XML (eXtensible Markup Language) for Geospatial Metadata

by

Ying Teng
Bradford G. Nickerson

TR00-131, January 2000

Faculty of Computer Science
University of New Brunswick
Fredericton, N.B. E3B 5A3
Canada

Phone: (506) 453-4566
Fax: (506) 453-3566
E-mail: fcs@unb.ca
www: http://www.cs.unb.ca

Contents

1. Introduction to XML

1.1 What is XML?

XML (Extensible Markup Language) is an ISO (International Organization for Standardization) compliant subset of SGML (Standard Generalized Markup Language) and is designed to make it easy and straightforward to use SGML on the Web [1]. XML is a meta-markup language that provides a format for describing structured data. This facilitates more precise declarations of content and more meaningful search results across multiple platforms. In addition, XML will enable a new generation of Web-based data viewing and manipulation applications [2, 3].

1.2 Advantages of XML

1.2.1 Extensible

In XML you can define an unlimited set of tags. While HTML tags can be used to display a word in bold or italic, XML provides a framework for tagging structured data. As XML tags are adopted throughout an organization's intranet, and by others across the Internet, there will be a corresponding ability to search for and manipulate data regardless of the applications within which it is found. Once data has been located, it can be delivered over the Internet and presented in a WWW browser in any number of ways, or it can be handed off to other applications for further processing and viewing.

1.2.2 Structural representation of data

XML provides a structural representation of data that has proved broadly implementable and easy to deploy. Industrial implementations in the SGML community and elsewhere demonstrate the intrinsic quality and industrial strength of XML's tree-structured data format. XML documents are easy to create. If you are familiar with HTML, you can quickly learn to author in XML. In the following example, XML is used to describe a weather report. This file can be saved with an extension of XML, like Weather.xml.

```
<weather-report>
         <date>March 25, 1998</date>
         <time>08:00</time>
         <area>
            <city>Seattle</city>
            <state>WA</state>
            <region>West Coast</region>
            <country>USA</country>
         </area>
         <measurements>
            <skies>partly cloudy</skies>
            <temperature>46</temperature>
            <wind>
               <direction>SW</direction>
               <windspeed>6</windspeed>
            </wind>
            <h-index>51</h-index>
            <humidity>87</humidity>
            <visibility>10</visibility>
            <uv-index>1</uv-index>
         </measurements>
 </weather-report>
```

Rather than describing the order and fashion in which the data should be displayed, the tags indicate what each item of data means (whether it is a <date> element, an <area> element, and so forth). Any receiver of this data can then decode the document, using it for their own purposes. For example, an individual might use it to make plans for the day, while a weather researcher might use it as data in a historical record of Seattle.

XML is defined by the World Wide Web Consortium (W3C), ensuring that structured data will be uniform and independent of applications or vendors. This resulting interoperability is kick-starting a new generation of business and electronic-commerce Web applications.

Once the data is on the client desktop, it can be manipulated, edited, and presented in multiple views, without return trips to the server. Servers now become more scalable, due to lower computational and bandwidth loads. Also, since data is exchanged in the XML format, it can be merged from different sources.

1.2.3 Data is separated from the presentation and the process

XML maintains the separation of the user interface from the structured data. Hypertext Markup Language (HTML) specifies how to display data in a browser, XML defines the content. For example, in HTML you use tags to tell the browser to display data as bold or italic; in XML you only use tags to describe data, such as city name, temperature, and barometric pressure. In XML, you use stylesheets such as Extensible Style Language (XSL) and Cascading Style Sheets (CSS) to present the data in a browser. XML separates the data from the presentation and the process, enabling you to display and process the data as you wish by applying different style sheets and applications.

This separation of data from presentation enables the seamless integration of data from diverse sources. Customer Information, purchase orders, research results, bill payments, medical records, catalogue data, and other information can be converted to XML on the middle tier. Data encoded in XML can then be delivered over the Web to the desktop. Legacy information stored in mainframe databases or documents can easily be adapted to XML format.

1.3 Applications of XML

1.3.1 XML as data

XML is gaining popularity as a data storage and exchange format as well as a document markup language. Extra tags are required in an XML file besides the content, so it may not be suitable for large sets of records in a spatial data warehouse. Developers may prefer to use a traditional database, and convert its content into XML on the fly or use XML to store data extents.

1.3.2 Describing "domain document"

XML is valuable on the Internet because it provides a universal standard mechanism for describing a "domain document". Different XML domain-specific vocabularies or XML dialects (e.g., mathematical, chemical, medical and geospatial metadata markup) can be invented for any purpose. These domain documents can be understood by all browsers with XML parsers so that data with specific meaning for this community can be freely exchanged and shared within the community of interest. This results in more efficient data exchange and more accurate searching on the web.

The following is an example of MathML [4]. It contains token element tags: <mi> --identifier, <mn> -- number, <mo> -- operator, fence, or separator and a general layout schemata tag: <mrow> -- horizontally group any number of subexpressions.

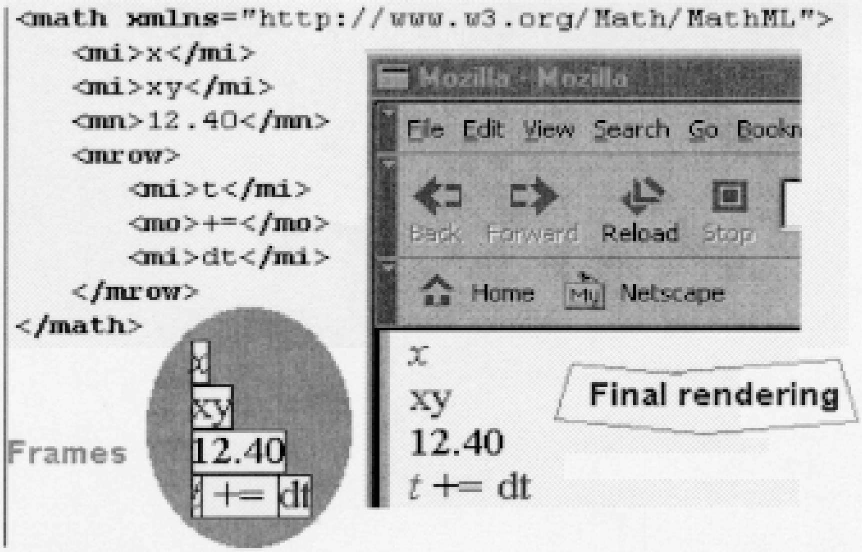

Figure 1. An example of MathML from [4]

1.3.3 Combining different sources

XML enables structured data from different sources to be combined in an efficient manner. Since XML is database-neutral, it is likely that it will play a major role in connecting heterogeneous databases. With XML content markup, queries are more likely to retrieve relevant files due to contextual information. Search engines could retrieve a specific portion of file; they also could be much faster if the added context eliminates numerous irrelevant matches.

1.3.4 Non-human interaction

We should not always assume that it will just be humans looking at XML pages in browsers over the web. XML files describe their contents, so it is possible for non-human user agents, such as the programs sent out by search engines, to process the information in the file. In the specific case of search engines, it will mean that they can provide more accurate results to queries.

2. FGDC (Federal Geographic Data Committee) CSDGM (Content Standard for Digital Geospatial Metadata)

2.1 General information

A metadata standard is simply a common set of terms and definitions that describe geospatial data. The Federal Geographic Data Committee (FGDC) recently adopted Content Standard for Digital Geospatial Metadata (CSDGM) as a metadata standard [5]. This standard specifies the information content of metadata for a set of digital geospatial data and provides a consistent approach and format for the description of data characteristics. The standard provides a way for data users to know what data are available, whether the data meet their specific needs, where to find the data and how to access the data.

2.2 Purpose and goals

The objectives of the standard are to provide a common set of terminology and definitions for the documentation of digital geospatial data. The standard establishes the names of data elements and compound elements (groups of data elements) to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements [5].

The standard was developed from the perspective of defining the information required by a prospective user to determine the availability of a set of geospatial data, to determine the fitness of the set of geospatial data for an intended use, to determine the means of accessing the set of geospatial data, and to successfully transfer the set of geospatial data. The standard does not specify the means by which this information is organized in a computer system or in a data transfer, nor the means by which this information is transmitted, communicated, or presented to the user.

2.3 What is the standard?

The standard is designed to describe all possible geospatial data. There are 334 different elements in the FGDC standard, 119 of which exist only to contain other elements. These compound elements are important because they describe the relationships among other elements. CSDGM uses both SGML and XML for structuring information.

Major sections of the CSDGM are: identification information, data quality information, spatial data organization information, spatial reference information, entity and attribute information, distribution information, and metadata reference information.

Minor sections of the CSDGM include: citation information, time period information, contact information.

2.4 Tools

Figure 2 shows the architecture of several tools available on the Web page of metadata tools [6]. Most of the tools are written in Standard C.

2.4.1 cns

Cns (chew and spit) is a pre-parser for formal metadata designed to assist metadata managers in converting records that cannot be parsed by mp into records that can be parsed by mp. It takes as input a poorly-formatted metadata file and, optionally, a list of element aliases, and outputs (1) a

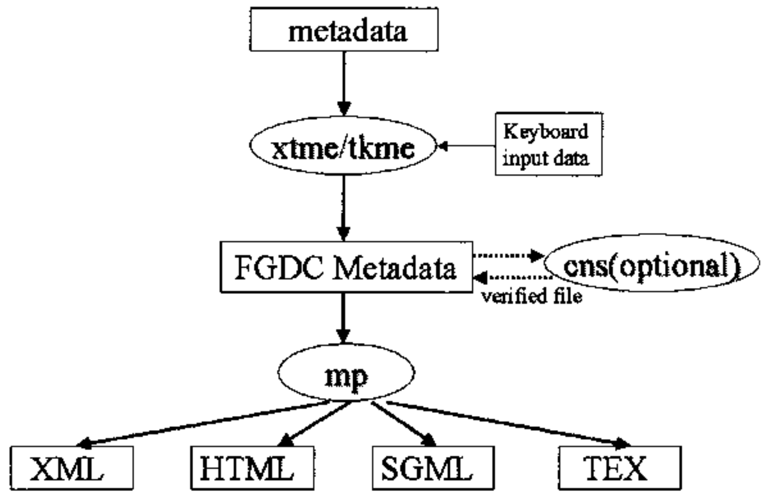

Figure 2. Metadata tools

metadata file that can be read by both mp and xtme or tkme and (2) a file listing all of the lines that it couldn't figure out where to put.

2.4.2 mp

mp is designed to parse metadata encoded as indented text, check the syntactical structure against the the FGDC Content Standard for Digital Geospatial Metadata, and reexpress the metadata in several useful formats (HTML, SGML, TEXT and XML). The output is suitable for viewing with a web browser or text editor. It runs on UNIX systems and on PCs running Windows 95, 98, or NT. mp generates a textual report indicating errors in the metadata, primarily in the structure but also in the values of some of the scalar elements (i.e. those whose values are restricted by the standard).

2.4.3 xtme/xkme

xtme is an editor for formal metadata, that is, structured documentation conforming to the Content Standard for Digital Geospatial Metadata developed by the Federal Geographic Data Committee (FGDC). This editor is written using Standard C to run on UNIX systems with the X Window System, release 11, version 5 or later. It utilizes the X Toolkit and the Athena widget set, which are available on nearly all X distributions. The editor is intended to simplify the process of creating metadata that conform to the standard. Its output format is the input format for mp.

Tkme is a port of xtme to run under Tcl/Tk, thus bringing the excellent UNIX Athena widget based tool to the PC platform. The characteristics, appearance and strengths of this tool are nearly identical to xtme, and the capabilities of xtme are generally applicable to tkme. The major differences from xtme are: tkme allows the user to 'Open' a metadata file from the tkme File menu and to recall recently edited files; configuration files to support metadata extensions can be loaded from the File menu before opening the metadata file; tkme has a menu to allow font customization for the tool window, menus and help. Figure 3 is an edit screen shot of tkme [6].

2.5 A small example

The FGDC metadata file (063c03.txt, see Appendix II) was obtained by using software developed by Ying Teng in the Faculty of Computer Science, University of New Brunswick. It converts the demonstration data set [7] (063c03.met in this example, see Appendix I) metadata in NTDB (National Topographic Data Base) format to FGDC format. It is written in Borland C++, using an ad hoc translation process. A more formal process using a context file grammar will be investigated in the future. The FGDC metadata file was converted to xml file (063c03.xml, see

Appendix III) using metadata tool mp. The procedure is illustrated in figure 4.

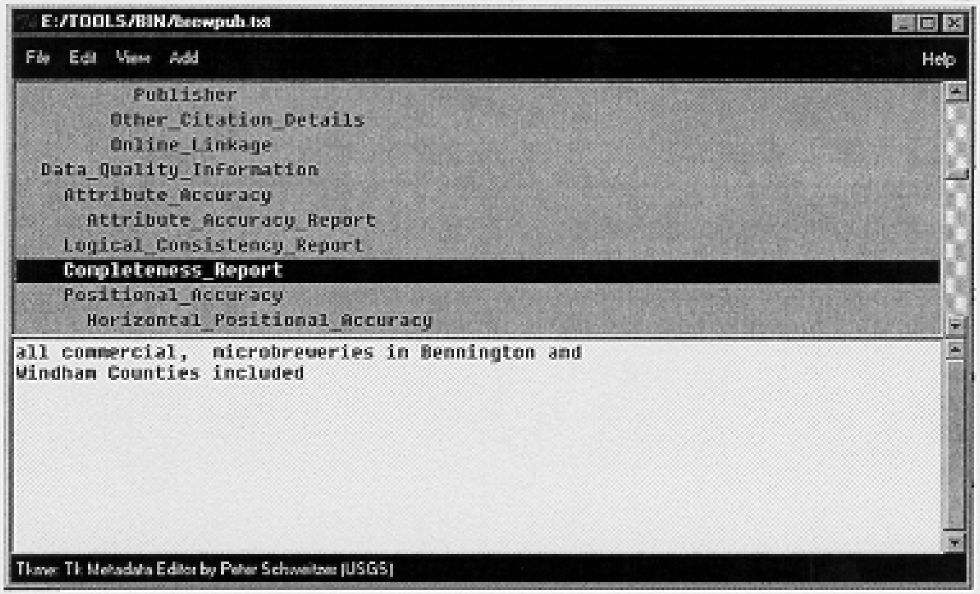

Figure 3. An edit screen shot of tkme from [6]

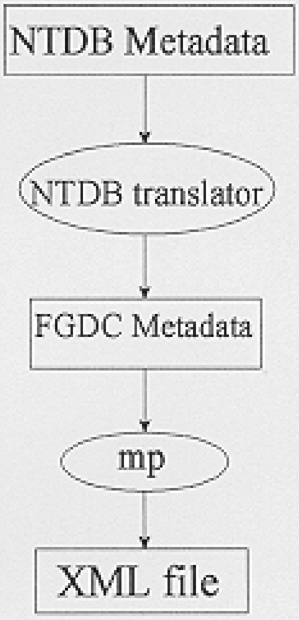

Figure 4. Process to translate NTDB metadata to FGDC XML

3. ISO (The International Organization for Standardization) Metadata Standard

3.1 General information

International Standard ISO 19115 was prepared by Technical Committee ISO/TC 211, Geographic information/Geomatics [8].This standard identifies the metadata required to describe digital geographic data. Metadata is applicable to independent datasets, aggregations of datasets, individual geographic features, and the various classes of objects which compose a feature. Metadata is mandatory (required) for each geographic dataset and may, optionally, be provided for aggregations of datasets, features, and parts of features.

This International Standard defines [8]:

- mandatory and conditional metadata packages, metadata entities, and metadata elements — the minimum set required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data).
- optional metadata elements — to allow for a more extensive standard description of geographic data, if required.
- a method for extending metadata to fit specialised needs.

Though this International Standard is applicable to digital data, its principles can be extended to many other forms of geographic data such as maps, charts, and textual documents.

3.2 Purpose and goals

The objective of the standard is to provide a structure for describing digital geographic datasets and a procedure to extend the existing structure so that users will be able to locate, select, purchase, and access geographic data, determine whether the data in a holding will be of use to them, and use it in the most efficient way. By establishing a common set of metadata terminology, definitions, and extension procedures, this standard will promote the proper use and effective retrieval of geographic data. Supplementary benefits of this standard for metadata are to facilitate the organization and management of geographic data and to provide information about an organization's dataset to others. This standard for the documentation of data furnishes data producers the appropriate information for them to characterize geographic data, and it makes possible dataset cataloguing enabling data discovery, retrieval and reuse.

3.3 What is the standard?

The metadata are specified using UML in a hierarchy to establish relationships and an organization for the information [9]. The metadata is categorised into eight main sections: Identification information, DataQuality information, Lineage information, Spatial Organization information, Spatial Reference information, Entity/Attributes information, Distribution information, and Metadata Reference information; and five supporting entities: Citation, Responsible Party, Address, Extent (BP, VT, Tmp), and On-line resource. The sections are subdivided into entities, which are further divided into metadata elements that carry the individual units of metadata. The supporting repeatable entities provide common information called for by main sections, and are never used alone [8].

Figure 5 defines the class "Metadata" and shows containment relationships with the other metadata classes which, in aggregate, define geographic metadata.

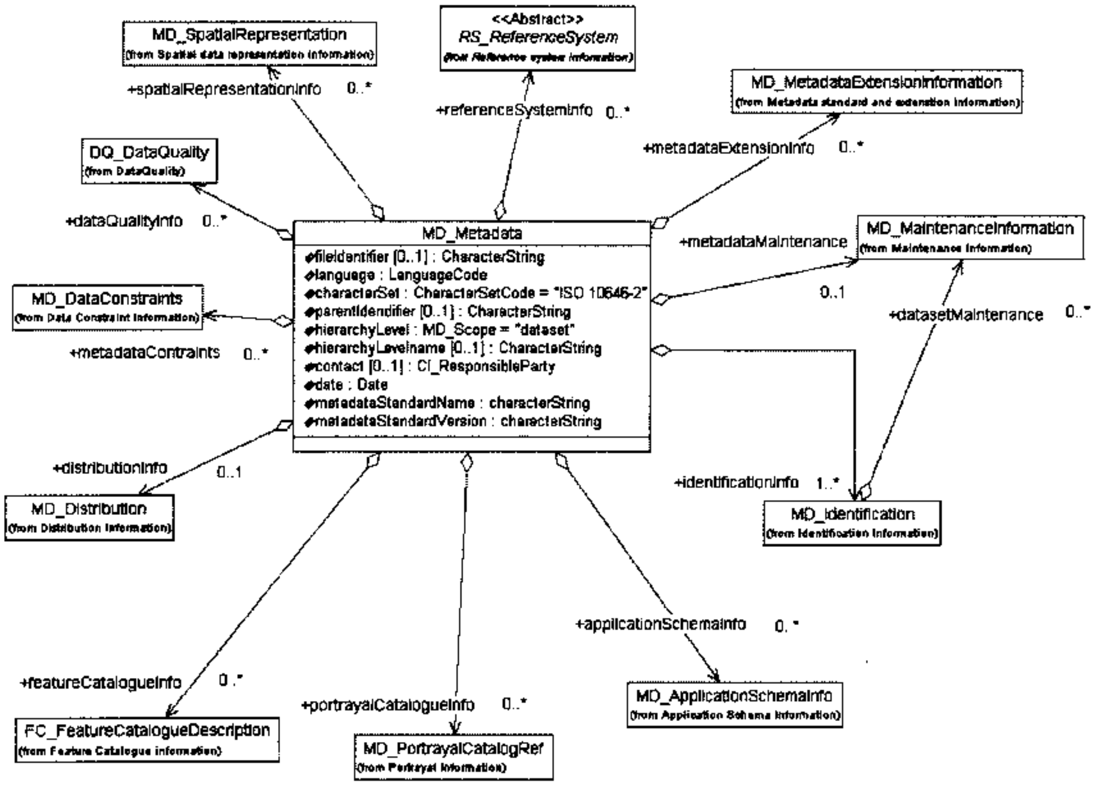

Figure 5. Metadata entity set information from [9]

Figure 6 defines the metadata classes required to identify a dataset. It also defines a specialization subclass for identifying images.

Figure 7 defines metadata required to describe the mechanism used to represent spatial information.

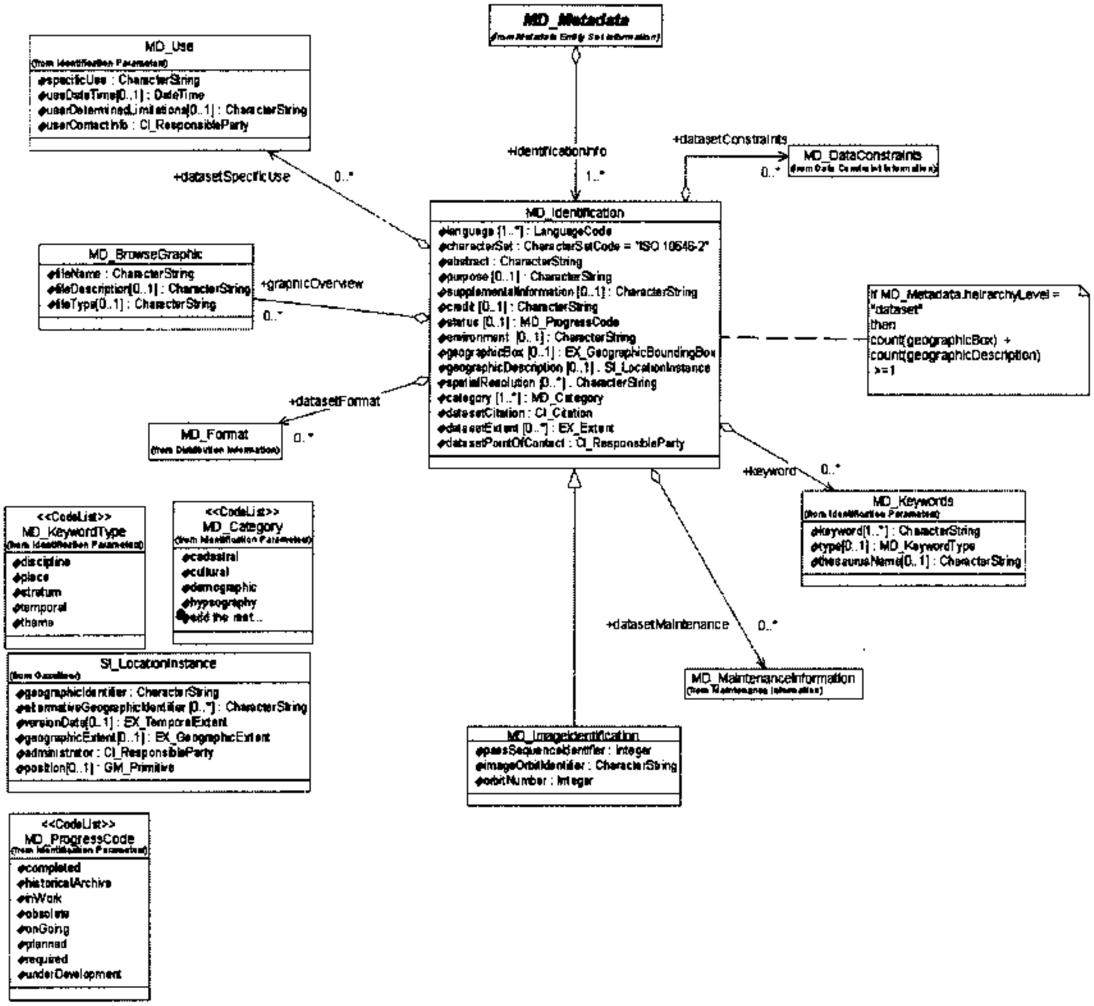

Figure 6. MD_Identification from [9]

3.4 A small example

The example (see Appendix IV) is metadata for a vector dataset - essential profile [9].

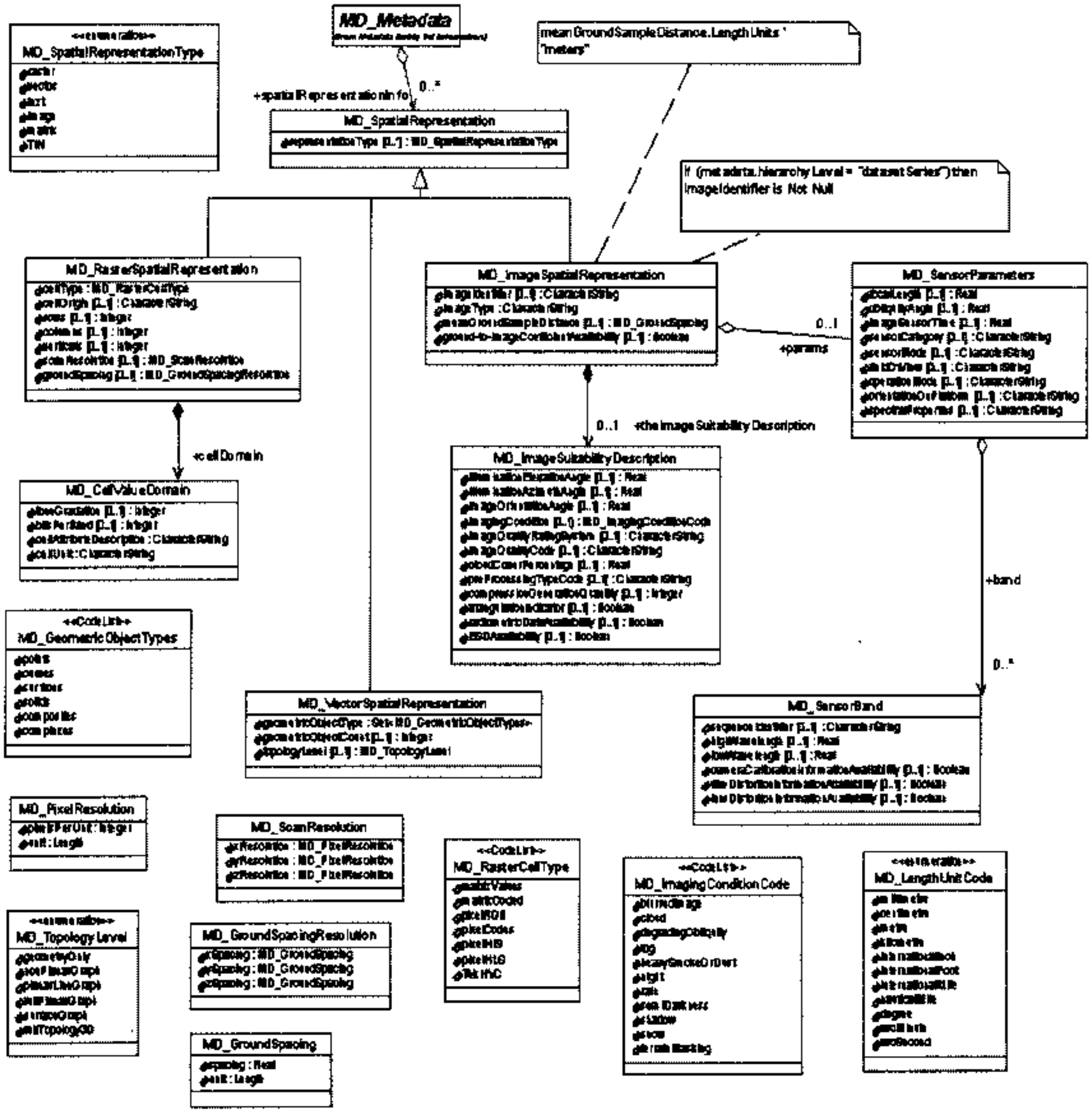

Figure 7 MD_SpatialRepresentation from [9]

4. FGDC/ISO Metadata Standard Harmonization

4.1 Comparison of the two standards

ISO and FGDC metadata standards have some differences.

The ISO standard has one more section (lineage information) than the FGDC CSDGM. This lineage information contains information about the usage, sources, and production processes used in producing a dataset [8]. The ISO standard has many more "Optional" elements. The ISO standard also addresses many known deficiencies in the FGDC CSDGM (e.g. Raster, Imagery). For example, the FGDC CSDGM does not provide a compound element that contains the coordinates of the four corners of satellite images.

ISO has 2 levels of compliance: "catalogue" and "complete/full". Conformance Level 1 is the minimum metadata required to uniquely identify a dataset (independent dataset, dataset series, or individual geographic features). This level of conformance shall be used ONLY for the purposes of cataloguing datasets and dataset series and to support data clearinghouse activities facilitating data discovery. Conformance Level 2 provides the metadata required to document a dataset completely (independent dataset, dataset series, or individual geographic features and attributes). This level of conformance fully defines the complete range of metadata required to identify, evaluate, extract, employ, and manage geographic information. A data producer would typically provide metadata at this level.

FGDC CSDGM and ISO standard also have terminology and element name differences as illustrated in the above examples.

4.2 Activities in progress

Since 1996, the International Organization for Standardization (ISO) Technical Committee

211, project item 19115 (formerly 15046-15), has been drafting an International Metadata Standard. The FGDC, through the United States ANSI L1 committee, has been an active participant in this project and has had a significant impact on the working drafts of the ISO Metadata Standard. Much of the draft ISO Metadata Standard is closely aligned with the CSDGM [5]. In June 1998, the status of the ISO Metadata Standard progressed from that of a working draft to a committee draft [8]. With the status changed to that of a committee draft, the proposed ISO Metadata Standard was reviewed in the United States by over 230 individuals; with over 500 comments submitted by the reviewers. The ISO Metadata Standard has been modified based upon the US and other national comments and has now been reissued as the 2nd Committee Draft [9]. This document is now ready to be reviewed again by the national bodies.

Numerous organizations plan to use the ISO Metadata Standard once it has been approved by the ISO Standards Committee. The FGDC is committed to harmonize the CSDGM with the ISO Metadata Standard and has numerous activities ongoing to assure that the harmonization is successful [5]. To protect the significant already existing metadata investment, it is important to assure that the proposed ISO Metadata Standard allows the maximum compatibility with existing FGDC compliant metadata records.

5. Conclusion

The ISO standard is necessary to meet "international" requirements. This International Standard defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. This International Standard is applicable to the cataloguing of datasets, clearinghouse activities, and the full description of datasets. This International Standard is also applicable to geographic datasets, dataset series, and individual geographic features and feature properties.

Since XML is a meta-markup language that provides a format for describing structured data, it is a suitable technique for the implementation of the ISO metadata standard.

Tools that are useful for the ISO metadata standard would be similar to those for the FGDC metadata: editors that can create the ISO metadata file from keyboard input or from FGDC metadata file; software that can create XML, text, HTML or SGML files from the ISO metadata files.

6. References

1. World Wide Web Consortium, "Extensible Markup Language (XML) Version 1.0", Recommendation 10-February-1998, http://www.w3.org/TR/1998/REC-xml-19980210.html

2. MSDN Online Web Workshop, "Why XML" http://msdn.microsoft.com/workshop/xml/toc.htm

3. Pontus Norman, "A study of eXtensible Markup Language (XML)", Thesis for Master of Science degree in Computer Science at the Department of Teleinformatics, Royal Institute of Technology in Sweden, Decerno AB, Näsby Park, Sweden. February 25, 1999

4. Roger B. Sidje, "Implementation of MathML in Mozilla: Progress Report", Sep 18 1999 http://www.mozilla.org/projects/mathml/update.html

5. Federal Geographic Data Committee, "Content Standard for Digital Geospatial Metadata", document FGDC-STD-001-1998, FGDC Secretariat, US Geological Survey, 590 National Center, Reston, VA, 1998, available from http://www.fgdc.gov/metadata/contstan.html6. "Metadata Tools for Geospatial Data" http://badger.state.wi.us/agencies/wlib/sco/metatool/mtools.htm

7. National Resources Canada, "National Topographic Data Base, 1:50 000 and 1:250 000" http://www.ccg.rncan.gc.ca/ext/html/english/products/ntdb/ntdb.html

8. ISO/TC 211 Secretariat, "ISO/TC 211 Geographic information/Geomatics CD 15046-15, Geographic information - Part 15: Metadata", Committee draft for comments, ISO/TC 211 N538, available from Norwegian Technology Standards Institution, Oslo, Norway, http://www.statkart.no/isotc211/, July 9th, 1998, 161 pages.

9. ISO/TC 211 Secretariat, "ISO/TC 211 Geographic information/Geomatics 2. CD 19115, Geographic information - Metadata", Committee draft for comments, ISO/TC 211 N831, available from Norwegian Technology Standards Institution, Oslo, Norway, http://www.statkart.no/isotc211/, November 30th, 1999, 118 pages.

APPENDIX I

The demonstration data set 063c03.met [7] metadata in NTDB (National Topographic Data Base) format.

```
BEGIN          FILE
BEGIN          TERRITORY_SECTION
NTS            063C03
DATA_SET_NAME  SWAN RIVER
PROVINCE       MB (Manitoba)
ZONE_NUMBER_1  14
ZONE_NUMBER_2  -1
PCT_OF_LAND    100
SPECIAL_LIMITS N (Normal)
END            TERRITORY_SECTION
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
BEGIN          DATA_SET_SECTION
EDITION_VERSIO 2.02
NTDB_SPEC      3.1
DATE_AVAILABLE 23-SEP-98
FORMAT         IFF-BNDT-3.0
FORMAT         CCOGIF-P3.0
UNIT_CONTOURS  P (Foot(feet))
CONTOUR_INTERV 25
CONT_AUXILIARY -1
DIMENSION      2D
MAP_EDITION    2
COMMENT
END            DATA_SET_SECTION
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
BEGIN          INTEGRATION_SECTION
NORTH_EDGE     O (Yes)
SOUTH_EDGE     I (Imperfect)
EAST_EDGE      O (Yes)
WEST_EDGE      O (Yes)
END            INTEGRATION_SECTION
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
BEGIN          POLYGON_SECTION
NB_POLYGONS    2
BEGIN          POLYGON
ID_POLYGON     0001
COORDINATES    362705 5762926 345545 5763428 328386 5763989 329344 5791789
COORDINATES    346408 5791229 363471 5790728 362705 5762926
```

```
ENTITIES        P 1-1849,1856-2047
ENTITIES        L 1-1809, 1825-2047
ENTITIES        S 1-1829, 1834-2047
SOURCE_TYPE     REPRO (Reprographic material)
SOURCE_NAME     063C03-ED2
VALID_DATE      1976/00
PLAN_ACCU_QUAL E (Estimated)
PLAN_ACCURACY   50
ALTI_ACCU_QUAL E (Estimated)
ALTI_ACCURACY   5
RES_PLAN_ACC    50
ACTION          ACQ.COMP.SCAN (Complete acquisition of entities by scanning)
IMPACT_ACT_C    S (Systematic)
IMPACT_ACT_P    N (No)
POL_ED_VER      2.00
COMMENT
END             POLYGON
!
BEGIN           POLYGON
ID_POLYGON      0002
COORDINATES     362705 5762926 345545 5763428 328386 5763989 329344 5791789
COORDINATES     346408 5791229 363471 5790728 362705 5762926
ENTITIES        P 1850-1855
SOURCE_TYPE     BDN+CARTE (Digital data base and paper map)
SOURCE_NAME     BDTC-MB-1995
VALID_DATE      1976/00
PLAN_ACCU_QUAL I (Unknown / Non-applicable)
PLAN_ACCURACY   -1
ALTI_ACCU_QUAL I (Unknown / Non-applicable)
ALTI_ACCURACY   -1
RES_PLAN_ACC    -1
ACTION          ACQ.TO (Acquisition of toponymy)
IMPACT_ACT_C    S (Systematic)
IMPACT_ACT_P    N (No)
POL_ED_VER      2.00
COMMENT
END             POLYGON
!
END             POLYGON_SECTION
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
BEGIN           THEMES_SECTION
NB_THEMES       14
BEGIN           THEME
NAME            AD (Designated Area)
AVAIL_THEME     O (Yes)
RESOLUTION      50000
NB_KM           17
```

```
NB_POINTS        2
END              THEME
!
BEGIN            THEME
NAME             CH (Roads)
AVAIL_THEME      O (Yes)
RESOLUTION       50000
NB_KM            207
NB_POINTS        0
END              THEME
!
BEGIN            THEME
NAME             CO (Manmade Features)
AVAIL_THEME      O (Yes)
RESOLUTION       50000
NB_KM            36
NB_POINTS        1737
END              THEME
!
BEGIN            THEME
NAME             FO (Relief and Landforms)
AVAIL_THEME      O (Yes)
RESOLUTION       50000
NB_KM            0
NB_POINTS        1
END              THEME
!
BEGIN            THEME
NAME             GE (General)
AVAIL_THEME      O (Yes)
RESOLUTION       50000
NB_KM            124
NB_POINTS        0
END              THEME
!
BEGIN            THEME
NAME             HD (Hydrography)
AVAIL_THEME      O (Yes)
RESOLUTION       50000
NB_KM            923
NB_POINTS        1
END              THEME
!
BEGIN            THEME
NAME             HP (Hypsography)
AVAIL_THEME      O (Yes)
RESOLUTION       50000
```

```
NB_KM          1769
NB_POINTS      0
END            THEME
!
BEGIN          THEME
NAME           LA (Administrative Boundaries)
AVAIL_THEME    N (No)
RESOLUTION     -1
NB_KM          0
NB_POINTS      0
END            THEME
!
BEGIN          THEME
NAME           RE (Power Network)
AVAIL_THEME    O (Yes)
RESOLUTION     50000
NB_KM          38
NB_POINTS      0
END            THEME
!
BEGIN          THEME
NAME           RF (Rail Network)
AVAIL_THEME    O (Yes)
RESOLUTION     50000
NB_KM          51
NB_POINTS      0
END            THEME
!
BEGIN          THEME
NAME           RR (Road Network)
AVAIL_THEME    O (Yes)
RESOLUTION     50000
NB_KM          670
NB_POINTS      0
END            THEME
!
BEGIN          THEME
NAME           SS (Water saturated soils)
AVAIL_THEME    O (Yes)
RESOLUTION     50000
NB_KM          34
NB_POINTS      0
END            THEME
!
BEGIN          THEME
NAME           TO (Toponymy)
AVAIL_THEME    O (Yes)
```

```
 RESOLUTION     50000
 NB_KM          0
 NB_POINTS      43
 END            THEME
!
 BEGIN          THEME
 NAME           VE (Vegetation)
 AVAIL_THEME    O (Yes)
 RESOLUTION     50000
 NB_KM          1381
 NB_POINTS      0
 END            THEME
!
 END            THEMES_SECTION
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 END            FILE
```

APPENDIX II

The 063c03.txt metadata file in FGDC format obtained from the software developed by Ying Teng in the Faculty of Computer Science, University of New Brunswick using 063c03.met as input file.

```
Identification_Information
   Citation:
      Citation_Information:
         Originator: Canadian Government Agencies: NRCan
         Publication_Date: 1998-09-23
         Title: SWAN RIVER
         Edition:2.02
         Online_Linkage:
   http://www.ccg.rncan.gc.ca/ext/html/english/products/ntdb/demontd5.html
   Description:
      Abstract:
       The National Topographic Data Base (NTDB) is a digital data base
   developed by Geomatics Canada. It covers the entire Canadian landmass
   and contains the features normally found on topographic maps at the
   scales of 1:50 000 and 1:250 000: hydrography, hypsography
   (contours), vegetation, the road network, roads, the rail network, the
   electric power network, designated area, land form, wetlands, and
   manmade features.

       In addition to allowing map production, the NTDB is highly useful to
   users of geographic information systems (GIS).

       This data set has been extracted from a complete NTDB Demonstration
   Data File at the scale of 1:50 000. It covers the area of SWAN RIVER ,MB
   Manitoba), Canada and bears the number 063C03 based on the National
   Topographic System (NTS).
      Purpose:
      To provide Canadians with topographic and toponymic information to
   support the sustainable development of natural resources, environmental
   protection and the management of Canadian territory.
   Spatial_Domain:
```

```
      Bounding_Coordinates:
          West_Bounding_Coordinate: -101.50
          East_Bounding_Coordinate: -101.00
          North_Bounding_Coordinate: 52.25
          South_Bounding_Coordinate: 52.00
    Keywords:
      Theme:
        Theme_Keyword: AD(Designated Area)
        Theme_Keyword: CH(Roads)
        Theme_Keyword: CO(Manmade Features)
        Theme_Keyword: FO(Relief adn Landforms)
        Theme_Keyword: GE(General)
        Theme_Keyword: HD(Hydrography)
        Theme_Keyword: HP(Hypsography)
        Theme_Keyword: LA(Administrative Boundaries)
        Theme_Keyword: RE(Rail Network)
        Theme_Keyword: RR(Road Network)
        Theme_Keyword: SS(Water Saturated Soils)
        Theme_Keyword: TO(Toponymy)
        Theme_Keyword: VE(Vegetation)
      Place:
        Place_Keyword:SWAN RIVER
        Place_Keyword:MB (Manitoba)
        Place_Keyword: Canada
        Place_Keyword: North America
    Access_Constraints: None
    Use_Constraints: None
  Data_Quality_Information:
    Positional_Accuracy:
      Horizontal_Positional_Accuracy:
        Horizontal_Positional_Accuracy_Report: 50
      Vertical_Positional_Accuracy:
        Vertical_Positional_Accuracy_Report: 5
    Spatial_Data_Organization_Information:
      Direct_Spatial_Reference_Method: Vector
    Spatial_Reference_Information:
      Horizontal_Coordinate_System_Definition:
        Planar:
          Map_Projection:
              Map_Projection_Name: UTM projection
        Geodetic_Model:
```

```
        Horizontal_Datum_Name: North American Datum of 1927
    Vertical_Coordinate_System_Definition:
      Altitude_System_Definition:
        Altitude_Datum_Name: Mean Sea Level
        Altitude_Distance_Units: 25.0
        Altitude_Encoding_Method: feet
Entity_and_Attribute_Information:
  Detailed_Description:
    Entity_Type:
      Entity_Type_Label: Polygon
Metadata_Reference_Information:
  Metadata_Date: 1999-08-18
  Metadata_Contact:
    Contact_Information:
      Contact_Person_Primary:
        Contact_Person: Ying Teng
        Contact_Organization: Faculty of Computer Science, University of
      New Brunswick
      Contact_Position: MCS Student
      Contact_Electronic_Mail_Address: n74v9@unb.ca
  Metadata_Standard_Name: FGDC Content Standards for Digital Geospatial
  Metadata
  Metadata_Standard_Version: CSDGM Version 2 - FGDC-STD-001-1998
```

APPENDIX III

The 063c03.xml obtained from mp using 063c03.txt as input file.


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE metadata SYSTEM"http://www.fgdc.gov/metadata/fgdc-std-001-1998.dtd">
<metadata>
<idinfo>
<citation>
<citeinfo>
<origin>Canadian Government Agencies: NRCan</origin>
<pubdate>1998-09-23</pubdate>
<title>SWAN RIVER</title>
<edition>2.02</edition>
<onlink>http://www.ccg.rncan.gc.ca/ext/html/english/products/ntdb/demontd5.htm
l</onlink>
</citeinfo>
</citation>
<descript>
<abstract>
The National Topographic Data Base (NTDB) is a digital data base developed by
GeomaticsCanada. It covers the entire Canadian landmass and contains the
features normally found on topographic maps at the scales of 1:50 000 and
1:250 000: hydrography, hypsography(contours), vegetation, the road network,
roads, the rail network, the electric powernetwork, designated area, land
form, wetlands, and manmade features.

In addition to allowing map production, the NTDB is highly useful to users of
geographic information systems (GIS).

This data set has been extracted from a complete NTDB Demonstration Data File
at the scaleof 1:50 000. It covers the area of SWAN RIVER ,MB (Manitoba),
Canada and bears the number 063C03 based on the National Topographic System
(NTS).
</abstract>
<purpose>
To provide Canadians with topographic and toponymic information to support the
sustainabledevelopment of natural resources, environmental protection and the
management of Canadian territory.
</purpose>
</descript>
```

<spdom>
<bounding>
<westbc>-101.50</westbc>
<eastbc>-101.00</eastbc>
<northbc>52.25</northbc>
<southbc>52.00</southbc>
</bounding>
</spdom>
<keywords>
<theme>
<themekey>AD(Designated Area)</themekey>
<themekey>CH(Roads)</themekey>
<themekey>CO(Manmade Features)</themekey>
<themekey>FO(Relief adn Landforms)</themekey>
<themekey>GE(General)</themekey>
<themekey>HD(Hydrography)</themekey>
<themekey>HP(Hypsography)</themekey>
<themekey>LA(Administrative Boundaries)</themekey>
<themekey>RE(Rail Network)</themekey>
<themekey>RR(Road Network)</themekey>
<themekey>SS(Water Saturated Soils)</themekey>
<themekey>TO(Toponymy)</themekey>
<themekey>VE(Vegetation)</themekey>
</theme>
<place>
<placekey>SWAN RIVER</placekey>
<placekey>MB (Manitoba)</placekey>
<placekey>Canada</placekey>
<placekey>North America</placekey>
</place>
</keywords>
<accconst>None</accconst>
<useconst>None</useconst>
</idinfo>
<dataqual>
<posacc>
<horizpa>
<horizpar>50</horizpar>
</horizpa>
<vertacc>
<vertaccr>5</vertaccr>

```xml
</vertacc>
</posacc>
<spdoinfo>
<direct>Vector</direct>
</spdoinfo>
<spref>
<horizsys>
<planar>
<mapproj>
<mapprojn>UTM projection</mapprojn>
</mapproj>
</planar>
<geodetic>
<horizdn>North American Datum of 1927</horizdn>
</geodetic>
</horizsys>
<vertdef>
<altsys>
<altdatum>Mean Sea Level</altdatum>
<altunits>25.0</altunits>
<altenc>feet</altenc>
</altsys>
</vertdef>
</spref>
</dataqual>
<eainfo>
<detailed>
<enttyp>
<enttypl>Polygon</enttypl>
</enttyp>
</detailed>
</eainfo>
<metainfo>
<metd>1999-08-18</metd>
<metc>
<cntinfo>
<cntperp>
<cntper>Ying Teng</cntper>
<cntorg>Faculty of Computer Science, University of New Brunswick</cntorg>
</cntperp>
<cntpos>MCS Student</cntpos>
```

<cntemail>n74v9@unb.ca</cntemail>
</cntinfo>
</metc>
<metstdn>FGDC Content Standards for Digital Geospatial Metadata</metstdn>
<metstdv>FGDC-STD-001-1998</metstdv>
</metainfo>
</metadata>

APPENDIX IV

The following example is metadata for a vector dataset - essential profile [9].

```
MD_Metadata
language:  en
characterSet:  ISO 10646-2
hierarchyLevel:  dataset
+identificationInfo
         MD_Identification
         language:  en
         characterSet::  ISO 10646-2
      abstract:  The Digital Chart of the World is a comprehensive 1:1,000,000
      scale vector basemap of the world. It consists of cartographic,
      attribute, and textual data stored on compact disc read only memory (CD-
      ROM). The primary source for the database is the Defense Mapping
      Agency's (DMA's) Operational Navigation Chart (ONC) series. This is the
      largest scale unclassified map series in existence that provides
      consistent, continuous global coverage of essential basemap features.
      The database contains more than 1,500 megabytes of vector data and is
      organized into 17 thematic layers.  The data includes major road and
      rail networks, major hydrologic drainage systems, major utility networks
      (cross-country pipelines and communication lines), all major airports,
      elevation contours  (1000 foot (ft)) with an index of geographic names
      to aid in locating areas of interest. The database can be accessed
      directly from the four optical CD-ROMs that store the database or can be
      transferred to a magnetic media.
      geographicBox:
            EX_GeographicBoundingBox
                  extentType:  inclusion
                  westBoundLongitude:  -180.0
                  eastBoundLongitude:  180.0
                  northBoundLatitude:  90.0
                  southBoundLatitude:  -90.0
      spatialResolution:  1,000,000
      category:  MD Category (haven't finalized this list yet)
      datasetCitation:
            CI_Citation
            title:  Digital Chart of the World
            date:  199208
            edition:  1
      datasetPointOfContact:
            CI_Responsibility
            mandantoryParty:
                  CI_MandantoryParty
                  organisationName:  National Imagery and Mapping Agency
            responsibility:  010 (pointOfContact)
            contactInfo:
                  CI_Contact
                  phone:
                        CI_Telephone
                        voice:  1-301-XXX-XXXX
                        facimile:  1-301-XXX-XXXX
            address:
                  CI_Address
                  deliveryPoint:  4600 Sangamore Road
                  city:  Bethesda
```

```
                    administrativeArea: Maryland
                    postalCode:  20816-5003
                    country:  US
              onlineResource:
                    CI_OnlineResource
                    linkage:  http://www.nima.mil
+ referenceSystemInfo
        RS_ReferenceSystem
        name:
              RS_Identifier
              identifier:  WGS 84
+ spatialRepresentationInfo
        MD_SpatialRepresentation
        representationType:  002 (vector)
+ lineageMetadata
        LI_Lineage
        statement::  The DCW was compiled from the following sources: the entire
        series of Operational Navigational Charts (primary source), Joint
        Navigation Charts (geographic features - Antarctica), AVHRR imagery
        (Vegetation), and NIMA Digital Flight Information File (airport
        information)
+digitalTransferOptions
        MD_DigitalTransferOptions
        onLine:
              CI_OnlineResource
              linkage:  http://www.nima.mil
+datasetFormat
        MD_Format
        name: VPF
        version: 1.0
```