

Identificadores de persona en registros de MARC 21 en un entorno de web semántica

Person Identifiers in MARC 21 Records in a Semantic Environment

Liliana Giusti Serra, Jodi A. Schneider & José Eduardo Santarém Segundo

Cataloging & Classification Quarterly, ISSN 1544-4554, Vol. 58, n.5, 2020, p. 505-519

Este artículo trata de cómo las bibliotecas pueden incluir identificadores de persona en el formato MARC. Sugiere que usar URIs en los campos y subcampos puede ayudar a la transición de los datos en un modelo RDF, y ayudar a preparar el catálogo para los datos enlazados. Analiza la selección de URIs para objetos del mundo real, y el uso de la etiqueta 024 para describir identificadores de persona en registros de autoridad. Cuando un creador o colaborador es identificado en una obra, los identificadores son transferidos del registro de autoridad al bibliográfico. Estos registros pueden usar el subcampo \$0 para almacenar un número de control de registro de autoridad o un número estándar, que incluya identificadores de personas. En 2017 la Library of Congress creó el subcampo \$1 y cambió el subcampo \$4 para adaptar URIs y preparar los datos bibliotecarios para una futura conversión de MARC 21 a RDF. El subcampo \$1 sirvió para crear un nuevo lugar en el que poner URIs para proyectos de web semántica y está diseñado para admitir URIs de objetos del mundo real, identificando personas, lugares, conceptos, cosas, etc., que tienen un nombre. Por su parte, el subcampo \$4 describe relaciones, usando vocabularios u ontologías como fuentes, tales como Friend of a Friend (FOAF), Dublin Core Metadata Initiative (DCMI), Resource Description and Access (RDA), etc. El subcampo \$e de las etiquetas 100 y 700 representa relaciones entre creadores intelectuales y contribuidores. Por ejemplo, un nombre en la etiqueta 100 puede representar un creador y un nombre en la etiqueta 700 un ilustrador. La misma representación en una estructura de web semántica se hace en el subcampo \$4, donde la información puede ser proporcionada en texto plano (el código par la relación), o por medio de un URI, haciendo que los datos sean legibles por máquina y compatibles con las recomendaciones de la web semántica. En 2019 el estándar MARC actualizó la etiqueta 024, incluyendo los subcampos \$0 y \$1 para registrar URIs, y estableció que la fuente de los identificadores debe ser registrada en el subcampo \$2. Si no se proporciona la fuente, solo se puede reconocer por los URI en \$0 y \$1, lo cual no es amigable ni para humanos ni para máquinas, ya que no siempre es posible identificar la fuente de los datos. El hecho de que los identificadores puedan ser registrados en los subcampos \$0 y \$1 puede ser confuso para los bibliotecarios, ya que deben comprender si un identificador representa una descripción o una cosa, y en qué subcampo debe registrarse cada uno. Esto puede añadir complejidad a la interpretación y reutilización de los datos. Las descripciones estructuradas deben ofrecer beneficios para

los usuarios en el catálogo. Incluir más datos ocultos o difíciles de localizar para los usuarios supone malgastar esfuerzos de los catalogadores, y no supone beneficios para los usuarios. Hay que centrarse en cómo la descripción puede proporcionar nuevas opciones a los usuarios, facilitando su uso del catálogo y acelerando el descubrimiento tanto en el catálogo como más allá.

<https://www.tandfonline.com/doi/full/10.1080/01639374.2020.1771499>

Resumen elaborado por Antonio Rodríguez Vela

El trabajo de PCC sobre URIs en MARC

PCC's Work on URIs in MARC

Jackie Shieh

Cataloging & Classification Quarterly, ISSN 1544-4554, Vol. 58, n.3-4, 2020, p. 418-427

En 2015 el Grupo de Trabajo (GT) de PCC sobre URIs en MARC fue encargado de investigar el uso de identificadores de datos enlazados en el actual formato MARC. A través de una prueba piloto, una encuesta, artículos, propuestas, etc., el GT indicó varios cambios para la codificación en MARC. Para facilitar el procesamiento por máquina (comunicación de máquina a máquina) en MARC es necesario registrar los datos RDF en un subcampo diferente del utilizado normalmente en MARC para los URIs, lo que también posibilita su interoperabilidad. Tras consultar la base de datos WorlCat y la documentación de MARC referente al uso de estos subcampos, el GT decidió redefinir el \$0 para acomodar los URIs desreferenciables. En la web semántica, los URIs identifican no solo documentos web, sino también objetos del mundo real como personas y coches, e incluso ideas abstractas que son cosas inexistentes, como un unicornio. A estos objetos del mundo real se los llama RWO. El modelo de datos enlazados para personas, lugares y conceptos descritos en los campos de autoridad normalmente se modelan siguiendo SKOS Concepts. El GT redactó un artículo para establecer un nuevo subcampo para registrar URIs desreferenciables para los ROWs. Tras una revisión de las opciones disponibles, se decidió proponer el \$1 para este propósito. De la observación de los informes de la prueba piloto, se dedujo que se necesitaba una aclaración sobre cómo los campos y subcampos de MARC se relacionan con los objetos RDF. El GT se dio cuenta de que podría ser imposible establecer una norma general aplicable a todos los campos MARC sobre los subcampos dedicados a objetos RDF. Se necesitaría que los catalogadores utilizaran sus conocimientos para proporcionar un análisis que ayudara a una conversión automatizada confiable de MARC a RDF. El GT discutió los problemas de la entidad Obra en relación con los campos de MARC para obras y expresiones,

incluyendo construcciones autor-título en los campos 1XX/240 y 7XX del formato bibliográfico y el campo correspondiente 1XX en el formato de autoridades. El consenso fue la necesidad de registrar la entidad de Obra en MARC aparte de los puntos de acceso establecidos en MARC. El GT tuvo éxito en establecer un nuevo campo MARC, el 758, para señalar los URIs, con el nombre del campo ahora cambiado a Identificador de Recurso, de acuerdo con el despliegue de un subcampo de relación (en \$4 o \$i), expresando la relación entre sujeto y objeto de RDF. Una de las tareas del GT fue consultar a los profesionales sobre las prioridades y problemas en la identificación. Se diseñó una encuesta, enviada a colegas de todo el mundo en 2017, y se informó de los resultados al comité pertinente de PCC. El GT supo por experiencia que los URIs desreferenciables no son un concepto fácil. Por lo tanto, realizó dos documentos de ayuda para que sus esfuerzos pudieran aliviar el peso y la confusión que rodea su aplicación en MAR: URI FAQs.

(<https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI%20FAQs.pdf>) y Formulating and Obtaining URIs.

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

<https://www.tandfonline.com/doi/full/10.1080/01639374.2019.1705951>

Resumen elaborado por Antonio Rodríguez Vela

[El futuro del control de autoridades: cuestiones y tendencias en el entorno de datos enlazados](#)

The Future of Authority Control: Issues and Trends in the Linked Data Environment

Lihong Zhu

Journal of Library Metadata, ISSN 1937-5034, Vol. 19, n. 3-4, 2019, p. 215-238

Los datos enlazados son un conjunto de buenas prácticas para publicar y conectar datos estructurados en la web. Las tecnologías que apoyan los datos enlazados son las URIs (Uniform Resource Identifier), RDF (Resource Description Framework), y SPARQL (Protocol and RDF Query Language). Los datos enlazados pueden ser datos abiertos, y Berners-Lee recomienda un esquema de cinco estrellas para su valoración. El control de autoridades es un conjunto de procedimientos para mejorar los archivos de autoridad y mantener su consistencia en forma de encabezamientos usados en registros bibliográficos. Los usuarios del catálogo de las bibliotecas deben ser capaces de confiar en

la consistencia y fiabilidad de los encabezamientos para una búsqueda efectiva. El propósito principal de este artículo es examinar las publicaciones académicas recientes sobre el control de autoridades y el entorno de datos enlazados y responder a la pregunta de cuáles son sus tendencias y los problemas actuales. En el artículo se identifican seis tendencias principales. 1, el cambio de un entorno basado en MARC a uno que utiliza datos enlazados. La Library of Congress está trabajando en convertir los registros de MARC21 en BIBFRAME, lo que ha llevado a la OCLC a construir capacidades que se apoyan en un paisaje más complejo. 2, cambio de secuencias a cosas. Para prepararse para los datos enlazados, es crítica una transición del trabajo de control de autoridades que se centra en crear encabezamientos autorizados ("secuencias" de texto) a uno basado en identificadores (URIs o "cosas"). 3, cambiar hacia vocabularios abiertos. La tendencia es usar la palabra "ontología" no en su sentido más complejo, una colección formal de términos, sino en un sentido más vago, en el que los vocabularios son las construcciones básicas para la construcción de la web semántica. 4, cambiar hacia la gestión de identidad. Esta enfatiza el proceso de asociación de un identificador registrado (o una URI) con una entidad simple y la diferenciación de nombres de encabezamientos solo tiene una importancia secundaria en la gestión de identidad. 5, cambio hacia un esquema facetado de terminología de materias. Se trata de una transformación de las secuencias precoordinadas a terminologías postcoordinadas de terminología de materias. 6, cambio para utilizar RDA como un estándar de contenido flexible. RDA es un lugar lógico desde el que empezar el trabajo cooperativo. Sin embargo, parece claro que RDA pueda no cumplir todos los requisitos de un entorno de datos enlazados. BIBFRAME y RDA tendrán que ser evaluados para que ambos estándares puedan ser utilizados en todo su potencial, permitiéndose la coexistencia de prácticas no-RDA.

<https://www.tandfonline.com/doi/full/10.1080/19386389.2019.1688368>

Resumen elaborado por Antonio Rodríguez Vela
