

LA REVOLUCIÓN DE LOS DATOS BIBLIOGRÁFICOS, CIENTÍFICOS Y CULTURALES

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 licenciente 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 licenciente.



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 la [misma licencia](#) del original.

La revolución de los datos bibliográficos, científicos y culturales

COLECCIÓN

METADATOS

INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS Y DE LA INFORMACIÓN

**La revolución de los datos
bibliográficos, científicos y culturales**

Coordinador
Ariel Alejandro Rodríguez García



Universidad Nacional Autónoma de México
2020

Z666.7

R47

La revolución de los datos bibliográficos, científicos y culturales / Coordinador Ariel Alejandro Rodríguez García.
- México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2020.

xv, 346 p. – Colección: Metadatos

ISBN: 978-607-30-2996-4

1. Metadatos bibliográficos. 2. Datos vinculados. 3. Big data 4. Repositorios institucionales. 5. BIBFRAME (Modelo conceptual). I. Rodríguez García, Ariel Alejandro, coordinador. II. ser.

Diseño de cubierta: Oscar Daniel López Marín

Primera edición, 2020 D.R. © UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, Ciudad de México

Impreso y hecho en México

ISBN: 978-607-30-2996-4

Publicación dictaminada

Contenido

INTRODUCCIÓN	ix
1. DATOS BIBLIOGRÁFICOS	1
I. Retro y prospectiva de la investigación bibliotecológica sobre metadatos	3
ARIEL ALEJANDRO RODRÍGUEZ GARCÍA	
II. <i>BIBFRAME</i> : Un modelo de metadatos para la web semántica	17
FILIBERTO FELIPE MARTÍNEZ ARELLANO	
III. Metadatos, datos enlazados e interoperabilidad: reflexiones en el dominio bibliográfico	33
FABIANO FERREIRA DE CASTRO	
IV. De los metadatos para la organización de la Información a la tecnología <i>middleware</i> para los servicios de las bibliotecas: la biblioteca inteligente	63
GERARDO BELMONT LUNA	
V. Enlazando datos: un modelo conceptual orientado a objetos para el diseño de bases bibliográficas	79
LIZBETH BERENICE HERRERA DELGADO	
VI. Informe desde las trincheras: Transformando unidades de catalogación para incluir servicios de repositorio	113
LISA FURUBOTTEN	

2. DATOS CIENTÍFICOS141

I. Factores determinantes para la implementación
del esquema de metadatos para repositorios de datos
de investigación de la Política de Ciencia Abierta en México143

MIGUEL ADOLFO GUAJARDO MENDOZA

II. Habilidades requeridas por el profesional de la Industria 4.0.
Aparición de un nuevo perfil para el área de producción,
en la cuarta Revolución Industrial.161

MARCO BRANDÃO

III. Implementación de un repositorio unificado para
conformar el Sistema de Inteligencia Institucional del INE183

ALEJANDRO VERGARA TORRES

IV. Calidad en los metadatos:
reto para asegurar la participación en repositorios
nacionales manteniendo los estándares institucionales205

ALMA BEATRIZ RIVERA AGUILERA, ELISA CRUZ ROJAS,

BENJAMÍN ALEJANDRO GUERRERO OLIVERA Y ÓSCAR ALFONZO PEREYRA

V. Interoperabilidad en plataformas
de publicación y distribución de libros digitales.....223

JENNY TERESITA GUERRA GONZÁLEZ

3. DATOS CULTURALES	245
I. Metadatos para preservación digital	247
JUAN VOUTSSÁS MÁRQUEZ	
II. El Sistema de Información para el Registro Universitario de Espacios y Activos Culturales (SI-RUEyAC)	265
CATALINA NAUMIS PEÑA, NATALIA VELAZCO PLACENCIA Y ARIEL ALEJANDRO RODRÍGUEZ GARCÍA	
III. La ciudadanía digital y sus mediciones: el caso del estado de ánimo de los tuiteros en México	283
HÉCTOR ALEJANDRO RAMOS CHÁVEZ	
IV. El papel de los metadatos en la construcción de datos enlazados para bibliotecas	301
EDER ÁVILA BARRIENTOS	
V. La integración del usuario final en la creación de metadatos	325
PATRICIA HERNÁNDEZ SALAZAR	

Introducción

Por cualquier parte que miremos, el uso de los datos está convirtiéndose en una nueva manera de intercambiar información. Nos hallamos inmersos en un gran proyecto que de alguna manera desafía al pasado. La evolución natural de los datos ha generado nuevos tipos y modelos que permean cada aspecto de nuestra vida. De ahí que, al incrementarse el cúmulo de datos, se esté llegando a lo que diversas disciplinas han comenzado a señalar como la revolución de los datos.

En un principio, la idea era que al aumento en el volumen de información se le refiriera como datos masivos y requería de algún tipo de procesamiento para poder analizarlos. En estos momentos, Mayer-Schönberger y Cukier refieren en su libro sobre el big data que con los datos masivos son ese con los que se puede trabajar solo a gran escala para transformar paradigmas de mercado o gubernamentales, así como culturales.

Diversos textos que han abordado el asunto de la revolución de los datos masivos, regularmente hacen mención a la suma de gigabytes, terabytes y exabytes de información almacenada; el aumento en las capacidades de procesamiento de datos, y la sobrecarga de información que recibe el usuario.

Pero la realidad es que no hay manera fácil de concebir la cantidad tan exorbitante de datos que está siendo observado como un “diluvio digital” que está invadiendo el planeta.

Para algunos, puede parecer novedoso disertar sobre la revolución de los datos, pero esta discusión viene desde la década de 1960, cuando se hacía referencia, como lo señalan Mayer-Schönberger y Cukier, a la “revolución de la información” y “era digital”. En aquellos momentos incipientes, la masa digital no era gran cosa puesto que el porcentaje existente de datos digitales era muy bajo en comparación con los datos analógicos. En las siguientes décadas, los datos se multiplicarían drásticamente debido al aumento en el procesamiento de datos, las constantes actualizaciones tecnológicas y la concepción de los datos masivos como una fuente de innovación.

Como apuntan los autores ya referidos, la revolución de los datos es la forma que varias disciplinas han adoptado para referirse al conjunto de datos masivos que consiste en ver y comprender las relaciones en el seno y entre distintos fragmentos de información que, hasta hace muy poco, nos esforzábamos por captar plenamente.

En 1997, en la bibliotecología se indicó que el dato, refiriéndose al dato bibliográfico, es aquel que proporciona un significado para que sea posible su almacenamiento, recuperación, búsqueda y registro con fines de dar acceso al dato de investigación. De modo que no es nuevo hablar sobre el dato bibliográfico, pero, como lo indican Khan Hamman Rauf y Yuntei Du en *What is a data librarian*, traerlo al entorno digital se vuelve un asunto que requiere nuestra atención porque el concepto sirve de fundamento cuando se hace referencia al dato científico. Es decir, es concebido como un término facilitador en todos los estadios de la investigación científica.

A propósito de lo anterior, en las tradiciones bibliotecológicas el postulado que se tiene sobre el dato es que éste sirve para crear estructuras que define con claridad un punto de acceso o sirven para crear conjuntos de datos con los elementos de los diversos recursos de información, razón por la cual no es desconocido para los bibliotecólogos hablar sobre el dato desde sus características generales y el dato bibliográfico en lo particular.

En esta era digital, los datos masivos están abriendo espacios a las disciplinas interesadas para que se sumen a esta nueva revolución. La bibliotecología, de acuerdo con Zhang Ming y Gunila Widén, es la más reciente en unirse, asumiendo que sus responsabilidades tradicionales de organizar el conocimiento; recuperar y diseminar la información, así como el mantenimiento de los sistemas de información han sido rebasadas por los nuevos tipos y modelados de datos. De ahí que se esté buscando la reformulación de sus patrones con el fin dar continuidad a sus tradicionales e incorporar aquellas surgidas por el uso, acceso y reúso de los datos masivos.

En la misma línea de cambio, Gordon-Mumane puntualiza que son cuatro las razones principales por las que en la bibliotecología se está interesado por ingresar al estudio de los datos masivos. El fácil acceso al Internet por su famosa accesibilidad y disponibilidad de los recursos de información; la asequibilidad y aplicación de dispositivos digitales; el incremento en la cantidad de sus tipos de recursos digitales; su avanzada tecnología para la colecta, registro, análisis y agregación de datos y su manera de difundir la utilización de los datos.

En definitiva, los bibliotecólogos que ingresan al entorno digital bajo las premisas de los datos masivos serán los referente del antes y el ahora en las habilidades, títulos de trabajo y responsabilidades profesionales

asociadas con la creación y uso de los datos. Como lo indican Khan Hamman Rauf y Yuntei Du, se distinguirán del catalogador tradicional y el creador de metadatos y tal vez los nombren como los bibliotecólogos de datos, haciendo referencia a sus responsabilidades y habilidades orientadas al tratamiento, administración, control y uso de los datos.

Cada vez que hacemos clic en una búsqueda de información, realizamos una transacción móvil; generamos contenidos utilizando las redes sociales o cualquier otra acción que implique el uso de la computadora, estamos siendo partícipes de una u otra forma en el uso intensivo de datos masivos. Por tanto, como indican Mayer-Schönberger y Cukier, una vez que se ha datificado el mundo, los usos potenciales de la información no tiene más límites que el ingenio personal.

Un componente importante de los datos masivos es el metadato. Con éste, la mayoría de los datos se podrán manipular los programas y especificaciones de las aplicaciones para lograr una correcta organización, uso y acceso a la información. Abundando en esta idea, los nuevos tipos de metadatos están más presentes en cada aspecto de nuestra vida.

Por citar algunos ejemplos, los dispositivos móviles en los automóviles, los posicionadores GPS, las redes sensoriales, entre otros han desarrollado el movimiento de la “metainformación” clara y orientativa. Los textos, el video, los objetos físicos y la fotografía han explicado el uso del metadato embebido. En los recursos textuales es posible encontrar el uso de signos o símbolos como los *emotronics*, los cuales representa un metadato semiótico. Por su parte, los videos están integrando metadatos que logran hacer un reconocimiento facial.

En definitiva, los metadatos han provocado un crecimiento exponencial en cuanto a la forma de acercarse al conocimiento y el medio en que vivimos, así como conozcamos con más detalle el dato masivo. En suma, los viejos y nuevos tipos de metadatos dan sustento a la teoría de la “cultura del uso de los datos.”

El presente texto contiene tanto las reflexiones como las investigaciones recogidas del Seminario de Investigación Metadatos e investigadores y especialistas en bibliotecología y estudios de la información, así como aquello que en sus disciplinas se encuentran trabajando sobre el impacto de los datos masivos.

La obra que tiene en sus manos reúne textos que muestran la manera en que los datos masivos están estudiándose como datos bibliográficos, datos científicos y datos culturales. La intención de agruparlos de esta manera es buscar que como lector no pierda de vista que los datos masivos toman como fundamento a los datos bibliográficos para después darle consistencia al dato científico y cerrar con el dato cultura como una forma de mirar cómo el patrimonio cultural está siendo digitalizado con el fin de ser datificado. Es decir, traspasar la barrera de ver al objeto físico como una copia digital y observarlo a través de un reconocimiento óptico que pueda ser procesado por el humano y las computadoras.

La parte referente a los datos bibliográficos inicia con el estudio retrospectivo y prospectivo de la investigación bibliotecológica sobre los metadatos. Después, se analiza el modelo de metadatos BIBFRAME como una nueva propuesta para la descripción de recursos que está acorde a la perspectiva de los datos enlazados recuperables de modo semántico. Posteriormente, se reflexiona sobre la interoperabilidad en el dominio bibliográfico, identificada

como una ontología generadora de datos semánticos que sirve para construir una Web diferente a la actual. Continúa con la revisión de la tecnología *Middleware* que facilita el intercambio de datos entre las aplicaciones utilizadas en los servicios de las bibliotecas. Finaliza con dos estudios más, uno relacionado con el modelo conceptual orientado a objetos y el otro relativo a la transformación de una unidad de catalogación a otra más amplia enfocada a trabajar con datos y metadatos.

La parte relativa a los datos científicos comienza con la implementación de esquemas de metadatos para una política de ciencia abierta. Continúa con el estudio de un nuevo perfil profesional para la producción de datos en la cuarta revolución industrial. Enseguida, se aborda el tema de los ciudadanos y su relación con el gobierno y la forma en que se obliga al gobierno a hacer disponible la información. Posteriormente se analiza la manera en que las instituciones de educación superior participan en un programa nacional de ciencia abierta. Esta sección concluye con dos trabajos, el primero relacionado con el análisis de la interoperabilidad de las plataformas de libros electrónicos y las exigencias de los usuarios como factor determinante en la distribución de este tipo de publicaciones. Finaliza con una reflexión sobre el metaanálisis del uso de los datos en las ciencias médicas y cómo los médicos buscan, recuperan y usan la información para el desarrollo de esta ciencia.

La parte final de esta obra inicia con un trabajo relativo a la preservación digital y cómo los metadatos deben adecuarse para el presente y las futuras migraciones de los recursos y las tecnologías, y se hace mención de que los metadatos son un factor esencial para la transferencia de información y el resguardo adecuado.

Se continúa con un proyecto cultural que tiene como fin identificar, ubicar y dar a conocer los distintos espacios y activos culturales de la Universidad Nacional Autónoma de México. Posteriormente, se encuentra la investigación relacionada con la medición del estado de ánimo de quien se reconoce como ciudadano digital. Esta sección finaliza con dos estudios, uno relativo a los elementos constitutivos de los datos enlazados bibliográficos y su relación con los datos científicos y culturales y el otro, referente al usuario como el eje de todo esquema de metadatos.

Queremos agradecer a quienes de manera directa e indirecta permitieron que los trabajos expuesto en el Seminario de Investigación Metadatos sean publicados, especialmente a la doctora Georgina Araceli Torres Vargas, directora del Instituto de Investigaciones Bibliotecológicas y de la Información, al personal del Instituto que con su trabajo cotidiano hace posible la publicación de esta obra, a los miembros del Seminario de Investigación Metadatos y a mi familia, la presente y ausente, que son los motores que me impulsan y motivan para que se continúe entusiastamente en el descubrimiento de nuevos temas de la bibliotecología y estudios de la información.

Ariel Alejandro Rodríguez García

I. DATOS BIBLIOGRÁFICOS

Retro y prospectiva de la investigación bibliotecológica sobre metadatos

ARIEL ALEJANDRO RODRÍGUEZ GARCÍA
Universidad Nacional Autónoma de México

INTRODUCCIÓN

La investigación bibliotecológica por varias décadas se concentró en el estudio de la catalogación descriptiva; el desarrollo del catálogo en sus diferentes transiciones, esto es desde el catálogo en tarjetas, pasando al catálogo electrónico, luego al catálogo en línea, hasta su versión más reciente denominado como descubridor de información, y en el control de autoridades de nombres personales y temáticas con el firme propósito de que el usuario localice, seleccione y obtenga los distintos trabajos de un mismo autor o contenido temático.

A propósito, Roe (2000) señala que los fundamentos de la investigación en la catalogación y clasificación han trazado las direcciones actuales de desenvolvimiento en este campo, razón por la cual es de comprenderse el arribo de las tecnologías como un hecho para que la investigación sea más fácil que en cualquier época. Por supuesto que los principios en este tipo de investigación siguen siendo los mismos, de ahí que con el paso del tiempo se piense que la localización

de una colección podrá hallarse a través de un catálogo en línea o un motor de búsqueda Web, o bien en un conjunto de documentos electrónicos colocados en una biblioteca digital.

Si bien es cierto que la base de la investigación en catalogación ha sido el perfeccionamiento de la catalogación descriptiva, en estos momentos sería pertinente preguntarse ¿por qué profundizar en la investigación sobre metadatos? Varios de los estudiosos estarían de acuerdo con lo que refiere Wendler (2000, 36) al señalar que *metadato* es la palabra de moda que está en boca de toda nuestra sociedad, fuertemente influenciada por la información, y que como bibliotecólogos se debe estar emocionado. Después de todo, siempre hemos creado metadatos para controlar y proporcionar acceso a nuestras colecciones.

En la misma línea de ideas, Caplan (2013) refiere que el término *metadato* fue acuñado por Jack E. Myers en los años sesenta y, posteriormente, registrado como marca comercial en 1986 por la compañía Metadata, la cual es proveedora de programas y servicios relacionados con la medicina.

En esta ocasión, no ahondaremos ni nos detendremos en revisar y analizar el concepto de metadatos porque no es intención de este trabajo. Lo que nos interesa es entrar en materia sobre los diversos asuntos relacionados con la investigación bibliotecológica sobre el tema. Las variaciones del término, de acuerdo con Caplan (2013) están fundamentadas en el Consorcio W3 (World Wide Web) y el acuerdo ha sido definir a los metadatos como la máquina que interpreta información sobre la web.

En la bibliotecología, se venido discutiendo sobre la importancia de los metadatos desde sus aspectos teóricos, metodológicos y prácticos. Así como los diversos tipos y esquemas de metadatos. Lo que nos proponemos documentar en este trabajo es hacer una revisión retro y prospectiva sobre

los diversos alcances que se ha tenido, tiene y tendrá el estudio de los metadatos.

También se considera oportuno examinar los estudios que se han gestado en torno a los metadatos como el medio para describir recursos de información. Posteriormente destacaremos los arquetipos de aquellas investigaciones que han señalado que en la bibliotecología los metadatos transitan por dos rutas, una que sigue la perspectiva tradicional de la disciplina y la otra encaminada a la administración de datos.

Los resultados que se muestran en esta investigación son preliminares y responden a un estudio más amplio relacionado con la investigación relativa a la interoperabilidad global de los datos bibliográficos, científicos y culturales que están colocando las bibliotecas, museos y archivos bajo los postulados de la web semántica.

DISEÑO DEL ESTUDIO DE LOS METADATOS

Los datos que se revelan más adelante provienen de una búsqueda realizada en diversos descubridores de información como Librunam, Clase, Periodica e Infobila, así como distintas bases de datos especializadas como Web o Science, Scopus, Science Direct. Esta búsqueda se llevó a cabo del 16 de marzo al 16 de mayo de 2018 y comprende el período de 1960, año en que surge el concepto de metadatos, hasta diciembre de 2017, año en que la investigación bibliotecológica ha desarrollado varias nociones teóricas y metodológicas sobre los metadatos y las estructuras de metadatos.

El concepto de *metadato* en idioma español y *metadata* en inglés fue la manera en que se buscaron, seleccionaron y recuperaron los estudios, trabajos e investigación sobre metadatos.

Los resultados se sistematizaron en una hoja de procesador de datos (Excel). El primer resultado fue que se reunieron 109 916 registros. De los cuales 69 500 registros con el término en inglés *metadata* provienen de las bases de datos especializadas, mientras que de los descubridores de información se obtuvo la cantidad de 38 545 registros en inglés y 811 registros en español, respectivamente.

El segundo resultado es que los 109 916 registros se consiguió agrupar los registros por décadas, de los cual se logró ver los siguientes datos. En la década de los sesenta se reunieron 46 registros. En la década de los setenta, 203 registros. Para la década de los ochenta, fueron 734 registros. En la década de los noventa, la cantidad de registros fue de 4 075. Mientras que en la década del 2000, los registros localizados fueron de 41 514. Finalmente, para el periodo que comprende 2010 a 2017, la cantidad localizados fue de 61 668 registros.

Lo que prosiguió a la agrupación de los registros por década fue la selección y obtención de aquellos textos que en cada década caracterizaron un avance, desarrollo, noción o derrotero para incrementar los estudios sobre metadatos.

METADATO, ALGO MÁS QUE DATOS

Una de los objetivos que está conduciendo este trabajo es indagar la evolución y desarrollo teórico, técnico y metodológico respecto a los metadatos, ya que en las últimas tres décadas los especialistas sobre metadatos han venido desplazando a los catalogadores.

Autores como Intner y Johnson (2008) afirman que el desarrollo de los metadatos ha influido en la apreciación que se tiene respecto a los recursos de información, así como

las prácticas seguidas por generaciones en la creación de los registros para los sistemas de información. Es por eso que la creación de metadatos exige un mayor conocimiento especializado tanto en las técnicas, como en los métodos a utilizar para generar o capturar un dato estructurado, así como la comprensión sobre la evolución del recurso de información.

Otro objetivo que soporta la revisión retro y prospectiva de la investigación bibliotecológica sobre metadatos es no perder de vista el futuro inmediato y a mediano plazo sobre la actividad de los metadatos, ya que de acuerdo con Miller (2011), quien está de acuerdo con Caplan (2013): esta actividad se vuelve difícil en cada disciplina debido a dos razones sustantivas. La primera es que la práctica de la catalogación se ha soportado y sustentando en un elevado número de normativas y estándares que tiene que aprobar con la creación de los metadatos. En segundo lugar, la generación de estructuras de metadatos actualmente emite teorías al interior de cada comunidad.

A propósito de la creación de los metadatos, como lo manifiestan Miller (2011), Coyle (2005), Caplan (2013) y Méndez Rodríguez (2002), entre otros más, ésta se soporta en cuatro tipos de metadatos básicos, los cuales dan validez a otros cuatro que permiten desarrollar la codificación e intercambio de datos, como los estándares estructuras de datos, de contenido, de valor y estándares de formato, intercambio técnico o codificación de datos.

La investigación bibliotecológica sobre metadatos hasta los primeros tres lustros del siglo XXI ha dependido del conocimiento adquirido por experiencias prácticas y hasta un poco empíricas que le han impedido llegar a generalizaciones como lo establece el método científico. Esto es, al final del proceso de investigación deberá llegarse a leyes o teorías científicas.

Con relación a esto último, cuando Smiraglia (2002) analiza el avance teórico sobre la organización del conocimiento, afirma que la organización ha generado teorías que se han movido hacia un estado epistémico del pragmatismo (basado en observaciones sobre el diseño de los mecanismos de recuperación) hacia el empirismo (sustentado en los hallazgos alcanzados por la investigación empírica) para llegar a un estado epistémico del racionalismo (construcción de metodologías de recuperación sustentadas en principios razonados).

La propuesta de Smiraglia (2002) sobre los estados epistémicos de la catalogación concluye que después de dos siglos de llevar a cabo diversos trabajos formales respecto a la elaboración del catálogo y la clasificación, los bibliotecólogos se encuentran en la cúspide puesto que han logrado un pensamiento racionalista y un largo conocimiento respecto a códigos con reglas pragmáticas.

En la misma línea de análisis, Miska (2009) aborda el asunto del futuro de la catalogación con base en la revisión de los textos elaborados por Chan y Taylor, autores clásicos en los libros de texto para la enseñanza de la catalogación y clasificación angloamericana. Estos autores han influido en la evolución de la tradición catalográfica americana y por muchos años sus textos sirvieron para el desarrollo de las mejores prácticas de catalogación, esencialmente en la descripción de algunas combinaciones sobre el control de autoridades de autor. A partir de sus escritos examinados desde la tesis de Smiraglia, se comprende que Chan como Taylor han fomentado el desarrollo epistémico del racionalismo en la catalogación.

En definitiva, el análisis que realiza Miska (2009) a los textos clásicos sobre catalogación de Chan y Taylor nos refiere que son problemáticos porque ambos autores en sus

ediciones más recientes no hacen mención de temas relacionados al surgimiento de Internet y la Web, siendo que ambos autores conocen y están enterados de estos dos fenómenos. De ahí que Miska afirme que la nueva era digital no es simplemente hablar de la representación de un nuevo tipo de recurso que deberá integrarse a un sistema, sino que es una nueva postura que deberá asumirse respecto al valor y el uso de la información.

Referirse a la investigación bibliotecológica sobre metadatos es imperioso desde antes de la aparición del Internet y la web, pero también lo es continuarlo con lo que ha venido sucediendo en la denominada era o futuro digital.

LOS CAMINOS RECORRIDOS POR METADATOS EN LA INVESTIGACIÓN BIBLIOTECOLÓGICA

Tomaremos como punto de salida el *Proverbio chino* que dice “Cuanto más grande es el caos, más cerca está la solución”. El interés por conocer más sobre los metadatos en la bibliotecología ha ido en aumento y en cada década el aparente caos ha permitido una estabilidad en los estilos y tipos de investigación.

Por ejemplo, la década de los sesenta la investigación era mínima y comenzaría a mostrar las diversas interpretaciones que iban surgiendo en torno al concepto de metadato. Históricamente el término se acuñó en 1968 por Bagley y se refería a la estructuración de datos que contenían datos.

Al respecto, Baofu (2015, 5) refiere que el término “meta-data en las bibliotecas” tenía su fundamento en las tarjetas catalográficas porque los metadatos se usaban en varias formas con un significado de información archivada en el catálogo.

En los años setenta, las investigaciones distinguieron entre lo que se venía realizando desde el control bibliográfico y reciente introducción de la computadora como medio automatizado de los datos en el registro bibliográfico.

Los años ochenta marcaron varias tendencias relativas a ver los metadatos desde dos vertientes, una orientada al control bibliográfico y la otra hacia la administración de datos.

En los años noventa, el estilo y tipo de investigación se orientó al desarrollo de modelos conceptuales basados en principios tecnológicos más que bibliotecológicos. Como es el caso de la aparición de modelo FRBR que a decir es un modelo abierto a diversas interpretaciones y aplicaciones como lo señala Tillett (2005), que es una forma de generalizar las entidades que se encuentran en el universo bibliográfico.

La arquitectura de los metadatos por su parte transcribe los datos inherentes a una entidad. Al respecto, Caplan citado por Smiraglia (2002) indica que los metadatos son información estructurada acerca de recursos de información de cualquier medio o tipo de formato.

Por lo anterior, estamos de acuerdo con Intner, Lazinger y Weihs (2006), quienes afirman que todas las nuevas particularidades distintivas entre los metadatos y la catalogación tradicional se vislumbran en la información estructurada que se usa para encontrar, acceder, usar y administrar principalmente aquellos recursos de información que están en un ambiente digital.

Otra investigación que refiere la arquitectura de los metadatos es la presentada por Taylor y Joudrey (2009), quienes dicen que los metadatos pueden clasificarse en tres niveles. El primero con un formato simple donde el metadato se encuentra desestructurado. El segundo nivel el metadato está estructurado e incluye un conjunto de elementos formales para otorgarle representatividad y descripción a los

recursos. Finalmente, está el tercer nivel que es un metadato enriquecido que puede ser empleado por cualquier sistema.

En esta década, la investigación sobre los metadatos tomó un rol preponderante en la organización de la información debido a que surgió la biblioteca digital como el medio por el cual se lograría una efectiva organización de los recursos de información. De ahí que se espere que las redes de información dependan de los metadatos como la mejor forma de administrar y organizar la información digital.

Autores como Brunett, Kwong, Bor Ng y Park en 1999 señalarían que esta década se caracteriza por el surgimiento de dos enfoques en el trabajo y desarrollo de los metadatos; estos serían el enfoque del Control bibliográfico y el de la administración de datos.

El enfoque del control bibliográfico surgido originalmente en la bibliotecología señala, por tradición, que los metadatos atienen la organización y el procuramiento de acceso a las entidades cargadas de información, así como a la descripción bibliográfica, el análisis temático y la construcción de notaciones clasificatorias. También al almacenamiento de los registros contenidos en los catálogos de la biblioteca y como la principal auxiliar para el usuario en el acceso a los objetos de información de su interés.

Por otra parte, el enfoque de la administración de datos, por tradición, nace en la computación y se orienta a la administración y organización de los datos; almacenaje de la información, la recuperación de datos textuales; seguridad de los datos y a las funciones de integración y seguridad de los datos. Es importante mencionar que en esta década de los noventa se marca la diferencia entre lo que se entiende por metadato y estándar de metadato.

El andamiaje construido por las cuatro décadas finales del siglo XX sobre la investigación de los metadatos responde

a las distintas necesidades creadas por los recursos de información, los sistemas de información, la recuperación y navegación en la información y otros temas que comenzaría a despuntar en el nuevo siglo. Pero hizo que las investigaciones futuras se preguntaran quién debería crear los metadatos: ¿los humanos, las máquinas o una combinación de ambos?

En la primera década del siglo XXI, autores como Smiraglia, Caplan, Méndez Rodríguez e Intner, Lazinger y Weish formaron los primeros manuales relacionados con el entendimiento de los metadatos y los estándares de metadatos. En ideas generales, cada texto señala que es vital conocer los principios y fundamentos de los metadatos, así como las indicaciones, los lineamientos y las buenas prácticas respecto a cómo se crean, aplican y usan. En ese mismo orden de ideas, refieren que estos textos no son libros de recetas que deban aplicarse invariablemente en todos los casos, sino que el creador de metadatos debe reflexionar y analizar sobre la manera en que los datos toman su lugar en la estructura de metadatos y luego sobre su posicionamiento en el entorno digital.

En suma, las rutas trazadas en esta década del 2000 ampliaron el horizonte y los alcances de los metadatos en la bibliotecología, disciplina que ingresó al estudio de los metadatos dentro de lo que se dio a conocer como datos vinculados y la web semántica. Intner y Johnson (2008) refieren que los metadatos han permitido cambiar la percepción que se tenía respecto a los recursos digitales y los impresos, debido a que trabajar con los sistemas de etiquetado es una forma distinta de descripción.

Con todos estos avances, cambios y nuevos derroteros establecidos por el estudio, análisis y propuestas, Olipant (2017) plantea la necesidad urgente de educar y preparar profesional que conozcan cómo se colecta, analiza,

transforma y presenta el dato metadatos, sin la necesidad de utilizar un código y obtener conocimientos de grandes conjuntos de datos, así como navegar y estructurar los datos para la toma de decisiones.

LA INVESTIGACIÓN BIBLIOTECOLÓGICA EN EL ESTUDIO DE LOS METADATOS

Este trabajo se propuso revisar una serie de documentos que por más de cinco décadas han marcado el camino en el estudio de los metadatos. El análisis se encuentra aún en la depuración. El refinamiento *a priori* del cúmulo de referencias obtenidas y otras recuperadas para su estudio arrojan por lo menos cuatro perspectivas a seguir, considerando sus diversas dimensiones de avance.

El enfoque del control bibliográfico; el enfoque en la administración de datos; el enfoque sobre los datos vinculados para la web semántica, así como la visión actual sobre la dataficación,¹ son algunos de los hitos que deban tener en cuenta, si se está interesados en el estudio sobre los metadatos.

Otra forma para referirse sobre la retrospectiva y prospectiva del estudio de los metadatos es la que autores como Lei Zeng y Qin (2016) señalan que en tiempos del *preinternet* los metadatos tenían dos funciones: proporcionar ricas descripciones bibliográficas y sus relaciones entre los datos heterogéneos de los ítems, y facilitar las formas de compartir los datos bibliográficos entre el universo de bibliotecas.

1 La dataficación emplea el reconocimiento óptico de caracteres para que puedan ser procesados por los seres humanos y las computadoras. Esto es de acuerdo con Viktor Mayer-Schönberger y Kenneth Cukier (2013).

Los mismos autores refieren que el desarrollo de los metadatos en tiempo de Internet en sus inicios fue atender sus particularidades, tipos y usos a través del *crosswalk*. Después, las comunidades científicas comenzaron a observar las soluciones para organizar el vertiginoso incremento de datos científicos y crear los diversos estándares de metadatos.

Desde la década del 2010 a la fecha, surgieron los servicios de vocabularios abiertos vinculados (Linked Open Vocabulary), la nube de datos vinculados (Linked Data Cloud), entre otros metadatos. En este sentido, la tendencia de los datos abiertos está generando esquemas de metadatos con *namespaces* específicos.

En síntesis, estamos siendo testigos de un continuo desarrollo y expansión sobre el estudio de los metadatos que se está llevando a la práctica en todas las disciplinas. La bibliotecología y los estudios de la información no están exentos de estudiar al dato, el cual se perfilar a ser “el nuevo petróleo” en la era del Internet.

BIBLIOGRAFÍA

- Bagley, Philip R. 1968. *Extension of programming language*. Philadelphia: University City Center.
- Baofu, Peter. 2015. *The future of post-human meta-data: toward a new theory of structure and process in information science*. Brisbane, Australia: Primrose Hall Publishing Group.
- Brunett, Kathleen, Kwong Bor Ng, Soyeon Park. 1999. "A comparison of two traditions of metadata development". *Journal of the American Society for Information Science*, vol. 50, núm. 3: 1209-1217.
- Caplan, Priscila. 2013. *Metadata fundamentals for all libraries*. Chicago: American Library Association.
- Coyle, Karen. 2005. "Library data in the web world" En: *Library Technology Report* (febrero-marzo): 5-11.
- Intner, Sheila y Johnson. 2008. *Impact of digital resource. Fundamental of Technical Services Management*. Chicago: American Library Association.
- . 2008. *Fundamentals of technical service management*. Chicago: American Library Association.
- Inter, Sheila S., Susan S. Lazinger, y Jean Weihs. 2006. *Metadata and its impact on libraries*. Wesport, Conn.: Libraries Unlimited.
- Lie Zeng, Marcia y Jian Qin. 2016. *Metadata*, 2nd. Editon. Chicago: Neal-Schuman.
- Martínez Arellano, Filiberto Felipe y Lina Escalona Ríos (comp.) 2000. *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica*. UNAM, CUIB; Infoconsultores.
- Mayer-Schönberger, Víktor y Kenneth Cukier. 2013. *Big data. La revolución de los datos masivos*. Madrid: Turner.

- Méndez Rodríguez, Eva. 2002. *Metadatos y recuperación de información: Estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón: Ediciones Trea.
- Miller, Steven. 2011. *Introduction to metadata for digital collection: a how-to-do-it- manual*. Nueva York: Neal-Schuman.
- Miska, Francis. 2009. "Chan, Taylor, and the future of cataloging text". *Library Quarterly*, vol. 79, núm. 1: 131-143.
- Olipant, Tami. 2017. "A case for critical data studies in library and information studies". *Journal of Critical Library and Information Studies*, vol. 1, núm. 1 DOI: 10.24242/ICLIS.V1I1.22.
- Roe, Sandy. 2000. "Acceso temático en línea". En: Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica. Filiberto Felipe Martínez Arellano, Lina Escalona Ríos (comps.). México: UNAM, CUIB; Infoconsultores.
- Smiraglia, Richard. 2002. "The progress of theory of knowledge organization". *Library Trends*, vol 50, núm. 3: 330-349.
- Taylor, Arlene y Daniel N. Joudrey. 2009. *The organization of information*. Third edition. Westport, Connecticut.
- Tillett, Barbara B. 2005. "FRBR and cataloging for the future. *Functional requirements for bibliographic records (FRBR): hype or cure-all?*" Patrick Le Boeuf, ed. *Cataloging & Classification Quarterly*, 39, núm. 3/4.
- Wendler, Robin. 2000. "Diversificación de actividades: habilidades y funciones catalográficas en la era digital". En: *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica*. Filiberto Felipe Martínez Arellano, Lina Escalona Ríos (comps.). México: UNAM, CUIB; Infoconsultores.

BIBFRAME: Un modelo de metadatos para la web semántica

FILIBERTO FELIPE MARTÍNEZ ARELLANO
Universidad Nacional Autónoma de México

INTRODUCCIÓN

BIBFRAME es un proyecto iniciado en 2011 por la Biblioteca del Congreso de los Estados Unidos con la finalidad de evolucionar de un entorno de datos bibliográficos descriptivos basados en el formato MARC a otro en donde se puedan aprovechar los beneficios de la web semántica para el uso de este tipo de datos, más allá del contexto actual de las bibliotecas.

BIBFRAME constituye un modelo que permite el manejo de las relaciones entre las entidades y los recursos de información y no solamente su descripción, como acontece actualmente en los registros del “catálogo” u otras herramientas de recuperación de información. BIBFRAME puede ser conceptualizado como un modelo de entidad-relación basado en la aplicación de los Datos Ligados (*Linked Data*) y RDF (*Resource Description*), elementos esenciales de la web semántica.

BIBFRAME establece las bases para el futuro de la organización de la información y la catalogación en las bibliotecas y otras instituciones que prestan servicios de información, lo que facilita la recuperación y navegación de los recursos

informativos, así como el intercambio de datos bibliográficos en el entorno de la web semántica. Por la importancia de este proyecto para la organización de la información, el objetivo que este documento pretende lograr es ofrecer una visión general sobre los fundamentos y características de BIBFRAME, así como una perspectiva sobre el futuro de la organización de la información y la catalogación en las bibliotecas y otras instituciones relacionadas con la recuperación de información.

¿QUÉ ES BIBFRAME?

BIBFRAME es el acrónimo de los términos *Bibliographic Framework* (estructura bibliográfica), cuyo origen se remonta a 2011 como un proyecto iniciado en la Biblioteca del Congreso de la Estados Unidos, cuya finalidad es lograr la evolución de los datos bibliográficos a un modelo de datos ligados y que la información bibliográfica sea de mayor utilidad dentro y fuera de las bibliotecas.

Respecto a BIBFRAME, Frank (2014) ha señalado que éste constituye la base para el futuro de la descripción bibliográfica; se convertirá en el principal medio de intercambio de datos bibliográficos, y reemplazará al Formato MARC. El principal beneficio de BIBFRAME para quienes buscan conocimiento es su capacidad para mejorar la exploración de la información a través del uso de enlaces y tecnologías de la World Wide Web, lo que crea la experiencia virtual de una “navegación múltiple” entre una gran cantidad de recursos, en comparación con la navegación y revisión física de ellos de manera tradicional.

Mediante la integración de los datos bibliográficos en un entorno de datos vinculados dentro de la World Wide Web,

BIBFRAME mejorará la recuperación de información y promoverá su navegación. Adicionalmente, reducirá los costos asociados con la catalogación tradicional porque disminuirá el tiempo asociado con el mantenimiento de los datos de autoridad. BIBFRAME se basa en las relaciones entre los recursos y no solamente en la descripción de éstos en registros bibliográficos.

DESCRIPCIÓN DE LOS RECURSOS DE INFORMACIÓN

A lo largo de su historia, las bibliotecas han centrado gran parte de sus esfuerzos en el desarrollo de catálogos. El objetivo fundamental del catálogo, al igual que el de cualquier otro sistema de información, es ofrecer a los usuarios de la biblioteca los materiales o recursos para satisfacer sus necesidades informativas.

Lo anterior implica que el catálogo debe permitir la localización, identificación, selección y obtención de aquellos recursos de utilidad para sus usuarios, anteriormente dentro de la biblioteca y actualmente en cualquier otro lugar donde éstos se encuentren. Evidentemente, la primera necesidad de los usuarios de la biblioteca, o de cualquier otra institución que ofrezca servicios de información, es localizar materiales o recursos informativos de utilidad. Esto es logrado por medio del establecimiento de los puntos de acceso del catálogo; los más comunes son el autor, el título y los temas.

La siguiente necesidad de un usuario al efectuar búsquedas en el catálogo, o en cualquier otro sistema de información, es “identificar” las características o los atributos de los materiales o recursos recuperados, tales como su título y subtítulo, la edición o versión, el lugar y la fecha de publicación o producción, la editorial o entidad productora,

su extensión, así como su formato y otros detalles técnicos. Lo anterior constituye la catalogación descriptiva de los recursos. La necesidad del usuario de reconocer y distinguir las características o atributos de los materiales, o recursos de información, se encuentra directamente relacionada con la siguiente acción que se efectúa dentro del proceso de búsqueda, seleccionar los materiales o recursos para necesidades particulares.

Una vez que el usuario ha identificado y seleccionado a través de los datos de los registros del catálogo aquellos materiales o recursos que le son de utilidad, el siguiente reto es brindarle los elementos que le permitan obtener aquellos que ha seleccionado. Esto es logrado a través de la identificación del lugar físico donde estos se encuentran, por medio de su clasificación y, actualmente, en algunos casos, a través de su URL o dirección electrónica. Sin embargo, el entorno actual de los recursos de información ha cambiado substancialmente, y ha traído consigo nuevos retos para la organización de la información y el desarrollo de los catálogos.

Los impresos dejaron de ser el único formato para generar y difundir información. Actualmente, los usuarios desean localizar en el catálogo, o en cualquier otro sistema de información, todos aquellos materiales o recursos que requieren, sin importar los formatos en que éstos hayan sido producidos, así como todos aquellos recursos derivados o relacionados con ellos. Esto implica que el catálogo y otros sistemas de recuperación de información deben ser capaces de llevar a cabo una nueva función, la de navegar o relacionar los distintos tipos de recursos, sin importar la forma o formatos en los que estos se manifiesten, así como todos aquellos recursos relacionados de una u otra forma, mostrando o visualizando en forma integral todas las relaciones existentes entre los recursos de información, así como

entre sus creadores. Esto únicamente puede ser logrado a través de la aplicación de los modelos de entidad-relación.

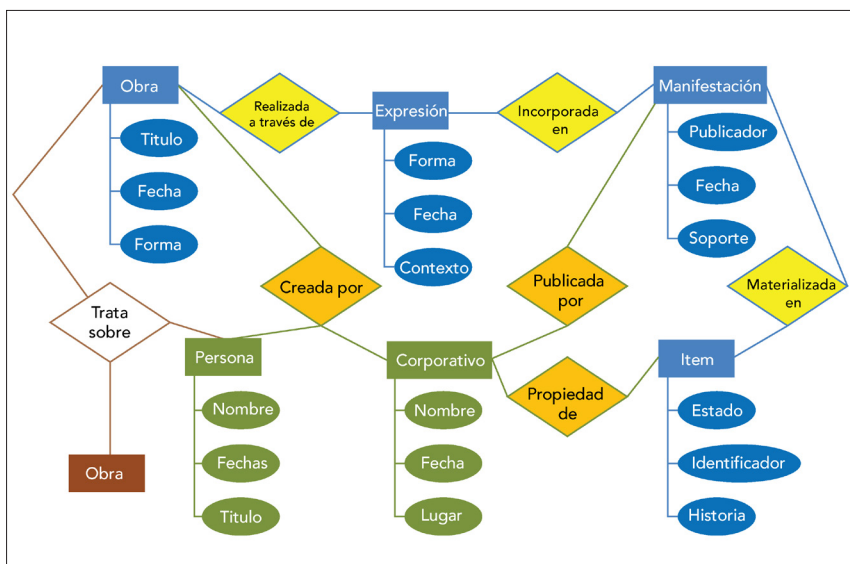
LOS MODELOS DE ENTIDAD-RELACIÓN

Peter Chen (1976) fue el creador del prototipo de los modelos de entidad-relación, los cuales son herramientas para el modelado de datos que permiten representar un grupo de entidades relevantes dentro de un sistema de información, así como sus atributos o características propias, pero además el establecimiento de las posibles relaciones entre las diferentes entidades que conforman dicho grupo.

Los modelos de entidad-relación se han aplicado en diversos ámbitos de la industria y los negocios y es hasta los noventa cuando este modelo se empezó a utilizar en el ámbito de las bibliotecas, al desarrollar el Grupo de Trabajo de la IFLA sobre los Requerimientos Funcionales para Registros Bibliográficos, bajo la dirección de la doctora Barbara Tillet, el modelo Functional Requirements for Bibliographic Description (FRBR), el cual ofrece una nueva perspectiva sobre la estructura y las relaciones en los registros bibliográficos, además de un vocabulario más preciso para ayudar a los catalogadores y diseñadores de sistemas a satisfacer las necesidades del usuario (Tillet 2004).

El modelo FRBR, el cual es ejemplificado en la Ilustración 1, se encuentra conformado por tres grandes componentes: entidades, atributos y relaciones. Las entidades pueden ser de tres tipos: productos intelectuales o materiales (obras, expresiones, manifestaciones e ítems); creadores o autores (personas y organizaciones), y materias o conceptos de los productos.

Ilustración 1



Fuente: Ruiz Figueroa, R. (2011).

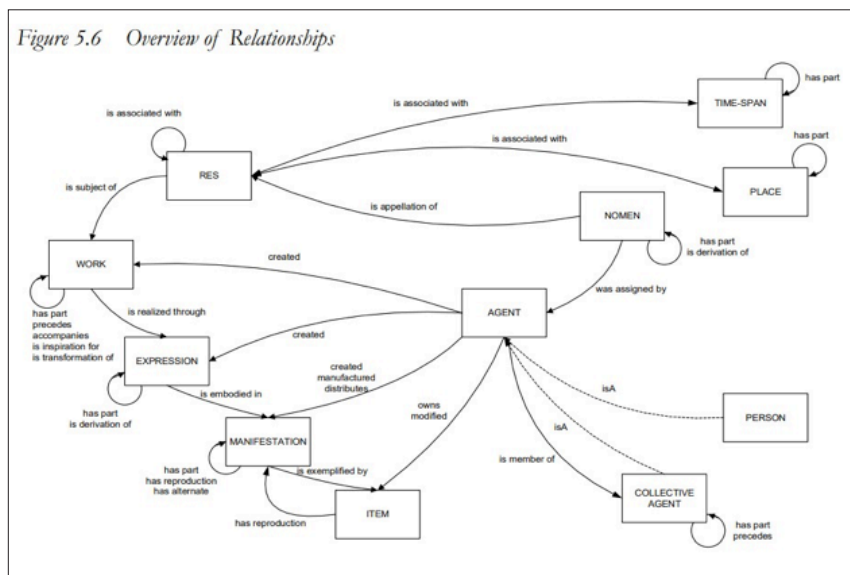
El segundo componente del modelo FRBR son los atributos o características de todas las entidades anteriormente mencionadas. El tercer componente del modelo FRBR, el cual puede ser considerado como el más importante dentro del nuevo enfoque de la organización de la información, son las relaciones, las cuales pueden ser, entre otras, las existentes entre los diferentes tipos de productos y sus creadores, entre los recursos mismos, o bien, entre las personas y organizaciones que intervienen en la creación de los diferentes productos. Posteriormente, este modelo fue complementado por dos modelos más, Functional Requirements for Authority Data (FRAD) y (Functional Requirements for Subject Authority Data (FRSAD).

FRAD fue desarrollado en 2009 con el objetivo de proporcionar un marco de referencia estructurado y claramente definido para relacionar los datos que son registrados en los registros de autoridad para los creadores con las necesidades de los usuarios de esos datos, así como para apoyar el uso potencial de los datos de autoridad en su intercambio internacional dentro del sector de las bibliotecas, así como en otros (IFLA 2013).

FRSAD fue desarrollado con el propósito principal de elaborar un esquema que proporcionara una comprensión clara y compartida del objetivo que persiguen los datos, registros y archivos de autoridad de materias en relación con la información temática y las expectativas que dichos datos deben alcanzar para responder a las necesidades del usuario (IFLA 2010).

Recientemente, la IFLA ha desarrollado el modelo LRM, el cual integra los tres modelos anteriores. LRM pretende ser un modelo de referencia conceptual de alto nivel, desarrollado dentro de un marco enriquecido del modelado de entidades-relación. El modelo comprende los datos bibliográficos, entendidos en un sentido general y amplio, adoptando el enfoque original de FRBR. Se utiliza una técnica de análisis de entidades que comienza aislando las entidades que son los objetos clave de interés para los usuarios de los registros bibliográficos. Se identifican las características o los atributos asociados con cada entidad, así como las relaciones entre las entidades que son más importantes para los usuarios al momento de formular búsquedas bibliográficas, interpretar respuestas a esas búsquedas y “navegar” en el universo de entidades descrito en los registros bibliográficos (IFLA 2017).

Ilustración 2. Relaciones en el modelo LRM



Fuente: IFLA. 2017. Library Reference Model: a conceptual model for bibliographic information.

Actualmente, las relaciones constituyen un elemento de primer orden en la organización y recuperación de la información puesto que éstas conectan diversos tipos de entidades y proporcionan un contexto para cada una ellas. En el modelo LRM, las relaciones son amplias, y como puede observarse en la ilustración 2, son establecidas en forma general y abstracta para permitir a sus implementadores la inclusión de aquellas que sean necesarias y relevantes para el entorno en donde se encuentran.

CARACTERÍSTICAS DE BIBFRAME

El formato MARC, creado en la mitad de los sesenta, ha sido empleado por las bibliotecas desde entonces hasta nuestro días para registrar y compartir datos bibliográficos; sin embargo, el desarrollo tecnológico ha cambiado la forma en que los datos pueden ser creados y compartidos. Esto ha eclipsado las capacidades, en el pasado revolucionarias, de MARC para compartir datos, lo que deja a las bibliotecas aisladas.

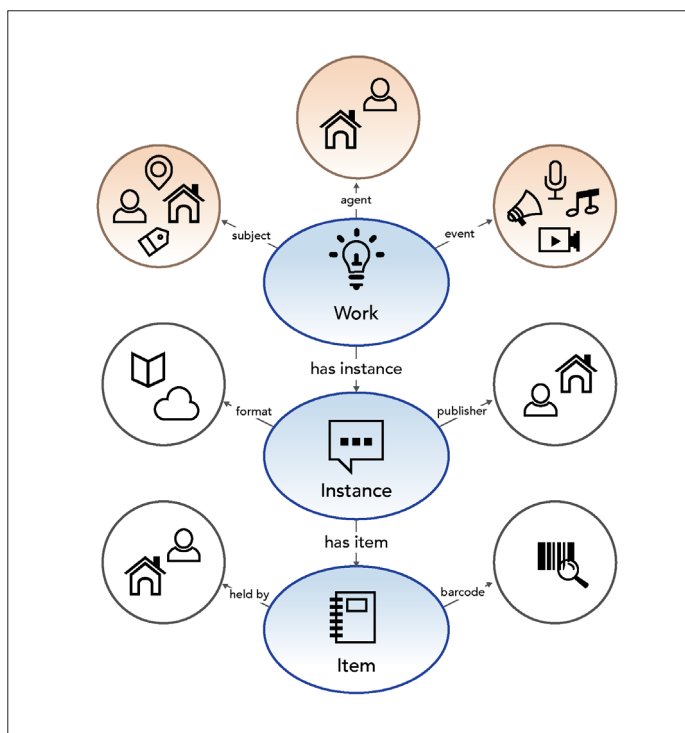
Los sistemas de recuperación de información como Google no pueden cosechar los datos codificados con MARC y hacerlos accesibles para otros. BIBFRAME presentará los datos bibliográficos de forma tal que en los sistemas de recuperación de información puedan ser manejados con un sentido semántico para presentarlos al usuario en forma ligada y de mejor manera, independientemente si dicho sistema es Google o el de una biblioteca como la del Congreso de los Estados Unidos (Frank 2014).

BIBFRAME no solamente es un remplazo del formato MARC, sino que va más allá. Establece las bases para el futuro de la descripción bibliográfica, considerando a ésta como parte de la web y el mundo de redes en que actualmente vivimos. Ha sido diseñado para integrarse y ser usado por amplias comunidades de información, aunque sin dejar de lado a las bibliotecas y otras instituciones similares. BIBFRAME pretende lograr la evolución de los estándares de descripción bibliográfica a un modelo de datos ligados con la finalidad de hacerlos de mayor utilidad dentro y fuera de la comunidad bibliotecaria.

El modelo BIBFRAME se encuentra conformado por tres clases principales: obras, instancias e ítems, tal y como puede ser notado en la ilustración 3. Las obras representan el

más alto nivel de abstracción en el contexto de BIBFRAME; contienen las propiedades conceptuales de los recursos catalogados: autores, idiomas y sobre lo que tratan (temas). Las instancias corresponden a las materializaciones de una obra en uno o más formatos; por ejemplo, determinado tipo de publicación. Una instancia posee información como su editor, lugar y fecha de su publicación, así como su formato. Un ítem es una copia (física o electrónica de una instancia y contiene información como su localización (física o virtual) a través de su clasificación y código de barras.

Ilustración 3. El modelo BIBFRAME



Fuente: Library of Congress (2016).

BIBFRAME también define conceptos relacionados con las clases principales que lo conforman, entre los que se encuentran los agentes y los temas. Los agentes son las personas u organizaciones asociadas con una obra, instancia o ítem, a través de diversos roles como autor, editor, artista, fotógrafo, compositor, ilustrador. Los temas son una o más materias o conceptos sobre las cuales trata una obra, pueden incluir tópicos, lugares, expresiones temporales, eventos, obras, instancias, ítems y agentes (Library of Congress 2016b).

Evidentemente, si analizamos el modelo BIBFRAME con los modelos FRBR y LRM, se puede identificar toda una serie de similitudes en su conceptualización, aunque también algunas diferencias, particularmente en la denominación de los elementos que los integran.

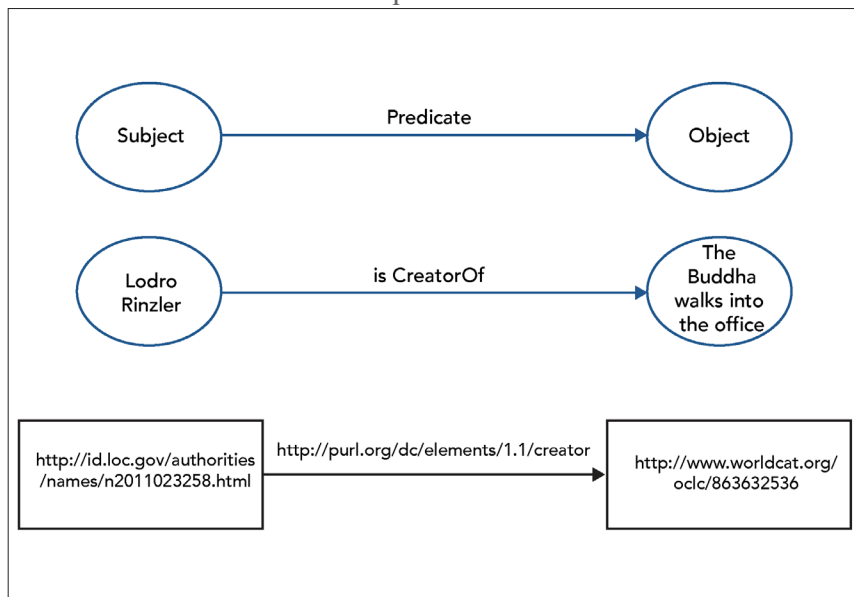
Por otro lado, si comparamos a BIBFRAME con el formato MARC, es notable que este último se enfoca a los registros del catálogo, cuyos datos son comprensibles pero de manera unívoca e independiente. Asimismo, MARC incluye información sobre aspectos conceptuales de las obras y sus atributos físicos, pero utilizando cadenas lingüísticas para identificar a los nombres personales, los nombres corporativos y los temas, entre otros elementos, los cuales son importantes y valiosos fuera del registro.

BIBFRAME, en lugar de combinar todo en un registro y potencialmente duplicar información en múltiples registros, depende en gran medida de las relaciones entre los recursos (relaciones obra-obra, relaciones obra-instancia, relaciones obra-agente), pero mediante el uso de identificadores únicos para diversos elementos (personas, lugares, idiomas, etcétera). MARC ya empleaba algunas de estas ideas (códigos geográficos, códigos de idioma), pero BIBFRAME busca que estos aspectos sean la norma en lugar de la excepción. El Modelo BIBFRAME es el punto de entrada de la comunidad

bibliotecaria para formar parte de una red de datos mucho más grande, en donde los vínculos entre diversos elementos son primordiales (Library of Congress 2016c).

La representación de las relaciones en BIBFRAME utiliza un modelo de datos vinculados (*linked data*) y, por lo tanto, la práctica del modelado RDF para identificar de manera única como recursos web a todas las entidades, los atributos y las relaciones (propiedades) entre las entidades. BIBFRAME utiliza el vocabulario o la sintaxis RDF para expresar las propiedades, clases y relaciones existentes entre todas ellas (Library of Congress 2016). Un ejemplo del uso de los aspectos mencionados puede ser observado en la ilustración 4, en donde a través de una tripleta RDF se

Ilustración 4. Tripleta RDF in BIBFRAME



Fuente: Billey (2015).

expresa la relación autor-título. Adicionalmente, es importante mencionar que las tripletas BIBFRAME pueden ser manejadas con estándares para el manejo de los datos en la Web como RDF-XML y JSON, entre otros (Billey 2015).

Finalmente, no puede dejarse de señalar que BIBFRAME y sus componentes se encuentran en desarrollo, está siendo experimentando por algunas instituciones y desarrolladores de sistemas; existen algunas herramientas para estudiar y probar su aplicación (Library of Congress 2016a). No existe aún una aplicación generalizada de este modelo de datos, aunque tarde o temprano esto tendrá lugar, por lo que como organizadores de la información es importante el conocimiento de sus características y la comprensión de la forma en que este modelo puede ser aplicado, tanto en las bibliotecas como en otras organizaciones y sistemas que almacenan y recuperan información.

CONSIDERACIONES FINALES

Tomando en consideración todo lo expuesto en este documento, BIBFRAME puede ser definido como un modelo para organizar y relacionar las entidades que existen dentro del contexto de la información bibliográfica, aprovechando las características y potencialidades de la web semántica con la finalidad de lograr una mayor recuperación de la información, pero sobre todo, de lograr una navegación o interrelación de los recursos existentes dentro y fuera de la biblioteca.

BIBFRAME tiene el potencial para lograr la evolución de la organización de la información y dejar atrás la conceptualización del manejo aislado de los datos bibliográficos en los registros del catálogo y compartir estos registro a un nuevo

entorno en donde los diferentes datos incluidos en los registros del catálogo cobran independencia e importancia por sí mismos, pudiendo ser compartidos a nivel global. En el entorno actual, las bibliotecas, así como otras instituciones que brindan servicios de información, necesitan intercambiar más datos en lugar de más registros.

Lo mencionado únicamente puede ser logrado a través de la comprensión de nuevas alternativas para el uso de datos en la web semántica como los modelos de entidad-relación, el modelo de datos vinculados, el modelado de datos en RDF la identificación única de recursos a través de URLs (*Uniform Resource Identifies*), lo cual es necesario para establecer una adecuada comunicación y trabajo interdisciplinario entre los bibliotecólogos y los profesionales de cómputo, quienes serán los responsables de la aplicación de estos nuevos modelos y técnicas de la web semántica.

BIBLIOGRAFÍA

- Chen, Peter. 1976. "The entity-relationship model--toward a unified view of data", *ACM Transactions on Database Systems*, vol.1, núm. 1: 9-36. Disponible el 17 de agosto de 2018 en <http://bit.csc.lsu.edu/~chen/pdf/erd-5-pages.pdf>.
- Billey, Amber. 2015. *BIBFRAME basics*. Disponible el 17 de agosto de 2018 en <https://netsl.files.wordpress.com/2015/12/bibframe-basics-nela20151.pdf>.
- Frank, Paul. 2014. *BIBFRAME: Why? What? Who?*. Disponible el 17 de Agosto de 2018 en [http://www.loc.gov/aba/pcc/bibframe/BIBFRAME paper 20140501.docx](http://www.loc.gov/aba/pcc/bibframe/BIBFRAME%20paper%2020140501.docx).
- IFLA. 2010. *Functional Requirements for Subject Authority Data (FRSAD): a conceptual model*. Disponible el 17 de agosto de 2018 en <https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>.
- _____. 2013. *Functional Requirements for Authority Data: a conceptual model*. Disponible el 17 de Agosto de 2018 en https://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.
- _____. 2017. *Library Reference Model: a conceptual model for bibliographic information*. Disponible el 17 de agosto de 2018 en https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.
- Library of Congress. 2016. *BIBFRAME 2.0 Vocabulary*. Disponible el 17 de Agosto de 2018 en <https://www.loc.gov/bibframe/docs/index.html>.

- _____. 2016a. *BIBFRAME Implementation, Tools, and Downloads*. Disponible el 17 de Agosto de 2018 en <https://www.loc.gov/bibframe/implementation/index.html>.
- _____. 2016b. *Overview of the BIBFRAME 2.0 model*. Disponible el 17 de Agosto de 2018 en <https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
- _____. 2016c. *What are the general differences between MARC and BIBFRAME?*. Disponible el 17 de agosto de 2018 en <http://www.loc.gov/bibframe/faqs/#q02>.
- Ruiz Figueroa, R. 2011. *El nuevo modelo FRBR y sus relaciones con FRAD*. http://web.uaemex.mx/REBICS/docs/publica_memorias/050_pres_ch_RDA_RRF.pdf
- Tillett, Barbara. 2004. *¿Qué es FRBR?: un modelo conceptual del universo bibliográfico*. Disponible el 17 de agosto de 2018 en <http://www.loc.gov/catdir/cpsd/Que-es-FRBR.pdf>.

Metadatos, datos enlazados e interoperabilidad: reflexiones en el dominio bibliográfico

FABIANO FERREIRA DE CASTRO
Universidade Federal de São Carlos, Brasil

Introducción

La sociedad contemporánea marcada por el uso intensivo de las Tecnologías de la Información y la Comunicación (TIC) ha generado una variedad de tipos de metadatos y patrones de metadatos, elementos esenciales para el modelado de entornos informativo-digitales que garantiza la interoperabilidad entre sistemas.

La evolución de las bibliotecas ha destacado a lo largo del siglo XX en lo que se refiere al desarrollo y uso de tecnologías, principalmente las TIC, pues potenciaron sus servicios con el ofrecimiento de nuevos recursos de acceso, con la utilización de formatos de intercambio de datos en la Catalogación Descriptiva, en el uso de ontologías, en la orientación para el modelado de catálogos, y en el proceso de búsqueda y recuperación de la información.

Los metadatos tienen como fundamento teórico y epistemológico los principios establecidos por la Catalogación Descriptiva, área de la bibliotecología y de la ciencia de la información, responsable de la descripción de recursos informacionales y representación de los datos y de la información disponible y almacenada en diversas estructuras.

La propuesta de creación de una web semántica, en la que los datos se procesan no sólo por su sintaxis, sino también por su semántica, ha impulsado el desarrollo de un conjunto de tecnologías para la representación de datos en la web, así como para la consulta a estos datos y el razonamiento computarizado a partir de ellos.

De esta manera, el contexto tecnológico vigente caracterizado por la iniciativa *linked data* y web semántica puede beneficiarse de los datos bibliográficos originarios de ambientes altamente estandarizados y estructurados, como las bibliotecas, lo que favorece la colaboración y la interoperabilidad en dominios del conocimiento heterogéneos.

Así, es posible identificar que existe una tendencia a la disponibilidad de recursos y contenidos informacionales en formato digital y la no utilización de estándares adecuados para representar y describir estos recursos. Se considera que el tratamiento dado al recurso informacional requiere una descripción de forma y contenido legible por máquinas con resultados comprensibles a los humanos y que cumple los requisitos de interoperabilidad entre los entornos informacionales.

El escenario contemporáneo de la catalogación descriptiva, marcado con el modelado conceptual de catálogos con los Requisitos Funcionales para Registros Bibliográficos (FRBR), y la propuesta del nuevo código de catalogación Resource Description and Access (RDA), y caracterizados por nuevas perspectivas tecnológicas, conocidas hoy por la web semántica, Web 2.0¹ y Web 3.0², requiere el estudio

1 La Web 2.0 se destaca por el ambiente colaborativo y de interacción para la construcción y el compartir el conocimiento. La sinergia creada a través de esa colaboración e interacción de los individuos acelera el proceso de socialización del conocimiento, en espacios o ambientes más interactivos y participativos. Los autores afirman que la Web 2.0 se constituye como “un nuevo espacio para acceder, organizar, gestionar, tratar y diseminar la información, conocimientos y saberes”(Blattmann y Silva 2007).

de las herramientas y métodos para el Tratamiento Descriptivo de la Información (TDI); en especial, para la representación y la descripción (forma y contenido) de recursos bibliográficos en el contexto digital.

En el desarrollo de un marco teórico y metodológico sobre las herramientas tecnológicas disponibles para la construcción de formas de representación de recursos informacionales en el ámbito digital, el presente trabajo se caracteriza por ser una investigación de análisis exploratoria y descriptiva del tema, en el intento de identificar el fundamento conceptual subyacente a las herramientas para la estandarización de la descripción de los aspectos de forma y del contenido de los recursos bibliográficos, en el ámbito digital, en el abordaje de la Ciencia de la Información.

Como resultados y reflexiones, se destaca en qué medida los entornos informacionales digitales necesitan ser repensados a partir del modelado conceptual, en la capa intangible definida en la representación y en la descripción de los datos bibliográficos para la promoción de la interoperabilidad.

DATOS ENLAZADOS Y EL DOMINIO BIBLIOGRÁFICO

En 2006, Berners-Lee partió de la comprensión de que “la web semántica no es sólo sobre colocar datos en la web. Está sobre establecer relaciones [*links*] para que una persona o máquina pueda explorar la web de los datos” (Berners-Lee 2006, s.p.)³. Se introdujo el concepto de datos enlazados al

2 Web 3.0 se constituye como una denominación para un periodo de evolución de la web marcado por la creación de ambientes informacionales altamente especializados y que sólo funcionarán efectivamente a partir de la implantación de la estructura de la web semántica (Santos y Alves 2009).

3 La traducción es del autor.

proponer cuatro reglas, o mejores prácticas, para que los datos publicados en la web sean considerados datos vinculados:

1. Utilice URI como nombres [identificadores] para las cosas.
2. Utilice URIs HTTP, para que las personas pueden acceder a estos nombres.
3. Cuando alguien accede a un URI, proporciona información útil utilizando los estándares (RDF, SPARQL).
4. Incluya enlaces a otros URI. Así podrán descubrir más cosas.

La propuesta de datos abiertos interconectados ofrece gran potencial al conectar recursos informativos a través de enlaces semánticos, que son significativos también para programas. Por el contrario, los enlaces convencionales no son más que una etiqueta textual significativa para los usuarios humanos, que los medios para que los programas navegadores, a partir de un recurso, accedan a otro sin explicitar cuál es el significado de la conexión entre los recursos. Al ser significativos para los programas, los enlaces semánticos pueden ser procesados de forma más rica por ellos, explorando y enriqueciendo cognitivamente el significado (legible por máquina) del vínculo entre ambos recursos.

Los datos enlazados proponen la conexión de datos representados en un formato estándar a partir del uso de enlaces semánticos y de la utilización de herramientas de prácticas propuestas por el movimiento. “Este proceso facilita la búsqueda de agentes humanos y no humanos y los dirigen en diferentes bases a partir de esos datos vinculados” (Arakaki 2016, 27).

Desde su introducción en 2006, el concepto de datos enlazados ha evolucionado y hoy ya se encuentra en la literatura un concepto derivado de él: Datos Abiertos Vinculados (DAV). “[...] datos enlazados que está disponible bajo una licencia abierta que no impida su libre reutilización” y presenta un sistema de clasificación para los datos disponibles en la web que en cuanto mayor sea la cantidad de estrellas más vinculadas y abiertas son los datos:

- ★ Disponible en la Web (en cualquier formato), pero como una licencia abierta [open license], para ser Open Data.
- ★★ Disponible como dato estructurado y legible por máquina (por ejemplo, una hoja de Excel en lugar de una imagen escaneada de una tabla).
- ★★★ Al igual que en la anterior, pero en un formato no propietario (por ejemplo, en formato CSV en lugar del formato de Excel).
- ★★★★ Todos los anteriores utilizando estándares abiertos del W3C (RDF y SPARQL) para identificar cosas, así las personas podrán referirse a sus cosas.
- ★★★★★ Todos los anteriores y: relacionar sus datos a los datos de otras personas para proveer contexto (Berners-Lee 2010, s.p.).⁴

En el caso de los datos adjuntos, se debe tener en cuenta que el concepto de datos enlazados se refiere a la interoperabilidad técnica de los datos, mientras que el de Datos Abiertos Vinculados tiene su foco en la interoperabilidad legal de estos datos, no siendo obligatorio que todos los datos enlazados sean abiertos (W3c Library Linked Data Incubator

⁴ La traducción es del autor.

Group 2011). Se destaca también que datos enlazados y web semántica no compiten entre sí, al contrario, “[...] los datos enlazados pueden ser considerados los bloques que constituyen la web semántica”.

En el escenario actual de la comunidad de la biblioteología y de la ciencia de la información, hay una preocupación y un reconocimiento creciente de que habrá la necesidad de un sucesor del formato bibliográfico MARC 21 (Machine Readable Cataloging), debido a las nuevas transformaciones en el dominio bibliográfico, permeada por el uso intensivo de las TIC (Tecnologías de la información y la comunicación). Según Coyle (2011), estas discusiones tienden a centrarse principalmente en las cuestiones estructurales, ¿el nuevo formato o modelo de datos será XML (eXtensible Markup Language), hará uso de RDF (Resource Description Framework) y patrones de datos vinculados?

Lo que estos temas no contemplan es la tarea mucho más compleja de traducir la semántica de los datos bibliográficos para la construcción de un nuevo estándar. De acuerdo con Thomale (2010), basta con una pequeña investigación de los datos codificados en el formato MARC 21 para revelar que las etiquetas y los subcampos son insuficientes para definir los elementos de los datos reales realizados por los registros de un catálogo bibliográfico. “El primer paso en la transformación de MARC 21 a otro formato es identificar cuáles son los elementos contenidos en un registro bibliográfico MARC 21, lo que no parece tan simple” (Thomale 2010, 3).⁵

MODELOS DE DATOS Y TECNOLOGÍAS DISPONIBLES

Con la heterogeneidad de datos y metadatos disponibles en el ambiente web, se hace necesario el desarrollo de un

5 La traducción es del autor.

esquema o una estructura que conlleve esta diversidad de metadatos, de modo que permita sistemas interoperables en la red. Este esquema o estructura se denomina arquitectura de metadatos, cuyos objetivos son representar y proporcionar una base de codificación de metadatos a través de estructuras flexibles.

En este escenario, la arquitectura de metadatos Resource Description Framework (RDF) utilizada de manera interdisciplinaria en la ciencia de la computación y en la ciencia de la información, posibilita el procesamiento y la gestión automatizada de recursos informativos, de acuerdo con áreas del conocimiento distintas, principalmente con relación: el descubrimiento de recursos informativos, la descripción de las relaciones entre los recursos informativos representados en la red, la ayuda a los agentes de software, el intercambio y el intercambio de información entre otros tipos de aplicaciones (Lassila y Swick 1999).

El RDF, como destacado por Ferreira y Santos (2013, 21),

[...] oferece a possibilidade para as comunidades de descrição de recursos definirem a semântica de seus metadados de maneira formal, isto é, definindo o significado dos elementos de metadados, conforme as suas necessidades específicas de descrição, em um modelo processável por máquinas.

El RDF es un marco para metadatos que permite la interoperabilidad. Se proporciona una infraestructura que permite la codificación, el intercambio, el uso y la reutilización de metadatos de una forma no ambigua, de manera que las máquinas puedan comprender la semántica de los metadatos y, por lo tanto, utilizarlos en el proceso llamado descubrimiento del recurso (Taylor 2004; Zeng y Qin 2008), lo que promueve

un mecanismo para la integración de múltiples esquemas de metadatos para la descripción de recursos en la web.

RDF es un grupo de especificaciones desarrolladas por el World Wide Web Consortium (W3C) como un modelo para la descripción de metadatos en la red. Es un lenguaje para la representación de información sobre recursos web.

El concepto de un recurso es generalizado en RDF para significar cualquier cosa que pueda describirse con metadatos. Esto permite a los metadatos que se aplican a cualquier cosa su identificación, incluso si no se puede recuperar directamente en la web.

Hillmann y sus colegas (2010) apuntan que RDF es, por lo tanto, compatible con el actual escenario del mundo real que RDA está destinado a abordar, donde los metadatos son en gran parte legibles por máquinas y los recursos descritos por sujetos humanos. Los mismos autores (2010) aclaran que RDF es una opción adecuada para la aplicación en vocabularios RDA por muchas razones:

- El modelo RDF se basa en las más simples estructuras de metadatos, una única declaración sobre una sola propiedad de un solo recurso. Tales declaraciones pueden ser agregadas de forma flexible para formar descripciones de alto nivel, o registros de metadatos, de un recurso específico.
- Al construir vocabulario para las entidades RDA en RDF, se cumple el objetivo de proporcionar instrucciones RDA para el registro de metadatos que pueden ser aplicadas independientemente de la estructura o sintaxis para el almacenamiento y la presentación de los datos.
- RDF asume un mundo abierto donde el almacenamiento y mantenimiento de metadatos se distri-

buyen; es decir, sus contenidos se destinan al compartir y no se organizan en silos cerrados. Esta apertura es esencial para que RDA rompa con los límites de las prácticas de las bibliotecas convencionales o tradicionales.

En relación al uso de ontologías y de los lenguajes de ontologías, Hillmann y colegas (2010) explican que cuando se expresan los elementos de RDA de una manera compatible con la Web Ontology Language (OWL) y el lenguaje RDF Schema (RDFS), posibilita soportar las reglas de inferencias tan importantes en el mundo RDF, y permite RDA beneficiarse de la extensibilidad del modelo RDF.

De la misma manera, los valores de vocabulario en el Simple Knowledge System (SKOS)⁶ y un vocabulario RDF construido en OWL permiten que los vocabularios se extiendan y se utilizan de acuerdo con las intenciones anheladas en RDA. El uso de SKOS también permite a la RDA estar mejor integrada al desarrollo general de ontologías y organización del conocimiento que son importantes para mejorar las aplicaciones de recuperación de la información centradas en los usuarios.

RDF requiere el uso de identificadores fácilmente procesables por las máquinas para entidades estructurales y de contenido. Estos identificadores son independientes de las consideraciones del lenguaje humano y permiten que los

⁶ El Sistema de Organización Simple del Conocimiento (SKOS) es una familia de lenguajes formales creados para la representación de tesauros, esquemas de clasificación, taxonomías, sistemas de encabezados de tema, o cualquier otro tipo estructurado de vocabulario controlado. SKOS se construye sobre RDF y RDFS, y su objetivo principal es permitir la fácil publicación de vocabularios controlados estructurados para la web semántica. Actualmente, SKOS es desarrollado por el grupo de trabajo y especificaciones del W3C. Más información en <http://www.w3.org/TR/skos-reference/skos.html#semantic-relations>.

vocabularios contruidos en RDA se traduzcan a diferentes lenguajes, sin necesidad de identificadores distintos. Ésta es una ventaja significativa para incentivar el uso de RDA además de la comunidad anglófona.

Los metadatos expresados en RDF pueden ser más fácilmente procesados y tratados para garantizar la semántica y la veracidad de sus contenidos, sin detenerse o aferrarse en la validación de su formato o sintaxis, y es significativamente diferente de “todo el mundo debe usar el mismo esquema o modelo de XML” (Hillmann *et al.* 2010, 2).⁷

Con base en la citación anterior, se puede decir que esto es importante en entornos informáticos digitales, donde la generación de metadatos heterogéneos por sujetos humanos no preparados e instruidos y también por no humanos (máquinas) es una constante.

De acuerdo con el editor de RDA, Tom Delsey (2010), la implementación del RDA se puede aplicar a las estructuras de varias bases de datos, estructurando sus metadatos según el vocabulario (reglas de contenido) determinado en RDA.

Para ello, las siguientes secciones se basarán en el pensamiento de Hillmann y sus colegas (2010), por sus relevantes contribuciones en el campo de la catalogación descriptiva, sobre todo en las áreas relacionadas al desarrollo, el uso y la aplicación de los metadatos en entornos digitales, además de formar parte de los miembros de la Dublin Core Metadata Initiative (DCMI).

Según Hillmann y colegas (2010), se describen tres escenarios en este contexto:

1. Una base de datos relacional y orientada a objetos.

⁷ La traducción es del autor.

2. Un banco de datos bibliográficos vinculados y registros de autoridades.
3. Y una base de datos de archivo simple.

RDA es optimizado para el primer y segundo escenarios, donde la estructura, es decir, el modelado de la base de datos se implementa según los modelos conceptuales del FRBR y de los Requisitos Funcionales para Datos de Autor (FRAD). Esto resulta en metadatos para un único recurso que se distribuyen a través de registros que contienen datos para la descripción y el acceso, con la duplicación de datos y un mejor potencial para el (re)uso. El grupo de trabajo asumió que este enfoque se desarrollaría, por ejemplo, por el tratamiento de valores de vocabulario RDA como “archivos de autoridad”. Esta descomposición *top-down* de registros de catálogos monolíticos es compatible con una agregación de instrucciones de RDF (Hillmann *et al.* 2010).

Para Hillmann y colegas (2010), RDA necesita ser compatible con las prácticas actuales a fin de incitar la adhesión, y la más prevalente implantación de base de datos en las bibliotecas de hoy son aquellas categorizadas por registros bibliográficos y de autoridad que son mejores ejemplificados por el estándar o formato de metadatos MARC 21. RDA busca ser compatible con los tres escenarios, pero esto dificulta la extensión del escenario de base de datos relacional en RDF puro para representaciones RDA, lo que implica una serie de desafíos por el grupo de trabajo RDA.

Un aspecto significativo del RDA es el uso del modelo entidad-relación datos en FRBR. El uso del modelo entidad-relación (ER) es un componente clave de la web semántica, pero es completamente nuevo en las reglas de catalogación en bibliotecas. Las revisiones del RDA realizadas por el JSC,

a lo largo de su desarrollo, hicieron que el comité repensara las bases conceptuales tradicionales establecidas por las reglas de catalogación. Es importante destacar que el desarrollo de los elementos de RDA refleja el compromiso del JSC junto a los principios de los FRBR, aunque la interacción de RDA con los FRBR no ha sido desarrollada de tal forma que traduciría directamente la correspondencia de las relaciones en RDF (Hillmann *et al.* 2010). En este sentido, el grupo DCMI / RDA ha trabajado en el desarrollo y registro de vocabularios para subsanar las principales lagunas que puedan existir en la correspondencia de los elementos RDA en las representaciones en RDF.

La integración entre RDA y FRBR significaba que RDA en RDF sería necesario para interactuar con un RDF compatible con el tratamiento de FRBR, preferentemente con un acuerdo oficial entre sus creadores. Sin embargo, la IFLA no pudo proporcionar a las entidades FRBR con los requisitos e identificadores necesarios estructurados en RDF a su debido tiempo. Además, el trabajo de la IFLA sobre el FRAD recién llegó a su fin. Por lo tanto, para realizar la creación de propiedades RDA en RDF fue necesario crear una versión transaccional de FRBR, preferentemente en el mismo registro de RDA. Esta versión específica RDA de FRBR incluye “Familia”, de los FRAD, y “Agente” del modelo orientado a objetos de FRBR, e incluye sólo el nivel de la entidad, con las entidades identificadas como clases y subclases. Una versión RDF de FRBR está oficialmente disponible a partir de la IFLA, las relaciones entre las mismas clases definidas en la versión RDA de FRBR y una versión IFLA de los FRBR se hará para indicar que éstas son realmente las mismas entidades (Hillmann *et al.* 2010).

Hacer la conexión entre las entidades FRBR y las propiedades RDA no fue simple (Delsey 2010). El primer paso fue definir elementos RDA como propiedades RDF, intentando

asignar cada propiedad a una y sólo una entidad específica de FRBR. Había preocupaciones significativas con este enfoque porque los elementos de RDA eran limitados y no correspondían como los FRBR podrían ser implementados en la práctica.

Las comunidades especializadas con una visión diferente y de acuerdo a las necesidades de sus usuarios no tendrían otra alternativa que crear nuevos elementos de datos (con diferentes relaciones para entidades FRBR) para expresar sus metadatos.

De acuerdo con Hillmann y colegas (2010), los estudios realizados por bibliotecarios de música y catalogadores de audiovisuales muestran que la definición de entidades del grupo 1 de los FRBR es altamente diferenciada cuando se trata de diferentes formas de expresión creativa. Además, hay elementos de RDA que están vinculados explícitamente a las orientaciones de RDA para una entidad FRBR, generalmente en virtud de la inclusión de la entidad en nombre del elemento. Ejemplos de estos últimos son elementos del tipo “Identificador de la Manifestación” y “Lenguaje de la Expresión”.

Los registros de metadatos han sido un tema de discusión desde hace más de una década. Con el aumento exponencial de informaciones, consecuentemente ocurre el crecimiento de los formatos o patrones de metadatos, y en ese contexto, la idea de reaprovechar y combinar los diversos elementos de la descripción es determinante.

Un registro público proporciona información sobre los patrones de metadatos en una estructura legible por máquinas, capaces de realizar la integración en aplicaciones específicas. La utilización de elementos de datos registrados aumenta la consistencia del uso de los mismos metadatos, sobre una variedad de servicios, ya que todas las aplicaciones se construyen bajo la misma configuración determinada por la máquina (Hillmann *et al.* 2010).

Según Hillmann (2010), un registro de metadatos eficiente facilita la declaración, la administración y el descubrimiento de esquemas de metadatos legibles por computadora, perfiles de aplicación y vocabularios controlados mejor estructurados. Con el registro de metadatos, aumenta la capacidad de interoperabilidad, así como la precisión de cualquier intercambio de elementos de datos y vocabularios. De esta manera, los registros proporcionan el significado para una red de información global, con una interacción mucho mayor entre los servicios de información.

De acuerdo con Hillmann *et al.* (2010), un aspecto clave de un registro que se debe considerar es que puede proporcionar un identificador único (URI) para cada elemento de datos y para cada elemento de un vocabulario.

Hillmann *et al.* (2010) aclaran que un objetivo importante del JSC RDA es destacado, donde las orientaciones pueden ser utilizadas más allá de la comunidad bibliotecaria, registrando elementos y conceptos que están disponibles para uso en aplicaciones en ambientes digitales con XML o RDF, lo que potencia la codificación futura de la descripción de los recursos informativos. Además, el RDA relacionado o vinculado con XML y RDF apoya la construcción de vocabulario y de las relaciones para la comunidad web semántica.

El desarrollo de RDA implica para las bibliotecas condiciones para avanzar rápidamente hacia un servicio de intercambio e interoperabilidad más amplio, reutilizando los datos bibliográficos de otros entornos informativos digitales. Los elementos y vocabularios RDA proporcionan la base para la migración del uso exclusivo del formato MARC, que es relevante y significativo en el dominio bibliográfico, transponiendo fronteras para otras comu-

nidades comprender, interpretar y utilizar la información más ampliamente.

Las discusiones recientes sobre el futuro de la biblioteca apuntan a la realidad que los grandes consumidores de metadatos bibliográficos como Amazon y Google Books han utilizado datos en formato MARC de forma que revelan cierta falta de comprensión de los metadatos de una biblioteca tradicional o convencional. (Hillmann *et al.* 2010).

Este hecho lleva una vez más a reforzar la tesis de que la comprensión y la aplicación del formato MARC 21 es ajustada solamente por un catalogador especialista (bibliotecario), pues éste posee las habilidades y herramientas necesarias en la construcción y la codificación de los metadatos, apoyado en reglas y esquemas de descripción que orientan la representación adecuada, estandarizada y unívoca de los recursos bibliográficos en el entorno digital.

Las metodologías de la catalogación descriptiva y una ontología para la descripción bibliográfica expresada o manifestada en los códigos y patrones de metadatos favorecen la recuperación de la información a partir del modelado del banco de datos y de la construcción de formas de representación de los recursos informacionales, posibilitando la interoperabilidad en entornos digitales. Se parte del principio de que los fundamentos epistemológicos y metodológicos de las formas de descripción de recursos ligados al contexto tecnológico, marcado por la iniciativa datos enlazados, garantizarán la consistencia y persistencia de los recursos informacionales en las capas delineadoras y estructurantes de los entornos informativos digitales, sobre todo, campo teórico de la construcción web semántica y en la efectividad de las relaciones semánticas.

MODELADO ESTRUCTURAL DE DATOS PARA LA CONSTRUCCIÓN
DE ENTORNOS DIGITALES INTEROPERABLES

Al pensar en la creación y el desarrollo de un modelo de datos, nos encontramos con las cuestiones de granularidad y análisis de los datos a ser catalogados. Esto no es algo nuevo en el dominio bibliográfico, donde se desarrolló un modelo de datos implícito en las reglas de descripción (AACR2) y en los formatos de intercambio de datos bibliográficos (MARC 21).

Actualmente, con el desarrollo de nuevas reglas y estándares para el modelado de los ambientes informativos como FRBR, FRAD y RDA, se reconoce que hay una tendencia en la estructuración y la definición de los datos a ser catalogados preparándolos para moverlos y hacerlos compatibles con la web semántica y los datos enlazados.

En las capas intangibles de los datos (definida en la representación y descripción de los recursos informacionales), hay un aumento en la estructura y la granularidad de los datos. Yee (2009) apunta que más estructura y más granularidad posibilitan presentaciones más sofisticadas a los usuarios de los sistemas y aumentan la posibilidad de producción de datos interoperables.

Cualquier cambio o asignación que se contrató para crear datos interoperables produciría un menor denominador común (los datos más simples y menos granulares), y una vez interoperables, no sería posible su recuperación en su totalidad debido a su pérdida. Los datos con menos estructura y menos granularidad podrían ser más fáciles y más baratos para aplicarlos y tener el potencial más simple para las comunidades involucradas (Yee 2009, 59)⁸.

⁸ La traducción es del autor.

Vamos a tomar como ejemplo un nombre personal. Conforme a las reglas de catalogación establecidas por el código AACR2 demarca el apellido del nombre propio, se registra primero el apellido, seguido por una coma y luego el nombre. Esta cantidad de granularidad puede representar a menudo un problema para el catalogador en una cultura desconocida, que no domina necesariamente las reglas. Más granularidad puede ocasionar situaciones ambiguas para los sujetos que están recogiendo los datos. “Otro ejemplo es en cuanto al género del creador, donde el catalogador podría encontrarse con una situación, que no sabría si lo definía como masculino o femenino” (Yee 2009, 59)⁹.

Yee (2009) comenta que si añadimos una fecha de nacimiento y muerte, sea cual fuere, las utilizamos juntas en un subcampo \$ d sin ninguna codificación separada para indicar cuál es el nacimiento y cuál es la muerte (aunque un ocasional “b” o “d” nos dirá ese tipo de información). Podríamos proporcionar más granularidad para fechas, pero haría el formato MARC 21 mucho más complejo y difícil de aprender.

En la representación del campo 100 (autor personal), del patrón de metadatos MARC 21, por ejemplo, la forma autorizada para describir el contenido se define de la siguiente manera:

100 1#\$a Adams, Henry, \$d1838-1918.

En este caso, el subcampo \$ d (fechas asociadas al nombre NR), 1838 corresponde a la fecha de nacimiento, mientras que en 1918 a la fecha de fallecimiento del autor.

De acuerdo con Yee (2009), granularidad y estructura también pueden crear “tensión” una con la otra. Más granularidad

⁹ La traducción es del autor.

puede conducir a una estructura más pequeña (o más complejidad para mantener la estructura junto con la granularidad). En la búsqueda de mayor granularidad de datos (RDA intenta apoyar RDF en la codificación XML), se han atomizado los datos para hacerlos útiles a las computadoras, pero eso no necesariamente hará los datos más comprensibles a los humanos.

Para ser útil a los humanos, debe ser posible agrupar y organizarlos de forma significativa para la catalogación, la indexación y la presentación de los mismos. Los desarrolladores del *Simple Knowledge Organization System* (SKOS) se refieren a la cantidad de información no estructurada (es decir, legible por humanos) en la web, y etiquetan bits de datos como relaciones semánticas de los registros en una máquina accionable, de forma que no necesariamente proporciona el tipo de estructura necesaria para hacer los datos legibles por humanos y, por lo tanto, útiles para las personas en la web (Yee 2009).

Para reforzar su pensamiento, Yee (2009, 59) afirma que:

Cuanto más granular los datos, menos el catalogador puede construir orden, secuenciación y la conexión de los datos; la codificación debe estar cuidadosamente diseñada para permitir el orden, la secuenciación y la conexión de los datos deseados, para que la catalogación, la indexación y la presentación sean posibles, lo que podría llamarse una codificación de los datos aún más compleja”.¹⁰

En lo que se refiere a la estructura de datos, Yee (2009) dice que existen varios significados ligados al término, conforme se puede observar.

¹⁰ La traducción es del autor.

- Estructura es el cuerpo de un registro (estructura de documento); por ejemplo, Elings y Waibel se refieren a “campos de datos [...] también referidos como elementos [...] que están organizados en un registro por una estructura de datos”.
- Estructura es la capa de comunicación, a diferencia de la capa de presentación o designación de contenido.
- Estructura es el registro, el campo y el subcampo.
- Estructura es la conexión de bits de datos en conjunto en la forma de varios tipos de relaciones.
- Estructura es la presentación de los datos de manera estructurada, ordenada y secuenciada para facilitar la comprensión humana.
- Estructura de datos es la forma de almacenamiento de los datos en una computadora para que pueda ser utilizada eficientemente (es decir, cómo los programas de computadoras utilizan el término).

Castro (2012) define la estructura de datos, la capa intangible de instanciación de los datos bibliográficos modelados para la representación y la descripción, al igual que los formatos o patrones de metadatos, para permitir la interoperabilidad de los entornos informativos digitales, por agentes humanos y no humanos, garantizando interfaces más accesibles a los usuarios para posterior recuperación, uso y (re)uso de los recursos informacionales.

Cuando se menciona estructura de datos en el dominio bibliográfico, se piensa en el modelo conceptual de datos establecido por los FRBR.

FRBR hace uso de un modelo entidad-relación, el cual consiste en dos principales conceptos: “cosas” y relaciones. FRBR define diez categorías “cosas”, las cuales se denominan entidades: Obra, Expresión, Manifestación, Elemento, Persona, Entidad colectiva, Concepto, Objeto, Evento y Lugar. Las entidades pueden ser interpretadas, por ejemplo, como una obra, un texto, un libro, etcétera. Los atributos corresponden a las características de los datos relacionados con la entidad y sirven para diferenciar el contenido intelectual o artístico. Las relaciones describen los vínculos entre una entidad y otra en la facilitación de manejo del recurso informacional por el usuario en un sistema (Moreno 2006).

Para Castro (2012), los bibliotecarios que se acostumbren a los FRBR probablemente no encontrarán mucha dificultad en la transición del modelo conceptual al modelo RDF. Es importante en ese momento destacar las principales terminologías empleadas por los FRBR, RDF y RDFS, de acuerdo con el Cuadro 1.

Cuadro 1. Diferencias terminológicas en los modelos de datos

FRBR	RDF	RDFS
Entidad	Asunto	Clase
Atributo	Objeto	Propiedad
Relaciones	Predicado	Relaciones/Relación Semántica

Fuente: Yee (2009, 64).

En este contexto, Riley (2010) profundiza el estudio comparativo terminológico del RDF en el contexto de la bibliotecología y de la ciencia de la información, especificando:

- **Asunto:** en bibliotecas, lo que un recurso informativo abarca en términos de contenido; en RDF, lo que dice una declaración sobre algo (recurso informacional).
- **Vocabulario:** en bibliotecas, implica un cierto tipo de vocabulario controlado (términos autorizados, estructuras jerárquicas, términos relacionados, etc.); en RDF, definiciones mucho más flexibles (incluye definiciones formales de clases y de dominio de un recurso informacional).
- **Clase:** en bibliotecas, un esquema de clasificación (Clasificación Decimal de Dewey - CDD, Clasificación Decimal Universal - CDU etc.) que indique el tópico general o área del conocimiento abarcado por el recurso informacional; en RDF, un tipo o categoría perteneciente a un objeto o recurso informacional.
- **“Schema”:** XML Schema define un conjunto de elementos destinados a ser usados juntos; RDF Schema define clases y propiedades destinadas a ser usadas en cualquier lugar, solas o en combinación.

La dificultad en cualquier ejercicio de modelado de datos, sobre todo en el dominio bibliográfico, se encuentra en decidir qué tratar como entidad o clase y qué tratar como un atributo o propiedad. FRBR decidió crear una clase llamada expresión para tratar cualquier cambio en el contenido de una obra (Castro 2012). Los FRBR en comparación con el

modelo de datos RDF, se encuentran en armonía, las entidades de los FRBR se registran como clases, mientras que las relaciones se registran como dominios.

FRBR en RDF¹¹ añade sólo tres clases. Dos de ellas (Endeavor y Entidad Responsable) son súper conjuntos de clases FRBR. Endeavor es una generalización que puede ser relacionada a la obra, expresión, o manifestación, o sea, una clase cuyos miembros son cualquiera de los productos de la actividad artística o creativa. Entidad Responsable es un término más general que puede relacionarse con una entidad colectiva o una persona. Estas clases especifican más claramente información sobre el contenido intelectual de un recurso, sin necesidad de proporcionar información adicional. “La tercera clase que se agrega es el asunto. Las tres incluyen alguna instancia del asunto en sus esquemas. FRBR trata claramente el asunto como una relación” (Davis y Newman 2005; Coyle 2012).

La condición de los entornos informativos digitales interopera sus datos es el modelado conceptual, está definida y codificada por una serie de requisitos funcionales establecidos por las arquitecturas de metadatos, reglas y esquemas de descripción bibliográfica y ontologías, que proporcionará entornos mejor estructurados, en la garantía de recuperación de información más efectiva a los usuarios (humanos y no humanos).

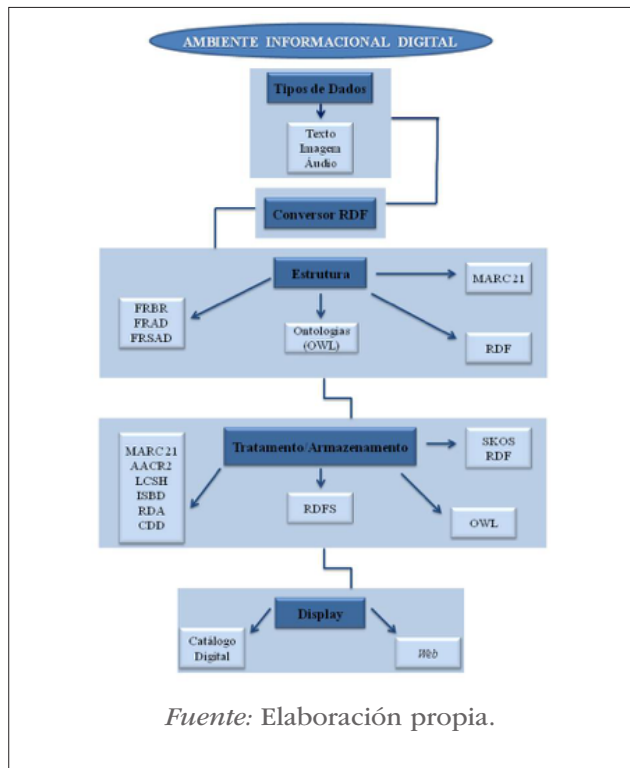
PRESENTACIÓN DE UN MODELO PARA LA DESCRIPCIÓN BIBLIOGRÁFICA SEMÁNTICA

Se percibe que la consigna en la constitución de entornos infor-

11 Las expresiones del núcleo de conceptos y vocabularios FRBR en RDF pueden ser mejor visualizados en el sitio <http://vocab.org/frbr/core.html>.

macionales digitales, sobre todo en el ámbito de las bibliotecas digitales, es la interoperabilidad. Como un punto de partida para garantizar la interoperabilidad, esta investigación, a partir de las observaciones y reflexiones en el campo de la catalogación descriptiva, destaca algunos requisitos y directrices funcionales que pueden utilizarse para establecer la interoperabilidad en entornos informacionales digitales de una forma más efectiva. Para ello, se pretende comprender la estructura intangible en una propuesta presentada en capas superpuestas conforme a la figura 1, ya que ellas deben trabajar en sinergia para la consistencia y el funcionamiento pleno del entorno digital.

Figura1. Propuesta de modelado para la interoperabilidad en el ámbito bibliográfico



Capa 1 - Tipología de los datos: en esta fase inicial, el proyectista (catalogador) define qué datos serán trabajados para alimentación y modelado del ambiente informativo, a partir del recurso bibliográfico a ser catalogado. Ejemplo: datos textuales, imágenes, audio, etcétera. Es importante destacar que en este trabajo se contemplan sólo en los datos textuales, explícitos en los códigos de catalogación (AACR2) y en los patrones de metadatos (MARC 21).

Capa 2 - Preparación de los datos: una vez definidos los datos bibliográficos a ser utilizados en el sistema, la preparación de los datos consiste en la adopción de herramientas para la conversión¹² de datos en RDF. Los datos que fueron extraídos de otras fuentes; es decir, datos no RDF, el W3C ha recomendado algunos convertidores que auxilian la transformación, como RDFizer. La adopción del RDFizer se justifica ya que no se tienen oficialmente representaciones de datos MARC en RDF. En ese sentido, tal convertidor transforma datos del patrón de metadatos MODS en RDF, iniciativa que promovió la Biblioteca del Congreso (MODS, 2011).

Capa 3 - Tratamiento y Almacenamiento de los datos: Después de la conversión de los datos en RDF, la siguiente capa consiste en la adopción efectiva por el catalogador de las reglas o de los esquemas de descripción bibliográfica (AACR2 y RDA); es decir, la catalogación de los recursos bibliográficos, en la confección estandarizada de metadatos; la definición de los patrones de metadatos (MARC 21), la

12 Un convertidor para RDF es una herramienta que convierte aplicaciones de datos de un formato específico en RDF para su uso con herramientas de RDF e integración con otros datos. Los convertidores pueden ser parte de un esfuerzo de migración o parte de un sistema en ejecución que proporciona una vista Web semántica de una aplicación concreta (W3C 2012). La traducción es del autor.

arquitectura de metadatos RDF para la estructuración de datos y RDF Schema para su validación. En este momento, el catalogador deberá también adoptar los FRBR, en consonancia con los FRAD y los Requisitos Funcionales para Datos de Asunto (FRSAD). Las ontologías aparecen en ese contexto para definir los conceptos de los elementos de un registro bibliográfico, por medio de las reglas y esquemas de descripción para la confección metodológica de los metadatos y de los patrones de metadatos.

Capa 4 - Presentación (*display*) de los datos: La fase final consiste en hacer disponibles los datos (*output*) y presentarlos a los usuarios del ambiente informacional. Los datos pueden aparecer de la forma en que fueron contruidos y almacenados (*input*) en las capas 1 y 3, para la capa tangible de recuperación y también visualizada en la web.

Se cree que estos requisitos y recomendaciones pueden propiciar un modelado de los catálogos mejor estructurados para posterior recuperación, uso y re (uso) de las informaciones, lo que garantizará la interoperabilidad y potenciará las relaciones bibliográficas semánticas, iniciativa que va en contra de los ideales vislumbrados por la web semántica y datos enlazados.

CONSIDERACIONES FINALES

Se verifica, en el panorama actual del dominio bibliográfico, que los modelos conceptuales propuestos por los FRBR y las bases ontológicas de los esquemas de metadatos pueden propiciar la claridad en las relaciones entre registros bibliográficos.

Es importante señalar que el modelo FRBR facilita el diseño de un modelo conceptual, en consonancia con las

ontologías y los metadatos, no sólo por su riqueza conceptual y estructural, sino porque se constituye en un marco de referencia para la comprensión de las relaciones bibliográficas y en el modelado de los catálogos digitales.

Los entornos informativos digitales al adoptar la arquitectura de metadatos RDF, junto con la ontología y los esquemas de metadatos, pueden encontrar un camino para el establecimiento de la interoperabilidad semántica entre sistemas heterogéneos y distintos.

El uso de los modelos conceptuales de datos, de las arquitecturas de metadatos y de las ontologías rediseñan los nuevos entornos informacionales digitales, definiendo conceptualmente los elementos de la descripción bibliográfica a ser representados por el catalogador, de modo que proporciona interfaces de búsqueda más comprensibles a los usuarios y en el establecimiento efectivo de la interoperabilidad.

El modelo RDF subyacente a las tecnologías, web semántica se describe a menudo como el futuro de metadatos estructurados. Su adopción en bibliotecas ha sido lenta, justificada por el hecho de que las diferencias fundamentales están en el enfoque de modelado que lleva RDF, que representa una arquitectura *bottom up* donde las descripciones se distribuyen y se puede hacer bajo cualquier característica considerada necesaria, mientras que la el enfoque “céntrico” de registros bibliográficos de las bibliotecas tiende a ser más *top down* al confiar en las funciones predefinidas determinadas por ellas (Lee y Jacob 2011; Yee 2009; Riley 2010; Coyle 2012).

La ontología y los metadatos utilizados de forma estratégica y sinérgica para garantizar la integridad de la información proporcionan elementos de descripción más ricos semánticamente y proporcionan las relaciones entre un conjunto de esquemas, originarios de entornos informales heterogéneos, lo que garantiza la interoperabilidad semántica

y multidimensiona las formas de acceso a los recursos para la posterior utilización y reutilización de la información.

La propuesta de modelado desarrollada en este estudio puede posibilitar el intercambio entre patrones de metadatos y entornos y sistemas informacionales distintos trabajando en una filosofía de colaboración entre los recursos informativos disponibles y las tecnologías que están abarcadas en su construcción, en el establecimiento de la interoperabilidad, en la optimización de las relaciones bibliográficas, en la conexión y vinculación de los datos y ampliados para la construcción estandarizada de recursos en la web.

BIBLIOGRAFÍA

- Arakaki, F. A. 2016. *Linked data: ligação de dados bibliográficos*. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.
- Berners-Lee, T. 2006. *Linked Data*. Cambridge: W3C. Disponible el 24 de septiembre de 2017 en <http://www.w3.org/DesignIssues/LinkedData.html>.
- _____. 2010. “Is your Linked Open Data 5 star?” En Berners-Lee, T. *Linked Data*. Cambridge: W3C. Disponible el 20 de noviembre de 2017 en <http://www.w3.org/DesignIssues/LinkedData.html>.
- Blattmann, U. y F. C. C. da Silva. 2007. “Colaboração e interação na web 2.0 e biblioteca 2.0”, *Revista ACB: Biblioteconomia em Santa Catarina*, Florianópolis, vol. 12, núm. 2: 191-215.
- Castro, F. F. de. 2012. *Elementos de interoperabilidade na catalogação descritiva: configurações contemporâneas para a modelagem de ambientes informacionais digitais*. 2012. 202 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.
- Coyle, Karen. 2018. *Bibliographic framework: RDF and linked data*. Berkeley, United States, Jan. 2012. Disponível em: < <http://kcoyle.blogspot.com/2012/01/bibliographic-framework-rdf-and-linked.html> >. Acesso em: 20 jan.
- _____. 2011. MARC 21 as data: a start. *Code4Lib Journal*, vol. 14.

- Davis, Ian y Richard Newman. 2005. *Expression of core FRBR concepts in RDF*. Disponible el 20 de enero de 2012 en <http://vocab.org/frbr/core.html>.
- Delsey, Tom. 2010. *RDA: resource description and access*. Disponible el 25 de febrero de 2012 em <http://www.rda-jsc.org/rda.html>.
- Ferreira, J. A. y P. L. V. A. da C. Santos. 2013. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. *Informação & Sociedade: estudos*, vol. 23, núm. 2: 13-23.
- Hillmann, Diane. *et al.* 2010. "RDA vocabularies: process, outcome, use", *The Magazine of Digital Library Research*, vol. 16, núm. 1/2.
- Lassila, O. y R. R. Swick. 1999. *Resource description framework* (RDF) model and syntax specification. Disponible el 3 de julio de 2017 en <http://www.w3.org/TR/1999/REC-rdfsyntax-19990222/>.
- Lee, Seungmin y Elin K. Jacob. 2011. "An integrated approach to metadata interoperability: construction of a conceptual structure between MARC and FRBR", *Library Resources & Technical Services*, vol. 55, núm. 1: 17-32.
- MODS (*Metadata Object Description Schema*). 2011. *Library of Congress*. Disponible el 21 de mayo de 2018 en <http://www.loc.gov/standards/mads/rdf/>.
- Moreno, F. P. 2006. *Requisitos funcionais para registros bibliográficos - FRBR: um estudo no catálogo da rede bibliodata*. 202 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

- Riley, Jenn. *RDF for librarians*. 2010. DLP Brown Bag Series, p. 1-38. Disponible el 15 de noviembre de 2017 en <http://www.dlib.indiana.edu/education/brownbags/fall2010/rdf/rdf.pdf>.
- Santos, Plácida, L. V. A. C.; Alves, Rachel C. V. 2009. "Metadados e web semântica para a estruturação da web 2.0 e web 3.0." en: *Datagramazero – Revista de Ciência da Informação*, vol. 10, núm. 6.
- Taylor, A. G. 2004. *The organization of information*. Westport: Libraries Unlimited.
- Thomale, Jason. 2010. *Interpreting MARC: where's the bibliographic data?* *Code4Lib Journal*, Vol. 11.
- Van Hooland, S. y R. Verborgh. 2015. *Linked Data for libraries, archives and museums: how to clean, link and publish your metadata*. Londres: Facet.
- W3C Library Linked Data Incubator Group. 2011. *Final report*. Cambridge. Disponible el 12 de diciembre de 2017 en <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025>.
- W3C. *Converter to RDF*. 2012. Disponible el 21 de enero de 2012 en <http://www.w3.org/wiki/ConverterToRdf>.
- Yee, Martha M. 2009. "Can bibliographic data be put directly onto the semantic web?". En: *Information Technology and Libraries*, vol. 28, núm. 2: 55-80.
- Zeng, M. L y J. Qin. 2008. *Metadata*. Nueva York: Neal-Schuman Publishers.

De los metadatos para la organización de la información a la tecnología *middleware* para los servicios de las bibliotecas: la biblioteca inteligente

GERARDO BELMONT LUNA
Universidad Iberoamericana, México

INTRODUCCIÓN

La presente investigación aborda el tema de las tecnologías *middleware*, entendidas como el acueducto de los datos. Su novedad radica en que no modifica ningún sistema de información ni exige tecnologías ideales. Por su arquitectura distribuida y el protocolo de Internet (IP) para la comunicación, su capacidad se potencia en la medida en que a la red se suman nuevos equipos personales y servidores. Su principal característica es integrar los datos y las aplicaciones existentes en una plataforma de manera distribuida. Esto permite resolver el indeterminismo tecnológico que los requisitos funcionales para los registros bibliográficos (FRBR, por sus siglas en inglés) ha presentado desde su reporte final, antes de su liberación e implementación en el año 2013. En un sentido más amplio, el uso de esta tecnología amplía las posibilidades sobre un tema que siempre ha distinguido a las bibliotecas: la cooperación.

¿Por qué las bibliotecas necesitamos de la tecnología *middleware*? El siglo XX es un hito en la historia respecto a los avances tecnológicos por la rapidez y presencia en la vida humana. El quehacer social es influido por la implementación de sistemas de almacenamiento de información, recuperación y difusión sin precedentes. La biblioteca recoge de forma exitosa estos avances y los incorpora a las tareas sustantivas. Pero el reto no terminó ahí; se ha extendido con mucha mayor fuerza durante este siglo XXI, lo que implica un trabajo continuo en la adopción de las tecnologías que se actualizan, lo que da nuevos bríos a la comunión biblioteca-tecnología.

Uno de los principales problemas es la representación de los datos entre las distintas plataformas tecnológicas, como los sistemas operativos, servidores, manejadores de bases de datos, lenguajes de programación, lenguajes de marcado, entre muchas otras tecnologías. Por otra parte, la biblioteología ha desarrollado estándares para la organización de la información que se ha mantenido con cierta estabilidad y uniformidad desde el siglo XIX hasta nuestros días. Con la renovación a partir de la descripción y acceso a los recursos (RDA por sus siglas en inglés), la práctica catalográfica toma un nuevo impulso, y se abren nuevas preguntas, nuevos horizontes de acción y nuevas posibilidades para cumplir con los objetivos tradicionales con esquemas tecnológicos actuales.

Entonces ¿qué tecnología requiere la biblioteca para competir frente a los buscadores, sitios y/o concentradores de información? ¿qué características y alcances debe tener el OPAC en una sociedad como la actual? ¿cómo constituirlo? El profesional de la información debe considerar que estas nuevas tecnologías están trazando un nuevo ecosistema tecnológico en el que conviven necesidades de información a varios niveles operativos. El usuario es

multiusuario, de modo que las habilidades del profesional se van multiplicando también en la organización de la información. Finalmente, un tema colateral pero no menos importante es la visualización que de ese universo tiene el usuario, lo cual significa nuevas ópticas y, por lo tanto, otras implicaciones en la organización y construcción del conocimiento.

En este orden de ideas, el objetivo de este trabajo es analizar las posibilidades tecnológicas de *middleware* con miras a la tecnología de grafos que el siglo XXI nos ofrece y que existen en el mercado para lograr la representación y obtención de datos no sólo cuantitativos, sino la construcción de consultas semánticas con una mejorada representación del universo de la información, dando oportunidad a los sentidos del usuario para la interpretación y descubrimiento del mismo.

Comenzaremos con la descripción general de la tecnología *middleware*, explicaremos la arquitectura posible de BSmart, así como su modelo de operación, que incluye el modelo de datos.

LA TECNOLOGÍA *MIDDLEWARE*

La historia de la tecnología *middleware* (MDWR) se remonta a la década de los sesenta, cuando en la NATO Software Engineering Conference se utilizó por primera vez este término para nombrar al software utilizado para adaptar la funcionalidad del sistema de archivos genérico a las necesidades específicas de la aplicación. El reporte dice que esto se debe a que “no importa lo bueno sea el *software* del fabricante para elementos como el manejo de archivos, simplemente no es adecuado, es ineficiente e inapropiado” (Software Engineering 1969,14).

A continuación explica con más precisión las tareas que el *software* no hace, y que el usuario debe volver a armar. Más recientemente, el término se entiende como “un software situado entre el nivel de aplicación (interfaz de usuario) y los niveles de transporte e inferiores (que suele presentarse como el interfaz o API de programación de los distintos sistemas operativos)” (García *et al.* 2005).

NickGall (citado en RedHat 2018) es todavía más pragmático y define el *middleware* como el software que se trata de un software. En resumen, la tecnología *middleware* es aquella que permite el flujo de datos de una aplicación a otra. Esto es posible gracias a que reúne tres elementos fundamentales: protocolo simple de acceso a objetos (SOAP), servicios web, transferencia de estado representacional (REST) y notación de objetos JavaScript (JSON). Entre las tecnologías *middleware* más extendidas, podemos destacar las siguientes: CORBA (Common Object Request Broker Addressing), JavaRMI (Remote Method Invocation) y .NET Remoting. Estas tecnologías son las facilitadoras para los robots de control y gestión de software.

Actualmente, la tecnología MDWR está presente en casi todas las actividades económicas, políticas, sociales y culturales. En las bibliotecas, los museos y los centros de información, por ejemplo, se hace presente cuando se compran o adquieren licencias de accesos a los servicios a través de agentes como EBSCO o Elsevier. Éstos se encargan de conectar a las bases de datos de corte científico y generan los índices para su comercialización. En general, la interacción con la tecnología MDWR cada vez es más compleja.

Los temas de big data se agudizan sobre todo cuando existe la necesidad de la seguridad y vigilancia en las transacciones electrónicas, así como en ambientes que se perfilan a

la Inteligencia Artificial (IA), en el seguimiento de personas (*people tracking* en inglés) y el Internet de las cosas (IoT, por sus siglas en inglés), aprovechando la minería de datos en diferentes formatos con diferentes algoritmos, lo que aumenta la capacidad de respuesta en tiempo real. Lo anterior hace posible que los sistemas de información adquieran heterogeneidad al garantizar la consistencia y la disponibilidad la información en períodos asíncronos.

METODOLOGÍA

Es cierto que existe una amplia variedad de tecnologías *middleware*, algunas de paga y otras de acceso libre. Para garantizar a los interesados la réplica de este trabajo, se adoptan tecnologías de *software* libre. Particularmente, se hace referencia al manejador de bases de datos NoSQL, útil para la resolución de complejidades tecnológicas de las bases de datos. Asimismo, al ser un *software* que trabaja sobre servidor, también resuelve el problema de comunicación entre los sistemas operativos.

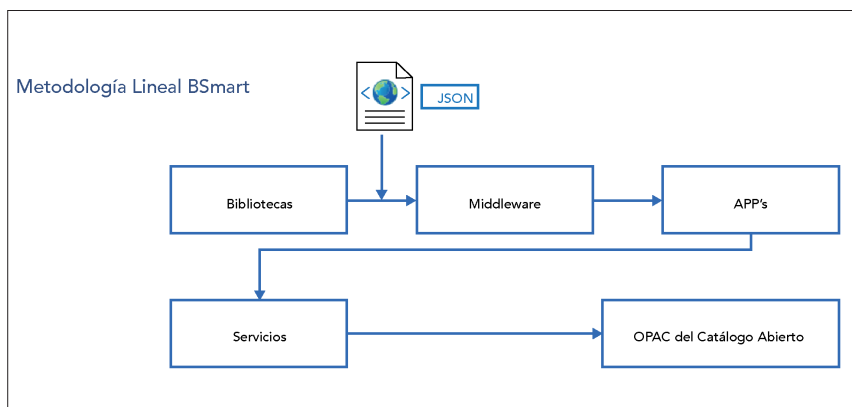
La metodología se caracteriza por el diseño de un proceso mejorado al que denominaremos BSmart (Biblioteca Inteligente). La definimos como una entidad que puede procesar información estableciendo relaciones y posibles vínculos con otros elementos de tipo externo que se entretajan entre sí dentro del universo de la información. Su propósito es identificar aquellos elementos de las entidades de información que hacen que el universo de la información se pueda relacionar bajo las siguientes premisas:

1. Garantizar la interoperabilidad de las bases de datos.

2. Que haga las funciones de interfaz con las capas superiores, lo que posibilita que las bibliotecas participantes puedan desarrollar su propia interfaz gráfica.
3. Hacer de traductor entre las bases de datos de bibliotecas y de cualquier otra fuente.

En la ilustración 1 se muestra la metodología sugerida:

Ilustración 1. Metodología lineal de la BSmart

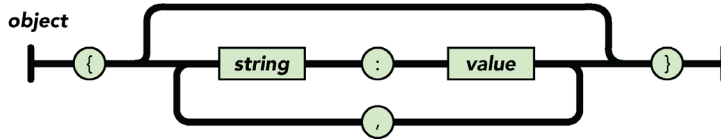


Fuente: Elaboración propia.

En donde:

- JSON, (JavaScript Object Notation - Notación de Objetos de JavaScript) es un formato ligero de intercambio de datos que en conjunto con NoSQL puede resolver el complejo de interoperabilidad del catálogo abierto. La ilustración 2 muestra el sistema de notación.

Ilustración 2. (ECMA-404 The JSON Data Interchange Standard)



- Bibliotecas es la base de datos propiamente de cada biblioteca; su característica principal, es que está dispuesta a través de la red bajo el Protocolo de Internet (IP). Esta IP debe ser pública.
- *Middleware* es el software que se ha definido para resolver la complejidad tecnológica y obtener la interoperabilidad de las bases de datos de las bibliotecas.
- APP's es la aplicación informática para la interacción de los servicios, y también, son los monitores y captadores de información de los usuarios.
- Los servicios son las acciones de la biblioteca que se establecen bajo un cuadro de normas y políticas que responden a las necesidades de información de los usuarios.
- OPAC, como sabemos, es el catálogo en línea.
- Catálogo abierto, considerado el

[...] instrumento que describe las relaciones de las entidades de información en donde es posible relacionar los elementos internos que dan identidad a una obra, así como a los elementos externos que hacen posible su vinculación con otros elementos de tipo externo que establecen una correspondencia entre sí (imaginemos una

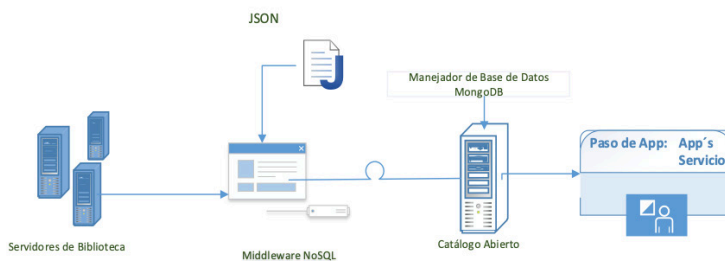
figura imposible). El atributo de figura imposible elimina las barreras tradicionales de la biblioteca, por lo que es posible establecer relaciones y procesos de innovación que se verán reflejados en los servicios de información de las bibliotecas y en general de las unidades de información, enfrentado la complejidad y el caos del universo de la información (Belmont 2014).

Finalmente, podemos decir que la interacción BSmart debe garantizar a las bibliotecas:

- Mantener su identidad; es decir, no debe alterar su infraestructura.
- Poseer una IP fija es la base de la comunicación para este sistema distribuido.
- Renovar su servicios; es uno de los objetivos de esta implementación es generar servicios basados en la transformación de las nuevas demandas que los usuarios prefieren de otras aplicaciones web.
- Trabajar con aplicaciones es la base para el seguimiento y predicción para acompañar al usuario.
- Acercarse al modelo de *business intelligence* es la forma de nombrar la revolución de los nuevos servicios de información. Basados en el aprovechamiento de *big data*.

1. ARQUITECTURA POR COMPONENTES BSMART Y MODELO DE OPERACIÓN

Ilustración 3. Arquitectura por componentes BSmart

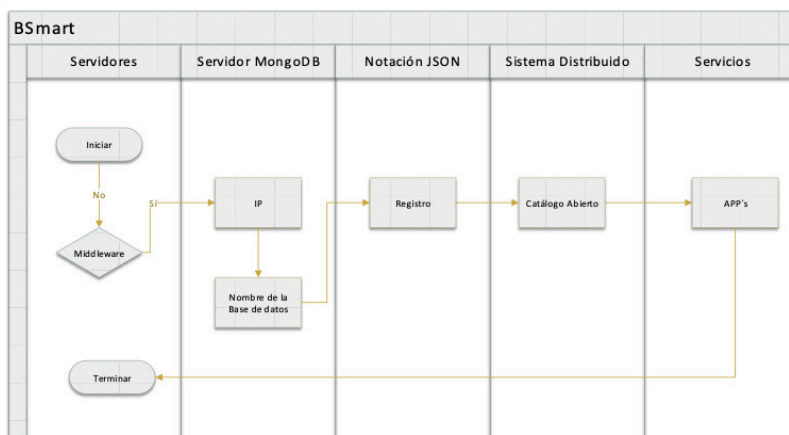


Fuente: elaboración propia.

Fuente: Elaboración propia.

Proceso de integración

Ilustración 4. Proceso de integración BSmart



Fuente: Elaboración propia.

El proceso de integración es simple: consta de el servidor 0 (cero), que dará los altas y permisos a las IP que se integren de manera distribuida al catálogo abierto. El metodo de notación es JSON, como se muestra en el siguiente ejemplo:

```
BSmart = {CA = array
[
    {"Id" : 0,},
    {"Id" : 1,},
    {"Id" : 2,}
]
```

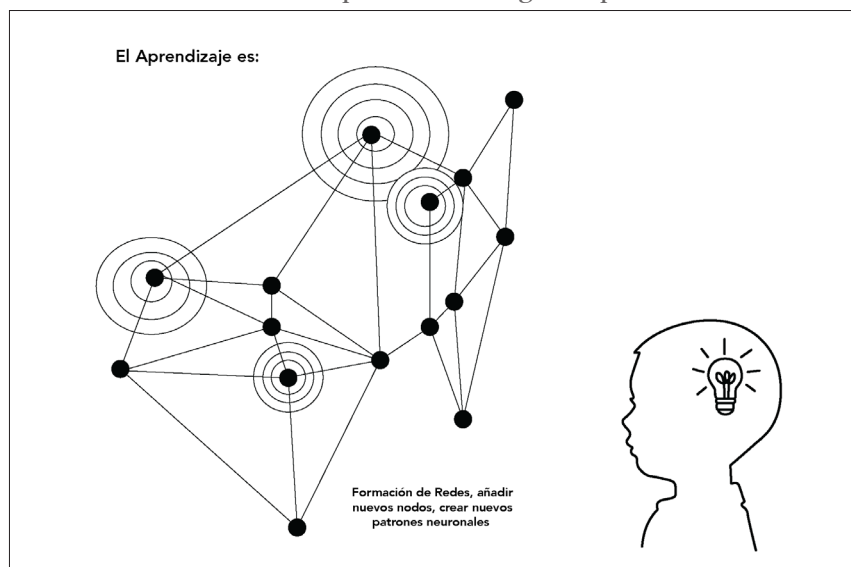
En un ejercicio real, la notación sería de la siguiente manera:

```
{
  "Id" : 0,
  "name" : "CA-0:27017",
  "health" : 1,
  "state" : 1,
  "stateStr" : "PRIMARY",
  "uptime" : 7420,
  "optime" : Timestamp(1433415310, 1),
  "optimeDate" : ISODate("2018-05-29T08:57:10Z"),
  "electionTime" : Timestamp(1433412721, 2),
  "electionDate" : ISODate("2018-05-29T08:57:10Z"),
  "self" : True
},
{
  "Id" : 1,
  "name" : "localhost:27017",
  "health" : 1,
  "state" : 1,
```

```
    "stateStr" : "PRIMARY",
    "uptime" : 7420,
    "optime" : Timestamp(1433415310, 1),
    "optimeDate" : ISODate("2018-05-29T08:57:10Z"),
    "electionTime" : Timestamp(1433412721, 2),
    "electionDate" : ISODate("2018-05-29T08:57:10Z"),
    "self" : True
  },
  { "Id" : 2,
    "name" : "192.258.258.***:27017",
    "health" : 1,
    "state" : 1,
    "stateStr" : "PRIMARY",
    "uptime" : 7420,
    "optime" : Timestamp(1433415310, 1),
    "optimeDate" : ISODate("2018-05-29T08:57:10Z"),
    "electionTime" : Timestamp(1433412721, 2),
    "electionDate" : ISODate("2018-05-29T08:57:10Z"),
    "self" : True
  },
```

En donde se nombra cada entidad integradora por su IP, su Id, que es el lugar consecutivo que van ocupando al momento de ser agregadas al sistema distribuido del CA. Cabe hacer mención de que el sistema de notación estará presente en cualquiera de las tareas y procesos de la BSmart. Finalmente, la BSmart propone la representación de la información basada en relaciones. Sabemos que cada día la información crece de manera exponencial, así como sus manifestaciones y soportes. Ante este fenómeno, creemos que los grafos son la tecnología recomendable para su comprensión. El tema merece su trabajo aparte.

Ilustración 5. La BSmart representada en grafos para el conocimiento



Fuente: Imágenes tomadas de Internet.

DISCUSIÓN

Esta contribución representa una oportunidad de innovación para las bibliotecas y da respuesta a las reflexiones y aportaciones que, en el marco del futuro de la catalogación, usando RDA, se han venido dando en los foros de especialistas.

Por otra parte, el catálogo abierto rebasa fronteras ideales y sesgos de cualquier índole con el fin de coincidir en la cooperación catalográfica, donde no exista el recurso profesional de un bibliotecólogo, o bien, es útil para que las bibliotecas recuperen a los usuarios frustrados por no encontrar la información requerida. Principalmente, las bibliotecas

iberoamericanas podrían beneficiarse de este tipo de desarrollos, que detonen la capacidad de crear y hacer catalogación con sentido para un bien común universal.

CONCLUSIONES

Este trabajo es una invitación a pensar en nuevas posibilidades de construir catálogos con ayuda de la tecnología de nueva generación. Ante una sociedad cada vez más demandante y productora de información y conocimiento, es necesario poner sobre la mesa todas las posibilidades en el tratamiento de la información registrada. El valor agregado que se imprima en la organización y representación del conocimiento, más allá de la descripción que hasta hoy se ha realizado, tal vez marcará el importante paradigma que está por insertarse en la práctica catalográfica. Puntualizando los aspectos más relevantes de este trabajo de investigación, se concluye:

- Las tecnologías Middleware pueden renovar la cooperación bibliotecaria con un alcance más amplio.
- El catálogo de una biblioteca puede potenciarse tanto como se quiera al concebirse como un catálogo abierto y considerar al dato como elemento multidimensional.
- Este trabajo es un punto referencia realizar una revisión crítica sobre el quehacer del bibliotecólogo en materia catálogos compartidos.

- El futuro paradigmático de las bibliotecas requiere mente abierta, creatividad e imaginación para enfrentarlo.

Trabajo futuro

- La disciplina bibliotecológica debe proveerse de profesionales con mayores habilidades y destrezas en ciencia de datos.
- El modelo FRBR tienen muchas aristas. Por esta razón, merece estudios desde la teoría de conjuntos y teoría de grafos.
- La sociedad del conocimiento, la sociedad del siglo XXI, demanda de las bibliotecas mayor presencia en sus actividades cotidianas.
- Finalmente, estamos lejos de alcanzar el cielo, la escasez de profesionales con nuevas habilidades será atendida en la manera en que nuestros planes de estudio se reformulen. Es inevitable.

BIBLIOGRAFÍA

- Belmont Luna, Gerardo. 2014. *La catalogación en el entorno de las tecnologías de la información: aplicación y desarrollo del modelo conceptual FRBR*. México: FFyL-UNAM. p.19.
- Fernández Rivas, Alexis; Oswaldo Andrés Pérez García y José Henández Palancar. 2016. *Estado del arte sobre aplicaciones de middleware ROS*. CENATAV. Disponible en http://www.cenatav.co.cu/doc/RTecnicos/RT%20SerieGris_033web.pdf.
- Functional Requirements for Bibliographic Records*. 1998. IFLA. K. G. Saur München.
- García Sánchez, F.; A. J. García Sánchez; P. Pavón Mariño y J. García Haro. 2005. *Tecnologías Middleware: soluciones actuales para grandes empresas y proyectos*. p.1.
- Introducción a JSON*. Disponible en <http://www.json.org/>.
- Jackson, Andy. 2015. "The Provenance of Web Archives." British Library UK Web Archive. Disponible en <http://britishlibrary.typepad.co.uk/webarchive/2015/11/the-provenance-of-web-archives.html>.
- Linares Columbié, Radamés. 2005. *Ciencia de la información: su historia y epistemología*. Bogotá: Rojas Eberhard.
- RedHat. 2018. "¿Qué es Middleware?". Disponible en <https://www.redhat.com/es/topics/middleware/what-is-middleware>.
- NATO Science Committee. 1968. "Software Engineering: report on a conference sponsored". Peter Naur y Brian Randell. Bruselas, Bélgica. Disponible en <http://homepages.cs.ncl.ac.uk/brian.randell/NATO/nato1968.PDF>.

Standard ECMA-404. 2017. *The JSON Data Interchange Syntax*. 2ª edición. Página 3. [en línea] <https://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>

Enlazando datos: un modelo conceptual orientado a objetos para el diseño de bases bibliográficas

LIZBETH BERENICE HERRERA DELGADO

Universidad Nacional Autónoma de México

INTRODUCCIÓN

Para hablar de una forma diferente de estructuración de datos y su representación, es necesario conocer el modelo conceptual orientado a objetos, y el impacto que éste pudiera tener en los sistemas de almacenamiento de información que maneja la biblioteca inscrita en un entorno digital; después, es necesario identificar las partes que le conforman y que se involucran en la presente investigación. Asimismo, es importante considerar a la web, la biblioteca digital y la información o los datos, como los temas generales, ya que dada la naturaleza de cada uno de ellos, es necesaria la consideración de aspectos relacionados con su conformación, sin olvidar que las bases de datos son el agente en común.

También es importante identificar que el aspecto que vincula los tres temas generales (por lo que inevitablemente debemos referirnos a ellas) son las bases de datos y la forma en cómo, a través de ellas, se organizan los datos y se gestiona la vinculación de información, lo que a su vez está estrechamente relacionado con la estructura y los que hasta la fecha existen y se aplican a su creación, determinando

así su funcionamiento. Es innegable que el surgimiento de la web obliga a la biblioteca digital a plantearse y realizar cambios sustanciales y necesarios, de los cuales dependerá su visibilidad, presencia y permanencia en el ámbito digital, así como su inscripción en la propia web, algunos de los cuales corresponden a las bases de datos, de ahí la importancia de pensar en estrategias eficaces a través de las cuales se establezcan las formas más idóneas de vinculación y enlace de datos que permitan el acceso, la localización y la explotación de la información requerida por el usuario.

Por otro lado, se ha de considerar que en la actualidad la biblioteca digital académica, en el caso particular de México, cuenta con una estructura de organización para la información documental la cual se fundamenta principalmente en las *International Standard Bibliographic Description* (ISBD), *Anglo-American Cataloguing Rules* 2ed. (AACR2), *Library of Congress Subject Headings* (LCSH), *MACHine Readable Cataloging* (MARC) reflejándose en las directrices de estructura para la generación de bases de datos bibliográficas (BD); y el establecimiento de relaciones jerárquicas como la vinculación entre los datos. No obstante, esta forma de organización y de establecimiento de relaciones de información comienza a parecer limitado no sólo ante los ojos de los usuarios, sino ante los de los bibliotecólogos (Herrera 2014, 41).

Es entonces que, desde el ámbito bibliotecológico y el de los estudios de la información, surgen propuestas para una forma distinta de organización y estructuración de la información, tales como Dublin Core (DC), los Functional Requirements for Bibliographic Records (FRBR), los Resource Description and Access (RDA), la identificación y uso de los metadatos, las ontologías, las redes semánticas y los Sistemas de Bases de Datos Orientados a Objetos como elementos que se han

de aplicar para traducirse en alternativas diferentes, viables y prometedoras en la conformación de aplicaciones más avanzadas en la identificación de aquella información demandada por el usuario y resguardada por la biblioteca digital, en lugar de los sistemas integrales de automatización aplicados a la biblioteca tradicional cuyo origen data de 1970, actualmente rebasados.

En este sentido tenemos que los FRBR actualmente constituyen el conceptual de dónde se crean las RDA con un nuevo proceso de descripción documental, calificado por muchos como muy útil para explicar la información bibliográfica bajo una visión estructural postulada por la web y sus cibernautas.

En función de lo anterior y bajo una mirada general, el planteamiento de generar un modelo conceptual orientado a objetos para las bases de datos —mismo que se encuentra presente en la arquitectura programática de la web—, regido por los principios establecidos en las FRBR y la aplicación de los metadatos conjuntamente con las ontologías¹ como una manifestación de la semánticidad² que se busca alcanzar en la web. Estos son temas que se están analizando desde el ámbito bibliotecológico y que resultan

1 Ambos considerados como piezas claves en el éxito de identificación y enlazado de datos.

2 De acuerdo con Tusón (2003, 16), Charles F. Hockett en 1958; desarrolló una lista de quince rasgos que el lenguaje humano comparte con otros sistemas de comunicación animal. Uno de ellos corresponde al término de semánticidad y en su definición hace referencia a la dimensión de las señales lingüísticas o signos y su función en la transmisión de significados y su significancia dentro de un sistema de lenguaje. De lo anterior y para fines de esta investigación, se entiende la semánticidad como la forma de referir una relación, vínculo u asociación que se define entre un signo y el contenido que se desea que represente de forma constante y sistemática, que permita la evocación del concepto y que vaya más allá de la representación de un término especializado.

del cambio en las características que actualmente poseen la información y los soportes documentales; van de la mano con el surgimiento de aplicaciones, programas y sistemas de cómputo como herramientas tecnológicas que responden tanto a esta nueva realidad, como a las tendencias de desarrollo de la web semántica.

Se identifica la necesidad de migrar tanto del Modelo programático entidad-relación (ME-R), como de las anteriores normas de organización aplicadas por la bibliotecología sobre la generación de registros bibliográficos de las fuentes de información. Éstas son algunas de las razones por las cuales los sistemas integrales de automatización existentes están siendo incapaces de explotar las bondades que las FRBR y las RDA proyectan para los recursos de información digital.

Así pues, este trabajo está perfilado en seis apartados comenzando con el análisis de los cuatro factores de cambio en el desarrollo de un modelo conceptual en la arquitectura de bases de datos bibliográficas, establecer los pasos en el desarrollo de un modelo conceptual para las bases de datos, presentar las quince clases y los objetos³ que cada una de ellas reúne en el diseño del modelo propuesto y ejemplificando la encapsulación de datos, dando paso a la presentación del esquema inicial de los Modelos Orientados a Objetos (MOO) desarrollados y ejemplificando las relaciones entre los elementos de una obra partiendo de un registro bibliográfico parcializado de la obra de Oscar Wilde para dar paso a las conclusiones.

3 Representación física de una obra, se utiliza en casos de dar a conocer la materia de la obra.

FACTORES DE CAMBIO CONSIDERADOS EN LA PROPUESTA Y DESARROLLO
DE UN MODELO CONCEPTUAL PARA LA ARQUITECTURA DE
BASES DE DATOS BIBLIOGRÁFICAS

Para aplicar el modelo conceptual orientado a objetos en el diseño de bases de datos bibliográficas, es necesario tomar en consideración al menos cuatro factores principales y directamente involucrados en la creación de la propuesta:

El **primero** corresponde a la web, su lenguaje de programación orientado a objetos, arquitectura, el desarrollo de la Inteligencia Artificial y más especialmente los metadatos y las ontologías. La *world wide web* derivó de la evidente carencia de estructura y la capacidad de establecer una conexión a mayor escala por Internet, gracias a la propuesta de protocolos universales y su desarrollo a cargo de Tim Berners-Lee entre los años de 1989 y 1991. De los elementos enunciados, se ha de mencionar que las ontologías, conjuntamente con los metadatos, se traducen en herramientas de semantización, además de ser piezas clave para la identificación de datos y la vinculación de información, al tiempo que dentro del MOO se dota al concepto de atributos, los cuales permiten la generación de clases, dominios como agentes activos en el desarrollo de la Inteligencia Artificial y el buen funcionamiento de los sistemas. El binomio ontología-metadato resulta trascendental para poder materializar el enlazado de datos dando cabida al desarrollo de dicha inteligencia.

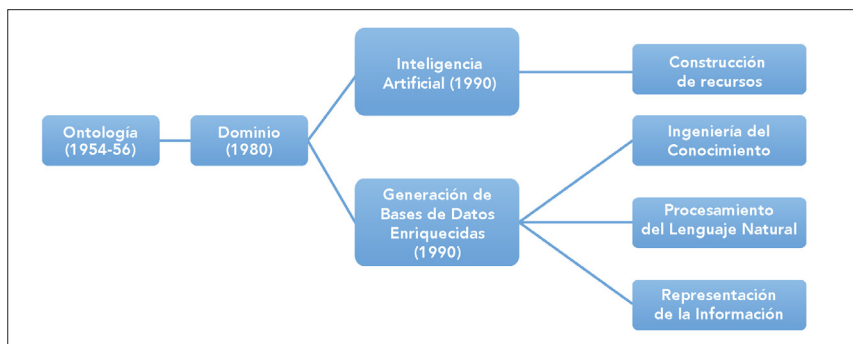
Es a partir de estos nuevos entornos digitales que las ontologías son vistas como una herramienta que da respuesta a la necesidad de descripción y ordenación en el manejo de información, tras haber identificado la pobreza existente en la indización o la señalización con la que se ha caracterizado

la web, aunado al resultado del balance costo-beneficio que se genera entre el tiempo empleado en las búsquedas y los tipos de resultados que estos arrojan.

A la fecha, las ontologías se presentan como la máxima expresión de los vocabularios controlados que se han ideado, aplicados a la forma de organización y descripción de la información y que conforman los principios de las taxonomías y los tesauros como sus antecesores. En este sentido, y al igual que los vocabularios controlados, se genera un proceso de asignación de conceptos⁴ seleccionados que describen documentos, objetos de contenido y materiales. Así, la aplicación de las ontologías es una forma de ayudar a las personas a explotar de manera más amigable, intuitiva y presumiblemente efectiva, de los recursos existentes en la web.

Con las ontologías, se busca ampliar el alcance de los

Figura 1. Conformación evolutiva de la ontología



Fuente: Elaboración propia.

⁴ Idea o noción abstracta que representa el tema de una obra; por ejemplo: campos de conocimiento, disciplinas, escuelas de pensamiento, teorías, procesos técnicos o prácticos, etcétera.

conceptos, concentrándose en mayores niveles de abstracción de los que se han establecido en las formas de ordenación previamente conocidas. Es decir, estas palabras son una representación simbólica del conocimiento (elevándose del nivel término al nivel concepto) que se expresa como un conjunto de axiomas y hechos, el cual puede jerarquizarse para establecer un sistema de clasificación temática con un lenguaje codificado y representar una organización que sea posible de enlazarse con la estructura y organización de las TIC, gracias a que permiten la obtención y recuperación de conocimientos a partir de modelos conceptuales jerárquicos o asociativos, compartidos para la agrupación de documentos, asignación, reutilización, generación de inferencias, compartición y formulación de criterios de ordenación.

La idea que rige el surgimiento, la creación y el aprovechamiento tanto de las ontologías como de los metadatos en los campos de la bibliotecología, los estudios de la información y la ingeniería de sistemas, responde a un nuevo intento de representación formal de los documentos, sus características y atributos propios mediante la aplicación de la semántica y los principios de lingüística dentro de la gestión de información en el ámbito digital, por lo que es necesario desarrollar una mejor predefinición formal, normalizada y universal de la información dentro de un contexto semántico específico.

El **segundo** factor de cambio corresponde a la biblioteca, sus modelos, estructura de organización interna, oferta de servicios, canales de comunicación y características de usuarios ocupándonos particularmente en los entornos digitales. En el caso de México, sabemos que existen múltiples realidades y es por ello que centramos la mirada en las bibliotecas académicas, ya que una buena cantidad de ellas se encuentran inscritas en el ámbito digital y el uso de las bases de datos que poseen, fueron desarrolladas bajo el

modelo programático E-R, incluidos los repositorios institucionales, como los desarrollos más recientes.

De igual forma, la biblioteca académica digital dispone de herramientas y recursos tecnológicos al alcance de los usuarios para la explotación y el aprovechamiento de sus servicios y recursos para consolidar su papel de agente facilitador en cuanto al acceso a la información respecta; en este sentido, podemos mencionar que además de contar con los sistemas integrales de automatización para bibliotecas y los catálogos en línea, implementa servicios basados en los sistemas de descubrimiento y cosechadores de URL para la recuperación en los recursos de información propios, suscritos y abiertos; se vincula con plataformas de gestión de archivos digitales; crea repositorios institucionales, entre otros, todo ello como estrategias de adaptación a los ambientes digitales muestra de un esfuerzo por mantenerse a la vanguardia.

Los recursos de información a los que se hace referencia se adquieren, registran, resguardan y ponen a disposición de la comunidad a través de los servicios de la biblioteca académica digital sin discriminación entre los formatos de registro (impresos o digitales), pues los objetivos de la biblioteca académica digital se enfocan en brindar acceso a la información observando aspectos como la amplitud, profundidad, la especialización y el alcance tanto de los contenidos como de los recursos de información que reúne, registra y oferta.

Asimismo, la biblioteca académica digital requiere brindar atención a un grupo de usuarios más heterogéneo en el cual se inscriben no sólo aquellos que se ubican dentro de su círculo inmediato de influencia (integrantes de su comunidad que acuden presencial o remotamente), sino también a los que se ubican en los círculos secundarios y subsecuentes.

En esta dinámica interactiva generada a través de los usuarios, la biblioteca académica digital funciona como el espacio propicio para la interacción entre usuarios, bibliotecólogos e información.

De este modo, la generación de servicios dentro y fuera de sus instalaciones operan en el amplio contexto global de la información, montados en el creciente ámbito de las TIC; ello se traduce a una realidad actual que resulta compleja y obliga a que la biblioteca académica digital se vuelva cada vez más dinámica, lo que consecuentemente conforma una serie de retos para continuar proveyendo sus servicios básicos, avanzados y especializados a su creciente y variado núcleo de usuarios, y se involucra en la investigación de los elementos tecnológicos emergentes, su aplicabilidad, viabilidad y utilidad, previo a la inclusión en su sistema de organización y operación, lo que registra una transformación en sus Modelos que de manera permanentemente transitan conformando un híbrido.⁵

El **tercer** factor de cambio corresponde a los modelos de descripción y los códigos empleados en el tratamiento de la información: organización, descripción e indización, los cuales fungirán como elementos de identificación de los recursos en las bases de datos.

En este sentido tenemos a los nuevos modelos teóricos de descripción o catalogación, gestados a últimas fechas conocidos como FRBR y FRAD y, los códigos RDA y FRAR —que se desprenden de dichos Modelos respectivamente—.

5 Un modelo de biblioteca híbrida pretende ofrecer un rango de tecnologías útiles aplicadas a las diferentes fuentes de información y a las formas de trabajo propias de la biblioteca tradicional mediante sistemas y servicios integrados localizados in situ pero asequibles en ambos ambientes (físicos y digital), lo que garantiza el acceso a todo tipo de recursos de información disponibles en su universo.

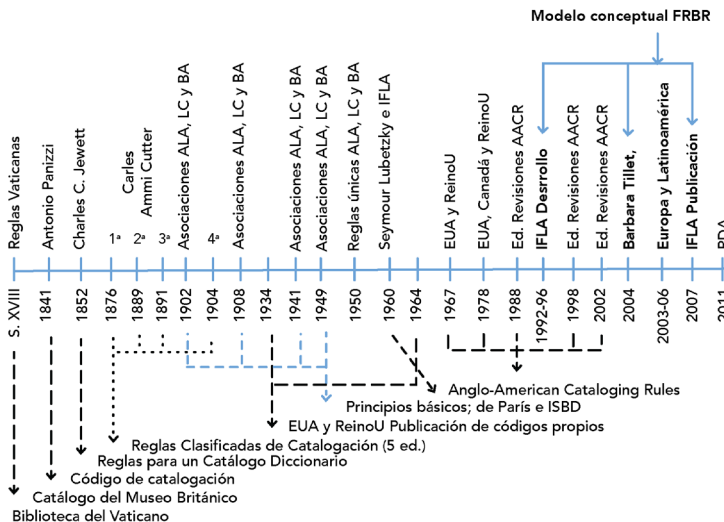
Estos códigos apoyan la generación de registros descriptivos de las fuentes de información y se verán enriquecidos al incluir elementos como los metadatos y las ontologías, que de manera conjunta se perciben como una nueva forma de estructuración del vocabulario controlado que funciona con cierto nivel de eficiencia en nuestro campo, particularmente en la descripción e identificación de la información y los documentos que la contienen.

Ambos elementos (los modelos y los códigos) son actualmente emergentes y aún se encuentran en una etapa de construcción, cohesión y maduración, con algunos ejemplos de aplicación en fase experimental dentro de bibliotecas como la Library of Congress (Estados Unidos), la Biblioteca Nacional de Colombia o la British Library, sin que éstos se reflejen en la descripción de los recursos de información existentes en las bases de datos comerciales (Herrera 2018, 41).

Los lineamientos y códigos de catalogación, en términos generales, son vistos como una forma de establecer un orden en el mundo del saber, las ideas, la información y las cosas que en él existen; el ser humano ha dedicado parte de su tiempo a desarrollar formas de estructuración basándose en múltiples y distintos criterios, siempre tratando de crear una fórmula infalible, eficiente y universal, de modo que éste es uno de los temas centrales en los que se ocupan algunos de los profesionales de nuestra área, quienes generan propuestas de lo que resulta el cambio y la variedad de herramientas ideadas, perfeccionadas o desechadas a través del tiempo, las cuales son presentadas a la comunidad científica que se encarga de aprobarla y, tras ello, se propicia el periodo de difusión dentro del gremio a todos los niveles para buscar su comprensión y aplicación en el ámbito laboral y la práctica profesional.

Un claro ejemplo de la evolución de reglas, normas y códigos desarrollados y empleados a lo largo de la historia se presenta en la figura 2, donde resaltan los años de 1992-2007, momento en que aparecen las FRBR como el modelo más actual y que dará paso al surgimiento de las RDA.

Figura 2. Evolución de las reglas y los códigos de catalogación



Fuente: Elaboración propia.

Se concibe a las FRBR como un marco conceptual terminológico y de datos bibliográficos que analiza el registro catalográfico como un conjunto de información sobre cuatro aspectos diferenciados del documento descrito,⁶ la base

6 Las entidades para la descripción bibliográfica son: obra (creación intelectual o artística distintiva, vista como una integralidad conceptual que se conjuga con el contenido intelectual y sus versiones lingüísticas, se le denomina también como objeto); expresión (refleja el contenido intelectual o artístico; su sinónimo es “versión”); manifestación (materialización de una obra en su(s) soporte(s), su sinónimo es “publicación”) e ítem (documento único de una manifestación que da a conocer la forma física de la obra).

para desarrollar catálogos innovadores, una actualización a los principios internacionales de catalogación y el motor generador de las RDA como código internacional de catalogación. Por ello los FRBR y las RDA serán elementos integrados en la presente propuesta del Modelo Conceptual Orientado a Objetos (MCOO) en el diseño de bases bibliográficas.

De forma sucinta, se ha de mencionar que los cambios que impactaron a la catalogación se deben, por un lado, a la proliferación de nuevos tipos, contenidos y formatos de publicaciones, y por el otro se encuentra la automatización de las operaciones y los servicios que desarrolla la biblioteca, aunado a la evolución y escalabilidad continua en el ámbito digital, lo que hizo patente la necesidad de modificar el tratamiento de los datos de descripción bibliográfica en cada fuente de información que la biblioteca como ente digital administra.

Elaine Svenonius también señala que estos cambios impactaron a los principios de control de autoridades o archivo de autoridad virtual e internacional sobre la forma y estructura de los encabezamientos corporativos, lo que obliga a dejar de lado la concepción tradicional del Control Bibliográfico Universal (CBU) y reconocer la necesidad que los usuarios manifiestan sobre una visualización más comprensible. Hay que subrayar que esta nueva versión de norma de contenido multinacional es para proveer descripción y acceso bibliográfico para todo tipo de medios, independientemente del formato usado para comunicar la información (Svenonius 2000, 6).

Por su, parte Matthew Beacom afirma que a pesar de que esta norma se desarrolla para ser usada en comunidades de habla inglesa, esto no limita su uso en comunidades que hablen otros idiomas, y recomienda que se prefiera su adop-

ción sin la intervención de una traducción (Joint Steering Committee for revision of AACR 2004, 1).

Karen Coyle, por otro lado, aporta sus reflexiones sobre el accionar de los bibliotecólogos y cuestiona que si existe una claridad respecto de la incompatibilidad entre FRBR y el formato MARC, ¿por qué se insiste en generar una conciliación? Tras lo que denota la necesidad de idear una nueva estructura de registro que apoye la experimentación y aplicación de los principios de las FRBR con una lógica distinta a la estructura que ofrece el formato MARC (Coyle y Hillmann 2007).

Por otra parte está Jeff Edmunds, quien en su disertación establece que en el desarrollo inicial de BIBFRAME se abandona una de las principales entidades establecidas por las FRBR *Manifestation* por dos razones: su comprensión resulta difícil y se le considerada inútil. El autor conserva solo las tres entidades principales restantes y centra su atención en *Work* (Edmunds 2017, 1). No obstante, *Work* para BIBFRAME se concibe de forma muy distinta que las RDA, lo que se traduce en una incongruencia entre ellos y ha generado que el Modelo de datos establecido por BIBFRAME, aunque funcional, presente muchas debilidades prácticas, de lo que se deduce que uno de sus problemas se relaciona con la intención de mantener una estructura programática de BD bajo el Modelo E-R, en vez de migrar al MOO.

Finalmente, se ha de comentar que en lo referente a los formatos de integración también se registra una evolución y un enriquecimiento gracias a la inclusión de los metadatos administrativos, que van en armonía con los formatos digitales de las fuentes de información y el ambiente digital.

El cuarto factor de cambio corresponde al estudio de la estructura programática de las bases de datos, su modelación, diseño y funcionamiento dentro de un sistema de

almacenamiento de información para la biblioteca, de donde además de tener en cuenta los diagramas de Modelos conceptuales E-R y OO, se tienen presentes los modelos de codificación bibliográfica desarrollados para la biblioteca; el más reciente es BIBFRAME en versión alfa.

En términos generales, una base de datos es un conjunto de datos homogéneos ordenados de una manera determinada que se presenta de forma legible por una computadora. No obstante, cuando se ve a la BD como diagrama, ésta se define como una descripción de los datos de interés. Sin embargo, en el caso de concebirse como un diagrama conceptual, se genera la representación de un modelo conceptual que tenga presente cualidades como expresividad, simplicidad, minimalismo y formalidad.

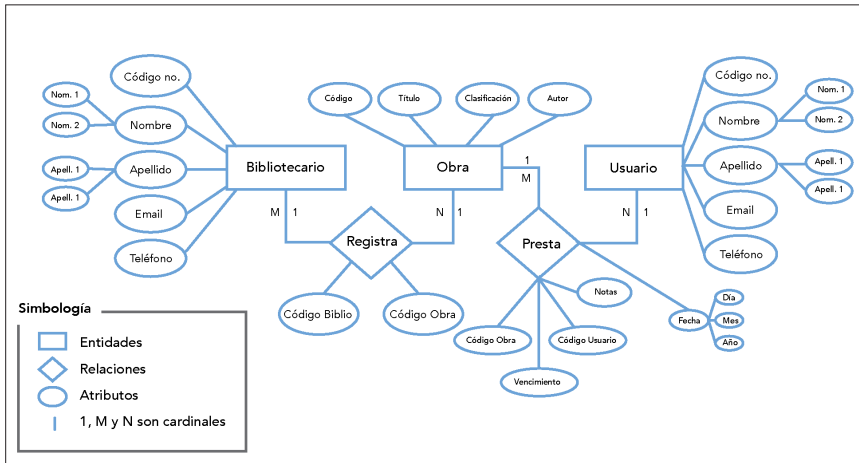
Es claro que de las concepciones presentadas que definen la base de datos (primero de forma estructural, después como diagrama y finalmente como diagrama conceptual de ellas) la última resulta la más enriquecida.

Ahora bien, en términos funcionales, una base de datos es concebida como una solución tecnológica para la gestión de información a diferentes niveles, ya sea individual u organizacional. Dicha solución tecnológica en la actualidad dota al hombre de mayores capacidades para el manejo de grandes masas de datos no estructuradas, con lo cual se abre la puerta a la integración de documentos que apoyen a los datos estructurados.

En las fases que dan origen a un Modelo de BD, se consideran por un lado los datos empleados por el sistema y, por otro, los procesos que realizan las tareas del sistema sobre esos mismos datos. De esta forma, se obtiene un Modelo completo de BD que incluye sus dos partes esenciales (datos y procesos) y describe claramente la arquitectura organizativa que contempla un punto de acceso a los datos.

Luego entonces, al ME-R se le concibe sólo como el diagrama con el cual se va a desarrollar la BD y para el desarrollo de análisis, la identificación de entidades fuertes y débiles diagramadas en Lenguaje Unificado de Modelado (UML por sus siglas en inglés), lo que posibilita distinguir las jerarquías de cada elemento⁷ y sus relaciones o dependencias, como se aprecia en la figura 3:

Figura 3. Diagrama estructural del Modelo entidad-relación



Fuente: Elaboración propia.

Actualmente, en los diagramas de sistemas modelados bajo E-R la ubicación de las bases de datos que conocemos y con las cuales interactuamos dentro de un sistema de información resulta ser el nivel inferior, definido como primera capa leída por la máquina, que interactúa con una segunda capa intermedia, en donde se ubican los procedi-

⁷ Los elementos básicos que componen el ME-R son entidad, relación (de participación, exclusividad, N:M, 1:N, 1:1) y atributos (superclave, calve candi-data, primaria, alternativa).

mientos (como objetos intermedios), que a su vez se comunican con la tercera capa superior en la que se localiza la interfaz a través de la cual los humanos visualizamos, solicitamos o administramos los datos.

Ahora bien, el MOO se basa en la existencia de los objetos y las clases de objetos. Los primeros son entidades físicas o sucesos que se describen en términos de sus datos y procesos. Es decir, se trata de una ocurrencia específica sin que sea el objeto físico mismo, sino más bien la representación conceptual del objeto que consiste básicamente en datos y procesos.

Por ejemplo: los objetos del mundo poseen dos características básicas que son comunes a todos ellos: *estado* y *comportamiento*. Un vehículo terrestre define su estado a partir de su velocidad actual, cantidad de cilindros, combustible, etcétera; su comportamiento se define por las siguientes acciones: encender, acelerar, frenar, cargar combustible, etcétera.

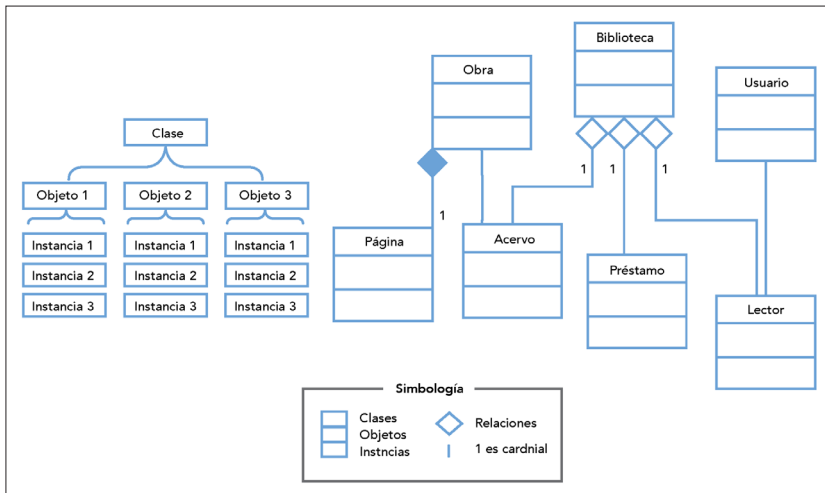
Los objetos de *software* se modelan de la misma forma que los objetos del mundo real, dado que también presentan un estado y un comportamiento. Un objeto de *software* mantiene su estado en una o más variables e implementa su comportamiento con métodos funcionales o procedimientos asociados con el objeto. El fundamento básico de la orienta-

8 Una imagen mental del encapsulamiento puede ser la de una caja cerrada con una etiqueta en el exterior en la cual se identifica el contenido del mismo. La idea del encapsulamiento y del ocultamiento de la información es fundamental para el análisis y el diseño OO, en ellos existe una “parte de información privada” y una “parte de información pública” dentro de cada clase encapsulada. Gracias al encapsulamiento, cada objeto se hace responsable de sus datos y no existe riesgo de que un algoritmo exterior a un objeto (por ejemplo, otro objeto) altere sus datos sin su consentimiento.

ción a objetos es el agrupamiento de datos y las acciones en un único paquete denominado objeto. Este concepto es el que provee el encapsulamiento y su cualidad más atractiva.

El MOO incorpora nuevas características a las BD y a los mecanismos de almacenamiento de objetos de longitudes indeterminadas para que sean capaces de encapsular⁸ métodos⁹ en su interior usando estos objetos como datos complejos o como funciones. Algunas de sus ventajas son la reusabilidad y la interoperabilidad, mientras que sus desventajas son el desarrollo más largo y el marcado contraste con los enfoques tradicionales de los conceptos y sus diferencias. Es importante señalar que este modelo y

Figura 4. Diagrama de Organización en el Modelo Orientado a Objetos



Fuente: Elaboración propia.

⁹ La implementación de comportamiento de un objeto en una base de datos se establece a través de métodos funcionales o procedimientos asociados con este; en otras palabras, es cumplir con las instrucciones que realiza un área determinada.

los elementos que lo componen han sido el más innovador y actual, gracias al refinamiento de las abstracciones clave y su identificación por nivel.

En términos generales, los diagramas de componentes presentan a los objetos, sus clases, atributos, relaciones, orden jerárquico y las interrelaciones o comunicaciones que deben establecerse para que sean funcionales dentro del sistema, tal como se muestra en la figura 4.

DESARROLLO DE UN MODELO CONCEPTUAL ORIENTADO A OBJETOS PARA BASES DE DATOS

Todo desarrollo implica cierta organización de las actividades y procesos que permitan alcanzar el objetivo deseado. En el caso que nos ocupa, se identifica un total de cuatro etapas y cada etapa se conforma por una o más fases, lo que suma un total de ocho en el siguiente orden:

Etapas	Fases requeridas en la creación de un programa o base de datos
1	1. Especificaciones de requisitos
2	2. Diseño conceptual 3. Diagrama conceptual
3	4. Diseño lógico 5. Diagrama lógico 6. Normalización
4	7. Diseño físico 8. Diagrama físico

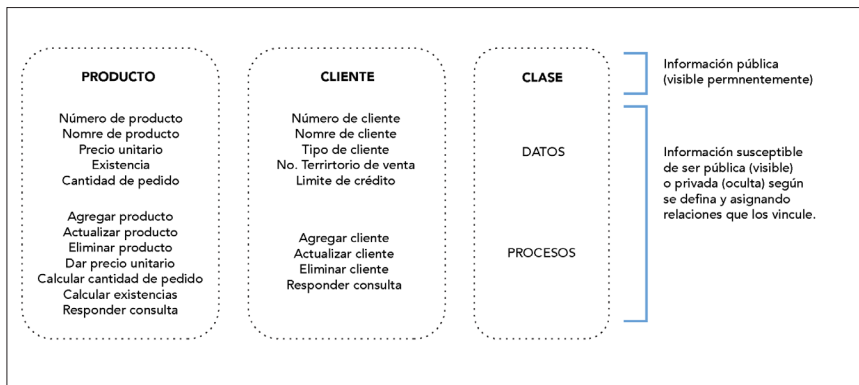
En función del objetivo fijado, es necesario cubrir las etapas uno y dos sin transitar las etapas restantes para obtener el diseño del Modelo conceptual orientado a objetos de base de datos.

Es gracias al análisis de los modelos conceptuales orientados a objetos de las bases de datos bibliográficas y en la web, que existe la identificación de los elementos funcionales y descriptivos que conformarán la base de datos y de todos los

factores que de manera directa intervienen, y que se cuenta con las *especificaciones de requisitos* que establecen las clases y sus variables, objetos y sus métodos, instancias y sus variables, encapsulamientos y finalmente refinar las abstracciones requeridas en la *fase de diseño conceptual*, lo que da paso al diseño de las clases como elementos que conformarán el diagrama del Modelo conceptual aprovechando las bondades del UML bajo la estructura que ilustra la figura 5.

Figura 5.

Diagrama de representación para una clase de Objeto en UML



Fuente: Elaboración propia.

Para identificar, nombrar y distribuir tanto las clases como los objetos de este Modelo conceptual orientado a objetos, se aprovecharon tres documentos: las tablas desarrolladas por el 6JSC (JSC 2015), la propuesta de la LC entorno al desarrollo de estándares (Library of Congress 2008) y el mapeo desarrollado por Joseph A. Kiegel, director de los servicios de catalogación y metadatos de la Biblioteca en la Universidad de Washington (Kiegel 2016), tras un cruce de datos y la conformación de una sola tabla que permitiera el registro

de los elementos existentes en bibliotecología concebidos en las FRBR y desarrollados en las RDA que reflejan esta lógica de organización de la información distinta a sus antecesores.

Para la distribución de los descriptores reunidos en las RDA, se establecieron ocho criterios de ordenamiento, a través de los cuales se establece el tratamiento correcto de los datos, la construcción de clases, objetos y elementos encapsulados, relaciones de comunicación, vinculación y dependencia jerárquica, así como la evaluación de las estructuras de clases y objetos; todo ello acorde tanto a la lógica de las bases de datos orientadas a objetos, como a lo establecido por el modelo FRBR y el código RDA.

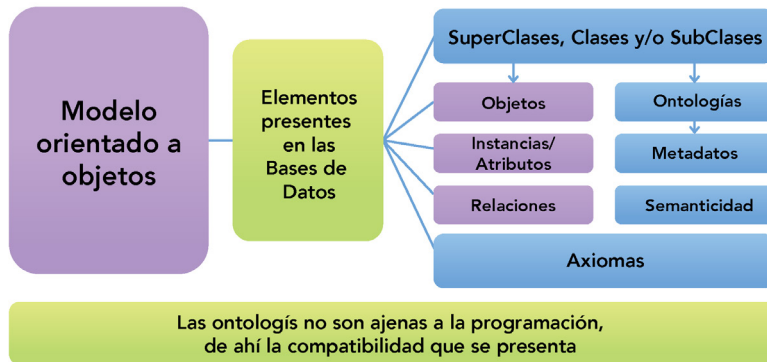
En la diagramación se debe contar con una jerarquía de clases basada en políticas y criterios a través de las cuales por un lado se erradique las redundancias, las clases irrelevantes o vagas, y por otro lado permitan identificar los atributos correspondientes, tomando en cuenta los mecanismos por los cuales los objetos colaboran para proporcionar algún comportamiento de nivel superior y expresan las decisiones estratégicas de diseño en concordancia con el funcionamiento correcto del programa.

Ahora bien, para enfatizar la incorporación de las ontologías en el modelo conceptual que nos ocupa, la figura 6 ilustra la ubicación e importancia que éstas guardan dentro de la orientación a objetos.

Este esquema presenta, por un lado la estructura de construcción de un objeto propio del MOO y, por otro, la arquitectura ontológica. Declara las similitudes existentes entre ambos mediante su ubicación ideal dentro de la propuesta de modelo, conformada por clases, subclases, instancias, relaciones, axiomas y reglas; elementos que denotan los diferentes niveles de abstracción tanto en el

aspecto semántico, como en la construcción programática. Estos elementos resultan indispensables dentro de la programación orientada a objetos y por ende son relevantes en una web semántica aún en desarrollo.

Figura 6. Elementos semánticos y programáticos en MCOO



Fuente: Elaboración propia.

DISEÑO ESTRUCTURAL DEL MODELO CONCEPTUAL ORIENTADO A OBJETOS

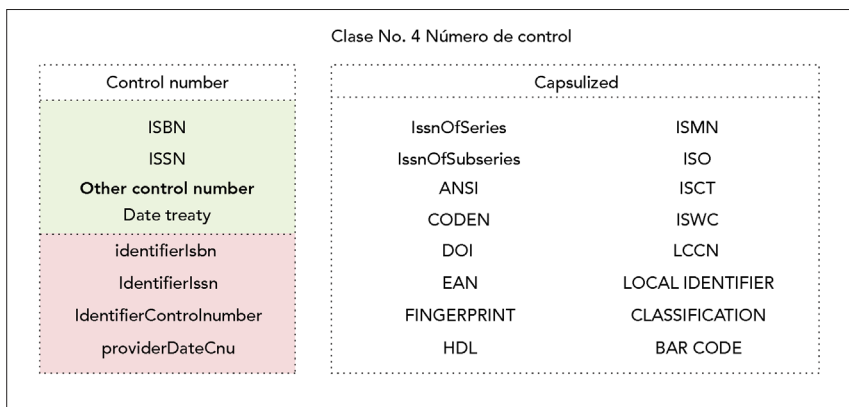
Así llegamos a la generación de objetos y la conformación de clases y objetos propias de la presente propuesta y el establecimiento de una jerarquía, objetos e instancias exportados o creados y observando siempre la sintaxis de SGML. El resultado obtenido tras estas actividades corresponde a la creación de quince clases de objetos —donde destacan las cuatro principales— en las cuales se reúnen todos los elementos considerados por las RDA, cuyas denominaciones en inglés¹⁰ son:

¹⁰ Son dos las razones por las cuales se consignan en inglés aquí: primero por ser el idioma origen del término y segundo por observar las políticas de práctica que se manifiestan en las RDA, donde se recomienda evitar la traducción.

- | | | | |
|-----------------------|-------------------------|---------------|-----------------|
| 1. Work | 5. Expresion | 9. Title | 13. Distributor |
| 2. Description | 6. Access | 10. Creator | 14. Performer |
| 3. Copyright | 7. Manifestation | 11. Publisher | 15. Subject |
| 4. Control number | 8. Manufacturer | 12. Editor | |

Y cuyo ejemplo de diagrama de clase, los objetos e instancias que la componen y aquellos elementos que fueron encapsulados.¹¹ Según el lenguaje UML y la sintaxis SGML tenemos lo siguiente.

Figura 7. Diagrama de representación clase control number



Fuente: Elaboración propia.

En este ejemplo, la clase 4, *control number*, reúne cuatro objetos con las correspondientes instancias siguiendo la definición que Mortier establece para la estructura de las clases de un MCOO, de donde:

- a) El primer apartado hace referencia a la clase, identificada por un **concepto específico** que

¹¹ Reiterando que el encapsulamiento es una de las características importantes del MOO.

se consideró como el más representativo para nombrar a la clase y que da cabida a los elementos del siguiente apartado.

- b) El segundo apartado enuncia y **reúne a los objetos**, ya sea de forma explícita o de forma implícita gracias al encapsulamiento.
- c) El tercer apartado hace referencia a la **instancia** de cada objeto previamente considerado como un paso que identifica al objeto en su etiqueta, lo que nos acerca a la fase de programación (Mortier 2005, 204).

De sus cuatro objetos, uno (*date treaty*) es resultado del cruce de datos de las tres fuentes base, mientras que los tres elementos restantes fueron creados. También tenemos los objetos que encapsulan a otros elementos bajo la lógica de correspondencia siguiente:

- a) ISSN **refiere y encapsula** a ISSNOfSeries para ampliar su alcance e incluir ISSNOfSubseries.
- b) OtherControlNumber **encapsula** todos los números de control susceptibles de ser considerados en diferentes tipos de documentos.

Finalmente, una forma sintética mediante la cual se puede ilustrar explicativamente su definición, los elementos encapsulados y las relaciones establecidas entre los objetos de cada clase se muestra en la tabla 1.

Hasta este punto, se han explicado e ilustrado las estrategias desarrolladas que enlazan los datos denominados como objetos e instancias dentro de las quince clases, lo que permite dirigir nuestra atención al diseño de un

Tabla 1. Definición de los elementos de la clase control number

Clase No. 4	CONTROL NUMBER
Nombre del objeto	Definición del objeto
ISBN	Este objeto reúne información sobre ISBN del recurso. Siendo su instancia y representación en lenguaje de marcado "identifierIsbn"
ISSN	Este objeto reúne información sobre ISSN del recurso. Siendo su instancia y representación en lenguaje de marcado "identifierIssn"
Other control number	Este objeto reúne información sobre los números de control del recurso, integrando variantes. Encapsula a Código regional, IssnOfSeries, IssnOfSubseries, ANSI, CODEN, DOI, EAN, FINGERPRINT, HDL, ISMN, ISO, ISCT, ISWC, LCCN, LOCAL IDENTIFIER CLASSIFICATION, BAR CODE, etc. Siendo su instancia y representación en lenguaje de marcado "IdentifierControlnumber"
Date treaty	Este objeto reúne información sobre fecha del número de control del recurso. Siendo su instancia y representación en lenguaje de marcado "providerDateCnu"

Fuente: Elaboración propia.

diagrama, en el que se presenten las vinculaciones de las quince clases.

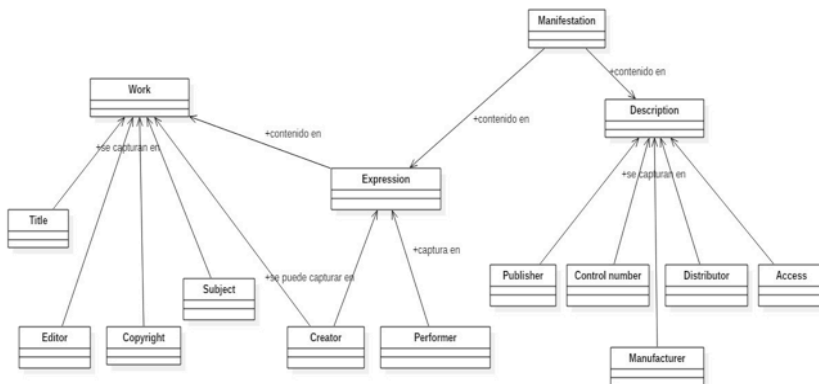
El diagrama del modelo "Flujo de datos" es uno de los cuatro diagramas generados. Este diagrama puede ser concebido como la representación visual de las clases conceptuales u objetos del mundo real en un dominio determinado, el gráfico pone de manifiesto las jerarquías y las relaciones que entre las quince clases han de existir, emulando la trayectoria del flujo de datos de la BD dentro de un sistema de información de la biblioteca, lo que en su momento deberá permitir un correcto funcionamiento; de ello se desprende lo siguiente:

- Contenido en: refiere a que las *tags* de los datos pueden ser localizadas en la pestaña de la clase a la cual se referencia y que por su naturaleza pueden ser consideradas como información adicional más detallada.
- Capturar en: permite determinar una pestaña o lugar específico para la presentación de las *tags* de los datos en cuestión.

- Se puede capturar en: hace referencia a *tags* de datos que se comparten entre clases.

La lógica aplicada en la jerarquización de clases nos permite determinar que work bien puede ser considerada como la clase principal o superclase de éste modelo, mientras que la lógica de funcionamiento y flujo de datos muestra la orientación de las relaciones y comunicaciones aplicables al sistema.

Figura 8. Diagrama del modelo conceptual flujo de datos



Fuente: Elaboración propia.

REGISTROS BIBLIOGRÁFICOS Y VINCULACIÓN DE DATOS

La forma más transparente de mostrar las diferencias en los datos enlazados que existen entre la propuesta y el registro bibliográfico actual se desglosa en las tablas que se presentan a continuación.

Tabla 2. Registro codificado bajo MARC 21 bibliográfico

Campo MARC21 Bibliográfico	Información
020=a	9780980968903
020=a	0980968909
100=a	Wilde, Oscar
100=d	1854-1900
100=e	Author
245=a	The important of being earnest
245=b	a trival comedy for serious people
245=c	By Oscar Wilde; witj wood-engraved ilustrations by Jarrett Morrison
264=a	North Vancouver, BC
264=b	The Bowler Press
264=c	MMVIII [2008]
264=c	c2008
300=a	100 pages
300=b	Ilustrations
300=c	24cm
336=a	text
336=a	still image
337=a	unmediated
338=a	volume
500=a	published... in an edition of sixty-five copies. The text was hand-set in Bembo roman with Bembo Titling, cast by Jim Rimmer and graciously provided by jan & Crispin Elsted of Barbarian Press. The stage directions were composed in Fairbank italic by The Letter-foundry of Michael & Winifred Bixler and manually inserted. The book was printed on Somerset Book Wove using an 8 by 12 Chandler and Price old Style jobber... Design, seting, ilustrstion, printing & binding were completed at the hands of Jarrett Morrison." --Colophon.
651=a	Englnd
700=a	Morrison, Jarrett
700=q	Jarrett Stephen
700=d	1973-
700=e	wood engraver
700=a	Rimmer, Jim
700=e	printer
710=a	Bowler Press
710=e	printer
710=a	Press and Letterfoundry of Michael & Einifred Bixler
710=e	compositor
710=a	Wookey Hole Mill
710=e	pepermarker

Fuente: Library of Congress Online Catalog (s.f.).
Disponible en <http://go.utlib.ca/cat/9758496>.

El ejemplo anterior pretende mostrar la forma como se han codificado y etiquetado cada uno de los datos que componen la descripción de una obra en formato MARC 21 bibliográfico.

Tabla 3. Registro descrito bajo los elementos de las RDA

Elemento RDA	Información
Identifier for Manifestation	9780980968903
Identifier for Manifestation	0980968909
Preferred Name for Person	Wilde, Oscar
Date Associated with Person	1854-1900
Relationship Designator	Author
Title Proper	The important of being earnest
Other title information	1854-1900
Statement of Responsibility relating to Title Proper	By Oscar Wilde; with wood-engraved illustrations by Jarrett Morrison
Place of Publication	North Vancouver, BC
Producer's Name	The Bowler Press
Date of Production	MMVIII [2008]
Copyright Date	c2008
Extent	100 pages
Illustrative content	Illustrations
Dimensions	24 cm
Content Type	text
Content Type	still image
Media Type	unmediated
Carrier Type	volume
Note on Publication Statement	published... in an edition of sixty-five copies. The text was hand-set in Bembo roman with Bembo Titling, cast by Jim Rimmer and graciously provided by jan & Crispin Elsted of Barbarian Press. The stage directions were composed in Fairbank italic by The Letter-foundry of Michael & Winifred Bixler and manually inserted. The book was printed on Somerset Book Wove using an 8 by 12 Chandler and Price old Style jobber... Design, seting, ilustrstion, printing & binding were completed at the hands of Jarrett Morrison." --Colophon.

Preferred Name for Person	England
Fuller Form of Name	Morrison, Jarrett
Date Associated with Person	Jarrett Stephen
Relationship	1973-
Designator	wood engraver
Preferred Name for Person	Rimmer, Jim
Relationship Designator	printer
Preferred Name for Corporate Body	Bowler Press
Relationship Designator	printer
Preferred Name for Corporated Body	Press and Letterfoundry of Michael & Einifred Bixler
Relationship Designator	compositor
Preferred Name for Corporated Body	Wookey Hole Mill
Relationship Designator	pepermarker

Fuente: Library of Congress Online Catalog (s.f.).

Disponible en: <http://go.utlib.ca/cat/9758496>

El ejemplo anterior pretende mostrar la forma como se describe cada uno de los datos que compone una obra usando las RDA.

Tabla 4. Registro modelado bajo las instancias del MCOO propuesto

Elemento RDA	Información
Identifierisbn	9780980968903
identifierisbn	0980968909
instanceCreator	Wilde, Oscar
instanceCreator	1854-1900
responsabilityStatementCr	Author
instanceTitle	The important of being earnest
instanceSubtitle	a trival comedy for serious people

hasAuthorityT	By Oscar Wilde; with wood-engraved illustrations by Jarrett Morrison
providerPlacepu	North Vancouver, BC
instancePublisher	The Bowler Press
providerDatepu	MMVIII [2008]
providerCopyrightdate	c2008
noteExtent	100 pages
asAnnotationM	Illustrations
dimensions	24 cm
contentCategoryW	text
contentCategoryW	still image
mediaCategory	unmediated
carrierCategory	volume
NoteD	published... in an edition of sixty-five copies. The text was hand-set in Bembo roman with Bembo Titling, cast by Jim Rimmer and graciously provided by jan & Crispin Elsted of Barbarian Press. The stage directions were composed in Fairbank italic by The Letter-foundry of Michael & Winifred Bixler and manually inserted. The book was printed on Somerset Book Wove using an 8 by 12 Chandler and Price old Style jobber... Design, seting, ilustrstion, printing & binding were completed at the hands of Jarrett Morrison." --Colophon.
providerTerm	England
instanceContributor	Morrison, Jarrett
asAnnotationlcr	Jarrett Stephen
instanceContributor	1973-
ResponsibilityStatementCr	wood engraver
instanceContributor	Rimmer, Jim
ResponsibilityStatementCr	printer
instanceContributor	Bowler Press
ResponsibilityStatementCr	printer
instanceContributor	Press and Letterfoundry of Michael & Einifred Bixler
ResponsibilityStatementCr	compositor
instanceContributor	Wookey Hole Mill
instanceContributor	pepermarker

Fuente: Library of Congress Online Catalog (s.f.).

Disponible en <http://go.utlib.ca/cat/9758496>.

El ejemplo anterior pretende poner de manifiesto la forma como se han nombrado y conjuntado las instancias, que se verán registradas en los metadatos descriptivos, para cada uno de los datos que componen la descripción de una obra en el MCOO descrito en la presente investigación.

CONCLUSIONES

La conjunción de los cuatro factores de cambio en los que se señala la injerencia de las TIC con sus desarrollos, la web con los metadatos, la semántica con las ontologías y la biblioteca académica digital con los códigos de descripción se conjugan para fundamentar el desarrollo de esta propuesta de modelo conceptual, en la cual se busca una nueva forma de enlazar los datos descriptivos de las fuentes de información en concordancia con la lógica de organización que establece el código RDA.

Estos factores de cambio son los responsables de gestar transformaciones importantes tanto en los profesionales de la información, como en los centros de información. Por esta razón, los elementos que les conforman se han convertido en objeto de estudio, mientras que en su visión global permiten generar escenarios posibles, de lo que se deriva el cuestionamiento al que se ven sometidas las estructuras de organización, descripción y vinculación de datos registrados en una base de datos bibliográfica ya no solo desde la perspectiva bibliotecaria, sino desde la ingeniería en sistemas.

En este sentido se considera ineludible la inclusión de elementos semánticos; es decir, ontologías como datos ubicados en la parte descriptiva como parte administrativa de la base de datos o sistema, consolidará la vinculación de datos tanto inherentes a la base de datos, como aquellos ubicados en la web permitiendo su recuperación y aprovechamiento semántico y su alto nivel de abstracción. La inclusión del Modelo orientado a objetos en el diseño de las bases de datos, como la forma a través de la cual se formula una organización distinta de enlazar datos y que presenta una

lógica distinta, resulta en la propuesta expresada acorde con el desarrollo que presenta la web.

En este sentido, la visión que se persigue con la web semántica es lograr la disposición de datos en la web, definidos y vinculados de tal forma que puedan ser enlazados por las máquinas y no solamente con el objetivo de ser visualizados, sino con la posibilidad de automatizar tareas, integrar datos y reutilizarlos con cualquier aplicación actual o emergente, creando servicios cada vez más inteligentes y adaptativos que aprovechen el uso de los agentes,¹² capaces de procesar y operar los recursos electrónicos de información a nivel semántico.

Estas razones son las que fundamentan la importancia que posee el Linked Open Data como estrategia de publicación estructurada incluido en la ecuación de la web; surge la aplicación de la metodología en favor de la reutilización de datos e información a través de las fases: sensibilización, análisis inicial de los datos, enriquecimiento semántico y, por último, exposición y reutilización; esto registra un cambio fundamental pues el dato puede ser leído, interpretado y procesado en forma automatizada apoyándose en el URI y sus variantes. El mejor ejemplo corresponde al sector público, gubernamental y el denominado *gobierno abierto*, al cual aún le hace falta el proceso de recogida de los factores existentes dentro de un contexto social determinado para completar el ciclo marcado por el Linked Open Data.

Con lo anterior, se pone de manifiesto que la generación de bases de datos con las características y bondades que posee la orientación a objetos es una forma de dar paso a importantes cambios como la aplicación de los agentes abiertos que escu-

12 Concebidos como entidades de software cuyas capacidades se abocan a los procesos de recolección, filtrado, procesamiento de información e inferencias de manera tanto semiautónoma como automática (Berners-Lee, Hendler y Lassila 2001).

driñan información pública en constante aumento. El cambio en estas especificaciones facilita el cumplimiento de tareas como la búsqueda, localización, identificación y recuperación en términos de mayor cobertura de la web respecto a la información; es decir, la generación de respuestas significativamente más pertinentes a la solicitud de un usuario.

En síntesis, si las razones que motivan la propuesta de un cambio no resultan suficientes, solo nos queda preguntarnos, ¿por qué continuamos haciendo lo mismos al tiempo que deseamos que las cosas cambien?; bien lo dijo Albert Einstein, “si buscas resultados distintos, no hagas siempre lo mismo”.

BIBLIOGRAFÍA

- Berners Lee, Tim, James Hendler y Ora Lassila. 2001. "The Semantic Web". *Scientific American*. Vol 285.
- Coyle, Karen y Diane Hillmann. 2007. "Resource Description and Access (RDA): Cataloging Rules for the 20th Century", *D-Lib Magazine* vol. 13, núm. ½. <https://doi.org/10.1045/january2007-coyle>.
- Edmunds, Jeff. 2017. "BIBFRAME as empty Vessel", Google Docs. Disponible en https://drive.google.com/file/d/0B1IKJYVwLwHyX1VnblJFZ3EtS1U/view?usp=embed_facebook.
- Herrera Delgado, Lizbeth Berenice, sustentante. 2018. *El modelo conceptual orientado a objetos y su alcance en los sistemas de almacenamiento de información de la biblioteca rumbo a la web semántica*.
- . 2014. "Modelación, web semántica y biblioteca". En A. A. Leiva Mederos & Y. Hidalgo Delgado (Presidencia). *Taller de Web semántica. Simposio llevado a cabo en el Congreso International de Información INFO 2014*, La Habana, Cuba. Disponible en <http://ceur-ws.org/Vol-1219/paper4.pdf>.
- Joint Steering Committee for revision of AACR. 2004. "Strategic plan for AACR". 4JSC/Chair/79/Rev/3. Disponible en <http://www.rda-jsc.org/archivedsite/docs/chair79rev3.pdf>.
- Library of Congress. 2008. "MARC to Dublin Core Crosswalk". Disponible en <http://www.loc.gov/marc/marc2dc.html>.
- Mortier, Gustavo Du. 2005. *Técnicas de programación : guía fundamental de desarrollo de software*. Manuales users 74. Buenos Aires: MP.

- Real Academia Española. 2001. *Diccionario de la lengua española*. Disponible en <http://lema.rae.es/drae2001/srv/search?id=HOaXZfVBvDXX2349DlRt>.
- Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Digital libraries and electronic publishing. Cambridge, Massachusetts: The MIT Press.
- Tusón Valls, Jesús. 2003. *Introducción al lenguaje*. Cataluña: Universitat Oberta de Catalunya.

Informe desde las trincheras: Transformando unidades de catalogación para incluir servicios de repositorio

LISA FURUBOTTEN
Texas A&M University

INTRODUCCIÓN

De acuerdo con la convocatoria para el Congreso Internacional sobre Metadatos 2018 que cita Oliphant, los catalogadores (entre los otros bibliotecólogos) “deberían de colocarse como los expertos y autoridades en la era de los datos.” Pero, ¿cuáles son los pasos pragmáticos de tal esfuerzo? ¿Cómo se puede transformar una típica unidad de catalogación tradicional compuesto de profesionales y técnicos en una biblioteca académica en Estados Unidos, a una que incorpora servicios de repositorio? Tales servicios incluyen la creación de metadatos non-MARC, gestión del sistema de repositorio, control de autoridades e ingestas masivas de colecciones digitales. ¿Cuáles son las tendencias que alientan o impiden el cambio y cómo podemos sobrepasar los retos que se presentan? ¿Cómo se puede animar a los catalogadores a tomar un papel fundamental en la preparación y gestión de las colecciones digitales? Aquí presentamos la experiencia reciente de Texas A&M.

En junio de 2017, el sistema bibliotecario de Texas A&M decidió trasladar la gestión del repositorio institucional¹ de la Unidad de Comunicaciones Académicas a la Unidad de Catalogación. Varias razones impulsaron tal decisión, principalmente porque se volvió evidente que el modelo existente para manejar las colecciones digitales no era adecuado para recibir y procesar el número creciente de colecciones que llegaban debido al esfuerzo exitoso de Comunicaciones Académicas en promover el repositorio en la comunidad universitaria.

La Unidad de Comunicaciones Académicas estaba aislada, y se usaba el modelo de asignar una sola persona para manejar todos los aspectos del sistema. Es decir, el bibliotecario encargado de DSpace, entre otras tareas, se ocupó de promover el repositorio, proporcionar capacitación sobre su uso a los 60mil miembros de la comunidad universitaria, responder a las preguntas sobre derechos de autor, lidiar con el Departamento de Sistemas relativo a cuestiones técnicas, gestionar las tablas de metadatos e índices, crear metadatos para colecciones, administrar ingestas masivas, atender el *digital help desk*, etcétera. No había una persona ayudante entrenada en las mismas tareas para auxiliar o sustituir si el encargado principal no estaba disponible.

En el modelo tradicional de las unidades de catalogación, hay personas que saben gestionar los diferentes sistemas que son utilizados, y todos comparten una misma comprensión de los estándares de metadatos y de contenido (generalmente nacionales o internacionales) utilizados. Este modelo requiere de inversión en capacitación estandarizada y puede ser menos eficiente en el sentido que se requiere alcanzar el consenso de un grupo (incluyendo los socios externos) para implementar cualquier cambio en los procedimientos,

1 OakTrust, un repositorio D-Space: <http://oaktrust.library.tamu.edu>.

políticas o normas. Sin embargo, el modelo ha permitido a los catalogadores procesar grandes cantidades de material en una forma consistente, así como poder intercambiar registros con socios externos. Así, las unidades de catalogación también pueden llevar a cabo proyectos de metadatos de manera eficiente debido al número de miembros con los que ya se dispone en las unidades (profesionales, técnicos y becarios). A continuación, se enumeran las habilidades que el personal ya posee.

- Conocimiento de metadatos descriptivos.
- Análisis temático.
- Experiencia construyendo índices.
- Experiencia construyendo filtros.
- Experiencia desarrollando entrenamiento.
- Experiencia redactando documentación.
- Creación de plantillas.
- Criterios para la presentación pública.
- Administración de catálogos en línea.
- Creación de vocabularios.
- Participación en comunidades de normas.
- Control de autoridades (desambiguación).
- Limpieza automatizada de datos .
- Migración de datos (*mapping/crosswalking*).
- Diseño de flujos de trabajo.

Nuestra tarea era pasar de la teoría a la práctica empírica e integrar la gestión de las colecciones digitales en la unidad de Catalogación, de forma que este trabajo fuera tan cotidiano como la catalogación tradicional y el manejo de nuestro catálogo Voyager.

Contrariamente a la visión de que el personal es reactivo al cambio, los catalogadores estaban muy entusiasmados.

Los verdaderos problemas eran los retos relacionados entre sí: el tiempo y la formación, además de las islas departamentales y redefinir la labor de los catalogadores.

RETO: TIEMPO

Los catalogadores en Texas A&M tienen años intentando participar con las colecciones digitales, pero no había sido posible por la falta de tiempo, además de otros factores. Por ejemplo, debido a limitaciones económicas en el ámbito nacional, los departamentos de catalogación se han reducido porque el personal que está por jubilarse no ha sido reemplazado. Esto ocasiona que el personal en activo intente cubrir o suplir las actividades que por ésta y otras causas están sin algún responsable de ellas. Otro factor es que aunque la cantidad de materiales impresos está disminuyendo, en contraste hay un número cada vez mayor de recursos electrónicos para ser catalogados. Ocurre, como en otras instituciones, que hay una gran acumulación de materiales de fondo antiguo y colecciones especiales (en Texas A&M el último conteo fue de 60mil piezas) que se hallan sin catalogar.

Otro factor es que el aprendizaje del uso de las nuevas tecnologías requiere bastante atención y tiempo. Ante este problema, el reto es idear cómo administrar el tiempo, de tal manera que sea posible facilitar que los catalogadores puedan entrenarse para reorientar sus habilidades para asumir la gestión del número creciente de colecciones digitales, pero también atender el tratamiento de los materiales tradicionales.

LEAN CONTINUOUS IMPROVEMENT JOURNEY²

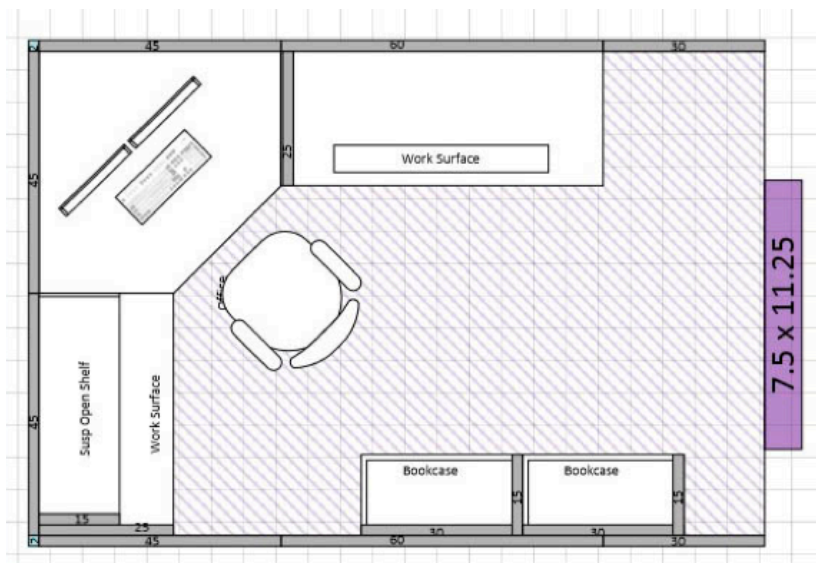
El personal del Departamento de Recursos Informáticos (Integrado por las áreas de Catalogación, Adquisiciones, Recursos Electrónicos y Control del Acervo) recibió entrenamiento intensivo de *Lean*, conocido también como “manufactura fina” o “producción sin desperdicios”. *Lean* es la versión creada por Toyota acerca de la metodología “estudio de tiempos y movimientos.” Involucra otros muchos aspectos laborales, además de la búsqueda de eficiencias en los flujos de trabajo; incluye técnicas para mejorar la cultura laboral, la formación de equipos de personal, promoción de innovación y la superación de las islas departamentales, etcétera. Es útil para poder analizar los procesos y flujos de trabajo y detectar embudos causados por libros en espera, transporte inadecuado de libros que van de un lado a otro, procedimientos manuales antiguos, errores innecesarios y otros factores que desaprovechan tiempo, pues era preciso lograr soluciones globales que involucraban todas las unidades en el departamento.

El Departamento de Recursos Informáticos ocupa el tercer piso del Edificio 3 de la Biblioteca Central. En él se hicieron las siguientes modificaciones:

- Los cubículos reducidos de tamaño fueron rediseñados para la mejor eficacia con paredes bajas para aumentar la visibilidad y la interacción entre los miembros de los equipos de trabajo (ilustración 1).

² <http://www.auburnworks.org/lean-continuous-improvement/>

Ilustración 1

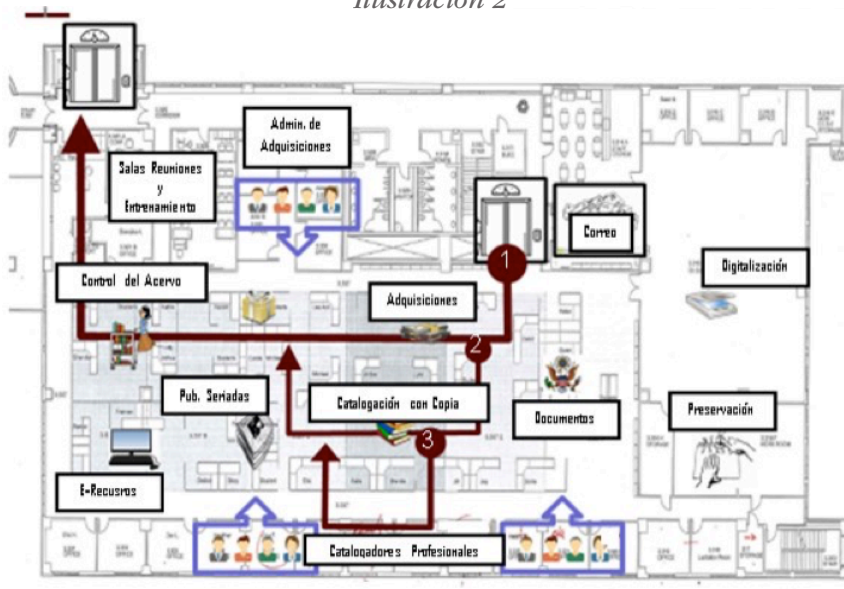


Fuente: Elaboración propia.

- También fue redimensionado el espacio ocupado por las unidades originales del Departamento. Se redujeron los cubículos y se prescindió de los cubículos en desuso por puestos vacantes. Dos unidades adicionales fueron incorporadas en el espacio liberado: Control del Acervo y Digitalización, que antes estaban aisladas en el cuarto piso.
- El plano de las instalaciones piso fue rediseñado: se reubicaron unidades afines entre sí con la intención de reducir el desplazamiento del personal de esas unidades. Por ejemplo, Recursos Electrónicos se trasladó a una esquina alejada para que no interfiriera el flujo de materiales físicos; por

otra parte, el personal de Publicaciones Seriadas se ubicó a su lado (de Recursos Electrónicos) porque esta unidad cataloga las publicaciones seriadas electrónicas; es decir, las dos unidades precisan colaborar con frecuencia.

Ilustración 2



Fuente: Elaboración propia.

- Derivado de la observación y el análisis, se conoció que las rutas recorridas por nuevos ítems durante su integración a la colección tenían un movimiento errático e ineficiente. Por este motivo, el nuevo plano de las instalaciones tiene la intención de establecer mejores flujos de trabajo. En la ilustración 2, se muestran las rutas mencionadas a continuación.

- 1) **Ruta 1:** los ítems precatalogados³ son trasladados del elevador de carga al cuarto del correo, luego a Adquisiciones para su recibimiento, pago y proceso físico. De allí, el personal de Control de Acervo integra los nuevos ítems en los acervos correctos.
- 2) **Ruta 2:** Si un ítem necesita catalogación con copia sencilla, pasa a los Procesos Técnicos del acervo con copia. En ese proceso se busca e importa un registro de OCLC que se estandariza para optimizar su uso local.
- 3) **Ruta 3:** Si no hay ningún registro disponible o hay un problema complejo (por ejemplo, el ítem está publicado en una lengua extranjera o hay encabezamientos erróneos), el ítem se envía a Procesos Técnicos, en dónde los bibliotecarios especializados en la materia habrán de crear los registros MARC originales o resuelvan alguna dificultad que se presente.

OTROS CAMBIOS

- Se contrató a una coordinadora de Catalogación. Desde su llegada, el personal de Procesos Técnicos ya le comunica a la nueva coordinadora sus actividades. Igual ocurre con el personal de la unidad de Adquisiciones. Esto tiene como finalidad facilitar que la unidad de Adquisiciones, cuyo personal está capacitado, trabaje los

3 Los ítems precatalogados son materiales con registros MARC aportados por el proveedor.

archivos MARC producidos por los proveedores, así como facilitar y agilizar la catalogación con copia en tanto se recibe el material y se aprueban los pagos. Los catalogadores ya no supervisan o realizan actividades de catalogación con copia, pero su consejo es valioso si ocurren dificultades en el procedimiento mencionado.

- Se ha aumentado el uso de registros MARC proporcionados por los proveedores, pues se han hecho modificaciones a los archivos modificaciones globales con MARCEdit antes de ingresarlos al catálogo. No se revisan ni se realizan modificaciones a registros individuales ya que la gerencia prefiere que el personal reciba los registros tal como llegan, es decir, sin normalizarlos o adecuarlos a las especificaciones locales.
- Dado que los registros de los proveedores se reciben por volumen y son aceptados tal como llegan por el personal en Adquisiciones, no es posible realizar ningún control de autoridades para encabezamientos de nombres o materias. Hemos contratado a la compañía Backstage⁴ para establecer el control de autoridades. Además, periódicamente se envían a Backstage los archivos de nuestros registros más recientes para que la compañía realice un análisis comparativo el Catálogo de Autoridades NACO para posteriormente corregir los encabezamientos en los registros bibliográficos y, si es el caso, Backstage envía cualquier registro de autoridad que no tengamos reconocido en el sistema local.

⁴ <http://www.bslw.com/cataloging/>

- Las bibliotecas con fondo antiguo, con colecciones especiales o de ciencias médicas contaban con unidades de catalogación propias. Éstas serán disueltas paulatinamente en la medida en que su trabajo se integre a la unidad de Catalogación Central. Además, la biblioteca de Ciencias Médicas tiene una instancia ajena del catálogo Voyager, que habrá de unificarse con el Catálogo Central. Con estas acciones, se espera lograr la eficiencia y el ahorro de costos, pues se empleará una sola instancia del catálogo y normalizarán las prácticas de catalogación que antes eran diferentes entre las mismas bibliotecas. Además de que se liberará al personal de las bibliotecas satélites para realizar otras funciones.
- Se ha hallado una nueva situación técnica para eximir a los catalogadores de algunas tareas técnicas. Esto incluye la coordinación con la unidad de Sistemas para la instalación de software que es utilizado en la unidad de Catalogación; el diseño, y la implementación de la capacitación a nuevos empleados para el manejo de Voyager, OCLC, Microsoft Office, etcétera; así también el desarrollo de soluciones automatizadas para tareas rutinarias del proceso de catalogación, procesos de depuración de datos, control de calidad, la gestión de estadísticas y otros informes sobre el catálogo.
- Derivado de la jubilación de algunos miembros del personal, se tenía solo un catalogador dedicado a actividades administrativas, cuatro bibliotecarios calificados y un catalogador también calificado, adscrito al fondo antiguo. Eviden-

temente, esto no fue suficiente, por lo que se procedió a la financiación para ampliar la unidad. Actualmente, existe un director que funge como Coordinadora de Catalogación, además de cinco catalogadores calificados en la unidad central, así como otros dos catalogadores en el fondo antiguo.

El flujo más lineal y eficaz de materiales, resultado de la modificación del plano de las instalaciones y la utilización por los técnicos de los registros de los proveedores, facilita que los catalogadores ahora se encarguen nada más de la catalogación original. Las otras iniciativas también fueron adoptadas con la intención de estandarizar el trabajo (en función de ahorrar tiempo), así como desligar a los catalogadores de quehaceres que otros pueden hacer. La meta principal fue reducir el tiempo dedicado a la catalogación tradicional para que los catalogadores pudieran integrar los servicios de repositorio a sus actividades cotidianas. Sin embargo, esto no implicaría el desatender el compromiso con el Programa de Catalogación Cooperativa (PCC)⁵ de la Biblioteca del Congreso, como miembros de NACO, CONSER y BIBCO.

Esta actividad, no obstante, lleva mucho tiempo debido a la inversión en capacitación para nuevos catalogadores y el estudio posterior requerido para mantenerse al día con respecto a los cambios en las políticas y los estándares PCC. Algo que sin duda se tiene presente es lo valioso de ser miembro del PCC. Para nosotros, PCC es la puerta de entrada a la participación nivel nacional con relación a las normas y políticas de catalogación. Debido a que muchos materiales locales ya serán procesados por técnicos de la unidad de Adquisiciones, los registros ya no serán exami-

⁵ <http://www.loc.gov/aba/pcc/>

nados por catalogadores profesionales para su conversión a las normas del PCC. Por lo tanto, nuestras futuras contribuciones de registros PCC tendrán que ser generadas a partir de materiales que requieran registros originales.

RETO: ENTRENAMIENTO

Aunque es difícil capacitarse en un periodo limitado, los catalogadores tenían años adquiriendo conocimientos teóricos sobre los esquemas de metadatos. Además, no tenían oportunidades de colaborar con las colecciones digitales, por lo tanto su experiencia derivada de la práctica era nula. Los técnicos no tenían experiencia y habilidad más allá de la práctica en la catalogación tradicional con MARC y AACR2/RDA. Era inobjetable que la biblioteca precisara de invertir tiempo y presupuesto para preparar el personal de catalogación para enfrentar las nuevas actividades con las colecciones digitales del repositorio.

¡OJO! BRECHA GENERACIONAL

Observamos que hay una brecha generacional entre los catalogadores experimentados que, por falta de tiempo, no estaban tan informados sobre las nuevas tecnologías como los catalogadores recién graduados. Texas A&M no había contratado nuevos catalogadores por más de una década, así que las diferencias en la formación de los recién graduados se acentuaron. Algunas habilidades deseadas fueron constantes: flexibilidad, buena comunicación, conocimiento de la economía de la biblioteca, habilidades de investigación,

atención a las necesidades y comportamiento del usuario, buena organización del tiempo, capacidad para la solución de problemas, buen criterio y adaptación a nuevas ideas.

Los catalogadores recién graduados tienen nuevas habilidades como el manejo de softwares de repositorio tales como Fedora, Islandora, CONTENTdm, etcétera, y conocimientos de nuevos formatos de metadatos y herramientas tales como XML, Turtle, RDF, XSLT, Python, JSON, editor del oxígeno XML, OpenRefine, XMLNotepad, etcétera. La realidad actual es que la catalogación tradicional con MARC ya no es una temática esencial en sus programas académicos como lo fue en antaño. Por lo tanto, tampoco es visto como algo necesario para avanzar en el conocimiento relativo a la organización documental. Por lo tanto, es evidente que cada grupo necesita orientar al otro en las habilidades de las que no se tiene conocimiento ni práctica.

Por ejemplo, sólo uno de los catalogadores sabía manejar el sistema de DSpace, y otro tenía un buen conocimiento teórico adquirido por su participación en los comités locales sobre el repositorio. Por lo tanto, fue indispensable entrenar a más personal, incluidos los técnicos, puesto que sin el conocimiento indispensable de saber cómo opera la plataforma es imposible diseñar o implementar proyectos de las colecciones digitales. Con la capacitación al personal, se busca que los técnicos familiarizados con DSpace o cualquier otra plataforma, de la misma manera que lo son con los sistemas de catálogo Voyager y OCLC.

Existen muchas herramientas de acceso abierto, de operación sencilla y por lo tanto fácil de usar que nos permiten crear, manejar y transformar metadatos sin la necesidad de aprender lenguajes de programación (es decir, Python, JSON, XSLT, SQL, etcétera). MARCEdit⁶, en particular, ofrece un

conjunto de herramientas para la migración y transformación de metadatos. Bajo este criterio, emprendimos una reunión semanal para la formación en MARCEdit, pero también observamos otras herramientas como Excel, Access, OpenRefine, MacroExpress, Regular Expressions y SharePoint. Esta acción nos llevó a descubrir que los técnicos son capaces de producir, migrar o transformar metadatos rápidamente con precisión exacta usando estas herramientas. Es decir, anteriormente estábamos, valga la expresión, “utilizando un cañón para matar moscas”. Se ha comprobado que no es recomendable y que es un derroche de recursos humanos emplear catalogadores profesionales o personal de sistemas para las tareas de transformar o migrar metadatos, esto salvo en la capacidad de los administradores de los proyectos.

Como se mencionó en líneas anteriores, se contrató a un asistente técnico para la catalogación, y una actividad adicional para este asistente es la búsqueda y evaluación de herramientas sencillas y amigables que permitan a los técnicos manejar metadatos para proyectos digitales, y el proceso de su ingreso al repositorio. Como ejemplo concreto, un profesor estaba agregando manualmente registros MARCXML a cada objeto de su colección digital en el repositorio; nuestro asistente técnico tardó solo dos horas en construir un macro que mecánicamente hace el mismo trabajo en lotes de trescientos registros. El profesor nos informó que esto le ahorró semanas de trabajo. No era necesario un *software* complicado o un profesional de Sistemas para automatizar esta tarea, sino un técnico entusiasmado y dotado de una herramienta sencilla.

Los catalogadores profesionales, y en cierta medida también los técnicos, fueron alentados y financiados para asistir eventos, conferencias y seminarios de capacita-

6 <https://marcedit.reeset.net/>

ción dedicados a las nuevas tecnologías, en adición de las reuniones tradicionales de catalogación. Nuestros catalogadores han participado en eventos relativos a DCMI Dublin Core, Open Repositories, Electronic Thesis and Dissertations, Linked Data, Joint Conference on Digital Libraries, Texas Conference on Digital Libraries, ASIS&T y eventos de difusión de Ruby on Rails, Python e ISNI. Además de las clases en línea de ALCTS⁷ y Library Juice Academy⁸, que son otras excelentes fuentes de educación a distancia.

RETO: ISLAS DEPARTAMENTALES

Encontramos dos barreras culturales para superar antes de trasladar la administración de DSpace a los catalogadores. La primera barrera fue la falta de comunicación entre los diferentes departamentos que tienen actuaciones afines en la gestión de colecciones digitales.

Las ubicaciones de las unidades de Catalogación en el tercer piso, de Comunicación Académica y Digitalización en el quinto piso y la unidad de Sistemas en el sexto piso. Eran causa de un deficiente trabajo de equipo interdisciplinario.

Ahora bien, no es propio el usar los estereotipos, pero hay algo de verdad, por ejemplo, en que los catalogadores no son propensos a la promoción de sus habilidades. Además de otro estereotipo que consiste en la resistencia a participar, y al cambio que implica dejar la catalogación tradicional con MARC. Por otra parte, los miembros de otros departamentos no sabían qué trabajo desempeñan los catalogadores o que podrían como aliados, aportar sus valiosas habilidades mencionadas anteriormente.

⁷ <http://www.ala.org/alcts/events>

⁸ <http://libraryjuiceacademy.com/>

No se tuvo la idea de consultar a los catalogadores sobre el flujo de trabajo de las colecciones digitales. Además de que los metadatos y las políticas para el repositorio fueron diseñados por individuos que no sabían al respecto. Y que tampoco sabían de los problemas con los que lidiaban con respecto a los metadatos del repositorio. Esas son cuestiones que los catalogadores vienen investigando durante décadas a través del manejo de los catálogos. Los catalogadores necesitaban aprender a prestigiarse por sí mismos como expertos en metadatos y control de autoridades.

Cabe señalar que las tareas inherentes a los metadatos deberían ser evidentemente asignadas a ellos, sin importar si la naturaleza de una colección es impresa o digital. A raíz de lo mencionado, los catalogadores empezaron a llevar a cabo reuniones internas para compartir a los demás sus conocimientos, y posteriormente participaron en conferencias como la Texas Conference on Digital Libraries.

*Ilustración 4. Scholarly
Communication, Preservation and Metadata (SPAM)t*



Fuente: Texas A&M Library

Para superar las islas departamentales e impulsar la comunicación, se creó la *Scholarly Communication, Preservation and Metadata* (SPAM), que consiste en un equipo interdisciplinario para coordinar las funciones distribuidas entre las diferentes unidades (ilustración 4).

SPAM evalúa nuevos proyectos, difunde informes de progreso y coordina la implementación de nuevos flujos de trabajo y herramientas. Por lo que, en vez de instituir un nuevo departamento dedicado a las colecciones digitales y el repositorio, se optó por un modelo descentralizado donde los departamentos ya existentes colaboran compartiendo conocimiento ya desarrollado por sus actividades cotidianas de procesar colecciones impresas, y llevan a cabo cada uno sus actividades asignadas. Se tiene la intención de perfeccionar este modelo descentralizado en razón de que convirtamos nuestro actual repositorio DSpace en un ecosistema de actividades digitales (Digital Asset Management Ecosystem) incorporando módulos adicionales de compilación y exhibición como Fedora y Spotlight. Las funciones están repartidas entre las unidades en la siguiente manera:

Unidad	Función
Catalogación y metadatos	Creación y mantenimiento de metadatos. Administración de DSpace. Creación y documentación de políticas del repositorio.
Preservación	Preservación digital. Administración de Archivemática. Escaneo/digitalización.
Comunicación académica	Promoción universitaria de acceso libre. Publicidad de colecciones digitales y el repositorio. Formación de usuarios. Asesoría sobre derechos del autor. Base de identidades VIVO.

¡OJO! TERMINOLOGÍA DISCIPLINARIA

Los miembros de las unidades de Comunicación Académica y Sistemas usan una terminología que es resultado de su formación en la informática. Con respecto al repositorio, muchos de los conceptos son los mismos utilizados por los informáticos al referirse acerca del catálogo, pero es difícil superar la barrera de una terminología que es confusa para un catalogador pues, en algunos casos, diferentes términos son utilizados por un catalogador y un informático para el mismo concepto, y hay otros casos donde el mismo término significa cosas diferentes para ambos o en otros casos no se entiende nada.

La mejor estrategia parece ser pedirle a quien se expresa que repita lo que dijo en diferentes términos, y en un acto de retroalimentación, expresar al orador lo que se entendió para asegurar que se está de acuerdo y así lograr una buena comunicación. Esto requiere tiempo y paciencia, pero con práctica los miembros de las diferentes unidades lograrán comunicarse más fácilmente.

RETO: REDEFINIR EL PAPEL DE LOS CATALOGADORES

La designación de la unidad de catalogación fue cambiada a Catalogación y Metadatos para aclarar la nueva labor de la unidad y señalar que la administración del sistema bibliotecario auspiciaba el traslado del repositorio a la unidad de Catalogación. Los títulos y las descripciones de los puestos de todos los catalogadores se cambiaron a *Cataloging and Metadata Librarian*, lo que autorizaba en forma oficial a los catalogadores a asumir el trabajo de metadatos y el manejo del repositorio, para asegurar que recibieron crédito y para

dar la oportunidad de participación a todos los catalogadores y no sólo unos pocos seleccionados por las otras unidades.

Anteriormente los catalogadores conceptuaban su actuar como especialistas y encargados de materiales asignados en varios formatos o idiomas. Las evaluaciones anuales de desempeño laboral se calificaban con base en los registros que un catalogador redactaba por año, contando además el número de registros hechos según las normas del PCC y el número de registros de autoridades contribuidos al programa NACO. Era necesario cambiar lo que se reconocía en las evaluaciones laborales para distinguir a los catalogadores y para que éstos fuesen partícipes en las nuevas tareas que demanda el repositorio. Así como fortalecer la meta de formar catalogadores con el perfil de administradores de proyectos para guiar y adiestrar a los técnicos y prestadores de servicio social, en vez de ser autómatas que solamente realizan cuotas de registros.

Se han modificado los criterios de evaluación laboral con la finalidad de que el número de registros producidos por un individuo deje de ser la base de su valoración y, más bien, se priorizarán las siguientes dimensiones:

- El número de colecciones procesadas o el número de proyectos exitosamente terminados, ya sea mediante el diseño de un nuevo flujo de trabajo para agilizar el trabajo o la administración satisfactoria de un equipo de trabajo compuesto de técnicos o prestadores de servicio social.
- Contribución de ideas o estrategias que optimiza la eficacia del flujo de trabajo, etcétera.
- Participación en un proyecto o comité tratándose de nuevas iniciativas como Open Access, Linked Data, BIBFrame, ISNI, FOLIO, etcétera.

- Adiestramiento para una o más actividades que auxilie a lograr el objetivo de transformar la unidad (es decir, XSLT, Python, SQL, etc., por medio de una clase de Library Juice Academy, etcétera).
- Participación en la gestión de colecciones digitales y DSpace.
- Evidencia del trabajo en equipo.
- Participación en el PCC produciendo registros de NACO, BIBCO o CONSER en OCLC.
- Contribución de registros a OCLC en sus formatos asignados. Si bien el criterio principal ya no involucra cuotas, tampoco un catalogador puede abandonar por completo la producción de registros.
- Demostración de habilidades de liderazgo. Como es un objetivo desarrollar habilidades de liderazgo en los catalogadores, serán reconocidos aquellos que coordinen una nueva temática, proyecto o comité que sirva como un recurso para los otros miembros del equipo. La biblioteca ha financiado entrenamiento de la metodología de *Project Management*.

CONCLUSIONES

La unidad de Catalogación y Metadatos se hizo cargo de DSpace en junio de 2017. Actualmente (agosto de 2018) hay casi veinte proyectos digitales en proceso y otros veinte en espera. Esto demuestra que no era viable el modelo anterior con un solo administrador para todas las funciones del repositorio porque es complicado que una sola persona pueda administrar tantos proyectos, por lo que ahora la unidad de

Catalogación ya tiene el personal con la experiencia necesaria de organización para procesar grandes cantidades de materiales. Es decir, todos los catalogadores profesionales y tres técnicos participan en la gestión de colecciones digitales.

Un proyecto ejemplar muestra la importancia de incorporar catalogadores en los proyectos digitales. Texas A&M aporta ítems digitales a la biblioteca digital de HathiTrust,⁹ que requiere que cada ítem tenga un registro MARC en OCLC y un archivo de esos mismos registros en formato MARCXML.

Simultáneamente, se añaden los ítems a la colección digital local de OakTrust, que requiere la descripción expresada en metadatos de Dublin Core. En 2016, un solo bibliotecario en Comunicaciones Académicas pudo contribuir trescientos 300 ítems a ambas colecciones. La cantidad de tiempo para crear sólo los metadatos de Dublin Core fue de poco más de veinte minutos por cada ítem.

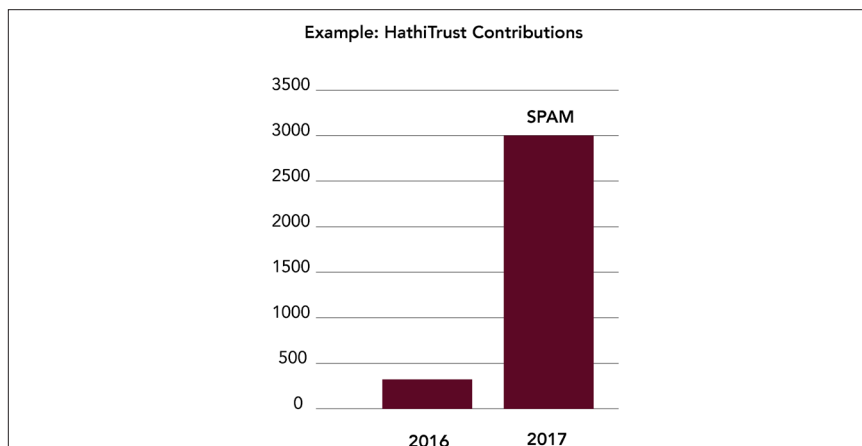
En 2017, dos catalogadores profesionales y dos técnicos se integraron al proyecto local de HathiTrust. Usando *katas* de *Lean*¹⁰ se desarrolló un procedimiento donde actualmente se utilizan entre tres y ocho minutos para crear un registro original de MARC en OCLC. Esos registros se recopilan en un archivo y utilizando MARCEdit se migran mecánicamente a MARCXML.

Según las especificaciones de HathiTrust y a la vez a Dublin Core en un archivo CSV requerido para importar los metadatos a Archivematica u OakTrust, en seis meses el equipo de catalogación creó registros en MARC, MARCXML, y Dublin Core para más de tres mil ítems, mientras seguían al tanto de sus otros quehaceres (gráfica1).

⁹ <https://www.hathitrust.org/>, una asociación de instituciones académicas y de investigación que ofrece una colección de millones de títulos digitalizados de bibliotecas de todo el mundo.

Gráfica 1.

Resultados de la producción con base en el proyecto HathiTrust



Fuente: Elaboración propia.

Además de continuar con sus deberes acostumbrados, durante el último año todos los catalogadores han asumido responsabilidades nuevas. Véase la siguiente tabla comparativa:

Catalogador	Tradicional	Nueva
1	Administración. Publicaciones seriadas. Coordinación CONSER. Coordinación NACO Series.	FOLIO - Chair. Entrenar técnicos para proyectos digitales. Migración de datos
2	Publicaciones seriadas. ECIP Almacenamiento Remoto.	Proyectos digitales. Supervisar asistentes estudiantiles.
3	Publicaciones seriadas.	Comité del Repositorio - Chair. Proyectos digitales. Digital Help Desk. Control de autoridades colecciones digitales.

10 La herramienta *kata* consiste en sistematizar la búsqueda de las condiciones de trabajo deseadas utilizando las capacidades humanas de una determinada manera. De este modo, se aprovecha el recurso más importante dentro de la filosofía *Lean*, mencionada en el apartado 1.

4	Monografías. Lengua china. BIBCO. ECIP.	Proyectos digitales.
5	Fondo antiguo. Coordinación BIBCO.	ISNI. Proyectos digitales. Entrenar técnicos.
6	Video. ECIP. Lenguas/Literatura. Coordinación NACO.	Casalini datos enlazados. Comité acceso libre. ISNI - Chair.
7	Video. E-Monografías. ECIP.	Comité del repositorio. Proyectos digitales. Backstage - Chair. ISNI.

Fuente: Elaboración propia.

Es notable que los técnicos han participado con entusiasmo en los proyectos digitales, a la vez que continúan desempeñando sus labores de procesar materiales impresos. Ellos atienden sus nuevas funciones de crear registros MARC necesarios para proyectos y migrarlos a MARCXML y Dublin Core.

También participan en la generación de resúmenes para tesis y disertaciones electrónicas y en preparar metadatos para realizar la exportación a DSpace. Sin duda, ellos están muy satisfechos con sus nuevas habilidades y vigilan cuidadosamente el control de calidad de los datos antes de mandarlos a Sistemas, con un sentido de responsabilidad de ser miembros del equipo de Catalogación.

Para aquellos que preguntan cómo es posible persuadir los catalogadores (profesionales y técnicos), a aceptar cambios tan formidables, la respuesta es que los catalogadores siempre se han adaptado al cambio, les encantan los retos intelectuales y lo aceptan por interés propio y no por otra razón.

Observando el mercado laboral, es evidente la escasez de vacantes para los catalogadores tradicionales, y las pocas

oportunidades disponibles dan prioridad y ofrecen una mejor remuneración a aquellos individuos que se hayan actualizado y que por ende tienen habilidades técnicas y experiencia con colecciones digitales.

Sin embargo, en el caso de los catalogadores experimentados y cercanos a su jubilación, pueden tomar la decisión pragmática de que ya no es conveniente una inversión fuerte y sobre todo difícil en actualizar sus habilidades. Lo cual no es un problema porque hay suficiente material impreso para emplear a estos catalogadores. Lo que resulta problemático es que los recién graduados muestran una resistencia con respecto a la catalogación tradicional.

Saben que las actividades de la gestión de colecciones digitales dan mayor soporte a sus currículos y a sus trayectorias profesionales. Si la visión es transformar y hacer progresar a todo un departamento, es necesario ser firme e insistir en que todos lleven su parte de trabajo tradicional. De otra forma, puede arriesgarse a crear en la unidad de catalogación un sistema de castas donde los recién graduados abarcan todo el trabajo con las colecciones digitales, dejando los catalogadores experimentados solo lo tradicional y sin oportunidades para crecer y adquirir nuevos conocimientos.

Concluyo con la siguiente anécdota curiosa: Un director de procesos técnicos que visitó Texas A&M expresó su frustración afirmando que era difícil conseguir que los catalogadores acepten el cambio porque “son como los gatos encerrados toda su vida en casa. Y que aun cuando se abre la puerta, no quieren salir a la calle debido a la indiferencia o al miedo”. Los catalogadores nos moríamos de la risa, y nuestra respuesta fue que, por el contrario, los catalogadores somos como gatos frustrados rascando la puerta aullando para salir. Es decir, los catalogadores tienen

muchas contribuciones valiosas que aportar, y esperan entusiasmados las oportunidades para colaborar con los procedimientos propios de las nuevas colecciones digitales y el repositorio institucional.

BIBLIOGRAFÍA

- American Library Association. 2017. *Cataloging competencies Task Force. Core Competencies for Cataloging and Metadata Librarians*. Disponible en <https://alair.ala.org/handle/11213/7853>.
- Glasser, Sally. 2007. "The Changing Face of Cataloging Positions at Academic Institutions", *The Serials Librarian*, vol. 51, núm. 3-4: 39-49. DOI: 10.1300/J123v51n03_04.
- Hill, Debra W. 1997. "Requisite Skills of the Entry-Level Cataloger: A Supervisor's Perspective", *Cataloging & Classification Quarterly*, núm. 23: 3-4, 75-83.
- Leffler, Jennifer y Pamela Newberg. 2010. "Re-Visioning Technical Services: A Unique Opportunity to Examine the Past, Access the Present, and Create a Better Future", *Cataloging & Classification Quarterly*, núm. 48:6-7, 561-571. DOI: 10.1080/01639374.2010.496306
- Liss, Jennifer A. 2016. "Core Competencies for Professional Catalogers. A Report of the Competencies and Education for a Career in Cataloging Interest Group Meeting. American Library Association Annual Conference", *Technical Services Quarterly* vol. 34: 1, 81-84. DOI: doi.org/10.1080/07317131.2017.1238207
- McGeachin, Robert. 2018. "Time Costs to Digitize Print Agricultural Serial Publications", *Journal of Agricultural & Food Information*, vol. 19: 2, 116-120. DOI: 10.1080/10496505.2018.1436443
- Mooney Gonzales, B. 2014. "Preparing LIS students for a career in metadata librarianship", *SLIS Student Research Journal*, vol. 4, núm. 1. Disponible en <http://scholarworks.sjsu.edu/slissrj/vol4/iss1/3>

- Oliphant, Tami. 2017. A case for critical data studies in Library and Information Studies. *Journal of Critical Library and Information Studies*. Vol. no. 1. DOI: 10.24242/jclis.1i1.22
- Valentino, Maura L. 2010. "Integrating Metadata Creation into Catalog Workflow", *Cataloging & Classification Quarterly*, vol. 48: 6-7, 541-550. DOI: 10.1080/01639374.2010.496304

2. DATOS CIENTÍFICOS

Factores determinantes para la implementación del esquema de metadatos para repositorios de datos de investigación de la Política de Ciencia Abierta en México

MIGUEL ADOLFO GUAJARDO MENDOZA
CONACYT

INTRODUCCIÓN

De acuerdo al Capítulo X, “Del Acceso Abierto, Acceso a la Información Científica, Tecnológica y de Innovación y del Repositorio Nacional” de la Ley de Ciencia y Tecnología (Capítulo adicionado DOF 20-05-2014), el Consejo Nacional de Ciencia y Tecnología (CONACYT) es el órgano instruido para el diseño e impulso de la estrategia nacional para la democratización de la información científica, tecnológica y de innovación nacional e internacional a texto completo, en formatos digitales, a través de repositorios construidos por las instituciones de educación superior y centros de investigación.

La regulación de esta estrategia parte de los lineamientos emitidos por el CONACYT, conformados actualmente por los Lineamientos Generales de Ciencia Abierta (LGCA), los Lineamientos Jurídicos de Ciencia Abierta (LJCA) y los Lineamientos

Específicos para Repositorios (LER), que sumaron los programas de Revistas, Conricyt, SIICYT, Comunicación Pública de la Ciencia y Conectividad al programa de Repositorios para la conformación de una Política de Ciencia Abierta en México.¹

Los lineamientos específicos para repositorios tienen por objeto coordinar las políticas, los recursos, los programas y las acciones realizadas por el CONACYT relacionadas con los repositorios Nacional e institucionales de ciencia abierta. Están conformados por veintitrés lineamientos específicos, tres transitorios y tres apéndices con la descripción de la estructura de los metadatos de literatura y datos para la interoperabilidad de los repositorios Nacional e institucionales.

La conformación de esta estructura se encuentra basada en el esquema *OpenAIRE* para el manejo de repositorios de literatura 3.0 y para el manejo de repositorios de datos 2.0, como se indica en los Apéndices 1 y 2, así como en los Lineamientos Décimo Quinto al Décimo Séptimo de los Lineamientos Específicos para Repositorios.

El crecimiento exponencial en la generación de los datos y su proliferación ha generado nuevas perspectivas para el uso, análisis e interpretación de la información que contienen, lo que ha aumentado la atención en el papel que tienen los metadatos en la organización de la información en herramientas como los repositorios de datos, lo que permite su localización y recuperación (Oliphant 2017).

Es necesario diferenciar que cuando se habla de datos, se pueden referir a diversos tipos de datos, como los administrativos, de gestión, los financieros o los de investigación.

¹ Repositorio Nacional. Política de Ciencia Abierta. Disponible en <https://repositorionacionalcti.mx/documentos>.

Para el caso de la Política de Ciencia Abierta en México y los repositorios que la integran, esta última tipología de los datos es la de particular interés, pues en su objetivo está la máxima diseminación del conocimiento científico, tecnológico y de innovación, resultado de las investigaciones en nuestro país.

La política de la biblioteca de la Universidad de Melbourne (The University of Melbourne s.f.) describe los datos de investigación como hechos, observaciones o experiencias sobre un argumento, teoría o prueba. Estos pueden ser numéricos, descriptivos o visuales y pueden presentarse de manera cruda, analizada, experimental u observacional.

Para Greenberg, los datos son “la esencia de la ciencia”, los portadores de la verdad científica cuando se revisan los hallazgos (Greenberg *et al.* 2009). Ella también retoma el comentario de Davis y Vickery, quienes mencionan que “los conjuntos de datos pueden abarcar desde los sistemas de información geográficos o geoespaciales (GIS), hasta datos genómicos o cualquier conjunto de datos que respalde a una publicación académica, como los datos de un censo”.

Desde el nacimiento de la estrategia de Acceso Abierto y su posterior evolución a la Política de Ciencia Abierta en México, se les ha denominado datos primarios de las investigaciones a aquellos conjuntos de información recolectada y utilizada para la investigación académica, científica y de innovación, y ha culminado con la diferenciación entre los repositorios de datos y los de literatura, con la implementación de un esquema de metadatos particular para cada tipo, con la intención de describir de la mejor forma la información de los recursos que exponen y son cosechados por el Repositorio Nacional, los cuales son descritos en los lineamientos mencionados anteriormente.

DE LA ESTRATEGIA DE ACCESO ABIERTO
A LA POLÍTICA DE CIENCIA ABIERTA EN MÉXICO

El Estado es el principal financiador de la investigación, que a través de los recursos públicos apoya el desarrollo de los países. En el caso de México, el gobierno federal a través del Conacyt distribuye dichos recursos públicos para la producción científica. Sin embargo, los productos de la investigación generalmente son publicados en editoriales con alto prestigio académico, pero con un elevado costo de suscripción. Esta situación en la que las investigaciones financiadas con recursos públicos y que para acceder a la misma se requiere invertir más recursos afecta el ciclo de producción científica no sólo en México, sino en el mundo.

Desde 2002, se han apoyado iniciativas a nivel global en beneficio del acceso a la información científica, entre las que se destaca la Declaración de Bethesda (2002), la Declaración de Berlín (2003) y la Iniciativa de Budapest (2003), desde las que nace el movimiento internacional denominado “Acceso Abierto” (*Open Access*), que es definido como

la disponibilidad gratuita en Internet público, para que cualquier usuario la pueda leer, descargar, copiar, distribuir, imprimir, buscar o usarla con cualquier propósito legal, sin alguna barrera financiera, legal o técnica, fuera de las que son inseparables de las que implica acceder al Internet mismo (Iniciativa de Budapest 2003).

Considerando esto, el 20 de mayo de 2014 se publicó en el *Diario Oficial de la Federación* el Decreto por el cual se modifica la Ley de Ciencia y Tecnología, en el cual se incluye el Capítulo X titulado Del Acceso Abierto, Acceso a la información Científica, Tecnológica y de Innovación

y del Repositorio Nacional. Este decreto establece que el CONACYT es el órgano encargado de la creación y operación del Repositorio Nacional, al cual define como “[...] la plataforma digital centralizada que, siguiendo estándares internacionales, almacena, mantiene y preserva la información científica, tecnológica y de innovación, la cual deriva de las investigaciones, productos educativos y académicos” (CONACYT Informe general...2017a).

Desde ese momento, el CONACYT adquiere la obligación de presentar los lineamientos y las disposiciones correspondientes para el funcionamiento de los Repositorios Nacional e Institucionales, así como para capacitar, convocar, organizar y coordinar a las instituciones e instancias en materia de acceso abierto y el funcionamiento de dichas plataformas a través de lo que se denominó la Estrategia de Acceso Abierto a la Información Científica, Tecnológica y de Innovación. Esto llevó a la publicación de los Lineamientos Generales para el Repositorio Nacional y los Repositorios Institucionales en noviembre de 2014 y los Lineamientos Técnicos para el Repositorio Nacional y los Repositorios Institucionales en noviembre del siguiente año.

Durante los siguientes dos años, se identificó una nueva tendencia a nivel internacional que se orientaba a hacer más transparente y colaborativo el proceso de generación del conocimiento científico (principalmente financiado con recursos públicos) denominado Ciencia Abierta (*Open Science*), paradigma que busca que cualquier interesado pueda acceder libre y gratuitamente a los materiales y recursos de información que resultan del proceso de investigación, en cualquiera de sus etapas, con la posibilidad de usarlos, reusarlos, modificarlos, compartirlos y difundirlos mediante la utilización de las Tecnologías de la Información y Comunicación (TIC).

En sincronía con este paradigma, el CONACYT revisó los lineamientos vigentes y la viabilidad de transición; dentro de su espectro de acción incluiría además del Repositorio Nacional y los repositorios institucionales a los programas de Comunicación Pública de la Ciencia, Revistas, Consorcio para la Adquisición y Disseminación de la información, Conectividad y el Sistema Integrado de Información Científica y Tecnológica. Esto permitiría diluir las barreras para compartir cualquier producto, recurso o herramienta generada en cualquiera de las etapas del proceso de investigación, desde la generación de datos, la publicación de los resultados en documentos arbitrados, hasta la divulgación de las investigaciones al público en general, ampliando el alcance originalmente planeado.

En consecuencia, el 9 de junio de 2017 se emitieron los Lineamientos Generales de Ciencia Abierta, que sustituirían a los Lineamientos Generales para el Repositorio Nacional y los Repositorios Institucionales; posteriormente, cada uno de los seis programas que componen la política ha realizado las adecuaciones necesarias en sus instrumentos con la intención

Imagen 1



Fuente: Elaboración interna para la presentación de la política en diferentes instancias.

de alinearse con lo establecido en los nuevos lineamientos, y han quedado articulados de la siguiente manera:

El CONACYT, a través de la política de Ciencia Abierta, busca asegurar la máxima diseminación del conocimiento científico, tecnológico y de innovación entre la población en general a través de cualquier medio, incluidos los digitales. La Ciencia Abierta será democrática y universal, y se regirá por los siguientes principios:

- I. Máxima apertura.
- II. Máxima captación y colaboración.
- III. Máxima facilidad de acceso.
- IV. Costos mínimos o gratuidad.
- V. Respeto a otros regímenes de Derecho, como la seguridad nacional, propiedad Intelectual, confidencialidad y reserva de datos, secretos protegidos, entre otros aplicables.²

EL REPOSITORIO NACIONAL Y LOS REPOSITORIOS INSTITUCIONALES

El programa de repositorios es el que da origen a la Política de Ciencia Abierta. Tiene por objetivo acopiar, preservar y asegurar el acceso abierto a los recursos de información científica, tecnológica y de innovación generados principalmente con recursos públicos. Para lograrlo, el programa se desagrega en dos componentes: 1) Repositorio Nacional y 2) Repositorios institucionales. Cada uno de estos componentes tiene un objetivo específico, que son:

² De acuerdo con lo establecido en el Lineamiento Cuarto. Componentes y principios de la Política de Ciencia Abierta en los Lineamientos Generales de Ciencia Abierta (2017).

- **Repositorio Nacional:** diseminar los recursos de información (publicaciones científicas, productos del desarrollo tecnológico y la innovación y los datos primarios de la investigación) provenientes de los Repositorios Institucionales para fomentar su utilización, reúso y acelerar la colaboración científica.

- **Repositorios Institucionales:** apoyar mediante una convocatoria pública a aquellas instituciones públicas o privadas que realicen investigación científica y tecnológica el desarrollo de repositorios institucionales que agreguen valor al Repositorio Nacional.

Los repositorios mencionados contendrán tres tipos de información sin perjuicio de las disposiciones en materia de patentes, protección de la propiedad intelectual o industrial, seguridad nacional y derechos de autor: I) Publicaciones científicas, II) Productos del desarrollo tecnológico y la innovación y III) Datos primarios de las investigaciones.³

Si bien muchos de los agregadores de repositorios tradicionales se concentran especialmente en los denominados repositorios de literatura, una de las particularidades del Repositorio Nacional es su capacidad para agregar también datos primarios de las investigaciones, que se definen en los LER como aquella información recolectada y utilizada para la investigación académica, científica y de innovación. Esta

³ De acuerdo con lo establecido en el Lineamiento Cuarto. Componentes y principios de la Política de Ciencia Abierta en los Lineamientos Generales de Ciencia Abierta (2017).

información será presentada en los formatos originales de su creación y deberá contar con licencias que permitan su libre reutilización a través de los repositorios institucionales que las alojen.⁴

Su depósito, además, está condicionado a que se cumpla con una serie de documentos que permitan al usuario conocer la información y las instrucciones relevantes sobre la recopilación de los datos, la unidad de observación o la información del ponderador, entre otros. Estos documentos adicionales son a) Metodología, b) libro de códigos, c) cuestionario (si aplica) y d) resumen de contenido.

Contar con la posibilidad de compartir este tipo de recursos de información permiten a los investigadores y estudiantes de las instituciones depositar en sus repositorios estos conjuntos de datos que fomentan el aceleramiento del proceso de la investigación científica, acortando los tiempos de la liberación de los datos posteriores a los procesos editoriales y compartiéndolo libre y gratuitamente.

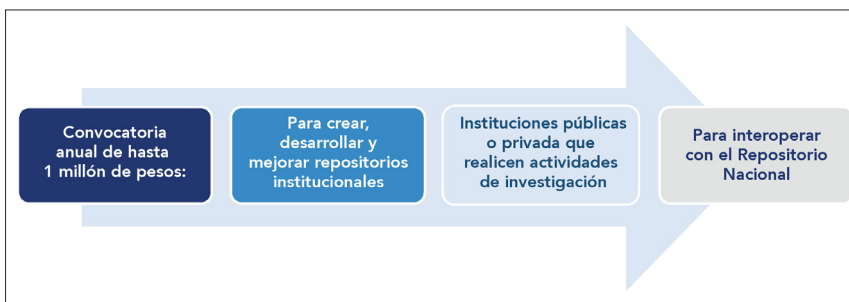
Aunado a la publicación de los lineamientos para los repositorios, desde diciembre de 2015 el CONACYT ha publicado tres convocatorias anuales con la finalidad de apoyar a las instituciones públicas y privadas que realicen actividades de investigación científica, tecnológica y de innovación y que busquen construir, mejorar o adecuar un repositorio institucional interoperable con el Repositorio Nacional, de acuerdo con lo establecido en los Lineamientos Específicos —antes Técnicos— para Repositorios.⁵ Dichas convocatorias han otorgado hasta \$1 000 000.00 (un millón de pesos) a diversas

⁴ Véase el Lineamiento Séptimo. Datos primarios de las investigaciones, de los Lineamientos Específicos para Repositorios.

⁵ De acuerdo con el objetivo de la Convocatoria 2017 para Desarrollar Repositorios Institucionales de Ciencia Abierta.

instituciones que poco a poco se han ido integrando al Repositorio Nacional a través de la interoperabilidad entre las plataformas, gracias al protocolo Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH por sus siglas en inglés), diseñado para la cosecha o recolección de metadatos de recursos de información, que permite redireccionar al usuario por medio de enlaces al recurso u objeto del repositorio de origen.

Imagen 2



Fuente: Elaboración interna para la presentación del proceso de las convocatorias en diversas instancias.

Al 26 de junio de 2018, se cuenta con más de cincuenta Repositorios Institucionales interoperando con el Repositorio Nacional, que aportan más de 22mil recursos de información y rondando el millón de visitas desde el lanzamiento de la versión 1.0 del RN a mediados de 2016.⁶

La meta para este año es que se encuentren interoperando más de ochenta repositorios institucionales y que se encuentren

⁶ Información disponible en el sitio del Repositorio Nacional <https://repositorionacionalcti.mx/>.

disponibles más de 50mil recursos de información, provenientes tanto de repositorios de literatura como de datos.

FACTORES DETERMINANTES PARA LA IMPLEMENTACIÓN DEL ESQUEMA DE METADATOS PARA REPOSITARIOS DE DATOS

Uno de los elementos clave para lograr la interoperabilidad es el manejo de esquemas de metadatos estandarizados y adaptables para el correcto intercambio de información. Existe una amplia variedad de esquemas de metadatos compatibles con el estándar del OAI-PMH que los trabaja de una manera muy flexible.

El estándar estipula que, como mínimo, los proveedores de datos —en este caso, los repositorios— deben incorporar y exponer metadatos en algún esquema revisado, aceptado y conciliado que sea apropiado para el intercambio de la información de acuerdo con las disciplinas o el contenido de las plataformas (Breeding 2002).

Dentro de la discusión sobre la funcionalidad de los metadatos, Park incluso hace una analogía relacionado los esquemas de metadatos con los catálogos y las bases de datos de bibliotecas en línea tradicionales para encontrar, identificar, seleccionar y obtener documentos (Park 2009).

Contar con una política pública que puede otorgar estímulos a diversos tipos de instituciones del sector público o privado como universidades o centros de investigación, que a su vez concentran recursos de información de diversas áreas del conocimiento y disciplinas, requiere de implementar esquemas estandarizadas y de gran flexibilidad que permitan a las instituciones adecuar la descripción de los recursos de acuerdo a las necesidades de sus comunidades, así como mantener una

homogeneización de la información para la interoperabilidad.

Force 11 (Force 11 s.f.) incluso propone los Principios FAIR para el manejo de datos de investigación, los cuales han sido considerados como elementos clave para la adaptación de los Lineamientos Específicos para Repositorios:

1. Findable (recuperable).
2. Accesible.
3. Interoperable.
4. Re-usable.

A nivel del principio *Findable* (recuperable), se hace hincapié en que se utilicen identificadores persistentes y únicos a nivel global, que haya una enriquecida descripción de los metadatos, así como que se especifique el esquema a utilizar.

Para la accesibilidad, refieren a que los metadatos sean recuperables a través del identificador antes mencionado, con protocolos de comunicación estandarizados, libres y universalmente aplicables (como OAI-PMH), lo que permita una autenticación y autorización de las peticiones de los datos entre las plataformas.

En la interoperabilidad, se refuerza la importancia de que los metadatos sean accesibles y permitan la representación del conocimiento, que para el caso del CONACYT, se han enfocado en el uso del catálogo de áreas del conocimiento de la UNESCO.

Para que los datos sean reusables, mencionan que los metadatos deben ser compartidos de manera clara y con un lenguaje accesible, pero, sobre todo, que sean relevantes para los estándares de la comunidad que los va a utilizar.

Algunos de los elementos clave más persistentes en la

literatura y los ejercicios de evaluación sobre la calidad de los metadatos son la consistencia y la persistencia (Stivila y Gasser *apud* Park 2009). Si bien la variedad de “estándares” es amplia, para el manejo de una gama de tipología de repositorios es necesario que los esquemas a utilizar sean persistentes y además sean lo suficientemente adaptables para el uso correcto de los metadatos por parte de las instituciones, de modo que se exponga la mayor cantidad de información y se limite lo más posible el margen de información no recuperable.

Greenberg (Greenberg *et al.* 2009) menciona que no tiene sentido crear esquemas totalmente nuevos cuando otras iniciativas se han tomado el tiempo para deliberar y formalizar las suficientes propiedades de los metadatos.

En este ejercicio de exploración en la literatura y la implementación de esquemas de metadatos en redes de repositorios, algunos de los factores determinantes para la selección han sido:

1. Sencillez
2. Interoperabilidad
3. Adaptabilidad
4. Arquitectura semántica
5. Consistencia

Esto llevó a que para los repositorios de literatura se haya establecido el Apéndice 1 de los Lineamientos Específicos para Repositorios, que se encuentra basado en OpenAIRE 3.0 para el manejo de repositorios de literatura que a su vez cuenta con elementos de DublinCore, mientras que el esquema de metadatos para repositorios de datos se

7 Para mayor especificidad sobre el uso de los DataCite se puede consultar el Apéndice 1 de los Lineamientos Específicos para Repositorios del CONACYT.

encuentra en el Apéndice 2 de los mismos lineamientos, basados en OpenAIRE 2.0 para el manejo de repositorios de datos, estructurados con DataCite.⁷

OpenAIRE ha demostrado a lo largo de los años una gran persistencia y consistencia en sus esquemas, tomando en cuenta las ventajas que ofrecen diversos modelos de metadatos como DublinCore y DataCite, por citar algunos, trabajando colaborativamente no sólo con la comunidad europea, sino con la participación de expertos en materia de repositorios de América del Norte y Latinoamérica, lo que ha permitido tomar estos modelos y flexibilizar el uso de algunos metadatos para el mayor beneficio de diversos tipos de repositorios.

Estos convenios de comunicación e intercambio de información comparten la meta de proveer un esquema de metadatos independiente del dominio y proporcionar interoperabilidad mediante un pequeño número de propiedades, de la manera más sencilla posible y manteniendo las barreras técnicas para la implementación lo más bajas posible, lo que fortalece las capacidades para el desarrollo e implementación de estas plataformas bajo el paradigma de la Política de Ciencia Abierta en México, gracias a la vasta documentación que la comunidad global pone en línea a través de diferentes foros y en el sitio de OpenAIRE.

VISIÓN 2.0 DEL REPOSITORIO NACIONAL

Los metadatos son “datos sobre los datos”, lo que vuelve a los repositorios herramientas que permiten la explotación e interconexión de datos más allá del contenido que contienen y describen. El Repositorio Nacional genera una gran cantidad de información gracias al suministro de estadísticas de uso de los Repositorios Institucionales

y a sus propias herramientas de registro de consumo, que al cruzarse puede generar información de gran valor esta creciente tendencia de análisis de datos.

La visión 2.0 del Repositorio Nacional se encamina a la adaptación y explotación de esta plataforma para exponer libremente estos datos y estadísticas sobre la información que se está consultando, las instituciones que están haciendo mayor uso de las plataformas a nivel de depósito y consultas, así como la generación de perfiles de usuarios que utilicen esta plataforma centralizada para su mejoramiento y adecuación a futuro, considerando también las nuevas tendencias tecnológicas y de preservación de los documentos.

Los datos primarios de las investigaciones juegan un papel importante en la potencialización de los repositorios y del desarrollo científico, académico y de innovación en México. La muestra está en la diversidad de repositorios de datos que actualmente interoperan o están en proceso de lograrlo, como son: de observaciones geoespaciales, de ciencias del mar y limnología, de especies de maíz y trigo, de genomas, especies naturales y de sismicidad en México, entre otros.

Actualmente, se encuentran interoperando más de setecientos conjuntos de datos en el Repositorio Nacional. Si bien el porcentaje relativo con el poblamiento total aún es bajo, se estima que esta cifra se multiplicará con la interoperabilidad de los repositorios que se encuentran en proceso de ser cosechados, siendo los datos primarios de las investigaciones el recurso de información que mayores números aporta al Repositorio Nacional.

La continuidad de la publicación de convocatorias y el otorgamiento de los recursos económicos son importantes para apoyar a las instituciones para desarrollar y fortalecer no solo las plataformas, sino sus capacidades, recursos

humanos y el *know how* para generar nuevos especialistas en materia de repositorios, que seguirán innovando con la implementación de nuevas herramientas, infraestructuras o conceptos a la escena de los repositorios institucionales, como ya ha sucedido desde hace dos años, cuando se lanzó la primera versión del Repositorio Nacional.

BIBLIOGRAFÍA

- Breeding, Marshall. 2002. "Understanding the Protocol for Metadata Harvesting of the Open Archives Initiative", *Computers in Libraries*, vol. 22, núm. 8. Disponible en <https://librarytechnology.org/document/9944>.
- CONACYT. 2017a. *Informe general del estado de la ciencia, la tecnología y la innovación 2016*. México: Consejo Nacional de Ciencia y Tecnología. Disponible en <http://www.siicyt.gob.mx/index.php/estadisticas/informe-general/informe-general-2016/3835-informe-general-2016/file>.
- 2017b. *Lineamientos Generales de Ciencia Abierta*. México: Sistema Integrado de Información sobre Investigación Científica, Desarrollo Tecnológico e Innovación. Disponible en <http://www.siicyt.gob.mx/index.php/normatividad/conacyt-normatividad/programas-vigentes-normatividad/lineamientos/lineamientos-generales-de-ciencia-abierta>.
- 2017c. *Lineamientos Jurídicos de Ciencia Abierta*. México: Sistema Integrado de Información sobre Investigación Científica, Desarrollo Tecnológico e Innovación. Disponible en <http://www.siicyt.gob.mx/index.php/normatividad/conacyt-normatividad/programas-vigentes-normatividad/lineamientos/lineamientos-juridicos-de-ciencia-abierta>.
- 2017d. *Lineamientos Específicos para Repositorios*. México: Sistema Integrado de Información sobre Investi-

- gación Científica, Desarrollo Tecnológico e Innovación. Disponible en <http://www.sicyt.gob.mx/index.php/normatividad/conacyt-normatividad/programas-vigentes-normatividad/lineamientos/lineamientos-especificos-para-repositorios>.
- FORCE11. *The FAIR data principles*. Disponible en <https://www.force11.org/group/fairgroup/fairprinciples>.
- Greenberg, J., H. C. White, S. Carrier y R. Scherle. "A metadata best practice for a scientific data repository", *Journal of Library Metadata*, vol. 9, núm 3–4 (2009): 194-212. <https://doi.org/10.1080/19386380903405090>
- Iniciativa de Budapest para el Acceso Abierto*. 2002. <https://www.budapestopenaccessinitiative.org/translation/spanish-translation>.
- Ley de Ciencia y Tecnología*. 2015. Capítulo X Del Acceso Abierto, Acceso a la Información Científica, Tecnológica y de Innovación y del Repositorio Nacional. Capítulo adicionado DOF 20-05-2014. México: Cámara de Diputados del H. Congreso de la Unión. Disponible en http://www.diputados.gob.mx/LeyesBiblio/pdf/242_081215.pdf.
- Oliphant, Tami. 2017. "A case for critical data studies in library and information studies", *Journal of Critical Library and Information Studies*, vol. 1, núm. 1.
- Park, Jung-Ran. 2009. "Metadata Quality in Digital Repositories: A Survey of the State of the Art", *Cataloging & Classification Quarterly*, vol. 47.
- The University of Melbourne. *Melbourne Policy Library. Management of Research Data and Records Policy* (MPF1242). Disponible en <https://policy.unimelb.edu.au/>.

Habilidades requeridas por el profesional de la Industria 4.0. Aparición de un nuevo perfil para el área de producción en la Cuarta Revolución Industrial

MARCO BRANDÃO

Universidade Federal Fluminense, UFF, Brasil

INTRODUCCIÓN

El historiador Yuval Noah Harari, profesor de la Universidad Hebrea de Jerusalén y el autor del libro *Sapiens: a brief history of humankind* (2015) profundiza en la cuestión del trabajo humano en relación con el avance de la inteligencia artificial: más que decir que los seres humanos serán sustituidos por máquinas en las actividades, considera que se crearán nuevas profesiones, pero no todas las personas lograrán calificarse para estas actividades. En su artículo “The meaning of life in a world without work”, publicado en *The Guardian* (2017), el escritor afirma que una nueva clase de personas emergerá hasta 2050: la de los inútiles. “Son personas que no sólo estarán desempleadas sino que no serán empleables” (Harari 2017).

Frente a todas esas transformaciones por las cuales pasa el mundo del trabajo, la que cambia el sentido y significado del trabajo de la humanidad en cuanto a forma y producto

en un mundo rodeado por tecnologías que ya sustituye el trabajo humano en muchas actividades es un gran reto. Sin embargo, la necesaria especialización profesional, según Harari, llega a un otro nivel, no sólo restringido al hecho de tener o no empleo, sino de que las personas estén listas para ocuparlos.

Uno de estos factores emerge con el concepto de Industria 4.0, un nuevo *modus operandi* del área de producción en la cual el profesional tendrá que desarrollar habilidades relacionadas a un nuevo perfil de trabajo. No es algo perteneciente solamente a la industria como tal, sino a la forma de trabajar actualmente, ya siendo parte de muchas actividades humanas.

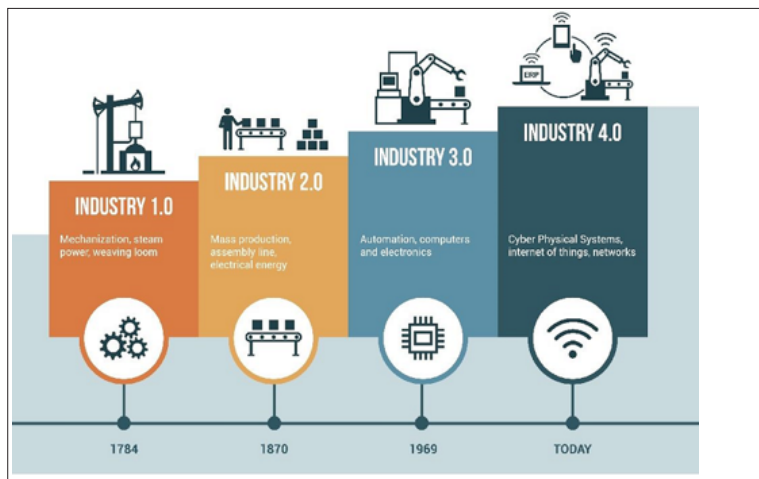
La Tecnología de Información (TI) aparece como un soporte indispensable de ese contexto. Esto no se refiere solamente a la existencia de computadoras automatizando o facilitando procesos, también a una actividad productiva que depende y genera muchos datos, información y conocimiento para su uso en fronteras de conocimiento creadas por la propia actividad productiva. El profesional, por lo tanto, ya no es solo un especialista y responsable por una parte del trabajo, debe tener la visión de su conjunto para adoptar medios adecuados para una producción eficiente y con calidad.

Como manera de ilustrar esta realidad, ese texto presenta lo que es la Industria 4.0 como un concepto y método de trabajo, identificando los aspectos que la constituyen, como las herramientas de TI, que requieren un nuevo perfil profesional, teniendo en vista su estructura técnica en el desarrollo de un producto. El artículo concluye presentando como ejemplo una experiencia en curso de un trabajo de producción musical basado en el concepto y método de la Industria 4.0.

¿QUÉ ES LA INDUSTRIA 4.0?

Industria 4.0 es el término creado para hacer referencia a una nueva forma de producción de la industria, es decir, la “Cuarta Revolución Industrial”, en una clara relación con las otras tres fases del desarrollo industrial. La expresión se originó en el año 2011 dentro de un proyecto de alta tecnología del gobierno alemán que planteaba, en principio, la informatización de la manufactura con la adición de los aspectos de interoperabilidad entre los sistemas humanos y de las fábricas a través de Internet y la *Cloud Computing*, de la virtualización creada por sensores de datos interconectados, de la descentralización de la toma de decisiones, especialmente sin la intervención humana, de la capacidad de recopilar, analizar datos y entregar conocimientos, de ofrecer servicios a través de Internet y de la adaptación flexible (Hermann, Pentek y Otto 2015).

Figura 1. Fases de la Revolución industrial



Fuente: Hammel Scale (2018).

De la misma manera que en la primera fase de la revolución industrial se movilizó la mecanización de la producción con máquinas de vapor, en la segunda fase se introdujo la producción la energía eléctrica y en la tercera fase inicio la automatización mediante el uso de aparatos y dispositivos electrónicos; la cuarta fase es una que avanza, sobre todo, con el intercambio de datos y el uso de sistemas ciberfísicos del Internet de las cosas y de la *Cloud Computing* como nuevas tecnologías de producción.

La Industria 4.0 también ha sido el camino a la competitividad del sector industrial no sólo por su modo de funcionamiento a través de las tecnologías digitales, sino por la generación de productos capaces de satisfacer las demandas de producción actuales.

En los países latinos, la 4.0 todavía no es imprescindible, ya sea por el retraso en la integración de las tecnologías físicas y digitales en las etapas de desarrollo de un producto, ya sea por las realidades económicas marcadas sobre todo por empresas que todavía no ven la ventaja competitiva en este nuevo modo de producción. Según la Confederación Brasileña de la Industria (CNI 2016), las industrias siguen conociendo la digitalización y los impactos que puede tener sobre la competitividad. La ignorancia va más allá cuando se tienen en cuenta otros aspectos necesarios para la adopción a gran escala de la industria 4.0, como los educativos y culturales.

Como un esquema general de una industria de 4.0, el intercambio de datos, el uso de sistemas ciberfísicos, el Internet de las cosas y la *Cloud Computing* como tecnologías de producción aparecen arreglados en conjunción con las etapas de desarrollo del producto como se ve en la figura 2:

Figura 2. Esquema general de una Industria de 4.0



Fuente: Silva (2017).

En ese escenario, un nuevo perfil profesional emerge con el concepto de Industria 4.0, en el cual para trabajar en el área de producción de la fábrica él tendrá que desarrollar habilidades relacionadas a la visión técnica, la multidisciplinariedad, la colaboración, el dominio lingüístico, el sentido crítico y la flexibilidad (ABC 2017), todo esto impregnado por entornos y herramientas virtuales que requieren un gran nivel de capacidades.

Por lo tanto, parte de los retos de la Cuarta Revolución Industrial consiste en integrar este perfil de la fuerza laboral sin apretar las cuestiones de empleabilidad de los ciudadanos

de países como los pertenecientes al bloque latinoamericano, un hecho crónico que se incrementa a medida que estos nuevos estándares de productividad son exigidos.

Factores como el gran y rápido incremento en el volumen de datos, del cómputo y de la conectividad, los cuales han forzado un avance en las capacidades analíticas, además de las nuevas formas de interacciones entre humanos y máquinas con innovaciones que permiten la transferencia de datos digitales para algo físicamente utilizable conducen esta transición y requieren cada vez más al profesional que los domine en la producción (Hermann, Pentek y Otto 2015). De hecho, el consultor Edson Miranda da Silva, de la Ronín Consultoría, nos señala que entre los retos principales de Industria 4.0 están las cuestiones de la falta de profesionales preparados (Silva 2017).

Y va más allá: cuando hablamos de “producción”, de “trabajo” hoy en día en muchos sectores que no son industriales, hablamos de procesos que toman en cuenta estos aspectos: la educación, la medicina, la ingeniería, el arte, etcétera, requieren de sus profesionales el mismo dominio de estos aspectos como factores de competitividad.

HABILIDADES REQUERIDAS AL PROFESIONAL DE LA INDUSTRIA 4.0

Las transformaciones incursadas por la Industria 4.0 llegan a todo el universo del trabajo. En ese particular, la necesaria especialización profesional llega a otro nivel, no restringido a un área o carrera, sino a la articulación del conocimiento de muchos sectores.

Hay muchas habilidades que uno puede requerir de un profesional hoy en día. En el área de producción, bajo el concepto de Industria 4.0, para que el profesional trabaje

tendrá que desarrollar, entre otros, habilidades relacionadas a la visión técnica, a la multidisciplinariedad, a la colaboración, al dominio lingüístico, al sentido crítico y a la flexibilidad (ABC 2017). Todo esto impregnado por entornos y herramientas virtuales que requieren otro gran nivel de capacidades.

Figura 3. Entornos y herramientas virtuales



El campo de la Tecnología de Información (TI) aparece como un requisito indispensable en la formación del profesional, sin lo que se ve obstaculizada su inserción en esta nueva realidad. Esto no se refiere a un uso genérico e indiscriminado de computadoras automatizando o facilitando procesos de trabajo, sino de un nuevo *modus operandi* del profesional en el cual su actividad productiva requerirá que él desarrolle habilidades relacionadas con el uso de las herramientas digitales y de información articulada en fronteras de conocimiento.

Para tener idea de lo que eso implica, es necesario mirar el país de origen de la Industria 4.0, Alemania, donde está más avanzada. El Boston Consulting Group (ABC 2017) indica, entre otras cosas, que la demanda por empleados que también puedan dominar las áreas de TI, especialmente en la parte de *software*, requerirá cada vez más un número de personas con altas calificaciones que tengan incorporado este aspecto de manejo de *software*. Esto se debe a la medida que se están desarrollando sistemas de producción y de trabajo para ser operados por *software*, ya sea para el control humano o para la automatización de tareas; es decir, será imposible estar en los sistemas de producción y de trabajo sin utilizarlos.

Y no sólo eso: en los *softwares* también está uno de los principales insumos para la producción y diferenciación competitiva para las estrategias de desarrollo empresarial, esto es, los datos e informaciones necesarios a la creación del producto y al posicionamiento estratégico de la producción. Así que trabajar proactivamente no es solamente dominar las herramientas, sino también articular el conocimiento, desarrollar nuevas habilidades, ejercer funciones más complejas y creativas con la responsabilidad y la visión de todo el proceso productivo (ABC 2017).

Por lo tanto, es necesario estar abierto a los cambios, ser flexible para añadir nuevos conocimientos y adaptarse a nuevas funciones, acostumbrarse al aprendizaje multidisciplinario continuo. Esto no significa que el conocimiento técnico haya perdido su importancia: simplemente ello no es más suficiente. “Es necesario especializarse en varios frentes y conocer un poco de todo. Tiene que gustar de la tecnología, de la innovación y, sobre todo, tener curiosidad de aprender y seguir una industria que siempre se reinventa” (ABC 2017).

LA TECNOLOGÍA DE INFORMACIÓN EN LA PRODUCCIÓN

Superando el hallazgo de las TI's como modificadores importantes de los procesos de trabajo y producción, queda ahora entender de otra manera cómo éstas transforman la industria 4.0, ya que en su fase anterior —de la Industria 3.0—también ya estaban presentes en los procesos de automatización.

Al sí revisar el proyecto original del gobierno alemán, su fuerte está en una informatización de la manufactura con la adición de la interoperabilidad entre los sistemas humanos y de las fábricas a través de la Internet y de la *Cloud Computing*, de la virtualización creada por sensores de datos interconectados, de la toma de decisiones sin la intervención humana y de la capacidad de recopilar, analizar datos y entregar conocimientos, servicios a través de la Internet y de la adaptación flexible (Hermann, Pentek y Otto 2015). Así que la TI sigue como un soporte indispensable de ese contexto.

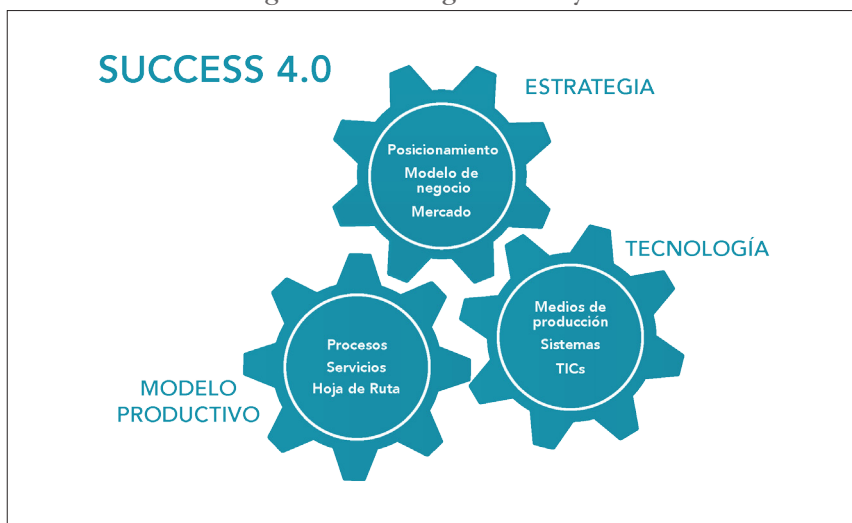
Sin embargo, para que esa adición sea exitosa, son cruciales un conjunto de factores, en particular la dimensión de la interoperabilidad entre los sistemas humanos y aquellos emulsionados por las TI's, es decir, el intercambio de datos de sistemas ciber-físicos, posible a través del Internet de las cosas y de la *Cloud Computing*. Esto no se refiere solamente a la existencia de computadoras automatizando o facilitando procesos, sino a una actividad productiva que depende y genera muchos datos, información y conocimiento. El profesional, por lo tanto, debe interactuar con dispositivos y *softwares* interconectados y capaces de ofrecer una visión del conjunto para adoptar medios adecuados para una producción eficiente y con calidad.

Esa nueva forma de producción es, por lo tanto, también una nueva forma de producción de datos, de información que, a su vez, necesitan ser recopilados, analizados tanto por los sistemas digitales, cuanto por los sistemas humanos. En ese sentido, hay que tener una robusta estructura de metadatos que permitan la interoperabilidad interna y externa de estos sistemas, incluso porque ellos podrán ser accedidos remotamente a través de las herramientas de gerenciamiento remoto y de *Cloud Computing*.

Los metadatos, sin duda, van asegurar esa interoperabilidad y también a definir la territorialidad de la producción: trabajar con el área de producción y con personas ubicadas en diferentes partes del mundo también es el entendimiento de una producción 4.0. Con tantos dispositivos físicos virtualizados a través del Internet de las cosas y operados por éste, esta forma de recopilar datos ha generado en las grandes bases de dato (big datas), una gestión cada vez más eficiente de la producción, sobre todo porque ellos también definen la estrategia (posicionamiento, mercado, modelo de negocio) que será de interés de la producción. En otras palabras, de los big datas se pueden extraer metadatos con información importante para saber “qué”, “cuándo”, “cuánto”, “dónde” y “para quién” producir.

A través de la European Factories of the Future Research Association (EFFRA), según Alzaga y Larreina (2016), Europa está desarrollando las “fábricas del futuro”, un conjunto de proyectos innovadores para reaccionar o analizar las oportunidades de este nuevo escenario de la Industria 4.0 con reflexiones realizadas en ese nivel estratégico y otros dos más:

Figura 4. Estrategia Industry 4.0



Fuente: Alzaga y Larreina (2016).

Así, además del nivel estratégico, es necesario analizar cómo la propuesta de valor será definida en el modelo productivo que va a responder a esa estrategia, definir y concretar una hoja de ruta en este sentido para identificar las tecnologías clave sobre las que apoyarse y cómo hacerlas interoperables para permitir el tratamiento de los datos y las informaciones necesarias al producto.

Por más de una razón se entiende que los profesionales de estos días requieren nuevas habilidades que la escuela, la formación debe fomentar y fortalecer: por un lado, aquellas asociadas al conocimiento de las tecnologías y, por otro, aquellas habilidades que permitan el trabajo interdisciplinario, imprescindible en la Industria 4.0.

LA APARICIÓN DE UN NUEVO PERFIL PARA EL ÁREA DE PRODUCCIÓN

Como manera de ilustrar esta realidad, aquí se describe una experiencia en curso de una producción musical basada en el concepto y método de la Industria 4.0.

¿Por qué una producción musical? Inicialmente porque la industria de la música es una de las más afectadas por los cambios tecnológicos. De los músicos y artistas involucrados en la creación a los profesionales de la producción y distribución, todos han tenido sus actividades profundamente afectadas por las TI's. Desde la creación del formato digital, no sólo se modificaron los soportes, sino también las formas de creación, producción, distribución y consumo de música.

Según un informe de la Representing the Recording Industry Worldwide (IFPI), en abril de 2017 el consumo de música digital en el mundo alcanzó el 50 por ciento debido al vertiginoso incremento del streaming, que es una tecnología de informaciones multimedia que utiliza Internet sin la necesidad de descargas de archivos, eso representa un 60,4 por ciento de incremento en comparación con el año de 2015. La descarga de archivos, otra forma de consumo de música, tuvo una caída de 20,5 por ciento (IFPI 2017).

Además del soporte del producto final, la música, su producción también está envuelta por cambios que están más allá del talento y el acto creativo de la creación musical, del músico o artista. Según Sandro Chagas (2015), el esquema abajo contempla el equipo y los recursos mínimos necesarios en la producción de una canción:

1. El músico o compositor: persona o conjunto de personas que empieza la producción, la cual puede o no ser el ejecutor de la obra. En la

mayoría de los casos, es un músico que participará con algún instrumento musical o como el cantante de la melodía creada, y debe tener conocimiento de las técnicas musicales, además de una propuesta o línea temática de composición que identifica el estilo/género musical, público, mensaje etcétera;

2. La música o melodía: la pieza compuesta por el músico/compositor que se producirá con el fin de corresponder a lo planteado por el artista;
3. La orquestación o instrumentación: es el proceso de componer la música o melodía con otros instrumentos. En esta etapa, se invita a otros músicos, en caso de que la creación de la obra fuera realizada por un músico o compositor individual;
4. El productor musical: es el profesional al cual se presentará la pieza musical para su grabación y distribución. Él puede modificar la orquestación o instrumentación y va trabajar bastante la parte timbre de los instrumentos, su captura, el soporte o formato de grabación, entre otras cosas. Él es comúnmente un músico (tiene conocimiento técnico, toca alguno o varios instrumentos), pero no es necesariamente el músico o compositor de la pieza que produce;
5. El editor: hecha la grabación, se pasa a la edición del sonido, que es el ajuste de lo que no suena bien o no fue bien capturado. Son cosas mínimas como algún desliz en el ritmo de algún instrumento, la voz del cantante que desafina en alguna nota o el recorte de las capturas de la grabación que no son necesarias a la música. La edición puede ser realizada por el propio productor

- musical o por un profesional de edición;
6. El *mixing*: es la etapa en la cual se pone cada instrumento de manera equilibrada para la uniformidad del sonido. Por ejemplo, se pone un piano a la izquierda, una guitarra a la derecha, la batería se coloca al centro, el cantante resaltado etcétera. Se realizan varios procesos de manipulación del material grabado y, por esta razón, se utilizan herramientas según la propuesta de la música: el ecualizador, el compresor, el Gate, el *limiter*, el *deEsser*, el *reverb*, el *delay*, etcétera. De la misma manera, eso puede ser hecho por el Productor o por un profesional de *mixing*;
 7. La premasterización: en la premasterización se finaliza el trabajo para enviarlo a la fábrica que realizará las copias. Es el toque final que dará un sonido con calidad para la reproducción en copias de todo lo que se hizo en los pasos anteriores;
 8. La masterización: hecha fuera del estudio, en la fábrica donde se crearán las copias para la venta de la obra después de toda la producción musical, la masterización precede a la distribución, lo que requiere *marketing* y publicidad;
 9. La distribución: según la propuesta del músico o compositor y del productor musical, es hacer llegar la producción a los mercados potenciales y a los consumidores según las estrategias establecidas de *marketing*, soporte, comercialización, etcétera;
 10. El consumo: es el objetivo final y tendrá en la estrategia un fuerte aliado. Es cierto que la música depende de gustos personales y afinidades de estilo, propuesta, momento y esto no sólo refuerza

la necesidad de una estrategia bien definida, sino que también proyecta un mercado para las futuras producciones.

La producción así tradicionalmente realizada contaba con un conjunto de profesionales especializados en cada etapa, en ambientes y con equipos específicos (estudios). Con la producción en formato digital, se inició un profundo proceso de cambio de ese esquema, que hoy depende principalmente de *softwares* para establecer los procesos identificados en las fases de 4 a 10.

Estos softwares, conocidos como Digital Audio Workstation (DAW), hacen todo el trabajo de procesamiento de audio y por lo tanto es extremadamente importante tener datos interoperables que puedan ser trabajados por DAWs de diferentes desarrolladores y estar disponibles en formatos accesibles. De ahí también la importancia de los metadatos, ya que la información musical digital es un formato binario, cuyo proceso de manipulación del material grabado y de las diversas herramientas utilizadas para ello, como el ecualizador, el compresor, el *limiter*, el *deEsser*, el delay, etcétera, son ahora virtuales. Esto es, desde el productor hasta el *mastering*, los profesionales deben ahora conocer a estas herramientas también. Esa forma de *mixing* se llama *mix in the box* (Chagas 2015).

En cuanto a los músicos, algo similar sucede en relación con la forma de composición de la música o melodía básicamente por dos razones: primero porque estos *softwares* ya les permiten realizar todas las fases de la producción musical (desde la composición hasta la distribución) sustituyendo el productor, el editor, el *mixing* etc.; segundo, que incluso el músico o compositor solista puede sustituir en el proceso de producción la necesidad de otros músicos

para la orquestación/instrumentación. Un tipo de *software* conocido como Virtual Studio Technology (VST), que son pequeños *softwares* (*plugins*) incorporados en la DAW, puede hacer el procesamiento de audio de instrumentos musicales, incluso de la voz. Así que una vez grabada una canción en este formato, lo que tenemos es una serie de metadatos que son manipulados por estos *softwares* para una propuesta sonora que definirá el estilo de la música y del artista.

Figura 5. DAW y plugins VST



Un músico que inicia una composición, graba y envía al entorno virtual para obtener las colaboraciones necesarias al proceso de producción, el cual es llevado a cabo por profesionales de diferentes partes del planeta, aportando su conocimiento y trabajo en un producto colectivo y sin territorio. Todos, por lo tanto, deben hacer su trabajo interoperable y capaz de ser manejado por DAWs y VSTs, es decir, un sistema ciberfísico que implica la transformación de la información del sonido en datos que serán manejados por diferentes personas y herramientas de TI/software a través de Internet para un producto posicionado estratégicamente. De ahí también la importancia de metadatos bien estructurados para garantizar este proceso.

CONSIDERACIONES

Aunque como experiencia en curso, ya son posibles algunas consideraciones que justificaron la realización de este trabajo y la continuidad de su análisis, las cuales son:

1. Nuevas habilidades han sido requeridas a todos los profesionales en la Cuarta Revolución Industrial. En el caso de la TI, no se trata más del puro y simple manejo de herramientas, sino de un conocimiento que rebasa los aspectos de la ejecución técnica de las tareas por humanos o por máquinas;
2. Con la Industria 4.0 se refuerza esas nuevas habilidades, especialmente aquellas que involucran el manejo de datos o metadatos, información en entornos digitales manejados a través de los

softwares para la definición estratégica, tecnológica y de modo productivo de la empresa;

3. Frente a esas transformaciones, la especialización profesional llega a un otro nivel, no más restringido a un área o carrera, sino a la articulación del conocimiento de muchos sectores. Para trabajar en el área de producción, el profesional tendrá que desarrollar habilidades relacionadas a la visión técnica, la multidisciplinariedad, la colaboración, el dominio lingüístico, el sentido crítico y la flexibilidad modelos de producción directamente asociados a la web;
4. La escuela debe, por lo tanto, fomentar y fortalecer por un lado habilidades asociadas al conocimiento de la tecnología y, por otro, las habilidades que permitan al individuo el trabajo asociado a grupos interdisciplinarios vinculados a la Industria 4.0 que actúen con la interoperabilidad permitida por las herramientas digitales en los sistemas ciberfísicos;
5. El campo de la TI conforma un nuevo *modus operandi* del profesional en el área de producción. Como se percibe en la experiencia en curso, en ella se identifican aspectos constituyen ese modo de producción basado en el concepto y método de la Industria 4.0 y, así como se revisa la importancia del uso de las herramientas de TI teniendo con base en su estructura técnica aplicada a un producto cultural.

BIBLIOGRAFÍA

- ABC, Estudio. 2017. “Como será o profissional da Indústria 4.0?”, *Exame*. Julio 07. Disponible el 30 de marzo de 2018 en <https://exame.abril.com.br/tecnologia/como-sera-o-profissional-da-industria-4-0>.
- Alzaga, Aitor y Jon Larreina. 2016. “La 4ª Revolución Industrial da lugar a la llamada Fábrica Inteligente”, *Interempresas – Metalmecánica*. Julio 01. Disponible el 24 de marzo de 2018 en <http://www.interempresas.net/MetalMecanica/Articulos/159161-La-4-revolucion-industrial-da-lugar-a-la-llamada-Fabrica-Inteligente-o-Industria-40.html>.
- Chagas, Sandro. 2015. “O passo a passo da Produção Musical em Home Studio”, *Ideaudio Home Studio*. Diciembre 15. Disponible el 25 de marzo de 2018 en <https://idaudio.com.br/producao-musical>.
- EBC, Agencia Brasil. 2016. “Pesquisa revela perfil da indústria 4.0 no brasil”. Mayo 30. Disponible el 24 de marzo de 2018 en <http://agenciabrasil.ebc.com.br/pesquisa-e-inovacao/noticia/2016-05/pesquisa.revela-perfil-da-industria-40-no-brasil>
- Hammel Scale. 2018. *Industry 4.0 Compliant Weighing Solutions*. Disponible el 30 de marzo de 2018 en <https://www.hammelscale.com/industry-4-0>.
- Harari, Yuval Noah. 2017. “The meaning of life in a world without work”, *The Guardian*. Mayo 08. Disponible el 30 de marzo de 2018 en <https://www.theguardian.com/technology/2017/may/08/virtual-reality-religion-robots-sapiens-book>.

- _____. 2015. *Sapiens: a brief history of humankind*. Gran Bretaña: Vintage.
- Hermann, Mario; Tobias Pentek y Boris Otto. 2015. "Design Principles for Industrie 4.0 Scenarios: a literatura review", *Business Engineering Institute St. Gallen*. Artículo en proceso. Disponible el 22 de marzo de 2018 en http://www.snom.mb.tu-dortmund.de/cms/de/forschung/Arbeitsberichte/Design-Principles-for-Industrie-4_0-Scenarios.pdf.
- Representing the Recording Industry Worldwide (IFPI). *IFPI Global Music Report 2017*. 2017. Abril 25. Disponible el 25 de marzo de 2018 en <http://www.ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2017>.
- Pesquisa Revela Perfil da Indústria 4.0 no Brasil. 2016. *Agência Brasil*. Mayo 30. Disponible el 24 de marzo de 2018 en <http://agenciabrasil.ebc.com.br/pesquisa-e-inovacao/noticia/2016-05/pesquisa-revela-perfil-da-industria-40-no-brasil>.
- Silva, Edson Miranda da. "Indústria 4.0: a 4ª Revolução Industrial", 2017. *Quality Way*. Noviembre 17. Disponible el 24 de marzo de 2018 en <https://qualityway.wordpress.com/2017/11/16/industria-4-0-a-4a-revolucao-industrial-por-edson-miranda-da-silva>.

Implementación de un repositorio unificado para conformar el Sistema de Inteligencia Institucional del INE

ALEJANDRO VERGARA TORRES
Instituto Nacional Electoral. México

INTRODUCCIÓN

El Instituto Nacional Electoral (INE) es el organismo autónomo encargado de organizar las elecciones presidenciales, de senadores y diputados federales en México; también es el encargado de administrar el listado nominal y padrón electoral mediante el cual a los ciudadanos se les provee de una credencial para votar que hoy en día funciona como el medio de identificación personal. El INE también se encarga de la capacitación de los ciudadanos que fungen como funcionarios de casilla el día de la Jornada Electoral, fiscaliza los recursos públicos que utilizan los partidos políticos en sus campañas y es el encargado de garantizar el uso de sus prerrogativas en radio y televisión, entre otras funciones que derivan en una gran acumulación de información. El presente capítulo da cuenta de las diferentes acciones realizadas para la implementación de un repositorio unificado de información a partir del cual se conformó el Sistema de Inteligencia Institucional del Instituto Nacional Electoral.

Diversos factores, entre ellos la acumulación de grandes volúmenes de datos, han hecho del análisis de datos una tecnología esencial y cada vez más necesaria. Las empresas y organizaciones que procesan su información mediante herramientas analíticas se vuelven más productivas y rentables que otras al explotar de mejor manera la información que producen, les rodea y poseen, y utilizarla a su favor para la toma de decisiones. Por ende, la adopción de herramientas analíticas se está acelerando y cobrando mayor relevancia en muchas organizaciones, y más aún, tomando en consideración los debates del gobierno abierto y la ciudadanía digital.

Sin embargo, uno de los retos mayores, que conlleva un gran esfuerzo, es el de recabar y concentrar la información histórica y presente en un solo repositorio; asegurar su homogeneidad, calidad y consistencia es tan solo el primer preparativo para poder explotarla. Por ello, un primer esfuerzo importante para analizar múltiples fuentes de información es contar con un repositorio unificado en donde se concentren y convivan diversas fuentes de datos con una estructura homogénea y correctamente relacionada para poder realizar análisis de esas grandes cantidades de información ya integradas.

El Sistema de Inteligencia Institucional (SII) a grandes rasgos se compone de: 1) un repositorio unificado de información o *dataware house*; 2) un componente analítico que permite el modelado y análisis de la información que alberga, y 3) un componente de visualización con el cual pueden producirse diversos gráficos y análisis visuales. Este sistema es una herramienta informática que favorece la cultura analítica dentro del Instituto, cuyo objetivo principal es integrar en un solo repositorio la información que se ubica en bases de datos históricas y presentes de los diversos sistemas operados por el Instituto Nacional Electoral, así como diversas

fuentes de datos externas (Instituto Nacional de Estadística y Geografía o bien información producida por el Instituto Federal de Telecomunicaciones), previamente depuradas y homogeneizadas para facilitar su consulta y análisis, lo que permite la explotación de información a partir de la construcción de diversos productos de información (tableros de control, gráficas, consultas dinámicas, mapas y herramientas visuales, entre otros).

El Sistema de Inteligencia Institucional le permite al Instituto:

- Contar con una base de conocimiento de la información que se produce en los sistemas de información que permite conocer y comparar información histórica o que provenga de otros procesos electorales.
- Lograr la homogeneidad de la información al cruzar datos cartográficos, relacionados con procesos electorales (locales o federales) y con base en un calendario proveniente de diversos sistemas de información.
- Obtener datos de diversos sistemas de información y de áreas operativas distintas.
- Contar con un modelo integral que permita acoplar los datos contenidos en los sistemas de información con los que cuenta el Instituto.
- Garantizar la permanencia del conocimiento concentrada en un repositorio institucional único.
- Contar con una herramienta que realiza diversos análisis de información que pueden ser de índole cuantitativo, predictivo, de comportamiento o ponderativos.

- Consolidar la información institucional a pesar de que los sistemas informáticos desarrollados para cada área del Instituto no guarden relación directa.
- Cotar a las áreas del Instituto de un herramienta con capacidades tecnológicas sin precedente que permita hacer eficiente y fortalecer la toma de decisiones.

El objetivo de este capítulo se centra en compartir las diferentes etapas realizadas para la implementación de este proyecto, las grandes dificultades que conlleva un proyecto de esta magnitud, su gestión y relevancia para que proyectos como éstos logren construir acervos de información de calidad explotables y permitan salvaguardar el conocimiento acumulado de la mejor manera para generar un valor agregado para las organizaciones y para la ciudadanía mediante una mejor toma de decisiones.

SOBRE EL ANÁLISIS DE DATOS, LA CIENCIA DE DATOS, LA INTELIGENCIA DE NEGOCIOS Y EL *BIG DATA*

Comprender el uso de herramientas analíticas, en particular aquellas relacionadas con la inteligencia de negocios, requiere conocer un poco del contexto en donde este tipo de soluciones se ubican dentro del amplio campo del análisis de datos, por ello a continuación se describirán los diferentes tópicos relacionados con el análisis de datos.

El primer término que es necesario comprender es el de la ciencia de datos, el cual relaciona y da origen a la gran mayoría de las herramientas analíticas hoy conocidas. David Donoho (2017), en su artículo “50 years of Data Science” da

cuenta del origen de la necesidad de analizar los datos; como el autor lo cuenta, hace más de cincuenta años, John Tukey propuso reformar la estadística académica y en su publicación “The Future of Data Analysis” puntualiza sobre la existencia de una aún no reconocida ciencia cuyo objetivo era “aprender de los datos” o bien del “análisis de los datos”. Diez o veinte años después, John Chambers, Bill Cleveland y Leo Breiman promueven dentro de círculo de la academia de la estadística expandir sus límites más allá del dominio clásico de la teoría estadística.

Chambers llamó la atención de la academia para enfatizar la preparación de los datos más que el modelado estadístico, y Beiman promovió un mayor énfasis en la predicción que en la inferencia. Cleveland incluso sugirió el pegajoso nombre Data Science para ese campo.

En este sentido y como John D. Kelleher y Brendan Tierney (2018) refieren en su libro *Data Science*, nunca ha sido más fácil que hoy en día para las organizaciones recopilar, almacenar y procesar datos. El uso de la ciencia de datos está impulsado por el aumento de los grandes datos y las redes sociales, el desarrollo de la informática de alto rendimiento y la aparición de métodos tan potentes para el análisis y modelado de datos como aprendizaje profundo. La ciencia de datos abarca un conjunto de principios, definiciones de problemas, algoritmos y procesos para extraer patrones no obvios y útiles de grandes conjuntos de datos y está estrechamente relacionado con los campos de la minería de datos y el aprendizaje automático, con un alcance aún más amplio.

Como Keheller y Tierney (2018) refieren en la introducción de su libro, el registro de información se distingue en dos grandes grupos, 1) datos transaccionales, los cuales incluyen información de cada evento, la venta un artículo, la entrega

de algún bien, etcétera, y 2) datos no transaccionales, como los datos demográficos.

El desarrollo de las computadoras ha contribuido a un masivo incremento de información que es recolectada y almacenada en bases de datos con tablas de un campo por instancia y una columna por atributo; en este sentido, las bases de datos son la tecnología natural para obtener transacciones estructuradas o datos operacionales. Sin embargo, hoy en día, en razón de la gran variedad de datos que se generan, empezaron a surgir dificultades para analizar esas grandes cantidades de información. Uno de los principales problemas era que los datos se encuentran almacenados en múltiples bases de datos independientes entre sí. Esto provocó el desarrollo de una tecnología llamada *data warehouses* o repositorios de datos en donde la información es obtenida a lo largo de la organización e integrada para que pueda ser analizada de una mejor manera. Éste es el origen de las herramientas analíticas de inteligencia de negocios, las cuales se dedican a procesar múltiples fuentes de datos estructurados provenientes de bases de datos de una organización para su cruce, integración y posterior análisis. Este tipo de tecnología es la que se empleó para la integración del Sistema de Inteligencia Institucional y que más adelante se detallará ampliamente.

Adicionalmente, en las últimas dos décadas el uso de dispositivos móviles conectados a Internet, el crecimiento y la masificación de las redes sociales combinado con diversas plataformas en línea han tenido un impacto dramático en la cantidad de datos recopilados. No solo ha aumentado la cantidad de datos, sino también su diversidad, fotografías, tuits, correos electrónicos, mensajes, videos, imágenes, etcétera, todos ellos relacionados con los metadatos (datos que describen la estructura y las propiedades de los datos en bruto) de cada uno de estos eventos, a partir de lo cual podemos empezar a comprender

el significado del término *big data*, que a menudo se define en términos de las tres “V”: el Volumen extremo de datos, la Variedad de los tipos de datos y la Velocidad a la que deben procesarse los datos. Esto ha derivado en el desarrollo de nuevas tecnologías de consulta y búsqueda de información que no se encuentre estructurada como sí lo están las bases de datos relacionadas, es decir, las nuevas tecnologías de *big data* se enfocan en procesar grandes bases de datos no relacionadas. Es decir, que la existencia de *big data* ha promovido el desarrollo de nuevos esquemas de procesamiento de datos ya que al manejarse grandes cantidades de información a velocidades muy grandes puede ser muy útil desde una perspectiva computación y de velocidad distribuir los datos en múltiples servidores.

El *big data* es sólo un gran volumen de datos concentrado, pero el valor agregado es el conocimiento que obtenemos a partir de estos, lo cual sólo se logra a partir de estructurarlos, modelarlos y explotarlos para influir valor en la toma de decisiones.

En resumen, podríamos decir que la ciencia de datos ha permitido el desarrollo de diversas tecnologías, algoritmos, modelos de procesamiento de información para trabajar tanto con información estructurada como no estructurada, y en ese sentido podemos decir que la ciencia de datos engloba múltiples disciplinas relacionadas con el estudio de los datos, y *big data* es una de ellas, pues se concentra en optimizar la explotación de fuentes de datos no relacionadas y relacionadas haciendo uso de algoritmos precisos de análisis y explotación de información y usando múltiples computadoras para su procesamiento mientras que las herramientas de inteligencia de negocios están enfocadas a explotar y consultar múltiples bases de datos de una organización, para lo cual hace uso de tecnologías como *data warehouses* para concentrar la

información en un repositorio único que resuelva la problemática de múltiples fuentes de datos que no guardan una relación estrecha.

Es así que el Sistema de Inteligencia Institucional pretende integrar en un solo repositorio las bases de datos relacionadas y diversas fuentes adicionales de información estructurada para que puedan ser consultadas, procesadas y analizadas de manera eficiente, bajo un modelo de datos multidimensional unificado; es decir, gobernado por un modelo de datos global que permita relacionar todas las fuentes que se desee analizar con la garantía de que sean homogéneas, que puedan intersectarse y que se eliminen las inconsistencias que surgen de contar con múltiples bases de datos dispersas.

IMPLEMENTACIÓN DE UN REPOSITORIO UNIFICADO

La inteligencia de negocios es una práctica que busca fortalecer la toma de decisiones a partir de la utilización de información relevante de alta calidad, y maximizar su valor a través del análisis, la definitividad y la interpretación.

En el contexto de los sistemas de información, la Inteligencia de negocios es una tecnología que facilita la recopilación y el procesamiento de datos para identificar correlaciones y patrones significativos para apoyar las decisiones dentro de una organización.

El Dataware House Institute describe que la inteligencia de negocios une datos, tecnología, análisis y conocimiento humano para optimizar las decisiones organizacionales y, en última instancia, impulsar el éxito y logro de metas en una organización. Los programas de inteligencia de negocios suelen combinar un almacén de datos de empresa y una plataforma de inteligencia de negocios o conjunto de

herramientas para transformar datos en información organizacional útil.

En este sentido, el Sistema de Inteligencia Institucional fue concebido como un proyecto que relaciona información entre el pasado, presente y futuro para la toma de decisiones estratégicas. Una institución de gobierno tan grande y relevante como el Instituto Nacional Electoral cuenta con áreas ejecutivas de administración (ejecutivos), funcionarios (personal operativo y de servicio), clientes o receptores de los servicios (ciudadanos, partidos políticos y sujetos regulados) y productos (servicios proporcionados; por ejemplo, la credencial para votar).

En esa labor diaria, el INE produce grandes cantidades de información que, como en la gran mayoría de instituciones y empresas, se encuentra dispersa en múltiples bases de datos independientes entre sí, por lo cual no es posible comparar el pasado con el presente fácilmente y en ese sentido la cualidad de los acervos no permite obtener información con valor analítico.

Por ello, el potencial del proyecto reside en relacionar la diversa información que el INE posee y ha generado en el pasado y sigue produciéndose en el presente, conciliarla y concentrarla en un único repositorio, relacionarla y permitir su procesamiento y análisis para la toma de decisiones estratégicas. Es decir, concebir un proyecto que permitiera entretener la diversa información con que cuenta el Instituto en sus bases de datos históricas y presentes para que puedan ser consultadas a partir de herramientas analíticas que doten a la institución de capacidades analíticas que más adelante permitan a la ciudadanía y otros sujetos la explotación de esas nuevas capacidades de análisis. Porque una forma de lograr la efectividad en la administración pública es utilizar tecnología que soporte y coadyuve en una mejor

toma de decisiones, y qué mejor que las decisiones se basen en información, hechos y datos existentes.

En resumen, las herramientas analíticas fortalecen y mejoran la toma de decisiones porque en la evolución de la sistematización de procesos el primer paso es el sistematizar procesos, y al hacerlo logramos sistematizar la operación diaria de una labor en particular, conseguimos que cada transacción quede registrada y con ello somos capaces de guardar la historia en la operación de un proceso operativo; con ello dotamos a las instituciones de sistemas operacionales que nos permiten conocer el ejercicio diario de sus acciones.

Pero eso no es suficiente para dotar a una institución de capacidades analíticas, la sistematización de información no es una labor suficiente para mejorar la toma de decisiones, la clave está en dar el siguiente paso, que es analizar la información; para ello es necesario contar con mecanismos y herramientas que nos permitan procesar múltiples fuentes de datos, conciliar la información histórica y presente de una organización, y asegurar que la obtención de respuestas a preguntas claves puedan ser viables. Para ello, la primera decisión era crear un repositorio único donde pudieran integrarse las múltiples bases de datos que posee el INE y generar un meta modelo, un modelo multidimensional, que permitiera cruzar y llevar a cabo diversas consultas de información para generar productos analíticos.

Al igual que en una gran biblioteca, esa labor fue necesaria para poder concentrar en una sola herramienta el acervo de información generado por las múltiples bases de datos que genera la institución en un primer ejercicio que le permita al INE fortalecer sus capacidades analíticas de la información que posee para que no sólo pueda saber “qué pasó” y “qué está pasando”, lo cual se logra en primer nivel implementando sistemas operacionales, sino poder

anticiparse al “qué puede pasar” y “qué debe planearse” con un enfoque estratégico y no solo operativo, con base en indicadores vinculados al negocio identificado en cada proceso sistematizado y vinculado.

IMPLEMENTACIÓN DEL PROYECTO

Un sistema de inteligencia de negocios está integrado de tres grandes componentes:

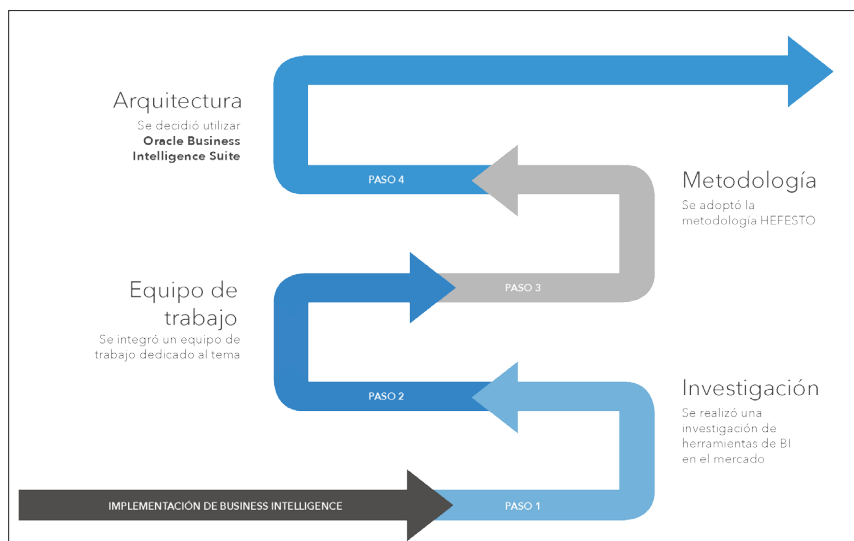
1. Un repositorio único o central (*data warehouse*), el cual sirve como una gran biblioteca centralizada de información donde se integran las múltiples fuentes de datos estructuradas (es decir bases de datos, información obtenida de archivos Excel, CSV u otras fuentes que tengan un estructura de orden previo). En este repositorio se diseñará un modelo multidimensional, el cual formará parte del gran metadato que describa las relaciones que vayan descubriendo al vincular cada una de las bases de datos que integren y cada uno de los datos de estas múltiples bases de datos.
2. Un componente analítico. Herramientas analíticas que llevan a cabo la extracción y carga de información al repositorio único que genera el modelo multidimensional y adicionalmente un modelo basado en cubos de información e indicadores a partir de Cubos OLAP (On Line Analytical Process) mediante el cual se pueden agilizar las consultas de grandes cantidades de información a partir de la estructura multidimensional y no sobre la tradicional estructura rela-

- cional de las bases de datos.
3. Un reporteador o visualizador. Herramienta orientada a usuarios de negocio para que exploren y analicen la información que necesitan de manera rápida y sencilla. Permite crear visualizaciones basadas en distintos conjuntos de datos y mostrar el resultado en la misma página, como si fuera una combinación entre respuestas a preguntas y tableros de control.

Para implementar un proyecto de inteligencia de negocios, fue necesario ejecutar los siguientes pasos:

A partir de la fase inicial del proyecto, se distinguieron las

Gráfico 1. Pasos previos
a la implementación de un proyecto de inteligencia de negocios



Fuente: Elaboración propia.

necesidades tecnológicas y del grupo de trabajo que llevaría a cabo la gestión del proyecto, y se resaltó la importancia de la ejecución de cada paso:

Paso 1. Es necesario realizar una investigación previa para conocer la situación actual en el mercado tecnológico sobre herramientas de inteligencia de negocio que puedan implementarse.

Paso 2. Integrar un equipo de trabajo es esencial, ya que un proyecto de estas características demanda amplios conocimientos en materia de análisis de datos, así como de administración de la infraestructura donde se ejecutarán estas herramientas.

Paso 3. Identificar metodologías y un marco de referencia permitirá tener una idea clara de ejercicios previos; para este proyecto se implementaron dos metodologías. La primera es HEFESTO, metodología creada por Bernabeu Ricardo Darío disponible bajo licencia GNU FDL, la cual se fundamenta en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. Consta de cuatro fases: análisis de requerimientos, análisis de las bases de datos orientadas al procesamiento de transacciones OLTP (On-Line Transactional Processing), modelo lógico del almacén de datos y proceso ETL (Extract-Transform-Load), que corresponde a las actividades de extraer los datos de las fuentes OLTP. Asimismo, se tomó como referencia a The Dataware Housing Institute (TDWI), organización que por más de veinte años se ha dedicado a la investigación y asesoría sobre inteligencia de

negocios (*business intelligence*), *data warehousing* y analíticos alrededor del mundo. El objetivo de investigación de este organismo se centra exclusivamente en diversas situaciones a las que se enfrentan líderes y ejecutantes para entregar un profundo y amplio entendimiento del negocio y los retos técnicos involucrados en la implementación y el uso de inteligencia de negocios, *data warehousing* y soluciones analíticas. El TDWI creó el Modelo de Madurez Analítico (Analitics Maturity Model) con el objetivo de ayudar a la necesidad que tienen las organizaciones de conocer de calidad de las implementaciones de soluciones analíticas en comparación con otras para identificar las mejores prácticas y basarse en ellas. Por lo tanto, este modelo sirve para determinar en qué nivel se encuentra una institución o empresa con respecto al grado de madurez de adopción de un modelo analítico al interior de la organización.

Paso 4. Definir la arquitectura tecnológica que se empleará para la explotación de información, el proceso completo de carga y la publicación de tableros y diversos análisis que puedan ser consultados, la cual involucre ambientes específicos de desarrollo, pruebas y producción.

A partir de este marco de referencia y tomando en consideración los pasos definidos por la metodología HEFESTO, se pueden iniciar los trabajos para integrar las múltiples fuentes de información y lograr la construcción de análisis de información identificados, para lo cual se ejecutarán las siguientes fases:

La primera es analizar e identificar los requerimientos analíticos y descubrir las preguntas clave a responder para así ubicar los indicadores y las perspectivas de análisis que será necesario mapear hacia las fuentes de datos origen y con ello lograr un modelo conceptual.

La segunda fase comprende el análisis de las fuentes de datos a cargarse en el repositorio único. Para determinar los indicadores a implementarse, se establecen las correspondencias de estos a cada uno de los campos de información que provienen de las fuentes de datos origen y se crean los niveles de granularidad; es decir, se ubican los campos que contendrá cada perspectiva, ya que a partir de ellos se examinarán y filtrarán los indicadores, lo que logrará un modelo ampliado.

La tercera fase corresponde a la generación de un modelo lógico del *dataware house*, en donde se diseñan las tablas de dimensiones, las tablas de hechos y se detallan las uniones; es decir, se genera el mapa lógico final que será creado en el repositorio único donde se construirá el modelo que albergue los datos provenientes de las diversas fuentes.

La cuarta fase corresponde a la creación de los procesos de carga y extracción de los datos origen para poblar el repositorio único tomando como base el modelo lógico recién creado en el repositorio.

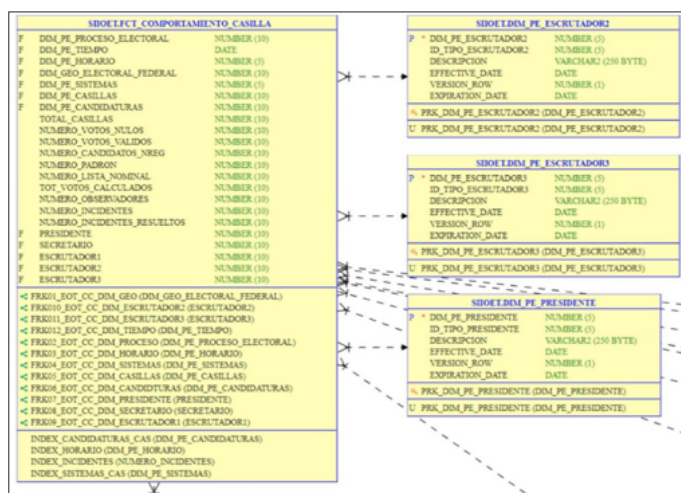
Este camino de cuatro fases deberá ejecutarse para cada una de las fuentes de información que pretenden colocarse en el repositorio único. Una labor esencial es realizar una correcta gestión del modelo multidimensional diseñado. Algo muy importante es ubicar indicadores que se identifiquen como datos maestros que son empleados por múltiples fuentes de datos y que gobiernan la relación entre ellos; por ejemplo, algunas dimensiones maestras que son comunes entre múltiples fuentes de datos son el tiempo, elementos

geográficos y catálogos genéricos que eran empleados por el modelo lógico para permitir las uniones entre múltiples fuentes de datos de manera homologada y bien definida.

Este trabajo permite contar con un modelo multidimensional capaz de relacionar múltiples fuentes de datos que ahora pueden ser consultadas a partir de una perspectiva de cubos de información en indicadores; es decir, se logró abstraer la complejidad de los modelos de datos operacionales y se implementó un modelo más afín al negocio, lo que resolvió así la complejidad de consultas y la dispersión de las fuentes de información.

A continuación, se presenta como ejemplo un fragmento del modelo multidimensional, en particular refiriendo a información obtenida de una fuente de datos histórica donde se registraron las cifras referentes al conteo de actas de escrutinio y cómputo de una elección:

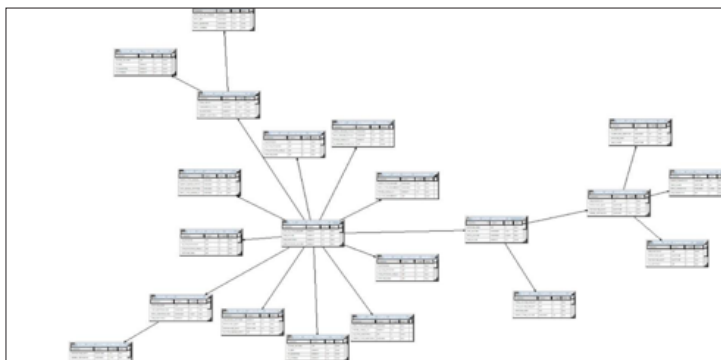
Gráfico 2. Fragmento del modelo multidimensional correspondiente al modelo de datos de información correspondiente al conteo de actas



Fuente: Elaboración propia.

Asimismo, se presenta una vista de una estrella contenida dentro de la constelación total del modelo multidimensional:

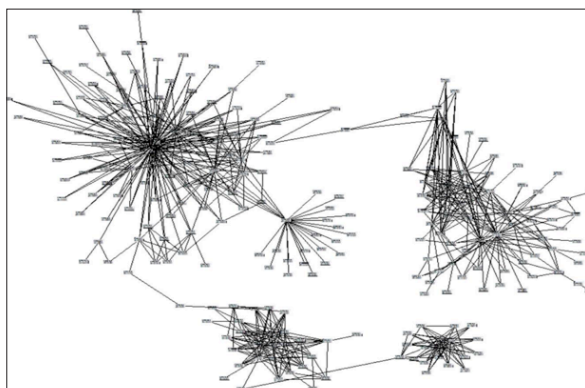
Gráfico 3. Fragmento del modelo multidimensional correspondiente al modelo de datos de información correspondiente al conteo de actas



Fuente: Elaboración propia.

Y finalmente, una visión del modelo físico completo solo como referencia de lo que se integra en un modelo multidimensional de un repositorio unificado.

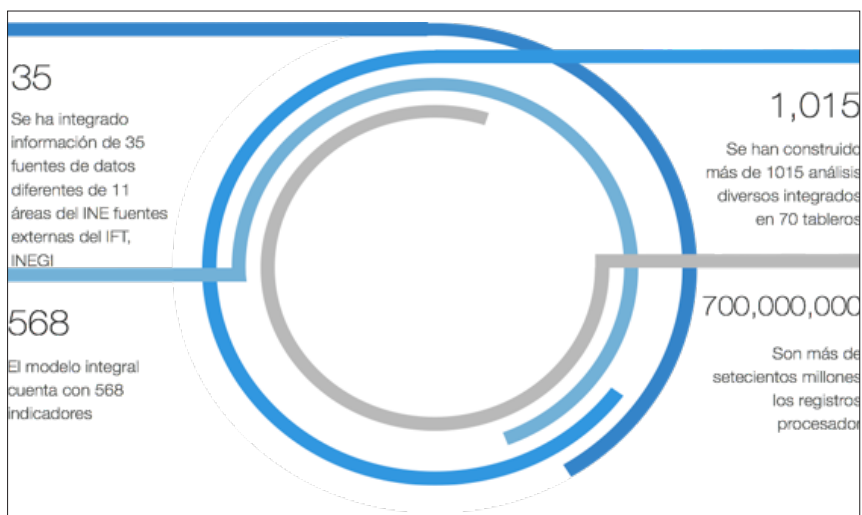
Gráfico 4. Fragmento del modelo físico completo del SII



Fuente: Elaboración propia.

Lo cual, en conjunto, engloba lo siguiente:

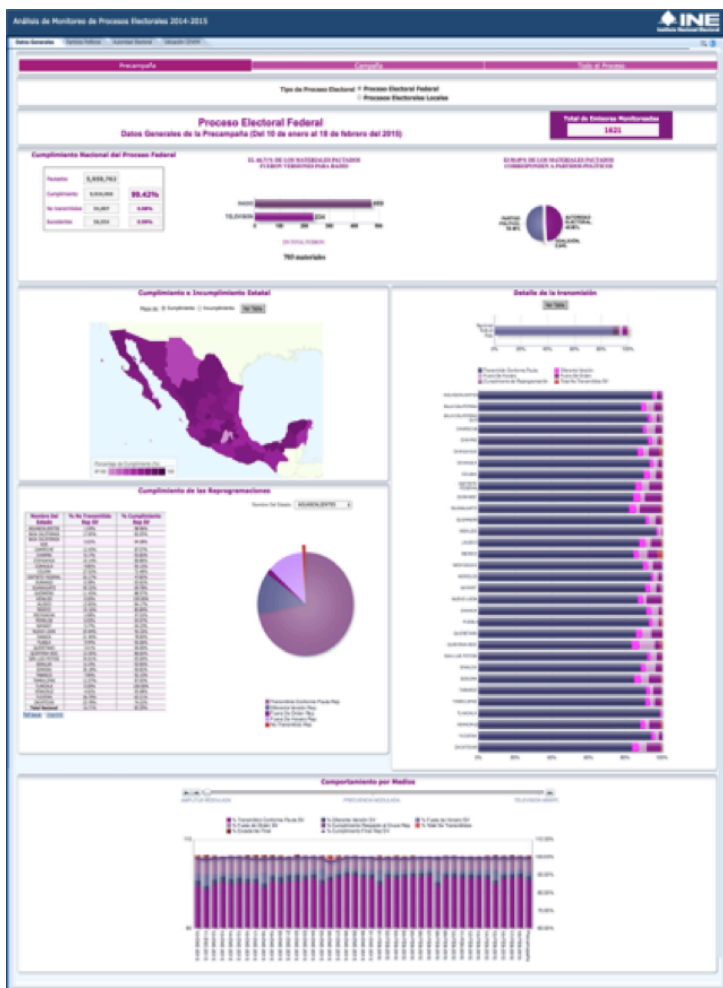
Gráfico 5. Información cargada al repositorio único del SII



Fuente: Elaboración propia.

Ejemplo de un análisis que puede ser generado desde el Sistema de Inteligencia Institucional:

Gráfico 6. Ejemplo de análisis sobre la transmisión de promocionales de radio y televisión para los partidos políticos en México



Fuente: Elaboración propia.

CONCLUSIONES:

El uso de herramientas analíticas cobra mayor relevancia en las organizaciones de gobierno y empresas privadas; estar informados no sólo implica acumular información diariamente, sino tener capacidades de explotar esa información de la mejor manera para la toma de decisiones.

Elegir una herramienta de análisis de información depende del tipo de aplicaciones y usos que se hará de la información; en particular para la inteligencia de negocios un pilar fundamental es contar con un repositorio unificado de información, que más adelante puede complementarse con herramientas de *big data* que permitan el análisis de información no estructurada. Sin embargo, en una organización de gobierno o empresa, mucha de la información que se encuentra en el inventario de los sistemas empleados corresponde a información estructurada, por ello implementar el uso de herramientas analíticas que exploten este tipo de acervo es sustancial para potenciar en análisis al interior de este tipo de organizaciones.

Es importante diferenciar entre una herramienta de inteligencia de negocios y herramientas de *big data*. La distinción se vuelve importante en el tipo de fuentes que buscan analizarse, ya sean estructuradas o no estructuradas, ya que cada una conlleva esfuerzos y objetivos distintos. El valor que hemos descubierto de trabajar en un repositorio unificado es que al integrar las fuentes estructuradas en un modelo multidimensional, se logra la madurez del modelo a largo plazo; es decir, todo el conocimiento que se ocupó para diseñar cada una de las bases de datos independientes y que se emplearon en procesos operativos reales dentro de la organización se han recuperado y fusionado en un modelo que permite hoy asegurar que ante cualquier

cambio futuro en las fuentes el modelo multidimensional será persistente y permitirá acumular conocimiento de diversos años de años previos, presentes y futuros permitiendo la convivencia de múltiples modelos de bases de datos en un modelo ya unificado y funcional, y con plena certeza en que los indicadores están plenamente validados y la información que producen es cierta y de calidad. *Big data* al contrario podría analizar y cruzar fuentes diversas, sean estructuradas y no estructuradas, y proponer hallazgos, identificar patrones que permitan concebir una fotografía mucho más clara de lo que esos datos representan; sin embargo, en una organización donde los procesos de negocio son modelados concebir un gran modelo multidimensional, será sin duda el mayor beneficio al reflejar la operación del negocio con un modelo confiable y bien estructurado.

Noam Chomsky decía en el prefacio de su libro *El conocimiento del lenguaje* que le intrigaban dos problemas referentes al conocimiento humano. El primero era explicar cómo conocemos tanto a partir de una experiencia tan limitada, cómo conocemos tanto teniendo en cuenta que los datos de los que disponemos son tan escasos (problema de Platón), y el segundo es el problema de explicar cómo conocemos tan poco considerando que disponemos de una experiencia tan amplia, cómo conocemos y comprendemos tan poco, a pesar de que disponemos de unos datos tan ricos (problema Orwell). Considero que hoy más que nunca, este segundo problema cobra aún mayor relevancia: el problema no es acumular información. sino lograr discernir dentro de ese gran cúmulo de datos los datos que nos otorguen valor, y es ahí cuando usar herramientas analíticas que nos permitan posicionar por encima de la capa de los datos podrán dotarnos de una inteligencia adelantada, afinada y certeza sobre lo que se debe discernir dentro del conjunto de datos no explorados.

BIBLIOGRAFÍA

- Bernabeu, Ricardo Darío y Mariano García Mattío. 2017. "Hefesto: Data Warehousing: Guía completa de aplicación teórico práctica; metodología Data Warehouse v 3.0". Disponible en https://sourceforge.net/projects/bi-hefesto/?source=typ_redirect.
- Breiman, Leo. 2001. "*Statistical Modeling: The two cultures*".
- Chambers, John M. 1993. "Greater or lesser statistics: a choice for future research. *Statistics and Computing*", vol. 3, núm. 4:182-184.
- Cleveland, William S. 2001. "Data Science: an action plan for expanding the technical areas of the field of statistics", *International statistical review*, vol. 69, núm. 1:21-26.
- Chomsky, Noam. 1985. *El conocimiento del lenguaje*. Madrid: Atalaya.
- Donoho, David. 2017. "50 years of Data Science", *Journal of Computational and Graphical Statistics*. <https://doi.org/10.1080/10618600.2017.1384734>
- Halper, Fern y David Stodder. 2014. "TDWI Analytics Maturity Model". The Data Warehouse Institute. Disponible en <https://tdwi.org/whitepapers/2014/10/tdwi-analytics-maturity-model-guide.aspx>.
- Kelleher y Tierney. 2018. *Data Science*. The MIT Press Essential Knowledge Series: MIT.

Calidad en los metadatos: reto para asegurar la participación en repositorios nacionales manteniendo los estándares institucionales

ALMA BEATRIZ RIVERA-AGUILERA

ELISA CRUZ ROJAS

BENJAMÍN A. GUERRERO OLVERA

ÓSCAR ALFONZO PEREYRA

Universidad Iberoamericana Ciudad de México

INTRODUCCIÓN

La consolidación de un Repositorio Institucional (RI) involucra factores como gestión técnica y financiera, renovación de plataformas tecnológicas, nuevas estrategias de poblamiento; normalización y adecuación de metadatos a los lineamientos del Repositorio Nacional (RN) para asegurar la cosecha, consideración de diversos estándares, mantenimiento de la calidad de los registros del RI, respeto a los derechos de autor y mejora en estrategias de preservación. Compartir experiencias es de mucho valor cuando se abordan proyectos innovadores como suelen ser los relacionados con RI y, en especial, cuando los metadatos de los RI serán cosechados por un repositorio más amplio.

Llevar a cabo una reflexión sistemática de lo que ha sucedido en un proyecto suele quedarse en una buena intención, pues el proyecto mismo y otras actividades consumen

el tiempo para un escrito colectivo, como el que aquí se comparte. En este texto, revisaremos los antecedentes de la organización y publicación de colecciones digitales propias en nuestra biblioteca. Abordaremos el proyecto que posibilitó la experiencia descrita, el poblamiento del RI a ser consolidado, los lineamientos en relación a los metadatos en RI para ser cosechados por el RN, la normalización que algunas etiquetas han requerido y los resultados de las pruebas de cosecha.

ANTECEDENTES

En 1998 se iniciaron en la Biblioteca Francisco Xavier Clavigero los esfuerzos para la creación de una biblioteca digital, tarea que continuaba la tradición de uso de tecnología para ofrecer mejores servicios de información que arrancó en la biblioteca en los años setenta, tal como señala en detalle Rivera-Aguilera (2010).

En el año 2004, se inauguró la primera colección digital en texto completo, que de forma sistemática e integral (metadatos y texto completo) explotaba la información digital. Esta colección fue la de tesis digitales, la cual estuvo basada en estándares de metadatos descriptivos para tesis ETDMS (Electronic Thesis and Dissertation Metadata Standard) y DC (Dublin Core) y utilizó una combinación de software comercial (Aleph) y software libre (Harvest). La colección de tesis digitales de la Ibero se conformó originalmente en un sitio en HTML que ligaba en una estructura jerárquica organizada por departamentos académicos los archivos PDF de las tesis almacenadas en un servidor Linux. Los trabajos

de titulación podían recuperarse, también, a través del catálogo en línea en Aleph 500 vía la etiqueta 856 del formato MARC. Una vez puesta al servicio de los usuarios, la aplicación fue evaluada, desde la perspectiva de usabilidad, y encontrada satisfactoria por éstos. La explotación de los textos completos de los archivos PDF se hacía hasta el 2007 a través del buscador Harvest.

En 2008, se consideró ampliar —no sustituir— el concepto de biblioteca digital al de repositorio institucional. En el caso de la BFXC de la Ibero México, la concepción incluye desde entonces, además de las colecciones bibliográficas, la producción docente de investigación, documentos institucionales y de vinculación social en un solo interfaz de búsqueda. El administrador de colecciones digitales, seleccionado para implementar esta nueva etapa, fue Greenstone, el cual de forma integral almacena y explota los metadatos, y el texto completo de los objetos digitales textuales. Greenstone también administra con eficiencia las imágenes digitales y genera estructuras jerárquicas de acceso vía HTML dinámico. Este modelo implicaba, para las tesis digitales, una transferencia de metadatos desde MARC en Aleph a DC en Greenstone. Para las colecciones de comunicaciones oficiales, patrimonio natural y cultural (de uso local); producción científica (en pruebas) y material didáctico (en pruebas), una descripción basada en DC, Learning Object Metadata (LOM) y algunas etiquetas locales. Nuestra colección de tesis en Greenstone ha sido cosechada por REMERI. Sin embargo, dada la poca dinámica en nuevas versiones en Greenstone, la falta de integración de interfaces de autoguardado —adecuados para usuario final— y la escasez de usuarios de Greenstone en México, se consideró que el paso a DSpace fortalecía a nuestra biblioteca digital/RI.

EL PROYECTO

El Consejo Nacional de Ciencia y Tecnología (CONACYT) ha abierto, al menos en cuatro ocasiones (2014-2017), convocatorias para financiar proyectos de inicio o consolidación de RI. Obtener fondos en convocatorias evaluadas ayuda a las unidades de información a adquirir recursos que en la operación normal no estarían disponibles y a poner a la biblioteca, como un actor central en la innovación tecnológica institucional. Otro beneficio de participar es promover las políticas de acceso abierto en las universidades y posicionar a los bibliotecarios como pares académicos. Por ello, se decidió participar en la convocatoria 2016 y se obtuvo financiamiento para nuestro proyecto de consolidación en 2017.

La experiencia que se comparte en este capítulo se enmarca en el desarrollo del proyecto antes mencionado. Éste implicó una serie de tareas simultáneas en ciertas medida y necesariamente coordinadas:

1. Consolidar nuestro repositorio de tesis existente, diseñar y ejecutar procedimientos de acopio de publicaciones académicas de acceso abierto, generadas por comunidad universitaria desde diferentes plataformas abiertas y sus correspondientes metadatos.
2. Definir una estrategia de verificación de contenidos en acceso abierto y de obtención de permisos por parte de los autores.
3. Transferir archivos y metadatos disponibles, y adecuarlos a los lineamientos del repositorio nacional.

4. Llevar a cabo un ejercicio crítico y propositivo en relación con la normalización planteada por los lineamientos del repositorio nacional.
5. Obtener recursos de *hardware*, instalar adecuadamente nueva versión de *software* gestor de colecciones digitales, transferir metadatos y contenidos, e integrar una nueva forma de almacenamiento de parte de los recursos en formatos que aseguren la preservación de textos a largo plazo (XML).
6. Gestionar el equipo interno y la vinculación grupal con consultores o proveedores para lograr la concreción de un proyecto en un plazo de tiempo muy limitado.

En este texto, nos vamos a enfocar en el rol de los metadatos en un proyecto que pretende compartirlos desde un repositorio institucional a uno nacional y los enormes retos que implica adecuarse a unos lineamientos nacionales manteniendo los estándares calidad institucionales.

LINEAMIENTOS PARA COSECHAR METADATOS DE RI AL RN

El RN, administrado por Conacyt, brinda la normalización y los lineamientos para el ingreso de los metadatos que serán cosechados de los RI en su manual de Lineamientos Específicos para Repositorios (LER), se encuentran basados en estándares internacionales, tales como el protocolo OAI-PMH para la parte de interoperabilidad entre repositorios y los campos de OpenAIRE para la definición de las etiquetas y contenidos de los metadatos a ingresar en el RN.

Los lineamientos proponen veintisiete campos de metadatos DC: son nueve campos de carácter obligatorio, nueve campos obligatorios cuando aplique y otros nueve campos solo como recomendados.

Estos lineamientos brindados por CONACYT, en conjunto con la sintaxis necesaria para cada elemento, generan un “manual” sencillo y básico para la construcción del repositorio. El RN también ofrece catálogos de lenguajes o términos controlados a consultar para los campos que lo requieran para una mayor normalización de los repositorios.

Sin embargo, ya en la práctica del asentamiento de metadatos para la conformación del repositorio, nos encontramos con ciertas dificultades; por ejemplo, el interface rústico del catálogo de autores. Este catálogo, si bien es una gran herramienta para la normalización, no ofrece una búsqueda de autores de manera ágil ni fácil, ya que es necesario probar con variaciones la búsqueda de un nombre, pues al asentar un creador existen diferentes maneras en las que se separan los apellidos. En algunos casos, se utilizan guiones bajos para separar el primer del segundo apellido, lo que genera caracteres con los que se deben de realizar múltiples búsquedas para obtener el nombre que se esté buscando y, en algunos casos, los registros de los autores no cuentan con los indicadores que el mismo Conacyt pide. Bajo esta misma línea, no existe una normalización en el uso de mayúsculas y minúsculas, ya que existen casos de los tres tipos: únicamente con los usos de letras mayúsculas, con el uso de solamente letras minúsculas y el uso de ambas.

El otro problema más complejo es la obtención de los identificadores de los creadores que ya no tienen relación con la institución. Si bien integrar los identificadores parece una manera eficiente de ligar la información y ser

un principio de un control de autoridades de personas en el ámbito de ciencia y tecnología, todavía se muestra un esfuerzo muy precario y poco coordinado con esfuerzos internacionales como NACO o VIAF. En la práctica, obtener los identificadores Clave Única de Registro de Población en México (CURP), Número de Curriculum Vitae Único asignado por Conacyt (CVU), *Open Researcher and Contributor* ID (ORCID) y Documento Nacional de Identidad de cualquier país (DNI) de los autores de productos de investigación particulares, se vuelve complejo. Esto debido a que los materiales con los que se va a colaborar dentro del repositorio nacional son tesis, pues la mayoría pertenecen a autores que son estudiantes sin un registro para obtener un CVU y, si bien se puede asentar utilizando el CURP, se fuerza a las instituciones a revisar sus compromisos en materia de privacidad. Otro reto es la múltiple autoría y, más específico, la autoría en conjunto con investigadores de otras nacionalidades en las que en algunos casos no se cuenta con información en primera mano sobre algún número identificador para su registro.

Por cuestión de los tiempos en los que se deben entregar resultados del repositorio —tomando en cuenta la experiencia previa con la que ya se contaba en la Biblioteca Francisco Xavier Clavigero (BFXC) en torno a la administración y asignación de metadatos, posterior a un mapeo—, se tomó la decisión de reducir los campos Dublin Core (DC) a doce elementos, los cuales fueron elegidos por su importancia dentro de la descripción y por ser los campos con los que se facilitaba la cosecha de los datos, y considerando la experiencia y las recomendaciones de los consultores externos que brindaron su apoyo.

POBLAMIENTO DEL REPOSITORIO CONSOLIDADO

Tesis: la transferencia de metadatos desde Aleph.

La transferencia de registros bibliográficos de tesis se presentó en dos fases: de nuestro catálogo bibliográfico hacia el RI y del RI al RN. Se preparó el mapeo de los campos MARC21 a DC para el paso al RI y se adecuó paso de los metadatos DC a LER en los registros de transferencia basados en el protocolo OAI para ser cosechados en el RN a fin de lograr la interoperabilidad entre diversos repositorios (Roy y Chandra 2017,74). De esta manera, los contenidos de las tesis tienen mayor visibilidad a nivel nacional e internacional al participar en el repositorio nacional CONACYT. La preparación de los campos MARC21 utilizados en la descripción de las tesis digitales de posgrado se llevó a cabo de manera detallada por medio de un mapeo con una correlación exacta de los contenidos requeridos en los Lineamientos Específicos para Repositorios (LER), para lo cual se realizaron varias versiones y se fueron afinando los detalles. De manera simultánea, se llevó a cabo un barrido de control de calidad de los registros de tesis en ALEPH con la intención de reducir el margen de error para lograr una transferencia lo más limpia posible.

Cabe mencionar la importancia de los metadatos integrados y fusionados de diversos esquemas adaptados a las necesidades para obtener un perfil de aplicación específico en este proyecto. Este perfil se ve reflejado en la transferencia de las tesis a la versión DSpace 6.0 y, así también, en la cosecha hacia el repositorio CONACYT. Recordemos que los metadatos apoyan la construcción de servicios coherentes para los usuarios a partir de componentes que son técnicamente distintas y administrados por diferentes organizaciones (Potvin 2016, 100).

El proyecto permitió examinar las divergencias que existen entre lo dicho por los lineamientos y la realidad. Nuestro ejercicio demostró la importancia de desarrollar conjuntos de metadatos eficientes en la conformación de repositorios colectivos con la finalidad de mantener la interoperabilidad, autenticidad y confiabilidad. Tal como lo señala la literatura especializada, los conjuntos de esquemas o perfiles de aplicación deben de ser capaces de aplicarse a diversas instituciones con la finalidad de obtener recolecciones y cosechas a través de protocolos de acceso

abierto. Los metadatos enfatizan la base para la recuperación eficiente entre recursos, así como, dando la debida importancia a descriptivos, técnicos, administrativos y de preservación a través del tiempo y las tecnologías (Potvin 2016, 101).

Para lograr el objetivo de la transferencia en dos niveles, es decir, a DSpace RI y Repositorio CONACYT, se desarrolló un archivo de Excel con los campos requeridos para ambos casos. Aproximadamente en novecientos registros de tesis se enriquecieron los campos y se agregaron los requeridos por LER. Algunos de estos campos fueron los autores y directores con los identificadores requeridos por Conacyt y el área de conocimiento con su identificador, tipo de documento, es decir, “dc.contributor.author”, “dc.creator”, “dc.contributor.director”, “dc.type”, entre otros.

Artículos, ponencias, capítulos de libros
y libros: identificación de metadatos en Internet.

La recuperación de documentos requirió una búsqueda en las páginas especializadas en investigación científica. La plataforma que nos dio la mayor cantidad de documentos subidos por los propios investigadores fue Researchgate.net,

en la que encontramos veintisiete perfiles de investigadores de la universidad con bastante producción; las cuentas pertenecían tanto a profesores miembros del Sistema Nacional de Investigadores, como a profesores de tiempo. También buscamos documentos en Academia.edu, en los que dimos con cuarenta y siete perfiles de investigadores, pero muchos de ellos sin producción reportada en esa plataforma. Es importante recalcar que al momento de descargar los archivos, nos encontramos con varios errores y no había una forma sencilla de recuperar los metadatos. Más tarde, observamos que este error provenía del tamaño de los archivos subidos a Researchgate.

El siguiente reto fue observar los derechos de autor de muchos de los documentos descargados. Diversas casas editoriales tienen reglas para la distribución de los artículos publicados por ellos. Para entender las reglas de embargo y los tiempos en los cuales es permitido distribuir un artículo, recurrimos a SHERPA/RoMEO, página en la que se pueden consultar las reglas de *copyright* según la editorial. Esto nos llevó a tener que eliminar hasta un 15 por ciento de los documentos encontrados. Para poder compensar esta pérdida, se espera enviar emails a los investigadores invitándolos a enviar una versión *preprint* de la cual tiene permiso de distribución.

Durante la recuperación de archivos en Google Académico, también tuvimos algunos problemas en la descarga. En ocasiones, un mismo artículo se encontraba en diversas páginas y el nombre podía llegar a ser ligeramente diferente o el autor tenía un nombre homónimo que la del investigador de la universidad. Gracias a la página de Google, nos dimos cuenta de que había más investigadores en la plataforma de Researchgate; el problema al momento de buscarlos fue que sus nombres o apellidos no estaban escritos por completo en el perfil del investigador.

Una vez que localizamos más investigadores en las diversas plataformas con un total de mil recursos de ciencia y tecnología que era la meta, nos dimos a la tarea de almacenar los datos en una tabla de Excel que sirviera como base para identificar la cantidad de documentos encontrados, en qué plataforma y qué tipo de material se trataba. A este archivo de control, se le fueron agregando datos que nos permitieron llevar un seguimiento del avance del proyecto. El archivo también sirve de referencia para saber qué profesores cuentan con material de investigación, la cantidad y la fuente de estos documentos.

Otro aspecto importante a reportar fue la eliminación de documentos que no contaban con la suficiente calidad para ser subidos a los repositorios. Esto debido a que los documentos estaban mal escaneados, era solamente la portada del libro en donde presentaron un artículo o simplemente no contaban con los suficientes metadatos. En varias ocasiones, se requirió una nueva búsqueda del mismo documento a fin de encontrar metadatos faltantes o rectificar si el documento era del autor. Cada área disciplinar supuso un reto debido a que en las ciencias naturales es común participar con una gran cantidad de autores, mientras que, por otro lado, en ciencias sociales se puede contener tanto una como varias autorías.

Una vez recuperados los documentos, el siguiente objetivo fue identificar la mejor forma de abordar la captura de metadatos. En ánimos de acelerar el proceso, probamos extraer los datos de forma automática mediante el uso del software Zotero. Gracias a la aplicación, logramos extraer los datos de varios documentos descargados desde las páginas académicas que nos permitieron la libre obtención de los datos. El problema que nos causó este método fue que nos arrojaba demasiados datos que no eran necesariamente solicitados

por los lineamientos del CONACYT, además de generar una búsqueda extra de cada archivo a fin de conseguir los metadatos de la página. Lo anterior nos llevó a descartar este método y, en su lugar, optamos por capturar los datos de forma manual.

Mediante el listado de control en Excel, logramos ubicar a los investigadores con la mayor cantidad de documentos recopilados, por lo que se empezó la captura de los metadatos de acuerdo con este orden. Tuvimos la oportunidad de revisar qué profesores aún se encontraban laborando en la universidad y quiénes se acaban de integrar al Sistema Nacional de Investigadores (SNI). También integramos los correos electrónicos de éstos y otros datos que nos ayudaron a generar los archivos con los metadatos.

Durante el trabajo de la captura de metadatos, generamos dos archivos de Excel. En uno de estos archivos, se introdujeron todos los datos de importancia sobre el investigador y el archivo. Al mismo tiempo de generar metadatos para este primer archivo, revisamos por segunda vez los documentos para verificar si contaba con alguna restricción y si el archivo presentaba la calidad suficiente para estar en el repositorio. También, se realizó una modificación a los nombres de archivos recuperados. Se utilizaron las iniciales del investigador, más un identificador que permitiera saber si el archivo era un artículo, libro, conferencia o datos de investigación. Para hacer la modificación de archivos en grupo, utilizamos el *software* de Advanced Renamer, gracias al cual pudimos acelerar los cambios de nombre.

Gracias a todos estos procesos generados durante el proyecto, automatizamos la futura inserción de nuevos archivos en el repositorio que encontramos interesantes sobre el estado de investigación de la universidad, las áreas que producen más y cuáles son los investigadores más prolíficos.

ADECUACIÓN A LOS LINEAMIENTOS LER PARA EL CASO DE LOS AUTORES DE TESIS

Como ya se comentó en el apartado de lineamientos para cosechar metadatos, el principal reto fue la normalización de los nombres de autores, directores de tesis y su localización dentro del catálogo CONACYT. El proceso de búsqueda en dicho catálogo tomó cierto tiempo en la curva de aprendizaje. Nuestra base de tesis, por ejemplo, era de novecientos registros, entonces eran aproximadamente mil ochocientos o más nombres que normalizar y agregar identificadores en un lapso de tiempo de dos meses.

Así pues, para adecuar los nombres e identificadores en el archivo Excel conformado a partir de los registros bibliográficos de tesis en Aleph, se llevaron varias acciones. En una primera etapa se solicitó el apoyo de ocho catalogadores para la investigación de autores e identificadores y el enriquecimiento del archivo de Excel. Esto fue posible revisando y transcribiendo información de manera simultánea de los registros bibliográficos en ALEPH y catálogo CONACYT hacia el archivo Excel. Para esto, se elaboraron procedimientos, flujos de trabajo, controles de calidad; entre otros, para optimizar el tiempo y las actividades de los catalogadores.

En una segunda etapa del proceso, se identificaron los autores no localizados en el catálogo de autores CONACYT y se solicitó que se dieran de alta. Siguiendo los lineamientos establecidos, se llenaron otros archivos de Excel y se subieron a través de la interfaz del portal de Conacyt de alta de autores. Posteriormente, se fue completando el archivo de Excel con los campos LER requeridos.

Hay que destacar la importancia de implementar un estándar común nacional, en este caso, aplicado a las tesis

electrónicas, lo cual ha favorecido la visibilidad y cooperación entre diversas instituciones hacia la conformación de un repositorio nacional. Por otro lado, ha sido una oportunidad de reevaluar los desacuerdos entre lo ideal y la realidad en la aplicación de los esquemas de metadatos en la descripción, preservación, herramientas y transferencias de las tesis electrónicas; asimismo, su impacto en los sistemas, implementación de mejores servicios y difusión de los contenidos en acceso abierto.

RESULTADOS PRUEBAS DE COSECHA

Se llevaron a cabo siete pruebas de cosecha de metadatos, en su mayoría tesis y artículos, retroalimentadas por el área informática y bibliotecológica del RN. Los resultados de las pruebas iniciales reportaron dos tipos de problemas:

1. Autores
 - a. Falta de un identificador válido para los autores en al menos uno de ellos. Este tema se ha desarrollado ampliamente en los apartados referidos a los LER. Si bien conocíamos de este requerimiento, al momento de las pruebas de cosecha nos dimos cuenta de que para el caso de las tesis no sólo se refería a que al menos un autor tuviese CVU, CURP, etcétera, sino que todos los autores o colaboradores de un recurso (por ejemplo, para tesis dc.contributor con datos de director o sinodales) que no tengan un identificador único, deberían tener un identificador de nombre llamado RN que se considera un Registro de Autor CONACYT, para lo cual hay que dar de alta a todos los *contri-*

butor. Nótese que en el caso de un registro que solamente tenga RN, este será rechazado si no cuenta con algún dc.creator o dc.contributor que posea identificador del tipo CVU, CURP, ORCID o DNI. Con respecto a este punto, cabe mencionar que la política de CONACYT sobre identificadores únicos tuvo algunas modificaciones que facilitaron la cosecha durante el transcurso del proyecto; sin embargo, hay una oportunidad de mejora para los facilitadores del RN en la redacción de las políticas y ejemplificación de la captura de metadatos en relación a autores, y en general en relación a todos los metadatos LER.

- b. Inconsistencias en el catálogo de autores CONACYT. Por ejemplo, dos registros de un mismo autor, asentados de forma distinta y que a pesar de tener el mismo CVU eran rechazados.

2. El tipo de material para los estudios de caso debió ser ajustado, pues el catálogo de captura indicaba que se podía usar **dc.type** con el valor **masterDegreeWork**, pero nos rechazaba esos trabajos de titulación. El reporte de la cosecha nos indicó que necesitamos una repetición de este metadato previa con el valor **other**. Este mismo metadato seguía dando problema hasta que se detectó que las listas de valores válidos se albergaban en repositorios semánticos diferentes.

`<dc:type>info:eu-repo/semantics/other</dc:type>`

`<dc:type>info:mx-repo/semantics/masterDegreeWork</dc:type>`

En la séptima prueba, finalmente, se indicó que se podía solicitar la cosecha oficial.

CONCLUSIONES Y PROSPECTIVA

La experiencia del proyecto aquí reportado nos ha dejado muchas enseñanzas y retos a futuro, pues los nueve meses de duraición nos dejan la satisfacción de ver el repositorio consolidado con una nueva plataforma e interoperando con otros repositorios como el RN y el repositorio AUSJAL (<http://repositorio.ausjal.org/>). Sin embargo, quedan todavía muchas tareas sobre todo las relacionadas con afinar las políticas de uso del Dublin Core y perfiles de OAI, enriquecer los metadatos y asegurar el uso correcto de las etiquetas para mantenernos en los estándares y al mismo tiempo responder a los requerimientos de algunos cosechadores como el RN.

BIBLIOGRAFÍA

- Potvin, Sarah y Santi Thompson. "An Analysis of Evolving Metadata Influences, Standards, and Practices in Electronic Theses and Dissertations", *Library Resources & Technical Services* vol. 60, núm. 2: 99–114. Disponible el 29 de julio de 2018 en <https://journals.ala.org/index.php/lrts/article/view/5963/7585>.
- Rivera-Aguilera, Alma Beatriz. 2010. "Perfiles de aplicación en Greenstone: integración de Diversos Esquemas de Metadatos Para La Catalogación de Las Colecciones Digitales de La Biblioteca Francisco Xavier Clavigero." En *IV Encuentro de Catalogación y Metadatos. Memoria 4 Al 6 de noviembre de 2009*, 215–32. México: UNAM/Centro de Investigaciones Bibliotecológicas. Disponible en <http://132.248.242.6/~publica/conmutarl.php?arch=3&idx=247>.
- Rivera Aguilera *et al.* 2011. "Propuesta de Un Perfil de Aplicación Para Material Didáctico Digital Basado En El Marco de Singapur." En *VI Encuentro de Catalogación y Metadatos*, 100–113. México: UNAM/Centro de Investigaciones Bibliotecológicas.
- Rivera Aguilera, Alma Beatriz y José Juan Téllez Bertadillo. 2011a. "Elementos Contextuales, Técnicos y de Diseño Que Posibilitan El Acceso Abierto a Colecciones Digitales Universitarias." En *Coloquio Acceso Abierto a La Información en las Bibliotecas Académicas de América Latina y El Caribe*, Mesa Redonda Acceso En Las Instituciones Latinoamericanas, 309–334. México: UNAM/Centro de Investigaciones Bibliotecológicas.

- Rivera-Aguilera, Alma Beatriz *et al.* 2011b. “Conjunto Descriptivo de Un Perfil de Aplicación En Lo Referente a Creadores de Material Didáctico Digital En Ambientes de Biblioteca 2.0.” En *5o Encuentro de Catalogación y Metadatos*, 321–47. México: UNAM/Centro de Investigaciones Bibliotecológicas. <http://132.248.242.6/~publica/conmutarl.php?arch=1&idx=258>
- Roy, Bijan Kumar; Subal Chandra Biswas y Parthasarathi Mukhopadhyay. 2017. “Designing Metadata Harvesting Framework for OAI-Based LIS Repositories: A Prototype”, *International Journal of Information Science and Management* vol. 15, núm. 1: 73–88. <https://ijism.ricest.ac.ir/index.php/ijism/article/view/938/294>

Interoperabilidad en plataformas de publicación y distribución de libros digitales

JENNY TERESITA GUERRA GONZÁLEZ
Universidad Nacional Autónoma de México

INTRODUCCIÓN

En el sector del libro digital, la interoperabilidad es una condición escasamente consolidada. Su poca viabilidad en el corto plazo es consecuencia de aspectos tales como la coexistencia de múltiples formatos de producción, distribución y comercialización del *ebook*; el inadecuado uso de esquemas de metadatos para la descripción, y la exigua factibilidad de la interoperabilidad de índole pragmática en las plataformas.

Partiendo de este escenario, en las siguientes páginas se analizan las condiciones evolutivas del libro digital primero en cuestión de formatos (PDF, Epub, Mobi) y posteriormente de estándares (International Digital Publishing Forum, World Wide Web Consortium (W3C)); las características y funcionalidades de las plataformas de publicación y distribución; la operación de la interoperabilidad en las plataformas mencionadas, así como los retos a enfrentar para mejorar aquella.

EL LIBRO DIGITAL: DE LOS FORMATOS A LOS ESTÁNDARES

El nacimiento del libro digital está anclado a la puesta en marcha del Proyecto Gutenberg en 1971. Diseñado con el propósito de disponer libros en dominio público por vía electrónica, Michael Hart, su creador, sentó las bases de un fenómeno que se ha tornado evolutivo y que si bien no ha logrado consolidarse del todo, ha contribuido a ampliar las posibilidades de ubicuidad informativa y documental del libro como producto/objeto y del campo de las publicaciones en general.

Los años noventa del siglo pasado marcaron el punto de quiebre en la primera fase de perfeccionamiento del libro digital. Aparejado con un mayor desarrollo y la masificación en el uso de la World Wide Web, este momento coyuntural viabiliza la aparición de *Amazon* como la primera gran librería en línea (1995); los grandes grupos editoriales empiezan a experimentar con la venta de colecciones de libros y enciclopedias en disco compacto (1996); un número mayor de bibliotecas brinda servicios remotos por medio de páginas web y algunas otras empiezan a crear sus propias bibliotecas digitales (Lebert 2009).

A la par del rápido crecimiento y la robustez de la *web*, los avances tecnológicos apuestan por innovar en materia de dispositivos móviles e inalámbricos –lector Rocket (1996), EveryBook, Hanlin (2000)– para la lectura de *ebooks*. Esos primeros ebooks o libros digitales eran textos en formato HTML, .DOC, Open eBook –de la asociación *Autoring Group*– y en el mejor de los casos, versiones digitalizadas en PDF, presentado por Adobe en 1994 (Armañanzas Sodupe 2013, 19).

En esta segunda fase evolutiva del libro digital, multinacionales del corte de Amazon, Apple, Google y Microsoft

cooptan una buena parte del negocio editorial a través de la digitalización de importantes fondos bibliográficos (Proyecto Google Print, posteriormente Google Books), así como del diseño de dispositivos de lectura propia y formatos únicamente compatibles con ellos y su *software*.

Si bien la innovación del libro digital seguía en buena medida una ruta comercial, quedaba claro que adolecía de orden y formalidad. En el afán de subsanar estas carencias y definir una serie de estándares, en 1998 se creó el International Digital Publishing Forum (IDPF), organización de comercio y normas para la industria editorial digital. El IDPF desarrolló el *epub* (acrónimo de Electronic Publication) lanzado en 2007 con base en OPS (Open Publication Structure), OPF (Open Packaging Format) y OCF (Open Container Format).

El *epub* se diseñó como un formato de archivo capaz de manejar el contenido de manera independiente a su forma, lo que permitía que el documento se adaptara a cualquier dispositivo de lectura automática y rápidamente.

Epub emplea codificación Unicode; trabaja como formato dinámico y expandible que admite texto, imágenes y vídeos que pueden incluirse en cualquier lugar. Opera como un medio para representar, empaquetar y codificar contenido web estructurado y semánticamente mejorado, incluidos HTML, CSS, SVG y otros recursos para su distribución en un contenedor de un solo archivo: ZIP; tiene soporte para la gestión de derechos (DRM). Empero, *epub* no es uno solo, sino que inició y mejoró su funcionamiento mediante el desarrollo de versiones con novedades y cambios sustanciales entre una y otra (versiones por número) y distintas adiciones o ajustes menores (versiones de parche). *Epub* 2.0 fue la primera de estas versiones.

Tras las mejoras alcanzadas por *Epub* 2.0.1, maquetación reajustable o fluida, en 2011 llegó *Epub* 3.0, que se basa en

HTML5, formato en código abierto que está especialmente adaptado a la creación de libros digitales enriquecidos ya que dispone de comandos que permiten incluir vídeo y audio.

Esta versión implementó además un mejor soporte de textos, mejor accesibilidad, metadatos enriquecidos, control de pantallas táctiles, soporte de ecuaciones matemáticas, manejo de caracteres no latinos (particularmente ideogramas asiáticos o escritura árabe) y se alinea con el estándar DAISY (Digital Accessible Information System) de presentación multimedia en la web compatible con todos los navegadores (Adell 2016; García Jiménez 2014).

Epub 3.0 fue adoptado en marzo de 2013 como estándar de la International Publisher Association gracias a que aporta elementos propios para la creación de la web semántica, lo que facilita y flexibiliza la indización de contenidos.

En junio de 2014 apareció *Epub* 3.0.1, que incorporó especificaciones de diseño fijo; para enero de 2017 quedó listo *Epub* 3.1, cuyo objetivo era consolidar el formato en el entorno de las plataformas *web* haciendo que las características fueran más simples y fáciles de leer. Ésta y las versiones precedentes de *Epub* fueron desarrolladas por el IPDF, organización que se fusionó con el W3C a comienzos de 2017.

Epub 3.2 es la última versión —en proceso de construcción— del formato. Su desarrollo ha sido pausado debido a que editores y otros proveedores de contenido continúan utilizando *Epub* 2 mientras que *Epub* 3 solo es implementado parcialmente por los proveedores de sistemas y ofrece diversas características que apenas necesita una publicación.

A ese respecto, en 2013 la European & International Booksellers Federation (EIBF) recomendó el uso de *Epub* 3 como formato/estándar interoperable (abierto) aunque con poco éxito. La iniciativa se vio condicionada por factores ajenos al ecosistema *Epub*; entre ellos, la diversidad de formatos

que convergen en torno al libro digital y una serie de factores concomitantes resumidos en la tabla 1 (Cramer 2018; *Euro-pean eBook Interoperability...* 2013; Miguel de Bustos y Casado del Río 2016; EPUB 3 3.0.1, 3.1, 3.2... 2018).

Tabla 1. Factores particulares del desarrollo del libro digital en la segunda década del siglo XXI

Factor	Descripción
Noción ampliada del concepto de libro determinada por el entorno digital	El libro de conforma por texto o cuenta con elementos multimedia e interactivos (<i>enhanced ebook</i>) expandibles en la web.
Su creación se da en un contexto de software y hardware diverso	Gigantes tecnológicos transnacionales controlan el mercado de la publicación digital en entornos cerrados (de software y hardware propietario) o parcialmente abiertos (soluciones de software y hardware libre en conjunción con formatos propietarios).
Implementación de medidas tecnológicas de protección que impiden prestar o compartir libremente el libro	Las medidas tecnológicas permiten a los autores y a otros titulares de derechos el ejercicio de las facultades que les otorga el derecho de autor mediante un control efectivo sobre el acceso y la utilización de las obras digitales
Estructura que facilita su descubrimiento y acceso en la web	Obligatoriedad en el uso de metadatos Posicionamiento SEO (Search Engine Optimization) Estrategias de Social Media Personalización
Múltiples opciones para su adquisición o uso	La propiedad de un contenido intangible (archivo) se diversifica en modelos de negocio/acceso como la suscripción, el préstamo, la compartición, el uso individual limitado o ilimitado, etcétera
Incorporación de pautas y estándares para su preservación	Los archivos y ficheros que conforman al ebook serán conservados para su uso futuro y se garantizará su integridad y autenticidad.

Fuente: Elaboración propia.

Un libro digital o *ebook* es interoperable cuando de manera independiente al sistema en que se adquiere o descarga puede leerse utilizando el mismo *software* o aplicación de lectura, incluidas las funcionalidades de administración y lectura social. En ese sentido, la interoperabilidad se entorpece cuando la lectura de diferentes libros requiere el uso de distintas aplicaciones de lectura, incluso si se ejecutan en el mismo dispositivo (de *hardware*). Del mismo modo, la interoperabilidad no se cumple si las herramientas de conversión permiten a los autores publicar libros digitales en diferentes ecosistemas. Por ejemplo, el *software* KindleGen de Amazon convierte *Epub*¹ ebooks en el formato *KF8* de la empresa. El resultado es dos ebooks diferentes capaces de distribuirse en diferentes ecosistemas. Además, esta opción solo se aplica a libros *epub* con características que también son compatibles con *KF8*. Esto se convierte en un serio problema ya que las versiones de *Epub3* proporcionan más ventajas al autor que *KF8*.

La existencia de formatos de libros digitales abiertos y propietarios condiciona la plena interoperabilidad que es imprescindible para que el ecosistema/mercado del ebook se expanda. Dicha interoperabilidad se traduce en la estandarización a todos los niveles, tanto de los libros digitales como de los dispositivos de lectura y el software de visualización. Por decir algo, no existen razones técnicas o funcionales que impidan la utilización del formato Epub 3 como estándar del libro digital a nivel global pero sí de política comercial, que finalmente se imponen (González San Juan; 2016, 70).

¹ Los editores y agregadores de contenido prefieren desde 2016 el Epub como formato de distribución primaria aun cuando lo envían a Amazon.

CARACTERÍSTICAS Y FUNCIONALIDADES DE LAS PLATAFORMAS DE PUBLICACIÓN Y DISTRIBUCIÓN DE LIBROS

Las plataformas en línea, entre las que se encuentran aquellas orientadas a la publicación y distribución de libros digitales, son definidas como sistemas de información proveedores de infraestructura tecnológica y reglas que trabajan en el entorno de la nube para el desarrollo de diversas aplicaciones. Permiten crear, distribuir, comercializar, almacenar y preservar productos y servicios digitales. Son los ecosistemas de la economía digital y base de la denominada industria 4.0 (clientes + productos o servicios + formas organizacionales simplificadas + innovación colaborativa) (Laudon y Price Laudon 2012). Parker, Van Alstyne y Choudary redondean estas ideas al concebir a la plataforma *online* como:

[...] un negocio basado en la capacidad de crear valor a partir de interacciones entre productores y consumidores externos. La plataforma provee una infraestructura abierta y participativa para aquellas interacciones y el conjunto de las condiciones que las regulan. El principal objetivo de una plataforma es contactar a los usuarios y facilitarles el intercambio de bienes, servicios o transacciones sociales, de esta manera será capaz de crear valor para todos los participantes (Parker, Van Alstyne y Choudary 2016, 11).

Estas plataformas deben diseñarse, desarrollarse y operarse considerando cuatro variables: 1) inteligencia para la captura de datos; 2) diseño como red abierta; 3) con hiperconexión, y 4) capacidad de crecimiento (escalabilidad). Tales sistemas buscan la interoperabilidad con otros a través del uso de

estándares que faciliten los intercambios de información y servicios. Sus componentes son tres:

- Contenido: Productos infocomunicacionales de índole digital, información autoagregada, portabilidad de formatos y empaquetamiento de contenidos generados por los usuarios, servicios.
- Infraestructura tecnológica: Sistemas robustos que sean escalables, dispongan de interfaces intuitivas y se basen en una arquitectura web de servicios en la nube. Incluyen herramientas colaborativas, flujo de operación automatizado, herramientas de recolección y análisis de datos.
- Usuarios: Manejo de variados niveles de interacción, comunicación personalizada, roles de actuación múltiples, efectos de red (Brand 2017).

Los actores que intervienen en la operación de una plataforma en línea son:

- a) Usuario: Cualquier persona física o jurídica que se registra en la plataforma y que interactúa con otras personas registradas.
- b) Consumidor: Tipo de usuario que se relaciona con profesionales en la plataforma.
- c) Proveedor: Tipo de usuario que aprovecha las plataformas como un canal de venta o distribución.
- d) Profesional: Tipo de usuario que lleva a cabo una o varias actividades o presta uno o varios servicios con ánimo de lucro y de manera habitual a través de una o varias plataformas. Un mismo usuario puede ser profesional en un ámbito y no profesional en otro.

- e) No profesionales: Tipo de usuario que lleva a cabo una actividad sin ánimo de lucro ni intención comercial a través de una o varias plataformas y con la única intención de compartir un bien infrautilizado, sus conocimientos o su tiempo y los costes asociados a la actividad en cuestión. Un mismo usuario puede ser no profesional en un ámbito y profesional en otro (*Los modelos colaborativos y bajo demanda...* 2017).

Una plataforma de publicación o distribución de libros debe permitir: crear el archivo del libro por parte del autor o editor; llevar a cabo la evaluación del material; efectuar los procesos de producción, distribución, comercialización, intercambio y preservación del *ebook*. En la actualidad, este tipo de plataformas son mayoritariamente de carácter comercial y están encabezadas por Amazon, multinacional que inició como librería y amplió su perfil como minorista, proveedor de servicios tecnológicos y finalmente reposicionó su papel en la industria editorial con el servicio de autopublicación Kindle Direct Publishing. Otras plataformas de alcance global que conjugan el negocio librero con el de autopublicación son Nook Press (EEUU), Apple Ibooks Store (EEUU), Kobo (Canadá), Tagus (España) y Smashwords (EEUU) (Truneau, 2017).

Los ebooks que se incorporen en esos sistemas con fines de comercialización deberán incluir metadatos ONIX (Online Information eXchange), además de los metadatos Dublin Core solicitados por el W3C. La importancia de los metadatos recae en su potencial de etiquetado, catalogación, descripción y clasificación de los recursos distribuidos en la web, lo que facilita su búsqueda y posterior recuperación. ONIX for Books es un estándar internacional en XML

diseñado para la codificación y el intercambio digital de información bibliográfica y comercial orientada a la industria del libro. Trabaja con códigos que proporcionan datos sobre ISBN, formato, título, autor, colaborador, sinopsis, diseño de cubierta, etcétera. Dispone de más de doscientos campos y fue propuesto por EDItEUR, grupo internacional que coordina el desarrollo de la infraestructura de normas para el comercio electrónico en los sectores de libros, libros electrónicos y publicaciones seriadas (Eguaras 2015).

La infraestructura y arquitectura de las plataformas de publicación o distribución de libros digitales establecen una estructura que integra colecciones de ebooks; espacios de trabajo para editores, lectores, autores y otros intermediarios; motores de búsqueda para localización de los materiales; opciones de administración de derechos; modalidades de pago; disponibilidad y acceso las 24 horas, los siete días de la semana; visualizadores de texto; información bibliográfica; conexión con redes sociales; herramientas de búsqueda dentro del ebook; opciones de citación del material; servicios de maquetación, diseño, traducción, animación; bibliotecas personales remotas para autores y lectores; variedad de formatos de descarga (que es la menos recurrente al tratarse de sistemas cerrados).

El informe *Neutralidad de las plataformas. Fomentar un entorno digital abierto y sostenible* (2014) del gobierno francés nos dice que “Como intermediarios, las plataformas pueden conectar, pero también interponerse y en ocasiones convertirse en competidores de sus propios clientes. Varias plataformas han adoptado modelos de desarrollo basados en la constitución de auténticos ecosistemas en cuyo centro se origin” (2014, 5). A ese respecto, datos del *Informe anual del libro digital 2017* demostraron que durante ese año el mercado del libro digital en lengua española se concentró

78 por ciento en plataformas internacionales, 12.3 por ciento en librerías y plataformas *online* independientes y 5.2 por ciento en plataformas de suscripción.

LA INTEROPERABILIDAD EN LAS PLATAFORMAS DE PUBLICACIÓN
Y DISTRIBUCIÓN DE LIBROS DIGITALES Y LOS RETOS A ENFRENTAR
PARA MEJORARLA

Las plataformas de publicación y distribución de libros digitales son interoperables cuando disponen de la facultad para intercambiar información con aquellas de su mismo tipo o con iguales condiciones o protocolos de comunicación. El ideario subyacente a la creación de esta clase de infraestructura tecnológica es compilar, organizar y proporcionar puertas de acceso/desarrollo a información o sistemas de información dispersos en la web pero centralizados en una interfaz que ahorra tiempo, dinero y esfuerzo a sus usuarios (*Marco Iberoamericano de Interoperabilidad* 2010; Martínez Tamayo *et al.* 2011). Ya en 2005, Lawrence Lessig afirmaba que “Quizá lo más importante que nos ha dado Internet es una plataforma sobre la cual la experiencia es interoperable”.

En plataformas *online*, la interoperabilidad posee una condición dual: semántica y sintáctica. Sintácticamente, la interoperabilidad de los sistemas se logra mediante el intercambio de datos; en sentido semántico, dar sentido a los datos de manera normalizada-estándar es una tarea más compleja y, a menudo, difícil. Empero, la interoperabilidad sintáctica puede concretarse utilizando estándares como el Lenguaje de marcado extensible (XML), que codifica datos de manera que se comprende su estructura pero no su significado. Por otra parte, el formato de metadatos

Dublin Core, utilizado en la catalogación de libros digitales, es un ejemplo de un estándar enfocado en la interoperabilidad semántica: el significado estandarizado se especifica en términos abstractos, con muchas codificaciones posibles (Amor, Fuentes y Pinto s.f.).

Si bien la interoperabilidad sintáctica y semántica son el punto de partida para la concreción de una “interoperabilidad real” en las plataformas de publicación o distribución de libros digitales, es necesario destacarlas como etapas iniciales de un proceso más complejo que se resume en esta serie de niveles:

- *Nivel 0. Los sistemas no tienen interoperabilidad.* Ejemplo: Páginas o sitios web de editores funcionan individualmente y de manera aislada. Son básicamente espacios para la difusión y venta directa de libros y servicios relacionados con la edición.
- *Nivel 1. Los sistemas disponen de interoperabilidad técnica.* Se establece una infraestructura de comunicación que permite a los sistemas intercambiar bits y bytes, y las redes y protocolos subyacentes se definen de manera inequívoca. Este nivel asegura la comprensión común de las señales. Ejemplo: Editores científicos transnacionales crean portales para centralizar su actividad y terciarizar servicios como la diagramación, lo que permite que sistemas con protocolos similares logren comunicarse.
- *Nivel 2. Los sistemas disponen de interoperabilidad sintáctica.* Se introduce una estructura común para intercambiar información; es decir, se aplica un formato de datos común. El formato

del intercambio de información está inequívocamente definido. Esta capa define la estructura y garantiza la comprensión común de los símbolos. Ejemplo: En la industria editorial digital, la interoperabilidad sintáctica va de la mano con la normalización de la información de los distintos sistemas a través del uso de estándares como ONIX y metadatos Dublin Core que describirán a los ebooks.

- *Nivel 3. Los sistemas utilizan un modelo de referencia de intercambio de información común.* Se alcanza el nivel de interoperabilidad semántica. En este nivel, el significado de los datos es compartido; el contenido de las solicitudes de intercambio de información está inequívocamente definido. Esta capa garantiza la comprensión común de los términos. Ejemplo: Al actuar como sistemas cerrados, la mayoría de las plataformas no disponen de un modelo de referencia para comunicarse con los editores —en el caso de las librerías como Amazon—, lo que sí ocurre con los repositorios digitales que intercambian información vía el protocolo OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting). Esto obliga a una ingesta directa de materiales por parte de los vendedores de productos o un ingreso a la plataforma por quienes proporcionan servicios;
- *Nivel 4. La interoperabilidad pragmática se alcanza cuando los sistemas interoperativos conocen los métodos y procedimientos que cada sistema está empleando.* En otras palabras, los sistemas participantes entienden el uso de los

datos, o el contexto de su aplicación; el contexto en el que se intercambia la información se define inequívocamente. Ejemplo: Plataformas de distribución y publicación de libros como OpenEdition han alcanzado este nivel de interoperabilidad al estar articuladas como consorcios o grupos de editores que circunscriben acuerdos para el manejo de estándares que faciliten la comunicación y disposición de sus recursos por medio de sistemas organizacionales que intercambian datos e información.

- *Nivel 5.* A medida que un sistema opera con datos a lo largo del tiempo, el estado de ese sistema cambiará y esto incluye las suposiciones y restricciones que afectan su intercambio de datos. Si los sistemas han logrado la *interoperabilidad dinámica*, comprenderán los cambios de estado que ocurren en las suposiciones y restricciones que cada uno está tomando a lo largo del tiempo y pueden aprovechar esos cambios. Los efectos de las operaciones se tornan cada vez más importantes; el efecto del intercambio de información dentro de los sistemas participantes está inequívocamente definido. Esta capa garantiza una comprensión común de los efectos en términos de cambios de estado, parámetros generados, etcétera. Ejemplo: Las plataformas comienzan a aprender, por medio de sus algoritmos² de trabajo, cómo se comportarán los intervinientes en la publicación y distribución de *ebooks* o de servicios relacionados con ellos. De este modo, son capaces de diversificar sus

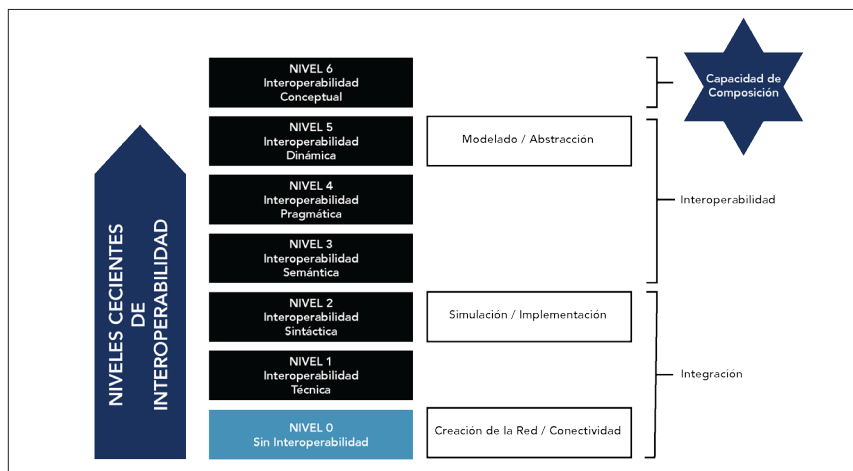
opciones de comunicación con sistemas para en conjunto proporcionar una nueva oferta de bienes intangibles. Este nivel está claramente vinculado al machine learning, proceso en el que las computadoras actúan sin estar programadas explícitamente; predicen resultados a partir de datos y análisis estadístico.

- *Nivel 6.* Si el modelo conceptual; es decir, las suposiciones y restricciones de la abstracción significativa de la realidad, está alineado, se alcanza el nivel más alto de interoperabilidad: la interoperabilidad conceptual. Ejemplo: Las plataformas de publicación y distribución de libros digitales pueden recombinar sus materiales y servicios, y dar lugar a nuevos productos, servicios y sistemas editoriales o de otra índole. Estos sistemas trabajan con total independencia del ser humano.

En el contexto internacional, la interoperabilidad en plataformas de publicación y distribución de libros digitales se encuentra mayoritariamente en los niveles 0 al 2, afianzando su conectividad interna y creando redes de sistemas externos locales o nacionales. Adolecer de interoperabilidad en cuanto a formatos de *ebooks* trae consecuencias a las plataformas y al resto de su infraestructura y operación.

2 Los algoritmos son series de operaciones detalladas y no ambiguas que actúan como reglas que resuelven cierta clase de problemas mediante aproximaciones paso a paso. Los algoritmos son los motores de las plataformas y un factor clave que determina la experiencia digital de los usuarios. Son clave en los procesos de Inteligencia Artificial (IA) y Machine Learning (aprendizaje de máquinas) (*Manifiesto por un nuevo...* 2018, 69).

Esquema 1. Niveles de interoperabilidad deseable en plataformas de publicación o distribución de libros digitales.



Fuente: Elaboración propia con información de *Niveles de Interoperabilidad* (2017) y Manso *et al.* (2008).

CONCLUSIONES

La interoperabilidad entre formatos y plataformas de publicación y distribución de libros digitales no se reduce a un principio deseable, sino a un factor necesario coadyuvante en la consecución de una verdadera sociedad digital. Esto quedó de manifiesto en la *Agenda Digital de la Comisión Europea* (2010), que contempla a la interoperabilidad efectiva de los productos electrónicos como un requisito a cumplir en ese sentido. En 2018, más del 70 por ciento del sector de *ebooks* a nivel mundial no ha trasladado sus

flujos de trabajo (publicación y distribución) a plataformas *online*. Solo un mínimo porcentaje de las plataformas de publicación o distribución de libros han alcanzado la interoperabilidad pragmática —nivel 4—. Este ecosistema se encuentra rezagado en términos de interoperabilidad; sus actores y organizaciones funcionan como entes aislados y escasamente integrados.

Una de las mayores trabas a las que se enfrentan las plataformas de publicación y distribución de libro digitales es de carácter legal y contractual. La existencia del *DRM* en sistemas de formatos cerrados implica la creación de acuerdos entre diferentes jugadores. Además, la distribución y publicación geográficas y por lenguas es un inconveniente porque favorece la creación y consolidación de ecosistemas y plataformas cerrados, lo que da lugar a oligopolios que segmentan el mercado y limitan la existencia de librerías independientes, así como la edición de libros orientados a minorías o comunidades muy específicas.

Finalmente, la falta de interoperabilidad perjudica gravemente a los usuarios o consumidores del ebook, quienes cuando adquieren un libro en una plataforma concreta se ven limitados a utilizarla fuera de ella por lo ilegal o complicado que resulta la conversión entre formatos propietarios de distinto origen.

BIBLIOGRAFÍA

- Adell, Fernan. 2016. *Edición en EPUB*. Barcelona: Editorial UOC.
- Amor, M., L. Fuentes y M. Pinto. s.f. *Interoperabilidad entre plataformas de agentes FIPA: una aproximación basada en componentes*. Montevideo: Instituto de Ingeniería Eléctrica de la Universidad de la República.
- Armañanzas Sodupe, Emy. 2013. "Breve historia y largo futuro del ebook". *Historia y Comunicación Social*, 18: 15-26. Disponible en <http://revistas.ucm.es/index.php/HICS/article/view/44307/41867>.
- Brand, Kurt. 2017. "Digital Business Models and Platform Economy". Blog personal de Kurt Brand. D 04 de noviembre de 2017. Disponible en <https://www.xing.com/news/insiders/articles/digital-business-models-and-platform-economy-1015352>.
- Cramer, Dave. 2018. "Why Specs Change: Epub 3.2 and The Evolution of The Ebook Ecosystem". (*Blog del W3C*). 16 de abril de 2018. Disponible en <https://www.w3.org/blog/2018/04/why-specs-change-epub-3-2-and-the-evolution-of-the-ebook-ecosystem/>
- Eguaras, Mariana. 2015. "Qué son los metadatos de un libro y cuál es su importancia". (*Blog de Mariana Eguaras. Consultoría editorial*). 06 de julio de 2015. Disponible en <https://marianaeguaras.com/que-son-los-metadatos-de-un-libro-y-cual-es-su-importancia/>.
- EPUB 3, 3.0.1, 3.1, 3.2: Das Versions-Chaos der eBook-Format. 2018. Múnich: Digital Publishing Competence. Disponible en <https://www.dpc-consulting.org/epub-3-3-0-1-3-1-3-2-das-versions-chaos-der-ebook-format/>.

- European eBook Interoperability Study Recommends EPUB 3*. 2013. Seattle: International Digital Publishing Forum. Disponible en <http://idpf.org/news/european-ebook-interoperability-study-recommends-epub-3>.
- García Jiménez, Jesús. 2014. *Libro electrónico_epub*. Disponible en https://jesusgarciaj.com/produccion-editorial-jesus-garcia/libro-electronico_epub/.
- González San Juan, José Luis. 2016. "Libro Electrónico e Interoperabilidad. Formatos y Medidas Tecnológicas de Protección". *Telos. Revista de Pensamiento sobre Comunicación, Tecnología y Sociedad*, vol. 104: 69-77. Disponible en <https://telos.fundaciontelefonica.com/revista/>.
- Informe anual del libro digital 2017*. 2018. Barcelona: Librandia. Disponible en <http://librandia.com/wp-content/uploads/2018/04/Informe-anual-libro-digital-2017.pdf>.
- Laudon, Kenneth y Jane Price Laudon. 2012. *Sistemas de información gerencial*. México: Pearson Educación.
- Lebert, Marie. 2009. *Una corta historia del ebook*. Toronto: Universidad de Toronto. Disponible en <http://www.etudes-francaises.net/dossiers/ebookES.pdf>.
- Los modelos colaborativos y bajo demanda en plataformas digitales*. 2017. Madrid: Asociación Española de la Economía Digital. Disponible en <https://www.adigital.org/media/plataformas-colaborativas.pdf>.
- Manifiesto por un nuevo pacto digital. Una digitalización centrada en las personas*. 2018. Madrid: Fundación Telefónica. Disponible en https://www.telefonica.com/manifiesto-digital/assets/un_nuevo_pacto_digital_infografia.pdf.

- Manso, M. A., M. Wachowicz, M. A. Bernabé, A. Sánchez y A. F. Rodríguez. 2008. "Modelo de Interoperabilidad Basado en Metadatos (MIBM)". *Proceedings JIDEE*, 1-20. Disponible en http://oa.upm.es/3604/1/INVE_MEM_2008_56085.pdf.
- Marco Iberoamericano de Interoperabilidad*. 2010. Caracas: Centro Latinoamericano de Administración para el Desarrollo.
- Martínez Tamayo, Ana M., Julia C. Valdez, Edgardo A. Stubbs, Yanina González Terán y María Inés Kessler. 2011. "Interoperabilidad de sistemas de organización del conocimiento: el estado del arte", *Información, cultura y sociedad*, 24: 15-37. Disponible en <http://www.scielo.org.ar/pdf/ics/n24/n24a02.pdf>.
- Miguel De Bustos, Juan Carlos y Miguel Ángel Casado del Río. 2016. "Google, Apple, Facebook y Amazon. Emergencia de los GAFA y cambios en el sistema comunicativo global", *Telos. Revista de Pensamiento sobre Comunicación, Tecnología y Sociedad*, 104: 38-48. Disponible en <https://telos.fundaciontelefonica.com/revista/>.
- Neutralidad de las plataformas. Fomentar un entorno digital abierto y sostenible*. 2014. París: Conseil National du Numérique.
- Niveles de Interoperabilidad*. 2017. Madrid: Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado. Disponible en <http://educalab.es/intef/tecnologia/infraestructuras-tecnologicas/interoperabilidad/niveles-de-interoperabilidad>.

- Parker, Geoffrey G., Marshall W. Van Alstyne & Sangeet Paul Choudary. 2016. *Platform Revolution. How networked markets are transforming the economy and how to make them work for you*. Nueva York: W. W. Norton & Company.
- Truneau, Valentina. 2017. *Plataformas de publicación digital. Ventajas y desventajas*. Maracaibo: Valentina Truneau.

3. DATOS CULTURALES

Metadatos para preservación digital

JUAN VOUTSSÁS MÁRQUEZ

Universidad Nacional Autónoma de México

Los metadatos existen desde mucho antes de que se acuñara este nombre en la década de los sesenta. Se encontraban ya en forma de descriptores de contenido en tabletas de arcilla de la biblioteca de Ebla en el tercer milenio antes de nuestra era, en la Biblioteca de Alejandría, a lo largo de la Edad Media y, por supuesto, en la edad contemporánea. Debido a esa antigüedad y a que siguen siendo sin duda de la mayor importancia, han sido los más tratados y desarrollados y se les presta mucha atención. Hoy en día existe un sinnúmero de definiciones, métodos, variantes, teorías, estándares, etcétera, relativos a los descriptores de contenidos. No obstante, existen muchos otros atributos de los documentos que también pueden ser representados con metadatos.

Este texto pretende tratar en especial aquellos metadatos para preservación de documentos digitales. Su importancia radica en el hecho de que con mucha frecuencia pasan desapercibidos y son ignorados por los responsables de esos documentos, lo cual pone en riesgo su adecuada preservación. Sin adecuados metadatos creados expresamente al efecto, será muy difícil, o casi imposible, lograr la preservación ulterior de un documento digital.

La preservación de documentos digitales dentro de un fondo documental, llámese este colección, repositorio, acervo, archivo, etcétera, es un proceso que debe ser diseñado al momento de concebir ese fondo. Muchos todavía piensan que la preservación de documentos digitales es exclusivamente un problema tecnológico, y que por eso su solución consiste en la adquisición de dispositivos y programas computacionales y algunos procesos y accesorios para respaldos de la información, y por tanto, es responsabilidad total del área de informática de la organización.

En consecuencia, el momento en que se comienza a pensar en la preservación es cuando el fondo documental está creado, estable y lleva buen tiempo operando. Ésa es una receta para el desastre. Las políticas, los procedimientos y las estructuras para la preservación documental digital se deben diseñar y construir al mismísimo inicio del fondo documental para que cuando este comience a construirse, ya contenga todos los elementos que permitirán a los documentos digitales ser preservados por todo el tiempo necesario. Pretender agregar después estos componentes de preservación es prácticamente imposible y cuando llega a hacerse, implica una enorme inversión de recursos económicos, técnicos y humanos. Es por tanto indispensable diseñar la preservación de inicio junto con el fondo documental correspondiente.

Para comenzar, conviene establecer que se preservan objetos digitales y no simples documentos digitales. De forma simple, puede definirse que

[...] un documento digital es una unidad de información persistente constituida por un mensaje fijado a un medio de manera sintácticamente estable, cuyos datos están representados por medio una sucesión de valores numéricos diferenciados, por lo

general secuencias de bits, de acuerdo con un código o convención preestablecido (Voutssás 2017, 206).

Esta definición puede ser complementada agregando que esa unidad de información puede contener cualquier tipo de ella: texto, imágenes, sonido, video, y sus combinaciones, en cualquiera de sus formas de expresión conocidas: libros, revistas, artículos, tesis, diarios, documentos de archivo, mapas, partituras, gráficas, música, cine, juegos, programas y aplicaciones informáticas, correos electrónicos, arte digital, etcétera. No obstante, conviene que este concepto de “documento digital” sea ampliado al de “objeto digital”. Existen muchas definiciones al respecto; por ejemplo, la Biblioteca Digital de California establece en su glosario: “Objeto digital: entidad en la que uno o más archivos de contenido y sus metadatos correspondientes están unidos físicamente y/o lógicamente mediante el uso de un contenedor digital”.

Habiendo algunas sutiles diferencias entre ambos conceptos relacionadas sobre todo al contexto, un objeto digital, para resumirlo de forma simple, es:

[...] un objeto documental representado numéricamente por una agregación o secuencia de valores numéricos en forma de una o más cadenas de bits las cuales –bajo ciertas reglas de representación– contienen datos dentro de ese objeto, así como metadatos acerca de las propiedades del mismo y –cuando sea necesario–, los métodos para realizar operaciones sobre el objeto. El significado de las cadenas de bits como objeto documental está en función de la interpretación que hagan de ellas los programas de su entorno digital (Interpares s.f.).

Puede verse que las definiciones de “documento digital” y “objeto digital” son semejantes; la diferencia fundamental

consiste en que en las definiciones de objeto digital se aclara que estos tienen, además de los datos o contenido, metadatos y eventualmente procedimientos. Básicamente, quiere decir que un documento digital es un contenido, mientras que el objeto digital incluye otros elementos más allá del contenido, y en esos elementos no pueden faltar los metadatos.

Conviene también establecer que para fines de este texto, la *preservación digital* tiene que ver con el contenido del documento, y su alcance siempre es el largo plazo; trata objetos lógicos. Como complemento, la *conservación digital* tiene que ver con el soporte y forma del documento digital, y su alcance es siempre el corto-mediano plazo; trata objetos físicos, es sinónimo de *mantenimiento*. Como corolario, podemos afirmar que conservamos objetos físicos (soportes) para preservar objetos lógicos (contenidos). *Preservación digital* es a la vez un objetivo y un resultado: como objetivo, “preservación digital” es en esencia “asegurar a largo plazo de la permanencia y acceso del contenido de documentos digitales confiables a lo largo del tiempo y las tecnologías”. Al instrumentar y lograr este objetivo, se obtiene como resultado –se logra– la “preservación digital” de los materiales; esto es: se obtiene “la protección del contenido intelectual de materiales digitales –datos, documentos o archivos” (Voutssás y Barnard 2014).

Para que un objeto digital se preserve, debe mantener a lo largo del tiempo ciertas características inherentes. Éstas varían de acuerdo a cada tipo de documento, así como su importancia y peso específico, pero en general pueden establecerse seis características principales (Voutssás 2009):

- Permanencia (la información existe). Esta característica está relacionada con el concepto de que el documento estará disponible por un lapso

considerable. Está asociado con su presencia, su seguridad y, por supuesto, con la duración y continuidad de su soporte. La permanencia depende en mucho del almacenamiento permanente seguro de los objetos digitales. Para ello son necesarios soportes, procedimientos y técnicas adecuadas para mantener esos objetos a largo plazo conservando las *cadenas de bits* que forman el documento digital para asegurar la presencia de esos datos. Si el soporte o los datos no permanecen, el documento ya no estará disponible.

- **Accesibilidad** (la información es reinterpretable y usable). Tiene que ver con la capacidad de poder acceder en un futuro a la información o datos contenidos en un cierto soporte. El hecho de lograr que el soporte y las cadenas de *bits* contenidas en él se conserven en buen estado no implica que podrán ser utilizables. Se debe poder asegurar que habiéndose conservado en buen estado se podrán ver o ejecutar estos archivos informáticos dentro de mucho tiempo reproduciendo correctamente las cadenas de *bits*. Si el soporte o formato del fichero informático pertenecen a tecnologías muy obsoletas y ya no pueden ser reproducidos porque se carece del equipo para leerlos o del sistema operativo o el programa informático que los operaba, la información existe, pero no puede accederse a ella. Independientemente del buen estado de los datos, es decir, de que haya habido *permanencia*, puede o no haber *accesibilidad*. Como puede verse, son dos características distintas

pero complementarias; se requiere que existan ambas para un adecuado acceso. La falta de cualquiera de los dos hará inaccesible al documento en cuestión.

- Disponibilidad (la información puede ser encontrada y accedida). Tiene que ver con que el documento está alojado en un soporte que seguirá siendo accesible a través del tiempo, y que la información está debidamente registrada en algún sistema de búsqueda para asegurar que podrá ser localizada y reproducida.
- Confidencialidad y privacidad. Se refiere al hecho de que el documento solo pueda ser visto o utilizado por las personas autorizadas para ello, y que los datos personales asociados al documento no puedan ser transferidos a terceros no autorizados ni extraídos por ellos.
- Autenticidad (atributos para que el documento ser considerado original y fiable). Cada tipo de documento debe tener previamente establecidos aquellos atributos que a lo largo del tiempo siga siendo considerado auténticos y veraces (autoría, identidad, integridad, estabilidad, contexto, custodia, forma fija, etcétera) y que, por tanto, pueda seguir siendo considerado confiable, aun a pesar de cambios que se hayan introducido al documento esporádicamente debido a actualizaciones tecnológicas.
- Aceptabilidad (no repudio). Tiene que ver con las características que permiten a un cierto documento seguir siendo aceptado por el emisor y los usuarios: números DOI, ISBN; firmas, sellos, y marcas de agua electrónicas; *hashings*, *checksums*,

depósito en repositorios oficiales fiables, cifrados, etcétera.

Si se crean de inicio y se mantienen, todas estas características en conjunto mantendrán la usabilidad y confiabilidad de la información de un cierto documento a lo largo del tiempo. Para saber que estos atributos existen y que se han conservado a lo largo del tiempo, se requiere indispensablemente de dotar al documento de metadatos específicos para ello que, como puede intuirse, no tienen ya que ver con la descripción de contenido del documento: son adicionales.

En efecto, los metadatos fueron creados para describir características inherentes a un documento; comenzaron como descriptores básicos del contenido de ese documento, pero después se fueron sofisticando para describir muchos otros atributos del mismo, y también se fueron especializando para describir características específicas de cada tipo de documento. De todos los tipos de documentos, pueden extraerse varios atributos generales, pero además puede extraerse una serie de características o atributos propios y específicos para cada uno de ellos; esas características propias son importantes para el sector que los produce y los utiliza. Además, el peso específico y la importancia de cada una de esas características varía de sector en sector y, por tanto, su tratamiento es muy distinto para cada tipo de documento: libro, archivo, artículo, filme, imagen, partitura, noticia, etcétera. Además, el advenimiento de los documentos digitales trajo nuevas listas de atributos a incluir. En términos generales y agrupando en una lista-resumen, al comparar y extraer diversas especificaciones y conceptos para cada tipo de documento, las características específicas inherentes a todos los tipos de documentos digitales pueden describirse en los siguientes grandes grupos:

- Contenido - Tipos y géneros de contenidos, multimedialidad (variedad de tipos de datos en un mismo documento), multidimensionalidad (múltiples aproximaciones o enfoques a un documento), etcétera.
- Contexto - Información acerca de cómo, cuándo, dónde, por quién y en qué circunstancias se produce el documento; por qué se produce; para qué se usa y su relación con otros documentos en un corpus.
- Proveniencia - Información acerca del origen del documento: persona(s) o autoría, organización, procedimiento, etcétera.
- Estructura - Información acerca de cómo debe ser la apariencia del documento, cómo se integran intelectualmente las partes de documentos complejos, hipervínculos, etcétera.
- Almacenamiento - Información sobre las características particulares del almacenamiento actual del documento, o cómo debe almacenarse (dispositivos especiales, algoritmos de compresión, etcétera).
- Accesibilidad - Medida en que el documento puede ser accedido, reproducido, copiado, modificado, transmitido, etcétera; su gratuidad o costo, etcétera.
- Preservación - Información acerca de cómo se debe preservar o se ha preservado ese documento; periodos de conservación, eliminación, etcétera.
- Autenticidad - Información acerca de los atributos que el documento tiene y debe conservar a lo largo del tiempo para ser considerado auténtico u original (autoría, identidad, integridad, estabilidad, custodia, firmas, etcétera).

- Modificabilidad - Información acerca del grado en que el documento puede ser modificado, en qué condiciones y por quién (wikis, folksonomías, etcétera).
- Tecnología - Datos acerca del entorno tecnológico de producción y uso del documento: sistemas, equipos, y programas informáticos; formatos, estándares, escalas, procesos técnicos efectuados, envolturas, firmas o sellos digitales, sumas verificadoras, cifrados, certificados, etcétera.
- Modularidad - Medida en que el documento puede ser dividido en varias partes para su uso o interpretación.
- Interoperatividad - Medida del grado en que el documento puede ser utilizado a lo largo de múltiples plataformas y estándares con pocas o nulas restricciones.
- Dinamismo - Medida en que el documento se actualiza y cambia con las experiencias del usuario (bases de datos, páginas web dinámicas, etcétera).

No todas las anteriores características existen en todos y cada uno de los documentos digitales, pero absolutamente todos los tipos de documentos digitales tienen varias de estas características. Éstas deben ser representadas por medio de metadatos específicos para cada una de ellas.

Además del contenido, algunos de estos atributos se consignan en forma inherente dentro del propio documento, como el nombre o identificador del archivo informático que lo contiene o el formato principal del documento: doc, pdf, jpg, wav, mp3, etcétera, y muchos otros atributos se agregan en forma de metadatos específicos. Para cada tipo de documento, se contemplan o recomiendan metadatos específicos

con mayor o menor énfasis en algunos de ellos dependiendo de la importancia que se les otorga dentro de ese tipo de documentos. Así, se encuentran metadatos específicos descriptivos, documentales, estructurales, de contexto, jurídicos, administrativos, funcionales, tecnológicos, de autenticidad, de ubicación, de privilegios o restricciones al acceso y uso, de procedimiento, y por supuesto de preservación, etcétera. En función de sus características y volumen, los metadatos pueden estar embebidos dentro del documento o crearse en documentos aparte con sus correspondientes hipervínculos. Más aún, en algunos tipos de documentos se recomienda ya el uso de “metametadata”; esto es, datos acerca del origen de los metadatos y su compilación, con el propósito de establecer su rigor, precisión, autenticidad, amplitud, etcétera, y por consiguiente, su confiabilidad; por ejemplo, nombre de personas u organizaciones que compilaron los metadatos; fechas de los mismos; si son metadatos generados en forma manual o automatizada; las normas y estándares, fuentes, herramientas, técnicas, procedimientos, actualizaciones, etcétera, utilizados para recopilarlos.

Para abundar más acerca del tema de la variedad de metadatos, consúltese el diccionario PREMIS (2016), así como la norma ISO 23081-1 (2011) - Información y documentación - Procesos para la gestión de documentos – Metadatos. Ambos son de índole general; para cada tipo de documentos existen estándares, normas y recomendaciones específicos. Éstos son hoy en día muy numerosos. A manera de ejemplos pueden mencionarse:

- Los estándares ISO/IEC 15489:2001 - “Información y Documentación - Gestión de Documentos - Parte 1. Generalidades”; ISO/IEC 15489:2006

- “Información y Documentación - Gestión de documentos - Parte 2: Directrices”.
- El estándar ISO TR 15801:2009 -Recomendaciones para la veracidad y confiabilidad de información almacenada electrónicamente.
- El estándar ISO 18492:2005- Conservación a largo plazo de información basada en documentos.
- ISO 19005-1:2008 – Partes 1, 2 y 3. Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1).
- El Modelo de Referencia OAIS –Open Access Information System– para archivos digitales, el cual generó la norma ISO 14721:2012 (OAIS, 2012).
- El Modelo MoReq: Modelo de Requisitos para la gestión de Documentos de Archivo - (DLM Forum, 2001).
- El documento “*Trusted Digital Repositories*” del RLG-OCLC Working Group on Digital Archives Attributes - (RLG-OCLC, 2002).
- La iniciativa NDIIPP - The National Digital Information Infrastructure and Preservation Program de la Biblioteca del Congreso de los E.U.A. - (NDIIPP, 2012).

Como ya se mencionó, el gran problema de la preservación de recursos documentales digitales a largo plazo no reside mayormente en factores de tipo técnico, sino de procedimiento. Independientemente del tipo o modelo de plan de preservación que una organización adopte y del tipo de documentos que maneje, es necesario establecer de entrada una política de preservación, la cual será acompañada de estrategias de *conservación y estrategias de preservación*

que deriven en un plan de acción. Esto obedece de origen a planteamientos ampliamente difundidos por la UNESCO en sus *Directrices para la Preservación del Patrimonio Digital*, que ofrecen un amplio marco de referencia para el establecimiento de estrategias de preservación para cada organización y sus tipos de documentos de archivo digitales (UNESCO 2003).

Estrategias de conservación (también llamadas estrategias de mantenimiento). Consiste en un conjunto coherente de objetivos y métodos para proteger y mantener la accesibilidad de los soportes de los documentos digitales a lo largo de las primeras etapas de una cadena de custodia (InterPARES 2010). En general, consisten en:

- 1) Clara asignación de responsabilidades dentro de la organización.
- 2) Adquirir la infraestructura técnica adecuada.
- 3) Mantenimiento, actualización y cambio de sistemas informáticos.
- 4) Transferencia de datos periódicamente hacia soportes de almacenamiento nuevos y vigentes.
- 5) Adopción de condiciones físicas apropiadas para el almacenamiento de los soportes.
- 6) Redundancia de medios en distintas ubicaciones físicas.
- 7) Plan de seguridad informática en los sistemas.

Estrategias de preservación. Además de las anteriores estrategias de conservación, todo preservador de documentos digitales es responsable de establecer también estrategias de preservación. Consisten en un conjunto coherente de objetivos y métodos para proteger y mantener la usabilidad y estabilidad tecnológica de los contenidos de documentos

digitales a lo largo de todas las etapas de una cadena de custodia (InterPARES 2010). En general, éstas son:

- 1) Preferir estándares abiertos y permanentes (pdf 1-A, odt, tiff, etcétera).
- 2) Usar formatos autodescriptivos como XML.
- 3) Restringir el número de formatos que pueden usarse en la organización; es decir, normalizar en lo posible estándares, estructuras y formatos.
- 4) Establecer e incluir en los documentos metadatos adecuados para el tipo de documentos en cuestión, sin olvidar los de preservación.
- 5) Migrar, refrescar, emular, etcétera (evitar obsolescencia tecnológica).

Como conclusiones, puede establecerse que la preservación de documentos u objetos digitales no consiste en el simple almacenamiento protegido de documentos en un soporte electrónico. Es toda una escuela de pensamiento cuyo objetivo central pretende hacer permanentes esos documentos digitales en el largo plazo —idealmente para siempre—; pretende mantenerlos accesibles y usables; auténticos, íntegros y fiables. La base de toda su teoría implica, además de la tecnología, la organización y observación cuidadosa de políticas y procedimientos de gestión documental que involucren planes y estrategias para una correcta y permanente preservación documental digital. Más aún, debe ser entendida no solo como un método para mantener cadenas de *bits* legibles y utilizables en un futuro, sino como todo un proceso cultural y social que interactúa en forma planeada y armónica con una completa variedad de necesidades y responsabilidades de organizaciones públicas y privadas, y que sirve para apoyar sus actividades cotidianas, su rendición de cuentas

y su memoria institucional como parte de una historia colectiva de una sociedad.

El proceso de preservación de documentos digitales para su resguardo y uso a largo plazo es un reto que implica amplios conocimientos tanto teóricos como metodológicos; decisiones delicadas, fuertes inversiones en recursos humanos calificados y algunas en tecnología; implicaciones legales, obsolescencia de dispositivos, reconversiones de tiempo en tiempo, y ciertos riesgos. Muchos responsables de colecciones, acervos, repositorios o archivos digitales han considerado la posibilidad de preservar o digitalizar algunos de sus fondos documentales pero han pospuesto la decisión de emprender estos proyectos confiando en que llegaría un momento en el que la tecnología y los costos se volviesen estables y competitivos, y los procesos fueran más simples, solo para darse cuenta de que a pesar de que la tecnología y sus costos mejoran cada día, la cantidad y variedad de materiales a digitalizar crece en forma exponencial y cancela de golpe estas ventajas.

Por lo mismo, no es recomendable esperar a que los parámetros se estabilicen totalmente y los riesgos desaparezcan; eso no está sucediendo y no pareciera ser así en un futuro cercano. Por otra parte, la demanda social y la legislación presionan cada día más para que las organizaciones públicas tengan colecciones, repositorios o archivos digitales dinámicos, actuales, funcionales y operativos, además de ser accesibles en forma remota por medios electrónicos. Es imperativo realizar estas actividades, y por ello será mejor estudiar a fondo el problema, entender sus fundamentos, características, ventajas y desventajas; prepararse bien y realizar un plan meticuloso y detallado para el desarrollo de una estrategia específica para estas actividades a fin de evitar cometer fallas al hacerlo bien desde la primera vez,

sin incurrir en costos o riesgos innecesarios, y sin errores que a la larga se deban lamentar y corregir. Como ha podido observarse a lo largo de este texto, los metadatos de preservación forman parte importante de ello.

BIBLIOGRAFÍA

- DLM Forum. 2001. Moreq -*Modelo de Requisitos Para la Gestión de Documentos Electrónicos de Archivo*. Véase también: Moreq2 - *Modelo de Requisitos Para la Gestión de Documentos Electrónicos de Archivo*, versión 2, 2004. Y también: MoReq 2010 - *Modelo de Requisitos Para la Gestión de Documentos Electrónicos de Archivo*, versión 2010. Disponible en http://ec.europa.eu/archival-policy/moreq/index_en.htm.
- InterPARES. 2010. (The International Research on Permanent Authentic Records in Electronic Systems). Guía del Preservador - *Preservación de Documentos de Archivo Digitales: Guía Para las Organizaciones*. México: UNAM - CUIB. Disponible en <http://iibi.unam.mx/archivistica/guias>.
- InterPARES 2. 2019. Glossary Database. Entrada por “Digital Object”. Disponible el 24 de octubre de 2019 en http://www.interpares.org/ip2/ip2_terminology_db.cfm.
- ISO 23081-1. 2011. Información y documentación. *Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1 - Principios; Parte 2 - problemas conceptuales y de implementación; parte 3 - Métodos de autoevaluación*.
- NDIIPP. 2012. (The National Digital Information Infrastructure and Preservation Program of the Library of Congress). *Preserving our digital heritage*. Disponible en http://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report_Post.pdf

- OAIS. 2012. (Reference Model for an Open Archival Information System). Version CCSDS 650.0-M-2. Management Council of the Consultative Committee for Space Data Systems (CCSDS). Washington DC, junio 2012. Disponible en <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- PREMIS Data Dictionary for Preservation Metadata, Version 3.0*. 2016. Library of Congress. Disponible en <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.
- RLG-OCLC Working Group. 2002. *Trusted Digital Repositories: Attributes and Responsibilities*. An RLG-OCLC Report. Mountain View, CA: OCLC. Disponible en <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.
- UNESCO. 2003. *Directrices para la Preservación del Patrimonio Digital*. División de la Sociedad de la Información de la UNESCO. Preparadas por la Biblioteca Nacional de Australia. Sitio Oficial de la UNESCO. Disponible en <http://www.unesco.org/new/es/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-the-preservation-of-digital-heritage/>.
- Voutssás M., Juan. 2017. *Confianza e Información Digital: Bibliotecas, Archivos y Web*. México: IIBI-UNAM. ISBN: 978-607-02-9757-1. Disponible en http://132.248.242.6/~publica/archivos/libros/confianza_informacion_digital_s.pdf.
- Voutssás-M., Juan y Alicia Barnard. 2014. *Glosario de preservación archivística versión 4.0*. México: IIBI-UNAM. ISBN: 978-607-02-5744-5. Disponible en http://132.248.242.6/~publica/archivos/libros/glosario_preservacion_archivistica_digital_v4.0.pdf.

Voutssás-M, Juan. 2009. *Preservación del Patrimonio Documental Digital en México*. México: CUIB-UNAM. ISBN: 978-607-02-0583-5. Disponible en http://132.248.242.6/~publica/archivos/libros/preservacion_patrimonio.pdf.

El Sistema de Información para el Registro Universitario de Espacios Activos y Culturales (SI-RUEyAC)

CATALINA NAUMIS PEÑA

NATALIA VELAZCO PLACENCIA

ARIEL ALEJANDRO RODRÍGUEZ GARCÍA

Universidad Nacional Autónoma de México

INTRODUCCIÓN

El conocimiento del patrimonio cultural de la Universidad Nacional Autónoma de México (UNAM) sin lugar a duda contribuye a su aprovechamiento. Las tecnologías de la información y la comunicación permiten disponer de registros referentes a la posesión de infraestructura cultural y recuperar esa información para programar actividades y abonar al mantenimiento y uso del patrimonio universitario.

El panorama anterior fue el detonante para que la STPP-CDC (Secretaría Técnica de Planeación y Programación de la Coordinación de Difusión Cultural) se diera a la tarea de iniciar el proyecto Sistema de Información para el SI-RUEyAC (Registro Universitario de Espacios y Activos Culturales), que consiste en la construcción de una plataforma en línea para el registro, la catalogación, la recuperación y la consulta de la infraestructura cultural, y la programación de las actividades

artísticas y culturales, que contribuyen al desarrollo integral de los universitarios a través de la difusión de expresiones artísticas y culturales de la UNAM. Un proyecto de esta naturaleza, sirve a la universidad y es un antecedente de utilidad para registrar, conocer y operar la infraestructura cultural de otras instituciones similares.

Se contemplan como usuarios principales de la información a personas involucradas en la labor cultural de la CDC de la UNAM, así como su subsistema y los coordinadores de actividades culturales en distintas sedes, instancias y planteles de educación superior y media superior de la universidad (entidades académicas). Esta herramienta permite hacer más fuerte a la universidad en el ámbito de la difusión cultural porque el registro de infraestructura ofrece la posibilidad de emplear los recursos en la materia con eficiencia.

La necesidad del uso de información oportuna para facilitar procesos de consulta y de toma de decisiones en el ámbito de la administración cultural exige el diseño de una estructura que permita la recuperación de información sistematizada mediante herramientas como los metadatos y los vocabularios controlados para que la comunicación entre responsables de planificar actividades fluya sin malas interpretaciones.

FUNDAMENTOS DE LOS ESPACIOS CULTURALES UNIVERSITARIOS

La cultura es considerada un elemento indispensable en el Plan de Desarrollo Institucional de la UNAM, tal como se especifica en el mismo “componente insustituible en la conformación de la identidad de toda colectividad, de ahí su relevancia y trascendencia, confirmadas por una larga trayectoria y una profunda tradición” (Graue 2017, 36).

Las universidades públicas como la UNAM están incluidas entre las organizaciones independientes no lucrativas en relación con la gestión de su patrimonio cultural. En esta condición,

La estructura interna de la organización sigue el modelo de pirámide jerárquica [...] la financiación de este tipo de entidades es compleja necesariamente ya que deben procurarse sus medios independientemente [...] dado su carácter no lucrativo... para emprender inversiones deben procurarse fondos extra (Ballart-Hernández y Tresserras 2001, 122)

La conformación de la plataforma de espacios y activos culturales de la UNAM incluye el financiamiento parcial del proyecto a través de un proyecto institucional de apoyo que permite la integración de dos dependencias universitarias para la consecución del trabajo y difusión del proyecto: la STPP-CDC y el IIBI (Instituto de Investigaciones Bibliotecológicas y de la Información).¹

La teoría del capital cultural desarrollada por Pierre Bourdieu trasciende como un telón de fondo en la perspectiva de la difusión cultural universitaria como fundamental en la creación de condiciones que amplíen el horizonte y las perspectivas de la comunidad.

El capital cultural puede existir bajo tres formas: en el estado incorporado, es decir, bajo la forma de disposiciones duraderas del organismo; en el estado objetivado, bajo la forma de bienes culturales, cuadros, libros, diccionarios, instrumentos, maquinaria, los cuales son la huella o la realización de teorías o de críticas a

¹ Agradecemos a la Dirección General de Asuntos del Personal Académico de la UNAM por el financiamiento otorgado para el desarrollo del Proyecto PAPIIT IT 400318 Sistema de Información para el Registro Universitario de Espacios y Activos Culturales.

dichas teorías, y de problemáticas, etc., y finalmente en el estado institucionalizado, como forma de objetivación muy particular, porque tal como se puede ver con el título escolar, confiere al capital cultural —que supuestamente debe de garantizar— las propiedades totalmente originales (Bourdieu 1987, 12).

Para resumir, los argumentos clave de Bourdieu en el plano de la cultura, Bennet los divide en tres aspectos: la importancia del capital cultural, la homologación entre los campos de la cultura y el rol de la cultura para crear ventajas sociales (Bennet 2009, 9)

En este sentido, existen estudios basados en la teoría de Bordieu con referencia a los jóvenes:

[...] se trata de un período en el que las personas adoptan decisiones muy importantes, de las que posteriormente pueden o no sentirse responsables, sobre las cuales puede o no juzgárseles culpables. E igualmente no cabe duda de que esas decisiones se han adoptado bajo regímenes de realidad social que desde el principio generaron condiciones radicalmente diferentes para unos y otros jóvenes (Gayo 2013, 168-169).

La juventud, definida en la encuesta que realizó Gayo para sustentar su hipótesis,

[...] como la edad comprendida entre los quince y los veintinueve años, contiene un rango de edades dentro del cual hay una notable diversidad. A este respecto, los más jóvenes entre los jóvenes son los más participativos, lo que muestra que el muchas veces comentado alargamiento de la juventud no oculta que importantes cambios se producen en lapsos de pocos años (Gayo 2013, 160).

La universidad trabaja precisamente con y para jóvenes en ese rango de edades y la difusión de la cultura es una de sus responsabilidades sustantivas. Sin lugar a duda, esa actividad cultural se extiende a la población en general, y cubre así un amplio espectro. El conocimiento y la manipulación de los espacios y sus posibilidades de programación son una necesidad dada la amplitud de la cobertura universitaria en materia de difusión.

Una programación y difusión efectiva de las actividades culturales de la universidad exige el conocimiento de sus posibilidades a través de mecanismos eficientes de comunicación. Por lo tanto, existe la necesidad de una información sistematizada sobre la infraestructura disponible que de pie a reforzar el crecimiento de nuevas formas de la difusión cultural. La razón de ser de los bienes culturales es social (Querol 2010, 13).

La integración de los espacios y activos culturales de la UNAM en un sistema de información es de utilidad social al sistematizar a través de metadatos las funciones, los atributos, los instrumentos y los programas al interior de la infraestructura cultural universitaria, y sirve a los responsables de eventos culturales que en forma permanente o esporádica utilizan o administran algún espacio o equipo cultural. Los objetivos del sistema son:

- Fomentar el aprendizaje institucional entre la Coordinación de Difusión Cultural, los responsables de actividades culturales en los planteles de la universidad y las direcciones del Subsistema, para detectar necesidades y oportunidades en la administración de información cultural.
- Realizar el registro de recintos especializados en la Difusión de la Cultura con el uso de estándares

internacionales para el levantamiento y la homologación de información.

- Facilitar a las direcciones del Subsistema de la Coordinación una herramienta para consultar información cultural que agilice los procesos de programación y apoye los criterios de planeación con información rápida y oportuna.
- Contribuir con un catálogo de espacios culturales y medios técnicos de los planteles de educación UNAM a la línea estratégica 10.13 del Plan de Desarrollo Institucional que establece “Integrar actividades culturales semanales in situ a las escuelas y facultades, facilitar su comprensión, cultivar su gusto, divertimento, permitiendo que las inclinaciones culturales se consoliden y se conviertan en un hábito de esparcimiento”.

El logro de estos objetivos es un desafío por varias razones. Los objetivos de la universidad son tres: docencia, investigación y difusión de la cultura. Los tres son atendidos en cada uno de los campus de la universidad por separado. Además, existe una infraestructura cultural central que no solo sirve a la universidad, sino también a la población interesada en las actividades que promueve.

En definitiva, las actividades que organiza la Coordinación de Difusión Cultural abarcan a la sociedad en general y a las diferentes dependencias universitarias. El número de espacios culturales que posee la universidad son muchos; se encuentran en diversos lugares de la ciudad de México, los estados de México e incluso en el exterior. Por esta razón, surge la necesidad de armonizar a través de un sistema con gran funcionalidad el conocimiento de los espacios y activos que existe en cada uno de ellos.

Otro aspecto es la dificultad de definir en una base de datos qué infraestructura debe estar representada y cómo, además de conocer y respetar los términos que reflejan los usos de la comunidad destinataria. Si bien es evidente a través de los objetivos definidos que la programación de actividades culturales es el interés central, no siempre es fácil separar lo que una coordinación puede ofrecer con respecto a lo que las diferentes instancias universitarias que no pertenecen a esa coordinación consideran de interés para ellos. Esta situación requiere coordinación para responder de muchas maneras a las necesidades de la universidad.

ELEMENTOS TEÓRICOS

Los recintos culturales son de diferente tipo y forma. Una sala para conciertos tiene un tamaño y una forma diferente a una sala de teatro. Cuando se programa un evento, es necesario conocer las características de los espacios y la infraestructura que contienen y que respondan a la pregunta ¿Dónde existe un espacio que permita disponer de un escenario con determinadas características y la infraestructura que facilite el espectáculo que quiero presentar? La respuesta del sistema de información explicará cada uno de los elementos de la cadena técnica que se requieren para determinado evento, como el diseño del espacio, el equipamiento y la cantidad de personal del que se dispone.

El sistema de información está conformado por registros técnicos que detallan la información que requiere una compañía externa para presentarse en determinado recinto, el modo de operar en el espacio disponible y planificar dicha presentación. Existen varios formatos de recintos culturales, donde la relación es diferente entre los actores y el público porque el diseño del espacio define en cierto modo esa relación.

Uno de los recintos culturales por excelencia es el teatro; explicar algunas de las partes que lo integran permite ejemplificar los puntos de acceso que se necesitan recuperar en una base de datos. La conformación del teatro se expone con claridad en un manual editado en Chile:

[...] existen tres áreas principales en un teatro, cuyo carácter y escala variarán según el tamaño y el tipo del recinto. Sin embargo, las características y la manera de relacionarse entre sí son muy similares en todos los casos. Éstas son:

- El escenario y la sala: es el corazón de un teatro donde se desarrolla la principal actividad de presentar y experimentar la magia de un **espectáculo**. Sin importar las dimensiones del recinto, el público se encuentra sentado para ver el escenario y el escenario es el lugar donde los **actores** se presentan ante el público. La relación entre estos dos espacios es crucial para el buen desarrollo de un espectáculo; esta relación puede variar, pero en cualquier caso debe estar pensada para que la mayor cantidad de espectadores vean y escuchen de manera óptima.
- Recepción o FOH (siglas en inglés de *front of house*, es decir la parte delantera de un recinto): es el área donde el espectador espera para entrar a la sala. Incluye también áreas de servicios como baños, cafetería y la boletería”
- Áreas de trabajo del escenario o *Backstage*: incluye las zonas de camerinos, talleres, bodegas, sala de eléctrica o de *dimmers*, sala de control, etc. En resumen, corresponde a las zonas de apoyo al trabajo en el escenario (Manual de escenotecnia, 2014: 14).

Uno de los aspectos fundamentales a cuidar en una base de datos es la terminología para transmitir información comprensible a los usuarios destinatarios. En el párrafo

anterior, se pueden apreciar las diferencias terminológicas que existen entre situaciones reales en diferentes países en que se habla la lengua española. En México, se generaliza el uso de vestíbulo en el teatro; para el lugar que en Chile se considera recepción, lo mismo sucede con la taquilla en México, llamada boletería en Chile. En México está más extendido el nombre de cabina para la llamada sala de control en Chile. Sin embargo, las denominaciones también varían en México y la definición de los elementos de la plataforma y de la infraestructura mediante vocabularios controlados para diseñar los metadatos y el llenado de estos es un aspecto importante en el que colaborará el IIBI.

La justificación terminológica de los usos lingüísticos que se integrarán al proyecto en la etapa siguiente ya están siendo trabajados y constituirán una parte importante del proyecto.

METODOLOGÍA Y DESARROLLO DE LA BASE DE DATOS

Por muchos años se daba por sentado que las bases de datos eran las que generaban productos y servicios como los catálogos, índices u otro medio que permitía el acceso a las colecciones de recursos; que apoyaban en distintas tareas y procesos administrativos. Pero el mismo desarrollo de esas bases de datos y la generación de nuevos formatos de los recursos de información llevó a comprender que los datos almacenados e interconectados entre sí son los que sirven para desarrollar productos y servicios.

El SIRUEyAC está desarrollando una base de datos que responde a un objetivo general y varios específicos. Además, atiende dos principios fundamentales: la estandarización y normalización de sus datos, razones por las cuales se observa que los datos culturales están ajustándose a lo

que Coyle (2010) refiere como la estructura fundamental de cualquier sistema de información:

- a) **Diseñar el modelo básico de los datos.** Éste responde a los requerimientos esenciales del sistema de información.
- b) **Definición de los elementos.** Cada elemento deberá establecerse de acuerdo con los requerimientos que permitan describir el concepto con miras a normalizarse para que pueda ser interoperable en el ambiente de la Web Semántica.
- c) **Definición del vocabulario.** Los datos deberán de crearse de una manera simple para que los distintos términos interactúen, pero su definición será lo más cercana a un identificador único.
- d) **Desarrollo de reglas de aplicación.** El empleo de reglas de aplicación no funcionará del todo bien en la Web si no se especifican las restricciones del metadato.

Tomando como referencia el modelo anterior, la fase que presentamos del desarrollo del SIRUEyAC es la que a continuación se menciona.

El modelo básico de los datos en este sistema de información tiene como propósito identificar los distintos tipos de espacios y activos culturales que pueden ser de utilidad a los gestores culturales como a la comunidad universitaria en general al momento de buscar entre sus datos vinculados tanto los espacios como los activos culturales con que cuenta la UNAM.

El grupo técnico de SIRUEyAC ha trabajado durante varios meses en la creación, el desarrollo y la aplicación de diversos instrumentos de recolección de datos de los espacios y

activos culturales de la Universidad Nacional con el apoyo de especialistas de la Secretaría Técnica de la Coordinación de Difusión Cultural. Los instrumentos de recolección de datos dieron el sustento técnico para la creación de plantillas y formularios, así con una estructura de metadatos inicial para crear la base de datos de SIRUEYAC.

En este mismo orden de ideas, los instrumentos de recolección se basaron en los formatos empleados por la Secretaría de Cultura para la creación de su registro de los recursos que se encuentra en el proyecto Mi Cultura.

También se revisó la plataforma del Centro de Información Cultural de Hidalgo, cuya orientación es la difusión de archivos fonográficos y hemero-bibliográficos. Otra fuente de apoyo fue Ibermuseos, que es una red iberoamericana de museos que reúne los datos museísticos; su consulta sirvió para formular el instrumento de recolección de datos para espacios culturales.

Finalmente, el grupo técnico del SIRUEYAC analizó los desarrollos de las necesidades técnicas que tiene la Coordinación de Difusión Cultural con relación a Carro de Comedias, el Festival Itinerante en Contacto Contigo, entre otras más.

La recolección de los datos se realizó en un lapso de seis meses consistentes en visitar, *in situ*, distintas entidades académicas de la UNAM. Los datos recabados se fueron almacenando en una hoja de Excel.

Siguiendo el planteamiento de Coyle (2010), la fase consistente a la definición de los elementos de la base de datos, el grupo técnico de SIRUEYAC obtuvo los datos a través del instrumento de recolección para que posteriormente se definieran las propiedades de los elementos, consistente en agrupar los datos obtenidos en dos grandes categorías: una relativa a los datos de los espacios culturales y la otra a los datos de activos culturales.

Con los datos almacenados en la hoja correspondiente, da inicio el proceso de normalización que consistió en analizar los diversos datos guardados que, a priori, sirvieron para categorizar la información dentro de la base de datos. Con un total de 757 reactivos distribuidos en nueve categorías principales es que se construye la base de datos del SIRUEyAC.

Las categorías más amplias que se rescatan de cada recinto cultural son:

- Actividades y atención administrativa.
- Infraestructura cultural.
- Acondicionamiento de las instalaciones.
- Ocupaciones.
- Medidas de accesibilidad y seguridad.
- Datos de ubicación y contacto.
- Recursos técnicos.
- Material gráfico y registro fotográfico.

Actualmente, esta base de datos se encuentra en desarrollo con base en las siguientes características técnicas: para la definición, manipulación, validez y relación de datos, se utilizó DataBase Management System (DBMS) como administrador de los datos y el My Structured Query Language (MySQL) como la estructura compleja de cientos de miles de registros.

Los procesos para obtener los datos referentes a los espacios y la infraestructura están sistematizados y definidos a partir de una batería de preguntas compuesta por 1,064 ítems con la que se implementó la metodología de OOAD (Object Oriented Analysis and Design) que resultó en el *mapeo* de todos los “conceptos abstractos” en términos de Entidades, Atributos, Relaciones y Roles que en su conjunto

definen los metadatos *descriptivos*, *estructurales* y *administrativos* del conjunto de datos.

Esta versión preliminar de la base de datos se constituye por los siguientes módulos:

- Módulo de Ingesta de datos. Relativa a los formularios y plantillas que sirven para almacenar los datos recogidos, ingresar, capturar y administrar los datos de cada recurso, sea este espacio o activo cultural.
- Módulo de administración. Contiene todas las indicaciones referentes al ingreso de usuario, sus perfiles, roles y acciones que pueden realizar tanto en la base de datos como en la consulta, selección búsqueda y obtención de datos.
- Módulo de consulta. Referente a las diversas opciones de búsqueda y recuperación de datos.
- Módulo Técnico. Contiene las diagramaciones para ilustrar los aspectos del sistema, las cuales están determinadas por el UML (Unified Modeling Language), traducido comúnmente como Lenguaje Unificado De Modelado.

Grosso modo, el UML no es un lenguaje de programación, sino una notación gráfica específica para un sistema orientado a objetos. Éste describe más de una docena de diagramas distintos; sin embargo, en este documento sólo se utilizan lo más comunes.

Las ventajas del UML, entre otros, es que se puede decir que es una manera sistemática de describir y definir los datos o la información relacionada a un “proceso de negocios” (*business process*). El proceso es modelado a través de componentes (*entidades*) que están vinculados entre sí por

medio de *relaciones* que expresan dependencias y requerimientos entre ellos, tales como: un edificio puede estar dividido en cero o más departamentos, pero un departamento sólo puede estar localizado en un único edificio. Las *entidades* pueden tener muchas propiedades (*atributos*) que las caracterizan. Los diagramas creados para representar tales *entidades*, *atributos* y *relaciones* de manera gráfica se conocen como *diagrama entidad-relación*.

Una primera definición fue realizada para establecer los tipos de espacio cultural que se distinguieron de acuerdo con los levantamientos realizados:

Figura 1. Fases de la Revolución industrial



Fuente: Hammel Scale (2018).

Por una parte, la información referente a la infraestructura de un espacio cultural: dimensiones del espacio, área total del recinto, altura de la plataforma del escenario, aforo, componentes físicos, conectividad (acceso a Internet y líneas telefónicas), un formulario de servicios y seguridad: si el recinto cuenta con salidas de emergencia, señalización en caso de sismos o incendios y recursos humanos, así como los datos de ubicación y contacto de los coordinadores de actividades culturales en la universidad. Por otro lado, se pueden consultar los recursos técnicos en audio, iluminación y video proyección e incluso instrumentos musicales. Con la idea de completar la información, se incluye la elaboración de diagramas de los espacios incluidos en la base de datos.

En suma, el SIRUEyAC de la UNAM que se encuentra en fase de desarrollo, permite que se llegue a dos supuestos, uno teórico y otro práctico. Sobre el supuesto técnico, es de observarse que en los estudios y trabajos de investigación relativos a los espacios y activos culturales están en ciernes, razón por la cual el sustento teórico es mínimo.

En lo que respecta al supuesto práctico, se advierte que sobre la creación y el desarrollo de sistemas de información en la bibliotecología y los estudios de la información hay suficientes casos en los cuales se muestran las buenas prácticas para el tratamiento de los datos culturales, museográficos y archivísticos, pero pocos relativos a los datos de espacios y activos culturales.

CONCLUSIONES

Un proyecto de esta naturaleza parte de objetivos muy generales, que desde el inicio de su desarrollo requieren rectificaciones a lo largo del camino, tal y como se comprobó

en la prueba piloto llevada a cabo en el sexto mes del diseño de la ficha. Por esta razón, cada etapa de progreso debe incluir una prueba piloto antes de seguir adelante.

La Coordinación de Difusión Cultural y los responsables de las actividades culturales de los diferentes campus de la universidad manifestaron su interés por el proyecto y colaboraron ampliamente en la complementación de la inscripción de espacios y bienes culturales.

Los órganos consultivos que desarrollaron actividades similares permitieron cumplir más y mejor las normas requeridas para la adopción de metadatos en la plataforma.

La utilización de los datos registrados en la plataforma ha supuesto el reconocimiento del trabajo realizado, así como de los comentarios realizados al respecto.

A un año de desarrollo, la RUEyAC ofrece la posibilidad de capturar, gestionar y consultar en línea un inventario sistematizado de bienes raíces, servicios, usos sociales de los espacios, actividades y medios técnicos de iluminación, proyección de audio y video, a través de formularios especializados para la recolección de datos sobre museos, explanadas y foros al aire libre, salones de baile, música y audiovisuales, cines, teatros y auditorios, cuya finalidad es elaborar, preservar, producir, programar o difundir contenidos culturales en la UNAM.

BIBLIOGRAFÍA

- Alonso Hierro, Juan y Juan Martín Fernández. 2013. “Activos culturales y desarrollo sostenible: la importancia económica del Patrimonio Cultural.” *Política y Sociedad* vol. 50, núm. 3: 1133-1147. Disponible en http://dx.doi.org/10.5209/rev_POSO.2013.v50.n3.41861.
- Ballart Hernández, Josep y Jordi Juan I Tresserras. 2001. *Gestión del patrimonio cultural*. Barcelona: Ariel.
- Bennet, Tonny, *et al.* 2009. *Culture, Class, Distinction*, Routledge.
- Bourdieu, Pierre. 1987. “Los tres estados del capital cultural”, *Sociológica* vol. 2, núm. 5
- Coyle, Karen. 2010. “Library data in the web world”, *Library Technology report* (febrero-marzo 2010).
- Gayo, Modesto. 2013. “La teoría del capital cultural y la participación cultural de los jóvenes. el caso chileno como ejemplo”, *Ultima década* vol. 21, núm. 38: (julio 2013). <http://dx.doi.org/10.4067/S0718-22362013000100007>.
- Graue, Enrique. 2017. *Línea estratégica 10. Desarrollo Integral de los estudiantes. Plan de Desarrollo Institucional 2015-2019*. Ciudad de México: Universidad Nacional Autónoma de México, 36-40.
- Manual de escenotecnia*. 2014. Publicación a cargo de Daniela Campos Berkhoff. Santiago de Chile: Consejo Nacional de la Cultura y las Artes.
- Querol, María Angeles. 2010. *Manual de gestión del Patrimonio Cultural*. Madrid: Akal.

La ciudadanía digital y sus mediciones: el caso del estado de ánimo de los tuiteros en México

ALEJANDRO RAMOS CHÁVEZ

Universidad Nacional Autónoma de México

INTRODUCCIÓN

Este capítulo se inscribe en las discusiones sobre el gobierno abierto y la ciudadanía digital, y relaciona el estudio más con el segundo de estos elementos, pues se pretende dar cuenta de la utilización social que se le da a las tecnologías de la información y comunicación. Sin embargo, se hace referencia a la vinculación existente entre este tipo de ciudadanía con el gobierno, así como con las mediciones gubernamentales que de cierta forma se intentan llevar a cabo para conocer el funcionamiento o comportamiento social que están generando a partir de la utilización de las redes sociales digitales.

En este sentido, la ciudadanía digital puede ser entendida como una que, utilizando las tecnologías de la información y la comunicación participa, esporádicamente o de forma frecuente, en la discusión y el intercambio de información relacionados con las formas en las que están siendo tratados los asuntos públicos. Esta participación también puede desencadenar procesos de intervención con el gobierno bajo el enfoque de la gobernanza para proponer alternativas para tratar los asuntos y problemas de interés social. En este punto,

bajo tendencias como la del gobierno abierto, se da cuenta de que el gobierno está cada vez más interesado en escuchar a sus ciudadanos por medio de estas herramientas digitales, y con ello facilitar y hacer más participativos, consultivos y transparentes las gestiones gubernamentales.

Para el desarrollo de este trabajo, se considera necesario abordar en un primer momento algunos elementos conceptuales de la temática general de la ciudadanía, con un mayor análisis al tipo de ciudadanía digital. Posteriormente, se abordarán algunas tendencias académicas en el estudio y abordaje teórico de los estados de ánimo desde una perspectiva más social que individual. Principalmente se analizan dos enfoques que se consideran pilares en este contexto: el de la antropología de las emociones y el de la economía de la felicidad. En tercer lugar, se analizará el caso específico de un estudio experimental que lleva a cabo el Instituto Nacional de Estadística y Geografía (INEGI) para conocer el estado de ánimo de los tuiteros en nuestro país. En el último apartado concluirá con algunas reflexiones finales.

SOBRE EL CONCEPTO DE CIUDADANÍA

Para entrar al primer tema es necesario definir la idea general de ciudadanía, para posteriormente entender al concepto de ciudadanía digital. Se puede entender el concepto de ciudadanía desde una visión más conservadora y atributiva al individuo, donde el gobierno concede a los individuos una serie de derechos y obligaciones, entre los que pueden destacar, como derechos, los de elección de representantes públicos, de asociación, de acceso a información tanto general como específicamente pública, así como el derecho de la libre manifestación de ideas.

Desde una visión más participativa, se podría entender al ciudadano como aquel individuo que constantemente interviene en el desarrollo de los asuntos públicos; es decir, como aquel individuo que no sólo es destinatario de derechos y obligaciones, sino que participa mediante la argumentación y la deliberación de ideas en el rumbo que adquieren tanto las políticas como los asuntos públicos.

En este sentido, en la última década del siglo pasado se comenzó a analizar (García-Canclini 1995; Lechner 2000; Winocur 2003a, 2003b) el tema de la ciudadanía no exclusivamente desde la perspectiva de la igualdad que conlleva la idea general del concepto, como se había hecho hasta entonces, sino desde el enfoque de las ciudadanías diferenciadas, relativas a la búsqueda del reconocimiento de los individuos a ser ciudadanos tomando en cuenta las diferencias, así como el reconocimiento de la existencia de diferentes identidades y problemáticas específicas de cada grupo, que los hace acercarse o alejarse a los principios igualitarios incluidos en el análisis general de la ciudadanía.

En este sentido, se podría ubicar a la ciudadanía digital como un tipo específico de ciudadanía que utiliza los medios tecnológicos tanto para hacer valer sus derechos y cumplir con sus obligaciones, como para informarse y participar activamente en comunidades virtuales que partiendo del mundo *online* puedan dar paso, en algunos de los casos, a actividades *offline*. En este último punto, parece haber un acuerdo general en definir precisamente a la ciudadanía digital como un ejercicio del individuo que, según Arcila, es capaz de ejercer su ciudadanía bajo un panorama de interactividad ofrecido por una plataforma digital (Arcila 2006, 18).

En este contexto, es cada vez más común el uso de aplicaciones como wikis, Facebook, Twitter, Instagram y Youtube, que están siendo utilizadas como medios para

obtener y generar información, pero también para llevar a cabo análisis políticos y sociales, pues de cierta forma se conciben cada vez más con atributos que anteriormente se aplicaban a aspectos no virtuales, como lugares de acceso a la información, encuentro social, debate, generación de opinión pública y espacios de presión política que en determinadas ocasiones pueden desatar la acción colectiva.

Algunos datos que ilustran la importancia que está adquiriendo el análisis de este tipo de ciudadanía son reflejados en los resultados del “Estudio sobre los hábitos de los usuarios de internet en México, en su edición del año 2018”. En este estudio, se menciona, entre otras cuestiones, que los usuarios de internet en México han crecido de 20.2 millones de usuarios en el año 2006 a 70 millones de usuarios en el 2016 (Asociación de Internet.mx 2018). Otro dato interesante extraído de una muestra de las 1626 entrevistas que abarcó este estudio es que los usuarios utilizan Internet en promedio ocho horas y un minuto diarios, lo que deja ya muy alejadas actividades que antes ocupaban gran parte del tiempo del ocio de las personas, como ver la televisión (con tres horas tres minutos) o escuchar la radio (con dos horas cincuenta minutos). Dentro de las actividades que más llevan a cabo los individuos en Internet está la revisión de redes sociales digitales con un 83 por ciento de menciones, enviar o recibir mails con un 78 por ciento, enviar o recibir mensajes con un 77 por ciento, y la búsqueda de información con un 74 por ciento. Estos datos son muy ilustrativos de la importancia que adquieren en la actualidad la utilización y el intercambio de información a través de las redes sociales.

Con relación al Internet y la democracia, seis de cada diez entrevistados consideraron que Internet los acerca

más a los procesos democráticos de México; la principal actividad llevada a cabo a este respecto es la búsqueda de información sobre procesos electorales, con un 92 por ciento de incidencia. De igual forma, los principales lugares para la búsqueda y la difusión de información son las redes sociales (97 por ciento), sitios de noticias (79 por ciento), buscadores (74 por ciento), además de videos o podcasts; blogs, foros y publicidad en línea con menor porcentaje (Asociación de Internet.mx 2018).

Las problemáticas relacionadas con las barreras de acceso al Internet siguen siendo: 1) la relativa poca oferta del servicio en algunas zonas del país, 2) los costos aún muy elevados de los servicios y 3) la falta de capacidades individuales para usar las tecnologías.

Con este tipo de análisis, nos damos cuenta de los esfuerzos, a manera experimental, para conocer y medir los usos políticos, democráticos y participativos que lleva a cabo la ciudadanía mediante la utilización de Internet y específicamente de las redes sociales digitales. En este trabajo interesan de forma particular los estudios de la felicidad y las mediciones de este tema y su vínculo con la ciudadanía digital.

Lo anterior da paso al segundo de los temas del trabajo, pues desde hace algunos años se ha intentado medir el bienestar y desarrollo de los pueblos y las naciones no solamente con base en datos productivos, económicos y financieros, sino con perspectivas que apuntan más a los estados de ánimo. Si bien existen críticas al mencionar que estos estudios son incompletos y subjetivos, han surgido análisis empíricos muy interesantes que tienen en cuenta la felicidad, con diversos indicadores como elementos valiosos para medir el bienestar.

ESTUDIOS SOBRE EL ESTADO DE ÁNIMO

Con el objeto de brindar un contexto en estas discusiones, se considera pertinente destacar principalmente dos de las teorías que han abordado el estudio de los estados de ánimo en su vinculación social o del desarrollo de las comunidades: la antropología de las emociones y la economía de la felicidad.

Con relación a la antropología de las emociones, existen aportes muy interesantes que han analizado el estado de ánimo en el desarrollo o atraso de diversas comunidades. Un referente es el desarrollado por Frida Jacobo (2013) en su análisis de la situación de los nahuas de Cuetzalan, Puebla, que ofrece un abanico muy amplio y detallado del desarrollo teórico de esta línea de investigación de la antropología de las emociones. Siguiendo a esta autora, se puede hacer referencia al trabajo de Evans-Pritchard (1937), en el que se describe un comportamiento emocional generalizado entre los grupos africanos de los Azande ante aspectos naturales o sociales compartidos. Desde esta perspectiva, autores como Solomon (1984) han ido más allá al sostener que las emociones pueden estar construidas más por la cultura que por cuestiones biológicas, al sostener que

[...] las emociones deben interpretarse como adquisiciones culturales, determinadas por las circunstancias y conceptos de una cultura particular, mucho más que por las funciones de la biología y, más específicamente, de la neurología. Puede haber emociones universales, pero este es un asunto que debe resolverse empíricamente, no mediante un pronunciamiento a priori (Solomon 1984, 4).

Bajo estas perspectivas, siguiendo con los aportes de Jacobo (2013), una emoción, más allá de un sentimiento individual,

también puede ser analizada como una construcción social, si partimos de que puede ser compartida mediante el lenguaje. Por lo tanto, a pesar de que los sentimientos pueden ser los mismos (amor, miedo, ira, felicidad, etcétera), existen variaciones en aspectos históricos y transculturales que los desatan, por lo cual pueden ser analizados, según Lutz (1986), como aspectos contruidos e incrustados como categorías sociales.

En términos generales, la antropología de las emociones, tomando en cuenta los aportes de Jacobo (2013), considera que las emociones se encuentran fuertemente establecidas en el ámbito estructural y funcional de lo social, pues su eje rector es visualizado en la cultura, principalmente por tres aspectos:

- 1) Las emociones cumplen una función tanto a nivel individual, como a nivel social pues poseen contenidos ideológicos que se transmiten culturalmente.
- 2) Las emociones pueden ser situadas dentro de un contexto histórico y social determinado.
- 3) Las emociones forman parte de la estructura social pues vinculan al individuo con su sociedad al mismo tiempo que la construyen.

Por otro lado, con relación a la “economía de las emociones”, se puede destacar un antecedente muy importante en el libro del premio nobel de economía en el año 2001, Joseph Stiglitz; se titula *Los felices 90. La semilla de la destrucción* (2003). En este libro, Stiglitz vislumbra un principio de siglo muy turbulento debido a los acontecimientos de la aparente estabilidad política, económica y financiera de la última década del siglo pasado en Estados Unidos de América. Ante esta situación, Stiglitz menciona que no sólo ese país sufriría los

embates de las crisis, sino que serían todos los países que dependen económicamente de él. Por lo tanto, podemos ubicar el análisis de la supuesta felicidad precisamente en la estabilidad de la última década del siglo pasado, pero a la postre sería el principio de las fuertes crisis venideras.

Para intentar definir este concepto, siguiendo las propuestas de Stordeur (2014), se puede mencionar que

La economía de la felicidad consiste en una rama de la economía de la conducta que desafía también algunas de las asunciones centrales de la economía tradicional y busca, básicamente por medio del empleo de reportes de utilidad obtenidos en encuestas, recabar información sobre las preferencias de las personas de una manera alternativa a la típica de la economía tradicional. Este enfoque se basa, entonces, en encuestas donde las personas reportan su nivel de felicidad en una escala finita para luego emplear herramientas estadísticas con la finalidad de establecer correlaciones entre diversas variables y niveles de felicidad (Stordeur 2014, 86).

Un elemento que desató el interés por medir el bienestar desde otros enfoques consistió en los cuestionamientos más serios que se les hizo a los economistas tradicionales, quienes veían relación directa y casi exclusiva entre el nivel de ingreso y el bienestar, pues un gran número de estudios demostraron que las personas podían ser cada vez más ricas, pero no necesariamente más felices. Bajo estas perspectivas, podemos destacar los aportes de otro premio nobel de economía, este de 1998, Amartya Sen, quien en parte de su obra se interesó por estudiar no sólo la disponibilidad de satisfactores, sino las capacidades de poderles sacar provecho, que nos llevarían a las realizaciones de satisfacción. Sen (2000), bajo estas perspectivas, se aleja

de la visión dominante en la economía de libertad negativa (relacionada con el no impedimento de acción) y avanza hacia la libertad positiva; es decir, las capacidades reales de un individuo de ser o hacer algo.

Estos planteamientos han sido tomados en consideración cada vez con mayor importancia por organismos internacionales para medir el grado de bienestar de las naciones. Un ejemplo específico son los estudios llevados a cabo por la Red para el Desarrollo Sostenible de las Naciones Unidas, que desde el año 2012 lleva a cabo un estudio denominado “Reporte de la felicidad” o “World Happiness Report”, cuya versión del año 2018 (Helliwell, Layard, y Sachs 2018) comparó a 156 naciones, tomando en consideración seis variables específicas: ingreso, expectativa de vida saludable, apoyo social, confianza, libertad y generosidad.

En los resultados de este último estudio, y teniendo en cuenta que México resulta un país muy mal calificado en cuanto a la percepción de corrupción¹, además de los problemas de pobreza, inseguridad y violencia, se ubica en sitio veinticuatro de los 156 países tomados en consideración. Es el segundo país más feliz de América Latina, apenas superado por Costa Rica, que ocupa la posición número trece. Según los especialistas, la felicidad de gran parte de los países de América Latina radica principalmente en la calidez y fortaleza de las relaciones familiares, las cuales son vistas como condicionantes para alcanzar la felicidad sobre otros aspectos económicos o políticos.

1 Según el Índice de Percepción de Corrupción llevado a cabo por Transparencia Internacional, México, en el reporte del año 2017, ocupa el lugar 135 de las 180 naciones analizadas (Transparency International 2018), mientras que en el 2010 se ubicaba en el lugar 98, en el 2000 en el 59 y en 1995 en el 32. Es decir, ha sido constante el incremento en la percepción de la corrupción en el país a lo largo de la realización de este estudio, lo que de cierta forma se contrapone con la idea general de la felicidad, en la cual México constantemente se encuentra en el primer cuartil de las posiciones.

Existen estudios similares que, a manera experimental, están intentando medir la felicidad de las comunidades virtuales. Tal es el caso del estudio del estado de ánimo de los tuiteros en México, que será abordado de forma general en el siguiente apartado del trabajo.

ESTUDIO DE CASO: ESTADO DE ÁNIMO DE LOS TUITEROS EN MÉXICO

Desde el año 2015, el Instituto Nacional de Estadística y Geografía (INEGI) lleva a cabo el estudio del estado de ánimo de los tuiteros en el país. Lo anterior representó, según las instituciones participantes en este proyecto,

[...] un paso hacia el uso de fuentes alternativas de *big data* para generar nuevas estadísticas experimentales. Se logró concretar una aplicación, para el ámbito de Bienestar Subjetivo, en la que se presentaba el resultado de todo el ciclo necesario en el uso de técnicas de *big data* utilizando fuentes no tradicionales como lo son las redes sociales. Este ciclo abarca desde la recolección de la fuente de datos, el análisis de éstos, su preprocesamiento, la clasificación y el procesamiento de cada dato, su almacenamiento, la generación de indicadores, hasta la representación visual para difundir dichos resultados” (Instituto Nacional de Estadística y Geografía 2017, 1).

En este análisis se incursiona en aplicaciones prácticas de Ciencia de Datos, definida por William Cleveland (2007) como un campo que incluye análisis de datos, modelos estadísticos, métodos de construcción de modelos, métodos de estimación para realizar inferencia estadística, sistemas

de *hardware* y *software*, algoritmos computacionales, herramientas estadísticas, entre otras.

Lo anterior fue utilizado por el INEGI con el objeto de brindar estudios de bienestar subjetivo utilizando Twitter y fuentes de *big data* para acercarse a determinar el estado de ánimo de los tuiteros del país.

En este contexto, en el documento metodológico del estudio se menciona que Twitter es una red social en donde los usuarios pueden escribir textos cortos de hasta doscientos ochenta caracteres (originalmente eran ciento cuarenta) que quedan visibles de forma pública. Los mensajes pueden ser georreferenciados para conocer la ubicación exacta desde donde sale el texto. Cabe señalar que para este estudio del estado de ánimo, sólo se utilizan mensajes que tengan ese atributo geográfico con objeto de no solo conocer el estado de ánimo general (del país), sino de cada una de las entidades federativas.

En este estudio, se siguió una metodología de *machine learning* (aprendizaje automático) muy interesante para generar la estadística del estado de ánimo de los tuiteros. En un primer momento, se etiquetaron múltiples veces, con objeto de alcanzar un consenso en la etiqueta, cuatro mil tuis por más de cinco mil estudiantes de la Universidad Tec Milenio. Posteriormente, se llevó a cabo una limpieza y normalización de tuits para disminuir el desorden en las calificaciones. En este punto, se eliminaron contradicciones y repeticiones y se identificaron los tuits con mayor consenso en su etiqueta. También se utilizó un proceso de normalización con objeto de corregir errores, anonimización de usuarios, aprovechamiento de emoticones e identificación de sintaxis. Con respecto a los *emoticones*, se clasificaron quinientos doce de uso popular en cuatro clases (positivo, negativo, neutro y ninguna) (INEGI 2017).

Posteriormente, se llevaron otros procesos de definición de conjuntos de entrenamiento y validación, desarrollo y entrenamiento de clasificadores automáticos, clasificación masiva de tuits, además de un procesamiento para distinguir tuiteros locales de visitantes. Esto representó para los desarrolladores retos específicos en cada etapa para generar la herramienta utilizada en la actualidad, para clasificar los tuits con carga positiva y aquellos con carga negativa.

Para aprovechar de mejor manera la aplicación, el programa no sólo brinda las gráficas del estado de ánimo de los tuiteros del país por día, semana o mes, sino que las gráficas son acompañadas con datos muy interesantes para, de cierta forma, conocer o intuir el porqué del estado de ánimo generalizado en el país en una región o en un estado determinado, así como la posibilidad de comparar el estado de ánimo de los tuiteros entre dos entidades fedrativas distintas o entre una de ellas con el global nacional.

Otro dato que también es puesto a disposición por el programa es un gráfico relativo al número de tuits que son escritos por hora del día, lo que ofrece una visualización de los horarios de mayor uso de esta red social en el país. De igual forma, es posible visualizar en una nube, realizada por día, los temas que son tendencia, mejor conocidos como *trending topics*, que se van formando a partir de las etiquetas más utilizadas o *hashtags*. Lo anterior da muestra del trabajo profundo de metadatos que son utilizados por estas estadísticas con la intención de identificar los temas del momento que pueden influir en el estado de ánimo de los usuarios de esta red social. Finalmente, otro recurso que brinda esta aplicación es el de conocer las noticias del día que pudieron influir en el sentimiento colectivo, por lo que la aplicación da la posibilidad de acceder a las noticias

que acontecieron en un día determinado directamente en el buscador de noticias de Google o en la página kiosko.net.

Sin lugar a dudas, este estudio abre una perspectiva innovadora en cuanto al análisis de la situación de la población del país, pues los datos no sólo se basan en aspectos económicos o censales, sino que toman en consideración la perspectiva de los sentimientos de la población, específicamente reflejada en la carga negativa o positiva en la que son escritos y son compartidos los tuits. Lo anterior abre una perspectiva amplia tanto en el uso, como en el estudio de los metadatos y su incidencia en el estudio de lo social.

REFLEXIONES FINALES

El concepto de ciudadanía ha evolucionado de su concepción como conjunto de atribuciones otorgadas por el gobierno (derechos y obligaciones), a uno mucho más activo que implica un involucramiento constante de los ciudadanos en la forma en la que son atendidos los asuntos públicos y formuladas las políticas. De igual forma, han surgido enfoques especializados que parten de la noción de ciudadanas diferenciadas, dependiendo del tipo específico y las condiciones particulares del ciudadano que sea analizado. En este trabajo interesó particularmente el ciudadano digital, entendido como aquel individuo que ejerce sus derechos y obligaciones mediante la utilización de las tecnologías de la información y la comunicación, de forma específica Internet, además de llevar a cabo procesos participativos vinculantes con la vida política y social de su comunidad; procesos que inclusive pueden llegar a generar una acción colectiva.

De igual forma, el tema de la medición del bienestar de los habitantes de las naciones ha transitado de su análisis en función, casi con exclusividad, del ingreso y las cuestiones ligadas a la riqueza (o su contraparte la pobreza), a estudios que intentan medir el bienestar de las poblaciones con otros indicadores. En este sentido, los estudios del estado de ánimo han cobrado vigencia e importancia en el análisis tanto académico como de agencias internacionales encargadas de medir el desarrollo de las naciones.

Finalmente, han surgido análisis muy interesantes, como lo es el caso analizado sobre la felicidad de los tuiteros en México, que en forma de estadística experimental están llevando a cabo trabajos que vinculan las temáticas de la ciudadanía y su medición en función del estado de ánimo, con lo que han generado indicadores muy valiosos en términos sociales y políticos brindados a partir de la utilización de *big data*.

En este contexto, se abre una posibilidad muy amplia de análisis de metadatos con el objeto de identificar patrones que reflejen la situación o las situaciones en las que se encuentren las poblaciones de una nación. En este trabajo, interesó de forma particular el estado de ánimo reflejado en la utilización de las redes sociales, específicamente en Twitter; sin embargo, estos estudios pueden extenderse tanto a otras redes sociales, así como a otros tópicos de análisis que pueden incluir minería de datos para el análisis político, de injerencia pública o de vinculación gubernamental con la ciudadanía para el tratamiento y cause de los asuntos públicos. Si bien estos temas quedan como derroteros de investigaciones futuras, van estableciendo en la actualidad nuevas formas de comportamiento que dan vida a las situaciones políticas, sociales, económicas y culturales

de las naciones, pues el potencial de la información para el análisis y entendimiento de lo social que pueden generar los metadatos, la utilización y apropiación de las herramientas digitales por parte de la ciudadanía, así como la utilización de estas herramientas por parte del gobierno para llevar a cabo la gestión pública y el acercamiento con sus ciudadanos son ya aspectos que forman parte integral del entendimiento de la situación de los pueblos y las naciones.

BIBLIOGRAFÍA

- Arcila, Carlos. 2006. "El ciudadano digital", *Revista Latinoamericana de Comunicación CHASQUI*, núm. 93: 18-21.
- Asociación de Internet.mx. 2018. "Estudio sobre los hábitos de los usuarios de Internet en México 2018". Disponible en <http://www.asociaciondeinternet.org.mx/es/component/remository/Habitos-de-Internet/lang,es-es/?Itemid=>.
- Cleveland, William S. 2007. "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics", *International Statistical Review*, núm.69, vol. 1: 21-26. <https://doi.org/10.1111/j.1751-5823.2001.tb00477.x>.
- Evans-Pritchard, Edward Evan. 1937. *Witchcraft, Oracles and Magic among the Azande*. Oxford: Oxford University Press.
- García-Canclini, Néstor. 1995. *Consumidores y ciudadanos*. Ciudad de México: Grijalbo.
- Helliwell, John, Richard Layard y Jeffrey Sachs. 2018. *World Happiness Report 2018*. Nueva York: Sustainable Development Solutions Network.
- Instituto Nacional de Estadística y Geografía. 2017. *Estado de ánimo de los tuiteros en los Estados Unidos Mexicanos*. Documento metodológico Segunda edición. Disponible en <http://www.beta.inegi.org.mx/app/biblioteca/ficha.html?upc=702825099718>.
- Jacobo-Herrera, Frida. 2013. *Hacia una antropología de las emociones. La atención de la envidia entre los nahuas de Cuetzalan, Puebla*. Tesis de doctorado, Ciudad de México: CIESAS.

- Lechner, Norbert. 2000. "Nuevas Ciudadanías", *Revista de estudios sociales*, núm. 5: 25-31.
- Lutz, Catherine. 1986. "Emotion, Thought, and Estrangement: Emotion as a Cultural Category", *Cultural Anthropology* núm.1, vol. 3: 287-309.
- Sen, Amartya. 2000. *Desarrollo y libertad*. Buenos Aires: Planeta.
- Solomon, Robert C. 1984. «Getting Angry: The Jamesian Theory of Emotion in Anthropology». En *Not Passion's Slave*, editado por Robert C. Solomon, 1-23. Oxford: Oxford University Press. Disponible en <http://oxfordindex.oup.com/view/10.1093/0195145496.003.0006>, <http://oxfordindex.oup.com/view/10.1093/0195145496.003.0006>.
- Stiglitz, Joseph E. 2003. Los felices 90: *La semilla de la destrucción*. Barcelona: Taurus.
- Stordeur, Eduardo. 2014. "Felicidad, bienestar y políticas públicas: algunas consideraciones preliminares", *Revista Teoría del Derecho de la Universidad de Palermo*.
- Transparency International. 2018. *Corruption Perceptions Index 2017*. Nueva York.
- Winocur, Rosalía. 2003a. *Ciudadanos en los medios. La construcción de lo público en la radio*. Buenos Aires: Gedisa.
- 2003b. "La invención mediática de la ciudadanía". En *Espacio público y reconstrucción de ciudadanía*, editado por Patricia Ramírez, 231-52. México: Miguel Ángel Porrúa.

El papel de los metadatos en la construcción de datos enlazados para bibliotecas

EDER ÁVILA BARRIENTOS

Universidad Nacional Autónoma de México

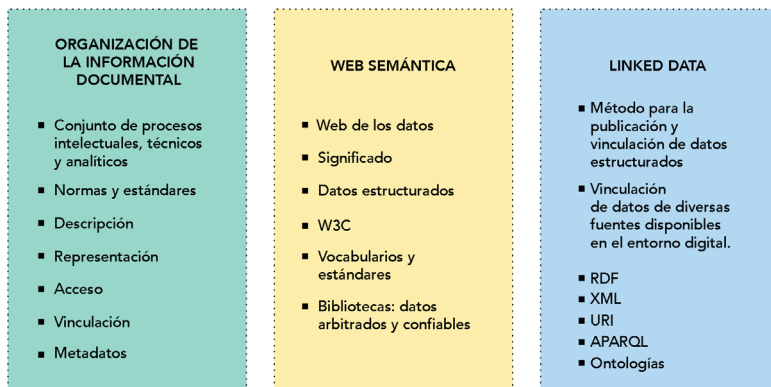
Introducción

El fenómeno de los datos enlazados puede abordarse desde dos campos de conocimiento de la Biblioteconomía y los Estudios de la Información, desde la perspectiva de las tecnologías de la información y desde la visión de la organización de la información documental.

En la última década, la comunidad científica se ha concentrado en el desarrollo de tecnologías y estándares que dinamicen el crecimiento de la web semántica, lo que ha permitido la construcción de una web dotada de mucho más significado, la cual puede ser procesado tanto por los humanos como por las computadoras.

Los datos enlazados ponen de manifiesto la generación de un modelo innovador para la organización y vinculación de la información. En este modelado, los metadatos juegan un papel esencial en la generación de datos enlazados, pues son el mecanismo que permite la descripción y vinculación de los datos. Es importante señalar que los datos enlazados pueden generarse en diversas fuentes o contextos. Uno de ellos es el de las bibliotecas.

Ilustración 1.
Elementos introductorios de los datos enlazados



Fuente: Elaboración propia 2018.

El objetivo de este trabajo es analizar el papel de los metadatos en la construcción de datos enlazados enmarcados en el entorno web de las bibliotecas. Para ello, se abordan los principios de datos enlazados y su presencia en el ambiente de las bibliotecas; se estudia la función de los metadatos y su rol en la generación de datos enlazados, y se proponen alternativas para la generación de datos enlazados desde las bibliotecas. Pues los metadatos y la generación de datos enlazados son una integración que otorga la posibilidad de generar nuevos métodos para el acceso a los datos y la información que están disponibles en el entorno de la web.

A su vez, se presentan los desarrollos actuales que tienen que ver con la interacción de los datos enlazados en las bibliotecas. Se abordan las propuestas y actividades de grupos de investigación a nivel internacional, los cuales tienen el propósito de proponer pautas para la generación e interoperabilidad de los datos enlazados en el ambiente de las bibliotecas.

REVISIÓN DE LA LITERATURA

¿Qué función tienen los metadatos en la generación de datos enlazados para las bibliotecas? La interacción de los datos enlazados y los metadatos en el ambiente de las bibliotecas ha generado estudios rigurosos de índole teórica y procedimental. Algunos de estos estudios exponen análisis a plataformas de datos enlazados existentes en la actualidad como es el caso de DBpedia, Wikidata, Europeana y Datos.BNE.es (Vila Suero y Gómez Pérez 2013). Estas plataformas son las más representativas respecto a la integración de datos enlazados y los metadatos en un ambiente de información documental.

Uno de los proyectos más significativos de la aplicación de datos enlazados en bibliotecas es Europeana. Este proyecto reúne contribuciones digitalizadas de reconocidas instituciones culturales de los países miembros de la Unión Europea. Se trata de una biblioteca digital que está soportada en un portal web que proporciona una interfaz que brinda acceso a millones de libros, pinturas, películas, objetos de museo y registros de archivos digitalizados de toda Europa. El proyecto inicial comenzó en el año 2005 en la Comisión Europea de Sociedad de la Información y Medios (EC i2010 Iniciativa de Bibliotecas Digitales) y su desarrollo actual, con una dotación presupuestaria de 120 millones de euros, corre a cargo del Community programme eContentPlus de la Comisión Europea y es supervisado y coordinado por la Fundación Biblioteca Digital Europea (EDL Foundation - European Digital Library Foundation) ubicada en Ámsterdam.

La organización de los datos de Europeana se lleva a cabo mediante la utilización de un modelo de datos enlazados que

han denominado “modelo de datos de Europeana” (EDM por sus siglas en inglés). La adopción de EDM permite a Europeana ser compatible con la web semántica utilizando los estándares del W3C. EDM está constituido por la utilización de normas como RDF, RDFS (Resource Description Framework Schema), OAI-ORE (Open Archives Initiative Object Reuse and Exchange), SKOS y DCmiterms (Dublin Core Metadata Initiative Terms).

En el 2011, se publicó el documento Library Linked Data Incubator Group Final Report, en el cual se planteaban las siguientes recomendaciones para trabajar con el desarrollo de datos enlazados en las bibliotecas:

That library leaders identify sets of data as possible candidates for early exposure as Linked Data and foster a discussion about Open Data and rights; That library standards bodies increase library participation in Semantic Web standardization, develop library data standards that are compatible with Linked Data, and disseminate best-practice design patterns tailored to library Linked Data; That data and systems designers design enhanced user services based on Linked Data capabilities, create URIs for the items in library datasets, develop policies for managing RDF vocabularies and their URIs, and express library data by re-using or mapping to existing Linked Data vocabularies;

That librarians and archivists preserve Linked Data element sets and value vocabularies and apply library experience in curation and long-term preservation to Linked Data datasets (W3C s.f., s.p.).

Este documento es altamente significativo para el estudio que se desarrolla en este trabajo, pues en él se plantea el estado de la implementación de los datos enlazados en el ambiente de las bibliotecas. Se explica la naturaleza de los datos enlazados que están disponibles en las biblio-

tecas y se mencionan sus atributos y fuentes de suministro. Derivado de este estudio, se han generado una amplia gama de estudios de caso en donde se describe la implementación de los principios de datos enlazados conjuntamente con los vocabularios y normas que son utilizadas en el contexto de las bibliotecas. Como parte de este estudio, pueden rescatarse las siguientes observaciones:

- a) Los datos de la biblioteca no están integrados con los recursos disponibles en la web.
- b) Los estándares de la biblioteca están diseñados solo para la comunidad bibliotecaria.
- c) La comunidad bibliotecaria y la comunidad de la web semántica tienen una terminología diferente para conceptos de metadatos similares.
- d) Los cambios tecnológicos de la biblioteca dependen del desarrollo de los sistemas pertenecientes a proveedores comerciales.

Las observaciones anteriores permiten obtener una visión compleja de la integración de los datos enlazados, los metadatos y las bibliotecas en un ecosistema digital común. Sin embargo, dentro de la literatura se han localizado algunos esfuerzos que pretenden alcanzar la interacción de los metadatos con los principios de datos enlazados en un entorno bibliotecario.

Por ejemplo, Baker (2012) planteaba la interacción entre los lenguajes de descripción de las bibliotecas y los principios de datos enlazados. Este estudio es pertinente pues plantea la integración de Dublin Core con la estructura de RDF para fomentar el intercambio de datos entre las bibliotecas. Asimismo, se plantea que “the translation of library standards into RDF involves the separation of languages of

description from the specific data formats into which they have for so long been embedded” (Baker 2012, 130).

En el 2013, Dydimus enfatiza en la creciente necesidad de contar con bibliotecarios especializados en el tratamiento de los datos y en el procesamineto de metadatos al señalar que

Libraries are encouraged to learn from others and also publish their data so that they benefit from “being of the web” and reach out to as many users as possible. It is therefore argued that there is great need for librarians to transform their expertise in working with metadata into expertise in working with ontologies or models of knowledge so that libraries can fully participate as information providers in the digital age (Dydimus 2013, 86).

Como parte de la preparación profesional del bibliotecario en lo que respecta a la gestión de datos y el uso de metadatos, surgieron diversos estudios que proponían la formación de personal altamente especializado que fuera encargado del desarrollo de proyectos de datos enlazados (O’Dell 2015). Son escasos los estudios que reúnen aspectos relacionados con la presentación de proyectos de datos enlazados en bibliotecas; la mayoría de ellos se concentra en el tratamiento de recursos de información documental mediante el uso de metadatos, sobre todo con el uso del esquema Dublin Core.

Mitchell (2016) realizó un estudio exploratorio de las tendencias de datos enlazados que fue aplicado a las bibliotecas. En este estudio, concluye que el avance de la aplicación de Linked Data en las bibliotecas se documenta en proyectos técnicos, generación de vocabularios y se sugieren prácticas para el desarrollo de datos enlazados en las bibliotecas (Mitchell 2016,33).

La función de los metadatos en la generación de datos enlazados para bibliotecas es un asunto que carece de una fundamentación teórica y procedimental; este aspecto puede analizarse dentro de la literatura especializada. Sin embargo, los escasos estudios que abordan este tópico sugieren que es posible llevar a cabo una vinculación de ambos elementos para alcanzar el desarrollo de datos enlazados de índole bibliotecaria.

METADATOS Y DATOS ENLAZADOS

Berners-Lee (2018) definió cuatro principios básicos para la publicación de datos enlazados:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs. So that they can discover more things (Berners-Lee 2018).

De esta manera es viable publicar datos que cumplan sólo los tres primeros principios, pero el hecho de no aplicar el cuarto los convierte en menos visibles y, como consecuencia, menos reutilizables.

Los datos enlazados pueden considerarse como un modelo estructurado para la publicación y vinculación de datos disponibles en diferentes fuentes de la web. Tal y como manifiestan Sakr y colegas, “Linked Data uses RDF, the Resource Description Framework, as its basic data model. RDF provides the means to describe resources in a

semi-structured manner. The information expressed using RDF can be exchanged and processed by applications” (Sakr et al. 2018, 12).

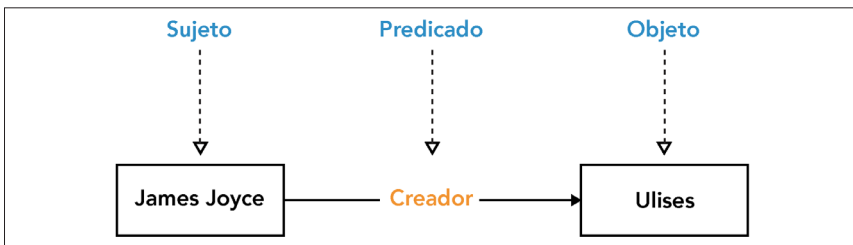
Los datos enlazados tienen una serie de componentes que permiten la construcción de datos estructurados con un significado previamente definido. A continuación se mencionan estos elementos:

- XML. Sintaxis superficial para documentos estructurados.
- XML Schema. Lenguaje para definir la estructura de los documentos XML.
- RDF. Modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos.
- RDF Schema. Vocabulario para describir las propiedades y las clases de los recursos RDF con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases.
- URI. Identificador de recurso uniforme. Se trata de una cadena de caracteres que identifica de manera unívoca los recursos disponibles en la web.
- SPARQL. Lenguaje de consulta sobre RDF que permite hacer búsquedas sobre los recursos de la Web Semántica utilizando distintas fuentes de datos.
- OWL. Web Ontology Language. Añade más vocabulario para describir propiedades y clases como relaciones entre clases, cardinalidad, igualdad, tipologías de propiedades más complejas, caracterización de propiedades o clases enumeradas.
- SKOS. Simple Knowledge Organization System. Sistema Simple de Organización del Conocimiento.

Se trata de un modelo para la representación de la estructura básica y el contenido de esquemas de conceptos como tesauros, esquemas de clasificación, listas de encabezamientos de materia, taxonomías, folksonomías y otros vocabularios controlados similares. Al tratarse de una aplicación de RDF, SKOS permite la creación y publicación de conceptos en la web, así como vincularlos con datos en este mismo medio e incluso integrarlos en otros esquemas de conceptos.

- **TRIPLES RDF.** La construcción de tripletes con RDF se basa en la idea de declarar recursos usando la expresión en la forma sujeto-predicado-objeto. Esta expresión se conoce en la terminología RDF como triple.
 - a) **Sujeto.** Persona, entidad corporativa o nombre de familia a la que nos referimos.
 - b) **Predicado.** Es la propiedad o vinculación que se desea establecer acerca del sujeto.
 - c) **Objeto.** Es el valor de la propiedad o del otro recurso con el que se establece la vinculación.

Ilustración 2. Representación gráfica de un triple básico



Fuente: Elaboración propia 2018.

La construcción de triples se logra mediante la aplicación e integración de los componentes mencionados con anterioridad. Al conjunto de triples se les identifica como *datasets* (conjunto de datos). Dichos conjuntos integran un contexto en específico que se relaciona con otros contextos mediante la vinculación de sus atributos. La construcción de estas relaciones tiene gran relevancia al momento de realizar una determinada búsqueda sobre un recurso, autor o tema en específico en el entorno web. En este sentido, los metadatos tienen una función muy importante, pues hacen posible representar y estructurar los datos que están disponibles en las bibliotecas.

Los metadatos descriptivos son un conjunto de elementos que hacen posible representar y describir los atributos de los recursos de información documental. Tradicionalmente, los metadatos han sido utilizados en las bibliotecas para organizar y describir a los recursos que forman parte de las colecciones documentales. Rodríguez García (2013) resume la utilidad de los metadatos de la siguiente manera:

- Permiten describir diversos atributos de los recursos.
- Describen el contenido, forma o atributos de un recurso de información.
- Proporcionan y proveen servicios de información a los usuarios.
- Su principal función es facilitar la identificación, localización, recuperación, manipulación y uso de los recursos de información digital.

El uso de datos enlazados hace posible la publicación de datos, los cuales pueden ser migrados desde diferentes fuentes en las que se encuentren almacenados, por ejemplo,

bases de datos, repositorios, hojas de cálculo y catálogos en línea, siempre y cuando se utilicen normas interoperables para comunicar datos que han sido debidamente estructurados. Los atributos con los que deben contar los datos enlazados son los siguientes:

- a) Deben ser datos consistentes, debidamente estructurados y libres de inexactitudes.
- b) Los datos deben contener un significado previamente establecido; para ello se utilizan lenguajes documentales, ontologías y normas computacionales.
- c) Los datos deben responder a una demanda, ya sea dentro de un contexto o fuente de datos en específico.
- d) Los datos siempre deberán vincularse con fuentes externas, fomentando con ello un ecosistema interoperable de datos en la web.

Para la conformación de datos enlazados es necesario utilizar esquemas de metadatos que permitan la estructuración uniforme de datos y su respectiva descripción, representación, accesibilidad y vinculación. La integración de datos enlazados y metadatos permite vincular los atributos de un recurso mediante su descripción y representación con elementos de un determinado esquema. Esto es posible mediante la estructura semántica que pueda tener un esquema de metadatos en particular. La semántica de los datos es proporcionada mediante el uso de normas y estándares de índole internacional. Los datos enlazados tienen una estructura básica llamada “Triple”, la cual está conformada por tres elementos esenciales: sujeto, predicado y objeto.

Cada dato tiene la capacidad de representar los atributos de un sujeto, predicado y objeto definido en una sentencia de vinculación. Los metadatos proporcionan la estructura descriptiva que tendrán los datos y la asignación de un URI a cada dato y ofrecen la posibilidad de vincular los datos disponibles en diferentes fuentes de la web que tengan patrones similares. Por lo tanto, es posible descubrir las diferentes expresiones y manifestaciones derivadas de una obra en específico.

Tabla 1.

Ejemplo de un triple con asignación de URI y elemento Dublin Core.

Sujeto	Predicado	Objeto
http://www.libraryld.edu/aut/JamesJoyce	http://purl.org/dc/elements/1.1/creator	"Ulises"

Fuente: Elaboración propia 2018.

En la tabla 1 puede apreciarse una ejemplificación de un triple al cual le han sido asignados los URIs de un sistema hipotético y se ha utilizado el elemento *creator* del esquema de metadatos Dublin Core para asignar el predicado que permite la vinculación entre los datos. Como puede observarse, cada uno de los datos del triple ha sido codificado y estructurado acorde a una estructura lógica.

El papel de los metadatos para la generación de datos enlazados radica en las siguientes acciones:

- Los metadatos permiten describir a los datos que conformarán los triples.
- Los metadatos pueden contener elementos semánticos en su estructura. Esta característica posibilita la capacidad de los datos para vincularse.

- c) Los metadatos permiten nombrar las vinculaciones que tendrán los datos que se desean estructurar.
- d) La estructuración y vinculación de los datos puede desarrollarse mediante la utilización de metadatos.

Por lo tanto, los metadatos son un eje fundamental para el establecimiento de datos enlazados en las bibliotecas; los datos enlazados pueden fomentar la vinculación de contenidos en el entorno web y en consecuencia incrementar la participación de las bibliotecas en una web más interoperable, usable y accesible para el usuario final.

GENERACIÓN DE DATOS ENLAZADOS EN BIBLIOTECAS

La disponibilidad de los datos en el entorno digital es una realidad en los campos científicos y académicos. Los datos son representaciones numéricas, alfabéticas, algorítmicas, etcétera, que permiten identificar a un determinado recurso de información. La riqueza de datos y las herramientas generadas para su búsqueda, análisis, recuperación y visualización, permiten a los investigadores académicos interpretar cada vez mayores cantidades de datos.

La generación de datos enlazados en bibliotecas tiene como objetivo crear un método para vincular datos estructurados y contenidos que puedan estar disponibles en el entorno de la web. La principal fuente de datos en las bibliotecas son los registros bibliográficos y de autoridades con los cuales se representan a los recursos de información disponibles en las diversas colecciones.

Algunos esquemas de metadatos han generado propuestas que integran la utilidad de los principios de datos enlazados para desarrollar datos enlazados de índole bibliotecaria. Uno de estos esquemas es Dublin Core, el cual proporciona una serie de “recomendaciones para expresar los metadatos de Dublin Core en sentencias RDF” (Dublin Core s.f., s.p.).

Tabla 2. Estructura descriptiva de datos enlazados

Estructura descriptiva para obtención de datos enlazados disponibles en un libro impreso

Sujeto	Predicado	Objeto
http://www.libraryld.edu/aut/JamesJoyce	http://purl.org/dc/elements/1.1/creator	Ullises
http://www.libraryld.edu/title/book/Ullises	http://purl.org/dc/elements/1.1/title	Ullises
http://www.libraryld.edu/literaturacontemporanea	http://purl.org/dc/elements/1.1/subject	Ullises
http://www.libraryld.edu/debolsillo	http://purl.org/dc/elements/1.1/publisher	Ullises
http://www.libraryld.edu/josemariavalverde	http://purl.org/dc/elements/1.1/contributor	Ullises
http://www.libraryld.edu/2004	or http://purl.org/dc/elements/1.1/date	Ullises
http://www.libraryld.edu/texto	http://purl.org/dc/elements/1.1/type	Ullises
http://www.libraryld.edu/impreso	http://purl.org/dc/elements/1.1/format	Ullises
http://www.libraryld.edu/974p	http://purl.org/dc/elements/1.1/extent	Ullises
http://www.libraryld.edu/español	http://purl.org/dc/elements/1.1/language	Ullises
http://www.libraryld.edu/isbn8497930967	http://purl.org/dc/elements/1.1/identifier	Ullises
http://www.libraryld.edu/DLB27493	http://purl.org/dc/elements/1.1/rights	Ullises

Fuente: Elaboración propia, 2018.

En la tabla 2, se pueden consultar los datos que han sido codificados en un sistema hipotético¹ de identificadores uniformes de recursos. Estos datos tienen el potencial de vincularse mediante una relación significativa previamente establecida, como es el caso de la utilización de los elementos descriptivos de Dublin Core.

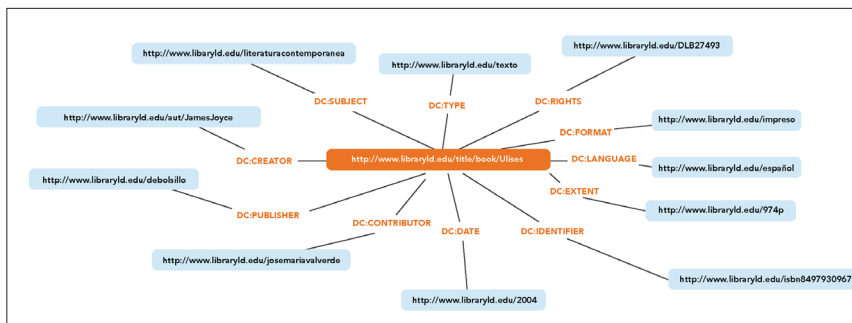
Se ha analizado que el esquema de metadatos Dublin Core además de proporcionar un esquema para la descripción de

¹ En la actualidad existen sistemas computacionales que permiten gestionar la generación de URIs en una determinada unidad de información.

recursos, también permite obtener una estructura de relaciones de significado al momento de aplicar e integrar los elementos de metadatos con los principios de Linked Data.

La estructura descriptiva de datos enlazados de un libro impreso está conformada por los elementos de metadatos básicos de Dublin Core. Dicha estructura puede extenderse y especificarse con mayor alcance dependiendo de los atributos del recurso de información que analizará. De esta manera, el establecimiento de vinculaciones permitirá descubrir los datos y contenidos que tienen patrones significativos de relación. La integración de los metadatos y los datos enlazados en un entorno común permite la generación de un mecanismo para descubrir el universo documental que rodea a un determinado dato.

Ilustración 3. Grafo de los datos enlazados con Dublin Core correspondientes a un libro impreso



Fuente: Elaboración propia 2018.

En la ilustración 3, se presenta el grafo de los datos enlazados con Dublin Core del libro impreso que fue descrito con anterioridad. Esta representación gráfica permite obtener una visión cercana al establecimiento de vinculaciones entre los datos disponibles en una descripción de un

libro. Es deseable que los datos representen el contenido de los recursos de información, pues de esta manera será posible establecer conexiones entre contenidos de diversa naturaleza pero son patrones y atributos similares.

El desarrollo de datos enlazados es un proceso sistemático e intelectual para la generación de datos estructurados capaces de vincularse significativamente. Asimismo, el desarrollo de datos enlazados de índole bibliográfica de autoridad mediante el análisis y procesamiento de registros pone de manifiesto el desarrollo de políticas que permitan planificar el procesamiento de los datos dentro de las bibliotecas.

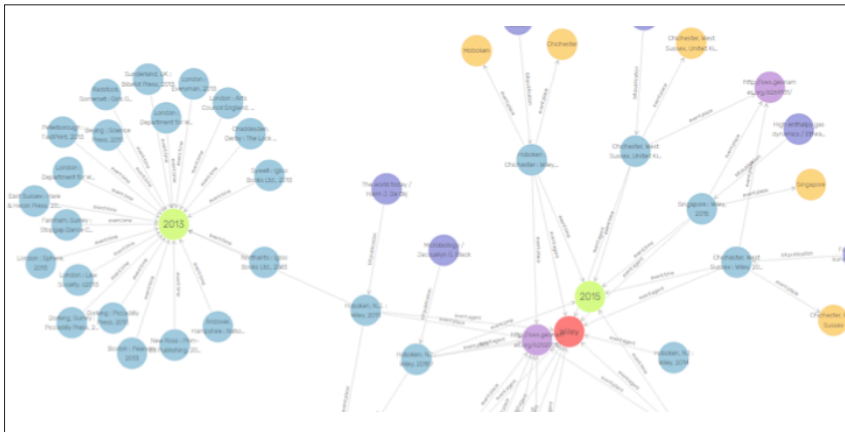
En este sentido, las bibliotecas además de incursionar en el ámbito de la web deberán adaptarse a los cambios vertiginosos de la tecnología y la generación de datos en diversas fuentes. Se ha identificado que la generación de datos enlazados en las bibliotecas requiere cumplir con las siguientes acciones:

- a) Considerar a los datos como una fuente para atender complejas demandas de información dentro de la biblioteca. En la actualidad los datos son productos de investigaciones científicas y académicas que son de gran relevancia para el desarrollo de innovaciones e investigaciones en diversos campos de conocimiento. Las bibliotecas deberán adaptar sus procesos y servicios para organizar, representar y proporcionar acceso a estos datos.
- b) Los datos disponibles en las bibliotecas deben estar perfectamente estructurados, pues de ello dependerá su vinculación con otras fuentes de datos disponibles en la web. De esta manera, será posible gestar redes de colaboración entre

bibliotecas y otras entidades mediante el uso y explotación de datos enlazados.

La generación de datos enlazados en bibliotecas pone de manifiesto la creación de perfiles semánticos de metadatos. Estos perfiles son estructuras descriptivas y representativas de los datos disponibles en registros bibliográficos y de autoridad que remiten a recursos de información documental.

Ilustración 4. British Library. Linked Data Set



Fuente: Elaboración propia 2018.

La Biblioteca Británica permite apreciar la aplicación de las tecnologías de los datos enlazados en sus servicios a través de sus servicios de metadatos. Esta biblioteca actualmente se encuentra desarrollando una versión de la Bibliografía Nacional Británica que se pone a disposición de los usuarios como datos abiertos enlazados a través de una plataforma. La propuesta inicial de esta plataforma incluye datos enlazados sobre monografías y publicaciones seriadas que corresponden a las colecciones de la Biblioteca Británica.

En el modelo de datos enlazados para monografías de la Biblioteca Británica, es posible apreciar el ejemplo de un recurso de información y las múltiples relaciones que pueden establecerse a través de sus atributos. Cada una de las relaciones permite descubrir nueva información a través de los atributos de un solo recurso. A su vez, puede apreciarse la categorización de estas relaciones mediante núcleos de tema, título, autor, datos de publicación y series. Cada uno de estos núcleos responde a los atributos de la monografía ejemplificada en el modelo.

El set de datos enlazados perteneciente a la bibliografía nacional de la Biblioteca Británica (véase ilustración 4) permite descubrir las vinculaciones entre datos de índole bibliográfica. Estos conjuntos de datos están codificados en RDF y utilizan una serie de vocabularios para describir los atributos del recurso. El procesamiento del conjunto de datos fue llevado a cabo mediante el Software GraphDB, el cual permite obtener una interfaz de captura de datos en donde se presenta la estructura básica de un triple, tal y como puede observarse en la ilustración 5.

Ilustración 5. Graph DB. Module description LD set

	subject	predicate	object	context
1	http://brb.data.bl.uk/id/agent/AcceleratedEducationPublications	rdf:type	dct:Agent	http://www.ontotext.com/explicit
2	http://brb.data.bl.uk/id/agent/AcceleratedEducationPublications	rdf:type	foaf:Agent	http://www.ontotext.com/explicit
3	http://brb.data.bl.uk/id/agent/AvantBooks	rdf:type	dct:Agent	http://www.ontotext.com/explicit
4	http://brb.data.bl.uk/id/agent/AvantBooks	rdf:type	foaf:Agent	http://www.ontotext.com/explicit
5	http://brb.data.bl.uk/id/agent/ExpressPublishing	rdf:type	dct:Agent	http://www.ontotext.com/explicit
6	http://brb.data.bl.uk/id/agent/ExpressPublishing	rdf:type	foaf:Agent	http://www.ontotext.com/explicit
7	http://brb.data.bl.uk/id/agent/FacetPublishing	rdf:type	dct:Agent	http://www.ontotext.com/explicit
8	http://brb.data.bl.uk/id/agent/FacetPublishing	rdf:type	foaf:Agent	http://www.ontotext.com/explicit
9	http://brb.data.bl.uk/id/agent/Ginn	rdf:type	dct:Agent	http://www.ontotext.com/explicit
10	http://brb.data.bl.uk/id/agent/Ginn	rdf:type	foaf:Agent	http://www.ontotext.com/explicit
11	http://brb.data.bl.uk/id/agent/Heinemann	rdf:type	dct:Agent	http://www.ontotext.com/explicit
12	http://brb.data.bl.uk/id/agent/Heinemann	rdf:type	foaf:Agent	http://www.ontotext.com/explicit
13	http://brb.data.bl.uk/id/agent/Ladybird	rdf:type	dct:Agent	http://www.ontotext.com/explicit

Fuente. Elaboración propia, 2018.

GraphDB (Ontotext s.f.) es desarrollado por Ontotext; se trata de un *software* libre para establecer datos enlazados de diferente naturaleza y magnitud. GraphDB permite consultar la descripción de los datos enlazados mediante un módulo específico. Como se muestra en la ilustración 5, el módulo de descripción utiliza URIs para asignar la representación, organización y descripción de los datos mediante una estructura fundamentada en las características de un triple simple. Además contiene un apartado para el contexto, ya que el significado de los datos debe estar relacionado con un contexto en específico. Cada contexto de datos representa a un dominio de triples y estos pueden conformar un conjunto de datos enlazados en específico.

DESARROLLOS FUTUROS

El desarrollo de los datos enlazados en las bibliotecas depende del grado de avance en la aplicación de procesos que permitan generar datos de esta índole en las bibliotecas, pues los datos enlazados son un desarrollo tecnológico que puede impactar en los servicios y la organización de la información disponible en éstas.

Se ha identificado que el avance de los datos enlazados depende de los siguientes factores:

- a) Desarrollo y adopción de vocabularios normalizados en las bibliotecas que permitan la construcción de datos enlazados.
- b) Integración de los esquemas de metadatos tradicionales con los principios de datos enlazados.

- c) Generación de *software* interoperable que permita la gestión y construcción de datos enlazados en las bibliotecas.
- d) Flexibilización de las bibliotecas para adoptar nuevas tecnologías en sus estructuras.
- e) Identificación de fuentes de datos con potencial para llevar cabo su vinculación en el entorno digital.

Asimismo, se ha identificado que el desarrollo de Bibframe favorecerá la generación de datos enlazados en las bibliotecas. Bibframe es un modelo de datos de descripción bibliográfica, que ha nacido como parte de una iniciativa de la Biblioteca del Congreso de los Estados Unidos. Bibframe se encuentra en su versión 2.0 y se presenta como una propuesta que pretende evolucionar las normas de descripción bibliográfica a un modelo de datos vinculados, con el fin de hacer que la información bibliográfica sea más útil tanto dentro como fuera de la comunidad bibliotecaria (Library of Congress s.f.).

El desarrollo y la adopción de Bibframe en el contexto de las bibliotecas facilitará la comprensión de los datos enlazados en este ambiente y permitirá descubrir el potencial que un dato bibliográfico tiene al momento de vincularse con una fuente disponible en el entorno digital, pues para la web semántica, los datos disponibles en las bibliotecas son gran valor debido a su calidad y relevancia para atender las complejas demandas informativas de los usuarios.

Para IFLA, el desarrollo de datos enlazados en las bibliotecas es un asunto de importancia para fomentar la participación de las bibliotecas en el desarrollo de la web semántica. Por ello, será necesario considerar el avance y las actividades futuras los grupos Linked Data for Libraries y Linked Data Special Interest Group de la IFLA.

Mediante el seguimiento de estos grupos, será posible descubrir los avances inmediatos de la implementación de los datos enlazados en las bibliotecas y mantener actualizaciones de las implementaciones significativas de datos enlazados en las bibliotecas.

CONSIDERACIONES FINALES

El papel de los metadatos en la construcción de datos enlazados en las bibliotecas consiste en proporcionar elementos para la construcción de una estructura semántica normalizada que permita la descripción, representación, accesibilidad y vinculación de los datos. Los datos disponibles en las bibliotecas son de gran valor para la generación de datos enlazados, pues ofrecen una fuente confiable, arbitrada y consistente de datos de índole bibliográfica, de autoridad y temática.

El desarrollo de la implementación de los datos enlazados en las bibliotecas depende de factores tecnológicos y técnicos relacionados con la adopción de nuevos sistemas digitales en las bibliotecas y de la flexibilidad normativa que las bibliotecas puedan ejercer en sus estructuras.

En este sentido, la organización de la información en las bibliotecas está evolucionando. La participación de las bibliotecas en la web semántica es necesaria pues mediante los principios de la organización de la información y la integración de los principios de datos enlazados pueden obtenerse entornos generados por las bibliotecas. Esta situación incrementará la visibilidad y participación de las bibliotecas en ecosistemas digitales de datos que fomenten la vinculación de fuentes confiables de datos que pueden ser de gran utilidad para las comunidades de usuarios.

BIBLIOGRAFÍA

- Baker, Thomas. 2012. "Libraries, languages of description, and linked data: a Dublin Core perspective", *Library Hi Tech*, vol. 30, no. 1: 116-133.
- Berners Lee, Tim. *Linked Data*. Disponible el 13 de junio de 2018 en <https://www.w3.org/DesignIssues/Linked-Data.html>.
- Dublin Core Metadata Initiative. *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. Disponible el 14 de junio de 2018 en: <http://dublincore.org/documents/dc-rdf/>.
- Dydimus, Zengenene. 2013. "Global interoperability and linked data in libraries", *New Library World*, vol. 114, núm. ½: 84-87.
- Library of Congress. *Overview of the BIBFRAME 2.0 Model* [en línea]. Disponible el 18 de junio de 2018 en <https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
- Mitchell, Erik T. 2016. "Library Linked Data: Early Activity and Development", *Library Technology Reports*, vol. 52, no. 1.
- O'Dell, Allison. 2015. "Planning for Linked Data: Recruitment, Training, and Workflow Design for Resource Description and Metadata Management". En: *Library Staffing for the Future*, 245-261. Estados Unidos: Emerald.
- Ontotext. *Graph Database*. Disponible el 18 de junio de 2018 en <https://ontotext.com/products/graphdb/>.
- Rodríguez García, Ariel Alejandro. 2013. "El aprovechamiento de los metadatos en las bibliotecas", *Revista e-Ciencias de la Información*, vol. 3, núm. 1: 1-13.

- Sakr, Sherif *et al.* 2018. *Linked Data: Storing, Querying and Reasoning*. Switzerland: Springer.
- Vila Suero, Daniel y Asunción Gómez Pérez. 2013. “Datos. bne.es and MARiMbA: an insight into library linked data”, *Library Hi Tech*, vol. 31, núm. 4: 575-601.
- W3C Incubator Group Report. 2005. *Library Linked Data Incubator Group Final Report* Disponible el 13 de junio de 2018 en: https://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/#The_library_community_and_Semantic_Web_community_have_different_terminology_for_similar_metadata_concepts.

La integración del usuario final en la creación de metadatos

PATRICIA HERNÁNDEZ SALAZAR

Universidad Nacional Autónoma de México

INTRODUCCIÓN

Una de las funciones esenciales de los metadatos es apoyar a los usuarios en el entendimiento de la complejidad de los datos y su utilización para su transformación en información. La creación de metadatos está pensada para cubrir las necesidades de al menos dos categorías de comunidades, los profesionales de los metadatos y los usuarios finales; en términos generales, estos últimos están totalmente ajenos al proceso de generación de metadatos. Algunos especialistas pertenecientes a empresas e instituciones académicas que crean metadatos han percibido como obstáculo para su adecuada explotación no considerar al usuario final.

Este capítulo tiene por objetivo determinar la pertinencia de que el individuo interesado en recuperar información o usuario final se incluya en los equipos que trabajan con metadatos. Para cubrirlo, se tocan tres temas principales.

El primer apartado de este trabajo, “Concepto y categorías de usuarios de metadatos”, delimita el sentido del término usuario de la información y presenta una primera clasificación de las comunidades iniciales, intermedias y finales vinculadas a los metadatos.

En la segunda sección, titulada “Importancia del usuario final en la creación de metadatos”, se analizan algunos textos que visualizan las necesidades y los comportamientos del sujeto como un elemento básico para generar metadatos. Por último, se examina un “Estudio sobre la integración de usuarios finales en la creación de metadatos”, el cual se ubica en un contexto de producción de conocimiento científico, la población estudiada es la comunidad de investigadores del Centro para la Detección Integrada en Red de la Fundación Nacional de Ciencia de Estados Unidos de América.

CONCEPTO Y CATEGORÍAS DE USUARIOS DE METADATOS

La investigación sobre cualquier tema requiere una delimitación conceptual de los términos básicos para estructurar la argumentación. Se entiende por usuario de la información a aquella persona o grupo de personas que utilizan o utilizarán información para cubrir una necesidad identificada. La categorización de este sujeto se realiza con base en alguna o algunas de sus características específicas como edad, nivel de estudios, actividad principal, objetivo de uso y situación económica, entre otras.

Al empezar a explorar el binomio usuarios-metadatos, se identificaron dos categorías, los profesionales de los metadatos y los usuarios finales. La particularidad que los crea es su participación en el proceso de generación de los metadatos;

derivadas de esta característica se identificaron además otras tres clases: iniciales, intermedias y finales.

¿Cómo se llegó a esta clasificación? Una primera búsqueda sobre el tema arrojó recursos que aluden a personas relacionadas con empresas que crean, administran y utilizan datos y metadatos para entornos de negocios; el manejo de lenguaje varía en lugar de base de datos hablan de almacén de datos (data warehouse). La mayoría de estas publicaciones emanan de diversas empresas dedicadas a producir herramientas que permiten la organización de información como SAS The Power to Know, International Business Machines (IBM) o The Data Warehousing Institute.

Un data warehouse es un almacén electrónico donde generalmente una empresa u organización mantiene una gran cantidad de información. Todos los componentes, procesos y datos son (o deberían ser) rastreados y administrados desde un repositorio de metadatos, éste sirve como una ayuda tanto al diseñador como al administrador del almacén de datos. Es un sistema muy complejo, el volumen de datos registrados es vasto y los procesos empleados para su extracción, transformación, limpieza, almacenamiento y agregado son numerosos y sensibles a los cambios y las variaciones temporales.

Los metadatos para negocios muestran “a los usuarios no técnicos [finales] donde encontrar la información dentro del almacén de datos, de donde viene y como llegó allí, describe sus cualidades, y proveen asistencia para interpretarlos” (Bentley s.f., s.p.), permiten confiar en los datos o en su manera de estar almacenados.

Entre sus funciones están “incrementar la accesibilidad, la calidad, la credibilidad y la usabilidad de almacenes de datos de múltiples maneras” (Bentley s.f., s.p.) y “decirles” a los usuarios (y programadores) dónde encontrar los datos exactos que necesitan y ayudarles a entender lo que significan.

Esto permite fundamentar las tres categorías de usuarios de metadatos mencionadas supra: iniciales, intermedios y finales. Los iniciales son creadores, diseñadores y programadores; los intermedios, administradores y técnicos y los finales, los sujetos que buscan y recuperan información para cubrir una necesidad específica y que está fuera del proceso de creación de metadatos. Los dos primeros corresponden a los profesionales de los metadatos.

Esta aproximación conceptual de los metadatos dentro de un contexto de negocios nos haría creer que en ese ámbito tienen clara la importancia de los metadatos; sin embargo, en una encuesta realizada en 1998 por el Instituto de Almacenamiento de Datos (The Data Warehousing Institute) entre gerentes de almacenes de datos, se les preguntó sobre la posibilidad de crear metadatos para resolver problemas de información para la toma de decisiones. En ese momento, sólo el 25 por ciento estaba realizando alguna solución con metadatos, el 21 por ciento tenía pensado hacerlo pero no lo había hecho y el 54 por ciento no tenía planeado implementar alguna solución en la que se consideraran los metadatos; el argumento para no hacerlo era que la recuperación de la inversión no era segura (Bentley s.f.). A dos décadas, se percibe que ha ido en aumento el interés, sobre todo por parte de grandes empresas montadas en la web, las facilidades tecnológicas de estas plataformas posibilitan el seguimiento de las huellas (registros) que dejan las personas que usan ciertos productos, y la creación de datos y metadatos.

IMPORTANCIA DEL USUARIO FINAL EN LA CREACIÓN DE METADATOS

La literatura sobre la importancia del usuario final de los metadatos es poca. A continuación se describen cuatro

textos, dos que ubican a los metadatos dentro del contexto de negocios y los *data warehouse* y dos en el campo de las bases de datos.

El primero de los relacionados con negocios se titula “Metadatos para usuarios-finales: una parte crucial de tu política de calidad en tu Data Warehouse” (s.f.). El objetivo fue discutir sobre los productos de la empresa SAS The power to know para resolver problemas del usuario final. Clasifica a los usuarios en dos categorías, administradores del almacén de datos y usuarios finales; apunta a que existen diferencias en cuanto al tipo de metadatos que requieren: Administradores: fuente, tablas de objetivos, formatos de conversiones, código de salida del usuario y cálculos.

Usuarios finales: origen de las variables, papel de la variable, su exactitud y sugerencias para utilizarla en los reportes; número o nombre de la pantalla del sistema operacional donde puedan ver valores. Existen dos puntos en los que los usuarios finales requieren de ayuda: la cláusula dónde y la selección de variables; por ejemplo, la variable que contiene los valores que se necesitan para algún reporte. Se evidencia que los requerimientos varían de acuerdo con el objetivo de uso y la actividad principal de los sujetos que utilizan los metadatos.

Dentro de este contexto de negocios, Neil Foshay publicó en 2005 un estudio auspiciado por IBM y Gavilan Asociados en Investigación titulado “La influencia del usuario final de metadatos sobre las actitudes del usuario y uso de, un *data warehouse*”, donde se identifican los siguientes propósitos:

- Entender los tipo(s) de información (metadatos) que el usuario final necesita.
- Desarrollar una teoría que considere la forma en que los metadatos apoyan a los usuarios finales.

- Determinar, mediante retroalimentación con usuarios-finales de negocios, si los metadatos contribuyen al éxito del data warehouse (Foshay 2005, 2).

Foshay realizó la investigación en dos etapas. La primera fue para apuntalar el marco teórico, entrevistó telefónicamente a autores de los documentos que analizó, las preguntas fueron:

- ¿El usuario final de los metadatos contribuye de forma importante para el éxito de una iniciativa de *data warehouse*?
- ¿Qué tipos o categorías de metadatos son más importantes para el usuario-final? (Foshay 2005, 4).

Los autores entrevistados coincidieron en que:

- [...] en la mayoría de los casos, el usuario final de los metadatos es importante para el éxito del almacén de datos, pero indirectamente.
- [...] una medida básica del éxito de los almacenes de datos es el grado en el que son adoptados y usados por la audiencia meta.
- [...] los metadatos pueden influir en las actitudes del usuario hacia los datos almacenados, lo cual puede intervenir en el nivel de uso del almacén de datos. (Foshay 2005, 4).

El autor tipifica los metadatos en definicional¹, de calidad², de navegación³ y de origen⁴ de los datos, lo que denomina linaje.

En la segunda etapa, aplicó una encuesta vía Internet a las dos categorías ya mencionadas, los profesionales de metadatos y los usuarios finales (sujetos dedicados a los negocios). Resalta la pregunta relacionada con la utilidad de los datos; al igual que en el estudio de Bentley, hubo diferencias en cuanto al tipo de metadatos que requieren, los primeros le dieron un mayor peso a los definicionales y los segundos a los de calidad.

En un rubro denominado otros factores, se refuerza la calidad de los datos, aparecen dos elementos la utilidad de la herramienta denominada inteligencia de negocios usada para acceder al almacén de datos y la capacitación recibida sobre su contenido, el usuario final bien a bien no se entera puntualmente para qué le sirven los metadatos, por lo que requiere ser formado en el entendimiento y explotación de los mismos.

Estos resultados permiten establecer que existe una diferencia clara entre los requerimientos de los profesionales de metadatos (creadores, técnicos, administradores) y los usuarios finales, por lo que resulta válida la propuesta de considerar al usuario final en la creación de metadatos. Esto

-
- 1 Cualquier información que comunica el significado de los datos a los usuarios finales: definiciones, cálculos, reglas de negocios y valores admisibles.
 - 2 Aconseja a los usuarios sobre la actualidad (frescura), exactitud, validez o totalidad de los datos. Permite cerciorarse si les interesa o no utilizar los datos para un propósito específico.
 - 3 Ofrece la posibilidad de buscar y localizar los datos que necesitan.
 - 4 Posibilita a los usuarios identificar la procedencia original de los datos y su transformación antes de formar parte del warehouse. (Foshay 2005, 21).

puede hacerse de tres maneras: el estudio de sus necesidades, la incorporación en el proceso de creación lo que los ubicaría como iniciales-finales o ambas.

Una de las conclusiones en la que se advierte la trascendencia de considerar a los usuarios finales es cuando afirma que “Los metadatos no son genéricos; necesitan ser hechos a medida para contextos organizacionales específicos” (Foshay 2005, 1).

Ya dentro del ámbito de las bases de datos, Jane Greenberg publicó en 2003 el artículo “Generación de metadatos: procesos, gente y herramientas”, referido a las bibliotecas digitales. Si bien no aparece puntualmente la frase usuarios finales, sí aparecen contenidos relacionados con ellos. Establece cuatro clases de personas que se involucran en la generación de metadatos: creadores profesionales, creadores técnicos, creadores de contenidos y comunidad o entusiastas del tema.

Entre los creadores profesionales están catalogadores, indizadores, personas con un nivel alto de entrenamiento en el área y administradores de web (*web masters*).

En cuanto a los técnicos en la creación de metadatos, son profesionales que han sido capacitados para crear metadatos. Con creadores de contenido se refiere a personas encargadas de generar contenido intelectual, como investigadores, artistas o autores de textos y obras que pueden proveer metadatos mediante una plantilla, esto se reconoce como metadatos generados por el autor, se han hecho investigaciones exploratorias sobre el tema y se ha comprobado que tienen un elevado grado de efectividad. La comunidad o los entusiastas del tema no tienen entrenamiento formal en la creación de metadatos, pero tienen un conocimiento sobre el tema y quieren apoyar en el proceso.

Se evidencia que hace más de una década ya existía un interés por considerar a los usuarios finales en el proceso de

elaboración de metadatos, los creadores de contenido y la comunidad o entusiastas del tema, los cuales pueden representar a los usuarios finales.

Estudios más recientes son los publicados por el Grupo de Trabajo de la Asociación para la Investigación sobre Metadatos de Archivos de la Web del Centro de Cómputo para Bibliotecas en Línea (OCLC por sus siglas en inglés de Online Computer Library Center). El primero fue el artículo titulado “Desarrollar metadatos de archivos web buenas prácticas para satisfacer las necesidades de los usuarios”, publicado en 2017, que reporta un estudio cuyo objetivo fue entender las necesidades y el comportamiento del usuario.

Mediante una revisión de la literatura sobre el tema, se identificaron buenas prácticas neutrales en tres aspectos, comunidad, programas y salidas, encontraron y examinaron veinticinco documentos; es de notar que el número es mínimo en relación con la producción general sobre metadatos.

Utilizaron artículos, reportes, publicaciones en blogs, encuestas, memorias de conferencias y presentaciones en diapositivas.

Los resultados se presentaron en dos grandes apartados: caracterización de los usuarios y sus áreas de necesidad. La característica básica identificada fue su actividad principal, de la que se desprendió una sola clase los académicos y cuatro subclases: investigadores académicos (en general); investigadores en derecho, humanistas digitales y analistas de datos. Las necesidades derivan de las preguntas de investigación que ingresan a los archivos; los comportamientos se agruparon en tres tipos: de navegación (búsqueda de sitios específicos); informativo (búsqueda de información sobre un tema), y transaccional (adquisición de algún elemento del sitio web).

Las necesidades más recurrentes fueron:

- Interoperabilidad, su falta impide agregar y analizar contenidos de sitios que son preservados en múltiples espacios.
- Formación para acceder a archivos de sitios que no están diseñados de forma amigable, su explotación muestra un alto grado de complejidad y pierden efectividad las herramientas de descubrimiento.
- Posibilidad de buscar y recuperar en múltiples archivos, que tengan un acceso unificado basado en sistemas de descubrimiento.
- Comunicación con archivistas y bibliotecarios que activamente trabajan en el área de creación de sitios, archivos y metadatos.
- Colaboración entre investigadores, entre investigadores y creadores de archivos, y de forma interinstitucional para evitar duplicar el trabajo y enriquecer la normalización de los archivos.

Entre los textos que analizaron, hubo algunos relacionados con los profesionales de los metadatos, entre las conclusiones resalta el hecho que falta compartir sus prácticas con la comunidad y la ausencia de la evaluación de algunos tipos de metadatos por parte del usuario. Los administradores de sitios web como usuarios perciben como un reto compartido “la falta de aproximaciones comunes para crear metadatos” (Dooley *et al.* 2017, 1)

El segundo texto del grupo de OCLC, publicado en 2018, se titula “Metadatos descriptivos para archivos web: revisión de la literatura de las necesidades del usuario”, producto ampliado de la investigación anteriormente mencionada. Aumentó el

número de documentos a 64, es fácil afirmar que sigue siendo una cantidad poco significativa para el grueso de literatura sobre metadatos. Los textos correspondían a investigaciones realizadas en esta segunda década en países como Portugal, Nueva Zelanda, Reino Unido y China, entre otros.

Su interés fue captar información sobre tipos de usuarios finales, sus métodos de investigación, barreras para utilizar los archivos web y los servicios de apoyo que necesitan. Consideraron dos grupos de comunidades, los profesionales de los metadatos y los usuarios finales. Aquí solo se retoma lo relacionado con usuarios finales.

Establecieron una primera clasificación: usuarios institucionales (bibliotecas nacionales, archivos, e instituciones públicas) y personales (investigadores académicos, postdoctorales, e independientes; profesores, graduados, estudiantes y público en general). En cuanto a disciplinas estuvieron historia, literatura, sociología, lingüística, periodismo, humanidades digitales, ciencias de la computación, derecho y profesionales de la información.

Otra clasificación fue con base en las funciones o actividades principales en el proceso de creación de datos y metadatos para la web: propietarios, diseñadores, analistas de datos, científicos de la web, especialistas en humanidades digitales, e instructores. Se reitera la clasificación de los tipos de comportamiento en la búsqueda: navegación, información y transaccional.

En relación con los tipos de metadatos que requerían, sobresalen: procedencia (cómo, por qué, cuándo y de quién) y preservación. Existe un énfasis especial en los siguientes aspectos: procedencia de los datos (les permite determinar su autenticidad y la confianza para utilizarlos); detalles técnicos de las capturas como la incoherencia temporal de capturas irregulares; curadurías, y acceso, reúso y publica-

ción de la información contenida, lo que nos lleva a un asunto de derechos de autor.

Se reiteran la formación en la explotación adecuada de los archivos y los metadatos de los archivos, y el diseño de herramientas amigables en las interfases para hacer posible que el contenido sea más rápidamente accesible y reutilizable.

El estudio de Venlet (Venlet *et al.* 2018) tiene un apartado titulado “Involucrar a la comunidad” en el que reafirman la exigencia de educar a los usuarios sobre el acceso y reuso de los datos contenidos en los archivos web y que haya colaboración y comunicación entre los usuarios finales y los creadores o diseñadores web “para crear metadatos”.

Entre los documentos que examinaron, hubo uno hecho a investigadores. Entre sus respuestas, afirmaron su deseo de “contribuir en la creación de sus propios metadatos de los archivos que utilizaban y que la mayoría de los usuarios quieren trabajar con aquellos objetos de la web, para enriquecerlos y hacer anotaciones en cualquier nivel que sea apropiado para su análisis” (Venlet *et al.* 2018, 11).

Por último, para que los usuarios finales tuvieran un mayor entendimiento sobre los metadatos descriptivos, solicitan que los archivos web contengan un apartado específico sobre metadatos.

ESTUDIO SOBRE LA INTEGRACIÓN DE USUARIOS FINALES EN LA CREACIÓN DE METADATOS

Si existe poca literatura sobre el tema de la importancia de considerar las necesidades del usuario final para la generación de metadatos, su inclusión como parte del equipo de trabajo es aún menor. Se examina una de las investiga-

ciones más acabadas titulada “Realidades de los metadatos para la ciberinfraestructura: autores de datos como creadores de metadatos”, publicada en 2011, que corresponde a una tesis de doctorado elaborada por Matthew Stephen Mayernik.

La importancia del texto radica en que estudia comunidades de investigadores dentro de un entorno académico universitario de producción científica. Estas comunidades crean y utilizan datos y metadatos de manera constante por su actividad. En términos generales, los metadatos se extraen de textos, situación que no siempre es aplicable en los datos científicos o el entorno de producción de conocimiento.

El trabajo explora las prácticas cotidianas de gestión de los datos y metadatos de los investigadores del Centro para la Detección Integrada en Red, de la Fundación Nacional de Ciencia de Estados Unidos de América con sede en la Universidad de California. Específicamente examinó las formas en que los investigadores crean, describen, anotan, organizan, administran, documentan y archivan metadatos en sus espacios de trabajo (laboratorios y escenarios de trabajo de campo).

Se plantearon las siguientes preguntas de investigación:

¿Cómo y dónde se crean los metadatos, por quién y con qué propósito [...]?

¿Cómo son aprendidas las actividades de creación de metadatos y distribuidas en los grupos de investigación?

¿Cómo se traducen las actividades de metadatos locales en la creación de metadatos para repositorios compartidos? (Mayernik 2011, 68, 89 y 90)

Realiza un estudio etnográfico mediante el seguimiento de un método y una técnica de corte cualitativo, la observación participante (esto pudo lograrse porque el autor forma parte del personal del centro) y la entrevista respectivamente, y una técnica cuantitativa consistente en un análisis bibliográfico de diversos textos.

Los sujetos de estudio formaban parte de grupos de investigación en ecología, biología acuática, ciencias del ambiente y sismología.

Clasificó a los responsables de la generación de metadatos en cuatro categorías, autores de datos (científicos, educadores, estudiantes y otras personas involucradas en la investigación que produce datos digitales), administradores de datos (organizaciones, sujetos), científicos de los datos (ingenieros, especialistas en aspectos computacionales, responsables de la operación y mantenimiento de las bases datos, curadores, bibliotecarios y archivistas), y usuarios de datos (comunidades científicas y de educación). En esta clasificación aparecen dos grupos que interesan: los autores de los datos, propiamente los investigadores o grupos de investigadores y los usuarios finales. Resulta difícil aceptar y formalizar estas categorías puesto que de suyo esta actividad es compleja, es una acción práctica que debería responder a necesidades específicas de grupos de sujetos, en este contexto los metadatos son:

1. [...] siempre indizables y selectivos.
2. [...] representados de acuerdo con ocurrencias particulares de situaciones inmediatas en una manera adecuada para realizar tareas de investigación próximas.
3. [...] abarcan la negociación de significados compartidos. Son creados con la expectativa que

el lector o usuarios de las descripciones tengan el conocimiento para leerlos e interpretarlos.

4. [...] “explicables” para sus creadores. Los científicos u otros investigadores tienen la posibilidad de “explicar” por qué las descripciones de los metadatos son o no son creadas para sus datos, y “explicar” la selectividad de esas descripciones. (Mayernik 2011, 67)

Esta caracterización de los metadatos permite apreciar que se vuelve crítico en el momento de su generación, pensar en lo sujetos que eventualmente los utilizarán.

Como primer acercamiento, se les solicitó que comentaran sobre los referentes de cada campo mediante la técnica de pensamiento en voz alta. Posteriormente se les preguntó sobre su experiencia para desarrollar la tarea y sobre los campos que les parecieron más y menos útiles, los campos que pudieran aumentarse, y los beneficios de crear metadatos. Entre los resultados, sobresalen:

- Para los investigadores, considerados como autores de datos, la creación de metadatos es una actividad empírica situada, que varía de proyecto en proyecto y depende de laboratorios, tecnología usada, normas disciplinarias y habilidades de los individuos.
- Los procesos de metadatos, y los productos de metadatos creados como parte de esos procesos, están íntimamente ligados a la gente, las organizaciones, al equipo, al programa y a los objetos digitales” (Mayernik 2011, 240).
- Existen diferencias de tipos de metadatos de

acuerdo con la disciplina, a manera de ejemplo se listan los de dos grupos de investigación:

- Sismología. Mapas, números asignados del sitio para cada estación, nombres de los directorios de archivos y guiones, lecturas de sistemas de posición global GPS (Global Position Systems), notas en cuadernos o en laptops, correos, registros de los digitalizadores, equipos para medir los sismos, entre otros.
 - Ciencia ambiental. Lista de embalaje, etiquetas de botellas, discusiones personales sobre el lugar para tomar muestras, unidades de medida escritas en cuadernos, nombres de los lugares de las muestras, ancho entre las medidas de profundidad de las corrientes.
- Los metadatos permiten a los investigadores utilizar sus datos para alcanzar sus metas de investigación inmediatas.
 - La distribución de las tareas de metadatos es un proceso informal en el que participan todos los integrantes de un grupo de trabajo, investigador(es), técnicos y estudiantes.
 - El aprendizaje para la creación de metadatos por parte de los estudiantes se hace en espiral; los estudiantes más avanzados les enseñan a los principiantes, y es a prueba y error.
 - Los metadatos generados por estas comunidades pueden ser agrupados en seis tipos: identidad, características, calidad, equipo, métodos, y análisis

de métodos.

- Es viable y recomendable la integración de usuarios autores de datos en los equipos de creación de metadatos; la falta de experiencia formal se compensa con el entendimiento de la disciplina y sus procesos de creación de conocimiento.
- Se requiere desarrollar programas de formación y explicación de la creación de sistemas de metadatos.
- La ambigüedad de los límites alrededor de los conjuntos de datos y la fluidez de la prospectiva de los usuarios y los usos de los datos sugiere que el material y las actividades de capacitación requerirán directrices que consideren el enfoque del proceso de creación de metadatos.
- Las tensiones entre el conocimiento individual y grupal permitieron percibir que se requiere investigar sobre métodos para la creación de metadatos que incluyan contribuciones individuales y grupales.
- El primer nivel de utilización de los metadatos es del grupo de trabajo, posteriormente se comparten, aunque no es su interés hacerlo.

CONCLUSIONES

Un área poco trabajada en el fenómeno metadatos es el usuario final, desde la identificación de sus necesidades y comportamientos como base para la generación de metadatos, hasta su integración en el proceso de creación.

Su inclusión apoyaría el principio que expresa que los metadatos “Necesitan ser lo suficientemente simples para

que los humanos puedan leerlos y entenderlos”(McCarthy 1982, 6).

Se requiere involucrar al usuario final en la creación de metadatos para verificar su entendimiento de lo que espera.

La búsqueda en metadatos debería ser transparente, como si fueran registros de índices, tesauros o comandos, empezar en cualquier punto específico y desplegar resultados.

Existen requerimientos críticos por parte de los usuarios, procedencia de los metadatos, su interoperabilidad y la colaboración entre diversos actores.

Es necesario incluir ayudas para el usuario mediante la consideración de información sobre cada herramienta de comando y cada aspecto de los metadatos como entidades explicativas separadas.

Las necesidades del usuario final requieren ser tomadas en cuenta para la generación de metadatos.

Es interesante percibir que los profesionales de la información son considerados como miembros de los equipos que crean metadatos en dos de los textos examinados; habrá que trabajar más en el reconocimiento de la importancia de su labor.

BIBLIOGRAFÍA

- Bentley, John E. "Metadata: Everyone talks about it, but what is it?", Disponible en <http://www2.sas.com/proceedings/sugi26/p125-26.pdf>.
- Dooley, Jackie M., Karen Stoll Farrell, Tammi Kim y Jessica Venlet. 2017. "Developing Web Archiving Metadata Best Practices to Meet User Needs", *Journal of Western Archives*, vol. 8, núm. 2. Disponible en <https://digital-commons.usu.edu/westernarchives/vol8/iss2/5/>.
- Foshay, Neil. 2005. *The influence of end-user metadata on user attitudes toward, and use of, a data warehouse*. Disponible en <ftp://public.dhe.ibm.com/software/data/integration/whitepapers/influence-metadata-wp-122905v2.pdf>
- Greenberg, Jane. 2003. "Metadata generation: processes, people and tools", *Bulletin of the American Society for Information Science and Technology*. Disponible en <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bult.269>.
- Kotler, Gad. "Meta-Data for End-Users A crucial part of your Data Warehouse quality policy". Disponible en <http://www2.sas.com/proceedings/sugi29/105-29.pdf>
- McCarthy, John L. 1982. *Metadata management for large statistical databases*. Presented at the Eighth International Conference on Very Large Data Bases, Mexico City, Mexico, September 8-10. <https://cloudfront.escholarship.org/dist/prd/content/qt5cc031cm/qt5cc031cm.pdf?t=p0kgv5>

- Mayernik, Matthew Stephen. 2011. *Metadata realities for cyberinfrastructure: data authors as metadata creators*. Tesis, 2011. Disponible en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2042653.
- Venlet, Jessica, Karen Stoll Farrell, Tammy Kim, Allison Jai O'Dell y Jackie Dooley. 2018. *Descriptive Metadata for Web Archiving: Literature Review of User Needs*. Dublin, OH: OCLC doi:10.25333/C33P7Z.

La Revolución de los Datos Bibliográficos, Científicos y Culturales. Instituto de Investigaciones Bibliotecológicas y de la Información / UNAM. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Valeria Guzmán González, revisión de pruebas, Carlos Ceballos Sosa, formación editorial, Oscar Daniel López Marín. . Fue impreso en papel cultural de 90 gr. en los talleres de Litográfica Ingramex, S. A. de C. V., Centeno 162-1, Col. Granjas Esmeralda, Alcaldía Iztapalapa, Ciudad de México. Se terminó de imprimir en noviembre de 2021.