Developing Web-Service Oriented Geoscientific Information Systems

Alistair Ritchie, Bruce Simons, Gavin Stilgoe, Deepthi Fernando
Minerals and Petroleum Division, Department of Primary Industries,Victoria, Australia
Corresponding author: Alistair.Ritchie@dpi.vic.gov.au

ABSTRACT : Scientific data collected by GeoScience Victoria, Minerals and Petroleum Division, were previously held in databases engineered around purpose-built software packages. A wealth of information was stored in free-text legend fields or geological reports. Analysis of these data using other software was difficult and the approach led to inconsistent and contradictory systems and content.

GeoScience Victoria has redesigned its geological information systems to resolve these problems. A data model was constructed around geological concepts. The model made no distinction between spatial and aspatial data-sets and ignored constraints imposed by particular vendors or their tools.

A single logical data store was built using Oracle and ESRI ArcSDE databases integrated with a brokering service. This service and an overarching layer of data management services were constructed using the Java platform and running under IBM's Websphere application server. Data are delivered using web services, including web feature services.

By using data models that describe true geological entities and concepts the information stored can be more readily used by a variety of users and applications for a variety of purposes. Adopting services oriented architecture allowed the reuse of technology in a number of contexts thereby reducing short-term development, and long-term maintenance overheads.

KEYWORDS : geology, geoscience, GML, GeoSciML, GIS, web service, web feature service, XML, Victoria, Australia

1. Introduction

Scientific data collected by GeoScience Victoria were previously held in databases engineered around purpose-built software packages. These focused on providing the content for geological maps and their accompanying legends. A wealth of information was stored in free-text fields or geological reports. Analysis of these data using other software was difficult and the approach led to inconsistent and contradictory systems and content. To address this GeoScience Victoria is undertaking a long-term project, GeoDATA.Vic, to redesign its information systems. The project is being implemented in stages and this paper summarises the results of stage one.

The project has created a data store based on geological principles, with an overarching application server layer to enable the maintenance of the data. Data are delivered to end users and client systems through a variety of World Wide Web Consortium and Open Geospatial Consortium (OGC) compliant web services. Figure 1 summarises the system architecture.

The services-oriented approach allows data storage to be engineered efficiently, with the services layer handling the manipulation of the data for presentation and analysis. As the construction of system components is modular, the impact of future changes in client software or vendors is minimised.

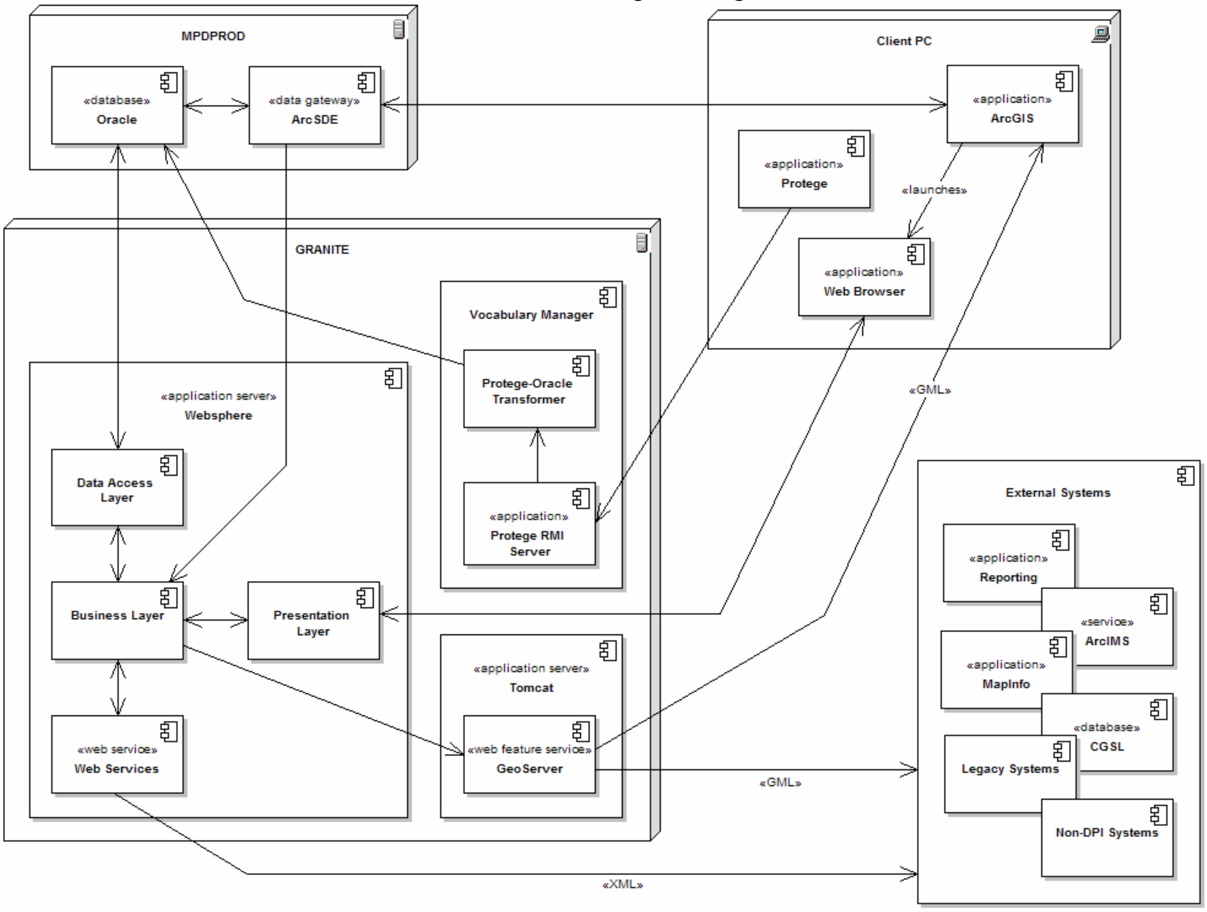

Fig. 1. Simplified GeoDATA.Vic project Component model. The MPDPROD server is the data store, the GRANITE server hosts the application server layer and the client PC shows clients developed for this project. External Systems is an incomplete grouping of non-project clients.

2. The System

2.1. Data storage

The North American Data Model Conceptual Design (NADM-C1.0 – North American Geologic Map Data Model Steering Committee, 2004) provided the foundation for the design of the data storage systems. NADM-C1.0 is not influenced by software vendor data models but instead focuses purely on geological concepts. The earth sciences Geography Mark-up Language (GML) exchange model, GeoSciML (Commission for the Management and Application of Geoscience Information, 2005), and GeoScience Victoria's business requirements further constrained the development of the data store.

The data model used makes no distinction between spatial and aspatial data-sets. It reflects real world information and has not been generalised to suit specific applications or subsequent use. A modular approach has been used to ensure extensibility of the data store during development.

Controlled vocabularies are used, wherever possible, to specify property values, minimising free text fields. The Protégé vocabulary manager is used to manage these data.

The majority of the data are stored in an Oracle database. The physical model used is somewhat abstract, allowing the same database table to be used to store a number of geological concepts.

An ESRI ArcSDE managed Oracle database stores the geometry data and their dependent attributes. The structure of these data is very simple, the only real complexity being a

structure imposed by ESRI to handle the storage of different geometry types and the topology rules used to assist in their maintenance.
The two databases are integrated in the services layer, with some integration components in the databases themselves.

2.2. Application server

The application server manages access to data stores, data transactions and the integration of systems and databases. To implement this we used IBM's Websphere application server and the Java platform to develop the software.
Conceptually, the application server can be divided into three layers: the data access layer, the business layer, and the presentation layer (Figure 1). These overlap, often using the same components.
The data access layer contains the java objects that interact with the data-stores. These objects are structured to mirror the underlying database tables and features. The data layer accesses Oracle using a Java database connection. Spatial data are compiled using requests made to the ArcSDE Application Programmable Interface (API).
There is a one to one relationship between the Oracle database tables and data access layer objects. Geometry objects acquired from ArcSDE are simplified to a standard geometry model. Under the ESRI data model, ArcSDE stores data separately depending on geometry type and topological considerations. Typically, at least three ArcSDE feature classes (one each for points, lines and polygons) are combined into a data access layer geometry object.
The business layer applies the business logic to the data. The java objects compiled here have a strong resemblance to the logical data model. Business rules are applied and maintained using Enterprise Java Beans. This layer provides data to the presentation layer and web services.
The presentation layer receives the business layer objects as extensible mark-up language (XML) documents, then uses extensible stylesheet language templates to render the data on a client's web browser. This method allows a change of platforms, for example from a browser on a PC to one on a small portable device, by adding a new template to suit the new platform.

2.3. Web services

Web services are portals that deliver structured raw data, formatted as XML, to clients without the presentation components. Web feature services (web services extended to handle spatial data and queries) are used to deliver spatial data.
Data are delivered according to a set of schema definitions based on GeoScience Victoria's business model or XML models defined elsewhere. These definitions may be for a particular purpose and the same data values can therefore turn up in different contexts. GeoSciML will be the primary schema definition used by the web feature service once it reaches a suitable level of maturity. As the current generation of GIS applications cannot easily handle complex schema like GeoSciML, a simple GML schema has also been defined. All web feature service schema must be compliant with OGC standards.

2.4. Clients

The GeoDATA.Vic project has developed a user interface to the presentation layer and integrated ESRI's ArcGIS. Both are run on a networked PC and are linked to allow the results of queries in one application to be displayed in the other.
The user interface is run on a web browser. It allows complex queries to be executed and is the primary means of viewing and editing all data not managed by ArcSDE.
ArcGIS allows use of spatial queries and analytical tools. The data is supplied through a direct connection to ArcSDE or the web feature services. The direct connection allows the editing of the purely spatial data using the ArcMap application. The web feature services

bring together the data from both Oracle and ArcSDE, creating complex objects that can be analysed or mapped.

In addition, a number of external (non-system) clients have been identified. These include GIS clients such as MapInfo, other databases within the organisation, legacy systems and external clients (such as other government departments or mining companies). Data are provided to these clients through the web services.

The benefits of the service oriented approach are exemplified by the development of a controlled vocabulary manager. GeoDATA.Vic depends on well structured, controlled vocabularies and a system was needed to maintain these. GeoScience Victoria adopted Protégé - a free, open source ontology editor - on merit, without concerns about significant changes to the broader set of systems.

Protégé was readily added to the system (Figure 1). The Protégé application is installed on client PCs and accesses a central XML-based data-store through a Java Remote Method Invocation server. A Protégé-Oracle transformer accesses the data store - through the Protégé API - and transfers the data into tables in Oracle. These data are then used by the GeoScience Victoria classification system.

By adopting this approach GeoScience Victoria did not have to develop a conceptually and technically complex vocabulary manager internally. Only the Protégé-Oracle transformer had to be constructed.

3. Summary and conclusions

The modular approach to development has enabled the use of reusable patterns throughout the system. Development overheads have therefore been reduced. In addition, changes to a system component will typically not be apparent to an end user or client system. Conversely, changing client applications, for example using software from a different vendor, should not impact upon the service or data layers of the system.

Adopting a services oriented architecture has allowed GeoScience Victoria to store data structured in a geologically meaningful way. The requirements of users and client applications can be met by manipulating the data in the services layer. When delivered in a consistent format the same data can therefore be used in many contexts, and in many ways.

Acknowledgments : The authors would like to thank Alan Willocks and Manoshan Domingo for their contribution to the preparation of this paper.

REFERENCES

Commission for the Management and Application of Geoscience Information, 2005. Conceptual Model/Interchange. http://www.bgs.ac.uk/cgi_web/tech_collaboration/data_model/task_groups/concept/concept.html

North American Geologic Map Data Model Steering Committee, 2004. NADM Conceptual Model 1.0—A conceptual model for geologic map information. *U.S. Geological Survey Open-File Report 2004-1334* 58p.