

Creación de datos abiertos enlazados en la Biblioteca Nacional de Costa Rica

Desiree Rodríguez Calvo*

Artículo recibido:
13 de junio de 2023
Artículo aceptado:
8 de septiembre de 2023
Artículo de investigación

RESUMEN

El propósito de este artículo es exponer la propuesta metodológica para el establecimiento de datos abiertos enlazados para la Biblioteca Nacional de Costa Rica «Miguel Obregón Lizano», con el fin de determinar si esta cuenta con los elementos necesarios para que su sistema automatizado de gestión permita –a través de los principios, normas y modelado de datos– enlazar sus datos bibliográficos para que sean accesibles, usables y reusables en el entorno de datos abiertos enlazados. Es un estudio de tipo descriptivo con enfoque cualitativo, utilizando el método del estudio de caso con una investigación basada en evidencias. Uno de los principales hallazgos de esta investigación es resaltar la importancia de que la Biblioteca

* Escuela de Bibliotecología y Ciencias de la Información, Facultad de Educación, Universidad de Costa Rica Costa Rica desiree.rodriguezcalvo@ucr.ac.cr

Nacional de Costa Rica incursione en este ámbito para que se posicione como pionera y principal referencia para otras bibliotecas en Costa Rica, en el uso de datos enlazados, pues estos brindan una plataforma sólida para mejorar la calidad y el alcance de los servicios que ofrecen en la actualidad.

Palabras clave: Datos abiertos enlazados; Catalogación; Normalización de la información; Web semántica.

Creation of Linked Open Data in the National Library of Costa Rica

Desiree Rodríguez Calvo

ABSTRACT

The purpose of this article is to show the methodological proposal for the establishment of linked open data for the National Library of Costa Rica «Miguel Obregón Lizano», in order to determine if it has the necessary elements so that its automated management system allows, through the principles, standards and data modeling link its bibliographic data so that it is accessible, usable and reusable in the environment of linked open data. This is a descriptive study with a qualitative approach, using the case study method with evidence-based research. One of the main findings of this research is to highlight the importance of the National Library of Costa Rica entering this field so that it can position itself as a pioneer and main reference for other libraries in Costa Rica, in the use of linked data, since these they provide a solid platform to improve the quality and scope of the services they offer today.

Key words: Linked Open Data; Cataloging; Information standardization; Semantic web.

INTRODUCCIÓN

En Costa Rica se han dado grandes avances en el área de organización de la información, ya que, desde las instituciones de educación superior—en donde se imparte la carrera de bibliotecología—hasta los grandes sistemas de bibliotecas, se ha incorporado el uso de herramientas de normalización actuales, como Recursos, Descripción y Acceso (RDA) o el formato de codificación MARC 21, para la realización y registro de sus recursos de información.

Considerando lo expuesto, surge la necesidad de fortalecer el área de procesamiento técnico de la información, lo cual se puede realizar mediante la utilización y expansión de datos abiertos enlazados en la gestión de registros bibliográficos, con el propósito de que dichos registros resulten accesibles en un entorno Web y sean aprovechados por otras instituciones bibliotecarias. Esta medida posibilitaría la interconexión de datos alojados en la esfera digital de manera abierta, permitiendo su recuperación eficaz en los sistemas bibliotecarios.

Tal como se menciona, los datos abiertos enlazados «permiten identificar las vinculaciones de significado entre un dato en particular con las obras, expresiones, manifestaciones y ejemplares que forman parte del universo de información, fomentando con ello la conexión de datos con fuentes externas disponibles en el entorno digital» (Ávila Barrientos 2022, 127). Estas vinculaciones generan que las personas usuarias de colecciones bibliográficas accedan a un mundo más amplio de información y así satisfacer sus necesidades informativas.

Al respecto, Feroso-García *et al.* (2018, 412) señalan que los datos enlazados abiertos «permiten conectar unos datos de la Web con otros con los que no existía una relación directa previamente. La relación se establece gracias al valor semántico de estos datos». Lo indicado por los autores anteriores hace alusión a un aspecto fundamental para establecer las vinculaciones entre los datos, que constituye el valor semántico, lo cual se puede realizar mediante la utilización de un modelo de datos como el Marco de Descripción de Recursos (RDF).

De igual manera, se debe recalcar que un aspecto relevante para el establecimiento de datos enlazados abiertos radica en la disponibilidad de un conjunto de datos que estén preferentemente normalizados, en donde dicha normalización resulta esencial para facilitar una estructuración que permita un tratamiento más apropiado de los mismos. El uso del estándar RDA puede contribuir de manera significativa a dicha normalización, ya que mediante su propuesta de relacionar entidades bibliográficas (obra, expresión, manifestación e ítem) y sus atributos buscan «representar el contenido intelectual de las obras, e intentar acercarse mucho más a las necesidades reales de los usuarios» (Anglada 2016, 156). Estas necesidades se asocian con la búsqueda de recursos informativos, la identificación de entidades, la aclaración de relaciones entre ellas y el acceso eficiente a los recursos utilizados.

Acorde al contexto planteado, las unidades de información deben incursionar en las tendencias que se den para garantizar a las personas usuarias las mayores posibilidades de descubrimiento de los recursos de información con los que cuentan, considerando que una de dichas tendencias es el desarrollo de datos abiertos enlazados.

Así, resulta necesario que las bibliotecas desarrollen proyectos de datos abiertos enlazados para ampliar el acceso y localización de los recursos de información con los que cuentan, resaltando que el tipo de bibliotecas que debería enfocarse en esto, en primer lugar, corresponde a las bibliotecas nacionales de cada país, ya que son las entidades encargadas de «conservar, organizar y divulgar el patrimonio bibliográfico de un país al servicio de la sociedad e impulsar los esfuerzos nacionales investigativos y de acceso a la información con miras a elevar el nivel social y educativo de la población» (Amaya 2018). Considerando esto, las bibliotecas nacionales, mediante la incursión en datos abiertos enlazados, pueden promover de forma más amplia el acceso a la información mediante el aprovechamiento y la distribución efectiva del conocimiento almacenado en sus colecciones bibliográficas.

De acuerdo con esta función que desempeñan las bibliotecas nacionales, resulta fundamental que se posicionen como pioneras en cualquier desarrollo o avance que se dé en las diferentes áreas de la disciplina bibliotecológica, por lo que en el caso de esta investigación se plantea que la Biblioteca Nacional de Costa Rica «Miguel Obregón Lizano» asuma un rol de liderazgo en el ámbito nacional en cuanto a los procesos de organización de la información, y en particular en la implementación de datos abiertos enlazados. Este liderazgo se manifestaría a través de la ejecución de un proceso de conversión de datos descriptivos de los registros bibliográficos a datos abiertos enlazados, estableciendo de esta manera una metodología que podría ser adoptada por otras bibliotecas en el país.

Así, el propósito de esta investigación se enfoca en el desarrollo de un modelo metodológico que establezca las bases teóricas para el cambio de datos descriptivos de los registros catalográficos a datos abiertos enlazados, en donde se puedan visualizar mediante grafos las relaciones entre los atributos de cada entidad presente en los registros de los recursos presentes en la Biblioteca Nacional de Costa Rica «Miguel Obregón Lizano».

De acuerdo con el objetivo expuesto, la investigación pretende abordar la siguiente pregunta de investigación:

¿La Biblioteca Nacional de Costa Rica «Miguel Obregón Lizano» cuenta en la actualidad con los elementos necesarios para que los datos descriptivos de los registros bibliográficos presentes en su catálogo automatizado en línea sean accesibles, usables y reusables en el entorno de los datos enlazados?

REFERENTES TEÓRICOS

Organización de la información

La organización de la información desde sus inicios ha desempeñado un papel primordial, sobre todo en la actualidad donde existe una creciente cantidad de información disponible en diversos formatos, generando la necesidad de desarrollar sistemas eficientes para describir y facilitar el acceso a los recursos desde las unidades de información.

Por lo anterior, una adecuada organización de la información permite a las poblaciones de usuarios localizar de manera eficaz los materiales que necesitan, independientemente de su formato; además, contribuye a mantener la integridad y la coherencia de las colecciones, asegurando la correcta identificación y descripción de los recursos.

Para dar esa correcta identificación y descripción de los recursos es necesario, tal como lo detalla Suárez Sánchez (2017, 3) «el uso de herramientas como encabezamientos de materias, sistemas de clasificación, tesauros o estándares de catalogación». Dentro de los estándares de catalogación se puede destacar *Recursos, Descripción y Acceso* (RDA) que, según lo definido por Salta (2021), «implica una nueva organización, totalmente diferente a la estructura de áreas y elementos de descripción comportada por AACR2 e ISBD, basándose íntegramente en los modelos conceptuales creados por IFLA» (p. 18). Esta nueva organización en los elementos de descripción establece una sintaxis para registrar datos catalográficos descriptivos así como para la creación de relaciones significativas entre los recursos descritos y otros.

Uno de los aspectos principales que se destacan en la implementación del estándar RDA radica en poner un «fuerte énfasis en registrar la existencia de relaciones y proporciona un vocabulario controlado muy detallado por medio del cual se puede registrar el tipo de relación, los cual está presente en los apéndices de designadores de relación tales como el I, J, K, M, que se pueden utilizar para identificar la relación específica entre entidades» (Wallheim 2016, 483). Esto resulta fundamental ya que se presenta una nueva posibilidad de establecimiento de relaciones entre datos que anteriormente no se llevaban a cabo, mejorando de esta manera los procesos de recuperación de la información.

Tomando en cuenta el tipo de normalización y de relaciones propuestas en RDA, así como los elementos fundamentales de la Web semántica para la interconexión de datos, es que los datos descriptivos dentro de los registros bibliográficos deberían estructurarse de manera que se puedan vincular, dando paso al mundo de los datos abiertos enlazados.

Datos abiertos enlazados

Los datos abiertos enlazados buscan compartir información de forma libre, accesible y estructurada mediante el uso de estándares Web que permiten la vinculación de diferentes fuentes de datos para facilitar su recuperación.

Lo anterior lo refuerza Ávila Barrientos (2022, 128) al indicar que los datos abiertos enlazados «emplean herramientas tecnológicas que hacen posible la construcción de un espacio común en donde se puedan consultar, recuperar y visualizar datos en diferentes fuentes de la Web».

Asimismo, permiten la creación de catálogos enriquecidos, donde la información de los recursos puede estar interconectada con datos externos, como autoridades o información temática. Además, facilitan la interoperabilidad entre diferentes sistemas, lo que agiliza la búsqueda y recuperación de información.

Lo anterior se lleva a cabo dentro del contexto de la Web semántica, la cual, de acuerdo con Berners-Lee, Hendler y Lassila (2001, 37) es «una extensión de la web actual, en la que se le da un significado bien definido a la información, lo que permite que las computadoras y las personas trabajen mejor en cooperación».

Ampliando lo mencionado, Moreno Mejía (2020, 412) indica que la Web semántica incluye «un conjunto de actividades desarrolladas en el seno del World Wide Web Consortium (W3C) con tendencia a la creación de tecnologías para publicar datos legibles por aplicaciones informáticas, añadiendo metadatos semánticos y ontológicos». Gracias a la utilización de estos metadatos semánticos u ontológicos es que se facilita los procesos de búsqueda de información en la Web y permite otorgar un significado lógico a los datos para realizar tareas de gestión del conocimiento y razonamiento automático de la información.

Lo anterior se refuerza con lo establecido por Berners-Lee, Hendler y Lassila (2001, 36) quienes se refieren a que la web semántica «aporta estructura al contenido significativo presente en la Web, creando un entorno en el que los agentes de software que se desplazan de una página a otra pueden realizar fácilmente tareas sofisticadas para los usuarios».

Para lograr lo planteado, es fundamental enfatizar en los principios de la Web semántica establecidos por Berners-Lee (2006), a saber:

- Utilización de URI para nombres de recursos.
- Utilizar el protocolo HTTP URI para localizar los nombres de los recursos.
- Utilización de estándares de metadatos o lenguajes de consulta.
- Enlazar diferentes URI para que los usuarios logren localizar otros recursos asociados.

Como se puede observar en la propuesta de Berners-Lee, resulta esencial la utilización de URIs, ya que estos generarán identificadores únicos para cada recurso que se enlace. De igual forma, utilizar los estándares de metadatos permiten comprender el significado intrínseco de cada conjunto de datos y utilizarlos de manera precisa para su adecuada recuperación.

Es importante indicar también algunas de las principales características de la Web semántica definidas por Moreno Mejía:

Es una web extendida y está centrada en dotar del significado, optimizar grandes volúmenes de contenido, se obtienen respuestas rápidas, la información se obtiene de forma definida y concreta, permite compartir, transferir y procesar archivos y usa metalenguajes como XML, RDF y OWL. (2020, 415)

Por su parte Sakr *et al.*, (2018, 3-4) señalan que las tecnologías que hacen posible que la Web semántica permita la vinculación de datos son:

- Tecnologías web HTTP: estas tecnologías contemplan el uso de identificadores internacionales de recursos (IRI) que representan una extensión de los identificadores uniformes de recursos (URI); *Unicode*, el cual es utilizado para representar texto en múltiples idiomas; y XML que se utiliza para codificar documentos en formato estructurado.
- Tecnologías semánticas: compuestas por *Resource Description Framework* (RDF) el cual representa información descrita en forma de tripletas (sujeto, predicado y objeto); *Web Ontology Language* (OWL) proporciona construcciones más avanzadas que RDF para modelar un conocimiento rico y complejo en la Web; y SPARQL que es un lenguaje de consulta para RDF y proporciona patrones de tripletas para consultar gráficos RDF.

Tomando en cuenta las tecnologías semánticas mencionadas resulta importante recalcar que Berners-Lee, Hendler y Lassila (2001, 38) refuerzan que para el trabajo en la Web semántica hay dos tecnologías que son esenciales: el lenguaje de marcado extensible (XML) y el marco de descripción de recursos (RDF).

Considerando lo anterior, estas tecnologías que permiten el establecimiento de datos abiertos enlazados deben conocerse a profundidad para que de esta manera los procesos de organización de la información bibliográfica que se llevan a cabo en las bibliotecas se enfoquen en facilitar la interoperabilidad entre sistemas y bibliotecas, fomentando el intercambio de recursos.

En definitiva, la incorporación de las tecnologías de la Web semántica a la organización de la información resulta trascendente para garantizar la eficiencia y la accesibilidad de los servicios bibliotecarios, permitiendo a los usuarios aprovechar al máximo el universo de conocimiento disponible.

Datos abiertos enlazados en el contexto bibliotecológico

El acceso abierto a la información representa un concepto que ha estado presente en las bibliotecas desde hace un tiempo atrás, principalmente para asegurar que las comunidades de usuarios obtengan un acceso libre a los recursos que se encuentran en ellas. Sin embargo, en las actualidad este acceso a la información se puede ver de manera más amplia con los datos abiertos enlazados, ya que suponen una transformación en la forma en que se organizan, acceden, relacionan y visualizan los datos.

Hoy en día existen varias iniciativas en Bibliotecas Nacionales que han incursionado –y son referentes– en el desarrollo de proyectos de datos enlazados abiertos para la gestión de sus recursos, dentro de estas se pueden mencionar las que agrupan los autores Torre-Bastida, González-Rodríguez y Villar-Rodríguez haciendo énfasis en el *Europeana Linked Open Data*, que ha permitido que diferentes bibliotecas europeas puedan «estructurar y representar los datos de instituciones de patrimonio cultural» (2015, 115). Aunque, mencionan también a la *British National Bibliography* (BNB), que cuenta con una «plataforma de datos enlazados que permite acceder al catálogo y las colecciones por varias interfaces» (2015, 116).

La Biblioteca Nacional de España (BNE) cuenta con una plataforma llamada Datos.bne.es que ha tomado su catálogo bibliográfico y de autoridades y «los registros en Marc21 han sido transformados a RDF y modelados utilizando principalmente las ontologías de la IFLA». La *Bibliothèque Nationale de France* (BnF) y la *Library of Congress Subject Headings* (LCSH) como bibliotecas líderes en el uso de datos enlazados, para la recuperación de información y se encuentra también en *Linked Data Service* de la Biblioteca del Congreso de Estados Unidos.

Cabe resaltar que pese a los esfuerzos de las bibliotecas mencionadas, la comunidad bibliotecaria tiene grandes retos que afrontar para lograr la vinculación de datos abiertos ya que como lo mencionan Peset, Ferrer-Sapena y Subirats-Coll (2011, 171) se deben considerar aspectos como «las herramientas tecnológicas para la conversión; esquemas de datos a utilizar por entidades que ya intercambian datos masivamente con otros formatos; acuerdos legales y licencias para permitir la utilización de estos datos; documentación detallada sobre los procedimientos a seguir».

Por su parte, Agenjo-Bullón y Hernández-Carrascal (2017, 322) mencionan que las dificultades se pueden presentar en «la falta de sistemas de soporte y encontrar programas que incluyan mecanismos de almacenamiento de tripletas, puntos de servicios SPARQL, indexación adecuada e interfaces de consulta y catalogación».

Asimismo, Saha y Das (2020) detallan que para que las bibliotecas implementen datos abiertos enlazados deben trabajar con el modelo de datos RDF, así como convertir sus servicios bibliotecarios existentes en servicios basados en tecnología semántica con la integración de servicios de datos enlazados.

Estos retos suponen la complejidad y amplitud de aspectos que la comunidad bibliotecaria debe considerar para así maximizar el potencial de los datos abiertos enlazados, ya que como Ávila Barrientos (2014, 71) indica «con los datos enlazados las bibliotecas pueden aumentar la visibilidad de búsqueda e interoperabilidad de información».

Lo anterior se refuerza con lo señalado por Peset, Ferrer-Sapena y Subirats-Coll (2011, 171) debido a que un beneficio sobresaliente que obtendrían las unidades de información con la incorporación de datos abiertos enlazados radica en que «compartir datos de forma masiva y estable ayuda en la detección de duplicados, la desambiguación terminológica, el enriquecimiento de los datos o en suministrar información en formatos más transparentes a diferentes comunidades de usuarios».

Se puede rescatar que con proyectos de datos abiertos enlazados en las bibliotecas se lograr una interconexión que enriquece la descripción y contextualización de los recursos bibliográficos, creando una red de relaciones significativas que potencia la experiencia de las comunidades de usuarios y fomenta la colaboración entre distintas unidades de información.

METODOLOGÍA

Esta investigación se basó en la aplicación de un estudio de tipo descriptivo con un enfoque cualitativo, el cual, de acuerdo con Hernández Sampieri, Fernández Collado y Baptista Lucio (2010, 7) «utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación».

Se utilizó el método del estudio de caso con una investigación basada en evidencias, ya que según como lo expone Durán:

es un proceso de indagación focalizado en la descripción y examen detallado, comprensivo, sistemático, en profundidad de un caso definido, sea un hecho, fenómeno, acontecimiento o situación particular. El análisis incorpora el contexto, lo que permite una mayor comprensión de su complejidad y, por lo tanto, el mayor aprendizaje del caso particular. (2012, 128)

Se tomó como caso los registros bibliográficos de la Biblioteca Nacional de Costa Rica para analizarlos y establecer los elementos metodológicos con la finalidad de su conversión al entorno de datos enlazados.

Considerando lo anterior, la propuesta de esta investigación de transformar los datos descriptivos de los registros bibliográficos de la Biblioteca Nacional de Costa Rica a datos enlazados abiertos tomó como base la propuesta metodológica utilizada para la conversión de registros bibliográficos a datos enlazados; en la cual se toma como base el procedimiento definido por el *World Wide Web Consortium* (por sus siglas W3C) y la metodología propuesta por Ávila Barrientos, en su libro *Los datos enlazados y su uso en bibliotecas*, publicado en 2020, la cual fue actualizada en su obra *Recuperación de información con datos abiertos enlazados*, publicada en 2022.

Tomando en cuenta lo mencionado, los pasos de la metodología propuesta para transformar los datos descriptivos de los registros bibliográficos de la Biblioteca Nacional de Costa Rica a datos enlazados abiertos se presentan en dos etapas.

Etapas 1. Normalización y estructuración del conjunto de datos

En esta etapa se desarrollan los siguientes procesos asociados a la selección, limpieza y estructuración inicial del conjunto de datos descriptivos de los registros bibliográficos que se quieren convertir en datos enlazados.

Preparación de las partes interesadas

La W3C establece que se debe informar a las partes involucradas en el proceso de desarrollo de datos enlazados, así como instituir los requerimientos iniciales sobre el tipo de datos a seleccionar, el tipo de estructuración con la que cuentan esos datos, el uso que se le darán a los datos y el tipo de visualización que se le dé a estos.

Selección del conjunto de datos

La selección del conjunto de datos deberá responder a un objetivo o proyecto específico institucional; el seleccionado puede encontrarse en una base de datos que posea la institución o puede ser un conjunto de datos creado específicamente para el desarrollo del proyecto.

Modelado de datos

Tomando como base lo establecido por Ávila-Barrientos (2020, 132), para el proceso de modelado de datos, es necesario establecer lo siguiente:

1. Utilización de normas internacionales para el modelado de los datos con RDF.
2. Tener una fuente de extracción de los datos de documentos textuales que idealmente se deben de obtener de obras digitales.
3. Contar con un conjunto de elementos de metadatos descriptivos tomados del esquema de metadatos seleccionado.
4. Sintaxis del perfil: con la utilización del lenguaje de marcado XML crear las configuraciones para vincular los datos y que sean accesibles en el contexto web, conformando de esta manera la estructura de hipervínculos, para comunicarlas en la web mediante el protocolo HTTP.

Uso de vocabularios normalizados

Se recomienda el uso de vocabularios normalizados que faciliten la interconexión de los datos publicados en la Web mediante el empleo de URI's. De acuerdo con las pautas establecidas por el Consorcio World Wide Web (W3C), estos vocabularios deben cumplir con características esenciales, tales como disponer de documentación clara y completa, deben ser autodescriptivos, lo que implica que su estructura y significado se encuentren definidos en el propio vocabulario, de igual forma, resulta crucial que estos recursos estén accesibles durante largos periodos de tiempo y cuenten con direcciones URL persistentes, asegurando su disponibilidad continua.

Etapa 2. Conversión de datos

Para esta etapa se llevan a cabo los siguientes procesos que permiten la conversión de datos descriptivos de los registros bibliográficos a datos enlazados, una vez que estos han sido estructurados.

Establecimiento de URI's

Es necesario crear URI's HTTP que permitan la identificación y vinculación de datos en la Web. Esta conversión es posible gracias a la generación de declaraciones RDF; para dicha conversión se requiere contar con una herramienta de software que permita el enlace del conjunto de datos y su visualización.

Las URI's establecidas se desarrollan a partir de las tripletas creadas en RDF utilizando los siguientes atributos de los datos bibliográficos: mención de título, creador, mención de edición, datos de publicación, descripción física, tipo de contenido, medio y soporte y descriptores temáticos. Para los datos de autoridades se consideraron los siguientes atributos: nombre, lugar asociado, campo de actividad, afiliación y ocupación.

Promoción del acceso a los datos enlazados

Según la W3C el acceso a los datos enlazados se puede asegurar mediante acceso directo a través de URI, descarga de archivos RDF, suministro de SPARQL endpoint y uso de API.

Publicación de los datos enlazados

Los datos enlazados deben estar accesibles en la Web para que puedan ser consultados por la comunidad de personas usuarias de la información.

Especificación de licencia de uso de datos

Se recomienda la utilización de licencias abiertas para las herramientas de software que son necesarias para el manejo y publicación de los datos enlazados.

Establecer un acuerdo de responsabilidad del editor

Este acuerdo consiste en la creación de un planeamiento que permita que los datos enlazados puedan ser reutilizados y estén disponibles en la Web de forma constante y estable.

PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

En esta investigación se siguió la metodología expuesta anteriormente tomando como caso de estudio los registros catalográficos del acervo de la Biblioteca Nacional de Costa Rica, obteniendo los siguientes datos.

Etapa 1. Normalización y estructuración del conjunto de datos

Preparación de las partes interesadas

Se procede a informar a la jefatura de la Unidad Técnica del Sistema Nacional de Bibliotecas de Costa Rica (SINABI, unidad encargada de realizar el registro catalográfico de los recursos bibliográficos con los que cuenta la Biblioteca Nacional de Costa Rica) sobre el proyecto a desarrollar para realizar la solicitud del conjunto de datos.

Selección del conjunto de datos

Para el 2022, año en el que se realizó el estudio, el acervo bibliográfico de la Biblioteca Nacional tenía registrado en la base de datos un total 116 822 títulos en las recuperables en la etiqueta 245, de Mención de título del Formato MARC; considerando las autoridades, la base de datos incluye un total de 97 040 autores. Sin embargo, se seleccionaron únicamente 77 registros bibliográficos y 200 registros de autoridad tomando como criterio de selección los registros ingresados en el sistema de gestión Janium (sistema de automatización de bibliotecas utilizado por el SINABI) a partir del 2021, que cuentan con las etiquetas actualizadas en MARC, las cuales incluyen los atributos de los registros que fueron ampliados con la utilización de RDA en la Unidad Técnica.

Se procedió con el proceso de depuración que consistió en eliminar etiquetas MARC que no se consideraron para enlazar, así como la revisión de aspectos ortográficos o de presencia de caracteres incorrectos presentes en los registros.

Con el conjunto de datos depurado en Excel se procede a importarlo al programa Open Refine el cual permite cargar datos de gran volumen para procesos de depuración o análisis.

Modelado de datos

El modelado de los datos se lleva a cabo en la herramienta mencionada Open Refine, en el cual se agregó el esquema RDF que permitió establecer las tripletas para asignar la semántica de los datos y posteriormente vincularlos.

Se utilizó el modelo de datos de descripción bibliográfica BIBFRAME, considerando dos de sus niveles de abstracción o clases principales, la obra y la instancia.

Uso de vocabularios normalizados

Mediante la incorporación del esquema RDF al sistema Open Refine se realizó la integración de vocabularios interoperables semánticamente. Considerando esto, se trabajó con Linked Open Vocabularies (<https://lov.linkeddata.es>) y se toma el vocabulario de Bibframe, el cual consiste en las propiedades que describen las clases principales. Las propiedades que se tomaron en cuenta son las siguientes: *responsability statement, date, agent, place, content, media, carrier y subject*.

Etapa 2. Conversión de datos

Establecimiento de URI's

A partir del vocabulario de BIBFRAME incluido al programa Open Refine se crearon las tripletas que permitieron establecer las URI's para relacionar los datos. Estas tripletas se definieron a partir del conjunto de datos de los registros bibliográficos seleccionados de la Biblioteca Nacional de Costa Rica.

Por ejemplo, para los registros bibliográficos se relacionaron datos como el título, autor, datos de publicación, campos de materia, entre otros. En el caso de los registros de autoridad se enlazaron mediante las tripletas datos como el nombre personal, lugar asociado, campo de actividad, afiliación, entre otros. Un ejemplo de la visualización de la conformación de las tripletas en el sistema Open Refine es el que se muestra en la *imagen 1*.

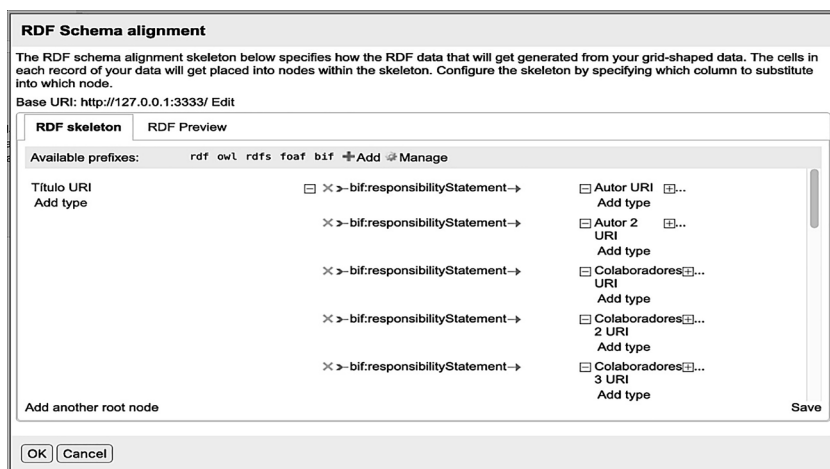


Imagen 1. Tripletas RDF en Open Refine
Fuente: Elaboración propia.

En la *imagen 1* se pueden visualizar las vinculaciones con el título, el cual dentro de la tripleta de RDF corresponde al sujeto; seguido, se realiza la vinculación del sujeto con el predicado que corresponde a la mención de responsabilidad. Finalmente se debe vincular el objeto, que corresponde a los autores y los colaboradores asociados a los títulos, en este caso a los objetos se les debe asignar un URI para que se pueda hacer la vinculación. Con el procedimiento anterior se tiene establecida una tripleta con los tres elementos principales de definidos en RDF, sujeto, predicado y objeto.

Al establecerse las tripletas mostradas en la *imagen 1* es que se crean las URI's para cada dato relacionado tal como se puede visualizar en la *imagen 2*.

RDF Schema alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: <http://127.0.0.1:3333/> Edit

RDF skeleton	RDF Preview
This is a sample Turtle representation of (up-to) the first 10 rows	
<pre> @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> . @prefix owl: <http://www.w3.org/2002/07/owl#> . @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> . @prefix foaf: <http://xmlns.com/foaf/0.1/> . @prefix bif: <http://id.loc.gov/ontologies/bifframe/> . <http://127.0.0.1:3333/reflexiones> bif:responsibilityStatement <http://127.0.0.1:3333/agular-bulgarelli-jos-francisco>; bif:date <http://127.0.0.1:3333/2017>; bif:agent <http://127.0.0.1:3333/editor-no-identificado>; bif:place <http://127.0.0.1:3333/san-jos-costa-rica>; bif:content <http://127.0.0.1:3333/texto>; bif:media <http://127.0.0.1:3333/no-mediado>; bif:carrier <http://127.0.0.1:3333/volumen>; bif:subject <http://127.0.0.1:3333/micalnea-costarricense>, <http://127.0.0.1:3333/-reflexiones> . <http://127.0.0.1:3333/instrumentos-de-poder-el-rol-de-las-emociones-en-la-opresin> bif:responsibilityStatement <http://127.0.0.1:3333/senz>; bif:date <http://127.0.0.1:3333/2021>; bif:agent <http://127.0.0.1:3333/editorial-costa-rica>; bif:place <http://127.0.0.1:3333/san-jos-costa-rica>; bif:content <http://127.0.0.1:3333/texto>; bif:media <http://127.0.0.1:3333/no-mediado>; bif:carrier <http://127.0.0.1:3333/volumen>; </pre>	
<input type="button" value="OK"/> <input type="button" value="Cancel"/>	

Imagen 2. Establecimiento de URI's en Open Refine
Fuente: Elaboración propia.

Conversión de datos

Para el establecimiento de los datos enlazados se procede a exportar el conjunto de datos en RDF a un software que permita la visualización de los datos mediante grafos semánticos, en este caso se utilizó el software GraphDB, el cual permite la gestión de bases de datos orientadas a grafos para su visualización, consulta y análisis. Lo anterior se refleja en la *imagen 3*.

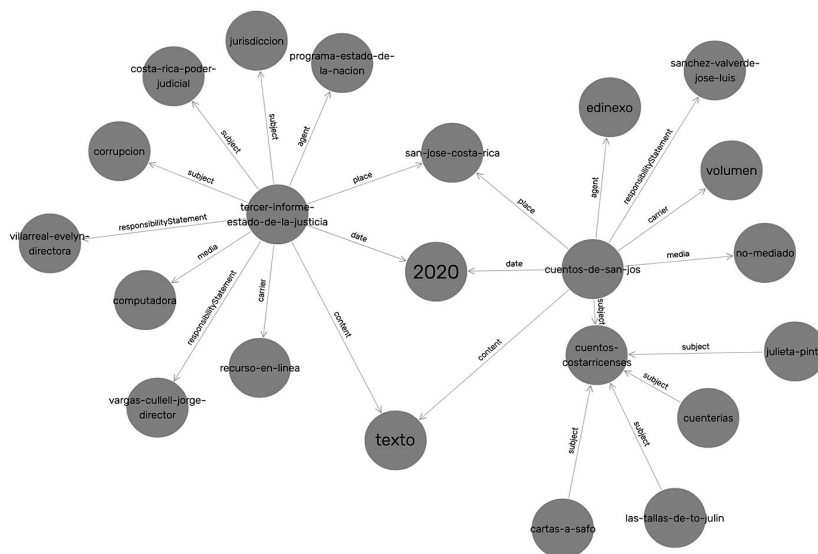


Imagen 3. Ejemplo de visualización de datos enlazados en GraphDB de los registros bibliográficos de la Biblioteca Nacional
Fuente: Elaboración propia.

Publicación y promoción del acceso a los datos enlazados

El acceso a los datos enlazados es posible gracias a los URI's con los que cuentan cada uno de los datos, debido a esto se pueden establecer las vinculaciones entre los datos y así pensar en la integración de estos en el catálogo automatizado en línea de la Biblioteca Nacional de Costa Rica.

Especificación de licencia de uso de datos

Se recomienda que para el proceso completo de transformación de registros bibliográficos a datos enlazados se utilicen programas de software libre para que así los datos puedan ser utilizados y compartidos de forma libre en la Web, pero con la atribución de los derechos de autor correspondientes, considerando el uso de licencias como las *Creative Commons*.

Establecimiento de acuerdos de responsabilidad del editor

Es necesario que en un proceso de desarrollo de datos enlazados se contemple la preservación y acceso a los datos en la Web a lo largo del tiempo mediante el establecimiento de planes o estrategias de conservación y preservación digital.

Analizando cada procedimiento de las dos etapas planteadas se puede considerar que la propuesta presenta una metodología para el manejo de datos descriptivos de los registros bibliográficos y de autoridad que se quieren convertir a datos enlazados.

DISCUSIÓN

Del análisis planteado se puede establecer que para la transformación de datos descriptivos a datos enlazados convergen diversos aspectos que más allá de los procesos que de forma tradicional se han llevado a cabo en bibliotecas para el tratamiento de la información; estos aspectos se relacionan de manera directa con la aplicación de elementos de la Web semántica y el uso de herramientas tecnológicas que permiten la vinculación de los datos en el contexto de ésta.

De lo anterior se resaltan los siguientes aspectos fundamentales para desarrollar el proceso de establecimiento de datos enlazados para la Biblioteca Nacional de Costa Rica:

- Respecto a la selección del conjunto de datos, es recomendable que se seleccionen únicamente datos que van a tener relaciones significativas y que a partir de dichas relaciones se pueda realizar análisis de información, ya que si se consideran la totalidad de los datos que se ingresan en un registro catalográfico no necesariamente se establecen vinculaciones que son de utilidad para las personas usuarias. En este caso los datos que representan relaciones significativas para la Biblioteca Nacional de Costa Rica corresponden a los atributos de mención de título, creador, mención de edición, datos de publicación, descripción física, tipo de contenido, medio y soporte y descriptores temáticos para los datos bibliográficos. Para los datos de autoridades se consideraron los siguientes atributos: nombre, lugar asociado, campo de actividad, afiliación y ocupación.
- El uso de herramientas de normalización tradicionales en el área de procesos técnicos sigue siendo fundamental para el desarrollo de un proyecto de datos abiertos enlazados, ya que por ejemplo si se utiliza el estándar RDA se puede tener un conjunto de datos estructurados y normalizados, lo cual facilita a otras unidades de información a comprender dichos datos. Estándares como este no solo contribuye a una organización consistente de los datos, sino que también posibilita que otras unidades de información comprendan y utilicen de manera efectiva estos datos normalizados. De esta manera, las herramientas de normalización tradicionales continúan siendo un pilar fundamental en la

gestión de datos que, una vez estructurados y estandarizados, pueden llegar a contribuir significativamente al desarrollo de proyectos de datos abiertos enlazados.

- Es necesario utilizar vocabularios normalizados que puedan vincularse en la Web semántica ya que de esta forma es que al conjunto de datos se le puede dar un significado en sus relaciones mediante creación de URI's. Los URI's permiten asignar una dirección única a cada recurso o dato presente en la Web, asegurando así su localización y acceso de manera precisa y consistente. Los URIs no solo actúan como «etiquetas» únicas para los recursos, sino que también son la base para establecer relaciones semánticas entre los datos enlazados.
- Para el proceso de establecimiento de datos enlazados se recomienda utilizar herramientas tecnológicas de acceso abierto para que no se presenten restricciones en la reutilización de los datos en la Web, garantizando una interoperabilidad global, sin restricciones de tipo económicas, legales o técnicas para la institución que desarrolle el proyecto. Una herramienta que permitió el trabajo con datos enlazados fue Open Refine que propició a la mejora de la calidad del conjunto de datos a trabajar, mediante la limpieza de estos al corregir errores de inconsistencias, valores vacíos o duplicados; además, facilitó la transformación de los datos mediante la integración de vocabularios semánticos que permitieron crear URI's de cada tripleta en RDF.
- Respecto al proceso de visualización de datos fue posible a la herramienta Graph DB la cual es una base de datos orientada a grafos que permite una visualización de datos en forma de grafos en donde cada uno se puede interrelacionar y recuperar mediante URI's asignados; Con dicha herramienta se pudo realizar consultas avanzadas para encontrar vinculaciones específicas entre datos, lo cual tiene una gran utilidad para su análisis ya que se pueden identificar así conexiones más evidentes entre los datos. En el conjunto de datos utilizados de la Biblioteca Nacional de Costa Rica, se pudieron establecer relaciones entre sus diversas entidades (obra, expresión, manifestación, personas y corporativas) y entre dichas entidades y sus atributos específicos.

CONSIDERACIONES FINALES

Los datos abiertos enlazados ofrecen un mundo de posibilidades respecto a los procesos de organización de información que se llevan a cabo en bibliotecas ya que estas pueden llegar a vincular sus recursos de información con otros datos que

se encuentren fuera de sus colecciones, enriqueciendo la experiencia de acceso a la información de su comunidad de usuarios. Con el desarrollo de proyectos de datos enlazados se le otorga significado a los datos que únicamente se tenían como referenciales en bibliotecas, lo que es fundamental para la gestión de información, mejorando los procesos de búsqueda y comprensión de esta.

Asimismo, se debe rescatar la importancia de la utilización de las tecnologías dentro del entorno de la Web semántica como lo es RDF, ya que proporciona un marco estructurado para modelar y representar relaciones entre diferentes datos mediante el uso de tripletas, siendo esencial en la creación de vinculaciones semánticas generando así interoperabilidad y enriquecimiento de conjuntos de datos.

El procedimiento desarrollado como parte de la propuesta metodológica planteada en esta investigación puede ser replicado en otras bibliotecas, ya que para dar inicio con un proyecto de implementación de datos abiertos enlazados lo necesario es contar con un conjunto de datos que se quiera vincular, sin dejar de lado el análisis de los beneficios que se quieran obtener de un proyecto así en función de las necesidades de información de la comunidad de usuarios que se atiende, considerando que dichos beneficios pueden incluir la recuperación eficiente de información, descubrimiento de recursos relacionados, navegación en la Web semántica, acceso a metadatos enriquecidos mejorando la comprensión y elección de recursos y el intercambio de datos entre diferentes bibliotecas o sistemas de información.

Con la implementación de datos abiertos enlazados en la Biblioteca Nacional de Costa Rica se pueden mejorar las funcionalidades de sus catálogos automatizados en línea; con procesos más eficientes de recuperación de información y una vinculación mayor entre sus datos y datos externos, impulsando de esta manera la reutilización de información.

Considerando el proceso de conversión de registros catalográficos propuesto, se puede concluir que la Biblioteca Nacional cuenta con grandes fortalezas para ejecutar un proyecto de este tipo, ya que al contar en este momento con gran parte de su colección normalizada con el estándar RDA permite que el uso de sus datos sea más eficiente y responda así a los procesos internacionales de normalización de información compartiéndose así con otras instituciones.

Por otro lado, queda en evidencia un reto que puede afrontar la Biblioteca Nacional de Costa Rica, y tiene que ver con el proceso de capacitación de su personal en el uso de la Web semántica, ya que si se quiere llevar a cabo un proyecto de datos enlazados se debe contar con conocimientos en el funcionamiento de ésta, en el uso de herramientas tecnológicas que permitan realizar vinculaciones semánticas, así como la configuración de sus catálogos automatizados para que puedan integrarse con los elementos de la Web semántica.

Pese a los posibles retos que se puedan presentar para el desarrollo de un proyecto de datos enlazados en la Biblioteca Nacional de Costa Rica, es recomendable que dicha institución incursione en este ámbito para que se logre posicionarse como pionera y principal referencia para otras bibliotecas en Costa Rica, pues estos brindan una plataforma sólida para mejorar la calidad y el alcance de los servicios que ofrecen en la actualidad y dan la posibilidad de emprender en nuevos servicios o productos de información, fomentando la colaboración, la innovación y el acceso libre al conocimiento.

REFERENCIAS

- Agénjo-Bullón, X. y F. Hernández-Carrascal. 2017. El estado de los datos vinculados en bibliotecas 2015. *Anuario Think EPI* 11: 321–23.
<https://doi.org/10.3145/thinkepi.2017.60>.
- Amaya, S. H. 2018. El rol de las bibliotecas nacionales en la era digital y los entornos virtuales. *Realidad y reflexión* (48): 53-61.
- Anglada, L. 2016. Encontrar, agrupar, enlazar: las promesas de las RDA. *Anuario ThinkEPI* 10: 152-157.
- Ávila Barrientos, E. 2014. Aplicación de *Linked Data* en la Biblioteca Digital Académica: Una alternativa para el apoyo a la investigación. Tesis de maestría, Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras.
- Ávila Barrientos, E. 2020. *Los datos enlazados y su uso en bibliotecas*. México: Instituto de Investigaciones Bibliotecológicas y de la Información, UNAM.
- Ávila Barrientos, E. 2022. Recuperación de información con *Linked Open Data*. *Investigación bibliotecológica* 36 (91): 125-146.
- Berners-Lee, T. 2006. Linked data – Design issues.
<https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., J. Hendler, y O. Lassila. 2001. The Semantic Web - A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Scientific American* 284 (5): 34-43.
- Durán, M. M. 2012. El estudio de caso en la investigación cualitativa. *Revista Nacional De Administración* 3 (1):121-34.
<https://doi.org/10.22458/rna.v3i1.477>.
- Fermoso-García, A.-M., M.-I- Manzano García, M. Mateos-Sánchez y C. Hernández-Tamayo. 2018. Sistema de modelado semántico para catalogación, clasificación, consulta y publicación en abierto de información bibliográfica. *El profesional de la información* 23 (2): 410-418.
- Hernández Sampieri, R., C. Fernández Collado, y P. Baptista Lucio. 2010. *Metodología de la investigación*. 5° ed. México: McGRAW-HILL.
- Moreno Mejía, J. F. 2020. La Web semántica y organización de información, en *Organización de la información con RDA: su presencia en los catálogos de bibliotecas de América Latina*, coordinado por F. Felipe Martínez Arellano, S. M. Salgado Ruelas y P. de la R. Valgañón, 402-419. México: Instituto de Investigaciones Bibliotecológicas y de la Información.

- Peset, F., A. Ferrer-Sapena e I. Subirats-Coll. 2011. Open data y linked open data: su impacto en el área de bibliotecas y documentación. *El profesional de la información* 20 (2): 165–73.
- Rodríguez Calvo, D. 2022. Conversión de los registros bibliográficos a registros de datos enlazados en la Biblioteca Nacional de Costa Rica. Tesis de maestría, Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras.
- Saha, B. y R. Das. 2020. Producing Linked Open Dataset from Bibliographic Data with Integration of External Data Sources for Academic Libraries. *Library Philosophy & Practice*, 1–12.
- Sakr, Sh., M. Wylot, R. Mutharaju, D. Le Phuoc e I. Fundulaki. 2018. *Linked Data Storing, Querying, and Reasoning*. Suiza: Springer.
- Salta, G. 2021. La implementación de RDA en la Argentina: Implicancias en las prácticas de organización de la información seguidas por las bibliotecas. *Información, cultura y sociedad* 45: 15-32.
- Salta, A. 2017. Sistemas Para La organización Del Conocimiento: definición y evolución histórica. *E-Ciencias de la información* 7 (2):1-18.
<https://doi.org/10.15517/eci.v7i2.26878>.
- Torre-Bastida, A.-I., M. González-Rodríguez y E. Villar-Rodríguez. 2015. Datos abiertos enlazados (LOD) y su implantación en bibliotecas: iniciativas y tecnologías. *El profesional de la información* 24(2), 113-120.
- Wallheim, H. 2016. From Complex Reality to Formal Description: Bibliographic Relationships and Problems of Operationalization in RDA. *Cataloging & Classification Quarterly* 54 (7): 483-503.
<https://doi.org/10.1080/01639374.2016.1200169>

Para citar este texto:

- Rodríguez Calvo, Desiree. 2023. “Creación de datos abiertos enlazados en la Biblioteca Nacional de Costa Rica”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 37 (97): 75-95.
<http://dx.doi.org/10.22201/iibi.24488321xe.2023.97.58809>