





CHARLES BLAIR

Aplicando el modelo de datos de Europeana a un repositorio de biblioteca digital

El Repositorio Digital de la Biblioteca de la Universidad de Chicago es un repositorio de preservación que comprende documentos digitales y los originalmente en soportes tradicionales del Archivo de la Universidad de Chicago, así como los digitalizados retrospectivamente procedentes de las Special Collections Research Center (Colecciones Especiales del Centro de Investigación). Los materiales son de tipo textual, gráficos, audio y audiovisuales. La estructura de los materiales oscila desde la relativamente simple, por ejemplo, un mapa, a otros de cierta complejidad, por ejemplo un documento de la Universidad o una publicación seriada de múltiples volúmenes, cada uno de los cuales consiste en una secuencia de ficheros master en formato TIFF, uno por cada página; de los que se derivan los formatos JPEG para cada página, archivos OCR para cada página en formato XML, un archivo PDF para todo el número y un registro de metadatos estructurales que anotan las características de la página (por ejemplo, “Página 1”).¹

La LDR se conecta a los procesos archivísticos de transferencia, ingreso y procesamiento de los materiales.² También, siguiendo la práctica archivística tradicional, cada ingreso se asigna a una colección particular, aun cuando, como en

Charles Blair (email: chas@uchicago.edu)
Digital Library Development Center (DLDC), University of Chicago

Traducción: Alicia Barnard y Luis Hernández Olivera

Recibido: 29-06-2015. Aceptado: 12-09-2015

Citación: Blair, Charles (2015). “Aplicando el modelo de datos de Europeana a un repositorio de biblioteca digital”.
Tábula, n. 18, pp. 105-108

el caso de algunos materiales digitalizados retrospectivamente, se genere una colección artificial para este propósito. Tras el ingreso se crean unos instrumentos de consulta preliminares en formato EAD³. El procesamiento se mapea en un modelo de información Sistema Abierto de Información Archivística (OAIS por sus siglas en inglés).⁴ Este proceso se describe a continuación.

Después de la transferencia de los materiales digitales a las Colecciones Especiales del Centro de Investigación, se registran los metadatos administrativos, acerca de los materiales que se transfieren, incluyendo quien solicitó la transferencia, el personal que llevó a cabo la misma y los tipos de derechos y permisos (restricciones) que se aplican a los materiales que se transfieren. Estos pueden incluir una restricción de acceso por un determinado período.

Cuando los materiales han sido transferidos, se procede a su incorporación. Además del número de entrada asignado localmente, se genera de forma automática un identificador ARK (clave de recurso archivístico, siglas en inglés) (Kunze 2003) para cada incorporación. Todos los ingresos deben pertenecer a una colección. Si no existiese la colección el archivero establecerá una nueva antes de que la incorporación sea efectuada. La experiencia ha demostrado que, aunque sea posible distribuir la función de incorporación entre diferentes miembros interesados en los materiales que se depositan, es mejor asignar la función a un miembro del personal dedicado a la misma: el especialista en ingreso digital (DAS por sus siglas en inglés). El DAS también activará el FITS (conjunto de herramientas de información del archivo)⁵ para generar los metadatos técnicos. El DAS también migrará los formatos de archivo que se encuentren en riesgo a otros que ofrezcan mejores garantías de preservación archivística por el paso del tiempo.

Después del ingreso los materiales son procesados. Esta función se realiza de manera programática por el Digital Library Development Center (Centro de Desarrollo de la Biblioteca Digital), utilizando la información que el DAS registra en la base de datos de ingresos cuando los materiales se reciben. Los paquetes de presentación de la información (SIP's en inglés) se crean de forma automática desde la base de datos de ingresos según el Modelo de Datos Europeana (EDM) en Turtle (Terse RDF Lenguaje Triple),⁶ lo que permite ajustar la variedad de tipos de contenido que existen en la LDR, y también permite configurar las diferentes estructuras de los recursos digitales que utilizan los elementos de datos que proporciona. Mediante el uso del modelo OAI-ORE (Iniciativa de Archivos Abiertos-Intercambio y Reutilización de Objetos, siglas en inglés)⁷, el subyacente EDM, los metadatos técnicos compatibles con PREMIS⁸ para recursos digitales se registran para los ficheros master. Los proxies de EDM se emplean para representar una variante descriptiva de los metadatos, utilizando un enfoque análogo al de "llamar por referencia" en lugar de "llamar por valor" de los lenguajes de programación. Es decir, en lugar de una representación en-línea para, por

ejemplo, una descripción TEI⁹ o una MARC¹⁰ de un recurso, se proporciona un puntero mediante una URL accionable a un fichero de metadatos descriptivos, haciendo que los metadatos estén disponibles para su utilización por una aplicación descendente (downstream). Esta aproximación vincula los recursos digitales a metadatos de dominio específico potencialmente ricos, por ejemplo metadatos VRA Core¹¹ a diferencia de Dublin Core.¹² Los que podríamos llamar “metadatos para subtítulos” en el Dublin Core no calificado, se proporcionan también en línea, como se especifica en el modelo EDM, para permitir una capa de descubrimiento uniforme que se establece a lo largo de todos los objetos en el LDR. Este enfoque también permite el mapeo de EAD a EDM como se describe en Gardasevic (2011), Hennieke, Olensky, de Boer et al. (2011), y Casarosa, Meghini, Gardasevic (2013). El control de las versiones se maneja al reemplazar un enlace con otro: una vez depositado, no hay necesidad de mover objetos, simplemente se registra un evento en el control de versiones.

Los paquetes de difusión de la información (DIP en inglés) se crean mediante consultas SPARQL una vez que los SIP's se han cargados en un RDF triplestore. Los DIP's contienen URL's procesables basadas en los identificadores ARK, asignados en el momento del ingreso; estos pueden ser recuperados desde la LDR utilizando herramientas como curl.¹³ Los DIP's se pueden adaptar a propósitos específicos, como por ejemplo, para transferir todos los recursos a otro repositorio para un almacenamiento duplicado, para construir una colección de ciertos recursos o para enviar recursos seleccionados a un académico con un interés investigador.

Esta rápida aproximación tiene una gran “capacidad de elasticidad” pues puede operar una carga de trabajo completa con una relativamente pequeña cantidad de esfuerzo invertido en su desarrollo y mantenimiento. El hecho de que cumple con los estándares, la implementación de todos los elementos requeridos por EDM y PREMIS, supone que no se requiera ningún conocimiento especial, distinto a esas normas y a los conceptos básicos del lenguaje de consulta SPARQL, también establecido por una norma, para extraer del LDR información relevante acerca de los recursos que contiene.

Referencias

- Consultative Committee for Space Data Systems. (2012). *Reference Model for an Open Archival Information System (OAIS). Recommended Practice*. CCSDS 650.0-M-2. Magenta Book. Issue 2.
- Casarosa, Vittore; Meghini, Carlo; Gardasevic, Stanislava. (2013). “Improving Online Access to Archival Data”. *Digital Libraries & Archives*, pp. 153-162.
- Encoded Archival Description Tag Library: Version 2002*. (2003). Chicago: The Society of American Archivists.

- Gardasevic, Stanislava. (2011). "Opening Archives to the General Public, a data modeling approach". Master thesis. International Master in Digital Library Learning.
- Hennicke, Steffen; Olenky Marlies; de Boer, Victor; Isaac Antoine; Wielemaker, Jan. (2011). "Conversion of EAD into EDM Linked Data". En: *Proceedings of the 1st International Workshop on Semantic Digital Archives*. <http://www-e.uni-magdeburg.de/predoiu/sda2011/sda2011_06.pdf>. [Consultado: 2015-06-29]
- Kunze, John. (2003). "Towards Electronic Persistence Using ARK Identifiers". *Proceedings of the ECDL Web Archiving Workshop*.
- ROE, Kathleen D. (2005). *Arranging & Describing Archives & Manuscripts*. Chicago: The Society of American Archivists.

Notas

¹ Sin embargo, aún un mapa puede tener estructura si es muy grande para caber completamente en un escaner que de tal forma que tendrá que escanearse en secciones.

² Ver, por ejemplo, Roe, 2015, pp.45-86

³ *Encoded Archival Description Tag Library: Version 2002*. 2003. *Passim*.

⁴ Consultative Committee for Space Data Systems, 2012, secs. 2.2.3, 2.3, pp. 2-7, 2-8.

⁵ <<http://projects.iq.harvard.edu/fits/home>>

⁶ <<http://www.w3.org/TeamSubmission/turtle/>>

⁷ <<http://www.openarchives.org/ore/>>

⁸ <<http://www.loc.gov/standards/premis/>>

⁹ <<http://www.tei-c.org/index.xml>>

¹⁰ <<http://www.loc.gov/marc/>>

¹¹ <<http://loc.gov/standards/vrarc/>>

¹² <<http://dublincore.org/>>

¹³ Esta es una herramienta estandar en distribuciones Unix/Linux

