

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	

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

GERARDO BELMONT LUNA
Universidad Iberoamericana, México

INTRODUCCIÓN

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

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

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

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

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

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

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

LA TECNOLOGÍA *MIDDLEWARE*

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

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

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

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

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

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

METODOLOGÍA

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

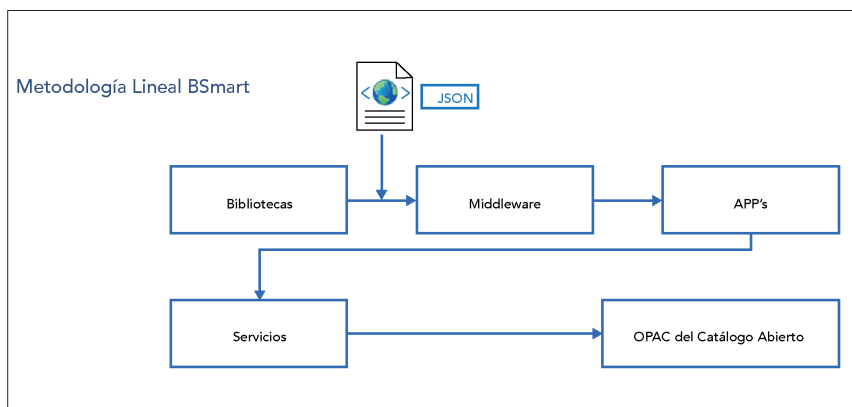
La metodología se caracteriza por el diseño de un proceso mejorado al que denominaremos BSmart (Biblioteca Inteligente). La definimos como una entidad que puede procesar información estableciendo relaciones y posibles vínculos con otros elementos de tipo externo que se entretajan entre sí dentro del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

1. Garantizar la interoperabilidad de las bases de datos.

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

En la ilustración 1 se muestra la metodología sugerida:

Ilustración 1. Metodología lineal de la BSmart

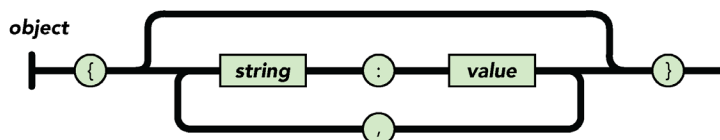


Fuente: Elaboración propia.

En donde:

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

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



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

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

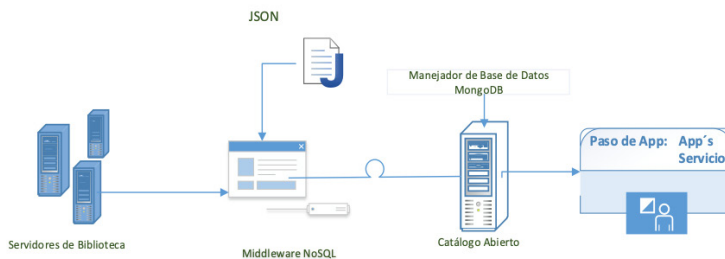
figura imposible). El atributo de figura imposible elimina las barreras tradicionales de la biblioteca, por lo que es posible establecer relaciones y procesos de innovación que se verán reflejados en los servicios de información de las bibliotecas y en general de las unidades de información, enfrentado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

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

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

1. ARQUITECTURA POR COMPONENTES BSMART Y MODELO DE OPERACIÓN

Ilustración 3. Arquitectura por componentes BSmart

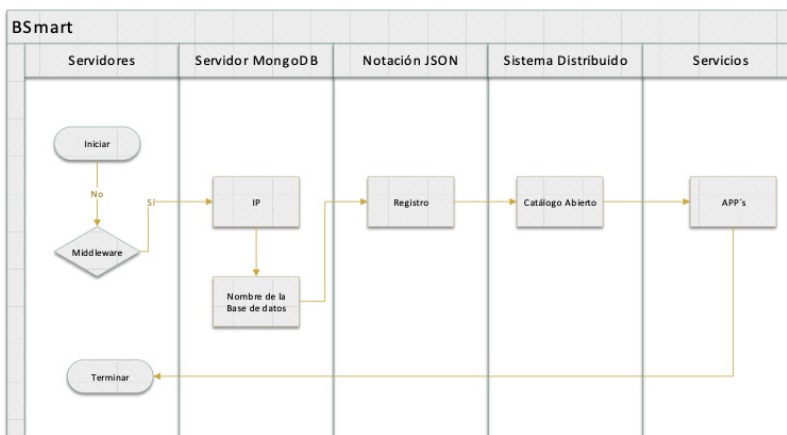


Fuente: elaboración propia.

Fuente: Elaboración propia.

Proceso de integración

Ilustración 4. Proceso de integración BSmart



Fuente: Elaboración propia.

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

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

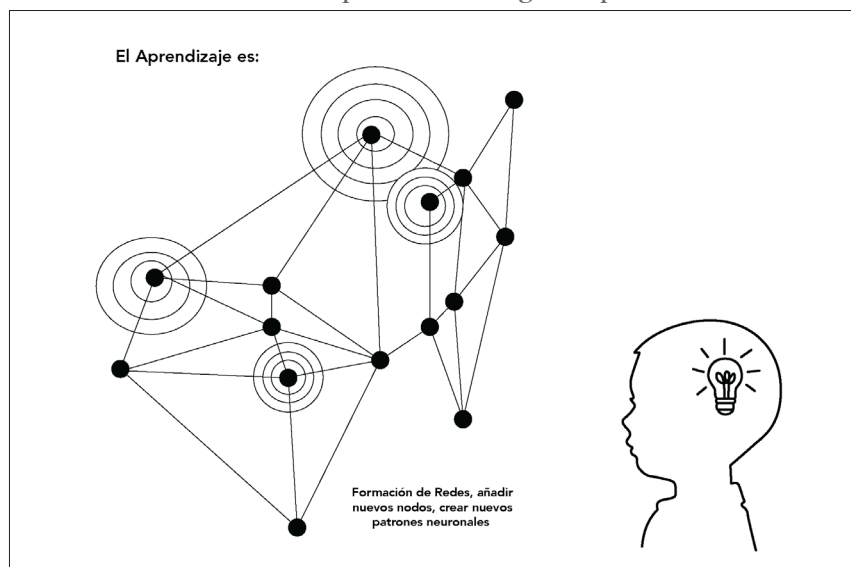
En un ejercicio real, la notación sería de la siguiente manera:

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

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

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

Ilustración 5. La BSmart representada en grafos para el conocimiento



Fuente: Imágenes tomadas de Internet.

DISCUSIÓN

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

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

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

CONCLUSIONES

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

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

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

Trabajo futuro

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

BIBLIOGRAFÍA

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

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

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.