XMML

STANDARDS-COMPLIANT TRANSPORT OF GEOSCIENTIFIC DATA ONLINE IN THE EXPLORATION AND MINING INDUSTRY

CSIRO

CSIRO Exploration and Mining
PO Box 437
Nedlands WA 6009
Attn: Dr Simon Cox
Simon.Cox@dem.csiro.au

FRACTAL GRAPHICS

Fractal Graphics p/l
P.O. Box 437
Nedlands WA 6009
Attn: Dr Nick Archibald
nja@fractalgraphics.com.au

2000-05-22

Table of Contents

1. Summary

Web-based information standards and geospatial data are maturing to a stage where they can support sophisticated and transparent exchange of technical information. By attending to a small number of connectors among generic technologies the exploration industry can take advantage of widely available tools in order to achieve a situation where

(a) exploration datasets will be organised according to well-understood common models, in file formats compatible with e-business protocols, leading to a situation where the corporate database extends transparently into multiclient and open file data on the World Wide Web

(b) data is available for immediate import into and transfer between a variety of applications software, allowing developers to concentrate on their distinctive functionality, and users to mix-and-match this according to their processing needs.

We propose to develop the eXploration and Mining Markup Language **XMML**, a web-compatible **XML** based exploration and mining data transfer format. This will use a sophisticated geology domain model built on the **ISO** geographic standards, **OpenGIS Consortium** implementations, and **World Wide Web Consortium** encoding recommendations. Because the geology model is built merely as a “schema” on top of a generic geospatial infrastructure, it will be compatible with both generic (e.g. GIS, CAD, DBMS, spreadsheet, web-browser) and specialised (geology modelling, mechanics and fluid-flow, resource estimation, mine-planning etc) software for analysis, modelling, visualisation and transfer. The system will be capable of describing rich 3D geology, including boreholes, geophysics and analytical data, so that data can easily be exchanged between software applications, between offices, and between explorers, contractors, data-managers and regulators on a transactional basis.

The self-describing plain-text form of XML documents also makes them ideal for archival purposes, overcoming the problem of loss of data because of software incompatibilities.

Keywords: XML, ISO/TC 211, geographic information, geomatics, geology, mining, online, web, WWW, e-commerce, e-business, standard

2. Introduction

The World Wide Web is causing rapid development of tools for information sharing. Our desktop, or even our corporate information system, no longer defines the limits of the information at our fingertips. At the same time affordable desktop computers are capable of data processing and visualisation that was until recently only possible on specialised systems.

These tendencies have yet to have a proper impact on the exploration industry. In most enterprises data is held locally in proprietary formats for use within each specialised software application, and re-use in another can be difficult and is often accompanied by net information loss. A plethora of file-formats means that databases are fragmented and effective cataloging is difficult. Not uncommonly this leads to duplicate processing and even acquisition. Data fusion is always a challenge due to incompatible data organisation, quality, survey basis, etc, so the full value of the information acquired sometimes at great expense is rarely realised.

Although email is now routine for messaging, and downloads of complete datasets relatively common, this is on a largely *ad-hoc* basis. Web-based e-business technologies, now taking over from the previous generation proprietary data interchange systems, are confined to commercial and production divisions of enterprises.

We look forward to an exploration industry with

?? Web-based exchange of technical data and models; seamless integration with desktop applications, database and GIS systems, visualisation, modeling and simulation software.

?? Discovery and download of data directly from multi-client and open-file databases; automatic management of metadata for catalogue and access control.

?? Upload of statutory reporting data direct to mines-department databases.

?? Online specification of lab analyses; results downloaded direct to database.

3. Context

A variety of relevant generic and specific technologies are maturing.

3.1 Internet and generic standards

Highlights are:

XML (eXtensible Markup Language) – is a text-based format for encoding structured data, designed to be compatible with web-based data transfer and e-business

ISO/TC 211 – is finalising a set of standards for geographic data, providing complete geometry for features, an extensible metadata model, and a method for defining domain specific feature models

OGC (OpenGIS Consortium) – is the interface to GIS vendors: producing implementations based on the ISO abstract specs (e.g. GML – geography markup language), and demonstrations of next-generation solutions (e.g. WMT – web mapping testbed)

Accelerated **3-D graphics** is now affordable, commodity hardware, controlled through standard software api's (Open-GL, Java-3D, and soon SVG, X3D)

See Appendix A: and Appendix B: for more detailed discussion.

3.2 Exploration and mining industry-based data specifications

Considerable work has been done on exploration and mining industry **datamodels**, including:

AMIRA P431 Geoscience Data Model
(primarily geochem sampling and drillholes)

GGIPAC guidelines for digital submission
(composite standard, with detailed geochem & drillhole formats, metadata)

Geochemistry sample analysis formats from WMC ExDiv, Genisys

POSC datamodel for petroleum, including the DTD based WellLogML
(strong commercial components)

Geophysics data and model-control formats from SEG, DFA, Encom, Fugro

In particular, several years' work by CSIRO and Fractal Graphics has led to the development of two significant and related applications: CSIRO's AMIRA-sponsored **Data Translator**, and Fractal Graphics **FracViewer**. These both use a consistent geometry model, which is based on spatial components drawn from the ISO/TC 211 model. The Data Translator has an interface called the **GeoEditor** based on SGI's OpenInventor, and its data model attaches attribute information to objects using a generic method, necessary for an application that imports and exports a wide variety of file-formats. On the other hand, FracViewer is a visualisation tool using OpenGL Optimiser, but accessing data served from an object-oriented database called **FracSIS**, in which the geological objects are defined according to a controlled feature model. The FracSIS model supports the major functions required for 3-D geological interpretation and analysis, and includes most of the capabilities of the public standards such as P431 and GGIPAC.

FracSIS and CSIRO's datamodel are implemented as Object Models within software. Data is stored in binary format, but may be exported in a variety of "legacy" file formats. However, this is a lossy process since the legacy formats are incapable of representing the data completely.

3.3 Aligning Exploration and Mining with the wider world

This is where the opportunity offered by the standards comes in. The ISO/TC 211 approach is explicitly designed to support a variety of applications, by providing basic geometry, to which domain- and application-specific attributes and behaviours are added (more details in Appendix B:). ISO uses UML for abstract models, which ensures consistent codable structures, and XML as the encoding for data-instances, which ensures compatibility with web-based transport protocols. Support for ISO based data is guaranteed because of the commitment of statutory bodies to the ISO process. OGC is facilitating this by leading implementations with the participation of all the major GIS vendors.

Models of information in the exploration and mining industry frequently go beyond the capabilities of mainstream GIS because of the true 3-D nature of the domain. Nevertheless the ISO model extends to 3-D, and the CSIRO and FracSIS geometry and feature models are compatible with the ISO feature model. Furthermore, the CSIRO and FracSIS models are not merely abstract ideas, but have been implemented in high-quality software and extensively tested, and are thus immediately available for use in certain applications. The efficiency and scalability of these representations is understood and appears to be good.

The intention of this project is to leverage these early achievements in order to enable more general use of data-interchange standards in the Exploration and Mining industry, derived from the ISO/TC 211 standards and transported using XML. Data-streams in a standard XML format corresponding to a standard ISO-based model will be available for input to and output from a rich suite of applications packages, such as simulation and modelling, visualisation, estimation. The aim is to provide a framework in which software developers can focus on the distinctive functionality of their packages, while maintaining interoperability with complementary applications through data interchange conforming to a robust common model. Data providers and users may then transport information between packages using the standard, and even store and archive data using the XML document format (see Appendix C: for some use-cases).

3.4 Uptake issues

Why should this project be different to all the previous "standards" and models that failed to have a widespread impact (SDTS, P431, etc)?

?? a critical mass of the GIS industry is participating

?? this is a direct way to align our domain (exploration/mining) with future developments in the broader geospatial information community, allowing us to take advantage of commodity software more easily

?? several of the statutory agencies that we deal with routinely (AUSLIG, AGSO) are already involved in the ISO and OGC initiatives, mainly concerning map-datasets

?? the internet is maturing so that it is now a routine part of business practice; we can ride on the back of generic developments in e-business, etc.

The project will have been successful when industry, service companies, vendors, regulators, and researchers are routinely shuffling data in near real-time using web-technology. This may be specifically facilitated by:

?? mining software vendors implementing import/export with the standard formats

?? survey companies, contractors, labs, mapping agencies, state mines-departments and AGSO moving to the standards for their usual provision of data

?? the statutory data submission framework, facilitated by GGIPAC, requiring use of the standard formats for upload.

4. Project plan

4.1 General

The goal of the project is to develop an ISO compliant geology model, including

1. ISO compliant geology/mining feature catalogue
2. ISO compliant geology/mining community metadata profile
3. XML encodings: the eXploration and Mining Markup Language **XMML**
4. demonstration software and API for import/export to legacy applications
5. geology model viewers (in conjunction with project sponsors)

The approach is to synthesise a large amount of prior work in a consistent framework. As far as possible the outcomes will be compatible with existing systems developed within resources industry at the model level, and basic interoperability will be possible through transformations to and from legacy file-formats.

We do not intend to produce finished, commercial-grade software. The primary deliverables will be model specifications, and text-based encodings of these that may be used for transfer or archiving. Some demonstration and proof-of-concept software will be developed and made available under an open source license, and sample API's may result, in the form of compiled (binary) libraries. Project sponsors will have advance access to project results and all source code, and will be assisted and encouraged to begin development of implementations based on the specifications during the project. Uptake of the transfer standard in commercial software will be a criterion of the success of the project. Furthermore, while e-business applications and interfaces are beyond the scope of this project, it is likely that they would be developed on top of the resulting tools.

The project will be managed by adapting the consortium methods established by W3C and OGC. The goal of this approach is to facilitate involvement by a variety of stakeholders, giving them influence in the outcomes appropriate to their interest and contribution towards supporting the project.

4.2 Prior IP

Since the project is largely concerned with drawing existing models together within a common framework and representation, we will be referring to a large number of prior standards and specifications. The most important of these are:

Phase	Standard	Source
Specification	Spatial Schema	**ISO 19107**
	Application Schema	**ISO 19109**
	Feature Catalogue	**ISO 19110**
	Simple Features	**OGC**
	FracSIS Geometry Model	**Fractal Graphics**
	FracSIS Geology Domain Model	**Fractal Graphics**
	Geology Datamodel	**CSIRO/AMIRA**
	AMIRA P431 schema	**AMIRA**
	Statutory submission format	**GGIPAC**
	Geochem format	**WMC ExDiv**
	Geochem formats	**Genisys**
	Geophysics	**SEG**
	Well Log ML	**POSC**
	Potential-field Geophysics	**Fugro**
	Gravity GDF2	**GGIPAC/DFA**
	EM Geophysics	**CSIRO/Encom/AMIRA**
Implementation	XML encoding standard	**ISO 19118**
	GML	**OGC**
	SOTF	**OGC**
	SVG, XSLT	**W3C**
	X3D	**Web3D**
	GeoEditor	**CSIRO**
	FracViewer, FracSIS	**Fractal Graphics**
	Data Translator	**CSIRO**
Metadata	Metadata Schema	**ISO 19115**
	Statutory submission format	**GGIPAC**
	FracInfo	**Fractal Graphics**
	ANZMeta	**ANZLIC**
	XML Encoding Standard	**ISO 19118**
Web-interfaces	SVG	**W3C**
	X3D	**Web3D**
	Web Map Server Interface	**OGC**
	ASDD	**ANZLIC/AUSLIG**

4.3 Timetable & resources

4.3.1 Deliverables

Date	Activity	Who*
Ongoing	Project website Periodic progress reports – online+hardcopy Liaison: industry stakeholders Liaison: OGC/ISO	SC SC SC, KC, NJA SC
2000Q3	Road-map through ISO/OGC standards explaining how a conformant domain-specific model or profile is developed	SC, **OGC**
2000Q3-4	Conversion of FG GDM to ISO compliant geology & mining feature catalogue (UML, also in RDF/S? if OGC continue down that track) ISO compliant application schema (outline only – if needed)	SC, RW, NJA
2000Q3-4 20001Q1-2	XML Schema definition of 3D-GML, using ISO encoding rules to convert from UML. XML Schema implementation of basic Geology Domain Model - XMML	SC, **OGC**
2001Q2	Add drill-holes to XMML – from P431, GGIPAC	SC, **FG** ++
2001Q2	Add geochem to XMML – from WMC, Genisys	*Sponsors* ++
2001Q3-4	Add maps to Geol-ML: 3D -> 2D	**FG** ++
2000Q4- 2001Q4	XMML DataTranslator i/o module XSLT for 3D-GML -> SVG (2D graphics) XSLT for 3D-GML -> X3D (3D graphics) XSLT for XMML -> X3D (3D graphics)	PH, GG KC, SC, **OGC** KC, SC, **OGC** KC, SC
2001Q1-2 2001Q3-4	import/export tools for desktop packages Excel/Access/ODBC?/SIF API for XMML - *Java? Open Source? JDBC?*	*Sponsors, COM programmer* *Sponsors, Java programmer*
2001Q1 2001Q2	ISO compliant XM metadata profile (UML, RDF/S) XML Schema definition of XM-Meta	SC, **OGC, GGIPAC, AGSO**
	Web interfaces: catalog/query specification getmetadata/getdataset specification (modelled on OGC "getmap")	SC, **OGC**, **AGSO**, *Sponsors* ++

*for topics indicated **OGC** we expect to be linked to work coordinated through OGC;
- for topics indicated **FG** the current expertise and interests of Fractal Graphics mean that the work will proceed most effectively if FG takes a lead role;
- deliverables concerning *implementations* will mainly be accomplished through collaboration with sponsors.

The deliverables listed are indicative, and depend on our current understanding of the scope of the problem and tools available. These will be revised as the project progresses. Many of the implementation-based deliverables will depend on contributions of resources from sponsors. These will be expanded as more resources or sponsors become available. Sponsors will be kept fully up-to-date with progress and modifications to plans through monthly revisions to the project plan as posted on the project website.

4.3.2 Personnel

Funded directly by this project:

Simon Cox – 100% 18 months

Kim Covil – 50% 18 months

New Java/COM programmer – 100% 12 months

Contributed by project sponsors

Fractal Graphics– Nick Archibald, Robert Woodcock

CSIRO in conjunction with "Glass Earth" – Peter Hornby, Gordon German

4.3.3 Travel

Part of the project activity will be education and advocacy within our community, and advice and assistance to early implementors. Building working relationships with vendors such as Datamine, Surpac, Maptek, Micromine, Fractal Graphics, Encom, ESRI-Australia, and data providers and data managers such as AGSO, the state Mines Departments, survey companies, as well as users from the resources industry, will be a key factor in ensuring the effectiveness of the project. This will require a reasonable amount of domestic travel for liaison with sponsors and other stakeholders.

Another significant liaison task is between our community and the broader standards organisations. Effective collaboration with OGC in particular will be important to the success of this project. This will require travel to meetings internationally three or four times per year, as well as some domestic travel to work with other members of the Australian SIG.

4.3.4 Budget (provisional)

Personnel	$245 000
Travel*	$45 000
Materials/Equipment	$15 000
Total (18 months)	**$305 000**

*SC 6 interstate trips, other staff 3 interstate trips @ $2000 each == $18000
SC 6 international trips @ $4500 each == $27000

This budget is designed to allow the project to deliver initial versions of the outcomes listed in 4.3.1. Sponsorship beyond these levels, as cash or in the form of in-kind collaborative software development efforts, will permit attention to be paid to more comprehensive versions of the specifications and tools, and to address the implementation activities towards the bottom of the table.

4.4 Sponsorship

Several distinct groups of stakeholders can be identified:

1. The lead organisations and project managers: CSIRO E&M, Fractal Graphics.

2. Vendors: early involvement by software developers will be important to ensure timely uptake of the standards, with the resulting benefits in reducing overlap and duplication of effort. Access to the deliverables and consortium management must be structured to particularly encourage participation by mining-industry and associated (e.g. GIS, graphics) vendors. For example, although the final model and XMML encoding will be made freely available (as befits a “standard”) drafts and early versions will only be available to consortium members. This will enable participating software vendors to commence development of implementations well in advance of the public release of the standards, so that they can be shipping ISO compliant, web compatible products almost immediately.

3. Resource companies: these will be the direct beneficiaries of the work and should be encouraged to take an interest from an early stage, particularly in providing requirements and example datasets. However, the resource companies are expected to be less involved in detailed technical analysis, and may not want to take full membership of the consortium. A group of resource companies could be organised to sponsor the XMML project through a consortium, e.g. through AMIRA. This precedent has worked well in OGC where a group of Australian interests are represented as the “Australian Web Mapping Consortium” with a single vote on the management committee.

4. Survey companies: those involved in collecting and providing data for multiple-clients have an obvious interest in minimising additional data processing demands by using standards. Survey companies are often involved in the development of their own data-management systems and fill a role comparable to vendors so are encouraged to be pro-active in their involvement.

5. Consultants: these perform similar analyses for a variety of clients, so will be able to focus on their value-adding activities better if data-management barriers are made lower. However, consultants are expected to largely accept solutions provided by software vendors and are unlikely to make direct input into standards development.

6. Facilitators: granting agencies such as MERIWA may sponsor the project but will not require direct input into the project.

7. Regulators and statutory data suppliers: the mines departments and surveys have a particular interest in the establishment of standards for archiving datasets. The regulators often follow users and commercial data-managers in setting data standards, but have additional requirements particularly concerning metadata which need to be built in to the design phase. GGIPAC is the most obvious body to coordinate this, though individual agencies might choose to participate directly or through AMIRA.

8. Other standards bodies: liaison with OGC and ISO/TC 211 will occur through SJDC's involvement in those activities.

4.4.1 Sponsorship rates:

Full: $100k p/a cash and in-kind (minimum $60k cash) - full participation in project management

Associate: $15k p/a – access to drafts and right to comment

Facilitation: grant agencies such as MERIWA

Rates are subject to change by the unanimous agreement of full sponsors.

4.4.2 Communication and decision-making

Active consultation on technical issues, and communication of progress will be essential to the project. Six-monthly sponsors meetings will be scheduled to allow face-to-face discussion. Project sponsors will also be polled from time to time on issues where a judgment-call is required regarding priority setting, for example. Otherwise, however, the project team, led by Dr Simon Cox, will make final decisions on technical issues, informed by feedback from sponsors. Where technical and engineering decisions are involved this approach is appropriate, and follows the precedent of W3C where the all recommendations (standards) are subject to final approval by the director.

The list of existing standards (4.3.3) that will be used as inputs gives an indication of the extent of consultation that we expect to be involved. An archived email-list will be used to canvas issues and for regular notification of results, new drafts, etc. A web-based document repository will be maintained with all sponsors notified of any new postings. We expect the specification documents to go through a series of minor versions (v0.1, 0.2, … 0.9) before the first public version (1.0) is released. Each draft of the various specifications will be made available for a minimum of three weeks before being given a minor-version-number, to allow any deficiencies to be discovered and objections to be raised by project sponsors.

4.5 Software licensing/distribution:

1. All final specifications (in UML models, RDF/S schema and XML Schema for XMML) will be made freely available under a public license. The specifications will include clearly identified mandatory and optional components, and will include mechanisms to generate derived standards through restriction or extension, similar to the methods for deriving domain models from the ISO standards. A license will be used that encourages use of the standard by software developers, but does not permit arbitrary non-standard derivations to be labelled as "compliant" (c.f the Sun Microsystems Java licenses).

2. Draft specs will be available to sponsors for download from the (password protected) website. Sponsors will be notified when minor versions of the draft spec, and bug-fix releases, are posted on the website.

3. Source code for standard conversion components (XSLT scripts) will be made public, as an effective example of user-software.

4. A module for i/o of XMML will be added to the public release (binaries) of the CSIRO Data Translator.

5. Some basic i/o tools (to ODBC, Excel, SIF) will be made publicly available.

6. We anticipate that a Java API will be developed in conjunction with one or more of the project sponsors, and will endeavour to arrange for that to be released publicly.

7. The project team will collaborate with sponsoring developers and vendors in implementation activity

Appendix A: Generic data modelling and encoding tools

A.1 XML – Encoding structured data for transfer via the internet

The popularity of the web-browser page format HTML has led on to the development of the related **eXtensible Markup Language** (XML) to support the recording of richer data and document structures. XML has a small but flexible set of plain-text notations. Labels or *markup*, which identify the different parts or *elements* of a *document*, are embedded as character-strings within the text to indicate the start and end of a section. The document designer defines the element names. Elements can be nested. Data occurs as element *content* or as *attributes*. The *document* is the unit of encapsulation, often corresponding to a disk-file. A simple document might look like this:

```
<report>
       <title>Report title</title>
       <section id=”1”>
              <paragraph>
                     Some text …
              </paragraph>
              <paragraph>
                     Some more text …
              </paragraph>
              <footnote symbol=”dagger”>
                     Some extra text …
              </footnote>
       </section>
</report>
```

The syntax allows complex data which has a tree structure to be encapsulated in text streams or documents for transfer. Additional syntax allows full normalisation and recording of more complex datamodels, such as graphs, by making links between different parts of documents or between documents. The self-describing plain-text form of XML documents also makes them ideal for archival purposes; media migration is still required, but loss of data because of software incompatibilities will no longer be a problem.

Many specific XML based formats have been designed. Languages for technical subjects such as math (for both manipulation and display) and molecular chemistry were early achievements, and there is now a big push into business transaction formats. The **Resource Description Framework** (**RDF**) metadata format uses XML for its standard encoding. Of immediate interest amongst the general purpose formats are those in the graphics area. The recently finalised **Scalable Vector Graphics** (**SVG**) is a drawing format for the web. Vector graphics allows proper zooming and crisp linework and lettering. This complements the image formats like gif, jpeg and png which until now have been the only easy way to present graphics in a web browser, but often lead to fuzzy looking results. There has also been substantial interest in 3-D graphics in the web community. These have been largely focussed on gaming, which means that support for arbitrary geometries at high-precision within large universes is incomplete. For example, 3DML is very fast and compact, but describes models based around “blocks”. Industrial-strength 3D is provided by the community that developed the VRML standard for 3D graphics and behaviours, which is now collaborating with the Java-3D developers on the XML based **X3D** data format.

A.2 Modeling and manipulating data –XML schemas, UML, XSLT

XML relieves the applications developer, and to a much lesser extent the user, of some of the burdens of managing data for transfer. Basic parsers for importing XML data are being built in to many software applications packages, and it can be reliably transported using basic web protocols. However, the availability of generic XML capabilities really only moves the design and development focus to a different meta-level. The data-stream still has to conform to a model which is useful to the particular application, even if that model is encoded using XML.

Generic tools supporting the interpretation of XML are provided in two areas. First *schema-languages* are provided to describe the structure of a document. Second, a document may be transformed from one form to another, changing element-names and omitting, combining and re-ordering elements, by applying a set of rules in a generalisation of the actions of a *stylesheet*. Both of these functions are managed by formal languages which mediate the behaviour of an XML processor. For example, XML editing software should use a schema as the reference point to check that the data entered conforms to the document model. And simple XML display or browsing software might use a style-sheet to transform the XML into HTML, for which generic display software is already available.

It is the schema that defines how the underlying model of the data for the particular application is expressed in XML, and instances of data, called documents, are validated if their structure and content matches the schema. Currently there is only one *recommended* schema definition language for XML: the **DTD** (document type definition) method, inherited in slightly restricted form from SGML (the predecessor to XML). This is too weak for most purposes since, reflecting its origins in the textual world, it has almost no data typing (the only primitive element content is character strings), no method for range-restrictions, no inheritance, etc.

However, since the database community recognised the potential of XML, there have been a number of proposals for richer schema languages for XML. Current attention is focused on **XML Schema**, available as a W3C Working Draft (as of late December 1999). It provides a basic set of data TYPES (including the obvious ones for various numeric types, dates, etc) and a method for deriving more complex types by combining or restricting the built-ins and other derived types. XML types are comparable to "classes" in an OO analysis. ELEMENT definitions then specify a content model in which the primitives are these data types. A DOCUMENT is constrained by the definition of the outermost element. Overall, XML Schema appears to provide a complete set of constructs necessary for the data-part of standards DBMS structures.

The schema is itself an XML document, conventionally given the filename extension .xsd .

So, now we have a full schema language for XML, which should allow us to express a data-model. It is *not*, however, a modelling language. You need to consider Express, **Unified Modelling Language** (**UML)**, or one of the other representations used by CASE tools for that job. Typically, an XML Schema definition will be an intermediate stage between a model expressed in an abstract schema language and an instance of the data. Similarly a C++ or Java class-library may be derived from UML, and it is software applications that use the libraries which actually get to manipulate data.

The cascade doesn't stop there, either: most examples of XML Schema definitions are accompanied by a translation of these on into a DTD - necessarily incomplete, but allowing processing using the current generation of XML tools which only go as far as DTD validation.

And going the other way, you might put RDF/S above the XML Schema providing a public interface to the semantic level.

There is an important point here - a good schema language (e.g. UML) will support valid implementations in a *variety* of programming environments, (C++, Python, Java, XML, RDBMS schema languages, even Visual Basic, etc), preferably through automated translation or code-generation. Different targets will be chosen for different reasons. XML is chosen here as the target for representation of data-instances because it is a good vendor-neutral serialisation for archiving and interchange of structured data, which is also highly compatible with "internet-generation" tools.

Because of the newness of the XML Schema spec, the rules for translating from the more abstract schema languages to XML Schema are still in the process of being worked out. Nevertheless, the general patterns are straightforward and experiments are already underway for UML->XML at least (in "internet-time" the implementors seldom wait for a standard to be finalised before they get started, as long there is a significant itch to be scratched).

Finally, the transformation of a document conforming to a particular XML format (defined by a schema) to another format (not necessarily XML) can be described by a style-sheet. If the elements in the XML document do not need re-ordering, but simply converted into a regularised HTML format, then **Cascading Style Sheets** (**CSS**) provide a simple way of describing the mappings from data-oriented to presentation-oriented elements. For more complex transformations, such as from the Geographic Markup Language (GML) to a presentation format like SVG, then the transformation may be described using the **eXtensible Stylesheet Language for Tranformations** (**XSLT**). Both XSLT and CSS use text files which are then provided as additional input to a generic processing application.

Appendix B: Geospatial data

B.1 ISO geospatial standards

ISO Technical Committee 211 has been working on standards for Geographic Information/Geomatics for several years. The scope of the work is a bit of a sprawl, but effectively they are defining the standards for what information has to be handled by GIS and other spatial information systems. There is evidence of influence from earlier standards efforts, such as SDTS, as well as implementation experience. There is strong buy-in from statutory organisations in all the significant national jurisdictions.

It is a somewhat unwieldy affair with 20+ separate standards-track documents wending their way through various sub-committees, many 100+ pages, and most now in their 2nd or 3rd major draft. (Although they are not public documents, SC has access to them through OGC and through being a Standards Australia expert.) Operationally TC 211 has had a tough time as the web-revolution has really arrived during their work, and this and other technical developments have tended to overtake the work while it was still in progress. Nevertheless, what is emerging is likely to be important to us in the following ways:

?? the ISO model for representing geometry is comprehensive, and is likely to form the basis for most future commercial GIS packages

?? the ISO spatial/geometry model includes 3-D (and even 4-D)

?? the ISO metadata model will be the basis for a forthcoming revision of the ANZLIC metadata standard

?? there is a clearly laid-out mechanism for defining domain models and community profiles

domain models are formed of an application schema and feature catalogue: “features” are the primary entities in the ISO model, and each particular community will specify which feature-types are required for them (e.g. drill-hole, fault, pit) and what its attributes, behaviours, and relationships with other features are

community profiles specialise the standard parts of the schema by restricting the range of a value, making something compulsory instead of optional, and by adding additional detail, etc.

?? (as of the middle of 1999) the models are all presented as UML class-diagrams

?? (as of late 1999) the standard encoding for instances is XML

B.2 GIS application standards

The other major international initiative, is the OpenGIS Consortium (**OGC**). After concentrating on conceptual modeling for a few years OGC have now hit their straps in two ways

?? the best-known GIS vendor ESRI is now on-board

?? they're focusing on implementations, and have largely ceded the abstract spec work to ISO (but remain actively engaged in that forum as well).

The major achievements of OGC to date include (i) the **Simple Features** Specification, which is a "Lite" version of the ISO model corresponding to the data model required to support basic GIS systems; (ii) an XML encoding for Simple Features, called the **Geographic Markup Language** (**GML**); (iii) the **Web Map Server Interface** specification.

The latter was demonstrated in the highly succesful **Web Mapping Testbed** (**WMT**) exercise in 1999, when OGC incited a number of vendors to implement http based "map-servers" on top of their proprietary GIS software, so that clients can seamlessly join/overlay data coming from divergent sources over the web. So far it is accomplished primarily by layering partially transparent gifs (with some querying), but it was a very obvious proof-of-concept of the idea that if you put the data into a neutral format then it can be used more easily and widely. The parts of the data that are not images mainly use XML.

Australian organisations have been active in OGC work through the Australian Web Mapping Consortium coordinated through AUSLIG, and an Australian Special Interest Group is in formation. CSIRO is also a member of OGC in its own right.

SC has been monitoring the Open GIS activities since they emerged from the GRASS community, and has recently been contributing to the development of GML, and has also been involved in reviewing ISO/TC 211 drafts through Standards Australia.

B.3 Datamodels for Exploration and Mining

Given the existence of several data-models that were developed in and around our industry the simplest way forward might appear to be to simply generate XML versions of these. The existing models might be expressed rigorously in UML or Express or similar, and then (asap) run through CASE software to generate XML Schema. Job done, as far as XML coding of our data is concerned. The outcome is limited interoperability between those groups who find each of the data-models applicable and are prepared to develop applications based on them.

But would this be the smartest thing to do? That depends on whether we are confident that the modelling done to date is adequate, and whether there are more recent developments that can help us. The answer to these questions depends on a number of assessments:

?? Do the existing models cover our domain of interest?
- do they cover all of the necessary maps, boreholes, estimation, geophysics, assays, simulation configuration, functions? etc

?? Can they be used together with each other?
- is the borehole model compatible with the GIS model?, etc

?? Is there any useful software around to process data encoded using any of them?
- is any company database configured using P431?
- do any GIS systems export GGIPAC format digital reports? etc

?? Do they provide a solid basis for the newer tasks that we would like to accomplish in the online world?
- submission of statutory reports online;
- automatic loading of statutory data into smart databases enabling easy retrieval;

- input into numerical simulation software;
- exchange of requests and data between companies and labs, etc

and perhaps most importantly:

?? Is there an easier/cheaper/faster way? (preferably leveraging off generic or commodity technology)

The prime candidate is customisation of the ISO model and OGC implementation.

B.4 Adapting the ISO Feature model for an application domain

The ISO/OGC conceptual model is **feature** based. A feature has a set of attributes or **properties**, including metadata, each of which are modeled separately. A **feature-type** is simply a definition of the set of properties that are required for a particular feature, some of which may be spatial. Properties involving geometry and location build on a standard **spatial model**, while other properties will use simple and complex data-types as required. Supporting and contextual information or **metadata** is based on a common core with domain sensitive restrictions and extensions.

The procedure for developing a **domain model** (e.g. for geology and mining) in the ISO context is not trivial. It requires a **feature-catalogue** enumerating the features required for operating in the application domain, some relationships between features, and operators on them. A comprehensive model may also include an **application schema**, though it is not clear if this is a requirement. A roadmap of the ISO standards and documentation and the relationship with the OGC implementations will be an early task of the project.

B.4.1 Spatial information

The ISO Spatial Schema is used as the basis for geometry and location properties in any domain-model. The schema is described in ISO 19107 and implements the usual point-line-arc-ring-polygon hierarchy. The schema definition is “abstract” and presented as a set of UML class-diagrams with supporting tables. OGC has defined a Simple Features profile corresponding to a simple 2D “coverage” as known from the GIS world. The most definitive version of this is in the form of an SQL interface, and following the decisions by ISO and OGC to standardise on XML encoding, Geography Markup Language (GML) has recently been proposed. In principle GML allows higher-dimensional spaces to be addressed.

Although there have been no implementations, the ISO spatial model appears to be complete in 3D. While we suspect that performance on a data structure matching the ISO model is unlikely to be very efficient or scalable, we will use the ISO model *as-is* for the purposes of a common model and transfer standard. Because of the successful existing implementations, the FracSIS and CSIRO spatial models might be useful as a benchmark for an audit of the ISO/OGC models, and the first test will be to map the FracSIS and CSIRO spatial classes to classes in the ISO schema. If we discover required 3D spatial classes missing from the ISO schema in the first instance we will use a place-holder in our implementation and will collaborate with OGC, ISO (to the extent possible), and in particular with the project sponsors on the most effective solution. We do not expect that any significant development of the spatial model will be necessary.

Methods for conversion between 3-D and 2-D representations of features are needed to support viewing in commonly available 2-D applications software.

B.4.2 Geology and mining domain information

Existing industry models provide a critical preview of the requirements, though none of the publicly available models or formats use a fully modern (oo) modeling method, and many do not even use a formal schema language so consistency and completeness are not ensured. Nevertheless, reference to the existing models will allow rapid early progress, particularly where they partition the information in a way that is compatible with the ISO Feature model. One of the outcomes of the project will be to re-cast selected models in ISO compatible form, and it is expected that re-usable components or classes will be uncovered in this way.

However, there are significant areas in exploration and mining data that are not managed by existing public models for exploration and mining. This particularly affects 3-D objects other than drill-holes, and general interpreted geology. Adaptation of models to include additional features will be required in order to take full advantage of the support offered by standards.

The FracSIS geology domain model is expected to provide the key input to the development of a consistent and comprehensive domain-model and encodings for it.

B.4.3 Metadata

Discovery of particular data from within a repository system requires descriptive information for indexing. Appropriate use of a particular dataset requires information concerning its origin and lineage, reference-frame, quality, any legal restrictions, etc. This contextual and supporting information, which largely lies outside the dataset itself, is referred to as metadata. Descriptive metadata is particularly important to support catalogue services, and has been the driving force behind community standards such as ANZLIC metadata, which is the basis of the Australian Spatial Data Directory (ASDD). Detailed specifications of data collection methods and file structure, necessary for re-use of archived data, is referred to as metadata in regulatory frameworks such as the statutory reporting system subject (in part) to recommendations from GGIPAC.

The ISO standard includes a rich metadata model, covering a basic core description expected to be common across all spatial datasets, and with explicit guidelines on how to customise this to produce a **profile** for a particular community or application domain. A revision of the ANZLIC standard following the ISO model is already planned. We will use the existing standards, such as ANZMETA and GGIPAC, as well as various configuration and directory files derived from the FracSIS and CSIRO implementations, as the basis for an ISO compliant exploration and mining metadata profile.

SC has also been strongly involved with standards activities in metadata, and is an active contributor to the Dublin Core Metadata Initiative where he is a member the Advisory Committee, and to the Australian Government Locator Service where he is an invited expert on the Working Group. He wrote the standard mapping between ANZMETA and AGLS.

Appendix C: Example applications of XMML

C.1 Geology modeling using multi-client potential-field data

While constructing a 3-D geology model of a prospect, it may be determined part way through a screen-session that some geophysics will provide a useful additional constraint. It is suspected that data to support the analysis is already available from a non-restricted source, such as a multi-client survey archived by a contractor, or from a statutory provider such as AGSO or a state survey, so a query is submitted through a web-based form interface to a multi-source catalogue. Having discovered the data, this is downloaded from the host database and imported directly as an object into the modelling environment, mid-session. Any required payment is authorised through an e-commerce wrapper around the technical data.

Most components of this scenario either already exist or have been demonstrated for 2-D data.

Clearinghouse, based on the library-oriented distributed query system z39.50, has been run by USGS for the past three years, exposing the catalogues of nearly 100 US based geospatial data custodians. An Australian version of this, the Australian Spatial Data Directory (ASDD) has been online for almost a year, providing an index of holdings by agencies from statutory lands and survey organisations in all the States and the Commonwealth. The index is built from metadata conforming to the FGDC and ANZMETA metadata standards, respectively, though the actual data is not available for download through the same interface.

Meanwhile, the Web Mapping Testbed was demonstrated by OGC in late 1999, through which data served by several Web Map Servers, each running different GIS systems, could be combined to-order in a single web-browser based viewing window, in real-time. The Catalogue Specification, which is under development by OGC, supports an enriched version of the Clearinghouse or Spatial Data Infrastructure, fully integrated with the data encapsulation and data delivery mechanism.

Since it will fully compliant with the necessary ISO and OGC standards, XMML will support both catalogues services and data-download implemented according to the OGC specifications, for 3-D exploration data. This will enable the scenario to be implemented, by cooperation of the various data custodians, with a minimum additional effort.

N.B. A map-server conforming the OGC's Web Map Server Interface has been implemented by SC and KC on a web-server at the CSIRO Nedlands site, serving data from an underlying GRASS database.

C.2 Statutory Reporting

Most of the state Mines Departments in Australia hold two major sets of exploration related data: (i) the data collected in geological and geophysical surveys, collated and interpreted as part of the activities of staff under direct supervision of the department; (ii) data submitted by tenement holders as required by the Mines Acts. The survey data are typically stored in formal database and GIS systems and are the basis of products such as maps. These are the main source of information provided to clients in support of exploration activities. The submitted data are normally archived in a different system, and become available to clients in due course through Open File when tenements are relinquished.

Even though the Open File reports include very valuable data, particularly large numbers of chemical analyses, various difficulties in managing the data means that this value tends not to be fully realised for subsequent investigations. The major impediments to use of Open File data are:

(i) they are not stored digitally, though *indexes* of them are now generally available digitally

(ii) where digital forms are used, they are frequently only “facsimile” forms, such as PDF, which makes re-use of any data in a processing or analysis package difficult

(iii) where structured forms have been introduced, such as recommended in the 1999 GGIPAC guidelines, these do not correspond with i/o formats from any widely used exploration and mining industry applications software, so specialised reformatting or parsers are still required.

Ideally, submitted data should be loaded directly into a structured database when submitted, and then delivered from this for re-use, in response to a structured query, which specifies the theme, location and other constraints on the data of interest.

XMML contributes an important component towards accomplishing this. Tenement-holders would prepare versions of their data in XMML and submit this by uploading it with a transaction initiated through a web form. Suitable authentication procedures would be used, in accordance with general e-business technology. Software on the server-side (i.e. the Mines Department) would verify that the data package conformed to the XMML Schema, and that all components and metadata were complete according to the local regulations. If not, the submission would be rejected in real-time. If accepted, the data would load direct into a corporate database, with the appropriate tenement-code attached. The data would become visible through an index automatically, according to the regulations pertaining to the tenement history. When requested subsequently, data items required would be delivered, online, in XMML for direct import into the clients processing software.

C.3 Drill-hole assays

One aspect of the development of XMML will be to incorporate a model of chemical-analysis data.

A single analysis may be delivered as a message from a laboratory to customer as an XML “document”, perhaps as an email message. However, since an XML document is defined simply as the outermost XML “element” in the instance, this element may in turn be stripped from the message and incorporated into a richer document unaltered, or direct into a database where it is associated with the relevant sample through the id or key.

This mode of operation is fully compatible with the generic model. The ISO feature-model is based on real-world objects, which have application-specific names such as fault, pit, bench, or drill-hole, as determined by the feature-catalogue. The parameters that give the objects specific meaning, such as shape, location, colour, density, field-strength, etc. are attached to the “feature” as properties. The properties can have complex types, i.e. they are not simply character-strings, numbers or even vectors. So in this model, a chemical-analysis may be treated simply as a property attached to a sample.

Appendix D: Reference websites

AMIRA	http://www.amira.com.au/
ANZLIC	http://www.anzlic.org.au/metaelem.htm Reference definition of ANZMETA metadata standard
ASDD	http://www.environment.gov.au/net/asdd/ Australian Spatial Data Directory – clearinghouse for geospatial data from Australian statutory providers
CSIRO	http://www.ned.dem.csiro.au/research/visualisation/DMGE/ Data Model for Geology
DFA	http://www.dfa.com.au/gravity.html
ENCOM	http://www.encom.com.au/
Fractal Graphics	http://www.fractalgraphics.com.au/
Fugro Survey	http://www.fugro.com.au/
Genisys	http://www.genconsult.com.au/
GGIPAC	http://www.dme.wa.gov.au/statdata/index.html#guide Requirements for Submission of Data in Digital Format
ISO	http://www.statkart.no/isotc211/
OGC	http://www.opengis.org/ Specificiations for Simple Features, GML, WMS, Catalog Services
POSC	http://www.posc.org/ebiz/WellLogML/ WellLogML DTD
SEG	http://seg.org/publications/tech-stand/index_body.html SEG Technical Standards
W3C	http://w3.org/TR/ Reference specifications for XML, XSLT, XML Schema, SVG, RDF, RDF/S
Web3D	http://www.web3d.org/technicalinfo/specifications/vrml97/index.htm http://www.web3d.org/x3d.html VRML, X3D