GI-CAT: A WEB SERVICE FOR DATASET CATALOGUING BASED ON ISO 19115

Lorenzo Bigagli, Stefano Nativi, Paolo Mazzetti, Giacomo Villoresi
Institute of Methodologies for Environmental Analysis
of the National Research Council (IMAA-CNR)
Area della Ricerca di Potenza, Contrada Santa Loja
Zona Industriale, I-85050 Tito Scalo (PZ), Italy
{bigagli, nativi}@imaa.cnr.it

TOPIC: Web services for spatial information, SDI software architectures

Abstract

In this article, we present the GI-Cat web service: a SOAP-based service which provides the basic functionalities for GI dataset cataloguing and access. The introduced GI-Cat data schema is an instance of the hierarchical, semi-structured data model conforming to the ISO 19115 data model.

GI-Cat reference model, data model (resources and accompanying metadata) and operations are described. Some of the technological aspects of the experimental implementation of the service are briefly described. Eventually, we relate GI-Cat to OWS specifications.

The proposed service is being currently experimented in the framework of the COS(OT) project, which is part of the Italian National Operating Program on Scientific and Technological Research and High Education.

1. INTRODUCTION

There have been several efforts in standardization of GI services and data models, the main ones being conducted by:

- the ISO Technical Committee ISO/TC 211, Geographic Information/Geomatics
  [1]
- the OpenGIS® Consortium (OGC) Web Services Initiative (OWS) [2]

The former aims at “specifying, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analyzing, accessing, presenting and transferring such data in digital/electronic form between different users, systems and locations”.[1]

The latter is a somewhat more pragmatic industrial effort aiming at designing a “complete interoperability framework for implementing multivendor enterprise solutions in government and business”, including specifications for Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), Web Object Service (WOS), Catalog Services-Web (CS-W, also known as Web Registry Service, WRS) and Geography Mark-up Language (GML).[2]

Although the above initiatives seem to be eventually converging on a common ground (e.g. OGC WMS specification is being standardized by ISO in ISO/DIS 19128), there exist differences in the technical choices as well as in the general approach that worth further investigation.

In this article, we present a contribution to the above subject: the specification of the GI-Cat web service, providing the basic functionalities for GI dataset cataloguing and access.

This work builds on the results of the SINOTS Project [3], targeted at the implementation of a Federated Information System for the Earth Observation from Space (EOS) scientific community and funded by the Italian Space Agency (ASI).

The proposed solution is being currently experimented in the framework of the COS(OT) project [4], that is part of the Italian National Operating Program (PON) on Scientific and Technological Research and High Education and aims at supporting the development of industrial districts for Earth Observation in southern Italy. The research is partially funded by the Italian Ministry of Education, University and Research, through the Institute of Methodologies for Environmental Analysis of the National Research Council (IMAA-CNR) and the Italian Inter-University Consortium for Telecommunications (CNIT).

In the following sections, we describe the GI-Cat service with regards to its reference model, its data model (resources and accompanying metadata) and its operations.

Lastly, we briefly describe some of the technological aspects of our experimental implementation of the service and we relate it to OWS specifications.
2. REFERENCE MODEL

The rationale of our service model takes into account the work done in the OWS framework [5, 6, 7, 8], but we mainly focused on full compliance with ISO principles, namely with ISO metadata standard specification [9], due to its generality and its broad area of application to GI (and possibly other kind of information).

This choice accounts for simplicity and flexibility in the GI-Cat interface design, enabling the implementation of different solutions and the deployment of diverse application scenarios, as described in the following sections.

2.1 Requirements

We reckon that the following requirements should be fulfilled by a GI dataset catalog service:
- Persistence capabilities
- Metadata browsing and querying
- User authentication, authorization and auditing
- Asynchronous messaging support (useful for notification or long transactions)
- Service chaining and aggregation support
- Proxy capabilities
- Data access support (including streaming data such as video or audio)
- Transport protocol and network topology independence

2.2 Functional Decomposition

GI-Cat functionalities for the fulfillment of the above requirements are grouped in the following modules:
- Discovery – providing browsing, querying and access functionalities
- Session – providing authentication, authorization, auditing capabilities, single sign-on and transaction support
- Messaging – allowing asynchronous communication pattern
- Repository – providing caching and persistence capabilities
- Distribution – providing routing, aggregation, load-balancing and registry capabilities
- Mediation – providing proxy and service adaptation capabilities

Figure 1 depicts such functional decomposition: the Discovery module is mandatory for every GI-Cat implementation; the other modules may be chained in any needed combination, according to the application scenario at stake (e.g. Figure 2 shows an example module chaining for a possible GI-Cat deployment scenario).
2.3 Reference Architecture

SOAP is the wire protocol connecting the various functional modules of the proposed service.

According to the guidelines of the Web Service Interoperability Organization (WS-I) [10], we have chosen the document/literal binding style, that guarantees a better decoupling of servers and clients with respect to their implementation language, avoiding the subtle problems of RPC. This choice also better enables service chaining.

Figure 2 shows the architecture of a possible GI-Cat deployment scenario, comprising a GI-Cat Repository (possibly an Intranet server) providing fat clients with authenticated, asynchronous access to an aggregation of disparate (GI) resources, by means of GI-Cat instances configured to act as proxies for other services implementation, by mediating their data and service models [11].

With this regard, we have implemented a WCS proxy, as well as a proxy to legacy SINOTS servers.

3. DATA MODEL

GI resource structure is usually highly heterogeneous with respect to schema. Actually, datasets are often schemaless (e.g. flat files).

Hence, we propose a simple data schema, instance of the hierarchical, semi-structured data model and conforming to the ISO 19115 data model.

Figure 3 shows the proposed schema. Noticeably, a GI-Cat service itself is a dataset (it is actually the "root" dataset) and the same metadata set is associated to resources at every level of the data hierarchy.

3.1 Metadata

As stated above, our metadata model is based on [9], that defines the schema required for describing geographic information and services. It provides information about digital geographic data identification, quality, spatial and temporal schema and extent, spatial reference, format and distribution (including the options for obtaining it).

[9] is "applicable to the cataloguing of datasets, clearinghouse activities, and the full description of geographic datasets, dataset series, individual geographic features and feature properties".

The standard defines:
- mandatory and conditional metadata sections, metadata entities, and metadata elements;
- the minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data);
- optional metadata elements to allow for a more extensive standard description of geographic data, if required;
- a method for extending metadata to fit specialized needs.

Moreover, [9] principles “can also be extended to other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data” (although certain mandatory metadata elements may not apply to these other forms of data).

4. OPERATIONS

Figure 4 shows the operations provided by the various GI-Cat modules, most of which have an intuitive meaning. Some of them are worth to be briefly commented:
- query operation is part of the Discovery module and is directly supported by the (mandatory) ISO 19115 core metadata set. The overall information space is considered as a four-dimensional space, segmented according to a 4-W metaphor: What, When, Where, Who. Hence, clients can express fifteen combinations of four intuitive criteria:
  - the topics that queried data deal with;
  - a time interval representing the temporal extent of acquisition of the queried data;
  - a rectangular spatial extent representing the area where the queried data insist;
  - the aggregated datasets containing the queried data;
• **publish/subscribe** are part of the Messaging module and enable the implementation of asynchronous communication patterns by allowing clients to register and receive messages produced by resources (e.g. upon the completion of lengthy operations, etc.)
• **add/remove** operations are part of the Distribution module and allow to aggregate data sources for the sake of performance (e.g. forking queries, etc.)
• **set/getMapping** operations are part of the Mediation module and allow to specify the (data and service model) mapping rules used to integrate heterogeneous resources.

5. TECHNOLOGICAL ASPECTS

The technologies used for the experimental implementation of the proposed solution include: Apache Axis as the enabling Web Service middleware; XML Schema, XSLT and XPath for data encoding and manipulation; Apache Xindice Native XML Database and the XML:DB API as the enabling repository solution; Java as the language of choice for APIs and software components implementation, along with AspectJ, supporting Aspect-Oriented Programming for cross-cutting concerns and enhanced modularization.

6. RELATIONS TO OWS SPECIFICATIONS

GI-Cat may be mainly related to the recent OWS CS-W specification [7] (a registry service with query capabilities), as well as, to some extent, to other specialized OWS access services, such as the Web Coverage Service [6] or the Web Object Service, (a repository service with storage capabilities of XML resources and other content).

However, OWSs were initially conceived to support quite simple data queries based on the HTTP transport protocol and a synchronous request-response communication pattern (i.e., usually, the HTTP GET method with a set of key-value pairs).

The use cases developed in OWS-Phase 1.2 have identified a number of requirements that are beyond the current specifications and demand for more advanced capabilities (e.g. conveying complex information such as multiple binary data, metadata of any formats, notification messages, spatio-temporal filters). A number of limitations have also been identified with the use of HTTP and a centralized network topology for web services.

OWS-Phase 1.2 activities included a SOAP Experiment, investigating the benefits of porting OWS to the Web Services distributed computing platform [12]. The experiment turned successful and current OWS-Phase 2 is working on refining that porting and strengthening SOAP support in OWS Common Architecture, although maintaining compatibilities with older OWSs specifications.

The synchronous communication patterns was also reckoned to be unsuitable to the fulfillment of such advanced requirements. An asynchronous messaging framework has been proposed as the OWS communication pattern [5].

On the other hand, GI-Cat has been designed since its inception as a W3C Web Service based on XML Protocol (SOAP, WSDL, etc.) and takes full advantage of such platform. In particular, it supports a message-based, publish-subscribe, asynchronous communication pattern.

Concerning metadata, OWSs support to the ISO 19115 standard appears to be growing: after various initial proposals (e.g. [13]), sections of the standard have been included in the description of the generic service capabilities [14]. The various OWS resources (map, coverage, and the like) feature specific metadata sets (arguably, to some extent, for historical reasons), but arbitrary metadata (supposedly ISO-compliant) can be associated to them through a reference (noticeably, the Catalog Services specification mandates the use of ISO 19115 for geographic dataset descriptions).
On the other hand, GI-Cat metadata model is fully compliant with ISO 19115 at every level, for the sake of uniformity and generality.

7. CONCLUSIONS AND FUTURE WORK

The implementation of a Spatial Data Infrastructure implies the implementation of interoperability among the vast diversity of GI systems, largely heterogeneous and autonomous under the physical, logical, methodological and technological point of view. Service-Oriented Architecture (SOA) approach is suited to address such problem, due to its inherent scalability and loose coupling. In particular, the W3C Web Services distributed computing platform [15], based on XML, SOAP, WSDL and related technologies, proved to be an excellent framework to built SOA solutions for GI.

The presented GI-Cat service implements GI cataloguing and access functionalities, resorting to ISO 19115 data model and using W3C web services solutions.

GI-Cat is currently the foundation of a nationwide interoperability solution, being developed in the framework of a project led by the Italian Ministry of Environment and Territorial Protection.

GI-CAT implementation choices are compatible with Grid Services technologies and the upcoming Web Service Resource Framework (WS-RF), currently under theoretical investigation by the ICT scientific community and supposedly the next step toward a truly "World Wide Lab", where geo-data producers and consumers alike interact and leverage each others' experience and results.

8. REFERENCES

[1] ISO TC 211
   http://www.opengis.org/initiatives/?iid=7
[4] Homepage of the COS(OT) Project
   http://www.imaa.cnr.it/n_cosot.htm
[6] OGC 03-065r6, Web Coverage Service (WCS), Version 1.0.0
[8] OGC 03-025, OpenGIS® Web Services Architecture
[12] OGC 03-014, OWS 1.2 SOAP Experiment Report
[14] OGC 03-088r6, OWS Common Implementation Specification
   http://www.w3.org/2002/ws