# Management of Petroleum Data Records in the Custody of Australian Archives

*Stephen Yorke*

**Stephen Yorke** has been with Australian Archives since 1978 and has worked in various positions mainly connected with appraisal of records and systems analysis. In recent years he has been a project officer with the Custody and Preservation Program. He has been concerned with petroleum data records related issues since 1991.

*There has been substantial debate recently in the archival community on custodial issues related to records in electronic media. A major issue is whether archival institutions should take electronic records into their custody or rather rely on the creating agency to fulfil all necessary management, servicing, preservation, and access requirements[1]. The purpose of this paper is not to discuss the merits or otherwise of such a distributed custody policy. Rather, it is to describe the problems associated with archival management of a significant volume of a particular type of electronic record. With the benefit of experience and recent developments in technology, many of the problems described in this paper could have been avoided or resolved sooner. In a sense, the situation described could be used to support arguments for or against archival institutions accepting custody of electronic records.*

This is a refereed article.

## Background

Exploration for petroleum and the development of petroleum resources in Australia occurs within a very complex legal and administrative context of responsibilities divided between the Commonwealth government and State and Territory governments. Under current arrangements, the Commonwealth

government is responsible for control of off-shore areas beyond the three-mile limit to the edge of the continental shelf as defined for the purposes of international law. Off-shore exploration is governed under the provisions of the *Petroleum (Submerged Lands) Act 1967* [P(SL)A]. State and Territory governments are responsible for on-shore areas.

The petroleum exploration, development, and production processes involve a high risk financial investment by explorers and the Government (in the form of taxation and other financial incentives). Under the P(SL)A, exploration permits are periodically released to exploration companies under a work program or (in highly prospective areas) a cash bidding system. The object of the exploration permit arrangement is to ensure an orderly and efficient exploration work program. If petroleum is found in an exploration permit area then a production license can be issued to the title holder providing that the discovery can be developed in the short term.

A requirement of the P(SL)A is that a copy of all exploration results including seismic data is provided to the Commonwealth government. In many countries such data remains the exclusive property of its owners. It largely comprises original field tapes gathered by survey ships undertaking large scale surveys. The data acquisition process is a complex operation resulting in thousands of data tapes for a single survey. The data is processed and analysed by the explorers to determine suitable drilling targets. An off-shore well costs an average of $7,000,000 to drill in an active exploration area (and most wells are dry). When a title is relinquished the data collected by the title holder can be used by other explorers to develop new exploration strategies or for other research purposes. New methods in reprocessing older data has led to new interpretations of prospectivity or different drilling strategies. It is obviously far more cost-effective to try and reprocess existing data than undertake a new survey[2].

The formats of seismic and other exploration data records have changed markedly since the 1950s. These changes reflect developments both in scientific knowledge about petroleum formation and exploration methods generally, and the technology that is adaptable for purposes of petroleum exploration. The wide variety of formats requires different technologies to access and copy. The most common media are 1 inch and $^1/_2$ inch digital magnetic tape (the one seen in film scenes of computer rooms from the 1960s through the 1980s). Some of the formats - particularly early analogue formats - are no longer accessible. This inaccessibility is not due to physical decay but rather the unavailability of the necessary hardware and software. A more detailed description of the data formats used for petroleum data and associated issues can be found in *Keeping Data*[3].

## The problems

In 1974, under an arrangement between the Australian Archives and the then Bureau of Mineral Resources, surrendered petroleum data records began to be stored in Sydney at the Villawood repository. A significant reason behind this decision was that at the time Villawood was the only facility available in Australia that could provide sufficient air-conditioned storage space to the standard necessary for magnetic media. Life expectancy of magnetic media is dependent on storage temperature, humidity levels, and careful machine usage when processing[4].

The volume of petroleum data records in custody has fluctuated over the years: at its peak there were some 600,000 data tapes in various formats. Currently, 12,500 shelf-metres are allocated for storage of petroleum records in all formats. The total floor area taken up by petroleum data storage and related activity is some 1,700 square metres. It is estimated that a similar volume of exploration data subject to eventual surrender to the Commonwealth is held by exploration companies.

### *Control and access*

The data records are recognised as Commonwealth Government records under the terms of the *Archives Act 1983*. This Act governs recordkeeping by Commonwealth agencies and the management of records belonging to the Commonwealth[5]. The petroleum data (for reasons which will perhaps become clearer later on) could not be seen purely as either 'data records' or 'electronic records' as the terms are now used. Similarly, the collection as a whole could not be seen as a 'data archive'. A data archive implies, inter alia, that the holdings are well organised by subject, the methodologies used for each data set collection are readily known, the data meets high standards in terms of consistency of contents and useability, and a common medium is used for data storage.

While the Villawood magnetic tapes, for example, contained data, they could not be meaningfully accessed or processed without reference to associated records, mainly in paper formats. In addition, because of the use of many different media for the data and the variety of data collecting techniques, the quality and usefulness of the petroleum data varied quite markedly from survey to survey. If there had been consistency in media and formats then it would have been possible to treat the petroleum data records more as a data archive than in the way that is described. It would appear the original intention of the agency controlling had been to develop something like a true data archive. However, for various reasons this proposal did not eventuate.

The twenty years of the Archives' direct responsibility for petroleum data has exposed many of the problems of trying to treat such records as 'items' in the traditional sense. In addition, it has revealed problems not conceived of at the time of the original decision to take the data into custody[6]. The basic approach to control and access adopted by the Archives was to treat the data records as being analogous to paper records. As a result, record series were created on the basic principle of treating all records for surveys off-shore from the coastline of a State as being one series . This system worked satisfactorily at the broad level to *control* the items in custody. However, this approach required much use of complex lists of individual items when responding to an enquiry. The system also posed problems because for any given area there could be many relevant data records gathered in several different surveys. Also, the degree of completeness of the data record for a given survey was not always readily apparent from the documentation available. The petroleum data for a survey could be transferred in a number of consignments over several years and the documentation accompanying each transfer could be equally incomplete or poorly prepared. This could cause another set of queries or exploration of other options. Consequently, a response to a request for information about the data records could take some time and relied very much on the skills of a few experienced staff who did not have any technical background in the relevant disciplines.

An important factor contributing to management problems was that petroleum data records were surrendered by explorers only when they were no longer of any use or when they chose to surrender their exploration titles. There was no incentive for them to ensure that the data records were complete, recorded on high quality magnetic tape, or that the metadata was comprehensive. The metadata sets out details such as data gathering processes used in the survey including navigation data, processing details, and so forth. And it is the quality of this information that determines in many instances whether any other user can extract any real value from the data records.

The short-comings of depositors were compounded (if not indirectly encouraged) by the Archives. The Archives did not have the technical capability in terms of hardware and software, and staff expertise to access the data itself either at time of deposit or later on. Normal archival skills were applied to the deposit and retrieval processes. However, there was a significant lack of the technical resources necessary for the Archives to manage them properly intellectually at an adequate level. For example, while descriptive errors in labels and paperwork submitted by depositors were detected, it was not possible to establish whether the tape and supporting documentation contained what it purported to contain or - in the case of tapes - contained anything at all.

Lending procedures required borrowers to copy the original data records first and to conduct any re-processing using the copy. While copying was a relatively inexpensive process, there was evidence in several cases that re-processing had occurred using the original tapes. Such actions place the unique data at risk. Also, there could be no certainty that the borrower returned the same tapes that had been borrowed - and that they were in the same condition as when borrowed.

*Appraisal and disposal*

This inability by the Archives to control and access the data in turn caused appraisal and disposal problems. As mentioned, the petroleum data had firstly come into custody because of the legislative requirement for surrender. A very significant backlog of data records had built up in exploration companies awaiting a decision about a point of deposit. Thus there were very large transfers from the very beginning of the deposit arrangement. In any event, it could not always be established whether the data records proposed for transfer were unique, duplicates, incomplete - or even blank - and if the metadata and paperwork were accurate. As a result, data records that came into custody were unappraised and unsentenced in terms of both the industry's and the Archives' requirements.

Putting it simply, the main appraisal principle applied to seismic data is that if the data is unique, ie there are no other data records covering the same area, then it should be retained on a continuing basis. If an area is re-surveyed then the earlier data is generally considered obsolete as the new survey will be far more detailed. But, for the reasons explored above, there was often no way of assessing whether the data records were unique or complete. To compound this problem, surveys can be conducted which overlap parts of other surveys. Thus there was no way to be certain if disposal decisions could be made with confidence about the bulk of the data records.

*Preservation*

However, control and access problems, and their consequences for appraisal and disposal were not the fundamental issue. The international nature of the petroleum industry and the competitiveness of the major explorers resulted in a number of differing acquisition and processing standards. These problems were accentuated by the use of proprietary hardware and software at a time when the computer industry was undergoing rapid changes in approaches to processing this type of data. Processes followed by organisations often represent the particular

business need at a given time and such needs generally stress short-term objectives and - in turn - cause their own set of problems.

An example of such a problem follows. In the late 1970s several very large marine seismic surveys were conducted using $^1/_2$ inch magnetic tape and typical data formats. Much thinner tape was used than was usual so that far more tape could fit on a standard reel and thus it would hold more data. However, this usage led within a few years to binder failure, ie the oxide layer on the tape started to peel off whenever data was accessed[7]. It was estimated at the time some 10,000 tapes in custody were possibly affected. Thus these tapes posed significant preservation problems that had to be addressed. This was done firstly by freezing all access to tapes of known affected surveys and conducting an extensive investigation to determine which were involved. These tapes were then treated and copied on to a more compact medium at a not insignificant cost to the Commonwealth. However, this crisis did result in the recognition by government that preservation of the data was costly and led to the development of a management database system to control the petroleum data records.

It is likely that a significant contributor to preservation problems was the storage of data records by explorers under unsuitable temperature and humidity conditions in office space or uninsulated sheds or other poor handling practices. These actions occurred even though data storage and handling requirements have been known for decades. For example, consignments were received with tapes smeared with oil, or covered in dust and other substances. Similarly, tapes were shipped across Australia by trucks with no consideration being given to required environmental controls for magnetic media. Somewhat ironically, older data records such as those in analogue formats may present no immediate preservation problems due to higher quality media. Their problem is that the hardware and software and technical expertise are no longer available to access the contents[8].

As changeable as business needs are the changing technologies for capturing and processing seismic data. Generally, the changes have involved collecting more detailed data during a survey and storing it on increasingly compact media[9]. These developments are not unique to the petroleum exploration industry but they do significantly complicate the management process, particularly when the records may have enduring value. Inter alia, it becomes necessary to establish the deterioration rates of various media and actual deterioration rates of individual items in all the different formats used, not an easy task. To take a simple example: the brand name might be on the carrier (reel) of a tape but that does not mean the actual tape is of the same brand and the difference is not readily discernible[10].

Thus it is not easy to assess the quality of particular brands to indicate likely problem areas and to know which records should be a priority for preservation.

If it can be established which data is unique, long-term preservation can be problematic. The only practicable way for ensuring longevity of data records in electronic formats is by periodically copying the contents at some cost to new media[11]. This has to be a phased process given the sheer volume and general age of the data. As mentioned, the deterioration rate (and thus the copying priority) is extremely difficult to determine or quantify. These preservation selection problems are compounded by the fact that while a retention decision may be based on uniqueness the actual *value* of a given survey is not. The value, including the monetary value, of a dataset rises and falls depending on external influences. The most common factors are the amount of exploration activity being undertaken generally and the perceived prospectivity of an area. In turn, this is linked to the world price of oil and the incentives being offered by governments: if exploration activity increases so does interest in existing data. In more practical terms, if a given area is considered non-prospective then the value of data relating to the area is low. However, if petroleum was to be found then all data records for that area could be extremely valuable - at least in the short-term.

In summary, while users of the data focus on their immediate business needs, such business needs may in a few years require data that is in need of active preservation now. The potential users may not see these problems as their concern let alone acknowledge the need to develop a common or co-ordinated approach to data issues. The costs of actively preserving all the data both in and out of custody is in the order of tens of millions of dollars. Thus a preservation program is difficult to structure, co-ordinate, and fund particularly when the data records are of disparate age, media, and source. Therefore it is very difficult to come up with an appraisal and preservation strategy that is likely to satisfy all the stakeholders in petroleum exploration within Australia.

## Towards a solution

It is important to remember that the problems described were not apparent from the beginning but only revealed bit by bit over a number of years. Similarly, recognition by government of the problems over the years has been limited and marked by a reluctance to provide the sums necessary for an ideal solution to all the problems: from storage through preservation. However, some funding was provided to address various problems such as the preservation of the stiction

affected data records. This was done by the Bureau of Resource Sciences (BRS) which is the agency functionally responsible for petroleum data matters.

In 1993-94 the Archives undertook a review of its complete involvement with the data records. This review was a complex process. For, as has been described, the problems are not only complex but inter-related such as the inability to identify preservation priorities without better item level control and analysis. The major issues identified during the review can be put in simple terms as:

- to establish adequate intellectual control;
- to come to terms with the constantly changing and improving technology for storing data;
- to solve the problem of preserving existing data; and
- to store adequately the physical volume of the data records.

Under the terms of an extensive negotiated agreement in 1995, a new management arrangement was entered into by the Archives and BRS. The Archives now only provides storage services. All other services and functions, including intellectual control, processing, handling of enquiries from potential borrowers, retrieval, culling of obsolete and duplicated data, and preservation, are the responsibility of the BRS. The agreement places particular emphasis on preservation needs and the sentencing requirements for obsolete data. To these ends, the actual storage charges for petroleum data records are linked not only to the physical volume in custody at Villawood but also to agency performance in undertaking preservation activities. The charges are based on achievement of performance targets that are set annually by the Archives in conjunction with the BRS.

The BRS has stationed technical staff at Villawood and the question of intellectual control is being progressively addressed. As well, it is reviewing all the data records in custody and entering details of all surveys and items into its database management system. The system will control all data records and it will assist identification of duplicate and other redundant material. It is envisaged that at some stage remote access to control data will be possible by potential borrowers. The question of developing and implementing comprehensive national standards for data gathering and management is also being addressed.

The BRS has undertaken the preservation copying of particular surveys as well as culling many obsolete and redundant data records. There is also a preservation strategy in place whereby users can borrow at minimal cost in exchange for

returning a copy of the data in a compacted format.

In the longer term, management of preservation and storage requirements appears to be converging with the adoption of new storage media generally by the exploration industry. This means that something smaller than a VHS cassette can hold 10 gigabytes of data or the equivalent of over 55 of the $^1/_2$ inch magnetic tapes and there is an expectation of far greater capacity in the near future. Other media such as optical tape promise to increase the data capacity of an item by the order of a magnitude at minimum. Thus problems presented by the sheer physical volume of new data may no longer be significant. There will, of course, be a continuing problem with existing data records which may take some time to resolve to the satisfaction of all stakeholders. The greatest problem is funding the compaction process for existing data records[12]. But the eventual result of these activities will be a collection which could be said to resemble a data archive along the lines proposed a generation earlier.

## The lessons

The main problems over the past twenty years of the Archives' involvement with petroleum data have been many and - indeed - often painful. But it is suggested some major lessons have been learned and that these are in some way applicable to all archives dealing with data records, and in some cases to all records.

1. If an archives cannot access the contents of data records in its custody then there is a severe limit to its ability to properly and efficiently manage them and assure their authenticity, eg to ensure that what was borrowed was returned in the same condition and has not been affected by access processes. If the archives cannot master the current technology used for creation and access, then what are the odds that it will be able to better manage future generations of data formats and media?

2. Data records grow to what can be termed a 'critical mass' or volume. This can equally be due to the complexity of their contents as to their physical volume. It is at this point an archives can be 'trapped' because either withdrawal from involvement is not possible or the set-up cost of alternative arrangements is too high. However, to continue managing existing data records will require additional resources - even if no new transfers occur. Eventually, these demands will distort the activities and services that the archives can provide to other clients and other records.

3. The roles and responsibilities of all stakeholders or other parties with a direct involvement or interest must clearly be understood and formally set out at the outset. The stakeholders and their staff will inevitably change over time and the greater the number of stakeholders then the greater the potential for disagreement about respective responsibilities (especially after changes in personnel). As each stakeholder's interests can conflict with those of others, a lack of coincidence of interests will cause significant if not intractable problems, eg in deciding whether the available resources should be concentrated on accessibility issues rather than preservation activities.

4. The most cost-effective and efficient way of managing data records and electronic records over the long term involves consistent use of common standards for creation, access, storage, and preservation by creators, users and whoever is responsible for their custody and preservation over time - whether the creator or an archival institution. The greater the number of different organisations involved, and the greater the lapse of time between creation and transfer (if this occurs), then the greater the problems of developing common standards and ensuring their consistent application by all parties. The less that standards are followed at the time of creation then inevitably the more the cost of access and preservation, and the greater the proportion of costs that will fall on the long-term custodian.

5. Retrospective establishment of proper management requirements for data and electronic records including item level control is very expensive, if not impossible, in terms of the resources necessary to develop proper control systems and capture the necessary metadata. A clear and considered approach to planning and consideration of possible consequences of custody over time is essential before any commitment or other involvement is negotiated.

## Conclusion

In all probability, most data records and electronic records will require some degree of active intervention by their custodian at some stage during their period of retention - whether to preserve them or merely to access their contents. All such involvement requires resources such as the technical skills and capabilities needed to provide access, to transfer data and records to new media, or to ensure the continuing availability of necessary hardware and software. This is in contrast to the basically passive approach that is used to manage more traditional paper records.

The need for some of this intervention can be predicted and quantified. But the amount and cost of intervention become more problematic as the retention period increases. The odds are that there will be some unforeseen intervention necessary due to totally unanticipated requirements. For example, the custodian may discover that a storage medium is vulnerable to unexpected decay. Similarly, technological innovation will continue to occur and at an even faster rate. The starting question then is how to determine what type of custodial 'environment' is best suited to which particular electronic data or records.

## Endnotes

1. Perhaps the best-known of the protagonists are David Bearman, who advocates postcustodial approaches and Luciana Duranti who defends the traditional custodial role of archival institutions. Much has been written by these and others on the matter. See, for example the articles by Duranti, Eastwood, O'Shea and Roberts, and Upward in the November 1996 issue of this journal. They provide a useful bibliography of the debate.
2. However, it should be pointed out that some data gathered will have a retention value for scientific purposes or other reasons. For example, some data records will remain unique due to changing circumstances and access conditions for particular areas. This situation applies for the seabed now included in the Great Barrier Reef Marine National Park and other areas which are covered by more recent international treaties.
3. Steve Stuckey, 'The Good Oil for Australia: Petroleum Data', *Keeping Data*, edited by Barbara Reed and David Roberts, ASA, Sydney, 1991.
4. Australian Standard AS 4390.6 - 1996, *Records management Part 6: Storage*, sets temperature and humidity requirements at 18°C ±2% and 40%RH ±5%. For a general discussion of the issues see Guy Petherbridge, 'Environmental and Housing Considerations for the Preservation Of Modern Records', *Proceedings of the 1991 Conference of the Records Management Association of Australia*, RMAA, Darwin, 1991.
5. The Act's definition of a 'record' is very wide. The definition includes 'magnetic tape or disc that is, or has been, kept by reason of any information or matter that it contains or can be obtained from it or by reason of any information or matter that it contains or can be obtained from it by reason of its connection with any event, person, circumstance or thing'.
6. For a brief pertinent discussion of the practical issues associated with archival custody and management of electronic records, see Hugh W. Shinn, 'The Electronic Archive', *Archivaria*, no. 36, Autumn 1993.
7. This process is known as 'stiction' or 'oxide shedding' in the petroleum industry, and more generally as 'sticky shed syndrome'. The technical name is 'binder hydrolysis' and it has also been found to occur with audio tape and videotape. Various companies have developed heat-based treatments which will temporarily reverse the process enabling tapes to be copied. For a detailed discussion of tape deterioration and related issues, see John W.C. Van Bogart, *Magnetic Tape Storage and*

*Handling: A Guide for Libraries and Archives*, The Commission on Preservation and Access, National Media Laboratory, June 1995.

8. There are very few places in the world now capable of reading analogue records because the machinery is no longer common. In addition, there are few people now with the technical knowledge to process such tapes.
9. The $^1/_2$ inch tape which reigned for some 30 years for data gathering has now been replaced with what are known as 3480 and 3590 tapes.
10. This type of problem is in reality even more complex: only a few companies manufacture data tape used for seismic records but it is marketed under many different brands by various other companies. The actual formulations used are proprietary information and can change without indication or other change to the label on the tape container.
11. The alternative for digital tape was to undertake periodic cleaning and re-tensioning. This is a resource intensive process and was never practised at Villawood. Copying to new tape/media is a more cost-effective process and is the best means to avoid problems of media obsolescence.
12. The amount of data being collected has, of course, increased by an order of magnitude to take advantage of the new technology. The many implications of the use of highly compacted data are outside the scope of this paper. However, one implication is that the value of a given record item would significantly increase (say 55 times) and the damage or cost of its loss would increase commensurately.