

Building Service Oriented Applications on top of a Spatial Data Infrastructure – A Forest Fire Assessment Example

Anders Friis-Christensen¹, Lars Bernard¹, Ioannis Kanellopoulos¹, Javier Nogueras-Iso², Stephen Peedell¹, Sven Schade³, Cathal Thorne¹

¹ Joint Research Centre, IES, I-21020 Ispra(VA), Italy

{anders.friis|lars.bernard|ioannis.kanellopoulos|stephen.peedell|cathal.thorne@jrc.it}

² Universidad de Zaragoza, E-50018 Zaragoza, Spain

{jnog@unizar.es}

³ University of Münster, D-48149 Münster, Germany

{schades@uni-muenster.de}

SUMMARY

In this paper we study the feasibility of using services offered by a Spatial Data Infrastructure as the basis for distributed service oriented geoprocessing. By developing a prototype we demonstrate that a Spatial Data Infrastructure facilitates rapid development of applications that solve real problems. The prototype provides users with a distributed application that enables the assessment of fire damage areas based on land cover data in a given area. The services involved in the application include: Web Feature Services, Web Map Services, and Gazetteer Service, Catalogue Service, and Geoprocessing services. We present the architecture of the application and describe details about implementations specific issues. We conclude that the OGC specifications provide a sound basis for developing service oriented architectures for geographic applications; however, in particular for geoprocessing applications, we question the feasibility of the use of Web Feature Services as data sources for larger amounts of data and call for further research in this direction.

KEYWORDS: *Spatial Data Infrastructures (SDI), web services, geoprocessing services, forest fire, Service Oriented Architecture (SOA)*

INTRODUCTION

In recent years the technology trend within information technology has made it possible to move towards service oriented architectures and distributed computing. A service can be defined as self-contained, self-describing collections of operations, and does not depend on the context or state of other services. A service oriented architecture (SOA) is essentially a collection of services, which communicate with each other. The communication can involve either simple data passing (function calls) or it can involve two or more services coordinating some activity or processes. The SOA approach also applies within the GIS domain where several initiatives have been launched (OGC, 2004a; OGC, 2004b; OGC, 2005b; OGC, 2005c; OGC, 2005d). This has created a technology evolution that moves from standalone GIS applications towards a more loosely coupled and distributed model based on self-contained, specialized, and interoperable GI services (GSDI, 2004), see also, e.g. ESA portal web site²⁴.

The benefits are several. From users' perspective, a SOA setting is an open and interoperable environment, which is based on reusability and standardized components. Basically a SOA creates an infrastructure for application development. Development is focused towards concrete applications (and thereby specific requirements and needs) and in contrast to standard GIS applications where normally only a small percentage of the functionalities in the software are used, applications based on SOA provide users with just the functionality they need. Another prominent intention of the design of a SOA, is that data used for a given processing activity are not stored locally, but rather decentralized close to the source of production. This means that inconsistency in local copies and repositories of

²⁴ <http://services.eoportal.org/>

data are avoided and, hence, the quality of the output is possibly increased in cases where data from various different sources are used. Furthermore, redundancy in the algorithms used for specific processing tasks are also avoided. The SOA approach to system development can produce systems that can be flexibly adapted to changing requirements and technologies, and offers easier maintainable and more consistent systems of data and functionality. However, whereas the technology is ready for SOA, the existing specifications, standards, and products available are still immature. Thus, there is a need to evaluate a SOA based on existing standards and specifications for services in various scenarios to investigate whether specific domain requirements are fulfilled.

This paper investigates distributed geoprocessing for a forest fire application and reports on a development of a SOA application using the resources provided by a Spatial Data Infrastructure (SDI) setting. SDIs provide the framework for the optimization of the creation, maintenance and distribution of geographic information at different organization levels (e.g., regional, national, or global level) and involve both public and private institutions (GSDI, 2004). Furthermore, the political support given at high governmental levels by legislations like the U.S. National Spatial Data Infrastructure (USFR, 1994) or the emerging INSPIRE (Infrastructure for Spatial Information in Europe) proposal for a directive (EC, 2004) have encouraged the development of SDIs. At present, most of these SDI initiatives are still at an early stage in their development, just starting to offer geoportals that integrate attractive on-line map viewers and search services for their data holdings (EC, 2005). However, the idea is that in addition to developing attractive geoportals, an SDI should be a valid mechanism for the development of useful applications in an easier and more flexible manner as opposed to stand-alone applications.

The idea behind the prototype presented in this work is to provide users with a distributed application that enables the assessment of fire damage areas based on land cover data in a given area. It should be possible for users to search catalogs for available land cover data and select those required for damage area assessment. Although a very simple application and use case, which can be done in any stand-alone GIS client, the aim and scope of this work is to create a prototype that runs in a distributed and interoperable environment and should exemplify the scalability, flexibility and reusability of the components in an SDI (Bernard et al, 2005).

The remaining paper is as follows. Section 2 describes the proposed architecture of the use case application. Section 3 describes details of the various components and provides details about specific implementations and how they relate to relevant OGC standards. Finally, in Section 4 we conclude on our findings and briefly outline future research topics.

ARCHITECTURE OF PROTOTYPE

An overview of the components in the overall architecture is depicted on Figure . There are various data sources which need to be accessed via WMS, WFS, and Gazetteer (WFS) services. For the forest fire application we have access to the following thematic data, which are potentially important for forest fire statistics:

- Natura 2000 (EC, 2001), which is a coherent ecological network of special areas of conservation across the European Union. The main objective of Natura 2000 is to contribute to the preservation of biological diversity on the territory of the European Union. It is based on two directives: the 1979 “Birds Directive” and the 1992 “Habitats Directive”. It is important in a management of protected areas to assess the amount of areas that are affected by fire.
- Corine Land Cover 2000, which is a fundamental thematic reference data set for spatial and territorial pan-European analyses. The Corine Land Cover is interesting in our use case as it is relevant to know which thematic data classes are affected when assessing forest fire damage.

A reason for doing statistics on a WFS instead of WMS is that a WFS offers possibilities to access attributes, which in further development of the application could be useful. Of other data relevant for our application are:

- Image 2000, which provides backdrop satellite images to the application.

- Place names used for locating specific geographic area based on geographic name input.
- Assessed boundaries of burned areas for various years, which is crucial for our application. This information is collected and assessed by the Joint Research Center under the European Commission and distributed via the European Forest Fire Information System (EFFIS²⁵).

In addition to the data services, the application comprises a catalog service and two different geoprocessing services (an area statistics service and coordinate transformations service), which are necessary for the application:

- The catalog service provides metadata of the various thematic data services.
- The area statistics service is responsible for calculating the areas affected by fire
- The coordinate transformation service is responsible for transforming coordinates into requested coordinate reference system.

The reason for including a coordinate transformation service is that the thematic data services provides data in geographic coordinates, which are not usable for area statistics (coordinate reference system is in degrees and not area true), hence, the coordinates transformations service transforms the coordinates into projected ones. The coordinate transformation service box is dotted in Figure as it is currently implemented as part of the area statistics service and not as a standalone service. In the future this will be implemented as a Web Coordinate Transformation Service (WCTS) to evaluate feasibility and performance issues. Finally, there is a client which provides the user interface for the application.

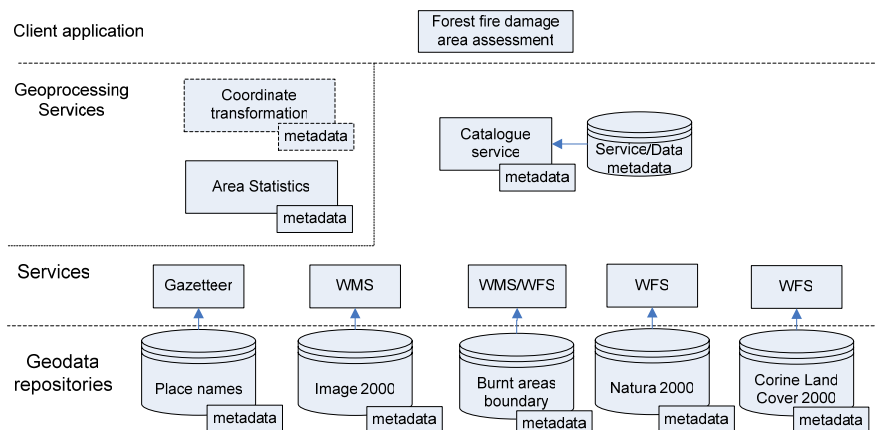


Figure 1: The components in the architecture of the prototype application

A component view of an application scenario is depicted in Figure .

It is seen here that the client requests the backdrop image for visualization (basically as guidance for a more feasible zoom level). Then the area of interest and year is selected in order to get parameters for selecting appropriate data and then, the Image 2000 data showing the area of interesting is visualized. The catalog is used to search and select those data used as source data (also called the mask) and target data. In the scenario, the source data is a specific layer of burned area and the target data could be Natura 2000 or Corine Land Cover data. In theory any type of thematic data layer could

²⁵ <http://inforest.jrc.it/effis/>

be used, but these are the ones we have access to and that are potentially important when assessing forest fire damage. After selection of data a parameter for statistics, which specifies whether statistics should be done per classes or as a whole, is selected by the user. For example the Corine Land Cover data is divided in thematic classes and it would be interesting to know which classes are affected by fire. For data not having several thematic classes this is not necessary, e.g., the Natura 2000 data only includes areas that are protected and the attributes associated here are name of area, reference to a textual description, and other attributes, which are not suitable for basing statistics on. The area statistics service is invoked after selecting all parameters necessary and a request for data is made (by the statistics service). If necessary, coordinates are transformed into projected coordinates and then the assessed area statistics are returned and visualized as a table in the client.

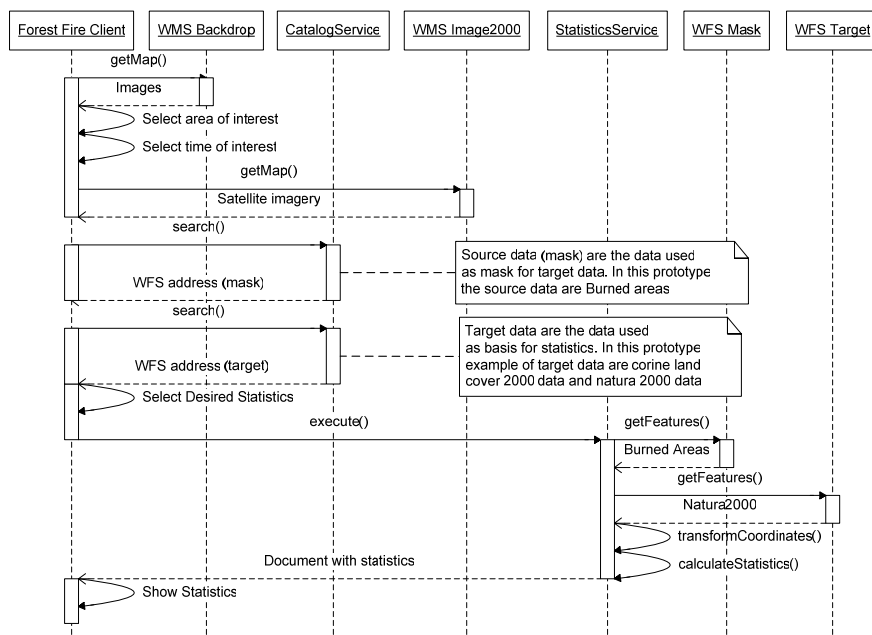


Figure 2: A simplified sequence diagram of an application scenario

COMPONENT DESIGN

In this section the various components in the architecture are presented. More specifically the statistics service, the mapping and feature services, the catalog service and client, and the forest fire client.

Statistics service

The interface of the statistics service follows the WPS specification discussion paper (version 0.3.0) which is continuously evolving (OGC, 2005c) and is implemented in Java 1.5 using the Geotools 2.1 API (Geotools, 2005). A conceptual model of the service and its interface is shown in Figure . The model is simplified and does not show detailed implementation aspects. The OGC WPS specification specifies three operations as mandatory: `getCapabilities`, `describeProcess`, and `execute`. The `getCapabilities` operation (which is common for all OGC web services) simply allows clients to retrieve service metadata from the service. The `describeProcess` describes a specific process (operation) that is supported by the specific WPS. We have not (yet) implemented this operation in

the statistics service. The process supported by a WPS can be called via the `execute` operation, which carries out the specific operation requested.

Two types of requests are supported by WPS: key valued pair (KVP – Get) requests and XML (Post) request. The Get request is a plain html request with all parameters specified. The Post request is a submitted XML document including all parameters in XML tags. The `execute` operation takes both Get and Post request; however, we only describe the Get request here. As seen in Figure there are several parameters to the request, which need to be passed to the statistics service. First, the `service` and `processname` have to be specified. The service is a WPS and it only supports one process: `AreaStatistics`. The `mask` is the http address for the mask data and `masktypename` is the feature type name. The same parameters hold for target data. The `bbox` is the bounding box in which the statistics need to be calculated. The `attribute` parameter specifies the specific attribute to be used, if there is a need to give statistics per thematic class. The `totalarea` parameter specifies if the total burned area should be given (default is true).

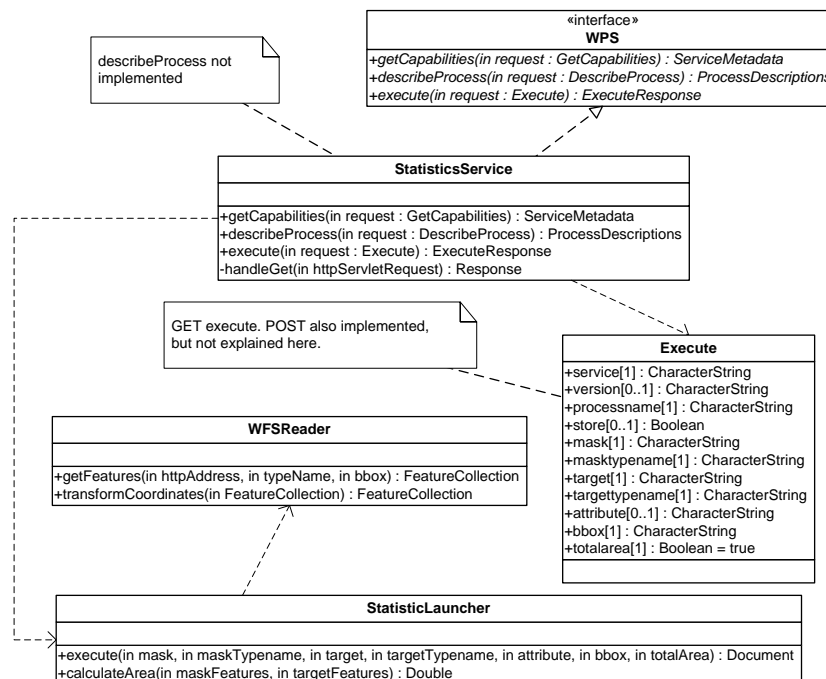


Figure 3: A simplified model of the statistics service

An example of a Get request is (line breaks are added to improve readability):

```

http://naturegis.h07.jrc.it:8090/StatisticsService/Process?
SERVICE=WPS&
REQUEST=Execute&
VERSION=0.3.0&
PROCESSNAME=AreaStatistics&
STORE=false&
MASK=http://naturegis.h07.jrc.it:8090/geoserver/wfs/&
MASKTYPENAME=INSPIRE:ba2003&
TARGET=http://naturegis.h07.jrc.it:8090/geoserver/wfs/&
  
```

```
TARGETTYPE=INSPIRE:clc&  
ATTRIBUTE=CLC_00&  
BBOX=-9,39.33,-8.56,39.7
```

When a request is made to the statistics service, it launches the `execute` operation in the `StatisticLauncher`, which uses the parameters from the request. In order to receive the features in the chosen area of interest (the bounding box) a `getFeatures` from a `WFSReader` is launched. Here it is determined from a `getCapabilities` request to the WFS, which coordinate reference system is provided. At the moment our WFSs only distribute data in a geographic coordinate system and we transform the coordinates in order to get an area true coordinate reference system. In the current implementation, there is only the possibility to transform to EPSG:3034 (ETRS89 Lambert Conformal Conic Coordinate Reference System), however we plan to implement the coordinate transformations to EPSG:3035 (ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System) in the future. Finally, the statistics are calculated and then an XML document including the statistics is returned from the service.

Mapping and feature services

As mapping and feature services we use standard implementations of OGC WMS 1.3 and WFS 1.0 specifications. We use the open source Geoserver version 1.3.0 as a WFS and the ArcIMS 9.1 with a WMS connector, which provides the backdrop and satellite images. For the gazetteer RedSpider Studio provides a simple mechanism to build a gazetteer service on top of an existing WFS service. The WFS used is available at ionic software website²⁶.

Catalog

As Catalog we use *con terra*²⁷ `terraCatalog`, which is an implementation of the OGC Web Catalog Service (OGC, 2004a) specification and makes it possible to store and retrieve information about spatial data and services. In particular, this implementation supports the ISO 19115/19119 profile for CSW 2.0 catalog services (OGC, 2005a). In order to access the catalog from the forest fire client (and not the standard *con terra* client) we used the standard catalog interface to develop a client, which can access metadata stored in the catalog. A simplified model of the client is shown in Figure . What the client basically offers is a `search` operation which takes title, bounding box, and year as parameters. The client supports two different protocol bindings using HTTP as transport mechanism (the Z39.50 protocol binding and the Catalog Services for the Web (CSW)). For the communication with the `terraCatalog` we use the `CSWCatalogClient`, which implements the CSW protocol binding. The title of the data set and a service URL is returned for the catalog and then the preferred data set can be selected in the forest fire client. The service URL is used as parameter for a request to the statistics server.

²⁶ http://webservices.ionicsoft.com/gazetteer/wfs/GNS_GAZ

²⁷ <http://www.conterra.de>

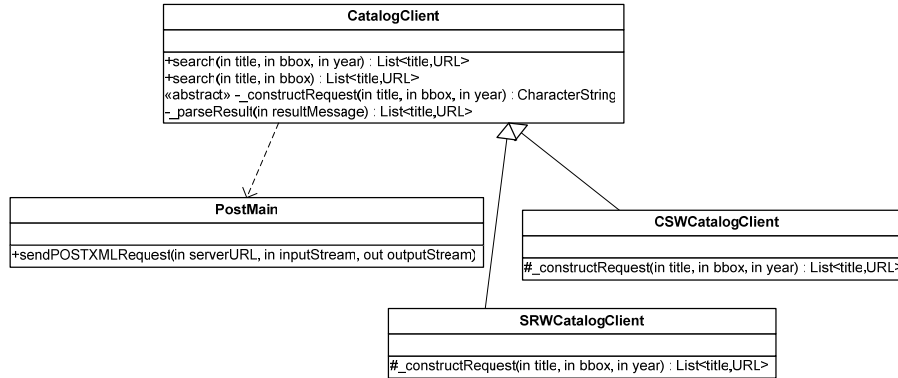


Figure 4: A model of the client for accessing the catalog

Forest Fire Client Application

The client was built using Dynamic HTML (DHTML) and the RedSpider Studio 3 which is a geospatial portal development solution for distributed OGC web services. More specifically, the JSP 'geotag' library allows for easy access to remote services which implement OGC specifications. Response times for user requests were reduced by utilizing the Asynchronous JavaScript and XML (AJAX) scripting technique. This allows users to call the catalog and statistics services and work with the results without requiring a full application refresh. As depicted on the screenshot in Figure users can zoom and pan or locate an area via a gazetteer service. Then, after selecting a year and keywords for searching burned areas (only this is shown) and target data, a damage area statistics report is generated (the bottom part in Figure).

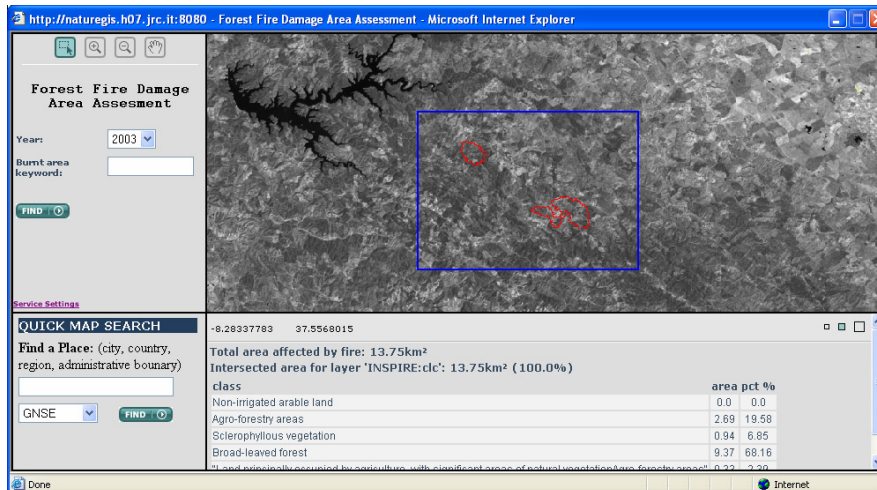


Figure 5: Screenshot of the forest fire damage area assessment client

CONCLUSIONS AND FUTURE WORK

In this paper we have reported on the development of an application that enables the assessment of fire damage areas based on land cover data in a given area. We have done this using and implementing important components in an SDI. As stated in the introduction, most SDI initiatives are still in an initial state and just starting to offer geoportals that integrate on-line map viewers and search services for their data. However, we have demonstrated here, that apart from this initial step, SDIs can be used to develop applications solving real problems in a more flexible and scalable manner than ad-hoc and stand-alone applications. Additionally it is important to remark that we built the prototype using software from various vendors and by doing this showed that interoperability can be achieved.

Our experiences show that the standards of OGC provide a necessary standardized base for services within an SDI. The discussion paper of the Web Processing Service provides a specification that sufficiently enables an implementation. Catalogs and service metadata are a backbone for an application involving distributed data sources and geoprocessing services. It shows the fundamental need to document not only data but also services and data sources in order to support interoperable application scenarios involving distributed geoprocessing.

In addition the work presented shows that there are considerable benefits by a distributed geoprocessing environment (e.g., non replicated data and reusability). However, there are fundamental performance issues stemming from technical limitations and the architectural design. This result in an application, which for certain uses is not feasible. It is evident that a WFS provides interoperable access to geographic data. The main problem is that access in the WFS context also means retrieval. Clearly the WFS is not suited for applications, which require transport and calculations of large amounts of data. For example if we require forest fires statistics for the whole of Spain and Portugal it would simply take too much time to perform the calculations because the GML need to be transported from the data sources (WFS) to the statistics service. An example is that the GML file for Corine Land Cover is > 1 GB. Obviously, transport of such amounts of data is not feasible. On the contrary for much smaller areas the prototype application performs well. So evidently there is a need to investigate the feasibility and architectural design for distributed geoprocessing. One could image server side generalization of data sources decreasing the quantity of data, e.g. (Lehto and Sarjakoski, 2005). Another possibility could be to distribute algorithms instead of data, considering the data sources they are capable of processing. These issues are pertinent topics for further research. Naturally, for our simple application, which executes overlay operations we should consider using raster format instead of vector, however, vector data offers many more possibilities for doing more advanced statistics.

Another issue we have addressed is the interoperability of implemented services following OGC standards. In general interoperability is achieved but several problems have been encountered. For example an application developed using the Geotools API has problems accessing the ArcIMS WFS connector. This just shows that even though vendors following the same standard implementation specifications, vendor specific implementation decisions can have impact on the interoperability.

BIBLIOGRAPHY

- Bernard L., I. Kanellopoulos, A. Annoni, P. Smits, 2005: The European Geoportal – one step towards the establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems*, 29 (2005), pp. 15–31
- EC, 2001: European Commission – Managing Natura 2000 Sites, ISBN: 92-828-9048-1
- EC, 2004: European Commission – Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE). COM(2004) 516 final, 2004/0175 (COD).
- EC, 2005. European Commission – Spatial Data Infrastructures in Europe: State of play Spring 2005. Summary report of Activity 5 of a study commissioned by the EC (EUROSTAT & DGENV) in the framework of the INSPIRE initiative. http://www.ec-gis.org/inspire/state_of_play.cfm

- Geotools, 2005: Geotools toolkit 2.1, <http://www.geotools.org/>
- GSDI, 2004: Developing Spatial Data Infrastructures: The SDI Cookbook. Editor: D. D. Nebert
- IES, 2003: Map Projections for Europe. Eds: A. Annoni, C. Lutzet, E. Gubler, and J Ihde. European Commission, Joint Research Center.
- Lehto L. and L.T. Sarjakoski, 2005: Real-Time generalization of XML-encoded spatial data for the Web and mobile devices. International Journal of Geographical Information Science, Vol.19, No.8-9, pp. 957-973
- OGC, 2004a: Open Geospatial Consortium Inc. Catalogue Services Specification, v2.0. OGC 04-021r2.
- OGC, 2004b: Open Geospatial Consortium Inc. Web Map Service specification 1.3. OGC 04-024
- OGC, 2005a: Open Geospatial Consortium Inc. Catalog Services Specification 2.0 - ISO19115/ISO19119 Application Profile for CSW 2.0. OGC 04-038r2
- OGC, 2005b: Open Geospatial Consortium Inc. Web Feature Service Implementation Specification 1.1. OGC 04-094
- OGC, 2005c: Open Geospatial Consortium Inc. Web Processing Service. Discussion paper. OGC 05-007r4
- OGC, 2005d: Open Geospatial Consortium Inc. OGC Web Services Common Specification. OGC 05-008
- USFR, 1994: U.S. Federal Register 1994 – Executive Order 12906. Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure (U.S.). The April 13, 1994, Edition of the Federal Register, 59(71):17671-17674