

Community Metadata ISO 19115 Adaptor

Qiping Yan^a, Michael J. McMahon Jr.^a, Sergiu Dascalu^a, Frederick Harris^a

^a Computer Science and Engineering Dept.

University of Nevada-Reno, Reno, Nevada 89557

qiping@aol.com, mcmahon@cse.unr.edu, dascalus@cse.unr.edu, fredh@cse.unr.edu

Abstract

In today's world, geographic data plays an increasing role in many areas, including academic research, government decision making, and in people's everyday lives. As the quantity of geographic data gets larger, making full use of the data in a distributed, heterogeneous network environment, like the Internet, becomes a major issue. To better utilize and share those valuable resources, metadata standards have been developed. Metadata makes it easier to discover, explore, and share geographic data, particularly for cataloguing geographic data in clearinghouses. But one of the new problems that came about with metadata is interoperability since multiple metadata standards exist. Important geographic metadata standards include: ISO 19115, Dublin Core, CSDGM (US) and prENV 12657 (Europe). Semantically, metadata standards are distinct but, rather, they overlap and relate to each other in diverse and complex ways. In this research, we propose a mechanism to transform different geographic metadata standards to solve the interoperability issue using the new popular web service and XML/XSLT technologies.

Key Words: Geographic metadata, metadata interoperability, ISO 19115 CSDGM, Dublin Core.

1 Introduction

1.1 Metadata

Metadata is commonly defined as "data about data", "structured data about data", or "data which describes attributes of a resource", or even "information about data". Metadata provides information about one or more aspects of the data, such as purpose of the data, means of creation of the data, owner or author of the data, point of contact of the data, etc.

Metadata is not a new concept. Business cards and library cards are examples of metadata in our ev-

eryday lives. Metadata is often used for photographs (IPTC Schema, XMP, Exif, PLUS, etc.), videos (transcript, text description of the scenes), and webpages (keywords, description, software used to create the page, author of the page).

Furthermore, metadata is commonly used for computer or software, rather than humans. Metadata can be stored internally, in the same file as the data, or externally, in a separate file.

1.2 Metadata Standards

As described in [4], "metadata standards are requirements which are intended to establish a common understanding of the meaning or semantics of the data, to ensure correct and proper use and interpretation of the data by its owners and users." A metadata standard is usually established by national and international standard communities like ANSI (American National Standards Institute) and ISO (International Organization for Standard).

1.3 Geographical Metadata Standards

In the geographic domain, the most common metadata standards are FGDC's Content Standard for Digital Geospatial Metadata (CSDGM) and the recently ratified ISO 19115 [12]. The CSDGM standard contains over 300 data and compound elements while the ISO 19115 has over 400 elements (divided into 14 metadata packages) in 86 classes that have 282 attributes and 56 relations. The ISO 19115 was developed by the geospatial community to address specific issues relating to both the description and the curation of spatial data [5]. The ISO 19115's abstract models are written using the UML (Unified Modeling Language). The accompanying XML schema, ISO/CD TS 19139, enables interoperable XML expression of ISO 19115 compliant metadata.

1.4 XML/XSLT

The XML (eXtensible Markup Language) is a markup language created to structure, store, and transport data. It defines a set of rules for encoding documents in a format that is readable by both machines and human readers. Its main purpose is to separate data from its presentation. The design goals of XML emphasize simplicity, generalization and usability over the Internet. It is a textual data format with strong unicode support for the languages of the world. XML is a very flexible markup language because users can create their own tags to structure and store their documents.

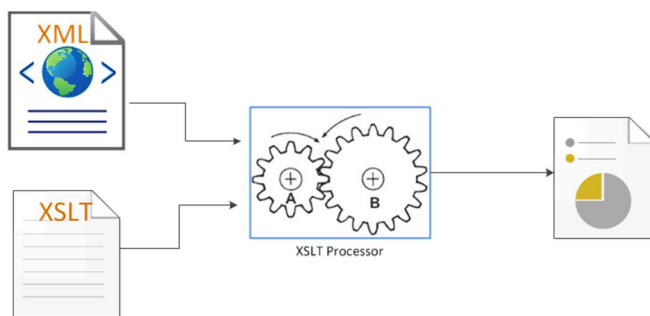


Figure 1: An XSLT processor takes one or more XML documents and one or more XSLT files as input and produces a new XML document as output.

On the other hand, XSLT (eXtensible Stylesheet Language Transformation) is a language for transforming one XML document into other XML documents, or other objects. As shown in Figure 1, the XSLT template processor takes one or more XML source files and one or more XSLT stylesheet modules, and processes them with the XSLT template processing engine to produce an output document.

2 Background

In the information age, data, especially geographic data, is exploding. Many companies produce and collect geodata in various formats from different sources with different equipments. The result of this heterogeneous geodata creates a major problem for data sharing. Geographic metadata describes the existing geodata. By reading the geo-metadata, users can get more information about the original dataset, like name, quality, ratio, data structure, etc. [11]. Because most geodata is large in size, written in various (standard) formats and stored in different file formats, it is hard to directly access the original dataset. Thus, it is necessary to utilize the geometadata to not only describe and catalogue

data, but also to discover, convert, manage, and use data in a network [9].

The more widely used geographic metadata standards include: CSDGM by FGDC and ISO 19115 by ISO/TC211 [4]. There are many differences between metadata standards. For example, the original version of DC (Dublin Core) defined only 15 elements known as the original set of 15 classic metadata set while the ISO 19115 has more than 300 elements, organized in 86 classes with 282 attributes and 56 relations.

Metadata standards are usually presented using a structural file. This concept is useful for standards with many elements defined, but it is hard to analyze and process the metadata. Moreover, there is not a single metadata definition language. Therefore, different standards are presented using different notations. For example, the ISO 19115 uses the UML while the CSDGM uses a formal file notation [13]. Finding a common way to present different metadata standards is a must in order to make it possible for computers to automatically recognize, analyze metadata, and share geo-data in different metadata standards. To this end, XML is a popular markup language which can define, present, verify, and index metadata [10].

3 Proposed Approach

3.1 Overall Solution

This effort is part of the metadata exchange project, which is, itself, part of Track II of the NSF EPSoR funded project "Collaborative Research: Cyberinfrastructure Developments for the Western Consortium of Idaho, Nevada, and New Mexico". On September 2008, NSHE (Nevada System of Higher Education) was awarded \$15 million by NSF EPSoR over five years to develop science, education, and outreach infrastructure at UNR, UNLV, DRI (Desert Research Institute), NSL (Nevada Seismological Laboratory), and NSHE community colleges for the study of climate change and its effects on Nevada. The Nevada Climate Change Project now comprises a web portal, the SENSOR data system (software, hardware, and database), and a high-speed TCP/IP network infrastructure.

The NCCP (Nevada Climate Change Portal) website (<http://sensor.nevada.edu>) provides information to project members, researchers, and the public [2]. Via search interfaces and web services, users can search and download collected from SENSOR data. The web portal also provides real-time videos and photos from monitoring sites. The SENSOR data collection system comprises numerous Campbell Scientific data loggers (CR1000 and CR3000), each with dozens of physical sensors that collect thousands of measurements per

minute. Through secure virtual private network (VPN) of the Nevada Seismological Laboratory [2], the data loggers transport collected data to the data center via the lossless TCP/IP protocol [3]. Then the data is stored on file servers and imported into a SQL server. By design, the SENSOR system uses a standards-neutral database schema, meaning that the database structure is not modeled after a specific metadata standard like FGDC or ISO 19115 [3] – the system simply collected more information than was needed by any one standard. The main purpose of this work is to provide a system to transform raw data to a chosen metadata standard.

As mentioned before, the NSF EPSoR-funded project involves three western states: Nevada, Idaho, and New Mexico. Idaho and New Mexico have their own data centers or data repositories implemented particular metadata standards. This creates a problem: how can we effectively share data between data centers? One solution is to build a central clearinghouse to which each data center submits metadata for cataloguing [13]. Users can then search the clearinghouse for geographic data from all participating data centers. Each participant web portal can query the clearinghouse for data too.

This paper proposes a practical mechanism to transform between different metadata standards. With this service, the clearinghouse will be able to handle data submissions accompanied by differing metadata standards, allowing each data center to utilize its own metadata standard, such as ISO 19115.

In this paper, we propose to create a community metadata ISO 19115 adaptor that will transform geographic metadata in different standards to the ISO 19115 metadata standard format and vice versa. The project will be implemented using standard SOAP web service technology, XML, XSLT, and C#.

3.2 Logical Model

The diagram in Figure 2 shows the logical model of the proposed approach.

In the diagram (Figure 2), a user initiates a web service request with proper parameters. The web service will validate the request, record user information (optional), process the request, and return proper output metadata.

The request parameters include: geospatial information, sampling interval, time period, and return data format. The return data is an XML document that complies with the requested metadata standard. See the web user interface in Figure 4 for more details.

3.3 System Design/Software Workflow

Figure 3 shows the generic software workflow chart. The request can be a web form request or script-generated web service request. The response is always an XML document in a standard metadata format.

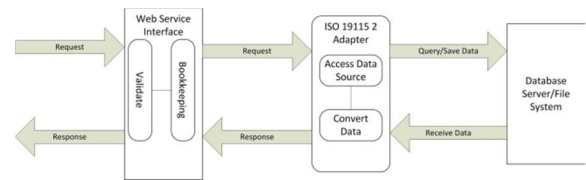


Figure 3: Software Workflow with web service interface module and ISO 19115 adaptor module

3.4 User Interface

Users will use a form to easily construct the request. The web form will provide the necessary data for the web service and the ISO 19115 adaptor to properly query data and form metadata (translate). For security purposes, a user must provide a valid user ID or developer reference number with each request. Although the web form user interface is easy to use, it is not the only way to access the service. A service request can be constructed by scripts supplying the same required request parameters. Figure 4 shows the user interface.

3.5 Transformation

In our proposed approach, transformation between two different metadata standards is carried out using XSLT. That means two XSLT template files are required for each pair of source and target formats: $T_{a \rightarrow b}$ and $T_{b \rightarrow a}$.

This approach is straightforward when the number of supported formats (n) is small because the number of XSLTs needed to transform between from a source format to a target format without using a transformation sequence ($T(n)$) is n factorial: $T(n) = n!$. For example, 6 XSLTs are required for 3 supported formats. This mechanism works well with

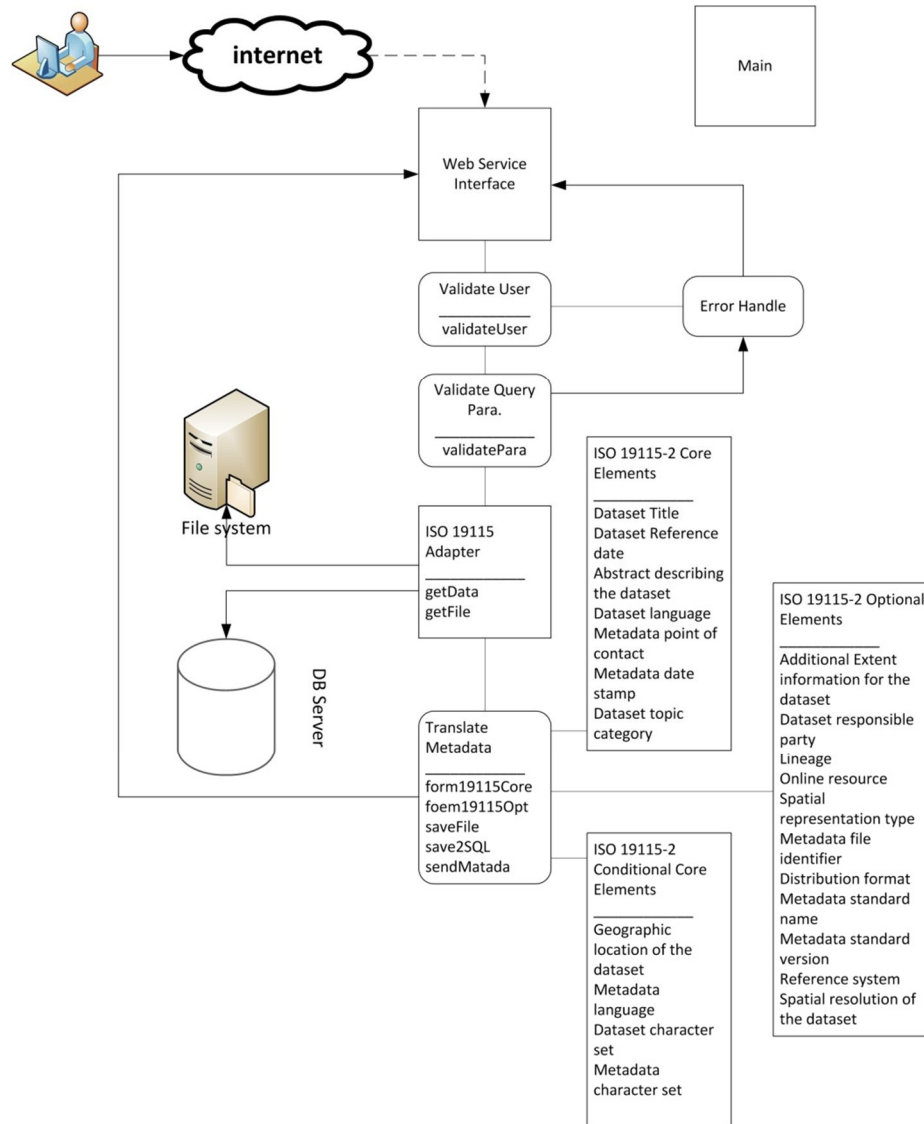


Figure 2: System Design showing basic functionalities of the web service based ISO 19115 adaptor

a small number of supported formats, but is clearly inefficient for a large number of formats. For example, $T(10) = 3,628,800$, $T(15) = 1,307,674,368,000$. This is clearly impractical. Some researchers suggested using the graph theory. By letting each vertex in the graph represent a one-way transformation, as long as a connected graph is created, any source formats can be theoretically transformed into any target formats by using a transformation sequence. Furthermore, weights can be assigned to each edge based on the transformation complexity, computation time, etc. [1]. Then the graph theory (Dijkstra's algorithm) can be implemented to find the shortest path to improve efficiency.

The graph makes the concept complex. Given a set of transformations, determining the most efficient transformation sequence between source format A and target format B is essentially equivalent to solving the shortest path problem for a graph in which the data formats are represented by vertices, and the transformations are represented by edges [6, 8]. We propose a new mechanism: centralized transformation. We pick a popular format as the center point (TC) and create the two-way XSLTs for each pair of formats in which one of the formats is the center point format TC. This way, only $2 * (n - 1)$ XSLTs are needed. A transformation

Figure 4: Web User Interface to construct and send a web service request to search for data or to request for a metadata standard transformation

between any two formats $T_{n \rightarrow m}$ can be carried out with a sequence of only two transformations: $T_{n \rightarrow m} = \text{XSLT}_{n \rightarrow TC} + \text{XSLT}_{TC \rightarrow m}$. See Figure 5.

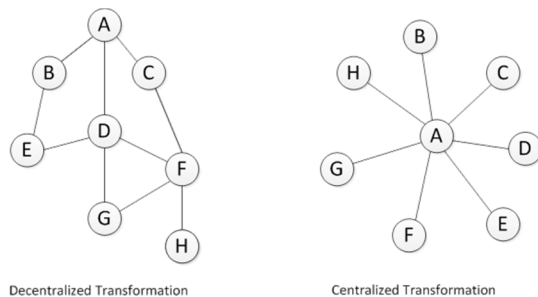


Figure 5: Concepts comparison between graph-based transformation (on the left side) and centralized transformation (on the right side)

4 Related Works

4.1 GeoNetwork Opensource

GeoNetwork Opensource is a standards-based and decentralized spatial information management system designed to enable access to geo-referenced databases and cartographic products from a variety of data providers through descriptive metadata, enhancing the spatial information exchange and sharing between organizations and their audiences using the capacities and the power of the Internet [7]. The GeoNetwork Opensource is an open source Java application developed following the principles of Free and Open Source Software (FOSS) and based on the International and Open Standards for Services and Protocols. The project started out as a Spatial Data Catalogue System for the Food and Agriculture Organization of the United Nation (FAO), the United Nation World

Food Programme (WFP), and the United Nations Environmental Programme (UNEP).

4.2 IDACT Transformation Manager

IDACT Transformation Manager project was part of a group research on the Arctic Pollution Issues conducted by AMAP (The Arctic Monitoring and Assessment Programme). It leverages the knowledge of a group of users, ensuring that once a method for a data transformation has been defined either through automated processes or by any single member of the group, it can be automatically applied to similar future datasets by all the members of the group. This prototype was primarily focused on the dataset itself instead of the metadata [6, 9].

5 Conclusions and Future Work

Data is valuable and expensive to collect. Organizations and institutes invest heavily in both equipment and labor to collect data, sometime just for a specific study. Thus, a method to reuse this valuable data becomes an important research topic. Different organizations use different software systems during the data collection phase. Therefore these data are in different formats, which makes it hard to reuse, discover, and catalogue them. By introducing metadata standards to describe data, researchers can easily search and discover existing data for their studies. In the geographic domain, there are multiple metadata standards. Thanks to the XML/XSLT technology, it is relatively easy and straightforward to represent those metadata standards in the XML format and perform different transformations among different metadata standards using XSLT.

In this paper, we proposed a mechanism to transform between metadata standards based on web services and the XML/XSLT technology. This will enable more effective sharing of geographic data represented in different data/metadata standards.

Future work includes implementing the full web service system and customizing it so it can be used in the NCCP portal to generate ISO 19115 metadata on the fly.

References

- [1] Y. Wei, L. Di, B. Zhao, et al. (2007). "Transformation of HDF-EOS metadata from the ECS model to ISO 19115-based XML." *Computers & Geosciences*, vol. 33, no 2, 2007, pp 238-247.
- [2] McMahon, Michael J., et al. "Architecting Climate Change Data Infrastructure for Nevada." *Advanced Information Systems Engineering Workshops CAISE-2011, Lecture Notes in Business Information Processing*. LNBP-83. 2011.
- [3] McMahon, M.J. and Dascalu, S.M., Harris, F. C., Jr., and Strachan, S., "SENSOR: Applying Modern Software Data Management Practices to Climate Research", *Proceedings of the 2011 Workshop on Sensor Network Applications*, Honolulu, HI, 2011.
- [4] Metadata standards, November 2012, , Wikipedia, http://en.wikipedia.org/wiki/Metadata_standards
- [5] Metadata standards, 2012, DCC (Digital Curation Centre), <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>
- [6] Hay, B. Idact transformation manager. *Diss. MONTANA STATE UNIVERSITY Bozeman*, 2005.
- [7] Ticheler, Jeroen, and Jelle U. Hielkema. "GeoNetwork opensource Internationally Standardized Distributed Spatial Information Management." *OS-Geo Journal*, 2.1 (2007).
- [8] Nance, Kara L., and Brian Hay. "Automatic transformations between geoscience standards using XML." *Computers & Geosciences* 31.9 (2005): 1165-1174.
- [9] Nance, K., and Brian Hay. "IDACT: automating data discovery and compilation." *Proceedings of the 2004 NASA Earth Science Technology Conference*, 2004.
- [10] Greenberg, Jane. "Metadata extraction and harvesting." *Journal of Internet Cataloging* 6.4 (2004): 59-82.
- [11] Kawtrakul, Asanee, and Chaiyakorn Yingsaeree. "A unified framework for automatic metadata extraction from electronic document." *Proceedings of The International Advanced Digital Library Conference*. Nagoya, Japan. 2005.
- [12] Nogueras-Iso, J., et al. "Metadata standard interoperability: application in the geographic information domain." *Computers, Environment and Urban Systems* 28.6 (2004): 611-634.
- [13] Bigagli, Lorenzo, et al. "GI-Cat: A web service for dataset cataloging based on ISO 19115." *Procs of the 15th IEEE International Workshop on Database and Expert Systems Applications*, 2004.