



27th International Cartographic Conference

Pre-conference workshop on

## **Spatial data infrastructures, standards, open source and open data for geospatial (SDI-Open 2015)**

jointly organized by the

Commission on Geoinformation Infrastructures and Standards,

Commission on Open Source Geospatial Technologies &

Open Geospatial Consortium (OGC)

20-21 August 2015

Brazilian Institute of Geography and Statistics (IBGE), Rio de Janeiro, Brazil

ISBN 978-1-77592-117-2



**International Cartographic Association**  
Association Cartographique Internationale



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

Serena Coetzee, Silvana Camboim, Antony K Cooper, Suchith Anand, Trevor Taylor,  
Rogerio Borba, Julia Strauch (Editors)

27th International Cartographic Conference

Pre-conference workshop on

## **Spatial data infrastructures, standards, open source and open data for geospatial**

### **SDI-Open 2015**



#### **Organizing committee**

Chair:

Silvana Camboim, Vice-Chair: ICA Commission on Open Source Geospatial Technologies

Members:

Rogerio Borba, Brazilian Institute of Geography and Statistics (IBGE)

Julia Strauch, Brazilian Institute of Geography and Statistics (IBGE)

Antony K Cooper, Chair: ICA Commission on Geoinformation Infrastructures and Standards

Serena Coetzee, Vice-Chair: ICA Commission on Geoinformation Infrastructures and Standards

Suchith Anand, Chair: ICA Commission on Open Source Geospatial Technologies

Trevor Taylor, Open Geospatial Consortium

Published in August 2015 by:

University of Pretoria, Pretoria, South Africa

International Cartographic Association (ICA)

ISBN 978-1-77592-117-2

The authors retain copyright over their work, while allowing the workshop organizers to place their unpublished work under a Creative Commons Attribution License, which allows others to freely access, use, and share the work, with an acknowledgement of the work's authorship and its initial presentation at this conference.

## Table of Contents

Foreword	iv
NSDI-BR - A historical and contemporary analysis <i>Rogério L. R. Borba, Wesley Silva Fernandez, Julia C. M. Strauch and Jano M. Souza</i>	1
The Australian National Map <i>Ana Belgun, Keith Grochow, Mats Henrikson, Peter Leihn, Alexander Mason, Michelle Nic Raghnaill, Kevin Ring and Bill Simpson-Young</i>	5
Open standards and formats used in the implementation of a marine spatial data infrastructure at the Directorate of Hydrography and Navigation of the Brazilian Navy <i>Juan Carballini and Flávia Mandarino</i>	9
SDI - National to Global: perspectives from the UK academic sector <i>Conor Smyth</i>	12
A decade of geoinformation sharing at ETH Zurich <i>Ionuț Iosifescu-Enescu, Cristina Iosifescu-Enescu, Nadia H. Panchaud, Angeliki Tsorlini and Lorenz Hurni</i>	19
Discussing the role of university in spatial data infrastructure construction: issues and challenges for UERJ-V-SDI <i>José Augusto Sapienza Ramos and Carlos Eduardo Gonçalves Ferreira</i>	24
Using open source software and open standards for implementing a European spatial data infrastructure via service cascade <i>Lassi Lehto</i>	29
Interoperable provision of geodata and services according to the INSPIRE Directive in German geoportal sites <i>Franz-Josef Behr</i>	33
Geoportal MTOP <i>Hebenor Bermúdez, Ana Fernández, Laura Marroni, Melissa Robert and Federico Godán</i>	35
Conceptual data modeling in the geospatial information technology toolkit (gittok) <i>Morishige Ota</i>	41
A solution for processing large files in the LASer (LAS) format using the message passing interface (MPI) and parallel file systems <i>Jeffrey Wendel, Michael P. Finn, John Kosovich, Jeff Falgout and Yan Liu</i>	45

## Foreword

The discovery, access, exchange and sharing of geographic information and services among stakeholders from different levels in the spatial data community is facilitated through a *spatial data infrastructure (SDI)*. *Standards* are key for the quality and development of interoperable geographic information and geospatial software. The drive for access to geographic information has led to its publication as *open data*, i.e. freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. According to a report by the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), the use of *open source* solutions is likely to increase significantly in the future as a viable alternative to proprietary suppliers. Open source software for geospatial, geographic information standards and open data policies are therefore significant for SDI development and implementation.

SDI researchers and practitioners were invited to submit extended abstracts that describe SDI case studies where

- open source software for geospatial is used in the technical implementation; and/or
- geospatial standards are implemented; and/or
- geographic information is accessible as open data.

Original contributions in English were reviewed by members of the organizing ICA Commissions. Upon acceptance, a revised abstract had to be submitted for inclusion in the workshop proceedings. One of the authors presented the abstract at the workshop. Not all submitted abstracts are included in the proceedings.

The proceedings of this workshop represent a variety of SDIs: there are examples of national (Behr 2015; Belgun et al. 2015; Bermudez et al. 2015; Borba et al. 2015) and international SDI initiatives (Lehto 2015); of an SDI in the hydrographic domain (Carballini and Mandarino 2015) and SDI initiatives in academia (Iosifescu-Enescu et al. 2015; Ramos and Ferreira 2015; Smyth 2015). Finally, Wendel et al. (2015) and Ota (2015) focus on technical solutions in the SDI context.

Borba et al. (2015) describe the activities leading to the decree that established the Brazilian National SDI and conclude with a critical reflection of the pros and cons of using open source and open standards in an SDI. According to Belgun et al. (2015), the Australian National Map's easy-to-use spatial standards-based web browser interface for sharing and exploring data from many different sources and in many different formats has facilitated the publication of open geospatial data. Carballini and Mandarino (2015) describe IHO standards-based database-driven cartographic production in a marine SDI.

Smyth (2015) and Iosifescu-Enescu et al. (2015) each present an academic SDI with a technical implementation comprising a mix of open source and proprietary components. Ramos and Ferreira (2015) make a case for universities to become producers of spatial (research) data in an SDI. Lehto (2015) describes the implementation of a specialization of the Web Feature Service (WFS) standard that aggregates data in real-time from multiple services. Behr (2015) aims to analyse the geoportal sites of German federal states with a focus on the INSPIRE Directive, license issues and open access to data, while Bermudez et al (2015) present a case study of an open source geoportal implementation for the Uruguay Ministry of Transport and Public Works.

Ota (2015) focuses on capacity building, an issue also raised by other authors, and describes a tool that can be used to teach geographic information standards. The proceedings are concluded with the interesting experiments conducted by Wendel et al. (2015). They significantly reduced the processing time of large lidar point clouds by implementing a message passing interface (MPI) in order to achieve parallel processing.

The motivations for using open source software in the SDI initiatives vary: maturing open source geospatial technologies are mentioned (Iosifescu-Enescu 2015; Smyth 2015), while the opportunity to extend open source software through novel research is particularly attractive in academic SDIs (Iosifescu-Enescu 2015). Borba et al. (2015) and Smyth (2015) comment on the challenge of finding a sustainable business model in the world of open source and open data.

Serena Coetzee

Associate Professor: Department of Geography, Geoinformatics and Meteorology, University of Pretoria

Vice-Chair: ICA Commission on Geoinformation Infrastructures and Standards



## **NSDI-BR - A historical and contemporary analysis**

Rogério L. R. Borba<sup>1,2</sup>, Wesley Silva Fernandez<sup>3</sup>, Julia C. M. Strauch<sup>1</sup>, Jano M. Souza<sup>2</sup>

[rogerio.borba@ibge.gov.br](mailto:rogerio.borba@ibge.gov.br), [wfernandes@anp.gov.br](mailto:wfernandes@anp.gov.br), [julia.strauch@ibge.gov.br](mailto:julia.strauch@ibge.gov.br), [jano@cos.ufrj.br](mailto:jano@cos.ufrj.br)

<sup>1</sup>*Federal University of Rio de Janeiro (UFRJ). Institute Alberto Luis Coimbra of Postgraduate Studies and Research in Systems Engineering and Computing Program – Rio de Janeiro, RJ – Brazil*

<sup>2</sup>*Geography and Statistics Brazilian Institute (IBGE) – Rio de Janeiro, RJ – Brazil*

<sup>3</sup>*National Agency of Petroleum, Natural Gas and Biofuels (ANP) – Rio de Janeiro, RJ - Brazil*

Keywords: NSDI, GIS, government

### **1. NSDI-BR: Background information and relevant policies**

Recognizing the importance of geospatial information in decision-making and the formulation of public policies, the National Spatial Data Infrastructure (NSDI) was established in Brazil, through the 6,666 decree, in November 2008. The presidential decree that created the NSDI had the following objectives (CONCAR, 2008):

1. Take care of the management, production and dissemination and use of geospatial data across all levels of government;
2. Promote and target the use of standards and norms approved by the National Commission of Cartography (CONCAR) for the production of geospatial data by public institutions in the federal, state, district and municipal spheres.
3. Avoid duplication of efforts and waste of resources to obtain geospatial data from public administration, through the dissemination of metadata relating to such data available in the public institutions of all levels of government.

However, for the establishment and implementation of the NSDI several factors, milestones and activities preceded its implementation, for example:

1. In 2001, a National Cartographic Plan was prepared by CONCAR, a collegiate body of the Planning and Budget Management Ministry (MOP), aiming at the integration of cartographic plans of public institutions producing geospatial information (CONCAR, n.d.);
2. Creation of specialized committees aimed at standardization of geospatial data and metadata, as in many cases the production of geospatial data and metadata were not following common standards (DELOU, 2006; CAMBOIM, 2013);
3. Some antecedents and parallel activities in public institutions and academia to promote, research and adopt Free and Open Source Software (FOSS) and development IDEs (CÂMARA *et al.*, 2006);
4. Legal framework of the NSDI and action plan for the implementation of the NSDI that not only aimed at management, but also guide the implementation and maintenance of the NSDI through well-defined cycles (CONCAR, 2010).

The following relevant policies are enumerated:

1. The very decree establishing the NSDI, which requires all federal institutions to share their spatial data and metadata and follow the standards, set for the NSDI and the rules

on the National Cartography. In addition, follow the Electronic Government Interoperability standards (e-PING, n.d.);

2. Brazil is a signatory of the Open Government Partnership (OGP) treaty aimed at giving more transparency to actions taken by governments, empower citizens, fight corruption, and harness new technologies to strengthen governance (OGP, 2011);
3. The OGP treaty influenced the enactment in 2011 of the Law of access to information (Law No. 12,527) at the federal level, which established a series of disclosure obligations of public information on the Internet from government agencies (federal, district, state and municipal) and citizen service procedures;
4. The civil framework of the internet provides that the Internet should treat equally any data packets, regardless of content, origin and destination of the service. In addition, this law removes the responsibility of sites on the content generated by third parties (Marco civil da internet, 2014). One consequence of this law is that it favors the establishment of more participatory and collaborative SDIs to society, as it removes the legal uncertainty because of content generated by third parties;
5. Creation of the National Open Data Infrastructure (INDA) which was also based on the NSDI. INDA is the government policy on open data (INDA, 2012).

## **2. A description of how open source, standards and open data are implemented in the SDI**

The CONCAR, in its mission of coordinating the implementation of the National Cartographic Policy, has among its responsibilities to approve norms and standards for NSDI (CONCAR, 2008). Currently there are five specialized committees in CONCAR (CAMBOIM, 2013):

1. National Digital Cartographic Library Structuring Committee (CEMND): aiming at structuring a National Digital Map Library (MND), shared use, in which the reference mapping products would be made available by the various components agencies of the National Cartographic System (SCN).
2. Geospatial Metadata Structuring Committee (CEMG): aims to standardize metadata. For this, a Brazilian geospatial metadata profile was created. It is noteworthy that the Brazilian profile is a subset of the ISO standard 19115.
3. Geographic Names Committee (CNGEO): promotes and disseminates the standardization of geographical names in the territory, as well as of foreign names.
4. Standardization of Cadastral Mapping Committee (CNMC): works towards the creation of norms and standards for cadastral mapping.
5. National Spatial Data Infrastructure Implementation Committee (CINDE): develops the necessary actions and activities for the implementation of NSDI, as the proposed courses of action in its action plan.

In relation to the creation or use of geospatial free/libre/open source software (FOSS), Brazil has good experiences as presented by CÂMARA *et al.* (2006). For example, TerraLib is a FOSS that allows its use in the development of various GIS tools. In addition, research and work has been done in other institutions such as IBGE using OGC standards and several FOSS technologies (BORBA *et al.*, 2007) and later served as basis for recommendation and use of free software adherent to OGC standards for NSDI (CONCAR, 2010). Also a symbol library was proposed to NSDI, taking into account FOSS, national and OGC standards, in-

cluding Symbology Encoding (SE) and Styled Layer Descriptor (SLD) (FERNANDES, 2013).

Currently, most of the institutions that are part of the NSDI use FOSS adherent to the OGC standards. For example, the Brazilian Institute of Geography and Statistics (IBGE), the main information provider of the country (IBGE, n.d.) and manager of the Brazilian Directory Geospatial Data (DBDG) of the NSDI (CONCAR, 2008) is one of the NSDI nodes using software such as I3geo (Mapserver- for viewing), GeoServer (for services), Geonetwork (for metadata) and PostgreSQL / PostGIS (Spatial-DBMS).

Finally, following the open data policy, geospatial data is made available as open.

### **3. An evaluation of the motivators, benefits, barriers and challenges concerning implementation of open source software, standards and open data**

INDE-BR had also as inspiration other SDI initiatives as evidenced in chapters one and two of the action plan (CONCAR, 2010). This was a result of the first and second initiatives of SDIs that served as reference (WILLIAMSON *et al.*, 2003) and there is no doubt that the recommendation and use of FOSS and open standards in building the NSDI-BR brought several advantages. On the other hand, there are many challenges that must be faced for NSDI-BR to thrive (BORBA *et al.*, 2014). For example, with some exceptions, there is no active society participation, since FOSS employed does not provide characteristics that allow the effective society engagement. There are also criticisms and controversies involving some Brazilian standards as technical specifications for Geospatial Vector Data Structuring (EDGV). There is concern with the support and development of FOSS as pointed out in (CÂMARA *et al.*, 2012). For this, new business models need to be proposed and implemented to give support to these initiatives.

Some policies and initiatives related to the use of open data, open source and open standards clearly show the path adopted by NSDI-BR. However, after seven years of existence, there is still much to be done.

### **References**

- BORBA, R., FERNANDES, W., DINIZ, V., FRANÇA, V., 2007. Aplicações de web mapping na Coordenação de Cartografia do IBGE. In: XXIII Congresso Brasileiro de Cartografia - I Congresso Brasileiro de Geoprocessamento, Rio de Janeiro.
- BORBA, R., STRAUCH, J., SOUZA, J., COLEMAN, D., 2014. Architectural and technological aspects for SDI of third generation. Presented at the AutoCarto, Pittsburgh, PA, USA.
- CÂMARA, G., FONSECA, F., MONTEIRO, A.M., ONSRUD, H., 2006. Networks of innovation and the establishment of a spatial data infrastructure in Brazil. *Inf. Technol. Dev.* 12, 255–272.
- CÂMARA, G., VINHAS, L., de SOUZA, R.C.M., 2012. Free and open source GIS: will there ever be a geo-Linux?, in: *Geospatial Free and Open Source Software in the 21st Century*. Springer, pp. 229–245.
- CAMBOIM, Silvana Philippi, 2013. ARQUITETURA PARA INTEGRAÇÃO DE DADOS INTERLIGADOS ABERTOS À INDE-BR.
- CONCAR (Comissão Nacional de Cartografia) [WWW Document], n.d. URL <http://www.concar.ibge.gov.br/> (accessed 3.14.15).
- CONCAR, 2008. Decreto no 6666.
- CONCAR, 2010. Plano de Ação para implantação da INDE.

- DELOU, Alberto, 2006. Estruturação de bases cartográficas para SIG um estudo comparativo dos modelos do IBGE e da DSG (Dissertação de Mestrado). Rio de Janeiro, RJ, Brasil.
- e-PING — Programa de Governo Eletrônico Brasileiro - Site Oficial [WWW Document], n.d. URL <http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padrees-de-interoperabilidade> (accessed 3.3.15).
- FERNANDES, W.S., 2013. Criação de uma biblioteca de símbolos cartográficos utilizando os padrões Symbology Encoding (SE) e Styled Layer Descriptor (SLD) do OGC. Revista Brasileira de Cartografia.
- IBGE (Instituto Brasileiro de Geografia e Estatística) [WWW Document], n.d. URL <http://www.ibge.gov.br/home/> (accessed 3.15.15).
- INDA - Infraestrutura Nacional de Dados Abertos — Programa de Governo Eletrônico Brasileiro - Site Oficial [WWW Document], 2012. URL <http://www.governoeletronico.gov.br/acoes-e-projetos/Dados-Abertos/inda-infraestrutura-nacional-de-dados-abertos> (accessed 3.8.13).
- Marco civil da internet [WWW Document], 2014. URL [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/112965.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm) (accessed 3.15.15).
- OGP - Open Government Partnership | [WWW Document], 2011. URL <http://www.opengovpartnership.org/> (accessed 3.16.15).
- WILLIAMSON, I.P., RAJABIFARD, A., Feeney, M.-E.F., 2003. Developing spatial data infrastructures: from concept to reality. CRC.



## **The Australian National Map**

*Ana Belgun, Keith Grochow, Mats Henrikson, Peter Leihn, Alexander Mason, Michelle Nic Raghnaill,  
Kevin Ring, Bill Simpson-Young*  
[{first.last}@nicta.com.au](mailto:{first.last}@nicta.com.au)

*National ICT Australia (NICTA)*

Keywords: Australia, Cesium, Leaflet, open source

### **1. Introduction**

This paper explores the implementation of the NationalMap spatial data infrastructure in Australia ([www.nationalmap.gov.au](http://www.nationalmap.gov.au)). Built on open source technologies such as Cesium and Leaflet, the NationalMap platform allows the Australian Government to simplify public access to the tremendous variety of geospatial data it possesses. The platform is an intuitive web interface in which spatial data can be visualised without the need for external software. The platform has proved popular and is being adopted by several Australian State Governments.

NICTA has developed NationalMap for the Australian Government's Department of Communications. NationalMap draws on the open data policies designed at both Federal and State level including: The Government's Plan for the Digital Economy & e-Government (Liberal Party of Australia 2013); Open public sector information: from principles to practice (OAIC) (Australian Government - Office of the Australian Information Commissioner 2013); The Open Data Network; NSW Open Data Policy (NSW Government - Dept. of Finance); and the ACT Open Data Policy (ACT Government) among others. A full list of open data policies and documentation is available (Australian Government - Dept. of Finance).

The Australian Minister for Communications launched NationalMap in his opening speech of the Locate15 Spatial Conference (Malcolm Turnbull 2015). According to the Minister "the inclusion of open government data from all levels of Australian government provides for the first time a demonstration of how seamless open data exploitation can be achieved".

Based on the tenets of open data, the NationalMap platform provides an improved data infrastructure and visualisation capability for Australia. It brings together dispersed information from all levels of Government, into an easily searchable, viewable and fully customisable map-based view. The map is designed and organised through the collaboration between a technical development team and a user experience team to create a geospatial web platform which meets the needs of a broad range of users.

### **2. Use of open source, standards, and data**

NationalMap was designed as a fully open framework. The web front-end connects directly to specific data custodian repositories using open protocols, including Open Geospatial Consortium (OGC) Web Map and Feature Services (WMS, WFS), open data formats such as OGC Keyhole Markup Language (KML) and GPS Exchange Format (GPX). The National Map also enables the

Federal government open data platform Data.gov.au to make its spatial data easily accessible, by directly accessing their Comprehensive Knowledge Archive Network (CKAN) catalogue.

Relying on the strong geospatial visualisation skills at NICTA, the spatial data infrastructure (SDI) is built with Cesium, an open source WebGL virtual globe and map engine that runs in a web browser without plugins. Cesium is a project which NICTA is co-developing with an international community of software engineers and the original developers, Analytical Graphics, Inc. (AGI). NICTA has developed upon the National Map project and produced an open source product named TerriaJS (Terria 2015), which is now the basis of the National Map and several other NICTA projects.

The National Map platform adds an intuitive data layer menu with interactive layer visualisation controls, presenting a powerful and user-friendly interface. The interface supports adding arbitrary WMS and WFS layers, as well as supporting drag and drop of common file formats straight into the browser window. To lower the barrier to producing useful spatial visualisations of ad hoc data, it is possible to use Comma Separated Value (CSV) files, with columns referring to latitude and longitude, postcode, state, or any of a number of different geographical division systems. These CSV files are then easily imported into the system using drag and drop or a file loading interface, and are displayed on the map with automatic legends and information retrieval popups.

The National Map platform also uses Cesium Language (CZML), a simple but extremely powerful JavaScript Object Notation (JSON) based data format developed for Cesium. CZML makes it relatively easy to visualise feature based data, but in particular, it is useful for displaying time-varying feature data.

### **3. Motivation, benefits, barriers and challenges**

The motivation for the creation of National Map came from the difficulties imposed on users of Data.gov.au in exploring geospatial data provided to them in a meaningful way, and in evaluating the relationships between these datasets. Furthermore, it is believed that by providing a powerful interface to explore geospatial data, the production and release of new datasets to the public would be accelerated. By making it simple for data custodians to explore their own data on the NationalMap before making it public, via internal services or drag and drop, they are encouraged to open their data to the public.

Several business sectors rely on geospatial data in order to make informed decisions about the future of their business, particularly in the Renewable Energy sector. NICTA is developing the Australian Renewable Energy Mapping Infrastructure (AREMI, [www.nationalmap.gov.au/renewables](http://www.nationalmap.gov.au/renewables)) platform for the Australian Renewable Energy Agency (ARENA). The aim of AREMI is to lower the cost and effort of sourcing early stage geospatial data when developing and prospecting for new renewable energy projects. The AREMI platform is based on TerriaJS and is now provided as an official Australian Government service together with the NationalMap. Through this project, previously non-public datasets and several datasets produced as a result of ARENA funded projects will be available on the AREMI platform. The availability of these datasets in one location is an excellent example of how NationalMap is

accelerating the release of new geospatial datasets. This project has been made possible by the open source nature of the NationalMap project.

One of the greatest barriers so far in creating AREMI is working with potentially sensitive data sources. Datasets which reveal fine grained population and census data, electricity infrastructure, and defence lands and infrastructure etc. can have commercial, privacy or national security implications. These must be addressed before the data can become publicly available.

Occasionally, data custodian organisations lack the expertise required to produce high quality datasets and web based services, which poses another challenge. In some instances, NICTA provides assistance to custodians to produce these datasets and services, with the intention that the custodians can subsequently update and maintain them.

#### **4. Recommendations**

As work continues on the National Map platform, there are still many obstacles to overcome. Our goal is to develop a sustainable platform which allows for ease of addition, and exploration of, new and current geospatial datasets. We aim to provide a platform which allows data custodians to be in charge of their own data, so that it can be kept as up to date as possible. In future projects, we will work with data custodians as early as possible, to ensure that they are headed in a direction which makes integration into open standard platforms as easy as possible. We believe that using open standards is a must for making these datasets available, but we also believe that it is important to encourage custodians to make the raw data freely accessible in simple formats so that it can be processed programmatically, to produce new tools which combine various datasets.

It is important to work alongside stakeholders to demonstrate the value of making their data available. To make this easier in the future, we will need to discover and develop ways of anonymising and aggregating data which might be sensitive in nature, so that we can produce useful tools without revealing personal or commercially sensitive data.

#### **5. Conclusion**

The National Map platform provides an improved data infrastructure and visualisation capability for Australian Government data. Its aim is to bring together dispersed information from all levels of Government, into an easily searchable, viewable and fully customisable map-based view.

Our goal is to develop a sustainable platform which allows for ease of addition and exploration of new and current geospatial datasets. We aim to provide a platform which allows data custodians maintain ownership of their own data, so that it can be kept as up to date as possible.

Since its inception, the NationalMap spatial data infrastructure has shown that opening government data and maximising its use is highly dependent on the ease of access. This includes the combination of technology (a simple web browser), content (various and numerous datasets

in one place) and promotion. By making vast amounts of geospatial data accessible to anyone through a modern web browser, we have made open geospatial data available to everyone.

### **Acknowledgements**

NationalMap is an initiative of the Australian Government, Department of Communications. We are grateful to the Department of Communications, Geoscience Australia (who host the official NationalMap service) and the many Australian Federal and State agencies who have provided their data services for access by the National Map. The AREMI project has been funded by the Australian Renewable Energy Agency (ARENA). NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

### **References**

- ACT Government, *Open Data Policy by ACT Government*. Available: <http://actgov.github.io/opendatapolicy/>. [16 March 2015].
- Australian Government - Dept. of Finance, *Australian Government Data Initiatives and Policies - An interactive mind map to help understand the landscape*. Available: [http://www.finance.gov.au/sites/default/files/The\\_Government\\_Data\\_Landscape\\_in\\_Australia\\_Export\\_2014\\_1\\_14.doc](http://www.finance.gov.au/sites/default/files/The_Government_Data_Landscape_in_Australia_Export_2014_1_14.doc). [16 March 2015].
- Australian Government - Office of the Australian Information Commissioner 2013, *Open public sector information: from principles to practice*. Available: <http://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-reports/open-public-sector-information-from-principles-to-practice>. [16 March 2015].
- Liberal Party of Australia, *The Coalition's plan for the digital economy & e-Government | Liberal Party of Australia*. Available: <http://www.liberal.org.au/latest-news/2013/09/02/coalition%E2%80%99s-plan-digital-economy-e-government>. [16 March 2015].
- Malcolm Turnbull, M.P., *Speech to the Locate 15 Conference: The Power of Open Data*. Available: <http://www.malcolmturnbull.com.au/media/speech-to-the-locate-15-conference-the-power-of-open-data>. [16 March 2015].
- NSW Government - Dept. of Finance, *DFS C2013-6 Open Data Policy*. Available: <http://www.finance.nsw.gov.au/sites/default/files/DFSC2013-6-CircularandOpen-Data-Policy.pdf>. [16 March 2015].
- Terria, *A spatial data platform that provides spatial predictive analytics* <http://terria.io> [2 July 2015].



## **Open standards and formats used in the implementation of a marine spatial data infrastructure at The Directorate of Hydrography and Navigation of the Brazilian Navy**

*Juan Carballini<sup>1</sup>, Cmdr. Flávia Mandarinino<sup>2</sup>*  
[juan.carballini@caris.com](mailto:juan.carballini@caris.com), [flavia@chm.mar.mil.br](mailto:flavia@chm.mar.mil.br)

<sup>1</sup>CARIS, 115 Waggoners Lane, Fredericton, NB, Canada, E3B2L4

<sup>2</sup>Brazilian Navy Hydrographic Centre, Rua Barão de Jaceguay, s/no. - Ponta da Armação, Niterói - RJ – Brazil - 24048-900

Keywords: marine spatial data infrastructure, ocean mapping

### **1. Introduction**

In the last few years, the Brazilian economy has been growing to the point that today Brazil is one of the biggest markets in the world. This is also due in great part to the attention the nation pays to its coast, ports and rivers in an effort to guarantee proper use of natural resources and safe transportation through its massive water bodies.

Hydrographic organizations around the world are responsible for the compilation and management of large volumes of marine spatial data, which is used primarily to ensure safe navigation within their national waters, as well as the proper management of natural resources, national security, environmental protection and many other tasks that are essential to the existence of a nation. The Directorate of Hydrography and Navigation of the Brazilian Navy (DHN) is responsible for mapping the Brazilian jurisdictional waters and creating and maintaining the corresponding nautical cartography and related products. Although hydrographic mapping can be financially costly it is a very important matter for nations, especially in a large country such as Brazil, where natural resources are one of the main components of the country's economy.

### **2. Mapping the Brazilian waters**

Charting the Brazilian waters is not a simple undertaking, especially given that Brazil has a coastline of approximately 8.500 km, its economic exclusive zone is approximately 3,500,000 km<sup>2</sup> and there are more than 40.000 km of navigable inland waterways that need to be mapped and maintained. The implementation of a marine spatial data infrastructure is helping DHN in mapping Brazilian waters in order to provide better data, products and services to the mariners and stakeholders, and also will be able to reduce the production times.

According to the International Hydrographic Organization (IHO), a nautical chart is defined as a cartographic product specifically designed to meet the requirements of marine navigation, showing depths of water, nature of bottom, elevations, configuration and characteristics of the coast, existing dangers and aids to navigation. These real world features can be represented in a paper chart, electronic navigational chart (ENC) or a raster navigational chart (RNC) which are also referred to as nautical charts, hydrographic charts, or simply charts. DHN has the mandate to produce paper charts, ENCs and RNCs for the maritime sector, as well as for the inland

waterways. In order to be able to tackle the creation and maintenance of these nautical products DHN has to follow different international open standards and specifications, such as IHO S-4 and INT1/INT2 for paper chart production and the IHO S-57 for ENC production. In addition, there are also standards and product specifications for Inland ENC production that are regulated by the Inland ENC Harmonization Group (IEHG). At the moment DHN has to maintain around 660 paper charts for the maritime and inland waters as well as 137 ENCs and there are plans to produce around 20 ENCs more in order to cover the main ports and the Brazilian territorial waters.

### **3. Data volumes**

The volume of hydrographic data that can be collected is huge and it creates new requirements to handle the increasing data volumes in processing, management, and analysis as well as product creation. For example a quick and coarse estimation suggests that 2TB of disk space may be required per 24 hours of high resolution data collection of imagery and bathymetry in order to use this information to determine depths and classify the nature of the seabed and identify obstructions and ship wrecks, which constitute in themselves potential hazards to navigation.

The bathymetric data from hydrographic surveys, until about 2010 was stored in "flat files", making it difficult to search the data and organize metadata. The use of bathymetric databases to store this information, along with associated metadata, has been very beneficial. The metadata can be used to search for bathymetry data within the database and can provide the criteria for data manipulation and analysis. Metadata compliant with the ISO19115 standard can accompany data exported from the system so that the data can be used by others with confidence.

### **4. Challenges and data storage**

With the increased volumes of data being collected and the demand for multiple products Hydrographic Organizations are entering the realm of "Big Data" and the proper management of this data constitutes a very big challenge for agencies. There is a myriad of occurrences in which the same marine spatial information can be used to satisfy requirements for a wide range of other applications including; defence, coastal zone management, environmental studies, natural resources management, fisheries, security, inundation modelling and many more. The implementation of a Marine Spatial Data Infrastructure (MSDI) using open industry formats and standards from organizations like the International Hydrographic Organization (IHO), International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC), allows DHN to ensure interoperability within the organization and with other stakeholders, as well as increase the use of the marine spatial information in the greater GIS community.

In 2009, DHN initiated the migration to a new solution for the cartographic production, based on a computing environment where all cartographic features are stored in a single Oracle database. The cartographic products (paper charts, raster charts and ENCs) are produced from the same source information, stored and validated in the database. This integrated solution ensures consistency between products eliminates redundancies in the validation and facilitates the better synchronization in the production of paper charts, raster charts and ENCs. The data is

stored using the S-57 data model that is an open format and standard for transfer digital hydrographic data. The changes in workflows from traditional file based chart production systems to the use of relational database management systems with open standards and formats as well as the implementation of metadata profiles allow DHN to gain efficiencies in the administration, visualization, production and distribution of data, and data storage.

## **5. Discovering marine data**

The hydrographic data is not being used only for nautical chart production and DHN would like to make the data more accessible for the general public and other stakeholders through open standards and formats for web mapping. In 2014 DHN started a pilot project to extend the workflows to include data distribution via web. The idea is to distribute through the web some of the data including, vector representations, raster charts, and bathymetry in the near future. The OGC services; WMS (Web Map Service) will be used to allow users to browse the map, WFS (Web Feature Service) will be used for query and features download, and WCS (Web Coverage Service) will be used for bathymetry data downloads.

## **6. The use of marine information beyond the nautical charts**

The use of marine information will be maximized in the near future with the adoption of the new Universal Hydrographic Data Model S-100, developed by the IHO. The S-100 standard will use and extend the ISO 19000 series of geographic standards for hydrographic, maritime and related information. A new series of product specifications based on the S-100 standards will be developed for nautical charting, Nautical Publications, Gridded Data, Maritime Boundaries, etc. The new S-100 data model is a more flexible standard than the previous S-57 one and will allow DHN to provide hydrographic and marine information in a more interoperable way and the data will be discoverable by more users because there will be more compatibility with web services for analysis and access to the data and products. The use of extended metadata included in the S-100 standard for the hydrographic data will also help DHN and the users to have more information about the data and the products.

## **7. Conclusion and benefits**

At the moment, there is a big effort to migrate the hydrographic related data from different formats and in single files to this database driven environment, because the data validation should be made during the data compilation process and not on the products. After this initial phase, the production of the first products, and their maintenance (update) becomes faster, since the system indicates to the reviewer only what has been modified and needs to be validated. In 2014, the first paper charts were produced using this new system. The paper and raster nautical charts of the Tocantins River were produced following a new production work-flow using databases for bathymetric data collected during the hydrographic surveys, and for the cartographic production. The systems are fully compatible and use interoperable open standards that allow optimizations in the production work-flows, reducing the time for the generation of various products and traceability of cartographic information because the features are stored once but used many times. The system is ready to extend the data model to support the new S-100 which will make the use of hydrographic data easier and allows preparing and integrating the additional layers of information into the web mapping application that is a key component in this MSDI implementation.

## **SDI – National to Global: perspectives from the UK academic sector**

*Conor G. Smyth, PhD, FRGS, CGeog. (GIS)*  
[conor.smyth@ed.ac.uk](mailto:conor.smyth@ed.ac.uk)

*EDINA, The University of Edinburgh, Scotland*

Keywords: academic SDI; geospatial standard; open source; open data; services

### **1. The UK national academic Spatial Data Infrastructure (SDI)**

Geodata Services, EDINA, based at the University of Edinburgh, develops and delivers a portfolio of world-class online services and expertise that benefit research and education in the United Kingdom (UK) and beyond and that form a comprehensive spatial data infrastructure (SDI) on behalf of the academic community.

As part of service delivery, Geodata Services have promoted and utilised the use of open source software (OSS), international open geospatial standards, and open geographic data as integral components in the development and promotion of the academic SDI typified by: the development and delivery of innovative online services (subscription/open); the provision of online resources; and extensive user support (helpdesk, training). Increasingly, work in the mobile application development domain is providing newer, innovative (often open) delivery platforms incorporating responsive design principles. This work extends into the citizen science domain too. The importance of digital infrastructures is highlighted by Smyth (2016).

Geodata expertise has also extended into other sectors with integral components of the UK national and regional government SDIs supported by EDINA, e.g. the UK Location Infrastructure metadata editor has been developed by EDINA as part of [www.data.gov.uk](http://www.data.gov.uk) (Locationmde.data.gov.uk, 2015).

From a UK policy perspective, the work of Geodata services at EDINA forms a significant part of the contribution that Jisc (a national education ICT charity), makes as a champion for the use of digital technologies in the UK education, research and skills sectors; acting as a Jisc-designated centre for digital expertise and online service delivery, EDINA is charged to fulfil the Jisc Strategic Objectives in their national Geospatial Solutions area (Jisc, 2015).

Given the nature and scale of service operations (active user bases > 250k+), there are many opportunities for components and tools of the EDINA infrastructure to inform, or have applications in other countries or sectors to empower, improve outcomes/benefits and increase impacts in a cost-effective way.

In 2014, EDINA was awarded the ‘RSGS Bartholomew Globe award’ for excellence in the assembly, delivery or application of geographical information through cartography, GIS and related techniques (RSGS, 2015).

### **2. The importance of effective SDI services at national level**

The principal Geodata services (Edina.ac.uk, 2015a) that form the basis of the UK academic SDI are:

- **Digimap** – online web mapping and data download service (with comprehensive mapping/data and learning resources) (Wikipedia, 2015).



- **Digimap for Colleges** - online web mapping service (Digimapforcolleges.edina.ac.uk, 2015).
- **Digimap for Schools** - online web mapping service (& learning resources) (Digimapforschools.edina.ac.uk, 2015).
- **Digimap Mapstream** – a WMS offering licensed Ordnance Survey (OS) data (Mapstream.edina.ac.uk, 2015).
- **Digimap Openstream** - a WMS offering Open OS data (Openstream.edina.ac.uk, 2015).
- **GoGeo** – comprehensive open geospatial resource ‘discovery’ service utilizing GeoNetwork (includes information, learning resources, events and services) (Gogeo.ac.uk, 2015a; Mathys, 2015; Geonetwork-opensource.org, 2015); includes a metadata editor tool, **GeoDoc** (Gogeo.ac.uk, 2015b). The service promotes the use of ISO standards for the discovery, sharing (**ShareGeoOpen** repository for open data access to freely contributed datasets), re-use and creation of standards compliant metadata (Sharegeo.ac.uk, 2015). Also, **GoGeo App** for iphone.
- **UNLOCK** – open middleware (gazetteer and geo-referencing) service to extract and locate place-names in text documents and metadata; also, search tool for place-names, locations and shapes across geographic information sources. **UNLOCK Places** (GitHub, 2014) is an open service using open data sources (Edina.ac.uk, 2015b) and **UNLOCK Text** is a RESTful API utilising parsers (Ltg.ed.ac.uk, 2015; GitHub, 2015a). UNLOCK powers data discovery in GoGeo.
- **Fieldtrip GB App** – Android/iPhone mobile app for data capture including schema compilation using open datasets.
- **Fieldtrip Open** – Android mobile app software development code kit (SDK) available on GitHub for global use (GitHub, 2015b).

## 2.1. Technical implementation of open source software

Whilst there are varied technical implementations, in terms of our main web mapping service – Digimap, the systems architecture is predominantly predicated on OSS use (Opengeospatial.org, 2015a; Opengeospatial.org, 2015b), with Shibboleth authentication (Shibboleth.net, 2015); Geodata services runs its own Shibboleth IdP to allow in-house access and to create teaching accounts to showcase our services externally. Open services are accessible to all, not requiring registration or subscription.

Specifically, Digimap is a multi-tiered application stack providing access to comprehensive geospatial information and maps with ‘mapping’ and ‘download’ client functionality. The main mapping stacks are: Topographic, Geology, Hydrosatial, Historic and Environmental.

The architecture is split over multiple (task specific) machines across two geographical locations, with concurrent LIVE-LIVE capability to load balance across sites as well as allowing engineers to deal with site specific problems and/or to upgrade applications seamlessly without user impact/downtimes. Overall, a DEV/BETA/LIVE setup environment exists to allow full unobstructed lifecycle development, testing and release.

A conceptual architecture for Digimap is shown in Figure 1.

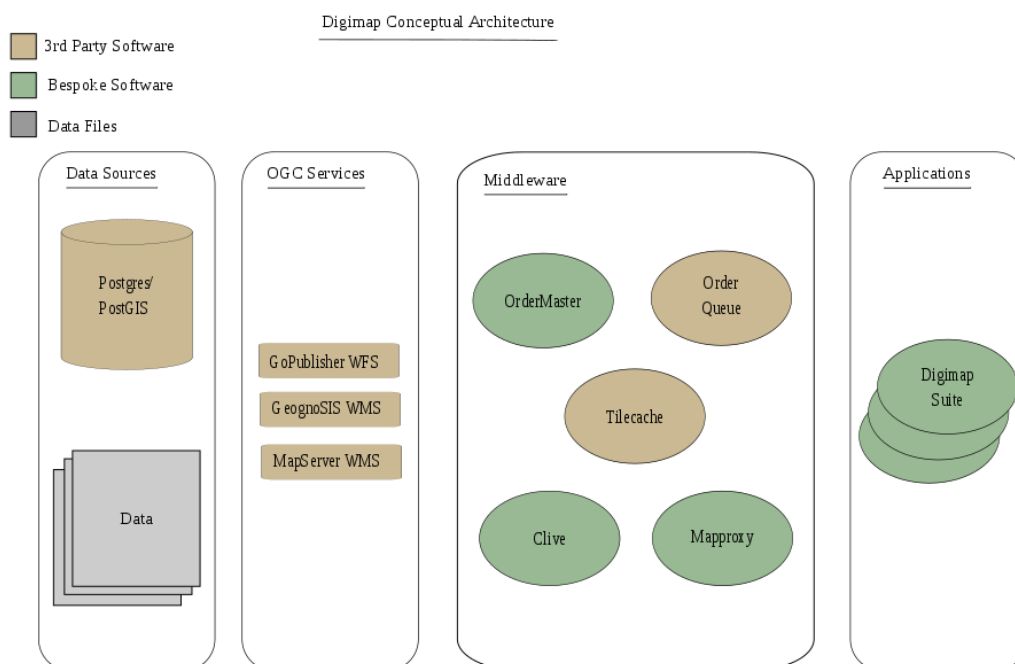


Figure 1. Digimap Conceptual Architecture

A number of open source tools, standards protocols, frameworks and script libraries are currently implemented to provide comprehensive GIS functionality: Java; Apache Tomcat & Webserver (Tomcat.apache.org, 2015; Activemq.apache.org, 2015; Group, 2015); Tilecache (Tilecache.org, 2015); OpenLayers (Openlayers.org, 2015); Mapserver (Mapserver.org, 2015); PostgreSQL (Postgresql.org, 2015); PostGIS (Developers, 2015).; Spring; and ExtJS, with a requirement for some proprietary software components, e.g. Cadcorp GeognoSIS (Cadcorp.com, 2015)., Snowflake GoPublisher software (Snowflakesoftware.com, 2015), and formerly, Oracle.

All our bespoke applications are written in Java SE8 (Docs.oracle.com, 2015) and run on Tomcat application servers, some of which are fronted by Apache web servers (depending on whether access to shared geographic information is required, or not).

## 2.2. Geospatial standards

Where applicable, Geodata Services utilise OGC standards (Openeospatial.org, 2015c), namely: WMS; WFS; and WCS, as well as a range of international metadata standards and profiles such as: ISO 19115 (Iso.org, 2015).; Federal Geographic Data Committee (Fgdc.gov, 2015); Dublin Core Metadata Initiative (Dublincore.org, 2015); UK AGMAP 2.1 (XML/PDF) (Gogeo.ac.uk, 2015c).; UK GEMINI 2.2 (BSI 1ST/36, CEN/TC287, ISO/TC211) (User, 2015); EU DIRECTIVE 2007/2/EC - INSPIRE – Infrastructure for Spatial Information in Europe (Inspire.ec.europa.eu, 2015). Workshops on metadata standards are an important part of educating those in the sector.

In terms of Digimap, Mapserver and GeognoSIS are OGC WMS applications that handle map generation. PostGres databases (formerly Oracle) hold all data that the WMS servers interact with to generate maps. Oracle was used to hold OS Mastermap data – however GoPublisher software now supports PostGres affording a more open software infrastructure.

### **2.3. Geographic information and open data**

A diverse range of geographic information and data is utilised and managed within Geodata services, furthermore, datasets are deposited in ShareGeo Open, accessible to anyone.

In addition to third-party national coverage licensed datasets from public and commercial data suppliers (via Jisc), there are many open and accessible datasets in utilization, ranging from organisations such as OS (Ordnancesurvey.co.uk, 2015a) and BGS (bgs.ac.uk, 2015a) (at times integrated into subscription services) to more diverse international open datasets.

With trends in open data, it is expected that more and more open data will form part of our offering. Importantly, the evaluation and prioritization of new digital data content is an on-going process influenced by community need (for subsequent delivery).

Key ‘licenced’ datasets are: OS, BGS, Historic (Landmark.co.uk, 2015), Marine - Hydrospatial (Seazone.com, 2015), Land-use (Ceh.ac.uk, 2015). Key ‘open’ datasets are: OS (Ordnancesurvey.co.uk, 2015b), BGS (bgs.ac.uk, 2015b), Open Street Map (Openstreetmap.org, 2015), contributed open data (in ShareGeo). In UNLOCK, a range OS open data is used, plus Natural Earth, DEEP (Digital Exposure of English Place-names) and GeoNames (Edina.ac.uk, 2015b).

## **3. Why a SDI and services matter to UK academia**

### **3.1. Motivators**

The central premise for establishing and maintaining a single national academic SDI supported by services has been to consolidate skills, expertise, technical, data and support infrastructures to benefit the UK academia; note that open services exist too.

### **3.2. Benefits**

The primary benefit of an established SDI is that of provision and (authenticated) access to standards based services, data and knowledge or support. EDINA has considerable expertise in the development and management of national online geospatial content services that eliminate duplication at the institutional level promoting economies of scale in areas such as data acquisition, licensing (where relevant), management, and access (lower overall costs). facilitate enhanced teaching and learning.

### **3.3. Barriers**

As a national SDI acting on behalf of Jisc, there are some barriers – relating to: (a) access (b) licensing and/or (c) funding. These can be at service and/or data content levels.

(a) access to services is either by subscription or entirely ‘open’. In the former, data by default is accessed via active institutional service subscriptions.

(b) licensing of content can be a barrier from a cost and/or service feasibility perspective.

(c) the ability to provide relevant services and content is often funding constrained.

### 3.4. Challenges

In an era of rapid sectoral and technology change, there are challenges: increased user expectations; reduced funding; and increasing demands on resources to continuously improve (multi-platform) services, are significant. Also, 'open' data necessitates business model re-evaluation to ensure future service sustainability.

### 3.5. Future developments

In terms of strategic vision, Geodata services seek to expand their expertise and provision of geospatial content services to support the sector as well as establish new collaborative partnerships to develop innovative projects and services in the UK and internationally; including a greater focus on mobile applications development. With maturity, OSS has the potential to remove the need for proprietary software.

## 4. Conclusion

The development of a national academic SDI has served to benefit users for over 15 years. The impact of EDINA's work is far-reaching and has been built on an enviable reputation for innovation and the adoption of largely open geospatial technologies and standards. Whilst primarily focussed on the academic sector, EDINA supports external partners, such as UK government. Many activities are exemplar models for wider applications elsewhere.

## Acknowledgements

The author would like to acknowledge: the Geoservices team; the University of Edinburgh; data content/software providers; and Jisc as a substantial funder.

## References

- Activemq.apache.org, (2015). *Apache ActiveMQ™ -- Tomcat*. [online] Available at: <http://activemq.apache.org/tomcat.html> [Accessed 30 Jun. 2015].
- bgs.ac.uk, (2015a). *British Geological Survey (BGS) | A world-leading geoscience centre*. [online] Bgs.ac.uk. Available at: <http://www.bgs.ac.uk/> [Accessed 29 Jun. 2015].
- bgs.ac.uk, (2015b). *Data collections | OpenGeoscience | Free data | British Geological Survey (BGS)*. [online] Bgs.ac.uk. Available at: <http://www.bgs.ac.uk/opengeoscience/collections.html> [Accessed 29 Jun. 2015].
- Cadcorp.com, (2015). *Interactive Web Mapping and Web GIS Software Service GeognoSIS | Cadcorp*. [online] Available at: <http://www.cadcorp.com/products/server/geognosis> [Accessed 30 Jun. 2015].
- Ceh.ac.uk, (2015). *Environmental Information Data Centre at CEH*. [online] Available at: <http://www.ceh.ac.uk/data/index.html> [Accessed 29 Jun. 2015].
- Developers, P. (2015). *PostGIS - " Spatial and Geographic Objects for PostgreSQL*. [online] Postgis.net. Available at: <http://postgis.net/> [Accessed 30 Jun. 2015].
- Digimapforcolleges.edina.ac.uk, (2015). *Digimap for Colleges*. [online] Available at: <http://digimapforcolleges.edina.ac.uk/> [Accessed 29 Jun. 2015].
- Digimapforschools.edina.ac.uk, (2015). *Digimap for Schools*. [online] Available at: <http://digimapforschools.edina.ac.uk/> [Accessed 29 Jun. 2015].
- Docs.oracle.com, (2015). *The Java® Language Specification*. [online] Available at:



- <http://docs.oracle.com/javase/specs/jls/se8/html/index.html> [Accessed 30 Jun. 2015].
- Dublincore.org, (2015). *DCMI Metadata Basics*. [online] Available at: <http://dublincore.org/metadata-basics/> [Accessed 29 Jun. 2015].
- Edina.ac.uk, (2015a). *Maps and Data*. [online] Available at: <http://www.edina.ac.uk/maps> [Accessed 29 Jun. 2015].
- Edina.ac.uk, (2015b). *Unlock - Unlock Places: Data Sources*. [online] Available at: <http://edina.ac.uk/unlock/places/sources.html> [Accessed 29 Jun. 2015].
- Fgdc.gov, (2015). *Standards - Federal Geographic Data Committee*. [online] Available at: <https://www.fgdc.gov/standards/> [Accessed 29 Jun. 2015].
- Geonetwork-opensource.org, (2015). *Home- GeoNetwork opensource*. [online] Available at: <http://geonetwork-opensource.org/> [Accessed 29 Jun. 2015].
- GitHub, (2014). *edina/unlock-places-api*. [online] Available at: <https://github.com/edina/unlock-places-api> [Accessed 29 Jun. 2015].
- GitHub, (2015a). *Berico-Technologies/CLAVIN*. [online] Available at: <https://github.com/Berico-Technologies/CLAVIN> [Accessed 29 Jun. 2015].
- GitHub, (2015b). *edina/fieldtrip-osm*. [online] Available at: <https://github.com/edina/fieldtrip-osm> [Accessed 29 Jun. 2015].
- Gogeo.ac.uk, (2015a). *GoGeo - The portal to spatial data and information*. [online] Available at: <http://www.gogeo.ac.uk/gogeo/> [Accessed 29 Jun. 2015].
- Gogeo.ac.uk, (2015b). *GoGeo - Metadata Learning Resources*. [online] Available at: <http://www.gogeo.ac.uk/gogeo/metadata/geodoc.htm> [Accessed 29 Jun. 2015].
- Gogeo.ac.uk, (2015c). *GoGeo - Metadata Workshops*. [online] Available at: <http://www.gogeo.ac.uk/gogeo/metadata/agmap.htm> [Accessed 29 Jun. 2015].
- Group, D. (2015). *Welcome! - The Apache HTTP Server Project*. [online] Httpd.apache.org. Available at: <http://httpd.apache.org/> [Accessed 30 Jun. 2015].
- Inspire.ec.europa.eu, (2015). *INSPIRE> WELCOME TO INSPIRE*. [online] Available at: <http://inspire.ec.europa.eu/> [Accessed 29 Jun. 2015].
- Iso.org, (2015). *ISO 19115-1:2014 - Geographic information -- Metadata -- Part 1: Fundamentals*. [online] Available at: [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=53798](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53798) [Accessed 29 Jun. 2015].
- Jisc, (2015). *Maps and geospatial data | Jisc*. [online] Available at: <http://www.jisc.ac.uk/content/maps> [Accessed 29 Jun. 2015].
- Landmark.co.uk, (2015). *Historic Data you can really get your teeth into*. [online] Available at: <http://www.landmark.co.uk/content/historic-data-you-can-really-get-your-teeth> [Accessed 29 Jun. 2015].
- Locationmde.data.gov.uk, (2015). *UK Location Metadata Editor*. [online] Available at: <https://locationmde.data.gov.uk/ukmelogin> [Accessed 29 Jun. 2015].
- Ltg.ed.ac.uk, (2015). *- Software*. [online] Available at: <http://www.ltg.ed.ac.uk/software.html> [Accessed 29 Jun. 2015].
- Mapserver.org, (2015). *Welcome to MapServer - " MapServer 7.0.0-beta1 documentation*. [online] Available at: <http://mapserver.org/uk/> [Accessed 30 Jun. 2015].
- Mapstream.edina.ac.uk, (2015). *MapStream for Schools*. [online] Available at: <http://mapstream.edina.ac.uk/> [Accessed 29 Jun. 2015].

- Mathys, T. (2015). GoGeo: A Jisc-funded service to promote and support spatial data management and sharing across UK academia. In: *International Association of Social Science Information Service and Technology*. Minneapolis: IASSIST.
- Opengeospatial.org, (2015a). *Web Feature Service | OGC*. [online] Available at: <http://www.opengeospatial.org/standards/wfs> [Accessed 30 Jun. 2015].
- Opengeospatial.org, (2015b). *Web Map Service | OGC*. [online] Available at: <http://www.opengeospatial.org/standards/wms> [Accessed 30 Jun. 2015].
- Opengeospatial.org, (2015c). *OGC Standards | OGC*. [online] Available at: <http://www.opengeospatial.org/standards/is> [Accessed 29 Jun. 2015].
- Openlayers.org, (2015). *OpenLayers 3 - Welcome*. [online] Available at: <http://openlayers.org/> [Accessed 30 Jun. 2015].
- Openstream.edina.ac.uk, (2015). *Digimap OpenStream*. [online] Available at: <http://openstream.edina.ac.uk/registration/> [Accessed 29 Jun. 2015].
- Openstreetmap.org, (2015). *OpenStreetMap*. [online] Available at: <http://www.openstreetmap.org/about> [Accessed 29 Jun. 2015].
- Ordnancesurvey.co.uk, (2015a). *Britain's mapping agency | Ordnance Survey*. [online] Available at: <https://www.ordnancesurvey.co.uk/> [Accessed 29 Jun. 2015].
- Ordnancesurvey.co.uk, (2015b). *OS OpenData Supply - Download or order OS OpenData*. [online] Available at: <https://www.ordnancesurvey.co.uk/opendatadownload/products.html> [Accessed 29 Jun. 2015].
- Postgresql.org, (2015). *PostgreSQL: The world's most advanced open source database*. [online] Available at: <http://www.postgresql.org/> [Accessed 30 Jun. 2015].
- RSGS, (2015). *Bartholomew Globe - RSGS*. [online] Available at: <http://rsgs.org/inspiring-people/medals-awards/bartholomew-globe/> [Accessed 29 Jun. 2015].
- Seazone.com, (2015). *Marine Maps & Data for Web Applications & GIS | SeaZone*. [online] Available at: <http://www.seazone.com/> [Accessed 29 Jun. 2015].
- Sharegeo.ac.uk, (2015). *ShareGeo Open*. [online] Available at: <https://www.sharegeo.ac.uk/> [Accessed 29 Jun. 2015].
- Shibboleth.net, (2015). *Shibboleth*. [online] Available at: <https://shibboleth.net/> [Accessed 29 Jun. 2015].
- Smyth, C.G., (2016). 'Where' matters: keeping apace with Geo-ubiquity in a digital world. In: Baker, D. and Evans, D., *Digital information strategies: from applications and content to libraries and people*. Oxford: Chandos Publishing Ltd. (in print).
- Snowflakesoftware.com, (2015). *GO Publisher*. [online] Available at: <http://www.snowflakesoftware.com/products/gopublisher/> [Accessed 30 Jun. 2015].
- Tilecache.org, (2015). *TileCache, from MetaCarta Labs*. [online] Available at: <http://tilecache.org/> [Accessed 30 Jun. 2015].
- Tomcat.apache.org, (2015). *Apache Tomcat - Apache Tomcat 7 Downloads*. [online] Available at: <https://tomcat.apache.org/download-70.cgi> [Accessed 30 Jun. 2015].
- User, S. (2015). *Standards Committee*. [online] Agi.org.uk. Available at: <http://www.agi.org.uk/join-us/agi-groups/standards-committee> [Accessed 29 Jun. 2015].
- Wikipedia, (2015). *Digimap*. [online] Available at: <https://en.wikipedia.org/wiki/Digimap> [Accessed 29 Jun. 2015].

## A decade of geoinformation sharing at ETH Zurich

*Ionuț Iosifescu-Enescu, Cristina Iosifescu-Enescu, Nadia H. Panchaud, Angeliki Tsorlini, Lorenz Hurni*

*{iosifescu,ciosifescu,nadia.panchaud,atsorlini,lhurni}@ethz.ch*

*Institute of Cartography and Geoinformation, ETH Zurich, Switzerland*

Keywords: GeoVITE, geospatial data sharing, corporate SDI, open source, standards, ETH Zurich

### 1. Background Information about the Spatial Data Infrastructure at ETH Zurich

ETH Zurich receives proprietary geodata from the Swiss Federal Office of Topography (swisstopo) on the basis of an annual, renewable contract. The organization of the received geodata is file-based, usually in folders named with the year and the quarter of delivery.

A file-based delivery system is not optimal for geodata sharing and the complexity of accessing geospatial data can be challenging for the non-geodata specialists, especially for researchers in other fields than geography, geomatics or cartography. Researchers and students are often interested in a specific region and they do not need nor want to download the complete dataset covering the entire Switzerland. Users spend unnecessarily precious time studying metadata in order to identify and select the correct data tile or layer, having also to deal with GIS tasks such as clipping, merging and reprojecting geospatial data.

As consequence, since 2004, the Institute of Cartography and Geoinformation in cooperation with the ETH Library has been building a corporate Spatial Data Infrastructure (SDI) at ETH Zurich, with the goal of providing ETH Zurich employees with direct access to Geoinformation and Geoservices through a user-friendly geoportal.

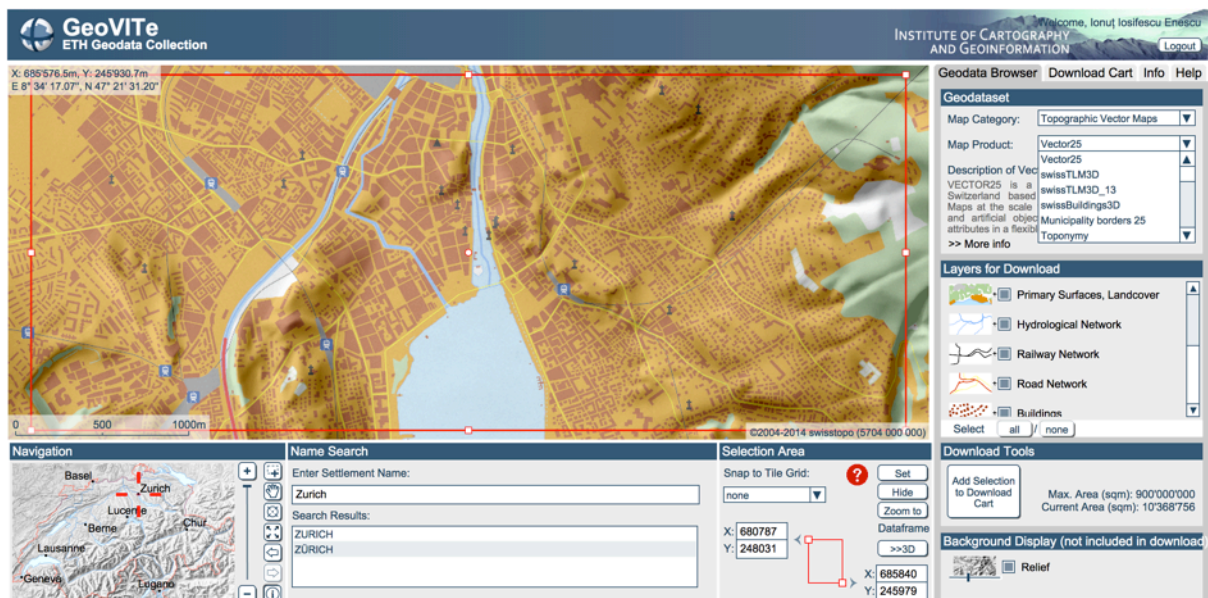


Figure 1. The GeoVITE portal

The GeoVITE (GEOdata Visualization and Interactive Training Environment) portal allows its users to access geodata through a standard Web browser as shown in *Figure 1*. The user can visually browse the spatial data, select the desired area and download the necessary

data in a straightforward manner. The portal provides a clear visualization of available geodata products, it facilitates the choice of the right data (map product and location), and it hides the complexities related to merging and extracting the needed data extent.

## 2. Architectural Components of the GeoVITe SDI

The SDI technical implementation used for GeoVITe is entirely service-driven. We use a three-tier-architecture, with a clear separation between presentation, services and data. The user has access only to a “thin” Web-based interface, which handles the majority of user interactions by sending requests and listening to responses from visualization and geoprocessing services. The services access the corresponding data in the database, do the necessary processing and send the responses back to the user interface. From a software engineering viewpoint, these background services are entirely transparent for the users: they only see that the interface reacts to their commands such as navigating through the different products available, zooming or panning the map, selecting the right area for download, and enabling them to straightforwardly download the desired data (Iosifescu et al, 2011; 2013).

As mentioned before, the corporate SDI at ETH Zurich has been in development since 2004 and the GeoVITe portal has been actively used for the download of geodata since 2010. Currently at version five, GeoVITe portal has experienced a long history of implementation technologies, migrating from proprietary technologies to open software and standards. The first prototype version, finished in 2008, was mainly based on proprietary technologies such as IBM’s DB2 (database), ESRI ArcSDE (database middleware) and ESRI ArcIMS (services). The subsequent four versions, from 2009 to 2012, continued to be mainly powered by proprietary software. However, the fifth version, operating since 2013, harmoniously blends open source and proprietary technologies as shown in *Figure 2*.

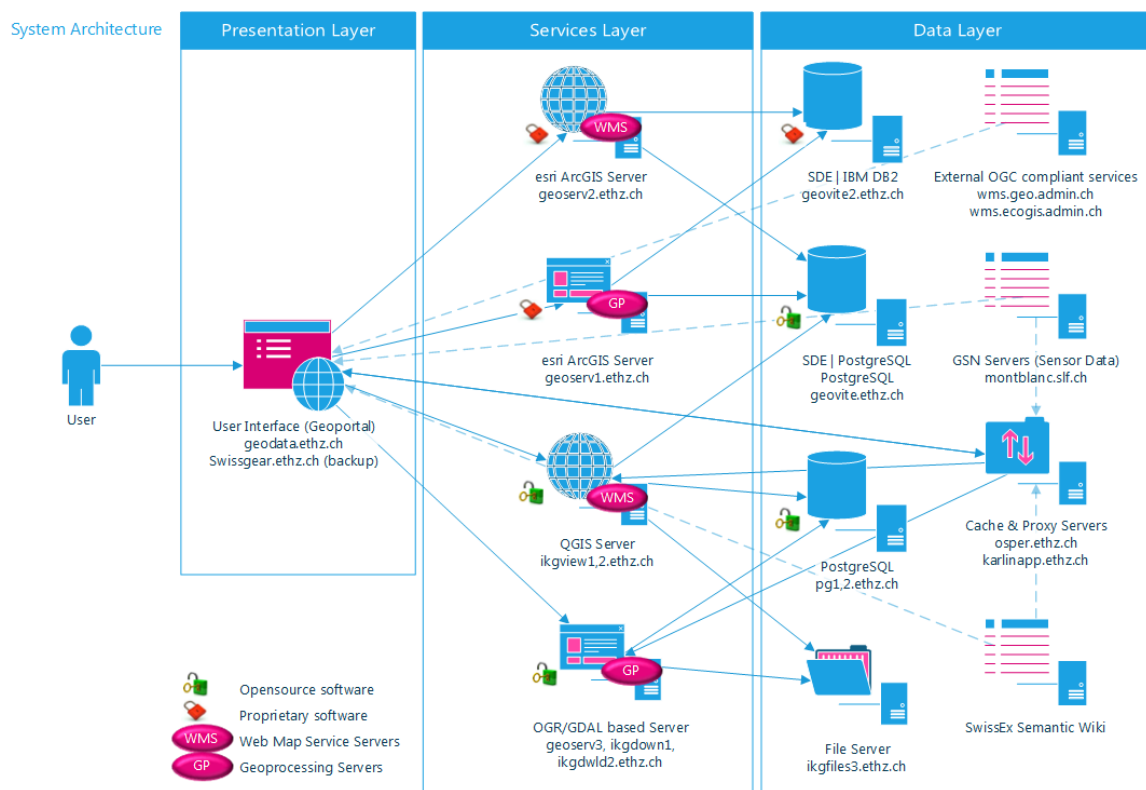


Figure 2. The three-tier architecture of the GeoVITe platform (Iosifescu et al., 2014)





The third motivation is the interest in high-quality open source software that lowers the costs of geodata publication in related SDIs, such as GEOIDEA.RO (shown in *Figure 3*) geoportal. The main objective of the project GEOIDEA.RO (GEodata Openness Initiative for Development and Economic Advancement in Romania) is to improve the scientific basis for the adoption of an open geodata model in Romania.

In the main GeoVITe framework, we aim to offer both proprietary and open source technologies an equal footing. However, according to the requirements of various related projects, we can quickly switch the underlying portal technologies from proprietary to open-source and vice-versa, especially due to the high interoperability of free and open source software. Therefore, the main benefit of integrating open source software into GeoVITe is the vendor-independence and the flexibility to adapt quickly to a fast-changing geospatial environment. Furthermore, the concept of openness has a serendipitous influence on several other activities, especially when taking into consideration the main duty of academia, namely to propagate knowledge. For example, we support the user-friendly dissemination of existing open geospatial datasets for which redistribution inside our organisation is permitted and we strive to establish new cooperation channels related to geospatial open source technologies with the establishment of an ICA-OSGeo-ISPRS Open Source Geospatial Laboratory at ETH Zurich.

Finally, we have to mention that although there are no barriers concerning the integration of open source technologies in GeoVITe, there is still a major challenge related to the sustainable use of open source technologies in this project. Unfortunately, in-depth open source know-how is still restricted to a minority of people from the ICA-OSGeo Open Source Geospatial Laboratory at ETH Zurich. While users of proprietary software can be easily sent to training courses, achieving an equally comprehensive understanding of open source technologies is not easy. There does not exist an all-encompassing “open source” training. The training courses that do exist are restricted to specific aspects of individual software packages, whereas one needs a solid comprehension in many aspects of the underlying open source software libraries. For example, before understanding and using the spatial extension PostGIS with various GIS software such as QGIS, a user should know first about managing and administering the geospatial database PostgreSQL, to have some IT background, a good knowledge of SQL, an understanding of spatial analysis and a general GIS know-how. Furthermore, a user should be aware of the many underlying libraries used in various open source GIS desktop and server software (such as GDAL/OGR) in order to correctly identify and fix problems that may arise. Therefore, a priority will be to extend the ICA-OSGeo-ISPRS Open Source Geospatial Laboratory at ETH Zurich as a hub for the transfer of knowledge and as well as a support group for the continuing development and use of open source software in GeoVITe.

#### **4. Conclusion**

Over the last decade the GeoVITe platform evolved through many technological changes. The most important change occurred in the latest version, when open source technologies were being introduced on an equal footing with proprietary software. As a result, PostgreSQL/PostGIS, GDAL/OGR, QGIS Server and QGIS have become important elements of the open source part of GeoVITe, which harmoniously coexist with proprietary software such as ESRI ArcGIS Server.

The open source technologies in the geospatial domain reached the tipping-point towards functional equivalence with proprietary software, and they represent a mature solution for the implementation of spatial data infrastructure and geoportals. Unfortunately, there is still an absence of good training possibilities for open source software and a solid understanding of open source technologies can only be gradually achieved with a lot of effort invested in apparently different, but interconnected directions. Furthermore, a correct use and development of open source software is driven by an individual's passion, which no training can provide.

### **Acknowledgements**

The integration of open source technologies in GeoVITE was and continues to be partially supported by a Swiss grant in the frame of the Swiss Contribution to the enlarged European Union through the GEOIDEA.RO project (Project Number: IZERZ0-142129) as well as by swissuniversities.ch through the Geodata4SwissEDU project.

### **References**

- Iosifescu, I., Hurni, L. (2010) Web-Service Driven Cartography. Proceedings of Autocarto 2010, Orlando, Florida, United States.
- Iosifescu, C., Iosifescu, I., Jenny, H., Hurni, L. (2011) GeoVITE - A Service-driven Solution for an On-demand, User-friendly Web Access to Geodata. Proceedings of the 25th International Cartographic Conference, Paris, France
- Iosifescu, C., Iosifescu, I., Hurni, L. (2013) "Customizable Map Content and 3D Visualization in the Online Platform GeoVITE". Proceedings of the 26th International Cartographic Conference, Dresden, Germany
- Iosifescu, I., Hurni, L. (2013) Integrated Cartographic and Programmatic Access to Spatial Data using Object-Oriented Databases in an Online Web GIS. Proceedings of the 26th International Cartographic Conference, Dresden, Germany.
- Iosifescu, I., Iosifescu, C., Panchaud, N. H., Eichenberger, R., Sieber, R., Hurni L (2013) Advances in Web Service Driven Cartography. Proceedings of the 26th International Cartographic Conference, Dresden, Germany.
- Iosifescu, I., Vescoukis, V., Iosifescu, C., Panchaud, N. H., Hurni, L. (2014) Open Source Technologies for Integrating Environmental Data in an Online GIS Platform for Interdisciplinary Environmental Research. Proceedings of Autocarto 2014, Pittsburgh, Pennsylvania, United States.



## **Discussing the role of university in spatial data infrastructure construction: issues and challenges for UERJ-V-SDI**

*José Augusto Sapienza Ramos, Carlos Eduardo Gonçalves Ferreira*  
*{sapienza, carlos.ferreira}@labgis.uerj.br*

*LabGIS System - Research Center of Geotechnology of Rio de Janeiro State University (UERJ)*  
*Multidisciplinary Institute of Human Formation with Technology of UERJ*

Keywords: academic community, data producer, final-work geospatial dataset, value-added data, volunteer, open data.

### **1. Introduction**

In the computer era, an unprecedented amount of data is being produced. This data is then organized in catalogs or repositories called “data infrastructures” (Kitchin, 2014), where information should be easily retrieved in order to be useful. It is noticeable that the performance of any new information paradigm has an impact over science and society, increasing and spreading the human knowledge more rapidly.

Geographic data can be defined as the way in which the knowledge of the Earth's surface is represented (Goodchild, 2011). Advancements in technology as well as the new paradigm in geographical information science and systems allow the production, visualization, analysis and sharing of large amounts of geographic data suitable to different kinds of audiences. To achieve that end proper scientific and technological debates are required.

With that in mind concepts such as “data infrastructure”, “open data”, “big data”, “open governmental data” are being used to describe some of the new aspects of today’s society – see Kitchin (2014) for a survey. On the other hand, data production and its usage have been decentralized from governments and companies to include the population in general, giving birth to terms such as “neocartographers” (Liu and Palen, 2010) and “Volunteered Geographic Information” (VGI) (Budhathoki and Nedovic-Budic, 2008).

Data infrastructure (DI) “*is the institutional, physical and digital means for storing, sharing and consuming data across networked technologies*” (Kitchin, 2014, p. 32). Following O’Carroll et al. (2013) and Kitchin (2014), the DIs can be divided into: (a) *data holdings*, as informal collections in data files; (b) *data archives*, as formal collections of data that are structured, curated and documented; (c) *catalogues, directories and portals*, as centralized resources linking different data holdings and archives; (d) *data repositories* which aim to ensure that each archive or holding meets a specific set of audited requirements in order to validate data integrity and ensure trust, where repositories can be single-site or multi-site; and (e) *cyber-infrastructures*, as a suite of dedicated and integrated hardware and network technologies, including interoperable software and middleware services, shared services, analysis tools, data visualization and shared policies (Cyberinfrastructure Council, 2007).

By its turn, the spatial data infrastructure (SDI) can be understood as a DI that incorporates geographic data and its specific technologies and standards, such as spatial databases, geovisualization, geoservices or geospatial metadata. At the beginning, the SDI discussion grew with a focus on implementation of cyber-infrastructures in federal

governments (National SDI or NSDI) and in larger companies. However, today, the SDI initiative spreads into other kinds of sectors, and the epistemological debates are changing to introduce, for instance, its social impacts and decentralized data production such as VGI – see Dessers (2012) for a survey of concepts and history.

## **2. Placing this case**

The University environment is an important data producer, which includes geographic data based on scientific approach. For instance, the Sirius Network – a network of libraries from Rio de Janeiro State University (UERJ) – had, in 2013, a collection of maps with 3,896 titles and 3,553 final work documents in Engineering and Natural Sciences (DATAUERJ, 2014), covering both analog and digital spatial datasets produced and documented in several formats. The majority of higher-level education institutions (if not all) create important data repositories to store academic documents such as thesis and papers generated by their students. However, few institutions maintain similar repositories to gather another type of academic production: the spatial data created by the very same students, which results in disperse data holdings and archives (with high data loss) that make these databases unfit for further use in academic research, by governments, by businesses or by any other potentially interested sectors.

In this scenario, the public universities have a considerable spatial data production in Brazil, as they centralize an important part of the academic production in fields such as geosciences and engineering, among others. According to *Plano de Ação da INDE* (BRASIL, 2010), the Brazilian NSDI, namely Infraestrutura Nacional de Dados Espaciais (INDE) in Portuguese, the academic sector is regarded as an agent “*responsible for fomenting and developing education, qualification, training and research in SDI*” (p. 61), i.e., limited to the role of non-producer of data. The same document defines that some producers of value-added data must be additionally identified among all SDI agents, where “*it is expected that the weight and the agent’s participation [as data producers] will increase considerably with the INDE evolution*” (p. 66). Today, it is noticeable that the academic sector is not perceived as a data producer by the INDE or in itself.

The “Open Government Working Group” (2015) stipulates the principles of Open Government Data, highlighting that it must be complete, primary, timely, accessible, machine-processable and license-free; the access must be non-discriminatory, and the compliance reviewable. The geographic data produced in a public university as UERJ should also follow these principles whenever possible. This way, data shall be open, so that it can be freely reused and redistributed among citizens (Pollock, 2006).

This work discusses and proposes the UERJ-V-SDI (UERJ Volunteer SDI) dedicated to the student community of UERJ, which is adherent to INDE standards, with its database searchable through the INDE portal. At the beginning, the academic community is sharing its spatial data voluntarily. However, it is expected that the university will develop further the discussion of an efficient data infrastructure for its production of spatial data. The proposal presented here is not definitive, but rather an initial discussion on this matter. An SDI is dynamic, and should evolve as the discussion evolves, but this first step is fundamental and urgent.

## **3. UERJ-V-SDI Components**

Aiming towards the goals described above, a SDI can be addressed in different ways, and it is common to find in the literature its division into different components and objectives (Rajabifard, 2008; Dessers, 2012). This paper follows the major division proposed in Warnest (2005), which is adherent to INDE components. The Figure 1 shows the division used in this work, i.e., to an academic context.

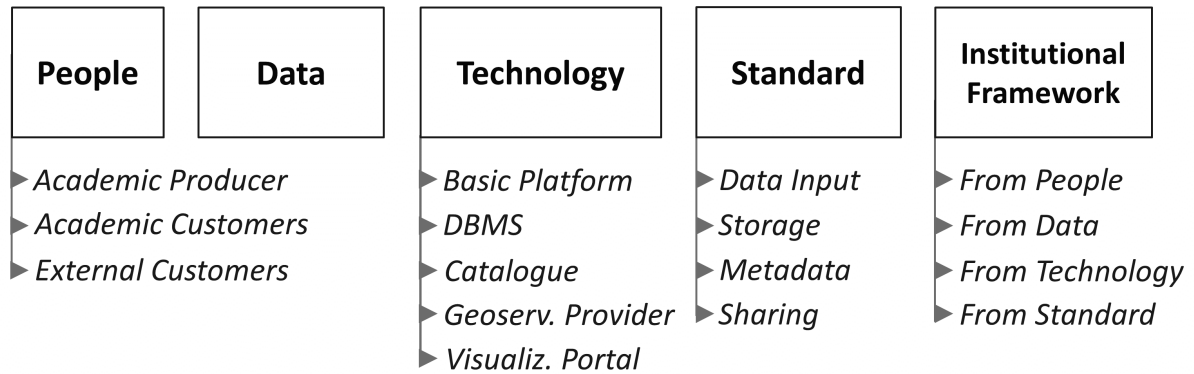


Figure 1 – UERJ-V-SDI components.

You will find below brief explanations about each component, considering the space constraints of this paper. The *People component* is divided into producers and customers. The customers can be diverse, that is, sets of users with different needs. This SDI may focus on academic and governmental users. The data producers are any university member that generates geographic data upon activities of teaching, research or extension. This way, it is necessary to identify potential data holdings and archives inside of the university's departments to seek an incremental integration with Sirius Network's databases. On the other hand, students, professors, directors and other members must be aware of the benefits and values behind their volunteer participation on SDI, since volunteers will produce new databases with the necessary procedures and documentation as metadata, for instance.

The *Data component* can be very diverse, not following a unique and rigid conceptual model, considering that countless themes and methodologies can be used. Furthermore, standards for character set, reference system, field names, data format, among others, need to be defined, and quality control encouraged.

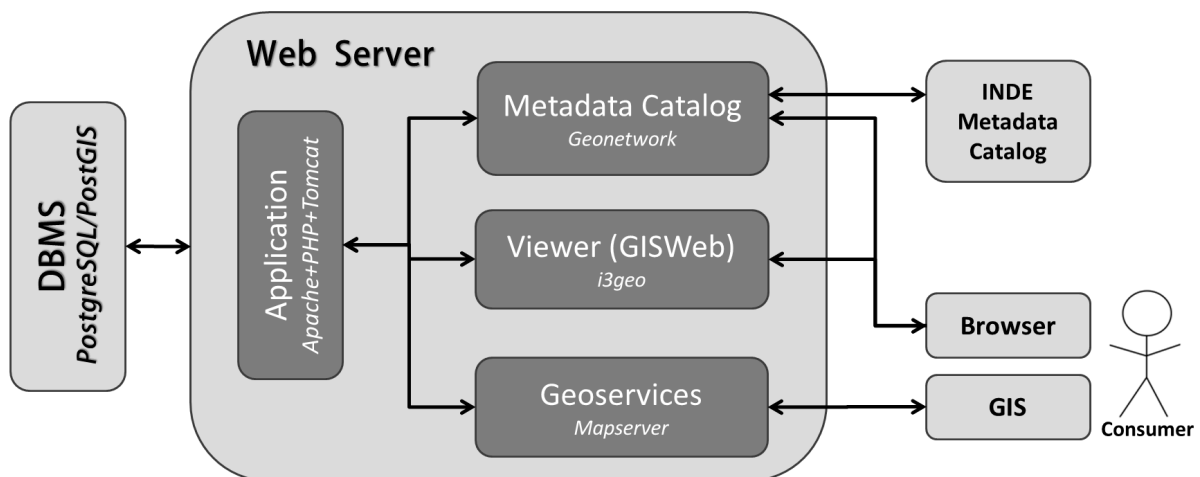


Figure 2 – The technological platform and its main communications.

The *Technology component* defines the network solutions that implement the digital means for storing, sharing and consuming data – see Figure 2. Note that this platform is fully open source, and the user can access the data using browsers or a GIS which implements the OGC standards. The user can discover the database through a metadata catalog, thus visualizing and analyzing the relevant data using the GISWeb or geoservices by interoperable means. In addition, the CSW (Catalogue Service for the Web) protocol is applied to share the metadata database with the INDE metadata catalog, thus enabling searches for the INDE catalog to return results in UERJ-V-SDI database. It is noteworthy that Figure 2 omits some communications between adopted solutions, showing the main links.

It is necessary to adopt standards for each step of the data lineage of the SDI, namely: input, storage, catalogue, access and sharing. Thus, the *Standard component* defines directives which tell how some activities of the SDI must be performed in a coordinated way. So far, the main standards adopted are: OGC standards as WMS, WFS, WCS, KML, and CSW to Web services, the core metadata of ISO 19115:2003, and format standards to store the data using the same reference system and character set, among others.

At last, the *Institutional component* can be defined as the driver force which maintains and guides the SDI and its components towards the current goal. In other words, the targets and policies must be frequently updated to reflect the current needs of the SDI. As of today, this component recognizes challenges to: (a) identify spatial data pools in the university; (b) identify and recognize data producers as end-of-course students or key professors who could volunteer; (c) promote the debate about the organization of spatial data production inside the university environment and the SDI values (as this paper); (d) put together an initial and valuable database; and (e) promote the university as an important data producer.

#### **4. Conclusion**

Data infrastructures are technological platforms to ease data retrieval, whereas SDI initiatives show the way governments and others groups enforce cyber-infrastructures which create interoperable environments to access and share the geographic information. When the university keeps spatial data produced by its members in data holdings or data archives, part of its intellectual production becomes unavailable or lost for the internal or external community. Therefore, the public university and several government agencies in Brazil must recognize this academic database as more relevant for the INDE objectives – it is necessary to promote a larger debate.

When the university shares its spatial data more efficiently as open data, it spreads its production deeper in society, fulfilling its social role. However, SDI proposals in Brazilian universities might face the same problems of several SDIs around the world, which are not of a technological nature, but rather related to the awareness of institutional parties. Because of that, this paper is sometimes more about context than technology.

This proposal is at its beginning, and needs to be discussed further, where the volunteer approach is a strategy to disseminate the SDI idea among the university environment and in external communities.

#### **References**

- BRASIL – Ministério do Planejamento, Orçamento e Gestão 2010, *Plano de Ação para implantação da INDE: Infraestrutura nacional de dados espaciais*, Comissão Nacional de Cartografia (CONCAR), Rio de Janeiro.
- BUDHATHOKI, NR, & NEDOVIC-BUDIC Z 2008, 'Reconceptualizing the role of the user of spatial data infrastructure', *GeoJournal*, v. 72, n. 3-4, pp. 149-160.
- CYBERINFRASTRUCTURE COUNCIL 2007, *Cyberinfrastructure vision for 21st century discovery*, National Science Foundation. Available from < [www.nsf.gov/pubs/2007/nsf0728](http://www.nsf.gov/pubs/2007/nsf0728) > [16 March 2015].
- DATAUERJ, *Anuário Estatístico 2014*. UERJ – Núcleo de Informação e Estudos de Conjuntura – UERJ/NIESC/VR, Rio de Janeiro. Available from <<http://www2.datauerj.uerj.br/>> [16 March 2015].
- DESSERS, E 2012. *Spatial Data Infrastructures at work. A comparative case study on the spatial enablement of public sector processes*. PhD Thesis, Leuven University.
- GOODCHILD, MF 2011, 'Challenges in geographical information science, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Science*, The Royal Society A 467(2133), pp.2431–2443.
- KITCHIN, R 2014, *The data revolution: Big data, open data, data infrastructures and their consequences*, Sage, London.
- LAURIAULT, TP, CRAIG, BL, TAYLOR, DR, & PULSIFER, PL 2008. *Today's data are part of tomorrow's research: Archival issues in the sciences*, *Archivaria*, n.64.
- LIU, SB & PALEN, L 2010, 'The new cartographers: Crisis map mashups and the emergence of neogeographic practice', *Cartography and Geographic Information Science*, 37(1), pp. 69-90.
- O'CARROLL, A, COLLINS, S, GALLAGHER, D, TANG, J & WEBB, S 2013, *Caring for Digital Content: Mapping International Approaches*, NUI Maynooth/Trinity College Dublin/Royal Irish Academy and Digital Repository of Ireland, Dublin.
- OPEN GOVERNMENT WORKING GROUP 2015, *Principles of open Government data*. Available from: <[www.openGovData.org](http://www.openGovData.org)> [16 March 2015].
- POLLOCK, R 2006, *The value of public domain*, IPPR. Available from <[http://rufuspollock.org/papers/value\\_of\\_public\\_domain.ippr.pdf](http://rufuspollock.org/papers/value_of_public_domain.ippr.pdf)> [16 March 2015]
- RAJABIFARD, A 2008, A Spatial Data Infrastructure for a Spatially Enabled Government and Society in *A Multi-View Framework to Assess Spatial Data Infrastructures*, eds CROMPVOETS, J, RAJABIFARD, A, VAN LOENEN, B & FERNÁNDEZ, TD, Wageningen University, RGI, pp.
- WARNEST, M 2005, *A collaboration model for national spatial data infrastructure in federated countries*. Ph.D. Thesis, University of Melbourne.

## Using OS Software and Open Standards for Implementing a European SDI via Service Cascade

Lassi Lehto  
[lassi.lehto@nls.fi](mailto:lassi.lehto@nls.fi)

*Finnish Geospatial Research Institute, Geodeetinrinne 2, FI-02430 Masala, Finland*

Keywords: service cascade, content integration, open source, spatial data infrastructure, web feature service

### 1. Introduction

The research described in this paper has been conducted in the context of a major EU project, called European Location Framework (ELF), initiated by EuroGeographics (EG), the co-operation organization of the European National Mapping and Cadastral Agencies (NMCAs) (EuroGeographics 2015). The ELF project aims at developing European-wide INSPIRE-compliant services from the geodata resources maintained by the EG's member organizations (European Location Framework 2015).

The ELF project started in March 2013 and will run for three years. The project has 30 participant organizations and 13 of them represent EU/EFTA member states as official NMCAs. Thus, the project has quite extensive spatial coverage extending from Finland to Spain and from Great Britain to Poland (see Figure 1).

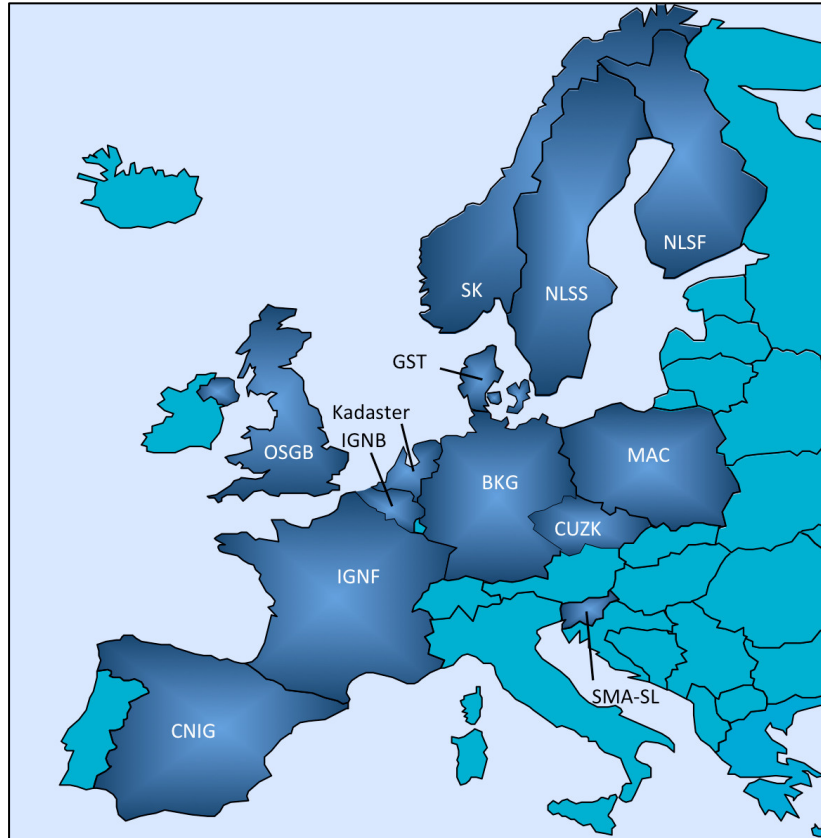


Figure 1. The countries participating in the European Location Framework project.

## **2. Aim of the research**

The ELF project includes a work package specifically dedicated for data provision and service development. In this work package there is a subtask responsible for investigating the issues related to service cascade. The approach presented in this paper focuses specifically on the provision of European level Download Services (Web Feature Service, WFS) based on data services maintained by local NMCAs on country level (Vretanos 2011). The goal is to support the end user in accessing data content directly from national services. Thus, the cascading approach aims at carrying out real-time aggregation of content from a set of distributed national data sources (Lehto et al 2014).

One of the new challenges encountered when accessing national services from European-level applications is the need to introduce spatial integration capabilities to the service cascade approach. At the moment only thematic integration is supported in the existing cascade mechanisms of the Open Geospatial Consortium's Web Map Service standard. In this setup every single map layer is to be served by one and only one back-end service. When implementing cascaded integration over a set of national services, one has to resolve the problem of spatial query distribution and cross-border fusion of content. These are the biggest challenges in the further service cascade developments of the ELF project, too.

The goal of the cascaded services work in the ELF project has been to facilitate access to national download services providing master level datasets. The work has focused on direct access download services that conform to OGC's Web Feature Service (WFS) service specification. The aggregation layer developed by the cascaded services subtask of the project on top of the national services can be seen as a Cascading WFS implementation. Service cascade is not officially specified as part of the WFS standard. For this reason ELF Cascading WFS must be taken as an experimental test service.

## **3. Implementation**

A content request coming from a client application is first analyzed by the Cascading WFS to determine, which national level services must be included into the process. Then the request is forwarded to the involved national services. The resulting datasets are merged together and finally returned back to the calling application. The analysis on the service inclusion is based on the bounding box of the query and on the requested feature types. The bounding box is overlaid on top of a dataset of national borders to determine, which countries the query overlap. The actual service inside the country is then selected depending on the requested feature type. ELF Cascading WFS does not support the full-fledged WFS service interface, but has certain significant restrictions, like the fact that the query must always contain a bounding box. The service has been developed as a Java Servlet-based web application. All the supporting software modules, like the DBMS for the country boundaries, the applied XML-processing libraries and the used software development tools are available as OS products. Most of the country level services are implemented using OS software implementations of the WFS standard.

The Cascading WFS component is running on the Amazon Cloud (EC2 small instance). The main tasks of the component is to decide to which national services the incoming WFS GetFeature –query must be forwarded, depending on the location of the query bounding box and the requested feature types. The bounding box of the query is overlaid with a national boundaries dataset stored in a separate PostgreSQL/PostGIS database. As a result, the



involved countries are listed. Then the requested feature types are checked to determine, which country level services have to be queried.

If only one backend service is involved in the Cascading query, the process is straightforward. The single service is accessed and the resulting dataset is returned to the calling application without further processing. If two or more services are involved, the Cascading WFS dispatches several parallel query threads to access the national services. The returning datasets are processed in the order they become ready. The datasets are merged into a single response message using a SAX-based XML-processing model.

The Cascading WFS has to create the root element (FeatureCollection) of the resulting dataset to include all required XML namespace declarations. In addition, the Cascading WFS fills in the `boundedBy` -element using the query's bounding box as the limits of the spatial extent. Then the contents of the individual background responses are written out as they become ready, taking away the FeatureCollection and `boundedBy` -elements. Finally the FeatureCollection element is closed by the Cascading WFS.

#### **4. Limitations**

The ELF Cascading WFS access interface has certain significant restrictions. Most importantly, it requires that a query must always have a bounding box defined. This can be presented as `BBOX` -parameter or as `<fes:BBOX>` -element inside either the `FILTER` -parameter of a GET-query or inside the `<Filter>` -element of a POST query. If there are several `<Query>` -elements in a single POST GetFeature request, they have to contain the same `<fes:BBOX>` (only the first one is taken into account). The ELF default Coordinate Reference System (CRS) is Web Mercator (EPSG:3857). This CRS can also be requested using the EPSG code 900913. Whether a given CRS is supported by the backend national level service varies from service to service.

Currently the GetCapabilities response message of the ELF Cascading WFS is configured manually. This could be based on dynamic GetCapabilities -querying of the backend national services. In any case there are certain limitations. All the backend services should support the same set of feature types, otherwise certain queries would result in an exception being returned. GetCapabilities response defines the spatial extent of each feature type by one single rectangle. As the spatial extents of the backend services are often distinct, the area definition in the GetFeature response becomes misleading. Because the other parameters, like the list of supported CRSs, must be composed as the greatest common denominator, some possibilities of the services remain unusable.

Feature identifiers must be globally unique for the Cascading WFS to work. In the case of the INSPIRE feature identifiers this is handled properly by using well-defined namespace mechanisms. Some of the ELF national services already apply an URI-based namespace label, which automatically ascertains global uniqueness. However, this becomes crucial in the case of the XML ID-typed `gml:id` -attributes used in feature and geometry elements. These values must also be prepended with the namespace label to keep them unique in the content aggregation phase, or some other mechanisms have to be employed.

#### **5. Conclusion**

The real-time content aggregation that is tested in the ELF project is based on the principle of Cascading WFS. ELF Cascading WFS does not support the full-fledged WFS service

interface, but has certain significant restrictions, like the fact that the query must always contain a bounding box.

Altogether 32 national download services have been included into the ELF Cascading WFS (Poland 5, Sweden 4, Norway 3, Czech Republic 10, Finland 4, France 1, Spain 5). These services provide access to 79 different feature types.

### **Acknowledgements**

The ELF project has been funded by the European Commission, Grant Agreement No 325140.

### **References**

EuroGeographics 2015, Home Page. <http://www.eurogeographics.org> Accessed 9 June 2015

European Location Framework 2015, Project Home Page. <http://www.elfproject.eu> Accessed 9 June 2015

Lehto, L., Latvala, P. and Kähkönen J. 2014, Service Cascade as a Means for Pan-European Access to National Geodata Content CASE: European Location Framework. The Sixth International Conference on Advanced Geographic Information Systems, Applications and Services, "GEOProcessing 2014", Mar 23 – 27, 2014, Barcelona, Spain, CD-ROM.

Vretanos P. A. [ed.] 2011, OpenGIS Web Feature Service 2.0 Interface Standard.  
[http://portal.opengeospatial.org/files/?artifact\\_id=39967](http://portal.opengeospatial.org/files/?artifact_id=39967) Accessed 9 June 2015

## **Interoperable provision of geodata and services according to the INSPIRE Directive in German geoportal sites**

*Franz-Josef Behr*  
[franz-josef.behr@hft-stuttgart.de](mailto:franz-josef.behr@hft-stuttgart.de)

*Department of Geomatics, Computer Science and Mathematics, University of Applied Sciences  
Stuttgart, Schellingstraße 24, 70174 Stuttgart, Germany*

**Keywords:** INSPIRE, open geodata, geoportals Germany, usability, spatial data infrastructure

The European INSPIRE directive introduces a legal framework and general rules to establish an infrastructure for spatial information in Europe primarily with focus on policies and having an impact on the environment. In this context 34 spatial data themes are listed comprising general geospatial base data, i.e. cadastral parcels, administrative units, and more ecological data themes, i.e. geology, species distribution, energy resources, etc.<sup>1</sup> The data are supposed to be provided through spatial data infrastructures (SDI) in the 28 Member States of the European Union and in the corresponding politically mandates organisations, i.e. the 16 federal state surveying authorities in Germany.

Besides this data catalogue INSPIRE specifies a set of mandatory interoperable services, i.e. Download Services, View Services, Discovery Services, Transformation Services, Schema Transformation Network Services, and Coordinate Transformation Services which are partially similar to OGC's services but partially go beyond (Bray & Ramage 2012). As well metadata have to be provided following a schema which extends the specifications of ISO 19115 (European Commission Joint Research Centre 2010).

With regard to these services requirements regarding the quality of service, e.g., service performance, are given. i. e. specific spatial reference systems have to be supported, and the services must offer multi-lingual support. The member states have to monitor and to report the availability and usage of the services. There is also a strong link to the Open Data policies in the member states.

In the contribution the relevant policies and architectural and technical requirements are explained.

The portal sites of the German federal state surveying authorities are analysed related to information about INSPIRE, license issues, and the Open Data movement. The assessment includes as well the availability of OGC services resp. INSPIRE conformant services, the open access to data according INSPIRE's spatial data themes, the accessibility, the level of interoperability and the quality of services. The portal sites of the different surveying authorities show significant levels of maturity (Ernst 2014). Based on this evaluation, recommendations for improvements will be given.

### **References**

- Bray, A. & Ramage S. (2012): OGC Market Report Open Standards and INSPIRE. [https://portal.opengeospatial.org/files/?artifact\\_id=48233](https://portal.opengeospatial.org/files/?artifact_id=48233) [2015-03-14]
- Weichand J. (2014): Moderne Geodatenbereitstellung im Kontext der europäischen Geodateninfrastruktur INSPIRE. Vortrag, FOSSGIS 2014, Berlin, <https://www.youtube.com/watch?v=Qj4-8OQv7Ng> [2014-11-27]
- Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE) - <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:DE:PDF>
- 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). <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0002&from=EN> [2015-03-15]

---

<sup>1</sup> [http://inspire.ec.europa.eu/documents/Metadata/INSPIRE\\_MD\\_IR\\_and\\_ISO\\_v1\\_2\\_20100616.pdf](http://inspire.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf)

- Verordnung (EG) Nr. 976/2009 der Kommission vom 19. Oktober 2009 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich der Netzdienste (es existiert noch ein Amendment) - <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:274:0009:0018:DE:PDF>
- Verordnung (EG) Nr. 1205/2008 der Kommission vom 3. Dezember 2008 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich Metadaten Text von Bedeutung für den EWR- <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:DE:PDF>
- Verordnung (EG) Nr. 1089/2010 der Kommission vom 23. November 2010 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich der Interoperabilität von Geodatenätzen und -diensten (es existiert noch ein Amendment)- <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:DE:PDF>
- European Commission Joint Research Centre (2010): INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Version 1.2) . [http://inspire.ec.europa.eu/documents/Metadata/INSPIRE\\_MD\\_IR\\_and\\_ISO\\_v1\\_2\\_20100616.pdf](http://inspire.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf) [2015-03-15]
- IOC Task Force for Network Services (2011): Technical Guidance for the implementation of INSPIRE View Services (07.11.2011). [http://inspire.ec.europa.eu/documents/Network\\_Services/TechnicalGuidance\\_ViewServices\\_v3.1.pdf](http://inspire.ec.europa.eu/documents/Network_Services/TechnicalGuidance_ViewServices_v3.1.pdf) [2015-03-15]
- Ernst, J. (2014): OpenGeoData – bisher nur ein (leeres) Versprechen. 13. Fachforum: „Verfügbare Geodaten – was bringt’s?“, [http://vorarlberg.at/pdf/4\\_ernst\\_geokomm.pdf](http://vorarlberg.at/pdf/4_ernst_geokomm.pdf) [2015-07-03]

## Geoportal MTOP

*Hebenor Bermúdez, Ana Fernández, Laura Marroni, Melissa Robert, Federico Godán*

[hebenor@gmail.com](mailto:hebenor@gmail.com), [anafernandez1@gmail.com](mailto:anafernandez1@gmail.com), [laura.marroni78@gmail.com](mailto:laura.marroni78@gmail.com),  
[melissarobsar@gmail.com](mailto:melissarobsar@gmail.com), [federicogodan@gmail.com](mailto:federicogodan@gmail.com)

*Fundación Julio Ricaldoni, Ministerio de Transporte y Obras Públicas*

Keywords: geoportal, SDI, Metadata, standards

### 1. The project

The Ministry of Transport and Public Works, MTOP (Ministerio de Transporte y Obras Públicas) of República Oriental del Uruguay, has been developing the core of geographical information which will be the provider of the MTOP geospatial information to the Uruguayan Spatial Data Infrastructure (IDEuy).

This project has as its first objectives to organize, structure and publish all the information inside MTOP who responds to a spatial location in order to bring this information to the entire institution, to all the dependencies, as well as to the entire public.

#### 1.1. Project development

IDEuy has been working since 2008. It had some rebuildings and different directors, but the structure, organization and politics of standards have been the same along the years. IDEuy suggests the use of free software and open data through a law enacted in 2008 (Ley 18381 "Derecho de acceso a la información pública"). IDEuy defines a structure based on cores which provides geographic data from different state institutions.

IDEuy diagram is displayed at Figure 1.

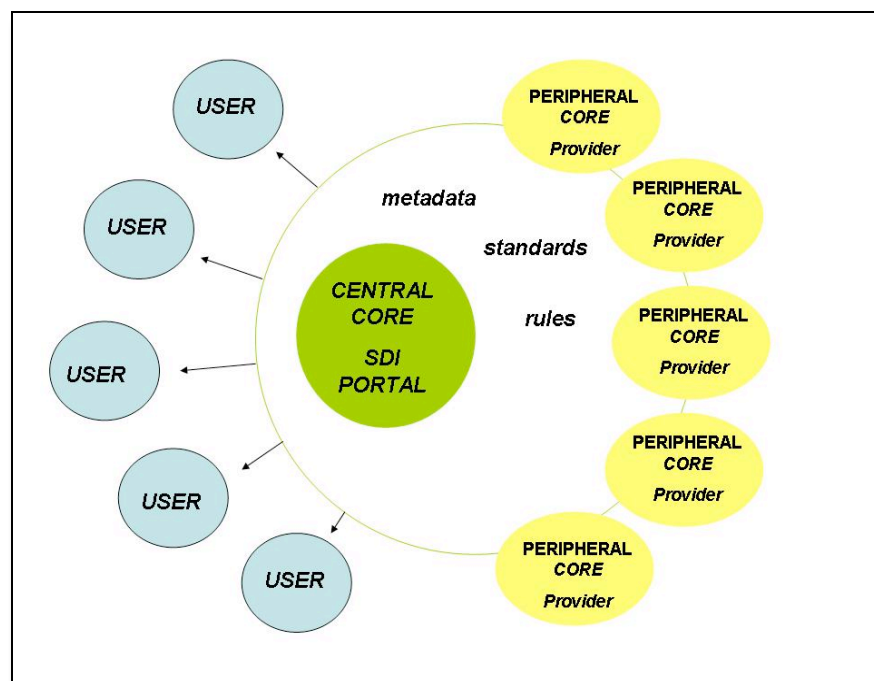


Figure 1. IDEuy diagram

In 2011, School of Engineering through an agreement signed with MTOP, studied the needs in the institution for implement a geographic information website to provide data to IDEuy, MTOP core.

In 2012 MTOP formed a team of professionals and assistants related to geomatics, such as Surveyors Engineers and Computer Engineers, because there was not enough trained staff to focus exclusively in this project.

After that, it has begun the work of writing the guidelines and schedule meetings with the most important actors inside the MTOP. Thanks to these meetings the team started to build the structure and determine the principal components of the information to generate the Feature Catalog for MTOP spatial information, according ISO standards (ISO/TC 211: ISO 19110, ISO 19126). Then, the team begins to develop a technological infrastructure to contain the different elements for the geoportal application, which provides services to access into the data stored in the spatial database. The geoportal has the ability to view, search and access to geospatial data or metadata. This applications are: Website (geoportal.mtop.gub.uy.), map server, metadata server and two maps viewers. Also can access to a lot of geoservices offered from the map server.

## **1.2. Standards and technologies**

Since the beginning of the project, the MTOP uses free software and technologies for development.

To fulfill the functions of the geoportal, a database management system PostgreSQL with PostGIS extension is installed, where the spatial data are stored. The maps are available through GeoServer map server. As map viewer we used Mapbender framework which supports all of the Open Geospatial Consortium (OGC) standards needed. It is a Web-GIS client built with JavaScript, PHP, which offers a configurable user interface not dependent on any particular server maps. Its orientation is a geoportal client OGC services. Currently supports WMS, WFS-T and WMC. It includes a fairly comprehensive user support groups and services OGC (OWS). MapBender provides access to the components of a Spatial Data Infrastructure (SDI) to adhere to international interoperability standards for geographic information services via the Web. It allows to create a lot of applications or widgets without the need or, in some cases, very low programming tasks.

To create, edit and publish geographic metadata it uses GeoNetwork platform, suggested by IDEUy. Also, the standards suggested by the IDEuy and the Uruguayan profile of Metadata, according ISO 19115, ISO 19139.

## **2. The data**

In the MTOP are 6 National Bureaus. The total staff at 2013 was 3200 according to the annual report of that year. In turn, each Bureau have different assignments which shows a great heterogeneity in MTOP functions.

The collection of information, its standardization and the definition of responsibilities are a long and laborious process. Knowing and understanding the MTOP's difficulties and needs about generation of spatial data is essential to being able to establish the starting point of the project. So, a series of activities to achieve this goal were performed.

The team begins the project with the Bureau of Surveying (DNTopo) which manage the project development. One of the major challenges is to change the way of work with geographic information. While the different Bureau are used to work with reference to the territory, GIS concepts, spatial database, geographic information layer, etc. are often not easily handled. This provokes many issues, so this project developing has to agree in some aspects to the effect that all users can meet your needs trough the geoportal and the spatial database. This is undoubtedly the major challenge to the project.

DNTopo has also undertaken to meet their specific purposes, doing several high resolution photogrammetric flights, generating altimetric information, basic cartography and orthophotos of some areas of interest.

The MTOP principal function resides in the administration, maintenance and management of the national routes. So, one of the most important actors is Bureau of Roads (DNV). Between 2002 and 2008, DNTopo carried out a survey of all national roads. This allows DNV to use this surveyed geometric data and put together all databases, either spatial as others. This background information is useful to other authorities and staffs working with the roads. For example, the Logistics and Planning Bureau (DINAPLO) has annual surveys of the cargo on the routes, so having organized information is a great contribution.

MTOP Geoportal allows access to the geographic information of Bureaus. This information is about aerial transport (heliport, airport), terrestrial transport (routes), rail transport (train track, train stations), fluvial transport (ports), logistic information, etc. The web site is published in 2014 and we are working on publicizing the Geoportal.

### 3. Geoportal MTOP components

As described, a geoportal must have some principal components. The website [geoportal.mtop.gub.uy](http://geoportal.mtop.gub.uy) has a link for each of these components, displayed at Figure 2.



Figure 2. Geoportal web homepage

In the Figure 2 can be seen the homepage with accesses to the applications. The first,



“Visualizador de Mapas” is a link to a map viewer which contains all the spatial information compiled in the first stage of this development.

The principal map viewer is displayed in Figure 3.



Figure 3. Principal map viewer

Another map viewer is “Planos de Balnearios” which shows the information of cadastral maps of seaside resorts on the coast of Canelones, Maldonado and Rocha State departments. In this application you can search the original survey maps by name of seaside resort. This search shows all published maps in high resolution JPG format, facilitating the information access. At Figure 4 is a view of “Planos de Balnearios” Map viewer, and Figure 5 displayed the result of a research in the application.



Figure 4. “Planos de Balnearios” Map viewer

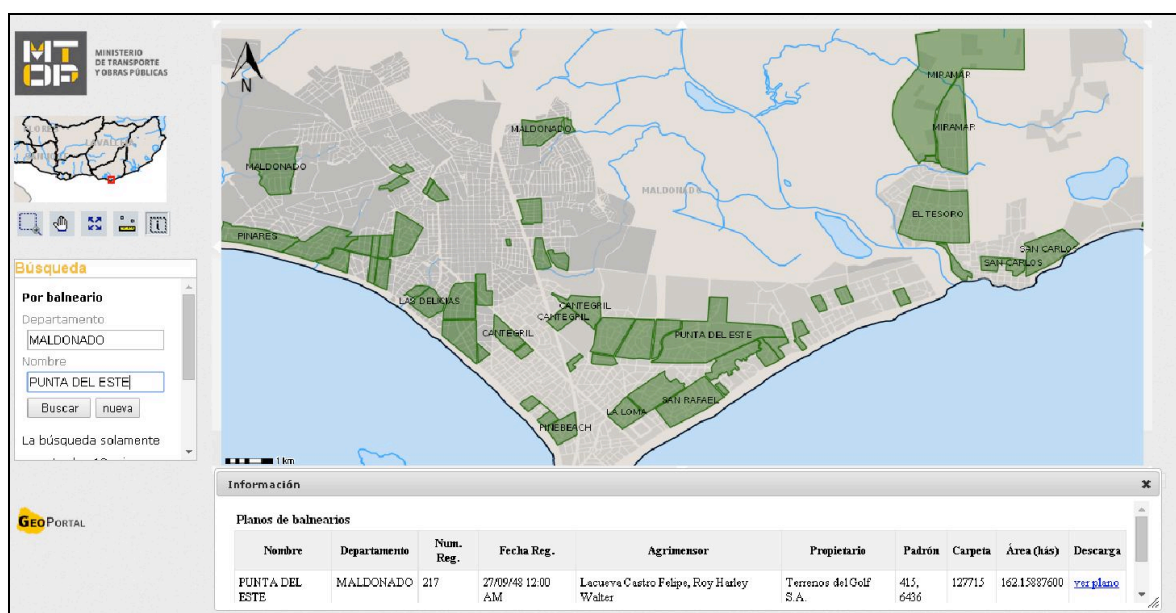


Figure 5. result of a research maps by name of seaside resort

Following this, there is an access to the Metadata Server where the information can be searched, accessed and be distributed thanks to the interoperability of the data provided by the OGC standards utilization. To this service use Geonetwork open source, which is based on the principles of Free and Open Source Software (FOSS) and International and Open Standards for services and protocols (ISO/TC211 and OGC). Its supports a lot of schemas, including the selected for the project, ISO19115/ISO19119/ISO19110 following ISO19139.

Finally, is possible to access some geoservices: WMS (Web Map Services) and WFS (Web Feature Services).

“Geoservicios Web” is displayed in Figure 6.

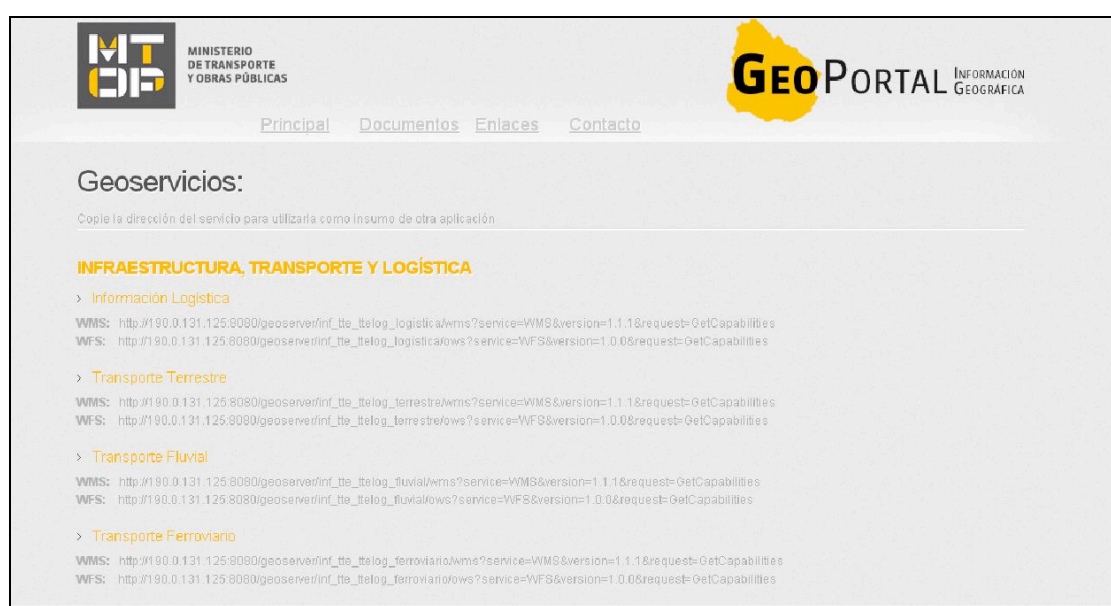


Figure 6. Geoservices page

#### 4. Conclusion and future work

Is a hard work introduce this kind of knowledge in spatial information into an organization which has a few workers with experience in GIS and not many in Spatial Data Infrastructure. But this project has been well received inside the organization, so we hope this work could continue growing a lot more with a political support.

In this way, the team is working on value added applications, like the incorporation of all the cadastral surveys into a map viewer which allows to search and view the information in JPG format, like the map viewer for seaside resorts.

Another application under development is a geographic searcher for passenger transport in interdepartmental services. It will be using the open data generated by MTOP in Bureau of Transport which contains bus stops, points of interest and timetables. It will be searching between two points and depending on the city, the transport line and the hour, it brings you a timetable and the bus stops or bus combinations to take you to your destination. Today, a similar application is available at the Municipality of Montevideo but only with metropolitan bus-line data.

The spatial information is constantly actualized and new information is generated. MTOP looks for participation of all of the actors involved with the geospatial information.

#### References

- IDEuy. 2012. "Sistemas de Referencias, Sistemas de proyecciones. Site: ide.uy [http://ide.uy/wps/wcm/connect/f747910040a1150ebbc7ffff29651486/Sistema\\_Referencia\\_Proyecciones.pdf?MOD=AJPERES](http://ide.uy/wps/wcm/connect/f747910040a1150ebbc7ffff29651486/Sistema_Referencia_Proyecciones.pdf?MOD=AJPERES) [Accessed: October 2014].
- IDEuy. 2012. "Lineamientos estratégicos para la información geográfica". Site: ide.uy [http://ide.uy/wps/wcm/connect/2772d80040a110a8bbbeffff29651486/lineamiento\\_Datos\\_Geograficos.pdf?MOD=AJPERES](http://ide.uy/wps/wcm/connect/2772d80040a110a8bbbeffff29651486/lineamiento_Datos_Geograficos.pdf?MOD=AJPERES) [Accessed: October 2014].
- IDEuy. 2012. "Información geográfica - Perfil de Metadatos". Site: ide.uy [http://ide.uy/wps/wcm/connect/5a82d00040a0e971bb3dffff29651486/Inf\\_Geografica\\_Perfil\\_de\\_Metadatos.pdf?MOD=AJPERES](http://ide.uy/wps/wcm/connect/5a82d00040a0e971bb3dffff29651486/Inf_Geografica_Perfil_de_Metadatos.pdf?MOD=AJPERES) [Accessed: October 2014].
- Open Geospatial Consortium Inc.2006. "Web Map Server Implementation Specification" <http://www.opengeospatial.org/> . [http://portal.opengeospatial.org/files/?artifact\\_id=14416](http://portal.opengeospatial.org/files/?artifact_id=14416) . [Accessed: February 2015].
- Open Geospatial Consortium Inc. 2015. "OGC Information Technology Standards for Sustainable Development". <http://www.opengeospatial.org/> . <https://portal.opengeospatial.org/files/60920> [Accessed: February 2015].
- Open Geospatial Consortium Inc.2005. "Web Feature Service Implementation Specification" <http://www.opengeospatial.org/> . [http://portal.opengeospatial.org/files/?artifact\\_id=8339](http://portal.opengeospatial.org/files/?artifact_id=8339) [Accessed: February 2015].
- Geonetwork, <http://geonetwork-opensource.org/> [Accessed: February 2015].

## **Conceptual data modeling in geospatial information technology tool kit (gittok)**

*Morishige Ota*  
[morishige\\_ota@kk-grp.jp](mailto:morishige_ota@kk-grp.jp)

*Kokusai Kogyo Co., Ltd., and Chuo University*

Keywords: teaching assistance tool, geospatial information technology, geospatial information standards, conceptual data modeling

### **1. Introduction**

Capacity building of people who require the knowledge of geospatial information technology (GIT) is a pressing problem for the development and the improvement of a spatial data infrastructure (SDI). Meanwhile, geospatial information standards (GI-Standards) are adopted as one of the knowledge units of geospatial information science and technology (GIS&T) Body of Knowledge proposed by countries such as the USA, Europe, South Africa, and Japan. However, at least in Japan, there is little education on GIT based on GI-Standards. Often, in an introductory course on GIS, students learn geo-data processing through exercises using a general-purpose GIS. The merit of this is that students become familiar with how to use GIS. Meanwhile, currently, people are becoming familiar with geospatial data processing, mainly because the availability of software systems and hardware such as global positioning systems, unmanned aerial vehicles and mobile devices is increasing. In addition, the open data policies promoted by governments, and the use of free and open-source software are becoming widespread. These facts encourage the enhancements to ubiquitous information services, appearance of volunteered geographic information (VGI), and SDI. For example, Cooper et al. (2011) have pointed out that the increasing costs of official mapping programs coupled with the availability of high volumes of quality and up-to-date VGI have led to the integration of VGI into some SDIs.

However, improvement of geo-data sharing is impossible without engineers who have the knowledge of GIT and GI-Standards. GIT is defined as an information technology for the modeling, acquisition, management, analysis, exchange, and representation of geospatial information (Ota 2012). According to the scope of ISO/TC 211, standardization for geospatial information aims to establish a structured set of standards for information concerning objects or phenomena that are associated with a location relative to the Earth. In order to teach the underlying technology of SDI, teaching materials on GIT and GI-Standards need to be prepared. However, software tools that meet these conditions are very few. This was the reason to develop the geospatial information technology tool kit (gittok, pronounced: jee-tock, dʒi:tɒk). Gittok is a free open-source “all-in-one” software tool that assists students learning about “what is GIT” on the basis of GI-Standards (Ota 2015). Gittok and the related teaching materials are available at <<http://stinfodesign.net/gittok/>>. Source code is uploaded on <<https://github.com/stinfodesign/gittok>>.

Gittok includes a module to design the application schema (AS) according to the gittok general feature model (GFM). Data acquisition, analysis, and mapping directly use the AS in gittok. This extended abstract focuses on conceptual data modeling after a brief introduction to gittok, and finally, the results of the first college-level course on the introduction to GIT based on GI-Standards at Chuo University, Japan, will be reported.

## 2. Introduction to Gittok

Gittok consists of six modules to realize the knowledge areas in the pages displayed on the screen. These areas are modeling, acquisition, management, analysis, exchange, and representation (Figure 1).

Gittok was implemented on the basis of the simplified profiles of GI-Standards to maintain the harmonization among the six modules.

**Conceptual data modeling** is a conceptualization from the universe of discourse to the AS that describes the feature types and their relationships. Gittok enables the modeling by using a module called “Modeler.” Students can generate ASs including definitions of feature types and association types. Modeler uses a simplified version of GFM provided by ISO 19109. Modeler also refers to ISO 19103, ISO 19107, and ISO 19155 for the definition of simple data types, geometric primitives, and place identifiers including postal address and URL.

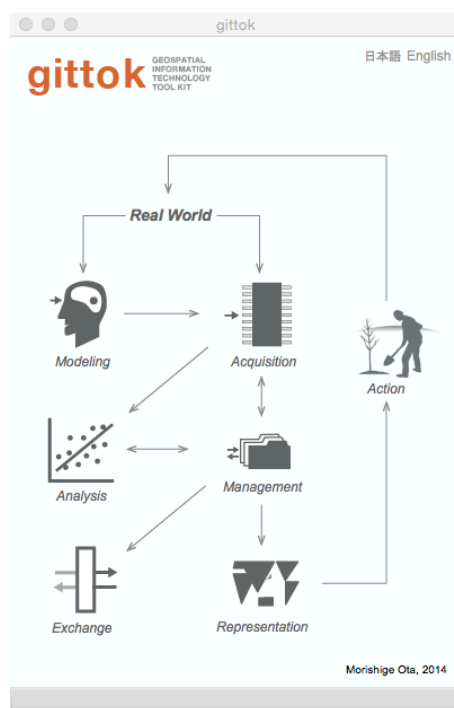


Figure 1. Title page of gittok.

**Acquisition** is a transformation from the real-world phenomena into a geospatial dataset called a “kit.” Gittok is not a general-purpose GIS, and therefore, it is impossible to use various acquisition methods. However, students can practice map digitizing to capture geometric attributes with topology and can embed images, movies, audio files, URLs, and other thematic data into feature instances through the software module called “Editor.” Gittok refers to ISO 19111 for the definition of coordinate reference systems and ISO 19118 for designing the instance model.

**Management** is a process of dealing with kit files. Gittok manages kit files through “Manager,” where the students can store and query kit files. Additional procedures such as updating, analysis, exchange, and representation can be practiced through “Manager.” Gittok refers to ISO 19115 for the schematization of the simple metadata.



**Geo-data analysis** derives the analysis result from a kit file. Gittok provides a limited number of operations applying basic computational geometry in a module called “Analyst.” However, the author recommends that teachers and students add additional analysis tools, because the last purpose of this education is the capacity building of engineers who can contribute to improve SDIs.

A **Data exchange** module called “Exchanger” encodes and decodes XML data such as an AS, a kit, metadata, and a set of parameters defining the coordinate reference system. ISO 19136 affects gittok in terms of the rules for describing XML documents.

Finally, **representation** uses a portrayal schema that specifies how to design maps. A module called “Cartographer” enables the students to design interactive maps by using a kit file and a portrayal schema. The author has referred to ISO 19117 in order to design the portrayal architecture of gittok.

### 3. Conceptual Data Modeling in Gittok

GFM establishes a basis for the definition of features and their relationships. AS is a conceptual schema for data required by one or more applications (ISO 19101-1) created in compliance with the GFM.

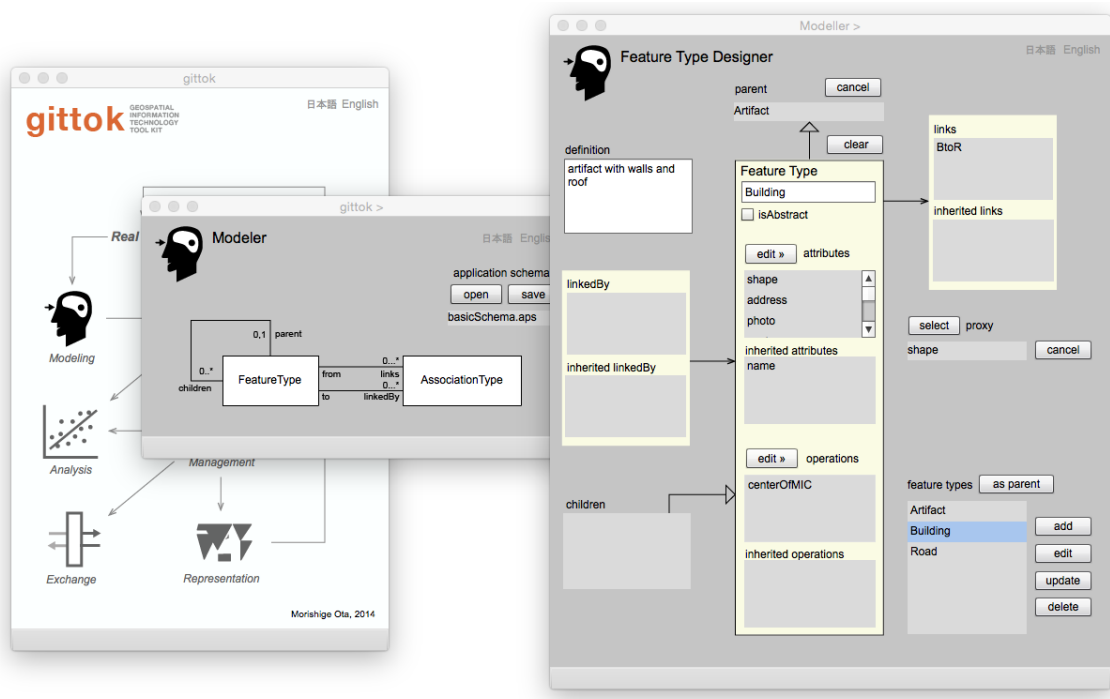


Figure 2. Modeler and a page for feature type designing.

According to ISO 19109, GFM distinguishes four aspects of defining feature types: the definitions or description used for grouping features into types, the attributes associated with each type, the relationships among the types, and the behavior of the features (Figure 2). Behavior is an operation procedure associated with a feature.

Gittok GFM is a subset of ISO GFM. First, gittok GFM does not define the GF\_Constrain type for the restriction of attribute values. Because instantiation of this class provides a high degree of freedom and investigation is required before the implementation. For example, it

can be an independent part in the data product specification. Second, an association type in gittok GFM does not inherit a feature type to make the association type simpler. Finally, the inheritance relation type is not defined in gittok GFM. Instead, inheritance is defined as a relationship between different feature types. Child features inherit the properties of parent feature by this relationship.

Gittok GFM is a practical meta-model of the AS to be used as a specification for the data acquisition. However, currently, only a few software programs directly apply the AS to data acquisition. This is attributed to the fact that an automatic translation from the AS into the data structure of RDBS is difficult because of the so-called impedance mismatch. This is one of the reasons that gittok does not use DBS; instead, it enables users to obtain geospatial data in accordance with an AS produced by “Modeler.”

#### **4. Practice of education**

The author conducted a semester-long course for 10 senior-year undergraduate students at Chuo University in Tokyo, Japan, in 2014. The program covered the design of an AS, geo-data acquisition, data management, spatial analysis, XML document encoding, and design of an interactive thematic map. Gittok was used each time after the preliminary explanation and introduction of the underlying knowledge such as that of GI-Standards. From the results of the final exercise, the author confirmed that it is possible to provide a formal introduction to GIT based on GI-Standards in the form of a college-level course to new students.

#### **5. Conclusion**

Gittok is educational support software of GIT based on GI-Standards. Gittok consists of six modules to introduce GIT; for instance, “Modeler” is a module to design the AS according to the gittok GFM, a practical profile of ISO GFM. Gittok was used in a semester-long course for undergraduate students and it was confirmed that it is possible to provide a formal introduction to GIT based on GI-Standards in the form of a college-level course.

#### **Acknowledgements**

I am grateful to the members of the GIT group (Leader: Koichi Kubota, Chuo University) established under the project supported by Grant-in-Aid for Scientific Research (A), Ministry of Education, Culture, Sports, Science and Technology (Project number: 21240075. Project leader: Yasushi Asami, The University of Tokyo), the Geospatial Information Authority of Japan, and the many people who provided their kind advice and made this study possible.

#### **References**

- Cooper, AK, Rapant, P, Hjelmager, J, Laurent, D, Iwaniak, A, Coetzee, S, Moellering, H & Düren, U 2011, ‘Extending the formal model of a spatial data infrastructure to include volunteered geographical information’, *Proceedings for International Cartographic Conference 2011*, Paris, France, July 3–8.
- Ota, M 2012, ‘Software tool development for higher education on geospatial technology’, *Proceedings for the 8th European GIS Education Seminar EUGISES*, Leuven, Belgium. September 6-9, 2012, pp.53-62. Available from: <<https://ees.kuleuven.be/eugises12/eugises12-seminar-proceedings.pdf>>.
- Ota, M & Plews, R 2015, ‘Development of a software tool as an introduction to geospatial information technology based on geospatial standards’, *Cartography and Geographic Information Science* ahead-of-print (2015), pp.1-16. DOI: 10.1080/15230406.2015.1031701.



## **A solution for processing large files in the LASer (LAS) format using the message passing interface (MPI) and parallel file systems**

Jeffrey Wendel<sup>1</sup>, Michael P. Finn<sup>1</sup>, John Kosovich<sup>2</sup>, Jeff Falgout<sup>2</sup>, Yan Liu<sup>3</sup>

<sup>1</sup>U.S. Geological Survey, Center of Excellence for Geospatial Information Science

[mfinn@usgs.gov](mailto:mfinn@usgs.gov)

<sup>2</sup>U.S. Geological Survey, Core Science Analytics, Synthesis, & Libraries

<sup>3</sup>CyberInfrastructure and Geospatial Information Laboratory, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

Keywords: parallel throughput architecture, lidar, MPI, geospatial data, message passing interface, point clouds, raster datasets

### **1. Introduction**

In recent high-performance computing (HPC) research, a persistent, restrictive case problem arises when designing scalable computational solutions for geospatial data with regard to input/output (I/O) (Behzad *et al.*, 2012; Finn *et al.*, 2015). We inspected high performance I/O for supporting parallel read and write of raster (grid) datasets, and more particularly, very large lidar point clouds that were interpolated to grid datasets. We illustrate a fresh solution for processing large lidar datasets by taking advantage of HPC power through the use of the Message Passing Interface (MPI) and the Lustre Parallel File System (Piernas *et al.*, 2007).

### **2. Study Area, data, and test design**

We acquired lidar point cloud data over areas of the Great Smoky Mountains and the Grand Canyon National Parks in the United States. We used the lasmerge application (Isenburg, 2014) to merge a subset of the Great Smoky Mountains data into one file of approximately 16 gigabytes (GB) with 572,693,051 points over a 40,000 X 20,000 meter area. Also, we used the lasmerge application to merge a subset of the Grand Canyon data into one file of approximately 120 GB with 4,294,967,295 points (maximum for LAS v.1.2) over a 25,000 X 30,000 meter area.

Producing a DEM typically involves filtering and transforming (e.g. reprojecting) LASer (LAS) file format (ASPRS, 2011) data, and using that result to produce a DEM. We named our programs `p_las2las` and `p_points2grid` and tested them using the two large test files on the Extreme Science and Engineering Discovery Environment (XSEDE). Initial testing of our compiled parallel implementations in this environment using both the 16 GB and 120 GB point cloud files provided good results, which will be described below.

### **3. Description, implementation, and results**

#### **3.1 *p\_las2las***

##### **3.1.1 Description:**

The las2las application and supporting LASlib library were extended with the MPI application programming interface (API) to allow the application to be run in parallel on a cluster. Our goal is an application that scales to arbitrarily large input, limited only by the

volume of disk space needed to store the input and output files. Figure 1 shows the high level view of the application. The processes across the top are run in parallel, while the vertical flow describes the job flow.

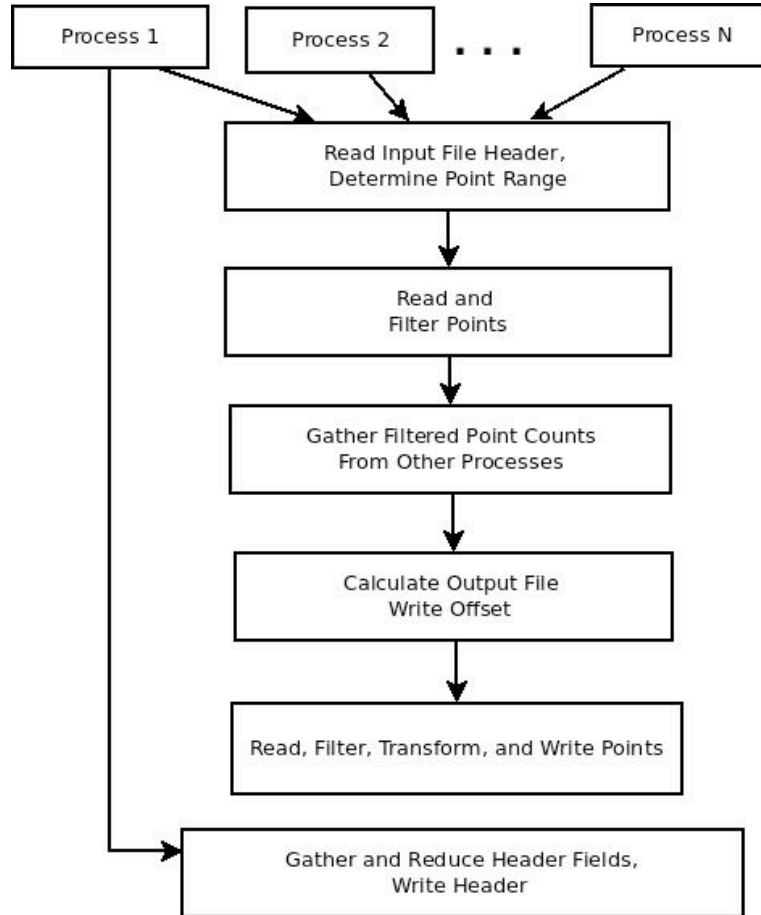


Figure 1. The high level view of the application. The vertical flow describes the job flow while the processes across the top are run in parallel on the flow.

### 3.1.2 Results

The test of *p\_las2las* on the XSEDE Stampede cluster used a Lustre parallel file system with 64 Object Storage Targets (OSTs – Factor *et al.*, 2005) available. We striped our 16 GB Smoky Mountains and 120 GB Grand Canyon test files over all 64 OSTs and specified a 4 megabyte (MB) stripe size; a common stripe size for large files managed by Lustre. The output directory was configured similarly. Table 1 shows the results of running the *p\_las2las* program on the Stampede supercomputer using the 16 GB Smoky Mountains dataset using various numbers of processes. The asterisk in the table refers to execution runs with native, “unmodified” *las2las* source code from *LAStools* compiled on Stampede with the Intel C++ compiler. Table 2 shows the same data for the 120 GB Grand Canyon dataset. For both sets of results, the tables describe the difference in elapsed time between the various test cases, as a function of number of processors.

Table 1. Smoky Mountains 16 GB File Results on Stampede.

Number of Processes	Filter / Transformation	Output Size	Elapsed Time (seconds)
Native*	None	16 GB	138
Native*	Keep Class 2	2 GB	73
Native*	Reproject	16 GB	502
64	None	16 GB	20
64	Keep Class 2	2 GB	6
64	Reproject	16 GB	26
256	None	16 GB	8
256	Keep Class 2	2 GB	4
256	Reproject	16 GB	9
1024	None	16 GB	8
1024	Keep Class 2	2 GB	5
1024	Reproject	16 GB	8

\* Native unmodified *las2las* source code from *LAStools* compiled on Stampede with the Intel C++ compiler.

Table 2. Grand Canyon 120 GB File Results on Stampede

Number of Processes	Filter / Transformation	Output Size	Elapsed Time (seconds)
Native*	None	120 GB	1211
Native *	Keep Class 2	25 GB	623
Native*	Reproject	120 GB	6969
64	None	120 GB	128
64	Keep Class 2	25 GB	59
64	Reproject	120 GB	150
256	None	120 GB	33
256	Keep Class 2	25 GB	18
256	Reproject	120 GB	42
1024	None	120 GB	18
1024	Keep Class 2	25 GB	9
1024	Reproject	120 GB	24

\* Native unmodified *las2las* source code from *LAStools* compiled on Stampede with the Intel C++ compiler.

### 3.2 *p\_points2grid*

#### 3.2.1 Description

Our goal is an application that scales to an arbitrarily large input, limited only by the amount of disk space needed to store the input and output files. When run on a cluster, the number of processes used by *p\_points2grid* is determined as a parameter to the scheduler. Figure 2 shows the high level view of the application. The job flow is described by the boxes on the right side while the processes along the left are the internal processes of the flow functions.

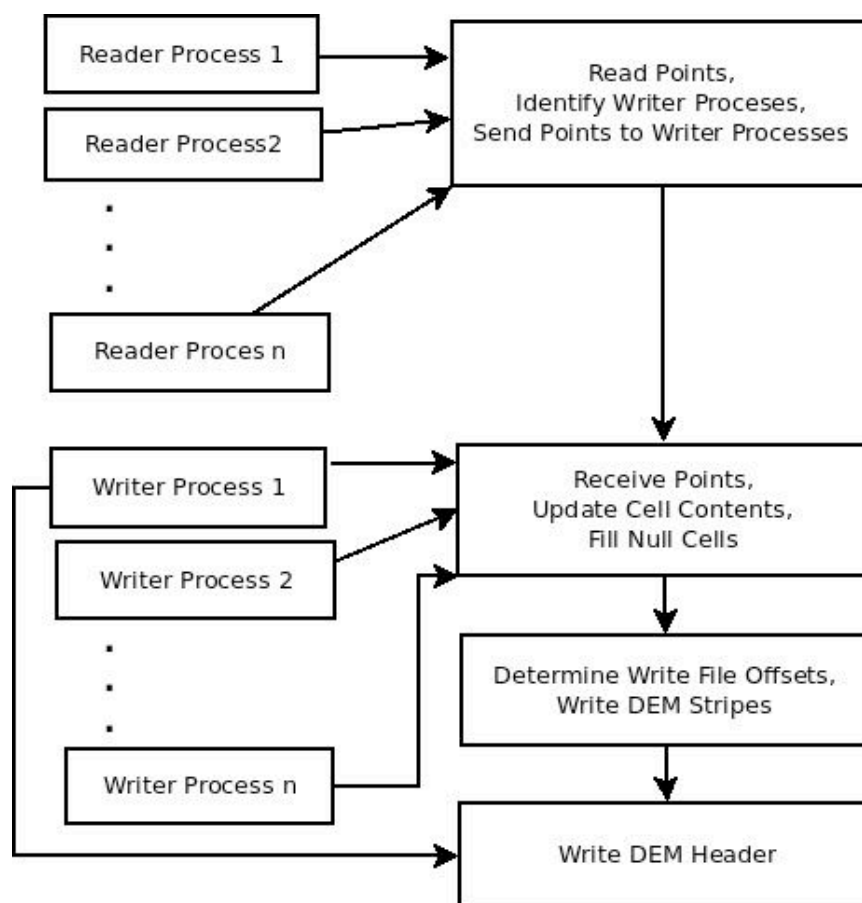


Figure 2. The high level view of the application. The job flow is described by the boxes on the right side while the processes along the left are the internal processes of the flow functions.

### 3.2.2 Results

The Smoky Mountains and Grand Canyon LAS input files were read from a Lustre File System on the “work” partition of Stampede. The files were striped over 64 OSTs with a 4MB strip size. The DEMs were written to the same directory that held the input files. The directory was configured to write files over 64 OSTs with a 4MB stripe size. Table 3 shows the results for the Smoky Mountains dataset and Table 4 shows the results for the Grand Canyon dataset. These results, in these two tables, show the varying time reading and communicating versus writing as the number of readers or writers are varied at execution time.

Table 3. Smoky Mountains 16 GB Input File Results, (12, 1 meter resolution DEMs totaling 70 GB of output for p\_points2grid runs, 12, 6 meter resolution DEMs totaling 2 GB of output for native run. Times are in seconds.)

Number of Processes	Number of Readers	Number of Writers	Time: Reading, Communication	Time: Writing	Elapsed Time (seconds)
Native	1	1	NA	NA	328
128	32	96	33	56	105

128	64	64	26	84	125
512	32	480	20	13	40
512	64	448	10	17	33
512	128	384	8	23	36
512	256	256	7	26	40
512	384	128	11	44	68
1024	64	940	10	11	32
1024	384	640	2	14	29
1024	768	256	6	28	46

Table 4. Grand Canyon 120 GB Input File Results, (12, 1 meter resolution DEMs totaling 71 GB of output for *p\_points2grid* runs, 12, 6 meter resolution DEMs totaling 2 GB of output for native run. Times are in seconds.)

Number of Processes	Number of Readers	Number of Writers	Time: Reading, Communication	Time: Writing	Elapsed Time (seconds)
Native	1	1	NA	NA	1548
512	64	448	104	15	135
512	128	384	55	21	90
512	256	256	60	30	110
1024	64	960	80	7	101
1024	128	896	39	11	62
1024	256	768	27	8	51
1024	384	640	24	15	53
1024	512	512	26	19	56
1024	768	256	47	24	90
1024	896	128	89	44	167
4096	256	3840	17	10	63
4096	512	3584	10	11	53
4096	1024	3072	8	8	46
4096	2048	2048	8	18	76

Our test runs of *p\_points2grid* specified a grid resolution of 1 meter. No output or cell value types were specified, so each run produced 12 1-meter-resolution DEMs. In the Smoky Mountains test case, each DEM has a dimension of 40,000 columns by 20,000 rows and the total size of all 12 files is approximately 70 GB. In the Grand Canyon test case each DEM has a dimension of 31,000 columns by 26,500 rows and the total size of all 12 files is approximately 71 GB. We also ran the native “unmodified” *points2grid* application against our test datasets. We had to specify a 6 meter grid resolution because the memory requirements for 1 meter resolution were well beyond what the native application supports. These runs produced 12 DEMS totaling about 2 GB, or 36 times smaller than the 1 meter grid resolution DEMS produced by *p\_points2grid*.

#### 4. Conclusions

By creating parallel processing algorithms based on the open source *las2las* and *points2grid* code bases, we have shown greatly reduced run times processing extremely large datasets (over 100 GB), both in classifying the points and in generating DEMs. Using these programs, *p\_las2las* and *p\_points2grid*, we have shown through preliminary testing approximately two or more orders of magnitude reduction in processing time. In addition, we have shown scalability up to 4,096 processes.

## Disclaimer

Any use of trade, product, or firm names in this paper is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## References

- ASPRS (American Society for Photogrammetry and Remote Sensing) (2008) *LAS Specification, Version 1.2*. Internet at [http://www.asprs.org/a/society/committees/standards/asprs\\_las\\_format\\_v12.pdf](http://www.asprs.org/a/society/committees/standards/asprs_las_format_v12.pdf). Last accessed 24 November 2014.
- Behzad, B., Y. Liu, E. Shook, M. P. Finn, D. M. Mattli, and S. Wang (2012). A Performance Profiling Strategy for High-Performance Map Re-Projection of Coarse-Scale Spatial Raster Data. Abstract presented at the *Auto-Carto 2012, A Cartography and Geographic Information Society Research Symposium*, Columbus, OH.
- Factor, M., K. Meth, D. Naor, O. Rodeh, and J. Satra (2005) Object storage: the future building block for storage systems. In *LGDI '05: Proceedings of the 2005 IEEE International Symposium on Mass Storage Systems and Technology*, pages 119–123, Washington, DC, USA. IEEE Computer Society.
- Finn, Michael P., Yan Liu, David M. Mattli, Babak Behzad, Kristina H. Yamamoto, Qingfeng (Gene) Guan, Eric Shook, Anand Padmanabhan, Michael Stramel, and Shaowen Wang (2015). High-Performance Small-Scale Raster Map Projection Transformation on Cyberinfrastructure. Paper accepted for publication as a chapter in *CyberGIS: Fostering a New Wave of Geospatial Discovery and Innovation*, Shaowen Wang and Michael F. Goodchild, editors. Springer-Verlag.
- Isenburg, Martin (2014) *lasmerge: Merge Multiple LAS Files into a Single File*. Internet at <http://www.liblas.org/utilities/lasmerge.html>. Last accessed 03 March 2015.
- Piernas, J., J. Nieplocha, and E. Felix (2007). Evaluation of active storage strategies for the lustre parallel file system. *Proceedings of the ACM/IEEE Conference on Supercomputing*. ACM, New York.
- rapidlasso GmbH (2014) *Lastools*. Internet at <http://rapidlasso.com/lastools/>. Last accessed 24 November 2014.