

# Una aplicación de la Wikimedia Semántica

Marcela Vegetti, Horacio Leone

INGAR (CONICET - UTN), Avellaneda 3657, Santa Fe, Argentina  
{mvegetti, hleone}@santafe-conicet.gov.ar

**Abstract.** Ontolog es una comunidad abierta virtual que trabaja colaborativamente para impulsar avances en el campo de la ingeniería ontológica y las tecnologías semánticas. Se encuentra alojada en una plataforma wiki que no posee las características necesarias para realizar búsquedas dentro de este cuerpo de conocimientos. Asimismo, la cantidad de información almacenada y la organización de la misma dificulta el acceso por parte de personas no familiarizadas con la estructura del contenido. Este artículo presenta una propuesta que incorpora semántica a las páginas de la wiki Ontolog y, además, provee una capa que permite el fácil acceso a la información relevante, reutilizando la información almacenada.

Keywords: ICOM; wikimedia semántica; ontolog

## 1 Introduction

El foro ontolog<sup>1</sup> es una comunidad abierta virtual dedicada a impulsar avances en el campo de la ingeniería ontológica y las tecnologías semánticas. Una plataforma wiki archiva las contribuciones de los miembros de la comunidad sirviendo de repositorio de conocimiento para la misma.

Una de las actividades impulsada por el foro Ontolog es la cumbre de ontologías, una serie anual de eventos que involucra a representantes de la comunidad de ontologías así como de las comunidades relacionadas con el tema elegido para la cumbre de cada año. Al igual que el resto de los proyectos del foro Ontolog, toda la información sobre las diferentes actividades que son llevadas a cabo durante cada cumbre anual son mantenidas en la infraestructura colaborativa de Ontolog.

La cantidad de información, y la forma en que está organizada, dificulta el acceso rápido a las partes más importantes de la misma para aquellas personas no familiarizadas con la estructura del repositorio. Es por esto que el comité organizador de la edición 2012 decidió que a partir de ese año debería existir, además de la información en la plataforma wiki, un sitio web que resuma la información más importante. En 2012, este desarrollo ha implicado la reescritura de parte del contenido, duplicando la información con el esfuerzo e inconvenientes que eso implica.

---

<sup>1</sup> <http://ontolog.cim3.net/>

Para evitar este inconveniente y teniendo en cuenta el objetivo que tiene Ontolog de impulsar la utilización de las tecnologías semánticas en las aplicaciones, surge la necesidad de implementar una plataforma que facilite el acceso, el reuso y la incorporación de semántica al contenido del foro Ontolog.

Para atender estos requerimientos, este trabajo presenta una propuesta, que se está implementando en una plataforma wikimedia semántica, denominada Ontolog PSMW. La ontología ICOM (Integrated Collaboration Object Model for Interoperable Collaboration Services)<sup>2</sup> es utilizada para definir la semántica del cuerpo de conocimiento y una capa de presentación sirve de máscara a la información almacenada, proveyendo un fácil acceso a la información relevante. El resto de este artículo se organiza como sigue. La sección 2 presenta una breve descripción de la organización de la serie de cumbres de ontologías y de su contenido, así como de la tecnología de la plataforma en la que está contenida. La sección 3 describe las características de la propuesta e ilustra cómo se ha aplicado al cuerpo de conocimiento de la cumbre de ontologías 2013. Finalmente, en la sección 4 se presentan las conclusiones y trabajos futuros .

## **2 Cumbres de ontología. Descripción y plataforma de soporte**

La serie de cumbres de ontologías se viene desarrollando como iniciativa del foro Ontolog y el NIST (US National Institute of Standards and Technology) desde el año 2006. Cada año se selecciona un tema que es discutido durante una serie de eventos que se extienden alrededor de tres meses, desde enero a abril. Estos eventos incluyen teleconferencias semanales, discusiones en línea sobre el tema elegido, así como otras actividades particulares a cada año. Cada cumbre finaliza con un simposio presencial en el que se resumen los resultados de las diferentes actividades. Además, cada año se presenta en un comunicado el mensaje de los participantes de la cumbre para la comunidad de ontologías.

El tema elegido cada año es abordado desde diferentes facetas, que son analizadas y estudiadas en distintos espacios de discusión, denominados "Tracks". Cada uno de éstos está a cargo de uno o dos coordinadores responsables. Cada track debe organizar una o más conferencias virtuales, las cuales comprenden entre 3 y 6 presentaciones cortas sobre temas dentro del alcance del mismo, seguida de una sesión de preguntas y respuestas entre participantes y panelistas invitados. Estas reuniones virtuales están soportadas por una audio conferencia y una sala de chat. Todo el proceso de la cumbre es coordinado por un comité organizador integrado por los responsables de las diferentes actividades y dos coordinadores generales.

Cada track produce como resultado una página wiki que sintetiza la discusión. Los coordinadores del track son los responsables de crear y mantener dicha página en base a: i) las discusiones por mail que se promueven en la lista [ontology-summit-list], ii) las contribuciones que los participantes hacen en una página destinada a tal fin (Track community input), iii) las presentaciones en las conferencias virtuales que

---

<sup>2</sup> <https://wiki.oasis-open.org/icom>

el Track organiza y iv) los resultados de actividades específicas propuestas por los coordinadores del track.

La cumbre de este año ha puesto el foco en "La evaluación de las ontologías a lo largo de su ciclo de vida" y ha sido organizada en cuatro tracks: i) Aspectos intrínsecos de la evaluación de ontologías, ii) Aspectos extrínsecos de la evaluación de ontologías, iii) Construcción de Ontologías para cumplir con criterios de evaluación, y iv) Entornos de software para la evaluación de ontologías.

Como se mencionara en la sección previa, el repositorio del foro Ontolog se encuentra alojado en una plataforma wiki. La misma permite, a través de los denominados "purple numbers" (números púrpura), referenciar a párrafos específicos dentro de una página utilizando un código alfanumérico único dentro de la página. Desde otras páginas es posible referenciar a estos nodos utilizando el identificador del nodo, denominado nid ("node identifier") o número púrpura, en el URL de la página. Esta plataforma wiki se constituye en un importante repositorio del conocimiento que se genera colaborativamente en las cumbres de ontología. Sin embargo, a pesar de su utilidad, la información allí contenida no es fácilmente utilizable por herramientas externas. No es sencillo recolectar información que se encuentra distribuida en diferentes páginas, como por ejemplo, cuáles son las conferencias virtuales en las cuales ha participado alguna persona. A pesar de que la información en una wiki se encuentra estructurada (cada conferencia tiene su página con links a miembros que participaron en ella), su significado no es claro para una computadora, ya que no se encuentra representado de manera procesable por máquina. Por lo cual, este contenido no puede ser consultado utilizando las nuevas tecnologías de la Web Semántica.

Asimismo, el cuerpo de conocimiento almacenado en la wiki Ontolog es muy voluminoso y posee una organización que impide que personas no familiarizadas con su estructura puedan acceder fácilmente al mismo. Si tomamos, por ejemplo la página de alguna conferencia virtual, la misma contiene bastante información sobre ella: el tema a tratar, quienes son los coordinadores, los panelistas, una pequeña reseña de cada uno de ellos y el enlace a sus transparencias, también se muestra la transcripción del chat (una vez que fue realizada la conferencia) e incluso las instrucciones para poder conectarse para participar de la misma (información que no es importante una vez finalizada).

### **3 Incorporación de Semántica y Reorganización del Contenido**

Como se mencionara en la sección previa, el contenido alojado en la wiki Ontolog está siendo migrado a una plataforma wikimedia que incorpora características semánticas. Sin embargo, la sola migración a esta nueva plataforma no agrega semántica al contenido, así como tampoco implica la reorganización del mismo. A fin de cubrir estas necesidades se presenta una propuesta que incorpora semántica al contenido a través de la ontología ICOM y una capa que provee acceso al repositorio de información mediante una vista reorganizada de la misma. La Fig. 1 muestra los componentes principales de esta propuesta, los cuales serán introducidos a continuación.

PSMW (Purple Semantic Media Wiki) es una plataforma que, mediante las extensiones SMW (Semantic Media Wiki) [1] y PMWX (Purple mediawiki) [2], permite agregar semántica y acceso de "grano fino" al contenido de páginas wiki. El contenido del foro Ontolog está siendo migrado a esta nueva plataforma, creando el foro Ontolog PSMW, que posibilitará la incorporación de semántica al contenido.

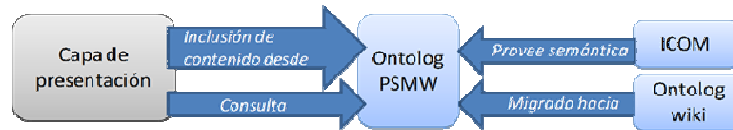


Fig. 1. Componentes de la propuesta

A través de la extensión SMW es posible hacer explícita la información acerca de las relaciones existentes entre las páginas wiki. En la mediawiki estas relaciones se establecen con los links entre las páginas. Cada enlace establece cierta relación entre dos páginas, sin explicitar qué tipo de relación se trata. SMW permite caracterizar estos links mediante la definición de propiedades, de manera que el destino de un link se convierte en el valor de una propiedad definida para la página en la que está el enlace. De esta manera, es posible formular consultas semánticas sobre el contenido de las páginas. Para conseguir esto se requiere la definición de 4 tipos de páginas diferentes: propiedades, plantillas, formularios y categorías, en sus correspondientes espacios de nombre: Property, Template, Form y Category.

Las propiedades, junto con los tipos, son las formas básicas de introducir datos semánticos en la wikimedia semántica. Una propiedad se utiliza para anotar una porción de información en una página. Como ejemplo, considere la Fig. 2 en la cual se muestra el código utilizado en la wiki original y en Ontolog PSMW para mostrar una porción de texto en una página wiki. En el segundo caso, se puede observar que, los enlaces de las páginas a los organizadores de Ontolog han sido anotados con la propiedad *hasOrganizer*.

**Texto que se ve en la página wiki:**

communities related to each year's theme chosen for the summit. The Ontology Summit program is now co-organized by Ontolog, NIST, NCOR, NCBO, IAOA, NCO\_NITRD along with the co-sponsorship of other organiza that are supportive of the Summit goals and objectives. See: (1A)

... The Ontology Summit program is now co-organized by `[[Ontolog]],` **Código**  
`[[NIST]], [[NCOR]], [[NCBO]], [[IAOA]], [[NCO_NITRD]]` along with **Ontolog wiki**  
 the co-sponsorship ...

**Código**  
**PSMW** → ... The Ontology Summit program is now co-organized by  
`[[hasOrganizer::Ontolog]], [[hasOrganizer:: NIST]], [[hasOrganizer::`  
`NCOR]], [[hasOrganizer:: NCBO]], [[hasOrganizer:: IAOA]],`  
`[[hasOrganizer:: NCO_NITRD]]` along with the co-sponsorship ...

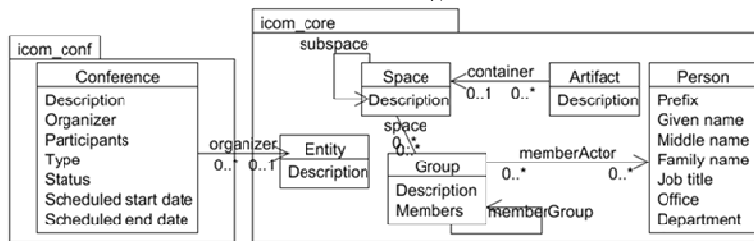
Fig. 2. Ejemplo de código para anotar texto en PSMW

Las plantillas son páginas wiki en el espacio de nombres Template cuyo contenido puede ser insertado en otra página. Una plantilla posibilita establecer la visualización

de los datos de una página y, además, permite convertir los datos en información semántica real ya que en ella es posible definir las propiedades de las páginas que la utilizan.

Los formularios permiten a los usuarios crear o editar páginas fácilmente. Se define un formulario por cada categoría y se asocia a una o más plantillas, las cuales permiten clasificar las páginas, que el formulario crea, bajo una Categoría. Las categorías son usadas como etiquetas para las páginas wiki e indican que una página pertenece a un tipo particular de página. Éstas son una característica de la wikimedia que permite el indexado automático de las páginas. Cada categoría establece un tipo de página diferente que se vincula con una plantilla y un formulario que permite la creación de sus correspondientes páginas.

Por su parte, la ontología ICOM, utilizada para definir la semántica de Ontolog PSMW, define un marco para la representación e integración de las actividades que se llevan a cabo en un entorno de colaboración [3]. Este vocabulario integra una amplia gama de actividades de colaboración, incluyendo y adaptando una serie de modelos que forman parte de normas y tecnologías existentes. ICOM posee una estructura modular para permitir la extensibilidad, a través del agregado de nuevos módulos. Los conceptos fundamentales y sus relaciones se incluyen en el núcleo (icom\_core), mientras que los conceptos y las relaciones específicas para cada área de actividades de colaboración se definen en los módulos de extensión separados (icom\_conf, por ejemplo). El subconjunto de los conceptos ICOM que son relevantes para la propuesta presentada en este artículo se muestran en la Fig. 3.



**Fig. 3.** Conceptos de ICOM utilizados en la propuesta

Una *Entidad* (Entity) es un objeto identificable que puede ser almacenado. Entity es la superclase de todas las otras clases definidas en ICOM (la relación superclase-subclase de Entity con los otros conceptos se ha omitido de la Fig. 3 para favorecer la claridad de la misma).

Un *Grupo* (Group) es una colección de individuos que comúnmente está asociado a un espacio de trabajo pero puede estar relacionado con más de uno (asociación *space* en la Fig. 3) y con cero o más subgrupos (relación *memberGroup* en la Fig. 3). Los individuos, Persona, (Person), miembros de un grupo, se vinculan a éste a través de la relación. *memberActor*.

Un *Espacio* (Space) es un ámbito que define un marco para que los actores trabajen o colaboren. *Espacio* la representación concreta de un área de trabajo para una colaboración y puede estar vinculado con un grupo y con otros espacios subsidiarios.

El concepto *Persona* (Person) representa a un individuo que participan en una colaboración. Una persona puede ser miembro de un grupo así como tener su propio espacio de trabajo.

El resultado de las diferentes actividades colaborativas llevadas a cabo en los distintos espacios se representa como un *Artefacto* (Artifact) que se vincula a un espacio a través de la asociación *container*. Un artefacto puede representar un documento o un conjunto de documentos relacionados que pueden estar incluidos en alguna página o almacenados en un repositorio.

Otro de los conceptos propuesto por ICOM es el de *Conferencia* (Conference), el cual representa una reunión que se lleva a cabo, de manera virtual o presencial, entre individuos. Una conferencia tiene un organizador, que puede ser un grupo o una persona, un conjunto de participantes, así como fechas de inicio y finalización tanto planificadas como reales.

La aplicación de estos conceptos al cuerpo de conocimiento relacionado con la cumbre de ontologías 2013 se muestra parcialmente en las Fig. 4 y 5. En particular, en la Fig. 4 se observa que la cumbre de ontologías del año 2013 se representa como un espacio de trabajo (*OntologySummit2013Space*) que está relacionado con el grupo (*OS2013Team*) que impulsa el desarrollo de las actividades de la cumbre. Este grupo, tiene a los coordinadores generales como miembros actores (relación *memberActor*) y a su vez, está relacionado (relación *memberGroup*) con los subgrupos que representan a los comités organizador (*OS2013OrgCommittee*) y asesor (*OS2013AdvCommittee*), a los patrocinadores (*OS2013Sponsors*) y a las instituciones organizadoras de la cumbre (*OS2013CoOrganizer*).

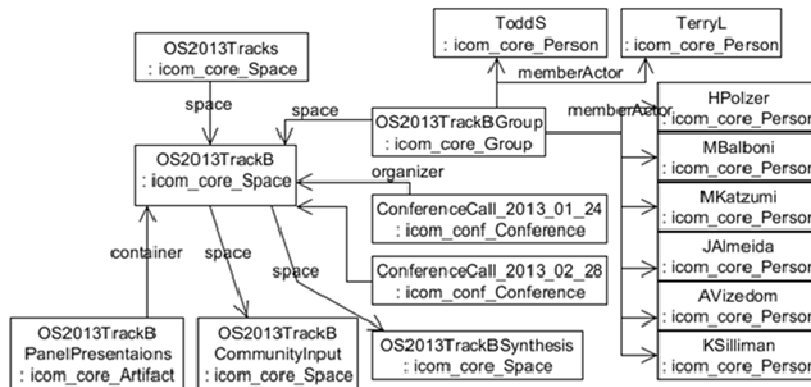
Por otra parte, la responsabilidad del espacio *OntologySummit2013* en la organización del simposio presencial, así como de las reuniones virtuales que no son específicas de un track, está representada también en la Fig. 4 mediante la relación *organizer* que se establece entre el espacio y las conferencias en cuestión. En la figura sólo se muestran dos de estas conferencias: *Phase2PhaseSymposium* y *ConferenceCall\_2013\_04\_25*. Asimismo, el comité organizador, tiene asociado un espacio de trabajo (*OS2013OrgCommittee*) y lleva adelante reuniones virtuales entres sus miembros, las cuales son representadas como conferencias (por ejemplo, *OrganizingCommitte meeting n.07-Fri 2013.03.15*).

Como se mencionó anteriormente, en la cumbre de ontologías se llevan adelante diferentes tipos de actividades, las cuales pueden clasificarse en espacios de discusión denominados tracks y espacios transversales de trabajo, en los que se busca la realización de un trabajo concreto que, preferentemente, involucre aspectos discutidos en todos los tracks. Este año, por ejemplo, se concretaron el desarrollo de una encuesta, la redacción del comunicado, la organización de la biblioteca digital, el desarrollo del sitio web, además de la organización de conferencias tipo hackathon. En todos los casos, las actividades, son representadas como un espacio (*icom\_core\_Space*) subsidiario de *OS2013Activities*. En la Fig. 4 se muestran los espacios creados para representar las actividades de este año: *SurveyDevelopment*, *Hackathon&Clinics*, *CommuniqueDevelopment*, *WebsiteDevelopment* y *CommunityLibraryDevelopment*. Como puede observarse en la Fig. 4 para *Hackathon&Clinics* y *CommuniqueDevelopment*, cada actividad: i) tiene un grupo



grabación del audio de la conferencia. Todos estos artefactos son almacenados en un repositorio al que los participantes de la cumbre y cualquier interesado tienen acceso.

Los conceptos ICOM que se presentaron en la Fig. 3 se mapearon a PSMW para poder ser utilizados para anotar las páginas wiki de Ontolog. En este mapeo se crearon: i) *propiedades* para representar las relaciones y los atributos, ii) *plantillas* y *categorías* para representar los conceptos ICOM, y iii) *formularios* que permiten la creación de las páginas. En las diferentes plantillas se incorporaron consultas acerca de las propiedades definidas en las páginas.



**Fig. 5.** Representación del Track B: Extrinsic Aspect of Ontology Evaluation

Para abordar el inconveniente de la complejidad del contenido y de la dificultad del acceso al mismo por parte de personas no familiarizadas con la estructura del repositorio, se propone incluir una capa de presentación sobre las páginas wiki que permita acceder a través de consultas y de sentencias de transclusión (incluir en una página wiki parte del contenido de otra página) al contenido que se pretende mostrar. En principio la tarea se focalizó solamente en el cuerpo de conocimiento asociado a la cumbre de ontologías de este año 2013. Sin embargo, la propuesta puede ser extendida sin mucho esfuerzo a cumbres de otros años y a otros eventos de Ontolog. Siguiendo la estrategia utilizada para el desarrollo del sitio web de la cumbre del año pasado, la propuesta divide el contenido de la cumbre en tres espacios: Actividades, Recursos y Entregables.

El espacio de Actividades presenta información sobre las diferentes actividades propuestas y posibilita el acceso a los respectivos espacios de trabajo. Los otros dos espacios, permiten a los usuarios acceder a la lista de las conferencias virtuales, la biblioteca digital, el archivo de la lista de discusión, en el espacio *Recursos* y a todos los artefactos que son resultado de las diferentes actividades que se llevan a cabo durante los tres meses de la cumbre en el espacio *Entregables*.

Las páginas que forman parte de esta capa no han sido migradas desde la vieja plataforma sino que se crearon utilizando las propiedades, plantillas, categorías y formularios definidos al mapear los conceptos de ICOM a PSMW. Además, con el fin de reutilizar el contenido existente en las páginas migradas, se hace uso de las



consultas que provee la extensión SMW y de la capacidad de transclusión provista por la wikimedia. El objetivo de la transclusión es obtener los últimos datos referenciales. En otras palabras, si se cambia la página de referencia, a continuación, la información incluida se actualizará automáticamente. PSMW soporta actualmente transclusión de toda una página, o de una sección de la página (a través de los números púrpura).

Con esta propuesta se ha incorporado semántica al cuerpo de conocimiento relacionado con la cumbres de ontologías 2013, lo cual permite realizar consultas como la planteada al final de la sección 2 de ese artículo: *¿Cuáles son las conferencias virtuales en las cuales ha participado una determinada persona?* La misma puede realizarse con la siguiente función #ask, en la cual se consulta por las conferencias en las que ha participado Marcela Vegetti:

```
{ {#ask:[[:Category:Icom conf Conference]]
[[[:com core participant::MarcelaVegetti]]
|?Icom core organizer
|?Icom core Description
|format=table }}
```

Esta consulta, podría ser incluida en el template *icom core Person* de manera que para cada página de esta categoría se muestre la lista de conferencias de las que participó. Para ello es necesario cambiar el nombre de la persona por la variable `{{PAGENAME}}` que almacena el nombre de la página donde la consulta está incluida. Sin embargo, también es posible ejecutarla desde `[Special:Ask]`. Esta página, cuya funcionalidad está implementada por la extensión SMW, permite realizar consultas definidas por los usuarios, más allá de las incluidas en las páginas wiki. En la Fig. 6 se muestra los resultados obtenidos al ejecutar la consulta mostrada desde esta página especial.



Fig. 6. Resultados de la consulta acerca de conferencias a las que asistió una persona dada

Es importante resaltar, que una vez creados los datos en una wikimedia semántica, éstos no necesitan mantenerse dentro de la wiki; pueden ser fácilmente exportados a otros formatos como archivos separados por comas, JSON y RDF. Esto posibilita que una wiki semántica sirva como fuente de datos de otras aplicaciones. En particular, utilizando la página especial `ontolog-02. cim3.net/wiki/Special:ExportRDF` es posible

exportar los datos de una o varias página como tripletas RDF, con lo cual el contenido de la cumbre de ontologías 2013 es ahora legible y descubrible por máquinas. Impulsando con esto el movimiento del contenido de ontologías PSMW hacia la obtención de las 5 estrellas que según Berners-Lee deben tener los datos para pertenecer a LinkedData [4].

#### **4 Conclusiones y Trabajos Futuros**

Ontolog es una comunidad abierta virtual que busca impulsar avances en el campo de la ingeniería ontológica y las tecnologías semánticas. Todas las contribuciones realizadas por los miembros de esta comunidad en las diferentes actividades que se proponen son almacenadas en una plataforma wiki que no tiene implementadas ninguna de las tecnologías semánticas que este foro trata de promover.

Otra de las actividades que se llevan a cabo en el marco de Ontolog, es el desarrollo de la plataforma PSMW, a la cual se está migrando el repositorio de Ontolog. El cambio de plataforma, sólo provee la capacidad de añadir semántica, lo cual debe ser complementado con la implementación de una ontología en la plataforma PSMW, para incorporar efectivamente semántica. En este artículo se propone la utilización de la ontología ICOM para representar la estructura y el contenido de la cumbre de ontologías 2013 y capturar el contenido de acuerdo a este vocabulario común. Adicionalmente, se propone la reorganización del contenido de la cumbre de manera que fomente y facilite el acceso y la reutilización del material generado.

La implementación se está llevando a cabo a nivel de prototipo. Una vez validados los resultados de esta propuesta, se procederá a extender la aplicación de la ontología para la representación de otras actividades de Ontolog.

#### **Agradecimientos**

Este trabajo ha sido financiado en forma conjunta por CONICET, la ANPCyT (PAE-PICT2315 y PIP 112-200801-02754) y la Universidad Tecnológica Nacional (PID 25/O156 y PID 25/O144). La comunidad Ontolog ha brindado su infraestructura para permitir la realización de este trabajo. Se agradece el apoyo brindado por estas instituciones, así como la colaboración de Tejas Parikh, Ken Baclawski, Ali Hashemi, Peter Yim y Soledad Sonzini para llevar adelante esta propuesta.

#### **Referencias**

- [1] Krötzsch, M., D. Vrandečić, M. Völkel, H. Haller, R. Studer (2007). Semantic Wikipedia. Web Semantics,5,
- [2] Baclawski, K., V. Gupta, T. Parikh, P. Yim, J. Cheyer. (2008). Purple MediaWiki: Fine-Grained Addressability of Wiki Content. Available at: [http://project.cim3.net/wiki/PMWX\\_White\\_Paper\\_2008](http://project.cim3.net/wiki/PMWX_White_Paper_2008)
- [3] OASIS ICOM. (2013). Integrated Collaboration Object Model (ICOM) for Interoperable Collaboration Services Version 1.0. 31. OASIS Committee Specification 01. Disponible on-line en: <http://docs.oasis-open.org/icom/icom-ics/v1.0/cs01/icom-ics-v1.0-cs01.html>
- [4] Tim Berners-Lee. (2009). Linked Data. Disponible en línea en: <http://www.w3.org/DesignIssues/LinkedData.html>