Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Assisted Generation and Publication of Geospatial Metadata

Sergio Trilles
INIT, Universitat Jaume I
Av. de Vicent Sos Baynat, s/n 12071
Castelló de la Plana, Espanya
strilles@uji.es

José Gil
INIT, Universitat Jaume I
Av. de Vicent Sos Baynat, s/n 12071
Castelló de la Plana, Espanya
jose.gil@uji.es

Laura Díaz
INIT, Universitat Jaume I
Av. de Vicent Sos Baynat, s/n 12071
Castelló de la Plana, Espanya
laura.diaz@uji.es

Joquín Huerta
INIT, Universitat Jaume I
Av. de Vicent Sos Baynat, s/n 12071
Castelló de la Plana, Espanya
huerta@uji.es

Abstract

Today, the existence of metadata is one of the most important factors for the effective discovery of geospatial data published in Spatial Data Infrastructures (SDI). However, due to the lack of efficient tools, integrated in the user?s daily workflow to assist users in generating metadata, little metadata is produced. This paper presents a mechanism for generating and publishing metadata with a Publication Service. This method is provided as a web service implemented with a standard-base interface, the Web Processing Service (WPS) specification to improve its interoperability with other SDI components. This paper extends previous research regarding the design of a Publication Service within the framework of the European Directive INSPIRE as a solution to assist users in publishing geospatial data and metadata automatically in order to improve, among others, SDI maintenance and usability.

Keywords: metadata, automated assistance, geoprocessing, discovery, SDI, INSPIRE

1 Introduction

Because of the current status of the SDI and the mechanisms deployed for discovery, metadata (necessary to describe the resources) and Catalogue Services are the key elements for the discovery and resource fusion possibilities [16] [10]. In this context, the INSPIRE Directive mandates the creation and maintenance of metadata and related Discovery Services [7]. According to the INSPIRE Directive, [12], it is necessary to generate and publish metadata for each data published in infrastructure. This metadata must be validated (meet standards) and accurately describe the referred data [15].

Currently, there are a lot of geographic data published online nevertheless, there is a lack of associated metadata published online, and thus finding it is difficult. [17]. One of the reasons is the implicit difficulty and arduous process of creating metadata manually as well as the lack of mechanisms combined with the creation of the data, which would be the best moment to collect all of this information. [22].

To this day there is no trivial automated mechanism that does not require technicians to assist in the process [11], as the generation and publication of data and metadata have to be done together in order to increase the information about the data when generating the metadata.

The aim of this paper is to bridge the gap between the generation and publication process of geographical data together with its metadata in an interoperable and scalable component.

We present a publication service that focuses on describing the capability of automatically generating and publishing the associated metadata [20]. We provide this functionality by implementing an SDI standard-base service that facilitates the creation and publication of metadata.

2 State of art

Metadata are created by data providers and stored in catalogs according to the standards dictated by international organizations such as ISO [1] which are adopted as standards by the community. The purpose of this process is for users to find specific data in distributed environments.

Today, the creation of metadata is not combined with the creation and publication of data, as metadata is difficult to create. In addition, there is no maintenance function of this metadata. [3]. Metadata creation is mainly a manual process normally separate from the process of data generation; this provokes metadata creation to be a time consuming task that in many cases lacks critical information such as where the data is published online and how can it be accessed [17].

In the literature reviews, there is research that addresses this issue by proposing automatic tools for metadata generation, which is a complex task that can be categorized according to the generation techniques [5] [14] [4]. These are generated manually; related works describe automatic and semi-automatic techniques [6] [8] [13]. [1]describes five categories: manually, extending the data stored with values obtained by consulting, using automatic measurements and observations, using data extracted

[1] International Organization for Standardization, http://www.iso.org.

and calculated and finally those inferred from other elements. While other authors categorize the automatic generation into two classes: metadata extraction and inference [3].

Tools such as CatMDEdi [22] are used for automatic extraction of metadata in different formats, as reflected in [19]; the amount of information that can be extracted fundamentally depends on the representation model used and its file format. Thus, there are elements that can only be extracted from certain types of data and files, while others, such as the size of the data, can be obtained in all circumstances.

Other works describe how to integrate the generation of metadata within Geographic Information System (GIS) also used to generate data. These tools already provide the functionality to be able to read formats from the data which thus facilitates the extraction of metadata. In this paper we can highlight the few existing GIS tools that offer an automatic deduction of metadata for raster and vector containers that are based on the analysis of these specific formats, as well as an application of ad-hoc mechanisms that process the data to extract information which is used in metadata construction [14]; also, the gvSIG metadata editor [2].

Besides generation techniques we can describe the nature of the tools which provide metadata generation functionality. Normally this functionality is included in Catalogues Services where metadata will be later stored, this is, for example, in the case of Geonetwork[2] and ESRI ArcCatalog, that is used for the automatic generation of metadata for geographical data. This tool enables the loading of basic fields automatically as well as updating data synchronization and metadata. In both cases, to improve these tools we can add extensions to transform metadata from one metadata standard to another. ARCCatalogue provides a Metadata Editor from the profile *Nucleo Español de Metadatos* (NEM). It is integrated into the ArcCatalog.

In order to provide a component for metadata generation and publication using extraction techniques we wrap this functionality to be provided as an interoperable service that can be combined with other services for data generation. We aim at having the data and metadata generation and publication in the same workflow as a standard services in SDI. This facilitates metadata generation so as to know more about the data itself, such as when it has been generated , where is it available; thus, we increase the linkage between data and metadata in order to facilitate further data visibility and accessibility.

Our approach is a Publication Service called GEOSS Service Factory [11]; in this paper we extend previous research work by adding the ability to generate metadata and publish it to this Publication service, this process when executed after data generation, increases the amount of data descriptions thereby generating more useful metadata.

In order to implement our service interface we choose the standard OGC WPS [18]. WPS allows the deployment of any functionality on the Internet. This provides standards to describe any computation (process) and how to make service requests and responses.

[2]Geonetwork OpenSource, http://geonetwork-opensource.org

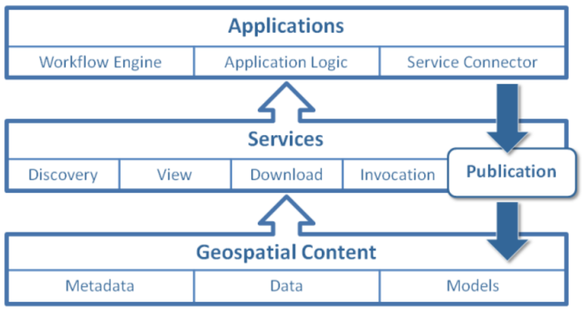
3 Context: extending SDI with a Publication service

The above mentioned process is part of the Geospatial Service Factory (GSF) [9] [11], a publication service that facilitates the provision of content and maintenance in the realm of SDI by all participants. The GSF provides, as an interoperable service implementing the OGC WPS specification, the functionality to publish geospatial data as a standard data service deployed in an SDI according to the European directive INSPIRE. GSF offers a solution for publishing geospatial content aimed at increasing data accessibility for improving data sharing, enabling automatic publication and providing a collaboration tool for the end-user.

The life cycle of content available in an information system, consists of four steps: publishing, discovery, access and processing [9]. GSF improves the publication stage of content, by simplifying and automating the publishing process. The publication is part of the natural workflow for processing information. After the discovery, access and/or data processing, the online publication of new content needs to be created to be reused by other interested stakeholders in the SDI.

In the previous section 2 we described the existing applications to create and publish metadata, which are mainly desktop applications. In contrast, we provide the functionality of generation and publishing metadata as an interoperable, scalable service that can be added to other service calls. Figure 1 shows where the GSF is deployed in an INSPIRE base SDI architecture, where all the functionality is provided as a service in the service layer.

Figure 1: Proposed extended GII architecture.



GSF is implemented with an OGC standard interface specification, particularly the GSF is an OGC WPS. GSF internal design uses the software design pattern: the factory pattern [21]. The factory pattern is a creational pattern that provides a scalable mechanism to create new entities according to particular criteria. In this case, the factory pattern is used to development, for each SDI service is assigned a different factory [11].

This allows the GSF to publish new content (data and metadata) by adding new entries to different service types (according to factory type) View, Download and Discovery. As such, we use inheritance to define the OGC standards, as defined by INSPIRE services. According to the INSPIRE implementation rules for these types of services that should be used for implementation there are ISO 19128:2005 Web Map Service (WMS) for visualization services, Web Feature Service (WFS) for download and CS-W for discovery.

The result of executing the GSF is metadata URLs that describe the recently published content; this metadata contains information about the download, visualization and discovery services where the data is available.

The GSF is associated with a profile or Service Publication Profile (SPP); this contains the necessary information for the publication and metadata generation. The SPP is an Extensible Markup Language (XML) document that can be sent as an input.

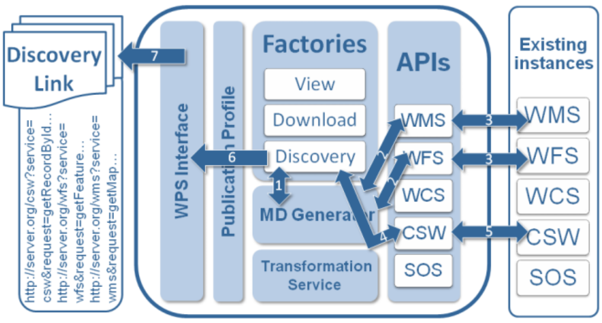
4 Generation and Publication of Metadata

The present paper aims at defining a module that is able to generate and publish metadata within GSF. To generate the metadata information, the data and the site of publication must be specified. Part of this information can be extracted from the services where the data has been published. These data services are mainly available in SDIs.

These services conform to the INSPIRE implementing standards as OGC services WMS, WFS, Web Coverage Service (WCS) and Sensor Observation Service (SOS). Other information that is more difficult to identify are the keywords, so we designed GSF with input parameters such as keywords, which can be indicated by the user. In general, the metadata generator will have to run when the data is published if not stated otherwise. The factory responsible for working with the metadata generator is the Discovery Factory.

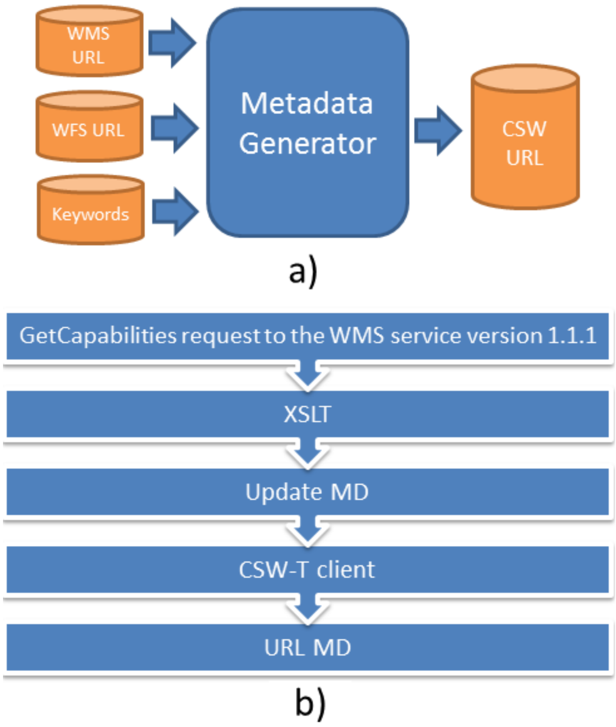
The first step is to consult WMS and WFS services to get the information needed to create the metadata. This is done through appropriate requests (GetCapabilities). After obtaining the necessary information the process applies metadata generation process and proceeds to their publication in the catalog (CSW). Finally, the WPS will return the URL result to the publication. This operation can be seen in Figure 2.

Figure 2: Steps for the publication and generation metadata in the GSF.



After publishing the content and if the WPS has not received a URL with the metadata (MD_URL), WPS call metadata generator. This requires the public services URLs to generate the requests, in addition to inputting keywords. The only output parameter of the GSF has a URL to access the metadata catalog service where it has been published. To further explain the process of metadata generation, we have the following image (Figure 3a) which shows the inputs and outputs of the Metadata Generator module. Figure 3b shows the steps involved.

Figure 3: a) Input and output parameters metadata generator b) Steps metadata generator.



In the first step the GetCapabilities request is performed to services of information where the previous content has been published. For example, if the content has a vector, this will be published for visualization in WMS and for download in WFS. This request returns a XML with the characteristics of the information that will be necessary for the metadata creation. Secondly, the same step is done by the WFS, to provide more information to the metadata.

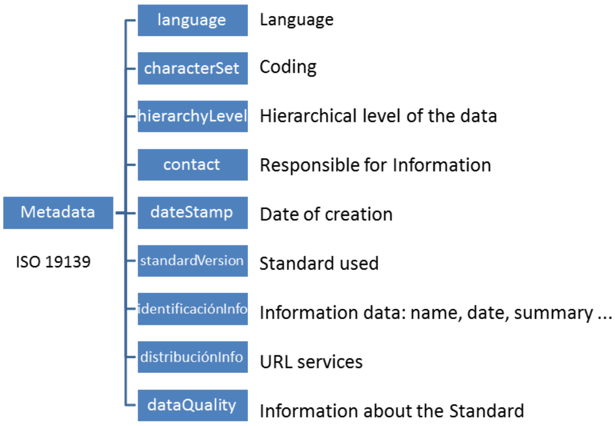
The following step must apply the Extensible Stylesheet Language Transformations (XSLT) to generate metadata. The transformation will take the XML entry obtained from the previous request. The result obtained after the transformation is another XML document, which is the metadata with the ISO 19139 standards and INSPIRE.

The third step consists of parsing the XML metadata for fields that have not been filled in the transformation, such as keywords or URLs to services. After this step, the metadata will be considered complete.

Afterwards, the metadata will be published in the catalogue that is defined in the SPP. This step is performed similarly to the publication of data but in this case to the metadata catalog. In our work we use transactional profile (CSW-T) according to the OpenGIS Catalogue Services specification. In this way, the system will have to implement all operations allowing the CSW-T protocol. These are: GetCapabilities, GetRecords, GetRecordById, GetDomain, DescribeRecord and the transactional.

The last step is to get the URL that identifies the metadata

Figure 4: Some metadata fields.



posted on the server. After obtaining the metadata URL, the generator metadata is considered as finished.

For the metadata generation has been chosen to create the template using XSLT. This is a standard from the World Wide Web Consortium (W3C) organization that transforms XML documents or other types of documents. As we have already stated previously metadata is nothing more than an XML following a standard, so that these transformations can be applied.

We should mention that the metadata that is generated does not have all the fields in the specification and some cannot be obtained automatically. We have chosen the fields that can be seen in the figure below (Figure 4). It shows the main fields that contain the generated metadata.

5 Experimentation

Currently, GSF is able to publish different raster and vector formats, to perform the experiment; we use a Shapefile as an example for vector data and a Geotiff file as an example for raster format. The data files, the content, is sent to the GSF as an input parameter in the WPS Execute request (as a mandatory parameter).

In particular, figures show the publication process of the vector data describing fire-burned areas in the province of Castellón, in Spain in the year 2005. Having this data available as an interoperable SDI service, improves its accessibility for further reuse by other user?s to carry out analysis of environmental impact caused by such events. Furthermore, in order to improve its visibility, it has to be published not only in a data service for visualization or download, but also its metadata has to be published in a metadata catalog, offering information about what it is, how it is and where it can be accessed.

One of the (open source) implementations of the OGC-based SDI services is Geoserver, which offers implementation for OGC WMS, WFS and WCS. We have chosen the Geoserver to test the published data Moreover, the metadata catalog GeoNetwork has been chosen for the metadata publication. As we have mentioned above, we use the CSW-T protocol for inserting metadata in GeoNetwork catalog. This does not include its authentication as it has been implemented specifically for this type of catalogue.

Figure 5: Example of a GetCapabilities layer.

```
<Layer queryable="1">
  <Name>eurogeoss:incend_2005_4326</Name>
  <Title>incend_2005_4326</Title>
  <Abstract/>
  <KeywordList/>
  <SRS>EPSG:4326</SRS>
  <!--WKT definition of this CRS: GEOGCS["WGS 84",
  DATUM["World Geodetic System 1984",SPHEROID["WGS
  84", 6378137.0, 298.257223563, AUTHORITY["EPSG",
  "7030"]], AUTHORITY["EPSG","6326"]], PRIMEM["Gre
  enwich", 0.0, AUTHORITY["EPSG","8901"]], UNIT["d
  egree", 0.017453292519943295], AXIS["Geodetic lo
  ngitude", EAST], AXIS["Geodetic latitude", NORTH
  ], AUTHORITY["EPSG","4326"]]-->
  <LatLonBoundingBox minx="-1.4795661150193993"
  miny="37.90006305852903" maxx="0.3527841387963444"
  maxy="40.68920370363359"/>
  <BoundingBox SRS="EPSG:4326" minx="-1.47956611501
  93993" miny="37.90006305852903" maxx="0.352784138
  7963444" maxy="40.68920370363359"/>
  <Style>
    <Name>polygon</Name>
    <Title>Default polygon style</Title>
    <Abstract>A sample style that just draws out
    a solid gray interior with a black 1px outline
    </Abstract>
    <LegendURL width="20" height="20">
      <Format>image/png</Format>
      <OnlineResource xmlns:xlink="http://www.w3.org
      /1999/xlink" xlink:type="simple" xlink:href=
      "http://elcano.dlsi.uji.es:8080/
      geoserver/wms?request=GetLegendGraphic&amp;
      format=image%2Fpng&amp;width=20&amp;height=20
      &amp;layer=incend_2005_4326"/>
    </LegendURL>
  </Style>
</Layer>
```

After the execution we see how the data has been published to the geographic data server Geoserver and, the getCapabilities is accessible to WMS and WFS services. Figure 5 shows the published layer information.

We have also published GeoNetwork in the metadata server, In Figure 6 we see the visual metadata information. The online resource of the data is inside the red box.

6 Conclusions

In this paper, a mechanism for the automatic generation and publication of metadata has been developed by extending the previous work of geographical information publication in SDIs [9].

The global component addresses the improvement of data availability in the environment of the SDI as well as its maintenance, the particular generation and publication of metadata facilitates where data is found, because it will generate and publish metadata which defines and facilitates its location. There has been a possible solution to the problem of availability of data and metadata, and most importantly we have linked the data and metadata using the link, so that every time we publish data and/or metadata, and improve not only the visibility and discovery of the data but also its access.

Different branches of expansion have been defined as future work. The first is continuing the current work to generate a client which will facilitate the implementation of the WPS, in fixed devices such as a phone. Another line of research is to automatically increase the elements of the metadata generated by using ontologies.

References

[1] K. Beard. A structure for organizing metadata collection. third international conference. *Workshop on Integrating GIS and Environmental Modelling*, 1996.

[2] A. Beltrán, L. Díaz, C. Granell, and J. Huerta. Geocrawler y gvsig: un tándem para la generación automática de metadatos. In *5th gvSIG User Conference (gvSIG 2009). Valencia, Spain, November 2009*. 2009.

[3] A. Beltrán, L. Díaz, C. Granell, J. Huerta, and C. Abargues. Description and publication of geospatial information. pages 133–152. 2009.

[4] A. Beltrán and C. Martín. Generando descripciones de recursos para gvsig desde geocrawler. In *6th gvSIG User Conference (gvSIG 2010). Valencia, Spain, November 2010*. 2010.

[5] D. C.A. Bulterman. Is it time for a moratorium on metadata? *IEEE Multimedia*, 11(4):10–17, December 2004.

[6] T. Campbell. Fostering a culture of metadata production. In *GSDI10: Tenth International Conference for Spatial Data Infrastructure*, 2004.

[7] M. Craglia, I. Kanellopoulos, and P. Smits. Metadata: where we are now, and where we should be going. In *Proceedings of 10th AGILE International Conference on Geographic Information Science*, 2007.

[8] S. Currier, J. Barton, R. O' Beirne, and B. Ryan. Quality assurance for digital learning object repositories: issues for the metadata creation process. *ALT-J*, 12:5–20, March 2004.

[9] L. Díaz, C. Granell, M. Gould, and J. Huerta. Managing user-generated information in geospatial cyberinfrastructures. *Future Gener. Comput. Syst.*, 27(3):304–314, March 2011.

[10] L. Díaz, C. Martín, M. Gould, C. Granell, and M.A. Manso. Semi-automatic metadata extraction from imagery and cartographic data. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 3051–3052, 2007.

[11] L. Díaz and S. Schade. GEOSS service factory: Assisted publication of geospatial content. *Advancing Geoinformation Science for a Changing World*, pages 423–442, 2011.

[12] INSPIRE EU Directive. Directive 2007/2/ec of the european parliament and of the council of 14 march 2007 establishing an infrastructure for spatial information in the european community (inspire). Technical report, Official Journal of the European Union, L 108/1, Volume 50, 2007.

[13] Jorum. The jisc online repository for [learning and teaching] materials: Jorum scoping and technical appraisal study. In *JORUM. Volume V: Metadata*, 2004.

[14] M.A. Manso, J. Nogueras-Iso, M.A. Bernabe, and F.J. Zarazaga-Soria. Automatic metadata extraction from geographic information. In *7th Conference on Geographic Information Science (AGILE 2004), Heraklion, Greece*, 2004.

[15] J. Nogueras-Iso, J. Barrera, A.F. Rodríguez, R. Recio, C. Laborda, and F.J. Zarazaga-Soria. Development and deployment of a services catalog in compliance with the INSPIRE metadata implementing rules. *SDI Convergence*, page 21, 2009.

[16] J. Nogueras-Iso, F. Zarazaga-Soria, R. Bejar, P. Alvarez, and P. Muro-Nedrano. OGC catalog services: a key element for the development of spatial data infrastructures. *Computers & Geosciences*, 31(2):199–209, March 2005.

[17] J. Nogueras-Iso, F.J. Zarazaga-Soria, and P.R. Muro-Medrano. *Geographic information metadata for spatial data infrastructures: resources, interoperability and information retrieval*. Springer, 2005.

[18] P. Schut. Opengis web processing service version 1.0.0. OpenGeospatial Consortium., 2008.

[19] M. Taussi. *Automatic production of metadata out of geographic datasets*. PhD thesis, master thesis, Department of Surveying, Helsinki University of Technology, 2007.

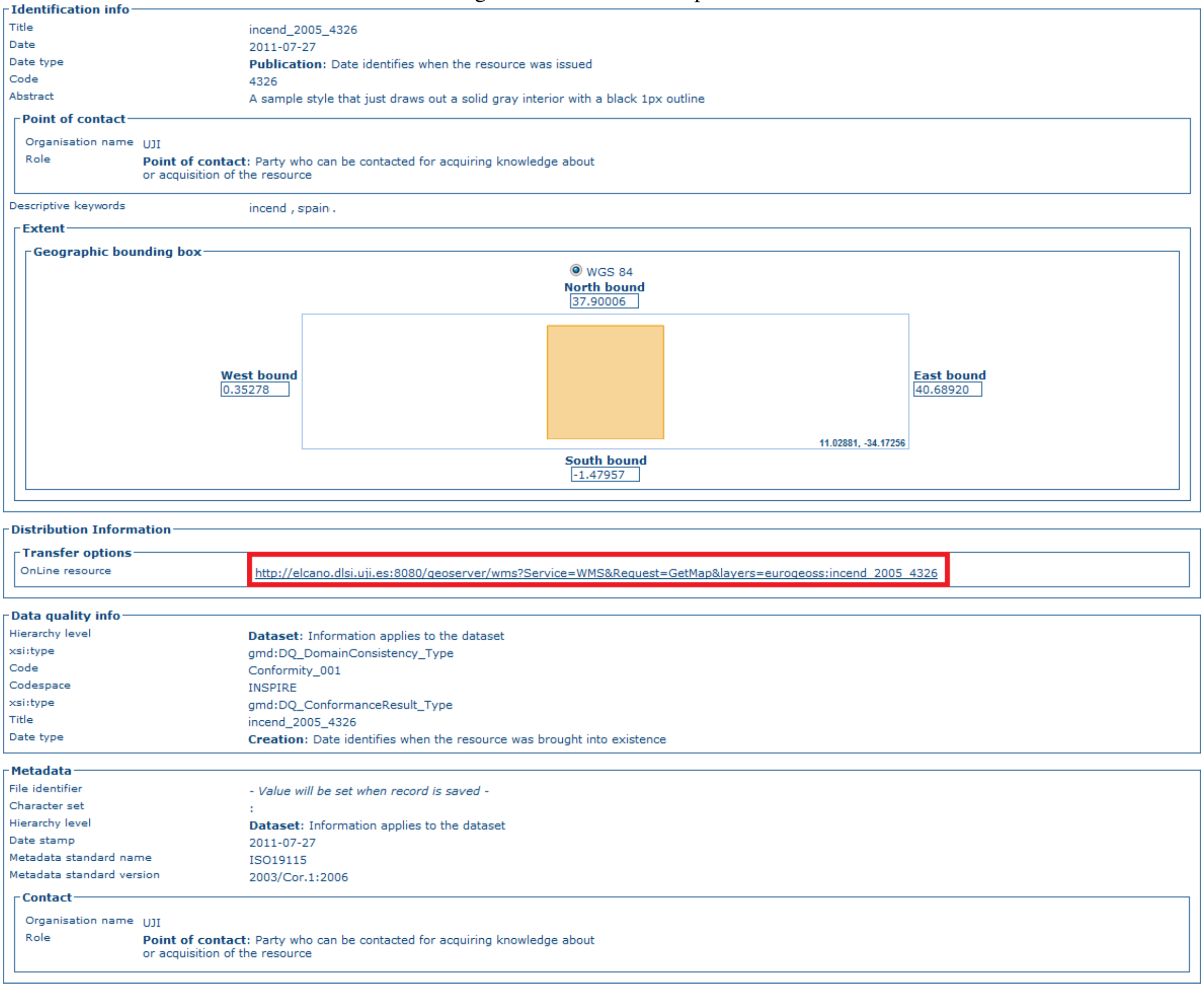
[20] S. Trilles, J. Gil, L. Díaz, and J. Huerta. Un geoprocesamiento para la publicación de metadatos geográficos. 2011.

[21] J. Vlissides, R. Helm, R. Johnson, and E. Gamma. Design patterns: Elements of reusable object-oriented software. *Reading: Addison-Wesley*, 1995.

[22] F.J. Zarazaga-Soria, J. Lacasta, J. Nogueras-Iso, M.P. Torres, and P.R. Muro-Medrano. A java tool for creating ISO/FGDC geographic metadata. *Geodaten-und Geodienste-Infrastrukturen-von der Forschung zur praktischen Anwendung. Beiträge zu den Mänsteraner GI-Tagen*, 26:27, 2003.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Figure 6: GeoNetwork capture.



ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke