

- [Contents](#) |
- [Author index](#) |
- [Subject index](#) |
- [Search](#) |
- [Home](#)

Linked open data en bibliotecas: estado del arte

**[Francisco Javier Martínez Méndez](#), [Juan-Antonio Pastor-Sánchez](#)
and [Rosana López Carreño](#).**

Introducción. Los datos enlazados forman parte del proyecto de web semántica. Los datos enlazados bibliotecarios son los conjuntos de datos relacionados con estas instituciones. Desde el principio del lanzamiento de esta idea se pensaba que las bibliotecas y archivos serían de las primeras organizaciones en hacer un uso masivo de esta tecnología.

Metodología. Se ha llevado a cabo una exhaustiva revisión bibliográfica especializada, dirigida a averiguar el estado de la cuestión tras varios años de desarrollo de esta tecnología.

Análisis. Se ha trabajado fundamentalmente con estudios de caso, obteniendo aquellas reflexiones y conclusiones más relevantes de los participantes en estos proyectos.

Resultados. Lo cierto es que la premisa inicial no se ha verificado de forma general. El desarrollo de esta tecnología es todavía parcial y ha encontrado cierto rechazo de una parte de los profesionales debido a su complicada curva de aprendizaje y a la dispersión de los vocabularios desarrollados. Además de contar con recursos económicos, técnicos y humanos, se necesita un compromiso claro para el desarrollo de esta tecnología.

Conclusiones. A pesar de todos los problemas identificados, parece claro que los datos enlazados definen las bases conceptuales y tecnológicas para incorporar los conjuntos de datos bibliotecarios dentro del entorno de la web de datos, siendo preciso superar las reticencias de los profesionales y los problemas con las licencias de uso de los datos.

[Abstract in English](#)

Introducción

El término *linked open data* (*datos abiertos enlazados* en español) es el resultado de la fusión de dos ideas: *linked data* (datos enlazados) y *open data* (datos abiertos). La primera consiste en usar la web para crear vínculos entre datos de diferentes fuentes o sistemas heterogéneos. Técnicamente, '*LOD se refiere a los datos publicados en la web para que sea legible por máquina, su significado se define de forma explícita y está vinculado a otros conjuntos de datos externos*' (Bizer et al., 2008, p. 2). La segunda idea es un movimiento que promueve la liberación de datos de las organizaciones, generalmente no textuales y en formatos reutilizables a

través de su publicación en la web, descritos en formato RDF (Resource Description Format) y con indicación de la relación que puede existir entre esos datos y otros (Peset et al., [2011](#), p. 165).

Con la tecnología de los datos abiertos enlazados es posible identificar, describir, enlazar y relacionar datos estructurados de manera análoga a la función relacional de los sistemas de bases de datos. Así entendido, es un meta-modelo que aporta un marco para definir, diseñar, implementar y mantener esquemas y vocabularios de cualquier tipo y tamaño en un dominio (Alemu et al., [2012](#), p. 553). Este concepto es cada vez más importancia en la gestión de información y se utiliza por muchas organizaciones para crear portales, plataformas digitales, servicios basados en internet y aplicaciones destinadas a la reutilización de estos conjuntos de datos. Berners-Lee ([2007](#)) señala que agregar una página proporciona contenido, pero que agregar un enlace proporciona organización, estructura y respaldo a la información publicada en la web, aumentando así su valor. Bauer y Kaltenböck ([2012](#)) lo presentan como la solución de dos problemas: (1) la falta de conectividad de las bases de datos, actualmente diseñadas como silos de información de difícil acceso y con pocas posibilidades de reutilización, y (2) los datos de muchas de estas aplicaciones son difícilmente interoperables al ser muy complicado extraerlos de las mismas porque los metadatos y esquemas de información no están separados debidamente de la lógica de la aplicación. En parte, esta situación deriva del clásico pensamiento en el que sólo unos expertos pueden comprender y gestionar los sistemas de bases de datos, algo obsoleto porque la competencia en los mismos por parte de la comunidad de usuarios de internet es mucho mayor que la existente en los inicios de estos sistemas.

Linked open data debe permitir que estos datos estén abiertos a cualquier tipo de aplicación por medio del uso de estándares abiertos (como, por ejemplo, RDF). Existe una cierta analogía a la época en la que se popularizó el uso de la web y muchas organizaciones no entendieron la necesidad y los beneficios de publicar documentos en ese incipiente entorno. Esta circunstancia fue superada con el paso del tiempo, lo que nos permite pensar que sería positivo para las organizaciones vislumbrar las ventajas competitivas asociadas a la publicación de los datos en la web de forma abierta y enlazada.

***Linked open data* en las bibliotecas: áreas de actuación.**

El World Wide Web Consortium (W3C) identifica a la comunidad bibliotecaria como uno de sus principales focos de interés. El Library Linked Data Incubator Group comenzó en 2010 a trabajar en la publicación de datos abiertos, estructurados y enlazados en bibliotecas (Peset et al., [2011](#), p. 168). Bauer y Kaltenböck ([2012](#), p. 27) las ponen como ejemplo de aplicación de *linked open data* debido al enorme potencial de la creación de datos interrelacionados y al intercambio y utilización conjunta de los mismos con instituciones no bibliotecarias, además de la confianza en el crecimiento de la web semántica. Las insuficiencias de las normas actuales de representación de datos bibliográficos han sido estudiadas por expertos y organizaciones internacionales del contexto bibliotecario, manifestándose interés en establecer una representación y comunicación significativa de los datos bibliográficos en el entorno de la web de datos. Esto es algo ya aparecido al explotar tecnologías semánticas para la publicación y el consumo de datos estructurados (Zapounidou et al., [2013](#)). El desarrollo de esta tecnología en el seno de las bibliotecas puede ser la principal innovación a introducir en este sector, constituyéndose en una apuesta necesaria para la adopción de las tecnologías de la web semántica (Latif et al., 2016, 156). Alemu et al. ([2012](#), p. 557-562) exponen cinco beneficios principales para su adopción:

1. Apertura e intercambio de metadatos de forma sencilla con datos reutilizables y redistribuibles usando identificadores únicos (obviando duplicación de datos y habilitando enlaces cruzados entre recursos de varias bibliotecas). Se podrán vincular documentos a fuentes de información no bibliotecarias (Google, Wikipedia o Amazon, por ejemplo). Wang y Yang ([2018](#), p. 12-13) recuerdan que el 82% de las búsquedas de información provienen de un motor de búsqueda, no parece lógico desaprovechar la posibilidad de acceder a los catálogos de las bibliotecas por esta vía.
2. Facilitar el descubrimiento fortuito de los recursos de información: Los actuales módulos de consulta en línea de los catálogos bibliotecarios poseen bajos niveles de pertinencia en sus búsquedas. A pesar del esfuerzo realizado en la descripción con metadatos de los registros, están igual de lejos de interactuar con la web como los catálogos manuales (Fallgren et al., [2014](#), p. 129). Los datos enlazados vincularían bases de datos y fuentes relevantes de información, disponiendo para su consulta objetos de información de fuentes externas gracias al uso de los identificadores internacionales de recurso. Esto permitirá saltar de una fuente a otra, introduciendo en la biblioteca digital los procesos de descubrimiento a manera y usanza de como las estanterías abiertas de libre acceso lo proporcionan a los usuarios en sala.

3. Identificación de patrones de uso de recursos, Zeitgeist y metadatos emergentes: Se tiende hacia una biblioteca digital respaldada por objetos de información mixtos (basados en estándares y construidos socialmente), aumentada por la semántica adicional y las capacidades tecnológicas de los datos enlazados (IRI, RDF, OWL, SPARQL principalmente). Esto aportará descripciones con metadatos más ricas y reflejará de manera más precisa las diversas acepciones terminológicas empleadas por los usuarios. Los datos abiertos también facilitarán el análisis de los patrones de uso de los recursos (número de clics, descargas totales, etiquetas de uso frecuente y recursos altamente recomendados).
4. Navegación basada en facetas: usar diferentes dimensiones y propiedades al navegar por un sistema de información, con el objetivo de mejorar la detectabilidad (Morville, [2005](#)). El uso de facetas mejora el uso de taxonomías porque un artículo estará en un único lugar en la jerarquía, limitando formas alternativas de navegación. El uso del lenguaje SKOS ('Simple Knowledge Organization System') resulta esencial para el modelado, reutilización e interconexión de sistemas de navegación facetados para clasificar contenidos de forma eficiente (Putkey, [2011](#)).
5. Metadatos enriquecidos con enlaces: un libro descrito de forma enriquecida podrá vincularse con recursos relacionados con su contenido (la biografía del autor, por ejemplo). Estos metadatos enriquecidos con enlaces contextuales y relevantes permitirán navegar entre catálogos distintos de la misma biblioteca (o de bibliotecas diferentes e incluso motores de búsqueda).

Tälleras ([2013](#)) estudió proyectos de datos abiertos enlazados desarrollados por las principales bibliotecas nacionales: (1) el marco para la descripción bibliográfica de recursos de la Biblioteca del Congreso (que terminó siendo el modelo BIBFRAME); (2) la conversión de los catálogos de datos en implementaciones RDF desarrollada por las bibliotecas nacionales británica, sueca y francesa y (3) la disposición por parte de la empresa OCLC ('Online Computer Library Center') de un conjunto de datos enlazados a partir de los términos presentes en los tres niveles más altos de la Clasificación de Dewey en 23 idiomas. Estos proyectos demuestran el deseo de cambio de estas instituciones, no del todo satisfechas con los metadatos producidos hasta ahora y con la esperanza de mejorar la calidad del registro bibliográfico, facilitando su conversión e interrelación.

En un principio, la adopción de la tecnología de datos abiertos enlazados aparentaba ser urgente (especialmente en las bibliotecas nacionales) y se iban a hacer accesibles los registros bibliográficos de forma libre y abierta (Wilson, [2010](#) en Alemu and Stevens, [2015](#), p. 72), esto no se ha conseguido del todo. A pesar del alto interés en esta tecnología, todavía existe debate sobre los cambios introducidos en los modelos tradicionales centrados en los formatos de registros y funciones bibliotecarias (la descripción de recursos y de acceso RDA, los requisitos funcionales para registros bibliográficos FRBR ('Functional Requirements for Bibliographic Records') y la catalogación en formato MARC ('Machine Readable Cataloging') y se mantiene la duda sobre la validez de estos sistemas. Alemu et al. ([2012](#), p. 555) fijaban el año 2011 (quizá con excesivo ímpetu) como clave para el impulso de los datos abiertos enlazados en las bibliotecas porque ese año coincidieron tres informes elaborados por la Biblioteca del Congreso, las bibliotecas de la Universidad de Stanford y, finalmente, el W3C. Estos tres informes estaban dirigidos a abordar la importancia de la apertura de los metadatos de las bibliotecas hacia formatos de fácil acceso y reusables. Saorín López et al. ([2013](#)) añadían otros hitos al despegue de esta tecnología en ese momento:

la web de datos ocupa gran parte de las ponencias de la Conferencia ELAG ('European Library Automation Group') bajo el lema 'It's the context, stupid!'; Europeana incorpora linked open data en su estrategia y desarrolla proyectos piloto, etc.

Los mismos autores avisaban ya de que se trataba de un movimiento arriba-abajo falto de la suficiente masa crítica y diversidad de actores.

Objetivo y método

El objeto de este trabajo es ofrecer una visión multinivel del grado de desarrollo de las tecnologías de datos abiertos enlazados en bibliotecas. Para ello se han analizado los fundamentos de los aspectos relacionados con la catalogación y representación de recursos bibliográficos, así como su integración en el ecosistema de publicación de conjuntos de datos mediante tecnologías de la web semántica. También se ha estudiado su despliegue e implantación a nivel internacional en cinco bibliotecas nacionales: Biblioteca del Congreso, Biblioteca Británica, Biblioteca Nacional de Francia, Biblioteca Nacional de España y la Biblioteca Nacional de Alemania. De forma más específica, como objetivo secundario, se realizó el mismo análisis en el ámbito de las bibliotecas españolas. La investigación se ha llegado a cabo a partir de una exhaustiva revisión

bibliográfica y de la visita a los sitios web de las bibliotecas que más han emprendido el camino de transición hacia el paradigma de los datos abiertos en sus instituciones.

Implantación de los datos abiertos enlazados en bibliotecas: estado de la cuestión

A pesar de los buenos deseos iniciales, el impulso proporcionado a la tecnología *linked open data* en las bibliotecas no ha sido concluyente. Casi al inicio de todo, el informe del W3C Library Linked Data Incubator Group (Baker et al., 2011) hablaba de la lentitud en su adopción, enfatizando la importancia de convertir en abiertos y accesibles los datos para hacerlos compartibles, extensibles y reutilizables (Alemu et al., 2012, p. 555). No se debe olvidar un problema de inicio: los estándares de uso extendido en bibliotecas – formato MARC y protocolo Z39.50 principalmente – se diseñaron para uso interno de la comunidad bibliotecaria y no resultan de sencilla reutilización fuera de este ámbito. Si se añaden los esfuerzos desarrollados por las instituciones, asociaciones bibliotecarias y proveedores de servicios de información (como OCLC) en el mantenimiento del formato MARC, la transición nunca será cuestión trivial, será necesario superar barreras aptitudinales y muchas actitudinales. El citado informe reflexionaba también sobre la disparidad terminológica existente en los estándares web y las normas bibliotecarias. Alemu et al. (2012, p. 557) mencionan la cierta dificultad del manejo de estas tecnologías por parte de los profesionales bibliotecarios, recordando que la web actual se ha expandido gracias a la simplicidad del proceso de edición de una página y enlazarla con otras, procesos elementales y rutinarios que no aportan nivel de complicación a la actividad profesional.

El punto de partida es un escenario donde el formato MARC desempeña un papel fundamental en el intercambio de registros bibliográficos. Esta situación de preponderancia se ve reforzada porque las principales normas de catalogación (ISBD, AACR2 y RC) y las directrices para los ficheros de autoridades están basadas en este formato (Cormenzana, 2017). La Library of Congress (LC) planteó en 2012 el desarrollo de BIBFRAME como modelo sucesor de este formato. Se trata de un proyecto de adaptación de los datos bibliotecarios hacia el universo de los datos enlazados por medio de la interconexión que propicia la web y sus tecnologías de búsqueda (Library of Congress, 2012). BIBFRAME puede expresarse en RDF (el estándar del consorcio W3C para describir recursos mediante datos estructurados) para la publicación y el intercambio de datos. Este marco extiende la estructura de los hipervínculos mediante el uso de tripletas o sentencias RDF, declaraciones sobre recursos web en forma de expresiones sujeto-predicado-objeto identificadas por medio de un identificador internacional de recurso. Estos conceptos generales representan la base del modelo BIBFRAME (O'Dell, 2014; Jin et al., 2016, Cormenzana y López-Burrull, 2018). Sus beneficios pueden presentarse en términos de viejo y nuevo paradigma en el acceso a la información de las bibliotecas. La conexión de los registros bibliotecarios con otros conjuntos de datos posibilitará su localización vía buscador web al asignarles encontrabilidad (el estándar RDF facilita que otras aplicaciones consulten, rastreen y usen los catálogos y los muestren a los usuarios sin necesidad de diseñar aplicaciones específicas de consulta como hasta ahora). En este nuevo paradigma, se podrá acceder a los registros del catálogo bibliotecario por medio de consultas a los motores de búsqueda y también por medio de consultas SPARQL (acrónimo recursivo de ‘SPARQL Protocol and RDF Query Language’), posibilitando su localización más allá de las bibliotecas (algo no posible en la web tradicional donde estos objetos de información forman parte de la “internet profunda”). Para Breeding (2015, p. 24), la tecnología de datos abiertos es una oportunidad para sacar a la luz más recursos dentro de los sistemas de información de las bibliotecas, brindando oportunidades para aprovechar el contenido y las relaciones más allá de un índice de descubrimiento a partir de unos términos clave en las operaciones de búsqueda.

La tecnología de datos abiertos es el ecosistema más razonable para los datos bibliográficos, pero creemos necesario comprender del todo la integración de BIBFRAME en el mismo. Una de las claves reside en la interoperabilidad, no solo entre los propios datos sino en el enfoque de desarrollo que, a partir de un modelo conceptual, conduce a un entorno de codificación. Para ello, hay que recordar los trabajos de IFLA (‘International Federation Library Associations and Institutions’) que crearon el modelo FRBR (‘Functional Requirements for Bibliographic Records’) y que evolucionaron hasta el modelo *library requirements model*. La experiencia acumulada en estos trabajos recomendó resolver determinadas inconsistencias entre los tres modelos conceptuales FRBR, FRAD (para datos de autoridades) y FRSAD (para datos de materias). Fruto de ello fue la propuesta (también de IFLA) del *library requirements model* actualizado en agosto de 2017. No obstante, como modelo conceptual que es, no aborda determinadas cuestiones relacionadas con los principios prácticos relacionados con la catalogación. Es aquí donde entran en juego las instrucciones ICP (‘International Cataloguing Principles’) actualizadas en 2016 y que conectan las prácticas de catalogación con el enfoque

conceptual para la identificación de las diferentes entidades del *library requirements model* (obra, expresión, manifestación, manifestación, ítem, agente, etc.). La experiencia adquirida evidencia que una cosa es definir un modelo y otra muy distinta es la operatoria para acometer la descripción de recursos bibliográficos a partir de aquel.

Por su parte, las normas RDA ('Resource Description and Access') ofrecen una serie de directrices para describir recursos bibliotecarios y culturales mediante metadatos adecuados para su uso en entornos de datos abiertos bibliotecarios. Estas normas se concibieron para sustituir las norteamericanas AACR2 ('Anglo-American Cataloguing Rules') cuyo origen se remontaba al año 1978. Las normas RDA se publicaron en 2009, basándose en los requerimientos del modelo FRBR siendo su última actualización de 2016 para adaptarse precisamente a los cambios introducidos en el modelo de requerimientos LRM. Es preciso considerar que este estándar, al tratarse de normas para catalogar, conserva su independencia respecto a su implementación y el entorno de aplicación. BIBFRAME desarrolla las directrices de catalogación y compartición de recursos bibliográficos de RDA partiendo de la incardinación de los datos bibliotecarios en el ecosistema de los datos abiertos enlazados en entornos bibliotecarios. Considerando todo lo expuesto anteriormente, es lógico concluir que la versión actual (la 2.0) se alinea con RDA y con el modelo LRM en última instancia. Dicha integración se puede comprobar en cómo el vocabulario RDF de BIBFRAME 2.0 expresa mediante clases y propiedades las correspondientes clases y atributos de los requerimientos del *library requirements model* (Cormenzana, [2017](#)).

Parece aceptado que BIBFRAME sustituirá al formato MARC aunque el cambio sea más lento de lo deseado. Por ello, el descubrimiento de recursos en las bibliotecas tendrá que hacer similar transición, optimizando la indexación de los registros bibliográficos apoyados en estructuras orientadas a datos enlazados. Se puede mejorar el contenido de los índices siguiendo los enlaces a los que se hace referencia e indexando también ese contenido, algo que no es posible con el formato MARC pero sí es habitual en el módulo de rastreo de un motor de búsqueda. Los identificadores únicos de recursos derivados de las estructuras BIBFRAME permitirán introducir mejoras en el acceso y la vinculación a los conjuntos de datos. Un beneficio derivado será que las colecciones estarán bajo el control de la biblioteca, sin limitaciones de acuerdos comerciales o de derechos de autor, cada vez más expuestas como datos abiertos enlazados (Breeding, [2015](#), p. 25). Si bien el desarrollo de BIBFRAME es crucial para poder percibir todo el potencial de las normas RDA, este modelo no es solo el reemplazo del clásico formato MARC, está destinado a ser un entorno bibliográfico completamente nuevo que hará posible la interacción de los metadatos (Fallgren et al., [2014](#), p. 129-131), convirtiendo el paradigma de los datos abiertos enlazados en algo mucho más accesible y práctico para las bibliotecas. En ningún momento debe considerarse como una amenaza para la catalogación, será útil para las bibliotecas la coexistencia entre el nuevo modelo y el contexto actual, compartiendo datos bibliográficos en la web y aportando mejores metadatos sobre los recursos de conocimiento basados en estándares internacionalmente reconocidos.

Así, los metadatos de los registros serán fácilmente aprovechables por los motores de búsqueda, transformando catálogos y colecciones de uso restringido en repositorios de conocimiento activos y susceptibles de ser recuperados por un motor de búsqueda (Tharani, [2015](#), p. 16). Para difundir el conocimiento de BIBFRAME entre la comunidad bibliotecaria la Biblioteca del Congreso desarrolló una ontología inicial siguiendo el consejo de las necesidades expresadas en el contexto del formato MARC, el estándar RDA (en los inicios de su implementación) y el marco de descripción de recursos RDF. Se desarrollaron varias herramientas de acceso universal basadas en esta ontología aportando una vista específica y concreta de los datos BIBFRAME, un transformador para la conversión del formato MARC a BIBFRAME y un editor simple (McCallum, [2017](#), 78). Gran parte de la literatura consultada afirma que en los próximos años la mayoría de los datos de las bibliotecas (millones de registros bibliográficos almacenados en silos independientes unos de otros), estarán disponibles como datos vinculados, libre y abiertamente localizables y accesibles en la web, tal como lo hacen ahora algunas bibliotecas nacionales. Sin embargo, estas bibliotecas todavía afrontan el desafío de obtener datos bibliográficos por medio de los motores de búsqueda:

con un imperativo para apoyar los nuevos medios de descubrimiento, y una gran cantidad de experiencia en la producción de datos estructurados de alta calidad, las bibliotecas son complementos naturales de los datos vinculados (Heath, [2011](#), p. 36).

Resulta muy positivo que casi todas estas bibliotecas alcancen el máximo nivel en la clasificación 5 estrellas de Berners-Lee ([2006](#)) en el desarrollo de datos vinculados abiertos, publicando datos bibliográficos con licencia abierta en formato legible por máquina y no propietario, utilizando estándares del W3C (el marco de descripción RDF y el lenguaje de consulta SPARQL) y vinculando datos a las fuentes externas. Por el contrario, se detecta una serie de problemas a resolver: la escasa disponibilidad de herramientas de soporte, los

mecanismos de control de calidad de los datos, la mejora de los interfaces de consulta, la falta de personal capacitado en este campo, las dificultades para definir los derechos de autor de los datos vinculados abiertos sin olvidar el escaso soporte de datos enlazados que aportan los actuales sistemas de gestión bibliotecaria (en parte por la indefinición y dispersión de los vocabularios). Además, los autores del estudio encontraron poca información sobre el uso de datos vinculados en cada biblioteca analizada. Hallo et al. (2016) esperaban un aumento del número de proveedores de los conjuntos de datos y también la proliferación de aplicaciones para consumir datos vinculados. Otra previsión era que los catálogos se enriqueciesen con recomendaciones y rankings de búsqueda basados en la popularidad de un artículo y la actividad de los usuarios, objetivos tampoco alcanzados. Los profesionales de las bibliotecas deben tener en cuenta la preservación de los conjuntos de datos vinculados y colaborar en la difusión de buenas prácticas para aumentar la participación de la biblioteca en la web semántica.

En el año 2011, el ‘Ontology Engineering Group’ de la Universidad Politécnica de Madrid y la Biblioteca Nacional de España iniciaron un proyecto de publicación de los catálogos bibliográficos y de autoridades de esta institución en conjuntos de datos RDF siguiendo los principios de datos abiertos enlazados. Es un experimento que propone un acercamiento y una exploración diferentes a los catálogos bibliográficos tradicionales por medio de una experiencia de navegación nueva que integra los distintos recursos de la biblioteca, al tiempo que se enriquecen los datos propios con otros externos. Uno de los principales resultados de este proyecto ha sido la gran interacción y compenetración alcanzada entre los especialistas en la web semántica y los expertos en bibliotecas. También se ha propuesto la conservación de fuentes de datos como una actividad transversal paralela a las actividades de especificación, modelado y generación. Al informar y solucionar problemas en las fuentes de datos, se aumenta la calidad de los datos del estándar RDF y de las fuentes, lo que permite ahorrar costos para la institución. La publicación en formato de datos abiertos enlazados puede ayudar a crear un ciclo virtuoso que pueda tener un impacto directo en la calidad de los datos de la biblioteca (Vila-Suero y Gómez-Pérez, 2013, p. 599). Para Santos et al. (2015, p. 17), a partir de ahora se deben seguir tres pasos:

1. Aportar más contenido y más entidades. El modelo de datos y la visualización basada en entidades ha demostrado ser esencial para construir una arquitectura de datos flexible y robusta así que es deseable diseñar más entidades y, con ellas, establecer más relaciones.
2. “Excavar” más hondo. Las fuentes externas de datos no están completamente explotadas. Se necesita un examen detallado de las fuentes disponibles (de entornos bibliotecarios y ajenos a ellos), recogiendo datos valiosos, diseñando y estableciendo puertas de enlace para intercambiar y extraer datos que podrían usarse.
3. Cambiar mentalidades. Los bibliotecarios deben pensar en la forma en que se modelan los datos cuando se procede a la catalogación diaria. La web semántica comienza con el creador de datos humanos facilitando el enlace al grabar datos. La capacitación y el conocimiento son herramientas absolutamente necesarias y la visualización de datos está llena de beneficios para usuarios de todos los niveles.

Smith-Yoshimura (2016), de la empresa OCLC, realizó una encuesta sobre la implementación de los datos abiertos en bibliotecas con el objetivo de averiguar qué instituciones estaban trabajando en proyectos de este tipo y por qué. Participaron 71 instituciones de 20 países (de ellas 22 universitarias y 14 nacionales). Se identificaron hasta 112 proyectos de datos abiertos, de los cuales 46 llevaban más de dos años en producción y 37 no habían comenzado a implantarse. La mayoría de estos proyectos (64) preveían consumir y publicar datos abiertos a la vez. Los datos más frecuentes eran registros bibliográficos, autoridades, metadatos de descripción y ontologías/vocabularios. Los principales problemas para publicar resultaron ser la curva de aprendizaje en el uso de estas tecnologías y los problemas legales relacionados con los derechos de autor. A la hora de consumir datos abiertos enlazados, las dificultades comenzaban con la coincidencia, desambiguación y alineación entre los datos de origen y los recursos de datos vinculados, junto con el mapeo del vocabulario y la reutilización del conjunto de datos publicados. A pesar de los resultados negativos de este informe, Agenjo y Hernández (2017) afirman que “aunque subsistan muchos problemas sobre la calidad de los datos, que subsistirán porque son endémicos en cualquier base de datos, son justamente los aspectos técnicos los menos preocupantes”. Los autores aconsejan centrarse en el objetivo a lograr y no solo en las dificultades técnicas.

Wang y Yang (2018) desarrollan un estudio sobre el uso de la tecnología de datos abiertos enlazados en las bibliotecas y su uso por parte de la comunidad bibliotecaria internacional, prestando especial atención a la empresa OCLC, a la Biblioteca del Congreso y a otras instituciones estadounidenses y del resto del mundo. Los autores afirman taxativamente que ya han pasado casi 20 años desde el inicio de las normas FRBR (entonces RDA) y que se ha esperado mucho tiempo la implantación de BIBFRAME (de hecho, seguirá siendo un

proyecto en desarrollo incluso con la próxima publicación de BIBFRAME Editor 2.0). Anteriormente, Godby y Denenberg (2015) analizaron las similitudes y diferencias entre BIBFRAME y los modelos de datos desarrollados por la empresa OCLC basados en Schema.org concebidos para optimizar las descripciones de los recursos de la biblioteca para el descubrimiento en la web más allá de los catálogos en línea. Tal y como se concluía en el párrafo anterior, el camino a los datos vinculados ha sido accidentado, pero no debe haber vuelta atrás.

En los niveles más altos, la propuesta de OCLC es similar a BIBFRAME, especialmente en la definición de entidades como Trabajo, Instancia, Organización y Persona. Pero los objetivos de estos modelos son diferentes: el primero pretende reforzar la descripción de los recursos bibliográficos a través de la curación de datos y el segundo se centra en aumentar los accesos desde la web a partir de un vocabulario más comprensible por los usuarios. No obstante, mediante el mapeado entre BIBFRAME y Schema.org, es posible que las dos iniciativas puedan participar en proyectos comunes, como desarrollar y testear modelos de descripción comunes para recursos de presencia habitual en bibliotecas (material audiovisual o mapas) complicados de describir actualmente utilizando un único modelo (Godby and Deneberg, 2015).

La Biblioteca del Congreso está particularmente interesada en estos tipos de recursos porque esta biblioteca posee colecciones grandes y variadas de ellos (McCallum, 2017, 81). Con el lanzamiento de BIBFRAME 2.0 en el año 2016 se puso en marcha otro proyecto piloto (más ambicioso que los anteriores) para crear un entorno donde un catalogador pudiera trabajar sin referencia alguna al formato MARC. Este proyecto se divide en dos partes principales: (1) la conversión del catálogo de la biblioteca de formato MARC a BIBFRAME sobre el cual el catalogador creará descripciones de nuevos recursos y (2) la creación de un asistente para interactuar con el catálogo BIBFRAME (McCallum, 2017, p. 82). Fuera del ámbito de la Biblioteca del Congreso, la empresa Zepheira trabaja con varias bibliotecas para convertir sus datos a un formato simplificado de BIBFRAME y hacerlos accesibles vía web (McCallum, 2017, p. 83-84). Esta empresa colaboró en el diseño inicial de BIBFRAME pero posteriormente se encontró con el problema de la complicada visibilidad de los recursos bibliográficos en la web al desarrollar otros proyectos fuera de esta biblioteca.

Jin et al. (2016) desarrollaron un proyecto experimental en las bibliotecas de la Universidad de Illinois convirtiendo cerca de 300.000 registros de libros electrónicos descritos en formato MARC a BIBFRAME, creando un modelo entidad-relación para los libros electrónicos y dos interfaces de búsqueda para los datos abiertos enlazados. Este proyecto sirvió para el enriquecimiento de los datos del catálogo y como base para el desarrollo de un nuevo proyecto que afectaba ya a siete millones de registros en formato MARC. Para estos autores, la catalogación se encuentra en una absoluta transición y BIBFRAME supone un profundo avance dentro la comunidad bibliotecaria, a la que habrá que enseñar a apreciar todas sus ventajas. Zaponidou et al. (2017) integraron monografías dentro de la web semántica comparando cuatro modelos conceptuales: FRBR, FRBR 'Object-Oriented' (FRBRoo), BIBFRAME y EDM ('Europeana Data Model'). La conclusión principal de su estudio fue que la web semántica proporciona una representación semánticamente rica de las descripciones, mientras que la interoperabilidad se ve afectada por los distintos tipos de materiales bibliográficos empleados. Las iniciativas relacionadas con los modelos bibliográficos están aún en desarrollo (como ocurre con los proyectos LD4P y LD4L promovidos por Philip E. Schreur desde la Universidad de Stanford).

Proyectos desarrollados en bibliotecas nacionales

Las bibliotecas nacionales de Francia, Reino Unido, Estados Unidos y España, Hallo et al. (2016) compararon varios proyectos de empleo de la tecnología de datos abiertos enlazados analizando los beneficios asociados a la adopción de esta tecnología: mejorar la visibilidad de los datos, el establecimiento de vínculos con otros servicios online, la transformación de los ítems al modelo SKOS, la recuperación abierta de información, la mejora de la interoperabilidad, la realización de búsquedas basadas en metadatos enlazados con múltiples instituciones y, finalmente, si se podían modelizar entidades de interés relacionadas con el recurso de información (tales como personas, sitios o eventos). También estudiaron qué vocabularios y ontologías publicados en formato de datos abiertos enlazados se utilizaban en estas bibliotecas y los problemas detectados en la publicación.

En su informe de junio de 2017, el grupo de trabajo de ALA sobre datos abiertos enlazados, Brian Rennick (Brigham Young University) comentaba que solo BIBFRAME llevará esta transición a buen puerto porque la utilización de técnicas de optimización en motores de búsqueda para visibilizar un poco más los registros

bibliográficos apenas ha sido tomada en consideración por Google. Tiziana Possemato, de Casalini Libris, presentó la primera fase del proyecto ‘Virtual Discovery Environment’ (SHARE-VDE) cuyo objetivo era apoyar a las bibliotecas en su adaptación a un entorno de datos abiertos enlazados (Cagnazzo, 2017). Los beneficios tienen que ver con el potencial de los datos abiertos enlazados para mejorar la visibilidad y la encontrabilidad de los datos, una amplia reutilización y reutilización de los datos, y la mejora intrínseca de la calidad de estos datos.

Con respecto a este último asunto, las bibliotecas nacionales de Letonia y Países Bajos conciben la publicación de datos abiertos enlazados como una forma de contrato social donde cada organización es responsable de los datos proporcionados, y tiene la tarea de garantizar que los mismos permanezcan disponibles de manera persistente, sin interrumpir el acceso por parte de los usuarios. Desde esta perspectiva, los datos abiertos enlazados implican una profunda reflexión sobre lo que se está haciendo y si se está haciendo de la manera adecuada. En las tablas I y II se presentan, de forma resumida, los beneficios y desilusiones detectadas en el seno de estas instituciones:

Tabla 1: Resumen de los beneficios que se vislumbran en las bibliotecas nacionales por el desarrollo de proyectos de datos abiertos enlazados. Fuente: Hallo et al. (2016, p. 8) y Cagnazzo (2017), con elaboración propia.

Beneficios	GE	SP	LC	FR	GR	LA	LU	WA	NE	PO	UK	SW
Colaboración con otras fuentes (visibilidad)	X		X	X	X	X	X	X	X		X	X
Multilingüismo		X			X							X
Conocimiento web semántica		X	X	X		X					X	
Responsabilidad en la gestión de los datos).						X			X			

Leyendas: GE: Alemania. SP: España. FI: Finlandia. GR: Grecia. LA: Letonia. LU: Luxemburgo. WA: País de Gales. NE: Países Bajos. PO: Portugal. UK: Reino Unido. SW: Suecia

El principal beneficio para las bibliotecas nacionales tiene que ver con las posibilidades de colaboración con otras fuentes de información (dBpedia, motores de búsqueda, Europea, otros catálogos externos, etc.). Esta colaboración situará en un lugar destacado a todos estos registros alojados en los catálogos bibliotecarios y expandirá la presencia de estas instituciones en la web (fundamentalmente en los resultados de los motores de búsqueda). Destaca también el soporte al multilingüismo que proporciona esta tecnología, especialmente interesante en países con varias oficiales (España y Bélgica, por ejemplo). Por otro lado, la posibilidad de aprender a usar las tecnologías semánticas y asumir la responsabilidad de gestionar cada institución sus conjuntos de datos abiertos enlazados se percibe todavía más un reto (incluso un problema) a asumir por parte de las bibliotecas que un beneficio.

Tabla 2: Resumen de los problemas detectados en bibliotecas nacionales europeas en el desarrollo de proyectos de datos abiertos enlazados. Fuente: Cagnazzo (2017) con elaboración propia.

Problemas detectados	GE	SP	LC	FR	GR	LA	LU	WA	NE	PO	UK	SW
Poca reutilización de información y escasos usuarios	X	X						X	X		X	
Ausencia de estrategia LLD		X	X								X	
Poca disposición de la tecnología y escasez de RRHH		X	X	X			X	X	X	X	X	
Falta disposición a compartir datos						X						
Dispersión de vocabularios						X						
Pruebas que justifiquen la inversión							X					

Leyendas: GE: Alemania. SP: España. FI: Finlandia. GR: Grecia. LA: Letonia. LU: Luxemburgo. WA: País de Gales. NE: Países Bajos. PO: Portugal. UK: Reino Unido. SW: Suecia

Los problemas detectados en la implementación de esta tecnología en las bibliotecas nacionales estaban previstos al principio y no parecen haber sido subsanados con el paso del tiempo. El bajo nivel de destreza en

su uso por parte de los profesionales era, y parece seguir siendo, el principal inconveniente para su desarrollo. Consecuencia de este problema es el pequeño número de usuarios que reutilizan muy escasamente los conjuntos de datos publicados (en Alemania consideran que hasta que el sector industrial no muestre interés no despegará esta tecnología). También es un problema la ausencia de una estrategia clara y definida para la implantación de la tecnología de datos abiertos enlazados (en el Reino Unido, pioneros en su uso, se quejan de haberlo tenido que hacer todo en solitario). Actualizar la formación en las escuelas de información y el reciclaje del personal bibliotecario se hace ahora más importante que nunca.

Otro aspecto a considerar es la evolución de estos proyectos a lo largo del tiempo. Smith-Yoshimura (2016b) estudió la implantación de proyectos de datos enlazados en el ámbito de las bibliotecas. Dicho trabajo destaca los casos de los portales id.loc.gov (Biblioteca del Congreso), data.bnf.fr (Biblioteca Nacional de Francia) y DNB Linked Data Service (Biblioteca Nacional Alemana). Posteriormente, la misma autora (2018) realizó un trabajo similar más actualizado y exhaustivo. Para los profesionales que desarrollaron dichos proyectos se ha producido un ajuste de las expectativas, la reutilización e integración de estándares ya existentes para representar datos bibliotecarios y una visión más holística en relación a la corrección y mapeado de metadatos mediante la explotación y procesamiento de otros conjuntos de datos.

Casos de implementación de datos abiertos enlazados en bibliotecas españolas

En España, la Lista de Encabezamientos para Bibliotecas Públicas sigue siendo una herramienta valiosa y por esa razón se apostó en el año 2011 por construir una lista de materias adaptada a los principios de los datos abiertos vinculados y en formato SKOS, actualizándola y ampliándola con encabezamientos de otros registros de autoridad y vinculándola con otras listas internacionales, además de traducir esos encabezamientos a todos los idiomas oficiales (Spain. *Ministerio...* 2016). Los encabezamientos pueden ser recuperados mediante búsquedas (simples, avanzadas y un punto de consulta SPARQL). Además, se pueden descargar ficheros con las distintas listas completas, tanto en el formato MARC21 como en SKOS (RDF/XML).

Vinculado a nuestra área de conocimiento, pero no dentro de las tareas principales desarrolladas en bibliotecas, Lauscher et al. (2018), proponen utilizar la tecnología de datos enlazados para crear un índice de citas denominado 'Linked Open Citation Database' (LOC-DB), en el cual los cómputos de citas resultantes estén disponibles como datos vinculados y abiertos, utilizando para ello el modelo de datos 'Open Citations' (Peroni & Shotton, 2016) fomentando la colaboración y el intercambio. Este modelo afecta al flujo de trabajo de catalogación en una biblioteca, distribuye la carga de trabajo entre varias instituciones e integra componentes como el extractor de referencias automáticas y diversas fuentes de datos para la sugerencia del enlace. Todos estos componentes han de optimizarse para que la conservación de los datos de citas sea factible, siendo recomendable dedicar tiempo a curar las citas y la calidad de los datos resultantes que se pueden lograr. Se considera que los recursos necesarios para abordar este proyecto son fácilmente asumibles por las bibliotecas de la red, resultando mucho más económicas que las suscripciones a los tradicionales índices de citas, si bien en términos cuantitativos podría ser cierto, no está tan clara esa afirmación en términos cualitativos.

En el entorno de las bibliotecas universitarias, y a partir de los proyectos citados anteriormente, Senso y Arroyo Machado (2018) realizaron un experimento con un conjunto de registros extraídos de la biblioteca de su universidad con el fin de estudiar, de primera mano, algunos de los problemas típicos de los proyectos de datos abiertos enlazados en el entorno bibliotecario. Propusieron un modelo para llevar a cabo la transición en seis etapas: (1) identificación y descripción de los datos; (2) su almacenamiento y corrección; (3) desarrollo de un vocabulario para describirlos según el estándar RDF; (4) generación de los recursos en ese mismo formato; (5) conectar el conjunto de datos a otros que permitieran su enriquecimiento y (6) publicación del conjunto de datos. Si bien solo se ha trabajado con mil registros, los autores están convencidos de que es posible automatizar completamente los procesos de extracción de datos, la limpieza y la generación de grafos RDF y, lo que es más importante desde el punto de vista de la perdurabilidad de una futura implementación, se pueden añadir datos nuevos a los obtenidos en estas transferencias. Ávila Alonso (2014) aplicó los principios de la tecnología de datos abiertos enlazados a la lista de encabezamientos de materia de la Universidad Politécnica de Madrid, migrando hacia un modelo semántico de representación del vocabulario mediante técnicas y tecnologías de datos abiertos enlazados utilizando SKOS.

Gomis Aliaga (2015) detalla la transformación a formato de datos abiertos enlazados de una colección de más 3000 registros de discos de vinilo de la biblioteca de la Universidad de Alicante usando el gestor de contenidos

Ximindex y aplicando la metodología Meloda para reutilizar datos abiertos. En la Universidad Pontificia de Salamanca se ha desarrollado el sistema Openbiblio (Fermoso et al., [2018](#)) basado en el diseño de dos ontologías que posibilitan modelar semánticamente obras y categorías en bibliotecas. Para ello, en primer lugar, se han apoyado en la ontología LD4L ('Linked Data for Libraries') promovida por instituciones de referencia del sector de las bibliotecas universitarias (Stanford, Harvard y Cornell). En segundo término emplean otra ontología de diseño propio para categorizar semánticamente un recurso bibliográfico. El resultado final provee al usuario de un sistema de consulta a los recursos bibliográficos por libro, autor o tema (categoría), con diferentes formatos visuales al mostrar los resultados de las consultas. Los autores consideran que el modelo ontológico y el proyecto es generalizable y reutilizable en cualquier biblioteca. Muchos de estos proyectos apenas han traspasado la fase de prototipado, especialmente si se comparan con los desarrollos que hemos citado anteriormente a nivel internacional.

Más calidad en los conjuntos de datos abiertos enlazados en bibliotecas

Es evidente que la calidad de los metadatos es determinante para un funcionamiento eficiente de un sistema de información digital. De hecho, podría afirmarse que este criterio es la base del éxito de una biblioteca digital. En un contexto donde las bibliotecas caminan hacia su integración en el ecosistema de los datos enlazados es preciso prestar especial atención a la calidad de los conjuntos de datos. Gonçalves et al. ([2007](#)) identifican una serie de conceptos de alto nivel en las bibliotecas digitales: Objetos Digitales, Especificaciones de Metadatos, Colecciones, Catálogos, Repositorios y Servicios. Los objetos digitales forman parte de colecciones cuyas representaciones de metadatos se almacenan en catálogos. Los repositorios permiten la búsqueda, recuperación y reutilización de los objetos digitales mediante servicios que explotan los metadatos disponibles en los catálogos. Los autores identifican una serie de dimensiones de calidad, consistencia e integridad, aplicadas a los conceptos de colección, catálogo y repositorio; las de integridad, exactitud y conformidad se asocian al concepto de especificaciones de metadatos.

Tani, Candela y Castelli ([2013](#)) intentan delimitar que es la calidad de los metadatos incluyendo un interesante análisis de siete modelos de evaluación. Los modelos analizados son muy heterogéneos, incluyendo diversos parámetros y métricas de análisis de calidad. Sigue siendo necesario disponer de un análisis más detenido que defina las correspondencias entre estos modelos de evaluación, algo apuntado por Park ([2009](#)), quien también señalaba como aspectos claves los criterios de completitud, exactitud y consistencia.

Debattista et al. ([2018](#)) tratan la calidad de los conjuntos de datos abiertos enlazados en bibliotecas digitales desde el punto de vista de la experiencia de los profesionales de la información. A través de una encuesta se identificaron los siguientes problemas o cuestiones de calidad:

- **Precisión semántica:** Existen conjuntos de datos que contienen valores incorrectos (uso de identificadores de los propios editores en vez del ISBN, datos extraídos mediante reconocimiento óptico de caracteres, etc.). Una posible solución sería la implantación de políticas del ciclo de vida de los conjuntos de datos que revisen la validez de su contenido.
- **Completitud (cobertura de datos):** La confianza que tienen los consumidores de un conjunto de datos en relación a la completitud de la información para la reutilización o la consistencia del conjunto de datos.
- **Interoperabilidad:** Fomento de la reutilización de los vocabularios de esquemas de metadatos, ontologías y conjuntos de datos de vocabularios controlados ya existentes. La falta de consenso sobre las alternativas más adecuadas, el rápido desarrollo de nuevas versiones que no contemplen la retrocompatibilidad de las ontologías y de los esquemas y el uso de formatos no estándares constituyen importantes obstáculos a superar.
- **Concisión:** La ambigüedad y duplicidad del contenido de los conjuntos de datos conlleva incertidumbre en los elementos a reutilizar.
- **Validez sintáctica:** Es preciso hacer un uso totalmente correcto de los formatos empleados en la representación de los conjuntos de datos para así garantizar su legibilidad por parte de máquinas.
- **Versatilidad del lenguaje:** El idioma en el que se expresa la información de los conjuntos de datos es un factor determinante para su reutilización. Para ello se recomienda el uso del multietiquetado en diferentes idiomas mediante estándares utilizados en la web.
- **Disponibilidad:** Los datos enlazados están descentralizados e interconectados mediante identificadores universales de recurso. Es esencial que estos identificadores perduren a lo largo del tiempo, evitando enlaces rotos que impidan el acceso a los datos.

- **Confianza:** La veracidad de los datos en función de su exactitud, imparcialidad y objetividad en la elaboración del conjunto de datos determina su confianza con vistas a su reutilización.
- **Interpretación:** El uso correcto y actualizado de vocabularios de valores de datos, ontologías y esquemas de metadatos estándares garantiza que los datos sean consistentes y sea factible su legibilidad por máquinas.
- **Licencia:** Los datos deben tener una descripción mediante metadatos de la licencia que informe de forma explícita a los consumidores sobre su reutilización.
- **Carácter imperecedero de los datos:** La actualización de los datos es un factor que influye directamente en la relevancia de las fuentes de datos.
- **Procedencia:** Se ofrece información para comprender el origen de los datos, quién los creó y cómo. Los editores deben proporcionar información de contacto y de procedencia de la fuente original de los datos. Esto influye en la confianza con vistas a su reutilización.
- **Interconexión:** Los enlaces entre recursos permiten la ejecución de procesos de descubrimiento de datos. Es uno de los requisitos de los datos vinculados enunciados por Berners-Lee (2006).
- **Documentación:** La reutilización de datos necesita entender cómo acceder y utilizar los mismos. Los editores de datos deberían publicar dicha documentación en forma de metadatos descriptivos del conjunto de datos. Dicha documentación hace que el conjunto de datos sea más comprensible y esto podría resultar en una mayor reutilización.

Los datos abiertos bibliotecarios se incardinan directamente con la web de datos. Por ello es preciso vincular los diferentes aspectos de calidad con las buenas prácticas de publicación en la web de datos que han sido elaboradas por el consorcio W3C (2017). Estas buenas prácticas responden a una serie de retos que los editores deben afrontar para la publicación de datos en la web:

1. El modo en el que se deben proporcionar los metadatos, tanto a las personas como a las máquinas.
2. Las licencias de los datos: permisos y restricciones para reutilizar los datos.
3. La procedencia y calidad establecen la forma de añadir confianza a los datos.
4. El correcto versionado de los datos permite realizar un seguimiento del histórico de versiones.
5. La identificación de datos permite distinguir diferentes conjuntos de datos y sus distribuciones individuales.
6. La definición de los formatos a utilizar para representar los conjuntos de datos.
7. Los vocabularios de datos permiten mejorar la interoperabilidad de los datos.
8. Hay que habilitar las herramientas que permitan el acceso a los datos.
9. Se deben establecer los métodos para el archivo y conservación de los datos.
10. La retroalimentación: permite que los usuarios contacten con los editores de los datos para corregir o completar los conjuntos de datos.
11. Es necesario definir cómo se puede añadir valor a los datos.
12. Hay que elaborar normas para la reutilización de datos, su transformación y redistribución.

Como puede verse, para publicar conjuntos de datos abiertos enlazados en la web, no es suficiente con disponer de tecnologías y recursos adecuados desde el punto de vista técnico. Es preciso implantar una verdadera cultura alrededor de todos estos procesos y aplicar una necesaria serie de pautas que sirvan de guía en este tipo de proyectos. La formación del personal bibliotecario que debe acometerlos no debe centrarse solo en las competencias del saber, hay que ampliarlas claramente al ámbito del saber hacer, deben formar parte del genoma vital de su tarea diaria, tal como ahora la forman los procesos de referencia o de formación de usuarios.

Conclusiones

Si bien existen desarrollos prometedores que vendrán seguidos de muchos más, la transición a las tecnologías de datos abiertos enlazados no ha sido tan rápida como se preveía. Se ha repetido varias veces que buena parte del problema tiene que ver con la curva de aprendizaje en el uso de estas tecnologías por parte de los bibliotecarios, lo que provoca cierta actitud reactiva por parte de este colectivo. También se han detectado problemas relacionados con los derechos de autor que, de momento, parecen destinar los proyectos de datos abiertos enlazados a conjuntos de datos con licencias creative commons. A la hora de consumir estos conjuntos de datos, los principales problemas son la coincidencia, desambiguación y alineación entre los datos de origen y los recursos de datos vinculados, junto con el mapeo del vocabulario y la reutilización de todo el conjunto publicado.

La dispersión de vocabularios RDF, catálogos de conjuntos de datos, etc., complica la adopción de criterios comunes. Las bibliotecas nacionales de España, Reino Unido y Francia han desarrollado sus propias ontologías y vocabularios para publicar sus conjuntos de datos abiertos enlazados que deben dar soporte al desarrollo de nuevos proyectos. No obstante, sigue siendo complicado localizar datos del ámbito bibliotecario en los principales catálogos de conjuntos de datos. Los vocabularios SKOS están consolidados, se ve la utilidad de su publicación como datos abiertos enlazados y se percibe claramente su potencial en la reutilización de datos. Sin embargo, no sucede lo mismo con los ficheros de autoridades y los catálogos, donde son las bibliotecas nacionales las que toman la iniciativa, ya sea por su capacidad en cuanto a recursos o por el desinterés de otras instituciones (especialmente fuera de Estados Unidos). En realidad se ha trabajado e investigado poco sobre las posibilidades reales o las necesidades de reutilización de los conjuntos de datos abiertos enlazados. Posiblemente, esta carencia de estudios influya en las decisiones a tomar por parte a nivel de dirección de las bibliotecas. Esto puede influir en que se valoren las posibilidades de esta tecnología (no porque no sea interesante, sino porque no se ha estudiado a fondo).

Los conjuntos de datos abiertos enlazados en entornos bibliotecarios deben estar sujetos a la aplicación de las buenas prácticas de publicación de datos en la web. Por otro lado, los diferentes modelos de evaluación de la calidad de los metadatos en este contexto deben armonizarse siguiendo la recomendación del consorcio W3C (2017). Aunque no es objeto del presente trabajo abordar de forma exhaustiva estos aspectos de calidad, sí podemos apuntar la necesidad de un mapeado y adaptación de los diferentes modelos existentes en el marco de la definición de los requisitos y buenas prácticas de publicación de datos en la web.

A pesar de todos estos condicionantes, que sin duda alguna han lastrado la evolución hacia el uso extendido de esta tecnología, nadie parece dudar de que es el camino a seguir. Los desarrollos de los sistemas de gestión de biblioteca parecen avalar esta tesis, puesto que entre sus funcionalidades se incorporan aquellas que están relacionadas con la publicación y acceso a los registros bibliográficos como datos abiertos enlazados (especialmente, y como han destacado autores y profesionales, desde los motores de búsqueda de la web). Sin duda alguna, la consolidación del *library requirements model*, el estándar RDA y BIBFRAME propiciará que sus desarrolladores encuentren una base estandarizada para la representación de los objetos bibliográficos. En este punto destacan las ventajas que aportaría la consulta de catálogos bibliotecarios cuyos ficheros locales de autoridades estén vinculados con otros de gran relevancia como VIAF, ULAN, o LCNAF. Tampoco hay que olvidar las expectativas creadas alrededor de la integración de los catálogos bibliotecarios en el ecosistema de la web de datos, especialmente con iniciativas como Wikidata o con contenidos audiovisuales disponibles en la web. En definitiva, la publicación de los catálogos como conjuntos de datos RDF no se limitaría a la simple disponibilidad de los mismos en la web de datos, sino que plantearía nuevas aplicaciones y retos para estos instrumentos que irían más allá de sus usos actuales en el entorno de las bibliotecas.

About the authors

Francisco Javier Martínez Méndez es profesor del Departamento de Información y Documentación en la Universidad de Murcia. Su email de contacto es javima@um.es. ORCID No. orcid.org/0000-0003-1098-9361

Juan-Antonio Pastor-Sánchez es profesor del Departamento de Información y Documentación de la Universidad de Murcia. Su email de contacto es pastor@um.es, ORCID No. orcid.org/0000-0002-1677-1059

Rosana López-Carreño es profesora del Departamento de Información y Documentación de la Universidad de Murcia. Su email de contacto es rosanalc@um.es. ORCID No. orcid.org/0000-0002-2097-9389

Abstract in English

Introduction. 'Linked data' is part of the Semantic Web project. Library linked data are datasets related to these institutions. From the beginning of the launch of this idea it was thought that libraries and archives would be the first organizations to use this technology massively.

Method. An exhaustive specialized literature review has been carried out to define the state of the matter after several years of development of this technology.

Analysis. Authors have worked primarily with case studies, obtaining those most relevant reflections and conclusions of the participants in these projects.

Results. The initial premise has not been verified in a general way. The development of this technology is still partial and has found some rejection by some of the professionals due to its complicated learning curve and the dispersion of the developed vocabularies. In addition to having economic, technical and human resources, a clear commitment is needed for the development of

this technology.

Conclusion. Despite the problems, it seems clear that 'linked data' defines the conceptual and technological bases for incorporating the library datasets within the data web environment, being necessary to overcome the reluctance of the professionals and the problems with the licences for the exploitation of data.

References

Note: A link from the title is to an open access document. A link from the DOI is to the publisher's page for the document.

- Agenjo Bullón, X. & Hernández Carrascal, F. (2017). [El estado de los datos vinculados en bibliotecas en 2015](https://doi.org/10.3145/thinkepi.2017.60). *Anuario ThinkEPI*, (1), 321-323. <https://doi.org/10.3145/thinkepi.2017.60>
- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Linked data for libraries. *New Library World*, 113(11/12), 549–570. <https://doi.org/10.1108/03074801211282920>
- Alemu, G., & Stevens, B. (2015). *An emergent theory of Digital Library Metadata: enrich then filter*. Chandos Publishing.
- Ávila Alonso, R. (2014). [Aplicación de los principios linked open data a la lista de encabezamientos de materia de la biblioteca de la Universidad Politécnica de Madrid](http://hdl.handle.net/10016/19674). [Unpublished master's dissertation]. Universidad Carlos III de Madrid. <http://hdl.handle.net/10016/19674>
- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, M., Schneider, J., Singer, R., Summers, E., Waites, W., Young, J., & Zeng, M. (2011). [Library Linked Data Incubator Group final report. 25 October 2011](https://www.w3.org/2005/Incubator/lld/XGR-ldd-20111025). W3C-the World Wide Web Consortium. <https://www.w3.org/2005/Incubator/lld/XGR-ldd-20111025> (Archived by the Internet Archive at <https://bit.ly/30UTdyJ>)
- Bauer, F., & Kaltenböck, M. (2012). [Linked open data: the essentials. A quick start guide for decision makers](https://www.scribd.com/document/77911038/Linked-Open-Data-The-Essentials-A-Quick-Start-Guide-for-Decision-Makers). Edition mono/monochrom. <https://www.scribd.com/document/77911038/Linked-Open-Data-The-Essentials-A-Quick-Start-Guide-for-Decision-Makers> (Archived by the Internet Archive at <https://bit.ly/3hDZWtI>)
- Berners-Lee, T., & Hendler, J. (2001). Publishing on the semantic web. *Nature*, 410(6832), 1023-1024. <http://dx.doi.org/doi:10.1038/35074206>
- Berners-Lee, T. (2006). [Linked data](https://www.w3.org/DesignIssues/LinkedData.html). W3C. <https://www.w3.org/DesignIssues/LinkedData.html> (Archived by the Internet Archive at <https://bit.ly/37HvNOY>)
- Berners-Lee, T. (2007). [Testimony of Sir Timothy Berners-Lee before the United States House of Representatives Committee on Energy and Commerce Subcommittee on telecommunications and the internet](http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html). CSAIL Decentralized Information Group. <http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html> (Archived by the Internet Archive at <https://bit.ly/30Tplms>)
- Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. (2008). [Linked data on the Web \(LDOW2008\)](https://dl.acm.org/doi/pdf/10.1145/1367497.1367760). In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, Beijing, China, April, 2008 (pp. 1265-1266). ACM. <https://dl.acm.org/doi/pdf/10.1145/1367497.1367760>.
- Breeding, M. (2015). [The future of library resource discovery. A white paper commissioned by the NISO discovery to delivery \(D2D\) Topic Committee](https://groups.niso.org/apps/group_public/download.php/15077/future_library_resource_discovery.pdf). NISO. https://groups.niso.org/apps/group_public/download.php/15077/future_library_resource_discovery.pdf
- Cagnazzo, L. (2017). [Linked data: Implementation, use, and perceptions across European national libraries](http://hdl.handle.net/10760/32004) [Unpublished doctoral dissertation] University of Strathclyde. <http://hdl.handle.net/10760/32004>
- Cormenzana López, R. (2017). [RDA y BIBFRAME: Hacia un nuevo paradigma en catalogación](http://hdl.handle.net/10609/84868). [Unpublished bachelor's dissertation]. Universitat Oberta de Catalunya. <http://hdl.handle.net/10609/84868>
- Cormenzana López, R., & López-Borrull, A. (2018). [Estudio de la adaptación a RDA y BIBFRAME en el ámbito de las bibliotecas españolas](https://doi.org/10.6018/analesdoc.21.2.323441). *Anales de Documentación*, 21(2). <https://doi.org/10.6018/analesdoc.21.2.323441>
<https://revistas.um.es/analesdoc/article/view/323441/250761> (Archived by the Internet Archive at <https://bit.ly/3fEtY87>)
- Debattista, J., McKenna, L., & Brennan, R. (2018). [Understanding information professionals. A survey on the quality of Linked Data Sources for digital libraries](https://doi.org/10.1007/978-1-4939-9888-8_10). In H. Panetto, C. Debruyne, H. Proper, C. Ardagna, D. Roman & R. Meersman, (Eds). *On the Move to Meaningful Internet Systems. OTM 2018*

- Conferences. OTM 2018* (pp. 537-545). Springer. (Lecture Notes in Computer Science, vol 11230). <http://doras.dcu.ie/22979/1/ODBASE2018Library.pdf>
- Fallgren, N., Lauruhn, M., Romano Reynolds, R., & Kaplan, L. (2014). [The missing link: the evolving current state of linked data for serials](#). *The Serials Librarian*, 66(1-4), 123-138. <https://doi.org/10.1080/0361526X.2014.879690>
 - Feroso-García, A. M., Manzano-García, M. I., Mateos-Sánchez, M., & Hernández-Tamayo, C. (2018). [Sistema de modelado semántico para catalogación, clasificación, consulta y publicación en abierto de información bibliográfica](#). *El Profesional de la Información*, 27(2), 410-418. <https://doi.org/10.3145/epi.2018.mar.20>
 - Godby, C. J., & Denenberg, R. (2015). [Common ground: exploring compatibilities between the linked data models of the Library of Congress and OCLC](#). Library of Congress and OCLC Research. <https://files.eric.ed.gov/fulltext/ED564824.pdf>
 - Gonçalves, M.A., Moreira, B.L., Fox, E.A., & Watson, L.T. (2007). "What is a good digital library?" - a quality model for digital libraries. *Information Processing & Management* 43(5), 1416–1437. <https://doi.org/10.1016/j.ipm.2006.11.010>
 - Gomis Aliaga, S. (2015). [Publicar datos abiertos enlazados \(LOD\) en entornos bibliotecarios mediante el gestor de contenido semántico Ximdex](#). *Cuadernos de Gestión de Información*, 5(1), 1-17. <https://revistas.um.es/gesinfo/article/view/232181> (Archived by the Internet Archive at <https://bit.ly/3hHYze>)
 - Hallo, M., Luján-Mora, S., Maté, A., & Trujillo, J. (2016). Current state of linked data in digital libraries. *Journal of Information Science*, 42(2), 117-127. <https://doi.org/10.1177/0165551515594729>
 - Heath, T., & Bizer, C. (2011). *Linked data: evolving the web into a global data space*. Morgan & Claypool. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>
 - Jin, Q., Hahn, J. F., & Croll, G. (2016). [BIBFRAME transformation for enhanced discovery](#). *Library Resources and Technical Services*, 60(4), 223-235. <https://journals.ala.org/index.php/lrts/article/view/6134/7929> (Archived by the Internet Archive at <https://bit.ly/2CnyOS>)
 - Latif, A., Scherp, A., & Tochtermann, K. (2016). LOD for library science: benefits of applying linked open data in the digital library setting. *Kunstliche Intelligenz*, 30(2), 149–157. <https://doi.org/10.1007/s13218-015-0420-x>
 - Lauscher, A., Eckert, K., Galke, L., Scherp, A., Rizvi, S. T. R., Ahmed, S., & Klein, A. (2018). [Linked open citation database: enabling libraries to contribute to an open and interconnected citation graph](#). In *JCDL '18: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. Fort Worth, Texas, USA, June, 2018. (pp. 109–118). Association for Computing Machinery. <https://ub-madoc.bib.uni-mannheim.de/44470/1/jcdl089-lauscherA-1.pdf> (Archived by the Internet Archive at <https://bit.ly/2YcC6qj>)
 - Library of Congress. (2012). [Bibliographic framework as a web of data: linked data model and supporting services](#). Library of Congress. <https://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf> (Archived by the Internet Archive at <https://bit.ly/2N8pUAP>)
 - Lóscio, B.F., Burle, C., & Calegari, N. (Eds.). (2017). [Data on the Web best practices. W3C Recommendation. 31 January 2017](#). W3C-World Wide Web Consortium. <https://www.w3.org/TR/2017/REC-dwbp-20170131> (Archived by the Internet Archive at <https://bit.ly/2zHpb6i>)
 - McCallum, S. (2017). [BIBFRAME development](#). *JLIS.it, Italian Journal of Library, Archives, and Information Science*, 8(3), 71-85. <http://doi.org/10.4403/jlis.it-12415>
 - Morville, P. (2005). *Ambient findability*. O'Reilly.
 - O'Dell, A. J. (2014, July 11). [On BIBFRAME interfaces: the application of linked data in libraries](#). *Medium* <https://medium.com/@allisonjaiodell/on-bibframe-interfaces-44fe5e0d4f57> (Archived by the Internet Archive at <https://bit.ly/37IZi2s>)
 - Park, J. R. (2009). Metadata quality in digital repositories: a survey of the current state of the art. *Cataloging & Classification Quarterly*, 47(3-4), 213-228. <https://doi.org/10.1080/01639370902737240>
 - Peset Mancebo, M. F., Ferrer Sapena, A., & Subirats-Coll, I. (2011). [Open data y linked open data: su impacto en el área de bibliotecas y documentación](#). *El Profesional de la Información*, 20(2), 165-173. <https://doi.org/10.3145/epi.2011.mar.06> <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2011.mar.06> (Archived by the Internet Archive at <https://bit.ly/2YHBVm0>)
 - Peroni, S., & Shotton; D. (2016). [The OpenCitations data model](#). Figshare. <https://doi.org/10.6084/m9.figshare.3443876>
 - Putkey, T. (2011). [Using SKOS to express faceted classification on the semantic web](#). *Library Philosophy and Practice (e-journal)*, paper 620. <https://digitalcommons.unl.edu/libphilprac/620> (Archived by the

- Internet Archive at <https://bit.ly/2BhQ55e>)
- Santos, R., Manchado, A., & Vila-Suero, D. (2015). [Datos.bne.es: a LOD service and a FRBR-modelled access into the library collections](#). In *Proceedings, IFLA World Library and Information Congress, 81st IFLA General Conference and Assembly*, 15-21 August 2015, Cape Town, South Africa. (18 p.) IFLA. <http://library.ifla.org/1085/1/207-santos-en.pdf> (Archived by the Internet Archive at <https://bit.ly/3hCecfP>)
 - Saorín, T., Peset, F. & Ferrer-Sapena, A. (2013). [Factores para la adopción de linked data e implantación de la web semántica en bibliotecas, archivos y museos](#). [Factors for the adoption of linked data and implementation of the semantic web in libraries, archives and museums]. *Information Research*, 18(1) paper 570. <http://InformationR.net/ir/18-1/paper570.html> (Archived by the Internet Archive at <https://bit.ly/3hIpMpu>)
 - Senso, J.A., & Arroyo Machado, W. (2018). [La publicación en linked data de registros bibliográficos: modelo e implementación](#). [The publication in linked data of bibliographic records: model and implementation]. *Revista española de Documentación Científica*, 41(4), 217. <https://doi.org/10.3989/redc.2018.4.1535> <http://redc.revistas.csic.es/index.php/redc/article/view/1023/1636> (Archived by the Internet Archive at <https://bit.ly/37Gr2F7>)
 - Smith-Yoshimura, K. (2016). [Linked data implementations - who, what and why?](#) [PowerPoint slides]. OCLC Research. <https://bit.ly/2Bi9piM>
 - Smith-Yoshimura, K. (2016b). [Analysis of international linked data survey for implementers](#). *D-Lib Magazine*, 22(7/8). <https://doi.org/10.1045/july2016-smith-yoshimura>
 - Smith-Yoshimura, K. (2018). [Analysis of 2018 international linked data survey for implementers](#). *code4lib Journal* (No. 42). <https://journal.code4lib.org/articles/13867> (Archived by the Internet Archive at <https://bit.ly/2UXEkb2>)
 - Spain. *Ministerio de Educación, Cultura y Deporte* (2016). [Lista de encabezamientos de materia para las bibliotecas públicas en SKOS](#). [List of subject headings for public libraries in SKOS]. Ministerio de Educación, Cultura y Deporte. <http://id.sgcb.mcu.es/>
 - Tallerås, K. (2013). [From many records to one graph: heterogeneity conflicts in the linked data restructuring cycle](#). *Information Research*, 18(3) paper CoLIS18. <http://InformationR.net/ir/18-3/colis/paperC18.html> (Archived by the Internet Archive at <https://bit.ly/3fCpLx>)
 - Tani, A., Candela, L., & Castelli, D. (2013). Dealing with metadata quality: the legacy of digital library efforts. *Information Processing & Management*, 49(6), 1194 – 1205. <https://doi.org/10.1016/j.ipm.2013.05.003>
 - Tharani, K. (2015). [Linked data in libraries: a case study of harvesting and sharing bibliographic metadata with BibFrame](#). *Information and Library Technologies*, 34(1), 6-19. <https://doi.org/10.6017/ital.v34i1.5664>
 - Vila-Suero, D., & Gómez-Pérez, A. (2013) datos.bne.es and MARiMbA: an insight into library linked data. *Library Hi Tech*, 31(4), 575-601. <https://doi.org/10.1108/LHT-03-2013-0031>
 - Wang, Y., & Yang, S. Q. (2018). [Linked data technologies and what libraries have accomplished so far](#). *International Journal of Librarianship*, 3(1), 3-20. <http://doi.org/10.23974/ijol.2018.vol3.1.62>
 - Wilson, N. (2010). [Linked data prototyping at the British Library](#). [PowerPoint slides]. Research Information Network. <https://bit.ly/2YbA21Y> (Archived by the Internet Archive at <https://bit.ly/2YgpbE2>)
 - Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2013, November). Highlights of library data models in the era of linked open data. In E. Garoufallou & J. Greenberg, (Eds.). *Metadata and Semantics Research, Proceedings 7th Research Conference, MTSR 2013*, Thessaloniki, Greece, November 19-22, 2013. (pp. 396-407). Springer. https://doi.org/10.1007/978-3-319-03437-9_38
 - Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017). Representing and integrating bibliographic information into the semantic Web. A comparison of four conceptual models. *Journal of Information Science*, 43(4), 525-553. <https://doi.org/10.1177/0165551516650410>

How to cite this paper

Martínez Méndez, F.J., Pastor-Sánchez, J-A., & López Carreño, R. (2019). Linked open data en bibliotecas: estado del arte. *Information Research*, 25(2), paper 862. Retrieved from <http://InformationR.net/ir/25-2/paper862.html> (Archived by the Internet Archive at <https://bit.ly/3fglZ0K>)

Find other papers on this subject

[Scholar Search](#)

[Google Search](#)

[Bing](#)

Check for citations, [using Google Scholar](#)

© the authors, 2020.

1809 Last updated: 14 June, 2020

- [Contents](#) |
 - [Author index](#) |
 - [Subject index](#) |
 - [Search](#) |
 - [Home](#)
-