Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Spatial data management for sustainable land-use

Hansen, Henning Sten

Published in:
Geographic information facilities supporting access to environmental information

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

SPATIAL DATA MANAGEMENT FOR SUSTAINABLE LAND-USE

Henning Sten Hansen
National Environmental Research Institute
Roskilde, Denmark
hsh@dmu.dk

SUMMARY
Land is essential for the provision of food, water and energy for many living systems, and is critical to human activity. However, in rapidly growing urban areas, access to land is restricted by the competing demands of housing, industry, commerce, infrastructure, transport, agriculture as well as the need for open spaces and green areas, and the protection of sensitive ecosystems. With on average 117.5 people per square kilometre in Europe, it is easy to see why land use planning and management is such an important environmental issue for the EU.

SENSOR (http://www.sensor-ip.org/) is funded under priority area 6.3 of the Sixth Framework Programme Global change and ecosystems, and it was launched in December 2004. The aim of the project is to develop integrated, computer-based, sustainability impact assessment tools to support the European Union in assessing more accurately the longer-term impact of its policies on multifunctional and sustainable land-use. Access to reliable and harmonised data across Europe is a fundamental precondition for realisation of the SENSOR project.
Interoperability and open architectures are core requirements for state of the art implementations of IT solutions (Klopfer, 2006). Service oriented architectures based on a commitment to using open standards enables a system of component based building blocks, which can be chosen, run and maintained according to their best match of user requirements, independent of vendor solutions or storage models.

A basic foundation for all data related work in SENSOR is the draft INSPIRE principles (INSPIRE, 2002), which by themselves build on various international standards. Standardisation bodies like ISO or CEN are developing de jure standards, whereas organisations like the Open Geospatial Consortium (OGC) develop specifications that by a consensus process and their common acceptance become de facto standards.

KEYWORDS: spatial data management, land-use, INSPIRE, spatial data mining.

INTRODUCTION

The objective under work package 5.3 ‘Data protocols and system requirements’ as set out in the Description of Work (DoW) of Sensor states the development of a GIS-based, quality assured and harmonised data management system for sustainability impact assessment of land use, which satisfies end-users needs and can be employed for regional assessments at EU25 scale beyond the lifetime of the project. This includes a framework for indicator sets and criteria for indicator selection ensuring a harmonised approach across the SENSOR project and indicator sets covering the environmental, social and economic dimensions of sustainable development. A key point concerning data management is the ability for the users to search for and evaluate data, which fit perfectly into the intended use. Therefore metadata is of great importance for all data management systems, and from the beginning of the SENSOR project we have focused very much on the ability to report metadata for data added to the SENSOR database.

Developing a common data infrastructure requires some degree of standardisation (interoperability) among the various data sets. Although, the standards of interest to the SENSOR project are not static but will evolve during the project period as technology changes, the draft specifications of the INSPIRE initiative on architecture, standards and metadata as the main guidelines in this task. Based on this foundation, an overall frame for the data infrastructure including Web-based catalogue services enabling participants to discover and download appropriate data for their work will be designed and a prototype developed (Figure 1). NERI, EFI and ARCSys will undertake these tasks using the mentioned knowledge of the thoughts behind the INSPIRE initiative and expertise on developing clearinghouse like systems. Standard (off-the-shelf) GIS software will be applied for analysis, modelling and visualisation purposes at the client side.

The main aim of the SENSOR Data Management System will be to support the project partners concerning data handling. To do this the system will include the following components

- Data Warehouse
- Geoportal (Clearinghouse mechanism)
- Metadata reporting system
- Upload and download of data
- Pre- and post processing tools

Besides these IT components the SENSOR Data Management system contains a defined set of Core data and a SENSOR Data Policy.

The SENSOR operational integrated GIS Data management system will not only need to take into consideration the technical system requirements but also be sure to serve clearly the operational needs of the SENSOR consortium and other targeted user groups outside the project. This implies active involvement of both the SENSOR coordinator and the module leaders. This process was initiated in a first step when investigating data needs of the consortium partners in the data policy paper.

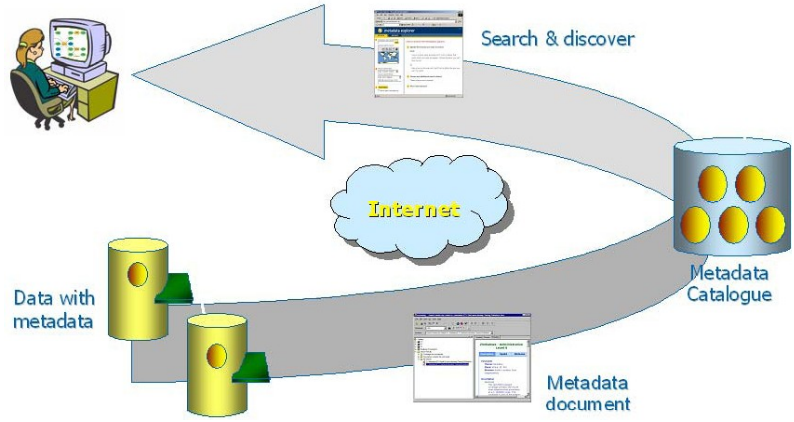

Figure 1. The Data Management System from a user's point of view.

During early spring we carried out a questionnaire among SENSOR partners in order to find out their expectations and requirements to the SENSOR Data Management System. The results of this questionnaire were presented at the SENSOR cluster meeting in Bratislava 24. – 26. April 2006. We got feedback – answers on the questionnaire – from 12 partners. However, only one partner's response addressed the important questions concerning the functionality of the SENSOR Data Management System. That single answer emphasises the ability to search for geographic data, to view table contents, to download selected data sets, to change to projection of spatial data, and to export to XML-format. This is certainly not impressive expectations and requirements.

Concerning expectations to data upload from partners there is nearly an equal division between spatial data and tabular data. Remark that tabular data must be associated to existing core reference data – primarily NUTS, but EuroGrid cells are also a possibility.

SENSOR Questionaire Administration Portal

Below you can see the entries of the SENSOR data questionaire so far.

Using file-based database:

Records Table

File format	Content	Coverage	Entity orCell Size	Time and Timerange	Data provided by	Data requested from	Data used by module	Input orOutput	Partner Abbreviation	Email
Shape File	NUTSx enitities	Europe 25	NUTSx	2005	NERI, IIASA	-	ALL	I	NERI	HSH@neri.dk
Excel Table	data on indicators related to IAG	Europe	NUTSx	50 years	BTUC	M6 Partners	M6	O	BTUC	dilly@tu-cottbus.de
Excel Table	Sensitive Area Case Studies on indicators related to IAG	Europe	NUTS3	50 years	BTUC	M6 Partners	M6	I	BTUC	dilly@tu-cottbus.de
Excel Table	Forest areas by NUTSx	Europe 25	NUTSx	1990	EFI	no idea	ALL	O	EFI	sergo@efi.fi
Excel Table	land claim for production forest coniferous/deciduous	Europe	NUTS0	from now to 2030 for all scenarios	EFI	WUR	M1	I	WUR	martha.bakker@wur.nl
Excel Table	landclaim per agricultural sectors	Europe	NUTS0	from now to 2030 for all scenarios and policies	Capri-Nemesis	WUR	M2	I	WUR	martha.bakker@wur.nl

Figure 2. Extract from the Data Questionnaire Portal

PRINCIPLES OF DISTRIBUTED GIS TECHNOLOGY

GIS technology is evolving beyond the traditional GIS community and becoming an integral part of the information infrastructure in many organisations. The unique integration capabilities of a GIS allow disparate data sets to be brought together to create a complete picture of a situation. Thus organisations are able to share, coordinate, and communicate key concepts among departments within an organisation or among separate organisations using GIS as the central Spatial Data Infrastructure. GIS technology is also being used to share information across organisational boundaries via the Internet and with the emergence of Web services.

An open GIS system allows for the sharing of geographic data, integration among different GIS technologies, and integration with other non-GIS applications. It is capable of operating on different platforms and databases and can scale to support a wide range of implementation scenarios from the individual consultant or mobile worker using GIS on a workstation or laptop to enterprise implementations that support hundreds of users working across multiple regions and departments. An open GIS also exposes objects that allow for the customisation and extension of functional capabilities using industry standard development tools.

The current chapter will describe some of the most important elements of distributed GIS, as we will use the concept in SENSOR.

Standards and operability

Interoperability and open architectures are core requirements for state of the art implementations of IT solutions (Klopfer, 2006). Service oriented architectures based on a commitment to using open standards enables a system of component based building blocks, which can be chosen, run and maintained according to their best match of user requirements, independent of vendor solutions or storage models.

A basic foundation for all data related work in SENSOR is the draft INSPIRE principles (INSPIRE, 2002):

- Data should be collected once and maintained at the level where this can be done most effectively
- It should be possible to combine seamless spatial information from different sources across Europe and share it between many users and application
- It should be possible for information collected at one level to be shared between all the different levels, detailed for detailed investigations, general for strategic purposes

- Geographic information needed for good governance at all levels should be abundant under conditions that do not refrain its extensive use
- It should be easy to discover which geographic information is available, fits the needs for a particular use and under which conditions it can be acquired and used
- Geographic data should become easy to understand and interpret because it can be visualised within the appropriate context selected in a user-friendly way.

Standards define the common agreements that are needed to achieve interoperability between IT components. Standardisation bodies like ISO or CEN are developing de jure standards, whereas organisations like the Open Geospatial Consortium (OGC) develop specifications that by a consensus process and their common acceptance become de facto standards.
Several ISO TC/211 standards are of high importance for building Spatial Data Infrastructures. Besides the ISO Standards the Open Geospatial Consortium (OGC) has developed implementation rules to ensure interoperability. Products and services compliant to OpenGIS interface specifications enable users to freely exchange and apply spatial information, applications and services across networks, different platforms and products. Besides these GI related standards, we will build the system on general IT standards like XML (extensible Mark-up Language, SOAP (Simple Object Access Protocol and WSDL (Web Services Description Language).

Data warehouse Architecture

A Data Warehouse is often defined as a subject-orientated, integrated, time-variant, non- volatile collection of data that support the decision-making process in an organisation (ESRI, 1998). In general a Data Warehouse is a large database organising data from various sources in a repository facilitating query and analysis. The database is well designed and contains key data, which are of high importance for the organisation.

However why do we need a spatial data warehouse in SENSOR. First we have to realise that the SENSOR project involves 35 partners from many countries, and the data sources are very wide spread. The main task for the central database is to facilitate access to data for all partners. Most common data set should be put into the Data Warehouse and harmonised so they fit into the overall architecture of the system. Data downloaded from EuroStat, ESPON, or the European Environment Agency are not usable immediately, but must be transformed in various ways. All tabular information associated with various NUTS classifications will be checked for consistency before being uploaded. Additionally, all data produced as part of the SENSOR project must be uploaded to the central Data Warehouse in order to obtain synergy.

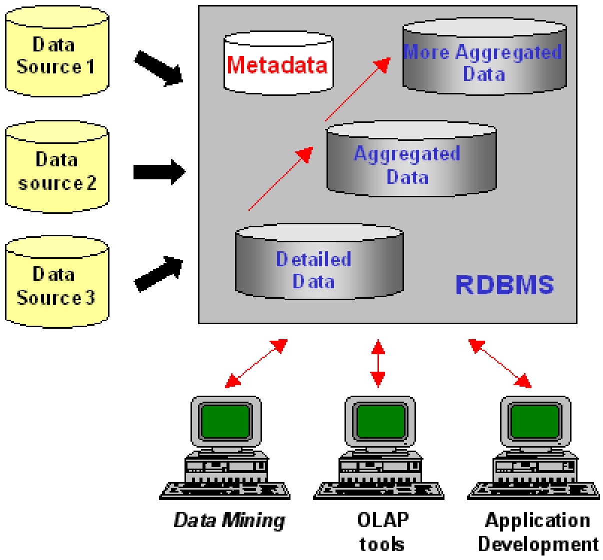

Figure 3. Data Warehouse Architecture.

Geoportals and clearinghouses

Efficient use of geographic information assumes access to documentation that describes origin, quality, age, ownership and fitness for purpose. This associated information is referred to as metadata (see paragraph 2.4). A key component of any spatial data infrastructure is a catalogue of metadata that can be used in searching for data using geographic location, time and thematic attributes.

The term Geoportal has nearly replaced the earlier term data clearinghouse. Technically the word portal refers to a web site acting as an entry point to other web sites (Tait, 2005). Further developing this definition a Geoportal will be a web site that represents an entry point to sites with geographic content.

Spatial portals was developed as the gateways to SDI initiatives and served as contact point between users and data providers. The Geoportal allow users to search and browse between huge amounts of data. One of the earliest attempts to develop a Geoportal was the US Federal Geographic Data Committee's Clearinghouse, and in Europe the INSPIRE proposal resulted in the development of a European Geoportal (Bernard et al., 2005).

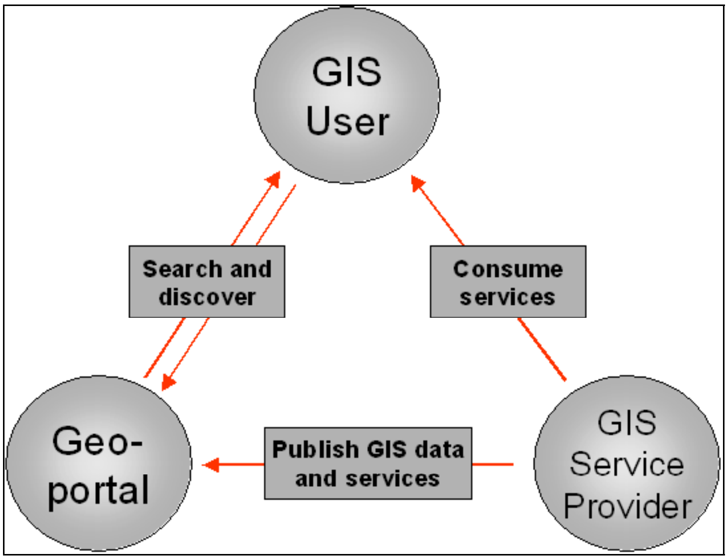

Figure 4. The relationships between Geoportal, user and service provider (After Tang & Selwood, 2005).

Geoportals can be divided into two groups: Catalogue Geoportals and Application Geoportals (Tang and Selwood). Catalogue portals create and maintain indexes describing available information services. Catalogue portals are useful when they provide information to a wide variety of services, data providers and user groups. Application portals combine information services into a Web based mapping application that generally focuses on a particular task. Their target community is well defined and they provide efficient access to data and functional services, which the portal manager selects to meet the user's needs. Often – like in the SENSOR project – some kind a combination between Catalogue and Application Portal is used.

Geoportals provide tools for searching, viewing, exploring, downloading and publishing (uploading) spatial information. The search tool can be based on attributes or key words returning a list of candidate items for final selection. Alternately, the user can draw a rectangle on a map returning a list of all data sets covering the requested area. In many cases both geographic and attribute criteria are allowed for searching. Attached to each data set you will have metadata, which gives the user deeper information about the data. General speaking it is not a prerequisite that a Geoportal provide map visualisation capabilities, but it will add value to the searching process, and very few – if any – do not posses the viewing capability.

The publishing process is the most important part – without any (meta) data. Publishing comprises addition, modification and deletion of metadata. The SENSOR project has focused much on this effort and a web based metadata publishing / reporting system has been available since August 2005 (SENSOR deliverable 5.1.1).

Geoportals are built using the World Wide Web infrastructure technology and GIS software, and the front end is typically sits on top of some kind of Internet Map Server, that delivers the services. The Geoportal contains three components: Web Portal, Web services and Data Management. Table 2.1 describes the components, their relationships to one another and the standards and technologies they are built upon.

Table 1. Geoportal architecture (After Tait, 2005)

Components	Elements	Environments	Functions
Web Portal	Web site	HTML, HTTP, XML, XSL, JSP, ASP	Search, View, Publish, Admin.
	Web controls	Java beans, .NET	Query, Map, Edit
Web services	Geo Web services	XML, SOAP, WSDL, WMS, WFS, GML	QUERY, Render, Transaction
Data Management	RDBMS Data	SQL	Vector Raster Tabular

Metadata

SENSOR realised from the very beginning, that in order to build a strong spatial data infrastructure and to establish integrity and consistency of all data, metadata would be crucial. Metadata and metadata servers enable users to integrate data from multiple sources, organizations, and formats. Metadata for geographical data may include the data source, its creation date, format, projection, scale, resolution, and accuracy.

Metadata will be specified and assembled on the Internet to existing, international standards (ISO). As a matter of routine, INSPIRE data sets will be documented to facilitate their identification, proper management and effective use across the Community, and to avoid collecting or purchasing the same data more than once. To provide an accurate list of data sets held by local, regional, national and EU institutions, metadata catalogues will be compiled. This will include discovery level metadata about content, geographic extent, currency, and accessibility of the data, together with contact details for further information about the data.

Due to the fact that our end users are the European Commission, it seems reasonable to take outset in existing metadata standards within the Commission. At first, we therefore took a look on the metadata profile from the EEA (European Environmental Agency) as an initial metadata set. The EEA metadata profile builds on the principles in ISO 19115 as well as INSPIRE. Currently, a Metadata Core Drafting team is working on a detailed metadata specification for INSPIRE, and the Data Management group will follow the results of this work. Several discussions during the Vienna Meeting / Feb. 2005, and the Grenoble Meeting / June 2005, and further e-mail correspondence lead to a final selection of metadata attributes which are considered as necessary information for data retrieval, download and merging suitable for the purpose of generating European-wide indicators regarding quality of landscape multi-functionality.

The attribute set was reduced for SENSOR in order to increase acceptance among SENSOR data deliverers to fill out the forms completely. The metadata set shall fulfil all needs within the project to fully inform all team members about the content of the data sets. The metadata is furthermore a precondition to assess the usability of the respective data. Therefore some additional attributes, which are not considered by ISO 19115 –standard, but seem to the M5-team important, have been added. The most important among them are the fields regarding spatial entities, which contain thematic statistical information (e.g. demographic or economic data based on NUTS-Regions) and further content regarding spatial characteristics (e.g. land use classes, elevation, terrain shape, environmental pollution etc.). More details can be found in SENSOR deliverable 5.1.1.

SENSOR DATA MANAGEMENT SYSTEM DESIGN

The overall aim of the SENSOR Data Management System will be to support all partners get access to data from various sources as well as data produced within the SENSOR project. Finally the SENSOR Data Management System will be the main provider of NUTS based tabular data for the SIAT system. The first element in the SENSOR Data Management System is the Metadata Publishing System aiming at reporting metadata for data related to the SENSOR project. This reporting tool has already been available since late summer 2005. Parallel to the metadata reporting the application facilitates the upload of data to the central server. Closely related to the upload procedure is a checking tool for NUTS based tabular data. Finding and discovering spatial data will be provided through two different kinds of user interfaces. First there will be a web-based application for searching metadata. Second, there will be a searching application based on either the ArcIMS based Metadata Explorer or the Geoportal kit developed by ESRI.

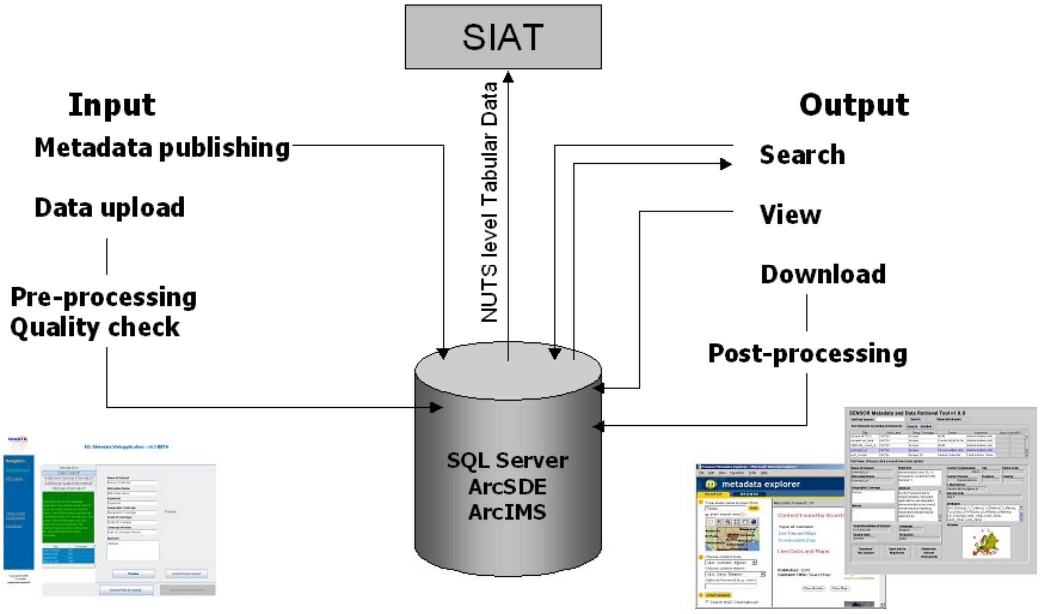

Figure 5. Principles for SENSOR Data Management System

As mentioned above the INSPIRE principles will be applied as an overall frame as well technically as organisational for the SENSOR Data Management System. However the technical foundation of INSPIRE is the Standards from ISO TC/211 and Open Geospatial Consortium.

SENSOR Data Warehouse

The main component in the SENSOR Data Management system will be the Data Warehouse storing pre-processed spatial data with associated metadata. All common data used in the SENSOR project as well as all data produced by SENSOR will be available from the Data Warehouse.

The Data Warehouse will be based on state-of-the-art database technology using ArcSDE 9.1 from ESRI. ArcSDE is an advanced data server, providing a gateway for storing, managing, and accessing spatial data in any of several leading RDBMS from any ArcGIS application. It is a key component in managing a shared, multi-user Geodatabase in a RDBMS. Currently ArcSDE support the following relational databases: Oracle, IBM DB2 Universal Database, IBM Informix Dynamic Server, and Microsoft SQL Server. Within SENSOR the underlying relational database system will be Microsoft SQL Server 2005, mainly because many partners already have SQL Server and thus experiences in using this platform.

Spatial data are stored in ArcSDE as either vector features or as raster data sets along with traditional tabular attributes. For example, an RDBMS table can be used to store a collection of features, where each row in the table represents a feature. A shape column in each row is used to hold the geometry of the feature. According to the Simple Feature specification from OGC the shape column holding the geometry is typically one of two column types:

- A binary large object (BLOB) column type
- A spatial column type

However, currently SQL Server only supports the Binary large object data type.

A homogenous collection of common features, each having the same spatial representation, such as a point, line, or polygon, and a common set of attribute columns is referred to as a feature class and is managed in a single table. Raster and imagery data types are managed and stored in relational tables as well. Raster data is typically much larger in size and requires an associated side table for storage. During the storage process the software cuts the raster into smaller pieces, called "blocks," and stores them in individual rows in the separate block table.

The column types that hold the vector and raster geometry vary from database to database. When the RDBMS supports spatial type extensions, the Geodatabase can readily use them to hold the spatial geometry (e.g., Oracle Spatial Type).

Topology—the spatial relationships between geographic features—is fundamental to ensuring data quality (ESRI, 2005; Silvertand, 2004). Topology in ArcSDE is implemented as a set of integrity rules that define the behaviour of spatially related geographic features and feature classes. Topology rules, when applied to geographic features or feature classes in a Geodatabase, enable GIS users to model such spatial relationships as connectivity and adjacency. Topology is also used to manage the integrity of coincident geometry between different feature classes – for example to chick if the coastlines and country boundaries are coincident. These rules are described in SENSOR Deliverable 5.1.2 and they will be applied to all data in the Data Warehouse. The SENSOR Core data set (Chapter 5) is already checked for topological errors.

The SIAT system will retrieve tabular information at NUTS level from the SENSOR Data Warehouse. This architecture will ensure a consistent delivering of data from the other modules to the SIAT system. The precise input data format to SIAT will be decided during late summer 2006, after the SIAT system specifications have been finalised. The most probable format will be DBF-files, which is an open and easy to handle file format.

Input – The SENSOR Metadata Publishing and Upload Application

The SENSOR Metadata Publishing System was build by ARCsys as the first part of the SENSOR Data Management System. Module 5 has developed the prototype of data upload/download Web based application. The purpose is to give the SENSOR community tools for uploading various NUTS related data to the Data Management System. This tool also allows exploring and visualising the data. Important feature of this application is automated data integrity checking.

On the front page of the application the user can select, whether he want to upload data to the system or explore existing datasets.

On the first step of uploading application the some necessary information will be asked from the user to fill the SENSOR metadata record. Also there is a link to a predefined SENSOR NUTS regions identification. This classification could be downloaded and used as reference information. On this page user also have to specify the file with data should be uploaded by the application.
Before the next page is coming some preliminary information check is applied:

- Mandatory fields (metadata) filled
- Uploading file format
- Data consistence (field names in the first row, are number of records in each row match to the number of fields, is there nonnumeric values for the data…).

In case of any error user will be informed and asked to proceed from the previous page. If checking is OK, next step will be performed.

On this page the additional information for the metadata record and data integrity check should be provided. Here a user should specify the NUTS resolution, and NUTS identification field, which assumed to be present in the uploading data. Also user can define a data, which will be displayed on the map for the "final" check.
At this stage of the application additional information check is provided:

- Are the mandatory fields (metadata) filled
- Is there already data with same metadata profile (the application queries the database)
- Does the NUTS identification field and NUTS resolution specified
- Is the NUTS identification key in each row valid (the application queries a special NUTS reference table in the database). It is very important for data integrity: as SENSOR data management system should be GIS oriented, we must be sure, that all data in the system could be linked to the corresponding map.

In case of any error user will be informed and asked to proceed from the previous page. If checking is OK, the last step will be performed.

At this stage the application displays a corresponded NUTS map where uploaded information can be displayed. To implement the "final" step of integrity checking user should to exam is the data linked to the map correctly.

Output – The SENSOR Geoportal and Data Retrieval Application

The general entrance to the system will be through the SENSOR Geoportal or the Data Retrieval Application, which is closely related to the Metadata Publishing System. The idea behind these applications is that the user can search and discover data in the Data Warehouse, explore their metadata and possibly select data for download. The Geography Network Explorer (fig. X) as well as the INSPIRE Geoportal are both examples on how to use an Internet Map Server based Geoportal for data searching, discovering and retrieval.

The SENSOR Geoportal will be based on an Internet Map Server and ArcExplorer, and the main competitors among Internet Map Servers are ArcIMS from ESRI and MapServer, which is an Open Source implementation. MapServer is free of charge and an open concept, with unlimited possibilities for developing targeted implementations. This is obviously an advantage. However, the implementation effort can be a rather tough job, because you have to develop much more by yourselves, and this is certainly a disadvantage. ArcIMS is a rather expensive product, but comes with built-in applications for administration and authoring as well as end user applications. But you still have the possibility to build your own end user application using Java. The choice between the two alternatives is at first sight not easy, but when we take outset in our existing environments, we have decided to use ArcIMS. NERI and ARCsys already have ArcIMS licenses, and all partners in the SENSOR project are using ESRI software.

ArcIMS works in a Java 2 environment, and requires a Web server, Java Virtual Machine and a servlet engine. Within SENSOR we use Apache. The Web server handles requests from a client based on HTTP (Hypertext Transfer Protocol) forwards the request to the appropriate application and sends a response back to the requesting client. The Java Virtual Machine provides the application-programming interface for running the Java 2 components of ArcIMS. The servlet engine is an extension to the Java VM and provides support for servlets through a servlet API.

Viewers

Viewers determine the functionality and graphic look of an ArcIMS Web site and offer tools for viewing and querying spatial and attribute data. ArcIMS provides three viewer choices:

- HTML Viewer

  The HTML Viewer, consisting of HTML and JavaScript, must be downloaded to the client. The HTML Viewer's functionality can be extended using a combination of Dynamic HTML (DHTML), JavaScript, XML, and other technologies. However, the HTML Viewer supports only Image Map Services. Image Map Services send a snapshot of the data in JPG, TIF, or

PNG format to the client. The data is not streamed as with Feature Map Services.

- Java Viewers

ArcIMS supplies two Java Viewers. The Java Custom Viewer uses Java applets to serve maps and information. A Java applet differs from a Servlet. It runs on the client, not the server, and must be downloaded to the client. Consequently, Java clients are thicker than the other viewers. To view a Web site that uses a Java Viewer, the user must initially download two plug-ins. Both Java Viewers can serve Image and Feature Map Services. Feature Map Services use data streaming, which allows user interaction and analysis.
Neither the tools nor the format of the Java Standard Viewer can be customized. The Java Custom Viewer can be customized through HTML and scripting to the applets using JavaScript. Because Netscape does not support applet scripting, the Java Custom Viewer will not work in Netscape browsers.

Map services
The OGC WMS connector produces maps of geo-referenced data in image formats (PNG, GIF, JPEG) and creates a standard means for users to request maps on the Web and for servers to describe data holdings.

The OGC WFS connector enables ArcIMS to provide Web feature services that adhere to the OpenGIS Web Feature Service Implementation Specification. The connector provides users with access to geographic (vector) data, supports query results, and implements interfaces for data manipulation operations on Geographic Mark-up Language (GML) features served from data stores that are accessible via the Internet. GML is an OpenGIS Implementation Specification designed to transport and store geographic information, and it is a encoding of Extensible Mark-up Language.

The main development environments for the SENSOR Geoportal are Java and ArcXML, which is the protocol for communicating with the ArcIMS Spatial Server (ESRI, 2002).

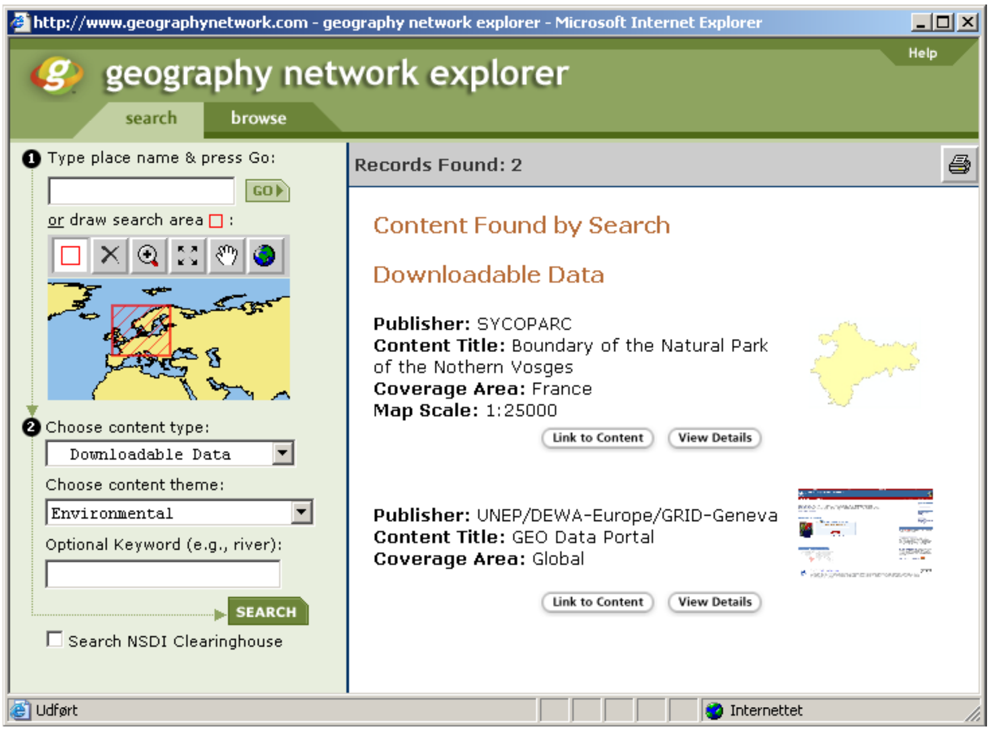

Figure 6. Geography Network Explorer – an example.

Spatial Data Mining

The immense amount of geographically referenced data occasioned by developments in digital mapping, remote sensing, and the global diffusion of GIS emphasises the importance of developing data driven inductive approaches to geographical analysis and modelling to facilitate the creation of new knowledge and aid the processes of scientific discovery (Openshaw, 1999). Spatial data mining aims to uncover spatial patterns and relations.

Data mining has the potential to equip users with extended analytical capabilities that can enable them to discover non-obvious relationships between datasets. By augmenting data discovery tools with spatial data mining, it is envisaged that users will discover related datasets that they would have otherwise overlooked. A major challenge for this final part of the SENSOR Data Management implementation is therefore to do research and development on effective methods for determining spatial and non-spatial relationships between datasets. The tools will be based on recent advances in spatial data mining and knowledge discovery as described by Ester, Kriegel and Sander (2001) and facilitate spatial characterisation, spatial clustering and spatial trend detection.

DATA POLICY

The data policy covers aspects of data mining (clearing house), data availability, data access, ownership, licensing, and Intellectual Property Rights (IPR) on the data used in the frame of the SENSOR project. The conditions for uploading data to and downloading data from the SENSOR Data Warehouse need to be detailed. The SENSOR data policy should follow the principles to be developed under the INSPIRE initiative. Currently, however, only a position paper on 'Data Policy and Legal Issues exists, which lacks relevant details. As a consequence the SENSOR data policy has been developed as a consensus among the SENSOR partners, following the indications given in the INSPIRE position document. It might need revision when more detailed guidelines become available under the INSPIRE initiative.

Following these principles, it will be important that all data used and generated in the frame of SENSOR are well documented following strictly the SENSOR metadata profile and that the relevant search facilities are available. It is further of importance that all data are available to the whole SENSOR community under clear conditions. Questions of data ownership, copyrights and conditions of have now been clarified in order to encourage the disclosure and upload of data available as well as their widespread use within the SENSOR community.

Upload policy

All partners are encouraged to upload metadata on data of common interest and possibly to upload the data themselves. The uploading institution will retain the ownership of the data and will specify the conditions of use of the data. For any dataset to be uploaded, a copyright statement must be included in the metadata.

By uploading the data, the data provider (owner) agrees that all SENSOR partners have free access to the data for their work within the SENSOR project. If not explicitly specified otherwise, all other uses will have to be authorised. It is strictly forbidden to deliver data to third parties outside the SENSOR project or to use the data for purposes outside the SENSOR project without the written consent of the data owner. Inquiries from third parties should be transferred to the data owner for clarification.

All datasets must be accompanied by metadata, which follow the SENSOR metadata profile, which complies with the ISO 19115 standard. Without a full set of mandatory metadata, the data will not be accepted. Metadata will be freely available also for further (public) distribution.

Data sets can be uploaded once the metadata are completely available and the data policy and copyright agreement has been accepted.

Download policy

All SENSOR partners have full access to the metadata system, where they can search for data and information on the conditions of their use. Available datasets can be downloaded for use within the SENSOR project. Before downloading the data, the user agrees on the conditions of use of the data (data policy and copyright agreement).

Data formats

On the outset of the project, it is not yet fully clear what data types can be expected from (or are needed by) the different modules. In principle, however, Module 5 expects georeferenced data (vectors, polygons, grids, points) and statistical data referenced to geographical entities (e.g., administrative regions).

Data submitted to the Data Management System should follow certain standards. XML is emerging as the international standard for exchange of information, and you ca easily import and export XML data in most modern GI software systems like ArcGIS. However due to the often huge size of geographic data sets, XML has had limited success in the GI Community. Instead native data formats from vendors like ESRI are used. In the SENSOR project, data should be exchanged in one of the following formats:

- ESRI Shapefiles
- ESRI Personal Geodatabases
- Erdas Imagine or TIFF
- ESRI Coverages and Grids via Exchange File Format (E00)
- XML
- Tabular data (e.g., statistics for administrative regions). These data need to be linked to a geographic entity via a common feature code.

Examples of data types, which will not be accepted, are the following:

- Text documents
- Newspaper clippings
- All data not referenced geographically

The Module 5 Data Management (M5 DM) team strongly recommends that data are provided in one of the formats listed above. If this should prove to be impossible for certain cases, the M5 DM team can try to help to solve this problem. However, we underline that this should not be the rule and that in principle it remains the task of the different modules to provide data in an easily manageable and acceptable format.

In principle, SENSOR data should comply with INSPIRE recommendations. This implies that data should be provided in a compliant reference and projection system (i.e. ETRS89 specifications (Annoni et al, 2003) and that grids should follow the INSPIRE grid specifications (JRC, 2003). This is very important in order to make these data readily available and useable for different applications. In case partners should have problems to convert the data, the M5 DM team can try to help to solve the problem, provided that the data provider is able to give a detailed and accurate description of the projection system of the data. However, we underline that this should not be the rule and that in principle it remains the task of the different modules to provide data in the correct projection system.

SENSOR CORE DATA

The INSPIRE Working Group on Reference Data and Metadata encourage establishing a reference or core data set as in instrument to harmonise data from various sources. The recommendations from this group were

- Geodetic reference data
- Units of administration
- Units of property rights (parcels, buildings)
- Addresses
- Selected topographic themes (hydrography, transport, height)
- Orthoimagery
- Geographical names

During the further work with INSPIRE, the reference data set was changed a little bit – now also including European Grid in the so-called annex 1 data (COM, 2004). Within SENSOR we have chosen geodetic reference system, administrative boundaries in form of NUTS, European Grid, CORINE Land cover, LANMAP and the European Digital Elevation model as our reference data set. By defining a SENSOR core data set we encourage partners to use for example the same NUTS map – although many different versions are available. Concerning the role as data harmonisation, the reference system / projection, NUTS boundaries and the European Grid play the most important role. Those data are described in detail below.

Reference system and projections

The European Terrestrial Reference System 1989 (ETRS89) is the geodetic datum for pan-European spatial data collection, storage and analysis. This is based on the GRS80 ellipsoid and is the basis for a coordinate reference system using ellipsoidal coordinates. For many pan-European purposes a plane coordinate system is preferred. But the mapping of ellipsoidal coordinates to plane coordinates cannot be made without distortion in the plane coordinate system. Distortion can be controlled, but not avoided.

For many purposes the plane coordinate system should have minimum distortion of scale and direction. This can be achieved through a conformal map projection. The ETRS89 Transverse Mercator Coordinate Reference System (ETRS-TMzn) is recommended for conformal pan-European mapping at scales larger than 1:500 000. For pan-European conformal mapping at scales smaller or equal 1:500 000 the ETRS89 Lambert Conformal Conic Coordinate Reference System (ETRS-LCC) is recommended.

With conformal projection methods attributes such as area will not be distortion-free. For pan-European statistical mapping at all scales or for other purposes where true area representation is required, the ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System (ETRS-LAEA) is recommended.

The ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System (ETRS-LAEA) is a single projected coordinate reference system for all of the pan-European area. It is based on the ETRS89 geodetic datum and the GRS80 ellipsoid. Its defining parameters are given in Table 1 following ISO 19111 Spatial referencing by coordinates.

With these defining parameters, locations North of 25º have positive grid northing and locations eastwards of 30 º West longitude have positive grid easting. Note that the axes abbreviations for ETRS-LAEA are Y and X whilst for the ETRS-LCC and ETRS-TMnz they are N and E.

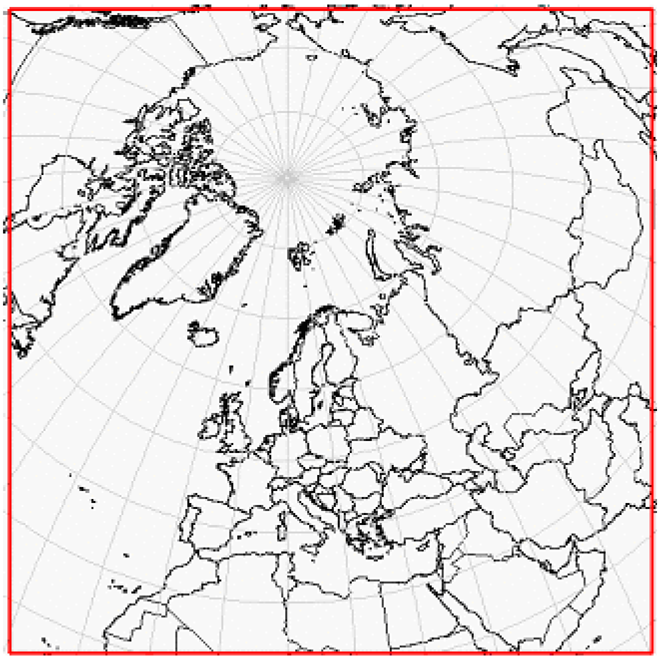

Figure 7. The Azimuthal Equal Area projection

NUTS

EuroStat established the Nomenclature of Territorial Units for Statistics (NUTS) more than 25 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. The NUTS classification has been used since 1988 in Community legislation. But only in 2003, after 3 years of preparation, a Regulation of the European Parliament and of the Council of NUTS was adopted. From 1. May 2004, the regions in the 10 new Member States have been added to the NUTS.

The NUTS nomenclature is defined only for the 25 member states of the European Union. For the additional countries comprising the European Economic Area (EEA) and also for Switzerland, a coding of the regions has been accomplished in a way, which resembles the NUTS.

Principles of the NUTS nomenclature
a) The NUTS favours institutional breakdowns.
Different criteria may be used in subdividing national territory into regions. These are normally split between normative and analytic criteria:

- Normative regions are the expression of a political will; their limits are fixed according to the tasks allocated to the territorial communities, according to the sizes of population necessary to carry out these tasks efficiently and economically, and according to historical, cultural and other factors;
- Analytical (or functional) regions are defined according to analytical requirements; they group together zones using geographical criteria (e.g., altitude or type of soil) or using socio-economic criteria (e.g., homogeneity, complementarity or polarity of regional economies).

For practical reasons to do with data availability and the implementation of regional policies, the NUTS nomenclature is based primarily on the institutional divisions currently in force in the Member States (normative criteria).

b) The NUTS favours regional units of a general character.
Territorial units specific to certain fields of activity (mining regions, rail traffic regions, farming regions, labour-market regions, etc.) may sometimes be used in certain Member States.
NUTS exclude specific territorial units and local units in favour of regional units of a general nature.

c) The NUTS is a three-level hierarchical classification
Since this is a hierarchical classification, the NUTS subdivides each Member State into a whole number of NUTS 1 regions, each of which is in turn subdivided into a whole number of NUTS 2 regions and so on.

The NUTS map in SENSOR is based on SABE (Seamless Administrative Boundaries in Europe), which is an official product developed by EuroGeographics. The data behind SABE is the official administrative boundaries prepared by the national mapping agencies. The scale is generally 1: 100,000.

Aiming at a more equal size of the NUTS-3 polygons, we have decided to develop a special version called NUTSx, where NUTS-3 is the baseic map, but some countries – e.g. Germany is represented by NUTS-2.

Because NUTS plays such an important role in SENSOR, we have decided to develop a generalised version with much less detail than the SABE based NUTS map. However, remark that the generalised version is only for visualisation purposes. For geospatial analysis and modelling it is recommended to use the detailed NUTS version.

European Grid

The Grid is based on an equal area projection. The European grid should be used mainly for European purposes, but it can be useful also for national purposes. The datum to be used is ETRS89 as previously identified by INSPIRE. The geographical location of the grid points is based on the Lambert Azimuthal Equal Area coordinate reference system (ETRS-LAEA). The cartographic projection is centred on the point N 52°, E 10°. The coordinate system is metric. The Grid is based on the projection system Lambert Azimuthal Equal Area (ETRS-LAEA). The square shape will appear when used in the defined projection, smaller or larger distortions will appear when re-projected to other projections.

Naming the individual cells can be done in several ways, but in SENSOR we have decided to use the so-called Direct Coordinate Coding System (DCCS), which concatenates the coordinates of Easting and Northing of a grid point. The length of the coordinates defines the precision of the grid. A grid with a precision of 1 m would require a maximum of 7 digits by each dimension. The resulting code would have 14 digits. A grid with a precision of 1 km would be defined by a code comprising 8 digits. Leading zeros are coded in order to preserve the precision information. Grid code identifies southwestern corner of a cell.

CONCLUSIONS

The SENSOR Data Management System will provide state-of-the-art core functionality for uploading data and metadata, storing data, searching and exploring data, selecting and downloading data. Use of standard off the shelf software complying with international standards like W3C, ISO TC/211 and the OGC will be the implementation platform.

When we talk about SENSOR Data Management we actually mean SENSOR Spatial Data Infrastructure dealing with all aspects of data management. Thus not only the technical aspects are included but also the economic and legal dimensions of data are addressed. The SENSOR community has agreed on the important topic of data sharing between partners.

The first part of the Data Management System was already developed during summer 2005 as part of the first deliverable (5.1.1). This first component comprises the SENSOR Metadata Publishing system, and closely related to this is the data upload application, which still is under improvement. This data upload application could play an important role in establishing, at some level, data harmonisation and integrity.

The second part of the system will be the implementation of the Data Warehouse with attached SENSOR Geoportal for searching, exploring, selecting and downloading data. During this second phase, the connections between the Data Management system and SIAT will be established.

The third part of the system will deal with the development of tools for Spatial Data Mining and necessary pre and post processing tools. Data mining has the potential to equip users with extended analytical capabilities that can enable them to discover non-obvious relationships between datasets. By augmenting data discovery tools with spatial data mining, it is envisaged that users will discover related datasets that they would have otherwise overlooked. A major challenge for this final part of the SENSOR Data Management implementation is therefore to implement do research and development on effective methods for determining spatial and non-spatial relationships between datasets.

Generally speaking, among the process of developing the overall design of the SENSOR Data Management system, some “working” prototypes of different parts mentioned above have been developed. The main task for the nearest future we see in the bringing them together and establishing the integrated system.

BIBLIOGRAPHY

Annoni, A., C. Luzet, E. Gubler, J. Ihde(Eds)(2003): Map Projections for Europe. EUR 20120 EN. (http://www.ec-gis.org/document.cfm?id=425&db=document)

Bernard, L., Kanellopoulos, I., Annoni, A. & Smits, P. (2005). The European geoportal—one step towards the establishment of a European Spatial Data Infrastructure. Computers, Environment and Urban Systems, vol. 29, pp. 15 – 31.

COM (2004). Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing an infrastructure for spatial information in the Community (INSPIRE)

ESRI (1998). Spatial Data Warehousing – An ESRI White Paper. Redlands, USA.

(ESRI, 2002). ArcXML Programmers Reference Guide. Redlands, USA

ESRI (2005). ArcGIS 9 – Building a Geodatabase. Redlands, USA.

Ester, M., Kriegel, H.-P. and Sander, J. (2001). Algorithms and Applications for Spatial Data Mining. Geographic Data Mining and Knowledge Discovery. Research Monographs in GIS. Taylor & Francis.

INSPIRE (2002 a). INSPIRE Architecture and Standards Working Group Position Paper.

INSPIRE (2002 b) INSPIRE Data Policy and Legal Issues Working Group Position Paper, October 2002. http://inspire.jrc.it

Klopfer, M. (Ed.) (2006). Interoperability & Open Architectures: An Analysis of Existing
Standardisation Processes & Procedures. OGC White Paper. Open Gepspatial Consortium 2006.

Mücher, C.A., Bunce, R.G.H., Jongman, R.H.G., Klijn, J.A., Koomen, A.J.M., Metzger, M.J., and Wascher, D.M., 2003. Identification and characterisation of environments and landscapes in Europe. (Alterra-rapport 832), Alterra, Wageningen (NL). 119pp

Openshaw, S. (1999). Geographical data mining: key design issues. Proc. Conference on GeoComputation, http://www.geovista.psu.edu/sites/geocomp99/Gc99/051/gc_051.htm

JRC, (2003). Proceedings of the 1st European Reference Grid workshop and Proposal for a European Reference Grid coding system, 15-10-2004, JRC ESDI Action. (http://inspire.jrc.it)

Silvertand, G. (2004) Storing and Maintaining Topology in an ArcSDE Geodatabase.. ArcUser July – September 2004, p. 36.

Tang, W. and Selwood, J. (2005). Spatial Portals – Gateways to Geographic Information. ESRI Press.