# Exposing CSW catalogues as Linked Data

Francisco J. Lopez-Pellicer, Aneta J. Florczyk, Javier Nogueras-Iso,
Pedro R. Muro-Medrano and F. Javier Zarazaga-Soria

Department of Computer Science and Systems Engineering
Universidad de Zaragoza, Spain
{fjlopez,florczyk,jnog,prmuro,javy}@unizar.es

## Abstract

The OpenGIS Catalogue Services (CS) specification defines a set of abstract interfaces for the discovery, access, maintenance and organization of metadata repositories of geospatial information and related resources in distributed computing scenarios, such as the Web. The CS specification also defines a HTTP protocol binding, which is called "Catalogue Services for the Web" or CSW. A fair description of CSW is a remote catalogue interface over the HTTP protocol, but not over the architecture of the mainstream Web where search engines are the users' gateway to information. This paper identifies some aspects of CSW that difficult the findability of metadata in the Web, and hence, the discovery of resources. This paper also presents a toolkit that exposes as Linked Data the content of metadata repositories offered through CSW with the purpose of improving the discovery of metadata records in search engines.

## 1 Introduction

A catalogue is a system that helps publish, query and retrieve items of information in a systematic way. The OpenGIS Catalogue Services (CS) specification provides discovery, access, maintenance and organization in-

terfaces for metadata catalogues of geospatial information and related resources, and allows users to find information in distributed systems (Nebert et al. 2007). The CS specification defines a HTTP protocol binding named Catalogue Services for the Web (CSW). Spatial Data Infrastructures (SDIs) use CSW as one of the gateways to their geospatial resources. An example of the relevance of CSW is the recommendation issued in the context of INSPIRE by the INSPIRE Network Services Drafting Team (2008) to SDIs in European Union to derive the base functionality of discovery services from the ISO profile of CSW defined in Voges et al. (2007).

However, CSW is not properly prepared for the mainstream Web where search engines are the users' gateway to information. Some features of the infrastructure for the discovery of information in the mainstream Web are:

- **Search engines try to browse and index Deep Web databases**. Surfacing Deep Web content is a research problem that concerns the search engine community since its description by Bergman (2001). The term Deep Web refers to the database content that is behind Web forms and applications. From this point of view, SDI metadata repositories are hidden behind catalogue applications; therefore, SDI metadata is part of the Deep Web. Hence, the findability in search engines depends on the success of crawling processes that require the analysis of the Web interface, and then the automatic generation of queries.

- **Applications ask for Linked Data**. The Linked Data community, which has blossomed in the last three years, promotes a Web of data based on the architectural principles of the Web (Bizer et al., 2008). Linked Data is a set of best practices to publish, share and connect data, information and knowledge using URIs that are resolved to Resource Description Framework (RDF) documents. RDF is a W3C recommendation for modelling and exchanging metadata (Miller et al., 2004). According to Bizer et al. (2009), in May 2009, the approximate amount of information released under this practice amounts to 4,700 million of RDF statements connected by 142 million of links and a growing number of relevant nodes.

- **Evolution of metadata vocabularies.** Well known metadata vocabularies have evolved to models based on RDF with an emphasis in the linking of metadata descriptions. The abstract data models of Dublin Core Metadata Initiative (DCMI) and the Open Archive Initiative (OAI) have evolved side of the RDF data model. This process has resulted in abstract models based on the RDF data model (Nilsson et al. 2008; Lagoze et al. 2008) that empathizes the use (and reuse) of entities rather than plain literals as the value of properties. This evolution enables the

effective hyperlink of metadata and traverse queries using query languages and protocols, such as SPARQL (Seaborne et al., 2008).

This paper identifies three drawbacks in CSW in relation with the depicted scenario:

- The protocol is hard to crawl by standard Deep Web crawlers.
- The remote procedure call style for accessing metadata is orthogonal to the linked data approach.
- The support of association links between metadata in queries is limited.

The most relevant consequence is that metadata published by SDIs have become part of the Deep Web content not surfaced by search engines. Therefore, the resources offered by SDIs are more difficult to be discovered in the mainstream gateway to information.

This paper proposes republishing CSW catalogues as Linked Data to make their content easily accessible for search engines and machine-to-machine applications aware of the Web of data. This paper also proposes the CSW2LD toolkit for republishing SDI metadata. The mission of the toolkit is to expose the content of standard CSW catalogues as Dublin Core metadata conform to the RDF data model and the principles of Linked Data.

The structure of this paper is as follows. Section 2 identifies related work. Section 3 presents CSW and the above drawbacks. Section 4 discusses the general approach that the CSW2LD toolkit follows to map catalogue information models to the RDF abstract data model. Section 5 describes the CSW2LD toolkit for publishing metadata. Finally, the conclusions review the ideas presented and sets the next research goals.

## 2 Related work

This section presents related work in the geographic information domain about the use of semantic descriptions and search engines in catalogue systems, and presents some publishing tools of the Linked Data community related with CSW2LD.

Egenhofer (2002) proposes the use of the Semantic Web to face problems of semantic heterogeneity in geo-resource discovery. Studies on geospatial catalogue usability, such as Larson et al. (2006), identify as a potential improvement the use of semantic techniques for knowledge description and discovery. Some authors have considered the use of search engines. For example, the approach of Oates et al. (2007) is to provide metadata encoded in KML about resources and make the KML files discoverable through Google search.

There is a variety of Linked Data publishing tools. Many of them are services that publish the content of relational databases as Linked Data (Bizer et al., 2009). Large geographical information providers are investigating how Linked Data and other Semantic Web technologies can assist the diffusion of geographic data. For example, Ordnance Survey is developing datasets in RDF and publishing them using the Linked Data principles (Goodwin et al. 2009). The Linked Data community has an increasing interest in the geospatial databases. In particular, the LinkedGeoData project maps OpenStreetMap data into linked data (Auer et al., 2009), and the GeoNames ontology describes the content of the GeoNames database (Vatant et al., 2007).

Haslhofer et al., (2008) is the only directly related work found in the literature but it does not belong to the geo community. It proposes a server that wraps the metadata protocol for digital libraries OAI-PMH (Lagoze et al. 2002), exposes metadata as Linked Data and provides metadata access via a SPARQL endpoint.

## 3 Catalogue Services for the Web

CSW defines the interaction between a catalogue client and a CSW server that exposes the contents of an opaque catalogue. Request and response messages must conform to the CSW specification or to application profiles derived from it.

### 3.1 The context

CSW is the HTTP protocol binding of the OpenGIS Catalogue Services (CS) specification. The CS specification defines interfaces for the management, the discovery and the access to collections of metadata about geospatial information resources. The management interface supports the ability to administer and organize collections of metadata in the local storage device. The discovery interface allows users to search within a catalogue and provides a minimum query language. Finally, the access interface facilitates access to metadata items previously found with the discovery interface. The CS specification also defines an abstract information model that includes a core set of shared attributes, a common record format that defines metadata elements and sets, and a minimal query language called CQL.

Additionally to the HTTP protocol binding, the CS specification includes binding implementation guidance for the application protocols Z39.50, a pre-Web protocol widely used in digital libraries, and CORBA/IIOP, a remote procedure call specification *in a niche of relative obscurity* (see Henning 2008).

## 3.2 Request and response example

CSW is quite complex. For example, the operation `GetRecordById` fetches representations of metadata records using the identifier of the metadata in the local metadata repository. The parameter `elementSet-Name`, if used, establishes the amount of detail of the representation of the source record. Each level of detail specifies a predefined set of record elements that should be present in the representation. The predefined set name `full` represents all the metadata record elements. By default, the operation `GetRecordById` returns a metadata record representation that validates against the information model of the metadata repository. The parameter `outputSchema` allows user agents to request for a response in a different information model, and the CSW implementations must support at least the representation of the common information schema defined in the CSW standard.

Figure 1 shows a sample `GetRecordById` request for a metadata record available in IDEE, the SDI of Spain, and the corresponding response. The request URI identifies the location of the CSW server, the operation, the identification of the metadata record (parameter `id`), the amount of detail of the representation (parameter `elementSetName`), and the output schema (parameter `outputSchema`). The XML response consists of a `<GetRecoredByIdResponse>` element that contains a record that conveys the information of the source metadata. When a `<SummaryRe-cord>` element is the conveyor, the retrieved representation contains a summary of the original metadata record. The value of the output schema identifies the subset that conforms to the common information schema defined in the CSW standard.

Request:

```
GET/csw/servlet/cswservlet?request=GetRecordById&id=
ESIGNMAPASRELIEVESERIE200701180000&elementSetName
full&outputSchema=http://www.opengis.net/cat/csw/
2.0.2 HTTP/1.1
Host: www.idee.es
```

Response:

```
HTTP/1.x 200 OK
Content-Type: application/xml;charset=ISO-8859-1
...

<?xml version = '1.0' encoding = 'ISO-8859-1'?>
<GetRecordByIdResponse
 xmlns="http://www.opengis.net/cat/csw/2.0.2"
 xmlns:dc="http://purl.org/dc/elements/1.1/"
 xmlns:dcterms="http://purl.org/dc/terms/"
 ...>
 <SummaryRecord>
  <dc:title>Mapas en Relieve</dc:title>
  <dc:identifier>
   ESIGNMAPASRELIEVESERIE200701180000
  </dc:identifier>
  ...
  <dc:format>PVC</dc:format>
  <dc:subject>elevation</dc:subject>
  <dc:subject>imageryBaseMapsEarthCover</dc:subject>
  <dc:type>dataset</dc:type>
  ...
  <dcterms:spatial>COUNTRIES.SPAIN</dcterms:spatial>
  ...
  <dcterms:spatial>
   northlimit=43.8;
   southlimit=37.83;
   westlimit=-9.32;
   eastlimit=0.72;
  </dcterms:spatial>
 </SummaryRecord>
</GetRecordByIdResponse>
```

**Fig. 1:** Sample CSW `GetRecordById` request and response.

### 3.3 Identified drawbacks

CSW is undoubtedly useful to enable the discovery and access to geographic information resources within the geographic community (Nogueras et al, 2005). However, it presents the following drawbacks:

- **Mismatch with operational model of Deep Web crawlers**. The search engines have developed several techniques to extract information from Deep Web databases without previous knowledge of their interfaces.

The operational model for Web crawlers, described in Raghavan (2001), based on (1) form analysis, (2) query generation and (3) response analysis is widely accepted. It models queries as functions with $n$ named inputs $X_1..X_n$. where the challenge is to discover the possible values of these named inputs that return most of the content of the database. This approach is suitable for CSW HTTP GET requests. However, the constraints are encoded in a single named input as a CQL string (see Nebert et al. 2007), or an XML Filter (Vretanos, 2004). This characteristic is incompatible with the query model of the Deep Web crawlers. Researchers working for search engines, such as Google (see Madhavan et al. 2008), discourage the alternative operational model that consists in the development of ad-hoc connectors as non-sustainable in production environments.

- **RPC approach to access metadata.** Metadata repositories are behind a proprietary RPC from the point of view of other communities. CSW does not define a simple Web API to query and retrieve metadata. Some communities that potentially can use CSW are accustomed to simple APIs and common formats. For example, many geo mashups and related data services (see Turner, 2006) use Web APIs to access and share data built following the REST architectural style (Fielding, 2000) and the vision of Berners-Lee et al. (2001) about the Semantic Web. These APIs are characterized by the identification of resources by opaque URIs, semantic descriptions of resources, stateless and cacheable communication, and uniform interface based on the verbs of the HTTP protocol in opposition to the RPC style.

- **Queries limited to same record properties**. The field based query model of the CS specification does not define the support for associations in the CQL or Filter syntax. CSW application profiles may describe the support of associations. For example, the ISO application profile (Voges et al. 2007) supports the linkage between services and data instances. However, the linkage is based in the equality of literal values of properties such as `MD_Identifier.code`, and the profile does not extend the CQL and the XML Filter syntax. Hence, association queries require being decomposed in parts. For example, in a metadata repository where metadata records about data and services instances are linked, a query that returns the services that serves data created by a producer requires (1) to query initially about the data created by this producer, (2) to retrieve their identifiers, and then, (3) to query about servers that serve data with these identifiers.

## 4 Mapping SDI metadata to RDF

The annotation of geographic resources is based on the concept of metadata. Metadata are information and documentation that enable data to be understood, shared and exploited effectively by all users over time. As mentioned in Nebert (2004), the geographic metadata help geographic information users to find the data they need and determine how to use.

One of the main goals of the creation of geographic metadata is the reuse of organization's data by publishing its existence through catalogue metadata records that conveys information about how to access and use the data (FGDC, 2000). In the context of European SDI, the information is conveyed as ISO 19115 / ISO 19119 metadata records represented in XML. However, RDF is the lingua franca for the metadata interchange in the Semantic Web. The publication of SDI metadata in the Semantic Web requires a mapping from the metadata schema to the RDF data model.

### 4.1 The RDF data model

RDF is a metamodel for expressing metadata about resources. A resource may be an abstract concept, a real world concept or a digital asset such as an entire Web site. The RDF provides a simple model to describe relationships between resources in terms of properties associated with a name and a set of values. The RDF conceptual model is a graph-based model with directed labelled arcs. The nodes of the graph are resources, named or blank, and values, also known as literals. Each named node has an associated URI that uniquely identifies the node. The rules of the arcs, known as triples, are:

- The subject, that is, the origin of the arc, is a resource.
- The property or predicate, that is, the label of the arc, is a named resource.
- The object, that is, the target of the arc, is a resource or a literal

There are two kinds of literals: plain and typed. A plain literal is a character string that optionally has a tag that documents the language of the character string. A typed literal is a pair composed by a value encoded as a character string, and the data type, which defines both the semantics of the value and the syntax of the encoding. For the declaration and the interpretation of these properties, the RDF Schema (RDFS) provides a language to define and restrict the interpretation of the RDF vocabularies

| DC property | ISO 19115:2003 property mapping | RDF property | RDF property range |
|---|---|---|---|
| Contributor | MD_Metadata.identificationInfo. MD_DataIdentification.credit | dct:contributor | Agent |
| Coverage | MD_Metadata.identificationInfo. MD_DataIdentification.extent. EX_Extent.geographicElement. EX_GeographicBoundingBox | dct:spatial | Location |
| Creator | MD_Metadata.identificationInfo. MD_DataIdentification.citation. CI_Citation.CitedResponsibleParty. CI_ResponsibleParty. OrganisationName[role="originator"] | dct:creator | Agent |
| Date | MD_Metadata.identificationInfo. MD_DataIdentification.citation. CI_Citation.date.CI_Date | dct:modified | Typed literal (date) |
| Description | MD_Metadata.identificationInfo. MD_DataIdentification.abstract | dct:abstract | Plain literal |
| Format | MD_Metadata.distributionInfo. MD_Distribution.distributionFormat. MD_Format.name | dct:format | MediaType |
| Identifier | MD_Metadata. MD_Distribution. MD_DigitalTransferOption.onLine CI_OnlineResource.linkage.URL | dct:identifier | Plain literal |
| Language | MD_Metadata.identificationInfo. MD_DataIdentification.language | dct:language | LinguisticSystem |
| Publisher | MD_Metadata.identificationInfo. MD_DataIdentification.citation. CI_Citation.CitedResponsibleParty. CI_ResponsibleParty. OrganisationName. [role="publisher"] | dct:publisher | Agent |
| Relation | - | dct:relation | Resource |
| Rights | - | dct:rights | RightsStatement |
| Source | MD_Metadata.dataQualityInfo. DQ_DataQuality.lineage. LI_Lineage. source. LI_Source.description | dct:source | Resource |
| Subject | MD_Metadata.identificationInfo. MD_DataIdentification.topicCategory. | dct:subject | Resource |
| Title | MD_Metadata.identificationInfo. MD_DataIdentification.citation. CI_Citation.title | dct:title | Plain literal |
| Type | MD_Metadata.hierarchyLevel | rdf:type | Class |

**Table 1:** CWA 14857: Crosswalk ISO 19115 – Dublin Core; the prefix *dct:* maps to the http://purl.org/dc/terms/ namespace; the entities *Agent*, *Location*, *MediaType*, *LinguisticSystem* and *RightsStatement* of RDF property range are DCMI terms classes.

### 4.2 Expressing geographic metadata in RDF: the Dublin Core crosswalk approach

There are several geographic metadata crosswalks to the Dublin Core vocabulary. Table 1 describes the crosswalk of the geographic metadata ISO 19115 to the Dublin Core vocabulary defined in CWA 14857 (Zarazaga-Soria et al., 2003). We propose the use of well-known Dublin Core crosswalks to implement uniforms mappings from geographic metadata schemas to the RDF data model. This approach consists of three steps:

- Apply a metadata crosswalk from the original metadata schema to the Dublin Core vocabulary.
- Add additional metadata such as provenance of the record, original information model or crosswalk identification.
- Apply the profile for expressing as RDF the metadata terms.

The output of the crosswalk can be augmented by adding additional metadata descriptions that log the crosswalk and the provenance of the metadata. Then, this metadata description is transformed to the RDF data model by applying a profile for expressing the metadata terms as RDF. Table 1 also includes an example of a profile. This table includes for each Dublin Core term its mapping to a RDF properties and its range. The RDF Dublin Core profiles are different from the XML Dublin Core profiles. The DCMI abstract model (DCAM) has a reference model formalized in terms of the semantics of the RDF abstract model since 2005 (Powell et al., 2007). One of the changes is that properties may have a formal range. In the RDF data model, this range can be literal, for example the property title, or a resource, for example the property creator. With this approach, when the object of a property refers to an entity, it can be properly identified and described.

## 5 The CSW2LD toolkit

Our approach to solve the drawbacks of CSW is the CSW2LD toolkit. The ideas behind the design of the CSW2LD toolkit are presented below.

### 5.1 Conceptual model for re-publishing metadata

The conceptual model can be decomposed as follows:
- **CSW interface model**. A metadata repository contains metadata about resources. Client applications use CSW requests to query metadata re-

positories. The CSW requests may generate metadata snapshots that are subsets of metadata at the time of the request. The CSW request determines the amount of information (user defined, brief, summary or full records) and the information schema of the metadata snapshot. The CSW response contains the realization of the metadata snapshot in a supported media format. XML is the only media format that all CSW implementations must support.

- **Harvest model.** The harvest produces a set of metadata snapshots realized in XML representations. The harvest process asks for metadata records whose information model can crosswalk to Dublin Core. The CS specification defines a common group of metadata elements expressed using the Dublin Core vocabulary. CSW defines a default mapping of the common group of metadata elements to XML that all CSW implementations must support. The harvest process queries for the common representation if no crosswalk is applicable to the information model of the catalogue.

- **Semantic publication model**. The harvested representation of the metadata snapshot is mapped to the RDF data model and published following the Linked Data principles. The base of the mapping is the DCMI recommendation for expressing Dublin Core using RDF (Nilsson et al., 2008). The result is a semantic description about a resource that is a version of the metadata snapshot that describes the same resource. This semantic description is published according to the best practices to publish Linked Data on the Web (Bizer et al., 2007). The model assumes that a dereferenceable URI, the semantic URI, can identify the resource that the semantic description describes. This semantic URI is owned by the responsible of the semantic publication and redirects to an URI where user agents can get a RDF representation of the semantic description. The semantic description, in turn, has the semantic URI as subject in its assertions. If the mapping process discovers links between the resources, it may replace the original RDF mapping by these semantic URIs. For example, the description of a service may include a brief description of the data. Then, this brief description can be replaced with the URI that identifies the semantic description of the data. The semantic description may contain a link that encodes a CSW HTTP GET request equals to the CSW request done in the harvest. Semantic browsers and search engines, such as Tabulator (Berners-Lee et al. 2006) and Sindice (Tummarello et al., 2007) respectively, can browse and index the semantic descriptions, and use the links to navigate to other resources or to retrieve transparently the original metadata description.

- **Non-semantic publication model**. Given the semantic descriptions described in the previous point, the model assumes that an URI identifies the human-readable representation in HTLM format. The semantic URI of a resource may be resolved to this URI if the agent requests a human-readable representation of its semantic description. This representation uses the HTLM element `<link>` to provide information to navigate alternative representations. At least, it includes a link that points to the semantic URI and a link that encodes the CSW HTTP GET request. Web browsers and search engines can browse and index respectively these representations. In addition, they can use the links to navigate to the semantic representations and to retrieve transparently the original metadata description.

## 5.2 Algorithm for harvesting and publishing a CSW server

Figure 2 summarizes the process in the context of SDIs where the information model of many catalogues is ISO 19115 / ISO 19119. The steps of the harvesting process are:

- Analyze the capabilities of the CSW service to discover the information models served and the levels of amount of information.
- Fetch identifiers of new and updated records with the CSW `GetRecords` operation.
- Retrieve new and updated records using the `GetRecordById` operation; request ISO 19115 / ISO 19119 information models if they are available.
- Crosswalk to the Dublin Core vocabulary if the requested information model is not the common information model.
- Map the set of Dublin Core metadata terms to the RDF data model.
- Generate or update the human readable and machine-readable representations from the RDF graphs.

The `GetRecords` operation does a search and returns piggybacked metadata. The harvest process uses the `GetRecords` operation to determine the number of metadata records to retrieve, and to obtain piggybacked unique identifiers for retrieve metadata records. Optionally, along with the identifier, the harvester process can ask for the creation or update date of the record within the catalogue. The identifiers, and, if available, the creation or update date, are compared with the previous harvest of the same repository to detect new and updated records to retrieve. Deleted records may be kept for archiving reasons.
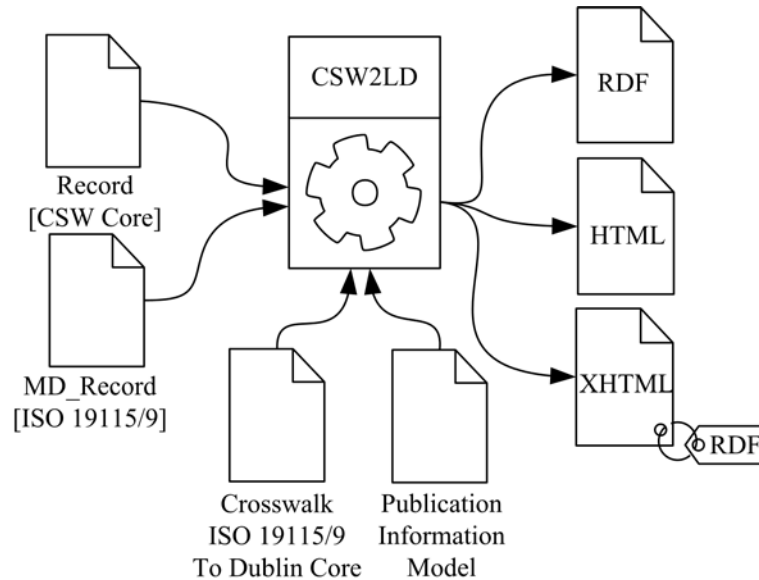
**Fig. 2:** Overview of the republish process of CSW served catalogues in terms of metadata representations.

The current implementation implements a crosswalk described in No-gueras-Iso et al. (2004). The available formats for the machine readable and human readable representations are RDF/XML, N3, TURTLE and XHTML with RDF annotations (RDFa) for the former, and HTML and XHTML for the later.

The harvest process configures an Apache HTTP server for publishing the representations following the conventions of Linked Data (Berrueta et al, 2008). The configuration enables the server to publish machine-processable and human-readable representations. Figure 3 shows the core logic of the redirection and content negotiation implemented in the con-figuration. If the URI matches the web folder, the server returns a `303 See other` response that locates an HTML that informs the user agent about the metadata records exposed in the folder. If the URI matches with a resource contained in the web folder, the server identifies the resource and returns with a `303 See other` the location of the representation that matches the kind of representation requested. The harvest process also creates the index document. It contains hyperlinks to the URIs of the rep-resentations of the semantic descriptions with a summary of the informa-tion such as title and keywords. If the catalogue is large, the harvest proc-ess creates multiple index documents linked each other simulating pagination and modifies the redirection logic.
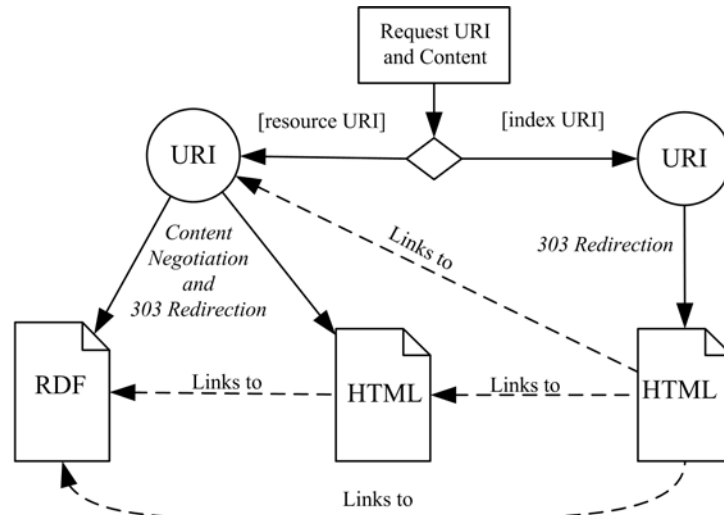
**Fig. 3:** Redirection and content negotiation algorithm.

## 5.3 Enabling transparent access to the metadata repository

One on the goals of the CSW2LD toolkit is to provide transparent access to CSW served repositories. Transparent access and provenance metadata are related concepts in the CSW2LD toolkit. Each semantic description includes a simple provenance description as a triple with an `rdfs:seeAlso` predicate whose subject is the semantic URI and the object is a CSW HTTP GET request. The information content of the semantic description may include additional information about the metadata snapshot, for example, the retrieval date.

   Figure 4 shows how semantic aware user agents can access transparently to the metadata repository. The user agent can discover the semantic URL of a resource in a semantic search engine. Then, it can retrieve its semantic description. After processing the content, the user agent can require additional information. The semantics of RDF says that this might be found traversing the `rdfs:seeAlso` property. As the target is a complete CSW HTTP GET request, the user agent can retrieve a XML representation of the original metadata.
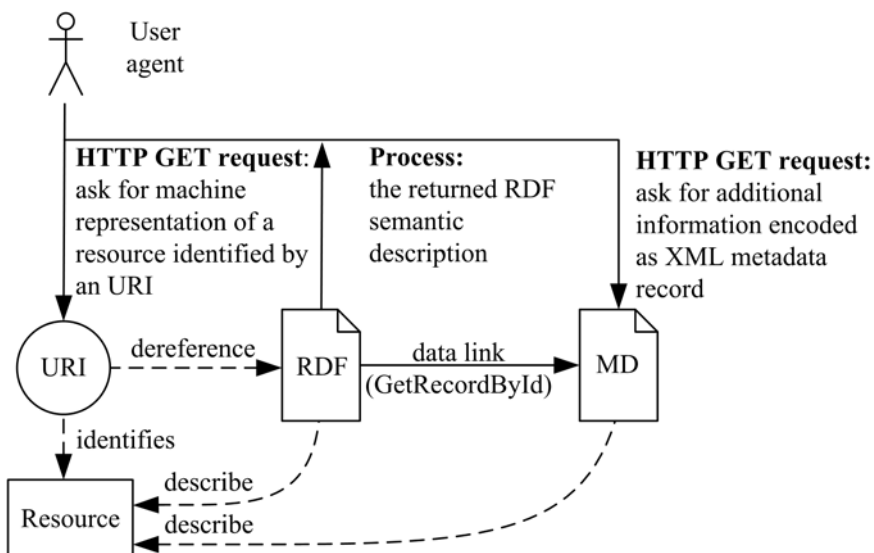
**Fig. 4:** A semantic user agent can access to the content of the metadata repository without previous knowledge of CSW.

## 6 Conclusions

This paper presents the CSW2LD toolkit: a software component that republishes according to the Linked Data metadata from repositories accessible through CSW. Applied to SDI metadata catalogues, the CSW2LD toolkit exposes the description of SDI assets as dereferenceable Web resources, and allows search engines to index them. On the other hand, the published RDF description of metadata records and resources is not standard, and can be semantically inaccurate. The main reasons lie on the lack of standards mappings from geographic metadata schemas to the RDF model, and the heterogeneity of communities targeted by CSW.

Future versions of the CSW2LD toolkit should include additional technical features, such as additional crosswalks, and functional features, such as the generation of links between the metadata and existing thesauri and ontologies, augment the meta-metadata available about the provenance and quality of the exposed information, and describing the exposed data as aggregations.

## Acknowledgements

## References

Auer, S.; Lehmann, J. and Hellmann, S. (2009) LinkedGeoData – Adding a spatial Dimension to the Web of Data Proceedings of 8th International Semantic Web Conference (ISWC)

Becker, C. and Bizer, C. (2008) DBpedia Mobile: A Location-Enabled Linked Data Browser Proceedings of the Linked Data on the Web Workshop, Beijing, China, April 22, 2008, CEUR Workshop Proceedings

Berners-Lee, T., Hendler, J. and Lassila, O (2001). The Semantic Web Scientific American , 284, 34-43

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A. and Sheets, D. (2006) Tabulator: Exploring and Analyzing linked data on the Semantic Web Proceedings of the 3rd International Semantic Web User Interaction

Berrueta, D. And Phipps, J (2008) Best Practice Recipes for Publishing RDF vocabularies [online]. W3C. Available from: http://www.w3.org/TR/swbp-vocab-pub/

Bizer, C., Cyganiak, R., and Heath, T. (2007) How to Publish Linked Data on the Web [online]. Freie Universität Berlin, Available from: http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

Bizer, C.; Heath, T.; Idehen, K. and Berners-Lee, T. (2008) Linked data on the web (LDOW2008) WWW '08: Proceeding of the 17th international conference on World Wide Web, ACM, 1265-1266

Bizer, C., Heath, T. and Berners-Lee (2009) Linked Data - The Story So Far International Journal on Semantic Web and Information Systems

Egenhofer, M. J. (2002) Toward the Semantic Geospatial Web. In GIS '02: Proceedings of the 10th ACM international symposium on Advances in geographic information systems, New York, NY, USA, ACM, pp. 1–4.

Federal Geographic Data Committee (2000) Content Standard for Digital Geospatial Metadata Workbook [online]. Technical report, Federal Geographic Data Committee, Washington, DC. Available from: http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf.

Fielding, R. T. (2000) REST: Architectural Styles and the Design of Network-based Software Architectures University of California, Irvine

Goodwin, J.; Dolbear, C. and Hart, G. (2009) Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web Transactions in GIS, doi: 10.1111/j.1467-9671.2008.01133.x

Haslhofer, B. and Schandl, B.(2008) The OAI2LOD Server: Exposing OAI--PMH Metadata as Linked Data Proceedings of the Linked Data on the Web Workshop, Beijing, China, April 22, 2008, CEUR Workshop Proceedings.

Henning, M. (2008) The rise and fall of CORBA Communications of the ACM, ACM, 51, 52-57

Jacobs, I. & Walsh, N. (2004) Architecture of the World Wide Web, Volume One. W3C

Lagoze, C. and de Sompel, H. V. (2002) The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) – version 2.0. Available from: http://www.openarchives.org/OAI/openarchivesprotocol.html

Lagoze, C. and de Sompel, H. V. (2008) ORE User Guide – Resource Map Implementation in RDF/XML Open Archives Initiative. Available from: http://www.openarchives.org/ore/1.0/rdfxml

Madhavan, J., Ko, D., Kot, L., Ganapathy, V., Rasmussen, A. and Halevy, A. (2008) Google's Deep Web crawl, Proceedings of the VLDB Endowment, VLDB Endowment, 1, 1241-1252

Miller, E. and Manola, F. (2004) RDF Primer. W3C, Available from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

Nebert, D.D. (2004) Developing Spatial Data Infrastructures: The SDI Cookbook [online]. Technical report, Global Spatial Data Infrastructure Version 2.0. Available from: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf.

Nebert, D., Whiteside, A. and Vretanos, P. A. (2007). Open GIS Catalogue Services Specification. OpenGIS Publicy Available Standard, Open GIS Consortium Inc.

Network Services Drafing Team (2009) Technical Guidance Document for INSPIRE Discovery Services v 2.0 [online]. Available from: http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical%20Gu idance%20Discovery%20Services%20v2.0.pdf

Nilsson, M.; Powell, A.; Johnston, P. and Naeve, A. (2008) Expressing Dublin Core metadata using the Resource Description Framework (RDF) [online] Dublin Core Metadata Initiative, DCMI Recommendation, Available from: http://dublincore.org/documents/dc-rdf/

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, 2004, 28, 611- -634

Nogueras-Iso, J., Zarazaga-Soria, F.J., Muro-Medrano, P.R. (2005) Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval. Springer-Verlag New York, Inc., Secaucus, NJ, USA

Powell, A., Nilsson, M., Naeve, A., Johnston, P. and Baker, T. (2007) DCMI Abstract Model [online] Dublin Core Metadata Initiative, Available from: http://dublincore.org/documents/abstract-model/

Raghavan, S. and Garcia-Molina, H. (2001) Crawling the Hidden Web VLDB '01: Proceedings of the 27th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., 129-138

Seaborne, A. & Prud'hommeaux, E. (2008) SPARQL Query Language for RDF W3C. W3C Recommendation. Available from: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/

Tummarello, G., Oren, E. and Delbru, R. (2007) Sindice.com: Weaving the Open Linked Data Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea, Springer Verlag, LCNS 4825, 547-560

Turner, A. (2006) Introduction to neogeography O'Reilly Media, Inc.

Vatant, B. and Wick, M.(2007) GeoNames Ontology. GeoNames, Accessed 6 June 2009. Avilable from: http://www.geonames.org/ontology/

Voges, U. and Senkler, K. (2007) Open GIS Catalogue Services Specification 2.0.2: ISO Metadata Application profile. Open GIS Consortium Inc.

Vretanos, P. A. (2004) OpenGIS Filter Encoding Implementation Specification Open Geospatial Consortium Inc.

Zarazaga-Soria, FJ, Nogueras-Iso, J., Ford, M. (2003) Mapping between Dublin Core and ISO 19115, Geographic Information – Metadata. CWA 14857, CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core