A Java Tool for Creating ISO/FGDC Geographic Metadata

F. Javier Zarazaga-Soria, Javier Lacasta, Javier Nogueras-Iso,
M. Pilar Torres, P.R. Muro-Medrano

Computer Science and Systems Engineering Department
University of Zaragoza

{javy | jlacasta | jnog | mtorres | prmuro}@unizar.es

ABSTRACT

Metadata are "data about data", that is to say, they describe the content, quality, condition, and other characteristics of data in order to help a person to locate and understand data. Nevertheless, maybe one of the main problems for launching a Spatial Data Infrastructure (SDI) is to have appropriate and well-defined contents for its catalogues, that is to say metadata. The creation of metadata is an arduous labour that must be facilitated by the adequate tools. At the University of Zaragoza, a Java tool for creating geographic metadata has been developed. This SDI component enables the different agents of a spatial data infrastructure to create consistent metadata describing geospatial data resources. It is being used in some SDI initiatives as the reference tool for the metadata-creation process. This paper presents this tool functionality.

INTRODUCTION

Metadata are "data about data", that is to say, they describe the content, quality, condition, and other characteristics of data in order to help a person to locate and understand data. The creation of metadata has three major objectives (FGDC 2000). The first one is to organize and maintain an organization's investment in data. As personnel change or time passes, information about an organization's data will be lost. And later workers may have little understanding of the content and uses for a digital database and may find that they cannot trust results generated from these data. Complete metadata descriptions of the content and accuracy of a geospatial data set will encourage appropriate use of the data. Such descriptions also may provide some protection for the producing organization if conflicts arise over the misuse of data. The second objective is to provide information to data catalogues and clearinghouses. Applications of geographic information systems often require the integration of data from different thematic sources. And few organizations can afford the creation of all data they

need. Furthermore, data created by an organization also may be useful to others. By making metadata available through data catalogues and clearinghouses, organizations can find data to use, partners to share data collection and maintenance efforts, and customers for their data. Finally, third objective of metadata is to provide information to aid data transfer. Metadata should accompany the transfer of a record. The metadata aids the organization receiving the data process and interpret data, incorporate data into its holdings, and update internal catalogues describing its data holdings.

Nevertheless, maybe one of the main problems for launching a Spatial Data Infrastructure (SDI) is to have appropriate and well-defined contents for its catalogues. The creation of metadata is an arduous labour that must be facilitated by the adequate tools.

At the University of Zaragoza, a Java tool for creating geographic metadata has been developed. This SDI component enables the different agents of a spatial data infrastructure to create consistent metadata describing geospatial data resources. Apart from the basic metadata-creation functionality, the component is enhanced with a set of tools to improve the quality of metadata: a thesaurus management tool and an automatic metadata generation tool. The first tool enables metadata originators to use look-up tables in order to fill some metadata elements with predefined lists of controlled keywords. The use of these controlled keywords facilitates the mapping between a selected vocabulary and a large collection of records. This way, the catalogue search tools make possible the discovery of data based on hierarchies of concepts. The thesaurus management tool integrates a relational database storing classifications provided by recognized authorities (some examples are found in UNESCO 2003, NASA 1996 or ADL 2003). The second optional tool is an automatic metadata generation tool which is able to derive metadata from the data sources by means of interconnection with commercial GIS tools or proprietary software. Examples of derived metadata are information about spatial reference systems, number and type of geographic features, extension covered by a record, or information about the entities and attributes of alphanumerical related data.

This application may work with a “cheap” database, like Access or mySQL (the software uniquely requires access by means of JDBC to any Relational Data Base Management System), which is responsible for the storage of the metadata entries using a SQL-92 metadata database model.

APPLICATION FUNCTIONALITY

The application integrates a series of components, each one making an independent task with respect to the others, which rely on a generic software kernel offering services of graphical visualization and communication between components. The components contained in the application and their functionality will be shown in following sections.

Metadata Edition tool

The *Metadata Edition* tool, as its own name indicates, enables edition and visualisation of metadata entries in our catalog. The metadata supported by this tool are based on ISO/DIS 19115 standard (ISO 2001), but it also allows the separate edition of a reduced subset of ISO/DIS 19115 metadata elements that corresponds with the Dublin Core Metadata Element Set (Dublin Core 2003). Besides, this tool allows the replication of metadata records to create new ones, thus avoiding the waste of time to fill again the fields if the new record is similar to an existing one. Its use is very simple; there is a window (*Record Selector* tool) where the list of records (stored in the catalog) is shown with some information to identify each record, e.g. title, reource type, creator, editor or security classification. With this tool, the user can create, delete and replicate records, but not modify the content of metadata element fields. The edition of metadata records is done with the *Detailed Editor* or *Dublin Core Editor* tool.

When the user selects a record in *Record Selector* Tool window, the Dublin *Core Editor* and *Detailed Editor* windows enable the user to modify the metadata content of the selected record. Although there are two edition tools, both modifying the selected record, the only difference is that the *Dublin Core Editor* only edits a subset of the metadata elements available at the *Detailed Editor* tool. The purpose of *Dublin Core Editor* is to facilitate a minimum cataloguing, while the *Detailed Editor* is thought for an specific and detailed use.

The *Dublin Core Editor* shows a graphic user interface that enables the edition of metadata elements according to Dublin Core specification. Apart from Dublin Core fifteen elements, this editor also includes other metadata elements that we consider of interest for a minimum cataloguing of resources: bounding box, scale, publication place, metadata record status, metadata and data security information and metadata creation date.

The *Detailed Editor* shows instead the metadata indicated in the specifications of ISO/DIS 19115 standard. The following packages of the standard have been included in the *Detailed Editor*:

- Identification Information
- Data Quality Information
- Spatial Representation Information
- Reference System Information
- Content Information
- Distribution Information
- Metadata entity Set Information
- Citation and Responsible party Information

The tool does not cover completely all these sections but have the main elements of each one; it covers all elements included in the subset "*Core metadata for Geographic datasets*", and all mandatory but not all of the optional elements included in the "*Comprehensive dataset metadata application profile*". Some sections like *Identification Information* are almost completely covered, while others (e.g. *Spatial Representation Information* section) include only the minimum required.

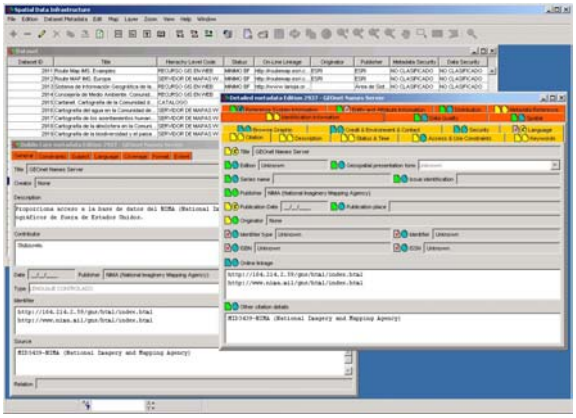

Fig. 1: Record Selector, Detailed Editor and Dublin Core Tools

ISO/DIS 19115 specifies whether a metadata element must contain free text or a controlled vocabulary. Therefore, the editor tool has followed these constraints and provides controlled lists or thesaurus browsing facilities when needed.

Thesaurus Management tool

The *Thesaurus Management* tool permits the management of thesauri supported by our Thesauri database. The main functions of this tool are: Creation/Deletion/Modification of thesauri; Edition/Visualisation of terms in a hierarchical and alphabetical structure; and Import/Export from/to text files in different formats.

The content of some metadata fields (keywords) is mainly oriented for discovery purposes. Therefore, it is important to have the content of these fields under control. However, a simple controlled list of values is not versatile enough to select the content of these fields. Therefore, we decided to use thesaurus as selection structure and this tool was created to facilitate the work with this kind of hierarchical structure.

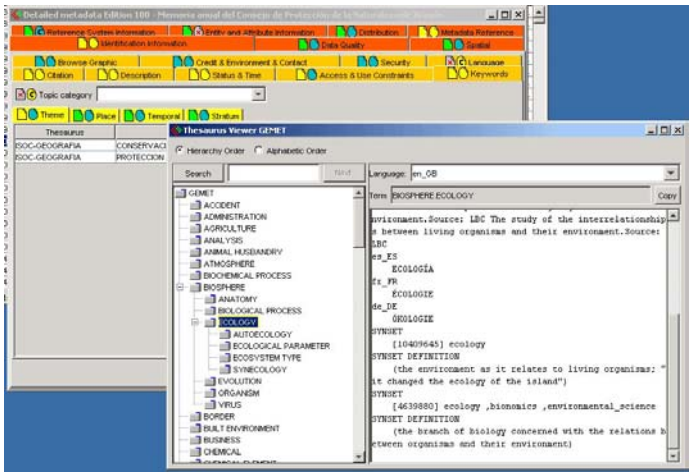

Fig. 2: Term Selector Tool

The tool is completely integrated as an independent component in the cataloguing tool. It is able to send to the thesaurus term that has been selected, so that the metadata edition tool can include it in the corresponding keyword field. The process of sending the term from a component to other one is performed by means of a simple copy-paste mechanism. The functionality of this tool is described in detail in (Nogueras-Iso, Bañares et al. 2003), also presented in this conference.

XML Import/Export tool

The *XML Import/Export* tool enables the exchange of metadata records in XML format (tagged plain text files) conforming to different standards such as CSDGM (FGDC 2000), ISO/DIS 19115 and Dublin Core Metadata Element Set. Besides, the tool also facilitates more readable presentations of metadata records in HTML format, e.g. English and Spanish FAQ, ESRI, and Geography Network style presentations.

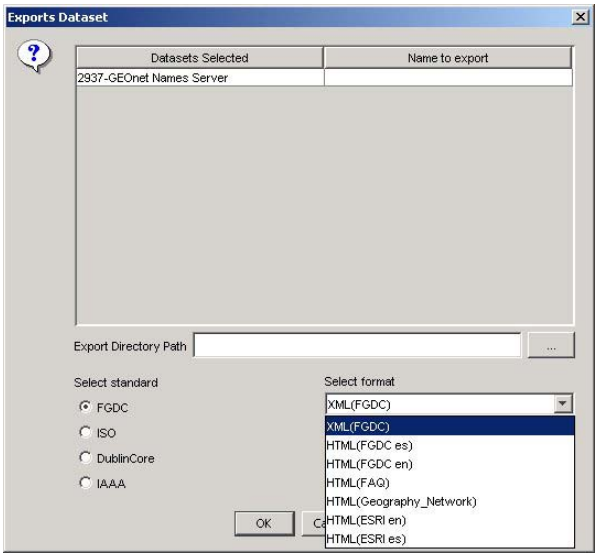

Fig. 3: XML Export Tool

The export tool is able to export the record selected in the *Edition Tool* to XML format following the structure imposed by the standards. The circumstance that the export format conforms to an aggreed standard facilitates the understanding and interoperation with other applications making use of metadata.

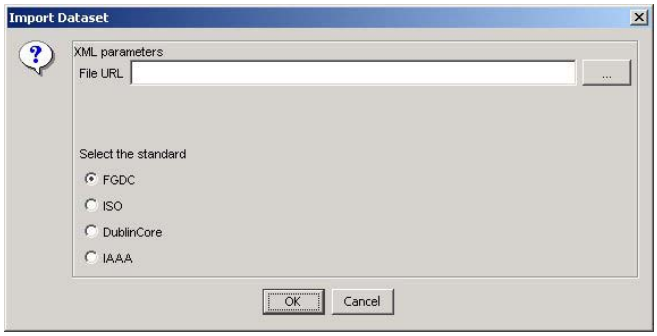

Fig. 4: XML Import Tool

The import tool can import XML files compliant with the same standards as the ones provided in the export tool. The import process has two possibilities: it can either create a new metadata record, or update the content of a metadata record previously selected in the *Record Selector* window.

Metadata Generation Tool

The *Metadata Generation* tool enables the semi-automatic generation of metadata for several types of resources. For instance, this tool is able to obtain descriptive information from ESRI SHAPE files (format typically used by ArcView). Additionally, the tool can also extract metainformation corresponding to the relational structure of tabular sources (e.g. Excel, Access, Oracle…).

Taking into account that ISO/DIS 19115 standard defines more than 300 metadata elements, it can be assured that the creation of appropriate content for the different metadata records is a hard and arduous process. Therefore, a tool that can extract metadata automatically from the data resources results very helpful. Apart from saving time in the cataloguing process, it prevents users from making frequent typing mistakes.

As concerns ESRI SHAPE files, the tool extracts the following metadata elements: bounding box, number and type of geometric objects, URL of the SHAPE file, and the relational structure of the associated attribute information. These metadata elements belong to *Identification Information*, *Spatial Representation Information* and *Content Information* sections.

Finally, regarding tabular sources, the tool extracts information about tables (name and constraints) and columns (name, constraints, and domain type);

and then, it maps this information to the metadata elements in *Content Information* section.

Collection Metadata Edition Tool

The *Collection Metadata Edition* tool enables the edition and visualization of metadata describing collections of datasets (later called units) that can be considered as a unique entity. Some examples of such collections are spatial aggregations, like mosaics of aerial images, which arise as a result of the fragmentation of geographic resources into datasets of manageable size and similar scale. Other example could be temporal series which aggregate geographic resources taken in similar conditions but at different instant times. The objective of this tool is to manage jointly the metadata at collection level (shared by the all the units in the collection) and the specific characteristics of each unit (Bejar, Muro-Medrano et al. 2003).

Once the user introduces the value "*series*" for the *hierarchy level* metadata element, the user is allowed to edit the *Aggregation Information* window that contains the extended meta-information to describe the characteristics of the collection (at present, only spatial aggregations are considered). In this *Aggregation Information* window the user will define: the metadata elements that may be overwritten for each unit; and the rules to combine metadata at the aggregated level and the specifics of each unit (in order to obtain complete metadata descriptions for units integrating the collection). The enumeration of the specific metadata elements of each unit and the merging rules are encoded in XML and typical templates for spatial aggregations are provided foreseen. In addition to this, the user must select the spatial pattern that explains the spatial distribution of units in the collection. This spatial pattern can be expressed, for instance, by means of a SHAPE file containing the polygons that correspond to the spatial extent of each possible unit in the collection. Furthermore, these spatial patterns are typically reused for different collections. Frequently, National Geographic Institutes define spatial distribution patterns for core geographic data at different scales, thus providing an established numbering and bounding box for units (also called tiles). Moreover, this *Collection Metadata Edition* includes a visualization tool that enables the supervision of the status of cataloguing. The tool generates a GML layer whose geographic features represent the metadata for each unit as follows: the geographic location of the feature corresponds to the bounding box of a unit; and the rest of feature attributes encode the specific metadata of each unit. Thanks to this, this tool can provide an approximate view of what has been already catalogued. In addition to this, through the visualization tool, it is possible to modify the

particular metadata of each unit, which is displayed as the attribute information contained in the GML feature.

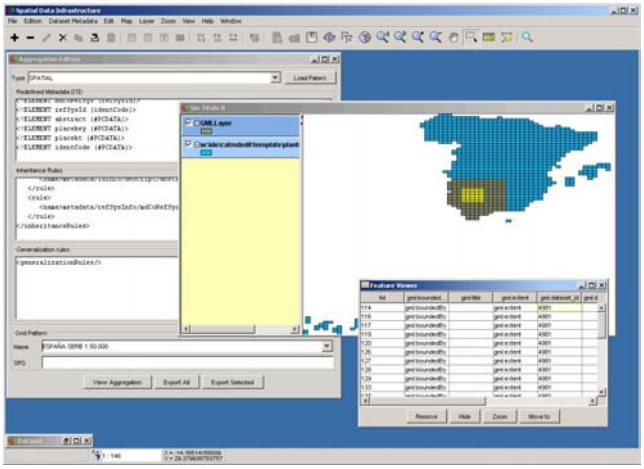

Fig. 5: Collection Metadata Edition Tool

The tool also enables the XML exchange of these extended metadata records to give support for collections. The XML generated is an extension of usual XML format generated *Import/Export* tool but including the aforementioned GML that encodes the specific metadata elements for the units. With this tool it is also possible to generate complete metadata descriptions of units as if we had created separate metadata records. The advantage of this approximation is the avoidance of metadata replication. Only a few metadata elements must be revised for each unit and this is particularly relevant if the size of the collection is quite large (e.g. a collection composed of thousand of files).

Metadata Validation Tool

The *Metadata Validation* tool facilitates the validation of metadata elements inserted by the user, i.e. checking the conditionality (optional, mandatory or mandatory-if-applicable) of metadata elements and reminding what mandatory elements have not been filled in. The verification can be done according to the standards FGDC and the Dublin Core Metadata Element Set.

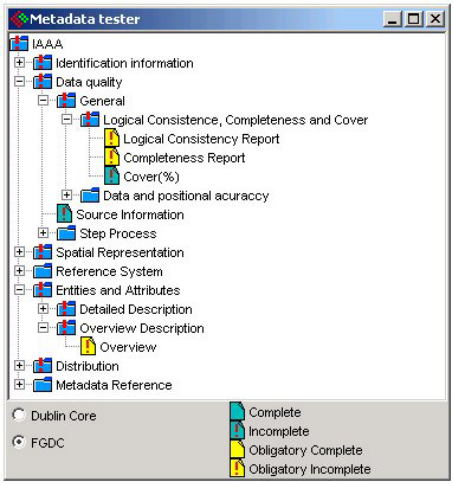

Fig. 6: Metadata Validation Tool

The geographic metadata standards such as CSDGM or ISO/DIS 19115 define hundred of metadata elements, hierarchically structured. Thus, when someone edits a record, he usually ignores whether he has completed all metadata elements defined as mandatory or not. This tool aims at overcoming this limitation, showing the hierarchical structure of the record and indicating with an advertence icon the incomplete elements. It has also a colour code to distinguish between mandatory elements and optional. The mandatory-if-applicable elements have been considered as optional because it is not possible to know automatically if they are applicable or not.

To determine if a section has been completed, the tool has only into account the mandatory items and others subsections. Therefore, if all subsections and mandatory items contained in a section have been filled, the broader section is complete. Otherwise the section is considered as incomplete. For this completeness test, it does not mind the status of optional elements.

Contact Management Tool

The *Contact Management* tool permits the reuse of contact information (e.g. name, address, telephone…), which is needed in several metadata fields. Thanks to this tool, the contact information about a person is only inserted once and used whenever it is required.

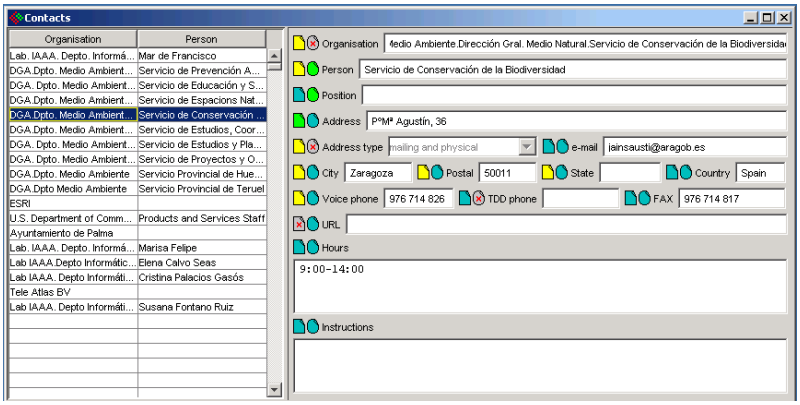

Fig. 7: Contact Management Tool.

The use of this tool is very simple. When the user needs to fill contact information inside a metadata record, instead of creating a new contact, the user verifies first whether the contact has been already created through the *Catalog Management* tool or not. If this contact already exists, the user simply includes a reference to the existent contact. Otherwise, the user will access the *Catalog Management* tool to create the new contact and later will include the appropriate reference. This tool facilitates consistency since contact information is modified uniquely in one place. Once you have updated the contact information, all the metadata records referencing it will be automatically updated.

User Control System

The *Users Control* system prevents users without the required permission from using the application. As most typical control systems, it asks for user id and a password, validates them, and returns a session identifier as a result if everything was correct. Then, this session identification is used in each operation that implies access to database to verify whether the operation can be executed or not.

Licence Control System

The *Licence Control* System enables the control of the distributions of the tool, by means of a licence that limits the days of use of the tool and the number of users that can use it simultaneously. For instance, this easy-configuration functionality allows SDI coordinators to distribute this appli-

cation among their partners and clients with marketing information and constrained access (e.g. evaluation license with limited functionality).

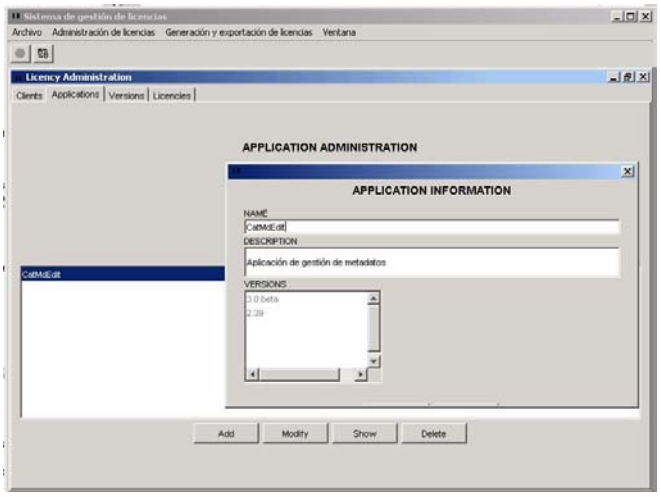

Fig. 8: Licence Administrator Tool

This system allows the creation of licenced components that may be later integrated into different applications. Apart from delivering licences, the system also aims at controlling the different distributions, versions and types of users of each product; thus enabling customized licenses according to each version and user needs.

USE EXPERIENCE

This application is being used by the Department of Environmental from the government of Galicia (Xunta) in the construction of its Spatial Data Infrastructure. It is also being used by the Ebro River Basin (CHE, http://www.chebro.es) for cataloguing their geographic information. And additionally, this tool will be used by the Spanish Geographic Institute (IGN, , http://www.ign.es) for maintaining their metadata.

Currently, it is being evaluated its adequacy for the Spatial Data Infrastructure of Spain and it is also in the selection process made by the Global Spatial Data Infrastructure (GSDI, http://www.gsdi.org) to adopt a reference implementation for the cataloguing of geographic information.

FUTURE WORK

We are actually working in the improvement of the metadata edition tool in order to manage multilingual versions of metadata records. As a first step, the goal is to store numeric codes for all metadata elements constrained to a controlled list of values and make possible the visualization and exchange of these values as text keywords in the desired language. And in addition to this, we are also concerned with the problem of storing multilingual content for free text metadata elements. The objective is to automate as much as possible the management and synchronization of multilingual versions of the same metadata record.

Other area of improvement is the *Collections Metadata* tool. We are improving this tool to support metadata for temporal series and collections of collections (collections whose units are also aggregations of datasets). The final purpose is to obtain a unified model than can be used for individual metadata records, records describing collections, or records describing collections of collections.

The last area of improvement in which we are working is the *Thesaurus Management* tool. The objective here is to enhance its disambiguation capabilities, so that catalogs can improve the precision and recall of their information retrieval algorithms.

ACKNOWLEDGEMENTS

The basic technology of this work has been partially supported by the Spanish Ministry of Science and Technology through the project TIC2000-1568-C03-01 from the National Plan for Scientific Research, Development and Technology Innovation, and by the project P089/2001 from the Aragón Government. Javier Lacasta's work has been partially supported by a grant from the Aragón Government and the European Social Fund.

REFERENCES

ADL (2003). *Homepage of the Alexandria Digital Library Project,* http://www.alexandria.ucsb.edu (last accessed, May 2003).

Béjar, R., P.R. Muro-Medrano, J. Nogueras-Iso, F.J. Zarazaga-Soria (2003): *Facilitating Uniform Intercommunity Natural Risk Management Through Metadata for High-Level Geographic Aggregations*. Proceed-

ings of the 6th AGILE Conference: 663-672. Lyon (France), 24–26 April 2003.

Dublin Core (2003): *Homepage of the Dublin Core Metadata Initiative: http://www.dublincore.org* (last accessed, May 2003).

FGDC (2000): *Content Standard for Digital Geospatial Metadata Workbook, version 2.0*. Federal Geographic Data Committee (USA), 2000.

ISO (2001): *Draft International Standard ISO/DIS 19115, Geographic information — Metadata*. ISO/TC 211 (http://www.isotc211.org).

NASA (1996): *Draft geospatial thematic keywords from the NASA Master Directory in short and long format for CSDGM of FGDC*. Version available from http://www.fgdc.gov/clearinghouse/reference/refmat.html (last accessed, May 2003).

Nogueras-Iso, J., J.A. Bañares, J. Lacasta, J. Zarazaga-Soria (2003): *A software tool for thesauri management, browsing and supporting advanced searches.* Proceedings of GI-days 2003. Münster (Germany), 26-27 June 2003.

UNESCO (2003): *UNESCO Thesaurus*. United Nations Educational, Scientific and Cultural Organization. Available from http://www.ulcc.ac.uk/unesco/ (last accessed, May 2003).

IfGI*prints*

Schriftenreihe des Instituts für Geoinformatik,
Westfälische Wilhelms-Universität Münster

Herausgegeben von
Werner Kuhn und Ulrich Streit

Schriftleitung: Jörn Möltgen

Institut für Geoinformatik
Westfälische Wilhelms-Universität
Robert-Koch-Str. 26-28
48149 Münster

Band 18

ISBN 3-936616-14-0

IfGI*prints* 18

Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung

Beiträge zu den Münsteraner GI-Tagen
26./27. Juni 2003

Lars Bernard, Adam Sliwinski und Kristian Senkler (Hrsg.)

ifgi
Institut für Geoinformatik
Universität Münster

This book is published in the series *Collection des sciences appliquées de l'INSA de Lyon*, edited by professor Bernard Balland

Other books published by Presses polytechniques et universitaires romandes:

Systèmes d'information géographique en mode image
Claude Collet

Cartographie des sols
Jean-Paul Legros

Environmental Geomechanics
Laurent Vulliet, Lyesse Laloui, Bernard Schrefler, Eds

Geostatistics for Pollution Data
Michel Maignan, Mikhael Kanevski

You can receive our general catalog on request:
Presses polytechniques et universitaires romandes,
EPFL – Centre Midi, CH-1015 Lausanne, Switzerland
E-Mail: ppur@epfl.ch
Phone: 021/693 41 40
Fax: 021/693 40 27.

www.ppur.org

Presses polytechniques et universitaires romandes,
CH-1015 Lausanne

ISBN 2-88074-541-1

Printed in Switzerland

Inhalt

GDI Architectures

E-Government

GDI Applications

GDI Organisation and Policy

Semantic in GDI

Practical Applications of GDI

GDI Architectures

GDI and Simulation