An Integrated Geospatial Data Management System in a Complex Public Research Environment using Free and Open Source Software

Christian Braun, Ulrich Leopold
Public Research Centre Henri Tudor,
Resource Centre for Environmental Technologies,
Esch-sur-Alzette, Grand Duchy of Luxembourg
e-mail: Christian.Braun@tudor.lu, Ulrich.Leopold@tudor.lu

***Abstract*—The interdisciplinary nature of environmental research centres, dealing with geospatial data, analysis and environmental modelling on a daily basis, requires specific methods and technologies in the field of geospatial information management. The large amount of generated information has to be stored, catalogued, visualised and treated effectively for further analysis. The Public Research Centre Henri Tudor has set up a prototype system to create an integrated geospatial data infrastructure, serving the needs of various user profiles from novice level to advanced and experienced data analysts and modellers. The paper will show solutions on how to give a broad range of users access to an integrated infrastructure. This is achieved by introducing different user interfaces: an easy to use web interface for beginners - advanced web mapping and feature services coupled to desktop GIS applications for intermediates - direct data base access, making use of cutting-edge geospatial tools and spatially distributed modelling algorithms for experts. The system is fully functional on all user levels and based on free and open source software. It is integrating current standards of the Open Geospatial Consortium, to assure exchange with stakeholders and to guarantee its further functional extensibility.**

Keywords - Geospatial data management; geospatial data infrastructure; Web services; Web GIS; INSPIRE.

I. Introduction

Since the enactment of the Infrastructure for Spatial Information in Europe (INSPIRE) directive [1] in 2007 the development of geospatial information services has increased. Stakeholders are obliged to create functional data and software infrastructures and to be compliant to defined data exchange guidelines. Web-enabled geospatial technologies are key components [2]; data base management systems with components to store spatial objects [3], data access libraries [4], map rendering engines [5] and web front ends are constantly adapted and developed to meet different needs.

These components, though interactive, are merely data viewers; the information flow has to be more open and bidirectional when users operate with the data by further geospatial analysis or by creating new data sets. In this case the focus is on the interaction between data storage, single software components and modelling tools. If a geospatial data management infrastructure has to handle operations like real-time mapping or automated interpolations [6], then further components have to be integrated. Examples for these are algorithms powered by geostatistical software tools like R and Gstat [7] or GIS software libraries that are working on data base level like TerraLib [8].

In the recent years, the Public Research Centre Henri Tudor has faced an increasing amount of geospatial information, generated and compiled by its daily work for more and more demanding models. This called for integrated geospatial data management techniques and a concept to design a solution based on free and open source software. The solution to this is a geospatial data infrastructure that is able to operate as a central hub, serving geospatial information to provide the necessary data to each collaborator.

This paper reflects on the possibilities of this proposition to build a fully functional integrated geospatial data infrastructure with free and open source components by the given constraints and requirements. This will be described in the next section, followed by an extensive description of used components and how they interact in Section III. Finally, a short description is given in Section IV on how the system is used in the institute so far and which major improvements will have to be made in the future.

II. System design, Requirements and Constraints

The topic of geospatial data was quite new at the institute and the collaborators were introduced to existing techniques and demo applications on the Internet illustrating how workflows could look like with advanced data management tools. In several brainstorming sessions certain requirements were identified and a prototype was designed. This led to a fully functional geospatial data infrastructure with advanced data management capabilities.

One major constraint was that the whole infrastructure had to be built on free and open source software products. The main aim of this is to reduce licensing costs and to increase interoperability and extensibility considering open

 ISBN: 978-1-61208-118-2

standards. With the interdisciplinary nature of an environmental research institute the system should be fully scalable to serve the needs of the various user profiles in particular.

An easy to use web based spatial information system should be an entry point for exploration (meta data search) and visualisation of all connected data pools. Furthermore, expert users should be able to make use of advanced geospatial data analysis and modelling with connected programming environments such as R [9], shown in Figure 1, and geographic information systems like GRASS GIS [10]; both embedded in high performance computing facilities.

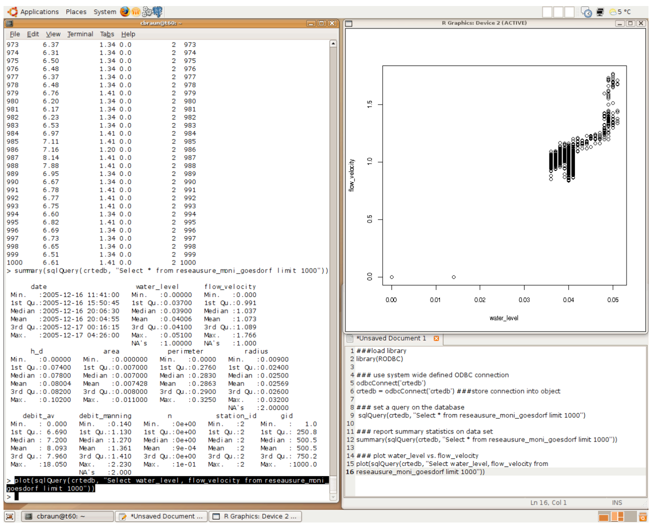

Figure 1. R implemented as modelling tool.

Another component is the meta data management and data exchange in standardised formats as defined by the Open Geospatial Consortium (OGC) [11]. This ensures the compliance with infrastructures on the national level (e.g. GeoPortal Luxembourg [12]) and the fulfilment of international reporting needs compatible with the INSPIRE directives. Meta data are a crucial part to clearly identify data sets and have information about extent, quality, spatial and temporal schema, spatial reference, and distribution of digital geographic data. One wide spread standard and "best practice" is ISO 19115 [13] with its XML implementation schema ISO 19139 [14]. In addition, the general effort in maintenance and supervision of the whole system should be as small as possible.

These requirements and constraints were leading to a first sketch of the infrastructure design shown in Figure 2 summarising the most important parts of the geospatial data infrastructure.

The central part is represented by the physical data storage system, which consists of one or more data bases and provides interfaces for exchange with software clients, such as GIS and other modelling environments used in the research centre. An application server is coupled to the physical storage and represents the high performance computing facilities of the institute. For data exploration, administration and maintenance the geospatial data infrastructure should provide a meta data system as well as administration tools.

An example workflow is the geospatial analysis and automated mapping of field measured soil properties (e.g. texture, content of organic carbon and water uptake rate) using automated mapping algorithms as developed in the INTAMAP project [15]. A user can upload the relevant data sets via a web interface to the central data base where he is able to set corresponding access privileges to his working group. If the gathered data does not need further analysis in a laboratory, field devices with compatible software could upload it via cellular network without effort. Later, different group members are able to access the data with the connected software products of their choice for an initial overview mapping or to check data for consistency. After positive data checks, further geostatistical analysis, such as interpolation with kriging methods [16] will be done and the data as well as analysis results are fed into the meta data base and the web mapping interface. This ensures fast access of necessary up-to-date information to all collaborators.

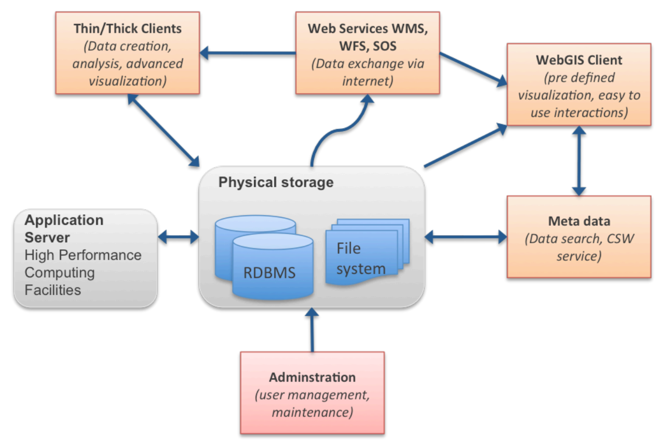

Figure 2. Overview of the main components of the prototype design.

With the implementation of Web Processing services (WPS) [17] users are able to access and run predefined models on the application server. Necessary data sets and parameters have to be included in the server request. This is a favourably tool for non-advanced users to run complex calculations with automated error checking capabilities and visualise outputs directly in the WebGIS as overlay maps. This is a recent achievement running a cast shadow calculation module in a city area implemented in PyWPS [18]. We are using the OGC WPS standard which offers native support for GRASS GIS and supports a generic WPS Java Script library.

III. THE GEOSPATIAL DATA INFRASTRUCTURE

The design of the geospatial data infrastructure can be seen in Figure 3, with special consideration of the given constraints and user needs. The figure shows the used software components and their arrangement and connection from a technical point of view. It is organized in five different layers with their unique assigned functions in the system.

The data pools at the bottom are representing the physical storage of different data sets. Vector and meta data are stored in a PostgreSQL [19] data base system with PostGIS [20] extension to allow the storage of spatial objects as well. Currently, raster data is stored simply file based, with ideas to move and couple it to the above mentioned PostgreSQL set up driven by PostGIS Raster [21]. This project is about to implement the raster data type as much as possible like vector type data in PostGIS. It will also offer a single set of overlay SQL functions operating seamlessly on vector and raster data. General user access and data security is handled on data base level by user login and password protection. With user and group privileges on table level inside the data base a granular access right system can be established to protect sensitive information.

Libraries like GDAL/OGR [22] make sure that access and coordinate transformation of geospatial objects is possible. A map server like MapServer [23] is able to create different OGC compliant services, such as Web Map Service (WMS), Web Feature Service (WFS) and Web Coverage Service (WCS) [18] and is providing the rendering of maps for the web client p.mapper [24]. This web interface is mainly based on PHP MapScript and is accessing data sets directly via the MapServer rendering engine.

Direct data access for data analysis software, such as the wide spread Microsoft Excel, or the more advanced environments like R and S-Plus could be realised through ODBC/JDBC interfaces. Desktop GIS clients, like GRASS GIS, uDig [25] or Quantum GIS [26] can access the data storage directly with built-in interfaces to visualize or edit geospatial data sets.

Another important component of the infrastructure is a catalogue service for meta data and a corresponding web interface like GeoNetwork [27]. It allows the user to interactively search for sets of any information in the system via related meta data. Furthermore it is a crucial part of compatibility to other infrastructures because GeoNetwork is acting as a CSW (Catalogue Service Web) service to publish data sets in higher-level infrastructures (e.g. Geoportal Luxembourg, INSPIRE services). Further it can connect to them as client as well. This feature enables harvesting of foreign data sets to link them into the own system by OGC web services.

The geospatial data infrastructure and its diverse options of data access enable the acquisition and compilation of relevant information in an integrated way for all types of user levels. It ensures easy and compatible data exchange and it is fully functional and well integrated in the daily work flow of collaborators.

A main tool for data research is the WebGIS and the web based meta data application where most basic information on available data can be retrieved. A most fundamentally work flow, such as search, selection, visualisation of geospatial information and printing of basic maps can easily be done via the web map interface adding full support to basic or novice users. Users with advanced geocomputation skills are using the WebGIS itself mainly for discussion purposes during meetings. Nevertheless, they are creating and analysing data sets by making use of powerful server hardware and further installed software products like PostGIS, which is offering spatial data queries and geoprocessing methodologies for vector data directly implemented into SQL (Structured Query Language).

Further applications could be automated mapping of environmental parameters or statistical analysis tools built into the web based map interface to provide further insight to data to the more inexperienced user. The Public Research Centre Henri Tudor is also involved in geospatial uncertainty analysis and modelling within the developments of the UncertWeb project [28] and its Uncertainty Markup Language (UncertML). These are of importance in the near future, e.g. in decision making at administrative levels.

All functional blocks are built on free and open source software (FOSS) components. FOSS is used in many applications and operational infrastructures, making use of advantages like the constant and continuous development and on-going support by the developer community. The monetary benefit of free and open source software could be reduced in the beginning phase because of extra expenses caused by an additional need in development of features that are not available out-of-the-box. In terms of desktop GIS software, standardized interfaces to OGC services are available due to the effectiveness of OGC standards and the active community support.

IV. Conclusion and Future Work

The presented infrastructure has the following benefits: improved data organisation and management, high data security and fast availability of information on all user levels.

The system is able to connect all GIS and modelling tools that are used in the institute seamlessly and present them in an integrated way to the user.

As European frameworks like INSPIRE are just being implemented, the integration of distributed data sources will continue in the future. This integration will be gradual, adding data when it is available and prepared for integration. The system provides all necessary interfaces to achieve an easy and straightforward integration of all OGC compliant data sources. At the same time it also complies with all OGC standards to provide data for many other geospatial data infrastructures and applications.

Future work will have to be done in the fields of a more seamless WPS integration and the set up of more calculation modules to assist collaborators. The storage of raster data should be integrated in the data base system by the recent development of PostGIS Raster. This would allow making use of basic raster map algebra, also in combination with vector data, without using dedicated GIS tools. Furthermore, the overall usability of the web frontend should be improved with better support for data queries and quick look summary statistics of available data sets. In addition, the system should be referenced in the national geospatial data infrastructure to assure an exchange of information and make use of meta data services.

REFERENCES

[1] Directive 2007/2/EC of the European Parliament and of the council of 4 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), http://inspire.jrc.ec.europa.eu 05/10/2010

[2] Anderson, G. and Moreno-Sanchez, R. (2003): Building web-based spatial information solutions around open specifications and open software, Transactions in GIS, vol. 7(4), pp. 447-466.

[3] Yeung, A.K.W. and Brent Hall, G. (2007): Spatial database systems: design, implementation and project management. Springer.

[4] Warmerdam, F. (2008): The Geospatial Data Abstraction Library. In: Open Source Approaches to Spatial Data Handling, vol. 2, pp. 87-104. Springer, Berlin.

[5] Vatsavai, R., Shekhar, S., Burk, T., and Lime, S. (2006): UMN-MapServer: A High-Performance, Interoperable, and Open Source Web Mapping and Geo-spatial Analysis System. In: Geographic, Information Science. Lecture Notes in Computer Science, vol. 4197, pp. 400-417. Springer, Berlin.

[6] Brenning, A. and Dubois, G. (2007): Towards generic real-time mapping algorithms for environmental monitoring and emergency detection. Stochastic Environmental Research and Risk Assessment, vol. 22(5), pp. 601-611.

[7] Bivand,R.S., Pebesma, J.E., and Gómez-Rubio, V. (2008): Applied spatial data analysis with R. Springer.

[8] Câmara, G., Vinhas, L., Ferreira, K., Queiroz, G., Souza, R.C.M., Monteiro, A.M. et al. (2008): TerraLib: An open-source GIS library for large-scale environmental and socio-economic applications. In: Open Source Approaches to Spatial Data Handling, vol. 2, pp. 247-270. Springer, Berlin.

[9] R Development Core Team, 2010, R: A Language and Environment for Statistical Computing, Vienna, Austria. http://www.R-project.org 05/10/2010

[10] GRASS Development Team, 2010, Geographic resources analysis support system (GRASS GIS) software, Trento, Italy. http://grass.osgeo.org 05/10/2010

[11] Open Geospatial Consortium, http://opengeospatial.org 05/10/2010

[12] Administration du Cadastre et de la Topographie, http://act.public.lu 01/10/2010

[13] ISO 19115:2003, http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020 06/12/2010

[14] ISO 19139:2007, http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557 06/12/2010

[15] Interoperability and Automated Mapping: INTAMAP, http://www.intamap.org 29/09/2010

[16] Chilès, J.-P. and Delfiner, P. (1999): Geostatistics: modeling spatial uncertainty. Wiley, New York.

[17] OGC Standards and Services, http://www.opengeospatial.org/standards 10/10/2010

[18] PyWPS, http://pywps.wald.intevation.org/ 06/12/2010

[19] PostgreSQL, http://www.postgresql.org 10/10/2010

[20] PostGIS, http://postgis.refractions.net 10/10/2010

[21] PostGIS Raster, http://trac.osgeo.org/postgis/wiki/WKTRaster 10/10/2010

[22] GDAL, http://www.gdal.org 15/09/2010

[23] MapServer, http://mapserver.org 10/10/2010

[24] p.mapper, http://pmapper.net 15/09/2010

[25] uDig – User-friendly Desktop Internet GIS, http://udig.refractions.net 03/09/2010

[26] Quantum GIS, http://www.qgis.org 15/09/2010

[27] GeoNetwork opensource, http://geonetwork-opensource.org 05/10/2010

[28] UncertWeb, http://www.uncertweb.org 12/10/2010

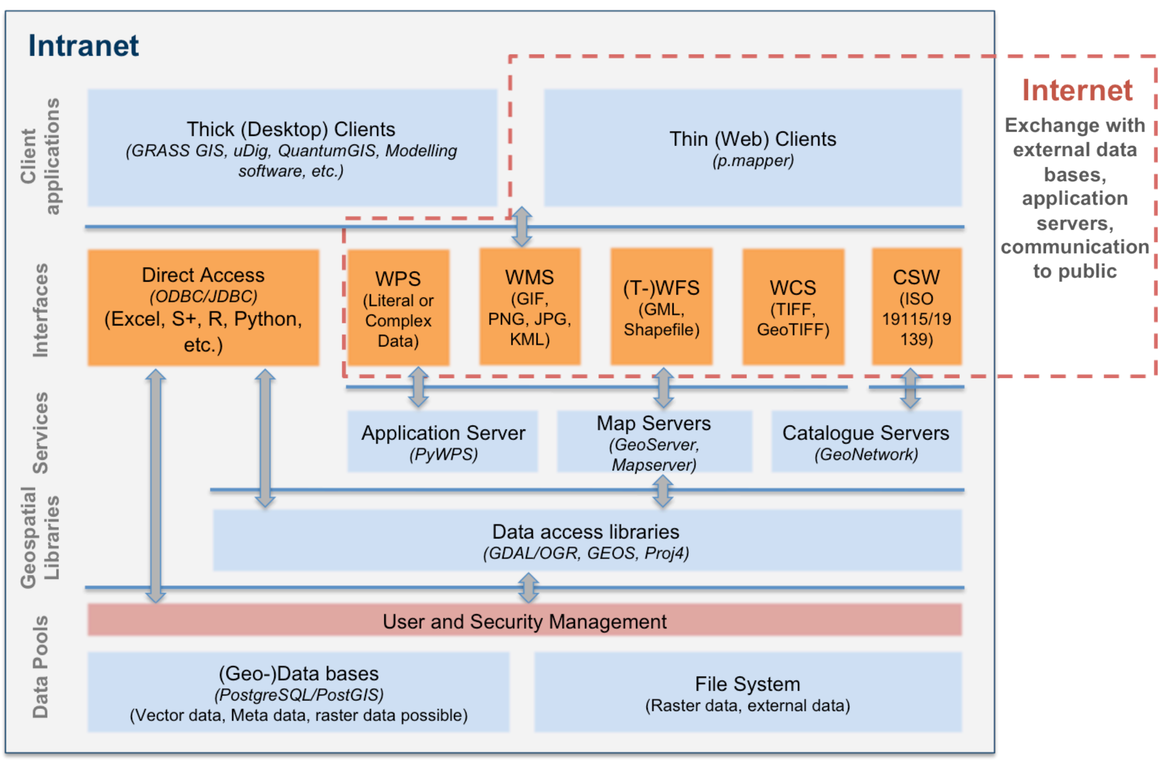

Figure 3. Detailed schema of components of the geospatial data infrastructure.