

GeoSciML – A GML Application for Geoscience Information Interchange

By Stephen M. Richard and CGI Interoperability Working Group

U.S. Geological Survey and Arizona Geological Survey
416 W. Congress #100
Tucson, AZ 85701
Telephone: (520) 770-3500
Fax: (520) 770-3505
e-mail: steve.richard@azgs.az.gov

CGI Interoperability Working Group is (in alphabetic order): Eric Boisvert, Boyan Brodaric, Simon Cox, Tim Duffy, Jonas Holmberg, Bruce Johnson, John Laxton, Tomas Lindberg, Stephen Richard, Alistair Ritchie, Francois Robida, Marcus Sen, Jean Jacques Serrano, Bruce Simons, and Lesley Wyborn.

INTRODUCTION

The GeoSciML application is a standards-based data format that provides a framework for application-neutral encoding of geoscience thematic data and related spatial data. GeoSciML is based on Geography Markup Language (GML, Cox et al., 2004) for representation of features and geometry, and the Open Geospatial Consortium (OGC) Observations and Measurements Best Practices (Cox, 2006) for observational data. Geoscience-specific aspects of the schema are based on a conceptual model for geoscience concepts and include geologic unit, geologic structure, and Earth material from the North America Data Model (NADMC1, North American Geologic-Map Data Model Steering Committee, 2004), and borehole information from the eXploration and Mining Markup Language (XMML, <https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome>). Development of controlled vocabulary resources for specifying content to realize semantic data interoperability is underway.

The intended scope for initial versions of GeoSciML includes information typically found on geologic maps as well as information typically recorded with boreholes. The possible uses for GeoSciML include transporting, storing, and archiving information. Amongst these, the most significant is transport—or information exchange—which enables information to be visualized, queried, and downloaded in spatial data infrastructures. This role for GeoSciML is particularly important, as geoscience information consumers are becoming more digitally sophisticated and are no longer satisfied with images and portrayals of data, but want digital data in standardized formats that can be used immediately in applications. Hours, days, or weeks spent merging data sets obtained separately from multiple agencies is time wasted. Use of a standardized

markup for serializing geoscience information supports a commitment by data providers to publish data to users in a standardized format. Thus, GeoSciML allows applications to utilize globally distributed geoscience data and information.

The GeoSciML (<https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML>) project was initiated in 2003 under the auspices of the [Commission for the Management and Application of Geoscience Information \(CGI\) working group on Data Model Collaboration](https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/CommissionfortheManagementandApplicationofGeoscienceInformation(CGI)workinggroup>DataModelCollaboration) (<https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/WebHome>). The CGI is a commission of the [International Union of Geological Sciences](https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/InternationalUnionofGeologicalSciences) and has the objective to enable the global exchange of geoscience information for legal, social, environmental, and geoscientific reasons. The project is part of what is now known as the [CGI Interoperability Working Group](https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/InteroperabilityWG) (<https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/InteroperabilityWG>), which has the specific objectives to:

- develop a conceptual model of geoscientific information that draws on existing data models,
- implement an agreed subset of this model in an agreed schema language,
- implement an XML/GML encoding of the model subset,
- develop a test bed to illustrate the potential of the data model for interchange, and
- identify areas that require standardized classifications to enable interchange.

GeoSciML draws from many geoscience data model efforts and from them establishes a common suite of feature types based on geological criteria (units, structures, fossils) or artifacts of geological investigations (speci-

mens, sections, measurements). Supporting objects are also considered (timescale, lexicons, etc), so that they can be used as classifiers for the primary objects. Predecessor projects that have had a strong influence on the development of GeoSciML include activities undertaken within national statutory bodies (e.g., the USGS/AASG National Geologic Map Database, British Geological Survey, and Japanese Geological Survey) in multi-jurisdictional contexts (the [North American Data Model](http://nadm-geo.org/), <http://nadm-geo.org/>, for geological maps), and activities oriented to an industry sector (eXploration and Mining Markup Language – XMML, <https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome>). Currently, several external projects are leveraging GeoSciML for more specific applications, including Water Resources monitoring and management, Soils, Geotechnical and Engineering, Assay Data, and Geochemistry.

This report summarizes the schema and instance documents as implemented in a test bed demonstrated at the IAMG meeting in Liege, Belgium in September, 2006. The working group met subsequent to the test bed demonstration and has identified a number of aspects of the model and schema in need of update, as well as model elements that need to be added. Anticipated changes are discussed here as well. Version 1.1 is the current version of the markup language, with schema available at <https://www.seegrid.csiro.au/subversion/xmml/GeoSciML/tags/1.1.0/schema/>. Planning is underway for evolution of the schema to version 2 to expand the scope and clarify some of the top level model issues. Working group activity is currently focused in several task-groups (pending formalization):

- Use-cases and requirements task group, responsible for setting technical goals.
- Design task group, responsible for the structural

and syntactic aspects of the “Information Model” of a GeoSciML-based service architecture.

- Service architecture task group, responsible for the “Computational Model” of GeoSciML-based service architecture.
- Concepts definition task group, responsible for the “Semantic Model,” which will be a standard set of concepts (ontology) for the content used to populate GeoSciML, and will facilitate semantic interoperability with GeoSciML.
- Implementation test bed task group, responsible for liaison with GeoSciML Design and Service Architecture task groups to ensure that requirements are satisfied and coordinate and deliver TestBed3 demonstrating the GeoSciML v.2 use-cases.
- Outreach and technical assistance task group, responsible for providing advice and assistance to direct collaborators, assisting them to deploy conformant GeoSciML services.

GEOLOGIC MAP DATA SCHEMATIC INTEROPERABILITY

The development of standardized markup languages is a critical step necessary to achieve interoperability, which is defined by ISO/IEC 2382-01 (SC36 Secretariat, 2003) to mean: “The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units.” Technical requirements to meet this goal include system-level shared protocols for network communication, resource discovery, and service invocation (Figure 1). Applications that use these protocols must communicate by way of a shared data language that defines how information will be encoded. Geography Markup Language



Figure 1. Multiple levels of interoperability (Brodaric and Gahegan, 2006).

(<http://www.opengeospatial.org/standards/gml>) is the data language adopted for GeoSciML development. GML provides a framework for encoding geometry, defining features and associating them with properties (including geometry), and constructing dictionaries in which controlled vocabularies can be defined.

GeoSciML is a GML application scheme, which is defined by a collection of XML schema that utilize and extend elements from GML to represent standard geologic observations and descriptions in a geospatial context. GeoSciML is not a database structure. GeoSciML defines a format for data interchange (Figure 2). Agencies can provide a GeoSciML interface onto their existing data base systems, with no restructuring of internal databases required.

The semantic level of interoperability (Figure 1) requires agreement on the meaning of words used to express property values contained in GeoSciML elements. Developing common meanings for GeoSciML contents that can be applied to various multi-lingual vocabularies is a planned future activity. At present, we anticipate that implementation of schematic interoperability will demonstrate the need for data content standards to enable semantic interoperability.

IMPLEMENTATION

GeoSciML was developed by representatives from an international group of geologic map data providers in a series of face-to-face meetings and online discussion (see Twiki at <https://www.seegrid.csiro.au/twiki/bin/view/CGI-Model/GeoSciML>). One design objective was to re-use, revise, and extend existing standards wherever possible. The design philosophy of this interchange format has focused on an accurate representation of geoscience information in a general way. This results in great representational flexibility at the price of complexity and verbose encoding. Fortunately, text-based XML compresses very efficiently, and the markup is designed for machine input and output, not human readability.

Model development has utilized UML notation with a UML profile to enable systematic mapping from UML to XML schema. The mapping from UML models to GML is described in <https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/UmlGml> and <https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/Uml2GMLAS>. A detailed procedure for generating a GML-compliant XML schema is summarized in <https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/HollowWorld> and

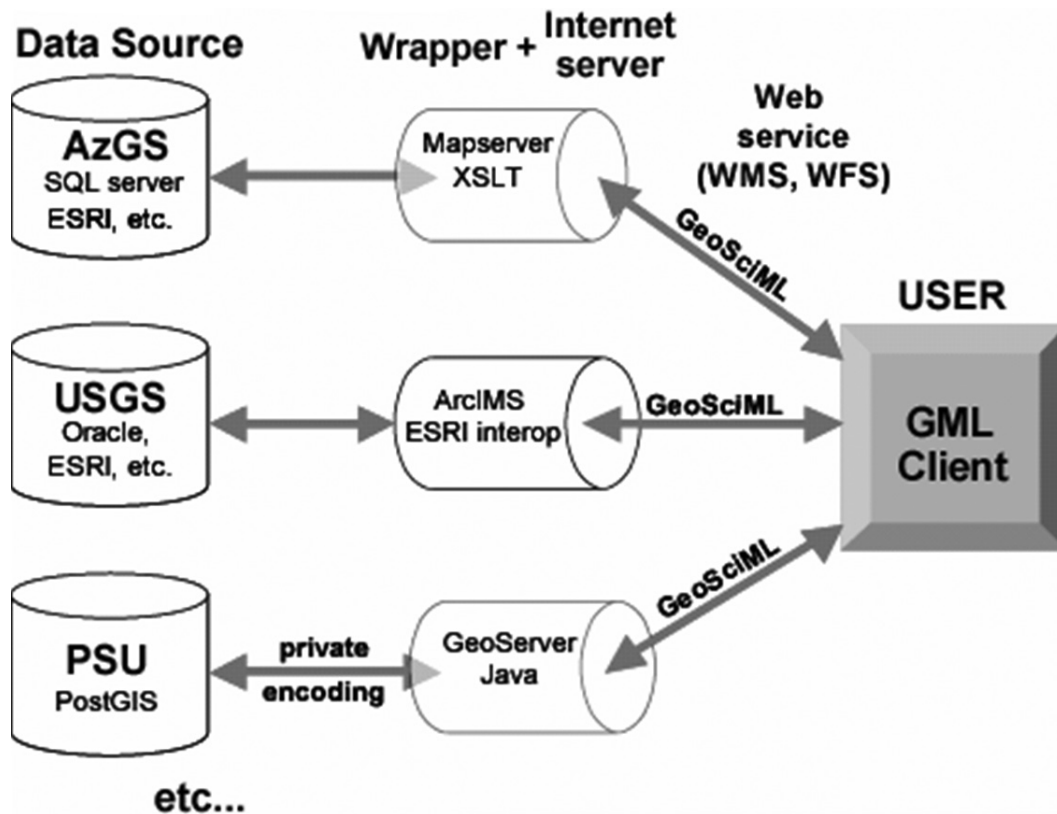


Figure 2. Communication between data providers and consumers utilizes standard GeoSciML schema. Clients that can interpret GeoSciML can operate with any GeoSciML-enabled data source.

<http://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/OandMCookbook>. See also Boisvert et al. (2004) from the USGS DMT 2004 workshop. Use of a standard graphical notation for model representation during development makes group analysis and review of the evolving model much easier.

Major Entities

Only a small part of the GeoSciML model is discussed here. See the GeoSciML Twiki (<https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML>) for more information about the full model. Figure 3 presents the logical framework that underlies the draft GeoSciML version 2 GeologicFeature implementation, a core aspect of GeoSciML. Starting from the center left, a MappedFeature associates a GeologicFeature with a GML_geometry that specifies a location on or within the Earth. The mapped feature may be the result of an Observation if observation-related metadata concerning identification of the mapped feature are recorded. Each GeologicFeature is associated with a ControlledConcept classifier that specifies the intention of what the GeologicFeature represents. A GeologicFeature may have one or more associated GeologicFeatureDescriptions that specify properties assigned to the feature. Each description may also be represented as the result of an Observation. Table 1 summarizes the packages included in the GeoSciML UML model. Each package is implemented as a separate XML schema.

Geologic Feature

In the draft GeoSciML version 2 model, GeologicFeature is an association class that binds mapped

feature(s) and description(s) with one or more classification concepts. Geologic feature is an entity that represents some particular phenomenon that may be observed in the Earth. It has a primary classification in terms of a controlled concept, and this association establishes a content model or concept space within which the feature is located/given identity by specification of a collection of properties in a description. A MappedFeature instance specifies a particular located occurrence of a geologic feature by associating it with a location (GML_geometry). GeologicFeatures may be classified by geologic unit or geologic structure ControlledConcepts terms. In addition to its primary classification (e.g. a lithostratigraphic designation), a feature may carry alternative classifications (e.g. geotechnical classification). GeologicFeature corresponds with a “legend item” from a traditional geologic map and with “occurrence” in conceptual models presented by Brodaric and Gahegan (2006) or Richard (2006). GeologicFeatures may have one or more associated GeologicFeatureDescriptions. Multiple descriptions associated with a feature may be the result of different observations (different observer, different time, different observation procedure...), or may specify different properties

Mapped Feature

A MappedFeature is a specific bounded occurrence, such as an outcrop or map polygon that carries a geometry or shape (through its samplingFrame association). It has an associated GeologicFeature instance that specifies what kind of thing is represented by the mapped geometry, both by classification with a vocabulary term (ControlledConcept) and through association with one or more description objects (GeologicUnitDescription) that specify property values.

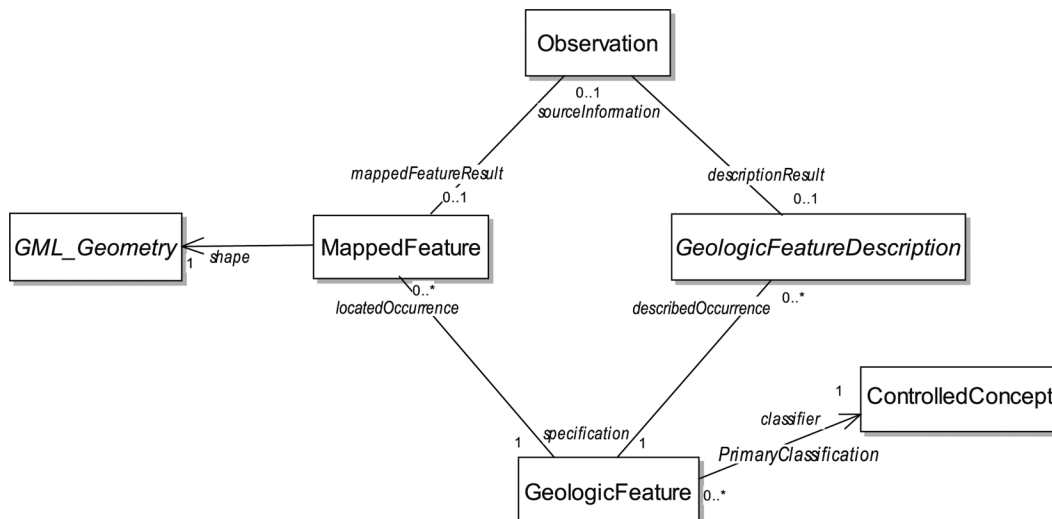


Figure 3. Core GeoSciML 2.0 logical model.

Table 1. Packages in GeoSciML model.

Package Name	Contents
TopLevel	The core model for mapped entities distinguishes between geologic features, mapped features, and controlled concepts
BasicTypes	Simple package, extends gml:MeasureType to represent quantification of measurements using relative comparisons, e.g. greater than, less than.
LiteralValue	The GeoSciML “value” model provides a generic way of encoding “literal” values, both textual and numeric, which have uncertainty and may be a range. These values are usually obtained as the result of an observation. The description of the associated observation event will provide more detail about the observation method, result quality, etc.
RootDoc	Generic collection element for packaging objects from the GeoSciML schema.
GeologicMetadata	Interim model for representation of dataset, feature, and attribute-level metadata. ISO 19115 metadata would be preferred, but the XML implementation (ISO 19139) is not yet supported by common software. The interim model is intended to have some forward compatibility with ISO. For example, the scope-codes are a subset of MD_ScopeCode from ISO 19115.
GeologicVocabulary	Model for controlled vocabularies of terms linked to normative descriptions, link to ontology. A GeologicVocabulary is a collection of terms (ControlledConcept) and relationships (VocabRelation).
BoreHole	Support for borehole data in GeoSciML is provided by XMML components. Borehole is modeled as a kind of sampling profile that may have various sorts of associated ‘logs’, modeled as kinds of coverages.
EarthMaterial	Earth Material is a class that holds a material description. A naturally occurring substance in the Earth. Earth Material represents substance, and is thus independent of quantity or location. Ideally, Earth Materials are defined strictly based on physical properties, but because of standard geological usage, genetic interpretations enter into the description as well.
GeologicAge	The age of a particular geological event or feature expressed in terms of years before present (absolute age), referred to the geological time scale, or by comparison with other geological events or features (relative age). A GeologicAge can represent an instant in time, an interval of time, or any combination of multiple instants or intervals. Specifications of age in years before present are based on determination of time durations based on interpretation of isotopic analyses of EarthMaterial (some other methods are used for geologically young materials). Ages referred to geological time scales are essentially based on correlation of a geological unit with a standard chronostratigraphic unit that serves as a reference. Relative ages are based on relationships between geological units such as superposition, intruded by, cross-cuts, or ‘contains inclusions of’.
GeologicRelation	Geologic Relations are typed, directed associations between geologic objects. Represents any of a wide variety of relationships that can exist between two or more GeologicFeatures. For example, the GeologicRelation ‘intrudes’ is a relationship between an intrusive igneous rock and some host rock. Includes spatial, temporal, sequence, correlation, and parent/child relations. Two or more GeologicFeatures are associated in a GeologicRelation; each has a role in the relationship. Examples of geological roles include “overlies”, “is overlain by”, “is younger”, “is older”, “intrudes”, “is intruded by”, and so forth. In a relationship where an igneous unit intrudes a sedimentary unit, the geological relationship is ‘intrudes’, the intruded sedimentary unit has the role ‘host’, and the igneous unit has the role ‘intrusion’. Many other types of relationships can also be accommodated via GeologicRelation, for example, topological relations between spatial objects could be described where they are scientifically significant.
GeologicTime	The GeoSciML Geologic Timescale model and encoding is described in detail in the paper ‘A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards’ (Cox and Richard, 2005). The classic “geological time scale” is a hierarchical ordinal system, in which the eras are ranked: “stages” nest within “series” within “systems” within “eras” within “eons” (in the most common version of the ranking system).
GeologicUnit	Package containing content model for geologic unit. Geologic unit is a notional unit, whose complete and precise extent is inferred to exist. Practically, spatial properties are only available through association with a MappedFeature. Includes both formal units (i.e. formally adopted and named in the official lexicon) and informal units (i.e. named but not promoted to the lexicon) and unnamed units (i.e. recognizable and described and delineable in the field but not otherwise formalized).
StructureObject	Package containing content model for geologic structure. Version 1 includes fault system, fault, contact, and fault displacement.

Controlled Concept

ControlledConcepts represent human concepts in computer form, typically as words (lexical objects) with an associated definition. Because GeoSciML extends GML, each controlled concept instance may have one or more `gml:names`, but the GeoSciML model adds a preferred name element that specifies one term that is used to identify the concept. Practically speaking, each preferredName should be associated with a unique concept, but in a distributed system, this cardinality cannot be enforced. ControlledConcepts are aggregated into GeologicVocabulary collections, which are derived from a GML dictionary. Data producers should ensure that preferredNames are unique within a particular vocabulary. A Controlled Concept may have an associated prototype entity (not shown in Figure 3) that can be a GeologicFeature, EarthMaterial, or Specimen. The prototype entity provides a mechanism to associate machine-analyzable properties with ControlledConcept terms. Similar functionality might be provided by links from the ControlledConcept to some other formal ontology representation.

Geologic Feature Description

Descriptions are collections of properties with assigned values (e.g. attributes) that characterize some feature. Different kinds of descriptions specify different properties. Descriptions may be associated with Observation elements that supply information on the origin of the property value assignments.

Observation

Observation describes the “metadata” associated with an information capture event, together with a value for the result of the observation. Observations are the basis for classified features, interpretations, and models. GeoSciML uses the Observation and Measurement model from the Open Geospatial Consortium (Cox, 2006), which models observation as a kind of event, in which a result value is assigned to some property of a feature of interest, using some procedure.

SOME SCHEMA DESIGN AND USAGE PATTERNS ISSUES

Names and Identifiers

Any GML Object or Feature may have an unlimited number of `gml:name` properties, which reflects the fact that the same object often has different identifiers assigned by different authorities. To assert “this is the name or identifier assigned by authority XYZ corporation,” use the `codeSpace` attribute on `gml:name` (i.e. the scope identifier).

If the codespace for the `gml:name` is not specified, then the value is implicitly under the authority of the organization or service that supplies the document, which should be indicated by associated document-level metadata.

Note that GML document elements also include a `gml:id` attribute, which plays a different role from the `gml:name` element. The value of the `gml:id` has `type="xsd:ID"`, so it must be unique within the (XML) document. It is a document fragment identifier that acts as a handle for an XML element in the scope of its appearance within a particular document, and is usually assigned by the information management system since it is primarily significant in that context. The `gml:id` supports cross-references within a document and references that involve individual nodes (elements) within a system of documents. The value of a `gml:name` has `type="gml:CodeType"`, which is a string with a “codeSpace” attribute. In the context of a GML object, the value of a `gml:name` is a label or identifier for the object described by the containing element, and is typically assigned by the data provider agency. The `gml:name` should be used for identifiers that are required to be persistent and are subject to constraints (e.g. uniqueness) applicable to a context wider than just the document scope. Different authorities may have different authoritative identifiers for the same item.

Namespace and Packaging

The namespace for GeoSciML version 2.0 schema is <http://www.cgi-iugs.org/xml/GeoSciML/2>. Versioning strategy for namespace evolution will follow practice described in [OGC 05-062r3](#). For future upgrades, each minor version of any such schema that retains the namespace of the predecessor shall not introduce any new XML types or elements that could not be safely ignored by existing application based on the previous minor version, which ensures a strong form of backward compatibility. Components from other namespaces (e.g., <http://www.opengis.net/om>) may also constitute a “canonical” part of GeoSciML but will be incorporated using the WXS import mechanism and, thus, retain their own namespace names.

The physical document location (path) for GeoSciML schema will include the complete version number—initially 1.0.0, moving to 1.0.x for bug-fix releases, and 1.1.x (etc.) for extensions that do not change the scope of the schema. Schema documents are hosted in the GeoSciML publish/build repository, which is at <https://www.seegrid.csiro.au/subversion/xml/GeoSciML/tags/>.

Use of Scoped Names

Use of scoped names, i.e., a term or word with an identifier for the source of the term, provides a method for linkage to formal controlled vocabularies (e.g. an ontology) that may then be used for semantic mediation. For

example, a GeoSciML file might have a property value specified by the following element:

```
<CGI_TermValue>
  <qualifier>common</qualifier>
  <value codeSpace="http://www.iugs-
    cgi.org/outcropCharacterVocabulary">
    ledge forming</value>
</CGI_TermValue>
```

The <value> element contains a scoped name “ledge forming” from the vocabulary specified by the codeSpace attribute. If the data interpreter is familiar with the “http://www.iugs-cgi.org/outcropCharacterVocabulary” codeSpace (vocabulary), then they may use the scoped name directly or by correlation with a preferred term in a different vocabulary. On the other hand, if the identified codeSpace (vocabulary) is not familiar and its identifier is a resolvable URL that points to service that can provide a definition of the term in a known format (e.g. free text, OWL, KIF...), it is possible to interpret the term. This may be as simple as someone studying a free text definition and determining the closest corresponding term in their vocabulary. An automated semantic mediator might be able to use a formal definition (e.g., OWL) to match with the closest subsuming term in a different formal vocabulary that is preferred by the data interpreter.

Value specification

The GeoSciML data model includes a flexible value specification scheme that is designed to capture value descriptions conventionally recorded by geologists. All

values may carry a qualifier. Numeric values include units of measure. Values may be specified in several manners:

- by a single numeric value with optional uncertainty, e.g., 5.24 +/- 0.12
- by a numeric range, e.g., 5.7-13.6
- by a term with an identifier for the source vocabulary, e.g., “thick-bedded (NADM SLTTs)”
- as a range with bounds assigned by terms or by a term and a numeric value, e.g. “fine- to medium-grained (Folk 1968)” or “Miocene (IUGS 2004)” – 1.7 Ma.

Instance Document Example

Example instance documents associated with each version of the schema in the subversion repository (<https://www.seegrid.csiro.au/subversion/xmlml/GeoSciML/tags/>) are stored in an “Instances” subdirectory in the directory for that version. The following listing provides an example usage of many of the elements for geologic unit description. The base element in the document is a GeoSciML collection (gsm!); each member of the collection starts with a <member> element. GeoSciML collection members may be:

1. Geologic features (a kind of GML feature)
2. GML geometry elements
3. Mapped Features (outcrops, sample locations, traverses/sections)
4. Controlled concepts (vocabulary definitions)
5. Geologic relationships
6. Dictionaries (collections of controlled concepts)

Comments in the following listing are delimited by ‘<!--’ and ‘-->’.

```
<?xml version="1.0" encoding="UTF-8"?>
<Gsm! xmlns="http://www.cgi-iugs.org/xml/GeoSciML/1" ... other namespace declarations>
  <!-- The lexicon would probably be in a separate file. The StratigraphicLexicon element extends GML dictionary (through GeologicVocabulary GeoSciML element) -->
  <member>
    <StratigraphicLexicon gml:id="AZGSGeologicUnits">
      <!-- This is a lexicon element that includes three units -->
      <gml:description>Collection of geologic units defined by State of Arizona</gml:description>
      <gml:name>Arizona stratigraphic unit lexicon</gml:name>
    </member>
    <ControlledConcept gml:id="MartinFormationConcept">
      <gml:description>lithostratigraphic formation defined by ... </gml:description>
      <gml:name>urn:x-cgi:def:lithostratigraphy:USGS:2006:Geolex:Martin-Formation </gml:name>
      <preferredName>Martin Formation</preferredName>
      <prototype xlink:href="#Feature2524"/>
```

```

        <vocabulary xlink:href="#AZGSGeologicUnits"/>
        <metadata/>
    </ControlledConcept>
</member>
<member>
    <ControlledConcept gml:id="LS2"> ... </ControlledConcept>
</member>
<member>
    <ControlledConcept gml:id="LS3"> ...</ControlledConcept>
</member>
</StratigraphicLexicon>
</member>

<member>
    <GeologicFeatureRelation gml:id="rel-100">
    <!-- This is a geologic relationship element-->
        <gml:name>urn:x-cgi:def:lithostratigraphy:USGS:2006:featureRelation:
            Stratigraphic position</gml:name>
        <role codeSpace="http://www.iugs-cgi.org/featureRelationVocabulary">overli
            es</role>
        <source xlink:href="#BeckersButteMemberPrototype"/>
        <target xlink:href="#JeromeMemberPrototype"/>
    </GeologicFeatureRelation>
</member>

<member>
<!-- GeologicFeature is derived from GML AbstractFeature, it associates a de-
description, a classifier (what is described) and an extent (where it was de-
scribed, if defined).The Classifier element defines the type of a feature. Mul-
tiple descriptions may be associated with a GeologicFeature -->
    <GeologicFeature gml:id="Feature2524"> <!-- This is a geologic unit GML feature,
        which is the basic container for geologic unit descriptions in GeoSciML v.
        1 -->
    <gml:description>The type section of the Martin Formation at Mt. Martin near
        Bisbee consists almost entirely of medium-gray to medium dark-gray
        aphanitic to fine-grained limestone. dolostone is entirely subordi-
        nate, ...
    </gml:description>
    <gml:name>urn:x-cgi:def:lithostratigraphy:USGS:2006:Geolex:TypeMartinForma-
tion</gml:name>
    <gml:boundedBy>
        <gml:Envelope>
            <gml:lowerCorner/> <!-- corners of a bounding box for type area of the
                Martin Formation; geometry specification elements not included
                here-->
            <gml:upperCorner/>
        </gml:Envelope>
    </gml:boundedBy>
    <purpose>typicalNorm</purpose>
    <age>
    <!-- Geologic age element includes a date value specification (see below), and an
event specification that explicitly identifies the event to which the age is as-
signed (e.g. deposition, cooling through biotite closure temperature...) -->
        <GeologicAge>
            <value>
                <CGI_TermValue>
                    <value codeSpace="http://www.iugs-cgi.org/geologicAgeVocabulary">Middle

```



```

        Devonian</value>
    </CGI_TermValue>
</value>
<event>
    <CGI_TermValue>
    <value codeSpace="http://www.iugs-cgi.org/EventVocabulary">deposition</
        value>
    </CGI_TermValue>
</event>
</GeologicAge>
</age>
<classifier xlink:href="#MartinFormationConcept"/> <!-- here's the link to the
    controlled concept that defines the intention of the Martin Formation.
    Link is reference to controlled concept instance in this document-->
<description>
<LithostratigraphicUnitDescription>
<metadata/> <!-- xlink to metadata for this description; this provides tie to
    Observation model-->

<partOf>
    <GeologicUnitDescriptionPart> <!-- 310- 340 thin bedded, non fossiliferous
        dolostone -->
    <unit>
        <LithostratigraphicUnitDescription gml:id="GeoUnitPart0235">
<!-- part is also a lithostratigraphic unit, uses same description schema as con-
    taining unit; it could have parts itself; partonomy is recursive. -->
        <descriptionSource xlink:href="reference to description source observa-
            tion" /> <!--Source observation element not included here -->
        <bodyMorphology xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <outcropCharacter xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <grossGenesisTerm xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <exposureColor xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <grossChemistry>
            <CGI_TermValue>
                <qualifier>always</qualifier>
                <value codeSpace="http://www.iugs-cgi.org/grossChemistryList">carb
                    onate</value>
            </CGI_TermValue>
        </grossChemistry>
        <rank codeSpace="http://www.iugs-cgi.org/Vocabulary">DescriptionPart</
            rank>
        <weatheringCharacter xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <metamorphicGrade/> <!-- not specified so implies same as containing
            unit -->
        <unitThickness>
            <CGI_NumericValue>
                <principalValue uom="meter">30</principalValue>
                <plusDelta uom="meter">20</plusDelta>
                <minusDelta uom="meter">10</minusDelta>
            </CGI_NumericValue>
        </unitThickness>
        <beddingStyle xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <beddingPattern xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
        <beddingThickness>
            <CGI_TermValue>
                <value codeSpace="http://www.iugs-cgi.org/ThicknessVocabulary">Thin-

```

```

        bedded</value>
      </CGI_TermValue>
    </beddingThickness>
  </LithostratigraphicUnitDescription>
</unit>
<role codeSpace="http://www.iugs-cgi.org/unitPartRoleVocabulary">Stratigraphic part</role>
<type>codeSpace="http://www.iugs-cgi.org/unitPartTypeVocabulary">DescriptivePart</role>
<proportion>
  <CGI_NumericValue>
    <qualifier>approximate</qualifier>
    <principalValue uom="percent">12</principalValue>
    <plusDelta uom="percent">0</plusDelta>
    <minusDelta uom="percent">0</minusDelta>
  </CGI_NumericValue>
</proportion>
</GeologicUnitDescriptionPart>
</partOf>
<!-- end of part descriptions. Following properties apply to entire described unit -->

<descriptionSource xlink:href="reference to description source observation" />
<bodyMorphology xlink:href="urn:x-ogc:def:nil:OGC:unknown" />
<outcropCharacter>
  <CGI_TermValue>
    <qualifier>common</qualifier>
    <value codeSpace="http://www.iugs-cgi.org/outcropCharacterVocabulary">ledge forming</value>
  </CGI_TermValue>
</outcropCharacter>
<grossGenesisTerm>
  <CGI_TermValue>
    <qualifier>always</qualifier>
    <value codeSpace="http://www.iugs-cgi.org/GenesisVocabulary">Sedimentary, marine</value>
  </CGI_TermValue>
</grossGenesisTerm>
<exposureColor>
  <CGI_TermValue>
    <qualifier>common</qualifier>
    <value codeSpace="http://www.color.org/ColorVocabulary">Light gray</value>
  </CGI_TermValue>
  <CGI_TermValue>
    <qualifier>common</qualifier>
    <value codeSpace="http://www.color.org/ColorVocabulary">Medium gray</value>
  </CGI_TermValue>
  <CGI_TermValue>
    <qualifier>rare</qualifier>
    <value codeSpace="http://www.color.org/ColorVocabulary">Pink</value>
  </CGI_TermValue>
</exposureColor/>
<grossChemistry>
  <CGI_TermValue>

```

```

    <qualifier>common</qualifier>
    <value codeSpace="http://www.iugs-cgi.org/grossChemistryList">carbonate</
      value>
  </CGI_TermValue>
</CGI_TermValue>
  <qualifier>occasional</qualifier>
  <value codeSpace="http://www.iugs-cgi.org/grossChemistryList">siliceous</
    value>
</CGI_TermValue>
</grossChemistry>
<rank codeSpace="http://www.iugs-cgi.org/Vocabulary">Formation</rank>
<weatheringCharacter xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
<metamorphicGrade>
  <CGI_TermValue>
    <qualifier>always</qualifier>
    <value codeSpace="http://www.iugs-cgi.org/metamorphicGradeVocabulary">not
      metamorphosed</value>
  </CGI_TermValue>
</metamorphicGrade>
<unitThickness>
  <CGI_NumericValue>
    <principalValue uom="meter">340</principalValue>
    <plusDelta uom="meter">10</plusDelta>
    <minusDelta uom="meter">10</minusDelta>
  </CGI_NumericValue>
</unitThickness>
<beddingStyle>
  <CGI_TermValue>
    <qualifier>common</qualifier>
    <value codeSpace="http://www.iugs-cgi.org/Vocabulary">Planar bedding</
      value>
  </CGI_TermValue>
</beddingStyle>
<beddingPattern xlink:href="urn:x-ogc:def:nil:OGC:unknown"/>
<beddingThickness>
  <CGI_TermRange>
    <lower>
      <CGI_TermValue>
        <qualifier>common</qualifier>
        <value codeSpace="http://www.iugs-cgi.org/Vocabulary">thin bedded</
          value>
      </CGI_TermValue>
    </lower>
    <upper>
      <CGI_TermValue>
        <qualifier>rare</qualifier>
        <value codeSpace="http://www.iugs-cgi.org/Vocabulary">thick bedded</
          value>
      </CGI_TermValue>
    </upper>
  </CGI_TermRange>
</beddingThickness>
</LithostratigraphicUnitDescription>
</description>
</GeologicFeature>
</member>
</Gsm1>

```

TEST BED DEMONSTRATION

Six national and two state geological survey agencies, in Australia, Europe, and North America, participated in a proof-of-concept demonstration of GeoSciML at the International Association of Mathematical Geologists (IAMG) meeting in Liege, Belgium, in September 2006. The demonstration showed that it is possible to access information in real time from globally distributed data sources. Geological map polygons and attribute information, and borehole data, were displayed, queried, and re-portrayed using web applications hosted by the Geological Survey of Canada and the French Bureau de Recherche Géologiques et Minières (BRGM). Functions demonstrated included continuous map portrayal with attribute query, reclassification according to attributes, and download of complex data structures encoded in GeoSciML.

Information delivery from different complex data stores using a community standard schema demonstrated that [GeoSciML](#) provides a data model and format capable of supporting transfer of geology data from multiple jurisdictions. This also demonstrated that a distributed data delivery system can be constructed by specifying standard interfaces, not limited to single vendor software. New services can be added easily, providing they conform to the interface. All of the services in the test bed used different data stores, wrapped by a variety of server software applications. Deployment requires configuration of server- and client-side software to conform to the data model, but does not require development of new software "from scratch."

Three use cases were demonstrated at the IAMG 2006 meeting in Belgium. Use Case 1 demonstrated display of map data and query for the description of a single

map object. When the client asks for the map, the server returns a map with default symbolization. A user can then click on any graphic feature from a layer to retrieve information for the feature, which can be presented to the user as raw GeoSciML or as a more clearly-rendered HTML version. Presentation formats other than HTML can be requested by the client if the server supports them. The types of features used must include at least one of the following: geologic units, faults, contacts or boreholes.

Use case 2 demonstrated selection and download of features; a geographic bounding box is specified and the contents downloaded as a GeoSciML document. The GeoSciML document can be reformatted (e.g. by XSLT for display in a browser) or serve as input for another process in a workflow. The GeoSciML document contains a collection of GeologicFeatures or Boreholes.

Use case 3 demonstrated dynamic query and re-symbolization of mapped features on the basis of age, using the IUGS standard geologic age color scheme, or on the basis of lithology, using a CGI defined lithology color scheme. The results of symbolization by lithology for data from Canada, the U.S., and Scandinavian countries is shown in Figure 4. A very simple lithologic classification and symbolization was used, with four classes and related colors: igneous (pink), sedimentary (green), metamorphic (purple), and unconsolidated (yellow). Each participant had to implement a mechanism to map from properties associated with the mapped features to the standardized lithology classes. It is the service provider's prerogative to determine the mapping from the data source to the classification.

SUMMARY

A standardized schema and syntax for information encoding is a fundamental requirement for interoperable infor-



Figure 4. Use Case 3 from Testbed 2, re-symbolization of geologic units by lithology for Canada, U.S. and Scandinavian countries: igneous (medium gray), sedimentary (light grey), metamorphic (dark gray), and unconsolidated (nearly white).

mation systems. The IUGS CGI Data Model collaboration working group has developed GeoSciML, an XML-based GML (geography markup language) application, to meet this requirement for the interchange of geoscience information. The schema for this application reuses existing markup languages where possible. Newly developed markup specifications are based on existing conceptual models in most cases. This standards-based data format provides a framework for application-neutral encoding of geoscience thematic data and related spatial data. It is intended for use in publishing or interchanging data between organizations that use different database implementations and software/systems environments. Full realization of data interoperability at the semantic level will require development of controlled vocabulary resources for specifying actual content. A Testbed demonstrated simple interoperability using web map and feature services (WMS, WFS) between geological surveys in several different countries. GeoSciML is being considered as a national standard for geoscience data exchange by federal and state geological surveys in Australia and the European Union Spatial Data Infrastructure (INSPIRE), and will be submitted in 2007 as an IUGS-CGI specification.

Development of GeoSciML is an open process with the intent to involve as many participants as possible. This will ensure development of a schema and services that will meet the needs of a wide variety of geoscience data producers and users. Three types of participation are available: 1) direct participation in GeoSciML development, 2) monitoring GeoSciML development via the web-collaboration tools and 3) deploying an internet server to provide data in GeoSciML format.

REFERENCES

- Boisvert, E., Johnson, B.R., Cox, S.J., and Brodaric, B.M., 2004, GML Encoding of NADM C1, in Soller, D.R., ed., Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-File Report 2004-1451, p. 95-103. Available at <http://pubs.usgs.gov/of/2004/1451/boisvert/index.html> (accessed 12/12/2006).
- Brodaric, B.M., and Gahegan, Mark, 2006, Representing geoscientific knowledge in cyberinfrastructure: some challenges, approaches and implementations, in Sinha, Krishna, ed., Geoinformatics-Data to knowledge: Geological Society of America Special Paper 397, p. 1-20.
- Cox, S.J.D., ed., 2006, Observations and measurements: Open Geospatial Consortium, Inc., document OGC 05-087r4, version 0.14.7, 168 pages. Available at http://portal.opengeospatial.org/files/?artifact_id=17038 (accessed 12/13/2006).
- Cox, S.J.D., Daisey, P., Lake, R., Portele, Clemens, and Whiteside, Arliss, eds., 2004, OpenGIS Geography Markup Language (GML) Implementation specification: Open GIS Consortium, Inc, document OGC 03-105r1, version 3.1.0, 580 p.
- Cox, S.J.D., and Richard, S.M., 2005, A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards: Geosphere, v. 1, p. 119-137, DOI: 10.1130/GES00022.1. Available at <http://geosphere.geoscienceworld.org/cgi/content/abstract/1/3/119> (accessed 12/12/2006).
- 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, 61p., available at <http://pubs.usgs.gov/of/2004/1334/>.
- Richard, S.M., 2006, Geoscience concept models, in Sinha, Krishna, ed., Geoinformatics-Data to knowledge: Geological Society of America Special Paper 397, p. 81-108.
- SC36 Secretariat, 2003, Proposed Draft Technical Report for: ISO/IEC 2382, Information technology—Learning, education, and training—Management and delivery—Specification and use of extensions and profiles: ISO/IEC 2382-01, ISO/IEC JTC1 SC36 N0646, available at <http://jtc1sc36.org/doc/36N0646.pdf> (accessed 12/12/2006).