

Cataloguing for distribution of environmental resources

Jean-Christophe Desconnets¹, Thérèse Libourel², Stéphane Clerc², Bruno Granouillac¹

1 IRD (Institut de Recherche pour le Développement), 34394 Montpellier Cedex 5, France,
jcd@teledetection.fr, bgranou@teledetection.fr

2 LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier)
UMR 5506, Université de Montpellier II, CNRS, 34392 Montpellier Cedex 5, France
libourel@lirmm.fr, clerc@lirmm.fr

ABSTRACT

Environmental applications have strengthened the case for efforts in the establishment of sharing and mutualisation infrastructures for georeferenced information. Within the framework of these initiatives, our work has led us to design and create a tool for cataloguing resources for environmental applications, MDweb. This tool can be used to catalogue different types of resources by using the metadata standard, and offers a search engine for these resources. The goal is to present this tool and make a focus on its originalities. After, giving the principles behind MDweb's design and its major features, we discuss about the interest to use associated thematic and spatial bases to give to the final users a powerful search engine and propose some components to assist and automatize the cataloguing.

At the end, from a perspective of integrating MDweb as a component for locating resources within spatial data infrastructures, we intend to very shortly to implement the Catalog Web service specified by OGC, and also plan for the architecture necessary to integrate other such components that could be used in the framework of non-spatial data infrastructures.

INTRODUCTION

Environmental applications are acquiring increasing importance in the existing land management context, which is being subjected to ever growing natural and human requirements (risk management, land management, etc.). All these applications require georeferenced data. The acquisition process for this type of data is complex and costly. The need for establishing a georeferenced-data infrastructure is universally acknowledged. Several initiatives towards this end have been undertaken; the major ones being FGDC, ANZLIC, NSDI, GSDI and, a recent addition, the INSPIRE initiative in the countries of the European Union.

The architecture of all these infrastructures assumes a cataloguing service for allowing data producers to distribute their data and for allowing the user to locate data sets and even access them.

A cataloguing service is one way of partially solving the problem of interoperability of distributed and heterogeneous resources. In fact, this service is based on the concept of metadata [DESC 01], [LIBOU 03]. Several standardization proposals have been advanced in the last few years in various fields (Dublin Core [DCMI 05], FGDC [FGDC 98], and ISO [ISO 03]). As far as geographic information is concerned, it is the ISO 19115 standard – Geographic Information – Metadata [ISO 03] (ISO TC/211 group) that has been widely adopted by the international community.

Existing search engines can potentially be used to provide a cataloguing service but the description of the metadata used remains incomplete and far from meaningful. The indexing techniques are proprietary to each search engine and so are the searches. It seemed necessary to us to construct a tool that is both generic and expandable, and conforms to the ISO 19115 standard and to the OGC specifications [OGC 05] (Open Geospatial Consortium, <http://www.opengis.org>). MDweb, tool for cataloguing and locating resources, accessible on the Web at <http://www.mdweb-project.org/>, was

thus born out of a collaboration of several partners (Desertification unit (IRD), JRU Tetis (Cemagref-ENGREF-Cirad), Cepralmar, LIRMM, Languedoc-Roussillon Region) within different common research projects.

In section 1, this article reviews the origin of several existing proposals on the subject of cataloguing and locating geographic information. The next section covers the principles behind MDweb's design and its major features. In sections 3 and 4, we will focus on the inclusion of thematic and spatial reference sets and their availability for assisting and facilitating the entry and search stages (section 4).

We shall conclude by presenting our work for the tool's expansion, especially in relation to its interoperability and the extension of the description of processing protocols and resource traceability in the context of spatial-data infrastructures.

EXISTING BODY OF WORK

The use of metadata and initiatives for cataloguing resources in the context of georeferenced information are widespread. The pioneering standards, notably the Federal Geographic Data Committee (FGDC), the Open Geospatial Consortium (OGC), and the European Committee of Standardization (CEN), have allowed their rapid development. The veritable explosion of activities on the Internet has confirmed the interest shown by scientific communities in the distribution and sharing of their resources and knowledge. We have selected those existing proposals that seem to us most relevant and present an overview in table 1, based on essential criteria: standard used, Web or stand-alone architecture, semantic coverage, cartographic interface, proprietary or open source, storage architecture, interoperability, etc.

Name	Standard incorporated	Distributor	User interface	OS	Language used
ArcCatalog 9.x	FGDC, ISO	ESRI ¹	Stand-alone	Windows	C++
M3Cat 1.5	FGDC, ISO	Intelec Géomatique ²	Web	Windows	ASP
Geonetwork 2.0	ISO	FAO ³	Web	Independent	Java
Reports V2	ENV 12657	CERTU ⁴	Stand-alone	Windows	Visual Basic
Nokis	ISO	BMBF ⁵	Web	Independent	Java
MDweb 1.5	Multi-standard	IRD ⁶ , LIRMM ⁷ , Cemagref ⁸	Web	Independent	PHP
Meta-Manager 4	FGDC, ISO	Compusult ⁹	Web	Windows	C++
Red Spider 3.3	ISO	IONIC ¹⁰	Web	Independent	Java

Table 1: Overview of existing cataloguing tools and their characteristics – part 1

¹ ESRI: <http://www.esri.com/>

² Intelec Géomatique: <http://www.intelec.ca/>

³ FAO: Food and Agriculture Organization, <http://www.fao.org/>

⁴ CERTU: Centre d'études français sur les réseaux de transport et l'urbanisme [French centre for study into transport networks and urbanism], <http://www.certu.fr>

⁵ BMBF: German Federal Ministry of Education and Research, <http://www.bmbf.de/>

⁶ IRD: Institut français de recherche pour de Développement [French institute for research into development], <http://www.ird.fr/>

⁷ LIRMM: Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier [The Montpellier laboratory of computer science, robotics, and microelectronics], <http://www.lirmm.fr/>

⁸ Cemagref: Institut de recherche pour l'ingénierie de l'agriculture et de l'environnement [Institute for agricultural and environmental engineering research], <http://www.cemagref.fr/>

⁹ Compusult Limited: <http://www.compusult.net/>

¹⁰ IONIC Software: <http://www.ionicsoft.com/>

Name	Translation	Licence	Cataloguing architecture	Cartographic interface	Interoperability	Associated reference bases
ArcCatalog 9.x	Multi-lingual	Proprietary	XML files	No	No	No
M3Cat 1.5	French, English	GPL licence	MS-Access, SQLServer	SoftMap	Z39.50 service	No
Geonetwork 2.0	English, French, Spanish	GPL licence	MySQL, PostgreSQL,	InterMap	CSW-2 service OGC	No
Reports V2	French	Free	MS-Access	No	No	No
Nokis	Multi-lingual	Free depending on version number	MySQL, PostgreSQL,	MapServer	Z39.50 service	GEMET
MDweb 1.5	French, English, Portuguese	CeCILL licence	MySQL, PostgreSQL	MapServer	Z39.50 service	AGROVOC, GEMET, own Geo DB
Meta-Manager 4	English	Proprietary	MS-Access, SQLServer, Oracle	Open	Additional components	No
Red Spider 3.3	English	Proprietary	Oracle, PostgreSQL	Open	CWS-2 service	No

Table 1: Overview of existing cataloguing tools and their characteristics – part 2

From this rapid overview, we observe that the majority of existing solutions implement, or are in the process of doing so, international standards (ISO, OGC) thus allowing catalogue structuring and interoperability. We also see that the architecture used to deploy these tools is the Web except for those tools included in wider applications (ESRI's ArcCatalog) or those designed a few years ago (CERTU's REPORT). As far as the tools' architectures are concerned, the existing offers are based on similar architectures: a DBMS component for storing data catalogues (Oracle, MS-Access or MySQL, PostgreSQL for free solutions) connected to a Web application server of type Tomcat/Java, Apache/PHP or IIS/ASP. In terms of user interactivity, most of these tools now include cartographic modules for helping construct geography-based queries and also, for some of them, to display the georeferenced data.

Finally, in terms of the distribution of the solutions, for the most part, existing solutions are proprietary and have to be paid for; most of these form part of larger solutions designed for creating and managing entire Geographical Information Systems. Those tools available under free licence (GPL, CeCILL: French version of the GPL licence) are based on free technologies (programming language, DBMS and cartographic server); their design focuses mainly on cataloguing and locating functions, using interactions with other components by implementing, most often, the OGC standards. On the other hand, NOKIS and MDweb are amongst the only tools that offer the use of thematic and spatial reference bases to help the data producer index his data sets, enhance the search engines by using the relationships between concepts included in these models and, finally, offer to the user the possibility of interacting with the search engine to refine or expand his search. The association of semantics with the cataloguing tool via the inclusion of knowledge-representation models of the concerned domain (thesaurus, geographic-objects database) is not much developed yet and leads to a problem of consistency of the indexing of the resources, a problem that can be fundamental in a context of a cataloguing project involving a large number of users. We wanted to offer a comprehensive tool based on free technologies that can be scalable and adaptable, especially in its availability and in the customization of the reference bases associated with the target application domain.

PRESENTATION OF MDWEB

Principles and architecture

Conscious of the variety and potential expandability of the standards, we designed the MDweb tool with genericity and modularity in mind. In fact, we want to offer more than a traditional cataloguing tool by providing a meta-tool, capable of creating, managing, administering and consulting multi-standard and multi-lingual catalogues over the Web. MDweb is a server-side application and its modular architecture is designed around a completely 'open source' solution. Its implementation (conforming to the n-tier principle) is based on three components: a database management system (DBMS), components for the dialogue between the HTTPD server and the DBMS, PHP and an HTTP server for publishing data (cf. figure 1). The tool has been developed to be independently deployed on the Windows or Linux operating systems.

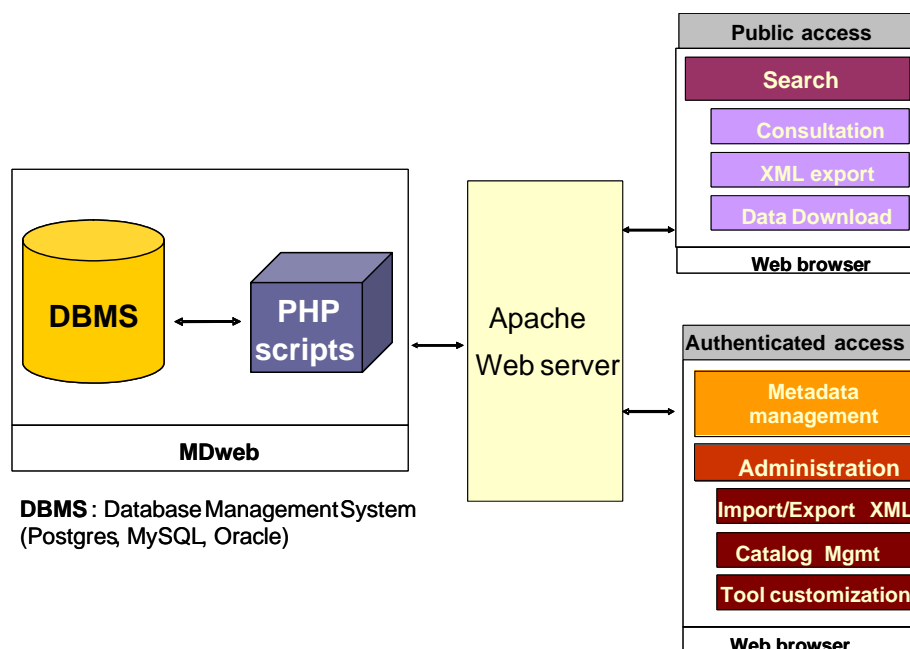


Figure 1: MDweb general architecture

To ensure interoperability with other cataloguing applications, MDweb notably uses the international standard for geographic-information metadata (ISO 19115 and its 19139 implementation in its next version) for structuring metadata and the Catalog Service specifications of OGC to ensure interoperability of the cataloguing service by implementing the Z39.50 protocol, GEO profile [ISO 98].

Genericity and modularity

The design and implementation of the tool draws on the concepts of metamodeling and modularity, first introduced in the software engineering field, which ensure portability and reusability. The genericity of the tool's constituent components enables its application in varied contexts.

These essential components consist of a generic database of metadata (metabase) and two thematic and spatial reference bases used for semantic control of metadata during the data-entry phase and during searches.

The genericity of the metadata database resides in the originality of the relational schema that describes a structure at a meta level coupled with a standard structure.

It can be broken up into five schemas (cf. figure 2):

- storage schema of the standard's dictionary,
- storage schema of the metadata-profile structures and their parameters inherited from a standard,
- storage schema of the metadata (metadata storage and catalog),
- storage schema of the contents of the user interfaces (labels, user parameters).
- storage schema of search index.

The use of the storage schema of the standards dictionary allows the generation via calculations of different structures (metadata profiles) and the ability to implement and change the tool without having to modify the relational schema.

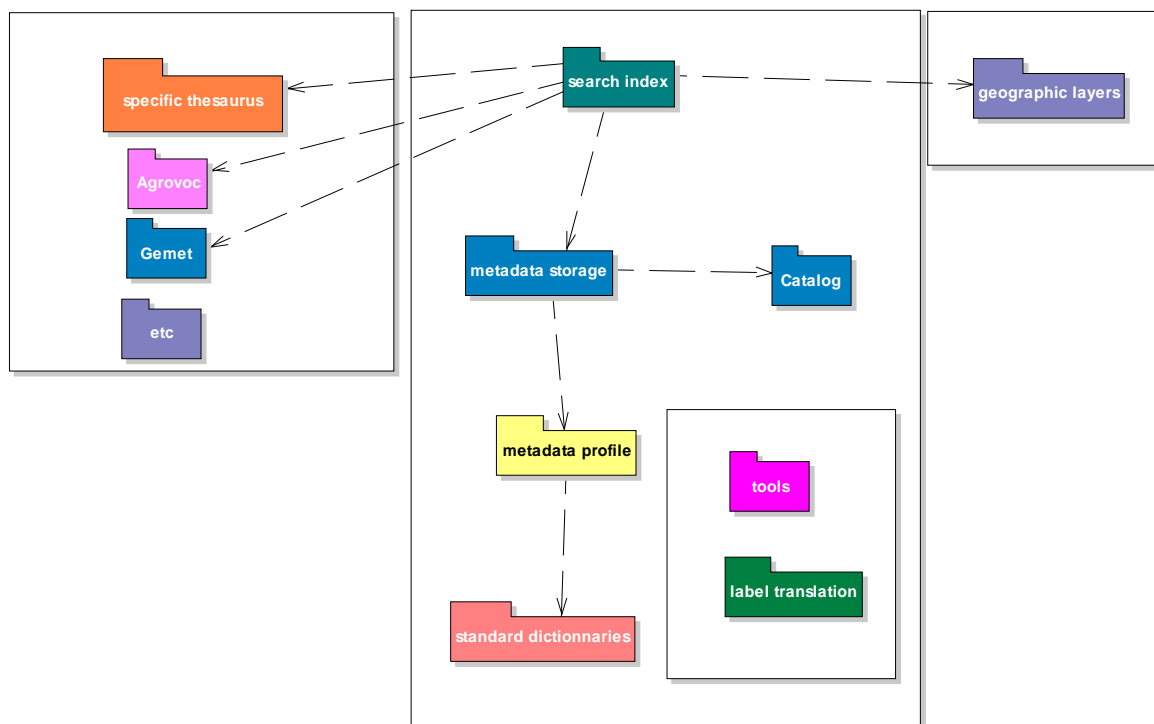


Figure 2: Schema of the generic relational model

Any user community can thus, with the help of this component, adapt the structure of the metadata description and configure the user interfaces as convenient.

Two other components, the spatial and thematic reference bases, help improve the geographic and thematic indexing of referenced resources. They fulfil the need of controlling metadata items that relate to the description of the resource's geographic extent and to the thematic description, both essential at the query level.

The spatial reference base is a spatial database dedicated to the storage of objects relevant to the concerned user community. The thematic reference base is also a database designed to store one or more existing thesauri (e.g Agrovoc, Gemet for environmental context) suitable to the utilization context. You can also create a new thesaurus for a specific context of use of MDweb.

Functionalities and associated actors

Our tool is designed to be a component of a data infrastructure. To begin with, we have delimited the scope of the infrastructure to that of a limited community of users. Within such a framework, several functionalities have been defined for users with predetermined roles (actors) [DESC 03]. Four distinct roles are proposed (cf. figure 3):

- The end user,
- The producer,
- The expert,
- The administrator.

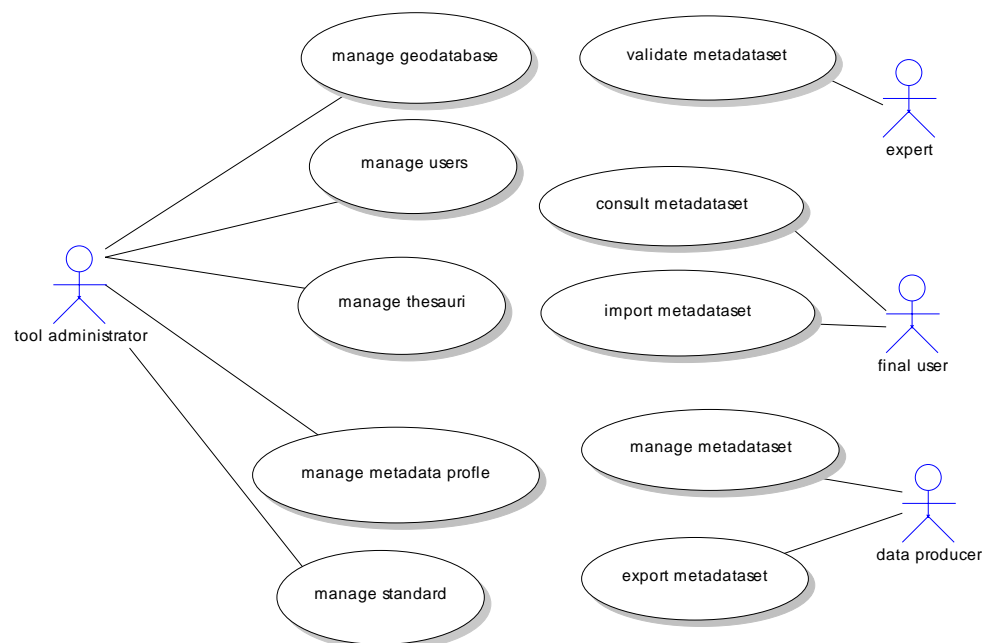


Figure 3: Usage cases (roles and functionalities)

MDweb’s modular architecture is designed to offer several usage levels depending on the role assigned to the user (previous section). It consists of:

- a **‘search’ engine** accessible by all users to locate data sets indexed in the catalogue(s). This engine consists of several modules each with different levels of complexity and interactivity during the formulation of query requests:

- *a module for searching by geographic localisation.* It is based on a cartographic component associated with the spatial reference base. It offers the user a set of geographical objects in his domain of interest, i.e., the geographical representations specific to the framework to

which the information-sharing context belongs. The searches are conducted by using spatial operators of the DBMS (intersection or inclusion) relying, whenever possible, on the represented objects or by the selection of a search zone.

- *a multi-criteria search module*: The user can construct a query by combining four criteria that respond to the following questions: What type of resource? To what topic does the resource relate to? Where is it located? and When was it created? SQL queries thus created are based on the filtering of contents of the following items: title, summary and keywords using the LIKE operator. The spatial requests use PostGIS's intersection operator by relying on the coordinates of a resource's spatial extent.

- a **'management' module for metadata records** allows an authenticated user to enter and update metadata records, to attach resource files, view the progress of the data-entry, to manage a metadata record's and its attached resource's privacy settings, to import to and export from the catalogue in the ISO 19115 XML format, to manage the address directories and to access the area for adding terms to complement the thesauri used.

- a **module for the 'administration' of the catalogues and the tool**. It includes different modules for configuring the tool (adding new languages, managing the record-publication rule, etc.). Associated with these functionalities, the administration module offers the administrator the ability to define metadata profiles for his user community (items of the standards and their properties to generate entry forms and the use of metadata for searches). It also offers the management, import (importing a reference thesaurus in the SKOS or RDF formats) and creation of thematic reference bases (addition of terms and relationships proposed by the users) and spatial reference bases (creation of a spatial database and import of information layers).

ASSOCIATED THEMATIC AND SPATIAL REFERENCE BASES

The associated thematic and spatial reference bases

Conscious of the fact that a cataloguing tool's success depends to a great extent on the final users obtaining relevant results to their searches, we have emphasized the contribution of semantics as an extension of related section present in the metadata structure. This has taken the form of extending the Keywords and DataIdentification sections of the ISO-19115 standard (cf. figure 4). Initially, the ISO standard allows for the resource description via keywords selected from a pre-established list specific to the geographic information of the application domain.

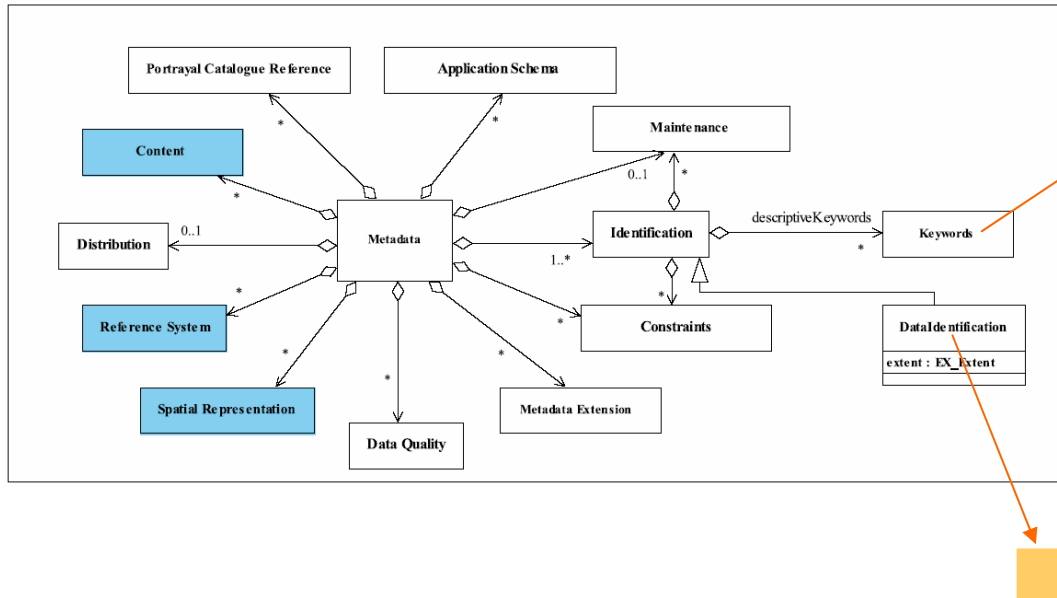


Figure 4: Extending the ISO 19115 standard via the Keyword and dataIdentification classes

However, it also recommends an extension for the use of a dedicated vocabulary (thesaurus). On the strength of this recommendation, we have introduced the use of thesauri within MDweb. A thesaurus is a set of terms arranged by semantic relationships (cf. figure 5) defined by the ISO 2788 [ISO 86] standard.

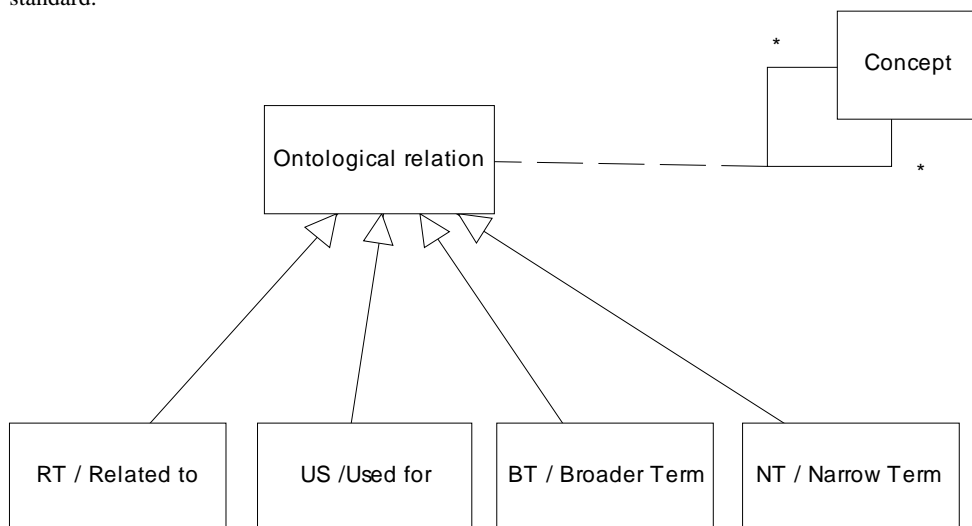


Figure 5: The principal semantic relationships between terms, ISO 2788

Several reference thesauri have been included with the MDweb tool: the multi-lingual agricultural-terms thesaurus of the FAO (Food and Agriculture Organization) called AGROVOC and the multi-

lingual environmental-terms thesaurus called GEMET from EIONET (European Environment Information and Observation Network). Other thesauri can be associated by importing them in the SKOS or RDF formats in the relational schema of MDweb's thematic reference base.

Going beyond these thesauri, we have introduced the concept of the thematic reference base. A thematic reference base describes, for a given community, the semantics of the domain under consideration by the intermediary of explicit models. A model translates a community's expertise. It is the *semantic interoperability* vector between actors for *sharing the understanding of the domain's concepts*. For example, in the context of integrated management, the model defines [Barde 05] the concepts of actor, resource, interaction and tool (cf. figure 6).

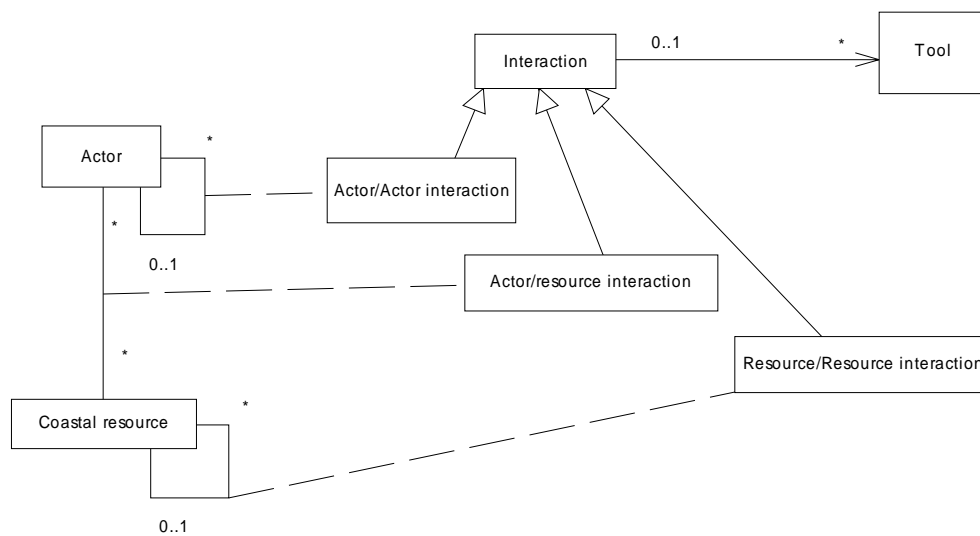


Figure 6: Example of a model for integrated management of coastal zones [BARDE 05]

Using this model, the standard terms and relationships of pre-existing thesauri can be 'filtered' in the MDweb application to construct a specific thesaurus suitable for use by the concerned community.

An approach similar to the one on the terminology was applied to the spatiality of information. We designate by spatial reference base a set of relevant geographic objects, which can act as referents for the concerned community.

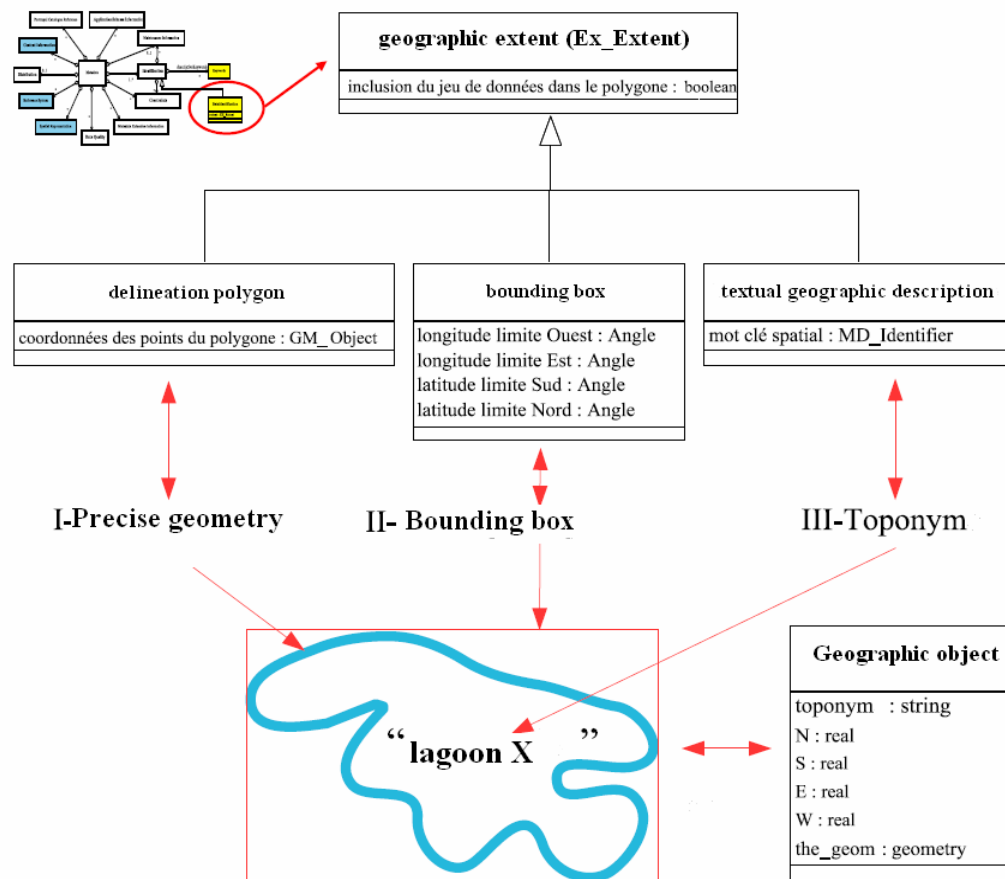


Figure 7: Illustration of three types of ISO 19115 geographic descriptors

The referent geographic objects can be represented in different ways: toponymic identifiers (term of a semantic reference base), bounding box describing the object's spatial extent, the precise geometry of the object.

The two thematic and spatial reference bases are tightly correlated (cf. figure 7), the spatial dimension constitutes a 'mediator' between the thematic search and the spatial search. A spatial term describes an intentional term (example: lagoon) of the semantic reference base and possesses an extension in the form of geographic objects that constitute a 'layer' of the spatial reference base.

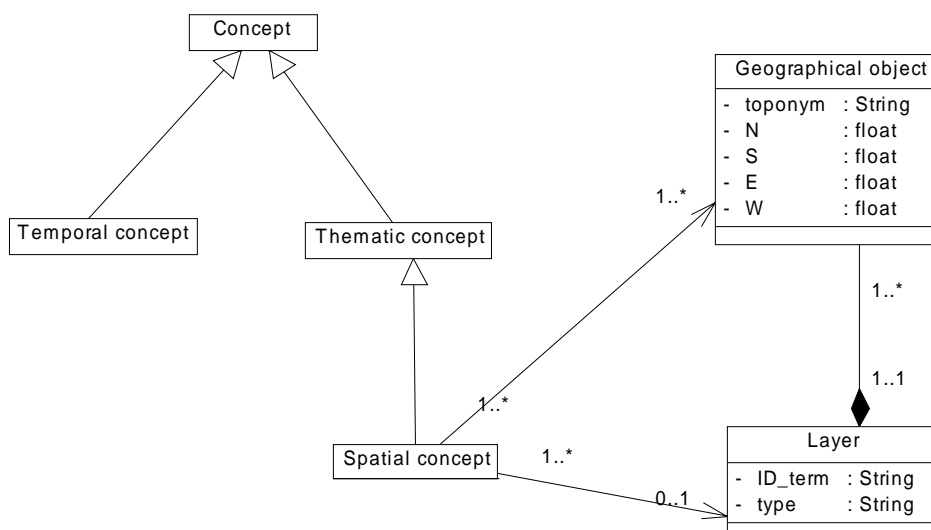


Figure 8: Link between thematic and spatial concepts [Barde 05]

CONCLUSION AND OUTLOOK

For the past several years, MDweb has been deployed and implemented in various application domains for the purposes of sharing and dissemination of data and knowledge (Integrated management of coastal zones <http://syscolag.teledetection.fr>, PRCM regional programme in West Africa <http://soga.univ-brest.fr/mdweb/>, Desertification observatory <http://mdweb.roselt-oss.org/>, National network for environmental monitoring in Tunisia and Egypt, etc.). In some implementations, the tool was deployed in a distributed architecture based on the Z39.50 communication protocol.

From a perspective of integrating MDweb as a component for locating resources within spatial data infrastructures, we intend to very shortly complement it by implementing the Catalog Web service specified by OGC, and also plan for the architecture necessary to integrate other such components that could be used in the framework of non-spatial data infrastructures. The semantic interoperability necessary to maintain consistency within catalogues and useful for conducting searches should be retained when the associated reference bases are shifted to a completely separate service.

From the point of view of the description of resources, we will have to expand the description to include processing protocols and data traceability on the one hand and, on the other, propose a description of quality to judge the resource's reliability, not only its intrinsic quality, dependent on the sensor, but also its quality in relation to the application domain.

BIBLIOGRAPHY

- Barde, Julien., 2005: Mutualisation de données et de connaissances pour la gestion intégrée des zones cotières. Application au projet Syscolag., PhD, Université Montpellier II, December 2005.
- Barde J., Libourel T., Maurel P., Desconnets J.C., Mazouni N., Loubersac L., 2005: A Metadata service to manage resources in coastal areas. CoastGis'.
- DCMI, 2005. Dublin core metadata initiative. <http://dublincore.org/>.

- Desconnets J.C., Libourel T., Maurel P., Miralles A., Passouant M., 2001: Proposition de structuration des métadonnées en géosciences. Spécificité de la communauté scientifique. Géomatique et espace rural, Journées Cassini 2001, 26-28 September 2001, Montpellier. pp. 69-82.
- Desconnets, J.C., Moyroud, N., Libourel, T., 2003: Methodologie de mise en place d'observatoires virtuels via les métadonnées. INFORSID actes du XXIeme congrés (2003), 253-267.
- FGDC, 1998: Content standard for digital geospatial metadata. FGDC-STD-001-1998.
- Libourel T., Desconnets J.C., Maurel P., Moyroud N., Pasouant M., 2003: Les métadonnées: pourquoi faire? (in CD-ROM) Géo-Evénement 2003, Paris.
- ISO 2788, 1986. Guidelines for the establishment and development of monolingual thesauri, ISO 2788. International Organization for standardization (ISO), Geneva, Switzerland.
- ISO 19115, 2003: Geographic Information Metadata, ISO 19115, International Organization for Standardization, Geneva, Switzerland.
- ISO 23950, 1998: Information and documentation – Information retrieval (Z39.50) Application service definition and protocol specifications, ISO 23950. International Organization for Standardization, Geneva, Switzerland.
- OGC, 2005: Catalog Service Specification. Open Geospatial Consortium, Inc.
- Boisson P., Clerc S., Desconnets JC, Libourel T. 2006 : Using a semantic approach for a Cataloguing Service. OTM workshops (2) 2006 : 1712-1722. LNCS Springer Heidelberg.