

B.3. Reutilización e intercambio de objetos digitales compuestos en la Web: el proyecto OAI-ORE

Por Enrique Orduña-Malea

28 julio 2008

Orduña-Malea, Enrique. "Reutilización e intercambio de objetos digitales compuestos en la Web: el proyecto OAI-ORE". *Anuario ThinkEPI 2009*, EPI SCP, pp. 45-48.



Resumen: La reutilización e intercambio de agregaciones de recursos web trae consigo nuevas posibilidades de creación y descripción de colecciones de objetos sin importar su ubicación, facilitando con ello nuevos flujos de trabajo académico. Este trabajo se centra en el proyecto OAI-ORE, cuyos objetivos se basan en la identificación y descripción de agregaciones de recursos web, así como en las relaciones entre éstos.

Palabras clave: OAI-ORE, Objetos digitales compuestos, Agregación de recursos, Mapa de recursos, Conjunto de datos.

Title: *Reuse and exchange of compound digital objects on the Web: the OAI-ORE project*

Abstract: The reuse and exchange of compound digital web resources offers new possibilities to create and describe collections of objects, regardless of their location, thereby facilitating new academic workflows. This paper is focused on the OAI-ORE project, the goals of which are based on the identification and description of aggregations of web resources and their interrelationships.

Keywords: OAI-ORE, Compound digital objects, Aggregated resources, Resource maps, Dataset.

1. Introducción

LOS OBJETOS DIGITALES COMPUESTOS son conjuntos o agrupaciones de documentos que forman juntos una entidad lógica nueva con significado propio. Por ello, bajo la óptica de la teoría general de sistemas (Bertalanfy, 1976), una agregación web podría considerarse un sistema.

En los documentos impresos las diferencias entre los elementos de una agregación son, salvo excepciones, meramente conceptuales pero no físicas. Un artículo impreso formado por imágenes, tablas, texto, etc., está formado en su origen por documentos distintos pero que al ser impresos (o incrustados previamente en un procesador de textos tipo *Word*), pasan a conformar un único documento físico (o lógico).

En la Web la situación es diferente. Los distintos elementos de una agregación pasan a tener una independencia lógica y física (se encuentran en ficheros diferentes y en algunos casos incluso en distintos servidores).

El futuro de la comunicación académica y científica depende en gran medida de la gestión de estas agregaciones, pues su uso trae consigo nuevas necesidades de citación así como la posibilidad de creación y descripción de colecciones de objetos sin importar su ubicación.

Este hecho pone en evidencia la necesidad de definir un estándar que permita identificar, describir e intercambiar agregaciones de recursos. Es decir, de acotar los límites del documento en la Web: su dominio.

La reciente especificación OAI-ORE, cuyos objetivos son la identificación y descripción de agregaciones de recursos web, constituye una novedosa aproximación a este problema.

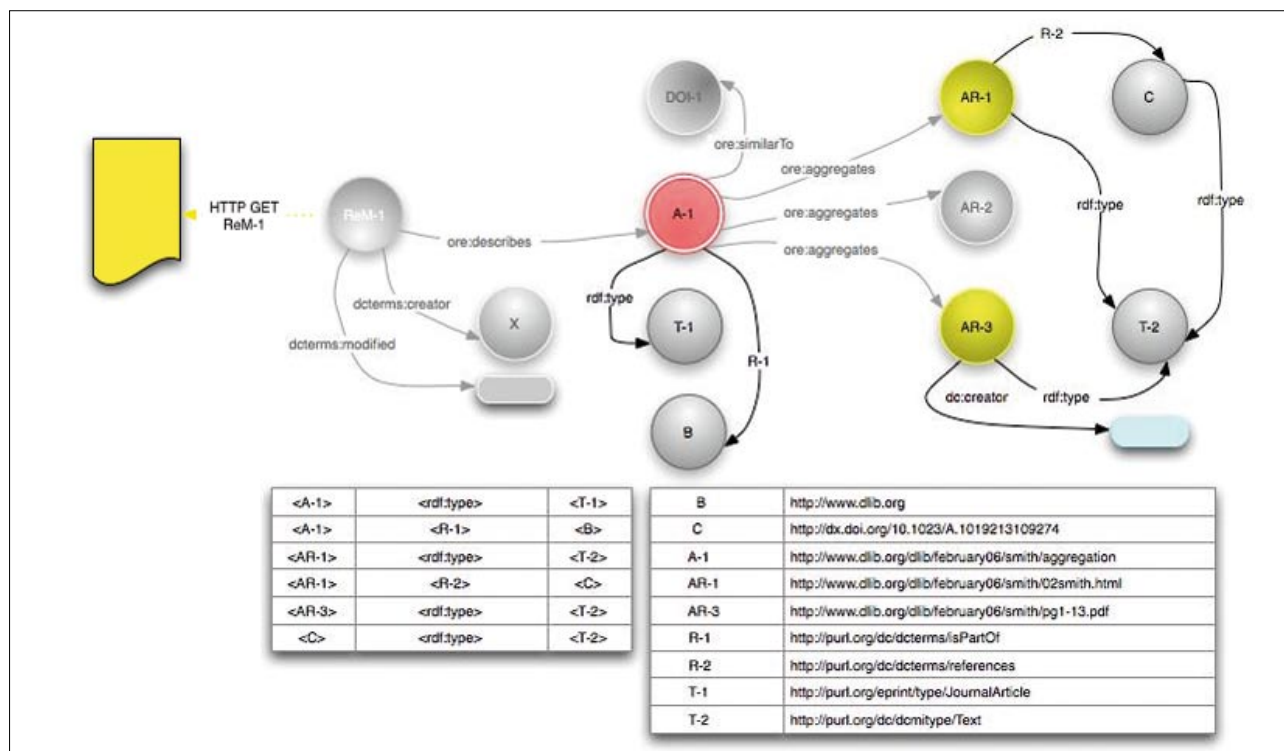
2. OAI-ORE

El proyecto OAI-ORE (*Open Access Initiative-Object Reuse and Exchange*), financiado por la *Andrew W. Mellon Foundation* y adicionalmente por la *National Science Foundation* y *Microsoft*, inicia su andadura en octubre de 2006 bajo la tutela de la OAI, estando coordinado por **Carl Lagoze** (*Cornell University Information Science*) y **Herbert van De Sompel** (*Los Alamos Research Library*).

Tras casi dos años de trabajo, el 2 de junio de 2008 se puso a disposición para revisión pública la versión beta del proyecto¹.

2.1. Objetivos y fundamentos

La implementación de este estándar precisa de las siguientes consideraciones:



Ejemplo de grafo OAI-ORE (fuente: OAI-ORE)

- Cómo puede un proveedor de servicios codificar y desglosar descripciones de agregaciones.
- Cómo puede un usuario/máquina descubrir e interpretar descripciones de agregaciones.

Para ello se deben establecer unos parámetros que posibiliten una interoperabilidad entre:

- sistemas que gestionen contenido (como repositorios), para proporcionarlo a los proveedores de servicios,
- y sistemas que aprovechen el contenido gestionado (como motores de búsqueda) para proporcionarlo al usuario, de forma que se permita una interoperabilidad más allá de los límites de los fondos de los repositorios.

OAI-ORE representa el dominio de una agregación a través de un grafo. Éste queda formado por nodos (recursos o propiedades) y arcos (relaciones).

La idea radica en la existencia de un “mapa de recurso” que describe al nodo que representa a la agregación en sí misma. Este mapa de recurso queda identificado con un *URI (URI-R)*, de forma que éste dé acceso vía http a un fichero donde en un determinado esquema de metadatos se proporcione a la persona/máquina todos los recursos y relaciones existentes en el grafo. Es decir, donde se explicita el dominio de la agregación de forma estructurada.

Ejemplos de agregaciones pueden ser:

- Una lista no ordenada de recursos (lista de favoritos).
- Una documento html multipágina.

“La especificación OAI-ORE tiene como objetivos la identificación y descripción de agregaciones de recursos web”

- Información recopilada en algún servicio 2.0, como Flickr².
- Diversas versiones de un artículo, como las ofrece ArXiv³.

- Una e-revista, formada por la agregación de artículos provenientes de distintos repositorios.

Finalmente, la idea de recursividad es fundamental. Cada material puede ser asimismo una agregación. De esta forma una agregación podría ser una revista, cuyos agregados podrían ser volúmenes, cuyos agregados serían números quienes a su vez agregarían artículos y, finalmente, éstos agregarían partes componentes como tablas, gráficas o texto.

Cada agregación podría tener objetos asociados con otro tipo de relaciones, no sólo de “parte-todo”, como descripciones bibliográficas, citas, versiones, notas, etc., lo que permitiría la inclusión y uso de datos brutos de investigación (*raw data*), tal como ya refleja la *Declaración de Berlín* (2003).

A pesar de que ya existen diversos métodos para enlazar conjuntos de datos científicos con sus publicaciones (basados en identificadores como DOI o en datos embebidos en las publicaciones mediante lenguaje de marcas), se precisa

que los elementos del agregado sean entendibles tanto por el humano como por la máquina (**Cheung et al., 2007**). Ahí es donde precisamente OAI-ORE aporta su verdadero valor.

Algunas de las ventajas de este sistema serían las siguientes:

- Los motores de búsqueda proporcionarían resultados en forma de conjuntos de recursos descritos.
- Los navegadores permitirían la navegación por agregados.
- Las descripciones de los agregados permitirían conocer las mejores estrategias de conservación para los recursos, al entender a cada uno de ellos como parte de un documento lógico superior.

2.2. Estructuración del mapa

Las especificaciones describen, de momento, tres formatos para serializar mapas de recursos: *RDF/XML*, *RDFa* y *Atom*.

Este último es una opción muy interesante, pese a que se pierda algo de expresividad. La filosofía de *Atom*, basada en la noción de *feed* al cual se le asocia un número de entradas, tiene una clara similitud con OAI-ORE. Así, un *feed* se puede hacer corresponder con una agregación y cada una de las entradas del *feed* con cada uno de los elementos agregados.

2.3. Localización

La localización de mapas de materiales, por personas o máquinas, es lo que da verdadero sentido a la especificación. El borrador OAI-ORE muestra varios métodos para llevar esto a cabo:

– Orientados a la localización de un mapa

Se basan en proporcionar la *URI-R* embebida en el recurso agregado para que la persona/máquina acceda a su serialización. Se pueden diferenciar métodos opacos al usuario (mediante el elemento *link* de html o el *http link header*), o visibles (mediante *links* en el cuerpo del html).

– Orientados a la localización de muchos mapas

Se basan en la recuperación en masa de mapas. Los métodos establecidos en la especificación de momento son: *OAI-PMH*, *Sitemaps* y redifusión de *feeds* (*rss* o *Atom*).

3. Aspectos interesantes a debatir

- Aunque la iniciativa nace en un contexto científico, ¿podrá expandirse hacia otras formas de publicación, como la periodística, con agregados como la prensa digital?

– Pese a que las especificaciones anuncian futuros desarrollos de nuevos formatos de estructuración de mapas de recursos, ¿en qué papel quedarán los esquemas de metadatos de objetos complejos ya existentes, como *Mets*?

– ¿Qué grado de solapamiento tendrá OAI-ORE con otras herramientas similares, como *Powder*⁴, desarrollado por el W3C?

– La elaboración de vocabularios específicos queda en manos de los diferentes sectores. ¿Debería la comunidad bibliotecaria definir los tipos de relaciones entre recursos basándose en la taxonomía de **Tillett** (1987)?

– El análisis automático de las descripciones de los mapas de recursos, ¿abrirá nuevos caminos en cibermetría?

– Como resulta más difícil movilizar personas que bytes, si ya es costoso que los autores ejerzan el autoarchivo, ¿qué ocurrirá cuando además deban aportar datos complementarios de sus trabajos?

– Proyectos como *Foresite*⁵ o *Scope*, que ya experimentan con OAI-ORE, ¿pasarán infravalorados o desapercibidos para el 80% de la profesión, mientras el 20% restante se lamenta?

– ¿Quién fijará las condiciones de reutilización de las agregaciones?

Para concluir, consideramos que el manejo de agregados en la Web (bajo estas especificaciones u otras que puedan surgir), es un asunto que debería ser tratado en los planes de estudio si deseamos que los futuros documentalistas salgan preparados para enfrentarse a un aspecto crucial de la realidad de su profesión, como es la documentación compleja en la Web.

4. Notas

1. La versión 1.0 fue publicada finalmente el 17 de octubre de 2008. Puede consultarse en la siguiente url: <http://www.openarchives.org/ore/1.0/toc.html>

2. <http://www.flickr.com/>

3. <http://arxiv.org/>

4. <http://www.w3.org/2007/powder/>

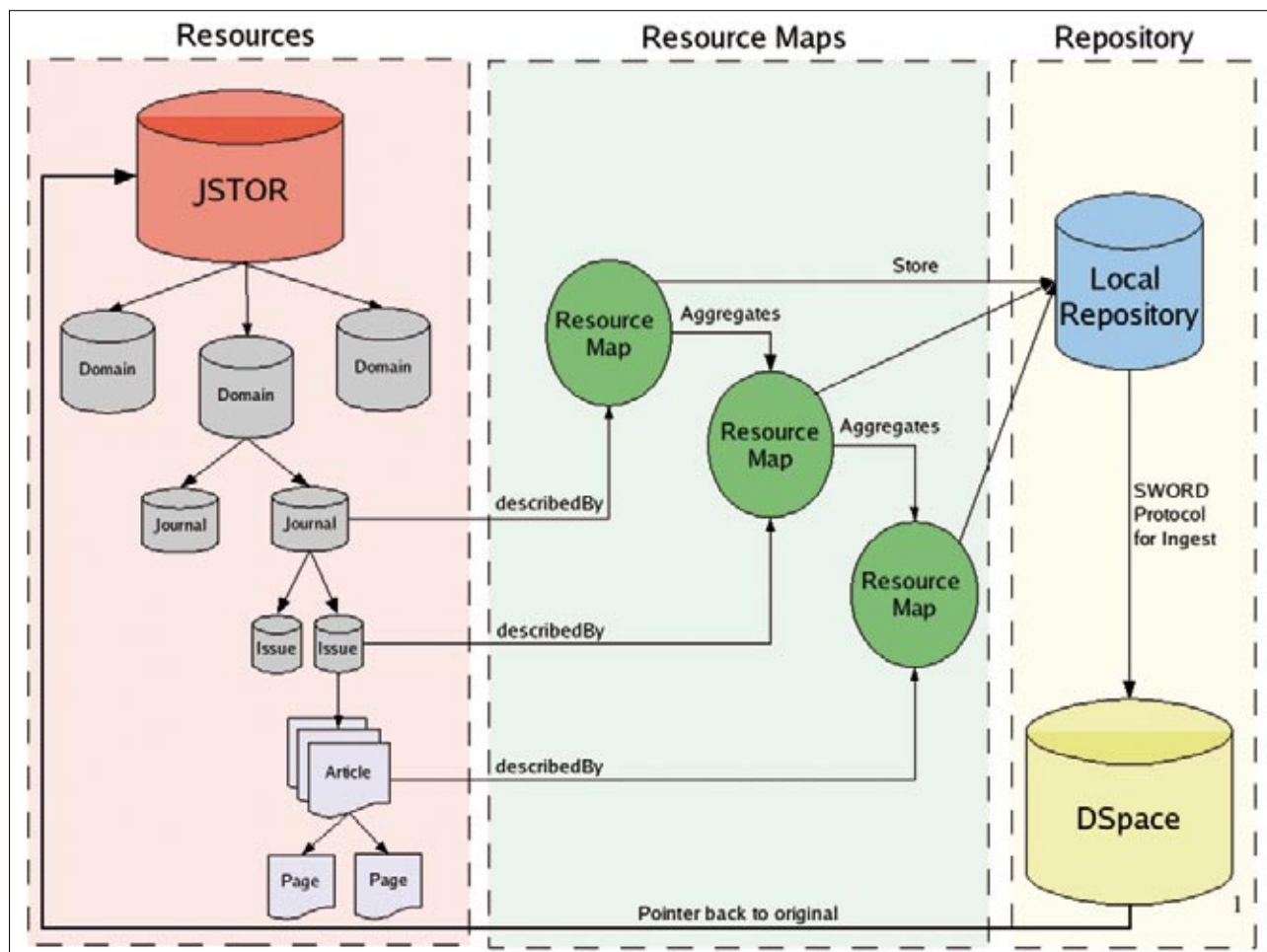
5. <http://foresite.cheshire3.org/>

5. Referencias

Berlin Declaration on open access to knowledge in the sciences and humanities. 22 octubre 2003. <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

Bertalanffy, Ludwig von. *Teoría general de los sistemas: fundamentos, desarrollos, aplicaciones*. Madrid: Fondo de cultura económica, 1976.

Cheung, Kwok; Hunter, Jane; Lashtabeg, Anna; Dren-



Esquema conceptual del proyecto Foresite

nan, John. "Scope: A scientific compound object publishing and editing system". En: *3rd intl digital curation conf "Curating our digital scientific heritage: a global collaborative challenge"*, 11-13 Dec 2007. Washington DC, USA. <http://espace.library.uq.edu.au/eserv/UQ:1513701/IDCC07.pdf>

De Sompel, Herbert van; Lagoze, Carl; Bekaert, Jeroen; Liu, Xiaoming; Payette, Sandy; Warner, Simeon. "An interoperable fabric for scholarly value chains". *D-Lib magazine*, October 2006, v. 12, n. 10. <http://dlib.org/dlib/october06/vandesompel/10vandesompel.html>

De Sompel, Herbert van; Payette, Sandy; Ericsson, John; Lagoze, Karl; Warner, Simeon. "Rethinking scholarly communication: building the system that scholars deserve". En: *D-Lib Magazine*, September 2004, v. 10, n. 9. <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>

Lagoze, Carl; De Sompel, Herbert van. Compound information objects: the OAI-ORE perspective, May 28, 2007.

<http://www.openarchives.org/ore/documents/CompoundObjects-200705.html>

Lagoze, Carl; De Sompel, Herbert van; Johnston, Pete; Nelson, Michael; Sanderson, Robert; Warner, Simeon (eds.). *ORE User guide – Abstract data model*, 02-06-2008.

<http://www.openarchives.org/ore/0.9/datamodel.html>

Lagoze, Carl; De Sompel, Herbert van; Johnston, Pete; Nelson, Michael; Sanderson, Robert; Warner, Simeon (eds.). *ORE User guide – Primer*, 11-06-2008. <http://www.openarchives.org/ore/0.9/primer.html>

Lagoze, Carl; De Sompel, Herbert van; Johnston, Pete; Nelson, Michael; Sanderson, Robert; Warner, Simeon (eds.). *ORE User guide – Resource map discovery*, 02-06-2008.

<http://www.openarchives.org/ore/0.9/discovery.html>

Tillett, Barbara B. *Bibliographic relationships: toward a conceptual structure of bibliographic information used in cataloging* [tesis]. University of California: Los Angeles, 1987.