés.

Seguidamente, se ofrece a los alumnos una exposición detallada sobre los conceptos y términos en los tesauros, mientras continúan trabajando en sus decisiones preliminares sobre el tesoro, que acabarán convirtiéndose en la introducción de su tesoro. También disponen de un par de sesiones para presentar los proyectos a sus compañeros y debatir entre ellos. Para aumentar la motivación y conseguir cierta aproximación a entornos reales, se pide a los alumnos que se consideren una empresa KOS, le den un nombre comercial y preparen una reunión de negocios en la que presentarán sus proyectos como profesionales. Suelen disfrutar mucho de esta parte.

Mientras se les explican las relaciones conceptuales comienzan su trabajo con las fuentes, que deben elegirse no sólo porque ofrecen conceptos potenciales, sino principalmente porque también proporcionan perspectivas de or-

37 Es importante que el profesor esté motivado y tenga experiencia en diferentes campos, si no le va a resultar difícil interactuar con tantos proyectos diferentes.

ganización alternativas para el dominio. Aunque esto se aleja de lo aconsejado en la norma ISO 25964-1, se ha comprobado que es muy pedagógico, porque los estudiantes pueden trabajar con las capas más generales del dominio con el que están trabajando, poniéndolas en relación con las herramientas de organización del conocimiento más abstractas, por ejemplo, facetas, categorías y árboles disciplinares. Por lo general, basta con una simple hoja de Excel para esbozar rápidamente la estructura general del tesoro e introducir descriptores en otros idiomas, términos no preferidos, términos relacionados y notas.

A continuación, se introduce la presentación y disposición de los tesoros, con algunos ejemplos pertinentes y seleccionados. Entretanto, terminan el esbozo de su jerarquía y comienzan su trabajo con la aplicación del tesoro, TemaTres, introduciendo conceptos, términos, relaciones y notas, y trabajando con las diferentes presentaciones que se proporcionan. Se fomenta y ayuda la importación por lotes utilizando un archivo Excel modificado.

Por último, se ofrece a los alumnos una introducción a la interoperabilidad en cuanto a su importancia, contexto con énfasis en la web semántica, herramientas y problemas, para que aprendan cuál es el siguiente paso que deben dar para mejorar sus capacidades de construcción y mantenimiento de tesoros. En cuanto a la web semántica, se explica su estructura completa en capas, análoga a la “documental” y se analiza con atención SKOS. Practican la exportación de su tesoro en al menos uno de los formatos de la web semántica, SKOS; y el grupo de alumnos analiza sus archivos de exportación y comenta el significado de las instrucciones de la cabecera y de una muestra del código.

El curso finaliza con cinco entregables: una presentación oral con apoyo de diapositivas en un programa de presentación como PowerPoint o similar; un informe tradicional del tesoro con su introducción y las dos presentaciones básicas, sistemática y alfabética; su versión en línea en TemaTres; la exportación SKOS y su comentario; y un informe detallado de la tarea con los tiempos dedicados al proyecto, los problemas encontrados y las soluciones aportadas.

4. CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO

Los resultados han sido siempre muy positivos y, tanto los informes de satisfacción de los alumnos como, sobre todo, los proyectos de los estudiantes, demuestran que se ha alcanzado el objetivo clave de promover la autocompetencia de los estudiantes en el diseño de tesoros. Sin embargo, es necesario seguir trabajando en varias direcciones, aunque las limitaciones temporales del curso pueden hacer necesario abordar futuros trabajos en el marco de otras asignaturas o de la formación permanente.

En la actualidad, la práctica de la interoperabilidad –que sería relativamente fácil utilizando TemaTres– no se ha abordado en profundidad, principalmente por falta de tiempo. Deberían dedicarse al menos dos semanas a explicar los fundamentos de la interoperabilidad según la norma ISO 25964-2 y hacer algunos mapeos entre los tesauros que han diseñado los alumnos. En esta fase, la introducción de Protégée parece una muy buena opción que debería tenerse en cuenta, como Zeng³⁸ ha demostrado eficazmente para los cursos de posgrado. Protégée es potente, está bien probado, es interoperable, profesional y proporciona el contexto mayor del desarrollo de ontologías. Introducirlo en cursos menos avanzados sería muy formativo. Los profesores deben evaluar a sus alumnos y el contexto de enseñanza para ver si este *software* puede adoptarse en cursos de grado sin sacrificar objetivos más básicos.

Además, para permitir un mejor trabajo en equipo y la integración con el resto de las actividades de aprendizaje del programa LIS, sería muy relevante prever salidas de importación y exportación en los formatos más comunes para autoridades bibliográficas. De esta forma, los alumnos podrían conectar mejor sus prácticas de catalogación con sus clases de diseño de tesauros.

Además de estas mejoras específicas, debería explorarse la integración de estrategias pedagógicas alternativas. Los sistemas de enseñanza programada, más directivos, como el propuesto por Irving³⁹ o más sencillamente test de auto-evaluación programados secuencialmente, pueden ser complementarios a un enfoque educativo basado en proyectos. Pueden ser muy útiles para garantizar que los alumnos no se pierdan ningún punto importante y que hayan completado su aprendizaje conceptual y normativo antes de abordar su práctica. Sin embargo, también pueden plantear problemas. En nuestro caso, aunque se ha diseñado ya un conjunto de tests, hasta el momento no se han utilizado porque, vista la disponibilidad de tiempo de los estudiantes, no se ha visto adecuado sobrecargarles más de trabajo y, sobre todo, se ha intentado evitar que este enfoque más directivo pueda contaminar el clima de clase, coartando su creatividad, ya que es frecuente que la exposición a una señal discriminante aprendida anteriormente –como la presentación de tests– desencadene los comportamientos normalmente asociados a ella –aprendizaje orientado a su resolución. No obstante, esta cuestión tendrá que someterse a investigación educativa experimental con grupo de control.

38 Zeng, Marcia Lei, “Using software to teach thesaurus development and indexing in graduate programs of LIS and IAKM”, *Bulletin of the ASIS&T* 31, n.º 6 (2005), 11-13.

39 Irving, H, “CAIT: Computer-assisted indexing tutor, implemented for training at NAL”, *Agric. Libr. & Inform. Notes* 21, n.º 4-6 (1995), 1-5.

Por último, es bien sabido que los alumnos despliegan diferentes estilos de aprendizaje y de pensamiento a la hora de abordar las materias que deben dominar.⁴⁰ Estos estilos de aprendizaje parecen estar muy relacionados con los rasgos de personalidad de los alumnos. En concreto, algunas personas prefieren un enfoque analítico, paso a paso, mientras que otras necesitan obtener una Gestalt del campo (pregnancia) para ocuparse después de los detalles. Esto podría estar relacionado con la preferencia por un enfoque ascendente o descendente en el aprendizaje del desarrollo del tesoro. Por lo tanto, hay que seguir investigando en este sentido para averiguar si un rasgo tan importante de la personalidad tiene un impacto real en la enseñanza y el aprendizaje de la construcción de tesoros. Por supuesto, estos estudios deberían realizarse también de forma controlada, obteniendo medidas objetivas y subjetivas sobre los estilos de aprendizaje de los estudiantes.

40 Sternberg, Robert J. *Thinking styles*. New York: Cambridge University Press, 1997.

REFERENCIAS BIBLIOGRÁFICAS

- Asociación Española de Normalización y Certificación, *UNE 25964-1:2014: Información y documentación. Tesauros y su interoperabilidad con otros vocabularios. Parte 1: Tesauros para la recuperación de información*, Madrid: Aenor, D.L. 2014-11-26, 47, equivalente a ISO 25964-1.
- Asociación Española de Normalización y Certificación, *UNE-ISO 25964-2:2016: Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 2: Interoperabilidad con otros vocabularios*, Madrid: Aenor, 2016-12-14. 116, equivalente a ISO 25964-2.
- Agencia Nacional de Evaluación de la Calidad y Acreditación, *Título de Grado en Información y Documentación: Libro Blanco*, Madrid: Agencia Nacional de Evaluación de la Calidad y Acreditación, 2004, http://www.aneca.es/media/150424/libroblanco_jun05_documentacion.pdf
- Aitchison, J. y S. D. Clarke, "The thesaurus: a historical viewpoint, with a look to the future", en Roe, S. K. y A. R. Thomas, *The Thesaurus: Review, Renaissance, and Revision*, 5-21, New York: Haworth Press, 2004.
- Aitchison, J. y A. Gilchrist, *Thesaurus construction: a practical manual*, London: Aslib, 1972.
- Broughton, Vanda, *Essential thesaurus construction*, London: Facet, 2006.
- Currás Puente, Emilia, Jane Aitchison y Alan Gilchrist, *Thesaurus: lenguajes terminológicos*, Madrid: Paraninfo, 1991.
- Currás Puente, Emilia, *Tesauros: Linguagens Terminológicas*, Brasilia: CNPq-IBICT, 1995.
- Currás Puente, Emilia, *Tesauros: manual de construcción y uso*, Madrid: Kaher II, D.L., 1998.
- Currás Puente, Emilia, *Ontologías, taxonomía y tesauros: manual de construcción y uso*, 3.^a ed., act. y amp. Gijón: Trea, 2005.
- Dextre Clarke, S. y M. L. Zeng, "From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling (2012)", *Information Standards Quarterly* 24, n.º 1 (2012), 20-26.

- España, Jefatura del Estado, “Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario”, *Boletín Oficial del Estado*, n.º 70 (23/03/2023), <https://www.boe.es/buscar/pdf/2023/BOE-A-2023-7500-consolidado.pdf>
- Euroguide LIS: the guide to competencies for European professionals in library and information services*, London: Aslib, the Association for Information Management, 2000, <http://www.aslib.co.uk/pubs/2001/18/01/foreword.htm>
- Ezzo, Jaenette, “Bella and Yakov and Tillie’s Panties: What I learned in «construction and maintenance of indexing languages and thesauri»”, *Bulletin of the American Society for Information Science and Technology* 31, n.º 4 (2005), 23-26, <https://www.doi.org/10.1002/bult.1720310408>
- Ferreira, Diego, “TemaTres”, <http://www.vocabularyserver.com/blog/contact/>
- García Marco, Francisco Javier, 25731-*Construcción y evaluación de tesauros: Guía docente para el curso 2015-2016*, Zaragoza: Universidad, 2015, <http://titulaciones.unizar.es/ asignaturas/25731>
- García Marco, Francisco Javier, “The evolution of thesauri and the history of knowledge organization: between the sword of mapping knowledge and the wall of keeping it simple”, *Brazilian Journal of Information Studies: Research Trends* 10, n.º 1 (2016), 1-11, <http://www.bjis.unesp.br/revistas/index.php/bjis/article/view/5786>
- García Marco, Francisco Javier, “Normas y estándares para la elaboración de tesauros de patrimonio cultural”, en *El lenguaje sobre el patrimonio: estándares documentales para la descripción y gestión de colecciones*, Secretaría General Técnica, Centro de Publicaciones, Ministerio de Educación, Cultura y Deporte (ed.), 29-46, Madrid: Ministerio de Educación, Cultura y Deporte, 2016, <https://sede.educacion.gob.es/publiventa/el-lenguaje-sobre-el-patrimonio-estandares-documentales-para-la-descripcion-y-gestion-de-colecciones/cultura-sociedad-lenguaje-patrimonio-historico-artistico/20819C>
- García Marco, Francisco Javier, “Teaching Thesaurus Construction: A Top-Down Approach for LIS Undergraduate Programmes”, In *Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific,*

- and Technological Sharing in a Connected Society: Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil*, Guimarães, José Augusto Chaves, Milani, Suellen Oliveira y Dodebei, Vera (eds.), 546-554, Würzburg: Ergon-Verlag GmbH, 2016.
- García Marco, Francisco Javier, “Las normas de tesauros se ponen al día: vocabularios estructurados para la recuperación de información en el entorno digital”, *Anuario ThinkEPI 2008: Análisis de tendencias en información y documentación*, 2008, 57-62.
- Gazan, Rich, *Controlled vocabulary & thesaurus design*, Association for Library Collections & Technical Services, 2013, <https://www.loc.gov/catworkshop/courses/thesaurus/index.html>
- Gil Urdiciáin, Blanca, “Origen y evolución de los tesauros en España”, *Revista General de Información y Documentación* 8, n.º 1, 1998, 63-110.
- Irving, H., “CAIT: Computer-assisted indexing tutor, implemented for training at NAL”, *Agric. Libr. & Inform. Notes* 21, n.º 4-6, 1995, 1-5.
- Lancaster, F. W., *Vocabulary control for information retrieval*, Washington: Information Resources Press, 1972.
- Lancaster, F. W., *Thesaurus Construction and Use: A Condensed Course*, Paris: United Nations Educational, Scientific and Cultural Organization, General Information Programme, 1985, 96, <http://unesdoc.unesco.org/images/0007/000703/070359EB.pdf>
- Moro Cabero, Manuela y Carmen Caro Castro, “Propuesta metodológica para la enseñanza de la utilización y elaboración de tesauros”, *Organización del conocimiento en Sistemas de Información y Documentación*, n.º 2, 1997, 159-167.
- Naumis Peña, Catalina, “Criterios teóricos y metodológicos para la construcción de un tesoro”, *Boletín de la ANABAD* 55, n.º 1-2, 2005, 97-106.
- Naumis Peña, Catalina, *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*, México: UNAM, CUIB, 2007.
- Nielsen, M. L., “Thesaurus Construction: Key Issues and Selected Readings”, *Cataloging & Classification Quarterly* 37, n.º 3-4, 2004, 57-74, doi:10.1300/J104v37n03_05

- Shearer, J. R., "A Practical Exercise in Building a Thesaurus", *Cataloging & Classification Quarterly* 37, n.º 3-4, 2004, 35-56, doi:10.1300/J104v37n03_04
- Slype, Georges van, *Les langages d'indexation: conception, construction et utilisation dans les systèmes documentaires*, París: les éditions d'organisation, 1987.
- Sternberg, Robert J., *Thinking styles*, New York: Cambridge University Press, 1997.
- TemaTres, "TemaTres: controlled vocabulary server", *Sourceforge*, 2006, <https://sourceforge.net/projects/tematres/>.
- Thomas, A. R., "Teach Yourself Thesaurus: Exercises, Readings, Resources", *Cataloging & Classification Quarterly* 37, n.º 3-4, 2004, 23-34, doi:10.1300/J104v37n03_03
- Zeng, Marcia Lei, "Using software to teach thesaurus development and indexing in graduate programs of LIS and IAKM", *Bulletin of the ASIS&T* 31, n.º 6, 2005, 11-13.

II. APLICACIONES PRÁCTICAS

La representación temática facetada: análisis y propuesta de conversión y representación de encabezamientos de materia a FAST

Faceted thematic representation: analysis and proposal for conversion and representation of subject headings to FAST

Noé Ángeles Escobar
Biblioteca Daniel Cosío Villegas, El Colegio de México

INTRODUCCIÓN

FAST (*Faceted Application of Subject Terminology*) es una propuesta de adecuación del lenguaje de encabezamientos de materia (EM) que busca una manera de describir principalmente recursos web de manera sencilla, a un bajo costo y con poco esfuerzo, que dicha descripción sea compatible con diferentes esquemas de metadatos, eficaz para la recuperación de información temática y con interoperabilidad semántica. Es decir, se busca establecer un sistema más sencillo de aprender y de aplicar, y con una forma más amigable de navegación en la consulta temática. Los EM se descomponen y reformulan en ocho facetas y adquieren una dinámica diferente. En este trabajo se presentan las características de la terminología FAST, su comparación con respecto a los EM, su definición estructural como sistema poscoordinado de representación temática, así como la codificación que adquiere en el formato MARC21. Se discuten las ventajas de FAST y se presenta el estado en el que se encuentra el “Proyecto sobre la implementación de la terminología FAST” de la Biblioteca Daniel Cosío Villegas (BDCV), de El Colegio de México.

ANTECEDENTES DEL SISTEMA FAST

Desarrollada desde 1998 como parte de WorldCat por OCLC, FAST (*Faceted Application of Subject Terminology*) es una propuesta de adecuación del lenguaje de encabezamientos de materia (EM) de la *Library of Congress* (LCSH)¹ que busca una manera de describir principalmente recursos web, de manera sencilla, a un bajo costo y con poco esfuerzo, que sea compatible con diferentes esquemas de metadatos, eficaz para la recuperación de información temática y con interoperabilidad semántica.² Es decir, se busca establecer un sistema que sea sencillo de aprender y de aplicar, y con una forma amigable de navegación en la consulta temática. En esencia, FAST rompe las cadenas de sintaxis pre-coordinada con las que se construyen los EM y propone un sistema facetado, poscoordinado, con una sintaxis muy simple y que elimina la redundancia en la asignación de EM; pero, ante todo, conserva el control terminológico establecido en los LCSH, lo que permite aprovechar todo el trabajo previo desarrollado en este sistema de representación temática.³

Los temas se agrupan en las diferentes facetas que describen de qué trata un recurso de información. En algunas fuentes se mencionan seis facetas: Nombres –incluye personales, corporativos y títulos–, Nombres geográficos, Eventos, Tiempo/Periodos, Tópicos y Forma bibliográfica/Género,⁴ mientras que en otras se señalan ocho,⁵ pero eso depende de si se agrupan o no los nombres de personas, nombres corporativos y títulos. Actualmente existen más de 1.8 millones de términos generados con la lógica FAST desde los EM presentes en WorldCat,⁶ los cuales están disponibles como datos vinculados (<http://>

1 Las propuestas que provienen de una derivación de los LCSH son relevantes en los ámbitos latinoamericanos porque la mayoría de los listados o sistemas de EM en la región tiene como base tanto la estructura como, en muchos casos, la traducción al español de los LCSH.

2 Corey A. Harper, y Barbara B. Tillett, “Library of Congress controlled vocabularies and their application to the Semantic Web”, *Cataloging & Classification Quarterly* 43, n.º 3-4 (2007), 47-68.

3 Marilyn Montalvo Montalvo, “LCSH, FAST y DELICIOUS: vocabularios normalizados y nuevas formas de catalogación temática”, *Anales de Documentación* 14, n.º 1 (2011), 1-13.

4 Janet Ashton y Caroline Kent. 2017, “New Approaches to Subject Indexing at the British Library,” *Cataloging & Classification Quarterly* 55, n.º 7-8 (2017), 549-559, <https://doi.org/10.1080/01639374.2017.1354345>

5 Edward T. O'Neill, y Lois Mai Chan, *FAST (Faceted Application of Subject Terminology), a simplified LCSH-based vocabulary*. Presented at World Library and Information Congress: 69th IFLA General Conference and Council, 1-9 August 2003, Berlin, https://archive.ifla.org/IV/ifla69/papers/010e-ONEILL_Mai-Chan.pdf

6 Chew Chiat Naun, Kerre Kammerer, Kim Mumbower, y Dean Seeman, *FAST Quick Start Guide* (Ohio: OCLC FAST Policy and Outreach Committee, 2022), 17, <https://www.oclc.org/content/dam/oclc/FAST/FAST-quick-start-guide-2022.pdf>

id.worldcat.org/fast/.), además de que cuenta con sendas aplicaciones para la búsqueda de términos FAST (<http://fast.oclc.org/searchfast/>) y para convertir de LCSH a FAST (<https://fast.oclc.org/lcsh2fast/>).

En cuanto a su recepción por los catalogadores, en 2013 OCLC realizó una encuesta entre 16 instituciones acerca de la decisión de usar FAST;⁷ las respuestas indicaron que en su adopción identifican como virtudes de FAST: la facilidad de uso, su sintaxis simple, su conveniencia para usarse por personal no altamente especializado, estructura uno a uno entre encabezamientos y registros de autoridad y su riqueza de vocabulario. Con menos menciones, pero también relevantes, señalan: la facilidad de aprender FAST, su disponibilidad como datos ligados y la posibilidad de lograr una navegación facetada. Por otro lado, aunque la documentación de aplicaciones es escasa, resalta la experiencia de la British Library, que desde 2014 desarrolló un proyecto piloto en el cual los catalogadores utilizaron FAST en lugar de LCSH en materiales seleccionados y cuyos resultados apuntan a muchas ventajas potenciales de la adopción de FAST.⁸ Al tiempo que mencionan que el personal bibliotecario señala preocupaciones profesionales y técnicas (relativos a la precisión en la asignación temática, el mantenimiento de los datos sólo en OCLC y su control externo), apuntan como beneficios de dicha adopción los siguientes: rapidez en la asignación de términos temáticos; que a partir de la aplicación de OCLC para convertir cadenas de OCLC a FAST es posible enriquecer registros existentes, de forma retrospectiva; el potencial de FAST en el descubrimiento de recursos y su aplicación como datos vinculados; en cuanto a la cuestión económica, el impacto observado tiene que ver con la menor inversión en términos de formación de profesionales; los resultados también sugieren que el cambio hacia FAST sería relativamente simple, comparado con la adopción de un sistema diferente que implique costos y tiempos, pues FAST requeriría un entrenamiento mínimo para personal experimentado y el tiempo para personal nuevo no sería muy extenso; otro beneficio observado es la compatibilidad con MARC, pues sigue los mismos principios. Asimismo, Smith-Yoshimura en su informe sobre una nueva generación de metadatos apunta que los términos FAST brindan una fácil transición a un entorno de datos vinculados, ya que cada término tiene un identificador único y señala que debido a que los términos se generan a partir de los LCSH, también pueden contener los sesgos de origen de dicha fuente, por lo que retoma la política de OCLC que indica que

7 Jeffrey Mixter y Eric R. Childress. *FAST (Faceted Application of Subject Terminology) Users: Summary and Case Studies* (Dublin, Ohio: OCLC Research, 2013), <http://www.oclc.org/content/dam/research/publications/library/2013/2013-04.pdf>

8 Janet Ashton y Caroline Kent. 2017, *New Approaches*, 557-558.

FAST será un vocabulario temático general con herramientas y servicios que satisfagan las necesidades de diversas comunidades y contextos.⁹

Visto lo anterior, los retos que implica FAST a la organización de la información temática en las bibliotecas tienen que ver con la ponderación de las ventajas mencionadas que aportaría su implementación frente a las implicaciones de la misma. Por ejemplo, resulta claro que si FAST conlleva un menor trabajo en la asignación de temas, los cuales se ofrecen al usuario de forma más amigable y simple, la instrumentación demanda también un reforzamiento mayor del control terminológico, sobre todo en el caso de las bibliotecas que trabajan con temas en español, pues si a partir de WorldCat ya hay más de 1.8 millones de términos establecidos, no existe aún tal cantidad en español. Junto con esto es necesario conocer y dominar exactamente cómo se dividen, de manera sistemática, y cuáles son las reglas para la conformación de los términos FAST desde las cadenas de EM. Varias fuentes explican el funcionamiento estructural de FAST¹⁰ y no hay mucho que discutir acerca de la catalogación de recursos web, pues FAST cumple requerimientos propios de este tipo de recursos: facilidad de asignación temática, en poco tiempo, sencillez en la representación, accesibilidad y compatibilidad con diferentes esquemas de metadatos. La gran cuestión tiene que ver con su implementación a nivel estructural desde el insumo que son los EM del catálogo de la biblioteca. La discusión debe darse a partir del desarrollo de pruebas que señalen la pertinencia de un cambio definitivo a FAST, dejando de lado los EM, o en hacer que coexistan ambos sistemas. El planteamiento tiene que hacerse considerando: una discusión colegiada institucional, la definición de una ruta de análisis, la toma de decisiones y las acciones relacionadas con la implementación de FAST. De acuerdo con lo anterior, en la Biblioteca Daniel Cosío Villegas (BDCV) de El Colegio de México, se desarrolló el “Proyecto sobre la implementación de la terminología FAST”.

9 Karen Smith-Yoshimura, *Transitioning to the Next Generation of Metadata* (Dublin, Ohio: oclc Research, 2020), 11, <https://doi.org/10.25333/rqgd-b343>

10 Véase: Lois Mai Chan y Edward T O’Neill. *FAST : Faceted Application of Subject Terminology : Principles and Applications* (Santa Barbara, Calif.: Libraries Unlimited, 2010); Haiyun Cao y Marcia Salmon. *FAST: Faceted Application of Subject Terminology*. Presented at OLA Super Conference, Thursday February 25, 2010, https://www.academia.edu/85012503/FAST_Faceted_Application_of_Subject_Terminology?uc-sb-sw=79242156; Qiang Jin, “Is FAST the right direction for a new system of subject cataloging and metadata?”, *Cataloging & Classification Quarterly* 45, n.º 3 (2008), 91-110, doi: 10.1300/J104v45n03_08

EL PROYECTO FAST DE LA BDCV

A partir del año 2020 se comenzó a analizar la pertinencia de emplear la terminología FAST en la descripción de recursos de información en la BDCV, habida cuenta del desarrollo de diferentes colecciones digitales, para las que los EM, aunque fueron utilizados, no resultaban pertinentes, principalmente debido a los esquemas de metadatos no MARC empleados para la descripción de los recursos de información contenidos en dichas colecciones y a la complejidad estructural inherente a los propios EM. Entonces, empezó a tomar forma el proyecto sobre FAST, el cual definió como objetivo general analizar y definir cursos de acción sobre la forma en que se puede implementar la iniciativa FAST en la BDCV, el cual toma forma en los siguientes objetivos específicos: 1. Analizar la propuesta FAST para la representación temática y sus implicaciones en la elaboración de registros bibliográficos y del trabajo de asignación temática en la BDCV; 2. Familiarizarse con la conversión de la representación temática EM hacia FAST; 3. Desarrollar un mecanismo que logre convertir EM al vocabulario FAST de manera automática; 4. Aplicar la conversión automática para la generación de términos FAST en registros existentes en la BDCV; junto con esto, determinar las conversiones globales que pueden hacerse y las que requerirán un tratamiento manual.

Para el cumplimiento de los objetivos se desarrollaron las siguientes acciones:

ETAPA 1. Comprender la dinámica de FAST.

Acciones: *a)* Revisión de literatura; *b)* Analizar la forma en que se construye la terminología FAST a partir de los EM; *c)* Asimilar la asignación de términos FAST en los registros bibliográficos; *d)* Generar esquemas de conversión desde EM, para las diferentes facetas de FAST.

ETAPA 2. Desarrollar reglas de normalización en el sistema integrado de gestión bibliotecaria (SIGB) de la BDCV.¹¹

Acciones: *a)* Determinar casos de conversión según las ocurrencias representadas en EM existentes en la BDCV; *b)* Generar instrucciones para conformar las reglas de normalización a partir de cada caso determinado; *c)* Probar el funcionamiento de las reglas de normalización en un conjunto de registros; *d)* Aprobar las reglas de normalización.

11 Las reglas de normalización, definidas de ese modo en el SIGB de la BDCV, son el mecanismo con algoritmos específicos mediante el cual se logra automatizar cambios globales en conjuntos de registros bibliográficos, de modo que permiten automatizar la conversión EM-FAST.

ETAPA 3. Aplicación de las reglas de normalización de manera global.

Acciones: *a)* Determinar el grado de conversión por cambios globales desde la aplicación de las reglas de normalización; *b)* Definir la forma de atender la conversión manual de EM a FAST cuando no es posible hacerlo por medio de las reglas de normalización y cuando hay cambios en la terminología o en la estructura.

Como parte de la Etapa 1 se documentó la forma en que se logra pasar de una representación temática tipo EM a una tipo FAST. Para el análisis de conversión de EM a FAST se retoman las ocho facetas establecidas en <http://FAST.oclc.org/searchFAST/> la cual es la fuente de autoridades para términos FAST, así como ejemplos provenientes de <https://www.worldcat.org/> que constituyen la aplicación en registros reales de la terminología FAST. La presentación de ejemplos y casos corresponde a los hallados en las fuentes referidas (en inglés). En todo momento la idea fue asimilar la dinámica que impone esta iniciativa de representación temática; junto con esto deben considerarse posteriormente las implicaciones para la conversión de términos de EM a FAST en español. A continuación, se presentan algunos casos prototípicos de conversión.

EL PASO DE EM A FAST

Como se podrá observar, no se trata únicamente de romper las cadenas pre-coordinadas de los EM, sino de retomar su vocabulario y disponerlo de manera significativa en las facetas establecidas. Esto puede derivar en:

- a. Términos que no cambian (pasan de manera transparente de EM a FAST).
- b. Ajustes terminológicos al pasar a FAST.
- c. Ajustes estructurales en la representación temática.
- d. Cambios en la codificación habitual en el formato MARC.

En las siguientes secciones se ejemplifican conversiones para la tipología mencionada y se señalan los ajustes requeridos. Los casos se retoman de OCLC, por lo que se presentan en inglés y debe presuponerse un subcampo \$a al inicio de los campos 6XX. Cada fila de la tabla debe asumirse como el o los EM asignados a un mismo registro. Nótese el uso del segundo indicador 7 y de los subcampos \$2 y \$0 para establecer que se trata de un término FAST y de su identificador en OCLC, respectivamente

1. Faceta de nombre de persona

No hay cambios en los términos ni en la estructura cuando se pasa de EM a FAST para los datos propiamente dichos de los nombres de las personas. Lo que sí cambia son las subdivisiones asociadas (subcampos \$v, \$x, \$y, \$z) con el nombre de la persona en un EM, las cuales se convierten a su faceta respectiva, por lo que es necesario observar los cambios en estas subdivisiones y comprender su lógica al pasar a FAST. Esta faceta incluye personajes ficticios y mitológicos. Las notas de ajustes presentadas para esta faceta deben revisarse también en el resto de las facetas. Para el caso de los EM de nombre/título (subcampo 600 \$t) véase más adelante la faceta Título.

EM	FAST	Ajustes
600 10 Álvarez Bravo, Lola.	600 17 Álvarez Bravo, Lola. \$2 FAST \$0 (OcoLC) fst00169007	Sin cambios en el paso del nombre personal a FAST
600 00 Pius \$b XII, \$c Pope, \$d 1876-1958.	600 07 Pius \$b XII, \$c Pope, \$d 1876-1958. \$2 FAST \$0 (OcoLC) fst00034590	Sin cambios en el paso del nombre personal a FAST
600 10 Lawrence, D. H. \$q (David Herbert), \$d 1885-1930 \$x Travel \$z New Mexico \$z Taos.	600 17 Lawrence, D. H. \$q (David Herbert), \$d 1885-1930. \$2 FAST \$0 (OcoLC) fst00031306 650 7 Travel. \$2 FAST \$0 (OcoLC) fst01155558 651 7 New Mexico \$z Taos. \$2 FAST \$0 (OcoLC) fst01208684	Sin cambios en el paso del nombre personal a FAST Se desagrega \$x en un campo 650 Se desagregan \$z en un campo 651. Nótese cómo se conserva una estructura jerárquica y no la forma habitual de 651 \$a que sería: Taos (N.M.)

EM	FAST	Ajustes
<p>600 10 Shakespeare, William, \$d 1564-1616 \$x Criticism and interpretation.</p>	<p>600 17 Shakespeare, William, \$d 1564-1616 \$2 FAST \$0 (OcoLC) fst00029048</p> <p>655 7 Criticism, interpretation, etc. \$2 FAST \$0 (OcoLC) fst01411635</p>	<p>Sin cambios en el paso del nombre personal a FAST</p> <p>En este caso, de EM a FAST se observa un cambio en el término, así como se verifica un cambio de \$x tópico a 655\$a término de forma.</p>
<p>600 10 Paz, Octavio, \$d 1914-1998.</p> <p>600 10 Paz, Octavio, \$d 1914-1998 \$x Criticism and interpretation.</p> <p>600 10 Paz, Octavio, \$d 1914-1998 \$x Homes and haunts \$z Mexico \$z Mixcoac.</p> <p>600 10 Paz, Octavio, \$d 1914-1998 \$x Friends and associates.</p> <p>651 0 Mixcoac (Mexico City, Mexico) \$x In literature.</p>	<p>600 17 Paz, Octavio, \$d 1914-1998 \$2 FAST \$0 (OcoLC) fst00038877</p> <p>650 7 Friendship. \$2 FAST \$0 (OcoLC) fst00935174</p> <p>650 7 Homes. \$2 FAST \$0 (OcoLC) fst01353235</p> <p><i>**cambio en el término</i></p> <p>650 7 Literature. \$2 FAST \$0 (OcoLC) fst00999953</p> <p><i>**cambio en el término</i></p> <p>651 7 Mexico \$z Mexico City \$z Mixcoac. \$2 FAST \$0 (OcoLC) fst01712073</p> <p><i>**cambio en la estructura</i></p>	<p>Sin cambios en el paso del nombre personal a FAST. Nótese que Paz aparece cuatro veces en los EM, mientras que en FAST sólo una vez</p> <p>Cambio en el término: “Friends and associates” pasa a “Friendship”</p> <p>Cambio en el término: “Homes and haunts” pasa a “Homes”</p> <p>Cambio en el término: “In literature” pasa a “Literature”</p>

EM	FAST	Ajustes
	<p>655 7 Criticism, interpretation, etc. \$2 FAST \$0 (OcoLC) fst01411635 <i>**cambio en el término y pasa a FORMA</i></p>	<p>Se desagregan \$z en un campo 651. Nótese cómo se conserva una estructura jerárquica y no la forma habitual de 651 \$a que sería: Taos (N.M.). Además, se dejan todas las subdivisiones presentes y no sólo dos, como en la dinámica de los EM. En este caso, de EM a FAST se observa un cambio en el término, así como se verifica un cambio de \$x, tópico, a 655\$a, término de forma.</p>

Tabla 1. Faceta de nombre de persona y cambios generales (casos tomados del sistema OCLC Connexion).

2. Faceta de nombre corporativo

Como en el caso de los nombres de persona no hay cambios en los términos ni en la estructura cuando se pasa de EM a FAST para los datos propiamente dichos de los nombres corporativos. Asimismo, lo que sí cambia son las subdivisiones asociadas (subcampos \$v, \$x, \$y, \$z) y para encabezamientos nombre/título (subcampo 610 \$t).

EM	FAST
<p>610 20 World Trade Organization.</p>	<p>610 27 World Trade Organization. \$2 FAST \$0 (OCoLC)fst00697669 <i>**sin cambios en el término, ni en la codificación</i></p>

EM	FAST
610 10 Mexico. \$b Secretaría de Relaciones Exteriores \$v Bibliography.	610 17 Mexico. \$b Secretaría de Relaciones Exteriores. \$2 FAST \$0 (OCoLC)fst00556526 <i>**sin cambios en el término, ni en la codificación</i> 655 7 Bibliographies. \$2 FAST \$0 (OCoLC)fst01423717 <i>**se desagrega y hay cambio en el término, pasa a plural</i>
610 20 Catholic Church \$x Doctrines \$x History \$y 20th century.	610 27 Catholic Church. \$2 FAST \$0 (OCoLC) fst01919895 650 7 Theology, Doctrinal. \$2 FAST \$0 (OCoLC)fst01149617 <i>**se desagrega y hay cambio en el término</i> 655 7 History. \$2 FAST \$0 (OCoLC) fst01411628 <i>**cambio en la codificación, pasa a forma/género</i> 648 7 1900-1999 \$2 FAST <i>**cambio en el término y en la codificación a nivel de campo</i>

Tabla 2. Faceta de nombre corporativo (casos tomados del sistema OCLC Connexion).

3. Faceta de título

Esta faceta abarca tanto los títulos sin autor reconocido (campo 630) como los títulos con autoría identificada (subcampos 6XX \$t). Nótese los cambios en la estructura del término que ocurren con FAST y en la codificación.

Ejemplos

EM	FAST
630 00 Amadís de Gaula (Spanish romance).	630 07 Amadís de Gaula (Spanish romance) \$2 FAST \$0 (OCoLC) fst01356403
600 10 Paz, Octavio, \$d 1914-1998. \$t Laberinto de la soledad.	630 07 Laberinto de la soledad (Paz, Octavio) \$2 FAST \$0 (OCoLC) fst01363143

EM	FAST
610 10 Mexico. \$t Constitución política (1917).	630 07 Constitución política (Mexico: 1917) \$2 FAST \$0 (OcoLC) fst01360749
630 00 Montevideo Treaty \$d (1980).	630 07 Montevideo Treaty (1980) \$2 FAST \$0 (OcoLC)fst01773892

Tabla 3. Faceta de título (casos tomados del sistema OCLC Connexion).

4. Faceta de nombre de reunión o evento

De acuerdo con Ashton y Kent (2017) “*Events: are distinct from topics, differing once more in this respect from lcsb. Events in FAST terms include meetings and conferences as well as specific historical occurrences*”.¹² Sin embargo, en <http://FAST.oclc.org/searchFAST/> aparecen agrupados los términos relativos a reuniones (es decir, conferencias, congresos, etcétera) y acontecimientos históricos (por ejemplo, revoluciones, huracanes, huelgas, accidentes, etcétera), aunque la codificación es diferente (en registros de autoridad MARC21, 111 para los primeros y 147 para los segundos). Es decir, que en sentido estricto, de acuerdo con FAST, todos son eventos o acontecimientos, no obstante, en este análisis se mantendrá la forma en que son agrupados por OCLC, así como su codificación. Nótese los cambios en la codificación para 611 y los cambios estructurales para eventos en 647.

EM	FAST
611 20 Paris Peace Conference \$d (1919-1920).	611 27 Paris Peace Conference (1919-1920) \$2 FAST \$0 (OCoLC) fst01696461
611 20 Encuentro Nacional de Arte Joven \$v Periodicals.	611 27 Encuentro Nacional de Arte Joven. \$2 FAST \$0 (OCoLC) fst01408793 655 7 Periodicals. \$2 FAST \$0 (OCoLC) fst01411641

¹² Ashton y Kent, *New Approaches*, 552.

EM	FAST
651 0 Mexico \$x History \$y Revolution, 1910-1920.	647 7 Revolution \$c (Mexico : \$d 1910-1920) \$2 FAST \$0 (OCoLC)fst01354542 651 7 Mexico. \$2 FAST \$0 (OCoLC)fst01211700 655 7 History. \$2 FAST \$0 (OCoLC)fst01411628
650 0 Mexico City Earthquake, Mexico, 1985 \$x Anniversaries, etcétera.	647 7 Mexico City Earthquake \$c (Mexico : \$d 1985) \$2 FAST \$0 (OCoLC)fst01755715 650 7 Anniversaries. \$2 FAST \$0 (OCoLC)fst00809757

Tabla 4. Faceta de reunión o evento (casos tomados del sistema OCLC Connexion).

5. Faceta de tiempo/periodos

Para la conversión, los periodos son expresados con numerales, no con palabras. Un catalogador puede establecer cualquier fecha, o rango de fechas, que considere adecuado. Para la conversión, en general se retoman las subdivisiones del subcampo \$y que expresa el periodo de cobertura de un recurso de información. Cuando aparece un término del tipo “Twentieth century”, o similares, en 650 \$a, este se considera tópico, pues indica al siglo considerado como tema y no como indicación de cuándo ocurrió algún hecho, y se escribe con palabras. De estos términos no se hicieron registros de autoridad en <http://FAST.oclc.org/searchFAST/>.

Ejemplos de posibles casos:¹³

- 1975.
- Since 1951.
- To 1856.
- 1939-1945.
- From 140 to 190 million years ago.
- 146 B.C. -323 A.D.
- 2001 (September 11).
- 1989 (December 1) -1990 (January 20).

13 Tomados de Cao y Salmon. “FAST: Faceted Application of Subject Terminology”.

En este caso, el ajuste principal tiene que ver con las formas en español que habrán de adquirir las menciones de estos periodos.

EM	FAST
610 20 Catholic Church \$x Doctrines \$x History \$y 20th century.	610 27 Catholic Church. \$2 FAST \$0 (OCoLC)fst00531720 650 7 Theology, Doctrinal. \$2 FAST \$0 (OCoLC)fst01149617 <i>** cambio en el término</i> 655 7 History. \$2 FAST \$0 (OCoLC) fst01411628 648 7 1900-1999 \$2 FAST <i>** cambio en el término</i>

Tabla 5. Faceta de tiempo o periodos (casos tomados del sistema oclc Connexion).

6. Faceta de tópico

Desde los EM, esta faceta está conformada por: 1. Los encabezamientos principales de tópico (subcampo 650 \$a) más sus subdivisiones de tópico (subcampos 650 \$x) cuando están presentes; 2. Los subcampos \$x presentes en campos 6XX que no sean 650.

Esta faceta incluye organizaciones imaginarias, por ejemplo:
150__ \$a Umbrella Corporation (*Imaginary organization*)

Asimismo, un campo 650 con una o más subdivisiones \$x, pasa a FAST tal como si estuviera precoordinado (como se observa en el primer caso); es decir, que un 650 \$a va a mantener sus respetivos \$x. Como en otros casos, hay que estar atentos a los cambios en terminología y en codificación que se presenten, de acuerdo con la dinámica de FAST.

EM	FAST
650 0 Education \$x Philosophy.	650 7 Education \$x Philosophy. \$2 FAST \$0 (OCoLC)fst00902721

EM	FAST
<p>650 0 Education \$z Mexico \$x History.</p>	<p>650 7 Education. \$2 FAST \$0 (OCoLC) fst00902499 651 7 Mexico. \$2 FAST \$0 (OCoLC) fst0121170 655 7 History. \$2 FAST \$0 (OCoLC) fst01411628 <i>**nótese que, aunque existe un \$x, como conceptualmente este pasa a la faceta forma, no se mantiene junto al 650\$a de origen, sino que genera una nueva faceta (655)</i></p>
<p>650 0 Education \$x Aims and objectives \$z Mexico \$v Congresses.</p>	<p>650 7 Education \$x Aims and objectives. \$2 FAST \$0 (OCoLC) fst00902507 651 7 Mexico. \$2 FAST \$0 (OCoLC) fst01211700 655 7 Conference papers and proceedings. \$2 FAST \$0 (OCoLC) fst01423772 <i>**cambio en el término</i></p>
<p>651 0 Greece \$x Kings and rulers \$v Biography \$v Juvenile literature.</p>	<p>651 7 Greece. \$2 FAST \$0 (OCoLC) fst01208380 650 7 Kings and rulers. \$2 FAST \$0 (OCoLC)fst00987694 655 7 Biographies. \$2 FAST \$0 (OCoLC)fst01919896 655 7 Juvenile works. \$2 FAST \$0 (OCoLC)fst01411637</p>
<p>630 00 Bible. \$p Old Testament \$x History of Biblical events.</p>	<p>630 07 Bible. \$p Old Testament. \$2 FAST \$0 (OCoLC)fst01808092 650 7 History of Biblical events. \$2 FAST \$0 (OCoLC)fst01865053</p>

EM	FAST
610 20 Texas A & M University \$x Alumni and alumnae.	650 7 Universities and colleges \$x Alumni and alumnae. \$2 FAST \$0 (OCoLC)fst01161636 <i>**el término en \$x adquiere un mayor sentido cuando se le antepone el término en \$a. Para llegar a esta definición es importante la revisión de los registros de autoridad</i> FAST 610 27 Texas A & M University. \$2 FAST \$0 (OCoLC)fst00549598

Tabla 6. Faceta de tópico (casos tomados del sistema OCLC Connexion).

7. Faceta de nombre geográfico

Desde los EM, esta faceta está conformada por: 1. Los encabezamientos principales de nombre geográfico (subcampo 651 \$a), convertidos a la estructura FAST; 2. Los subcampos \$z presentes en campos 6XX. En general, se establece una estructura en orden indirecto, del área geográfica más general a la más específica. Por ejemplo:

651_0 \$a Mixcoac (Mexico City, Mexico) pasa a
651_7 \$a Mexico \$z Mexico City \$z Mixcoac \$2 FAST \$0 (OCoLC)
fst01712073

La razón es que se decidió seguir el orden jerárquico que establece la *marc Code List for Geographic Areas*, el cual consta de tres niveles de referencia. Véanse también a continuación los cambios que operan para áreas geográficas de regiones consideradas grandes dentro del sistema LCSH (concretamente, ciudades o pueblos dentro de Estados Unidos y similares) y también para áreas geográficas internacionales.

EM	FAST
651 0 Pluto (<i>Dwarf planet</i>).	651 7 Pluto (<i>Dwarf planet</i>) \$2 FAST \$0 (OCoLC)fst01710146
651 0 San Diego (Calif.).	651 7 California \$z San Diego. \$2 FAST \$0 (OCoLC)fst01205232 <i>**No se emplean abreviaturas</i>

EM	FAST
651 0 Mexico \$v Biography.	651 7 Mexico. \$2 FAST \$0 (OCoLC) fst01211700 655 7 Biographies. \$2 FAST \$0 (OCoLC)fst01919896
650 0 Latin American literature \$v Bibliography.	650 7 Latin American literature. \$2 FAST \$0 (OCoLC)fst00993031 655 7 Bibliographies. \$2 FAST \$0 (OCoLC)fst01919895
651 0 Africa \$x Social life and customs \$v Fiction.	651 7 Africa. \$2 FAST \$0 (OCoLC) fst01239509 650 7 Manners and customs. \$2 FAST \$0 (OCoLC)fst01007815 655 7 Fiction. \$2 FAST \$0 (OCoLC) fst01423787 655 7 Short stories. \$2 FAST \$0 (OCoLC)fst01726740

Tabla 7. Faceta de nombre geográfico (casos tomados del sistema OCLC Connexion).

8. Faceta de forma/género

Esta faceta se convierte a partir de las subdivisiones codificadas en 6XX \$v, además del listado *Library of Congress Genre/Form Terms*, el cual toma preminencia en muchos casos. De ahí que se observe que las formas habituales de \$v en LCSH pasen a registrarse en plural. Por ejemplo, \$v Bibliography, pasa a ser 655 \$a Bibliographies. En algunos casos se crea un término que no existía en EM como tal; por ejemplo, la forma “Short stories” no existía en autoridades EM y se codificaba como \$v Fiction; con FAST, se crea esta forma y existe tanto 655 \$a Fiction, como 655 \$a Novels y 655 \$a Short stories. El planteamiento aquí es que ya es posible identificar formas específicas, aunque en la conversión de registros preexistentes sólo podría codificarse la forma *Fiction*.

EM	FAST
651 0 Mexico \$v Biography.	651 7 Mexico. \$2 FAST \$0 (OCoLC) fst01211700 655 7 Biographies. \$2 FAST \$0 (OCoLC)fst01919896

EM	FAST
650 0 Latin American literature \$v Bibliography.	650 7 Latin American literature. \$2 FAST \$0 (OCoLC)fst00993031 655 7 Bibliographies. \$2 FAST \$0 (OCoLC)fst01919895
651 0 Africa \$x Social life and customs \$v Fiction.	651 7 Africa. \$2 FAST \$0 (OCoLC)fst01239509 650 7 Manners and customs. \$2 FAST \$0 (OCoLC)fst01007815 655 7 Fiction. \$2 FAST \$0 (OCoLC)fst01423787 655 7 Short stories. \$2 FAST \$0 (OCoLC)fst01726740

Tabla 8. Faceta de forma o género (casos tomados del sistema OCLC Connexion).

DESARROLLO DE REGLAS DE CONVERSIÓN EM-FAST EN REGISTROS MARC

De acuerdo con el análisis anterior, y a partir de la documentación revisada, se desarrolló un mapeo de los campos MARC21 con su correspondencia codificada en la iniciativa FAST. Se puede observar que resulta una codificación en campos más específicos para FAST que para EM, en cuya estructura las diferentes facetas pueden encontrarse aglutinadas en una sola etiqueta MARC.

EM	FAST	MARC FAST
600 \$abcdq	Nombre de persona	600
610 \$abndc	Nombre corporativo	610
611 \$andc	Reunión	611
651 \$y (cuando hay una frase como revolución, sitio, etcétera)	Evento	647
650 \$a (cuando conceptualmente refiere un acontecimiento histórico. Por ejemplo, un sismo, un huracán, etcétera)	Evento	647

EM	FAST	MARC FAST
630 \$adp	Título uniforme	630
6XX \$t	Título uniforme	630
650 \$a + \$x	Tópico	650
6XX \$x (excepto 650 \$x)	Tópico	650
6XX \$y (cuando no hay frase)	Tiempo	648
651 \$a	Nombre geográfico	651
6XX \$z	Nombre geográfico	651
6XX \$v	Forma	655
6XX \$x (en algunos casos)	Forma	655

Tabla 9. Mapeo EM-FAST (elaboración propia).

A partir de este mapeo, y de las definiciones adoptadas para cada faceta, se operó en casos y ejemplos reales de registros de la BDCV con el objetivo de desarrollar dos reglas de normalización que permitieran probar y establecer procesos de conversión automática. En esta etapa se desarrollaron múltiples pruebas, por campo y por caso, se desarrollaron las instrucciones de modo que al activar ambas reglas los EM existentes en los registros bibliográficos cambiasen a FAST. El siguiente paso después de la generación de las reglas fue aplicarlas en un número mayor de registros bibliográficos con el fin de detectar situaciones que requieran afinar algún detalle de su funcionamiento.

De ese modo, una vez hechas suficientes pruebas con un conjunto reducido de casos, las reglas fueron aplicadas a 3,157 registros bibliográficos, los cuales contienen 10,845 EM y de los mismos se derivan 17,976 términos FAST. Al revisar por separado los términos FAST, se detectó que en 1,808 casos no se había logrado la conversión requerida (es decir, el resultado de la conversión no correspondía a la forma establecida por FAST), lo cual representó el 10% del total de los casos convertidos, lo que significa que las reglas funcionaron bien en un 90% de los casos. Asimismo, cuando se tomaron por separado los 1,808 casos que no lograron una conversión adecuada se pudo determinar que en 1,288 de ellos se debió a errores ya sea de codificación en MARC, de estructura del EM o de digitación, los cuales no tienen que ver con las reglas de conversión sino con la historia del catálogo en sí mismo; los 520 casos restantes sí correspondieron a una necesidad de ajuste en las reglas de normalización, principalmente debido a que para ese momento las reglas no consideraron tales casos, los cuales tenían que ver con: la estructura de nombres geográficos en más de un

país (2 casos), los nombres de eventos (49 casos) o formas que pasan de 650\$x a 655 (469 casos) con cambio de codificación o terminología. Para subsanar, se desarrolló una tercera regla de normalización, la cual va integrando los casos aislados detectados, particularmente los que tienen que ver con un cambio terminológico o en la codificación; por lo que con esta regla la efectividad de las reglas de normalización alcanza al momento un 97% a reserva de revisar los casos identificados con errores de codificación o digitación (1288).

El proyecto continúa en curso y la idea sigue siendo localizar casos de cambios que no logran hacer las reglas de normalización pero que, una vez establecidos, logren impactar en conversiones adecuadas a partir de cambios globales en grandes cantidades de registros. La idea también es lograr la conversión en una mucho mayor cantidad de registros bibliográficos, aunque el foco de atención es contar con un sistema aplicable a la descripción temática de colecciones digitales principalmente, que logre integrarse de manera transparente en sus propios esquemas de metadatos.

Como corolario, deben destacarse algunas consideraciones a tener en cuenta en la implementación de la iniciativa FAST:

1. La efectividad de la conversión depende de la correcta codificación MARC y asignación adecuada de los EM. Asimismo, la aplicación en múltiples registros permitirá fundamentar la decisión de emplear FAST como el sistema de asignación de temas.
2. Para asimilar, consultar y corroborar la creación de términos FAST, tener presente <http://FAST.oclc.org/searchFAST/> y WorldCat como fuentes de control de autoridad y de aplicación en registros bibliográficos.
3. Para la faceta “tópico” mapeada desde 6XX \$x, en ocasiones la conversión no lleva necesariamente a 650, sino a 655. Por ejemplo, en el caso de 6XX \$x History o de \$x Criticism, interpretation, etcétera.
4. Considerar los cambios en la terminología, aunque se conserve la faceta en la conversión. Por ejemplo: 6XX \$v Congresses, pasa como 655 \$a Conference papers and proceedings.
5. Como se observa en los dos puntos anteriores es necesario tener presentes los cambios a nivel conceptual y terminológico, así como a nivel estructural, esto último particularmente considerando las políticas locales de una biblioteca en particular, como es el caso de la BDCV que tiene una política propia sobre la representación de nombres geográficos.
6. Otro aspecto relevante tiene que ver con la traducción y adecuación de los términos FAST a la lengua española, aunque justamente el trabajo acumulado hasta ahora en la asignación de EM es el que se retoma como base para FAST.

7. Debe tenerse presente el desarrollo del control de autoridades de términos FAST para garantizar la consistencia de su aplicación.

CONCLUSIONES

Se ha mostrado la lógica y las necesidades subyacentes a la formulación de la propuesta FAST. A partir de esto, se pone sobre la mesa la valoración de su desarrollo y utilización en sistemas de información bibliográfica. Particularmente, es relevante la mejor adecuación de esta propuesta de recuperación de información temática, más allá del sistema que imponen los EM, los cuales de muchas formas han entrado en desuso.¹⁴ La dinámica de facetas, y de aplicación de filtros y relaciones entre ellas, se ajusta mejor a sistemas de descubrimiento y a proyectos como repositorios institucionales u organización de colecciones digitales; los términos FAST pueden hacer más llevaderos, ágiles y económicos los procesos de análisis y síntesis de información temática a partir de una dinámica más acorde con las necesidades de procesamiento de información en el entorno digital, en donde se busca que los términos, vistos como metadatos, sean sencillos, ágiles, rápidos de producir, con poco esfuerzo y que resulten reutilizables. Nos toca revisar sus posibilidades. No debe descartarse que incluso a nivel de la creación de registros bibliográficos tradicionales en una biblioteca se plantee la sustitución de los EM por FAST.

¿Qué se puede decir que nos corresponde ahora hacer en tanto organizadores de información? Aquí algunas valoraciones:

- Para empezar, la ponderación de por qué en estos tiempos resulta mejor una asignación temática tipo keywords (términos sueltos facetados) por sobre una cadena de significado tipo EM, anclada mayormente esta en entornos de indización manual.
- Nos corresponde comprender la dinámica de generación y de aplicación de la terminología FAST en sistemas de información, sobre todo frente al hecho de que cada vez más nos vemos involucrados en proyectos de organización de información que van más allá de los procesos habituales de las bibliotecas.
- Nos corresponde analizar si procede, y en qué condiciones, una sustitución de los EM por FAST: ventajas, desventajas e implicaciones. Una pista

14 Noé Ángeles Escobar, *La relevancia del acceso temático con encabezamientos de materia en los catálogos de biblioteca* (Tesis de maestría, Universidad Nacional Autónoma de México, 2011).

que arroja luz es el hecho de la manera en que han caído en desuso los EM, particularmente la consulta de su estructura habitual, en la organización y recuperación de información en el entorno digital. Entonces, nos tocará evaluar la obsolescencia de los EM.

- Nos corresponde analizar qué necesidades de representación de información puede resolver el sistema FAST. También, nos corresponde hacer lo anterior en español, de acuerdo con la revisión presentada aquí, esto será más sencillo que el control de vocabulario y la asignación temática que se realiza con los EM.
- Nos corresponde explorar las aplicaciones de FAST para la asignación y conversión de términos.
- Nos corresponde hacer pruebas de implementación.

Hasta aquí se ha mostrado una ruta para hacer operativa la manera de pasar de EM a FAST, de acuerdo con las justificaciones presentadas.

REFERENCIAS BIBLIOGRÁFICAS

- Anderson, James D., y Melissa A. Hofmann. 2006, "A Fully Faceted Syntax For Library of Congress Subject Headings", *Cataloging & Classification Quarterly* 43, n.º 1, 2006, 7-38, https://doi.org/10.1300/j104v43n01_03
- Ángeles Escobar, Noé, *La relevancia del acceso temático con encabezamientos de materia en los catálogos de biblioteca*, Tesis de maestría, Universidad Nacional Autónoma de México, 2011.
- Ashton, Janet y Caroline Kent, "New Approaches to Subject Indexing at the British Library", *Cataloging & Classification Quarterly* 55, n.º 7-8, 2017, 549-59, <https://doi.org/10.1080/01639374.2017.1354345>
- Chan, Lois Mai y Edward T O'Neill, *FAST: Faceted Application of Subject Terminology: Principles and Applications*, Santa Barbara, Calif.: Libraries Unlimited, 2010.
- Cao, Haiyun y Marcia Salmon, *FAST: Faceted Application of Subject Terminology*, Presented at OLA Super Conference, Thursday February 25, 2010, <https://yorkspace.library.yorku.ca/xmlui/bitstream/handle/10315/2996/FAST.pdf>
- Chew, Chiat Naun, Kerre Kammerer, Kim Mumbower y Dean Seeman, *FAST Quick Start Guide*, Ohio: OCLC FAST Policy and Outreach Committee, <https://www.oclc.org/content/dam/oclc/FAST/FAST-quick-start-guide-2022.pdf>
- Harper, Corey A., y Barbara B. Tillett, "Library of Congress controlled vocabularies and their application to the Semantic Web", *Cataloging & Classification Quarterly* 43, n.º 3-4, 2007, 47-68.
- Jin, Qiang, "Is FAST the right direction for a new system of subject cataloging and metadata?", *Cataloging & Classification Quarterly* 45, n.º 3, 2008, 91-110, doi: 10.1300/J104v45n03_08
- Mixer, Jeffrey y Eric R. Childress, *FAST (Faceted Application of Subject Terminology) Users: Summary and Case Studies*, Dublin, Ohio: OCLC Research, 2013, <http://www.oclc.org/content/dam/research/publications/library/2013/2013-04.pdf>.
- Montalvo Montalvo, Marilyn, "LCSH, FAST y DELICIOUS: vocabularios normalizados y nuevas formas de catalogación temática", *Anales de Documentación* 14, n.º 1, 2011, 1-13.

- O'Neill, Edward T. y Lois Mai Chan, *FAST (Faceted Application of Subject Terminology), a simplified lcsb-based vocabulary*, Presented at World Library and Information Congress: 69th IFLA General Conference and Council, 1-9 August 2003, Berlin, https://archive.ifla.org/IV/ifla69/papers/010e-ONeill_Mai-Chan.pdf
- O'Neill, Edward T., Rick Bennett y Kerre Kammerer, "Using authorities to improve subject searches", *Cataloging & Classification Quarterly* 52, n.º 1, 2014, 6-19, doi: 10.1080/01639374.2013.850018
- Smith-Yoshimura, Karen, *Transitioning to the Next Generation of Metadata*, Dublin, Ohio: OCLC Research, 2020, <https://doi.org/10.25333/rqgd-b343>

Los sistemas de anotación como instrumento de organización del conocimiento

Annotation systems as an instrument for knowledge organizations

Jesús Tramullas
Depto. Ciencias de la Documentación,
Universidad de Zaragoza

INTRODUCCIÓN

La organización de la información a nivel individual ha sido objeto de estudio desde la década de 1980, en el marco de lo que la bibliografía especializada ha denominado como “gestión de información personal” (*personal information management, PIM*).¹ En la gestión de información personal desempeñan un papel fundamental las tareas de *organización* de información, en especial en aspectos relativos al etiquetado de la información, y a la creación de sistemas y pautas para ello, y su aplicación a la organización y recuperación de las colecciones creadas por los usuarios y usuarias. Tareas que no sólo contemplan la organización de la información para su uso inmediato, sino que también se desarrollan anticipando posibles usos futuros.

En paralelo al estudio académico, han ido apareciendo herramientas de *software* que pretendían facilitar la integración y organización de la información personal. En las décadas de 1980 y 1990 estas herramientas se centraban en integrar información proveniente de diferentes fuentes (correo electrónico, documentos ofimáticos, agendas...), a través de interfaces más o menos homogéneas. Es en la década de 2000 cuando empiezan a aparecer herramientas que facilitan la creación y organización de información, e incorporan funcio-

1 Abdus Sattar Chaudhry y Bibi M. Alajmi, “Personal information management practices: how scientists find and organize information”, *Global Knowledge, Memory and Communication*, ahead-of-print (1 de enero de 2022), <https://doi.org/10.1108/GKMC-04-2022-0082>

nalidades de etiquetado de información, que posteriormente se amplían a la creación de relaciones entre los diferentes ítems informacionales. A ello hay que unir una creciente tendencia a compartir diferentes facetas de la información personal en grupos colaborativos, en especial en entornos de trabajo y desarrollo.

Los sistemas de anotación se han revelado durante la pasada década como una herramienta clave para la organización y estructuración de información personal, ya que facilita la integración de la producción de información personal con la organización de la misma a través del uso de etiquetados, y con la creación de estructuras bi y tridimensionales de relaciones relevantes entre las propias informaciones. A su vez, han incorporado las prestaciones de acceso a información externa que ofrecían los gestores de información personal de décadas anteriores. De la misma forma, muchos de ellos están preparados para la publicación e intercambio de información en entornos colaborativos abiertos.

LA GESTIÓN DE INFORMACIÓN PERSONAL

La definición comúnmente aceptada de la gestión de información personal, expresión que aparece en la década de 1980 en contexto de la digitalización de actividades y tareas, establece que:

Personal Information Management (PIM) refers to the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use, and distribute information in each of its many forms (paper and digital, in e-mails, files, Web pages, text messages, tweets, posts, etc.) as needed to meet life's many goals (everyday and long-term, work-related and not) and to fulfill life's many roles and responsibilities (as parent, spouse, friend, employee, member of community, etc.).²

Esta formulación teórica pretende abarcar todos los tipos de información, en diferentes medios o soportes, que las personas necesitan en su quehacer diario, tanto en la esfera profesional como en la personal. Es el usuario o

2 William Jones, "Personal Information Management", *Annual Review of Information Science and Technology* 41, n.º 1, 2007, 453-504, <https://doi.org/10.1002/aris.2007.1440410117>; William Jones, Jesse David Dinneen, Robert Capra, Anne R. Diekema, y Manuel A. Pérez-Quñones, "Personal Information Management", en *Encyclopedia of Library and Information Sciences*, 4.ª ed. (Taylor y Francis, 2017), <https://doi.org/10.1081/E-ELIS4-120053695>

usuaria quien decide, en cada caso y para cada contexto, qué es información personal. La realidad y el comportamiento informacional de los usuarios y usuarias se ha encargado de demostrar que una solución global y completa al fenómeno de la información personal es ilusoria: siempre se toman decisiones en virtud de necesidades específicas, se seleccionan y priorizan unas soluciones frente a otras, la información no siempre se encuentra en el momento y lugar precisos, y no siempre está disponible. El “olvido” es uno de los factores clave que planean sobre el desarrollo de la gestión de información personal. En el momento actual, hay que considerar que la gestión de información personal es eminentemente operacional, individual, mediada tecnológicamente, y su material de trabajo es cualquier tipo de información digital.

Para enfrentar este problema, la gestión de la información personal ha adoptado, en su formulación clásica, la idea de la colección de información personal, formada por elementos o ítems informativos sobre los cuales se aplican principios clásicos de organización de la información. A tal fin, se creaban esquemas organizativos sobre estas colecciones, pero la variabilidad de los elementos que las formaban acababa dando como resultado colecciones dispersas, no integradas, que llevaban a la denominada fragmentación de la información. El intento de superar esta fragmentación ha llevado a la simplificación de los procesos de gestión de la información personal y ha favorecido la aparición de nuevas herramientas de *software*.

Los elementos o ítems de información pueden ser muy variados. Como cualquier documento, pueden ser creados, almacenados, distribuidos, transformados, borrados y, como no podía ser menos, organizados y descritos. Los conjuntos de elementos o ítems se organizan, en un primer nivel, en colecciones. Estas colecciones están formadas, además de por los elementos o ítems, por su representación o descripción, sus propiedades, por descriptores o etiquetas, y por las relaciones existentes entre ellos, que pueden responder a diferentes parámetros lógicos.

Los investigadores sobre la cuestión han establecido que se llevan a cabo tres tipos principales de actividades sobre la información personal:³

- Recopilación: las tareas de adquirir, consumir o tratar y almacenar la información, en virtud de necesidades presentes y futuras. La variabilidad de las necesidades hace que sean tareas que muestran múltiples facetas y posibilidades.

3 Jorge Franganillo, “Gestión de información personal: elementos, actividades e integración”, *Profesional de la información* 18, n.º 4 (8 de agosto de 2009), 399-406, <https://doi.org/10.3145/epi.2009.jul.06>

- **Búsqueda/rebúsqueda:** las actividades de búsqueda de información, bien a través de búsquedas clásicas mediante ecuaciones, bien a través de tareas de ordenación, navegación y exploración de los contenidos de la colección o colecciones. Suele tratarse de actividades iterativas.
- **Mantenimiento:** las tareas relacionadas con el mantenimiento técnico de las colecciones, el flujo de información, la seguridad, la evaluación y la organización. Son tareas básicas, pero en muchas ocasiones son consideradas secundarias, o no urgentes, por los usuarios.

Con el desarrollo del trabajo colaborativo mediado tecnológicamente en la década de 1990 y la popularización de entornos sociales llegados de la mano, el web 2.0 en la década siguiente, se ha introducido en el ámbito de la gestión de información personal la posibilidad de compartir e intercambiar elementos y colecciones con otros usuarios en espacios digitales comunes. En estos contextos, las actividades pueden incorporar una división de tareas entre los participantes, o en compartir, con diferentes niveles de acceso, las colecciones individuales de los miembros. En entornos organizativos, en muchas ocasiones, estas acciones se llevan a cabo de manera independiente a los procesos y procedimientos institucionalizados. La gestión de la información personal da paso a la gestión del conocimiento personal, que supera los límites clásicos del tratamiento de información, integrando los elementos o bloques de conocimiento, un modelo organizativo semántico de los mismos, y el marco estructural de interrelación de los elementos.⁴

HERRAMIENTAS DE SOFTWARE PARA LA GESTIÓN DE INFORMACIÓN PERSONAL

Las herramientas de *software* para la gestión de información personal⁵ se desarrollan a partir de la década de 1980. En una primera generación, estas herramientas ofrecían funcionalidades que permitían gestionar y organizar conjuntos de documentos a los que se asignaba una somera descripción, y que podían organizarse en estructuras de carpetas, siendo acompañados en ocasiones con la posibilidad de asignarles etiquetas descriptivas. El desarrollo

4 Ulrich Schmitt, "Shortcomings of the web of documents and data for managing personal knowledge and collaboration", en *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 2017, 23-29, <https://doi.org/10.1109/NEXTCOMP.2017.8016171>

5 William Jones, *Transforming Technologies to Manage Our Information: The Future of Personal Information Management, Part 2*, Synthesis Lectures on Information Concepts, Retrieval, and Services (Cham: Springer, 2014), <https://doi.org/10.1007/978-3-031-02329-3>

acelerado de las aplicaciones de internet en la segunda mitad de la década de 1990 y comienzos de 2000 trajo consigo que los llamados “gestores de información personal” pasasen a ofrecer una interfaz en la que se incluían el correo electrónico y la gestión de calendarios y agendas de trabajo. De esta manera, se pretendía ofrecer un entorno único de trabajo que integrase todos los tipos de información que los usuarios podían necesitar para el desarrollo de sus tareas.

Sin embargo, este enfoque, aunque produjo algunas herramientas de *software* interesantes, se reveló como insatisfactorio para los usuarios, lo que explica el escaso éxito de las aplicaciones. Se trataba de una integración a nivel de interfaz de usuario, pero no ofrecía una integración de información relevante ni útil que permitiese la creación y el uso de relaciones entre los diferentes elementos informativo-documentales. Por otra parte, no habían sido desarrollados instrumentos de visualización de la información que ofreciesen un acceso a la información relevante, superando el clásico listado de datos.

En la década de 2000 aparecieron en el mercado aplicaciones que combinaban la organización de documentos y de otros elementos informativo documentales, con visualizaciones gráficas basadas en la noción de asociación (el principio clásico del hipertexto). Esto trajo un nuevo interés por los gestores de información personal, beneficiados por la difusión de herramientas basadas en la asociación de conceptos, como los mapas conceptuales y los *topic maps*.⁶ A ello contribuyó la mayor disponibilidad de librerías y aplicaciones que eran capaces de generar visualizaciones bi y tridimensionales basadas en el procesamiento de datos. Sin embargo, estas aplicaciones seguían adoleciendo de una carencia de funcionalidades para la incorporación de información relacionada por parte del usuario final. Al igual que sucedió en el periodo anterior, el éxito e impacto de estas herramientas fue limitado, dado que el modelo teórico subyacente se centraba en la organización de conceptos y de sus relaciones, antes que en el contenido y representación de documentos.

Hay que esperar a finales de la década de 2010 para la aparición de gestores de información personal que superan las limitaciones previas y que suponen un cambio de enfoque. La unidad informativo-documental, la unidad de trabajo para el usuario, es la anotación textual, y sobre esta se pueden construir estructuras organizativas del conocimiento registrado por el usuario. A la anotación como núcleo acompañan las funcionalidades clásicas de organización de documentos en estructuras de carpetas, el etiquetado, y el establecimiento de relaciones entre anotaciones o documentos. El acceso a la información se beneficia de la generación en tiempo real de visualizaciones gráficas interac-

6 Jesús Tramullas, Ana-I Sánchez-Casabón, y Piedad Garrido-Picazo, “Gestión de información personal con *software* para mapas conceptuales”, *El Profesional de la Información* 18, n.º 6 (noviembre de 2009), 601-12, <https://doi.org/10.3145/epi.2009.nov.03>

tivas de las estructuras creadas, así como de búsquedas sobre texto completo. Además de la nota o anotación textual, se ha comenzado a incorporar prestaciones de pizarra gráfica, que facilitan el “dibujo” de esquemas y otros tipos de representaciones de las estructuras de información.⁷

Las anotaciones se escriben usando el lenguaje de marcado *Markdown*. Se trata de un lenguaje de marcado ligero, con una curva rápida de aprendizaje, que se ha expandido rápidamente en un buen número de contextos de uso, desde la programación web hasta el procesamiento de textos y la edición técnica.⁸ Dado que es extensible, y que puede incorporar metadatos y enlaces hipertextuales, se adecúa a las necesidades de creación rápida de estructuras de conocimiento relacionales.

Una característica importante, y en la actualidad imprescindible, de estas herramientas de *software* es su capacidad para sincronizar las estructuras de información entre diferentes plataformas (sobremesa, dispositivos móviles...) lo que implica su utilización con diferentes interfaces. Además de la sincronización, que hace posible que el usuario siempre tenga acceso a la información estructurada y organizada, algunas de ellas añaden capas de seguridad, lo que resulta relevante en cuanto se habla de información personal. Junto a la sincronización también se han añadido, en algunas herramientas, capas de funcionalidad que hacen posible el trabajo colaborativo entre diferentes usuarios, lo que significa la capacidad de compartir y trabajar en grupo sobre estructuras de información.

Estas herramientas ofrecen a los usuarios y usuarias la posibilidad de crear y gestionar esquemas personales de organización del conocimiento de manera interactiva, en un entorno real y cambiante, y ajustados a sus necesidades individuales. La actual generación de sistemas de anotación (*Roam*, *Obsidian*, *Notion*... por citar las más conocidas) incorporan prestaciones de los sistemas hipertextuales clásicos que amplían las capacidades comunes de organizar y recuperar información. Curiosamente, estas herramientas son identificadas entre los usuarios como *note-taking apps* o *annotation apps*, dado que la función principal es la de toma de notas o la de anotación textuales. Sin embargo, en las descripciones más detalladas se hace referencia a las mismas como gestores personales de información o de conocimiento.

7 Williams Ezinwa Nwagwu y Chidiebube Blossom Williams, “Knowledge Mapping and Visualization of Personal Information Management Literature, 1988-2020”, *IFLA Journal*, 22 de febrero de 2022, <https://doi.org/10.1177/03400352211065487>

8 Jesús Tramullas, “Desarrollos en elaboración de documentación técnica: los lenguajes de marcado ligero”, *Anuario ThinkEPI*. 13, 2019, e13f03, <https://doi.org/10.3145/thinkepi.2019.e13f03>

LA ANOTACIÓN

La anotación es una técnica básica de incorporación de información a un documento, que puede responder a diferentes criterios y objetivos. Históricamente, la anotación es tan antigua como el documento escrito;⁹ en la actualidad, el desarrollo tecnológico de las herramientas de *software* ha hecho posible que la anotación desborde el ámbito original, facilitando la creación de estructuras de información relacionada que combinan diferentes tipos de información, distribuidas en diferentes tipos de elementos y organizadas según los intereses y criterios específicos del usuario o grupo de usuarios interesados en la cuestión. La anotación es una práctica que debe contemplarse siempre en el marco de su contexto social y cultural, lo que explica sus diferentes usos y acepciones históricas.¹⁰ La noción de anotación será diferente para un estudiante, un investigador, un informático o un periodista, atendiendo a este hecho.

Para una perspectiva operacional básica, una anotación es una nota añadida a un texto o a otro tipo de documento. Una nota puede ir desde un término, una fecha, una referencia o una frase, hasta un texto explicativo más o menos extenso o un reenvío a otros documentos diferentes que ofrecen información relacionada (y que no tienen por qué ser documentos textuales, a su vez). Los diferentes tipos de medios que los usuarios integran en su actividad diaria y sobre los que pueden hacer anotaciones multiplican los contextos y usos de las notas y de la anotación. Desde el ámbito de la bibliografía, Beatty y Cochran han diferenciado cinco tipos diferentes de anotaciones que identifican como descriptivas, sumativas, evaluativas, reflexivas y combinadas.¹¹ El uso de la anotación y de los métodos de anotación es constante en el campo de la investigación científica.¹²

La anotación significa, además, que el usuario participa en el desarrollo de una estructura informativa nueva, superando el documento fuente, al que anota, y al cual expande mediante sus propias contribuciones. Se puede crear una situación de interacción que lleva a la creación de intertextualidad mediante el establecimiento de un diálogo. No es el objetivo de este texto analizar en profundidad las características de la intertextualidad que puede generarse

9 Elaine Leong, "Read. Do. Observe. Take Note!", *Centaurus* 60, n.º 1-2, 2018, 87-103, <https://doi.org/10.1111/1600-0498.12203>

10 Remi Kalir y Antero García, *Annotation* (Cambridge, MA: MIT University Press, 2021), 5-7.

11 Luke Beatty y Cindy Cochran, *Writing the Annotated Bibliography* (New York: Routledge, 2021), 12.

12 Joanna Pitura, "Digital Note-Taking for Writing", en *Digital Writing Technologies in Higher Education: Theory, Research, and Practice*, ed. Otto Kruse et al. (Cham: Springer, 2023), 101-19, https://doi.org/10.1007/978-3-031-36033-6_7

a partir de las anotaciones, así que se limitará a señalar precisamente el desarrollo y potencial de la misma en virtud, precisamente, de su capacidad para generar estructuras de conocimiento que amplían el contenido informativo y documental, al mismo tiempo que proponen nuevas formas de organización y acceso a la información.

Por último, cabe señalar que, tradicionalmente, las anotaciones podían quedar circunscritas a un entorno personal o privado. El desarrollo de las herramientas de *software* específicas para estas tareas ha hecho posible que las anotaciones y los documentos a los que hacen referencia puedan compartirse y puedan trabajarse en entornos colaborativos mediados tecnológicamente. Es posible establecer relaciones no sólo entre documentos y anotaciones sino también entre usuarios, y de esta interacción, que retoma la idea de diálogo en la organización del conocimiento, desarrollar nuevos esquemas dinámicos y evolutivos de organización de información con la capacidad de adaptarse a contextos y necesidades cambiantes.¹³

Kalir y García han identificado cinco usos y entornos principales para la actividad de anotar:¹⁴

1. Suministrar información: las anotaciones pueden suministrar información que amplía el documento original, aportando datos sobre el contexto en el que se ha producido y se recibe, y sobre el modo en que se interpreta. Un ejemplo clásico de este tipo de anotación son las etiquetas, que facilitan la comprensión del documento o de su contenido en el contexto. De la misma forma pueden entenderse los metadatos, que pueden aportar información sobre origen, aspectos administrativos, relaciones, etcétera. Otra cuestión a considerar es que estas anotaciones pueden estar enfocadas a diferentes grupos de usuarios o interés, además que sus procesos de creación puedan ser individuales o colaborativos.
2. Compartir comentarios: el proceso de diálogo sobre los documentos mediado por las anotaciones ha sido históricamente asincrónico. Un usuario anotaba, para que otros usuarios accediesen al contenido posteriormente, generalmente de manera individual. A su vez, estos nuevos usuarios podían convertirse en nuevos anotadores del original. Se genera de esta forma una cadena de anotaciones que adoptan la forma de comentarios. La mediación tecnológica ha difuminado en el momento

13 Deborah N. Brewis y Sarah Taylor Silverwood, "Annotation", en *Writing Differently*, ed. Alison Pullen, Jenny Helin, y Nancy Harding, vol. 4, *Dialogues in Critical Management Studies* (Emerald, 2020), 67-90, <https://doi.org/10.1108/S2046-60722020000004008>

14 Kalir y García, *Annotation*, 33-160.

actual los límites entre anotación y comentario, como puede apreciarse en las conversaciones en redes sociales, al mismo tiempo que permite diálogos sincrónicos interactivos entre usuarios.

3. Ampliar la conversación: el punto anterior introduce la importancia de la anotación para crear y expandir conversaciones o diálogos alrededor del documento, su contexto y sus usos. Un ejemplo clásico de conversación es el uso de anotaciones en el ámbito académico, que pueden ser de diferente tipo, como comentarios, referencias bibliográficas, revisiones... o el intercambio de interacciones en un web abierto a comentarios de los usuarios (de cine, productos comerciales, etcétera).
4. Expresar poder: compartir anotaciones sirve para expresar diferentes puntos de vista, para exponer diferentes discursos de manera pública, y supone un empoderamiento de los autores. Las herramientas de *software* han posibilitado una expansión de las anotaciones y las referencias cruzadas, dando voz y presencia a diferentes enfoques y aproximaciones que pueden adquirir influencia y penetración entre las concepciones de los usuarios.
5. Ayudar al aprendizaje: la anotación es un instrumento básico en los procesos educativos y ha sido usada históricamente, pudiendo encontrarse manuales y guías de métodos y técnicas. Las actividades suelen llevar asociadas diferentes tipos de anotaciones que responden a diferentes propósitos. El aprendizaje puede beneficiarse de los resultados de la anotación, y se han creado y desarrollado plataformas colaborativas de anotación orientadas al aprendizaje. La anotación social en estos entornos facilita el acceso e intercambio de información, la colaboración y la producción de conocimiento.

LA ORGANIZACIÓN DE INFORMACIÓN CON HERRAMIENTAS DE ANOTACIÓN

La gestión de la información personal, la anotación, en sus diferentes tipos y usos, y el desarrollo de herramientas de *software* se han combinado para crear un entorno, mediado tecnológicamente, en el cual pueden crearse, compartirse y desarrollarse sistemas y esquemas de organización del conocimiento y de la información. Desde su mismo inicio, la anotación ha estado relacionada con las tecnologías disponibles en cada momento.

Las herramientas de *software* para anotación se han desarrollado en los últimos años, incorporando funcionalidades y prestaciones que hacen posible la creación de espacios de información en los cuales aplicar diferentes esquemas de organización de la información. Estos esquemas integran tanto los propios

documentos como las estructuras desarrolladas alrededor de ellos, configurando lo que se ha denominado bases de información personal,¹⁵ que recuerdan las ideas delineadas por precursores como P. Otlet, V. Bush, T. Nelson o D. Engelbart, y que intenta dar respuesta a los problemas de la gestión de la información y del conocimiento personal señaladas en un apartado anterior.

Este enfoque se encuentra en herramientas de *software* que reciben denominaciones diversas y cuyas funcionalidades pueden ser también diferentes.¹⁶ Gestores de información personal, aplicaciones de toma de notas, gestores de conocimiento personal, herramientas de anotación... Algunas herramientas o plataformas son propietarias, otras son libres; son aplicaciones web o herramientas de escritorio (o combinadas). En ocasiones pueden formar silos aislados de información; en otros casos, atienden a estándares para facilitar el acceso y la preservación. Esta variabilidad indica que se trata de un conjunto de herramientas que se encuentran en proceso de desarrollo y consolidación, de las que puede esperarse una progresiva estandarización a corto plazo.

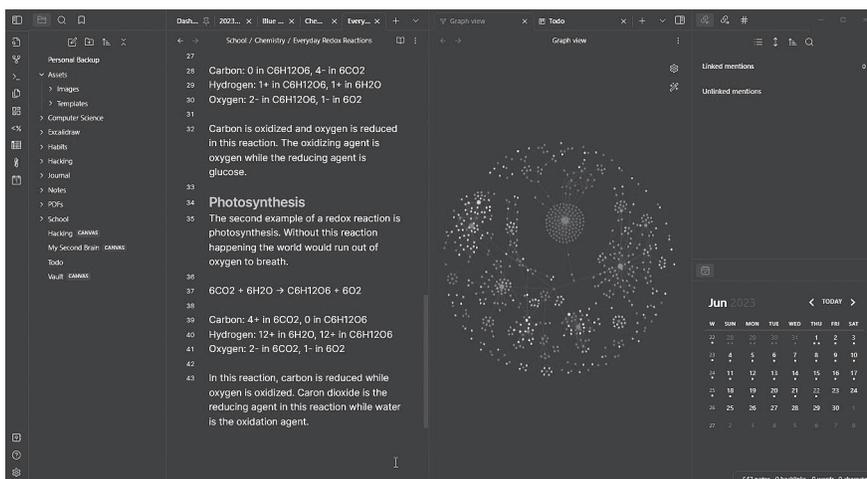


Ilustración 1. Interfaz de *Obsidian* (Fuente: Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Obsidian_desktop_demo_workflow.jpg).

15 Stephen Davies, “Still building the memex”, *Communications of the ACM* 54, n.º 2 (1 de febrero de 2011), 80-88, <https://doi.org/10.1145/1897816.1897840>

16 Mark A. Finlayson y Tomaž Erjavec, “Overview of Annotation Creation: Processes and Tools”, en *Handbook of Linguistic Annotation*, ed. Nancy Ide y James Pustejovsky (Dordrecht: Springer, 2017), 167-91, https://doi.org/10.1007/978-94-024-0881-2_5

Independientemente de su uso, generalmente se identifican tres tipos principales de herramientas de *software* para organización de información personal que se basan en el uso de la anotación:

1. Aplicaciones de anotación o para toma de notas (*note-taking*), aquellas que se orientan a la elaboración de anotaciones, principalmente textuales.
2. Wikis personales: entornos destinados a un uso individual, que replican los modelos clásicos de organización hipertextual de los wikis.
3. Organizadores personales: aplicaciones que permiten integrar y estructurar la información personal de diferentes fuentes, siguiendo los esquemas de los gestores de información personal tradicionales.

Sin embargo, debe llamarse la atención sobre el hecho de que, independientemente de su denominación, todos los tipos ofrecen funcionalidades que hacen posible la descripción de los documentos o elementos informativos, su etiquetado y el establecimiento de relaciones entre los elementos. Esto permite delinear las funcionalidades básicas de tratamiento y organización de la información que ofrecen:

1. Edición de notas: funcionalidades de creación, edición y borrado de notas, usando lenguajes de marcado ligero.
2. Etiquetado: creación y gestión de etiquetas significativas para el usuario, así como procesos de búsqueda y selección basados en ellas.¹⁷
3. Organización de notas: organización de los conjuntos de anotaciones con criterios jerárquicos (carpetas o subcarpetas), temporales, etcétera. Junto a la anterior, son las dos funcionalidades más utilizadas por los usuarios.¹⁸
4. Relaciones: creación de enlaces hipertextuales entre anotaciones y con documentos. Algunas aplicaciones ofrecen enlaces bidireccionales, como define el modelo hipertextual clásico.
5. Visualización de relaciones: generación y navegación interactiva de representación visual de las estructuras de organización de la información creada, normalmente atendiendo al etiquetado y a las relaciones definidas previamente. No todas las herramientas de anotación ofrecen esta funcionalidad.

17 Tiziana Catarci *et al.*, “Structure Everything”, en *Personal Information Management*, ed. William Jones y Jaime Teevan (Seattle: Univ. of Washington Press, 2007), 108-26.

18 Ofer Bergman *et al.*, “Folder versus tag preference in personal information management”, *Journal of the American Society for Information Science and Technology* 64, n.º 10 (octubre de 2013), 1995-2012, <https://doi.org/10.1002/asi.22906>

6. Integración de bibliografía: funcionalidades para recuperar e integrar referencias bibliográficas provenientes de gestores de referencias, como *Zotero*.
7. Intercambio de información: capacidad de importación/exportación de las anotaciones.

Las herramientas y sus funcionalidades adquieren todo su potencial cuando se usan en el marco de un proceso sistemático de organización de información. El método actualmente más extendido para la gestión de anotaciones (no bibliográficas) es el conocido método *Zettelkasten*, que fue formulado por el investigador N. Luhmann.¹⁹ A los efectos de este texto interesa señalar que Luhmann identificaba cuatro tipos de anotaciones correspondientes a notas rápidas, notas que contenían referencias a bibliografía o fuentes de información, notas estructurales que creaban estructuras organizativas sobre notas relacionadas, y notas permanentes que consolidaban la información (derivadas del tratamiento de las notas rápidas). El principio organizador de las diferentes notas es la relación o asociación. Luhmann también estableció cinco tipos de relaciones entre las anotaciones. Se trata de las relaciones de yuxtaposición, anidación, etiquetas o palabras clave, directas y estructurales.

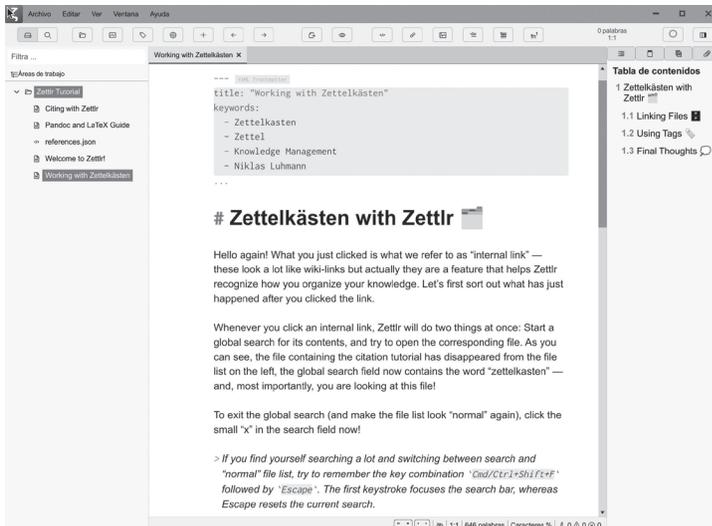


Ilustración 2: Interfaz de Zettlr. Fuente: elaboración propia.

19 Sönke Ahrens, *El método Zettelkasten: Cómo tomar notas de forma eficaz para impulsar la escritura y el aprendizaje de estudiantes, académicos y escritores de no ficción*, 2020.

La combinación de las herramientas de anotación con un método de organización de información y conocimiento basado en notas, como *Zettelkasten*, hace posible la creación de espacios de información personal con esquemas de organización del conocimiento, que se ajustan a las necesidades y requerimientos situacionales de los usuarios. La versatilidad de uso permite que los niveles de organización puedan ir desde un sencillo etiquetado de contenidos hasta la creación de estructuras hipertextuales complejas.

CONCLUSIONES

La anotación conecta y relaciona documentos, conecta y relaciona información, y es una técnica empleada desde el inicio de la producción de documentos escritos. Las herramientas de *software* para anotación facilitan la creación de un entorno en el cual los procesos cognitivos de creación y organización de información y de conocimiento, tanto individuales como colaborativos, pueden desarrollarse y generar estructuras y espacios de información. Los principios y esquemas de organización del conocimiento aplicables en cada situación hacen posible que los usuarios puedan crear estructuras y modelos específicos y especializados de organización del conocimiento de manera independiente a las recomendaciones y a la práctica profesionales, cuestión sobre la que es necesario reflexionar desde los ámbitos de la investigación, la actividad profesional y la educación, dado el impacto que ello tiene en la progresiva digitalización de la sociedad.

REFERENCIAS BIBLIOGRÁFICAS

- Ahrens, Sönke, *El método Zettelkasten: Cómo tomar notas de forma eficaz para impulsar la escritura y el aprendizaje de estudiantes, académicos y escritores de no ficción*, Sönke Ahrens, 2020.
- Beatty, Luke y Cindy Cochran, *Writing the Annotated Bibliography: A Guide for Students & Researchers*, New York: Routledge, 2021.
- Bergman, Ofer, Noa Gradovitch, Judit Bar-Ilan, y Ruth Beyth-Marom, “Folder versus tag preference in personal information management”, *Journal of the American Society for Information Science and Technology* 64, n.º 10, octubre de 2013, 1995-2012, <https://doi.org/10.1002/asi.22906>
- Brewis, Deborah N. y Sarah Taylor Silverwood, “Annotation”, En *Writing Differently*, editado por Alison Pullen, Jenny Helin, y Nancy Harding, 4: 67-90, Dialogues in Critical Management Studies, Emerald Publishing Limited, 2020, <https://doi.org/10.1108/S2046-607220200000004008>
- Catarci, Tiziana, Luna Dong, Alon Halvey y Antonella Poggi, “Structure Everything”, en *Personal Information Management*, editado por William Jones y Jaime Teevan, 108-26, Seattle: Univ. of Washington Press, 2007.
- Chaudhry, Abdus Sattar y Bibi M. Alajmi, “Personal information management practices: how scientists find and organize information”, *Global Knowledge, Memory and Communication* ahead-of-print, n.º ahead-of-print, 1 de enero de 2022, <https://doi.org/10.1108/GKMC-04-2022-0082>
- Davies, Stephen, “Still building the memex”, *Communications of the ACM* 54, n.º 2, 1 de febrero de 2011, 80-88, <https://doi.org/10.1145/1897816.1897840>
- Finlayson, Mark A. y Tomaz Erjavec, “Overview of Annotation Creation: Processes and Tools”, en *Handbook of Linguistic Annotation*, editado por Nancy Ide y James Pustejovsky, 167-91, Dordrecht: Springer Netherlands, 2017, https://doi.org/10.1007/978-94-024-0881-2_5
- Franganillo, Jorge, “Gestión de información personal: elementos, actividades e integración”, *Profesional de la información* 18, n.º 4, 8 de agosto de 2009, 399-406, <https://doi.org/10.3145/epi.2009.jul.06>

- Jones, William, "Personal Information Management", *Annual Review of Information Science and Technology* 41, n.º 1, 2007, 453-504, <https://doi.org/10.1002/aris.2007.1440410117>
- Jones, William, *Transforming Technologies to Manage Our Information: The Future of Personal Information Management, Part 2*, Synthesis Lectures on Information Concepts, Retrieval, and Services, Cham: Springer International Publishing, 2014, <https://doi.org/10.1007/978-3-031-02329-3>
- Jones, William, Jesse David Dinneen, Robert Capra, Anne R. Diekema y Manuel A. Pérez-Quinones, "Personal Information Management", en *Encyclopedia of Library and Information Sciences*, 4.a ed. CRC Press, 2017.
- Kalir, Remi y Antero García, *Annotation*, Cambridge: MIT, University Press, 2021.
- Leong, Elaine, "Read. Do. Observe. Take Note!", *Centaurus* 60, n.º 1-2, 2018, 87-103, <https://doi.org/10.1111/1600-0498.12203>
- Nwagwu, Williams Ezinwa y Chidiebube Blossom Williams, "Knowledge Mapping and Visualization of Personal Information Management Literature, 1988-2020", *IFLA Journal*, 22 de febrero de 2022, <https://doi.org/10.1177/03400352211065487>
- Pitura, Joanna, "Digital Note-Taking for Writing", en *Digital Writing Technologies in Higher Education: Theory, Research, and Practice*, editado por Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt y Antonette Shibani, 101-19, Cham: Springer, 2023, https://doi.org/10.1007/978-3-031-36033-6_7
- Schmitt, Ulrich, "Shortcomings of the web of documents and data for managing personal knowledge and collaboration", en *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 23-29, 2017, <https://doi.org/10.1109/NEXTCOMP.2017.8016171>
- Tramullas, Jesús, "Desarrollos en elaboración de documentación técnica: lenguajes de marcado ligero", *Anuario ThinkE-PI* 13, marzo de 2019, <https://doi.org/10.3145/thinkepi.2019.e13f03>

Tramullas, Jesús, Ana-I Sánchez-Casabón y Piedad Garrido-Picazo, "Gestión de información personal con *software* para mapas conceptuales", *El Profesional de la Información* 18, n.º 6, noviembre de 2009, 601-12, <https://doi.org/10.3145/epi.2009.nov.03>

Hacia una sistematización de la terminología de las ediciones: una contribución desde la organización del conocimiento a la industria editorial

*Towards a systematization of the terminology of editions:
a contribution from the organization of knowledge to
the publishing industry*

Mario Barité
Carolina Saravia Rebollo
*Facultad de Información y Comunicación
Universidad de la República, Uruguay*

1. INTRODUCCIÓN

La industria editorial cuenta con una antigua y poderosa comunidad de productores, usuarios y actores, y un espacio en el que se cruzan el trabajo profesional, la técnica y el arte. La edición es entendida, en general, como el proceso de producción de ejemplares de textos, obras artísticas, documentos visuales o audiovisuales emitidos, de una vez y al mismo tiempo, bajo determinadas características comunes, en forma impresa, digital u otra modalidad alternativa. Sus diferentes aspectos son de extrema importancia para la Bibliotecología y la Ciencia de la Información, particularmente en lo relativo a la descripción, la clasificación, la indización, el desarrollo de colecciones y la valoración de fuentes de información.

A modo de ejemplo, en los registros catalográficos de una obra en las bases de datos de las bibliotecas existe un segmento descriptivo llamado 'área de edición', en el cual se aplican reglas normalizadas para establecer las características particulares que puede tener, justamente, la edición de la obra de que se trate (tercera edición, modificada, corregida y aumentada, del bicentenario, conmemorativa, etcétera).

La diferencia entre una edición y otra de la misma obra puede ser esencial incluso, para justificar o descartar su adquisición en una biblioteca. Por ejem-

plo, el libro *La estructura de las revoluciones científicas* de Thomas Kuhn fue publicado primero como monografía en la *International Encyclopedia of Unified Science* y luego como libro bajo la responsabilidad de la editorial de la Universidad de Chicago en 1962. Sin embargo, unos años después, el propio Kuhn decidió impulsar una nueva edición¹ en la que incorporó un *Postscript-1969* con el que buscó responder al conjunto de cuestionamientos que había suscitado la primera edición. Como es de suponer, en medios académicos es especialmente aconsejable contar con la segunda edición antes que la primera, y por ello la indicación de la edición en el registro de la base de datos es tan relevante.

Desde la perspectiva de la organización del conocimiento, y también desde la Terminología, la edición puede ser tomada como un dominio, esto es, un área de trabajo (“field of work”, en palabras de Schmidt and Wagner²), delimitada por las prácticas e intereses propios de una comunidad de practicantes, que termina produciendo una forma particular de lenguaje y de comunicación especializada.

Con el fin de proporcionar distintas herramientas para realizar un mapeo o una organización nocional de un dominio del saber, Hjørland³ propuso once abordajes metodológicos, a los que genéricamente denomina ‘análisis de dominio’. Estos once abordajes fueron contestados parcialmente o ampliados por Smiraglia,⁴ Guimarães and Tognoli⁵ y Barité y Rauch⁶. De este modo, la enumeración de formas de análisis de dominio en la actualidad comprende dieciséis modalidades, en la mayoría de las cuales existe consenso: guías de literatura, clasificaciones especializadas y tesauros, investigación sobre competencias en indización y recuperación de la información, estudios de usuarios en áreas temáticas, estudios bibliométricos, estudios históricos, estudios de documentos y géneros documentales, estudios epistemológicos y críticos, es-

1 Thomas S. Kuhn, *The nature of scientific revolutions* (Chicago: University of Chicago, 1970).

2 Kjeld Schmidt e Ina Wagner, “Ordering Systems: Coordinative Practices and Artifacts in Architectural Design and Planning”, *Computer Supported Cooperative Work CSCW* 13, 2004, 349-408.

3 Birger Hjørland, “Domain analysis in information science: eleven approaches-traditional as well as innovative”, *Journal of documentation* 58, n.º 4, 2002, 422-462.

4 Richard P. Smiraglia, *Domain analysis for Knowledge Organization: tools for ontology extraction*. (Oxford: Chandos Publishing, 2015).

5 José Augusto Chaves Guimarães y Natália Bolfarini Tognoli, “Provenance as a domain analysis approach in archival knowledge organization”, *Knowledge Organization* 42, n.º 8, 2015, 562-569.

6 Mario Barité y Mirtha Rauch, “Terminological studies as domain analysis: a critical exploration”, *Brazilian Journal of Information Science* 16, n.º 19, 2022.

tudios terminológicos, lenguajes para fines específicos, análisis del discurso, estudios de estructuras e instituciones en comunicación científica y profesional, estudios sobre cognición profesional, representación del conocimiento en computación e inteligencia artificial, bases de datos semánticas, y el criterio de procedencia en el campo de la Archivología.

De entre todas las posibilidades de análisis de dominio, en el caso de la presente investigación, se optó por realizar un abordaje desde los estudios terminológicos, con la finalidad de reunir la terminología, bastante dispersa en lengua castellana, relacionada con la edición.

Son dos las metodologías típicas de investigación en Terminología: estudios puntuales o abordajes sistemáticos. Si la pesquisa se enfoca en un término o en un conjunto muy pequeño y emparentado de términos se habla de estudios de terminología puntual. En este entendimiento, dos son los criterios que identifican un trabajo de terminología puntual: “el número de términos afectado” y “la motivación inicial que provoca la puesta en marcha del trabajo”.⁷ En efecto, el foco puesto en el análisis de un término o unos pocos de ellos, y la demanda que proviene de usuarios específicos de una institución o de una especialidad, configuran este tipo de estudio. La terminología puntual, por sus características, suele enfocarse en el significado y la presentación gráfica de los neologismos.

La investigación sistemática, por su parte, pretende reunir el conjunto de la terminología de un dominio y presentarla en forma de banco de datos terminológicos, o en alguno de sus productos clásicos: diccionarios, clasificaciones, glosarios y vocabularios especializados.

En este caso se prefirió el formato de diccionario, atendiendo a la dimensión cuantitativa de los términos propios del dominio elegido: más de tres centenas. Ya desde principios de este siglo, varios autores (Hodge;⁸ Soergel⁹) mencionan a los diccionarios como sistemas de organización del conocimiento, aunque todavía no son utilizados en forma regular para operaciones de indización o de representación temática. De todos modos, nada impide que más allá de las funciones típicas de consulta que pueda ofrecer el diccionario en construcción, también pueda ser utilizado para seleccionar –entre las entradas– términos candidatos a descriptores o encabezamientos de materia.

Tras una revisión bibliográfica que comprendió diversas obras de referencia y repertorios terminológicos se encontró que existe una carencia en los sistemas

7 María Teresa Cabré, *La terminología: teoría, metodología, aplicaciones* (Barcelona: Antártida; Empúries, 1993), 339.

8 G. Hodge, *Systems of knowledge organization for digital libraries: Beyond traditional authority files* (Washington: The Council on Library and Information Resources, 2000).

9 D. Soergel, *Knowledge organization systems: overview* (Alexandria: Dsoergel, 2009).

de organización del conocimiento especializado con relación a la terminología de las ediciones. No se encontraron antecedentes de diccionarios específicos en la materia, tanto en idioma castellano como en otras lenguas de alcance y conocimiento internacional. Sí se ubicaron, generalmente en formato digital, una serie de glosarios o vocabularios que tienen cierto desarrollo, pero que rara vez superan la cincuentena de términos identificados y definidos. También aparecen términos relacionados al tema, en distintos glosarios o diccionarios relacionados con el universo temático de las Artes Gráficas, así como también en manuales especializados en las disciplinas de Bibliotecología y Documentación. En general, como se anticipó, prevalecen los glosarios de pequeño o mediano porte sobre los tipos de edición en internet.

Esta constatación dio fundamento mínimo suficiente para establecer la importancia de construir un repertorio terminológico especializado en la edición en lengua castellana, con pretensión de exhaustividad, que pueda ser de utilidad para distintos colectivos profesionales, así como también en espacios de educación formal e informal.

2. OBJETIVOS

La investigación que se describe aquí, ya en fase final de su desarrollo, se propone reunir de una forma estructurada un conjunto amplio de términos relacionados entre sí bajo la perspectiva común de la edición, para su utilización por distintos sectores de actividad vinculados con la industria editorial, la organización bibliográfica y documental, el control bibliográfico y la curaduría de objetos de valor cultural e histórico.

En el marco del antedicho objetivo general se espera alcanzar al menos dos objetivos específicos: *i*) cumplir con una investigación terminológica sistemática original, que aportará una obra de referencia con un alcance exhaustivo, en un espacio de conocimiento en el cual prevalece la existencia de glosarios de alcance limitado y cierta dispersión del registro de términos y definiciones, en una gran variedad de documentos antiguos y modernos; *ii*) enriquecer la enseñanza universitaria de grado y posgrado, a través de la transmisión de conocimientos teóricos y metodológicos en la materia, aplicados en un caso de la realidad.

Se espera que el cumplimiento de estos objetivos contribuya a identificar, por añadidura, algunas claves de la evolución propia del dominio de la edición hasta el presente, teniendo en cuenta la continua transformación del proceso de producción, distribución y consumo de productos y servicios editoriales,

así como los cambios semánticos sobre términos existentes y el surgimiento de neologismos en el área.

3. METODOLOGÍA

Para el desarrollo de la investigación terminológica sistemática sobre el dominio de la edición se cumplieron las siguientes fases metodológicas:

3.1. Recuperación de un proyecto antecedente

El antecedente de esta investigación es un proyecto inconcluso, centrado en una metodología para la clasificación de los tipos de edición desde una perspectiva bibliotecológica, presentado en 2001 por el licenciado en Bibliotecología Vicente Volpe, hoy fallecido, a cuya tutoría fue llamado el profesor Mario Barité. En 2019 se decide por parte de los actuales integrantes del equipo, retomar este trabajo desde un enfoque estrictamente terminológico. Esta primera decisión impacta sobre esa idea original al modificar el abordaje disciplinario inicial, al tiempo que varía la propuesta de organización y la dimensión del producto terminológico que había imaginado Volpe. Más allá del redimensionamiento y recaracterización de su proyecto, es de justicia reconocer la existencia de ese primer trabajo, y la inspiración que ofreció para el actual desarrollo de la investigación.

La reformulación inicial del avance realizado por Volpe dejó como resultado una lista de setenta y seis (76) términos candidatos, algunos de ellos con definiciones. Desde esa base se realizó una exploración preliminar a partir de distintas fuentes documentales de tipo académico, profesional y de divulgación sobre el mercado editorial, para relevar el panorama general del área de estudio y delimitar el foco de la investigación.

3.2. Delimitación conceptual del dominio de la edición

La palabra 'edición' es un término polisémico que acoge varias acepciones en función del uso que se hace de ella en distintos contextos. Tradicionalmente, se utiliza la palabra con el sentido de proceso de producción o elaboración de un documento. En el caso de procesadores de texto, se llama edición a la funcionalidad que permite modificar el contenido de un texto o un documento. Pero también suele hablarse de edición o nueva edición de festivales, concursos, productos.

En la presente investigación, a efectos de utilizar sentidos operativos del término ‘edición’ se tomaron como referencia la segunda, la tercera, la quinta y la octava acepción del vocablo, tal como aparecen en la vigésimo tercera edición del Diccionario de la Lengua Española:¹⁰

2. Conjunto de ejemplares de una obra impresos de una sola vez, y por ext., la reimpresión de un mismo texto. *Edición del año 1732. Primera, segunda edición.* [...]
3. Colección de libros que tienen características comunes como su formato, el tipo de *edición*, etcétera. Edición de bolsillo, de lujo.
[...]
5. Cada una de las sucesivas tiradas de un periódico o de sus versiones locales, regionales o internacionales.
[...]
8. *Ecd.* Texto preparado de acuerdo con los criterios de la ecdótica y la filología. [Las negritas y las cursivas están en el original].

Las únicas excepciones a este criterio de inclusión corresponden a algunas expresiones que, si bien no corresponden estrictamente a tipos de edición, aparecen frecuentemente en diferentes contextos y cuyo conocimiento podría ser útil para los lectores (inedición, edición hipotética, edición autografiada).

De igual forma, aunque el área de estudio no es la industria editorial, se considera el proceso de edición en un sentido amplio. Por tanto, además de incluir términos relacionados con la tipología de las ediciones en función de sus atributos –formato, soporte, contenido, destinatarios y finalidad, entre otros–, también se incluyen términos que guardan relación directa con la edición, tal como ha sido delimitado su concepto en esta investigación, en el marco del conjunto de operaciones intelectuales y técnicas realizadas antes y después de la publicación –diseño, producción, distribución y comercialización.

3.3. Caracterización del producto final

Se estableció la realización de un producto terminológico, bajo la forma de un diccionario, en esta instancia monolingüe en idioma castellano, que está en su fase final de construcción. Si bien a lo largo de la investigación se fueron recogiendo datos sobre las equivalencias de numerosos términos en lengua

¹⁰ Real Academia Española, *Diccionario de la lengua española* 23.^a ed.: Edición del Tricentenario (Buenos Aires: Espasa, 2014), 838.

portuguesa, inglesa, francesa e italiana, se prefirió mantener en esta etapa la característica monolingüe en idioma castellano, debido a que no se pudo lograr una cobertura exhaustiva de las equivalencias de los términos de la edición en otros idiomas, lo que deberá requerir una investigación complementaria.

El diccionario fue construido conforme a una concepción descriptiva. Esto quiere decir que las formas de los términos se corresponden exactamente con las encontradas en la literatura que ofició de corpus. Las definiciones que se elaboraron ponen el énfasis en proposiciones razonables sobre el alcance de cada término, antes que en juicios de valor.

Eventualmente, el diccionario que se publicará podrá servir de base para una lista de descriptores, de uso en bibliotecas y sistemas de información.

Los destinatarios o usuarios del diccionario que se obtendrá como producto pueden ser profesionales que necesiten etiquetas temáticas relacionadas con las distintas variedades o manifestaciones de la edición —como bibliotecólogos, bibliotecarios, bibliófilos, coleccionistas, periodistas, gestores culturales, curadores y museólogos—, así como agentes de la industria editorial —editores, correctores de estilo y libreros, que requieran herramientas de precisión terminológica en el desarrollo de sus actividades, entre otros interesados en el estudio y análisis de la industria editorial.

3.4. Constitución del corpus textual

Se entiende por corpus en Lingüística y, por derivación, en Terminología “o acervo de textos escritos ou orais que constituem o conjunto documental da pesquisa”.¹¹ El corpus para una investigación puede integrarse con cualquier tipo de obras (diccionarios, obras literarias, colecciones de revistas científicas, textos de prensa digital, mapas, etcétera), dependiendo de los objetivos que se persigan. Por otra parte, en algunos casos suele requerirse que el corpus se integre con obras de un solo género, para dar homogeneidad al análisis y a los resultados; por ejemplo, los estudios de diccionarización se concentran particularmente en *corpora* integrada por diccionarios, glosarios, vocabularios y similares.¹² En cualquier caso, lo relevante es que el corpus constituya un

11 Anna Maria Becker Maciel, “Terminologia e corpus” en *Corpora na terminologia*, orgs. Stella Tagnin y Cleci Bevilacqua (São Paulo: Hub, 2013), 30.

12 Mario Barité, Mirtha Rauch y Varenka Parentelli, “Knowledge organization systems in reference works specialized in knowledge organization: study of punctual terminology” en *Knowledge Organization for Resilience in Times of Crisis: Challenges and Opportunities. Proceedings of the Eighteenth International ISKO Conference*, 39-50, Wuhan, China, 20-22, march, 2024.

número suficiente y representativo de obras que permitan la mejor realización de la investigación.¹³

Por las características de la pesquisa en curso, se buscó integrar el corpus con todos los diccionarios y glosarios especializados (en versión impresa o digital), en los que pudiera recabarse información sobre tipos de ediciones, sin limitación de origen, idioma o fecha de publicación. Este énfasis en la búsqueda preferente de fuentes formales partió del supuesto de que, para que un término sea incluido y definido en una obra de referencia, los respectivos autores tomaron decisiones razonables en lo relativo a la existencia y estabilidad de los términos, su ortografía y su alcance conceptual.

Cuando se entendió que las obras de referencia del corpus no aseguraban la información suficiente sobre un término, se apeló a fuentes heterogéneas, mayoritariamente disponibles en internet. Para este corpus complementario se realizó en cada caso un análisis de fiabilidad de las fuentes, a partir de la experiencia adquirida en la frecuentación de obras sobre edición e industria editorial. El acceso a fuentes complementarias se llevó a cabo a partir de búsquedas específicas sobre términos insuficientemente informados en los motores de búsqueda Google y Google Scholar.

Por su extensión, no es posible transcribir en este trabajo el conjunto de las referencias del corpus. Baste decir que en total, las fuentes utilizadas, superan las 80 obras.

3.5. Diseño de ficha terminológica

Para el registro de los datos terminológicos se elaboró una ficha con los siguientes campos de información:

- Término en castellano (autorizado con definición en mayúscula, no autorizado sin definición en minúscula). Los términos no autorizados son sinónimos no elegidos para ser entrada de la información.
- Equivalencias en otras lenguas (por lo dicho más arriba, no trascienden al diccionario en esta instancia).
- Definiciones extraídas del corpus.
- Indicación de fuente de cada definición.

13 H. Picht, "Korpora als Ausgangspunkt für die Extraktion von terminologischen daten", *Synaps*, n.º 8, 2001, 38-48.

- Definición original (para este fin se siguió el modelo aristotélico de elaboración de definiciones: identificación del género y de las diferencias específicas).
- Atribución de relación de véase (control de sinónimos, cuasi sinónimos y variantes).
- Atribución de relación de véase además (control de términos relacionados).
- Notas o comentarios.
- Numeración de cada término.

3.6. Selección y extracción de términos

Fueron seleccionados y extraídos del corpus los términos cuyas definiciones permitieron encartarlos en alguna de las cuatro acepciones operativas del término 'edición', que fueron tomadas del diccionario de la Real Academia Española.

La validación se realizó mediante la denominada garantía literaria (*literary warrant*), entendida como el proceso de justificación terminológica basado en la aparición de los términos en un corpus textual especializado (Beghtol;¹⁴ Bullard¹⁵). De este modo, todos los términos relacionados con la edición que aparecen en el diccionario cuentan con el registro previo de su existencia en el corpus. Se estableció que, para su inclusión, los términos seleccionados debían comparecer en dos o más fuentes de autores diferentes. El proceso de validación se llevó a cabo a través del cumplimiento de las siguientes premisas:

1. Se comprueba que el término existe.
2. Se reconoce que el término queda comprendido, de un modo razonable, dentro del dominio de la edición, tal cual se ha delimitado.
3. Se encuentran formas consistentes desde el punto de vista sintáctico y ortográfico.
4. Se ubican equivalencias más o menos literales del término en otros idiomas.
5. Se comprueba que el término cuenta con una o más definiciones.
6. Se verifica que las definiciones tienen similitudes en sus alcances.
7. Se encuentra que, en algunos casos, existen sinónimos.
8. Se identifican relaciones no sinonímicas entre términos.

14 C. Beghtol, "Semantic validity: concepts of warrant in bibliographic classification systems", *Library resources & technical services* 30, n.º 2 (1986), 109-125.

15 J. Bullard, "Warrant as a means to study classification system design", *Journal of Documentation* 73, n.º 1, 2017, 75-90.

9. Se encuentran otros datos del término, relacionados –por ejemplo– con información histórica de su acuñación, antigüedad y contexto de uso.

3.7. Elaboración de definiciones originales

Todas las definiciones establecidas en el diccionario son originales. Para la elaboración de estas definiciones se siguió el método aristotélico de género próximo y diferencia específica, que es el más comúnmente utilizado por los responsables de diccionarios y glosarios. Este método establece que la estructura sugerida de una definición es la que comienza con la referencia al género del *definiens*, que es seguida por una secuencia descriptiva de atributos, cada uno más específico que el anterior. De este modo cada atributo implica señalar una diferencia específica respecto a otros objetos o *definiendum* similares.

Por otra parte, para la elaboración de las definiciones originales, se utilizó el método de la grilla que categoriza los atributos mencionados explícitamente en definiciones preexistentes, para luego compararlos, y a partir de los resultados redactar una nueva definición, diferente pero conectada con las anteriores.¹⁶ Esto quiere decir que, en todos los casos, las definiciones originales que fueron redactadas toman como punto de partida las definiciones preexistentes que fueron encontradas en el corpus.

3.8. Control de vocabulario

Se desarrollaron un conjunto de operaciones de control de vocabulario, destinadas a proporcionar parámetros normalizados de organización del conocimiento que cumplen la función de restricción de sentido, desambiguación, normalización, jerarquización y relación.¹⁷

En ese marco, se establecen referencias cruzadas que remiten un término a otro u otros y vinculan conceptos entre sí con los que mantienen distintas relaciones, por ejemplo, de identidad, de sinonimia o funcionales:

- Referencia de *véase*: Se remiten los sinónimos no autorizados a los términos preferidos, como ya fue dicho más arriba.

16 Mario Barité y Mirtha Rauch, “Propuesta metodológica para la elaboración de definiciones terminológicas” [ponencia]. *Congreso de RITERM*. Montevideo, Uruguay, 2006.

17 Mario Barité et al., *Diccionario de organización del conocimiento: Clasificación, Indización, Terminología*, 6.^a ed. corregida y aumentada (Montevideo: CSIC, 2015).

- Referencia de *véase además*: Se remiten términos usados o preferidos a otros términos usados o preferidos, los cuales guardan entre sí relaciones de afinidad o proximidad semántica.

Toda vez que se identifican dos o más sinónimos, se elige uno de ellos como entrada principal. Bajo este sinónimo preferido se incluye la definición y otra información asociada. Los otros sinónimos se registran igualmente, pero solo conservan valor de conocimiento y referencia. Como criterio de consistencia, se elige como descriptor o sinónimo preferido aquel con mayor cantidad de resultados en Google Scholar, o subsidiariamente en Google, al momento de elaboración de la lista.

3.9. Elaboración de un árbol de dominio

Un árbol de dominio es una variedad de taxonomía, utilizada en los últimos años en el ámbito de la Terminología, que distribuye el dominio en sus principales ramas y que, en su modalidad más desarrollada, sitúa todos y cada uno de los términos del dominio bajo esas ramas. Se trata, por tanto, de una herramienta de control terminológico que ayuda a equilibrar las subseries de términos, identificar lagunas, y reunir las expresiones que son afines desde el punto de vista de su semántica. Como en toda taxonomía, los términos o taxones se presentan priorizando las relaciones jerárquicas que existen entre ellos.

En esta investigación, una vez culminada la recopilación, validación, sistematización y definición de los términos, estos fueron agrupados conforme a atributos generados en forma inductiva, siempre dentro del enfoque amplio del dominio que fue establecido por las cuatro acepciones del Diccionario de la Real Academia. En función de ese análisis inductivo se identificaron nueve atributos organizadores de la terminología, de modo tal que cada término fue asociado a alguno de esos nueve atributos o ramas del dominio de la edición: formato, soporte, presentación gráfica, origen, contenido, destinatarios, tecnología utilizada, finalidad de la edición, y situaciones de la realidad. El árbol de dominio se presentará en los preliminares del diccionario.

3.10. Revisión del texto prefinal

Una vez culminada la investigación y redactado el diccionario en su totalidad, el texto se ofrecerá para la opinión de al menos un (o una) especialista uruguayo/a del ámbito editorial.

3.11. Publicación del diccionario

Luego de realizar los ajustes que correspondan conforme a la opinión experta, el diccionario será presentado para su publicación ante entes financiadores de la investigación en Uruguay. Se espera alcanzar esta última etapa antes de que finalice el año 2024.

4. RESULTADOS

Todas las obras que integran el corpus fueron productivas, ya que cada una de ellas aportó información pertinente a la investigación. En su conjunto, permitieron entender adecuadamente los alcances conceptuales de cada término, así como resolver las dudas que se fueron generando.

Siempre con el norte en la recopilación exhaustiva, se registró un total de 302 términos pertenecientes al dominio de la edición conforme a los criterios establecidos. De ellos, 213 tienen una entrada directa en el diccionario (es decir, cuentan con una definición), y 89 son utilizados solo con valor de referencia y no cuentan con definición, pues se trata de sinónimos no preferentes. En términos porcentuales, esto se expresa así:



1. Términos registrados (Fuente: elaboración propia).

Las referencias de *véase* consignadas desde los sinónimos no preferentes a los preferidos ascienden entonces a 89. Por su parte, las referencias de *véase además*, que se establecen entre términos que tienen algún tipo de conexión semántica, suman un total de 174 instancias.

En muchos casos fue difícil establecer si dos términos eran sinónimos o no, en virtud de su extrema proximidad semántica, y de las explicaciones ambi-

guas –o incluso contradictorias– de las fuentes. A riesgo de cometer algún error de apreciación, esta dificultad se resolvió a partir de un análisis pormenorizado de la información contenida en las fuentes respectivas del corpus, un intercambio entre los autores del diccionario y, eventualmente, una consulta a especialistas. La dilucidación de estos casos no es menor, en la medida en que determina el establecimiento de una relación de *véase*, o de *véase además*, según el caso.

El árbol de dominio se constituyó con nueve ramas, en las cuales se distribuyeron los 302 términos registrados. A continuación, se mencionan las nueve ramas y se agregan algunos ejemplos de términos comprendidos en cada una de ellas:

- Formato; ejemplos: “edición de bolsillo”, “edición en octavo”, “edición en miniatura”.
- Soporte; ejemplos: “edición en papel”, “edición de hojas intercambiables”, “edición digital”.
- Presentación; ejemplos: “edición de lujo”, “edición ilustrada”, “edición numerada”.
- Origen; ejemplos: “edición aldina”, “edición áurea”, “edición elzeviriana”.
- Contenido; ejemplos: “edición abreviada”, “edición adaptada”, “edición vulgata”.
- Destinatarios; ejemplos: “edición de bibliófilo”, “edición juvenil”, “edición universitaria”.
- Tecnología utilizada; ejemplos: “autoedición”, “edición fresada”, “edición mimeografiada”.
- Finalidad; ejemplos: “edición conmemorativa”, “edición para el teatro”, “festschrift”.
- Situaciones de la realidad; ejemplos: “edición confiscada”, “edición príncipe”, “inedición”.

La utilización de estos nueve atributos es la que vuelve operativa la idea planteada en el objetivo general de realizar este repertorio bajo la perspectiva común de la edición.

Por otra parte, pudo cumplirse con la elaboración de definiciones para cada uno de los doscientos trece (213) términos utilizados como entradas. En algunos casos, se identificaron dos o más acepciones.

En cada definición se estableció el género de procedencia, seguido de los diferentes atributos esenciales que permiten dar identidad a cada concepto. Por las particularidades del dominio, el género suele estar representado por una expresión que representa al sustantivo ‘edición’, tal como ‘la’ o ‘aquella’.

A continuación, se organizan los atributos adjetivos, siguiendo una secuencia más o menos lógica. En una segunda frase se agregan habitualmente enunciados informativos. Así, por ejemplo, la definición de ‘edición consolidada’ dice: “La que reúne en un mismo volumen distintas obras o versiones de obras sobre determinado tópico, publicadas previamente, con el fin de crear una nueva estructura de contenido. Esto permite incorporar modificaciones y actualizaciones de la materia. Suele estar relacionada con normas, reglamentos y manuales técnicos”.¹⁸

En algunas definiciones se incorporaron también algunos enunciados accidentales e históricos, con distinto énfasis según las circunstancias. Por ejemplo, la definición de ‘edición aldina’ es “Cualquiera de las realizadas por la imprenta del veneciano Aldo Manucio también conocido como Aldo Manuzio, Aldus Pius Manutius, Teobaldo Mannucci o Aldo el Viejo (Bassiano 1449-Venecia 1515), o sus sucesores. A él se debe sobre todo la edición sistemática de los clásicos griegos, la creación del formato en octavos y de la tipografía aldina, conocida hoy como cursiva o itálica”.¹⁹ En este caso, se prefirió enfatizar sobre la información histórica, que es la que determina la existencia misma del término y su denominación.

El conjunto de los resultados y su consistencia confirman que el diseño metodológico elegido fue el más apropiado para el proceso de investigación y el cumplimiento de los objetivos establecidos.

5. CONCLUSIONES

La investigación emprendida se encuentra ya muy avanzada y, como se anticipó, se aspira a publicar el diccionario sobre edición en el correr del año 2024, previo pasaje por un régimen de referato.

Más que en una compilación valiosa por la contemporaneidad de sus términos, la pretensión de la presente investigación se ha enfocado en su reunión exhaustiva, sin límites temporales. Por ello, el diccionario resultante está en condiciones de ofrecer un panorama diacrónico de la terminología de las ediciones, que alcanza incluso a expresiones utilizadas desde hace varios siglos en adelante.

Se cree que el avance logrado demuestra ya la tendencia a generar nueva terminología en esta área, cada vez que se ha producido un cambio tecnológico importante. El mundo de los bibliófilos, por su parte, ha introducido una

18 Mario Barité y Carolina Saravia, *Diccionario de las ediciones* (Montevideo: inédito, 2024).

19 *Ibidem*.

serie de términos vinculados con los libros antiguos, que siguen utilizándose con cierta regularidad en la documentación.

En la medida en que cada nueva edición representa un cambio con relación a su antecedente, también se encuentran términos que refieren a las modificaciones realizadas entre una edición y otra.

Estas y otras situaciones que se han dado a lo largo del tiempo, son las que quedan a la vista apenas se visualizan los resultados de la investigación en forma global.

Finalmente, cabe destacar que este estudio ha constituido una desafiante experiencia de trabajo, en especial por la ausencia de antecedentes de esta ambición y envergadura. Se espera, por tanto, que tenga una utilidad recurrente como obra de referencia, considerando al conjunto de actores que se mueven en torno al mundo de la industria editorial.

NOTA:

Se deja constancia de que este trabajo se realiza en el marco del apoyo financiero que recibe el Grupo de Investigación en Terminología y Organización del conocimiento (GTERM), radicado en la Facultad de Información y Comunicación de la Universidad de la República del Uruguay, por el cuatrienio (2023-2027), para el desarrollo de la línea de investigación “Sistematización crítica de la terminología en áreas especializadas”. El apoyo financiero para el citado cuatrienio es aportado por la Comisión Sectorial de Investigación Científica (CSIC) de la Universidad de la República del Uruguay.

REFERENCIAS BIBLIOGRÁFICAS

- Barité, M. y C. Saravia, *Diccionario de las ediciones*, Montevideo: inédito, 2024.
- Barité, M. et al., *Diccionario de organización del conocimiento: Clasificación, Indización, Terminología*, 6.ª ed. corregida y aumentada, Montevideo: CSIC, 2015.
- Barité, Mario, Mirtha Rauch y Varenka Parentelli, "Knowledge organization systems in reference works specialized in knowledge organization: study of punctual terminology", en *Knowledge Organization for Resilience in Times of Crisis: Challenges and Opportunities*, Proceedings of the Eighteenth International ISKO Conference, Wuhan, China, 20-22, march, 2024, 39-50.
- Barité, M. y M. Rauch, "Terminological studies as domain analysis: a critical exploration", *Brazilian Journal of Information Science* 16, n.º 19, 2022.
- Barité, Mario y Mirtha Rauch, "Propuesta metodológica para la elaboración de definiciones terminológicas" [ponencia], *Congreso de RiTERM*, Montevideo, Uruguay, 2006.
- Beghtol, C., "Semantic validity: concepts of warrant in bibliographic classification systems", *Library resources & technical services* 30, n.º 2 (1986), 109-125.
- Bullard, J., "Warrant as a means to study classification system design", *Journal of Documentation* 73, n.º 1, 2017, 75-90.
- Cabré, María Teresa, "Contexto y evolución de la terminología: de una aproximación nominalista a una teoría comunicativa", en *Teoría y Praxis en Terminología*, coord. S. Álvarez Catalá, M. Barité, 9-22, Montevideo: UCUR, 2017.
- Cabré, María Teresa, *La terminología: teoría, metodología, aplicaciones*, Barcelona: Antártida; Empúries, 1993.
- Guimarães, José Augusto Chaves y Natália Bolfarini Togno-
li. "Provenance as a domain analysis approach in archival knowledge organization", *Knowledge Organization* 42, n.º 8, 2015, 562-569.
- Hjørland, Birger, "Domain analysis in information science: eleven approaches - traditional as well as innovative", *Journal of documentation* 58, n.º 4, 2002, 422-462.

- Hodge, G., “Systems of knowledge organization for digital libraries: Beyond traditional authority files”, The Council on Library and Information Resources, Washington DC, 2000, <http://www.clir.org/pubs/reports/pub91/contents.html/>
- Kuhn, Thomas S., *The nature of scientific revolutions*, Chicago: University of Chicago, 1970.
- Maciel, Anna Maria Becker, “Terminologia e corpus”, en *Corpora na terminologia*, orgs. Stella Tagnin y Cleci Bevilacqua, 29-45, SãoPaulo: Hub, 2013.
- Picht, H., “Korpora als Ausgangspunkt für die Extraktion von terminologischen daten”, *Synaps*, n.º 8, 2001, 38-48.
- Real Academia Española, *Diccionario de la lengua española*, 23.^a ed.: Edición del Tricentenario, Buenos Aires: Espasa, 2014.
- Schmidt, Kjeld e Ina Wagner, “Ordering Systems: Coordinative Practices and Artifacts in Architectural Design and Planning”, *Computer Supported Cooperative Work CSCW* 13, 2004, 349-408.
- Smiraglia, Richard P., *Domain analysis for Knowledge Organization: Tools for ontology extraction*, Oxford: Chandos Publishing, 2015.
- Soergel, D., *Knowledge organization systems: overview*, Dsoergel, Alexandria, 2009.

Consolidando un ecosistema de documentación
del patrimonio histórico y cultural de la UNAM:
propuesta metodológica *ad hoc* para el desarrollo
de un microtesauro

*Consolidating an ecosystem of documentation of the historical
and cultural heritage of the UNAM: ad hoc methodological
proposal for the development of a microthesaurus*

Claudio Molina Salinas
Instituto de Investigaciones Estéticas
Universidad Nacional Autónoma de México

INTRODUCCIÓN

En los últimos diez años, desde una perspectiva GLAM (*Galleries, Libraries, Archives and Museums*),¹ académicos y personal de algunos museos, facultades e institutos de la Universidad Nacional Autónoma de México (UNAM) emprendimos un trabajo enfocado en consolidar la documentación del patrimonio histórico y cultural que custodiamos en la Universidad, hacer trabajo de investigación en torno a este, y tomar acciones como utilizar un sistema de gestión de colecciones de licencia abierta (CollectiveAccess), que permite la gestión de estas y su publicación en línea;² implementar buenas prácticas de catalogación, adoptando procedimientos y estándares internacionales como VRA Core 4;³ implementar reglas de catalogación de bienes de interés cultural

1 Marina Salse *et al.*, “GLAM Metadata in Museums and University Collections: A State-of-the-art (Spain and other European Countries)”, *Global Knowledge, Memory and Communication* 73, 4/5, 2024, 478.

2 “CollectiveAccess”, *collectiveaccess.org*, 2022, <https://www.collectiveaccess.org/>.

3 Visual Resources Association, “VRA Core 4.0 Element Description”, Visual Resources, 2007, https://www.loc.gov/standards/vracore/vra_Core4_Element_Description.pdf

(BIC)⁴ como Cataloguing Cultural Objects (CCO),⁵ entre otros. Estas son las dependencias universitarias representadas en esta comunidad:

- Antigua Academia de San Carlos (ASC), Facultad de Artes y Diseño (FAD).
- Archivo Fotográfico Manuel Toussaint (AFMT), Instituto de Investigaciones Estéticas (IIE).
- Centro de Documentación Arkheia (CDA).
- Centro de Interpretación Xaltillo (CIX), Centro Cultural Universitario Tlatelolco (CCUT).
- Dirección General de Repositorios, UNAM (DGRU-UNAM).
- Museo Universitario de Arte Contemporáneo (MUAC).
- Museo Universitario del Chopo (MUCH).
- y la Unidad de Información para las Artes (UNIARTE), IIE.

Luego de varios años de arduo trabajo y de cosechar notables avances individuales en la documentación de nuestras colecciones, debido a que encontramos una amplia dispersión en el uso de los términos descriptores de estas, a principios del año 2023 nos propusimos desarrollar un microtesauro de uso común que sea, a la vez, un control terminológico *ad hoc*, un sistema local de organización del conocimiento y un instrumento de desambiguación conceptual para los catalogadores de nuestras dependencias.

En este capítulo se presenta el proceso de creación y publicación de un microtesauro que recoge términos relativos al tipo de objeto, únicamente; también se ahonda en algunas decisiones tomadas al momento de: primero, integrar la información de todas nuestras bases de datos (BD); segundo, normalizar las terminologías que se usan en los distintos museos y centros de investigación, poniendo particular atención a los que denominan tipos de objetos; tercero, definir los términos preferidos y los alternativos; cuarto, explicar qué modelo de definición se ha adoptado y la estrategia de redacción

4 En España se usa el término legalmente para denominar a los bienes muebles o inmuebles del patrimonio cultural de la nación, y con la publicación en línea de “Mexicana: el repositorio cultural de México”, se ha incorporado a nuestro contexto este concepto para describir al conjunto de bienes de la nación, pese a no tener implicación legal alguna. Ministerio de Educación, Cultura y Deporte, “Definición de Bien de Interés Cultural”, Página del Ministerio de Educación, Cultura y Deporte, *s/f*, <https://www.cultura.gob.es/cultura/patrimonio/bienes-culturales-prottegidos/definicion.html>; José Manuel Morales-del-Castillo, Pedro Ángeles Jiménez, y Claudio Molina Salinas, “Mexico’s Tradition and Culture Entering the Digital Age: The Mexican Cultural Heritage Repository Project”, *Heritage 2*, 2019, 356-65.

5 Murtha Baca *et al.*, eds., *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images* (Chicago: American Library Association, 2006).

de estas; quinto, revisar las definiciones y delimitar las relaciones jerárquicas entre los términos; y, por último, se argumentan las razones por las que hemos optado por utilizar un *software* y solución tecnológica como TemaTres para la publicación del microtesauro.

1. UN ECOSISTEMA DE CATALOGACIÓN DE BIC EN LA UNAM

En el año 2012, Renato González Mello (director del IIE), en coordinación con funcionarios de dependencias de órganos desconcentrados del Gobierno Federal (Centro Nacional de Conservación y Registro del Patrimonio Artístico Mueble del Instituto Nacional de Bellas Artes y Literatura; Dirección General de Sitios y Monumentos del Patrimonio Cultural, y Registro Público de Monumentos y Zonas Arqueológicas e Históricas, ambos del Instituto Nacional de Antropología e Historia), junto con autoridades de dependencias universitarias (Instituto de Investigaciones Bibliotecológicas y de la Información; y Unidad de Informática para la Biodiversidad (UNIBIO) del Instituto de Biología (IB), actualmente la DGRU-UNAM), organizaron un encuentro intensivo para conocer y discutir sobre los métodos usados en México para la estandarización y organización de registros de BIC.⁶

En seguimiento a los trabajos de esta reunión, en el IIE se fundaría la UNIARTE, institucionalizada hasta el año 2016, que es, a su vez, un punto de encuentro, un seminario permanente y una unidad de información que tiene como finalidad integrar todos los datos organizados de nuestro Instituto, así como impulsar las mejores prácticas de catalogación y publicación de datos científicos sobre el arte y llevarlos, en su momento, hasta un formato de datos abiertos vinculados.^{7 y 8}

En la UNIARTE, además de plantearnos discusiones teóricas y revisión de diferentes estándares y metodologías, adoptamos un modelo de catalogación del patrimonio utilizando el *software* de licencia abierta CollectiveAccess (creado por Whirl-i-Gig), que sirve para gestionar y difundir colecciones de museos, archivos y bibliotecas.⁹ Usando este *software*, en 2015, en la UNIARTE acom-

6 Comunicación personal con Renato González Mello, exdirector del IIE-UNAM, abril de 2024.

7 Comunicación personal con Pedro Ángeles Jiménez y Gabriela Betsabé Miramontes Vidal, uniarte, abril de 2024.

8 Unidad de Información para las Artes (UNIARTE), “Unidad de Información para las Artes, Instituto de Investigaciones Estéticas”, Página web del Instituto de Investigaciones Estéticas, 2020, <http://www.esteticas.unam.mx/uniarte>.

9 Rubén Alcaraz Martínez, “CollectiveAccess, un sistema de gestión y difusión de colecciones de museos, archivos y bibliotecas”, *BiD: textos universitaris de biblioteconomia i documentació*, 33, 2014.

pañamos la migración de la antigua BD del AFMT a una instalación de CollectiveAccess; a finales del 2016 comenzaríamos a administrar tanto una BD del patrimonio de México con más de 20 000 registros y, debido a los sismos del 2017, se crearía otra BD de BIC afectados por desastres naturales, que contiene más de 2 000 registros.¹⁰

En los últimos ocho años otros miembros de nuestro seminario también adoptarían CollectiveAccess y publicarían exitosamente sus colecciones; por ejemplo: en 2017, la ASC publicaría el catálogo digital de sus colecciones;¹¹ desde 2021, las colecciones del MUAC se han ido documentando en una instalación de CollectiveAccess y están próximas a publicarse en 2024; y, en mayo de 2024, se integró a las actividades de nuestro seminario el MUCH, que publicó, en 2021, *Desobediente: Archivo digital del Museo Universitario del Chopo*.¹²

Por otra parte, el CIX no usa *CollectiveAccess*, sino Tainacan,¹³ como otra alternativa de *software* de código abierto en el que se pueden administrar, gestionar y publicar colecciones digitales;¹⁴ asimismo, en el CDA comenzarán próximamente con el proceso de publicación de sus colecciones (aún no definen el *software* gestor para las mismas).¹⁵ Estos dos casos sirven para aclarar que CollectiveAccess no es “la solución definitiva”, ya que reconocemos más importante para nuestra comunidad el mantener la implementación de buenas prácticas de documentación, que el uso de un *software*.

Claramente, aún hay mucho trabajo por hacer, no obstante, desde enero de 2023 nos hemos propuesto resolver una necesidad de forma colectiva, es decir: contar con un microtesauro para normalizar el uso de la terminología con la que describimos nuestras colecciones, en el que, en principio, incluiremos únicamente a tipos de objetos (caso práctico descrito en este capítulo) y, en etapas posteriores, continuaremos con materiales, técnicas, etcétera. A continuación se explicarán los pasos de la metodología desarrollada *ad hoc* para el microtesauro y se ilustrarán algunas decisiones tomadas para su construcción.

10 Comunicación personal con Pedro Ángeles Jiménez, coordinador de uniarte, abril de 2024.

11 <https://coleccionessancarlos.fad.unam.mx/>

12 <https://archivodesobediente.chopo.unam.mx/>

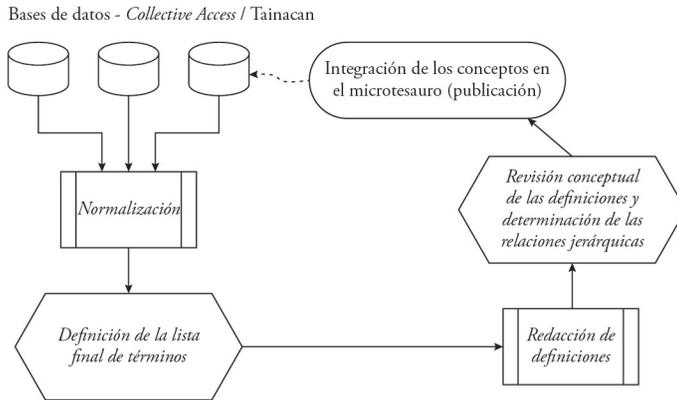
13 Tainacan es un “open source, powerful and flexible repository platform for WordPress. Manage and publish your digital collections as easily as you publish a post on your blog while having all professional repository platform tools available”. “Tainacan Wiki”, The Tainacan Project Web’s page, 2020, <https://tainacan.github.io/tainacan-wiki/#/?id=tainacan-wiki>

14 Comunicación personal con Lucía Sánchez de Bustamante, coordinadora del CIX, abril de 2024.

15 Comunicación personal con Pedro Ángeles Jiménez, coordinador de UNIARTE, abril de 2024.

2. FLUJO DE TRABAJO PARA LA CONSTRUCCIÓN DEL MICROTESAURO

El método seguido para la construcción del microtesauro consta de cinco pasos diferenciados, consecutivos, que llevan a una o varias unidades léxicas a través de ciertos procesos de cotejo y documentación hasta su publicación en línea. Los pasos propuestos son: (1) integración de las BD y normalización de los términos, (2) definición de la lista final de términos, (3) redacción de definiciones, (4) revisión conceptual de las definiciones y determinación de las relaciones jerárquicas, y (5) integración de los conceptos en el microtesauro (publicación en TemaTres). Lo anterior se puede ver gráficamente en el siguiente esquema:



1. Propuesta metodológica para la creación de un microtesauro (diagrama de flujo).

Adoptar este flujo de trabajo acorta notablemente los tiempos de revisión, dictaminación y publicación del microtesauro. A continuación, explicaré cada uno de estos momentos del diagrama anterior y señalaré algunas decisiones tomadas para un par de casos excepcionales.

2.1. Integración de las BD y normalización de los términos

Para el primer paso de la construcción del microtesauro hay que concentrar los términos usados en las BD o fuentes de información disponibles para el proyecto y también normalizarlos. De tal manera que integramos los términos o descriptores sancionados en el elemento de información “tipo de objeto” de todas las BD en un solo archivo csv, anotando debidamente la fuente de

procedencia de cada uno de los registros. Este proceso dio forma a un listado terminológico con 718 691 ocurrencias (*tokens*) y 991 tipos (*types*).¹⁶ Como se verá más adelante, hacer esta distinción entre ocurrencias y tipos, y registrar la fuente de procedencia de los datos fue relevante para la toma de decisiones y metodologías a lo largo de todo el flujo de trabajo.

En el listado de 718 691 ocurrencias nos encontramos casos ilegibles debido a problemas con la codificación, así como algunos errores de captura (con espacios al inicio o al final de la unidad léxica o formas con espacios consecutivos), términos con “errores de dedo” o escritos en su totalidad en mayúsculas. Para resolver esto, el terminólogo del seminario trató estos datos con OpenRefine,¹⁷ *freeware* usada para limpiar y transformar BD desordenadas. En la imagen siguiente, a la izquierda, vemos los datos antes de ser procesados en OpenRefine y, a la derecha, el resultado de las transformaciones hechas con la herramienta

657087	VISTA		
657088	VISTA		
657089	VISTA		
657090	VISTA		
657091	VISTAS		
657092	VISTAS		
657093	vitral		
657094	Zona arqueológica		
657095	Zona arqueológica		
657096	zona arqueológica		
657097	Zona arqueológica		
657098	Zona arqueológica		
657099	Zona arqueológica		
657100	Zona arqueológica		
657101	Zona arqueológica		
657102	Zona arqueológica		
657103	Zona arqueológica		
657104	Zona arqueológica		
657105	Zona arqueológica		

O352			
	A	B	C
334	universidad		
335	universidad estatal		
336	urna		
337	utensilio		
338	vasija		
339	vasija (contenedores por forma)		
340	vasija para beber		
341	vaso		
342	vasp		
343	via cotidiana		
344	vida cotidiana		
345	vida cotidiana/tipos populares		
346	vista		
347	vistas		
348	vitral		
349	zona arqueológica		

2. Listado de términos del AFMT antes y después del refinamiento de datos hecho con OpenRefine.

Como se puede observar en el resultado del refinamiento mostrado en la imagen anterior, aún persisten errores ortográficos o “de dedo” como “via cotidiana” o “vasp”. En este caso, se corrigieron manualmente. También se eliminaron de la lista voces como “universidad estatal”, “vida cotidiana” y “vida cotidiana/tipos populares”, ya que no son unidades morfológicamente fijas deberían estar en otro elemento de información, y son un error de captura por parte del documentalista, respectivamente. Asimismo, descartamos los cuali-

16 “Ocurrencia” es cada una de las unidades léxicas que aparecen en el listado de términos que concentramos y “tipo” es la unidad léxica registrada eliminando sus repeticiones. Luis Fernando Lara Ramos, Curso de lexicología (México, D.F.: El Colegio de México, 2006), 155-56.

17 “OpenRefine”, The Open Refine Web’s Page, 2012, <https://openrefine.org/>.

ficadores, como en el caso: “vasija (contenedores por forma)”, conservando únicamente “vasija”, ya que la gran mayoría de las terminologías de nuestras BD no utilizan este recurso de desambiguación y porque nosotros optamos por enumerar las terminologías para resolver la ambigüedad.

Finalmente, terminamos la normalización de los términos ajustándonos a los criterios usados para definir vocablos, según la usanza de la lexicografía española, llevando plurales a singulares y femeninos a masculinos cuando fue el caso.¹⁸ Además, revisamos los términos cotejando las reglas ortográficas del español y contra algunos diccionarios especializados.¹⁹ Todos estos procesos derivaron en una lista de 991 tipos de objetos con los que trabajamos en la etapa siguiente de esta metodología.

2.2. Definición de la lista final (o cercana) de términos

Para definir una lista de términos validados por especialistas es necesario reflexionar sobre la autenticidad de estos en un seminario y, simultáneamente, hacer un análisis cuantitativo que defina la frecuencia de los términos para establecer un orden de definición de estos, así como establecer los términos preferidos (descriptores) y los alternativos (no descriptores).

Para ello, previo a algunas de las reuniones del seminario permanente, el terminólogo del grupo convirtió la lista de 991 candidatos a términos en un archivo .DOCX, se puso en contacto con los colegas de los museos y dependencias universitarias involucradas en el seminario y les hizo llegar solo los términos que utilizan para catalogar sus colecciones. Vía correo electrónico y en el seminario se socializaron las dudas que resultaron de estas revisiones, se aclararon algunos casos que no fueron del todo claros y, por último, se generó una lista final de 715 términos aceptados de forma colegiada.

Por otro lado, podría pensarse que para un proyecto de microtesauro como este, precisar el orden de redacción de las definiciones de los términos es irrelevante, ya que hay que definirlos todos; sin embargo, debido a la naturaleza del trabajo del seminario, la forma inmediata en la que se publican las colecciones en CollectiveAccess y Tainacan, el modo de publicación del microtesauro en el *software* TemaTres y la necesidad de contar con el microtesauro, hemos optado

18 Luis Fernando Lara Ramos, *Teoría del diccionario monolingüe* (México, D.F.: El Colegio de México, 1997), 118-119, 170.

19 Algunos recursos consultados son los Tesauros-Diccionarios del patrimonio cultural de España, Art & Architecture Thesaurus, Tesoro de la UNESCO, Tesoro de Patrimonio Histórico Andaluz, Tesoro de bienes muebles e inmuebles por destino de recintos religiosos, entre otros.

por ordenar nuestros términos basándonos en su frecuencia, esto es: primero el término más frecuente, luego el segundo más frecuente y así sucesivamente.

Considerando la forma en la que documentamos las terminologías del proyecto, la naturaleza de nuestras fuentes de información (catálogos/bases de datos relacionales) y el progreso en la documentación de nuestras colecciones no se puede aplicar ninguna de las técnicas tradicionales de estimación de frecuencias del léxico, como: la frecuencia relativa,²⁰ la frecuencia relativa por millón de palabras,²¹ el índice normalizado de dispersión,²² la frecuencia corregida, propuesta original de Lanke, según reporta Ham,²³ etcétera, ya que no es posible conocer el número total de ítems en las colecciones hasta que estas estén completamente catalogadas y, por ende, se desconoce el número total de términos que las describen.

Una alternativa plausible y aplicable a estos casos es estimar la frecuencia de documento, método que permite cuantificar el uso de una unidad léxica, forma o giro lingüístico a partir de su dispersión o recurrencia en fuentes de información, es decir, la frecuencia de documento de una unidad léxica es la frecuencia absoluta de la aparición de esta en diferentes fuentes de información.²⁴ Esto encuentra fundamento en lo explicado por Sinclair, de que solo los hechos léxicos más prominentes, que se repiten en una variedad de textos y autores de forma independiente deben considerarse para estar en una obra de referencia, principio que él llama *recurrence*.²⁵

Para aclarar el alcance de todo esto, pensemos en este ejemplo: el AFMT aporta al total de la muestra analizada 691 505 ocurrencias y 319 tipos; de estos, los cinco términos más frecuentes son “fotografía”, “transparencia”, “reprografía”, “diapositiva” e “imagen”; en oposición de las 10 347 ocurrencias y 72 tipos de la BD de la ASC, en los que los cinco términos más frecuentes son “dibujo”, “fotografía”, “estampa”, “postal” y “escultura”.

20 Charles Muller, *Estadística lingüística* (Madrid: Gredos, 1973), 27.

21 Douglas Biber, Susan Conrad, y Randi Reppen, *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge: Cambridge University Press, 1998), 38.

22 Luis Fernando Lara Ramos y Roberto Ham Chande, “Base estadística del Diccionario del Español de México”, en *Investigaciones lingüísticas en lexicografía*, ed. Luis Fernando Lara Ramos, (Ciudad de México: El Colegio de México, 1974), 37-38.

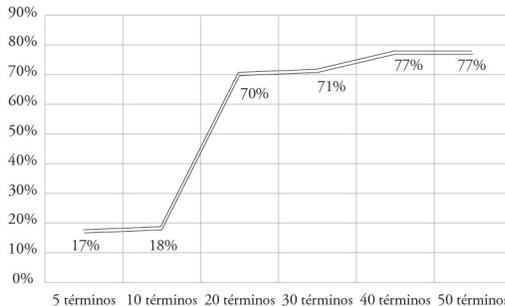
23 Lara Ramos y Ham Chande, 35-37.

24 Claudio Molina Salinas y Gerardo Eugenio Sierra Martínez, “Hacia una normalización de la frecuencia de los corpus crea y corde”, *Revista Signos. Estudios de lingüística* 48 (89), 2015, 318-324.

25 John Sinclair, “Corpora for lexicography”, en *A practical guide to lexicography*, ed. Piet Van Sterkenburg (Amsterdam/Philadelphia: John Benjamins Publishing Company, 2003), 167.

Si el primer término incluido y definido en el microtesauro hubiera sido “reprografía” y se usara para la catalogación en todos los museos y dependencias involucradas en el seminario, este término describiría 45 575 ítems del AFMT, CDA y UNIARTE, pero no le serviría a la ASC, el MUAC, el CIX o al MUCH, impidiendo, con esto, que el trabajo de documentación de las colecciones avance simultáneamente en todos los museos y dependencias.

Consecuentemente, aplicar la frecuencia de documento a nuestros listados terminológicos nos permitió establecer un ranquin en el que los primeros términos definidos servirían por igual a todos los participantes del seminario. En este caso, los cinco términos más frecuentes (por frecuencia de documento) que sirven para describir a los ítems de todas nuestras colecciones son “dibujo”, “escultura”, “pintura”, “fotografía” y “moneda”. Tan solo estos cinco términos describen al 17% de los registros de los catálogos de todas nuestras colecciones, es decir, 124 027 objetos. A continuación se puede ver gráficamente la cobertura que tienen los cinco, diez, veinte, treinta, cuarenta y cincuenta términos más frecuentes de nuestra lista respecto al total de ítems en nuestras colecciones.



3. Cobertura que tienen los términos, resultado de la aplicación de la frecuencia de documento.

Con base en la información de la gráfica anterior tomamos una decisión: comenzaríamos publicando en el microtesauro los cuarenta términos más frecuentes de nuestro listado, ya que con ello describimos dos terceras partes de los registros de nuestras BD, es decir, el 77% de registros capturados de nuestras colecciones o 552 102 ítems. Desde luego, estos cuarenta términos solo describen a los tipos de objetos artísticos y arqueológicos pero, más adelante, trabajaremos específicamente con tipos de materiales, técnicas y otras terminologías.

Estimar la frecuencia de documento de nuestros términos e integrarlos en una lista de frecuencias o ranquin de recurrencia también nos permite tomar decisiones sobre cuáles son los términos preferidos y cuáles los alternativos, esto con base en la posición que estos ocupen dentro de nuestra lista. De esta forma, tenemos evidencias cuantitativas de que “tarjeta postal” debe ser el término preferido por tener una frecuencia de documento mayor en nuestra lista, mientras que “postal” o “postcard” serían términos alternativos, ya que tienen una frecuencia menor.

En este momento no hemos identificado términos cancelados, pero suponemos que habría dos condiciones para reconocerlos, primero, una muy baja frecuencia en las BD; y, segundo, hacer un trabajo colegiado, como el que se describió al inicio de este apartado; aún falta explorar ambos procedimientos.

Por último, en esta etapa de trabajo es posible decidir si un término no fue debidamente normalizado; si es el caso, simplemente se regresa a la etapa anterior para corregir el procedimiento de normalización y se retoma el proceso.

2.3. Redacción de definiciones

Para redactar las definiciones es necesario contar con un corpus de referencia, precisar el tipo de definición que se va a aplicar y poner manos a la obra. Por una parte, en UNIARTE contamos con dos recursos documentales: el *Tesoro Terminológico de las Artes y Arqueología en México* (TTAAM), que es un corpus de definiciones en español que concentra 18 190 artículos terminográficos relativos a la historia del arte y arqueología;²⁶ y, por otro lado, el *Corpus de Documentos de la Historia del Arte y Arqueología en México* (CHAAM), que cuenta con 2 164 documentos relacionados con este dominio. Ambos recursos documentales son propiedad del IIE, están en línea y son de acceso restringido. En estos dos corpus se documentan los términos y, a partir de los contextos en los que aparecen, se infieren sus significados.

Sobre la definición, tomando en cuenta nuestras necesidades, hemos adoptado un tipo analítico o aristotélico en el que describimos la clase a la que pertenece nuestro definido (*genus*) y lo que lo cualifica, lo especifica o lo ca-

26 Claudio Molina Salinas, “Un Tesoro Terminológico de las Artes y Arqueología en México (TTAAM) desde el IIE-UNAM”, *Revista Electrónica Imágenes*, 2023.

racteriza (*differentia specifica*).²⁷ Esta diferencia específica siempre enuncia los materiales y técnicas que conforman al tipo de objeto, mientras que el *genus* es otro tipo de objeto más general (hiperónimo) o una superclase.

La estrategia planteada para facilitar la redacción de las definiciones, y que sirve para homogenizarlas, independientemente del redactor de estas, se basa en una propuesta basada en preguntas del tipo “completing”, planteamiento de Molina, Ruiz y Hernández.²⁸ A continuación se muestra la parte que retomamos de esta estrategia

La definición consta de dos partes:

1. *Genus* o género próximo. Puede ser una unidad léxica o una frase que identifique el tipo o la clase a la que esta unidad léxica pertenece. Para establecerlo es útil completar esta oración: “El (poner aquí al definido) es un tipo o clase de..._____”.

Ejemplos:

El cantor es un tipo o clase de cargo u. oficio.

El bajón es un tipo o clase de instrumento musical.

2. *Differentia* o diferencia específica. Califica o caracteriza a la unidad léxica definida; tiene la función de distinguirla de otros conceptos de la misma clase que se encuentran en relación horizontal. Generalmente se liga con el *genus* por medio de las formas léxicas: “que, cuyo, cual, entre otras, o, incluso, de cualquier participio”.

Ejemplo:

capellán de coro. Cargo u oficio *de quien* encabezaba el canto litúrgico a cambio de un pago fijo.

4. Modelo de preguntas que sirven para redactar definiciones analíticas (fragmento).

27 Robert Lew, “Identifying, Ordering and Defining Senses”, en *The Bloomsbury Companion to Lexicography*, ed. Howard Jackson (London/New Delhi/New York/Sydney: Bloomsbury Publishing, 2013), 296-97.

28 Claudio Molina Salinas, Antonio Ruiz Caballero, y Salvador Hernández Pech, “Un modelo de definiciones terminográficas para un glosario de documentos litúrgicos virreinales de México”, *Textos en Proceso* 6(1), 2020, 116.

Con base en todo lo anterior, comenzamos con la redacción de los borradores de las definiciones para, posteriormente, presentarlas a los especialistas del seminario.

2.4. Revisión de las definiciones y determinación de las relaciones jerárquicas

Los especialistas del seminario (dictaminadores de las definiciones en todos los casos) deben revisar las definiciones propuestas atendiendo a dos dimensiones: primero, preguntarse si el *genus* de la definición es, efectivamente, un tipo de objeto o una clase más generales que podrían recoger o incluir al definido; y, segundo, revisar cuidadosamente si los materiales y las técnicas que conforman la *differentia specifica*, en su conjunto, describen lo que particulariza o hace único al definido. El cotejo contra otras fuentes de autoridad no se descarta en este momento. Considerando estas dos dimensiones del significado, el dictaminador de la definición ya podría tener algunos elementos, adicionales a su experiencia profesional, para valorarla.

Naturalmente, la valoración y validación de las definiciones requiere un trabajo colegiado y una toma de decisiones de forma consensuada, por ello consideramos que el seminario permanente es un lugar conveniente para la discusión sobre la idoneidad de estas.

En esta etapa de trabajo es posible que se identifiquen problemas graves con la definición, por ello podría regresar a la etapa anterior y redactarse nuevamente; incluso, este análisis del significado puede hacer notar al analista problemas con la normalización del definido, así que también podría regresar el término a esa etapa de trabajo.

Por último, el *genus* de las definiciones determina las relaciones jerárquicas verticales en el microtesauro; de esta forma organizamos a “pintura al fresco” como un hipónimo de “pintura mural” y este, a su vez, como un hipónimo de “pintura”:

TOMATE: microtesauro de tipos de objetos, materiales y técnicas de manufactura que describen el patrimonio artístico y cultural de la UNAM

The screenshot shows the TOMATE microthesaurus interface. At the top, there is a navigation bar with 'Inicio', 'Menú', 'Agregar término', a search box with a 'Buscar' button, 'Búsqueda avanzada', and 'Sobre...'. Below the navigation bar is a horizontal menu with letters A, C, D, L, M, O, P, S, T. The main content area displays a hierarchical menu structure for 'Tomate: microtesauro':

- <Tomate: microtesauro> ▼
 - <Agentes>
 - <Conceptos asociados>
 - <Deterioros o amenazas>
 - <Materiales>
 - <Periodos espacio-temporales>
 - <Técnicas>
 - <Tipos de objetos> ▼
 - <Según los materiales y/o técnicas que los forman> ▼
 - pintura ▼
 - pintura mural ▼
 - pintura al fresco

5. Ejemplo de las jerarquías en el microtesauro.

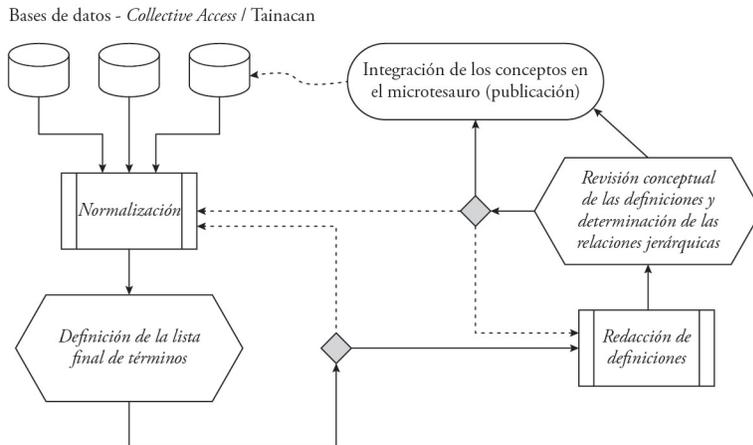
2.5. Integración de los conceptos en el microtesauro (publicación en TemaTres)

Para publicar el microtesauro hemos optado por utilizar el *software* TemaTres, que sirve para crear y editar todo tipo de vocabularios controlados: desde una lista de control terminológico simple hasta un tesaurus. TemaTres nos permite, entre otras, declarar relaciones entre términos (de equivalencia, jerárquicas y asociativas), redactar definiciones y distintos tipos de notas (de alcance, históricas, bibliográficas, etcétera), y administrar y gestionar distintos usuarios y editores; además, se puede documentar cada uno de los pasos explicados directamente en dicho *software* e ir publicando los avances paulatinamente.²⁹

2.6. Últimos ajustes a la propuesta metodológica

Después de poner en práctica de forma global esta metodología y hacer una evaluación de los cinco pasos originales de ella, descubrimos la necesidad de replantearla, sin afectar la definición de las etapas, pero implicando dos momentos de toma de decisiones al definir la lista final de términos y revisar las definiciones.

²⁹ Maria Antonia Osés, *Manual de TemaTres. La consulta y gestión de un vocabulario controlado* (Ciudad Autónoma de Buenos Aires, 2017), *passim*.



6. Propuesta metodológica para la creación de un microtesauro (diagrama de flujo, reelaborado).

Como se puede ver en el esquema anterior, esta toma de decisiones podría llevar a uno o varios casos trabajados tanto a la etapa de normalización como a la de redacción de definiciones (puntos señalados con los rombos).

3. RESUMEN Y CONCLUSIONES

En este capítulo hemos explicado el origen y contexto del trabajo de nuestro seminario, así como puesto en perspectiva la necesidad de contar con un microtesauro que complemente el uso de los *softwares* de administración, gestión y publicación de colecciones que usamos (*CollectiveAccess* y *Tainacan*).

Además, hemos propuesto una metodología de cinco pasos para publicar un microtesauro, explicada e ilustrada con ejemplos específicos que, a su vez, tiene dos momentos de toma de decisiones: justo cuando se definen la lista final de términos y se revisan las definiciones, respectivamente (véase antes: “6. Propuesta metodológica para la creación de un microtesauro (diagrama de flujo, reelaborado)”).

En nuestra opinión y experiencia, el proceso de documentación léxica y los tiempos de publicación del microtesauro se acortan notablemente con esta propuesta, misma que podría adoptarse en lugar de metodologías tradicionales y aplicarse a otros proyectos de catalogación semejantes a este.

Sabemos que incluir solo tipos de objetos es discutible y se podría cuestionar la representatividad del microtesauro en esta etapa, pero no hay que

perder de vista que el objetivo de este proyecto es representar a nuestras colecciones y no al dominio de la historia del arte y arqueología en su totalidad. Finalmente, como seminario, hemos argumentado que este procedimiento de compilación de las terminologías es un muestreo por selección intencionada o de conveniencia que, pese a ser subjetivo, sesgado y que es difícil determinar su representatividad³⁰ es suficiente para los alcances de nuestros proyectos individuales.

Por último, creemos que nuestro microtesauro debería tener un nombre más atractivo, que también describa sus alcances; por ello lo hemos llamado, tentativamente, *TOMATE: microtesauro de tipos de objetos, materiales y técnicas de manufactura que describen el patrimonio artístico y cultural de la UNAM*. TOMATE es el acrónimo de: tipos de objetos, materiales y técnicas.

30 Jordi Casal y Enric Mateu, “Tipos de Muestreo”, *Revista de Epidemiología y Medicina Preventiva* 1, 2003, 5.

4. REFERENCIAS BIBLIOGRÁFICAS

- Alcaraz Martínez, Rubén, “CollectiveAccess, un sistema de gestión y difusión de colecciones de museos, archivos y bibliotecas”, *BiD: textos universitarios de biblioteconomía i documentació* 33, 2014, <https://doi.org/10.1344/BiD2014.33.23>
- Baca, Murtha, Patricia Harpring, Elisa Lanzi, Linda Mcrae, y Ann Whiteside, eds., *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*. Chicago: American Library Association, 2006.
- Biber, Douglas, Susan Conrad y Randi Reppen, *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press, 1998.
- Casal, Jordi y Enric Mateu, “Tipos de Muestreo”, *Revista de Epidemiología y Medicina Preventiva* 1, 2003, 3-7.
- “CollectiveAccess”, [collectiveaccess.org](https://www.collectiveaccess.org/), 2022, Consultado el 30 de abril de 2024, <https://www.collectiveaccess.org/>
- Lara Ramos, Luis Fernando, *Curso de lexicología*, México, D.F.: El Colegio de México, 2006.
- Lara Ramos, Luis Fernando, *Teoría del diccionario monolingüe*, México, D.F.: El Colegio de México, 1997.
- Lara Ramos, Luis Fernando y Roberto Ham Chande, “Base estadística del Diccionario del Español de México”, en *Investigaciones lingüísticas en lexicografía*, editado por Luis Fernando Lara Ramos, 7-30, Ciudad de México: El Colegio de México, 1974.
- Lew, Robert, “Identifying, Ordering and Defining Senses”, en *The Bloomsbury Companion to Lexicography*, editado por Howard Jackson, 284-302, London/New Delhi/New York/Sydney: Bloomsbury Publishing, 2013.
- Ministerio de Educación, Cultura y Deporte, “Definición de Bien de Interés Cultural”, página del Ministerio de Educación, Cultura y Deporte, *s/f.*, consultado el 30 de abril de 2024, <https://www.cultura.gob.es/cultura/patrimonio/bienes-culturales-prottegidos/definicion.html>
- Molina Salinas, Claudio, “Un Tesoro terminológico de las artes y arqueología en México (TTAAM) desde el IIE-UNAM”, *Revista Electrónica Imágenes*, 2023, <http://www.revistaimagenes.esteticas.unam.mx/un-tesoro-terminologico>

- Molina Salinas, Claudio, Antonio Ruiz Caballero y Salvador Hernández Pech, “Un modelo de definiciones terminográficas para un glosario de documentos litúrgicos virreinales de México”, *Textos en Proceso* 6(1), 2020, 96-117.
- Molina Salinas, Claudio y Gerardo Eugenio Sierra Martínez, “Hacia una normalización de la frecuencia de los corpus CREA y CORDE”, *Revista Signos. Estudios de lingüística* 48, n.º 89, 2015, 307-31, <https://doi.org/10.4067/S0718-09342015000300002>
- Morales-del-Castillo, José Manuel, Pedro Ángeles Jiménez y Claudio Molina Salinas, “Mexico’s Tradition and Culture Entering the Digital Age: The Mexican Cultural Heritage Repository Project”, *Heritage* 2, 2019, 356-365, <https://doi.org/10.3390/heritage2010024>
- Moure Romanillo, Alfonso, *Patrimonio cultural y patrimonio natural: una reserva de futuro*, Santander: Universidad de Cantabria, 2003.
- Muller, Charles, *Estadística lingüística*, Madrid: Gredos, 1973.
- Osés, María Antonia, *Manual de TemaTres, La consulta y gestión de un vocabulario controlado*, Ciudad Autónoma de Buenos Aires, 2017, consultado el 30 de abril de 2024, <http://vocalarios.saij.gob.ar/portalthes/descargas/ManualTematresSAIJ2017.pdf>
- Salse, Marina, Javier Guallar-Delgado, Núria Jornet-Benito, María Pilar Mateo Bretos y Josep Oriol Silvestre-Canut, “GLAM metadata in museums and university collections: a state-of-the-art (Spain and other European countries)”, *Global Knowledge, Memory and Communication* 73(4/5), 2024, 477-95, <https://doi.org/10.1108/GKMC-06-2022-0133>
- Sinclair, John, “Corpora for lexicography”, en *A practical guide to lexicography*, editado por Piet Van Sterkenburg, 167-78, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2003.
- “OpenRefine”, The Open Refine Web’s Page, 2012, consultado el 30 de abril de 2024, <https://openrefine.org/>.
- “Tainacan Wiki”, The Tainacan Project Web’s page, Consultado el 30 de abril de 2024, <https://tainacan.github.io/tainacan-wiki/#/?id=tainacan-wiki>

Unidad de Información para las Artes (UNIARTE), “Unidad de Información para las Artes | Instituto de Investigaciones Estéticas”, Instituto de Investigaciones Estéticas, 2020, consultado el 30 de abril de 2024, <http://www.esteticas.unam.mx/uniarte>

Visual Resources Association, “VRA Core 4.0 Element Description”, Visual Resources, 2007, consultado el 30 de abril de 2024, https://www.loc.gov/standards/vracore/vra_Core4_Element_Description.pdf

III.
PERSPECTIVAS INNOVADORAS

Gestión de datos digitales: desafíos de la organización de información

Digital data management challenges of information organization

Lourdes Feria Basurto
*Consultora Independiente en Gestión de Datos,
Información y Conocimiento*
Elizabeth Nashielli Márquez Sánchez
*Dirección General de Archivo y Documentación
del Consejo de la Judicatura Federal*

1. SOCIEDAD RED

La doctora Li Fei Fei, considerada por el corporativo Time como una de las cien personas más importantes del mundo en el tema de inteligencia artificial,¹ afirma:

Aunque fue necesario más de medio siglo para que las condiciones previas requeridas se alinearan (hitos históricos en la evolución de los algoritmos, datos a gran escala y potencia informática bruta, todos convergiendo en los albores de la década de 2010), tomó menos de un lustro desarrollar las capacidades que desataron un cambio tecnológico. Las empresas se transformaron, se invirtieron miles de millones de dólares y todos, desde analistas de la industria hasta comentaristas políticos y filósofos, se encontraron luchando por encontrarle sentido a una tecnología que parecía explotar, de la noche a la mañana, de un nicho académico a una fuerza para el cambio global. El surgimiento de la inteligencia artificial fue de una velocidad y alcance posiblemente sin precedentes en toda la historia...²

1 Li Fei Fei, Time 100/AI, <https://time.com/collection/time100-ai/>

2 Li Fei Fei, *The Worlds I See* (New York: Flatiron Book, 2023), 8 y 9.

Las tecnologías de información, incluso antes del auge de la inteligencia artificial (IA) han provocado la construcción de una nueva estructura social a la que el sociólogo Castells³ ha denominado *Sociedad-Red*, y que, si bien tiene como plataforma la tecnología, también involucra redes de producción, relaciones y maneras de accionar que demandan un cambio en las organizaciones, empresas, instituciones y otros entornos, incluidas las bibliotecas.

1.1. Renovarse o morir: la transformación digital en bibliotecología

Dentro de esa sociedad-red el desafío es incorporar la transformación digital, es decir, el proceso para “responder a la ‘presión digital’ que sienten las organizaciones desde múltiples direcciones”⁴ para ello se identifican cuatro vertientes: la evolución de la competencia, los nuevos actores, los agentes internos en las organizaciones (en el caso de las bibliotecas serían los usuarios, los bibliotecarios, las autoridades, los proveedores de servicios) y los procesos/servicios tanto tradicionales como innovadores.

En este punto es importante explorar la noción de digitalización y reflexionar sobre lo que se entiende por ese término. En principio, la idea de digitalizar se orientaba exclusivamente hacia contenidos e información, se limitaba a la idea de trasladar del papel a la computadora una determinada información utilizando un escáner para convertir documentos y pasarlos a formato PDF, es decir, en transformar un documento de contenido analógico a digital mediante el uso de *hardware*, *software* y herramientas de visualización.

Sin embargo, en la actualidad, según la Real Academia Española (RAE), el concepto general de digitalización es registrar datos en forma digital. Esto significa que no se trata únicamente de convertir documentos, el concepto de digitalización abarca mucho más, ya que implica el registro de datos en formato digital y se está aplicando a acciones tan diversas como compras en línea, actividades deportivas, procesos de carga de combustible, mantenimiento de vehículos y contratación de servicios de comida a domicilio, por mencionar algunos.

Aplicado a la academia, más allá de la necesidad de digitalizar los cursos y los procesos educativos contempla, en cambio, el registro en formato digital de todo el proceso de enseñanza y aprendizaje, sin perder la parte humana. Todos los actores educativos se encuentran inmersos en un proceso de aprendizaje, incluso las propias instituciones y sus bibliotecas, intentando mejorar conti-

3 Manuel Castells, *La sociedad red* (México: Siglo XXI, 2006).

4 Luis Miguel Uriarte y Manuel Acevedo Ruiz, “Sociedad Red y transformación digital”, *Economía industrial*, n.º 407, 2018, 37.

nuamente cómo transferir el conocimiento y fomentar la interacción humana, a pesar de la distancia. Más que transformar lo analógico a lo digital se trata de nuevas maneras de vivir porque la sociedad va cada vez más a la convivencia en entornos de red. La digitalización se refiere a la conversión de contenidos y procesos, pero lo que realmente se está experimentando es una transformación digital, un efecto social y global de la tecnología.

Y aún más allá de eso, este fenómeno implica conectar y mantener digitalmente todas las operaciones que antes se realizaban de forma manual o automatizando solo algunos procesos bibliotecarios. Hoy en día la realidad muestra la necesidad de aprovechar el internet de las cosas, el *big data*, el aprendizaje de máquina (*machine learning*) y otras tecnologías como el *blockchain* para el manejo de transacciones en línea. Corresponde asumir que ahora el mundo bibliotecológico se tiene que adaptar a nuevas formas de hacer las cosas, de trabajar, comunicarse e interactuar, fortaleciendo habilidades digitales, tanto a nivel personal como organizacional.

Tener una página web, utilizar herramientas de biblioteca digital o contar con bases de datos no constituye transformación digital. Estas son herramientas fundamentales, pero para realmente ingresar a la transformación digital se tiene que reconocer que el actual punto de partida se encuentra en un momento similar al advenimiento de internet o incluso antes, cuando en las primeras versiones de computadoras de escritorio tuvieron que desarrollarse sistemas como los OPAC.

1.2. Cambio tecnológico o cambio cultural

La transformación digital implica un cambio de acciones, prácticas y actitudes; es necesario que los nuevos modelos tecnológicos se apliquen activamente en la vida digital de las bibliotecas. Esto significa desarrollar las capacidades para integrar el contenido en cursos, MOOC (cursos en línea masivos abiertos, por sus siglas en inglés), e-aprendizaje y en fomentar la interacción de los usuarios en sus redes. La biblioteca debe trascender la simple publicación en plataformas sociales y llegar al nivel de interactividad que se observa en otras industrias, como por ejemplo la moda, donde se puede explorar cada detalle de una prenda y recibir retroalimentación directa del público. La transformación digital implica que la biblioteca ya no es un único sitio, sino que debe estar presente en innumerables lugares y plataformas para asegurar que las múltiples facetas de la biblioteca se integren naturalmente en el funcionamiento general de su entorno.

Ello representa un desafío significativo que ha de ser abordado con determinación, encontrando la manera de evolucionar y estar al tanto de los cambios. Desde chefs hasta diseñadores de moda, pasando por expertos en logística, compras y contabilidad, todas las profesiones se ven afectadas por la transformación digital; ejemplos cotidianos como los códigos QR en los restaurantes que permiten ver los menús en línea, las aplicaciones de entrega que ofrecen servicios de compra de medicamentos de la farmacia y artículos del supermercado tendrían que convertirse en referentes para optimizar procesos de pedido y suministro de información a las comunidades que atienden las bibliotecas.

Estos cambios, a su vez, provocarán un impacto directo en las necesidades de formación de los nuevos profesionales, “los bibliotecarios hace décadas que dejamos de ser custodios de la información y la documentación. Somos facilitadores de esta información y documentación”⁵ y es momento de ir reconociendo que los modelos y contenidos educativos deben evolucionar para apoyarse en nuevas y mejores formas de obtener y utilizar la información que estos profesionales requerirán en un entorno cambiante que, claramente, se muestra influenciado por una vertiente que día a día está presente y que continuará incrementándose en el futuro: los datos digitales.

2. GESTIÓN DE DATOS: UNA TEMÁTICA EMERGENTE

En 2007, Weinberger en su libro *Everything Is Miscellaneous: The Power of the New Digital Disorder*, ampliamente reconocido en círculos académicos, argumenta que las diversas categorías mediante las cuales se solían etiquetar, organizar y clasificar han estado limitadas de una u otra manera por el mundo físico, sin embargo, con el vasto poder de la informática moderna y las diversas formas en que se puede utilizar la web, ya no es necesario utilizar tales categorías y gracias a la imaginación y comprensión las nuevas maneras de estructurar los recursos y contenidos digitales se están dando a través de una *miscelánea* estructurada. En esa publicación lo que se pretende es hacer una invitación a repensar los caminos tradicionales de procesamiento de información, pero sobre todo de su componente nuclear: los datos.

Por ello, la gestión de esos datos ahora tiene que ser repensada en el mundo bibliotecológico lo mismo que en empresas, en organismos públicos, en servicios de entretenimiento como Netflix, en compañías que se basan en el manejo de

5 Francisco Javier García, “Lenguaje verbal y lenguaje no verbal: esta es la cuestión... también en la Gestión de la Atención al Usuario en Bibliotecas”, entrevista hecha por Francese García Grimau, *Desiderata*, n.º 22 (abril, mayo y junio de 2023), <https://www.auxiliardebiblioteca.com/wp-content/uploads/2023/04/Desiderata-22.pdf>

inventarios como Amazon, Shein, o cualquier otro modelo de ventas con o sin almacenes físicos. En la sociedad-red todo es información y por ello, todo es susceptible de ser catalogado, clasificado y codificado con descriptores y metadatos.

De ahí la importancia de tomar consciencia de la necesidad del habilitamiento digital basado en la gestión de datos para la entrega de valor a través de una mirada incluyente, actual y que permita accionar en otros formatos. Los elementos clave que componen la transformación digital en los procesos de manejo de información, sea una aplicación móvil, un sitio web, un portal de conocimiento o, bien, todos ellos, inician con los datos, por ello a continuación se hace un repaso de ese concepto, seguido del fenómeno *Big Data* para de ahí pasar a comentar el ciclo de vida de los datos con el objetivo de precisar esos términos que constituyen la nueva plataforma de trabajo para las colecciones en las bibliotecas y otros entornos telemáticos del siglo 21.

2.1. ¿Qué son los datos?

Dentro de las prácticas contemporáneas actuales los datos han adoptado un puesto de suma importancia, redefiniendo la forma en que las organizaciones, ya sean empresariales, gubernamentales o académicas gestionan, protegen y aprovechan el vasto flujo de información que las rodea. Para comprender a profundidad en qué consisten estas actividades es necesario comenzar entendiendo la materia prima con la cual funcionan: los datos.

En este sentido, el dato es la unidad elemental e indivisible de información que representa un hecho, una medida o una descripción específica.

Se trata del elemento constitutivo del conocimiento. Comprenden hechos, representaciones o mecanismos por los cuales es posible medir e identificar algún aspecto de nuestro mundo-realidad la cual nos enfrenta a un universo de fuentes y elementos factuales en espera de que desarrollemos las técnicas para medirlos e identificarlos, es decir, convertirlos en datos.⁶

Los datos al ser procesados y contextualizados adquieren significado y contribuyen al vasto conjunto de conocimientos disponibles y son presentados en diversas formas, incluyendo números, texto, imágenes o cualquier otra representación simbólica.

6 Ennio Prada, “Los insumos invisibles de decisión: datos, información y conocimiento”, *Anales de Documentación*, n.º 11, 2008, 184.

Los datos son la unidad fundamental de información en el entorno digital, su valor radica en la capacidad que tienen para ser almacenados, manipulados y transmitidos, son la materia prima esencial que alimenta los sistemas de información, por lo tanto, su calidad y veracidad son fundamentales para la toma de decisiones en cualquier proceso interno, en la empresa o institución. La calidad de la información que resulte de los datos dependerá en gran medida del análisis y procesamiento meticuloso al que fueron sometidos.

2.2. Big Data

El aumento significativo en la cantidad de datos generados, almacenados y procesados en diversas plataformas y dispositivos ha dado lugar al fenómeno denominado *Big Data*, el cual se caracteriza por el enorme volumen de información digital que se ha multiplicado exponencialmente en las últimas décadas debido al creciente número de aparatos conectados, aplicaciones en línea y sistemas de almacenamiento informático. Esta avalancha de datos proviene de diversas fuentes, como redes sociales, apps, trabajo en el entorno conocido con el nombre de “la nube” y otros instrumentos.

La densidad de datos, por otro lado, se refiere a la concentración de información en un espacio o unidad de almacenamiento determinado. A medida que aumenta el volumen de aquellos, su concentración también tiende a crecer, ya que más información se almacena en los mismos recursos físicos o virtuales. Un ejemplo claro de cómo la producción de datos y el rol de la tecnología se potenció lo mostró el confinamiento del año 2020 cuando la crisis de salud replanteó las maneras de accionar y se pasó a contextos más digitales, que abrieron nuevas vías para trasladar los procesos analógicos y mecánicos a escenarios basados en apoyos tecnológicos y *software* que alimentaron aún más el universo de la *big data* desarrollados con algoritmos con programación informática.

Este crecimiento exponencial en el volumen y corpus de datos presenta tanto desafíos como oportunidades para las organizaciones y la sociedad en general. Por un lado, el manejo y análisis eficiente de grandes volúmenes de *bytes* requiere infraestructuras tecnológicas robustas, herramientas de análisis avanzadas y personal capacitado. Por otro lado, estos datos masivos ofrecen un tesoro de información que puede utilizarse para mejorar la toma de decisiones, descubrir patrones ocultos, personalizar servicios y productos, y generar nuevos conocimientos en áreas como la medicina, la investigación científica, el desarrollo empresarial, e incluso, las artes y las humanidades.

2.3. Ciclo de vida

Por lo anterior, los nuevos acervos ahora son los datos digitales y, para contribuir a un cuidado saludable de los datos, se debe comprender el ciclo de vida de los mismos. Este proceso se enfoca en aspectos cruciales como la clasificación, almacenamiento, procesamiento, análisis, distribución, acceso, seguridad y privacidad de los mismos. “Se trata de conocer todo lo que ocurre con el dato: desde cómo se crea o se incorpora a un proceso, hasta el descarte final, reciclaje o venta del mismo”.⁷

Dicho ciclo es adaptable a la institución, autoridad o empresa que lo esté utilizando y presenta bastantes similitudes con el ciclo de vida de los libros (procesos técnicos) en una biblioteca física, pero a grandes rasgos el ciclo de vida de los datos comprende en su gran mayoría los siguientes criterios:

Captura: Se recopila la información desde diversas fuentes, como formularios en línea, listas de Excel, entre otras fuentes; cabe mencionar que las fuentes deben ser confiables.

Almacenamiento y Organización: El almacenamiento se enfoca en la elección de los medios físicos o virtuales adecuados para conservar la información de manera segura y confiable, mientras que la organización de datos se centra en la estructuración y categorización de la información de manera coherente y comprensible. Esto incluye la aplicación de técnicas como la normalización de bases de datos y la creación de índices.

Procesamiento y Análisis: En este apartado se lleva a cabo la transformación de datos crudos en información significativa mediante técnicas de registro y clasificación, revelando patrones, tendencias y vínculos.

Distribución y Acceso: Facilita el encuentro y localización de la información relevante a través de sistemas y plataformas adecuadas, asegurando que las personas tengan caminos confiables para llegar a los datos correctos.

Seguridad de Datos: Se refiere a la implementación de medidas orientadas a proteger la información contra accesos no autorizados, pérdida de datos o alteraciones, abordando aspectos como autenticación y la anonimización.

Privacidad de Datos: De la mano con el anterior, garantiza el respeto a la privacidad de los individuos, abordando cuestiones éticas y legales asociadas con la recopilación y uso de información personal.

Eliminación Responsable: Proceso de descarte de datos obsoletos de manera segura, evitando riesgos y cumpliendo con regulaciones y políticas de privacidad.

⁷ Marcos Pérez, *El ciclo de vida del dato* (España: Universitat Oberta de Catalunya, 2017), 11.

Organización del conocimiento...

El principal propósito del ciclo reside en transformar datos crudos soportados en contenedores digitales en información significativa y finalmente en conocimiento útil. La gestión eficaz de estos activos informativos fortalecerá las prácticas educativas, científicas y permitirá, en las empresas, la toma de decisiones informadas, la amortización de procesos, la innovación y el cumplimiento normativo.

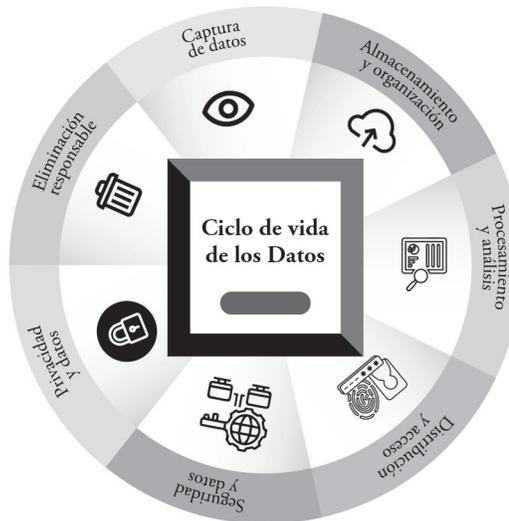


Figura 1. Ciclo de vida de los datos (Fuente: Elaboración personal).

2.4. Los datos en empresas, gobierno y academia

La seguridad, la organización, así como la preservación de datos, son prioridades críticas en entornos empresariales, gubernamentales y académicos, donde la información sensible y estratégica es susceptible a diversas amenazas.

Al tocar estos temas se advierten dos miradas igualmente importantes, una que se da desde el ámbito del *software* y la informática, pero otra, igualmente necesaria, es desde la perspectiva de las ciencias de la información. Ambas se complementan. En la primera, la gestión de datos se convierte en un elemento estratégico, permitiendo la identificación de patrones, tendencias y relaciones que sustentan la formulación de estrategias de programación sólidas, así como funciones dirigidas a la *seguridad tecnológica de datos*, esencial para salvaguardar la propiedad intelectual, la información financiera y la confianza de usuarios.

En la segunda, el objetivo fundamental es evitar “una dispersión de datos e información que, en ocasiones, duplica tareas y hace imposible su recuperación por parte de quienes la necesitan en el momento oportuno”,⁸ para ello es necesaria una buena planificación, para que la información no solo pueda ser recolectada y almacenada sino que también se pueda proyectar e interpretar; es decir, tener certeza sobre lo que existe y lo que no existe, así como las opciones a futuro. Una adecuada organización de los datos permitirá que estos cumplan su función, por lo tanto deben seguir los principios *FAIR*, sigla en inglés que representa los términos *Findable, Accesible, Interoperable* y *Reusable*,⁹ lo cual quiere decir que la gestión de las colecciones de datos ha de permitir que estos sean fácilmente localizables, digitalmente legibles, coherentes, recuperables en plataformas de *software* similares y permitan el manejo, lectura y reutilización de los materiales.

En el ámbito académico, la organización de datos se concentra en la gestión eficiente de recursos educativos y de investigación; por ejemplo, la catalogación precisa y el uso de metadatos en bibliotecas digitales y repositorios académicos; las prácticas de preservación de datos de investigación mediante la utilización de repositorios digitales y estándares de formato duraderos asegura la accesibilidad a largo plazo; la colaboración e intercambio de datos entre instituciones académicas para fomentar el avance de la investigación.

Por último, en otros sectores, como el gubernamental, la organización de datos se vincula con la transparencia, la memoria de las instituciones, la rendición de cuentas y la eficiencia en la prestación de servicios públicos. En ese ámbito, los datos han de pasar por procesos de clasificación, según su importancia y sensibilidad, junto con su jerarquización; se debe contar con políticas de retención de registros para garantizar la conservación adecuada; controles de acceso para que solo individuos autorizados tengan acceso a información sensible para proteger información clasificada, garantizar la privacidad y mantener la confianza pública.

3. LA GESTIÓN DE DATOS EN BIBLIOTECOLOGÍA

Compañías como Pixar y Disney hoy en día buscan *data librarians* en sus anuncios de vacantes de puestos de trabajo porque están conscientes de que la falta de organización de su información, específicamente de los sets de datos

8 Yuniet, Rojas, “Organización de la información: un factor determinante en la gestión empresarial”, *ACIMED* 12, n.º 2 (marzo/abril de 2004), 4.

9 The Future of Research Communications and e-Scholarship (FORCE11), 2019, <https://www.force11.org/fairprinciples>

que manejan, pueden causar la pérdida de importantes recursos creativos. Estas empresas están contratando profesionales con conocimientos en el manejo de datos, lo que tendría que representar una ampliación de las oportunidades laborales para quienes egresan de las carreras de documentación, bibliotecología, archivología y afines.

Ocurre lo mismo en universidades. Tal es el caso en Estados Unidos, de las instituciones académicas más destacadas, las *R1* (*Research One*, es decir, investigación de primer nivel) que son conocidas por su trabajo de vanguardia, tienen al menos un *data librarian*, pero llegan a incluir en plantilla hasta veinte de ellos porque saben que los profesionales de la información desempeñan un papel crucial.

A continuación, se revisan algunas áreas de la bibliotecología en las que el trabajo con datos es cada vez más una necesidad ineludible.

3.1. Colecciones digitales

Este es uno de los principales temas en la gestión de datos y se refiere a la construcción, preservación y acceso a los recursos digitales. “La gestión de colecciones es una labor en evolución que continúa su metamorfosis y su etimología va evolucionando junto con su término, desde la selección al desarrollo de la colección, a la gestión de la colección y, ahora, a la gestión de colecciones digitales”.¹⁰

La gestión de colecciones se rige por dos principales bases “uno es el término ‘principios’, pues son las bases que rigen cómo se desarrolla la colección y bajo los cuales se administra la biblioteca, manteniendo de esta manera la consistencia y cohesión de la misma; y el otro término ‘políticas’, por ser el producto que se deriva del trabajo que se realiza, documentos que a la larga se convierten en el sustento teórico llevado a la práctica de cómo se conforma la colección”.¹¹

La gestión de colecciones satisface las necesidades de una comunidad, esto mediante una colección sólida y completa que permita una amplia variedad de áreas y temas de interés para la comunidad con la que se esté trabajando, de igual manera se busca complementar estas colecciones con recursos externos de la biblioteca con el fin de responder a las necesidades de la comunidad de manera eficaz y eficiente.

10 Amelia García, Idania Licea y Riselis Martínez, “Del desarrollo de colecciones a la gestión de colecciones digitales: mirada a la evolución histórica del término a partir de apuntes teóricos”, *Bibliotecas. Anales de Investigación* 19, n.º 3, 2023, 2.

11 *Ibid.*, 3.

Para que la biblioteca, empresa o institución marche de manera adecuada, acorde a la colección que tiene, se deben establecer fines y objetivos claros y que identifiquen los grupos de usuarios y los servicios que se habrán de ofrecer. Dentro de este servicio también debe tomarse en cuenta la facilidad de acceso a otros acervos, y las principales actividades a desarrollar son la adquisición y descripción de objetos digitales, su almacenamiento y su difusión en entornos digitales, de igual manera deben estar involucrados estándares y protocolos que describan los metadatos, la aplicación de políticas de acceso y uso, junto con la adopción de tecnologías de almacenamiento y preservación.

Al igual que la gestión de datos, la gestión de colecciones digitales también se enfrenta a desafíos únicos, como la obsolescencia tecnológica, el cuidado y protección a largo plazo de formatos digitales y la protección de la integridad y autenticidad de los objetos digitales; de ahí la necesidad de “conocer la naturaleza de los materiales digitales y las nuevas circunstancias de trabajo en las que deben desenvolverse, cada vez más en un entorno en red”.¹² Para estos desafíos los profesionales en este campo deben tener enfoques multidisciplinarios que combinen conocimientos de ciencia de la información, tecnología de la información, preservación digital y políticas de gestión de la información.

Las tendencias en la publicación y la gestión de colecciones están y seguirán experimentando transformaciones diversas, por lo tanto, no todas podrán ser tratadas desde un punto de vista bibliotecario a la manera tradicional. Es el caso, por ejemplo, de la serie *Stranger Things*, donde Netflix aprovechó su colección de datos para construir una serie que resonara con su audiencia, utilizando elementos que se alineaban con los gustos y preferencias de los espectadores. Esto plantea a quienes gestionan colecciones de datos digitales nuevos caminos para explorar la información de los usuarios, con la finalidad de ayudar a la biblioteca a comprender mejor el interés de sus comunidades y atiende con estrategias más efectivas el desarrollo de sus acervos.

3.2. Catalogación y metadatos

¿Cómo influyen las nuevas tecnologías, por ejemplo, la inteligencia artificial y el aprendizaje automático, en la automatización y mejora de los procesos de catalogación y creación de metadatos en entornos digitales, particularmente en la sistematización de los conjuntos de datos? La catalogación, como se sabe, es el proceso de describir y organizar recursos de información de manera ordenada y coherente y asignar etiquetas descriptivas a esos recursos para facilitar

¹² *Ibid.*, 3.

su identificación y recuperación. Tradicionalmente se ha basado en normas y esquemas como el AACR2 (*Anglo-American Cataloging Rules*) y el RDA (*Resource Description and Access*). Sin embargo, en el mundo de los datos la catalogación se enfrenta a desafíos únicos debido a la diversidad de formatos y la dinámica de los recursos digitales.

En este punto cobra especial relevancia el manejo de metadatos, “no es un tema nuevo en la bibliotecología, puesto que el término fue acuñado en la década de los sesentas por Myers con el propósito de referirse a la descripción de un conjunto de datos. Actualmente referirse a lo que son los metadatos es con el firme propósito de significar algo tan sencillo, pero a la vez complejo que está relacionado con la estructura y los datos de los recursos (entidades) que permiten describir, explicar, identificar, usar y recuperar los contenidos”.¹³

La catalogación está en una nueva etapa que va más allá del trabajo, con piezas documentales clásicas, como pudieran ser los libros tanto físicos como electrónicos o artículos científicos en formatos PDF, XML, HTML, epub y otros. El espectro ahora se amplía y observamos que los grandes consorcios de venta de bienes y servicios son conglomerados de ítems que deben, previamente, ser catalogados. Un ejemplo podría ser Amazon, que registra sus objetos digitales mediante un sistema de metadatos que incluye información sobre el producto, como título, autor, descripción, categoría, formato, fecha de lanzamiento, calificaciones de los clientes, precios y otros. Utiliza, además, algoritmos y herramientas de análisis de datos para optimizar continuamente sus catálogos, precisar sus recomendaciones y ofrecer a quienes usan la plataforma lo que se conoce como “experiencia de compra”.

Estos escenarios, junto con la proliferación de recursos digitales y la dinámica cambiante de la información en línea, marcarán una transición en la catalogación digital y el uso de metadatos en las prácticas y en la normatividad.

3.3. Preservación digital

Ante este contexto y considerando la rápida evolución de la tecnología y los formatos, otro de los retos que la transformación pone sobre la mesa es el del cuidado de los documentos digitales a largo plazo. Esto se conoce como preservación digital y se refiere al: “conjunto de principios, políticas, normas y estrategias diseñadas para asegurar que un objeto digital permanezca accesible,

13 Ariel Rodríguez, “Sobre la catalogación y la actividad de los metadatos aproximaciones a su coexistencia en la era del acceso”, *Biblioteca Universitaria* 11, n.º 1 (enero-junio de 2008).

inteligible y usable a través del tiempo y de los cambios tecnológicos, su fiabilidad y exactitud están protegidas y su autenticidad es verificable”.¹⁴

Ante la fragilidad y obsolescencia de los medios digitales, así como la rápida evolución de la tecnología y los formatos de archivo, los nuevos profesionales de la información tendrán que aprender y concientizarse de estrategias que definan metas y procedimientos para salvaguardar y mantener los datos, sus elementos, sus componentes y toda la información asociada.

Al ser de constitución mutable, los recursos digitales son más vulnerables y, por lo mismo, requieren que su preservación sea pensada con antelación, incluso desde el momento mismo de su creación. La atención a lo largo de la vida del documento digital debe de ser pensada y construida desde su nacimiento para que sea fiable, exacto y auténtico en el presente y aun con el paso del tiempo.

Para que la preservación de datos tenga un buen comienzo es importante partir de esa sensibilización aplicando conocimientos de procesamiento de información. Hablar de ello es abordar el tema de las colecciones digitales que se mencionaron en párrafos anteriores y considerar la inclusión, en los nuevos programas de estudios, de modelos de referencia estándar como el conocido por la sigla OASIS, aprobado por la ISO, que describe seis actividades centrales para todo repositorio: planeación de la preservación, administración, gestión de datos, ingesta o ingreso, almacenamiento archivístico y acceso. Estas actividades incluyen desde recibir y registrar la información digital hasta proporcionar servicios de almacenamiento, mantenimiento, recuperación y acceso a largo plazo.¹⁵

3.4. Análisis de uso y usuarios

“Cuando muchos bibliotecarios ingresaron a la profesión, probablemente nunca imaginaron que serían analistas de datos en su carrera. Sin embargo, la capacidad de capturar y evaluar información sobre sus colecciones, programas y usuarios es un conjunto de habilidades cada vez más importante que deben

14 Consejo Internacional en Archivos e InterPARES, *Un marco de referencia para la preservación digital*, trad. Alicia Barnard, Alejandro Delgado y Juan Voutssás (Ciudad de México: Archivo General de la Nación, 2017), 17.

15 *Ibid.*, 27.

tener”.¹⁶ De ahí la necesidad de ingresar en temáticas como la analítica de usuarios, especialidad que se hace cargo de gestionar los datos de las comunidades de una biblioteca con el propósito de comprender qué es lo que desean, buscan y necesitan. Idealmente se debería contar incluso con datos de no-usuarios. Su insumo de trabajo son los informes estadísticos, historias de uso de información y publicaciones, videos, audios o fotografías que contribuyan al conocimiento del entorno y que permitan analizar cómo acercar más la biblioteca a las personas.

El trabajo con datos en este sentido debería permitir responder preguntas como las siguientes:

- ¿Cuáles son las metodologías más efectivas para recopilar y gestionar datos de usuarios en bibliotecas?
- ¿Qué variables son relevantes para caracterizar la población de usuarios y no usuarios en una biblioteca?
- ¿Cómo influye la analítica de usuarios en la toma de decisiones para mejorar los servicios bibliotecarios?
- ¿Cuál es el impacto de la analítica de usuarios en la satisfacción de las personas y en el éxito estudiantil en las bibliotecas académicas?
- ¿Cómo influyen los datos de uso de los repositorios electrónicos en la toma de decisiones sobre renovaciones de suscripciones y adquisiciones de recursos en bibliotecas?
- ¿Qué estrategias pueden implementarse para fomentar la participación de los usuarios en la recolección de datos y mejorar la calidad de la información obtenida?

Sistematizar adecuadamente ese tipo de información y aprender a interpretar los datos que de ella emanan permitirá plantear mejoras, soluciones y cumplir de maneras más efectivas con la misión de las bibliotecas.

3.5. Alfabetización en datos

Una de las fortalezas importantes de la profesión la constituyen los modelos ALFIN, que constantemente buscan guiar a las personas en las competencias para la alfabetización informacional. Llevar esto al ámbito de los datos no

16 Library IQ, “Using Data Is Essential to Library Operations. Here’s Why”, *Library Journal* (4 de diciembre de 2023), <https://www.libraryjournal.com/story/using-data-is-essential-to-library-operations.-heres-why-lj231204#:~:text=Using%20data%20can%20help%20librarians,data%20analysts%20in%20their%20career>

significa convertir a los bibliotecólogos en científicos de datos, sino en impulsar una de sus principales competencias que es la habilidad de gestionar los contenidos para generar conocimiento significativo.

Las bibliotecas académicas, por ejemplo, en sus programas ALFIN podrían incluir un área de atención a las buenas prácticas para la gestión de datos, trabajando el tema lo mismo con estudiantes de todos los niveles (preparatoria, grado, posgrado), que con docentes e investigadores en ejercicio. Se pueden conformar charlas de una hora o sesiones de hasta veinte horas en las que se les muestre cómo se elabora un plan de datos y cómo pueden estructurar su base de conocimiento desde el inicio de la investigación; otros temas serían la relevancia de los repositorios de datos, la compartición e intercambio de datos y también recomendaciones acerca de cómo llevar un diario de investigación o bitácora de laboratorio. Estas actividades serían de beneficio no solo en disciplinas relacionadas con las ciencias sino también para las humanidades y las ciencias sociales, ya que el manejo de datos permea todos los ámbitos.

En las bibliotecas públicas un modelo ALFIN especializado en datos ayudaría a capacitar a la ciudadanía para ir aclarando conceptos. En la sociedad-red será muy valioso contar con cuatro habilidades fundamentales: comprensión de conceptos digitales, habilidades de pensamiento crítico y capacidad de colaboración en línea para navegar y contribuir de manera significativa en el entorno digital contemporáneo. La enorme actividad digital que ocurre constantemente en internet, desde el número de personas conectadas hasta la cantidad de correos electrónicos y búsquedas en Google y en aplicaciones de inteligencia artificial en tan solo un minuto, es impresionante (véase <https://www.internetlivestats.com/>).

Trabajar en, con y para una cultura de datos y desarrollar habilidades para entender, analizar y argumentar con ellos es una responsabilidad que suma a la profesión. En este punto parece muy pertinente la reflexión en torno a la necesidad de atender la temática en las instituciones educativas, por lo cual la propuesta “la alfabetización en datos está buscando una sede ¿por qué no la biblioteca?”¹⁷ no pierde vigencia.

3.6. Necesidad de formación y capacitación

Para apoyar la adopción de competencias en gestión de datos que requieren los profesionales de la información, las autoras llevaron a cabo en 2023 la compilación de un inventario de los cursos existentes, tanto virtuales como

17 Alex, Merril, “Library+”, *Public Services Quarterly* 7, n.º 3-4, 2011.

presenciales para bibliotecarios académicos, basado en un perfil de necesidades que identificó aquellas opciones que permitiesen subsanar las brechas en su momento.

Para realizar la compilación se tomaron en cuenta los siguientes temas:

- Fundamentos de la gestión de datos.
- Introducción a los datos de investigación.
- Difusión y comunicación de la importancia de la gestión de datos.
- Implementación de los servicios de datos en bibliotecas.
- Gestión de planes de datos y repositorios de datos.
- Conocimientos básicos sobre herramientas de análisis, visualización y programación para la gestión de datos.

Luego de identificar los cursos y talleres existentes, tanto virtuales como presenciales para bibliotecarios académicos, se integró un listado de 49 ítems que a continuación se incluyen. Todos ellos pueden ser localizados en la web para ver, mediante el vínculo para su acceso en internet, para conocer los detalles, tales como el contenido específico, la duración y si tienen un costo o están disponibles de manera gratuita.

Nombre	Autor
RDA y las humanidades digitales, RDA y biblioteconomía archivística y ciencias de la información	Research Data Alliance
Investigador Profesional Certificado (PCI)	ASIS Internacional
Datos de investigación	Association of college & research libraries
Datos de investigación	Digital Research Alliance of Canada
Gestión y publicación de datos de investigación	Universidad de Tartu (UT)
Formación RDM para investigadores en Etiopía	EIFL y el Consorcio de Bibliotecas Académicas y de Investigación de Etiopía (CEARL)

Nombre	Autor
Capacitación de RDM	Serbia.RDM
Estudios de casos de gestión de datos de investigación	The University of British Columbia
Guía de conservación y gestión de datos de ICPSR	Inter-university Consortium for Political and Social Research (ICPSR)
Data Management	Australian National
Directrices para planes efectivo de gestión de datos	International Association for Social Science Information Service and Technology (IASSIST)
Módulos DataONE	DataONE
Gestión de datos	Universidad de Wisconsin
Gestión de datos	Blog Datos Ab Initio
Gestión de datos	Bibliotecas de datos
Gestión de datos	DataCure
Gestión de datos	DataQ
Gestión de datos	DMP Assistant
Plan de gestión de datos	UK Research and Innovation
Kit de herramientas	Universidad de Penn State
Guía de datos de Investigación	Stony Brook University
Conjunto de herramientas	DMPTool
Gestión de datos	Research Compliance and Integrity Office (RCI)

Tabla 1. Cursos sobre gestión de datos (Fuente: Elaboración personal).

Nombre	Autor
Creación de planes de gestión de datos de la investigación	Pontificia Universidad Católica de Chile
Introducción a la gestión de datos de investigación	Universidad Pablo de Olavide
La gestión de datos de investigación	Pérez Aliende, María Luisa

Nombre	Autor
Curso Gestión de datos de investigación en la práctica científica	Secretaría Nacional de Ciencia Tecnología e Innovación de la República de Panamá (SENACYT-Panamá)
Acceso a la Información y gestión de datos para la comunicación	Facultad de ciencias economicas
<i>Big Data</i> Simplificado	NIC, Innovation, Integration
Gestión práctica de datos de investigación - Pasos Principales y Herramientas Esenciales	NIC, Innovation, Integration
Siete buenas prácticas para gestión de datos en las bibliotecas	Lourdes Feria Basurto
Capacitación en prácticas de investigación de apertura y reproducibilidad	Center for open science
Academia de bibliotecarios de gestión de datos de investigación (RDMLA)	Elsevier y LIS
Aplicaciones abiertas para RDM 102: Más allá de la gestión de datos de investigación para bibliotecarios de ciencias biomédicas y de la salud	National Library of Medicine (NLM) and the National Network of Libraries of Medicine Training Office (NTO)
Gestión de datos	Mantra
Gestión de datos de investigación	Cornell University
Proyecto RDMRose	University Sheffield, JISC
Gestión de datos de investigación	UK Data Archive
Plan de estudios de gestión de datos colaborativos de Nueva Inglaterra	Biblioteca Lamar Soutter de la Facultad de medicina de la Universidad de Massachusetts en asociación con varias bibliotecas de la región de Nueva Inglaterra
Gestión e intercambio de datos de investigación	Universidad de Carolina del Norte en Chapel Hill y la Universidad de Edimburgo en conjunto con Coursea
Gestión de datos	Comunidad para la Integración de Datos del Servicio Geológico de EE. UU., la Asociación de Información de Ciencias de la Tierra (ESIP) y DataONE

Nombre	Autor
Certificado Académico de Posgrado en Gestión de Datos y Curación Digital (GAC)	Departamento de Ciencias de la Información, Universidad del Norte de Texas
Gestión de datos de investigación (RDM) curación digital y desarrollo de servicios RDM	The Digital Curation Centre (DCC)
Guía de curación de DH	Digital Humanities Data Curation (DHDC)
Guía de expertos en gestión de datos (DMEG)	CESSDA- free online expert tour guide to data management
Gestión de Datos	MIT Libraries
Introducción a la gestión de datos	Library Juice Academy
Ciclo de vida de los datos, ¿Qué son los datos? Escencial Python	LinkedIn Learning
Analista de datos	Udacity Nanodegree

Tabla 2. (Continuación) Cursos sobre gestión de datos (Fuente: Elaboración personal).

4. Conclusión

De la noche a la mañana en el mes de marzo de 2020 las sociedades de los distintos países del orbe migraron sus acciones cotidianas a transacciones en línea, compras, educación, estilos de vida y los trabajos tuvieron que realizarse por medios tecnológicos, trayendo como consecuencia una gran transformación digital. Así, el desarrollo del entorno de la sociedad-red se aceleró con nuevas maneras, se pasó de los encuentros y formatos exclusivamente presenciales a nuevas opciones en línea soportadas en datos. El ecosistema digital se fortaleció y sigue haciéndolo; la masificación del acceso a la tecnología también, en este sentido las cifras que registra el reporte anual de la Unión Internacional de Telecomunicaciones muestran que el 67% de la población mundial cuenta con internet, y es por ello que el desarrollo de las economías ya no solamente considera los insumos como el petróleo y otros recursos naturales sino los datos, las dinámicas de tráfico y circulación en la red, la necesidad de enrutadores de Wi-Fi y el acceso al uso de la nube para respaldos de información, salvaguarda de archivos o para trabajo en equipo.

De acuerdo con un estudio de la CEPAL del año 2020¹⁸ las tecnologías en el periodo de pandemia ayudaron a mitigar el impacto de las pérdidas económicas hasta en un 75% y a pesar de la brecha digital se observan avances en la cobertura tecnológica. En ese mismo documento se muestran indicadores de resiliencia en cuatro rubros: la inversión en infraestructura debido a que el espectro de telecomunicaciones se saturó de un día para otro y fue necesario hacerlo más robusto y flexible; la adaptabilidad en los hogares, especialmente por el hecho de que los niños tuvieron que tomar sus clases por internet y pasar mucho más tiempo en casa y, en el mismo tenor, la producción fue otro de los puntos que determinó ajustes en las políticas de Estado.

Aunque aún parece insuficiente, la inversión pública tanto en fibra óptica como en liberación de la cobertura, en ajustes para la regulación de servicios y hasta en materia de derechos humanos, se abre cada vez más para que la tecnología sea parte cotidiana de la vida, lo cual lleva a reflexionar profundamente en el manejo y acceso a la información como uno de los componentes fundamentales, puesto que impacta en los servicios de salud, educación, el trabajo y la economía.

Las profesiones bibliotecológicas y, en general las ciencias de la información, ya se replantean las estrategias de recopilación, curaduría, integración, recuperación y acceso que propicien la utilidad, seguridad y preservación de datos; experimentan innovadoras maneras de accionar ante las tecnologías, el internet de las cosas, la inteligencia artificial y la escuela en línea gracias a una capacidad que no se había visto antes.

Por ello, la formación de las actuales y las próximas generaciones de profesionales en el área han de tomar en cuenta los retos que plantea esta realidad, prepararse para estas tendencias, adaptar sus programas de estudio a las expectativas cambiantes y contribuir al desarrollo de habilidades necesarias para las plataformas de información del presente y futuro.

18 Naciones Unidas. CEPAL-CAF, “Las oportunidades de la digitalización en América Latina frente al covid 19” (7 de mayo de 2020), <https://repositorio.cepal.org/items/608db1ec-a680-452b-800b-8c6df582be0d>.

REFERENCIAS BIBLIOGRÁFICAS

- Castells, Manuel, *La sociedad red*, México: Siglo XXI, 2006.
- Consejo Internacional en Archivos e InterPARES, *Un marco de referencia para la preservación digital*, Traducido por Alicia Barnard, Alejandro Delgado y Juan Voutssás, Ciudad de México: Archivo General de la Nación, 2017 (Trabajo original publicado en 2012).
- Fei Fei, Li, *The Worlds I See*, New York: Flatiron Book, 2023, <https://itunes.apple.com/WebObjects/MZStore.woa/wa/viewBook?id=0>
- Fei Fei, Li, *Time 100/AI*, <https://time.com/collection/time100-ai/>
- García, Francisco Javier, “Lenguaje verbal y lenguaje no verbal: esta es la cuestión... también en la Gestión de la Atención al Usuario en Bibliotecas”, entrevista hecha por Francesc García Grimau, *Desiderata*, abril, mayo y junio de 2023, <https://www.auxiliardebiblioteca.com/wp-content/uploads/2023/04/Desiderata-22.pdf>
- García, Amelia, Idania Licea y Riselis Martínez, “Del desarrollo de colecciones a la gestión de colecciones digitales: mirada a la evolución histórica del término a partir de apuntes teóricos”, *Bibliotecas. Anales de Investigación* 19, n.º 3, 2023, 1-18.
- Library IQ, “Using Data Is Essential to Library Operations. Here’s Why”, *Library Journal*, 4 de diciembre de 2023, <https://www.libraryjournal.com/story/using-data-is-essential-to-library-operations.-heres-why-lj231204#:~:text=Using%20data%20can%20help%20librarians,data%20analysts%20in%20their%20career>
- Merril, Alex, “Library+”, *Public Services Quarterly* 7, n.º 3-4, 2011, 144-148.
- Naciones Unidas. CEPAL-CAF, *Las oportunidades de la digitalización en América Latina frente al COVID 19*, 7 de abril de 2020, <https://repositorio.cepal.org/items/608db1ec-a680-452b-800b-8c6df582be0d>
- Pérez, Marcos, *El ciclo de vida del dato*, España: Universitat Oberta de Catalunya, 2017.

- Prada, Ennio, “Los insumos invisibles de decisión: datos, información y Conocimiento”, *Anales de Documentación*, n.º 11, 2008, 183-196.
- Rodríguez, Ariel Alejandro, “Sobre la catalogación y la actividad de los metadatos aproximaciones a su coexistencia en la era del acceso”, *Biblioteca Universitaria* 11, n.º 1, 2008, 31-39.
- Rojas, Yuniét, “Organización de la información: un factor determinante en la gestión empresarial”, *ACIMED* 12, n.º 2, 2004, 1-12.
- The Future of Research Communications and e-Scholarship, *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version B1.0*, 2019, <https://www.force11.org/fairprinciples>
- Time, *Time 100/AI*, 2024, <https://time.com/collection/time100-ai/> artículo
- Uriarte, Luis Miguel y Manuel Acevedo, “Sociedad Red y transformación digital”, *Economía industrial*, n.º 407, 2018, 35-49.
- Weinberger, David, *Everything is Miscellaneous: The Power of the New Digital Disorder*, New York: Henry Holt and Company, 2007.

Repositorios institucionales, redes y organización del conocimiento: un estado de la cuestión en Hispanoamérica

Institutional repositories, networks and knowledge organization: a state of the art in Hispanic America

Gema Bueno-de-la-Fuente
María del Carmen Agustín-Lacruz
*Departamento de Ciencias de la Documentación e
Historia de la Ciencia
Universidad de Zaragoza*

1. INTRODUCCIÓN

Los repositorios institucionales son plataformas digitales diseñadas para depositar, recopilar, almacenar, preservar y proporcionar acceso a la producción científica y académica de una institución o de una red, con el propósito de facilitar la difusión abierta y gratuita del conocimiento y aumentar su visibilidad.

Surgieron a finales del siglo xx como una nueva estrategia de las instituciones académicas, con el objetivo de preservar y difundir colecciones de recursos producidos por ellas mismas, junto con los metadatos que los describen.¹ Fueron impulsados por el movimiento de acceso abierto a la ciencia, fundamentado en las declaraciones internacionales de Budapest (2002), Berlín (2003) y Bethesda (2003), para fomentar el acceso libre, sin restricción de tipo económico, técnico o legal a la información científica, académica, técnica y cultural.

1 Mariângela Spotti Lopes Fujita, Carmen Agustín-Lacruz, Jéssica Beatriz Tolare, Ana Lúcia Terra y Gema Bueno de la Fuente, "Institutional Repositories and Knowledge Organization: A Bibliographic Study from Library and Information Science", *Education for Information* 39, n.º 1, 2023, 51-66, <https://doi.org/10.3233/EFI-220015>

Los repositorios comenzaron a participar en el proceso de comunicación académica, desarrollando el papel de los editores² y mejorando la diseminación y difusión de la producción intelectual de sus instituciones. Estos sistemas de información están gestionados por bibliotecas universitarias o instituciones y centros de investigación, y son entornos de trabajo habitual para bibliotecarios y profesionales de la información. A su desempeño profesional se suman nuevos roles³ que requieren aplicar procesos y técnicas de organización del conocimiento tanto para organizar las colecciones como para describir temáticamente los recursos y objetos digitales.⁴ En la actualidad, los repositorios son imprescindibles para la gestión de la producción intelectual de una institución académica⁵ y se han convertido en infraestructuras esenciales y determinantes del avance científico.

En el ámbito hispanoamericano ha avanzado notablemente el cambio de paradigma en los modos de producir, evaluar y comunicar la ciencia, impulsado por el movimiento de ciencia abierta y el desarrollo de los repositorios y redes institucionales. En los últimos años se ha afianzado la idea del conocimiento como bien público y del acceso abierto gestionado por la comunidad académica sin fines de lucro, y los repositorios institucionales se han convertido en herramientas privilegiadas por la regulación legal y el establecimiento de políticas de acceso abierto a la ciencia.⁶

2 Ana Paula de Oliveira Villalobos y Fabio Andrade Gomes, “Análise dos repositórios das universidades federais brasileiras”, *Ponto de Acesso* 12, n.º 3, 2018, 126-144, <https://doi.org/10.9771/rpa.v12i3.27929>

3 Tatiana Sanches, “Bibliotecários do ensino superior: o desafio da atualização de competências” en XIV Jornadas APDIS-Be OPEN: Inspirar, Capacitar, Investigar, *Inovar & Partilhar* (Porto: apdis, 2020), 1-16, <http://hdl.handle.net/10451/45788>

4 Gema Bueno-de-la-Fuente, Carmen Agustín-Lacruz, Mariângela Spotti Lopes Fujita y Ana Lúcia Terra, “Competencias profesionales para la Organización del conocimiento en Repositorios Institucionales” en *Organización del conocimiento en el ámbito documental, patrimonial y social. Actas del VI Congreso ISKO España-Portugal*, 16.º ISKO España (Madrid: Universidad Complutense, 2024, en prensa).

5 Alexia Dini-Kounoudes y Marios Zervas, “Best practices and policies in institutional repositories development: The Ktisis case”, en *3rd International Conference on Qualitative and Quantitative Methods in Libraries*, Atenas, 2011, <https://ktisis.cut.ac.cy/bitstream/20.500.14279/4601/3/Best%20practices%20and%20policies%20in%20institutional%20repositories%20development.pdf>

6 Dominique Babini y Laura Rovelli, *Tendencias recientes en las políticas científicas de ciencia abierta y acceso abierto en Iberoamérica* (Buenos Aires: CLACSO; Fundación Carolina, 2020). <https://www.clacso.org/wp-content/uploads/2020/12/Ciencia-Abierta-1.pdf>

Como señala Babini,⁷ diversos países latinoamericanos han establecido políticas y normas de distinto rango, tanto leyes nacionales para regular el acceso abierto a las publicaciones y datos científicos (Perú y Argentina en 2013 y México en 2014); como declaraciones, manifiestos y propuestas legislativas en favor del acceso abierto (Brasil en 2007 y 2011); recomendaciones para el acceso y preservación de información científica y datos de investigación (Chile en 2014) y directrices sobre política de ciencia abierta (Colombia en 2019). Todas ellas han contribuido al crecimiento y fortalecimiento de los repositorios institucionales, como evidencia la información proporcionada por el Directorio de Repositorios de Acceso Abierto (en inglés, *Directory of Open Access Repositories*, OpenDOAR).

En este contexto, este capítulo ofrece una visión panorámica de las redes y repositorios institucionales en los países del área hispanoamericana y su ecosistema de normas, recomendaciones e iniciativas. Se plantea como un estado de la cuestión que sirva a los estudiantes y profesionales de Bibliotecología, Ciencias de la Información y la Documentación, para conocer y afrontar la labor de la organización del conocimiento (en inglés, *Knowledge Organization*, ko) en repositorios institucionales. Está estructurado en seis apartados dedicados a: conceptos generales; definición de objetivos y metodología del estudio; las redes nacionales de repositorios institucionales; las redes internacionales de repositorios institucionales; directrices sobre metadatos y organización del conocimiento en redes y repositorios institucionales en Hispanoamérica y conclusiones.

2. OBJETIVOS Y METODOLOGÍA

Este trabajo se propone analizar, ofrecer una visión panorámica general de las redes de repositorios institucionales en los países del área hispanoamericana, tanto desde su dimensión nacional como internacional, para identificar las características de cada uno de estos sistemas y su estado de desarrollo.

De forma específica, se estudian con detalle las recomendaciones sobre metadatos de descripción temática y organización del conocimiento recogidas en las directrices, guías y manuales que adoptan estas redes de repositorios en el área hispanoamericana.

7 Dominique Babini, “Toward a global open access scholarly communications system -a developing region perspective” en *Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access*, ed. Martin Paul Eve, Jonathan Gray (Cambridge, Massachusetts: MIT, 2020), <https://doi.org/10.7551/mitpress/11885.003.0033>

Se trata de un estudio exploratorio y descriptivo, con enfoque cualitativo. La colecta de datos ha incluido la revisión de documentos, declaraciones, normativa y bibliografía sobre el tema tanto a nivel internacional como regional y nacional. Se ha llevado a cabo en cuatro etapas:

1. Identificación y descripción de las redes nacionales de repositorios institucionales de los países de Hispanoamérica.
2. Análisis de las políticas y normativas (leyes, decretos, reglamentos) que impulsan los repositorios digitales de acceso abierto a la ciencia en cada uno de los países.
3. Identificación y descripción de las redes internacionales de repositorios y de sus directrices técnicas.
4. Análisis de las recomendaciones específicas sobre descripción temática y organización del conocimiento.

3. REDES NACIONALES DE REPOSITARIOS INSTITUCIONALES, UNA PANORÁMICA

En el área hispanoamericana, el desarrollo de repositorios y la adhesión a las políticas de ciencia abierta a través de políticas nacionales muestra un grado de implantación desigual (Babini y Rovelli 2020,⁸ Belli *et al.* 2023,⁹ Bueno-de-la-Fuente *et al.* 2023¹⁰ y Comisión Europea 2023¹¹). Este apartado describe la situación de los repositorios en los diferentes países, ordenados alfabéticamente, y recoge una síntesis general en la Tabla 1.

Argentina cuenta con la red del Sistema Nacional de Repositorios Digitales (SNRD), fruto de una iniciativa conjunta del Ministerio de Ciencia, Tecnología e Innovación y del Consejo Interinstitucional de Ciencia y Tecnología

8 Babini y Rovelli, *Tendencias recientes en las políticas científicas de ciencia abierta y acceso abierto en Iberoamérica*.

9 Simone Belli, Rubén Díez, María Luisa Ramos, Jorge Resina y Marta Romero Delgado, *Desafíos para las políticas de ciencia abierta en Iberoamérica: redes, repositorios y multilingüismo* (Madrid: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura, 2023), <https://oei.int/publicaciones/desafios-para-las-politicas-de-ciencia-abierta-en-iberoamerica-redes-repositorios-y-multilinguismo>.

10 Gema Bueno de la Fuente, Carmen Mariângela Spotti Lopes Fujita y Ana Lúcia Terra, “Knowledge organisation in institutional repositories: a case study on policies and procedures manuals in the Ibero-American environment”, *The Electronic Library*, 41, n.º 6 2023, 770-786, <https://doi.org/10.1108/EL-05-2023-0128>

11 Comisión Europea, Dirección General de Investigación e Innovación, *Políticas de acceso abierto en América Latina, el Caribe y la Unión Europea: avances para un diálogo político*, Oficina de Publicaciones de la Unión Europea, 2023, <https://data.europa.eu/doi/10.2777/162>

(CICYT). El objetivo del SNRD es conformar una red interoperable de repositorios digitales a partir del establecimiento de políticas, estándares y protocolos comunes. Agrupa a 48 repositorios de universidades e institutos de investigación. Las bases legales de la red se definieron en la *Ley n.º 26.899 de Repositorios digitales institucionales de acceso abierto*¹² de 2013, que establece la obligatoriedad de desarrollar repositorios digitales institucionales de acceso abierto, propios o compartidos, por parte de todos los organismos e instituciones públicas que componen el Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) que reciben financiación pública.

Bolivia por el momento carece de un sistema nacional de repositorios digitales, no obstante, participa en la *Red de Repositorios Latinoamericanos* con cinco instituciones.¹³

En *Chile*, el Repositorio de la Agencia Nacional de Investigación y Desarrollo (ANID) difunde y preserva la producción científica y los resultados de los proyectos financiados por ella. Surgió en 2012 y es el Nodo Nacional de Acceso y socio fundador de LA Referencia. En 2022 se publicó la *Política de Acceso Abierto de la ANID*,¹⁴ donde se recogen los principios rectores de las políticas públicas para garantizar la disponibilidad y acceso a las publicaciones generadas en proyectos financiados con fondos públicos.¹⁵

Colombia cuenta con la Red Colombiana de Información Científica (Red-Col), creada por el Ministerio de Ciencias, Tecnologías e Innovación (Minciencias) mediante resolución de 2019¹⁶ en la que se estableció como nodo central el Centro de Documentación y Biblioteca (CENDOC). Esta Red es responsable de consolidar, gestionar y visibilizar la información científica del país, así como el nodo Colombia para participar en proyectos de intercambio de información, productos, datos científicos y proyectos de Ciencia Abierta

12 Se puede consultar en https://repositoriosdigitales.mincyt.gov.ar/files/Boletin_Oficial_Ley_26899.pdf

13 Víctor Hugo Perales Miranda y Gabriela Amarilis Perales Miranda, “Repositorios y revistas científicas de acceso abierto en Bolivia”, *Educación Superior* 9, n.º 1, 2022, 49-60. Estos autores han realizado un estudio sobre diez universidades bolivianas que cuentan con repositorios institucionales, de un total de más de medio centenar, de las que solo tres están registradas en los directorios internacionales OpenDOAR y ROAR (*Registry of Open Access Repositories*).

14 Documento disponible en https://s3.amazonaws.com/documentos.anid.cl/estudios/Politica_acceso_a_informacion_cientifica_2022.pdf

15 Esta normativa se desarrolla al amparo de la Ley 20285, promulgada en 2008 por la Secretaría General de la Presidencia sobre acceso a la información pública en <https://www.bcn.cl/leychile/navegar?idNorma=276363>

16 Se trata de la Resolución 0166 del 20 de febrero de 2019, por la que se conforma la Red Colombiana de Información Científica, disponible en <https://minciencias.gov.co/normatividad/resolucion-0166-2019>

de carácter internacional, según establecieron los *Lineamientos para una política de Ciencia Abierta en Colombia*.¹⁷ La *Política Nacional de Ciencia Abierta 2022-2031* fue aprobada en 2022.¹⁸

Costa Rica cuenta con el Repositorio Nacional Kimuk, impulsado en 2016 por el Consejo Nacional de Rectores (CONARE) a través del Comité de Conocimiento Abierto. En su sitio, declaran contar con unas directrices propias, basadas en las de OpenAIRE (*Open Access Infrastructure for Research in Europe*), si bien su enlace no está disponible. El repositorio nacional costarricense Kimuk (cooperación, en lengua bribri) permite la asociación de otros repositorios institucionales y actualmente cosecha siete repositorios de universidades, centros de investigación y portales de revistas. Además de ser parte de LA Referencia, está asociado a la Confederación Internacional de Repositorios (COAR). Costa Rica no cuenta aún con una política de acceso abierto a nivel nacional, aunque el CONARE trabaja en su desarrollo.

Cuba cuenta con 16 repositorios en OpenDOAR y 2 agregadores, 16 repositorios en OpenAIRE Explore y ROAR y tan solo 11 en el portal de la Red Cubana de la Ciencia. Hasta el momento carece de mecanismos de coordinación nacional para el desarrollo de sus repositorios institucionales y tampoco cuenta con un agregador o repositorio nacional,¹⁹ aunque se han realizado análisis para su implementación.²⁰ Se han llevado a cabo iniciativas interesantes dentro del proyecto ELINF “*ICT Supporting the educational processes and the knowledge management in higher education*” en colaboración con el Consejo de Universidades Flamencas VLIR-UOS (Bélgica) de la que ha surgido un anteproyecto,²¹ en la actualidad pendiente de aprobación.

En Ecuador, la Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEDIA) es la entidad responsable de gestionar la Red de Repositorios de Acceso Abierto del Ecuador (RRAAE) y el Repositorio Ecuatoriano de Investigadores (REDI). La RRAAE es el agregador y buscador nacional,

17 Disponible en https://minciencias.gov.co/sites/default/files/upload/reglamentacion/resolucion_0167-2019.pdf

18 El documento se encuentra en https://minciencias.gov.co/sites/default/files/upload/reglamentacion/resolucion_0777-2022.pdf

19 Babini y Rovelli, *Tendencias recientes en las políticas científicas de ciencia abierta y acceso abierto en Iberoamérica*.

20 Mirelys Puerta-Díaz, Yohannis Martí-Lahera y Edgar Bisset Alvarez, “Federación de repositorios universitarios: realidad y retos en Cuba”, *Informação & Tecnologia* 5, n.º 2 2018, 4-23.

21 El *Anteproyecto de Estrategia y Política para la Ciencia y la Educación Abiertas en las instituciones del Ministerio de Educación Superior (MES) de la República de Cuba*, de 2019 está disponible en https://olinfer2023.vliired.cu/wp-content/uploads/2023/06/Anteproyecto_de_Estrategia_y_Politica_para_la_Educacion_y_la_Ciencia.pdf

implementado en 2015, aglutina 128 repositorios de 68 instituciones y actúa como nodo nacional de LA Referencia. Carece de política nacional de acceso abierto, aunque cuenta con iniciativas significativas recogidas en el *Código Orgánico de Economía Social del Conocimiento e Innovación*, aprobado en 2016²² y en el documento preliminar del *Plan de Ciencia, Tecnología, Innovación y Saberes Ancestrales* (CTIYSA)²³ de 2021, que contiene un apartado específico sobre ciencia abierta.

El Salvador cuenta con un agregador nacional de repositorios institucionales denominado Acceso a la Ciencia y la Cultura de El Salvador (ACCESS) que cosecha los fondos de los repositorios de siete universidades y del propio CBUES (Consortio de Bibliotecas Universitarias de El Salvador). Este consorcio y el Ministerio de Educación, Ciencia y Tecnología (MINED) son los responsables de su implementación. Además, cuenta con el Repositorio Digital de Ciencia y Cultura de El Salvador (REDICCES) (<http://www.redicces.org.sv/jspui/>) desarrollado por CBUES.

En *España* el Recolector de Ciencia Abierta (RECOLECTA) es la plataforma que agrupa a todas las infraestructuras digitales en las que se publican o depositan resultados de investigación en acceso abierto y el agregador nacional de repositorios de acceso abierto. Surgió en 2007 como resultado de la colaboración entre la Fundación Española para la Ciencia y la Tecnología (FECYT) y la Red de Bibliotecas Universitarias (REBIUN). Actualmente recolecta 183 repositorios, siendo el nodo nacional en LA Referencia. Dispone de una Guía para la evaluación de repositorios institucionales (2021),²⁴ dirigida a los responsables de la gestión de repositorios científicos. Cuenta con una Estrategia Nacional de Ciencia Abierta (ENCA) 2023-2027, establecida en 2023.²⁵

Guatemala y *Honduras* carecen de sistemas nacionales de repositorios y no forman parte de LA Referencia. Algunas de sus universidades guatemaltecas cuentan con repositorios institucionales y una de ellas está incluida en OpenDOAR, donde también se relacionan cinco repositorios institucionales hondureños.

22 Disponible en https://www.gob.ec/sites/default/files/regulations/2019-02/Documento_C%3%B3digo-Org%C3%A1nico-Econom%C3%ADa-Social-Conocimientos-Creatividad-Innovaci%C3%B3n.pdf

23 Elaborado por la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación se encuentra en https://www.bivica.org/files/5879_Plan%20Nacional%20SENESCYT.pdf

24 José Manuel Barreco Cruz, Pilar Rico-Castro y Laura Valeria Bonora Eve (coords.), *Guía para la evaluación de repositorios institucionales* (Madrid: Fundación Española para la Ciencia y la Tecnología; FECYT, 2021).

25 Elaborada por la Secretaría General Técnica del Ministerio de Ciencia e Innovación, está en https://www.recolecta.fecyt.es/sites/default/files/documents/Estrategia%20Nacional%20de%20Ciencia%20Abierta_0.pdf

México estableció la Red Mexicana de repositorios (REMEDI) en 2013, una red federada de repositorios institucionales y temáticos de las instituciones universitarias y de investigación, que actúa como nodo nacional en LA Referencia bajo la coordinación del Consejo Nacional de Ciencia y Tecnología (CONACYT), con el apoyo de la Corporación Universitaria para el Desarrollo de Internet (CUDI). Poco después, un decreto de 2014²⁶ modificó la *Ley de Ciencia y Tecnología*, añadiendo un capítulo sobre “Acceso Abierto, Acceso a la Información Científica, Tecnológica y de Innovación y del Repositorio Nacional”. Actualmente, México cuenta con 56 repositorios registrados en OpenDOAR, 48 de ellos en LA Referencia.

Nicaragua carece de una política nacional de repositorios, aunque algunas universidades sí las han establecido para sus repositorios.²⁷ El país cuenta con el Repositorio Universitario de Nicaragua,²⁸ promovido por el Consejo Nacional de Universidades (CNU) que recolecta registros de nueve repositorios. La Red Nacional de Gestión de la Información para el Desarrollo de la Investigación (REGIDI), del CNU, con su Consorcio Programa para el Fortalecimiento de la Información para la Investigación (peri-Nicaragua) está impulsando el desarrollo de una red nacional de repositorios institucionales, desarrollando el *Plan Operativo Anual (POA) 2023*.²⁹ En OpenDOAR se incluyen doce repositorios y agregadores, pero todavía no forma parte de LA Referencia.

Panamá cuenta con el Portal de Repositorios Institucionales de Ciencia, Tecnología e Innovación de Literatura Abierta (PRICILA), promovido por la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT) en 2018. En 2020 se aprobó el *Plan Estratégico Nacional de Ciencia, Tecnología e Innovación (PENCIYT) 2019-2024*³⁰ para promover la ciencia abierta. Es nodo de LA Referencia.

Paraguay carece de sistema nacional de repositorios digitales con tan sólo tres repositorios registrados en OpenDOAR y al menos otros diez no registrados. El repositorio institucional de la instancia superior de investigación,

26 Se trata del Decreto por el que se reforman y adicionan diversas disposiciones de la Ley de Ciencia y Tecnología, de la Ley General de Educación y de la Ley Orgánica del Consejo Nacional de Ciencia y Tecnología, disponible en <https://www.repositorionacionalecti.mx/docs/LCyT.pdf>

27 Es el caso de las Políticas del Repositorio Institucional de UNAM-Managua (RIUMA) <https://biblio.unan.edu.ni/index.php/2017/03/13/politicas-del-repositorio-institucional-unan-managua> o de la Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense (URACCAN) <https://www.uraccan.edu.ni/institucionalfile/politica-del-repositorio-institucional>

28 Se puede consultar en <https://repositorio.cnu.edu.ni/>

29 Está en https://cnu.edu.ni/wp-content/uploads/2024/01/Plan-Operativo-Anual_POA-2023.pdf.

30 En https://www.senacyt.gob.pa/wp-content/uploads/2020/01/GacetaNo_28936b_202001081-1.pdf

el Consejo Nacional de Ciencia y Tecnología (RI.CONACYT) promueve el depósito de los resultados de investigación financiados por sus programas. En 2023 se aprobó la *Ley n.º 7064, que crea el sistema nacional de investigadores (SISNI) y establece la carrera del investigador científico*³¹ en la que se promueve la creación de repositorios institucionales de acceso abierto y de los mecanismos que permitan conectarlos a nivel nacional e internacional.

Perú estableció su política nacional de acceso abierto en 2013 en la *Ley n.º 30035 que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto*³² y cuenta con la Red Nacional de Repositorios Digitales de Ciencia, Tecnología e Innovación de Acceso Abierto (RENARE), coordinada desde CONCYTEC (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica) y con el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación ALICIA (Acceso Libre a la Información Científica), que actúa como agregador nacional de la investigación financiada por el estado.³³ Cuenta desde 2019 con la *Guía Alicia versión 2.0. Directrices para repositorios institucionales de la Red Nacional de Repositorios Digitales de Ciencia, Tecnología e Innovación de Acceso Abierto* (RENARE).³⁴

Uruguay cuenta con SILO, nombre del sistema nacional y de la red de repositorios de acceso abierto de ciencia y tecnología. Fue creado en 2019, impulsado por la adhesión de la ANII (Agencia Nacional de Investigación e Innovación uruguaya) a LA Referencia, para promover el acceso, la preservación y la visibilidad de la producción de ciencia y tecnología uruguaya (Aguirre-Ligüera, Maldini y Fontans 2019³⁵ y Seroubian 2022³⁶). SILO, el nodo de Uruguay en esta red, reúne los resultados de investigación de los repositorios de acceso abierto de las instituciones adheridas -actualmente 13- y proporciona datos a agregadores regionales y globales mediante la adopción de políticas, estándares y protocolos comunes.

31 Disponible en <https://www.bacn.gov.py/archivos/11313/LEY+7064.pdf>

32 Está publicada en <https://portal.concytec.gob.pe/images/stories/images/2013/portal/areas-institucion/dsic/ley-30035.pdf>

33 Está regulado por una directiva de 2020, publicada en <https://cdn.www.gob.pe/uploads/document/file/1948054/DIRECTIVA%20N%C2%B0002-2021-CONCYTEC-P.pdf>

34 Publicada en https://cdn.www.gob.pe/uploads/document/file/747504/guia_ALICIA_version_2.0.pdf

35 Natalia Aguirre-Ligüera, Juan Maldini y Exequiel Fontans, “Acceso abierto a la producción científica de Uruguay: poca historia en 10 años (2009-2018)”, *Palabra Clave (La Plata)* 9, 1, 2019, <https://doi.org/10.24215/18539912e079>.

36 Mabel Seroubian, “Acceso abierto y ciencia abierta. Experiencia desde la gestión del repositorio institucional COLIBRI de la Universidad de la República”, *Informatio* 27, n.º 1, 2022, 284-305, <https://doi.org/10.35643/info.27.1.6>

Venezuela formó parte del proyecto de LA Referencia en sus fases iniciales, aunque no está actualmente en esta red.³⁷ Cuenta con 15 repositorios registrados en OpenAIRE, si bien carece de una red federada de repositorios o un agregador (Martínez-Guerrero y García Romero 2018).³⁸

En la Tabla 1 se recoge una relación completa de todos los países hispano-americanos, con el nombre y acrónimo de la red nacional de repositorios, su fecha de constitución, si forman parte de la red internacional LA Referencia, y citando sus documentos más relevantes (políticas, directrices, guías, planes o legislación).

País	Red y acrónimo	URL	Fecha	Documentos relevantes
Argentina	Sistema Nacional de Repositorios Digitales (SNRD)	http://repositoriosdigitales.mincyt.gob.ar	2013	<i>Directrices smrd. Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales (2015)</i>
Chile	Repositorio ANID	https://repositorio.anid.cl/about	2012	<i>Política de acceso abierto a la información científica y a datos de investigación financiados con fondos públicos de la anid (2022)</i>
Colombia	Red Colombiana de Información Científica (RedCol)	https://redcol.minciencias.gov.co	2019	<i>Lineamientos para una política de Ciencia Abierta en Colombia (2019); Política Nacional de Ciencia Abierta 2022-2031 (2022)</i>
Costa Rica	Repositorio Nacional Kimuk	https://kimuk.conare.ac.cr/	2016	

37 Así se recoge en <https://www.lareferencia.info/legacy/realizan-primeros-acuerdos-regionales-sobre-repositorios.html>

38 Christian Alexander Martínez-Guerrero y Marisol García Romero, “Estado de los repositorios institucionales venezolanos y sus revistas científicas”, *Información, Cultura y Sociedad* 38, 2018, 89-106, <https://doi.org/10.34096/ics.i38.3731>

País	Red y acrónimo	URL	Fecha	Documentos relevantes
Ecuador	Red de Repositorio de Acceso Abierto del Ecuador (RRAAE)	https://rraae.cedia.edu.ec/	2015	
El Salvador	Acceso a la Ciencia y la Cultura de El Salvador (ACCES)	http://www.acces.org.sv/vufind		<i>Políticas de uso del cosechador nacional ACCES Acceso a la Ciencia y Cultura de El Salvador: Propuesta (2021)</i>
España	Recolector de Ciencia Abierta (RECOLECTA)	https://www.recolecta.fecyt.e	2007	<i>Guía para la evaluación de repositorios institucionales (2021); Estrategia Nacional de Ciencia Abierta (ENCA) 2023-2027</i>
México	Red Mexicana de repositorios (REMERI)	http://www.remeri.org.mx/portal/index.html	2013	<i>Ley de Ciencia y Tecnología, con un capítulo dedicado al Acceso Abierto, Acceso a la Información Científica, Tecnológica y de Innovación y del Repositorio Nacional (2014)</i>
Nicaragua				<i>Plan Operativo Anual (POA) 2023 del Subsistema de Educación Superior (2023)</i>
Panamá	Portal de Repositorios Institucionales de Ciencia, Tecnología e Innovación de Literatura Abierta (PRICILA)	http://pricila.senacyt.gob.pa/vufind/	2018	<i>Plan Estratégico Nacional de Ciencia, Tecnología e Innovación (PENCIYT) 2019-2024 (2020)</i>

País	Red y acrónimo	URL	Fecha	Documentos relevantes
Perú	Acceso Libre a la Información Científica (ALICIA)	https://alicia.concytec.gob.pe/	2016	<i>Directrices para repositorios institucionales de la Red Nacional de Repositorios Digitales de Ciencia,</i> <i>Tecnología e Innovación de Acceso Abierto (RENARE) (2019)</i>
Uruguay	Repositorios abiertos de Ciencia y Tecnología (SILO)	https://silu.uy/vufind/	2019	

Tabla 1. Relación alfabética de países hispanoamericanos y sus redes nacionales de repositorios institucionales (Fuente: Elaboración propia).

4. REDES INTERNACIONALES DE REPOSITARIOS

En el ámbito latinoamericano se han desarrollado, en los últimos años, varias redes internacionales de repositorios con distinto alcance, orientación y composición.

Destaca LA Referencia (Red Latinoamericana para la Ciencia Abierta),³⁹ auspiciada en 2010 en el marco del proyecto “Estrategia Regional y Marco de Interoperabilidad y Gestión para una Red Federada Latinoamericana de Repositorios Institucionales de Documentación Científica”, liderado por organismos de ciencia, tecnología e innovación de América Latina, el Caribe y España. Con el apoyo de la Cooperación Latinoamericana de Redes Avanzadas (redCLARA),⁴⁰ el proyecto fue financiado por el Banco Interamericano de Desarrollo (BID) y apoyado por la Red Comunidad Latinoamericana de Bibliotecas y Repositorios Digitales (CoLaBoRa), en la actualidad desaparecida.

LA Referencia se fundó formalmente en 2012 con el *Acuerdo de Cooperación entre Altas autoridades de Ciencia, Tecnología e Innovación de América Latina para la constitución de LA Referencia*.⁴¹ En su constitución formaron parte ocho socios nacionales: Argentina, Brasil, Chile, Colombia, Ecuador, México, Perú y Venezuela, a los que posteriormente se sumaron El Salvador, Costa

39 Su página web se puede consultar en <https://www.lareferencia.info>

40 Una amplia información está disponible en <https://www.redclara.net/es>

41 El texto del acuerdo se puede consultar en <https://www.lareferencia.info/legacy/sites/default/files/AcuerdoBuenosAires.pdf>

Rica, Panamá, Uruguay y España, mientras que Venezuela, que fue socio original, no lo es actualmente.

Desde 2014, cada país ha gestionado, a su ritmo, una infraestructura centralizada de repositorios que, a su vez, escala a otra plataforma regional. Esto ha implicado avances en el modelo de la red, tanto desarrollando buenas prácticas en la gestión de repositorios para hacerlos interoperables a nivel nacional y regional, como implementando un buscador regional de publicaciones científicas⁴² (LA Referencia, 2014).

Un mapa de los países integrantes (Figura 1) y un balance global de todos los logros alcanzados en sus diez años de vida está recogido en la *Declaración 10 años de LA Referencia: Hacia un ecosistema de Ciencia Abierta no comercial*.⁴³



Figura 1. Mapa de países latinoamericanos integrantes de la referencia (Fuente: <https://www.lareferencia.info/es/institucional/quienes-somos>).

También es interesante el portal de la Red de Repositorios Latinoamericanos⁴⁴ surgido en 2006 por iniciativa de la Universidad de Chile a través de su Dirección de Servicios de Información y Bibliotecas (SISIB), que agrupa más de 500 instituciones de 21 países de América Latina y El Caribe. Las fechas

42 Una información detallada sobre estos aspectos se puede leer en un boletín especial, disponible en <https://dev.lareferencia.info/wp-content/uploads/2023/05/edicion-especial12.pdf>

43 La noticia completa está publicada en <https://www.lareferencia.info/es/blog-masonry/noticias/item/307-declaracion-10-la-referencia>

44 Información procedente de <https://repositorioslatinoamericanos.uchile.cl>

del pie de página, 2006-2018, señalan que ha dejado de actualizar su mantenimiento.⁴⁵



Figura 2. Repositorio Centroamericano SIIDCA-CSUCA (Fuente: <https://repositoriosiidca.csuca.org>).

En el ámbito centroamericano existe una iniciativa federada constituida como Red SIIDCA-CSUCA (ver Figura 2). Se estructura en torno al Repositorio Centroamericano SIIDCA-CSUCA,⁴⁶ un descubridor lanzado en mayo de 2016 (Velía Gómez y Correa 2021)⁴⁷ que agrupa 17 repositorios de universidades centroamericanas (Costa Rica, Nicaragua, Honduras, Panamá y Guatemala) y cuenta con el apoyo del Consejo Universitario Centroamericano (CSUCA). El objetivo de este portal es difundir el conocimiento científico generado en Centroamérica, lo mantiene la Universidad Nacional Agraria de Nicaragua.

45 No se cuenta con políticas específicas para la participación en el agregador más que los tipos de contenidos y el uso del protocolo OAI-PMH, aunque se indica que los metadatos recolectados serán validados y los participantes pueden recibir observaciones sobre metadatos u otros aspectos técnicos.

46 Ofrece una interesante información en <https://repositoriosiidca.csuca.org>

47 Ruth Velía Gómez y Ernesto Correa, "Repositorio Centroamericano: La experiencia de un trabajo colaborativo", 2021, https://eulacfoundation.org/sites/default/files/attachments/presencacionrepositorio13abril2021_0.pdf

5. DIRECTRICES SOBRE METADATOS Y ORGANIZACIÓN DEL CONOCIMIENTO EN REDES Y REPOSITARIOS INSTITUCIONALES

Al igual que las bibliotecas digitales, los repositorios institucionales se ocupan de la gestión de la información y necesitan organizar el conocimiento mediante técnicas de representación y recuperación de la información.⁴⁸ Mientras que las bibliotecas están orientadas al proceso técnico de las colecciones –selección bibliográfica, catalogación y servicio de referencia– los repositorios se centran en la preservación digital, la disponibilidad y la accesibilidad. Se diferencian de los catálogos de bibliotecas porque proporcionan acceso al propio recurso digital junto a los metadatos estructurados que lo describen y, al orientarse al acceso a los objetos, los repositorios prestan menos atención a los aspectos de organización del conocimiento en sus colecciones de recursos.

Aunque originalmente respondían a necesidades diferentes, en la actualidad bibliotecas académicas y repositorios coexisten y suelen compartir funciones y estructuras administrativas. Ambos sistemas de información tienen estrategias comunes en materia de interoperabilidad y adopción de normas y comparten procesos básicos de organización del conocimiento, como la atribución de metadatos para apoyar el acceso a las colecciones.⁴⁹

Además de su carácter disciplinar, la organización del conocimiento ($\kappa\omicron$) tiene una orientación aplicada incuestionable, como precisa Hjørland.⁵⁰ Para este autor, la $\kappa\omicron$ abarca tanto los procesos como la indización, el etiquetado, la clasificación, la descripción y la organización de documentos e información; como los propios sistemas de organización del conocimiento, como los tesauros, los sistemas de clasificación, las ontologías y las taxonomías. Por lo tanto, los dos aspectos principales abordados por esta disciplina son los procesos de organización del conocimiento y los sistemas de organización del conocimiento.⁵¹

48 Birger Hjørland, “Information Retrieval and Knowledge Organization: A Perspective from the Philosophy of Science”, *Information* 12, n.º 3, 2021, 135, <https://doi.org/10.3390/info12030135>

49 Bueno-de-la-Fuente *et al.*, “Knowledge organisation in institutional repositories: a case study on policies and procedures manuals in the Ibero-American environment”.

50 Birger Hjørland, “Information Retrieval and Knowledge Organization: A Perspective from the Philosophy of Science”.

51 Birger Hjørland, “Knowledge Organization ($\kappa\omicron$)”, *Knowledge Organization* 43, n.º 6, 2016, 475-484, <http://dx.doi.org/10.5771/0943-7444-2016-6-475>

5.1. Políticas y directrices sobre descripción temática en las redes de repositorios internacionales

En este apartado se estudian las directrices internacionales seguidas -declaradas expresamente o no- por las redes de repositorios en relación a los metadatos temáticos, pues son estos metadatos las principales herramientas de las que disponen los repositorios en la actualidad para la organización conceptual del conocimiento.

Desde sus inicios en 2012,⁵² los acuerdos técnicos de LA Referencia adoptaron las directrices DRIVER 2.0 (*Digital Repository Infrastructure Vision for European Research*) de 2008. Estas directrices establecen el uso de Dublin Core no cualificado, “oai_dc” y el elemento de metadatos “dc:subject” es considerado obligatorio si procede (MA- Mandatory if Applicable).

Además, las directrices de DRIVER 2.0⁵³ señalaban que los metadatos de OAI-PMH (en inglés, *Open Archives Initiative Protocol for Metadata Harvesting*) permitían indicar la clasificación de materias o sistema de encabezamiento utilizado, generalmente en el elemento “dc:subject”. Esta información de clasificación se solía utilizar para agrupar ítems de un repositorio por áreas disciplinares y, por tanto, tal información aparecía con frecuencia en el elemento OAI “setSpec”. Como esta forma de codificación generaba problemas a los recolectores, se recomendaba transportar la información del elemento “dc:description” de la respuesta “Identify” al campo de materia; es decir, “dc:subject” y que estuviese “URI-ficada”,⁵⁴ utilizando un espacio de nombres de autoridad para respaldar el reconocimiento del esquema de clasificación. Así mismo, las directrices establecían que, si los términos se toman de un esquema de clasificación estándar, se codificase cada término en un elemento separado.

En el campo “dc:subject” indicaban que se pueden recoger palabras clave o códigos de clasificación. Establecen que los términos controlados pueden

52 El acta del acuerdo del comité técnico, correspondiente a la 5.ª Reunión LA Referencia, sobre el Plan Piloto Componente III, celebrado en Bogotá, el 27 y 28 de agosto de 2012, está disponible en <https://www.lareferencia.info/legacy/sites/default/files/ActaAcuerdoTecnicoBogota.pdf>

53 Estas directrices, denominadas *Digital Repository Infrastructure Vision for European Research. DRIVER Guidelines 2.0. Guidelines for content providers Exposing textual resources with OAI-PMH*, se publicaron en 2008 y están disponibles en: <https://wiki.surfnet.nl/display/driverguidelines/Home>

54 Este neologismo en español está relacionado con el acrónimo inglés URI (identificador uniforme de recurso). Su significado hace referencia a que el valor del elemento de metadatos es una URI que enlaza con el término en el espacio de nombres del vocabulario controlado.

proceder de tesauros, listas de encabezamiento de materia y de los números de clasificación.

En ese momento, los esquemas de clasificación más frecuentes en el contexto OAI eran la *Clasificación Decimal Dewey* (DDC, en inglés), *Clasificación Decimal Universal* (UDC), *Clasificación de la Biblioteca del Congreso* (LCC) y entre los sistemas de encabezamientos de materia, los *Encabezamientos de Materia de la Biblioteca del Congreso* (LCSH), los *Schlagwortnormdatei* (SWD, en alemán). También se empleaban códigos de clasificación de áreas disciplinares como *Mathematics Subject Classification* (MSC) y *Medical Subject Headings* (MESH) y otros sistemas de clasificación local. Si no se utilizaba ningún esquema de clasificación específico, DRIVER recomendaba por defecto la *Clasificación Decimal Dewey* (DDC).

Classification Schemes

Namespace	Type	Term	Description
info:eu-repo	/ classification	/ ddc	Dewey Decimal Classification (DDC) http://www.oclc.org/dewey/
info:eu-repo	/ classification	/ udc	Universal Decimal Classification http://www.udcc.org/http://www.udcc.org/
info:eu-repo	/ classification	/ lcc	Library of Congress Classification http://www.loc.gov/catdir/cpsol/lcco/http://www.loc.gov/catdir/cpsol/lcco/
info:eu-repo	/ classification	/ lcsch	Library of Congress Subject Headings (LCSH)
info:eu-repo	/ classification	/ swd	Schlagwortnormdatei (SWD)
info:eu-repo	/ classification	/ bicssc	Book Industry Communication - Standard Subject Categories http://www.bic.org.uk/7/BIC-Standard-Subject-Categories/
info:eu-repo	/ classification	/ nbc	Nederlandse Basisclassificatie (Dutch Basic Classification) http://www.kb.nl/vak/basis/bc04.pdf
info:eu-repo	/ classification	/ jel	Journal of Economic Literature (JEL) Classification System Articles in economics journals are usually classified according to the JEL classification codes. JEL code (sub)categories, including periodic updates, are referenced at Journal of Economic Literature (JEL) Classification System, published quarterly by the American Economic Association (AEA) http://www.aeaweb.org/journal/jel_class_system.php

Figura 3. “Classification Schemes”. INFO-EU: repo, Application profiles (Fuente: Standards wiki of surffoundation, <https://wiki.surfnw.nl/display/standards/info-eu-repo#infoeu-repo-ClassificationSchemes>).

Para registrar los metadatos de los esquemas de clasificación se empleó el perfil de aplicación “info:eu-repo”. En la Figura 3 se muestra el listado de diversos esquemas de clasificación en la columna de descripción tal y como aparecen en la wiki de estándares de *SURF Foundation*,⁵⁵ en las que aún se alojan las directrices de DRIVER.

Estas directrices de DRIVER se fundieron en 2012 con las de OpenAIRE, el proyecto que lo continuó y se constituyó formalmente en 2018 como organización. OpenAIRE gestiona una infraestructura europea que ofrece un

55 Esta interesante wiki se encuentra disponible en <https://wiki.surfnw.nl/display/standards/info-eu-repo#infoeu-repo-ClassificationSchemes>

conjunto diverso de servicios públicos, para acelerar la adopción de la Ciencia Abierta, y cuenta con el apoyo de una red de expertos ubicados en organizaciones nacionales clave de todos los países europeos, los *National Open Access Desks*.⁵⁶

Las *Directrices de OpenAIRE para los administradores de repositorios institucionales y temáticos*, en su versión 4.0⁵⁷ propone un perfil de aplicación de metadatos basado en los esquemas de “dc” (Dublin Core simple), “dcterms” (términos de Dublin Core), “datacite” (Figura 4), y “oaire”. En la elaboración de la versión en español participaron activamente los miembros de LA Referencia.

```
1 <datacite:subject>
2 <datacite:subject>Earth sciences and geology</datacite:subject>
3 <datacite:subject subjectScheme="DDC" schemeURI="http://dewey.info/" valueURI="">
4 551 Geology, hydrology, meteorology
5 </datacite:subject>
6 </datacite:subjects>
```

Figura 4. Ejemplo de codificación de materias en el elemento “datacite:subject” (Fuente: Guía de OpenAIRE 4.0, https://guiasopenaire4.readthedocs.io/es/latest/field_subject.html#dci-subject).

En 2015, LA Referencia publicó sus propias directrices de metadatos y políticas de cosecha,⁵⁸ indicando las recomendaciones para todos los campos presentes en DRIVER, incluyendo las precisiones y cambios que se derivan de OpenAIRE, y aprobando el plan para adoptar las directrices de OpenAIRE.⁵⁹

En la tabla resumen de compatibilidad de las distintas directrices empleadas por LA Referencia,⁶⁰ el elemento de metadatos “dc:subject” ha pasado de ser recomendado en sus primeras directrices de 2013-14, a ser “obligatorio si es aplicable” en 2015, en consonancia con las directrices de DRIVER y OpenAIRE 3.0.⁶¹

Las directrices de metadatos de LA Referencia (2015) desarrollan de manera extensa la información sobre el elemento de metadatos materia. Indican en su definición que este elemento recoge la temática del recurso que se expresa

56 Se puede encontrar información en el apartado “About us” de la web de OpenAIRE, disponible en <https://www.openaire.eu/about>

57 Estas Directrices, publicadas en 2018, están disponibles en <https://guiasopenaire4.readthedocs.io/es/latest/#directrices-de-openaire-para-administradores-de-repositorios-de-literatura-v4>

58 La Referencia, *Metadatos y políticas de cosecha*, 2015, publicada en <https://recursos.lareferencia.info/download/metadatos-y-politicas-de-cosecha-interoperables-para-los-nodos-nacionales/#>

59 “Latinoamérica interoperable con los estándares internacionales de acceso abierto”, LA Referencia, accedido 20 mayo, 2024, <https://www.lareferencia.info/es/servicios/directrices>.

60 “Guidelines Compatibility Overview”, Recursos, LA Referencia, junio de 2015, <https://recursos.lareferencia.info/download/guidelines-compatibility-overview/>

61 Las *OpenAIRE Guidelines for Literature Repositories v 3*, se encuentran disponibles en https://guidelines.openaire.eu/en/latest/literature/index_guidelines_lit_v3.html

típicamente en palabras clave, descriptores y/o códigos de clasificación que describen el contenido intelectual del recurso.

En las instrucciones del uso del elemento señalaban que, siguiendo Dublin Core, es posible tener dos tipos de valores que describen el contenido (p. ej. palabra clave y/o código de clasificación). Cuando ambos valores están disponibles se recomienda registrarlos en distintas ocurrencias. Para las palabras clave o términos del lenguaje natural (no controladas por un vocabulario o tesauro) y/o descriptores (términos controlados por un vocabulario o tesauro), se aconseja repetir el elemento tantas veces como sea necesario, poniendo primero los términos en el idioma original y luego en el/los idioma/s en que se encuentren disponibles.

En cuanto a los esquemas de codificación siguen las mismas instrucciones establecidas por DRIVER y OpenAIRE 3.0, mencionando el uso de tesauros, listas de encabezamiento de materia, etcétera, para términos controlados; y los esquemas de clasificación más comunes CDD, CDU, LCC, etcétera, para los números de clasificación.

En general, si no hay un esquema de clasificación en uso (muy común para repositorios disciplinarios o propuestas de países), recogen la recomendación de algunas entidades en cuanto al uso de Dewey. Las directrices también reflejan que algunos países de la región utilizan Frascati de la OECD (*Organisation for Economic Cooperation and Development*) como vocabulario controlado (OECD 2018).⁶² En cualquier caso, LA Referencia no establece una norma al respecto, considerándose una decisión nacional.

Por otra parte, la organización del conocimiento mediante todo tipo de vocabularios controlados y ontologías es una de las piezas que sustentan la web semántica. Ya desde 2012,⁶³ LA Referencia apostó por participar en ella gracias a un convenio firmado por el Observatorio Iberoamericano de la Ciencia, la Tecnología y la Sociedad de la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OCTS-OEI) y la redCLARA, a través de la herramienta Intelligo.⁶⁴ Es un explorador semántico de documentos científicos y tecnológicos de repositorios iberoamericanos, donde los usuarios pueden experimentar una ruta de búsqueda no sólo para acceder a los documentos finales, sino también mediante mapas temáticos que agregan contexto a las

62 OECD, *Manual de Frascati 2015: Guía para la recopilación y presentación de información sobre la investigación y el desarrollo experimental* (OECD, París; FECYT, Madrid: 2018), <https://doi.org/10.1787/9789264310681-es>

63 La noticia se publicó el 2 de julio de 2014 en <https://observatorioocts.oei.org.ar/2014/07/02/llego-la-web-semantica-a-la-referencia-a-traves-de-intelligo/>

64 La información sobre esta herramienta se encuentra en <https://www.explora-intelligo.info/>

búsquedas y que se basan en los metadatos asociados a los documentos (véase Figura 5).

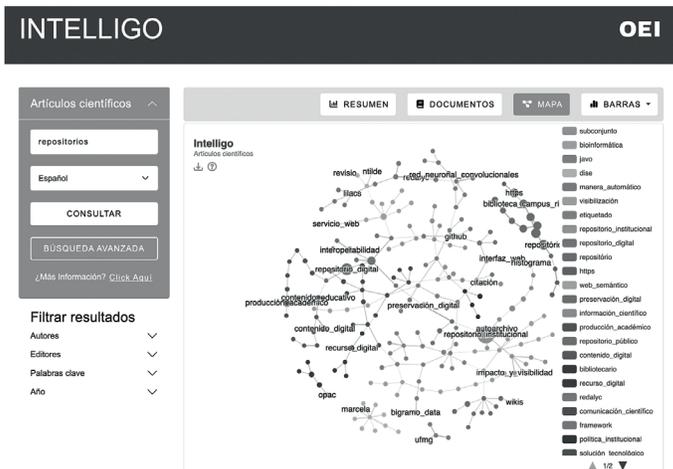


Figura 5. Mapa de conceptos de Intelligo, búsqueda por “repositorios” (Fuente: <https://www.explora-intelligo.info/repositorios/es?search=repositorios>).

Por su parte, OpenAIRE implementó en 2012 su grafo de conocimiento, una forma avanzada de organizar y representar dicho conocimiento.⁶⁵

Un grafo de conocimiento es una forma de conectar y unificar de manera significativa la información y hacerla interrogable de manera natural para las personas, con el fin de lograr una web más inteligente. Se suelen usar para enlazar entidades (objetos reales o abstractos) entre sí mediante propiedades semánticas. En el grafo OpenAIRE, estas entidades son resultados y productos de la investigación, como artículos, conjuntos de datos y *software*, y las relaciones entre estas entidades representan datos informativos como citas, financiación y colaboraciones.

Actualmente, el grafo de OpenAIRE es uno de los grafos del conocimiento más grande, en comparación con los de las bases de datos de *Scopus*, *WOS* o el buscador *Google Scholar*.⁶⁶ Se emplea en *OpenAIRE Explorer* para la búsqueda y navegación de contenidos (ver Figura 6). En el conjunto de datos que alimenta el grafo también se ha integrado una taxonomía de campos del conocimiento, denominada *Field-of-Science* (FoS) (Kotitsas, Pappas, Manola

65 El grafo de OpenAIRE está disponible en <https://graph.openaire.eu/>

66 “What is the OpenAIRE Graph?” OpenAIRE Graph <https://graph.openaire.eu/what-is-the-openaire-graph>

y Papageorgiou 2023),⁶⁷ basada en la clasificación *FORD* (*Fields of research and development*) de la OECD y en los campos del conocimiento de la clasificación de revistas de *SCIENCE-METRIX*.⁶⁸

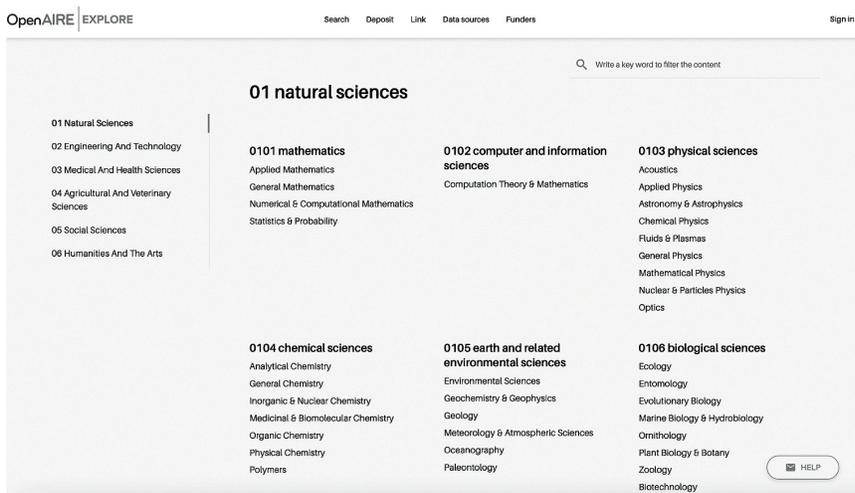


Figura 6. Taxonomía de campos de conocimiento, *Fields of Science*, OpenAIRE Explore (Fuente: <https://explore.openaire.eu/fields-of-science>).

El buscador avanzado de OpenAIRE Explore también ofrece la posibilidad de filtrar los resultados de búsqueda de productos de investigación por campo de conocimiento, empleando dicha clasificación (Figura 7).

67 Sotiris Kotitsas, Dimitris Pappas, Natalia Manola y Haris Papageorgiou, “SCINOBO: a novel system classifying scholarly communication in a dynamically constructed hierarchical Field-of-Science taxonomy”, *Frontiers in Research Metrics and Analytics* 8, 2023, <https://doi.org/10.3389/frma.2023.1149834>

68 Es una entidad internacional, propiedad de Elsevier, dedicada a la evaluación de políticas y actividades de ciencia y tecnología, utilizando métodos bibliométricos. Información disponible en <https://science-metrix.com/classification/>

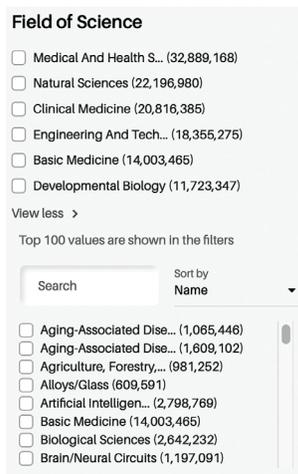


Figura 7. Filtro de búsqueda por *Fields of Science*, OpenAIRE Explore Search (Fuente: <https://explore.openaire.eu/search/find/>).

5.2. Políticas y documentos sobre descripción temática en repositorios nacionales

Los países que participan en LA Referencia han adoptado de forma generalizada también sus directrices de metadatos. Por otra parte, al adherirse esta red a OpenAIRE, las respectivas redes nacionales también han adoptado paulatinamente sus directrices. Así, aquellos que cuentan con sus propias prescripciones técnicas de metadatos para la organización y descripción temática de los repositorios, siguen lo dictado por LA Referencia y las directrices de alcance internacional, como OpenAIRE.⁶⁹ En los siguientes párrafos se analizan algunas de estas iniciativas.

En 2015 aparecieron en Argentina las *Directrices SNRD. Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales*,⁷⁰ como norma que garantiza la interoperabilidad nacional e internacional de los repositorios digitales institucionales argentinos. Sirven como orientación para que los administradores de los repositorios puedan definir e implementar políticas lo-

69 En casos como el de Ecuador no se han localizado directrices propias de la red de repositorios rraae, aunque en su página web <https://traae.cedia.edu.ec/Content/about> indican que para formar parte de la red las instituciones deben comprometerse a cumplir con la estandarización de sus metadatos a las directrices internacionales. Declaran que siguen el protocolo OAI-PMH y las directrices de Driver 2.0 y OpenAire 3.0.

70 Disponibles en https://repositoriosdigitales.mincyt.gob.ar/files/Directrices_SNRD_2015.pdf

cales de administración de datos acordes al Sistema Nacional de Repositorios Digitales (SNRD) y para proporcionar estándares internacionales, facilitando su inclusión en redes internacionales de repositorios.

El actual esquema de metadatos o perfil de aplicación propuesto en las *Directrices SNDR* sigue el modelo establecido por LA Referencia, por lo que considera el elemento de los metadatos de materia “dc:subject”, como obligatorio en el caso de ser aplicable. En sus pautas no establece vocabularios controlados para este campo, aunque sí para el elemento de los descriptores de cobertura “dc.coverage”, empleando el *Getty Thesaurus of Geographic Names Online* (TGN),⁷¹ como ya lo hicieron las directrices de DRIVER 2.0. No obstante, muchos repositorios incluyen en sus metadatos de materia las *Disciplinas FORD*⁷² para la clasificación temática de las colecciones, enlazando con su versión en español disponible en la web de la Biblioteca Electrónica de Ciencia y Tecnología del Ministerio de Ciencia, Tecnología e Innovación de Argentina.

En cuanto a Chile, desde 2024, el proyecto Nodo Nacional de Acceso cuenta con unas *Directrices de metadatos y mecanismos de interoperabilidad*,⁷³ en las que se recogen los lineamientos generales sobre el uso de metadatos para la descripción de los objetos digitales en los repositorios. En ellas se establecen una serie de elementos y su vocabulario asociado, basándose en las *OpenAIRE Guidelines for Literature Repository Managers*, v. 4.0, el esquema de metadatos Dublin Core, el protocolo de intercambio OAI-PMH y el vocabulario COAR. En particular, el elemento de metadatos temáticos “Materia” se basa en el elemento “datacite:subject”.

En Colombia, en 2020 se aprobaron las *Directrices para repositorios institucionales de investigación de la Red Colombiana de Investigación Científica*.⁷⁴ Más que un texto cerrado, constituyen un documento dinámico a modo de propuesta en constante evolución, que se retroalimenta y ajusta según las necesidades de la comunidad académica colombiana. Su objetivo es crear lineamientos para adecuar los datos de la producción científica a las infraestructuras establecidas por las instituciones y se fundamentan en estándares internacio-

71 Las versiones de las Directrices SNRD de 2013 se basaron en las “Directrices DRIVER 2.0”, habiéndose analizado también las Directrices OpenAIRE 1.1, las Directrices de la Biblioteca Digital Colombiana “Modelo de Interoperabilidad para BDCOL” y las “Directrices LUCIS/MODS” a propuesta de la Biblioteca Digital de la Universidad Nacional de Cuyo.

72 “Disciplinas FORD”. Biblioteca Electrónica de Ciencia y Tecnología, Ministerio de Ciencia, Tecnología e Innovación, disponibles en <https://biblioteca.mincyt.gov.ar/ford/>

73 La Agencia Nacional de Investigación y Desarrollo, dependiente del Gobierno de Chile, ha elaborado el “Proyecto nodo nacional de acceso. Directrices de Metadatos y Mecanismos de Interoperabilidad”, disponible en https://acceso-abierto.anid.cl/wp-content/uploads/sites/4/2024/05/Metadatos_para_la_Interoperabilidad_de_los_Repositorios_2024.pdf

74 Se pueden consultar en <https://redcol.readthedocs.io/es/latest/index.html#>

nales de calidad de datos. Se pretende que estas directrices puedan ser utilizadas en diferentes fases, para incorporar a los repositorios los resultados obtenidos por los grupos de investigación de sus actividades de generación de nuevo conocimiento; de desarrollo tecnológico e innovación; de apropiación social del conocimiento y de Formación de Recursos Humanos en ciencia, tecnología e innovación. Estas directrices declaran el seguimiento de las reglas de catalogación RDA (*Research Description and Access*) y presentan un perfil de aplicación de metadatos propio. El perfil toma elementos de los esquemas de OpenAIRE (oaire), Dublin Core, y DataCite, adaptado del esquema de metadatos *DataCite Metadata Schema*, versión 4.2.⁷⁵ En particular, se establece el uso del elemento “datacite:subject” para los metadatos temáticos. Aunque no se establece como obligatorio, se aporta mucha información de cómo emplear este campo, en cuanto a sus propiedades y atributos, de forma incluso más detallada que en las directrices de OpenAIRE. En los atributos se puede indicar el vocabulario normalizado al que pertenecen las materias seleccionadas, y las URI (*Uniform Resource Identifier*) tanto de los vocabularios controlados como de los descriptores. Se aportan ejemplos y menciones a vocabularios controlados.

Cuba carece de una red de repositorios nacional y, por tanto, tampoco dispone de directrices técnicas de metadatos y cosecha. No obstante, al menos uno de sus repositorios, el Repositorio Nínive de la Universidad de Moa, sí que realiza recomendaciones sobre metadatos temáticos en sus *Políticas de uso del Repositorio, Política de metadatos*.⁷⁶ Incluso indica tesauros y otros repertorios léxicos para el control de términos de indización en su Anexo 1.

En cuanto a México, en sus primeras directrices de 2013⁷⁷ establecía la necesidad del seguimiento de las directrices DRIVER 2.0. En noviembre de 2015 se publicaron sus *Lineamientos técnicos para el repositorio nacional y los repositorios institucionales*,⁷⁸ que en su esquema de metadatos indican que se seguirán las directrices OpenAIRE 3.0 y la síntesis de metadatos de LA Referencia. De esta manera, en este caso también el elemento “dc:subject” es recomendado y simplemente se indica “obligatorio si es aplicable”. Siguiendo las directrices

75 Está disponible en <https://doi.org/10.5438/bmjt-bx77>

76 Políticas de uso del Repositorio Institucional del UMOA están publicadas en <https://ninive.ismm.edu.cu/page/about#Anexo%202>

77 Requisitos Técnicos y Directrices de Interoperabilidad para REMERI. <http://www.remeri.org.mx/portal/documentos.html#prettyPhoto%5Biframe%5D/2013-doc-05/>

78 Publicados por CONACYT en 2015, los *Lineamientos técnicos para el repositorio nacional y los repositorios institucionales* se pueden consultar en <https://www.sicyt.gob.mx/index.php/normatividad/conacyt-normatividad/conacyt/1499-lineamientos-tecnicos-para-el-repositorio-nacional-y-los-repositorios-institucionales/file>

de OpenAIRE se recomienda el uso de vocabularios controlados como los encabezamientos *LCSH* o los *MESH*.

Un caso destacado es el de *Nicaragua*, con su Repositorio Universitario de Nicaragua que actúa de agregador de la red nacional. Su sistema de búsqueda avanzada por materias utiliza un listado combinado de términos procedentes de la Clasificación Decimal Dewey, códigos de la Clasificación Médica, Encabezamientos de materia precoordinados, y el sistema de Clasificación Dewey-AGRIS, basado en el tesaurus AGROVOC (ver Figura 8).

Existen otras opciones de navegación o búsqueda por materias en determinados repositorios institucionales, empleando Dewey con extensiones disciplinares adaptadas a la temática de cada repositorio, con combinaciones con el Sistema de Clasificación Médica, o el Tesaurus de Ingeniería Sanitaria y Ambiental.

Repositorio Universitario de Nicaragua

INICIO ACERCA DE REPOSITARIOS IDIOMA

Buscar / Avanzado

Consejos de búsqueda

Ayuda Búsqueda Avanzada

Ayuda Operadores de Búsqueda

Búsqueda Avanzada

Buscar: [] [Todos los Campos] [Coincide: [Todos los Términos]]

⊕ Agregar Campos de Búsqueda

⊕ Agregar Grupos de Búsqueda

Buscar Limpiar

Limitar

Institución:	Formato:	Materia:	Idioma:
Bluefields Indian & Caribbean U	Animation	658.87 Gerencia	Alemán
Universidad Americana	Artículo	664 Tecnología de alimentos	Español
Universidad Católica Redempto	Artículo	720 Arquitectura	Inglés
Universidad de Ciencias Coner	Artículo de conferencia o taller	CONOCIMIENTOS, ACTITUDES	Otro
Universidad de las Regiones Au	Audiom	DERECHO COMPARADO	
Universidad Internacional Antor	Book chapter	EDUCACIÓN SECUNDARIA	
Universidad Nacional Agraria	Experiment	F01 Cultivo	
Universidad Nacional Autonom	Image	F04 Fertilización	
Universidad Nacional Autonom	Imagen	F08 Arreglo y sistemas de culti	
Universidad Nacional de Ingeni	Learning Object	F30 Genética vegetal y fitomejoramiento	

Figura 8. Formulario de búsqueda avanzada del Repositorio Universitario de Nicaragua, destacando las opciones de límites por Materia (Fuente: <https://repositorio.cnu.edu.ni/Search/Advanced>).

En el nodo de *Panamá*, para participar en PRICILA, el Repositorio Nacional de Publicaciones desarrolló en 2019 las *Directrices técnicas de metadatos para los repositorios institucionales de acceso abierto para su cosecha por el repositorio nacional*.⁷⁹ Este documento de nuevo se basa en las directrices OpenAIRE 3.0

⁷⁹ La Secretaría Nacional de Ciencia, Tecnología e Innovación del Gobierno de Panamá las publicó en 2019 en <http://pricila.senacyt.gob.pa/vufind/themes/root/images/Lineamientos%20t%C3%A9cnicos%20para%20el%20repositorio%20nacional%20y%20los%20repositorios%20institucionales%201.pdf>

para repositorios documentales y establece las mismas recomendaciones en cuanto a metadatos temáticos y uso de vocabularios controlados, o la clasificación Frascati de la OCDE.

Paraguay ha publicado en 2024 sus *Políticas del repositorio institucional del CONACYT*, y en su política de metadatos sólo indican que se emplee Dublin Core cualificado sin detallar los elementos.⁸⁰

En Perú se elaboraron en 2021 las *Directrices para repositorios institucionales de la Red Nacional de Repositorios Digitales de Ciencia, Tecnología e Innovación de Acceso Abierto (RENARE) [Guía Alicia]*. Definen un perfil de aplicación para que los repositorios sean recolectados, basado en las directrices de OpenAIRE 3.0. Establecen como obligatorio el elemento de metadatos “dc.subject”, para recoger las palabras clave o descriptores del contenido. Además, incluyen un elemento de metadatos para los campos del conocimiento conforme al vocabulario de *Campos de la Investigación y el Desarrollo* (FORD) de la OCDE,⁸¹ disponible en SKOS (ver Figura 9).

↑	5.08.03 -- Bibliotecología
URI	https://purl.org/pe-repo/ocde/ford#5.08.03
Preferred label (es)	Bibliotecología
Preferred label (en)	Library science
Broader	https://purl.org/pe-repo/ocde/ford#5.08.00

Figura 9. Concepto 5.08.03-Bibliotecología, en la Clasificación de Campos de la Investigación y el Desarrollo OCDE (Fuente: <http://purl.org/pe-repo/ocde/ford>).

Por último, El Salvador, en su propuesta de *Políticas de uso del cosechador nacional ACCES Acceso a la Ciencia y Cultura de El Salvador*⁸² de 2021, también se basan en las directrices de LA Referencia. Incluyen las mismas indicaciones en cuanto al elemento “dc:subject” y al uso de vocabularios controlados, no estableciendo una norma al respecto de un vocabulario en concreto y considerándose una decisión de cada repositorio.

80 Las Políticas del Repositorio institucional del CONACYT se han publicado en https://www.conacyt.gov.py/sites/default/files/upload_editores/u489/Res-120-2024-Politicar-RI-conacyt.pdf

81 Están disponibles en https://conocimiento.concytec.gob.pe/vocabulario/ocde_ford.html.

82 *Políticas de uso del cosechador nacional ACCES Acceso a la Ciencia y Cultura de El Salvador: Propuesta*, Consorcio de Bibliotecas Universitarias de El Salvador están disponibles en <http://www.redicces.org.sv/jspui/handle/10972/4382>

6. CONCLUSIONES

Esta investigación ha proporcionado una visión panorámica general de las redes de repositorios institucionales en los países del área hispanoamericana, tanto desde su dimensión nacional como internacional, y ha permitido conocer el origen, las características y el desarrollo de cada uno de estos sistemas. Además, se han podido analizar con detalle las indicaciones sobre metadatos temáticos recogidas en las directrices, guías y manuales, adoptados por estas redes.

Por otra parte, ha sido notable el impulso que diferentes acuerdos gubernamentales regionales y políticas nacionales de acceso abierto han proporcionado al desarrollo de repositorios que se han mostrado, así como las herramientas más adecuadas para cumplir con las políticas y la legislación de acceso abierto aprobadas en varios países.

Las evidencias disponibles corroboran que existe un vasto número de repositorio, redes, agregadores y recolectores en el ámbito latinoamericano: El directorio OpenDOAR registra 789 repositorios de acceso abierto. La Red LA Referencia congrega a 12 nodos nacionales y casi 5 millones de documentos; Recolecta, el agregador de Ciencia Abierta en España, es uno de los nodos de la red más grande, con 183 repositorios y más de 1.600.000 documentos, y la Red de Repositorios Latinoamericanos, que aglutina a más de 500 instituciones y 8 millones de documentos. Y se trata tan sólo de algunos de los muchos ejemplos posibles.

Por lo que respecta a las directrices sobre metadatos y organización del conocimiento en redes y repositorios institucionales es destacable el impulso proporcionado por las *Directrices de LA Referencia* y de *OpenAIRE*, han desarrollado de manera extensa la información sobre los metadatos temáticos, que se expresa típicamente en palabras clave, descriptores y/o códigos de clasificación para describir el contenido de los recursos.

Sin embargo, todavía quedan retos pendientes, como extender y normalizar el uso de vocabularios de materias en la descripción de los recursos científicos para ampliar la disponibilidad del conocimiento abierto, entendido como bien social y contribuir a la construcción de la web semántica, aprovechando al máximo todas las posibilidades que la organización del conocimiento proporciona.

REFERENCIAS BIBLIOGRÁFICAS

- Aguirre-Ligüera, Natalia, Juan Maldini y Exequiel Fontans, “Acceso abierto a la producción científica de Uruguay: poca historia en 10 años (2009-2018)”, *Palabra Clave (La Plata)* 9, 1, 2019, <https://doi.org/10.24215/18539912e079>
- Babini, Dominique, “Toward a global open access scholarly communications system -a developing region perspective” en *Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access*, ed. Martin Paul Eve, Jonathan Gray, Cambridge, Massachusetts: MIT, 2020, <https://doi.org/10.7551/mitpress/11885.003.0033>
- Babini, Dominique y Laura Rovelli, *Tendencias recientes en las políticas científicas de ciencia abierta y acceso abierto en Iberoamérica*, Buenos Aires: CLACSO; Fundación Carolina, 2020, <https://www.clacso.org/wp-content/uploads/2020/12/Ciencia-Abierta-1.pdf>
- Barrueco Cruz, José Manuel, Pilar Rico-Castro y Laura Valeria Bonora Eve (coords.), *Guía para la evaluación de repositorios institucionales*. Madrid: Fundación Española para la Ciencia y la Tecnología (FECYT), 2021, https://calidadrevistas.fecyt.es/sites/default/files/informes/2021guiaevaluacionrecolecta_vf.pdf
- Belli, Simone, Rubén Díez, María Luisa Ramos, Jorge Resina y Marta Romero Delgado, *Desafíos para las políticas de ciencia abierta en Iberoamérica: redes, repositorios y multilingüismo*, Madrid: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura, 2023, <https://oei.int/publicaciones/desafios-para-las-politicas-de-ciencia-abierta-en-iberoamerica-redes-repositorios-y-multilinguismo>
- Bueno-de-la-Fuente, Gema, Carmen Agustín-Lacruz, Mariângela Spotti Lopes Fujita y Ana Lúcia Terra, “Knowledge organisation in institutional repositories: a case study on policies and procedures manuals in the Ibero-American environment”, *The Electronic Library* 41, n.º 6, 2023, 770-786, <https://doi.org/10.1108/EL-05-2023-0128>
- Bueno-de-la-Fuente, Gema, Carmen Agustín-Lacruz, Mariângela Spotti Lopes Fujita y Ana Lúcia Terra, “Competencias profesionales para la Organización del conocimiento...

- to en Repositorios Institucionales” en *Organización del conocimiento en el ámbito documental, patrimonial y social. Actas del VI Congreso ISKO España-Portugal (16.º ISKO España)*, Madrid: Universidad Complutense, 2024, en prensa.
- Comisión Europea, Dirección General de Investigación e Innovación, *Políticas de acceso abierto en América Latina, el Caribe y la Unión Europea: avances para un diálogo político*, Oficina de Publicaciones de la Unión Europea, 2023, <https://data.europa.eu/doi/10.2777/162>
- Dini-Kounoudes, Alexia y Marios Zervas, “Best practices and policies in institutional repositories development: The Ktisis case”, en *3rd International Conference on Qualitative and Quantitative Methods in Libraries*, Atenas, 2011, <https://ktisis.cut.ac.cy/bitstream/20.500.14279/4601/3/Best%20practices%20and%20policies%20in%20institutional%20repositories%20development.pdf>
- Fujita, Mariângela Spotti Lopes, Carmen Agustín-Lacruz, Jéssica Beatriz Tolare, Ana Lúcia Terra y Gema Bueno-de-la-Fuente, “Institutional Repositories and Knowledge Organization: A Bibliographic Study from Library and Information Science”, *Education for Information* 39, n.º 1, 2023, 51-66, <https://doi.org/10.3233/EFI-220015>
- Hjørland, Birger, “Knowledge Organization (ko)”, *KO Knowledge Organization* 43, 6, 2016, 475-484, <http://dx.doi.org/10.5771/0943-7444-2016-6-475>
- Hjørland, Birger, “Information Retrieval and Knowledge Organization: A Perspective from the Philosophy of Science”, *Information* 12, n. 3, 2021, 135, <https://doi.org/10.3390/info12030135>
- Kotitsas, Sotiris, Dimitris Pappas, Natalia Manola y Haris Papageorgiou, “SCINOBO: a novel system classifying scholarly communication in a dynamically constructed hierarchical Field-of-Science taxonomy”, *Frontiers in Research Metrics and Analytics* 8, 2023, <https://doi.org/10.3389/frma.2023.1149834>
- Martínez-Guerrero, Christian Alexander y Marisol García Romero, “Estado de los repositorios institucionales venezolanos y sus revistas científicas”, *Información, Cultura y Sociedad* 38, 2018, 89-106, <https://doi.org/10.34096/ics.i38.3731>

- OECD, *Manual de Frascati 2015: Guía para la recopilación y presentación de información sobre la investigación y el desarrollo experimental*, OECD, París; FECYT, Madrid: 2018, <https://doi.org/10.1787/9789264310681-es>
- Perales Miranda, Víctor Hugo y Gabriela Amarilis Perales Miranda, “Repositorios y revistas científicas de acceso abierto en Bolivia”, *Educación Superior* 9, n.º 1, 2022, 49-60, http://www.scielo.org.bo/pdf/especies/v9n1/v9n1_a06.pdf
- Puerta-Díaz, Mirelys, Yohannis Martí-Lahera y Edgar Bisset Álvarez, “Federación de repositorios universitarios: realidad y retos en Cuba”, *Informação & Tecnologia* 5, n.º 2, 2018, 4-23, <https://doi.org/10.22478/ufpb.2358-3908.2018v5n2.44415>
- Sanches, Tatiana, “Bibliotecários do ensino superior: o desafio da atualização de competências” en *XIV Jornadas APDIS-Be OPEN: Inspirar, Capacitar, Investigar, Inovar & Partilhar*, Porto: APDIS, 2020, 1-16, https://repositorio.ul.pt/bitstream/10451/45788/1/2020%20APDIS_Sanches_Bibliotec%3%a1rios%20do%20ensino%20superior.pdf
- Seroubian, Mabel, “Acceso abierto y ciencia abierta. Experiencia desde la gestión del repositorio institucional COLIBRI de la Universidad de la República”, *Informatio* 27, n.º 1, 2022, 284-305, <https://doi.org/10.35643/info.27.1.6>
- Velia Gómez, Ruth y Ernesto Correa, “Repositorio Centroamericano: La experiencia de un trabajo colaborativo”, 2021, https://eulacfoundation.org/sites/default/files/attachments/presentacionrepositorio13abril2021_0.pdf
- Villalobos, Ana Paula de Oliveira y Fabio Andrade Gomes, “Análise dos repositórios das universidades federais brasileiras”, *PontodeAcesso* 12, n.º 3, 2018, 126-144, <https://doi.org/10.9771/rpa.v12i3.27929>

IV.
ORGANIZACIÓN DEL CONOCIMIENTO DIGITAL

Mapas tópicos y ontologías terminológicas: características y tendencias en la organización temática de colecciones digitales

*Topical maps and terminological ontologies:
characteristics and trends in the organization
of digital collections*

Adriana Suárez Sánchez
*Instituto de Investigaciones Bibliotecológicas y de la Información
Universidad Nacional Autónoma de México*

INTRODUCCIÓN

El universo de los sistemas de organización del conocimiento integra variados instrumentos: listados de temas, taxonomías, encabezamientos de materia, clasificaciones bibliográficas, entre otros. Recientemente, derivado de las tecnologías de la información y comunicación, han emergido sistemas como mapas tópicos (MT) y ontologías terminológicas (OT) que integran los preceptos esenciales de las prácticas bibliotecarias (denominación, categorización, relaciones) y consideran aspectos tecnológicos relacionados con la web semántica, el universo de los metadatos, la hipervinculación como fundamento de exploración en el ciberespacio y los datos enlazados.

Los mapas tópicos y las ontologías terminológicas, en comparación con los encabezamientos de materia o los tesauros que nacieron en el seno de la bibliotecología, fueron proyectados en las ciencias computacionales. Históricamente, las ontologías emergieron en las ciencias informáticas en la década de 1980 con el claro propósito de ser herramientas para simbolizar dominios de conocimiento de manera comprensible para los agentes artificiales; mientras que, los mapas tópicos en el año 2000 se asentaron como un lenguaje de indicación idóneo para el espacio digital, bajo la premisa de tópicos que organizan temáticamente un dominio y direccionan hacia recursos de información.

Si bien las raíces de los mapas tópicos y las ontologías terminológicas se anclan en las ciencias computacionales, en la última década se han expandido a múltiples áreas de saber que representan, organizan y recuperan información. Tal es el caso de la bibliotecología donde se señala que ofrecen grandes ventajas en la organización temática de dominios de conocimiento y sus recursos vinculados. Tomando en cuenta tales afirmaciones, el presente capítulo analiza los mapas tópicos y las ontologías terminológicas desde la perspectiva de la organización temática de dominios de conocimiento altamente especializados, por ejemplo: anatomía de himenópteros, arqueología islámica, ciencias de los alimentos, endometriosis, enfermedades neurológicas, vasijas de cerámica china, etcétera.

Con fines de presentación del tema, el capítulo se divide en cuatro secciones: 1) se caracterizan los MT y OT en el marco de los sistemas de organización de conocimiento; 2) se analiza su definición, propósito, estructura, tecnologías y estándares de construcción; 3) se esboza un análisis comparativo entre ambos sistemas y, 4) se establecen sus ventajas/desventajas en la organización temática de dominios de conocimiento.

MAPAS TÓPICOS Y ONTOLOGÍAS TERMINOLÓGICAS EN EL UNIVERSO DE LOS SISTEMAS DE ORGANIZACIÓN DEL CONOCIMIENTO

Los sistemas de organización del conocimiento (SOC) [*KOS-Knowledge Organization Systems*] son un conjunto de estructuras para gestionar la información y promover el conocimiento.¹ Entre ellos es posible mencionar glosarios, diccionarios, tesauros, taxonomías, listas de temas, encabezamientos de materia, mapas tópicos, ontologías, entre otros.

Además de ser instrumentos para gestionar el conocimiento, los SOC comparten ciertas características: hacen referencia al saber especializado, entrañan funciones representadoras del universo cognoscible, poseen una estructura coherente-lógica, se fundamentan en preceptos lingüísticos, identifican las entidades de un dominio, evidencian las relaciones entre entidades, son referentes en contextos académicos, científicos o sociales y favorecen la búsqueda/recuperación de información.

De igual forma, como los SOC poseen rasgos comunes es posible identificar elementos diferenciadores. Mientras un grupo (encabezamientos de materia, clasificaciones bibliográficas y tesauros) floreció asociado a los recursos im-

1 Gail Hodge, *Systems of knowledge for digital libraries: beyond traditional authority files* (Washington: Council on Library and Information Resources, 2000).

presos y la biblioteca física, los mapas tópicos y las ontologías terminológicas surgieron asociadas a la web. En cuanto a cobertura temática, algunos sistemas (encabezamientos de materia y clasificaciones bibliográficas) integran un panorama amplio del universo de conocimiento mientras que mapas tópicos y ontologías terminológicas tratan segmentos de conocimiento reducidos. Por último, mientras unos actúan sólo como vocabularios controlados, a partir de etiquetas lingüísticas, los mapas tópicos y las ontologías terminológicas posibilitan el intercambio de información, integran una sintaxis especial e incorporan altos niveles semánticos, con el claro objetivo de ser comprensibles tanto para humanos como para las máquinas.

Según los diferentes rasgos que caracterizan a los SOC, Zeng² habla de: listas de términos, modelos basados en metadatos, categorizaciones y modelos de relación. Orbst³ prefiere agruparlos según su capacidad significativa: las listas de temas y los encabezamientos de materia con semántica baja, las taxonomías y las clasificaciones con semántica media, los tesauros con semántica avanzada y los mapas tópicos y ontologías terminológicas con semántica alta. Bergman⁴ prefiere una tipología según las tecnologías que emplean: Procesadores de texto/HTML- Lenguaje de Marcas de Hipertexto, técnicas empleadas por listas de temas; UML- Lenguaje Unificado de Modelado/XMI-Intercambio de Metadatos, para la codificación de vocabularios controlados; XML- Lenguaje de Marcado Extensible/XLST- Transformaciones XSL, aplicable en la estructuración de tesauros y taxonomías; RDF- Marco de Descripción de Recursos/XML para mapas tópicos y RDF/OWL- Lenguaje de Ontologías Web ontologías terminológicas. Un último aspecto de los SOC, que ha sido poco mencionado, es su forma de despliegue donde es posible diferenciar: listas alfabéticas, estructuras jerárquicas, modelos semánticos y redes gráficas.

El amplio espectro de rasgos señalados permite que los SOC sean útiles en variadas aplicaciones. Si bien uno de sus propósitos es ser herramientas auxiliares en la recuperación de información -mejorando la búsqueda, desambiguación, expansión o reformulación de consultas- otros usos contemplan representar dominios de conocimiento, servir como rutas o mapas semánticos y apoyar el esclarecimiento terminológico-conceptual.

2 Marcia Lei Zeng, "Knowledge organization systems (KOS)", *Knowledge Organization* 35, n.º 2-3, (Enero, 2008), 160-182.

3 Leo Orbst, "Introduction to ontologies and semantic technologies", consultado abril 22, 2024, https://stids.c4i.gmu.edu/STIDS2011/presentations/STIDS2011_Tutorial_Parts1-2.pdf

4 Michael Bergman, "An intrepid guide to ontologies," consultado Noviembre 14, 2023, <https://www.mkbergman.com/374/an-intrepid-guide-to-ontologies/>

A la luz del marco diferenciador de los SOC antes esbozado, los mapas tópicos y las ontologías terminológicas pueden ser caracterizadas del modo siguiente (Figura 1):

Mapas tópicos	Ontologías terminológicas
<ul style="list-style-type: none">• Asociados a la web• Tratan dominios de conocimiento reducidos• Modelos de relación• Semántica alta• RDF• XML• Redes gráficas• Plantillas tópicas para la recuperación de recursos	<ul style="list-style-type: none">• Asociados a la web• Tratan dominios de conocimiento reducidos• Modelos de relación• Semántica alta• RDF• OWL• Redes gráficas• Plantillas tópicas para la recuperación de recursos (no es su función primordial, pero pueden actuar como tal)• Sistemas para el razonamiento artificial

Figura 1: Mapas tópicos y ontologías terminológicas-Características (Fuente: Elaboración propia, 2024).

Ambos sistemas son modelos de relación, esto implica que mientras algunos SOC operan a nivel terminológico (listas de temas o encabezamientos de materia) o a nivel categorial (sistemas de clasificación), los mapas tópicos y las ontologías terminológicas poseen niveles terminológicos, categoriales y relacionales. En el segundo rubro encontramos que ambos son esquemas con semántica alta, lo que implica gran potencial para definir cada uno de sus elementos y contextualizarlo mediante relaciones jerárquicas, causales, todo/parte, materia prima/producto, acción/agente, etcétera.

En cuanto a tecnologías, ambos sistemas emplean como base RDF-Marco de Descripción de Recursos. En los modelos relacionales con visualización gráfica, las tripletas Sujeto-Predicado-Objeto constituyen la sintaxis que permite marcar y vincular las entidades. También en el rubro tecnológico se observa que los mapas tópicos emplean codificación XML mientras que las ontologías emplean OWL, un lenguaje de expresión con mayor capacidad. Por último, ambos sistemas son redes gráficas, motivo por el que han sido denominados SOCR-Sistemas de Organización del Conocimiento en Red (*SKOS-Networked Knowledge Organization Systems*).

MAPAS TÓPICOS

Según *ISO/IEC 13250:2000-Tecnologías de la información-SGML Aplicaciones-Mapas tópicos*,⁵ la designación mapa tópico entraña dos acepciones:

- Estándar: es una norma para la gestión distribuida de la información y el conocimiento.
- Plantilla/Colección documental: es un conjunto de objetos documentales limitados, cuyo núcleo central es una plantilla tópica afianzada en XML.

Aunque la norma ISO/IEC 2000 es el documento esencial sobre los mapas tópicos, algunos autores han esbozado definiciones y proposiciones alternativas. Biezunski⁶ establece que son una tecnología para la representación y el intercambio de conocimiento, una definición formal de temas, asociaciones y ocurrencias. García Martul⁷ coincide al respecto cuando asume que “son una solución de continuidad al paradigma de la web semántica, al permitir la navegación contextual en línea y la recuperación entre sistemas de información distribuidos e interoperables.” Miller, Nilanka, Song, Zhu y Hua⁸ apuntan que los mapas tópicos poseen la capacidad de organizar y recuperar información, ya que ofrecen modelos y fundamentos sólidos para organizar de manera semántica redes de conexiones, convirtiéndose así en una alternativa valiosa para la organización y exploración de grandes colecciones de documentos en constante expansión.

A modo de síntesis, es posible afirmar que los mapas tópicos son un sistema de organización del conocimiento con rasgos terminológicos, conceptuales y relacionales. Han sido propuestos como mapas visuales que integran las denominaciones temáticas de un dominio de conocimiento, por ejemplo: “Enfermedades neuronales= “Enfermedad de Alzheimer” / “Enfermedad de Parkinson” / “Esclerosis Múltiple” / “Enfermedad de Huntington” / “Esclerosis Lateral Amiotrófica”. En agregación, resuelven problemas de heterogeneidad

5 Organización Internacional de Normalización, *ISO/IEC 13250:2000-Tecnologías de la información-SGML Aplicaciones-Mapas tópicos* (Ginebra, Suiza: ISO, 2000).

6 Michel Biezunski, “Topic maps and the essence of indexing”, *The Indexer* 36, n.º 4 (Enero, 2018), 157-161.

7 David García Martul y Guillermina Franco Álvarez, “¿Tesauros, mapas conceptuales o topic maps para la Biblioteca Digital Agustín Millares?”, *Boletín Millares Carlo* 28 (Enero, 2009) 150.

8 Les Miller, Sree Nilakanta, Yunan Song, Lei Zhu y Ming Hua, “Managing knowledge in organizational memory using topic maps”, *International Journal of Knowledge Management* 4, n.º 1, enero 2008, 17-28.

semántica, posibilitan compartir datos y son una potente herramienta para navegar en colecciones de recursos de información asociados con un dominio.

Los mapas tópicos nacieron como un lenguaje de indización. Al principio fueron empleados en proyectos editoriales, como esquemas temáticos con vinculación a recursos (libros) sobre un tópico establecido. Luego, se expandieron hacia contextos electrónicos (bases de datos) y, finalmente, a la gran web, espacio en el que mantienen sus principios esenciales: ser esquemas temáticos hipervinculados hacia recursos de información.

Su estructura ha sido definida en los elementos ulteriores:

- **Tópicos:** son datos de naturaleza temática (un concepto, un lugar, una persona, etcétera). Constituyen etiquetas para describir el contenido temático de un dominio que posee una colección de recursos digitales.
- **Asociaciones:** son vinculaciones entre los tópicos. Pueden ser jerárquicas o facetadas.
- **Ocurrencias:** son los recursos de información relacionados con un tópico, se direccionan mediante identificadores en la web (URI- Identificador de Recursos Uniforme, URL- Localizador de Recursos Uniforme, URN- Nombre de Recurso Uniforme). Establecer una relación entre tópico y ocurrencia implica que el recurso trata sobre el asunto establecido por el tópico.

En comparación con algunos SOC que suelen operar con exclusividad en la bibliotecología, los mapas tópicos poseen un espectro de acción más vasto. Se ha probado su utilidad y aplicación en portales semánticos, aprendizaje electrónico, modelado de procesos de negocios, estructuras para el ordenamiento de productos, esquemas para la integración de la información, en funciones de metadatos, como rutas de portales web, en función de mapas de Intranets, como plantillas de interoperabilidad temática entre sistemas diversos, como estructuras para la navegación y lenguajes de indización en unidades informativas (bibliotecas, archivos, museos, etcétera).

ONTOLOGÍAS TERMINOLÓGICAS

Las ontologías son un marco común o una estructura conceptual útil para organizar información con miras a su posterior recuperación.⁹ En cuanto a las

9 Thomas R. Gruber, "Ontology", consultado febrero 17, 2024, <http://tomgruber.org/writing/ontology-definition-2007.htm>

ontologías terminológicas deben ser entendidas como una colección de conceptos, arreglados en una jerarquía de categorías, combinadas con los atributos y las relaciones entre tales conceptos, de modo que reflejen el vocabulario de un área. Una ontología terminológica abarca preceptos del control del vocabulario, la esquematización categorial, la semántica relacional y la presentación de propiedades para cada clases o instancias. Con una estructura robusta, “una ontología funciona como una plantilla de trabajo que permite el almacenamiento, la búsqueda y el razonamiento basado en instancias y reglas.”¹⁰

Pieterse y Kourie¹¹ señalan dos aspectos que diferencian las ontologías de otros SOC. Primero, son representaciones basadas en formalismos y, segundo, sus axiomas o reglas de inferencia que facilitan el razonamiento de los sistemas recuperadores de información y otros agentes artificiales. La estructura y razonamiento de las ontologías excede la de cualquier otro SOC, de ahí que sean aplicadas como lenguajes de indización y, al mismo tiempo, operen como bases de datos para el esclarecimiento terminológico y el razonamiento de los sistemas recuperadores de información.

A inicios del presente siglo, López Alonso¹² indicaba que los lenguajes documentales tenían limitaciones: no intervenían directamente en la interpretación de los conceptos y se esperaba que los sistemas futuros pudieran manejar el conocimiento y la información de manera más dinámica, utilizando una semántica más amplia de objetos y sus relaciones. Estos sistemas se han materializado gracias a las tecnologías actuales y, en este tono, las ontologías terminológicas son consideradas lenguajes documentales para organizar temáticamente la información y los recursos hipermedia, debido a que tienen la capacidad de servir como vocabularios controlados y sistemas de clasificación virtuales.

En lo tocante a su estructura, las ontologías contemplan los siguientes elementos:

- Entidades: algo que “es” de una manera determinada, puede ser que exista, haya existido, se crea que puede existir o forme parte del imaginario colectivo. Son referentes que con fines comunicativos se concretan en signos de alguna naturaleza (icónica, lingüística, etcétera).

10 *Op. cit.* Zeng: 161.

11 Vreda Pieterse y Derrick G. Kourie, “Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle”, *Knowledge Organization* 41, n.º 3 (Enero 2014): 217-229.

12 Miguel Ángel López Alonso, “Las estructuras conceptuales de representación del conocimiento en Internet”, *Scire: representación y organización del conocimiento* 6, n.º 1 (junio 2000): 107-123.

- **Conceptos:** son la esencia de las ontologías terminológicas. Constituyen unidades de pensamiento que remiten a una entidad. Se concretan en términos que los simbolizan. Aunque la esencia de las ontologías son los conceptos, su concreción está asociada al uso del vocabulario, en una relación indisoluble equivalente a la dicotomía significante/significado que está presente en el lenguaje.
- **Clases:** son conjuntos de elementos que comparten características similares. Estas permiten segmentar campos de objetos, personas, disciplinas, actividades, personalidades, entre otros. En los sistemas de clasificación, las categorías son las divisiones que componen el conjunto del sistema.
- **Instancias:** son cada una de las entidades particulares presentes en un dominio de conocimiento. Son identificables a partir de sus características únicas que las hacen ser de manera inconfundible. Suelen tener denominaciones bien definidas, por ejemplo: “Museo de Louvre”, “El nacimiento de Venus”, “Pirámide de Keops”.
- **Atributos o Propiedades:** son cada una de las características que posee una clase o instancia. Constituyen rasgos que definen a las entidades, las dotan de esencia y las diferencian en un mundo de existencias. La identificación de propiedades o atributos contribuye en la formación de clases y en la conceptualización que se tiene de una entidad.
- **Relaciones:** son vinculaciones significativas entre las entidades que forman un dominio de conocimiento. Las relaciones responden a tres funciones: explicitar la naturaleza y comportamiento del dominio para reconstruir el conocimiento, presentar modelos organizativos que sean significativos y capturar de manera semántica (conceptual y relacional) procesos o dominios de conocimiento, cualquier tipo de relación entre entidades.

MAPAS TÓPICOS Y ONTOLOGÍAS TERMINOLÓGICAS: ANÁLISIS COMPARATIVO

Mapas tópicos y ontologías terminológicas son sistemas con características cercanas, de ahí que persista una constante confusión entre ambos. En la siguiente tabla (Tabla 1) es posible percibir algunas de sus características que las aproximan y/o diferencian:

Características ¹³	Mapas tópicos	Ontologías terminológicas
Objetivo	Ser índices que, mediante tópicos establecidos, redireccionen hacia recursos de información	Representar y organizar el conocimiento terminológico que favorezca el entendimiento entre expertos, posibilite la comunicación consensuada sobre dominios o permita a indización de recursos de información
Origen	2000, Ciencias computacionales	Década de 1980, Ciencias computacionales
Cobertura	Dominios de conocimiento	Dominios de conocimiento
Entorno	Digital	Digital
Fuentes	Autorizadas (literatura científica y lenguaje de especialistas)	Autorizadas (literatura científica y lenguaje de especialistas)
Lenguaje	Lenguaje natural, lenguaje controlado, lenguaje formal	Lenguaje natural, lenguaje controlado, lenguaje formal
Estructura	Terminológica, jerárquica, relacional	Terminológica, conceptual, atributiva, jerárquica, relacional
Costo de elaboración	Alto	Alto
Tipos de relaciones	Equivalencia, jerárquicas, asociativas y cualquier otro tipo de relación (temporales, familiares, causa-efecto, síntomas-tratamiento) que sea posible de estructurar mediante RDF	Equivalencia, jerárquicas, asociativas y cualquier otro tipo de relación (temporales, familiares, causa-efecto, síntomas-tratamiento) que sea posible de estructurar mediante RDF
Inferencias	Sí, nivel medio	Sí, nivel alto
Definiciones	No	Sí
Reutilización	Sí	Sí
Axiomas	No	Sí
Normativa de construcción	ISO/IEC. 13250: Topic Maps	No cuentan con normativa oficial
Presentación	Recomendaciones ISO/IEC 13250	Recomendaciones W3C

13 Algunas categorías de análisis de la columna 1 de la Tabla 1, fueron retomadas del artículo *Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías* de autoría Soler Monreal y Gil Leiva.

Editores de construcción	Ontopia, Wandora, OntopicStudio, Crystallize	Protégé, Graphite Taxonomy and Ontology Management, TopBraid Composer, Ontosaurus, Ontopia
Lenguajes de construcción	RDF, XML, XMT	RDF, OWL, Turtle, OBO Format, OIL, DAM-OIL, Manchester OWL Syntax, Latex Syntax, JSON-LD
Productores	Expertos en cómputo, lingüistas, profesionales de la información	Expertos en cómputo, lingüistas, profesionales de la información
Usuarios	Usuarios humanos y agentes artificiales	Usuarios humanos y agentes artificiales
Entidades	Asentadas y estables	Asentadas y estables
Autoría	Institución o persona creadora	Institución o persona creadora
Actualización	Modificable según requerimientos	Modificable según requerimientos
Precisión en la recuperación	Media	Alta
Presentación	Índice alfabético Gráfico-visual	Jerarquía Gráfico-visual

Tabla 1. Mapas tópicos-Ontologías terminológicas-Comparación (Fuente: Elaboración propia, 2024).

Con base en los rasgos de cada sistema se detectó que los mapas tópicos y las ontologías terminológicas convergen en elementos referenciales, terminológicos, conceptuales y relacionales. En cuanto a sus divergencias, las ontologías terminológicas superan a los mapas tópicos en rasgos atributivos y en su capacidad para operar en el marco de la web semántica, no solo como plantillas temáticas sino como bases de datos que afianzan la recuperación de información, esto es, como herramientas subyacentes en motores de búsqueda. También se identificó que, mientras los mapas tópicos tienen por objetivo servir como rutas tópicas con hipervinculación a los recursos de información digitales, las ontologías terminológicas -si bien pueden cumplir la misma función- han sido gestadas como bases de datos para el razonamiento de las inteligencias artificiales, lo que implica funciones más allá de simples rutas tópicas.

MAPAS TÓPICOS Y ONTOLOGÍAS TERMINOLÓGICAS: VENTAJAS/DESVENTAJAS EN LA ORGANIZACIÓN TEMÁTICA DE DOMINIOS DE CONOCIMIENTO

Durante las últimas décadas, las tecnologías han renovado el universo de los sistemas de organización del conocimiento y, como resultado, las bibliotecas y repositorios digitales han registrado enormes cambios. Nisheva y Pavlov¹⁴ señalan que las bibliotecas digitales deben integrar las mejores prácticas de creación, organización, recuperación y uso de contenidos digitales. De igual modo, recomiendan la aplicación de tecnologías semánticas que promuevan nuevas formas de búsqueda y navegación para las comunidades de usuarios. Es tiempo de que los espacios digitales bibliotecológicos ofrezcan una búsqueda basada en palabras clave (nivel léxico) y búsqueda semántica (nivel jerárquico y relacional) y, en tales funciones, los mapas tópicos y las ontologías terminológicas ofrecen grandes beneficios.

Las ventajas de los mapas tópicos y las ontologías terminológicas en la organización temática de colecciones digitales son las siguientes:

- Siguen la tendencia del control de vocabulario.
- Integran una estructura jerárquica que incide en la organización/recuperación de los recursos.
- Crean asociaciones a partir de tripletas RDF-Marco de Descripción de Recursos, bajo la sintaxis Sujeto-Predicado-Objeto.
- Generan grafos para el descubrimiento de los recursos.
- Se adhieren al modelo FRISAD-Requisitos funcionales para datos de autoridad de materia, basado en entidades-atributos-relaciones y al estándar SKOS-Sistemas para la Organización del Conocimiento Simples.
- Aportan en la construcción de la web semántica.
- Forman parte de los vocabularios de datos enlazados.

Si bien poseen muchas ventajas, también integran algunos aspectos desventajosos:

- Construcción completa, desde su planeación, desarrollo, codificación tecnológica, implementación, evaluación, visualización, etcétera.
- Alto costo de recursos humanos, tecnológicos y económicos en su planeación, desarrollo e implementación.

14 Maria M. Nisheva-Pavlova y Pavel Iliev Pavlov, "Ontology-based search and document retrieval in a digital library with folk songs", *Information Services & Use* 31, n.º 3-4 (Octubre, 2012): 157-166.

- Poca vinculación con los sistemas para la gestión de bibliotecas o repositorios digitales, casi siempre su visualización final se coloca en una interfaz web agregada.
- Lenta integración en sitios web, repositorios, bibliotecas y otras plataformas digitales.

A partir de sus ventajas se observan como sistemas de organización del conocimiento con amplio potencial en la organización de colecciones digitales. Tanto mapas tópicos como ontologías terminológicas están diseñados para mejorar la navegación y recuperación. Si bien desde el año 2010 se observa un crecimiento paulatino de implementación en sitios web, bibliotecas, repositorios y plataformas digitales, su integración, con mayor fuerza, es un asunto pendiente que sólo el tiempo esclarecerá.

CONCLUSIONES

El desarrollo de las tecnologías de la información y comunicación modifica el universo documental y las unidades informativas. De igual forma, también transforma las teorías y procesos organizadores del conocimiento y los sistemas que le sirven de apoyo. Mapas tópicos y ontologías terminológicas son parte de este enriquecimiento tecnológico que en el espacio digital intenta ordenar temáticamente los recursos de manera precisa, categorizada, relacionada y visual-gráfica. Su objetivo es simular temáticamente los dominios de conocimiento en el ciberespacio con miras a permitir un acceso temático más significativo, dinámico y navegable.

Como señalan García Martul y Franco Álvarez,¹⁵ se ha hecho necesario tanto para los profesionales de la información como para otros profesionales concebir sistemas de organización del conocimiento más robustos en sus propiedades, con el claro propósito de adaptarlos no sólo a la recuperación de los actuales objetos informativos sino a las nuevas formas de acceso a los mismos y a las capacidades del usuario para navegar en el espacio hipermedia.

Actualmente, en la representación, organización y recuperación temática de los recursos digitales albergados en el ciberespacio es posible emplear encabezamientos de materia y descriptores registrados en campos de metadatos temáticos, pero también es posible crear mapas tópicos u ontologías terminológicas con altos niveles de precisión terminológica, asociaciones complejas y

15 David García Martul y Guillermina Franco Álvarez, “¿Tesauros, mapas conceptuales o topic maps para la Biblioteca Digital Agustín Millares?”, *Boletín Millares Carlo* 28 (Enero 2009): 133-155.

funciones hipertextuales que permiten integrar los conceptos y redireccionar recursos a través de enlaces, posibilitando entornos interactivos. Estamos ante una nueva perspectiva que combina sistemas de organización del conocimiento, hipertextualidad, estructuras de navegación, procesos de indización y técnicas de recuperación de información acordes al entorno digital.

Organizar adecuadamente recursos de información en la web implica un ordenamiento eficaz que satisfaga las necesidades del usuario y mejore el descubrimiento de información. Resulta preciso identificar y vincular los metadatos temáticos con el uso de técnicas e instrumentos tecnológicos. Por último, se concluye sobre la importancia de crear mapas tópicos y ontologías terminológicas en sistemas de información digitales (repositorios, bibliotecas y otras plataformas en la web) debido a que la aparición y el desarrollo de la web semántica, desde principios del siglo, está cambiando drásticamente la forma de organización, producción y distribución de los recursos. Tales sistemas se hacen más necesarios tras la aparición de skos-Sistemas de Organización del Conocimiento Simples y las recomendaciones del W3C-World Wide Web Consortium que promueven el desarrollo de vocabularios controlados con gran aproximación al lenguaje ontológico; esto es, representados en RDF y codificados en OWL u otros lenguajes ontológicos.

REFERENCIAS BIBLIOGRÁFICAS

- Bergman, Michael, "An intrepid guide to ontologies", consultado noviembre 14, 2023, <https://www.mkbergman.com/374/an-intrepid-guide-to-ontologies/>
- Biezunki, Michel, "Topic maps and the essence of indexing", *The Indexer* 36, n.º 4, enero, 2018, 157-161.
- García Martul, David y Guillermina Franco Álvarez, "¿Tesauros, mapas conceptuales o topic maps para la Biblioteca Digital Agustín Millares?", *Boletín Millares Carlo* 28 (Enero 200): 133-155.
- Gruber, Thomas R., "Ontology", consultado febrero 17, 2024, <http://tomgruber.org/writing/ontology-definition-2007.htm>
- Hodge, Gail, *Systems of knowledge for digital libraries: beyond traditional authority files*, Washington: Council on Library and Information Resources, 2000.
- López Alonso, Miguel Ángel, "Las estructuras conceptuales de representación del conocimiento en Internet", *Scire: representación y organización del conocimiento* 6, n.º1 (Junio, 2000): 107-123.
- Miller, Les, Sree Nilakanta, Yunan Song, Lei Zhu y Ming Hua, "Managing knowledge in organizational memory using topic maps", *International Journal of Knowledge Management* 4, n.º 1 (Enero, 2008): 17-28.
- Nisheva-Pavlova, Maria M. y Pavel Iliev Pavlov. "Ontology-based search and document retrieval in a digital library with folk songs", *Information Services & Use* 31, n.º 3-4 (Octubre, 2012): 157-166.
- Orbst, Leo, "Introduction to ontologies and semantic technologies", consultado abril 22, 2024, https://stids.c4i.gmu.edu/STIDS2011/presentations/STIDS2011_Tutorial_Parts1-2.pdf
- Organización Internacional de Normalización, *ISO/IEC 13250:2000-Tecnologías de la información-SGML Aplicaciones-Mapas tópicos*, Ginebra, Suiza: ISO, 2000.
- Organización Internacional de Normalización *ISO/IEC 13250:2003-Tecnologías de la información-Aplicaciones SGML- Mapas tópicos*, Ginebra, Suiza: ISO, 2003.

- Organización Internacional de Normalización, *ISO/IEC 13250-2:2006-Tecnologías de la información-Mapas tópicos-Parte 2: Modelos de datos*, Ginebra, Suiza: ISO, 2006.
- Organización Internacional de Normalización, *ISO/IEC 13250-3:2007-Tecnologías de la información-Mapas tópicos-Parte 3: Sintaxis XML*, Ginebra, Suiza: ISO, 2007.
- Organización Internacional de Normalización, *ISO/IEC 13250-4:2009-Tecnologías de la información-Mapas tópicos-Parte 4: Estandarización*, Ginebra, Suiza: ISO, 2009.
- Organización Internacional de Normalización, *ISO/IEC 13250-6:2010-Tecnologías de la información-Mapas tópicos-Parte 6: Sintaxis compacta*, Ginebra, Suiza: ISO, 2010.
- Organización Internacional de Normalización, *ISO/IEC 13250-5:2015-Tecnologías de la información-Mapas tópicos-Parte 5: Modelo de Referencia*, Ginebra, Suiza: ISO, 2015.
- Pieterse, Vreda y Derrick G. Kourie, “Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle”, *Knowledge Organization* 41, n.º 3 (Enero, 2014): 217-229.
- Soler Monreal, Concha e Isidoro Gil Leiva, “Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías”, *Revista Interamericana de Bibliotecología* 33, n.º 2, diciembre 2010, 361-377.
- Zeng, Marcia Lei, “Knowledge organization systems (KOS)”, *Knowledge Organization* 35, n.º 2-3 (Enero, 2008): 160-182.

Las relaciones terminológicas entre los sistemas de información y los sistemas informáticos

The terminological relationships between information systems and computer systems

Hugo Alberto Guadarrama Sánchez
Instituto de Investigaciones Bibliotecológicas y de la Información
Universidad Nacional Autónoma de México

1. INTRODUCCIÓN

El término *sistema de información* usualmente presenta una polisemia, dependiendo de los diferentes enfoques disciplinarios y campos de estudio, como por ejemplo, en el sector tecnológico y en la organización de la información la definición o cuando menos la percepción de un sistema de información siempre será distinta, porque los intereses disciplinarios están orientados en direcciones, objetivos y metas disímiles. Asimismo, cada formación disciplinaria posee sus propios presupuestos teóricos y técnicos, de manera que si alguna persona que desconozca acerca de la naturaleza de cada dominio de conocimiento podría confundirse o solamente considerar una perspectiva disciplinaria.

En este tenor, las estructuras terminológicas son interpretaciones que expresan algunas de las características más representativas de las actividades disciplinarias o conjuntos de medios tecnológicos a partir de su importancia a nivel conceptual, así como su prolijidad de manera ordenada y organizada para garantizar, en la medida de lo posible, un entendimiento por parte de las personas interesadas en conocer la composición terminológica de ciertas áreas del conocimiento o agrupación de objetos tangibles, pero sobre todo las estructuras terminológicas deben de reflejar de manera cabal, objetiva y precisa las propiedades de algún dominio de conocimiento con base en la garantía literaria procedente de los propios contenidos disciplinarios.

La terminología¹ como composición estructurada de los términos favorece el entendimiento del discurso, ya que cada concepto se encuentra asociado a un campo del conocimiento u oficio en particular, ello permite crear una diferenciación mutua entre dominios similares o relacionados. La terminología, además de proporcionar un soporte al discurso especializado, reduce el ruido en las comunicaciones en favor de la sistematicidad de los textos especializados, reflejándose las prácticas disciplinarias que se llevan a cabo en las aulas, los laboratorios, las bibliotecas y en los centros de investigación, de manera que el valor de los términos disciplinarios se distingue por una densidad de conocimiento acumulado al paso del tiempo.

Una estructura terminológica² se define por un conjunto de palabras altamente especializadas en un determinado campo de estudio o dominio de conocimiento, la importancia de las estructuras terminológicas radica en conocer, comunicar y comprender la composición temática y teórica de los dominios de conocimiento; en este sentido, un término pertenece a una esfera de estudio y de investigación, aunque en ocasiones los términos se encuentran en más de un dominio, pero su relevancia tiende a ser diferente, dependiendo del contexto académico, además de ello, en ocasiones se presenta el problema de sinonimia y polisemia, aunque el propio contexto de la estructura terminológica coadyuva a esclarecer la ambigüedad.

Las estructuras terminológicas funcionan a manera de diagramas con la capacidad de mostrar las partes que integran la totalidad, en cierto modo las estructuras terminológicas son una descripción reducida de los dominios de conocimiento por el hecho de mostrar las propiedades de un conjunto temático ordenado y jerarquizado, en este aspecto de representación la descripción³ se orienta en observar, percibir e inspeccionar la naturaleza de un objeto físico y/o abstracto de forma cualitativa o cuantitativa, utilizando para ello un instrumento textual, mecánico o, bien, alguna referencia que provea de contexto para analizar las conexiones que hacen viable la integración y operatividad de las partes de la estructura terminológica, puesto que no se puede observar a las entidades desde un punto ciego o de una mera suposición.

1 María Teresa Cabré, “Contexto y evolución de la terminología: de una aproximación nominalista a una teoría comunicativa”, en *Teoría y praxis en terminología*, coords. Sara Catalá y Mario Barité (Uruguay: Ediciones Universitarias, 2017), 2.

2 Birger Hjørland, “Encyclopedia of Knowledge Organization”, noviembre 23, 2022, <https://www.isko.org/cyclo/terminology>

3 Birger Hjørland, “Description: Its meaning, epistemology, and use with emphasis on information science”, *Journal of the Association for Information Science and Technology* 74, n.º 13 (2023), 1535, <https://doi.org/10.1002/asi.24834>

Las estructuras terminológicas⁴ poseen una variación en sus ramificaciones en función del contexto social, temporal o comunicativo, a este fenómeno se le conoce como multidimensionalidad, ello significa que no todas las comunidades disciplinarias utilizan los términos de la misma manera y sus jerarquías terminológicas poseen diferencias notables a pesar de utilizar las mismas unidades léxicas posicionadas en diferentes niveles, de esta manera las variantes terminológicas muestran los intereses disciplinarios en relación a la relevancia de ciertos tópicos tecnológicos o abstractos, según sea el caso, la variación terminológica no sólo tiene una variación cognitiva sino que tiene una funcionalidad en la interpretación de los textos.

En el caso de los sistemas de información y los sistemas informáticos al ser dominios de conocimiento relacionados pero diferenciados por sus respectivos enfoques, sus estructuras terminológicas comparten algunas unidades léxicas, aunque su posicionamiento jerárquico depende directamente de su naturaleza como sistemas de información y sistemas informáticos, a pesar de manifestar una diferenciación notable en el significado de sus unidades pueden llegar a ser similares para aquellas personas ajenas a estos dominios, provocando confusión en cuanto a sus propiedades.

2. SISTEMAS DE INFORMACIÓN Y SISTEMAS INFORMÁTICOS

Un sistema de información es el resultado de una serie de documentos almacenados, procesados y organizados que hacen viable la recuperación de la información. En los sistemas de información la tecnología electrónica y digital pueden llegar a ser prescindibles, puesto que el acceso a los documentos puede realizarse de forma directa, sin la necesidad de consultar alguna base de datos o interfaz digital; no obstante, en ocasiones es indispensable contar con alguna persona que conozca la ubicación de algún documento en específico; por otro lado, un sistema informático se compone a partir de un conjunto de dispositivos tecnológicos interconectados que dan respuesta a alguna actividad institucional, académica, laboral o profesional.⁵

4 Sabela Fernández-Silva y Nelson Becerra Rojas, “La variación terminológica en la comprensión y producción de textos académicos: propuesta de representación en un diccionario especializado de aprendizaje de Psicología”, *Ibérica*, n.º 30 (2015), 187.

5 Wiki Informática. 2024. “Sistema informático y sistema de información”, https://informatica.fandom.com/wiki/Sistema_inform%C3%A1tico_y_sistema_de_informacion

Las instituciones, compañías y empresas utilizan sus propios sistemas de información,⁶ los cuales no dependen de los sistemas informáticos, empero en situaciones de competitividad o alta demanda de la información pertinente y relevante, los medios y dispositivos tecnológicos se hacen necesarios. Cuando los sistemas de información convergen con los sistemas informáticos emergen conceptualmente los sistemas automatizados, caracterizados por generar resultados a partir de secuencias programadas que atienden diferentes variables y situaciones para generar servicios y productos de información a sus respectivas organizaciones y/o comunidades.

Un sistema informático también puede ser identificado como una Tecnología de la Información (TI) basada en un conjunto de instrumentos, medios y dispositivos que dan sustento a la infraestructura institucional. Los responsables del Departamento de Tecnologías de la Información (DTI) se encargan de diseñar, desarrollar y mantener las redes, el *hardware* y el *software* del sistema informático; en cambio, los sistemas de información se distinguen por optimizar los procesos referentes al acopio, adquisición, almacenamiento, análisis y difusión de la información, de acuerdo con los objetivos institucionales y las necesidades de información en cada organización, asegurando que la información sea verificada, precisa y accesible para quienes la necesitan. La diferencia entre los sistemas informáticos y los sistemas de información se distingue por los recursos y las funciones administrativas.⁷

2.1 Sistemas informáticos

Chacón define al sistema informático como la combinación entre distintos componentes electrónicos que realizan un procesamiento abstracto, lógico y matemático, como instrumentos receptores de los dispositivos de entrada, utilizados para transcribir y registrar los datos que serán almacenados en el disco duro de las computadoras y, posteriormente, visualizados con dispositivos de salida; tanto los componentes de la computadora como los programas de productividad utilizados en las instituciones se dividen en dos categorías: *hardware* y *software*.⁸

6 Juan Armada, “Sistemas de Información vs Sistemas Informáticos”, septiembre 11, 2020, <https://juanarmada.com/sistemas-de-informacion-vs-sistemas-informaticos/>

7 Francisco Velázquez, “No es lo mismo tecnología de la información que sistemas de información: conoce las diferencias”, mayo 3, 2023, <https://es.linkedin.com/pulse/es-lo-mismo-tecnolog%C3%ADa-de-la-informaci%C3%B3n-que-sistemas-vel%C3%A1zquez-cruz>

8 Francisco Chacón, “Sistemas informáticos: estructura y funciones”, 2 (2007): <https://www.preparadores.eu/temamuestra/PTecnicos/PComerciales.pdf>

Para fines prácticos, los sistemas informáticos⁹ usualmente están segmentados en *hardware* y *software*; sin embargo, el usuario es un elemento crucial en la toma de decisiones, de modo que es el recurso que comúnmente vincula, en un sentido operativo, el *hardware* y el *software*; dichos recursos (humano-interfaz-programa) hacen posible la obtención de objetivos institucionales e individuales, aunque cabe destacar que al paso del tiempo el *software* tiende a rebasar las capacidades del *hardware*, e incluso, a disminuir la accesibilidad hacia el usuario.

Un sistema informático¹⁰ además de analizar los datos recopilados, interactúa con otros sistemas o tecnologías de envío y recepción de mensajes mediante el uso de satélites y redes que rodean al planeta; estos sistemas, ajenos a las computadoras convencionales, envían y reciben datos que son analizados de manera que los usuarios puedan obtener resultados concretos mediante el uso de diferentes medios tecnológicos que realizan procesos complejos, los cuales llevarían mucho tiempo de forma manual y aislada, en este sentido se puede considerar que los sistemas informáticos se constituyen a través de un ensamble de diferentes componentes y dispositivos.

Los sistemas informáticos¹¹ no sólo están basados en *hardware*, *software* y usuarios sino también en documentación, la cual incluye un conjunto de manuales o tutoriales que explican la operatividad y el mantenimiento de cada sistema; los sistemas informáticos requieren de un mantenimiento periódico con la finalidad de no perder los datos y la información almacenada, así como evitar gastos económicos y extender su utilidad física, abstracta y funcional mediante la ejecución práctica de los manuales que explican los procesos de mantenimiento preventivo.

Aunque los sistemas informáticos¹² tienen su propósito de ayudar a los humanos en el procesamiento de la información y de los datos, la potencia de los sistemas informáticos depende del grado de prestación de sus componentes electrónicos que se encuentran en el interior de las computadoras de escritorio, computadoras portátiles y estaciones de trabajo, condicionando la veloci-

9 Marisel Bernad y Mabel Rodríguez, “Sistemas informáticos, tipos y clasificación”, 1 (2020): https://epet1.caucete.com/admin/noticia_img/Carpeta%20guia2%20web/bernad212223inform%C3%A1tica.pdf

10 Juan Sepúlveda, Carlos Lago, Alejandro Rasete, Roberto Sepúlveda y Javier Lorenzo, “Sistema Informático para análisis y procesamiento de datos de los computadores de a bordo de la serie AGM-200X”, *Revista Cubana de Ciencias Informáticas* 2, n.º 1 (2008), 13.

11 Francisco Calle, “Mantenimiento de Sistemas Informáticos” (2020): https://iessoterohernandez.es/wp-content/uploads/2020/05/Tareas04_FPB_1%C2%BA_MMSCI.pdf

12 Sotirios Ziavras, “Computer systems” 1, (2012): <https://www.eolss.net/sample-chapters/c15/E6-45-02-08.pdf>

dad y la capacidad tanto de almacenamiento, gestión, edición y recuperación de la información a un nivel técnico, pero no logístico.

2.2 Sistemas de información

Un sistema de información¹³ es una secuencia de actividades enfocadas en: registrar, administrar, manipular y almacenar datos y documentos con información, con el propósito de que las organizaciones puedan realizar sus funciones, alcanzar sus objetivos y tomar decisiones asertivas. Un sistema de información surge debido a las necesidades de información de las organizaciones y se desarrolla como un conjunto de procesos vinculados, no necesariamente tecnológicos, que hacen viable la gestión de la información.

Cada sistema de información¹⁴ se conforma por un subsistema técnico que incluye un componente tecnológico y otro vinculado a los procesos; en adición, cuenta con un subsistema social constituido por la estructura y las personas; cabe mencionar que en la mayoría de las ocasiones el subsistema técnico recibe mayor atención, ello debido a que la tecnología integrada por las computadoras de escritorio, computadoras portátiles, los teléfonos móviles y los accesorios inteligentes son considerados instrumentos indispensables en la edición y difusión de los contenidos digitales, así como el *software* de programación, sin dejar de lado la infraestructura externa de telecomunicaciones.

El componente del proceso organizacional¹⁵ en los sistemas de información está enfocado en una mejor productividad y control de los recursos y servicios, la finalidad de un proceso es eliminar costos, tiempos y reducir el ruido en las comunicaciones para proporcionar un mejor nivel de satisfacción. Cada proceso requiere de su documentación para garantizar el control en la estandarización de los productos y servicios. Los procesos suelen documentarse mediante listados de pasos a seguir y diagramas que explican la secuencia de las actividades.

13 Francisco Gonzalez-Longatt, “Introducción a los Sistemas de Información: fundamentos”, 3 (2017): <https://www.uv.mx/personal/artulopez/files/2012/08/FundamentosSistemasInformacion.pdf>

14 Richard Watson, *Information Systems* (United States: Global Text Project, 18, 2007), <https://ufdcimages.uflib.ufl.edu/AA/00/01/17/04/00001/InformationSystems.pdf>

15 David Bourgeois, “Information Systems for Business and Beyond”, 2019, <https://opentextbook.site/informationssystem2019/chapter/chapter-1-what-is-an-information-system-information-systems-introduction/>

Huelga decir que, un sistema de información¹⁶ tiene la capacidad de conectar las diferentes actividades de cada departamento o área de trabajo, además de generar informes diversificados en texto, hojas de cálculo, gráficos y datos estructurados complejos que pueden manifestarse en diferentes formatos decodificables y legibles en diferentes dispositivos tecnológicos. Los sistemas de información pueden recopilar datos de las búsquedas, consultas, préstamos, devoluciones y multas que muestran el movimiento de los documentos y el estado de las colecciones para analizar, sintetizar, calcular y ordenar los datos de acuerdo con las necesidades de análisis de los administradores. Los procesos de un sistema de información también son esenciales para determinar el funcionamiento del sistema informático.

Usualmente, un sistema de información¹⁷ recopila, procesa, almacena y analiza la información de su entorno a través de una entrada que recibe datos para generar informes en su salida, los informes en formato impreso son considerados como una herramienta para identificar los riesgos en las instituciones, de manera que no sólo es necesario el uso de las computadoras y de los programas informáticos, sino también es necesario contar con impresoras para leer y entender la información; en este sentido, los documentos impresos son una tecnología primigenia y eficaz en las instituciones académicas y gubernamentales para determinar los avances y los gastos en comparación con los años anteriores.

Los sistemas de información destacan las capacidades del subsistema técnico, el cual contiene y fundamenta los procesos que permiten la consolidación de las actividades necesarias para generar un nivel óptimo en la satisfacción de los usuarios, control de los recursos de información y actualización de los registros en las bases de datos. Un sistema de información también posee componentes abstractos que se encuentran interconectados para garantizar la recolección, almacenamiento y difusión de los datos, teniendo como mecanismo a la retroalimentación, que provee indicadores para mejorar los productos y servicios de información, así como para tomar mejores decisiones.¹⁸

16 TechTarget Contributor, "Information systems", 2023, <https://www.techtarget.com/whatis/definition/IS-information-system-or-information-services>

17 Jorge Ribeiro de Toledo y Paulo Elias, "An information sub-system model for the utilization of the receipt printer as a management instrument in the supermarket sector in the state of Santa Catarina", *Journal of Information Systems and Technology Management* 7, n.º 3 (2010), 593, <https://www.tecsi.org/jistem/index.php/jistem/article/view/10.4301%252FS1807-17752010000300004/227>

18 Leila Zemmouchi-Ghomari, "Basic Concepts of information systems", en *Contemporary issues in information systems: a global perspective*, ed. Denis Reilly (Republika Hrvatska: IntechOpen, 2022), 7, <https://www.intechopen.com/chapters/76540>

En suma, los sistemas de información¹⁹ están integrados por *hardware* y *software*, pero sobre todo por individuos que tienen como objetivo recopilar, almacenar y registrar documentos valiosos tanto para sus respectivas organizaciones, a las que pertenecen, como para sus comunidades que hacen uso de los recursos de información; asimismo, el almacenamiento de los documentos y de los registros requiere de medios tecnológicos que permiten la transferencia y la preservación digital de los bancos de datos y contenidos digitales, ello con el fin de garantizar el control de las colecciones y generar certidumbre en los administradores y directivos de cada organización.

3. RELACIONES TERMINOLÓGICAS ENTRE LOS SISTEMAS DE INFORMACIÓN Y LOS SISTEMAS INFORMÁTICOS

A partir de las referencias y autores mencionados anteriormente se realizó una composición terminológica de los sistemas de información y los sistemas informáticos, ello ha permitido detectar algunas diferencias en sus estructuras (ver anexo A y B), como por ejemplo, la categoría de “Componente” y “Hardware” o viceversa, dependiendo de cada estructura terminológica. Además de la relevancia de los términos posicionados como categorías, resulta evidente el cambio de singular a plural en los términos “Componentes” y “Componente tecnológico”. Esta particularidad en ambas estructuras denota la relevancia que existe para ambos dominios de conocimiento en torno a los medios tecnológicos, como se muestra en la Tabla 1.

Sistemas informáticos			Sistemas de información		
	Hardware			Componente tecnológico	
		Componentes			Hardware

Tabla 1. Diferencias en las estructuras terminológicas.

En la estructura terminológica de los sistemas informáticos (anexo A), el *hardware* contiene un desglose amplio de términos, a diferencia de los sistemas de información (anexo B), que tienden a limitar la ramificación terminológica sin desarrollar sus términos específicos. La “documentación” es un término

19 Forat Hasan, “A review study of information systems”, *International Journal of Computer Applications* 179, n.º 18, 2018, 15, <https://www.ijcaonline.org/archives/volume179/number18/hasan-2018-ijca-916307.pdf>.

relacionado en ambas estructuras, puesto que dicha actividad proporciona certidumbre en la operatividad tanto a nivel tecnológico como organizacional, presentándose en dos diferentes categorías: “Hardware” en los sistemas informáticos, y “Componente de procesos” dentro de los sistemas de información. Con base en lo anterior se puede interpretar que la documentación tecnológica y administrativa es un activo primordial en las organizaciones que constituyen las instituciones, las compañías privadas y las empresas dedicadas a los negocios.

La documentación, entendida como un conjunto de impresos y/o contenidos digitales organizados, posee un alto valor intelectual para sus comunidades; sin importar la tecnología electrónica y/o digital, la documentación es un elemento que provee de conocimiento a los usuarios para afrontar las situaciones complejas de carácter lógico y social. Razón por la cual, la organización de la información adquiere un mayor sentido en las instituciones gubernamentales, académicas y de investigación, por el hecho de disminuir los costos administrativos y las inversiones económicas. De esta manera se puede entender que los sistemas de información y los sistemas informáticos, a pesar de estar estrechamente vinculados, poseen diferencias que al mismo tiempo los complementan como conjuntos en intersección, pero sin perder su composición heterogénea como dominios.

4. CONCLUSIONES

Las relaciones terminológicas entre los sistemas de información y los sistemas informáticos pueden llegar a presentarse de manera explícita a medida que sus ramificaciones terminológicas se desarrollen, ello se debe a que comparten un conjunto de dispositivos tecnológicos y procedimientos de trabajo en el tratamiento de la información manifestada en impresos, discos ópticos, objetos culturales y contenidos digitales. Los sistemas de información se enfocan mayormente en los procedimientos que involucran la transferencia de los documentos y las prioridades institucionales en su adquisición y resguardo; mientras que los sistemas informáticos se orientan en identificar los diferentes medios tecnológicos que hacen viable el registro y la recuperación de la información; sin embargo, cuando la complejidad administrativa se incrementa en las organizaciones las estructuras terminológicas convergen progresivamente, formando una nueva estructura terminológica en cada organización, expandiendo y modificando el dominio de conocimiento.

A pesar de la particularidad de los dominios de conocimiento en las organizaciones que cuentan con diferentes miembros procedentes de diferentes dis-

ciplinas, resulta importante distinguir las diferencias y similitudes terminológicas, porque de ello depende el nivel de entendimiento que se puede obtener en las comunicaciones; a manera de ejemplo, los departamentos institucionales dedicados al resguardo de los archivos tendrían que conocer algunos de los términos utilizados por el departamento de informática y, a su vez, dicho departamento, de forma recíproca, con el archivo histórico institucional, pero esta comprensión mutua depende tanto de la disposición como de las interacciones que se producen al interior de las organizaciones. El uso de la terminología no tiene la finalidad de crear conflictos o fricciones entre las personas, sino que debe de procurar la ampliación del vocabulario en beneficio de las actividades laborales, disciplinarias y de la investigación.

REFERENCIAS BIBLIOGRÁFICAS

- Armada, Juan, “Sistemas de Información vs Sistemas Informáticos”, consultado febrero 13, 2024, <https://juanarmada.com/sistemas-de-informacion-vs-sistemas-informaticos/>
- Bernad, Marisel y Mabel Rodríguez, *Sistemas informáticos, tipos y clasificación*, Argentina, 2020, https://epet1cauce-te.com/admin/noticia_img/Carpeta%20guia2%20web/bernad212223inform%C3%A1tica.pdf
- Bourgeois, David, “Information Systems for Business and Beyond”, 2019, consultado febrero 13, 2024, <https://opentextbook.site/informationssystem2019/chapter/chapter-1-what-is-an-information-system-information-systems-introduction/>
- Cabré, María, “Contexto y evolución de la terminología: de una aproximación nominalista a una teoría comunicativa” en *Teoría y praxis en terminología*, coordinado por Sara Catalá y Mario Barité. Uruguay: Ediciones Universitarias, 2017.
- Calle, Francisco, *Mantenimiento de Sistemas Informáticos*, 2020, https://iessoterohernandez.es/wp-content/uploads/2020/05/Tareas04_FPB_1%C2%BA_MMSCI.pdf
- Chacón, Francisco, *Sistemas informáticos: estructura y funciones*, Madrid, 2007, <https://www.preparadores.eu/temamuestra/PTecnicos/PComerciales.pdf>
- Fernández-Silva Sabela y Nelson Becerra Rojas, “La variación terminológica en la comprensión y producción de textos académicos: propuesta de representación en un diccionario especializado de aprendizaje de Psicología”, *Ibérica*, n.º 30, 2015, 183-208.
- Gonzalez-Longatt, Francisco, “Introducción a los Sistemas de Información: fundamentos”, 2017, <https://www.uv.mx/personal/artulopez/files/2012/08/FundamentosSistemasInformacion.pdf>
- Hasan, Forat. “A review study of information systems”, *International Journal of Computer Applications* 179, n.º 18 (2018): 15-19, <https://www.ijcaonline.org/archives/volume179/number18/hasan-2018-ijca-916307.pdf>
- Hjørland, Birger. “Description: Its meaning, epistemology, and use with emphasis on information science”, *Journal of the Association for Information Science and Technology*

- 74, n.º 13 (2023): 1532-1549, <https://doi.org/10.1002/asi.24834>
- Hjørland, Birger, “Encyclopedia of Knowledge Organization”, consultado febrero 13, 2024, <https://www.isko.org/cyclo/terminology>.
- Ribeiro de Toledo, Jorge y Paulo Elias, “An information subsystem model for the utilization of the receipt printer as a management instrument in the supermarket sector in the state of Santa Catarina”, *Journal of Information Systems and Technology Management* 7, n.º 3 (2010): 579-598, <https://www.tecsi.org/jistem/index.php/jistem/article/view/10.4301%252FS1807-17752010000300004/227>
- Sepulveda, Juan, Carlos Lago, Alejandro Rasete, Roberto Sepúlveda y Javier Lorenzo, “Sistema Informático para análisis y procesamiento de datos de los computadores de a bordo de la serie AGM-200X”, *Revista Cubana de Ciencias Informáticas* 2, n.º 1 (2008): 183-208.
- TechTarget Contributor, “Information systems”, 2023, consultado febrero 13, 2024, <https://www.techtarget.com/whatis/definition/IS-information-system-or-information-services>
- Velázquez, Francisco, “No es lo mismo tecnología de la información que sistemas de información: conoce las diferencias”, consultado febrero 13, 2024, <https://es.linkedin.com/pulse/es-lo-mismo-tecnolog%C3%ADa-de-la-informaci%C3%B3n-que-sistemas-vel%C3%A1zquez-cruz>
- Watson, Richard, *Information Systems*, United States: Global Text Project, 2007, <https://ufdcimages.uflib.ufl.edu/AA/00/01/17/04/00001/InformationSystems.pdf>
- Wiki Informática, “Sistema informático y sistema de información”, 2024, consultado febrero 13, 2024, https://informatica.fandom.com/wiki/Sistema_inform%C3%A1tico_y_sistema_de_informacion
- Zemmouchi-Ghomari, Leila, “Basic Concepts of information systems” en *Contemporary issues in information systems: a global perspective*, editado por Denis Reilly, Republika Hrvatska: IntechOpen, 2022, <https://www.intechopen.com/chapters/76540>
- Ziavras, Sotirios, *Computer systems*, 2012, <https://www.eolss.net/sample-chapters/c15/E6-45-02-08.pdf>

ANEXO A

Sistemas informáticos				
	Hardware A, D, C, M			
		Documentación C		
		Componentes D		
			Tarjeta madre	
			Tarjeta gráfica	
			Tarjeta de sonido	
			Procesador	
			Memoria RAM	
			Disco duro	
				SSD
				HDD
		Dispositivos de entrada I		
			Teclado G	
				Mecánico
				Membrana
			Micrófono G	
			Mouse G	
				Alámbrico
				Inalámbrico
			Lector de óptico G	
		Dispositivos de salida I		
			Escáner G	

Organización del conocimiento...

			Webcam G	
			Monitor G	
				SVGA
				Táctil
				LED
				OLED
			Impresora G	
			Audífonos G	
				Diadema
				Micrófono
			Proyector G	
			Bocinas G	
			Dispositivos de red I	
				Módem
				Switch
		Router		
	Software A, C, D, M			
		Programación		
		Aplicación		
		Sistemas Ope- rativos		

*Estructura terminológica de los sistemas informáticos, con base en los autores consultados.

ANEXO B

Sistema de información				
	Subsistema técnico K			
		Componente tecnológico E		
	Hardware E			
			Servidores	
			Computadoras de escritorio	
			Computadoras portátiles	
			Estaciones de trabajo	
			Teléfonos móviles	
			Accesorios inteligentes	
	Software E			
			Programación	
			Aplicación	
	Componente de procesos A, B, E, F, H, J, L			Sistemas Operativos
		Infraestructura de telecomunicaciones E		
			Redes	
		Entrada		
			Recopilación	
		Recolección		
		Adquisición		
		Organización		

				Análisis	
				Registro	
				Almacenamiento	
			Salida		
				Informes	
				Datos estructurados	
				Resultados de búsquedas	
			Documentación B		
	Subsistema social K, L				
		Componente estructural K			
			Jerarquías organizacionales		
		Componente de personas K			
			Relaciones humanas		

*Estructura terminológica de los sistemas de información, con base en los autores consultados.

**Autores consultados*

- A. Bernad y Rodríguez, 2020.
- B. Bourgeois, 2019.
- C. Calle, 2020.
- D. Chacón, 2007.
- E. Gonzalez-Longatt, 2017.
- F. Hasan, 2018.
- G. Rebollo, 2011.
- H. Ribeiro de Toledo y Elias, 2010.
- I. Sepúlveda *et al.*, 2008.
- J. TechTarget Contributor, 2023.
- K. Watson, 2007.
- L. Zemmouchi-Ghomari, 2022.
- M. Ziavras, 2012.

La organización de la información en la era digital

The organization of information in the digital age

Blanca Estela Sánchez Luna
Colegio de Bibliotecología y Archivología
Facultad de Filosofía y Letras, UNAM

INTRODUCCIÓN

Las bibliotecas nutren sus colecciones, ya no solo con fuentes de información tradicionales sino con recursos de información electrónicos, esto les otorga ahora el adjetivo de bibliotecas híbridas; es decir, coexisten entre libros, fondos antiguos, materiales audiovisuales, fuentes digitales, incluso recursos de información derivados de la inteligencia artificial como el ahora conocido chat GPT. Chowdhury¹ señala que los usuarios obtienen información desde una gran variedad de fuentes a través de un amplio rango de canales que van desde los recursos bibliográficos convencionales, como libros y revistas a los que se puede acceder desde el OPAC, hasta una infinidad de recursos electrónicos. Estos incluyen bases de datos, revistas y libros electrónicos, sin mencionar la gran cantidad de recursos web y de internet, a cada uno de los cuales se puede acceder a través de interfaces de búsqueda específicas, diseñadas por los proveedores de servicios (p. 1). En este contexto, las unidades de información evolucionan y adaptan sus servicios y procesos internos de manera constante y según los requerimientos actuales de la sociedad con el firme propósito de satisfacer las necesidades de información de los usuarios, tanto para los que aún acuden físicamente a la biblioteca como para los usuarios que consumen información desde la red.

Por lo anterior, el objetivo de este capítulo se centra en esbozar un panorama actualizado de los recursos de información digitales, tomando como punto de

1 G.G. Chowdhury y Sudatta Chowdhury, *Organizing Information: From the shelf to the Web* (London: Facet, 2007), 1.

partida los fundamentos intelectuales de la organización de la información con la intención de analizar y fortalecer el rol que juega la organización documental en esta era digital, así como el papel del bibliotecólogo documentalista o profesional de la información.

Para alcanzar el objetivo planteado el trabajo se estructurará en tres partes: la primera, dedicada a resaltar la importancia y vigencia de los fundamentos sobre los que descansa la organización de la información; la segunda parte está dedicada a describir el panorama de los recursos de información actuales; y la tercera consiste en analizar el papel que juega la organización documental y los retos a los cuales se enfrenta. En este punto es importante destacar el comportamiento del usuario actual ante sus necesidades de información.

Finalmente, se ofrecen una serie de conclusiones con el afán de resaltar la importancia y vigencia no solo de la organización documental sino de las bibliotecas en esta era digital en la que recuperar información resulta relativamente simple; sin embargo, se recupera demasiada información que no siempre es fiable y tampoco cubre las necesidades específicas de los usuarios. De aquí se resalta el papel de la biblioteca ante la sociedad actual.

I. LOS FUNDAMENTOS INTELECTUALES DE LA ORGANIZACIÓN DE LA INFORMACIÓN

Para hablar de organización de la información, a partir de un entorno digital, es preciso remitirnos a los fundamentos sobre los cuales descansa esta teoría cuya funcionalidad y eficacia está basada en un grado considerable en el uso y aplicación de modelos conceptuales, herramientas y técnicas que norman la descripción de los documentos impresos, audiovisuales y digitales. Además, es de destacar que la organización de la información es una disciplina bibliotecológica que se ha caracterizado a lo largo del tiempo por evolucionar y adaptarse a las tecnologías de información de cada tiempo, basta echar un vistazo general a la transformación de los soportes documentales y los dispositivos para su consulta. Véase Tabla 1.

Soportes	Ejemplos	Dispositivo para su consulta
Papiro, pergamino	Códices, fondos de archivo	Ninguno
Papel	Libros, revistas, folletos, periódicos, fondos de archivo	Ninguno
Plásticos	Microfichas, discos de acetato	Soportes mecánicos (Lector de microfichas, tocadiscos)

Magnéticos	Casetes de audio, cintas beta y vhs, disquetes de 3 ½ y 5 ¼	Reproductor (grabadoras, televisiones, computadoras)
Ópticos	Discos compactos	Lector de discos compactos, computadora
Digitales o electrónicos	Libro electrónico, revista electrónica	Lector electrónico, tableta, computadora Para algunos se requieren aplicaciones específicas
Entorno web	Redes sociales (Facebook, Instagram), plataformas educativas	Tableta, computadora, teléfono inteligente, mediante aplicaciones específicas
Inteligencia artificial (IA)	Chat GPT	Tableta, computadora, teléfono inteligente

Tabla 1. Evolución de los soportes documentales (Fuente: Elaboración personal).

Ahora bien, retomando los fundamentos de la organización de la información, Svenonius² afirma que el marco conceptual de la organización documental parte de un contexto histórico-filosófico, aunque su característica más sobresaliente es que la información se organiza mediante la descripción física y de contenido a través del uso de lenguajes especializados. En cuanto al antecedente histórico más relevante, Svenonius lo coloca en la tradición de la catalogación angloamericana descriptiva y temática en lo que respecta del último siglo y medio; y añade que, si bien desde el 2000 a.C. se practicó alguna forma de organización sistemática, se suele considerar que la historia moderna de la catalogación data de mediados del siglo pasado con Antonio Panizzi para organizar los libros de la British Library. En el periodo posterior a Panizzi se sentaron las bases para los principales sistemas bibliográficos que se utilizan en las bibliotecas actuales: Clasificación decimal Dewey, la clasificación LC, la clasificación decimal universal, los encabezamientos de materia de la Biblioteca del Congreso (LCSH) y las Reglas de Catalogación Angloamericanas (pp. 1-2). A este punto vale añadir las herramientas actuales, como RDA (*Resource Description Access*) y los modelos conceptuales como FRBR, FRAD, Bibframe, LRM y datos vinculados.

2 Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, Massachusetts: The MIT Press, 2009), 1-2.

Con respecto al trasfondo filosófico, Svenonius³ afirma que resultan relevantes para la base intelectual de la organización de la información los puntos de vista adoptados por tres movimientos filosóficos que han permeado el pensamiento académico y popular durante el siglo xx: la filosofía de los sistemas; la filosofía de la ciencia, y la filosofía del lenguaje. En el primero de ellos recae el pensamiento de Charles A. Cutter (1876), ya que fue el primero en reconocer la importancia de los objetivos del catálogo.

En tal contexto, la biblioteca desde un punto de vista documental o bibliográfico busca seguir cumpliendo con su misión de antaño que es proporcionar al usuario la información que necesite en cualquier soporte documental. Esto recae en las tareas del catálogo planteadas por Cutter en 1876 y que, en términos generales, Spedalieri⁴ resume:

- a) “Permitir a una persona encontrar un libro del cual conoce el autor, el título o el tema.
- b) Mostrar qué posee la biblioteca sobre un autor, un tema o un género literario determinados.
- c) Ayudar en la elección de un libro (por la edición bibliográfica) o por su carácter (literario o temático)”.

Ahora bien, desde el punto de vista bibliotecológico, y a manera de contextualizar el párrafo arriba citado y con miras a adaptarlo a nuestro presente, en primer lugar ya no hablamos de catálogo como tal y tampoco solamente de libros, tampoco hablamos del mismo perfil y comportamiento de usuarios que solían acudir físicamente a la biblioteca en busca de satisfacer su necesidad de información. A esto se añade que ahora, además de libros, existen otros recursos de información que van de lo impreso a lo digital.

En suma, el usuario de información en la actualidad recupera información desde cualquier medio, desde cualquier lugar geográfico y en el momento que lo necesite, además obtiene grandes cantidades de información sin costo alguno, pues con una conexión a internet le basta para navegar en un mar de recursos de información. Sin embargo, lo que aún sigue vigente es que cuando un usuario requiere de información la busca principalmente por cualquiera de los tres puntos propuestos por Cutter (1876), y sus preferencias de elección

3 Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, Massachusetts: The MIT Press, 2009), 3-4.

4 Graciela Spedalieri, Los objetivos del catálogo. *Información, cultura y sociedad*, (15), 51-69, consultado abril 28, 2024.
http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17402006000200004&lng=es&tlng=es

las basará de acuerdo con su perfil y gustos concretos: libros, videos, audios, etcétera.

En este sentido es que los modelos conceptuales de organización de la información, como FRBR, nos remiten no a la función del catálogo sino a las tareas del usuario; Tillet⁵ las enuncia así:

Encontrar: Involucra satisfacer los criterios de búsqueda del usuario a través del atributo o la relación de una entidad.

Identificar: Permite al usuario confirmar que ha encontrado lo que buscaba diferenciando entre recursos similares.

Seleccionar: Involucra satisfacer los requerimientos del usuario con respecto al contenido, formato físico rechazar sino satisface las necesidades del usuario.

Obtener: Permite al usuario adquirir una entidad mediante compra, préstamo, etcétera o acceso electrónico remoto.

A lo arriba expuesto, Svenonius⁶ sostiene que vale la pena considerar dos características en particular con respecto de los sistemas bibliográficos: el primero de ellos son las soluciones que estos sistemas han aportado a la problemática que dificulta el acceso eficiente a la información. Algunos problemas de acceso son causados por las nuevas tecnologías, otros se deben a la variedad de información, a las múltiples facetas de sus usuarios y a las anomalías que caracterizan el lenguaje de recuperación. Por ejemplo, si los usuarios buscan en los estantes o en internet, algunos recuperarán demasiado, otros muy poco y algunos ni siquiera serán capaces de formular peticiones adecuadas de búsqueda. La segunda característica que hace que los sistemas bibliográficos tradicionales continúen siendo dignos de interés es la visión expresada en sus ideologías, pues la eficacia de un sistema para organizar la información es en parte función de una ideología que establece las ambiciones de sus creadores y lo que esperan lograr. Los sistemas producidos durante la segunda mitad del siglo XIX fueron sistemas ambiciosos, diseñados para satisfacer las necesidades de información de los usuarios más exigentes, ya que un buen diseño de sistemas comienza postulando objetivos visionarios, aunque solo sea para hacer conscientes a los usuarios del grado sobre el cual se compromete el sistema.

5 Barbara Tillet, *¿Qué es FRBR?: Un modelo conceptual del universo bibliográfico* (Washington, DC: Cataloging Distribution Service). Consultado mayo 2, 2024, <https://www.loc.gov/catdir/cpsd/Que-es-FRBR.pdf>

6 Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, Massachusetts: The MIT Press, 2009), 2-3.

Los sistemas bibliográficos del pasado reflejan lo que se puede lograr mediante la organización inteligente de la información.

En suma, para hacer frente a la gran variedad de recursos de información, a los múltiples comportamientos informativos de los usuarios y a la gran cantidad de conocimiento que debe ser organizado, es necesario conocer y comprender los fundamentos de la organización de la información, así como ciertas características de los sistemas bibliográficos tradicionales y las herramientas normativas, ya que este marco conceptual sienta las bases para organizar y recuperar la información digital emanada por las diferentes vías, como son el mercado editorial, el acceso abierto y, en general, todos los recursos de información electrónicos que circulan en la red.

II. RECURSOS DE INFORMACIÓN

En la sociedad actual, la información se ha convertido en un recurso de vital importancia al grado de que ya no solo es una herramienta o un bien imprescindible en todos los campos del conocimiento, sino que prácticamente forma parte de la cotidianidad del ser humano en todas las etapas de su vida, incluso desde antes de nacer. La información genera conocimiento y esto a su vez abre de manera casi infinita líneas de investigación que se convierten en nuevos conocimientos que se plasman en una variedad de soportes documentales tradicionales y digitales, mismos que el ser humano consulta, lee, analiza o consume⁷ en cada ámbito de su vida.

En este punto vale la pena retomar el concepto de recursos de información o, como antes se les llamaba, fuentes de información. A este respecto, Moscoso⁸ afirma que los recursos de información son una evolución del concepto

7 En la actualidad se habla de consumismo y sociedades consumistas. La RAE define al consumismo como: “Tendencia inmoderada a adquirir, gastar o consumir bienes, no siempre necesarios”. También se afirma que vivimos en una sociedad de consumo, por lo que el concepto *consumo de información* o *consumismo informacional* no queda ajeno a la bibliotecología. De acuerdo con Martínez Gemma (2003) el consumo de información se refiere a: “La acción de buscar y recuperar datos e información con la finalidad de dar respuesta a una necesidad o interés informativo” (sección 2). Derivado de esto y, sin ánimo de ahondar en ello, surgen conceptos como Infoxicación e Infobesidad, que tienen relación directa con el consumo inmoderado de información y que, por supuesto, tiene serias consecuencias en el ser humano de cualquier edad.

8 Purificación Moscoso Castro, “Reflexiones en torno al concepto recurso de información”, *Revista General de Información y Documentación* 8, n.º 1 (Madrid: Universidad Complutense, 1998), 328-329, consultado mayo 14, 2024, <https://www.google.com/search?client=safari&rls=en&q=purificacion+moscoso+reflexiones&ie=UTF-8&oe=UTF-8>

fuentes de información, y los define como: “medios y bienes que permiten adquirir, precisar o comunicar conocimientos con el fin de resolver una necesidad o llevar a cabo una empresa; representan la herramienta principal en el acceso y recuperación de la información”. Otros autores aluden a que el concepto fuentes de información se refiere a los recursos tradicionales, mientras que recursos de información se refiere más bien a los soportes digitales o electrónicos. Tal distinción se considera que ahora ya no es necesaria, pues la información se concibe en cualquier soporte de información ya sea este impreso, audiovisual o digital, por lo que para efectos de este trabajo se tomarán como sinónimos tanto recursos de información como fuentes de información.

Por lo tanto, los recursos de información o fuentes de información son aquellos materiales y documentos impresos, audiovisuales y de carácter digital o electrónico cuyo contenido está representado por texto, imagen o multimedia, y que han sido publicados por los diversos sectores editoriales que Frania Hall⁹ propone: sector de obras de referencia, sector del mercado académico, sector del mercado educativo y sector del mercado de consumo. Adicionalmente, dentro de estos sectores, Sánchez Luna¹⁰ señala que también se ubican aquellos productores de información que han encontrado en el acceso abierto alternativas de publicación y difusión de contenidos como son: editoriales independientes, organizaciones no gubernamentales (ONGs), organismos internacionales y organismos gubernamentales, incluso lo que actualmente es muy común, la autopublicación (*self publishing*).

Ahora bien, el universo bibliográfico actual deambula entre lo tradicional y lo digital, por lo que los recursos de información se pueden clasificar en primer lugar:

- a) Recursos de información impresos.
- b) Recursos de información audiovisuales.
- c) Recursos de información digitales.

9 Frania Hall, *El negocio de la edición digital* (México: Fondo de Cultura Económica, 2014), 65-238.

10 Blanca Estela Sánchez Luna, *El usuario como elemento central del desarrollo de colecciones en el entorno digital* (Tesis de Doctorado, UNAM, Facultad de Filosofía y Letras 2022), 102, consultado mayo 28, 2024, https://tesiunam.dgb.unam.mx/F/DMDR7V4UPVFEB1NQYASJIUHCRY6CQXTYT7JSAPFG431XKLNUQD-10837?func=find-b&local_base=TES01&request=Sanchez+luna+blanca+estela&find_code=WAT&adjacent=N&filter_code_2=WYR&filter_request_2=&filter_code_3=WYR&filter_request_3=

De lo anterior se desprende la siguiente tipología documental digital:

- a) Libro electrónico.
- b) Revista electrónica.
- c) Obras de referencia.
- d) Recursos educativos abiertos.
- e) Repositorios.
- f) Datos abiertos.
- g) Redes sociales (Facebook, Twitter, Instagram, Tik Tok, Whatsapp, Telegram).
- h) Recursos de información derivados de la inteligencia artificial (IA) como chat GPT.

Como se puede observar, las tendencias en materia de recursos de información representan un fenómeno que aumenta de forma acelerada, por lo que el bibliotecólogo actual debe estar preparado para incluirlos dentro de su proceso de desarrollo de colecciones, el cual implica considerar colecciones equilibradas temáticamente, de actualidad y con variedad en soportes documentales, sobre todo si se considera que estamos hablando de bibliotecas híbridas. En tal contexto, la organización documental cobra un sentido especial, ya que aunque los fundamentos ayudan a mantener los principios que sostienen los preceptos de Cutter (1876), las 5 famosas leyes de Ranganathan (1931)¹¹ y las tareas del usuario (2004), siempre será necesario conocer y dominar el entorno digital en el que se produce el conocimiento, el mercado editorial y las alternativas por las cuales se pueda ofrecer al usuario la información que necesita y en el tiempo que la requiere. Todo esto sin descuidar la fiabilidad y veracidad de los recursos de información que se proporcionan al usuario.

III. ORGANIZACIÓN DE LA INFORMACIÓN EN LA ERA DIGITAL

La organización de la información es una disciplina intelectual que se relaciona con actividades como la descripción, indización y clasificación, con la finalidad de proporcionar sistemas de representación y así ordenar la información y el conocimiento.

11 Las leyes son: 1) Los libros son para usarlos; 2) A cada lector su libro; 3) A cada libro su lector; 4) Ahorra el tiempo del lector; 5) La biblioteca es un organismo en crecimiento.

En palabras de Glushko,¹² un sistema de organización es una colección de recursos intencionalmente ordenada. Añade que es posible organizar objetos físicos y, también ahora, objetos digitales; en este contexto, los catálogos de las bibliotecas digitales, los portales web de agregadores de contenido organizan los enlaces que llevan a otros recursos digitales, por lo tanto, los motores de búsqueda web realizan análisis de contenido y enlaces junto con calificaciones de relevancia y de ahí organizar billones de páginas web que compiten por la atención del usuario.

De acuerdo con Chowdhury,¹³ en una biblioteca híbrida los profesionales de la información deben estar preparados para organizar y describir recursos de información tanto impresos como digitales. Además de considerar que cada organización/institución ahora produce una gran cantidad de información digital y que sus actividades diarias dependen en gran medida del acceso y uso de los recursos de información digital adecuados. Esto refuerza la necesidad de una organización adecuada de los recursos de información institucionales para un mejor acceso y gestión. De hecho, en los últimos años han surgido muchos enfoques, herramientas y nuevas técnicas para tales actividades, por lo que los bibliotecólogos deben estar preparados para organizar la información en los diferentes dominios, el tradicional, el entorno web y el entorno digital.

En este contexto se habla del proceso de organización de documentos con características peculiares, pues son documentos que no necesariamente residen en la biblioteca, incluso ni el servidor de la misma, pues como se ha mencionado, desde hace ya varias décadas la información y el conocimiento se socializan y se publican desde una gran variedad de fuentes de información y muchos de ellos, incluso, residen en la llamada nube. Por esta razón es que la tarea de organización documental adquiere nuevos enfoques que el profesional de la información debe considerar. A saber estos enfoques pueden ser:

Naturaleza de la información digital ubicua y efímera

Los recursos de información ya no se encuentran estáticos en un soporte como puede ser el libro, por lo tanto la ubicuidad de la información consiste en la disponibilidad de la misma desde cualquier lugar; es decir, la información y el conocimiento pueden estar en todas partes y al mismo tiempo. Esto obliga al documentalista a desarrollar estrategias para garantizar al usuario la disponibilidad y acceso seguro a los recursos de información.

12 Robert J. Glushko, *The Discipline of Organizing* (Cambridge, Massachusetts: The MIT, 2013), 2.

13 G.G. Chowdhury y Sudatta Chowdhury, *Organizing Information: From the shelf to the Web* (London: Facet, 2007), xiii.

Susceptibilidad de la información

La información digital al no estar en un medio o soporte fijo puede ser sensible y vulnerable, por lo que se ha visto la necesidad de desarrollar candados electrónicos o bardas electrónicas o barreras tecnológicas que establecen permisos de acceso y garantizan que solo usuarios autorizados accedan a la información. Esta iniciativa lo que busca es proteger los derechos de autor y de propiedad intelectual. Este aspecto es muy importante, ya que garantiza que el sistema de información, con recursos de información organizados en diferentes entornos, cuente con la protección necesaria para que su consulta y descarga por el usuario sea bajo lo que legalmente se establece por los derechos de autor del país.

Conocimiento y manejo de los modelos y estándares para organizar la información digital

Uno de los mayores retos que enfrentan las bibliotecas y los profesionales de la información es encontrar la mejor forma de organizar los recursos de información disponibles, a partir de sus formatos y modos en los que se presentan hasta el día de hoy. En este sentido, el documentalista o catalogador debe conocer y dominar las herramientas y técnicas, incluso conocer la literatura que le permitirá entender la complejidad y vastedad del universo bibliográfico digital junto con sus enlaces y relaciones que se producen entre entidades de información. Actualmente, existen varios enfoques que apoyan en el diseño de los sistemas de información a organizar y facilitar el acceso a la información y conocimiento.

Metadatos

Ontologías

Arquitectura de la información

Web semántica

Datos enlazados o datos vinculados (*linked data; linked data flavors*)

Modelos conceptuales como LRM (*Library Reference Model*), BIBFRAME; modelos de intercambio de datos en la web como RDF.

La norma RDA (*Resource Description Model*)

En suma, los bibliotecólogos dedicados a la organización de la información deben estar a la vanguardia con los temas arriba mencionados, con el objetivo de estar preparados para garantizar el acceso a la información dentro de una infraestructura tecnológica adecuada. Deben responder a las necesidades de información de su comunidad de usuarios a partir de los preceptos de la organización documental.

Almacenamiento y preservación digital de la información

En la actualidad existen una diversidad de modelos para garantizar la accesibilidad y disponibilidad de los recursos de información, así como lo que representa una infraestructura tecnológica adecuada. Para la UNESCO,¹⁴ la preservación digital se define como

los procesos destinados a garantizar la accesibilidad permanente de los objetos digitales, y se define como el conjunto de los procesos destinados a garantizar la continuidad de los elementos del patrimonio digital durante todo el tiempo que se consideren necesarios. El objetivo de la preservación de los objetos digitales es mantener su accesibilidad, es decir, la capacidad de tener acceso a su mensaje o propósito esencial y auténtico.

En este punto vale la pena destacar lo que Sánchez Luna¹⁵ comenta al respecto: “esfuerzos encaminados, algunos por parte de universidades, para administrar recursos electrónicos, que sirven como medios de almacenamiento y acceso; y además que funcionan como medio de preservación, de aquí surgen sistemas como DSpace, Portico, LOCKSS (*Lots of copies Keeps Stuff Safe*), CLOCKSS, PANDORA.”

Comunidad de usuarios

El conocimiento de la comunidad implica conocer a la sociedad red, identificar sus comportamientos de búsqueda ante la necesidad de información y, por lo tanto, implica conocer sus preferencias en cuanto a soportes documentales. En este sentido los estudios de usuarios permiten identificar la tipología de usuarios de acuerdo con la generación en la que se enfrentaron por vez primera a los medios digitales y en qué circunstancias. Se trata de recolectar información de la comunidad para conocer tendencias con el fin de identificar y resolver problemáticas en la búsqueda y recuperación de información. Para lograr lo anterior se requiere de un adecuado sistema de información, que permita identificar escenarios futuros y áreas de oportunidad y así tomar de-

14 Biblioteca Nacional de Australia, *Directrices para la preservación del patrimonio digital* (París: UNESCO, División de la Sociedad de la Información, 2003).

15 Blanca Estela Sánchez Luna, *El usuario como elemento central del desarrollo de colecciones en el entorno digital* (Tesis de Doctorado, UNAM, Facultad de Filosofía y Letras, 2022), 225, consultado mayo 28, 2024, https://tesiunam.dgb.unam.mx/F/DMDR7V4UPVFEB1NQYASJIUHCRY6CQXTYT7JSAPFG431XKLNUQD-10837?func=find-b&local_base=TES01&request=Sanchez+luna+blanca+estela&find_code=WAT&adjacent=N&filter_code_2=WYR&filter_request_2=&filter_code_3=WYR&filter_request_3=

cisiones estratégicas con respecto de la organización de la información digital. Se trata de proporcionar a cada usuario el contenido que requiere o desea en el medio perfecto, de acuerdo con sus preferencias, ya sean estos medios tradicionales, medios digitales o ambos.

Cooperación bibliotecaria y consorcios de bibliotecas

Finalmente, otro aspecto a destacar en este tema de la organización documental digital es la importancia de la cooperación bibliotecaria desde los consorcios de bibliotecas, ya que como Pérez Aarranz señala, “han surgido ante las dificultades para acceder a los cada vez más numerosos, costosos y diversos nuevos recursos de información en formato electrónico”, por lo tanto, los consorcios representan un apoyo para las bibliotecas en cuanto a compartir recursos de información, alternativas de organización y almacenamiento, diseño de políticas, optimización de presupuesto e incluso modelos de negociación con el mercado editorial digital.

CONCLUSIONES

La constante evolución tecnológica en materia de información es un reto al que se enfrenta el bibliotecólogo, ya que cada día surgen nuevos recursos de información digital que se publican tanto de manera tradicional como a través de las nuevas alternativas, como el acceso abierto, la autopublicación o a través de plataformas digitales. Por lo anterior, los sistemas de información deben anticiparse y, prácticamente, predecir el futuro para mantener la vigencia de las bibliotecas dentro de una sociedad en donde la información se obtiene en grandes cantidades, de manera rápida y desde cualquier lugar. Por lo tanto, el rol del documentalista radica en reconocer los fundamentos intelectuales de la organización de la información para de ahí replantear y rediseñar nuevas formas de organizar la información desde un entorno web de datos enlazados, con la finalidad de brindar a los usuarios información pertinente y relevante en diferentes soportes documentales. Se trata de que el usuario pueda navegar por el mar de información impresa, audiovisual y digital, y establecer relaciones entre las diversas entidades de la información.

Por todo lo arriba expuesto, se finiquita que existen una gran cantidad de argumentos que dotan a la organización de la información y el uso de herramientas y estándares de un valor incalculable, ya no solo en el ámbito bibliotecológico sino desde el ámbito mundial, que reconoce que la información y el conocimiento son bienes a los que todos los ciudadanos deben tener derecho. Por lo tanto, un adecuado sistema de organización garantiza el acceso a los

recursos de información ya no solo de la biblioteca sino a los recursos de la red que valen la pena ser recuperados por los usuarios de la información.

REFERENCIAS BIBLIOGRÁFICAS

- Biblioteca Nacional de Australia, *Directrices para la preservación del patrimonio digital*, París: UNESCO, División de la Sociedad de la Información, 2003.
- Chowdhury, G.G. y Chowdhury, Sudatta, *Organizing Information: From the shelf to the Web*, London: Facet, 2007.
- Glushko, Robert J., *The Discipline of Organizing*, Cambridge, Massachusetts: The MIT Press, 2013.
- Hall, Frania, *El negocio de la edición digital*, México: Fondo de Cultura Económica, 2014.
- Martínez, Gemma, *El consumo de información: Una primera aproximación al concepto*, Barcelona: Universitat Pompeu Fabra, 2003, <https://arxiu-web.upf.edu/hipertextnet/numero-1/consumo-info.html>
- Moscoso Castro, Purificación, Reflexiones en torno al concepto recurso de información, *Revista General de Información y Documentación* 8, n.º 1, Madrid: Universidad Complutense, 1998, <https://www.google.com/search?client=safari&rls=en&q=purificacion+moscoso+reflexiones&ie=UTF-8&oe=UTF-8>
- Pérez Aarranz, Fernando, Los consorcios de bibliotecas universitarias españolas en el ámbito de la edición electrónica, *Boletín de la Asociación Andaluza de Bibliotecarios* (22, 90-91, enero-junio), 2008, 11-28, <https://www.redalyc.org/pdf/353/35313089002.pdf>
- Sánchez Luna, Blanca Estela, *El usuario como elemento central del desarrollo de colecciones en el entorno digital*, Tesis de Doctorado, UNAM, Facultad de Filosofía y Letras, 2022, https://tesiunam.dgb.unam.mx/F/DM-DR7V4UPVFEb1NQYASJIUHCRY6CQXTYT7JSA-PFG431XKLNUQD-10837?func=find-b&local_base=TES01&request=Sanchez+luna+blanca+estela&find_code=WAT&adjacent=N&filter_code_2=WYR&filter_request_2=&filter_code_3=WYR&filter_request_3=

Spedalieri, Graciela, *Los objetivos del catálogo. Información, cultura y sociedad* 15, 51-69, http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17402006000200004&lng=es&tlng=es

Svenonius, Elaine, *The Intellectual Foundation of Information Organization*, Cambridge, Massachusetts: The MIT Press, 2009.

Tillet, Barbara, *¿Qué es FRBR?: Un modelo conceptual del universo bibliográfico*, Washington, DC: Cataloging Distribution Service, <https://www.loc.gov/catdir/cpsol/Que-es-FRBR.pdf>

Organización del conocimiento. Bibliotecología, Terminología e Información. Instituto de Investigaciones Bibliotecológicas y de la Información/UNAM. Edición digital. Coordinación editorial: Sergio Javier Sepúlveda H. y Angélica Valenzuela; revisión especializada: Diana Martínez Aguilar; corrección de pruebas: Carlos Ceballos Sosa y Marcos Emilio Bustos Flores; formación editorial y portada: Carlos Papaqui Landeros. Se terminó en febrero de 2025.