

---

# PROPUESTA DE UN VOCABULARIO DE MICRODATOS BASADO EN SCHEMA.ORG PARA LA DESCRIPCIÓN SEMÁNTICA DE PUBLICACIONES CIENTÍFICAS Y LA MEJORA DE SU VISIBILIDAD EN ENTORNOS ABIERTOS DEL CONOCIMIENTO.

*Proposal of a microdata vocabulary based on Schema.org for semantic description of scientific publications and improvement of their visibility on open environments of knowledge*

---

**María Ester Rubio Lucas**

Universidad de Murcia – mariaester.rubio@um.es

## Resumen

La web hace ya tiempo que es el medio principal de comunicación científica y los contenidos de las publicaciones científicas no pueden quedar al margen de esta tendencia y deben aprovecharse también de sus ventajas y deben adentrarse en el campo de la descripción semántica promovido por Shotton quien introduce el concepto de "Publicación Semántica". Este concepto engloba las pautas y recomendaciones para el enriquecimiento semántico, así como la propuesta de diferentes ontologías con las que describir publicaciones científicas y recursos bibliográficos. Nuestro trabajo tiene como objetivo proponer un vocabulario de microdatos basado en Schema.org que permita identificar y describir publicaciones científicas. Para elaborar la propuesta se han seguido las pautas marcadas por la comunidad Schema.org (promovida por Google, Yahoo, Bing! y Yandex) y se ha confeccionado un listado de propiedades y términos (básicos y opcionales) que permiten la descripción semántica de los artículos, revistas, autores, estructura de artículos y referencias bibliográficas. Como herramienta de descripción hemos utilizado microdatos debido a su gran auge y facilidad de implementación en la web y a que su uso favorece que los motores de búsqueda puedan inferir información y ser capaces de establecer relaciones entre documentos. El vocabulario propuesto facilita a los motores de búsqueda la tarea de recuperar documentos relevantes proporcionando mayor efectividad a las búsquedas. Además, la propuesta permitirá la descripción semántica integral de publicaciones científicas, abarcando desde aspectos formales como la estructura del documento a la justificación de las referencias bibliográficas contenidas.

**Palabras clave:** Ciencia 2.0; Comunicación Científica; Edición de Revistas Científicas; Microdatos, Publicación Semántica; Schema.org

## Abstract

The web is since long the main channel for scientific communication and scientific publication contents cannot be left out of this trend, and must profit also from the advantages and break into the semantic description field promoted by Shotton, who introduces the concept of "Semantic Publication". This concept encompasses patterns and recommendations for the semantic enrichment, and a proposal of different ontologies that may be used to describe scientific publications and bibliographic resources. Our work has the objective of proposing a microdata vocabulary based on Schema.org that allows identifying and describing scientific publications. In order to elaborate the proposal, the guidelines marked by the Schema.org community (endorsed by Google, Yahoo, Bing and Yandex) have been followed and a list of properties and terms (basic and optional) that allow the semantic description of articles, journals, authors, article structure and bibliographic references has been drafted. As a description tool we have used microdata due to its great rise and ease of implementation that favors search engines to infer information and be able to establish semantic relationships between documents. The proposed vocabulary eases the task of retrieving relevant documents to search engines, allowing more efficiency to searches. Furthermore, the proposal will permit the integral semantic description of scientific publications, encompassing formal aspects, document structure and the justification of the bibliographic references it contains.

**Keywords:** Science 2.0; Scientific Communication; Scientific Journals Edition; Microdata; Semantic Publishing; Schema.org.

## 1. Introducción

Actualmente, a pesar de los grandes avances introducidos en la gestión del proceso editorial principalmente por el uso del gestor OJS (1), la difusión de las revistas científicas (y por tanto, de su contenido) sigue en algunos casos anclada en los tradicionales procesos editoriales presentes en el formato en papel. El proceso de publicación no ha sufrido grandes cambios editoriales. Con la generalización del uso de internet para la difusión de información en formato

digital, la publicación de las revistas cambió el formato y el medio, no habiendo llegado aún a aprovecharse del todo de las inherentes ventajas de ese medio, más allá de las ventajas económicas y de inmediatez en el acceso (más si cabe cuando la publicación es de acceso abierto). De hecho, una gran parte de las revistas científicas comenzaron a editarse solo en digital después de un periodo de transición entre el papel y el formato digital.

“La edición de revistas científicas en muchos casos aún tiene pendiente abordar la revolución digital con todas sus consecuencias. Hace falta “pensar en digital”, abandonar toda dependencia de la versión impresa, que puede mantenerse si es necesario como estrategia de distribución, pero que no debe frenar la explotación y difusión a través de internet”. (Rodríguez-Yunta, 2013, p.207)

Existe una gran variedad de formatos de publicación en la web, siendo los más predominantes el formato PDF y el lenguaje HTML para la publicación de revistas electrónicas. Hay algunos casos excepcionales, como la Revista Española de Documentación Científica (2) que permite consultar los últimos números en formato XML, aportando una mayor interoperabilidad y facilidad para ser exportados sus contenidos. Hasta ahora, las tecnologías de la información no han sido aprovechadas del todo para mejorar el posicionamiento de los estos artículos científicos dentro de las búsquedas que se llevan a cabo en los motores de búsqueda generalistas (Google, Bing, Yahoo, etc.) o bien de búsqueda académica que emplean similares criterios de búsqueda sobre colecciones más específicas (Google Scholar o Microsoft Academic Search), sistemas de búsqueda que constituyen una fuente de información de peso considerable entre los estudiantes e investigadores (especialmente en el caso de los primeros) y entre los investigadores que tienen limitado su acceso a las publicaciones de pago de las editoriales científicas (cada vez más caras y menos accesibles para aquellos países en vías de desarrollo).

Como ocurre con otros tipos de sitios y contenidos de la web, un buen posicionamiento repercutiría directamente en su visibilidad y, por tanto, en el impacto y difusión del trabajo científico desarrollado. Internet ha supuesto una gran facilidad de publicación y la reducción del presupuesto necesario para la edición de revistas científicas. Tanto en la web como en los repositorios hay una gran cantidad de documentación científica que en ocasiones resulta difícil recuperar. La ingente cantidad de información hace más complicada la recuperación precisa y relevante a una determinada consulta o petición de información.

En este campo concreto, la propuesta de Web Semántica elaborada por Berners-Lee et al. (2001) y la posibilidad de describir semánticamente la información en la web (Berners-Lee and Hendler, 2001) abre un mundo lleno de posibilidades para la publicación y recuperación de información en general, que la edición de revistas científicas puede y debería aprovechar. Es

por ello que dotar de significado semántico a la información científica disponible en la web es actualmente más necesario que nunca. La descripción o marcado de estos recursos de información de manera semántica abre un nuevo espacio lleno de posibilidades a la hora de facilitar la recuperación de información de una manera más precisa y relevante. Así como, favorece la interoperabilidad y la reutilización de información por parte de las máquinas y/o motores de búsqueda. Para ello, el marcado semántico se sirve de propiedades, clases, relaciones o etiquetas de marcado que haga referencia a un vocabulario concreto (en el caso de esta propuesta, Schema.org) y a una sintaxis, como microdatos, microformatos o RDFa, etc.)

El marcado semántico de las publicaciones no solo se limita a la identificación de revistas o autores, Rodríguez-Yunta (2013) apunta una serie de posibles elementos a identificar semánticamente: tablas, imágenes, gráficos, figuras, fotografías, o anexos de los artículos; considerando que “el artículo no es la única unidad de información objeto de una posible recuperación documental”. (Rodríguez-Yunta, 2013, p. 208).

Un ejemplo de esta idea lo encontramos en el vocabulario basado en Schema.org propuesto por Gaurav, et al (2015), para la descripción de contenidos educativos en la web. En ella se pretende aumentar la visibilidad y los resultados de los contenidos educativos publicados en la web en las búsquedas a realizar en los motores de búsqueda, para ello se enriquecen los portales con microdatos utilizando las propiedades y clases disponibles en Schema.org y además, proponen nuevos términos para enriquecer la descripción. Con el vocabulario propuesto pretenden solucionar dos problemas principales: reducir el ruido y los resultados irrelevantes en las búsquedas y facilitar los enlaces a las webs con contenidos educativos.

Las mejoras que aporte la Web Semántica al mundo de las revistas científicas irán en la línea de introducir un marcado (o etiquetado) semántico en las publicaciones, ampliando la idea de publicar digitalmente a un concepto más amplio que se ha dado en llamar Publicación Semántica (del inglés *Semantic Publishing*), concepto en torno al cual se concentra, en los últimos años, el trabajo de David Shotton (3), principal autor sobre este campo. El autor (2009, p.2) define este concepto como todo aquello realizado para mejorar el significado de los artículos de las revistas científicas, facilitando su descubrimiento, habilitando su enlazado semántico con artículos relacionados, proporcionando acceso a los datos dentro del artículo en forma procesable y/o facilitando la integración de datos entre

artículos. Esto implica principalmente enriquecer el artículo científico con metadatos que puedan ser procesados y analizados automáticamente, lo que mejora la verificabilidad de la información publicada, proporcionando la capacidad de descubrimiento de manera automatizada. La utilización del marcado semántico incrementa el valor intrínseco de los artículos, aumentando implícitamente el valor de la información, haciéndola más comprensible y fácil de recuperar por las personas y por las máquinas. Otro aspecto importante es que no solo se potencia la recuperación de información por parte de las máquinas, sino que se aporta a los lectores una mayor profundidad informativa, y un acceso más completo y seguro a la información.

Para Shotton, el desarrollo e implementación del enriquecimiento semántico de las publicaciones ayuda a crear nuevas oportunidades de negocio en forma de valor añadido en los servicios de publicación, algo que si bien no parece ser el objetivo primigenio de un editor de revistas científicas, puede convertirse en una ventaja competitiva. Además, los lectores y autores se podrán beneficiar de un acceso más rápido y mejor a los datos relativos a las publicaciones.

Del conjunto de herramientas disponibles para la descripción semántica de información en la web (RDF, OWL, etc.), optamos por la utilización de los microdatos para la realización de nuestra propuesta. Los microdatos son una forma de etiquetar contenido para describir estructuras de información específicas (personas, eventos, calendarios, etc.) en la web. Han sido creados en el contexto de la especificación HTML5 (4) descrita por la W3C (5), para asignar nombres cortos y descriptivos a elementos y propiedades (en epígrafes posteriores de este trabajo se presentan de forma más detallada). Además, pueden aplicarse en XHTML, lo que resulta de gran importancia para favorecer la interoperabilidad entre tecnologías de la web semántica como OWL o RDF. Usar microdatos persigue posibilitar a los robots de los motores de búsqueda el procesamiento y la extracción de información semántica de las páginas web analizadas para poder responder y mostrar los resultados de manera más precisa a las necesidades de información planteadas por los usuarios en la ecuación de búsqueda.

Hay otras alternativas para aportar marcado semántico, como es el caso de los microformatos. Consisten en un tipo de metadatos creados para el marcado de páginas web escritas en HTML. Ayudan a los sistemas de información a identificar estructuras de información relevantes en esas páginas con la consulta formulada. No se trata de un nuevo lenguaje de marcado, sino

de una forma o método de marcado a través de las etiquetas HTML de un tipo concreto de información (opiniones, eventos, productos, personas, etc.). En general, los microformatos hacen uso del atributo *class* en las etiquetas HTML (suelen ser las etiquetas `<span>` o `<div>`), para asignar nombres breves y descriptivos a las entidades y a sus propiedades. Los atributos *class* se utilizan para crear clases y englobar dentro determinados elementos. Lo que se consigue creando una "clase" es que todos los elementos de esa clase tengan los mismos valores o estilos que se le atribuyan a esa clase. Las etiquetas `<div>` se utilizan para crear divisiones dentro de un documento HTML para diferenciar bloques de contenido a los que les afecte operaciones dirigidas hacia ese bloque específico. Las etiquetas `<span>` también sirven para agrupar texto al que se le desee aplicar otro estilo diferente al del párrafo.

La razón de elegir los microdatos y el vocabulario Schema.org se basa en dos razones: su sencillez y el gran auge que están teniendo en los últimos años (seguramente debido a esa sencillez). De entre los vocabularios disponibles elegimos Schema.org por su sencillez, interoperabilidad y un futuro previsiblemente prometedor gracias al apoyo de los principales buscadores web a esta iniciativa. Estamos convencidos de que la utilización de este vocabulario permite ampliar el horizonte de la edición y publicación de revistas científicas y acercar los procesos editoriales a la era digital modernizándolos, permitiendo así una mayor interoperabilidad semántica y funcional entre las editoriales y facilitando las tareas de recuperación y el descubrimiento de información por parte de los lectores e investigadores.

El objetivo principal del presente trabajo es realizar una propuesta de etiquetado semántico de publicaciones científicas para una mejor descripción de las publicaciones científicas en la web.

El presente trabajo sigue la siguiente estructura: (1) introducción en donde se expone el contexto por el que surge la idea del trabajo, (2) descripción de la metodología llevada a cabo para su elaboración, (3) descripción de herramientas para el marcado semántico en la web, (4) descripción del concepto de publicación semántica de Shotton, (5) técnicas aplicadas a la descripción de revistas, (6) propuesta del vocabulario y (7) conclusiones.

## 2. Metodología

Para la elaboración del estado de la cuestión se han consultado diversas fuentes científicas (en

español y en inglés), entre las que se encuentran: Google Scholar, Web of Science, Scopus, Dialnet, E-Lis, Catálogo ALBA de la Biblioteca Universitaria de la Universidad de Murcia, búsquedas en el buscador Google genérico, Wikipedia, etc.

Se ha realizado una extensa búsqueda de información en estas fuentes para conocer el estado de la cuestión de los microdatos y para conocer en profundidad sus capacidades de descripción de recursos digitales en la web, y la utilización de Schema.org.

En la búsqueda bibliográfica se ha observado una gran escasez de desarrollos previos en el campo de la comunicación científica que creemos que refuerza la pertinencia de nuestro trabajo: llevar a cabo una propuesta de esquema de microdatos orientados a la publicación y difusión de las publicaciones científicas, en definitiva, una propuesta de publicación semántica apoyada en microdatos.

Para obtener una primera aproximación de la propuesta del vocabulario se han tomado como ejemplos vocabularios publicados en Schema.org (como el vocabulario Article Periodical, etc.) u ontologías como FaBIO o CiTO y. Posteriormente, para su elaboración se ha redactado una lista de posibles propiedades a incluir en el vocabulario y se ha evaluado su necesidad en básicas u optativas según su importancia dentro del artículo. Además, se han englobado en diferentes categorías dependiendo de su propósito o funcionalidad en:

1. Descripción de revistas: términos o propiedades empleadas para la identificación formal y de contenido de las publicaciones periódicas.
2. Descripción de artículos: propiedades para la descripción de los artículos de revistas. Se trata de identificadores básicos que permitan localizar y recuperar aquellos artículos referenciados.
3. Identificación de partes del documento: Propiedades para la descripción estructural y formal del artículo científico. Pretenden ser una guía de localización para conocer dónde se encuentra la cita, los datos del autor, etc.
4. Justificación y propósito de las citas: propiedades que permitan la justificación, argumentación, contextualización, definición de propósitos, etc. de la cita incluida en el texto.

Tras la clasificación del primer borrador de las propiedades a incluir en la propuesta, se han

consultado los vocabularios publicados en Schema.org en busca de propiedades o relaciones que cubran la misma funcionalidad de las expresadas en el borrador con el fin de no crear propiedades, clases o relaciones ya existentes y proponer duplicados. Por ejemplo: la clase Article ya está definida dentro del corpus de Schema.org, por lo que la propuesta de la creación de una nueva clase Article, se ha desestimado, otro ejemplo es el caso de la propiedad URL que ha sido reutilizada en numerosas ocasiones dentro del vocabulario propuesto con el fin de no crear diferentes propiedades para la URL que ofrezcan la misma funcionalidad. Sin embargo, para aquellas propiedades o relaciones que Schema.org no haya establecido, se propone la creación de dicha propiedad o relación.

### 3. Microdatos y otros formatos de enriquecimiento semántico

Con el objeto de compartir un vocabulario y sintaxis común, los motores de búsqueda Google, Bing, Yahoo! y Yandex han colaborado para apoyar la propuesta Schema.org que pretende desarrollar la extensión de un vocabulario común basado en microdatos que los principales motores de búsqueda sean capaces de entender y de procesar la información semánticamente descrita:

“Muchos sitios han sido generados con datos estructurados, los cuales están almacenados en bases de datos. Muchas aplicaciones, especialmente los motores de búsqueda, pueden beneficiarse enormemente del acceso directo a estos datos estructurados”. (Schema.org, 2014a)

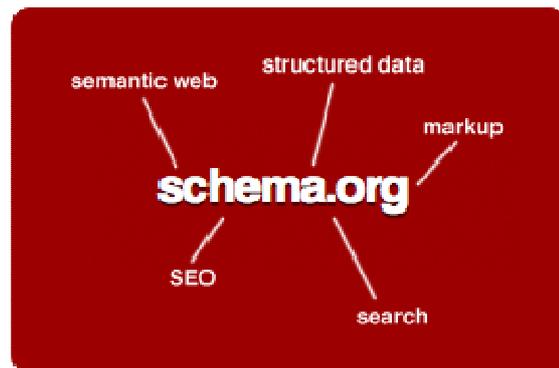


Figura I. Logo de Schema.org rodeado de los principales elementos que describen su vocabulario. Fuente: <http://www.analizosuweb.com>

Los microdatos son una forma de etiquetar contenido para describir estructuras de información específicas (personas, eventos, calendarios, etc.) en la web. Utilizan la especificación HTML5 (19) (basada en RDFa) descrita por la W3C (20) para asignar nombres breves y descriptivos a

elementos y propiedades. Su objetivo fundamental es posibilitar a los motores de búsqueda, robots y navegadores el procesamiento y la extracción de información semántica de la web para poder responder y mostrar los resultados de manera más precisa a las necesidades de información de los usuarios (ver Imagen 2 en Anexo 5).

Para el etiquetado se utilizan diferentes etiquetas como la etiqueta *itemscope* la cual indica que el contenido incluido en la división (<div>) es un elemento del vocabulario "http://data-vocabulary.org/Person" y que por lo tanto es un elemento de una descripción de persona. Las propiedades para describir el elemento de persona se pueden definir con el atributo *itemprop*, como por ejemplo: *itemprop="title"* para indicar que se trata de un ingeniero.

Con el objetivo de compartir vocabulario y sintaxis común, Google, Bing, Yahoo! y Yandex colaboraron en el desarrollo de esta propuesta.

"Muchos sitios han sido generados con datos estructurados, los cuales están almacenados en bases de datos. Muchas aplicaciones, especialmente los motores de búsqueda, pueden beneficiarse enormemente del acceso directo a estos datos estructurados." (Schema.org, 2014a)

A la pregunta de por qué estos buscadores han optado por los microdatos y el vocabulario Schema.org y no han optado por utilizar RDFa o microformatos, desde el blog de Google argumentan lo siguiente:

"Existen argumentos para preferir cualquiera de los estándares existentes, pero creemos que los microdatos ofrecen el equilibrio necesario entre la posibilidad de ampliación de RDFa y la sencillez de los microformatos, así que ese es el formato por el que hemos optado." (Google, 2014)

Hasta la fecha, Schema.org no se ha constituido como un vocabulario normalizado, pero probablemente solo sea una cuestión de tiempo que así lo sea. Ronallo (2012), apunta que es poco realista para Google, Bing y Yahoo! apoyar a todos los vocabularios que se utilizan hoy en día, sin embargo, con Schema.org pretenden definir un vocabulario amplio, compartido e in-crustado la web de forma definitiva.

Schema.org se basa en un vocabulario jerárquico en donde su ítem superior es 'Thing', compuesto por cuatro propiedades que serán heredadas por todos los ítems del vocabulario: 'name', 'description', 'url', e 'image'. Actualmente existen diversos vocabularios ya definidos y publicados en Schema.org, y están disponibles

para que los creadores de sitios web los implementen en sus códigos. En la sede web de Schema.org (21) se puede visualizar y consultar todos los vocabularios publicados.

Dentro de la comunidad de la W3C existen grupos especializados de debate y colaboración sobre un tema concreto. Sobre Schema.org se ha creado el 'Schema.org Community Group (22)', donde se propone un foro de debate y discusión sobre los cambios realizados en Schema.org. Además, podemos encontrar el grupo 'Schema BIB Extend Community Group' (23) donde se discuten y preparan propuestas para la extensión de Schema.org a través de nuevas propuestas de esquemas que permitan la representación de información bibliográfica. Este grupo mantiene además una wiki (24) con ejemplos ilustrativos, guías para la utilización del esquema e información de seguimiento de las reuniones del grupo. Este grupo y sus futuras publicaciones resultan de gran importancia para el desarrollo de un vocabulario para la descripción semántica de publicaciones periódicas.

García-Marco (2013) considera que la utilización de Schema.org responde a una reinención de la catalogación tradicional de las webs:

"Schema.org supone una iniciativa y un avance importante en dos grandes direcciones: por un lado, la incorporación de la WWW a la normalidad en el ámbito de recuperación de información; y por el otro, la democratización de la Web Semántica o, si se quiere, el lanzamiento de una Web Semántica fácilmente incorporable por cualquier webmaster [...]"

Los metadatos utilizados por Schema.org introducidos con HTML5 son los siguientes: (1) *Itemscope*: para crear un nuevo ítem, (2) *Itemtype*: para especificar un tipo de ítem Y (3) *Itemprop*: para proporcionar propiedades al ítem (ver Imagen 3 en anexo 5).

Otro de los vocabularios utilizado en la web para la descripción semántica es RDFa (25), una extensión de HTML5 y XHTML (26) recomendada por el W3C para poder enriquecer contenidos en la web referentes a personas, eventos, recomendaciones, etc. Su objetivo es facilitar a los buscadores el listado de sitios web que responde a la consulta planteada y favorecer la visibilidad de los contenidos en la web.

A diferencia de Schema.org, un vocabulario fácil de implementar, RDFa es un vocabulario complejo (a la par que potente) que necesita de conocimientos avanzados de RDF para poder ser utilizado. Quizás sea ésta la principal diferencia

entre RDFa y Schema.org, tal como recogen Rosati y Mayernik (2013) como conclusión de su comparativa de ambos vocabularios para evaluar la capacidad de descubrimiento y conectividad de los recursos en la web: “por encima de todo, la principal diferencia entre RDF y los microdatos es la facilidad de uso vs la potencia, la extensibilidad y la complejidad”.

Sin embargo, algunos autores como ManuSporny (2011), acusan a Google, Yahoo!, Bing y Yandex de imponer un vocabulario que no es tan bueno como puede parecer y defienden el uso de microformatos y RDFa al mismo nivel que el uso de Schema.org. Este autor expone que la utilización de un vocabulario u otro no debería ser impuesta por ningún buscador, dejando esa elección a los webmasters.

Además, cabe destacar que RDF es un vocabulario de propósito general, mientras que Schema.org (a pesar de su gran extensión en los últimos años) su dominio de aplicación es más reducido.

Otra comparativa realizada entre RDF y Schema.org, en este caso RDFa Lite 1.1 (27) es la realizada por Bernardis (2014) en donde analiza el impacto del SEO semántico en los motores de búsqueda aplicando ambos vocabularios. A pesar de las limitaciones del experimento (falta de tiempo, bajo nivel de contenido o escaso PageRank de las páginas objeto del estudio), caben destacar las conclusiones de esta autora:

Las webs enriquecidas con datos estructurados obtienen mejor posicionamiento que las que solo han sido optimizadas con SEO. La utilización de Schema.org favorece el posicionamiento en un plazo relativamente corto. La utilización de RDFa Lite 1.1 no favorece el posicionamiento.

RDFa Lite (28) es una simplificación de RDFa con las siguientes propiedades propuestas: (1) vocab: indica el vocabulario utilizado, (2) typeof: tipo de objeto sobre el que se realiza la descripción, (3) property: propiedades del elemento descrito, (4) resource: para expresar relación con otros recursos, (5) prefix: para definir nuevos vocabularios.

Los 'rich snippets' (fragmentos enriquecidos), son pequeñas piezas de información que aparecen de manera personalizada en los resultados de las búsquedas de los motores de búsqueda. Sirven para dotar al usuario de cierta información antes de decidir si entrar al enlace ofrecido por el buscador. El uso de estos fragmentos de información sirve para personalizar la presentación que el buscador va a hacer del sitio web localizado. *Imagen 4. Ver anexo 5.*

La Ayuda de Herramientas para webmasters de Google (29) recomienda tres pasos para enriquecer estos fragmentos: (1) elegir un formato de marcado: microdatos, microformatos o RDFa. La opción recomendada por Google es la de los microdatos, (2) marcar el contenido. Google es capaz de identificar fragmentos enriquecidos de opiniones, personas, productos, empresas y organizaciones, recetas, eventos y música, (3) probar el marcado. Google pone a disposición de usuarios y webmasters una herramienta de pruebas de datos estructurados (30).

Los microformatos son un tipo de metadatos creados para el marcado semántico de sitios web en HTML. Ayudan a los sistemas de información a identificar estructuras de información relevantes en una página web. No se trata de un nuevo lenguaje de marcado, sino de una forma o método de marcado a través de etiquetas HTML para un tipo concreto de información (opiniones, eventos, productos, personas, etc.):

“En general, los microformatos hacen uso del atributo class en las etiquetas HTML (que suelen ser <span> o <div>) para asignar nombres breves y descriptivos a las entidades y a sus propiedades” (Google, 2014b).

A continuación mostramos un ejemplo de un bloque HTML breve con la información de contacto básica de Roberto Sánchez.” (Google, 2014b) (ver Imagen 5 en anexo 5).

Mientras que con el marcado con el microformato *hCard* para describir personas (31):

```

<div class="vcard">

  <strong class="fn">Roberto Sánchez</strong>

  <span class="title">Editor principal</span> de <span class="org">Ediciones AQME</span>

  <span class="adr">
    <span class="street-address">Calle Mayor 4</span>

    <span class="locality">Alcalá de Henares</span>, <span class="region">Madrid</span>

    <span class="postal-code">28801</span>

  </span>
</div>

```

Figura 6. *Ejemplo de código.* Fuente: <https://support.google.com/webmasters/answer/146897?hl=es>

La etiqueta “vcard” indica que el código incluido en la etiqueta corresponde a la descripción de

una persona. Con `class="photo"` se indica una imagen relacionada con la persona. A través de `class="fn"` especifica el nombre de la persona descrita. `Class="title"` especifica el cargo de la persona en una organización (`class="org"`). Finalmente, crea una nueva división (`<span>`) para incluir los datos de la dirección postal: `class="adr"` que estará compuesto por la calle: `"street-address"`, la localidad: `"locality"`, la región `"región"` y el código postal: `"postal-code"`.

Los principios definidos para su implementación definidos en la wiki 'microformats.org' (32) son los siguientes: (1) resolver un problema específico, (2) comenzar de la manera más fácil posible, resolver los problemas más fáciles primero y a partir de ahí seguir mejorando, (3) los microformatos están diseñados para humanos en primer lugar, en segundo lugar para las máquinas, (4) reutilizar bloques de microformatos ya utilizados y ampliamente adoptados, (5) diseñar microformatos de manera que puedan ser reutilizables y embebidos, y (6) habilitar y facilitar la descentralización y distribución del desarrollo, los contenidos y los servicios.

Para comprobar si el marcado se ha realizado correctamente, Google habilita una herramienta web de pruebas de datos estructurados.

#### 4. El concepto de Publicación Semántica

Se ha citado anteriormente al profesor Shotton como el autor de referencia sobre esta nueva concepción de enriquecimiento semántico en las publicaciones científicas. El autor (2009) define a la Publicación Semántica como todo aquello llevado a cabo para mejorar el significado de los artículos de las revistas científicas, facilitar su descubrimiento, habilitar su enlazado semántico con artículos relacionados, proporcionar acceso a los datos dentro del artículo en forma procesable y/o facilitar la integración de datos entre distintos artículos. Principalmente aplica el enriquecimiento del artículo científico con metadatos para su proceso y análisis automático. Esto permite mejorar la verificabilidad de la información publicada, proporcionando la capacidad de descubrimiento de manera automatizada por parte de los motores de búsqueda. El uso del marcado semántico incrementa el valor intrínseco de los artículos aumentando el valor de la información contenida en los mismos, haciéndola más comprensible y fácil de recuperar por las personas y por las máquinas. Otro aspecto importante a destacar es que no solo se mejora la recuperación de información por los motores de búsqueda, sino que se aporta a los lectores una mayor profundidad informativa, y un acceso más completo y seguro a la información.

Shotton (2009) establece 6 reglas para el publicado semántico:

1. Comenzar de manera sencilla e ir implementando incrementalmente mejoras.
2. Esperar grandes cosas de los autores.
3. Explotar las habilidades existentes de las que se disponen en casa.
4. Utilizar estándares siempre que sea posible.
5. Publicar en la web el conjunto de datos "en crudo".
6. Publicar metadatos de los artículos, particularmente listas de referencias en un lenguaje capaz de ser leído por los sistemas de búsqueda automatizados.

Para este autor, el desarrollo e implementación del enriquecimiento semántico de las publicaciones puede ayudar a crear nuevas oportunidades de negocio en forma de valor añadido en los servicios de publicación. Además, los lectores y autores se podrían beneficiar de un acceso más rápido y mejor a los datos relativos a las publicaciones y mejorar sus búsquedas.

Tal y como apunta el mismo Shotton et al. (2009), los pioneros en este campo fueron la Royal Society of Chemistry, la cual dispone de algunas revistas con marcado semántico en HTML, como es el caso de la publicación *Molecular Biosystems*. Este marcado introduce información adicional e hipervínculos a nombres químicos, o términos de la International Union of Pure and Applied Chemistry Compendium of Chemical Terminology (34) (The Gold Book) y términos definidos en la Gene Ontology (35), Cell Ontology (36) y en Sequence Ontology(37). Otro ejemplo de similares características es el marcado semántico que realiza la revista *Acta Crystallographica A: Foundations of Crystallography* (publicada por la International Union of Crystallography), con su marcado de términos con enlace al citado The Gold Book y al IUCr Online Dictionary of Crystallography(38).

Shotton et al. (2009) escogieron el trabajo de Reis (2008) sobre el riesgo de contraer leptospirosis por parte de los habitantes de las favelas en la población de Pau da Lima (Brasil) y que ya estaba en formato XML, para poner en práctica la publicación semántica. El enriquecimiento semántico de este artículo se llevó a cabo aplicando de forma sucesiva las siguientes acciones:

1. Hacer accesibles y procesables los datos sin editar o brutos contenidos en el artículo. Normalmente los datos conte-

- nidos en figuras, tablas o imágenes pueden ser descargados tan solo en el formato de imagen, siendo imposible su manipulación o procesamiento. Una de las mayores ventajas de publicar los datos es la posibilidad de ser procesados y presentados de manera diferente.
2. Fusión simple de datos geoespaciales (Google Maps o similar, imagen 7, ver anexo 5).
  3. Fusión de datos geoespaciales a través de diferentes publicaciones.
  4. Mapeo de las localizaciones de los diferentes estudios de leptospirosis en el espacio y tiempo.
  5. Fusión de datos serológicos entre diferentes publicaciones.
  6. Resaltado de términos claves. Esta acción permite hacer una lectura rápida por los términos que más y mejor describan el documento, viene a ser como resaltar parte del texto de unos apuntes con rotuladores pero usando tecnología semántica (imagen 8, ver anexo 5).
  7. Enlazar nombres de entidades a fuentes de información externas, lo que permite ampliar información sobre las instituciones mencionadas.
  8. Incluir citas en contexto a través del sistema Supporting Claims Tooltip que permite mostrar información sobre las citas a través de una caja de texto o ventana emergente que aparece cuando el lector mueve el cursor del ratón a la parte del texto en donde está incrustado (imagen 9, ver anexo 5).
  9. Proveer de un sumario al documento, de una nube de etiquetas y/o de un árbol de etiquetas.
  10. Enlace a una ontología sobre enfermedades infecciosas.
  11. Estadísticas del documento.
  12. Análisis de las citas.
  13. Tipificación de las citas utilizando CiTO (Citation Typing Ontology).
  14. Inclusión del resumen en un idioma alternativo.
  15. Añadir información sobre la procedencia del artículo, más allá del nombre del autor y el título.
  16. Insertar figuras interactivas.
  17. Opción para reordenar el listado de referencias bibliográficas (por ejemplo: en orden alfabético, año de publicación, frecuencia en el texto o por orden numérico).
  18. Enlace directo a los recursos citados y a sitios externos al artículo.

19. Incorporación de metadatos embebidos procesables por las máquinas (RDFa).

Como se puede observar en esta serie de acciones, no se crea nada de la nada sino que se incorpora a las publicaciones un conjunto de herramientas y recursos ya disponibles en la web. Son pequeños cambios que modifican la publicación electrónica de las revistas científicas, tal y como la conocemos actualmente.

Peroni y Shotton (2012), dentro del marco de la publicación semántica, han desarrollado dos ontologías para la descripción de recursos bibliográficos y de citas llamadas FaBiO, la FRBR-aligned Bibliographic Ontology y CiTO, la Citation Typing Ontology. FaBiO está basado en Dublin Core, PRISM (39), FRBR(40) y SKOS (41) y puede ser utilizada para la descripción semántica de una amplia variedad de recursos bibliográficos. Dado que implementa FRBR, se pueden emplear *FRBR categorías* para describir claramente la separación entre las partes que componen el proceso editorial. Además, permite incluir fácilmente elementos de otros vocabularios para describir entidades particulares involucradas en el proceso editorial, como por ejemplo el uso de FOAF para describir personas u organizaciones. FaBiO y BIBO (42) (Bibliographic Ontology) son muy parecidos entre sí, sin embargo FaBiO está estructurado siguiendo el modelo FRBR y dispone de una colección de clases y propiedades más extensa que BIBO.

Por otro lado, CiTO ha sido creada por Shotton (2009) con la intención de capturar la intención del autor de la cita cuando la lleva a cabo. Permite indicar las motivaciones del autor para citar otro documento ya sea de manera retórica o factual. Este vocabulario contiene dos objetos principales: *cites* y *isCitedBy* que a su vez tienen 32 subpropiedades. Estas propiedades pueden ser clasificadas como retóricas o de hecho. Las retóricas a su vez pueden ser agrupadas en 3 sets: positivas, informativas o neutrales y/o negativas (ver Imagen 10 en anexo 5).

Peroni, Dutton, Gray y Shotton (2015) en un reciente artículo exponen la necesidad de publicar las referencias bibliográficas en acceso abierto (open access, OA):

“En esta década de OA, pensamos que es un escándalo que las bibliografías de los artículos científicos, las cuales son elementos principales de la comunicación científica porque permiten la atribución de crédito e integrar nuestras investigaciones independientes, no están todavía disponibles libremente para su uso científico. Pa-

ra rectificar esto, las referencias necesitan ser reconocidas como parte de Creative Commons y estar localizada en un repositorio abierto, donde deberían ser almacenados en formatos apropiados para su lectura automatizada, para ser fácilmente reutilizada por máquinas para asistir a las personas a producir nuevas publicaciones. Por lo que queda trabajo por hacer.”

Los autores proponen como recurso para solucionar estas lagunas la utilización de The OCC (43), un repositorio de acceso abierto que contiene información sobre referencias bibliográficas. Su objetivo es el de facilitar la publicación de las referencias para que otros puedan abiertamente reutilizar y mejorarlas con cualquier propósito. Según los autores esto es posible depositando las referencias en una licencia Creative Commons.

Para conseguir esto, Peroni, Dutton, Gray y Shotton (2015) sugieren los siguientes propósitos:

1. Crear una infraestructura semántica que haga posible la descripción de las citas, referencias y entidades bibliográficas en RDF.
2. Extender la infraestructura semántica para poder soportar datos sobre citas y entidades así como citas bibliográficas y entidades bibliográficas.
3. Proveer de ejemplos de cómo estas ontologías pueden ser aplicadas al mundo real de los datos.
4. Publicar las referencias bibliográficas como Linked Open Data en el repositorio The OCC.

Los mencionados autores hacen hincapié en la necesidad de invertir esfuerzos en el desarrollo de modelos semánticos para la publicación de información bibliográfica y referencias de manera que puedan ser procesadas por máquinas, y de desarrollar herramientas de anotación que permitan a los autores utilizar los modelos definidos para enriquecer sus documentos con definiciones semánticas.

## 5. Tecnologías aplicadas a la edición de revistas

La utilización de metadatos para la descripción de revistas científicas digitales es actualmente algo común dentro de las editoriales. Tal y como exponen Rovira y Marcos (2006), el grupo Digi-Doc de la Universitat Pompeu Fabra analizó 61 revistas digitales de libre acceso del campo de la Información y Documentación para averiguar

qué porcentaje disponían de metadatos y cuáles no. Aunque los resultados de esta estadística no son extrapolables al periodo actual, los datos por aquel entonces eran poco alentadores, tan solo el 69% de las revistas tenían artículos en HTML y de estos, solo el 13% disponía de metadatos específicos para cada artículo.

Existen diversos esquemas de metadatos que pueden ser utilizados para la descripción semántica de las revistas o artículos científicos. A continuación se presentan algunos ejemplos:

Dublin Core (DC) creado en 1995 como esquema de metadatos para catalogar cualquier tipo de documento digital. Consta de un conjunto de elementos básicos (15), para que sean los propios autores del documento los que realizaran la catalogación del mismo. Se trata de un esquema muy versátil, simple, flexible y extensible. Un ejemplo de la utilización de DC en revistas es la revista “BiD: textos universitarios de bibliotecología i documentació” que desde el año 2000 incorpora en sus artículos metadatos. Además, existe un portal dedicado a la recuperación de revistas por metadatos en DC llamado Temaria. Presenta un problema importante que tiene que ver con el objeto de nuestro trabajo: “la lentitud con que se ha desarrollando una fórmula para incluir la cita bibliográfica de la revista que incluye el artículo —imprescindible, por otro lado, para aplicar el DC a este tipo de recursos”. (Estivill et al., 2005, p. 3)

Publishing Requirements for Industry Standard Metadata (PRISM) es un esquema de metadatos para la creación, administración, agregación, producción, distribución y reutilización de contenidos publicados. Está orientado a la descripción de publicaciones. PRISM incorpora elementos en XML, RDF o DC. Con PRISM se pueden describir las entidades que intervienen en el proceso creativo (autores, editores, etc.), localizaciones, organizaciones, fechas de publicación, derechos de autor, etc.

Friend of a Friend Ontology (FOAF) es un vocabulario para la descripción de personas e instituciones. En publicaciones puede ser útil para la normalización de autores o entidades. De propósito parecido es The Semantic Web for Research Communities (SWRC), que permite la descripción de personas, instituciones, organizaciones y publicaciones académicas.

De uso más extendido en la actualidad es FRBR (Requerimientos funcionales para registros bibliográficos), modelo conceptual desarrollado por la IFLA para la descripción de documentos y su evolución. El modelo se estructura en diferentes entidades: obra, expresión, manifestación e ítem.

La ontología Bibliographic Ontology Specification (BIBO) proporciona los conceptos y propiedades principales necesarias para la descripción de citas y referencias bibliográficas en la Web Semántica. Incluye DC, PRISM y FOAF. Otra ontología es Swan Citations Ontology (SWAN), módulo disponible para la descripción de recursos bibliográficos. Está incluido en CiTO.

Tous, Guerrero y Delgado (2011), proponen una nueva arquitectura de metadatos en RDF y OWL cuyo objetivo es proveer de fiabilidad tanto a las citas como al seguimiento como ciclo de vida de la cita. Esta arquitectura está basada en el uso de metadatos semánticos en las diferentes fases del proceso editorial de la revista. El objetivo de la arquitectura es permitir a los sistemas de análisis verificar la procedencia y la confianza de los metadatos de lectura mecánica sobre las citas, antes de su incorporación en los repositorios.

Además, con la utilización de metadatos los autores y editores pueden almacenar evidencias sobre las diferentes fases del proceso editorial. Esto es posible porque utilizan un sistema de firma digital basado en el esquema PKI ('public key infrastructure'). Los autores tendrán un certificado que los identifique como autor a través de una URI (Universal Resource Identifier) y una marca de tiempo que indica la fecha de creación del certificado. Además, dispone de un sistema antirrepudio que permite a los autores certificar sus publicaciones en sus evaluaciones de curriculums. En cuanto a las revistas, tendrán un certificado que contenga información relativa a la revista tal como los editores, miembros de la junta directiva, datos sobre la aceptación o rechazo de los artículos (con sus respectivos comentarios del editor).

La utilización de esta propuesta de esquema de metadatos supone las siguientes ventajas:

1. Los autores pueden certificar ante cualquier entidad de evaluación que se evaluará su curriculum con las publicaciones que ha realizado.
2. Las entidades evaluadoras podrán solicitar el análisis de citas y obtener las publicaciones que el autor haya realizado.
3. Los autores no podrán falsificar las notificaciones de sus publicaciones.
4. Los editores no podrán repudiar la publicación de artículos una vez enviado el certificado.
5. Dos o más autores no podrán intentar hacer creer al sistema que son la misma persona para obtener más publicaciones a su cuenta.

### 5.1. Microdatos para la edición de revistas electrónicas

La OCLC creó en 2012 su propia extensión de Schema.org con clases que recogen los diferentes tipos de recursos bibliográficos presentes en la mayoría de las bibliotecas: material de archivo, archivos de ordenador, juegos, imágenes, páginas web, material audiovisual, etc. Las clases añadidas de primer nivel son en realidad subclases del vocabulario de Schema.org denominado CreativeWork. Esta propuesta resulta muy reveladora para este trabajo, en cuanto que recoge los principales recursos de una institución documental, sin embargo, no profundiza en la descripción de las clases, ni define propiedades con las que poder extender la descripción de publicaciones periódicas más allá de definir la clase 'Periodical' como subclase de CreativeWork.

Otra institución de reconocido prestigio, la Library of Congress (LOC), está desarrollando el proyecto BIBFRAME basado en RDF como futura forma de descripción bibliográfica en la web. Pretende reemplazar a los formatos MARC con este vocabulario basado en los principios de Linked Data. Con BIBFRAME se pretende construir un vocabulario que permita la representación y el intercambio de información a través de las siguientes medidas (LOC, 2015a):

1. Diferenciar claramente entre el contenido conceptual y las manifestaciones físicas o digitales.
2. Identificar de manera no ambigua entidades (por ejemplo, autoridades).
3. Hacer uso y exponer las relaciones entre entidades.

Las clases principales de BIBFRAME son las siguientes (LOC, 2015b):

- Creative Work: representa la esencia conceptual del recurso a catalogar.
- Instance: un recurso que refleja una forma de realización individual o material de la obra.
- Authority: el recurso refleja las autoridades claves que a las cuales le han sido definidas relaciones en Work e Instance. Por ejemplo: personas, lugares, materias, organizaciones, etc.
- Annotation: un recurso que enlaza nuestro conocimiento sobre otro recurso. La autoría de la anotación es importante. Por ejemplo: fondo al que pertenece el recurso, portadas, cubiertas o comentarios sobre la obra.

Otro proyecto colaborativo donde podemos encontrar metadatos para publicaciones periódicas es Freebase. En su portal podemos encontrar conjuntos de tipos y propiedades para describir entidades que serán conectadas a través de grafos:

“Almacenando los datos en formato de grafos, Freebase puede rápidamente realizar conexiones arbitrarias entre los topics y añadir fácilmente nuevos esquemas sin necesidad de cambiar la estructura de los datos.” (Google Developers, 2015)

Su licencia es Creative Commons, por lo que sus esquemas son reutilizables. Fue desarrollado por Metaweb (empresa comprada en 2010 por Google). A mediados de 2015, Freebase dejará de ser un servicio independiente, uniéndose a Wikidata para publicar datos estructurados.

Mika y Potter (2012), a través de un estudio estadístico sobre la inserción y penetración de los metadatos en la web ponen de manifiesto un progreso importante. Más del 30% de los metadatos contenidos en la web son microformatos, RDFa o microdatos. La utilización de RDFa ha crecido significativamente comparada con otros estudios previos, mientras que la adopción de los microdatos ha crecido muy rápido debido al éxito de Schema.org. De igual manera, Mendes, Muhleisen y Bizer (2012) exponen en su estudio que la utilización de microdatos se ha visto incrementada entre los diferentes formatos de datos estructurados embebidos en comparación con otros.

Ambiah y Lukose (2012) han propuesto una herramienta llamada Schema.org Microdata Creator (ScheMicCr) para la generación automática de microdatos de descripción de las páginas web (existentes o de nueva creación). Google dispone de su propia herramienta de marcado de datos creada en 2012 con el objeto de facilitar el etiquetado de datos estructurados de un sitio web ya rastreado por Google a webmasters. Para su utilización no son necesarios conocimientos de HTML o lenguajes de programación, a través de un asistente se puede proceder al etiquetado de manera fácil e intuitiva con el uso del ratón. Herramientas como las mencionadas anteriormente, facilitan la incorporación de microdatos a la web agilizando los procesos de etiquetado, lo que facilita su usabilidad y fácil empleo.

Por otro lado, Nogales et al. (2013) han explorado el potencial de mapear Schema.org con LOD ('Linked Open Data'), concluyendo que resulta relativamente sencillo realizar un mapeo

entre las clases y propiedades de estos dos esquemas. La posibilidad de realizar un mapeo resulta de vital importancia durante el periodo en el que los dos esquemas convivan, ya que favorece la interoperabilidad y el intercambio de información. Muchos webmasters todavía andan indecisos sobre qué esquema de marcado utilizar para describir semánticamente su web, por lo que la posibilidad de una futura adaptación o transición podría favorecer la colaboración de los creadores de webs en la implantación de los esquemas de marcado:

“Para que los usuarios se beneficien realmente de las técnicas semánticas, los autores (de las webs) necesitamos visualizar y explorar información desestructurada y semántica de una manera amigable”. (Khalili & Auer, 2013)

Para ello, estos últimos autores presentan el concepto WYSIWYM ('What You See Is What You Mean') como la posibilidad de poder manipular el contenido semántico de manera estructurada y convencional. Para Khalili y Auer (2013), este concepto formaliza la unión entre los modelos de representación semántica y elementos para la edición, la visualización y la exploración. Un ejemplo que los autores utilizan es la interfaz de código abierto RDFaCE para editar contenido semántico basado en RDFa.

## 6. Propuesta de vocabulario

Nuestra propuesta de vocabulario se basa en Schema.org y sirve para la descripción de publicaciones periódicas. Se presenta como una herramienta útil para favorecer la identificación y recuperación de publicaciones periódicas en principales motores de búsqueda. Estamos convencidos de que facilitando la descripción de las publicaciones en un lenguaje que los motores de búsqueda puedan leer y analizar, la localización de artículos por parte de los investigadores se verá afectada positivamente. Con el vocabulario propuesto se presenta la oportunidad de poder justificar y contextualizar las citas incluidas en los artículos de las revistas científicas. El principal objetivo de estas propiedades es proporcionar al lector la posibilidad de conocer más sobre el por qué (la razón) de la inclusión de la cita, incluir enlaces al texto del artículo para indicar dónde poder conseguirlo, qué ha motivado al autor a incluirla, la razón de no incluir otro autor o comentarios subjetivos sobre la obra citada, etc. Por otro lado, también permite ampliar información a través de hiperenlaces que el autor del documento quiera introducir a modo de hipertexto documental:

“Estamos partiendo de la base de que el usuario tiene una serie de necesidades de información para la adquisición de conocimientos que luego serán usados de un modo u otro, sin embargo el hipertexto no debe dar una impresión de estructura pasiva en este proceso, todo lo contrario, el hipertexto es una herramienta activa en la búsqueda de información ya que puede ofrecer una información elaborada que responda a las necesidades de información de un usuario: el hiperdocumento.” (Pastor Sánchez y Saorín Pérez, 1995 p. 10)

Aunque el concepto del hipertexto no es para nada novedoso, su inclusión frecuente en la lectura de documentos científicos si puede considerarse algo innovador. Su utilización para realizar enlaces a páginas oficiales o complementarias para adquirir información suplementaria sobre conceptos (enciclopedias y otras fuentes de información), ofrece al lector una experiencia de lectura enriquecida con más información de la que el autor por razones externas a él, haya incluido en el artículo (los microdatos se convierten así en representantes de esa “metainformación” de la que se habla en las definiciones indicadas en apartados anteriores). Este vocabulario permite, en definitiva, tanto al lector como al autor, extender la información proporcionada más allá de los límites definidos dentro del propio artículo.

En el blog de Schema.org (2014b), se habla de la inclusión de propiedades para la descripción de relaciones bibliográficas y de publicaciones periódicas: Article, Periodical, PublicationIssues y PublicationVolumes. También incluye la posibilidad de utilizar las propiedades hasPart y isPartOf para describir las relaciones conceptuales y/o de representación de una obra o representación. Estas extensiones fueron realizadas por el W3C Schema Bib Extend Community Group. El mecanismo de extensión de los vocabularios en Schema.org fue modificado en mayo de 2015. Los mecanismos de extensión tienen como finalidad la creación de vocabularios adicionales a los del núcleo de Schema.org. La extensión se puede llevar a cabo mediante dos procesos: extensiones revisadas/alojadas o externas. Sin embargo, en ambos procesos se añaden subclases o propiedades a las definidas en el núcleo. Las propiedades pueden ser añadidas a clases nuevas o a clases ya existentes. Además, se señala como aspecto importante que las extensiones deben ser consistentes con el núcleo de Schema.org y no tener duplicados o términos sinónimos. Las extensiones revisadas/alojadas: en este tipo de extensión, los íte-

mes son creados y mantenidos por los creadores de la extensión. Para convertirse en una extensión de este tipo debe de ser discutida en la comunidad y contar con un respaldo fuerte en la comunidad. Mientras que las extensiones externas: creadas por un tercero o desarrollador de 'apps' para crear una extensión específica para su aplicación. Este tipo de extensión puede estar alojada en la web del tercero o en la del desarrollador de 'apps'. Una vez que la extensión ha sido aprobada, se deberá transferir un fichero a un directorio especificado por los administradores de Schema.org.

## 6.1. Vocabulario propuesto

Nuestra propuesta de vocabulario está compuesta por términos ya presentes en los vocabularios de Schema.org junto a otro conjunto de términos que se proponen para su inclusión. Aquellos ya representados se informan en las tablas con la correspondiente URL que permite acceder a la propiedad identificada. El motivo de reutilizar propiedades ya definidas es el de adaptar el vocabulario propuesto al existente para aumentar la interoperabilidad entre vocabularios y reutilizar propiedades. Asimismo, se disminuye la ambigüedad entre propiedades parecidas entre sí. Con el fin de que la descripción de los documentos no sea una tarea pesada e interminable para los autores (algo que podría inducirles a no llevarla a cabo), se han dividido las propiedades propuestas en dos categorías: básicas para la descripción (aquellos términos que permitan describir los aspectos más importantes de los documentos) y opcionales (términos que aunque sean importantes, puede llevarse a cabo una descripción mínima del contenido sin su presencia).

En cuanto al propósito de los términos para la descripción se han identificado cuatro categorías:

1. Descripción de revistas.
2. Descripción de artículos.
3. Identificación de partes del documento.
4. Justificación y propósito de las citas.

### Descripción de revistas:

Son aquellos términos o propiedades empleadas para la identificación formal y de contenido de las publicaciones periódicas. Estas propiedades recogen los aspectos identificativos de las revistas tales como nombre, ISSN, volumen, etc. en definitiva, datos importantes para cualquier referencia bibliográfica (se muestra la propiedad que

describe el nombre de la revista, el resto de esta categoría está en el anexo 1).

#### Descripción de artículos:

Se trata de propiedades para la descripción de los artículos de revistas. Se trata de identificadores básicos que permitan localizar y recuperar aquellos artículos referenciados (se muestra la propiedad que describe el título del artículo, el resto de esta categoría está en el anexo 2). Posiblemente esta propiedad debiera llamarse "title" pero así es como está definida en Schema.org y se reutiliza con su nombre original.

PROPERTY	<b>name</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Thing
BA-SIC/OPTIONAL	Basic
DESCRIPTION	The name of the article in the main language.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/name">http://schema.org/name</a>

#### Justificación y propósito de las citas:

Esta categoría es la más innovadora de la propuesta. Se compone por propiedades que permitan la justificación, argumentación, contextualización, definición de propósitos, etc. de la cita incluida en el texto. Además, conociendo la ubicación de citas se podrá evaluar la relevancia o peso dentro del artículo, por ejemplo, las menciones en conclusiones o resultados que para algunos autores suelen poseer mayor relevancia que las contenidas en la introducción (se muestra la propiedad quizá más original de la propuesta, aquella en la que el autor indica el objetivo de la cita, el resto de la propiedades de la categoría está en el anexo 3).

PROPERTY	<b>referenceAim</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Reference
BA-SIC/OPTIONAL	Optional
DESCRIPTION	This property indicates the aim of the reference in order to know the intentions of the author with the quote.
ALREADY DEFINED? Y/N	No
LINK	

#### Partes del documento

Propiedades para la descripción estructural y formal del artículo científico. Pretenden ser una guía de localización para conocer dónde se encuentra la cita, los datos del autor, etc. En la siguiente tabla se muestra la propiedad correspondiente al resumen del artículo en su lengua original, el resto de la propiedades de la categoría está en el anexo 4.

PROPERTY	<b>articleAbstract</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BA-SIC/OPTIONAL	Optional
DESCRIPTION	This property represents the abstract in the main language of the article.
ALREADY DEFINED? Y/N	No
LINK	

#### 6.2. Aplicación de la propuesta para el enriquecimiento de artículos

Al igual que hicieron Shotton et al. (2009) con el artículo de Reis (2008) sobre el riesgo de contraer leptospirosis por parte de los habitantes de las favelas de Pau da Lima en Salvador de Bahía, hemos procedido a aplicar nuestra propuesta de enriquecimiento a un artículo publicado en esta misma revista. Hemos elegido precisamente el primero de los publicados en el volumen 1, cuyo autor es el insigne profesor británico Thomas Daniel Wilson y se titula 'Information Management: today and tomorrow', publicado en el año 2011.



Figura XI. Fragmento de la primera página del artículo 'Information Management, today and tomorrow', escrito por Tom Wilson y publicado en el primer número de la revista *Cuadernos de Gestión de información*. Fuente:

<http://revistas.um.es/gesinfo/article/view/207461>

Este artículo se publicó en formato PDF, hemos realizado una versión en HTML 5 y sobre la misma hemos procedido a aplicar una serie de microdatos a partir de la nuestra propuesta de

trabajo. En la siguiente ejemplo podemos ver los microdatos aportados para describir a la revista.

Ejemplo1:

```
<div itemscope item-
type="http://schema.org/Periodical">
<h1 itemprop="journalName">Cuadernos de Ges-
tión de Información</h1>
>span itemprop='volumeNumber'>2011 </span>
<span itemprop='hasArticle'>Information Manage-
ment, today and tomorrow</span>
<span itemprop="URL">http://revistas.um.es/gesinfo/issue/view/
13051</span>
<span itemprop="issn">2253-8429</span></span>
<span itemprop="journalRegularity">Anual</span>
<span itemprop="editor">Facultad de Comunicación
y Documentación, Universidad de Murcia</span>
<span itemprop="addressCountry">Espanya</span>
<div itemprop="address" itemscope item-
type="http://schema.org/PostalAddress">
<span itemprop="addressLocality">Murcia</span>,
<span itemprop="addressRegion">Murcia </span>
<span itemprop="postalCode">30100</span>
<span itemprop="streetAddress">Campus Universi-
tario de Espinardo</span>
</div> </div>
```

En la siguiente ejemplo se recoge un fragmento del enriquecimiento de datos efectuado sobre la introducción del artículo, su propósito y el método de trabajo desarrollado.

Ejemplo 2:

```
<span itemprop="articleIntroduction">These are
important questions for anyone work-ing in this field
today because we know that it changes very rapidly.
To know how the subject is perceived today helps us
to cope with the immediate changes and to identify
gaps and weaknesses in our own information man-
agement activities. Similarly, guessing how the fu-
ture is going to turn out can help us to prepare for
that future: designing new teaching programmes in
academia, and plotting our own career structure in
the world at large. [...] </span>
<span itemprop="articleAim">El objetivo principal de
esta investigación es.... [...] </span>
<span itemprop="articleMethod">Para realizar el
estudio se ha utilizado los siguientes métodos es-
tadísticos [...], procesados por [...], con una escala
[...], agrupados en [...], etc.</span>
```

Y en el siguiente ejemplo podemos observar el enriquecimiento semántico realizado sobre una de referencias del artículo, viendo su descripción, propósito y argumentación.

Ejemplo 3:

```
<span itemprop="referenceBiblio">Enterprise Con-
tent Management. (2011). What is enter-prise con-
tent management? Retrieved 25 July, 2011 from
http://www.contentmanager.eu.com/ecms.htm (Arc-
hived by WebCite® at
http://www.webcitation.org/62hiDeFXh)</span>
<span itemprop="referenceAim">The purpose of this
aim is to highlight this acronym of other denomina-
tions that are used to introduce Document Manage-
ment Systems into organizations. El propósito de
esta cita es resaltar este acrónimo del resto de de-
nominaciones que se emplean para presentar a los
sistemas de gestión de documentos en una organi-
zación</span>
<span itemprop="referenceArgumentation">The
ECM acronym is perhaps the most widely used in
scientific and technical literature . Its scope is broa-
der .El acrónimo ECM es quizá el más utilizado en la
literatura científica y técnica especializada. Su al-
cance es más amplio. Fuente:
https://www.laserfiche.com/ecmblog/17-acronyms-
you-need-to-know-before-researching-ecm/ </span>
```

Para ver el ejemplo de aplicación completo hay que consultar el anexo 6. Esta versión HTML del artículo será publicada en breve en la revista Cuadernos de Gestión de Información y en fases posteriores de trabajo se intentará medir el impacto y visibilidad de la misma en comparación con la tradicional versión en PDF.

## 7. Conclusiones

En el presente trabajo se ha presentado la propuesta de un vocabulario de microdatos basado en Schema.org para describir las publicaciones científicas de manera que permita enriquecer a nivel semántico la información sobre revistas, artículos, autores, estructura, referencias bibliográficas y otras piezas de esta información. Este vocabulario pretende mejorar el nivel de descripción semántica de las publicaciones científicas y facilitar su recuperación por parte de los motores de búsqueda, al mismo tiempo que persigue describir los documentos en lenguajes que las máquinas puedan interpretar. La utilización de este vocabulario creemos que amplía el horizonte de la edición de revistas científicas y acerca un poco más sus procesos editoriales a la era digital modernizándolos, propiciando una mayor interoperabilidad semántica y funcional entre editoriales y facilitando las ta-

reas de localización y redifusión de la información a lectores e investigadores.

La elección del esquema de microdatos Schema.org está motivada principalmente por el gran potencial de crecimiento y respaldo que aportan los grandes motores de búsqueda: Google, Yahoo, Bing y Yandex. Contar con estos gigantes tecnológicos como participantes y potenciadores del proyecto garantiza recorrido a estos vocabularios de microdatos dentro de la web actual, cada vez más semántica. Otro de los aspectos que justifica la elección de Schema.org para la elaboración de la propuesta es que se trata de un esquema de microdatos libre y de acceso abierto, respaldado además por grupos de la W3C, lo que implica la presencia de una gran comunidad que lo impulse y continúe desarrollando, incluyendo mejoras y solucionando errores.

Por otro lado, la utilización de esquemas de metadatos actualmente cobra una importancia especial debido a la gran necesidad de descripción de documentos en la web. La utilización de este tipo de herramientas semánticas junto a perfiles de aplicación de metadatos y/o ontologías para describir semánticamente la información plasmada en los documentos web, es algo indispensable para la interoperabilidad y reutilización de información en la web. Asimismo, esto favorece que las máquinas y motores de búsqueda puedan inferir información y ser capaces de establecer relaciones entre documentos mejor descritos semánticamente. El vocabulario Schema.org se atisba como una buena apuesta para este tipo de descripción, debido a la facilidad de implementar microdatos. Este hecho hace posible la expansión de Schema.org de mano de los desarrolladores web que encuentren la tarea de describir semánticamente sus portales de manera menos tediosa y más fácil de implementar.

El concepto de Publicación Semántica impulsado por Shotton recoge las principales acciones a llevar a cabo en este campo de trabajo. Resulta interesante su propuesta ontológica para la descripción bibliográfica en lenguaje RDF: BiBO, CiTO, etc. y el repositorio The OCC desde el cual publicar referencias y citas bibliográficas en acceso abierto. Este autor y su trabajo en publicación semántica sientan un precedente para la elaboración de esta propuesta, y marca el camino a seguir en la publicación semántica.

Nuestra propuesta no sólo persigue facilitar su tarea a los motores de búsqueda para aumentar su efectividad en la búsqueda de información científica en la web, también aporta información adicional a los lectores de los artículos dotándo-

los de una mayor profundidad informativa y de un acceso más completo y seguro a la información. El vocabulario propuesto intenta cubrir todos los aspectos básicos y una amplia diversidad de opciones en la descripción semántica de las publicaciones científicas en la web. Uno de los aspectos recogidos que se ha considerado más relevante e innovador es incluir la justificación de las citas y de las referencias bibliográficas en el texto (lo que podría llamarse "bibliografía en contexto"). La descripción de las intenciones, argumentaciones, discusiones o justificaciones es un aspecto poco o nada explotado hasta en la edición de publicaciones científicas, por lo que se ha considerado una de las piedras angulares de nuestro vocabulario propuesto y el punto de partida de futuros trabajos relacionados.

Estas futuras líneas de investigación van en la línea de compartir en acceso abierto nuestra propuesta y someterla a debate en foros especializados, al mismo tiempo que se plantea la necesidad de implementar a nivel informático un asistente de marcado semántico para artículos y revistas científicas del campo de la Información y Documentación. Tras este desarrollo, corresponderá llevar a cabo el proceso de diseño de un sistema de evaluación del posible impacto (positivo esperamos) de esta propuesta, tanto en términos de visibilidad y ubicuidad de la información como en la mejora de la experiencia del usuario (los investigadores que consultan habitualmente las revistas científicas). Con base en esos resultados podremos determinar si la descripción semántica a través de microdatos de las publicaciones científicas digitales mejora ambos aspectos y justifica el esfuerzo adicional que se exigirá a los editores de las revistas y, seguramente, a los autores de los trabajos científicos. Además, se evaluará si el marcado de elementos como imágenes, figuras, tablas y referencias o citas en los artículos, se transforma en un valor añadido de información para los lectores, que permita ampliar la información contenida en el artículo.

## Notas

- (1) OJS son las siglas del gestor 'Open Journal System', una solución informática basada en un sistema de administración y publicación de revistas y documentos periódicos en Internet. El sistema está diseñado para reducir el tiempo y energías dedicadas al manejo exhaustivo de las tareas que involucra la edición de una publicación seriada. Más información en <https://pkp.sfu.ca/>
- (2) Más información sobre esta revista en: <http://redc.revistas.csic.es/index.php/redc>

- (3) Profesor Emérito del Departamento de Zoología de la Universidad de Oxford especializado en la utilización de las tecnologías semánticas en actividades de publicación de datos vinculados dentro de la Web Semántica, con la idea de mejorar la comunicación científica, en general, y más en particular, facilitar un libre acceso e interoperabilidad de las citas bibliográficas. Más información en: [http://www.zoo.ox.ac.uk/people/view/shotton\\_dm.htm](http://www.zoo.ox.ac.uk/people/view/shotton_dm.htm)
- (4) HyperText markup Language versión 5. Es la última versión del lenguaje de marcas HTML con el que se crean las páginas web. Esta nueva versión contiene dos variantes: HTML clásico y XHTML. De esta versión destacamos la posibilidad que incorpora para la descripción semántica de información en la web. Para más información: [http://www.w3schools.com/html/html5\\_intro.asp](http://www.w3schools.com/html/html5_intro.asp)
- (5) HTML microdata: <http://www.w3.org/TR/microdata/>
- (6) LISA: Library and Information Science Abstracts. Para más información: <http://www.proquest.com/products-services/lisa-set-c.html>
- (7) Dublin Core: es un vocabulario de metadatos promovido por la DCMI (Dublin Core Metadata Initiative) para la descripción de recursos digitales a través de 15 elementos o descriptores básicos. Este vocabulario está plasmado en la norma ISO 15836/2003. Más información en: <http://dublincore.org/>
- (8) Schema.org: conjunto de microdatos para etiquetar webs, fácilmente procesable por los principales motores de búsqueda. Más información en: <http://schema.org/>
- (9) Resource Description Framework (RDF): es un estándar de metadatos simple para el intercambio de información en la web. Describe la información a través de tripletas: sujeto, objeto, predicado. Posibilita la interoperabilidad y la reutilización de vocabularios o esquemas. <http://www.w3.org/RDF/>
- (10) MARC21: conjunto de etiquetas para representar los aspectos formales y de contenido de documentos bibliográficos para su posterior recuperación por un ordenador. Facilita el intercambio de registros a través de la catalogación corporativa. Permite la descripción de datos bibliográficos, autoridades, ejemplares, etc. Más información en: <http://www.loc.gov/marc/>
- (11) IBERMARC: conjunto etiquetas para la catalogación de recursos bibliográficos adaptado al modelo español. Está basado en el modelo MARC. Más información en: <http://www.bne.gob.es/es/Inicio/Perfiles/Bibliotecarios/NormasInternacionales/FormatoIBERMARC/index.html>
- (12) Modelo CIDOC: se trata de un modelo semántico que constituye una ontología del patrimonio cultural. Para más información: [http://www.bnf.fr/es/profesionales/modelizacion\\_ontologias/a.modelo\\_cidoc\\_crm.html](http://www.bnf.fr/es/profesionales/modelizacion_ontologias/a.modelo_cidoc_crm.html)
- (13) Encoded Archival Description: estándar basado en XML para la descripción archivística. Para más información: <http://www.loc.gov/ead/>
- (14) General International Standard Archival Description: norma para la descripción archivística multi-nivel publicada por el Consejo Internacional de Archivos. Para más información: <http://www.ica.org/download.php?id=1745>
- (15) International Standard Archival Authority Record for Corporate Bodies, Persons, and Families: Norma internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias publicada por el Consejo Internacional de Archivos. Para más información: <http://www.agn.gob.mx/menuprincipal/archivistica/normas/pdf/asaar.pdf>
- (16) Content Standard for Digital Geospatial Metadata: estándar para la descripción de metadatos sobre información digital geoespacial. Se compone de un conjunto de definiciones y terminología y establece los nombres de los elementos de datos y los grupos de elementos que se quieren describir. Para más información: <http://www.fgdc.gov/metadata/csdgmg/>
- (17) Government Information Locator Service: estándar para ayudar a los investigadores a localizar información gubernamental. Para más información: <http://www.gils.net/>
- (18) Media Error Monitoring and Reporting Information: [http://www.iso.org/iso/catalogue\\_detail.htm?number=24942](http://www.iso.org/iso/catalogue_detail.htm?number=24942)
- (19) HyperText markup Language version 5. Es la última versión del lenguaje de marcas HTML con el que se crean las páginas web en la WWW. Esta nueva versión contiene dos variantes: HTML clásico y XHTML. De esta versión destacamos la posibilidad que incorpora para la descripción semántica de información en la web. Para más información: [http://www.w3schools.com/html/html5\\_intro.asp](http://www.w3schools.com/html/html5_intro.asp)
- (20) HTML microdata: <http://www.w3.org/TR/microdata/>
- (21) Catálogo de vocabularios publicados: <http://schema.org/docs/full.html>
- (22) Schema.org Community Group: <https://www.w3.org/community/schemaorg/>
- (23) Schema BIB Extend Community Group: <https://www.w3.org/community/schemabibex/>
- (24) Schema BIB Extend Wiki: [https://www.w3.org/community/schemabibex/wiki/Main\\_Page](https://www.w3.org/community/schemabibex/wiki/Main_Page)
- (25) Web de información sobre RDFa con documentación, ejemplos y ayudas: <http://rdfa.info/>
- (26) eXtensible HyperText Markup Language. Se trata del lenguaje HTML expresado en XML. Surge para impulsar el etiquetado semántico en la web y facilitar la separación entre contenido y estructura.

- Para más información sobre las similitudes y diferencias de estos dos lenguajes de marcas: [http://www.w3schools.com/html/html\\_xhtml.asp](http://www.w3schools.com/html/html_xhtml.asp)
- (27) RDFa Lite 1.1: versión simplificada del vocabulario RDF con un conjunto de propiedades mínima para para expresar datos legibles por máquina en documentos Web como HTML, SVG y XML. Para más información: <http://www.w3.org/TR/2012/REC-rdfa-lite-20120607/>
- (28) Más información en: <http://skos.um.es/TR/rdfa-lite/>
- (29) <https://support.google.com/webmasters/answer/99170?hl=es>
- (30) Herramienta de pruebas de datos estructurados: <http://www.google.com/webmasters/tools/richsnippets>
- (31) Fragmentos enriquecidos para personas: <https://support.google.com/webmasters/answer/146646>
- (32) Wiki dedicada a los microformatos en donde puede encontrar más información sobre su implementación y las etiquetas a utilizar: <http://microformats.org/wiki/about>
- (33) Herramienta Google para la prueba de datos estructurados: <http://www.google.com/webmasters/tools/richsnippets>
- (34) IUPACC: vocabulario estandarizado y normalizado con nomenclaturas, unidades y símbolos químicos. Para más información: <http://goldbook.iupac.org/>
- (35) Gene Ontology: ontología que describe las relaciones entre genes de cualquier organismo. Para más información: <http://geneontology.org/>
- (36) Cell Ontology: ontología para la descripción de de tipos de células. Para más información: <http://bioportal.bioontology.org/ontologies/CL>
- (37) Sequence Ontology: es una ontología colaborativa para la descripción de secuencias biológicas. <http://www.sequenceontology.org/>
- (38) Vocabulario de términos de cristalografía: [http://reference.iucr.org/dictionary/Main\\_Page](http://reference.iucr.org/dictionary/Main_Page)
- (39) Publishing Requirements for Industry Standard Metadata. Para más información ver epígrafe 6.
- (40) Requerimientos funcionales para registros bibliográficos. Para más información ver epígrafe 6.
- (41) SKOS Simple Knowledge Organization System. Para más información: <http://www.w3.org/2004/02/skos/>
- (42) Para más información sobre BIBO ver epígrafe 6.
- (43) The Open Citations Corpus <https://opencitations.wordpress.com/>
- (44) Article: subclase para describir artículos de revista o de periódico, etc. <http://schema.org/Article>
- (45) Periodical: subclase para la descripción de publicaciones periódicas <http://schema.org/Periodical>
- (46) PublicationIssue: subclase para la descripción de publicaciones periódicas que contienen agrupaciones de obras como artículos. <http://schema.org/PublicationIssue>
- (47) PublicationVolume: subclase para la descripción de publicaciones periódicas en varios volúmenes numerados. <http://schema.org/PublicationVolume>
- (48) hasPart: propiedad que indica que un CreativeWork pertenece a otro CreativeWork <http://schema.org/hasPart>
- (49) isPartOf: propiedad que indica que un CreativeWork es parte de otro Creative Work. <http://schema.org/isPartOf>

## Referencias

- Ambiah, N. & Lukose, D. (2012). Enriching Webpages with Semantic Information. *International Conference on Dublin Core and Metadata Applications*, Kuching, Sarawak, Malaysia. Recuperado de <http://dcpapers.dublincore.org/pubs/article/view/3663> [8-12-2014]
- Bernardis, S. (2014). Estudio del impacto del SEO semántico en los motores de búsqueda: aplicando microdatos y RDFa Lite 1.1 en el ámbito de Schema.org. *Cuadernos de Gestión de Información*, 3(1), 85–104. Recuperado de <http://revistas.um.es/gesinfo/article/view/207771> [12-12-2014]
- Berners-Lee, T. & Hendler, J. (2001). Publishing on the semantic web. *Nature*, 410 (6832), 1023–1024. doi.org/10.1038/35074206
- Estivill, A., Abadal, E., Franganillo, J., Gascón, J. y Rodríguez-Gairín, J.M. (2005a). Uso de metadatos Dublin Core en la descripción y recuperación de artículos de revista digitales = Use of Dublin Core metadata for describing and retrieving digital journals. *DC-2005 International Conference on Dublin Core and Metadata Applications*. Recuperado de <http://eprints.rclis.org/6760/> [10-01-2015]
- García-Marco, Francisco-Javier. (2013). Schema.org: la catalogación revisitada. *Anuario ThinkEPI*, 7, 169–172. Recuperado de <http://www.thinkepi.net/schema-org-catalogacion-revisitada> [04-04-2015]
- Gaurav, S., Jithendranath, Y., Adil, A., Yadav, S. & Kasturi, B. (2015). A study to assess and enhance educational specific search on web for school children. In *Proceedings of the 20 IS IEEE 9th International Conference on Semantic Computing (IEEE ICSC 20 IS)* (p. 260–263). California, USA: IEEE.
- Google (2014). *Preguntas frecuentes sobre schema.org*. Recuperado de <https://support.google.com/webmasters/answer/1211158?hl=es> [08-12-2014]

- Google (2014b) *Acerca de los microformatos*. Recuperado de <https://support.google.com/webmasters/answer/146897?hl=es> [08-12-2014]
- Google Developers. (2015). *Basic Concepts - Freebase API*. Recuperado de [https://developers.google.com/freebase/guide/basic\\_concepts](https://developers.google.com/freebase/guide/basic_concepts) [19-04-2015]
- Khalili, A. & Auer, S. (2013). WYSIWYM Authoring of Structured Content Based on Schema.org. In Lin. X., Manolopoulos, Y., Srivastava, D. & Huang, G. (Eds.). *Web Information Systems Engineering – WISE 2013* (p. 425–438). Springer Berlin Heidelberg. Recuperado de [http://link.springer.com/chapter/10.1007/978-3-642-41154-0\\_32](http://link.springer.com/chapter/10.1007/978-3-642-41154-0_32) [04-01-2015]
- Library of Congress. (2015a). *BIBFRAME Frequently Asked Questions (Bibliographic Framework Transition Initiative, Library of Congress)*. Recuperado de <http://www.loc.gov/bibframe/faqs/> [11-04-2015]
- Library of Congress. (2015b). *Overview of the BIBFRAME Model (BIBFRAME - Bibliographic Framework Initiative, Library of Congress)*. Recuperado de <http://www.loc.gov/bibframe/docs/model.html> [11-04-2015]
- ManuSporny. (2011). *The False Choice of Schema.org | The Beautiful, Tormented Machine*. Recuperado de <http://manu.sporny.org/2011/false-choice/> [04-01-2015]
- Mendes, P.N., Mühleisen, H. & Bizer, C. (2012). Sieve: Linked Data Quality Assessment and Fusion. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops* (p. 116–123). New York, NY, USA: ACM. doi.org/10.1145/2320765.2320803
- Mika, P. & Potter, T. (2012). Metadata Statistics for a Large Web Corpus. *Linked Data Workshop at the International World Wide Web Conference*, Lyon. Recuperado de <http://ceur-ws.org/Vol-937/ldow2012-inv-paper-1.pdf> [11-01-2015]
- Nogales, A., Sicilia, M.A., García-Barriocanal, E. y Sánchez-Alonso, S. (2013). Exploring the Potential for Mapping Schema.org Microdata and the Web of Linked Data. In Garoufallou, E. & Greenberg, J. (Eds.). *Metadata and Semantics Research* (p. 266–276). Springer International Publishing. Recuperado de [http://link.springer.com/chapter/10.1007/978-3-319-03437-9\\_27](http://link.springer.com/chapter/10.1007/978-3-319-03437-9_27) [04-01-2015]
- Pastor Sánchez, J.A. & Saorín Pérez, T. (1995). El hipertexto documental como solución a la crisis conceptual del hipertexto. El reto de los documentos cooperativos en redes. *Cuadernos de Documentación Multimedia*, (4), 41–52. Recuperado de <http://pendientedemigracion.ucm.es/info/multidoc/multidoc/revista/cuadern4/hiperdoc.htm> [21-06-2015]
- Peroni, S. & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17, 33–43. doi.org/10.1016/j.websem.2012.08.001
- Peroni, S., Dutton, A., Gray, T. & Shotton, D. (2015). Setting our bibliographic references free: towards open citation data. *Journal of Documentation*, 71 (2), 253-277.
- Reis, R.B., Ribeiro, G.S., Felzemburgh, R.D., Santana, F.S., Mohr, S., Melendez, A. X., & Ko, A. I. (2008). Impact of environment and social gradient on *Leptospira* infection in urban slums. *PLoS Negl Trop Dis*, 2 (4), e228.
- Rodríguez-Yunta, L. (2013). Indización en profundidad y aplicación de metadatos a materiales suplementarios en la edición de revistas. *Anuario ThinkEPI*, 8, 207–210. Recuperado de <http://www.thinkepi.net/indizacion-en-profundidad-y-aplicacion-de-metadatos-a-materiales-suplementarios-en-la-edicion-de-revistas> [04-01-2015]
- Ronallo, J. (2012). HTML5 Microdata and Schema.org. *The Code4Lib Journal*, (16). Recuperado de <http://journal.code4lib.org/articles/6400> [05-01-2015]
- Rosati, A., & Mayernik, M. (2013). Facilitating Data Discovery by Connecting Related Resources. *Semantic Web Journal*. Recuperado de <http://semantic-web-journal.net/system/files/swj484.pdf> [05-01-2015]
- Rovira, C., y Marcos, M.C. (2006). Metadatos en revistas-e de Documentación de libre acceso. *El Profesional de La Información*, 15 (2), 136–144. Recuperado de <http://www.elprofesionaldelainformacion.com/contenidos/2006/marzo/6.pdf> [20-03-2015]
- Schema.org. (2015). *Extending Schemas - schema.org*. Recuperado de <http://schema.org/docs/extension.html> [21-06-2015]
- Schema.org. (2014a). *Home - schema.org*. Recuperado de <http://schema.org/> [21-06-2015]
- Schema.org. (2014b). *schema blog: Schema.org Support for Bibliographic Relationships and Periodicals*. Recuperado de [http://blog.schema.org/2014/09/schemaorg-support-for-bibliographic\\_2.html](http://blog.schema.org/2014/09/schemaorg-support-for-bibliographic_2.html) [21-06-2015]
- Shotton, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, 22(2), 85–94. doi.org/http://dx.doi.org/10.1087/2009202
- Shotton, D., Portwin, K, Klyne, G. & Miles, A. (2009). Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Computational Biology*, 5 (4). doi.org/10.1371/journal.pcbi.1000361
- Tous, R., Guerrero, M. y Delgado, J. (2011). Semantic Web for Reliable Citation Analysis in Scholarly Publishing. *Information Technology and Libraries*, 30 (1), 24–33. doi.org/10.6017/ital.v30i1.3042

## Anexo 1. Properties from Journals

CLASS	Periodical
SUBCLASS OF	CreativeWorkSeries
DESCRIPTION	This class includes all kind of journals such as scientific journals or popular science magazine, etc.
ALREADY DEFINED? Y/N	Yes
LINK	<a href="http://schema.org/Periodical">http://schema.org/Periodical</a>

PROPERTY	<b>journalName</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Periodical
BASIC/OPTIONAL	Basic
DESCRIPTION	This property represents the name of the journal, could be the code name or the extended name.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>volumeNumber</b>
EXPECTED TYPE	Text or Integer
SUBCLASS OR CLASS	PublicationVolume
BASIC/OPTIONAL	Basic
DESCRIPTION	This property represents the number of the journal.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/volumeNumber">http://schema.org/volumeNumber</a>

PROPERTY	<b>URL</b>
EXPECTED TYPE	URL
SUBCLASS OR CLASS	Thing
BASIC/OPTIONAL	Optional
DESCRIPTION	URL of the item.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/URL">http://schema.org/URL</a>

PROPERTY	<b>issn</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Periodical
BASIC/OPTIONAL	Basic
DESCRIPTION	The International Standard Serial Number (ISSN) that identifies this periodical. You can repeat this property to (for example) identify different formats of this periodical.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/Periodical">http://schema.org/Periodical</a>

PROPERTY	<b>journalRegularity</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Periodical
BASIC/OPTIONAL	Optional
DESCRIPTION	This property indicates the regularity of the journal.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>editor</b>
EXPECTED TYPE	Person
SUBCLASS OR CLASS	Person
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the Person who edited the journal.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/editor">http://schema.org/editor</a>

PROPERTY	<b>addressCountry, addressLocality, addressRegion, postalCode, streetAddress</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	PostalAddress
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the mailing address
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/PostalAddress">http://schema.org/PostalAddress</a>

## Anexo 2. Properties from article

CLASS	Article
SUBCLASS OF	CreativeWork
DESCRIPTION	An article, such as a news article or piece of investigative report. Newspapers and magazines have articles of many different types and this is intended to cover them all.
ALREADY DEFINED? Y/N	Yes
LINK	<a href="http://schema.org/Article">http://schema.org/Article</a>

PROPERTY	<b>name</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Thing
BASIC/OPTIONAL	Basic
DESCRIPTION	The name of the article in the main language.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/name">http://schema.org/name</a>

PROPERTY	<b>keywords</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	CreativeWork
BASIC/OPTIONAL	Basic
DESCRIPTION	Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/keywords">http://schema.org/keywords</a>

PROPERTY	<b>name_alt</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Thing
BASIC/OPTIONAL	Optional
DESCRIPTION	The alternative name of the item. This could be used in order to represent the title of the article in different languages.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>keywords_alt</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	CreativeWork
BASIC/OPTIONAL	Optional
DESCRIPTION	The alternative keywords for the creative work. This could be used in order to represent the keywords of the article in different languages.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>articleURI</b>
EXPECTED TYPE	URL
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	This property indicates the URI (URL or DOI) of the Article.
ALREADY DEFINED? Y/N	No
LINK	-

### Anexo 3. Properties from references

En la versión actual, se puede encontrar la propiedad *Citation*<sup>1</sup> correspondiente a la clase CreativeWork, sin embargo, para nuestro propósito la utilización de esta propiedad es insuficiente. Por ello se propone crear la clase Reference tal y como se indica a continuación y sus correspondientes propiedades.

CLASS	Reference
SUBCLASS OF	CreativeWork
DESCRIPTION	This class represents the references or quotes that the authors include in an article or another creative works.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>URL</b>
EXPECTED TYPE	URL
SUBCLASS OR CLASS	Thing
BASIC/OPTIONAL	Optional
DESCRIPTION	Link to additional information about concepts, wikipedia, institutions, persons, etc.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/URL">http://schema.org/URL</a>

PROPERTY	<b>quote</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Reference
BASIC/OPTIONAL	Optional
DESCRIPTION	Text that identified the quoted text in the document.
ALREADY DEFINED? Y/N	No
LINK	

PROPERTY	<b>referenceBiblio</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Reference

<sup>1</sup> Citation: <http://schema.org/citation>

BASIC/OPTIONAL DESCRIPTION	Optional Bibliographic reference that corresponds with the quote identified previously. This property includes all information about the reference with the standards of reference used by the author (APA, Vancouver, ISO 690, etc.).
ALREADY DEFINED? Y/N	No
LINK	
PROPERTY	<b>referenceAim</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Reference
BASIC/OPTIONAL DESCRIPTION	Optional This property indicates the aim of the reference in order to know the intentions of the author with the quote.
ALREADY DEFINED? Y/N	No
LINK	
PROPERTY	<b>referenceArgumentation</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Reference
BASIC/OPTIONAL DESCRIPTION	Optional This property indicates the author argumentation with the quoted text. For example: if the author is for or against the quoted or the reason of the quoted.
ALREADY DEFINED? Y/N	No
LINK	
PROPERTY	<b>referenceDiscussion</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Reference
BASIC/OPTIONAL DESCRIPTION	Optional This property indicates possible discussion about the quoted text.
ALREADY DEFINED? Y/N	No
LINK	

#### Anexo 4. Properties from structure.

PROPERTY	<b>articleAbstract</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL DESCRIPTION	Optional This property represents the abstract in the main language of the article.
ALREADY DEFINED? Y/N	No
LINK	
PROPERTY	<b>articleAbstract_alt</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL DESCRIPTION	Optional This property represents the abstract in the alternative language of the article.
ALREADY DEFINED? Y/N	No
LINK	

PROPERTY	<b>articleBody</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	The actual body of the article.
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/articleBody">http://schema.org/articleBody</a>
PROPERTY	<b>pagination</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Any description of pages that is not separated into pageStart and pageEnd; for example, "1-6, 9, 55" or "10-12, 46-49".
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/pagination">http://schema.org/pagination</a>
PROPERTY	<b>wordCount</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Any description of pages that is not separated into pageStart and pageEnd; for example, "1-6, 9, 55" or "10-12, 46-49".
ALREADY DEFINED? Y/N	Yes
LINK	- <a href="http://schema.org/wordCount">http://schema.org/wordCount</a>
PROPERTY	<b>articleIntroduction</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the introduction part of an article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleAim</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the aims of an article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleMethod</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the methodology part of an article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleAttached</b>
EXPECTED TYPE	Text

SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the attached part of an article, in case that it exists.
ALREADY DEFINED? Y/N	No
LINK	-
<b>PROPERTY</b>	<b>articleSoftware</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the software used in the methodology of the article (in case that it is relevant).
ALREADY DEFINED? Y/N	No
LINK	-
<b>PROPERTY</b>	<b>articleStatisticsTechniques</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies statistics techniques used in the methodology of the article (in case that it is relevant).
ALREADY DEFINED? Y/N	No
LINK	-
<b>PROPERTY</b>	<b>articleSample</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the sample used in the methodology of the article (in case that it is relevant).
ALREADY DEFINED? Y/N	No
LINK	-
<b>PROPERTY</b>	<b>articlePopulation</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the population used in the methodology of the article (in case that it is relevant).
ALREADY DEFINED? Y/N	No
LINK	-
<b>PROPERTY</b>	<b>articleResults</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the results part of an article.
ALREADY DEFINED? Y/N	No
LINK	-
<b>PROPERTY</b>	<b>articleConclusion</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the conclusion part of an article.

ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleGraph</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the graph included in the article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleImagen</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the images included in the article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleTables</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the tables included in the article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleTablesXi</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the content of the columns of the tables included in the article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleTablesYj</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the content of the rows of the tables included in the article.
ALREADY DEFINED? Y/N	No
LINK	-
PROPERTY	<b>articleTablesXiYj</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the content of a specific cell of the tables included in the article.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>articleFootNote</b>
EXPECTED TYPE	Text
SUBCLASS OR CLASS	Article
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the foot notes included in the article.
ALREADY DEFINED? Y/N	No
LINK	-

## RELACIONES

PROPERTY	<b>hasArticle</b>
EXPECTED TYPE ORIGIN	Journal
EXPECTED TYPE DESTINY	Article
SUBCLASS OR CLASS	hasPart
INVERSE OF	isPartOfJournal
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the articles that are included in a journal.
ALREADY DEFINED? Y/N	No
LINK	-

PROPERTY	<b>isPartOfJournal</b>
EXPECTED TYPE ORIGIN	Article
EXPECTED TYPE DESTINY	Journal
SUBCLASS OR CLASS	isPartOf
INVERSE OF	hasArticle
BASIC/OPTIONAL	Optional
DESCRIPTION	Specifies the articles that are included in a journal.
ALREADY DEFINED? Y/N	No
LINK	-

## Anexo 5. Imágenes

Figura 2. Ejemplo de microdatos de Google. Fuente:  
<https://support.google.com/webmasters/answer/176035?hl=es>

```
<div itemscope itemtype="http://data-vocabulary.org/Persona">
  Me llamo <span itemprop="name">Roberto Sánchez</span>, pero me llaman <span
  itemprop="nickname">Rober</span>.
  Esta es mi página web principal:
  <a href="http://www.example.com" itemprop="url">www.example.com</a>
  Vivo en Alcobendas, Madrid, y trabajo de <span itemprop="title">ingeniero</span>
  en <span itemprop="affiliation">ACME Corp</span>.
</div>
```

Figura 3. Ejemplo de Schema.org en Google. Fuente: <http://schema.org/docs/gs.html>

```
<div itemscope itemtype = "http://schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <span>Director: <span itemprop="director">James Cameron</span> (born August 16, 1954)</span>
  <span itemprop="genre">Science fiction</span>
  <a href=" ../movies/avatar-theatrical-trailer.html" itemprop="trailer">Trailer</a>
</div>
```

Figura 4. Ejemplo de rich snippets en una búsqueda en Google. Fuente: Búsqueda en [www.google.com](http://www.google.com)

**Biblioteca Nacional de España** ✓  
[www.bne.es/](http://www.bne.es/) ▾  
 La BNE es la institución bibliotecaria superior del Estado; su misión es recoger y conservar el Patrimonio Bibliográfico de España. Colecciones, servicios en ...  
 4,6 ★★★★★ 74 reseñas de Google · [Escribir una reseña](#)

📍 Paseo de Recoletos, 20-22, 28001 Madrid, España  
 +34 915 80 78 00  
 Has visitado esta página muchas veces. Fecha de la última visita: 1/06/14

<p><b>Catálogos</b> ✓              En el catálogo general se pueden consultar las referencias ...</p>	<p><b>National Library of Spain</b> ✓              La BNE es la institución bibliotecaria superior del ...</p>
<p><b>Hemeroteca Digital</b> ✓              La Hemeroteca Digital forma parte del proyecto Biblioteca Digital ...</p>	<p><b>Servicios</b> ✓              Actividades · Conócenos · Museo · Prensa. Usted está aquí ...</p>
<p><b>Museo</b> ✓              El Museo de la Biblioteca Nacional de España, antiguo Museo del ...</p>	<p><b>Colecciones</b> ✓              Fotografía - Archivo de la Biblioteca - Manuscritos - Carteles</p>

[Más resultados de bne.es »](#)

Figura 5. Ejemplo de código. Fuente: <https://support.google.com/webmasters/answer/146897?hl=es>

```
<div>
  

  <strong>Roberto Sánchez</strong>

  Editor principal de Ediciones ACME

  Calle Mayor 4

  Alcalá de Henares, Madrid 28801

</div>
```

Figura 7. Shotton (2009) Ejemplo de superposición de una figura de la expansión de una enfermedad sobre la imagen del área urbana de la ciudad de Salvador de Bahía que ofrece Google Maps. Así se proporciona al lector una mejor ubicación de la propagación del problema.



Figura 8. Ejemplo de “subrayado” semántico. Cada color pertenece a una clase terminológica (fechas, institución, organismos, etc.)

turn all highlighting off | date | disease | habitat | institution | organism | person | place | protein | taxon

Top | Abstract | Author Summary | Introduction | Methods | Results | Discussion | Supporting Information | Acknowledgements | References | Data Fusion Supplements

### Introduction

At present, one billion of the world's population resides in **slum settlements** [1]. This number is expected to double in the next 25 years [1]. The growth of large urban populations which are marginalized from basic services has created a new set of global health challenges [2],[3]. As part of the Millennium Development Goals [4], a major priority has been to address the underlying poor sanitation and environmental degradation in slum communities which, in turn, are the cause of a spectrum of neglected diseases which affect these populations [2],[3],[5].

**Leptospirosis** is a paradigm for an urban health problem that has emerged due to recent growth of **slums** [6],[7]. The disease, caused by the **Leptospira spirochete**, produces life-threatening manifestations, such as **Weil's disease** and severe **pulmonary hemorrhage syndrome** for which fatality is more than 10% and 50%, respectively [7]-[9]. **Leptospirosis** is transmitted during direct contact with animal reservoirs or water and soil contaminated with their urine [8],[9]. Changes in the urban environment due to expanding slum communities has produced conditions for rodent-borne transmission [6],[10]. Urban epidemics of **leptospirosis** now occur in **cities** throughout the developing world during seasonal heavy rainfall and flooding [6],[11]-[18]. There is scarce data on the burden of specific diseases that affect slum populations [2], however **leptospirosis** appears to have become a major infectious disease problem in this population. In **Brazil** alone, more than 10,000 cases of severe **leptospirosis** are reported each year due to outbreaks in **urban centers** [19], whereas roughly 3,000, 8,000 and 1,500 cases are reported annually for **meningococcal disease**, **visceral leishmaniasis** and **dengue hemorrhagic fever**, respectively, which are other infectious diseases associated with urban poverty [20]-[22]. Case fatality (10%) from **leptospirosis** [19] is comparable to that observed for **meningococcal disease**, **visceral leishmaniasis** and **dengue hemorrhagic fever** (20%, 8% and 10%, respectively) in this setting [20],[23],[24]. Furthermore, **leptospirosis** is associated with extreme weather events, as exemplified by the El Niño-associated outbreak in **Guayaquil** in 1998 [25]. **Leptospirosis** is therefore expected to become an increasingly important slum health problem as predicted global climate change [26],[27] and growth of the world's slum population [1] evolves.

Urban **leptospirosis** is a disease of poor environments since it disproportionately affects communities that lack adequate sewage systems and refuse collection services [6],[10],[11]. In this setting, outbreaks are often due to transmission of a single serovar, **L. interrogans serovar Copenhageni**, which is associated with the **Rattus norvegicus** reservoir [6], [28]-[30]. Elucidation of the specific determinants of poverty which have led to the emergence of urban **leptospirosis** is essential in guiding community-based interventions which, to date, have been uniformly unsuccessful. Herein, we report the findings of a large seroprevalence survey performed in a Brazilian slum community (*favela*). Geographical Information System (GIS) methods were used to identify sources for **Leptospira** transmission in the **slum environment**. Furthermore, we evaluated whether relative differences in socioeconomic status among slum residents contributed to the risk of **Leptospira infection**, in addition to the attributes of the environment in which they reside.

Figura 9. Ejemplo de “cita en contexto”: como puede observarse, se incluye la cita al trabajo Albert I Ko et al. (1999) porque se ha considerado pertinente y además se remite al lector a las partes del trabajo que han motivado la cita.

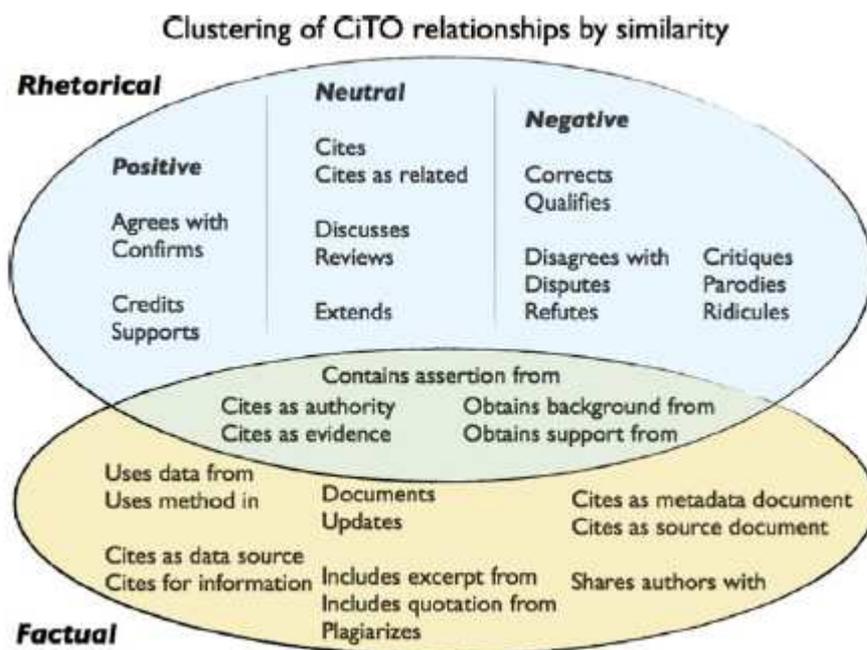
Leptospirosis is a paradigm for an urban health problem that has emerged due to recent growth of slums [6],[7]. The disease, caused by the **Leptospira spirochete**, produces life-threatening

[6] Albert I Ko et al. (1999). Urban epidemic of severe leptospirosis in Brazil *Lancet* **354**: 820–825.

#### Supporting claims:

- **Introduction:** "...the creation of urban slums (favelas) where the lack of basic sanitation favours rodent-borne transmission of leptospirosis..."
- **Discussion:** "...Individuals at highest risk for severe leptospirosis were the urban poor living in the slums on the city's periphery, which lack basic sanitation..."

Figura 10. Object Properties agrupados según su categorización en retórica o de hecho y en según su connotación en positivo, neutral y negativo.



## Anexo 6. Ejemplo de aplicación

```

<div itemscope itemtype="http://schema.org/Periodical">
  <h1 itemprop="journalName">Cuadernos de Gestión de Información</h1>
    <span itemprop='volumeNumber'>2011 </span>
    <span itemprop='hasArticle'>Information Management, today and tomorrow</span>
    <span itemprop='URL'>http://revistas.um.es/gesinfo/issue/view/13051</span>
    <span itemprop="issn">2253-8429</span></span>
    <span itemprop="journalRegularity">Anual</span>
    <span itemprop="editor">Facultad de Comunicación y Documentación, Universidad de Murcia</span>
    <span itemprop="addressCountry">Espanya</span>
    <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
      <span itemprop="addressLocality">Murcia</span>,
      <span itemprop="addressRegion">Murcia</span>
      <span itemprop="postalCode">30100</span>
      <span itemprop="streetAddress">Campus Universitario de Espinardo</span>
    </div>
  </div>
</div>
<div itemscope itemtype="http://schema.org/Article">
  <span itemprop="name">Information Management, today and tomorrow</span>
  <span itemprop="name_alt">Speech of the Workshop in Information Management (1), University of Murcia, 4th September 2011</span>
  Palabras clave: <span itemprop="keywords">Gestión de Información, Mercado de la Información, Gestión de Contenidos Integrada</span>
  KeyWords: <span itemprop="keywords_alt">Information Management, Information Market, Integrated Content Management.</span>
  <span itemprop="articleURI">http://revistas.um.es/gesinfo/article/view/207461</span>
  <span itemprop="articleAbstract">Se intenta ofrecer una panorámica actual de la Gestión de Información a partir de las actividades propias de este campo, intentando llevar a cabo una prospectiva sobre el futuro del perfil profesional de los gestores de Información. Hoy en día existe una relación tangible entre los conceptos de Gestión y Sistemas de Información, así, el gestor de información debe tener una sólida formación en el diseño de sistemas de información y en las TICs. Al mismo tiempo, debe considerarse que las organizaciones necesitan integrar sistemas de gestión de información, tanto a nivel departamental, así como la información del exterior de la organización. Informes y estudios resaltan que pese a la existencia de herramientas para el desarrollo de sistemas integrados, muchas organizaciones carecen de un enfoque unificado en la gestión de información. Esta carencia se debe a la limitación de los recursos económicos para invertir en sistemas adecuados y a la inadecuación de estos sistemas a las necesidades reales de la organización, como por ejemplo, la seguridad física y lógica de los datos, gestión de información no estructurada y descentralización de los procesos de gestión de información. Se identifica una tendencia futura hacia la integración de la Gestión de Documentos, la Gestión de Contenidos Web y la Gestión de Registros. Esta integración se realiza por medio de mecanismos de trabajo en grupo y flujo de trabajo, teniendo en cuenta la digitalización de documentos en formato papel a formato digital. Se concluye resaltando la necesidad de supervisar el mercado de sistemas de información para identificar temas de investigación y mejorar los programas de formación académica.</span>
  <span itemprop="articleAbstract_alt">It aims to provide a panoramic view of the current Information Management, from the point of view of the activities in this field. It tries to make a forecast about the future of the professional profile of the managers of information. Today there is a tangible relationship between the concepts of Information Management and Information Systems. Thus, the information manager should have a strong background in the design of information systems and Information Technology. At the same time, it should be that organizations need to integrate information management systems, both at the departmental level, as well as information from outside the organization. Different reports and studies highlight that despite the existence of tools for developing embedded systems, many organizations lack a unified approach to information management. This deficiency is due, on

```

one hand the limited financial resources to invest in appropriate systems and the other to the inadequacy of these systems to the real needs of the organization, such as physical and logical security of the data, management of unstructured information and decentralization of information management processes. It identifies a future trend towards the integration of Document Management, Web Content Management and Records Management. This integration is done through groupware and workflow mechanisms and taking into account the digitization of paper documents to digital format. It concludes by emphasizing the need to monitoring marketplace of information systems to identify areas of research and improve academic programs.

These are important questions for anyone work-ing in this field today because we know that it changes very rapidly. To know how the subject is perceived today helps us to cope with the immediate changes and to identify gaps and weaknesses in our own information management activities. [...]

5 p.

3.187 palabras

These are important questions for anyone work-ing in this field today because we know that it changes very rapidly. To know how the subject is perceived today helps us to cope with the immediate changes and to identify gaps and weaknesses in our own information management activities. Similarly, guessing how the future is going to turn out can help us to prepare for that future: designing new teaching programmes in academia, and plotting our own career structure in the world at large. [...]

El objetivo principal de esta investigación es.... [...]

Para realizar el estudio se ha utilizado los siguientes métodos estadísticos [...], procesados por [...], con una escala [...], agrupados en [...], etc.

Anexo 1, anexo 2, etc.

Excel 2007

Distribución de Pearson

300 niños y 300 niñas

Niños en edades comprendidas entre 5 y 10 años

Los resultados obtenidos son... [...]

We cannot know the future, but we can try to identify trends, to “follow the money” and see what organizations are buying and what vendors are selling. By monitoring the marketplace for information management systems in this way we can identify research topics of potential interest and perhaps identify commercial partners in the research – an aspect of continuing importance when putting together teams bidding for Euro-pean project funds. From a teaching perspec-tive, this kind of monitoring will also help us to keep our programmes up to date and relevant in the marketplace.

I have said that we cannot know the future, and that is true; but we can try to identify trends and follow them, thereby ensuring that we are at least as up to date as everyone else!

Editor's note: This job is an invited paper. It is adapted from the speech opening of the 1st Workshop of the Master in Information Management in Organizations, an educational innovation project designed to train students in the field of scientific communication. Nobody knows better than Professor Tom Wilson, the true 'guru' of this discipline to be the first speaker in this event and for being the author of the first article of this journal.

Cuadernos de Gestión de Información

</div>

<div itemscope itemtype="http://schema.org/Person">

Tom Wilson

University of Sheffield

Profesor Emerito

<http://sheffield.ac.uk/is/staff/wilson>

[wilsontd@gmail.com](mailto:wilsontd@gmail.com)

</div>

```

<div itemscope itemtype="http://schema.org/Reference">
    El concepto de information management, <span item-
prop="URL">https://en.wikipedia.org/wiki/Information_management</span>
    <span itemprop="authorInstitution">University of Sheffield</span>
    <span itemprop="quote">The notion of content means that this term covers all media. Perhaps the most
rigorous definition of enterprise content management comes from the Gartner consultancy organization, which uses
the following criteria to define a vendor of ECM systems (Enterprise... 2011):</span>
    <span itemprop="referenceBiblio">Enterprise Content Management. (2011). What is enter-prise content
management? Retrieved 25 July, 2011 from http://www.contentmanager.eu.com/ecms.htm (Archived by WebCite@
at http://www.webcitation.org/62hiDeFXh)</span>
    <span itemprop="referenceAim">The purpose of this aim is to highlight this acronym of other denomina-
tions that are used to introduce Document Management Systems into organizations. El propósito de esta cita es
resaltar este acrónimo del resto de denominaciones que se emplean para presentar a los sistemas de gestión de
documentos en una organización</span>
    <span itemprop="referenceArgumentation">The ECM acronym is perhaps the most widely used in scientif-
ic and technical literature . Its scope is broader .El acrónimo ECM es quizá el más utilizado en la literatura científica
y técnica especializada. Su alcance es más amplio. Fuente: https://www.laserfiche.com/ecmblog/17-acronyms-you-
need-to-know-before-researching-ecm/ </span>
    <span itemprop="referenceDiscussion">Controversias o discusiones sobre la veracidad, originalidad, etc.
de la cita. Texto subjetivo.</span>
</div>

```