Geo-Science GML Encoding Project

Final Report

TR2002-213-02

Submitted to Natural Resources Canada by Galdos Systems Inc

Date: May 6, 2002

Galdos Systems Inc.
Suite 200, 1155 West Pender Street
Vancouver, BC V6E 2P4
Canada

Contact: Tim Wilson
Phone: (604) 484-2750
Fax: (604) 484-2755
E-mail: twilson@galdosinc.com

GLOSSARY OF TERMS

CSIRO	Commonwealth Scientific and Industrial Research Organisation (Australia).
GML	Geography Markup Language
GSC	Geological Survey of Canada
GXF	Grid eXchange Format (Geosoft)
MO	Mineral Occurence
NGR	National Geochemical Reconnaissance
srsName	Spatial Reference System Name
SVG	Scalable Vector Format
uom	Units Of Measure
WWW	World Wide Web
XML	eXtensible Markup Language

REVISION HISTORY

Version	Date	Pages	Description	Release
1/0	2002-03-30	All	First Issue	
1/1	2002-05-06	All	Updated for revised schemas	

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report outlines the development process, schema design issues, and intended uses for the following Geographic Markup Language (GML) application schemas:

- Mineral Occurrence
- Geochemical Reconnaissance
- Aeromagnetic/Radiometric grid data (based on Grid Exchange Format).

The primary purpose of these national application schemas is to support the mapping and transfer of provincial data to a federated content model. The federated schemas are intended to support the collection, analysis, and transfer of consistent information among provinces, while allowing provincial autonomy in the management of independent database models and systems.

For convenience, each set of application schemas (Mineral Occurrence, Geochemical Reconnaissance, and GXF) are described in individual summary documents. These are:

- *Mineral Occurrence Application Schemas [Document R-03]*
- *Geochemical Reconnaissance Schemas [Document R-04]*
- *GXF Schema, GXF2GML and GML2GXF Conversions [Document R-05]*

Any related references and appendices (e.g. HTML documentation) are attached to each report.

The schemas can serve within the Canadian Geo-Science Knowledge Network (CGKN) as a means of promoting the exchange of Geo-Science data and services between Canadian geology and mining agencies over the Internet. It is expected that the schemas will be refined over time to meet the changing data model requirements of the Canadian Geo-Science community, and to keep pace with the development of new GML core schemas (e.g. upcoming OGC GMLv 3).

1 INTRODUCTION

1.1 Background

The motivation for the Geo-Science GML Encoding Project is described in the *Geo-Science Encoding Technical Specifications [Document R-1]* document. Historically, provincial Geo-Sciences agencies have maintained their own database systems and models. While there have been attempts to promote national standards for information sharing, these have not been widely embraced. Provinces are unlikely to make significant changes in their existing systems and hence convergence is more likely to occur through the development of a national schema to which each provincial schema is mapped in an appropriate fashion. Such an approach preserves provincial autonomy while allowing for national integration and querying.

1.2 Terms of Reference

Galdos Systems Inc. entered into an agreement with Natural Resources Canada ("Client") to develop GML application schemas for the Geo-Science sector. The development of the schemas was guided by the Client through ongoing discussions of design priorities and schema review. This report provides a summary of all final application schemas, a GXF XSLT demonstration application, and recommendations for further development.

1.3 List of Deliverables

Galdos has met the terms of this agreement by conducting or providing the following:

- Producing draft versions of Mineral Occurrence (MO), Geochemistry, and GXF schemas for review by NRCAN staff.
- Soliciting internal and external review comments and/or clarifications from Geo-Science and GML domain experts. The list of reviewers includes:
    - John Glynn (NRCAN)
    - Stephen Adcock (NRCAN)

  - Eric Grunsky (CGKN)
  - Simon Cox (CSIRO)
  - Robert Deklerk (Yukon Geology Program)
  - Ron Lake (Galdos Systems Inc)
  - Richard Martell (Galdos Systems Inc)
  - Milan Trninic (Galdos Systems Inc)

- Hosting an information seminar at the office of Galdos Systems Inc. on January 21st, 2002 for NRCAN/CGKN staff.
- Delivery of final application schemas and GXF XSLT application with sample data.
- Hosting a simple browser viewer for national MO data.
- Delivery of final report.

1.4 Reference Documents

The following documents are referenced in this document:

R-1	TR2001-213-01	Geo-Science GML Encoding Project: Technical Specifications
R-2	TR2001-213-03	Geo-Science GML Encoding Project: Mineral Occurrence Application Schemas
R-3	TR2001-213-04	Geo-Science GML Encoding Project: Geochemical Reconnaissance Schemas
R-4	TR2001-213-05	Geo-Science GML Encoding Project: GXF Schema - GXF2GML Transformation
R-5	OGC 02-009	Geography Markup Language Implementation Specification, version 2.1.1. http://www.opengis.org/techno/specs/02-009/GML2-11.html , 25 April 2002. OGC

2 APPLICATION SCHEMAS

2.1 Viewing and Validating GML Data/Schemas

The Geo-Science application schemas have been developed using XMLSpy v4.3. These schemas can be viewed using any Unicode-capable text editors and/or browser software (e.g. IE5.5). End-users may prefer using specialized XML editing software (e.g. XMLSpy) for the larger schemas and sample datasets due to the ability to quickly navigate and manipulate XML data structures.

The application schemas and sample data have been supplied in the attached CD-ROM. Should these schemas be transferred to a local machine, the directory structure must be maintained in order for the sample data and schemas to validate correctly.

All schemas have been validated using the XMLSpy v4.3 and Xerces 2.0.1 validating parsers.

2.2 Mineral Occurrence

A federated Mineral Occurrence (MO) application schema has been developed to support the direct mapping of provincial data to a national content model. The federated schema contains:

- Content that is common to all examined provincial MO databases; or,
- Content that is provided by only one or some provincial MO databases, but is considered highly valuable. In other words, it helps to formulate a "best of" content model where the content is treated as optional elements, such that the federated schema is not tightly coupled to any specific provincial MO databases.
- Content that shares semantic similarities between provincial MO databases. In this case, semantic qualifiers are carried as attributes to maintain unique provincial meaning and/or interpretation.

Generalization mechanisms (extension/restriction, element substitution, weak typing, semantic refinement using attributes, and optional elements) are used to

maximize the commonality between provinces, while maintaining a coherent and structured content model. As a result, the federated schema is sufficient to capture the entire British Columbia (BC) MINFILE, Newfoundland and Labrador (NF) MODS, and Yukon MINFILE data models, and the majority of the Ontario MDI2 database.

The federated schema supports the primary purpose for the MO application schema, i.e. the transformation of provincial data to a national content model for the purposes of querying and displaying data. Since the federated schema is a generalized schema, it does not contain sufficient business rules or constraints to support data loading into provincial databases (updates, insert, etc.). However, it is possible to model provincial database constraints, using XML technologies such as Schematron, to apply business rules to provincial schemas (that extend the federated schema).

The MO schemas are based on draft OGC GMLv3.0 core schemas as they offer better support for the federated schema requirements. The GML 3.0 schemas used are the OGC versions that were 'frozen' on April 26, 2002. It is anticipated that slight changes to the GMLv3.0 core schemas will occur prior to their official release, anticipated at the end of 2002. Some updating to the MO application schemas will likely be required at that time.

For a detailed description of the MO application schemas and sample data, see *Mineral Occurrence Application Schemas [Document R-3].*

2.3 Geochemical Reconnaissance

The Geochemical Reconnaissance (GR) application schemas are based on the Geological Survey of Canada (GSC) "Standardized geochemical data model" (DB_Schema-1.11.pdf data schema). This model was first captured as a UML logical model in consultation with Stephen Adcock, the database designer, of NRCAN. The application schemas have been developed using this model.

The GR application schemas are based on GMLv2.1.1, since GML3.0 was not sufficiently advanced at the time of production. However, a supplementary demonstration version of GR schemas using GMLv3.0 has been produced and included with the deliverables.

For a detailed description of the GR application schemas and sample data, see *Geochemical Reconnaissance Schemas [Document R-4].*

2.4 Grid Exchange Format (GXF)

Federal aeromagnetic and radiometric data is currently stored in proprietary GeoSoft™ binary format. To facilitate the transfer of data to a non-proprietary, accessible, and standardized format, a GXF GML application schema has been developed. The application schema is based on the GeoSoft™ Grid eXchange Format (GXF), as it provides intermediary text-based data format that can be readily exported or imported from existing GeoSoft™ software.

Once again, the GXF application schema is based on GMLv2.1.1, since GML3.0 was not sufficiently advanced with respect to handling coverage data. However, a supplementary demonstration version of the GXF schema using GMLv3.0 coverage schema has been produced and included with the deliverables.

For a detailed description of the GXF schema and transformation tool, see *GXF Schema, GXF2GML and GML2GXF Conversions [Document R-5].*

3 XSLT DATA TRANSFORMATION

One of the inherent advantages of GML (and many other XML vocabularies) is that it separates data content from presentation. This means that data can be "transformed" to multiple target formats, such as graphical format (SVG), XHTML, and other alternative XML encodings.

To demonstrate XML data transformation, a bi-directional XSLT GXF2GML application has been produced to transform sample GXF data into sample GML data. This transformation tool provides a mechanism for NRCAN to publish its GeoSoft Grid data in GML format, using exported GXFdata as an intermediate data format.

For a detailed description of the GXF2GML transformation application, see *GXF Schema, GXF2GML and GML2GXF Conversions [Document R-5].*

Data visualization using Adobe's free Scalable Vector Graphics (SVG) plug in is demonstrated for the sample MO data. The graphical SVG data, and associated HTML properties, have been transformed from the sample MO data using XSLT scripts. The SVG/HTML data is viewable at
http://www.galdosinc.com/demos/geoscience/

4 XML LANGUAGE SUPPORT

GML is rapidly becoming an international "lingua franca" for modeling geo-spatial data and information. However, as an XML-based content model, it goes one step further by allowing GML application schemas and datasets to be written or translated into any language supported by a valid XML character-encoding scheme (e.g. UTF-8, ISO-8859-1).

As Galdos Systems Inc. did not receive any French-language MO, GR, or GXF data (e.g. Quebec provincial data), no specific analysis for French data and schemas could be conducted. However, Galdos Systems Inc. has built internationalization support into other application schemas and WFS systems for other NRCAN projects (e.g. Centre for Topographic Information). The following mechanisms have been employed to support multi-lingual datasets:

1. An XSLT transformation of normative English application schemas is applied to produce an XML localization file containing element names and

enumerated content that requires translation. The actual process of translating vocabulary must be conducted manually. Québécois French terms are used wherever possible, as provided by l'Office de la langue française (http://www.olf.gouv.qc.ca/index.html?/charte.html).

2. An XSLT stylesheet imports the localization file to substitute French vocabulary for English element and attribute names, and enumerated values. The XSLT stylesheet uses parameterized templates to substitute terms within element and attribute references and declarations, type definitions, and annotations. The resulting French CTIS application schemas can be used to validate a French CTIS GML dataset.

3. In a similar fashion, an XSLT stylesheet is used to transform between English and French datasets. Element names and attribute names, and enumerated content, are translated.

Note that the imported GML base schemas (feature.xsd, geometry.xsd) and XML schemas (xlinks.xsd, xml.xsd) are not localized; this is because only English versions of these schemas are considered normative.

The Geo-Science application schemas currently carry the xml:lang attribute on the parent feature or feature collections as a basis for future internationalization support.

5 RECOMMENDATIONS

In order to promote the use and integration of Canadian Geo-Science data using GML application schemas and technology, Galdos recommends that NRCAN/ CGKN:

- Update all existing application schemas to GMLv3.0 following its final publication (OGC estimates final approval by December 2002).

- Provide GML/XML training courses and/or material tailored to the Canadian Geo-Science community. This will promote the awareness, participation, and adoption of GML technologies by Geo-Science business leaders and domain experts. The successful long-term maintenance and development of future GML application schemas requires the participation of agency and industry geoscientists, system administrators, database analysts, etc.

- Develop a "proof-of-concept" Geo-Science Web Feature Server (WFS) application to automate the delivery of provincial data to a federated or national GML format. This would serve as an opportunity to demonstrate the many technical advantages of GML, and support the business case for provinces to pursue similar Geo-Science WFS/GML initiatives.

- Develop remote sensing application schemas using the GML3.0 coverage schema. This would provide a similar non-proprietary, open-source standard for the exchange of such data between Canadian agencies and industries.

Galdos Systems Inc. would be glad to provide assistance with any or all of the above-mentioned initiatives.

