

# Cuadros de mando para gestionar el uso de bibliotecas digitales sobre datos enlazados

María Hallo<sup>1</sup>, Sergio Luján-Mora<sup>2</sup>, Alejandro Mate Morga<sup>2</sup>

[maria.hallo@epn.edu.ec](mailto:maria.hallo@epn.edu.ec), [sergio.lujan@ua.es](mailto:sergio.lujan@ua.es), [amate@lucentia.lab.es](mailto:amate@lucentia.lab.es)

<sup>1</sup> Escuela Politécnica Nacional, Quito, Ecuador

<sup>2</sup> Universidad de Alicante, Alicante, España

DOI: [10.17013/risti.22.57-72](https://doi.org/10.17013/risti.22.57-72)

**Resumen:** Este trabajo describe un proceso para la publicación de cuadros de mando para evaluar el uso de las bibliotecas digitales mediante tecnologías de datos enlazados. Actualmente, los indicadores de los cuadros de mando publicados están dispersos, lo que dificulta la reutilización y combinación con otros indicadores para tomar mejores decisiones. El desarrollo de cuadros de mando usando tecnologías de datos enlazados ayudará a definir mejores indicadores para monitorear el uso de las bibliotecas digitales determinando los grados de avance o retroceso respecto a los objetivos planteados. Adicionalmente, los indicadores podrán ser enriquecidos con información externa. Este trabajo incluye un estudio de los usos actuales de los datos enlazados en las bibliotecas digitales y los procedimientos para la generación y publicación de cuadros de mando en la web semántica. Los resultados de la investigación se han probado con metadatos para un tipo de biblioteca digital, una revista científica digital de acceso abierto. Se han incluido nuevas características en el sistema sin afectar a las estructuras de datos existentes.

**Palabras clave:** Datos enlazados; web semántica; cuadros de mando; bibliotecas digitales.

## *Dashboards for management of use of digital libraries on linked data*

**Abstract:** This paper provides a process for publishing scorecards to evaluate the use of digital libraries using Linked Data technologies. Currently, the published scorecard indicators are dispersed, which makes difficult the reuse and easy combination with other indicators to make better decisions. The development of scorecards using Linked Data will help to define better indicators to monitor the use of digital libraries determining the progress level towards the objectives. Moreover, the indicators can be enriched with external information. This work includes a study of current uses of Linked Data in digital libraries, and the procedures for generating and publishing scorecards in the Semantic Web. The research results have been tested with metadata for a specific type of digital library, a digital open access journal. New features have been included without affecting existing data structures.

**Keywords:** Linked data; Semantic web; Scorecards; Digital Libraries.

## 1. Introducción

Este artículo propone un proceso para la integración de funcionalidades orientadas al almacenamiento y publicación en la web semántica de cuadros de mando, facilitando la evaluación del uso de las bibliotecas digitales. La evaluación de bibliotecas digitales es muy importante, no solo por las restricciones presupuestarias que obligan a la optimización y priorización de servicios, sino también por el interés de los promotores en ofrecer servicios de calidad.

Los cuadros de mando o *dashboards* son una herramienta comúnmente utilizada en entornos de inteligencia de negocio para monitorear el estado de la empresa. (Allio, 2012); (Eckerson, 2010); (Freixo & Rocha, 2014); (Yigitbasioglu & Velcu, 2012). De esta manera, pueden detectarse problemas y corregir errores en los procesos empresariales. Son especialmente útiles para llevar un seguimiento adecuado de los objetivos estratégicos en un negocio u organización (Banker, Chang, & Pizzini, 2004). Estos cuadros de mando se pueden preparar a partir de metadatos enlazados para enriquecer los análisis y publicar sus datos en la web semántica para facilitar consultas más complejas.

La web semántica es una extensión de la *World Wide Web* en la que el significado de la información y de los servicios está definido en un formato que permite tanto la consulta de información por parte de humanos como de los computadores (Ontology Engineering Group, 2015).

Las tecnologías de datos enlazados son utilizadas en la web semántica para vincular los datos distribuidos en la Web, permitiendo un paso de la Web de documentos a la Web de datos enlazados (Library Linked Data Incubator Group, 2011). Las nuevas funcionalidades de explotación de enlaces se pueden integrar en las bibliotecas digitales sin afectar las estructuras ya existentes.

El trabajo presentado en este artículo se inició con la revisión de las tecnologías de datos enlazados más utilizadas y sus aplicaciones en el gobierno electrónico. Por otra parte, se analizó el uso de datos enlazados en un grupo de bibliotecas digitales importantes por su uso a nivel mundial. El estudio continuó con la definición y aplicación de guías de publicación de cuadros de mando para la evaluación del uso de bibliotecas digitales en la web semántica.

Los lineamientos definidos en este estudio cubren las fases de generación, almacenamiento y publicación en la web semántica de metadatos de registros bibliográficos e indicadores obtenidos de cuadros de mando. Nuestra propuesta apoya a las funcionalidades básicas de las bibliotecas digitales, que son la búsqueda y recuperación de documentos junto a la navegación sobre documentos relacionados y a la evaluación del uso de las bibliotecas digitales.

Este artículo está estructurado como sigue: la sección 2 presenta un resumen de tecnologías de datos enlazados y un estudio de los usos de datos enlazados en importantes bibliotecas digitales; la sección 3 describe la importancia de la evaluación del uso de bibliotecas digitales utilizando cuadros de mando y datos enlazados; la sección 4 presenta una propuesta de proceso para definir y publicar un cuadro de mando para la evaluación

de bibliotecas digitales usando el formato RDF (*Resource Description Framework*); finalmente, la sección 5 describe las conclusiones y los trabajos futuros.

## 2. Trabajos relacionados

### 2.1. Las tecnologías de datos enlazados y sus aplicaciones

El término datos enlazados se refiere a un conjunto de mejores prácticas para publicar y enlazar datos estructurados en la Web de tal forma que puedan ser accesibles por humanos y computadores (Health, 2011). Está basado en URI (*Uniform Resource Identifiers*) y en especificaciones RDF. Un URI es un identificador global para recursos web. RDF es un modelo estándar para especificación de datos en la Web que permite hacer declaraciones sobre los recursos en forma de expresiones sujeto-predicado objeto, lo cual posibilita organizar bases de datos de conocimiento y usar lenguajes de consulta con mayor capacidad de inferencia como SPARQL<sup>1</sup>.

En el año 2006, Tim Berners-Lee escribió una nota de diseño en la que proponía soluciones a los problemas que impedían el enlace de los datos, mediante la aplicación de los siguientes principios (Tim Berners-Lee, 2006):

- Usar URI como nombres de recursos.
- Usar HTTP URI de manera que la gente pueda buscar esos nombres.
- Presentar información útil usando los estándares RDF o SPARQL, cuando alguien busque un URI.
- Incluir enlaces a otros URI de manera que la gente pueda encontrar más recursos relacionados.

Hay reportes de desarrollo de nuevas técnicas que tienen un gran potencial para publicar y consumir datos enlazados en la web semántica (Health, 2011); (Hallo et al, 2012). En el organismo internacional W3C (*World Wide Web Consortium*), se han creado grupos especializados en varias tecnologías como RDF, SPARQL y OWL para coordinar su desarrollo (Hendler, 2009). La adopción de las mejores prácticas de datos enlazados ha contribuido a que se extienda el espacio global de datos conectando datos de diversos dominios.

A partir del 2009 se ha producido un avance más rápido en el desarrollo de las tecnologías de datos enlazados. Se han publicado y usado algunas especificaciones del W3C tales como: SPARQL, GRDDL, RDFa, VoID; se ha formado la comunidad del proyecto *Linking Open Data* y se observa una creciente actividad en gobierno electrónico con numerosos catálogos de datos publicados y aplicaciones con funcionalidades específicas orientadas a dominios que combinan datos de varias fuentes de datos enlazados (Ontology Engineering Group, 2015).

Para publicación de datos enlazados se han desarrollado varias herramientas tales como *D2R server*, *Open Link Virtuoso*, *Talis*, *Sésame*, *Jena* y *4Store*, entre otros.

<sup>1</sup> <https://www.w3.org/TR/rdf-sparql-query/>

Estas herramientas permiten el almacenamiento de datos en formato RDF y la consulta usando el lenguaje estándar SPARQL.

Para consumo de datos se han desarrollado navegadores especializados tales como *Disco*, *Tabulator*, *Marbles*, etc. Por otra parte, se han desarrollado motores de búsqueda de datos enlazados tales como *Sig.ma*, *Falcons* y *Watson*, entre otros (Hartig, Sequeda, et al., 2010). Algunos gobiernos están usando estas tecnologías.

Las experiencias impulsadas por el grupo de trabajo de datos enlazados del Reino Unido, han demostrado ser de mucha utilidad para encontrar información relacionada a un caso dado, a un tipo de dominio o información histórica. Sin embargo, aún no existen patrones de publicación y consumo simples para una aplicación a gran escala a nivel de gobierno. Se continúan desarrollando alternativas de solución que faciliten la implementación de esta tecnología como es el caso del desarrollo de interfaces de programación de aplicaciones para facilitar el consumo de datos enlazados (Shadbolt, O'Hara, Gibbins & Berners-Lee, 2012).

Se ha reportado que en las administraciones públicas en general las aplicaciones son desarrolladas para cubrir servicios de agencias administrativas jerárquicas, mientras que los ciudadanos tienen necesidades horizontales de acceso a servicios de múltiples agencias (Hallo et al., 2012). Las tecnologías de datos enlazados posibilitan satisfacer los requerimientos de integración horizontal y búsquedas semánticas de aplicaciones de administración electrónica, que algunos países ya empiezan a utilizar.

La estrategia de publicar datos a la mayor brevedad posible y paralelamente evolucionar hacia un modelo cercano a la web semántica es la escogida por algunas administraciones como Estados Unidos, Reino Unido o España (Sheridan, 2010). El hecho de poder construir capas semánticas sobre la Web actual facilita el trabajo en esta dirección.

## 2.2. Datos enlazados y bibliotecas digitales

En los tiempos actuales las bibliotecas están dando importancia a la web semántica en una variedad de formas. Para las bibliotecas digitales se crean modelos de metadatos, y publican datos enlazados desde archivos de autoridades, catálogos bibliográficos e información de proyectos digitales. Además, se ha transformado a RDF la información extraída de otros proyectos tales como la Wikipedia originando la Dbpedia (Yang et al., 2015). El uso de datos enlazados en las bibliotecas digitales fue analizado para las implementaciones más importantes expuestas en el reporte del grupo *Library Linked Data Incubator* de la W3C en el 2011.

Las bibliotecas analizadas fueron:

- Biblioteca Nacional de Francia (data.bnf.fr).
- Biblioteca Nacional de España (data.bne.es).
- Biblioteca Británica (www.bl.uk).
- Biblioteca Europea de la Unión Europea (data.europeana.eu).
- Biblioteca del Congreso de Estados Unidos (www.loc.gov).

La revisión bibliográfica se realizó siguiendo parcialmente una guía metodológica de revisión sistemática de la literatura (Kitcherman et al, 2009). Se inició con la definición

de las preguntas de investigación sobre aplicaciones de datos enlazados en bibliotecas digitales: ¿cuáles son los principales problemas a los que se enfrentan las bibliotecas digitales?, ¿qué beneficios pueden obtener de la web semántica y la publicación de cuadros de mando?, ¿cuáles son los principales vocabularios y ontologías usados?, ¿cuáles son las tendencias futuras de investigación?

En segundo lugar se definió el proceso de búsqueda que cubrió publicaciones realizadas en los últimos cinco años. Se realizaron las búsquedas en Google Scholar, Scopus, ACM Digital Library, Springer, DBLP, IEEE Xplorer. Los términos de búsqueda utilizados fueron *Linked Data* y el nombre de cada biblioteca.

Los datos extraídos de cada estudio fueron enfocados a responder las preguntas de investigación y sintetizados. Se resumen a continuación los principales resultados obtenidos.

Las bibliotecas nacionales analizadas usan diferentes arquitecturas para obtener y publicar datos enlazados. Los vocabularios y ontologías utilizados varían en cada implementación, pero es posible establecer las equivalencias y publicar las relaciones respetando los modelos de datos de origen.

Los vocabularios y ontologías más usados son: DUBLIN CORE, BIBO, BIO, FOAF, FRBR, FRAD, FRASAD, IFLA, ISBD, INTERMARC, MADS / RDF, XML EAD, OAI\_ORE, RDF, RDF SCHEMA, ORG, OWL, RDA, SKOS, WGS84 y EVENT.

Los principales beneficios reportados de la publicación y uso de datos enlazados de las bibliotecas digitales son:

- Mejora la visibilidad de los datos.
- Facilita establecer enlaces a otros datos.
- Posibilita la anotación colaborativa de recursos.
- Mejora la interoperabilidad sin afectar las fuentes de datos.
- Permite descubrir nuevas relaciones.
- Soporta razonamiento e inferencia sobre los datos, especialmente por sistemas basados en el conocimiento.

Este estudio también detectó que existen problemas pendientes de resolver para utilizar las tecnologías de datos enlazados en bibliotecas digitales:

- Necesidad de herramientas de apoyo.
- Falta de aplicaciones que consuman datos enlazados.
- Ausencia de acuerdos para proveer los datos.
- Falta de mecanismos de control de calidad de datos.
- Poco personal técnico con conocimientos de estas nuevas tecnologías.

Además se espera en el futuro:

- Nuevos conjuntos de datos aportados por la comunidad.
- Lograr una participación ciudadana en enriquecer los conjuntos de datos mediante procesos de anotación de datos.
- Más aplicaciones que consuman los datos enlazados.
- Enriquecer los catálogos en línea de las bibliotecas con recomendaciones basadas en la popularidad de un ítem y en la actividad de un usuario.

Por último, como resultado de este estudio se pudo observar que hay pocos reportes sobre evaluación de uso de las bibliotecas digitales utilizando tecnologías de datos enlazados.

### 3. Evaluación del uso de bibliotecas digitales

La forma de evaluación del uso de bibliotecas digitales varía en las diferentes instituciones. Algunas usan criterios tradicionales tales como precisión, tiempo de búsqueda, tasa de errores, etc. Pocas evalúan el beneficio que proporcionan las bibliotecas digitales a los usuarios.

Un cuadro de mando es una herramienta que puede ser usada para evaluar las estrategias y productividad de una organización desde diferentes perspectivas, pero pocas bibliotecas digitales los usan (Hallo, Luján-Mora & Mate, 2015).

Los principales elementos de un cuadro de mando para monitorear los objetivos estratégicos de una organización se presentan en la Tabla 1. Para cada objetivo estratégico se definen indicadores con sus valores actuales y esperados. De esta forma se puede medir en el tiempo la variación entre el valor medido y el valor esperado y tomar decisiones para los casos críticos. Un caso especial de cuadro de mando es el cuadro de mando integral en el que se definen objetivos e indicadores en cuatro perspectivas: clientes, finanzas, procesos y aprendizaje. Los cuadros de mando pueden elaborarse en forma interactiva e incremental según las prioridades de análisis que se definan. Por otra parte el seguimiento de los indicadores de un cuadro de mando es fundamental para la gestión del conocimiento organizacional (Illescas, Sanchez-Segura, & Canziani, 2015).

Indicadores de rendimiento	Valor actual	Valor esperado	Tiempo
Número de usuarios	1000 visitas por mes	10000 visitas por mes	6 años

Tabla 1 – Fragmento de cuadro de mando para revistas de acceso abierto

Las tecnologías de datos e.nlazados permiten agregar información a los indicadores de rendimiento enlazándolos a otros conjuntos de datos relacionados que están publicados en la Web de datos. Por ejemplo, si por una parte tenemos información del número de visitas por palabra clave y por localización, y por otra parte tenemos información del número de pacientes por enfermedad y por localización, podemos encontrar si la biblioteca digital está contribuyendo con información sobre una enfermedad específica. Hay estudios que permiten estimar epidemias como la de la gripe usando datos disponibles en Google (Yang, 2015). Existen además herramientas que permiten la extracción semántica de información para obtener conjuntos de datos útiles para nuevos enlaces (Guillén, Lloret, & Gutiérrez, 2016).

### 4. Publicación de un cuadro de mando en RDF

El proceso de publicación de cuadros de mando en RDF propuesto consta de seis etapas:

- Modelado del cuadro de mando.
- Modelado del *datamart*.
- Modelado de datos RDF.
- Generación RDF.
- Enlazamiento.
- Publicación.

La Figura 1 presenta el proceso de publicación de cuadros de mando en la web semántica con sus diferentes etapas. Para cada etapa se definen actividades concretas que guían todo el proceso y ayudan a asegurar la correcta publicación de los cuadros de mando.

En las siguientes secciones se describe cada etapa del proceso. Como caso de estudio, el método se aplicó a la Revista Politécnica, revista científica digital de acceso abierto de la Escuela Politécnica Nacional de Quito, Ecuador. El software de administración de esta revista es *Open Journal System (OJS)*.

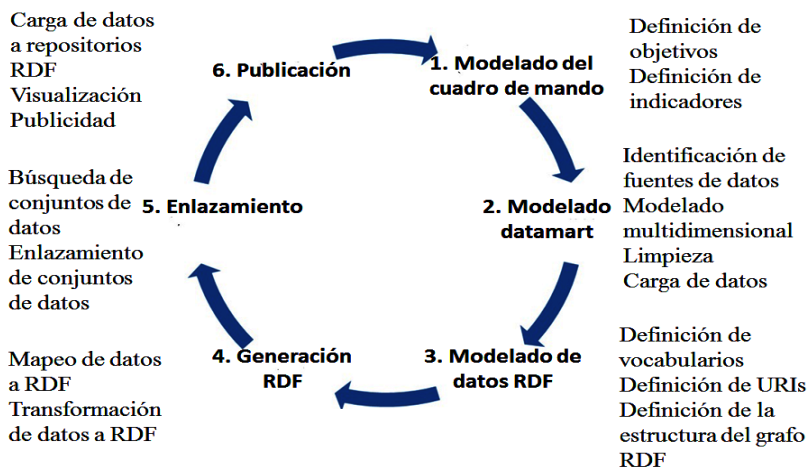


Figura 1 – Proceso de publicación de datos de cuadros de mando en RDF

#### 4.1. Modelado del cuadro de mando

Para modelar un cuadro de mando orientado a evaluar las estrategias y rendimientos de una organización se definen y priorizan las perspectivas de análisis. Para cada perspectiva se deben definir los objetivos estratégicos y los indicadores de rendimiento.

##### **Definición de objetivos**

En esta etapa se definen los objetivos estratégicos de la perspectiva en análisis.

Para el caso en estudio se definieron los objetivos estratégicos orientados al uso de las bibliotecas digitales:



- Atraer a los usuarios.
- Retener a los usuarios.
- Convertir a los usuarios en clientes.
- Comprometer a los usuarios.

### **Definición de indicadores**

Para cada objetivo se deben definir indicadores de rendimiento que permitan medir el cumplimiento de los objetivos.

Los indicadores del cuadro de mando propuesto para medir el uso de bibliotecas digitales se presentan en la Tabla 2 y son un subconjunto de los presentados en los estándares ISO 2789:2013 (ISO, 2013) e ISO 11620:2014 (ISO, 2014) para evaluar el uso de documentos electrónicos. Estos indicadores fueron escogidos a partir de entrevistas con bibliotecarios, autoridades locales y evaluaciones de los datos disponibles. Los indicadores fueron agrupados en varias categorías dependiendo del tópico que se quería medir.

Las categorías de los indicadores permiten diseñar vistas detalladas del cuadro de mando focalizadas cada una en un tópico concreto. Así, mientras el cuadro de mando permite obtener una visión global del estado de la biblioteca digital, las vistas detalladas permiten a los gestores entender qué es lo que está ocurriendo en relación a un objetivo de la biblioteca, dónde se está teniendo éxito y dónde se está fallando. Los indicadores deben analizarse por varios criterios o dimensiones.

Las dimensiones de análisis escogidas para el ejemplo presentado en este artículo fueron:

D1: Tiempo.

D2: Documento.

D3: Autor.

D4: Localización geográfica del usuario.

D5: Palabras clave.

D6: Objetivo estratégico.

<b>Atraer usuarios</b>	<b>Convertir usuarios en clientes</b>
<ul style="list-style-type: none"> <li>• Número de visitas</li> <li>• Número de documentos digitales almacenados</li> <li>• Número de documentos digitales añadidos</li> </ul>	<ul style="list-style-type: none"> <li>• Número de descargas</li> <li>• Número de descargas por documento</li> <li>• Porcentaje de items no usados</li> </ul>
<b>Retener usuarios</b>	<b>Comprometer</b>
<ul style="list-style-type: none"> <li>• Porcentaje de satisfacción del usuario</li> <li>• Visitas por usuario</li> </ul>	<ul style="list-style-type: none"> <li>• Número de anotaciones</li> </ul>

Tabla 2 – Indicadores de rendimiento



Las dimensiones pueden ser compartidas para varios cuadros de mando dependiendo de la granularidad de los indicadores y el alcance del análisis. Por ejemplo, para un cuadro de mando asociado con retener usuarios sería interesante definir como dimensión de análisis Usuario con atributos tales como género y profesión entre otros.

#### 4.2. Modelado del *datamart*

Un *datamart* es un almacén de datos orientado a un tema o área del negocio. En nuestra propuesta cada *datamart* se desarrolla siguiendo los pasos indicados a continuación:

##### *Identificación de fuentes de datos*

En esta etapa se identifican las fuentes de datos para los indicadores y las dimensiones de análisis definidos en el cuadro de mando.

El primer *datamart* fue realizado para el análisis del número de visitas dentro del conjunto de indicadores para atraer usuarios. Los datos fueron tomados de la base de datos del sistema OJS del caso de estudio.

##### *Modelado multidimensional*

El *datamart* es modelado usando un modelo multidimensional de datos. Este modelo se representa mediante una tabla de medidas conectadas a tablas de dimensiones formando una estructura en copo de nieve (*snowflake*) (Luján & Tujillo, 2003). Esta estructura está orientada a facilitar las posteriores consultas analíticas.

El *datamart* del caso de estudio fue modelado utilizando una estructura en copo de nieve, seleccionando los indicadores que comparten el mismo grado de granularidad.

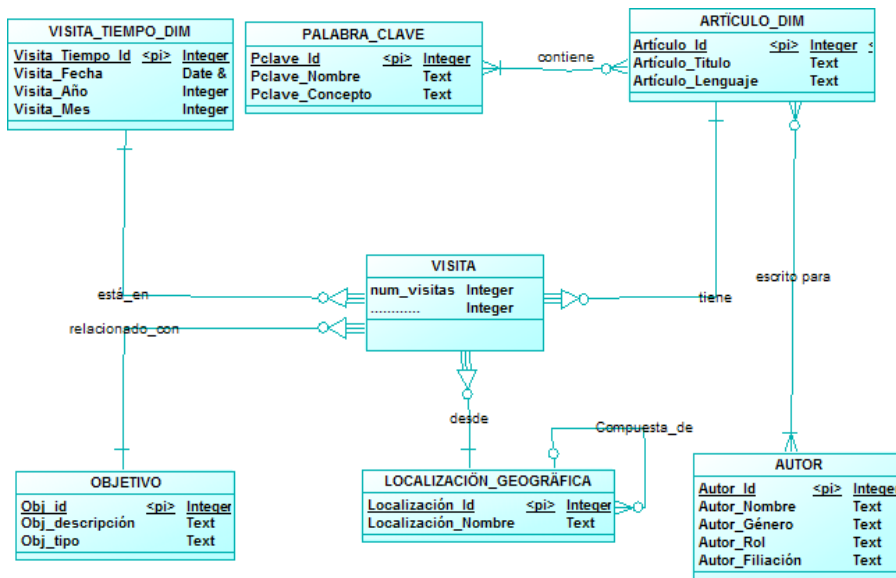


Figura 2 – Modelo multidimensional del *datamart* para el caso de estudio

En la Figura 2 se presenta el modelo multidimensional transformado al modelo relacional. Este modelo consta de dos tipos de elementos: las dimensiones de análisis (VISTA\_TIEMPO\_DIM, ARTICULO\_DIM, LOCALIZACIÓN\_GEOGRÁFICA y OBJETIVO) y los indicadores de análisis contenidos en la entidad VISITA. Las dimensiones de análisis, entre ellas OBJETIVO, se compartirán entre los diferentes *datamarts*, lo cual hace posible analizar el grado de cumplimiento de los objetivos estratégicos.

### ***Extracción, transformación y carga de datos***

Esta etapa contiene las actividades de extracción de datos de fuentes de datos heterogéneas, la transformación (conversión, limpieza, reducción de inconsistencias, normalización, etc.) y la carga de datos desde las fuentes de datos al *datamart* (Luján & Palomar, 2001).

Los metadatos fueron extraídos de la base de datos de la biblioteca digital y sometidos a un proceso de limpieza de datos y carga de datos al *datamart* usando *Spoon*<sup>2</sup>, que es el diseñador gráfico de transformaciones de la herramienta *ETL Pentaho Data Integration*, para realizar las tareas de extracción, transformación y carga de datos desde las fuentes de datos al modelo multidimensional. Otra alternativa de extracción de datos probada fue utilizar OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), un protocolo usado para la extracción y transmisión de metadatos por internet (Maali & Cyganiak, 2012).

### **4.3. Modelado de datos RDF**

El modelado de datos RDF tiene como objetivo representar los elementos del modelo multidimensional en RDF, lo cual permite referenciar esta información en la Web mejorando la interoperabilidad entre aplicaciones.

Las actividades que comprende el modelado de datos en RDF son:

- Definición de vocabularios.
- Definición de los URI.
- Definición de la estructura del grafo RDF.

#### ***Definición de vocabularios***

El objetivo de esta actividad es diseñar un vocabulario para describir las fuentes de datos en RDF.

En este trabajo se han reusado algunos vocabularios y ontologías existentes, tales como:

- *RDF data cube* para representar datos multidimensionales.
- *Dublin Core*, un conjunto de términos para describir recursos web. Permite identificar datos como el título, el autor o la fecha de creación de un recurso.
- BIBO (*The Bibliographic Ontology*) proporciona recursos para describir recursos bibliográficos.
- FOAF (*Friend of a Friend*) es una ontología para describir personas y sus relaciones con otras personas en RDF.

<sup>2</sup> <http://pentaho.almacen-datos.com/kettle-spoon.html>

- SKOS (*Simple Knowledge Organization System*) es un conjunto de términos para compartir y encadenar conceptos.
- VoID (*Vocabulary of Interlinked Datasets*) contiene términos para expresar metadatos de conjuntos de datos en RDF.

### Definición de los URI

Para poder ser referenciados, cada elemento del modelo multidimensional debe ser identificado por un URI. Los URI de los elementos del *datamart* del caso de estudio se definieron con la siguiente estructura:

Dimensiones:

{URI base}/cubos/{nombre del cubo}/prop/ {nombre de la dimensión}

Por ejemplo: [http://opendata.epn.edu.ec/cubos/revista\\_epn/prop/titulo](http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo)

Medidas:

{URI base}/cubos/{nombre del cubo}/prop/ {nombre de la medida}

Por ejemplo: [http://opendata.epn.edu.ec/cubos/revista\\_epn/prop/visita](http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita)

Definición de la estructura del grafo RDF

En esta actividad se elabora la estructura del grafo definiendo sus componentes: dimensiones y medidas con sus respectivos URI. Para el caso de estudio la estructura del grafo RDF fue definida usando la extensión RDF Refine de Open Refine (Maali & Cyganiak, 2012).

La Figura 3 presenta un fragmento del proceso de definición de la estructura de grafo RDF con sus componentes: dimensiones de análisis y medidas.

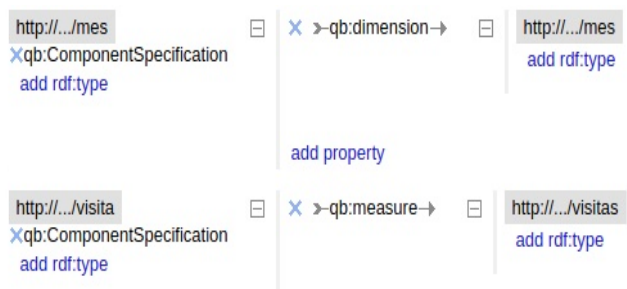


Figura 3 – Fragmento de la definición de la estructura del grafo RDF usando RDF Refine

El prefijo qb corresponde al espacio de nombres del vocabulario usado para representar los elementos de un modelo multidimensional en RDF<sup>3</sup>. Los puntos en los URI corresponden a la estructura ya indicada en esta sección.

<sup>3</sup> <http://purl.org/linked-data/cube#>

#### 4.4. Generación RDF

Esta etapa del proceso comprende las actividades de mapeo de fuentes de datos a conceptos de RDF y la transformación en RDF.

En el caso de estudio el mapeo y la transformación de datos a RDF fueron realizados usando Open Refine. La Figura 4 presenta un fragmento de triples RDF generadas en Turtle. Turtle es un formato para expresar datos en RDF que presenta la información de las tripletas RDF en forma abreviada, lo cual es más legible y fácil de editar que otros formatos como RDF/XML.

```
ds:obs1 a qb:Observation;
  prop:título "Aplicaciones de Procesamiento de Lenguaje Natural";
  prop:visitas "250"^^xsd:int;
```

Figura 4 – Fragmento de triples RDF generadas en Turtle

Unos ejemplos de tripletas representadas en la Figura 4 son:

```
<ds:obs1><a><qb:Observation>
<ds:obs1><prop:título><"Aplicaciones de Procesamiento de Lenguaje Natural">
<ds:obs1><prop:visitas><"250"^^xsd:int;>
```

El símbolo ^^xsd se usa para definir tipos de datos numéricos o fechas.

La Tabla 3 contiene los espacios de nombres representados en los prefijos del ejemplo anterior.

Prefijo	URI
ds	http://opendata.epn.edu.ec/cubos/revista_epn/dataset/
prop	http://opendata.epn.edu.ec/cubos/revista_epn/prop/
qb	http://purl/linkeddata/cube#

Tabla 3 – Espacios de nombres

#### 4.5. Enlazamiento

El objetivo de esta actividad es mejorar la conectividad con conjuntos de datos externos.

Los pasos en esta actividad son:

- Búsqueda y selección de conjuntos de datos externos. Se realizó esta búsqueda a partir de conjuntos de datos contenidos en el sitio web datahub.io<sup>4</sup>.

<sup>4</sup> <https://datahub.io/>

- Creación de enlaces a conjuntos de datos externos. Para la creación de estos enlaces se usó el software Silk<sup>5</sup> que permite entre otros: especificar reglas de enlazamiento, generación de enlaces en RDF, limpieza y transformación de datos.

#### 4.6. Publicación

El objetivo de esta actividad es hacer los conjuntos de datos RDF disponibles a los usuarios en la Web.

En el caso de estudio se utilizó para el almacenamiento de datos el software OpenLink Virtuoso<sup>6</sup> que contiene funcionalidades para manejo de bases de datos relacionales, de repositorios RDF y de datos en formato XML. Para la visualización de datos se utilizó CubeViz<sup>7</sup> y Ontowiki<sup>8</sup>. CubeViz es un navegador para datos estadísticos representados con el vocabulario RDF Data Cube, además permite filtrar en forma interactiva los indicadores para su visualización. Ontowiki permite la administración del conocimiento mediante interfaces que posibilitan manipular clases, propiedades y recursos además permite importar y exportar datos en formato RDF. Para la visualización de cubos de datos tiene el módulo CubeViz.

La Figura 5 presenta un ejemplo de consulta sobre el cubo de visitas de la revista digital tomada como caso de estudio almacenado en formato RDF. En la Figura 5 se pueden observar en el eje vertical los títulos de los artículos y en el eje horizontal se representa el número de visitas en el período de tiempo seleccionado (en este ejemplo un mes).

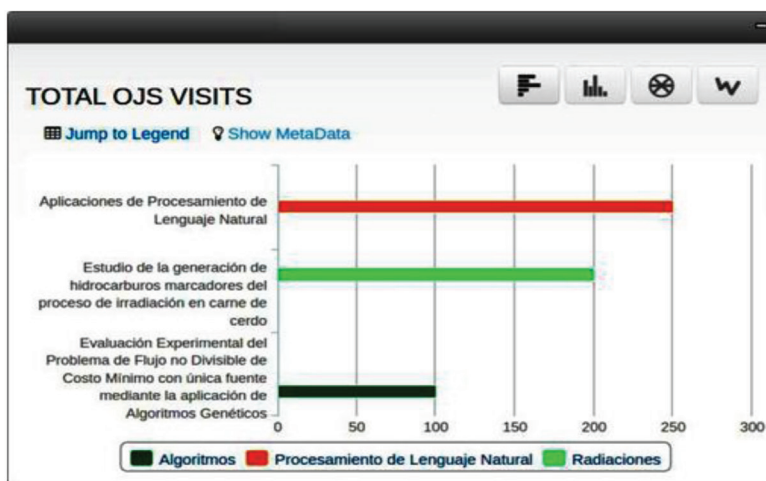


Figura 5 – Ejemplo de consulta sobre el cubo de visitas del OJS

<sup>5</sup> <http://silkframework.org/>

<sup>6</sup> <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/>

<sup>7</sup> <http://cubeviz.aksw.org/>

<sup>8</sup> <http://aksw.org/Projects/OntoWiki.html>

## 5. Conclusiones y trabajos futuros

El desarrollo de un proceso para la evaluación del uso de documentos de bibliotecas digitales usando cuadros de mando y tecnologías de datos enlazados ha permitido la generación y publicación de conjuntos de datos en RDF, lo cual permite obtener mejores indicadores para el proceso de toma de decisiones en las bibliotecas digitales.

Se ha definido una clasificación de indicadores de uso de las bibliotecas digitales y sus dimensiones de análisis que permite desarrollar los cuadros de mando de manera iterativa e incremental, modelarlos con RDF y visualizarlos facilitando la evaluación del uso de bibliotecas digitales.

Además, los elementos del cuadro de mando han sido representados como recursos, permitiendo que puedan ser enriquecidos con información externa mejorando los tipos de análisis que puedan realizarse.

Los resultados de la investigación se han aplicado a una revista científica digital, generándose conjuntos de metadatos en RDF y publicándolos en el punto SPARQL asociado al sitio web. Por otra parte se han generado en RDF conjuntos de datos de cuadros de mando con indicadores de uso de la revista analizada. Estos conjuntos de datos con sus metadatos se han publicado en datahub.io. El método desarrollado ha sido aplicado con éxito observándose un incremento en las visitas al sitio de la revista usada como caso de estudio, facilitando el seguimiento de los objetivos marcados por la revista digital.

Para el futuro se plantea: ampliar el uso de SKOS para enlazar temas y disciplinas con la definición de perfiles de usuarios y patrones de búsqueda; ampliar el modelo de datos con actividades del proceso de publicación; integrar información de investigación usando tecnologías de datos enlazados; obtener datos anotados colaborativamente; implementar módulos de anotación semántica; añadir nuevas dimensiones de análisis como tipo de recurso, tópicos, usuarios y finalmente continuar con las validaciones prácticas del modelo presentado.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad de España (MINECO/FEDER), bajo el proyecto SEQUOIA-UA (*Management requirements and methodology for Big Data analytics*) (TIN201563502C33R) y por la Escuela Politécnica Nacional de Quito, Ecuador en el marco del proyecto: Método de evaluación de bibliotecas digitales usando tecnologías de *Linked Data*.

## Referencias

Allio, M. K. (2012). Strategic dashboards: designing and deploying them to improve implementation. *Strategy & Leadership*, 40(5), 24–31. DOI: 10.1108/10878571211257159.

- Banker, R., Chang, H. & Pizzini M. (2004). The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review*, 79(1), 1–23. DOI: 10.2308/accr.2004.79.1.1.
- Berners-Lee, T. (2006). *Linked data-design issues*. Retrieved from <http://www.w3.org/DesignIssues/LinkedData>. Html.
- Bogza, R. M. (2014). Datamarts architecture and design. In *Proceedings of the 9th edition of the International Conference on Accounting and Management Information Systems (AMIS 2014)*, 381.
- Eckerson, W. W. (2010). *Performance dashboards: measuring, monitoring, and managing your business*. New Jersey, NJ: John Wiley & Sons
- Freixo, J., & Rocha, Á. (2014). Arquitetura de informação de suporte à gestão da qualidade em unidades hospitalares. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (14), 1–15. DOI: 10.17013/risti.14.1-15.
- Guillén, A., Lloret, E., & Gutiérrez, Y. (2016). TLH Suite: herramienta para la anotación semántica de información. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (18), 99–103. DOI: 10.17013/risti.18.99-113.
- Kitchenham B., Pearl O., Budgen D., Turner M., Bailey J., and Linkman S. (2009). Systematic literature reviews in software engineering. *Information and software technology*, 51(1), 7–15. DOI: 10.1016/j.infsof.2010.03.006.
- Hallo, M., Martínez-González, M., & de la Fuente Redondo, P. (2012). Las tecnologías de Linked Data y sus aplicaciones en el gobierno electrónico. *Scire: Representación y Organización del Conocimiento*, 18(1), 49–61.
- Hallo M., Luján-Mora S., & Trujillo J. (2015). An Approach to Publish Statistics from Open-access Journals using Linked Data Technologies. In *Proceedings of the 9th International Technology, Education and Development Conference (INTED2015)*, 5940–5948.
- Hartig, O., Sequeda, J. et al. (2010). How to consume Linked Data on the Web. Retrieved from [http://iswc2009.semanticweb.org/wiki/index.php/ISWC\\_2009\\_Tutorials/](http://iswc2009.semanticweb.org/wiki/index.php/ISWC_2009_Tutorials/).
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1–136. DOI: 10.2200/S00334ED1V01Y201102WBE001.
- Hendler, J. (2009). Web 3.0 Emerging. *Computer*, 42(1), 111–113.
- Illescas, G., Sanchez-Segura, M. I., & Canziani, G. A. (2015). Métodos de Pronóstico por Indicadores dentro de la Gestión del Conocimiento Organizacional. *RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação*, (SPE3), 29–41. DOI: 10.17013/risti.e3.29-41.
- International Organization for Standardization. (2013). ISO 2789:2013: Information and documentation-International library statistics. ISO.



- International Organization for Standardization. (2014). ISO 11620:2014: Information and documentation-Library performance indicators. ISO.
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N. & Bizer, C. (2015). DBpedia—a largescale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2), 167–195.
- Library Linked Data Incubator Group. (2011). W3C Incubator Group Report 25 October 2011. Retrieved from <http://www.w3.org/2005/Incubator/lld/XGRlld20111025>.
- Luján-Mora, S., & Palomar, M. (2001). Reducing inconsistency in integrating data from different sources. In *Proceedings of Database Engineering and Applications Symposium*, 209–218.
- Luján-Mora, S & Trujillo, J. (2003). A UML based approach for modeling ETL processes in data warehouses. In *Proceedings of 22nd International Conference on Conceptual Modeling*, 307–320. DOI: 10.1007/978-3-540-39648-2\_25.
- Maali, F., & Cyganiak, R. (2012). RDF Refine-a Google Refine extension for exporting RDF. Retrieved from <http://refine.deri.ie/>.
- Ontology Engineering Group. (2015). Web Semántica y Linked Data. Retrieved from <http://mayor2.dia.fi.upm.es/oegupm/index.php/es/researchareas/4semanticweb>.
- Open Knowledge Foundation. (2006). Datahub. Retrieved from <http://datahub.io/>.
- Shadbolt, N., O’Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, 27(3), 16–24. DOI:10.1109/MIS.2012.23.
- Sheridan, J. (2010). Legislation.gov.uk. Retrieved from <http://blog.law.cornell.edu/voxpath/tag/legal-linked-data/>.
- Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. In *Proceedings of the National Academy of Sciences*, 112(47), 14473–14478. DOI:10.1073/pnas.1515373112.
- Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1), 41–59.