GeoSciML: a standards-based encoding for transfer of geoscience information from IUGS/CGI

S.J.D.Cox[1], E. Boisvert[2], B. Brodaric[3], T.R., Duffy[4], B.R. Johnson[5], J.L. Laxton[4], S.M. Richard[6], B. Simons[7]

1 CSIRO Exploration & Mining, PO Box 1130, Bentley, WA 6102, Australia
2 Geological Survey of Canada, 490 rue de la Couronne, Quebec, G1K 9A9 Canada
3 Geological Survey of Canada, 615 Booth St, Ottawa, K1A 0E9 Canada
4 British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, UK
5 US Geological Survey, 954 National Centre, Reston, Va. 20192, USA
6 Geological Survey of Arizona, 416 W Congress St, #100, Tucson, Az, 85701, USA
7 Geoscience Victoria, GPO Box 4440, Melbourne, 3001, Australia
Corresponding author: Simon.Cox@csiro.au

ABSTRACT : GeoSciML is an information model and XML encoding for geoscience information. It is intended that GeoSciML will ultimately encompass a broad range of geoscience concepts. Version 1.0 (released in January 2006) is concerned with a set of objects and concepts underlying geologic maps, including geologic units, structures, materials and boreholes. GeoSciML is designed using UML, and the XML implementation is a GML Application Schema, making it compatible with standard service interfaces for geospatial data published by Open Geospatial Consortium (OGC). The conceptual design builds on work done in some important predecessor projects, in particular NADM and XMML. The modular implementation also adopts observations and sampling components developed in the OGC Sensor Web Enablement projects, geological timescale developed in collaboration with CHRONOS, and Assay Data Exchange developed with statutory and commercial organizations from the minerals sector. GeoSciML is a collaborative project operating under the auspices of IUGS Commission for Geoscience Information, and involving representatives of a significant number of organisations primarily from the statutory sector. The prototype language is being used as the basis for an interoperability testbed being reported on in a companion paper (Duffy et al.).

KEYWORDS : data, information GIS, XML, UML, standards.

1. Introduction

The development of web-based data access interfaces together with increased expectations on agencies for delivery of re-usable geoscience data has led to interest in standardising data-exchange formats. The IUGS Commission for the Management and Application of Geoscience Information (CGI) is coordinating an initiative to develop GeoSciML (GeoScience Mark-up Language) which is a geoscience data model and exchange format based on GML (Geography Mark-up Language, Cox et al., 2004).
A number of predecessor projects strongly influenced the development of GeoSciML. These include multi-jurisdictional activities by the North American Geologic Map Data Model (NADM Steering Committee, 2004) and Australian Government Geologists Information Policy Advisory Committee (GGIPAC, 2004), the eXploration and Mining Mark-up Language work (XMML) (Cox, 2004) and individual agency work at the British Geological Survey (Sen and Duffy, 2005) and GeoScience Victoria (Simons et al, 2005).

Developing GeoSciML as a GML Application Schema ensures compatibility with standard Open Geospatial Consortium service interfaces. To demonstrate the relevance of GeoSciML, a WFS/WMS interoperability testbed has been undertaken, reported on in the companion paper by Duffy et al.

2. Design approach

2.1. Scope

The scope of GeoSciML is primarily the kinds of geological concepts conventionally portrayed on geological maps. This involves a suite of feature types based on geological criteria (geological units, geological structures, fossils, geological relationships, earth materials, geological fabrics) or artefacts of geological investigations (specimens, sections, observations, measurements). Supporting objects, such as timescales and lexicons, are also included so that they can be used as classifiers for the primary objects.

2.2. Methodology

The model has been designed using the methodology specified in ISO 19109 and ISO DIS 19136. This specifies that the model shall be represented using a profile of UML as the conceptual schema language, leading to a literal serialization in XML with W3C XML Schema validation.

3. The model

3.1. Core model

The core concepts for mapped features are shown in Figure 1. This top-level conceptual model is adapted from NADM.

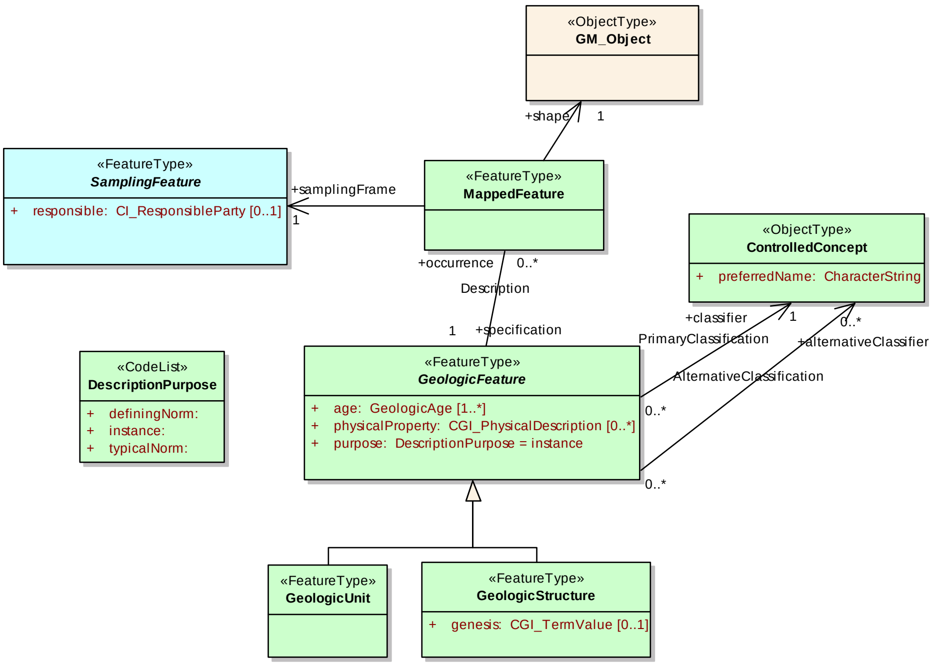

Fig. 1. Top level feature-types relating to geological map data

The descriptions of geologic features are modelled separately to generic feature-types representing mapped occurrences of these. This strategy reflects the longstanding convention of using a structured *legend* to supply detail about one or more features portrayed in spatial context on the *map*. Each geologic feature must be classified according to a concept from a controlled vocabulary (e.g. stratigraphy) and may also carry alternative classifications (e.g.

geotechnical rating). Each mapped feature occurs in the context of some sampling frame, such as a section, borehole, or mapping horizon.

3.2. Geologic Unit Model

The properties of geologic units are shown in Figure 2.

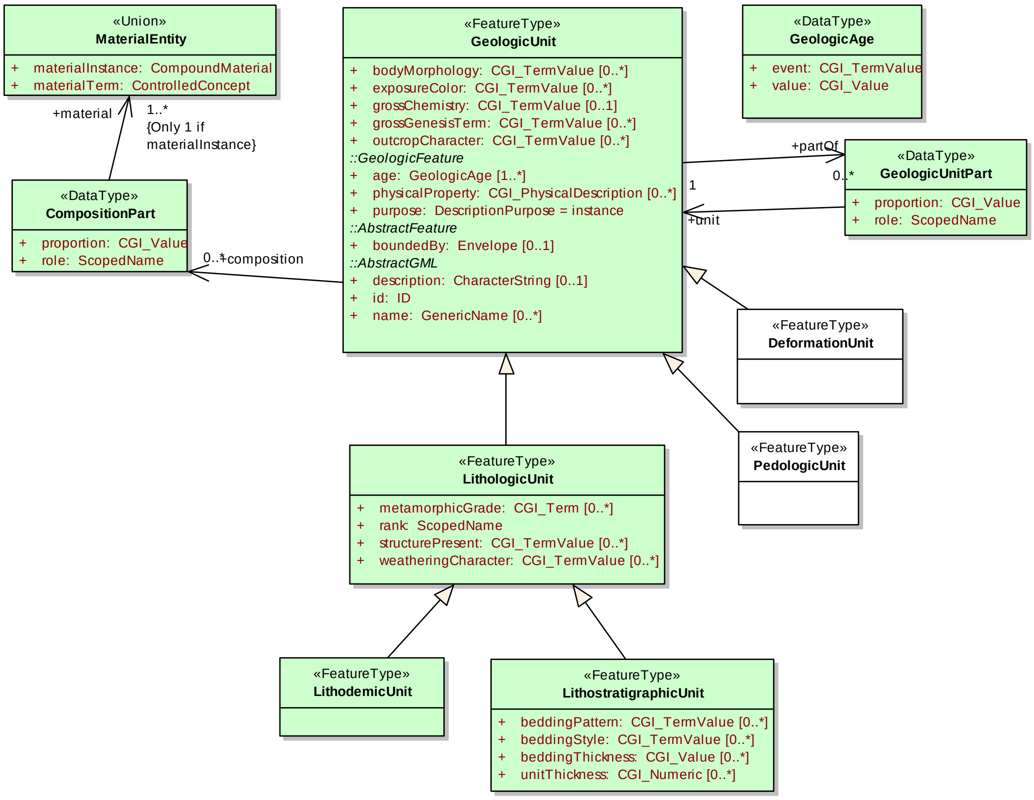

Fig. 2. Geologic units specializations and geologic material

The generic GeologicUnit class adds several attributes and associations to the properties inherited from GeologicFeature. The association with GeologicUnitParts allows units to be composed proportionally into super-units, and the association with CompositionParts allows a unit to be composed proportionally of material types. A number of specialized unit types are envisaged, though only lithostratigraphic and lithodemic units (types of LithologicUnits) have been modelled in GeoSciML version 1.

3.3. Property Values and Observations

Many of the feature properties are intended to have values taken from controlled vocabularies. However, GeoSciML does not enforce the use of a single vocabulary. Rather, a method for constructing a vocabulary is provided. When a term appears in a feature description instance, it is required to also identify the vocabulary from which it is taken.

It is also intended that GeoSciML documents be able to capture descriptions of geologic concepts using all the detail that might be included as part of field and lab observations. To enable this, a model for encoding “values” has been defined, which supports numeric (e.g. 2mm) or textual values (e.g. silt) optionally from controlled vocabularies, with uncertainty, and optionally combined into ranges (e.g. 2mm-to-boulder). Many of the feature properties are specified to have types defined by one of these value types.

In addition, GeoSciML will incorporate the Observations and Sampling Features defined in the Observations and Measurements model from OGC (Cox, 2006).

3.4. Other concepts

In addition to the feature, unit and material models shown here, GeoSciML models controlled vocabularies linked to their prototypes, geologic structures, relations between features, geologic timescale (Cox & Richard, 2005) and boreholes. For more information consult the website https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML

4. Encoding in XML

The UML model shown (in part) here serves as the design for XML document instances, which are constructed according to the rules defined in Annex E of ISO DIS 19136 (GML 3.2). This leads to an explicit document in which both class and property names appear as XML elements, thus preserving the model in the serialization. An XML Schema representation of the model is used for document validation.

5. Conclusions

We have described a feature model for geoscience information. While the scope is primarily the generation of geologic maps, the design is conceptual, not cartographic. Hence, GeoSciML data should be useful for other applications.

The project has been managed as a collaboration between representatives of agencies and research organizations in a variety of jurisdictions, so represents the best opportunity for international standardization in this topic.

REFERENCES

Cox, S.J.D., 2004, XMML – a standards conformant XML language for transfer of exploration data: Proceedings, *ASEG/PESA Geophysical Conference and Exhibition*, Sydney 2004

Cox, S.J.D., 2006, Observations and Measurements. OpenGIS® Discussion Paper, OGC document 05-087r3

Cox, S.J.D., Daisey, P.W., Lake, R., Portele, C. & Whiteside, A., 2004, *Geography Markup Language (GML) 3.1.0*, OpenGIS® Recommendation Paper, OGC document 03-105rl, xxi+ 580 pp.

Cox, S.J.D, & Richard, S.M. (2005) A formal model for the geologic timescale and GSSP, compatible with geospatial information transfer standards, *Geosphere, Geological Society of America* **1**/3, **119–137. doi:10.1130/GES00022.1 .** Available from <http://www.gsajournals.org/pdfserv/10.1130%2FGES00022.1>

Government Geologists Information Policy Advisory Committee, 2004, *National Geological Model Version 1.0* Explanatory Notes. http://www.geoscience.gov.au/geoportal/standards.html

North American Geologic Map Data Model Steering Committee, 2004: NADM Conceptual Model 1.0—A conceptual model for geologic map information: *U.S. Geological Survey Open-File Report* 2004-1334, 58 p., accessed online at URL http://pubs.usgs.gov/of/2004/1334. Also published as Geological Survey of Canada Open File 4737, 1 CD-ROM.

Sen, M. & Duffy, T.R., 2005 GeoSciML: Development of a generic GeoScience Markup Language. *Computers & Geosciences*, 31, pp1095–1103.

Simons, B., Ritchie,A., Bibby, L., Callaway, G., Welch, S., & Miller, B., 2005, Designing and Building an Object–Relational Geoscientific Database using the North American Conceptual Geology Map Data Model (NADM-C1) from an Australian Perspective. *Proceedings of IAMG'05: GIS and Spatial Analysis*, Vol.2, pp929–934.