The development of GeoSciML as a transfer standard for geoscience

J.L Laxton[1], T R Duffy[1], S J D Cox[2],
CGI Interoperability Working Group Members[3]

[1]British Geological Survey, Edinburgh, UK; [2]CSIRO Exploration and Mining, Bentley, WA, Australia;
[3]https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/WebHome

Background

It is becoming increasingly important to query and exchange geological information between geological data providers for legal, social, environmental and geoscientific reasons. Coincidentally, technological opportunities arising from the evolution of geospatial and geological information standards are making interoperability a viable endeavor.

In recognition of these trends the GeoSciML Interoperability Working Group was formed under the auspices of the IUGS Commission for the Management and Application of Geoscience Information (CGI), which aims to develop relevant and timely standards and test them within the context of individual agency activities. The ultimate objective of the working group is to enable seamless web integration of select information hosted at different agencies in varied formats.

More specific objectives are:

- to develop a conceptual model of geoscientific information drawing on existing data models
- to implement an agreed subset of this model in an agreed schema language
- implement an XML/GML encoding of the model subset
- develop a testbed to illustrate the potential of the data model for interchange
- identify areas that require standardised classifications in order to enable interchange

Materials developed by the Interoperability Working Group, including the GeoSciML UML schema, XML/GML schema, conformance criteria, instance documents and documentation, will be made available from the CGI website (http://www.cgi-iugs.org/tech_collaboration/data_model/downloads.html)

GeoSciML development work is mainly carried out on a Twiki site (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/WebHome) where detailed information about all aspects of the work can be found. We welcome participation in these discussions from all those interested in the development of geoscience interchange standards. A password is required for such participation and you can apply for one by mailing jll@bgs.ac.uk giving your accreditation and outlining your interest and involvement in geoscience interoperability.

Occasional face to face meetings are also held. These are by invitation to those who have been active in the relevant Twiki discussions and are designed to provide a concentrated period of development and decision making.

This poster will describe the development of GeoSciML as a data transfer standard for geoscience, including its underpinning data model, the development of web services to test the use of GeoSciML and demonstrate its potential, and initial work in developing standard geoscience concept definitions. There are four levels of interoperability, as shown in **Figure 1**, of which GeoSciML is designed to address schematic interoperability.

Conceptual Model

The scope of the current version of GeoSciML is the information generally shown on geological maps along with some observations, in particular those made using boreholes. The logical model has been built in UML and the model includes packages for mapped features, geologic units, earth material and geologic structures. The model inherits from the Open Geospatial Consortium (OGC) GML standard, for spatial information, and the OGC Observations and Measurements (O&M) standard for modelling boreholes and outcrops. At present the scope of the model is largely interpreted information, but the intention is to extend it to include more observational data.

Central to the model is the MappedFeature which can be considered an occurrence, such as a polygon on a geologic map, of a real-world geologic feature the full extent of which is unknown. A Geologic Feature is independent of geometry, so can have different MappedFeature instances representing mapped polygons at different scales or a modelled volume, for example. Each MappedFeature is associated with a SamplingFrame that indicates the spatial reference frame within which the MappedFeatures have been observed, such as a surface of mapping or a borehole

Figure 2 shows the relationship between a polygon on a geological map, the equivalent package of the conceptual model in UML, and the GeoSciML derived through a literal and explicit mapping of the UML using the encoding rules specified in ISO 19136.

The model incorporates a structure for controlled concepts which can be defined in terms of normative descriptions of geologic units or earth materials. These can be built into Vocabularies, such as stratigraphic lexicons. The GeosciML Vocabulary package (**Figure 3**) uses a specification for vocabularies derived from gml. A GeologicVocabulary is made up of a collection of ControlledConcepts and VocabRelations. A VocabRelation describes the relationship between ControlledConcepts (eg parent-child). LocalizedGenericName enables different names to be given to the same ControlledConcept eg for different languages.
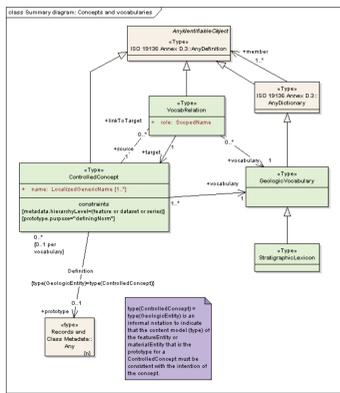

Figure 3: GeoSciML vocabulary package

Web services to test the model and its schema

Testbeds have been put up to test the implementability and interoperability of the model and its generated schema using OGC Web Mapping Services (WMS) and Web Feature Services (WFS) with queries. Eight geological surveys located across the planet have used corporate data and databases to test data exchange using GeoSciML (**Figures 4 and 5**) as the community agreed exchange schema. Defined Use Cases are tested using these OGC services. A testbed rendering and query client and registry lists registered GeoSciML conformant services (**Figure 6**). A GeoSciML WFS query and download Use Case for borehole data is displayed in **Figures 7a, 7b, 8a and 8b**. GeoSciML data download is also available from the OneGeology (www.onegeology.org) portal (**Figure 9**). The OneGeology initiative involves 80+ geological surveys around the world aiming to publish and share their data on the web using OGC WMS and WFS services based on GeoSciML.

Once testing is complete and GeoSciML V2.0 is released it has the potential to be used in Web Processing Service (WPS) environments and the GRID.
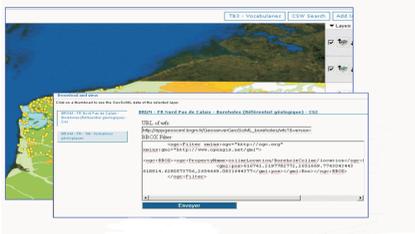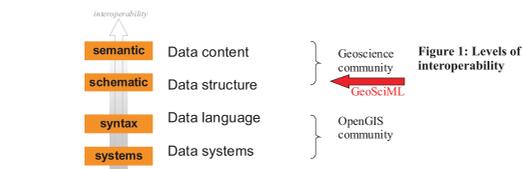
The Open Standards Development Approach

The development process for GeoSciML is an exemplar of the community-centric but standards-based model-driven-architecture approach defined in standards from ISO/TC 211 and mandated for INSPIRE. The model was developed in a collaborative process by a task group representing the stakeholder community. GeoSciML was formalized in UML, with standard components used through links to the ISO model hosted in a remote repository. The GeoSciML model is itself hosted in an online repository, allowing more specialized models (e.g. GroundWaterML, VirtualCoreLibraryML) to be derived directly from GeoSciML in turn. A GML-conformant XML implementation was generated automatically following rules defined by ISO and implemented in an Xquery processing framework. GeoSciML data transfer services can thus be deployed following the Open Geospatial Consortium's Web Feature Service interface spec.

Figure 10 shows the standards governing different components of the GeoSciML model. **Figure 11** shows the governance relationships of GeoSciML.
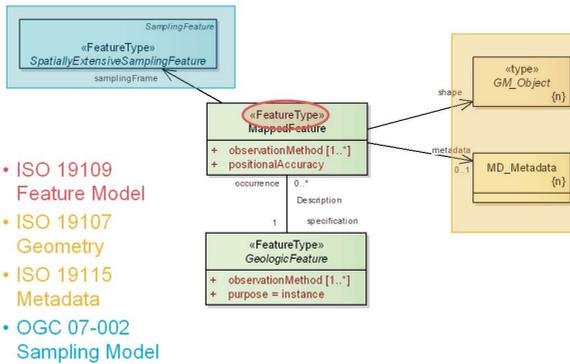

Figure 1: Levels of interoperability

Figure 11: The governance relationships of GeoSciML

- ISO 19109 Feature Model
- ISO 19107 Geometry
- ISO 19115 Metadata
- OGC 07-002 Sampling Model

Figure 10: The standards governing components of the GeoSciML model
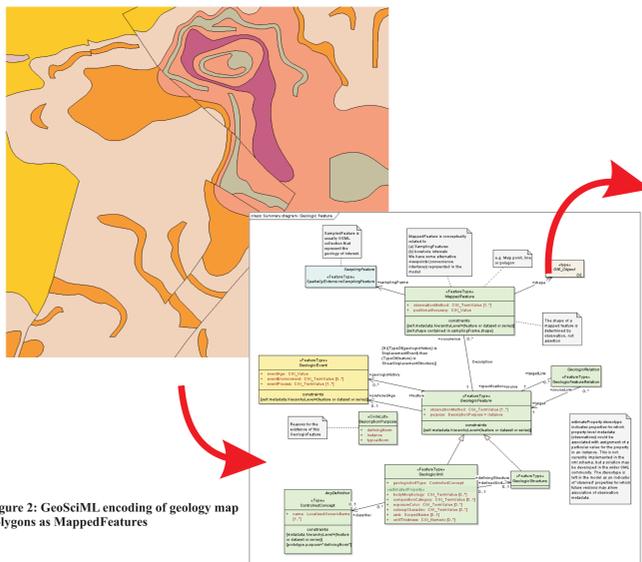

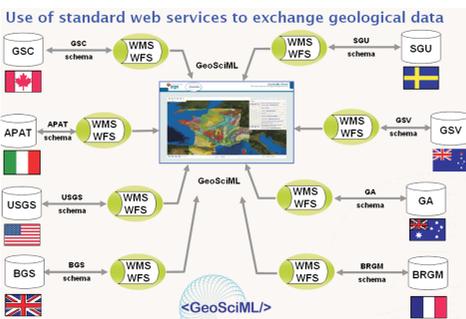**Figure 2: GeoSciML encoding of geology map polygons as MappedFeatures**

Figure 4: Interoperable exchange of Geoscience data from organisation specific backend database schemas via the community agreed GeoSciML exchange language
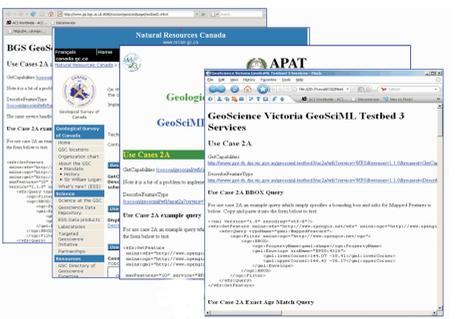

Figure 5: Testbed services from the geoscience community

Figure 7a: GeoSciML release candidate 2 bounding box query on a borehole WFS
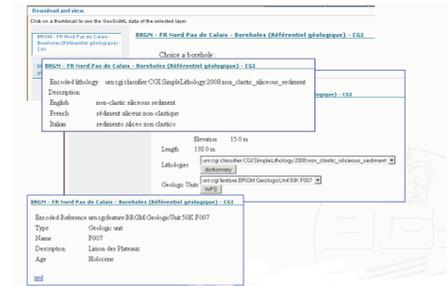

Figure 7b: Boreholes resulting from queries using international community agreed semantic multi-lingual vocabularies
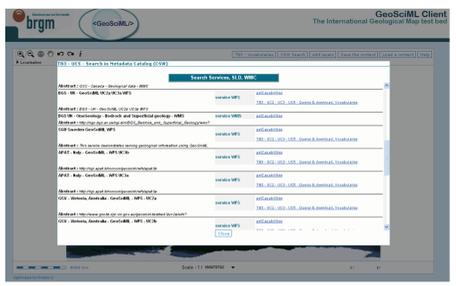

Figure 6: Available GeoSciML testbed services in a testbed registry

Figure 8a: A borehole WFS feature collection returned by a query
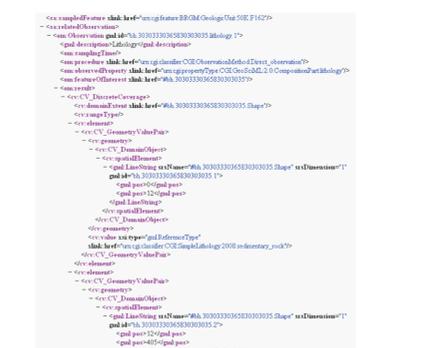
Figure 8b: The returned borehole log or interval observation from the borehole's downhole depth into the rock
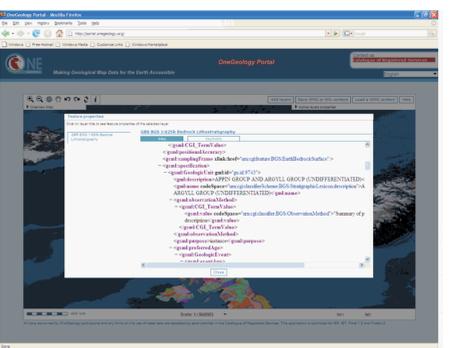

Figure 9: A GeoSciML WFS data exchange service behind a rendered WMS in the OneGeology portal