

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 CIENTÍFICOS 141

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	

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.

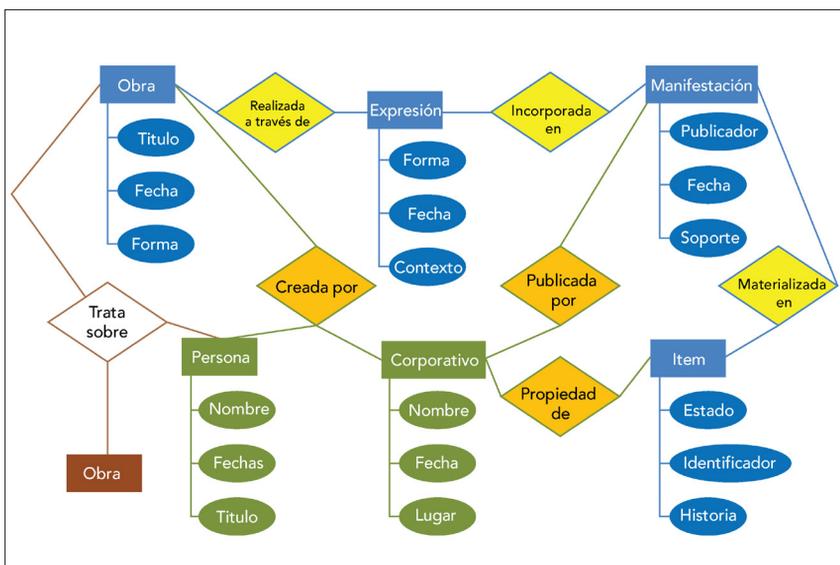
LOS MODELOS DE ENTIDAD-RELACIÓN

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).

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.

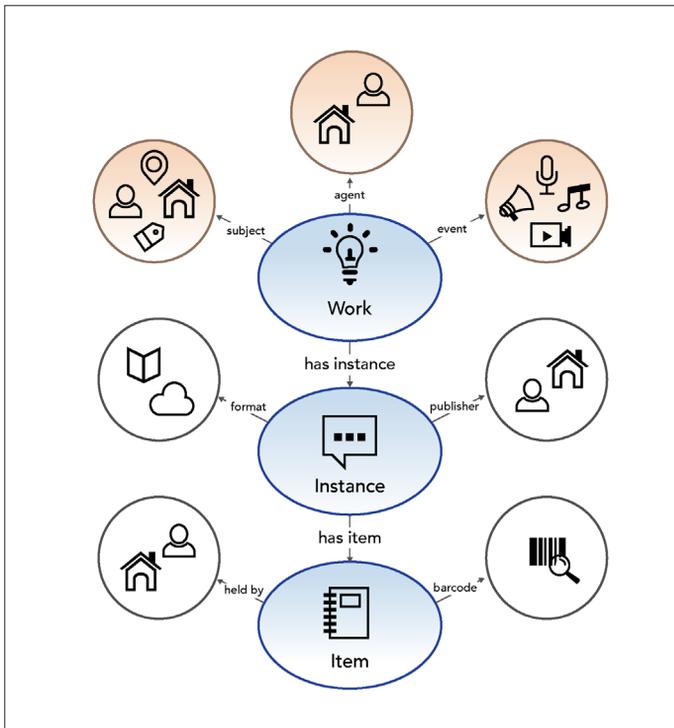
Los sistemas de recuperación de información como Google no pueden cosechar los datos codificados con MARC y hacerlos accesibles para otros. BIBFRAME presentará los datos bibliográficos de forma tal que en los sistemas de recuperación de información puedan ser manejados con un sentido semántico para presentarlos 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.

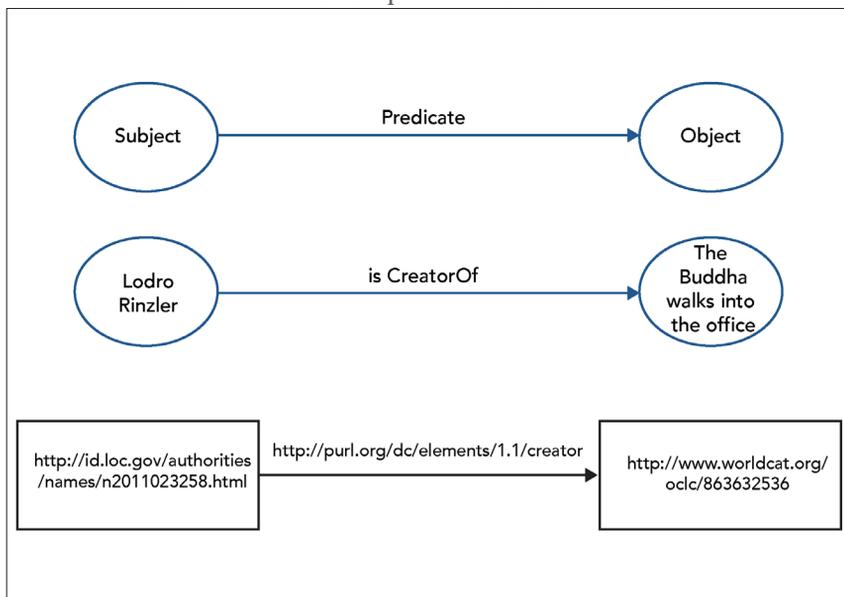
Por otro lado, si comparamos a BIBFRAME con el formato MARC, es notable que este último se enfoca a los registros del catálogo, cuyos datos son comprensibles pero de manera unívoca e independiente. Asimismo, MARC incluye información sobre aspectos conceptuales de las obras y sus atributos físicos, pero 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 em <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_paper_20140501.docx).
- IFLA. 2010. *Functional Requirements for Subject Authority Data (FRSAD): a conceptual model*. Disponible el 17 de agosto de 2018 em <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>.

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.