



MINISTERIO  
DE CULTURA



Fecha de la presentación: 03/08/2010

Acceso abierto al conocimiento - Promover el Progreso Sostenible

## **Conversión de los vocabularios bibliotecarios a la Web Semántica, y viceversa**

### **Un viaje en el que todos ganan**

**Bernard VATANT**

Senior Consultant

Mondeca

París, Francia

<http://www.mondeca.com>

E-mail: [bernard.vatant@mondeca.com](mailto:bernard.vatant@mondeca.com)

**Reunión: 149. Tecnologías de la Información, Catalogación, Clasificación e Indización con la Gestión del Conocimiento**

WORLD LIBRARY AND INFORMATION CONGRESS: 76<sup>a</sup> Conferencia General y Asamblea de IFLA

10 al 15 Agosto, 2010, Gotemburgo, Suecia

<http://www.ifla.org/en/ifla76>

#### **Resumen:**

*Desde hace tiempo el papel de los vocabularios es fundamental en la sinergia entre la Web y el patrimonio bibliotecario. La Web Semántica debería aprovechar los vocabularios existentes en vez de reinventarlos, pero las características específicas de los vocabularios bibliotecarios los hacen más o menos trasladables a la Web Semántica. Basándose en los resultados preliminares en el marco del proyecto TELplus, sugerimos pautas para que se den las evoluciones necesarias para hacer a los vocabularios más eficientes y utilizables en el ámbito de la Web Semántica, para evaluar las decisiones realizadas hasta ahora por las grandes bibliotecas de publicar vocabularios conforme a normas y buenas prácticas, y analizar como las herramientas de la Web Semántica pueden ayudar a gestionar esos vocabularios.*

## **Introducción**

A pesar de la metáfora habitual de la Web como la Biblioteca Mundial Suprema, a las tecnologías de búsqueda en la Web y a los sistemas de clasificación bibliotecarios se les ha considerado durante mucho tiempo como competidores en el mercado de la organización del conocimiento, o rivales a la hora de abordar de forma independiente las demandas específicas de los usuarios. La corta historia de la Web (menos de veinte años) en comparación con los siglos de la experiencia bibliotecaria, la rápida eclosión y el ocaso de muchos negocios en la web, y el ciclo vital frenético de las tecnologías de la información se ha visto por los bibliotecarios con una mezcla de fascinación y desprecio. Sin embargo, ha surgido lenta pero inexorablemente la idea de que era deseable la cooperación y sinergia entre los "viejos" sistemas bibliotecarios y las "nuevas" tecnologías Web, y probablemente al final todo esto dará lugar a grandes beneficios para ambas partes. Más allá de los logros técnicos espectaculares en la federación y la catalogación de los datos como WorldCat [1], la gran cantidad de lenguajes y tecnologías de la Web Semántica [2] está dando lugar al crecimiento cada vez mayor de nuevas oportunidades para la valoración de bibliotecas tradicionales, nuevas perspectivas en métodos de clasificación mejorados por la biblioteconomía, así como nuevas herramientas para la gestión patrimonial.

Entre los diversos aspectos de esta convergencia semántica, nos centraremos en el papel central de los vocabularios controlados. En primer lugar, recordaremos cómo los vocabularios encajan en la evolución general de la Web y de los sistemas de información; a continuación, explicaremos por qué y cómo la Web Semántica debería aprovechar los vocabularios existentes en vez de reinventarlos; posteriormente, intentaremos determinar qué características de cada vocabulario son más o menos trasladables para su uso en la Web Semántica, y finalmente sugeriremos directrices para que sean más eficientes en este nuevo escenario.

Los ejemplos prácticos, problemas y lecciones aprendidas se basan en la experiencia en el marco del proyecto TELplus [3], donde la Biblioteca Nacional de Francia (BNF) pidió a un consorcio de empresas de Web Semántica comparar una serie de marcas semánticas y herramientas de búsqueda, integrando vocabularios de grandes bibliotecas como RAMEAU y LCSH, así como mapeos entre estos vocabularios.

Por otro lado, señalamos los beneficios de la utilización de las herramientas y el software de la Web Semántica para la gestión de vocabularios, basado en nuestra experiencia con EUROVOC [5] y otros vocabularios gestionados y publicados por la Oficina de Publicaciones de la Unión Europea.

Terminamos con una revisión de los actuales esfuerzos que se están dando hacia la organización global de esta convergencia, el papel de liderazgo de la Library of Congress Authorities and Vocabularies [4], la convergencia de diversos proyectos en el marco de la Comunidad Europea, y la iniciativa del W3C Library Linked Data Incubator Group.

## Los vocabularios de la Web, una historia de amor y odio

Los vocabularios controlados y estructurados tales como las clasificaciones y los tesauros se han desarrollado por los bibliotecarios durante siglos para servir de soporte a la búsqueda y la

*indización humana*, mientras que la búsqueda en la web se ha apoyado cada vez más en los sistemas automáticos utilizando *algoritmos estadísticos* basados en la búsqueda inteligente a texto completo. Los motores de búsqueda no necesitan vocabularios estructurados para llevar a cabo su tarea, y los que utilizan las tecnologías de búsqueda desde hace mucho tiempo sostienen la idea de que la indización y la búsqueda automáticas podrían alcanzar un alto nivel de rendimiento basándose en algoritmos cada vez más inteligentes, y que las herramientas y métodos de clasificación bibliotecaria tradicionales se deberían declarar obsoletas de forma continua, al menos en la escala de la Web. Por otro lado, los bibliotecarios y especialistas de la organización del conocimiento son de la opinión de que los motores de búsqueda no pueden tratar de manera adecuada la ambigüedad de los términos y la búsqueda inteligente debería aprovechar el conocimiento estructurado y el valor añadido de lo humano embebido en los vocabularios estructurados.

### **El auge de los motores de búsqueda**

El debate estuvo bastante abierto durante la primera década de la Web, cuando el nivel de rendimiento de los motores de búsqueda era todavía excesivamente pobre [6]. Pero su impresionante auge y el éxito de Google durante la segunda década de la Web parecen poner sobre la mesa un argumento demoledor a favor de la indización automática y los campos de búsqueda. Sobre la base de este éxito, algunos han ido más lejos con este argumento hasta el extremo de declarar la obsolescencia de los vocabularios controlados en beneficio de las tecnologías de búsqueda. Por supuesto, la obsolescencia de los vocabularios significaría también la obsolescencia de los metadatos utilizados en ellos, así como la propia clasificación en general, y al fin y al cabo, como quien dice, la muerte de la biblioteconomía.

### **Los seres humanos no lo hacen mucho mejor**

Por otro lado, en los primeros años de la Web surgieron tentativas de organización de la misma bajo esquemas de clasificación de forma análoga a las grandes bibliotecas, a la medida de la propia Biblioteca Global. El directorio Yahoo!, seguido por su alternativa colaborativa Open Directory, éste último copiado en cientos de diversos avatares incluyendo a Google Directory, ha encontrado rápidamente graves problemas de escalabilidad, dada la gran magnitud, variedad y total inestabilidad de los recursos Web. A pesar de la frase en negrita "Los seres humanos lo hacemos mejor" [7], el eventual fracaso de estas tentativas en cuanto a su coste de mantenimiento en comparación con la realización de la indización automática se ha traducido en un descrédito general de las herramientas y servicios de clasificación de la Web global, en beneficio de los algoritmos de los motores de búsqueda. Sin embargo, muchos de los directorios verticales se han mantenido a flote, y todavía están vivos.

Un daño colateral de la falta de paradigma de los directorios Web podría ser la imposición en la mente del usuario final medio de la vaga noción de que los vocabularios estructurados son instrumentos heredados de una edad oscura de la clasificación bibliotecaria, antes de que Google se inventara. Por otra parte, este usuario no es consciente de las diferencias entre un tesoro, un sistema de encabezamientos de materia, una clasificación, un índice, una tabla de contenidos, una taxonomía, un directorio... todos los esquemas de vocabularios que supuestamente ayudan a organizar y recuperar cosas. Por ejemplo, el usuario de Wikipedia se

verá frente a una variedad de "listas" y "categorías" que crece de forma orgánica, sin un plan general, basado en las directrices difusas de personas no siempre bien adiestradas en el arte de la clasificación [8].

Con todo, la experiencia del usuario medio en cuanto a vocabularios se refiere en la Web es que están tan desordenados como lo está la propia web, con lugares bien organizados a nivel local en medio de un caos general. En lugar de ayudar a organizar y encontrar cosas de forma clara y rápida, los vocabularios en la Web a menudo sólo se añaden al caldo de conocimiento global [9], aportando a lo sumo algún tipo de descubrimiento afortunado.

### **El auge creciente de las taxonomías**

Mientras tanto, en el mundo cerrado de los sistemas de información empresarial, la importancia de los vocabularios controlados y estructurados es cada vez más reconocida y manejar tales vocabularios se ha convertido en lo suficientemente importante como para sostener el modelo de negocio de una serie de empresas. En el ámbito empresarial, los vocabularios controlados por regla general, aparecen con el nombre genérico de "taxonomías". Heather Hedden escribe en la introducción de "The Accidental Taxonomy" [10]:

*Actualmente en la gestión de la información, el término taxonomía se utiliza tanto en un sentido estricto, significando una clasificación jerárquica o un sistema de categorización, como en un sentido amplio, haciendo mención al significado de organización de conceptos del conocimiento. A algunos profesionales incluso no les gusta usar este término, por considerarlo demasiado ambiguo y frecuentemente mal utilizado. Sin embargo, ha ido ganando bastante popularidad, y no parece existir un término práctico alternativo.*

Así que, nos guste o no, vuestros vocabularios, una vez publicados, e independientemente de su estructura original y uso específico, es probable que se les conozca en su ámbito como "taxonomías" [11].

## **Hacia una reconciliación semántica**

### **Apertura de los almacenes de datos y la necesidad de la interoperabilidad semántica**

En los últimos años ha ido surgiendo una nueva tendencia: la apertura general de los sistemas de información. Es por todos reconocido que la duplicación y gestión de datos en muchos almacenes de información y bases de datos cerrados tienen un coste increíble, cuando una gran cantidad de datos puede estar disponible al público desde servidores "en la nube" y pueden ser consumidos bajo demanda. La distinción entre la Web abierta y el sistema de información de empresa es cada vez más y más difusa. SaaS (Software como un Servicio) ya ha llegado a su mayoría de edad, continuada ahora por DaaS (Datos como un Servicio). Por supuesto, el intercambio de datos entre sistemas necesita no sólo de formatos de intercambio de datos comunes (como XML), sino también un acuerdo sobre la semántica de estos datos. La *interoperabilidad semántica* se ha convertido en el aspecto clave, y para ello es evidente que son necesarios vocabularios de referencia común, con semánticas declaradas

formalmente.

### **Semántica básica: resolver la ambigüedad y la correlación**

El conjunto de la Web Semántica se presenta a menudo como una tecnología de la inteligencia artificial, donde el punto principal es la representación formal de las cosas en lenguajes basados en descripciones lógicas, tales como OWL. Es evidente que esta es la parte más avanzada del conjunto, utilizando restricciones y reglas complejas para resolver incoherencias y deducir nuevo conocimiento. Pero antes de eso, y ciertamente más importante, los lenguajes y herramientas de la Web Semántica proporcionan modos de tratar la *ambigüedad* y la correlación. Para ser breves, permiten expresar de forma explícita si dos cosas, que llevan o no el mismo nombre, son iguales o no; por ejemplo si dos documentos tratan de la misma materia, si tienen el mismo autor, si dos recursos se encuentran en el mismo lugar, etc. A este respecto, los principios de la Web Semántica siempre han estado más cerca de la biblioteconomía que de la infraestructura de la llamada Web 1.0. Considerando que la Web 1.0 se centra en el acceso a los recursos de información a través de URLs y enlaces de hipertexto estáticos, la Web Semántica se ocupa primero y sobre todo de *identificar recursos de cosas que están relacionadas*.

La identidad de las cosas, manteniendo la precisión y la correlación, se garantiza en la Web Semántica a través de los identificadores universales (URIs) distintos de los términos del lenguaje natural utilizados en vocabularios bibliotecarios. Sin embargo, los principios de la Web semántica están más cerca de la biblioteconomía que de la lógica de los motores de búsqueda. La incorporación de la información en ambos universos es "un concepto central", incluso si los conceptos se representan en la Web Semántica por URIs y descripciones formales, y los vocabularios de bibliotecas a través de términos del lenguaje natural. Lo que se necesita simplemente es reunir a estos dos universos y trabajar en sinergia en un lenguaje que enganche los vocabularios de biblioteca a los identificadores Web (URI).

### **SKOS: ¿un vocabulario para gobernarlos a todos?**

El lenguaje SKOS [12] se ha desarrollado desde el año 2005 a través de un diálogo fructífero entre las bibliotecas y las comunidades de la Web Semántica, con el objetivo declarado de permitir una fácil migración de los vocabularios tradicionales. Desde 2009, SKOS, una recomendación de W3C, se ha utilizado rápidamente para llevar a cabo la migración de los vocabularios tradicionales, incluyendo RAMEAU [13], LCSH [14], AGROVOC [15], EUROVOC y muchos otros. La mayoría de los editores de vocabularios es probable que sigan este ejemplo en los próximos meses y años.

SKOS no se ha diseñado específicamente para vocabularios bibliotecarios, sino que pretende utilizarse para una gran variedad de Sistemas de Organización del Conocimiento, que abarca las "taxonomías" en el sentido amplio comentado anteriormente. Los vocabularios estructurados normalizados como los tesauros se pueden traducir a SKOS con bastante facilidad, pero para algunos vocabularios con características más complejas debe ampliarse la expresividad de SKOS.

Por otra parte, la traducción de un vocabulario a un formato SKOS no significa que proporcione un uso eficaz en la Web Semántica. Veremos en la siguiente sección algunas cuestiones prácticas, que tratan de identificar lo que hace que un vocabulario sea útil, y por lo tanto cómo se puede mejorar para una mayor eficiencia.

## Construyendo vocabularios bibliotecarios listos para la Web Semántica

La integración de vocabularios tiene un claro valor añadido para las aplicaciones de la Web Semántica, tales como notaciones, herramientas de búsqueda y navegación, extensión semántica de búsqueda, etc. Pero parece que algunos vocabularios son más eficientes que otros en este sentido, muy adecuados para la utilización humana y no necesariamente tan bien ajustados para las máquinas. Las máquinas no tienen la inteligencia sutil de los usuarios humanos, necesitan definiciones y reglas explícitas precisas, en concreto más aún para entornos abiertos denominados de interoperabilidad a gran escala. Las máquinas no presentan excesivos requisitos. Lo único que necesitan es una semántica explícita, e interpretan la misma sintaxis de la misma manera, siempre y en todas partes. Un primer efecto positivo de la conversión de un vocabulario a la Web Semántica es generalmente el de descubrir que no siempre significa lo que se supone que significa eso en la cabeza de sus autores y usuarios humanos.

Revisemos ahora varias cuestiones y lecciones aprendidas de la comparación utilizando RAMEAU, LCSH y SDW en el marco de TELplus. Los temas que realmente se destacan son los de la evolución de los vocabularios, a fin de hacerlos realmente eficaces y útiles. Se presentan en forma de normas de buenas prácticas, de las cuales algunas se aplicarían a la migración de datos hacia aplicaciones semánticas, y otras son específicas de vocabularios de bibliotecas.

### 1. Identificar y describir conceptos no ambiguos

Los supuestos básicos de la Web Semántica en lo que respecta a los conceptos, o más concretamente una "cosa" definida por un URI es simple, pero fuerte.

1. La semántica (significado del URI) es independiente del contexto. Siempre que el URI se encuentre, significa (denota) la misma cosa. Los URIs son nombres únicos universales.

2. La descripción formal del concepto, si es posible obtenida dereferenciando el URI a través de un protocolo web (la mayoría de las veces, HTTP), transmite esta semántica y hace al concepto preciso.

En otras palabras, la descripción formal del concepto debe ser suficientemente explícita para que sea claramente distinguible de, o claramente idéntica a, cualquier otro concepto definido por cualquier otro URI en la Web. Esto es una exigencia bastante rigurosa, pero en este sentido testar el vocabulario publicado es la primera tarea del potencial usuario de la web semántica.



Se necesita una política de identificadores únicos, ya que con frecuencia los vocabularios bibliotecarios generales dependen de la identificación de los nombres en su contexto. Es fundamental hacer distintos los conceptos y los términos utilizados para nombrarlos, por lo que los identificadores deben ser notaciones independientes de las etiquetas. Si el sistema de gestión de un vocabulario utiliza claves únicas internas para los conceptos, una buena práctica es construir URIs para estas claves, así como para los URIS de LCSH que se presentan a continuación.

## 2. Explicitar la semántica de la sintaxis del vocabulario

Cuando se utilizan vocabularios en un contexto bibliotecario, se puede obtener gran precisión por el usuario humano. Un término puede ser ambiguo en el lenguaje natural, pero ya no en el contexto de un vocabulario controlado. "Table" es ambiguo en inglés (y en francés también) pero lo será menos si se define en el contexto de "muebles de casa" o "base de datos". Para aplicaciones de la Web Semántica los dos conceptos "Table (muebles para el hogar)" y "Table (base de datos)" deberían identificarse como conceptos diferentes, por diferentes URIs y enlazarse de manera formal a sus respectivos contextos. Estos contextos tienen que ser formalmente identificados y declarados ya sea como un concepto genérico o relacionado, o como un esquema conceptual (en función del nivel de granularidad). No es suficiente el simple indicio del contexto por un calificador como en la práctica de los tesauros.

De forma similar a los calificadores, existen otras estructuras sintácticas específicas que ocultan semánticas implícitas que deben hacerse explícitas. El formato MARC utilizado por LCSH y RAMEAU contiene construcciones sintácticas como "Actores - Psicología" (<http://id.loc.gov/authorities/sh85000748>). Aunque la relación parece lógica para los usuarios humanos, la descripción formal de este concepto no proporciona enlaces ni a "Actores" (<http://id.loc.gov/authorities/sh85000744>) ni a "Psicología" (<http://id.loc.gov/authorities/sh85108459>). Para las aplicaciones de la Web Semántica, esta relación debe ser explícita para ser utilizable.

Una consecuencia indirecta de la elaboración de estos ejemplos para el gestor de vocabularios es obligar a una reflexión sobre la semántica subyacente de la sintaxis, que se debería haber tenido en cuenta hace años. Para la migración automática al formato de Web Semántica, las construcciones sintácticas idénticas deben ser interpretadas y procesadas de la misma manera. Pero parece que averiguar una semántica general para dichos "--" separadores, va más allá de un general y vago "SKOS: related", y que dicha semántica sea válida totalmente en un vocabulario grande, como RAMEAU o LCSH, no es una tarea obvia. Incluso habiendo pautas para los administradores de vocabularios explicando en lenguaje natural lo que debería ser la utilización correcta de cada construcción sintáctica, en general no hay manera de hacer cumplir esas directrices en el sistema de gestión de vocabularios, por no hablar de la evaluación de las consecuencias de su uso incorrecto.

## 3. Comprobar la actual transitividad de las jerarquías

La estructura jerárquica de los vocabularios, es probable que se traduzcan en las propiedades SKOS: broader o SKOS: narrower, que por lo general serán utilizadas por motores semánticos



para realizar la ampliación o restricción de la búsqueda para conceptos más genéricos o más específicos, a fin de reducir el ruido o el silencio en los resultados. Dicha ampliación es probable que siga la jerarquía de una manera transitiva. Aunque ha habido un gran debate en SKOS sobre este tema, finalmente cerrado por la decisión de que esas propiedades no son en general transitivas, probablemente muchas aplicaciones las consideran como tales.

Si se mira de cerca a las jerarquías de gran tamaño como LCSH o RAMEAU, está claro que la extensión semántica por transitividad es "hasta cierto punto" o "localmente" válida, por así decirlo. Desde que los gestores de vocabulario no tienen necesariamente una visión global del enorme tamaño de un vocabulario dado, es probable que exista una especie de "deriva semántica" en la jerarquía de arriba hacia abajo, provocando una extensión semántica con errores. El ejemplo siguiente, tomado de LCSH no es tan extremo como parece, ya que hemos encontrado un montón de casos similares. En RAMEAU incluso descubrimos que algunos ciclos que abarcan hasta trece conceptos, vuelven locos a los motores de extensión semántica.

Ciencias auxiliares de la historia

. Civilización

...Aprendizaje y Educación

... Humanidades

.... Filosofía

.....Psicología

.....Atención

... .. Escuchar

.....Escuchas

... .. Escuchas telefónicas

Si uno mira localmente esta jerarquía, no existe ningún fallo en cualquiera de las semánticas genérico-específicas. Cada relación local tiene sentido. Pero el uso de la extensión semántica de forma global podría llevarnos a resultados extraños.

Por tanto, pidiendo la semántica explícita llevará finalmente al administrador de vocabularios a cuestionar la pertinencia de cada gran jerarquía y la deriva semántica que pueda implicar, incluso dentro del sistema bibliotecario, por no hablar de la Web abierta.

#### 4. Mantenerlos tan pequeños y simples como sea posible

Los ejemplos anteriores muestran suficientemente bien que los grandes vocabularios en general no están bien adaptados a las tecnologías semánticas, y deberían estar mejor organizados en dominios verticales, donde es más fácil la precisión. Un vocabulario eficiente es o *poco profundo en significado y amplio en el ámbito de aplicación*, o de *poco alcance y profundo en significado*. Los intentos de conseguir tanto amplio alcance como profundo significado en esos grandes vocabularios, con cientos de miles de conceptos, los llevan a dificultades para mantener su uso. Esta es una lección básica aprendida a las malas, ya que los grandes vocabularios son parte de la tradición, y han sido muy utilizados para la indización. No obstante para aprovechar los índices tradicionales, una tarea prioritaria debería ser mapear estos grandes vocabularios generales a unos más pequeños verticales. Por supuesto que no se hará en un día, pero pueden ayudar las herramientas de gestión construidas bajo los



lenguajes de la web semántica. Volveremos a este punto en la última sección.

## 5. Mapear con otros vocabularios

Mapear los conceptos equivalentes en otros vocabularios, en especial en entornos multilingües, proporciona un importante valor añadido para el vocabulario. Incluso en entornos monolingües, los vocabularios generales se pueden extender a otros más especializados, desarrollados por grupos independientes. SKOS ofrece la expresividad para estos mapeos, y los motores de búsqueda semánticos pueden aprovechar tales mapeos para la extensión semántica. En este sentido, el trabajo realizado en el proyecto de alineación MACS [16] está marcando la pauta.

En particular, como se ha definido anteriormente el mapeo de grandes vocabularios generales con vocabularios más pequeños facilitará su utilización por taxonomías de navegación proporcionando puntos de vista simplificados y personalizados en índices y catálogos de una biblioteca.

## 6. Los conceptos tienen un ciclo de vida, pero los inflexibles URIs no cambian

Los URIs publicados para los conceptos en los vocabularios deben ser sólidos en el tiempo; las aplicaciones de la Web Semántica requieren que el URI se utilice como algo "robusto" [17] en la medida de lo posible. Aunque los conceptos cambien en un vocabulario, se aplicarán métodos estrictos para asegurar la persistencia de los URIs y las descripciones adjuntas, incluso para conceptos obsoletos. Se pueden implementar los mecanismos de rechazo, así como los mecanismos de redirección utilizando lenguajes y protocolos de la Web Semántica.

## 7. Exponer y promover tu Vocabulario como un Servicio

La Web Semántica no es sólo un montón de lenguajes, es también una arquitectura de servicios. Para ser plenamente útil, el vocabulario debe estar disponible para aplicaciones en diversas formas, todo ello sustentado en la arquitectura Web. Sin entrar en detalles que están bien descritos en excelentes tutoriales [18], vamos a resumir las principales recomendaciones.

1. Cada URI de un concepto debe sostener la negociación del contenido, proporcionando descripciones formales RDF para las máquinas y HTML a los navegadores habituales para el entendimiento humano.
2. Hacer que los vocabularios estén disponibles en paquetes descargables, un paquete por Esquema Conceptual.
3. Establecer un criterio de valoración SPARQL con el fin de permitir a los usuarios extraer partes del vocabulario adecuado a demandas específicas.
4. Exponer el contenido de los vocabularios utilizando por ejemplo la ontología VOID [19].
5. Por último, pero no menos importante, promover su publicación a través de diversos foros de la comunidad de la Web Semántica.

## 8. Usar el software de la Web Semántica para la gestión del vocabulario

Por último, pero no menos importante, el desarrollo del software manteniendo los lenguajes de la Web Semántica trae consigo nuevas herramientas para gestionar vocabularios, facilitando las tareas como el control de la coherencia, la mejora de la expresividad utilizando ontologías con más relaciones específicas que las simples relaciones jerárquicas y asociativas, el manejo de versiones, el rechazo y la evolución de los conceptos, el mapeo con otros vocabularios, la importación y exportación original en RDF, y por último la integración en SPARQL.

En este sentido, la Comunidad Europea ha dado ejemplo en la adopción del paquete de software de la Web Semántica para administrar y publicar EUROVOC y en el futuro muchos otros vocabularios de la Oficina Europea de Publicaciones. La solución necesitó de un amplio número de requisitos de procesos y gestión para un vocabulario multilingüe (más de veinte idiomas), de gestión de términos a alto nivel, incluidos sinónimos, traducciones y siglas, de la organización en micro-tesauros con diferentes restricciones, y desde luego de la publicación en formato SKOS, incluyendo control de versiones y marcas de tiempo.

## El trabajo en curso y las perspectivas

Vamos a terminar con un rápido repaso de las iniciativas en curso en el espacio de la Biblioteca de la Web Semántica. Esto por supuesto es una lista abierta y no exhaustiva, que puede ampliarse en los próximos meses y años.

### Library of Congress Authorities and Vocabularies

La Library of Congress ha llevado la pauta al iniciar la publicación de sus vocabularios en formatos conformes a los estándares de la Web Semántica y las buenas prácticas de Linked Data. El primer conjunto de vocabularios publicado en 2009 fue LCSH, y se han añadido otros en 2010. La transformación automática de vocabularios existentes en SKOS, como se mencionó anteriormente, plantea una serie de cuestiones, y esta iniciativa tiene que ser evaluada cuidadosamente. Pero es, por supuesto, una referencia indiscutible, el espacio de nombre id.loc.gov está aquí para quedarse, y aún más cualquier publicación en el espacio de la biblioteca de la Web Semántica debería mirar en este espacio y conectarse a él en la medida de lo posible

### Los Vocabularios Europeos

Los estudios preliminares realizados en el marco de TELplus han dado lugar a intercambios entre la Biblioteca Nacional de Francia y sus socios técnicos, a fin de definir la ruta de migración de RAMEAU hacia la usabilidad de la Web Semántica. Esperamos que este trabajo finalmente se extienda a otros socios en el marco de Europeana y la Biblioteca Europea. Los desafíos multiculturales y multilingües son altos, pero es un espacio muy prometedor. Hemos mencionado anteriormente las iniciativas de la Oficina Europea de Publicaciones, y desde luego de la sinergia y el intercambio de buenas prácticas entre todas las iniciativas en Europa que deben llevar finalmente al surgimiento de una abundancia de vocabularios públicos

multilingües, interconectados para el acceso a recursos públicos europeos de cualquier naturaleza: culturales, reguladores, económicos.

### **W3C Library Linked Data Incubator Group**

Iniciado en mayo de 2010, el W3C Library Linked Data Incubator Group [20] se fletó en un año para reunir los esfuerzos en este ámbito, identificar y promover buenas prácticas, y ayudar a la sinergia entre la Biblioteca y la Web Semántica y llevar a ambas a su " máximo potencial". Además de recoger a los principales actores de la comunidad bibliotecaria, este grupo también trabajará en la sinergia con el grupo de trabajo de ISO en la nueva norma de Tesoros (ISO 25964) y el Dublin Core Metadata Initiative.

## **Conclusión**

Esperamos haber mostrado que ahora es el momento para los gestores de vocabularios para que participen activamente en una cooperación productiva con la comunidad de la Web Semántica, con el fin de ver su precioso legado tanto mejorado como aplicado de nuevas maneras. Se creó un marco para esta tarea por el W3C, y citando la carta del W3C Library Linked Data Incubator Group:

*Es necesaria una reorientación en la perspectiva bibliotecaria sobre la interoperabilidad de la información, basándose en la existencia de una arquitectura Web y unas normas, a fin de llevar estos contenidos a la Web. Ya están disponibles una gran cantidad de datos estructurados dentro de los sistemas bibliotecarios y podrían ser mostrados como Linked Data, utilizando tecnologías de la Web Semántica. Las instituciones del patrimonio cultural podrían ser un importante proveedor de conjuntos de datos de autoridad (personas, materias...) para el Linked Data Web.*

Los conjuntos de datos de autoridad son aquí claramente vocabularios tradicionales. El conjunto de tecnologías de la Web Semántica debería permitir a los bibliotecarios manejarlos de una manera más eficiente, poniéndolos disponibles a través de interfaces de servicios semánticos. En resumen, la evolución que se espera conseguir son editores de vocabularios abiertos y extender su uso tradicional a la Web en general, a través de una arquitectura distribuida y en red del Vocabulario como un Servicio.

## **Referencias y enlaces**

- [1] OCLC WorldCat <http://www.oclc.org/us/en/worldcat/default.htm>
- [2] W3C Semantic Web Activity <http://www.w3.org/2001/sw/>
- [3] TELplus <http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/>
- [4] Library of Congress Vocabularies and Authorities <http://id.loc.gov/>
- [5] EUROVOC <http://europa.eu/eurovoc/>
- [6] Search Engine History <http://www.searchenginehistory.com/>



- [7] Humans Do It Better: Inside the Open Directory Project, Chris Sherman, ONLINE Mag, July 2000 <http://www.onlinemag.net/ol2000/sherman7.html>
- [8] Wikipedia discussion about the "Uncategorized categories" list [http://en.wikipedia.org/wiki/Wikipedia\\_talk:Special:UncategorizedCategories](http://en.wikipedia.org/wiki/Wikipedia_talk:Special:UncategorizedCategories)
- [9] Representing Knowledge Soup in Language and Logic, John Sowa, 2002 <http://www.jfsowa.com/talks/souprepr.htm>
- [10] The Accidental Taxonomist, by Heather Hedden, Information Today, May 2010, ISBN 978-1-57387-397-0 <http://www.hedden-information.com/accidental-taxonomist.htm>
- [11] The Taxonomy Warehouse - <http://www.taxonomywarehouse.com/>
- [12] Simple Knowledge Organization System <http://www.w3.org/2004/02/skos/>
- [13] RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) <http://rameau.bnf.fr/>
- [14] Library of Congress Subject Headings <http://id.loc.gov/authorities#ConceptScheme>
- [15] AGROVOC Thesaurus <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>
- [16] Integrating MACS initial data and new alignments into TEL framework [http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/documents/TELplus\\_D3.4\\_04012010.pdf](http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/documents/TELplus_D3.4_04012010.pdf)
- [17] Cool URIs don't change, Tim Berners-Lee, 1998 <http://www.w3.org/Provider/Style/URI>
- [18] How to Publish Linked Data on the Web, C.Bizer, R. Cyganiak, T. Heath. <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>
- [19] Vocabulary of Interlinked Datasets <http://semanticweb.org/wiki/VoiD>
- [20] Library Linked Data Incubator Group charter <http://www.w3.org/2005/Incubator/lld/charter>

## Sobre el autor

Se graduó en la Ecole Normale de l'Enseignement Technique (ENSET) en 1975; Bernard Vatant ha enseñado matemáticas en la *Education Nationale* de Francia desde 1975 hasta 1997. Reorientó su actividad hacia las nuevas tecnologías de representación, publicación y distribución del conocimiento.

A finales del año 2000, se integró en el equipo Mondeca como consultor en el ámbito de las ontologías y lenguajes de representación del conocimiento. Su experiencia en la modelización y migración de datos ha ido aumentando desde esa fecha, gracias a la diversidad de clientes y proyectos de Mondeca, por lo que su campo abarca ámbitos tan diversos como terminología científica y médica, turismo y gobierno local o publicaciones jurídicas.

Bernard Vatant ha sido el representante de Mondeca en varios grupos de trabajo u organismos de normalización como la ISO (ISO 13250 norma Materia para Mapas, y grupos de trabajo actuales sobre la futura norma ISO 25964 sobre Tesoros), o W3C (OWL, SKOS). Presidió de 2001 a 2003 el Comité Técnico de OASIS sobre «Materias publicadas».

Es un personaje conocido de la investigación sobre el uso industrial de tecnologías de la Web Semántica, y como tal es participante habitual en las conferencias especializadas de los Program Committees ya sea en Francia, como en IC (Journées francophones d'Ingénierie des Connaissances) o en el extranjero, tales como LDOW (Linked data on the Web), y ha sido conferenciante invitado a los talleres organizados por organizaciones como el INRIA en Francia



MINISTERIO  
DE CULTURA



o ISKO en el Reino Unido.

Su experiencia es reconocida, en particular en el ámbito de la modelización, migración e interoperabilidad de vocabularios de referencia tradicionales, y en este sentido ha trabajado con instituciones como la Oficina de Publicaciones de la Unión Europea (vocabulario EUROVOC), o la Biblioteca Nacional de Francia (BNF) en el marco del proyecto TelPlus (evolución e integración del vocabulario RAMEAU). Es desde mayo de 2010 un Invited Expert en el W3C Library Linked Data Incubator Group.

Traducción: Pascual Jiménez Huerta  
Biblioteca Nacional de España