



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília



**CULTURA
ACADÊMICA**
Editora

La Evolución de la Complejidad Tecnológica de la Web Semántica

Juan-Antonio Pastor-Sánchez

Como citar: PASTOR-SÁNCHEZ, J. A. La Evolución de la Complejidad Tecnológica de la Web Semántica. *In:* JORENTE, M. J. V.; PADRÓN, D. I. (org.). **Una Mirada a la ciência de la información desde los nuevos contextos paradigmáticos de la posmodernidad**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2017. p. 197-224.
DOI: <https://doi.org/10.36311/2017.78-85-7983-904-7.p197-224>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

CAPÍTULO 8

LA EVOLUCIÓN DE LA COMPLEJIDAD TECNOLÓGICA DE LA WEB SEMÁNTICA

Juan-Antonio Pastor-Sánchez

1 INTRODUCCIÓN

La Web actual es muy diferente de aquella diseñada por Tim Berners-Lee que fue presentada al mundo en 1992. Desde entonces hasta hoy día, este servicio de Internet se ha convertido en una plataforma cliente-servidor universal sobre el que se desarrollan todo tipo de aplicaciones e intercambian contenidos muy diversos. En la actualidad la Web se ha convertido en una realidad incardinada en nuestro devenir diario. La mayor parte de los usuarios son poco conscientes que múltiples tipos de servicios y contenidos utilizan la Web como núcleo tecnológico de desarrollo (GUINARD; TRIFA; WILDE, 2010). Redes sociales, medios de comunicación digitales, contenidos multimedia, comercio electrónico, gestión financiera o herramientas de gestión corporativas son algunos de los ámbitos en los que la Web constituye la plataforma de funcionamiento.

Por lo tanto, la Web es algo más que un entorno en el que los usuarios interactúan con contenidos e información. La propia arquitectura de la Web permite su ampliación mediante nuevas tecnologías que hacen

posible crear nuevas aplicaciones que van más allá de la gestión o generación de documentos HTML¹³². Los desarrolladores de contenidos y servicios vieron el potencial de la web como plataforma para la publicación y el intercambio de información estructurada.

Una primera reflexión nos podría acercar a la idea que Internet en general y la Web en particular han cambiado para siempre el concepto de documento. De hecho, desde mediados de los años 90 del siglo pasado la mayoría de los documentos nacían digitalmente en entornos ofimáticos. A partir de este momento, con el auge de Internet comenzaron a manejarse de forma masiva los documentos digitales, fundamentalmente a través del correo electrónico y la publicación distribuida en múltiples formatos y niveles de agregación realizada mediante la Web (JAEGER; BERTOT, 2010). La auténtica revolución de la Web no resulta tan visible para el gran público. El documento digital, en diferentes formatos, como paradigma de la Web, es ampliamente superado en volumen y heterogeneidad por los datos estructurados publicados en este sistema (HEATH; BIZER, 2011). El desarrollo de tecnologías y su aplicación para que grandes cantidades de datos estructurados estén disponibles a través de la Web se agrupan bajo el nombre de Web Semantica (BERNERS-LEE; HENDLER; LASSILA, 2001). Denominaciones como *Linked Open Data*, *Hyperdata* o *Data on the Web* se refieren a distintos “sabores” de una misma realidad: la publicación de grandes volúmenes de datos altamente estructurados en la Web.

Se trata de un fenómeno que tiene profundas implicaciones que afectan tanto a los usuarios, como a las instituciones que tienen en la Web un soporte para sus servicios. Las bibliotecas o archivos, por poner un ejemplo, comienzan a plantearse publicar sus catálogos, autoridades y materias en forma de conjuntos de datos abiertos, disponibles para su consulta, descarga y reutilización (HASLHOFER; ISAAC, 2011). Entran en juego aspectos técnicos, pero también otros relacionados con la descripción y la representación de los recursos informativos, el acceso, la calidad, preservación y autenticidad de los datos disponibles o las licencias de uso entre otros.

¹³² *HyperText Markup Language*. Es un lenguaje de marcado usado para la estructuración y el formato de páginas web.

Por otro lado, la Web requiere herramientas eficientes de recuperación de información. No solamente existen más de 900 millones de sitios web en todo el planeta¹³³, sino que el número de páginas en cada uno de ellos puede ser de varios cientos e incluso miles. A lo anterior cabe añadir que un conjunto de datos puede tener un número indeterminado de elementos. Por poner un ejemplo Datahub¹³⁴ recoge más de 10.000 conjuntos de datos, entre los cuales DBpedia¹³⁵ ofrece 1.200 millones de tripletas o más “modestamente” GeoNames¹³⁶ contiene casi nueve millones de nombres geográficos. En consecuencia, tal y como afirma Spivack (2007) con la aparición de los contenidos generados en la web 2.0 se produjo un cambio en el funcionamiento de los motores de búsqueda que también se está produciendo con la incorporación de grandes volúmenes de datos accesibles directamente a través de la web. Con la Web 2.0 se incorporaron entradas y comentarios de blogs, perfiles de redes sociales o foros de debate, siendo preciso adaptar la dinámica de los motores de búsqueda a estos contenidos. La existencia de datos estructurados con un modelo de descripción semántica no solamente demanda un nuevo enfoque en la recuperación de información en la Web, sino que perfila nuevas posibilidades para que dichos datos puedan hacer viable el desarrollo de servicios de valor añadido partiendo de procesos de inferencia o el descubrimiento de información.

La Web conforma un ecosistema de gran complejidad y sujeto a continuos cambios. El W3C¹³⁷ siendo consciente de ello, no solo centra su trabajo en el desarrollo de estándares y tecnologías. También pretende ser un centro aglutinador, en el que usuarios, desarrolladores, investigadores, empresas y fabricantes participen e intercambien la experiencia resultante de aplicar y desplegar dichas tecnologías. Este enfoque colaborativo provoca que haya cambiado en cierto modo el modelo de gestión de los conjuntos de datos: frente a la centralización y los procesos de recolección/agregación de datos u objetos digitales se plantea la gestión distribuida de los mismos

¹³³ Según datos de Internet Live Stats: <<http://www.internetlivestats.com/total-number-of-websites/>>.

¹³⁴ Disponible en: <<http://datahub.io>>. Acceso en: 10 dez. 2015.

¹³⁵ Disponible en: <<http://wiki.dbpedia.org/>>. Acceso en: 10 dez. 2015.

¹³⁶ Disponible en: <<http://www.geonames.org/>>. Acceso en: 10 dez. 2015.

¹³⁷ *World Wide Web Consortium*. Consorcio internacional cuya misión es la creación de recomendaciones y estándares Web.

y la identificación, localización, acceso y uso a través de tecnologías abiertas interoperables. Sin duda dicho enfoque coincide en gran medida con los principios del paradigma postcustodial (UPWARD, 1996), motivo por el que las tecnologías y prácticas de la Web Semántica podrían tener una aplicación directa en la práctica actual del Record Management o la Gestión Archivística (WU; HEOK; TAMSIR, 2006). Comprender la Web Semántica supone en gran medida comprender una nueva realidad en la gestión y difusión de datos en cualquier contexto, puesto que la Web se ha convertido en soporte “de facto” en donde la publicación y gestión descentralizada de grandes conjuntos de datos supone una alternativa más eficaz a los procesos de recolección y agregación para la creación de grandes bancos de datos centralizados.

En este capítulo se expondrán de forma comprensiva los principales elementos y aproximaciones de aplicación de la Web Semántica. En primer lugar se contextualiza algunos aspectos conceptuales como son la interoperabilidad, la reutilización, los metadatos y las ontologías. A continuación se realiza una breve descripción de las tecnologías utilizadas para dar soporte de dichos aspectos. Más adelante se contrapone la agregación de datos para generar repositorios centralizados frente al enfoque de Linked Open Data en el que la gestión de los datos se realiza de forma descentralizada al tiempo que se establecen conexiones entre los mismos. También se analizan los trabajos relacionados con las buenas prácticas de publicación de datos en la web y se concluyen con una serie de consideraciones finales a modo de reflexión.

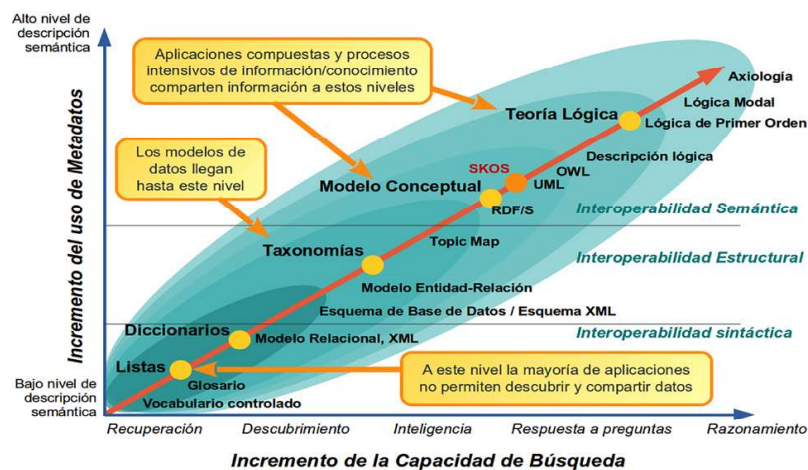
2 INTEROPERABILIDAD, REUTILIZACIÓN, METADATOS Y ONTOLOGÍAS

La interoperabilidad es la capacidad de un sistema de información y los procesos que desarrolla para compartir datos para que sea posible el intercambio de información y conocimiento de forma directa con otros sistemas existentes o futuros y sin restricción de acceso o implementación (SHETH, 1999). Se trata de un concepto que supera los procedimientos de importación y exportación de datos que precisan de procesos de transformación y definición de equivalencias entre estructuras diferentes.

Esto permite la compatibilidad de sistemas al compartir directamente la información que gestionan. En igual medida la interoperabilidad garantiza que en un futuro los sistemas de información no encuentren problemas para incorporar datos gestionados por sistemas desarrollados previamente cuyas funciones quedan obsoletas con el paso del tiempo. En esta perspectiva diacrónica, la interoperabilidad también es uno de los componentes de la preservación digital (DAY, 2003).

Por otro lado, tal y como puede apreciarse en la figura 1 (DAVIS, 2006), existe una relación entre diferentes grados de interoperabilidad la capacidad de uso de metadatos y el incremento de la capacidad de búsqueda de los sistemas de información.

Figura 1 - La interoperabilidad y la relación entre el incremento de la capacidad de búsqueda y el uso de metadatos.



Fuente: DAVIS, 2006, traducido y adaptado.

La interoperabilidad enlaza directamente con el concepto de reutilización. Aquella se centra en el modo en el que los sistemas están diseñados para la implementación de modelos de descripción, gestionar estructuras los datos que los concretan y aplicar formatos. Por su parte, la reutilización está asociada a los procedimientos de acceso, calidad y utilidad de los datos, licencias de uso y transformación y la capacidad para generar nuevos datos, contenidos, conocimiento, productos y servicios.

En consecuencia estamos ante una perspectiva en el que los datos, con sus correspondientes licencias, constituyen un recurso esencial para el desarrollo de nuevos productos y servicios en la sociedad de la información. Son la clave para establecer procedimientos para compartir datos, intercambiar información y conocimiento. Es evidente que se precisa una coordinación e interrelación de los agentes clave: administración pública, empresas, investigadores, educadores y ciudadanos. En los lugares comunes en los que estos agentes generan sinergias de reutilización mediante técnicas de interoperabilidad con estándares abiertos se generan nuevas actuaciones y posibilidades de negocio (RAMOS SIMÓN et al., 2012). A todo esto subyace la voluntad política de adoptar y ejercitar el principio de neutralidad tecnológica que desarrollen y fomenten el uso de dichos estándares.

Para que la interoperabilidad y la reutilización puedan desarrollarse en la práctica es preciso adoptar una serie estándares tecnológicos. Pero también es necesario comprender que los datos estructurados siempre requieren de modelos conceptuales que permitan representar abstracciones de ideas u objetos del mundo real. Esta es la base de los modelos descriptivos que caracterizan los objetos de información según determinados atributos y definen las relaciones lógicas entre objetos (SOLODOVNIK, 2011).

En la Web Semántica los planteamientos anteriores se llevan a la práctica mediante la descripción de recursos de información aplicando vocabularios que implementan esquemas de metadatos y ontologías.

Los metadatos son elementos que describen un determinado objeto siguiendo algún modelo o conjunto de reglas. La clasificación de metadatos de Gilliland-Swetland (2003) mostrada en la siguiente tabla, permite su agrupación en función de aplicación.

Tabla 1 - Ámbitos de aplicación de los metadatos

| <i>Tipo</i> | <i>Definición / Aplicación</i> | <i>Ejemplos</i> |
|---------------------------|------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Metadatos Administrativos | Usados en la gestión y administración de recursos de información | Adquisición de Información Derechos y reproducción Requerimientos legales para el acceso Localización de información Criterios de selección para la digitalización Control de versiones |

| | | |
|---------------------------|------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Metadatos descriptivos | Utilizados para representar recursos de información | Registros de catalogación Proporcionar ayuda durante la búsqueda Índices especializados Describir relaciones entre recursos usando enlaces explícitos Anotaciones de los usuarios |
| Metadatos de preservación | Permiten salvaguardar los recursos de información | Informar sobre las condiciones de uso de los recursos (físicos y digitales). Informar sobre las acciones llevadas a cabo para preservar los recursos (físicos y digitales). |
| Metadatos técnicos | Relativos a cómo funcionan los sistemas o el comportamiento de los metadatos | Documentación de hardware y software Digitalización de información (formato, compresión) Autenticación y seguridad (encriptación, claves, etc.) Control de tiempo de respuesta de sistemas |
| Metadatos de uso | Relativos al nivel y tipo de uso que se hace de los recursos | Información sobre versiones Reutilización del contenido del recurso |

Fuente. Gilliland-Swetland.

Así pues, los metadatos son descripciones estructuradas y codificadas que describen características y propiedades de objetos y recursos para facilitar su localización, recuperación, valoración, administración, persistencia e interoperabilidad. No solo son capaces de realizar una descripción de dichos recursos, sino que también pueden hacer lo propio con los procesos en los que intervienen, sus componentes, las restricciones y las relaciones que se establecen entre ellos. La dispersión de los recursos, así como el gran volumen de información contenida en ellos y la diversidad en cuanto a su actualización y estructura son algunos de los factores que han hecho que los modelos tradicionales de descripción bibliográfica se hayan demostrado ineficaces al aplicarse a los recursos disponibles en la Web. En el caso de la Web Semántica, los metadatos son pares de atributo-valor que describen recursos de información digital identificables en Internet. Los elementos descriptivos conforman un vocabulario que se especifica en un esquema de metadatos. La representación del vocabulario y de los metadatos se realiza, como veremos más adelante, siguiendo el modelo de datos RDF. En definitiva, con los metadatos se puede convertir la gran esfera de datos en la que se ha convertido la Web en un núcleo

de información estructurada y descrita de forma precisa que pueda ser localizada y reutilizada de un modo más eficaz.

Sin embargo, Los metadatos únicamente permiten la descripción de los recursos web en forma de propiedades y los consiguientes valores que toman para cada recurso. Para complementar los metadatos de forma que sea posible realizar representaciones lógicas de las relaciones lógicas entre objetos es preciso utilizar ontologías.

Dicho de forma sencilla: en el contexto de la Web Semántica, las ontologías permiten expresar tipos de objetos y el modo en el que se relacionan entre sí. Una ontología web realiza descripciones de objetos, basadas en los principios de interoperabilidad semántica, mediante la definición de clases, propiedades, relaciones y axiomas. Esto permite diseñar estructuras muy elaboradas de información que además de facilitar el intercambio y explotación de datos posibilita la ejecución de inferencias para descubrir datos no declarados explícitamente.

Para Swartout et al. (1996, p. 138) “una ontología es un conjunto de términos estructurados jerárquicamente para describir un dominio que puede ser utilizado como un marco inicial de una base de conocimientos”. Podría afirmarse que una ontología es una abstracción, más o menos esquemática, de algún objeto o hecho del mundo real. Por lo tanto podría afirmarse que una ontología es un modelo abstracto de una realidad concreta que se obtiene tras haber identificado y definido los conceptos relevantes de la misma de forma explícita.

Las ontologías proporcionan un modelo lógico-conceptual común en un área de conocimiento, definiendo, a diferentes niveles de formalización, el significado de las clases de objetos, las relaciones entre ellos y las propiedades que los caracterizan. Cualquier formalismo usado para materializar ontologías debe contener elementos que permitan representar los conceptos, atributos y sus relaciones, utilizando siempre un conjunto básico de axiomas que establecen los parámetros y reglas de representación. El conocimiento en ontologías se formaliza principalmente usando seis tipos de componentes: clases, atributos, relaciones, funciones, axiomas e instancias.

Las ontologías deben planificarse en un proceso que contemple dos perspectivas:

- La Semántica de la ontología, que formaliza, expresa y estructura el contenido con niveles de detalle flexibles para una descripción de objetos y hechos del mundo real, tanto en sus propiedades como en sus relaciones.
- La Pragmática de la ontología: En donde se definen las posibles aplicaciones de una ontología con métodos de diseño y mantenimiento adecuados y teniendo en cuenta en qué contextos y ámbitos se va a utilizar.

En la Web Semántica se hace un uso intensivo de las ontologías y algunas de ellas, como SKOS o la ontología Dbpedia, resultan esenciales para su comprensión. Además, las ontologías también están sujetas a los principios de interoperabilidad y reutilización afectando a los aspectos lógicos de la descripción de recursos y a las interrelaciones entre ellos.

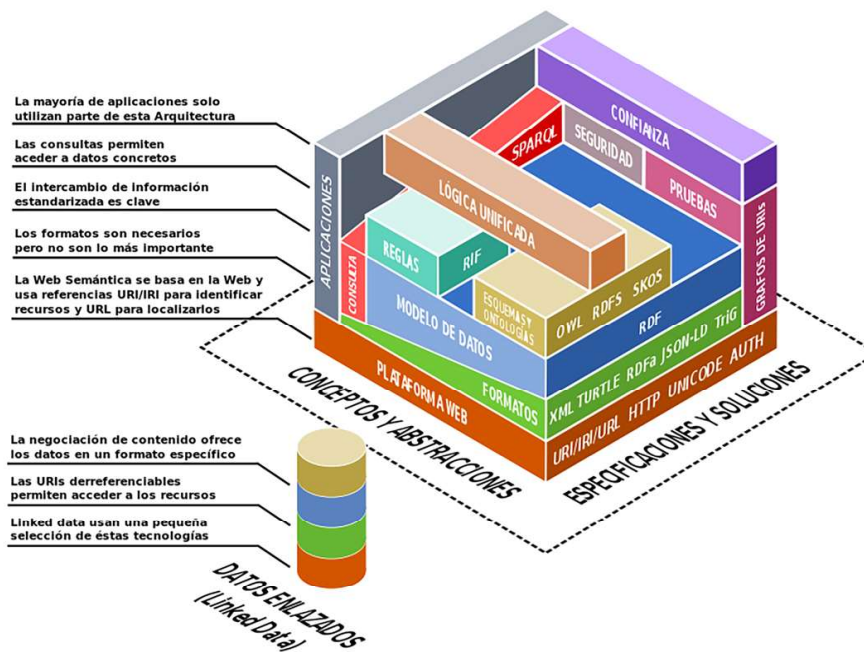
3 LAS TECNOLOGÍAS

Los aspectos conceptuales como la interoperabilidad, reutilización, metadatos y ontologías precisan de tecnologías concretas que las lleven a la práctica. En este sentido, la evolución de la Web Semántica se plantea desde una arquitectura en la que se han desarrollado diversas capas tendentes a dotar a los datos de mayor expresividad semántica. En todo momento se ha tenido en cuenta la interoperabilidad de los datos desde tres puntos de vista:

- Sintáctico: el formato utilizado para la descripción de los recursos.
- Estructural: el modo en el que se puede realizar la descripción de los recursos, estableciendo los mecanismos para definir los atributos que los pueden caracterizar y los posibles valores que pueden tomar.
- Semántico: el significado de los atributos que describen los recursos, así como de las relaciones y restricciones lógicas en los valores de los atributos o las relaciones entre los propios recursos.

La siguiente figura muestra la estructuración de las diferentes tecnologías de la Web Semántica.

Figura 2 - Arquitectura de la Web Semántica.



Fonte: Traducido y adaptado de: <<http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake>>

La Plataforma de la Web convencional continúa siendo la base de la Web Semántica. Se utiliza el protocolo HTTP¹³⁸ para la petición y envío de los datos y se asignan identificadores URI¹³⁹ a los recursos, cuyo alcance de aplicación es mayor que la localización mediante URL¹⁴⁰.

Los lenguajes de marcado y los formatos de datos son uno de los ejes en torno a los cuales ha girado el ciclo evolutivo de la Web. En un principio el lenguaje HTML era adecuado para cubrir las necesidades básicas de estructuración de documentos cuyo uso estaba pensado para el usuario final. Cuando se pensó en la Web como un medio para la publicación y el intercambio de datos fue preciso desarrollar otro

¹³⁸ *HyperText Transfer Protocol*. Es el protocolo de comunicaciones que permite la transferencia de información a través de la Web.

¹³⁹ *Uniform Resource Identifier*. Es una cadena de caracteres que identifica un recurso disponible en un entorno de red, generalmente a través de la Web.

¹⁴⁰ *Uniform Resource Locator*. Es un tipo de URI que especifica la localización de un recurso en la red para poder acceder al mismo.

lenguaje de marcado que permitiera representar información altamente estructurada de un modo muy flexible. Tal es la finalidad de XML.

Desde su creación en 1998, el lenguaje XML¹⁴¹ supuso un punto de partida para nuevos usos y enfoques de la Web. Con XML es posible definir la sintaxis de lenguajes de marcado para propósitos específicos, puesto que se trata de un metalenguaje para definir vocabularios orientados al intercambio de datos correctamente estructurados, con independencia de la plataforma en la que se ejecuten las aplicaciones en los que se procesan dichos datos.

Por este motivo, la familia de tecnologías XML es muy amplia y diversa para cubrir diversos aspectos y funciones. Con XML Schema se definen los elementos estructurales del vocabulario o vocabularios XML utilizados en un documento. Mediante XML Namespace es posible identificar de manera inequívoca a diferentes esquemas y vocabularios (W3C, 2009). Por ejemplo: Dublin Core dispone de un término “publisher” al igual que EAD¹⁴². Para evitar que se produzcan ambigüedades en el uso de los elementos de idéntico nombre se utilizan los espacios de nombres XML. De esta forma se agrupan los elementos de un esquema en torno a una URI a la cual se le asigna un prefijo y que a su vez se utiliza para diferenciar a los elementos:

ead = <http://www.loc.gov/ead/ead.xsd> → ead:publisher

dcterms = <http://purl.org/dc/terms/> → dcterms:publisher

Por su parte XSL¹⁴³ permite definir estilos visuales y establecer transformaciones de partes de un documento XML seleccionadas mediante XPATH y XQUERY.

XML ofrece una serie de soluciones para la interoperabilidad de la información a nivel sintáctico. De este modo se garantiza una serie de reglas muy definidas que permiten desarrollar mecanismos comunes

¹⁴¹ *eXtensible Markup Language*.

¹⁴² *Encoded Archival Description*. Es un modelo de codificación para la descripción de documentos de archivos basado en el lenguaje XML.

¹⁴³ *eXtensible Stylesheet Language*. Es una familia de lenguajes que permite transformar y formatear documentos XML para su visualización en un medio concreto.

para identificar y estructurar estructuras de pares atributos-valor. El uso conjunto de diferentes tipos vocabularios o lenguajes de marcado definidos con XML permite describir y representar diferentes tipos de objetos, mediante especificaciones normalizadas. En consecuencia XML supone un salto cualitativo pues permite separar estructura, contenido y formato visual.

En el ámbito de las bibliotecas y archivos XML ha encontrado una aplicación directa. En este campo son muchas las iniciativas, como MARC-XML¹⁴⁴, EAD, ISAAD(G)¹⁴⁵, DACS¹⁴⁶ o EAC-CPF¹⁴⁷ entre otros, que se disponen de su correspondiente esquema XML para generar documentos válidos con respecto a estos estándares.

Sin embargo, aunque XML constituye una pieza esencial en la interoperabilidad, únicamente ofrece una solución desde el punto de vista sintáctico. A lo sumo, mediante un esquema XML se consigue un nivel básico de interoperabilidad estructural. Para alcanzar los objetivos de la Web Semántica se precisan herramientas que ofrezcan mecanismos estándares para definir de un modo preciso los aspectos conceptuales y lógicos de los modelos de descripción. El uso de esquemas XML hace necesario conocer la semántica particular de cada esquema y los desarrolladores de software son quienes deciden la interpretación de dicha semántica. En consecuencia se precisa un modelo de datos adecuado que defina tanto los aspectos estáticos (representación) como los dinámicos (procesos). (BERNERS-LEE, 1998).

RDF¹⁴⁸ es un modelo de datos que aporta un marco semántico para la descripción de recursos. En RDF los elementos a describir se definen como recursos. Este modelo representa las descripciones de

¹⁴⁴ Esquema que permite representar los registros bibliográficos MARC en forma de documentos XML. Más información en: <<http://www.loc.gov/standards/marcxml/>>.

¹⁴⁵ *General International Standard Archival Description*. Es la norma internacional de descripción archivística publicada por el Consejo Internacional de Archivos. Más información en: <<http://www.ica.org/download.php?id=1745>>.

¹⁴⁶ *Describing Archives: A Content Standard*. Norma de Descripción Archivística de EE.UU desarrollada y publicada por la Sociedad de Archiveros Americanos. Más información en: <http://www2.archivists.org/standards/DACS>.

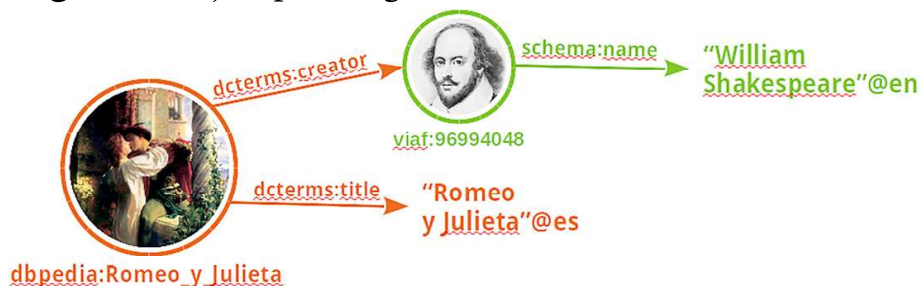
¹⁴⁷ *Encoded Archival Context - Corporate bodies, Persons and Families*. Estándar de codificación basado en XML para representar la información sobre diversos tipos de autores de material archivístico y adoptado por la Sociedad de Archiveros Americanos. Más información en: <<http://eac.staatsbibliothek-berlin.de/>>.

¹⁴⁸ *Resource Description Framework*.

recursos mediante sentencias que adoptan la forma de tripletas (*triplets*) compuestas por un sujeto, un predicado y un objeto.

Cada recurso se identifica de manera unívoca mediante una URI y se describen mediante la asignación de valores a determinados atributos o con el establecimiento de relaciones entre dos recursos (RDF 1.1 CONCEPTS AND ABSTRACT SYNTAX, 2014)¹⁴⁹. En el primer caso los recursos se vinculan con los valores con propiedades de datos. En el segundo caso los recursos se relacionan entre sí a través de propiedades de objeto. Cada propiedad está asociada con una URI, de manera que al igual que los recursos se identifican de manera unívoca. Por lo tanto, La tripletas de descripciones conforman grafos, en donde los recursos se representan como nodos, lo valores como literales y las propiedades mediante arcos.

Figura 3 - Ejemplo de grafo RDF.



dbpedia: <http://es.dbpedia.org/resource/>
viaf: <https://viaf.org/viaf/>
schema: <http://schema.org/>
dcterms: <http://purl.org/dc/terms/>

Fuente: Elaboración propia.

La figura anterior es un ejemplo de grafo RDF muy sencillo. Tanto los recursos, como las propiedades están definidos dentro de un espacio de nombres XML, de forma que aquellos elementos con idéntico nombre se distinguen entre sí por tener un espacio de nombres diferente. Los espacios de nombres pueden abreviarse utilizando prefijos. Tal y

¹⁴⁹ *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation, 25 February 2014. Disponible em: <<http://www.w3.org/TR/rdf11-concepts/>>. Acceso em: 10 dez. 2015.

como puede observarse en el ejemplo, el recurso identificado como “dbpedia:Romeo_y_Julieta” se corresponde en realidad con el recurso cuya URI es “http://es.dbpedia.org/resource/Romeo_y_Julieta”. Puede verse como el término de Dublin Core `dcterms:creator` se utiliza como una propiedad de objeto para vincular la obra con su autor. Por su parte, `dcterms:title` y `schema:name` se consideran propiedades de datos ya que se utilizan para especificar valores literales.

Los grafos de sentencias RDF tienen un alto grado de expresividad para las personas, pero lo cierto es que para que se puedan intercambiar y procesar por máquinas¹⁵⁰ es preciso utilizar una representación textual. Este proceso de codificación (serialización) precisa aplicación de un determinado formato para expresar las sentencias RDF según una sintaxis normalizada. Existen diversos formatos para la serialización de RDF, tales como RDF/XML¹⁵¹, Turtle¹⁵², JSON-LD¹⁵³ o RDFa¹⁵⁴ entre otros.

Figura 4 - Formatos para la serialización de RDF.

| | | |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| RDF/XML | <pre><rdf:Description rdf:about="http://www.wikidata.org/entity/Q83186"> <dcterms:creator rdf:resource="https://viaf.org/viaf/96994048"/> <dcterms:title>Romeo and Juliet</dcterms:title> </rdf:Description></pre> | Permite aplicar las tecnologías XML (XSLT, XPath, XQuery, etc.) |
| Turtle | <pre><http://www.wikidata.org/entity/Q83186> dcterms:creator <https://viaf.org/viaf/96994048"> ; dcterms:title "Romeo and Juliet"@en .</pre> | Compacto y ligero. Ideal para su uso por máquinas. |
| JSON-LD | <pre>@id:"http://www.wikidata.org/entity/Q83186", "dcterms:creator": "https://viaf.org/viaf/96994048", "dcterms:title": {"@language": "en", "@value": "Romeo and Juliet"</pre> | Diseñado para el consumo de RDF por JavaScript. |
| RDFa | <pre><div resource="http://www.wikidata.org/entity/Q83186"> <h1 property="dcterms:title">Romeo and Juliet</h1> <p>Author: William Shakespeare</p> </div></pre> | Marcado semántico de documentos HTML insertando sentencias RDF. |

Fuente: elaboración propia.

RDF tiene un doble propósito: por un lado describir los recursos y por otro definir los vocabularios a partir de los cuales se realizan tales descripciones. Dichos vocabularios deben incluir categorías o clases para clasificar los recursos. Igualmente la definición de propiedades precisa establecer sobre qué clase de

¹⁵⁰ En este contexto por “máquina” se entiende a programas informáticos capaces de tomar ciertas decisiones de forma autónoma durante el procesamiento de datos en función del contenido de los mismos.

¹⁵¹ Disponible em: <<http://www.w3.org/TR/rdf-syntax-grammar/>>. Acceso em: 10 dez. 2015.

¹⁵² Disponible em: <<http://www.w3.org/TR/turtle/>>. Acceso em: 10 dez. 2015.

¹⁵³ Disponible em: <<http://www.w3.org/TR/json-ld/>>. Acceso em: 10 dez. 2015.

¹⁵⁴ Disponible em: <<https://www.w3.org/TR/rdfa-primer/>>. Acceso em: 10 dez. 2015.

recursos puede aplicarse una determinada propiedad y los posibles valores en el caso de las propiedades de datos o los recursos con los que puede establecerse una relación en las propiedades de objeto.

Para ello RDF se amplía con RDF Schema y con OWL. Mediante RDF Schema¹⁵⁵ no solamente es posible definir clases y propiedades, sino jerarquías de estas derivando subclases y subpropiedades. De este modo es posible definir la clase “Persona” y de ésta la subclase “Artista” y de ésta a su vez la subclase “Escritor”. RDF Schema permite desarrollar ontologías “ligeras” que pueden aplicarse en la descripción de recursos mediante esquemas sencillos. OWL¹⁵⁶ aporta una mayor expresividad semántica, de forma que es posible establecer restricciones de cardinalidad, equivalencia entre clases de distintas ontologías, definir propiedades inversas entre sí, etc.

Como todo modelo de datos que se precie, RDF también dispone de un lenguaje de consulta para recuperar sentencias que respondan a ciertas condiciones de búsqueda. Dicho lenguaje se denomina SPARQL y permite definir patrones de concordancia con sentencias RDF. La especificación también define el formato de los datos de las sentencias recuperadas. Desde la especificación SPARQL 1.1 también es posible utilizar operaciones para el mantenimiento (inserción y borrado) de sentencias dentro de un grafo RDF¹⁵⁷.

En resumidas cuentas podría decirse que SPARQL es para RDF lo que SQL es para el modelo relacional. No obstante esta tecnología supone un enorme avance en el despliegue de la Web Semántica. SPARQL supera el acceso a las sentencias RDF mediante la descarga de conjuntos de datos completos en ficheros de texto mediante alguno de los formatos de serialización. Utilizando SPARQL es posible almacenar las sentencias en bases de datos RDF (*triplestore*) para que mediante un SPARQL Endpoint puedan recuperarse datos concretos sin que sea

¹⁵⁵ *RDF Schema 1.1*. W3C Recommendation 25 February 2014. Disponible em: <<http://www.w3.org/TR/rdf-schema/>>. Acceso em: 10 dez. 2015.

¹⁵⁶ *OWL 2 Web Ontology Language Primer (Second Edition)*. W3C Recommendation 11 December 2012. Disponible em: <<http://www.w3.org/TR/owl-primer/>>. Acceso em: 10 dez. 2015.

¹⁵⁷ *SPARQL 1.1 Query Language*. W3C Recommendation, 21 March 2013. Disponible em: <<http://www.w3.org/TR/sparql11-query/>>. Acceso em: 10 dez. 2015.

necesario procesar todo el conjunto de datos. Con SPARQL se reducen considerablemente las dificultades para desarrollar aplicaciones que sirvan datos semánticos, puesto que ya no resulta imprescindible diseñar APIs personalizadas para interrogar conjuntos de datos.

Los datos publicados mediante ontologías están expresados según un alto nivel de abstracción y representación lógicas. Por este motivo una de las principales aplicaciones derivadas del uso de estos datos es la ejecución de procesos de inferencia. Este tipo de procesos puede permitir descubrir informaciones implícitas a las sentencias RDF de un conjunto de datos. Las inferencias se realizan a partir de reglas. RIF¹⁵⁸ es un lenguaje que permite expresar este tipo de reglas para que puedan intercambiarse.

4 LINKED OPEN DATA

Linked Open Data (Datos Abiertos Enlazados, también conocido por sus siglas en inglés LOD) se refiere a la técnica publicar datos estructurados en la Web. En esencia, el mantenimiento de los conjuntos de datos se realiza de forma descentralizada, por diferentes entidades, y posteriormente puede ser referenciados y reutilizados libremente. En realidad se trata del mismo enfoque que los enlaces que se definen entre documentos HTML, pero en este caso los datos se conectan entre sí gracias a las relaciones que se establecen entre recursos. De este modo se definen interconexiones a través del establecimiento de enlaces entre recursos ubicados en distintos conjuntos de datos.

Estos enlaces pueden ser procesados por máquina, utilizando diversos estándares y formatos abiertos, junto con el protocolo HTTP, para acceder a conjuntos de datos con la finalidad de su reutilización y su análisis para descubrir conexiones con otros conjuntos de datos a partir de los enlaces que se hayan podido definir.

Autores como Berners-Lee (2006) y Haslhofer y Schandl (2010), definen una serie de reglas que deben contemplarse para la publicación LOD:

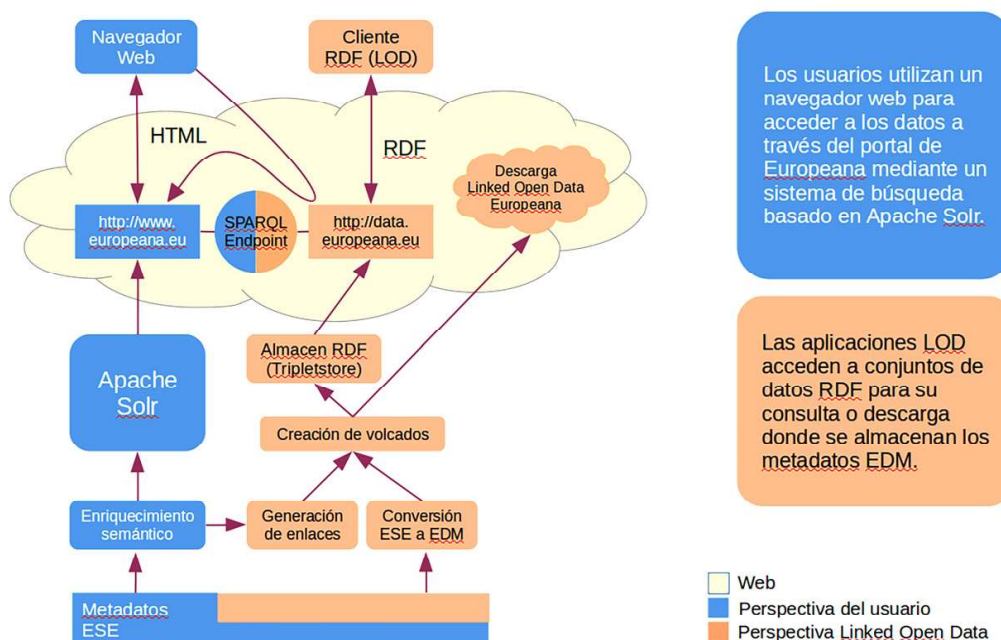
¹⁵⁸ RIF Primer (Second Edition). W3C Working Group Note 5 February 2013. Disponible em: <<http://www.w3.org/TR/rif-primer>>. Acceso em: 10 dez. 2015.

1. Utilizar URIs dereferenciables como mecanismo que permita tanto la identificación como el acceso a objetos o recursos. Esto es de gran importancia para facilitar la definición de enlaces entre diferentes fuentes de datos. Un ejemplo de URI dereferenciable sería una URL que además de identificar inequívocamente un recurso también proporciona una dirección para su recuperación.
2. A partir del uso de URIs dereferenciables es posible utilizar el protocolo HTTP para la transmisión de datos entre cliente y servidor, pudiendo indicar el primero el formato en el que éste último debe suministrar los datos. En unos casos el servidor detectará que la petición se está realizando a través de un navegador web, en cuyo caso la información se suministrará al usuario en formato HTML/XHTML. En otros la petición será realizada por parte de una aplicación informática y se utilizará un formato legible por máquina para suministrar los datos.
3. Incluir enlaces a URIs de recursos externos. De esta forma los procesos de publicación de datos son mucho más eficientes, utilizando fuentes externas, conectándolas con las propias y evitando duplicidades o procesos de replicación. Esto permite descubrir nuevos recursos partiendo de un conjunto de datos diferente.
4. Es recomendable que las instituciones que publiquen LOD, ofrezcan un SPARQL Endpoint para construir consultas que permitan recuperar, de un modo más selectivo, únicamente aquellos datos que se precisen.
5. Desde el punto de vista de la apertura de los datos, es importante que estos dispongan de una licencia libre que no esté sujeta a ningún tipo de copyright y que permita su reutilización flexible, sin contraprestaciones de ningún tipo y no restrictiva.

Podría decirse que Linked Open Data en cierta medida se contraponen a las soluciones basadas en el protocolo OAI-PMH. En este

caso se crean repositorios centralizados cuyo contenido se obtiene a partir de agregadores que “recolectan” datos de fuentes externas e incluso otros repositorios. La aproximación con LOD supone hacer accesible, en un formato estándar (normalmente a través de algún vocabulario RDF), aquellos datos que una determinada organización desea hacer públicos y en los que podemos encontrar vínculos con fuentes de datos externas. Es paradigmático el caso de Europeana que en un principio se basaba únicamente en procesos de agregación OAI-PMH y ha cambiado su enfoque hacia LOD (EUROPEANA, 2011).

Figura 5 - Esquema de generación y uso de datos EDM.



Fuente: <<http://pro.europeana.eu/tech-details>>.

En cuanto al ámbito de bibliotecas, archivos y museos (Libraries, Archives, Museums, LAM) los beneficios son claros tanto para investigadores y usuarios en general, instituciones, así como para profesionales y fabricantes del sector. No obstante, hay que resaltar que LOD/LAM todavía precisa de un desarrollo ya que actualmente está en un estado muy inicial debido a aspectos políticos y operativos más que a problemas técnicos (WORLD WIDE WEB, 2011).

Un indicador de la expansión de esta tecnología puede encontrarse en *Datahub*¹⁵⁹. Se trata de un catálogo colaborativo con información descriptiva de conjuntos de datos disponibles en Internet. *Datahub* contiene información sobre la localización, tipo de contenido, forma de acceso a los datos, volumen o relaciones con conjuntos de datos externos entre otras características. A día de hoy este catálogo referencia más 10000 conjuntos de datos.

5 LAS BUENAS PRÁCTICAS

Pese a todo lo visto hasta ahora, puede afirmarse que, para la publicación de datos estructurados en la Web, no es suficiente con disponer de tecnologías suficientemente desarrolladas desde el punto de vista técnico. La práctica ha demostrado que es necesario aplicar una serie de pautas que sirvan de guía en este tipo de proyectos.

El grupo de trabajo DWBP del W3C de buenas prácticas de datos en la web (Data on Web Best Practices, DWBP) fomenta el desarrollo de un ecosistema de datos abiertos de forma que desarrolladores y editores de los datos se entienda entre sí. Para ello es imprescindible que los editores mejoren la coherencia de los conjuntos de datos. Por este motivo Para ello el grupo de trabajo enfoca sus tareas en dos líneas: la definición de buenas prácticas para aplicar en la implementación tecnológica durante la publicación de los datos y la definición de vocabularios para representar indicadores de calidad y uso de los conjuntos de datos¹⁶⁰.

Resulta esencial entender que la noción de conjunto de datos o dataset ha cambiado en función de la práctica. En efecto, el desarrollo de proyectos aplicados ha hecho posible identificar nuevas necesidades en los procesos de publicación de datos estructurados. Ahora el centro del desarrollo de este tipo de proyectos no son tanto las soluciones tecnológicas de edición y publicación como la reutilización y consumo de dichos datos, su fiabilidad o el tipo de licencia (SAORÍN;

¹⁵⁹ Disponible em: <<http://datahub.io>>. Acesso em: 10 dez. 2015.

¹⁶⁰ *Data on the Web Best Practices*: <https://www.w3.org/2013/dwbp/wiki/Main_Page>.

PESET; FERRER-SAPENA, 2013). En definitiva, el desarrollo y mantenimiento de los conjuntos de datos ha de ser independiente de su posible explotación a posteriori.

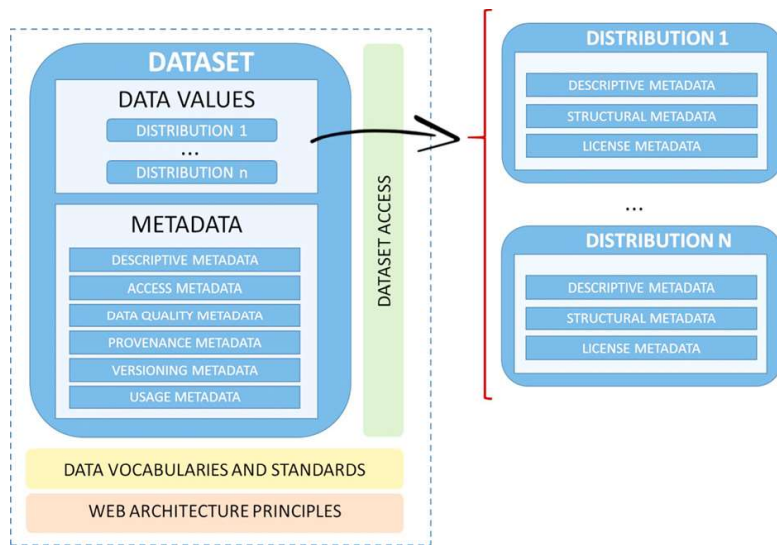
Un conjunto de datos publicado en la web está rodeado de los elementos expuestos en la sección anterior: la propia arquitectura de la web, estándares y recomendaciones, los vocabularios utilizados para representar los datos y los mecanismos para acceder a los mismos.

Además de los datos propiamente dichos (data values), un conjunto de datos debe contener ciertos metadatos con información sobre su contenido, estructura, procedencia, indicadores de calidad o medios para su acceso y uso. Los valores de datos han de organizarse en diferentes distribuciones, que responderían a diferentes necesidades de explotación o consumo.

Evidentemente, los aspectos de modelado o representación son esenciales, pero es igual de importante contemplar a un conjunto de datos en función de su ciclo de vida dentro el contexto Linked Open Data. Por ello, han de considerarse otras operaciones, como la interconexión con otros conjuntos de datos, la aplicación de vocabularios controlados para clasificar los datos, el enriquecimiento de los mismos, el análisis de su calidad. Todo ello al tiempo que se tiene en cuenta explotación mediante mecanismos para buscar, descubrir, extraer y consultar los datos (AUER et al., 2012).

Con el tiempo, la publicación de datos en la Web, ha planteado nuevos retos sobre su privacidad, seguridad, calidad, procedencia y licencias de explotación. De igual manera, disponer de diferentes versiones de los datasets, convenientemente enriquecidos, mantenidos y preservados son tareas incardinadas en el ciclo de vida de los datos.

Figura 6 - Composición de los conjuntos de datos (valores de datos y metadatos) junto con otros componentes relacionados con su uso y publicación.



Fuente: <<https://www.w3.org/TR/dwbp>>.

El grupo de trabajo DWBP ha actualizado las buenas prácticas para la publicación LOD¹⁶¹ gracias a un documento de buenas prácticas que identifica de un modo muy concreto los retos anteriormente esbozados¹⁶²:

- Utilizar URIs persistentes para identificar tanto a las distintas versiones de un dataset como a los datos que contiene.
- Uso de vocabularios de datos para incrementar la interoperabilidad semántica, reutilizando vocabularios y escogiendo niveles adecuados de modelado y formalización.
- Usar estándares para que los datos estén disponibles en diferentes formatos que hagan más flexible su reutilización.
- Enriquecer los datos para aportar valor añadido a través del uso de otros conjuntos de datos, bases de datos u otros recursos.
- Suministrar metadatos del conjunto de datos legibles tanto por máquina como por personas. Estos metadatos, ya sean descriptivos, estructurales o de aspectos más concretos, se deben representar con los términos de vocabularios estándar.

¹⁶¹ *Best Practices for Publishing Linked Data*. W3C Working Group, Note 09 January 2014. Disponible em: <<https://www.w3.org/TR/ld-bp>>. Acceso em: 10 dez. 2015.

¹⁶² *Data on the Web Best Practices*. W3C Working Draft, 12 January 2016. Disponible em: <<http://www.w3.org/TR/dwbp>>. Acceso em: 15 jan. 2016.

- Proporcionar acceso en tiempo real a datos actualizados en la medida de lo posible, no solamente descargando el conjunto de datos en su totalidad en algún formato estándar, sino también a través de servicios web y con APIs adecuadamente documentadas.
- Garantizar la privacidad y la seguridad de los datos sensibles, identificando qué partes de los conjuntos de datos no están disponibles públicamente e indicando sobre el motivo de dicha limitación.
- Informar a los usuarios y consumidores de los datos acerca del modo para contactar con los editores y publicando los resultados de dicha retroalimentación como datos abiertos.
- Incluir una licencia de uso del *dataset* para que los usuarios estén informados sobre las limitaciones y posibilidades de reutilización de los datos.
- Incluir información sobre la procedencia del contenido del conjuntos de datos para que pueda evaluarse con la finalidad de medir su fiabilidad y confianza.
- Gestionar adecuadamente las distintas versiones de un conjunto de datos, evitando modificar las APIs o servicios de acceso a los datos que obliguen a introducir cambios en los clientes que ya hagan uso de los mismos.

Dicho documento establece 32 buenas prácticas clasificadas en función de los beneficios que pueden aportar en los proyectos de publicación de datos en la Web:

- **Comprensión:** las personas pueden tener una mejor comprensión sobre la estructura, naturaleza y significado de los datos, así como del conjunto de datos en su globalidad y los metadatos que lo acompañan.
- **Procesamiento:** las máquinas son capaces de procesar y manipular el contenido del conjunto de datos.
- **Descubrimiento:** las máquinas pueden descubrir de forma automática conjuntos de datos o datos específicos dentro de estos.

- **Reutilización:** se incrementan las posibilidades de reutilización de datos por parte de diferentes grupos de consumidores.
- **Confianza:** mejora la confianza de los consumidores en el conjunto de datos.
- **Conectividad (Linkability):** es posible crear vínculos entre los recursos, tanto a nivel de conjunto de datos como de elementos individuales dentro de estos.
- **Acceso:** tanto las personas como las máquinas pueden acceder a datos actualizados en diferentes formatos y niveles de detalle.
- **Interoperabilidad:** facilita el acuerdo entre editores y consumidores de datos.

Puesto que tanto la calidad como el uso que se haga de los conjuntos de datos actualmente se están desarrollando sendos vocabularios. El primero de ellos amplía la funcionalidad del vocabulario DCAT¹⁶³, para representar determinados aspectos, indicadores y métricas sobre la calidad de los conjuntos de datos¹⁶⁴. El otro vocabulario¹⁶⁵ tiene como objetivo representar los usos que se realiza de un *dataset* por parte de los consumidores de datos.

6 CONSIDERACIONES FINALES

El desarrollo de la Web Semántica puede entenderse como una evolución lógica de la Web convencional. Sin embargo, la publicación de datos resulta evidentemente mucho más compleja que la publicación de documentos, puesto que la interoperabilidad semántica de los datos es mucho más exigente que la interoperabilidad sintáctica que se exige en el caso de los protocolos y los lenguajes de marcado. Por esta razón, no

¹⁶³ Disponible en: <<http://www.w3.org/TR/vocab-dcat/>>. Acceso en: 10 dez. 2015.

¹⁶⁴ *Data on the Web Best Practices: Data Quality Vocabulary*. W3C Working Draft, 17 December 2015. Disponible en: <<https://www.w3.org/TR/vocab-dqv/>>. Acceso en: 15 jan. 2016.

¹⁶⁵ *Data on the Web Best Practices: Dataset Usage Vocabulary*. W3C Working Draft, 28 January 2016. Disponible en: <<http://www.w3.org/TR/vocab-duv/>>. Acceso en: 3 fev. 2016.

hay que confundir la dispersión tecnológica con la propia naturaleza del ecosistema de la Web Semántica. Exigirle características propias de un conjunto compacto e integrado de tecnologías y funcionalidades o un sistema planificado no tiene sentido en una realidad en constante cambio como la Web. Son muchas las propuestas que desaparecen o se funden con otras fruto de la propia naturaleza auto-organizativa de la Web, en especial cuando participan múltiples agentes en procesos colaborativos abiertos proponiendo recomendaciones, trabajando sobre casos de uso y elaborando estándares.

Pero sin lugar a dudas, hablamos de una apuesta por la interoperabilidad y los estándares abiertos en una arquitectura tecnológico-organizativa basada en las sinergias y tensiones existentes en el binomio dispersión-integración de datos en un entorno distribuido. La clave de la reutilización en el panorama actual no es el control centralizado, la normalización extrema y la transformación de los datos en función de su uso, sino la colaboración y la interconexión.

Los Archivos y Bibliotecas, sensibles y conscientes de esta realidad, están abrazando rápidamente los desarrollos de la Web Semántica y las aplicaciones prácticas como Linked Open Data. El incremento en el número y volumen de conjuntos de datos es un reflejo de las nuevas (y no tan nuevas) tendencias en estas instituciones. Las entidades públicas juegan un rol crucial en todo ello a través de la publicación de conjuntos de datos. La misma Sociedad aporta recursos económicos y humanos para la creación y gestión de estos datos y las administraciones públicas están obligadas a su publicación en formatos que garanticen su interoperabilidad para que ciudadanos y organizaciones públicas y privadas incorporen valor añadido definiendo nuevos vínculos, creando nuevas plataformas de consulta y ofreciendo nuevos servicios y productos.

Tampoco olvidemos que se precisan ciertas condiciones que contribuyan a incrementar su despliegue y ofrezcan incentivos para que el sector privado participe en este panorama. Un mercado de datos semánticos resulta atractivo por las nuevas oportunidades que sugiere, pero también conlleva definir un modelo que considere aspectos tales como los derechos y licencias de explotación de los datos, la inclusión de publicidad o el

respecto a la privacidad. Aquí tienen mucho que ver las políticas públicas y corporativas y los modelos de negocio que apoyen y fomenten su uso. El camino es largo y las nuevas tecnologías han de aplicarse teniendo en cuenta que la mejor solución no es la que aparente ser la definitiva, sino la que aparente ser fácilmente ampliable y modificable en el futuro y a ser posible que no vuelvan a pasar por un camino ya recorrido.

En definitiva, estamos ante un conjunto de tecnologías, aplicaciones, servicios y productos emergentes en continuo desarrollo y en consecuencia su maduración ha estado sujeta al cumplimiento de ciertas expectativas. Y estas expectativas, como toda tecnología emergente, han evolucionado en función de desarrollos concretos. La práctica ha arrojado la conclusión que las tecnologías en sí, aunque brillantes, no son el único elemento, ni siquiera el más determinante, para asegurar su éxito y su correspondiente supervivencia. Por este motivo la aplicación de las tecnologías de la Web Semántica trasciende los aspectos meramente técnicos y conocer en profundidad otros relacionados con el ciclo de vida de los datos en la web. Si algo hemos aprendido es que las dinámicas de publicación, acceso y reutilización de los datos son tanto o más importantes. Esto es así, puesto que durante la puesta en práctica de los procesos de publicación de datos se han identificado dificultades concretas, siendo preciso elaborar una serie de pautas (buenas prácticas) que orienten a los actores que participan en la publicación de datos en la Web y mejoren la calidad y reutilización de los conjuntos de datos.

Pero hay que considerar que la propia naturaleza cambiante y heterogénea exigirá una revisión de los trabajos del DWBP, tan útiles hoy, pero que sin duda quedarán desfasados con la identificación de nuevas problemáticas, la aparición de nuevas tecnologías y el desarrollo de aplicaciones prácticas aún por imaginar. No esperemos en ningún momento de la Web Semántica y sus aplicaciones un mundo ordenado e inmutable. Y es que la misma Web, como cualquier otra realidad humana, está inmersa en un mundo cuya realidad social, cultural y tecnológica se encuentra en estado de “beta” continuo. Por mucho trabajo ya realizado, casi todo está por hacer, por pensar, por idear, así es el mundo en el que nos ha tocado vivir en todos sus aspectos.

REFERÊNCIAS

- AUER, S. et al. Managing the Life-Cycle of Linked Data with the LOD2 Stack. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 11., 2012, Boston. **Proceedings...** New York: Springer, 2012. 2. part., p. 1-16. http://dx.doi.org/10.1007/978-3-642-35173-0_1. Disponível em: <<http://iswc2012.semanticweb.org/sites/default/files/76500001.pdf>>. Acesso em: 10 dez. 2015.
- BERNERS-LEE, T. **Why RDF model is different from the XML model**. 1998. Disponível em: <<https://www.w3.org/DesignIssues/RDF-XML>>. Acesso em: 10 dez. 2015.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, Los Angeles, v. 284, n. 5, p. 28-37, 2001.
- BERNERS-LEE, TIM. Linked Data. In: **Design Issues**. W3C, 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 10 dez. 2015.
- DAVIS, M. **Semantic Wave 2006: Executive Guide to the Business Value of Semantic Technologies**. Tradução e adaptação. Semantic Interoperability Community of Practice (SICoP), 2006. 56 p. White Paper Series Module 2. Davis, M. Disponível em: <<http://semanticcommunity.info/@api/deki/files/6142/=sicopsemwave2006v1.0.doc>>. Acesso em: 10 dez. 2015.
- DAY, M. Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 3., 2003, Seattle. **Proceedings...** Maryland: Dublin Core Metadata Initiative, 2003. Disponível em: <http://opus.bath.ac.uk/23599/1/101_paper38.pdf>. Acesso em: 10 dez. 2015.
- EUROPEANA. **Definition of the europeana data model elements**. Version 5.2.2, 2011. Disponível em: <<http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22>>. Acesso em: 10 dez. 2015.
- GILLILAND-SWETLAND, A. Metadata - Where Are We Going? In: GORMAN, G.E. (Ed.). **International yearbook of library and information management 2003-2004: metadata applications and management**. London: Facet Publishing, p. 17-33, 2003.
- GUINARD, D., TRIFA, V., WILDE, E. A resource oriented architecture for the Web of Things. In: INTERNET OF THINGS CONFERENCE, 2., 2010, Tokyo. **Proceedings...** New York: IEEEExplore, 2010. DOI: 10.1109/IOT.2010.5678452. Disponível em: <<https://vs.inf.ethz.ch/publ/papers/dguinard-things-2010.pdf>>. Acesso em: 10 dez. 2015.
- HASLHOFER, B.; ISAAC, A. Data.europeana.eu: The Europeana Linked Open Data Pilot. In: INTERNATIONAL CONFERENCE DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Haia. **Proceedings...** Maryland: Dublin Core Metadata Initiative, 2011. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/view/3625/1851>>. Acesso em: 10 dez. 2015.

HASLHOFER, B.; SCHANDL, B. Interweaving OAI-PMH data sources with the linked data cloud. **International Journal of Metadata, Semantics and Ontologies archive**, Ginebra, v. 5, n. 1, 2010. Disponível em: <http://eprints.cs.univie.ac.at/73/1/ijmso2010_haslhofer_schandl.pdf>. Acesso em: 10 dez. 2015.

HEATH, T.; BIZER, C. **Linked Data**: evolving the web into a global data space. New Jersey: Morgan & Claypool publishers, 2011. DOI: 10.2200/S00334ED1V01Y201102WBE001. Disponível em: <<http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>>. Acesso em: 10 dez. 2015.

JAEGER, P. T.; BERTOT, J. C. Transparency and technological change: Ensuring equal and sustained public access to government information. **Government Information Quarterly**, Amsterdam, v. 27, n. 4, p. 371-376, 2010.

RAMOS SIMÓN, L. F. et al. De la reutilización de la información del sector público a los portales de datos abiertos en Europa. **BiD: textos universitaris de biblioteconomia i documentació**, Catalunya, n. 29, 2012. Disponível em: <<http://bid.ub.edu/29/pdf/ramos2.pdf>>. Acesso em: 10 dez. 2015.

RDF 1.1 CONCEPTS AND ABSTRACT SYNTAX. **W3C Recommendation**. W3C, 25 February 2014. Disponível em: <<http://www.w3.org/TR/rdf11-concepts/>>. Acesso em: 10 dez. 2015.

SAORÍN, T.; PESET, F.; FERRER-SAPENA, A. Factores para la adopción de linked data e implantación de la web semántica en bibliotecas, archivos y museos. In: **Information Research**, Lund, v. 18, n. 1, p. 570, 2013. Disponível em: <<http://InformationR.net/ir/18-1/paper570.html>>. Acesso em: 10 dez. 2015.

SHETH, A. P. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: GOODCHILD, M.F. et al (Ed.). **Interoperating geographic information systems**. Massachusetts: Kluwer Academic Publishers, p. 5-29, 1999. Disponível em: <<http://knoesis.wright.edu/library/download/S98-changing.pdf>>. Acesso em: 10 dez. 2015.

SOLODOVNIK, I. Metadata issues in Digital Libraries: key concepts and perspectives. **JLIS.it Italian Journal of Library, Archives and Information Science**, Florence, v. 2, n. 2, 2011. DOI: 10.4403/jlis.it-4663. Disponível em: <<http://dialnet.unirioja.es/descarga/articulo/3945416.pdf>>. Acesso em: 10 dez. 2015.

SPIVACK, N. Making sense of the semantic web. **Blog nova spivack**, 21 nov. 2007. Disponível em: <<http://www.novaspivack.com/technology/powerpoint-deck-making-sense-of-the-semantic-web-and-twine>>. Acesso em: 10 dez. 2015.

SWARTOUT, B. et al. Toward distributed use of large-scale ontologies. In: WORKSHOP ON KNOWLEDGE ACQUISITION FOR KNOWLEDGE-BASED SYSTEMS, 10., 1996, Marina Del Rey. **Proceedings...** Marina Del Rey: Information Science Institute, 1996, p. 138-148. Disponível em: <<http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-018.pdf>>. Acesso em: 10 dez. 2016.

UPWARD, F. Structuring the records continuum (Series of two parts) Part 1: post custodial principles and properties. *Archives and manuscripts*, Sydney, v. 24, n. 2, 1996.

W3C Recommendation. **Namespaces in XML 1.0**. 3. ed. W3C, 8 dezembro 2009. Disponível em: <<http://www.w3.org/TR/xml-names/>>. Acesso em: 10 dez. 2015.

W3C Recommendation. **OWL 2 Web Ontology Language Primer**. 2. ed., W3C, 11 December 2012. Disponível em: <<http://www.w3.org/TR/owl-primer>>. Acesso em: 10 dez. 2015.

_____. **RDF Schema 1.1**. W3C, 25 February 2014. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 10 dez. 2015.

_____. **SPARQL 1.1 Query Language**. W3C, 21 March 2013. Disponível em: <<http://www.w3.org/TR/sparql11-query>>. Acesso em: 10 dez. 2015.

W3C Working group note. RIF Primer. 2. ed. W3C, 5 February 2013. Disponível em: <<http://www.w3.org/TR/rif-primer>>. Acesso em: 10 dez. 2015.

WORLD WIDE Web consortium. **Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets**. W3C Incubator Group Report, 25 de octubre de 2011. Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>

WU, P. H. J.; HEOK, A. K. H.; TAMSIR, I. P. Annotating the Web Archives: An Exploration of Web Archives Cataloging and Semantic Web. In: SUGIMOTO, Shigeo et al (Ed.). **Digital Libraries: achievements, challenges and opportunities**. New York: Springer-Verlag Berlin Heidelberg, p. 12-21, 2006.