

# METADATOS SOCIALES: iniciativas, tecnologías, aplicaciones y softwares



Ariel Alejandro Rodríguez García  
COORDINADOR



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.

---

---

**Metadatos sociales: iniciativas, tecnologías,  
aplicaciones y softwares**

COLECCIÓN  
METADATOS

Instituto de Investigaciones Bibliotecológicas y de la Información



---

---

**Metadatos sociales: iniciativas, tecnologías,  
aplicaciones y softwares**

**Coordinador**

**Ariel Alejandro Rodríguez García**



**Universidad Nacional Autónoma de México  
2024**

**Z666.7**  
**M48**

Metadatos sociales : iniciativas, tecnologías, aplicaciones y softwares / Coordinador Ariel Alejandro Rodríguez García. - México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2024.

xv, 293 p. - (Metadatos)  
ISBN: 978-607-30-8624-0

1. Metadatos. 2. Datos vinculados. 3. Indización - Aspectos sociales. 4. Tecnología de la información - Aspectos sociales. 5. Contenidos generados por los usuarios. I. Rodríguez García, Ariel Alejandro, coordinador. II. ser.

Diseño de cubierta: Mario Ocampo Chávez  
Imagen: Kishore Newton - stock.adobe.com

Primera edición: Mayo de 2024  
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: 978-607-30-8624-0

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

Impreso y hecho en México

# Contenido

INTRODUCCIÓN .....	ix
--------------------	----

## INICIATIVAS

CIUDADANÍA Y <i>DATAFICACIÓN</i> : EL ANÁLISIS SOCIOLÓGICO EN EL CONTEXTO DE LA INFORMACIÓN DIGITAL .....	3
Alejandro Ramos Chávez	

LOS METADATOS EN LOS PLANES DE ESTUDIO DE LOS GRADOS EN INFORMACIÓN Y DOCUMENTACIÓN: UN ENFOQUE COMPARATIVO ENTRE PORTUGAL Y ESPAÑA .....	17
Ana Lúcia Terra	

METADATOS EN LA FORMACIÓN PROFESIONAL EN CC II .....	33
Julio César Rivera Aguilera	
Luis Roberto Rivera Aguilera	
Brenda Lucero Campos Monreal	

FLUJO DE INFORMACIÓN Y USUARIOS DE REDES SOCIALES UNIVERSITARIAS: CARACTERÍSTICAS, PERFILES, NECESIDADES E IMPACTOS EN LA ORGANIZACIÓN .....	57
Marco Brandão	

## TECNOLOGÍAS

METADATOS PARA DOCUMENTOS FÍLMICOS: INICIATIVAS Y ESTÁNDARES .....	77
Hilda Gabriela Lobatón Cruz	

CURACIÓN DE METADATOS PARA RECURSOS EDUCATIVOS DIGITALES .....	91
Ana Carolina Simionato Arakaki	

METADATOS BIBLIOGRÁFICOS Y METADATOS SOCIALES: CONEXIONES EN ENTORNOS DE DATOS VINCULADOS .....	113
Fabiano Ferreira de Castro	

METADATOS Y SEGURIDAD DE LA INFORMACIÓN: DESAFÍOS Y SOLUCIONES .....	129
Javier Moncayo García	

## APLICACIONES

METADATOS SOCIALES Y PRESERVACIÓN DIGITAL: CINCO RETOS PARA LAS INSTITUCIONES DE LA MEMORIA .....	147
Arien González Crespo	

LOS RETOS DE LOS RECURSOS EDUCATIVOS ABIERTOS Y SU CATALOGACIÓN: CREACIÓN DE METADATOS PROFESIONALES Y SOCIALES .....	175
Alma Beatriz Rivera Aguilera	
Elisa Cruz Rojas	
María Guadalupe Barrera Galán	

EL SENTIDO SOCIAL DEL DATO CIENTÍFICO GENERADO POR LA BIBLIOTECA UNIVERSITARIA DESDE LA PRÁCTICA DE LA DIVULGACIÓN ACADÉMICA .....	199
Luisa Coral Acosta Cruz	

LA CATALOGACIÓN SOCIAL, SU PRÁCTICA PROFESIONAL Y EMPÍRICA .....	213
Ariel Alejandro Rodríguez García	

## SOFTWARES

METODOLOGÍA PARA ESTABLECER RELACIONAMIENTO AUTOMATIZADO DE PATRONES COMUNES EN TESTIMONIOS ESCRITOS DE VÍCTIMAS DEL CONFLICTO ARMADO EN COLOMBIA .....	231
Fabián Orlando Baena Henao	



MODELOS Y TECNOLOGÍAS PARA LA VISUALIZACIÓN DE ONTOLOGÍAS TERMINOLÓGICAS EN EL CONTEXTO DE LA WEB SEMÁNTICA .....	243
Adriana Suárez Sánchez	
EL OBJETO VIRTUAL DE APRENDIZAJE (OVA) COMO PRODUCTO DE APROPIACIÓN SOCIAL DEL CONOCIMIENTO DEL BANCO DE DATOS TERMINOLÓGICOS DE LAS CIENCIAS DE LA INFORMACIÓN .....	265
María Teresa Múnera Torres	
APRENDIZAJE MÁQUINA EN LA BIBLIOTECOLOGÍA .....	277
Guadalupe Vanessa Carolina Gutiérrez Hernández Jorge Gómez Briseño	

## Introducción

**D**esde que la Organización Mundial de la Salud nos alertó sobre la crisis sanitaria mundial que se avecinaba, como humanidad comenzamos a cambiar a pasos agigantados. Unos cambios fueron buenos, otros no tanto, pero todos ellos encaminados a modificar nuestro modo de pensar y actuar, porque entendimos que todo lo que conocíamos ya era pasado y, de repente, nos encontrábamos en un mundo completamente nuevo.

El motor de cambio en este siglo XXI se ha acogido de distintas maneras, unas más generales y otras más específicas, tales como la transformación digital, el futuro digital, la era digital o la era del acceso. Pero en todas esas transformaciones hay varios denominadores comunes como son las Tecnologías de la Información y la Comunicación, el internet, los datos y la web.

En el término de *transformación digital* se hace referencia a la reinención de las operaciones de los negocios a través de la explotación de las tecnologías de la información y las facilidades que se otorgan en la cadena laboral por medio de la colaboración. Mientras que en el término de *futuro digital* se señala que los cambios en la clase media, el trabajo y educación han mejorado continuamente nuestro acceso a la información, las capacidades para

conectarnos entre nosotros y las nuevas capacidades para concebir e implementar soluciones. En tanto que en *La era del acceso* Rifkin menciona que “las ideas serán los objetos con los que comercien las instituciones, y las personas comprarán cada vez más acceso a esas ideas y a las formas físicas en las que se encaramen esas ideas. La habilidad para controlar y vender pensamientos es la última expresión de las nuevas capacidades comerciales”.<sup>1</sup>

El cambio digital ha permitido la expansión de la información y potenciado la convivencia y comunicación tradicional. Hoy en día, por ejemplo, ya no se puede explicar y enseñar como en los tiempos en donde “todo el conocimiento venía por la vía de la escuela. La enseñanza en general consistía en eso: en informar a los niños de las cosas que no sabían”,<sup>2</sup> de modo que tampoco será posible seguir con el esquema de uno-a-uno o de uno-a-muchos o de muchos-a-uno porque la comunicación web se sustenta en el principio de la conectividad simultánea de mucho-a-muchos.

En relación con el tema que engloba este libro, *Metadatos sociales: iniciativas, tecnologías, aplicaciones y softwares*, se puede observar en los distintos capítulos que lo integran, que la participación de la sociedad y las personas será cada vez mayor, debido a que habrá nuevas formas de convivir digitalmente.

Dicho lo anterior, los metadatos sociales juegan un rol importante en el entorno digital y, particularmente, en el soporte que se debe dar a la finalidad de las bibliotecas, archivos y museos, que es proporcionar la localización y descubrimiento de los recursos de información tanto por los usuarios como el personal que los genera. Es decir, los metadatos sociales son el auxiliar invisible que sirve para conectar a los usuarios para que interactúen con la información digital.

La creación de los metadatos sociales, en teoría, es un proceso continuo en el que se encuentran involucrados creadores, editores, bibliotecarios, usuarios, así como programadores y webmasters. De ahí que en los tiempos actuales se esté hablando de los

---

1 Jeremy Rifkin, *La era del acceso*, 84.

2 Fernando Savater, *Ética de urgencia*, 31.

metadatos enriquecidos o también referenciados como los metadatos socialmente construidos.

La tesis principal de esta obra es enfatizar que los metadatos sociales se han venido construyendo desde hace varias décadas, pero sólo eran examinados desde la perspectiva del creador profesional y se dejaba de lado a quienes carecían de un conocimiento formal o especializado, pero aportaban significativamente al enriquecimiento en los contenidos de los recursos de información. Dicho de otra manera, el etiquetado social en su forma más común se viene practicando desde finales del siglo pasado, pero toma un gran impulso cuando se empiezan a utilizar las aplicaciones de la web 2.0. De manera que, con ello, han surgido también una gran cantidad de estudios que han contribuido al análisis del etiquetado y los sistemas de etiquetación de contenidos.

En definitiva, para el momento que vivimos tanto los profesionales que crean los metadatos como los usuarios quienes los usan y enriquecen, están cada vez más familiarizados con los distintos sistemas de etiquetación social, lo que repercute en hacer más fácil la colaboración y seguimiento de acciones relativas a lo que se conoce como navegación social. De ahí que la composición de este libro responda a conocer las iniciativas, las tecnologías, aplicaciones y softwares que permitan al lector introducirse o profundizar en el estudio de los metadatos sociales y formarse un criterio con relación al metadato socialmente enriquecido.

En el primer apartado, las *Iniciativas*, el capítulo inicial parte la discusión teórica del concepto de la sociología de la información haciendo énfasis en lo que se ha denominado como masificación a través de las tecnologías de la información, y cómo el valor de los datos permite generar productos y servicios basados en datos proporcionados por los propios individuos, mediante la interacción con los medios digitales, conocidos también bajo el neologismo *dataficación*. En el segundo capítulo se presenta un estudio comparado sobre los planes de estudio de los grados en Ciencias de la Información y Documentación en Portugal y España. En él se analizan las asignaturas relacionadas con la temática de metadatos, y se halla que los metadatos han introducido nuevos enfoques en



la estructura y sintaxis de la descripción de los recursos de información, así como la identificación de tendencias en la organización de la información. En el tercer capítulo se muestra un estudio sobre el plan de estudios de Gestión de la Información y Gestión Documental y Archivística y se aborda el tema de los metadatos desde las competencias de la formación profesional en dicha licenciatura, para lo cual se encontró que es necesario fortalecer los tópicos de los metadatos en los planes y programas de estudios de las instituciones de educación superior en México, porque está siendo un tema de actualidad y relevancia para la formulación del perfil profesional en Ciencias de la Información. El capítulo final de este apartado presenta los resultados de una investigación que se encargó de analizar las redes sociales Facebook, X (antes Twitter) y Youtube con el propósito de encontrar el perfil e impacto de estas redes en el logro de los objetivos institucionales y la satisfacción de las necesidades de información de los internautas que las siguen.

En el segundo apartado, *Tecnologías*, el capítulo con el que se inicia resalta la importancia y necesidad de normalización de los datos para hacerlos interoperables entre las distintas instituciones que resguardan el documento fílmico, con el firme propósito de que este tipo documental cumpla con estándares de análisis de contenido y las posibilidades de recuperación de información cinematográfica por parte de los usuarios. El siguiente capítulo trata el tema de la curación de datos para recursos educativos digitales y hace énfasis en que la curaduría digital a través de los metadatos puede maximizar los procesos de recuperación, representación, organización, acceso, uso, reutilización y preservación de estos recursos. Además, los resultados de este estudio identificaron un proceso de clasificación, evaluación y recuperación para los recursos educativos. En contraste con lo anterior, el siguiente capítulo nos invita a hacer una revisión a la evolución que han tenido los metadatos bibliográficos en el entorno de los datos vinculados, con la finalidad de encontrar hallazgos que permitan identificar en qué momento los metadatos sociales son introducidos en la práctica común entre los metadatos introducidos por un profesional y

aquellos que provienen de las contribuciones de los usuarios, llegando a la conclusión de que los metadatos se construyen socialmente y requieren de enfoques orientados a la comunidad y la colaboración social. El capítulo final de este apartado nos refiere al estudio de los metadatos y la seguridad de la información. En él se indica que la presencia de datos masivos en el entorno web presenta un desafío tanto para los usuarios como para las organizaciones en materia de seguridad, y propone que algunos agentes reguladores en materia de privacidad y ciberseguridad deben usar nuevas tecnologías para el análisis y limpieza de metadatos.

En el tercer apartado, *Aplicaciones*, el capítulo con el que se inicia hace referencia a los metadatos sociales y los cinco retos para las instituciones de la memoria y señala que estos metadatos son un conjunto de aplicaciones basadas en la web 2.0, y que entre sus características fundamentales se encuentra la gran cantidad de hipervínculos externos y diversos tipos de información embebida que fomenta el intercambio de información entre quien la genera y quien la consume. El capítulo siguiente investiga sobre la creación de metadatos profesionales y sociales aplicados a los recursos educativos abiertos, y se señala que la enorme complejidad de este tipo de recursos, así como su diversidad y formatos involucrados ha permitido documentar políticas internas, homogeneizar los registros y establecer estrategias de formación en la asignación de los metadatos sociales. Posteriormente, se encuentra un capítulo sobre el sentido social, que nos reseña que los contenidos de divulgación científica permiten abrir un abanico de posibilidades a la población en general porque hacen entendible la información proveniente de la comunidad científica. El apartado *Aplicaciones* cierra con un capítulo sobre la catalogación social, su práctica profesional y empírica, en donde se infiere que este tema está en auge debido a los cambios experimentados en el catálogo en línea por la incorporación de las tecnologías de la web 2.0, así como los nuevos modelos de producción, distribución y consumo de la información. Arguye que los sistemas de etiquetación y práctica social colaborativa han propiciado que la práctica catalográfica debe encaminarse hacia la creación del principio del metadato construido socialmente.

El cuarto apartado y final de este texto, *Softwares*, inicia con un capítulo sobre las metodologías automatizadas de patrones comunes a través de algoritmos de *Machine Learning*, Procesamiento de Lenguaje Natural y Minería de Texto, utilizadas para transcribir los testimonios escritos de las víctimas del conflicto armado en Colombia, y donde se concluye que con esta metodología se crearon herramientas que facilitaron el entendimiento de la configuración de la guerra en Colombia. El siguiente capítulo se concentra en el estudio de modelos y tecnologías para la visualización de vocabularios controlados asociados al contexto de la web semántica, y concluye que en la última década las tecnologías para la visualización de tecnología se han incrementado debido a las ventajas que los grafos ofrecen en la exploración de los datos e información. El Objeto Virtual de Aprendizaje, tema del siguiente capítulo, nos da a conocer cómo un producto de apropiación social del conocimiento puede generar estrategias formativas, pedagógicas y de comunicación, a partir de la estructuración de la terminología especializada en las ciencias de la información y así crear una Base de Datos Terminológica en las áreas de la Archivística, Bibliotecología, Museología y Documentación. El capítulo con el que se cierra este apartado y el libro, se refiere al aprendizaje máquina en la Bibliotecología y nos invita a reflexionar sobre la importancia o no que tiene la acumulación de datos y aprovechar su potencial para la extracción de conocimiento y el valor de un conjunto de datos. Además, los autores mencionan que el aprendizaje máquina es un campo de estudio multidisciplinario en donde todas las disciplinas, incluyendo la Bibliotecología, trabajan de manera conjunta para obtener la mejor visión de los datos y con ello una mejor toma de decisiones.

Para finalizar esta introducción, deseo agradecer a quienes de manera directa o indirecta permitieron que esta obra llegue a sus manos y con ello conseguir que los temas emergentes en la actividad de los metadatos, en la Bibliotecología y Estudios de la Información, vayan formado un cuadro del antes y ahora en el desarrollo de la organización de la información. Se debe agregar que este agradecimiento se hace extensivo a las autoridades del

Instituto de Investigaciones Bibliotecológicas y de la Información, así como a los miembros del Seminario de Investigación Metadatos por su colaboración y participación.

*Ariel Alejandro Rodríguez García*



## INICIATIVAS

# Ciudadanía y *dataficación*: el análisis sociológico en el contexto de la información digital

ALEJANDRO RAMOS CHÁVEZ

*Universidad Nacional Autónoma de México, México*

## INTRODUCCIÓN

**E**n el presente trabajo se propone poner a discusión teórica al concepto de sociología de la información. En este sentido, convendría partir del supuesto de que todo análisis sociológico se da en un contexto donde la información, socialmente construida, permite no sólo el desarrollo de una sociedad determinada, sino también su análisis. En este punto se propone centrar el análisis en un tema que convendría ser explorado, relacionado con el análisis sociológico que se puede llevar a cabo en un contexto marcado por la masificación en la utilización de las tecnologías de la información y comunicación, así como por el aumento en el valor de los datos que permiten generar productos y servicios basados en información que van generando los propios individuos, mediante su interacción con los medios digitales, conocido también bajo el neologismo *dataficación*.

En este contexto, el presente capítulo está dividido en los siguientes apartados generales: en primer lugar, partiendo de una contextualización del uso del concepto de ciudadanía y ciudadanía digital, se propone cuestionar la pertinencia analítica del concepto

de sociología de la información, para lo cual se considera pertinente abordar su desarrollo, desde sus postulados hasta las críticas que pudiera tener su aplicabilidad en el análisis sociológico y de la información. En segundo lugar, abordar el análisis del concepto de *dataficación*, y algunas áreas específicas de su impacto (como la *dataficación* de la educación, la *dataficación* de la salud o la *dataficación* social), proponiendo como pieza nodal el escenario marcado por la gran influencia de las tecnologías de la información y comunicación, principalmente internet, no sólo para formar la opinión pública, sino para la construcción de la comunicación social y política en la actualidad. En tercer lugar, se propone abordar el concepto específico de *dataficación* en su vinculación directa con la ciudadanía. Se concluye con algunas reflexiones sobre la temática analizada.

#### CIUDADANÍA, CIUDADANÍA DIGITAL Y SOCIOLOGÍA DE LA INFORMACIÓN

Para entrar en el primer elemento tratado en este capítulo, se podría entender a la ciudadanía como el conjunto de individuos que, al habitar un mismo Estado-nación, tienen derechos y obligaciones que comparten. En este sentido, para alcanzar la plenitud enmarcada en la idea de ciudadanía, se debería generalizar las oportunidades para alcanzar, reclamar y hacer valer los derechos, así como responsabilizarse y cumplir de forma homogénea con las obligaciones. Tomado en cuenta esto, en la actualidad es cada vez más común encontrarnos con una utilización importante de las tecnologías de la información y comunicación para llevar a cabo estos procesos.

Desde esta perspectiva, gran parte de los derechos y las obligaciones enmarcadas en la idea general de ciudadanía son ejercidas en la actualidad a través de la utilización de herramientas tecnológicas, principalmente internet; por tanto, han surgido nuevos enfoques para entender y adaptar el concepto de ciudadanía. Tal es el caso del surgimiento del concepto de ciudadanía digital,

del cual se ha mencionado que “hace referencia a un nuevo paradigma, cuyos criterios orientadores dan cuenta del camino hacia un tipo de sociedad emergente y en construcción, resultante de la acción de los sistemas tecnológicos y de los progresivos procesos de digitalización”.<sup>1</sup> En otras palabras, la ciudadanía digital “se refiere a un tipo de individuo que ejerce sus derechos, obligaciones y en gran medida lleva a cabo su participación social, todos estos elementos enmarcados en la concepción de ciudadanía clásica, mediante la utilización de las TIC, como lo es internet”,<sup>2</sup> o como lo señala Arcila,<sup>3</sup> el ciudadano digital puede ser definido como aquel individuo “capaz de ejercer su ciudadanía bajo un panorama de interactividad ofrecido por una plataforma digital”.

Es oportuno mencionar que la ciudadanía digital se inscribe en el entendimiento de una ciudadanía diferenciada, la cual “hace referencia tanto a los derechos individuales de las personas, como al vínculo de pertenencia a una comunidad específica, no sólo limitándola a la idea genérica de nacionalidad, sino además de ello a un grupo más fácilmente identificable tal como lo son las mujeres, los jóvenes, los adultos mayores”<sup>4</sup> o los usuarios de las herramientas tecnológicas, para los efectos de esta presentación.

A la par de este desarrollo conceptual de la ciudadanía, la noción de la sociología de la información comienza a cobrar relevancia en diversos trabajos académicos. Si bien la información como construcción social de intercambio de ideas no podría escapar del análisis sociológico, pues no está desconectada del sistema social, su análisis no resulta tan simplista o banal como ha sido abordado al cuestionar si se trata de un fenómeno emergente o una realidad existente.<sup>5</sup> Buena parte del análisis sociológico de

---

1 Raúl Bernal-Meza y Gustavo Masera, “Sociedad de la información: etapa posterior de la globalización/mundalización...”, 94.

2 Alejandro Ramos, “Información y ciudadanía, una propuesta desde la gobernanza”, 605.

3 Carlos Arcila, “El ciudadano digital”, 18.

4 Alejandro Ramos, *op. cit.*, 119.

5 Manuel Alejandro Romero, “Sociología de la Información: ¿fenómeno emergente o realidad existente?...”.

la información, en la era contemporánea, se da con los análisis de Manuel Castells,<sup>6</sup> en el surgimiento de la nueva estructura social basada en la facilitación del comercio y la organización productiva a nivel global, gracias al desarrollo y a la apropiación social de las tecnologías, que propician la creación acelerada de formas de interacción, interconexión y comunicación social y económica. Convendría, no obstante, repasar brevemente también los aportes de Lyon, en Calhoun, Rojek y Turner,<sup>7</sup> quienes argumentan que

la información en sí misma requiere un análisis sociológico si queremos comprender su conexión con cuestiones cruciales, desde la identidad y la desigualdad hasta la materia y el significado. La información, en este contexto, es fundamental para las transformaciones sociales que comenzaron después de la Segunda Guerra Mundial y ahora son visibles en las sociedades globalizadas basadas en la alta tecnología en todo el mundo.

Todos estos planteamientos cobran relevancia al analizar ese nuevo contexto marcado por el desarrollo y uso tecnológico, donde a la par de derribarse las fronteras espaciotemporales, también van surgiendo novedosas formas de generar políticas, productos y servicios que impactan directamente el desarrollo de la noción de ciudadanía, y para los fines específicos de este trabajo, la ciudadanía digital.

Bajo este orden de ideas, se han propuesto conceptos como el de:

[...] innovación impulsada por datos, que se refiere a los cambios dados y fundamentados tanto por la exponencial producción, así como por los intercambios y asimilación de datos que permiten a los ciudadanos, los gobiernos y las empresas, tomar mejores decisiones con respecto a los asuntos de interés general, con objeto

---

6 Manuel Castells, *La sociedad red: una visión global*.

7 Craig Calhoun *et al.*, *The SAGE Handbook of Sociology*, 223.

de impactar positivamente en la economía de las sociedades, que, partiendo de lo local, puede tener efectos en un ámbito global.<sup>8</sup>

En este sentido, convendría tener en cuenta las discusiones teóricas de la innovación impulsada por datos y su vinculación con la temática de la *dataficción*, elementos puestos a consideración en el siguiente apartado del capítulo.

### *DATAFICCIÓN Y ALGUNAS ÁREAS DE SU INCIDENCIA (SALUD, EDUCACIÓN Y SOCIAL)*

En este punto es en donde se considera pertinente la inclusión del tema del neologismo *dataficción*, que en una primera instancia se relacionaría con la utilización de los datos obtenidos de la interacción de las personas en el mundo digital, por medio de su análisis y su reorganización, para la generación de servicios y productos que redunden, a final de cuentas, en la mejor toma de decisiones personales y colectivas. Cabe hacer la precisión que tanto en inglés como en español se suelen usar indistintamente tanto los términos *datafication (dataficción)*<sup>9</sup> o *datification (datificación)*<sup>10</sup> para dar cuenta del mismo fenómeno; sin embargo, en este trabajo

- 
- 8 Alejandro Ramos, “Innovación impulsada por datos para el mejor funcionamiento de los gobiernos...”, 33.
- 9 Jose van Dijck, “Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology...”, 197-208; Juliane Jarke y Andreas Breiter, “Editorial: the datafication of education”; Mark Lycett, “‘Datafication’: making sense of (big) data in a complex world”; Ulises A. Mejías y Nick Couldry, “Datafication”; Minna Ruckenstein y Natasha Dow Schüll, “The Datafication of Health”; Jathan Sadowski, “When Data Is Capital: Datafication, Accumulation, and Extraction”.
- 10 R.D. Galliers *et al.*, “Datification and its human, organizational and societal effects”; M. Lynne Markus, “Datification, Organizational Strategy, and IS Research: What’s the Score?”; Sue Newell y Marco Marabelli, “Strategic Opportunities (and Challenges)...”; Mari-Klara Stein *et al.*, “Datification and the Pursuit of Meaningfulness in Work”.

### ***Metadatos sociales: iniciativas...***

se considera pertinente utilizar el de *datafificación* (*datafication* en inglés), pues es el que tiene un uso más extendido entre los trabajos científicos realizados en estos temas.

Según los análisis de Sadowski (2019),

[...] los datos son en la actualidad elementos centrales y esenciales para cada vez más sectores del capitalismo contemporáneo. Capitalismo en el que, por cierto, se destaca la valoración económico-política de los datos y su interrelación con el propio sistema económico capitalista, bajo ideas como la del capitalismo de vigilancia, el capitalismo informativo, el capitalismo comunicativo o el capitalismo de plataforma. Diversas industrias enfocadas en temáticas de tecnología, infraestructura, finanzas, manufactura, seguros y energía están tratando, en la actualidad, a los datos como una forma de capital o de recursos.<sup>11</sup>

Lo anterior contrasta con lo que acontecía hace sólo algunos años,

[...] en donde los datos eran materia de preocupación o análisis casi con exclusividad de los científicos, además de que las empresas tanto públicas como privadas optaban por no recopilarlos o directamente eliminarlos debido a que se consideraba una mala inversión pagar para su almacenamiento. Ahora, debido a tendencias como la economía de datos o la economía digital, las empresas están más interesadas en recopilar tantos datos como puedan donde sea que puedan.<sup>12</sup>

Siguiendo con las ideas de Sadowski, en la actualidad eliminar datos debido a los costos de almacenamiento sería equivalente a quemar montañas de dinero o tirar barriles de petróleo por el desagüe.<sup>13</sup>

Algunos ejemplos de áreas en donde se ha utilizado la *datafificación* para la generación de servicios públicos han sido en temas

---

11 Jathan Sadowski, "When Data Is Capital", 1.

12 *op. cit.*, 1.

13 *Ibid.*

como el de la salud, bajo el entendido de *datafication of health* o *dataficación* de la salud, en donde autores como Ruckenstein o Dow (2017) han argumentado que se desarrolla en varias escalas de registros diferentes, incluida la investigación médica basada en datos y las infraestructuras de salud pública, la atención clínica de la salud y las prácticas de autocuidado. En este mismo sentido, estos autores mencionan que “Durante la última década, la capacidad de recopilar, almacenar y analizar datos fisiológicos, conductuales y de geolocalización de las personas ha llegado a afectar una amplia gama de dominios de la vida cotidiana, desde la formulación de políticas hasta la vigilancia de temas relacionados con la salud, y que pueden llegar a impactar, por ejemplo, la planificación urbana de prestación de servicios de salud”.<sup>14</sup>

Otro ejemplo más es el de la *dataficación* de la educación, en donde autores como Jarke y Breiter (2019) mencionan que

El sector educativo es uno de los dominios más afectados por la *dataficación*, porque transforma no sólo las formas en que se organizan la enseñanza y el aprendizaje, sino también las formas en que las generaciones futuras [construirán] la realidad con y a través de los datos. La *dataficación* de la educación comprende la recopilación de datos en todos los niveles de los sistemas educativos (individual, aula, escuela, región, estado, internacional), potencialmente sobre todos los procesos de enseñanza, aprendizaje y gestión escolar.<sup>15</sup>

Estos son sólo algunos ejemplos en los que la *dataficación* ya ha impactado en algunos países y, dependiendo del uso y aprovechamiento que se les dé a los datos por parte de agencias públicas y privadas, podría influir en algunos más a corto plazo. De igual forma, un ejemplo que ilustra de buena representación el potencial en el uso de la *dataficación* sería el relacionado con la instrumentalización de políticas y programas públicos relativos al fomento y desarrollo de las capacidades cívicas y ciudadanas, sobre todo

---

14 Ruckenstein y Dow Schüll, “The Datafication of Health”, 261.

15 Jarke y Breiter, “Editorial: the datafication of education”, 1.



las relacionadas en sus planos digitales, elementos tomados en consideración en el siguiente apartado del trabajo.

## *DATAFICACIÓN Y SU VÍNCULO CON LA CIUDADANÍA DIGITAL*

Para entrar al último apartado de este trabajo, los datos pueden ser entendidos como el recurso subyacente para la inteligencia en la toma de decisiones públicas o de negocios que impacten de diversas formas a los ciudadanos, lo anterior principalmente por tres elementos, mejor conocidos como las tres “V” propuestas por Lycett (volumen, velocidad y variedad):<sup>16</sup>

En cuanto al volumen se refiere, se propone que hay un beneficio clave en la capacidad de procesamiento de grandes cantidades de datos: la tesis analítica paralela a este primer planteamiento es la de que entre mayor cantidad de datos se analicen, se podrá llegar a mejores modelos de prestación de servicios o la generación de productos para la ciudadanía. Las consideraciones clave aquí se relacionan con la escalabilidad, la distribución y la capacidad de procesamiento, entre otros aspectos. Dentro de este volumen de datos no se debería dejar de lado a las minorías o ciudadanías diferenciadas,<sup>17</sup> pues muchas veces son los grupos más vulnerables los que tienen menos acceso a las tecnologías digitales, sin embargo, también son los que más dificultades y obstáculos afrontan para acercarse a la plenitud de la idea de ciudadanía. En otras palabras, no por prestar o desarrollar servicios o políticas basados en la *dataficación*, se deben olvidar la prestación de servicios o políticas focalizadas a los grupos que se encuentran en un nivel más vulnerable, pues muchas veces estos mismos grupos coinciden

---

16 Mark Lycett, “Datafication’: making sense of (big) data in a complex world”, 381.

17 Álvaro Bello, “Derechos indígenas y ciudadanías diferenciadas en América Latina y el Caribe. Implicancias para la educación”; Sonia Fleury, “Construcción de ciudadanía en entornos de desigualdad”; Guillermo O’Donnell, “Estado, Democratización y ciudadanía”.

con los que tienen un menor acceso a las tecnologías digitales, por lo que su voz en la *dataficación* pudiera ser escasa o nula.

En cuanto a la velocidad se refiere, se propone que la prontitud de la obtención del flujo de datos es importante, sobre todo en relación con el ciclo de retroalimentación para la acción.

Las consideraciones clave aquí incluyen la prontitud del análisis de los flujos de datos, la comprensión de lo que se puede descartar y lo que resulta importante de tomar en cuenta de esos mismos datos, para la mejor y más pronta toma de decisiones y adopción de medidas que impacten a la ciudadanía. Un ejemplo que podría ilustrar de buena forma este elemento sería en la capacidad de manejo y uso que se le pueda dar a la *dataficación* de la salud, con el tema específico de la pandemia, pues una más pronta recuperación y asimilación de datos relacionados con la pandemia podría reflejarse en un mejor manejo y en aminorar los efectos de ésta.

Finalmente, en cuanto a la variedad se refiere, se parte del supuesto de que en muchas ocasiones los datos resultan confusos para entender la realidad de una situación determinada, pues provienen de muchas fuentes y en muchos formatos diferentes, a menudo desestructurados, llenos de errores y de naturaleza inconsistente. Lo que resultaría clave aquí es tanto la capacidad de tomar en cuenta el grado de pérdida de información en el proceso de limpieza y depuración de los datos, así como las capacidades para su integración y el entendimiento de su versatilidad en la representación de problemáticas o necesidades ciudadanas, para que puedan ser atendidas o tomadas en cuenta. En este punto se torna fundamental el desarrollo de tecnologías adecuadas que puedan procesar grandes cantidades de datos que son dejados por los usuarios en su registro de actividad en línea, así como a la traducción de esos resultados arrojados por las tecnologías en las capacidades reales de los tomadores de decisiones públicas, con objeto de que sean traducidos en políticas, programas, proyectos o servicios que impacten de forma positiva la prestación de servicios públicos, como por ejemplo en los relacionados con la salud, la educación o el desarrollo ciudadano, planteados en este trabajo.

Estos elementos se vinculan, en definitiva, con el desarrollo del concepto y ejercicio de la ciudadanía digital relacionado con el uso de las tecnologías de la información y la comunicación tanto para hacer valer los derechos, así como para cumplir con las obligaciones enmarcadas en la idea tradicional de ciudadanía. Este uso tecnológico dejaría como resultado un número mayúsculo de datos que, mediante su adecuado procesamiento y uso, podría generar productos y servicios innovadores basados en la *datafificación*. Desde este punto de vista, sería fácilmente comprensible que, entre mayor capacidad de uso y asimilación de datos generados por la ciudadanía, se pudiera llegar a la mejor formulación de políticas y programas *dataficados*.

En este sentido, y siguiendo con los aportes de Lycett, “La *datafificación* se puede conceptualizar a través de tres nociones innovadoras que permiten repensar la lógica de la creación de valor: la desmaterialización, la condensación y la densidad”.<sup>18</sup>

La desmaterialización destaca la capacidad de separar el aspecto informativo de un dato (bajo su entendimiento como activo o recurso) y su uso en un contexto específico del mundo físico. La condensación destaca el hecho de que, una vez desmaterializada, la información puede ser manipulada y entendida fácilmente (siempre y cuando se cuente con una infraestructura adecuada), lo que permita que los recursos y los conjuntos de actividades que estaban estrechamente vinculados físicamente se desagreguen y se “reagrupen”, en formas que sean más manejables para una acción determinada, al eliminarse aspectos que anteriormente resultaban difíciles de entender, lentos y costosos. La densidad es la mejor (re)combinación de recursos movilizados para un contexto particular, en un momento y lugar determinados, pues es el resultado del proceso de creación de valor.

En términos generales, parafraseando a Hintz (2019), el mundo social *dataficado* se describe como el mundo en el que vivimos y se identifica como una abstracción de ciudadanos donde nuestras

---

18 Mark Lycett, *op. cit.*, 382.

experiencias vividas y recreadas a través de la interacción tecnológica y la generación de datos, crean nuevas bases para una mejor toma de decisiones públicas tanto gubernamentales como ciudadanas. Este escenario no está exento de problemáticas o retos a los cuales se les debe hacer frente. Entre estos estaría nuevamente la desigualdad o inequidad en el desarrollo tecnológico y de capacidades de las naciones y los gobiernos para la gestión de políticas, servicios y programas basados en la *dataficación*, pues nos veríamos relegados a una situación en la que las potencias económicas y sociales serían nuevamente las que generarían mejores herramientas para la gestión de proyectos *dataficados*, mientras que las economías y gobiernos que no estén al nivel de las potencias, nuevamente quedarían relegados de estas herramientas por falta de infraestructura y capacidades. De igual forma, la privacidad en el uso de las tecnologías es un tema interesante para tener en cuenta, pues el manejo de datos personales se debería de llevar de una forma segura que permita la información y las preferencias de la ciudadanía.

## REFLEXIONES FINALES

En el presente trabajo se puso a discusión teórica el concepto de sociología de la información. En este contexto, se partió del supuesto de que todo análisis sociológico se da en un contexto donde la información, socialmente construida, permite no sólo el desarrollo de una sociedad determinada, sino también su análisis. En este punto, se propuso centrar el análisis de la sociología de la información en un contexto marcado por la masificación en la utilización de las tecnologías de la información y comunicación, así como el aumento en el valor de los datos, porque permiten crear innovadores productos y servicios basados en información que los propios individuos van generando en su contante interacción tecnológica, ello conocido bajo el neologismo *dataficación*.

Esta *dataficación* ya ha generado ejemplos en algunas latitudes donde se ha utilizado, por parte de agencias públicas y privadas,

### *Metadatos sociales: iniciativas...*

para la generación de políticas y servicios relacionados con la salud, la educación y con el desarrollo social y ciudadano. En estos ejemplos se observan las potencialidades que tienen los registros y el cúmulo de datos que la ciudadanía va generando en su interacción constante con las tecnologías de la información y comunicación. Si embargo, a la par de todas estas ventajas, convendría preguntarse sobre las problemáticas y los retos que se tienen para un mejor uso de esos datos y la generación de políticas y servicios basados en la *dataficación*.

### REFERENCIAS BIBLIOGRÁFICAS

- Arcila, Carlos. "El ciudadano digital". *Revista Latinoamericana de Comunicación CHASQUI*, núm. 93 (2006): 18-21.
- Bello, Álvaro. "Derechos indígenas y ciudadanías diferenciadas en América Latina y el Caribe. Implicancias para la educación". En *Interculturalidad, educación y ciudadanía. Perspectivas latinoamericanas*, editado por Luis Enrique López. Bolivia: FUNPROEIB Andes. 2009.
- Bernal-Meza, Raúl y Gustavo Masera. "Sociedad de la información: etapa posterior de la globalización/mundalización. Desafíos y riesgos para América Latina". *Realidad Económica* 8, núm. 227 (2007): 90-116.
- Calhoun, Craig, Chris Rojek y Bryan Turner. *The SAGE Handbook of Sociology*. London: SAGE. 2005. <https://doi.org/10.4135/9781848608115>.
- Castells, Manuel. *La sociedad red: una visión global*. Madrid: Alianza Editorial. 2006.
- Dijk, Jose van. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology". *Surveillance & Society* 12, núm. 2 (9 de mayo de 2014): 197-208. <https://doi.org/10.24908/ss.v12i2.4776>.

- Fleury, Sonia. “Construcción de ciudadanía en entornos de desigualdad”. *Instituciones y Desarrollo* 16, núm. 1 (2005): 133-70.
- Ficury, Sonia *La expansión de la ciudadanía. VVAA: Inclusión social y nuevas ciudadanías*. Bogotá: Pontificia Universidad Javeriana. 2003.
- Galliers, Robert D., Sue Newell, Graeme Shanks, y Heikki Topi. “Datification and its human, organizational and societal effects”. *The Journal of Strategic Information Systems* 26, núm. 3 (1 de septiembre de 2017): 185-90. <https://doi.org/10.1016/j.jsis.2017.08.002>.
- Jarke, Juliane, y Andreas Breiter. “Editorial: the datafication of education”. *Learning, Media and Technology* 44, núm. 1 (2019): 1-6.
- Lycett, Mark. “‘Datafication’: making sense of (big) data in a complex world”. *European Journal of Information Systems* 22, núm. 4 (2013): 381-86.
- Markus, M. Lynne. “Datification, Organizational Strategy, and IS Research: What’s the Score?”. *The Journal of Strategic Information Systems* 26, núm. 3 (1 de septiembre de 2017): 233-41. <https://doi.org/10.1016/j.jsis.2017.08.003>.
- Mejias, Ulises A., y Nick Couldry. “Datafication”. *Internet Policy Review* 8, núm. 4 (29 de noviembre de 2019). <https://policyreview.info/concepts/datafication>.
- Newell, Sue, y Marco Marabelli. “Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of ‘Datification’”. *The Journal of Strategic Information Systems* 24, núm. 1 (1 de marzo de 2015): 3-14. <https://doi.org/10.1016/j.jsis.2015.02.001>.
- O’Donnell, Guillermo. “Estado, Democratización y ciudadanía”. *Nueva Sociedad*, núm. 128 (1993): 62-87.

**Metadatos sociales: iniciativas...**

- Ramos, Alejandro. "Información y ciudadanía, una propuesta desde la gobernanza". *Investigación bibliotecológica* 29, núm. 67 (2015): 113-40.
- Ramos, Alejandro. "Innovación impulsada por datos para el mejor funcionamiento de los gobiernos y la ciudadanía digital". En *Las prácticas sociales en la producción, distribución y acceso a la información mediadas por las tecnologías digitales*, editado por Araceli Torres, 30-37. Ciudad de México: IIBI-UNAM. 2022.
- Romero, Manuel Alejandro. "Sociología de la Información: ¿fenómeno emergente o realidad existente? Su comportamiento en Infotecarios". IFT, 20 de noviembre de 2015. <https://www.infotecarios.com/sociologia-de-la-informacion-fenomeno-emergente-o-realidad-existente-su-comportamiento-en-infotecarios/>.
- Ruckenstein, Minna y Natasha Dow Schüll. "The Datafication of Health". *Annual Review of Anthropology* 46 (2017): 261-78. <https://www.annualreviews.org/doi/10.1146/annurev-anthro-102116-041244>.
- Sadowski, Jathan. "When Data Is Capital: Datafication, Accumulation, and Extraction". *Big Data & Society* 6, núm. 1 (1 de enero de 2019): 2053951718820549. <https://doi.org/10.1177/2053951718820549>.
- Stein, Mari-Klara, Erica L. Wagner, Pamela Tierney, Sue Newell, y Robert D. Galliers. "Datification and the Pursuit of Meaningfulness in Work". *Journal of Management Studies* 56, núm. 3 (2019): 685-717. <https://doi.org/10.1111/joms.12409>.
- Vieira, Liszt. "Ciudadanía y control social". En *Lo público no estatal en la reforma del Estado*, editado por Nuria Cunill y Luiz Carlos Bresser, 215-56. Argentina: Paidós Ibérica, 1998. <http://dialnet.unirioja.es/servlet/articulo?codigo=610311>.

# Los metadatos en los planes de estudio de los grados en Información y Documentación: un enfoque comparativo entre Portugal y España

ANA LÚCIA TERRA  
*Universidade de Coimbra, Portugal*

## INTRODUCCIÓN

La palabra metadatos (*metadata*) entró en la lengua inglesa en 1968, pero la idea que la sustenta se remonta a la primera biblioteca. Así que, aunque la palabra “metadatos” sólo tiene unas pocas décadas de antigüedad, los bibliotecarios llevan miles de años trabajando con metadatos al crear catálogos con elementos que permiten a los usuarios encontrar los recursos de las colecciones de las bibliotecas.<sup>1</sup> Desde esta perspectiva, algunos autores utilizan el término metadatos para referirse a la descripción de los recursos de información.<sup>2</sup> Desde este punto de vista, los bibliotecarios trabajan en lo que tradicionalmente se denomina organización y representación de la información, una de las tres áreas centrales de la Ciencia de la Información, junto con la Gestión de la Información y el Comportamiento de la Información.<sup>3</sup>

---

1 Jeffrey Pomerantz, *Metadata*, 5-6.

2 Philip Hider, *Information Resource Description: Creating and Managing Metadata*, 4.

3 Armando Malheiro da Silva y Fernanda Ribeiro, *Das “Ciências” Documentais à Ciência Da Informação: Ensaio Epistemológico Para Um Novo Modelo Curricular*.



La organización de la información forma parte del plan de estudios de las titulaciones en Información y Documentación o en Ciencia de la Información<sup>4</sup> desde la creación de la formación académica especializada en esta materia, a finales del siglo XIX.<sup>5</sup> En un estudio longitudinal (2000-2013), publicado en tres partes, Daniel Joudrey traza un panorama de la evolución más reciente de los estudios relacionados con la organización de la información, la catalogación y los metadatos en las facultades de Biblioteconomía y Documentación de Estados Unidos.<sup>6</sup> En el primer estudio, publicado en 2002, el autor constató que las disciplinas relacionadas con los metadatos y la catalogación de materiales no librarios estaban aumentando y subrayó que deberían seguir creciendo a medida que aumenten también los recursos en formato electrónico. Al mismo tiempo, señaló que las disciplinas de catalogación descriptiva, tecnologías de catalogación y temas de catalogación avanzada estaban disminuyendo.<sup>7</sup> En 2008, el mismo autor destacó que los estudiantes de Library and Information Science (LIS), independientemente de su área de especialización, necesitan los fundamentos de la catalogación y de los enfoques orientados a los metadatos. Sólo así los estudiantes podrán comprender el carácter central de la organización de la información en todos los aspectos de las profesiones de la información.<sup>8</sup> En 2014, Joudrey

---

4 Dado que en España y Portugal se utilizan ambas denominaciones para los cursos en estudio, se utilizarán las dos expresiones como equivalentes.

5 Armando M. da Silva y Ribeiro, *op. cit.*, 141-149; Karen Snow, "Shifting Sands and the Prophet's Dream: Exploring the Future of Information Organization Education".

6 Daniel N. Joudrey, "A New Look at US Graduate Courses in Bibliographic Control"; Daniel N. Joudrey, "Another Look at Graduate Education for Cataloging and the Organization of Information"; Daniel N. Joudrey y Ryan McGinnis, "Graduate Education for Information Organization, Cataloging, and Metadata".

7 Daniel N. Joudrey, "A New Look at US Graduate Courses in Bibliographic Control".

8 Daniel N. Joudrey, "Another Look at Graduate Education for Cataloging and the Organization of Information".

consideraba que la disminución de cursos sobre organización de la información en los 58 cursos acreditados por la ALA, en 2012, suponía el riesgo de privar a los futuros profesionales de la información de un área de conocimiento fundamental. Pero en este contexto, también subrayó el crecimiento de las asignaturas relacionadas con los metadatos, señalando que en el año 2000 había seis escuelas con asignaturas en esta área y que en 2013 había 36 escuelas. Desde su punto de vista, los planes de estudio de Ciencia de la Información deben garantizar que los estudiantes, futuros profesionales de la información, sean capaces de trabajar en amplios contextos informativos; dominar los conceptos básicos de organización de la información, las normas aplicables y las herramientas utilizadas en diversos entornos informativos (de los que las bibliotecas y los servicios de información tradicionales son sólo una parte); ser capaces de interpretar y trabajar con una amplia variedad de normas de metadatos; y que conozcan y sean capaces de mantenerse al día con los cambios en la organización de la información.<sup>9</sup> La transversalidad y utilidad de la organización de la información, basada naturalmente en los metadatos, es destacada por Snow, que la presenta como una actividad crucial no sólo para el buen funcionamiento de las bibliotecas, sino de cualquier organización que crea, mantenga y difunda información.<sup>10</sup>

Desde este marco, el presente estudio se ha centrado en la inclusión de la enseñanza de temas relacionados con los metadatos, y la organización del conocimiento/información, en los cursos de grado en Ciencia de la Información, en España y en Portugal, en el año 2022.

---

9 Daniel N. Joudrey y McGinnis, "Graduate Education for Information Organization, Cataloging, and Metadata", 546.

10 Karen Snow, "Shifting Sands and the Prophet's Dream: Exploring the Future of Information Organization Education".

## METODOLOGÍA Y OBJETIVOS

Desde el punto de vista metodológico, se realizó un estudio exploratorio que siguió un enfoque cualitativo a partir de una recopilación de datos, disponibles en línea, sobre los planes de estudio de las titulaciones de grado en Ciencia de la Información en España y Portugal. En abril de 2022, se analizaron las páginas web de las titulaciones y se identificaron 14 carreras (11 en España y 3 en Portugal). Este estudio inicial de las titulaciones de grado/licenciatura en Ciencia de la Información, en España y Portugal, se ha complementado con el análisis de la información genérica disponible en las páginas web de las titulaciones (descripción general, plan de estudios y programas de las asignaturas). En la descripción general de los cursos se buscaron referencias a metadatos y a la organización del conocimiento en sentido amplio. A continuación, se identificaron las disciplinas relacionadas con los metadatos y la catalogación, y se consideró también una perspectiva más amplia de la organización del conocimiento, a partir de una revisión genérica del temario de las asignaturas de los planes de estudio. Estos aspectos fueron analizados usando una hoja de recogida de datos que contenía los siguientes criterios: Fundamentos de la normalización descriptiva, Normas de contenido (AACR, RDA, normas nacionales de catalogación, etcétera), Normas de estructura (Dublin Core, ISBD, MARC, etcétera), Normas de codificación, formato e interoperabilidad (MARC, XML, etcétera), Modelos conceptuales para datos bibliográficos (FRBR, FRISAD, etcétera), Tendencias en catalogación (datos enlazados, VIAF, interoperabilidad, etcétera). Estos criterios se definieron según la categorización establecida por el referencial “Core competencies for Cataloging and Metadata Professional Librarians”.<sup>11</sup> En la última fase, se codificaron y analizaron los datos.

Con esta opción metodológica se trazó un marco de referencias a la organización del conocimiento/información y a los metadatos

---

11 ALA, “Core Competencies for Cataloging and Metadata Professional Librarians”.

en la presentación y en los planes de estudio de las titulaciones en Información y Documentación. Para cumplir con este objetivo general, se realizó una encuesta sobre la inclusión de los temas de organización del conocimiento/información como uno de los elementos centrales de la presentación y posicionamiento de las carreras de Ciencia de la Información existentes en España y Portugal. Además, se ha intentado rastrear la presencia de los temas de metadatos en los planes de estudio de los cursos de licenciatura en Ciencia de la Información. En estos objetivos específicos también se incluyó un análisis más fino para verificar qué tipos de contenidos relacionados con los metadatos se impartían, según la categorización de la ALA de 2017.

## PRESENTACIÓN Y DISCUSIÓN DE LOS RESULTADOS

En lo que respecta al cumplimiento del primer objetivo específico, que pretendía identificar las referencias a los términos “organización” y “organización del conocimiento” en las descripciones generales/presentación inicial de los cursos, se encontró que es una referencia existente en todos los cursos, con mayor o menor énfasis.

Un análisis más fino muestra que en la descripción/presentación general del curso siempre se hace referencia al término “organización”, en el sentido de organización del conocimiento/información, normalmente como componente de un proceso global de gestión/procesamiento de la información. Así, por ejemplo, en la descripción del Grado de Gestión de Información y Documentación Digital, de la Universidad de Barcelona, se habla de “Objetivos: Desarrollar competencias específicas de captura, selección, **organización**, preservación y difusión de información y documentación mediante la aplicación de las tecnologías”.<sup>12</sup> Y entre las competencias específicas figuran “capacidad para **definir criterios y**

---

12 <https://www.ub.edu/portal/web/informacion-medios-audiovisuales/grados/-/ensenyament/detallEnsenyament/6063329/0>.

**aplicar las normativas** y otros instrumentos de **estructuración, descripción y representación de datos e información** en diferentes entornos y contextos”.<sup>13</sup> En la Universidad de A Coruña, se destaca el apartado de la organización, en un sentido amplio y luego en un enfoque más dirigido a la creación de metadatos descriptivos, y se afirma que el eje central del Grado en Gestión Digital de Información y Documentación “lo forman materias de Información y Documentación que proporcionan formación en todo lo relativo a los procesos de producción, **organización**, recuperación, evaluación y preservación de la información y la Documentación digital y digitalizada”, y se hace hincapié en “la publicación de información de forma interoperable mediante **lenguajes de descripción de información digital**”.<sup>14</sup> Por su parte, el Grado en Gestión de Información y Contenidos Digitales, de la Universidad de Murcia, afirma en su justificación que “incide en la **dimensión tecnológica** de la **producción y publicación de información, su organización mediante esquemas de metadatos multidominio** para posibilitar su descubrimiento, acceso y consumo desde plataformas digitales, así como su gestión en marcos institucionalizados para su preservación y servicio”.<sup>15</sup> Este énfasis en los metadatos se refuerza en las palabras clave que se asocian a la descripción del curso, en el que se incluyen:

Diseño y programación web | Innovación digital con datos y contenidos | Publicación interactiva | Visualización de información | **Taxonomías** | **Metadatos** | Gestión de contenidos | **Organización del conocimiento** | Diseño de interacción | Administración digital | Privacidad | Comunidades Virtuales | Digitalización de la difusión de la cultura | Normalización | Posicionamiento web | Búsqueda y descubrimiento de información | Nuevas profesiones

---

13 <https://www.ub.edu/portal/web/informacion-medios-audiovisuales/grados/-/ensenyament/detallEnsenyament/6063329/1>.

14 <https://www.udc.es/es/gradoinfodigital/>.

15 <https://www.um.es/web/estudios/grados/contenidos-digitales/descripcion>.

digitales | Emprendimiento | Transformación digital en comunicación social | Economía de los datos | Patrimonio cultural digital | Derechos digitales.<sup>16</sup>

En la Universidad de Salamanca, el Grado en Información y Documentación no incluye ninguna referencia a la organización del conocimiento o a los metadatos en su página de inicio, pero en las competencias específicas menciona “**organizar y estructurar los datos relativos a la descripción de documentos y colecciones** de documentos en cualquier soporte; crear y explotar las herramientas de acceso a los datos, documentos o referencias”.<sup>17</sup> La Universidad de León, por su parte, incluye en la presentación inicial de su Grado en Información y Documentación esta mención: “se demandan cada vez más profesionales en Información y Documentación formados y con experiencia en la **organización del conocimiento**, el tratamiento y la recuperación eficaz de información”.<sup>18</sup> Mientras, en las habilidades, se centra en los aspectos más tradicionales de la Biblioteconomía y de la Documentación, incluyendo

**definir y estructurar los datos relativos a la descripción de documentos y colecciones** de documentos en cualquier soporte para hacerlos accesibles al usuario y facilitar su control por el profesional. **Elaborar registros bibliográficos por medio de la catalogación, indización y clasificación.** **Evaluar los productos documentales derivados de la catalogación, indización, clasificación y descripción** sustancial de la información.<sup>19</sup>

---

16 <https://www.um.es/web/comunicacion/grado-en-gestion-de-informacion-y-contenidos-digitales>.

17 <https://www.usal.es/grado-en-informacion-y-documentacion/objetivos>.

18 <https://online.unileon.es/estudios/grado-informacion-documentacion/>.

19 <https://online.unileon.es/estudios/grado-informacion-documentacion/>.

Tanto en la Universidad de Granada,<sup>20</sup> como en la Universidad Zaragoza,<sup>21</sup> en la Universidad de Valencia<sup>22</sup> o en la Universidad Complutense,<sup>23</sup> el término organizar, aplicado a la información o a los documentos, se utiliza también en el texto de presentación de sus respectivos grados de Información y Documentación, lo que lo enmarca más ampliamente en las competencias y funciones que desempeñarán estos titulados. Por último, en lo que respecta a las universidades españolas, observaremos que en la Universidad de Extremadura se opta por referirse a asignaturas concretas en las que la organización y los metadatos están implícitos pero sin utilizar estos términos en concreto: “cuenta con materias tales como diseño y dirección de sitios web, creación y gestión de bases de datos, localización de recursos informativos, edición digital o análisis de redes, junto a otras de carácter tradicional, como **catalogación y clasificación bibliográfica, descripción de documentos** de archivo, paleografía o diplomática”,<sup>24</sup> lo que busca asociar este último con temas de carácter más tecnológico.

En el contexto portugués, la Universidad de Coimbra, en la página de presentación del grado en Ciencia de la Información, define como uno de los objetivos del curso que el alumno “adquiera conocimientos sobre los principios teóricos y metodológicos de los procesos de selección, **organización, representación**, recuperación, acceso y conservación de la información”.<sup>25</sup> La Universidad de Oporto también hace referencia a que el curso está dirigido a estudiantes “interesados en el estudio de la información, es decir, en su producción, **organización**, recuperación, evaluación

---

20 <https://fcd.ugr.es/docencia/grados/graduadoa-informacion-y-documentacion/motivacion>.

21 <https://fyl.unizar.es/grado/grado-en-informacion-y-documentacion>.

22 <https://www.uv.es/uvweb/grado-informacion-documentacion/es/se-estudia/estudiar-este-grado/interes-academico-cientifico-profesional-1285956919260.html>.

23 <https://www.ucm.es/informacionydocumentacion/competencias-y-objetivos>.

24 <https://www.unex.es/conoce-la-uex/centros/alcazaba/titulaciones/info/presentacion?id=1706>.

25 <https://apps.uc.pt/courses/PT/course/5543>.

y conservación, así como en el análisis, desarrollo y gestión de sistemas y servicios de información”.<sup>26</sup> Por último, el Grado en Ciencias y Tecnologías de la Documentación y la Información, del Instituto Politécnico de Oporto, también menciona la organización de la información en la página de presentación del curso, orientándola hacia la gestión de la información.<sup>27</sup>

En cuanto al segundo objetivo específico, que era sondear la presencia del tema de los metadatos en los planes de estudio de las 14 titulaciones de Grado en Documentación e Información / Ciencia de la Información existentes en España y Portugal, se identificaron 66 asignaturas en las que se tratan temas relacionados con los metadatos; existen cursos con un máximo de siete asignaturas en estas áreas y cursos con sólo dos asignaturas (tabla 1).

**Tabla 1:** Asignaturas con temas relacionados con los metadatos en los grados de Información y Documentación / Ciencia de la Información en España y Portugal

Universidad	Curso	Disciplinas
Universidad de Barcelona	Grado de Gestión de Información y Documentación Digital	Representación de la Información: Metadatos   2º año, Sem. 1
		Representación de la Información: Catalogación   2º año, Sem. 2
		Web Semántica   4º año, Sem. 2
		Tratamiento del Material Audiovisual   3º año, Sem. 1
		Información y Formatos Digitales   1º año, Sem. 1
Universidad Carlos III	Grado en Gestión de la Información y Contenidos Digitales	Lenguajes de Marcado   1º año, Sem. 1
		Organización y representación del conocimiento   1º año, Sem. 2
		Accesibilidad de contenidos digitales   3º año, Sem. 2
		Metadatos Avanzados. Optativa   4º año
		Web Semántica. Optativa   4º año
		Ciencia de Datos. Optativa   4º año
Datos Abiertos. Optativa   4º año		

26 [https://sigarra.up.pt/flup/pt/cur\\_geral.cur\\_view?pv\\_ano\\_lectivo=2021&pv\\_origem=CUR&pv\\_tipo\\_cur\\_sigla=L&pv\\_curso\\_id=454](https://sigarra.up.pt/flup/pt/cur_geral.cur_view?pv_ano_lectivo=2021&pv_origem=CUR&pv_tipo_cur_sigla=L&pv_curso_id=454).

27 <https://www.iscap.ipp.pt/cursos/licenciatura/808>.



**Metadatos sociales: iniciativas...**

Universidad da Coruña	Grado en Gestión Digital de Información y Documentación	Representación y Recuperación Digital de la Información I   2º año, Sem. 2
		Representación y Recuperación Digital de la Información II   3º año, Sem. 1
Universidad de Murcia	Grado en Gestión de Información y Contenidos Digitales	Metadatos y Descripción de Recursos   1º año, Sem. 1
		Vocabularios y Taxonomías   2º año, Sem. 1
		Edición Digital II: Lenguajes de Marcado   1º año, Sem. 2
Universidad de Salamanca	Grado en Información y Documentación	Metadatos y descripción de archivos   3º año, Sem. 1
		Descripción y acceso de recursos bibliográficos   2º año, Sem. 1
		Representación y organización de contenidos documentales   2º año, Sem. 2
		Introducción al análisis documental   1º año, Sem. 2
Universidad de León	Grado en Información y Documentación	Catalogación en bibliotecas y centros de documentación   2º año, Sem. 1
		Control de autoridades   2º año, Sem. 2
		Catalogación en archivos   2º año, Sem. 2
		Sistemas de representación normalizada del contenido   3º año, Sem. 1
		Colecciones digitales   4º año, Sem. 1
Universidad de Granada	Grado en Información y Documentación	Catalogación Descriptiva   2º año, Sem. 1
		Sistemas para la Organización del Conocimiento I   2º año, Sem. 1
		Normalización para la Organización de la Información   2º año, Sem. 2
		Sistemas para la Organización del Conocimiento II   2º año, Sem. 2
		Sistemas de Metadatos. Optativa   3º año, Sem. 2
		Tecnologías Web para Servicios de Información   4º año, Sem. 1
Universidad Complutense	Grado en Información y Documentación	Catalogación descriptiva   2º año, Sem. 2
		Catalogación automatizada II   3º año, Sem. 1
		Edición digital en el entorno web   3º año, Sem. 1
		Lenguajes de marcado, semántica y metadatos   3º año, Sem. 2

**Los metadatos en los planes...**

Universidad de Zaragoza	Grado en Información y Documentación	Edición digital   1º año, Sem. 1
		Fundamentos de indización   1º año, Sem. 2
		Catalogación   2º año, Sem. 1
		Indización precoordinada y clasificación bibliográfica   2º año, Sem. 2
		Aplicaciones para recursos de información digital   3º año, Sem. 1
		Construcción y evaluación de tesauros. Optativa   3º o 4º año, Sem. 1
Universidad de Valencia	Grado en Información y Documentación	Informática II   1º año, Sem. 2
		Arquitectura de la Información en la Web   2º año, Sem. 2
		Catalogación Automatizada   2º año, Sem. 2
		Catalogación General   2º año, Sem. 2
		Sistemas de Representación de la Información y Conocimiento   2º año, Sem. 2
Universidad de Extremadura	Grado en Información y Documentación	Catalogación descriptiva   2º año, Sem. 2
		Organización y descripción de archivos   2º año, Sem. 2
		Catalogación automatizada   3º año, Sem. 1
		Diseño y dirección de websites   3º año, Sem. 2
Universidade de Coimbra	Licenciatura em Ciência da Informação	Fundamentos de Controlo da Informação. Optativa   Sem. 2
		Normas e Aplicações de Metadados   Sem. 2
		Organização da informação em bibliotecas   Sem. 2
		Organização da Informação: Controlo de Autoridades   Sem. 2
		Organização da Informação: Descrição de Recursos Bibliográficos   Sem. 2
		Representação da Informação: Linguagens Vocabulares   Sem. 2
		Representação da Informação: Resumos e Linguagens Codificadas   Sem. 2
Universidade do Porto	Licenciatura em Ciência da Informação	Metainformação Descritiva   2º año, Sem. 2
		Controlo de Autoridade e Indexação   2º año, Sem. 1

### **Metadatos sociales: iniciativas...**

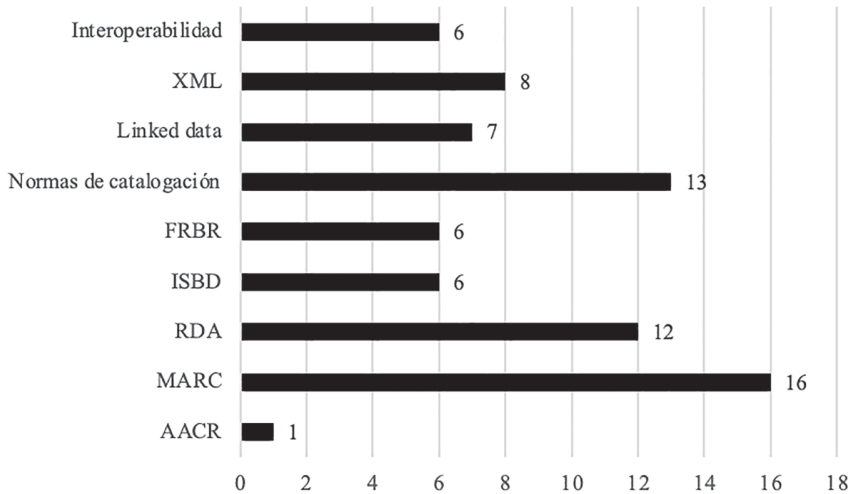
Instituto Politécnico do Porto	Licenciatura em Ciências e Tecnologias da Documentação e Informação	Descrição da Informação I   1º año, Sem. 2
		Análise e Representação da Informação I   2º año, Sem. 1
		Descrição da Informação II   2º año, Sem. 1
		Análise e Representação da Informação II   2º año, Sem. 2
		Bibliotecas e Arquivos Digitais   3º año, Sem. 1
		Interoperabilidade de Sistemas   3º año, Sem. 2

Fuente: Elaboración propia, 2022.

El análisis de las denominaciones de las asignaturas identificadas en los planes de estudio muestra que sólo seis de ellas incluyen directamente el término metadato o metainformación; optan en su lugar por denominaciones más tradicionales como catalogación, descripción o representación, semánticamente especificada, con alusiones al contexto de los archivos o la información, entre otras. Además, se adoptan nuevas denominaciones más orientadas al aspecto tecnológico como interoperabilidad de sistemas, arquitectura de información web, accesibilidad a los contenidos digitales o web semántica.

Un primer análisis de los contenidos impartidos en las unidades del curso muestra que las normas nacionales de catalogación (españolas y portuguesas, además de las AACR) son un tema muy frecuente en el conjunto de disciplinas analizadas, lo que apunta a una visión más tradicional. Sin embargo, es interesante que la RDA sea el siguiente contenido más identificado. El formato MARC aparece como el estándar de estructura más presente en los planes de estudio. También hay que señalar que esta preocupación por presentar a los estudiantes perspectivas más tradicionales y visiones más actuales es claramente visible en el caso de las referencias a la ISBD, que es equivalente a las referencias a los FRBR. En este análisis preliminar de los contenidos de las asignaturas relacionadas con los metadatos en los cursos de licenciatura en Ciencia de la Información, en España y Portugal, también se identificaron nuevas tendencias como el XML o los temas de interoperabilidad.

**Figura 1:** Estudio preliminar de los contenidos impartidos en las asignaturas



Fuente: elaboración propia, 2022.

Así, en la actualidad, los cursos de Ciencia de la Información están integrando en sus planes de estudio, en las disciplinas destinadas a la inclusión de metadatos, estándares que no provienen del contexto bibliotecario en el temario, y que son utilizados y entendidos por los no bibliotecarios, como el Extensible Markup Language (XML) o el Dublin Core. Además, están integrando la *Resource Description & Access* (RDA), publicada en 2010, y basada en los *Functional Requirements for Bibliographic Records* (FRBR), los *Functional Requirements for Authority Data* (FRAD) y los *Functional Requirements for Subject Authority Data* (FRSAD), mantenidos por la IFLA, y que se ajusta al *Library Reference Model* (LRM), la norma de la IFLA que los consolida. La introducción a RDA y/o a estos diferentes modelos de la IFLA inicia a los estudiantes en la creación de metadatos que pueden ser utilizados en la World Wide Web, preferentemente como datos enlazados. Así, los planes de estudio de Ciencia de la Información se encaminan a promover

las competencias en la creación de datos estructurados, que facilitan la puesta en común de la información y su localización en la Web, sin limitar su enfoque a los servicios de información tradicionales, como las bibliotecas, los archivos o los centros de documentación. Pero incluso para estos contextos están presentando nuevas tendencias y propuestas más tecnológicas y orientadas a contextos web.

## CONCLUSIÓN

La inclusión de disciplinas o temas relacionados con los metadatos en los planes de estudio de los cursos de Grado en Ciencia de la Información ilustra el paso de un ámbito profesional y de estudios muy centrado en el material bibliográfico a un escenario que abarca una gran diversidad de recursos y normativas. Estas disciplinas, centradas en la organización del conocimiento y los metadatos, introducen nuevos enfoques en la estructura y la sintaxis de la descripción de los recursos de información. Esta ampliación permite ir más allá de los campos de la catalogación o de las normas de descripción para dominios específicos, preparando a los estudiantes para nuevos contextos de organización y representación del conocimiento, en un entorno sobrecargado de información, donde los diferentes tipos de indización, clasificaciones, modelos de datos, sistemas de información y tecnologías son componentes fundamentales para dotar a los profesionales de la información de nuevas visiones y de la capacidad de intervenir en diferentes escenarios.

Una vez esbozado este marco general, se hace necesario, para futuros trabajos, un análisis más detallado y fino de los temas de metadatos que se imparten en las carreras de Ciencia de la Información en España y Portugal, con el fin de identificar tendencias y aspectos mejorables, y centrarse también en un análisis comparativo basado en estudios relativos a otros países.

REFERENCIAS BIBLIOGRÁFICAS

- ALA. “Core Competencies for Cataloging and Metadata Professional Librarians”. 2017.
- Hider, Philip. *Information Resource Description: Creating and Managing Metadata*. 2nd ed. London: Facet Publishing. 2018.
- Joudrey, Daniel N. “A New Look at US Graduate Courses in Bibliographic Control”. *Cataloging and Classification Quarterly* 34, núm. 1–2 (2002): 57–99. [https://doi.org/10.1300/J104v34n01\\_06](https://doi.org/10.1300/J104v34n01_06).
- . “Another Look at Graduate Education for Cataloging and the Organization of Information”. *Cataloging and Classification Quarterly* 46, núm. 2 (2008): 137–81. <https://doi.org/10.1080/01639370802177588>.
- , y Ryan McGinnis. “Graduate Education for Information Organization, Cataloging, and Metadata”. *Cataloging and Classification Quarterly* 52, núm. 5 (2014): 506–50. <https://doi.org/10.1080/01639374.2014.911236>.
- Pomerantz, Jeffrey. *Metadata*. Cambridge: MIT Press, 2015.
- Silva, Armando Malheiro da, y Fernanda Ribeiro. *Das “Ciências” Documentais à Ciência Da Informação: Ensaio Epistemológico Para Um Novo Modelo Curricular*. Porto: Afrontamento, 2002.
- Snow, Karen. “Shifting Sands and the Prophet’s Dream: Exploring the Future of Information Organization Education”. *Journal of Education for Library and Information Science* 60, núm. 2 (2019): 139–51. <https://doi.org/10.3138/jelis.2018-0043>.

## Metadatos en la formación profesional en CC II

JULIO CÉSAR RIVERA AGUILERA  
LUIS ROBERTO RIVERA AGUILERA  
BRENDA LUCERO CAMPOS MONREAL  
*Universidad Autónoma de San Luis Potosí, México*

### INTRODUCCIÓN

**E**s importante señalar que la sociedad de la información y del conocimiento<sup>1</sup> en la que nos encontramos, se caracteriza, entre otras cosas, por la excesiva producción de información, resultado de los avances en la ciencia y la tecnología, lo que ha generado una gran explosión documental, que causa en los individuos el fenómeno conocido como infoxicación,<sup>2</sup> al tener acceso a grandes cantidades de datos e información. Frente a este problema, los metadatos juegan un papel trascendental, ya que proveen estándares y herramientas que coadyuvan y facilitan la organización, uso y preservación de información digital disponible en la actualidad.

El término metadato tiene un sentido bidimensional, por un lado “meta”, que significa más allá de, y “datum”, que en latín significa dato, entendiéndolo entonces como “más allá del dato”, lo que hace referencia a un proceso de representación, aunque su uso no

---

1 Tópico estudiado por autores como Daniel Bell, Manuel Castells y Armand Mattelart.

2 Término que hace referencia a una gran cantidad de información, difícil de procesar por su volumen.

es exclusivo del ámbito de la informática, y es en las Ciencias de la Información donde ha encontrado una mayor utilización, sobre todo en procesos que tienen que ver con organización y recuperación de información.

Los metadatos son utilizados para describir el contenido de un documento, y para lograrlo, existen dos posibilidades en su utilización, puede ser de manera intrínseca o extrínseca, es decir, como parte del documento o como complemento al mismo, pero siempre buscando representar su contenido, contexto y estructura con miras a mantenerlo estable y accesible en el tiempo.

En este sentido, es preciso señalar algunas de las ventajas que se pueden lograr al hacer uso de los metadatos en la gestión de datos e información en las organizaciones; algunas de ellas reflejan procesos como: describir y codificar recursos de información, incrementar el acceso a los documentos, mantener el control de versiones en archivos digitales, gestionar el aspecto legal de los objetos digitales, facilitar la recuperación de información, preservar información digital, entre otros.

Es importante conocer las tendencias en el campo de los metadatos, una publicación reciente hace mención del perfil del profesional de la información y de las posibilidades de desempeño con relación a los metadatos, y hace alusión a las actividades que puede desarrollar (Bookriot):<sup>3</sup> analista de datos, bibliotecario de datos, analista de inteligencia de negocios, especialista en minería de datos, técnico del sistema de información global (SIG), entre otros.

## FORMACIÓN PROFESIONAL Y METADATOS

### Competencias profesionales

La investigación parte de la propuesta de competencias profesionales que se mencionan en el libro blanco: título de grado en

---

3 Bookriot, "50+ Alternative Careers for Librarians".



Información y Documentación, creado por ANECA (2005)<sup>4</sup> en el Espacio Europeo de Educación Superior.

El documento referido es resultado de dos informes básicos del área de Biblioteconomía y Documentación, el primero de ellos fue a raíz de la cumbre de Berlín de septiembre de 2003 y los directorios UNESCO Libraries Portal y World list of Departaments and Schools of Information Studies, Information Management, e Information Services, los cuales dan fundamento a lo propuesto en el libro blanco. Algunos de los países que participaron en su elaboración fueron Alemania, Austria, Dinamarca, Francia, Italia, Noruega, Países Bajos, Polonia, Reino Unido, Suecia, Suiza. Los participantes compartieron las entidades formadoras en el área, los títulos de grado que ofertan, duración y equivalencias.

El perfil profesional en Información y Documentación que propone el libro blanco de Documentación, en lo referente a aspectos sobre Información y Documentación, se muestra en la tabla 1.

En lo que respecta a las competencias profesionales específicas de formación disciplinar y profesional, el libro blanco propone las siguientes (ANECA 2005, 61):<sup>5</sup>

- E01 Interacción con los productores, los usuarios y los clientes de la información
- E02 Conocimiento del entorno profesional de la Información y la Documentación
- E03 Conocimiento del marco jurídico y administrativo nacional e internacional de la gestión de la información
- E04 Identificación, autenticación y evaluación de fuentes y recursos de información
- E05 Gestión de colecciones y fondos

---

4 Agencia Nacional de Evaluación de la Calidad y Acreditación, *Título de grado en Información y Documentación*.

5 *ibid.*

**Tabla 1. Competencias específicas de Información y Documentación**

<b>Competencias específicas</b>	<b>Conocimientos</b>	<ul style="list-style-type: none"> <li>• Conocimiento de la naturaleza de la información y de los documentos, de sus diversos modos de producción y de su ciclo de gestión, de los aspectos legales y éticos de su uso y transferencia, y de las fuentes principales de información en cualquier soporte.</li> <li>• Conocimiento de los principios teóricos y metodológicos para la planificación, organización y evaluación de sistemas, unidades y servicios de información.</li> <li>• Conocimiento de los principios teóricos y metodológicos para la reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio de la información.</li> <li>• Conocimiento de los principios teóricos y metodológicos para el estudio, el análisis, la evaluación y la mejora de los procesos de producción, transferencia y uso de la información y de la actividad científica.</li> <li>• Conocimiento de las tecnologías de la información que se emplean en las unidades y servicios de información.</li> <li>• Conocimiento de la realidad nacional e internacional en materia de políticas y servicios de información y de las industrias de la cultura.</li> </ul>
	<b>Competencias profesionales</b>	<ul style="list-style-type: none"> <li>• Capacidad de aplicar y valorar las técnicas de planificación, organización y evaluación de sistemas, unidades y servicios de información.</li> <li>• Capacidad de usar y aplicar las técnicas, las normativas y otros instrumentos utilizados en la reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio de la información.</li> <li>• Habilidades en el manejo de las tecnologías como medio indispensable en los procesos de tratamiento y transferencia de la información.</li> <li>• Habilidades en la autenticación, el uso, el diseño y la evaluación de fuentes y recursos de información.</li> <li>• Habilidades para analizar, asesorar y formar a productores, usuarios y clientes de servicios de información, así como habilidades en los procesos de negociación y comunicación.</li> <li>• Habilidades en la obtención, tratamiento e interpretación de datos sobre el entorno de las unidades y servicios de información, y el estudio, la gestión y la evaluación de los procesos de producción, transferencia y uso de la información y de la actividad científica.</li> </ul>
	<b>Competencias académicas</b>	<ul style="list-style-type: none"> <li>• Comprender y aplicar los principios y las técnicas para la planificación, organización y evaluación de sistemas, unidades y de servicios de información.</li> <li>• Comprender y aplicar los principios y las técnicas para la reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio de la información.</li> <li>• Utilizar y aplicar herramientas informáticas para la implantación, desarrollo y explotación de sistemas de información.</li> <li>• Comprender y aplicar las técnicas de evaluación de las fuentes y recursos de información.</li> <li>• Comprender y aplicar las técnicas de gestión y marketing de sistemas, unidades y servicios de información.</li> </ul>

Fuente: Elaboración propia. Adaptado de libro blanco de Documentación.

- E06 Preservación, conservación, y tratamiento físico de documentos
- E07 Análisis y representación de la información
- E08 Organización y almacenamiento de la información
- E09 Búsqueda y recuperación de la información
- E10 Elaboración y difusión de la información
- E11 Tecnologías de la información: informática
- E12 Tecnologías de la información: telecomunicaciones
- E13 Técnicas de producción y edición
- E14 Técnicas de gestión administrativa
- E15 Técnicas de marketing
- E16 Técnicas comerciales
- E17 Técnicas de adquisición
- E18 Técnicas de gestión microeconómica
- E19 Técnicas de instalación, acondicionamiento y equipamiento
- E20 Técnicas de planificación y gestión de proyectos
- E21 Técnicas de diagnóstico y evaluación
- E22 Técnicas de gestión de recursos humanos
- E23 Técnicas de formación

Las competencias que se tomaron en cuenta para el presente estudio, por ser las que se relacionan con la temática de la investigación, son las siguientes:

**Análisis y representación de información (E07).** Identificar y representar, en el lenguaje documental adoptado o en otro sistema simbólico, el contenido semántico de un documento o de una colección de documentos o de un fondo de archivo.

**Organización y almacenamiento de información (E08).** Organizar y estructurar los datos relativos a la descripción de documentos y colecciones de documentos en cualquier soporte; crear y explotar las herramientas de acceso a los datos, documentos o referencias.

**Búsqueda y recuperación de información (E09).** Buscar y recuperar la información por métodos que permitan dar respuesta

a las expectativas de los demandantes en condiciones óptimas de coste y tiempo.

**Elaboración y difusión de información (E10).** Hacer disponibles y explotables las informaciones tratadas y facilitar su uso mediante el suministro de productos y servicios documentales.

**TIC: Informática (E11).** Utilizar y poner en práctica métodos, técnicas y herramientas informáticas (hardware o software) para la implantación, desarrollo y explotación de sistemas de información.

**TIC: Telecomunicaciones (E12).** Utilizar y poner en práctica los métodos, las técnicas y las herramientas (hardware o software) para la implantación, desarrollo y explotación de sistemas de telecomunicación.

## METADATOS

### Importancia

Para González,<sup>6</sup> la importancia de los metadatos radica en que permiten describir y clasificar el contenido de un documento digital; es posible afirmar, además, que sin metadatos sería casi imposible encontrar un documento sin abrir y leer toda una carpeta o varios directorios; por otro lado, conviene tener presente que los metadatos posibilitan la descripción de las propiedades o atributos de los materiales digitales.<sup>7</sup>

En ese mismo sentido y de acuerdo con Viltres, Rodríguez y Pedro (2018),<sup>8</sup> los metadatos favorecen las tareas de recuperar y almacenar información de recursos digitales para su procesamiento y representación a los usuarios cuando realizan una consulta;

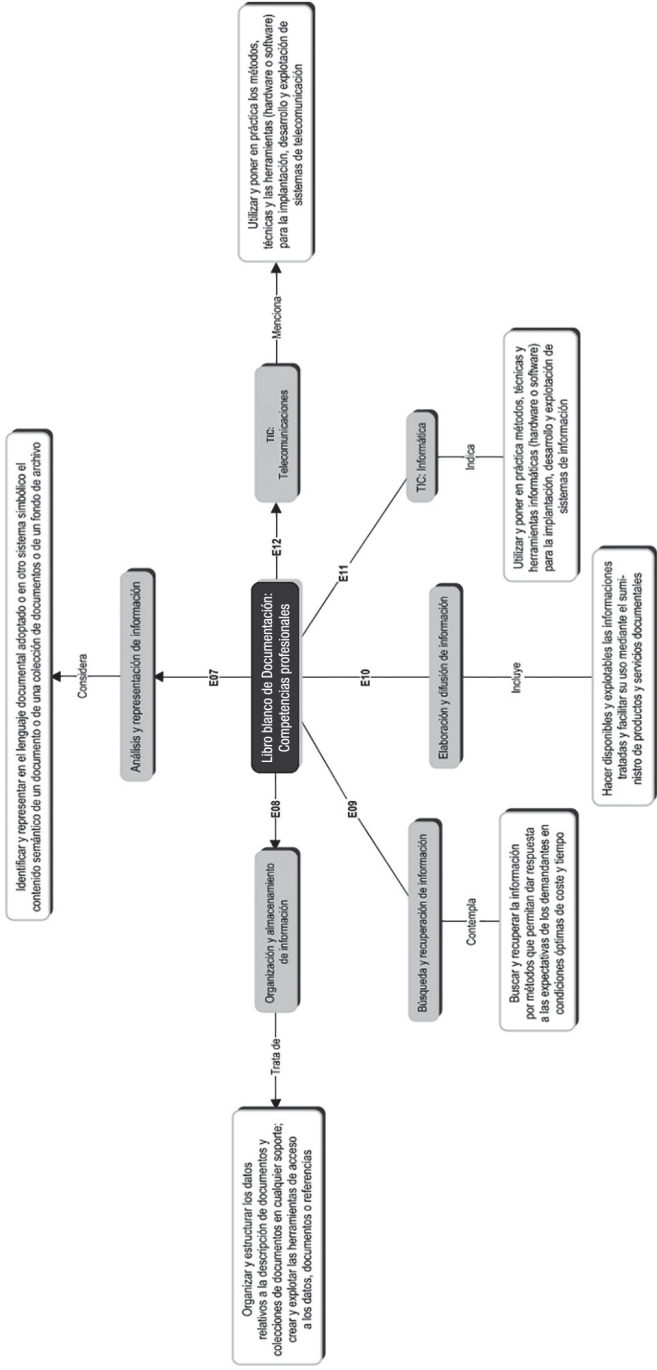
---

6 Renato González, “Información, datos y metadatos”, 17.

7 Archivo General de la Nación, *Desarrollo de políticas y procedimientos para la preservación digital*.

8 Hubert Viltres *et al.*, “Procesamiento semántico de información en sistemas de recuperación de información”.

Figura 1. Competencias profesionales específicas



Fuente: Elaboración propia. Adaptado del libro blanco de Documentación.

### *Metadatos sociales: iniciativas...*

además, proporcionan modos comunes e interoperables de acceso, uso y reutilización de recursos digitales (Ninin y Simionato 2018).<sup>9</sup>

### Objetivos

Destaca Voutssás (2020)<sup>10</sup> los siguientes:

- Definir estrategias de conservación (también llamadas estrategias de mantenimiento)
- Establecer estrategias de preservación
- Hacer permanente el acceso a documentos / objetos digitales en el largo plazo
- Preservar objetos lógicos (contenidos)

Por su parte, Ferreira (2020)<sup>11</sup> complementa señalando los siguientes:

- Representar y proporcionar una base de codificación a través de estructuras flexibles
- Procesar y gestionar recursos informativos
- Describir recursos informativos y sus relaciones con otros recursos en la red
- Intercambiar información digital

### Clasificación

Los metadatos se pueden clasificar en tres niveles [Taylor y Joudrey (2009)<sup>12</sup> y Rodríguez (2020)<sup>13</sup>]:

---

9 Débora Ninin y Ana Simionato, "Publicación de datos abiertos".

10 Juan Voutssás, "Metadatos para preservación digital".

11 Fabiano Ferreira, "Metadatos, datos enlazados e interoperabilidad: reflexiones en el dominio bibliográfico Universidade Federal de São Carlos, Brasil".

12 Arlene Taylor y Daniel N. Joudrey, *The organization of information*.

13 Ariel Rodríguez, "Retro y prospectiva de la investigación bibliotecológica sobre metadatos".

*Primero.* Con un formato simple donde el metadato se encuentra desestructurado.

*Segundo.* El metadato está estructurado e incluye un conjunto de elementos formales para otorgarle representatividad y descripción a los recursos.

*Tercero.* Un metadato enriquecido que puede ser empleado por cualquier sistema de información.

Adicionalmente, como parte de la clasificación de los metadatos, resulta de suma importancia considerar distintos estándares que contribuyen para la interoperabilidad, por ejemplo, véanse los *Criterios básicos para valorar sistemas*, publicado por la UNAM:<sup>14</sup>

Tabla 2. Estándares para interoperabilidad con metadatos

Formatos:	Normas:
<ul style="list-style-type: none"><li>• Marc21</li><li>• Dublin Core</li><li>• Eads</li><li>• Mods</li><li>• Mets</li><li>• Json</li><li>• Xml</li><li>• Cvs</li><li>• Mix</li></ul>	<ul style="list-style-type: none"><li>• ISO2709</li><li>• ISAD-G</li><li>• RCAA2</li><li>• RDA</li></ul>

Fuente: UNAM 2020, 38.

## Tipología

En la tabla 3, se muestran de manera sintética los tipos de metadatos, algunos ejemplos de sus propiedades y el uso primario que cada uno representa:

---

14 Instituto de Investigaciones Bibliograficas-unam. *Criterios básicos para valorar sistemas de preservación digital*, 38.

**Tabla 3. Tipos de metadatos**

Tipo de metadatos		Ejemplo de propiedades	Uso primario
Descriptivos		Título Autor Materia (tema) Género Fecha de publicación	Descubrimiento Presentación Interoperabilidad
Administrativos	Técnicos	Tipo de archivo Tamaño de archivo Fecha/hora de creación Esquema de compresión	Interoperabilidad Gestión de objetos digitales Preservación
	Preservación	Cotejo de datos Acción de preservación	
	Derechos de autor	Estatus Términos de licencia Titular de los derechos	
Estructurales		Secuencia Posición en la jerarquía	Navegación
Lenguajes de marcado		Párrafo Cabecera Lista Nombre Fecha	Navegación Interoperabilidad

Fuente: Elaboración propia. Adaptado de Riley (2017).<sup>15</sup>

## Contexto del estudio

El estudio se sitúa en la Facultad de Ciencias de la Información de la Universidad Autónoma de San Luis Potosí, México.

## Programas académicos

A continuación, se señalan las distintas versiones y denominaciones, desde su creación, del programa académico ofertado por la universidad autónoma potosina para el área de Ciencias de la Información.

---

15 Jenn Riley, *Understanding metadata what is metadata, and what is it for?*, 10-11.



- 1980. Lic. en Biblioteconomía
- 2000. Lic. en Bibliotecología e Información
- 2006. Lic. en Bibliotecología y Lic. en Archivología

A partir de 2014 y a la fecha, ofrece las siguientes licenciaturas:

- Gestión de la Información
- Gestión Documental y Archivística

### Mapa curricular

En las tablas 4 y 5 se muestra el mapa curricular para cada licenciatura.

**Tabla 4. Mapa curricular de la Licenciatura en Gestión de la Información**

SEMESTRE	Licenciado en Gestión de la Información										
<b>Primero</b>	Ciencias de la información documental (194)	Lógica (195)	Teoría administrativa (196)	Introducción a las TIC (197)	Cultura escrita (198)	Metodología de la Investigación (199)	Inglés Básico I				
<b>Segundo</b>	Tipología documental (200)	Comunicación (201)	Estadística (202)	Técnicas de registro documental (203)	Patrimonio documental (204)	Técnicas de investigación (205)	Inglés Básico II				
<b>Tercero</b>	Usuarios y clientes de la información (210)	Administración de recursos humanos (211)	Diseño bases de datos (212)	Descripción documental (214)	Lenguajes documentales e Indexación (213)	Contabilidad <b>Optativa 1</b> (215)	Inglés Intermedio I	Cultura bibliográfica en México <b>Optativa 2</b> (216)			
<b>Cuarto</b>	Servicios de información documental (223)	Teoría de sistemas (224)	Clasificación decimal Dewey (225)	Organización de los recursos de información (226)	Normalización de la información (227)	Redes y sistemas de información <b>Optativa 3</b> (233)	Inglés Intermedio II	Diseño de sitios web <b>Optativa 4</b> (234)			
<b>Quinto</b>	Servicios de consulta (244)	Unidades y sistemas de información (245)	Sistemas de información (246)	Clasificación LCC (247)	Recursos continuos (248)	Estrategias didácticas <b>Optativa 5</b> (251)	Inglés Avanzado	Taller de conservación y restauración <b>Optativa 6</b> (252)			
<b>Sexto</b>	Servicios especializados de información (301)	Desarrollo de colecciones (302)	Gestión de la información y del conocimiento (303)	Taller procesamiento de la información (305)	Bibliografía (306)	Digitalización y documentos electrónicos <b>Optativa 7</b> (292)	Servicio social (309)	Organización de documentos especiales I <b>Optativa 8</b> (308)			
<b>Séptimo</b>	Planación de unidades y sistemas de información (315)	Sistemas de gestión de calidad (316)	Políticas de información (319)	Seminarios de investigación (320)	Taller de proyectos tecnológicos (317)	Organización de Documentos Especiales II <b>Optativa 9</b> (323)	Prácticas profesionales (322)	Redacción de documentos <b>Optativa 10</b> (324)			
<b>Octavo</b>	Alfabetización informativa (333)	Taller de documentación (334)	Economía de la información (335)	Seminario de titulación (339)	Diseño de unidades y sistemas de información (336)	Desarrollo de habilidades directivas <b>Optativa 11</b> (340)		Preservación digital <b>Optativa 12</b> (341)			

Fuente: <http://www.fci.uaslp.mx/OE/LGIDEF/OELGIMC>.

**Tabla 5. Mapa curricular de la Licenciatura en Gestión Documental y Archivística**

SEMESTRE	Licenciado en Gestión Documental y Archivística						
<b>Primero</b>	Ciencias de la Información Documental (194)	Lógica (195)	Teoría administrativa (196)	Introducción a las TIC (197)	Cultura escrita (198)	Metodología de la investigación (199)	Inglés Básico I
<b>Segundo</b>	Teoría del conocimiento (206)	Historia de la archivística y de los archivos (207)	Estadística (202)	Eriquo de sistemas (208)	Patrimonio documental (204)	Teoría de la gestión documental (209)	Inglés Básico II
<b>Tercero</b>	Legislación archivística y normativa documental (217)	Diseño bases de datos (taller) (212)	Procesos archivísticos (218)	Archivos administrativos y archivos históricos (219)	Gestión de archivos (220)	Derecho Constitucional <b>Optativa 1</b> (221)	Inglés Intermedio I
<b>Cuarto</b>	Metodología de investigación cuantitativa (228)	Acceso a la documentación y protección de datos (229)	Archivos y administración pública (230)	Lenguajes documentales (231)	Sistemas de gestión documental (232)	Diseño de sitios web <b>Optativa 3</b> (234)	Inglés Intermedio II
<b>Quinto</b>	Sistemas de información (taller) (253)	Preservación y conservación del patrimonio documental (254)	Usuario de la Información (255)	Diagnóstico archivístico (256)	Taller de instrumentos de control y descripción archivística I (257)	Equidad y género <b>Optativa 5</b> (260)	Inglés Avanzado
<b>Sexto</b>	Deontología de la profesión (291)	Digitalización y documentos electrónicos (292)	Taller de instrumentos de control y descripción archivística II (296)	Organización de acervos especiales (295)	Paleografía (297)	Museografía <b>Optativa 7</b> (298)	Servicio social (608)
<b>Séptimo</b>	Seminario de titulación I (300)	Políticas de información (332)	Servicios de información (327)	Diplomática (328)	Sistemas institucionales de archivos (329)	Prospectivas de la archivística <b>Optativa 9</b> (330)	Práctica profesional (708)
<b>Octavo</b>	Seminario de titulación II (342)	Taller de diseño de proyectos tecnológicos (317)	Mercadeo y difusión de servicios de Información (345)	Planeación de edificios y acondicionamiento de espacios (346)	Administración de proyectos (347)	Diseño de cursos de capacitación <b>Optativa 11</b> (348)	Estudio de fondos históricos <b>Optativa 12</b> (349)

Fuente: <http://www.fci.uaslp.mx/OE/OELDGA/OELGDAMP>

## Ejes de formación

**Tabla 6. Ejes de formación**

Gestión de la Información	Gestión Documental y Archivística
<ul style="list-style-type: none"><li>• Servicios</li><li>• Administración</li><li>• Tecnologías de Información y Comunicación</li><li>• Patrimonio Documental</li><li>• Normalización</li></ul>	<ul style="list-style-type: none"><li>• Fundamentos</li><li>• Tecnologías de Información y Comunicación</li><li>• Patrimonio documental</li><li>• Servicios</li><li>• Organización</li></ul>

Fuente: Elaboración propia.

## Asignaturas

A continuación, se muestra la relación de asignaturas de ambos programas educativos en las que se aborda el tema de metadatos y tópicos relacionados.

**Tabla 7. Asignaturas que abordan los metadatos y tópicos relacionados**

Gestión de la Información	Gestión Documental y Archivística
<ul style="list-style-type: none"><li>• Digitalización y documentos electrónicos</li><li>• Gestión de la información</li><li>• Introducción a las TIC</li><li>• Normalización de la información</li><li>• Organización de acervos especiales</li><li>• Preservación digital</li><li>• Recursos continuos</li><li>• Redes y sistemas de información</li><li>• Taller de proyectos tecnológicos</li></ul>	<ul style="list-style-type: none"><li>• Digitalización y documentos electrónicos</li><li>• Diseño de sitios web</li><li>• Introducción a las TIC</li><li>• Organización de acervos especiales</li><li>• Taller de proyectos tecnológicos</li></ul>

Fuente: Elaboración propia.

## Temas

En seguida, se comparte el listado sobre temas vinculados con los metadatos, los cuales se abordan en las asignaturas referidas en la tabla 7.

**Tabla 8.** Temas de asignaturas donde se abordan los metadatos

Gestión de la Información	Gestión Documental y Archivística
<ul style="list-style-type: none"><li>• Catalogación</li><li>• Dublin Core</li><li>• Marc 21</li><li>• Metadatos descriptivos</li><li>• Metadatos técnicos</li><li>• Normas para el registro y organización</li><li>• Normas para transferencia y acceso</li><li>• Normas y estándares</li><li>• Normativa</li><li>• Organización de recursos</li><li>• PREMIS</li><li>• Recursos electrónicos</li><li>• Registro de fondos</li><li>• Repositorios digitales</li><li>• XML</li></ul>	<ul style="list-style-type: none"><li>• Descripción documental</li><li>• Dublin Core</li><li>• Estándares</li><li>• Marc 21</li><li>• Metadatos</li><li>• Normas y estándares</li><li>• Repositorios digitales</li><li>• XML</li></ul>

Fuente: Elaboración propia.

## Insumos para el estudio

Para llevar a cabo el estudio fue necesario el empleo de los siguientes materiales de trabajo:

- Libro blanco de Documentación e Información
- Modelo Universitario de Formación Integral (UASLP)
- Programas de asignatura (LGI / LGDyA)
- Mapas curriculares (versión sintética de los planes de estudio)
- Mallas curriculares (versión analítica de cada programa educativo)

### *Metadatos sociales: iniciativas...*

Como parte del análisis de dichos documentos, fue posible identificar distintos elementos indispensables, para observar la presencia de los metadatos entre la temática que se aborda en los programas académicos que oferta la UASLP, en el área de Ciencias de la Información. Al examinar los materiales referidos, se detectaron las competencias profesionales específicas que se fomentan entre los estudiantes, así como los ejes curriculares, asignaturas, temas y subtemas que guardan relación estrecha con los metadatos, de forma tal que se logró también determinar el nivel de inclusión de este asunto en el currículo.

## RESULTADOS

Entre los hallazgos de este estudio, fue posible identificar una serie de elementos que permitieron conocer el nivel de atención y abordaje sobre el tema de metadatos, en los planes y programas de estudio en Ciencias de la Información que se ofertan en la UASLP. Derivado de este análisis, se presenta a continuación una síntesis de los resultados encontrados.

Tabla 9. Competencias profesionales y síntesis sobre análisis del programa académico en Gestión de la Información y su relación con el estudio de metadatos

Competencia Libro Blanco	Competencia Disciplinar	Línea curricular	Asignaturas	Temas	Semestre
E-08. Organización y almacenamiento de información	Implementar procesos y técnicas para la organización, control y recuperación de información	Normalización	Normalización de la información	<ul style="list-style-type: none"> <li>• Normas para el registro y organización Marc 21</li> </ul>	4
E-14. Técnicas de gestión administrativa	Aplicar el proceso administrativo en las organizaciones	Administración	Redes y sistemas de información	<ul style="list-style-type: none"> <li>• Marc 21</li> <li>• Catalogación</li> <li>• Organización de recursos</li> <li>• Registro de fondos</li> </ul>	5
E-08. Organización y almacenamiento de información	Implementar procesos y técnicas para la organización, control y recuperación de información	Organización	Organización de acervos especiales	<ul style="list-style-type: none"> <li>• Marc 21</li> <li>• Dublin Core</li> <li>• Normativa</li> <li>• Recursos electrónicos</li> </ul>	7
E-11. TIC: Informática	Identificar, evaluar y emplear herramientas tecnológicas para la gestión de información	TIC	Digitalización y documentos electrónicos	<ul style="list-style-type: none"> <li>• Normas y estándares</li> <li>• Metadatos descriptivos</li> <li>• Metadatos técnicos</li> <li>• Formato Marc 21</li> <li>• Dublin Core</li> <li>• XML</li> </ul>	6
			Taller de proyectos tecnológicos	<ul style="list-style-type: none"> <li>• Repositorios digitales</li> <li>• Dublin Core</li> </ul>	7
				<ul style="list-style-type: none"> <li>• XML</li> <li>• PREMIS</li> </ul>	8

Fuente: Elaboración propia.

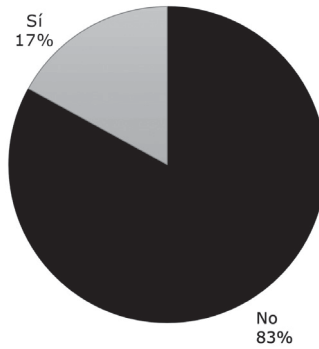
Tabla 10. Competencias profesionales y síntesis sobre análisis del programa académico en Gestión Documental y Archivística y su relación con el estudio de metadatos

Competencia Libro Blanco	Competencia Disciplinar	Línea curricular	Asignaturas	Temas	Semestre
E-08. Organización y almacenamiento de información	Planear e implementar la gestión documental para la operación del sistema institucional de archivos	Organización	Organización de acervos especiales	<ul style="list-style-type: none"> <li>• Dublin Core</li> <li>• Descripción documental</li> </ul>	6
E-11. TIC: Informática	Aplicar las tecnologías de la Información para la sistematización de la gestión documental	TIC	Introducción a las TIC	<ul style="list-style-type: none"> <li>• XML</li> <li>• Normas y estándares</li> </ul>	1
			Diseño de sitios web	<ul style="list-style-type: none"> <li>• Estándares</li> <li>• XML</li> </ul>	4
			Digitalización y documentos electrónicos	<ul style="list-style-type: none"> <li>• Metadatos</li> <li>• Formato Marc 21</li> <li>• Dublin Core</li> <li>• XML</li> </ul>	6
			Taller de proyectos tecnológicos	<ul style="list-style-type: none"> <li>• Repositorios digitales</li> <li>• Dublin Core</li> <li>• Formato Marc 21</li> </ul>	8

Fuente: Elaboración propia.

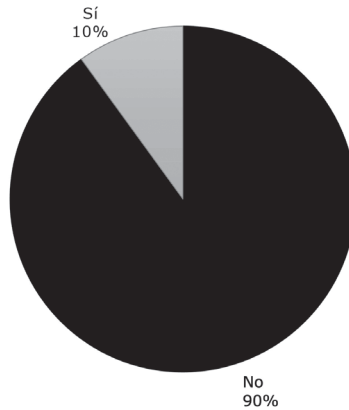


**Figura 2.** Porcentaje de asignaturas en LGI enfocadas al estudio de metadatos



Fuente: Elaboración propia.

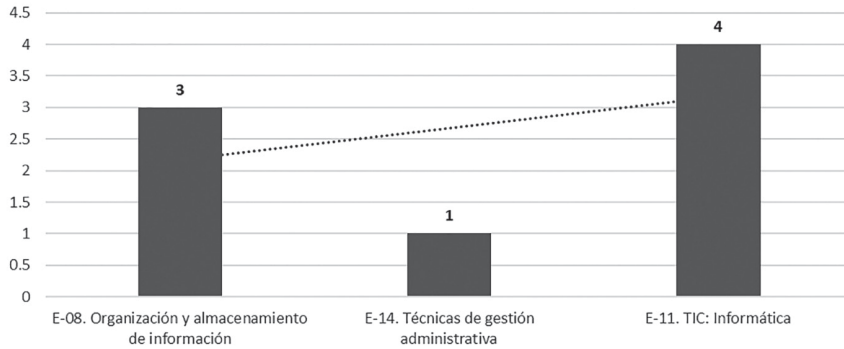
**Figura 3.** Porcentaje de asignaturas en LGDyA enfocadas al estudio de metadatos



Fuente: Elaboración propia.

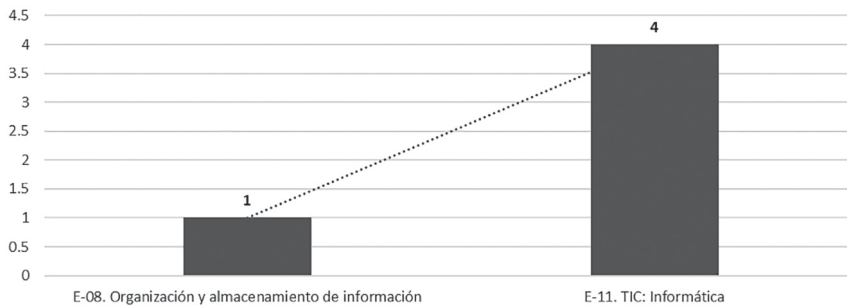
### **Metadatos sociales: iniciativas...**

**Figura 4. Cantidad de asignaturas en LGI que abonan al desarrollo de competencias profesionales declaradas en libro blanco**



Fuente: Elaboración propia.

**Figura 5. Cantidad de asignaturas en LGDyA que abonan al desarrollo de competencias profesionales declaradas en el libro blanco**



Fuente: Elaboración propia.

### **CONCLUSIONES**

- Los metadatos son un tema de alta relevancia en la actualidad, ya que hacen referencia a estándares y herramientas

que permiten la representación de información en ambientes digitales, al permitir la descripción, almacenamiento, recuperación y preservación de información a lo largo del tiempo.

- Es importante considerar las propuestas hechas por instituciones y organizaciones internacionales, en lo que a desarrollo de competencias profesionales se refiere, ya que marcan las tendencias emergentes que deben integrarse en los planes y programas de estudio de las instituciones de educación superior, con miras a satisfacer los requerimientos, por un lado, de la sociedad actual y por otro, del mercado laboral.
- Consideramos necesario fortalecer el tema de metadatos y tópicos relacionados en el perfil del profesional en Ciencias de la Información, a fin de desarrollar competencias que permitan su manejo y aplicación al momento del diseño, desarrollo e implementación de proyectos en materia de gestión y sistematización de recursos y contenidos digitales en las organizaciones.
- Después del estudio, reconocemos la importancia de actualizar periódicamente los programas de asignatura donde se abordan los metadatos y tópicos emergentes relacionados, así como ampliar el número de cursos con esta temática, desde un enfoque eminentemente teórico-práctico.
- Finalmente, debemos mencionar que, en los programas académicos ofertados en la UASLP para el área de Ciencias de la Información, en su versión 2014, se contribuye al desarrollo de competencias profesionales específicas, declaradas en el ámbito internacional a través del libro blanco de Documentación, particularmente, en lo que a metadatos se refiere, situación que permite a los egresados disponer de elementos que les facilite incursionar en proyectos vinculados a esta actividad.

## REFERENCIAS BIBLIOGRÁFICAS

- Agencia Nacional de Evaluación de la Calidad y Acreditación. Libro blanco: *Título de grado en Información y Documentación*. España: Agencia Nacional de Evaluación de la Calidad y Acreditación. 2005.
- AGN. *Desarrollo de políticas y procedimientos para la preservación digital*. México: AGN. 2017. [https://www.gob.mx/cms/uploads/attachment/file/228990/InterPARES\\_2\\_020617.pdf](https://www.gob.mx/cms/uploads/attachment/file/228990/InterPARES_2_020617.pdf)
- Bookriot. “50+ Alternative Careers for Librarians”. <https://bookriot.com/alternative-careers-for-librarians/> (consultada el 13 de julio de 2022).
- Ferreira de Castro, Fabiano. “Metadatos, datos enlazados e interoperabilidad: reflexiones en el dominio bibliográfico Universidade Federal de São Carlos, Brasil”. En *La revolución de los datos bibliográficos, científicos y culturales*, coordinado por Ariel Rodríguez, 33-64. México: UNAM-IIBI. [https://ru.iibi.unam.mx/jspui/handle/IIBI\\_UNAM/8](https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/8) (consultada el 13 de julio de 2022).
- González, Renato. “Información, datos y metadatos para la conservación del patrimonio cultural”. *Intervención* (enero-junio): 6-21. <http://www.scielo.org.mx/pdf/inter/v9n17/2007-249X-inter-9-17-6.pdf> (consultada el 13 de julio de 2022).
- Instituto de Investigaciones Bibliográficas-UNAM. *Criterios básicos para valorar sistemas de preservación digital*. México: Instituto de Investigaciones Bibliográficas-UNAM. 2020. <https://preservaciondigital.iib.unam.mx/docs/productos/Criterios-Basicos-Sistemas-Preservacion-Digital.pdf> (consultada el 13 de julio de 2022).

- Ninin, Débora y Carolina Simionato. “Publicación de datos abiertos en instituciones de patrimonios culturales”. *Palabra Clave (La Plata)* 8. <https://www.redalyc.org/articulo.oa?id=350556354007> (consultada el 13 de julio de 2022).
- Riley, Jenn. *Understanding metadata. What is metadata, and what is it for?* Baltimore: National Information Standards Organization. 2017. <http://groups.niso.org/higherlogic/ws/public/download/17446/Understanding%20Metadata.pdf> (consultada el 13 de julio de 2022).
- Rodríguez García, Ariel Alejandro (coord.). *La revolución de los datos bibliográficos, científicos y culturales*. México: UNAM-IIBI. 2020. [https://ru.iibi.unam.mx/jspui/handle/IIBI\\_UNAM/8](https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/8) (consultada el 13 de julio de 2022).
- Taylor, Arlene y Daniel N. Joudrey. *The organization of information*. Westport, Connecticut: Libraries Unlimited. 2009.
- Viltres, Hubert, Paúl Rodríguez, Juan Pedro Febles y Vivian Estrada. “Procesamiento semántico de información en sistemas de recuperación de información”. *Revista Cubana de Ciencias Informáticas* 12: 102-116. <http://scielo.sld.cu/pdf/rcci/v12n1/rcci08118.pdf> (consultada el 13 de julio de 2022).
- Voutsás Márquez, Juan. “Metadatos para preservación digital”. En: *La revolución de los datos bibliográficos, científicos y culturales*, coordinado por Ariel Rodríguez, 247-264, México: UNAM-IIBI. 2020. (consultada el 13 de julio de 2022).

# Flujo de información y usuarios de redes sociales universitarias: características, perfiles, necesidades e impactos en la organización

MARCO BRANDÃO

*Universidad Federal Fluminense, Brasil*

## INTRODUCCIÓN

La Tecnología de la Información ya no necesita probar los cambios que ha causado en todos los aspectos de la vida, especialmente con la información en la web. Los metadatos digitales han protagonizado gran parte de estos cambios, impactando directamente el mundo del trabajo al incorporar nuevos componentes para el tratamiento de las actividades.

Algunas áreas han sido muy afectadas por considerar significativo el comportamiento y los hábitos de los consumidores, especialmente a través de las Redes Sociales (RS): ellas han estado presentes de diferentes formas, congregando a las personas y las organizaciones alrededor del mundo para múltiples finalidades. Y debido a sus características, agregan un poder al usuario más allá de la simple capacidad de consumo: lo convierten en un *prosumidor* de información.

Las RS también han roto las barreras en cuanto a los procesos de comunicación e internacionalización, debido a que la información en ella está más allá de las fronteras territoriales, culturales

e ideológicas a través de los flujos de información internacional;<sup>1</sup> también es por eso que Contreras Orozco considera que las RS “se han convertido en un espacio con muchas oportunidades para cualquier ámbito de actividad productiva, y pueden ser un instrumento fundamental para construir nuevas relaciones de cercanía y participación social en las actividades gubernamentales”.<sup>2</sup>

Entre los objetivos que tienen las personas, las organizaciones y las propias RS, Celaya (2008)<sup>3</sup> propone una clasificación que considera a las RS como (1) Profesionales (e. g.: LinkedIn, Academia.edu), (2) Generalistas (e. g.: Facebook, X —antes Twitter—) y (3) Especializadas (E.G.: películas, música, deportes, grupos específicos de personas como padres, personas que están por casarse, abuelos, etcétera); así que sus seguidores eligen entre una u otra por sus necesidades de información.

Entre otros, el estudio *The Global State of Digital in 2019* de la plataforma de gestión de RS Hootsuite<sup>4</sup> ha considerado a México el tercer país del mundo en uso de RS: registraba en aquel año más de 83 millones de usuarios, lo que representaba más de la mitad de la población de ese país. Las RS más utilizadas por esa población son del tipo Generalista, y figuraba en primero YouTube, con un 95% de usuarios activos, y en segundo Facebook, con un 93% de usuarios activos (en el quinto lugar estaba X —antes Twitter—, con 57% de usuarios activos).<sup>5</sup>

Los datos de la realidad mexicana inducen a creer que las RS impactan en la vida de la población; sin embargo, el carácter Generalista de las principales RS en uso en el país convierte difícil

---

1 Egbert J. Sánchez Vanderkast, “Los flujos de información: algunas reflexiones”, 58.

2 Leticia Contreras Orozco, “Las redes sociales digitales como factor de proximidad entre la sociedad y los gobiernos subnacionales en México”.

3 Harold Hütt Herrera, “Las redes sociales: una nueva herramienta de difusión”, 8.

4 Simon Kemp, *Digital 2019: essential insights into how people around the world use the internet, mobile devices, social media, and e-commerce*.

5 AIMX, *15° Estudio sobre los Hábitos de los Usuarios de Internet en México 2019*.

y complejo su análisis. Desde la comunicación al entretenimiento, movilidad, actividades laborales, comerciales, gubernamentales, académicas etcétera, las RS Generalistas son un universo que tiene usuarios que pertenecen a varias de ellas, con distintas relaciones y objetivos.

En el presente trabajo se busca ofrecer una mirada a los usuarios de RS universitarias a partir de los resultados de una investigación que ha analizado las redes Facebook, X (antes Twitter) y YouTube de una organización universitaria, así como determinar el perfil y el impacto de ellas en los objetivos de la institución y en las necesidades de información de los usuarios con base en los flujos identificados.

#### CARACTERÍSTICAS DE LA INVESTIGACIÓN Y DEL FLUJO DE INFORMACIÓN DE LA RS UNIVERSITARIA

La investigación<sup>6</sup> fue realizada entre los años de 2019 y 2020 como un estudio exploratorio,<sup>7</sup> el cual hizo el establecimiento y la adaptación de indicadores para las RS de organizaciones universitarias y se propuso una metodología de evaluación para identificar el perfil de usuarios y determinar el impacto de la comunidad de las RS en la organización con sus flujos.

Las bases teórico-prácticas de ese estudio y los indicadores para las RS universitarias fueron diseñados en el artículo “Una revisión sociotécnica de las redes sociales de una organización universitaria”,<sup>8</sup> en el cual, se hizo (1) un análisis socio-técnico de las RS en conformidad con los abordajes de Bamforth (1951),<sup>9</sup> Fred

---

6 Marco Brandão, “El flujo de Información a través de las Redes Sociales”..

7 Augusto Nivaldo Silva Triviños, *Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação*.

8 Marco Brandão, 2021

9 Katy Manrique Valenzuela y Miguel Domingo González Álvarez, “Desarrollo de Sistemas Socio Técnicos en el área de Seguridad y Salud Ocupacional de una empresa de servicio”.



Emery y Eric Trist,<sup>10</sup> Manrique Valenzuela y Gonzalez Álvarez,<sup>11</sup> Caldevilla Dominguez,<sup>12</sup> Contreras Orozco,<sup>13</sup> Hütt Herrera;<sup>14</sup> (2) el estudio de los lineamientos de la organización universitaria para las RS;<sup>15</sup> (3) el análisis del flujo de información con base en los autores Keith Devlin<sup>16</sup> y Egbert Sánchez Vanderkast;<sup>17</sup> y, finalmente, (3) la proposición de los indicadores de eficiencia, los *Key Performance Indicator* (KPI),<sup>18</sup> adaptados para las RS de entidades y organizaciones universitarias públicas.

El objetivo de la organización con las RS es enriquecer la interacción y colaboración con la comunidad para que se logre una mayor visibilidad y apreciación del quehacer universitario.<sup>19</sup>

Con eso definido (el sistema gerencial) y considerados la comunidad universitaria (el sistema social) y las RS de la organización (el sistema técnico), se determinaron las siguientes métricas para la evaluación del impacto de las RS:

---

10 Fred E. Emery y Eric L. Trist, *Socio-technical Systems. Management Sciences Models and Techniques*.

11 Manrique Valenzuela y González Álvarez, *op. cit.*

12 David Caldevilla Dominguez, “Las Redes Sociales. Tipología, uso y consumo de las redes 2.0 en la sociedad digital actual”.

13 Contreras Orozco, *op. cit.*

14 Hütt Herrera, *op. cit.*

15 UNAM, *Instructivo de uso para Redes Sociales Institucionales de la UNAM*.

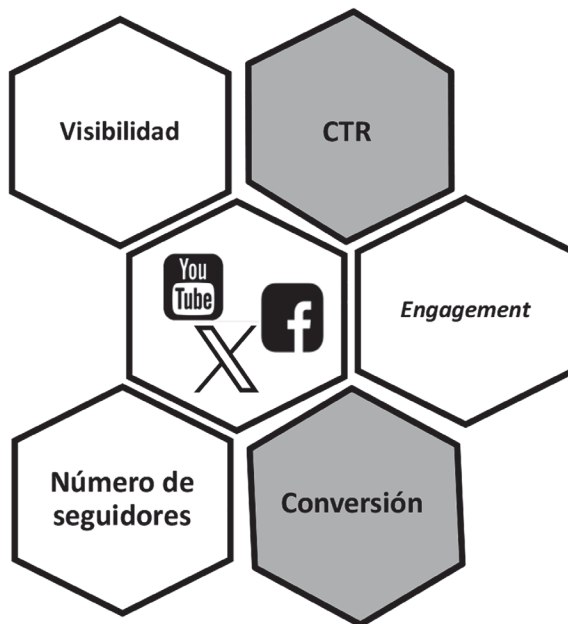
16 Keith Devlin, *Logic and Information*.

17 Sánchez Vanderkast, *op. cit.*

18 Los *Key Performance Indicator* (KPI) son métricas para evaluar procesos de gestión. En función de los distintos objetivos en las RS, las organizaciones tendrán distintos objetivos de acuerdo con su sistema gerencial y social, es decir, en función de aquello que se quiera con la información de las RS (Brandão, “Una revisión socio-tecnica...”, 26).

19 UNAM., *op. cit.*, 3.

Figura 1. Métricas de los KPIs para las RS de organizaciones universitarias



Fuente: Marco Brandão, “Una revisión sociotécnica...”, 28.

La visibilidad se refiere a la suma entre el número de personas que ha visto una publicación (alcance) y el número de veces que las publicaciones fueron vistas (impresiones); el número de seguidores se refiere al número de total de seguidores en la RS en la hora de su obtención; el *engagement* es la suma del número de *likes* con el número de *shares* y el número de comentarios; el *Click Through Rate* (CTR) indica el número de clics en cada publicación de la RS; y la conversión es el porcentaje de usuarios que han participado de los eventos de la entidad/organización universitaria.

Los flujos de información han sido definidos en conformidad con la tabla 1 que se presenta a continuación:

**Metadatos sociales: iniciativas...**

**Tabla 1.** Tipos de información de las RS del IIBI/UNAM

Institucionales	Emitidas por la alta gestión de la universidad, la rectoría, secretarías, direcciones.	Boletín, informes periódicos, campañas, noticias de carácter institucional.
Notas informativas	Emitidas por la alta gestión de la universidad y por la entidad.	Información de respuesta a situaciones internas y externas a la universidad/entidad, instrucciones a la comunidad.
Eventos	Emitidas por la entidad.	Difusión de talleres, cursos, congresos, libros, participación de la entidad en organismos nacionales e internacionales.

Brandão, *ibid.* 28.

Los KPIS consisten en la observación de los 28 días de transcurso de las publicaciones en las RS, periodo en el cual también se aplicó una encuesta que ha permitido identificar el perfil de usuarios y determinar el impacto del KPI conversión.

## LOS RESULTADOS DE LA INVESTIGACIÓN

### El perfil de usuarios de las RS

Las RS de la entidad/organización investigada tenían en 2020 más de 34 000 usuarios (figura 2). Tomadas aisladamente, era un total de 14 230 seguidores en Facebook,<sup>20</sup> 18 342 en X<sup>21</sup> y 2 070 (6%) seguidores en Youtube.<sup>22</sup> Estos números son los de base para la medición de los KPIS, y son ellos el primer indicador (número

---

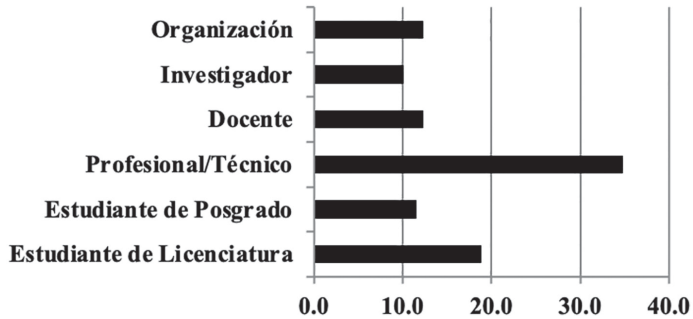
20 Datos de Facebook UNAM/IIBI capturados el 24/02/2020. <https://www.facebook.com/IIBI.UNAM.MX>

21 Datos de X (antes Twitter) UNAM/IIBI capturados el 24/02/2020. [https://twitter.com/IIBI\\_UNAM\\_MX](https://twitter.com/IIBI_UNAM_MX)

22 Datos de Youtube UNAM/IIBI capturados el 24/02/2020. <https://www.youtube.com/user/CUIB100>

de seguidores). Los datos de ocupación de estos usuarios apuntan a una comunidad así compuesta:

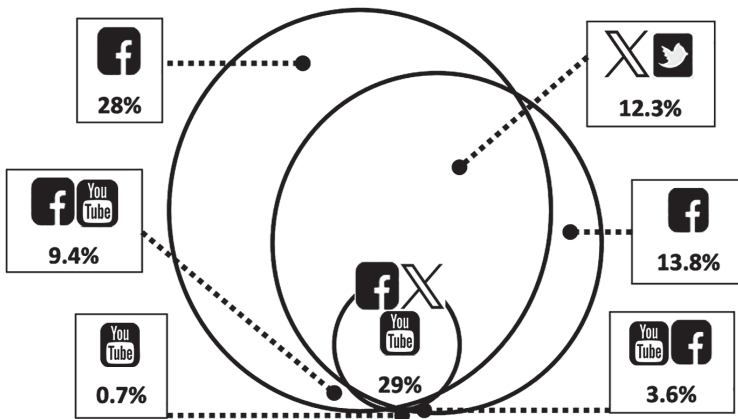
*Figura 2. Ocupación de usuarios de Facebook, X y Youtube*



Fuente: Elaboración personal.

En relación con la distribución de los usuarios que contestaron a la encuesta en las RS estudiadas, la visualización es la siguiente:

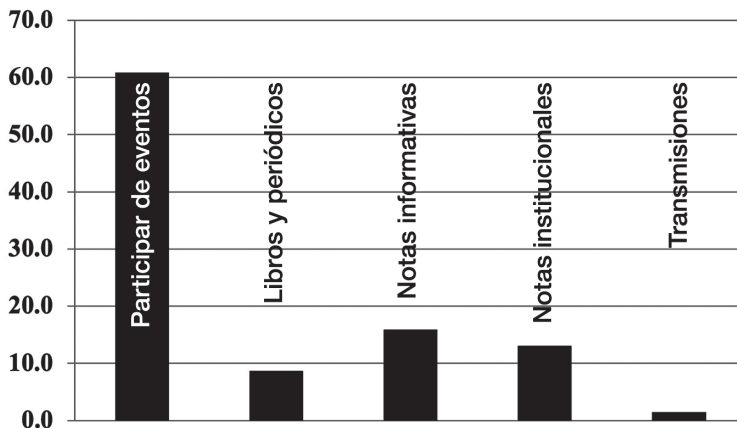
*Figura 3. Distribución de los usuarios que contestaron la encuesta en las RS*



Fuente: Elaboración personal.

Finalmente, los objetivos de esta muestra de usuarios con el seguimiento de estas RS están así clasificados:

Figura 4. Objetivos con el seguimiento de las RS



Fuente: Elaboración personal.

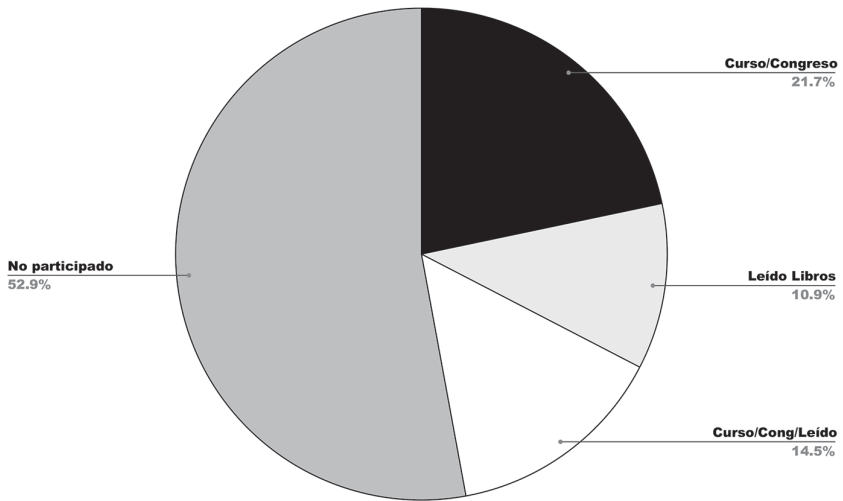
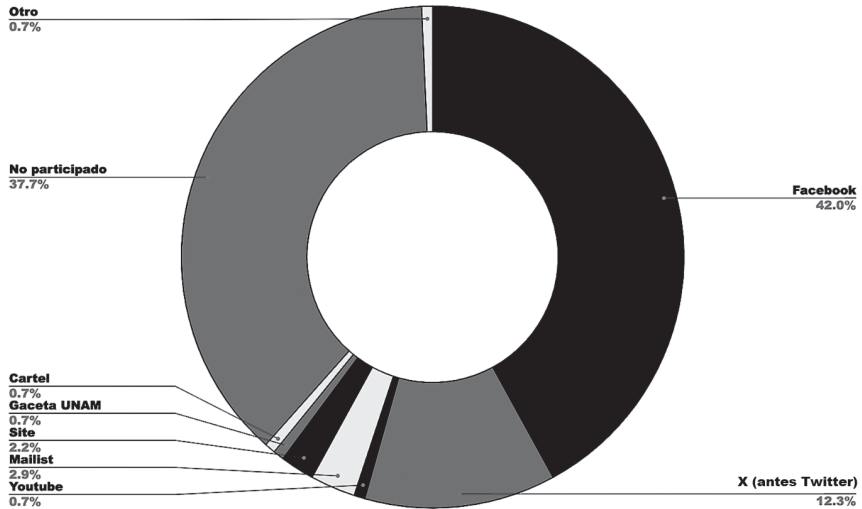
## El impacto de las RS en la entidad/organización

Con el perfil de usuario señalado anteriormente ya fue posible un análisis de impacto. Un primer análisis se refiere a Facebook que, con sus 14 230 seguidores impacta más que X (antes Twitter), con sus 18 342, en las actividades de la entidad. Youtube, por su finalidad (una RS especializada en video y *streaming*) resulta en un impacto asociado a las otras RS, puesto que un 99.3% de sus seguidores son también usuarios de Facebook y X, y a sus publicaciones se accede en buena medida a través de ligas en estas RS.




En los resultados de los KPIS eso se va confirmando. Al mirarse la tasa de conversión a partir de la adhesión a la encuesta, del total de seguidores que la contestaron, un 55.9% lo hizo a partir de Facebook (figura 5). De los eventos incursionados por la entidad/organización en el año de 2019, 42% han participado enterándose por esa red, lo que supera a medios como el *site* de la entidad

(2.2%) y el *mailist* (2.9%). La figura 6 completa la percepción de impacto con los KPIS *visibilidad*, *engagement* y *CTR*.

Figura 5. Participación del público en eventos y tasa de conversión de las RS

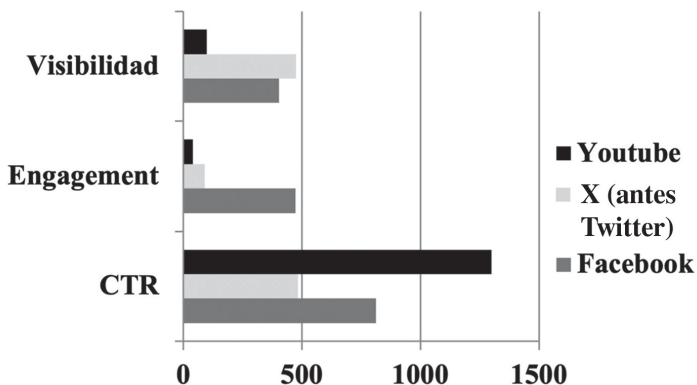


### Metadatos sociales: iniciativas...

Conversión de la encuesta		
		
55.9%	25.1%	0.0%

Fuente: Elaboración personal.

Figura 6. Visibilidad, engagement y CTR de las RS de la entidad/organización



Fuente: Elaboración personal.

## EL ANÁLISIS

La investigación ha constatado un perfil de usuarios de las RS de la entidad/organización, compuesto, en su mayoría, de profesionales o técnicos externos a la institución, los estudiantes de licenciatura son el segundo público con más expresión y son seguidos por los estudiantes de posgrado, docentes y otras organizaciones. Los investigadores aparecen en último.

Una posible explicación para estas características iniciales se puede ofrecer con la mirada a los objetivos de estos usuarios con

el seguimiento de las RS del IIBI: un 60% afirma que es para participar en los eventos promocionados por la entidad. Esto debido a que la universidad, al no ser el ambiente habitual de los profesionales/técnicos externos, se constituye como un lugar para obtener formación y actualización de conocimientos y, entre otros, las RS son los medios para mantenerse enterados de sus ofertas.

La cantidad de usuarios se distribuye de la siguiente manera en las RS mexicanas: Facebook concentra la participación de más de 40% de la muestra y X el 26%. A partir de ello, hay que destacar dos fenómenos observados: el primero se refiere al tamaño de la comunidad Facebook de la entidad/organización que, con un número de seguidores menor que los 18 342 de X, impacta más en las actividades, teniendo en vista la tasa de conversión de la encuesta (55%), el nivel de *engagement* y CTR; el segundo fenómeno es que se percibe un comportamiento distinto del usuario en estas RS: un 12.3% asegura estar tanto en Facebook como en X, pero en una se muestra más *engagement* y actividad y en otra menos.

Una explicación a eso se constata en el trabajo de Julia Llanos Alonso,<sup>23</sup> en el cual afirma que dentro de todas las redes sociales, una de las más complejas para el estudio de la personalidad, pero a la vez con mayor potencial, es X. En trabajos previos<sup>24</sup> se ha demostrado la utilidad de algunas de sus características frente a otras como Facebook. Algunas de sus características más relevantes son su elevado número de usuarios, la existencia de relaciones unidireccionales y bidireccionales entre usuarios, la alta frecuencia con la que se crean relaciones entre personas desconocidas, y la alta cantidad de datos totalmente públicos.<sup>25</sup>

---

23 Julia Llanos Alonso, “Análisis de personalidad en Twitter mediante modelos de usuario basados en extracción de características y aprendizaje colaborativo”.

24 Véanse las Referencias Bibliográficas, principalmente “Una revisión socio-técnica de las redes sociales...”, 28, 29.

25 Llanos Alonso, *op. cit.*, 6-7.



Sin intenciones de plantear o reducir aquí el estudio a uno de personalidad de usuarios, esa afirmación conduce al entendimiento de una de las razones del bajo *engagement* y CTR de X entre estos usuarios, es decir, con un público especialmente enfocado en las relaciones unidireccionales y bidireccionales entre usuarios y personas desconocidas, no hay motivaciones para una interacción más allá que las visualizaciones y la pertenencia a esta RS.

Youtube, por su finalidad especializada en video y *streaming*, tiene un impacto asociado al Facebook y X y ha sido identificada como la RS de mayor CTR entre las tres por ser una red caracterizada en la entidad/organización como un repositorio de sus transmisiones y producciones audiovisuales. Así, aunque con solamente 13 publicaciones de transmisiones del instituto en el transcurso de 28 días, se accedieron a ellos desde Facebook y X y su visibilidad y *engagement*, por lo tanto, están asociados a la visibilidad y *engagement* de estas RS, explicando el alto número de clics.

En estas percepciones, las tasas de conversión de estas RS ubican a Facebook como la red que más participación de usuarios en los eventos de la entidad genera: casi la mitad (42%) registraron haber participado en el año de 2019 en algún evento de la entidad (taller, curso, congreso, leído o comprado libros); la tasa de conversión en X fue de 25%, lo que también confirma la diferente relación que tiene el usuario en esa RS.

Finalmente, entre los eventos de la entidad/organización que los usuarios informaron participar enterándose por las RS, la participación en talleres, cursos y congresos representa un 21.7%, la lectura o compra de libros, 10.9% y ambos 14.5%, lo que demuestra el impacto que tiene la comunicación vía RS en la toma de decisión de sus usuarios para enterarse y participar de una actividad. Hay que señalar que, de hecho, ese impacto puede ser más grande si tenemos en cuenta las transmisiones y videos de eventos de la institución no considerados en el cálculo de la tasa de conversión, pero que permitieron la participación en línea.

## CONSIDERACIONES Y RECOMENDACIONES ESTRATÉGICAS

Sin presumir un final en ese estudio, ni la ciega universalización de sus resultados, la investigación nos permitió hacer algunas consideraciones muy pragmáticas y recomendaciones estratégicas fácilmente visualizadas en sus evidencias.

La primera es la aportación al campo de la Bibliotecología y Estudios de la Información de conocimientos necesarios al perfil profesional, que ha avanzado incorporando las nuevas dimensiones de trabajo con la información, sus características y usos. Las RS son espacios de información que deben ser asumidos y considerados por los expertos en sus horizontes de trabajo, especialmente porque cada día tenemos generaciones más conectadas y que cuentan con el ciberespacio como el medio principal de obtener información y de comunicarse.

La segunda es que la presencia en el ciberespacio de personas interesadas en ciertos temas, grupos y organizaciones permite, a través de las RS, una agregación más enfocada y dinámica de necesidades de información, lo que supone más interacción y eficiencia de la comunicación entre pares. Los seguidores de una organización tienen con ella alguna relación de identidad e interés en sus productos, contenidos, actividades, etcétera.

La tercera es que la investigación, al cumplir con sus objetivos, ha demostrado cierto grado de impacto de las RS en la organización universitaria, permitiendo una mejora de su visibilidad, interacción con el público y acceso a sus productos y servicios. A su vez, eso nos permite afirmar que las RS de la entidad investigada están cumpliendo con el objetivo gerencial de la organización de “enriquecer la interacción y colaboración con la comunidad para que se logre una mayor visibilidad y apreciación del quehacer universitario”.<sup>26</sup>

Una cuarta consideración es que, en conformidad con las tendencias de los usuarios de las RS de México, Facebook se ha

---

26 UNAM, *op. cit.*, 2.

confirmado como la red que más impacta en la comunidad de la entidad, sobre todo en materia de *engagement* y conversión. X tiene mayor visibilidad dado el tamaño de la comunidad; sin embargo, por las características comportamentales de los usuarios de esa RS, se observan niveles de *engagement* y conversión más bajos en la entidad. Y Youtube, como una RS especializada en videos y *streaming*, tiene un impacto asociado a las otras RS de la organización, ya que sus videos y transmisiones son accedidos por ligas publicadas en ellas. De ahí que entre las tres RS estudiadas es la que tiene mayor CTR.

Al presentar un perfil de usuarios, fue posible identificar necesidades de información en el flujo de datos de las RS analizadas y, con ello, obtener algunas recomendaciones estratégicas. La primera resalta la gran presencia de profesionales técnicos externos a la organización que tienen como objetivo la actualización en la disciplina a través de la participación en talleres, cursos, congresos y la compra o lectura de libros. Eso refuerza la oferta de educación continua a ese público, que puede servir igualmente al público docente y a los estudiantes de grado y de posgrado. Por otro lado, la entidad como un instituto de investigación, debe también incrementar los seguidores investigadores, ya que ese es el grupo con la menor expresión en sus RS. Una posibilidad sería crear segmentos que permitan la publicidad de los trabajos y actividades de sus Seminarios de Investigación más allá de los congresos, libros y periódicos por ellos producidos.

Queda evidente que Facebook, al ofrecer más impacto, debe seguir como un importante medio de comunicación de las actividades de la entidad. Sin embargo, hay que reconocer y pensar formas de hacer para que X pueda funcionar también estratégicamente, teniendo en cuenta la visibilidad proporcionada por su gran número de seguidores y las características comportamentales de esa red. Una propuesta sería que hubiera más publicaciones de las ligas de video y transmisiones de Youtube en esa RS, ya que la alta visibilidad de X, asociada al alto CTR de Youtube puede resultar en mayor *engagement* y conversión en ambas. También se puede ocupar X prioritariamente con las publicaciones que traigan

ligas externas para el sitio institucional, para eventos y recursos de la entidad disponibles en sus servidores.

Para concluir, ese trabajo abre también muchas posibilidades de profundizar más cuestiones y problemáticas que involucran el universo de las RS, no solamente relacionadas a las organizaciones universitarias y a la República Mexicana, sino también a otras organizaciones y países que seguramente perciben el impacto de esa tecnología de información y comunicación. Aparte de la evidente necesidad de más estudios de usuarios de RS, otro gran reto que se impone está asociado al desarrollo de recursos y capacidades profesionales que permitan un uso más eficiente, por las organizaciones, de las informaciones proporcionadas por las RS.

## REFERENCIAS BIBLIOGRÁFICAS

Aimx (Asociación de Internet MX). *15° Estudio sobre los Hábitos de los Usuarios de Internet en México 2019*. Ciudad de México: Aimx. <https://www.asociaciondeinternet.mx/estudios/habitos-de-internet> (Consultado el 30 de noviembre de 2019).

Brandão, Marco. “El flujo de información a través de las redes sociales”. Proyecto Individual de Investigación del Instituto de Investigaciones Bibliotecológicas y de la Información. Ciudad de México: UNAM/IIBI. 2019.

———. “Una revisión sociotécnica de las redes sociales. La propuesta de indicadores de eficiencia bajo el flujo de información de una organización universitaria en la web”, Informe Final de los becarios extranjeros que realizan una estancia de investigación de posgrado. Ciudad de México: SRE/AMEXCID. 2020.

———. “Una revisión sociotécnica de las redes sociales de una organización universitaria”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35, núm. 86 (2021): 161-184.

### ***Metadatos sociales: iniciativas...***

- Caldevilla Dominguez, David. "Las redes sociales. Tipología, uso y consumo de las redes 2.0 en la sociedad digital actual". *Documentación de las Ciencias de la Información* 33: 45-68. <https://revistas.ucm.es/index.php/DCIN/article/view/DCIN1010110045A/18656> (Consultado el 14 de enero de 2020).
- Contreras Orozco, Leticia. "Las redes sociales digitales como factor de proximidad entre la sociedad y los gobiernos subnacionales en México". *Revista Iberoamericana de las Ciencias Sociales y Humanísticas* 6, núm. 12 (Julio/Diciembre). Acceso el 13 de enero de 2020. <http://www.redalyc.org/articulo.oa?id=503954320009>.
- Devlin, Keith. *Logic and Information*. Cambridge: Cambridge University Press. 1991.
- Emery, Fred E. y Eric L. Trist. *Socio-technical Systems. Management Sciences Models and Techniques*. London: Pergamon Press. 1960.
- Hütt Herrera, Harold. "Las redes sociales: una nueva herramienta de difusión". *Reflexiones* 91, núm. 2 (2012). <http://www.redalyc.org/articulo.oa?id=72923962008> (Acceso el 13 de enero de 2020).
- Kemp, Simon. *Digital 2019: essential insights into how people around the world use the internet, mobile devices, social media, and e-commerce*. Vancouver: Hootsuite. 2019. <https://hootsuite.com/resources/digital-in-2019> (Consultado el 30 de noviembre de 2019).
- Llanos Alonso, Julia. "Análisis de personalidad en Twitter mediante modelos de usuario basados en extracción de características y aprendizaje colaborativo". Trabajo de fin de máster en Ingeniería Informática. Madrid: Universidad Autónoma de Madrid, Departamento de Ingeniería Informática. 2015. [https://repositorio.uam.es/bitstream/handle/10486/669358/Llanos\\_Alonso\\_Julia\\_tfm.pdf?sequence=1&isAllowed=y](https://repositorio.uam.es/bitstream/handle/10486/669358/Llanos_Alonso_Julia_tfm.pdf?sequence=1&isAllowed=y) (Consultado el 23 de abril de 2020).

- Manrique Valenzuela, Katy y Miguel Domingo González Álvarez. “Desarrollo de Sistemas Socio Técnicos en el área de Seguridad y Salud Ocupacional de una empresa de servicio”. *ALTEC 2013*. Lisboa: ALTEC. 2013. [http://www.altec2013.org/programme\\_pdf/429.pdf](http://www.altec2013.org/programme_pdf/429.pdf) (Consultado el 15 de enero de 2020).
- Sánchez Vanderkast, Egbert J. “Los flujos de información: algunas reflexiones”. *Investigación Bibliotecológica* 16, núm. 32 (Enero/Junio). <http://rev-ib.unam.mx/ib/index.php/ib/article/view/3995/3546> (Consultado el 04 de febrero de 2020).
- Torres Vargas, Georgina Araceli, Egbert J. Sánchez Vanderkast, Ariel Alejandro Rodríguez García y Jenny Teresita Guerra González. “The Blended Librarian and the Disruptive Technological Innovation in the Digital World”. *Open Access Library Journal* 2, e1764 (2015). <http://dx.doi.org/10.4236/oalib.1101764> (Consultado el 04 de febrero de 2020).
- Triviños, Augusto Nivaldo Silva. *Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação*. São Paulo: Atlas. 1987.
- UNAM (Universidad Nacional Autónoma de México, Dirección General de Comunicación Social, Comité de Comunicación de la Dirección General de Comunicación Social). *Instructivo de uso para Redes Sociales Institucionales de la UNAM*. Ciudad de México: UNAM/DGCS. 2014. <https://www.visibilidadweb.unam.mx/normateca/normaunam/Lineamientosredessociales.pdf> (Consultado el 28 de noviembre de 2019).

# TECNOLOGÍAS

# Metadatos para documentos fílmicos: iniciativas y estándares

HILDA GABRIELA LOBATÓN CRUZ  
*Cineteca Nacional de México, México*

## INTRODUCCIÓN

**E**l cine, si bien inicialmente fue considerado un mero espectáculo, en la actualidad tiene un papel de suma relevancia para la sociedad. A través de las obras cinematográficas, hemos podido no sólo entretenernos, sino comunicarnos y conectarnos por medio de historias de la vida real y hemos conocido y revivido hechos que han marcado las diferentes épocas de la humanidad. En este sentido, se puede decir que el cine es un reflejo de la sociedad y que obtiene su sustento, ideas e imaginación de sus alrededores y de los hechos y acontecimientos que se están llevando a cabo en el momento de la creación de su argumento, por lo que las dinámicas sociales son indispensables para el contenido de las obras fílmicas.

Por este motivo, lo audiovisual ha llegado a convertirse en un instrumento valioso para la investigación, dado que puede funcionar como documento o testimonio.

Por ello, los archivos fílmicos, como entidades encargadas del resguardo, preservación y difusión de la memoria fílmica, deben establecer procesos que ayuden a la gestión, acceso y uso de las



obras cinematográficas, pero para ello, deben transitar por diversas fases del análisis documental al igual que otros tipos de documentos.

Sin embargo, el procesamiento sistemático de una obra fílmica es complejo dadas sus características particulares, por lo que se requieren de estructuras de metadatos que puedan representar a estas obras con la finalidad de facilitar la búsqueda y recuperación de información en las imágenes en movimiento.

La asignación de metadatos ofrece la posibilidad de recuperación de información de las obras fílmicas, además de posibilitar el relacionar un recurso con otro en el contexto de una o varias colecciones. Asimismo, dentro de la industria fílmica, se apoyan los procesos de creación, producción, descripción e identificación de contenidos, protección de derechos, intercambio, distribución y consumo.

Por lo tanto, esta información contextual permitirá asegurar la autenticidad, integridad, fiabilidad, usabilidad y valor de los documentos cinematográficos.

## PATRIMONIO CINEMATOGRAFICO

El papel cada vez mayor que cobran los medios audiovisuales en la construcción de la memoria y la cultura de los pueblos, sobre todo para las nuevas generaciones, hace surgir la necesidad de crear consciencia en instituciones académicas y organizaciones de todo tipo acerca de la importancia de preservar y definir métodos que faciliten el acceso público a estos materiales.

Las recomendaciones para la salvaguarda y conservación de las imágenes en movimiento, dictadas por la UNESCO en 1980, supusieron un hito importante, tanto para los archivos como para la documentación fílmica, al reconocerse mundialmente de forma “oficial”, por primera vez, el valor patrimonial de los documentos audiovisuales.

Como bien lo menciona Ray Edmondson, “los documentos audiovisuales no son menos importantes que otros tipos de

documentos o artefactos, y en algunos contextos lo son más. Su carácter a menudo populista y su vulnerabilidad a una tecnología en rápida evolución no les restan importancia. Su conservación y accesibilidad deben garantizarse”.<sup>1</sup>

Aunque el reconocimiento fue tardío, muchos cinéfilos y estudiosos del cine ya tenían previamente preocupación por la conservación y resguardo de los documentos fílmicos. Así es como en 1938 se funda la Federación Internacional de Archivos Fílmicos (FIAF). El propósito de esta nueva organización era facilitar el intercambio de filmes históricos, educacionales y artísticos entre las diferentes instituciones miembro, como la British National Film Library, el German Reichsfilmarchiv, la Cinémathèque Française y The American Museum of Modern Art; sólo estos cuatro archivos participaron en su fundación. Esta organización no se encontraba involucrada con películas comerciales; de hecho, ser miembro no estaba dirigido a cualquier institución u organización y menos aún a las que pudieran hacer uso de las películas con propósitos comerciales.<sup>2</sup>

Una vez comenzados los trabajos de la FIAF, y con un número creciente de miembros, los archivos fílmicos empezaron a vislumbrar una mayor necesidad de estudiar y dar solución a problemas derivados de la preservación, conservación, restauración y organización de las colecciones fílmicas.

Derivado de estas problemáticas, en 1968 se creó la Comisión de Catalogación y Documentación. Esta comisión se enfocó en primera instancia en la organización de documentos impresos relacionados a la cinematografía.<sup>3</sup> Su programa inicial incluía un estudio para la indización de publicaciones periódicas especializadas en cine, sin embargo, también se llevaron a cabo reuniones para estudiar los problemas comunes del proceso de catalogación de películas. Como resultado de estas reuniones, la comisión preparó el manual titulado “Film Cataloguing”, el cual es una recolección de

---

1 Ray Edmondson, *Filosofía y principios de los archivos audiovisuales*, 8.

2 Christophe Dupin, “The origins of FIAF, 1936-1938”, 43.

3 Raymond Borde, *Los Archivos Cinematográficos*, 12.

### *Metadatos sociales: iniciativas...*

las experiencias de los archivos miembros de la FIAF y una discusión detallada de los principios básicos de catalogación de archivos fílmicos. Presenta también numerosos ejemplos que servirían de guía para los archivos que pretendieran establecer un sistema de catalogación. Pero fue hasta 1970 que a nivel mundial se logró normalizar la catalogación de las imágenes en movimiento con las Reglas de Catalogación de la FIAF para Archivos Fílmicos.

En algunos archivos se creó una metodología tendiente a hacer una descripción de contenido más detallada para poder hacer frente a las demandas de los investigadores que solicitaban planos, tomas o secuencias específicas.

### CINE DIGITAL

A medida que los días de las películas analógicas llegan a su fin, los archivos fílmicos deben confrontar una gran cantidad de nuevos asuntos: los aparatos y sistemas de comunicación entre el archivo digital y el equipo utilizado para su proyección, su almacenamiento, consulta, los formatos de archivos de imagen y sonido, el sistema de procesamiento de la imagen, etcétera.

Esta complejidad no representa un gran problema a corto plazo, pero para los propósitos de los archivos fílmicos comienza a convertirse en una preocupación real.

En cuanto al análisis documental de las películas digitales, los metadatos cumplen un rol crucial para mantener un acceso a largo plazo a materiales digitales y digitalizados. Esto significa que la información debe incluir una descripción de su contenido, su formato, qué hardware y software fueron utilizados para su creación, de qué manera fue codificado, etcétera. Además, saber que los datos digitales necesitarán ser administrados para siempre, por lo que su organización y procesos de preservación deben estar constituidos para perdurar.

En este sentido, la FIAF aún tiene mucho trabajo por realizar. En sus reglas de catalogación se contemplan las copias digitales, sin embargo, no se concibieron para considerar todos los aspectos

inherentes al documento digital. Por tanto, es importante que se plantee una pronta revisión y actualización del Moving Image Cataloguing Manual para incluir la catalogación de los documentos fílmicos de origen digital.

Si bien este panorama puede vislumbrarse complicado para los archivos fílmicos, lo ideal es seguir las mejores prácticas posibles, utilizando normas y estándares que garanticen la estabilidad y permanencia de estos documentos en cualquiera de sus soportes o formatos.

## METADATOS PARA DOCUMENTOS AUDIOVISUALES

Los metadatos para los materiales audiovisuales son cruciales cuando se considera el enorme volumen de bits dentro de las representaciones de una imagen en movimiento digital. Incluyen una amplia información que se puede utilizar para identificar, autenticar y contextualizar estos documentos.

En la industria audiovisual, los pioneros en el diseño y aplicación de sistemas de metadatos fueron las televisoras. Posteriormente, estos sistemas fueron adaptados por el resto las industrias, incluida la cinematográfica.

A continuación, se presentan las principales iniciativas de metadatos utilizadas por las televisoras:

- SMEF (Standard Media Exchange Framework).  
Creada por la BBC (British Broadcasting Corporation) para integrar su sistema de contenido de información audiovisual. Propuesta muy interesante ya que recoge información desde la creación del proyecto hasta su difusión.<sup>4</sup>

---

<sup>4</sup> Diane McGregor, "Developing and Implementing the Standard Media Exchange Framework".

### *Metadatos sociales: iniciativas...*

- Video Active.  
Trabajo que tiene como finalidad el acceso a los contenidos de varias cadenas de televisión. Este proyecto surge de eCONTENTplus de la Unión Europea, cuyo objetivo era poner a disposición de la Unión Europea los contenidos audiovisuales.<sup>5</sup>
- Echo (European Chronicles On-line).  
Proyecto financiado por la Unión Europea con el objetivo de lograr la creación de una infraestructura software que pueda ser mantenida por los archivos audiovisuales con colecciones digitalizadas, con el fin de facilitar la accesibilidad de material con interés histórico. Entre algunas de sus características se encuentra la posibilidad de llevar a cabo indización automática por medio de la extracción de términos significativos, así como el acceso inteligente. El formato de metadatos utilizado aunaba los elementos más importantes de los formatos Dublin Core y MPEG-7 basado en RDF.<sup>6</sup>

Otros estándares de metadatos aplicados en la industria audiovisual son:

- Open Video Digital Library.  
Iniciativa estadounidense que trabaja en el acceso a información audiovisual de diversas fuentes gubernamentales. Este proyecto utiliza los metadatos para el acceso y descripción del material almacenado en su librería digital.<sup>7</sup>
- Advanced Media Workflow Association (AMWA).  
Nace para mejorar el intercambio de imágenes y sonido por medio de diferentes aplicaciones. Trabaja con grandes

---

5 Video Active, "IASA World Conference 2009".

6 European Chronicles On-Line, "Fact sheet".

7 Gary Marchionini y Gary Geisler, "Open video digital library".

colecciones visuales y de documentales audiovisuales en la fase de creación, únicamente.<sup>8</sup>

- Dublin Core.

En el año 2000 se crea el grupo especializado en archivos de imágenes, que intenta normalizar los metadatos para este tipo documental. Una vez analizadas las diferentes propuestas existentes, intenta adaptarlas a Dublin Core. Son muchas las iniciativas que se han puesto en marcha utilizando Dublin Core, destacando a Australian Center for Moving Image que utiliza Dublin Core expresados en XML. Es además la base para formatos tales como MPEG 7 y 21, SMPTE, P-META y P-FRA.<sup>9</sup>

- Lista de Datos Mínimos de la Federación Internacional de Archivos de Televisión (FIAT-IFTA).

Aunque no puede ser considerado exactamente como un estándar de metadatos como los anteriores al no ser legible por máquina, sí aporta todos los elementos necesarios para describir el contenido de los documentos audiovisuales para programas de entretenimiento. Aunque es muy limitado y bastante antiguo, es una herramienta pionera en la descripción de elementos necesarios y que agrupa por epígrafes.<sup>10</sup>

## METADATOS PARA DOCUMENTOS FÍLMICOS

La catalogación digital con metadatos sólidos es una parte vital del proceso de conservación y archivo de películas. Cuanto mayor es el tamaño de la colección, más crítico se vuelve este proceso; sin embargo, a menudo los custodios de estas colecciones lo pasan por alto y lo subestiman.

---

8 Advance Media Workflow Association, <https://www.AMWA.tv/>.

9 Dublin Core, <https://www.dublincore.org/>.

10 Federación Internacional de Archivos de Televisión, <https://fiatifta.org/>.

### *Metadatos sociales: iniciativas...*

Muchos estudios cinematográficos importantes y otros poseedores de grandes colecciones fílmicas desarrollan y mantienen sus propios estándares de metadatos. La cantidad de campos en sus esquemas de metadatos es variable y puede estar determinada por algunos factores, como el software utilizado para almacenar los datos y el uso previsto de la imagen.

En cine, los metadatos más usuales son:

- Título de la obra fílmica
- Director
- Productor
- Reparto o elenco
- Sinopsis
- Género
- Clasificación por audiencia

Los estándares de metadatos implican principios de codificación y gestión de información altamente estructurados. Ya no es suficiente sólo contar con metadatos descriptivos como los mencionados arriba. Adicionalmente, se requiere de la comprensión humana sobre el contenido de la obra fílmica, es decir, la intervención de personas calificadas con un amplio conocimiento sobre cine y que, además, sean capaces de identificar personas, lugares y que comprendan e interpreten el lenguaje cinematográfico.

Iniciativas de metadatos para documentos fílmicos:

- Metadata Standards for Cinematographic Works CEN-EN 15744 Film Identification - Minimum Set of Metadata for Cinematographic Works.  
EN 15744 es un conjunto de metadatos para la identificación básica de obras cinematográficas preparadas por el Comité Europeo de Normalización (CEN).  
Es un conjunto de metadatos que contiene quince elementos básicos para la identificación de obras cinematográficas. La descripción que se puede obtener con su implementación es una simple identificación de los elementos más

básicos de una obra fílmica, dejando de lado los elementos relacionados son los soportes, visionado, requisitos técnicos, así como sus condiciones de acceso. Tiene una estructura abierta, ya que da la posibilidad de añadir otros datos. Esta característica está retomada de Dublin Core, el cual también presenta una estructura compuesta por un determinado número de elementos.<sup>11</sup>

- Metadata Standards for Cinematographic Works CEN - EN 15907 Film Identification - Enhancing Interoperability of Metadata - Element Sets and Structures.

EN 15907 es un conjunto de metadatos para la descripción completa de obras cinematográficas, incluidas las diversas versiones que puedan surgir durante su ciclo de vida.

Esta norma europea especifica un conjunto de metadatos para la descripción de obras cinematográficas, así como una terminología para uso de las partes que deseen intercambiar dichos metadatos descriptivos. También define algunas entidades y relaciones básicas útiles para definir modelos de datos, así como para estructurar representaciones serializadas y ordenadas jerárquicamente de metadatos sobre obras cinematográficas, incluidas sus variantes, manifestaciones y elementos. Esta norma define un modelo de datos mediante entidades y relaciones del mismo modo que FRBR, estableciendo atributos y elementos para cada entidad y relación.<sup>12</sup>

- FIAF Moving Image Cataloguing Manual

En mayo de 2016 los miembros de la comisión realizaron la revisión de las Reglas de Catalogación de la FIAF para Archivos Fílmicos ahora tituladas FIAF Moving Image Cataloguing Manual, en donde al igual que en RDA utilizan el

---

11 CEN, European Committee for standardization, "EN 15744:2009 Film identification - Minimum set of metadata for cinematographic works".

12 CEN, European Committee for standardization, "EN 15907:2010 Film identification - Enhancing interoperability of metadata - Element sets and structures".



modelo conceptual FRBR y la norma en 15907 para la catalogación de las imágenes en movimiento. Este manual toma en cuenta la evolución de la tecnología de la información y se alinea con los estándares de metadatos existentes y emergentes. Está diseñado para ser compatible con una variedad de estructuras de datos. Proporciona gráficos, árboles de toma de decisiones, ejemplos y otras herramientas para ayudar a expertos y no expertos en la catalogación de colecciones de imágenes en movimiento.<sup>13</sup>

## CONCLUSIONES

En conclusión y respecto a lo expuesto, se puede considerar que cualquier iniciativa encaminada a lo que sería una mayor normalización de los materiales fílmicos parece, más que conveniente, una necesidad. Cabe aclarar que los metadatos aquí presentados no son los únicos, pero son los que mejor aplicación han tenido en la industria audiovisual.

En los últimos años ha habido un desarrollo y proliferación en las normas y estándares de metadatos que han permitido un mayor control y coherencia de los datos en el entorno de las producciones fílmicas. La aplicación y uso de dichos estándares se ha desarrollado en diferentes entornos y sectores para la descripción, recuperación e intercambio de recursos mantenidos, producidos y salvaguardados por las productoras y los archivos fílmicos internacionales. Sin embargo, en los países latinoamericanos aún hay mucho trabajo por hacer en cuanto a la aplicación de metodologías que puedan posibilitar el acceso a sus colecciones fílmicas.

Aunado a esto, la producción digital cada vez es mayor y, por ende, también las problemáticas respecto a su organización crecen. La importancia de crear una serie de criterios y normas que sean de aplicación común para facilitar las tareas de análisis

---

13 International Federation of Film Archives, "Reglas de catalogación de la FIAF".

documental son indispensables para llevar a cabo proyectos e iniciativas de cooperación, ya que sin un consenso en la forma de descripción de la documentación fílmica los proyectos futuros no podrán realizarse de forma satisfactoria.

Asociar metadatos al material en muchos casos ha traído consigo problemas añadidos, tales como la subjetividad propia del análisis de este material, la dificultad de interpretación y los requerimientos propios de los usuarios y de la especificidad de los sistemas de información.

Aunque es cierto que en las últimas décadas se ha progresado significativamente en el reconocimiento al documento fílmico como fuente de información y, por ende, la importancia de su organización, aún no se ha logrado lo mismo en el análisis de contenido y en las posibilidades de recuperación de información cinematográfica por parte de los usuarios.

Debemos considerar que es justo y necesario que los usuarios de archivos fílmicos también tengan oportunidad de recuperar información de las imágenes en movimiento. Esto significaría un gran apoyo, ahorro de tiempo y de trabajo para encontrar lo que buscan, algo muy importante en un mundo dominado por la saturación informativa y las nuevas tecnologías. Por ello, es relevante que se abran nuevas líneas de investigación sobre este tema, manteniendo un contacto continuo entre los investigadores que se desempeñan en el campo de la documentación fílmica, los bibliotecarios que trabajan en archivos fílmicos y los usuarios que visitan estos archivos, con la finalidad de hacer más accesible el patrimonio fílmico y que se promueva su uso en la investigación.

Finalmente, se puede constatar que existe una gran distancia entre las iniciativas, las normas, la legislación cinematográfica y las tecnologías, así como las necesidades informacionales instauradas en la sociedad, y la situación actual de las instituciones responsables del patrimonio cinematográfico. Muestra de ello es la falta de presencia en internet, la carencia de catálogos en línea y la falta de normalización de los datos que se ofrecen en las plataformas. Como contrapunto, tal y como hemos visto, sí existen iniciativas promovidas por las entidades relacionadas con la industria

### *Metadatos sociales: iniciativas...*

audiovisual, algunas actualmente se encuentran en fase de gestación, pero el reto está en que dichas iniciativas permanezcan para que en un futuro permitan el acceso al patrimonio fílmico nacional e internacional.

### REFERENCIAS BIBLIOGRÁFICAS

Advance Media Workflow Association. <https://www.amwa.tv/> (Consultado el 3-08-2022).

Borde, Raymond. *Los Archivos Cinematográficos*. Valencia: Filmoteca de la Generalitat Valenciana. 1991.

CEN (European Committee for standardization). “EN 15744: 2009 Film identification - Minimum set of metadata for cinematographic works”. [http://filmstandards.org/fsc/index.php/EN\\_15744](http://filmstandards.org/fsc/index.php/EN_15744) (Consultado el 28-07-2022).

———. “EN 15907:2010 Film identification - Enhancing interoperability of metadata - Element sets and structures”. [http://filmstandards.org/fsc/index.php/EN\\_15907](http://filmstandards.org/fsc/index.php/EN_15907) (Consultado el 03-08-2022).

Dublin Core. <https://www.dublincore.org/> (Consultado el 1-08-2022).

Dupin, Christophe. “The origins of FIAF, 1936-1938”. *Journal of Film Preservation* 88, Abril (2013).

Edmondson, Ray. *Filosofía y principios de los archivos audiovisuales*. México: UNESCO. 2008.

European Chronicles On-Line. “Fact sheet”. <https://cordis.europa.eu/project/id/IST-1999-11994> (Consultado el 1-08-2022).

Federación Internacional de Archivos de Televisión. <https://fiatifta.org/> (Consultado el 2-08-2022).

- International Federation of Film Archives. "Reglas de catalogación de la FIAF". París: FIAF. 2016. <https://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/CDC-resources/20160920%20Fiaf%20Manual-WEB.pdf> (Consultado el 10-07-2022).
- Marchionini, Gary y Gary Geisler. "The Open video digital library". *D-Lib Magazine* 8, núm. 12 (december 2002). <https://www.dlib.org/dlib/december02/marchionini/12marchionini.html> (Consultado el 2-08-2022).
- McGregor, Diana. "Developing and Implementing the Standard Media Exchange Framework". *Finding Clip* 7, Paper 5071 (1999).
- Video Active. "IASA World Conference 2009". <https://www.slideshare.net/mrendina/videoactive-iasa-world-conference-2009> (Consultado el 2-08-2022).

# Curación de metadatos para recursos educativos digitales

ANA CAROLINA SIMIONATO ARAKAKI  
*Universidad Federal de São Carlos, Brasil*

## INTRODUCCIÓN

Cada año, los estudiantes de primaria, educación juvenil y de adultos y preparatoria de escuelas públicas inscritos en el Censo Escolar del Instituto Nacional de Estudios e Investigaciones Educativas Anísio Teixeira (INEP) tienen derecho a recibir libros de texto distribuidos por el Fondo Nacional para el Desarrollo de la Educación (FNDE), de manera sistemática, regular y gratuita.

Estos materiales adquiridos parten del Programa Nacional de Libros y Material Didáctico (PNLD). El PNLD se basa en el art. 208, VII, la Constitución Federal de 1988, la Ley n° 9.394/1996 y el Decreto n° 9.099/2017 y es administrado por el Ministerio de la Educación (MEC), representado por la Secretaría de Educación Básica (SEB) y la Secretaría de Modalidades Especializadas de Educación (Semesp). El Programa es integral y comprende un conjunto de acciones orientadas a la distribución de obras didácticas, pedagógicas y literarias, entre otros materiales de apoyo a la práctica educativa, dirigidas a estudiantes y docentes de escuelas públicas de educación básica del país. El PNLD también incluye instituciones

comunitarias, confesionales o filantrópicas sin fines de lucro y está contratado con las autoridades.<sup>1</sup>

La producción de obras didácticas para el PNLD presenta un movimiento significativo del mercado editorial brasileño. En cada convocatoria, las editoriales presentan sus propuestas y a partir de la selección por parte de SEB/MEC se construye una guía con algunos metadatos y un resumen de los trabajos, para que profesores y directores de centros públicos puedan comentar sus contenidos y comprobar cuáles son los más adecuados para el trabajo en el aula.

Además de la selección de materiales bibliográficos, existen Recursos Educativos Digitales (REDS). Los REDS consisten en objetos digitales que ayudan en las clases y el aprendizaje de los estudiantes, con una característica de ser materiales autónomos y pueden presentarse en diferentes formatos y extensiones. De acuerdo con la Ordenanza No. 451 del 16 de mayo de 2018, los Recursos Educativos Digitales incluyen materiales de enseñanza, aprendizaje, investigación, pedagógicos y de gestión escolar en apoyo digital, incluidos *libros electrónicos*, folletos, guías, aplicaciones, software, plataformas, videojuegos y contenido digital.<sup>2</sup>

Entre esta selección de REDS, destacamos el gran reto del aviso y la elección de los REDS a la hora de planificar los detalles específicos respecto a definiciones mínimas de calidad de objeto, obsolescencia de la extensión, estandarización de metadatos para una eficiente recuperación y acceso, así como condiciones mínimas para su acceso, uso y reutilización. Esto se debe principalmente a la recomendación y responsabilidad del propio aviso al indicar que los tintos comprados estarán disponibles posteriormente en licencia abierta, y se ofrecerán en repositorios o plataformas

---

1 Brasil, *Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União*; Brasil, *Dispõe sobre o Programa Nacional do Livro e do Material Didático. Diário Oficial da União*.

2 Brasil, Ministério da Educação, *Define critérios e procedimentos para a produção, recepção, avaliação e distribuição...*

públicas, como el ejemplo de MEC RED, disponible para su consulta por enlace (<https://plataformaintegrada.mec.gov.br/>).

Así, los atributos para la descripción de los REDS ya están indicados por el aviso complementario de PNLD y los editores deben entregar una ficha técnica conjunta al RED, que contenga los metadatos ordenados por 1) Título; 2) Descripción; 3) Palabras clave; 4) Área de conocimiento o Tema Integrador; 5) Competencias de la Base Curricular Nacional Común (BNCC); 6) Habilidades BNCC; 7) Autor; 8) Descriptivo; 9) Código alfanumérico BNCC y 10) Tipo de licencia.<sup>3</sup>

Sin embargo, existe una inconsistencia en relación con su estructuración en los metadatos del MEC RED, identificada por la falta de estandarización del uso de la estructura predefinida por el aviso, así como en la finalización de su contenido para su disponibilidad a todas las partes interesadas. A partir de este contexto presentado, existe la necesidad de identificar una dirección curatorial de metadatos para la plataforma MEC RED y para la estructuración de los atributos de cada RED.

En este sentido, la curación de metadatos se basa en los principios de la curación digital y estos pueden ofrecer un conjunto de acciones para la gestión de los objetos digitales, con el fin de garantizar el mantenimiento de características como autenticidad, fiabilidad, integridad y usabilidad para su futura reutilización de datos.<sup>4</sup> Esto se debe a que la curaduría digital designa como campo de estudio la definición de estrategias y procedimientos para el tratamiento y organización de los datos, con el objetivo de su acceso y preservación.

El objetivo de este trabajo es que la aplicación de los conceptos de curaduría digital puede maximizar los procesos de recuperación, representación, organización, acceso, uso, reutilización y preservación de los Recursos Educativos Digitales. La propuesta

---

3 Brasil, Ministério da Educação, *Define critérios e procedimentos para a produção, recepção, avaliação e distribuição.*

4 Robert K. Glushko, *The Discipline of Organizing*; Sarah Higgins, “Using Metadata Standards”.

es desarrollar un modelo de metadatos que aborde los conceptos de curación digital para describir los recursos educativos digitales.

## CONTRIBUCIÓN TEÓRICA

Higgins señala que los metadatos constituyen la “columna vertebral” de Digital Curation y señala que los metadatos pueden ser utilizados para “mejorar la accesibilidad y el descubrimiento de recursos informativos, así como controlar los procedimientos de autenticación mediante la creación de pistas de auditoría para garantizar que el material no pueda ser accedido o alterado por aquellos que no están autorizados a hacerlo”.<sup>5</sup> Estas características son posibles debido a que “los metadatos consisten en una serie de elementos que se pueden clasificar en las diferentes funciones que ofrecen”,<sup>6</sup> y pueden dividirse en cuanto al propósito de su uso.

Los metadatos son fundamentales para la publicación de datos, junto con los métodos de modelado de información conceptual. Como señalan Isotani y Bittencourt,<sup>7</sup> los metadatos pueden definirse como información adicional que describe los datos contenidos en los recursos, por lo tanto, son datos sobre datos.

Los metadatos no son exclusivos de las áreas relacionadas con la información y la tecnología, estos fragmentos informativos permean las tablas de nutrición alimentaria a formatos más complejos, como un conjunto de datos gubernamentales. Es decir, los metadatos son el contenido de datos resumidos, contextualizados, estructurados, interrelacionados y con marca de procedencia (información histórica y de orígenes), y agregan relevancia y propósito a los datos y permiten la identificación de datos similares en diferentes recopilaciones de datos.<sup>8</sup>

---

5 Sarah Higgins, *op. cit.*, 79, nuestra traducción.

6 *Idem.*

7 Seiji Isotani y Ig Ibert Bittencourt, *Dados Abertos Conectados*.

8 National Science Foundation, “US NSF - NSF 07-28, Cyberinfrastructure Vision...”.



Por lo tanto, los metadatos son estructuras intrínsecas al tratamiento de la información en los medios digitales. Pomerantz<sup>9</sup> conceptualiza que los metadatos son “[...] un medio por el cual la complejidad de un objeto se representa en una forma más simple”. Riley,<sup>10</sup> en la publicación más reciente de la National Information Standards Organization (NISO) divide los metadatos en cinco tipos según sus funciones: a) metadatos administrativos (incluye aquí a los metadatos técnicos, que se refieren a la información sobre archivos digitales necesarios para decodificarlos y renderizarlos); b) metadatos descriptivos: tienen la función de representar o encontrar un recurso; c) metadatos de preservación: se refieren a la gestión a largo plazo y futuras migraciones o emulación de archivos digitales; d) metadatos de derechos: se enumeran como una licencia de los derechos de propiedad intelectual asociados con el contenido; por último, el documento establece una categoría final de metadatos, e) lenguaje de marcado: son lenguajes asignados junto a metadatos y contenido en *forma de etiquetas* que se insertan para marcar características notables, palabras de señalización con información semántica o elementos estructurales.

Para ser más efectivos en su rol, los metadatos asumen una estructura estandarizada, componiendo un patrón de metadatos. Estos estándares admiten una serie de funciones definidas, que le permiten especificar elementos que hacen posible la suplantación y describen la característica de información que permite a los usuarios identificar, encontrar y recuperar, al tiempo que facilita la administración de contenido y acceso.<sup>11</sup>

Zeng y Qin<sup>12</sup> plantean que, al utilizar la misma estructura de estandarización de metadatos, se agiliza la recuperación automática de información y, por lo tanto, esto promueve la consistencia

---

9 Jeffrey Pomerantz, *Metadata*, 11. Nuestra traducción.

10 Jenn Riley, “Understanding Metadata: what is metadata...”, 6. Nuestra traducción.

11 Sarah Higgins, *op. cit.*

12 Marcia Lei Zeng y Jian Qin, *Metadata*.

de las bases de datos, lo que facilita el intercambio de información entre ellas, es decir, la interoperabilidad y el intercambio de datos.

En este sentido, surgieron los primeros esfuerzos para asegurar la preservación digital del material informativo. Sin embargo, esta preservación, al señalar sólo la integridad y autenticidad de los datos, implicaba poco acceso. Con el tiempo, el enfoque cambió para que el material digital pudiera conservarse sin comprometer su acceso, y luego se administró a lo largo de su ciclo de vida.<sup>13</sup>

Oliver y Harvey<sup>14</sup> explican que la curaduría digital es una nueva disciplina centrada en las actividades de preservación digital, pero que tiene como objetivo principal garantizar su disponibilidad, uso y reutilización de materiales digitales a lo largo de su ciclo de vida. De esta manera, se entiende que el proceso de agregar valor en la curaduría digital está destinado a representar datos basados en metadatos del contexto al que pertenece o está relacionado un recurso informativo. Este proceso está destinado a garantizar la validez y utilidad de los recursos, es decir, la autenticidad, fiabilidad, integridad y usabilidad de los datos para adaptarse a cualquiera que sea su finalidad informativa a lo largo del tiempo.<sup>15</sup>

En vista de esto, es importante recordar que incluso la expresión curaduría digital es de uso reciente. Glushko<sup>16</sup> puntea que la curaduría ocurre en todos los sistemas de organización, desde la reorganización de una estantería en un hogar para acomodar nuevos libros hasta a escala institucional; cuando un museo diseña una nueva exposición y, también, cuando las personas seleccionan y organizan fotos para cargar en redes sociales o quien mantiene un sitio web para difundir contenidos comerciales. En este sentido, Glushko<sup>17</sup> adhiere que cualquier persona puede tomar decisiones y emplear tecnología que mantenga el contenido con calidad

---

13 Sarah Higgins, *op. cit.*

14 Gillian Oliver y Ross Harvey, *Digital curation*.

15 Maria Lúcia Triques *et al.*, "Aspectos da representação", 15.

16 *The Discipline of Organizing*.

17 *Ibid.*

y carácter a lo largo del tiempo. Así, ejercer la curaduría, a juicio de Glushko,<sup>18</sup> no es más que definir “políticas claras para recopilar recursos y mantenerlos a lo largo del tiempo, lo que permite a las personas y los procesos automatizados garantizar que los datos de los recursos y sus representaciones sean confiables, precisos, completos, consistentes y no redundantes”.

La curaduría digital aborda toda la gama de procesos aplicados a los objetos digitales a lo largo de su ciclo de vida, comenzando incluso antes de que se creen, a través de la planificación y el establecimiento de estándares para la recopilación de datos, lo que apunta así a la mejor condición posible para garantizar que puedan mantenerse y utilizarse en el futuro.<sup>19</sup> Así, la curaduría digital implica todas las acciones de gestión de datos, desde su planificación, desde la recogida hasta la disponibilidad y preservación para su uso futuro.

Por lo tanto, este concepto cubre una amplia gama de funciones, y emplea procedimientos que buscan la gestión de cualquier material que tenga valor informativo para garantizar su disponibilidad y acceso. Por ello, cuenta con una gran capacidad de gestión de datos de colecciones digitales, ya que ofrece acciones que aseguran que dichos datos sigan contribuyendo a la generación y transmisión de conocimiento. Para Yamaoka,<sup>20</sup> un programa de curaduría digital tiene como objetivo salvaguardar los objetos digitales, permitiendo el acceso y la reutilización a lo largo de su ciclo de vida. Esto permite: a) mantener el documento intacto y accesible, siempre que tenga valor legal (evidencia); b) extraer nuevos conocimientos (valor informativo y de investigación); c) preservar la memoria de la sociedad (valor histórico); d) evitar retrabajos de recreación de los datos ya producidos anteriormente.

---

18 Robert K. Glushko, *The Discipline of Organizing*, 194. Nuestra traducción.

19 Gillian Oliver y Ross Harvey, *op. cit.*

20 “Ontología para mapeamento da dependência tecnológica”, 5.

Frente a esto, las obras de Oliveira y Lóscio<sup>21</sup> argumentan que los modelos de datos tradicionales indicados por la curaduría digital se pueden adaptar a los metadatos, sin embargo, estos modelos curados tuvieron un impacto limitado en el mantenimiento de los datos. Este argumento también lo hacen los autores Qin, Ball y Greenberg;<sup>22</sup> Kouper *et al.*;<sup>23</sup> Mauthner y Parry;<sup>24</sup> y Dallas,<sup>25</sup> quienes complementan en sus trabajos la necesidad de crear formas de reducir este impacto para la garantía de calidad y gestión de metadatos, sin abandonar las sistematizaciones y la calidad que ofrecen las acciones de la curaduría digital.

Por lo tanto, la curación de metadatos debe llevarse a cabo con base en los principios de la curación digital y las prácticas ágiles de desarrollo de software.<sup>26</sup> A través de estas acciones, la curación de metadatos puede contribuir a abordar la falta de centralización y organización de los responsables, así como la falta de actores dedicados a curar los metadatos.

## MATERIALES Y MÉTODOS

Planteado ya el propósito, esta investigación se configura como exploratoria, cualitativa y de carácter teórico-aplicado, y aborda como tema principal y como universo de investigación el uso de la curación de metadatos para recursos educativos digitales.

Para la contextualización teórica, se utilizarán fuentes bibliográficas como base para los resultados y, por esta razón, la investigación es de carácter bibliográfico. Además, la investigación tiene un carácter exploratorio, que dará una base teórica para la

---

21 Marcelo Iuri de Sousa, "A metadata curation framework..."; Marcelo Iuri de Sousa Oliveira y Bernardette Lóscio, "Ecosistemas de Dados na Web..."; Oliveira y Lóscio, "Louvre: A framework...".

22 "Functional and architectural requirements for metadata: Supporting..."

23 "Collaborate, Automate, Prepare, Prioritize: Creating..."

24 "Open access digital data sharing: Principles, policies and practices".

25 «Digital curation beyond the "wild frontier": A pragmatic approach».

26 Marcelo Iuri de Sousa, "A metadata curation...".

construcción de un conocimiento sobre la curación de metadatos y su práctica.

Se consultarán fuentes primarias, secundarias y terciarias en las bases de datos de Ciencias de la Computación y Ciencias de la Información, tales como: Biblioteca Digital Brasileña de Tesis y Disertaciones (BDTD), OASIS.br, Repositorio Científico de Acceso Abierto de Portugal (RCAAP), Base de Datos de Referencia de Artículos de Revistas en Ciencias de la Información (BRAPCI), Biblioteca Electrónica Científica en Línea SciELO, Brazilian Digital Computer Library (BDBComp), Digital Bibliography & Library Project (DBLP), ACM Digital Library, Scopus, Web of Science y EBSCO.

Para contemplar las publicaciones relevantes para el tema, las estrategias de búsqueda que se utilizarán inicialmente consistirán en el uso de términos como '*metadata curation*', '*digital curation for digital educational resource platforms*', '*data lifecycle*', en los siguientes idiomas: portugués, inglés y español. Sin embargo, es importante destacar que el término se ha ido consolidando durante los últimos años, y puede haber variables terminológicas que se destacarán durante la encuesta bibliográfica. El periodo de búsqueda estimado incluirá publicaciones a partir del año 2000, lapso de tiempo en el que se observa el inicio del crecimiento de los debates sobre el tema.

El estudio se desarrollará con el fin de aplicar en la Plataforma Integrada de Recursos Educativos Digitales MEC RED. Esta plataforma se establece en un espacio dinámico, con herramientas de búsqueda eficientes y fáciles de manipular para los profesionales de la educación, y combina un modelo de entorno de contenido digital con un modelo de red social, lo que hace que la navegación sea intuitiva y rápida.<sup>27</sup> La plataforma MEC RED ayuda a sus usuarios a encontrar recursos digitales en los portales del Ministerio de Educación y otros socios.

El software utilizado por la plataforma es Dspace, un paquete para la gestión de repositorios de código abierto para la

---

27 Brasil, Ministério da Educação, *op. cit.*

publicación de varios tipos de datos. En Dspace, los metadatos se representan en forma de propiedad-valor, con mecanismos de asignación de metadatos, a través de reglas XSLT, que permiten la generación de representaciones de metadatos en estructuras jerárquicas XML a partir de representaciones planas de valor de propiedad. Por esta razón, el software no permite la representación de elementos con estructuras cuyos contenidos también se desplieguen en nuevos elementos.

## RESULTADOS

De acuerdo con el propósito del trabajo y la contextualización teórica presentada, se muestra qué lematización de los pasos para la definición del Perfil de Aplicación de Metadatos puede ayudar en la formulación e identificación de datos para la complejión y definición de la curación de metadatos. Esto resulta de la propia sistematización definida para la construcción del perfil, que tiene en cuenta en la encuesta de requerimientos funcionales las necesidades informativas de los usuarios, las características representacionales del objeto y por describir, además de las demandas institucionales.

En este sentido, se propuso una Aplicación de Perfil y Metadatos para el ámbito de la red MEC, denominada PAM-MECRED. El perfil se basa en las directrices de la *Dublin Core Metadata Initiative* (2008) y fue creado por la necesidad de proporcionar metadatos que permitan la ubicación, identificación, acceso, almacenamiento y preservación de red en la plataforma para ayudar en la toma de decisiones para seleccionar los materiales que serán adquiridos y utilizados por la RED de educación pública en Brasil. Así, la construcción de políticas públicas exige la construcción de un edicto para la selección de este material y el aviso se basó en los pasos 1) los requisitos funcionales; 2) el modelo de dominio, y 3) la descripción conjunta de perfiles y pautas de uso, al concepto de evidencia, para la toma de decisiones.

Con ello, el perfil también proporcionará la capacidad de interoperar con sistemas nacionales e internacionales, además de la propia plataforma MEC RED, que almacenará todos los recursos educativos desarrollados y adquiridos por MEC. Con este fin, el MEC RED promueve una formulación necesaria para la concentración y recolección de datos para la construcción de evidencias y toma de decisiones para la evaluación y reevaluación de materiales de acuerdo con los Avisos PNLD.

Para la definición de requerimientos funcionales, PAM-MECRED abarca la descripción de los REDs a ser archivados y difundidos por la plataforma, encuentra las tipologías de: materiales didácticos, aprendizaje, investigación, pedagógica y gestión escolar en soporte digital, tales como e-books, folletos, guías, aplicaciones, softwares, plataformas, juegos electrónicos y otros tipos de contenidos digitales.

El acceso a la plataforma es abierto y tiene como público principal a profesores y estudiantes de escuela primaria, educación juvenil y de adultos y escuelas públicas de secundaria, y usuarios que tengan interés en los REDs. Entre las tareas del usuario, se considera necesario que encuentre uno o más REDs de interés, comprenda claramente la tipología de los REDs encontrados, los seleccione según las necesidades informativas; asimismo, se espera que el contenido de cada descripción y cada RED, además, pueda explorar y descubrir otros recursos digitales que estén relacionados con su necesidad informativa. Otro tipo de usuario son los editores que deben enviar los REDs en la plataforma para su selección y crear los registros de metadatos para la evaluación de los gestores. Los gerentes son responsables de administrar, insertar y administrar colecciones y elementos en la plataforma. Respecto de los REDs, la descripción de cada recurso tendrá metadatos obligatorios, dependiendo de la necesidad del recurso informativo, algunos elementos pueden ser “repetibles” y no “repetibles”, todos los recursos que se hayan publicado en acceso abierto deben estar disponibles para su descarga y cumplir con las políticas del repositorio. La plataforma debe promover la interoperabilidad con otros repositorios, facilitar la identificación de acceso abierto y almacenar documentos.

### ***Metadatos sociales: iniciativas...***

El modelo de dominio para PAM-MECRED prevé la responsabilidad de la apelación por parte de los editores y, sobre la base del objetivo de adquisición de estos recursos, forma parte del PNLD. Para el modelo, es necesario que los recursos estén soportados en licencias abiertas y caracterizados y localizados por las competencias y habilidades enumeradas por el BNCC y por asignaturas que se agrupan, denominadas como temas integradores.

Después de definir el modelo de dominio, se establecieron metadatos para describir las entidades. Por lo tanto, la definición de los términos de metadatos para el perfil se basó en el *Conjunto de Elementos de Metadatos Básicos de Dublín*, *Schema.org* y la *Iniciativa de Metadatos de Recursos de Aprendizaje*, mantenida por la *Iniciativa de Metadatos Básicos de Dublín* (DCMI) y algunos metadatos se crearon para satisfacer las necesidades de la plataforma MEC RED.

En el Aviso PNLD 2021, los metadatos administrados para la descripción de los REDs fueron: 1) título; 2) descripción; 3) palabras clave; 4) área de conocimiento o tema integrador; 5) Competencias BNCC; 6) Habilidades BNCC; 7) autor; 8) descriptivo; 9) Código alfanumérico BNCC y 10) tipo de licencia.

En este sentido, se realizó una comparación entre los metadatos ya utilizados por la plataforma para describir los REDs, las especificaciones contenidas en el Aviso PNLD 2021, los metadatos simples Dublin Core y los metadatos sugeridos, con el fin de proporcionar una representación más completa de los REDs.

A partir de esta comparación, PAM-MECRED estableció un conjunto mínimo necesario para describir los REDs. Se insertaron otros elementos basados en la encuesta de requisitos funcionales y que son fundamentales para ayudar en la toma de decisiones para la elección y selección de REDs por parte del MEC. La tabla 1 presenta los metadatos sugeridos con más detalle agregando sus categorías y descripciones. Presenta el nombre de cada metadato, cómo se presentará al usuario (etiqueta), las validaciones de los términos (URI), la definición de estos metadatos, el tipo de valor, la ocurrencia mínima y máxima, el valor (literal o no literal) y si es un metadato obligatorio o no.



Tabla 1. Descripción de PAM-MECRED

Nombre de metadatos	Presentación al usuario	Término URI	Definición	Tipo de valor	Ocurrencia mínima	Ocurrencia máxima	Valor (literal o no literal)	Obligatorio, especializado u opcional
dc.creator	Creador	http://purl.org/dc/elements/1.1/creator	Entidad(es) como el principal responsable de la creación del recurso	Propiedad	1	N	literal	Obligatorio
dc.contributor	Colaborador	http://purl.org/dc/elements/1.1/contributor	Entidad(es) responsable(s) de hacer contribuciones al recurso	Propiedad	0	N	literal	Obligatorio
dc.cobertura	Cubrir	http://purl.org/dc/elements/1.1/coverage	El tema espacial o temporal del recurso, la aplicabilidad espacial del recurso o la jurisdicción bajo la cual el recurso es relevante	Propiedad	0	N	literal	Especializado

**Metadatos sociales: iniciativas...**

dc:terms:emitido	Fecha de publicación	http://purl.org/dc/terms/issued	Punto o periodo asociado al ciclo de vida del recurso. Pueden tener varias fechas, como publicación, edición, versionado, etcétera	Propiedad	1	1	xsd:año	Obligatorio
dc:description	Descripción	http://purl.org/dc/elements/1.1/description	Información general sobre el recurso	Propiedad	1	N	literal	Obligatorio
dc:format	Formato	http://purl.org/dc/elements/1.1/format	El formato del archivo, el medio físico o las dimensiones del recurso	Propiedad	0	N	literal	Especializado
dc:identifier	Identificador	http://purl.org/dc/elements/1.1/identifier	Una referencia inequívoca al recurso en un contexto determinado	Propiedad	0	N	literal	Obligatorio
dc:language	Idioma	http://purl.org/dc/elements/1.1/language	Idioma del recurso	Propiedad	1	1	literal	Obligatorio
dc:publisher	Editor	http://purl.org/dc/elements/1.1/publisher	Entidad responsable de poner a disposición el recurso	Propiedad	1	N	literal	Opcional

dc.relation	Relación	http://purl.org/dc/elements/1.1/relation	Recurso(s) relacionado(s) de alguna manera con el recurso descrito	Propiedad	0	N	literal	Opcional
dc.source	Fuente	http://purl.org/dc/elements/1.1/source	Característica(s) relacionadas de las que se deriva el recurso descrito total o parcialmente	Propiedad	0	1	literal	Especializado
dc.subject	Palabra clave	http://purl.org/dc/elements/1.1/subject	Términos que ayudan a describir los temas cubiertos por el recurso y ayudan en el proceso de búsqueda/descubrimiento de recursos	Propiedad	1	N	literal	Obligatorio
dc.title	Título	http://purl.org/dc/elements/1.1/title	Nombre dado al recurso. Campo para la definición del título	Propiedad	1	1	literal	Obligatorio
dc.type	Tipo de recurso	http://purl.org/dc/elements/1.1/type	La naturaleza o el género del recurso	Propiedad	1	1	literal	Obligatorio

**Metadatos sociales: iniciativas...**

dc.rights	Licencia de uso	<a href="http://purl.org/dc/elements/1.1/rights">http://purl.org/dc/elements/1.1/rights</a>	Declaraciones sobre varios derechos de propiedad asociados con la apelación, incluidos los derechos de propiedad intelectual	Propiedad	1	1	literal	Obligatorio
Imri.typicalAge-Range	Nivel de educación	<a href="http://purl.org/dcx/irmi-terms/typicalAgeRange">http://purl.org/dcx/irmi-terms/typicalAgeRange</a>	Nivel de educación al que está destinado el recurso	Propiedad	1	N	literal	Obligatorio
Imri.educational-Framework	Componente curricular	<a href="http://purl.org/dcx/irmi-terms/educationalFramework">http://purl.org/dcx/irmi-terms/educationalFramework</a>	Componente curricular que cubre el recurso	Propiedad	1	N	literal	Obligatorio
schema.access-Mode	Accesibilidad	<a href="https://github.com/schemaorg/schemaorg/issues/1100">https://github.com/schemaorg/schemaorg/issues/1100</a>	Mecanismos de accesibilidad disponibles para el recurso	Propiedad	0	N	literal	Notas
Imri.educational-Alignment	Área de conocimiento / Tema Integrador	<a href="http://purl.org/dcx/irmi-terms/educationalAlignment">http://purl.org/dcx/irmi-terms/educationalAlignment</a>	Área de conocimiento o BNCC Integrator Tema abordado por el recurso	Propiedad	0	N	literal	Notas

Lmri.enseña	Competencias BNCC	<a href="http://purl.org/dc/lmri-terms/teaches">http://purl.org/dc/lmri-terms/teaches</a>	Competencias del BNCC contempladas en el recurso	Propiedad	0	N	literal	Notas
schema.skills	Habilidades BNCC	<a href="https://schema.org/skills">https://schema.org/skills</a>	Habilidades BNCC contempladas por el recurso	Propiedad	0	N	literal	Notas
dc.identifier:BNCC	Código alfanumérico BNCC	<a href="http://purl.org/dc/elements/1.1/identifier">http://purl.org/dc/elements/1.1/identifier</a> BNCC	Códigos alfanuméricos de habilidades y competencias BNCC contempladas por el recurso	Propiedad	0	N	literal	Notas

Fuente: Elaboración propia.

### ***Metadatos sociales: iniciativas...***

El nombre de cada metadato consiste en los términos ya indicados por *el Dublin Core Metadata Element Set*, Schema.org y la *Learning Resource Metadata Initiative*. El nombre de las presentaciones de los atributos (etiqueta) fue elaborado de manera objetiva para que todos los usuarios puedan describir adecuadamente el RED. La validación de términos (URI) es la especificación actualizada y definitiva de todos los términos de metadatos tal como se definen. La columna de definición comprende la descripción de cada atributo. Las columnas de tipo de valor informan de la propiedad de los atributos. La ocurrencia mínima y máxima demuestra la cardinalidad de cada atributo. También se muestra el valor (literal o no literal) y si se trata de metadatos obligatorios, especializados, opcionales o de anotación.

En función de los metadatos definidos, puede ayudar a qué RED se adquirirán, cuáles son los más apropiados para ciertos niveles escolares, qué características de accesibilidad se ofrecen, actualización de RED y asignatura. Así, toda esta información se utilizará como evidencia para la toma de decisiones de dónde puede invertir el MEC y qué REDS deben seleccionar, establecer cuántos REDS ha aportado cada editor, los tipos de REDS, además de establecer políticas centradas en el partir de las demandas, en qué áreas de conocimiento hay más inversión, en qué área hay déficit de REDS, y qué habilidades y competencias buscan contemplar los REDS.

### **CONSIDERACIONES FINALES**

Durante el trabajo se pudo identificar que la descripción adecuada puede facilitar los procesos de clasificación, evaluación y recuperación. De esta manera, se pueden organizar repositorios y plataformas como MEC RED, y permitir a los usuarios acceder a estas funcionalidades con mayor facilidad, así como en contextos como PNLD los gestores y evaluadores entienden lo que propone el recurso evaluado o puesto a disposición antes de interactuar con su propio contenido.

En este sentido, el artículo trajo fundamentos que validan el objetivo de la investigación, el PAM-MECRED, basado en *las Guidelines of Dublin Core Metadata Initiative* (2008), define los elementos esenciales y mínimamente necesarios para la descripción adecuada de un RED, es decir, sin ellos, la calidad de la descripción se vería comprometida. Los impuestos a deseables complementan la descripción esencial con atributos que, dependiendo de la situación, pueden o no existir. Los atributos especializados, por otro lado, aportan información que no es absolutamente necesaria, sino que profundizan en la descripción del recurso, y hacen que la representación sea más completa e informativa. Finalmente, las anotaciones son atributos auxiliares que complementan otros ítems del esquema, por ejemplo, los atributos referidos al BNCC complementan los atributos obligatorios del componente curricular y en BNCC.

Como un trabajo futuro, pon había dado la implementación de PAM-MECRED y su sintaxis incorporada en la plataforma MEC RED. Otra rama posible es la realización de un estudio empírico con especialistas en metadatos y con evaluadores de RED, con el fin de investigar sus percepciones con respecto a las contribuciones de contenido de metadatos propuestas en esta investigación en comparación con el punto de partida encontrado originalmente en el aviso de PNL D. Finalmente, puntuamos las reflexiones y el desarrollo de otros enfoques en los metadatos de relación o partición, lo que posibilita la conexión entre uno o más REDs.

## AGRADECIMIENTOS

Agradezco la acogida y supervisión del Prof. Dr. Seiji Isotani durante la pasantía de posdoctorado en la Universidad de São Paulo (USP) y al grupo de investigación “Datos y Metadatos” porque las discusiones llevadas a cabo ahí fueron sumamente beneficiosas para el desarrollo de este trabajo.

## REFERENCIAS BIBLIOGRÁFICAS

- Brasil. *Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União*. 1996. [http://www.planalto.gov.br/ccivil\\_03/leis/l9394.html](http://www.planalto.gov.br/ccivil_03/leis/l9394.html).
- . *Dispõe sobre o Programa Nacional do Livro e do Material Didático. Diário Oficial da União*. 2017. [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2017/decreto/D9099.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/decreto/D9099.htm).
- . Ministério da Educação. *Define critérios e procedimentos para a produção, recepção, avaliação e distribuição de recursos educacionais abertos ou gratuitos voltados para a educação básica em programas e plataformas oficiais do Ministério da Educação. Diário Oficial da União*. 2018. [https://www.in.gov.br/materia/-/asset\\_publisher/Kujrw0TZC2Mb/content/id/14729210/do1-2018-05-17-portaria-n-451-de-16-de-maio-de-2018-14729206](https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/14729210/do1-2018-05-17-portaria-n-451-de-16-de-maio-de-2018-14729206).
- . “Plataforma integrada MEC RED”. Plataforma Integrada de Recursos Educacionais Digitais. 2021. <https://plataformaintegrada.mec.gov.br/>.
- Dallas, Costis. «Digital curation beyond the “wild frontier”: A pragmatic approach». *Archival Science* 16 (4): 421–57.
- Dublin Core Metadata Initiative. “The Singapore framework for Dublin Core application Profiles”. 2008.
- Glushko, Robert J. *The Discipline of Organizing: Core Concepts Edition*. Sebastopol, EUA: O’Reilly Media. 2014. <https://www.overdrive.com/search?q=70B2FB07-B1BE-458C-8BC5-5E7FBBEF946E>.
- Higgins, Sarah. “Using Metadata Standards”. Digital Curation Centre. 2007. <https://www.dcc.ac.uk/guidance/briefing-papers/standards-watch-papers/using-metadata-standards>.



- Isotani, Seiji e Ig Ibert Bittencourt. *Dados Abertos Conectados: em busca da Web do Conhecimento*. São Paulo: Novatec Editora. 2015.
- Kouper, Inna, Stacy R. Konkiel, Jennifer A. Liss y Juliet L. Hardesty. “Collaborate, Automate, Prepare, Prioritize: Creating metadata for legacy research data”. En *International Conference on Dublin Core and Metadata Applications*, 41–46. 2013.
- Mauthner, Natasha Susan y Odette Parry. “Open access digital data sharing: Principles, policies and practices”. *Social Epistemology* 27 (1): 47–67.
- National Science Foundation. “US NSF - NSF 07-28, Cyberinfrastructure Vision for 21st Century Discovery”. 2007. <https://www.nsf.gov/pubs/2007/nsf0728/index.jsp>.
- Oliver, Gillian y Ross Harvey. *Digital curation*. Chicago: American Library Association. 2016.
- Oliveira, Marcelo Iury de Sousa. “A metadata curation framework for data ecosystems”. Tesis de doctorado en Ciencias de la Computación. Pernambuco: Universidade Federal de Pernambuco (UFPE). 2019. <https://repositorio.ufpe.br/handle/123456789/33910>.
- Oliveira, Marcelo Iury y Bernardette Farias Lóscio. “Ecosistemas de Dados na Web: da teoria aos desafios”. *Sociedade Brasileira de Computação*, 68-92. 2019. DOI: <https://doi.org/10.5753/sbc.6251.1.3>
- . “Louvre: A Framework for Metadata Curation in Data Ecosystem”. Em *Proceedings of the XV Brazilian Symposium on Information Systems*, 1–8. 2019. <https://dl.acm.org/doi/abs/10.1145/3330204.3330248>.
- Pomerantz, Jeffrey. *Metadata*. The MIT Press essential knowledge series. Cambridge, Massachusetts; London, England: The MIT Press. 2015.

***Metadatos sociales: iniciativas...***

- Qin, Jian, Alex Ball y Jane Greenberg. "Functional and architectural requirements for metadata: Supporting discovery and management of scientific data". En *International Conference on Dublin Core and Metadata Applications*, 62–71. 2012.
- Riley, Jenn. "Understanding Metadata: what is metadata, and what is it for?". National Information Standards Organization (NISO). NISO Primer Series. 2017. [http://www.niso.org/apps/group\\_public/download.php/17446/Understanding%20Metadata.pdf](http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf).
- Triques, Maria Lúgia, Ana Carolina Simionato Arakaki y Fabiano Ferreira de Castro. "Aspectos da representação da informação na curadoria digital". *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação* 25 (mayo): 1–21. <https://doi.org/10.5007/1518-2924.2020.e69898>.
- Yamaoka, Eloi Juniti. "Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital". *AtoZ: novas práticas em informação e conhecimento* 1 (2): 65. <https://doi.org/10.5380/atoz.v1i2.41313>.
- Zeng, Marcia Lei y Jian Qin. *Metadata*. New York: Neal-Schuman Publishers. 2008.
- . *Metadata*. Second edition, UK edition. London: FP, Facet Publishing. 2016.

# Metadatos bibliográficos y metadatos sociales: conexiones en entornos de datos vinculados

FABIANO FERREIRA DE CASTRO  
*Universidad Federal de São Carlos, Brasil*

## INTRODUCCIÓN

Las bibliotecas tienen una larga historia de estandarización y colaboración. Panizzi y Cutter ofrecieron los primeros enfoques para sistematizar la descripción en el siglo XIX. Sin embargo, las bibliotecas pronto se dieron cuenta de la necesidad de una mayor colaboración, no sólo para que los usuarios pudieran esperar un enfoque estandarizado para encontrar material en las existencias institucionales, sino para que pudieran compartir descripciones de recursos para ahorrar tiempo y dinero y reducir la duplicación de esfuerzos.

El desarrollo de reglas, esquemas de descripción y códigos de catalogación para la descripción y acceso a los recursos de información tiene eco, sobre todo, en la formalización e institucionalización de los principios que guían la descripción de los recursos, denominados Principios de París, en 1961, los cuales establecen la base de la cooperación internacional.

En este sentido, la idea de colaboración sucede con el enfoque compartido para la creación de datos bibliográficos, desarrollado por la Biblioteca del Congreso, en 1901, con la creación de la hoja

de catálogo de 7,5 X 12,5 cm, que reflexiona sobre el escenario de colaboración en red al mismo tiempo, como resultado directo de la estandarización internacional en décadas pasadas.

Otro hito histórico que consolida aún más los ideales de cooperación y estandarización en el mundo de las bibliotecas es el desarrollo del formato Machine Readable Cataloging (MARC), en la década de 1960, por Henriette Avram, que, según Zavalin e Miksa,<sup>1</sup> permitió codificar y transmitir descripciones automáticas y compartidas de los recursos de información.

De esta forma, se identifica la idea de compartir y colaborar, especialmente en la práctica profesional del bibliotecario, en la creación de metadatos por parte de los usuarios internos de los catálogos institucionales, es decir, por profesionales especialistas. Ante este escenario, la pregunta es: ¿cuál es el impacto de extender los metadatos bibliográficos con curación social y hacerlos visibles como datos vinculados?

El objetivo es verificar en la literatura científica disponible las investigaciones que abordan los metadatos, los metadatos sociales y sus relaciones con los datos enlazados. Por lo tanto, la metodología adoptada se caracteriza por un mapeo sistemático de la literatura, lo que permitió la identificación y categorización de los resultados a partir de las variables: a) Ventajas; b) Desventajas, y c) Iniciativas/Proyectos/Aplicaciones.

## CONSIDERACIONES SOBRE VENTAJAS Y DESVENTAJAS EN LA CONSTRUCCIÓN DE METADATOS SOCIALES

Con respecto a las ventajas de la categoría de curación social con metadatos visibles como datos vinculados, Alemu<sup>2</sup> discute que los metadatos contruidos socialmente reflejan mejor la terminología

---

1 Vyachoslav Zavalin, I., y Shwane D. Miksa, "Collecting and evaluating large volumes of bibliographic metadata...".

2 Getaneh Alemu, "A theory of metadata enriching and filtering: challenges and opportunities to implementation".

del usuario, mejoran la capacidad de localización, mejoran la casualidad, identifican el espacio, el tiempo y el contexto y hacen uso de vocabularios emergentes.

La investigación de Manzo *et al.*<sup>3</sup> señala que las contribuciones de los usuarios son un beneficio neto, mientras que los metadatos construidos socialmente, cuando se usan junto con los metadatos bibliográficos, aumentan la capacidad de búsqueda, corrigen errores de búsqueda evitables y, en general, son precisos. Además, los resultados del estudio muestran que, dadas las mismas condiciones de etiquetado, los bibliotecarios y los no bibliotecarios producen una distribución similar de metadatos potencialmente útiles.

Sobre las desventajas, Alemu<sup>4</sup> discute que los metadatos producidos por las bibliotecas no representan la cosmovisión diversa de sus usuarios. “Sus términos y vocabularios controlados son inadecuados: están desactualizados, carecen de contexto y requieren que los expertos en metadatos asuman la tarea de ‘descubrir’ lo que creen que es importante”. Además, los bibliotecarios y el personal bibliotecario tienen experiencia en el oficio de actividades especializadas como catalogación, clasificación e indexación, y no son necesariamente expertos en describir semánticamente el contenido que surge de fuentes de conocimiento heterogéneas. Todas estas críticas apuntan a la necesidad de facilitar una contribución más amplia a la creación de metadatos, especialmente con una participación comunitaria más efectiva.

Si tenemos en cuenta las notas de Alemu,<sup>5</sup> desde el punto de vista de los usuarios expertos, los bibliotecarios pueden estar convencidos de que se necesitan metadatos más abiertos, pero aún están limitados por aspectos prácticos. Los sistemas necesitan datos consistentes para funcionar bien, los datos deben interoperar con otros datos, las personas que describen cosas en todo el mundo requieren instrucciones sobre cómo hacerlo, pero los usuarios

---

3 Christina Manzo *et al.*, “By the people, for the people: assessing the value of crowdsourced, user-generated metadata”.

4 Getaneh Alemu, *op. cit.*

5 *Idem.*

prefieren algún tipo de presentación lógica de los recursos de una biblioteca y las bibliotecas necesitan compartir descripciones para ahorrar costos.

Si las bibliotecas consideran que el propósito de los metadatos es principalmente de naturaleza práctica, criticarían las contribuciones abiertas y sociales del público por motivos prácticos. Las bibliotecas deben cuestionar la veracidad de los metadatos enviados a través de medios abiertos y sociales: “cuando estos metadatos se combinan con metadatos creados por la biblioteca, se producen consecuencias no deseadas, como información errónea y metadatos en conflicto”.<sup>6</sup>

Otras críticas a los metadatos proporcionados por los usuarios señalan que “carecen de estructura y confiabilidad”, contienen “ambigüedad de significado” y, a menudo, presentan más “ruido que señal”.<sup>7</sup> Esta crítica lleva a pensar que la falta de autoridad final para la creación de metadatos puede generar datos superficiales, caóticos, falsos y, en última instancia, poco confiables.<sup>8</sup> Además, ¿cuánto valor agregan realmente estas contribuciones? Algunos estudios de metadatos construidos socialmente concluyeron que el marcado (etiquetado) sólo agregó “semántica de bajo nivel” para corregir errores y agregar detalles narrativos.<sup>9</sup> Desde este punto de vista, se concluye que el etiquetado tiene más que ver con el compromiso que con una contribución sustancial de manera efectiva.

El trabajo de Seeman y Dean,<sup>10</sup> sin embargo, advertiría contra el reemplazo del vocabulario de acceso controlado por tema con búsqueda de palabras clave y folksonomía; los autores encontraron que aproximadamente el 27% de las búsquedas de palabras clave

---

6 Getaneh Alemu y Brett Stevens, *An emergent theory of digital library metadata: enrich then filter*, 313.

7 *Idem.*

8 *Idem.*

9 Getaneh Alemu *et al.*, “Toward an emerging principle of linking socially-constructed metadata”.

10 Dean Seeman y Heather Dean, “Open social knowledge creation and library and archival metadata”.

en el catálogo de una biblioteca fallarían si no se realizaran mediante vocabularios controlados por materias. Esto trae a colación la noción de que los aspectos prácticos que proporciona el entorno bibliotecario estandarizado no son sólo para la facilidad de aplicación por parte del personal de la biblioteca o para el uso de máquinas. En última instancia, estos aspectos prácticos también ayudan a los usuarios a encontrar lo que buscan de manera más confiable y eficiente.

Pero ¿es posible reducir los metadatos de la biblioteca a únicamente aspectos prácticos? Algunos argumentan que los metadatos de las bibliotecas no sólo facilitan el descubrimiento y el conocimiento, sino que tienen un mayor efecto en el mundo más amplio de la organización del conocimiento y sus funciones sociales y culturales.<sup>11</sup> Uno necesita determinar cómo los metadatos reflejan nuestra interacción con el patrimonio cultural en un sentido más amplio e indirecto dentro de la sociedad.<sup>12</sup> Es decir, ¿cómo reflejan y dan forma a la realidad las descripciones de los recursos en las bibliotecas, además de actuar como marcadores sociales para encontrar elementos específicos?

Una forma clara de resolver la ‘tensión’ entre la práctica actual de los metadatos de las bibliotecas y la apertura es admitir que los registros de metadatos no son neutrales y no alcanzan a satisfacer las necesidades de una base de usuarios diversa. Claramente, los metadatos producidos por bibliotecas tienen limitaciones, pero los metadatos construidos socialmente a menudo carecen de integridad sintáctica y semántica. La biblioteca y sus usuarios claramente necesitan ambos. Para establecer esto, la biblioteca necesita reconocer sus limitaciones. La institución debe estar abierta a otros puntos de vista, permitir la participación de expertos en el dominio y adoptar un enfoque iterativo para que los metadatos se

---

11 Getaneh Alemu, “A theory of metadata enriching and filtering: challenges and opportunities to implementation”.

12 Dean Seeman y Dean, *op. cit.*

puedan cambiar en cualquier momento, desestabilizando la idea de un registro ‘perfecto’, ‘autoritario’ y ‘completo’.<sup>13</sup>

Asimismo, la posibilidad de enriquecer los metadatos bibliográficos con contenido generado por los usuarios, como clasificaciones, etiquetas y reseñas, por ejemplo, para facilitar la búsqueda de recursos conocidos, así como la investigación exploratoria, se presenta como una buena alternativa al intercambio, la apertura de metadatos en entornos informativos digitales contemporáneos.

#### COLABORACIÓN Y COMPARTIR EN ENTORNOS DIGITALES: DE LAS BIBLIOTECAS A LA WEB

Contribuir a la colaboración en el entorno de los metadatos bibliográficos suele requerir el acceso a una plataforma o servicio compartido, que normalmente se realiza a través de una biblioteca en lugar de una membresía individual.

Al mismo tiempo, las bibliotecas deben recordar que los metadatos casi siempre están mediados por máquinas, e incluso si los metadatos construidos socialmente son preferibles, su ‘deestructuración’ no permite que las computadoras los manejen y procesen de manera eficiente.

Las tecnologías web 2.0 han permitido comentar, etiquetar y agregar folksonomías en catálogos de bibliotecas y plataformas de bibliotecas digitales. Muchos otros proyectos digitales colaboran con transcripciones y otros metadatos relacionados con un objeto digital. Estas iniciativas permiten contribuciones abiertas a los metadatos autorizados producidos por las bibliotecas, pero ¿cuáles son las implicaciones?

Un enfoque que permite una multiplicidad de vistas usando la máquina son los datos vinculados. Este enfoque permite que cualquier persona haga una declaración legible por máquina sobre cualquier cosa, ya que desafía las ideas preconcebidas de que los

---

13 *Ibid.*



bibliotecarios son los últimos guardianes y permite que los datos bibliográficos participen en el mundo más amplio del consumo y la creación de metadatos.<sup>14</sup>

Los ejemplos de creación de datos sociales, como MusicBrainz y Wikidata, muestran que muchos usuarios pueden contribuir con rdas. datos estructurados, que se pueden usar como datos vinculados. De hecho, Wikidata tiene el beneficio adicional de ayudar a superar la complejidad y los gastos que antes eran prohibitivos, evitando la necesidad de crear estructuras de almacenamiento más colaborativas. En un sistema como Wikidata, cualquier persona con una cuenta de Wikimedia puede crear datos estructurados inmediatamente utilizables en un contexto de datos vinculados. Luego, la biblioteca puede optar por conectarse a esta creación de conocimiento abierto y social, lo que sucede a una escala mucho mayor que los intentos poco entusiastas de involucrar a más personas en la creación de metadatos a través de comentarios y etiquetas. La pregunta clave es si la biblioteca está lista para aceptar el caos desbocado de la creación de conocimiento social.<sup>15</sup>

Seeman y Dean<sup>16</sup> agregan además que un entorno en el que la biblioteca ‘pierde’ el control y la energía, pero se alinea más estrechamente con la forma en que los usuarios actuales crean y consumen datos, es una amenaza para la comunidad de práctica, pero probablemente inevitable. El hecho de que la biblioteca no tenga en cuenta esta realidad puede servir para marginarla, es decir, para ser excluida de una estructura más globalizada de intercambio de datos y colaboración.

En el escenario contemporáneo, las tecnologías emergentes como los teléfonos inteligentes, la internet inalámbrica, así como la web y los servicios móviles permiten a los usuarios crear, anotar y compartir datos sociales en la web a un ritmo cada vez mayor y acelerado. Estas tecnologías transformaron a los usuarios de consumidores de datos estáticos durante la década de 1990 (es decir,

---

14 Christina Manzo *et al.*, *op. cit.*

15 Dean Seeman y Dean., *op. cit.*

16 *Idem.*

accediendo a páginas web estáticas) a productores inteligentes y sensores proactivos de información durante la década de 2010 (es decir, produciendo blogs, publicando y anotando imágenes y videos, comentando tuits, publicando opiniones, etcétera), donde la mayor parte de la información compartida es multimedia y está asociada a otros datos. Sin embargo, alcanzar la siguiente etapa de la web, a saber, la denominada web inteligente o web semántica, que permite una colaboración significativa entre humanos y máquinas y entre máquinas dentro de un entorno informático ubicuo, requiere otro avance: permitir compartir y compartir la organización del conocimiento colectivo, en la combinación de todos los datos, información y metadatos conocidos relacionados con un dominio determinado.<sup>17</sup>

La comunidad debe trabajar en metadatos junto con editores y proveedores para abordar la catalogación desde la perspectiva del usuario y perfeccionar el conjunto de habilidades, así como producir metadatos de calidad. Los profesionales de la Biblioteconomía y las Ciencias de la Información no sólo deben ser usuarios de los estándares, sino participar activamente y contribuir en su desarrollo y mejora, para que los datos puedan conectarse de manera eficaz y eficiente de forma globalizada. Esta colaboración es necesaria no sólo para los bibliotecarios y los proveedores de metadatos, sino también para los usuarios en el desarrollo de un entorno de catalogación eficiente y para metadatos bibliográficos más utilizables.

Alemu y Stevens<sup>18</sup> mencionaron varias ventajas de hacer que el contenido generado por el usuario forme parte de los catálogos de las bibliotecas. Eso incluye:

i) mejorar la funcionalidad de los metadatos seleccionados profesionalmente al hacer que los objetos de información sean detectables;

---

17 Minale A. Abebe *et al.*, "Generic metadata representation framework for social-based event detection, description, and linkage".

18 Getaneh Alemu y Stevens, *op. cit.*

- ii) eliminar las limitaciones impuestas por los principios de suficiencia y necesidad sobre los metadatos seleccionados profesionalmente;
- iii) acercar a los usuarios a la biblioteca ‘involucrándolos de manera proactiva’ en la clasificación, el etiquetado y la revisión, etcétera, siempre que los usuarios también participen en la gestión y el control de las entradas de metadatos; y
- iv) la ‘sabiduría de la multitud’ resultante beneficiaría a todas las partes interesadas de este crecimiento masivo de metadatos seleccionados socialmente.

Sin embargo, esta combinación sólo se puede utilizar de manera óptima si podemos vincularla semántica y contextualmente a recursos internos y externos; se accede abiertamente a los metadatos resultantes, se comparten y se reutilizan; los usuarios pueden agregar fácilmente metadatos y son parte del control de calidad, lo que les permite informar la actividad de spam a los expertos en metadatos.<sup>19</sup>

Las redes sociales expandieron la biblioteca a espacios sociales al explotar etiquetas y nubes de etiquetas como herramientas de navegación y enriquecer las descripciones bibliográficas al integrar contenido generado por el usuario. Cierra la brecha de comunicación entre la biblioteca y sus usuarios, donde los usuarios participan activamente en la descripción, el descubrimiento y el acceso a los recursos.

De esta forma, las bibliotecas están adoptando el aspecto de colaboración social de la catalogación que tendrá lugar entre autores, administradores de repositorios, bibliotecas, consorcios de colecciones electrónicas, editores y proveedores. Los bibliotecarios deben mejorar sus habilidades de catalogación en línea con los avances en las tecnologías disponibles para exponer y hacer visibles sus metadatos bibliográficos como datos abiertos y vinculados (*Linked Open Data*).

---

19 *Idem.*

En el escenario actual, los profesionales bibliotecarios generan y utilizan los datos vinculados a la biblioteca. Los metadatos construidos socialmente actuarán como un valor agregado en la recuperación precisa de artefactos de conocimiento. La adición de metadatos socialmente construidos e impulsados por la comunidad (usuarios externos) a las estructuras de metadatos existentes, los vocabularios controlados y los sistemas de clasificación brindan una visión holística de estas estructuras a medida que agregan significado generado por la comunidad a las estructuras de metadatos seleccionadas profesionalmente.<sup>20</sup>

Alemu *et al.*,<sup>21</sup> a partir del análisis de veintiuna entrevistas en profundidad con profesionales de la Biblioteconomía y la Información, descubrieron cuatro principios de los metadatos: enriquecimiento, vinculación, apertura y filtrado de metadatos. Este análisis reveló que la ausencia de metadatos seleccionados socialmente beneficia poco el potencial de los datos abiertos y vinculados en las bibliotecas. Alemu *et al.*,<sup>22</sup> en su análisis, abogan por un enfoque de metadatos mixtos, en el que los metadatos sociales (etiquetas, calificaciones y reseñas) aumentan y/o califican los metadatos bibliográficos involucrando proactivamente a los usuarios y ofreciendo una plataforma de catalogación colaborativa social.

Los principios de los metadatos deben reconceptualizarse y los datos abiertos vinculados deben explorarse para abordar los desafíos de los metadatos bibliográficos existentes.

Por lo tanto, los esfuerzos actuales en datos abiertos vinculados deben considerar completamente los metadatos sociales. Los catálogos de las bibliotecas deben enriquecerse combinando metadatos profesionales y sociales, así como interconectados semántica y contextualmente con recursos de información internos y externos, para poder ser utilizados de manera óptima en diferentes escenarios de aplicación.

---

20 Minale A. Abebe *et al.*, *op. cit.*

21 Getaneh Alemu *et al.*, *op. cit.*

22 *Idem.*

Para aprovechar al máximo este vínculo, se debe reducir la duplicación de metadatos. Deben ser de acceso abierto para que puedan compartirse, reutilizarse y combinarse. Los metadatos enriquecidos deben filtrarse de acuerdo con los requisitos del usuario utilizando una interfaz flexible, personalizada, contextual y reconfigurable.<sup>23</sup>

Zavalin y Miksa<sup>24</sup> sugieren un ‘cambio de paradigma’ en el futuro de los metadatos, es decir, de simples a enriquecidos; desde metadatos desconectados, invisibles e inaccesibles hasta metadatos bien estructurados, comprensibles por máquina, interconectados, visibles y más visualizados; y desde una sola interfaz OPAC hasta interfaces de metadatos reconfigurables y adaptables.

Al involucrar a los usuarios en el proceso de selección de metadatos, el enfoque mixto (profesionales expertos y no expertos) aportará diversidad en los metadatos y hará que los recursos de información sean más fáciles de descubrir, utilizar y centrarse en el usuario, con la plataforma de datos abiertos más amplia y mejor respaldada y vinculada.

Por lo tanto, la fusión de metadatos socialmente seleccionados con metadatos profesionales basados en estándares es esencial desde el punto de vista del paradigma de catalogación centrado en el usuario, que tiene el potencial de ayudar en el descubrimiento y acceso a recursos y abrir nuevas oportunidades para los científicos de la información que trabajan con datos abiertos y vinculados, así como catalogadores que hacen la transición a la web de datos o la web semántica para hacer que sus metadatos sean visibles, reutilizables y conectados a otros recursos en la web.

---

23 Vyacheslav Zavalin y Miksa, *op. cit.*

24 *Idem.*

## INICIATIVAS, PROYECTOS Y APLICACIONES DE METADATOS CURADOS SOCIALMENTE COMO DATOS ABIERTOS Y VINCULADOS

LibraryThing for Libraries (LTFL) hace que el catálogo de una biblioteca sea más informativo e interactivo al mejorar el OPAC, al brindar acceso a metadatos profesionales y sociales y permitiéndoles buscar, navegar y descubrir los fondos de la biblioteca de una manera más atractiva. Es uno de los ejemplos prácticos de cómo enriquecer los catálogos de las bibliotecas con contenido generado por los usuarios. Esta tendencia de fusionar metadatos sociales y profesionales innova la catalogación bibliotecaria al disolver los límites entre la ‘esfera social’ y los recursos bibliotecarios.

El papel potencial de los metadatos seleccionados socialmente en la descripción, el descubrimiento y el acceso a los recursos también es evidente en la encuesta de la Social Book Search en el marco a Initiative for XML Retrieval (INEX), donde tanto los metadatos bibliográficos seleccionados profesionalmente como los metadatos generados socialmente por el usuario se explotan para recuperación y recomendación, para apoyar tanto la búsqueda de recursos conocidos como la búsqueda exploratoria. Al experimentar con conjuntos de datos (Amazon/LibraryThing) de 2,8 millones de registros de libros, que contienen metadatos profesionales y sociales, los resultados concluyen que enriquecer los metadatos profesionales con metadatos sociales, especialmente etiquetas, mejora significativamente la búsqueda y la recomendación. Abebe *et al.*<sup>25</sup> también señalaron que los metadatos sociales, especialmente las etiquetas y las reseñas, mejoran significativamente el rendimiento de la búsqueda, ya que los metadatos seleccionados por profesionales “a menudo se limitan” a describir libros y sus atributos.

---

25 Minale A. Abebe *et al.*, *op. cit.*

Otras iniciativas defendidas por Abebe *et al.*,<sup>26</sup> Yadav, Duhan y Bhatia<sup>27</sup> y Xião, Qian y Qin<sup>28</sup> se refieren a la elaboración de metadatos sociales, especialmente en relación con los datos geoespaciales.

Un primer movimiento consiste en el desarrollo de un modelo genérico para describir eventos semánticos y sus relaciones extraídas de metadatos sociales en la web. Es una estructura genérica para detectar, describir y vincular eventos basada en redes sociales llamada SEDDaL, que toma como entrada una colección de objetos de redes sociales de fuentes heterogéneas (por ejemplo, Flickr, Youtube y X –antes Twitter–) y que produce una colección de objetos interconectados semánticamente significativos, eventos con relaciones espaciales, temporales y semánticas.

SEDDaL consta de cuatro módulos principales para: i) describir objetos de redes sociales en un modelo espacial de representación de metadatos (MRSM) genérico que consta de tres dimensiones compuestas: temporal, espacial y semántica; ii) evaluar la similitud entre las descripciones de objetos de redes sociales siguiendo el MRSM; iii) detectar eventos de objetos de redes sociales similares utilizando un algoritmo de aprendizaje no supervisado personalizado, donde los eventos se representan como grupos de objetos en el MRSM; y iv) identificar relaciones direccionales, métricas y topológicas entre eventos siguiendo las dimensiones del MRSM.

Se puede apreciar en los proyectos e iniciativas presentados la preocupación por trabajar con metadatos construidos socialmente y, sobre todo, por utilizar el potencial de las tecnologías disponibles para compartir, interoperar y vincular datos y metadatos de estructuras y dominios heterogéneos.

---

26 *Idem.*

27 Usha Yadav, Neelam Duhan y Komal Kumar Bhatia, “Dealing with pure new user cold-start problem in recommendation system...”.

28 Kejing Xiao, Zhiaopeng Qian y Biao Qin, “A survey of data representation for multi-modality event detection and Evolution”.

## REFLEXIÓN FINAL

El último siglo de práctica bibliotecaria ha mostrado una tendencia continua hacia el intercambio, la colaboración y la estandarización a escala internacional. Como marcadores sociales, podemos destacar el estándar de metadatos MARC21, que consolidó la automatización de las colecciones institucionales, y potenció los servicios de intercambio bibliográfico.

Los metadatos se construyen socialmente y requieren enfoques orientados a la comunidad y el aspecto de colaboración social de la catalogación que involucra a autores, administradores de repositorios, bibliotecarios, consorcios de colecciones digitales, editores, proveedores y usuarios. Esta es una tendencia emergente, que está disolviendo gradualmente los límites entre la “esfera social” y los recursos bibliográficos y llenando la brecha de comunicación entre las bibliotecas y sus usuarios, donde los usuarios finales contribuyen a las descripciones bibliográficas, lo que resulta en una diversidad de metadatos y, además, centrados en el usuario y utilizables.

Adoptar un “enfoque de metadatos mixtos” al considerar que los metadatos bibliográficos y el contenido generado por el usuario son complementarios y esenciales entre sí, sugiere un “cambio de paradigma” en el futuro de los metadatos, de simples a enriquecidos; desde silos de datos legibles por humanos hasta comprensibles por máquina, bien estructurados y reutilizables; de lo invisible y restringido a lo visible y abierto; y desde un OPAC único hasta interfaces web reconfigurables.

La tendencia actual en los esfuerzos de catalogación colaborativa social es esencial para explotar completamente el potencial de los datos abiertos y vinculados. Sin embargo, si observamos algunos dominios, encontramos cuatro grupos, incluidos los bibliotecarios, expertos en datos enlazados, investigadores de recuperación de información y usuarios, todos yendo por caminos separados con una colaboración y comunicación mínimas. Más específicamente, no se benefician entre sí, pero si lo hicieran, podría resultar en mejores posibilidades de descripción, descubrimiento y acceso a los recursos.



Al revisar la literatura científica disponible de dominios relacionados, incluyendo Bibliotecología, Ciencia de la Información, Recuperación de Información y la web semántica, el enfoque de las discusiones está puesto sobre la importancia y el posible impacto de hacer metadatos con curación social como parte de metadatos bibliográficos o curados profesionalmente. Abonando a ello, nos parece relevante que las comunidades de práctica adopten enfoques de catalogación colaborativa social para trabajar con otras partes interesadas, a fin de que sus datos y metadatos bibliográficos estén disponibles y visibles como datos abiertos y vinculados, lo que maximizaría el acceso, el descubrimiento, el uso y la reutilización en múltiples dominios.

## BIBLIOGRAFÍA

- Abebe, Minale A., Joe Tekli, Fekade Getahun, Richard Chbeir y Gilbert Tekli. "Generic metadata representation framework for social-based event detection, description, and linkage". *Knowledge-Based Systems* 188 (January 2020).
- Alemu, Getaneh, Brett Stevens y J. Chandler. "Toward an emerging principle of linking socially-constructed metadata". *Journal of Library Metadata* 14, núm. 2 (2014): 103–29.
- Alemu, Getaneh y Brett Stevens. *An emergent theory of digital library metadata: enrich then filter*. Waltham, MA: Chandos Publishing, Elsevier Ltd., 2015.
- . "A theory of metadata enriching and filtering: challenges and opportunities to implementation". *Qualitative and Quantitative Methods in Libraries* 5, núm. 2 (2017): 311–34.

***Metadatos sociales: iniciativas...***

- Manzo, Christina *et al.* “By the people, for the people: assessing the value of crowdsourced, user-generated metadata”. *Digital Humanities Quarterly* 9, núm. 1 (2015): s.p.
- Seeman, Dean y Heather Dean. “Open social knowledge creation and library and archival metadata”. *KULA: knowledge creation, dissemination, and preservation studies* 3 (2019): 1-11.
- Xiao, Kejing, Zhaopeng Qian y Biao Qin. “A survey of data representation for multi-modality event detection and Evolution”. *Appl. Sci.* 12 (2022): 1-25.
- Yadav, Usha, Neelam Duhan, y Komal Kumar Bhatia. “Dealing with pure new user cold-start problem in recommendation system based on linked open data and social network features”. *Mobile Information Systems* 20 (2020): 1-20.
- Zavalin, Vyacheslav I. y Shawne D. Miksa. “Collecting and evaluating large volumes of bibliographic metadata aggregated in the WorldCat database: a proposed methodology to overcome challenges”. *The Electronic Library* 39, núm. 3 (2021): 486-503.

# Metadatos y seguridad de la información: desafíos y soluciones

JAVIER MONCAYO GARCÍA  
*Suments, España*

## INTRODUCCIÓN

Los metadatos, como datos que describen a otros datos, sirven para ofrecer información descriptiva sobre un recurso. Esta información puede ser sobre la calidad, condición o características del recurso en sí, como puede ser el autor de un libro o la fecha de modificación de un archivo digital. El uso de los metadatos es también de gran utilidad en muchos ámbitos científicos, académicos o sociales, pues permite buscar, recuperar y analizar información relevante. Además, los metadatos están presentes en los archivos que utilizamos en nuestra vida diaria, como documentos tipo .pdf, archivos de imagen .jpg o canciones en formato .mp3. Sin embargo, la presencia de metadatos de gestión de documentos en entornos digitales y sitios web presenta un riesgo en materia de seguridad de la información para las organizaciones y los usuarios, debido a su potencial para revelar datos personales o información sensible.

En este artículo se presenta un análisis del desafío que supone el crecimiento de los metadatos de gestión de documentos en materia de privacidad y protección de datos personales, la relación

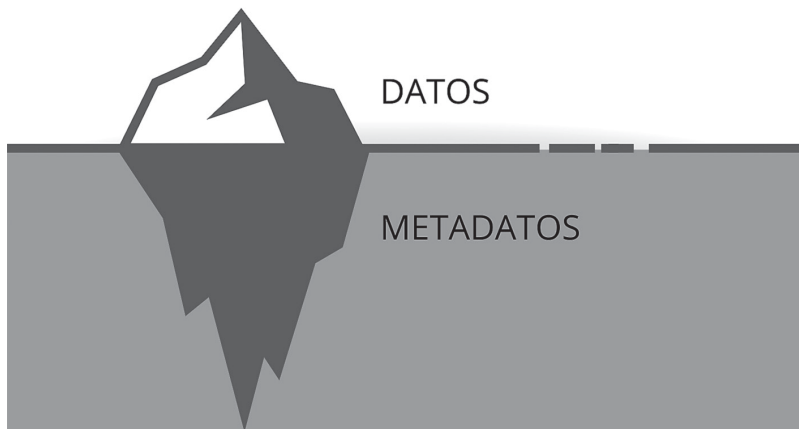
### *Metadatos sociales: iniciativas...*

de los metadatos con algunos de los conceptos del ámbito de la seguridad de la información, así como las soluciones que buscan contribuir al desafío que estos presentan para organizaciones y usuarios en nuestro tiempo.

### METADATOS, DESCONOCIDOS E INFINITOS

Los metadatos están presentes en todo el universo de internet con diferentes formas. No son visibles a primera vista, por eso la mayoría de los usuarios y organizaciones suele ignorar su existencia. En el mundo digital los metadatos son una gran masa de información que no es visible pero no por ello dejan de ser una materia prima muy importante para algunos y olvidados por otros muchos. Toda la actividad que los usuarios desarrollan en el universo digital genera metadatos, ya sea a nivel de documentos (metadatos de gestión de documentos electrónicos) o a nivel de comunicaciones (metadatos de comunicaciones electrónicas).

*Figura 1. Datos y metadatos en la web.*



Fuente: Elaboración propia.

La creación de los metadatos para la gestión de documentos es un proceso continuo en el que se encuentran involucrados creadores, editores, bibliotecarios, usuarios, programadores, webmasters y cualquier usuario responsable del diseño y creación de contenidos en entornos web. De ahí que en los tiempos actuales se esté hablando de los metadatos enriquecidos o los metadatos socialmente construidos. Los metadatos socialmente construidos nos invitan al estudio y el análisis, considerando sus múltiples interpretaciones por parte de quienes usan los recursos de información; además, tienen la capacidad de generar distintos metadatos. En Suments Data ponemos la privacidad, la protección de datos y la seguridad web en el centro de nuestras operaciones y para ello usamos los metadatos como materia prima de nuestras investigaciones y nuevos desarrollos de productos. Todo ello con el fin de crear tecnologías accesibles para facilitar el cumplimiento normativo y la seguridad de la información en las organizaciones.

Los metadatos se organizan con etiquetas (también conocidas como metaetiquetas o *meta tags*). Una etiqueta o metaetiqueta da nombre a un metadato. Por ejemplo, la fecha de creación de un documento o la última modificación de éste. También encontramos etiquetas relativas a las coordenadas GPS donde se ha tomado una fotografía, incluso sobre la velocidad a la que se mueve un dispositivo cuando se graba un video. Al ser la creación de metadatos un proceso continuo, las posibilidades de creación de nuevos metadatos son infinitas, ya que cada usuario puede determinar las etiquetas que considere oportunas para clasificarlos. Se pueden definir tantas metaetiquetas como sea necesario. En Suments Data, llevamos identificadas más de 250 000 metaetiquetas distintas a través de nuestra tecnología Verics para el análisis de metadatos y detección de filtraciones de datos personales en sitios web.

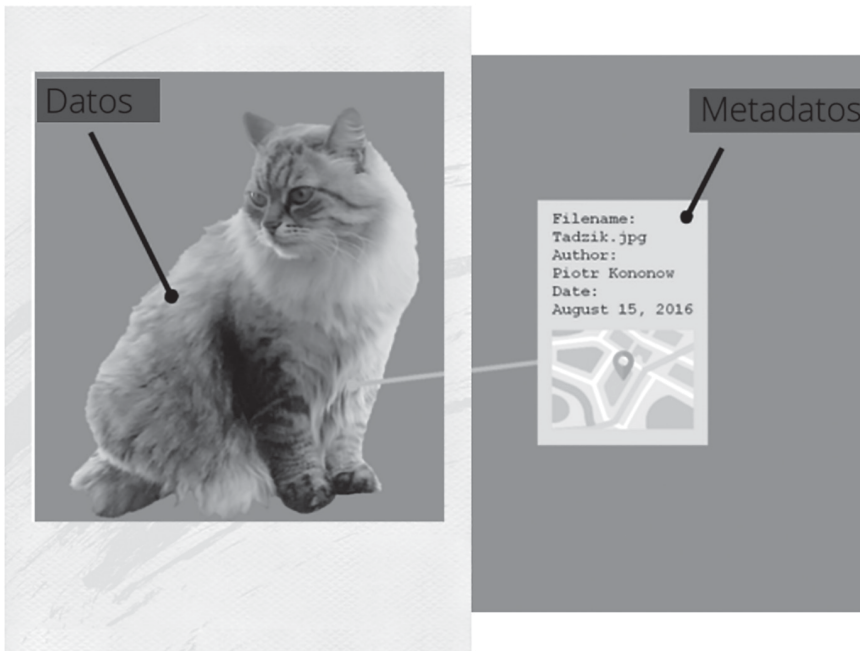
## METADATOS Y LA SEGURIDAD DE LA INFORMACIÓN

Los metadatos de gestión de documentos pueden contener información sensible o datos personales identificables. A partir del

### ***Metadatos sociales: iniciativas...***

análisis de los metadatos de un documento, que nos indican información como el nombre del autor, fecha de creación o dispositivos vinculados, podemos obtener bastante información sobre el contexto de ese archivo: quién ha creado el archivo, cuándo se ha creado, cuándo se ha modificado, con qué dispositivos, con qué sistemas y sobre qué ruta de directorio... lo que da información sobre la estructura interna de carpetas y servidores de un usuario o de una organización.

*Figura 2.* “Los metadatos pueden contener información sensible o datos personales”.



Fuente: Elaboración propia.

Esto es sólo un ejemplo de lo que los metadatos revelan sobre un único archivo. Este mismo potencial revelador de los metadatos asciende a una escala mayor cuando varios archivos se almacenan

dentro de un mismo sitio web. Esta información, agrupada con malas intenciones, es un caldo de cultivo para ataques de *phishing* y *ransomware*, ciberamenazas con las que los cibercriminales pueden utilizar información sensible sobre los usuarios y sus sistemas para lograr sus objetivos, como conseguir credenciales de acceso o infectar equipos con algún virus informático. En este sentido, la presencia de metadatos en sitios web y la exposición de los datos personales e información sensible que estos contienen presenta un problema de seguridad de la información, que hace referencia al conjunto de medidas preventivas y reactivas de las organizaciones y sistemas tecnológicos que permiten resguardar y proteger la información buscando mantener la confidencialidad, la disponibilidad e integridad de datos. Concretamente, la exposición de esta información sensible rompe el principio de confidencialidad.

Se conoce como fuga de datos o fuga de información a la pérdida de la confidencialidad de la información privada de una persona o empresa.<sup>1</sup> Una fuga de información en una organización es la pérdida de confidencialidad de la información que ocurre cuando personal no autorizado accede a información que no le pertenece. Cada vez que la información perteneciente a una empresa acaba en las manos de gente no deseada podemos decir que esta empresa ha sufrido una fuga de información. La fuga de información en una empresa también se conoce como fuga de datos o brecha de datos.

Existen tres principios básicos para garantizar la seguridad de la información: confidencialidad, integridad y disponibilidad. Una fuga de información en una organización indica que se ha roto el primer principio, ya que la información deja de ser confidencial cuando personal no autorizado accede a la misma. Esta fuga puede ser ocasionada de forma interna o externa a la organización. Las fugas de información de causa interna son causadas por el personal o empleados de la empresa, mientras que una causa externa

---

1 Agencia Estatal Boletín Oficial del Estado, “Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica”.

### ***Metadatos sociales: iniciativas...***

podría ser un incidente de seguridad causado por un proveedor o la intervención externa como un robo o un ataque de ciberseguridad a la empresa. Además, la fuga de datos puede ser intencionada o involuntaria. En el caso de una fuga de información por metadatos, normalmente se trata de una fuga de información interna e involuntaria, ya que se realiza por empleados de la empresa sin conocimiento previo de las consecuencias de subir documentos a un sitio web sin hacer una correcta limpieza de metadatos.

La fuga de información por metadatos permite que cualquier usuario tenga acceso a datos personales e información sensible como nombres y apellidos, números de teléfono, nombres de usuario de correo electrónico, rutas de directorios y coordenadas GPS (estos últimos son carne para ataques de *phishing*). Algunas consecuencias de la fuga de metadatos en organizaciones son las sanciones económicas, el riesgo de recibir ciberataques, exponer información confidencial y la pérdida de confianza de los consumidores. Los metadatos son un concepto importante en la norma ISO 27001,<sup>2</sup> que detalla los requisitos para establecer, implantar, mantener, supervisar y mejorar el sistema de gestión de la seguridad de la información de una organización.

## **METADATOS Y BRECHAS DE SEGURIDAD DE DATOS PERSONALES**

El impacto de una fuga de información puede ser aún más grave si se comunican datos personales identificables, lo que podría poner en peligro la seguridad de la organización o la privacidad de sus usuarios y, de esta forma, entrar en conflicto con la normativa o legislación vigente en materia de protección de datos personales, como es el Reglamento General de Protección de Datos (RGPD) o la Ley Federal de Protección de Datos Personales en Posesión de los Particulares de México.

---

2 Centro Criptológico Nacional (CCN), “Guía de Seguridad de las TIC CCN-STIC 835”.



Si un metadato contiene un dato personal (nombre, dirección de correo electrónico o número de teléfono) este metadato pasa a ser un dato en sí mismo atendiendo a la información que contiene (información personal o sensible). De esta forma, si un metadato revela un dato personal, el acceso y comunicación de esta información debe regularse por la normativa vigente. En el apartado 1 del artículo 4 del RGPD se incluye la definición de dato personal como “toda información sobre una persona física identificada o identificable”.<sup>3</sup> La difusión de datos personales identificables, sea voluntaria o accidental, puede suponer un incumplimiento del Reglamento General de Protección de Datos (RGPD), que en su artículo 4, apartado 12, hace referencia al concepto de “violación de la seguridad de los datos personales”, que se define como “toda violación de la seguridad que ocasione la destrucción, pérdida o alteración accidental o ilícita de datos personales transmitidos, conservados o tratados de otra forma, así como la comunicación o acceso no autorizados a dichos datos”.

El RGPD es el reglamento europeo relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos. Es una normativa a nivel de la Unión Europea, por lo que cualquier empresa de la unión, o aquellas empresas que tengan negocios en la Unión Europea, que manejen información personal de cualquier tipo, deberán acogerse a ella. Las multas por el incumplimiento del RGPD pueden llegar a los 20 millones de euros.

---

3 European Union Law, “Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)”.

## METAETIQUETAS Y RGPD

En Suments llevamos más de 40 000 000 de documentos únicos analizados como parte de nuestras labores de análisis y de entrenamiento de nuestros sistemas. A partir del uso de técnicas Big Data para la correlación de estos datos y la obtención de conocimiento, nuestro equipo ha desarrollado distintos casos de estudio de investigación en torno a los metadatos y la privacidad, lo que nos ha permitido destapar fugas de información graves así como desarrollar nuevos modelos de negocio basados en modelos de Inteligencia Artificial.<sup>4</sup> Debido a que la creación de metadatos es un proceso continuo y con distintos agentes creadores, las posibilidades de creación de nuevos metadatos son infinitas. Durante el desarrollo de nuestras tecnologías, a partir nuestra infraestructura para el análisis y recopilación de datos de fuentes públicas, en Suments hemos identificado más de 250 000 metaetiquetas distintas, y hemos realizado un análisis de datos para identificar el porcentaje de estas etiquetas que presentan un incumplimiento del RGPD por el hecho de contener datos personales o información sensible.

Los resultados de nuestro análisis muestran que un 0.13% de estos *meta tags* tienen potencial para revelar datos personales o información sensible. Dentro de este 0.13%, un 41% revela datos sensibles y un 56% revela datos personales (figura 3).

Aunque esta cifra parece pequeña, no es insignificante, pues este pequeño porcentaje de metaetiquetas es ampliamente utilizado en distintos documentos.

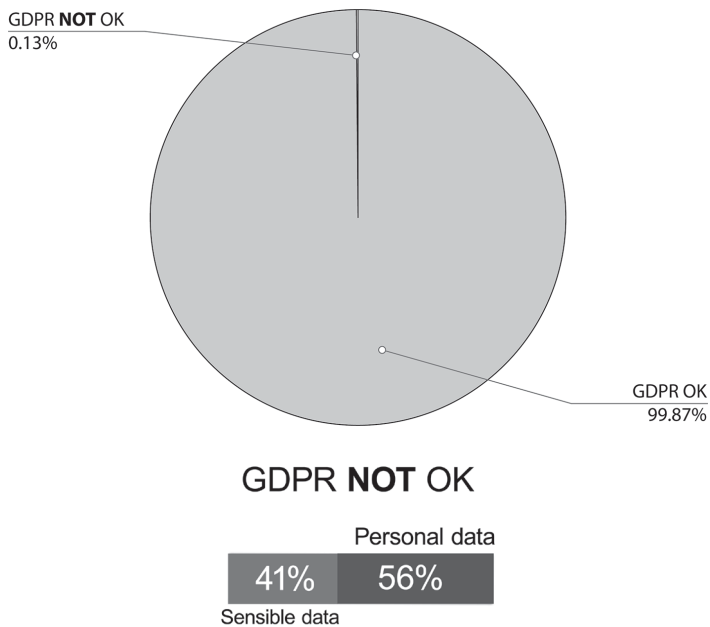
## ADMINISTRACIONES PÚBLICAS EN ESPAÑA, OBLIGADAS A ELIMINAR METADATOS

El Real Decreto 3/2010 de 8 de enero, actualizado por el Real Decreto 951/2015, del 23 de octubre, por el que se regula el Esquema

---

4 INCIBE, “Glosario de términos de ciberseguridad - Una guía de aproximación para el empresario”.

Figura 3. “0.13 de estos meta tags tienen potencial para revelar datos personales...”.



Fuente: Elaboración propia.

Nacional de Seguridad en el ámbito de la Administración Electrónica en España (en adelante, ENS)<sup>5</sup> establece la política de seguridad en la utilización de medios electrónicos para lograr una protección adecuada de la información en el ámbito de las administraciones públicas españolas.

El ENS establece una serie de medidas de seguridad en su anexo II. Dentro de estas medidas de seguridad, definidas en el ENS como [mp], encontramos una medida relativa a la “Limpieza de documentos”, [mp.info.6]. Esta medida determina lo siguiente: “En

5 International Standard Association (ISO), “ISO/IEC 27001:2013 Information technology - Security techniques - Information security management systems - Requirements”.

### ***Metadatos sociales: iniciativas...***

el proceso de limpieza de documentos, se retirará de estos toda la información adicional contenida en campos ocultos, metadatos, comentarios o revisiones anteriores, salvo cuando dicha información sea pertinente para el receptor del documento”.

Como se indica en la [mp.info.6], dicho proceso deberá aplicarse especialmente “cuando el documento se difunde ampliamente, como ocurre cuando se ofrece al público en un servidor web y otro tipo de repositorio de información”. En este contexto, el Centro Criptológico Nacional de España (CCN-STIC), publicó la guía “La gestión segura de metadatos”,<sup>6</sup> con el objetivo de proporcionar una guía de buenas prácticas para realizar la inspección y borrado tanto de metadatos como de otros datos ocultos asociados a los documentos electrónicos. Este documento forma parte del conjunto de normas desarrolladas por el CCN para la implementación del Esquema Nacional de Seguridad (CCN-STIC-800), siendo de aplicación obligatoria para el sector público y teniendo como objeto la protección de los servicios prestados a los ciudadanos y entre las diferentes administraciones públicas.

Esta normativa, actualizada en 2022, es un avance pionero en materia de metadatos y seguridad de la información que ya está actuando contra el desafío que plantean los metadatos en materia de privacidad y ciberseguridad en las organizaciones públicas y privadas.

### **METADATOS, FACTOR DE RIESGO SEGÚN LA AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS**

La Agencia Española de Protección de Datos (AEPD) ha publicado recientemente la guía “Gestión del riesgo y evaluación de impacto en tratamientos de datos personales”.<sup>7</sup> Este documento, dirigido

---

6 “Blog de protección de datos”.

7 Agencia Española de Protección de Datos, “Gestión del riesgo y evaluación de impacto en tratamientos de datos personales”.

principalmente a responsables, encargados de tratamientos y delegados de protección de datos (DPD), ofrece una información actualizada con orientaciones necesarias para realizar una Evaluación de Impacto para la Protección de Datos (EIPD) y llama la atención sobre el riesgo de los metadatos.

La EIPD es una herramienta clave que permite a las organizaciones evaluar los potenciales riesgos a los que están expuestos los datos personales en función de las actividades de tratamiento que se realizan. En este sentido, la prevención es la mejor arma, por eso es importante conocer bien cuál es el riesgo que tienen los metadatos a la hora de hacer una EIPD. En el punto IV del documento, en la “lista de factores de riesgo identificados en la normativa RGPD”, la AEPD menciona los metadatos hasta en tres ocasiones.

En primer lugar, el documento identifica que los metadatos en sí, en cualquiera de sus formas, son un factor de riesgo de nivel medio. También indica que otro factor de riesgo medio son los “datos y metadatos de las comunicaciones electrónicas y datos inferidos de las comunicaciones electrónicas”, como los derivados del tratamiento de datos mediante correos electrónicos, mensajes instantáneos, llamadas de teléfonos o videollamadas.

Por otro lado, la AEPD señala otro factor de riesgo de nivel medio que son los “identificadores únicos”. Algunos ejemplos de identificadores únicos pueden ser una dirección IP, un número de teléfono o la matrícula de un vehículo o un dígito de DNI. En este caso la AEPD menciona el riesgo de los metadatos cuando son “Identificadores únicos añadidos a archivos (por ejemplo, metadatos de fotografías subidas a redes sociales)”. Este punto es importante pues, según la experiencia de Suments Data, en la detección de filtraciones de metadatos con Verics, el riesgo de los metadatos, cuando no están controlados, puede alcanzar grandes dimensiones: son capaces de generar miles de filtraciones de información sensible (identificadores únicos como direcciones de correo electrónico o incluso DNIS). Realmente nadie es ajeno al riesgo de los metadatos, incluso en aquellas organizaciones que parecen ser las más “seguras” hemos encontrado filtraciones de datos personales con nuestra tecnología. En el apartado “Medidas de protección de

### ***Metadatos sociales: iniciativas...***

datos desde el diseño”, de la página 111 de la guía, la AEPD destaca la importancia de la “depuración de entrada de datos y metadatos”, al seguir el objetivo de “minimizar” riesgos evitando el tratamiento de datos personales innecesarios. Para ello, las tácticas de la AEPD a seguir con este tipo de filtraciones son claras: seleccionar, excluir, podar y eliminar.

## **SOFTWARE PARA EL ANÁLISIS Y LIMPIEZA DE METADATOS**

La mayoría de las plataformas de redes sociales eliminan los metadatos de las fotos y archivos que los usuarios suben a la nube de forma automática antes de ser publicadas. Sin embargo, esto no es así en sitios web, que necesitan la responsabilidad activa del propietario del sitio web para hacer esta limpieza de metadatos. A partir del uso de técnicas de Inteligencia Artificial aplicadas a los metadatos, en Suments hemos desarrollado distintas soluciones que permiten garantizar la seguridad de la información en internet y combatir las fugas de información por metadatos en los entornos web de las organizaciones.

Verics es un software para el análisis de los metadatos de los sitios web que permite la identificación de las filtraciones de datos personales e información sensible. Las organizaciones y los usuarios pueden auditar sus sitios web para identificar fugas de datos personales e información sensible sin necesidad de conocimientos técnicos ni de financiar servicios de ciberseguridad.

Verics analiza los diferentes archivos y documentos hechos públicos en los dominios y subdominios de la web. Es en estos mismos donde aparecen los metadatos, y con alta probabilidad algunos de ellos serán datos personales que infringen las leyes gubernamentales en materia de protección de datos.

*Figura 4. “...algunos de ellos serán datos personales que infringen las leyes gubernamentales...”.*



Fuente: Elaboración propia.

Metawash es una solución de software para la limpieza activa de metadatos en tiempo real. Elimina las fugas de datos personales e información sensible en entornos web y ayuda a las organizaciones a cumplir las normativas de seguridad de la información y protección de datos. Metawash funciona en la nube y los usuarios pueden utilizarlo con integración de API o sin necesidad de código con integración en los gestores de contenidos web WordPress o Joomla.

## CONCLUSIONES

Los metadatos de gestión de documentos juegan un rol importante en todo el entorno digital y, particularmente, en el ámbito de las bibliotecas digitales, pues permiten localizar y descubrir recursos de información. Sin embargo, el uso de estos en entornos digitales debe atender a ciertas precauciones de privacidad y protección de datos para responder a los desafíos de la seguridad de la información de nuestro tiempo.

### *Metadatos sociales: iniciativas...*

La creación de metadatos, como proceso continuo y con diferentes creadores, ha dado lugar a los llamados metadatos sociales o metadatos socialmente contruidos que, al ser un proceso de creación de carácter infinito, ha aumentado considerablemente el desafío de los metadatos ante la seguridad de la información.

Los metadatos de gestión de documentos de un archivo tienen un gran potencial para revelar información sensible y datos personales. El riesgo de este potencial revelador aumenta en entornos web, donde se almacenan muchos tipos de archivos y se exponen de forma pública en internet.

Las organizaciones deben realizar una correcta gestión de los metadatos de gestión de documentos con el objetivo de prevenir las fugas de información por metadatos y sus consecuencias negativas.

Los metadatos pueden considerarse datos personales siempre que su contenido revele información sobre una persona física identificada o identificable. Si esta información se comunica incorrectamente o se accede a la misma de forma no autorizada, se incurrirá en una brecha de seguridad de datos personales o violación de la seguridad de datos personales.

Sólo un 0.13% de las metaetiquetas analizadas por Suments suponen un riesgo de exposición de datos personales o información sensible. Sin embargo, este pequeño porcentaje de metaetiquetas es utilizado en la mayoría de los documentos que circulan en el entorno digital.

Ya existen normativas legales que están actuando para abordar el desafío que presentan los metadatos para la seguridad de la información de las administraciones públicas. En el caso de España, la limpieza de metadatos de documentos es una medida de aplicación obligatoria para todas las administraciones públicas.

Las organizaciones públicas y privadas pueden acceder a herramientas tecnológicas que permiten evitar el incumplimiento de normativas legales de protección de datos. Ejemplo de ello son las herramientas Verics y Metawash de Suments.



## REFERENCIAS BIBLIOGRÁFICAS

- Agencia Española de Protección de Datos. “Gestión del riesgo y evaluación de impacto en tratamientos de datos personales”. <https://www.aepd.es/sites/default/files/2019-09/guia-evaluaciones-de-impacto-rgpd.pdf> (consultado el 14 de agosto de 2022).
- Agencia Estatal Boletín Oficial del Estado. “Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica”. <https://www.boe.es/eli/es/rd/2010/01/08/3> (consultado el 14 de agosto de 2022).
- Centro Criptológico Nacional (CCN). “Guía de Seguridad de las TIC CCN-STIC 835”. <https://www.ccn-cert.cni.es/pdf/guias/series-ccn-stic/800-guia-esquema-nacional-de-seguridad/2031-ccn-stic-835-borrado-de-metadatos-en-el-marco-del-ens/file.html> (consultado el 14 de agosto de 2022).
- European Union Law. “Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)”. <http://data.europa.eu/eli/reg/2016/679/oj> (consultado el 14 de agosto de 2022).
- INCIBE. “Glosario de términos de ciberseguridad - Una guía de aproximación para el empresario”. [https://www.incibe.es/sites/default/files/contenidos/guias/doc/guia\\_glosario\\_ciberseguridad\\_2021.pdf](https://www.incibe.es/sites/default/files/contenidos/guias/doc/guia_glosario_ciberseguridad_2021.pdf) (consultado el 14 de agosto de 2022).

***Metadatos sociales: iniciativas...***

International Standard Association (ISO). “ISO/IEC 27001: 2013 Information technology - Security techniques - Information security management systems - Requirements”. <https://www.iso.org/standard/54534.html> (consultado el 14 de Agosto de 2022).

Suments Data. “Blog de protección de datos”. <https://suments.com/es/blog-de-proteccion-de-datos/> (consultado el 14 de agosto de 2022).

## APLICACIONES

# Metadatos sociales y preservación digital: cinco retos para las instituciones de la memoria

ARIEN GONZÁLEZ CRESPO  
*El Colegio de México, México*

## INTRODUCCIÓN

**E**n una encuesta realizada por Li Liew,<sup>1</sup> en la que respondieron 94 instituciones de información de 27 países, se encontró que, aunque casi todas las bibliotecas y archivos reconocían la importancia de trabajar con metadatos sociales, la razón principal por la que no lo hacían era precisamente por la “falta de experiencia o conocimiento al respecto”.<sup>2</sup> Espacios profesionales como este libro son necesarios para cambiar esta realidad, muy particularmente en América Latina y el Caribe, pues no sólo nos permiten reconocer la existencia de estos entresijos de la información, sino compartir colectivamente lo necesario para una transformación de nuestras prácticas.

La preservación digital de los contenidos de la web 2.0, por otra parte, debería ser también uno de los objetivos más importantes de trabajo para las instituciones de información en la sociedad

---

1 «Social Metadata and Public-Contributed Contents in Memory Institutions: “Crowd Voice” Versus “Authenticated Heritage”?».

2 *Ibid.*, 125.

actual. En primer lugar, por el hecho de que una parte cada vez más creciente y significativa de la comunicación humana, en todos los sectores de la sociedad, se produce a través de estos medios. Académicos e investigadores, periodistas, políticos, economistas y negociantes estudiosos del mercado, abogados, funcionarios públicos, maestros: todos ellos y otros más, han reconocido la importancia de la web 2.0 como fuente primaria de investigación, y ya no es extraño, en nuestras instituciones de información, encontrar entre las donaciones manifestaciones impresas o digitales de medios sociales, como capturas de pantalla o transcripciones de X (antes Twitter), Facebook, Instagram, etcétera. Por otro camino muy relacionado, desde las humanidades digitales y la ciencia de datos se desarrollan y difunden metodologías para su estudio, incluso desde las propias bibliotecas académicas. Los resultados del análisis de datos de las redes sociales, calificados y tratados como Big Data en muchas ocasiones, poseen altísimo valor para el mercado y las compañías de medios sociales exitosas han resultado empresas millonarias. Para las bibliotecas y archivos, entonces, dejar de lado estos cuerpos de información y de memoria social significaría, para un futuro no muy lejano, quedar ellas mismas al margen y dejar de servir en su función histórica como repositorios de conocimiento y memoria.

Y si bastara con la resolución de emprender la tarea, no tendríamos mayor problema. Pero es que resulta que preservar la web 2.0 o los medios sociales, es también uno de los trabajos más difíciles y menos conocidos que se pueden emprender por parte de cualquier institución de información; ellas han estado por cientos de años preparadas para lidiar con objetos materiales concretos, tangibles, fijos, desde una posición de autoridad intelectual sobre su cuidado, custodia y representación, pero no para el mundo digital.

Filosóficamente hablando, el tratar de reconciliar preservación y medios sociales o web 2.0, resulta un intento de reconciliar a dos contrarios dialécticos: lo permanente con lo efímero. Los medios sociales, diseñados para correr, apurarse, avanzar, transcurrir, transformarse, no durar... y la preservación, una disciplina

cuyo sentido es intentar fijar los objetos y retenerlos, conquistar para ellos “la eternidad”. Preservar las redes sociales equivale, en sentido poético, a atrapar “el momento fugaz y todas las sugerencias de eternidad que contiene”, como diría Baudelaire.

La preservación digital, desde sus inicios, ha partido de la pregunta de qué es la cosa que se debe preservar, operacionalizando esta visión “esencialista”, entre otras formas, a través de la definición de “características fundamentales o esenciales” del objeto a preservar. La definición de las características esenciales de un objeto de información se torna conflictiva cuando una de las propias características esenciales es el cambio permanente, la interacción, la transitoriedad; como en el caso de los medios sociales, objetos de los cuales se podría decir que son más devenir que ser. Este intento de buscar “armonía entre tensiones opuestas”, remueve las bases de la disciplina de la preservación con las que usualmente trabajamos. Esto reconocía el Proyecto InterPARES 2 desde hace más de una década, cuando emprendió la investigación de la preservación digital de objetos interactivos, dinámicos y experienciales, tal vez el mejor cuerpo investigativo para encontrar bases teóricas para realizar un abordaje práctico de este tema.

Si hacemos un análisis económico del asunto el panorama no mejora, más bien empeora. El hecho de que se dediquen millones a desarrollar tecnologías y plataformas a un ritmo acelerado, que resultan en la producción de cantidades enormes, masivas, de objetos digitales, cada vez más complejos, establece desde afuera determinaciones sobre los sistemas de preservación de la memoria, sobre las bibliotecas y los archivos. Lo que usa y produce la sociedad es lo que nos corresponde preservar por fuerza, o sea que emprender la preservación de los medios sociales no es una opción, es más bien una obligación de instituciones de información y memoria.

La producción de objetos digitales en la web 2.0 se aceleró mucho más después de la monetización de los contenidos generados por los usuarios y sus interacciones con los sistemas, entre ellos, los metadatos sociales. La preservación digital ha desarrollado sobre todo un enfoque normativo para lidiar con la diversidad

creciente, la rápida obsolescencia y las enormes cantidades de objetos. Es decir, les decimos a los productores cómo deben producir para que nuestros sistemas sean capaces de preservar. En el caso de los medios sociales este enfoque no es factible. Son determinaciones externas al sistema de preservación, que parten de fuerzas sociales que atraviesan nuestro campo de estudio y trabajo; y los enormes recursos dedicados a la producción de esas plataformas y contenidos digitales contrastan contra la cantidad muy escasa de recursos dedicados a una rama específica que suele ser colocada (tendría que apuntar que incorrectamente) en el ámbito del consumo final de estos objetos, de su acumulación en espacios sociales de la cultura, es decir, en la rama de las bibliotecas, archivos y otras instituciones o personas que socialmente cumplen o podrían cumplir esta función desde una perspectiva incluso postcustodial.<sup>3</sup>

Para llevar este contraste a pequeña escala, pensemos hoy en día en cada una de nuestras instituciones cuántas personas están dedicadas a la creación directa o al apoyo de la creación de productos de información digitales que tienen su destino en la web, y compáremoslo con cuántas personas y recursos se dedican a valorar, describir y preservar los contenidos que ellos producen. La tarea práctica de preservar los productos digitales de la web 2.0, complejizada con la producción de sus altos volúmenes, es cuando menos intimidante, por no llegar a decir que en muchos casos paralizante.

En ese panorama, piezas muy importantes dentro del funcionamiento de los medios sociales y la web 2.0 son precisamente los metadatos sociales. Dentro de la preservación digital de los contenidos de los medios sociales, el trabajo con los metadatos sociales resulta ser otro de los grandes espacios en blanco. En una investigación que tal vez sea la más extensa y abarcadora que se ha hecho sobre metadatos sociales en bibliotecas, archivos y museos,

---

3 Lo postcustodial es un término procedente de la Archivística y se define como lo relativo a “situaciones en las que los creadores de los documentos continúan manteniendo los documentos” en su poder y custodia. Véase <https://dictionary.archivists.org/entry/postcustodial.html>.

realizada por Karen Smith-Yoshimura y otros, y publicada en tres partes por OCLC entre 2011 y 2012, la preservación presenta apenas dos menciones, una en el título de un trabajo citado y otra de refilón, en el análisis de los resultados de la valiosa encuesta realizada por ellos. Otras bibliografías consultadas sobre metadatos sociales reafirman la misma situación. La preservación digital es el gran ausente en estos trabajos.

Se presenta en lo que sigue un análisis de las relaciones entre los metadatos sociales y la preservación digital. Se partirá de un estudio de los conceptos relacionados y se llegarán a identificar posteriormente los cinco retos principales que, considero, podrían encontrar las instituciones de información para abordar este asunto en la práctica.

#### ALGUNAS PRECISIONES CONCEPTUALES

Los medios sociales son un conjunto de “aplicaciones basadas en internet o dispositivos móviles que permiten a los usuarios formar redes en línea o comunidades”.<sup>4</sup> Day Thomson<sup>5</sup> señala que, usualmente, web 2.0 y medios sociales son tratados como sinónimos, siendo la web 2.0 una generación posterior derivada de los primeros medios sociales como Usenet, y después Friendster y MySpace. La web 2.0 tiene como una de sus características fundamentales el hecho de presentar una gran cantidad de hipervínculos externos y diversos tipos de información embebidos (información multimedia, sonora, visual, audiovisual, textual) y, como está dirigida a fomentar el intercambio de información entre usuarios, genera una alta cantidad de información secundaria en la forma de metadatos.

Entre los metadatos generados se encuentra la subclase de los metadatos sociales. Estos son “información adicional sobre un recurso, que resulta de la **contribución de los usuarios y de la**

---

4 Helen Hockx Yu, “Archiving social media in the Context of Non-Print Legal Deposit”, 2.

5 Sara Day, *Preserving social media*.



**actividad online** (tales como etiquetado, comentarios, reseñas, imágenes, videos, ratings, recomendaciones) que ayudan a las personas a encontrar, **comprender** o evaluar el contenido”.<sup>6</sup> Dos rasgos de este concepto podrían ser claves: el primero es que entre las funciones de estos metadatos sociales está garantizar precisamente la **comprensibilidad** del objeto, el segundo es que estos son producidos por “otros” que no son el propio sistema que contiene al objeto de información, es decir, usuarios, grupos de usuarios, público, derechohabientes, interesados, creadores, editores, etcétera. En otras palabras, es producido por la o las **comunidades**. Todos menos quienes tienen el control del poder sobre el propio sistema en que se presentan los objetos. Así, los metadatos sociales siempre los producen eso que las Ciencias Sociales han venido a llamar “un otro”, alguien que no tiene el control mayoritario del poder sobre el sistema de información.

Ahora veamos el sistema de conceptos relacionados con la preservación digital que propone la norma ISO 14721:2012 Reference Model for an Open Archival Information Systems (OAIS), la más importante norma para la generación de repositorios confiables para la preservación digital a largo plazo. En ésta, preservación digital es “el acto de mantener la información, en una forma **independientemente comprensible** (*understandable*) por una Comunidad Designada, con evidencia que soporte su autenticidad, en el largo plazo”,<sup>7</sup> entendiendo a su vez el largo plazo como “un periodo para el que exista preocupación acerca de los impactos de las tecnologías cambiantes sobre la información que es resguardada en un repositorio, incluyendo el soporte para nuevos medios y formatos de datos, y **de una comunidad cambiante de usuarios**”.<sup>8</sup> Este periodo “se extiende a un futuro indefinido”. Esta comunidad cambiante posee lo que OAIS llama una Base de Conocimiento,

---

6 Karen Smith-Yoshimura y Cindy Shein, *Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews*, 10.

7 Consultative Committee for Space Data Systems (CCSDS), *Reference Model for an Open Archival Information System...* 1-13.

8 *Ibid.*, 1-12.

que es “un conjunto de información, incorporado por una persona o sistema, que permite a esa persona o sistema comprender la información recibida”.<sup>9</sup>

Atendamos a dos rasgos de este sistema de tres conceptos: el primero es la definición de la preservación en función de la comprensibilidad del objeto, es decir, como el mantenimiento o garantía de la comprensibilidad de un objeto en un futuro indefinido, independientemente del sistema que lo preserva. El segundo rasgo es la definición de esa comprensibilidad no en abstracto, sino por parte de una comunidad específica, un “otro” externo al sistema de información, llamado Comunidad Designada, que a su vez cambia y que posee un conjunto de conocimientos en función del cual se define si el objeto es comprensible o no. Cambia el entorno tecnológico del objeto digital, pero también cambia la comunidad que lo debe comprender y cambian las condiciones de comprensibilidad, todo esto sin que necesariamente el objeto cambie. Esto significa que podemos garantizar que un archivo esté disponible, pero no que esté comprensible, y de aquí la importancia de la retención y el trabajo con los metadatos sociales para la preservación digital que ya hemos mencionado. Radica ahí la diferencia entre una preservación a nivel de bits, y una preservación a nivel intelectual de los contenidos, sin convertir el mantenimiento de copias (como hasta el momento fue en la historia de nuestras organizaciones), en un sinónimo de preservación digital. Para decirlo en términos comunes: podemos, por ejemplo, mantener perfectamente un documento en un .pdf/a no transformado tal y como éste es, y aun así no estaríamos garantizando la preservación digital del mismo.

Para recapitular, el conjunto de procesos que comprende la preservación digital tiene en su centro al concepto de comprensibilidad para un conjunto predeterminado de personas, exteriores al sistema de información, que poseen una base de conocimiento propia en función de la cual se formula esa comprensibilidad.

---

9 *Ibid.*, 1-12.

A simple vista, pues, los conceptos de preservación digital y de metadatos sociales tienen enormes resonancias: ambos colocan en el centro a la **comprensibilidad del objeto** y a la o las **comunidades** que se acercan a este objeto partiendo de su base de conocimiento, que puede ser distinta, otra, de la del sistema de información. ¿Qué supondría entonces emprender las cuestiones de preservación digital considerando la existencia de los metadatos sociales? Propongo examinar algunos elementos a través de la exposición de cinco retos, que seguramente no serían los únicos en un análisis más profundo y detallado.

#### CINCO RETOS, ENTRE OTROS...

##### *1. Identificar a los metadatos sociales como objeto o parte del objeto de la preservación digital*

Volvamos a la encuesta mencionada al inicio. Otra de las causas identificadas por las que las instituciones no trabajan con metadatos sociales es que lo consideran innecesario o inútil, es decir, la no existencia de interés en hacerlo. Entre las motivaciones que las instituciones ofrecieron a Smith-Yoshimura y Shein<sup>10</sup> para sí trabajar con metadatos sociales estuvieron:

- Obtener puntos de vista diferentes,
- Responder a directivas del gobierno y otras autoridades,
- Obtener puntos de vista a partir de comunidades específicas,
- Crear presencia en el espacio digital,
- Incrementar el involucramiento con el público,
- Comprender los intereses del público,
- Desarrollar la recuperabilidad de los ítems o colecciones (actualizar, complementar o enriquecer los metadatos institucionales existentes).

---

10 Smith-Yoshimura y Shein, *op. cit.*, 125.

En ningún caso el apoyo a los procesos de preservación del patrimonio digital fue identificado ni siquiera como una causa menor. Y aquí radica un reto que debemos enfrentar: la necesidad de identificar a los metadatos sociales como un objeto de la preservación digital.

Si se profundiza en el asunto conceptual, al clasificar los metadatos sociales atendiendo a quién los gestiona, podemos ubicar al menos dos tipologías relevantes para bibliotecas, archivos o museos. Una de ellas incluye aquellos metadatos sociales gestionados por los sistemas de información de sus metabuscadores, sus catálogos, sus sistemas de autodepósito, sus repositorios digitales o exposiciones virtuales. En la otra tipología ya suelen incorporar metadatos sociales a través de los cuales sus públicos pueden contribuir en la web para describir, clasificar, identificar, valorar, contextualizar los objetos que ya se encuentran dentro de éste o van a formar parte del mismo. Esta clase de metadatos sociales, aunque existen sobre tecnologías web, pueden referirse a cualquier tipo de objeto digital no web, e incluso a cualquier tipo de objeto de información no digital.

En estos sistemas se aprovechan las tecnologías participativas para disminuir la hegemonía intelectual de estos centros sobre la representación del conocimiento, de la memoria, etcétera o, cuando menos, como un intento de promoción de un diálogo entre miembros de las comunidades y el sistema de información y/o entre los miembros de las propias comunidades (pueden ser también una forma de “crear comunidad”), ofreciendo una alternativa de cocreación de contenidos (en el caso de los autodepósitos, por ejemplo). Estos están dentro del aliento de promover una cultura participativa en el espacio de información o memoria. Entre ellos, los metadatos más utilizados son las etiquetas descriptivas, los comentarios cortos y la recomendación del contenido.

Generalmente cuando se aborda el tema de los metadatos sociales en bibliotecas, archivos y museos, se suele referir a esta tipología mencionada, a las implicaciones de los metadatos sociales para las colecciones que se encuentran ya dentro del sistema de información, es decir, de la creación de datos socialmente

### ***Metadatos sociales: iniciativas...***

construidos sobre los objetos patrimoniales, o sobre los procesos asociados al patrimonio, ya cuando éste se encuentra capturado por el sistema de preservación.

La otra tipología incluye los metadatos sociales que incorporan productos de la web y las redes sociales, fuera de la administración o gestión de las instituciones de memoria, cuyo sentido y significado primario es establecer características del objeto por parte de sus creadores fundamentalmente, o por parte de tecnologías puestas en función de la identificación, clasificación, distribución y presentación de los contenidos en línea. Entre los metadatos que más comúnmente se usan están:

- Title
- Description
- Image
- Url
- Type

A diferencia de la tipología anterior, no tanta atención se ha prestado a las implicaciones de esta clase de metadatos sociales para bibliotecas, archivos y museos; los que están referidos a los objetos de información antes de que estos sean ingestados por el sistema son parte del objeto mismo de colección. En la literatura, frecuentemente ellos son tratados como *records*, es decir, información primaria que posee un valor evidencial, tanto para la construcción de conocimiento académico, para las ciencias y las humanidades, como un valor evidencial para propósitos jurídicos y regulatorios.

Sobre estos, Day Thomson comienza su resumen ejecutivo re-marcando lo siguiente:

[...] las plataformas de medios sociales brindan una alta cantidad de datos legibles por máquina que sirven como base para análisis computacionales, una práctica emergente utilizada por un conjunto de disciplinas académicas, periodísticas y de otros profesionales. Con el objetivo de recolectar y preservar este tipo de datos

“ricos” de una forma significativa, las instituciones de investigación y colección necesitan nuevos acercamientos y métodos. Estos métodos van a necesitar capturar los datos y sus contextos afiliados a gran escala [léase aquí metadatos sociales], en lugar de copias y capturas de pantalla de las páginas web.<sup>11</sup>

En una gran cantidad de casos la preservación de los metadatos sociales es tan importante como la de los datos a los que ellos refieren y, por tanto, desde el punto de vista de la preservación ellos también se consideran parte propia del objeto a preservar. En tanto formen parte del objeto de preservación, no cabe duda de que se trata de una cuestión relativa al carácter íntegro del objeto y que, por tanto, su resguardo tiene importantes consecuencias sobre la confiabilidad de la información y su autenticidad a largo plazo. Puesto que se trata de la integridad del objeto, se trata entonces directamente también de su confiabilidad y autenticidad. Y esto es válido no sólo para las instituciones de memoria, también lo es para los creadores de contenidos, quienes en muchos casos estarían obligados a preservar estos metadatos sociales como parte de su comunicación institucional para propósitos de rendición de cuentas y auditoría. La Financial Industry Regulatory Authority (FINRA), por ejemplo, establece en sus reglas la obligatoriedad de preservar la comunicación en redes sociales como si fueran *records* y se plantea allí la importancia de que se capturen los metadatos como parte de la confiabilidad de esa información.<sup>12</sup>

Parte de la importancia de estos metadatos para la preservación digital deriva de dos elementos fundamentales: su función y su origen o procedencia. En relación con su función, se suele hacer énfasis en su valor para la localización y descubrimiento de la información. Pero estos tienen una importancia fundamental para garantizar una adecuada presentación visual de los contenidos y

---

11 Sara Day *op. cit.*, 2.

12 Véase la siguiente página para las cuestiones específicas sobre el tema de los medios sociales en las reglas de FINRA: <https://www.finra.org/rules-guidance/key-topics/social-media>.

en la autenticación de los mismos, así como para determinar la circulación de esa información, lo que los hace directamente relevantes para la preservación digital. Implícitamente se comprende que la circulación de un objeto informativo determina también parte de su valor y la prioridad para su preservación.

En relación con su origen, aunque son varios los actores que pueden participar en la generación de los metadatos sociales, y en todos los casos hay implicaciones para el sistema de preservación, es fundamental la participación, entre otros, del propio creador de los contenidos digitales. Pensemos, por ejemplo, en los metadatos sociales del protocolo Open Graph o Twitter Cards, que determinan metapropiedades de los contenidos publicados en la web. Cuando estos metadatos están presentes, el creador puede determinar cómo plataformas como Facebook o X (antes Twitter) tienen que mostrar su contenido. De no estar, Facebook o X igualmente podrían mostrar los contenidos, pero no necesariamente en la manera en que el creador lo desea. O sea, los metadatos sociales son una participación más del creador dejando un testimonio más claro de su intención, ofreciendo orientaciones de presentación y/o representación de los contenidos compartidos. No deja a la plataforma la libertad total de, a través de una heurística interna, mostrar e interpretar su contenido, sino que se podría decir que finaliza su creación, que hay una participación del creador como curador de su propio contenido.

El hecho de que estos metadatos sociales describan propiedades nos remite a la mención al inicio de este texto de que la preservación digital está basada en la identificación de las características esenciales de su objeto, entre todas sus propiedades. Estamos entonces ante información que podría ayudar al sistema de preservación a definir las características que podrían ser consideradas como esenciales o no, y ante la opinión o el criterio del creador acerca de algunas que podrían ser consideradas como tal.

Los metadatos sociales enclavan la interacción entre el objeto y su contexto de información, por una parte, y entre el creador y su contexto de creación, por otra. Esta contribución a la definición de los contextos del objeto y de cómo éste se debe comportar en esos

contextos contribuye también en gran medida a la comprensibilidad del contenido informativo, ya que añade información adicional que podría no ser explícita en el contenido explícito del objeto y que es más importante precisamente cuando no es explícita. Parte de la información que los usuarios pueden incorporar es el historial de cambios u otras anotaciones acerca de los mismos. Ello puede servir al sistema de preservación para documentar los eventos del devenir de un objeto digital y la participación de todas las personas involucradas en el ciclo de vida de un objeto. Los metadatos sociales tienen un enorme potencial en este sentido, para documentar aquellos eventos que ocurren mientras los objetos se encuentran fuera del sistema de preservación, lo que puede constituir una manera de aportar a la construcción de esta cadena de acontecimientos.

## *2. Aceptar la participación de otros actores en el sistema de preservación*

Cuando en la bibliografía se discute la primera de las tipologías de metadatos sociales presentadas, es decir aquellos metadatos generados dentro y gestionados por la institución de información, es lugar común remarcar la limitación de los sistemas clásicos de representación en instituciones de información, y de sus procesos de producción y reproducción de significados, en favor de la participación de comunidades o individuos históricamente marginalizados en la interpretación y producción de conocimiento acerca del conocimiento a través de los metadatos sociales. Para su encuesta, Smith-Yoshimura y Shein<sup>13</sup> categorizaron un conjunto de razones por las cuales las instituciones trabajan con los metadatos sociales, lo que podría ser muy alentador:

- a. Potenciación de los datos—Desarrollo de la descripción
- b. Potenciación de los datos—Desarrollo del acceso por materias

---

13 Smith-Yoshimura y Shein, *op. cit.*



**Metadatos sociales: iniciativas...**

- c. Construcción de colecciones y contenidos
- d. Ratings y revisiones—Opiniones subjetivas
- e. Compartir y facilitar la investigación
- f. Construcción de redes y comunidades
- g. Promoción de actividades fuera de sitio

Una de ellas, que se podría considerar como parte de la razón “a”, es lidiar con las cantidades enormes de producción de contenidos que necesitan ser representados, para los cuales no dan abasto los sistemas tradicionales de procesamiento y catalogación, con una cantidad limitada de recursos para dedicar al procesamiento de la información. Entre estos metadatos de representación se encuentran parte de lo que se considera Información de Representación para la Preservación, como la considera OAIS.<sup>14</sup>

Li Liew<sup>15</sup> resume que estas tecnologías brindan una posibilidad de “diálogo, negociación, creatividad, y valor añadido” respecto a la información, la infraestructura a través de la cual se podrían brindar perspectivas e interpretaciones paralelas a las de estos objetos. Podrían permitir la reducción del sesgo cultural a la hora de interpretar los contenidos, al ofrecer la visión e interpretación del otro, de algún modo haciendo coexistir la manera profesional de interpretar y representar con otras que en ocasiones podrían representar “lo popular”, “lo marginal”, “lo diferente”. Creo que no se trata de una diversificación de la función de la biblioteca, el archivo o el museo, sino de una diversificación de cómo estas instituciones pueden cumplir sus funciones y es a la vez un desafío a la manera en que las han cumplido históricamente: proponiendo a sus públicos una mirada autoritativa única sobre los recursos que presenta y representa.

Pero todo lo anterior es sólo una posibilidad de la tecnología, que también lleva implícita la posibilidad contraria, de censura y cancelación de esas voces, y de apropiación del conocimiento,

---

14 Smith-Yoshimura y Shein *op. cit.*

15 Liew Chern Li, “Social Metadata and Public-Contributed Contents in Memory Institutions”.

bajo el manto de lo experto curatorial. En ocasiones en la valoración sobre los metadatos sociales prima el hecho de que los profesionales de la información sienten que su autoridad es disminuida o que su conocimiento es devaluado con esa participación de las comunidades en la producción de metadatos.

La tecnología, cualquier tecnología, es intrínsecamente social, esto significa que lleva en sí condensadas las fuerzas sociales que atraviesan su creación y existencia. La propia tecnología no es determinante socialmente, por tanto; en sí misma no garantiza que estos procesos de participación se den. Son las instituciones de información las que pueden o no aprovechar las tecnologías para la producción de metadatos sociales para promover el vínculo con la comunidad y extender su alcance como espacio para la memoria social.

Hay que añadir, además, que, en el contexto actual, lo que sí está representando una diversificación de funciones es que plataformas de web 2.0 o redes sociales están desarrollando modos de preservar sus contenidos y ofreciendo herramientas de investigación, destinadas a prestar servicios a usuarios académicos e investigadores, para la consulta y uso de sus contenidos. Aunque suelen atribuirse a este cambio los objetivos de mantener el control máximo sobre la propiedad, esto es, los derechos sobre sus contenidos (lo que tiene de fondo en parte una cuestión económica), es también parte de un fenómeno social más amplio que se plantea para la información digital: la consideración de alternativas postcustodiales para el almacenamiento y acceso a la información. Se podría decir que de alguna manera los propios propietarios o creadores se convierten en una suerte de “biblioteca” o “archivo”. Es el caso, por ejemplo, de X, Facebook o Google. Lo que es muy relevante desde el punto de vista de la preservación, es que ésta no es su primera función ni obligatoria, es una función secundaria o lateral, y podrían simplemente dejar de cumplirla un día, o cumplirla con discrecionalidad. También es relevante que estos creadores o propietarios son “partes interesadas” y no poseen la neutralidad tradicional de bibliotecas y archivos que da garantía sobre la preservación de sus contenidos de manera íntegra y auténtica. En ese

panorama de diversificación, incluso estos medios sociales han reconocido de algún modo la función de instituciones de memoria respecto a la preservación digital de sus contenidos. Se conoce que X (antes Twitter), por ejemplo, realizó una donación de todo su contenido de tuits a la Biblioteca del Congreso de los Estados Unidos, y al Massachusetts Institute of Technology (MIT), aportando también una significativa cantidad de financiamiento a este lugar para investigar el mejor modo de preservar sus contenidos.<sup>16</sup>

Otro aspecto a tener en cuenta es que los metadatos sociales poseen gran potencialidad para establecer las tramas documentales, las tramas de información en las cuales son producidas actualmente las ideas o “verdades”, así como la participación de los públicos en la valoración de la información a preservar. Esto es relevante no sólo para del campo de la política o la cultura, lo que resulta evidente, sino también dentro del campo de la ciencia, en la que una diversidad considerable de actores puede converger y crear tales tramas de documentación e información, esto es, tramas de conocimiento, al fin y al cabo.

### *3. La construcción del conocimiento necesario para tratar esa realidad*

Pero no basta con identificar el objeto, y aceptar y gestionar la participación de otras personas en la construcción de sentido para los objetos digitales de la cultura; el tercer reto es que debemos tener la capacidad intelectual profesional para realizar los trabajos necesarios. Estos conocimientos se requieren de parte del personal especializado en las instituciones de la información y también de parte de usuarios o creadores.

Volvemos aquí al punto de partida recordando que la falta de capacidades en el equipo de trabajo es una de las razones fundamentales por las cuales algunas instituciones de información o memoria no trabajan con metadatos sociales. Reconocer la falta de

---

16 Véase, por ejemplo: Elisabeth Fondren y Meghan M. McCune, “Archiving and Preserving Social Media at the Library of Congress: Institutional and Cultural Challenges to Build a Twitter Archive”.

experiencia o conocimiento para lidiar con los metadatos sociales, con el fenómeno de la participación en la representación, con las diversidades, con el reconocimiento de formas alternativas de conocimiento y experticia, podría ser un buen punto de partida. Para el personal especializado en funciones de preservación es difícil encontrar posibilidades de formación en estos temas. La bibliografía, además de no ser abundante, adolece de detalles necesarios o imprescindibles para llevar proyectos a la práctica real. Y este es sólo un tipo más dentro del maremágnum de información digital que los preservadores deben trabajar, y que usualmente realizan ellos solos o entre muy pocas personas dentro de los equipos de trabajo de bibliotecas o archivos.

Parte de este reto es aceptar que el hecho de que se requieran metadatos para la preservación digital no significa que el propio sistema de preservación o el preservador en funciones debe crear los metadatos a la manera tradicional en que procesadores han creado históricamente los metadatos descriptivos. Más bien, se trata de un proceso de carácter continuo, gestionado desde el sistema de preservación, que debe comenzar por coleccionar metadatos desde antes del momento de la creación de los propios objetos. La información que se requiere para la preservación digital de un objeto suele poseer múltiples fuentes, con la participación de varias personas a lo largo de todo el ciclo de vida de la documentación, desde la creación de los contenidos. El conocimiento y la sensibilidad para que estas varias personas involucradas participen de la manera correcta es fundamental para el éxito de cualquier programa de preservación y lo es mucho más para el caso de la web 2.0 o los medios sociales.

En lo que corresponde a los creadores o usuarios, en algunos casos no se requiere de ninguna preparación o conocimientos, porque los sistemas resultan intuitivos, fáciles y están preparados para esa participación. Pero en otros casos se requiere de estos usuarios formación y conocimientos, en particular para el uso de medios sociales para la investigación académica, y mucho más si se desea utilizar sus metadatos sociales. Hay plataformas que requieren de conocimientos altamente especializados como son el

uso, instalación o creación de códigos en Python u otros lenguajes de programación para coleccionar la información necesaria, incluyendo sus metadatos sociales o la gestión de las mencionadas APIS. Este es un tema que podría incorporarse en los programas de Alfabetización Informacional, o en los programas de formación de los investigadores y estudiantes en Gestión de Datos de Investigación.

En tanto los contenidos de los medios sociales y la web 2.0 son considerados cada vez más valiosas fuentes de información académica, se requieren una o varias de estas tres opciones: 1) que los investigadores mismos obtengan el conocimiento necesario para de manera independiente coleccionar y gestionar de manera adecuada esta clase de contenido, 2) que exista personal con el conocimiento necesario para que los estudiantes e investigadores reciban la asistencia o el servicio de sus sistemas bibliotecarios o archivísticos, a la manera de servicios puntuales o del “bibliotecario embebido”, por ejemplo, o 3) que estos reciban el ya mencionado servicio directo de las plataformas de web 2.0, en cuyo caso se requeriría a la par el conocimiento para canalizar estas solicitudes e interpretar los resultados tales como estos les son entregados. En muchos casos los propios usuarios o creadores no reconocen la necesidad de estos conocimientos y se limitan a coleccionar y utilizar los mensajes de modo tal que no es posible con el tiempo considerar que estos son fuentes confiables de información para la investigación o la ciencia.

La solución tecnológica más recomendada para coleccionar web 2.0 o medios sociales es la descarga mediante APIS. Esta opción suele ser la mejor para redes como X o Facebook, quienes ofrecen ellos mismos las APIS para acceder a sus contenidos con un objetivo investigativo. Pero no siempre es el caso de que existan ya APIS para la recolección de los datos junto a la totalidad o parte suficiente de sus metadatos, y es posible que sea necesario desarrollar localmente la infraestructura necesaria, o utilizar desarrollos o aplicaciones de terceros. Al menos en América Latina y el Caribe esto representa un importante obstáculo, porque la mayoría de las instituciones de memoria no cuentan con personal especializado que pueda realizar estas funciones de desarrollo o adaptación de herramientas.

Es necesario también difundir la importancia que tiene incluir metadatos sociales en los productos web, de modo que estos sean interpretados, visualizados y circulen de la manera en que ellos desean.

#### *4. El desarrollo de una infraestructura adecuada de herramientas y/o sistemas*

El trabajo con metadatos sociales depende de que los sistemas estén preparados para estos. Un importante matiz es que no existe una única solución válida para lidiar con todos los casos de web 2.0 y metadatos sociales. Casi cada plataforma de datos, con sus metadatos sociales, va a requerir soluciones específicas en sus detalles, si bien para todas ellas se cumple el mismo principio general de preservación: que no es posible preservar adecuadamente estos medios a través de capturas de pantalla e impresiones en .pdf o sobre papel de los contenidos o mensajes, porque se deja fuera una cantidad significativa de datos relacionados con el contexto de esa documentación, entre ellos, por supuesto, todos o parte de los metadatos sociales. Se necesitará “capturar los datos y sus contextos afiliados a gran escala, en lugar de copias o *snapshots*”.<sup>17</sup> Como ya se ha dicho, una de las soluciones más adecuadas para esto es el uso de APIS de las plataformas, desarrolladas por el sistema de información o por terceros.

En cuanto al almacenamiento, la mayoría de las plataformas de medios sociales limita el hecho de que sus contenidos, una vez en la institución de memoria, puedan ser almacenados en nubes, toda vez que se trata generalmente de transferir datos a una tercera persona moral. Así que la infraestructura de almacenamiento para preservación digital de estos contenidos suele tener que ser local, lo que en algunos casos podría aumentar los costos de los proyectos de preservación digital. Esta situación refuerza la importancia de emprender estos retos desde proyectos colaborativos donde puedan compartirse los gastos en infraestructura de todo tipo y

---

<sup>17</sup> Sara Day Thomson, *op. cit.*, 2.

aumentar la velocidad en los procesos de desarrollo de las herramientas necesarias al “multiplicar” a través de la colaboración, el alcance del grupo de trabajo.

En las organizaciones de información es necesario poder acomodar nuevos flujos de trabajo para nuevos flujos de información, desarrollando o aplicando las herramientas para ello, por ejemplo, las herramientas para que los usuarios agreguen metadatos sociales en los instrumentos de recuperación de la información en bibliotecas y archivos. Aunque en un sentido esto aumenta la carga, en algunos casos puede ser considerado como una inversión en función de la sostenibilidad de un sistema, con el objetivo de aumentar la cantidad de recursos disponibles y mejorar la representación de estos, entre otras cosas. La contribución pública y la creación de metadatos sociales puede realizarse en un variado espectro de posibilidades técnicas. El grado de conocimiento que se espera del público, el usuario, la comunidad, es algo a tener muy en cuenta cuando se diseñan los sistemas de información en los que se va a trabajar con metadatos sociales. Por ejemplo, es necesario que aquellos de autodepósito en los repositorios digitales, se hagan de modo tal que la comunidad pueda participar o contribuir, con la necesidad de variar en lo mínimo posible su base de conocimiento, es decir, sin convertirlos en especialistas.

##### *5. Las cuestiones éticas y los derechos involucrados*

La relación entre las instituciones de memoria y las comunidades que contribuyen con metadatos sociales en sus sistemas debe construirse como una relación de mutua confianza; el público debe saber que en muchos casos sus aportes son generalmente moderados o editados, transformados de algún modo y, por tanto, este proceso en la institución de información debe ser transparente. La institución debe poder ofrecer públicamente sus políticas y procedimientos en cuanto a estas prácticas. En muchos casos los sistemas de las instituciones de información solicitan a los usuarios que quienes contribuyen con metadatos sociales o contenidos en general lean, conozcan y acepten declaraciones formales de términos y condiciones. Lejos de molestar a los usuarios, la

existencia de este tipo de declaraciones ayuda a las instituciones a construir confianza por parte de sus comunidades.

Por otra parte, el sistema de información debe romper el miedo a que los contenidos públicamente contribuidos o metadatos sociales no brinden representaciones adecuadas de los recursos, que estén “equivocados”. Debe considerar, lo que significa la confiabilidad de la información, de la que ya hemos hablado, en cuanto a los metadatos sociales incorporados en su propio sistema, el alcance y limitaciones que el aporte de los metadatos sociales incorpora; y en cuanto a los metadatos sociales que los propios creadores habían incorporado en contenidos web coleccionados por la institución de memoria, el sistema de información debe considerar las especificaciones establecidas por estos para la representación de sus contenidos.

Entre todos los retos mencionados tal vez el más grande para la preservación de medios sociales y sus metadatos sociales, es el hecho de que la mayoría de las plataformas que los incorporan presentan también complejas declaraciones de términos y condiciones. Estas compañías suelen establecer en sus contratos limitaciones sobre qué puede ocurrir con los datos una vez que estos se encuentran ya en las manos de otros, tales como instituciones de la memoria o investigadores, y entre estas limitaciones suele estar el hecho de compartir los datos en repositorios. Los acuerdos de no transferibilidad alcanzan, como ya hemos mencionado, a los sistemas de almacenamiento para la preservación de los contenidos obtenidos. Adicionalmente, los mencionados términos y condiciones legales para el trabajo con estas colecciones cambian con cierta frecuencia, de modo que es difícil mantenerse al tanto de las implicaciones de cada uno de estos cambios, y es necesario considerar que muchas veces estos contenidos poseen otros contenidos embebidos sujetos a derechos de autor u otros derechos de propiedad intelectual que no pertenecen a la plataforma como tal.

Las instituciones de la memoria, que tendrían entre sus fines preservarlos, acceden a los datos producidos en estos medios en igualdad de condiciones que las compañías que lucran con el análisis de estos, y tienen que respetar los términos y condiciones



que estas empresas establecen para evitar la pérdida del control de los datos y poder garantizar para sí mismos su uso comercial y seguridad.

Para mayor contraste, en primer lugar, la mayoría de las instituciones de memoria no cuentan en su personal con abogados o especialistas en derechos relacionados con la documentación digital, y en segundo lugar nuestros países de América Latina y el Caribe no poseen sistemas jurídicos maduros para lidiar con los detalles de la información digital y no están establecidas todas las políticas públicas necesarias. Ante esta realidad, la cooperación y las alianzas vuelven a resultar herramientas muy valiosas para avanzar aún con tantos espacios en blanco. Como matiz específico, se requiere que nuestros países desarrollen el marco jurídico y en general, normativo, que considera a la preservación de su patrimonio digital en todas sus manifestaciones antes de que sea demasiado tarde y que este marco no reproduzca como ideas preconcebidas las formas tradicionales con las que hemos lidiado con el patrimonio en otros soportes y formatos.

Cuando se utilizan las APIS de las propias plataformas para la colección de los contenidos, es usual que estas limiten además la cantidad total de datos a los que se puede acceder y la cantidad de metadatos sobre un objeto que pueden obtenerse. Para el sistema de preservación esto significa que podría no tener los datos suficientes para la representación de determinadas materias o asuntos o incluso para preservar los objetos de manera íntegra. Los investigadores, por otra parte, también podrían necesitar cantidades mayores de ambas cosas. Ante las solicitudes de información, los medios sociales suelen responder utilizando algoritmos propios que no comparten o hacen públicos y, por tanto, los sistemas de preservación o los investigadores no reciben resultados que sean verificables.

Este acercamiento al tema de los metadatos sociales nos permite concluir que estos sí resultan relevantes para la preservación digital al menos en dos sentidos: como blancos de la preservación ellos mismos, y como apoyo fundamental para la preservación de los contenidos digitales a los que ellos se refieren, o sea de los

datos de los que ellos resultan ser metadatos. En este último caso, los metadatos sociales se podrían considerar elementos que son parte de la integridad de los objetos a preservar y podrían utilizarse de manera planificada con fines de preservación, y que se incluya esta perspectiva en la gestión de los mismos desde la creación de los propios objetos digitales de la web 2.0.

El conocimiento y las herramientas para trabajar con los metadatos sociales en los medios sociales o la web 2.0 deben ser desarrollados y compartidos en todo lo posible. Hasta el momento no es abundante la información con la que se cuenta sobre proyectos a cuyas metodologías y conclusiones es posible tener acceso, o los debates teóricos sobre temas relacionados. Las herramientas para el trabajo local en la preservación de la web 2.0 o los medios sociales que incluyen metadatos sociales no están ampliamente disponibles más allá de las propias plataformas propietarias de los contenidos y de los propios metadatos por extensión.

El desarrollo en el conocimiento y las herramientas de trabajo debe complementarse con el desarrollo de los marcos jurídicos y las políticas públicas para hacer posible la preservación digital de los medios sociales, incluso pueden considerarlos parte del patrimonio nacional y local. Los intereses por la preservación de la cultura actual en nuestras sociedades deben ser variables importantes a considerar en la preservación de estos contenidos, los cuales hasta ahora enfrentan una significativa cantidad de limitaciones para ser coleccionados e incluso, estudiados e interpretados como parte de una producción que tiene un contexto más amplio que el de las propias plataformas propietarias, o de las herramientas tecnológicas en los que se producen y gestionan. La alternativa de la preservación de estos contenidos en las manos de los propietarios o creadores, o sea, la perspectiva postcustodial para su preservación, contiene en sí retos que deben ser resueltos aún desde el punto de vista técnico, pero también desde el punto de vista jurídico.

## CONCLUSIONES

Son varios los retos y dificultades presentadas, no obstante, el emprendimiento de esta clase de trabajo es a mediano y largo plazo inevitable en nuestras instituciones de memoria, en tanto que estos contenidos constituyen fuentes ya casi consolidadas para los estudios académicos, para el sistema jurídico, y objetos con interés para la cultura de individuos, grupos y sociedades. En la medida en que aumenta la comunicación humana con medios sociales o web 2.0 que incluyen metadatos sociales, en esa misma medida aumentará su importancia e interés para las instituciones de memoria e información, y para la sociedad como un todo a las que estas deben su servicio.

La cooperación en redes de trabajo, que incluso podrían incluir personas naturales y jurídicas de muy diverso orden, representa una estrategia importante para lidiar con varios de los retos presentados. La alianza entre plataformas privadas, instituciones públicas y grupos de individuos con determinados intereses podría ser un camino importante a explorar, para dar continuidad a la experiencia lograda en iniciativas puntuales tales como la experiencia de X (antes Twitter) con la Biblioteca del Congreso de los Estados Unidos, algunos archivos comunitarios o de grupos minoritarios, y otros casos muy específicos y limitados en número y difusión de este tipo de alianzas.

## REFERENCIAS BIBLIOGRÁFICAS

“4 Tips For Collecting And Preserving Social Media Evidence”. En <https://www.logikcull.com/blog/4-tips-for-collecting-and-preserving-social-media-evidence> (consultado el 16 de mayo de 2022).

Begum, R. A. *Preserving social media: A Case Study at IITU Archives*. Luleå: Luleå University of Technology. 2015.

- Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data System Standards (Recommended Practice, Issue 2, CCSDS 650.0-M-2 Magenta Book)*. CCSDS Secretariat. Program Integration Division (Code M-3). Washington: National Aeronautics and Space Administration. 2012.
- Day Thomson, Sara. *Preserving social media*. Great Britain: Digital Preservation Coalition. 2016.
- Day Thomson, Sara y William Kilbride. "Preserving social media: The Problem of Access". *New Review of Information Networking* 20 (1-2: "Shaping our Legacy: Preserving the Social and Cultural Record"): 261-275.
- Facebook. "A Guide to Sharing for Webmasters". <https://developers.facebook.com/docs/sharing/webmasters> (Consultado el 19 de mayo de 2022).
- Financial Industry Regulatory Authority (FINRA). "Rules". En <https://www.finra.org/rules-guidance/rulebooks/finra-rules> (consultado el 16 de mayo de 2022).
- Fondren, Elisabeth y Meghan Menard McCune. "Archiving and Preserving Social Media at the Library of Congress: Institutional and Cultural Challenges to Build a Twitter Archive". *Preservation, Digital Technology and Culture* 47(2): 33-44. 2018.
- Green, Ann, JoAnn Dionne y Martin Dennis. *Preserving the Whole: a Two-Track Approach to Rescuing Social Science Data and Metadata*. Washington, DC: Digital Library Federation. 1999.
- Hockx-Yu, Helen. "Archiving social media in the Context of Non-Print Legal Deposit". *IFLA WLIC Libraries, Citizens, Societies: Confluence for Knowledge in Lyon*. 2014. <https://library.ifla.org/id/eprint/999/1/107-hockxyu-en.pdf>.

***Metadatos sociales: iniciativas...***

- Jones, Shawn M., Valentina Neblitt-Jones, Michele C. Weigle, Martin Klein y Michael L. Nelson. "It's All About the Cards: Sharing on social media Probably Encouraged HTML Metadata Growth". *Proceedings of the 2021 ACM/IEEE Joint Conference on Digital Libraries*, 110-119. 2021.
- Liew, Chern Li. «Social Metadata and Public-Contributed Contents in Memory Institutions: "Crowd Voice" Versus "Authenticated Heritage"?». *Preservation, Digital Technology and Culture* 45(3): 122-133.
- Madhava, Rakesh. "10 Things to Know About Preserving social media". *Information Management* 45(5): 33-37, 54.
- Mayernik, Matthew S. *Metadata Realities for Cyberinfrastructure: Data Authors as Metadata Creators*. Los Angeles: University of California. 2011.
- Skageby, Jörgen. *Gifting Technologies: Ethnographic Studies of End-Users and Social Media Sharing*. Stockholm: Stockholm University. 2008.
- . "Exploring Qualitative Sharing Practices of Social Metadata: Expanding the Attention Economy". *The Information Society: An International Journal* 25(1): 60-72.
- Smith, Gene. *Tagging: People-Powered Metadata for the Social Web*. Thousand Oaks, California: New Riders. 2007.
- Smith-Yoshimura, Karen y Cindy Shein. *Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews*. Dublin, Ohio: OCLC Online Computer Library Center. 2011.
- , Carol Jean Godby, Helice Koffler, Ken Varnum, y Elizabeth Yakel. *Social Metadata for Libraries, Archives and Museums. Part 2: Survey Analysis*. Dublin, Ohio: OCLC Online Computer Library Center. 2011.

- , y Rose Holley. *Social Metadata for Libraries, Archives and Museums. Part 3: Recommendations and Readings*. Dublin, Ohio: OCLC Online Computer Library Center. 2012.
- Spiteri, Louise F. “Using Social Discovery Systems to Leverage User Generated Metadata”. *Bulletin of the American Society for Information Science and Technology* 37(4): 27-29.
- “To Preserve or Not to Preserve: social media”. En <https://siarchives.si.edu/blog/preserve-or-not-preserve-social-media> (consultado el 16 de mayo de 2022).
- Zarro, Michael A., y Robert B. Allen. “User-Contributed Descriptive Metadata for Libraries and Cultural Institutions”. En *Research and Advanced Technology for Digital Libraries. ECDL 2010*, editado por Mounia Lalmas, Joemon Jose, Andreas Rauber, Fabrizio Sebastiani e Ingo Frommholz. Berlin: Springer.

# Los retos de los recursos educativos abiertos y su catalogación: creación de metadatos profesionales y sociales

ALMA BEATRIZ RIVERA AGUILERA

ELISA CRUZ ROJAS

MARÍA GUADALUPE BARRERA GALÁN

*Universidad Iberoamericana Ciudad de México, México*

## INTRODUCCIÓN

Los Recursos Educativos Abiertos (REA) de acuerdo con la UNESCO son materiales didácticos, de aprendizaje o investigación, usualmente digitales, que se encuentran en el dominio público o que se publican con licencias de propiedad intelectual que facilitan su uso, adaptación y distribución gratuitos. Los REA representan un reto especialmente complicado para su organización por la gran diversidad de formatos y tipos en los que se plasman. A lo anterior se añade que dentro de las instituciones educativas no necesariamente existen políticas para almacenar, organizar, solicitar y valorar la producción didáctica de los académicos. Otro de los retos es asegurar la interoperabilidad de estos<sup>1</sup> y algunos investigadores se han acercado a la problemática de

---

1 Dora María Tovar Gutiérrez, Alejandro López Ibarra y María Soledad Ramírez Montoya, “Estrategias de comunicación para potenciar el uso de Recursos Educativos Abiertos (REA)...”.

la calidad de los recursos mismos desde la visión no sólo de los creadores sino también de los estudiantes.<sup>2</sup>

En los últimos quince años la Universidad Iberoamericana, como parte de las actividades de excelencia académica, ha generado una gran cantidad de material didáctico digital que potencialmente puede convertirse en REA. La Biblioteca Francisco Xavier Clavigero (BFXC) de la Universidad Iberoamericana Ciudad de México, desde el 2008 ha observado la producción y reutilización de los materiales didácticos digitales de las diversas entidades de la Universidad y ha desarrollado perfiles de aplicación de metadatos para su tratamiento. Este fenómeno ha sido similar en otras instituciones educativas como las españolas reportadas por Colome<sup>3</sup> y en Cuba por Rodríguez Delís, Campaña Jiménez y Gallego Arrufat.<sup>4</sup>

A raíz de la pandemia por COVID-19, muchas instituciones de educación superior trasladaron sus actividades educativas al medio digital y por ende se generaron muchos más contenidos educativos digitales que en años anteriores. A partir de 2020 la BFXC, como parte de sus funciones, ha ponderado la tarea de organizar y preservar los contenidos de este tipo de recursos y ponerlos a disposición de sus usuarios a través del Repositorio Institucional. Los repositorios de recursos educativos tienen una tradición que data de varios años atrás,<sup>5</sup> aunque el tema se sigue tratando más que nunca en años recientes, como lo reportado a manera de este tipo de esfuerzos por Sherriff, Benson y Atwood, en 2019.<sup>6</sup>

---

2 Aldo Gordillo *et al.*, “Estimación de calidad de objetos de aprendizaje en repositorios de recursos educativos abiertos...”.

3 Dunia Colome, “Objetos de Aprendizaje y Recursos Educativos Abiertos en educación superior”.

4 Yomé Rodríguez *et al.*, “Iniciativas para la adopción y uso de recursos educativos abiertos en Instituciones de Educación Superior”.

5 Javiera Atenas-Rivera *et al.*, “Repositorios de recursos educativos abiertos”.

6 Graham Sherriff *et al.*, “Practices, policies and problems in the management of learning...”.



Como respuesta al incremento de los REA y otros contenidos digitales, las áreas de catalogación se han involucrado cada vez más con la asignación de metadatos de recursos digitales<sup>7</sup> y, en especial, con recursos educativos abiertos.<sup>8</sup> Desde 2008 en la BFXC se iniciaron las experiencias de descripción de materiales didácticos digitales tal como se registra en Rivera Aguilera, Cruz Rojas, Barrera Galán, Pozo Marrero.<sup>9</sup> Sin embargo, las prácticas de utilizar recursos digitales en clase no eran muy comunes en nuestra institución en la primera década del siglo XXI. Es a partir de la cuarentena por COVID-19 que este tipo de material se produce en mayor cantidad y se involucra a la biblioteca en su organización. Un aspecto importante por señalar es que constantemente existen nuevas versiones de material didáctico,<sup>10</sup> en especial en el caso de los webinars generados por la biblioteca y en las áreas de formación de profesores, por ejemplo, con recursos de apoyo a la docencia como las bases datos especializadas y los softwares educativos. Esta situación implica la definición de una política que establezca si se genera un registro por cada versión o un solo registro con diversas versiones asociadas como archivos.<sup>11</sup> En el siguiente apartado se comparte la experiencia en los últimos dos años con los REA integrados al repositorio institucional que gestiona nuestra biblioteca.

Con respecto a los metadatos sociales, estos son definidos por Karen Smith-Yoshimura y Cyndi Shein como la “información adicional sobre un recurso, que resulta de la contribución de los usuarios y de la actividad online, tal como etiquetado, comentarios, reseñas, imágenes, videos, ratings, recomendaciones, etcétera,

---

7 Lisa Furubotten, “Informe desde las trincheras: transformando unidades de catalogación para incluir servicios de repositorio”.

8 Gema Santos-Hermosa *et al.*, “Recursos educativos abiertos: repositorios y uso”.

9 Alma Beatriz Rivera *et al.*, “Propuesta de un perfil de aplicación para material didáctico digital basada en el Marco de Singapur”.

10 Alma Beatriz Rivera Aguilera, “La concepción didáctica del docente y los materiales didácticos...”.

11 Isabel Pérez-Ortega, “Creación de Recursos Educativos Digitales: Reflexiones sobre innovación educativa con TIC”, 262.

que ayuda a las personas a encontrar, comprender o evaluar el contenido”.<sup>12</sup>

En relación con los metadatos sociales, los REA representan una gran oportunidad de involucrar a los creadores en la asignación de los metadatos y a los usuarios en hacer explícita su opinión sobre los mismos a través de los metadatos de tipo comentario o valoración.<sup>13</sup> La experiencia de nuestra biblioteca a nivel de metadatos de creadores inició con las tesis digitales, en 2004, fecha desde la cual se solicita a los tesisistas datos relacionados con sus perfiles de interés al momento de entregar su tesis. Dichos metadatos se refieren a otros grados académicos previos, áreas de interés y experiencia docente. En la actualidad nuestro sistema Koha contiene más de 1000 registros con dichos metadatos, los cuales en un inicio fueron pensados para generar una red de temas de interés que no se llegó a implementar. Lo anterior era una idea original hace casi 20 años cuando todavía no existía Academia.edu ni ResearchGate, que hoy en día cubren con creces el desarrollo de redes académicas internacionales.

## MATERIALES Y MÉTODOS

Dada la diversidad y complejidad de los materiales didácticos digitales, y su integración al repositorio institucional el equipo de trabajo (compuesto por varios bibliotecarios del área de Organización de la Información y Repositorio Institucional), se llevaron a cabo varias fases de catalogación.

La **fase uno** inició con una pequeña muestra generada al inicio de la pandemia por COVID-19, con diferentes tipos y formatos de material educativo para profesores y alumnos universitarios (videos, infografías, textos en pdf, webinar, etcétera), elaborados por la biblioteca, la Dirección de Desarrollo y Acompañamiento

---

12 Karen Smith-Yoshimura y Cyndi Shein, “Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews”.

13 Alma B. Rivera Aguilera *et al.*, *op. cit.*, 2011.

Educativa (DiDAE) y la Dirección de la Enseñanza y Aprendizaje Mediados por Tecnología (DEAMeT). Aunado a ello, se sumó una solicitud en 2020 de integrar a nuestro repositorio institucional videos de talleres, cursos, clases y conferencias de profesores visitantes por parte de la Dirección de Cooperación Académica (DICA), en específico con el Programa Institucional de Académicos Visitantes (PIAV). Con el afán de tener una descripción más útil para el usuario, una bibliotecaria especializada en temas educativos y sistemas de información hizo recomendaciones de mejora e incluyó recomendaciones en el uso de metadatos especializados.

Posteriormente, en la **fase dos** se reconsideraron las etiquetas iniciales y los despliegues para diseñar una descripción bibliográfica más allá de las prácticas tradicionales, en donde se tomaron en cuenta las expectativas de docentes y alumnos en relación con la recuperación de este material educativo. En esta segunda fase se revisó un ejercicio llevado a cabo en 2009-2010,<sup>14</sup> se consultó literatura actualizada y se recuperó un análisis de los distintos esquemas de metadatos dedicados específicamente a recursos educativos abiertos,<sup>15</sup> con un estudio de esquemas como Dublin Core (DC) y Learning Object Metadata (LOM-ES),<sup>16</sup> cuya base es el estándar original en inglés de 2002 revisado en 2009,<sup>17</sup> a partir de lo cual se incluyó a la lista de etiquetas descriptivas de DC el metadato especializado LOM 5.2 Tipo de Recurso Educativo en el contexto de las posibilidades de la gestión en DSpace.

En la selección de la lista de metadatos descriptivos a utilizar, se buscó sobre todo que los registros fueran transferibles e interoperables. Los metadatos descriptivos apoyan la promoción, disponibilidad y uso de las prácticas educativas, a través de los

---

14 Alma B. Rivera *et al.*, *ibid.*

15 Alma Rivera Aguilera, "Metadatos de contenidos educativos: problemáticas y tendencias".

16 <http://educalab.es/documents/10180/40863/1LOM-ES.pdf/67a11fe2-edc0-487f-b6d5-6a87dc258668> <https://www.dublincore.org/>

17 IEEE. *IEEE Standard for Learning Object Metadata.*

### **Metadatos sociales: iniciativas...**

servicios generados en los repositorios.<sup>18</sup> Por ello, se revisaron los esquemas mencionados en el párrafo anterior y alguna literatura relacionada con experiencias de varios países.<sup>19</sup>

En lo referente a la plantilla DSpace para el tratamiento de los recursos educativos abiertos se elaboró una propuesta, que, además de los metadatos que aplican para cualquier tipo de material en un repositorio, incluía metadatos relevantes y específicos para este tipo de recurso en nuestra institución basado en el dominio Dublin Core. La lista de metadatos específicos para añadir a los REA más allá de la descripción genérica se muestra a continuación (tabla 1):

**Tabla 1.** Muestra la información descriptiva y metadatos adecuados para los REA.

Información descriptiva	Metadato
Tipo de recurso educativo. Basado en el vocabulario LOM-ES, metadato 5.2	dc.type.rea
Nombre de la institución asociada al autor para el caso de académicos visitantes	dc.contributor
Programa de estudio asociado	dc.contributor.other
Asignatura asociada	dc.relation.ispartof
Es parte de. Para el caso de los recursos que incluyen varias partes y/o secciones, o bien, cuando forman parte de una serie	dc.relation.ispartofseries
Categoría de calidad de los REA	dc.description

Fuente: Elaboración propia.

Se muestran los metadatos específicos para la descripción de los REA, los cuales, como se ha mencionado en las fases, provienen de una revisión específica para este tipo de recursos. A continuación, se muestra un ejemplo de despliegue sencillo y completo de un registro en el Repositorio Institucional de la Universidad

---


18 Dora M. Tovar *et al.*, *op. cit.*


19 Erla Mariela Morales-Morgado *et al.*, “Guía de metadatos educativos para agregar Recursos Educativos Abiertos (REA) en repositorios basados en DSpace”.

Iberoamericana Ciudad de México (<http://ri.iberomx/handle/iberomx/3628>):

Figura 1. Despliegue de registro sencillo de un REA del Repositorio Institucional

## Pensando el capitalismo: Karl Polanyi, la escuela de Fráncfort y Hannah Arendt 1 [video]




**Ver/**  
 RSD\_Curso\_01.pdf (294.1Kb)

**Ver texto completo**  
Enlace 1 al recurso

**Fecha**  
2020

**Autor**  
Rimoch Saraba, David  
Universidad de Pennsylvania,  
Universidad de Columbia, London  
School of Economics and Political  
Science

**Metadatos**  
Mostrar el registro completo del ítem



Clase 1: Introducción & la gran transformación I

**URI**  
<http://ri.iberomx/handle/iberomx/3628>

**Area de conocimiento**  
HUMANIDADES Y CIENCIAS DE LA CONDUCTA

**Recurso abierto educativo**  
Curso

**Colecciones**  
Recursos Educativos Abiertos [30]

Excepto si se señala otra cosa, la licencia del ítem se describe como <http://creativecommons.org/licenses/by-nc-nd/4.0>

Fuente: Elaboración propia.

La figura 1 muestra el despliegue breve de un registro de un REA que pertenece al Programa Institucional de Académicos Visitantes (PIAV). En éste se aprecia el tipo de recurso, el creador, la institución de procedencia, una portada del video, así como los enlaces al recurso.

## Metadatos sociales: iniciativas...

Figura 2. Despliegue de registro completo de un REA del Repositorio Institucional

dc.contributor.author	Rimoch Saraba, David
dc.contributor.author	Universidad de Pennsylvania, Universidad de Columbia, London School of Economics and Political Science
dc.contributor.other	Departamento de Historia, Programa de Cultura Judaica
dc.creator	Rimoch Saraba, David
dc.date.accessioned	2020-11-19T21:53:51Z
dc.date.available	2020-11-19T21:53:51Z
dc.date.issued	2020
dc.identifier.uri	<a href="http://ri.iberomx.com/handle/iberomx/3628">http://ri.iberomx.com/handle/iberomx/3628</a>
dc.description.abstract	Clase 1: Introducción & la gran transformación I
dc.description.provenance	Submitted by Benjamin Olvera (benjamin.olvera@iberomx.com) on 2020-11-19T21:53:51Z No. of bitstreams: 1 Pensando el capitalismo.pdf: 301180 bytes, checksum: da3ca4ce1178bf679f88c8732265bfb8 (MD5)
dc.description.provenance	Made available in DSpace on 2020-11-19T21:53:51Z (GMT). No. of bitstreams: 1 Pensando el capitalismo.pdf: 301180 bytes, checksum: da3ca4ce1178bf679f88c8732265bfb8 (MD5) Previous issue date: 2020
dc.format	mov
dc.language.iso	spa
dc.publisher	Universidad Iberoamericana, Dirección de Cooperación Académica
dc.relation.ispartofseries	Pensando el capitalismo
dc.relation.uri	<a href="https://iberomx-my.sharepoint.com/:v/g/personal/repositorio_piav_iberomx/ETSSoheXzfiNkXTKDfI3A4MBolioV6F1U7Y8vYQWILomYw?e=HO7POY">https://iberomx-my.sharepoint.com/:v/g/personal/repositorio_piav_iberomx/ETSSoheXzfiNkXTKDfI3A4MBolioV6F1U7Y8vYQWILomYw?e=HO7POY</a>
dc.rights.uri	<a href="http://creativecommons.org/licenses/by-nc-nd/4.0">http://creativecommons.org/licenses/by-nc-nd/4.0</a>
dc.subject.classification	HUMANIDADES Y CIENCIAS DE LA CONDUCTA
dc.title	Pensando el capitalismo: Karl Polanyi, la escuela de Fráncfort y Hannah Arendt 1 [video]
dc.title.alternative	Introducción & la gran transformación I [video]
dc.identifier	4
dc.rights.access	openAccess
dc.type.rea	Curso

Fuente: Elaboración propia.

La figura 2 muestra el registro con todos los metadatos que se han diseñado para cualquier registro que es albergado en el Repositorio

Institucional. En este caso se marcan en verde los metadatos específicos para un REA. Se incluyen los metadatos para registrar la institución de procedencia, el programa de estudio asociado, y para registrar el caso de los recursos que incluyen varias partes y/o secciones, o bien, cuando forman parte de una serie.

También se propusieron algunas reglas de contenido de los metadatos como fue el agregar a los títulos (dc.title) una designación general del material entre corchetes, para una pronta identificación por parte de los usuarios del tipo de recurso recuperado.

Se elaboró un manual para la creación de metadatos en formato excel a ser llenado por los creadores de los contenidos, ya que se visualizaba que los contenidos didácticos podrían contener elementos descriptivos que enriquecieran los registros integrados en el repositorio. Este manual se piloteó en la fase siguiente. Es importante mencionar que los metadatos propuestos para los REA están en constante revisión y ajuste para una descripción, recuperación y visualización por parte de los diversos actores involucrados en la gestión de las colecciones de los REA: usuarios, creadores, catalogadores y responsables de la difusión de los contenidos educativos de las diferentes áreas académicas.

En la **fase tres**, que inició en verano 2021, se recopiló mayor cantidad de materiales educativos de las áreas ya mencionadas que generan recursos didácticos para la formación de docentes y de alumnos. Se identificó que el manual es valioso, pero no es suficiente para que los autores elaboren los metadatos de acuerdo con las especificaciones, por lo que es necesario diseñar una capacitación tipo taller en donde se aclaren las dudas. Dicha capacitación se orientará a la descripción de metadatos sociales específicos para los REA, es decir, elaborados por autores no bibliotecarios relacionados con la autoría, edición, etcétera, que tengan la calidad que asegure la recuperación de los contenidos educativos.

La **fase cuatro** de este proceso incluyó el análisis de la propuesta de etiquetas y la revisión de los registros iniciales, para lograr una normalización de los registros de este piloto, que establezcan la base para la integración futura de los REA al RI. Esto será posible a partir de la invitación a los departamentos académicos

### *Metadatos sociales: iniciativas...*

del envío de sus recursos a la biblioteca para su ingesta masiva al Repositorio.

## ANÁLISIS DE LOS REGISTROS

El proceso del análisis inicia con la identificación de los Recursos Educativos Abiertos (REA) en el RI y su revisión a nivel de metadatos, lo cual fue llevado a cabo durante el mes de febrero 2022 y actualizado en mayo del mismo año.

El análisis parte del total de 59 registros de REA (46 en febrero y 13 en mayo de 2022) almacenados en el gestor de contenidos digitales de nuestra institución, que es DSpace. Dichos materiales fueron generados por cuatro entidades de la universidad identificados como comunidades en el contexto de DSpace:

- Biblioteca Francisco Xavier Clavigero (BFXC) - 11 recursos.
- Dirección de Cooperación Académica (DICA), en específico con el Programa Institucional de Académicos Visitantes (PIAV) - 30 recursos.
- Dirección de Desarrollo y Acompañamiento Educativa (Di-DAE) - 9 recursos.
- Dirección de Enseñanza y Aprendizaje Mediados por Tecnología (DEAMeT) - 9 recursos.

Actualmente todos los recursos del Repositorio son de acceso abierto y contienen el metadato de permisos indicando la licencia Creative Commons BY-NC-ND <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es> (Atribución/no comercial/sin derivados), esto implica que los usuarios pueden bajar el material y redistribuirlo si lo desean, siempre y cuando citen al autor, no comercialicen con el contenido y no lo modifiquen.



## Agrupaciones basadas en metadatos

En las siguientes tablas se muestran los resultados de la revisión de los registros en cuanto a tipo (tabla 2), formatos (tabla 3), entidades creadoras (tabla 4) de los REA analizados.

**Tabla 2.** Se muestra el tipo de documento educativo de REA, de acuerdo con LOM, por entidad académica de la universidad incluidas en el repositorio hasta la fecha del análisis. El término comunidad corresponde a la forma que DSpace agrupa colecciones.

Tipo de recurso educativo abierto (REA)	Comunidad BFXC	Comunidad DICA	Comunidad DiDAE	Comunidad DEAMeT	Totales por tipo de recurso educativo
Tutorial	8			1	9
Presentación multimedia	1			1	2
Seminario web	1				1
Instructivo	1				1
Conferencia		21			21
Taller		3			3
Curso		6			6
Sugerencias para docentes			5	3	8
Infografía			4	2	6
Guía				2	2
Total de REA en RI por comunidad	11	30	9	9	59

Fuente: Elaboración propia.

La Dirección de Cooperación Académica (DICA), a través del Programa Institucional de Académicos Visitantes (PIAV) es quien más recursos ha aportado. El tipo de recurso educativo de mayor presencia es la conferencia, seguido por los tutoriales (tabla 2). Cuando asignamos los tipos de material, utilizamos el catálogo de tipos que ofrece 5.2 LOM-ES en combinación con los propuestos por la biblioteca. Esta lista ampliada fue necesaria por la diversidad de los REA y las diferencias sutiles entre los materiales. Un ejemplo de

### **Metadatos sociales: iniciativas...**

esto es cuando el catalogador se enfrenta a asignar el tipo de material a un objeto digital, que podría ser un instructivo o una guía, que aparentemente son lo mismo. Para resolver esa duda se analiza con profundidad el REA a describir y se hace una revisión de las definiciones que contiene el estándar; éstas se han enriquecido con algunas otras definiciones locales para tomar una decisión semántica, es decir, qué contenido se le asignará al metadato tipo.

**Tabla 3.** Se muestra el formato de los documentos, de acuerdo con la lista en DSpace, por comunidad o entidad de la Universidad

Formatos de los recursos educativos abiertos (REA)	Comunidad BFXC	Comunidad DICA	Comunidad DiDAE	Comunidad DEAMeT	Totales por el tipo formato
mov	9	30		1	40
html	1			1	2
pdf	1		9	7	17
Total de REA por comunidad	11	30	9	9	59

Fuente: Elaboración propia.

El formato más usado para los REA es el .mov, por tratarse de una extensión de archivo en la creación de videos educativos como recurso didáctico preferido por los académicos visitantes de la Comunidad DICA y de la Biblioteca (FXC). Los resultados para los 59 recursos analizados demuestran que el video es el más utilizado en el campo académico universitario (tabla 3).

**Tabla 4.** Se muestra la frecuencia de uso de los formatos por tipo de REA de manera global en RI.

Tipo de rea global en RI	Formato .mov	Formato .pdf	Formato html
Tutorial	9	0	
Conferencia	21	0	
Sugerencias para docentes	0	7	
Curso	6	0	

Infografía	0	7	
Presentación multimedia	0	0	2
Taller	4		
Guía		2	
Seminario web	1		
Instructivo		1	

Fuente: Elaboración propia.

En la tabla 4 puede observarse que además de las conferencias en formato .mov los REA que deberíamos conocer más a fondo para asignar metadatos de manera adecuada son: el tutorial en formato .mov (dc.type=MovingImage – dc.format=mov), las sugerencias a docentes en texto – (dc.type=text y dc.format=pdf) y las infografías en multimedia – (dc.type=image y dc.format=pdf).

Las entidades de la universidad creadoras de los REA contribuyen desde diferentes aspectos a los procesos relacionados con la optimización, formación y apoyo a los docentes y estudiantes. Se reconoce la importancia del aprovechamiento de los recursos generados y la necesidad de ser incluidos en el RI. Por lo cual, a partir del 2021 se han realizado trabajos para incluirlos utilizando la propuesta antes mencionada en el apartado *Materiales y métodos* fase 2 con los metadatos DC y la adopción del vocabulario LOM-ES 5.2, adecuados con el objetivo de ampliar las posibilidades de identificación del tipo de recurso educativo. Los gestores de los repositorios señalan que los metadatos asociados a los REA facilitan las búsquedas en los repositorios, seguido de la pertinencia, accesibilidad y usabilidad.<sup>20</sup>

El análisis exploratorio de las colecciones de recursos educativos abiertos, asociadas a cuatro comunidades del RI, pone en evidencia varios aspectos sujetos a una revisión más profunda, en cuanto al enriquecimiento del vocabulario de los tipos de REA y la gestión del nombramiento de los archivos en el RI. El equipo

---

20 Dora M. Tovar *et al.*, *op. cit.*

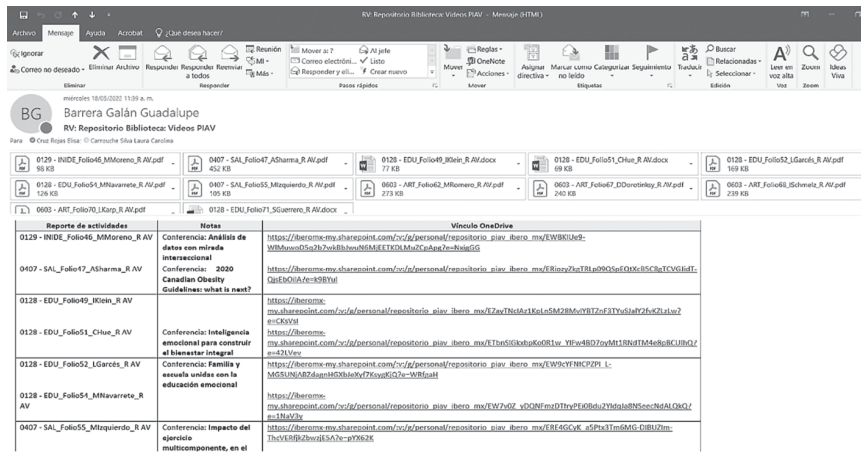
**Metadatos sociales: iniciativas...**

de catalogadores ha incrementado el vocabulario para incluir los tipos de recursos que han surgido. Es labor propia del ámbito de la catalogación la normalización de la información registrada. Esto ha representado un reto para la gestión de los registros de este recurso educativo.

La revisión de este trabajo nos llevó a identificar que los primeros registros no contienen todas las etiquetas que se definieron en la tabla 1, lo cual es lógico, debido a que esos primeros registros se capturaron previos a dicha propuesta.

Actualmente, los paquetes de información por parte de las unidades académicas que desean que les organicen sus contenidos educativos y se integren al RI, son enviados a la Coordinación de la Organización de la Información de la biblioteca. Por ejemplo, los paquetes del PIAV incluyen pdf, word y un excel con los datos específicos de cada REA, como autor, entidad académica de procedencia, país de origen, unidad académica anfitriona, programa asociado, actividades desarrolladas durante la visita, formato, título, ligas, etcétera, como se muestra en las siguientes figuras:



**Figura 3.** Correo electrónico con paquetes de información para su ingreso al repositorio



Fuente: Elaboración propia.

En la figura 3 se muestra un ejemplo de un correo electrónico con los paquetes de información enviados, por parte de la entidad PIAV, a la Coordinación de Organización de la Información para su ingreso al Repositorio Institucional. El envío incluye una tabla con información, enlaces y los documentos con las fichas de los académicos visitantes.

Figura 4. Muestra de una ficha de académico visitante como fuente de información para los metadatos descriptivos

 <b>REPORTE DE ACTIVIDADES</b> <b>Programa Institucional de Académicos Visitantes</b> (Para ser llenado por el AV) Muchas gracias por su colaboración	
Nombre del (de la) académico(a) visitante (incluya grado académico: Dr., Mtro., etc.):	
Dra. Belén Romero	
Institución de procedencia: Universidad de Valencia	
País: España	
Cargo actual en la Institución de procedencia: Académica	
Unidad Académica anfitriona: Departamento de Arte	
Programa(s) beneficiado(s) y/o programa(s) que organizan (favor de especificar): Doctorado en Historia y Teoría Crítica del Arte	
Contacto en la IBERO (nombre y dirección electrónica): María Sicarú Vásquez Orozco maria.vasquez@ibero.mx	
Actividades que desarrolló durante su visita: Conferencia "Maniobras ecológicas"	
Fecha(s) de realización de las actividades: 26 y 27 de agosto	
Formato de participación (conferencista, asistente, moderador, asesor, tutor, colaborador, etc.): Conferencista	
Título(s) de (los) trabajo(s) expuesto(s) (anexar copia de los mismos o página web para su consulta): Conferencia "Maniobras ecológicas"	
Comentarios o información adicional sobre su participación en la actividad realizada:	
La actividad se realizó vía zoom	
Firma del Académico Visitante	
 Firma de Ma. Sicarú Vásquez (p/a)	

Fuente: Elaboración propia.

### *Metadatos sociales: iniciativas...*

La ficha mostrada en la figura 4 incluye los datos generales del académico visitante, como son nombre, entidad anfitriona, el nombre del programa beneficiario de su participación, las actividades realizadas, fecha de la actividad, formato de participación, título del trabajo expuesto y tipo de REA. Esta ficha acompaña a la tabla con los nombres de archivos e hipervínculos, con los cuales se elaboran las descripciones para los metadatos específicos que alimentan los registros de los REA en las diversas comunidades del Repositorio Institucional.

### Análisis del enriquecimiento de los registros elaborados por los catalogadores

Uno de los principales problemas a los que nos hemos enfrentado es la asignación de un tipo específico de recurso, ya que, en muchas ocasiones las diferencias son muy sutiles. Tal es el caso, por ejemplo, de guía, instructivo, sugerencias a los docentes, tutoriales o seminario web, por mencionar algunos.

Otro problema enfrentado es que la normalización del nombramiento de los archivos sea consistente con la política de la gestión del RI. Estos 59 registros han permitido establecer algunas tendencias de los recursos educativos generados en la universidad. Por ejemplo, destacar que el video y el pdf son formatos muy utilizados para diversos tipos de recursos educativos. Por otro lado, los tutoriales, conferencias, cursos e infografías son tipos de REA ampliamente usados por las entidades creadoras de la universidad.

También, el análisis de los registros pone de manifiesto que son elementos sujetos a revisión y enriquecimiento constante por parte de los catalogadores responsables de alimentar el RI. Este trabajo permite la identificación de aspectos sujetos a mejora continua.

Como resultado preliminar se ha desarrollado una lista de etiquetas orientadas a dar visibilidad a los REA utilizando el sistema DSpace y un análisis de los registros existentes a fin de homologar sus contenidos. Para ello ha sido fundamental el uso de los metadatos correctos para generar puntos de acceso eficientes, interoperables y susceptibles de preservación a largo plazo.

## Análisis de la experiencia de asignación de metadatos por parte de becarios de la DEAMeT

Como se mencionó en la fase 3, se elaboró un manual para que las áreas donde se originaban los REA enviaran los metadatos correspondientes en un archivo de excel. La DEAMeT fue la única área que respondió a esta invitación, quien asignó a becarios analizar los REA y capturar la información en un archivo excel de acuerdo con el manual. Esta primera experiencia nos mostró algunos problemas en el proceso. Algunos aspectos que se identificaron como puntos a enriquecer y sistematizar son: puntuación y temas ortográficos, fechas de publicación de cuando el recurso se hace público, tipo de formato del recurso determinado por los creadores y sistematización en el llenado de los campos de la hoja de cálculo, es decir, no dejar huecos de información o filas en blanco. También resultó necesario registrar un pequeño esbozo de los contenidos de los recursos, así como palabras clave sobre los contenidos. Y, sobre todo, especificar el tipo de recurso educativo en cuestión.

Es importante destacar que la información recabada, realizada por parte de los becarios que participaron en la asignación de metadatos dentro del flujo de trabajo, nos proporcionó retroalimentación a partir de la cual pudimos determinar puntos a fortalecer, como mejorar de manera contundente los manuales y guías de llenado de los archivos; también establecer estrategias de comunicación rápida con los becarios y sobre todo implementar talleres de orientación. Esto, con la intención de optimizar los archivos que sirven de guía a los catalogadores para alimentar el repositorio de una manera eficiente y rápida.

## Análisis de la retroalimentación de los usuarios

Al analizar la experiencia de la integración de los REA al RI y la asignación de los metadatos correspondientes no hemos tenido mayores comentarios. El único que hemos recibido se relaciona con el acceso a los videos de conferencias de la colección del PIAV. Esta colección se alimenta con la entrega de información de

videos de conferencias de profesores visitantes a ser catalogados en el RI de acuerdo con los metadatos acordados. La Dirección de Informática y Telecomunicaciones (DIT) se encarga de almacenar en un espacio en la nube dichos videos, para asegurar un acceso rápido a los mismos.

El problema reportado se relacionó con la liga de acceso a los videos que se almacenan en el metadato, `dc.relation.uri`, la cual en su configuración inicial permitía al usuario llevar a cabo acciones que podían poner en riesgo la integridad de la liga. Este tema está en proceso de solución por parte de los responsables de generar la liga, quienes se asegurarán de que el usuario final no tenga posibilidad de vulnerar la seguridad solicitada.

## DISCUSIÓN Y CONCLUSIONES

- La situación de pandemia generó que se creara una gran cantidad de recursos educativos abiertos. Las áreas académicas identificaron a la biblioteca como un espacio para organizarlos.
- El proceso de creación, difusión y organización de los REA implica a diversos agentes que se involucran en la generación de metadatos de forma consciente e inconsciente:
  - Profesores creadores
  - Responsables de la difusión
  - Catalogadores
  - Usuarios finales
- La información que fluye asociada a estos materiales no siempre es completa y, aunque lo esté, no siempre se toma en cuenta para su catalogación. Para lograr una catalogación completa para el usuario final es necesario una revisión constante, no sólo por parte de los catalogadores sino por todos los actores involucrados.
- Los recursos educativos abiertos son tan complejos que representan un reto para los catalogadores, quienes, a pesar de tener experiencia en el manejo de recursos digitales,



necesitan aplicar estándares especializados, normalización de registros, sistematización de nombres de archivos, así como optimizar los flujos de trabajo y ofrecer capacitación. Los cambios de las formas de trabajo involucran no sólo la dificultad de salir de la zona de confort, sino por la complejidad de modificar estrategias de trabajo.

- Los flujos para organizar este tipo de trabajo tan complejo son diversos y con muchos actores involucrados, lo que requiere una visión de constante retroalimentación y mejora en el proceso. Esto demanda catalogadores flexibles, abiertos al cambio para enfrentar la explosión de los documentos y la dinámica de la tecnología.
- Se vuelve imperativa la capacitación de los diversos agentes involucrados en la construcción de metadatos, no sólo a través de manuales, sino también, por medio de talleres y/o videos con ejemplos concretos que sirvan de guía a los creadores de recursos.
- La revisión de registros a la luz de los estándares nos permitió identificar que nuestra instalación actual de repositorio está orientada a literatura científica, lo que nos hace considerar si fuera adecuado tener un repositorio especializado en recursos educativos abiertos a futuro, o bien, adecuar plantillas e interfaces a diversas colecciones desde un mismo repositorio.

Se puede concluir que es vital organizar, describir e indizar para su reutilización este tipo de materiales en los ambientes educativos contemporáneos y una responsabilidad para las bibliotecas universitarias encargadas de repositorios institucionales. Todo esto es un proceso extremadamente complejo que requiere una constante reflexión y revisión con involucramiento de todos los agentes.

### Prospectiva

- Elaborar la política de captura y metadatos para los REA a partir de las evidencias mostradas

### *Metadatos sociales: iniciativas...*

- Actualizar el manual
- Diseñar la formación de los colaboradores de metadatos con una estrategia pedagógica adecuada
- Optimizar y documentar los flujos de trabajo
- Incrementar el poblamiento de los REA en el RI, a partir de una mayor vinculación con más entidades académicas de la universidad
- Intercambiar experiencias con otras bibliotecas académicas a cargo de colecciones de los REA
- Integrar la asignación de metadatos de los usuarios con criterios que permitan evaluar la utilidad del material
- Integrar criterios de valoración como metadatos de los REA basados en modelos de reutilización<sup>21</sup> y criterios de calidad<sup>22</sup> que puedan ser asignados por los creadores, departamentos académicos y usuarios finales

### AGRADECIMIENTOS

La realización de las fases piloto que fundamentan la sistematización y reflexión de la experiencia no habría sido posible sin la solicitud de integración al RI de conferencias de profesores visitantes por parte de Rocío de Luna de la Dirección de Cooperación Académica. La integración de los diversos tipos de materiales educativos fue una solicitud de Indira Ochoa de la Dirección de Enseñanza y Aprendizaje Mediados por Tecnología. Asimismo, ha sido muy valiosa la participación en discusiones y pruebas de metadatos de las autoras con Benjamín Guerrero, Lourdes González y Laura Carrouché, quién además leyó el texto e hizo algunas sugerencias; todos ellos de la BFCX. Agradecemos también la solución del almacenamiento eficiente de los videos de las conferencias por

---

21 Alma Beatriz Rivera *et al.*, “Learning materials reusability in higher education: elements...”.

22 Irene Aguilar *et al.*, “Análisis de criterios de evaluación para la calidad de los materiales didácticos digitales”.

parte de Eugenio Fajardo Talavera, de la Dirección de Informática y Telecomunicaciones. La revisión de la interfaz del usuario de los REA fue hecha por Cristina Díaz de la DICA. La responsabilidad del contenido corresponde exclusivamente a sus autoras.

## REFERENCIAS BIBLIOGRÁFICAS

Aguilar Juárez, Irene, Joel Ayala de la Vega, Oziel Lugo Espinosa, y Alfonso Zarco Hidalgo. “Análisis de criterios de evaluación para la calidad de los materiales didácticos digitales”. *Revista iberoamericana de ciencia tecnología y sociedad* 9, núm. 25 (2014): 73-89.

Atenas-Rivera, Javiera, Francisco Rojas-Sateler y Mario Pérez-Montoro. “Repositorios de recursos educativos abiertos”. *El profesional de la información* 21, núm. 2 (2012): 190-193.

Colome, Dunia. “Objetos de Aprendizaje y Recursos Educativos Abiertos en educación superior”. *EDUTEC. Revista Electrónica de Tecnología Educativa* 69 (2019): 89-101.

Furubotten, Lisa. “Informe desde las trincheras: transformando unidades de catalogación para incluir servicios de repositorio”. En *La revolución de los datos bibliográficos, científicos y culturales*, de Ariel Alejandro Rodríguez García, 113-139. Ciudad de México: UNAM, IIBI, 2019.

Gordillo Méndez, Aldo, Enrique Barra Arias y Juan Quemada Vives. “Estimación de calidad de objetos de aprendizaje en repositorios de recursos educativos abiertos basada en las interacciones de los estudiantes”. *Educación XXI* 21, núm. 1 (2018): 285-302.

### ***Metadatos sociales: iniciativas...***

IEEE Standards Association. *IEEE Standard for Learning Object Metadata*. Editado por The Institute of Electrical and Electronics Engineers. 2022. <https://standards.ieee.org/ieee/1484.12.1/7699/> (último acceso: 20 de mayo de 2022).

Morales-Morgado, Erla Mariela, Rosalynn Argelia Campos-Ortuño y Tránsito Ferreras-Fernández. “Guía de metadatos educativos para agregar Recursos Educativos Abiertos (REA) en repositorios basados en DSPACE”. *XVIII Workshop rebium de Proyectos Digitales / VIII Jornadas OS Repositorios*. León, 25-27 de septiembre, 2019.

Pérez-Ortega, Isabel. "Creación de Recursos Educativos Digitales: Reflexiones sobre innovación educativa con TIC". *Revista Internacional de sociología de la educación* 6.2 (2017): 243-268.

Rivera Aguilera, Alma Beatriz. “La concepción didáctica del docente y los materiales didácticos digitales: voz, texto y producción de profesores universitarios”. Tesis doctoral. Ciudad de México: Universidad Iberoamericana. 2009.

———. “Metadatos de contenidos educativos: problemáticas y tendencias”. *Congreso Internacional sobre Metadatos 2016*. Ciudad de México: UNAM, IIBI. 2016.

———. “Biblioteca digital universitaria y el acceso a materiales didácticos digitales”. En *Bibliotecas universitarias, contenido digital y educación en línea: algunas experiencias iniciales en la Universidad Iberoamericana Ciudad de México*, coordinado por Alma Beatriz Rivera Aguilera, 25-38. Ciudad de México: Universidad Iberoamericana. 2019.

- Rivera Aguilar, Alma Beatríz, José Juan Téllez Bertadillo y Víctor Manuel Harari Betancourt. "Learning materials reusability in higher education: elements for designing digital collections from a knowledge management perspective". *Proceedings of the American Society for Information Science and Technology* 47, 1-5. Association for Information Science and Technology. 2010. 1-5.
- , Elisa Cruz Rojas, María Guadalupe Barrera Galán y Acalia Pozo Marrero. "Propuesta de un perfil de aplicación para material didáctico digital basada en el Marco de Singapur". *6o Encuentro de Catalogación y Metadatos*. Ciudad de México: UNAM, Centro de Investigaciones Bibliotecológicas. 2011.
- Rodríguez Delís, Yomé Marjories, Rafael Luis Campaña Jiménez y María Jesús Gallego Arrufat. "Iniciativas para la adopción y uso de recursos educativos abiertos en Instituciones de Educación Superior". *Educación Médica Superior* 32, núm. 4 (2018). <https://ems.sld.cu/index.php/ems/article/view/1524/734>
- Santos-Hermosa, Gema, Núria Ferran-Ferrer y Ernest Abadal. "Recursos educativos abiertos: repositorios y uso". *El profesional de la información* 21, núm. 2 (2012): 136-145.
- Sherriff, Graham, Daisy Benson y Gary S. Atwood. "Practices, policies and problems in the management of learning data: a survey of libraries' use of digital learning objects and the data they create". *Journal of Academic Librarianship* 45, núm. 2 (2019): 102-109.
- Smith-Yoshimura, Karen y Cyndi Shein. *Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews*. Dublin, Ohio: OCLC Research. 2011. <http://www.oclc.org/research/publications/library/2011/2011-02.pdf>.

***Metadatos sociales: iniciativas...***

Tovar Gutiérrez, Dora María, Alejandro López Ibarra y María Soledad Ramírez Montoya. “Estrategias de comunicación para potenciar el uso de Recursos Educativos Abiertos (REA) a través de repositorios y metaconectores”. *Innovar* 24, núm. 52 (2014): 67-78.

# El sentido social del dato científico generado por la biblioteca universitaria desde la práctica de la divulgación académica

LUISA CORAL ACOSTA CRUZ  
*Seminario de Investigación Metadatos*

## INTRODUCCIÓN

Los recientes acontecimientos en el mundo como los desplazamientos por conflictos bélicos, la pandemia por COVID-19 o las desapariciones forzadas son eventos que no pueden pasar desapercibidos por la academia universitaria. Las investigaciones realizadas en el ámbito universitario y sus resultados suman a la solución mundial, pero estamos lejos de cumplir con los Objetivos de Desarrollo Sostenible, ODS, de la Organización de las Naciones Unidas para 2030. Los integrantes de la comunidad bibliotecaria apoyan de manera sustancial en el desarrollo de las investigaciones universitarias, además de que están interesados en ser parte de la solución de las problemáticas mundiales.

La biblioteca universitaria tiene una agenda donde la divulgación es parte de las actividades que fomentan sus integrantes. Los productos generados por su comunidad tienden a ser divulgados hacia su población universitaria y son replicados fuera de su comunidad. Los repositorios son un buen ejemplo de ello. Las plataformas digitales de los repositorios necesitan que los metadatos albergados tengan una etiquetación precisa para que la

recuperación de información sea eficaz. La experiencia de los integrantes de la comunidad bibliotecaria permite obtener resultados favorables fomentando la colaboración multidisciplinaria.

En la actualidad es de suma importancia que los integrantes de la comunidad bibliotecaria pongan énfasis en divulgar el conocimiento, en particular, sobre temas que están integrados en los ODS. La divulgación del dato científico es una buena práctica para fortalecer el beneficio que tienen las investigaciones para la sociedad.

## DESARROLLO

La emergencia mundial ante los acontecimientos adversos que enfrenta la humanidad desde el siglo pasado hasta nuestros días impone retos a los gobiernos, así como a las organizaciones encargadas de paliar las deficiencias mundiales que azotan al planeta. La Agenda 2030 para el Desarrollo Sostenible reúne los objetivos del mundo por disminuir los riesgos a los que se enfrenta la humanidad. En 2015, las naciones reunidas en la cumbre mundial de las Organización de las Naciones Unidas acordaron trabajar en 17 objetivos: 1. Fin de la pobreza, 2. Hambre cero, 3. Salud y bienestar, 4. Educación de calidad, 5. Igualdad de género, 6. Agua limpia y saneamiento, 7. Energía asequible y no contaminante, 8. Trabajo decente y crecimiento económico, 9. Industria, Innovación e Infraestructura, 10. Reducción de las desigualdades, 11. Ciudades y comunidades sustentables, 12. Producción y consumo responsables, 13. Acción por el clima, 14. Vida submarina, 15. Vida de ecosistemas terrestres, 16. Paz, justicia e instituciones sólidas, 17. Alianza para lograr los objetivos.



**Figura 1. 17 objetivos de la cumbre mundial de la ONU, 2015**



Fuente: ONU.

Un reporte anual de los compromisos adquiridos por las naciones involucradas en alcanzar el cumplimiento de los Objetivos de Desarrollo Sostenible, ODS, pone en evidencia que estamos lejos de la meta de los Objetivos, como menciona el reporte correspondiente a 2020. A continuación, la mención de las conclusiones de dicho informe:

- Se estima que aproximadamente 71 millones de personas volverán a caer en la extrema pobreza en 2020, lo que supondría el primer aumento de la pobreza mundial desde 1998. La pérdida de ingresos, la limitada protección social y el incremento de los precios podrían poner en riesgo de pobreza y hambre incluso a personas que anteriormente estaban a salvo.
- El subempleo y desempleo derivados de la crisis implican que aproximadamente 1.600 millones de trabajadores ya vulnerables en la economía sumergida (la mitad de la fuerza laboral mundial) pueden verse considerablemente afectados, con un descenso estimado del 60% en sus ingresos durante el primer mes de la crisis.

*Metadatos sociales: iniciativas...*

- Los más de 1.000 millones de residentes de barrios marginales de todo el mundo están en grave situación de riesgo a causa de los efectos de la COVID-19, debido a la falta de viviendas adecuadas y agua corriente en las viviendas, baños compartidos, la escasez o ausencia de sistemas de gestión de residuos, la saturación de los transportes públicos y el acceso limitado a las instalaciones sanitarias oficiales.
- Las mujeres y los niños se encuentran asimismo entre las personas más afectadas por las consecuencias de la pandemia. La interrupción de determinados servicios sanitarios y de vacunación, así como el limitado acceso a los servicios de nutrición y alimentación, podrían ocasionar cientos de miles de fallecimientos adicionales entre los niños menores de cinco años y decenas de miles de muertes maternas adicionales en 2020. En muchos países se han disparado las denuncias de violencia doméstica contra mujeres y niños.
- Los cierres de las escuelas han afectado al 90% de los estudiantes de todo el mundo (1.570 millones) y han provocado que más de 370 millones de niños se salten comidas escolares de las que dependen. Dada la falta de acceso a ordenadores y a internet en casa, el aprendizaje remoto queda fuera del alcance de muchos. Alrededor de 70 países notificaron interrupciones de moderadas a graves o la suspensión total de los servicios de vacunación infantil durante los meses de marzo y abril de 2020.
- A medida que más familias caen en la extrema pobreza, los niños de las comunidades pobres y desfavorecidas corren un riesgo mucho mayor de verse involucrados en el trabajo infantil, el matrimonio infantil y el tráfico infantil. De hecho, es probable que los progresos logrados a nivel mundial en la reducción del trabajo infantil se vean invertidos por primera vez en 20 años.<sup>1</sup>

---

1 Organización de las Naciones Unidas, Informe sobre los progresos en el cumplimiento de los ODS 2020, “Algunas de las principales conclusiones”.

El informe señala asimismo que el cambio climático continúa produciéndose a una velocidad mucho mayor de lo previsto. El 2019 fue el segundo año más cálido del que se tenga constancia y marcó el final de la década más cálida jamás registrada: 2010-2019. Al mismo tiempo, la acidificación de los océanos se acelera, la degradación del suelo continúa, hay especies masivas en riesgo de extinción y siguen predominando los patrones insostenibles de consumo y producción.<sup>2</sup>

En este sentido, el Secretario General de las Naciones Unidas, Antonio Guterres, mencionó que

Como estados miembros reconocidos en la Cumbre de los ODS celebrada el pasado septiembre, los esfuerzos mundiales llevados a cabo hasta la fecha han sido insuficientes para lograr el cambio que necesitamos, lo que pone en riesgo el compromiso de la Agenda con las generaciones actuales y futuras. Ahora, debido a la COVID-19, una crisis sanitaria, económica y social sin precedentes amenaza vidas y medios de subsistencia, lo que dificulta aún más la consecución de los Objetivos.<sup>3</sup>

Con la llegada de la pandemia en 2019 fue significativo el impulso de la producción científica en dirección de la ciencia médica, la que aceleró los procesos de investigación con repercusión en otras ciencias más. La comunicación científica tuvo una dinámica inmediata entre pares que, por la importancia significativa del acontecimiento a nivel global, a su vez generó información que tuvo que divulgarse rápidamente para la protección de las personas contra la COVID-19.

Algunos especialistas afirman que los resultados de las investigaciones deben salir del ámbito académico hacia la comunidad como una obligación informativa para la sociedad. La divulgación

---

2 *Ibidem.*

3 Antonio Guterres, Informe sobre los progresos en el cumplimiento de los ODS 2020.

científica debe comunicar a la población en general contenidos científicos en un lenguaje comprensible para la sociedad. Los formatos de comunicación electrónicos así como digitales permiten llegar a una diversidad de personas, por lo que la inmediatez de propagar la información impulsa a una sencilla presentación de contenidos.

La Organización Mundial de la Salud contiene en su página digital, [www.who.int/es](http://www.who.int/es), un nicho especializado sobre la enfermedad por coronavirus (COVID-19) en español. En este sitio el lector encuentra documentos, videos, carteles e infografías. En cuanto al sitio en idioma inglés, [//covid19.who.int/](https://covid19.who.int/), éste contiene mapas en tiempo real del comportamiento diario de la enfermedad. La población mundial obtiene un seguimiento estadístico de la enfermedad de la coronavirus.

En la actualidad uno de los objetivos significativos de los ODS por cumplir tras la pandemia es el número tres, que corresponde a Salud y bienestar, así como el número cuatro, referente a Educación de calidad. Este último tiene para los países involucrados un reto importante, como en México, donde los motivos para que los alumnos de nivel superior no concluyeran el año escolar en 2019-2020 fue en primer lugar por la COVID-19, con un 44.6%; el segundo motivo, porque tenían que trabajar, con un 17.1%; otras razones con 30.3% y por último el 8.0% por falta de recursos, según la Encuesta para la Medición del Impacto COVID-19 en la Educación (ECOVID-ED) 2020, realizada por el Instituto Nacional de Estadística, Geografía e Informática, INEGI.<sup>4</sup>

Con respecto al uso de TIC, señala este documento que para realizar actividades escolares en una población entre 3 y 29 años, en el ciclo escolar 2019-2020, fue el teléfono inteligente el recurso tecnológico más usado, con un 65.7%, seguido de una computadora portátil, con un 18.2%, luego de la computadora de escritorio,

---

<sup>4</sup> Instituto Nacional de Estadística, Geografía e Informática, INEGI, “Encuesta para la Medición del Impacto COVID-19 en la Educación (ECOVID-ED) 2020”.

con un 7.2%, acto seguido está la televisión digital y por último la *tablet* con un 3.6%.<sup>5</sup>

Este panorama no es ajeno a la Universidad Nacional Autónoma de México, UNAM. En ella, los integrantes del Sistema Bibliotecario de la Dirección General de Bibliotecas y Servicios Digitales de Información, DGBSDI, contribuyeron a impactar de manera positiva a la comunidad durante el confinamiento por la pandemia de la COVID-19 procurando el servicio de sus recursos digitales a distancia para la comunidad universitaria. Esta acción está estrechamente relacionada con el compromiso de desarrollo y acceso a la información que las bibliotecas suman para la promoción de los ODS con respecto a Educación de calidad, como muestra el documento *Development and Access to Information 2019*, por parte de la International Federation of Library Associations and Institutions, IFLA (<https://da2i.ifla.org/>).

Los eventos organizados por la DGBSDI-UNAM durante el periodo de confinamiento sobresalieron por su sentido social para contribuir al fortalecimiento de la comunidad universitaria. La visibilidad de los eventos como el II Foro Virtual *Nuevos desafíos de las bibliotecas universitarias ante la COVID-19*, así como el VIII Foro Virtual DGBSDI-UNAM, *Bibliotecas accesibles e inclusivas para el acceso a la información de personas con discapacidad: La contribución de la biblioteca universitaria en la educación inclusiva*, organizado en el marco del Día Internacional de las Personas con Discapacidad, son un par de ejemplos de acceso incluyente por parte de la DGBSDI-UNAM hacia la comunidad universitaria.

El albergar los productos bibliotecarios en plataformas digitales institucionales de propia manufactura o en medios sociales que comunican a un público mayor y heterogéneo significa no sólo establecer un canal o varios canales de difusión, sino la buena práctica de divulgación de datos científicos.

La actividad bibliotecaria maneja un número significativo de metadatos científicos puestos en plataformas digitales como los

---

5 *Ibidem.*

repositorios. Uno de ellos, es el Repositorio Institucional de la UNAM (RI-UNAM), [www.repositorio.unam.mx](http://www.repositorio.unam.mx) (figura 2).

El RI-UNAM es el principal punto de consulta en línea de los contenidos digitales en acceso abierto producidos y resguardados por la universidad; funciona como una plataforma integradora de los contenidos albergados en los numerosos repositorios universitarios de las diversas entidades y dependencias de la institución. Asimismo, fue creado también para ser un nodo de conectividad con plataformas digitales de otras instituciones académicas y de gobierno, como el Repositorio Nacional del Conahcyt y el de la Coordinación de la Memoria Histórica y Cultural de México.

La RI-UNAM es interoperable con una semántica basada mediante la cosecha de metadatos OAI-PMH y una interfaz de programación de aplicaciones Application Programming Interface, API. Los metadatos albergados se encuentran con el sistema Dublin Core, que proporciona la descripción de recursos de información interdisciplinarios o con el formato MARC 21 para datos bibliográficos. Este repositorio cuenta con una caja de búsqueda que permite acceder a recursos digitales ya sean metadatos o texto completo. Las búsquedas pueden ser transversales, es decir, es capaz de enlazar varios acervos o un grupo en específico. Asimismo, cuenta con filtros para delimitar la búsqueda por repositorio, acervo, tipo de recurso, tipo de contenido, entidad aportante de la UNAM, área del conocimiento, año de producción, institución aportante, colección o autor, como se muestra en la figura 3.

La etiquetación social requiere de conocimientos del ámbito bibliotecario, como metodologías, normativas, así como políticas para el buen funcionamiento y uso de los productos finales. Hemos mencionado el esquema de metadatos Dublin Core como uno de los descriptores de recursos más populares en la práctica. Las etiquetas utilizadas permiten asignaciones flexibles, por ejemplo, en la temática o la cobertura.

Figura 2. Repositorio Institucional de la UNAM

UNAM
Repositorio Institucional de la UNAM
Inicio
Normatividad

Artículos x · · · dirección general de bibliotecas y servicios digitales de información UNAM · x

Artículos, tesis, libros, periódicos, registro de colecciones universitarias

Todo ▾

1 - 50 de 256 resultados

1 / 6 > <

**Repositorio**

Revistas UNAM 256

**Acervo**

Artículos 256

**Tipo de recurso**

Artículo 256

**Tipo de contenido**

Q, I, I 233

Artículo Técnico-Profesional 233

Artículo de Investigación 23

**Entidad aportante de la UNAM**

Dirección General de Bibliotecas y Servicios Digitales de Información, UNAM 233

**Artículo**

Hacia un modelo de servicios en la biblioteca digital

Tomás Vargas, Cecilia Anselmi, Instituto de Investigaciones Bibliotecológicas y de Información, UNAM, Ciencias Sociales y Económicas

Revistas UNAM

**Artículo**

Mensajería, servicios y reflexiones para bibliotecas

Dr. Anselmi, Susi, Anselmi, Dirección General de Bibliotecas y Servicios Digitales de Información, UNAM, Multidisciplina

Revistas UNAM

**Artículo**

Aplicación Móvil de Servicios de Información del Sistema Bibliotecario y de Información de la UNAM (AVCSBI-UNAM)

Ulrichs, Daniel, Miguel Ángel Durán, Martínez, Israel, Ibarra Contreras, Rafael - Dirección General de Bibliotecas y Servicios Digitales de Información, UNAM

2016-12-01

Revistas UNAM

**Artículo**

Bibliotecas Universitarias y de Servicios bibliotecarios y de información

Lemus, Alberto, Estrada Lemus, Osáez Jiménez, Daniel - Dirección General de Bibliotecas y Servicios Digitales de Información, UNAM

2022-08-13

Multidisciplina

Revistas UNAM

Fuente: UNAM.

Figura 3. Dublin Core

DUBLIN CORE		
Nombre del elemento	Elemento	Descripción
<b>Etiqueta (Label)</b>	Título	Palabra clave con la cual se identifica el recurso
<b>Definición (Definition)</b>	Definición	Una palabra(s), frase(s) u oración(es) que el creador asigno para reconocer e identificar su recurso
<b>Comentario (Comment)</b>	Comentario	Regularmente el título se encuentra, por ejemplo, en los archivos de texto (.doc, .txt, rtf) al inicio del escrito. También puede ser utilizada como el nombre con el cual se guardo el archivo.
<b>Repetibilidad (Repeatable)</b>	Repetitividad	SI
<b>Obligatoriedad (Mandatory)</b>	Obligatoriedad	SI
<b>Calificador (Qualifier)</b>	Calificador	

Fuente: Dublin Core.

Figura 4. Registro Dublin Core

```

32
REGISTRO DUBLÍN CORE

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description>
    <dc:title>Building digital libraries : a how-to-do-it manual /</dc:title>
    <dc:creator>Reese, Terry.</dc:creator>
    <dc:creator>Banerjee, Kyle.</dc:creator>
    <dc:type>text</dc:type>
    <dc:publisher>New York : Neal-Schuman Publishers.</dc:publisher>
    <dc:date>c2008.</dc:date>
    <dc:language>eng</dc:language>
    <dc:description>Includes bibliographical references and index.</dc:description>
    <dc:description>Planning a digital repository -- Acquiring, processing, classifying, and describing digital content -- Choosing a repository architecture -- General purpose technologies useful for digital repositories -- Metadata formats -- Sharing data : metadata harvesting and distribution. -- Federated searching of repositories -- Access management -- Planning for the future -- Conclusions.</dc:description>
    <dc:subject>Digital libraries.</dc:subject>
    <dc:identifier>http://www.loc.gov/catdir/toc/ecip0720/2007024103.html</dc:identifier>
  </rdf:Description>
</rdf:RDF>

```

Fuente: Dublin Core.



En cuanto al esquema MARC, éste es utilizado para codificar documentos bibliográficos en las bibliotecas universitarias. Es aceptado por la comunidad especializada en la materia de Documentación porque permite relacionar temáticas mediante un lenguaje controlado. Los catálogos LIBRUNAM y SERIUNAM son ejemplo de la puesta en práctica de este tipo de formato.

La Biblioteca del Congreso de los Estados Unidos produjo un esquema denominado Metadata Object Description Schema, MODS, para describir objetos digitales a partir de 2015. Las bibliotecas utilizan este esquema con mayor frecuencia para describir recursos digitales, ya que cuenta con una plantilla de etiquetas conjugadas a subelementos como se muestra en la imagen.

Como resultado de investigaciones sobre conflictos sociales como los desplazamientos de personas por la guerra, el tráfico de sustancias nocivas para la salud, trata de personas o desapariciones forzadas, integrantes del ámbito bibliotecario tienen a bien realizar o colaborar en proyectos que apoyen a las personas en conflicto. Uno de los proyectos significativos puesto en acción es la plataforma del Repositorio de Documentación sobre Desapariciones en México, RDDM, cuya presentación fue en el Colegio de México como institución participante, CDMX, los días 23 y 24 de marzo de 2022, con una conferencia titulada “Verdad, justicia, memoria: evidencia documental en la era digital”.

La doctora Guadalupe Vega Díaz, coordinadora de Organización de la Información de la Biblioteca Daniel Cosío Villegas del Colegio de México e integrante del RDDM, explicó que la plataforma exigió la realización de una plantilla con un esquema de metadatos especializada en el tema de desapariciones forzadas en México, para que la documentación ayude a los familiares involucrados o a los interesados en el tema. En cuanto a la elaboración de plantillas para un nicho en específico, están las instituciones del ámbito cultural como son los museos. Un ejemplo de ello es Getty Museum, que utiliza los esquemas ya mencionados en anteriores líneas, además de albergar metadatos relacionados con aquellos referentes al arte.

## CONCLUSIÓN

Los Objetivos de Desarrollo Sostenible de la Agenda 2030 son una oportunidad para mejorar las condiciones de vida de quienes habitamos el planeta y, a su vez, para el cuidado de éste. La biblioteca y sus integrantes son un componente básico para formar parte de la solución de los conflictos sociales y el bienestar común. Considero que la experiencia de los profesionales de la información documental proporciona herramientas que permiten el análisis de tópicos actuales para una recuperación de datos veraz y oportuna. Los esfuerzos realizados en este sentido fueron significativos durante el confinamiento de la pandemia de la COVID-19, porque aportó espacios de reflexión sobre los retos que tenemos que enfrentar en los años venideros. La biblioteca y sus integrantes tuvieron a bien realizar actividades de divulgación con sentido social, donde los productos presentados contienen metadatos con etiquetado social debido a su origen digital. El resultado es positivo porque enfatiza la vertiente de generar un impacto benéfico para la sociedad mediante los productos elaborados por la comunidad universitaria.

## REFERENCIAS BIBLIOGRÁFICAS

- Espinosa Santos, Victoria. “Difusión y divulgación de la investigación científica”. *IDESIA* (Chile) 28(3): 5-6.
- Guterres, Antonio. “Informe sobre los progresos en el cumplimiento de los ODS 2020”. ONU. <https://www.un.org/sustainabledevelopment/es/progress-report/>.
- Instituto Nacional de Estadística, Geografía e Informática, INEGI. “Encuesta para la Medición del Impacto COVID-19 en la Educación (ECOVID-ED) 2020”. <https://inegi.org.mx/investigacion/ecovided/2020/>.

- Kimura, Masahiro, Saito Kazumi, y Ryohei Nakano. "Extracting influential nodes on a social network for information diffusion". *Association for the Advancement of Artificial Intelligence* 7 (2007): 1371-76.
- Organización de las Naciones Unidas, ONU. Informe sobre los progresos en el cumplimiento de los ODS 2020. "Algunas de las principales conclusiones". <https://www.un.org/sustainabledevelopment/es/progress-report/>
- Sánchez-García, Sandra, y Santiago Yubero. "Función social de las bibliotecas públicas: nuevos espacios de aprendizaje y de inserción social". *Profesional de la Información* 24(2), 103.
- Santo Domingo, "La función social de las bibliotecas universitarias". *Boletín de la Asociación Andaluza de Bibliotecarios* 20(80), 43-70.

# La catalogación social, su práctica profesional y empírica

ARIEL ALEJANDRO RODRÍGUEZ GARCÍA

*Universidad Nacional Autónoma de México, México*

## INTRODUCCIÓN

**E**n el año de 2013 se publicó el libro intitulado *Nuevas direcciones en la Organización de la Información (New directions in organization information)*. El autor de este texto hizo ahí una revisión literaria desde el surgimiento hasta el tercer lustro del siglo XXI sobre lo que acontecía en la organización de la información.

Subraya que la catalogación social, actualmente, está tomando auge debido a los cambios que han sufrido los catálogos en línea de acceso al público (OPAC, por sus siglas en inglés) y lo que actualmente se está entendiendo por catalogación social es la interacción que se debe dar entre el catalogador profesional y el catalogador empírico (usuario), con el fin de descubrir y acceder a los recursos de información que se encuentran en las bases de datos de los sistemas bibliotecarios integrales.

Hay otros autores que refieren que la catalogación ha estado presente en los catálogos desde antes de que surgieran las tecnologías de información, pero su repunte se da cuando ingresamos a la era digital y comienzan a utilizar las aplicaciones de la web

2.0, así como los nuevos modelos de producción, distribución y consumo de la información 2.0. Estos fueron anexados a los OPAC, y con ello, éste cambió su nombre a SOPAC (Social Online Public Access Catalog).

Llegamos a este punto para señalar que un ingrediente que está presente en este proceso, además, es el etiquetado social, el cual, desde el año 2000 ha contribuido a la realización de diversos estudios que contribuyen a observar al etiquetado como una folksonomía e indicativo para incorporarlos a los sistemas de etiquetado en los SOPAC.

Esta inclusión trae como ventajas que los recursos de información estén más cercanos a los usuarios, porque ellos participarían de un trabajo que anteriormente era privativo sólo de los catalogadores profesionales. Habría que decir también que la visión social del catálogo aparece en la década de 1970, cuando Shera incorpora la epistemología social a la Bibliotecología, al suscribir que el valor de un sistema se ejerce cuando influye altamente en la comunicación del conocimiento entre la sociedad y el medio que la sociedad utiliza para obtenerlo.

La finalidad de este trabajo es determinar el alcance que tiene la catalogación social en los tiempos que hemos pasado y pasaremos en estos tiempos de crisis sanitaria mundial, ya que se ha multiplicado el uso de los recursos de información, así como las tecnologías de información y las tecnologías móviles, que hacen, tanto al catalogador profesional como empírico, “prosumidores” de información.

Para que se alcance el objetivo anterior, nos proponemos resolver cuestionamientos tales como: ¿Cuál es el potencial de un sistema de etiquetación en el que interactúan los catalogadores profesionales y empíricos? ¿Cuál ha sido el progreso de los sistemas de etiquetación social en los SOPAC? ¿Cuáles han sido las experiencias y resultados del trabajo conjunto entre catalogadores profesionales y empíricos? ¿Hacia dónde se dirigen los estudios de la catalogación social?

## DE DÓNDE PARTIMOS Y HACIA DÓNDE NOS DIRIGIMOS

En el año 2000, Michael Gorman, el editor de las Reglas de Catalogación Angloamericanas, segunda edición, presentó un trabajo intitulado “¿Metadatos o catalogación? Un cuestionamiento erróneo”, en el XVI Congreso Internacional de Investigación “Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica”. En su texto menciona que “la catalogación es una actividad central en el campo del control bibliográfico. Consiste en un número de pasos complejos”<sup>1</sup> y los describe cada uno de ellos detalladamente, lo que rebasa las intenciones de este trabajo, pero quedémonos con el hecho de que recalca que “la catalogación es una actividad profundamente laboriosa y profesional por naturaleza”.<sup>2</sup>

De igual modo, encontramos el razonamiento de Escamilla,<sup>3</sup> quien refiere que la “catalogación descriptiva parte del supuesto de que los usuarios del catálogo conocen de antemano algún rasgo de la obra en cuestión, que les va a permitir identificarla mediante la descripción que de ella haga el catalogador”.

Otros autores, como Joudrey, Taylor y Miller,<sup>4</sup> advierten que hay tres componentes básicos alrededor del catálogo: el catálogo de acceso al público, una lista de estantería y los archivos de autoridad. Habrá que decir también que los catálogos en línea se han erigido de manera distinta a los catálogos en tarjetas, ya que en los primeros se dejaron de usar las listas de estantería, y se cambiaron por registros con patrones (formatos) automatizados organizados internamente por los sistemas de información.

---

1 Michael Gorman, “¿Metadatos o catalogación? Un cuestionamiento erróneo”, 5.

2 *idem*.

3 Gloria Escamilla, *Manual de catalogación descriptiva*, 15.

4 Daniel N. Joudrey, Arlene G. Taylor y David P. Miller, *Introduction to cataloging and classification*, 13.

Desde otro punto de vista, Gorman,<sup>5</sup> por su parte y Schottlaender,<sup>6</sup> por la suya, refieren que en las publicaciones aparecidas en el primer lustro del siglo XXI relativas a los metadatos se decía que estos estaban presentando una gran ironía referente a que se esté investigando. Esto es, la relación cercana que hay entre el control bibliográfico “tradicional” y al mismo tiempo el advenimiento de los recursos electrónicos.

El arribo de los recursos electrónicos provocaría cambios en la concepción que había con respecto a la organización, recuperación, autenticidad y estabilidad, debido a que hay grandes diferencias entre un documento físico y el recurso electrónico.

A la fecha es sabido que los recursos electrónicos, ahora recursos de información digital, se multiplican de manera diferente a los impresos; gran cantidad de esos recursos y sitios web, tal vez, no han sido o no son de interés para los bibliotecarios y los usuarios; su temporalidad puede ser efímera o permanente, así como sus niveles de estandarización son distintos a los que hay con relación al encontrar un libro u otro material tangible de la biblioteca.

Por su parte, Schottlaender señala que una de las palabras que más se escucha en el campo de los metadatos es “esquema”, la cual hace referencia al conjunto de reglas para codificar la información que es empleada en una comunidad de usuarios.

En la actualidad hay una gran cantidad de esquemas que nos permiten codificar los datos bibliográficos, científicos, culturales o de cualquier otro tipo y con ello establecer diferentes métodos de descripción, de acuerdo con la estructura del documento. De ahí que en su momento Schottlaender se cuestionara ¿por qué a mí?, ante la presencia de los metadatos y se respondiera con la siguiente frase “la catalogación es un proceso invisible de hacer pedidos”, es decir, los registros que se crean son el resultado de la aplicación de un conjunto de reglas y es un proceso que se ha realizado por muchas décadas, si no es que siglos, en el cual se ha acumulado mucha experiencia que se ha compartido con otras comunidades.

---

5 Michael Gorman, “Cataloguing in an electronic age”, 4.

6 Brian E. C. Schottlaender, “Why metadata, why me? Why now?”.

Definitivamente, hablar de la catalogación de antes o “tradicional” es algo que nos llevaría tiempo y reconocimiento a quienes forjaron todas las bases para lo que ahora se está haciendo, por ejemplo, a partir de la introducción del modelado de datos y la adopción de nuevas normativas, así como la evolución del catálogo en línea de acceso al público (OPAC, por sus siglas en inglés).

Para avanzar en nuestro razonamiento respecto a lo que deseamos presentar referente a la catalogación en su práctica social y empírica, diremos que ésta despunta cuando se habla de la siguiente generación de catálogos (*next generation of catalog*). Breending<sup>7</sup> refiere que el catálogo ha evolucionado, acompañado de los avances en el mundo de computación, a lo largo de varias generaciones. En efecto, las bibliotecas han adoptado esos avances al grado de introducir a la web.

El surgimiento del catálogo en línea web se dio en 1995. Éste ofrecía una interfase simple que permitía al usuario ingresar términos y ver sus resultados, incluyendo la localización y la situación que guardaba el ítem. Cabe señalar que de la web de 1995 a la que tenemos actualmente, hay una enorme diferencia y ésta se acentuó más con el momento en que Tim O'Reilly en 2004 dio inicio a la era de la web 2.0.

Al respecto, Chambers<sup>8</sup> nos dice que surge el catálogo 2.0 y con ello los catálogos del futuro, los cuales han venido transformándose en los últimos 25 años y no sólo ellos han cambiado, sino también la práctica de la catalogación, las habilidades básicas para la descripción bibliográfica y, por lo tanto, la creación misma del catálogo.

Las ideas anteriores nos dan la pauta para observar cómo es que se llega a concebir la catalogación social en la era de los nuevos catálogos conocidos como los catálogos 2.0.

---

7 Marshall Breending, *Next-gen library catalogs*, 4.

8 Sally Chambers, editora, *Catalogue 2.0: the future of library catalogue*.



## LOS USUARIOS COMO CONTRIBUIDORES DE CONTENIDO

Hemos avanzado dos décadas y casi un lustro de este siglo XXI en los que el proceso de catalogación ha avanzado en la creación, adecuación y apropiación de nuevas teorías, modelos, paradigmas y prácticas en la descripción de los recursos de información. Uno de estos nuevos paradigmas es que los modelos para la representación y descripción pueden crearse y actualizarse tanto por los profesionales como por los usuarios que acceden y usan los datos de los recursos de información.

A más de 150 años que Cutter plasmó los objetivos del catálogo y 30 años de que el modelado de datos de los requerimientos funcionales formuló las tareas básicas del usuario frente al catálogo (“encontrar, identificar, seleccionar y obtener”), en estos momentos observamos cómo los usuarios están incidiendo en lo que se está denominando etiquetación social. Ésta permite que el usuario sea capaz de emplear de forma local o vinculatoria los datos y relaciones entre las instancias específicas<sup>9</sup> de una posible obra, expresión, manifestación e ítems.

Ahora podemos decir que el etiquetado social o comunitario recomienda que el usuario final (empírico) pueda equilibrar el funcionamiento entre los metadatos y datos que ingresa un profesional y aquellos que adiciona el usuario final. De ahí que entre los sistemas de etiquetación se diga que hay por lo menos cinco motivaciones.

Al respecto, Smith<sup>10</sup> refiere que en un sistema de etiquetado, para que sea más fácil de usar e invertir un mínimo de tiempo, se deben establecer: etiquetas simples, etiquetas flexibles, etiquetas extensibles, etiquetas que puedan agregar información de otros sitios y recomendaciones sobre el uso de la etiqueta.

---

9 En el modelado de datos FRBR, una instancia es un atributo de alguna entidad, mientras que en la catalogación descriptiva se le conoce como elemento descriptivo.

10 Gene Smith, *Tagging people-powered metadata for the social web*, 12-13.

En contraste con lo anterior, se tienen opiniones discordantes entre lo que es la práctica de la catalogación y el etiquetado social. Sobre la primera se ha dicho que en las últimas décadas ha evolucionado en varias direcciones que ha propiciado que se deje de pensar en que sigue siendo una práctica artesanal con procesos tediosos, complejos, especializados y costosos. Además, la tendencia creada por el modelado de datos ha modificado el paradigma del seguimiento de instrucciones específicas para pensar ahora en los tipos de recursos, sus instancias (atributos) y darle una gran validez al juicio del catalogador.

Con relación al etiquetado social, según Taylor,<sup>11</sup> se parte de que en la actividad de los metadatos deben interpretarse tres partes fundamentales. Debe proveerse de una descripción al recurso para que éste pueda ser administrado, preservado y estructurado; se deben proporcionar los accesos a cada una de esas descripciones y se debe codificar, entendiéndose esta actividad como la disposición de los elementos de acuerdo con la sintaxis del metadato.

Desde otro punto de vista, la participación del usuario como contribuidor de contenidos formalmente se perfecciona a partir de la introducción de las tecnologías de la web 2.0 en los catálogos de acceso al público (OPAC, en inglés), así como la conversión de las bases de datos bibliográficas y el surgimiento de los repositorios institucionales. Así, la adopción de estas tecnologías 2.0 reconocen que la participación del usuario crea una nueva forma de compartir y enriquecer los registros en los OPAC.

De acuerdo con Maness,<sup>12</sup> se debe agregar que la biblioteca 2.0, centrada en el usuario, permite que tanto los servicios como los contenidos se enriquezcan, lo que propicia experiencias multimedia y que se vea a la biblioteca como una entidad socialmente rica. Es decir, los servicios comunitarios estarán en constante cambio y tratando de encontrar nuevos servicios que no sólo

---

11 Arlene Taylor, "A world of disappearing boundaries: traditional organization of information in an electronic environment".

12 Jack Maness, "Library 2.0 theory: web 2.0 and its implications for libraries".

sirvan para buscar y encontrar o utilizar información, sino que respondan a las necesidades emergentes de la comunidad.

Todas estas observaciones se relacionan también con discernimientos que hay con relación al fenómeno de la catalogación social, la cual se conjetura que dio inicio en el 2003, momento en que los sitios web personales e institucionales iban en aumento. Jeffries<sup>13</sup> señala que la mayoría de esos sitios empleaba el catálogo para difundir sus colecciones personales, para dar a conocer a todos los usuarios web sus objetos y conectar con otros usuarios con propósitos similares.

Por su parte, Choi y Joo<sup>14</sup> refieren que el auge de los sitios de catalogación social ha crecido en popularidad tanto para el uso personal como en el bibliotecario y este rápido crecimiento ha atraído la atención de los investigadores con el fin de analizar la situación en que se encuentra el desarrollo de estos sitios sobre el uso de las etiquetas temáticas y su incorporación en los OPAC. Asimismo, les ha interesado explorar la ruta que sigue la catalogación social en las actividades de las bibliotecas, y han encontrado que este tipo de sitios promueve las colecciones y crea listas de recomendaciones para los clubes de libros.

En el contexto que estamos revisando, el usuario como contribuidor de contenidos, advertimos que la catalogación social ha creado un sentido de comunidad compartida en donde, de acuerdo con Choi y Joo, los usuarios se sienten miembros importantes dentro de la comunidad porque comparten las mismas necesidades e inquietudes con los demás miembros.

Con el propósito de facilitar la recuperación de los ítems y compartir opiniones y comunicarse con otros usuarios, nos refieren Lim y Kim,<sup>15</sup> se han creado cinco categorías de etiquetas: las que

---

13 Scott Jeffries, "Social cataloging tools: a comparison and applications for librarians".

14 Namjoo Choi y Soohyung Joo, "Booklover's world: an examination of factors affecting continued usage of social cataloging sites".

15 Hyewon Lim y Hyoung-Joo Kim, "Item recommendation using tag emotion in social cataloging services", 180.

están relacionadas con aquellas que describen el contenido o la categorización de un objeto; las etiquetas que están basadas en el contexto, la cuales representan tiempo y ubicación en dónde fue creado el objeto; los atributos de las etiquetas que muestran propiedades del objeto; las etiquetas subjetivas, las cuales ocupan la opinión o emociones de los usuarios, y las organizacionales, que son de uso personal.

En resumen, el usuario como contribuidor de contenidos está encontrando un nicho de investigación en la actividad de los metadatos y propiciando que emerjan estándares y prácticas comunes entre la catalogación y la etiquetación social, así como la oportunidad de que los OPAC y los catálogos de los sitios web sociales puedan interactuar de manera conjunta y compartir sus datos. El etiquetado profesional y empírico deben convivir en estos tiempos en donde estamos inundados de contenidos, tanto los normalizados por las instituciones académicas como aquellos provenientes de sitios que practican la catalogación social.

#### SOBRE EL PRINCIPIO DEL METADATO SOCIALMENTE CONSTRUIDO

En este siglo XXI se ha generado una gran cantidad de investigaciones sobre lo que son y para qué sirven los metadatos, y se da por entendido que todos los que nos leen tienen conocimientos bastos sobre el tema. Para los fines del trabajo sobre la catalogación social, diremos que los metadatos son estructuras codificadas de datos que describen las características o atributos de las entidades. Asimismo hay que señalar que cualquier metadato sirve para identificar, describir y localizar redes de recursos de información.

Ahora contamos con métodos y teorías relativas a los metadatos a grado tal que podemos crear principios sobre ellos. Esta tendencia ha entretejido los conjuntos de datos vertidos por los catalogadores y usuarios, de ahí que la catalogación social surja,

como lo dice Miska,<sup>16</sup> de la suma de esfuerzo tanto de los catalogadores como los usuarios para que los recursos de información puedan relacionarse dentro de un sistema.

Los anteriores conceptos se esclarecen con lo que, en la década del setenta, mencionó Shera sobre la epistemología social, esto es, que debería proporcionarse un escenario donde la indagación de un problema complejo requiera de un estudio tanto de la sociedad como todos sus actores para que sean percibidos con relación a un entorno total.

Esta idea de Shera, puesta en el contexto de la catalogación social en este siglo XXI, nos invita a que reflexionemos respecto a que la práctica de la catalogación tradicional deba renovarse con los conocimientos empíricos de los usuarios. De ello se infiere que, con las prácticas empíricas, los OPAC de las bibliotecas puedan verse y emplearse como los de aquellos sitios sociales como, por ejemplo, LibraryThing o Amazon, respectivamente.

Ahora bien, sobre el asunto del principio del metadato socialmente construido, encontramos en la literatura especializada estudios que están a favor y otros en contra de la participación del usuario final. Los trabajos que están en contra señalan que es de vital importancia hacer diversas revisiones técnicas sobre la forma en que participarían las comunidades de usuario.

Al respecto, se dice que desafortunadamente el uso de las tecnologías de la web social<sup>17</sup> obstaculizan la creación y gestión de los metadatos, ya que se centran más en la calidad de los productos que en la colaboración que pueda haber por parte de los usuarios y los creadores de metadatos. Con base en lo anterior, se ha demostrado que hay una subestimación en la práctica de la catalogación profesional, a grado tal de invisibilizar el arduo trabajo profesional a cambio de hacer partícipe a los usuarios en este trabajo de etiquetación de contenidos.

---

16 Shawne Miska, "Social cataloging: social cataloger", 102.

17 Seth van Hooland *et al.*, "Between commodification and engagement: on the double-edged impact o...".

De las posturas que están a favor de la participación del usuario, Avery<sup>18</sup> menciona que esta democratización de los metadatos ha propiciado el surgimiento de sistemas como son las folksonomías, como un proceso de etiquetación colectiva que motiva el establecimiento de una nueva gestión de datos a través de la visión de la *indización colectiva*, que está produciendo métodos actuales para la información organizada.

Hay que mencionar, además, que este asunto de la etiquetación social lo hemos desarrollado en otros momentos con relación a las nuevas perspectivas de los sistemas de etiquetación social de los contenidos digitales.<sup>19</sup> Ahí se argumentaba que el etiquetado social sería la pieza clave inexcusable de la teoría de la web 2.0 y que tanto los sistemas de información como las plataformas digitales deberían de interactuar con el usuario, los servicios y los productos. También señalábamos que, con las experiencias adquiridas en estos últimos lustros, los usuarios no son perjudiciales en los entornos colaborativos, siempre y cuando los servicios sean los que requieren y se orienten a sus necesidades de manera efectiva.

Para volver al tema que nos ocupa, el principio del metadato socialmente construido, varios autores han argumentado que en estos momentos las bibliotecas están pasando por la etapa de la incorporación de los sistemas de metadatos de la etiquetación social, lo que ha provocado un aumento en el rol social del metadato y con ello la incorporación de sistemas expertos, tales como los vocabularios controlados. Es así como Alemu y otros más<sup>20</sup> se dieron a la tarea de entrevistar a los catalogadores, para conocer su opinión respecto al surgimiento de cuatro nuevos principios.

Los nuevos cuatro principios son: el principio del metadato enriquecido, el principio del metadato vinculado, el principio del

---

18 Joshua M. Avery, "The democratization of metadata: collective tagging, folksonomies, and Web 2.0".

19 Ariel A. Rodríguez, "Nuevas perspectivas de los sistemas de etiquetación social de los contenidos digitales".

20 Getaneh Alemu *et al.*, "Toward an emerging principle of likening socially-constructed metadata".

metadato abierto y el principio del metadato filtrado. Es importante señalar que la organización de la información desde que introdujo la actividad de los metadatos como práctica, ingresó a una corriente teórica sustentada en el estructuralismo y con ello a seguir los patrones que se iban generando a través de la percepción de las realidades, así como en la visión de la epistemología social de Shera de 1951 contra la postura de la práctica dominante de nuestros días. Este apunte se trae a colación ya que la visión epistemológica social fue concebida en relación con el control bibliográfico, como lo dice Hartel,<sup>21</sup> al desviar la atención de la bibliografía microcósmica, que se preocupaba por el acceso individual a la información, a la bibliografía desde una perspectiva macroscópica, en sintonía con los grandes flujos de conocimiento.

Entonces, el porqué de estos nuevos principios, nos refieren Alemu y otros,<sup>22</sup> es derivado de que los actuales principios y estándares son el resultado del análisis del documento con eje central, de los metadatos legibles por humanos en lugar de procesables por máquina, y lo que se está requiriendo actualmente es que se formulen metadatos para crear y aprovechar metadatos compartibles, combinables y reutilizables, un concepto de cambio que se puede lograr con la finalidad de ajustar los datos a los actuales modelos de biblioteca.

Todavía cabe señalar que los cambios en los principios, estándares y protocolos deben mirarse desde dos perspectivas: una conceptual y otra técnica. La conceptual va enfocada a los fundamentos de la contemporaneidad en los estándares de metadatos, y la técnica se relaciona con todos los aspectos emergentes de la adopción de los datos vinculados, como son los lenguajes RDF, SPARQL y OWL, por citar algunos.

---

21 Jenna Hartel, "Social epistemology as theoretical foundation for information science: supporting a cultural turn".

22 Getaneh Alemu, Bret Stevens, Penny Ross y Jane Chandler, "Linked data for libraries: benefits of a conceptual...".

Ya hemos dicho en este escrito que sin los metadatos, las colecciones impresas y digitales de la biblioteca son invisibles para los usuarios. De ahí que sean fundamentales para que se cumpla con todas esas tareas que hay asociadas a la organización y recuperación de la información. De igual forma, el propósito de que los catalogadores usen cualquier estándar de metadatos es para que aseguren que la información catalogada cuente con las propiedades necesarias para que se desplieguen en los servicios de descubrimiento de información.

Para condensar lo dicho hasta aquí, el metadato socialmente construido está influenciado en gran medida por constructos socioculturales como el idioma, la educación, el contexto de uso e intereses personales. Ello nos lleva a observar lo que señala Alemu<sup>23</sup> respecto a no sólo pensar en la etiquetación, sino en la incorporación de otros factores que influyen en los usuarios como pueden ser sus revisiones, ratings y recomendaciones, factores que han sido reconocidos en la teoría de la web 2.0 como la inteligencia colectiva.

También, la aproximación a un metadato mixto requiere del soporte de diversas interpretaciones provenientes de los usuarios. A este suceso, Alemu<sup>24</sup> le ha denominado metadato *post-hoc*, en donde los recursos de información cuentan con una serie de metadatos creados por los expertos bibliotecarios (metadatos *a priori*). En suma, un metadato construido socialmente invita a que se revise a mayor detalle el paradigma conceptual y la infraestructura técnica de la web 2.0

## CONCLUSIONES

En este trabajo se presentó una revisión, no exhaustiva, respecto a los alcances que tiene la catalogación social en los tiempos

---

23 Getaneh Alemu, "A theory of metadata enriching and filtering: challenges and opportunities to implementation".

24 *idem*, 315.



actuales que vive la organización de la información y, con ello, se hace una invitación a que no se pierda de vista la importancia que tiene trabajar de manera conjunta entre profesional y usuario en la etiquetación de los recursos de información.

Se mostraron una serie de acontecimientos históricos del porqué hemos llegado hasta aquí y la trascendencia de realizar los cambios y ajustes en lo concerniente a la apropiación que debe hacerse de las teorías y metodologías de la web 2.0, la información 2.0 y la biblioteca 2.0, con el fin de comprender que deben desarrollarse y ponerse en práctica nuevos principios sobre los metadatos, los cuales estarán más cercanos a comprender el avance en la actividad de los metadatos que serán creados bajo protocolos y técnicas del dato vinculado.

## REFERENCIAS BIBLIOGRÁFICAS

- Alemu, Getaneh. "A theory of metadata enriching and filtering: challenges and opportunities to implementation". *Qualitative and Quantitative Methods in Libraries* 5 (2016): 311-334.
- , Bret Stevens, Penny Ross, Jane Chandler. "Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data model". IFLA 2012, Helsinki. World Library and Information Congress: 78th General Conference and Assembly. Recuperado de: <http://conference.ifla.org/ifla78>.
- , Bret Stevens, Penny Ross, Jane Chandler. "Toward an emerging principle of likening socially-constructed metadata". *Journal of Library Metadata* 14 (2014): 103-129. DOI: 10.1080/19386389.2014.914775.
- Avery, Joshua M. "The democratization of metadata: collective tagging, folksonomies, and Web 2.0". *Library Students Journal* 5, 8 (February 2010).

- Breeding, Marshall. *Next-Gen Library Catalogs. The Tech Set*. New York, NY: Neal-Schuman Publishers. 2010.
- Chambers, Sally, ed. *Catalogue 2.0: the future of library catalogue*. London: Facet Publishing. 2013.
- Choi, Namjoo y Soohyung Joo. "Booklover's world: an examination of factors affecting continued usage of social cataloging sites". *JASIST* 62, 2 (2016): 3022-3023.
- Escamilla, Gloria. *Manual de catalogación descriptiva*. México: UNAM. 1988.
- Gorman, Michael. "Cataloguing in an electronic age". En *Electronic cataloging: AACR2 and Metadata for serials and monographs*, editado por Sheila S. Intner, Sally C. Tseng y Mary Lynette Larsgaard. Boca Ratón, Florida: CRC Press. 2003.
- . "¿Metadatos o catalogación? Un cuestionamiento erróneo". En *Internet, Metadatos y Acceso a la Información en Bibliotecas y Redes en la Era Electrónica*, compilado por Filiberto Felipe Martínez Arellano y Lina Escalona Ríos. México: UNAM-Centro Universitario de Investigaciones Bibliotecológicas. 2000.
- Hartel, Jenna. "Social epistemology as theoretical foundation for information science: supporting a cultural turn". *Knowl. Org.* 45, 1 (2018). DOI: 10.5771/0943-7444-2018-1-79.
- Hooland, Seth van, Eva Méndez Rodríguez e Isabelle Boydew. "Between commodification and engagement: on the double-edged impact of user-generated metadata within the Cultural Heritage Sector". *Library Trends* 54, 4 (spring, 2011): 707-720.
- Jeffries, Scott. "Social cataloging tools: a comparison and applications for librarians". *Library Hi Tech New* 10 (2008).

**Metadatos sociales: iniciativas...**

- Joudrey, Daniel N., Arlene G. Taylor, y David P. Miller. *Introduction to cataloging and classification*. Santa Barbara, California: ABC-CLIO. 2015.
- Lim, Hyewon y Hyoung-Joo Kim. "Item recommendation using tag emotion in social cataloging services". *Expert Systems with Applications* 89 (2017): 180.
- Maness, Jack. "Library 2.0 theory: web 2.0 and its implications for libraries". *Webology* 3(2). Article 25. <http://www.webology.org/2006/v3n2/a25.html>.
- Miska, Shawne. "Social cataloging: social cataloger". En *New directions in information organization*, editado por Jung-Ran Park y Lynne C. Howarth. United Kingdom: Emerald. 2013.
- Rodríguez García, Ariel Alejandro. "Nuevas perspectivas de los sistemas de etiquetación social de los contenidos digitales". En *Contenidos digitales: convergencia, conectividad, modelos y nuevas características*, coordinado por Ariel Alejandro Rodríguez, 113- 129. México: UNAM-Instituto de Investigaciones Bibliotecológicas y de la Información. 2022.
- Schottlaender, Brian E. C. "Why metadata, why me? Why now?". En *Electronic cataloging: AACR2 and Metadata for serials and monographs*, editado por Sheila S. Intner, Sally C. Tseng y Mary Lynette Larsgaard, 19-29. Boca Ratón, Florida: CRC Press. 2003.
- Smith, Gene. *Tagging people-powered metadata for the social web*. Berkeley, Ca.: New Raiders. 2008.
- Taylor, Arlene. "A world of disappearing boundaries: traditional organization of information in an electronic environment". Recuperado de: <https://sites.pitt.edu/~agttaylor/world-disap/sld014.htm> (Consultado el 12 noviembre de 2022).

## SOFTWARES

# Metodología para establecer relacionamiento automatizado de patrones comunes en testimonios escritos de víctimas del conflicto armado en Colombia

FABIÁN ORLANDO BAENA HENAO  
*Universidad de Antioquia, Colombia*

## INTRODUCCIÓN

Desde finales de la década del 50, Colombia ha estado inmersa en una violencia justificada, entre otras causas, por la lucha política entre el Estado y las guerrillas, que se fue agudizando conforme se fueron sumando otros actores bélicos tales como paramilitares, narcotráfico, e incluso el mismo Estado. Según datos compilados por el Centro Nacional de Memoria Histórica,<sup>1</sup> se estima que entre 1958 y 2012 el conflicto armado en Colombia dejó por lo menos 220 000 muertos.

Los vestigios de la guerra en Colombia han generado la victimización de otro número considerablemente alto de personas a causas que guardan relación, entre las que se encuentran la desaparición forzada, la violencia sexual, el reclutamiento de menores, y el desplazamiento forzado, que, según datos tomados a

---

1 Grupo de Memoria Histórica, “BASTA YA! Colombia: Memorias de guerra y dignidad. Informe General Grupo de Memoria Histórica”.

2013 del Registro Único de Víctimas, asciende a más de 4 744 046 personas;<sup>2</sup> estas cifras y formas pueden aumentar dado que muchos datos se encuentran en subregistros no oficiales, así como personas que no fueron reportadas.

Las experiencias recopiladas de las víctimas, de manera particular los testimonios escritos, contienen gran cantidad de información a través de patrones clave de espacio, tiempo, género, individuos, grupos, entre muchos otros; con dichos patrones, es posible establecer conexiones que pueden ayudar a entender la configuración del conflicto en factores tales como sucesos, motivos o modos, sólo por mencionar algunos, además de facilitar el esclarecimiento de la verdad y evitar la repetición. Ahora bien, procesar cada testimonio implica un considerable esfuerzo en términos de recursos y tiempo, dado que, para establecer relaciones entre ellos, se involucra el trabajo de profesionales que lo interpreten, para encontrar patrones representados a través de frecuencias de palabras con sentido semántico dentro de un corpus, que de acuerdo con la definición de Parodi,<sup>3</sup> este último término consiste en focalizar datos observables a modo de evidencia científica.

Para el investigador, establecer un relacionamiento entre los testimonios es una tarea dispendiosa cuando se presentan grandes volúmenes de información, pues al ser tradicionalmente una tarea manual, poco repetitiva y en su mayoría subjetiva, los datos generalmente no se encuentran etiquetados con estándares de organización para su consulta. Por tanto, es en este punto donde la Ingeniería y las Ciencias Sociales convergen, proponiendo metodologías para dar solución a problemas en el estudio de fenómenos sociales, tales como el procesamiento de lenguaje natural a través del uso de algoritmos,<sup>4</sup> y así lograr la identificación de

---

2 *ibidem*.

3 Giovanni Parodi, "Lingüística de Corpus: Una introducción al ámbito".

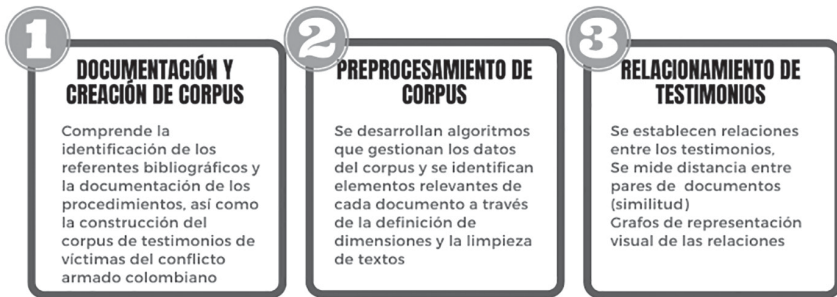
4 Manuel Montes *et al.*, "Minería de texto empleando la semejanza entre estructuras semánticas".

elementos representativos en los textos, lo que facilita la recuperación de información y la construcción de nuevo conocimiento.<sup>5</sup>

## FASES DE LA INVESTIGACIÓN

Para establecer la relación entre testimonios escritos, que en este caso particular corresponden a las narraciones de hechos violentos por parte de las víctimas del conflicto armado en Colombia, la investigación se desarrolla en 3 fases consecutivas como se describe a continuación y se muestra en la figura 1:

*Figura 1. Fases del relacionamiento automático de textos*



Fuente: Elaboración propia.

La fase uno establece la documentación de los procedimientos y la construcción del corpus de testimonios de víctimas del conflicto armado colombiano. La segunda fase se subdivide a su vez en tres etapas, pues se desarrollan algoritmos de PLN y ML que gestionan los datos del corpus previamente consolidado para identificar los elementos relevantes de cada documento y que apuntan a ser comunes entre los demás y así establecer relaciones entre ellos.

5 Erika T. Duque, “Metodología para la extracción de metadatos semánticos de textos en español...”.

La fase final corresponde al momento en el que se establecen las relaciones entre los testimonios, al identificar la afinidad entre pares de documentos y su aplicación en grafos que muestran visualmente la relación.

### Fase 1: Documentación y creación de corpus

Con el fin de tener un acercamiento a los diferentes enfoques que guardan relación con el objeto de investigación, se realizaron rastreos documentales en bases de datos bibliográficas tales como Scopus, Science Direct, Springer, Redalyc, Scielo y Google Académico. Se utilizó como punto de partida en la estrategia de búsqueda frases que identifican los temas de impacto tales como “corpus analysis”, “natural language processing” y “text mining”, asociados a los campos de la Ingeniería, Ciencias de la Computación y Ciencias Sociales. Posteriormente, luego de identificar casos de aplicación, se ajustaron los criterios de búsqueda a técnicas de ML que se acercaban más a los objetivos planteados, es allí donde los resultados arrojaron investigaciones que daban cuenta de la aplicación de “text similarity” y “clustering”.

La etapa de documentación, además de identificar los casos de aplicación, también consideró las herramientas y técnicas para el procesamiento de datos textuales, lo que dio paso a desarrollar los algoritmos en el lenguaje de programación Python, referenciado por la comunidad de programadores como ideal para realizar el análisis de los datos de la investigación; por ello, de manera particular, librerías como NLTK toolkit, Pandas y Numpy, resultaron fundamentales en el trabajo.

Por su parte, la construcción del corpus de testimonios escritos, que a una primera vista puede representar una tarea simple cuando es considerada sólo como el almacenamiento de documentos, es decir, la unificación de archivos de texto en una misma ubicación conformando una base de datos textual, realmente es una actividad que reviste mayor complejidad desde la misma consecución de los testimonios, pues aunque sólo por mencionar algunos, los procesos explícitos en la Ley de Reparación de



Víctimas<sup>6</sup> o la Jurisdicción Especial para la Paz JEP<sup>7</sup> consideran registrar los testimonios de las víctimas, estos tienen limitaciones para su uso público al tratarse de información sensible, tanto por los datos que consignan, como por el derecho a la privacidad y confidencialidad; todo ello sumado a los trámites para la autorización de uso.

Así pues, aunque se sabe que son varias las instituciones que cuentan con gran número de documentos que recogen las voces de las víctimas (entre ellas las ya mencionadas), se presentan barreras de uso, por lo que fue necesario rastrear testimonios que cumplieran con las características del corpus y estuvieran disponibles públicamente. Así, el sitio web del Centro Nacional de Memoria Histórica, en la sección Podcast,<sup>8</sup> el cual aloja series de testimonios de víctimas del conflicto armado colombiano, formó pieza clave en esta fase. Sin embargo, dado que los testimonios se encuentran en formato de grabaciones sonoras, particularmente podcast, fue necesario realizar la transcripción de tales archivos a texto; para lo cual, los archivos fueron extraídos en formato audio y posteriormente transcritos en procesadores de texto utilizando herramientas de conversión automática de audio a texto (Speech to Text STT). Los archivos resultantes fueron finalmente revisados y ajustados en un proceso de revisión manual cuando se encontraban caracteres no coincidentes o fuera de contexto.

## Fase 2: Preprocesamiento de corpus

En esta fase se desarrollaron actividades centrales en el procesamiento de la información en varias etapas, que van desde la consolidación de los elementos a identificar, hasta la implementación de algoritmos que realizan trabajos de tratamiento, limpieza y marcado

---

6 Colombia, Congreso de la República, Ley de reparación de víctimas.

7 Colombia, Unidad para la Atención y Reparación Integral a las Víctimas, “Jurisdicción especial para la paz”.

8 Colombia, Centro Nacional de Memoria Histórica, Podcasts.

### *Metadatos sociales: iniciativas...*

de los elementos representativos para establecer relaciones de los datos en el corpus.

#### *Dimensiones de relacionamiento*

Las dimensiones corresponden a los elementos centrales sobre los que se establecen las relaciones, es decir, los temas que agrupan y representan coincidencia entre cada testimonio. Al comprender la naturaleza de los testimonios, se definen 4 dimensiones: Entidades, Afectaciones, Temporalidad, Georreferenciación.

- *Entidades.* Esta dimensión comprende nombres de personas, grupos o instituciones sobre los que recaen las acciones, tales como grupos armados, víctimas, victimarios, entre otros. Entre las unidades que componen esta dimensión se encuentran palabras como “civiles, familia, compañero, hermano, guerrillero”.
- *Afectaciones.* Esta dimensión corresponde a las marcas que dejan los vestigios de los hechos violentos en las víctimas, las cuales pueden ser de tipo económicas, sociales, físicas o psicológicas. Entre las unidades que integran esta dimensión se pueden encontrar palabras como “terror, estrés, psicosis, asesinato”.
- *Temporalidad.* En esta dimensión se enmarcan las fechas, épocas o periodos. Se pueden ubicar palabras tales como “viernes, marzo, año”.
- *Georreferenciación.* Esta dimensión comprende referencias geográficas o de espacio, tales como ciudades, barrios, sitios, parajes, entre otros. Algunas de las palabras que se encuentran en esta dimensión pueden ser “Medellín, bodega, acera, vía pública”.

Para cada una de las dimensiones se realizó un trabajo de marcado manual en un documento al cual, a partir de la lectura de varios testimonios, se fueron agregando palabras y frases que identificarán el elemento o dimensión. Las unidades que identifican las dimensiones pueden ser complementadas con nuevas palabras y

frases cuando el investigador lo considere, pues adicionar nuevas unidades que describan las dimensiones aporta a la precisión de la metodología, haciendo más precisa la descripción de cada testimonio. Es de resaltar que la duplicidad de palabras en la identificación de las dimensiones no implica afectación en la metodología, es decir, no representa alteraciones de precisión si alguna palabra o frase ya se encuentra registrada previamente, pues esta etapa aplica un algoritmo que descarta duplicados, lo que facilita el trabajo sólo con palabras y frases únicas.

### *Limpieza y preprocesamiento del corpus*

Dado que los testimonios hacen parte de narraciones escritas de las víctimas que corresponden a sujetos individuales, los archivos que las condensan no cuentan una estructura definida, es decir, el texto se encuentra desestructurado, por lo que se hace necesario establecer condiciones de igualdad tales como la conversión de caracteres de mayúsculas a minúsculas, pues en el procesamiento de texto, una palabra que semánticamente sea idéntica a otra, pierde tal propiedad al representarse con caracteres diferentes ( $M \neq m$ ).

Asimismo, es preciso eliminar el ruido que puedan generar las denominadas palabras vacías o *stopwords*, las cuales representan poca relevancia o no ofrecen información al proceso, entre ellas se pueden resaltar los elementos más comunes del lenguaje, como lo son las preposiciones y artículos, ya que de omitir este paso, se puede incurrir en falsos positivos, es decir, establecer el relacionamiento de dos testimonios que no tienen ninguna conexión en sus dimensiones sólo por coincidir con elementos del tipo referenciado. A su vez, en este proceso, se expulsan tanto los símbolos o caracteres especiales como lo son asteriscos, símbolos de número, guiones, etcétera, como espacios en blanco extra que ocupan caracteres innecesarios en el texto.

### *Identificación de dimensiones*

En esta etapa, el algoritmo individualiza los elementos de cada testimonio, es decir, divide en tokens los elementos del corpus y coteja cada unidad con las unidades de las dimensiones

previamente establecidas, que al encontrar una coincidencia la separa para posteriormente construir un documento con el total de unidades por cada dimensión. El resultado final de esta etapa es que, a partir de un testimonio, se genera un nuevo documento con las unidades separadas (palabras y frases) que hacen parte de cada dimensión.

### Fase 3: Relacionamiento de testimonios

La fase final del proceso consiste en establecer las relaciones entre pares de testimonios que cuenten con elementos comunes; éstas se dan por la comparación entre coincidencias de las unidades de cada testimonio en sus dimensiones. Ello quiere decir que, si un par de testimonios cuentan con elementos comunes en una misma dimensión, de manera automática se establece un vínculo entre ambos.

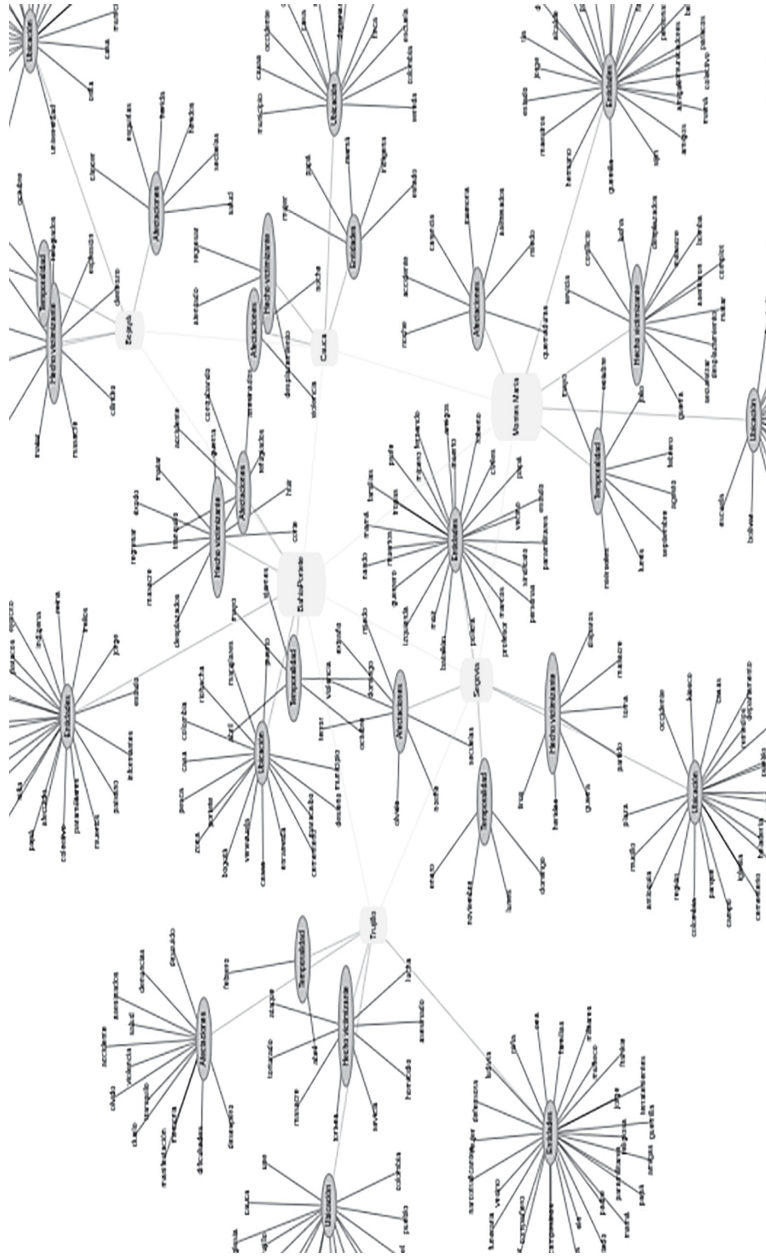
Ahora bien, si la relación de un testimonio está dada en más de un par, es importante determinar la distancia que existe entre pares de documentos, o dicho de otra forma, se considera importante determinar el nivel de semejanza de un testimonio comparado con los otros; para hallar esta similitud, el algoritmo utiliza la distancia de coseno, la cual es una medida que compara las distancias a través de valores trigonométricos considerando las frecuencias de los elementos que componen los testimonios.

Finalmente, dado que se cuenta con la relación de pares de documentos y una distancia entre cada uno de ellos, es posible representar gráficamente en un plano tales valores, lo que hace más perceptible la interpretación de los datos y sus relaciones, como se muestra en la figura 2.

## APLICACIÓN

La identificación de patrones comunes en los testimonios escritos de las víctimas del conflicto armado en Colombia permite aportar herramientas que facilitan el entendimiento de la configuración

Figura 2. Grafo de relacionamiento entre textos por dimensiones



Fuente: Elaboración propia.

### *Metadatos sociales: iniciativas...*

de la guerra en Colombia, así como los actores y su participación, todas éstas enmarcadas en procesos actuales del posconflicto como la Comisión de la Verdad, La Justicia Especial Para la Paz y la Búsqueda de Personas Desaparecidas. Asimismo se convierte en herramienta fundamental para el análisis de testimonios escritos de las víctimas, ya que facilita el procesamiento y análisis de textos desde un lenguaje natural.

El procesamiento del lenguaje natural se plantea como un ejercicio en el que la estructuración (o preprocesamiento) de textos se convierte en paso fundamental para aplicación de técnicas de minería de texto y posteriores análisis de información y extracción de eventos, que facilitan la identificación de patrones que permiten el relacionamiento, tanto entre unidades sintácticas como entre documentos de un corpus.

El procesamiento de lenguaje natural es un campo de la ingeniería y la lingüística que tiene aplicación en todas las áreas del conocimiento, pero que, de manera particular, beneficia las Ciencias Sociales, entre muchos otros casos, al encontrar, como en el caso de esta investigación, patrones comunes y semejanzas entre textos a través del etiquetado automático. Se hace entonces relevante para los profesionales de las Ciencias Sociales, la habilitación y ampliación de un corpus, así como etiquetado de texto y ontologías, que permitan facilitar las tareas y generar nuevo conocimiento por medio de técnicas computacionales como las utilizadas en este trabajo.

### REFERENCIAS BIBLIOGRÁFICAS

Colombia. Centro Nacional de Memoria Histórica. Podcasts. Centro Nacional de Memoria Histórica. <https://centrodememoriahistorica.gov.co/podcasts> (Consulta el 21 de octubre de 2021).

Colombia. Congreso de la República. Ley de reparación de víctimas. Pub. L. No. Ley 1448 de 2011 (2011).

- Colombia. Unidad para la Atención y Reparación Integral a las Víctimas. “Jurisdicción especial para la paz”. JEP. Jurisdicción especial para la paz. <https://www.jep.gov.co/JEP/Paginas/Jurisdiccion-Especial-para-la-Paz.aspx> (Consultado el 21 de octubre de 2021).
- Duque Bedoya, Erika Teresa. “Metodología para la extracción de metadatos semánticos de textos en español utilizando procesamiento de lenguaje natural: subaplicación para la identificación de contextos espaciales y temporales en textos que describan interacciones entre actores”. Medellín: Universidad EAFIT. Departamento de Informática y Sistemas. 2009. <http://hdl.handle.net/10784/1261>.
- Grupo de Memoria Histórica. “BASTA YA! Colombia: Memorias de guerra y dignidad. Informe General Grupo de Memoria Histórica”. Bogotá: Centro Nacional de Memoria Histórica. 2013.
- Montes y Gómez, Manuel, Alexander Gelbukh y Aurelio López López. “Minería de texto empleando la semejanza entre estructuras semánticas”. *Computación y sistemas* 9, núm. 1 (2005): 63-81.
- Parodi, Giovanni. “Lingüística de Corpus: Una introducción al ámbito”. *RLA. Revista de Lingüística Teórica y Aplicada* 46, núm. 1 (2008): 93-119. <https://doi.org/10.4067/S0718-48832008000100006>.

# Modelos y tecnologías para la visualización de ontologías terminológicas en el contexto de la web semántica

ADRIANA SUÁREZ SÁNCHEZ

*Universidad Nacional Autónoma de México, México*

## INTRODUCCIÓN

**E**n las últimas décadas las ontologías terminológicas han llamado la atención de los bibliotecólogos, debido a sus ventajas en la representación, organización y recuperación del conocimiento, la información y los recursos de información. Las ontologías poseen un rol primordial en las bibliotecas y otras unidades de información,<sup>1</sup> debido a sus funciones de esquematización del universo del saber, control del vocabulario y en el tratamiento temático de dominios de conocimiento y recursos de información.

Tradicionalmente, la Bibliotecología ha desarrollado diversos vocabularios controlados y a principios de milenio las ontologías se agregaron como uno más,<sup>2</sup> compartiendo características y funciones con encabezamientos de materia, clasificaciones bibliográficas y tesauros. Aunque las ontologías comparten características con la gran mayoría de los sistemas antes señalados, también se

---

1 David Stuart, *Practical ontologies: for information professionals*, 49.

2 Gail Hodge, *Systems of knowledge organization for digital libraries: beyond traditional authority files*, 1.



diferencian, pues poseen una estructura más compleja, vinculada con contextos digitales en los que, mediante interfaces, se despliegan como estructuras gráficas-visuales.

La última década ha sido un periodo de gran incremento en la investigación sobre ontologías terminológicas en asociación con las necesidades de abordar el problema terminológico-conceptual en los ámbitos documentales y el control del vocabulario en la organización de la información bajo los preceptos de la web semántica. Durante la primera década de presente siglo, la literatura sobre ontologías despegó<sup>3</sup> y de esos años a la fecha su investigación tanto teórica como aplicada ha sido constante desde perspectivas disciplinares diversas que integran las ciencias computacionales, las ciencias de la información y las ciencias documentales.

Desde la perspectiva bibliotecológica, la investigación sobre ontologías cuenta con numerosos trabajos que han tratado principalmente sus aplicaciones como herramientas para representar el conocimiento de un dominio y su potencial para proporcionar conceptualizaciones compartidas. Empero, como señalan Netto y Lima,<sup>4</sup> los estudios sobre el tema han dejado de lado el estudio de su estructura y su consecuente visualización en el ámbito de la búsqueda y recuperación de la información. Ramakrishnam y Vijayan<sup>5</sup> señalan que la investigación sobre visualización y navegación en bibliotecas digitales a partir de ontologías requiere atención, en vista de que, una vez creada la ontología, emergen preguntas sobre cómo será visualizada y empleada en una interfaz.

Después de considerar lo anterior, el presente capítulo tiene por objetivo:

- 
- 3 Francisco J. García, “El nacimiento y el despegue de la investigación moderna sobre ontologías”, 122.
  - 4 Cristiane Mendes Netto y Gercina Â. Lima, “Visualização de ontologias: estudos e perspectivas”, 59.
  - 5 Sivakumar Ramakrishnam y Arivoli Vijayan, “A study on development of cognitive support features in recent ontology visualization tools”, 598.

- Analizar los métodos y tecnologías más actuales para la visualización de ontologías terminológicas, formalizadas mediante editores ontológicos y codificadas bajo estándares RDF (Marco de Descripción de Recursos) y OWL (Lenguaje de Ontologías Web).

El supuesto del que parte la investigación asume que la visualización de ontologías cuenta con numerosos modelos y tecnologías bajo los cuales las herramientas pueden observarse gráficamente en interfaces de forma terminológica, jerárquica y relacional, en respuesta a tendencias recientes de despliegue de los vocabularios controlados en el contexto de la web semántica. Un segundo supuesto asume que los vocabularios controlados con rasgos de exploración visual se vislumbran esenciales en los sistemas de información web, lo que permite la navegación en entramados conceptuales semánticos que favorecen la búsqueda y recuperación de información.

## METODOLOGÍA

Para cumplir con el objetivo establecido, la metodología de estudio fue cualitativa con rasgos exploratorios-descriptivos. Tal acercamiento permitió obtener datos sobre el tema de interés, descubrir información sobre los modelos y tecnologías para la visualización de ontologías y definir las características de cada uno de estos. Como técnicas de estudio se utilizaron hermenéutica del discurso y análisis de casos.

- Hermenéutica del discurso: permitió examinar, de manera crítica, textos teóricos sobre el tema, con miras a identificar las tendencias sobre los modelos y tecnologías para la visualización de ontologías.
- Análisis de casos: permitió detectar modelos específicos y tecnologías que se han empleado en la visualización de ontologías generadas en dominios diversos (ciencias biomédicas, arqueología, derecho, espacios culturales, etcétera). Para

recuperar estos casos se emplearon los descriptores: “Ontologías-Visualización” y “Ontologías-Modelos gráficos” tanto en español como en sus correspondencias en inglés en los campos de búsqueda de título y temas de tres bases de datos bibliotecológicas: *Library Science Abstracts*, *Library Science Source* y *Library Science Database*.

Dado que las tecnologías para la visualización de ontologías son un asunto en constante cambio, en el estudio sólo se consideraron textos teóricos y casos publicados en el periodo temporal 2015-2022. Tanto los textos teóricos como los casos detectados fueron analizados con ayuda del programa MAXQDA, que permitió el marcado de etiquetas relevantes sobre qué son las ontologías terminológicas, cuál es su estructura, los editores para su diseño, los modelos de visualización que posibilitan y las tecnologías asociadas a este proceso de despliegue visual. Una vez analizada la literatura seleccionada, se contabilizaron los modelos de visualización de ontologías y las tecnologías detectadas. Los resultados del estudio se presentan en tres apartados: en el primero se habla sobre las ontologías y su estructura que, a la vez, incide en su modelado gráfico; en el segundo se tratan los estándares de modelado ontológico, que son el pilar para su estructuración en red y en el tercero se presentan los modelos y tecnologías actuales para concretar visualizaciones de ontologías.

## RESULTADOS

### Ontologías terminológicas: definición, estructura y visualización

Las ontologías terminológicas son un vocabulario controlado o lenguaje documental<sup>6</sup> que comparte funciones y características

---

6 Rafael Pedraza *et al.*, “Web semántica y ontologías en el procesamiento de la información documental”, 574.

con sistemas como encabezamientos de materia, tesauros, clasificaciones bibliográficas. Pese a ello, según Peña Vera, las ontologías están modificando las formas tradicionales de organización y representación del conocimiento:

Las ontologías agrupan términos junto con sus definiciones y relaciones, en entornos siempre informatizados. Para que el computador pueda interpretarlos debe contar con una síntesis, es decir, símbolos y reglas de interpretación y relación; también, con una semántica que le proporcione el significado de las expresiones construidas [...] En términos generales, implican una evolución de los instrumentos de ORC [Organización y Representación del Conocimiento], y un cambio de paradigma, en cuanto al seguimiento de los estándares tradicionales y establecimiento de asociaciones entre conceptos que no se manejaban anteriormente.<sup>7</sup>

En cuanto a funciones, las ontologías, a semejanza de otros vocabularios controlados o lenguajes documentales, posibilitan explicitar temáticamente dominios de conocimiento especializados: “química”, “dulces típicos mexicanos”, “suculentas”, etcétera. Aunque comparten rasgos y funciones con otros sistemas (encabezamientos de materia, tesauros, clasificaciones), su estructura es más compleja y se adhieren a los principios de la web semántica, como estructuras totalmente comprensibles tanto para los humanos como para las computadoras. Otra gran diferencia entre las ontologías y los lenguajes documentales de mayor tradición en el ámbito bibliotecario reside en sus posibilidades de visualización: poseen una estructura particular y se codifican en estándares que les permiten crear visualizaciones gráficas y navegables.

Abbas<sup>8</sup> enfatiza este despliegue gráfico cuando menciona que se comparan con otros vocabularios controlados desarrollados para su uso en un dominio, disciplina o comunidad de práctica específicos. Sin embargo, en lo que difieren de los vocabularios

---

7 Tania Peña Vera, *Organización y representación del conocimiento: incidencia de las tecnologías de la información y comunicación*, 192.

8 June Abbas, *Structures for organizing knowledge*, 165.

## Metadatos sociales: iniciativas...

controlados, es en las diversas formas en que se representan: en lugar de utilizar una estructura jerárquica estricta para mostrar las relaciones, la ontología puede estructurarse utilizando un gráfico no lineal que bosqueja los términos y las relaciones a partir de facetas.

En comparación con otros lenguajes documentales, las ontologías poseen una estructura rica que retoma el control del vocabulario de los encabezamientos de materia, la categorización de los sistemas de clasificación y las relaciones semánticas de los tesauros.

Entre sus componentes encontramos (figura 1):

Figura 1. Elementos compositivos de las ontologías terminológicas

Conceptos	• Unidades de pensamiento concretadas en términos.
Clases	• Consisten en el orden que le es asignado a las entidades a partir de sus características.
Instancias	• Cada uno de los individuos particulares e inconfundibles que forman la ontología.
Atributos	• Características de una entidad o variables que las definen.
Relaciones	• Vinculaciones significativas entre objetos o entidades conceptuales.
Axiomas	• Proposiciones que restringen los elementos conceptuales de las ontologías.

Fuente: elaboración propia, 2022.

A partir de tales elementos compositivos, las ontologías permiten un modelado complejo que, por una parte, hace uso de las tecnologías visuales de la web y, por otro, retoma los preceptos de FRASAD-Requisitos Funcionales para Datos de Autoridad de Materia,<sup>9</sup> que promueven la creación de vocabularios controlados más semánticos,

---

9 Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas, *Requisitos funcionales para datos de autoridad de materia: un modelo conceptual*, 15, 20 y 26.

apuntalados en entidades, atributos y relaciones. De tal forma, en comparación con los lenguajes documentales que son meras listas alfabéticas, los modelos semánticos permitirán: el control del vocabulario, la explicitación temática de los dominios de conocimiento en estructuras jerárquicas y el establecimiento de relaciones específicas (causa/efecto, todo/parte, materia prima/producto, etcétera) entre elementos conceptuales de un campo de saber.

De igual forma, las visualizaciones de las ontologías, adecuadamente denominadas en red, se observan altamente útiles en la búsqueda y recuperación de información, de forma exploratoria, paliando, incluso, algunos problemas que se presentan entre las solicitudes de búsqueda del usuario y los encabezamientos de materia o descriptores del vocabulario controlado que los bibliotecólogos empleamos en la organización temática de los recursos.

### Estándares tecnológicos para la visualización de ontologías

La visualización de estructuras de datos en la web es una tendencia actual en la que los sistemas cuentan con numerosas interfaces que ofrecen a los usuarios. Las ontologías terminológicas se adhieren a tales preceptos de visualización en la web; no obstante, su modelado gráfico es un poco más complejo que las bases de datos lineales: una ontología es algo más que una jerarquía de conceptos, pues está enriquecida con relaciones de rol entre conceptos y cada concepto tiene varios atributos relacionados.

En este orden de ideas, las ontologías se afianzan en una lógica informática fundamentada en el Marco de Descripción de Recursos (RDF), el Lenguaje de Ontologías Web (OWL) y el Lenguaje de Marcado Extensible (XML), estándares que les posibilitan soportar la inferencia computacional mediante descripciones y reglas.

RDF es un estándar para describir los recursos y sus relaciones en la web. Contempla la estructuración de datos bajo el principio entidad-relación, cuya idea es hacer declaraciones de los materiales que imitan la sintaxis del lenguaje natural. Sirve para manifestar declaraciones que sean legibles por los agentes artificiales y el modelado gráfico o en red de estructuras de conocimiento.

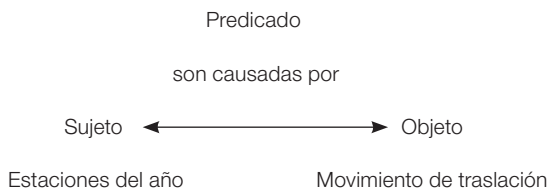
### **Metadatos sociales: iniciativas...**

El modelo RDF se cimienta en tres elementos:<sup>10</sup>

- Recursos: los recursos son entidades de información que poseen un título, autor, lugar de publicación, etcétera. Siempre son objetos digitales (texto, audio, imágenes, páginas web, objetos multimedia, etcétera) que deberán estar albergados bajo un Identificador de Recursos Uniforme (URI), Localizador de Recursos Uniforme (URL) o Dirección Web (WA).
- Atributos: son características informativas del recurso (clases, propiedades o relaciones) usadas para describirlo.
- Declaraciones: son vinculaciones entre el recurso, que consideran una propiedad nombrada y un segundo recurso, que es considerado una entidad asignada.

Las tripletas RDF, usadas en la codificación de ontologías terminológicas, son declaraciones sujeto-predicado-objeto expresadas del modo siguiente (figura 2):

**Figura 2.** Tripletas RDF en ontología terminológica



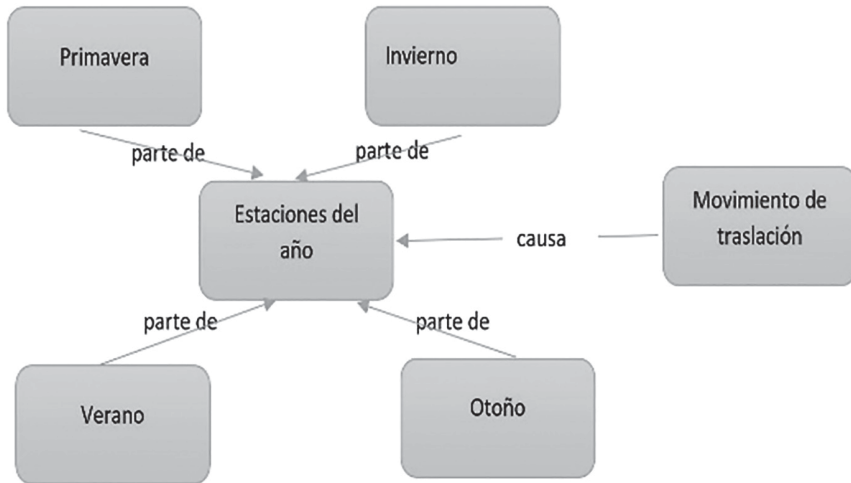
Fuente: elaboración propia, 2022.

Tal estructuración permite el trazado de redes o nodos semánticos que se vinculan progresivamente para construir entramados de datos (figura 3):

---

10 Dieter Fensel *et al.*, *Spinning the semantic web: bringing the World Wide Web to its full potential*, 116.

Figura 3. Tripletas RDF que configuran nodos y redes en ontología terminológica



Fuente: elaboración propia, 2022.

Además de codificar de modo comprensible tripletas para las computadoras, la codificación de ontologías en tripletas RDF forma un grafo dirigido y etiquetado, en el que las aristas representan el enlace entre dos recursos. Tal vista de grafo se utiliza a menudo en explicaciones visuales en el contexto de la web semántica y los datos enlazados.

El segundo estándar para la codificación de ontologías, que incide directamente en su visualización es OWL. El Lenguaje de Ontologías Web (OWL) es un estándar para publicar y compartir los datos que forman una ontología. En agregación con el entorno XML y RDF posibilita la comprensión de estructuras significativas para las máquinas. Actualmente, es el lenguaje de ontologías más empleado y se encuentra disponible en la mayoría de los editores ontológicos comerciales y de fuente abierta.



OWL se forma de tres niveles o capas estructurales:<sup>11</sup>

- Nivel de datos: las especificaciones OWL ayudan a representar clases, instancias y propiedades a través de las descripciones de las entidades.
- Nivel de esquema: las especificaciones de OWL permiten la especificación de relaciones entre clases: subclases, desuniones y equivalencias.
- Nivel de interrogación: los lenguajes de interrogación basados en OWL, como OWL-QR se encuentran en etapas de desarrollo hacia la optimización más alta. También existen aproximaciones que se combinan con SPARQL para mejorar los resultados.

OWL permite codificar conceptualizaciones explícitas y formales sobre los dominios de conocimiento. Sus principales características son: sintaxis definida, razonamiento eficiente, semántica formal, poder expresivo y expresiones adecuadas. En la construcción de ontologías, OWL se complementa con RDF, pues mientras RDF es una estructura sintáctica tripartita para establecer clases, propiedades, restricciones, etcétera, OWL es un lenguaje para la codificación de tales tripletas en una sintaxis comprensible para las máquinas, en general, y los programas de despliegue visual que, posteriormente, concretarán las ontologías terminológicas.

Mediante OWL es posible modelar las siguientes inferencias asociadas a un dominio de conocimiento:

- Miembros de clases: si A es una instancia de la clase B y B es una subclase de C, entonces se puede inferir que A es una instancia de C.
- Equivalencia de clases: si una clase A es equivalente a la clase B y la clase B es equivalente a la clase C, entonces A es equivalente a C también.

---

11 Vipul Kashyap *et al.*, *The semantic web*, 37 y 296.

- Clasificación: si se ha declarado que una propiedad o varias son condiciones para que un miembro sea agregado a una clase A y un miembro X cumple con el requisito, se agregará inmediatamente como parte de la clase.
- Desarticulación: las clases deberán diferenciarse para ser comprensibles en los sistemas recuperadores de información. Así, A deberá establecerse como una clase separada de B y tal separación será heredada a todas sus subclases.
- Implicaciones: supongamos que hemos declarado que X es instancia de la clase A y la clase A está desarticulada de la clase B, entonces la instancia X automáticamente se colocará en desarticulación con la clase B. La asociación y desarticulación en las ontologías es un proceso metódico que tiene por objetivo definir sentencias lógicas en los sistemas computacionales.
- Alcance de propiedades: cada propiedad deberá asociarse a la clase exacta que la contiene. Toda propiedad asignada a la clase A es heredada por todas sus subclases.
- Propiedades especiales: asignadas a entidades específicas, pueden ser propiedades transitivas, únicas o inversas.

Finalmente, XML o el Lenguaje de Marcado Extensible es el tercer estándar modelador de ontologías. Fue desarrollado en 1999 por el World Wide Consortium y es un metalenguaje para almacenar datos. En primera instancia, es un patrón para la expresión de la información en la web; no obstante, va más allá, pues constituye una norma para el intercambio de información en diversas plataformas.<sup>12</sup>

Las tripletas RDF unidas a la potencialidad de OWL y XML abren un universo de oportunidades para la explicitación temática de campos de conocimiento en agentes computacionales, mediante el establecimiento de clases, subclases, instancias, propiedades de clases, propiedades de subclases, propiedades de instancias, vinculaciones, desarticulaciones, etcétera.

---

12 World Wide Consortium, *Extensible Markup Language (XML)*.

## Modelos y programas para la visualización de ontologías terminológicas

El proceso de visualización de la información consiste en la transformación de datos abstractos en imágenes, refuerza la comprensión y los conocimientos a través de modelos gráficos y permite a los usuarios examinar en detalle las estructuras, propiedades, patrones, tendencias y tipos de interacción de los datos. En adhesión, toda visualización gráfica proporciona una perspectiva cualitativa de conjuntos de datos estructurados que ayuda a los usuarios a identificar bloques, elementos conceptuales y relaciones sobre las entidades temáticas.

Las visualizaciones de ontologías buscan “aumentar la capacidad de percepción humana para que, a partir de las relaciones espaciales mostradas, el usuario interprete y comprenda la información presentada y, además, deduzca nuevos conocimientos”.<sup>13</sup> Los modelos y programas para la visualización de las ontologías terminológicas han sido un asunto central de las ciencias computacionales en las últimas dos décadas en las que estándares como XML, RDF y OWL han sentado un precepto. Forman parte de una reciente tendencia de los vocabularios controlados, bajo la cual

[...] el uso de simples vocabularios controlados se ha ido desplazando progresivamente por el uso de modelos de conocimiento [...] los modelos de conocimiento almacenados en papel (taxonomías, tesauros) de las bibliotecas y otras instituciones se han informatizado y transformado en modelos ontológicos más formales para proporcionar un mayor nivel de semántica.<sup>14</sup>

En la organización temática de dominios de conocimiento, las posibilidades tecnológicas que integran las ontologías son una gran

---

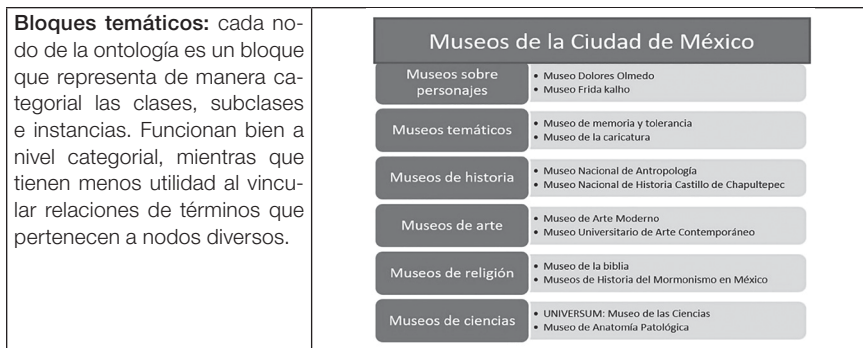
13 Cristiane Mendes y Gercina Â. Lima, “Visualização de ontologias: estudos e perspectivas”, 60.

14 Javier Lacasta *et al.*, *Terminological ontologies: design, management and practical applications*, viii.

ventaja pues, a diferencia de los listados de temas y tesauros que consisten en listados alfabéticos con algunos tipos de relaciones, éstas generan estructuras desplegadas y navegables.

A partir del estudio llevado a cabo, fue posible detectar diez modelos de visualización de ontologías. Tales propuestas fueron retomadas de Halilaj, Petersen, Grangel-González, Lange, Auer, Coskun, Lohmann,<sup>15</sup> Lembo, Pantaleone, Santarelli, Fabio Savo<sup>16</sup> y Saghafi.<sup>17</sup> A continuación se muestran los diez modelos en la figura 4.

Figura 4. Diez modelos de visualización de ontologías



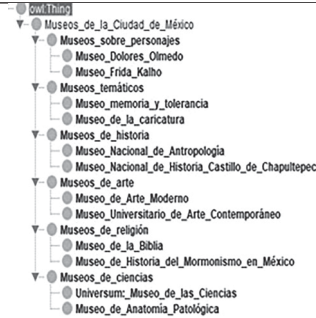
15 Lavdim Halilaj *et al.*, “VoCol: an integrated environment to support version-controlled vocabulary development”, 303.

16 Domenico Lembo *et al.*, “OWL drawing with the graphol visual ontology language”, 573.

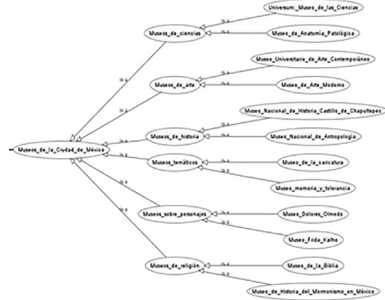
17 Arash Saghafi, “Visualizing ontologies. A literature survey”, 204.

**Metadatos sociales: iniciativas...**

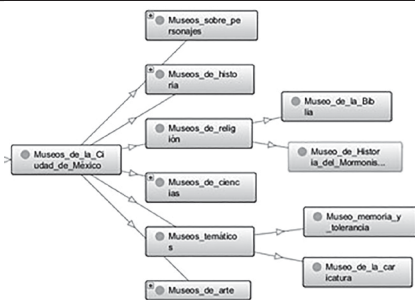
**Estructuras jerárquicas:** son estructuras en las que se observan los términos y las clases que forman una ontología. Permiten compendiar con precisión la estructura categorial de un dominio de conocimiento.



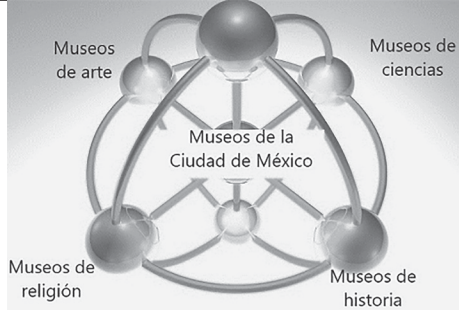
**Grafos 2D:** se forman de árboles de nodos que ofrecen una buena visión de los términos, la jerarquía y las relaciones que forman una ontología. Son esquemas fijos para visualizar ontologías cortas.



**Grafos 2D con ampliaciones:** se forman de árboles de nodos que se amplían o reducen, según las necesidades de quienes los exploran. Suelen mostrar ventajas en la navegación y exploración de nodos específicos y ontologías amplias.

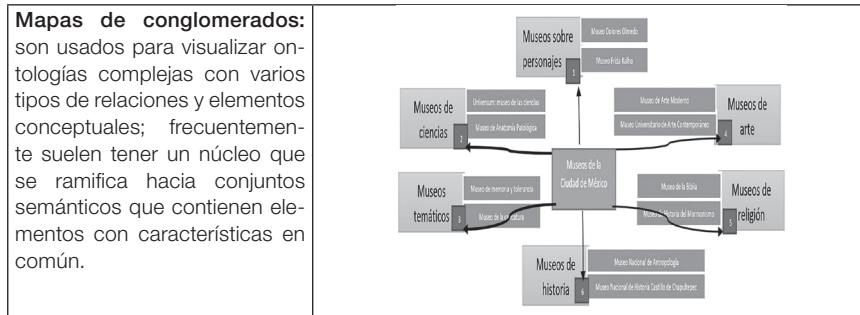


**Grafos 3D:** son mapas de nodos que se despliegan de manera tridimensional. Su desplazamiento es algo más complejo que los grafos 2D. Posibilitan mayor visualización de las relaciones presentes en una ontología.



<p><b>Visualizaciones radiales:</b> facilita la comprensión visual de los términos y clases de la ontología. En contraste, no suelen mostrar las relaciones presentes en la herramienta. Facilitan la exploración visual de grandes ontologías en segmentaciones temáticas.</p>	
<p><b>Círculos parcelados:</b> son conjuntos semánticos anidados que se ramifican según rasgos de los elementos que representan. Son de suma utilidad para visualizar ontologías que contienen muchas clases y elementos anidados en éstas.</p>	
<p><b>Parcelas de conjuntos:</b> son grafos anidados que permiten dividir jerárquicamente las clases de una ontología en conjuntos que de manera visual permiten contextualizar las entidades y las clases. Tienen pocas ventajas en la presentación de relaciones.</p>	
<p><b>Ruedas gráficas:</b> a partir de los axiomas se crean ruedas gráficas en las que se muestran los conceptos, las clases y las relaciones de la ontología. La ontología se asume como un conjunto de capas nodales que se despliegan de forma concéntrica.</p>	

## Metadatos sociales: iniciativas...



Fuente: Elaboración propia a partir de los autores mencionados previamente.

Entre los modelos para visualizar ontologías, algunos se observan ideales para el modelado de dominios de conocimiento pequeños que pueden ser representados mediante círculos parcelados, parcelas de conjuntos o bloques temáticos. Sin embargo, en la medida que las ontologías crecen, incorporan nodos y relaciones, su representación exige modelos más complejos como estructuras jerárquicas, gráficos 2D con ampliaciones, grafos 3D, mapas de conglomerados o visualizaciones radiales.

Una vez revisados los modelos con los que actualmente contamos para visualizar ontologías, es necesario abundar sobre los programas que los materializan. Como señalan Sicilia Nemirovski y Nolle, en los últimos años “se han desarrollado varias herramientas para ayudar a los expertos a especificar los mapeos entre las fuentes de datos y las ontologías”.<sup>18</sup>

A partir del estudio realizado se encontró que desde 2015 hasta la actualidad han surgido diversos programas o extensiones para la visualización de ontologías estructuradas en RDF y OWL. De manera precisa, se detectaron 32 herramientas (figura 5):

Algunas de las aplicaciones tecnológicas son programas y otras son extensiones. Los programas funcionan de manera independiente, una vez que se carga la base de datos u ontología

18 Álvaro Sicilia *et al.*, “Map-On: a web-based editor for visual ontology mapping”, 970.

Figura 5. Programas o extensiones para visualizar ontologías terminológicas



Fuente: Elaboración propia, 2022.

terminológica que se desea visualizar. En comparación, las extensiones son aplicaciones diseñadas para usar en combinación con un programa específico (editor de ontologías) en el que la ontología ha sido modelada previamente. Entre los programas es posible mencionar Gephi y Memograph, programas generales para modelar datos estructurados. Entre las extensiones es posible mencionar OntoIndex y OntoPlot, que son extensiones del programa R, y Ontograph y Ontoviz, que son extensiones del editor Protégé.

Como resultado de la investigación, también se identificó que en la visualización de ontologías se encuentran implicados diversos elementos: el editor en el que codifica la ontología, el uso adecuado, lógico y coherente de las tripletas RDF, la correcta codificación de la ontología en lenguaje OWL y el programa visualizador para su despliegue gráfico en una interfaz. Al considerar lo anterior, cuando construimos una ontología debemos planear su



visualización desde el inicio del proyecto, así estaremos en posibilidad de enlazar adecuadamente nuestras tecnologías de codificación y despliegue gráfico.

Por último, es importante señalar que cada programa o extensión para la visualización de ontologías responde a uno o varios de los modelos de despliegue. OWLViz genera grafos 2D, OntoGraph nos ofrece grafos 2D extensibles, CropCircles deriva círculos parcelados, mientras que OntoSphere produce grafos 3D. La elección de algún programa para visualizar este tipo de vocabularios controlados o lenguajes documentales es decisión de los creadores, quienes determinan cuál es la mejor forma de despliegue de la herramienta, en función del sistema de información en el que la ontología se integrará y las necesidades de los usuarios.

## CONCLUSIONES

En los últimos tiempos, muchos campos del conocimiento están construyendo ontologías terminológicas para aclarar o establecer la estructura conceptual y relacional de su área de saber. De igual modo, han detectado que los rasgos gráficos de las ontologías posibilitan una comprensión modelada del área en la que usuarios expertos y en formación aprenden sobre el campo. Como señalan Gahegan, Agrawal, Banchuen y DiBiase,<sup>19</sup> varias organizaciones en ámbitos tan diversos como la cocina o la física nuclear han empezado a utilizar las ontologías para describir el contenido de su área, colecciones, datos, etcétera, de manera gráfica, en una tendencia que hace uso de interfaces visuales.

En el caso de las bibliotecas y repositorios digitales, las ontologías terminológicas y su visualización son un asunto de interés. La teoría esbozada por autores como Bikakis y Sellis,<sup>20</sup> Cheng y

---

19 Mark Gahegan *et al.*, "Building rich, semantic descriptions of learning activities to facilitate reuse in digital libraries", 84.

20 Nikos Bikakis y Timos K. Sellis, "Exploration and visualization in the web of big, linked data: a survey of the state of the art", 1.

Chou<sup>21</sup> y Greene, Richardson y Turro<sup>22</sup> permite afirmar que, dado que las bibliotecas digitales se organizan en torno a metadatos estandarizados, los despliegues gráficos de las ontologías terminológicas pueden vincularse a los metadatos temáticos de una colección de recursos de información digitales. En tales casos, las ontologías terminológicas servirán como un vocabulario controlado a la vez que se concretará en una estructura gráfica que permitirá localizar contenidos relevantes asociados con el contenido tópico de las entidades bibliográficas.

Se concluye que las ontologías terminológicas han recibido mucha atención con el auge de la web semántica como forma de dar un significado bien definido a la información. Hoy en día se utilizan en muchos contextos diferentes para estructurar y organizar el conocimiento, la información y aun los recursos de información y, en consecuencia, su estudio es un asunto de suma importancia para los bibliotecarios, quienes día a día se enfrentan a mayores retos asociados a la organización del conocimiento y la información en la red.

## REFERENCIAS BIBLIOGRÁFICAS

June Abbas. *Structures for organizing knowledge*. Nueva York: Neil Schuman. 2010.

Bikakis, Nikos y Timos K. Sellis. "Exploration and visualization in the web of big linked data: a survey of the state of the art". *ArXiv* (2016): 1-8.

---

21 Yu-Jung Cheng y Shu-Lai Chou, "Using digital humanity approaches to visualize and evaluate the cultural heritage ontology", 83.

22 Daniel Greene *et al.*, "OntologyX: A suite of R packages for working with data", 1104.

**Metadatos sociales: iniciativas...**

- Cheng, Yu-Jung y Chou Shu-Lai. "Using digital humanity approaches to visualize and evaluate the cultural heritage ontology". *The Electronic Library* 40, núm. 1-2 (2022): 83-98. <https://doi.org/10.1108/EL-09-2021-0171>.
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas. *Requisitos funcionales para datos de autoridad de materia: un modelo conceptual*. Washington: IFLA. 2010.
- Fensel, Dieter, James A. Hendler y Henry Lieberman. *Spinning the semantic web: bringing the World Wide Web to its full potential*. Londres: Massachusetts Institute of Technology. 2005.
- Gahegan, Mark, Ritesh Agrawal, Tawan Banchuen y David DiBiase. "Building rich, semantic descriptions of learning activities to facilitate reuse in digital libraries". *International Journal of Digital Libraries* 7 (2007): 81-97. <https://doi.org/10.1007/s00799-007-0021-x>.
- García Marco, Francisco Javier. "El nacimiento y el despegue de la investigación moderna sobre ontologías". *Anuario ThinkEPI* 2 (2008): 122-125.
- Greene, Daniel, Sylvia Richardson y Ernest Turro. "OntologyX: A suite of R packages for working with data". *Bioinformatics* 33, núm. 7 (2017): 1104-1106. <https://doi.org/10.1093/bioinformatics/btw763>.
- Halilaj, Lavdim, Niklas Petersen, Irlán Grangel-González, Christoph Lange, Sören Auer, Gökkan Coskun y Steffen Lohmann. "VoCol: an integrated environment to support version-controlled vocabulary development". En *European Knowledge Acquisition Workshop*, editado por E. Blomqvst, P. Ciancarini, F. Poggi y F. Vitali, 303-319. Berlín: Springer. 2016.
- Hodge, Gail. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: The Digital Library Federation Council on Library and Information Resources. 2000.

- Kashyap, Vipul, Christoph Bussler y Matthew Moran. *The semantic web*. Heidelberg: Springer-Verlag. 2008.
- Lacasta, Javier, Javier Nogueras-Iso y F. Javier Zarazaga-Soria. *Terminological ontologies: design, management and practical applications*. Nueva York: Springer. 2010.
- Lembo, Domenico, Daniele Pantaleone, Valerio Santarelli y Domenico Fabio Savo. "Easy OWL drawing with the graphol visual ontology language". En *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 573–576. Berlín: Springer. 2016.
- Mendes Netto, Cristiane y Gercina Ângela Lima. "Visualização de ontologias: estudos e perspectivas". *Informação & Sociedade: Estudos* 27, núm. 3 (2017): 59-72.
- Pedraza Jiménez, Rafael, Lluís Codina y Cristòfol Rovira. "Web semántica y ontologías en el procesamiento de la información documental". *El profesional de la información* 16, núm. 6 (2007): 569-578.
- Peña Vera, Tania. *Organización y representación del conocimiento: incidencia de las tecnologías de la información y comunicación*. Buenos Aires: Alfagrama. 2011.
- Ramakrishnam, Sivakumar y Arivoli Vijayan. "A study on development of cognitive support features in recent ontology visualization tools". *Artificial Intelligence Review* 41, núm. 4 (2014): 595-623. <https://doi.org/10.1007/s10462-012-9326-2>.
- Saghafi, Arash. "Visualizing ontologies. A literature survey". En *International Conference on Conceptual Structure*, editado por Olivier Haemmerlé, Gem Stapleton y Catherine Faron Zucker, 204-221. Berlín: Springer, 2016. [https://doi.org/10.1007/978-3-319-40985-6\\_16](https://doi.org/10.1007/978-3-319-40985-6_16).

***Metadatos sociales: iniciativas...***

Sicilia, Álvaro, German Nemirovski y Andreas Nolle.  
“Map-On: a web-based editor for visual ontology mapping”. *Semantic Web* 8, núm. 6 (2017): 969-980.

Stuart, David. *Practical ontologies: for information professionals*. Chicago: Neal-Schuman. 2016.

World Wide Consortium. *Extensible Markup Language (XML)*. <https://www.w3.org/TR/xml/> (Consultado el primero de Julio de 2022).

# El Objeto Virtual de Aprendizaje (OVA) como producto de apropiación social del conocimiento del Banco de Datos Terminológicos de las Ciencias de la Información

MARÍA TERESA MÚNERA TORRES  
*Universidad de Antioquia, Colombia*

## INTRODUCCIÓN

Los metadatos constituyen componentes claves de los procesos investigativos, desde la perspectiva de la elaboración de herramientas de apropiación social del conocimiento y de divulgación de las metodologías que se llevan a cabo en el desarrollo de la dinámica investigativa.

En la investigación “Banco de Datos Terminológicos de las Ciencias de la Información (BDT-CI)” se han desarrollado herramientas que permiten la difusión de las metodologías utilizadas, tales como el Objeto Virtual de Aprendizaje (OVA), mediante el cual se difunde la dinámica que se lleva a cabo para la elaboración de bancos de datos terminológicos en diferentes ámbitos del conocimiento.

Con esta investigación se pretende dar a conocer el protocolo que se emplea para la elaboración de un Objeto Virtual de Aprendizaje (OVA) dentro del contexto de la investigación anteriormente relacionada y cómo los metadatos forman parte importante dentro de esta herramienta.

## INVESTIGACIÓN “BANCO DE DATOS TERMINOLÓGICOS DE LAS CIENCIAS DE LA INFORMACIÓN (BDT-CI)”

Se trata de una investigación que presenta la opción de sistematizar la terminología especializada en el ámbito de las Ciencias de la Información, conformadas por áreas como la Archivística, la Bibliotecología, la Museología y la Documentación. Esta iniciativa investigativa se viene desarrollando por etapas que corresponden a la elaboración de la Base de Datos Terminológicos en cada una de las áreas anteriormente descritas, desde el año 2018, y con la financiación de la Universidad de Antioquia.

Es de aclarar que un Banco de Datos Terminológicos se constituye en

un producto resultante de la elaboración de un trabajo terminológico. Su función consiste en recopilar términos y organizar la información relacionada, agrupándola en fichas terminológicas que siguen un modelo uniforme, diseñado específicamente. Los datos suelen corresponder a varios ámbitos temáticos y normalmente a varias lenguas. Se presentan siempre en formato electrónico y suelen incorporar remisiones, es decir, enlaces entre sus fichas (por ejemplo, entre un término y sus sinónimos, o entre un término y sus equivalentes).<sup>1</sup>

Hasta el momento se han llevado a cabo dos etapas de la investigación, traducidas en la elaboración de la Base de Datos Terminológicos de la Archivística (BDT-A) y la Base de Datos Terminológicos de la Bibliotecología (BDT-B) y en la actualidad, se ha recibido la aprobación por parte de la Fundación Universidad de Antioquia, entidad que financiará la ejecución de la tercera etapa de la investigación que se enfoca en la elaboración de la Base de Datos Terminológicos de la Museología (BDT-M). A continuación, se registra el enlace mediante el cual se puede acceder a la consulta del BDT-CI.

<https://biblioteca.udea.edu.co/terminologia/sitio/>

---

1 M. Amor Montané March, “El trabajo terminológico sistemático: su aplicación para la elaboración de bancos de datos terminológicos”, 115.

## Qué es un Objeto Virtual de Aprendizaje (OVA)

Los OVA son nuevas opciones de enseñanza y de apropiación del conocimiento por parte de diferentes audiencias, sobre todo en el ámbito educativo superior. En tal sentido, el Ministerio de Educación Nacional de Colombia (MEN) los define como un “material estructurado de una forma significativa, asociado a un propósito educativo (en este caso para la educación superior) y que corresponde a un recurso de carácter digital que puede ser distribuido y consultado a través de la internet”.<sup>2</sup>

Por su parte, Massa y Pesado definen como la “mínima estructura independiente que contiene un objetivo, un contenido, una actividad de aprendizaje, un metadato y un mecanismo de evaluación, el cual puede ser desarrollado con Tecnologías de Información y Comunicación (TIC), lo que posibilita su reutilización, interoperabilidad, accesibilidad y duración en el tiempo”.<sup>3</sup>

El OVA, al ser una herramienta digital que se utiliza dentro del contexto de la internet, posibilita que los conocimientos que se imparten mediante esta modalidad lleguen a un número importante de personas, con lo cual todas las propuestas del saber que se incorporen dentro de esta nueva opción de aprendizaje se constituyen en productos de apropiación social del conocimiento.

Por otra parte, Cabrera, Sánchez y Rojas<sup>4</sup> anotan que un Objeto de Aprendizaje es cualquier entidad digital o no digital que puede ser usada, reusada o referenciada para el aprendizaje soportado en tecnología. En América Latina se convoca, desde el 2006, a instituciones, investigadores y docentes interesados en conformar la Comunidad Latinoamericana de Objetos de Aprendizaje (LACLO), iniciativa a la que se han ido uniendo representantes de varios países entre los que se encuentra Colombia.

---

2 Colombia, Ministerio de Educación Nacional, párr. 3.

3 Stella Massa y Patricia M. Pesado, “Evaluación de la usabilidad de un objeto de aprendizaje por estudiantes”.

4 Jaime M. Cabrera-Medina e Irlesa I. Sanchez-Medina, “Uso de objetos virtuales de aprendizaje, OVAs, como estrategia...”.



En síntesis, el OVA tiene como finalidad que, al ser utilizado como herramienta de enseñanza, los estudiantes aprendan a su propio ritmo y en forma independiente.<sup>5</sup>

### Componentes del OVA

Vega y Chica<sup>6</sup> anotan que los componentes de un OVA son los contenidos, las actividades de aprendizaje y los elementos de contextualización, los cuales tienen una estructura (externa) de información que facilita su identificación, almacenamiento y recuperación. Por su parte, Cabrera, Sánchez y Rojas (2016) complementan lo expresado anteriormente precisando que “el Objeto de Aprendizaje debe tener una estructura de información externa (metadatos) que facilite su almacenamiento, identificación y recuperación”.

En términos generales, todo OVA debe estar constituido por un contenido que se debe divulgar, con base en la definición de un propósito u objetivo preciso a alcanzar como objeto de consulta, unas actividades de aprendizaje, así como unas dinámicas de evaluación que posibiliten su uso en diferentes tiempos y contextos, además de posibilitar el ser usado en el tiempo y disposición de quienes se beneficien de este nuevo recurso del proceso de enseñanza-aprendizaje.

Además de la conformación general con la que debe contar todo OVA, es importante tener en cuenta también unos componentes fundamentales, tales como:

- a. La enseñanza como proceso mediante el cual se comunican o transmiten conocimientos especiales o generales sobre un ámbito del saber.

---

5 Luz Y. Morales *et al.*, “Guía para el diseño de objetos virtuales de aprendizaje...”.

6 Carlos A. Vega *et al.*, “Diseño y validación de un Objeto Virtual de Aprendizaje que permita el aprendizaje de heurísticas y metaheurísticas”.

- b. El aprendizaje como acción de instruirse y el tiempo que demanda esta acción en sí misma. Proceso mediante el cual una persona es entrenada para brindar una solución a determinadas situaciones. Tal mecanismo va desde la adquisición de datos hasta la forma más compleja de recopilar y organizar la información.
- c. Aprendizaje autónomo como proceso de apropiación crítica de la experiencia vital, intelectual y cultural, a partir del reconocimiento de la realidad personal y social, mediante la profundización teórica de conceptos básicos, principios explicativos y valores fundamentales, generados en forma metódica, sistemática y autorregulada para transferirlos comprensivamente a diferentes contextos y aplicarlos creativamente en la solución de problemas de la vida cotidiana, en el desarrollo de procesos académicos y en la promoción del desarrollo humano.
- d. Aprendizaje colaborativo, es decir, aquel que surge de la colaboración entre personas que hacen parte de grupos donde se comparten la discusión y la realización de trabajo en equipo y en la que se permite la interacción entre sus miembros para alcanzar sus objetivos y deja en los participantes un aprendizaje.

### Características del OVA

Morales, Gutiérrez y Ariza,<sup>7</sup> y a la luz del *Manual de buenas prácticas para el diseño de objetos de aprendizaje* de APROA,<sup>8</sup> precisan entre algunas de sus características las siguientes:

- a. **Es autocontenido:** O sea, por sí solo debe permitir lograr el objetivo para el cual fue diseñado.
- b. **Es interoperable:** Garantizar su uso en plataformas con

---

7 Martín Morales *et al.*, *op. cit.*

8 APROA *Proyecto educacional de Chile: Aprendiendo con objetos de aprendizaje. Manual...*, 5.

- distintos ambientes de programación, en tanto que su estructura debe basarse en un lenguaje de programación XML, y contar con un estándar internacional de interoperabilidad (SCORM), que garantice su utilización en plataformas con distintos ambientes de programación.
- c. **Es reutilizable:** Ser usado por diversos educadores y estudiantes en distintos contextos de enseñanza y en diferentes tiempos.
  - d. **Es secuenciable con otros Objetos:** Facilita la relación con otros OVA dentro de un mismo contexto de enseñanza.
  - e. **Es durable y actualizable en el tiempo:** Se encuentra respaldado en una estructura (repositorio) que permite incorporar nuevos contenidos y hacer modificaciones según las condiciones y los objetivos de aprendizaje.
  - f. **Incorpora vínculos hacia documentos digitales que complementen el contenido.**

## Metodología de elaboración del OVA del BDT-CI

La metodología y la tecnología, según Morales, Gutiérrez y Ariza,

[...] constituye[n] el soporte para la integración pedagógica de los contenidos, las actividades de aprendizaje y la evaluación; se relaciona[n] directamente con la construcción del OVA que será implementado en una plataforma virtual de aprendizaje o ambiente virtual de aprendizaje. Esta dimensión la proporcionan generalmente las instituciones o las diferentes plataformas de uso libre en internet, y se requiere del apoyo técnico de profesionales en el campo de la multimedia y el diseño web.<sup>9</sup>

Para el caso del OVA del BDT-CI, su elaboración se inicia con la recolección del contenido sobre todo lo relacionado con dicho banco de datos, y los procesos que se deben seguir para su

---

9 Martín Morales *et al.*, *op. cit.*, 136.

estructuración y que, a su vez, cumpla con el objetivo de ser un OVA. Se diligencia luego un formato en el cual se consideran los siguientes aspectos:

- a. Tener la convicción de que se está escribiendo para una persona que no tiene conocimiento específico en el tema y a la cual le debe quedar lo más claro posible cada uno de los pasos a seguir para desarrollar la investigación.
- b. Definición de subtemas que pueden responder al interés de una persona sobre el Banco de Datos Terminológicos.
- c. Como va a ser un interactivo, y no necesariamente un documento de lectura, se recomienda redactar el texto lo más concreto posible.
- d. Se diligencia un formulario inicial en el que se define la información fundamental que tendrá el OVA, que para nuestro caso ha sido:
  - Presentación de esta guía (antecedentes, objetivos, alcance, destinatarios, créditos de los autores y de la institución, etcétera)
  - Instrucciones de uso de esta guía
  - Conceptos básicos
  - ¿Qué son las Ciencias de la Información?
  - ¿Qué es una terminología?
  - ¿Qué son las bases de datos terminológicos?
  - Guía para la elaboración de bases de datos terminológicos
  - Definición y delimitación del trabajo
  - Preparación del trabajo
  - Elaboración del recurso terminológico
  - Supervisión del trabajo terminológico
  - Tratamiento y solución de casos problemáticos
  - Presentación del trabajo terminológico

## Estrategias formativas del OVA

- a. Vinculación a los estudiantes en formación en la dinámica investigativa del proyecto de BDT-CI
- b. Presentación del OVA en sesiones de clase de universidades nacionales e internacionales
- c. Presentación del OVA en congresos, simposios, seminarios, etcétera, en el ámbito nacional e internacional
- d. La dimensión de interactividad del OVA

## Estrategias pedagógicas del OVA

Según Cabrera, Sánchez y Rojas,

[...] los OVAS han adquirido especial trascendencia e importancia en los últimos años dada la forma como consiguen conectar los procesos educativos con las Tecnologías de la Información y la Comunicación (TIC). Además, se les considera como herramienta esencial para potenciar los procesos de educación (a tal punto que la UNESCO se ha comprometido en su análisis y desarrollo bajo el esquema de formatos de acceso abierto conocidos como Open Educational Resource -OER-). Estas estrategias pedagógicas deben guardar estrecha relación con el mundo de la informática y tienen que estar a la par del avance tecnológico. En otras palabras, la tecnología debe estar al servicio de la escuela y de los profesionales egresados de sus aulas y el docente debe contar con recursos informáticos para desarrollar sus procesos de formación.

Así, por su importancia y relevancia, se da paso al concepto de los Objetos Virtuales de Aprendizaje. En el trabajo con OVAS, en el aula o fuera de ella, se propicia el aprendizaje colaborativo. En efecto, al estar disponible el mismo recurso para todos, se favorece el intercambio de ideas y el trabajo en equipo. El estudiante aprende a su propio ritmo. Un ejemplo de aplicación donde se comparten datos es el laboratorio virtual, que incluye la comunicación síncrona entre usuarios.<sup>10</sup>

---

10 Jaime Cabrera Medina *et al.*, *op. cit.*, 5.

Estos mismos autores añaden que “desde el enfoque de modelos pedagógicos, los OVAs promueven el uso del constructivismo, lo que se manifiesta en el aprendizaje autónomo, el ejercicio de análisis de casos y pensamiento crítico”.<sup>11</sup>

A través de un recurso como éste se pretende que el usuario tenga una ruta clara de navegación que le permita comprender e incorporar un saber en específico. Por esta razón, es importante considerar elementos como:

- La interactividad que permite la inmersión del usuario en el tema a través de la toma de decisiones de navegación y el desglose de la información.
- La multimedialidad que le ofrece al lector la posibilidad de ver información que se complementa en diferentes formatos, lo que capta más fácilmente su atención. En este caso en particular, el uso de videos y audios le dio una cara más humana a un proceso técnico.
- La aplicabilidad del recurso que posibilita el uso de la información allí contenida en diferentes escenarios de la vida académica y laboral.

En particular, al consistir este recurso en la sistematización de una experiencia, se buscaba brindar con la mayor claridad la información sobre los pormenores del proceso de elaboración de un BDT con la intención de motivar a los lectores a construir un recurso similar desde la concepción de sus propias áreas del saber.

De otro lado, y de acuerdo con las orientaciones de Ude@, dependencia de la Universidad de Antioquia con la que se elaboró el OVA del BDT-CI, las estrategias pedagógicas se evidencian desde la siguiente perspectiva:

- a. Conexión de los procesos educativos con las Tecnologías de la Información y la Comunicación (TIC).

---

11 *idem*, 6.

***Metadatos sociales: iniciativas...***

- b. Herramienta esencial para potenciar los procesos de educación.
- c. Fomento del aprendizaje colaborativo, así como del intercambio de ideas y el trabajo en equipo. El estudiante aprende a su propio ritmo.
- d. Los OVAs promueven el uso del constructivismo, manifestándose en el aprendizaje autónomo, el ejercicio de análisis de casos y pensamiento crítico.<sup>12</sup>

**Estrategias de comunicación del OVA**

Se presentaron avances en eventos académicos, en sesiones de clase en asignaturas relacionadas con la terminología y áreas en las que se elaboran las bases de datos terminológicos, así como divulgación del OVA en artículos de revistas, capítulos, de libros, informe de investigación, páginas web, entre otros medios. A continuación, se presenta el enlace mediante el cual se puede acceder al OVA del BDT-CI:

<https://view.genial.ly/609ac41d2642ab0cf47b44d1>

**CONCLUSIONES**

Los Objetos Virtuales de Aprendizaje (OVA) se están convirtiendo en importantes productos de apropiación social del conocimiento, que además evidencian nuevas tendencias educativas, a través de las cuales se articulan procesos pedagógicos con las Tecnologías de la Información y Comunicación (TIC) y, de esta forma, propician dinámicas de aprendizaje acopladas al ritmo y contexto de los estudiantes.

Los metadatos, además de ser componentes de procesos bibliotecológicos de organización de la información, vienen destacando como aspectos fundamentales de los productos digitales y,

---

12 *idem.*

en especial, como herramientas que son susceptibles de ser utilizadas de manera óptima en procesos de apropiación del conocimiento en diferentes escenarios educativos.

En la investigación “Banco de Datos Terminológicos de las Ciencias de la Información, BDT-CI” se evidencia la importancia del Objeto Virtual de Aprendizaje (OVA) como producto de apropiación social del conocimiento, mediante el cual se da a conocer la metodología que se lleva a cabo para elaborar un Banco de Datos Terminológicos en diferentes ámbitos del saber y, a su vez, se constituye en una herramienta de carácter educativo que permite un aprendizaje personalizado de los conocimientos incluidos en el mismo y que propician la fácil asimilación del conocimiento.

Los metadatos, como parte de herramientas tecnológicas de difusión del conocimiento y rutinas de aprendizaje en línea, contribuyen con el desarrollo de nuevas tendencias educativas, tales como los Objetos Virtuales de Aprendizaje, los cuales a su vez se están incorporando en las dimensiones educativas a todos los niveles.

## REFERENCIAS BIBLIOGRÁFICAS

APROA. *Proyecto educacional de Chile: Aprendiendo con objetos de aprendizaje. Manual de buenas prácticas para el diseño de objetos de aprendizaje*. Chile: APROA. 2005. <https://silo.tips/download/manual-de-buenas-practicas-para-el-desarrollo-de-objetos-de-aprendizaje-version>.

Cabrera-Medina, Jaime Malqui, Irlesa Indira Sánchez-Medina y Fernando Rojas-Rojas. “Uso de objetos virtuales de aprendizaje, OVAS, como estrategia de enseñanza-aprendizaje inclusiva y complementaria a los cursos teóricos-prácticos. Una experiencia con estudiantes del curso de física de ondas”. *Revista Educación en Ingeniería* 11 (22) 2016: 4-12. <https://educacioneningenieria.org/index.php/edi/article/view/602>.



**Metadatos sociales: iniciativas...**

- Colombia. Ministerio de Educación Nacional. *Portal Colombia aprende. La red del conocimiento*. Bogotá: MEN. 2005. [http://www.colombiaaprende.edu.co/html/directivos/1598/article-368.html#h2\\_1](http://www.colombiaaprende.edu.co/html/directivos/1598/article-368.html#h2_1).
- Massa, Stella y Pesado Patricia. “Evaluación de la usabilidad de un objeto de aprendizaje por estudiantes”. *Revista TE&ET* 8 (2012): 65-76. <http://sedici.unlp.edu.ar/handle/10915/25537>.
- March, M. Amor Montané. “El trabajo terminológico sistemático: su aplicación para la elaboración de bancos de datos terminológicos”. En *Teoría y praxis en terminología*, coordinado por Sara Álvarez Catalá y Mario Bartié, 113-133. Uruguay: Universidad de la República. 2017.
- Morales Martín, Luz Yolanda, Lucía Gutiérrez Mendoza y Luz Mary Ariza Nieves. “Guía para el diseño de objetos virtuales de aprendizaje (OVA). Aplicación al proceso enseñanza-aprendizaje del área bajo la curva de cálculo integral”. *Revista Científica General José María Córdova* 14 (18) 2016, 127-147. DOI: <https://doi.org/10.21830/19006586.46>.
- Vega, A., Carlos A. y Juan A. Chica U. “Diseño y validación de un Objeto Virtual de Aprendizaje que permita el aprendizaje de heurísticas y metaheurísticas”. *Revista Avances en Sistemas e Informática* 7, 3 (2010): 103-108. <http://revistas.unal.edu.co/index.php/avances/article/view/26657>.

# Aprendizaje máquina en la Bibliotecología

GUADALUPE VANESSA CAROLINA GUTIÉRREZ HERNÁNDEZ

JORGE GÓMEZ BRISEÑO

*Universidad Nacional Autónoma de México, México*

## INTRODUCCIÓN

**E**l manejo de los datos es un tema fascinante, extremadamente útil para mirar con una perspectiva distinta el entorno, problemáticas y patrones interesantes, que sin duda resultan de interés para las humanidades, las cuales requieren comprender los fenómenos, problemas y situaciones para las disciplinas que la integran.<sup>1</sup> Tal como lo menciona Meneses, a diferencia de otras áreas de conocimiento, las Ciencias Sociales no visualizan los datos de manera neutral, sino como una construcción en la que ocurren diversas mediaciones, donde se pretende encontrar patrones a partir de los datos, elaborar modelos y proyectar fenómenos sociales mediante técnicas analíticas específicas, entre las que destacan el aprendizaje automatizado y la minería de datos, de las cuales se obtienen correlaciones e incluso proyecciones.<sup>2</sup>

---

1 María Elena Meneses Rocha, “Grandes datos, grandes desafíos para las Ciencias Sociales”, 434.

2 *ibidem*.

### *Metadatos sociales: iniciativas...*

A continuación, se enlistan algunos ejemplos de proyectos sociales que se apoyan en el procesamiento de los datos:<sup>3</sup>

- UN Global Pulse. Iniciativa de las Naciones Unidas orientada al uso de datos, para obtener una mejor comprensión de los cambios en el bienestar humano y retroalimentación en tiempo real acerca del funcionamiento de las respuestas políticas.<sup>4</sup>
- UN Women. Dentro de sus planes de acción se encuentra el desarrollo de investigaciones e informes basados en datos sobre las brechas existentes, además, apoya la producción y el uso de estadísticas de género de alta calidad en la formulación de políticas. El trabajo de este grupo ha incluido el desarrollo de un conjunto mínimo de 52 indicadores de género, así como nuevas métricas para medir la violencia contra las mujeres y las niñas.<sup>5</sup>

Impulsado por el desarrollo de las TIC, en los últimos años, ha habido un avance en la ciencia de datos, gracias a la reducción de costos de almacenamiento de los datos, el desarrollo de algoritmos de aprendizaje máquina, así como la consolidación de los lenguajes de programación o software para el desarrollo de modelos, lo que permite dar soluciones a problemas y fortalecer la toma de decisiones.

## DESARROLLO

La ciencia de datos es el método para obtener valor de diversas fuentes y grandes volúmenes de datos. Se encuentra conformado por un campo interdisciplinario que involucra matemáticas,

---

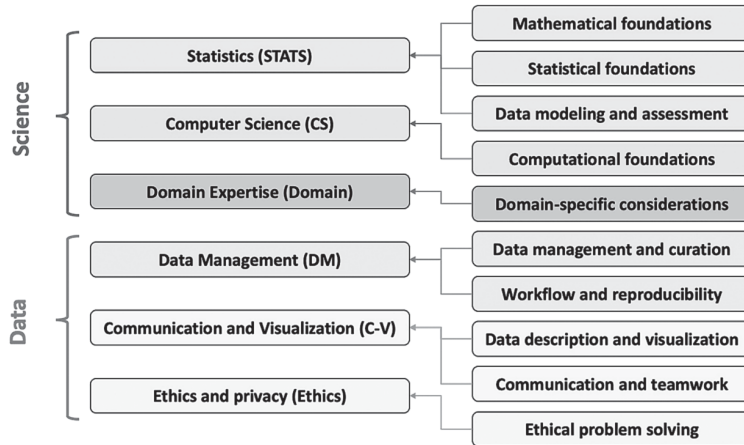
3 Massimo Lapucci y Ciro Cattuto, eds., *Data Science for Social Good*.

4 UN Global Pulse - Big Data for development and humanitarian action: <https://www.unglobalpulse.org>

5 UN Women <https://www.unwomen.org/en>

estadística, aprendizaje máquina, así como conocimiento del ámbito de estudio, entre otras habilidades.<sup>6</sup> Se integra por:<sup>7</sup>

Figura 1. Elementos a considerar en la ciencia de datos



Fuente: Gang Shao *et al.*, *op. cit.*, 23.

- 1) Fundamentos matemáticos. Es necesaria la construcción de modelos, demostrar hipótesis y validar si estas últimas representan los problemas del mundo real.
- 2) Fundamentos estadísticos. Ciencia de recopilar, analizar, presentar e interpretar los datos.<sup>8</sup>
- 3) Modelado de los datos y evaluación. En este apartado se debe de identificar el modelo apropiado para el tipo de problema o pregunta de investigación.

6 Dhanurjay Patil, "A Memo to the American People from U.S. Chief Data Scientist Dr. DJ Patil: whitehouse.gov".

7 Gang Shao *et al.*, "Exploring potential roles of academic libraries in undergraduate data science education curriculum development", 23.

8 Facultad de Economía, "Estadística...".

*Metadatos sociales: iniciativas...*

- 4) Fundamentos de computación. Habilidades para acceder y organizar datos en bases de datos, recuperarlos de los sitios web, procesamiento del lenguaje natural, administrar el almacenamiento de los datos y protegerlos desde el punto de vista informático.
- 5) Dominio del tema. Los involucrados en la aplicación de la ciencia de datos deben de tener amplio conocimiento del tema para formular las preguntas de investigación adecuadas e interpretar los resultados.
- 6) Administración y curaduría de datos. Parte fundamental de la ciencia de datos es la recuperación, preparación y acceso a los datos.
- 7) Flujo de datos y reproducibilidad. En la ciencia de datos se utilizan los flujos de datos, que muestran el camino seguido en los procesos, para resolver problemas. Documentar tales flujos permite entender el uso y análisis de los datos.
- 8) Descripción de los datos y visualización. Elemento importante para comunicar acerca de los descubrimientos del proceso de ciencia de datos. Como resultado del procesamiento, la visualización es importante para transmitir los resultados del proceso de aprendizaje máquina. La visualización es el proceso de crear imágenes que permitan entender tendencias, variaciones o descubrimiento significativo por parte de los datos.
- 9) Comunicación y trabajo de equipo. Los equipos de ciencia de datos deben formular preguntas de investigación y comunicar los resultados de manera entendible para la comunidad tanto experta como no experta. Esto implica un trabajo interdisciplinario que permita proponer soluciones creativas a diversos tipos de problemas.<sup>9</sup> Otros apartados dentro de la comunicación son:

---

9 Engineering and Medicine National Academies of Sciences, *Data Science for Undergraduates: Opportunities...*

- Habilidad para entender las necesidades de los clientes
  - Proporcionar reportes claros
- 10) Resolución de problemas éticos. El uso potencial de los datos incrementa las posibilidades de intrusiones en la privacidad de los usuarios, por ello es importante entender el papel de la ética a través del ciclo de vida de la ciencia de datos. Este apartado se integra por:
- Principios éticos y código de conducta
  - Privacidad y confidencialidad

La Bibliotecología es una disciplina relacionada con la recopilación, administración y tratamiento de registros de información.<sup>10</sup> La aplicación de ciencia de datos en esta rama de conocimiento tiene antecedentes importantes, entre el que destaca la propuesta del término *bibliomining*, que se define como: la minería de datos aplicada a bases de datos en bibliotecas.<sup>11</sup>

## APRENDIZAJE MÁQUINA

En 1950, en el artículo denominado “Computing Machinery and Intelligence”, Alan Turing se preguntaba si las computadoras podían pensar.<sup>12</sup> A partir de ese momento, el término de aprendizaje máquina ha ido evolucionando de acuerdo con el tiempo. En 1959 Arthur Samuel definió el aprendizaje máquina como el campo de estudio que proporciona a las computadoras la habilidad de aprender sin ser explícitamente programadas.<sup>13</sup> Por su parte,

---

10 Tingting Zhu y Lili Zhang, “Application of data mining in the analysis of needs of university library users”.

11 Scott Nicholson, “The basis for bibliomining: Frameworks...”.

12 Alan M. Turing, “Computing Machinery and Intelligence”.

13 Arthur L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers”.

Tom Mitchell<sup>14</sup> resaltó tres elementos: la tarea, la experiencia y el desempeño, es decir, cuando una computadora ejecuta un conjunto de tareas, la experiencia debe de estar encaminada al incremento del desempeño.

Una definición más reciente indica que el aprendizaje máquina es una rama de la inteligencia artificial, y que, utilizando procesamiento computacional, es posible diseñar sistemas que puedan aprender de los datos de forma que puedan ser entrenados.

La inteligencia artificial es la teoría, métodos y tecnologías para simular inteligencia biológica en la resolución de problemas. Existe cierta interacción entre la inteligencia artificial y la ciencia de datos, pero una no es subconjunto de otra. La inteligencia artificial incluye todo lo que permite a las computadoras aprender a resolver problemas y realizar decisiones inteligentes, y la ciencia de datos se puede apoyar de dichos algoritmos.<sup>15</sup> El aprendizaje máquina se enfoca en la detección de patrones y teoría computacional fundamentada en la inteligencia artificial.<sup>16</sup>

Dichos sistemas pueden aprender y mejorar con la experiencia y con el tiempo, y se puede refinar un modelo que puede ser usado para predecir salidas, basadas en el aprendizaje previo.<sup>17</sup> De manera complementaria, se refiere a la construcción e implementación de un algoritmo que permita aprender de los datos y a partir de ello predecir los atributos.<sup>18</sup>

---

14 Tom Michael Mitchell, *Machine Learning*.

15 IBM, "What is data science?".

16 Zongben Xu *et al.*, "Data science: connotation, methods, technologies, and development".

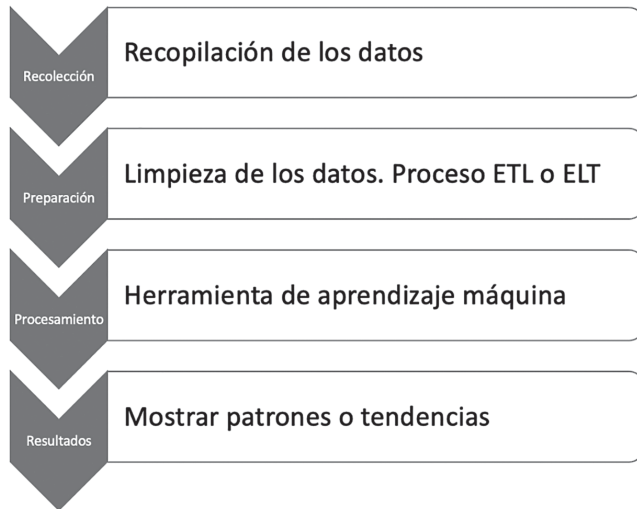
17 Jason Bell, *Machine Learning*.

18 Xu *et al.*, "Data science: connotation, methods, technologies, and development".

## PROCESO DE APRENDIZAJE MÁQUINA

Si bien el término de aprendizaje máquina se refiere a la aplicación de algoritmos, es importante describir el proceso, para que dicha aplicación sea exitosa:

*Figura 2. El proceso de aprendizaje máquina*



Fuente: Jason Bell, *op. cit.*

- a) **Recolección.** Cuando se quiere iniciar con el uso de aprendizaje máquina, es posible preguntarse acerca de las fuentes de datos a utilizar. Actualmente existen repositorios de datos abiertos que pueden ser útiles para comenzar a experimentar.
- **Kaggle.** Es un repositorio donde es posible encontrar código y conjunto de datos aplicables a la ciencia de datos. Se integra de más de 50,000 conjuntos de datos públicos. Entre los conjuntos de datos que es posible utilizar, destacan los de la Biblioteca de San Francisco, o la MTS Library, que



contiene información de la interacción entre sus usuarios y el uso de los libros, o el del inventario de la colección de la Biblioteca Pública de Seattle. En general, Kaggle incluye 110 conjuntos de datos relacionados con bibliotecas.<sup>19</sup>

- UCI Machine Learning Repository. El repositorio de aprendizaje automático de la University of California Irvine (UCI) es una colección de bases de datos, teorías de dominio y generadores de datos que utiliza la comunidad de aprendizaje automático para el análisis de los algoritmos de aprendizaje automático.<sup>20</sup>
- Datos abiertos. El portal de datos abiertos del Gobierno de la República integra el conjunto de datos de diversas instituciones gubernamentales. Los conjuntos de datos relacionados sobre bibliotecas son aproximadamente 41, entre los que es posible encontrar asistentes a bibliotecas públicas, ubicación, directorios de bibliotecas, catálogos de colecciones digitales, entre otros.<sup>21</sup>

Si bien los repositorios son una fuente confiable de conjunto de datos, es importante establecer las categorías de datos aplicables a las bibliotecas. La fuente de datos de la Bibliotecología se integra de una gran variedad de bases de datos, repositorios institucionales, datawarehouse, los cuales representan una gran oportunidad para la extracción de conocimiento.<sup>22</sup> Los datos pueden ser estructurados o no estructurados y pueden componerse de las búsquedas realizadas por los usuarios en OPAC (Online Public Access Catalog), las estadísticas de uso de los recursos digitales o los movimientos de circulación de sus materiales bibliográficos, por mencionar algunos.

---

19 “Kaggle: Your Machine Learning and Data Science Community”.

20 “UCI Machine Learning Repository: <https://archive-beta.ics.uci.edu>.”

21 “Datos Abiertos de México”, [datos.gob.mx](https://datos.gob.mx).

22 Alexandre Ribas *et al.*, “Data science in data librarianship: Core competencies of a data librarian”.

- b) **Preparación.** Una vez identificado el conjunto de datos, el siguiente paso es prepararlo para que pueda ser usado en el aprendizaje máquina. Esta parte del proceso es fundamental, y la que en ocasiones lleva más tiempo dado que aquí se establece la calidad de los datos. Sin un conjunto de datos tratado de manera adecuada, la aplicación de herramientas de aprendizaje máquina no funcionará de manera óptima. Existen diferentes visiones para preparar los datos, entre las que destacan:
- ETL (Extract, Transform, Load - Extracción, Transformación, Carga). Es un proceso de integración de datos que combina datos de múltiples fuentes en un solo medio (normalmente un datawarehouse) u otro sistema de destino.
  - ELT (Extract, Load, Transform). La principal diferencia con ETL es el orden. ELT copia o exporta los datos de las ubicaciones de origen, pero en lugar de cargarlos en un área de preparación para la transformación, carga los datos sin procesar directamente en el datawarehouse para transformarlos según sea necesario.

Los tipos de archivo más utilizados en la extracción de datos son los siguientes:<sup>23</sup>

- Texto en formato raw.
- Archivo separado por comas.
- JSON (Java Script Object Notation)
- XML (eXtensible Markup Language)
- Hojas de cálculo
- Bases de datos
- Imágenes

Un lenguaje de programación es una serie de instrucciones contenidas dentro de un archivo denominado código fuente y, mediante

---

<sup>23</sup> Jason Bell, *Machine Learning*.

entorno tecnológico, dichas instrucciones se traducen en programas de computadora. Estos son empleados para diseñar e implementar programas encargados de definir y administrar el comportamiento de los dispositivos físicos y lógicos de un equipo de cómputo.<sup>24</sup>

Existe una gran variedad de lenguajes de programación y con diferentes propósitos. Los que actualmente están enfocados a la ciencia de datos son:

- **Python.** Liderado por la Fundación de Software de Python (PSF Python Software Foundation), es un lenguaje de alto nivel, interpretado, modular, orientado a objetos. Su filosofía de lenguaje abierto permite el desarrollo de paquetes que aumentan su potencialidad. Su versión actual es la tres y es uno de los lenguajes de programación más utilizados en la ciencia de datos.
- **R.** Es un lenguaje de programación de alto nivel, usado para el cómputo estadístico, de aprendizaje máquina, y desarrollo de gráficos. Forma parte del proyecto de software libre GNU (“GNU no es UNIX”) y de la Fundación de Software Libre (Free Software Foundation FSF). Actualmente se encuentra la versión 4.2. Su nombre proviene de las iniciales de los dos primeros autores, quienes fueron Robert Gentleman y Ross Ihaka.<sup>25</sup> Usualmente R es más utilizado en la academia en áreas de la Estadística, Matemáticas y minería de datos. Entre sus características destaca la gran variedad de paquetes que permiten dirigir análisis complejos de datos y la integración óptima con otros lenguajes de programación tales como C++, Java, C., .net y Python.
- c) **Procesamiento.** Con los datos listos, es posible aplicar aprendizaje máquina. Existen diferentes tipos de algoritmos que pueden ser usados en el aprendizaje máquina, aunque

---

24 Francisco J. Ceballos, “Enciclopedia del lenguaje C++”.

25 Kurt Hornik y R Core Team, “R. Preguntas Frecuentes”.

su selección dependerá mucho del tipo de pregunta a contestar. Se clasifican en:

- Aprendizaje supervisado. Se refiere a trabajar con un conjunto de datos que se encuentran etiquetados. Por cada ejemplo en los datos de entrenamiento, existen campos de entrada y un campo de salida.<sup>26</sup> Un ejemplo aplicable a biblioteca sería tratar de dar respuesta a la pregunta ¿cuáles son las carreras que tienden a no regresar material bibliográfico? ¿Existe una correlación entre la carrera y el comportamiento de devolución de material bibliográfico?
  - Aprendizaje no supervisado. En este caso el algoritmo encuentra patrones ocultos de un conjunto de datos. Normalmente no existe una respuesta correcta o incorrecta, el objetivo es ejecutar el algoritmo y visualizar los posibles patrones de salida.<sup>27</sup> En este caso se podría aplicar para agrupar diferentes tipos de usuarios, por carrera, por tipo de material, y descubrir tendencias en el uso de las colecciones.
- d) **Mostrar patrones o tendencias.** Como resultado del procesamiento, la visualización es importante para transmitir los resultados del proceso de aprendizaje máquina. La visualización es el proceso de crear imágenes que permitan entender tendencias, variaciones o descubrimiento significativo por parte de los datos.<sup>28</sup>

Existen diferentes formas de visualizar los datos. Dependiendo del tipo de problema, es posible seleccionar el mejor medio para representar la visualización. La siguiente imagen muestra el comportamiento de variables involucradas en el préstamo de material bibliográfico.

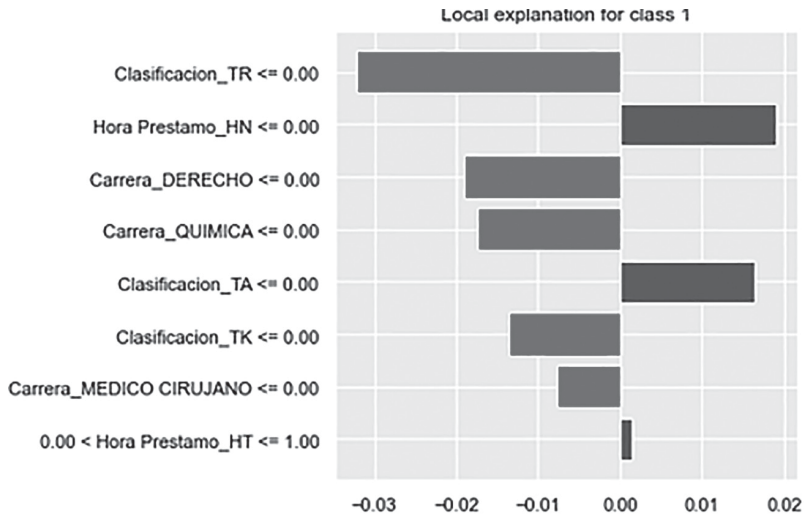
---

26 Jason Bell, *Machine Learning*.

27 Jason Bell, *idem*.

28 David Stuart, *Practical Data Science for Information Professionals*.

Figura 3. Ejemplo de representación de variables



## EL ROL DEL BIBLIOTECÓLOGO EN EL APRENDIZAJE MÁQUINA

Si bien es relevante el conocimiento en disciplinas tales como Matemáticas o de computación, es igual de importante el dominio del objeto de estudio y la habilidad de contar una historia. Diversos autores coinciden en que plantearse preguntas respecto a los datos o la curiosidad para explorarlos, es indicio de un buen comienzo para obtener respuestas mediante la aplicación de un procedimiento de ciencia de datos.

El profesional de la información aplica ciencia de datos en bibliotecas mediante la implementación de esquemas para la organización de estos. De hecho, existen roles asociados al científico de datos, en los que se integra a los profesionales de la información:<sup>29</sup>

---

29 Alma Swan y Sheridan Brown, “The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs report to the JISC”.

- Creador de datos (*data creator*). Investigadores con un nivel alto del dominio del tema, productores de datos.
- Científico de datos (*data scientist*). Profesionales que trabajan en el análisis de los datos producidos por la investigación.
- Administrador de datos (*data manager*). Ingenieros en cómputo o informáticos quienes toman la responsabilidad de proveer de facilidades tecnológicas.
- Bibliotecario de datos (*data librarian*). Bibliotecólogos, especializados en la curación, preservación y almacenamiento de los datos.

El rol de bibliotecario de datos ha evolucionado y deben ser referidos como científicos de datos cuando desarrollan de forma teórica y práctica habilidades para la administración y curaduría de datos.<sup>30</sup>

## CONCLUSIONES

La aplicación de la ciencia de datos en la Bibliotecología es un campo interdisciplinario de amplio desarrollo. El aprendizaje máquina es la aplicación de algoritmos computacionales, pero dicha aplicación forma parte de un proceso en el que cada etapa es importante para obtener descubrimientos interesantes.

Si bien la formación del bibliotecólogo está más orientada al área de las Ciencias Sociales, existen puntos que representan un área de oportunidad:

- El aprendizaje máquina requiere la interpretación humana, con el fin de valorar los resultados proporcionados por los algoritmos. El bibliotecario de datos, como experto del dominio, tiene la capacidad de interpretar dichos patrones, así como guiar el modelo de aprendizaje máquina.

---

30 Alexandre Semeler, Adilson Pinto y Helen Rozados, "Data science in data librarianship: Core competencies of a data librarian".

### ***Metadatos sociales: iniciativas...***

- La aplicación de aprendizaje máquina en equipo es mejor. Normalmente los equipos de ciencia de datos, orientados a bibliotecas, consideran a especialistas en Matemáticas, Informática y expertos en el dominio de conocimiento bibliotecológico.
- Aplicar aprendizaje máquina en las bibliotecas es totalmente viable, y la mejor forma de aprender es con datos públicos o con datos privados. Además, se debe fomentar el desarrollo de más repositorios de datos, con el fin de poder realizar ciencia en dichos conjuntos y con ello replicar ejemplos relacionados al ámbito bibliotecológico.

### **REFERENCIAS BIBLIOGRÁFICAS**

- Bell, Jason. *Machine Learning*. Indianapolis: John Wiley & Sons, Inc. 2014. <https://doi.org/10.1002/9781119183464>.
- Ceballos Sierra, Francisco Javier. *Enciclopedia del lenguaje C++*. Madrid: RA-MA. 2009, 899.
- Datos Abiertos de México - datos.gob.mx. <https://datos.gob.mx/> (Consultado el 19 de agosto de 2022).
- Dhar, Vasant. “Data science and prediction”. *Communications of the ACM* 56, núm. 12 (1 de diciembre de 2013): 64–73. <https://doi.org/10.1145/2500499>.
- Facultad de Economía. “Estadística. Algunos Conceptos”. <http://herzog.economia.unam.mx/profesores/blopez/eestadistica-descriptiva.pdf> (Consultado el 5 de junio de 2022).
- Hornik, Kurt y R Core Team. “R. Preguntas Frecuentes”. <https://cran.r-project.org/doc/FAQ/R-FAQ.html> (Consultado el 12 de abril de 2022).
- IBM. “What is data science?” (Consultado el 25 de febrero de 2024). <https://www.ibm.com/topics/data-science>.

- “Kaggle: Your Machine Learning and Data Science Community”. <https://www.kaggle.com/> (Consultado el 19 de agosto de 2022).
- Lapucci, Massimo y Ciro Cattuto, eds. *Data Science for Social Good*. Cham: Springer International Publishing. 2021. <https://doi.org/10.1007/978-3-030-78985-5>.
- Mitchell, Tom Michael. *Machine Learning*. McGraw-Hill series in computer science. McGraw-Hill. 1997.
- National Academies of Sciences, Engineering and Medicine. *Data Science for Undergraduates: Opportunities and Options*. *Data Science for Undergraduates*. National Academies Press. 2018. <https://doi.org/10.17226/25104>.
- Nicholson, Scott. “The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services”. *Information Processing and Management* 42, núm. 3 (1 de mayo de 2006): 785–804. <https://doi.org/10.1016/j.ipm.2005.05.008>.
- Patil, Dhanurjay. “A Memo to the American People from U.S. Chief Data Scientist Dr. DJ Patil: [whitehouse.gov](http://whitehouse.gov)”. 2015. <https://obamawhitehouse.archives.gov/blog/2015/02/19/memo-american-people-us-chief-data-scientist-dr-dj-patil>.
- Rocha, María Elena Meneses. “Grandes datos, grandes desafíos para las Ciencias Sociales”. *Revista Mexicana de Sociología* 80, núm. 2 (2018): 415–44. <https://doi.org/10.22201/iis.01882503p.2018.2.57723>.
- Samuel, Arthur L. “Some Studies in Machine Learning Using the Game of Checkers”. *IBM Journal of Research and Development* 3, núm. 3 (1959): 210–29. <https://doi.org/10.1147/rd.33.0210>.



**Metadatos sociales: iniciativas...**

- Semeler, Alexandre Ribas, Adilson Luiz Pinto, y Helen Beatriz Frota Rozados. "Data science in data librarianship: Core competencies of a data librarian". *Journal of Librarianship and Information Science* 51, núm. 3 (26 de septiembre de 2019): 771–80. <https://doi.org/10.1177/0961000617742465>.
- Shao, Gang, Jenny P. Quintana, Wei Zakharov, Senay Purzer, y Eunhye Kim. "Exploring potential roles of academic libraries in undergraduate data science education curriculum development". *Journal of Academic Librarianship* 47, núm. 2 (el 1 de marzo de 2021): 102320. <https://doi.org/10.1016/j.acalib.2021.102320>.
- Stuart, David. *Practical Data Science for Information Professionals*. Facet publishing. 2020.
- Swan, Alma, y Sheridan Brown. "The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs report to the JISC", 2008. [www.keyperspectives.co.uk](http://www.keyperspectives.co.uk).
- Turing, Alan M. "Computing Machinery and Intelligence". *Mind* 59, núm. 236 (1950): 433–60. <http://www.jstor.org/stable/2251299>.
- "UCI Machine Learning Repository". Consultado el 19 de agosto de 2022. <https://archive-beta.ics.uci.edu/>.
- "UN Global Pulse - Big Data for development and humanitarian action". Consultado el 17 de agosto de 2022. <https://www.unglobalpulse.org/>.
- "UN Women". Consultado el 17 de agosto de 2022. <https://www.unwomen.org/en>.
- Xu, Zongben, Niansheng Tang, Chen Xu, y Xueqi Cheng. "Data science: connotation, methods, technologies, and development". *Data Science and Management* 1, núm. 1 (marzo de 2021): 32–37. <https://doi.org/10.1016/j.dsm.2021.02.002>.

Zhu, Tingting, y Lili Zhang. "Application of data mining in the analysis of needs of university library users". *2011 6th International Conference on Computer Science & Education (ICCSE), Computer Science & Education (ICCSE)*, <https://doi.org/10.1109/ICCSE.2011.6028662>.

***Metadatos sociales: iniciativas, tecnologías, aplicaciones y softwares.*** Instituto de Investigaciones Bibliotecológicas y de la Información/UNAM. La edición consta de 100 ejemplares. Coordinación editorial, Sergio J. Sepúlveda H., revisión especializada: Marcos Emilio Bustos Flores; corrección de pruebas: Carlos Ceballos Sosa, Marcos Emilio Bustos Flores; formación editorial, Mario Ocampo Chávez. Fue impreso en papel cultural de 90 g en Editorial Albatros, Av. Benito Juárez M 26 L 14, Col. El Molino Tezonco, c.p. 09960, CdMx. Se terminó de imprimir en mayo de 2024.