DISCOVERY SERVICES FOR THE SIBERIAN EARTH SYSTEM SCIENCE CLUSTER

LORENZO BIGAGLI AND ROMAN GERLACH

Abstract. In the framework of a joint research concerning the study of advanced geomatics services to facilitate GMES applications in Spatial Data Infrastructures, we performed the analysis, design and prototype implementation of discovery services for the the Siberian Earth System Science Cluster, a large database of datasets and value-added products spanning central Siberia. The region represents a strong climate change hot spot in Northern Eurasia.

1. Introduction

The Department of Earth Observation of the Friedrich-Schiller-University, Jena (Germany) and the Earth and Space Science Informatics Laboratory at the Institute of Methodologies for Environmental Analysis of the Italian National Research Council (Italy) established a research collaboration to benefit from each others expertise and exchange knowledge, technical developments as well as research personal.

The collaboration aims at identifying and experimenting new methods and technologies for environmental monitoring, to support the scientific community, as well as to complement the traditional resource management systems and decision-support tools in general.

One aspect of the collaboration concerns the study of advanced geomatics services to facilitate GMES applications in Spatial Data Infrastructures (SDIs), with particular regards to the cataloguing and discovery of geospatial information.

The Department of Earth Observation of the Friedrich-Schiller-University is managing the Siberian Earth System Science Cluster (SIB-ESS-C), a large database of datasets and value-added products spanning the central Siberian region, and is currently pursuing the implementation of advanced access, discovery and processing services to these data, to generate and distribute products and information about central Siberia, along with advanced analysis support for Earth Sciences.

In this context, during a research visit supported by the 2008 Short-Term Mobility Program of the Italian National Research Council, the following activities have been carried out:

- Analisys of SIB-ESS-C intended users, use-cases, and data types and identification of the requirements for their cataloguing, discovery and access.
- Evaluation of the in-place infrastructure and solutions, as well as of the current international standards for cataloguing geospatial data, services and products.

Date: 13/04/2008.

This work was supported by the 2008 Short-term Mobility Program of the Italian National Research Council.

- Design, implementation and experimentation of a prototype catalogue service solution for SIB-ESS-C.

The next sections provide background information about the project topic and summarize the results of the above activities and the future development of the collaboration.

2. Background

It is estimated that there are currently around 100.000 in-situ stations and 50 environmental satellites.[1] Scientific and technological advancements in sensors, remote sensing and aerospace industry are set to increase exponentially the availability of geospatial information, in the near future.

Likewise, the advancements of research in environmental sciences, supported by the increasing capacity of computational platforms and telecommunication infrastructures, will allow to deepen our understanding of natural phenomena.

The United Nations Conference on Environment and Development, held at Rio de Janeiro from 3 to 14 June 1992, and the nineteenth special session of the General Assembly for the purpose of an overall review and appraisal of the implementation of Agenda 21, held in New York from 23 to 28 June 1997, explicitly stated that geospatial information plays a critical role in supporting the decision process and the management of environmental issues at a national, international and global scale. In fact, it is commonly estimated that 90% of the information in use by governments has a spatial connotation.

However, until now, traditional Geographic Information Systems in use by public administrations have scarcely integrated with scientific data, e.g. products stemming from Earth Observation, that are of great importance, for example, in scenarios of environmental risk. Likewise, real-time or near real-time applications, typical of monitoring and security scenarios, have been hindered by the quality of service of communication infrastructures, or by the available computational power.

Ensuring that the enormous future amount of knowledge be available to the Society (and in particular to decision-makers for risk mitigation, environmental protection and optimization of natural resources), calls for the identification of new methods and technologies for geospatial information management, capable to implement and effective integration among the existing Geographic Information Systems, Decision-Support Systems and Earth Observation Systems.

The continuous development of sophisticated ICT solutions provides the fundamental tools to tackle the above problem. On that premises, the international research community is strongly pursuing the specification and the standardization of frameworks (e.g. data and service models, with related profiles and extensions) of ICT solutions for geospatial information management, including Earth Observation and Environmental Monitoring.

The research focuses on the design and implementation of enabling infrastructures that support geospatial resources sharing by means of a minimum set of protocols, standard specifications and best practices. Such facilities, known as Spatial Data Infrastructures

[1] `http://ec.europa.eu/research/environment/themes/article_1357_en.htm`

(SDIs), can be defined as the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. Different hierarchical levels of SDIs are reckoned, e.g. global, regional, national, local. A SDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general.[3]

2.1. **State of play.** Several initiatives are related to the implementation of components of an SDI in various application sectors, including:

INSPIRE: the Infrastructure for Spatial Information in Europe,[2] recently adopted by the European Parliament as Directive 2007/2/EC. INSPIRE establishes a regional SDI in Europe, also addressing some aspects of environmental monitoring;

GEOSS: the Global Earth Observation System of Systems,[3] a worldwide effort to build upon existing national, regional, and international systems to provide comprehensive, coordinated Earth observations from thousands of instruments worldwide, transforming the data they collect into vital information for society;

GMES: the European Global Monitoring for Environment and Security,[4] promoted by the European Community and the European Space Agency to bring data and information providers together with users, so they can better understand each other and make environmental and security-related information available to the people who need it through enhanced or new services. GMES is one of the main European contributions to GEOSS.

2.2. **Open issues and scope of work.** The implementation of the above scenario entails addressing issues like:

- The difficulty of cataloguing and, hence, discovering geospatial information.
- The difficulty of accessing geospatial information, particularly in real-time or near real-time.
- The frequent incompleteness of observations (e.g. the presence of spatiotemporal gaps).
- The difficulty of tailoring geospatial information to the user needs.
- The limitations of processing systems (e.g. with respect to performance).
- The limitations of archiving systems (e.g. limited and/or short-lived storage, affecting processing of time series).
- The lack of interoperability between Earth Observation Systems, resource management systems and Decision-Support tools.

The presented research focuses on the first issue and, indirectly, on the last one. Referring to an ideal process of decision-making, the research scopes the analysis and experimentation of possible improvements to the Data Distribution/Management aspect, with special regards to:

- Data and service discovery;
- Earth observation data management (e.g. modelling, presentation);

[2] `http://www.ec-gis.org/inspire/`

[3] `http://earthobservations.org/geoss.shtml`

[4] `http://www.gmes.info`

- Geospatial information interoperability and mediation.

3. The Siberian Earth System Science Cluster

SIB-ESS-C is the follow-on activity to the EU funded SIBERIA-II project (Multi-Sensor Concepts for Greenhouse Gas Accounting of Northern Eurasia, EVG2-2001-00008).[5, 4, 6]

SIBERIA-II was a joint Russian-European remote sensing project that improved greenhouse gas accounting over a 300 Million ha area in the central Siberian region.[1] The overall objective of the SIBERIA-II project was to demonstrate the viability of full carbon accounting including greenhouse gases (GHG) on a regional basis using state-of-the-art environmental methods, biosphere modelling and advanced remote sensing technologies. The tools and systems which have been employed include a selected yet spectrally and temporally diverse set of 15 Earth observation datasets from 8 satellites, detailed GIS databases and two Dynamic Global Vegetation Models (the Lund-Potsdam-Jena LPJ-DGVM and the Sheffield-DGVM) to account for fluxes between land and atmosphere.

Following the SIBERIA-II project there was a clear need to preserve the results and data products created and make them accessible to a broad community (e.g. the global change and the modelling communities) for further research and exploitation.

Thus the SIB-ESS-C project was started, with the purpose of providing web-based tools and comprehensive information products derived from Earth Observation to support environmental and earth system research in Siberia, including:

- access to data products through standard interfaces (e.g. OGC CSW, WCS, WFS).
- continuous data product generation to build up time series.
- online visualization tools for spatio-temporal data analysis.
- integration wih Earth science models.

SIB-ESS-C is being funded by the Friedrich-Schiller University Jena (Germany) for the period commencing January 2006 until December 2009. Funding is granted for hard- and software as well as labour cost.[5]

Starting from the region of interest of SIBERIA-II, SIB-ESS-C is intended to cover the whole area of Siberia, spanning a vast number of ecosystems of in northern Eurasia ranging from the tundra, the boreal and temperate forests, mountainous areas and grasslands. The region is believed to play a critical role in global climate change and has been also defined by the International Geosphere-Biosphere Programme (IGBP) as one of the Boreal transects, representing a strong climate change hot spot in Northern Eurasia. Figure 1 shows the regions of interest of SIBERIA-II and SIB-ESS-C.

3.1. **Users and applications.**

The SIB-ESS-C system has been designed to provide data products and tools to the scientific community and international organisations working on global, continental or national scale. The target users come from a range of fields including:

- Earth system science modelling (input to models, validation of model results).
- Modelling of biogeochemical cycles.

[5]For more information, see `http://www.sibessc.uni-jena.de`.

FIGURE 1. The region of interest of SIBERIA-II (green-border areas) and SIB-ESS-C (red bounding box).

- Monitoring and modelling of vegetation dynamics (e.g shifting of tree line).
- Assessment of land-atmosphere interaction.
- Ecosystem response to global change.
- Support to convention implementation (e.g. Kyoto Protocol).
- Assessing the environmental impact of socio-economic development.

From the GEOSS perspective SIB-ESS-C data products and services are expected to contribute to the following societal benefit areas:

- Ecosystem.
- Climate.
- Water.
- Energy.
- Heath.
- Disasters.

3.2. **Use-cases.** The following actors are identified for SIB-ESS-C:

Contributors: group of researchers, producer of data to be published, provided with ad-hoc software instruments.

User: final (possibly remote) consumer of the system services, provided with ad-hoc software instruments.

Chain of acquisition: automatic process of data acquisition and generation of (part of) the associated metadata.

System administrator: manager of the system services, provided with ad-hoc software instruments.

The following main use-cases are identified:

(1) Data Publication

The system administrator or a contributor integrates a data product into the system by creating a record in the Metadata Database and copying the actual data file onto the Product Server. The Metadata record is being validated by the system and, if correct, inserted in the database. If the validation is unsuccessful, or if the dataset is already published, the system answers with an error message.

(2) Data Discovery

A user specifies a query and transmits it to the system. The system evaluates the query against the Metadata Database and responds with a list of data products that meet the search criteria. The query interface provides a minimum of three criteria: spatial, temporal and thematic.

Variant: If there is only small number of data products, only a list of all data products is provided (no querying).

(3) Data Visualization

From the list of data products returned by the Data Discovery process a user selects one or more data sets. The system then generates a map view (within the Web browser) of these data sets along with auxiliary data supporting orientation and navigation within the view.

(4) Data Access

The system provides access to data products through OGC web services (WCS, WFS, WMS). Users can directly incorporate a service into there application or receive and store the data as a file in common formats. Access is granted free of charge after a user registration procedure.

(5) Data Analysis

The system provides a Web interface to investigate spatial and temporal characteristics (e.g. changes/trends over time) of data products and to compare data products from multiple sensors and algorithms. A user selects one or two datasets (similar to the Data Discovery use case) and specifies the spatial and temporal coverage as well as the analysis method. According to the analysis method selected the system returns a graphical representation of the data set (e.g. time series plot, map). The service will be available for existing SIB-ESS-C data products, but also for external data sets if they are provided as an OGC service (WCS, WFS).

(6) Data Generation

Data Products are being derived from raw data by executing a predefined processing chain. During this process metadata is automatically created and stored in the Metadata database of the system. The final data product is being stored on the Product Server after successful validation by the system. The generation of new (or updated) data products is being triggered by an operator (contributor) or automatically by the system if new raw data becomes available to the system.

3.3. **Data types.** Table 1 summarizes the dataset and value-added products created within the SIBERIA-II project. In the initial phase of the SIB-ESS-C project, those data sets will form the basic set of products to be disseminated. Extending the existing data products to build up time series is one of the goals of SIB-ESS-C. As research is advancing and new algorithms and data products are being developed additional data sets of the region shall be included. In order to provide a comprehensive spectrum of data sets relevant for earth systems research collaboration with other data providers and research organisations to share data sets is highly desired.

3.4. **Prototype catalogue solution for SIB-ESS-C.** SIB-ESS-C comprises a legacy catalogue holding metadata of all data products. This catalogue is implemented onto

TABLE 1. Dataset and products of the SIBERIA-II project. The spatial extent of any listed data product spans the entire SIBERIA-II region of interest (see figure 1).

EO Product	Source	Temporal ext.	Spatial res.	Responsible Party
Phenology	SPOT-VGT, AVHRR	2000-2003 annual	1km, 10km	Center for the Study of the Biosphere from Space (CESBIO), France
Disturbances	MODIS, AVHRR ATSR-2	1992-2003 on yearly basis	1 km	Centre for Ecology and Hydrology Monks Wood, UK
Freeze/ Thaw	QuikSCAT	2000-2003	10km	TU Wien, Institute of Photogrammetry and Remote Sensing (IPF), Austria
Water bodies	ASAR WS	2003/2004	150m	TU Wien, Institute of Photogrammetry and Remote Sensing (IPF), Austria
Snow Depth	SSM/I	2000-2003	25km	Center for the Study of the Biosphere from Space (CESBIO), France
Snow Melt	SSM/I	2000-2003	25km	Center for the Study of the Biosphere from Space (CESBIO), France
Land cover	MODIS	2001-2004 annual	500m	University of Wales Swansea, UK
Continuous Field Land Cover	MODIS VCF, MODIS LC	2003	500m	Friedrich-Schiller-University Jena, Institut for Geography, Germany
Topography	SRTM/ GTOPO	2000	$3\,arcsec < 60°N$ $1\,km > 60°N$	Gamma Remote Sensing AG, Switzerland

an RDBMS, and users can perform data discovery through a web interface based on HTML.

This solution is not sufficient to implement the above use-cases. In particular, given the interdisciplinary nature of the intended users and applications, the adoption of interoperable interfaces and technologies was a driver for the design and experimentation of a catalogue solution for SIB-ESS-C.

The OGC CSW is the emerging standard for cataloguing services in geomatics, and its ISO Application Profile is the INSPIRE candidate recommendation for discovery services. Hence, these standards have been selected as the interface of the SIB-ESS-C

catalogue. The SIB-ESS-C catalogue is being implemented based on the GI-cat server[2], with an ad-hoc module that directly accesses the underlying database.

Moreover, a lightweight web interface to the system is being realized, based on AJAX technologies.

4. Conclusions and future Work

After the analysis of the target users and applications, the identification of the main use-cases and the evaluation of the existing standard and technologies, we have designed and experimented a catalogue solution implementing discovery services for the the Siberian Earth System Science Cluster, based on the OGC CWS ISO Application Profile.

When the system will be fully deployed, SIB-ESS-C will be a federated catalogue service complying to the ISO 19115/19139 standards. This will enable users to perform queries on external catalogues and in turn allow other registries to harvest information about SIB-ESS-C data holdings and services.

Access to SIB-ESS-C data products wil be provided through OGC Web Coverage, Feature, and Map Services, allowing users to directly integrate the data into their application or retrieve a file of the requested data product.

An advanced feature of SIB-ESS-C will be the online analysis tool to investigate spatial and temporal characteristics (e.g. changes/trends over time) of data products and their relationships (e.g. cross-correlation) or to assess uncertainty of parameters by intercomparing data products from multiple sensors and algorithms. This service will be availble for existing SIB-ESS-C data products, but also for external data sets if they are provided as an OGC service (WCS, WFS).

To provide an effective contribution to the field, the present research activity will participate in the current international frameworks for standardization and interoperability in Earth Observation and geospatial information management.

In particular, SIB-ESS-C will be contributed to the GEOSS Registry, as well as the lessons learnt in the design and implementation of the solution will be contributed to the relevant working groups of the mentioned initiatives.

A few research aspects to be possibly investigated in the future are: distributed discovery services based on P2P technologies; advanced acccess services for multidimensional data; integration of models via the OGC WPS interface. Another possible topic of further activity is the visual presentation of Earth Observation data, typically coverages, that is hard to implement in the general case, due to the complexity and heterogeneity of data structures and formats.

References

1. C. Beer, W. Lucht, C. Schmullius, and A. Shvidenko, *Small net carbon dioxide uptake by russian forests during 1981-1999*, Geophysical Research Letters **33** (2006), no. L15403.
2. L. Bigagli, S. Nativi, P. Mazzetti, and G. Villoresi, *Gi-cat: a web service for dataset cataloguing based on iso 19115*, DEXA '04: Proceedings of the Database and Expert Systems Applications, 15th International Workshop on (DEXA'04) (Washington, DC, USA), IEEE Computer Society, 2004, pp. 846–850.
3. Doug Nebert (Ed.), *The sdi cookbook*, GSDI, 2.0 ed., January 2004.

4. C. Schmullius, C. Beer, R. Gerlach, S. Hese, D. Knorr, M. Santoro, L. Skinner, and A. Luckman, *Siberia-ii - multi-sensor land cover products for greenhouse gas accounting of northern eurasia*, ForestSAT Conference (Umea, Sweden), 2005.
5. C. Schmullius, S. Hese, H. Balzter, W. Cramer, F. Gerard, R. Kidd, T. LeToan, W. Lucht, A. Luckman, I. McCallum, S. Nilsson, A. Petrocchi, S. Plummer, S. Quegan, A. Shvidenko, L. Skinner, S. Venevsky, S. Voigt, W. Wagner, U. Wegmüller, and A. Wiesmann, *Sensor systems and data products in siberia-ii - a multi-sensor approach for full greenhouse gas accounting in siberia*, ForestSAT Conference (Edinburgh, UK), 2002.
6. C. Schmullius, S. Hese, and D. Knorr, *Siberia-ii: A multi-sensor approach for greenhouse gas accounting in northern eurasia,*, Petermanns Geographische Mitteilungen **147** (2003), no. 2003/6.

National Research Council - Institute of Methodologies for Environmental Analisys (Italy)
URL: `http://www.imaa.cnr.it`
E-mail address: `bigagli@imaa.cnr.it`

Friedrich-Schiller-University - Department of Earth Observation, Jena (Germany)
URL: `http://www.eo.uni-jena.de`
E-mail address: `gerlach@uni-jena.de`