

LA REVOLUCIÓN DE LOS DATOS BIBLIOGRÁFICOS, CIENTÍFICOS Y CULTURALES

ARIEL ALEJANDRO RODRÍGUEZ GARCÍA
Coordinador



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 OLVERA 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	

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).

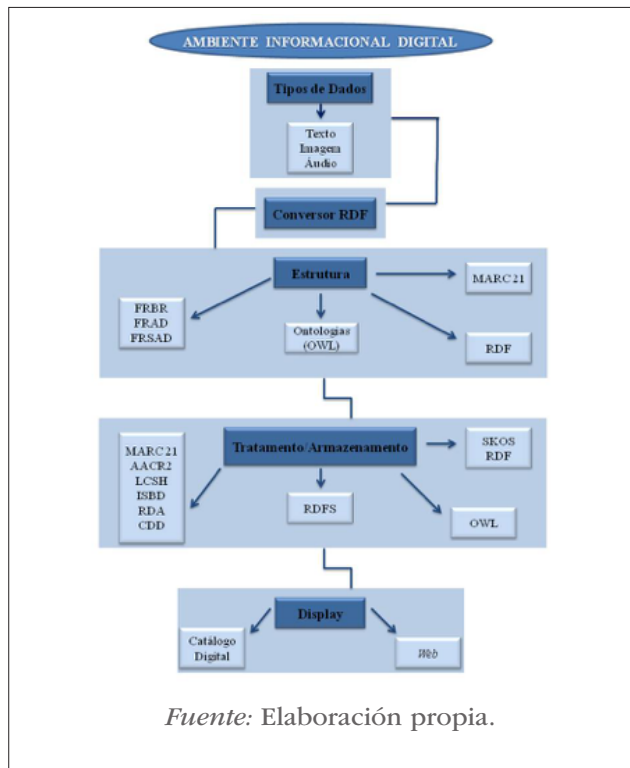
PRESENTACIÓN DE UN MODELO PARA LA DESCRIPCIÓN BIBLIOGRÁFICA SEMÁNTICA

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.

La Revolución de los Datos Bibliográficos, Científicos y Culturales. 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. Instituto de Investigaciones Bibliotecológicas y de la Información /UNAM. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en julio de 2020.