GeoSciML: development of a generic GeoScience Markup Language

Marcus Sen[1] and Tim Duffy[2]

British Geological Survey[1] , Kingsley Dunham Centre, Nicker Hill, Keyworth, Nottingham, NG12 5GG. Corresponding Email : mase@bgs.ac.uk

British Geological Survey[2], Murchison House, West Mains Road, Edinburgh, EH9 1ED

Abstract

It is possible for a geoscience data providing organization simply to design an eXtensible Markup Language (XML) schema for a particular data exchange problem at hand and make it generally available. Though not ideal, this is an improvement on having many varieties of non XML data formats. This is because the XML format is partially self-documenting and provides common methods for parsing files, obtaining their structure and transforming them to alternative formats. If communities share a common data transfer model for their domains of interest however , data exchange becomes even easier and more likely to take place efficiently.

The British Geological Survey (BGS) believes it is important for work to develop some common ML for the exchange of generic geoscience information. This is not being done from scratch but, by using XML Schema modularity capabilities, is being developed as an application of the OpenGIS Consortium's (OGC) Geography Markup Language (GML) and is building upon the applied geoscience domain focussed eXploration and Mining Markup Language (XMML). These are both fairly complex specifications and BGS is developing from them in an incremental and piece-wise manner to cover some particular geoscientific domains of interest, such as boreholes, text and structural geology, all within one overarching

GeoSciML schema. With the support of suitable software tools the XML model development can proceed in an iterative manner with rapid prototyping incorporating the comments of experts in those chosen geoscientific domains or disciplines. We propose that a similar development process be extended to the wider geoscience community, with the support of bodies such as the International Union of Geological Sciences' (IUGS) Commission for the Management and Application of Geoscience Information and the OGC, so that GeoSciML style schemas, based on internationally agreed geoscience conceptual modelling, could become a common language for generic geoscience information exchange using XML based web processes.

Keywords: data exchange; standards; geoscience information; XML; GeoSciML

1. Introduction

The British Geological Survey (BGS) has been collecting geoscience data since it was founded in 1835, and is the UK national centre for earth sciences covering all the major disciplines within geoscience. It operates both in the UK and internationally . A major role of the BGS is one of data custodian, and in recent years it has put major effort into making its data more accessible both to internal and external users. A key new component to this is the use of web technologies including XML for data transfer.

Most corporate non-spatial data is managed in relational databases (Oracle), spatial data is also handled using CAD & GIS systems (Intergraph and ESRI for example). Individual projects use a whole host of heterogeneous formats from Excel to custom formats [1]. Considerable effort has been expended over the years to move data from heterogeneous project formats to properly designed, managed and accessible corporate data stores.

The use of XML is gradually starting to spread for both corporate projects and in work for particular clients. In order to help promote the development of an interchange language for geoscientific information BGS became sponsors of the eXploration and Mining Mark-up Language (XMML) consortium (Cox, 2004) [2]. As part of the UK Department of Trade and Industry (DTI) DEAL project XML has been used to update well header information from the DTI's Well Online Notification System, which is used by North Sea oil and gas operators to apply for drilling licences and to report drilling activity [3].

In the geosciences, it is not only numerical, categorical etc. type information that needs exchanging, but also descriptive text fragments, scientific reports and geological memoirs. This type of data has not been traditionally well suited to corporate databases but is the original raison d'être of XML (from its origins in the publishing world's ISO and W3C standard Standard Generalized Markup Language SGML [4]). Thus BGS has also been involved in the "text-style" use of XML for a project called the Textbase. This is a BGS sponsored system for the delivery of reports, field notes and other scientific text that has been attributed with subject metadata and linkages to corporate database. This allows both searching for relevant text data and following references in the retrieved text to other corporate data.

The requirement to exchange data between different organisations is not new. Successful data exchange involves agreeing on the format in which to exchange data at a number of levels. The basic file format might be binary or ASCII, use various field delimiters or have fixed length fields. At a higher level there will be some form of data model (formal or implicit, well-defined or ill-defined) behind the way data is stored. Whether to use a proprietry or open format is also an issue.

The problem at the file format level is often reasonably straightforward. For example, many cases are satisfied by simply using tab or comma separated ASCII file formats understood by a wide range of software. However, there is still the problem of converting between different data models. This may be straightforward if the models are isomorphic just using, for example, different names for the same items, but may be impossible if the models are incompatible. One approach is to try to get prospective data exchangers to agree to using standard data models but this can take a long time, standards tend to multiply and one size doesn't fit all.

2. XML for data exchange

XML is receiving a lot of attention and the question for geoscientists is what does it bring new to the problems of data exchange? For the low-level file format, using XML means file parsing can be carried out by common tools eliminating part of the work involved in data exchange. The human readability and partially self-documenting nature of XML documents also makes it easier to get an idea of the structure of someone else's data. At the data model level, common tools (e.g. DTDs, XML Schemas) can ensure data conforms at least partially to a particular data model. Of course, we still can't exchange data between incompatible models but common tools (e.g. eXtensible Stylesheet Language Transformations)(XSLT) exist to make the translation process easier if a mapping or partial mapping between models exists.

These features do not automatically enable exchange of data between different parties but they do reduce the amount of work that needs to be done. However, it does still seem reasonable to suppose that having a standard intermediate model or library of model pieces for a particular domain of interest would reduce the number of translations that need to be coded and thus promote data exchange. This latter point is not universally accepted and some people believe that attempts to build common data models are not really viable [5]. Alternative approaches emphasise the development of tools to help discern similarity of instances and enable (possibly partial) transformation between them (e.g. Tennison, 2002).

3. Existing relevant XML developments

The BGS has decided to build upon some already existing related major standards efforts.

Geography Mark-up Language (GML) is an OpenGIS Consortium (OGC) standard using W3C XML Schema (Cox et al. 2003). It provides components for representation in XML of many of the types of data commonly stored in geoscience GIS systems such as geometries, coordinate systems and topological relationships. It is intended to be used as a basis on top of which you create your own application schema using components from GML as part of the definition of features of interest in your own particular domain of interest (e.g. roads, rivers, boreholes...). GML version 3.0+ appears to have the potential to represent any form of 2D or 3D geology that might be required to be exchanged.

eXploration and Mining Mark-up Language (XMML) (Cox, 2004) [2] is a GML application with a geoscience scope that is focussed on the needs of the exploration and mining industries. As such it already addresses much of the domain of interest of a more generic geoscience mark-up language. Rather than unnecessarily duplicate existing work BGS decided to develop GeoSciML as a GML application with a wider geoscience scope that imports and builds on XMML. This takes advantage of the inherent modularity and building block approach of XML Schema and facilitates inclusion of developing namespaces and schemas .

3.1 GML design methodology

The core of GML is based on the feature model developed by ISO TC/211 (ISO 19109). This involves making a catalogue of feature types for your domain of interest (e.g. road, river,...) and then defining the properties that each feature type can have (e.g. length, name,...). For most feature types, at least one of these properties would be expected to be a spatial property such as a point location, a line string in space. This should be contrasted with the traditional GIS "geometry-first" approach where you first define whether you have a point feature, line feature or polygon feature and then define what further attributes it may have.

GML provides components for features, 0, 1, 2 or 3D geometry coordinate reference systems, topology, temporal information, definitions and dictionaries, units, measures, values and directions, observations (see also separate Observations & Measurements standard, Cox, 2003), coverages and styling. This is a library of components some of which provide alternative views on what could be the same data. Thus a single monolithic database structure is not enforced on everyone but they can share data stored in their private databases with other people using a number of common views that other people can understand. It is intended that different information communities with particular domains of interest should develop "GML application languages" that build on the components supplied by GML to create their own particular feature types.

The GML specification assumes a Model Driven Architecture development approach where a data model is designed in UML and then transformed according to certain conventions into a normative XML Schema specification. There is a mismatch between conventional object models and XML Schema (see e.g. [5]) . The GML conventions for how to transform a UML model into an XML Schema do address this mismatch. For example, the tag names for feature types and their properties are "striped" into alternate nested layers. The draft GML v3.1 documentation [7] contains an explicit UML profile for the transformation between UML models and XML Schemas. Although this enables an object-oriented design there are few (at the time of writing) XML tools that provide full information on the type hierarchy of a validated document to a processing application.

This is clearly still trying to develop a standard for people to agree to use and it won't satisfy all requirements but: it is more a library of useful model components than a completely defined data model. It does allow more than one way of organizing data. If there is a logical transformation from one model to another, standard XML processing tools like XSLT can often be used to define and perform the transformations.

3.2 XMML

XMML has been designed as a particular GML application focussed on the exploration and mining industries with the detailed areas of development being driven by the interests of consortium sponsors. It has included customisation of existing GML components (features, observations, temporal components,...), creating geologically useful geometric constructs from GML components (e.g TIN surfaces) and feature and property types for subject areas such as geochemistry, earth material description and mineral exploration projects.

4 GeoSciML Development Approach

Much development work has so far taken place by feeding directly into the XMML project which has covered a number of the major geoscience areas. A namespace has been declared with the OpenGIS consortium: http://www.opengis.net/GeoSciML for GeoSciML which is intended to be more generic and not just focussed on the exploration mining area but may end up being a fairly shallow wrapper substantially re-using XMML components and other relevant geoscientific and other GML application schemas that are developed. Thus a GeoSciML Schema such as in Figure 1 will usually also declare the XMML as well as GML namespaces and import Schemas from these as required. In the first phase of development of GeoSciML, rather than attempt to build sophisticated models the focus has been on using the existing XMML and GML structures to deliver examples of real BGS data. This process brings to light details of the implementation that need refining in the design patterns. To date a few trial geoscience subject areas have been selected which cover our currently identified data exchange requirements with BGS customers: boreholes, structural geology and text. Existing XML software has been used to generate prototype GeoSciML instances from existing corporate data. This has included XSLT processors, Oracle XSQL and the ESRI ArcIMS OGC WFS connector.

GML and XMMLaware software is under development but is not well-advanced so the work has concentrated on checking whether BGS data can be transformed to fit in the proposed formats rather than doing any sophisticated processing of the resulting XML.

Fig. 1. **<u>GeoSciML</u> Schema sample**

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema targetNamespace="http://www.opengis.net/GeoSciML"
xmlns:gsml="http://www.opengis.net/GeoSciML"
xmlns:xmml="http://www.opengis.net/xmml"
xmlns:gml="http://www.opengis.net/gml"
xmlns:sch="http://www.ascc.net/xml/schematron" xmlns="http://www.w3.org/2001/XMLSchema"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
attributeFormDefault="unqualified" version="3.1.0">
    <xsd:annotation>
        <xsd:documentation>
            Top level GeoSciML schema
        </xsd:documentation>
    </xsd:annotation>
    <!-- ====================================================================== -->
    <xsd:include schemaLocation="gsmlBase.xsd"/>
    <xsd:include schemaLocation="geologicalMaps.xsd"/>
    <xsd:import schemaLocation="../XMML/borehole.xsd" namespace="http://www.opengis.net/xmml"/>
    <!-- ====================================================================== -->
</xsd:schema>
```

Note that all schemas and instance documents have been validated against the draft GML v3.1 Schemas and current versions of the XMML Schemas at the time of writing. Also, BGS namespaces (of the form http://ns.bgs.ac.uk/...) in the example instance documents are for illustrative purposes only and have not been officially assigned particular roles by BGS.

4.1 Boreholes

BGS holds data for over a million boreholes, from simple index information on (almost) all boreholes drilled in the UK to geological log, water level, geochemical etc. data from boreholes. It is currently delivered by a variety of front-ends: web-forms leading to HTML tables, SVG graphical log viewer and Internet GIS (ArcIMS based).

As there already existed a well-developed borehole model in XMML we have concentrated on transforming data from our borehole databases into this format. An example, including some geological log information is shown in Fig. 2. Certain properties can be transformed directly into the XMML representation such as the start position of the borehole in the position property. Others that exist in our databases but not in the XMML model may use a soft-typed property such as the comment property so that the information isn't lost, although the XMML does not define its significance or how it is supposed to be used. An alternative would be to define a BGS specific Schema building on the XMML borehole model that defines the additional properties only present in the BGS databases. Other properties such as the lithostratigraphic log values have a defined place in the log property but the full significance of the values depends on familiarity with the dictionary being used. The example instance uses various codeSpace attributes to specify certain BGS specific dictionaries. Applications that receive data encoded in this way could understand it at different levels. A generic GML application would be able to plot the locations and names of the boreholes on a map and note that they were features called "Boreholes" without knowing anything about the XMML. A more specific XMML application might be able to plot the logs with the text values given. Thus this is not an "all-or-nothing" encoding format which applications either understand or don't understand, but rather one in which both the level of detail that is encoded and the level that is "understood" at the other end can vary. This flexibility may help to ease the introduction of these formats as simple levels of implementation can still be useful; but useful software tools are still scarce at the moment.

Fig. 2 Encoded example of borehole data from BGS database

```
<?xml version="1.0" encoding="UTF-8"?>
<gml:FeatureCollection xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="http://www.opengis.net/GeoSciML"
xmlns:gml="http://www.opengis.net/gml" xmlns:sch="http://www.ascc.net/xml/schematron"
xmlns:xmml="http://www.opengis.net/xmml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/GeoSciML ../../GSML/gsml.xsd">
    <gml:boundedBy><gml:Null>missing</gml:Null></gml:boundedBy>
    <gml:featureMembers>
        <xmml:Borehole gml:id="bh.121911">
            <gml:metaDataProperty><xmml:AccessRestrictions>
                    <xmml:text>NO CONDITIONS APPLIED (NON-CONFIDENTIAL)</xmml:text>
            </xmml:AccessRestrictions></gml:metaDataProperty>
            <gml:name codeSpace="http://ns.bgs.ac.uk/bgs.sobi.pk">SE63NE/BJ/7/.</gml:name>
            <gml:name codeSpace="http://ns.bgs.ac.uk/bgs.sobi.bore_name">NEW COUNCIL
HOUSES</gml:name>
            <xmml:comment>NORTH DUFFIELD YORKS</xmml:comment>
            <xmml:comment>Instigator: URBAN OR RURAL DISTRICT COUNCIL</xmml:comment>
            <xmml:comment>Purpose: NOT ENTERED</xmml:comment>
            <xmml:comment>Paper record stored at: WLKW</xmml:comment>
            <gml:position>
                <gml:Point gml:id="bh.121911.start" srsName="epsg:7405"><gml:pos>468410 0437120
10</gml:pos>
```

```
                </gml:Point>
            </gml:position>
            <xmml:length uom="#m">55.17</xmml:length>
            <xmml:shape><xmml:BoreholeSurvey gml:id="bh.121911.shape">
                    <xmml:origin xlink:href="#bh.121911.start"/>
                    <gml:direction>
                        <gml:DirectionKeyword                codeSpace="http://ns.bgs.ac.uk/bgs.dic_incl">NOT
ENTERED</gml:DirectionKeyword>
                    </gml:direction>
            </xmml:BoreholeSurvey></xmml:shape>
            <xmml:log>
                <xmml:IntervalLog dimension="1">
                    <gml:boundedBy>
                        <gml:Null>missing</gml:Null>
                    </gml:boundedBy>
                    <xmml:intervalSetDomain>
                        <xmml:MultiInterval srsName="#bh.121911.shape">
                            <xmml:intervalMembers>
                                <xmml:Interval>
                                    <gml:pos>0</gml:pos>
                                    <gml:pos>4.57</gml:pos>
                                </xmml:Interval>
                                <xmml:Interval>
                                    <gml:pos>4.57</gml:pos>
                                    <gml:pos>6.09</gml:pos>
                                </xmml:Interval>
...
                            </xmml:intervalMembers>
                        </xmml:MultiInterval>
                    </xmml:intervalSetDomain>
                    <xmml:range>
                        <xmml:BandList>
                            <xmml:categoryList          property="http://ns.bgs.ac.uk/thesaurus/lithostratigraphy"
codeSpace="http://ns.bgs.ac.uk/bgs.lex_lithostrat_unit_v3">null TILL null    GLLD    GFDU    GLLD    SSG    SSG
</xmml:categoryList>
                            <xmml:categoryList                property="http://ns.bgs.ac.uk/thesaurus/lithology"
codeSpace="http://ns.bgs.ac.uk/bgs.dic_rock_name">S CV S C SAGR C S SDST </xmml:categoryList>
                            <xmml:categoryList      property="http://ns.bgs.ac.uk/thesaurus/description">including
!SUB_SOIL
possibly a "boulder clay"?
null
?GLLD "tight"
?GFDU
?GLLD !SOFT
!WEATHERED_SOLID
including "Calais Sand" and "Calais Gravel" !SOFT
</xmml:categoryList>

                        </xmml:BandList>
                    </xmml:range>
                </xmml:IntervalLog>
            </xmml:log>
        </xmml:Borehole>
    </gml:featureMembers>
</gml:FeatureCollection>
```


4.2 Text

The BGS has developed an in-house application (Textbase) to retrieve report fragments and other text based on subject meta-data and what features it describes. The subject terms can be taken from a general geoscience thesaurus or corporate dictionaries of lithostratigraphy, lithology or chronostratigraphy. The linked features can be boreholes and computer models that have been archived in the corporate database. There can also be attribution with a geographically defined area. We wanted to add descriptive text as data available for retrieving alongside other geoscience data as part of an OGC Web Feature Service (WFS) (Vretanos 2002).

Textual knowledge is an important part of the potential knowledge repository of geoscience information in an organisation such as the BGS. Only in recent times is such text being captured as data, which enables immediate access for query and delivery via digital databases and web services. Practical knowledge management schemes that have focussed on bringing the written word on-line have found that "knowledge summaries" – defined as encompassing the best accumulation of expertise on a particular topic built up over a period of time - can be very effective e.g. [9]. The traditional BGS Geological memoir on a defined area can be considered an example of such knowledge summaries but in geological science and funding timescales are only updated infrequently which means that texts like these are prime candidates for being made available as part of the digital knowledge offerings held as marked up text. Rather than returning an entire volume, as in a traditional text web search, this marked up text can be used to return pieces that are specific to the topic and geographical area of interest.

Text therefore is an important part of geoscientific data but is different in nature from the usual kinds of data encoded in GML. In particular it requires the use of XML mixed content which is normally avoided in GML. The OGC has previously trialled an application called Location Organizer Folders [8] to bring together assorted media types including text in a GML framework.

This has two types of text. The first is a narrative element which describes some event or story and is marked up to refer to multimedia raw source material such as video clips, sound files or text extracts (e.g. from an on-line newspaper). The latter raw source text fragments are the second type of text. This is for a different sort of application to the delivery of already attributed text which is the object of the BGS Textbase. The subject metadata and described feature attribution of the Textbase are naturally presented as subject and relatedFeature properties of a NarrativeText feature as illustrated in Fig. 3. The geographic attribution can be encoded in a subjectLocation property which can hold a variety of standard GML spatial location types from the simple text string illustrated to points, polygons etc. The textContent property is allowed to contain a single element from any other namespace which may contain mixed content text marked-up according to some other schema, but this is opaque and not processed by a GML level application. The example in Fig. 3 has text marked up according to the BGS Textbase schema but other text markup such as XHTML or DocBook could be used instead. A GML/GeoSciML level application can process the text subject and feature attribution and deliver the marked-up text, a more specialist text application would be needed to process the text content according to how it was marked-up. This model does not allow the attribution of text at multiple hierarchical levels of a document (such as book, chapter, section, etc.) which the Textbase application does. Implementing this in a GML pattern, although theoretically possible, would be clumsy and is judged to be inappropriate at this stage.

Fig. 3. Encoded example of text from BGS Textbase

```xml
<?xml version="1.0" encoding="UTF-8"?>
<gsml:NarrativeText gml:id="_TR_99_3.sect1.5" xmlns:gsml="http://www.opengis.net/GeoSciML"
xmlns:xmml="http://www.opengis.net/xmml" xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/GeoSciML ../../GSML/gsml.xsd">
    <gsml:subjectLocation>
        <gml:LocationString>within the Wyville Thomson Ridge, Faroe-Bank Channel and surrounding
area.</gml:LocationString>
    </gsml:subjectLocation>
    <gsml:subject codeSpace="http://ns.bgs.ac.uk/Textbase/AMF">2154</gsml:subject>
    <gsml:subject codeSpace="http://ns.bgs.ac.uk/Textbase/AMF">2189</gsml:subject>
    <xmml:relatedFeature xlink:href="http://urn.bgs.ac.uk/SOBI/TQ43NE6BJ."/>
    <gsml:textContent>
        <para xmlns="http://ns.bgs.ac.uk/Textbase">
            <span id="auto837394">Well 163/6-1A</span> is the only commercial well in the area covered by
seismic reflection data (Fig. 3) and is located on the NW flank of the Darwin Igneous Complex (Morton et al. 1988;
Abraham and Ritchie 1991). It was drilled by a consortium of companies in 1980, to a depth of 3686.3 m below RT
(rotary table), and proved 1252. 4 m of mainly Paleogene and some Neogene sedimentary rocks resting on 689
metres of Paleogene basaltic lavas and terminated in 356 m of Paleogene acidic igneous rocks. Unfortunately,
operational difficulties forced a premature abandonment of the well and a calibrated velocity log was not run.
</para>
    </gsml:textContent>
</gsml:NarrativeText>
```

4.3 Structural Geology

A major part of BGS output has historically been the traditional geological map. As technology has progressed the processes have developed into digital cartography [1] with the possible delivery of attributed GIS data as well as printed maps. BGS is moving towards a more geological GIS database system designed to model more closely the scientific observations and inferences of field mapping geologists. This can be used to produce more than just printed maps and will be the input for the development of fully three-dimensional models. This kind of data also needs exchanging so draft representation of data from this database has been prototyped in GeoSciML. The "Geoscience Spatial Database” (GSD2) currently has an internal draft implementation but has not reached its final form and there is not a large amount of data in this format yet. Thus experimentation with GeoSciML representations of this data has used data from the existing digital cartography system with some view of the likely structure of the new GSD2 system.

The scope of the system is mainly 2D mapping data such as mapped solid geology, altered ground, surface fault traces etc. with geological and some structural information such as relationship with contacted units. There are also features to deal with the mapped landforms, point observations and measurements used by field geologists in the construction of their mapped interpretations. The main features are what are called geology extents which represent the mapped polygons common on a gelogical map showing, for example, the area where a particular geological unit is present at surface (the bedrock surface in the case of solid geology). There are a number of ways these can be represented in a GML style, illustrating how GML can provide some common points of reference to make the process of exchanging data easier but doesn't enforce a rigidly defined structure which means that it can cope with data stored in varying ways. However, there is still some work to do when exchanging data between different parties.

The simplest method is to define a feature type with a list of dictionary properties and one spatial property for the polygon as illustrated in Fig. 4. One can imagine a generic GML application or toolkit could be used to easily plot these on a map and allow the properties to be queried although the exchanging parties would have to share a common understanding of the dictionary values. Even at this simple level there are some design decisions for the exchange format. For example, the BGS DigMap system (a digital geological map of the UK, [1]) specifically has four geology extent feature types (bedrock, superficial deposits, artificial ground and areas of mass movement) which all share the properties such as: lithostratigraphic classification, rock classification, bed, member, formation, group, age, stage, series, system etc.

There are two properties that really describe the mapped unit; the rest are just properties associated with the unit and rock type, and are not specific to the particular polygon. Clearly in the underlying relational database design these are not redundantly repeated. However, for the purposes of data exchange it may be useful to supply a view of the data as polygons with attributes for all the properties of interest. This would allow, for example, someone with a simple GIS to plot attributed polygons without the need to have available a full relational dictionary model of lithostratigraphic units and their properties. So this is a legitimate model for a data exchange "view" even if it wouldn't necessarily be a good model for a relational database implementation.

One needs to ask which properties are of a generic type likely to be useful to all geology users and which should be left to be added by organization-specific schemas built on top of GeoSciML. A dictionary defined mapped unit is clearly going to be common to all users, a form of rock description is likely to be quite common. The other properties may or may not be present in different cases, the question is whether it is possible to define optional properties which would satisfy anyone wanting to specify, for example, an age, or if it is best to leave different users to define these more specific properties in their own schemas based on the generic feature type.

Further development to represent the relationship of the mapped polygon onto the surface on which it is being projected (e.g. ground surface or bedrock surface), including surface data so that this can be used as a component in developing a 3D model and representing the topological relationships explicitly using GML patterns is under investigation.

Fig. 4 Encoded example of geological map data from BGS DigMap

```
<gml:FeatureCollection  xmlns:gml="http://www.opengis.net/gml"  xmlns:gsml="http://www.opengis.net/GeoSciML"
xmlns:xmml="http://www.opengis.net/xmml"                xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/GeoSciML
../../GSML/gsml.xsd">
    <gml:boundedBy>
        <gml:Envelope srsName="epsg:27700">
            <gml:lowerCorner>443554.612 1179784.018</gml:lowerCorner>
            <gml:upperCorner>466908.386 1217077.305</gml:upperCorner>
        </gml:Envelope>
    </gml:boundedBy>
    <gml:featureMember>
        <gsml:BedrockUnitExtent>
            <gsml:mappedUnit
codeSpace="http://ns.bgs.ac.uk/bgs.lex_lithostrat_unit_v3">NMMUS</gsml:mappedUnit>
```

```
            &lt;gsml:shape&gt;
                &lt;gml:Polygon srsName="epsg:27700"&gt;
                    &lt;gml:exterior&gt;
                        &lt;gml:LinearRing&gt;
                            &lt;gml:posList&gt;455313.386    1195067.109    455314.533    1195067.697    455316.358
1195064.138 455313.386 1195067.109&lt;/gml:posList&gt;
                        &lt;/gml:LinearRing&gt;
                    &lt;/gml:exterior&gt;
                &lt;/gml:Polygon&gt;
            &lt;/gsml:shape&gt;
            &lt;gsml:lithology codeSpace="http://ns.bgs.ac.uk/bgs.dic_rock_name"&gt;SCHH&lt;/gsml:lithology&gt;
        &lt;/gsml:BedrockUnitExtent&gt;
    &lt;/gml:featureMember&gt;
&lt;/gml:FeatureCollection&gt;
```

5 Geoscience disciplines that could expand the scope of generic GeoSciML

The BGS is also interested in the geochemical, geophysical and geotechnical disciplines that may in the future need XML schema for data exchange. There is currently active development by the XMML project of geochemical and geophysical schemas which look likely to cover all the BGS needs.

GeoSciML aims to be generic both in the sense that it could cover a general purpose variety of geoscience disciplines taking advantage of the inherent modularity gained by using XML schema and namepaces , and also in the sense that experience so far indicates that the level of detail of data exchange within any specific geoscience discipline is likely to be pitched somewhere between a broad brush level and the very detailed level actually stored in a geoscience data providers datastore. The latter level of detail is likely to contain institution specific elements that external users may not be interested in. GeoSciML will naturally move towards a more conceptual model of internationally agreed geoscience discipline information requirements that would help to both define the sub-set of actually stored data that it is meaningful to exchange and also to express it. This could lead to a geoscience model that was understood both by the client in terms of requesting and receiving data and the provider in terms of what they can offer.

6 Future work

Designing and understanding a comprehensive geoscience data exchange model is difficult. There is a need to iterate between geoscientists' use-cases and the developing prototype exchange formats. The BGS proposes that a similar development process be extended to the wider geoscience community, with the support of bodies such as the International Union of Geological Sciences' (IUGS) Commission for the Management and Application of Geoscience Information and the OGC, so that GeoSciML style schemas could become a common language for generic geoscience information exchange using XML based web processes. Geoscience data serving web-services (e.g. WFS) are envisaged that will allow the client to see the extent of geoscience disclipline data that is available, expressed concisely in a conceptual model. This will allow users to choose the pieces they desire according to discipline and forms of content (tables, polygons, logs, texts, etc.) and to geographic extent, giving the customer of the future precisely what they specify.

Acknowledgements

The authors acknowledge the extensive contribution made to the development of GeoSciML by Simon Cox of CSIRO, and to their many BGS colleagues involved with data architecture, design, modelling and implementation**.** This paper is published with the permission of the Executive Director of the British Geological Survey.

References

Cox, S. (Ed.) 2003. Observations and Measurements. OGC Document Number: 03-022r3.

Cox, S., 2004. A standards conformant XML language for geology and mineral exploration. Computers & Geosciences, this issue.

Cox, S., Daisey, P., Lake, R., Portele, C., Whiteside, A. (Eds.) 2003. OpenGIS® Geography Markup Language (GML) Implementation Specification (Version 3.00). OGC Document Number: 02-023r4.

Tennison, J., 2002. Comparing Markup Languages. Proc. Extreme Markup Languages, 4-9 August 2002, Montreal, Canada.

Vretanos, P.A., 2002. Web Feature Service Implementation Specification. OGC Document Number: 02-058.

Internet References

[1] BGS Earthwise magazine number 16 2003.http://www.bgs.ac.uk/magazine/magazines/docs/ewise16.pdf

[2] XMML project Web Home. https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome

[3] Description of Schemas used by DEAL (http://www.ukdeal.co.uk/schemas/)

[4] SGML home page. http://www.w3.org/MarkUp/SGML/

[5] xml-dev mailing list. http://www.xml.org/xml/xmldev.shtml

[6] Meijer, E., Schulte, W., Bierman, G. Programming with Circles, Triangles and Rectangles. (http://www.research.microsoft.com/~emeijer/Papers/XML2003/xml2003.html)

[7] Draft Geography Markup Language Version 3.1 documentation and Schemas package. (http://portal.opengis.org/files/?artifact_id=4700)

[8] Location Organizer Folders proposal.http://www.opengis.org/docs/01-037.pdf

[9] Knowledge Summaries within the TWI (The Welding Institute) World Centre for Materials and Joining Technology. http://www.twi.co.uk