

GeoNetwork OpenSource: From Index Cards to Spatial Data Infrastructures

Robert S. Nuske¹ and Jan C. Thiele²

¹Dept. of Forest Growth, Northwest German Forest Research Station, Grätzelstr. 2, 37079 Göttingen, Germany

²Dept. of Ecoinformatics, Biometrics and Forest Growth, University of Göttingen, Buisenweg 4, 37077 Göttingen, Germany

Abstract

Spatial data management is of particular importance to projects related to natural resources, because of their interdisciplinary nature, the large and high amount of datasets and the need for data exchange. Since project members are often distributed between different locations, but their success depends on a consistent and easy exchange of data, central data repositories are often web-based. In this article, we describe the use of GeoNetwork OpenSource for building spatial data infrastructures and report on experiences from practical application.

Introduction

To manage or study forest ecosystems one does not only need to know their inherent properties but also their context (cf. Fig. 1). Most of the parameters describing the characteristics and the drivers of the forest ecosystems and their context have a location on the Earth's surface. Thus, it becomes important not only to capture the "what" and "how much", but also the "where" of the phenomenon under consideration. Such spatial information does not only enable us to match different data sources by their common location but also to study a new set of questions such as the influence of the neighborhood or the dispersion of objects. In recent decades, spatial information has become crucial for decision-making in areas such as natural resources, urban development, facilities and even geomarketing (Buehler and McKee, 1996; Nogueras-Iso et al., 2005).

Although, large amounts of spatial data are gathered at an increasing speed, thanks to Earth observation satellites, Global Positioning Systems (GPS), autonomous sensor networks and an increased interest by individuals and institutions (Nogueras-Iso et al., 2005); the public availability and exchange of spatial information grows at a slower pace. Low awareness of data existence, missing conditions of use, data inconsistencies and poor documentation impede the effective use of the available spatial data (Craglia et al., 1999, Official Journal of the European Union, 2003, Nogueras-Iso et al., 2005).

To overcome these shortcomings a catalogue of the available data is needed. Such a catalogue was built up of index cards in the old times and is nowadays an online database containing the authors, titles, keywords etc. of the in stock datasets. If the dataset has a geographic reference, the coordinate reference system, its position in space, production date, quality and data format are also of importance for a prospective user. Such descriptions of datasets are called metadata. Metadata are not only relevant for finding datasets, but are also required to ensure the traceability and quality and to help preserve the meaning of a dataset over a long period of time (e.g. long-term archiving). The actual data may be stored in a central spatial data infrastructure (SDI) or in distributed local data repositories. Depending on the needs of the operator, a pure metadata catalogue can be extended to a SDI by adding functionalities such as interactive visualization, a metadata editor, rights management, and access control.



Fig. 1: Forest Stand in Northwest Germany (middle) and some of its properties and context (from left to right: stand table, forest map, aerial imagery, digital terrain model). Source: Created by the authors, using DEM and Aerial Photographs (c) GeoBasis-DE/BKG 2011.

Standard compliant metadata catalogues can query each other and link to datasets in other collections. The most important standard covering how to publish metadata for spatial data and services is the “[Catalogue Service for the Web](#)” (CSW) (Nogueras-Iso et al., 2005, Voges and Senkler, 2008) of the [Open Geospatial Consortium](#) (OGC). The OGC is a non profit organization with the mission to “advance the development of international standards for geospatial interoperability. This standard became important in connection with the [INSPIRE](#) Directive “establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment” (European Parliament, Council, 2007).

GeoNetwork OpenSource

A wealth of different software solutions are currently available to create spatial data infrastructures, such as [terraCatalog](#), [eXcat](#), [CatalogCube](#), [deegree](#), or [GeoNetwork OpenSource](#). We will focus on GeoNetwork OpenSource, since it is an already widespread and established solution to operate a SDI. Furthermore it is the OGC reference implementation of the CSW 2.0 standard and as an open source project available under the [GNU General Public License](#) [GPL, free of charge and adaptable even on source code level (Hielkema and Ticheler, 2007, Prunayre, 2010)].

The development of GeoNetwork began in 2001 at the [Food and Agriculture Organization](#) (FAO). The main objective back then was to produce a catalogue system for systematic publishing and archiving of geospatial data. In 2003, the World Food Programme (WFP) became involved in the development of the software and the first version was released. A year later the United Nations’ Environment Programme (UNEP) joined the project. GeoNetwork is now an approved project of the [Open Source Geospatial Foundation](#) (OSGeo) and deployed at many large organizations (Hielkema and Ticheler, 2007), such as the Consultative Group on International Agricultural Research (CGIAR), the European Space Agency (ESA) and the U.S. Federal Geographic Data Committee (FGDC). It is also used increasingly for developing national SDIs in Europe, e.g. [Germany](#), [Netherlands](#), [Switzerland](#), and various German federal states, such as [Lower Saxony](#) and Bavaria (Sanders and Weichand, 2011). Thus, there are strong and reliable partners and a diverse and vibrant developer community behind the GeoNetwork project to ensure the continuity and quality of the development.

Practical Experience

For the project NaLaMa-nT a central data infrastructure was established to provide a consistent and obligatory basis of datasets for the entire team using GeoNetwork. This data infrastructure shall ensure at the same time that only authorized persons gain access to the metadata and their associated datasets, since most datasets are not for public consumption. NaLaMa-nT will develop a knowledge and decision basis for a sustainable land management for the North German Plain against the background of climate change and increasingly globalised markets based on a transdisciplinary analysis of current land use systems such as forestry, agriculture, and water management, and their interactions. According to the broad research approach the different working groups employ diverse data and methods. Primary information is derived from monitoring networks, literature, and statistics. Those will be complemented by regional data, like ecological monitoring, inventory data, and other forms of data collection. Moreover, additional data will be gathered from experiments, indicator areas, cooperating companies, and choice-experiments.

The installation of [GeoNetwork](#) in a servlet container like Apache Tomcat on a server using a web application archive (WAR-file) was easily possible. After successful installation, the main website of the portal is available to be accessed via a web browser. Adjustments of the layout of GeoNetwork can be achieved fast by inserting a logo or a complete redesign of the interface via xml/xsl files. Since the project team is solely based in Germany, it was expected from the data infrastructure to talk a pleasant German. The existing German translations had to be tweaked to achieve that target. A PostgreSQL database with the [PostGIS](#) extension was set up to store the metadata and spatial indexes. The spatial data themselves were for the most part uploaded directly to GeoNetwork while editing the metadata and are therefore mainly stored in the file system of the server. That way they are directly downloadable for the registered user from GeoNetwork. However, there are also references to services such as [Web Map Service](#) (WMS) and [Web Feature Services](#) (WFS) or links to other data collections. Of particular importance to user of the data infrastructure is the information about the area covered, coordinate

reference system and the responsible party for the resource or the originator of the metadata. So far, we made no use of the option to “harvest” other metadata catalogues (include other sources in their own portal, such as the very large publically available [FAO Catalogue](#)) nor advertise our catalogue for “harvesting.”

For data descriptions GeoNetwork natively supports the three most important standards for spatial data: [ISO19139](#) (from the International Standards Organization, ISO; it is the xml implementation specification of ISO19115/119 which is also used for INSPIRE, Prunayre, 2010), [Dublin Core](#) (from the Dublin Core Metadata Initiative) and the standards of the [Federal Geographic Data Committee](#) (Hielkema and Ticheler, 2007). GeoNetwork support also self defined metadata schemas via XML files, if necessary.

Metadata can be easily edited with the convenient metadata editor. The (mandatory) input fields are derived from the metadata schema. They can be adjusted as needed (usually reduced) and pre-filled with default values. For each dataset collected in GeoNetwork an XML statement is stored in the database. This includes a description of the thematic content, the geographical location, the producer, the coordinate reference system, the scale, and quality. Important for the long-term use of the datasets are also information about the production date, the period of validity, the temporal and spatial validity of data and its permanent accessibility.

About 50 members of the NaLaMa-nT project use GeoNetwork to search for and download suitable data. After receiving their user name and password, they were able to operate the system intuitively without further instructions. The full text search function and the location based search performed very well according to our users (cf. Fig. 2). The built-in indexing system ensured a sufficient response time even with a large number of entries. Furthermore, GeoNetwork provides an interactive map for the visualization of local and remote geodata and services. In addition to the metadata management, GeoNetwork, delivers nearly all of the components needed to set up a spatial data infrastructure in accordance with the guidelines of the OGC (Rose, 2004).

Conclusion

Sustainable data management does not only include the long term preservation of data but also guarantees the continued maintenance, support, and further development of the infrastructure itself. Open source applications offer the advantage to build upon proven software components from previous projects and to pass on own enhancements. Experiences gained from the use of GeoNetwork Open Source in the forestry sector confirm that the software meets the requirements for data management and is flexible enough for customization.

The number of data infrastructures will grow considerably in the future, among other things because of the INSPIRE Directive, binding for public administrations, and the requirements for research projects to consider more closely the area of data management, for example enforced by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG; Winkler-Nees, 2011). A standard-compliant and customizable Open Source solution will help to implement data infrastructures faster and to avoid unnecessary in-house developments. The use of established and well-documented solutions, like GeoNetwork OpenSource, facilitates the sustainable exchange of data across different projects and a straightforward maintenance and development even after changes in staff.

Acknowledgement

We thank an anonymous reviewer for valuable comments on an earlier version of the manuscript. This study was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) in the BMBF-Funding Measure “Sustainable Land Management.”

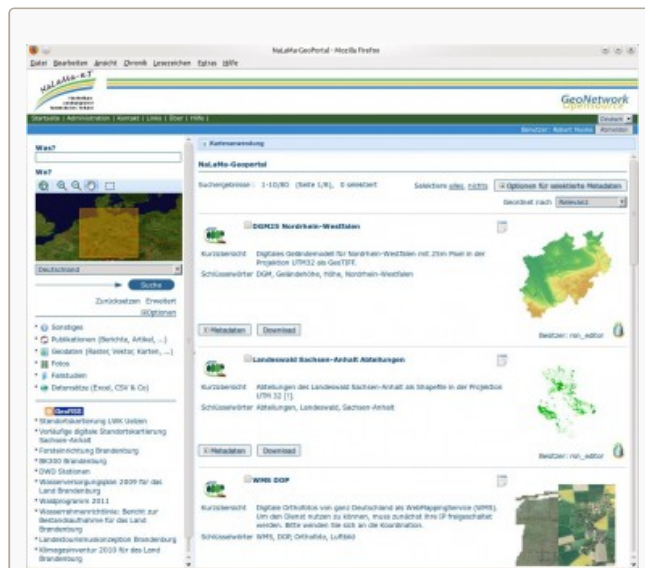


Fig. 2: Result Set, produced by a local GeoNetwork Portal run by the NaLaMa-nT Project. Source: Created by the authors.

References

- Buehler, K. and McKee, L. (1996): The OpenGIS Guide. Introduction to Interoperable Geoprocessing. Part I of the Open Geodata Interoperability Specification (OGIS). OGIS TC Document 96-001, OGIS Project 6 Technical Committee of the OpenGIS Consortium Inc.
- Craglia, M., Annoni, A., Masser, I. (Eds.) (1999): Geographic Information Policies in Europe: National and Regional Perspectives. EUROGI-EC Data Policy Workshop, Amersfoort, 15 November 1999. European Commission—Space Applications Institute, European Communities, <http://www.ec-gis.org/reports/policies.pdf> (last accessed 2012-06-01).
- European Parliament, Council (2007): Directive 2007/2/EC of the European Parliament and of the Council of 14th March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).
- Hielkema, J.U., Ticheler, J. (2007): FAO: Eine weltweite Geodaten-Plattform, GIS-Business 1/2: 17-19.
- Nogueras-Iso, J., Zarazaga-Soria, F.J.; Béjar, R.; Álvarez, P.J. and Muro-Medrano, P.R. (2005): OGC Catalog Services: a key element for the development of Spatial Data Infrastructures. Computers & Geosciences, (31) 2: 199-209.
- Official Journal of the European Union (2003): Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. L 345, 2003: 90–96.
- Prunayre, F.-X. (2010): INSPIRE support in GeoNetwork opensource. Dezember 2010 <http://www.neogeo-online.net/blog/archives/679> (last accessed 2012-06-01).
- Rose, L.C. (2004): Geospatial Portal Reference Architecture: A Community Guide to Implementing Standards-Based Geospatial Portals, OGC 04-039, 2004.
- Sanders, M., Weichand, J.(2011): Geoportal Bayern – Tor zur Welt der Geodaten. Mitteilungen des DVW Bayern, 3/2011: 215-225.
- Ticheler, J., Hielkema, J.U. (2007): GeoNetwork opensource, OSGeo Journal Vol. 2, September 2007: 15 – 19.
- Voges, U., Senkler K. (2008): OpenGIS Catalogue Services Specification 2.0.2 -ISO Metadata Application Profile. OGC: 07-045, Open Geospatial Consortium.
- Winkler-Nees, S. (2011): Promoting Accessibility to Research Data in Germany: Funding Initiatives, Projects and Perspectives. Workshop Making Scientific Research Data Accessible: Current Trends and Perspectives in Germany. Washington DC, June 21, 2011.
- Robert S. Nuske studied forest sciences at the University of British Columbia, Canada, and the University of Göttingen, Germany, where he received an MSc in Forest Ecosystem Analysis and Information Processing. He is interested in quantitative methods to analyze the dynamics of (near) natural forests (such as remote sensing, photogrammetry, and spatial statistics) and the management and processing of spatial data in general. Within the joint research project “[Sustainable Land Management for the North German Plain](#),” he is responsible for the central data and information management.
- Jan C. Thiele received a diploma degree in economics (University of Applied Sciences of Bremen, Germany) and an MSc in forestry sciences (University of Göttingen, Germany). Currently, he is on the way to finish his Ph.D. in the graduate program of environmental informatics at the University of Göttingen. He is interested in (web-based, spatial) decision support/information systems and data management and has a second focus on agent-based modelling, especially regarding tools and standards for building and analyzing models.