

Enabling Efficient Discovery of and Access to Spatial Data Services

Karel Charvat, Premysl Vohnout, Michal Sredl,
Stepan Kafka, Tomas Mildorf
Czech Centre for Science and Society
Prague, Czech Republic
ccss@ccss.cz

Andrea De Bono^{1,2}, Gregory Giuliani^{1,2}
¹Institute for Environmental Sciences, enviroSPACE
University of Geneva 1227 Carouge, Switzerland,
²United Nations Environment Programme
Global Resource Information Database
1211 Châtelaine, Switzerland
{debono, gregory.giuliani}@unepgrid.ch

Abstract—Spatial data represent valuable information and a basis for decision making processes in society. The number of specialisms that use spatial data for such purposes is increasing. Increasing is also the number of services enabling to search, access, process, analyse or visualise spatial data. Standardisation activities of the Open Geospatial Consortium (OGC) support standardised sharing of services through the Web. However, many services declared as OGC compliant do not respond or they are not available. The paper introduces an innovative solution for efficient discovery of and access to spatial data services compliant with OGC specifications. The research was performed in the context of the EnviroGRIDS geoportal. Several thousands of harvested services were quality checked and the summary of the testing including the identified problems are presented.

Keywords—EnviroGRIDS; web services; discovery; metadata; geoportal; SDI; INSPIRE; OGC; SuperCAT

I. SPATIAL DATA AND INTEROPERABILITY

Spatial data, sometimes referred to as geographic data, geodata or geospatial data, are defined by INSPIRE (Infrastructure for Spatial Information in the European Community) as "data with a direct or indirect reference to a specific location or geographic area." [1] It has been estimated that over 80% of all data have a spatial component. Spatial references enable to locate objects, processes and other phenomena; to model their shape and to analyse their relation to other data [2].

Spatial data are collected by various organisations all over the world, from local to global level. Data are collected using different techniques. The purposes of data collection also vary. Then there are issues for example of data storage, processing, analysing and visualisation. All of these aspects and many others contribute to heterogeneity of spatial data. Due to these aspects, it is not easy to combine data from various resources. In order to make spatial data usable in cross border activities, interoperability framework must be agreed.

Interoperability is defined by the International Organisation for Standardization (ISO) as "capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." [3]

Recent activity of the European Commission paid due attention to data interoperability in a document describing the European Interoperability Framework (EIF) for European public services. EIF distinguishes four levels of interoperability including legal, organisational, semantic and technical levels. As shown in Figure 1, the political context underlines all the interoperability levels and creates the environment for successful and meaningful cooperation.

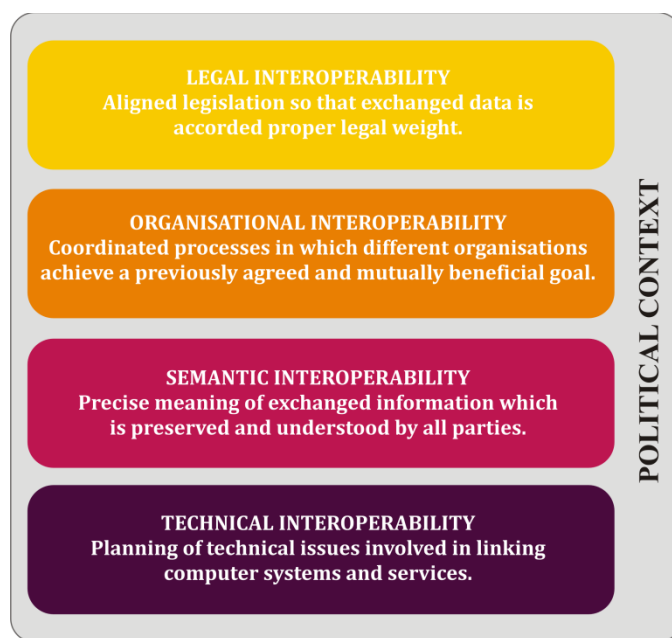


Fig. 1. Levels of interoperability (adapted from [4])

Interoperability on all levels can be achieved through adoption of common standards, specifications and other agreements. The most important international and well respected standards in the field of spatial data are those created by the Technical Committee 211 of the International Organization for Standardization (ISO/TC 211) and by the Open Geospatial Consortium (OGC). Together with the INSPIRE specifications and national standards they create the core of the European spatial data infrastructure (SDI), mainly from the technical and semantic points of view.

Spatial data represent one of the key components of SDI. Spatial data infrastructure, sometimes referred to as spatial information infrastructure, is generally understood as "a computerised environment for handling data that relate to a position on or near the surface of the earth." [5] There are many definitions of SDI. INSPIRE defines SDI as "the metadata, spatial data sets and spatial data services; network services and technologies; agreements on sharing, access and use; and coordination and monitoring mechanisms, processes and procedures, established, operated or made available in an interoperable manner." [6]

This paper presents an innovative solution for efficient discovery of and access to spatial data services and the OGC interoperability standards. "Spatial data services means the operations which may be performed, by invoking a computer application, on the spatial data contained in spatial data sets or on the related metadata." [6] Chapter II describes the rationale for the research in the context of the EnviroGRIDS project. Chapter III describes the innovative solution SuperCAT and Chapter IV provides the results of the spatial data service testing using SuperCAT.

II. ENVIROGRIDS GEOPORTAL

EnviroGRIDS is an FP7 project that aims at building capacities in the Black Sea region to use new international standards to gather, store, distribute, analyse, visualise and disseminate crucial information on past, present and future states of the Black Sea region in order to assess its sustainability and vulnerability. To achieve its objectives, EnviroGRIDS built a modern SDI that became one of the components of the Global Earth Observation System of Systems (GEOSS), compatible with the INSPIRE Directive.

GEOSS is being built by the Group on Earth Observations (GEO). GEOSS is focused on user needs and support better utilisation of environmental data and decision-support tools by users. GEOSS is focused on global infrastructure, supplying near-real-time environmental data, information and analyses. GEOSS supports utilisation of information by wide range of users. There are nine areas of interests in GEOSS: disasters, health, energy, climate, water, weather, ecosystems, agriculture and biodiversity. Potential user groups include decision makers in the public and private sectors, resource managers, planners, emergency responders, and scientists. [7]

An important part of the EnviroGRIDS SDI is a geoportal. The geoportal was designed and implemented as a virtual database. It uses the principles of web services, Uniform Resource Management (URM) [8], social media, Geoportal4everybody [9] and semantic web. The geoportal integrates social networking tools supporting social assessment. These services are not implemented directly on the EnviroGRIDS geoportal but as virtual services on different places all over Europe.

The design of the EnviroGRIDS geoportal is based on the analysis of the INSPIRE and GEOSS principles and the principles of the Service Oriented Architecture (SOA) that is INSPIRE compliant. The INSPIRE requirements give to the overall system architecture a loosely coupled integration based

on OGC standards, which allows to use any OGC compliant software components.

The EnviroGRIDS geoportal allows management of spatial and non-spatial data across the Black Sea catchment and integration of different existing resources in this area. The geoportal is not only a set of client applications but also a gate to all data and services registered on the geoportal and interconnected servers. The interconnection with other servers is achieved by using the OGC specifications for data and service interoperability. The geoportal enables for example to connect Web Map Services (WMS), Web Feature Services (WFS), Web Coverage Services (WCS) as well as Catalogue Services for the Web (CSW) from other servers.

III. METADATA CATALOGUE SUPERCAT

The central part of the EnviroGRIDS geoportal is a metadata catalogue. The catalogue enables harvesting of external catalogues published by other servers using the OGC Catalogue Service for the Web (CSW). User can then search other interconnected servers and discover and access available services through metadata records (Figure 2).

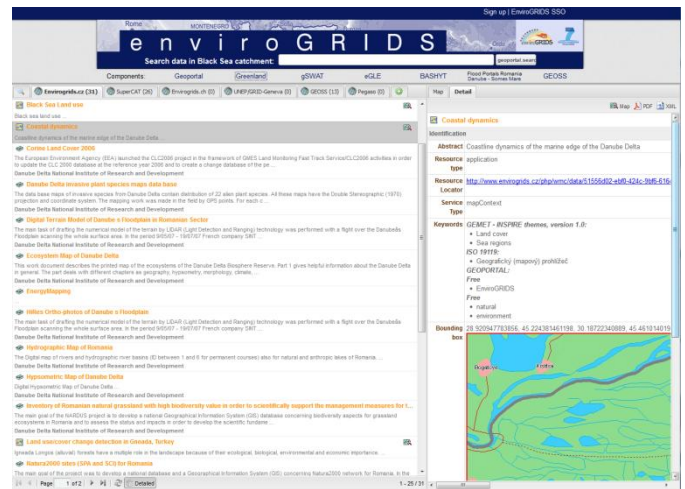


Fig. 2. Metadata search of the EnviroGRIDS geoportal showing the metadata records

There are many OGC web services available worldwide which may be widely used in different applications. In order to make SDI or other applications interoperable, it is crucial to have the services available and reliable at any time.

In real life, many problems occur with regard to operability and access to services. Several examples based on the authors experience can be mentioned:

- Some services do not have a public metadata record that can be used to search the services. The services are not catalogued.
- Services are usually registered in catalogues in order to discover them. However, the services cannot be simply searched by common search engines like Google or Bing. The services can be then searched only through catalogues in which the services are registered or by knowing the Uniform Resource Locator (URL) of the service.

- In many cases, metadata records are not up-to-date and links to services are invalid or not provided at all. Sometimes, the provided URL links to a web page or a viewer rather than to the service itself.
- Some catalogues do not respond.
- Service metadata can be catalogued in different ways. There are no common "global" rules on how to uniquely code certain elements, e.g. serviceType. Querying a catalogue may then result with an error.

The implementation of the EnviroGRIDS geoportal revealed that many OGC services from external servers show the above mentioned errors and this leads to user dissatisfaction. This experience led to the design and implementation of SuperCAT - a metadata catalogue for discovery of reliable services that are OGC compliant. In the first phase, the OGC Web Map Service (WMS) compliance testing was implemented and is further described in the next chapter.

SuperCAT is an independent and extended installation of the Micka metadata catalogue. Micka is a complex system for metadata management used for building Spatial Data Infrastructure (SDI) and geoportal solutions. It contains tools for editing and management of metadata for spatial information, web services and other sources (documents, web sites, etc.). It includes online metadata search engine, portrayal of spatial information and download of spatial data to local computer. Micka is compatible with obligatory standards for European SDI building (INSPIRE). Therefore it is ready to be connected with other nodes of prepared network of metadata catalogues.

SuperCAT is independent on the EnviroGRIDS geoportal and enables verification of the services which are accessed within the catalogue and other catalogues connected through the OGC Catalogue Services for the Web (CSW) 2.0.2.

IV. SERVICE TESTING USING SUPERCAT

A. Catalogue Interoperability Problems

As shown in the previous chapter, there are many problems with accessing web services. The authors performed an analysis of selected catalogues with regard to their functionality and compliance to the OGC CSW 2.0.2 and the INSPIRE specifications. The following problems were identified:

- The catalogue is not functioning. The catalogue was moved to another address, is temporarily unavailable or is password protected.
- The catalogue is not properly implemented according to the CSW 2.0.2 specification (older version, errors, etc.) or it is based on another standard.
- Many aspects in the current version of the CSW 2.0.2 specification are unclear or missing. As a result, service vendors implement them in different ways (e.g. harvesting of catalogues, behavior of different typenames, queryables).

- CSW should support Dublin Core (csw:Record) and other profiles are optional. There are implementations of ebRIM, ISO, FGDC and others. For INSPIRE the ISO AP 1.0 standard is mandatory. However, not all catalogues support it.
- CSW should support GET, POST and optionally SOAP protocol. In most implementations, POST and SOAP are not used. Not all catalogues implement these protocols for the GetRecord and GetRecordById operations.
- Query languages. CQL and OGC Filter should be supported. Some catalogues have errors in implementation or do not support full language properties.
- There is a mandatory set of queryables. INSPIRE requires additional ones which are usually not implemented by the service vendors.

B. Central Catalogue Implementation and Testing

In order to provide users easy access to services that are correctly implemented, a service metadata repository was built on our server. The repository:

- is CSW 2.0.2 ISO AP 1.0 compliant;
- supports the INSPIRE metadata profile and queryables;
- enables to register remote catalogues (CSW) and harvest them periodically;
- enables to harvest other services (WMS, WFS, WCS) and individual metadata files;
- enables verification of registered services.

The service metadata repository was tested using the following set of metadata catalogues. The most important catalogues of INSPIRE and GEOSS we complemented by other catalogues which are of knowledge by the authors:

- INSPIRE national catalogues of Austria, Belgium, the Czech Republic, Finland, France, Germany, Luxembourg, Poland, Portugal, Slovakia and the United Kingdom;
- GEOSS
- EuroGEOSS
- European Environmental Agency SDI
- EnviroGRIDS
- Habitats (<http://www.habitats.cz/>)
- One Geology Europe
- Plan4all (<http://www.plan4all.eu/>)
- World Health Organization

These resources are daily harvested and services are filtered (type=service) using the SuperCAT catalogue. A harvesting protocol form is generated for checking the availability of the

catalogues. Mail notifications are sent to corresponding users to ensure feedback.

The test of service availability is performed on daily basis for all services (only WMS at this phase). If a service is not running, the corresponding metadata record is not deleted but only hidden. This can ensure that temporarily unavailable services can be tested in the future.

C. Testing Results

2222 services were harvested from the registered catalogues. WMS services were analysed in detail. See the results in Table I. 87% of all the services were responding.

TABLE I. TESTING RESULTS

Service type code	Number	Responding (number)	Responding (%)
WMS	96	88	92
OGC:WMS	1418	1351	95
view	343	190	55
View	1	1	100
VIEW	2	2	100
View Service	73	73	100
Total	1860	1632	87

The following problems were identified during the testing:

- serviceType is in many cases ambiguous (see service type codes in Table 1). The INSPIRE Directive brings more confusion into service classification.
- There is no thematic classification for services and metadata are of poor quality (missing elements such as abstract, wrong bounding boxes, etc.). As a result, catalogue queries are not efficient. At least the INSPIRE theme keywords or other commonly used code lists would be a good step to introduce thematic classification.
- Unique service URL is not stated and in many cases it is coded in different ways.

V. CONCLUSIONS

The EnviroGRIDS portal is now part of the GEOSS infrastructure and is an important tool for capacity building in the Black Sea region. Currently, it offers a list of basic services for data integration, supports harvesting of available data, metadata and services and is an integrated access point for data in the region. The architecture enables to combine spatial data, metadata and services from different sources.

The core component of the system is represented by the metadata system management allowing to manage any type of information contained in the geoport and to use catalogue

services for sharing this information with other portals and social network sites.

The paper presented the innovative solution for efficient discovery of and access to spatial data services which are performed using SuperCAT - a metadata catalogue for discovery of reliable services that are OGC compliant.

The test demonstrates that many metadata catalogues including the GEOSS registry do not guarantee that registered services are operational. A better situation is with the INSPIRE national catalogues where the services are mostly guaranteed. But the test of accessibility and operability of services is still needed.

The future work includes implementation of an OGC Web Processing Service (WPS) client to execute external services and also for implementation of Sensor Observation Services (SOS).

ACKNOWLEDGMENT

The paper was prepared on the basis of the outputs of the EnviroGRIDS project - the solution was achieved with the financial co-funding of the European Commission within the Seventh Framework Programme with registration number 226740 and the name "Building Capacity for a Black Sea Catchment Observation and Assessment System supporting Sustainable Development".

REFERENCES

- [1] European Commission, 2010b. INSPIRE Glossary. Available at: <http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY> [Accessed February 17, 2012].
- [2] Charvát, K. et al., 2013. SDI, INSPIRE and other initiatives. In INSPIRE and Social Empowerment for Environmental Sustainability, Results from the HABITATS project. Spain: TRAGSA.
- [3] International Organization for Standardization, 1993. ISO/IEC 2382-1 Information technology -- Vocabulary -- Part 1: Fundamental terms. Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=7229 [Accessed June 15, 2011].
- [4] European Commission, 2010a. European Interoperability Framework (EIF) for European public services.
- [5] International Organization for Standardization, 2011. CEN/TR 15449 Geographic information - Standards, specifications, technical reports and guidelines, required to implement Spatial Data Infrastructures. Available at: <http://eSearch.cen.eu/eSearch/Details.aspx?id=6880372> [Accessed August 31, 2012].
- [6] European Parliament, 2007. 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). Available at: <http://eurlex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML> [Accessed May 31, 2012].
- [7] <http://www.earthobservations.org/geoss.shtml>
- [8] Janečka, K., Raitis, B. & Charvát, K., 2011. URM GeoPortal4Everybody: the modern platform for vocational education. In E-Learning, Distance Education or ... the Education of 21st Century. Sofia, Bulgaria.
- [9] Charvát, K. et al., 2011. Social Space for Geospatial Information. In IST-Africa 2011 Conference Proceedings. IST-Africa 2011. Gaborne, Botswana: HMC International Information Management Corporation.