Towards national geoscience data standards

Rod Ryburn[1] & Ian O'Donnell[1]

Geoscience data standards as a field of research may seem like a dull peripheral issue of little relevance to the domain of many geoscientists. However, the subject is gaining rapidly in importance as the information revolution takes hold, because ultimately billions of dollars worth of information are at stake. In this article we take a look at what has happened recently in this field, where we think it is heading, and AGSO's role in establishing national data standards.

Petroleum exploration data

The global petroleum exploration and production industry has established a clear lead in the quest for geoscience data standards. About 10 years ago, a contraction of profits in the oil industry forced companies to look carefully at their data-handling efficiency. They found that incompatibilities between software packages and data structures were major inefficiencies. Whereas most professionals were spending ~60 per cent of their time locating, sifting, 'massaging', and transcribing data, many others were being employed just to write programs to transcribe data from one 'stand-alone' system to another.

The magnitude of the industry-wide inefficiencies led to the formation in 1990 of a non-profit organisation — POSC (Petrotechnical Open Software Consortium, see *www.posc.org*), based in the United States and Europe. The original members (BP, Mobil, Elf, Texaco, and Chevron) head a list that has since grown substantially; AGSO, for example, has subscribed for the last 6 years. POSC has established a series of specifications, standards, and logical data models, including both object and relational data models. The POSC data models do not include important geoscience topics, such as biostratigraphy and organic geochemistry, but there are plans to expand their scope. Although these data models are complex, and are fully understood only by the largest companies and petrotechnical software houses, POSC claims that they already result in a saving of US$1–3 per barrel of oil. This represents a saving of about a billion dollars on the cost of production from the recent oil discoveries on Australia's North West Shelf.

Another similar consortium to which AGSO subscribes is PPDM (Public Petroleum Data Model Association, see *www.ppdm.org*), based in Calgary, Canada. PPDM's data models are more concrete than POSC's big-picture logical models, but the two organisations are not in competition.

Mineral exploration data

The mineral exploration industry lags a long way behind the example set by its petroleum counterpart. Nevertheless, some steps have been taken to formulate mineral data standards. For years, GGIPAC (Government Geoscience Information Policy Advisory Committee; formerly GGDPAC), which advises the Australian Chief Government Geologists (CGG), has been grappling with the problem, particularly the standards for reporting mineral exploration data to State governments. The problem is tripartite:

- **Legacy data.** 'Hard-copy' exploration reports, maps, and sections submitted to State mines departments must be captured as electronic bit-images, and appropriate metadatabase and storage facilities established so that the images can be distributed online. The Department of Mineral Resources (NSW) has shown the way here via its successful DIGS ('Digital imaging geological survey system') project (see *www.slnsw.gov.au/ILANET/clients/mineral_resources/about/gsgenerl.htm#digs*), which has invested millions of dollars electronically capturing over two billion dollars worth of exploration data.
- **Future reports.** The next step is to ensure that future reports are submitted in computer-readable form for text-searchable online access. The initial recommendation to GGDPAC that SGML (Standard Graphics Markup Language) be used (see draft guidelines at *www.dme.nt.gov.au/library/ggdpac96.html#INTR*) was rejected by NSW. A possible solution is to allow exploration companies to submit reports as word-processed documents suitable for translation to PDF format for web viewing. The State authorities would still need to maintain a proper metadatabase of reports, and large maps, etc., would require scanning as images.
- **Hard data.** The greatest difficulty is maximising the usefulness of all the hard data consigned to maps, tables, and diagrams in exploration reports and their appendices — i.e., the valuable data concerned with sample locations, drillholes, core logs, lithologies, chemical analyses, geophysical surveys and many other aspects of mineral exploration. To this end, these data must be placed in a proper information management system — i.e., a standardised database system — and submitted in this way by the exploration companies. Simpler schemes, such as spreadsheets and ASCII files, can never provide the required degree of standardisation and data integrity. Unfortunately, an agreed standard database system to match this requirement does not yet exist.

The AMIRA geoscience data model

As long ago as 1992, State mines departments, via GGDPAC, recognised that standard geoscience data models were needed across the mineral exploration industry, and arranged for AMIRA (Australian Mineral Industries Research Association, see *www.amira.com.au*) to manage project P431 — 'The geoscience data model'. The Australasian Spatial Data Exchange Centre was contracted to formulate a data model. Major sponsors were the State governments and AGSO. Disappointingly, only two industry sponsors, BHP Exploration and North Exploration, came forward. The themes that were included in the model were geology, drilling, geochemistry, and mineral resource/reserve.

The model, which was delivered in April 1998, reflects a great deal of consultation, including input canvassed from existing database models. It consists of entity-relationship diagrams (cf. Fig. 25) and a data dictionary for the agreed themes. Several practicality tests were applied to it, including a Microsoft Access database covering the Broken Hill region constructed from data supplied by government agencies. These were successful, and proved the feasibility of transferring existing databases to the new model.

The project has since been criticised for not delivering a working solution to the problem of company data submission. This was never on the agenda, and is a far more complex task that requires a much higher order of commitment — probably several million dollars worth, in terms of software development. At the same time, the AMIRA model, useful as it already is, needs to be expanded to cover a wider range of geoscience data.

Where to now?

For mineral exploration data standards to advance, resources must first be pooled — as in the oil industry under POSC. The importance of such a pooling of resources was acknowledged at the 'Gold round table' conference, held in Canberra in February 1998. In a joint paper prepared for submission to the Australia & New Zealand Mineral Exploration Council in July 1988, the Western Australian Mines Department and the CGG Conference estimated that about $50 million would be needed from the States and Commonwealth to capture electronically the $30 billion of exploration data held in Australia. If each State attempts to 'go it alone', the ultimate cost will be far higher, and the opportunity to standardise will be lost. With its excellent system of reporting exploration results to government, Australia should lead the world in setting geoscience data standards. We stand to reap rich rewards if we can solve this issue. The technical aspects are not difficult. However, industry (as opposed to government) needs to be much more involved in these issues than it presently is.

Notwithstanding their government-reporting obligations, the mineral exploration industry would benefit from the adoption of

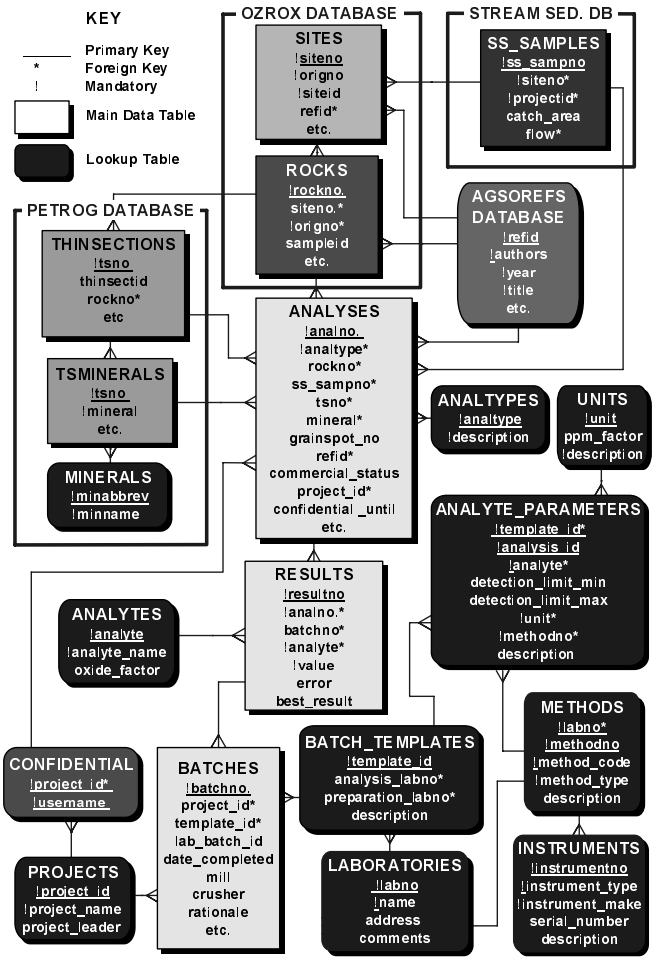

Fig. 25. A table-relationship diagram of the OZCHEM database together with the relevant parts of related databases. The rounded boxes are lookup tables. 'Crows feet' indicate multiple end-of-one-to-many table linkages.

standard data models and software by eliminating the time wasted by employees dealing with non-standard data — an inefficiency probably of a similar magnitude to that reported by the 'pre-standards' oil industry. No one company or government organisation can solve this problem on its own. We have to cooperate to compete!

AGSO's role in data standards

Although AGSO has no statutory role in collecting mineral exploration data, it has responsibility for national geoscience datasets. Through its Spatial Information & Mapping Services, AGSO takes a keen interest in all national geoscience standards, and maintains many national standard databases and authority tables. The 'National stratigraphic names' database (see *www.agso.gov.au/information/structure/isd/database/stratnames.html*) is an example of a standards database, while our geological timescale (GEOTIME) is representative of our numerous authority tables (*see www.agso.gov.au/information/structure/isd/database/lookups.html*). Most of our standard databases and lookup tables are now visible on our worldwide web site (*www.agso.gov.au*). The development of classification schemes and 'domains' (lists of all possible values an attribute can assume) are as much a part of national standards as are the logical data models and data dictionaries.

In addition to promoting geoscience data standardisation for the benefit of petroleum and mineral exploration, AGSO also addresses data standards for land management, marine science, the environment, and geohazards. Our recently compiled GWATER database is an example of our attempts to adopt AGSO standards in groundwater and water quality, and AGSO and the States are currently involved in a major effort to define national standards there. Databases that AGSO has compiled for earthquakes, landslides, and tsunamis represent the beginnings of national geohazards data standards. We also support metadata standards, the most important being the Australia New Zealand Land Information Council spatial metadata standard.

The impact of good data models

The impact of good data models is best explained with an example from AGSO's stable. For many years, AGSO maintained several diverse laboratory analytical databases typically consisting of wide tables corresponding to various analytical instruments and techniques. Thus, trace-element geochemistry was recorded in over 60 columns, one for each element, and organic geochemistry had over 700 different columns for each analyte. For every newly added method or analyte, new columns had to be added to the tables, or even whole new tables built. We recently restructured some of these databases into the OZCHEM database (Fig. 25).

OZCHEM is a fully 'normalised' database in which a long 'skinny' table called RESULTS substitutes for several previous 'fat' tables. For each sample analysed, RESULTS has many rows — usually one for each element or analyte processed. The identity of each analyte, the numerical analytical result, and pointers to other tables with errors, methods, unit, etc., are entered in RESULTS. This structure allows for multiple determinations of the one analyte by various laboratories. Also, any number of new analytes and methods can be added without having to change the database's structure. With the help of the latest database software, OZCHEM performs well, and attractive user interfaces are easily built.

OZCHEM can now manage whole-rock chemistry, stream-sediment geochemistry, and microprobe analyses of minerals and rocks. In theory it could be made to handle isotopes, organic geochemistry, and water chemistry as well — a universal geochemical database — but we have not integrated all these disparate analytical disciplines (and maintain separate analytical databases for groundwater and organic geochemistry).

From a corporate perspective, the efficiencies to be gained by pooling all analytical results in the one generalised, analytical database are spectacular, and an example of how good generalised database models can save a lot of time and effort.

Conclusions

- The global petroleum exploration industry has shown the way to uniform data standards by pooling resources in their POSDC and PPDM non-profit organisations.
- The mineral exploration industry needs to do the same. It must cooperate to compete.
- Australia is in a good position to lead the world in mineral exploration data standards, but both industry and government need to be involved.
- Australian States must standardise their mineral exploration reporting requirements.
- AGSO has a strong role in promoting national geoscience data standards.
- Good data models save time and money. They establish the *lingua franca* of geoscience.

Acknowledgments

Thanks to Murray Hazell for the diagram of the OZCHEM database, and to John Bradshaw, Murray Hazell, and Frank Brassil for reviewing the manuscript.

[1] Spatial Information & Mapping Services, Science & Survey Support Division, Australian Geological Survey Organisation, PO Box 378, Canberra, ACT 2601; tel. +61 2 6249 9605 (RR), +61 2 6249 9100 (IO); fax +61 2 6249 9942 (RR), +61 2 6249 9984 (IO); email Rod.Ryburn@agso.gov.au, Ian.Odonnell@agso.gov.au.