Kéna-Cohen, S. & Y. Bédard, 2001, M3Cat, Limiting Aggravation in Geospatial Metadata Cataloguing, Conférence GIS2001, 22-févr., Vancouver, Canada

M^{3}Cat, Limiting Aggravation in Geospatial Metadata Cataloguing

Kéna-Cohen Serge
Partner
Intelec Geomatics Inc.
696, Ste-Croix Boul., Suite 200, Saint-Laurent, Canada, H4L 3Y2
Tel. (514) 744-2029, Fax (514) 744-4495, E-mail kena@intelec.ca
Bédard Yvan,
Laval University, Centre for Research in Geomatics
Quebec City, Canada G1K 7P4
Tel. (418) 656-2131 (3694), Fax (418) 656-7411, E-mail yvan.bedard@scg.ulaval.ca

ABSTRACT

Timely access to accurate geospatial data is crucial to any GIS application. Stakeholders involved in distributing and accessing geospatial data agree that the first stage in achieving this goal is the documentation of data sets or the entry of metadata. Standards organisations (FGDC, ISO/TC211, OpenGIS) have identified metadata as a priority in their work and geospatial data infrastructures often start with the implementation of services related to metadata (discovery, evaluation, access).

However, the gathering of metadata is often a tedious and frustrating exercise that discourages most organisations that have started the cataloguing of their data sets. They are confronted with complex standards that do not handle all of their particularities and existing tools are not user-friendly. These tools have a steep learning curve, limited flexibility with regards to accepting metadata based on different standards, different versions of a same standard, or user-specific metadata, and usually do not work properly in any language other than English.

The M^{3}CAT (Multistandard, Multilingual Metadata Cataloguing) tool was developed under the Canadian Geospatial Data Infrastructure GeoInnovations program to resolve these problems. Based on more than 10 years of research and practical experiences, M3CAT has been designed and developed to offer flexibility and to benefit from the World Wide Web environment. First, it allows users to configure their own metadata fields using the ones delivered with the tool (FGDC, ISO/TC211, GILS and NBII), i.e. to build their own profile of a standard. Second, users can customise the user interface for standard parameters such as pull-down lists, establish default values for numerous metadata fields (organisation name, projection system, etc.), use templates that correspond to the type of data set they document and then take advantage of cataloguing aids such as Wizards, metadata inheritance between data sets, automatic metadata extraction from existing geospatial databases and translation into another language to facilitate the development of metadata databases. Third, it is ready to support the translation between standards (ex. for a ISO querying of FGDC metadata base) and among different versions of a same standard. Fourth, M^{3}CAT also integrates a semantic analysis tool to manage keywords and a multi-projection, multi-datum geographic viewer to enter data set bounding co-ordinates.

M^{3}CAT is provided free to potential users and has been tested with half a dozen governmental organisations involved in the management of geospatial data. It works in a Web environment, supports XML and SGML and can interface to the ANSI/NISO Z39.50 FGDC GEO profile.

CONTEXT

Timely access to accurate geospatial data is crucial to any GIS application. Stakeholders involved in distributing and accessing geospatial data agree that the first stage in achieving this goal is the documentation of data sets or the entry of metadata. Standards organisations (FGDC, ISO/TC211, OpenGIS) have identified metadata as a priority in their work and geospatial data infrastructures (GeoConnections, NSDI, ANZLIC, etc.) often start with the implementation of services related to metadata (discovery, evaluation, access). It is one of the few areas where ISO/TC211, FGDC and OpenGIS have agreed to collaborate.

Many organisations have implemented metadata servers to allow discovery and access of their geospatial information. Many tools exist to perform this work. A recent survey made by the Ohio Geographically Referenced Information Program identified fifty tools supporting the FGDC metadata standard (Wisclink et al 2000). However, the gathering of metadata is often a tedious and a frustrating exercise that discourages most organisations that have started the cataloguing of their data sets. They are confronted with complex standards that do not handle all of their particularities and existing tools are not user-friendly. The tools have a steep learning curve, limited flexibility with regards to accepting metadata based on different standards, different versions of a same standard, or user-specific metadata, and usually do work properly in any language other than English.

The M^{3}CAT (Multistandard, Multilingual Metadata Cataloguing) tool was developed under the Canadian Geospatial Data Infrastructure GeoInnovations program to resolve these problems. Based on more than 10 years of research and practical experiences, M3CAT has been designed and developed to offer flexibility and to benefit from the World Wide Web environment. M^{3}Cat is a tool that assists users in entering and managing metadata about geospatial data sets. Metadata consist of information that characterises data. Metadata are used to provide documentation for data. In essence, metadata answer **who**, **what**, **when**, **where**, **why**, and **how** about every facet of the data that are being documented.

M^{3}CAT is provided free to potential users and it allows users to enter metadata using any standard (Multistandard) and any language (Multilingual). It is presently provided with the FGDC, GILS, NBII and ISO 19115 metadata standards and in English and French. Functions are available to add other standards and/or languages.

The tool was developed by Intelec Geomatics (Intelec) and its partners, SoftMap Technologies, CrossDraw, and the Centre for Research in Geomatics (CRG) of Laval University. It has been tested with four governmental organisations involved in the management of geospatial data: Elections Canada, Environment Canada, Natural Resources Canada and the Plan géomatique of the Quebec Government.

M^{3}Cat is built on a rich experience on both the theoretical and practical sides as well as on several collaborative projects between industry and academia for more than 6 years. On the one hand, in 1995, Intelec had developed, within the GEOSCOPE Network project (financed by the CANARIE program) a metadata cataloguing tool with a number of interesting features (Intelec 2000).

This GEOSCOPE tool is currently used by Natural Resources Canada (National Atlas), Environment Canada, Transports Canada, Quebec's ministry of Natural Resources, as well as in Egypt and Venezuela. An average of 25 agencies download the tool every month.

On the other hand, Laval's CRG has made fundamental R&D (including participation to GEOSCOPE), practical experiences and published several thesis and papers since 1989 on the topic of geospatial metadata management (see Bédard 2001). They have implemented what is known as the first Geospatial Data Infrastructure in the world geared specifically for academic purposes (Larrivée et al 2000) and have given continuing education courses on this topic in Canada and Europe.

The experience gained over the years in the development and support of such type of tool allowed the project team to identify what would be the features of a new improved metadata cataloguing tool.

FEATURES

M3Cat offers a number of features, some of which have required the use of innovative concepts:

- The capability to catalogue metadata using various standards in particular FGDC, GILS, NBII and ISO 15046-15 (TC211).
- The capability for users to develop templates for each standard which take into account the types of data sets, the particularities of the organisation processes or information.
- Multilingual support with semi-automatic translation of metadata fields.
- Cataloguing aids, in particular, a Wizard and a Help menu, as well as the ability to obtain the significance of each metadata element; online validation of metadata elements according to each profile; offline validation that verifies metadata completeness (mandatory fields) and an approval process that can be performed by a supervisor; a map interface to assist in entering data sets bounding co-ordinates by panning and zooming on a map; and an automatic metadata extraction feature that allows users to automatically extract metadata fields from digital files (through OGDI).
- The capability to store metadata on different levels (parent and child) of data sets (granularity) such as at data base, map sheet, layer or entity levels with the capability for child data sets to inherit the metadata values of their parent.
- The capability to store standard values (parameters) in pick lists, such as information about an organisation, reference systems, etc. These pick-lists can be defined and modified by the user.
- A Semantic Analysis Tool that manages semantic relations between keywords. These keywords can be part of a defined thesaurus.
- The capability to make the metadata a clearinghouse node using either ISITE or MetaManager as a ANSI/NISO Z39.50 server.
- The capability to import or export metadata. Import and export files are in ASCII formats.
- The capability to manage data either in Oracle or Access.

MAIN CONCEPTS

The main features and concepts used in M3Cat are described in the following sections.

Multi Standard

Tools for handling metadata need to rely on their being predictable in both form and content. Predictability is assured only by conformance to standards. There are many standards for metadata such as the FGDC Content Standard for Digital Geospatial Metadata, the GILS standard (www.gils.net/), the NBII standard (www.nbii.gov/datainfo/metadata/standards/index.html) and the ISO 15046-15 (TC211) metadata standard.

Organisations also tend to define a profile of metadata for their own applications, encompassing validation rules, mandatory elements and customisation which take into account their particularities.

Users that document metadata in M^{3}Cat must use a **standard**. M3Cat is delivered with 4 metadata standards: FGDC, GILS, NBII and ISO 15046-15, but others can be developed.

For a specific standard, an organisation identifies a **profile.** A profile adds the cultural aspect to a standard, in particular a language and other particularities. While an organisation cannot change the standard, it can modify its profile using M^{3}Cat label management functions.

Data managers can also define **templates** for specific types of data sets. For example a template for raster data set will only include the metadata elements relevant to this type of data set. A template can also add the particularities, or rules, that an organisation wants their metadata elements to follow. A template may identify specific mandatory fields.

M^{3}Cat uses the concept of metamodel with some extensions to support these features. It contains all the elements proposed by the OpenGIS Abstract Specification for metadata (see Figure 1).

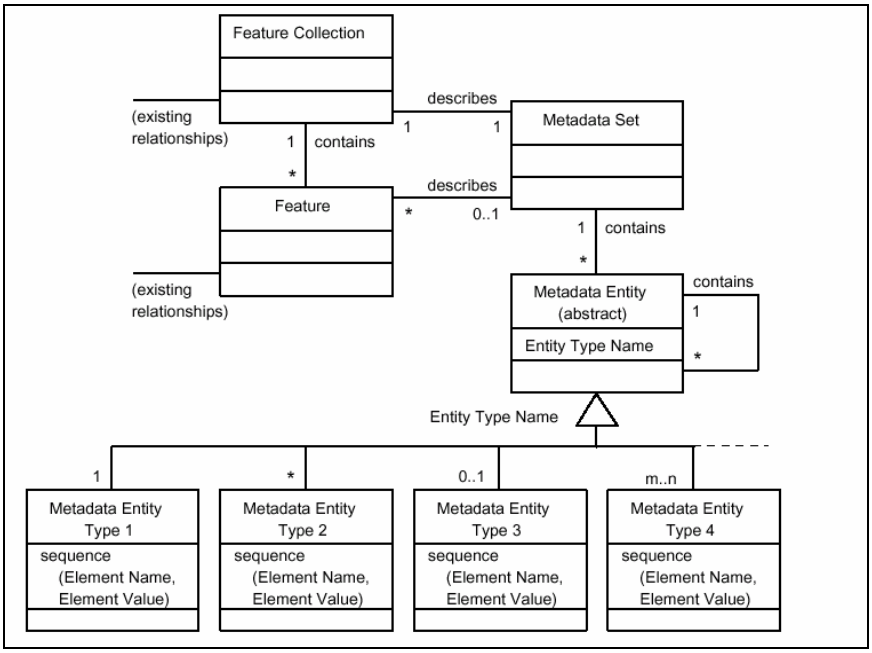

Figure 1- OpenGIS Metadata Classes and Relationships

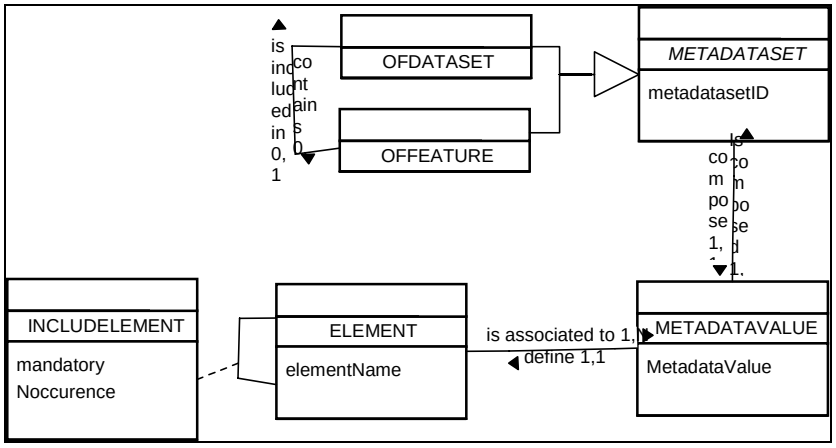

Figure 2-M^{3}Cat MetaModel

Multilingual

Currently, very few tools allow metadata cataloguing in a language other than English. Moreover, there does not exist to our knowledge of tools that:

- Allow cataloguing in more than one language;
- Take into account cultural differences.

The concept in M^{3}Cat is that the majority textual contents as well as the domain values for textual data and their units are coded and refer to tables of reference which give translations in each language (see Figure 3). For example for " Frequency of update " the domain values are " mensuel ", " quotidien", "au besoin ", in French and " monthly ", "daily " and " as needed" in

English. According to the language of cataloguing chosen, the tool will identify the contents to be used for cataloguing.

This is also the concept used by ISO JTC1 (Information Technology).

M^{3}Cat also offers the capability to translate the metadata elements into another language. Translation is a two step process : the first step automatically translates standards values into their equivalencies in the new language, the second step presents the text values not translated and allows the user to edit them. Once the process completed, the user can change the status of the translation flag to completed.

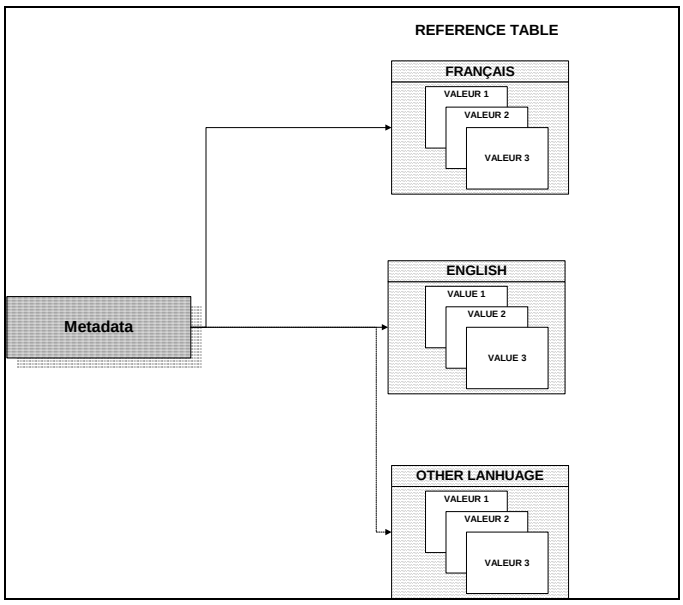

Figure 3- The Use of Reference Tables

Cataloguing Aids

A number of concepts and technologies are used to implement the cataloguing aids:

- A Wizard is active by default when starting the tool and conducts the user through a step by step cataloguing process;
- For validation, a number of features are available:
  - validation that all mandatory fields are entered;
  - validation that formats correspond to the standard (or template);
  - validation that fields with domain values use the appropriate reference table;
  - contextual validation of interrelated fields;
  - reporting of various types as well as the display of data set coverages on a geographic base.

- The SoftMap Developer toolkit is used to allow the entry of data sets coverages with an interactive multi-scale, multi projection map interface;
- For the automatic entry of metadata fields, the OGDI API is used to extract these fields from digital files (see Figure 4);
- Data sets can be organised in a hierarchical manner with lower level data sets inheriting metadata from the higher levels.

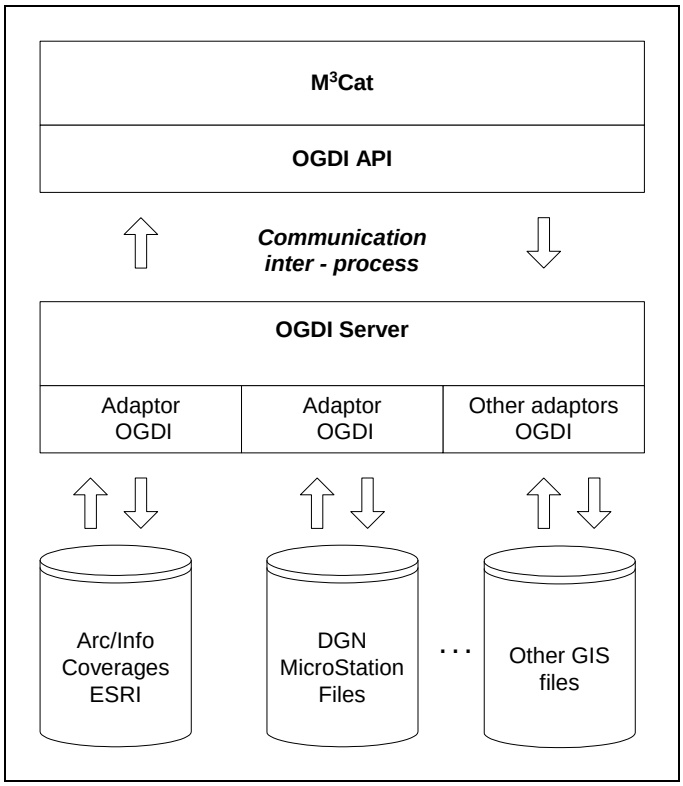

Figure 4 – OGDI Interface

Semantic Network Tool

A Semantic Tool, manages the keywords. The Tool maintains the semantic links between the keywords by essentially using Table Keyword and the recursive relation To Link to (see Figure 5).

The Keyword Table maintains the type (theme, entity, attribute), the name and the definition of each keyword. When appropriate, the Keyword Table refers to a Thesaurus Table. The relation To link maintains the type of link (e. g. synonym, included, polygamy quasi-synonym, etc.) and the Semantic Network Number. This number allows semantic relations to be built per application domain.

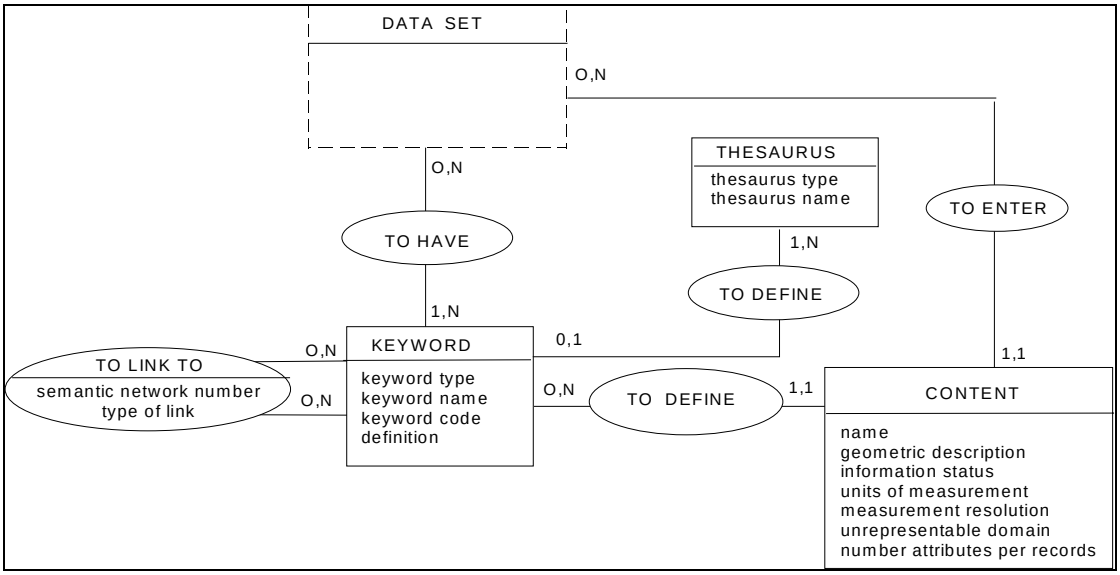

Figure 5 – Semantic Network Structure

Technologies

M3Cat can be used either in standalone mode (Windows NT, 95 or 98) or within a network in a multiple users mode with appropriate privileges defined for each type of user.

It works within a browser, Microsoft Internet Explorer version 3 and after or Netscape Navigator version 4.0 or after.

Metadata elements are stored in either Access or Oracle and can be exported in HTML, XML or SGML (interface with FGDC/MP).

M3Cat is provided with interfaces for ISITE (CNIDR) and MetaManager (Compusult). This allows users to connect to any search engine or clearinghouse using the FGDC GEO profile and the ANSI/NISO Z39.50 protocol.

WHAT'S NEXT

M3Cat is currently undergoing beta testing and should be available by the end of March.

The next step is to look at the integration of geospatial data and metadata. The objective is to forge a dynamic link between geospatial data servers and metadata servers. By integrating data and metadata management, these tools will enable data providers to directly update their metadata and to more easily configure their data for the Canadian Geospatial Data Infrastructure (CGDI). Integrated management will allow providers to devote more time to improving data accuracy, updating databases, and gathering new data. Users of geospatial data also will benefit, since the integration tools will make relevant information easier to find and its quality and detail easier to assess.

BIBLIOGRAPHY

Bédard 2001. Personnal web site, publications section. Web site as published on January 4th 2001, located at http://sirs.scg.ulaval.ca/yvanbedard.

Intelec 2000. Intelec web site, technologies section. Site located at http://www.intelec.ca

Larrivée, S., Bédard, Y. and Branger, E., 2000. Infrastructure de données géospatiales: concepts et application à l'enseignement pratique de la géomatique (IDG). National Symposium Geomatics 2000, Canadian Institute of Geomatics, Palais des congrès de Montréal, Québec, March 8-10.

WISCLINC, Wisconsin SCO, NSGIC and FGDC 2000. Metadata Tools for Geospatial Data. Web site published on December 29th 2000, located at http://www.state.wi.us/agencies/wlib/sco/metatool/mtools.htm#TOOLBOX

BIOGRAPHY

Mr. Kéna-Cohen earned his bachelor's and master's degrees in electrical engineering from École Polytechnique in Montreal. He has over 25 years experience with information technologies. Since 1981 he has worked primarily in GIS analysis, development and implementation.

In the past, he has managed or worked on a number of GIS projects for:

- Federal organisations (NRCan; Environment Canada; Transports Canada; Oceans and Fisheries; National Defence; Elections Canada);
- Provincial organisations (Quebec Environment, Forestry, Mines and Communications ministries);
- Municipalities (National Capital Commission; Montreal Urban Community; and a dozen municipalities);
- Utility Companies (Bell Canada, Hydro-Quebec, Gaz Metropolitan, Videotron).

Mr. Kéna-Cohen has also worked actively on GIS projects abroad: in Eastern Europe (Poland, Belarus, the Ukraine), Latin America (Argentina, Costa Rica, Venezuela, Mexico, Nicaragua), Asia (Indonesia, Thailand, the Philippines) and Africa (Egypt, Morocco, Ethiopia, Cameroon, Senegal, Burkina Faso).