

# Tales from “THE disK FILES”: Lessons Learnt from a Data Recovery Project in 2003–2006 at the National Archives of Australia

David Pearson, FSA, and James Doig

## ABSTRACT

This case study re-evaluates a large-scale project carried out by the National Archives of Australia (NAA) between 2003 and 2006. The project aimed to identify obsolete digital media (physical data carriers) in its collection and to describe and recover the data from the carriers using a third-party data recovery provider.<sup>1</sup> A detailed process for data recovery was developed that included the capture of a full audit trail of steps in the data recovery process. The project was completed in four stages: phase 1 obtained bit-level images from the carriers; phase 2 extracted individual bit-files from the carriers; phase 3 identified duplicate files and proprietary or complex file formats; and phase 4 was a final report that documented processes, made recommendations on future processes, and provided lessons learned. Recent work described in this article indicates that files extracted from the carriers in 2004–2005 can be accurately rendered in current computer environments. The ongoing significance of the project is that it is an early demonstration of the success of bit-level preservation and the need to create disk images as part of a preservation workflow, suggesting a sustainable methodology for digital preservation. The project also influenced archival policy at the NAA and influenced the development of subsequent software tools that became widely known in the broader digital preservation community. The focus on archival principles of authenticity, integrity, chain of custody, and provenance of the recovered records were key learnings to ensuring long-term access and usability. Finally, the metrics resulting from the project, for example, rates of readable carriers and rates of data recovery by carrier type, are useful data from a point in time that correspond quite closely to similar data recovery projects undertaken by other institutions at about the same time and provide a benchmark for future research.

© David Pearson, FSA, and James Doig.



## KEY WORDS

Digital preservation, Digital media, Physical carriers,  
Legacy carriers, Obsolete carriers, Data recovery

## 2003

*A pair of shadowy figures, clothed in thick, fleece coats to protect them from the cold, pull open the sliding air-lock door and enter the cold room vault. They nevertheless feel the near-zero temperature as they venture into a large, open space and peer into the gloom. One of them flicks a light switch, and rows of fluorescent bulbs illuminate compactus shelving on either side of a narrow corridor leading away into the depths of the repository. Moving past the large, heavy-duty drawer units that contain hundreds of microfilm reels, the figures move further into the vault, conscious of the unsettling stillness and silence around them. The focus of their investigation is an open row of shelves that contain an odd assortment of differently shaped boxes, each with strange hieroglyphics written on its side, in fact, the numbering system controlling the boxes in the series. They open the first box, a square, flat one that looks like a pizza box, and take out a large plastic reel with a scribble of barely decipherable writing on an age-darkened label that is starting to peel away. They huddle over the reel and make some notes in an exercise book, including a transcription of the writing on the label. The next box is a standard archival container that is packed full of thin, square plastic objects in paper sleeves. For the rest of the day, they carefully open each box in the compactus bay and make detailed notes of their contents—tapes, disks, cartridges, of all shapes and sizes. Finally, their work finished, they head out through the air-lock into the warmer air.*

*“You know, Dave,” the shorter one says, pulling a handkerchief out of the pocket of his corduroy trousers and cleaning his spectacles. “I wonder how much, if any, data we’ll be able to recover from this stuff?”*

*Dave shrugs. “Who knows Brendan? But there’s going to be an awful lot of this work to do in the future, so we’d better get it right this time.”*

In telling this story of shadowy figures, lost archives, and secret vaults, we will take you on a trip into the past to a time when digital preservation theory and practice were still in their infancy, when universally agreed digital preservation principles and workflows were still being developed, and before familiar standards such as PREMIS<sup>2</sup> and OAIS<sup>3</sup> were developed.

Like other government archives, the NAA collects official Commonwealth government records and the personal records of significant individuals closely connected with the government in an official capacity, such as governors-general and prime ministers. The NAA holds over 365 kilometres (226.8 miles) of physical records and over 2 petabytes of digital material (primarily AV material), which is growing rapidly as a result of large-scale digitization work. Some of this digital material is held on obsolete physical carriers such as floppy disks, magnetic tapes, and data cassettes. As long ago as 1991, it was recognized that managing digital

material on carriers in an offline storage environment poses a serious risk to ongoing access.<sup>4</sup> Published research such as that by Rothenberg<sup>5</sup> and Garrett and Walters<sup>6</sup> provided further impetus to address risks of carrier and file format obsolescence.

In 2003, the NAA commenced a project to identify digital content on obsolete carriers and describe and recover the data from them. A detailed process for data recovery was developed that included the capture of a full audit trail of steps in the data recovery process to ensure fixity, provenance, authenticity, and the chain-of-custody for archival management.<sup>7</sup> The data recovery project was classified into a four-phase process: phase 1—obtain bit-level<sup>8</sup> disk images<sup>9</sup> of all of the content on each physical carrier; phase 2—extract individual files from each of the physical carriers; phase 3—analyse and identify duplicate files and proprietary or complex formats; and phase 4—document the results for future archival reference and preservation processes.<sup>10</sup> An additional fifth phase was proposed but not enacted at the time, which was to investigate and use appropriate software to render or display the files recovered in phase 2 or disk images from phase 1 if the files were unrecoverable. These steps are described in detail in the Methodology section following. While the project has been referenced in a number of published articles since the project was completed, this article is the first detailed description of the scope, methodology, results, and lessons learned.<sup>11</sup>

Recent work by the NAA has demonstrated that some of the files extracted from obsolete carriers as a result of this project, and subsequently stored in a preservation system, can be accurately rendered in current computer environments. A key message is that recovering the bits and ensuring they are properly cared for when it is still possible to do so will lead to positive outcomes that may not be fully realized for years into the future.

## Agency to Researcher Project

In 2002–2003, as part of a broader digital preservation project called the Agency to Researcher Project, the NAA commenced a number of research studies designed to inform its overall digital preservation approach, to test its assumptions, and to fully understand the environment in which the archive was operating.

The research studies comprised

- A report setting out the conceptual understanding of digital records that form the basis of the NAA's approach to digital preservation. The result of this work was the digital preservation Green Paper, *An Approach to the Preservation of Digital Records*;<sup>12</sup>
- The design and construction of a purpose-built digital repository, open-source XML normalization software (the software was called Xena, which stands for XML normalizing of archives), and workflows to ingest digital records;<sup>13</sup>

- An investigation of researchers' expectations of preserved digital records;<sup>14</sup>
- A test transfer of digital records from a Commonwealth government agency (Australian Wool Research and Promotion Organisation);<sup>15</sup> and
- A project, formally named the Legacy Media Project, to identify existing digital records on legacy media (physical carriers) already in the NAA's custody and to make those records accessible to researchers.<sup>16</sup>

One of the objectives of the Legacy Media Project was to develop a methodology for recovering data from legacy physical carriers and to implement that methodology on known legacy carriers in the custody of the NAA. Many of the records on the carriers related to high-profile public inquiries, such as Royal Commissions and Commissions of Inquiry, and so had high secondary value.<sup>17</sup> Within the NAA at the time, the speculative expectation was that about 30% of data could be recovered from carriers dating from the 1970s, 1980s, and early 1990s, even though they were stored while in NAA custody in environmentally controlled repositories. It was unknown if the data or the carriers themselves, especially the 9-track ½" magnetic tapes, had become compromised and could be read more than once due to

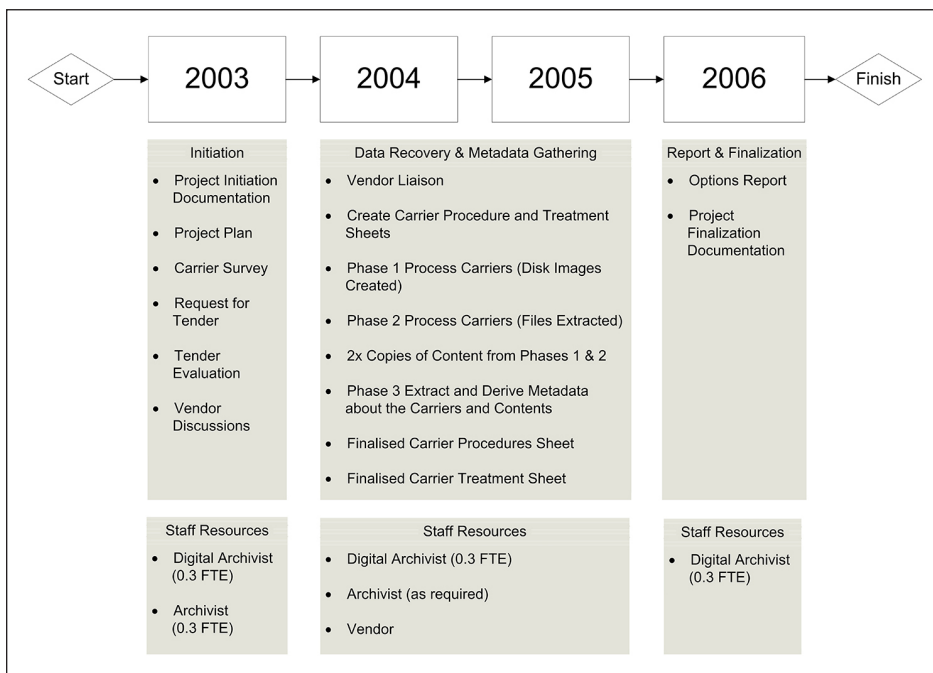
- Obsolescence of the carrier type;
- Decay of the data on the carriers;
- Obsolescence of the hardware/software mechanisms to access the data on the carriers;
- Incorrect storage resulting in the failure during of the data recovery process from deterioration such as "stiction," where the tape substrate had bonded together causing friction resulting in the stretching and/or breaking of the tape(s).<sup>18</sup>

It was also unknown how much of the material on the physical carriers was duplicated in paper form or if it was all original archival records. Although uncertainties existed about whether outcomes could be achieved, the NAA decided to undertake the project, first, because doing nothing was not an option because the records were high value and leaving them on the carriers risked record loss, and second, because the project provided the opportunity to test hypotheses and to develop workflows for digital archiving, storage, and long-term preservation.

## Legacy Media Project

The project team consisted of two staff, an archivist, and a digital archivist, each working on the project at 0.3 full-time equivalent. The archivist carried out the collection survey to identify obsolete carriers for recovery treatment at the beginning of the project in 2003 and was available for consultation on a needs basis afterward. The digital archivist managed the project from 2003 to 2006, including developing project documentation and controls, establishing the contract with the successful vendor, liaising with the vendor, and overseeing data recovery and quality control, as

well as developing the control and recording documentation. The project consisted of a number of phases in which the data stored on obsolete carriers was progressively extracted from the physical carriers and a detailed description of each phase produced. In this way, each phase resulted in not only the recovered data, but also a complete record of the recovery process, including establishing a fixity point and a digital verifiable chain of custody. Work on the project came in peaks and troughs: 2003 consisted of the project initiation, research, and tender process; the recovery work took place between 2004–2005; and the lessons-learned documentation was produced in 2006 (see Figure 1). At the time the project was initiated in 2003, external (vendor-developed) data recovery processes were highly proprietary and the domain of specialist operators, and there was limited in-house expertise, capacity and equipment at the NAA. Besides pioneering work such as those by Woodyard<sup>19</sup> and Ross and Gow<sup>20</sup> on data recovery, there was little literature on how to conduct a data recovery process in the GLAM sector (in contrast with digital forensics in the legal domain). In the years since, the increase in published accounts of data recovery projects and the development of open-source tools discussed below<sup>21</sup> have been exponential. A considerable amount of work has been published on the application of digital forensics in the GLAM sector.<sup>22</sup> However, the pioneering work at the NAA is notable because the lessons learned contributed to the requirements for a number of purpose-built preservation workflows to manage some elements of data recovery from physical carriers, including Prometheus<sup>23</sup> and BitCurator,<sup>24, 25</sup> and the



**FIGURE 1.** Timeline of the project phases, timeframes, deliverables, and staff

development of purpose-built knowledge bases on carriers and file formats and their dependencies such as Mediapedia,<sup>26</sup> as well as the National Library of Australia’s Digital Preservation Knowledge Base.<sup>27</sup> The need for institutional knowledge bases and registries continues to be relevant as indicated by recent projects undertaken by the Social Sciences and Humanities Research Council of Canada and the University of Illinois at Urbana-Champaign Library.<sup>28</sup>

Audit of Legacy Carriers

The initial phase of the project involved an audit of legacy physical carriers in the NAA collection.<sup>29</sup> At the commencement of the project, the full spectrum of carrier types was not known and the audit attempted to identify all digital carriers existing in the collection. The audit was carried out by querying the NAA’s descriptive catalog, an in-house developed archival management system called RecordSearch, for terms such as “disk,” “tape,” and “floppy” and by querying the series-level descriptor, Predominant Form, with the attribute “electronic record,” which picked up series whose predominant physical form was a digital carrier. Other sources of information, such as transfer documentation, were also checked. This work was carried out by an archivist, and the results, including any descriptive metadata captured in the catalog, were tabulated for action. Three hundred carriers were identified (see Table 1 and Appendix A, Figures 1–4). Even so, without a full, physical survey of the collection, it was impossible to confirm that all obsolete carriers were identified.

Relevant information such as series and item titles, date range, security classification, location, and any technical information was recorded in a register. Notable was the lack of technical information about the carriers, for example, almost no information about the creating application was recorded for any series, presumably not seen as relevant at the time of acquisition. Following the initial data collection, the carriers were physically checked, and any metadata or information located on the outside or stored with the carriers was recorded in the audit checklist (see Figures 2 and 3).

The audit captured as much descriptive information as possible both from the descriptive catalog and any information stored with the carriers, such as labels or other markings.

Table 1. Summary of the Carrier Types Used in the Project

Carrier Type	3½" Macintosh Floppy	Burroughs B20 5¼" Floppy Disks	Data Cassettes	Wang 8" Floppy Disks (Wang DOS)	Wang 8" Floppy Disks	9-track ½" Magnetic Tapes
Number	60	32	20	22	49	137

*Note: Information on Wang DOS, Wang OIS, and Burroughs B20 formatting was only available after data recovery and was not captured in the original audit.*

Downloaded from https://prime-pdf-watemark.prime-prod.pubfactory.com/ at 2025-06-28 via Open Access.

To reduce costs, a conscious decision was made at the outset to exclude digital materials described in the catalog as "backups," as well as personal records collections, that is the official and private records of significant individuals who served within, or were closely associated with, the Australian Commonwealth government. Nevertheless, some of the materials eventually recovered from the carriers were subsequently found to be backups or duplicates of paper records, once again a failure of the transfer process to record this information when the knowledge of it was readily available. Twenty data cassettes dating from the late 1970s were also identified in the audit, however the source equipment required to deal with these types of carriers could not be sourced at the time of the project.

As the capability to undertake the work in-house was limited, the NAA issued a request for quote (RFQ) process for recovering data from the identified physical carriers. The RFQ outlined the proposed methodology for data recovery and also stipulated that two copies of the recovered data would be burned to Mitsui Gold brand 650Mb CD-Rs (optical discs), an industry standard at the time. One vendor was selected in 2003, and the data recovery was carried out between 2004 and 2005 in three phases. Subsequently, the recovered data were ingested into the NAA's digital archive when it came into production in 2007. Since that time, additional legacy carriers have been identified, either transferred to the NAA at a later time, or not identified in the original audit, having been obscured/hidden by information disassociation (one of the ten agents of deterioration/change).<sup>30</sup> In addition, over 250 5.25" (5¼") and 3.5" (3½") floppy disks were identified in personal record collections at the time of the project but were excluded because seeking the agreement of the donor or the donor's estate could be time consuming. These floppy disks remain in personal records collections and still require data recovery for the data contained on them to be usable.

## Methodology

The methodology adopted was a "belts and braces" approach designed to mitigate against the expectation of data loss resulting from the perceived instability of the obsolete carriers. A cautious approach was therefore adopted: in phase 1, a disk image was taken of the whole contents of the carrier (which included all the data on the disk, including unwanted and possible deleted content which had not been overwritten); and, in phase 2, access to the file system allowed the individual files (that were identified for recovery) to be extracted from their individual carriers. This process resulted in two copies of the same content (the disk images and the files), as well as backup copies of each. The proprietary processes developed by the provider, and the equipment and software used to extract the data, were recorded for each of the carrier types on Carrier Treatment Procedure Sheets, and the results of each of the treatments were recorded on a Carrier Treatment Check Sheet. The treatment



procedure sheets and check sheets were very detailed templates developed by the project team that captured full treatment data to be able to prove the authenticity, integrity, chain of custody, and provenance of the recovered records.

## DOCUMENTING THE RECOVERY PROCESS

The data captured on the Carrier Treatment Procedure Sheets and the Carrier Treatment Check Sheets were determined by the project team using a risk-based approach: more data captured about the process would reduce the risk of the evidential value of the records being questioned in the future. The procedure sheet for each process provided a full inventory of the hardware, software, and propriety processes used by the provider, including computers, hard disk drives, operating systems, network details, emulators, checksum algorithms, and so on, as well as a step-by-step description of the treatments for each phase of the project. Some of this information was proprietary to the data recovery vendor. This detailed metadata and descriptive information was designed to be ingested at a later date into a digital preservation or archival management system.

The check sheet was a spreadsheet listing each digital file/object recovered with descriptive and technical metadata, such as date recovered, operator, series, carrier ID, carrier label (if any), carrier type, carrier density, file name, file size, object type (i.e., format), character encoding (if known), and checksum created at the time of recovery, effectively the *terminus post quem* for proving fixity. These data were used in 2020 to confirm the integrity of the files before some of them were examined. These sheets enabled a consistent stratigraphic view of the relationship of the file system, the file tree, and individual files on each of the carriers. Interestingly, the metadata captured in the check sheets and procedure sheets later maps quite closely with elements of PREMIS, for example Object Characteristics, Environment, and Storage Medium.<sup>31</sup>

This thorough documentation of process constituted an audit trail of the recovery treatments and was a necessary activity to prove the authenticity and integrity of the recovered information, useful then and for future reference.

## PHASE 1

The aim of this phase of the project was to obtain exact bit-level images of the contents of each carrier. The disk images were created as a precautionary measure due to the age and potentially degraded state of the physical carriers and their data. To mitigate the risk that a carrier might fail during or just after the first attempt to access the data, this process created a whole disk image of each carrier using a one-read process, which was subsequently copied onto the CD-Rs.



Notwithstanding concerns about the age of the carriers and storage conditions, after the disk imaging process was completed, the legacy carriers were found to be quite stable. The results of this process identified

- 257 (86%) carriers with 100% data recovery;
- 14 (4.7%) with system or known duplicate data;
- 13 (4.3%) with partial recovery; and
- 15 (5%) failed the process (see Table 2 and Appendix B).

**Table 2. Summary of Results after Phase 1 by Carrier Type**

Results description	Total carriers with 100% reads	Total carriers with 100% reads but identified as system or duplicate data	Total carriers with partial reads or errors	Total carriers with 100% fail
Total number and percentage	257 (86%)	14 (4.7%)	13 (4.3%)	15 (5%)

## PHASE 2

The aim of phase 2 was to extract individual files from their carriers into a format more acceptable for storage and future access. This phase consisted of copying all the viewable (not hidden) data objects that were able to be recovered from their respective original carrier and copying them onto Mitsui Gold brand 650Mb CD-Rs. As in phase 1, two CD-Rs containing recovered data—a master and a copy disk—were obtained.

As a result of the redundancy gained from phase 1, the process of extracting the native file system and contents using a multiple read process upon each carrier could be conducted with less concern for damage to the original carrier (i.e., the bit-level image copies of all carriers provided redundancy). Another result of phase 2 was that some of the file systems and digital objects could be examined, although most could not be opened due to inherent software dependencies. After the file extraction process was completed, the finer granularity resulted in a different outcome compared with the results obtained in phase 1 (see Table 3 and Appendix C, Tables 1-3). The results after phase 2 were

- 245 (81.9%) carriers with 100% data recovery;
- 20 (6.7%) system or known duplicate data;
- 14 (4.7%) with partial recovery;
- 14 (4.7%) found to be blank; and
- 6 (2%) failed the process.

Table 3. Summary of Results after Phase 2 by Carrier Type

Results description	Total carriers with 100% reads	Total carriers with 100% reads but identified as system or duplicate data	Total carriers with partial reads or errors	Total carriers which were found to be blank	Total carriers with 100% fail
Total number and percentage	245 (81.9%)	20 (6.7%)	14 (4.7%)	14 (4.7%)	6 (2%)

Comparison of results between phases 1 and 2 is shown in Figure 2. As mentioned, the difference between the results obtained in phases 1 and 2 is due to the granularity of the processes, operator observations, and access to the file system. Differences in the results were also due to

- More carriers being identified as containing system data or duplicate data;
- More carriers that were initially identified as failing but on investigation of the file system were subsequently found to be blank. If a disk was formatted but contained no data, it was still imaged in phase 1; and
- Lack of clarity about what the vendor defined as a carrier failure in phase 1, when in phase 2 some data were found to have been recovered from the “failed” carrier. It was subsequently apparent that data had been partially recovered from Series AA1979/319/0<sup>32</sup> and C379p1<sup>33</sup> before the carrier failed.

Even considering these factors, the results were of an order of magnitude better than expectations at the start of the project that only 30% of data could be recovered.

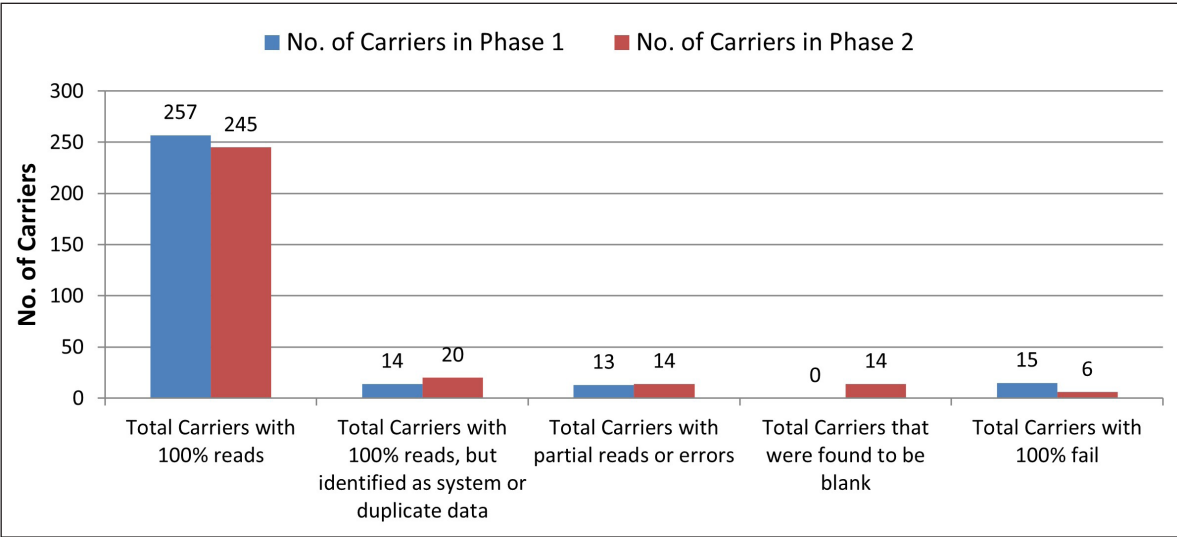


FIGURE 2. Comparative results after phases 1 and 2

PHASE 3

The aim of phase 3 was to examine the results obtained from phase 2 and cull duplicate files, system files, and blank carriers. This phase also included determining other problems such as the presence of data objects not identified in the transfer documentation (e.g., data that were found on the carrier in addition to the records identified for transfer to the NAA). Problem formats such as proprietary Landsat<sup>34</sup> data formats were also identified. On the basis of file name, file type, and the operator comments recorded on the Carrier Treatment Check Sheets, decisions could be made about triaging digital objects for preservation actions (see Table 4 and Appendix D). It should be noted that, at this stage, most of the individual file content had not been examined in any detail. However, as much of the content, especially on the older carriers, was encoded in ASCII<sup>35</sup> or EBCDIC,<sup>36</sup> some of the content could be accessed at the time easily with simple text editors and proprietary file analysis software used by a vendor called InterMedia.<sup>37</sup> Therefore, while the work undertaken in phase 3 demonstrated that the recovered data were usable, it was understood that further work was required to effectively render the data and provide access to them. In particular, data encoded in proprietary database formats and unknown geophysics formats, such as the Landsat data, would require further analysis to understand and identify options for accessing the recovered files.

**Table 4. Summary of Results from Phase 3 by Potentially Usable Files**

Carrier type	3½" Macintosh Floppy Disks	Burroughs B20 5¼" Floppy Disks	Wang 8" Floppy Disks (Wang DOS)	Wang 8" Floppy Disks (Wang OIS)	9-track ½" Magnetic Tapes
Amount of data	21.896 MB	18.514 MB	4.694 MB	5.492 MB	5111.17 MB
Potentially usable files	547	100	56	127	408
Identifiable nonarchival records/ files	47	Nil	Nil	Nil	521 Header/footer & 129 other files

The results from phase 3 also revealed the magnetic tapes contained identifiable information on each carrier, such as header and footer files, used by the original creating/accessing software, that did not affect the meaning of the content or the ability to access the content using current software applications. There was debate within the NAA about whether this information needed to be retained or securely disposed of. Given the uncertainty surrounding the future research value of this information, and probable technical dependencies to access and render the data,

the decision was made to retain it. It is also worth noting the total amount of data recovered from the process was relatively small, and the amount of data recovered from more modern carriers was likely to be exponentially larger.

#### PHASE 4

The aim of phase 4 was to document processes, make recommendations, and provide lessons learned for future archival workflows at the NAA. The outcome of this phase was an internal NAA report, "Options Paper to Determine How to Proceed with the Legacy Media Project,"<sup>38</sup> which provided a good deal of analysis and statistics of methods and outcomes, which have formed the basis of this article. It also provided recovery costs per carrier and per gigabyte, which provided a basis to determine resource requirements for future data recovery projects. Given the high cost associated with relatively small volumes of recovered data, the NAA developed transfer requirements for digital records that prohibit the transfer of legacy or obsolete carriers.<sup>39</sup> As part of a broader management regime for digital recordkeeping and information management, agencies will need to decommission systems and migrate data in a timely way to prevent software and carrier obsolescence from occurring.<sup>40</sup> Another outcome of the phase 4 work was that the amount of detail required to populate the procedure sheets and check sheets was too resource-intensive for a human operator and that machine-generated technical metadata was the preferred option going forward.

#### Results

The results of each phase of the project are tabulated in Appendixes B, C, and D. They indicate a high level of success for recovering data from carriers of this age and type. Broadly, of the 300 carriers treated, 257 (86%) achieved 100% reads in phase 1 (i.e., disk images were obtained), and 245 (82%) achieved 100% reads in phase 2 (i.e., complete digital object recovery). Partial recovery of files was achieved in about 5% of cases. A similar result was found in a slightly later project at the British Library involving data recovery from 8", 5¼", and 3½" floppy disks: "... there have been relatively few cases where disks have been entirely unreadable: occasionally degradation can be seen in the physical condition of the disk, ie a light reddish brown surface indicative of oxidation."<sup>41</sup> Additionally, in the case of the NAA project, analysis of the recovered data indicated that about 5% of the carriers were blank, and about 7% of the carriers contained data duplicated in another form, such as paper printouts. Not surprisingly, the blank carriers were from personnel computers and not part of a centralized corporate IT function and showed that they had not been examined by the agency or the NAA before taken into custody.

The 8" and 5¼" floppy disks achieved 100% data recovery. Data on seventeen 3½" floppy disks could not be recovered, while twelve 9-track ½" magnetic tapes could not be read. These carrier failures belonged to two series,<sup>42</sup> and the carrier degradation may relate to how those series were managed and stored prior to transfer to the NAA. Luckily, none of the disks with spanned backup data had failed; if any had, it would have rendered any future recovery process most likely impossible, given that all the data on each consecutive carrier are required in the order of the original backup process. However, rendering these data is still problematic, as without access to the original backup software (if known and/or available), the bits currently cannot be rendered in a meaningful way.

## RENDERING THE RECOVERED DIGITAL FILES

A fifth phase of the project was envisaged but not acted on at the time due to changed organizational priorities. In this phase, rendering or interpretation software was to be identified and used to obtain a usable copy of the files obtained in phase 2 that could be characterized, preserved, and rendered by the NAA digital preservation software. Although phase 5 was not enacted, the fact that the disk images and files were appropriately preserved and stored to ensure their integrity means that this phase can be commenced at any time.

For example, in early 2020, the NAA revisited the Landsat data recovered from two 9-track ½" magnetic tapes, which were part of the 1983 Royal Commission into the use and effects of chemical agents on Australian personnel in Vietnam (see Figures 3 and 4).<sup>43</sup> Figure 4 shows one of the bit sequences from the files opened in a modern hex editor, but very little information about the format can be extracted. These recovered files were sent to the US Geological Survey (USGS), the organization responsible for Landsat, which was able to extract the image data and display them using ENVI,<sup>44</sup> widely used image analysis software. Note that the recovered images are not perfect, as there is a significant offset between band 1 and the other bands that the USGS was not able to resolve, and there were "garbage" artifacts of some sort on the ends of each line. Nevertheless, as shown in Figure 5, the images were recoverable, and the individual bands exported as TIFF files, which can also be preserved and accessed. This example shows that files recovered during phases 1 and 2 in 2004–2005, properly stored and managed, can be effectively rendered by analysis and rendering software in 2020. Extracting the bits from obsolete carriers and processing them through robust digital preservation processes has allowed data to be readily available forty years after it had been created in a completely different technological access environment. Today, there are open-source preservation workflows like BitCurator<sup>45</sup> and hardware like Kryoflux,<sup>46</sup> but these technologies do not allow access to the original carriers without some form of working hardware interface and readable carrier (e.g., bits cannot be removed from physical carriers without the correct access technology), however one might wish it not to be the case!

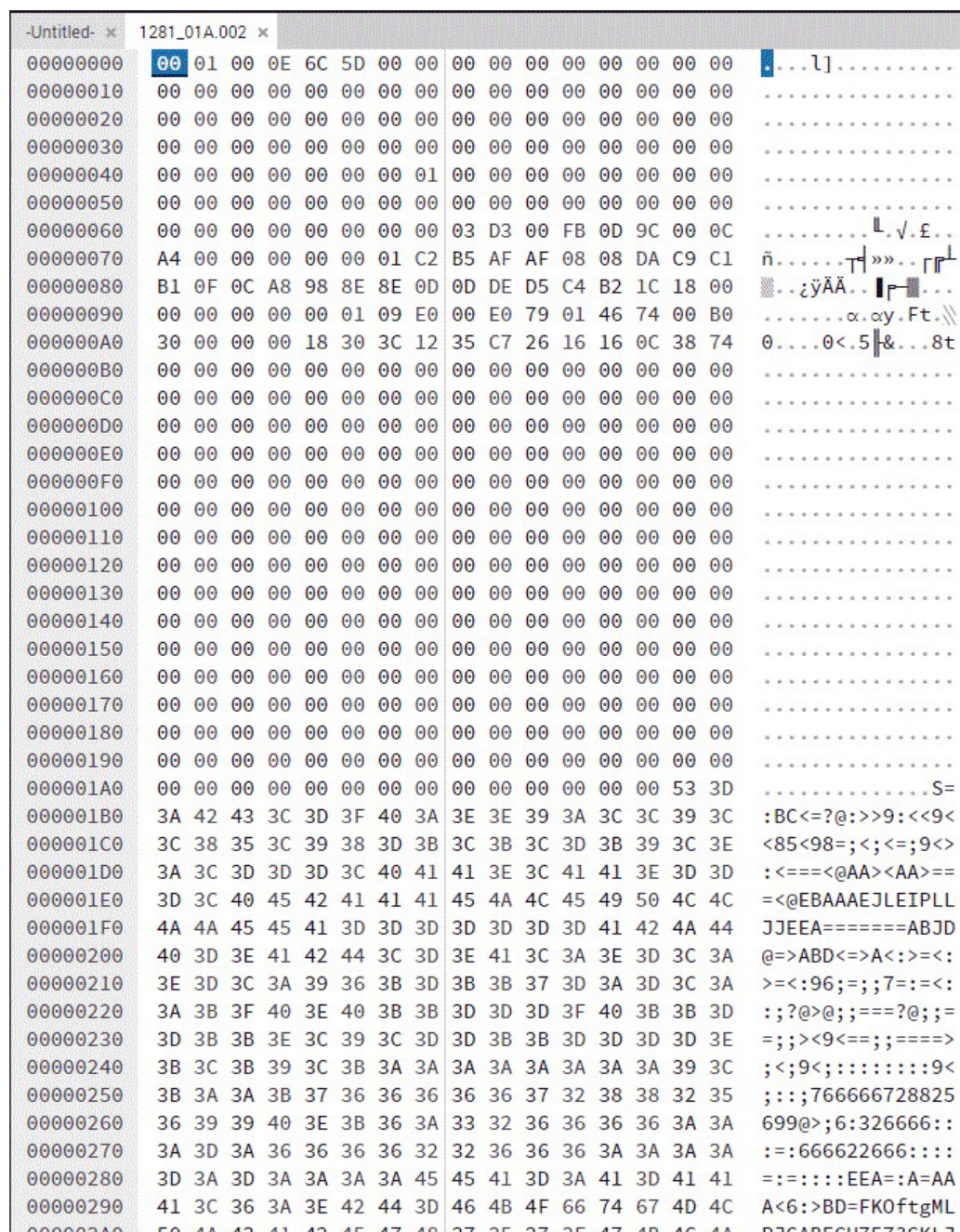


**FIGURE 3.** An image of the 9-track 1/2" magnetic tape that stored the original data (scale in cm). Descriptive metadata labels on the tape include the NAA accession numbers "C1281\_2" and "CRS C1281"; information about the collection of the data on the tape (for details of the tape label, see Figure 4); and technical metadata about the manufacturer and the carrier type and capacity: "Wabash Quadronix 1" and "Certified For Use Up To 6250 bpi (bits per inch)." This type of information was collected and entered into the Carrier Treatment Check Sheets.

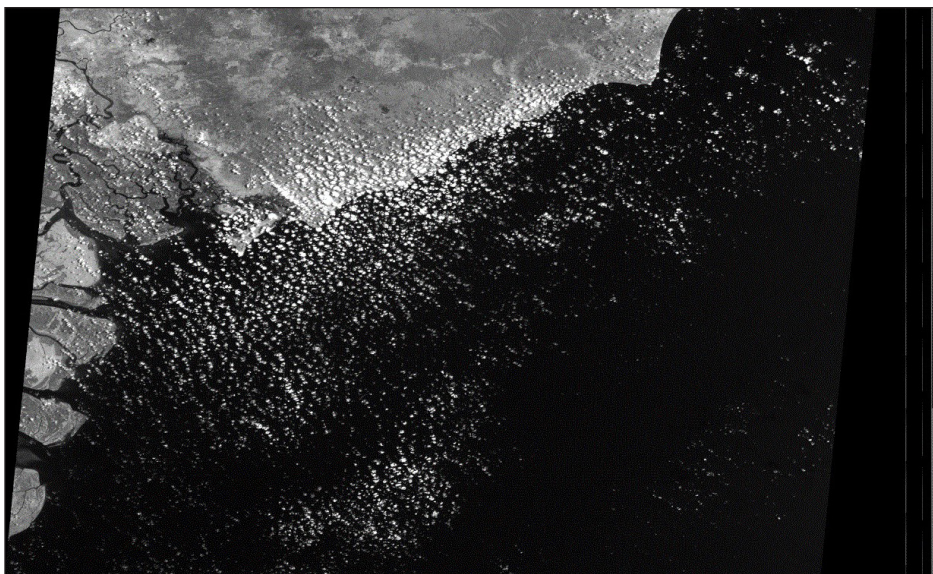


**FIGURE 4.** Close up of details on the faded tape label: "[undecipherable logo] THAILAND REMOTE SENSING CENTER/ LANDSAT 4 MSS BAND 4, 5, 6, 7, Data/ WRS PATH 125 ROW 53 DATE 01/06/83/ SCENE ID 4032002435/ FORMAT TYPE CCRS-AM/ TAPE DENSITY 9 TK 1600 PPI (pixels per inch) SEQ 1 OF 1/ 196 Phahonyothin Rd. Bangkok BKK 10 00/ Te 5790116 TEL X 82213 NRCTSD TH/ Date received from TLS 30/01/84." This type of information was collected and entered into the Carrier Treatment Check Sheets.









**FIGURE 6.** The same image in Figure 3, now presented as a rendered TIFF file in 2020 (image process by US Geological Survey)

Future work on the recovered files at the NAA could focus on emulation techniques to reconstruct the performance of the digital records when they were in active use, such as those proposed by the University of Freiburg's bwFLA and the current Emulation as a Service Infrastructure (EaaSI) Project using either the files or the disk images.<sup>47</sup> For example, files recovered from 9-track ½" magnetic tape relating to the Costigan Royal Commission into the Painters and Dockers Union (1980–1984) could provide important insights into early computerized records and information management practices.<sup>48</sup> Costigan was the first Royal Commission in Australia to use a computer information management system to manage and provide access to a wide range of investigative materials. The system allowed names and crimes to be cross-checked and referenced; and, according to Scott Prasser: "The Costigan Commission pioneered the use of computerised data to trace connections between personnel and transactions."<sup>49</sup> Very little information is available about this information management system; almost no information about the computer system is recorded in transfer or series documentation. However, this is a very fundamental archival expectation: future users of these records will want to understand how they were created, managed, and accessed, and how computerization contributed to what was an extraordinarily controversial public inquiry. Emulation may provide an opportunity to access and understand the recovered files in their original computer environment without altering the files and thus the fixity and provenance data.

The project also raised some important issues and had lessons for future policy development at the NAA.

## DESCRIPTIVE AND TECHNICAL METADATA

A key finding of the project was the need to capture detailed information about the carrier and its hardware and software dependencies at the point of transfer into the custody of the archive, for example, the specific descriptive and technical metadata, such as carrier type and version, required drives, operating systems, and other software dependencies.

The large amount of audit and integrity metadata recorded on the Carrier Treatment Procedure Sheet and the Carrier Treatment Check Sheet was resource intensive, and metadata acquisition was not very scalable without automatic tools and therefore costly. Although an audit trail of treatment is essential, there is a trade-off in cost, capture (supplied, extracted at the point of ingest, or derived from the file afterward), storage, and management and consistency of the metadata. In this case, although such metadata was developed well before current metadata standards were accepted, it still has high utility (and maps to the later standards). Standards such PREMIS may provide a baseline of essential audit metadata for projects such as this at the NAA, but they require fit-for-purpose tools and trustworthy registries of technical information to ensure preservation metadata is systematically captured and can be understood and used over time.<sup>50</sup>

A related finding was that the NAA's metadata standard for archival control, the Australian Series System,<sup>51</sup> was not designed for the management of digital carriers; additional metadata is required to ensure the accessibility of the digital content into the future. Metadata captured at the point of transfer must include additional information, such as the combination of software and hardware used to create and manage government data, as these data might not be extractable or derived from the file metadata. This information is essential for understanding the context of digital records when in active use and, if necessary, for data recovery and also for future access if, for example, an emulation strategy is used. It is worth noting that the development of descriptive standards in the software preservation domain highlights the need for similar technical properties to be captured by collecting institutions.<sup>52</sup>

Another key finding was that a significant amount of archival description work on the recovered data was necessary to facilitate discovery, to understand their context, and to document their relationship to other records, in particular analog records. In many cases, the carrier was transferred into the custody of the NAA with computer printouts and other paper records, but because the data could not be accessed, it was not possible to determine the relationships between different records, including whether or not the data were duplicated in paper form.

## COMMONWEALTH GOVERNMENT-WIDE ISSUE OF LEGACY DIGITAL CARRIERS

There was and continues to be an urgent need to understand the scope of the legacy carrier problem in Australian Commonwealth government agencies. No audit of legacy carriers in agencies has been undertaken, so very little information is available to quantify the risks of data loss.

Since 2011, the NAA has issued a series of rolling government-wide policies, in effect rolling five-year plans, to push government agencies along in digital transition, that include actions, targets, and pathways; online self-assessment kits; annual surveys; and other means to measure progress. The first of these, the Digital Transition Policy, was developed by the Department of Prime Minister and Cabinet, with the NAA as the lead agency, and released in 2011.<sup>53</sup> The Digital Continuity 2020 Policy was released in 2015, and the latest policy, Building Trust in the Public Record, came into effect in 2021.<sup>54</sup> The current policy emphasizes digital preservation and the risks associated with legacy information assets, including assets stored on obsolete or legacy carriers. A release schedule of products developed by the NAA for government agencies includes advice on identifying, managing, and disposing of legacy information assets.

## MANAGING LEGACY CARRIERS IN CUSTODY

Accessing data on legacy carriers is still an issue for the NAA. The project described in this article recovered data from 300 carriers (described in Appendix A, Figures 1-4), but excluded carriers in personal records collections and some carrier types, such as data cartridges. The current register of obsolete carriers includes over 250 items, mainly comprising 5.25" floppy disks in personal records collections. In addition, legacy carriers are still being found in paper files and will continue to be transferred to the NAA in this way. Data recovery tools such as Kryoflux, BitCurator, and others will probably form part of an in-house approach to deal with data recovery of remaining carriers in a more cost-effective way than the outsourced approach adopted in the 2003–2006 project.

## ACCESS AND DELIVERY

Providing meaningful access to the recovered data remains a pressing issue. Preservation actions, such as identifying suitable migration paths for the recovered files, even if the format and format version can be identified, in many cases may be difficult or impractical. For example, even records in formats created using early word processing software such as WordStar and Corel WordPerfect are not easy to render in modern software, and studies about conducting the preservation actions

necessary for accessing the content suggest that migration is also problematic (complex, costly, time consuming, etc.).<sup>55</sup> However, advances in scalable emulation services may provide a more viable means of meaningfully accessing complex data in obsolete formats over time.

Of course, if data are lost, access is not possible. The mechanisms and the extraction process of data from legacy carriers is like a game of Russian roulette, it may or may not work, or only partially work at the time of processing, and the process may destroy the carrier in some instances. However, the longer the material is left unattended on the carrier, the more problematic it will be. Recovering the bits from legacy carriers in a timely manner remains the critical risk mitigation for catastrophic loss.

## 2023

*The digital archivist logs into the workbench and calls up the emulation environment. Today, she is working on a group of files that had been removed from twenty-five 5.25" floppy disks almost twenty years before—records from the Royal Commission into British nuclear tests in Australia in the 1950s and 1960s, an important group of records about events that had long-term consequences for the Indigenous people who had lived there. The series information recorded in the archival control system gives no information about the operating system and application software or the hardware on which they ran, a serious gap in the information gathered about the disks when they were transferred into custody in 1985. Fortunately, useful technical metadata resulting from the data recovery project provides some guidance, the rest is her job to figure out and determine how to make these records accessible to the public. She carries out a fixity check to ensure the data files are the same as at the time of their recovery from the disks, and commences work. . . .*

The Legacy Media Project is a case study with many lessons learned. It informed policy decisions, for example the transfer policy, which listed the types of carriers the NAA would accept. The project also gathered important metrics on data recovery, including costs and rates of recovery by carrier type (see Appendixes B, C, and D). The success of the project in recovering data of national significance confirmed that the belts and braces approach adopted was warranted; although recovery processes and tools are very different today, they are still based on the need to ensure the authenticity, trustworthiness, and integrity of data. Similarly, at the time of the project, metadata standards for the preservation of digital records were in their infancy and not widely used, consequently the project developed what was in effect a default standard for managing the recovered data that is still useful today (i.e., the

working assumption is that it is better to have some metadata, even though it does not completely conform to modern standards, than to have none). Additionally, the project highlighted the need for an archival approach to data recovery, which led to the creation of or influenced a number of software tools and knowledge bases that are still relevant in 2022. Therefore, the discussion on the antiquity of digital process “history” is important to understand the development of digital forensics and preservation in the field of archival and library science, which is rightly considerably different today, as well as to provide a benchmark for future research.

Although the prospective fifth and final phase of the project—to provide the means to meaningfully render the recovered bits—was not commenced at the time, the fact that the bits were recovered and that metadata was extracted, derived, or provided by an operator and has been managed and stored according to early digital preservation principles means that the files can still be rendered. The risk-based approach to data recovery involving extracting multiple copies of the data and recording detailed information about the process, while resource intensive, may prove critical; for example, the disk images obtained in phase 1 could be critical for accessing data via emulation strategies. The key message is that recovering and preserving the bits while the opportunity for recovery exists is essential for future access.

If a recovery project was carried out today on the same carriers, the results would doubtless be different, even assuming the working instances of the access technology were available and the data were recoverable. The bits would still be the same bits (assuming that they had not degraded), but most of the metadata extracted would be the result of more automated processes and no longer the results of an artisan activity, but part of a more industrial process as described by Peter McKinney.<sup>56</sup>

In her book, *Romances of the Archive*, the American academic Suzanne Keen explores the many representations of archival research in twentieth-century fiction, from the ghost stories of H. P. Lovecraft and M. R. James, in which the labors of unwitting antiquarians unlock bogeys from the distant past, to the detective stories of Colin Dexter and P. D. James, in which insoluble crimes are solved in police archives by intrepid detectives, to the literary revelations of A. S. Byatt’s *Possession*, in which academic research in manuscript archives unlocks startling literary secrets.<sup>57</sup> The tale told in this article has also uncovered various chimeras, bogeys, and revelations in the data unlocked from the obsolete carriers of the past—and looks forward to new revelations in the future. For that reason, it is a tale worth telling from a different but not so distant past. This is one of the tales from THE disK FILES.

# Appendix A

Table 1: Series Title and No. of Carriers Involved in the Data Recovery Project

	Carrier type	Format type	Series	Date range	Total carriers
1	3 1/2" Floppy Disks	Macintosh	D4094	1989–1990	60
	Subtotal				60
2	Burroughs B20 5 1/4" Floppy Discs	Burroughs B20	A6459	1984–1985	25
		Burroughs B20	C1153	1984–1984	7
	Subtotal				32
3	Wang 8" Floppy Discs	Wang DOS	A6050	1982–1983	8
		Wang DOS	A6051	1982–1983	6
		Wang DOS	A6052	1982–1983	4
		Wang DOS	A6053	1982–1983	4
		Wang OIS	A6264	1982–1983	28
		Wang OIS	B3490	1981–1982	21
	Subtotal				71
4	9-Track 1/2" Magnetic Tape	9-Track	A10804	1970–1970	2
		9-Track	AA1979/319/0	1975–1975	33
		9-Track	B3971	1981–1983	3
		9-Track	C379p1	1981–1985	63
		9-Track	C379p2	1981–1985	25
		9-Track	C1281	1983–1983	2
		9-Track	C2920/1	1984–1989	4
		9-Track	C2920/3	1984–1989	5
	Subtotal				137
	Total		17		300

continued on next page

continued from previous page

	Series description
1	Miscellaneous Underlying Issues of the Aboriginal Issue Unit
2	Floppy discs of RC (Royal Commission) into British Nuclear tests in Austra during the 1950s and 1960s
	Floppy discs of final report of the National Road Freight Industry Inquiry
3	Data files from Independent Committee of Inquiry into Commonwealth Government Laboratory Services and Facilities
	As above
	As above
	As above
	Report of the Independent Committee of Inquiry into Commonwealth Government Laboratory Services and Facilities
	Word Processor Disks of the Royal Commission into Australian Meat Industry
4	Aboriginal Population Register
	'Scratch tapes' of raw data from two questionnaires relating to the Career Service Survey—RC [Royal Commission] into the Australian Public Service
	Computer Tapes relating to the Builders' Labourers' Federation deregistration case
	Magnetic tapes of program listings and source documents—Agent Orange Scientific Investigation
	As above
	Landsat satellite computer tapes—RC [Royal Commission] into User and Effects of Chemical Agents on Australian personnel in Vietnam
	Magnetic tape Binet activity details—Australian Bicentennial Authority
	As above



### 3 1/2" Floppy Disk



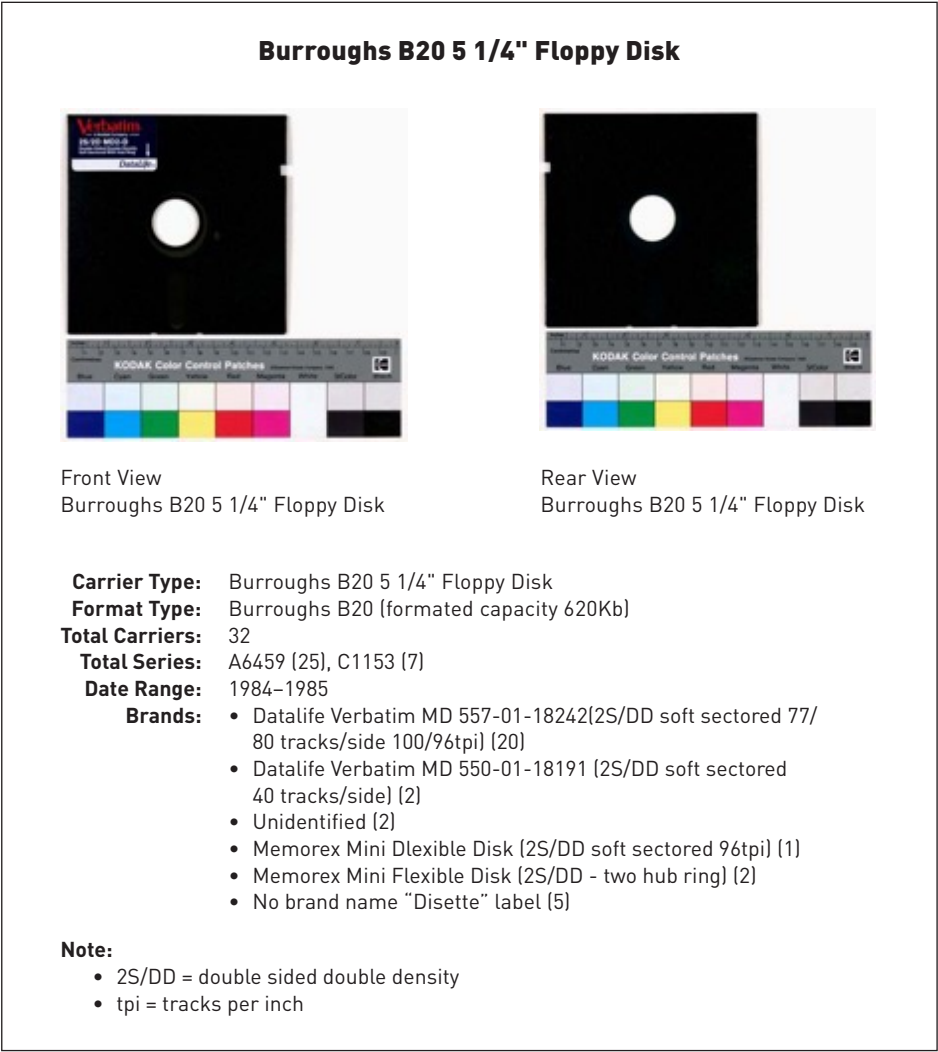
Front View  
3 1/2" Floppy Disk



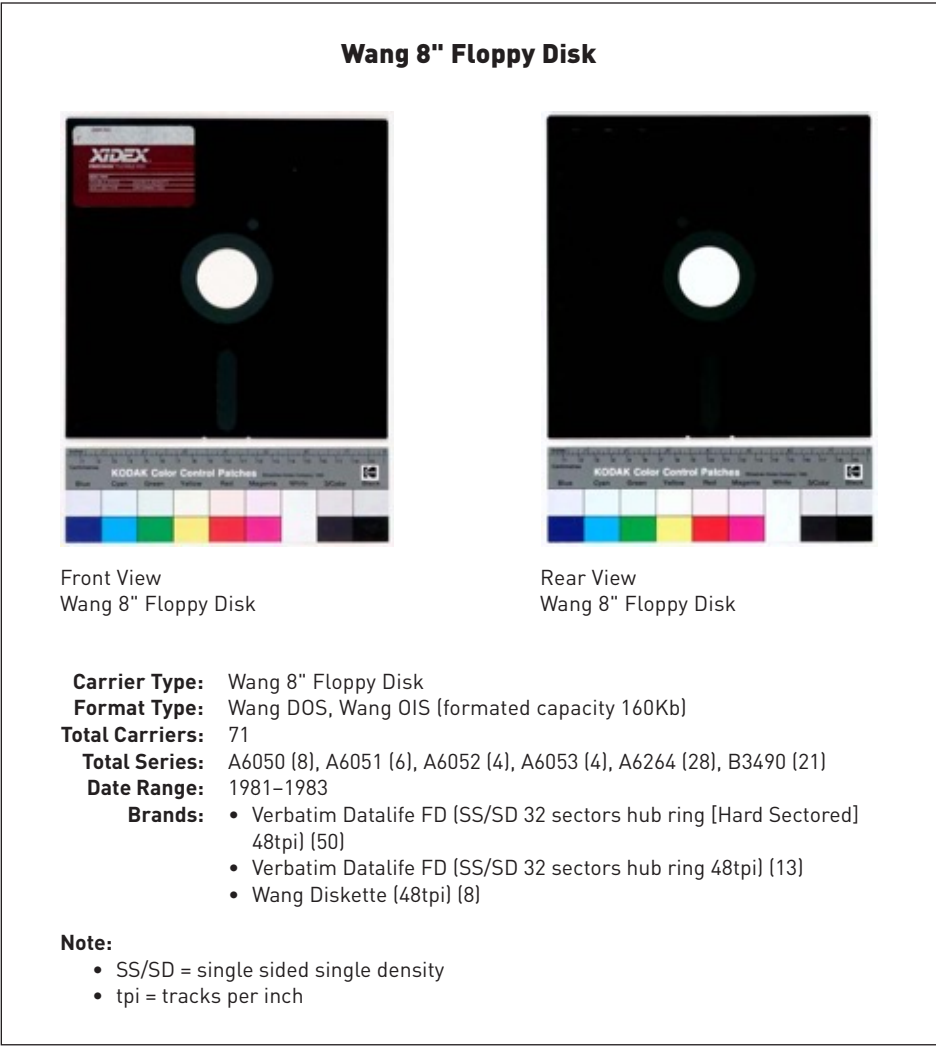
Rear View  
3 1/2" Floppy Disk

- Carrier Type:** 3 1/2" Floppy Disk  
**Format Type:** Macintosh (formatted capacity 800Kb)  
**Total Carriers:** 60  
**Total Series:** D4094 (60)  
**Date Range:** 1989–1990  
**Brands:**
- Nashua (2S/DD 135 tpi) (54)
  - SKC Micro FD (2S/DD 135tpi) (1)
  - Computer Land Solutions Microfloppy Disk (2S/DD 135tpi) (1)
  - BASF Flexy Disk (2S/DD 135tpi) (1)
  - Verbatim Data Life Microdisk MF 360 (2S/DD 135tpi) (1)
  - Sony Microfloppy disk (2S/DD) (1)
  - No Brand (1)
- Note:**
- 2S/DD = double sided double density
  - tpi = tracks per inch

**FIGURE 1.** Details of 3 1/2" floppy disk types in the data recovery project (including brands, format density, etc.)



**FIGURE 2.** Details of Burroughs B20 5 1/4" floppy disk types used in the data recovery project (including brands, format density, etc.)



**FIGURE 3.** Details of Wang 8" floppy disk types used in the data recovery project (including brands, format density, etc.)



**FIGURE 4.** Details of 1/2" magnetic tape (7- and 9-Track) types used in the data recovery project (including brands, format density, etc.)

Appendix B

Table 1: Detailed Results after Phase 1 (Based on Total Carrier Data Recovery, Partial Reads, and Fails)

Phase 1						
	Carrier type	Format type	Series	Date range	Total carriers	Total carriers examined in phase 1
1	3 1/2" Floppy Disks	Macintosh	D4094	1989–1990	60	60
	Subtotal				60	60
2	Burroughs B20 5 1/4" Floppy Discs	Burroughs B20	A6459	1984–1985	25	25
		Burroughs B20	C1153	1984–1984	7	7
	Subtotal				32	32
3	Wang 8" Floppy Discs	Wang DOS	A6050	1982–1983	8	8
		Wang DOS	A6051	1982–1983	6	6
		Wang DOS	A6052	1982–1983	4	4
		Wang DOS	A6053	1982–1983	4	4
		Wang OIS	A6264	1982–1983	28	28
		Wang OIS	B3490	1981–1982	21	21
	Subtotal				71	71
4	9-Track 1/2" Magnetic Tape	9-Track	A10804	1970–1970	2	1
		9-Track	AA1979/319/0	1975–1975	33	33
		9-Track	B3971	1981–1983	3	3
		9-Track	C379p1	1981–1985	63	63
		9-Track	C379p2	1981–1985	25	25
		9-Track	C1281	1983–1983	2	2
		9-Track	C2920/1	1984–1989	4	4
		9-Track	C2920/3	1984–1989	5	5
	Subtotal				137	136
	Total		17		300	299

\* = note: parts of the carriers in these instances could be read before failing

^ = note: some of these are blank carriers

continued on next page

continued from previous page

Phase 1						
	No. of potentially usable	No. of unusable carriers phase 1	Total carriers with 100% reads (after phase 1)	Total carriers with 100% reads but system or duplicate data (after phase 1)	Total carriers with partial reads (after phase 1)	Total carriers with 100% fail (after phase 1)
1	46, 11 Errors	3 Fail^	46 (76.6%)	0	11 (18.3%)	3 (5%)
	57	3	46 (76.6%)	0	11 (18.3%)	3 (5%)
2	25	0	25 (100%)	0	0	0
	7	0	7 (100%)	0	0	0
	32	0	32 (100%)	0	0	0
3	8	0	8 (100%)	0	0	0
	6	0	6 (100%)	0	0	0
	4	0	4 (100%)	0	0	0
	4	0	4 (100%)	0	0	0
	28	0	14 (50%)	14 (50%)	0	0
	21	0	21 (100%)	0	0	0
	71	0	57 (80.3%)	14 (19.7%)	0	0
4	1	0	1 (100%)	0	0	0
	26, 1 Errors	6 Fail*	26 (78.8%)	0	1 (3%)	6 (18.1%)
	3	0	3 (100%)	0	0	0
	59, 1 Errors	3 Fails*	59 (93.6%)	0	1 (1.6%)	3 (4.7%)
	22	3 Fails	22 (88%)	0	0	3 (12%)
	2	0	2 (100%)	0	0	0
	4	0	4 (100%)	0	0	0
	5	0	5 (100%)	0	0	0
	124	12	122 (89.7%)	0	2 (1.5%)	12 (8.8%)
	284	15	257 (86%)	14 (4.7%)	13 (4.3%)	15 (5%)

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-06-28 via Open Access.

## Appendix C

**Table 1: Detailed Results after Phase 2 (Including Carrier Blanks, Failure, Numbers of Files, Size of Contents, etc.)**

Phase 2						
	Carrier type	Format type	Series	Date Range	Total carriers	Total carriers examined in phase 2
1	3 1/2" Floppy Disks	Macintosh	D4094	1989–1990	60	60
	Subtotal				60	60
2	Burroughs B20 5 1/4" Floppy Discs	Burroughs B20	A6459	1984–1985	25	25
		Burroughs B20	C1153	1984–1984	7	7
	Subtotal				32	
3	Wang 8" Floppy Discs	Wang DOS	A6050	1982–1983	8	8
		Wang DOS	A6051	1982–1983	6	6
		Wang DOS	A6052	1982–1983	4	4
		Wang DOS	A6053	1982–1983	4	4
		Wang OIS	A6264	1982–1983	28	28
		Wang OIS	B3490	1981–1982	21	21
	Subtotal				71	71
4	9-Track 1/2" Magnetic Tape	9-Track	A10804	1970–1970	2	1
		9-Track	AA1979/319/0	1975–1975	33	33
		9-Track	B3971	1981–1983	3	3
		9-Track	C379p1	1981–1985	63	63
		9-Track	C379p2	1981–1985	25	25
		9-Track	C1281	1983–1983	2	2
		9-Track	C2920/1	1984–1989	4	4
		9-Track	C2920/3	1984–1989	5	5
	Subtotal				137	136
	Total		17		300	299

\* = note: parts of the carriers in these instances could be read before failing

^ = note: some of these are blank carriers

continued on next page



continued from previous page

Phase 2						
	No. of potentially usable carriers phase 2	No. of unusable carriers phase 2	Total no. of data objects	No. of potentially useable data objects	Total size (Bytes)	Total size (MB)
1	42	12 blank, 6 system	592	545	22421631.00	21.896
	42	18	592	545	22421631.00	21.896
2	24	1 blank	25	25	17463060.00	17.053
	7	0	75	75	1496064.00	1.461
	31	1	100	100	18959124.00	18.514
3	8	0	41	17	1886720.00	1.842
	6	0	20	5	1256448.00	1.227
	4	0	12	4	948736.00	0.926
	4	0	12	4	715776.00	0.699
	14	14 copies	68	34	3993513.00	3.899
	20	1 blank	91	91	1632139.00	1.593
	56	15	244	155	10433332.00	10.186
4	1	0	9	8	37810.00	0.369
	30	3 Fail	96	93	49005082.00	47.856
	3	0	9	3	26725760.00	26.099
	63	0	691	216	1907679880.00	1862.968
	22	3 Fail	97	35	2953093805.00	2883.88
	2	0	6	2	69159240.00	67.538
	4	0	129	42?	117431480.00	114.679
	5	0	39	12	110367884.00	107.781
	130	6	1076	369	5233500941.00	5111.17
	259	40	2948	1969	5285315028.00	5161.766

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-06-28 via Open Access.

**Table 2: Detailed Results after Phase 2 (Including Probable Formatting, Duplicate Data, and Total Recovery)**

Phase 2						
	Carrier type	Format type	Series	Date range	Total carriers	Vendor notes
1	3 1/2" Floppy Disks	Macintosh	D4094	1989–1990	60	Word Processing
	Subtotal				60	
2	Burroughs B20 5 1/4" Floppy Discs	Burroughs B20	A6459	1984–1985	25	Unknown Database
		Burroughs B20	C1153	1984–1984	7	Word Processing
	Subtotal				32	
3	Wang 8" Floppy Discs	Wang DOS	A6050	1982–1983	8	Unknown Database
		Wang DOS	A6051	1982–1983	6	Unknown Database
		Wang DOS	A6052	1982–1983	4	Unknown Database
		Wang DOS	A6053	1982–1983	4	Unknown Database
		Wang OIS	A6264	1982–1983	28	Word Processing (14 Disk & 14 Duplicates)
		Wang DOS	B3490	1981–1982	21	Word Processing
	Subtotal				71	
4	9-Track 1/2" Magnetic Tape	9-Track	A10804	1970–1970	2	Failed NAA Conversion (1996) of 7 track tape
		9-Track	AA1979/319/0	1975–1975	33	EBCDIC File
		9-Track	B3971	1981–1983	3	EBCDIC Database
		9-Track	C379p1	1981–1985	63	ASCII File
		9-Track	C379p2	1981–1985	25	EBCDIC & ACSII
		9-Track	C1281	1983–1983	2	EBCDIC Data File
		9-Track	C2920/1	1984–1989	4	Unknown
		9-Track	C2920/3	1984–1989	5	EBCDIC Printer File
	Subtotal				137	
	Total		17		300	

continued on next page

continued from previous page

Phase 2					
	Total carriers with 100% reads (after phase 2)	Total carriers with 100% reads but system or duplicate data (after phase 2)	Total carriers with partial reads (after phase 2)	Total carriers which were found to be blank (after phase 2)	Total carriers with 100% fail (after phase 2)
1	37 (61.7%)	6 (10%)	5 (8.3%)	12 (20%)	0
	37 (61.7%)	6 (10%)	5 (8.3%)	12 (20%)	0
2	24 (96%)	0	0	1 (4%)	0
	7 (100%)	0	0	0	0
	31 (96.9%)	0	0	1 (3.1%)	0
3	8 (100%)	0	0	0	0
	6 (100%)	0	0	0	0
	4 (100%)	0	0	0	0
	4 (100%)	0	0	0	0
	14 (50%)	14 (50%)	0	0	0
	20 (95.2%)		0	1 (4.8%)	0
	56 (78.9%)	14 (19.7%)	0	1 (1.4%)	0
4	0	0	1 (100%)	0	0
	26 (78.8%)	0	4 (12.1%)	0	3 (9.1%)
	3 (100%)	0	0	0	0
	59 (93.7%)	0	4 (6.3%)	0	0
	22 (88%)	0	0	0	3 (12%)
	2 (100%)	0	0	0	0
	4 (100%)	0	0	0	0
	5 (100%)	0	0	0	0
	121 (89%)	0	9 (6.6%)	0	6 (4.4%)
	245 (81.9%)	20 (6.7%)	14 (4.7%)	14 (4.7%)	6 (2%)

**Table 3: Carrier Log Showing Errors and Failures on Problematic Carriers**

Phase 2					
	Carrier type	Series	FD no.	Carrier label	Image/verify (phase 1)
1	3 1/2" Floppy Disks	D4094	409406	Sutty's field trips	Errors/Fail
			409408	Marcia	Errors/Ok
			409409	RID V2.0 Disk 1	Errors/Ok
			409410	[1/10]	Errors/Ok
			409411	[1/11]	Fail
			409412	RID V2.0 Segment 3	Fail
			409420	Police	Errors/Ok
			409421	Chip's Disk No. 1	Errors/Ok
			409428	[Some files faulty] Marcia on consultations	Errors/Ok
			409444	[4/6]	Errors/Ok
			409445	N/a	N/a
			409446	N/a	N/a
			409447	[4/9]	Errors/Ok
			409448	N/a	N/a
			409449	Copy consultations	Errors/Ok
			409455	CSO Laynhapy Resc, Police Report others	Errors/Ok
			409460	RID 8/6/90 CEG 1	Errors/Ok
	Subtotal				
3	1/2" Magnetic Tape	AA1979/319/0	197902	RCAGA15	Fail
			197903	RCAGA 14	Fail
			197904	RCAGA 13	Fail

continued on next page

continued from previous page

Phase 2						
	Vendor description (phase 1)	Total no. (phase 1)	Image/ verify (phase 2)	Vendor description (phase 2)	Total no. (phase 2)	NAA note
1	Disk has 2 data errors. Able to read file system		Ok	Nil		Data Objects retrieved— can't determine errors
	Disk has 5 data errors. Corrupted file system		Ok	Blank Disk		Blank
	Disk has 11 data errors. Corrupted file system		Ok	Blank Disk		Blank
	Disk has 24 data errors. Corrupted file system		Ok	Blank Disk		Blank
	Unable to image disk		Ok	Blank Disk		Blank
	Unable to image disk		Ok	Blank Disk		Blank
	Disk has 2 data errors. Able to read file system		Ok	Nil		Data Objects retrieved— can't determine errors
	Data		Ok	Nil		Data Objects retrieved— can't determine errors
	Data		Ok	Nil		Data Objects retrieved— can't determine errors
	Disk has data errors. Able to read file system		Ok	Blank Disk		Blank
	N/a		Ok	Blank Disk		Blank
	N/a		Ok	Blank Disk		Blank
	Disk has 5 data errors. Able to read file system		Ok	Blank Disk		Blank
	N/a		Ok	Blank Disk		Blank
3	Disk has 10 data errors. Able to read file system		Ok	Blank Disk		Blank
	Disk has 3 data errors. Able to read file system		Ok	Nil		Data Objects retrieved— can't determine errors
	Disk has 2 data errors. Able to read file system		Ok	Blank Disk		Blank
		14			12	
	Fail—unable to read tape		Ok	Nil		Data Objects retrieved— can't determine errors
	Fail—unable to read tape. Continued attempts risk damaging the carrier and tape drive		Ok	Nil		Data Objects retrieved— can't determine errors
	Fail—unreadable to read tape. Continued attempts risk damaging the carrier and tape drive		Ok	Carrier Failure		Fail

continued on next page

continued from previous page

Phase 2					
	Carrier type	Series	FD no.	NAA no.	Image/verify (phase 1)
			197905	RCAGA 12	Fail
			197906	RCAGA 11	Fail
			197915	RCAGA 23	Fail
			197921	RCAGA 30	Errors/Ok
		C379p1	379p124	A050 [BASF]	Errors/Ok
			379p139	A164 [DEC]	Fail
			379p151	361 [Gtape]	Fail
			379p163	CB 3112?	Fail
		C379p2	379p205	A041 [Q]	Fail
			379p207	A036 [Q]	Fail
			379p210	A031 [CD]	Fail
	Total				

continued on next page

continued from previous page

Phase 2						
	Vendor description (phase 1)	Total no. (phase 1)	Image/ verify (phase 2)	Vendor description (phase 2)	Total no. (phase 2)	NAA note
	Fail—unable to read tape. Continued attempts risk damaging the carrier and tape drive		Ok	Nil		Data Objects retrieved—can't determine errors
	Fails—can't read at all—NAA Label and documentation indicate the tape is RCAGA 11. We cannot find any evidence to support this		Fail	Carrier Failure		Fail
	Tape is blank		Ok	Carrier Failure		Fail
	Errors read ok		Ok	Nil		Data Objects retrieved—can't determine errors
		7			3	
			Ok	Nil		Data Objects retrieved—can't determine errors
			Ok	Nil		Data Objects retrieved—can't determine errors
			Ok	Nil		Data Objects retrieved—can't determine errors
			Ok	Nil		Data Objects retrieved—can't determine errors
	Fail—unreadable to read tape. Continued attempts risk damaging the carrier and tape drive		Fail	Failed Tape		Fail
	Fail—unreadable to read tape. Continued attempts risk damaging the carrier and tape drive		Fail	Failed Tape		Fail
	Fail—unreadable to read tape. Continued attempts risk damaging the carrier and tape drive		Fail	Failed Tape		Fail
		3			3	
		27			18	

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-06-28 via Open Access.



Appendix D

Table 1: Detailed Results after Phase 3

	Carrier type	Format type	Series	Date range	Total carriers	Total usable carriers (after Phase 3)
1	3 1/2" Floppy Disks	Macintosh	D4094	1989–1990	60	45
						7 errors
						12 blank
						3 system
	Subtotal				60	45
2	Burroughs B20 5 1/4" Floppy Discs	Burroughs B20	A6459	1984–1985	25	24
						1 blank
		Burroughs B20	C1153	1984–1984	7	7
	Subtotal				32	31
3	Wang 8" Floppy Discs	Wang DOS	A6050	1982–1983	8	8
		Wang DOS	A6051	1982–1983	6	6
		Wang DOS	A6052	1982–1983	4	4
		Wang DOS	A6053	1982–1983	4	4
					22	22
		Wang OIS	A6264	1982–1983	28	14
						14 copies
		Wang DOS	B3490	1981–1982	21	20
						1 blank
	Subtotal				49	34
4	9-Track 1/2" Magnetic Tape	9-Track	A10804	1970–1970	1	0
		9-Track	AA1979/ 319/0	1975–1975	33	26
						7 errors/fail
		9-Track	B3971	1981–1983	3	3
		9-Track	C379p1	1981–1985	63	58
						5 errors/fail
		9-Track	C379p2	1981–1985	25	22
						3 fail
		9-Track	C1281	1983–1983	2	2
		9-Track	2920/1	1984–1989	4	4
		9-Track	2920/3	1984–1989	5	5
	Subtotal				136	
	Total		17		300	

continued on next page

continued from previous page

	Total data objects (after phase 3)	Total usable? data objects (after phase 3)	Headers/ footers (after phase 3)	Empty, 0 length, file mark, previous data files, system files (after phase 3)	File type (if known)
1	594	547	0	47	Mac Word Processing
	594	547	0	47	
2	25	25	0	0	Backup spanned across a number of disks
	75	75	0	0	
	100	100	0	0	
3	41	41	0	0	Unknown/Database
	24	24	0	0	Unknown/Database
	16	16	0	0	Unknown/Database
	16	16	0	0	Unknown/Database
	56	56	0	0	
	36	36	0	0	Burroughs Word Processing Data
	91	91	0	0	Word Processing
	127	127	0	0	
4	8	0	0	0	Failed NAA conversion from 7 track Mag Tape
	96	91	0	5	Fixed length EBCDIC file
	9	3	3/3	0	EBCDIC Data files with header & footers
	687	220	173/187	107	
	98	39	21/26	12	EBCDIC & ASCII
	6	0?	0	0	Unknown EBCDIC Data File/ not a flat fixed length file/ possibly some form of propriety backup [Probably can't read—remote sensing Landsat data]
	127	43	42/42	2	ASCII Printer file header/footer
	39	12	12/12	3	Fixed length EBCDIC Printer file/ may contain field compression
	1070	408	521	129	

## NOTES

Thank you to Cal Lee and Andrew Long for their assistance and encouragement in writing this article. The authors also appreciate the feedback of the reviewers which undoubtedly made this a better article.

- <sup>1</sup> While the project used the term "media" to describe the objects of the data recovery project, the authors believe that the term is too ambiguous as it has multiple meanings depending on the context. To avoid confusion, this article will use the term "carriers" throughout.
- <sup>2</sup> Library of Congress, Premis, "The PREMIS Data Dictionary for Preservation Metadata," <https://www.loc.gov/standards/premis/index.html>.
- <sup>3</sup> Open Archival Information System (OAIS) Reference Model (ISO 14721), <http://www.oais.info>.
- <sup>4</sup> Steve Stuckey, "The Good Oil for Australia: Petroleum Data," in *Keeping Data: Papers from a Workshop on Appraising Computer-Based Records*. Australia Council of Archives and Australian Society of Archivists, ed. B. Reid and D. Roberts (Canberra: Australian Society of Archivists, 1991), 95–104. The risks associated with managing digital records in an offline storage environment led the NAA to adopt a distributed custody model between 1996 and 2000. See Don Boadle, "Reinventing the Archive in a Virtual Environment: Australians and the Non-Custodial Management of Electronic Records," *Australian Academic & Research Libraries* 35 (2004): 242–52, <https://doi.org/10.1080/00048623.2004.10755274>.
- <sup>5</sup> Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272, no.1 (1995): 42–47.
- <sup>6</sup> John Garrett and Donald Waters, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (Washington, DC: Commission on Preservation and Access and the Research Libraries Group, 1996).
- <sup>7</sup> Sarah Slade, David Pearson, and Steve Knight, "An Introduction to Digital Preservation in 2019," in *Preventive Conservation: Collection Storage*, ed. L. Elkin and C. A. Norris (New York: Society for the Preservation of Natural History; American Institute for Conservation of Historic and Artistic Works; Smithsonian Institution; The George Washington University Museum Studies Program, 2019), 810.
- <sup>8</sup> Bit level sends and receives data one bit at a time rather than in packets.
- <sup>9</sup> A disk image (or disk image file) is an exact binary copy of the contents of an entire disk or drive. Disk image files contain *all* the data stored on the source drive, including not only its files and folders but also its boot sectors, file allocation tables, volume attributes, and any other system-specific data. A disk image is not merely a collection of files or folders but is an exact duplicate of all the raw data of the original disk, sector by sector. Because disk images contain the raw disk data, it is possible to create an image of a disk written in an unknown format or even under an unknown operating system ([www.undisker.com/disk-images.html](http://www.undisker.com/disk-images.html)).
- <sup>10</sup> David Pearson, "Options Paper to Determine How to Proceed with the Legacy Media Paper" (National Archives of Australia, A14195, Corporate Records From the Electronic Recordkeeping System of the National Archives of Australia, 1998–; 2006/385). Much of this article is based on this internal report finalized in 2006 during phase 4 at the end of project.
- <sup>11</sup> David Pearson, "Preserve or Preserve Not, There Is No Try: Some Dilemmas Relating to Personal Digital Archiving" (presented at Digital Curation Practice, Promise and Prospects, University of North Carolina at Chapel Hill, North Carolina, April, 2009), Preserve or Preserve Not ([slideshare.net](http://slideshare.net)); Ross Harvey and Martha R. Mahard, *The Preservation Management Handbook: A 21st-Century Guide for Libraries, Archives, and Museums* (Lanham: Rowman & Littlefield, 2014), 17; Ross Harvey and Martha R. Mahard, *The Preservation Management Handbook: A 21st-Century Guide For Libraries, Archives, and Museums* 2nd ed., rev. D. Conn (Lanham: Rowman & Littlefield, 2020), 20.
- <sup>12</sup> Simon Davis, Helen Heslop, and Andrew Wilson, "An Approach to the Preservation of Digital Records" (National Archives of Australia, 2002), [https://www.ltu.se/cms\\_fs/1.83844!/file/An\\_approach\\_Preservation\\_dig\\_records.pdf](https://www.ltu.se/cms_fs/1.83844!/file/An_approach_Preservation_dig_records.pdf).
- <sup>13</sup> Cornell Platzer and David Pearson, *Digital Preservation: Illuminating the Past, Guiding the Future* (Canberra: National Archives of Australia, 2006), [https://web.archive.org/web/20070829144700/http://naa.gov.au/recordkeeping/preservation/digital/XENA\\_brochure.pdf](https://web.archive.org/web/20070829144700/http://naa.gov.au/recordkeeping/preservation/digital/XENA_brochure.pdf).
- <sup>14</sup> James Doig, "Digital Preservation at the National Archives of Australia: Achievements and Directions," in *iRMA: Information and Records Management Annual: Official Journal of the RMAA* (Queensland: Records Management Association of Australia, 2008), 167–68.

- <sup>15</sup> NAA Commonwealth Agency (CA) 8116: Australian Wool Research and Promotion Organisation; Series A13211: Electronic Records Media.
- <sup>16</sup> Doig, "Digital Preservation at the National Archives of Australia," 168. The Agency to Researcher Project outcomes did not result in many academic papers, and this article is an attempt to disseminate the results of this subproject beyond the project "Options Paper" (see endnote 10).
- <sup>17</sup> Scott Prasser, *Royal Commissions and Public Inquiries in Australia* (Chatswood, New South Wales: Butterworth, 2006).
- <sup>18</sup> Mary Feeney, *Digital Culture: Maximising the Nation's Investment: A Synthesis of JISC/NPO Studies on the Preservation of Electronic Materials* (London: British Library Board, 1999), 64.
- <sup>19</sup> Deborah Woodyard, "Farewell My Floppy: A Strategy for Migration of Digital Information," in *Proceedings of the 9th Vala Conference* (Melbourne, 1998), 331–40, <https://www.vala.org.au/vala1998-proceedings>.
- <sup>20</sup> Seamus Ross and Ann Gow, *Digital Archaeology: Rescuing Neglected and Damaged Data Resources* (London: British Library, 1999), <http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf>.
- <sup>21</sup> For example, Douglas Elford, Nicholas Del Pozo, Snezana Mihajlovic, David Pearson, Gerard Clifton, and Colin Webb, "Media Matters: Developing Processes for Preserving Digital Objects on Physical Carriers at the National Library of Australia," in *74th IFLA World Library and Information Congress* (Quebec City, Canada, August 10–14, 2008), <http://www.ifla.org/IV/ifla74/papers/084-Webb-en.pdf>; Jeremy L. John, "Adapting Existing Technologies for Digitally Archiving Personal Lives. Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools," in *iPRES 2008: Proceedings of the 5th International Conference on Preservation of Digital Objects* (London: The British Library, 2008), 48–55, <https://bl.iro.bl.uk/work/ns/1e331593-1eb1-45f2-a7ab-5509caf47b40>; Kam Woods and Geoffrey Brown, "From Imaging to Access—Effective Preservation of Legacy Removable Media," in *Archiving 2009: Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organisations: Final Program and Proceedings* (Springfield, VA: Society of Imaging Science and Technology, 2009), 213–18, <https://library.imaging.org/archiving/articles/6/1/art00047>; Kam Woods, Christopher A. Lee, and Simson Garfinkel, "Extending Digital Repository Architectures to Support Disk Image Preservation and Access," in *JCDL 11: Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* (New York, 2011), 57–66, <https://doi.org/10.1145/1998076.1998088>; Maureen Pennock, Michael Day, Peter May, Kevin Davies, Simon Whibley, Akiko Kimura, and Edith Halvarsson, "The Flashback Project: Rescuing Disk-Based Content from the 1980's to the Current Day," in *Proceedings of the 11th Digital Curation Conference*, (Amsterdam: IPRES, February 2016), 1–11, <https://doi.org/10.5281/zenodo.1321629>; Johan Van der Knijff, "Recovering '90s Data Tapes: Experiences From the KB Web Archaeology Project," in *iPRES2019: Proceedings of the 16th International Conference on Digital Preservation* (Amsterdam: IPRES, 2019), 25–36, <https://services.phaidra.univie.ac.at/api/object/o:1079683/diss/Content/get>; Monya Baker, "Disks Back from the Dead," *Nature* 545 (2017): 117–18, <http://dx.doi.org/10.1038/545117a>.
- <sup>22</sup> Luciana Duranti, "From Digital Diplomats to Digital Records Forensics," *Archivaria* 68 (2009), 39–66, <https://archivaria.ca/index.php/archivaria/article/view/13229>; Frederick B. Cohen, "Digital Diplomats and Forensics: Going Forward on a Global Basis," *Records Management Journal* 25 (2015), 21–44, <https://doi.org/10.1108/RMJ-03-2014-0016>; Michael Moss, David Thomas, and Tim Gollins, "The Reconfiguration of the Archive as Data to Be Mined," *Archivaria* 86 (2018): 118–51, <https://archivaria.ca/index.php/archivaria/article/view/13646>; Mark Wolverston, "Digital Forensics: From the Crime Lab to the Library," *Nature* 534 (2016): 139–40, <https://doi.org/10.1038/534139a>.
- <sup>23</sup> Prometheus Digital Preservation Workbench, <http://prometheus-digi.sourceforge.net> and <http://prometheus-digi.sourceforge.net/faq.html>. David Pearson, the coauthor of this article, was project manager of the project that developed Prometheus and was the manager of the Digital Preservation Section at the National Library of Australia between 2008 and 2015.
- <sup>24</sup> BitCurator, <https://bitcurator.net>. David Pearson, the coauthor of this article, was an inaugural member of the BitCurator Digital Advisory Group between 2010 and 2013.
- <sup>25</sup> Elford et al., "Media Matters"; Pearson, "Preserve or Preserve Not, There Is No Try"; Nicholas Del Pozo, Douglas Elford, and David Pearson, "Prometheus: Managing the Ingest of Media Carriers," in *Proceedings of DigCCurr 2009, Digital Curation Practice, Promise and Prospects* (University of North Carolina at Chapel Hill, 2009), 73–75, <https://www.slideshare.net/natlibaryofaustralia/prometheus-13399577>; David Pearson, "DigCCurr 2009, Digital Curation Tools and Demos II—

- Mediapeda: Managing the Identification of Media Carriers, and Prometheus: Managing the Ingest of Media Carriers at the NLA (presentation & demo)" (presented at DigCCurr 2009, Digital Curation Practice, Promise and Prospects, University of North Carolina at Chapel Hill, April, 2009), <https://www.slideshare.net/natlibraryofaustralia/dig-c-curr>; Christopher A. Lee, "Archival Application of Digital Forensics Methods for Authenticity, Description and Access Provision," *Commis*, no. 2 (January 2012): 133–40, <https://ils.unc.edu/calce/p133-lee.pdf>; Christopher A. Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassanoff, *From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions* (BitCurator Project, 2013), <https://bitcurator.net/files/2018/08/bitstreams-to-heritage.pdf> and <https://ils.unc.edu/news/2013/bitcurator-white-paper>.
- <sup>26</sup> Mediapeda, <https://www.nla.gov.au/mediapeda>.
- <sup>27</sup> Douglas Elford and David Pearson, "What Is the Mediapeda?" (presented at the Innovative Ideas Forum, National Library of Australia, April 10, 2008), <https://www.slideshare.net/natlibraryofaustralia/what-is-the-mediapeda>; Nicholas Del Pozo, Douglas Elford, and David Pearson, "Mediapeda: Managing the Identification of Media Carriers," in *Proceedings of DigCCurr 2009, Digital Curation Practice, Promise and Prospects* (University of North Carolina at Chapel Hill, 2009), 76–78, <https://www.slideshare.net/natlibraryofaustralia/mediapeda>; Pearson, "Preserve or Preserve Not, There Is No Try"; Mark Pearson and Gareth Kay, "National Library of Australia Software and File Formats Knowledge Base," in *iPRES2014, Proceedings of the 11th International Conference on Digital Preservation* (iPRES: State Library of Victoria, Melbourne, 2014), 383–84, <https://phaidra.univie.ac.at/o:378066>; Gareth Kay, Libor Coufal, and Mark Pearson, "Backing Up Digital Preservation Practice with Empirical Research: The National Library of Australia's Digital Preservation Knowledge Base," *Alexandria: The Journal of National and International Library and Information Issue*, 27, no. 2 (2017): 66–82, <https://doi.org/10.1177/0955749017724630>.
- <sup>28</sup> Sherry L. Xie, "Building Foundations for Digital Records Forensics: A Comparative Study of the Concept of Reproduction in Digital Records Management and Digital Forensics," *American Archivist* 74, no. 2 (2011): 576–99, <https://www.jstor.org/stable/23079051>; Kyle R. Rimkus, Bethany Anderson, Karl E. Germeck, Cameron C. Neilson, Christopher J. Prom, and Tracy Popp, "Preservation and Access for Born-Digital Electronic Records: The Case for an Institutional Digital Content Format Registry," *American Archivist* 82, no. 2 (2020): 397–428, <https://doi.org/10.17723/0360-9081-83.2.397>.
- <sup>29</sup> In 2013, the Society of American Archivists ran a series of projects called Jump In, which encouraged repositories to carry out similar audits, with the resulting inventories made publicly available, <https://www2.archivists.org/groups/manuscript-repositories-section/jump-in-initiative-0>.
- <sup>30</sup> Richard Waller, "Collection Risk Assessment," in *Preventive Conservation: Collection Storage*, 59–90. There are many other examples in the literature since 1994 outlining the ten agents of deterioration for physical collections encompassing direct physical forces, criminals, fire, water, pests, pollutants, light, incorrect temperature, incorrect relative humidity, and dissociation. There is a direct parallel with digital collections with the disassociation of information about the carriers, the files contained on them, and information required to access these assets. Also see Slade et al., "An Introduction to Digital Preservation," 817–18.
- <sup>31</sup> See the PREMIS Data Dictionary for Preservation Metadata, Version 3.0, <https://www.loc.gov/standards/premis/v3>.
- <sup>32</sup> NAA AA1979/319/0: Scratch tapes of raw data from two questionnaires relating to the Career Service Survey.
- <sup>33</sup> NAA C379p1: Magnetic tape of program listings and source documents relating to the Agent Orange Scientific Investigation.
- <sup>34</sup> USGS, "Landsat," <https://www.usgs.gov/land-resources/nli/landsat>.
- <sup>35</sup> ASCII (American Standard Code for Information Interchange) is a character encoding standard for electronic communication. ASCII codes represent text in computers, telecommunications equipment, and other devices. Most modern character-encoding schemes are based on ASCII (<https://en.wikipedia.org/wiki/ASCII>).
- <sup>36</sup> EBCDIC (Extended Binary Coded Decimal Interchange Code) is an 8-bit character encoding used mainly on IBM mainframe and IBM midrange computer operating systems. It descended from the code used with punched cards and the corresponding 6-bit binary-coded decimal code used with most of IBM's computer peripherals of the late 1950s and early 1960s (<https://en.wikipedia.org/wiki/EBCDIC>).

- <sup>37</sup> InterMedia was a UK company that specialized in supplying media and data conversion systems for over 2,000 floppy disk and hardware and operating system combinations, see John, "Adapting Existing Technologies," 52.
- <sup>38</sup> NAA A14195, 2006/385.
- <sup>39</sup> Australian Government National Archives of Australia, "Digital Preservation," [https://web.archive.org/web/20070829202518/http://naa.gov.au/recordkeeping/preservation/digital/digital\\_repository.html](https://web.archive.org/web/20070829202518/http://naa.gov.au/recordkeeping/preservation/digital/digital_repository.html).
- <sup>40</sup> Australian Government National Archives of Australia, "Digital Recordkeeping: Guidelines for Creating, Managing and Preserving Digital Records," <https://web.archive.org/web/20070830083332/http://www.naa.gov.au/recordkeeping/er/guidelines.html>.
- <sup>41</sup> John, "Adapting Existing Technologies," 52.
- <sup>42</sup> NAA AA1979/319, "Scratch tapes" of raw data from two questionnaires relating to the Career Service Survey, and NAA C379 (parts 1 and 2): Magnetic tapes of program listings and source documents relating to the Agent Orange Scientific Investigation.
- <sup>43</sup> NAA CA 3641: 1983 Royal Commission into the use and effects of chemical agents on Australian personnel in Vietnam, Series C1281: Landsat satellite computer tapes together with negative and positive prints.
- <sup>44</sup> ESRI Australia, "Products," <https://esriaustralia.com.au/envi>.
- <sup>45</sup> BitCurator.
- <sup>46</sup> Kryoflux, <https://www.kryoflux.com>.
- <sup>47</sup> Isgandar Valizada, Klaus Rechert, and Dirk von Suchodoletz, "Emulation-as-Service—Requirements and Design of Scalable Emulation Services for Digital Preservation," in *Hochleistungsrechen in Baden-Württemberg—Ausgewählte Aktivitäten im bwGRiD 2012*, ed. J. Schulz and S. Herman (Karlsruhe: Scientific Publishing, 2014), 103–16; Euan Cochrane, Klaus Rechert, Seth Anderson, Jessica Meyerson, and Ethan Gates, "Towards a Universal Virtual Interactor (UVI) for Digital Objects," in *iPRES2019: Proceedings of the 16th International Conference on Digital Preservation*, 191–200, [https://ipres2019.org/static/pdf/iPres2019\\_paper\\_128.pdf](https://ipres2019.org/static/pdf/iPres2019_paper_128.pdf).
- <sup>48</sup> NAA CA 3144: Royal Commission into the Activities of the Federated Ship Painters and Dockers Union, Series A7759: Electronic data and paper appendices relating to the Costigan Royal Commission on the Activities of the Federated Ship Painters and Dockers Union.
- <sup>49</sup> Prasser, *Royal Commissions*, 287.
- <sup>50</sup> Slade et al., "An Introduction to Digital Preservation," 824.
- <sup>51</sup> National Archives of Australia, "Australian Commonwealth Record Series System," <https://www.naa.gov.au/help-your-research/getting-started/commonwealth-record-series-crs-system>.
- <sup>52</sup> See, for the example, the resources developed by the Software Preservation Network's Fostering Communities of Practice, <https://www.softwarepreservationnetwork.org/fcop/resources>.
- <sup>53</sup> National Archives of Australia, "Previous Policies—Information Management," <https://www.naa.gov.au/information-management/information-management-policies/digital-continuity-2020-policy/digital-transition-policy>.
- <sup>54</sup> National Archives of Australia, "Building Trust in the Public Record," <https://www.naa.gov.au/information-management/information-management-policies/building-trust-public-record-policy>.
- <sup>55</sup> Jay Gattuso and Peter McKinney, "Converting WordStar to HTML4," in *iPRES2014, Proceedings of the 11th International Conference on Digital Preservation*, 149–59, <https://natlib.govt.nz/files/digital-preservation/WordStar-ipres2014-4.pdf>; Oxford University Research Archive, "Digital Preservation at Oxford and Cambridge Training Programme Pilot," <https://libguides.bodleian.ox.ac.uk/digitalpreservation/migration>.
- <sup>56</sup> Peter McKinney, "From Hobbyist to Industrialist: Challenging the DP Community" (presented at iPRES2012: 9th International Conference on Preservation of Digital Objects, Open Research Challenges in Digital Preservation Workshop, Toronto, October 1–5, 2012), <https://natlib.govt.nz/files/digital-preservation/mckinney.pdf>. Also see Michael Moss et al., "The Reconfiguration of the Archive as Data to Be Mined," 120, 146.
- <sup>57</sup> Suzanne Keen, *Romances of the Archive in Contemporary British Fiction* (Toronto: University of Toronto Press, 2001).

## ABOUT THE AUTHORS

---



**David Pearson**, FSA, is an archaeologist and senior curator in the Military Heraldry and Technology Section at the Australian War Memorial Museum in Canberra. He has published extensively on archaeological, digital preservation, and military historical issues and his main research interests concern the technological and social contexts of conflict archaeology. He has published four books and is a fellow of the Society of Antiquaries of London. Pearson has twenty-eight years of experience working in national cultural institutions and worked in digital preservation at the National Archives of Australia, the National Library of Australia, and the Department of Parliamentary Services from 2002 to 2019.



**Dr. James Doig** has worked at the National Archives of Australia for more than twenty years. In that time, he has worked in many roles in collection management, including digital preservation, transfer, description, and collection review. He was involved in the design and construction of the National Archives preservation facility, the Peter Durack building, in Canberra and oversaw the relocation of 115 kilometers of records into the facility. He has a PhD in medieval history from Swansea University, and his latest book is *Beyond the Orbit: An Anthology of Australian Science Fiction to 1950*.