METADATA REQUIREMENT FOR GIS: A CANADIAN EXPERIENCE

Gordon Plunkett
GeoAccess Division, Geomatics Canada, Natural Resources Canada
615 Booth Street, Ottawa, Ontario, K1A 0E9
Ph: (613) 992-0389; Fx: (613) 952-0916
E-Mail: gordon.plunkett@geocan.nrcan.gc.ca

Commission II, Working Group 3

KEYWORDS: GIS, Database, Standards, Archiving, Organization.

ABSTRACT:

During the development of Geographic Information Systems (GIS) projects, there is usually a requirement to perform data exchanges between different agencies or at least between different systems. Various distinct kinds of data are required by most GIS application projects in order to complete the project successfully. With the growth in the use of GIS and in the voracious data needs of GIS systems, there is a great requirement for common and consistent descriptions of the diversified kinds of data that can be used in GIS. Information describing data is often called metadata or data about data. Because of the diverse nature and multiplicity of data formats and standards, a need for some type of metadata standardization is required. Within Canada, a working group of the Canadian General Standards Board has developed a standard that can be used for the description of various digital georeferenced data sets. This metadata once collected can be used for the creation of data set directories, which can in turn be used by users for locating the various data sets required by their application. This paper describes the new Canadian metadata standard and several methods that are being used for the dissemination and maintenance of this metadata. Some of this work is being performed by the GeoAccess Division of Geomatics Canada under the auspices of the Inter-Agency Committee on Geomatics (IACG) within the Canadian federal government. The paper describes earlier work performed by Roger Tomlinson for the IACG, including both the data collection phase and the data distribution phase. Also the new Canadian standard for the collection of data describing digital georeferenced data sets is documented.

1. INTRODUCTION

GIS application project managers will generally spend over half of their project resources on data related issues. During the development of GIS projects, data related issues such as data location, data conversion and data storage can become messy and expensive tasks. Most GIS projects require various different kinds of data in order to complete the project successfully. The following list outlines many of the most common types of data required for typical GIS application projects.

- Base cartographic data (most often derived from topographic maps) with good spatial accuracy;
- Thematic data that is derived or conceived from human interpretation;
- Spatial statistical data that is derived mathematically from some type of data collection or census;
- Remote Sensing imagery data that is or can easily be geocoded to the base cartographic data; and
- Other kinds of data that is collected or portrayed in a GIS.

Project managers must often decide on whether to purchase or obtain data sets without ever seeing the data set. This means that a clear concise description of the data set must be available for the manager to make his decision. A standard for depicting this type of information describing digital georeferenced data sets should contain fields for the identification of the data set; contact persons; a description of the data set; a description of the coverage (including spatial, vertical and time); data revision; data set size; data collection details; data quality; the availability of the data set; the medias available; the structure of the data set and any reference material regarding the data set. The metadata standard should contain sufficient information for a researcher or manager to decide if they are interested in obtaining more information on the data set or perhaps the data set itself.

In the past, metadata was collected and then a directory was constructed that would allow users to peruse the catalog for information regarding their requirements. This previous work in data directories led to the development of data schemas in commercial DBMS for storage of the metadata. The metadata was then provided to users on diskette in this commercial format which the user would then load into his own DBMS or use directly. Users could then perform searches, produce reports, locate specific data set etc.

This method of data distribution was adequate at the time, but it suffered from several problems in the data revision area. The problem included the fact that users who received a diskette were not automatically sent a revised diskette because it was an expensive task to deliver updated diskettes to all the users. In addition, it was difficult for the data producing agencies to update their data easily and this needed to be done at a central site. So the result was a good metadata product, but it was difficult to maintain and disseminate. However, the solution to many of these problems is to develop a capability using the Internet to allow users to perform both data search and update functions.

2. INTER-AGENCY COMMITTEE ON GEOMATICS

The Inter-Agency Committee on Geomatics (IACG) was formed in 1988 from members of various Canadian federal

government departments and agencies to facilitate discussion, coordinate developments in GIS technology and to promote GIS activities. The Government Data Bases working group (now known as the Access and Marketing working group) of the IACG developed a metadata collection in cooperation with Tomlinson and Associates for the collection of digital and analogue geographic data within the Canadian federal government. The results of the data collection are described in the Report on the current Status and Trends in Federal Digital Geographic Data in Canada [IACG01].

The working group completed the survey of federal government geomatics data holdings. The survey results were an invaluable source of information for the many users who require GIS data for their analysis. Also, by knowing of the existence of significant datasets, duplication of data collection could be minimized. The working group drafted a list of information to be collected, administered the effort on the survey, and evaluated the resulting report. The survey was conducted through the completion of a questionnaire by a large number of selected agencies and a set of criteria was developed and applied for dataset inclusion.

Information on each dataset such as identification, description, availability, structure and applications was collected, verified and entered into the survey results database. Indexes on dataset name, acronym, agency, category, type and coverage were produced with additional analysis of the resulting metadata base. Following the significant distribution of this report and metadata database, it was decided that perhaps the most effective method of maintaining the information in the database was to have the producing agency maintain their data entries. However, with the proliferation of metadata databases, it was decided that a standard method of metadata collection was required so that data producers could then provide this information to data suppliers. The IACG then assisted in the development of a common metadata collection standard through the Canadian General Standards Board.

3. CANADIAN GENERAL STANDARDS BOARD METADATA STANDARD

The Canadian General Standards Board Directory Information Describing Digital Geo-Referenced Data Sets (Document No. CGSB-171.3) was developed and adopted for general use in 1994. The standard [CGSB01] describes the data that should be collected to describe data sets, but it does not contain a database schema that could be used for implementing this standard in an information system. The standard provides a printed form that can be used to describe digital georeferenced data sets.

The CGSB standard contains the information given in Table 1.

1.Identification	•Title of the Data Set •Owner •Contact Person/Position
2.Description	•Category of the Data Set •Information on the Data Set Content •Aerial Coverage •Vertical Coverage
	•Time Coverage •Data Updating •Size of the Data Set •Details of Data Collection •Data Quality
3.Availability	•Access Policy •Charges •External Communication •Data Set Media
4.Structure	•Host Computer •Structure Type •Data Structure
5.References	•Documentation Available •Demonstration/Tutorial Available

Table 1 : Contents of CGSB Directory Information

Of particular interest for the users of any metadata standard is the description of the category of the data set. For the CGSB standard the following categories are used to describe the principal theme or themes of the dataset. If none of the themes listed are appropriate then the standard allows for a word description of the data set theme. The categories are:

- Agriculture
- Anthropology
- Aquaculture
- Archaeology
- Arts and Crafts
- Biology
- Botany
- Boundaries
- Buildings
- Climatology and Meteorology
- Commerce
- Communications
- Cryogenics
- Ecology
- Economics
- Education
- Engineering
- Environmental Protection
- Environmental Pollution
- Folklore, manners and customs
- Forestry
- Geodesy
- Geography
- Geology
- Geomorphology
- Geophysics
- Health and Medicine
- Human Ecology
- Hydrocarbons
- Hydrology
- Hydrography
- Industry
- Land Use
- Language
- Manufacturing
- Military
- Mining and Metallurgy
- Natural disasters
- Natural resources

- Oceanography
- Pests and diseases
- Physics
- Political science
- Population
- Recreation
- Remote sensing
- Sociology
- Soils
- Transportation
- Utilities
- Waste
- Water
- Wildlife
- Zoology
- Other

4. METADATA DISSEMINATION AND MAINTENANCE

Both simple and sophisticated software for the dissemination of metadata has been developed by a number of organizations and using a number of different approaches over the past few years. The type and method of metadata dissemination differs with the organizations needs, but in general, in countries where there is readily available access to the Internet, metadata access tools are provided using Internet protocols. In countries or applications where there is less access to the network, the software and metadata are distributed on rigid media such as floppy disk or CD-ROM.

Figure 1 indicates the current thinking with regards to metadata database dissemination and maintenance. This structure allows for multiple users to access the metadata database search and query engine. It also allows the user to access numerous databases. These databases do not necessarily need to be in identical formats, however when the databases are not in analogous formats, this complicates the search engine software.

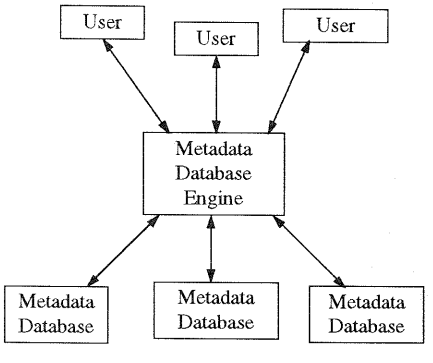

Figure 1: System Structure for metadata dissemination and maintenance.

4.1 Canadian Activities

Several metadata standards have been developed that use the Canadian CGSB standard. These include the following.

4.1.1 MetaView/GIS

The Geographic Information Systems and Services Division of Geomatics Canada has developed the MetaView/GIS [EFFA01] spatial browser for browsing and locating spatial data sets. The MetaView software tool provides a user friendly interface for searching and querying of a metadata base in the Canadian CGSB Metadata standard. The software is configured in the client server model to run over the network and the databases themselves can also be networked.

One of the main differences between the MetaView/GIS and other directory services is the ability for the user to select data based on a region or area. This means that the software must have two databases, one which contains the metadata and the other containing map data. The system supports spatial queries from the global level to the local level using the map display.

4.1.2 Data Catalogue Viewer

The Manitoba Land Related Information System (MLRIS) which is sponsored by a number of Manitoba based organizations has developed a centralized Information Utility for managing spatial information [LINN01]. The MLRIS employs a Data Catalogue Viewer to allow the user to locate and select data for the users requirements. The viewer has a search engine, a browse capability, a keyword browse and a coverage browse capability.

The Data Catalogue Viewer utilizes the Canadian CGSB metadata standard for storing and retrieving of the spatial metadata. The system also has some capability to disseminate spatial information as well. There does not yet appear to be any capability to store the metadata in a distributed or network environment (i.e. the metadata base is stored on a single system).Other capabilities for data security and maintenance are available.

4.1.3 GCNet directory

Geomatics Canada through the Canada Centre for Remote Sensing has developed a metadata browser for use across the World Wide Web [GCNE01]. The user can use either an alphanumeric interface of a graphical user interface such as Netscape or Mosaic to access the system. The user can query for the location of remote sensing data and products in Canada and abroad.

The GCNet directory service function does not use the CGSB standard, but it uses the International Directory Network (IDN) Master Directory (MD) developed by NASA. The MD is primarily used for describing remotely sensed imagery and is used in several countries and in many disciplines. The GCNet is coordinated through the Committee on Earth Observation (CEOS) group internationally.

4.2 International activities

Internationally, there are a number of directory services using various browsers, but these systems are not coordinated with respect to exchanging metadata between the systems.

4.2.1 UNEP GRID

UNEP plans to develop a version of "One-Stop Shopping" for environmental information via the World Wide Web. The UNEPNet concept will include the use of metadata and metadata browser software tools developed among the network of UNEP GRID centres. The Grid Metadata directory [MCKE01], both data and software will be disseminated publicly according to the facilities available to the different users. For those with Internet access, the service will be provided on the World Wide Web through links such as URL, by Telnet or by Email. Other users will be served by self-contained software and reference data that will be provided on request for PCs or workstations.

The UNEP metadata directory will serve as a library card catalogue of environmental information. It will contain "card entries" (or metadata descriptions) of institutes and data sets. Similar to the library card catalogue, the metadata directory will allow users to search for environmental information by institute name or data set (title), contact person (author), theme, keyword and location (subject), as well as other criteria.

The UNEP metadata directory is a PC reference system that provides environmental information to users via a user-friendly interface. The tool is designed to allow simple data entry (there are only ten mandatory fields for an institute entry or a dataset entry), uncomplicated data queries and easy data exchange between any organizations that use a basic international metadata standard, such as the CEOS Directory Interchange Format (DIF).

4.2.2 Spatial FreeWAIS

The US Geological Survey has developed a software tool called Spatial FreeWAIS [NEBE01], for the indexing, querying and retrieval of information in WAIS using spatial and free-text constructs. Some enhancements were made to the software in Germany and the software was made to conform the US FGDC Metadata standard. Spatial queries are made using Boolean type operators to select information within a bounding rectangle.

With the requirement for spatial data producers to also provide a standard descriptive metadata outline of their data, the availability of the metadata for the metadata database will provide the users with a substantial database from which to select information. This development is using the WWW tools for access to the various metadata sets and thus a single user should have access to a large number of metadata entries.

4.2.3 GeoWeb project

The State University of New York at Buffalo has developed software tools under the GeoWeb project [PLEW01] in order to implement a working prototype of the data clearinghouse concept. The project describes the four elements of the clearinghouse initiative. These elements include: the spatial data, the metadata, the metadata index and the search interface. This search interface allows the user to enter queries in a number of fashions, from direct entry to pointing to an area of interest.

The desire to have an open interface system and to have remote access, motivated the project to make the World Wide Web as the implementing technology for the metadata browser. The desire is also to have the browser access all on-line metadata databases as if they were one massive database. In addition the user interface will have a number of different metaphors for querying including map based spatial queries, place name queries or keyword search. The concept of using the Internet allows the user to operate the query engine on a computer that has only links to external metadata databases.

5. SUMMARY

There appears to be no single body responsible for the international coordination of either metadata or metadata search engines. There are a number of localized efforts, but little progress has been made in the area of interoperability. Perhaps the best method of obtaining some coordination on the metadata standardization front is through ISO/TC 211. This is an internationally accredited body that develops standards that are recognized in many countries around the world.

There is probably no need to standardize on the search or query engine. These simply only need to communicate via the network in some ordered fashion. There will likely be some requirement for local differences or language changes.

In summary, there are a number of different efforts ongoing in Canada and throughout the world. There are several good metadata standards and perhaps there are possibilities of harmonizing these in the future years. Canada, the US and UNEP among others have selected and adopted standards for metadata. The trick will be to coordinate the development of these standards to allow for the harmonization of these databases over the upcoming years.

6. ACKNOWLEDGMENTS

The author would like to thank the following people for their input to the paper generally through personal communication: James Mckenna, United Nations Environment Programme; Valerie Hume, Department of Indian Affairs and Northern Development; and Roger Tomlinson, Tomlinson and Associates.

7. BIBLIOGRAPHY

[CGSB01] Canadian General Standards Board; *"Directory Information Describing Digital Geo-Referenced Data Sets";* Document No C**/CGSB-171.3; CGSB, Ottawa, Canada, K1A 1G6.

[EFFA01] Samual Effah, Cherian Chaly and Wenjing Zhu; *"MetaView/GIS: The GIS Spatial Data Browser, Its Design, Implementation and Operation";* Proceedings of the ISPRS, Commission II Symposium, June 6-10, 1994,

[GCNE01] R. Boudreau, A. Buffam, and J. Guenette; *"Earth Observation Data Management at the Canada Centre for Remote Sensing";* Proceedings of the ISPRS, Commission II Symposium, June 6-10, 1994, Ottawa; ISPRS Archives Vol 30, Part 2.

[IACG01] Inter-Agency Committee on Geomatics under contract to Roger Tomlinson Assoc.; *"Report on Current Status and Trends in Federal Digital Geographic Data in Canada";* IACG, Geomatics Canada, Ottawa, K1A 0E9.

[LINN01] Linnet Geomatics International Inc; *"The Information Utility - Providing land related information and services to Manitoba and the World";* LGI, 1600-444 St Mary's Ave, Winnipeg, Manitoba, R3C 3T1.

[MCKE01] James McKenna; *"Data and Information Management";* UNEP - EarthViews, Vol 2, No. 2; May 1995.

[NEBE01] Douglas D Nebert; *"Serving Digital Map Information through the World Wide Web and Wide Area Information Server Technology";* Spatial Data Support Unit, USGS, Reston Virginia, USA.

[PLEW01] Brandon Plewe; *"The GeoWeb Project; Using WAIS and the World Wide Web to aid location of Distributed Data Sets";* Campus Information Services, State University of New York, Buffalo, New York, USA.