

Hacia la Publicación Abierta de Objetos de Aprendizaje

Jacqueline Guzmán
Universidad de la República
Uruguay

Regina Motz
Universidad de la República
Uruguay

Alberto Rodrigues da Silva
Instituto Superior Técnico
Portugal

jjguzman@adinet.com

rmotz@fing.edu.uy

alberto.silva@acm.org

RESUMEN

Actualmente la publicación de datos abiertos en la web posibilita la integración de datos de forma automática, la derivación de nuevas relaciones entre datos existentes de forma independiente en la web, así como el rápido desarrollo de aplicaciones personalizadas sobre estos datos. En áreas como gobierno electrónico o bibliotecas digitales la publicación de datos abiertos viene desarrollándose activamente. Sin embargo, para eLearning no se encuentran aún suficientes datos abiertos publicados en la web. En ese sentido, este trabajo reporta una experiencia sobre la publicación de objetos de aprendizaje como datos abiertos en la web, con base en la plataforma BOA y soportada por la herramienta D2R.

Palabras Claves:

Datos Enlazados, Objetos de Aprendizaje

1. INTRODUCCION

La evolución de la Web Semántica ha ido en los últimos años hacia el concepto de una Web de Datos Enlazados o de su término en inglés: Linked Data [1,2,3]. La construcción de esta Web de Datos Enlazados se sostiene en la publicación abierta y el intercambio de grandes volúmenes de datos en la web con una tecnología de base que hace uso de URIs, RDF [12] y SPARQL [13]. El enorme éxito y la adopción generalizada del enfoque de Datos Enlazados ha llevado a la disponibilidad de grandes cantidades de datos públicos, como DBpedia, RDF WordNet o la iniciativa data.gov.uk. A pesar que el enfoque Datos Enlazados no se ha adoptado aún ampliamente en el dominio de e-Learning, se observan las posibilidades que brinda de proporcionar una integración significativa, automatizada y personalizada de los diversos recursos de aprendizaje, por ejemplo, los formales (repositorios de datos) e informales (redes sociales) que se encuentran en la web. Si bien el intercambio de recursos educativos en la web se ha convertido en práctica común, proliferando para esto una gran variedad de estándares el foco ha estado dedicado a los repositorios de e-learning interoperables basados en tecnologías semánticas. Pero la interoperabilidad de datos es aún más crucial, ya que el intercambio de recursos en línea podría hacerse independiente de su fuente, ya sea ésta un repositorio de objetos de aprendizaje, un Learning Management System, Moodle, Clarolines, etc, o Redes Sociales como Flickr, YouTube o SlideShare. El problema es que estos tipos de intercambios requieren de formatos normalizados. El uso de RDF, en particular, podría resolver este problema y permitir el

intercambio y la integración de información a través de los límites de las diversas plataformas. Como primer paso hacia la Web de Datos Enlazados para e-Learning, nuestro trabajo describe una forma de realizar la publicación de objetos de aprendizaje en RDF. Nuestro caso de estudio es el repositorio de objetos de aprendizaje BOA [5], que tiene funcionalidades extras de red social. De esta forma nuestro estudio intenta ligar la tecnología actual de la Web Semántica con los aspectos relevantes del e-Learning sobre recomendación y evaluación de los recursos educativos. Los beneficios de la publicación abierta de datos enlazados permiten por ejemplo, que la información adicional relacionada a un tema de estudio existente en Flickr o en la base de datos geográfica, se pueda recuperar según la palabra clave determinada y recomendar al usuario en forma de información adicional. Existen Algunos trabajos en curso que tratan actualmente esta problemática [6-9] pero el grado de utilización de herramientas de la web semántica en conjunto con redes sociales es todavía muy escaso. En este escenario el trabajo que presentamos en este artículo se restringe a la publicación de objetos de aprendizaje en RDF y describe las herramientas utilizadas en este proceso. En especial se está trabajando con los objetos de aprendizaje del repositorio BOA que lo hacen muy interesante por poseer metadatos de valor de recomendación asociados.

El resto de este trabajo se organiza de la siguiente forma. En la Sección 2 se presentan las herramientas para la publicación de datos abiertos en la web. En la Sección 3 se describe brevemente el repositorio de objetos de aprendizaje BOA usado como fuente de datos a publicar. En la sección 4 se presenta la experiencia de publicación realizada. Finalmente en la Sección 5 se comentan algunas conclusiones y trabajos futuros.

2. PUBLICAR DATOS ABIERTOS

Antes de decidir sobre las herramientas a utilizar para la publicación de los datos, se debe conocer el dominio de datos a publicar, su nivel de cambios, su grado de confidencialidad y por supuesto su origen. Para datos existentes en bases de datos relacionales, planillas excel o páginas web existen varias herramientas accesibles para generar su transformación a la representación RDF deseada. Los datos deben ser descritos mediante RDF [12] que es el modelo de datos de la web semántica. Existen diferentes representaciones RDF, la serialización en XML (RDF/XML) la Turtle-RDF, la N-Tuplas y la N3 son las más difundidas. Al menos debe estar disponible una representación para cada recurso. Cada recurso debe ser unívocamente identificable a través del uso de URI (*Uniform*

Resource Identifier). Se utiliza el protocolo HTTP para nombrar y resolver la ubicación de los datos identificados mediante esas URIs. Al nombrar los recursos mediante URIs, se ofrece una abstracción del lenguaje natural y así se consigue evitar ambigüedades y ofrecer una forma estandar y unívoca para referirnos a cualquier recurso. Los términos utilizados en los nombres de las URIs son seleccionados desde vocabularios consensuados u ontologías. Ejemplos de vocabularios son Dublin Core¹ para describir propiedades genéricas de recursos, FOAF² para describir personas, SIOC³ para describir comunidades online, o VoID⁴ para describir conjuntos de datos enlazados, entre otros. En el caso de objetos de aprendizaje los vocabularios más extendidos para describir sus propiedades son Dublin Core y LOM. [4]. Estos vocabularios son usados para simplificar la interconexión de datos. La idea es además aplicar un patrón de diseño de URIs de forma que usando diferentes bases de URI pero con iguales identificadores se acceda a información del mismo recurso desde diferentes URIs.

El siguiente ejemplo presentado en [15] ilustra esta situación. La BBC ha creado URIs para acceder a información de artistas que son algorítmicamente relacionadas a URIs presentes en MusicBrainz usando una clave compartida común. Los identificadores creados por MusicBrainz son contruidos a partir de un ID definido por ellos. Las URIs para el mismo artista en los diferentes sitios serían las siguientes:

BBC: www.bbc.co.uk/music/artists/a74b1b7f-71a5-4011-9441-d0b5e4122711

MusicBrainz: musicbrainz.org/artist/a74b1b7f-71a5-4011-9441-d0b5e4122711

Otra posible forma de vincular recursos de distintos conjuntos de datos enlazados es mediante el establecimiento de correspondencias de equivalencia entre los mismos, por ejemplo usando ontologías, pero preservando sus identificadores y URIs. En cualquiera de los casos el mantenimiento de las correspondencias y el monitoreo de los posibles cambios en las bases URIs es un problema que no trataremos en este trabajo.

El siguiente paso luego de diseñar el patrón de la URI a utilizar es decidir cuánta información de la que se dispone sobre un recurso, por ejemplo sobre un ejercicio interactivo en Moodle, va a ser ofrecida como dato abierto. Esto puede ser un punto muy sensible considerando las leyes de protección de datos existentes en distintos países. Por ejemplo puede resultar en violación a la ley de protección de datos personales si se publica junto con el ejercicio las calificaciones que obtuvieron los estudiantes que lo realizaron. Una vez completado este análisis los datos seleccionados se publican en un formato RDF. La forma de publicación más utilizada es generar los metadatos correspondientes a los documentos de forma dinámica cuando llega una petición para una URI concreta. El W3C ha creado la especificación SPARQL [13], que define un lenguaje de consultas para RDF. Típicamente, un servidor semántico hará público este punto de consultas a través del protocolo HTTP. Cuando el servidor recibe una petición para la URI de un recurso concreto

esta petición es transformada en una consulta SPARQL sobre los datos de ese recurso, que son devueltos al cliente serializados en RDF/XML. Una vez que hemos definido las descripciones de nuestros recursos hay que decidir dónde serán alojados. Existen organizaciones especializadas que ofrecen plataformas para este tipo de servicios, como por ejemplo Talis [10]. Para publicar y mantener nuestros datos en nuestro propio servidor hay dos diferentes opciones: almacenar nuestros datos en un formato nativo de tripletas RDF o bien generarlos dinámicamente en base a otros formatos. Por ejemplo, (i) Virtuoso⁵ que es un completo servidor con soporte para almacenar modelos RDF y publicación mediante protocolo SPARQL; (ii) D2R, que realiza la correspondencia y publica de forma semántica los datos de bases de datos relacionales existentes; o (iii) Pubby [14], herramienta que genera y gestiona las URIs de los recursos cuando ya se dispone de un punto SPARQL con datos publicados.

3. EL REPOSITORIO BOA

El sistema BOA (Bolsa de Objetos de Aprendizagem) [4] es un repositorio de objetos de aprendizaje flexible y configurable, concebido para ser aplicado a diferentes escenarios y situaciones de utilización. BOA es una plataforma web que pretende promover la colaboración entre sus usuarios. Los elementos centrales del BOA son el Objeto de Aprendizaje (OA) y el Grupo, entendido éste como la noción de “comunidad electrónica” de los usuarios de los OAs. Un Grupo agrega lógicamente un conjunto de usuarios y de OAs categorizados siguiendo una jerarquía de Temas (definidos en el nivel de cada grupo). Cada grupo tiene asociados usuarios, pudiéndoles asignar distintos roles: (1) consultor, que permite buscar y comprar, (2) miembro, las mismas atribuciones que el consultor pero además puede someter OA en ese grupo, y (3) gestor, que permite administrar los aspectos funcionales del grupo. Entre otras, es responsabilidad del gestor del grupo, la definición de una jerarquía de temas, la asignación de revisores, y la publicación de OA.

El OA contiene la siguiente información sometida por el respectivo autor: (1) el contenido, i.e. un archivo según un formato aceptado por el sistema, (e.g. PDF, DOC, PPT) y (2) los respectivos metadatos que lo describen. Los metadatos siguen el estándar Dublin Core con algunas extensiones que completan la información de los OA (e.g. El valor inicial y el valor mínimo aceptado para la venta del objeto). Adicionalmente, el OA puede también contener información extra: (1) agregada por los usuarios que lo adquirieron previamente, tales como información relativa a la clasificación del OA, comentarios, sugerencias de mejora, experiencias educativas, buenas prácticas de utilización, y (2) en caso de existir, información sometida por los revisores.

Una de las principales innovaciones del BOA, que lo distinguen de otros repositorios de OA, es la adopción de la metáfora de “bolsa de valores” que permite traducir fácilmente el valor de cada OA través del tiempo, como medida de su utilización y popularidad. Esta metáfora se concretiza por un mecanismo de créditos que permite asignar inicialmente un valor a cada OA y que periódicamente ese valor se actualice según con su popularidad. Posibilita también cuantificar la colaboración de los usuarios, no sólo por la creación y publicación de OA, sino también a través de la evaluación de OA existentes.

¹ <http://dublincore.org/>

² <http://xmlns.com/foaf/spec/>

³ <http://sioc-project.org/>

⁴ <http://semanticweb.org/wiki/VoID>

⁵ <http://virtuoso.openlinksw.com/>

Nuestro trabajo se estructura en dos fases. La primera se centra en publicar los metadatos de los objetos de aprendizaje de BOA como datos abiertos. La segunda fase atiende la publicación de los datos relativos a las funcionalidades de red social de los grupos de BOA pero manteniendo la privacidad de datos. Este trabajo es un reporte inicial de la primera fase de publicación abierta de los OA.

4. LOS OA COMO DATOS ABIERTOS

Para realizar la publicación en RDF de los OA del repositorio BOA como datos abiertos se compararon las herramientas Virtuoso y D2R [11]. Sin embargo, se encontró que Virtuoso Open tiene la limitación de no trabajar con links a repositorios de datos externos, por tal motivo no fue elegida esta herramienta para la publicación RDF. Igualmente esta versión puede ser utilizada para consultar y manipular (mediante aplicaciones o interfases) los datos RDF publicados en algún sitio, pudiendo así realizar la publicación RDF con otra herramienta y aprovechar las prestaciones de Virtuoso en cuanto a manipulación de datos. Por estos motivos, y además por la facilidad de generar un archivo de correspondencias entre RDF y la base de datos relacional se trabajó con D2R. Este archivo está escrito en un lenguaje que permite personalizar aspectos por el filtrado de información, pudiendo manejar condicionales. Estas correspondencias se realizan mediante funcionalidades brindadas por D2R (*generate-mapping*) [16], las cuales analizan el esquema de la base de datos y generan las correspondencias de la base de datos completa. Para el caso de publicación de los OA del BOA interesa especialmente la personalización de esta correspondencia por ser una de las formas de poder limitar los nodos RDF a publicar, evitando así datos sensibles o confidenciales como el rol de los usuarios dentro del BOA, entre otros aspectos. Otra característica de la personalización del archivo de mapeo es que brinda la posibilidad de vincular el conjunto de datos a publicar con vocabularios estándares. Debido a que BOA contempla los metadatos Dublin Core para la descripción de sus OAs, se vincularon las características de los OA que correspondían a elementos DC mediante la personalización del archivo de mapeo. Teniendo la correspondencia a utilizar, D2R permite levantar una instancia generada a partir de dicho mapeo. Luego de levantado el servidor D2R con la publicación de la información de BOA, se utiliza la aplicación SPARQL Explorer (SNORQL) para realizar consultas RDF sobre la información. Esta aplicación permite, mediante navegabilidad con vínculos, consultar de una manera amigable los datos publicados, o si se prefiere se puede escribir la consulta usando el lenguaje SPARQL. También permite elegir si se quiere ver en el browser el resultado de la consulta o si quiere que la salida se vuelque a un archivo XML, XML/XSL o un archivo JSON⁶ (que trata los recursos como objetos java).

En la Figura 1 se muestra uno de los principales recursos de la publicación BOA, la clase que modela los objetos de aprendizaje (*clase DO*). Además se muestran propiedades representativas de relaciones con otros recursos, como la propiedad *submitter* que vincula los OA con sus autores (*userboa*). También se pueden ver relaciones funcionales tales que dado un OA permiten establecer el valor que tienen ciertas propiedades del mismo. Estas propiedades pueden referirse a vocabularios estándares como *Title* de Dublin Core o pueden referirse a aspectos propios del *data set*

tratado como por ejemplo *currentValue* que vincula al OA con un valor numérico que cuantifica la calidad del objeto.

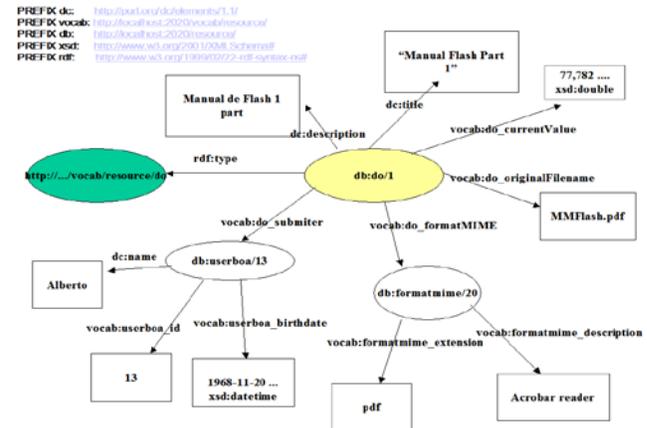


Figura 1: Vista parcial de Grafo RDF de BOA.

Si se desea por ejemplo listar desde la publicación abierta del BOA los OA que pueden estar vinculados a 'Manuales', podemos entonces escribir la siguiente consulta:

```
SELECT DISTINCT * WHERE { ?s ?p ?o
FILTER (REGEX(?o, "Manual")) }
```

SPARQL results:

s	p	o
db:image/10	vocab:image_name	"Manual de Utilização"
db:do/35	vocab:do_originalFilename	"Manual de Utilizacao.pdf"
db:do/37	vocab:do_originalFilename	"Manual de Instalacao.pdf"
db:do/166	vocab:do_originalFilename	"Manual - BOA.zip"
db:do/1	dc:title	"Manual Flash Part 1"
db:do/35	dc:title	"Manual de Utilização :Hot Potatoes versão 6.2"
db:do/37	dc:title	"Manual de Instalação do HotPotatoes"
db:do/166	dc:title	"Manual - BOA"
db:do/1	dc:description	"Manual de Flash 1 parte "
db:do/37	dc:description	"Para descarregar a última versão do programa Hot Potatoes deverá ligar acesso ao programa em regime de freeware, poderá também obter mais planos, tutoriais, etc. A figura seguinte mostra a secção de downloads de Manual."
db:do/166	dc:description	"Manual rápido para utilização do BOA"
db:do/1	vocab:do_keywords	"Flash Adobe Manual Animação"
db:do/37	vocab:do_keywords	"Manual de Instalação Construção de Materiais Didáticos com Recurso à"
db:do/38	vocab:do_keywords	"PowerPoint 2007 Manual Guia"

Observar que obtener esta información consultado una base de datos relacional hubiese requerido conocer el esquema de la base y la realización de varias consultas.

5. CONCLUSIONES

Este trabajo reporta una experiencia sobre la publicación de objetos de aprendizaje como datos abiertos en la web. Se trabajó sobre el repositorio BOA que se caracteriza además por su carácter de red social. Para la publicación se utilizó la herramienta D2R que produce un conjunto de correspondencias entre el esquema del repositorio de los objetos de aprendizaje y RDF. Considerando el carácter de red social del repositorio, se atendió especialmente que la publicación sea sólo de aquella información no sensible ni confidencial para el usuario. Este aspecto 'parcial' de la publicación se obtuvo generando un mapeo adecuado que permite acceder a los elementos que se quieren publicar, constatando que es posible hacerlo junto con la herramienta D2R. El filtro se realizó principalmente sobre clases/recursos, y sobre algunas propiedades con el objetivo de evaluar su factibilidad.

⁶ <http://json.org/json-es.html>

También se identificaron propiedades que debían referirse a vocabularios estándares como Dublin Core, y se establecieron correspondencias para algunas de estas propiedades a modo de prueba. Se obtuvo una publicación sobre la cual se pueden realizar consultas SPARQL y navegar por el modelo RDF mediante SNORQL (Aplicación SPARQL Explorer brindada por D2R). Se realizaron pruebas de consultas sobre esta colección de datos publicada. Se observa que las URIs a recursos se presentan en forma poco clara, la herramienta los identifica con una combinación de los id de las tuplas de las tablas originales. Para mejorar esto se plantea continuar trabajando en la preparación de datos RDF para el dominio eLearning, especialmente identificando aquellas propiedades que deberían referirse a LOM y FoaF. Otro de los aspectos en los que se planea trabajar es en el establecimiento de correspondencias con otros *Data Set* vía ontologías y en la administración de posibles evoluciones y cambios de los mismos. Parece interesante considerar el mantenimiento dinámico de la publicación. Cambios en los datos (valores) se reflejarían automáticamente al realizar cada consulta o navegación, pero debido a que se busca publicar solo parte de la base de datos el mapeo entre el esquema y RDF debe ser parcial, esto hace que la publicación no sea dinámica. Cambios en el esquema de la base de datos requerirá generar y personalizar el mapeo nuevamente. Por otro lado, los vínculos a vocabularios estándares como DC, FOAF o LOM, y los vínculos a otros *data set*, condicionarían a que el mantenimiento de la publicación requiriese una monitorización de los tipos de cambios, estableciendo si los cambios afectaron a los vínculos ya establecidos o si se agregaron recursos que podrían requerir agregar nuevos vínculos, situación que es recurrente en estos procesos abiertos de Linked Data. Sería interesante, implementar mecanismos de publicación de los metadatos de los OAs de forma regular (e.g. diaria) luego de una re-evaluación también regular (e.g. diaria) del valor de los OAs.

La disponibilidad lograda es parcial, por ahora es accesible desde terminales internas a la red donde se encuentra el servidor, pero una vez verificada la no violación de la ley de protección de datos, se piensa publicar con accesibilidad externa.

AGRADECIMIENTOS

Este proyecto se realiza con el apoyo del proyecto SOLITE-CYTED y del proyecto LACCIR_R1210LAC007.

6. REFERENCIAS

- [1] Christian Bizer: The Emerging Web of Linked Data. IEEE Intelligent Systems 24(5): 87-92 (2009)
- [2] Christian Bizer, Tom Heath, Tim Berners-Lee: Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst. 5(3): 1-22(2009)
- [3] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann: DBpedia - A crystallization point for the Web of Data. J. Web Sem. 7(3): 154-165 (2009)
- [4] LOM: IEEE LOM (del inglés, Learning Object Metadata) (ver IEEE LOM 2002) <http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html>
- [5] Patrícia Dinis, Alberto Rodrigues da Silva, Application Scenarios for the Learning Objects Pool, in Journal of Universal Computer Science, 15(7), pp 1455-1471, 2009.
- [6] Selber Softic, Behnam Taraghil and Wolfgang Halb. Weaving Social E-Learning Platforms into the Web of Linked Data. Proceedings of I-Know'09 and I-Semantics'09, Austria, september 2009.
- [7] Vincenzina Guzzi. SLOOP Project Sharing Learning Objects in an Open Perspective. Rustica ISBN: 8890311509, 2007.
- [8] F. Zablith, M. Fernandez, M. Rowe. The OU Linked Open Data: Production and Consumption, Proceedings of the Workshop on eLearning Approaches for the Linked Data Age at the 8th Extended Semantic Web Conference (ESWC), Heraklion, Crete (to appear) (PrePrint: www.fouad.zablith.org/docs/ESWC2011LinkedLearning.pdf)
- [9] Stokić, D.; Pata, k.; Devedžić, V.; Jovanovic, J.; Urošević, L.; Gašević, D.; Kieslinger, B.; Wild J. [2008]: Intelligent Learning Extended Organizations. TELearn 2008, Hanoi, Vietnam (intelLEO Project: <http://intelleo.eu/>).
- [10] Tiropanis, T., Davis, H., Millard, D., Weal, M. and White, S. (2009) Linked Data as a Foundation for the Deployment of Semantic Applications in Higher Education. In: SWEL'09: Ontologies and Social Semantic Web for Intelligent Educational Systems, 7 July 2009, AIED'09 Conference, Brighton, UK (ver también: <http://www.talis.com/platform/>)
- [11] Christian Bizer, Richard Cyganiak: D2RQ - Lessons Learned (Slides). Position paper at the W3C Workshop on RDF Access to Relational Databases, Cambridge, USA, October 2007.
- [12] Dean Allemang, James Hendler. Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL . Elsevier, Morgan Kaufmann Eds., 2008.
- [13] SPARQL Query Results XML Format, Dave Beckett and Jeen Broekstra (editors), W3C Recommendation, 15 January 2008.
- [14] Olaf Hartig, Christian Bizer, and Johann-Christoph Freytag: Executing SPARQL Queries over the Web of Linked Data. International Semantic Web Conference (ISWC2009), Westfields, USA, October 2009. (<http://www4.wiwiss.fu-berlin.de/pubby/>)
- [15] Leigh Dodds and Ian Davis. Linked Data Patterns. On-Line book licenced under the Creative Commons Attribution 2.0 UK. Accesible at <http://patterns.dataincubator.org>