

Publicación Semántica de Información Geográfica

Francisco J. Lopez-Pellicer¹, Jesús Barrera², Rodolfo Rioja², Miguel Usón¹,
Pedro R. Muro-Medrano¹, y F. Javier Zarazaga-Soria¹

¹ Universidad de Zaragoza {fjlopez,muson,prmuro,javy}@unizar.es

² GeoSpatiumLab S.L {jesusb,rodolfo}@geoslab.com

Resumen El mundo de las Infraestructuras de Datos Espaciales ha progresado muy deprisa en los últimos años de la mano de importantes avances tecnológicos y de conocimiento. En este crecimiento los servicios de publicación y búsqueda de información geográfica han tenido un papel preponderante gracias a los estándares de catálogo promovidos por Open Geospatial Consortium. No obstante, estas aproximaciones se apoyan en modelos y paradigmas que, aunque suficientemente probados y útiles para la resolución de problemas de integración en el contexto geoespacial, no fueron concebidos para la integración de las Infraestructuras de Datos Espaciales con otras comunidades tecnológicas. En los últimos años estas comunidades tecnológicas han creado nuevas formas de utilizar la información geográfica, muchas de las cuales han sido incorporadas a las Infraestructuras de Datos Espaciales para permitir la interoperabilidad. Esto es evidente en el caso de la Neo Geografía desarrollada alrededor de portales como Google Map y Open Street Map que ha impulsado la adopción generalizada por las Infraestructuras de Datos Espaciales de los servicios de mapas tileados y de los formatos de publicación GeorSS y KML. Ahora, el reto es integrarse con la Web Semántica pero desde una perspectiva que capture las peculiaridades de la información espacial. Este trabajo presenta una aproximación que posibilita la integración de metadatos geográficos publicados en las Infraestructuras de Datos Espaciales con los desarrollos y sistemas que operan en el contexto de la Web Semántica. Las aportaciones de este trabajo son dos: un esquema RDF basado en Dublin Core para la publicación de metadatos geográficos y una solución arquitectural que permite que el contenido de los catálogos de metadatos sea un recurso más de Web Semántica.

1. Introducción

El término “Infraestructuras de Datos Espaciales” (IDE, en inglés Spatial Data Infrastructure, SDI) suele utilizarse para denotar el conjunto básico de tecnologías, políticas, estándares y acuerdos entre instituciones destinados a facilitar la disponibilidad y el acceso a información espacial. Las IDEs proporcionan una base para la búsqueda, evaluación y explotación de la información espacial para usuarios y proveedores de todos los niveles de la administración, el comercio y la academia, así como a las organizaciones sin ánimo de lucro y a los ciudadanos. En Europa, avanzando un paso más, la Comisión Europea ha impulsado la

iniciativa INSPIRE³ (“INfrastructure for SPatial InfoRmation in Europe”) con el objetivo de establecer las bases, tanto técnicas como políticas, para poder crear una IDE Europea basada en la federación de las IDEs de cada uno de los países miembros. Dicha iniciativa se ha materializado en la Directiva Europea INSPIRE. Una de las primeras decisiones del grupo de trabajo de INSPIRE fue la elección de un modelo conceptual hacia el cual enfocar la iniciativa. El modelo elegido fue el Modelo de Referencia del proyecto Digital Earth (ver Figura 1) dado su alto nivel de simplicidad y síntesis. La arquitectura que plantea INSPIRE para las IDEs se basa en cuatro grupos principales de componentes: Repositorios de información (datos y metadatos), Catálogos y Servicios de catálogo, Redes de Servicios de Geoprocesamiento y Aplicaciones de usuario.

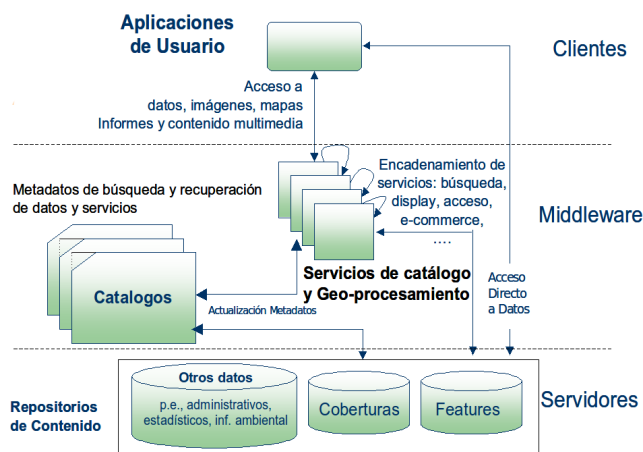


Figura 1. Modelo conceptual de INSPIRE

Al amparo de la transposición de INSPIRE a las legislaciones de cada país se están poniendo en marcha numerosas iniciativas que ofertan servicios de metadatos y catálogos según las arquitecturas y propuestas establecidas por Open Geospatial Consortium (OGC). Estas aproximaciones resultan de difícil encaje con las prácticas que en materia de Web Semántica se están poniendo en marcha. Una de las prácticas más relevantes es la denominada Linked Data⁴. Es un conjunto de buenas prácticas para publicar, compartir y conectar datos, información y conocimiento utilizando URIs que se resuelven a documentos RDF. Según [1], en mayo de 2009 la cantidad aproximada de información publicada bajo esta práctica asciende a 4.700 millones de triples RDF interconectados por 142 millones de enlaces.

³ <http://www.ec-gis.org/inspire/>

⁴ <http://linkeddata.org/>

En este trabajo se presenta una aproximación que posibilita la integración de los metadatos geográficos publicados en la Web por iniciativas surgidas al amparo de la implantación de INSPIRE con los desarrollos y sistemas que operan según las prácticas de la Web Semántica. Para ello se plantea el uso de RDF (Resource Description Framework) como herramienta de base. Este trabajo centra sus aportaciones en dos aspectos básicos. Una propuesta de un esquema RDF basado en Dublin Core que posibilita la publicación automática de metadatos geográficos como parte de la Web Semántica, y una solución arquitectural que permite el desarrollo de servicios de búsqueda sobre las informaciones así publicadas.

El resto del trabajo se estructura como sigue. La sección 2 presenta los diferentes enfoques que existen para describir semánticamente los recursos geográficos. La sección 3 incide en cómo se publica esta información en el ámbito geográfico utilizando catálogos. En la sección 4 se describe el esquema RDF propuesto para publicar esta información como parte de la Web Semántica. En la sección 5 se articula una solución arquitectural para la publicación de los contenidos de los catálogos. Finalmente, en las conclusiones se revisan las ideas presentadas y se plantean los siguientes trabajos a realizar.

2. Anotación semántica de recursos geográficos

La anotación semántica de recursos geográficos en el ámbito de una IDE se sustenta sobre el concepto de metadato. Los metadatos se definen comúnmente como “datos estructurados acerca de los datos” o “datos que describen los atributos de un recurso” o más simplemente “información acerca de los datos”. Los metadatos son la información y la documentación que permiten que los datos sean bien entendidos, compartidos y explotados de manera eficaz por todo tipo de usuarios a lo largo del tiempo. Tal como se menciona en [2], los metadatos geográficos ayudan a las personas involucradas en el uso de información geográfica a encontrar los datos que necesitan y a utilizarlos apropiadamente.

Dentro del mundo de la información geográfica se han ido definiendo recomendaciones para la creación de metadatos, cuya finalidad principal es proporcionar una estructura “jerárquica y concreta”. Aunque existen diferentes aproximaciones y propuestas en materia de estándares de metadatos geográficos, en el ámbito de las IDEs existe ya un claro consenso en la adopción del estándar ISO 19115:2003 como modelo de referencia. Aunque esta norma está principalmente orientada a la catalogación de conjuntos de datos geográficos en formato digital, puede ser extendida a otras formas de datos como mapas, documentos textuales, datos no geográficos o servicios.

De especial interés resultan aquellos que se han centrado en fomentar la interoperabilidad sobre la base del uso de Dublin Core como estándar de referencia [3,4]. La Iniciativa de Metadatos Dublin Core⁵, simplemente Dublin Core o DC, es actualmente el modelo de metadatos más aceptado para describir, recuperar e intercambiar información electrónica, independientemente del dominio

⁵ <http://dublincore.org>

científico o disciplinar. En sus orígenes (marzo 1995), Dublin Core surge como un modelo de metadatos dirigido a la descripción embebida en HTML por parte de los autores de los recursos, para una recuperación más eficaz y cualificada en motores de búsqueda en la Web, liderando el desarrollo de metadatos estructurales para la recuperación de información en Internet [5]. Con el tiempo, Dublin Core ha ido evolucionando hacia un formato de registro para el intercambio de información y a un estándar básico para la interoperabilidad entre repositorios de información científica, gracias a la integración de Dublin Core con el protocolo OAI-PMH, a la versatilidad del esquema y al nivel de estandarización formal que ha adquirido.

Respecto a la utilización de XML como lenguaje de representación de metadatos en el contexto de los metadatos geográficos, cabe destacar que hay un consenso bastante generalizado. Tal como se menciona en [2], los sistemas de catalogación de metadatos deben soportar tres formatos de metadatos: el formato de implementación, el formato de exportación o codificación, y el formato de presentación. Cabe destacar la especificación técnica ISO 19139:2007, la cual define la forma de serializar en XML modelos UML conformes con las normas ISO de la serie 19100.

Si hablamos de lenguajes de representación de metadatos en contextos más generales, no podemos olvidarnos del uso extensivo de RDF. El uso de RDF como lenguaje de representación de metadatos se debe a su utilización para expresar modelos de metadatos muy extendidos como Dublin Core y a su utilización como lengua franca de la Web Semántica. RDF es una recomendación W3C para el modelado e intercambio de metadatos, que se expresa en formato XML. La mayor ventaja de RDF es su flexibilidad. RDF no es un estándar de metadatos, es un metamodelo que permite la definición de esquemas de metadatos o la combinación de ellos. El núcleo básico de RDF ofrece un modelo simple para describir relaciones entre recursos en términos de propiedades con un nombre asociado y una serie de valores. Para la declaración y la interpretación de esas propiedades, RDF Schema (RDFS) proporciona un amplio conjunto de constructores que permiten definir y restringir la interpretación de los vocabularios en una comunidad de información particular. RDF permite la codificación, el intercambio y la reutilización de metadatos estructurados. Permite la interoperabilidad de los metadatos por medio del diseño de mecanismos que soporten convenciones comunes de semántica, sintaxis y estructura.

3. Publicación de metadatos geográficos

Una de las especificaciones líderes para la interoperabilidad entre catálogos es la especificación de Servicios de Catálogo de OGC (OGC Catalogue Services Implementation Specification 2.0.2:2007⁶) que describe el conjunto de interfaces de las operaciones que soportan la gestión, el descubrimiento, y el acceso a los recursos de información geográfica [6]. Un aspecto importante en esta especificación es que se proporcionan diferentes perfiles de implementación de las

⁶ <http://www.opengeospatial.org/standards/cat>

interfaces de acuerdo a la plataforma y protocolo de transporte que se va a utilizar. En particular, esta especificación de OGC define perfiles de implementación para CORBA, el protocolo Z39.50 y un protocolo definido por OGC sobre HTTP conocido como Catalog Services for the Web (CSW). Esta propuesta de OGC se ha constituido en el elemento clave que ha permitido desarrollar sistemas que posibilitan la localización y acceso a recursos de información geográfica en el marco de las IDEs.

Sin embargo, esta aproximación es de difícil encaje con las iniciativas que en la Web Semántica se están poniendo en marcha como es el caso de Linked Data⁷. Posiblemente, uno de los mayores inconvenientes radica en que la especificación de catálogo OGC ha sido concebida para permitir tanto la publicación, como la búsqueda de metainformación, mediante protocolos propios que resultan bastante alejados de aproximaciones más generales. Por ejemplo, los “motores de búsqueda” de la Web no han incorporado a sus heurísticas de búsqueda de información el acceso a los catálogos OGC por su especificidad. El resultado más inmediato es que todos los metadatos publicados en las diferentes IDE (el Instituto Geográfico Nacional tiene publicados más de 60.000 registros a través del catálogo de la IDE Española, IDEE⁸) han pasado a formar parte de la Web profunda (Deep Web) [7].

4. Modelo RDF para la publicación de metadatos geográficos

Tomando como punto de partida la especificación RDF de Dublin Core, se propone crear un perfil de aplicación que permita aportar una visión de la metainformación más amigable desde el punto de vista de la Web Semántica. Dublin Core ha demostrado poder ofrecer una visión de la metainformación geográfica expresada en ISO 19115:2003 de gran utilidad para posibilitar la interacción con otros sistemas de contextos en los que la carga geográfica es menor [3]. Tomando como base esta propuesta y la aproximación realizada por el CEN [8], se ha establecido el siguiente emparejamiento (Tabla 1) entre Dublin Core e ISO 19115:2003.

Como complemento, se ha fijado un campo de RDF que es el que posibilitará la navegación hacia el metadato completo. Éste será `dc:relation` y sus valores se fijan a partir del lenguaje de interrogación y la interfaz del servicio de catálogo OGC CSW 2.0.2. Dentro de las diferentes alternativas que esta interfaz permite, se ha seleccionado la operación de tipo GET del protocolo `http` por resultar la de más fácil comprensión humana. De este modo, el acceso al metadato original se efectuaría de acuerdo al siguiente patrón:

```
http://<host>:<puerto>/<app>?  
request=GetRecordById&Id=<Identificador>&elementSetName=full
```

⁷ <http://linkeddata.org/>

⁸ <http://www.ideo.es>

ISO 19115:2003	Elemento DC
MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.title	dc:title
MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.CitedResponsibleParty > CI_ResponsibleParty.OrganisationName [role="originator"]	dc:creator
MD_Metadata.identificationInfo > MD_DataIdentification.topicCategory	dc:subject
MD_Metadata.identificationInfo > MD_DataIdentification.descriptiveKeywords > MD_Keywords [type="theme"]	dc:subject
MD_Metadata.identificationInfo > MD_DataIdentification.abstract	dc:description
MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.CitedResponsibleParty > CI_ResponsibleParty.OrganisationName [role="publisher"]	dc:publisher
MD_Metadata.identificationInfo > MD_DataIdentification.credit	dc:contributor
MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.date > CI_Date	dc:date
MD_Metadata.hierarchyLevel	dc:type
MD_Metadata.distributionInfo > MD_Distribution.distributionFormat > MD_Format.name	dc:format
MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource.linkage > URL	dc:identifier
MD_Metadata.dataQualityInfo > DQ_DataQuality.lineage > LI_Lineage.source > LI_Source.description	dc:source
MD_Metadata.identificationInfo > MD_DataIdentification.language	dc:language
N/A	dc:relation
MD_Metadata.identificationInfo > MD_DataIdentification.extent > EX_Extent.geographicElement > EX_GeographicBoundingBox	dc:coverage
N/A	dc:rights

Cuadro 1. Correspondencia ISO 19115:2003 - Dublin Core

Donde <host>:<puerto>/<app> representan la URL de acceso al servicio de catálogo compatible OGC 2.0.2, e <Identificador> es el identificador único del registro dentro del catálogo. Seguidamente se muestra un ejemplo para el caso del catálogo del la IDEE:

```
http://www.idee.es/csw/servlet/cswservlet?
request=GetRecordById&Id=ESIGNMAPASRELIEVESERIE200701180000&
elementSetName=full
```

El resultado es una operación del tipo `GetRecordByIdResponse` que debe ser procesada para poder recuperar el metadato completo. A continuación se muestra un ejemplo de un metadato antes de extraer la información especificada conforme a la norma ISO 19139:2007.

```
<gmd:MD_Metadata [...]>
  <gmd:fileIdentifier>
    <gco:CharacterString>af88922d-f792-11dd-80b9-000475efbcdc</
gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      <gmd:citation>
        <gmd:CI_Citation>
          <gmd:title>
            <gco:CharacterString>Estuario del Guadalquivir</gco:CharacterString>
          </gmd:title>
          [...]
          <gmd:pointOfContact>
            <gmd:CI_ResponsibleParty>
              <gmd:individualName>
                <gco:CharacterString>Manuel Pérez López</gco:CharacterString>
              </gmd:individualName>
            [...]
          <gmd:abstract>
            <gco:CharacterString>Descripción [...]</gco:CharacterString>
          </gmd:abstract>
        [...]
      </gmd:MD_DataIdentification>
    </gmd:identificationInfo>
  </gmd:MD_Metadata>
```

El siguiente listado muestra la descripción del metadato en RDF obtenida tras aplicar una hoja de estilo conforme con los criterios definidos en la tabla 1.

```
<rdf:Description rdf:about=" http://www.myIDEE.org/RDF_metadata/af88922d-
f792-11dd-80b9-000475efbcdc">
  <dc:title xml:lang="es">Estuario del Guadalquivir</dc:title>
  <dc:creator xml:lang="es">Manuel Pérez López</dc:creator>
  <dc:subject>geoscientificInformation</dc:subject>
  <dc:description xml:lang="es">Descripción de la comunidad macrobentónica del
estuario del Guadalquivir</dc:description>
  <dc:identifier>af88922d-f792-11dd-80b9-000475efbcdc</dc:identifier>
  [...]
  <rdfs:seeAlso>
    <rdf:Description rdf:about="http://www.idee.es/csw/servlet/
cswservlet?request=GetRecordById&Id=af88922d-f792-11dd-80b9-
000475efbcdc&elementSetName=full">
      <dc:format>
        <dc:MediaType>
          <dc:identifier>ISO/TS 19139:2007</dc:identifier>
        </dc:MediaType>
      </dc:format>
    </rdf:Description>
  </rdfs:seeAlso>
</rdf:Description>
```

El documento, que describe un recurso con URI http://www.myIDEE.org/RDF_metadata/af88922d-f792-11dd-80b9-000475efbcdc, será publicado se-

gún las convenciones de Linked Data [9,10]. Además el mismo documento describe cómo se puede obtener la versión original del recurso utilizando la propiedad `rdfs:seeAlso`.

5. Arquitectura básica para la publicación y búsqueda semántica

El establecimiento de un emparejamiento directo entre la visión RDF propuesta y el modelo de metadatos en ISO 19115:2003 posibilita que la creación del primero se pueda llevar a cabo de manera automática una vez que se ha identificado la fuente de información (ésta debe ser un catálogo que cumpla con la interfaz OGC CSW 2.0.2). Para ello basta con construir un componente que se encargue de ejecutar el algoritmo que se presenta a continuación (componente `RDF_Publisher` de la figura 2).

```
response = csw::GetRecords(elementName=dc:identifier)
each(record_brief in response.searchResults) do {
  reponseById = csw::GetRecordById(record_brief.dc:identifier)
  ISO19115_record = reponseById.record
  RDF_record = ISO19115_to_RDF(ISO19115_record)
  store(RDF_record)
}
```

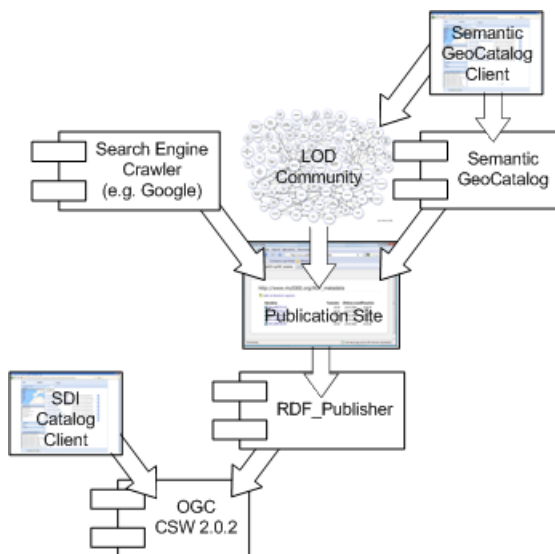


Figura 2. Mecanismo de publicación

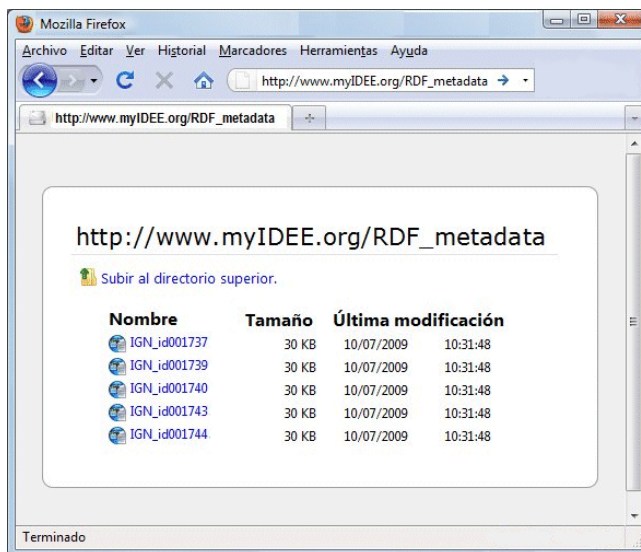


Figura 3. Recursos publicados

La publicación se llevará a cabo sobre un directorio Web de acuerdo al ejemplo que se muestra en la figura 3. Este mecanismo de publicación abre las puertas a la incorporación de las IDEs a comunidades como Linked Data. Adicionalmente es posible abordar la construcción de servicios semánticos de búsqueda de información geográfica usando como base tecnología de búsqueda, por ejemplo, el motor de Sesame⁹.

6. Conclusiones

Este trabajo ha presentado una propuesta de un esquema RDF basado en Dublin Core que posibilita la publicación automática de metainformación ya generada en el contexto de las IDEs. Adicionalmente, se ha propuesto una solución arquitectural que posibilita el desarrollo de servicios de búsqueda sobre las informaciones publicadas. De este modo se consigue un doble objetivo de facilitar la incorporación de las Infraestructuras de Datos Espaciales al mundo de la Web Semántica, así como abrir estas infraestructuras para que sus contenidos dejen de ser parte de la Web Profunda.

La aproximación presentada busca como objetivo posibilitar una publicación inmediata de la metainformación que está actualmente disponible. El siguiente paso a abordar será la explotación de las relaciones que en estos momentos la información en ISO 19139:2007 contiene. Vinculaciones entre el metadato y los tesauros y ontologías utilizados para su creación, información del publicador de los datos, o la pertenencia a colecciones de recursos son solamente algunos de los

⁹ <http://www.openrdf.org/>

aspectos candidatos a ser trasladados el modelo RDF con el fin de aprovechar su potencia en materia de gestión y explotación de relaciones.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Gobierno de España a través de los proyectos CENIT 2008-1030 y TIN2007-65341; y el Gobierno de Aragón a través del proyecto PI075/08. El trabajo de Rodolfo Rioja ha sido co-financiado por el Ministerio de Ciencia y Tecnología (programa Torres Quevedo ref. PTQ06-2 0790).

Referencias

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data – the story so far. *International Journal on Semantic Web and Information Systems* (2009) To appear. See <http://linkeddata.org/docs/ijswis-special-issue>.
2. Nebert, D.D.: Developing Spatial Data Infrastructures: The SDI Cookbook [online]. Technical report, Global Spatial Data Infrastructure (2004) Version 2.0. Available from: <http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf>.
3. Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R., Muro-Medrano, P.R.: Metadata standard interoperability: application in the geographic information domain. *Computers, Environment and Urban Systems* **28**(6) (2004) 611–634
4. Tolosana-Calasanz, R., Nogueras-Iso, J., Bejar, R., Muro-Medrano, P.R., Zarazaga-Soria, F.J.: Semantic interoperability based on dublin core hierarchical one-to-one mappings. *International Journal of Metadata, Semantics and Ontologies* **1**(3) (2006) 183–188
5. Weibel, S.L., Koch, T.: The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions. *D-Lib Magazine* [online] **6**(12) (2000) Available from: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>.
6. Nogueras-Iso, J., Zarazaga-Soria, F.J., Muro-Medrano, P.R.: Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)
7. Bergman, M.K.: The Deep Web: Surfacing hidden value. White paper, BrightPlanet, Sioux Falls, SD, USA. (2001) This white paper appeared in *The Journal of Electronic Publishing* from the University of Michigan (<http://www.press.umich.edu/jep/07-01/bergman.html>), July 2001.
8. Zarazaga-Soria, F.J., Nogueras-Iso, J., Ford, M.: Mapping between Dublin Core and ISO 19115, *Geographic Information – Metadata*. CWA 14857, CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core (September 2003)
9. Bizer, C., Cyganiak, R., Heath, T.: How to Publish Linked Data on the Web [online]. Available from: <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/> (2007)
10. Phipps, J., Berrueta, D.: Best practice recipes for publishing RDF vocabularies. W3C working draft, W3C (January 2008) Available from: <http://www.w3.org/TR/2008/WD-swbp-vocab-pub-20080123/>.