

## La minería de datos y el catálogo abierto en las bibliotecas: Hacia una sociedad del conocimiento

BELMONT-LUNA, Gerardo†

*Universidad Iberoamericana*

Recibido 5 de Junio, 2017; Aceptado 4 de Septiembre, 2017

### Resumen

Sin lugar a dudas, la revolución y evolución tecnológicas de nuestros días permiten extender los servicios con valor agregado; el plus de los servicios que se ofrecen en instituciones como las bibliotecas, museos, archivos y centros de información es: la generación del conocimiento. Pero, sobre todo, el de representar e interpretar el sentido del Universo de la Información (UI). Al mismo tiempo, se auxilia de tecnologías de bases de datos NoSQL Lo que nos lleva a proponer un OPAC con representación en grafos. Finalmente, en la bibliotecología, la minería de datos se ha traducido a bibliometría. En ese contexto, se propone la utilización del término CA para nombrar la conjunción de tres elementos presentes en la discusión actual en el campo bibliotecológico: acceso abierto a la información, RDA y tecnología. El objetivo del siguiente trabajo es mostrar la construcción de un catálogo abierto a partir de RDA utilizando un lenguaje de programación, que reúna tres características: que por sí mismo sea un lenguaje de programación, que sea un metadato y que sea documento. Para ello, el paper presenta: la justificación teórica del término catálogo abierto, la utilidad de su construcción en la disseminación de información científica y el esquema de la propia construcción.

**Minería de datos, Catálogo Abierto, Bibliotecas, OPAC, Ciencia de datos, Ciencia de Redes, Bases de datos NoSQL**

### Abstract

Undoubtedly, the technological revolution and evolution of our days allow us to extend services with added value; The most of the services offered in institutions such as libraries, museums, archives and information centers are: the generation of knowledge. However, the systems of these institutions are overtaken by traditional practices and schemas of description. But above all, to represent and interpret the meaning of the Information Universe (UI). At the same time, it assists with NoSQL database technologies. This leads us to propose an OPAC with representation in graphs. Finally, in librarianship, data mining has been translated into bibliometrics. The objective of the following work is to show the construction of an open catalog from RDA using a programming language, that has three characteristics: that by itself is a programming language, that is a metadata and that it is document. For this, the paper presents: the theoretical justification of the term open catalog, the utility of its construction in the dissemination of scientific information and the scheme of the construction itself.

**Data Mining, Open Catalog, Libraries, OPAC, Data Science, Network Science, NoSQL Databases**

**Citación:** BELMONT-LUNA, Gerardo La Minería de Datos y el Catálogo Abierto en las Bibliotecas: Hacia una sociedad del conocimiento. Revista de Tecnologías Computacionales 2017. 1-3: 23-47.

† Investigador contribuyendo como primer autor.

## Introducción

La minería de datos y el catálogo abierto en las bibliotecas. El siglo XXI se ha designado como la sociedad del conocimiento. En este siglo, los sistemas de información que tradicionalmente conocemos; han cedido paso.

La minería de datos que anteriormente conocimos hoy se transforma por una minería mucho más inteligente, con mayores recursos tanto en la capacidad de almacenar como de recuperar información, pero sobre todo de representar e interpretar. Esto último gracias a los grafos. Estos tienen su utilidad en cualquier disciplina, pues bajo esta nueva minería de datos es posible saber de dónde viene la información, su comportamiento y su tendencia, mostrando así, un abanico de posibilidades inmediatas. Pues bien, estamos comenzando la era del Big Data. Sin lugar a dudas, es un fenómeno que ninguna disciplina o quehacer humano podrá desatender.

El fenómeno se caracterizará por su habilidad para identificar patrones de comportamiento ya no solo de los datos ocultos, sino en la toma de decisiones influida cada vez más por el desarrollo de la inteligencia artificial. En la bibliotecología la minería de datos se ha traducido a bibliometría. Empero, para refrescar el concepto y el alcance es necesario reencontrarnos con la minería de datos bajo la propuesta del Catálogo Abierto.

El CA es una propuesta en espera de ser construida formalmente, ya se han dado las bases para su construcción. Fundamentado en el código de catalogación FRBR-RDA y este a su vez, en tecnologías de nueva generación para incursionar en el fenómeno del Big Data con miras a la generación del conocimiento. De acuerdo a cuatro factores: (DTI, 1998):

- El desarrollo de las Tecnologías de la Información y la Comunicación (TIC), que permiten una circulación de la información y el conocimiento en cantidades y velocidades inimaginables hasta hace unos años.
- El incremento en los avances científicos y tecnológicos, los cuales se difunden más rápidamente por el motivo anterior.
- La competencia global, la cual ha llevado a abrir nuevos mercados y a poder establecer nuevos tipos de relaciones.
- El cambio en la demanda de los consumidores, en cuanto a productos, inciden más en factores como la calidad, la personalización, el compromiso medioambiental, entre otros. Lo que conlleva a las empresas a la necesidad de reducir su dependencia con los componentes físicos de la producción, dándole mayor importancia a los activos intangibles y a factores como la innovación. (Citado en: Pérez-Soltero, y otros: 2009)

Nos faltaría solo resaltar la importancia que tiene el modelo conceptual FRBR y su propuesta derivada RDA. Dicho binomio es la fundamentación del código que permite la construcción de Catálogos Abiertos.

La fundamentación obedece a la necesidad de sentar las bases para la descripción y recuperación de información, bajo el binomio FRBR-RDA. Basta que se ubique en cualquier tipo de entidad de información en el ambiente del Big Data y la fundamentación de descripción para general la minería de datos.

Antes de continuar, cabe la siguiente pregunta: ¿Por qué debemos pensar en estas tecnologías si lo que tenemos en las bibliotecas hasta hoy funciona?

Tal vez la respuesta sea que se debe estudiar el alcance del binomio FRBR-RDA desde otra perspectiva, y resaltar el estado en que nos encontramos en la era de los objetos como elementos clave de la catalogación, que a diferencia de MARC (Machine Reading Access Cataloging) fue posible bajo la arquitectura de las bases de datos relacionales, y que sin duda alguna, el reto hoy es atender los datos no estructurados descritos en: entidades y atributos, entidades y relaciones, pues la demanda de uso y el exponencial crecimiento de la información nos lleva a pensar en nuevas tecnologías que acompañen al Catálogo Abierto.

Aunado a esto, cabe hacer mención de la opinión de algunos expertos que aseguran que “estadísticas de crecimiento de información generada para el 2016 alcanzando las cifras cercanas al zetabite” (El Mundo.es, 2012).

Lo anterior da muestra de la potencia que se necesita para almacenar y recuperar la información y más aún; para acercarnos a la minería de datos de nueva generación, que, dicho sea de paso ya no es posible interpretar en un software con tecnología limitada, pues tal parece que la característica principal de los datos de valor<sup>10</sup>, es la demanda mayor sobre la capacidad de almacenamiento, lo que nos lleva a pensar que las tecnologías para

<sup>10</sup> A diferencia de los datos de tipo simple que sólo pueden almacenar un valor, los datos estructurados o estructuras de datos pueden recolectar varios valores simultáneamente. Lo que nos concede la evolución de la minería de datos. Más información en: <http://ocw.upm.es/ciencia-de-la-computacion-e-inteligencia-artificial/fundamentos-programacion/contenidosteoriacos/ocwfundamentosprogramaciontema6.pdf/view>

almacenamiento tradicionales no pueden ser capaces de procesar la gran cantidad de datos generada en el instante.

Según Margaret Rouse --tecnologías como NoSQL, Hadoop y MapReduce, forman el núcleo del software de código abierto que soporta el procesamiento de grandes conjuntos de datos a través de sistema de clúster<sup>11</sup> para generar el conocimiento.

Finalmente, la integración de recursos de información en múltiples soportes que llegan a las bibliotecas se va diversificando, ya no solo en formatos de texto, sino también: video, audio, redes sociales y algunos otros. Por lo anterior, podemos decir entonces que: es posible pensar en la opción del Catálogo Abierto, y consecuentemente en la minería de datos.

## Revisión Literaria

Para fines de una mejor comprensión de este trabajo, a continuación, se definen los términos que se utilizan.

### *¿Qué es BIBFRAME?*

FRAME es un marco y podríamos sustituirlo por sinónimos como: dentro de, en el ámbito de o, sencillamente, en. Es decir, en la biblioteca, en el ámbito de la biblioteca.

<sup>11</sup> Clúster del conocimiento es una estructura de interacción y cooperación en gestión que se articula a través de procesos de aprendizaje colectivo e intercambio de conocimiento de los agentes del mundo de la gestión. Desde el clúster del conocimiento se promueven iniciativas tales como foros de intercambio de conocimiento, grupos de trabajo y visitas entre empresas, y se impulsan foros de reflexión para difundir las tendencias en gestión empresarial (Letreros, 2004)”.(Citado en: Pérez-Soltero, y otros: 2009).

De lo anterior podríamos entender a BIBFRAME como en la bibliotecología. Dicho de otra manera; En el ambiente de las bibliotecas.

Dentro del marco se encuentra el modelo conceptual FRBR; siendo entonces este conjunto un el modelo normativo, este debe ser de carácter universal para las bibliotecas y unidades de información.

¿Qué es FRBR?

La International Federation Library Association (IFLA) define a FRBR como el modelo conceptual del universo bibliográfico, que establece cuáles son las entidades, atributos y relaciones existentes en el universo bibliográfico, y qué elementos deben reflejarlo en los registros bibliográficos.

Sus objetivos son:

1. Debe proporcionar un registro bibliográfico, y qué es lo que se espera de un registro para satisfacer las necesidades del usuario
2. Recomendar un nivel básico de funcionalidad y requerimientos básicos de datos para los registros creados por agencias bibliográficas nacionales.

El propósito del Modelo conceptual es servir de base para relacionar atributos y relaciones específicos con las diversas tareas que el usuario realiza al consultar registros bibliográficos. (Tillet, 2009)

El FRBR es un modelo conceptual con la Propósito de mejorar los registros de catalogación (un producto), Catalogación (un proceso), y catálogos (una tecnología). (Caryle, 2011)

¿Qué es RDA?

Resource Description and Access (Descripción y Acceso de Recursos)

RDA son “un conjunto de directrices e instrucciones sobre la formulación de datos para apoyar la búsqueda de recursos (...) proporciona instrucciones que cubren todos los tipos de contenido y medios de comunicación” (Joint Steering Committee, 2013, párr. 2).” (Ramírez Méndez, 2015 pág. 2)

¿Qué es la Minería de Datos?

Desde los años sesenta en el campo del análisis de datos acuñó términos como: “Data Fishing, Data Mining (DM) o Data Archaeology con la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido.”(Pautsch y otros, 200?)

A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro entre otros, empezaron a consolidar los términos de Minería de Datos y KDD.(ídem)

La evolución de la minería de datos puede representarse según tabla no 1.

Etapa	Tecnología
Colección de Datos (1960).	Solo almacén de datos
Acceso de Datos (1980).	Base de datos relacional
Almacén de Datos y Apoyo a las Decisiones (principios de la década de 1990).	Base de datos relacional y software apropiado
Minería de Datos Inteligente finales de la década de los 1990	Web 2.0
Big Data	Núcleo de tecnologías (NoSQL, Hadoop, MapReduce)

**Tabla 1** Evolución de Minería de datos

Fuente *Elaboración propia*

¿Qué es el universo de la información?

El universo de la información se puede entender de la siguiente manera: Primero - El Universo es todo, sin excepciones: materia, energía, espacio y tiempo. Estos cuatro elementos los podemos definir como los objetos tangibles e intangibles portadores de información y forma parte del Universo Bibliográfico (UB) -. Segundo – La información es todo aquel mensaje por descifrar que porta cada objeto de información.

Podemos decir entonces que: El Universo de la Información, está formado por todos los objetos tangibles o intangibles que portan información y está en espera de ser descifrada.

### Antecedentes

Gleick muestra que la información es omnipresente. Turing codificaba las instrucciones como números (ceros y unos) e inventó una máquina ideal (la “máquina de Turing”) que suministró los fundamentos teóricos sobre los que se asientan los ordenadores actuales, habilidades que utilizó para descodificar mensajes criptografiados que los alemanes empleaban durante la Segunda Guerra Mundial. Shannon creó códigos para los genes y cromosomas y para los relés y los interruptores, y produjo un resultado de importancia descomunal, el teorema de la codificación sin ruido.

La teoría de la codificación se convirtió en un elemento esencial de la ciencia de la computación, y sin la corrección de errores y la compresión de datos no existirían los módems, los CD (aunque estos están, en realidad, desapareciendo), ni la televisión digital.

La idea de Shannon de codificar la vida resultó cierta: la réplica del ADN (herencia) es una copia de información, mientras que la fabricación de proteínas es una transferencia de información (ARN mensajero). “Si existiera algo parecido a una guía de los seres vivos”, señala Gleick, “creo que su primera línea diría como un mandato bíblico: que se multiplique la información. El gen no es una macromolécula portadora de información. El gen es la información”. (Sánchez Ron, 2012)

Después de la escritura, el lenguaje de las matemáticas ha sido el preferido por la ciencia: todo cuanto es cuantificable es verdadero, parece decir la sentencia. Desde tiempos de Pitágoras hasta 1900, cuando David Hilbert expuso los 23 problemas de las matemáticas, la congruencia y lógica de su estructura fue irreprochable hasta que, en 1931, Kurt Gödel demostró a partir de sus teoremas de incompletud, que en las matemáticas también había paradojas; es decir, podía haber construcciones y resultados inesperados.

En la misma línea de pensamiento se inscriben Alan Turing y Claude Shannon. En 1936 Turing ideó la máquina universal cuyo comportamiento se tradujo en el procesamiento de datos y expresión de resultados a través de la construcción adecuada de algoritmos. Turing asumió, en consonancia con Gödel, que la máquina nunca haría lo mismo que la mente humana. Esto es, la máquina perfecta era incompleta.

Por su parte, en 1948 Shannon publica la teoría matemática de la comunicación, mejor conocida como la Teoría de la Información, en la que Shannon también demuestra su incompletud al declarar “la información está estrechamente relacionada con la incertidumbre” (Gleick, 2011).

Esta forma no-lineal de pensar se hizo muy popular en el siglo XX. Así, son visibles los ejemplos en diversos campos de la ciencia: Ilya Prigogine en la química describió las estructuras disipativas y los sistemas auto-organizados; Einstein estableció la relatividad del tiempo; George Gamow definió el big-bang como el origen del Universo, que desde 1929 había referido Edwin Hubble. La Teoría del Caos es producto de esta forma de mirar los fenómenos. Se atribuye a Edward Lorenz el término efecto mariposa para entender cómo un pequeño evento puede ser la causa de un fenómeno de grandes proporciones.

A partir de ahí, la lista de publicaciones tanto científicas como en el cine y la literatura ha ido incrementándose. Tales referencias no son ajenas a la bibliotecología porque hay un elemento común de primer orden en todas estas teorías, y ese elemento es la información, que desde el momento en que Shannon le asignó una medida, el bit, se habla de procesamiento de información, almacenamiento de información y recuperación de información.

En su ensayo Elogio a la historia de la bibliotecología, Alfaro (2011) hace notar que las investigaciones de la profesión se encaminan equivocadamente a tratar exclusivamente el tema de las tecnologías, en detrimento de la historia, pero la biblioteca definitivamente es una institución ligada fuertemente a los avances de la tecnociencia<sup>12</sup> en virtud de sus funciones primordiales, tales como la preservación, organización y difusión de la información.

<sup>12</sup> Aunque el término lo usa por primera vez Bruno Latour en 1983, en 2004 Echeverría lo actualiza y lo vincula principalmente con el proyecto *Europe* efectuado en 2001 en Lisboa, en donde se habla del desarrollo de la sociedad del conocimiento.

Si la información es un producto de la cultura escrita, así como un producto de la generación y acumulación del conocimiento, desde la invención de la escritura y la imprenta, así como la explosión de la información y la creación de redes, la biblioteca está permanentemente ocupando los avances tecnocientíficos, precisamente en la realización de sus funciones. Tener presente esta premisa es importante para la concepción del catálogo como el medio por el cual, la comunidad tiene noticia del acervo de una biblioteca o unidad de información. Evidentemente hay una historia, que en el caso de la catalogación se ha explotado considerablemente, razón por la cual, consideramos pertinente identificar el aspecto tecnocientífico del tema que nos ocupa.

### Hipótesis

En la medida en que se despejen las dudas respecto a las tecnologías que giran alrededor de FRBR-RDA y se muestren las posibilidades de aplicación. Si esto consiste, no solo en mencionar esas tecnologías, sino que además analizar su estructura y modelos de datos y operación; límites y alcances. Entonces podremos tener mejores propuestas de desarrollo para los OPACS de las bibliotecas, en donde se identifica, el ejercicio de la minería de datos de nueva generación. Lo anterior deja una la siguiente pregunta ¿será realmente la aproximación a la sociedad del conocimiento?

### Estado del arte

Actualmente la Library of Congress en adelante LC, a través de la IFLA encargo a organizaciones como OCLC, y Zephyra desarrollar soluciones de tecnología que correspondan al modelo de datos FRBR.

Zephyra presento a OCLC tecnología de FRAME, quien acuño el nombre de BIBFRAME. “En dicho frame se presentó el catalogo en línea con el modelo de datos FRBR y la organización de la información.” (Santana, 2011) Tras la propuesta, en el 2015.

Se presentaron los avances y las perspectivas del nuevo modelo bibliográfico (Agenjo, 2015) Por su parte, OSMAN en el 2016 presenta un notable cuestionamiento acerca de que si las bibliotecas están listas para BIBFRAME y el futuro que el modelo debe tomar. Finalmente, en el mismo año, Perchi y Miller proponen el modelo de Linkend Data, como solución al problema, este último recomendado por Erick Miller, líder de Zephyra.

De lo anterior, se entiende que el proyecto de BIBFRAME bajo el desarrollo de Zephyra, es el logro más importante que la Library of Congress ha presentado para el binomio FRBR-RDA. No obstante, existen algunos otros intentos se puede atender esta paradoja.

Nos remontamos a la década de los años 90's. Década donde surge el concepto de Recursos Uniformes para Nombres URNs por sus siglas en inglés. Fueron la primera muestra de interés para trabajar con metadatos y lograr su organización; eran épocas de la Web 1.0. La arquitectura constaba de tres partes: Uniform Resource Locators (URL) y Uniform Resource Characteristics (URC).

Los URNs fueron muy bien creados tenían la característica de ser persistentes. Dicha virtud los condeno a desaparecer. Su persistencia fue mayor que el objeto que referenciaban, a tal grado que la dirección se encontraba, pero la información ya había desaparecido.

Lo que presentaba enlaces rotos. Otro aspecto técnico es: si un URL se volvía famoso o importantes el servidor se veía saturado y terminaba colapsando. Finalmente, esto derivó en que los enlaces no eran tolerantes a la demanda o cambio de dirección.

Para dar solución a este problema llegaron los identificadores de persistencia, que según Baleani, Cervin y Rodriguez: “La identificación permanente de los documentos digitales permite garantizar la citación correcta de este tipo de obras, puesto que su URN (Nombre Uniforme de Recurso) siempre es el mismo, aunque haya sufrido un cambio de ubicación a otro servidor o directorio.

De esta manera, se impide la existencia de citaciones bibliográficas que remitan a enlaces rotos.” (Baleani, Cervin y Rodriguez, 1999).

Algunos sistemas para identificadores de URL son:

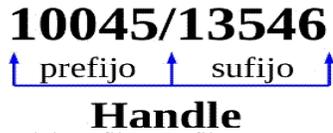
#### Sistema Handle

El Sistema Handle es un sistema que permite la asignación de identificadores persistentes a los recursos de información u objetos digitales existentes en Internet (artículos, revistas, imágenes, etc.).

Desarrollado por CNRI (Corporation for National Research Initiatives), se estructura en dos partes:

- Prefijo (Prefix): identifica al productor del identificador (universidad, editorial, revista, etc.)
- Sufijo (Suffix): identifica a cada uno de los documentos u obras digitales (artículo, libro, capítulo, etc.)

La suma del prefijo y el sufijo conforma el identificador persistente, en este caso llamado "handle". Figura 1.



**Figura 1** Suma del Prefijo y Sufijo  
*Fuente Idem*

### Sistema DOI (Digital Object Identifier)

El sistema DOI proporciona un marco para la identificación persistente de recursos electrónicos, la gestión de contenido intelectual, la gestión de metadatos, vincular a los clientes con los proveedores de contenidos, facilitar el comercio electrónico y permitir la gestión automatizada de los medios de comunicación. El sistema está gestionado por la International DOI Foundation, aunque el registro se realiza a través de Agencias de Registro.

Tal y como sucede con el sistema Handle, el sistema DOI no cambia con el paso del tiempo, aunque el artículo sea reubicado en una dirección distinta, puesto que lleva la información incorporada en forma de metadatos.

Este sistema es el más extendido entre los editores y revistas comerciales a la hora de asignar identificadores persistentes a sus obras digitales.

Al igual que el sistema Handle, el DOI se estructura en dos partes:

- Un prefijo que corresponde al productor de DOI.
- Un sufijo asignado a cada documento o ítem.

En el caso del siguiente ejemplo, 10.4100/jhse.2010.52.15, la primera secuencia numérica sería el prefijo que identifica al productor (10.4100), y la siguiente el sufijo que identifica el documento digital (jhse.2010.52.15). (Ídem) Como lo muestra la imagen no. 2.



**Figura 2** Suma del Prefijo y Sufijo  
*Fuente Idem*

### PURL (Persistent Uniform Resource Locator)

El PURL es un localizador de recurso uniforme (URL) basado en el Uniform Resource Identifier o URI, que no describe directamente la ubicación del recurso que se va a recuperar, sino que describe una ubicación intermedia (más persistente) que sirve de redireccionamiento al recurso final.

Cuando se entra a una PURL en un navegador, este se encarga de enviar el requerimiento de página a un servidor PURL que devuelve la verdadera URL de la página. Los PURL son persistentes porque una vez que el PURL es establecido no necesita ser cambiado nunca. La dirección real de una página web puede cambiar, pero el PURL sigue manteniéndose igual. Los PURLs son gestionados por el OCLC (Online Computer Library Center).

Uno de los principales retos es que la descripción en cualquier sistema debe ser entendida por cualquier ser humano, lo que no sucede con esta aplicación el caso contrario de JSON que si es entendible para cualquier ser humano.

En 1995 Keith Moore, de la Universidad de Tennessee. Formo un grupo de investigadores.

“Dentro de los grupos asistentes estuvieron los implementadores de:

- Sistema de Handles: David Ely y William Arms
- x-dns-2: Paul E. Hoffman y Ron Daniel
  - Path URN: Dan LaLiberte y Michael Shapiro
- OCLC: Keith E. Shafer, Eric J. Miller, Vincent Tkac y Stuart L. Weibel

A esta reunión le siguieron una serie de encuentros informales en Texas. El resultado de todas estas reuniones fue un compromiso llamado "Knoxville Frumework".[URN Report]”

En el 2007 OCLC encarga a Zephyra una herramienta para administrar direcciones y alias de Internet. El desarrollo llamado PURL. “El software OCLC PURL fue desarrollado por la Oficina de Investigación de OCLC para ayudar a los usuarios de Internet a localizar recursos. Dado que Internet está en constante expansión y cambio, los localizadores uniformes de recursos (URL) a veces cambian.

Una vez que una URL falla, todas las instancias de esa URL se vuelven inválidas. Incluso cambios menores en las URLs impiden que los usuarios accedan a los recursos de Internet, y administrar estos cambios con el tiempo es difícil y requiere mucho tiempo.” (Plutchak, T. S. Vision Sessions.)

Lo anterior hizo pensar que bajo esta tecnología se podían atender los grandes problemas de la bibliotecología, como es la semántica y las ontologías. Sin embargo, no dio frutos como se esperaba.

Pues tal vez, en este punto que hemos descrito, el problema ya no sea la tecnología. Si no, las formas de trabajo que las bibliotecas y unidades de información mantienen de manera individual. Donde se nota el dominio de las preferencias a la medida, por encima de los estándares en que se desarrollaron las propias tecnologías como las que ya hemos presentado.

Toca el turno a Zephyra, esta fue fundada por pioneros activos en las redes Web, Web Semántica, datos vinculados, XML, Open Source. Dando vida al frame Ethos. Se destaca la especial participación de arquitectos de la Library of Congress de la iniciativa FRAME (BIBFRAME). Miembros del Grupo Asesor Técnico de la Organización Mundial de la Salud para definir la Clasificación Internacional de Enfermedades (CIE) y los Estándares de Informática de la Salud. Colaboradores del World Wide Web Consortium (W3C).

Un cofundador de Semantic Web Science Association, una organización sin fines de lucro para promover e intercambiar el trabajo académico en la Web Semántica y campos relacionados en todo el mundo, incluyendo el establecimiento de la Conferencia Internacional de la Web Semántica (ISWC).

El XML Guild fundador, XML Guild es un consorcio de consultores XML independientes. Python Software Foundation (PSF) Miembro IBM developerWorks, XML.com y el autor de la columna de Tendencias de desarrollo de aplicaciones.(Zephyra Communiti) A partir de este congreso se hicieron intentos como:

- Exhibit 3.0, se centró en el desarrollo de un marco de publicación para páginas web interactivas de gran escala y ricas en datos.

- El MIT Simile Project, se centró en el desarrollo de herramientas robustas de código abierto basadas en tecnologías de Web Semántica que mejoran el acceso, la gestión y la reutilización entre los activos digitales.
- Akara, una plataforma de código abierto para desarrollar servicios de datos disponibles en la Web, utilizando la arquitectura REST.
- Recollection, una aplicación basada en Django<sup>13</sup> para generar y personalizar vistas, (mapas interactivos, líneas de tiempo, facetas, nubes de etiquetas) que permiten a los usuarios crear su propia experiencia con sus colecciones digitales. La Plataforma de Viewshare.org de la Biblioteca del Congreso es una instancia de Recollection.

Los representantes de Zepheira también forman parte de juntas directivas de diversas organizaciones cuyo objetivo es facilitar la búsqueda, el intercambio y la reutilización de datos en la web, entre ellos:

- Grupo de asesoramiento técnico de Creative Commons
- Grupo Asesor de la Organización Mundial de la Salud sobre el modelado de información sanitaria
- El Centro Nacional de Ontología Biomédica Consejo Consultivo Científico

- Grupo consultivo científico de GenomeSpace
- Junta consultiva de la Iniciativa de Metadatos de Dublin Core

#### Algunos desarrollos

- Para ebsco: Novelist <https://www.ebscohost.com/novelist>
- The Library.Link Network <http://library.link>, y finalmente

Por su parte, Library of Congress ha hecho algunos intentos para RDA (Resource Description Y Access), y schema.org, etc. (Pedersen, 2015)

#### Desarrollos

- Open Library Experience
- The Sierra History
- The Open Library Stack
- RIMM
- BIBFRAME

Este último, aparece en el año 2011, su objetivo es renovar tecnológicamente la estructura de datos bibliográficos y adaptarla a los requerimientos de la web actual, la web de los datos o web semántica, en conjunto con el binomio FRBR-RDA. Donde ya no solo es el intercambio de registros bibliográficos –este, se venía trabajando en conjunto con el formato MARC- El objetivo de BIBFRAME es renovar tecnológicamente la estructura de datos bibliográficos y adaptarla a los requerimientos de la web actual, la web de los datos o web semántica. (Agenjo-Bullón & Hernández-Carrascal, 2016)

<sup>13</sup> Django ya corresponde a las nuevas tecnologías que se están adoptando fuertemente en el mercado. Lo mismo NODESjs, y muchas otras son tecnologías con crecimiento imponente, para atender las demandas de información de los nuevos sistemas. Y que, además, es posible insertar combos tecnológicos para un mejor desempeño.

Instituciones que han implementado BIBFRAME según Tabla 2

Institución	Año
Library of Congress	2014
Colorado College	2014
German National Library	2014
George Washington University	2014
Princeton University Library	2014
Stanford University National Library of Music	2014
Cornell University Library	2014
Columbia University Libraries	2014
Biblioteca Nacional de Cuba "José Martí" (BNJM)	2014
University College London Department of Information Studies University of Illinois at UrbanaChampaign Library 26th March	2015
Library of Alexandria was added to register as the 1st Arab Library to join the Experiment	2015

**Tabla 2** Instituciones que han implementado BIBFRAME

*Fuente Elaboración propia*

### Marco Teórico

A continuación, se explican cómo se relacionan los conceptos relevantes en la descripción del problema, también se presentan algunos grados de complejidad derivados del propio modelo del problema. Anteriormente se describió la conceptualización para poder abordar el problema, por último, la aproximación al modelo de datos con base al binomio FRBR-RDA para el Catálogo Abierto.

### ¿Cómo se relacionan?

La imagen No. 3 corresponde al proceso de MeLiCa, esta muestra cómo se relacionan los elementos definidos para la construcción del Conocimiento en la minería de datos a través del catálogo abierto, bajo el modelo de datos FRBR y RDA.



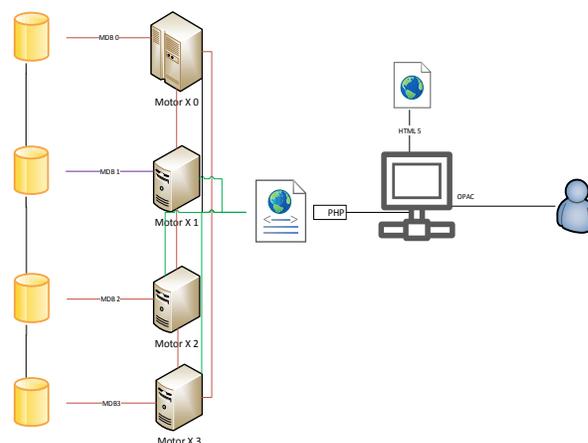
**Figura 3** MeLiCa

*Fuente Elaboración propia*

### Tecnología para MeLiCa

La imagen No. 4 muestra la conceptualización del modelo de datos, este debe ser montado sobre tecnologías cuyo desempeño sea eficiente y de grandes posibilidades de almacenamiento, escalabilidad y que atienda el marco bibliotecológico.

La imagen No. 5.1 Muestra las tecnologías que hacen posible al CA, entre ellas se encuentran: Base de datos MongoDB, Servidor NginX, lenguaje de marcado JSON, lenguaje de programación PHP, Leguaje de programación HTML5, los que a continuación se describen.



**Figura 4** Tecnología para MeLiCa

*Fuente Elaboración propia*

## Base de datos MongoDB

MongoDB es un sistema de base de datos multiplataforma orientado a documentos, de esquema libre. Como ya os expliqué, esto significa que cada entrada o registro puede tener un esquema de datos diferentes, con atributos o “columnas” que no tienen por qué repetirse de un registro a otro. Está escrito en C++, lo que le confiere cierta cercanía al bare metal, o recursos de hardware de la máquina, de modo que es bastante rápido a la hora de ejecutar sus tareas. Además, está licenciado como GNU AGPL 3.0, de modo que se trata de un software de licencia libre. Funciona en sistemas operativos Windows, Linux, OS X y Solaris.

Las características que más destacaría de MongoDB son su velocidad y su rico, pero sencillo sistema de consulta de los contenidos de la base de datos. Se podría decir que alcanza un balance perfecto entre rendimiento y funcionalidad, incorporando muchos de los tipos de consulta que utilizaríamos en nuestro sistema relacional preferido, pero sin sacrificar en rendimiento. (Introducción a MongoDB)

## Servidor NginX

NGINX (pronunciado como “engine X”) es un servidor web HTTP de código abierto que también incluye servicios de correo electrónico con acceso al Internet Message Protocol (IMAP) y al servidor Post Office Protocol (POP). Además, NGINX está listo para ser utilizado como un proxy inverso. En este modo, NGINX se utiliza para equilibrar la carga entre los servidores back-end, o para proporcionar almacenamiento en caché para un servidor back-end lento.

Se trata de un servidor web/proxy completamente inverso, que tiene como principal característica ser sumamente ligero, lo que lleva a su otro gran atractivo, su velocidad, lo que nos permite servir aplicaciones web con una velocidad muy superior a la de sus competidores más directos.

## Lenguaje de marcado JSON

JSON (JavaScript Object Notation - Notación de Objetos de JavaScript) es un formato ligero de intercambio de datos. Leerlo y escribirlo es simple para humanos, mientras que para las máquinas es simple interpretarlo y generarlo. Está basado en un subconjunto del Lenguaje de Programación JavaScript, Standard ECMA-262 3rd Edition - diciembre 1999. JSON es un formato de texto que es completamente independiente del lenguaje, pero utiliza convenciones que son ampliamente conocidos por los programadores de la familia de lenguajes C, incluyendo C, C++, C#, Java, JavaScript, Perl, Python, y muchos otros. Estas propiedades hacen que JSON sea un lenguaje ideal para el intercambio de datos.

JSON está constituido por dos estructuras:

1. Una colección de pares de nombre/valor. En varios lenguajes esto es conocido como un objeto, registro, estructura, diccionario, tabla hash, lista de claves o un arreglo asociativo.
2. Una lista ordenada de valores. En la mayoría de los lenguajes, esto se implementa como arreglos, vectores, listas o secuencias.

Estas son estructuras universales; virtualmente todos los lenguajes de programación las soportan de una forma u otra. Es razonable que un formato de intercambio de datos que es independiente del lenguaje de programación se base en estas estructuras.



El proyecto que dio lugar a PHP fue desarrollado por Rasmus Lerdorf en 1994, como un conjunto de scripts en Perl que utilizaba para gestionar su página web personal. Llamó a esto PHP/FI: Personal Home Page/Forms Interpreter. Tras esto, liberó el PHP Tools versión 1.0 para acelerar la detección de bugs y mejorar el código. Tras ello, la comunidad tomó el mando del desarrollo y en 1997 el ya lenguaje de programación pasó a tener como nombre un acrónimo recursivo PHP: Hypertext Preprocessor. Actualmente, el lenguaje tiene tres usos principales:

- Scripting en el lado de servidor.
- Scripting en línea de comandos.
- Creación de aplicaciones de escritorio, utilizando el paquete PHP-GTK.

PHP es soportado por todos los sistemas operativos mayoritarios y permite escoger el paradigma de programación: programación procedural, orientada a objetos o una mezcla de ambas. Entre las características más destacadas de PHP, podemos enumerar, además de lo ya nombrado:

- Soporte para un amplio rango de bases de datos.
- Soporte para comunicación con otros servicios mediante protocolos como LDAP, IMAP, SNMP, POP3, HTTP, etc.
- Soporte para apertura de sockets de red e interacción mediante cualquier otro protocolo.
- Capacidad de procesamiento de texto, entre los que destaca la compatibilidad con expresiones regulares de Perl y herramientas para parsear documentos XML.
- Soporte (y existencia) de incontables extensiones al lenguaje [8][9].

## Leguaje de programación HTML5

“HTML es un lenguaje de marcado que se utiliza para describir y categorizar la estructura y el contenido las páginas web. La sintaxis de este lenguaje consiste mayoritariamente en una serie de elementos con las que "etiquetamos" el contenido que queremos mostrarle al mundo a través de internet.

Estos elementos constan de una etiqueta de apertura (por ejemplo: <p>) y otra de cierre (por ejemplo: </p>), formando así una especie de contenedor que nos da información acerca del tipo de contenido hay dentro de estas etiquetas, como se muestra en el siguiente ejemplo.

```
<!DOCTYPE html>
<html>
  <head>
    <title>
      Esto es el título de mi página web
    </title>
  </head>

  <body>

    <h1>
      Esto es un titular
    </h1>

    <p>
      Esto es un párrafo
    </p>

    <p>
      Esto es otro párrafo
    </p>

  </body>
</html>
```

Como se puede observar, las etiquetas están contenidas unas dentro de otras, formando así una estructura jerárquica y semántica.

Así, la pareja de etiquetas <html> </html> indican que todos aquellos elementos que están contenidos dentro de ella son código html. Dentro de la etiqueta <body> </body> se encuentra el cuerpo de la página web, lo contenido dentro de la etiqueta <p> </p> son párrafos y así sucesivamente.” (Aprende html.com)

**Resultados**

**El Catálogo Abierto**



**Figura 10** Proceso MeLiCA  
Fuente Elaboración propia

*Elementos importantes*

¿Cuáles son?

De acuerdo a FRBR, los elementos importantes son los tres grupos que a continuación se mencionan:

Grupo 1

Obra, Expresión, Manifestación e Ítem

Grupo 2

Personas, Identidades Corporativas

Grupo 3

Concepto, Objeto, Acontecimiento, Lugar

¿Cómo funcionan?

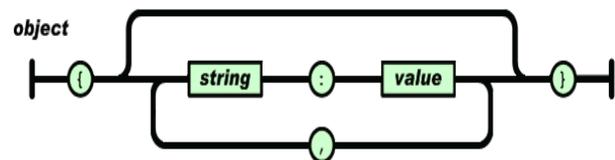
Para comprender.

Para lograr que un sistema entre en producción es necesario representar el problema y representar los momentos en los que se asignan las tareas para entender cuál es su función y que se desea obtener de este.

De este modelado se logra alcanzar las relaciones que establecen los futuros OPACs

La forma de escribir la catalogación:

Donde la Figura 11 muestra el modo de la escritura. El arreglo de esta representación quedaría de la siguiente manera:



**Figura 11** Modo de Escritura  
Fuente ECMA-404 The JSON Data Interchange Standard.

Ejemplo:

array

{

Título: “Ejemplo de Arquitectura MongoDB”,

Mención de responsabilidad: “Gerardo Belmont Luna”,

Mención de edición: “1ª edición”,

Número de serie: “RDA: No. 2”,

Entidad de producción: “Quien produce”,

Entidad de distribución: “Quien distribuye”,

Entidad de fabricación: “Quién fabrica”,

Entidad Fecha de copyright: “Fecha”,

Modo de isuance: “Modo de uso”,

Frecuencia: “Frecuencia de aparición”,

Identificador de la manifestación:

“Asigne identificador según sus políticas”,

```

Cita preferida: "Use la cita preferida o
asiento preferido",
Custodia e historia del ítem: "Narre la
historia del ítem",
Identificador para el ítem: "Asigne el
identificador de acuerdo a sus políticas"
}
    
```

Es un ejemplo simple, pero cumple con su propósito: demostrar un primer documento, con muchos documentos y posibilidades. De acuerdo a la estructura de RDA (Grupo 1, Obra, Expresión, Manifestación e Ítem), según muestra la Figura 9 y de la combinación de la estructura de JSON según muestra la Figura 8.

### Objetos de Información y Clases

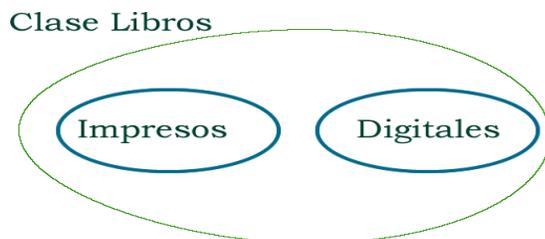
#### Objetos de Información OI

Cualquier entidad del universo contiene información.

Ejemplo: Libros, DVD's, Documentos PDF, Fotografías, y mucho más.

### Clases

Son el conjunto o naturaleza a la que pertenecen, por ejemplo (Figura 12):



**Figura 12** Clase Libros  
Fuente Elaboración propia

### Atributos y Relaciones

#### Atributo

Tomando como referencia la Imagen No. 10 se presenta la Clase Libros. El atributo Son las características intrínsecas de un objeto, es independiente de otros objetos. Tiene un nombre y rango posible de valores. (Tabla 3).

Creador	Título	Lugar	Año	País	Idioma	ISBN	Manifestación
Miguel de Cervantes Saavedra	El ingenioso Don Quijote de la Mancha	Madrid	1541	España	Español antiguo		Impreso
Miguel de Cervantes Saavedra	El Quijote de la Mancha	Ciudad de México	2015	México	Español	123456789	Electrónico
Miguel de Cervantes Saavedra	El Quijote y sancho Panza	D.F.	2010	México	Español antiguo		Película
Miguel de Cervantes Saavedra	Las Aventuras de Don Quijote	Barcelona	1985	España	Inglés	123456781	Audio libro

**Tabla 3** Clase Libros

Fuente Elaboración propia

### Relaciones

Una relación es una característica que vincula conceptualmente a varios OI. Cada OI juega un rol conceptual en ese vínculo.

Las relaciones se dan de acuerdo a las bases de datos

1:0 Elemento relacionado

0:1 Uno o ningún elemento relacionado

1:N Uno o varios elementos relacionados

N:N Muchas relaciones entre muchos elementos

Según lo muestra la Figura 13



Figura 13 Relaciones Fuente Elaboración propia

Caso 1.- Catálogo Abierto como modelo de datos en Excel

Caso 2 NMP

### Metodología

#### Iniciado de servidor NginX en Windows 10

Para este caso particular debemos descargar manualmente la última versión del comprimido en la página oficial de Nginx. Enseguida, descomprimiremos el archivo que descargamos y usando la consola navegaremos al nuevo directorio para iniciar el proceso: (Figura 14)

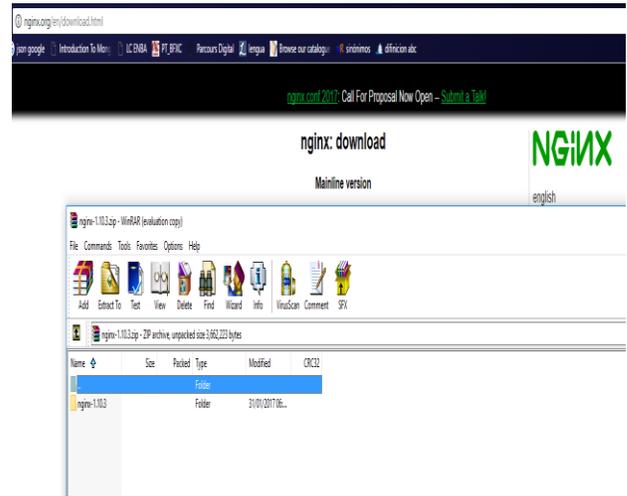


Figura 14 Proceso de descarga y descompresión de Ngnex.Fuente Elaboración propia

Lo descomprimimos en el directorio de C:\\ creamos una carpeta con el nombre de nginx y finalmente entramos al bin de la carpeta y ejecutamos enginx.exe

#### Inicio de MongoDB

En Windows 10, se invoca al Shell desde el buscador en la barra de inicio. (Figura 15)

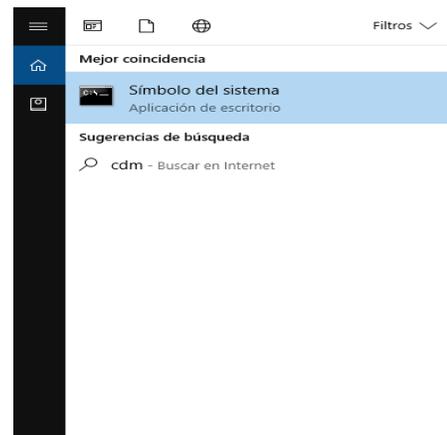


Figura 15 Búsqueda de Shell Fuente Elaboración propia

En seguida no aparece el Shell. Donde nos vamos a la siguiente ruta C://mongodb/bib

Enseguida corremos el mongod (base de datos) para que nos reciba la información que se cargará. (Figura 16)

```

C:\>cd mongodb
C:\mongodb>cd bin
C:\mongodb\bin>mongod.exe
2017-05-08T12:54:56.008-0500 I CONTROL [initandlisten] MongoDB starting : pid=14376 port=27017 dbpath=C:\data\db\ 64-bit
host=81841087388
2017-05-08T12:54:56.014-0500 I CONTROL [initandlisten] targetMinOS: Windows 7/Windows Server 2008 R2
2017-05-08T12:54:56.015-0500 I CONTROL [initandlisten] db version v3.4.4
2017-05-08T12:54:56.015-0500 I CONTROL [initandlisten] git version: 888390515874a9deb16c5d36559ca80644badd
2017-05-08T12:54:56.016-0500 I CONTROL [initandlisten] OpenSSL version: OpenSSL 1.0.1u-fips 22 Sep 2016
2017-05-08T12:54:56.016-0500 I CONTROL [initandlisten] allocator: tcmalloc
2017-05-08T12:54:56.016-0500 I CONTROL [initandlisten] modules: none
2017-05-08T12:54:56.016-0500 I CONTROL [initandlisten] build environment:
2017-05-08T12:54:56.016-0500 I CONTROL [initandlisten]   distmod: 2008plus-ssl
2017-05-08T12:54:56.016-0500 I CONTROL [initandlisten]   distarch: x86_64
2017-05-08T12:54:56.017-0500 I CONTROL [initandlisten]   target_arch: x86_64
2017-05-08T12:54:56.017-0500 I CONTROL [initandlisten] options: {}
2017-05-08T12:54:56.067-0500 I - [initandlisten] Detected data files in C:\data\db\ created by the 'wiredTiger' s
storage engine, so setting the active storage engine to 'wiredTiger'.
2017-05-08T12:54:56.069-0500 I STORAGE [initandlisten] wiredtiger.open config: create,cache_size=3537M,session_max=2000
0,eviction(threads_min=4,threads_max=4),config_base=false,statistics=(fast),log=(enabled=true,archive=true,path=journal
),compressor=snappy,file_manager=(close_idle_time=100000),checkpoint=(wait=60,log_size=2GB),statistics_log=(wait=0),
2017-05-08T12:54:57.175-0500 I CONTROL [initandlisten]
2017-05-08T12:54:57.176-0500 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2017-05-08T12:54:57.176-0500 I CONTROL [initandlisten] **      Read and write access to data and configuration is u
restricted.
2017-05-08T12:54:57.176-0500 I CONTROL [initandlisten]

```

**Figura 16** Búsqueda de Shell  
Fuente Elaboración propia

La Figura 16, se muestra como hemos levantado el servicio de base de datos de MongoDB.

2° paso. – Corremos la consola para ejecutar la interface de mongo.exe como lo muestra la imagen no. 17

```

C:\>cd mongodb
C:\mongodb>cd bin
C:\mongodb\bin>mongo.exe
El nombre del directorio no es válido.
C:\mongodb\bin>mongo.exe
MongoDB shell version v3.4.4
connecting to: mongodb://127.0.0.1:27017
MongoDB server version: 3.4.4
Server has startup warnings:
2017-05-08T12:54:57.175-0500 I CONTROL [initandlisten]
2017-05-08T12:54:57.176-0500 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2017-05-08T12:54:57.176-0500 I CONTROL [initandlisten] **      Read and write access to data and configuration is u
restricted.
2017-05-08T12:54:57.176-0500 I CONTROL [initandlisten]

```

**Figura 17** Búsqueda de Shell  
Fuente Elaboración propia

En esta consola, podemos realizar consultas como ver las bases de datos cargadas según lo muestra la imagen no. 15. Algunos comandos para empezar son:

## Configuración del PHP

Tal vez, la configuración más complicada es la de configurar el controlador de MongoDB para PHP, pues existen una gran cantidad de versiones. Y aun que se identifique la versión de PHP, no es posible aun tener la versión apropiada. Por lo que hay que estar probando entre versiones, aun cuando ya lo hemos instalado una y otra vez.<sup>14</sup>

## Instalando el driver para PHP

Después de haber instalado MongoDB, debemos configurar nuestro PHP, pues este es el lenguaje que hemos elegido para el CA. como debe comunicarse con él, para ello debemos descargar el driver adecuado y activarlo en el php.ini como una extensión.

Desde el repositorio de MongoDB descargamos la versión que nos puede ayudar. Es recomendable utilizar la antepenúltima. No usar más viejas ni la última porque suelen traer errores.

1. Descargamos el driver desde el repositorio en la siguiente ruta: <http://s3.amazonaws.com/drivers.mongodb.org/php/index.html> donde tendremos varias opciones.
2. Identificar la extensión

<sup>14</sup> Es una de las actividades que hacen que valga la pena toda esta metodología. Pues, configurar verdaderamente nuestra plataforma da muestra que el proceso realmente se entiende y además se disfruta. La repetición nos da la habilidad y la comprensión de jugar con los parámetros y verdaderamente personalizar nuestro sistema.

Una vez descargado el archivo, descomprimos y buscamos la extensión que se adapte a nuestra versión de PHP, en mi caso como estoy corriendo la versión 5.4 utilizaré php\_mongo-1.6.0RC2-5.4-vc9.dll y lo renombramos como php\_mongo.dll.

Luego debemos mover el archivo al directorio de nuestras extensiones que sería: C:\XAMPP\php\ext\php.ini

3. Damos de alta el servicio agregando la siguiente línea de comando Con la extensión añadida solo nos faltaría agregar en el php.ini la siguiente línea: extension=php\_mongo.dll

Finalmente reiniciamos nuestro servidor Apache y ejecutamos phpinfo() para verificar que la extensión este habilitada, buscamos con CTRL + F la palabra mongo y veremos la información de nuestra extensión:

En este caso es necesario crea un archivo con terminación PHP, para poder ver la información como a continuación se presenta.

mhash

MHASH support	Enabled
MHASH API Version	Emulated Support

mongo

MongoDB Support		enabled
Version	1.5.1	
Streams Support	enabled	
SSL Support	enabled	
Supported Authentication Mechanisms		
MONGODB-CR	enabled	
SCRAM-SHA-1	enabled	
MONGODB-X509	enabled	
GSSAPI (Kerberos)	enabled	
PLAIN	enabled	

Directive	Local Value	Master Value
mongo.allow_empty_keys	0	0
mongo.chunk_size	261120	261120
mongo.cmd	5	5

Figura 18 Búsqueda de Shell Fuente Elaboración propia

http://localhost:8080/mongojs/info.php, esta carpeta es la que se configuro para obtener la información de PHP.

Finalmente ya esta corriendo el servicio de MongoDB y PHP. Ahora nos hace falta empezar a cargar datos desde php. Y se hace de la siguiente manera:

1. Crear el index.html

```

<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport"
content="width=device-width, user-
scalable=no, initial-scale=1.0, maximum-
scale=1.0, minimum-scale=1.0">
    <link rel="stylesheet"
href="css/estilos.css">
    <link rel="stylesheet"
href="https://neocolas.github.io/normalize.css/3.0.2/normalize.css">
    <title>Lista de Tareas</title>
</head>
<body>
    <div class="principal">
        <div class="wrap">
            <form class="formulario"
action="">
                <input type="text"
id="tareaInput" placeholder="Agrega tu tarea">
                <input
type="button" class="boton" id="btn-agregar"
value="Agregar Tarea">
            </form>
        </div>
    </div>
    <div class="tareas">
        <div class="wrap">
            <ul class="lista"
id="lista">
                <li><a href="#">1
                Lorem ipsum dolor sit amet.</a></li>
                <li><a
href="#">2 Lorem ipsum dolor sit
                amet.</a></li>

```

```

        <li><a href="#">3
Lorem ipsum dolor sit amet.</a></li>
    </ul>
</div>
</div>
<script src="main.js"></script>
</body>
</html>

```

2. Crear conexión con el archivo conexión.php y escribimos el siguiente código:

```

<?php
$conexión = new MongoClient(
"mongodb://localhost:8080" ); // localhost es mi
equipo y 8080 es el puerto que estoy utilizando
para poder verme en mi red local.

```

```

<?php
$mongo = new MongoClient();
$db = $mongo->biblioteca; <--Aquí le digo a la
base que empiece a cargar los datos-->
$coleccion = $db->autores;
$autor = array('autorid' => 1, 'nombre' => 'Dan
Brown');
$coleccion->insert($autor);
$autor = array('autorid' => 2, 'nombre' =>
'Chuck Palahniuk');
$coleccion->insert($autor);
$autor = array('autorid' => 3, 'nombre' => 'Max
Brooks');
$coleccion->save($autor);
?>

```

El código anterior aún no tiene formulario, solo se está captando la información como prueba para saber que el servidor nos escucha, que la base de datos está recibiendo información, y que también envía respuesta. (Figura 19)



**Figura 19** Captación de Información

*Fuente Elaboración propia*

La imagen muestra que los datos cargados con página conexión.php, ya se están recibiendo a través de mongodb, php y html.

## Crear colecciones

Una vez que hemos instalado la base de datos MongoDB

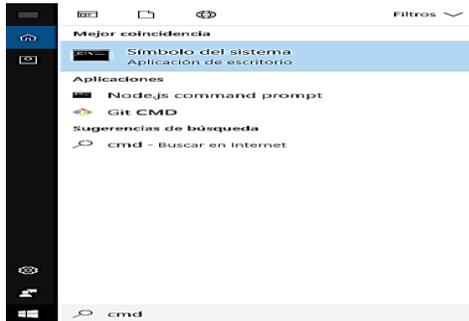
Abrimos el Shell de Windows<sup>15</sup>

Ejecutamos en el buscador de Windows 10 el buscador como lo muestra la Figura 20:

Ejecutamos o llamamos al Shell de Windows desde el buscador. Seleccionamos: Símbolo del sistema.

<sup>15</sup> Windows Shell es la interfaz gráfica de usuario principal del sistema operativo Microsoft Windows. Es el contenedor dentro del cual toda la interfaz de usuario se presenta, incluyendo la barra de tareas, el escritorio, el explorador de Windows, así como muchos de los cuadros de diálogo y controles de interfaz, pero también se describe el pasado de conchas, como MS-DOS ejecutivo y gestor de programas.

El valor por defecto se llama shell de Windows Explorer, este es el programa que determina el aspecto del escritorio (es decir, se crea la barra de tareas, el área de notificación, el menú de inicio, etc).



**Figura 20** Ejecución de buscador  
Fuente *Elaboración propia*

Nos abre el Shell

```

2017-04-27T12:11:54.811-0500 I CONTROL
[initandlisten]
2017-04-27T12:11:54.811-0500 I CONTROL
[initandlisten] ** WARNING: Access control
is not enabled for the database.
2017-04-27T12:11:54.811-0500 I CONTROL
[initandlisten] **      Read and write access to
data and configuration is unrestricted.
2017-04-27T12:11:54.812-0500 I CONTROL
[initandlisten]
> show dbs
admin      0.000GB
biblioteca 0.000GB
local      0.000GB
unTestDB   0.000GB
> use dbbiblioteca
switched to db dbbiblioteca
> show colleitions
2017-04-27T12:13:03.133-0500 E QUERY
[thread1] Error: don't know how to show
[colleitions] :
shellHelper.show@src/mongo/shell/utils.js:906:
11
shellHelper@src/mongo/shell/utils.js:659:15
@(shellhelp2):1:1
> db.createCollection('obras');
{ "ok" : 1 }
> db.createCollection('expresion');
{ "ok" : 1 }
> db.createCollection('manifestacion');

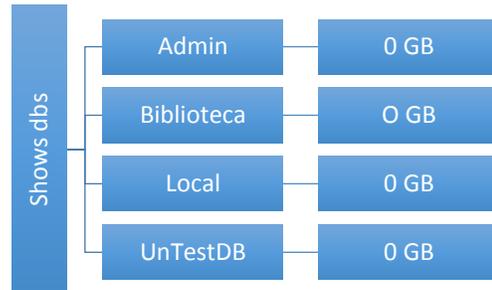
```

```

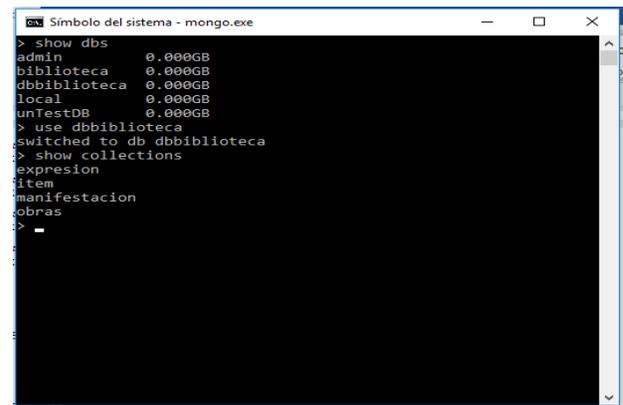
{ "ok" : 1 }
> db.createCollection('item');
{ "ok" : 1 }
>

```

Ahora, se realiza la consulta para mostrar las bases de datos que se lograron crear. – se ejecuta el comando de consulta: Show dbs-



La Figura 21 muestra la consulta sobre la base y las colecciones que hemos creado.



**Figura 21** consulta sobre la base  
Fuente *Elaboración propia*

La figura 21 muestra las bases de datos que se crearon en nuestro manejador de base de datos MongoDB. Dentro de la misma imagen se hace la consulta de las colecciones que se crearon para la base de datos biblioteca. Donde claramente muestra que ya está creadas las colecciones: obra, expresión, manifestación e ítem.

```
> show dbs
admin      0.000GB
biblioteca 0.000GB
dbbiblioteca 0.000GB
local      0.000GB
unTestDB   0.000GB
> use dbbiblioteca
switched to db dbbiblioteca
> show collections
expresion
item
manifestacion
obras
>
```

**Figura 22** Show dbs  
Fuente Elaboración propia

Al mismo tiempo, se presenta el curado de datos, que bajo las colecciones del Grupo 1 se logra, para un mejor funcionamiento y apego a los conjuntos que se plantearon. Finalmente hemos creado nuestra base de datos y también nuestras colecciones con las que trabajaremos con este ejercicio, y según se menciona en la imagen no. 6, correspondiente al Grupo 1 (Obra, Expresión, Manifestación e Ítem)

Recordar que para que este segundo Shell funcione, es necesario tener el primero corriendo en el primer Shell.

**Resultados**

**Curado de datos**

La tabla 4 muestra la información que se extrajo de la base de datos FXC (Francisco Xavier Clavigero), en esta se observa un número considerado de manifestaciones de la misma obra.

Z13_CALL_NO	Z13_CALL_NO	Z13_AUTHOR	Z13_AUTHOR	Z13_TITLE	Z13_TITLE	Z13_IMPRIN	Z13_IMPRIN	Z13_ISSN	Z13_ISSN	Z13_US_Z13	Z13_UPD	T13	
8204	841.45	C47	A34	1001	Cervantes	24510	Aventuras de Don Quijote / ad	260	León, España	SYS	238956	2E+14	
0504	PQ	6324	R6.1917	1001	Cervantes	24510	Cervantes: novelas ejemplare	260	Madrid	: Edici	SYS	454406	2E+14
0504	PQ	6324	A1.v.1	1001	Cervantes	24510	Cervantes: novelas ejemplare	260	Madrid	: Edici	SYS	635396	2E+14
0504	PQ	6324	A1.v.2	1001	Cervantes	24510	Cervantes: novelas ejemplare	260	Madrid	: Edici	SYS	635396	2E+14
0504	PQ	6323	A1.C6.1922	1001	Cervantes	24510	Comedias: el trato de Argel / e	260	Madrid	: Cal	SYS	459480	2E+14
0504	PQ	6323	A1.C6.1940	1001	Cervantes	24510	Comedias y entremeses / M. d	260	Madrid	: Cal	SYS	459477	2E+14
0504	PQ	6323	A1.C6	1001	Cervantes	24510	Comedias y entremeses: Com	260	Madrid	: Cal	SYS	459477	2E+14
0504	PQ	6323	A1.C6.1922	1001	Cervantes	24510	Comedias y entremeses: tom	260	Madrid	: Cal	SYS	459478	2E+14
0500	PQ	6323	A1.2001	1001	Cervantes	24510	Don Quijote de la Mancha / Mi	260	Barcelona	: C	20	8484322831	2E+14
0500	PQ	6323	A1.2004	1001	Cervantes	24510	Don Quijote de la Mancha / Mi	260	México	: San	20	9707703610	2E+14
0504	PQ	6323	A1.2005	1001	Cervantes	24510	Don Quijote de la Mancha / Mi	260	[Madrid]	: PI	20	8480658215	2E+14
0504	PQ	6323	A1.2000	1001	Cervantes	24510	Don Quijote de la Mancha / Mi	260	España	: Edir	20	8484038685	2E+14
0504	PQ	6323	A1.D6	1001	Cervantes	24510	Don Quijote de la Mancha / Mi	260	Barcelona	: I	20	8427807074	2E+14
82	824.46	C4	D65	1001	Cervantes	24510	Don Quijote de la Mancha / Mi	260	México	: Edir	20	8473193920	2E+14
0504	PQ	6323	A1	1001	Cervantes	24510	Don Quijote de la Mancha / tr	260	New York	: L	SYS	453955	2E+14
0504	PQ	6324	A.23.2003	1001	Cervantes	24510	Dos novelas ejemplares y un e	260	Madrid	: Cal	SYS	353605	2E+14
0504	PQ	6325	C54	1001	Cervantes	24513	El cerco de Numancia / Miguel	260	México	: Sig	SYS	48892	2E+14

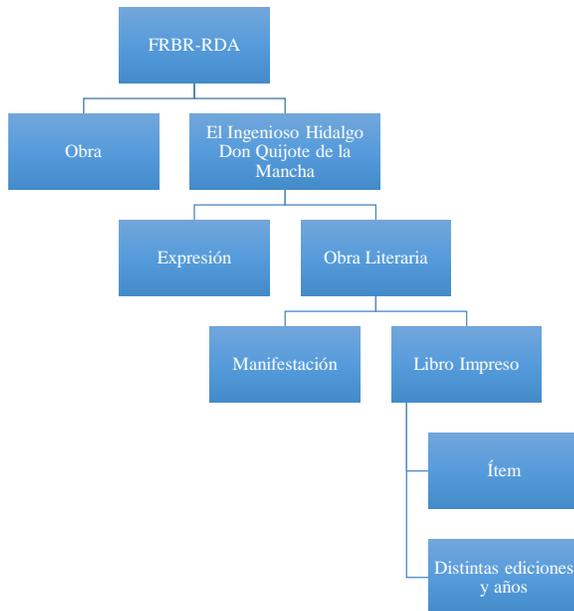
**Tabla 4** Proceso de curado  
Fuente Elaboración propia

De acuerdo a la Tabla 4, se realiza el proceso de curado de la base de datos, este consiste en dividir los datos por las siete áreas de la catalogación.

**Pasos para el curado:**

1. Se inserta un índice a la base de datos con el propósito de poder mover los datos de manera confiable.<sup>16</sup>
2. Se identifica que todas las manifestaciones pertenezcan a la misma obra y expresión – según la descripción del grupo 1.
3. La agrupación para estas cuatrocientas manifestaciones de la obra: el quijote; quedaran de la siguiente manera (Según estructura del grupo 1):

<sup>16</sup> Esta inserción la recomiendo siempre, pues al mover datos siempre hay campos en blanco saltos de línea que muchas veces no permiten seleccionar todo el combo de datos. La otra técnica es trabajar con el filtro de Excel, pero no es mi técnica.



**Figura 23** Agrupación de manifestaciones de obra  
Fuente *Elaboración propia*

4. Datos con estructura Json:

Work: El Ingenioso Hidalgo Don Quijote de la Mancha

Expretion: Obra literaria

Manifestation: Libro Impreso

Item: Ediciones

5. La sintaxis de la obra quedaría de la siguiente manera:

```

{
  "_id" :
  ObjectId("591688cbd8d2761b99f68a72"),
  "creator" : "Miguel de Cervantes Saavedra",
  "obra" : "El Ingenioso Hidalgo Don Quijote de la Mancha",
  "place" : {
    "country" : "España",
    "editorial" : "Imprenta de Don Cervantes",
    "year" : 1504
  },
},

```

```

"ontologi" : {
  "class" : "Literatura Universal",
  "subclass" : "Literatura Española",
  "relation" : {
    "materia" : "Literatura Española",
    "Time" : "Siglo XVI"
  }
},
"año" : 1504
}

```

6. Carga de ítems: en esta última parte, es donde se debe hacer la carga de datos particulares, incluso, en el análisis de la misma obra, podríamos insertar la estructura o capitulado. Según convenga, como lo muestra el siguiente código:

```

{
  "_id" :
  ObjectId("591688cbd8d2761b99f68a72"),
  "creator" : "Miguel de Cervantes Saavedra",
  "obra" : "El Ingenioso Hidalgo Don Quijote de la Mancha",
  "place" : {
    "country" : "España",
    "editorial" : "Imprenta de Don Cervantes",
    "year" : 1504,
    "Expretion" : "Obra literaria",
    "Manifestatio" : "Obra Impresa",
    "item" : {
      "1956" : {
        "editorial" : "Espasa-Calpe",
        "place" : "Esapña"
      },
      "1980" : {
        "editorial" : "La Lectura",
        "place" : "Esapña"
      },
      "1990" : {
        "editorial" : "McGraw-Hill",
        "place" : "México"
      }
    }
  }
}

```

```

},
"ontologi" : {
  "class" : "Literatura Universal",
  "subclass" : "Literatura Española",
  "relation" : {
    "materia" : "Literatura Española",
    "Time" : "Siglo XVI"
  }
}
}
}

```

## Conclusiones

Este trabajo es una invitación a pensar en nuevas posibilidades de construir catálogos con ayuda de la tecnología. 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. Reformular los conceptos de la catalogación y el catálogo, sin perder la esencia de la importancia del dato, permite descubrir y ampliar las fronteras de la descripción documental, de cara a una nueva generación de sociedad usuaria. Al abrir los márgenes de acción, se construye un universo infinitamente más grande al acervo de una biblioteca.

Es también una invitación a crear grupos de trabajo para llevar a cabo la comprobación de la propuesta en donde los bibliotecarios son protagonistas y líderes de proyectos tecnológicos, pues el catálogo abierto rebasa fronteras y propicia la cooperación catalográfica, lo cual fortalecería las prácticas de acceso abierto a la información, generaría oportunidades de cooperación en la región y actualizaría los paradigmas en la práctica bibliotecaria en la perspectiva de la información como un bien común universal.

Finalmente, vale decir también que el hecho de desarrollar habilidades tecnológicas no hace perder el humanismo de la profesión bibliotecaria, antes bien, se reconoce la identidad social de nuestro papel y agregamos valor a la misión de interceder entre el usuario y su información aprovechando las fortunas de la ciencia.

## Recomendaciones

Esta contribución representa y da respuesta a muchas de las reflexiones y aportaciones que en el marco de las discusiones sobre el futuro de la catalogación y las RDA se han publicado, ninguna opinión sobra e invita crear grupos de trabajo para llevar a la comprobación y calificar de acertada o errónea dicha propuesta, se requiere de infraestructura, colegas, bibliotecas y gente que quiera participar y poder llevar el proyecto a cualquier lugar del mundo, pues si bien, el CA rebasa fronteras ideales y sesgos de cualquier índole con el fin de coincidir en las cooperativas catalográfica, donde no exista el recurso profesional de un bibliotecólogo.

Esperando poder realizar un proyecto sólido de cooperación, donde no existen líderes ni absolutismo; solo la capacidad de crear y hacer catalogación con sentido, para un bien común universal.

## Referencias

Alfaro López, Héctor Guillermo (2011) Elogio a la historia de la bibliotecología. Revista Bibliotecas. Vol. XXIX no. 2 Julio- Diciembre. [en línea] Consultado el 31/01/2017 <http://www.revistas.una.ac.cr/index.php/bibliotecas/article/viewFile/3704/3557>

Baleani, Cervini, & Rodríguez, (1999). Identificadores Persistentes para Obras Digitales. [en línea] <https://biblioteca.ua.es/es/propiedad-intelectual/entorno-digital/identificadores-persistentes-para-obras-digitales.html>

Carlyle, A. (2011). Understanding FRBR As a Conceptual Model. *Library Resources & Technical Services*, 50(4), 264-273. [en línea] <https://journals.ala.org/index.php/lrts/article/view/5444>

ECMA-404 The JSON Data Interchange Standard. [en línea] <http://www.json.org/json-es.html>

Gleick, James (2011). *The information: a history, a theory, a flood.*

Introducción a MongoDB. [en línea] <https://www.genbetadev.com/bases-de-datos/una-introduccion-a-mongodb>

Méndez, S. R. (2014). ¿Es necesario cambiar las Reglas de Catalogación Angloamericanas por (RCAA) por Recursos, Descripción y Acceso (RDA)? *Revista electrónica semestral ISSN*, 1659, 4142. [en línea] [http://s3.amazonaws.com/academia.edu/documents/37335582/5-1-2.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1490811024&Signature=U5LRbrmf54df4KPgP3wcC%2FbhpGg%3D&response-content-disposition=inline%3B%20filename%3DEs\\_necesario\\_cambiar\\_las\\_Reglas\\_de\\_Cata.pdf](http://s3.amazonaws.com/academia.edu/documents/37335582/5-1-2.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1490811024&Signature=U5LRbrmf54df4KPgP3wcC%2FbhpGg%3D&response-content-disposition=inline%3B%20filename%3DEs_necesario_cambiar_las_Reglas_de_Cata.pdf)

El Mundo.es: El tráfico de Internet conquistará en 2016 el Zattabyte gracias al video. Jueves 31 de mayo de 2012.

Pedersen, (2015). *La Biblioteca Del Congreso, RDA (Resource Description Y Access), y schema.org, etc.*

Pautsch, Germán A., Kuna, Horacio D., Caballero, Sergio D., Rambo, Alice R., Meinl, Evaldo, Steinhiber, Andrés. (200?). *Detección de datos con ruido en base de datos utilizando herramientas OLAP.* Universidad Nacional de Misiones, Arg. Departamento de Informática, Facultad de Ciencias Exactas Químicas y Naturales.

Pérez Soltero, Armando, Castillo Navarro, Adolfo, Barcelo Valenzuela, Mario y León Duarte, Jaime. (2009). Importancia de los clúster del conocimiento como estructura que favorece la gestión del conocimiento entre organizaciones. *Intangible Capital*. Vol. 5, no. 1, p33-34.

Sánchez Ron, J. M. (31 del 03 del 2012) *Los reinos de la información.* (E. País, Ed) [en Línea] [http://cultura.elpais.com/cultura/2012/03/29/actualidad/1333029918\\_558392.html](http://cultura.elpais.com/cultura/2012/03/29/actualidad/1333029918_558392.html)  
OCLC Research 1996 OCLC Makes PURL Software Available free of charge

Pérez, T. H., Mateos, D. R., & De la Fuente, G. B. (2008, February). Open Access: el papel de las bibliotecas en los repositorios institucionales de acceso abierto. In *Anales de documentación* (Vol. 10, pp. 185-204).

Plutchak, T. S. *Vision Sessions.* [En línea] <http://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=1777&context=nasig>

Tillet, Barbara B. (2009). “FRBR: Antecedentes, estructura e impacto” presentación disponible con permiso de Graciela Spedalieri. Library Of Congress. [en línea] [https://www.loc.gov/catdir/cpsd/frbryfrad/frbr-instructor\\_oct09.pdf](https://www.loc.gov/catdir/cpsd/frbryfrad/frbr-instructor_oct09.pdf)

Zephira. *Communiti* [en línea]: <http://zephira.com/about/professional-communities/>