



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



**DATOS ABIERTOS ENLAZADOS *LINKED OPEN DATA* (LOD)
EN DOCUMENTACIÓN CIENTÍFICA**

Tesina final de estudios Master Oficial CALSI

Contenidos y Aspectos Legales
en la Sociedad de la Información

Presentada por
Alicia García García

Bajo la dirección de las doctoras
Antonia Ferrer Sapena
Fernanda Peset

Valencia, Septiembre 2012

Índice

CAPITULO 1	3
Justificación	3
Objetivos	4
Fuentes y metodología	5
Estructura	7
Agradecimientos	7
CAPITULO 2	8
Introducción: la web	8
CAPITULO 3	15
Acceso a la información científica en la web	15
Modelo de publicación OA	19
Autoarchivo en repositorios abiertos	24
Interoperabilidad	30
CAPITULO 4	38
Linked Open Data	38
Ontologías	44
CAPITULO 5	66
Ejemplo de LOD aplicado en información bibliográfica	66
CAPÍTULO 6	73
Conclusiones	73
BIBLIOGRAFÍA	78
ÍNDICE DE IMÁGENES	89

CAPITULO 1

Justificación

Con el presente trabajo se pretende situar el concepto de datos abiertos enlazados, Linked Open Data (LOD) y su relevancia como elemento base en los procesos de preservación, recuperación e intercambio de información en el ámbito de la web. Las siguientes páginas quieren realizar una revisión de las actuales formas de comunicación de la información científica en el ámbito digital y el cambio que supone la publicación de esta información en LOD en el contexto del acceso abierto.

Desde esta perspectiva se analiza el uso de vocabularios de descripción de esta tipología de datos, su almacenamiento y transmisión, en el contexto de una web semántica; y su aplicación en los sistemas de información en los que está siendo empleado. Para ello, será necesario revisar el estado de desarrollo de los estándares promovidos por la institución que encabeza el proyecto de definición de la información en datos abiertos enlazados.

Esta tesina revisa la situación actual de la información en cuanto a su publicación y su gestión en los medios digitales, teniendo en cuenta diversos aspectos del proceso que sigue la información en la cadena documental. Procesos como la publicación, la descripción de los documentos digitales a través de los metadatos, en la fase de almacenamiento, con miras a la recuperación de la información en el contexto digital de la web. También se quiere hacer una revisión de los protocolos existentes para la normalización y estandarización de algunos de los elementos que intervienen en estos procesos en el ámbito digital como por ejemplo el protocolo OAI-PMH para la transmisión de metadatos, enmarcado en el proceso documental de descripción y transmisión de objetos digitales.

El trabajo observa las formas actuales de publicación de la información

científica en el ámbito web su evolución hacia al acceso abierto. Este análisis es necesario ya que la realidad actual es que el soporte para el almacenamiento de la información y la transmisión del conocimiento producido por la humanidad, ha cambiado del papel al formato digital. A lo largo de la historia los soportes de almacenamiento del conocimiento han ido evolucionando, desde las tabletas de arcilla hasta los pergaminos, pasando por los papiros y el primer papel a base de pasta de telas. Se pasó del papel vegetal a los soportes analógicos como las cintas, a los soportes digitales como los disquetes, el cd y los archivos de almacenamiento. Hasta alcanzar su visualización en la web. Así mismo, la web como medio de comunicación evoluciona, y surge así un nuevo paradigma de transferencia de la información, de una forma distribuida en lo que se denomina web de datos.

De la misma manera, el soporte de la información digital y su almacenamiento continúa evolucionando, y debido a la gran variedad y a la multitud de versiones que van naciendo y quedando obsoletas, su preservación de forma indefinida constituye una de las cuestiones más importantes en la actualidad en la sociedad de la información.

Dentro del proceso de tratamiento de la información en el entorno digital, que se lleva a cabo por los profesionales tecnológicos, gestores de la información, y profesionales del mundo de las bibliotecas y la documentación, la preservación, la recuperación y el intercambio de información entre las personas constituyen fases sobre las que se está trabajando para su normalización. De forma que sea posible que se den fenómenos como la interoperabilidad entre los diversos sistemas existentes actualmente en la red.

Objetivos

El trabajo pretende alcanzar los siguientes objetivos:

- Obtener una visión general de la evolución de la web hasta la fecha actual en la que se desarrollan los datos enlazados.
- Introducir el concepto de datos abiertos enlazados (LOD) y su aportación a la nueva web.
- Analizar aspectos como las actuales formas publicación de la información científica en el ámbito digital y el cambio que puede suponer la publicación de esta información en LOD.
- Contemplar los actuales protocolos para la interoperabilidad y la continuidad que pueden tener para su aplicación también en los LOD.
- Revisar los estándares para la descripción de los documentos digitales a través de los metadatos.
- Revisar los vocabularios propuestos por las instituciones que encabezan el desarrollo de la web semántica

Fuentes y metodología

Para realizar este trabajo se ha seguido una metodología de consulta de fuentes de información, interrogando a los sistemas con dos ecuaciones de búsqueda y su correspondencia en lengua inglesa: “Web semántica”, “Datos enlazados”, “*Semantic web*” y “*Linked data*”.

En las bases de datos del CSIC, con la ecuación de búsqueda “Web semántica” en ICYT - Ciencia y Tecnología ha recuperado 41 documentos y en ISOC - Ciencias Sociales y Humanidades 73 documentos. Con “Datos enlazados” un documento, la ecuación “Datos abiertos enlazados” no recupera ningún documento, y “*Linked data*”, “*Linked open data*” y “*Semantic web*” tres documentos en ISOC.

El número de documentos se ha visto considerablemente aumentado al interrogar

la base de datos del *Web of Knowledge*, y las fuentes especializadas en Biblioteconomía y Ciencias de la información: el repositorio E-Lis (*Eprints in Library and Information Science*) y la Base de datos LISA (*Library and Information Science Abstracts*).

También se ha consultado la plataforma Recolecta de la Fundación Española para la Ciencia y la Tecnología (FECYT), que permite el acceso libre a toda la producción científica depositada en abierto en los repositorios científicos españoles. El directorio DOAJ (*Directory of Open Access Journals*), la Red de Revistas Científicas Españolas RECIVEN y Dialnet.

El catálogo de la Biblioteca Nacional de España, la base de datos de libros editados en España ISBN, del Ministerio de Cultura, y las bases de datos de tesis doctorales TESEO y TDX. Así como también la Red de bibliotecas universitarias REBIUN.

Así mismo, se han consultado numerosos documentos que el *World Wide Web Consortium* (W3C) ha definido y están relacionados con los datos enlazados, como la guía sobre Linked Data, la guía del vocabulario RDF, OWL, SKOS y SPARQL.

Además se han realizado otras búsquedas en estas mismas fuentes sobre conceptos que están relacionados y que son necesarios describir para entender mejor el propósito de los LOD, como Open Access, los protocolos de intercambio de información OAI-PMH, OAI-ORE, interoperabilidad, metadatos, repositorios, y los estándares de descripción normalizados como DC, METS, MODS, MADS. Y otros conceptos específicos relacionados con los datos enlazados como ontología, FOAF, DBPedia, EDM, Europeana.

Estructura

En el capítulo primero se hace un breve repaso a la historia de la web. Desde sus inicios, con los primeros proyectos de norteamericanos, pasando por la web actual, la de las redes sociales; hasta la web semántica, que crea la polémica entre los autores más escépticos, quienes instan a que sea únicamente un concepto teórico y los que trabajan en el desarrollo real aplicado, como la transferencia de datos a LOD.

En el cuerpo central del análisis, se desarrollan dos capítulos. Por un lado el capítulo sobre el concepto teórico de LOD, su relación con filosofías de publicación de información científica como OA, OAI, el protocolo OAI-PMH y OAI-ORE, y la relación conceptual con los metadatos. Por otro lado, el capítulo que vincula los LOD con los lenguajes necesarios que los describen y un capítulo en el que se describen ejemplos de sistemas que emplean LOD para compartir recursos.

Por último, en el capítulo de conclusiones se ponen de relevancia una serie de ideas, a las que se llega tras el desarrollo del trabajo.

Agradecimientos

La autora quiere dar las gracias a las personas que han ayudado en la elaboración de esta tesina, especialmente a las tutoras Dra. Antonia Ferrer y Dra. Fernanda Peset por su rigor científico y su metodología, así como por sus consejos e indicaciones.

CAPITULO 2

Introducción: la web

Actualmente la web es un entorno digital que se ha convertido en ámbito de trabajo, ocio, y negocios. La web es una herramienta de comunicación tan común en la vida diaria de las personas como los teléfonos móviles, la televisión o la radio. Facilita muchas de las actividades cotidianas, y es un medio de comunicación con un coste económico reducido, que se ha convertido en una herramienta para la consulta de información, para las transacciones comerciales y los trámites con la administración.

Según indican los datos estadísticos recogidos en 2011 por el Instituto Nacional de Estadística, en la encuesta sobre Equipamientos y Uso de Tecnologías de la Información y Comunicación en los hogares en 2011, el 18.9% de la población española realizan compras a través de Internet (octubre de 2011), un total de 3.392.601 personas realizan compras en la red de películas, música, libros, revistas, periódicos, material formativo *on line*, programas de ordenador y actualizaciones. El 98.5% de los hogares tienen acceso a Internet a través del ordenador personal, portátil o de sobremesa, y el 21,6% a través de sus teléfonos móviles. En definitiva, el 71,4% de la población española utiliza Internet diariamente, al menos cinco días por semana. Esto quiere decir que más de la mitad de la población española utiliza diariamente la web.

Estos datos constatan que la web proporciona numerosas ventajas cuya principal es el acceso a la información de forma remota, añadido a la inmediatez para conocerla y para comunicarla, que constituye un elemento fundamental en el desarrollo de las actividades y la toma de decisiones en un mundo cambiante como el actual. Hasta hace relativamente poco tiempo, para tener acceso a la información técnica, científica y docente, era necesario desplazarse a la biblioteca para consultar el catálogo, solicitar la publicación científica para acceder al

artículo, o consultar el manual o la obra de referencia físicamente.

En la actualidad, en la era de Internet, la situación ha cambiado gracias a la versión digital de los materiales y las herramientas digitales disponibles a través de la web, que facilitan la comunicación y difusión de la información, y donde las personas consumen gran cantidad de datos, pero también los producen. La sociedad actual demanda cada vez más una actualización constante de la información. La web constituye un elemento clave para atender esta demanda, de la misma manera que ha sido la creadora de esta necesidad de actualización continua.

Los sitios web tienen numerosas vertientes, y una de ellas consiste en ser una herramienta de docencia y de investigación, una aplicación a través de la que los usuarios pueden consultar los catálogos de las bibliotecas, los manuales digitales, los recursos docentes o las bases de datos y sus contenidos documentales, que van a contribuir al desarrollo de la producción científica o la formación académica. Tal cantidad de información académica, científica y técnica, principalmente de revistas especializadas, manuales técnicos, bases de datos, repositorios, etc. hace necesaria que la información haya de estar estructurada y organizada para que los usuarios, tanto estudiantes, como docentes e investigadores, puedan acceder de forma sencilla y rápida a los contenidos que satisfagan sus necesidades informativas.

Para cubrir las necesidades de los usuarios que cada vez son más exigentes, el medio digital ha ido evolucionando, ha mejorado sus características, ha ampliado sus servicios y las herramientas que ofrece. Pero hasta alcanzar el momento actual, se ha ido desarrollando a lo largo de todo el siglo XX.

El precursor de la idea de web fue Vannevar (1890 – 1974), ingeniero y profesor del *Massachusetts Institute of Technology* (MIT) y asesor de Investigación Científica y Desarrollo en la Administración del presidente Roosevelt de EEUU durante la Segunda Guerra Mundial.

En 1945 describió la primera aplicación hipertextual, el proyecto MeMex “Memory Extended” descrito en *As we may think*. “Cómo podríamos pensar” publicado en la revista *Atlantic Monthly* (Vianello, 2004). La idea descrita era la de un sistema distribuido de información accesible a través de un escritorio (interfaz), en el que los archivos de texto e imagen estuvieran ligados y pudieran ser consultados a través de los enlaces. Así los usuarios podrían visualizar la información a través de distintos recorridos (Bush, 2001).

Este es el origen del hipertexto, que fue la base para que en la década de los años 60 Theodor Nelson definiera lo que treinta años más tarde se desarrollaría como la primera web. En 1965 acuñó el término “Hypertext” que describió como “una escritura no secuencial”, un texto multilínea que permite al lector elegir qué itinerarios (nexos) seguir para continuar la lectura (Nelson, 1993). En 1980 publicó el libro *Literary Machines* “Máquinas literarias” en el que describía el proyecto *Xanadú*, en el que concebía la idea de una red que enlazara los documentos de todo el mundo. Se trataba de fundar un universo formativo (*docuverse*) en el que los textos contenidos se relacionaran entre sí mediante unos enlaces que permitieran navegar de unos a otros (Vianello, 2004).

Paralelamente, Douglas Carl Engelbart, del centro *Augmentation Research Centre* (ARC) de la Universidad de Stanford, trabajó sobre la idea de Bush, desarrollando un software capaz de implementar hipertexto o referencias cruzadas automáticas a otros documentos. El proyecto *oN Line System* (NLS) consistía en un sistema de navegación en línea, una red local que enlazaba los documentos publicados por el Instituto, para poder acceder a ellos y trabajar en grupo.

Basándose en estos proyectos, fueron surgiendo numerosos sistemas de hipertexto como *Hypertext Editing System (HES)* en 1967 por IBM y la Brown University, *File Retrieval and Editing System (FRESS)* en 1968, *ZOG* en 1972 de la Carnegie Mellon University, *Knowledge Management System (KMS)* en 1983, *Intermedia* en 1985 en la Brown University, *Note Cards* en 1985 por la Xerox Parc o *Smart Text* en 1988.

Hasta que en 1990 Tim Berners-Lee y Robert Cailliau dieron a conocer la web como software de navegación que se emplea actualmente para el uso de Internet. Presentaron el proyecto “World Wide Web” en el CERN (Organización Europea para la Investigación Nuclear) en Ginebra (Suiza), que constituyó la primera propuesta de Web. Un programa cliente (navegador / editor) basado en el lenguaje de etiquetas de hipertexto HTML (*Hyper Text Markup Language*), el protocolo de transferencia de hipertexto HTTP (*Hyper Text Transfer Protocol*), el lenguaje que los ordenadores usarían para comunicarse por Internet, y el identificador de recursos universal URI (*Uniform Resource Identifier*), el esquema para las direcciones de documentos (Berners-Lee, 2000).

El navegador/editor *World Wide Web* era capaz de seguir vínculos en los archivos de los servidores http, y también en los artículos de noticias de Internet, que comenzaron a ser accesibles como páginas de hipertexto gracias al protocolo de transferencia de archivos de un servidor a otro: FTP (*File Transfer Protocol*) que en los años setenta había puesto en marcha el grupo de trabajo *Network Workign Group* (NWG) en los laboratorios del ARPA (Agencia de Investigación de Proyectos Avanzados) del Departamento de Defensa de los EEUU.

De esta forma, la información disponible en Internet se hizo disponible en la Web, y los ordenadores podían comunicarse unos con otros y transferir información, desde cualquier ordenador en cualquier parte del mundo. Esta web del hipertexto es una web dinámica, multidireccional, que permite la exploración rápida de grandes bloques de conocimiento. Proporciona al usuario la capacidad de elegir hacia dónde dirigir su lectura, saltar de un contenido a otro, enlazando textos distribuidos y asociados mediante enlaces electrónicos.

Este cambio filosófico requería hacer comprender a los usuarios la importancia de aportar datos a la web. La posibilidad que ofrecía la web no sólo era la consulta, sino la también la oportunidad de incluir conocimiento que será consultado por otros.

A lo largo de una década muchas organizaciones a nivel mundial han ofrecido sus datos, de forma que el crecimiento de información ha crecido de forma exponencial. La web se ha expandido, se ha ido consolidando su uso en todo el mundo y por toda clase de usuarios. Ha evolucionado de forma natural hacia la adaptación a las necesidades de comunicación del ser humano, hacia una sociabilidad online. Esto se ve reflejado en que surgen una serie de nuevos servicios que cuentan con un alto grado de interactividad con los usuarios, aplicaciones orientadas a la comunicación colectiva.

A finales de 2004 Tim O'Reilly, propietario de la empresa editorial estadounidense de libros de informática, *O'Reilly Media*, acuñó el término Web 2.0 para hacer referencia a estos cambios (O'Reilly, 2005).

En la web definida como social destaca el alto nivel de participación de los usuarios para crear contenidos y establecer relaciones para compartirlos, toda una gama de servicios de colaboración para el intercambio de información entre usuarios: redes sociales, blogs, microblogging, wikis, folcsonomías, selecciones de música, de favoritos, donde la valoración de los contenidos depende de las relaciones entre los individuos (O'Reilly, 2009).

El hecho de que la web se haya transformado para desarrollar una función más social, se debe a que todo el concepto de la web ha cambiado, pasando a ser una importante herramienta de comunicación a través de plataformas para grupos y comunidades. El uso de estas plataformas ha supuesto un cambio en el modelo de negocio puesto que la mayoría de los negocios alojados o publicitados a través de la web han apostado por utilizar estas herramientas para acceder a un mayor número de usuarios.

Indistintamente las instituciones públicas, culturales, y las empresas privadas, han querido formar parte de esta nueva web en la que el *feedback* con los usuarios y la colaboración y transferencia de información entre los miembros de su comunidad

constituyen canales de comunicación muy válidos, y cuentan con espacios como blogs, wikis, etc.

El año 2009 se consideró el año de la eclosión de las redes sociales en España, que continuó teniendo su auge en 2010. Se trata de un fenómeno reciente que puede ser permanente en los próximos años hasta que cambien las tendencias. Algunos autores afirman la continuidad de estos servicios de redes sociales, ampliándose, como ya ha ocurrido a dispositivos móviles (Margaix, 2010).

Pero la generación de nueva información sumada al volumen ya existente en la web, que continúa creciendo a un ritmo vertiginoso debido a que los consumidores de información se constituyen a su vez en productores de contenidos, genera una serie de problemas de gestión.

Estos problemas derivados de la masificación de la información en el entorno web, son principalmente la exahustividad en la recuperación de información, la veracidad de las fuentes y la calidad de la información.

La falta de estructuración de contenidos en los recursos digitales produce ambigüedades en el proceso de recuperación de información en la web y, en los motores de búsqueda, los resultados obtenidos suelen ser múltiples y poco relevantes. De esta forma el usuario pierde mucho tiempo localizando la información que realmente le resulta relevante.

Para superar las limitaciones que se encuentran en la web 2.0, Tim Berners Lee en 1998 escribió su visión de lo que sería un nuevo concepto de web, y en 2009 presentó en TED (*Technology Entertainment and Design*), la idea de la web de los datos enlazados, que concebía como una nueva forma de contenido en la que se compartirían datos puros en la red. Los datos son las unidades más pequeñas de información con sentido lógico que en estructuras más grandes conforman conocimiento (Berners-Lee, 2009).

En el supuesto en el que los datos enlazados constituyen las bases de la web, las instituciones gubernamentales y las empresas privadas compartirían los datos, para que los técnicos, científicos e investigadores puedan utilizarlos, combinarlos y analizarlos para crear nueva información.

De la misma manera los individuos también compartirían sus datos puros, no los documentos. Para que puedan reutilizarse y combinarse para generar nuevo conocimiento. Es decir, que la idea de los datos abiertos enlazados constituyan la nueva web, se dirige a que estos datos se definan y se vinculen de forma que puedan ser usados por las máquinas, no solo en la forma de mostrarlos a través de la interfaz sino también para tareas de automatización, interpretación y reutilización de los datos a través de distintas aplicaciones (Berners-Lee y Hedler, 2001).

CAPITULO 3

Acceso a la información científica en la web

En la actualidad, a gran parte de la información existente en Internet se puede acceder de forma gratuita, pero otra parte muy importante tiene el acceso restringido. Esto se debe a distintos factores, ejemplo de ello es la información producida por una institución u organismo únicamente para su uso interno, con lo que, no se pondría a disposición pública en la red. Otro caso es la información que ha pasado por un proceso de tratamiento y su coste inicial ha aumentado de forma que el nuevo producto se ponga a la venta.

La información científica se comunica a través de Internet, es la forma más inmediata de dar visibilidad a la investigación y por lo general, su acceso es restringido. La literatura científica y académica es propiedad de las instituciones de carácter superior a las que pertenecen los investigadores que la producen, pero pasa a ser propiedad de las editoriales al publicarla. Editoriales y distribuidores son los grandes beneficiarios del negocio de la comunicación científica, pues venden a estas instituciones tales como universidades, centros de investigación, ministerios y bibliotecas, a unos costes elevados, el acceso a este conocimiento científico.

Según Borgoños, (2010):

“el conocimiento no debería tener unos costes tan elevados, máxime cuando se trata de publicaciones que emanan de organismos financiados con fondos públicos, cuyos investigadores reciben sueldo y subvenciones de la administración pública. Parece más lógico que lo público

revierta en público”;

Pues es muy importante que el trabajo que se realiza en las universidades y los centros de investigación favorezca a los ciudadanos, y sea ampliamente difundido, porque los resultados de las investigaciones tienen repercusiones sociales. Por un lado, es positivo para fomentar una sociedad del conocimiento y por otro lado, es beneficioso porque proporciona mejoras y resuelve problemas a los ciudadanos.

Siguiendo esta línea de pensamiento, un concepto relacionado con el valor de la información científica y su aplicación en la sociedad, es el de su rentabilidad o retorno de la inversión (ROI o *Return on investment*) “*la información se compra y se paga por ella según la fiabilidad de la fuente, y la percepción del valor que va a tener para el comprador*” (Baiget 2010). Es lo que ocurre con el acceso a la información científica, que tiene costes muy elevados, pero la fiabilidad de las fuentes es indiscutible, pues han pasado por un proceso de evaluación por pares. En este caso el retorno de la inversión también será positivo, pues los investigadores, tras haber utilizado la información producen nuevo conocimiento.

En este sentido se desarrolla el movimiento Open Access (OA), movimiento del acceso libre al conocimiento. Esta filosofía incluye la información y la cultura libre, el software libre y el acceso abierto a la ciencia, sin barreras económicas ni restricciones de derechos de autor, con miras a la difusión a un público más amplio y a la reutilización del conocimiento.

Para Peter Suber (2004), profesor en la Universidad de Stanford y autor de algunos de los textos más citados del estudio del acceso abierto a la ciencia:

“la literatura científica de acceso abierto es digital, en línea, gratuita y se encuentra eximida de la mayoría de derechos de autor y restricciones de licencias. Lo que la hace posible es Internet y el consentimiento del autor o del titular del

copyright... En la mayoría de campos del conocimiento, las revistas especializadas no pagan a los autores, quienes por consiguiente, pueden autorizar el acceso abierto son que ello repercuta en sus ingresos... El acceso abierto es absolutamente compatible con la revisión por pares de expertos, y la mayoría de iniciativas de acceso abierto destacables en el ámbito de la literatura académica insisten en la importancia de este punto”.

El OA en la información científica es un modelo vinculado a la forma de comunicación de la actividad investigadora. En esta definición se alude a tres características de los contenidos de libre acceso: su carácter digital y su disponibilidad en línea gracias a Internet, su carácter gratuito para los usuarios, y la ausencia de restricciones de derechos de autor. Otro elemento destacable es que se mantiene la revisión por pares igual que en el modelo tradicional de publicación, para asegurar la calidad de la publicación.

Existen una serie de declaraciones emitidas en consenso por grupos de expertos y científicos, sociedades profesionales e instituciones internacionales que sostienen y perfilan el movimiento OA. Las más importantes son:

- La Declaración de Budapest de 2002 (*Budapest Open Access Initiative, BOAI*): *“La literatura que debe ser accesible gratuitamente en la red es aquella que los científicos y estudiosos entregan al mundo sin esperar remuneración. (...) Por acceso abierto a la literatura científica erudita, entendemos su disponibilidad gratuita en Internet, para que cualquier usuario la pueda leer, descargar, copiar, distribuir o imprimir, con la posibilidad de buscar o enlazar al texto completo del artículo, recorrerlo para una indexación exhaustiva, usarlo como datos para software, o utilizarlo para cualquier otro propósito legal, sin otras barreras financieras, legales o técnicas distintas de la fundamental de acceder a la propia Internet. El único límite a la reproducción y distribución de los*

artículos publicados, y la única función del copyright en este marco, no puede ser otra que garantizar a los autores el control sobre la integridad de su trabajo y el derecho a ser acreditados y citados”.

- La Declaración de Bethesda de 2003 (*Bethesda Statement on Open Access Publishing*):

“ La publicación en acceso abierto ha de cumplir las dos condiciones siguientes:

El/los autor/es y el/los propietario/s de los derechos de propiedad intelectual otorgana los usuarios un derecho libre, irrevocable, universal y perpetuo de acceso y licencia para copiar, utilizar, distribuir, transmitir y presentar el trabajo públicamente y hacer y distribuir obras derivadas, en cualquier soporte digital para cualquier finalidad responsable, sujeto a la apropiada atribución de la autoría, así como el derecho de hacer una pequeña cantidad de copias impresas para su uso personal.

Una versión completa de la obra y todos los materiales suplementarios, incluyendo una copia de los permisos citados anteriormente, en un formato electrónico estándar apropiado se depositará de forma inmediata a la publicación inicial en al menos un repositorio en línea apoyado por una institución académica, una sociedad de intelectuales, una agencia gubernamental, o cualquier otra organización debidamente establecida que persiga facilitar el acceso abierto, la distribución sin restricciones, la interoperabilidad y el archivado a largo plazo.”

- La Declaración de Berlín de 2003 (*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*), en la que se avala este paradigma y se describe que las dos condiciones que deben satisfacer las contribuciones de acceso abierto son las dos citadas en la Declaración de

Bethesda.

Para alcanzar el acceso abierto a la literatura científica se recomiendan dos estrategias: el modelo de publicación OA y el autoarchivo en repositorios abiertos.

Modelo de publicación OA

En su definición se contempla que este modelo no cobrará derechos de suscripción o acceso sino que buscará otros métodos de financiación, en fondos de entidades gubernamentales, instituciones y universidades. No obstante, hay que tener en cuenta los gastos de edición y revisión por pares que se deriva del proceso de publicación de la actividad investigadora. Por lo que en este modelo surgen cuatro tipos de publicación de acceso abierto (Melero y Abad, 2008):

- Gratuitas y libres para lectores y autores: los costes son asumidos por los editores que suelen ser departamentos universitarios, centros de investigación subvencionados u organismos públicos. Suele tratarse de revistas en las que la suscripción de la versión en papel proporciona los ingresos suficientes para que la versión en línea sea absolutamente gratuita.
- Pago por publicación en revistas OA: es el autor el que asume los costes de la revisión del material y su publicación. Esta modalidad se fundamenta en que en escasas ocasiones es el autor el que verdaderamente paga las tasas, siendo más habitual que asuma los gastos la institución a la que está adscrita o la biblioteca. Algunos editores afirman que este sería el modelo que puede equipararse a la tradicional forma de publicación.
- Pago por publicación en revistas comerciales: es un modelo denominado híbrido porque coexisten artículos abiertos con artículos de acceso por suscripción. Se trata de la posibilidad que ofrecen las editoriales

tradicionales de publicar trabajos en OA mediante el pago de unas tasas por parte de los autores. Grandes editoriales como Springer y Oxford University Press cuentan con él. De esta forma las editoriales mantienen sus ingresos por suscripción e incorporan ingresos por publicación, mientras contribuyen a facilitar el acceso abierto.

- Acceso gratuito con periodo de embargo: se trata de revistas comerciales que siguen el sistema tradicional con acceso por suscripción, tanto en la versión en papel como en la digital, que tras un periodo de embargo ofrecen acceso gratuito la publicación. El periodo de embargo oscila entre los seis meses y los dos años. Aunque hay que destacar que este modelo no cede los derechos de explotación con lo que no está permitido copiar o distribuir los contenidos. Este modelo no cumpliría con la característica de publicación de acceso abierto de forma inmediata.

Estas formas de publicación son sistemas que se han ido implantando en los últimos años, y que van ganando mercado, pues cada vez son más los editores que ofrecen la versión de las revistas que publican también en OA.

A pesar de su continua expansión, existe discrepancia entre los autores, pues no todos están de acuerdo en ofrecer sus resultados en libre acceso. Algunos consideran que el inconveniente de la limitación en el uso de la obra no es elemento clave para el desarrollo de esta filosofía, sino que el triunfo del OA puede depender solamente de que se asegure la realización de *peer review* (Franklin 2003), para asegurar la calidad de la publicación. Pero este tipo de publicaciones pasa por el mismo proceso de revisión por pares que las publicadas en el modelo tradicional, lo que garantiza su calidad al mismo nivel.

Otros autores y editores que no apoyan la filosofía OA mantienen que este modelo de publicación no es sostenible económicamente, y que el sistema comercial es el que permite pagar a los revisores, los gastos de publicación y mantenimiento de

las plataformas digitales.

Mientras unas opiniones se dirigen en la línea de las editoriales tradicionales, otras, manifiestan en la literatura que la forma de publicación en OA proporciona numerosas ventajas, entre las que se destaca el aumento de la visibilidad de los trabajos y como consecuencia de sus autores. Añadiendo a esto la reducción de costes que supone la publicación con este sistema, surge la necesidad de la aplicación de este modelo de publicación en el ámbito de la financiación pública de la producción científica.

Estas afirmaciones se basan en la declaración de Budapest:

“distintas iniciativas han mostrado que el acceso abierto es económicamente viable, que proporciona a los lectores un poder extraordinario para acceder a literatura relevante, y que brinda a los autores y a sus trabajos una dimensión nueva, una nueva visibilidad, un nuevo impacto, y un público más amplio.

(...) Si bien la literatura de revistas científicas y técnicas dotadas de comité de selección debe ser accesible en red sin coste para los lectores, su producción sin costes no es posible. No obstante, y de forma experimental, se ha establecido que los costes totales para dar acceso abierto a la literatura publicada en línea son mucho más bajos que los costes de las formas tradicionales de difusión. Con la oportunidad de ahorrar dinero y al mismo tiempo ampliar el ámbito de la difusión, se presenta ahora un gran incentivo para asociaciones profesionales, universidades, bibliotecas, fundaciones y similares, adoptando el acceso abierto como un medio para avanzar en su misión. Poner en práctica el acceso abierto requerirá nuevos modelos de recuperación de costes y mecanismos de financiación, pero el coste total de la difusión, que resulta significativamente más bajo, es una buena razón para pensar que se trata de una meta alcanzable y no utópica”.

Además, el sistema de publicación de acceso libre se ve respaldado ya desde el año 2006 con las recomendaciones europeas que marca el EURAB (*European Research Advisory Board*), comité asesor creado por la Comisión Europea para la implementación de la política científica europea. Según el informe final sobre las políticas de acceso abierto para las publicaciones científicas de diciembre de 2006 “Scientific publication: Policy on Open Access” (EURAB, 2006), se recomienda que todos los investigadores financiados bajo el Séptimo Programa Marco de Investigación (FP7) estén obligados a depositar los resultados publicados de su investigación en repositorios digitales de acceso abierto en el plazo de seis meses.

Con la puesta en marcha en agosto de 2008, del Piloto de Acceso Abierto en su 7º PM, a los beneficiarios de subvención de las siete áreas (Energía, Medio Ambiente, Salud, parte de Tecnologías de la Información y las Comunicaciones, las Infraestructuras de Investigación, la Ciencia en la sociedad, y las Ciencias Sociales y Humanidades) que abarca el 20% del presupuesto total, se les requiere dos labores: en primer lugar, depositar los artículos de investigación revisados por pares o manuscritos finales resultado de sus proyectos del 7PM en un repositorio en línea.

Y en segundo lugar, que se esfuercen por garantizar el acceso abierto a estos artículos en el plazo de seis meses para Salud, Energía, Medio Ambiente, Tecnologías de la Información y Comunicaciones, Infraestructuras de Investigación, o doce meses en Ciencias Sociales y Humanidades, Ciencia en la sociedad) tras su publicación (Comisión Europea, 2008).

De este modo, se consigue alcanzar el objetivo de mejorar y promover la difusión del conocimiento, mejorar la eficiencia del descubrimiento científico y la maximización de retorno de la inversión en I+D financiada por organismos de investigación pública.

Por otro lado, la publicación de conocimiento científico en acceso abierto en el ámbito nacional, se ve reconocida con la aprobación de la nueva Ley de la Ciencia y la Tecnología de España (2011) que incluye un artículo sobre la publicación en acceso abierto, que dispone que todos los investigadores cuya actividad haya sido financiada a través de los presupuestos generales del estado está obligado a depositar en acceso abierto una versión digital de la publicación de la investigación. Ley 14/2011, de 2 de junio, de la Ciencia, la Tecnología y la Innovación, que deroga a la antigua "Ley 13/1986 de 14 de abril, de Fomento y Coordinación General de la Investigación Científica y Técnica", en el "*Capítulo II. Transferencia y difusión de los resultados de la actividad de investigación, desarrollo e innovación y cultura científica, tecnológica e innovadora*" y en concreto en su "*Artículo 37. Difusión en acceso abierto*":

1. Los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación impulsarán el desarrollo de repositorios, propios o compartidos, de acceso abierto a las publicaciones de su personal de investigación, y establecerán sistemas que permitan conectarlos con iniciativas similares de ámbito nacional e internacional.
2. El personal de investigación cuya actividad investigadora esté financiada mayoritariamente con fondos de los Presupuestos Generales del Estado hará pública una versión digital de la versión final de los contenidos que le hayan sido aceptados para publicación en publicaciones de investigación seriadas o periódicas, tan pronto como resulte posible, pero no más tarde de doce meses después de la fecha oficial de publicación.
3. La versión electrónica se hará pública en repositorios de acceso abierto reconocidos en el campo de conocimiento en el que se ha desarrollado la investigación, o en repositorios institucionales de acceso abierto.
4. La versión electrónica pública podrá ser empleada por las Administraciones Públicas en sus procesos de evaluación.

5. El Ministerio de Ciencia e Innovación facilitará el acceso centralizado a los repositorios, y su conexión con iniciativas similares nacionales e internacionales.
6. Lo anterior se entiende sin perjuicio de los acuerdos en virtud de los cuales se hayan podido atribuir o transferir a terceros los derechos sobre las publicaciones, y no será de aplicación cuando los derechos sobre los resultados de la actividad de investigación, desarrollo e innovación sean susceptibles de protección.

Esta ley constituye un avance en materia de acceso abierto a la información científica con miras a su reutilización y fomento de la actividad investigadora. Esto significa que, si el Estado Español tiene en cuenta la importancia que supone la liberación de la información científica y el libre acceso a la misma a través de Internet, es posible que se desarrolle la política de los datos abiertos enlazados. Que las administraciones públicas pongan en libre acceso datos de carácter científico, estadístico, matemático... que pueda ser de interés para la sociedad. Pero principalmente para que la inversión de gasto público que se ha realizado en estos proyectos de investigación revierta sobre la sociedad.

Autoarchivo en repositorios abiertos

Por otro lado para alcanzar el acceso abierto a la literatura científica, en la Declaración de Budapest se recomienda el autoarchivo, que permite depositar los artículos en archivos electrónicos abiertos, que se lleva a cabo siguiendo la iniciativa *Open Archives Initiative* (OAI) que contempla la tecnología necesaria para la creación de repositorios compatibles de publicaciones electrónicas.

Los repositorios son sitios web que preservan y difunden la producción académica y científica de una institución, denominados institucionales, o de una disciplina, denominados temáticos. Entre los contenidos que preservan se encuentran datos,

libros, recursos docentes, *e-prints* que incluyen tres tipologías: los *Pre-prints*, trabajos pendientes de revisión por pares (*peer review*), los *Post-prints*, trabajos que ya han sido evaluados y los *Reprints*, separatas o copias de los documentos ya publicados que la editorial facilita a los autores para su difusión (Silió, 2005).

Por lo general son los propios investigadores quiénes autoarchivan sus trabajos, y posteriormente los profesionales se encargan de la inclusión de los metadatos. No obstante, existe lo que se denomina carga media, que se lleva a cabo por los gestores del repositorio, y la carga masiva de forma automática a través de la recolección de otros repositorios.

La iniciativa OAI impulsó en 1999 en la Convención de Santa Fe (Estados Unidos), la interoperabilidad entre repositorios a través del protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). En dicha convención se buscaba un servicio universal de autoarchivo de *e-prints* para los autores (Fox, 2000), y la clave era la interoperabilidad entre los sistemas, para poder entenderse entre ellos y compartir así los recursos que almacenan.

OAI-PMH es un protocolo orientado a la transmisión de información en Internet que recolecta metadatos de distintos archivos distribuidos por la red, y que permite a los buscadores procesarlos como si estuvieran alojados en un mismo repositorio. Suber (2007) lo define de esta forma:

“La Iniciativa de Archivos Abiertos (OAI) define un protocolo para recoger metadatos de ficheros de datos que residen en archivos separados. Cuando el protocolo es utilizado por servicios de datos como motores de búsqueda, éstos pueden procesar los datos de archivos separados como si residiesen en un solo archivo. En términos técnicos, el protocolo de recolección de metadatos soporta la interoperabilidad”.

Es decir, que ante la multiplicidad de archivos, la utilización de metadatos permite el intercambio de registros y la accesibilidad a documentos de diferente naturaleza y de distintas disciplinas. Los metadatos constituyen la información descriptiva de un recurso almacenado en un repositorio, cuya finalidad es facilitar su recuperación y contribuir a la interoperabilidad entre los sistemas. *“Son un conjunto de elementos que se utilizan para ayudar a la identificación, descripción y localización de los recursos electrónicos por medio de una representación de la descripción bibliográfica de los mismos”* (Juárez Santamaría, 2007).

Su funcionamiento lo explican Barrueco y Subirats de la forma siguiente:

“OAI-PMH utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según determinados criterios como, por ejemplo, la fecha de creación de los datos. En respuesta, el primero le devuelve un conjunto de registros en formato XML, incluyendo identificadores (URL por ejemplo) de los objetos descritos en cada registro” (Barrueco y Subirats 2003).

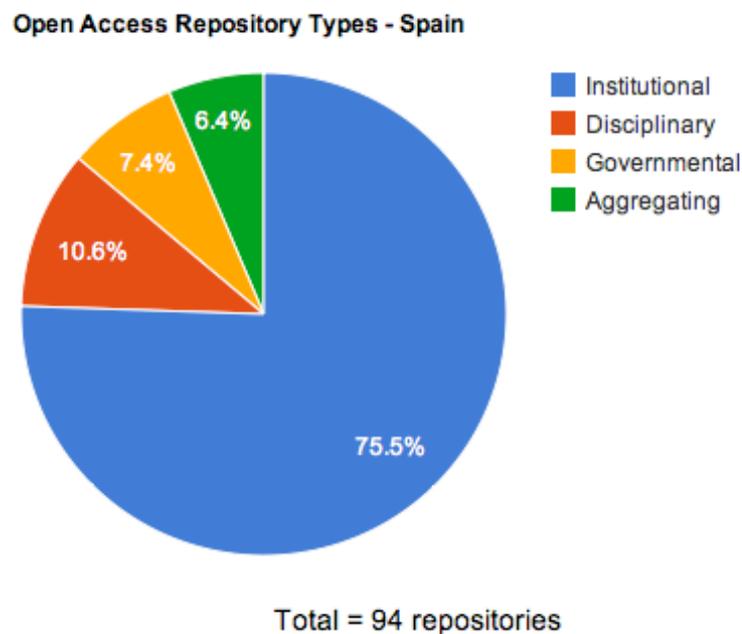
Desde los inicios de la implementación de la iniciativa, han sido numerosas las instituciones que han abogado por su uso, y otras que las que lo han apoyado, como la Federación de bibliotecas Digitales (DLB), la Liga para la Información en Red (CNI) y la Fundación Nacional para el Desarrollo de la Ciencia (NSF), (Gómez y Arias, 2002).

Según Barrueco y Subirats (2003) se trata de una tecnología cuya adopción inicial es paulatina, pero la respuesta positiva a su adopción aumenta de forma exponencial, superando en más de 100 las instituciones que han creado archivos abiertos en 2003. Pero en la actualidad esta cifra se ha visto superada, pues solo

en España, la mayoría de las universidades cuentan con su propio repositorio institucional. Según el Registro de Repositorios de Acceso Abierto ROAR, que recoge un total de 2952 repositorios a nivel mundial, 127 pertenecen a instituciones y organismos en España; y según el Directorio de Repositorios de Acceso Abierto OpenDOAR, que recoge 2199, 94 pertenecen a instituciones españolas.

Como se observa en el gráfico, de los 94 repositorios españoles, el 75,5% son institucionales, principalmente universidades y centros superiores, junto con grandes empresas. El 10,6% son repositorios temáticos, especializados por disciplinas. Y un 6,4% constituyen repositorios agregados y otro 7,4% gubernamentales.

Ilustración 1: Gráfico de las tipologías de repositorios de acceso abierto en España



Fuente: OpenDOAR, 10 septiembre 2012

Estos datos ponen de manifiesto que la opción de publicar la producción de la actividad científica de forma abierta es una actividad alternativa utilizada por un

gran número de investigadores en todo el mundo, y que aumenta de forma progresiva, frente a la tradicional forma de publicación. A esto contribuye que algunos de los grandes editores ofrecen también la opción de publicar en OA y cuentan con su propio repositorio de libre acceso.

El otro gran proyecto incluido en OAI es el de reutilización e intercambio de objetos digitales OAI-ORE (*Object Reuse and Exchange*), que define un conjunto de normas para la descripción e intercambio de agregaciones de recursos web entre repositorios. Se debe a que las colecciones que forman los archivos digitales están compuestas por agregaciones de materiales interrelacionados, denominadas también objetos digitales complejos. Esta agregación de recursos se almacena en ficheros diferentes y en muchas ocasiones están ubicados en distintos servidores, y mediante estas normas es posible tratar estas agregaciones de materiales como un recurso unitario.

OAI-ORE nace en 2006, bajo el amparo de la Fundación Andrew W. Mellon, con el propósito de aumentar la interoperabilidad, pero no únicamente a través de metadatos sino de objetos digitales complejos, que son el conjunto de unidades resultantes de cualquier investigación, es decir, los *datasets*, tales como tablas, textos, software, imágenes... albergados en uno o varios repositorios, pero relacionados de forma consistente como uno solo. (Borillo y Gambau, 2008).

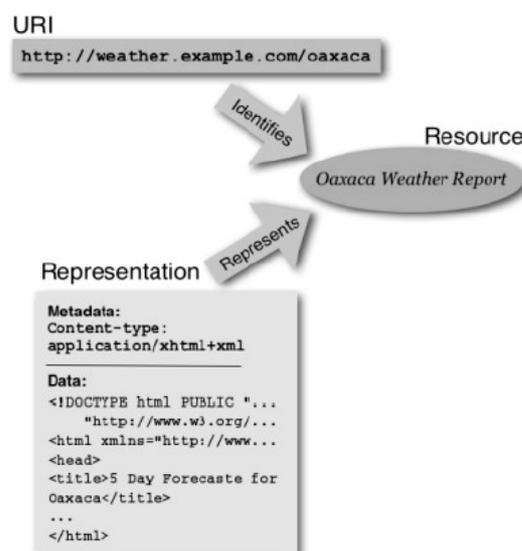
La especificación ORE define proveer una estructura que identifica a la agregación y describe las relaciones entre los componentes de ésta, de forma que los usuarios, máquinas y personas, puedan entender el significado y significado del objeto, y de cómo conecta con otros recursos (Kaplan, 2011). La representación del recurso, en la que se describen sus propiedades y los valores de las mismas, puede ser definida mediante tres formatos, en RDF/XML (*Resource Description Framework*) / (*Extensible Markup Language*) y RDFa siguiendo los principios de *Linked Open Data* (LOD). Cada una de estas unidades que componen un objeto digital tiene una URI asociada para ser referenciado.

También emplean los estándares RSS y Atom, que son formatos XML de redifusión de información. Se trata de dos especificaciones para la descripción de contenidos sindicados, es decir, que permiten la distribución de contenidos como texto, imagen, audio, video, etc., y de sus metadatos descriptivos. En este sentido existe similitud con OAI-ORE porque trabaja con la agregación y con cada uno de los elementos agregados.

Tal como Orduña-Malea (2009) expresa, OAI-ORE representa el dominio de una agregación a través de un grafo que está formado por nodos (los recursos) y arcos (las relaciones). Define la existencia de un nodo principal llamado “mapa de recurso” que describe al nodo que representa a la agregación en sí misma. Este mapa de recurso se identifica con una *URI (URI-R)* que da 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.

La estructura del mapa del recurso está formada por el propio recurso, la representación y la URI como muestra el siguiente gráfico:

Ilustración 2: Relación entre identificador, recurso y representación.



Fuente: Borillo Domenech, Ricardo y Gumbau Mezquita, José Pascual. (2008) “Arquitecturas

orientadas a servicios para la preservación de objetos digitales”

Existe una similitud entre OAI-ORE y OAI-PMH con LOD en cuanto a la importancia que tiene el uso de URIs en sus estructuras de definición. Aunque OAI-PMH y ORE las emplean únicamente con propósito de identificación, LOD las utiliza también con el papel de desreferenciar (Haslhofer, 2009). Es decir, si una aplicación que consume LOD encuentra datos descritos con un vocabulario o esquema semántico que no le es familiar, la aplicación puede desreferenciar la URI que identifica al vocabulario y encontrar su definición (Luján et al. 2011).

Estos protocolos son adaptables a LOD por los elementos que los constituyen y definen. OAI-PMH utiliza transacciones HTTP, para la transferencia de contenidos web, en la que está definida sintaxis y semántica, al igual que uno de los principios básico de los LOD (Heath y Bizer, 2011).

Otro elemento de relación con la estructura de LOD, es que la OAI ha impulsado la interoperabilidad con estos dos protocolos que están orientados a la libre difusión e intercambio y accesibilidad de documentos de diferente naturaleza, alojados en distintas fuentes, especialmente en repositorios, como ya se ha dicho. Este ideal es el que promueve la web semántica y su definición de LOD.

Interoperabilidad

La interoperabilidad se refiere a la capacidad del software empleado en la biblioteca digital o en el archivo digital, para comunicar e intercambiar información con otros software externos. Es fundamental que los repositorios y bibliotecas digitales puedan intercambiar datos para que sea posible ofrecer consultas conjuntas de contenidos alojados en diversas fuentes. Como ya se ha dicho, que sea posible la agregación de contenidos a través de los recolectores de metadatos.

Gómez Dueñas (2005) la define como:

“La capacidad que presenta un sistema de información de comunicarse y compartir información efectivamente con otro mediante una conexión libre y transparente (compartir metadatos, documentos y objetos digitales)”.

Según Martínez Equihua (2007):

”La interoperabilidad es la capacidad de un sistema o de un producto de trabajar con otros sistemas o productos sin esfuerzo especial por parte del cliente. Desde este punto de vista computacional, la interoperabilidad permite generar un enlace entre sistemas de trabajo para las diferentes tecnologías de información promoviendo una sana convivencia y operatividad...”

Para que sea posible que se produzca interoperabilidad entre diferentes sistemas de preservación, los metadatos de los objetos digitales tienen que estar elaborados con el mismo metalenguaje, el XML, con gran capacidad para estructurar y describir recursos. Lo que pone de manifiesto la importancia de la estandarización para una eficaz funcionalidad.

El concepto de metadato según la ISO/IEC 11179 se define como datos que describen otros datos o procesos. Es decir, que permiten aportar información sobre los recursos electrónicos, para facilitar el proceso de búsqueda y recuperación en la web. Definen una estructura y añaden un significado a la misma, permitiendo así identificar y localizar recursos, y establecer enlaces.

Se clasifican según su tipología en tres categorías (Elba y Caballero, 2009):

- Metadatos descriptivos: aquellos que describen e identifican los recursos

de información, facilitan la búsqueda y recuperación. Los que aplican los estándares Dublin Core, MARC, etiquetas meta de HTML.

- Metadatos estructurales: son aquellos que proporcionan la información sobre la estructura interna de los recursos y la relación entre sus elementos o archivos. Contribuyen a facilitar la navegación y la presentación. Los aplican los estándares SGML, XML/RDF y EAD.
- Metadatos administrativos: o de conservación, son aquellos que aportan información relativa a la gestión de derechos, uso del contenido digital, información sobre la autenticidad, historia de la custodia del recurso, estructura y características técnicas.

Los metadatos son un conjunto de elementos propuestos por un organismo, y se consolidan como estándares debido a su reiterada utilización. En el ámbito bibliotecario el estándar consolidado es el Formato MARC21, que incluye todos los elementos de datos esenciales que se necesitan para crear descripciones bibliográficas de información de los ítems.

El conjunto empleado hasta el momento como estándar interdisciplinar de descripción de objetos digitales a nivel general es el Dublin Core. Es producto de un esfuerzo internacional surgido en 1995 en la convocatoria de un taller de trabajo en Dublin, Ohio, para mejorar la normativa relacionada con la recuperación de información en el ámbito de la web que aumentaba de forma vertiginosa. DCMI (*Dublin Core Metadata Initiative*) es la agencia encargada de su mantenimiento.

DC a través de un conjunto básico de quince elementos permite describir todos los recursos existentes en la red, aunque también existe la versión ampliada Qualified Dublin Core con más de cincuenta propiedades que pueden ser ampliados.

Los quince descriptores básicos (ninguno de ellos obligatorios y todos repetibles) se agrupan en tres categorías:

- Contenido: *Title, Subject, Description, Source, Lenguaje, Relation, Coverage*
- Propiedad intelectual: *Creator, Publisher, Contributor, Rights*
- “Instantiation” / ejemplo: *Date, Type, Format, Identifier*

La incorporación de metadatos a los recursos permite aumentar la visibilidad de los recursos, mejorar la recuperación, que ya empezaba a ser difícil en un voluminoso almacén de recursos web, y fomentar la interoperabilidad. La principal característica de DC es la sencillez, por lo que era perfecto para erigirse como norma internacional. DC es un estándar sencillo, extensible e interoperable, es norma ISO 15836-2007, y permite su integración con XML, HTML, XHTML, y RDF/XML, por ello se ha convertido en una infraestructura operacional del desarrollo de la web semántica.

A pesar de la adaptabilidad de DC, es un esquema muy general que no está orientado a un dominio específico, y algunos materiales procedentes de archivos, bibliotecas y museos necesitan una descripción más profunda y específica, por lo que resulta necesario el uso de otros vocabularios adicionales. Un elemento importante es que los esquemas de metadatos permiten el uso conjunto de más de un esquema, lo que favorece utilizarlos en combinación. Existe un amplio conjunto de lenguajes definidos según las particularidades de cada tipo de material. Los más empleados para cubrir el conjunto de materiales disponibles son:

- METS (*Metada Encoding and Transmission Standard*) para el envío de ficheros, imágenes y objetos multimedia de una biblioteca digital.

Surgió en 2002 dirigido a la transmisión e intercambio de objetos digitales entre repositorios OAI. Es un estándar de codificación y recuperación de metadatos,

creado por la *Digital Library Federation*, y el organismo que lo mantiene actualmente es la *Library of Congress*. Es flexible para codificar metadatos descriptivos, administrativos y estructurales para un objeto digital. Relaciona los ficheros que componen el objeto digital y su estructura, y permite expresar sus relaciones. Se basa en XML y se estructura mediante etiquetas.

Un objeto digital descrito con METS, consta de siete secciones, compuestas por varios elementos y sus atributos (METS, 2011):

- La cabecera: <mets:Hdr> contiene los metadatos de descripción con la información del documento relativos a su creador, editor...
 - Los metadatos descriptivos: <mets:dmdSec> contiene la descripción del objeto al que se hace referencia en el documentos METS. Pueden ser tanto internos como externos al documento.
 - Los metadatos administrativos: <admSec> relativos a la creación y almacenamiento de los archivos del objeto digital, gestión de derechos, de uso y copias del contenido, pueden ser externos o codificarse dentro del propio documento.
 - La sección de archivo: <fileSec> es el registro de ficheros que componen el objeto digital. Se agrupan en <fileGrp>
 - El mapa estructural: <structMap> define la estructura jerárquica del objeto, es decir, las secciones que componen el objeto digital, las relaciones entre estos ficheros y la secuencia de los mismos.
 - Los enlaces estructurales: <smLink> son registros de los hiperenlaces entre las secciones del mapa.
 - Los comportamientos: <mets:Behaviour> hacen referencia a mecanismos que identifican un módulo de código ejecutable en la interfaz.
- MODS (*Metadata Object Description Schema*) es más completo que DC y está orientado a las descripciones bibliográficas.

Es una versión simplificada del formato MARC21 de registros bibliográficos definida en XML. Lo desarrollaron la Oficina del estándar MARC y la Biblioteca del Congreso. Consiste un esquema XML simplificado pero más amplio que DC, para la descripción bibliográfica y acceso al recurso web de libros, capítulos de libros, Series, artículos seriados, música...

Permite definir relaciones jerárquicas entre las descripciones y proporciona una codificación detallada de las partes del recurso (volumen, número, capítulo, sección, párrafo, pista, etc.), pudiendo llegar hasta la codificación de secciones o párrafos dentro de un determinado texto. Es una de las extensiones de metadatos METS ampliamente utilizada en bibliotecas y colecciones digitales (Ajenjo y Hernández, 2010).

- EAD (*Encoded Archival Description*) para las descripciones de colecciones de documentos de archivo.

Tiene sus orígenes en 1993 en la biblioteca de Universidad de California en Berkeley. Está mantenido por la *Network Development and MARC Standards Office* de la Biblioteca del Congreso de Estados Unidos en colaboración con la *Society of American Archivists (EAD Working Group)*.

Permite la representación digital de los elementos que componen los documentos de archivos. Es un estándar para codificar instrumentos de descripción archivística por medio de SGML (*Standard Generalized Markup Language*) y XML.

Es el primer estándar de estructura de datos para facilitar la distribución en Internet de información detallada sobre colecciones y fondos archivísticos.

- MADS (*Metadata Authority Description Schema*) para la descripción de autoridades.

Es un complemento de MODS, y consiste en un esquema XML para proporcionar metadatos de autoridades: personas, organizaciones, eventos y términos de materias, geográficos, etc. que está relacionado con el formato MARC 21 de registros de autoridad.

La institución encargada de su mantenimiento es la Biblioteca del Congreso. Es uno de los formatos de metadatos más extendidos en la descripción de autoridades de objetos digitales en los repositorios institucionales, ya que facilita el registro de formas autorizadas y formas de consulta de nombres para entidades.

- VRA CORE es otro estándar de metadatos que se emplea para la descripción de imágenes y trabajos de arte y cultura. La institución que se encarga de su mantenimiento es la *Visual Resources Association* y la Biblioteca del Congreso.

Aunque existen otras estructuras de definición de metadatos, con estos estándares es posible representar la descripción bibliográfica de los recursos electrónicos mayoritarios de Internet. Daudinot (2006) recoge las principales ventajas que proporciona la utilización de metadatos para la recuperación de información en la web:

- Constituyen instrumentos para describir el contenido semántico de un recurso y están mejor preparados para soportar la recuperación de información que el propio documento.
- Los recursos de información no son capaces de facilitar por sí mismos sus propias relaciones semánticas si no dispone de una caracterización (metadatos) en la que se describa.
- Los datos requieren contener información auxiliar “datos sobre datos” que describan la información relativa al recurso, que permita identificar el contexto de la información.

La importancia del uso de metadatos en la descripción de objetos digitales reside en posibilitar su interoperabilidad entre espacios de preservación de información académica y científica. De modo que el producto de la actividad investigadora y académica pueda ser depositada en un repositorio o biblioteca digital y compartida por otros.

CAPITULO 4

Linked Open Data

En el capítulo segundo se definió la idea que presentó Berners-Lee en TED para construir la nueva web, a la que denominó Web Semántica. La idea principal es la de exponer la gran cantidad de información disponible en la red como datos, para que mediante aplicaciones puedan combinarse y reutilizarse con distintas finalidades en todas las áreas de conocimiento.

La web actual tiene la característica de que la información es comprendida por las personas y no por las máquinas, de modo que el hecho de publicarla en forma de datos tiene la finalidad de que pueda ser interpretada por las máquinas de forma automática.

Para publicar la información en datos en la web hay que seguir una serie de procesos para que cumplan unas características determinadas que permitan enlazarlos. Los datos que forman la web semántica son denominados Linked Open Data (LOD) o Datos Abiertos Enlazados.

Las características de estos datos son que están conectados con otros, cuya tipología puede ser idéntica o diferente, y que provienen de entidades externas distintas. Los datos enlazados son datos puros, de diversas tipologías, por ejemplo datos estadísticos, matemáticos, geográficos, compuestos químicos, el genoma humano... exceptuando los sensibles, que las instituciones publican en Internet. De esta forma se puede acceder a ellos, pueden ser combinados y utilizados por todos de forma libre, con el fin de producir nuevo conocimiento y aplicaciones innovadoras en el ámbito digital.

La utilización de esta especificación de datos puros permite enriquecer los datos iniciales existentes en una fuente con información experta procedente de otras fuentes externas, por ejemplo, otras bases de datos. Se trata de formar un nuevo objeto digital creado mediante la agregación de terceros objetos procedentes de una o varias fuentes, estableciendo entre ellos relaciones semánticas que estructurasen sus contenidos (Ajenjo y Hernández, 2010). Los datos, que se encuentran alojados en servidores distintos y mantenidos por entidades independientes, permiten su reutilización y combinados, referenciados desde sus distintas ubicaciones, permiten la publicación de nueva información en la web de una forma estructurada.

Linked Open Data surge en el marco de la filosofía de Open Data (OD), Datos Abiertos, un movimiento cuyo principal objetivo es que determinados datos estén disponibles de forma libre en la web, sin restricciones de propiedad intelectual. Son principalmente datos no textuales y en formatos reutilizables como CSV (*Comma Separated Values*) que representa datos sencillos en forma de tabla, y que normalmente proceden de organizaciones (Peset y Ferrer, 2011).

Es necesario que los datos que se liberan en la web estén estructurados en un formato estándar, para que las relaciones entre los datos estén disponibles y permitan así crear una red entre ellos. De esta forma el conjunto de datos interrelacionados pueden considerarse datos enlazados (*Linked Data*).

Hans Rosling (2006) dijo *“Es importante tener gran cantidad de datos accesibles para que expertos, técnicos y usuarios, puedan hacer uso de ellos, juntarlos, estudiarlos, analizarlos, para obtener nueva información, que pueda ser útil en distintos ámbitos”*.

La tecnología de los datos enlazados trata de conectar los datos puros depositados en la red desde multitud de orígenes en todo el mundo, para establecer relaciones

semánticas entre ellos. Se trata de vincular los datos a través de relaciones específicas con contenido. Para ello se describen en lenguaje RDF para representaciones gráficas. Es un modelo de descripción de metadatos creado por W3C, que aunque no se considera un lenguaje ontológico porque no permite realizar inferencias, constituye la base para otros lenguajes ontológicos.

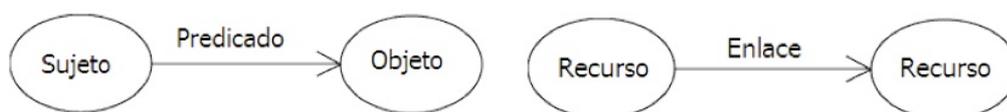
Esta tipología de datos enlazados están pensados para su lectura por parte de personas, pero también para su interpretación y procesado de forma automática por parte de agentes software. Para este propósito, Tim Berners-Lee (2006) definió los cuatro mecanismos para conseguir conectar los recursos LOD:

1. Utilizar URIs para identificar los recursos. De este modo se consigue identificar de forma unívoca y sin ambigüedades los recursos a los que se quiere hacer referencia. Por ejemplo se evitarían así las ambigüedades que puede producir el lenguaje natural en varios idiomas para referirse al mismo recurso.
2. Utilizar URIs http. Existen varios tipos de identificadores de recursos digitales, se recomienda la utilización del protocolo http de transferencia de hipertexto que defina la sintaxis y la semántica de los elementos de la web para comunicarse (cliente-servidor). En la que el cliente (navegador del usuario) realiza una petición a un servidor (base de datos) que devuelve un documento que se muestra a través de la interfaz del navegador. El uso de este protocolo permite buscar cualquier recurso almacenado en un servidor y presentarlo a través de la web.
3. Ofrecer información sobre los recursos usando RDF para que sea comprensible para el sistema informático. Al utilizar el estándar RDF para la descripción del recurso que la URIs http ha recuperado, permite al sistema reconocer la información relativa sobre la información que está representando.
4. Incluir enlaces a otros URIs. Se trata de enlazar un recurso que está identificado con su URI http y descrito en RDF, a otros recursos externos a través de sus identificadores URIs http, y que también han de estar

descritos en RDF. Los recursos enlazados pueden estar alojados en un mismo servidor, o pueden enlazar a objetos distribuidos por toda la red. Esto permite enriquecer el sitio web local con otra información experta procedente de otros dominios en la red, como también favorece evitar la duplicidad.

Una de las diferencias que surgen con la utilización de los LOD en la web en comparación con los enlaces hipertextuales se constituye en la forma de relacionar los elementos. Los LOD enlazan objetos a través de relaciones explícitamente diferenciadas. Para diferenciar estas relaciones llamadas tripletes, se escriben en lenguaje RDF. Es un lenguaje de etiquetado, con sintaxis XML, para describir los recursos a través de una estructura de triplete. Están formadas por sujeto, predicado y objeto, en las que se define tanto la estructura de la información como el significado, es decir, se define el tipo de recurso y el tipo de relación (W3C, 2012).

Ilustración 3: Representación gráfica de un triplete y representación gráfica de un enlace hipertextual.



Fuente: W3C (2010). RDF concepts (<http://www.w3.org/TR/rdf-concepts/>)

Por el contrario, en los enlaces hipertextuales o referencias cruzadas automáticas de un documento escrito en HTML las relaciones entre los recursos no están definidas, o están definidas de la misma forma a todos los recursos, de tal manera que no se especifica qué tipo de relación mantienen los recursos enlazados.

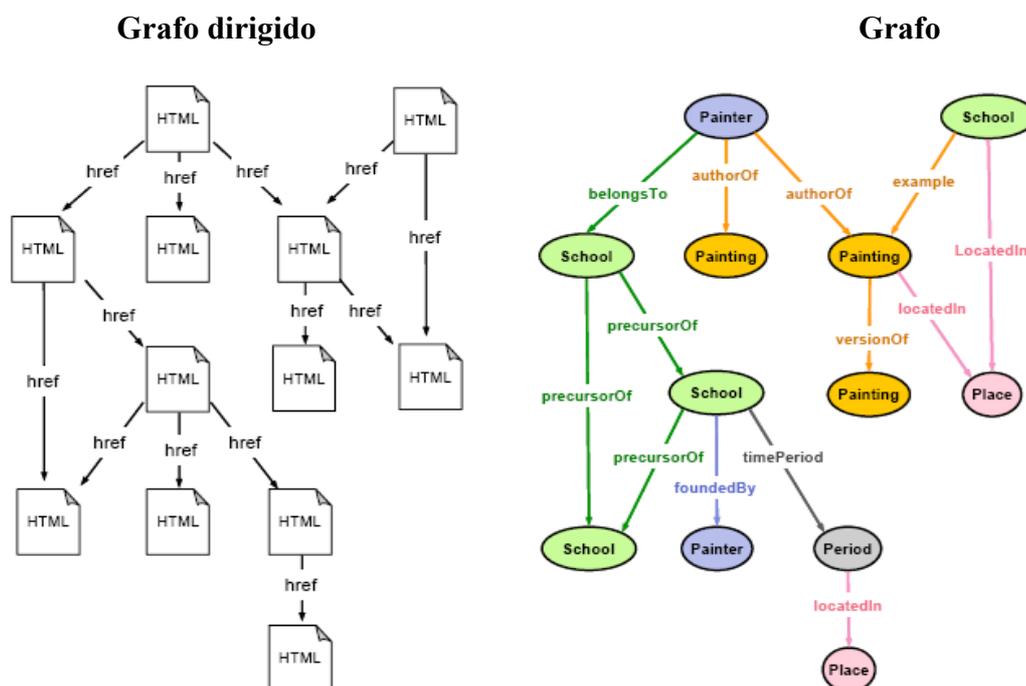
Estas relaciones se representan gráficamente en forma de grafos. El grafo cuyas relaciones entre recursos son hipervínculos es conocido como grafo dirigido, formado por nodos del mismo tipo (recursos) y arcos definidos de la misma

manera (hiperenlaces) (Castells, 2003).

En esta web el agente software realiza la presentación visual a través de una interfaz, de los recursos descritos en lenguaje HTML sin llegar a comprender el significado del contenido representado. El software interpreta el lenguaje de etiquetas HTML, que describe la estructura y parte la apariencia del documento, y muestra el contenido a los usuarios a través de la interfaz, pero no es capaz de comprender el significado de la información que ofrece. Solo los usuarios que navegan son capaces de interpretar el contenido presentado informacional presentado.

Esto quiere decir que un software interpreta de la misma manera un sitio web de una tienda *on line* que un sitio web de una asignatura de una carrera universitaria, mientras que en la web estructurada en RDF los recursos tienen un tipo especificado y las relaciones están también diferenciadas, definidas por una ontología (Castells, 2003).

Ilustración 4: Representación gráfica de las relaciones entre datos de un documento HTML (grafo dirigido), y representación gráfica de las relaciones entre datos abiertos enlazados (grafo).



Fuente: Castells, P. (2003) La web semántica.

En el grafo descrito en RDF, la estructura y el significado están definidos y cada nodo (recurso) está diferenciado, tiene un tipo de identificación (autor, pintor,...) y los arcos representan las relaciones explícitamente entre cada uno de los recursos que relacionan (autor de, localizado en, pertenece a) (Castells 2003).

En este grafo, los tres elementos de la tripleta, dos nodos (sujeto y objeto) y el arco (tipo de relación) hacen referencia a los recursos mediante las URI. Pues RDF utiliza las referencias URI para la localización de los recursos. Como se muestra en el siguiente ejemplo de Pastor Sánchez (2009) “La película Brazil fue dirigida por Terry Gilliam”

Ilustración 5: Representación gráfica de una tripleta.



Fuente: Pastor Sánchez, JA. (2009). *Diseño de un sistema colaborativo para la creación y gestión de tesauros en Internet basados en SKOS*.

Esta definición de datos permite al software identificar la relación que existe entre cada recurso e interpretar semánticamente el significado de los datos de un recurso. De esta forma un agente software es capaz de identificar relaciones semejantes y realizar deducciones lógicas, combinar información por tipologías, y generar información nueva a partir de una existente.

La sintaxis XML del lenguaje RDF facilita la interoperabilidad entre diferentes modelos de metadatos, y constituye la base de la estructura de capas de la web semántica. Este lenguaje de etiquetas permite la lectura del código no solo por la computadora, sino también por las personas. RDF se emplea en la creación de

nuevos datos que se quiere que sean semánticos, y también en la transformación de datos a LOD, pero cuando la información de un sitio web HTML, se quiere convertir a semántica, este tipo de información se puede incluir dentro del código XHTML. Se consigue a través del lenguaje RDFa, (*Resource Description Framework –in– Atributtes*) que es RDF+HTML. Consiste en insertar declaraciones RDF en las etiquetas de marcado de documentos XHTML, permitiendo así que los software puedan interpretar semánticamente el contenido. De esta forma en una misma URI hay código para la visualización por parte de las personas y código procesable por los agentes software.

Ontologías

Con las definiciones en RDF se consigue que el contenido sea comprensible también para las máquinas, con la finalidad de que puedan ser capaces de deducir conocimiento lógico. Para que las máquinas puedan inferir conocimiento se necesitan especificaciones de los conceptos definidas en vocabularios concretos. Estas conceptualizaciones del conocimiento se definen mediante unos modelos de organización del conocimiento, que hasta el momento han sido los lenguajes controlados basados en términos, pero que para proporcionar contenido semántico, son necesarios unos lenguajes más potentes basados en conceptos, son los vocabularios ontológicos.

Una ontología es un esquema de conceptos al igual que los vocabularios controlados, pero con una complejidad conceptual que las hace necesarias para que las computadoras puedan comprender semánticamente.

Las ontologías, a diferencia de los lenguajes controlados, no solo asignan descriptores a los documentos y categorizan los contenidos, sino que regulan que la especificación de los vocabularios sea correcta para que la máquina sea capaz de deducir conocimiento correctamente.

Son sistemas de organización del conocimiento, una herramienta equivalente a la forma que tienen los tesauros o las taxonomías de representar la información, junto con la especificación que aporta un sistema de metadatos. Pero como los metadatos únicamente llegan a la descripción estructurada de los recursos a través de propiedades y sus valores, para que las máquinas puedan inferir deducciones lógicas de conocimiento, son necesarios conceptos complementarios como los axiomas de las ontologías.

Según Pastor Sánchez (2009) una ontología se define como:

La descripción de tipos de entidades del mundo real y la forma en la que se relacionan entre si. Una ontología web realiza una descripción de objetos definiendo clases, propiedades, relaciones y axiomas. Las ontologías permitirían desarrollar estructuras complejas para simplificar tareas de comunicación entre personas y máquinas con la intermediación de aplicaciones informáticas, que realizan procesos de intercambio de datos y simplifican la interoperabilidad de diferentes representaciones de datos con herramientas de unificación, traducción y mapeado.

Las ontologías sirven para definir el lenguaje natural en lenguaje matemático, para que las máquinas puedan buscar contenidos semánticos. Son herramientas que mejoran el sistema de recuperación de información y relacionan el contexto, y sirven para calcular en tiempo razonable respuestas a consultas en sistemas de recuperación.

Se especifican mediante la lógica descriptiva, que a través de los axiomas, las especificaciones de las propiedades y las relaciones permiten definir las reglas que debe cumplir el lenguaje, las reglas que validan la información, comprueban la compatibilidad y permiten al sistema realizar deducciones de conocimiento razonable.

Proporcionan el vocabulario y definen los niveles jerárquicos de los conceptos, el significado de los términos y las relaciones entre ellos. Sus seis componentes son:

- *Clases*: son los conceptos que se organizan en taxonomías (clasificación jerarquizada y sistemática). Son entidades del mundo real, pueden ser objetos o conceptos, por ejemplo un *libro* o una *teoría científica*. Además, una clase puede tener subclases, normalmente la relación de subordinación más empleada en estas estructuras jerárquicas establece “tipos” de esa clase, por ejemplo la clase *libro* puede dividirse en las subclases *obra de referencia*, *manual*, *libro de texto*... Por lo general, las subclases heredan las propiedades de la clase a la que pertenecen.
- *Atributos*: son las propiedades que describen las características de los conceptos y permiten asignarles valores concretos, consisten en pares atributo/valor. Describen las características de las entidades que forman las clases.
- *Instancias*: es un ejemplar particular de cada una de las clases, es decir, un ejemplar de libro o una teoría científica específica.
- *Relaciones*: son un tipo de interacción, un enlace entre conceptos de un mismo dominio, de forma que los conceptos forman estructuras y las relaciones establecen el valor entre los conceptos. Los tipos de relaciones existentes pueden ser:
 - *Clase-instancia*: esta relación asocia un ejemplar (instancia) con la clase a la que pertenece. Es una relación de tipo *es un*, por ejemplo *obra de referencia-es un-libro*
 - *Instancia-propiedad*: las instancias (ejemplares) toman valores asociados a sus características o atributos. Esta relación es de tipo *tiene el valor*, por ejemplo la instancia *obra de referencia-tiene el valor-145* para el atributo (propiedad) *extensión*.
 - *Clase-propiedad*: la clase tiene propiedades que comparten todos sus atributos por pertenecer a ese conjunto. Se denominan *restricciones* que excluyen a otros atributos de pertenecer a esa

clase. La clase *libro* tiene la propiedad de *tiene capítulos* que excluye de esta clase a las revistas.

- *Clase-subclase*: Las clases pueden tener subclases y la relación de subordinación existente entre ellas es de tipo *es_un*.
- *Funciones*: es un tipo de relación que se establece a través del cálculo de una función de varios elementos de la ontología.
- *Axiomas*: son expresiones verdaderas, condiciones impuestas, que siempre deben cumplir los elementos de la ontología, y que sirven para inferir nuevo conocimiento de forma automatizada.

Por ejemplo “si A y B pertenecen a la clase C, entonces A no puede ser subclase de B” (Pastor Sánchez, 2009). El proceso de generar nuevo conocimiento de forma automática se basa en extraer conclusiones a partir de la información en forma de axiomas, sentencias descritas en Lógica de Primer Orden (LPO), o cálculo de predicados de primer orden (CPPO), que determinan que determinadas clases cumplen ciertas condiciones.

Además de este tipo de relaciones, existen otras entre conceptos (clases) de sinonimia, hiponimia, antonimia y meronimia. Aunque las ontologías sean lenguajes controlados estructurados en varios niveles, hay que diferenciarlas de los tesauros o taxonomías porque son más que eso, son una especificación formal y explícita de una conceptualización compartida, es decir, es una conceptualización de un fenómeno del mundo real, que especifica conceptos, propiedades y relaciones cuya finalidad es que sean comprensibles por los agentes software y soporten procesos de inferencia automáticos (Codina et al, 2009).

Las ontologías se encuentran en un nivel cognitivo superior, son una conceptualización de objetos del mundo real, y se describen a través de dos modelos de esquemas que son una extensión de RDF. Estos dos lenguajes ontológicos complementarios son *RDF Schema* (RDFS) y *Web Ontologic Language* (OWL), que se describen a continuación.

RDFS constituye un mecanismo cuya principal función es describir las ontologías simples, permite describir las clases de los recursos y las subclases, sus atributos y las relaciones entre ellos. Pero no permite declarar axiomas (Peis et al, 2003). Los principales elementos asociados a los recursos son:

Clases:

- *Rdfs:Resource* para la definición de cualquier objeto descrito mediante una tripleta RDF
- *Rdfs:Class* para la definición de clases de los recursos
- *Rdf:Property* permite definir nuevas propiedades para la descripción de los recursos

Propiedades:

- *Rdfs:subClassOf* para la definición jerárquica de las clases, que pueden ser organizadas en cuanto a generalización o especificación.
- *Rdfs:subPropertyOf* para la definición jerárquica de las propiedades. En función del nivel jerárquico que ocupen las propiedades se aplicarán a los recursos a los que implique por su posición.
- *Rdf:type* define relaciones del tipo instancia entre clases y recursos.

Restricciones de propiedades

- *Rdfs:Domain* es una instancia para definir las clases sobre las que se aplica una propiedad
- *Rdfs:Range* es una instancia para definir los posibles valores que puede tomar una propiedad

OWL es el otro lenguaje de representación de ontologías, basado en lógica descriptiva, con el que es posible inferir razonamientos en función de su nivel de expresividad y complejidad. Es un estándar propuesto por W3C, para que las máquinas puedan interpretar relaciones entre documentos, que permite definir

clases, propiedades y relaciones. Su función es dotar de significado las tripletas de RDF especificando a qué tripletas se les asigna un significado, y cuál es. Está compuesto por tres subconjuntos de lenguajes: OWL-Lite, OWL LD y OWL Full. Se asemeja a RDFS, pero es mucho más preciso y permite crear ontologías a un nivel mucho más detallado, resulta un complemento para RDF y RDFS. (Codina et al. 2009).

Los tres sub-lenguajes se caracterizan por tener distintos niveles de expresividad para representar el conocimiento. OWL-Lite es el más sencillo y menos expresivo, emplea un número menor de expresiones, proporciona una jerarquía de conceptos y unas restricciones más sencillas, para representar tesauros, taxonomías y otros esquemas conceptuales. Lo que permite adaptarse mejor a un mayor número de necesidades.

OWL LD es más extenso, tiene una gran expresividad y computabilidad, y permite inferir conocimiento automatizado. Incluye todas las construcciones del lenguaje OWL pero limitado bajo ciertas restricciones. Garantiza que todos sus resultados puedan ser computables, es decir, calculados en un tiempo finito.

OWL Full no es un sub-lenguaje sino que contiene todas las construcciones del lenguaje OWL, lo que permite una expresividad mucho más elevada y la libertad sintáctica de RDF sin restricciones. Pero al ser un lenguaje con mayor capacidad de deducción es más complejo y no asegura que los resultados puedan ser calculados en un tiempo finito.

También son diferentes entre sí en relación con el nivel de complementación que tienen con RDFS. Las clases en RDF no son equivalentes a las clases en cada uno de los sub-lenguajes OWL. Por ejemplo en OWL Full el recurso *owl:Class* es equivalente a *rdf:Class*, mientras que en OWL DL y OWL Lite *owl:Class* es una subclase de *rdf:Class* (W3C 2004).

Los elementos básicos del lenguaje OWL son las clases y las propiedades. Las clases son los conceptos más básicos, las raíces de los árboles taxonómicos *owl:Class*. Cada individuo es un miembro de *owl:Thing*. Y las clases vacías son *owl:Nothing*. Las clases más específicas y más generales se definen mediante *Rdfs:subClassOf* y las instancias con *Rdf:type* (W3C, 2004).

```
<owl:Class rdf:ID="PotableLiquid">
  <rdfs:subClassOf rdf:resource="#ConsumableThing" />
</owl:Class>
```

Por ejemplo, *PotableLiquid* es subclase de *ConsumableThing* (W3C, 2004).

```
<owl:Thing rdf:about="#CentralCoastRegion">
  <rdf:type rdf:resource="#Region"/>
</owl:Thing>
```

Por ejemplo, *rdf:type* es una propiedad RDF que une a un individuo a una clase de la cual es miembro (W3C, 2004).

Las propiedades son relaciones binarias, las hay de dos tipos, *owl:DatatypeProperty*, relaciones entre instancias de clases y literales RDF o tipos de datos XML Schema. Y *owl:ObjectProperty*, relaciones entre instancias de dos clases.

Para definir el dominio *rdfs:domain*, y el rango *rdfs:range* de una propiedad

```
<owl:ObjectProperty rdf:ID="madeFromGrape">
  <rdfs:domain rdf:resource="#Wine"/>
  <rdfs:range rdf:resource="#WineGrape"/>
</owl:ObjectProperty>
```

Fuente (W3C, 2004).

El rango puede ser usado para inferir un tipo

```
<owl:Thing rdf:ID="LindemansBin65Chardonnay">
  <madeFromGrape rdf:resource="#ChardonnayGrape" />
</owl:Thing>
```

Fuente (W3C, 2004).

Se puede inferir que “LindemansBin65Chardonnay” es un vino porque el dominio de madeFromGrape es vino.

Las propiedades de los recursos pueden ser de varios tipos, transitivas: *owl:TransitiveProperty*, simétrica *owl:SymmetricProperty*, funcional *owl:FunctionalProperty*, inversa *owl:InverseOf*, e inversa funcional *owl:InverseFunctionalProperty*. Y con las restricciones de propiedad es posible limitar aún más las propiedades que se aplican a los recursos, lo que permite inferir conocimiento. Estas son: *owl:allValuesFrom*, para todas las ocurrencias y *owl:SomeValuesFrom* para algunas ocurrencias de una propiedad, tengan en todas las instancias de una clase la propiedad de las instancias de otra clase.

También es posible definir la cardinalidad de las clases, es decir, el número de elementos que pueden componer una clase mediante *owl:cardinality*.

```
<owl:Class rdf:ID="libro">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasISBN"/>
      <owl:cardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Por ejemplo, la clase “libro” tiene la propiedad de que “tiene ISBN” y esta propiedad tiene una restricción de cardinalidad “&xsd;nonNegativeInteger” que permite indicar que tiene un valor entero positivo de 1.

Owl:HasValue asigna un valor de propiedad que permite definir instancias de clases, es decir, que determina el valor de una propiedad de una clase.

```
<owl:Class rdf:ID="tarta">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#es_dulce" />
      <owl:hasValue rdf:resource="#con_miel" />
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Donde la clase “tarta” tiene una propiedad “es_dulce” cuyo valor de propiedad es “con_miel”

Este lenguaje permite establecer un mapeado de la ontología, es decir, hacer una equivalencia de las clases y sus propiedades con otra ontología, lo que tiene una gran importancia puesto que ello implica su adaptación para su reutilización. Mediante *owl:equivalentClass* en OWL DL se indica que dos clases tienen las mismas instancias, y con *owl:sameAs* en OWL Full se expresa que dos clases son equivalentes en todos los sentidos. Con *owl:equivalentProperty* se hacen equivalentes las propiedades de las clases. Del mismo modo, los mecanismos para designar propiedades contrarias son *owl:differentFrom* y *owl:AllDifferent*.

Es posible formar nuevas clases a través de las expresiones de clases y un conjunto de operaciones sencillas como la unión *owl:unionOf*, la intersección

owl:intersectionOf y el complemento, *owl:complementOf*. Y especificar la extensión de una clase a través de la enumeración de sus miembros con el atributo *owl:OneOf*. También es posible, definir conjuntos de clases disjuntas entre sí, expresar que los miembros de una clase no pueden ser simultáneamente miembros de otra, a través del atributo *owl:DisjointWith*.

OWL también tiene en cuenta las versiones de las ontologías y permite vincular una versión anterior con la ontología que se está definiendo a través del enlace que proporciona el atributo *owl:priorVersion* dentro del elemento *owl:ontology*. Además atiende a las incompatibilidades que se pueden producir en las diversas versiones y permite identificarlas con los elementos *owl:backwardCompatibleWith* and *owl:incompatibleWith*.

Para una mayor especificación de este lenguaje el W3C tiene disponibles una serie de documentos como *OWL Guide*; *OWL Reference*; *OWL Use cases and requirements*; *OWL Semantic and abstract syntax*; *OWL Test*; *OWL XML Presentation Syntax*.

Esta versión de OWL se encuentra disponible desde 2004, pero el grupo de trabajo de OWL del W3C ha definido una nueva versión ampliada y revisada, definida como OWL2 de octubre de 2009. Esta versión tiene una estructura general muy similar con OWL1, y son perfectamente compatibles, las ontologías permanecen válidas con inferencias idénticas.

Pero OWL2 presenta nuevas funcionalidades, algunas de sus nuevas características son sintácticas, por ejemplo la unión disjunta de las clases; y otras, ofrecen nueva expresividad, como cadenas de propiedad, rangos de datos, y restricciones de cardinalidad cualificada entre otras. Además, algunas de las restricciones aplicables a OWL DL se han definido de forma menos estricta, lo que conlleva que el conjunto de grafos RDF que pueden ser manejados por las

descripciones lógicas sean mucho más amplios. (W3C, 2009). Todas estas ampliaciones se recogen en *OWL 2 Web Ontology Language: New Features and Rationale*.

Una ontología definida con OWL Full es *Simple Knowledge Organisation System* (SKOS). Su aplicación es importante porque SKOS es un estándar de migración que se ha diseñado para trasladar los sistemas tradicionales de organización del conocimiento *Knowledge Organization System* (KOS) a la web semántica. Su uso es necesario porque es un paso intermedio entre el bajo nivel de estructuración de la web con el formalismo descriptivo de las ontologías definidas con OWL (Pastor Sánchez, 2010).

Consiste en un lenguaje que se ha convertido en estándar y ha sido creada y es mantenido por W3C. Se emplea para representar conceptos, que establece la relación semántica entre los conceptos de los sistemas de organización del conocimiento, tales como tesauros, esquemas de clasificación, glosarios, listas de encabezamiento, taxonomías, folksonomías y cualquier vocabulario controlado.

Los elementos que lo forman son clases y propiedades.

Codifica la información en RDF y se puede completar con elementos de otros vocabularios como DC o FOAF, de forma que permite transmitirla entre agentes software de forma interoperable.

El uso básico de SKOS permite identificar los recursos mediante URIs, etiquetarlos con literales de uno o varios idiomas, documentarlos con diversos tipos de notas, relacionarlos entre sí mediante estructuras jerárquicas informales o redes asociativas, y agregarlos a esquemas de conceptos. (W3C, 2009).

El uso avanzado de SKOS permite mapear recursos conceptuales de distintos esquemas de conceptos y agruparlos en colecciones etiquetadas u ordenadas. Es

posible definir relaciones entre etiquetas de conceptos. Finalmente, el vocabulario de SKOS puede ser ampliado para adaptarse a las necesidades prácticas de comunidades de usuarios concretas o combinadas con otros vocabularios de modelado (W3C, 2009).

El elemento base de este lenguaje es el concepto como unidad de significado. Los conceptos se pueden definir a través de esquemas conceptuales y agrupar en colecciones etiquetadas. Estos conceptos se expresan mediante *skos:Concept* siguiendo dos procesos, el de crear una URI para identificar el concepto y el de definir en RDF utilizando la propiedad *rdf:type*.

El resultado sería:

```
<http://www.ejemplo.com/documentos> rdf:type skos:Concept.
```

La asignación de URIs a cada concepto SKOS permite referenciar de forma unívoca y crear relaciones semánticas entre conceptos de diferentes esquemas y previamente existentes.

Las agregaciones de uno o más conceptos que conforman esquemas se definen mediante *skos:ConceptScheme*. Las estructuras jerárquicas de relaciones entre conceptos y colecciones de conceptos se definen mediante *skos:Collection* (W3C, 2009).

Las etiquetas son las expresiones que se utilizan para referirse a los conceptos en lenguaje natural. Cada etiqueta expresa un estado, los tres tipos existentes son subpropiedades de *rdfs:label* y se utilizan para asociar un *skos:Concept* con una cadena de caracteres RDF.

Skos:prefLabel: es una etiqueta preferente que se aplica a un recurso, y sólo puede tener asignada una en cada idioma del lenguaje natural. Se emplean para

representar los términos utilizados como descriptores en un sistema de representación del conocimiento en varios idiomas, por ejemplo en español e inglés sería:

```
ex:documentos rdf:type skos:Concept;  
skos:prefLabel "documentos"@es;  
skos:prefLabel "documents"@en.
```

Skos:altLabel: es una etiqueta alternativa, se utiliza para proporcionar diversas entradas a un mismo concepto en un vocabulario. Sería equivalente al término relacionado en un tesoro. Se emplea con sinónimos, abreviaturas, acrónimos.

```
Ex:entidades rdf:type skos:Concept;  
skos:prefLabel "W3C"@en;  
skos:altLabel " World Wide Web Consortium"@en;
```

skos:hiddenLabel es una etiqueta oculta para determinar que los conceptos sean visibles o no a aplicaciones software. Se emplea para incluir variantes de errores ortográficos de otras etiquetas léxicas:

```
Ex:documentos rdf:type skos:Concept;  
skos:prefLabel "libros"@es;  
skos:altLabel "diccionarios"@es;  
skos:hiddenLabel "diccionarios"@es.
```

Como se acaba de ver, en los sistemas de representación del conocimiento, se representan los conceptos y las relaciones que éstos tienen con otros. Las relaciones semánticas organizan jerárquicamente los conceptos y definen así categorías. Las relaciones semánticas jerárquicas entre un género y sus especies o

un todo y sus partes, se definen mediante *skos:broader* que indica que un concepto tiene un significado más amplio que otro, y *skos:narrower* para definir que un concepto tiene un significado más específico que otro:

ex:animales rdf:type skos:Concept;	Ex:mamíferos rdf:type skos:Concept;
skos:prefLabel “animales”@es;	skos:prefLabel “mamíferos”@es;
skos:narrower ex:mamíferos.	skos:broader ex:animales..

Para las relaciones semánticas no jerárquicas que representan vínculos asociativos de dos categorías en la que ningún concepto es más genérico que otro se emplea *skos:related*.

```
Skos:related:
ex:bibliotecas rdf:type skos:Concept;
skos:prefLabel “bibliotecas”@es;
skos:related ex:biblioteconomía
```

Skos:related no es una propiedad transitiva, por tanto no permite inferir declaraciones a partir de vínculos asociativos: por ejemplo A(Química) está directamente relacionado con B (H2o), y B directamente relacionado con C (Ecologías), no debe establecer se que A y C estén relacionadas.

Para las relaciones jerárquicas transitivas que permiten realizar inferencias se emplean *skos:broaderTransitive* y *skos:narrowerTransitive*.

El lenguaje SKOS tiene características como aportar información sobre el significado de conceptos, el uso, etc, comprensible para las personas, tanto para usuarios como para gestores de la información, a través de las notas. Notas de diferentes naturalezas como *skos:definition*, *skos:example*, *skos:scopeNote*, *skos:historyNote*, *skos:editorialNote*, *skos:changeNote*.

Otra característica es que permite el uso de propiedades de otros lenguajes para completar informaciones, por ejemplo la propiedad `dc:creator` de Dublin Core, que indica la persona que creó el concepto águila pescadora de Madagascar:

```
ex:águilaPescadoraMadagascar dct:creator [ foaf:name "John Smith" ]
```

Fuente: (W3C, 2009).

Esta característica permite emplear las propiedades de otros lenguajes para utilizar vocabularios controlados y esquemas de clasificación externos, específicos de distintas materias, y vincularlos mediante la clase `skos:conceptScheme`; y `skos:inScheme`, que vincula un mismo concepto en varios esquemas.

El ejemplo recogido en el manual de SKOS del grupo de trabajo del W3C, recoge el uso de un tesoro sobre animales, emplea la propiedad `dc:creator` y `dc:title` de Dublin Core, para identificar el tesoro que se utiliza, y las propiedades `skos:concept` y `skos:inScheme` para vincular un concepto al tesoro:

```
ex:tesauroAnimales rdf:type skos:ConceptScheme;  
dct:title "Sencillo tesoro sobre animales";  
dct:creator ex:antoineIsaac.
```

```
Ex:mamíferos rdf:type skos:Concept;  
skos:inScheme ex:tesauroAnimales.
```

```
Ex:vacas rdf:type skos:Concept;  
skos:broader ex:mamíferos;  
skos:inScheme ex:tesauroAnimales.
```

```
Ex:peces rdf:type skos:Concept;  
skos:inScheme ex:tesauroAnimales
```

Al igual que en los vocabularios controlados, se indican los puntos de entrada de las jerarquías, asignando las propiedades `skos:topConceptOf` y

skos:hasTopConcept a los conceptos que son encabezamiento de una jerarquía. *Skos:hasTopConcept* permite vincular un esquema de conceptos a uno o varios conceptos cabecera, y *skos:topConceptOf* a la inversa, del concepto al esquema.

Representar un Sistema de Organización del Conocimiento mediante SKOS sirve también como un mecanismo que permite participar en una red de esquemas de conceptos, para definir estas relaciones de mapeado se utilizan las propiedades *skos:exactMatch*, cuando dos esquemas son idénticos, y *skos:closeMatch* cuando tienen un significado cercano aunque no idéntico (W3C, 2009).

Para una mayor definición del lenguaje, el W3C ofrece varios documentos, *SKOS Guide Reference*; *SKOS Primer*; *SKOS Use Cases and Requirement*.

Otro lenguaje que es necesario para describir en RDF a las personas y entidades es FOAF (*Friend of a Friend*). Surgido en el año 2000, constituye una ontología para describir personas y las relaciones existentes entre ellas. Se trata de una especificación de vocabulario que permite describir las relaciones escritas en lenguaje RDF a través de clases y propiedades. Permite describir informaciones sobre las redes sociales y de colaboración entre los individuos.

Declara las siguientes clases y sus propiedades (Brickley, 2010):

- Clases: Agent, Document, Group, Image, LabelProperty, OnlineAccount, OnlineChatAccount, OnlineEcommerceAccount, OnlineGamingAccount, Organization, Person, PersonalProfileDocument, Project.
- Propiedades: Account, AccountName, accountServiceHomepage, age, aimChatID, based_near, birthday, currentProject, depiction, depicts, dnaChecksum, familyName, family_name, firstName, focus, funddBy, geekcode, gender, givenName, givenname, holdsAccount, homepage, icqChatID, img, interest, isPrimaryTopicOf, jabberID, knows, lastName, logo, made, maker, mbox, mbox_sha1sum, member, membershipClass,

msnChatID, myersBriggs, name, nick, opend, page, pastproject, phone, plan, primaryTopic, publications, schoolHomepage, shal, skypeID, status, surname, theme, thumbnail, tipjar, title, topic, topic_interest, weblog, workInfoHomepage, workplaceHomeplace, yahooChatId.

Las ontologías constituyen una herramienta de representación del conocimiento, similar a las empleadas en Ciencias de la Información, pero mediante su definición con lógica descriptiva permiten deducir conocimiento, pueden relacionar el contexto y consultar tanto por el valor de sus propiedades como por los elementos de la relación, de forma que consiguen mejorar las respuestas de los sistemas de recuperación de información, con mayor exhaustividad en los resultados obtenidos.

Para que la recuperación de información sea selectiva, es decir, únicamente los datos que se precisen, hay que consultar las fuentes de datos enlazados mediante una interfaz de SPARQL.

SPARQL (*SPARQL Protocol and RDF Query Language*) es un lenguaje de interrogación que se utiliza para construir la consulta de grafos RDF almacenados. Con este lenguaje se definen las tripletas con las que se interroga al sistema, o en este caso la web, y como resultado recupera las sentencias RDF de un grafo que responden a los criterios de búsqueda definidos. Esto permite recuperar las sentencias RDF que se encuentran distribuidas por la web y reutilizar las que resulten necesarias para otros propósitos. Forma parte de los estándares del W3C desde 2008 (W3C, 2008).

SPARQL *Endpoint* es la interfaz general de consulta que recupera las sentencias RDF almacenadas en los *triplestore*, o repositorios de tripletas, que son sistemas de gestión similares a las bases de datos. Existen *endpoints* generales, y específicos de conjuntos de datos de instituciones. En los generales, cuando se

define la consulta hay que identificar la fuente de datos a la que tiene que interrogar, y es posible combinar varias fuentes mediante la sentencia FROM. Estas interfaces permiten la descarga de los datos recuperados en diversas sintaxis como HTML, XML, RDF/XML, JSON, Spreadsheet, Javascript, Ntriples.

La mayoría de las consultas contienen un conjunto de patrones triples, similar a las tripletas RDF que se van a recuperar, en el que cada uno de los elementos sujeto, predicado y objeto son una variable. Las partes fijas que se consultan son los arcos, que son lo que se compara en la búsqueda para extraer del grafo rdf los contenidos de los nodos que cumplan con los requisitos especificados.

Las clausulas existentes para realizar las consultas son: *select*, *construct*, *ask* y *describe*.

SELECT: devuelve todas o un subconjunto de variables que coinciden con el patrón expresado, y necesita la clausula WHERE que expresa el patrón gráfico para coincidir con el gráfico de datos. Las consultas pueden incluir tantas restricciones como sea necesario.

El resultado de esta consulta se expresa en forma de tabla. Como se describe en el ejemplo del W3C 2008:

Datos:

<http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title>
“SPARQL Tutorial”.

Consulta:

```
SELECT ?title
WHERE
{
  <http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title> ?title.
}
```

Resultados:

title
“SPARQL Tutorial”

Los modificadores de SELECT son:

- **DISTINC:** para que no devuelva resultados repetidos
- **ORDER BY:** para ordenar los resultados por algún criterio definido

Los filtros que se pueden emplear:

- Restricciones en los valores de los atributos: xsd:integer, xsd:decimal, xsd:float, xsd:double, xsd:string, xsd:boolean, xsd:dateTime
- Operadores matemáticos y lógicos: <, >, =, A*B, A/B, A+B, A-B, &&(y), II (o)
- Funciones: STR(A), LANG(A), DATATYPE(A), LangMATCHES(A,B), sameTerm(A,B), isURI(A), isLiteral(A)...

CONSTRUCT: devuelve un grafo RDF debido a que genera nuevas tripletas RDF a partir de los datos de las sentencias recuperadas que cumplen el patrón especificado en la consulta. También necesita la clausula WHERE para especificar las condiciones de la consulta. Como puede observarse en el siguiente ejemplo del W3C 2008:

Datos:

@prefix org: <http://example.com/ns#>.

_:a org:employeeName "Alice" .

_:a org:employeeId 12345 .

_:b org:employeeName "Bob" .

_:b org:employeeId 67890 .

Consulta:

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX org: <http://example.com/ns#>

CONSTRUCT { ?x foaf:name ?name }

WHERE { ?x org:employeeName ?name }

Resultados:

@prefix org: <http://example.com/ns#>.

_:x foaf:name "Alice" .

_:y foaf:name "Bob" .

ASK: Devuelve un operador booleano que indica si hay o no coincidencias con el patrón de consulta expresado. No necesita la cláusula WHERE y puede tener restricciones.

DESCRIBE: Devuelve un grafo RDF con los datos de los recursos que cumplen las restricciones de la consulta. La cláusula WHERE es opcional. Esta cláusula por sí misma puede contener un IRI (*Internationalized Resource Identifier*), un subconjunto de referencias del protocolo URI. Es una cadena compacta de caracteres con sintaxis controlada, del Conjunto Universal de Caracteres, Unicode/ISO10646, para identificar de forma única un recurso físico o abstracto:

DESCRIBE <http://example.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

DESCRIBE ?x

WHERE { ?x foaf:mbox <mailto:alice@org> }

Es posible realizar consultas sobre varias colecciones simultáneamente y los resultados pueden ser diferentes de lo que se obtendría al realizar la consulta por separado a cada colección:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc:
<http://purl.org/dc/elements/1.1/> SELECT ?g ?who ?mbox FROM
<http://example.org/aliceFoaf>

FROM <http://example.org/bobFoaf>

WHERE { ?g dc:publisher ?who . ?who foaf:mbox ?mbox } }
```

También es posible indicar las colecciones sobre las que se quiere ejecutar cada parte de la consulta

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX data:
<http://example.org/foaf/> SELECT ?nick { GRAPH
<http://example.org/foaf/aliceFoaf> {

?x foaf:mbox <mailto:bob@work.example>. ?x foaf:nick ?nick}}
```

Y se pueden aplicar restricciones diferentes a cada colección que se va a consultar:

```
PREFIX data: <http://example.org/foaf/> PREFIX foaf:
<http://xmlns.com/foaf/0.1/> SELECT ?mbox ?nick WHERE {

GRAPH data:aliceFoaf { ?alicia foaf:mbox <mailto:alice@work.example> ;
?alice foaf:knows ?whom .} .

GRAPH data:bobFoaf{ ?whom foaf:mbox ?mbox. ?whom foaf:nick ?nick }}
```

Una característica particular y novedosa de estos sistemas de consulta es que la tecnología semántica responde “no lo sé” cuando no hay datos para responder la consulta, es decir, cuando con la información del modelo no puede inferir ese conocimiento. Y cuando se introduce nueva información en el sistema, las

respuestas siempre son consistentes con las anteriores, porque gracias a las ontologías el sistema es capaz de identificar los errores incoherentes. A diferencia de los sistemas de consulta sintácticos, como las bases de datos, cuyas respuestas son negativas cuando falta información en el sistema.

Para una descripción más amplia del lenguaje de consulta SPARQL, W3C pone a disposición de los usuarios una recomendación de 2008 de la versión 1.0 y la versión 1.1 de enero de 2012 (W3C, 2012), que incluye extensiones de los documentos de la versión anterior, la descripción de servicios como modelos de inferencia, y consultas federadas: consultas a servicios *endpoint*, para suplir las nuevas necesidades como la inferencia implícita de la primera versión y la falta de operaciones como el anidamiento.

CAPITULO 5

Ejemplo de LOD aplicado en información bibliográfica

El proceso de evolución de la web junto con las estructuras de información tales como las ontologías, y los lenguajes de descripción de los datos enlazados que se han relacionado hasta aquí, constituyen los elementos básicos para los cambios que se están experimentando en los sistemas de datos propios de las bibliotecas.

El desarrollo de estos elementos está estrechamente vinculado a su aplicación en el ámbito bibliotecario, debido a que la identificación unívoca, la desambiguación de términos, así como las relaciones entre los términos y el modo normativo de referenciarlas, necesarias en el uso de LOD para la correcta creación de lo que los expertos han denominado como web semántica, forman parte de las prácticas bibliotecarias y bibliográficas tradicionales. Y son el alimento básico de casi cualquier ontología, esquema o aplicación semántica (Ajenjo y Hernández, 2010).

La primera iniciativa desarrollada sobre la aplicación de LD fue Dbpedia, un proyecto iniciado en 2007 por la Universidad de Leipzig y la Universidad Libre de Berlín, en el que el objetivo era la transformación de forma automática, de los contenidos de Wikipedia en datos estructurados, en RDF, y su libre disponibilidad en la web.

Según la organización DBPedia (<http://dbpedia.org/About>) en 2012, los *datasets* describen más de 3.64 millones de objetos digitales, de los cuales 1.83 millones están organizados jerárquicamente en la denominada *Ontología de DBPedia*, incluyendo 416,000 de personas, 526,000 lugares, 106,000 álbumes de música, 60,000 películas, 17,500 videojuegos, 169,000 organizaciones, 183,000 especies y 5,400 enfermedades. Estos elementos se encuentran en 97 lenguas distintas.

Como se observa en el fragmento del diagrama de la nube de LOD, Dbpedia ocupa el lugar central debido a que es el *dataset* publicado en formato LOD más enlazado con otros *datasets* externos en la actualidad, 2,724,000 enlaces a imágenes, 6,300,000 de enlaces a páginas web externas, y 6,200,000 enlaces externos a otros *datasets* en RDF.

Está enlazado con proyectos de importantes instituciones como:

Proyecto Gutenberg (<http://thedatahub.org/dataset/fu-berlin-project-gutenberg>),

British National Bibliography (BNB) de la British Library (<http://thedatahub.org/dataset/bluk-bnb>),

Open Library (<http://thedatahub.org/dataset/openlibrary>),

Europeana (<http://thedatahub.org/dataset/europeana-lod>),

PubMed (<http://thedatahub.org/dataset/bio2rdf-pubmed>),

LCSH de la Libray of Congress (<http://thedatahub.org/dataset/lcsh>),

Biblioteca de la Univeresidad de Manheim (<http://thedatahub.org/dataset/ub-mannheim-linked-data>),

VIAF: Fichero de Autoridades Virtual Internacional (*Virtual International Authority File*) (<http://thedatahub.org/dataset/viaf>), que unifica ficheros de autoridad en múltiples idiomas y procedentes de múltiples fuentes.

New York Time, *CERN Library*, entre otros muchos.

Ilustración 6: Fragmento del diagrama de la nube LOD.

de datos propio para estructurar y enlazar semánticamente los recursos de las instituciones que forman Europeana.

EDM es una versión nueva, que se basa en los principios LOD para sustituir al modelo empleado hasta el momento ESE (Elementos Semánticos de Europeana).

EDM integra los estándares de descripción de LOD para establecer relaciones semánticas entre los objetos, para llevar a cabo el proceso de agregación de recursos digitales europeos por medio de la recolección de metadatos. La definición de la estructura de objetos digitales viene dada por OAI-ORE, el modelo de representación que emplea es RDF, incluye ontologías como SKOS, que define los conceptos y los representa en un modelo de representación del conocimiento, DC que define las características esenciales de los objetos culturales, y FOAF.

Las instituciones que forman parte de Europeana son proveedoras de contenidos que proporcionan metadatos descriptivos de los objetos digitales y todas sus relaciones, con el autor, con las versiones de la obra, los eventos asociados, las copias, temáticas... La diversidad de los estándares de descripción de cada una de estas comunidades: bibliotecas, bibliotecas digitales, archivos, museos y el sector audiovisual, pueden ser satisfactoriamente mapeados por EDM debido a la flexibilidad y adaptación a las clases y propiedades de cada comunidad.

Otra iniciativa desarrollada de uso de LOD en información bibliográfica y de autoridades es el Proyecto de Datos Enlazados en la Biblioteca Nacional Española, que lleva a cabo el Ontology Engineering Group (OEG) de la Universidad Politécnica de Madrid. Este proyecto consiste en la transformación de los catálogos bibliográfico, descritos en el formato Marc21, y catálogo de autoridades descrito en FOAF-EAC, a su equivalencia en el lenguaje RDF.

La finalidad es proporcionar de forma estandarizada la descripción de datos

bibliográficos que faciliten el intercambio de registros bibliográficos a nivel internacional. Para ello han utilizado los vocabularios y ontologías de la IFLA como el estándar internacional para descripción bibliográfica ISBD (*International Standard for Bibliographic Description*) y el modelo de requisitos funcionales para registros bibliográficos FRBR (*Functional Requirements for Bibliographic Records*). Han enlazado con otros conjuntos de datos disponibles en la nube como VIAF (*Fichero de Autoridades Virtual Internacional*), SUDOC, GND (German National Library) LIBRIS y DBPedia (Facultad de Informática, Universidad Politécnica de Madrid, 2011).

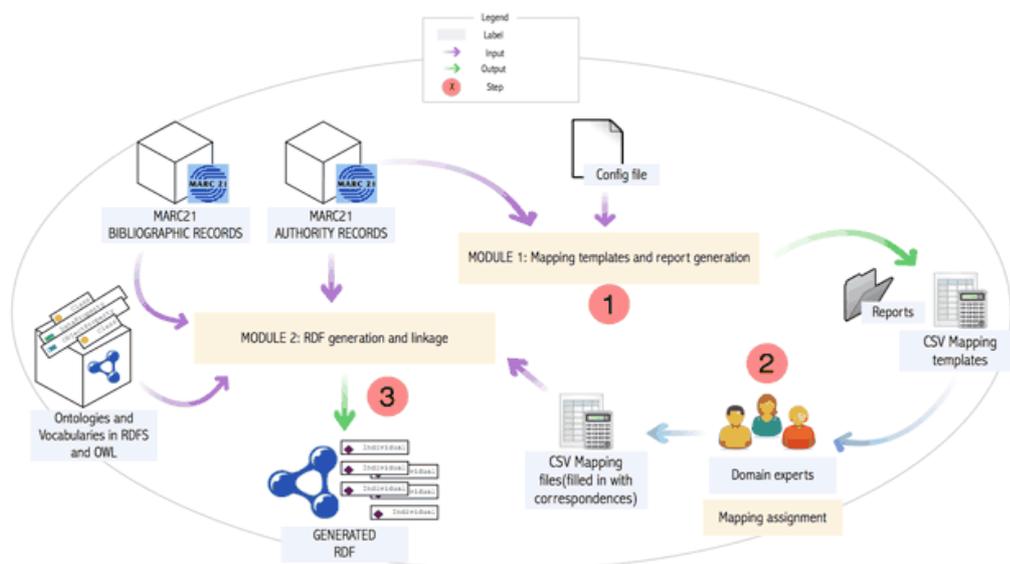
El proceso de migración ha producido cuatro millones de registros de autoridad y 2,4 millones de registros bibliográficos migrados, 58 millones de tripletas RDF y 600K de enlaces con recursos externos. Se ha llevado a cabo mediante un software llamado Marimba (MARC Mappings and rdf generAtor), una herramienta de conversión de MARC21 a RDF, desarrollada por este mismo grupo de trabajo y diseñada para el uso bibliotecario.

Esta herramienta de transformación facilita la realización de las actividades necesarias para establecer las equivalencias entre los atributos de MARC21 y las correspondencias entre los atributos y clases de RDF. Estos procesos son: el análisis de los registros MARC 21, su clasificación y asignación de un tipo o clase de los modelos RDFS/OWL elegidos para modelar los datos, lo que genera un recurso de un determinado tipo por cada registro MARC 21. La anotación o descripción de los recursos generados, usando los metadatos de los registros MARC 21, y la generación de relaciones entre los recursos generados y anotados.

La herramienta permite utilizar cualquier modelo o vocabulario RDFS/OWL, asignar las correspondencias sin necesidad de utilizar tecnologías complejas, a través de una serie de hojas de cálculo. Y proporcionar información analítica en los propios documentos de correspondencias para facilitar las decisiones. Esta información analítica permite conocer la cobertura de la transformación a RDF, en

otras palabras permite conocer aquellos metadatos que no están siendo transformados a RDF (Biblioteca Nacional Española, 2012).

Ilustración 7: Proceso y flujo de trabajo seguido por la herramienta.



Fuente: Biblioteca Nacional Española, Tecnologías, 2012.

El portal es *datos.bne.es*, y también cuenta con un SPARQL Endpoint (<http://datos.bne.es/sparql>) para las consultas en RDF. A través de la herramienta PUBBY es posible añadir interfaces de Linked Data a SPARQL endpoints. Gran parte de los datos de la Web Semántica residen dentro de las triples stores (repositorios de tripletas) y sólo se puede acceder mediante el envío de consultas SPARQL a un SPARQL endpoint. Pubby hace que sea fácil convertir un SPARQL endpoint en un servidor de Linked Data. Se implementa como una aplicación web en Java (Biblioteca Nacional Española, 2012).

Pero aunque ya son muchas las herramientas disponibles en la web para publicar datos en RDF, es necesaria una herramienta que permita encontrar relaciones entre recursos pertenecientes a distintas fuentes de Linked Data, que consiga unir los enlaces RDF de nuevas colecciones de datos propios, con otras fuentes de

datos de la web. La herramienta que realiza esta función es SILK, que permite a los editores de datos enlazados definir el tipo de relaciones RDF que hay que buscar en las fuentes externas. SILK accede a las fuentes de datos a través del protocolo SPARQL (Volz et al, 2009. Isele et al. 2011).

CAPÍTULO 6

Conclusiones

Podría decirse que nos encontramos en un periodo de transformación del ámbito digital y es posible referirse a este cambio como un nuevo paradigma tan importante que afecta no solo a las bibliotecas digitales, sino a todas las instituciones de carácter superior, a las instituciones públicas y a las privadas, en definitiva a la mayoría de organismos con contenidos en el ámbito web.

El proceso que se ha ido desarrollando con la evolución de la web, el aumento de su uso en la actividad diaria, su utilización como fuente de información y la adaptación de sus servicios a las características particulares de sus usuarios está desembocando en la transformación de la forma de publicar y compartir los contenidos digitales. En el ámbito académico y científico, la transformación es visible en lo que constituyen las actuales formas de publicación del conocimiento científico y técnico, y la tendencia actual hacia una mayor publicación en acceso abierto, que continúa evolucionando hacia la publicación de datos abiertos enlazados.

El movimiento de OA continúa siendo uno de los puntos de debate que más literatura ha generado, especialmente relacionado con el conflicto con el modelo tradicional, los beneficios que conlleva su aplicación para profesionales y usuarios, la viabilidad de su implantación, la adopción por parte de las editoriales y la preservación en repositorios.

Según el informe de John Whitfield de 2011 publicado en la revista *Nature* (Whitfield, 2011) el número de trabajos en acceso abierto incrementa desde el año 2000. El número de artículos ha tenido un crecimiento de un 20% anual, y las

revistas de acceso abierto un crecimiento del 15% frente al 5% de las de acceso por suscripción (Aleixandre, 2012).

Podría hablarse de consolidación del movimiento OA, que se ve respaldado por las políticas públicas nacionales e internacionales, como la Ley Española de la Ciencia de 2011, la ley Europea de 2003 (Directiva 2003/98/CE), el proyecto PEER (*Publishing and the Ecology of European Research*), financiado por la Comisión Europea dentro del programa eContentplus, que abogan por el almacenamiento en repositorios de la producción científica desarrollada con financiación pública.

Estos mandatos y directrices se están adoptando por instituciones de carácter superior, e incluso las principales editoriales publican algunos de los artículos académicos en acceso abierto como Elsevier, Springer, Taylor & Francis, Oxford University Press, Cambridge University Press o el grupo Nature, entre otros.

Es la comunidad científica la que está invirtiendo un mayor esfuerzo en el desarrollo de proyectos vinculados al acceso abierto y a los datos enlazados, pero desde la Ley 37/2007 han ido aumentando las iniciativas sobre datos abiertos y reutilización liderados por la Administración General del Estado y lo que se conoce como *Open Government Data*. Estos esfuerzos constituyen una labor para evitar duplicar recursos y trabajo, y favorecer la reutilización en el sector público.

Así mismo, desde la comunidad bibliotecaria existe un gran interés por el uso de datos enlazados, donde se están llevando a cabo numerosos procesos de transformación de catálogos de autoridades y bibliográficos al lenguaje RDF como en la Biblioteca Nacional de España, British Library y Deutsche National Bibliothek. Europeana es otro fructífero ejemplo de descripción de objetos digitales bajo su modelo de datos para el intercambio de datos entre sus participantes.

Los proyectos desarrollados muestran que los datos enlazados constituyen una realidad, aunque los niveles de compartición aún sean bajos y crezcan con

lentitud, se puede afirmar que este proceso ha superado el marco únicamente teórico y se encuentra en el punto de transformación de los datos.

Los esfuerzos se están dirigiendo a la transformación de datos biográficos y bibliográficos de sistemas de descripción tradicionales a lenguajes propios de la web semántica, de una forma casi automática, siguiendo un proceso de especificación en el que definen las URIs, un modelado en el que se definen las ontologías y taxonomías, la generación de RDF y links, la publicación y la consulta a través de las interfaces SPARQL.

Estos procesos de transformación han tenido su auge durante el 2011, y se han visto reforzados por numerosos congresos y debates que han tenido lugar en nuestro país sobre la web semántica. En estos congresos, cursos y jornadas se ha puesto de manifiesto la importancia que supone la normalización para el intercambio de datos de las instituciones participantes. Así como también los beneficios que proporciona la transformación de los datos en datos abiertos enlazados para alcanzar una serie de objetivos en el intercambio de información y su reutilización. Consiguiendo así evitar duplicar esfuerzos, y ahorrar costes compartiendo información de fuentes de datos interrelacionadas entre sí, y alojadas en ubicaciones distintas en la web.

Otros esfuerzos se están dirigiendo al desarrollo de mecanismos para la transformación de metadatos OAI-PMH en tripletas RDF e identificadores URI desreferenciables, para permitir preservar *datasets* LOD en repositorios institucionales.

En definitiva, en lo que se está el trabajo en estos momentos es en la publicación, en la creación de vocabularios controlados, en la adaptación de los sistemas existentes para generar fuentes de conjuntos de datos descritos en los vocabularios semánticos, y en su transformación para preservarlos en repositorios OA. Pero este parece ser el primer paso, y en segundo lugar habría que plantearse el consumo de estos datos, lo que podría definirse realmente como reutilización.

A pesar de la importancia que supone la utilización de los estándares para la descripción de los recursos y para su informatización, no hay estandarización en cuanto a la reutilización de datos. En ello ha trabajado el Grupo Incubadora sobre Datos Abiertos Bibliotecarios del W3C, y el 25 de agosto de 2011 publicó su informe final en el que establece directrices a llevar a cabo por las bibliotecas para integrar sus datos como *linked data*. Y sobre cómo buscar y utilizar *linked data* externos para enriquecer los servicios bibliotecarios en la web semántica, mediante el uso de los vocabularios ontológicos de organización del conocimiento, orientados a la inferencia de conocimiento por parte de los sistemas para realizar búsquedas de datos enlazados en los archivos *triplestores*.

Con el informe final del W3C se plantean las directrices para alcanzar el objetivo de consumir los datos. Debido a que la idea final que se desea conseguir a través de todas estas acciones de publicar datos abiertos, es la utilización de conocimiento distribuido. Es decir, utilizar contenidos producidos en la actividad investigadora de otros, para generar nuevos servicios o productos.

Estos elementos marcan el grado de desarrollo de los datos y tecnologías semánticas en la sociedad actual. Desde el marco regulatorio y las políticas que establecen recomendaciones, la implantación de sistemas se va consolidando, los vocabularios controlados disponibles se amplían a áreas técnicas y científicas, el número de conjuntos de datos que se publican aumenta exponencialmente, y el número de repositorios que acepta *datasets* para su preservación, está en aumento. Y por último el desarrollo de aplicaciones de consumo de datos.

En este sentido, cabe preguntarse por modelos de negocio. Saorín (2012), resalta la perfecta adaptación de las tecnologías *linked data* a una forma de innovación en el sector industrial y bibliotecario, definida como Innovación abierta, término acuñado por Henry Chesbrough en su libro *Open business models how to thrive in the new innovation landscape 2006*. Este término revela, en el ámbito empresarial, la estrategia de emplear conocimiento, objetos de propiedad intelectual, y resultados de actividad científica, procedente de entidades externas para mejorar

los propios servicios y procesos de investigación. Y de esta forma crear un valor añadido y generar innovación a partir de recursos que otras empresas han puesto a disposición de terceros (Chesbrough, 2006).

Si bien es cierto que este puede ser un factible modelo de negocio, la reutilización es posible que llegue antes al ámbito público que a la empresa, tanto por las regulaciones que establece el sector público, como por la barrera que supone para las empresas el compartir sus productos de propiedad intelectual con agentes externos a la organización.

En conclusión, todos estos procesos se están desarrollando al mismo tiempo, con el objetivo de integrar los datos de fuentes heterogéneas, en diferentes lenguajes y diferentes países, para compartir y reutilizar la información. La clave es establecer equivalencias en la terminología, empezar por reutilizar los modelos y vocabularios consensuados por las autoridades competentes, reutilizar las URIs que han definido de las fuentes autorizadas, y emplear estos modelos estandarizados para aplicarlos a los casos particulares. Y siguiendo las mismas bases, conseguir entre todos los participantes de la web de los datos, ir tejiendo poco a poco y enlazar, compartir y reutilizar la información disponible en lo que se está definiendo como la nueva web.

BIBLIOGRAFÍA

- AENOR. Información y documentación. Conjunto de elementos de metadatos Dublin Core. UNE-ISO 15836. Madrid: AENOR, 2007
- Aleixandre Benavent, Rafael. (2012). La situación científica en 2011. *Anuario ThinkEPI*, v.6 pp112-118
- Ajenjo Bullón, Xavier y Hernández Carrascal, Francisca. (2010). La Biblioteca Virtual: función y planteamiento. <http://eprints.rclis.org/bitstream/10760/14352/1/La_biblioteca_virtual_fin_al.pdf> [consultado: 22/09/2011]
- Agenjo Bullón, Xavier y Hernández Carrascal, Francisca. (2010). La Biblioteca Virtual Larramendi: fuente de información bibliográfica para el pensamiento iberoamericano en la Web 3.0 <http://www.redcid.org/index.php?option=com_mtree&task=att_download&link_id=66&cf_id=24> [consultado: 22/09/2011]
- Baiget, Tomás. (2010). Valor económico de la información. *Anuario ThinkEPI*, v. 4, pp.120-124
- Barrueco Cruz, José Manuel, y Subirats Coll, Imma. (2003). Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo. *El profesional de la información*, vol. 12, n 2, pp.99-106. <<http://eprints.rclis.org/archive/00000177/02/cardedeu.pdf>> [consultado: 22/07/2011]

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Disponible en: <http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf> [consultado: 22/07/2012]

- Berners-Lee, Tim. (2000). *Tejiendo la Red: el inventor del World Wide Web nos descubre su origen*. Madrid: siglo XXI de España editores.

- Berners-Lee, Tim. (2009). On the next web. Ted.com
<http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html>
[consultado: 12/08/2011]

- Berners-Lee, Tim. (2006). Linked Data.
<<http://www.w3.org/DesignIssues/LinkedData.html>> [consultado: 10/12/2011]

- Berners-Lee, Tim; Hendler, Lassila. (2001). The Semantic Web. *Scientific American*, Vol. 284, num. 5, pp. 34-43

- Bethesda Statement on Open Access Publishing (2003).
<<http://www.earlham.edu/~peters/fos/bethesda.htm>> [consultado: 22/07/2012]

- Biblioteca Nacional Española. (2012) Tecnología.
<<http://www.bne.es/es/Catalogos/DatosEnlazados/Tecnologia/>>
[consultado: 17/04/2012]

- Borgoñós martínez, MD. (2010). El movimiento Open Archives Initiative (OAI) y su repercusión en la difusión del Conocimiento. *Anales de Documentación*, vol.13, pp.23-40

- Borillo Domenech, Ricardo. Gumbau Mezquita, José Pascual. (2008). Arquitecturas orientadas a servicios para la preservación de objetos

digitales. *El documento electrónico: aspectos jurídicos, tecnológicos y archivísticos*, Alicante: Servicio de publicaciones de la Universitat Jaume I, pp. 215-236

- Brickley, Dan; Miller, Libby. (2010). FOAF Vocabulary specification 0.98 Namespace Document 9 August 2010. <<http://xmlns.com/foaf/spec/>> [consultado: 17/04/2012]
- Budapest Open Access Initiative (BOAI). Open Society Institute (OSI). Disponible en: <<http://www.soros.org/openaccess/>> [consultado: 22/07/2012]
- Bush, Vannevar. (1945). As We May Think. *The Atlantic Montly*, july (Traducción al castellano en Vannevar Bush. Cómo podríamos pensar. *Revista de Occidente*, N° 239, marzo 2001. Disponible en: <<http://sindominio.net/biblioweb/pensamiento/vbush-es.pdf>>.
- Castells, Pablo. (2003). La web semántica <<http://www.ii.uam.es/~castells>> [consultado: 22/07/2011]
- Chesbrough, Henry (2006). *Open busines models how to thrive in the new innovation landscape*. Cambridge: Harvard Business Review Pres.
- Codina, Lluís; Marcos, Mari-Carmen; Pedraza, Rafael. (2009). *Web semántica y sistemas de información documental*. Gijón: Edicione Trea.
- Daudinot Founier, Isabel. (2006). Descripción de los recursos de información en Internet: formato Dublín Core. *Acimed: revista cubana de los profesionales de la información y la comunicación en salud*, Vol. 14, N°. 4, Disponible en: <http://bvs.sld.cu/revistas/aci/vol14_4_06/aci09406.htm> [consultado: 15/10/2011]

- Elba García, Nelida. Caballero, Sergio. (2009) Metadatos: necesidad e importancia de integrar estándares.
<http://www.bn.gov.ar/descargas/catalogadores/ponencias/251109_10a.pdf> [consultado: 15/10/2011]

- España. Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Boletín Oficial del Estado, núm. 131 de 2 de junio de 2011, páginas 54387 a 54455
<<http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf>>[consultado: 10/11/2011]

- España. Ley 37/2007, de 16 de noviembre, sobre reutilización de la información en el sector público. Boletín Oficial del Estado, núm. 276 de 17 de noviembre de 2007, páginas 47160 a 47165 <<http://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814>>[consultado: 10/11/2011]

- Europa. Directiva 2003/98/ce del Parlamento Europeo y del Consejo de 17 de noviembre de 2003 relativa a la reutilización de la información del sector público. Diario Oficial de la Unión Europea. Num. 345 de 31 de diciembre de 2003. páginas 90-96
<http://www.datos.gob.es/datos/sites/default/files/files/17nov03DirectivaESP_200398CE.pdf> [consultado: 10/11/2011]

- European Research Advisory Board. (2006). *Final Report. Scientific Publication: Policy on Open Access*. Disponible en:
<http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec_06_en.pdf> [consultado: 10/11/2011]

- European Comision. 2008. *Open Access Pilot in FP7*. Disponible en:

- <http://www.uv.es/operuv/guies/open-access-pilot_en.pdf> [consultado: 15/10/2011]
- Facultad de Informática. (2011). Universidad Politécnica de Madrid. *La Biblioteca Nacional de España enlaza sus datos* 14 de Diciembre. Disponible en: <http://www.tendencias21.net/La-Biblioteca-Nacional-de-Espana-enlaza-sus-datos_a9010.html?com> [consultado: 17/04/2012]
 - Fox, Edward. (2000). “Open archives initiative”. *D-lib magazine*, v.6 n.6 <<http://www.dlib.org/dlib/june00/06inbrief.html#FOX>> [consultado: 15/10/2011]
 - Franklin, Jack. (2003). Open access to scientific and technical information: the state of the art. *Information Services and Use* Vol 23, n 2,3 pp.67-86
 - Gómez Dueñas, Laureano Felipe. (2005) La iniciativa de Archivos Abiertos (OAI) un nuevo paradigma en la comunicación científica y el intercambio de información. *Revista Códice*, vol. 2, n. 1, pp. 21-48.
 - Gómez, Nancy Diana; y Arias, Olga Margarita. (2002). El cambio de paradigma en la comunicación científica. *Información, cultura y sociedad* n6. Disponible en: <<http://eprints.rclis.org/archive/00003649>>
 - Haslhofer, Bernhard; Schandl, Bernhard. (2009). Interweaving OAI-PMH Data Sources with the Linked Data Cloud. *Int. J. Metadata, Semantics and Ontologies*, Vol. 5(1) pp.17-31
 - Heath, Tom; Bizer, Christian. (2011). *Linked data: evolving the Web into a*

- global data space*. (1st edition). Synthesis lectures on the semantic web: theory and technology, February, Vol. 1, n°. 1 pp. 1-136. Morgan & Claypool <<http://linkeddatabook.com/editions/1.0/>> [consultado: 14/07/2012]
- Instituto Nacional de Estadística. <<http://www.ine.es/>>
 - Isele, Robert; Jentzsch, Anja; Bizer, Chris; Volz, Julius. (2011) Silk a link discovery framework for the web of data <<http://www4.wiwiss.fu-berlin.de/bizer/silk/>> [consultado: 18/04/2012]
 - Juárez Santamaría, Beatriz (2007). Uso de los metadatos en el orden documental. *Información, producción y servicios*. Vol. 10, n 42 pp.8
 - Kaplan, Deborah; Saunes, Anne; Wilczek, Eliot. (2011) Archival description in OAI-ORE. *Journal of Digital Information* Vol 12, n 2 <<http://journals.tdl.org/jodi/article/view/1814/1769>>
 - Luján Gurmedi, María de. Piñeiro, Isabel. Marmonti, Emiliano. Jornada Virtual del Acceso Abierto Argentina 2011 Disponible en: <http://accesoabiertoargentina.caicyt.gov.ar/files/biblioteca/Ponencia_Piñeiro-Marmonti-Gurmendi.pdf> [consultado: 11/06/2012]
 - Library of Congress. (2010). MADS: Metadata Authority Description Schema <<http://www.loc.gov/standards/mads/mads-doc.html>> [consultado: 11/06/2011]
 - Library of Congress. (2011). METS: Metadata Encoding and Transmission Standard. Introducción y tutorial. <http://www.loc.gov/standards/mets/METSOverview_spa.html> [consultado: 18/01/2012]

- Library of Congress. (2012). MODS: Metadata Object Description Schema. <<http://www.loc.gov/standards/mods/>> [consultado: 11/06/2012]
- Margaix Arnal, Dídac (2010). “Redes sociales y web 2.0 en 2009” *Anuario Think Epi*, v4, pp. 211-216
- Martínez Equihua, Saúl. (2007) Biblioteca Digital Conceptos, recursos y estándares Buenos Aires: Alfagrama. 170 p. (Biblioteca Alfagrama / dirigida por Pablo A. Somma).
- Melero, Remedios; Abad, Francisca. (2008). Revistas open Access: características, modelos económicos y tendencias. *BiD: textos universitaris de Biblioteconomia i Documentació*, juny, n. 20. Disponible en: <<http://www.ub.edu/bid/20meler2.htm>> [consultado: 22/07/2012]
- Nelson, Theodor. (1993). *Literary Machines*. Sausalito: Mindful Press,
- O'Reilly, Tim. (2005). What is web 2.0 Disponible en <<http://oreilly.com/web2/archive/what-is-web-20.html>>
- O'Reilly, Tim; Battelle, John. (2009). Web Squared: web 2.0 five years on. *Web 2.0 Summit* Disponible en <<http://www.web2summit.com/web2009/public/schedule/detail/10194>>
- Open Archives Initiative. Object Reuse and Exchange. Disponible en: <<http://www.openarchives.org/ore/>> [consultado: 15/01/2012]
- OpenDOAR. <<http://www.opendoar.org/find.php>> [consultado: 12/09/2012]
- Orduna-Malea, Enrique. (2009). Reutilización e intercambio de objetos digitales compuestos en la Web: el proyecto OAI-ORE. *Anuario*

ThinkEPI, EPI SCP, pp.45-48

- Pastor Sánchez, Juan Antonio. (2009). *Diseño de un sistema colaborativo para la creación y gestión de tesauros en Internet basado en SKOS*. Murcia: Universidad de Murcia.
- Pastor Sánchez, Juan Antonio. (2011) *Tecnologías de la web semántica*. Barcelona: Editorial UOC.
- Pastor Sánchez, Juan Antonio; Martínez Méndez, Francisco Javier. (2010). Manual de skos (simple knowledge organization system, sistema para la organización del conocimiento simple) *Anales de Documentación*, vol. 13, pp. 285-320.
- Peis Redondo, Eduardo; Herrera Viedma, Enrique; Hassan Montero, Yusef y Herrera, Juan Carlos. (2003). Ontologías, Metadatos y Agentes: Recuperación ‘Semántica’ de la Información. *Actas de las II Jornadas de Tratamiento y Recuperación de la Información (JOTRI)*, pp. 157-165.
- Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma. (2011). Open Data y Linked Open Data: su impacto en el área de bibliotecas y documentación. *El profesional de la Información*, marzo-abril, v.20, n.2, pp.165-173
- Registry of Open Access Repositories. < <http://roar.eprints.org/>> [consultado: 12/09/2012]
- Rosling, Hans. (2006) shows the best stats you’ve ever seen. TED, <http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html>
- Saorín Pérez, Tomás. (2012). Cómo *linked open data* impactará en las

bibliotecas a través de la innovación abierta. *Anuario ThinkEPI*, v. 6 pp. 288-292

- Silió, Teresa. (2005) Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archives System. *El profesional de la información*, v.14, n. 5 pp. 365-380
<<http://www.elprofesionaldelainformacion.com/contenidos/2005/septiembre/8.pdf>> [consultado: 10/11/2011]
- Suber, Peter. (2007). Budapest Open Access Initiative: Frequently Asked Questions. <<http://www.earlham.edu/~peters/fos/boaifaq.htm>> [consultado: 14/11/2011]
- Suber, Peter. (2004). A very brief Introduction to Open Access. <<http://www.earlham.edu/~peters/fos/brief.htm>> [consultado: 14/11/2011]
- Vianello Osti, Marina. (2004). *El hipertexto entre la utopía y la aplicación: identidad, problemática y tendencias de la Web*. Gijón: TREA.
- Volz, Julius. Bizer, Christian. Gaedke, Martin. Kobilarov, Georgi. (2009). Discovering and Maintaining Links on the Web of Data. *International Semantic Web Conference (ISWC2009)*, Westfields, USA, October. <<http://www.wiwiss.fu-berlin.de/en/institute/pwo/bizer/research/publications/VolzBizerGaedkeKobilarov-ISWC2009-Silk.pdf>> [consultado: 14/11/2011]
- Volz, Julius. Bizer, Christian. Gaedke, Martin. Kobilarov, Georgi. (2009). Silk – A Link Discovery Framework for the Web of Data. *2nd Workshop about Linked Data on the Web (LDOW2009)*, Madrid, Spain, April. <http://events.linkedata.org/ldow2009/papers/ldow2009_paper13.pdf> [consultado: 14/11/2011]

- Whitfield, John. (2011). Open access come of age. *Nature*, June v.474, n. 428, <<http://www.nature.com/news/2011/110621/full/474428a.html>> [consultado: 20/08/2012]
- World Wide Web Consortium. (2004). OWL Guide. Recommendation 10 February <<http://www.w3.org/TR/2004/REC-owl-guide-20040210/#BasicDefinitions>> [consultado: 28/01/2012]
- World Wide Web Consortium. (2004). OWL Reference. Recommendation 10 February. <<http://www.w3.org/TR/owl-ref/>>, [consultado: 07/02/2012]
- World Wide Web Consortium. (2009). OWL2 *Web Ontology Language: New Features and Rationale* <<http://www.w3.org/TR/2009/REC-owl2-new-features-20091027/>>. [consultado: 20/02/2012]
- World Wide Web Consortium. (2010). Resource Description Framework RDF: concepts and abstract syntax. W3C Recommendation 10 February. <<http://www.w3.org/TR/rdf-concepts/>> [consultado: 10/01/2012]
- World Wide Web Consortium. (2009). *SKOS Guide Reference* <<http://www.w3.org/TR/2009/REC-skos-reference-20090818/>>, [consultado: 12/03/2012]
- World Wide Web Consortium. (2009). *SKOS Primer* Manual de Introducción a SKOS <<http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>> [consultado: 22/03/2012]
- World Wide Web Consortium. (2008). SPARQL Query Language 1.0 for RDF. W3C Recommendation 15 January. <<http://www.w3.org/TR/rdf-sparql-query/>> [consultado: 12/03/2012]
- World Wide Web Consortium. (2012). SPARQL Query Language 1.1 for

RDF W3C Recommendation January 2012

<<http://www.w3.org/TR/sparql11-protocol/>> [consultado: 22/5/2012]

ÍNDICE DE IMÁGENES

Ilustración 1: Gráfico sobre los tipos de repositorios de acceso abierto	27
Ilustración 2: Relación entre identificador, recurso y representación.....	29
Ilustración 3: Representación gráfica de un triplete y representación gráfica de un enlace hipertextual.	41
Ilustración 4: Representación gráfica de las relaciones entre datos de un documento HTML (grafo dirigido), y representación gráfica de las relaciones entre datos abiertos enlazados (grafo).....	42
Ilustración 5: Representación gráfica de una tripleta.	43
Ilustración 6: Fragmento del diagrama de la nube LOD.....	67
Ilustración 7: Proceso y flujo de trabajo seguido por la herramienta.....	71