5

Data Preservation

Introduction

5.1 Commonwealth recordkeeping is in the midst of a major shift in focus. For many years recordkeeping in Commonwealth agencies was overwhelmingly paper-based. In recent years, however, the rapid growth in the use of electronic messaging and storage systems has drastically changed the requirements.

> The National Archives of Australia has the task of ensuring that there is a full and accurate record of Government business. The changes flowing from the rise of electronic recordkeeping have presented NAA with the additional task of radically changing the way that government records are maintained.[1]

5.2 In reassessing its role, NAA produced a Green Paper in 2002, to promote discussion among Commonwealth agencies and to make government employees aware of the importance (and difficulty) of the task. The paper said that the rate of change had increased with the widespread introduction of computers into the workplace and had '… dramatically altered the way in which employees work, communicate and share information.' It continued:

> These changes have made recordkeeping both more difficult and more significant. For many years lack of attention to recordkeeping has been mitigated by the existence of long-standing, well known practices for the use of paper records. Paper

---

1 Helen Heslop, Simon Davis and Andrew Wilson, *An Approach to the Preservation of Digital Records*, National Archives of Australia, December 2002, pp. 5-6.

> records also have a robustness that enables them to survive long periods of neglect. In contrast, the sometimes haphazard use of electronic systems for communication and storing recorded information is more fragile.[2]

5.3 Another problem facing those responsible for setting the standards for Commonwealth recordkeeping is that, to date, the preservation of electronic records has not been treated with the same importance as the preservation of paper records. In a survey commissioned by the NAA, the results showed clearly that the bulk of electronic information in Commonwealth agencies was '… being created, managed and possibly disposed of without the benefit of the knowledge and expertise of trained records management staff.'[3]

5.4 The NAA Green Paper described the new challenges presented by the proliferation of electronic information by saying that electronic systems offer many advantages but agencies must ensure that these records are captured, survive as long as they are needed, and can be read and understood. The main example is e-mail messages, which must be captured into corporate recordkeeping systems where they can be preserved securely and found easily.[4]

5.5 In addition, the data must be stored in such a way that it can '… be migrated forward with hardware and software changes so that the records are still accessible.'

> Such records provide evidence that the transaction occurred and essential details about it. Nonexistent or poor quality records will prevent online business being conducted successfully ...[5]

5.6 A significant amount of the Government's electronic information needs to be preserved for very lengthy periods. FaCS stressed the importance of this factor and the need for a concerted approach to the problem.

5.7 FaCS considered that the Commonwealth needed to pay greater attention to the long term preservation of, and access to, its data holdings. It said '… there is no whole of government strategy or resources for identifying data sources across agencies that need to be preserved over long periods of time.' And then it added: 'There is also a need to ensure that such data remain accessible over changes of technology including software.'[6]

---

2 Heslop et al., *Preservation of Digital Records*, p. 5.

3 NAA, *Submission No. 22*, p. 2.

4 Heslop et al., *Preservation of Digital Records*, pp. 5-6.

5 Heslop et al., *Preservation of Digital Records*, pp. 5-6.

6 FaCS, *Submission No. 21*, p. 7.

Archival Integrity

5.8 In addition to the problem of insufficient importance being given to electronic records, there are two additional factors that make the long term preservation of electronic records difficult: media degradation and application obsolescence.

Media Degradation

5.9 The various types of information storage media all degrade over time. Magnetic and optical media, such as floppy disks, tape cartridges, CD ROMs and DVD ROMs, have an archival life of about 20 years.[7] After this time, the information becomes less and less readable and may be lost entirely. The NAA Green Paper commented that these storage media '… decay relatively rapidly compared to other media. They are not designed for long term use and are therefore extremely susceptible to short and medium term decay.'[8]

5.10 Alternatively, data may become unreadable through technological obsolescence, when the hardware no longer exists to read the media used for storage. As an example, data stored on 5¼ inch floppy disks is becoming less and less accessible, because few modern computers are equipped with 5¼ inch disk drives.

5.11 NAA noted that the pace of market driven innovation means that '… without the intervention by archivists to preserve the source and process, the performance cannot be guaranteed.' It also commented, however, that:

> The problems of decay and obsolescence do not make the job of preserving digital material impossible. … As long as the essential parts of the performance can be replicated over time, the source and process can be replaced.[9]

5.12 One solution to media degradation is to store all archival data in live storage. This involves copying the data to a modern storage device, such as a hard drive or storage silo. If the chosen device is regularly maintained to prevent degradation, the data stored on it will remain accessible as long as the device remains in working order. This method requires that, periodically, when the storage device is upgraded, all data must be transferred to the new storage device.

---

7 AUUG Inc., *Submission No. 13*, pp. 7-8.

8 Heslop et al., *Preservation of Digital Records*, p. 11.

9 Heslop et al., *Preservation of Digital Records*, p. 11.

Application Obsolescence

5.13 The pace of computer software development is such that each version of a program, such as a word processor or a spreadsheet, is rapidly replaced by a new version.[10] For users, the problem in this process is that most vendors only support the superseded format for a limited time. This can occur through the policy decisions of software companies or through the failure of the company itself.

5.14 The result of this process of obsolescence is that users need to establish a comprehensive forward plan to prevent the inadvertent loss of valuable data.

5.15 One technique would be to use a long-term storage format; a solution already under consideration by a number of Commonwealth agencies. The format adopted would be selected both for its suitability and the expectation that applications capable of reading its files would still be available in 50 to 100 years. All information to be archived would then be converted into the chosen format as part of the standard archiving procedure.

5.16 An advantage of long-term storage formats is that because their rules are known and freely available, anyone can write an application that can read the files. If no such application existed, then a new one could be written from the specifications.

5.17 The National Archives has chosen XML (eXtensible Markup Language) as its long-term storage format. This format is open and non-proprietary, so many applications exist that can read it, and new applications can be written at any time.[11]

5.18 The Victorian Electronic Records Strategy (VERS) has chosen Adobe's Printable Document Format (PDF) as its long-term storage format.[12] This program was created by a private company but it is an open format, with its standards freely available and widely known. Hence, it has the same sort of advantages as XML. The differences between PDF and XML are easily dealt with because each format can be converted to the other without loss of information.

5.19 All agencies should consult NAA, to ensure that that their plans are compatible with the national archival plans for Commonwealth data. Of

10 Heslop et al., *Preservation of Digital Records*, p. 11.

11 Mr Stuckey, *Transcript*, 1 April 2003, p. 98.

12 Victorian Electronic Records Strategy, *Final Report*, Chapter 3, p. 16, http://www.prov.vic.gov.au/vers/published/final/final3.pdf, 28 October 2003.

necessity, this would include arrangements for the long-term storage of data in a format that:

- will continue to be accessible for the foreseeable future; and
- uses an information storage medium that will:
  - (i) also remain accessible for a long period, and
  - (ii) allow a simple transfer procedure when a new medium becomes necessary.

5.20 The Committee notes the NAA use of open source and notes the reason for using it; namely the expense associated with maintaining a long term licence if the NAA was reliant on proprietary software and accessibility.[13]

Recommendation 7

5.21 **The Australian Government Information Management Office (AGIMO), with support from the National Archives of Australia (NAA), ensure that Commonwealth agencies implement knowledge management and archival policies such as e-permanance which give equal priority to preserving electronic and paper-based records. AGIMO to advise the Committee, in an Executive Minute, of the status of these arrangements. The NAA to be resourced properly.**

Business Continuity and Disaster Recovery

5.22 In addition to preparing defences against attacks on electronic data systems and degradation of records over time, it is also important for each Commonwealth agency to have a program in place to quickly re-establish normal operations after events such as fires or earthquakes.

5.23 In an audit report released in July 2003, the ANAO gave attention to the current state of preparedness in Commonwealth agencies. Its findings indicated that although most agencies were aware of the necessity for such preparations, the actual arrangements are, in most cases, far from complete. The report commented:

> Many of the entities reviewed either have no business continuity plan, or are in the process of developing one … However their ability to recover from a disaster may be limited.[14]

13 Mr Stuckey, *Transcript*, 1 April 2003, p. 98.

5.24 The ANAO report made the assessment that '… only about 30 per cent of the entities reviewed had a business continuity plan, a disaster recovery plan or use high availability equipment to deal with the threat'. In other cases, where plans do exist, ANAO found that:

> … there is insufficient scope or detail in the plans to be able to conclude confidently that they could recover from a disaster. In addition, some … have not tested these plans to ensure that they can recover from a disaster.[15]

5.25 ANAO itself reported that its '… Disaster Recovery Plan was reviewed in February 2002 and incorporated in an updated Business Continuity Plan…' The plans were tested by external consultants at the end of that year and recommendations from that review accepted and put in place. A further check by ANAO's internal auditor, confirmed that the plans were an appropriate way to minimise threats to the data it held.[16]

5.26 At the ANAO, data on the servers is backed up daily and the tapes stored off site. In the event of a disaster, the contractor is required to provide backup facilities at its premises for the restoration of ANAO's operations. This plan restricts the potential loss of data to a maximum of one day.[17]

5.27 Similarly, NOIE advised that its data is backed up and a copy stored offsite by TES.[18] The issue of the security of information stored offsite was addressed in Chapter 2.

5.28 As an example of how Commonwealth departments addressed this issue in response to this inquiry, FaCS told the Committee that it had developed a Business Continuity Framework to manage and recover from major disruptions. The Framework includes a command structure, recovery teams, high level strategies and detailed plans and procedures to cover key risk areas.[19]

5.29 The IT area of FaCS has a disaster recovery plan as part of a Business Continuity Management Project. This plan has been subjected to disaster scenario testing, to identify weaknesses in the system. A number of

---

14 ANAO, *Control Structures as part of the Audit of Financial Statements of Major Commonwealth Entities for the Year Ending 30 June 2003,* Audit Report No. 61 2002-03, 30 June 2003, p. 71.

15 ANAO, *Control Structures as part of the Audit of Financial Statements of Major Commonwealth Entities for the Year Ending 30 June 2003,* Audit Report No. 61 2002-03, 30 June 2003, p. 71.

16 ANAO, *Submission No. 42,* p. 7.

17 ANAO, *Submission No. 42,* p. 7.

18 NOIE, *Submission No. 60,* p. 4.

19 FaCS, *Submission No. 21,* p. 6.

recommendations for improvement arising from these tests will be addressed by the department.[20]

5.30 The evidence before the Committee indicates that, although some agencies have put adequate preparations in place, taken as a whole, Commonwealth agencies are not well prepared to cope with large scale interruptions to, or losses of, their IT capacity.

5.31 The Committee believes that all Commonwealth agencies should assign a high priority to the completion of comprehensive plans for business continuity and disaster recovery. DSD has offered its assistance in preparing such plans and agencies should take advantage of that assistance.

Recommendation 8

5.32 **The Australian Government Information Management Office (AGIMO), in consultation with the Australian National Audit Office, ensure that Commonwealth agencies have in place comprehensive and tested business continuity and disaster recovery plans for their electronic records networks and services. AGIMO to advise the Committee, in an Executive Minute, of progress with the implementation and testing of these plans.**

20 FaCS, *Submission No. 21*, p. 6.