

La transición a la siguiente generación de metadatos

Karen Smith-Yoshimura

Miembro directivo de Programas

OCLC

Traducción

Iván Pérez Marinas

Departamento de Proceso Técnico

Biblioteca Nacional de España

Ésta es la expresión española de la obra. Para consultar la expresión original acceda a este [enlace](#).

ÍNDICE

Resumen ejecutivo	4
Introducción.....	6
La transición a datos enlazados e identificadores.....	11
La ampliación del uso de identificadores permanentes	12
El paso del “control de autoridades” a la “gestión de identidades”	20
La respuesta a la necesidad de vocabularios múltiples y a la equidad, diversidad e inclusión	23
Los desafíos de los datos enlazados	30
La descripción de colecciones “de dentro hacia afuera” y “facilitadas”	33
Colecciones archivísticas.....	34
Sitios web archivados	35
Colecciones audiovisuales	37
Colecciones de imágenes.....	40
Datos de investigación.....	44
La evolución de “los metadatos como servicio”	50
Estadísticas	50
Asesoramiento	51
Nuevas aplicaciones	51
Bibliometría	53
Indización semántica	54
Los preparativos para los requisitos del personal del futuro	57
El cambio cultural.....	57
Oportunidades de aprendizaje	58
Nuevas herramientas y habilidades	59
Autoformación.....	61
El tratamiento de la rotación de personal	62
Consecuencias	65
Agradecimientos	66
Apéndice	67
Notas	68

IMÁGENES

IMAGEN 1	“Los cambios de las dinámicas de trabajo en la descripción de recursos” por OCLC Research.....	11
IMAGEN 2	Algunos de los 300 nombres de autor abreviados para un artículo de cinco páginas de <i>Physical Review Letters</i>	16
IMAGEN 3	Ejemplos de algunos identificadores DOI y ARK.....	19
IMAGEN 4	Un identificador de Wikidata enlaza a otros identificadores y etiquetas en diferentes idiomas.....	21
IMAGEN 5	Extracto del informe de resultados de la encuesta EDI de 2017 del Research Library Partnership.....	27
IMAGEN 6	Respuestas de la encuesta de 2019 sobre desafíos relativos a la gestión de colecciones audiovisuales.....	40
IMAGEN 7	El IIF Explorer de OCLC ResearchWorks recupera imágenes de “mapas de París” en colecciones de CONTENTdm.....	44
IMAGEN 8	Distribución de los 465 códigos de lenguas indígenas de la Australian National Bibliographic Database procedentes del Austlang National Codeathon.....	53
IMAGEN 9	“River of Authors (“Río de autores”) de Hachette UK generado a partir de los metadatos del catálogo de la British Library	54

RESUMEN EJECUTIVO

El OCLC Research Library Partners Metadata Managers Focus Group (Grupo de Debate de Gestores de Metadatos de los Socios de Bibliotecas de Investigación), creado en 1993, es uno de los grupos que más han perdurado en el OCLC Research Library Partnership (RLP; Sociedad de Bibliotecas de Investigación), una red internacional de bibliotecas de investigación. El Focus Group aporta un foro para que directivos que se ocupan de crear y gestionar metadatos en sus instituciones compartan información sobre asuntos de interés común e identifiquen problemas de gestión de metadatos.

El presente informe, *La transición a la siguiente generación de metadatos*, sintetiza seis años (2015-2020) de debates del OCLC Research Library Partners Metadata Managers Focus Group y lo que se podría pronosticar para la “siguiente generación de metadatos”. Todos los debates del Focus Group están impregnados de la firme creencia de que los metadatos subyacen en todas las opciones de búsqueda y descubrimiento independientemente del formato, ahora y en el futuro.

Sin embargo, los metadatos están cambiando. Tanto por motivos conceptuales como por cuestiones técnicas, pronto va a quedar obsoleta la gestión de metadatos de formato específico que se basan en cadenas de caracteres suministradas en registros bibliográficos que sólo son entendidas por sistemas bibliotecarios. Las innovaciones en biblioteconomía están presionando en las prácticas habituales de la gestión de metadatos para que éstas evolucionen, ya que a los bibliotecarios se les está exigiendo que proporcionen metadatos para muchos más recursos de varios tipos y que colaboren en proyectos institucionales o multiinstitucionales con menos personal. El presente informe indica cómo están evolucionando los metadatos y examina el impacto que esta transición puede tener en servicios bibliotecarios, planteando preguntas como:

- **¿Por qué cambian los metadatos?**
- **¿Cómo cambia el proceso de creación?**
- **¿Cómo cambian los metadatos en sí?**
- **¿Qué impacto tendrán estos cambios en los futuros requisitos de personal y cómo pueden prepararse las bibliotecas?**

El futuro de los datos enlazados está ligado al futuro de los metadatos; los metadatos que las bibliotecas, los archivos y otras instituciones de patrimonio cultural han creado y crearán proporcionarán el contexto para futuras innovaciones de datos enlazados como “menciones” asociadas con esos enlaces. Las consecuencias serán globales, afectando a cómo los bibliotecarios y archiveros describirán las colecciones de dentro hacia afuera y facilitadas, inspirando nuevas ofertas de “metadatos como servicio” e influyendo en los futuros requisitos del personal.

La transición a la siguiente generación de metadatos es un proceso evolutivo, entrelazado con normas, infraestructuras y herramientas cambiantes. Los miembros del Focus Group han llegado a un entendimiento común de los desafíos, han compartido posibles formas de afrontarlos y han transmitido estas ideas a otras comunidades con las que interactúan.

INTRODUCCIÓN

El OCLC Research Library Partners Metadata Managers Focus Group (en adelante mencionado como el Focus Group),¹ creado en 1993, es uno de los grupos que más ha perdurado en la OCLC Research Library Partnership (RLP),² una red internacional de bibliotecas de investigación. El Focus Group aporta un foro para que directivos que se ocupan de crear y gestionar metadatos en sus instituciones compartan información sobre asuntos de interés común e identifiquen problemas de gestión de metadatos. Los problemas planteados por el Focus Group son abordados por OCLC Research en apoyo del RLP e inspiran los productos y servicios de OCLC.

Todos los debates del Focus Group están impregnados de la firme creencia de que los metadatos subyacen en todas las opciones de búsqueda y descubrimiento independientemente del formato, ahora y en el futuro. Los metadatos aportan las infraestructuras de investigación necesarias para los “sistemas de difusión de contenidos” de todas las bibliotecas, cumpliendo las peticiones de información y recursos por parte de su comunidad. Los metadatos son cruciales para la transición a las nuevas generaciones de sistemas bibliotecarios y de búsqueda y descubrimiento. Los buenos metadatos creados actualmente pueden ser reutilizados con facilidad en un entorno de datos enlazados en el futuro.³ Como se apuntó en las *British Library's Foundations for the Future (Bases de la British Library para el futuro)*: “Nuestra visión es que alrededor del 2023 los recursos de metadatos de la colección de la British Library se unificarán en una sola infraestructura sostenible y normalizada que ofrezca opciones mejoradas de acceso, colaboración y reutilización abierta”.⁴

Tanto por motivos conceptuales como por cuestiones técnicas, pronto va a quedar obsoleta la gestión de metadatos de formato específico que se basan en cadenas de caracteres suministradas en registros bibliográficos que sólo son entendidas por sistemas bibliotecarios. Las innovaciones en biblioteconomía están presionando en las prácticas habituales de la gestión de metadatos para que éstas evolucionen, ya que a los bibliotecarios se les está exigiendo que proporcionen metadatos para muchos más recursos de varios tipos y que colaboren en proyectos institucionales o multiinstitucionales con menos personal. “Los métodos tradicionales de generación,

gestión y difusión de metadatos”, indica la Collection Management Strategy (Estrategia de Gestión de Colecciones) de la British Library, “no son extensibles o apropiados para una era de rápidos cambios digitales, con un aumento de las expectativas de los usuarios y una disminución de los recursos”.⁵ Los miembros del Focus Group están ansiosos por liberar el poder de los metadatos en los registros antiguos para que se produzcan interacciones y usos diferentes tanto de máquinas como de usuarios en el futuro. Los metadatos homogéneos creados de acuerdo con reglas o normas anteriores deben ser transformados en nuevas estructuras.

¿Por qué cambian los metadatos?

Los metadatos bibliotecarios tradicionales fueron creados y son creados por bibliotecarios de acuerdo a las reglas que principalmente los bibliotecarios utilizan y comprenden. Están centrados en el registro, son caros de producir y tienen históricas limitaciones de tamaño. Los metadatos están limitados en su cobertura; en particular, no incluyen artículos de revistas académicas u otras publicaciones académicas. La infraestructura no ha sido adecuada para la gestión de correcciones y mejoras, induciendo un énfasis en la perfección que ha exacerbado la lentitud de la creación de metadatos. En resumen, los metadatos podrían ser mejores, no hay suficientes y los metadatos que existen no se utilizan ampliamente fuera del ámbito bibliotecario.

¿Cómo cambia el proceso de creación?

Los metadatos ya no son sólo creados por personal bibliotecario. Actualmente editores, autores y otros agentes están igualmente involucradas en la creación de metadatos. La creación de metadatos también se ha adelantado en el transcurso de creación y publicación de documentos del ámbito académico, ya que los editores crean registros de metadatos con anterioridad al proceso de catalogación tradicional. Los metadatos ahora se pueden mejorar o corregir mediante máquinas o mediante *crowdsourcing*.

¿Cómo cambian los metadatos en sí?

Machine-Readable Cataloging (MARC; Catalogación Legible por Máquinas) fue creado para reproducir los metadatos que habitualmente se encontraban en las fichas de los catálogos bibliotecarios. Estamos pasando de los registros MARC a agregaciones de componentes bien codificados, compartibles y enlazables, con énfasis en las referencias,

y estamos eliminando las anacrónicas abreviaturas que las máquinas no comprenden. En lugar de depender únicamente de vocabularios bibliotecarios, como encabezamientos de materia y listas codificadas, las agregaciones en desarrollo pueden incorporar vocabularios creados para ámbitos específicos, aumentando los usuarios potenciales de los metadatos.

La composición del Focus Group ha variado a lo largo del tiempo y en la actualidad cuenta con representantes de 63 miembros de la RLP de 11 países de cuatro continentes.⁶ El grupo incluye responsables, tanto pasados como futuros, del Program for Cooperative Cataloging (PCC; Programa para la Catalogación Cooperativa),⁷ proporcionando un intercambio de ideas entre los dos. Los temas para los debates del grupo pueden ser propuestos por cualquier miembro del Focus Group y son seleccionados por el Planning Group (Grupo de Planificación; ver apéndice), compuesto por ocho miembros y que escribe las “declaraciones de contextualización” para explicar por qué se considera que el tema es oportuno e importante y luego desarrolla la batería de preguntas que profundiza en el tema. Después las declaraciones de contextualización y la batería de preguntas se hacen llegar a todos los miembros del Focus Group, que tienen de tres a cinco semanas para enviar sus respuestas. Las compilaciones de las respuestas del Focus Group inspiran los debates presenciales que se llevan a cabo, junto con las conferencias de la American Library Association⁸ y las reuniones virtuales posteriores.

Como moderadora del Focus Group, he resumido y sintetizado estos debates en una serie de publicaciones del blog *Hanging Together (Permanecemos Juntos)* de OCLC Research.⁹ En este foro he publicado alrededor de unas 40 entradas sobre temas de lo más variado relativos a metadatos a lo largo de los últimos seis años.

El Metadata Managers Focus Group es sólo una actividad dentro de la más amplia OCLC Research Library Partnership, que se dedica a las amplias oportunidades de desarrollo profesional del personal bibliotecario. Los miembros del Focus Group valoran su afiliación a la Research Library Partnership como un canal para llegar a ser los “agentes del cambio” de la gestión de metadatos del futuro.¹⁰ Las respuestas de los miembros del Focus Group a las preguntas han facilitado debates intrainstitucionales y

han ayudado a los gestores de metadatos a comprender cómo se compara la situación de sus instituciones con las de sus compañeros dentro de la Research Library Partnership. Estos debates del Focus Group han identificado una amplia variedad de problemas relativos a metadatos, documentados en el presente informe. La transición a la siguiente generación de metadatos es un proceso evolutivo, entrelazado con normas, infraestructuras y herramientas cambiantes. Los miembros del Focus Group han llegado a un entendimiento común de los desafíos, han compartido posibles formas de afrontarlos y han transmitido estas ideas a otras comunidades con las que interactúan.

De forma colectiva los miembros del Focus Group disponen de una amplia variedad de experiencias con datos enlazados. El gran interés del Focus Group en las implementaciones de datos enlazados desencadenó las sucesivas International Linked Data Surveys for Implementers (Encuestas Internacionales de Datos Enlazados para Implementadores) de OCLC Research.¹¹ Una parte de los miembros del Focus Group ha participado en varios proyectos de datos enlazados, como el Project Passage y el piloto CONTENTdm Linked Data de OCLC Research, la Shared Entity Management Infrastructure (Infraestructura de Gestión de Entidades Compartidas) de OCLC, la Bibliographic Framework Initiative (BIBFRAME; Iniciativa de la Estructura Bibliográfica) de la Library of Congress, el proyecto Linked Data for Production (LD4P, Datos Enlazados para Producción) con la subvención económica Mellon, la iniciativa Share-VDE y las Shareable Local Name Authorities (Autoridades de Nombres Locales Compartibles) con una subvención económica del IMLS, que presentaron problemas planteados por los centros coordinadores de identificadores en el entorno de datos enlazados.¹² Además, los miembros del Focus Group contribuyen a los grupos de trabajo del PCC que abordan aspectos de la labor de datos enlazados, como el PCC Task Group on Linked Data Best Practices (Grupo de Trabajo del PCC sobre las Buenas Prácticas de Datos Enlazados), el Task Group on Identity Management (Grupo de Trabajo sobre la Gestión de Identidades), el Task Group on URIs in MARC (Grupo de Trabajo sobre los URIs en MARC) y el PCC Linked Data Advisory Committee (Consejo Asesor de Datos Enlazados del PCC).¹³ Este intercambio de ideas ha incitado al Focus Group a examinar problemas entorno a las *entidades* representadas en recursos institucionales.

El presente informe sintetiza seis años (2015-2020) de debates del OCLC Research Library Partners Metadata Managers Focus Group y lo que se podría pronosticar para la “siguiente generación de metadatos”. El documento se organiza en las secciones siguientes, ya que cada una representa una tendencia emergente que ha sido identificada en los debates del Focus Group:

- La transición a datos enlazados e identificadores: ampliando el uso de los identificadores permanentes como parte del cambio desde el “control de autoridades” hacia la “gestión de identidades”
- La descripción de las colecciones “de dentro hacia afuera” y las colecciones “facilitadas”: desafíos en la creación y gestión de metadatos para recursos únicos creados o suministrados por instituciones en varios formatos y compartidos con consorcios
- La evolución de “metadatos como servicio”: una mayor participación en la creación de metadatos más allá del tradicional catálogo de biblioteca
- Los preparativos para los requisitos del personal del futuro: el panorama cambiante aboca a la necesidad de nuevas habilidades tanto para los nuevos profesionales que entran a este mundo como para los catalogadores expertos

El documento concluye con algunas observaciones sobre las consecuencias previstas de la siguiente generación de metadatos en la comunidad bibliotecaria en general.

LA TRANSICIÓN A DATOS ENLAZADOS E IDENTIFICADORES

Los datos enlazados ofrecen la capacidad de aprovechar los *datos estructurados* con un énfasis en el contexto. Se basan en identificadores independientes del lenguaje que indican objetos, centrándose en “cosas” que reemplazan las “cadenas de caracteres” inherentes a los registros actuales de autoridades y catálogos. Estos identificadores se pueden conectar a datos, vocabularios y términos relacionados en otros idiomas, disciplinas y ámbitos, incluidos ámbitos no bibliotecarios. Las aplicaciones de datos enlazados pueden absorber las contribuciones de otros y, por lo tanto, liberar a los especialistas de metadatos de tener que volver a describir cosas ya descritas en otro lugar, lo que les permite centrarse en dar acceso a las colecciones únicas y distintivas de sus instituciones. Esto promete una experiencia de usuario más rica y una mayor capacidad de búsqueda y descubrimiento con más relaciones de contexto de lo que es posible con nuestros sistemas actuales. Además, los datos enlazados ofrecen la oportunidad de ir más allá del ámbito de la biblioteca al aprovechar información sobre entidades procedentes de diversas fuentes.¹⁴



IMAGEN 1. “[Los cambios de las dinámicas de trabajo en la descripción de recursos](#)” por OCLC Research¹⁵

Se espera que los datos enlazados permitan a las bibliotecas ofrecer nuevos servicios de calidad que los modelos actuales no pueden mantener, que las personas de fuera puedan hacer un mejor uso de las descripciones de los recursos bibliotecarios y que los datos

sean más ricos porque más personas comparten su creación. El paso a un entorno de datos enlazados presagia cambios en los flujos de trabajo de descripción de recursos, como se muestra en la imagen 1.

La intención de trasladar operaciones de metadatos a datos enlazados depende de la disponibilidad de herramientas, el acceso a fuentes de datos enlazados para su reutilización, las mejores prácticas documentadas sobre identificadores y las descripciones de metadatos asociadas con ellos (“menciones”) y una masa crítica de implementaciones en un nivel de red.

LA AMPLIACIÓN DEL USO DE IDENTIFICADORES PERMANENTES

El Focus Group debatió la “garantía de futuro” de la catalogación, que se refiere a las oportunidades para liberar el poder de los metadatos en registros antiguos para diferentes interacciones y usos en el futuro. Los identificadores permanentes se consideraron cruciales para la transición de los metadatos actuales a las aplicaciones futuras.¹⁶ Los identificadores, en forma de cadenas alfanuméricas independientes del lenguaje, sirven como una forma abreviada de ensamblar los elementos necesarios para describir de forma única un objeto o recurso. Pueden resolverse a través de redes con protocolos específicos para encontrar, identificar y usar ese objeto o ese recurso. En el ámbito no bibliotecario, los números de la Seguridad Social o de los DNIs son ejemplos de tales identificadores. En el ámbito bibliotecario y en el ámbito académico el Focus Group señaló a ORCID (Open Researcher and Contributor ID; Identificador Abierto de Investigadores y Contribuidores)¹⁷ como el “pegamento” que permite que se mantengan juntos los cuatro brazos de una obra académica: publicación, repositorio, catálogo bibliotecario e investigadores; sin embargo, ORCID se limita sólo a investigadores *vivos*. ORCID se utiliza cada vez más en revistas de ciencias, tecnología, ingeniería y matemáticas para todos los autores y contribuidores¹⁸ y se incluye en los sistemas de gestión de información investigadora (Research Information Management) de las instituciones. ISNI (International Standard Name Identifier; Identificador Normalizado Internacional de Nombres)¹⁹ identifica de forma unívoca personas y organizaciones involucradas en actividades creativas utilizadas por bibliotecas, editoriales, bases de datos y organizaciones de gestión de derechos, y abarca a creadores no vivos.

Los identificadores permanentes son utilizados por grupos como Google y HathiTrust para la integración de servicios.²⁰ Más instituciones están utilizando coordenadas geoespaciales en metadatos o URIs (Uniform Resource Identifiers; Identificadores de Recursos Uniformes), que apuntan a coordenadas geoespaciales que admiten llamadas API (Application Programming Interface; Interfaz de Programación de Aplicaciones) a GeoNames,²¹ lo que permite las búsquedas y descubrimientos de mapas. Las instituciones de investigación también están adoptando identificadores de personas como ORCID para agilizar la agregación del registro de investigación institucional, generalmente a través de un sistema de gestión de información investigadora (Research Information Management), como se documenta en el informe de OCLC Research de 2017 *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information Management (Conveniencia y cumplimiento: estudios de casos sobre identificadores permanentes en la gestión de información investigadora europea)*.²²

Si bien los editores son actores fundamentales en el flujo de trabajo de metadatos, los datos de los editores no siempre cumplen con los requisitos de las bibliotecas. Por ejemplo, los datos de los editores de monografías no suelen incluir identificadores. La British Library está trabajando con cinco editoriales del Reino Unido para agregar ISNIs²³ a sus metadatos como una prueba conceptual prometedora para los editores y bibliotecas que trabajan conjuntamente al principio de la cadena de producción. En la lista de deseos de los miembros de Focus Group para el futuro está la capacidad de cargar por lotes o agregar identificadores algorítmicamente.

No existe ningún identificador de nombre que abarque todos los casos de uso. Los nombres de los investigadores se han incluido parcialmente en los ficheros nacionales de autoridades de nombres que identifican a personas vivas y muertas. Una cantidad considerable de nombres antiguos figuran sólo como cadenas de caracteres en registros bibliográficos. Los registros de autoridad son creados solamente por instituciones que participan en el Name Authority Cooperative Program (NACO; Programa Cooperativo de Autoridades de Nombres) del PCC²⁴ o en programas de bibliotecas nacionales. Incluso entonces, los registros de autoridades se crean de forma selectiva para ciertos encabezamientos o a veces cuando hay referencias implicadas. El fichero de autoridades

de nombres LC/NACO contenía el 30% del total de nombres reflejados en los puntos de acceso de registros bibliográficos del WorldCat (9 millones de registros de LC/NACO en comparación con los 30 millones de nombres totales que aparecen declarados en la página del proyecto WorldCat Identities, que es de 2012).²⁵ En 2020 este porcentaje se redujo al 18%: 11 millones de registros de autoridades de LC/NACO en comparación con los 62 millones de WorldCat Identities. Estas estadísticas demuestran que el número de nombres que figuran en los registros bibliográficos aumenta más rápidamente que los que están bajo control de autoridades.

Los ficheros de autoridades se centran en la “forma preferida” de un nombre, que puede variar según el idioma, la disciplina, el contexto y la época. Ciertos académicos se han opuesto al concepto mismo de una “forma preferida”, ya que el nombre se puede denominar de manera diferente según el contexto.²⁶ Cuando un nombre tiene varias formas, los historiadores necesitan conocer la procedencia de cada nombre siguiendo las prácticas de citación comúnmente utilizadas en su campo. Este embrollo se podría resolver con un identificador que enlazara a las diferentes formas de un nombre, cada una asociada con su procedencia y su contexto.

Los nombres de investigadores son sólo un ejemplo de una necesidad que los sistemas de identificación actuales no satisfacen. Las instituciones han estado creando sus propios “identificadores locales” para satisfacer dicha necesidad. Los identificadores locales se usan, entre otros, en los siguientes casos: el registro de todos los investigadores del campus; la inclusión de entidades que estén poco presentes en los ficheros nacionales de autoridades, como los autores de disertaciones y tesis electrónicas, los intérpretes, los acontecimientos, los nombres de lugares locales y los edificios del campus; la identificación de entidades en proyectos de bibliotecas digitales y repositorios institucionales; el reflejo de las necesidades multilingües de la comunidad; y el apoyo a las tareas internas, como el registro de títulos de colecciones archivísticas.²⁷

La necesidad constante de los miembros del Focus Group de eliminar la ambigüedad de los nombres en distintas disciplinas y formatos impulsó la creación del grupo de trabajo de OCLC Research para el Registering Researchers in Authority Files (Registro de

Investigadores en Ficheros de Autoridades).²⁸ La necesidad de registrar con precisión las afiliaciones institucionales de los investigadores, con los propósitos de reflejar la producción académica de la institución, promover colaboraciones interinstitucionales y llevar a una contratación y a una obtención de subvenciones más exitosas, condujo a la creación de otro grupo de trabajo para el Addressing the Challenges with Organizational Identifiers and ISNI (Tratamiento de los Desafíos con Identificadores de Organizaciones e ISNI),²⁹ que presentó nuevos modelos de datos de organizaciones que otras podrían adaptar para sus propios usos. Desde entonces se lanzó el Research Organization Registry (ROR; Registro de Organizaciones de Investigación) para desarrollar un identificador abierto, sostenible, utilizable y único para todas las organizaciones investigadoras del mundo.³⁰

La desambiguación de nombres es la parte más laboriosa del trabajo de autoridades y seguirá siendo un requisito previo para la asignación de identificadores únicos. Dados los diferentes sistemas de identificación de nombres que ya se utilizan, las bibliotecas necesitan un servicio de conciliación de nombres. El trabajo de autoridades y los algoritmos basados en la coincidencia de cadenas de caracteres tienen sus límites; los resultados aún necesitarán una revisión manual por parte de expertos. Puede ayudar el aprovechar la experiencia en las comunidades de usuarios para verificar si dos identificadores representan a la misma persona.

La desambiguación es particularmente difícil para los autores o los contribuidores que figuran en artículos de revistas, donde los nombres a menudo se abrevian y puede haber docenas o incluso cientos de contribuidores. Por ejemplo, un artículo de *Physical Review Letters*, [“Precision Measurement of the Top Quark Mass in Lepton + Jets Final State”](#), tiene aproximadamente 300 nombres de autor abreviados para ser sólo un artículo de cinco páginas (imagen 2).³¹ Esto sirve de ejemplo de las diferentes prácticas entre disciplinas. Por el contrario, otros materiales con muchos contribuidores, como largometrajes y grabaciones de orquesta, habitualmente figuran con muy pocos nombres asociados en los metadatos antiguos de las bibliotecas.³² Tales diferencias convierten en un desafío la creación de metadatos uniformes, comprensibles y ampliamente reutilizables.



IMAGEN 2. Algunos de los [300 nombres de autor abreviados](#) para un artículo de cinco páginas de *Physical Review Letters*

Las formas abreviadas de los nombres de autores en los artículos de revistas dificultan, y a menudo imposibilitan, emparejarlos con la forma de autoridad correcta o con un identificador, si existe. La asociación de ORCID con autores de artículos facilita la diferenciación de autores con las mismas formas abreviadas. Los sistemas de gestión de información

investigadora (Research Information Management; RIM) aplican la gestión de identidades de investigadores locales para que estén correctamente asociados con los artículos que han escrito. Sus artículos se muestran como parte de sus perfiles (véase, por ejemplo, los perfiles de investigación de Experts@Minnesota o de University of Illinois at Urbana-Champaign’s Experts).³³

Para que funcione la gestión de identidades de investigadores, ellos mismos deben crear y mantener sus propios ORCID. Las instituciones han estado animando a sus investigadores en sus perfiles. Los investigadores están más incentivados en adoptar un ORCID para cumplir los requisitos de subvenciones económicas privadas y de ayudas públicas, como los de la National Science Foundation y los National Institutes of Health en los Estados Unidos.³⁴ Los sistemas de gestión de información investigadora (RIM) recolectan metadatos de bases de datos de resúmenes e indexaciones como Scopus, Web of Science y PubMed, cada uno de los cuales tiene sus propios identificadores de persona que ayudan con la desambiguación; también pueden estar enlazados al ORCID de un autor. Los datos enlazados podrían acceder a información en muchos entornos, incluidos los de los sistemas de información investigadora (Research Information Systems), pero requerirían que se enlazara con precisión varios identificadores de la misma persona entre sí.

Algunos miembros del Focus Group están realizando un trabajo de conciliación de metadatos, como es el buscar términos coincidentes en fuentes de datos enlazados y agregar sus URIs en registros de metadatos como un primer paso necesario hacia un entorno de datos enlazados o como parte del trabajo de mejora de metadatos.³⁵ La mejora de la calidad de los datos mejora la experiencia de los usuarios a corto plazo y ayudará con la transición a los datos enlazados más adelante. La mayor parte de la conciliación de metadatos se realiza sobre nombres de persona, materia y lugar geográfico. Las fuentes utilizadas para dicha conciliación son el Virtual International Authority File (VIAF; Fichero de Autoridades Internacional Virtual) de OCLC; el servicio de datos enlazados de la Library of Congress (id.loc.gov); ISNI; la Union List of Artist Names (ULAN; Lista Colectiva de Nombres de Artistas), el Art and Architecture Thesaurus (AAT; Tesoro de Arte y Arquitectura) y el Thesaurus of Geographic Names (TGN; Tesoro de Nombres Geográficos) de Getty; la Faceted Application of Subject Terminology (FAST; Aplicación por Facetas de la Terminología de Materias) de OCLC; y varios ficheros nacionales de autoridades. La selección de la fuente depende de la fiabilidad de la organización responsable, el tema en cuestión y la riqueza de la información. Este trabajo de conciliación de metadatos requiere de mucha mano de obra y no es muy escalable.

Algunos miembros del Focus Group han experimentado con la obtención de identificadores (URIs permanentes de fuentes de datos enlazados) para posteriormente reemplazar su dependencia actual de cadenas de caracteres. Las instituciones concluyeron que es más eficiente crear URIs en los registros de autoridades desde el principio que reconciliarlos más adelante. La Universidad de Michigan ha desarrollado un programa de LCNAF Named Entity Reconciliation (Reconciliación de Entidades con Nombre de LCNAF)³⁶ utilizando Open Refine de Google, que busca el fichero VIAF con la API de VIAF para hallar coincidencias, busca registros de origen de la Library of Congress dentro de un grupo de VIAF y extrae el encabezamiento autorizado. Esto da como resultado un conjunto de datos que empareja el encabezamiento autorizado del LC Name Authority File (Fichero de Autoridades de Nombres de la Library of Congress) con el encabezamiento original y un enlace al URI del servicio de datos enlazados del LCNAF. Este servicio podría modificarse para incorporar en su lugar al identificador

VIAF; obtiene buenos resultados a pesar de que utiliza la coincidencia de cadenas de caracteres.

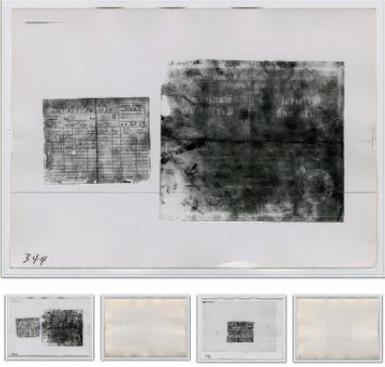
Se ha identificado una larga lista de fuentes no bibliotecarias que podrían mejorar los datos de autoridad actuales o que podrían ser útiles para poner enlaces a ellas en determinados contextos. A la cabeza de esta lista se encuentran Wikidata y Wikipedia. Otras fuentes son AllMusic, sitios web de autores y fans, Discogs, EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons, and Families; Contexto de Archivos Codificados para Entidades Corporativas, Personas y Familias), EAD (Encoded Archival Description; Descripción de Archivos Codificados), árboles genealógicos, GeoNames, GoodReads, IMDb (Internet Movie Database; Base de Datos de Películas de Internet), Internet Archive, Library Thing, LinkedIn, MusicBrainz, ONIX (ONline Information eXchange; Intercambio de Información en Línea), Open Library, ORCID y Scopus ID. El documento del Task Group on URIs in MARC (Grupo de Trabajo sobre los URIs en MARC) del PCC *Formulating and Obtaining URIs: A Guide to Commonly Used Vocabularies and Reference Sources* (Formulación y obtención de URIs: una guía para vocabularios y fuentes de referencia de uso común)³⁷ proporciona una guía útil para la agregación de datos a partir de estas otras fuentes. Wikidata es considerada como una fuente importante para ampliar la variedad de idiomas y proporcionar metadatos multilingües con más facilidad que con los sistemas bibliotecarios actuales.³⁸

Los identificadores de “obras” suponen un desafío particular, ya que no hay consenso sobre lo que es una “obra distintiva”.³⁹ Los identificadores de obra locales no se pueden compartir ni reutilizar. Los miembros del Focus Group expresaron su preocupación de que las diferentes interpretaciones de lo que es una “obra” podrían obstaculizar la capacidad de reutilizar los datos creados en otro lugar, y tienen puestas las esperanzas en un repositorio central fiable como OCLC para publicar identificadores de obras permanentes que podrían ser utilizados por toda la comunidad.

Ejemplos de algunos identificadores DOI y ARK

Registration Agency	Examples
DataCite	<p>DOI names for accessing registered research datasets:</p> <p>Sets & Subsets:</p> <ul style="list-style-type: none"> • Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo. [doi:10.1594/PANGAEA.726855] <p>Earthquake Event, Authored by Automated System:</p> <ul style="list-style-type: none"> • Geofon operator (2009): GEOFON event gfz2009kciu (NW Balkan Region) GeoForschungsZentrum Potsdam(GFZ). [doi:10.1594/GFZ.GEOFON.gfz2009kciu]
Multilingual European DOI Registration Agency mEDRA	<p>Sampling of resources identified using mEDRA DOI names:</p> <ul style="list-style-type: none"> • Journal Article: Prodi, Romano. "L'industria dopo l'euro", <i>L'industria-Rivista di economia e politica industriale</i> 4, 559-566 (2002); [doi:10.1430/8103] • Monograph: Attanasio, Piero. "The use of DOI system in eContent value chain: The case of Casalini Digital Division and mEDRA", White Paper (PDF). [doi:10.1392/BC1.0]
Japan Link Center (JALC)	<p>DOI names for Japanese Journal articles:</p> <ul style="list-style-type: none"> • Journal Article: 竹本 賢太郎, 川東 正美, 久保 信行, 左近 多喜男, 大学におけるWebメールとターミナルサービスの研究, <i>標準化研究</i> Vol.7(2009), No.1 p.11-20 [doi:10.11467/iss2003.7.1_11] • Journal Article: 川崎 努, 植物における免疫誘導と病原微生物の感染戦略, <i>ライフサイエンス 領域融合レビュー</i>, 2, e008 (2013), [doi:10.7875/leading_author.2_e008]

[Photographs of Identification Cards]



• **Archival Resource Key:** [ark:/67531/metaph346793](https://n2t.org/ark:/67531/metaph346793)

Relationships

- [Identification Cards, Photograph #1], DSMA_91-001-photo-344-001, [ark:/67531/metaph346554](https://n2t.org/ark:/67531/metaph346554)
- [Identification Cards, Photograph #2], DSMA_91-001-photo-344-002, [ark:/67531/metaph346846](https://n2t.org/ark:/67531/metaph346846)
- [Identification Cards, Photograph #3], DSMA_91-001-photo-344-003, [ark:/67531/metaph346928](https://n2t.org/ark:/67531/metaph346928)

IMAGEN 3. Ejemplos de algunos identificadores [DOI](#) (izquierda) y [ARK](#) (derecha)⁴⁰

Los identificadores deben ser invariables a lo largo del tiempo y deben ser independientes del lugar en dónde se encuentra el objeto digital o del lugar en dónde se almacenará. Por ejemplo, los identificadores para grupos de datos como los recursos digitales y las colecciones de repositorios institucionales son identificadores generados por el sistema, identificadores creados localmente, handles PURL, DOIs (Digital Object Identifiers; Identificadores de Objetos Digitales), URIs, URNs y ARKs (Archival Resource Keys; Claves de Recursos de Archivos). En la imagen 3 se muestran algunos ejemplos de identificadores DOI y ARK. Los recursos pueden tener varias copias y versiones que cambian con el tiempo. Los repositorios institucionales utilizados como espacios de colaboración pueden conducir a diversas publicaciones a partir de los mismos grupos de datos, un problema agravado por los autodepósitos de los coautores de diferentes instituciones en diferentes repositorios. Además, las bibliotecas, así como las agencias de evaluación y financiación nacionales, quieren poder enlazar partes relacionadas, como *preprints*, datos complementarios e imágenes, con la publicación. Varios DOIs que apuntan al mismo objeto suponen un problema. Algunas bibliotecas usan DataCite o Crossref para crear y publicar identificadores únicos a largo plazo y así minimizar la posibilidad de que haya enlaces de citas rotos.⁴¹ En el mejor de los casos

las bibliotecas contribuirían a un centro coordinador para los metadatos que describen los conjuntos de datos de sus investigadores con independencia de dónde se almacenen.

EL PASO DEL “CONTROL DE AUTORIDADES” A LA “GESTIÓN DE IDENTIDADES”

La prioridad en el trabajo de autoridades está cambiando desde la construcción de cadenas de caracteres hacia la *gestión de identidades*: la diferenciación de entidades, la creación de identificadores y el establecimiento de relaciones entre entidades.⁴² El trabajo intelectual que se necesita para diferenciar los nombres es el mismo tanto para el actual trabajo de autoridades como para la gestión de identidades. Los miembros del Focus Group están de acuerdo en que el futuro está en la gestión de identidades y en alejarse de la “gestión de cadenas de caracteres” como fundamento para controlar los encabezamientos de los registros bibliográficos.⁴³ Pero la gestión de identidades plantea un cambio de enfoque, desde proporcionar puntos de acceso en las descripciones de recursos hasta describir las *entidades* del recurso (obra, personas, entidades corporativas, lugares, eventos) y establecer las relaciones y enlaces entre ellas.

La transición desde el “control de autoridades” y los “puntos de acceso autorizados” de nuestros sistemas antiguos hacia la gestión de identidades requiere que separemos los identificadores de sus etiquetas asociadas. Un identificador único podría asociarse con una agregación de atributos que permitiría a los usuarios distinguir una entidad de otra.⁴⁴ En el mejor de los casos las bibliotecas podrían aprovechar los identificadores y atributos de otras fuentes no bibliotecarias. Wikidata, por ejemplo, agrega una variedad de identificadores, así como etiquetas en diferentes idiomas, como se muestra en la imagen 4.

Un identificador de Wikidata enlaza a otros identificadores y etiquetas en diferentes idiomas

Wikidata Identifier Q19526



Maya Angelou

Angelou reciting her poem "On the Pulse of Morning" at US President Bill Clinton's inauguration, January 20, 1993

Born Marguerite Annie Johnson
April 4, 1928
St. Louis, Missouri, U.S.

Died May 28, 2014 (aged 86)
Winston-Salem, North Carolina, U.S.

Occupation Writer · poet · civil rights activist

Period 1951–2014

Subject Memoir · poetry

Notable works *I Know Why the Caged Bird Sings*
"On the Pulse of Morning"

Spouses Tosh Angelos (m. 1951; div. 1954)
Paul du Feu (m. 1974; div. 1983)

Children 1

Website
www.mayaangelou.com

Other Identifiers

BnF ID

Freebase ID

GND ID

IMDb ID

ISNI ID

Library of Congress ID

Music Brains Artist ID

National Thesaurus for AuthorNames ID

Viaf ID

Other Labels

ar	مايا أنجيلو	ja	マヤ・アンジェロウ
azb	مايا آنجلو	ko	마야 앤젤로
bg	Мая Анджепов	ml	മായ (ആഞ്ചലോ)
el	Μάγια Αγγέλου	ne	माया एन्जोलो
fa	مايا آنجلو	si	මායිසා ඇන්ජලෝ
he	מאיה אנג'לו	zh	马娅·安杰卢
hy	Մայա Անջելոու		



IMAGEN 4. [Un identificador de Wikidata enlaza a otros identificadores y etiquetas](#) en diferentes idiomas

Proporcionar información de contexto es más importante que proporcionar una etiqueta única. Las etiquetas pueden diferir según comunidades, como en diversas ortografías de nombres y términos, diferentes idiomas y sistemas de escritura, y diferentes disciplinas, sin requerir que se prefiera una forma sobre otra. En las etiquetas se ha vuelto prioritario lo local en lugar de lo homogéneo para su uso global.

Una barrera crucial para pasar de las cadenas de caracteres a la gestión de identidades es la falta de tecnología e infraestructura que la respalde. Se necesitan nuevas herramientas para indexar y mostrar información sobre las entidades descritas con enlaces a las fuentes de los identificadores. Dado que varios identificadores pueden indicar la misma entidad, también se necesitarán herramientas para reconciliarlos. Algunos sistemas indexan sólo los puntos de acceso controlados, lo que es un problema cuando se trata de nombres que figuran en diferentes idiomas. ¿Se pueden reconfigurar los sistemas

bibliotecarios para que los *identificadores* sean los puntos de conexión para que se muestre e indexe cualquier etiqueta asociada a ellos?⁴⁵

Algunos miembros del Focus Group están haciendo experimentos con Wikidata como otra opción para asignar identificadores a nombres que no figuran en ficheros de autoridades, lo que ampliaría el conjunto potencial de contribuidores.⁴⁶ Muchas bibliotecas están barajando Wikidata y Wikibase, la plataforma de software que subyace a Wikidata, para resolver algunos de los viejos problemas afrontados por los departamentos de servicios técnicos, las unidades de archivos, etc.⁴⁷ Wikidata/Wikibase está siendo considerada como una posible alternativa al control de autoridades tradicional y tiene otros posibles beneficios, como el soporte multilingüe integrado y el puente entre los silos que describen los recursos institucionales. Los experimentos de los miembros del Focus Group con los proyectos de Wikidata y OCLC que utilizan la plataforma Wikibase indican que Wikibase es una estructura plausible para realizar implementaciones de datos enlazados. Esta infraestructura podría permitir que el Focus Group y las comunidades bibliográficas y de archivos se centren en las entidades que se deban crear, sus relaciones entre sí y la mejor manera de aumentar la capacidad de búsqueda y descubrimiento por parte de los usuarios.

Debido a que Wikidata se inició originalmente extrayendo datos de Wikipedia, la representación de los libros en Wikidata tiene un enfoque en “obras” y sus autores. Este enfoque de obras y autores podría considerarse como una versión alternativa de las entradas tradicionales de autor-título en los ficheros de autoridades. Los libros que son “insignes” tienen más probabilidades de figurar en Wikidata. Recientemente, una iniciativa que da soporte a las citas de los artículos de Wikipedia, WikiCite,⁴⁸ demuestra que es necesario registrar y dar soporte a los identificadores que componen esas citas, incluyendo la información sobre una edición o documento específico.

Uno de los aspectos más prácticos, y potentes, de la gestión de identidades es reducir los “copia y pega” en las dinámicas de trabajo de metadatos bibliotecarios cuando un identificador se custodia en una ubicación externa. Los identificadores podrían proporcionar un puente entre entornos MARC y no MARC y hacia recursos no bibliotecarios. Los bibliotecarios no tendrían que ser expertos en todos los ámbitos.⁴⁹

Muchos recursos conservados o gestionados por bibliotecas no están bajo el control de autoridades, como las colecciones digitales y de archivos, los repositorios institucionales y los datos de investigación. Los identificadores podrían proporcionar enlaces a estos recursos. La gestión de identidades también podría servir de puente entre las variantes de nombres que se encuentran en artículos de revistas, servicios de perfiles académicos y catálogos de bibliotecas, trascendiendo estos ámbitos ahora compartimentados. Este puente es un requisito para cumplir las promesas de los datos enlazados.

LA RESPUESTA A LA NECESIDAD DE VOCABULARIOS MÚLTIPLES Y A LA EQUIDAD, DIVERSIDAD E INCLUSIÓN

Los conceptos o los encabezamientos de materia son particularmente espinosos, ya que su terminología puede diferir según la época y la disciplina. En algunos casos, algunas comunidades pueden considerar los términos peyorativos, dañinos o incluso racistas. Abordar los problemas del idioma es importante, ya que las bibliotecas buscan desarrollar relaciones y generar confianza con las comunidades marginadas. Las cuestiones relativas a la equidad, la diversidad y la inclusión son complejas, el vocabulario utilizado en los encabezamientos de materia es tan sólo un aspecto, y los identificadores, que son independientes en cuanto al lenguaje, suponen una manera de abordar todo esto.

La cuestión de admitir encabezamientos de materia “alternativos” pasó a primer plano cuando no se llevó a cabo la solución inicial de la Library of Congress de cambiar su encabezamiento de materia “Illegal aliens” (“Extranjeros ilegales”) a “Undocumented immigrants” (“Inmigrantes indocumentados”). Esto llevó a un miembro del Focus Group a comentar: “Ser rehenes de un sistema nacional lento para los cambios ante una semántica cambiante es perjudicial para las bibliotecas, ya que generalmente nos enorgullecemos de ser acogedores e inclusivos”. Los usuarios responsabilizan a sus bibliotecas por lo que aparece en sus catálogos. Aunque los LCSH son los encabezamientos de materia de la Library of Congress (*Library of Congress Subject Headings*), se utilizan en todo el mundo, a veces perdiendo su contexto.⁵⁰

Hay quienes ven la Faceted Application of Subject Terminology (Aplicación por Facetas de la Terminología de Materias; FAST)⁵¹ como un medio para involucrar a la comunidad para mitigar los problemas que han impulsado los intentos de desarrollar encabezamientos de materia alternativos para los LCSH. Un subgrupo del Focus Group ha estado aplicando FAST a registros que de otro modo carecerían de materias. FAST fue desarrollado originalmente por OCLC como un punto medio entre descriptores no controlados, en un extremo del espectro, y cadenas de caracteres de materias precoordinadas difíciles de aprender y aplicar, en el otro extremo.⁵² Los encabezamientos de FAST proporcionan una transición sencilla a un entorno de datos enlazados, ya que cada encabezamiento de FAST tiene un identificador único. Puesto que los encabezamientos de FAST se generan a partir de los encabezamientos de materia precoordinados de la Library of Congress, también pueden incluir la misma terminología que algunos consideran inapropiada o irrespetuosa.

El recientemente conformado Policy and Outreach Committee (Comité de Política y Difusión) de FAST⁵³ representa a los usuarios de FAST para supervisar la participación de la comunidad, las contribuciones terminológicas y los procedimientos y para recomendar mejoras. Su declaración de objetivos dice:

FAST será un vocabulario general de materias apoyado totalmente, adoptado ampliamente y desarrollado de forma comunitaria que derivará de los LCSH con herramientas y servicios que satisfagan las necesidades de diversas comunidades y contextos.⁵⁴

Ya existen varios vocabularios superpuestos y, a veces, contradictorios en los datos bibliotecarios antiguos.⁵⁵ Por ejemplo, los miembros del Focus Group de Nueva Zelanda añaden términos del tesoro de encabezamientos de materia maoríes (Ngā Upoko Tukutuku) a los mismos registros de los encabezamientos de materia de la Library of Congress; los miembros del Focus Group de Australia añaden términos autorizados en los tesauros del Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS; Instituto Australiano de Estudios Aborígenes e Isleños del Estrecho de Torres).⁵⁶ Puede que no existan equivalencias satisfactorias entre idiomas. Los diferentes conceptos de los vocabularios de las bibliotecas nacionales no siempre se

pueden asignar de manera inequívoca a los conceptos del inglés. El plurianual MACS (Multilingual Access to Subjects; Acceso Multilingüe a las Materias)⁵⁷ estableció relaciones a través de tres vocabularios de materias: los encabezamientos de materia de la Library of Congress, el fichero de autoridades integrado alemán GND y el francés RAMEAU (Répertoire d’Autorité-Matière Encyclopédique et Alfabétique Unifié; Repertorio de Autoridades de Materia Enciclopédico y Alfabético Unificado). Ha sido un proceso laborioso y no se sabe si se implementará de manera generalizada.⁵⁸

Un porcentaje creciente de datos en las opciones de búsqueda y descubrimiento de las instituciones no proviene de fuentes bibliotecarias ni de MARC. Los metadatos que describen los datos y materiales de investigación universitarios en los repositorios institucionales generalmente se tratan de manera diferente, y por separado. ¿Cómo deberían las instituciones proporcionar normalización y acceso a las entidades descritas para que los usuarios no experimenten la “colisión de espacios de nombres” y los términos ambiguos (o sea, términos que significan cosas diferentes según la fuente)? En este contexto parece prometedora para la creación y gestión de varios tipos de vocabularios controlados la herramienta Ontology Management-Graphite de Synaptica Knowledge Solutions.⁵⁹

Los miembros del Focus Group citaron ejemplos de vocabularios o grupos de datos establecidos que se han quedado obsoletos o no satisfacen las necesidades o sensibilidades locales. A causa de los modelos de mantenimiento para vocabularios establecidos que van lentos o no responden, hay quienes han prestado consideración a los modelos distribuidos. Los altos umbrales de formación necesarios para poder participar en los modelos actuales han favorecido la pretensión de alternativas.⁶⁰

Los datos enlazados podrían proporcionar los medios para que las comunidades locales escojan una etiqueta diferente para el término autorizado de un vocabulario establecido para un concepto o entidad. Se puede hacer referencia a una descripción local de un concepto o entidad que no figura, o no figura satisfactoriamente, en vocabularios establecidos o fuentes de datos enlazados. Si este tipo de modificaciones y adiciones se hacen posibles en un entorno de datos enlazados, otros podrían estar de acuerdo, o en desacuerdo, con el punto de vista al enlazar al nuevo recurso. Como modelo distribuido

para la gestión tanto de la terminología como de la descripción de la entidad plantea problemas en torno a las expectativas de estabilidad de los metadatos, la interoperabilidad de los metadatos y el mantenimiento de los metadatos. ¿Cómo podría un modelo distribuido evitar que la gente duplique el trabajo en la misma entidad o concepto? ¿Cómo obtendría un modelo distribuido la confianza de los contribuidores, o cómo determinaría a quiénes se les permite colaborar?

Los numerosos vocabularios creados para proyectos específicos que, una vez finalizada la subvención económica, permanecen congelados en el tiempo han hecho aflorar los problemas de estabilidad y permanencia. Como señaló un miembro del Focus Group: “No hay nada más triste que dejar que un vocabulario que alguien inventó se quede trasnochado”. Estos ejemplos proporcionan una razón importante para que los bibliotecarios quieran confiar en ficheros de autoridades internacionales en lugar de soluciones locales. También ilustran la valía de que la Library of Congress asuma el coste total de crear y mantener los LCSH.

El informe del OCLC Research sobre los resultados de una encuesta de 2017 de la Research Library Partnership sobre equidad, diversidad e inclusión (EDI)⁶¹ estimuló los debates sobre la complejidad de incorporar la equidad, la diversidad y la inclusión en los vocabularios controlados de los catálogos de bibliotecas.⁶² Las instituciones educativas y las bibliotecas han emprendido iniciativas EDI, y los departamentos de metadatos han estado esforzándose por apoyarlas. El extracto de la encuesta EDI en la imagen 5 muestra que los metadatos de los catálogos de bibliotecas están a la zaga de otras áreas en apoyo de los objetivos y principios EDI de esta institución.

¿Qué áreas ha cambiado o planea cambiar debido a los objetivos y principios EDI de su institución?

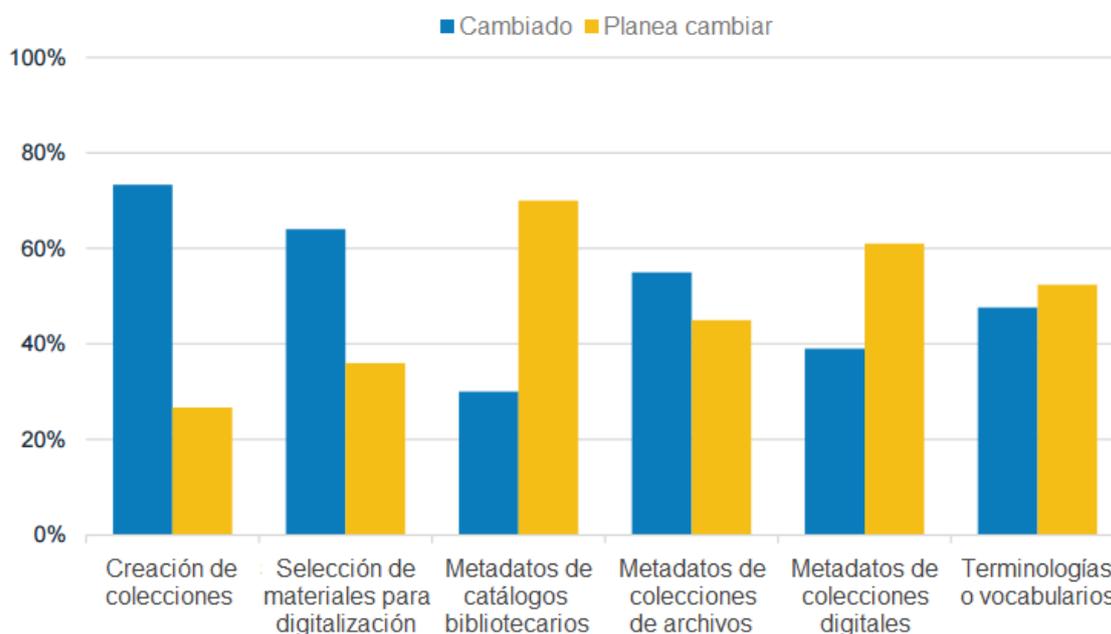


IMAGEN 5. Extracto del informe de resultados de la encuesta EDI de 2017 del Research Library Partnership⁶³

Los miembros del Focus Group están impacientes por proporcionar un acceso a materias más detallado que el que ofrecen actualmente los sistemas nacionales de encabezamientos de materia, como los LCSH, que tienen más nivel de detalle para los lugares de Europa Occidental que para el Sudeste Asiático y África. Ven necesario ofrecer términos más precisos y actuales y reemplazar los términos que reflejan sesgos o se consideran ofensivos por términos más neutrales.

Los desafíos que identificaron los miembros del Focus Group al ofrecer una terminología más respetuosa en el acceso a las materias para los usuarios:

- **Búsqueda y descubrimiento:** el uso de otros vocabularios menos ofensivos localmente puede dividir las colecciones en el catálogo de la biblioteca, lo que dificulta la búsqueda y descubrimiento de todos los materiales relevantes.
- **Falta de consenso:** los miembros del Focus Group dudan de que alguna vez pueda haber un consenso completo sobre una cadena de texto determinada. Es posible que los términos que puedan ser ofensivos para una comunidad no

siempre resulten claros para otras (por ejemplo, “Arte disidente” en lugar de “Arte inconformista”).

- **Velocidad:** el proceso de cambiar los encabezamientos de materia normalizados puede ser muy lento.
- **Capacidad:** Cambiar los encabezamientos de los registros existentes puede suponer una tarea inmensa. El mantenimiento específico de puntos de acceso se produce en el contexto del mantenimiento general de puntos de acceso. Por ejemplo, la Library of Congress cambió recientemente el encabezamiento “Mentally handicapped” (“Minusválidos mentales”) a “People with mental disabilities” (“Personas con discapacidades mentales”). La implementación de dichos cambios en el catálogo puede implicar una combinación de métodos correctivos automatizados y manuales, así como decisiones sobre la asignación de recursos.⁶⁴ Hay quienes señalaron que sería menos laborioso presentar un mensaje de “sensibilidad cultural” como parte del interfaz de búsqueda para alertar a los usuarios de que los términos y anotaciones que encuentren en un catálogo pueden reflejar la actitud del creador o de la época en la que se crearon y pueden considerarse inapropiados hoy en día en ciertos contextos.
- **Compartir:** Los vocabularios locales no se pueden compartir con otros sistemas.
- **Mantenimiento:** Quienes han intentado utilizar vocabularios locales más adecuados para su contexto y sus comunidades los encontraron demasiado difíciles de mantener y los abandonaron.
- **Barreras lingüísticas:** El idioma de nuestros vocabularios controlados puede resultar excluyente para destinatarios que no leen ese idioma. Las bibliotecas de la Universidad Estatal de Ohio han tratado de abordar esto desarrollando equivalencias de escritura no latina de los términos de materia en inglés.
- **Clasificación:** Los sistemas de clasificación actuales tienden a segregar grupos étnicos. En lugar de incluirlos como parte de un concepto general como historia, educación o literatura, tienden a agruparse como un solo grupo. A medida que las instituciones almacenan más publicaciones externamente, la necesidad de archivar los materiales juntos y tener una sola clasificación en un registro ha disminuido, pero pocas aplican clasificaciones múltiples en un registro.

Los requisitos para un sistema distribuido que se adapta a vocabularios múltiples y que también podría admitir EDI convergieron en torno a la necesidad de admitir relaciones semánticas entre diferentes vocabularios. Las comunidades de práctica necesitan un centro coordinador para agregar y reconciliar términos dentro de sus propios ámbitos. Se señaló que diferentes comunidades de práctica podrían utilizar términos que entren en conflicto con la terminología de otros o que signifiquen cosas diferentes. *Linked Data Infrastructure Models: Areas of Focus for PCC Strategies (Modelos de infraestructuras de datos enlazados: áreas de enfoque para las estrategias del PCC)* del Linked Data Advisory Committee (Comité Asesor de Datos Enlazados) del PCC⁶⁵ describe requisitos funcionales de alto nivel y un espectro de modelos anticipados a medida que las instituciones de patrimonio cultural adopten datos enlazados como una estrategia para compartir datos. El modelo debe ser ampliable y extensible, con la capacidad de hospedar la proliferación de nuevos temas y términos propios de las humanidades y las ciencias y facilitar las colaboraciones de los propios investigadores. Debe ser lo suficientemente flexible para coexistir con otros vocabularios.

El reemplazo de cadenas de caracteres por identificadores estables y permanentes facilitaría el uso de diferentes etiquetas según el contexto. Esto daría cabida tanto a diferentes idiomas y escrituras, y diferentes ortografías dentro de un idioma, como el inglés americano frente al inglés británico, así como a términos que son más respetuosos con las comunidades marginadas. El seminario web del OCLC Research Works in Progress del 19 de octubre de 2017 sobre “Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries, and Archives” (“Descolonizar las descripciones: encontrar, nombrar y cambiar las relaciones entre pueblos indígenas, bibliotecas y archivos”)⁶⁶ describió el proceso iniciado por la Association for Manitoba Archives (Asociación para los Archivos de Manitoba) y las bibliotecas de la Universidad de Alberta para examinar encabezamientos de materia y esquemas de clasificación y considerar cómo podrían ser más respetuosos e inclusivos con las experiencias de los pueblos indígenas.

La ampliación de los vocabularios para incluir aquellos que se utilizan en otras comunidades requiere construir relaciones de confianza. Podría ser más inclusivo un modelo de “colaboración comunitaria” para los nuevos términos y votación

comunitaria. El actual “entorno de consenso” de las bibliotecas excluye a mucha gente. Muchos metadatos se crean actualmente de acuerdo con las construcciones de conocimiento occidentales y en torno a ellos se han diseñado los sistemas. Comunicar el historial de cambios y la procedencia de cada término nuevo o modificado proporcionaría una transparencia que podría contribuir a la confianza en la fuente. Un posible modelo a seguir sería el historial de edición y las páginas de debate que forman parte de la descripción de cada entidad de Wikidata. Exigir la procedencia como parte de un modelo de vocabulario distribuido puede ayudar a crear un entorno alternativo que sea más equitativo, diverso e inclusivo.

LOS DESAFÍOS DE LOS DATOS ENLAZADOS

Los identificadores y vocabularios son sólo dos componentes necesarios en la transición a datos enlazados. Una parte fundamental de la descripción de entidades son los datos asociados que se transcriben. ¿Cómo resolverán o conciliarán las bibliotecas los conflictos entre estos datos?⁶⁷ Pueden aparecer diferentes tipos de inconsistencias ahora, por ejemplo, diferentes fechas de nacimiento de persona. La procedencia de cada dato se vuelve más esencial. Incluso en el entorno actual ciertas fuentes son más fiables y dan a los catalogadores confianza en su precisión. Las bibliotecas suelen tener una lista de “fuentes preferidas”.⁶⁸ OCLC Research investigó cómo las bibliotecas podrían aplicar la “Knowledge Vault” (“Bóveda del Conocimiento”) de Google para identificar datos que pueden ser más “veraces” que otras en “Works in Progress Webinar: Looking Inside the Library Knowledge Vault” (2015; “Works in Progress Webinar: Mirando dentro de la bóveda bibliotecaria del conocimiento”).⁶⁹ Los miembros del Focus Group postularon que la comunidad bibliotecaria considera como de mayor confianza los datos de fuentes agregadas como WorldCat, Virtual International Authority File (VIAF; Fichero de Autoridades Internacional Virtual) y Wikidata. Los bibliotecarios podrían compartir su experiencia estableciendo las relaciones entre datos de diferentes fuentes.

Sin embargo, los datos enlazados de calidad requieren de metadatos de calidad. Los directivos son muy conscientes del conflicto entre brindar acceso a las colecciones bibliotecarias de manera rápida y proporcionar una descripción de buena calidad. Las descripciones de metadatos deben ser lo suficientemente completas para permitir que las bibliotecas administren sus colecciones y ofrezcan al usuario accesibilidad y capacidad

de visualización. Muchas bibliotecas necesitan equilibrar entre velocidad y precisión, velocidad y exhaustividad, o brevedad y nada. Estos equilibrios se reflejan en el uso de registros de proveedores inadecuados, en la creación de descripciones de nivel mínimo o incompleto para ciertos tipos de recursos y en la limitación del trabajo de autoridades. La catalogación de nivel mínimo se usa comúnmente como una alternativa a dejar los materiales sin catalogar, a menudo debido al gran volumen de materiales y a la insuficiencia de recursos humanos.⁷⁰ Estas descripciones incompletas darán como resultado menos datos enlazados y de peor calidad.

Presenta muchos desafíos el período de transición desde los sistemas de catalogación antiguos que dependen de MARC hacia un nuevo entorno de datos enlazados con entidades y atributos ya que tanto las normas como la práctica cambian constantemente. No está claro cómo las bibliotecas compartirán datos en lugar de registros en un entorno de datos enlazados. Los miembros del Focus Group estaban divididos sobre si se necesitaría un depósito de datos enlazados centralizado para proporcionar una “procedencia de confianza” o si los datos deberían distribuirse compartiéndose de forma privada entre unos y otros.⁷¹ Datos diferentes podrían ser correctos en sus propios contextos. Los “datos contradictorios” podrían representar diferentes visiones del mundo. Seleccionar datos según su procedencia podría ser un desafío para nuestros principios de equidad, diversidad e inclusión.

Los miembros del Focus Group se preguntaron cómo involucrar a los numerosos proveedores que suministran o procesan registros MARC en la transición a datos enlazados. En el Reino Unido la iniciativa de Jisc “Plan M” (donde “M” significa “metadatos”) busca agilizar el suministro de metadatos entre bibliotecas, editores, suministradores de datos y proveedores de infraestructuras.⁷² Entre las implicaciones, citadas por los colaboradores de la National Bibliographic Knowledgebase (NBK; Base Bibliográfica Nacional de Conocimiento) del Reino Unido en su previsión a 10 años del Plan M, “Deberán crearse y mantenerse en NBK casos de uso de datos enlazados convincentes sobre el impacto que tiene en la investigación”.⁷³ Trabajar con otros actores en el entorno de datos enlazados involucra a personas que no están familiarizadas con el entorno de las bibliotecas, lo que supone que los especialistas en

La transición a la siguiente generación de metadatos

metadatos expliquen cuáles son sus necesidades en términos que los no bibliotecarios puedan entender.

LA DESCRIPCIÓN DE COLECCIONES “DE DENTRO HACIA AFUERA” Y “FACILITADAS”

El vicepresidente y estratega jefe de OCLC Lorcan Dempsey se refiere al cambiante énfasis de las bibliotecas en ofrecer apoyo para la creación, conservación y capacidad de visualización de recursos institucionales como la “colección de adentro hacia afuera” (en contraste con la “colección de afuera hacia adentro”, en la que la biblioteca obtiene licencia o compra materiales de proveedores externos para que los destinatarios locales tengan acceso a ellos). Proporcionar acceso a una variedad más amplia de recursos locales, externos y colaborativos en torno a las necesidades de los usuarios es la “colección facilitada”.⁷⁴ Las actividades de los miembros del Focus Group se han centrado cada vez más en los metadatos que proporcionarán acceso a los recursos exclusivos de sus instituciones, así como de sus consorcios o redes nacionales.

Todos los recursos agregados, creados y conservados por bibliotecas necesitan metadatos para posibilitar su búsqueda y descubrimiento. Sin embargo, los miembros del Focus Group se concentraron en los desafíos y problemas relativos a unos formatos concretos:

- Colecciones archivísticas
- Sitios web archivados
- Colecciones audiovisuales
- Colecciones de imágenes
- Datos de investigación

Todos estos tipos de contenidos se pueden clasificar como pertenecientes a colecciones “de adentro hacia afuera” y presentan diferentes desafíos. Por ejemplo, los miembros del Focus Group describieron los esfuerzos para recuperar metadatos de sistemas completamente diferentes como “súper desafiantes”. Además, muchos de estos recursos no están bajo ningún control de autoridades. La conciliación de los puntos de acceso de diferentes tesauros y el trabajo de mapeo de metadatos requiere de experiencia y habilidades en servicios técnicos.⁷⁵ Esta conciliación también será necesaria en el entorno de datos enlazados que se ha tratado anteriormente.

Esta sección resume los debates sobre estos tipos de formatos.

COLECCIONES ARCHIVÍSTICAS

Las colecciones archivísticas son, en muchos sentidos, la joya de la corona de las colecciones, ya que son recursos de investigación únicos que brindan información sobre el mundo a lo largo de muchos siglos y lugares, proporcionando las fuentes primarias para la creación de nuevos conocimientos. El aumento de la visibilidad de estas colecciones genera importantes beneficios tanto para los académicos como para las bibliotecas y archivos. Sin embargo, los archivos son complejos y presentan problemas diferentes de metadatos en comparación con las colecciones de bibliotecas tradicionales. Como las instituciones recurren a ArchiveSpace y otros sistemas de gestión de contenidos para proporcionar infraestructuras para metadatos estructurados de archivos, están surgiendo varios problemas.⁷⁶

Los archivos han tenido más autonomía que las bibliotecas dentro de sus instituciones porque tienen colecciones únicas con su propio tipo de usuarios, sus propias normas de metadatos y sus propios sistemas. Si bien algunas instituciones han integrado el procesamiento de archivos dentro de los servicios técnicos, la mayoría mantiene una unidad separada. Los archiveros no tienen la tradición de crear registros de autoridad y compartir identificadores para la misma entidad, como es común entre los bibliotecarios. También tienden a usar la forma más completa de un nombre basándose en la información que se encuentra en las colecciones, mientras que los bibliotecarios se centran en la forma “preferida” que se encuentra en las publicaciones. Aun así, se ha estado produciendo un cambio significativo desde los métodos archivísticos artesanales hasta la normalización de metadatos.

Entonces, ¿cómo pueden los archiveros y bibliotecarios integrar mejor sus metadatos y prácticas de autoridades? La cantidad de nombres personales en las colecciones archivísticas puede ser tan grande que la mayoría no están controlados y no tienen identificadores. Sin embargo, la información de contexto que los archiveros brindan a las entidades de personas y organizaciones podría enriquecer la información provista en los ficheros de autoridades; una práctica que se exploró en el piloto de Project Passage

de 2017-2018⁷⁷ y que se examinó con más detalle en 2019-2020 por el Archives and Special Collections Linked Data Review Group (Grupo de Revisión de Datos Enlazados de Archivos y Colecciones Especiales) de OCLC Research Library Partners.⁷⁸

La mayor dependencia en recursos electrónicos y digitales durante la pandemia de la COVID-19 probablemente acelerará que las instituciones digitalicen colecciones archivísticas y sus colecciones singulares que han estado disponibles sólo en forma física hasta el momento.⁷⁹ Se pueden crear más metadatos a partir de versiones digitalizadas de estos recursos.

SITIOS WEB ARCHIVADOS

Durante algunos años los archivos y las bibliotecas han estado archivando recursos web de interés académico o institucional para garantizar su acceso continuo y su supervivencia a largo plazo. Algunos sitios web son efímeros o intencionalmente temporales, como los creados para un evento específico. A las instituciones les gustaría archivar y preservar el contenido de sus sitios web como parte de su registro histórico. Los rastreadores web recolectan la gran mayoría del contenido web, pero los metadatos generados tan sólo por recolección se consideran insuficientes para dar soporte a la búsqueda y descubrimiento.⁸⁰

Algunos sitios web archivados son colecciones temáticas institucionales que apoyan un área de investigación específica, como las de derechos humanos, preservación histórica y planificación urbanística y religiones de la Ciudad de Nueva York de la Universidad de Columbia.⁸¹ Las bibliotecas nacionales archivan sitios web dentro de su ámbito nacional. Por ejemplo, los sitios web archivados de la Biblioteca Nacional de Australia, desde 1996 hasta ahora,⁸² agregan sitios web en asociación con instituciones culturales del entorno de Australia, sitios web gubernamentales que antes eran accesibles a través del Australian Government Web Archive (Archivo Web del Gobierno de Australia) y sitios web del dominio .au agregados anualmente mediante recolecciones de rastreo a gran escala. Estas colecciones ordenadas por temas brindan instantáneas de la historia cultural y social de Australia. Ejemplos de sitios web archivados en consorcio son el Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY; Archivo Web Colaborativo de Arquitectura, Urbanismo y Sostenibilidad) y el

Contemporary Composers Web Archive (CCWA; Archivo Web de Compositores Contemporáneos) de la Ivy Plus Libraries Confederation y el New York Art Resources Consortium (NYARC; Consorcio de Recursos de Arte de Nueva York), que captura versiones dinámicas de web de catálogos de subastas y sitios web de artistas, galerías y museos.⁸³

El Focus Group debatió los desafíos de crear y administrar los metadatos necesarios para mejorar los metadatos recolectados por robots procedentes de sitios web. Algunos de los desafíos identificados son:

- **El tipo de sitio web es importante.** Los requisitos de los metadatos descriptivos pueden depender del tipo de sitio web archivado (por ejemplo, sitios efímeros, datos de investigación, redes sociales o sitios de organizaciones). A veces, sólo se archiva el contenido de los sitios web cuando no se considera relevante la experiencia de usuario (su “apariencia y sensación”).
- **Las prácticas varían.** Algunas características de los sitios web no se abordan en las reglas de descripción existentes, como RDA (Resource Description and Access; Recursos: Descripción y Acceso) y DACS (Describing Archives: A Content Standard; Descripción de Archivos: una Norma de Contenidos). Los metadatos tienden a seguir las tradiciones de descripción bibliográfica o las prácticas archivísticas, dependiendo de quiénes crean los metadatos.
- **Ten en cuenta el tamaño y el uso previsto.** Los requisitos de los metadatos pueden diferir según el tamaño del material que se archiva y su uso previsto. Por ejemplo, los humanistas digitales ven el contenido web como datos y lo analizan para fines como la identificación de tendencias, mientras que otros usuarios simplemente necesitan unas páginas en particular. El nivel de detalle de los metadatos (agregación, semilla/URL, documento) también puede variar según las necesidades anticipadas de los usuarios, el tamaño del material que se rastrea y el personal disponible.
- **Frecuencia de actualización.** Muchos sitios web se actualizan repetidamente, por lo que es necesario volver a rastrearlos cuando ha cambiado el contenido. Algunos tipos de cambios pueden provocar errores en la captura.

- **Sitios web multiinstitucionales.** Algunos sitios web están archivados por varias instituciones. Cada uno puede haber capturado el mismo sitio web en diferentes fechas y con diferentes especificaciones de rastreo. ¿Cómo se pueden buscar y utilizar en conjunto?

Una encuesta de 2015 de la OCLC Research Library Partnership reveló que la “falta de pautas de metadatos descriptivos” es el mayor desafío relacionado con el archivado de sitios web, lo que llevó a la formación del Web Archiving Metadata Working Group (Grupo de Trabajo de Metadatos de Archivado Web) de la OCLC Research Library Partnership.⁸⁴ Los desafíos que identificó el Focus Group fueron analizados en profundidad por este grupo de trabajo, que emitió un informe con sus recomendaciones en 2018, *Descriptive Metadata for Web Archiving (Los metadatos descriptivos para el archivado web)*.⁸⁵

COLECCIONES AUDIOVISUALES

Los miembros del Focus Group informaron que sus instituciones tenían depósitos llenos de grandes cantidades de materiales audiovisuales, que a menudo constituyen colecciones locales únicas.⁸⁶ Sin embargo, como afirma Chela Scott Weber en la publicación *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries (Plan de investigación y aprendizaje para colecciones archivísticas, especiales y singulares en bibliotecas de investigación)*, “durante décadas los materiales audiovisuales de nuestras colecciones fueron en general separados del material escrito relacionado, a menudo cambiados de sitio para ocuparse de ellos en una fecha posterior, o fueron tratados a nivel de ejemplar. Tanto una forma como otra han servido para crear una considerable acumulación de materiales audiovisuales sin cuantificar ni describir”.⁸⁷ Gran parte de este material audiovisual requiere con urgencia de conservación, digitalización, aclaración de condiciones de uso y descripción.

Además, las habilidades y los colaboradores necesarios en todas las instituciones son complejos. La naturaleza de la gestión de los recursos audiovisuales requiere del conocimiento del contexto de uso, así como de los problemas técnicos de metadatos, lo que aporta un entorno complejo para pensar en los requisitos de descripción y acceso.

Asimismo, las bibliotecas deben lidiar con los actuales multimedia de duración temporal que se producen localmente como parte de la investigación y el aprendizaje, o con los multimedia de *streaming* que tienen una licencia comercial.

Los debates del Focus Group se centraron en los recursos audiovisuales dentro de las colecciones archivísticas, a menudo en formatos en deterioro, en grandes nidos y que a veces requieren de equipos poco comunes y costosos para acceder y evaluar los archivos. Para el contenido generado localmente las instituciones prefieren que los creadores describan sus propios recursos.

El desafío principal fue cuánto esfuerzo se debe invertir para describir estos materiales audiovisuales porque son únicos. Las instituciones han utilizado estructuras jerárquicas para agregar materiales similares con *finding aids* que están recogidas en la norma Encoded Archival Description (Descripción de archivo codificado),⁸⁸ que proporciona información de contexto útil para cada ejemplar dentro de una colección específica. Pero a menudo un método de agregación en las descripciones puede carecer de detalles importantes sobre cada ejemplar que son necesarios para la búsqueda y descubrimiento, como el título transcrito y la fecha de emisión. Éste es un problema particularmente grave para los datos antiguos que describen registros de años anteriores. Los metadatos que describen los mismos materiales audiovisuales pueden diferir en los sistemas de gestión de recursos bibliotecarios, archivísticos y digitales. Hay quienes esperan que unas mejores opciones de búsqueda y descubrimiento aliviarán la necesidad de repetir la misma información en todas las bases de datos, pero presentar la información a los usuarios requeriría del uso de puntos de acceso uniformes en todos los sistemas. Lo mismo ocurrirá en un entorno de datos enlazados. Pero seguirá existiendo el desafío de enlazar ejemplares y la *finding aid* y de mantener los enlaces a lo largo del tiempo a pesar de los cambios en los sistemas.

Los metadatos para los materiales audiovisuales deben incluir información técnica importante, como detalles sobre el proceso de captura y digitalización audiovisual al igual que la compresión, el año de digitalización, la tecnología utilizada y la compatibilidad de archivos. Estos datos son fundamentales para garantizar el acceso constante a archivos tan enormes y a formatos de reproducción volátiles. Algunos

miembros del Focus Group han implementado PREMIS (Preservation Metadata: Implementation Strategies; Metadatos de Preservación: Estrategias de Implementación),⁸⁹ la norma internacional de metadatos para mantener la preservación de objetos digitales y garantizar su capacidad de uso a largo plazo, para algunos de sus materiales audiovisuales.

Chela Scott Weber, miembro directivo de Programas de OCLC, continúa trabajando con la Research Library Partnership en las necesidades y desafíos de la gestión de colecciones audiovisuales, que se resumen en las publicaciones del blog *Hanging Together* de OCLC Research: “Assessing Needs of AV in Special Collections” (“Evaluación de las necesidades de los audiovisuales en las colecciones especiales”) y “Scale & Risk: Discussing Challenges to Managing A/V Collections in the RLP” (“Escala y riesgo: debatiendo los desafíos de gestionar colecciones audiovisuales en la RLP”).⁹⁰ Una parte de los miembros del Focus Group respondió a la encuesta de Weber de 2019 para evaluar las necesidades de materiales audiovisuales en colecciones especiales dentro de la Research Library Partnership; las incorporaciones de colecciones audiovisuales en las dinámicas de trabajo tanto de archivos como de colecciones digitales fueron dos de los desafíos que más interesaron a los encuestados, como se muestra en la imagen 6.

¿Qué desafíos relativos a la gestión de colecciones de audiovisuales estarías interesado que abordara la RLP? (n=137)



IMAGEN 6. Respuestas de la encuesta de 2019 sobre desafíos relativos a la gestión de colecciones audiovisuales

COLECCIONES DE IMÁGENES

Los miembros del Focus Group gestionan una amplia variedad de colecciones de imágenes que presentan desafíos para la gestión de metadatos. En algunos casos las colecciones de imágenes que se desarrollaron fuera de la biblioteca y sus modelos de datos deben integrarse con otras colecciones o en nuevos entornos de búsqueda. Dependiendo de la naturaleza de la colección y sus usuarios surgen preguntas sobre la identificación de obras, representación de entidades, cronología, geografía, procedencia, género y materias (“de lo que trata” y “lo que representa”). Las colecciones de imágenes también ofrecen oportunidades para el *crowdsourcing* y la investigación interdisciplinaria.⁹¹

Muchas bibliotecas describen sus recursos de imágenes digitales a nivel de colección mientras que describen ejemplares de manera selectiva. En la medida de lo posible las mejoras se realizan por lotes. Hay quienes hacen trabajo de autoridades, dependiendo de

la calidad de los metadatos adjuntos. Algunas bibliotecas han difundido pautas de metadatos para ayudar a aportar más coherencia a los datos.

Entre los desafíos debatidos por el Focus Group se encuentran:

- **La variedad de sistemas y esquemas.** Las colecciones de imágenes creadas en diferentes partes de la institución, como los departamentos de arte o antropología, tienen diferentes propósitos y utilizan diferentes sistemas y esquemas que los utilizados por la biblioteca. Los metadatos a menudo vienen en hojas de cálculo o en datos adjuntos no estructurados. A menudo los metadatos creados por otros departamentos requieren de mucha edición, manipulación y revisión manual. La situación es más simple cuando toda la digitalización se maneja a través de una ubicación centralizada y la biblioteca se ocupa de toda la creación de metadatos. Algunas bibliotecas utilizan Dublin Core para los metadatos de sus colecciones de imágenes y otras utilizan MODS (Metadata Object Description Schema; Esquema de Descripción de Objetos de Metadatos).⁹² Hay quienes empaquetan los registros de metadatos en METS (Metadata Encoding and Transmission Standard; Norma de Codificación y Transmisión de Metadatos),⁹³ un esquema mantenido por la Library of Congress diseñado para expresar la naturaleza jerárquica de los objetos de las bibliotecas digitales, los nombres y ubicaciones de los archivos que componen esos objetos y los metadatos asociados. Hay quienes sugirieron que MODS se utilice junto con MADS (Metadata Authority Description Schema; Esquema de Descripción de Autoridades de Metadatos).⁹⁴
- **Los metadatos duplicados para diferentes objetos.** Los metadatos de dibujos escaneados pueden ser idénticos incluso aunque existan ligeras diferencias en esos dibujos. Es probable que la duplicación de metadatos en objetos similares se deba a la limitación de personal. Posiblemente los profesores o los responsables del escaneo podrían añadir más detalles.
- **La falta de procedencia.** Un desafío común es la recepción de colecciones de imágenes con escasos metadatos y sin información sobre su procedencia. Por ejemplo, el personal de metadatos de una institución recibió un texto de OCR recuperado por un investigador del HathiTrust. Millones de imágenes carecían

de la ubicación del material fuente original y, por lo tanto, limitaban, si no desacreditaban, cualquier uso posterior.

- **El mantenimiento de enlaces entre metadatos e imágenes.** ¿Cómo deberían las bibliotecas almacenar imágenes y mantenerlas sincronizadas con los metadatos? Puede haber problemas de derechos al depender de una plataforma específica para mantener enlaces entre metadatos e imágenes. ¿Dónde deberían alojarse las miniaturas?
- **La vinculación de vistas y versiones diversas del mismo objeto.** Varias versiones del mismo objeto tomadas a lo largo del tiempo pueden ser muy útiles para disciplinas como la ciencia forense. Por ejemplo, la Universidad de Brown decidió describir un conjunto de varias imágenes de la misma cosa en diferentes formatos y luego describir las versiones específicas incluidas. Este trabajo se realizó a pesar de que todavía no existe un sistema que muestre las relaciones entre imágenes, como los componentes de una pieza, incluso cuando los metadatos de los registros están empaquetados y almacenados en METS.
- **La gestión de las relaciones con profesores y conservadores.** Es importante asegurarse de que los profesores sientan que se satisfacen sus necesidades. La colaboración es necesaria entre los poseedores de los materiales, los especialistas en metadatos y los desarrolladores, ya que todos provienen de perspectivas diferentes. El desafío es apoyar simultáneamente un fin específico, grupos de personas y la búsqueda y descubrimiento a gran escala.
- **La agregación de colecciones digitales.** Las instituciones han estado compartiendo los metadatos de sus colecciones digitales con servicios de búsqueda y descubrimiento nacionales e internacionales. Dentro de cada organización los bibliotecarios crean y vuelven a crear metadatos para recursos digitales y digitalizados en una gran cantidad de sistemas: el catálogo de la biblioteca, la gestión del archivo, los sistemas de recursos digitales y preservación, el repositorio institucional, los sistemas de gestión de la investigación y los repositorios externos de suscripción. Los destinatarios de compartir estos metadatos van desde servicios de búsqueda y descubrimiento digitales temáticos hechos por encargo para agregaciones nacionales e internacionales, como Google Scholar, HathiTrust, Digital Public Library of America (DPLA), Internet Archive, Trove y WorldCat, hasta exposiciones en

línea, como Google Arts and Culture, o bancos de imágenes, como Flickr o Unsplash. Dichas agregaciones pueden ayudar a informar del desarrollo de la colección propia de una institución, ya que los bibliotecarios pueden ver sus contribuciones en el contexto del contenido de otros e identificar los vacíos que puedan desear llenar localmente.⁹⁵

Los agregadores suelen tener diferentes pautas y formatos de entrada. El argumento muy razonable de los agregadores de que no pueden admitir muchas variaciones en los metadatos enviados entra en conflicto con el argumento muy razonable de los contribuidores de que no pueden satisfacer las diferentes necesidades de una amplia variedad de agregadores. La difusión de correcciones o actualizaciones entre la fuente y la agregación puede resultar problemática. La información que pueda haber sido corregida en la cadena que lleva a la incorporación en la agregación no puede ser devuelta a la fuente, por lo que los mismos errores deben corregirse repetidamente. A menudo no está claro qué elementos de datos se han actualizado, cuándo o por quién.

La agregación de imágenes y la agrupación de diferentes imágenes o versiones del mismo objeto fue la meta de los Europeana Innovation Pilots (Programas Piloto de Innovación Europea) de OCLC Research de 2012-2013,⁹⁶ que desarrollaron un método para estructurar jerárquicamente objetos culturales en diferentes niveles de similitud para encontrar “grupos semánticos”, los que incluyen términos con un significado similar. En 2017 OCLC implementó el protocolo Presentation Manifest (Manifiesto de Presentaciones) de la International Interoperability Image Framework (IIIF; Estructura de Imágenes de Interoperabilidad Internacional)⁹⁷ en su sistema de administración de contenido digital CONTENTdm, una agregación que contiene más de 70 millones de registros digitales aportados por más de 2500 bibliotecas de todo el mundo. En 2019 OCLC Research desarrolló un prototipo experimental de IIIF Explorer (Explorador de IIIF) para pruebas y evaluación que busca en todas las imágenes de CONTENTdm utilizando el protocolo Presentation Manifest de IIIF,⁹⁸ como se muestra en la figura 7. La agregación de contenido en los sistemas que cumplen con IIIF puede facilitar la búsqueda y descubrimiento en la gran cantidad de plataformas susodichas que contienen contenido digital.

La transición a la siguiente generación de metadatos

En 2020 OCLC Research lanzó el CONTENTdm Linked Data Pilot (Programa Piloto de Datos Enlazados en CONTENTdm),⁹⁹ centrado en desarrollar métodos y enfoques escalables para producir representaciones de entidades y relaciones que puedan ser leídas por máquinas y para hacer visibles las conexiones que antes eran invisibles. Los existentes metadatos basados en registros se están convirtiendo en datos enlazados al reemplazar cadenas de caracteres con identificadores de ficheros de autoridades conocidos y vocabularios definidos por bibliotecas locales; los gráficos resultantes de entidades y relaciones pueden recuperar información de contexto de fuentes como GeoNames y Wikidata. Este programa piloto, que se completará en agosto de 2020, aborda muchos de los desafíos susodichos identificados por el Focus Group.

El IIF Explorer de OCLC ResearchWorks recupera imágenes de “mapas de París” en colecciones de CONTENTdm

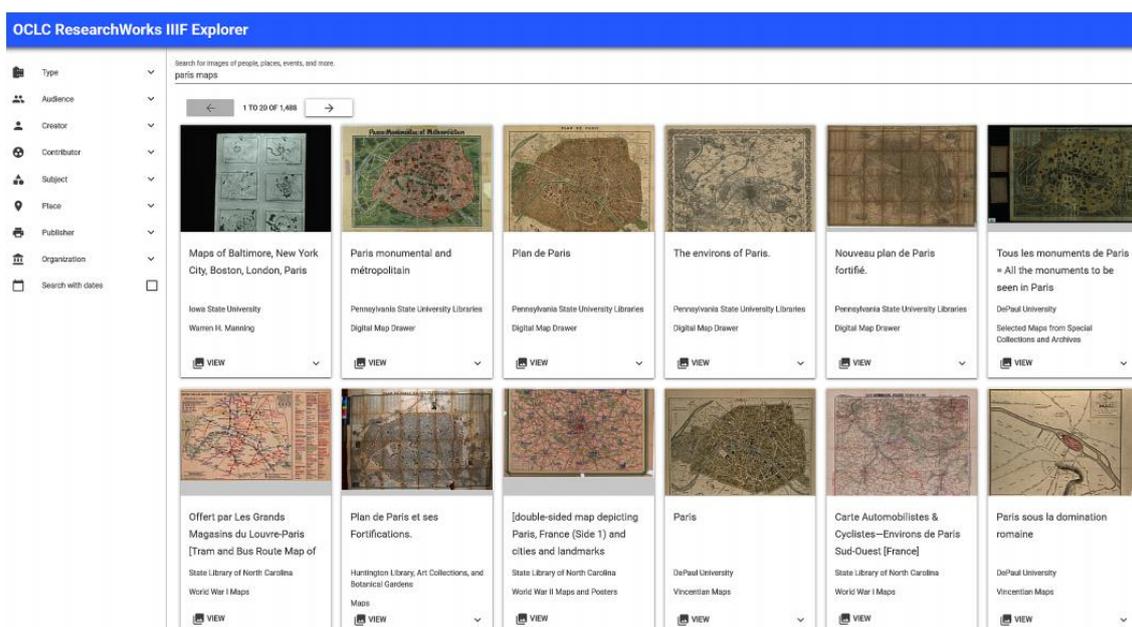


IMAGEN 7. El [IIF Explorer de OCLC ResearchWorks](#) recupera imágenes de “[mapas de París](#)” en colecciones de CONTENTdm

DATOS DE INVESTIGACIÓN

Cuando se subvenciona una investigación se espera que los datos resultantes se guarden y estén disponibles para otros. Las instituciones han asignado mayores recursos para

reunir y conservar este recurso académico, con el fin de su reutilización dentro del archivo académico. La publicación, en dos partes, en el blog de la científica investigadora de OCLC Ixchel Faniel, de la entrada “Data Management and Curation in 21st Century Archives” (“Gestión y conservación de datos en archivos del siglo XXI”; septiembre de 2015)¹⁰⁰ generó un debate entre los miembros del Focus Group sobre los metadatos necesarios para la gestión de datos de investigación.¹⁰¹ Para maximizar las posibilidades de que los metadatos de los datos de investigación se puedan compartir, es decir, se puedan comparar suficientemente, y puedan resultar útiles para quienes estén considerando reutilizar los datos, sería beneficioso para nuestras comunidades compartir ideas y debatir planes que satisfagan las emergentes necesidades de búsqueda y descubrimiento.

Los metadatos son importantes tanto para la búsqueda y descubrimiento como para la reutilización de datos. El informe de OCLC Research de 2016 *Building Blocks: Laying the Foundation for a Research Data Management Program (La construcción de bloques: sentando las bases para un programa de administración de datos de investigación)* señaló:

Los datos son útiles sólo cuando se pueden entender. Animad a los investigadores a proporcionar información estructurada sobre sus datos, proporcionando contexto y significado y permitiendo que otros encuentren, usen y citen adecuadamente los datos. Como mínimo aconsejad a los investigadores que cuenten claramente la historia de cómo agregaron y utilizaron los datos y con qué propósito. Esta información se coloca mejor en un archivo `readme.txt` que incluya información del proyecto y metadatos a nivel de proyecto, así como metadatos sobre los datos en sí (por ejemplo, nombres de archivos, formatos de archivo y software utilizado, título, autor, fecha, financiador, titular de los derechos, descripción, palabras clave, unidad de observación, clase de datos, tipo de datos e idioma).¹⁰²

Los cuatro seminarios web de la serie de 2017-2018 *The Realities of Research Data Management (Las realidades de la gestión de los datos de investigación)*¹⁰³ dirigidos

por Rebecca Bryant, miembro directivo de Programas de OCLC, mencionan la importancia de los metadatos. La infraestructura de la información de investigación recurre a muchas de las fortalezas clave de la profesión bibliotecaria. Los metadatos son fundamentales para nuestro complejo entorno de investigación, comenzando con la planificación que hacen nuestros investigadores antes y durante la creación de datos, continuando con la gestión de los datos, prosiguiendo luego con la difusión de los conocimientos adquiridos y acabando finalmente con la comprensión del impacto, el compromiso y la reputación resultante de las instituciones a las que pertenecemos.¹⁰⁴

La experiencia de las bibliotecas en normas de metadatos, identificadores, datos enlazados y sistemas de intercambio de datos, así como en sistemas técnicos, puede ser de valor incalculable para el ciclo vital de la investigación. Faniel destacó este valor en la publicación del blog *Next* de noviembre de 2019 “Let’s Cook Up Some Metadata Consistency” (“Preparemos algunos elementos uniformes de metadatos”):

Es crucial que la catalogación para la búsqueda y descubrimiento use términos y definiciones que sean uniformes en todos los repositorios, si queremos que los datos y sus metadatos asociados se puedan buscar y descubrir para su reutilización de cualquier forma imaginable. [...] Los bibliotecarios y los archiveros pueden ayudar a crear elementos uniformes en los metadatos que tiendan puentes entre los investigadores y los repositorios, aumentando así en gran medida la búsqueda y descubrimiento, la reutilización y el valor de las inversiones en investigación de sus instituciones.¹⁰⁵

La situación es diferente en cada país. Por ejemplo, nuestros colegas australianos pueden aprovechar la australiana National Computational Infrastructure (Infraestructura Computacional Nacional) para los *big data* y el Australian Data Archive (Archivo Australiano de Datos) para las ciencias sociales.¹⁰⁶ Canadá ha lanzado una red nacional llamada Portage para la “administración compartida de datos de investigación”.¹⁰⁷

Algunas instituciones han desarrollado plantillas para la captura de metadatos de forma estructurada. Algunos miembros del Focus Group señalaron la necesidad de mantener esos formularios lo más simples posible, ya que puede resultar difícil conseguir que los

investigadores los completen. Todos los creadores de datos que se hubieran convenido debían ser la principal fuente de metadatos. Pero, ¿qué inspirará a los creadores de datos a producir metadatos de calidad? Se necesitan nuevas formas de formación y divulgación, un área de exploración dentro del proyecto Research Communications (Comunicaciones de Investigación) de Metadata 2020.¹⁰⁸

Los miembros del Focus Group estuvieron en general de acuerdo en los elementos de datos necesarios para amparar la reutilización: licencias, pasos de procesamiento, herramientas, documentación de datos, definiciones de datos, administrador de datos, números de concesión y datos geospaciales y temporales, cuando sean relevantes. Los esquemas de metadatos utilizados son Dublin Core, MODS (Metadata Object Description Schema; Esquema de Descripción de Objetos de Metadatos) y DDI (la norma de metadatos de la Data Documentation Initiative; Iniciativa de Documentación de Datos). En Reino Unido el Digital Curation Center (Centro de Conservación Digital) proporciona un catálogo enlazado de normas de metadatos.¹⁰⁹ El Metadata Standards Directory Working Group (Grupo de Trabajo del Directorio de Normas de Metadatos) de la Research Data Alliance (Alianza de Datos de Investigación) ha creado un directorio de normas de metadatos mantenido de forma comunitaria para diferentes disciplinas.¹¹⁰ La disparidad de esquemas de metadatos entre disciplinas representa un obstáculo en las opciones de búsqueda y descubrimiento de las instituciones.

Se ha vuelto mucho más reconocida la importancia de los identificadores tanto para los datos de investigación como para los creadores de datos. Se han utilizado DOIs, Handles y ARKs (Archival Resource Key; Clave de Recursos Archivísticos) para proporcionar acceso permanente a colecciones de datos. Los identificadores están disponibles en el nivel de colecciones de datos al completo y para las partes que las componen, y se pueden usar para rastrear descargas y de forma potencial para ayudar a medir el impacto. Tanto ORCID como ISNI se utilizan para identificar a los creadores de datos de forma única y se continúa trabajando en el Research Organizational Registry (Registro de Organizaciones de Investigación) para abordar las afiliaciones institucionales.

Algunos miembros de Focus Group han comenzado a analizar los requisitos de metadatos para el ciclo vital de los datos de investigación, no sólo el producto final, haciendo preguntas como: ¿Quiénes son los colaboradores?¹¹¹ ¿Cómo varios proyectos utilizan diferentes archivos de datos? ¿Qué tipo de herramientas de análisis utilizan? ¿Cuáles son las relaciones de los archivos de datos en un proyecto, entre proyectos relacionados y con otros productos académicos, tales como artículos de revistas relacionados? Los servicios de apoyo a la investigación, como los que se ofrecen en la Universidad de Michigan,¹¹² se están desarrollando para ayudar a los investigadores durante todas las fases del ciclo vital de los datos de investigación, a menudo mediante la colaboración con otras unidades del campus. Entre los problemas más cruciales que se han identificado por los miembros del Focus Group se encuentra el de que los especialistas en metadatos deben participar más en las primeras etapas del ciclo vital de la investigación. Los investigadores deben comprender la importancia de los metadatos en sus planes de gestión de datos. La falta de “gobernanza de metadatos” en una institución hace que la integración de dinámicas de trabajo entre repositorios y opciones de búsqueda y descubrimiento sea problemática.

Algunas bibliotecas han comenzado a proporcionar apoyo en la gestión de datos de investigación de diversas formas. Por ejemplo, los especialistas en metadatos trabajan con el departamento de comunicaciones académicas y publicaciones de sus instituciones, que también administra el repositorio institucional. Estos repositorios institucionales pueden tener tan sólo los registros de “citas bibliográficas” o “sólo metadatos” con un enlace al texto completo o al conjunto de datos que está depositado en un repositorio temático. Se pueden proporcionar “servicios de consulta de metadatos” que sirvan de asesoramiento en el plan de gestión de datos, que incluye normas de metadatos apropiadas y vocabularios controlados, una estrategia para organizar eficazmente sus datos y un enfoque que facilitará la reutilización de los datos años después de que se complete la investigación. La colección de informes *The Realities of Research Data Management (Las realidades de la gestión de datos de investigación)* de OCLC Research clasifica el mantenimiento de metadatos como parte de la función de “habilidades profesionales” y señala algunas variaciones en sus estudios de casos prácticos.¹¹³ En la Universidad de Illinois, campus de Urbana-Champaign, los consultores de metadatos ayudan a los investigadores con los metadatos

independientemente de dónde se depositen los datos de la investigación; la Universidad Monash realiza la conservación de metadatos sólo para depósitos locales.¹¹⁴

La comunicación es clave para que los investigadores comprendan la importancia de los metadatos a lo largo del ciclo vital de la investigación. Algunas universidades ofrecen “*sprints* de investigación” en los que los investigadores se asocian con un equipo de bibliotecarios expertos que pueden abarcar la creación, gestión, análisis y preservación de metadatos. El “Shared BigData Gateway for Research Libraries” (“Portal de *big data* compartidos para bibliotecas de investigación”), organizado por la Universidad de Indiana y subvencionado parcialmente por el Institute of Museum and Library Services (Instituto de Servicios de Museos y Bibliotecas), está desarrollando una plataforma en nube para compartir datos y habilidades profesionales entre instituciones, entre los que se incluyen conjuntos de datos como los registros de la Oficina de Patentes y Marcas de EEUU y del Microsoft Academic Graph.¹¹⁵

La conservación de datos de investigación como parte del currículum académico requiere de nuevas habilidades, entre las que se incluyen el conocimiento más profundo del área y la experiencia con el modelado de datos y el desarrollo de ontologías. Las bibliotecas están invirtiendo más esfuerzos para convertirse en parte del proceso de investigación de su profesorado y están ofreciendo servicios que ayudan a garantizar que sus datos de investigación sean accesibles, además de preservados. Los buenos metadatos ayudarán a guiar a otros investigadores hacia los datos de investigación que necesitan para sus propios proyectos, y los creadores de datos tendrán la satisfacción de saber que sus datos han beneficiado a otros.¹¹⁶

LA EVOLUCIÓN DE “LOS METADATOS COMO SERVICIO”

Los metadatos son la base de la capacidad de buscar y descubrir todos los recursos en las colecciones de adentro hacia afuera y facilitadas. Los miembros del Focus Group prevén una mayor intervención en la creación de metadatos más allá del catálogo bibliotecario tradicional y nuevos servicios que aprovechen tanto los metadatos antiguos como los futuros.

ESTADÍSTICAS

En los objetivos estratégicos de la biblioteca a menudo aparecen frases clave como “fomentar tanto la búsqueda y descubrimiento como el uso”, “enriquecer la experiencia de los usuarios” y “explorar nuevas formas de apoyar todo el ciclo vital de la investigación”; todo lo cual se basa en metadatos de calidad. Las estadísticas de uso, como la de la frecuencia con la que se han prestado, se han citado, se han descargado o se han solicitado ejemplares, se podrían utilizar para crear una amplia variedad de servicios y actividades bibliotecarias. Los miembros del Focus Group identificaron algunos posibles servicios: apoyar las decisiones de gestión de colecciones sobre los proyectos de expurgo y la identificación de materiales para su almacenamiento externo; evaluar las suscripciones; comparar las citas que aparecen en las publicaciones de sus investigadores con lo que la biblioteca no está adquiriendo; y mejorar la clasificación de relevancia, personalizar los resultados de búsqueda, ofrecer servicios de recomendaciones en la capa de visualización y medir el impacto del uso de la biblioteca en el éxito investigador o estudiantil o el análisis de aprendizaje.¹¹⁷ La Universidad de Minnesota realizó un estudio para investigar las relaciones entre el uso de estudiantes de grado de primer año de la biblioteca universitaria, su rendimiento académico y su retención de conocimientos.¹¹⁸ Los resultados sugieren una fuerte correlación entre el uso de los servicios y recursos de la biblioteca universitaria (en particular los inicios de sesión en bases de datos, préstamos de libros, inicios de sesión en revistas electrónicas e inicios de sesión en estaciones de trabajo de la biblioteca) y los promedios más altos de calificaciones. En Reino Unido el Library Impact Data Project (Proyecto de Datos de Impacto Bibliotecario) de Jisc llegó a una correlación similar.¹¹⁹

ASESORAMIENTO

El valor de los metadatos se demuestra al integrarlos en la estructura tanto de la biblioteca como de otras unidades del campus. Por ejemplo, los especialistas en metadatos pueden proporcionar “metadatos como servicio”: asesoramiento en las primeras etapas de proyectos bibliotecarios y de investigación.¹²⁰ Una tendencia emergente es que los departamentos de humanidades digitales soliciten consejo de especialistas en metadatos sobre normas de metadatos y cómo utilizar vocabularios controlados. Se observa cada vez más esta función de consultor de metadatos en las recientes ofertas de trabajo de las bibliotecas. En una oferta de trabajo de bibliotecario de metadatos de la Universidad de Cornell,¹²¹ uno de los cometidos mencionados fue 20% para “divulgación y consulta de metadatos”: “mantiene estrechas relaciones de trabajo y se comunica regularmente con el personal de Cornell, fomentando los esfuerzos de colaboración entre Servicios de Metadatos y el resto de la comunidad universitaria de Cornell”. El Instituto Tecnológico de Georgia quiere contratar a un bibliotecario de metadatos que “servirá como consultor de metadatos para proyectos e iniciativas bibliotecarias superiores. Trabajará en estrecha colaboración con otros departamentos de la Biblioteca, las bibliotecas de la Universidad de Emory, GALILEO, las bibliotecas del Sistema Universitario de Georgia y otros miembros involucrados en proyectos conjuntos.”¹²²

NUEVAS APLICACIONES

El uso compartido y consistente de los campos MARC permite nuevas aplicaciones. Actualmente las bibliotecas usan identificadores en registros bibliográficos para recuperar índices, resúmenes, reseñas e imágenes de portada y para generar mapas de pisos para la localización de recursos en un cuerpo de estanterías de clasificación específico, como en la integración de OCLC con StackMap.¹²³ Los metadatos bibliográficos se utilizan para completar los sistemas de gestión de recursos digitales y los repositorios institucionales y, con herramientas como Tableau y OpenRefine, pueden permitir un análisis más completo de las colecciones y una perspectiva de las colecciones. Los metadatos de MARC están conectando a los académicos con los datos bibliográficos para sus proyectos y pueden generar relaciones entre recursos relacionados y aplicaciones como Yewno.¹²⁴ Los metadatos de MARC también se están utilizando para servir de apoyo a los índices de rendimiento institucionales y al control

del organismo y sirven como fuente para construir historiales de la organización. La procedencia implícita en los metadatos bibliográficos de una institución ha demostrado ser útil para documentar casos de robo. El análisis de los datos del catálogo mediante la minería de datos también se puede utilizar para enriquecer los metadatos, como generar códigos de idioma que faltan en registros relacionados o identificar los títulos originales de obras traducidas. Los datos de MARC también han permitido la generación de mapas temáticos para buscar y descubrir relaciones que de otro modo no resultarían explícitas en los metadatos de catalogación.¹²⁵

Las búsquedas y descubrimientos constituyen otro tipo de servicio de metadatos. Un ejemplo sorprendente es el del Auslang Nacional Codeathon celebrado en 2019, una colaboración entre la National Library of Australia (Biblioteca Nacional de Australia), el Australian Institute of Aboriginal and Torres Strait Islander Studies (Instituto Australiano de Estudios Aborígenes e Islámicos del Estrecho de Torres), Trove, Libraries Australia y las bibliotecas estatales y territoriales: un maratón de códigos (*code-a-thon*) nacional para identificar elementos en idiomas indígenas australianos.¹²⁶ La imagen 8 muestra los resultados, un mapa que indica las 465 lenguas indígenas de la Australian National Bibliographic Database (Base Bibliográfica Nacional Australiana de Datos) etiquetadas como resultado del *code-a-thon*, y un ejemplo de participación de la comunidad para mejorar los metadatos bibliográficos.

Distribución de los 465 códigos de lenguas indígenas de la Australian National Bibliographic Database

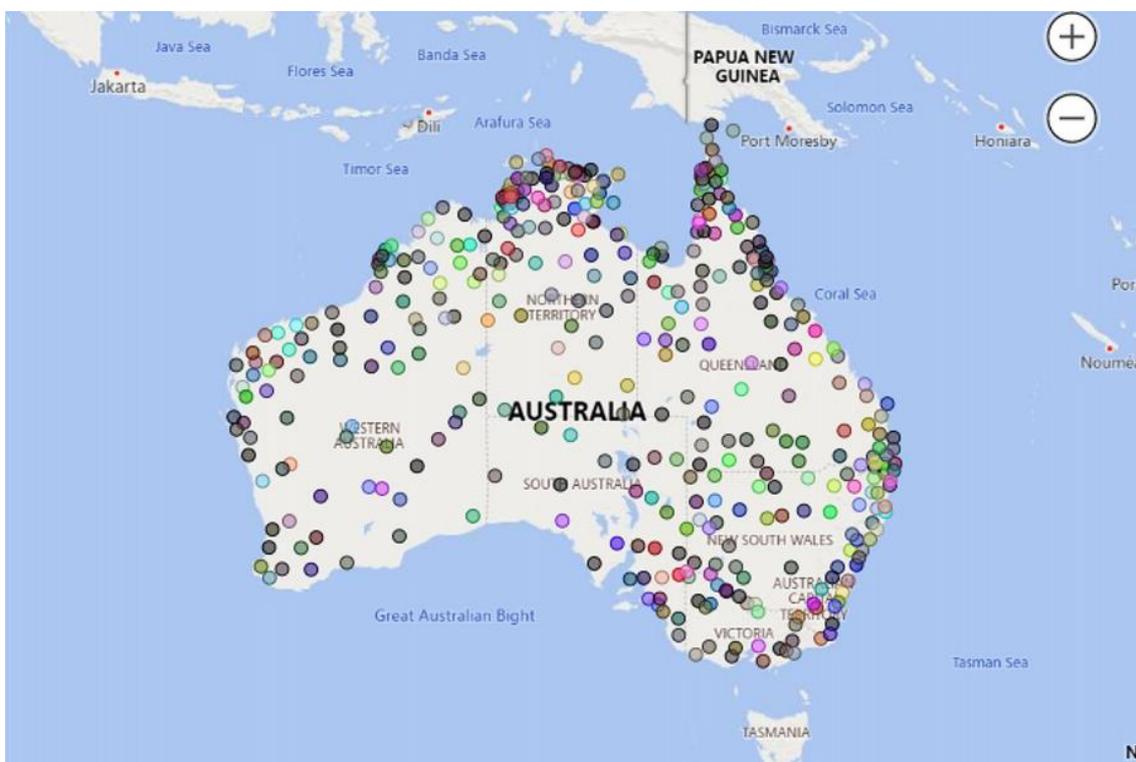


IMAGEN 8. Distribución de los 465 códigos de lenguas indígenas de la [Australian National Bibliographic Database](#) procedentes del Austlang National Codeathon

BIBLIOMETRÍA

Los metadatos de la biblioteca también se utilizan para generar datos bibliométricos, métodos estadísticos para analizar libros, artículos y otras publicaciones. Tiene mucho potencial el uso de metadatos de bibliotecas para proyectos de investigación de humanidades digitales. Por ejemplo, un investigador de la Library of Congress utilizó metadatos bibliográficos para rastrear la historia de las publicaciones y los derechos de autor; los investigadores de la UCLA han utilizado metadatos de catalogación para rastrear la comercialización de inventos como la insulina.

Un uso novedoso de los metadatos de catalogación fue realizado por Hachette UK, la segunda librería más grande del Reino Unido, que encargó a The Graphic History Company (La Compañía de Historia Gráfica) que le revelara los historiales de las nueve editoriales de Hachette y los entretajara en un relato coherente, al preguntar a la British Library por cada autor y cada título de libro publicado por sus nueve editoriales a lo

largo de 250 años. La British Library proporcionó una lista de más de 55000 autores, de los cuales se seleccionaron 5000 de los más destacados para crear quizás el ejemplo más hermoso de uso de metadatos: un mural gigantesco de ocho pisos con los 5000 autores en orden cronológico (la imagen 9 muestra una parte del mural; para ver más imágenes del mural, véase River of Authors (Río de Autores) de Hachette).¹²⁷

“River of Authors” (“El río de autores”) de Hachette UK generado a partir de los metadatos del catálogo de la British Library



IMAGEN 9. “[River of Authors](#)” (“Río de Autores”) de Hachette UK generado a partir de los metadatos del catálogo de la British Library

INDIZACIÓN SEMÁNTICA

Cuando los vocabularios controlados y los tesauros se convierten en datos abiertos enlazados y se comparten públicamente, se desvanecerá su función tradicional de facilitar la exploración de colecciones, pero podría encontrar un propósito renovado dentro de los sistemas de organización del conocimiento en la web.¹²⁸ Como señala Marcia Zeng en “Knowledge Organization Systems (KOS) in the Semantic Web: a multidimensional review” (“Sistemas de organización del conocimiento en la web semántica: una revisión multidimensional”),

un vocabulario de sistemas de organización del conocimiento es más que la fuente de valores que se utilizará en las descripciones de metadatos: al modelar las subyacentes estructuras semánticas de un dominio, los sistemas de organización del conocimiento actúan como mapas de ruta semánticos y hacen posible una orientación común para los indexadores y los futuros usuarios, ya sean humanos o máquinas.¹²⁹

Buenos ejemplos de dicha reutilización son los vocabularios de Getty que permiten navegar por la representación del conocimiento de Getty y también ayudan a los usuarios a generar sus propias consultas SPARQL que pueden integrarse en aplicaciones externas. Otro ejemplo es Social Networks and Archival Context (SNAC; Redes Sociales y Contexto Archivístico),¹³⁰ que permite navegar por entidades y relaciones independientemente de sus colecciones de origen. En tales casos la herramienta de búsqueda y descubrimiento pivota para centrarse en la persona, en la familia o en la materia en lugar de (sólo) centrarse en la colección.

En lugar de un “dominio único” los especialistas en metadatos podrían proporcionar un valor añadido al añadir puentes desde los metadatos de las bases de datos del ámbito bibliotecario a otros ámbitos. Wikidata es un ejemplo de una plataforma que agrega entidades de diferentes fuentes y las enlaza a más detalles en wikipedias de diferentes idiomas. Algunas instituciones han contratado a wikimedistas residentes para acelerar este proceso.

Los miembros del Focus Group esperan que la inteligencia artificial, o al menos el aprendizaje mecánico, pueda reducir el actual esfuerzo manual para enlazar nombres y conceptos en los datos de investigación. Quizás se podrían usar algoritmos para unir nombres basados en metadatos o fuentes relacionadas, relacionar temas entre sí en función del contexto, eliminar la ambigüedad de nombres basados en otros metadatos disponibles y analizar conjuntos de datos para identificar posibles sesgos en una colección.¹³¹ Algunos miembros de la Research Library Partnership participan en Artificial Intelligence for Libraries, Archives & Museums (AI4LAM; Inteligencia Artificial para Bibliotecas, Archivos y Museos),¹³² una “comunidad participativa

internacional centrada en promover el uso de la inteligencia artificial en, para y por bibliotecas, archivos y museos”.¹³³ Algunas recomendaciones de alto nivel sobre cómo mejorar las descripciones a medida y optimizar la búsqueda y descubrimiento se encuentran en el documento de posición de OCLC Research de 2019 de Thomas Padilla *Responsible Operations: Data Science, Machine Learning, and AI in Libraries* (Operaciones responsables: la ciencia de los datos, el aprendizaje mecánico y la inteligencia artificial en las bibliotecas).¹³⁴

Los preparativos para los requisitos del personal del futuro

Los cambios previstos a partir de la transición a la próxima generación de metadatos también variarán los requisitos del personal con el fin de prepararse para el futuro. Los miembros del Focus Group identificaron nuevas habilidades que son necesarias tanto para los nuevos profesionales como para los catalogadores experimentados, impulsadas por el panorama cambiante de la tecnología de la información y por el aumento del desgaste del personal. Los miembros del Focus Group caracterizaron a los profesionales como “los pioneros de las innovaciones”, que luego se hacen rutinarias para los no profesionales. Estos debates refuerzan las recomendaciones de Padilla sobre la investigación de las competencias básicas, el compromiso con el talento interno y la expansión de la formación empírica.¹³⁵

EL CAMBIO CULTURAL

Los miembros del Focus Group informaron sobre un delicado equilibrio entre la asignación de personal a “actividades de catalogación tradicionales” (como la catalogación original y por copia, el trabajo de autoridades) con proyectos de I+D más exploratorios, como proyectos de datos enlazados, exploración de nuevos modelos de datos y tecnologías como Wikidata y el aprendizaje de normas e identificadores emergentes. Se necesita un **cambio cultural**, que vaya desde la satisfacción por la producción sin más hacia la valoración de oportunidades de aprender, explorar y probar nuevos enfoques para el trabajo de metadatos. Los especialistas en metadatos deben comprender que mejorar *todos* los metadatos es más importante que las cifras de productividad de cualquier individuo. Este cambio cultural requiere de la participación de los directivos para que apoyen los programas de formación con el fin de que el personal aprenda nuevas dinámicas de trabajo para procesar varios formatos y vea a los especialistas en metadatos como algo más que “máquinas de producción”.

Los gestores de metadatos que se enfrentan a reducciones de personal a la vez de que se espera que mantengan niveles de producción deben justificar la asignación de tiempo del personal para I+D, o “tiempo de recreo”, para explorar preguntas del tipo: ¿Qué podemos dejar de hacer? ¿Qué cosa aprendiste y que todos necesitamos hacer más? ¿Qué necesitas para seguir adelante? ¿Qué software de código abierto podría ayudarnos a hacer el trabajo de manera más eficiente? ¿Qué nuevos métodos podrían mejorar la

búsqueda y descubrimiento, el acceso y el uso de nuestras colecciones facilitadas? Los gestores deben incorporar metas para el éxito que no se basen únicamente en números.¹³⁶

Entre los indicios de este cambio cultural se encuentran las instituciones que subcontratan parte del trabajo de metadatos o forman al personal de apoyo para crear metadatos para las “cosas más fáciles”, al tiempo que exigen que los catalogadores sólo hagan lo que sólo los humanos bien formados pueden hacer. Los gestores de metadatos podrían analizar aquellos materiales que requieren de metadatos que el personal de apoyo o los estudiantes puedan manejar, proporcionando plantillas cuando sea posible. Si eliminamos estas tareas, la mayoría de lo que queda requiere de especialistas en metadatos altamente formados con experiencia en idiomas, formatos físicos, y desambiguación y descripción de personas, organizaciones y otras entidades.

OPORTUNIDADES DE APRENDIZAJE

Para alentar el cambio cultural entre los especialistas en metadatos con el fin de cambiar su forma de pensar sobre cómo trabajan y de estimular su interés en las oportunidades de aprendizaje, los miembros del Focus Group han utilizado varios planteamientos:

- Identificar quiénes de su equipo tienen la aptitud de adquirir nuevas habilidades. En una institución una integrante del personal compartió lo que había aprendido y toda la unidad se “animó” a causa de lo que ella aportó a sus colegas. Esto creó reconocimiento a favor del “aprendizaje continuo” y el personal presentó sus actividades en conferencias nacionales.
- Convocar debates grupales entre equipos para analizar los metadatos problemáticos y hallar soluciones, alentando al personal a avanzar juntos. El personal menos interesado en las nuevas habilidades puede asumir parte de la producción de aquellos que aprendan nuevas habilidades y sean menos productivos.
- Lanzar “clubes de lectura” donde todo el personal lea un artículo y responda a tres preguntas de debate para inspirar a los especialistas en metadatos a pensar en cuestiones más amplias de metadatos fuera de su trabajo diario.

- Organizar semanalmente para el personal almuerzos informales para ver videos sobre nuevos desarrollos, como los datos enlazados, para que el personal pueda “ver y aprender” juntos.
- Participar en proyectos multiinstitucionales para colaborar con compañeros para resolver problemas e intercambiar ideas.
- Fomentar la participación en conferencias profesionales y desarrollo de normas.

La educación y la formación de los catalogadores han estado en la primera línea de muchos debates en la comunidad de metadatos. Tanto los nuevos profesionales como los catalogadores experimentados necesitan nuevas habilidades para realizar una transición exitosa al entorno emergente de datos enlazados. Los catalogadores están aprendiendo y experimentando con BIBFRAME sin dejar de ser responsables del control bibliográfico tradicional de las colecciones. Los especialistas en metadatos utilizan herramientas para la representación, corrección y mejora de metadatos. Identifican y representan relaciones semánticas entre una variedad de taxonomías para hacer que varios tesauros sean inteligibles para los usuarios finales. Para los aspectos más técnicos de la gestión de metadatos ha ido en aumento la competición por el talento por parte de otros sectores. Esto puede intensificarse a medida que los metadatos se vuelvan más importantes para diversas áreas de la Administración, las organizaciones sin ánimo de lucro y la empresa privada.¹³⁷

NUEVAS HERRAMIENTAS Y HABILIDADES

En cada institución varía el alcance de la colaboración de los especialistas en metadatos con el personal de tecnología de la información o de sistemas. Esta colaboración es necesaria por muchas razones, incluida la gestión de datos que se encuentran fuera del control de la biblioteca. Hay quienes señalaron que existen “diferencias culturales” entre las profesiones: los desarrolladores tienden a ser más dinámicos y se centran en la creación rápida de prototipos y en la reiteración, mientras que los bibliotecarios se centran primero en documentar lo que se necesita y son más “esquemáticos”. ¿Qué tiene más probabilidades de tener éxito? ¿Enseñar habilidades de tecnología de la información a especialistas en metadatos o enseñar principios de metadatos al personal de tecnología de la información? El “santo grial” es contratar a alguien con experiencia en tecnología de la información interesado en servicios de metadatos. Es difícil retener

al personal con habilidades de tecnología de la información; son demandados para trabajos mejor pagados en el sector privado. Las experiencias de los miembros del Focus Group han demostrado que es más fácil para los bibliotecarios aprender habilidades de programación que contratar especialistas en tecnología de la información para aprender la “mentalidad de los servicios técnicos”. Lo ideal para los miembros del Focus Group sería contar con algunos miembros del personal que tengan habilidades técnicas de realizar acciones por lotes en los datos, o al menos que sepan cómo utilizar las herramientas externas disponibles para automatizar tantas tareas como sea posible.

Durante muchos años los miembros del Focus Group han estado utilizando MarcEdit y/u otras herramientas como OpenRefine, lenguajes de programación (por ejemplo, Python, Ruby o Perl) y macros para la reconciliación de metadatos y el procesamiento por lotes.¹³⁸ MarcEdit es la herramienta más utilizada y tiene una comunidad de usuarios grande, global y activa, como se indica en su resumen de uso de 2017.¹³⁹ Terry Reese, desarrollador de MarcEdit, estima que alrededor de un tercio de todos los usuarios trabajan en entornos que no son MARC y dos tercios de los usuarios más activos son miembros de OCLC. Los miembros del Focus Group informaron que utilizan MarcEdit para la transformación de datos, la mejora de los registros de proveedores, la creación de registros MARC a partir de hojas de cálculo, la reconciliación de datos enlazados, la eliminación de registros duplicados en un archivo, la fusión de dos o más registros en uno, la recolección de Z39.50 y la reconciliación de metadatos antes de enviar registros a otros sistemas. MarcEdit 7, la versión de 2017, incluye nuevas características, como la función básica de agrupamiento, que proporciona una manera potente de encontrar relaciones entre datos sin introducir una gran curva de aprendizaje. También tiene mecanismos que permiten datos enlazados.¹⁴⁰ Reese ha creado una serie de tutoriales de YouTube que están disponibles en su lista de reproducción MarcEdit.¹⁴¹

Los gestores quieren centrarse menos en esquemas concretos y más en principios de metadatos que se puedan aplicar a una variedad de formatos y entornos diferentes. Serían deseables las siguientes habilidades blandas: la resolución de problemas, la colaboración eficaz, la voluntad, incluso el entusiasmo, de probar cosas nuevas, la comprensión de las necesidades de los investigadores y la recomendación. Aunque algunos especialistas en metadatos siempre han disfrutado de experimentar con nuevos

métodos, a menudo carecen de tiempo para aprender nuevas herramientas o metodologías mientras se mantienen al día con sus rutinarios cometidos de trabajo. Las bibliotecas deben promover los metadatos como una opción de carrera emocionante para los nuevos profesionales en lugares como escuelas bibliotecarias y la New Members Roundtable (Mesa Redonda de Nuevos Miembros) de ALA. Puede aumentar su atractivo haciendo hincapié en que los metadatos abarcan mucho más que la catalogación de bibliotecas (identificación de entidades; normas de descripción utilizadas en diversas disciplinas académicas; la descripción de datos de formato digital original, de archivo y de investigación que pueden interactuar con la web semántica). Como señaló un miembro del Focus Group, “traemos orden a partir de la nada”.¹⁴²

AUTOFORMACIÓN

Los metadatos se crean cada vez más fuera de la biblioteca por profesores y estudiantes con una formación mínima, lo que lleva a la necesidad de más catalogadores con habilidades de mantenimiento de registros. Los miembros del Focus Group señalaron la necesidad de habilidades técnicas como la programación simple, la corrección de datos y la gestión de identidades para unir las equivalentes en varios registros. Las fuentes de formación que se mencionan con frecuencia son Library Juice Academy, tutoriales de MarcEdit, LinkedIn Learning (adquirida por Lynda.com), seminarios web de formación de la Library of Congress, seminarios web de ALCTS, Code Academy, Software Carpentry y conferencias como Code4Lib y Mashcat.¹⁴³ Se recomendó la lectura de *Data on the Web Best Practices (Las mejores prácticas de datos en la web)* y *Semantic Web for the Working Ontologist (La web semántica para el que trabaja de ontólogo)* de W3C.¹⁴⁴ Para el éxito de dicha formación es fundamental la capacidad de aplicar rápidamente lo aprendido. Si no se aplican las nuevas habilidades, la gente olvida lo que ha aprendido. El personal se siente frustrado cuando ha invertido tiempo para aprender algo que no puede usar en su trabajo cotidiano.

Los miembros del Focus Group han visto un gran cambio de depender de las formaciones de la Library of Congress a la autoformación a partir de varias fuentes. Algunos planteamientos mencionados por los participantes son:

- Acentuar la continuidad de los principios de metadatos al introducir un mayor alcance de trabajo.
- Aprovechar el Library Workflow Exchange (Intercambio de Dinámicas de Trabajo Bibliotecarias),¹⁴⁵ un sitio web diseñado para ayudar a los bibliotecarios a compartir dinámicas de trabajo y mejores prácticas entre instituciones, incluidos los *scripts*.
- Procedente de la Electronic Resources and Libraries Conference (Conferencia de Bibliotecas y Recursos Electrónicos) de 2017: “No esperes; ¡reitera!” En otras palabras, en lugar de esperar hasta que el personal tenga todas las habilidades necesarias, déjalos hacer las tareas de forma reiterativa, aprendiendo sobre la marcha, para que estén preparados para nuevas tareas cuando llegue el momento.
- Tener grupos pequeños de especialistas en metadatos que tomen cursos de programación juntos, después de lo cual pueden continuar reuniéndose y debatir formas de aplicar sus nuevas habilidades para automatizar tareas rutinarias.
- Alentar al personal a participar en eventos como los seminarios web DevConnect de OCLC¹⁴⁶ para aprender de las bibliotecas utilizando las APIs de OCLC para mejorar sus operaciones y servicios bibliotecarios.
- Crear grupos de lectura y estudio que incluyan personal de todo el campus o de todas las divisiones.
- Ampliar el alcance del trabajo actual para permitir que los especialistas en metadatos apliquen sus habilidades a nuevos ámbitos o terminología, como el uso de Dublin Core para colecciones digitales. Involucrar al personal en proyectos digitales desde la etapa conceptual hasta el desarrollo de descripciones del proyecto, prácticas de control de la calidad y selección de herramientas. Como beneficio adicional esto fomenta las relaciones colaborativas de trabajo en equipo.
- Contratar a estudiantes graduados en ciencias informáticas para tareas a corto plazo, como crear *scripts*.

EL TRATAMIENTO DE LA ROTACIÓN DE PERSONAL

La rotación de un puesto profesional dentro de una unidad de catalogación o metadatos ahora conlleva el significativo riesgo de que puede ser imposible convencer a los directivos de que mantengan el puesto en la unidad y lo vuelvan a cubrir. Esto es

particularmente así cuando el titular que se va realizó una alta proporción de trabajo “tradicional”, como la catalogación original en MARC. Las probabilidades de conservar el puesto son mucho mayores si se analiza con detenimiento cómo podría reconfigurarse o reorientarse el puesto para satisfacer las necesidades emergentes.¹⁴⁷

La mayoría de los miembros del Focus Group han tenido que hacer frente a diferentes cifras de rotaciones, ya sea por jubilaciones o por cambiarse el personal a otros puestos. La mitad de ellos necesitaba rediseñar los puestos de los bibliotecarios que se iban. Mirar las ofertas de trabajo de otras instituciones ayuda a crear una descripción del puesto que sea atractiva. Muchos puestos de catalogación no requieren de un título de máster en Biblioteconomía y Ciencias de la Información, por lo que la contratación de profesionales se ha centrado en la adaptabilidad, alineando los nuevos puestos con las prioridades de la universidad, y en el afán por aprender y tomar la iniciativa en áreas como metadatos para resultados de investigación, acceso abierto, colecciones digitales y datos enlazados. El trazado de estrategias futuras y el diseño de formas de hacer que los metadatos interoperen entre sistemas han sido elementos de incorporaciones recientes. Se busca nuevo personal con habilidades de programación, ya que pueden aplicar técnicas de lotes a los metadatos que pueden compensar la pérdida de personal. El uso de la tecnología a favor del servicio bibliotecario ayuda a los catalogadores a “hacer más con menos”.

Los miembros del Focus Group quieren que el nuevo personal conozca tanto la comunidad de catalogación compartida como los cometidos coincidentes con otras organizaciones del patrimonio cultural, como los archivos y los museos. El ámbito bibliotecario sigue evolucionando y los bibliotecarios han tenido que reflexionar sobre sus prioridades para avanzar. Los gestores de metadatos deben repensar las funciones de los especialistas en metadatos más allá del trabajo de catalogación “tradicional”. Los candidatos potenciales que tienen habilidades más adaptables se han vuelto más interesantes que aquellos que tienen una formación tradicional de catalogación y que, por ello, posiblemente no se adapten bien a trabajar en nuevos entornos. Es posible que sea necesario reescribir y reestructurar muchas funciones y descripciones de catalogación. Quizás las únicas actividades que se mantengan para siempre entre las tareas profesionales sean la gestión, la búsqueda de nuevas tendencias, la elaboración de

estrategias, la participación en nuevas normas internacionales, la iniciativa y la implementación de cambios, y el pensar en el panorama general.

Consecuencias

La próxima generación de metadatos se centrará aún más en las entidades que en las descripciones basadas en registros de las colecciones de una institución. Las actividades de datos enlazados de los miembros del Focus Group, entre las que se incluye su participación en los pilotos Project Passage y CONTENTdm Linked Data de OCLC Research, contribuyeron a que OCLC obtuviera la subvención económica Andrew W. Mellon para su proyecto bianual Shared Entity Management Infrastructure (Infraestructura de Gestión de Entidades Compartidas),¹⁴⁸ lanzado en enero de 2020. Once de los miembros del Grupo Asesor de la Shared Entity Management Infrastructure también son miembros del Focus Group. El proyecto se basa en el trabajo de datos enlazados de OCLC Research y proporcionará una *infraestructura de producción* con identificadores de autoridades permanentes para personas y obras. Se basará en gran medida en API, lo que permitirá a los bibliotecarios adaptar a su medida sus dinámicas de trabajo en torno a la infraestructura de datos enlazados. Esta infraestructura ha sido anhelada por los miembros del Focus Group desde hace mucho tiempo, ya que abordará muchos de los desafíos mencionados anteriormente en torno a los identificadores permanentes, especialmente los identificadores de “obras”.

Los identificadores de autoridades permanentes aportados por la Shared Entity Management Infrastructure proporcionarán los enlaces independientes del lenguaje que se necesitan a fuentes fiables. Los metadatos que las bibliotecas, los archivos y otras instituciones de patrimonio cultural han creado y crearán proporcionarán el contexto para estas entidades, como “menciones” asociadas con esos enlaces. Las consecuencias serán globales, afectando a la forma en que los bibliotecarios y archiveros describirán las colecciones de adentro hacia afuera y facilitadas, inspirando nuevas ofertas de “metadatos como servicio” e influyendo en los requisitos del personal del futuro.

Agradecimientos

OCLC Research desea agradecer a todos los miembros del Research Library Partners Metadata Managers Focus Group que han compartido sus experiencias y pensamientos que se resumen aquí. Además, agradecemos al Metadata Managers Planning Group, que inició los temas y proporcionó las declaraciones de contextualización y la batería de preguntas, cuyas respuestas sirvieron de base para nuestros debates. Además, apreciamos particularmente los comentarios en profundidad de los siguientes miembros del Focus Group, que revisaron una versión previa de este documento; sus comentarios mejoraron este ensayo de síntesis.

- Charlene Chou, Universidad de Nueva York
- Suzanne Pilsk, Instituto Smithsonian
- Greg Reeve, Universidad Brigham Young
- Alexander Whelan, Universidad de Columbia
- Helen K. R. Williams, Escuela de Economía de Londres

También doy las gracias a los compañeros actuales y anteriores de OCLC: Rebecca Bryant, Jody DeRidder, Annette Dortmund, Rachel Frick, Janifer Gatenby, Jean Godby, Shane Huddleston, Andrew Pace, Merrilee Proffitt, Nathan Putnam, Stephan Schindehette y Chela Weber por su cuidadosa revisión total o parcial de las versiones previas de este documento. Gracias a Erica Melko por su edición, a Jeanette McNicol por el diseño de este informe y a JD Shipengrover por la portada artística.

A título personal me he beneficiado enormemente de mis interacciones con el OCLC Research Partners Metadata Managers Focus Group y me ha encantado desempeñar un pequeño papel en esta transición a la próxima generación de metadatos.

Apéndice

OCLC Research Library Partners Metadata Managers Planning Group

2015-2020

Los miembros del Planning Group seleccionaron los temas para los debates de OCLC Research Library Partners Metadata Managers, redactaron las declaraciones de contextualización de por qué el tema era importante y oportuno y desarrollaron la batería de preguntas a las que respondieron los miembros del Focus Group. Los que iniciaron cada tema en el Planning Group también revisaron los borradores de los resúmenes que luego se publicaron en el blog *Hanging Together* de OCLC Research.

Los actuales miembros del Planning Group están resaltados en negrita; las afiliaciones institucionales que aparecen son las que tenían en el momento en que formaron parte del Planning Group:

- **Jennifer Baxmeyer, Universidad de Princeton**
- Sharon Farnel, Universidad de Alberta
- Steven Folsom, Universidad de Harvard y Universidad Cornell
- **Erin Grant, Universidad de Washington**
- Dawn Hale, Universidad Johns Hopkins
- Myung-Ja Han, Universidad de Illinois, campus de Urbana-Champaign
- Kate Harcourt, Universidad de Columbia
- Corey Harper, Universidad de Nueva York
- **Stephen Hearn, Universidad de Minnesota**
- **Daniel Lovins, Universidad de Yale**
- **Roxanne Missingham, Universidad Nacional Australiana**
- Chew Chiat Naun, Universidad Cornell y Universidad de Harvard
- **Suzanne Pilsk, Instituto Smithsonian**
- **John Riemer, Universidad de California, campus de Los Ángeles**
- Carlen Ruschoff, Universidad de Maryland
- Philip Schreur, Universidad Stanford
- Jackie Shieh, Universidad George Washington
- **Melanie Wacker, Universidad de Columbia**

Notas

¹ OCLC Research Library Partnership Metadata Managers Focus Group.
<https://www.oclc.org/research/areas/data-science/metadata-managers.html>.

² OCLC Research. “The OCLC Research Library Partnership”.
<https://www.oclc.org/research/partnership.html>.

³ Smith-Yoshimura. 2017. “Metadata Advocacy”, *Hanging Together: the OCLC Research Blog*, 17 de octubre de 2017. <https://hangingtogether.org/?p=6282>.

⁴ British Library. 2019. *Foundations for the Future: The British Library’s Collection Metadata Strategy 2019-2023*. London: British Library.
<https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2019-2023.pdf>.

⁵ Ibid, 4.

⁶ Datos a 1 de junio de 2020.

⁷ Library of Congress. “Program for Cooperative Cataloging”.
<https://www.loc.gov/aba/pcc/>.

⁸ Excepto en junio de 2020, cuando todos los debates se desarrollaron tan sólo virtualmente a causa de la pandemia de la COVID-19.

⁹ Véase *Hanging Together: The OCLC Research Blog*, categoría de búsqueda “Metadata”.
<https://hangingtogether.org/?cat=81>.

¹⁰ Los beneficios de afiliarse a la RLP son citados en Smith-Yoshimura. 2018. “What Metadata Managers Expect from and Value about the Research Library Partnership,” *Hanging Together: The OCLC Research Blog*, 16 de abril de 2018.

<https://hangingtogether.org/?p=6683>.

¹¹ Los análisis de las tres International Linked Data Surveys for Implementers de 2014-2018 y la hoja de cálculo de todas las respuestas a las encuestas están disponibles. Véase OCLC Research. 2020. “Linked Data.” International Linked Data Survey.

<https://www.oclc.org/research/themes/data-science/linkedata/linked-data-survey.html>.

¹² Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Holly Tomren y Craig Thomas. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/faq3-ax08>;

OCLC Research. 2020. “CONTENTdm Linked Data pilot.”

<https://www.oclc.org/research/themes/data-science/linkedata/contentdm-linked-data-pilot.html>;

OCLC. 2020. “WorldCat®: OCLC and Linked Data.” Shared Entity Management Infrastructure.

<https://www.oclc.org/en/worldcat/linked-data/shared-entity-management-infrastructure.html>;

Library of Congress. “BIBFRAME.” Bibliographic Framework Initiative.

<https://www.loc.gov/bibframe/>;

Futornick, Michelle. 2019. “LD4P2 Linked Data for Production: Pathway to Implementation.” LS4P2 Project Background and Goals. Lyasis. Publicado el 14 de enero de 2019.

<https://wiki.lyasis.org/display/LD4P2/LD4P2+Project+Background+and+Goals>;

Share-VDE (Share Virtual Discovery Environment). “An Effective Environment for the Use of Linked Data by Libraries.” Consultado el 17 de septiembre de 2019.

<https://www.share-vde.org/sharevde/clusters?l=en>;

Casalini, Michele, Chiat Naun Chew, Chad Cluff, Michelle Durocher, Steven Folsom, Paul Frank, Janifer Gatenby, Jean Godby, Jason Kovari, Nancy Lorimer, Clifford Lynch, Peter Murray, Jeremy Myntti, Anna Neatrou, Cory Nimer, Suzanne Pilsk, Daniel Pitti, Isabel Quintana, Jing Wang y Simeon Warner. 2018. *National Strategy for*

Shareable Local Name Authorities National Forum: White Paper. Ithaka, New York: Cornell University Library *eCommons digital repository*.

<https://hdl.handle.net/1813/56343>.

¹³ Library of Congress. 2019. PCC (Program for Cooperative Cataloging) Task Group on Linked Data Best Practices. 2019. PCC Task Group on Linked Data Best Practices Final Report: Submitted to PCC Policy Committee 12 September 2019. Washington DC: Library of Congress.

<https://www.loc.gov/aba/pcc/taskgroup/linked-data-best-practices-final-report.pdf>;

Library of Congress. 2018. “Charge for PCC Task Group on Identity Management in NACO,” 5. American Bar Association, Program for Cooperative Cataloging, revisado el 22 de mayo de 2018.

<https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf>;

Library of Congress. 2020 “PCC Task Group on URIs in MARC.” Programs of the PCC. Charge. Consultado el 19 de septiembre de 2020.

<https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html>;

Library of Congress. 2018. “PCC Linked Data Advisory Committee: Linked Data Advisory Committee Charge.” PCC Task Groups 2018. Task Groups. Revisado el 24 de julio de 2018. [Documento Word; 28KB].

<https://www.loc.gov/aba/pcc/taskgroup/task-groups.html>.

¹⁴ Smith-Yoshimura, Karen. 2015. “Shift to Linked Data for Production.” *OCLC Research Hanging Together Blog*, 13 de mayo de 2015.

<https://hangingtogether.org/?p=5195>.

¹⁵ OCLC Research. 2020. “Linked Data.” Linked Data Overview.

<https://www.oclc.org/research/areas/data-science/linkddata/linked-data-overview.html>.

[Todas las imágenes en CC BY 4.0]

¹⁶ Smith-Yoshimura, Karen. 2019. “‘Future Proofing’ of Cataloging.” *OCLC Research Hanging Together Blog*, 10 de noviembre de 2019.

<https://hangingtogether.org/?p=7526>.

¹⁷ ORCID: Connecting Research and Researchers. “What is Orcid. Our Vision”. Consultado el 19 de septiembre de 2020. <https://orcid.org/about/what-is-orcid/mission>.

¹⁸ Véase, por ejemplo, la lista de firmantes de editores de revistas que piden identificadores ORCID para los autores. ORCID. “ORCID Open Letter - Publishers.” Consultado el 19 de septiembre de 2020.

<https://orcid.org/content/requiring-orcid-publication-workflows-open-letter>.

¹⁹ ISNI. “What is ISNI.” Consultado el 19 de septiembre de 2020.

<https://isni.org/page/what-is-isni/>.

²⁰ HathiTrust es una colaboración sin ánimo de lucro de bibliotecas universitarias y de investigación que preservan más de 17 millones de elementos digitalizados. Véase: HathiTrust Digital Library. “Welcome to HathiTrust.” Consultado el 19 de septiembre de 2020. <https://www.hathitrust.org/about>.

²¹ GeoNames. “Browse the Names.” Consultado el 19 de septiembre de 2020.

<https://www.geonames.org/>.

²² Bryant, Rebecca, Annette Dortmund y Constance Malpas. 2017. *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C32K7M>.

²³ ISNI tiene actualmente 11,02 millones de identidades: 10,26 millones son personas, de las cuales 2,91 millones son investigadores, y 933.039 son organizaciones. Estadísticas tomadas de ISNI. Véase ISNI. “Key Statistics.” Consultado el 5 de mayo de 2020. <https://isni.org/>.

²⁴ Library of Congress. 2020. “NACO – Name Authority Cooperative Program.” Documents and Updates. Programs for Cataloging and Acquisitions (PCC). Consultado el 19 de septiembre de 2020. <http://www.loc.gov/aba/pcc/naco/index.html>.

²⁵ Smith-Yoshimura, Karen. 2015. “Getting identifiers Created for Legacy Names.” *Hanging Together: The OCLC Research Blog*, 30 de octubre de 2015. <https://hangingtogether.org/?p=5463>.

²⁶ Smith-Yoshimura, Karen. 2013. “Irreconcilable Differences? Name Authority Control & Humanities Scholarship” *Hanging Together: The OCLC Research Blog*, 27 de marzo de 2013. <https://hangingtogether.org/?p=2621>.

²⁷ Smith-Yoshimura, Karen. 2017. “Use Cases for Local Identifiers.” *Hanging Together: The OCLC Research Blog*, 5 de mayo de 2017. <https://hangingtogether.org/?p=5938>.

²⁸ OCLC Research. 2020. “Registering Researchers in Authority Files.” <https://www.oclc.org/research/themes/research-collections/registering-researchers.html>.

²⁹ Smith-Yoshimura, Karen, Janifer Gatenby, Grace Agnew, Christopher Brown, Kate Byrne, Matt Carruthers, Peter Fletcher, Stephen Hearn, Xiaoli Li, Marina Muilwijk, Chew Chiat Naun, John Riemer, Roderick Sadler, Jing Wang, Glen Wiley y Kayla Willey. 2016. *Addressing the Challenges with Organizational Identifiers and ISNI*. Dublin, Ohio: OCLC Research. <https://doi.org/10.25333/C3FC9Q>.

³⁰ Research Organization Registry (ROR). “About.” <https://ror.org/about/>.

³¹ V. M. Abazov, B. Abbott, B. S. Acharya, M. Adams, T. Adams, J. P. Agnew, G. D. Alexeev et al. (2014) 2020. “Precision Measurement of the Top-Quark Mass in Lepton+jets Final States.” (Archivado el 24 de febrero de 2020) *ArXiv.org*: 1501.07912. <https://arxiv.org/pdf/1405.1756>.

³² Smith-Yoshimura, Karen. 2017. “How Much Metadata Is Practical?” *Hanging Together: The OCLC Research Blog*, 14 de noviembre de 2017.

<https://hangingtogether.org/?p=6328>.

³³ Universidad de Minnesota. 2020. “Experts@Minnesota.” Find Profiles.

<https://experts.umn.edu/en/persons/> o

Universidad de Illinois, campus de Urbana-Champaign. 2020. “Illinois Experts.” Find U of I Research, View Scholarly Works, and Discover New Collaborators.

<https://experts.illinois.edu/>.

³⁴ El National Institute of Health (NIH; Instituto Nacional de Salud): National Institute of Allergy and Infectious Diseases (NIAID; Instituto Nacional de Alergias y Enfermedades Infecciosas) el 7 de abril de 2020 exige ORCIDs para premios de formación, becas universitarias, educación o desarrollo profesional en el año fiscal 2020. Véase NIH: NIAID. 2019. “ORCID iD: Required for Some, Encouraged for All.” *NIAID Funding News*. Última revisión el 7 de agosto de 2019.

<https://www.niaid.nih.gov/grants-contracts/orcid-id-required-some-encouraged-all>;

Véase también Lyrasis. 2020. “SciENCv and ORCID to Streamline NIH and NSF Grant Applications.” *LyrasisNow* (blog), 8 de abril de 2020.

<https://lyrasisnow.org/sciencv-and-orcid-to-streamline-nih-and-nsf-grant-applications/>.

³⁵ Smith-Yoshimura, Karen. 2016. “Metadata Reconciliation.” *Hanging Together: The OCLC Research Blog*, 28 de septiembre de 2016. <https://hangingtogether.org/?p=5710>.

³⁶ Carruthers, Matt. (2014) 2020. *mcarruthers/LCNAF-Named-Entity-Reconciliation*. GitHub Repository.

<https://github.com/mcarruthers/LCNAF-Named-Entity-Reconciliation>.

³⁷ Deliot, Corine, Steven Folsom, Myung-Ja Han, Nancy Lorimer, Terry Reese y Adam Schiff. 2019. *Formulating and Obtaining URIs: A Guide to Commonly used Vocabularies and Reference Sources*. Library of Congress PCC Task Group on URIs in

MARC.

https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf.

³⁸ Smith-Yoshimura, Karen. 2019. “New Ways of Using and Enhancing Cataloging and Authority Records.” *Hanging Together: The OCLC Research Blog*, 2 de abril de 2019. <https://hangingtogether.org/?p=5710>.

³⁹ Smith-Yoshimura, Karen. 2015. “Persistent Identifiers for Local Collections.” *Hanging Together: The OCLC Research Blog*, 27 de octubre de 2015. <https://hangingtogether.org/?p=5445>.

⁴⁰ Véanse ejemplos de DOIs de forma detallada en: DOI. 2020. “DOI System Examples.” Consultado el 20 de septiembre de 2020. <https://www.doi.org/demos.html>;

y
Véanse ejemplos de ARKs de forma detallada en: Department, Dallas (Tex) Police. 1963. “[Photographs of Identification Cards].” Collection. University of North Texas. *The Portal to Texas History digital repository*. <https://texashistory.unt.edu/ark:/67531/metaph346793/>.

⁴¹ DataCite. “Assign DOIs.” <https://datacite.org/does.html>;
Wilkinson, Laura J. 2020. “Constructing your DOIs.” Crossref: The Crossref Curriculum. Última actualización el 8 de abril de 2020. https://www.crossref.org/education/member-setup/constructing_your_dois/.

⁴² “Gestión de identidades” refleja aquí su uso entre los especialistas de metadatos (véase, por ejemplo, Library of Congress. 2018. “Charge for PCC Task Group on Identity Management in NACO,” 5. American Bar Association, Program for Cooperative Cataloging. Revised 22 May 2018. <https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf>.) Pero este término tiene otros significados dependiendo de los destinatarios; por ejemplo, la gestión de accesos a identidades tal como se describe en: Wikiwand. “Identity Management.”

https://www.wikiwand.com/en/Identity_management.

⁴³ Smith-Yoshimura, Karen. 2018. “The Coverage of Identity Management Work.” *Hanging Together: The OCLC Research Blog*, 8 de octubre de 2018. <https://hangingtogether.org/?p=6805>.

⁴⁴ Smith-Yoshimura, Karen. 2017. “Beyond the Authorized Access Point? *Hanging Together: The OCLC Research Blog*, 10 de octubre de 2017. <https://hangingtogether.org/?p=6271>.

⁴⁵ Smith-Yoshimura, “Coverage of Identity Management.” (Véase nota 43)

⁴⁶ Véase el seminario web altamente calificado de Andrew Lih y Robert Fernandez. 2018. “Works in Progress Webinar: Introduction to Wikidata for Librarians: Structuring Wikipedia and Beyond.” Producido por OCLC Research, 12 de junio de 2018. Presentación vídeo MP4, 1:1:51. <https://www.oclc.org/research/events/2018/06-12.html>.

⁴⁷ Smith-Yoshimura, Karen. 2020. “Experimentations with Wikidata/Wikibase, *Hanging Together: The OCLC Research Blog*, 18 de junio de 2020. <https://hangingtogether.org/?p=8002>.

⁴⁸ Wikimedia. “WikiCite.” Home. <https://meta.wikimedia.org/wiki/WikiCite>.

⁴⁹ Smith-Yoshimura, Karen. 2016. “Impact of Identifiers on Authority Workflows. *Hanging Together: The OCLC Research Blog*, 22 de marzo de 2016. <https://hangingtogether.org/?p=5603>.

⁵⁰ Smith-Yoshimura, Karen. 2019. “Strategies for Alternate Subject Headings and Maintaining Subject Headings. *Hanging Together: The OCLC Research Blog*, 29 de octubre de 2019. <https://hangingtogether.org/?p=7591>.

⁵¹ OCLC 2020. “FAST (Faceted Application of Subject Terminology).”

<https://www.oclc.org/en/fast.html>.

⁵² Smith-Yoshimura, Karen. 2016. “Faceted Vocabularies.” *Hanging Together: The OCLC Research Blog*, 31 de octubre de 2016. <https://hangingtogether.org/?p=5739>.

⁵³ OCLC 2020. “FAST.” (Véase nota 51)

⁵⁴ OCLC 2020. “FAST (Faceted Application of Subject Terminology).” Heading #3, FAST Policy and Outreach (FPOC) Committee, : <https://www.oclc.org/en/fast.html>.

⁵⁵ Smith-Yoshimura, Karen. 2017. “Vocabulary Control Data in Discovery Environments.” *Hanging Together: The OCLC Research Blog*, 5 de octubre de 2017. <https://hangingtogether.org/?p=6264>.

⁵⁶ Biblioteca Nacional, Gobierno de Nueva Zelanda. “Ngā Upoko Tukutuku / Māori Subject Headings” <http://mshupoko.natlib.govt.nz/mshupoko/>;
AIATSIS Pathways: Gateway to the AIATSIS Thesauri. “Pathways.” <http://www1.aiatsis.gov.au/>.

⁵⁷ Deutsche Nationalbibliothek. 2019. “MACS - Multilingual Access to Subjects.” (Archivado el 13 de enero de 2019) https://web.archive.org/web/20190113003823/https://www.dnb.de/EN/Wir/Kooperation/MACS/macs_node.html.

⁵⁸ Smith-Yoshimura, Karen. 2019. “Knowledge Organization Systems.” *Hanging Together: The OCLC Research Blog*, 17 de marzo de 2019. <https://hangingtogether.org/?p=7135>.

⁵⁹ Synaptica. “Ontology Management – Graphite.” <https://www.synaptica.com/graphite/>.

⁶⁰ Smith-Yoshimura, Karen. 2018. “Are Distributed Models for Vocabulary Maintenance Viable?” *Hanging Together: The OCLC Research Blog*, 12 de abril de 2018. <https://hangingtogether.org/?p=6672>.

⁶¹ OCLC Research. 2020. “Equity, Diversity, and Inclusion in the OCLC Research Library Partnership Survey.” Overview. Consultado el 20 de septiembre de 2020. <https://www.oclc.org/research/areas/community-catalysts/rlp-edi.html>.

⁶² Smith-Yoshimura, Karen. 2018. “Creating Metadata for Equity, Diversity, and Inclusion.” *Hanging Together: The OCLC Research Blog*, 7 de noviembre de 2018. <https://hangingtogether.org/?p=6833>.

⁶³ Smith-Yoshimura. “Distributed Models.” (Véase nota 60)

⁶⁴ Smith-Yoshimura, Karen. 2019. “Strategies for Alternate Subject Headings and Maintaining Subject Headings.” *Hanging Together: The OCLC Research Blog*, 29 de octubre de 2019. <https://hangingtogether.org/?p=7591>.

⁶⁵ Baxmeyer, Jennifer, Karen Coyle, Joanna Dyla, MJ Han, Steven Folsom, Phil Schreur y Tim Thompson. 2017. *Linked Data Infrastructure Models: Areas of Focus for PCC Strategies*. Library of Congress PCC Linked Data Advisory Committee. <https://www.loc.gov/aba/pcc/documents/LinkedDataInfrastructureModels.pdf>.

⁶⁶ Bone, Christine, Sharon Farnel, Sheila Laroque y Brett Lougheed. 2017. “Works in Progress Webinar: Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries and Archives.” Producido por OCLC Research, 19 de octubre de 2017. Presentación vídeo MP4, 54:35.00. <https://www.oclc.org/research/events/2017/10-19.html>.

⁶⁷ Smith-Yoshimura, Karen. 2015. “Shift to Linked Data for Production.” *Hanging Together: The OCLC Research Blog*, 13 de mayo de 2015. <https://hangingtogether.org/?p=5195>.

⁶⁸ Smith-Yoshimura, Karen. 2015. “Working in Shared Files.” *Hanging Together: The OCLC Research Blog*, 7 de abril de 2015. <https://hangingtogether.org/?p=5091>.

⁶⁹ Bruce Washburn y Jeff Mixter, 2018. “Works in Progress Webinar: Looking Inside the Library Knowledge Vault.” Producido por OCLC Research, 12 de agosto de 2018. Presentación vídeo MP4, 57:45:00. <https://www.oclc.org/research/events/2015/08-12.html>.

⁷⁰ Smith-Yoshimura, Karen. 2019. Systematic Reviews of Our Metadata, *Hanging Together: The OCLC Research Blog*, 10 de abril de 2019. <https://hangingtogether.org/?p=7117>.

⁷¹ Smith-Yoshimura, Karen. 2015. “Working in Shared File.” *Hanging Together: The OCLC Research Blog*, 7 de abril de 2015. <https://hangingtogether.org/?p=5091>.

⁷² Jisc Library Services. Sin fecha. “What Is ‘Plan M’?” Consultado el 21 de septiembre de 2020. <https://libraryservices.jiscinvolve.org/wp/2019/12/plan-m/>;

Smith-Yoshimura, Karen. 2020. “Knowledge Management and Metadata.” *Hanging Together: The OCLC Research Blog*, 9 de abril de 2020.

<https://hangingtogether.org/?p=7845>;

Para aprender más sobre la fase actual del “Plan M” (mayo–noviembre de 2020), véase Grindley, Neil. “Moving Plan M Forwards – We Need Your Help!” *Library Services (PlanM)* (blog), Jisc, 6 de mayo de 2020.

https://libraryservices.jiscinvolve.org/wp/2020/05/planm_nextphase/.

⁷³ Grindley, Neil. 2019. “Plan M: Definition, Principles and Direction.” Jisc. (Word DOCX) <http://libraryservices.jiscinvolve.org/wp/files/2019/12/Plan-M-Definition-and-Direction-1.docx>.

⁷⁴ Dempsey, Lorcan. 2016. “Library Collections in the Life of the User: Two Directions.” *LIBER Quarterly* 26(4): 338–359. <http://doi.org/10.18352/lq.10170>.

⁷⁵ Smith-Yoshimura, Karen. 2019. “Presenting Metadata from Different Sources in Discovery Layers. *Hanging Together: The OCLC Research Blog*, 16 de abril de 2019. <https://hangingtogether.org/?p=7880>.

⁷⁶ Smith-Yoshimura, Karen. 2017. “Metadata for Archival Collections.” *Hanging Together: The OCLC Research Blog*, 30 de mayo de 2017. <https://hangingtogether.org/?p=5903>.

⁷⁷ Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas y Holly Tomren. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*, 49-51. Dublin, OH: OCLC Research. <https://doi.org/10.25333/faq3-ax08>.

⁷⁸ El Archives and Special Collections Linked Data Review Group del OCLC Research Library Partnership se describe en <https://www.oclc.org/research/partnership/working-groups/archives-special-collections-linked-data-review.html>.

⁷⁹ Smith-Yoshimura, Karen. 2020. “Metadata Management in Times of Uncertainty.” *Hanging Together: The OCLC Research Blog*, 15 de junio de 2020. <https://hangingtogether.org/?p=7998>.

⁸⁰ Smith-Yoshimura, Karen. 2016. “Metadata for Archived Websites.” *Hanging Together: The OCLC Research Blog*, 14 de marzo de 2016. <https://hangingtogether.org/?p=5591>.

⁸¹ Archive-It. 2008. “Human Rights.” Columbia University Libraries Collection. (Archivado en mayo de 2008). <https://archive-it.org/collections/1068>;
Archive-It. 2010. “New York City Places and Spaces.” Columbia University Libraries Collection. (Archivado en enero de 2010). <https://archive-it.org/collections/1757>;

Archive-It. 2010. “Burke Library New York City Religions.” Columbia University Libraries Collection. (Archivado en mayo de 2010).

<https://archive-it.org/collections/1945>.

⁸² NLA. “Trove.” Archived Websites. Sub Collections. Consultado el 20 de septiembre de 2020. <https://trove.nla.gov.au/website>.

⁸³ Archive-It. 2014. “Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY).” Ivy Plus Libraries Confederation Collection. (Archivado en junio de 2014) <https://archive-it.org/collections/4638>;

Archive-It. 2013. “Contemporary Composers Web Archive (CCWA).” Ivy Plus Libraries Confederation Collection. (Archivado en octubre de 2013) <https://archive-it.org/collections/4019>;

NYARC: New York Art Resources Consortium. “Web Archiving.”

<http://www.nyarc.org/content/web-archiving>.

⁸⁴ OCLC Research. 2020. “Web Archiving Metadata Working Group” The Problem, Addressing the Problem, Outputs. <https://www.oclc.org/research/themes/research-collections/wam.html>.

⁸⁵ Dooley, Jackie, y Kate Bowers. 2018. *Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3005C>.

⁸⁶ Smith-Yoshimura, Karen. 2018. “Metadata for Audio and Videos.” *Hanging Together: The OCLC Research Blog*, 29 de octubre de 2018.

<https://hangingtogether.org/?p=6814>.

⁸⁷ Weber, Chela Scott. 2017. *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3C34F>.

⁸⁸ Library of Congress. “Standards.” Sitio web oficial de la Encoded Archival Description (EAD). Consultado el 21 de septiembre de 2020. <https://www.loc.gov/ead/>.

⁸⁹ Library of Congress. “Standards.” Preservation Metadata Maintenance Activity (PREMIS). Consultado el 21 de septiembre de 2020. <https://www.loc.gov/standards/premis/>.

⁹⁰ Weber, Chela Scott. 2019. “Assessing Needs of AV in Special Collections.” *Hanging Together: The OCLC Research Blog*, 23 de julio de 2019. <https://hangingtogether.org/?p=7405>;

Weber, Chela Scott. 2019. “Scale & Risk: Discussing Challenges to Managing A/V Collections in the RLP.” *Hanging Together: The OCLC Research Blog*, 1 de octubre de 2019. <https://hangingtogether.org/?p=7479>.

⁹¹ Smith-Yoshimura, Karen. 2015. “Managing Metadata for Image Collections.” *Hanging Together: The OCLC Research Blog*, 9 de abril de 2015. <https://hangingtogether.org/?p=5130>.

⁹² Library of Congress. “Standards.” Metadata Object Description Schema (MODS). Consultado el 21 de septiembre de 2020. <http://www.loc.gov/standards/mods/>.

⁹³ Ibid.

⁹⁴ Library of Congress. “Standards.” Metadata Authority Description Schema (MADS).” Consultado el 21 de septiembre de 2020. <http://www.loc.gov/standards/mads/>.

⁹⁵ Smith-Yoshimura, Karen. 2016. “Sharing Digital Collections Workflows.” *Hanging Together: The OCLC Research Blog*, 2 de noviembre de 2016. <https://hangingtogether.org/?p=5744>.

⁹⁶ OCLC Research. 2020. “Europeana Innovation Pilots.” Consultado el 20 de septiembre de 2020.

<http://www.oclc.org/research/themes/data-science/europeana.html?urlm=168921>.

⁹⁷ IIF (International Image Interoperability Framework): Enabling Richer Access to the World’s Images. “Home.” Consultado el 20 de septiembre de 2020. <https://iif.io/>.

⁹⁸ OCLC Research. 2020. “OCLC ResearchWorks IIF Explorer.”

<https://www.oclc.org/research/themes/data-science/iif/iifexplorer.html>.

⁹⁹ OCLC Research. 2020. “CONTENTdm Linked Data Pilot.” Introduction.

<https://www.oclc.org/research/themes/data-science/linkedata/contentdm-linked-data-pilot.html>.

¹⁰⁰ Smith-Yoshimura, Karen. 2015. “Data Management and Curation in 21st Century Archives – Part 1.” 21 de septiembre de 2015. <http://hangingtogether.org/?p=5375>.

¹⁰¹ Smith-Yoshimura, Karen. 2016. “Metadata for Research Data Management.” *Hanging Together: The OCLC Research Blog*, 18 de abril de 2016.

<https://hangingtogether.org/?p=5616>.

¹⁰² Erway, Ricky, Laurence Horton, Amy Nurnberger, Reid Otsuji y Amy Rushing. 2015. *Building Blocks: Laying the Foundation for a Research Data Management Program*, 8. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C39P86>.

¹⁰³ Véase la Data Management Planning Guide (Guía de planificación de la gestión de datos) de OCLC Research en

<https://www.oclc.org/research/areas/research-collections/rdm/guide.html>.

¹⁰⁴ Smith-Yoshimura, Karen. 2020. “Knowledge Management and Metadata.” *Hanging Together: The OCLC Research Blog*, 9 de abril de 2020.

<https://hangingtogether.org/?p=7845>.

¹⁰⁵ Faniel, Ixchel M. 2019. “Let’s Cook Up Some Metadata Consistency.” *Next* (blog), *OCLC*, 21 de noviembre de 2019. <http://www.oclc.org/blog/main/lets-cook-up-some-metadata-consistency/>.

¹⁰⁶ NCI (National Computational Infrastructure): Australia. “Home.” Consultado el 21 de septiembre de 2020. <http://nci.org.au/>;

ADA (Australian Data Archive). “Home.” Consultado el 21 de septiembre de 2020. <https://www.ada.edu.au/>.

¹⁰⁷ Portage Network. “Home.” Consultado el 21 de septiembre de 2020.

<https://portagenetwork.ca/>.

¹⁰⁸ Metadata 2020 es una “colaboración que aboga por metadatos más enriquecidos, conectados, reutilizables y abiertos para todas las producciones de investigación” (<http://www.metadata2020.org/>). El proyecto Researcher Communications de Metadata 2020 se describe aquí:

<http://www.metadata2020.org/projects/researcher-communications/>.

¹⁰⁹ Digital Curation Centre. “Disciplinary Metadata.” List of Metadata Standards. Consultado el 21 de septiembre de 2020. <http://www.dcc.ac.uk/resources/metadata-standards/list>.

¹¹⁰ RDA Metadata Directory. “Metadata Standards Directory Working Group.” GitHub Repository. Consultado el 21 de septiembre de 2020.

<http://rd-alliance.github.io/metadata-directory/>.

¹¹¹ NISO está a punto de convertir CRediT (Contributor Roles Taxonomy; Taxonomía de Roles de Contribuidores), que identifica 14 roles que describen la contribución específica de cada contribuidor a la producción universitaria, en una norma. CRediT fue desarrollado por CASRAI, los Consortia Advancing Standards in Research Administration Information (Consortios Promovedores de Normas en la Información de

la Administración de Investigación). Véase CASRAI. “CRediT – Contributor Roles Taxonomy.” Consultado el 21 de septiembre de 2020. <https://casrai.org/credit/>.

¹¹² Biblioteca de la Universidad de Michigan. 2020. “Data Services.” <http://www.lib.umich.edu/research-data-services>.

¹¹³ OCLC Research. 2020. “The Realities of Research Data Management.” Overview. <https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html>.

¹¹⁴ Bryant, Rebecca, Brian Lavoie y Constance Malpas. 2017. *Scoping the University RDM Service Bundle. The Realities of Research Data Management*, Part 2, pp. 16, 21. Dublin, OH: OCLC Research. <https://doi.org/10.25333/C3Z039>.

¹¹⁵ Universidad de Indiana. 2018. “La Universidad de Indiana liderará una sociedad de 2 millones de dólares para ampliar el acceso a los datos de investigación: las bibliotecas de la Universidad de Indiana y el Instituto de Ciencias de la Red de la Universidad de Indiana lideran una sociedad público-privada para crear Shared BigData Gateway for Research Libraries”, *News at UI, (Science and Technology.) Indiana University*, 18 de octubre de 2018. <https://news.iu.edu/stories/2018/10/iu/releases/18-shared-bigdata-gateway-for-research-networks.html>;

Microsoft. 2020. “Microsoft Academic Graph.” Creado el 5 de junio de 2015. <https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/>;

Para más detalles véase la grabación de agosto de 2019 “Democratizing Access to Large Datasets through Shared Infrastructure.” Véase Wittenberg, Jamie, y Valentin Pentchev. “Works in Progress Webinar: Democratizing Access to Large Datasets through Shared Infrastructure.” Producido por OCLC Research, 8 de agosto de 2019. Presentación de vídeo MP4, 58:34:00.

<https://www.oclc.org/research/events/2019/080819-democratizing-access-large-datasets-shared-infrastructure.html>.

¹¹⁶ El distintivo de reproducibilidad y las definiciones de NISO, ahora disponibles a valoración pública, también pueden ayudar a los investigadores a ampliar el beneficio de su investigación a otros. Véase “Taxonomy, Definitions, and Recognition Badging Scheme Working Group | NISO Website.” Sin fecha. Consultado el 22 de septiembre de 2020. <https://www.niso.org/standards-committees/reproducibility-badging>.

¹¹⁷ Smith-Yoshimura, Karen. 2015. “Services Built on Usage Metrics.” *Hanging Together: The OCLC Research Blog*, 30 de septiembre de 2015. <https://hangingtogether.org/?p=5430>.

¹¹⁸ Krista M. Soria, Jan Fransen, Shane Nackerud. 2014. “Stacks, Serials, Search Engines, and Students’ Success: First-Year Undergraduate Students’ Library Use, Academic Achievement, and Retention.” *Journal of Academic Librarianship* 40: 84-91. <https://doi.org/10.1016/j.acalib.2013.12.002>.

¹¹⁹ Véase Jisc. “Library Impact Data Project (LIDP).” Consultado el 21 de septiembre de 2020. <http://www.activitydata.org/LIDP.html>.

¹²⁰ Smith-Yoshimura, Karen. 2019. “Alternatives to Statistics for Measuring Success and Value of Cataloging.” *Hanging Together: The OCLC Research Blog*, 15 de abril de 2019. <https://hangingtogether.org/?p=7122>.

¹²¹ DLF (Digital Library Federation; Federación Bibliotecaria Digital). 2015. “Metadata Librarian, Cornell University Library.” *DLF* (blog), 11 de junio de 2015. <https://www.diglib.org/metadata-librarian-cornell-university-library/>.

¹²² Salary.com. (2019) 2020. “Metadata Librarian.” Publicado por la Universidad Tecnológica de Georgia el 13 de noviembre de 2019. (Archivado el 2 de septiembre de 2020) <https://web.archive.org/web/20200903061830/https://www.salary.com/job/gt-library/metadata-librarian/e5644ece-c847-4cfb-994f-c4c80fa81e3d>.

¹²³ OCLC. 2020. “Locate Items in the Library with StackMap.”

https://help.oclc.org/Discovery_and_Reference/WorldCat_Discovery/Search_results/Locate_items_in_the_library_with_StackMap.

¹²⁴ Yewno: Transforming Information into Knowledge. 2020. “Home.”

<https://www.yewno.com/>.

¹²⁵ Smith-Yoshimura, Karen. 2019. “New Ways of Using and Enhancing Cataloging and Authority Records” *Hanging Together: The OCLC Research Blog*, 2 de abril de 2019. <https://hangingtogether.org/?p=7805>.

¹²⁶ National Library of Australia (NLA). “Austlang National Codeathon.” Consultado el 21 de septiembre de 2020. <https://www.nla.gov.au/our-collections/processing-and-describing-thecollections/Austlang-national-codeathon> [Map of Australia. 2020 HERE, Bing, Microsoft Corporation];

NLA. “Trove.” Search. Uncover. Australia. Consultado el 21 de septiembre de 2020. <https://trove.nla.gov.au/>.

¹²⁷ The Graphic History Company – Hachette UK. “River of Authors.” Consultado el 21 de septiembre de 2020. <http://theghc.co/project.php?project=hachette-uk-a-river-of-authors>.

¹²⁸ Smith-Yoshimura, Karen. 2019. “Knowledge Organization Systems.” *Hanging Together: The OCLC Research Blog*, 17 de abril de 2019.

<https://hangingtogether.org/?p=7135>.

¹²⁹ Zeng, Marcia Lei, y Philipp Mayr. 2019. “Knowledge Organization Systems (KOS) in the Semantic Web: A Multi-dimensional Review.” *International Journal on Digital Libraries* 20: 209-230. <https://doi.org/10.1007/s00799-018-0241-2>.

¹³⁰ SNAC (Social Networks and Archival Context). “About SNAC.” What is SNAC? <https://portal.snaccooperative.org/about>.

¹³¹ Smith-Yoshimura, Karen. 2020. “Knowledge Management and Metadata.” *Hanging Together: The OCLC Research Blog*, 9 de abril de 2020.

<https://hangingtogether.org/?p=7845>.

¹³² AI4LAM (Artificial Intelligence for Libraries, Archives & Museums). Actualizado el 18 de mayo de 2020. <https://sites.google.com/view/ai4lam/home>.

¹³³ La misión de AI4LAM es organizar, compartir y ensalzar el conocimiento y el uso de la inteligencia artificial por parte de bibliotecas, archivos y museos. Fue fundada en 2018, inspirada por el éxito del International Image Interoperability Framework (IIIF) coordinando la colaboración a gran escala en tecnología interoperable para avanzar en los LAMs.

Véase AI4LAM. “About.” Our Mission. <https://sites.google.com/view/ai4lam/about>.

¹³⁴ Padilla, Thomas. 2019. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/xk7z-9g97>.

¹³⁵ *Ibid*, 17-19.

¹³⁶ Smith-Yoshimura, Karen. 2019. “Alternatives to Statistics for Measuring Success and Value of Cataloging.” *Hanging Together: The OCLC Research Blog*, 15 de abril de 2019. <https://hangingtogether.org/?p=7122>.

¹³⁷ Smith-Yoshimura, Karen. 2017. “New Skill Sets for Metadata Management.” *Hanging Together: The OCLC Research blog*, 17 de abril de 2017. <https://hangingtogether.org/?p=5929>.

¹³⁸ Smith-Yoshimura, Karen. 2018. “MarcEdit and Other Tools for Batch Processing and Metadata Reconciliation.” *Hanging Together: The OCLC Research Blog*, 26 de marzo de 2018. <https://hangingtogether.org/?p=6646>.

¹³⁹ Reese, Terry. 2018 “MarcEdit 2017 Usage Information.” *Terry’s Worklog* (blog), 9 de septiembre de 2020. <http://blog.reeset.net/archives/2572>.

¹⁴⁰ Reese, Terry. 2020. “Working with Linked Data In MarcEdit.” *MarcEdit Development* (blog). Consultado el 21 de septiembre de 2020. <https://marcedit.reeset.net/working-with-linked-data-in-marcedit>.

¹⁴¹ Reese, Terry. 2018. “MarcEdit Playlist.” 139 vídeos de YouTube. Última actualización el 26 de diciembre de 2018. <https://www.youtube.com/playlist?list=PLrHRsJ91nVFScJLS91SWR5awtFfpewMWg>.

¹⁴² Smith-Yoshimura, Karen. 2017. “New Skill Sets for Metadata Management.” *Hanging Together: The OCLC Research blog*, 17 de abril de 2017. <https://hangingtogether.org/?p=5929>.

¹⁴³ “XML and RDF-Based Systems Archives.” Sin fecha. *Library Juice Academy* (blog). Consultado el 22 de septiembre de 2020.

<https://libraryjuiceacademy.com/certificate/xml-and-rdf-based-systems/>;

Reese, Terry. 2013. “Tutorials.” YouTube (selección). *MarcEdit Development* (blog). 14 de marzo de 2013. <http://marcedit.reeset.net/tutorials>;

“Lynda: Online Courses, Classes, Training, Tutorials.” Sin fecha. Lynda.com - from LinkedIn Learning. Consultado el 22 de septiembre de 2020. <https://www.lynda.com/>;

“Learn to Code - for Free.” Sin fecha. Codecademy. Consultado el 22 de septiembre de 2020. <https://www.codecademy.com/>;

Software Carpentry. “Teaching Basic Lab Skills for Research Computing.” Upcoming Workshops. Consultado el 22 de septiembre de 2020. <https://software-carpentry.org/>.

¹⁴⁴ “Data on the Web Best Practices.” Sin fecha. Consultado el 22 de septiembre de 2020. <https://www.w3.org/TR/dwbp/>;

Semantic Web for the Working Ontologist. (2008) 2020. <http://workingontologist.org/>.

¹⁴⁵ Library Workflow Exchange. Sin fecha. “About.” Consultado el 21 de septiembre de 2020. <http://www.libraryworkflowexchange.org/about/>.

¹⁴⁶ OCLC Developer Network. 2020. “DevConnect Webinars.” <https://www.oclc.org/developer/events/devconnect-workshops.en.html>.

¹⁴⁷ Smith-Yoshimura, Karen. 2019. “Stewardship of Professional FTEs In Metadata Work and Turnover.” *Hanging Together: The OCLC Research Blog*, 18 de octubre de 2019. <https://hangingtogether.org/?p=7580>.

¹⁴⁸ OCLC. 2020. “WorldCat®: OCLC and Linked Data.” Shared Entity Management Infrastructure. <https://www.oclc.org/en/worldcat/linked-data/shared-entity-management-infrastructure.html>.