

The Geoscientific Data Warehouse — delivering NSW geoscience data

Abstract

The Geoscientific Data Warehouse (GDW) was first released in 2012. It is the culmination of the COGENT II project, which aimed to secure and centralise the vast repositories of geoscientific data held by the Mineral Resources branch of NSW Trade & Investment, and provide a platform that would make the data easily accessible to industry, other government departments and the public. It is funded under the state government's New Frontiers initiative to attract mineral and petroleum exploration investment in New South Wales, and inform land use decision making.

This paper aims to document methodologies and technologies deployed in the GDW, present case studies to highlight the ease of access to geoscientific data, and promote the use of the GDW.

The GDW has been developed in-house by the Geological Survey of New South Wales using both open-source and commercial off-the-shelf-products. A strategy of minimising the use of coded software has reduced the cost of development and ongoing maintenance and provided the Mineral Resources branch with a flexible platform that can be expanded and improved upon into the future without major expense.

The main design features are the geoscientific 'single point of truth' database, nightly replication to an internal and internet-facing data warehouse, the display of data in Google Earth, simple searches on all datasets, and single-click downloads of data in the form of shape files, MapInfo files and CSV files.

Four case studies show how the GDW facilitates access to data, how the Google Earth or geological map data display enhances the physical context of the data and how the GDW enables the download of surface geology maps and geophysics imagery onto mobile phones and tablets.

Keywords: geoscience, data warehouse, spatial, web, geology, geophysics

AUTHORS

D.B. Collins, P.J. Gilmore, J.E. Greenfield and T. R. Barlin

Geological Survey of New South Wales,
516 High Street, Maitland, NSW 2320

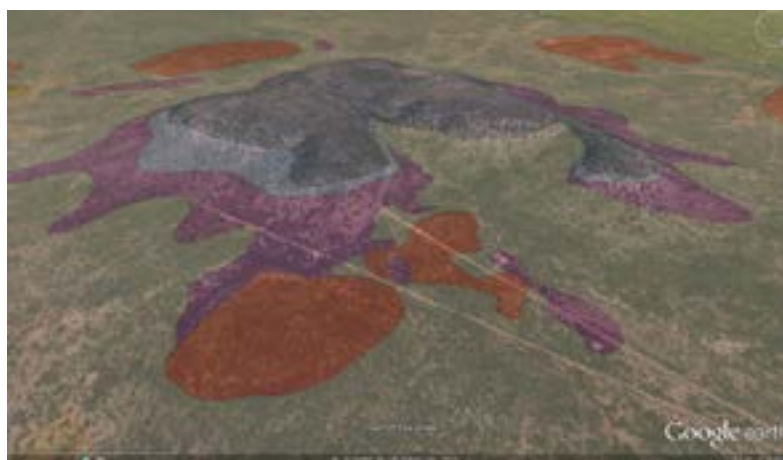
© State of New South Wales through Department of Trade and Investment, Regional Infrastructure and Services 2013.

Papers in Quarterly Notes are subject to external review. External reviewer for this issue was Laszlo Katona. His assistance is appreciated.

Quarterly Notes is published to give wide circulation to results of studies in the Geological Survey of New South Wales. Papers are also welcome that arise from team studies with external researchers.

Contact: john.greenfield@industry.nsw.gov.au

ISSN 0155-3410



Also in this issue (p. 23)

The Geoscience Product Catalogue

— a new way of finding GSNSW products online

CONTENTS

THE GEOSCIENTIFIC DATA WAREHOUSE

— delivering NSW geoscience data

Abstract	1
Introduction	2
The GDW	2
Broad design strategies	3
Case Study 1: Drillhole data	6
Major design features	7
Case Study 2: Structural readings	10
Case Study 3: Radiogenic isotope data	14
Internal system tools	15
Case Study 4: Maps for phones and tablets	18
Acknowledgements	18
References	18
Appendix 1	19
Appendix 2	21
Appendix 3	22

THE GEOSCIENCE PRODUCT CATALOGUE

— a new way of finding GSNSW products online

Abstract	23
Introduction	23
Business requirements	24
Conceptual design	25
Technical design	26
Search methodologies	26
Previewing the product	27
Acknowledgements	27

*Production co-ordination
and general editing :*

*Simone Meakin
and Geneve Cox*

Geospatial information:

*Cheryl Hormann
and David Collins*

Layout:

Nicole Edwards

Cover image: Geological map draped over Google Earth image of Mount Oxley.



**Trade &
Investment**
Resources & Energy

Disclaimer

The information (and links) contained in this publication is based on knowledge and understanding at the time of writing (June, 2013). However, because of advances in knowledge, users are reminded of the need to ensure that the information upon which they rely is up to date and to check the currency of the information with the appropriate officer of the Department of Trade and Investment, Regional Infrastructure and Services or the user's independent adviser. The product trade names in this publication are supplied on the understanding that no preference between equivalent products is intended and that the inclusion of a product name does not imply endorsement by the department over any equivalent product from another manufacturer.

Copyright

© State of New South Wales through Department of Trade and Investment, Regional Infrastructure and Services 2013. You may copy, distribute and otherwise freely deal with this publication for any purpose, provided that you attribute the Department of Trade and Investment, Regional Infrastructure and Services as the owner.

Introduction

The Geological Survey of New South Wales (GSNSW), and coal and petroleum sections of Mineral Resources branch (in the Resources & Energy division of NSW Trade & Investment) have vast repositories of data which are the legacy of decades of fieldwork, research and corporate archiving. The COGENT I project (1997–1999) was initiated as part of the state government's Information Systems Strategy, which recognised the need to provide electronic services to the community. Data was secured centrally in a corporate Oracle database, with a web-based data entry and querying system for internal department use. The COGENT II project, commenced in 2008, aimed to provide a platform which would make validated, non-confidential, corporate, geoscientific data easily accessible to industry, other government departments and the public. The Geoscientific Data Warehouse (GDW) was released on the internet in 2012 and is the culmination of the COGENT II project. It represents the first step in fulfilling the vision of providing on-demand access to geoscientific data from a single spatial-based interface.

COGENT II and the GDW are funded by the New Frontiers initiative, which continues the state government's program of pre-competitive geophysical surveys, data compilation and delivery, frontier mapping and interpretation. The GDW facilitates access to geoscientific information, helping to attract investment in mineral and petroleum exploration and inform land use decision making.

A glossary of terms and acronyms used in this paper is given in Appendix 1, where care has been taken to use [™] and [®] symbols to provide proper attribution to trade marks. These symbols are generally not shown elsewhere in this article (unless not in Appendix 1) and any mistakes are not intentional. All trade marks have only been used by the authors under terms of fair use.

The GDW

The GDW uses Google Earth to enrich the functionality and experience of discovering geoscientific data in NSW. Integrating the 3D and layer transparency features of Google Earth with geo-referenced field photographs provides a 'virtual field' experience for the user. Please refer to Appendix 2 for the URL of the GDW.

The geoscientific data held in the GDW is mainly point data, such as field observations, structural readings, geochronology, geochemical assays and drillhole locations. The data is displayed using a Google Earth interface and are listed as summary

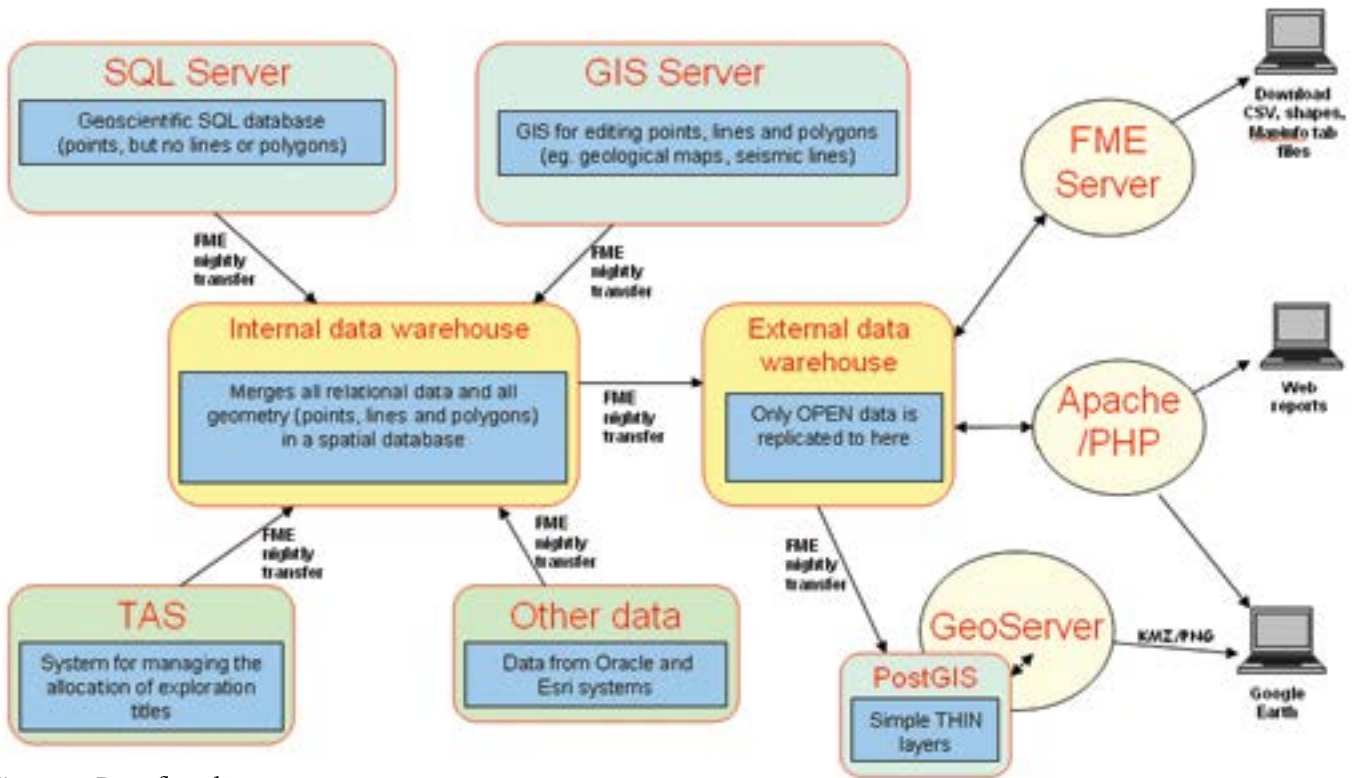


Figure 1. Data flow diagram.

reports. The Google Earth interface also has layers for geological and metallogenic maps, geophysical imagery, current mineral, coal and petroleum titles, and title applications. The layers can be draped over the topography in Google Earth. Data can overlay exploration titles which display pop-up information boxes with links to reports in DIGS. The GDW also contains a link to the new online facility, the Geoscience Product Catalogue (GPC). Please refer to Appendix 3 for a detailed list of current and planned datasets.

Comprehensive text-based search facilities bypass the Google Earth interface and generate detailed lists of data descriptions including references to reports and images stored in the DIGS database. These are complemented by a download facility that will support downloads in multiple formats including shape files, MapInfo files and CSV files.

A major feature of the GDW is the facility to download a statewide geological map and geophysical imagery to smartphones and other mobile devices. This is a one-off file download via an internet connection. The feature then does not require an internet connection to display the data, which means the maps are available in areas with no internet reception. GPS technology embedded in smartphones allows the user to instantly view maps or imagery at their location. This has proven to be a useful tool for geologists and others in the field. The list of maps available for download will continue to grow.

GDW also incorporates an advanced text and spatial search of DIGS. The spatial search of DIGS produces a list of reports in the selected area. Clicking on a title in the report links directly to the actual report in DIGS.

Feedback about the GDW is encouraged and a feedback form is available for comments and suggestions for improving the site. A list of data custodians in the metadata file for each dataset provides contact details for enquiries about the data.

The data warehouse is a data replication system. Data is replicated nightly from various sources, including the geoscientific database and the titles administration system. Figure 1 is a simple schematic of the system.

Broad design strategies

Ownership

Data custodians, who in the case of the GDW are geologists and geophysicists, need to understand the GDW systems and data structures. The systems should empower these scientists, not place a technological barrier between them and the data. If the database systems and data structures are understood and owned by the geoscientists, then they will be sustainable into the future. Part of this process is training staff in database design and administration and giving them sufficient access privileges. A significant strategy has been to employ geoscientists with IT skills or aptitude in design and development.

Minimising the generation of code

Coded systems have proven to be expensive to develop and maintain in the past. They generally require external resources and high management overhead to develop, fix bugs or even upgrade. In response to these types of issues arising from previous projects, the COGENT II project team decided to use off-the-shelf software where practical.

Geobank was purchased for entering and managing geoscientific data on a SQL Server. This provided several benefits, including a starting data model for drillhole and geochemical sample data, the ability to create data entry screens fairly quickly using the skills of internal staff, and a good import utility for uploading data from spreadsheets into SQL Server. Having configurable data entry systems contrasts starkly with previous systems such as MetIndEx and COGENT I, where data was entered using web pages built on thousands of lines of Java code. The new simplicity enabled the addition of datasets to Geobank in medium-sized projects, rather than requiring major organisation and expense. Being confined to the limitations of the Geobank data entry system (or any other off-the-shelf system) does sometimes mean that the exact functionalities of a previous coded system cannot be duplicated. However, this is overcome by redesigning the processes to suit the limitations of the software when required.

Software research

Building the GDW required a significant number of software systems to be established. The project team reviewed all software options thoroughly and cautiously against a set of criteria and constructed trial systems. Downloading and trialling of the software produced results that sometimes differed greatly from initial impressions gained from simply reading the literature. The experience gained by setting up a trial allowed the project team to analyse literature more critically. This process was assisted by the fact that good software vendors and open source communities generally make it relatively easy to install software for trial purposes and often include good example data. User-acceptance testing was carried out on the preferred software before it was purchased.

The project team used an iterative approach to development. This recognises that sometimes when development proceeds, information comes to hand that was not available at the planning stage and better ways of doing things come to light. Having the flexibility to change the plan as development proceeded allowed the project team to make ongoing improvements to the design of the system.

Choosing software

The project team used the following checklist throughout the project.

Is the software solving problems and fulfilling the project's requirements?

- What are the requirements? (a system might be good at what it does, but it may lack relevance to the project)

Is the product being actively developed?

- Is the product keeping pace with new technologies of the time?
- Is the vendor focussed on this particular product?
- Are there user forums?

Is the product taking the project towards standards?

- Is the vendor resisting standards and pushing proprietary formats for its own benefit?

Will the product empower staff?

- Will it give staff independence or just build dependence on the vendor?
- Can staff develop the skills to use and develop the product?

Once implemented, can staff maintain it?

- Will consultants be needed for even minor changes?
- Does the annual maintenance bill stifle systems development?

Open source vs closed source software

The project team took a pragmatic approach to the consideration of open source software. If the software was good quality, being actively developed and was well-supported then it was considered as an option for the project. In some instances commercial providers advertise their product as being open source for promotional purposes. According to the Open Source Initiative 'open source' means more than access to the source code. The distribution terms of the software ideally should comply with ten criteria that are listed by the Open Source Initiative (<http://opensource.org/osd-annotated>). If the source code is made available to the public but ongoing development is effectively closed or controlled solely by commercial considerations, then the benefits are not the same as those from genuine open source software.

The project team settled on a combination of open source and closed source software as listed below.

Open source software

GeoServer	for delivering spatial layers as KMZ to Google Earth
PostGIS	data layer store for GeoServer
Apache™ HTTP Server	web servers
PHP	language for creating active server pages
CodeIgniter®	application framework for structuring and modularising PHP further
Python	language for processing photograph metadata – using the pyexiv2 module
Eclipse™	development environment for PHP, Python, HTML, JavaScript, CSS
Aptana® Studio	user-friendly module for Eclipse™
Dokuwiki	for documenting software configurations and system designs
Notepad++	for editing various format files including XML
GIMP	for editing images, screen shots, photographs (similar to Photoshop)

Commercial or closed source software

Google Earth™	for displaying datasets over satellite imagery in a semi-3D environment
Geobank®	for data entry and uploading of CSV data into SQL Server®
SQL Server®	for storing entered data and replicated data
Citrix server (XenDesktop®)	for running the Geobank client
Windows Server®	for running Apache™/PHP, FME® Desktop, FME® Server and GeoServer/PostGIS
FME® Desktop	for transforming and replicating data from Geobank® and other sources to data warehouses
FME® Server	for the data download facility
ModelRight™	for modelling relational databases
dbForge Query Builder	for internal ad hoc queries of the geoscientific data warehouses

Case study 1

Drillhole data

The GDW provides access to non-confidential stratigraphic, coal, mineral and petroleum drillhole information across NSW. Currently over 56 000 coal, 44 000 mineral and 850 petroleum drillholes, with over 2 500 000 assay analyses from drilling are stored in Geobank (see Appendix 3).

The first priority of COGENT II is to store location information for drillholes. Data captured includes coordinates, title, company, date, purpose, end-depth and GS report number (a reference for further information via DIGS). The next priority is to capture down-hole information such as assay, lithology, geophysical and survey data. To date, the focus for capture of this data for minerals drilling has been the Cobar Peneplain Bioregion, to assist mineral explorers and land use planning.

In the GDW, drillhole locations may be viewed by type (coal, mineral, petroleum); whether they are stored in department core facilities (WB Clarke Geoscience Centre, Londonderry and EC Andrews Core Facility, Broken Hill); whether they have been hyperspectrally scanned; or whether they have associated lithology data. As with other GDW datasets, the data can be queried, viewed in Google Earth, and downloaded into GIS or database software. All data is stored in tables in Geobank that allow migration of data from both historic sources and the current data template. Data is replicated nightly from Geobank to the GDW.

The availability of drillhole data will allow exploration companies to easily access existing drillhole and associated assay and lithology information for their area of interest. Similarly, it allows departmental staff access to sub-surface geological information across NSW.

Department drilling

Drilling by the department has been undertaken for coal resource evaluation, petroleum wells and regional stratigraphic programs. For example, a regional drilling program was conducted under the Discovery 2000 initiative to assess the petroleum potential of the Darling Basin. Also, 138 shallow air core holes were drilled east of Cobar in the late 1990s as part of the Exploration NSW initiative to improve geological understanding of the area.

Exploration drilling

Reporting to the department of exploration drilling in coal, mineral and petroleum title areas is mandatory. Since the late 1990s, drilling reports have been submitted to the department in digital form using data templates with national standards consistent with the 'Australian Requirements for the submission of Digital Exploration Data'. This digital data is now routinely uploaded to Geobank, with back capture of digital exploration reporting data completed back to 2000. In addition, under the New Frontiers initiative, drilling information has been extracted from pre-digital submissions, which were scanned from hard copies and entered into DIGS, and entering this data into Geobank.



Exploration company RAB drilling in the Koonenberry Belt, western NSW. (Photographer P. Gilmore).

Major design features

Single point of truth (SPOT)

At the start of the COGENT II project, geoscientific data was held in a variety of locations including spreadsheets (.xls files) and MS Access databases (.mdb files). Various versions of these files were on the file server and it was not possible for any one person to know where the most recent data was held. The project team decided to store all suitable geoscientific data in a single SQL Server database called GBIS. This database is now regarded as the ‘master’ or ‘single point of truth’ location for the datasets. This is where data is modified, updated or added, because all data flows out from this point.

Apart from being the single point of truth for data, the GBIS database is also the master location for lookup lists. This includes the geological time scale, stratigraphic units and over 100 other lookup lists.

Data warehousing

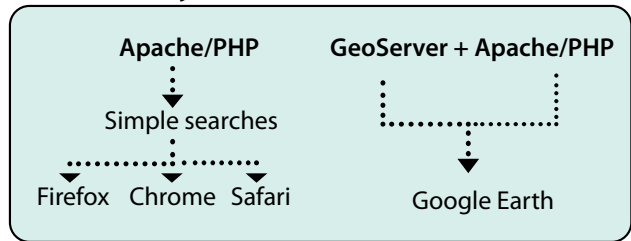
One downside of storing data in a relational database is that the number of tables (including lookup lists) is large and the relationships between the tables are complex. This is necessary as data has to be stored in a ‘normalised’ form so that no information is duplicated. To make the data more accessible, the data is transformed and replicated nightly into a data warehouse using FME spatial ETL (extract–transform–load) software. During this process, all codes are replaced with their descriptive names extracted from the appropriate lookup list. This means no lookup lists need to be replicated to the data warehouse, more than halving the number of tables that need to be transferred to the data warehouse. In this way, the table structure of the data warehouse is ‘de-normalised’ making the querying of the data much simpler and faster. In addition, all grid coordinates are converted to GDA94 latitudes and longitudes so that there is a single, consistent way of locating all points.

Data discovery and downloads

Data discovery and downloads are data user concepts. First users need to be able to discover what data is available. Then they need to be able to download selected data (Figure 2). Discovery mechanisms are provided by servers running Apache/PHP and Geoserver and these are accessible by users running any of the major browsers and the Google Earth application. Downloadable data is provided by FME Server, and Mobile Atlas Creator is used to create tiled

maps for phones and tablets. The data and maps are then usable in various databases, GIS systems and the two main phone operating systems, Android™ and iOS™. Note that the list of destination systems illustrated is not exhaustive; these are simply the most common.

Discovery



Data download

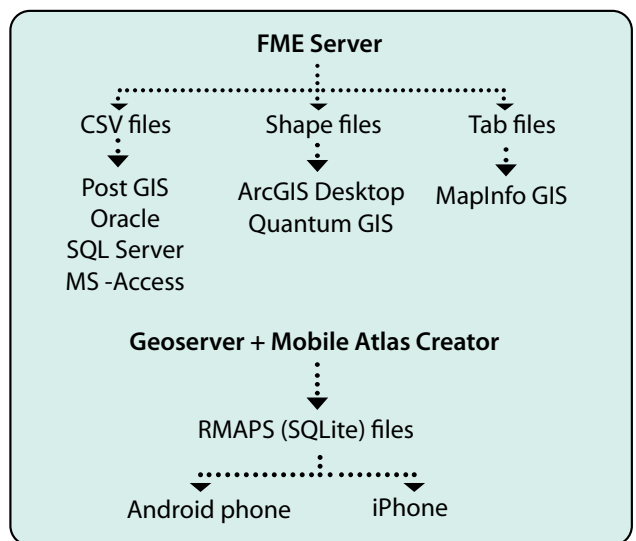


Figure 2. Schematic diagram summarising discovery and download of data (see Appendix 1 for trademark attributions).

The GDW provides two discovery mechanisms: simple queries and Google Earth.

Simple queries

Each dataset has a simple query page that allows users to enter text criteria and use a spatial search to retrieve a ‘Summary List’ of records. Clicking on a particular record brings up its full details. Producing this detailed report involves running several queries against all tables for the dataset and arranging the information in a logical order. As an extra feature, most query pages offer a choice of ‘Summary List’ or ‘Full Report’. The latter will return full reports for all the records for that search. This is done in such a way that all tables for the dataset (e.g. 17 tables for the MetIndEx database) are each queried only once. The returned data is sorted and merged and sent to the user’s web browser while the queries are still running. By using this logic, and disabling buffering on the PHP server, the top of the report appears almost immediately on the user’s web

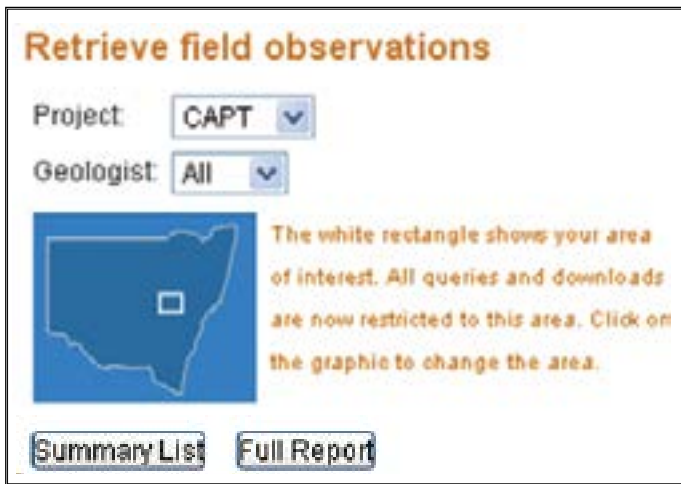


Figure 3. Screen grab from the GDW, illustrating query functionality by project, geologist and spatial extent.

page while the remainder of the report is still being generated. This unbuffered system of output works well with current versions of Firefox®, Chrome™ and Safari® browsers but does not work as designed in the versions of Internet Explorer® that were tested. Figure 3 shows an example simple query page.

Google Earth

All datasets can be viewed in Google Earth, partnered with various maps. This offers a very user-friendly environment for displaying data over satellite imagery and over geological maps or geophysical imagery without developing code other than that for creating the layer definitions in the left pane of Google Earth.

After trialling GeoServer, Esri and FME Server products, the project team chose GeoServer to deliver the data to Google Earth. GeoServer provides a high

performance system due to its dynamic interaction with Google Earth. It delivers the data differently depending on the amount of data available for the current view. To do this it uses the [kmscore](#) system. When there are more than 1000 points to display in the current view, GeoServer creates a transparent PNG raster file which, in simplified terms, is a georeferenced image of the points expressed as coloured dots. It then sends this image to Google Earth, where the image is draped correctly over the satellite imagery. When there are many tens of thousands of points involved, this is many times faster than sending all the vector data points to Google Earth. When zooming in to where there are less than 1000 points to display in the current view, the vector data points themselves are written into a KML file and this is zipped and sent to Google Earth. However, only a minimal amount of attribute information is sent with each point — just enough information to symbolise and sometimes sub-sample the layer. When one of these points is clicked in Google Earth, a request is sent to an Apache/PHP web server to run a PHP report. Only at this time is all the detailed data for that point retrieved by running queries against all data tables in that dataset. A detailed report is then displayed in a Google Earth balloon. This system very efficiently sends the minimum amount of data between GeoServer and Google Earth, with data only being sent across the internet when it is actually required (Figure 4). If the report needs to be printed, a link in the balloon opens the report in a browser where it can be printed. The PHP report is the same as that used in the simple query pages, so no extra PHP coding is required to achieve this.

The project team imposed a limit on how sophisticated to make the two discovery mechanisms. More resources could have been applied to deliver what would effectively be a web-based GIS system and complex database querying facility. However, we chose to keep the system simple to use and maintain. If users need more data analysis than the web site delivers, they can download the data and use their own analysis tools.



Figure 4. Example of data supply in GDW. Map © Google Inc., 2012.

Downloads

The download facility presents all datasets for download on one web page (Figure 5). The customer can choose to download data for the whole state or for a selected rectangular area within the state. There is a download button for each dataset. When the button is clicked, a request is sent to an FME server to generate the download file. While this is happening, the web user can continue browsing other pages or requesting other downloads. The download requests are queued and processed one at a time to maximise use of modest hardware. As each downloadable ZIP file is created, an email is sent to the requestor containing a link to the file, which can then be clicked on to download the file. The data can be formatted as CSV files, shape files or MapInfo tab files. The longitudes and latitudes for located data are included in the CSV files and as attributes in the GIS files for convenience.



Figure 5. Download buttons in the GDW.

Area of interest

Download statistics from a Mineral Resources web-based map and tenement viewing system called MinView demonstrate that most users are interested in data for only selected areas within the state. Obviously, many explorers focus on the area that includes their exploration lease and maybe some surrounding land. (This is reflected by current statistics from the GDW which demonstrate that about half of data downloads are for a chosen area within the state rather than for the whole state.)

Consequently, the spatial search caters for the concept of an 'area of interest'. This works by defining with a mouse a rectangular area of the state on a Google map base. Alternatively, geologists working on map sheet areas can select a map sheet from a drop down list. This 'area of interest' is then applied to all areas of the website and in Google Earth. All simple query pages and the download page are constrained by the chosen 'area of interest' and Google Earth opens up zoomed into the 'area of interest' (Figure 6).



Figure 6. Screenshot of the GDW illustrating spatial extent search by map sheet area, in this example the Dubbo 1:250 000 map sheet. Google and the Google logo are registered trademarks of Google Inc., used with permission.

Unlocated points raise interesting problems in a spatial search. How should unlocated data be treated when spatial criteria are being applied? After trying several unsuccessful ideas, the project team adopted the system that when more than half the state is chosen as the area of interest, then this will be treated as if the whole of the state has been selected. In this case, no spatial criteria are applied to the query and all

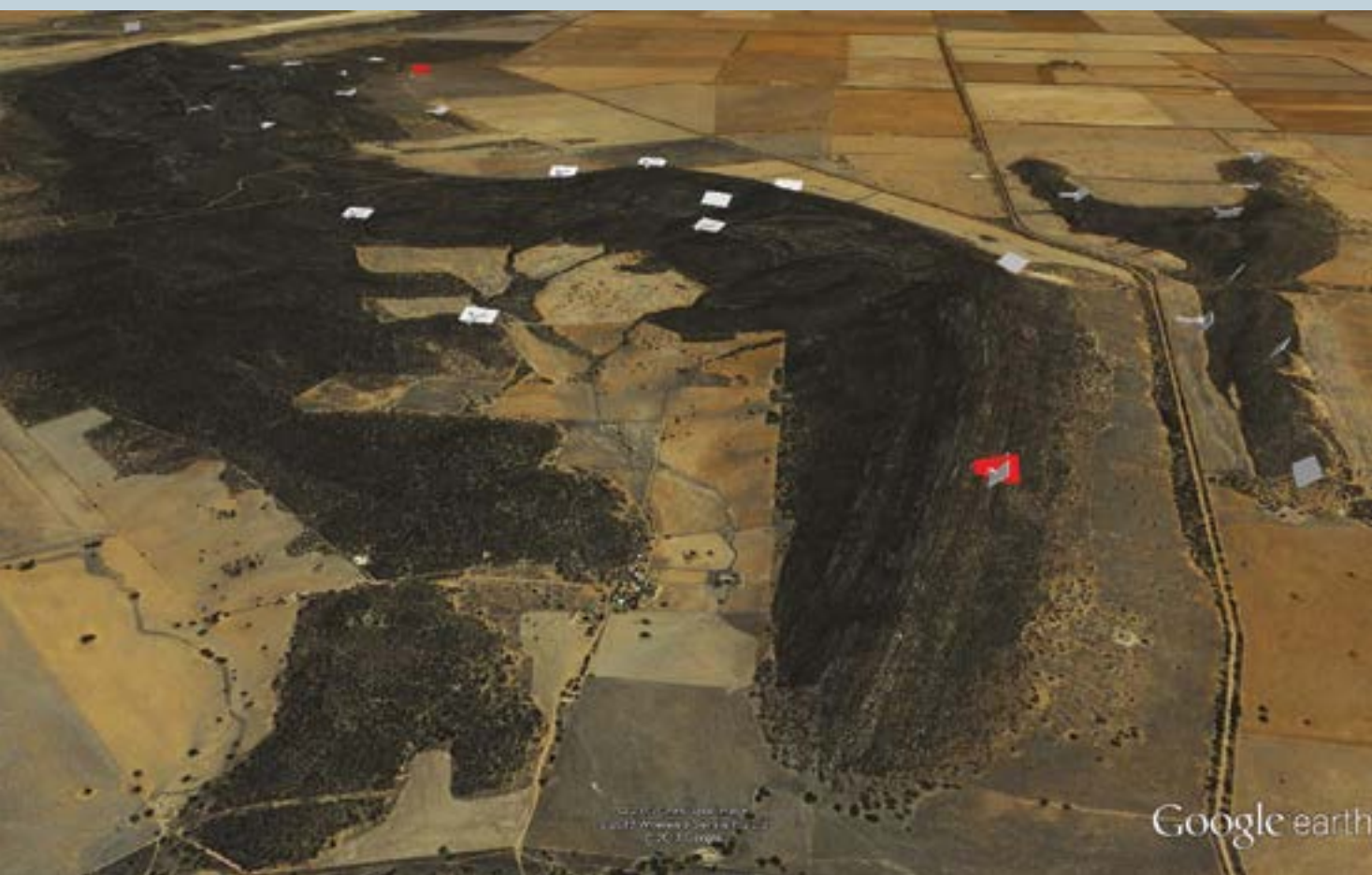
unlocated and located data is searched, using only the text criteria. If an area of less than half the state is chosen, then the search will only operate on data which has assigned latitudes and longitudes that occur within the area of interest. For downloads, located data is restricted to the area of interest, but all unlocated data for the whole state is also included in the download in a separate clearly named file.

Case study 2

Structural readings

Structural readings are among the most basic geological measurements obtained during field work. For example, the dip and strike of bedding planes and/or cleavage orientations, and/or the plunge of a fold axis are recorded during field work using a compass and inclinometer. Structural readings captured in the GDW are stored in the Field Observations dataset. In this way they are related to all other information recorded at the location. At the time of writing, most structural readings captured in Geobank were harvested from published maps in eastern NSW (Zone 56, MGA).

In the accompanying figure, structural data from the GDW has been converted into 3D shapes using a program developed by Blenkinsop (2012). This is loaded as a .kml file into Google Earth. The 3D terrane view, combined with a transparent geological map, provides a great visual context to the 3D structural symbols. In the figure, structural data from the GDW is plotted on outcropping ridges of Devonian Mulga Downs Group rocks that were deformed in the Carboniferous Kanimblan Orogeny. Bedding orientations are compatible with the strike and dip direction of the eroded ridges. However, the red upright cleavage symbol is axial planar to the fold and graphically highlights the contrast of cleavage and bedding.



3D structural symbols from the GDW shown on Google Earth imagery over the Rankins Springs 1:100 000 map sheet area. Google and the Google logo are registered trademarks of Google Inc., used with permission.

Systems documentation and source control

The project team uses Team Foundation Server (TFS), which is supported by the corporate IT section (IT Support), for tracking issues and source control of code. The TFS client software requires a major installation and is therefore not suitable for running on servers where the code for the GDW is developed and maintained. To cater for this, the IT Support established a SVN 'bridge' to the TFS server. This allows submission of updated code to the TFS server using simple SVN commands instead of using the TFS client software. SVN commands are contained within batch files that submit updated code to the TFS server when the developer logs out of the server where code is being updated.

'Issues tracking' is used to record problems and requests, to ensure that jobs do not get lost, and to manage the assignment of resources to tasks. This is accessed through the TFS web interface, so that only a browser is required, not the TFS client software.

Documentation relating to the installation and configuration of server software systems such as GeoServer, FME Server and Apache, is stored in a wiki using the open source Docuwiki product. Initially, this documentation was saved as Word documents and text files on servers and networked drives. However, Docuwiki provides a single-point documentation system which also conveniently allows editing of documentation from servers where Microsoft Office is not installed.

Support structure

The corporate IT section supports the underlying services to the GDW such as networking, proxy servers, hardware and operating systems. It also provides file servers, database servers, Citrix servers, licence servers and backup systems. Geoscience Information (a sub-unit of the GSNSW) is supplied with virtual Windows servers to use as application servers and has full operating system administrator access to these to enable applications to be installed. This means that all software above the operating system layer is the responsibility of Geoscience Information. These include Apache/PHP, Apache Tomcat, Geoserver (with local PostGIS data store), FME Desktop, FME Server, database client software and Esri client software. Taking over these responsibilities in Geoscience Information greatly reduces the amount of interaction with IT Support, resulting in a much more flexible and efficient development and maintenance environment.

Database design

Referential integrity is used extensively in SQL Server and all code fields have referential integrity rules linked to the appropriate lookup list. This ensures that code fields cannot contain invalid codes. In addition, foreign keys to other data tables generally will have referential integrity rules linked to the appropriate data table to ensure that all foreign keys only contain valid values. For example, the table holding drillhole intervals can only have a SITE_ID code that corresponds to an existing drillhole record in the main drillhole table. That is, a drillhole interval cannot exist for a drillhole that does not exist. This sounds like a very simple idea but requires discipline to implement and maintain globally, and is important in terms of maintaining the integrity of the data.

Primary keys are defined in all tables, and any other fields that are unique have a unique index applied. Fields are strongly data-typed. Character fields have strict length limits, the DATE or DATETIME data type is always used for dates and times, and fields that should only contain integers are data-typed as such. Range checks (e.g. a number should be between 1 and 100) are generally performed in the Geobank application rather than at a database level. When a validation rule is too complex to be applied in the Geobank application, a trigger is defined at the database level. Triggers are also used to populate some records with the name and date of the data entry person creating or modifying a record.

Database naming standards

A strict protocol is applied to naming tables. This provides an intuitive understanding of what the table is and its relationship to a dataset. A two-character prefix in the table name groups tables belonging to the same dataset when they are displayed in alphabetical order. The prefix represents the dataset to which the table belongs, for example, FO_SITE is the FieldObs (field observations) site table. Lookup tables have an additional 'L' in the prefix. For example, FOL_LITHOLOGY is the lithology lookup list in the FieldObs dataset. SQL Server views have an additional 'V' in the prefix, for example, FOV_SITE is a view of the FieldObs site table.

Field names are less than 10 characters for the following reasons.

- SQL statements need to be short and readable, particularly when embedded in a procedural language such as PHP
- The associative array indexes automatically created within PHP are also short and readable

- Long column names widen the columns more than necessary for the displayed data
- Aliases will not let users correlate data columns in their reports with published metadata
- Some file formats such as shape files truncate long field names

Lookup lists consistently use the field names 'CODE', 'NAME' and 'DESCR' and also create a 'SORTBY' field if sort order needs to be specified. The 'CODE' is typically no more than four to six characters long. The 'DESCR' field name is only used at data entry time and is displayed in the pop-up list of values to provide guidance to the data entry person. It will not appear in reports. The use of the 'CODE' field is a deliberate deviation from the common convention of prefixing 'CODE' or 'ID' with the table name. For example, if the code field in the 'FOV_LITHOLOGY' table was named 'LITHOLOGY_CODE', the full name in a SQL statement becomes 'FOV_LITHOLOGY.LITHOLOGY_CODE'. Using the 'CODE' name reduces the full field name to the more usable 'FOV_LITHOLOGY.CODE'.

GeoServer

GeoServer is a spatial data server which delivers a variety of [Open Geospatial Consortium \(OGC\)](#) protocols including WMS, WFS and WCS. The Geological Survey uses GeoServer to deliver KML/KMZ to Google Earth and uses a small local PostGIS database to deliver data to GeoServer.

For simplicity, the [CSS Styling extension](#) provides scale-dependent colouring and symbolisation within GeoServer replacing SLD, the native styling language for GeoServer (Figure 7).

```
* (
  mark: symbol(square);
)
/* All points are size=3 */
:mark (
  size: 3;
)
/* @title Petroleum */
[PROJECT = 'PET'] :mark {
  fill: blue;
  stroke: blue;
}
```

Figure 7. Portion of the CSS style description for the drillhole layers.

About 15 data layers are stored in PostGIS tables, but these are often divided into sub-layers in Google Earth. This is achieved by using [Common Query Language \(CQL\)](#) queries within the 'networked link' URL (Figure 8).

Apache/PHP reports are configured to run in Google Earth balloons (pop-ups) using [KML Placemark Template files](#) (Figure 9). This allows an Apache/PHP page to be displayed in an iframe in the lightweight browser of the Google Earth balloon.

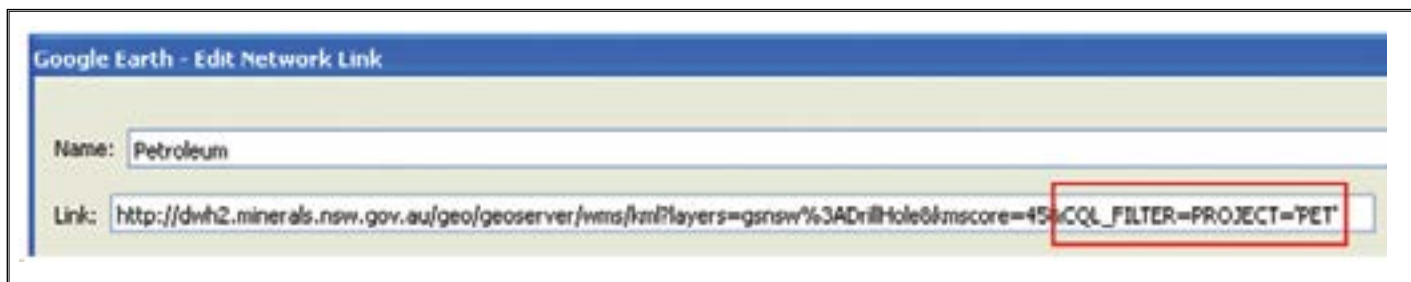


Figure 8. Network link for the petroleum drillhole sub-layer showing the query within the URL.

```
<div align="center">
<b>${PROJECT.value}<br><br>${SITE_ID.value}</b><br><br><br><br>
<a href="http://.../CI/warehouse/report/FieldObs?project=${PROJECT.value}&site_id=${SITE_ID.value}"
  target="_blank">View in Browser</a></div>
<iframe src="http://.../CI/warehouse/report/FieldObs?project=${PROJECT.value}&site_id=${SITE_ID.value}"
  width="300" height="400" frameborder="1">
```

Figure 9. Example of a description.ftl file for displaying a FieldObs PHP report inside a Google Earth balloon.

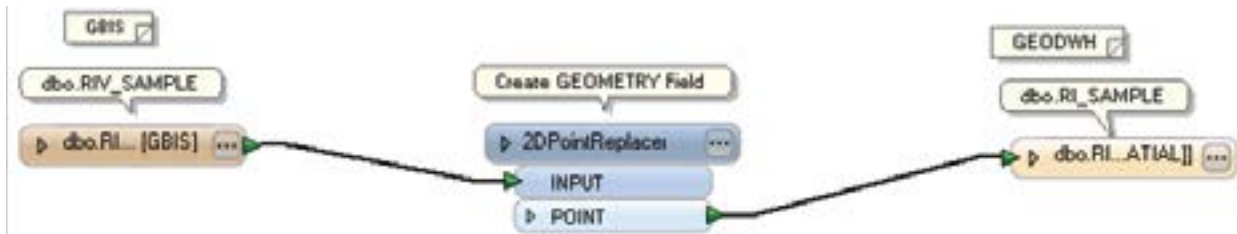


Figure 10. Very simple replication using FME Desktop — reader, transformer, writer.

FME Desktop — a ‘spatial ETL’

An ETL (‘extract, transform, load’) transforms and replicates data to and from data warehouses. It has a graphical interface that displays ‘readers’, ‘transformers’ and ‘writers’ and these are combined to achieve the desired data transformation and flow. The project team decided to use an ETL instead of writing scripts, so as to create a more easily maintained system. After trialling various software packages, we chose FME Desktop and supplemented it with SQL Server views outside of FME Desktop to join tables. Several staff were trained in the use of FME software. Figure 10 shows an example of an FME job.

Web site development

The GDW uses PHP as its web development language because it is an easy-to-use, high level language that is sufficiently powerful to achieve the project’s goals. Later stage use of the MVC based CodeIgniter framework improved the structure, modularity and security of the code. To achieve consistency of appearance with the corporate web site, CSS style sheets were copied from the corporate content management system. HTML generated by the PHP code is deliberately plain so, together with the CSS style sheets, the corporate style is largely achieved.

GDW has an internal and a public data warehouse web site. However, the GDW uses one code base so there is less code to maintain, and the system is tested quite

well internally through daily use before it is updated on the public site. The public site needs to be W3C-compliant so the system is not written specifically for the corporate browser, Internet Explorer, which is installed on all corporate PCs. The system works best using a recent version of Firefox, Chrome or Safari browser.

Spatial indexing of DIGS reports

A spatial search for DIGS reports and an advanced DIGS search page (Figure 11) are powerful features of the GDW. They are available to the public through the GDW website. There were two options considered in designing the spatial search process. Either create a polygon for a report from information in the ‘TENEMENT’ field or create a polygon from information in the ‘MAPNAME’ field. Initially, the former option seemed more attractive but was discarded in favour of the ‘MAPNAME’ field. This is because only a small percentage of ‘TENEMENT’ fields could be processed to form a valid shape and also tenement boundaries change over time, making the process quite complicated and possibly unreliable.

The DIGS database (without attached files) is replicated to the internal and external data warehouse each night. Each weekend, a script reads through all the non-empty ‘MAPNAME’ values and creates a shape for each report. This shape is stored in a field with the ‘GEOMETRY’ data type in SQL Server and it is this field that is referenced by the DIGS advanced search.

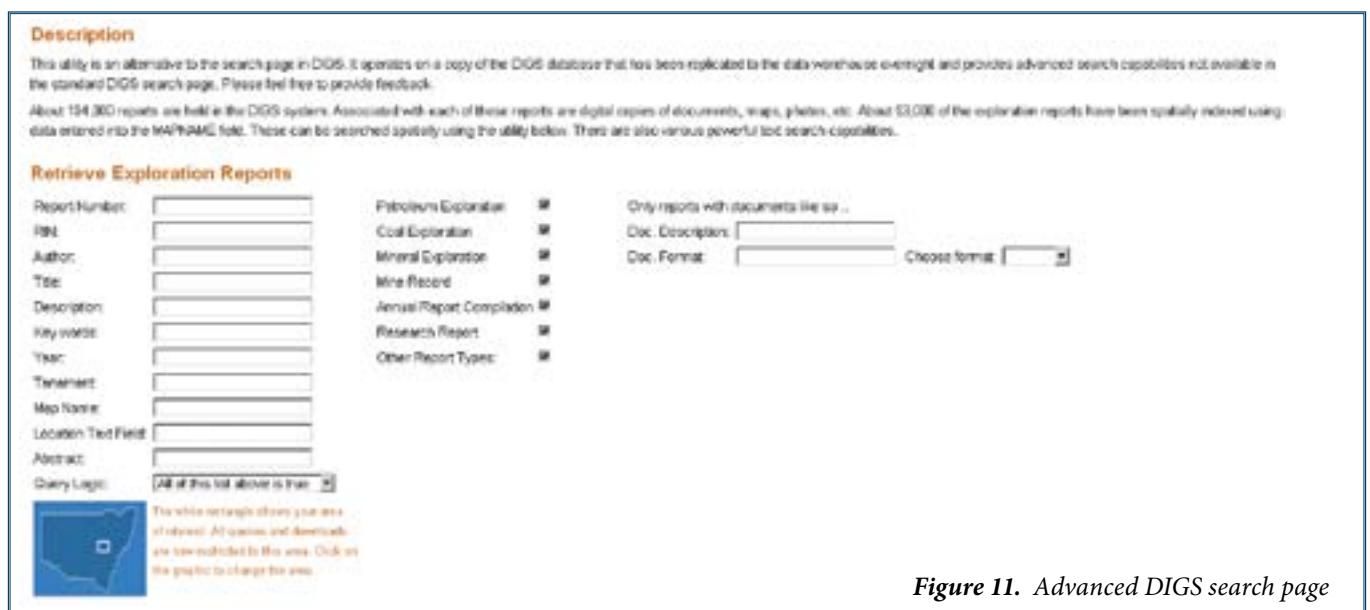


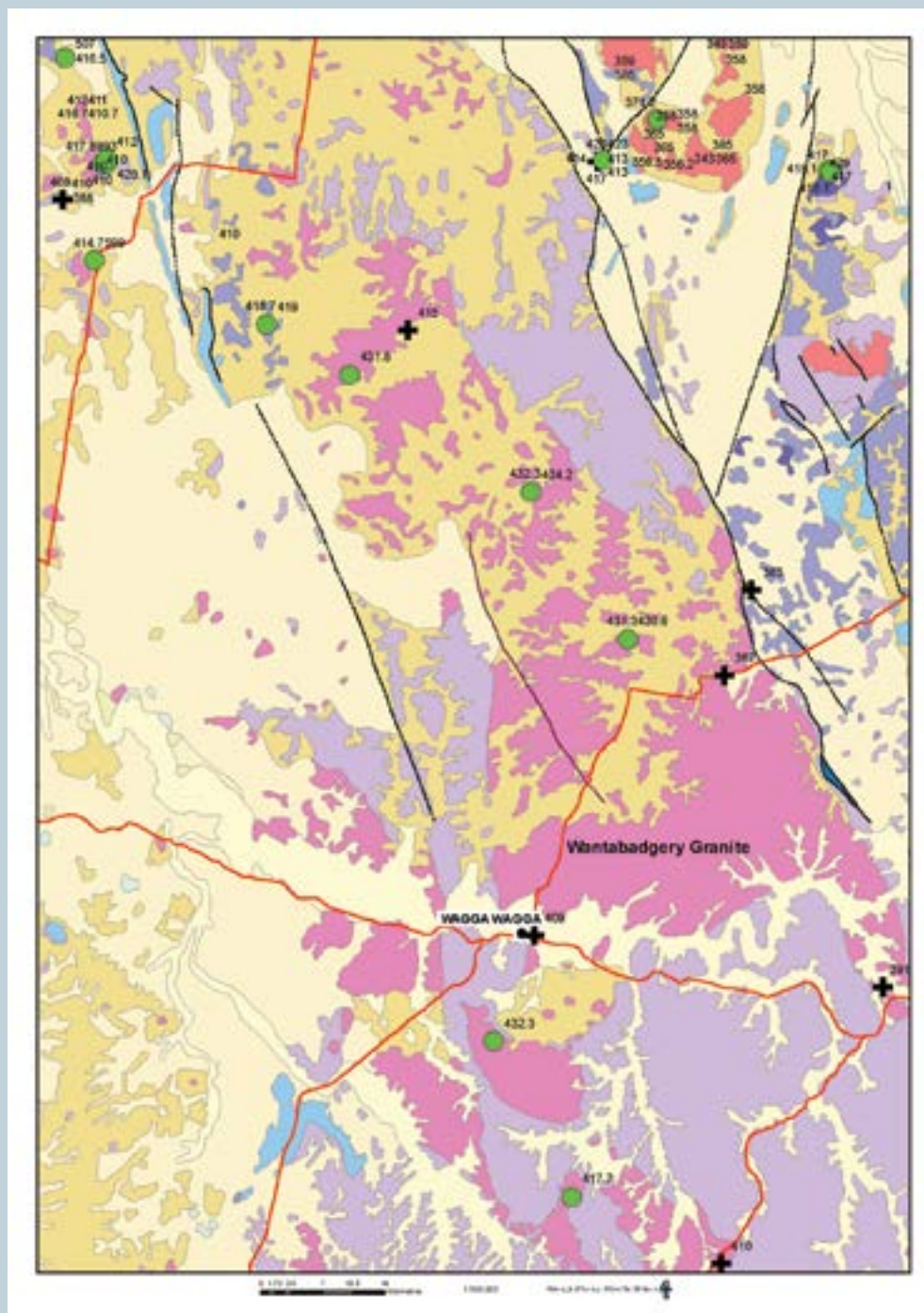
Figure 11. Advanced DIGS search page

Case study 3

Radiogenic isotope data

The radiogenic isotope dataset contains geochronological data that has been collected by the GSNSW or harvested from published research papers. The isotopic systems represented are U–Pb, K–Ar, Ar–Ar, Rb–Sr, Pb–Pb, Re–Os, Sm–Nd and U–Pb provenance. The database stores location and sample collection, lithological description, analytical and interpretation information for geological samples in NSW. The database includes links to stratigraphic unit, mineral occurrence and drillhole data.

An example of the radiogenic isotope data in the Wagga Wagga region is shown in the accompanying figure. Sample localities for U–Pb SHRIMP II analyses are shown as green circles, and K–Ar sample localities are shown as black crosses, labelled with their interpreted age (in millions of years). For the Wantabadgery Granite, the interpreted ages determined by the K–Ar technique are clearly younger than the U–Pb SHRIMP II ages for the same granite. This is because the closure temperature of the K–Ar isotopic system is lower than the U–Pb system (e.g. Williams 2001). The U–Pb ages are therefore more likely to be more representative of the magmatic crystallisation age, and the K–Ar ages reflect either protracted cooling of the granite or later metamorphism.



Geochronological data from the radiogenic isotope dataset, overlying a surface geology map (Geoscience Australia 1:1 000 000 surface geology) in the Wagga Wagga region.

Internal system tools

Security

Rather than creating a new system of usernames and passwords for access to the GDW, the internal web server is integrated with the corporate Active Directory system using an Apache module called SSPI. Access is controlled via membership to Active Directory groups. There are instructions on the internal system describing how to set up Firefox and Chrome for 'single-sign-on' so that the browsers do not need to prompt the user for their password. Using this system, access to all web pages on the internal GDW is limited to Minerals Resources staff. Access is also restricted to some of the internal utilities, for example the FieldObs synchronisation page.

FieldObs field data collection system

Mineral Resources geoscientists require a specialised relational database for recording geological observations in the field. The criteria for sourcing tablets to display the FieldObs interface were that they be under one kilogram and have a screen resolution of 1024 x 768 pixels. The FieldObs database model comprises seven data tables and 28 lookup lists. The interface design includes many 'cascading' drop-downs. This means that many of the lists of values that are valid for a particular field are subsets of the total list of valid values, and this subset is determined by a choice in a previous drop-down list field. Another consideration is choosing colours that are suitable for use in sunlight and that will not impact greatly on

battery power. The final design has dark greys for most of the screen, with white fields and large bold fonts for text (Figure 12). The development work was conducted in Microsoft Access to minimise coding. Even with this system, about 2000 lines of VBA code were required to customise the system.

The project team is currently monitoring developments on 7" Android and Apple ARM-based tablets to look for opportunities to use smaller tablets that are lighter and have a longer battery life than the current Windows Intel-based tablets. The current 'ruggedised' tablets have a battery life of about four hours in the field when the batteries are almost new. So two sets of batteries need to be carried in the field and even this can be insufficient for a long day's work. The battery life deteriorates further after only a year.

The system cannot rely on an internet connection in the field. When field geologists return to the office and connect to the network, a PHP system synchronises copies of FieldObs.mdb on the tablet with the master data in Geobank. This relies on an 'UPDATED' field in each of the FieldObs data tables to determine for each row which system holds the most recent data.

A customised DLL written in C++ for ArcGIS® Desktop allows a new observation to be created in FieldObs (with the current latitude and longitude) while using ArcGIS Desktop. This was adapted and simplified for the new FieldObs system from the old 'SITES' system. Note that this DLL is only required because ArcGIS Desktop prevents other applications from reading the GPS coordinates if it is reading the GPS. The project team is currently investigating performing all location requests via the 'Windows Location Provider' in Windows® 7 and 8. This would result in all location-aware applications being able to determine the GPS location simultaneously.

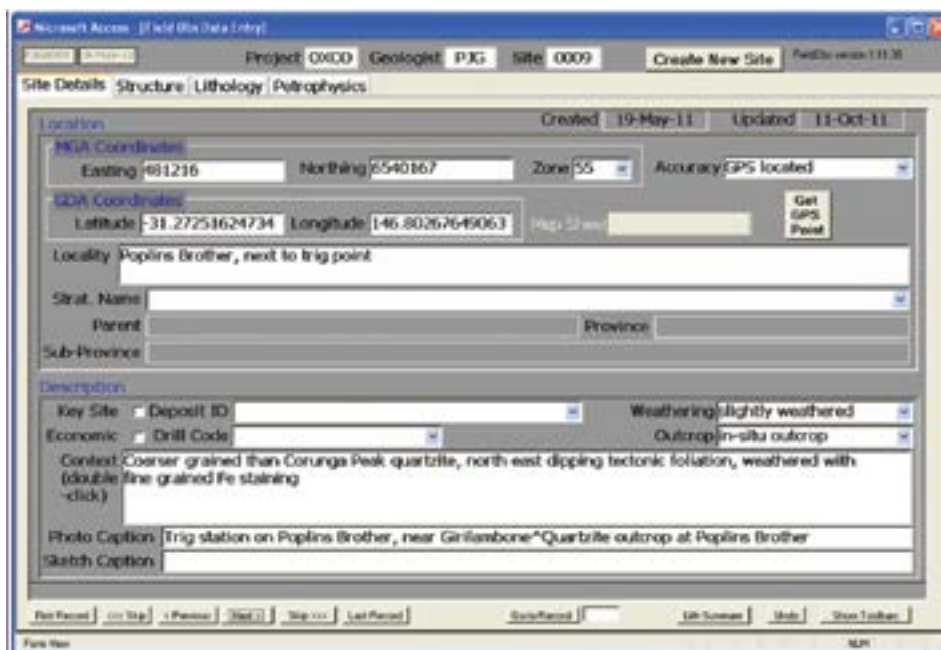


Figure 12. Screen grab of the MS Access® form for collecting field observations data.

Photographs — geo-location and other metadata

Information about photographs is stored within them in the 'header' of the JPG files. The standards for storing this 'metadata' in the photographs are fairly immature at present. Different commercial software systems commonly write the same data, such as title and descriptions, to different locations or in different formats. Currently, longitude and latitude data is stored in the EXIF part of the header (as is standard across all software); captions are in IPTC format (so that these can be edited easily using the XnView program), and the remainder of fields such as description, title and photographer, are in XMP format. This system can be reviewed when more rigorous standards are in place.

The 'single point of truth' for photographic information is regarded as the photograph itself, so this data is not held in Geobank, but is replicated to the GDW each night for use in web pages and Google Earth. This is done using the open source pyexiv2 library within Python scripts. Unfortunately, the exiv2 library is not available in a PHP module.

No commercial, graphically driven software was suitable for writing our photograph metadata into JPG files. Any such system needs to extract data such as coordinates and captions from a SQL database. It needs to allow editing of that data, perform validation on some fields, and enforce the writing of the data to the design specified locations in the JPG files. To achieve this, the web-based Photo Information Editor (PIE) was developed using PHP and is now used for editing

field observations, drillhole and petrology photographs (Figure 13). The processed photographs are stored on a file server to ensure that they are captured by the corporate backup system. The photos are generally readable, but only writable by the Photo Information Editor. XnView is a very useful desktop tool for checking that photos have been processed correctly.

Reports

Two different types of internal reports are required — well-formatted, printable reports for standard use, and ad hoc queries for often urgently required data. [Crystal reports](#) is used for the former within [SAP® Business Objects Explorer](#) and is supported by the corporate IT section. [dbForge Query Builder for SQL Server](#) is used for ad hoc queries. It has a graphical query builder similar to that of Microsoft Access but recognises all SQL Server data types and functions. The application is currently being set up on a Citrix server. All developed queries are saved to a specific area on the local file server.

GetDIGSData

The project team developed a PHP program called 'GetDIGSData' to process CSV data files submitted by exploration companies with their reports to Mineral Resources. The files relate to assay and Niton data in the mandatory ['Exploration Reporting Templates'](#) which accompany the company reports. 'GetDIGSData' processes drillhole location files, drillhole survey

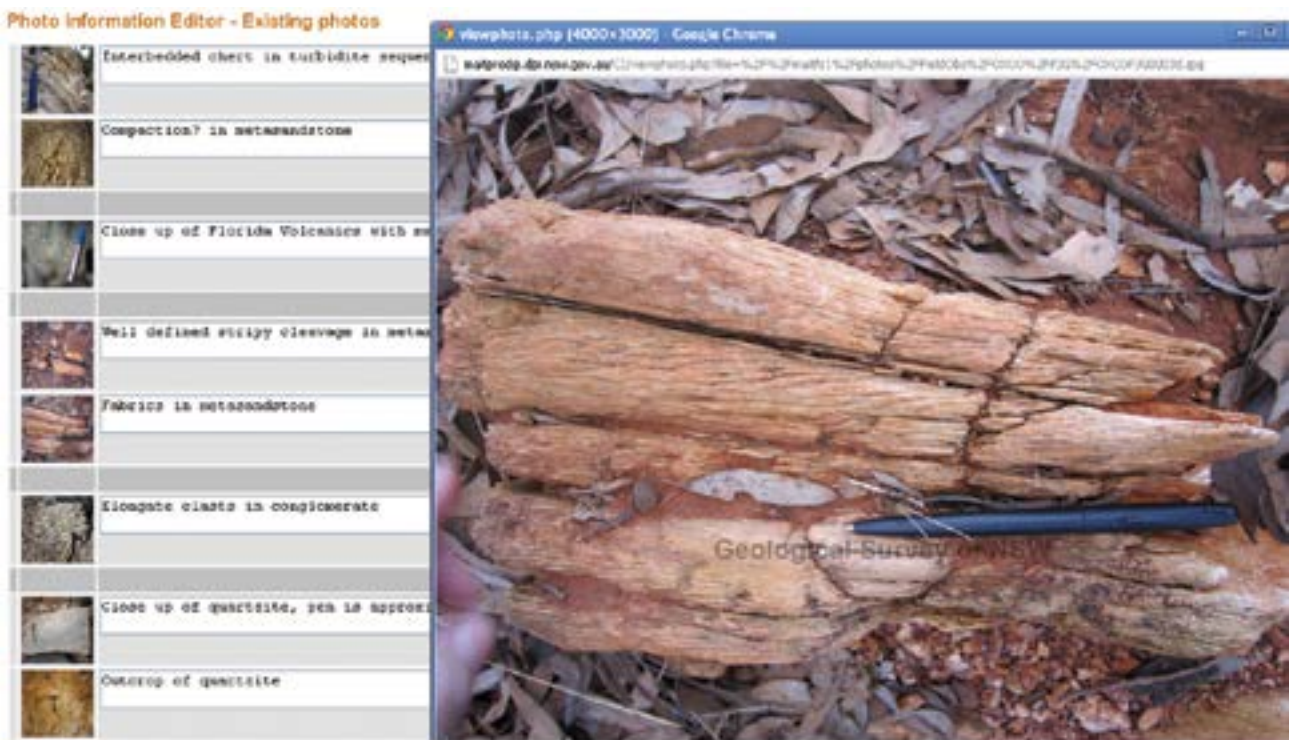


Figure 13. Example of the photograph upload facility, where photographs are matched with their FO_ID and caption.



Figure 14. Google Earth image from the GDW showing drillhole location that is probably incorrect. Google and the Google logo are registered trademarks of Google Inc., used with permission.

files, and downhole and surface geochemistry data, resulting in the data and metadata being written to a dedicated table structure in SQL Server. An important part of the quality assurance process is the displaying of all data points in Google Earth along with the exploration lease boundaries. When the operator zooms in, it becomes obvious if points are wrongly located in the middle of dams, on top of spoil heaps or outside of the lease area (Figure 14). The error often is that the grid coordinate system or the grid

coordinate zone has been wrongly specified in the CSV file. This can be adjusted or, if there is some doubt about the correction, then the data can be rejected and the exploration company asked to correct it.

The code becomes quite complex when dealing with format variations in historical CSV files. To deal with this, the program was developed iteratively as the operators processed thousands of files. Even with the built-in robustness of the program, many of the files need to be massaged by the operators. To avoid repeating this process, we updated the templates to clarify which fields need to be modified by the explorer and how the submitted information needs to be formatted. The improved templates and continued feedback to mineral explorers should improve the quality of submitted data substantially over time.

Currently there are about 30 million assay results. This volume of data means that nightly replication tasks need to be designed to transfer only data that has been updated recently and full replication of all assay data is done on Sunday. AMG and MGA coordinates are transformed to GDA94 latitudes and longitudes when the data is transferred to the internal data warehouse. Figure 15 shows a screenshot of the program.

For further information and feedback on the GDW contact david.collins@industry.nsw.gov.au

DOWNHOLE GEOCHEM FILE							
DG_RPT_GENERAL							
D_REMARKS Remarks: Depths_measured_from Hole collar elevation							
PROCESSOR							
DG_RPT_CODE							
CODE	FIELD1	FIELD2	RPT_ID	FILETYPE	CODETYPE		
RAB			542	DOC	SAMPL		
DD	DD Diamond Core	Description of sampling technique in the field	542	DOC	SAMPL		
eg Job Number: OR05890758							
AA25	Fire assay	ALS Chemex	542	DOC	ASSAY		
ME-ICP610	Induction Coupled Plasma ; Fusion, ICPAES, Pyromet		542	DOC	ASSAY		
DG_DCH_META							
RPT_ID	ELEMENT	UNITS	ASSAYCODE	DETLIMIT	ACCURACY	UPDETLIMIT	SURFDOWN
542	Hole ID	units	Assay_code	0	0	0	DH
542	From	metres		0	0	0	DH
542	To	metres		0	0	0	DH
542	BaSO4	%	ME-PK084	0.1	0.01	0	DH
542	SrO	%	ME-PK084	0.1	0.01	0	DH
542	SiO2	%	ME-PK084	0.1	0.01	0	DH
542	Fe	%	ME-PK084	0.1	0.01	0	DH
542	Mn	%	ME-PK084	0.1	0.01	0	DH
542	G.O.	unity	ME-PK084	0.1	0.01	0	DH

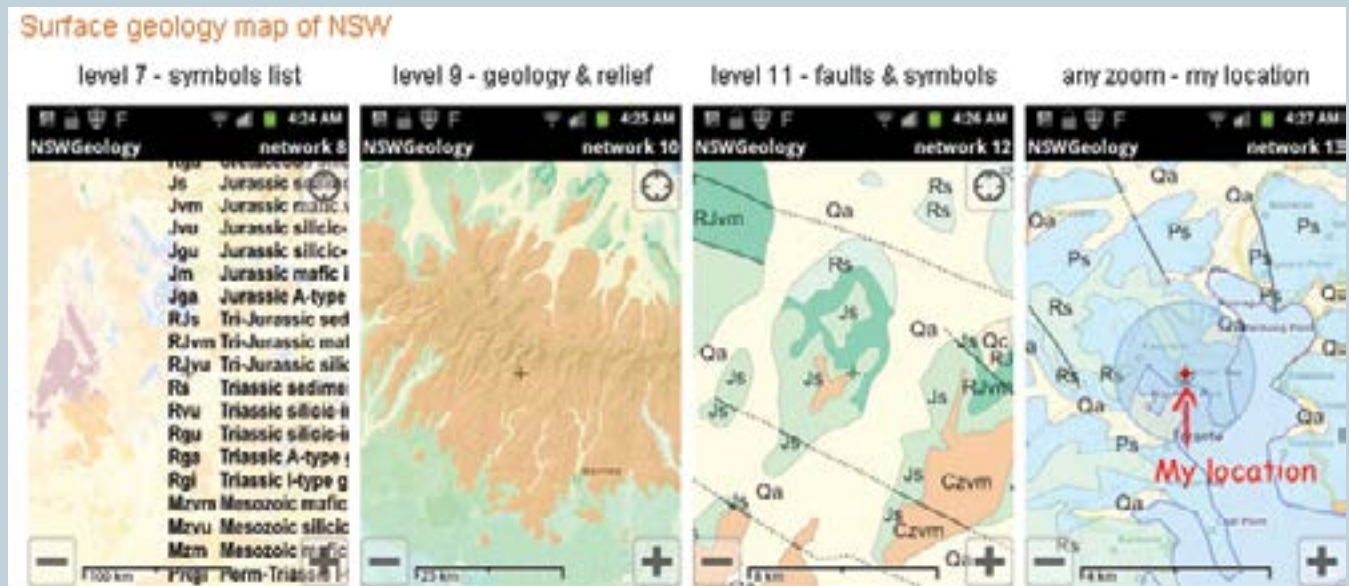
Figure 15. Screen shot of the GetDIGSData processing system.

Case study 4

Geology and geophysics maps for phones and tablets

SQLite® map tile files are provided for mobile devices. The NSW 1:1 500 000 surface geology map, Radiometrics Image of NSW and Total Magnetic Intensity Map of NSW were processed using Geoserver and the Mobile Atlas Creator (MOBAC) to produce an RMaps format SQLite file that can be viewed using RMaps on Android™ phones and tablets, Galileo on iPhone™ and iPad™ devices and many more mobile device applications. The surface geology map and geophysical images are downloaded to the mobile devices permanently so that no internet connection is required after the initial download. This means that users can see what the local geology and geophysics is wherever they are located in NSW if they have a GPS receiver on their device.

The mobile maps are available for download from the GDW main page linked to the URL <http://dwh.minerals.nsw.gov.au/CI/warehouse?content=mobileapps>



Screenshots from the mobile app illustrating different zoom levels available for the NSW 1:1 500 000 surface geology map.



By using GPS capability in a smartphone, the app will work outside of mobile phone range. (Photographer P. Gilmore).

Acknowledgements

The authors would like to acknowledge the work of Kevin Capnerhurst and Catherine Shirley in database design and migration; Marta Vega Faundez and Paul Parker for their thoroughness in compiling and checking data; members of the COGENT II steering committee (Brad Mullard, Adrian Delaney, Lindsay Gilligan, Boyd Wymer, David Trowbridge) and members of the COGENT II Project Board (John Watkins, Graham Butt, Gary Colquhoun, Trisha Moriarty, Kevin Ruming, Jola Jaworska, Peter Downes, Phoebe Hines, Phil Blevin, Guy Fleming, Ken McDonald, Adam King, David Robson). Joan Henley is thanked for reviewing and editing the original manuscript.

References

- BLENKINSOP T. G. 2012. Visualizing structural geology: From Excel to Google Earth. *Computers & Geosciences* **45**, 52–56.
- WILLIAMS I. S. 2001. Response of detrital zircon and monazite, and their U–Pb isotopic systems, to regional metamorphism and host-rock partial melting, Cooma Complex, southeastern Australia. *Australian Journal of Earth Sciences* **48**, 557–580.

Appendix 1: Glossary of terms

Apache™ HTTP Server	open source HTTP web server
Apache Tomcat™	provides a 'pure Java' HTTP web server environment for Java code to run
Aptana® Studio	an open source integrated development environment for various languages
ArcGIS Desktop	a geographic information system by Esri® for maps and geographic information
Citrix server (XenDesktop®)	a desktop virtualisation solution
CodeIgniter®	an open source rapid development web application framework, for use in building dynamic web sites with PHP
Crystal Reports®	a business intelligence application used to design and generate reports from a wide range of data sources
CQL	Contextual Query Language: a formal language for information retrieval
CSS	Cascading Style Sheets: a style sheet language used to describe the look and format of a document written in a markup language
CSV	text file containing comma separated values
DIGS®	Digital Imaging of Geological Systems: an interactive GSNSW database for exploration and geoscience information in scanned and digital documents
Dokuwiki	open source wiki application aimed at small companies' documentation needs
Eclipse™	an integrated development environment for PHP, Python, HTML, JavaScript™, CSS
Esri®	an international supplier of Geographic Information System (GIS) software and geodatabase management applications
ETL	Extract-Transform-Load: a database process used especially in data warehousing
EXIF, XMP, IPTC	3 different standards for storing metadata within digital photos
FME®	spatial data transformation software by Safe Software Inc.
Galileo	Galileo Offline Maps is a map browsing app that can be used offline and runs on iPhone® and iPad® devices
GBIS	Micromine® geological data management software now superseded (by Geobank®). Original data management software purchased under the COGENT II project
GDW	Geoscientific Data Warehouse: a web-based data discovery and delivery system using Google Earth™, simple web queries and file downloads; developed by GSNSW
Geobank®	Micromine® geological data management software. Superseded GBIS. Used to store relational geoscientific data (i.e. the Geoscience Product Catalogue)
GeoServer	open source server written in Java®. It allows users to share, process and edit geospatial data
GIMP	Gnu Image Manipulation Program: an open source, image retouching and editing tool
Google Earth™ mapping service	a virtual globe, map and geographical information program

GPC	Geoscience Product Catalogue: an online interactive spatial and textual search for Mineral Resources branch (NSW Trade & Investment) retail products and for downloading geological and metallogenic maps and their vector data.
iframe	an inline frame which places another HTML document in a frame
Java®	a general purpose, concurrent, class-based, object-oriented computer programming language that is specifically designed to have as few implementation dependencies as possible
KML/KMZ	the XML-based markup language used for exchanging geographic information for use with Google Earth™
kmscore	KML scoring which limits the number of features shown on a map based on certain criteria
MapInfo®	a Microsoft Windows-based mapping and geographic analysis application
MetIndEx®	the GSNSW mineral occurrence database for NSW
MinView	MinView: a Mineral Resources online interactive spatial search for exploration tenement and geoscience information for NSW
ModelRight™	a database design and data modelling tool
MVC	Model-View-Controller: a software architecture pattern that separates the business logic from the presentation logic
Notepad++	an advanced text editor and source code editor for use with MS Windows™ operating system
Open Geospatial Consortium (OGC)	an international voluntary consensus standards organisation encouraging development and implementation of open standards for geospatial content and services, GIS data processing and data sharing
Open Source Initiative	a non-profit corporation dedicated to promoting open source software
PIE	Photo Information Editor: a picture browser and viewer with photograph metadata and file renaming features (developed by GSNSW)
PHP	a server-side scripting language designed for web development and as a general purpose programming language
placemark template	used in GeoServer to specify what content appears in the popup balloon when a point is clicked in the Google Earth™ program
PostGIS	program that helps servers increase the amount of data migration
Rmaps	an on and offline map-based navigation program that runs on Android™ phones and tablets
SAP® R/3® Enterprise	an integrated enterprise resource planning software solution
SDE	Spatial Database Engine: ArcSDE™ is an Esri® system for spatial relational data
Shapefile	an industry-standard geospatial vector data format for geographic information system software regulated by Esri®
SPOT	an acronym for Single Point of Truth, referring to the concept of storing each data element only once, to avoid duplication or multiple versions
SQL	Structured Query Language: programming language used in designing relational database management systems

SQLite	a software library that implements a self-contained, server-less, zero-configuration, transactional SQL database engine
SVN	abbreviation for Apache Subversion™: a software versioning and revision control system distributed under an open source licence
TFS	Team Foundation Server™: a Microsoft® product offering source control, data collection, reporting and project tracking — intended for collaborative software development projects
URL	Uniform Resource Locator: address of a webpage or document on the world wide web (WWW)
VBA	Visual Basic for Applications® is an implementation of the Microsoft® event-driven programming language Visual Basic® 6 and its associated integrated development environment (IDE)
W3C -	The World Wide Web Consortium is the main international standards organization for the World Wide Web
WCS	Web Coverage Service: provides access, subsetting, and processing on coverage objects
WFS	Web Feature Service: for retrieving or altering feature descriptions
wiki	a website which allows its users to add, modify or delete its content via a web browser, usually using a simplified markup language or a rich-text editor
WMS	Web Map Service: provides map images
XnView	a cross-platform image viewer used for viewing, converting, organising and editing graphical and video files

Appendix 2: GDW URL

URL of the public GDW: <http://dwh.minerals.nsw.gov.au/CI/warehouse>

The screenshot shows the home page of the Geoscientific Data Warehouse (GDW). The page is titled 'Minerals and Petroleum' and is part of the NSW Government's 'Trade & Investment Resources & Energy' section. The main content area is titled 'Data discovery' and includes an 'Introduction' section, a 'Supported browsers and devices' section, and an 'Area of interest' section. The 'Area of interest' section features a map of NSW with a white rectangle indicating the area of interest. The page also has a sidebar with navigation options like 'DISCOVER OUR DATA', 'DOWNLOAD OUR DATA', and 'FEEDBACK & HELP'.

Figure A1. Home page of the Geoscientific Data Warehouse.

Appendix 3: data

The COGENT II project team prioritised the data with respect to value and risk. Radiogenic isotopes were considered the highest priority due to the cost of recollection and reanalysing samples, coupled with storage of records in multiple versions of spreadsheets on the computer network and on individual computers. Resources were allocated to compile, model and migrate data based on its priority. All data is secured and has a work plan for implementation into the GDW yet. Due to finite resources, not all data is in the GDW. (See Tables A1 and A2.)

Table A1. Data implemented as of 28 February 2013

Dataset	Description	Records
Radiogenic isotopes — except Pb–Pb	Sample and analytical data for U–Pb, K–Ar, Ar–Ar, Re–Os and Sm–Nd isotopic studies	2169 samples
Petrological collection	Catalogue and description of thin sections of rock	93 126 thin sections
Drillholes — minerals	Minerals drilling (mostly exploration data plus departmental stratigraphic) including collar, survey, lithology data	44 117 collars
Drillholes — petroleum	Petroleum wells — department and industry (CSG, oil, gas) — including collar, survey, lithology, other downhole data	864 collars
Drillholes — coal	Coal drillholes (department and industry) including collar, survey, lithology, other downhole data	56 303 collars
Exploration geochemistry — down hole assay	Analytical data for down drillhole assays reported by exploration companies	2 668 997 samples (from 96 717 drillhole collars)
Exploration geochemistry — surface samples (stream sed., soil, rock chips, Niton)	Sample and analytical data for geochemical surveys reported by exploration companies	773 817 samples
Field photographs	Embedded location and metadata for image bank	2500 photos
Field Observations	Geological observations, measurements and sample data by department geoscientists (Figure A2)	114 143 locations

Table A2. Data planned for implementation

Dataset	Description	Records
Seismic	Shot points, line coverage, SEGY and imagery from onshore and offshore seismic surveys (~2500 surveys for ~45 000 line km)	~45 000 line km (2500 surveys)
3D Geology	Vector data representing geological (geophysical) features below the Earth's surface in 3D space	
Petrophysics	Petrophysical properties of rocks (magnetic susceptibility, density, radioelements)	~5000 records
Whole rock geochemistry	Sample and analytical data for whole rock geochemical samples	~25 000 records
Palaeontology	Sample and descriptive data for fossil samples to aid stratigraphic and age discriminations	~65 000 records
Stable isotopes	Sample and analytical data for S, O and other stable isotopic systems	~3500 records
Polished blocks	Sample and descriptive data for polished blocks of mineral samples	~1775 records
Economic rocks	Sample and descriptive data for economic mineral samples	~30 000 records
Radiogenic isotopes: Pb–Pb	Sample and analytical data for Pb–Pb isotopic studies	~ 2000 records

The Geoscience Product Catalogue — a new way of finding GSNSW products online

Abstract

The Geological Survey of New South Wales is a state government geoscience agency that is responsible for supplying geoscience information and data in a range of product formats including books, journals, maps, DVDs, digital datasets and images. These products are promoted on the web and in printed catalogues and flyers. The Geoscience Product Catalogue and product database were designed to provide a comprehensive list of all products and a user-friendly spatial search tool enabling these products to be discovered, reviewed and purchased. This article documents the requirements and design decisions used to create the catalogue in order to address the issues of product promotion and maintaining an up to date product list.

Keywords: *product database, Geological Survey of New South Wales, information management, online catalogue, maps*

Introduction

The Geological Survey of NSW (GSNSW) is the state government's geoscience agency that provides information and advice to government, the mining and exploration industry and the community on the state's geology, mineral resources and land use planning.

The survey is responsible for supplying geoscience information and data in a range of product formats including books, journals, maps, DVDs, digital datasets and images. These products are promoted on the Division of Resources & Energy website and in printed catalogues and flyers. However, there is currently no single database or catalogue which has a complete record of all products. The Geoscience

Product Catalogue (GPC) and product database were designed to provide a comprehensive list of all GSNSW products and a user-friendly spatial search tool enabling these products to be discovered, reviewed and purchased.

This article will explain the design and implementation of the Geoscience Product Catalogue and the database that underpins it. It will document the business requirements and then describe in detail the functional and technical design of the application. Lastly, the article will outline some of the search methodologies available to search for and discover the rich vein of geoscience information that is maintained by the GSNSW.

AUTHOR

T.M. Moriarty

*Geological Survey of New South Wales,
516 High Street, Maitland, NSW 2320*

© State of New South Wales through Department of Trade and Investment, Regional Infrastructure and Services 2013.



Business requirements

— what it will do for customer service

With all software development projects it is critical, from the outset, to have a clear understanding of what the new application aims to achieve — both identify the product's features and understand the business processes it is being built to assist. These business requirements are then used to guide all subsequent decisions that are made throughout the design and build phases. This section outlines the business requirements for the GPC.

It is important that the GPC has a user-friendly, modern, intuitive interface. It will be used by the public, who have varying degrees of technical or domain knowledge, and internal staff with highly developed technical skills.

The GPC is intended to reduce, if not eliminate, the need for customers to seek assistance directly from the department regarding their information needs. Therefore, the GPC's search functionality is vitally important for its success. It needs to provide simple yet flexible tools that produce quality, relevant search results. A key feature to achieve this is the inclusion of the ability to search by location.

With the proliferation of map-based search tools on the internet, it is anticipated that many users are familiar with this type of searching. It is also a logical choice to use a map as the primary search tool, considering all of the department's products are inherently spatial and many requests for products are based on an area of interest.

The success of any search tool is determined by the quality of the data and the database model on which the search is applied. To support the GPC search interface a comprehensive database was constructed containing records of all product types, traditional and non-traditional. This also provided the opportunity to capture metadata on products that were previously not recorded, such as posters and presentations. These can now be systematically archived in the database.

The aim of the GPC is to improve the customer's experience by improving access to the department's publications and data products, including not only the discovery (search interface and database) but also the delivery of products. A further aim of the GPC is to streamline the ordering process — allowing the user to go straight from discovery to ordering products within the one interface.

Many of the department's data products were already available for download directly from the department's website and for viewing in Google Earth. These services have been maintained and the links are integrated within the GPC's design.

An additional requirement is to provide a mechanism for promoting the release of new products. The department's existing website is large and complex, containing large volumes of information. Its existing navigational structure, which reflected the department's organisational structure, did not assist the discovery of the new products and could make it difficult when updating the site with new content. The creation of the database allows for a simple query to automatically extract the list of new products and publish them on a dedicated 'New Products' page — providing a controlled workflow to manage new product releases and highlight them on the website.

It is also important that the system is robust and reliable with low ongoing maintenance costs.

Key features

— functional requirements

- Flexible spatial (map-based) search tool. The user will be able to select an area on a map and find all products in that area.
- Text (title/keyword/author) search tool — a very common search method.
- Comprehensive listing of all products — a complete suite of metadata for each product and the product edition.
- Relational database design allows different generations of products (i.e. editions) to link different types of product (for example, a map and its explanatory notes) and allows the categorisation of products by theme. It also relates products to the events where they were released/promoted/presented (particularly relevant for posters and presentations).
- Visual display of product coverage on a base map
- Ability to review product details and preview an image of the product (particularly useful for map products)
- Ability to review all products listed alphabetically by title
- Ability to see all new products
- Online self-serve facility for clients looking for geoscience information. Provide alternative displays of data (e.g. hierarchical listing by type and scale as is currently available on the website).

Benefits to external users

- A single comprehensive, searchable database for all products.
- A choice of different search methods, including combining and refining search criteria.
- The ability to search by location (not previously available despite the inherent spatial nature of the products).
- An immediate response to a query, allowing the user to conduct a direct search of the product database.
- 24/7 access to the system.
- An increase in the number and visibility of products that are promoted online.
- The ability to visualise the extent of NSW which the product relates to.
- The ability to preview maps before ordering.
- Easy access to lists of related products with links between Explanatory Notes and their maps, available at the preview screen.

Benefits to internal users

- Many of the benefits experienced by the external user also apply to internal users, but there are added benefits to managers and custodians of the data products. These include:
- A consistent description for each product
- The capturing of metadata for non-traditional products such as scientific posters and presentations
- A controlled workflow for publishing new product information on the web
- A record of what products were produced for, or displayed at, particular events
- A reduction in web content writing and editing, simplifying the promotion of products online
- Being able to respond more quickly to client enquiries
- Making older publications more visible
- Enabling internal analysis of product usage, determining most popular products for reprinting, identifying changing trends in the type of product formats preferred by users.

Conceptual design — what it looks like

The aim is to cater for the increasing sophistication of users, who are familiar with searching for purchasing information and products online, and who use maps in range of web-based activities, and to do this whilst adhering to the department's style for its websites.

There is a deliberate focus on the map search interface to encourage the use of a spatial search (Figure 1). This reflects the move away from the traditional metadata search engines that rely on entering text into the search fields. Instead a spatial search allows the user to specify their area of interest by drawing a shape on a map.

Users can access the GPC from a range of devices such as tablets and smart phones (not just desktops), so it is important to provide the ability to refine the search results interactively with minimal typing. Drop-down menus are used to provide filtering options.

The implementation of filters allows the system to cater for users with different search goals. For example, the user may be only looking for traditional printed products or, alternatively, only looking for digital vector data. Providing different search options, allows the user to conduct the most appropriate search based on their skills and search parameters, from simple one-step searches to combined options for more advanced searches. For example, users could choose options such as 'Show me all products that cover the region around Cobar' or 'Show me only geology available as digital vector data in Esri® shapefile format for Cobar'.



Figure 1. System search interface.

Technical design — how it was built

The GPC search external interface uses the corporate content management system Squiz Matrix®, in order to create a maintainable system without additional resources. Google Maps™ is used for the base map of the spatial search interface and PostGIS for the data at the Squiz Matrix® end. The internal database was built into the existing GSNSW geoscientific data warehouse, which employs the GeoBank® SQL database product by Micromine®. Figure 2 outlines the system architecture and Table 1 summarises the systems used to create the GPC and its database.

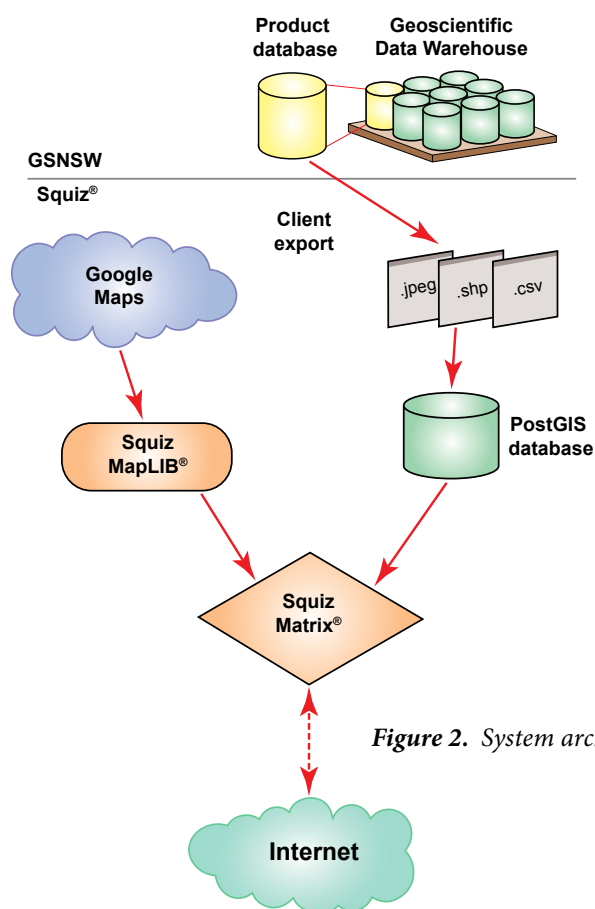


Figure 2. System architecture.

Table 1. System components

System name	Data type	Use
PostGIS	SQL	The main product catalogue, containing: <ul style="list-style-type: none"> All product information (e.g. product id, edition id, name, theme, etc.) GIS boundary data for each product coverage area (extent) GIS centroid data for each product Catalogue indexes for use in performing keyword searches
Squiz MapLIB	Javascript™	A client-side Javascript library designed to assist with the manipulation and rendering of map elements.
Google Maps™	SVG	A service providing the background map layer

Search methodologies — how to use the GPC

Multiple search methodologies are available to cater for a range of search parameters and summarised below.

1. Search by product title
2. Search by product abstract (keyword)
3. Search by author
4. Filter product searches by theme, type or format using the filter menu
5. Search by location by entering locality and distance
6. Search by latitude and longitude
7. Search by map sheet number
8. Search using an A–Z product index
9. Search by defining an area on the map using the polygon draw tool. Desktop and tablet devices provide a rectangle, a multi-sided polygon and a circle drawing tool.

These searches can also be combined. For example, an area can be selected on the map using the rectangle tool and then a title search applied to that area only.

Previewing the product — what you will see

Search results are displayed as pins on a map (marking the centroid of the product's extent) and in tabular form as a list below the map (Figure 3). The user can then select a pin on the map to reveal the product's name, scale, publication date and extent. More detailed information, including a thumbnail image of the product, can be accessed from either the pins or the list below the map.

Different symbols are used at different zoom levels to control the display of a large number of pins or pins that are coincidental. First level pins are used to display the location of products. At a small scale (zoomed out), coincident pins are clustered together and displayed as a coloured circle (except in Internet Explorer® 8 where it is a square), with the number of products shown.

When zooming in, the symbols change from the circles/squares to pins. The cursor can be used to rollover a pin to display the title. After selecting (clicking on) a pin, more information such as title, product type and scale is displayed in a popup balloon. Further details can be accessed by selecting the 'View Details' button in the popup balloon. Once selected, a new window opens displaying all relevant metadata including catalogue number, publication date, edition and references as well as a thumbnail image of the product if available. For many of the mapping products, the new window includes links to Google Earth™ and Google Maps™ (to view the maps) and links to download the GIS and image files.

The search results also list all products in a table below the map (Figure 3). These products can be sorted by title, edition, type, scale, publication date and cost.

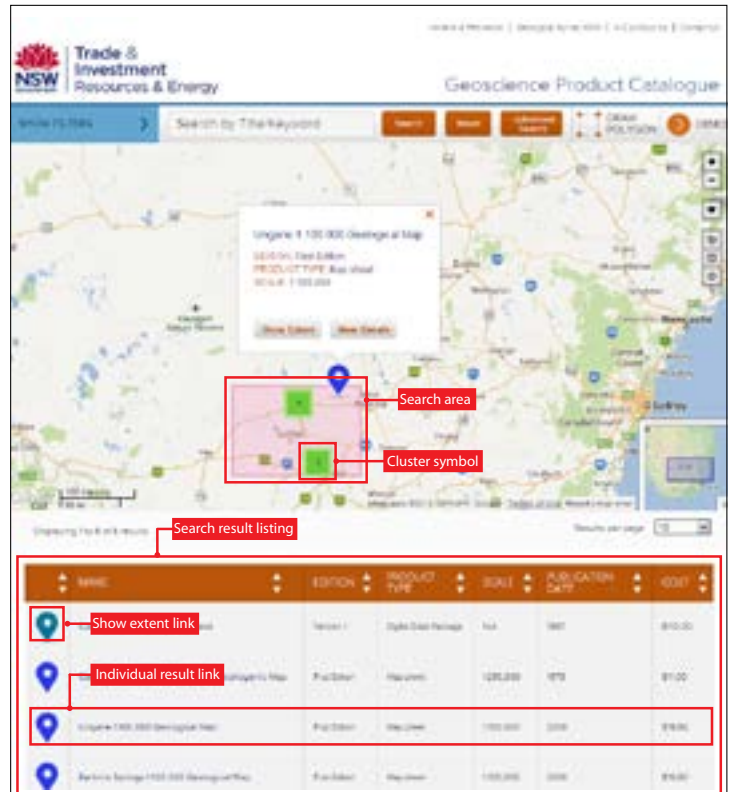


Figure 3. Search results displayed as pins on the map and as a list below the map.

Acknowledgements

The author would like to acknowledge the work of Scott Amort (technical leader/designer) and Ross McDowell and Lukas Bower (project managers) at Squiz; and Lynne Shields (user manual and demonstration video) at Fordcom. The following staff also contributed: Phil Carter (data custodian, system design), David Collins (database design), Catherine Shirley (database administrator), Joan Henley (content editor) and Cassie Bambach, Tammy Dennis and Eliza Wiberly (data collators).

FREE NSW DATA



Surface geology
Ternary radioelement potassium(K) /thorium(Th)/uranium(U) channels
Pseudocolour Total Magnetic Intensity (TMI)
First Vertical Derivative of Total Magnetic Intensity (1VD TMI)
200 more maps coming soon

using mobile devices...

For more information go to

<http://dwh.minerals.nsw.gov.au/CI/warehouse>

DIGS (Digital Imaging of Geological Systems) allows users to locate exploration data from company reports, Geological Survey reports, publications, cartographic plans and geological maps.

EROL (Exploration & Environmental Reports Online Lodgement) is an extension to DIGS® and enables exploration title holders and operators to lodge reports via the internet.

MIN VIEW (Exploration Tenement & Geoscience Data Viewer) allows users to spatially display, query and download geological and exploration titles information. It provides links to reports in DIGS® to access drillhole data, mineral deposit information and exploration titles.

GDW (Geoscientific Data Warehouse) provides access to spatial and non-spatial geoscientific relational data held by Resources & Energy. Drillhole, assay and geochronological data, mineral deposits, quarry locations and field observations can be downloaded as CSV and GIS formats.

GPC (Geoscience Product Catalogue) provides access to a database of the Geological Survey's retail products and free downloads using a keyword and a modern spatial search with filters. Items available include maps, books, periodicals, digital data packages and geophysical images and data.



enquiries

Enquiries:
geoinfo.admin@industry.nsw.gov.au

General information and access: <http://dwh.minerals.nsw.gov.au/CI/warehouse>

Feedback & help:
<http://dwh.minerals.nsw.gov.au/CI/warehouse/entry/feedback>



Trade &
Investment
Resources & Energy

NSW Trade & Investment, Division of Resources & Energy
516 High Street, Maitland NSW 2320
PO Box 344 Hunter Region Mail Centre NSW 2310
T:1300 736 122 T: (02) 49316666

www.resources.nsw.gov.au