BIG DATA, SMALL LIBRARY

ABSTRACT

Introduction: Shell Australia's Technical Librarians successfully combine the roles of the traditional special librarian and the new data librarian. They are a vital part of Shell Australia's multidisciplinary Subsurface and Wells Technical Data Management team, working collaboratively with colleagues across data management, geoscience and information technology (IT) disciplines to ensure the increasing volume, velocity, and variety of the company's geoscientific technical data – very big data - is managed efficiently.

Method: Shell Australia's Technical Librarians were key stakeholders in a recent project to review and improve the existing databases and processes used to manage Shell Australia's geophysical data. Geophysical data is challenging to manage; it is complex "big" data and defies traditional cataloguing. Shell Australia holds petabytes of geophysical data in a variety of file formats, on a wide variety of media, in many versions - from raw acquisition data, through to processed and interpretive data. During the project the Technical Librarians worked closely with geophysicists, geophysical data analysts, IT support specialists and database architects. Their contributions as library and information science (LIS) professionals included providing advice on required metadata fields, developing controlled vocabularies and naming conventions, defining required search parameters, identifying opportunities for added functionalities,

undertaking database testing and providing feedback, populating the database, and developing workflow procedures. The Technical Librarians also brought a long-term perspective to the review of data management tools and processes, essential for the preservation of Shell Australia's valuable geophysical data. Engagement with colleagues outside the LIS profession provided the Technical Librarians with an opportunity to learn more about geophysical data and its lifecycle.

Results: This cross-disciplinary engagement resulted in the implementation of tools and processes that provide improved metadata capture, clearer connections between projects and data, improved search functionality, better data management, and stronger relationships with stakeholders.

Conclusion: The professional skills of the technical librarian remain relevant and valuable in an era of big data, however cross-disciplinary collaboration with stakeholders is essential to communicate this value and develop additional data management skills.

Relevance statement: Engage – this presentation will demonstrate the value of working with colleagues outside the library and information profession in the sphere of data management.

PAPER

INTRODUCTION

Shell Australia's librarians were key stakeholders in a recent project to create a metadata database to manage Shell Australia's geophysical data. Geophysical data is challenging to manage; it is complex "big" data and defies traditional cataloguing. Shell Australia holds petabytes of geophysical data in a variety of file formats, on a wide variety of media, under a range of licensing conditions - from raw acquisition data, through to processed and interpretive data. During the project the librarians worked collaboratively with geophysicists, geophysical data analysts, information technology (IT) specialists and a database developer. Their contributions as library and information science (LIS) professionals included providing advice on required metadata fields, developing controlled vocabularies and naming conventions, defining required search parameters, identifying opportunities for added functionalities, undertaking database testing and providing feedback, importing metadata, developing user guidelines and providing training. The librarians also brought a long-term perspective to the review of data management tools and processes, essential for the preservation of Shell Australia's valuable geophysical data. Engagement with colleagues outside the LIS profession provided the librarians with an opportunity to learn more about geophysical data and its lifecycle.

BACKGROUND AND CONTEXT

Geoscience information management as a discipline with a role for librarians

The use of computers to generate and manipulate geoscience data increased during the mid-1960s and geoscience information professionals grappled with issues arising from managing the geoscience data explosion. Geoscience information established a foothold as a separate discipline with the creation of the Geoscience Information Society in the US in 1966, and the discipline achieved international recognition in 1978 at the *1st International Conference on Geological Information* in London (Ward & Walker, 1986). These early forums recognised the need for cooperation between geoscience data and information management stakeholders (Harvey & Diment, 1979). From the 1970s through to the early 1990s there was strong professional engagement by LIS professionals in geoscience data and information management: in Australia Tellis and colleagues published prolifically, detailing their role in the development of the Australian Earth Sciences Information System (AESIS) reference database (Parkin & Tellis, 1977a, 1977b; Tellis & Crowe, 1982; Tellis & Gerdes, 1986); Gerdes and Smith developed the *Guideline for geoscience bibliographical databases* (1990); and Bruce (1991) provided a detailed account of the involvement of the LIS professionals in the creation of a relational database at the Western Mining Corporation to manage a wide range of petroleum exploration data.

Since the mid-2000s there has been a surge in interest around digital research data management by the broader library and scientific community and discussion about how the “data deluge” (Lyon, 2007, p. 5) should be managed. A report to the National Science Foundation (Association of Research Libraries Workshop on New Collaborative

Relationships, 2006) in the US identified partnerships between scientists and information professionals as a critical component of data management. The report acknowledged Hey and Hey's (2006, p. 516) observation that "this new data dominant era brings new challenges for the scientists and they will need the skills and technologies of both computer scientists and of the library community to manage, search and curate these new data resources".

In the Australian open data environment there has been strong support by the Australian National Data Service (ANDS) in assisting LIS professionals in their efforts to provide repository services and data management training and support to researchers (Australian National Data Service, 2016b; Treloar, 2009). Through Research Data Australia, ANDS provides a gateway to the data curation efforts of LIS professionals: examples from the geoscience domain include the Curtin Digital Mineral Major Open Data Collection (Australian National Data Service, 2016a) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Data Access Portal (Cook, 2014) – in both cases LIS professionals engaging with IT specialists and data stakeholders to develop metadata solutions to maximize data access.

Despite the enthusiasm for collaboration shown in the literature, in reality LIS professionals are often not considered when developing data management solutions: at the Massachusetts Institute of Technology "some researchers are sceptical that the Libraries can help with their expensive/large/hard data needs and question why the Libraries would want to engage in that arena in the first place" (Soehner, Steeves, &

Ward, 2010). In the mining sector there is a perception that IT professionals are responsible for data management issues (Yacopetti & Mundell, 2010) and a survey of geological surveys in the US reveals that only one-third employ an information professional – “among geologists, hiring a professional librarian remains an option” (Foote, 2010, p. 63).

A survey of Western Australian geoscience information professionals showed considerable input into the management of non-digital geoscience data, but less involvement with digital data: there was no engagement by information professionals in the management of digital geophysical (seismic) data (Johnson, 2011, fig. 6.6 & 6.7).

Geophysical database development

Anderson (2006) identifies some of the challenges encountered in developing a metadata framework to accurately describe scientific and technical data: the enormous size and volume of data; complexity of origin, acquisition parameters and versions; tracking of applications needed to access and use the data; diversity of data types, formats, media and standards. All of these challenges apply to describing geophysical data. Progress has been made with broad acceptance of international geospatial metadata standards found in the ISO suite of standards 191** (The Federal Geographic Data Committee, 2014), but this is only a very small piece of the metadata required to describe geophysical data.

Some discussion of geophysical database development–metadata and architecture–can be found in the collaborative domain occupied by organisations such as Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Diviacco et al., 2015) and Geo-Seas (Diviacco, Lowry, & Schaap, 2012). The latter recognised the need to balance community and technological issues to develop metadata layers and define "a metadata parameter list that makes sense yet is light enough for project partners…who do not specialize in data management" (Diviacco et al., 2012, p. 247).

Information about geophysical database development in the petroleum exploration and production (E&P) sector is tightly guarded to maintain competitive advantage. In a geophysical database patent held by Halliburton (Rao, Coe, Gibson, Gleitman, & Rector, 2015, col. 3), one of the world's largest providers of products and services to the E&P industry:

> the current models for managing and distributing collected geophysical data are not in and of themselves efficient… data often resides in multiple locations. The exact location and ownership of the data may be uncertain. As a result of transmission, storage and processing, the quality of the data may be questionable. Data sets may be so large that it becomes inefficient if not impossible to sort them, identify their contents and distribute them to potential new users...there exists huge quantities of costly, collected geophysical data having potentially significant value to various users above and beyond those who sponsored the initial collections. However, significant problems exist with organizing, identifying, maintaining the integrity of and distributing such data

The E&P industry acknowledge the importance of managing their exploration data and information, but struggle to develop and deploy effective solutions. A 2015 survey of 1328 organizations engaged in exploration activities showed that only 24% had deployed a solution to manage their geophysical data assets and only 8% use a commercially available product (Geosoft Inc., 2016).

The Shell Australia Technical Library

There are many types of work units engaged in geoscience data management, including specialist information services staffed predominantly by librarians, multi-disciplinary teams and data management units that contain no information professionals (Johnson, 2011). Shell Australia's two librarians are part of a multi-disciplinary team (the Team), working alongside geophysical (seismic) data analysts, geological (well) data analysts, IT specialists, project managers and workflows and applications specialists. The Library had its beginning in Melbourne in the 1940s and has been through many changes: relocated, downsized, archived, revived, migrated and repatriated. The information and data collections managed by the Library comprise technical reports, well data, well logs, seismic data, geological samples, reference items and digital files on disk (see Figure 1). The collections encompass a wide range of digital and analog formats, on a wide variety of media. The librarians take the long view of managing the technical data and information collections under their care, although the utopian ideal of data immortality, as discussed by Baker and Yarmey (2009, p. 14) sometimes seems a long way from commercial reality:

An implicit assumption—perhaps a goal of the information age—is that responsible data stewardship will ensure data immortality, a preserved measurement, dataset or data collection will, through complete and accurate contextual description, be useful and usable far into the future. It is an ambitious goal of mythic proportions.

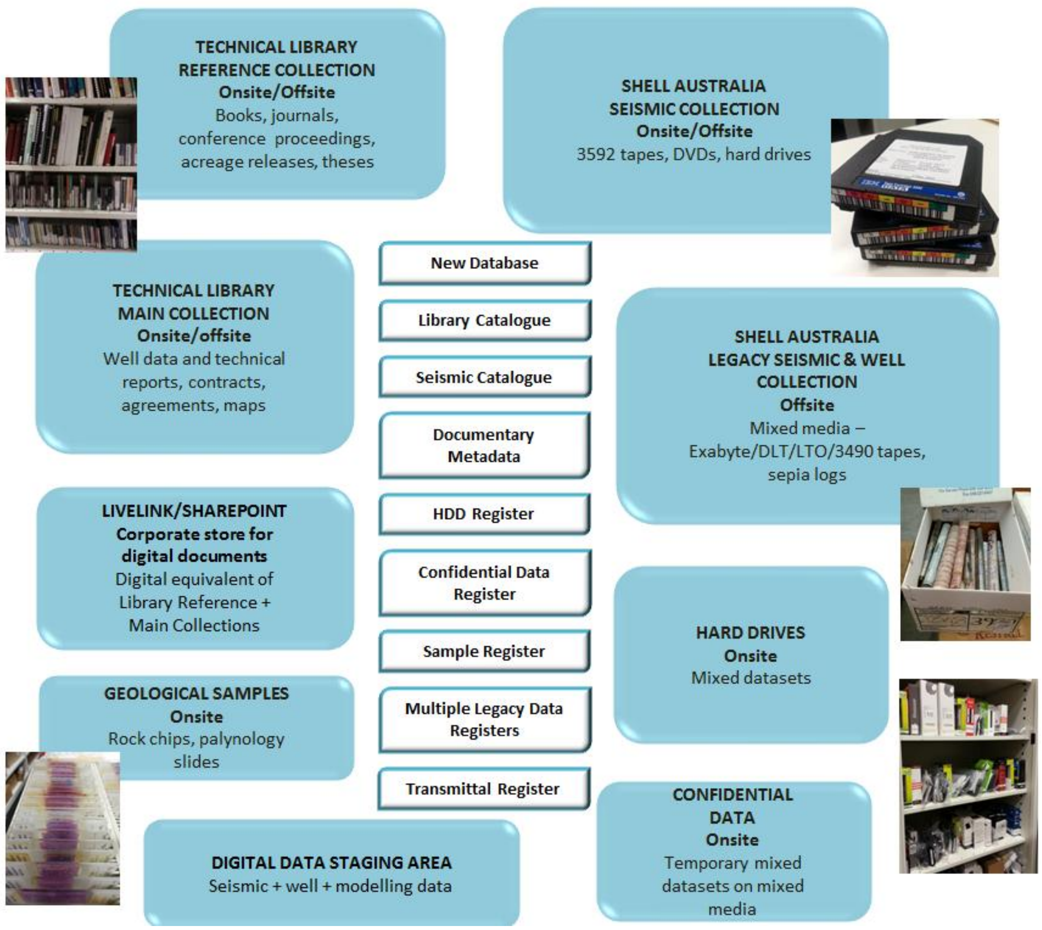


Figure 1: Shell Australia Technical Library - collections and metadata tools

The Library does not "own" the collections, it has a custodial role: data and information is generated by or purchased by the Business (e.g. the Exploration Department), and the Business retains ownership. The production and acquisition of geoscience data is a complex, time consuming and expensive task, and some data is irreplaceable (Cutler & Maples, 2002; Jones, 1986). Digital data are particularly fragile and vulnerable to loss through inadequate storage, technological change, poor metadata and deficient discovery tools (Joseph, 2008). The main role of the Library is to preserve and provide access to this valuable data and information, since older data and information can be as valuable to the modern geoscientist (Pruett, 1982). Geological and geophysical data analysts within the Team are responsible for quality control of incoming data and loading data to corporate data stores and applications. The collaboration within the Team to effectively manage Shell Australia's holdings of geoscience data and information facilitates sound decision making by the Business and compliance with regulatory requirements.

The offshore petroleum and gas exploration process

Petroleum and gas exploration and production in Australia is controlled by numerous pieces of state and federal legislation (National Offshore Petroleum Administrator, n.d.-b), most prominently the Commonwealth Government's *Offshore Petroleum and Greenhouse Gas Storage Act 2006* (Office of Parliamentary Counsel, n.d.).

Data management activities peak at three main points in the offshore petroleum and gas exploration cycle (see Figure 2). There are many permutations of this cycle, but for simplicity the example of the release of vacant acreage for exploration will be used.

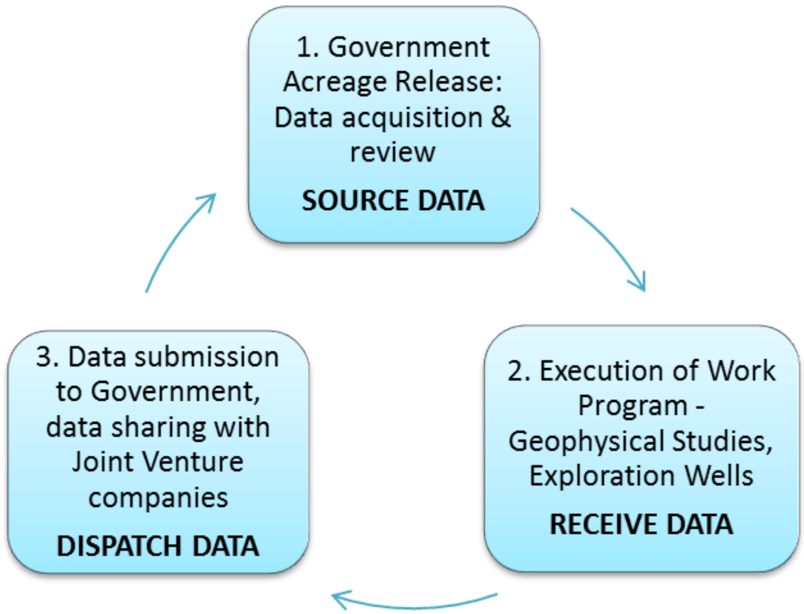


Figure 2: Peak data management demand in the Australian offshore petroleum and gas exploration cycle

1. The Australian Government encourages investment in petroleum exploration through annual Offshore Petroleum Exploration Acreage Releases (Geoscience Australia, 2015). At this stage the Library is involved in sourcing data relating to previous exploration activities from external sources–vendor and government–and in-house collections, enabling geoscientists to scope areas of interest. Once a favourable area is identified, an application for an Exploration License is submitted to government with a proposed Work Program (see Figure 3). It is not uncommon for companies to form joint ventures with other oil and gas companies to spread the risk when bidding for acreage
2. After a company has been awarded an Exploration License over a permit area they are required to complete the activities outlined in the Work Program. These

activities generate large amounts of data and technical reports which are accessioned by the Library, who in turn notify stakeholders of data availability

3. During exploration activities companies share data with joint venture partners, and on completion of the Work Program activities the permit operating company is required to submit data and reports to government in accordance with *Offshore Petroleum and Greenhouse Gas Storage (Resource Management and Administration) Regulations 2011* (National Offshore Petroleum Administrator, n.d.-a). After a period of confidentiality–dependent on the type of License granted and the type of data–these reports are made public. This “open file” data is then available from government sources such as Geoscience Australia, and is an important source of data and information used in exploration activities. The Library plays a key role in the stewardship of outgoing datasets

Year of Term of Permit	Permit Year Starts	Permit Year Ends	Minimum Work Requirements	Estimated Expenditure Constant Dollars (indicative only) $A
1	16/09/2011	15/09/2012	750km^2 New 3D Seismic Survey Geotechnical Studies	4,750,000
2	16/09/2012	15/09/2013	Geotechnical Studies	250,000
3	16/09/2013	15/09/2014	One (1) Exploration Well Geotechnical Studies	30,250,000
4	16/09/2014	15/09/2015	Geotechnical Studies	200,000
5	16/09/2015	15/09/2016	Geotechnical Studies	200,000

Figure 3: WA-365-P Title Instrument - Work Program (National Offshore Petroleum Administrator, n.d.-c)

As you can see from the Work Program in Figure 3, the exploration process is expensive and the costs shown are typical–A$5 million to conduct a seismic survey and A$30 million to drill a well, just to see what lies beneath–around 20% of wells provide results which can be classified as "successful" (Department of Mines and Petroleum, Petroleum Division, 2010).

Geophysical data

One of the most common types of geophysical data handled by the Shell Australia Technical Library is generated by 3D marine seismic surveys. These surveys use compressed air guns to generate low frequency, long wavelength sound waves which enable geoscientists to create images and models of geological structures many kilometres below the earth's surface (like an ultrasound or MRI image of your body). In an offshore environment, seismic acquisition vessels tow between 8 and 16 recording cables around 6km in length, at 50-100 metres apart (Wilkinson, 2006, Chapter 2) (see Figure 4).

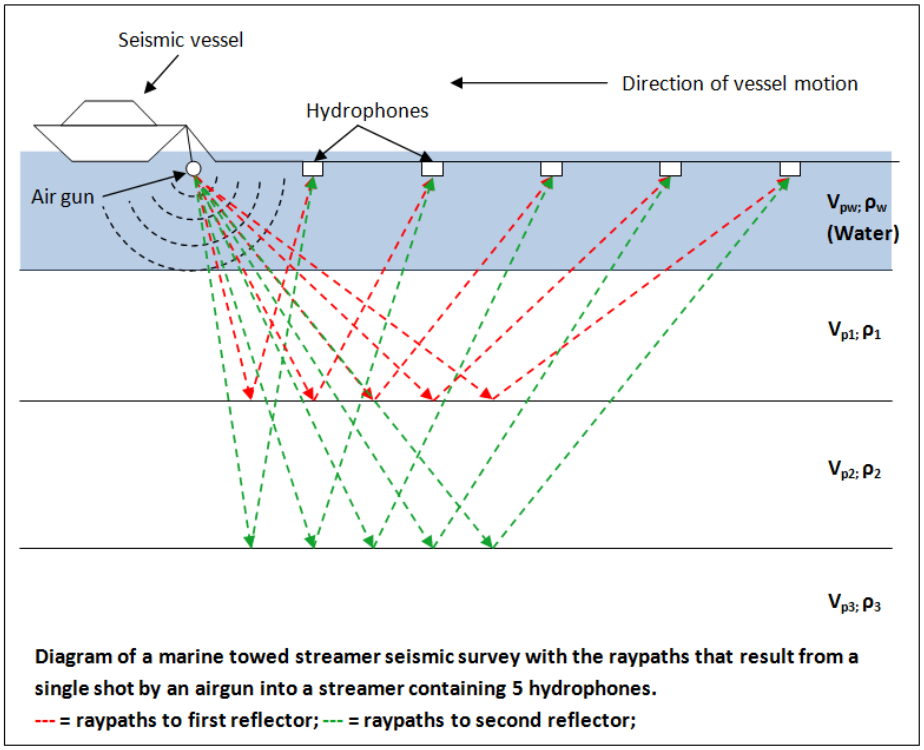


Figure 4: Marine seismic survey using a towed streamer ("Reflection seismology," 2016)

Depending on the size of the area, survey acquisition can take several months. The raw field data is processed and then interpreted by geophysical interpreters to determine potential reservoirs suitable for further geological investigation – specifically the drilling of exploration wells (see Figure 5). The data generated by survey acquisition-processing-interpretation typically arrives in batches over a period of up to 18 months on an assortment of 3592 magnetic tapes, hard drives, USB thumbdrives, DVDs, electronic file transfers and occasionally even paper reports. Multiple copies of datasets are received – one set for the operating company and additional sets to share with joint venture partners and for submission to government. The size of seismic datasets presents a data management challenge – it is truly big data. As an example, the raw field data alone for the Ceduna 3D Marine Seismic Survey totals 223.5 terabytes

(Cronin, 2016). As a rough estimate, Shell Australia's geophysical collections–on physical media–totals approximately 6 petabytes: to give a sense of the scale of the collection, the total size of the National Library of Australia's digital collections is around 3.8 petabytes (National Library of Australia, n.d.).

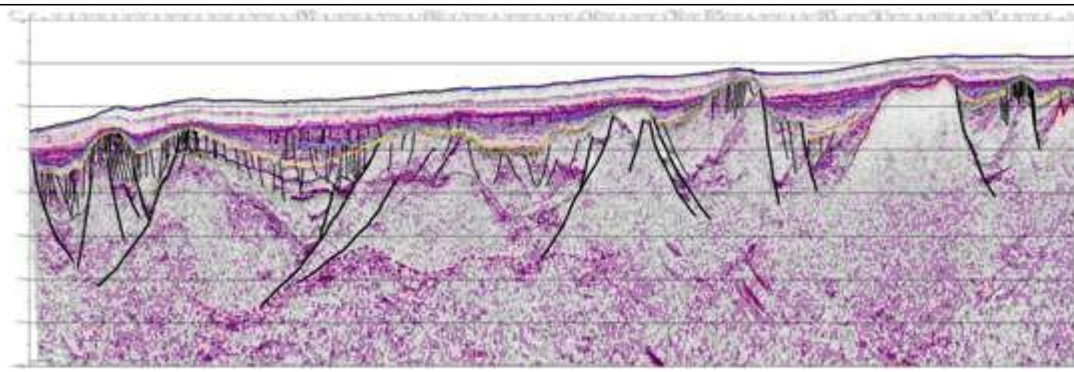


Figure 5: Interpreted Seismic Section (Geoscience Australia, 2016)

METHOD

A survey of Shell Australia's geoscientific stakeholders identified an opportunity to improve geophysical data management. The Geophysical Data Remediation Project was initiated, and part of this Project was the implementation of a new metadata tool. The metadata tool was required to meet requirements of cost, confidentiality, control and customization. In initial Project discussions it was assessed that no vendor off-the-shelf product could satisfy all four of these requirements. Shell prides itself on being an innovation company (Shell, 2016) so the decision to leverage in-house expertise to develop a bespoke metadata tool was a natural one.

The Project went through a series of five development phases.

1. Project Proposal – Metadata Database

In order to secure the necessary resources to create and implement the Database the Team's project manager developed a project proposal. This included elements such as scope, stakeholders, objectives, key activities, factors critical to success, deliverables, benefits and measures.

2. Stakeholder Workshops

Geophysical data stakeholders were invited to participate in workshops to identify pain points, integration opportunities and required metadata fields:

- Team stakeholders – technical librarians, geophysical data analysts, geophysical data management team lead, IT support specialists, project manager, workflows consultant; and
- other Shell Australia stakeholders – geophysicists, geomatics (Global Information Systems (GIS) specialists)

The initial workshop included all stakeholders. As owners of the Project, the Team led the workshop, first outlining the scope of the Project and its goals. A brainstorming session allowed all participants to list metadata fields required to describe geophysical datasets, indicating fields that were "must have" and "nice to have".

At a subsequent Team stakeholder workshop this list of fields was refined (see Appendix 1), taking into consideration recognised metadata fields used in current metadata tools managed by the Library, and those used by the Department of Mines and Petroleum's Western Australian Petroleum & Geothermal Information Management System (WAPIMS) (Department of Mines and Petroleum, 2016) (see Appendix 2).

As experienced users and custodians of geoscience bibliographic databases, the librarians provided valuable input into the metadata required for access and retrieval of geophysical datasets, particularly in relation to media. The librarians were instrumental in determining the format of individual fields within the Database – free text, controlled vocabularies and radio buttons.

Team stakeholders also developed a list of Database functions to overcome the difficulties encountered in the management of large, complex datasets to create a one-stop-shop for geophysical data discovery. While the main driver for the Project was to link data to Business projects, the following functionalities–strongly lobbied for by the librarians–were seen as crucial for effective data management:

- iteration and version management: the data from a survey is sometimes reprocessed–either a whole survey area or a portion of the original survey–generating multiple iterations of processing and interpretation. Entire surveys or parts of surveys can also be merged and reprocessed and interpreted. It is not unusual to hold more than one version of a dataset, where versions have been received via various sources such as joint venture partners, data trades, geophysical data vendors and government sources. It is important to be able to track the provenance of these many versions and iterations;
- capture licensing information: each dataset is subject to a licensing arrangement determined by the nature of its acquisition. The data license has a direct impact on how the data can be shared within Shell and externally so it is important to capture this information as data is received so access is only granted to those

entitled to access. This comes into sharp focus when data is sent outside the company, and where multiple versions of the same dataset from various sources are held;

- capture survey name aliases: surveys and their various versions and iterations can be known by multiple names. Different names can be applied by different survey stakeholders – the government, acquisition company, processing company, geophysical data vendors, in-house geophysical data stakeholders;
- track associated datasets: when surveys are reprocessed or merged to create new data products it is important to be able to link to the original datasets to provide an accurate description of the new dataset. Similarly, it is valuable to be aware of new data products generated from an earlier survey when scoping areas of interest;
- capture transmittal information: all incoming (and outgoing) datasets should be accompanied by a transmittal which is signed and returned to the source of the data to verify receipt. Transmittals are vital for tracking data movements;
- capture multiple media identifiers (ID): as well as a Shell Media ID, incoming media will have an ID generated by the source company, and sometimes additional IDs applied by subcontracted vendors. It is important to capture all of these IDs as any or all may be supplied as a search parameter;
- capture legislative metadata: it is important to know what legislative framework governs the acquisition of a particular dataset since this determines what dataset components must be submitted to government, and the period of confidentiality for submitted datasets;

- media remastering alerts: media has a limited archival lifespan and must be remastered to maintain accessibility. Automated alerts generated based on type of media and media creation date help manage this process;
- extensive search functionality: all metadata parameters need to be searchable. Boolean searching required;
- cater for a variety of dataset formats: existing metadata tools were very media centric (e.g. magnetic tape, hard drive, paper) which historically meant multiple searches across multiple media registers when searching for datasets. The Database should capture metadata for datasets on all types of media, as well as datasets with no media (such as those received via file transfer), eliminating the need for multiple registers;
- bulk metadata import/export via .csv/.xls; and
- data loading information: capture details of dataset components loaded to corporate data stores and applications

Following the workshops, a framework was developed which consisted of four relational metadata subsets: Project, Survey, Media, and Seismic (loading). These subsets would be hierarchical: for the purpose of geophysical data management the Survey would be the main subset (see Figure 6). Taking design inspiration from the CSIRO's Data Access Portal (CSIRO, n.d.), each of the four metadata subsets would be displayed on a separate tab of the Database interface, and navigable via browse and search functions, as well as internal hyperlinks (see Appendix 3).

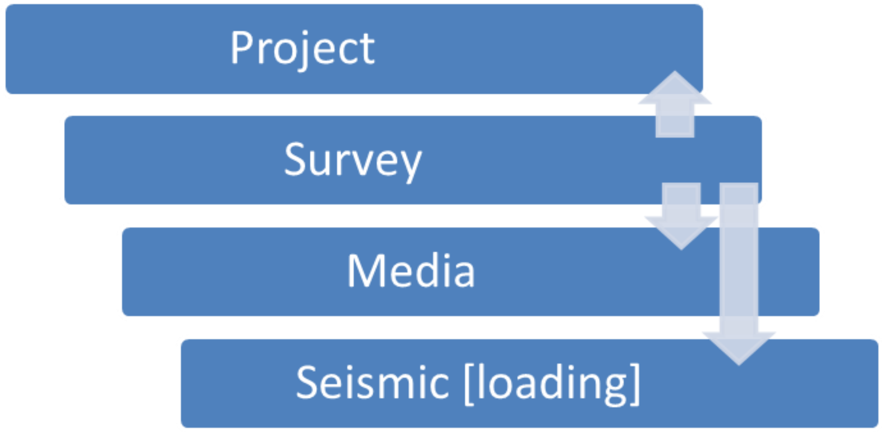


Figure 6: Metadata subsets

3. Database construction and testing

The structure and metadata field requirements were provided to a developer who created a prototype MySQL database. Team IT specialists made the Database available in a testing environment. Extensive testing using real-world datasets was undertaken by the librarians and geophysical data analysts in close collaboration.

At this point the Scrum Framework (see Figure 7) was used to control the development of the Database – this is a common tool used in software development (Scrum Alliance, 2016). A Kanban board was used to capture the four stages of the Scrum process: Backlog (wish list of features), Sprint backlog (priority list of features), Sprint (features under development), Done (features completed). Each Database feature was itemized on a post-it note and moved across the Kanban board as it moved through the development process. The Kanban board provided a quick visual snapshot of the Project's progress.

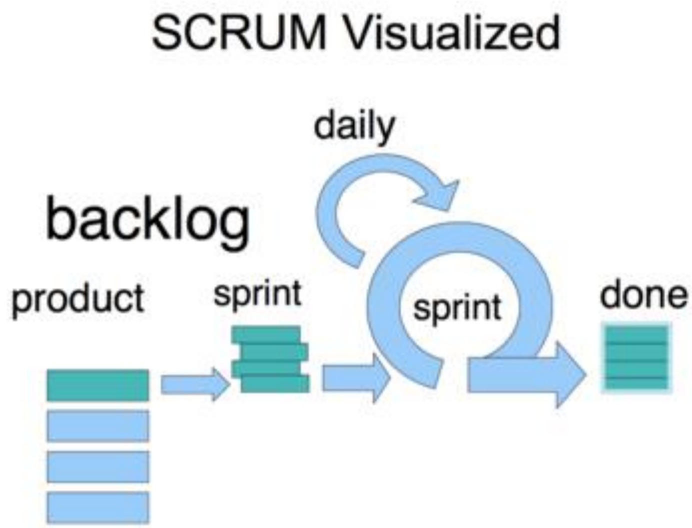


Figure 7: Scrum Visualized (Heller, 2011)

Following testing, a list of additional functions and bug fixes were added to the Backlog section of the Kanban board. From this list priority items were identified and added to the Sprint Backlog, and these were communicated to the database developer and placed under the Sprint phase. The developer worked on these items and delivered the revised code to Team IT specialists, who would apply updates to the Database for testing by librarians and geophysical data analysts. This cycle was repeated until all required changes and features were applied and moved to the Done phase of the Kanban board.

Scrum was also used for Team tasks relating to database development. Two examples of development tasks undertaken by the librarians:

- the visibility of the large offsite media collection needed to be improved. The librarians tested various label scanning procedures and image file types, ultimately choosing the in-house scanner to generate .tif files. Consultation with Team IT specialists and the database developer resulted in the application of a code patch; media label scans are named with the Media ID and the Database searches a specific file location to display the label image for each media record;
- the librarians sourced and developed controlled vocabularies for fields such as [Sedimentary] Basin, Permit, Company Name, Survey Type and Media Location in consultation with geophysical data analysts and GIS specialists. These controlled vocabularies were sourced from internal resources such as the Library Catalogue and external resources such as the WAPIMS database (Department of Mines and Petroleum, 2016)

4. <u>Guideline development and education</u>

Shell Global Seismic Standards include naming conventions for geophysical data and these are in use in Shell Australia, with some adjustments to allow for local anomalies. During the extensive testing procedure it became evident that additional naming standards were required for some incoming datasets, particularly those arriving electronically with little or no supporting documentation. Particularly challenging was applying a Media ID to an electronic file which does not exist on physical media. Through trial and error the librarians settled on a convention where the Media ID used in the Database uses the unique electronic identifier for that item within the relevant corporate data store.

The Database is bespoke and came with no set of instructions. Following their extensive experience of testing the Database the librarians and geophysical data analysts collaborated closely to develop a draft user guide. The draft user guide was tested by Team members with no experience of using the Database, and their feedback was incorporated into the final document. The user guide includes step-by-step procedures with screenshots showing the metadata entry and search screens, scope notes for metadata field contents, naming conventions, and a summary of database functions such as bulk metadata import/export and automated alerts. The user guide also contains a list of roles and responsibilities, detailing which Team members are responsible for updating the various metadata subsets. Once the Database and the user guide were completed, the new metadata tool was formally rolled out to the Team. One-to-one training was provided to Database users by the librarians and geophysical data analysts. The broader community of Shell geophysical data stakeholders who were involved in the initial Project workshop received a Database demonstration at an informal show and tell session.

5. Populating the Database

As new geophysical data arrives at Shell Australia its metadata is added to the Database, ensuring all relevant details are captured across the Project, Survey, Media and Seismic subsets.

Populating the Database with historical metadata has been more complicated.

It is important to capture pre-existing geophysical media holdings as it is common practice in the petroleum industry to revisit previously explored areas and re-work existing data using new processing and interpretation applications. Having metadata for historic holdings in the Database provides a complete picture of available datasets, saving time and money that might otherwise be spent searching for or re-purchasing data.

Due to the vast quantity of data held by Shell, importing historic metadata is progressing in stages. Importing metadata for Project and Survey subsets was achieved relatively simply with a bulk import via .csv file from existing databases.

Media metadata is being imported by the librarians, one survey at a time, with data associated with current projects given the highest priority. Scripting of this process was considered, however the complex web of historic metadata made scripting impractical. The librarians collated metadata from multiple existing media registers to a Media .csv template. As the Database Media metadata fields are slightly different to pre-existing media registers, the metadata needs to be correctly mapped and cleaned, and additional metadata added from documentary records. The librarians' extensive knowledge of the geophysical media collection has been vital in the painstaking process of gathering and accurately completing metadata. Once the Media .csv template for each survey has been completed it is imported into the Database. Geophysical data analysts are undertaking a similar phased approach to importing Seismic [loading] metadata into the Database.

RESULTS

The implementation of the Database has delivered a great improvement in many facets of geophysical data management at Shell Australia.

Clearer connections between Business projects and data

When a Survey is completed the metadata for the Survey is added to the Database and associated with a Business project. There is now a direct link between Business projects and datasets held by Shell Australia and a simple search of the Database provides a complete listing of all geophysical data held for a particular project.

Improved metadata capture

Detailed metadata is now captured in a central repository rather than across a number of media-based registers. The use of controlled vocabularies in the Database improves the consistency of metadata capture. The inclusion of a Survey Name Alias field has assisted in the identification of incoming datasets and with searching.

Improved search functionality

The ability to search consistent and detailed metadata via a single database has dramatically reduced search times for geophysical data for current Business projects. Searching and filtering within search results provides highly targeted results. The ability to export search results to .xls and share them with geophysicists is a valuable feature.

Better data management

Not all datasets received by Shell Australia are immediately loaded to corporate data stores and applications. The inclusion of the Seismic [loading] metadata subset enables a comparison between data held in-house (usually on media) with what has been loaded. Email alerts received by the Library enable more efficient tracking of media requiring remastering. It is now possible to easily audit geophysical data that has been submitted to government for individual work programs. The ability to view media labels for items held in offsite storage provides an additional layer of information and has reduced the number of items retrieved from archive.

Stronger relationships with stakeholders

The process of developing the Database developed the relationship between librarians, geophysicists and geophysical data analysts. These relationships have continued to grow as the need to accurately capture metadata encourages ongoing knowledge sharing about incoming datasets and technical issues relating to geophysical data.

DISCUSSION

The value of librarians in the development of the database

The involvement of Shell Australia's librarians in the development of the Database was an opportunity to demonstrate the value of information professionals in data and information management.

As the owners of Shell Australia's media registers and experienced users of a wide range of external geoscience databases, the librarians provided valuable input into metadata required for access and retrieval of geophysical datasets. The librarians' deep knowledge of the existing geophysical media collection has proved invaluable in extracting additional detailed metadata from historical records to ensure the metadata captured in the Database is rich and accurate.

Shell Australia's librarians recognise the long term value of data and information collections, beyond the relatively short Business project lifecycle. There is a temptation when developing new databases to populate it with data 'point forward'. While this is a simple and cheap option, it creates legacy datasets, and removes any possibility of developing a 'one-stop-shop' approach to data searching. The librarians successfully argued the value of bringing over as much metadata as possible from existing registers and databases to provide a complete view of Shell Australia's geophysical data holdings.

<u>The development process</u>

The workshops were a valuable information gathering tool in the early stages of the development process. Including all geophysical data stakeholders in the initial workshop ensured a wide range of perspectives were considered, and it also made these stakeholders easier to approach with questions as the Team moved through the development process.

The Scrum method was an effective way to manage a large and complex project by breaking it down into smaller, more manageable tasks. The Kanban board was an excellent visual tool which ensured all wish list tasks were dealt with, and provided a sense of satisfaction as items moved through the Scrum process to “Done”.

What did the librarians learn?

Extensive collaboration with geophysical data analysts and geophysicists provided the librarians with valuable insight into the acquisition, processing and interpretation of geophysical data.

Learning to use the SCRUM method has been an excellent addition to the librarians’ toolkit and has been applied to subsequent Library projects to great effect.

Future developments

Now that the Database is in full operation, its usefulness as a data management tool has been recognised by Team members and other Shell Australia geoscience data stakeholders. Linking datasets to Geospatial Information System (GIS) files would provide valuable added functionality when searching for data over a particular area and provision for this was made in the Survey metadata subset – implementing these links will require further collaboration with Shell Australia’s GIS specialists.

There is an opportunity to incorporate other datasets managed by the Team as additional metadata subsets. The simplest data type to add would be geological (well) data. More complex datasets such as modelling data will require more work. Including all datasets managed by the Team in the Database would provide a 360 degree view of all of Shell Australia's Business project data holdings.

CONCLUSION

The professional skills of the technical librarian remain relevant and valuable in an era of big data, however cross-disciplinary collaboration with data stakeholders is essential to communicate this value and develop additional disciplinary knowledge and data management skills.

The implementation of the Database has moved Shell Australia one step closer to the utopian ideal of data immortality (Baker & Yarmey, 2009).

APPENDIX 1: PROPOSED METADATA FIELDS

MEDIA METADATA	SURVEY METADATA
Survey name	Survey name
Media ID (multiple)	Survey alias
Data type e.g. Acquisition report, navigation, field	Survey type
Data details e.g. Inline, Xline	GA survey ID
Data format e.g. SEGD, SEGY, ASCII	State survey ID
Media location	Basin
Media type e.g. 3592 tape, DVD	State
Data license	Country
Media source	Terrain
Media listing	Permit No
Media label	QC By Geomatics
Transmittal	2D/3D/4D/Other
Media creation date	Line prefix
Media review date	Acquisition start/end
Destroyed	Reprocessing dates
Remarks	Operating company
	Processing company
	Acquisition contractor
	Legislative type
	Confidentiality
	Open file dates
	Gazettal information
	Remarks
	Associations

APPENDIX 2: WAPIMS METADATA FIELDS USING EXAMPLE "ADELE TREND TQ3D 3D M.S.S."

SURVEY METADATA FIELDS

Survey name	Survey id	Survey acquisition id	Acquisition name	Survey type	Line prefix	Start date	End date	Operator	Contractor	Basin	Terrain	Titles	Legislation	Total line length (km)	Area (sqkm)
Adele Trend TQ3D 3D M.S.S.	S001758	S001758A	Adele Trend TQ3D 3D M.S.S.	3D	GPAT	11/11/2000	15/02/2001	WesternGeco Australia Pty Ltd		Browse	MARINE	SPA 1SL/00-1,WA-35-P	P(SL)A67		3197.992
Adele Trend TQ3D 3D M.S.S.	S001758	S001758B	Adele Trend TQ3D 3D M.S.S.	2D	GPAT	20/10/2000	11/11/2000	WesternGeco Australia Pty Ltd	Western Geco	Browse	MARINE	SPA 1SL/00-1,WA-35-P	P(SL)A67	197.5	

REPORT DATA METADATA FIELDS

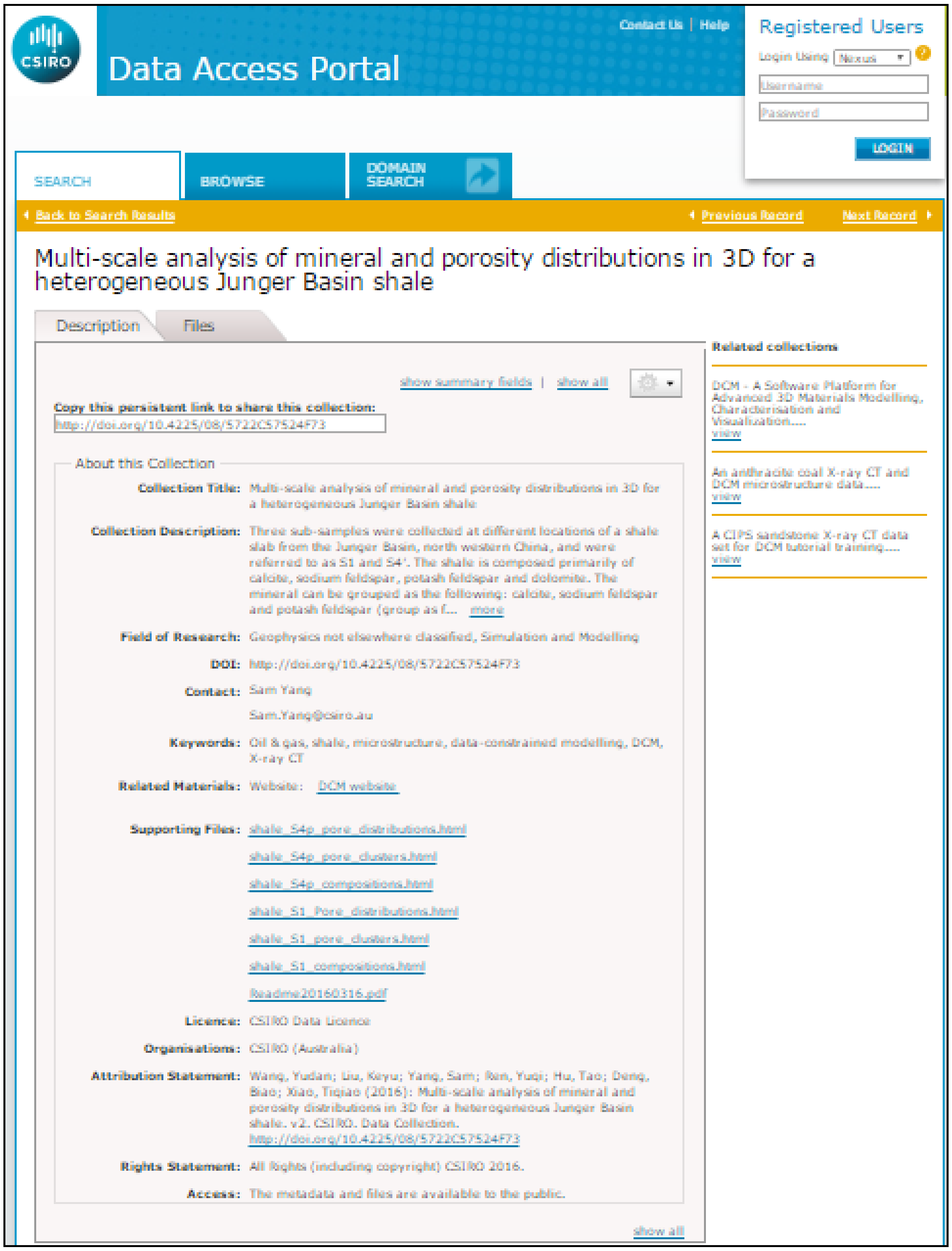
Survey name	Acquisition id	Title	Archive id	Company	Primary relation	Confidential
Adele Trend TQ3D 3D M.S.S.	S001758A	Adele Trend TQ3D 3D M.S.S.; Phase 1; Acquisition Report	S10452 A1	WesternGeco	S001758A	N

DATA METADATA FIELDS

Survey name	Acquisition id	Data description	Data types	Company	Tape number	Primary relation	Confidential	First shot point	Last shot point
Adele Trend TQ3D 3D M.S.S.	S001758A	Adele Trend TQ3D 3D M.S.S.; Every 10th inline/xline; Phase 1 Bin Centre NAV	BIN CENTRES	WesternGeco	S10452 A1	S001758A	N		

Note: These metadata fields are the public view and do not include more extensive "back end" fields used by the Department of Mines and Petroleum to manage Geophysical data. (Department of Mines and Petroleum, 2016)

APPENDIX 3: THE CSIRO DATA ACCESS PORTAL INTERFACE



(CSIRO, n.d.)

REFERENCES

Anderson, W. (2006). Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data. *Data Science Journal*, *3*, 191–201. http://doi.org/http://doi.org/10.2481/dsj.3.191

Association of Research Libraries Workshop on New Collaborative Relationships. (2006). *To stand the test of time: Long-term stewardship of digital data sets in science and engineering.* (pp. 1–160). Alrington, VA: National Science Foundation. Retrieved from http://www.arl.org/pp/access/nsfworkshop.shtml

Australian National Data Service. (2016a, February). The award winning team. *Share Newsletter*, *24*, 6–7.

Australian National Data Service. (2016b, February). The data librarians. *Share Newsletter*, *24*, 8.

Baker, K., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *The International Journal of Digital Curation*, *4*(2), 12–27.

Bruce, B. (1991). An holistic approach to data management in the petroleum industry and its appplication to reference database design for the next century. In *Proceedings: National Conference on the Management of Geoscience Information and Data: Adelaide, 22-25 July 1991* (Session 4: 15–27). Adelaide, South Australia: Australian Mineral Foundation.

Cook, S. (2014, July). *CSIRO Data Access Portal*. Presented at the AGIA Winter Warmer, Perth, Western Australia. Retrieved from https://www.dropbox.com/sh/o2kydtej1cn15qj/AAB9PFKLJgEluJOihZBE_lDua?dl=0

Cronin, P. (2016, May 13). Big seismic. Retrieved from Personal correspondence

CSIRO. (n.d.). CSIRO Data Access Portal. Retrieved from https://data.csiro.au/dap/landingpage?execution=e3s2&_eventId=viewDescription

Cutler, P., & Maples, C. (2002). Resources in peril. *Geotimes*, *47*(6), 16.

Department of Mines and Petroleum. (2016). WAPIMS. Retrieved from https://wapims.dmp.wa.gov.au/wapims

Department of Mines and Petroleum, Petroleum Division. (2010). Petroleum exploration,production and development activity in Western Australia in 2009. *Petroleum in Western Australia*, *April 2010*, 7.

Diviacco, P., Lowry, R., & Schaap, D. (2012). Marine seismic metadata for an integrated European scale data infrastructure: the FP7 Geo-Seas project. *Bollettino Di Geofisica Teorica Ed Applicata*, *53*(2), 243–252. http://doi.org/10.4430/bgta0051

Diviacco, P., Wardell, N., Forlin, E., Sauli, C., Burca, M., Busato, A., … Pelos, C. (2015). Data rescue to extend the value of vintage seismic data: The OGS-SNAP experience. *GeoResJ*, *6*(June 2015), 44–52. http://doi.org/http://dx.doi.org/10.1016/j.grj.2015.01.006

Foote, J. B. (2010). State geological survey libraries: A disparity in resources, services, access and professionalism. *Science and Technology Libraries*, *29*(1), 53–68. http://doi.org/10.1080.01942620903579385

Geoscience Australia. (2015). Offshore Petroleum Exploration Acreage Release Australia 2015. Retrieved from http://www.petroleum-acreage.gov.au/2015

Geoscience Australia. (2016). Seismic. Retrieved from http://www.ga.gov.au/scientific-topics/disciplines/geophysics/seismic

Geosoft Inc. (2016). *Exploration information management report 2015*. Retrieved from http://www.geosoft.com/exploration-information-management-survey-2015

Gerdes, L., & Smith, K. (1990). *Guidelines for geoscience bibliographical databases*. Perth, WA: Australian Geoscience Information Association.

Harvey, A., & Diment, J. (1979). *Geoscience information: A state-of-the-art review: Proceedings of the 1st International Conference on Geological Information, London, 10-12 April, 1978*. Heathfield, England: The Broad Oak Press Ltd.

Heller, G. (2011, November 1). Scrum visualized [Web log post]. Retrieved from http://home.gregoryheller.com/scrum-diagram

Hey, T., & Hey, J. (2006). e-Science and its implications for the library community. *Library Hi Tech*, *24*(4), 515–528. http://doi.org/1168375141

Johnson, V. (2011). *The role of information professionals in geoscience data management* (Unpublished master's thesis). Curtin University, Bentley, Australia.

Jones, R. (1986). Geoscience data gathering - are we getting value from the harvest. In *Proceedings of the 3rd International Conference on Geoscience Information* (Vol. 2, pp. 31–37). Adelaide, South Australia: Australian Mineral Foundation.

Joseph, L. E. (2008). Digital data curation: Investigating potential collaboration between librarians and researchers. In L. Johnston (Ed.), *Libraries in transformation,* *exploring topics of changing practices and new technologies* (pp. 3–9). USA: Geoscience Information Society.

Lyon, L. (2007). *Dealing with data: Roles, rights, responsibilities and relationships* (Consultancy report) (p. 65). Joint Information Systems Committee. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report.pdf

National Library of Australia. (n.d.). Facts and figures. Retrieved from https://www.nla.gov.au/facts-and-figures

National Offshore Petroleum Administrator. (n.d.-a). Data management. Retrieved from http://www.nopta.gov.au/data-mgmt/index.html

National Offshore Petroleum Administrator. (n.d.-b). Legislation, determinations, guidelines & fact sheets. Retrieved from http://nopta.gov.au/legislation/index.html

National Offshore Petroleum Administrator. (n.d.-c). National Electronic Approvals Tracking System (NEATS). Retrieved from http://neats.nopta.gov.au/

Office of Parliamentary Counsel. (n.d.). Offshore Petroleum and Greenhouse Gas Storage Act 2006. Retrieved from http://www.comlaw.gov.au/comlaw/management.nsf/lookupindexpagesbyid/IP200613750?OpenDocument

Parkin, L., & Tellis, D. (1977a). Australian Earth Sciences Information System. *Proceedings of the Australasian Institute of Minerals and Metallurgy*, (262), 7–23.

Parkin, L., & Tellis, D. (1977b). Australian Earth Sciences Information System. In *Proceedings no.262* (pp. 7–23). Australasian Institute of Mining and Metallurgy.

Pruett, N. J. (1982). Data on the rocks: A cross-section of user needs. In *Second International Conference on Geological Information: Proceedings* (Vol. 1, pp. 29–43). Golden, Colorado: Oklahoma Geological Survey.

Rao, M. V., Coe, J., Gibson, J., Gleitman, D., & Rector, S. (2015, August 16). Method, systems, and program product for selecting and acquiring data to update a geophysical database. Retrieved from https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US6931378.pdf

Reflection seismology. (2016). Retrieved from https://en.wikipedia.org/wiki/Reflection_seismology

Scrum Alliance. (2016). Learn about Scrum [Web log post]. Retrieved from https://www.scrumalliance.org/why-scrum

Shell. (2016). Future of energy. Retrieved from http://www.shell.com.au/future-energy.html

Soehner, C., Steeves, C., & Ward, J. (2010). *E-Science and data support services: A study of ARL member institutions* (pp. 1–74). Washington DC: Association of Research Libraries. Retrieved from http://www.arl.org/bm~doc/escience_report2010.pdf

Tellis, D., & Crowe, D. (1982). The Australian experience with information transfer in the geosciences. In *Second international conference on geological information: Proceedings* (Vol. 2, pp. 285–307). Golden, Colorado: Oklahoma Geological Survey.

Tellis, D., & Gerdes, L. (1986). Innovative developments in national and regional earth sciences reference data bases in Australia. In *Proceedings of the 3rd International Conference on Geoscience Information* (Vol. 1, pp. 71–94). Adelaide, South Australia: Australian Mineral Foundation.

The Federal Geographic Data Committee. (2014, September 12). Geospatial metadata standards. Retrieved from http://www.fgdc.gov/metadata/geospatial-metadata-standards

Treloar, A. (2009). Design and implementation of the Australian National Data Service. *The International Journal of Digital Curation*, *4*(1). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/107/83

Ward, D., & Walker, R. (1986). Twenty years of geoscience information - An analysis of coverage and content. In *Proceedings of the 3rd International Conference on Geoscience Information* (Vol. 2, pp. 52–68). Adelaide, South Australia: Australian Mineral Foundation.

Wilkinson, R. (2006). *Speaking Oil & Gas*. BHP Billiton Petroleum Pty Ltd.

Yacopetti, M., & Mundell, S. (2010). Improving the quality of geoscientific information. In *Back in the Black* (pp. 97–100). Retrieved from http://www.agia.org.au/docs/BBGG2010.pdf