



REFLECTION

Post-Custodialism, Distributed Custody, and Big Data

James Doig*

Digital Preservation Manager, National Archives of Australia, Canberra, Australian Capital Territory

Abstract

This reflection piece describes the outcomes of a research project undertaken by the National Archives of Australia that aimed to gather information from other government archives and selected Australian government agencies about their approach to archiving and preserving large-big datasets in the government sector. Big data collections pose a challenge for government archives around the world. Many of these archives have a role in information management in their government domains and provide guidance and advice to their government agency clients on ensuring the integrity and trustworthiness of data over time. The article examines the nexus between theory and practice, exploring issues related to the post-custodial ideas developed by Terry Cook and others in the 1990s and their practical implementation.

Keywords: *Distributed custody; Post-custodialism; Digital preservation; Big data*

This article reflects on the results of a research project carried out by the National Archives of Australia (NAA) in 2022 on large and big data producing agencies (i.e. in the high terabyte and petabyte size) and the challenges posed by big data collections to archives. By telling the story, I want to highlight the important work accomplished in the 1990s and early 2000s regarding recordkeeping standards and guidance – work that involved close collaboration between theorists and practitioners who espoused post-custodial approaches to archival and records management. Much of this work, especially around distributed custody, appears to have been forgotten, and it is worth drawing attention to it again in the context of big data.

Theory and practice in the 1990s

The second half of the 1990s is surely the high-water mark in collaboration between theorists and practitioners in Australia. In particular, the records continuum model developed at Monash University by Frank Upward and others influenced the development of foundational standards, policy guidance, and advice on recordkeeping and information management. Theorists and practitioners participated in the development of the world-first records management standard, AS 4390. Released by Standards Australia in 1996, it was the starting point for the

*Correspondence: James Doig, Email: james.doig@naa.gov.au

International Standards Organization's 2001 standard on records management, ISO 15489.¹ National Archives developed the 'e-permanence' suite of advisory products that Commonwealth government agencies could use to build best practice recordkeeping environments.² The cornerstone of e-permanence, released in 2000, was the *Designing and Implementing Recordkeeping Systems (DIRKS) Manual*.³ A joint development of the State Records Authority (RA) of New South Wales and the National Archives, the DIRKS Manual built on AS 4390 to provide comprehensive practical guidance on designing and implementing a recordkeeping system via an eight-step methodology that was included in outline in AS 4390 and which heralded the brave new world of functions-based appraisal.⁴

e-permanence was itself heavily influenced by the post-custodial theory that had informed the continuum model and that was most eloquently championed by the Canadian archivist and theorist, Terry Cook. Cook gave an invitational lecture tour of Australia in 1993 (see Figure 1) and his seminal article published the following year, 'Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era', was based on a lecture he delivered several times during his tour.⁵

When the National Archives implemented a distributed custody policy for digital records in 1996, the policy intent was succinctly expressed as follows: 'the preferred arrangement is for agencies to retain custody of electronic records of ongoing value, but under a management regime worked out with the Australian Archives'.⁶ The standards, policy guidance, and advice developed often in close partnership with theorists constituted the management regime. Although the development of a digital preservation capability brought an end to the distributed custody policy after 4 years, the management frameworks for post-custodialism and distributed custody were in place by the turn of the millennium and are still with us today, though the tools have continued to evolve.

This period also saw the development of practical rules for the management of digital records subject to distributed custody arrangements. As early as 1993, the National Archives of Canada had policies and rules for distributed custody in place, including a clear articulation of the circumstances in which archival value digital records would be left with the creating government institution. These were listed by Terry Cook in a 1995 article:

1. Where the cost of transfer of the record or other technical considerations (software copyright, data complexity, software and hardware dependency, etc.) make it impossible for the Archives to acquire the record at this time and/or
2. Where the institution has a continuing and long-term operational need for the record, which includes the provision of elaborate and extensive reference services and/or
3. Where because of the nature of the record reference services can best be provided by the institution rather than by the Archives and/or
4. Where there are statutory provisions that prevent transfer to the Archives.⁷

Interestingly, the main categories of records identified as candidates for distributed custody quite closely reflect the current big data environment in government, including cumulative and longitudinal systems such as scientific, environmental, and social data. Some of the terms and conditions developed by the Canadians also remain relevant today. For example, the National Archives may exercise the right to transfer the records into custody if there are major systems changes or where systems are to be decommissioned. That said, the preservation requirements reflect a period when information tended to be stored on 9-track magnetic tape in off-line storage environments⁸, and where Cook was talking about reference services involving people, in particular telephone inquiry services, these days agencies like the Bureau of Meteorology



Figure 1. Terry Cook [centre, with Mark Stevens and Ann Pederson] visits Australia, State Library of New South Wales, Macquarie Street Sydney, 1993. Image courtesy of City of Sydney Archives: A-00028118.

(BOM), Geoscience, and the Australian Bureau of Statistics (ABS) provide sophisticated on-line access to their data.

Digital preservation theory and practice have developed enormously since the mid-1990s, but the now well-known digital preservation principles and approaches – multiple independent storage, integrity checking, migration strategies, and so on – impose a hefty cost for petabyte size data and may be a barrier to distributed custody agreements with agencies.

Big data project

The National Archives' research project, conducted in 2022, was a response to issues raised by government agencies that are creating and managing large-big datasets. This includes agencies managing massive datasets in single systems (for example, BOM) and agencies with data assets distributed across many medium-sized systems (for example, ABS). In particular, the following three key business problems were identified:

1. *Information management challenges:* applying information management standards and requirements, particularly disposal requirements, to large-big datasets can be a challenge for agencies.
2. *Transfer, preservation, and access:* under the *Archives Act 1983*, Australian government agencies must transfer records sentenced Retain as National Archives (RNA) either as soon as practicable after business use has ceased or at the latest 15 years after creation. However, the size and complexity of large-big datasets pose a challenge for transfer.
3. *Distributed custody:* section 64 of the *Archives Act* allows for permanent value records to remain in the custody of the controlling agency subject to certain conditions; however, not a single section 64 agreement has been developed for digital records.

Ultimately, the aim of the project was to inform NAA decision making and shape our guidance, especially around distributed custody arrangements for digital records.

Project method

The project was carried out by the Digital Archives Innovation and Research (DAIR) section of National Archives, in partnership with Governance Records Assurance, the section responsible for government recordkeeping. DAIR operates as something of a research hub for the National Archives and undertakes short-term projects of a few days (e.g. rapid evidence reviews) to longer-term projects typically up to 6 months.

The approach adopted for the Big Data Project was to interview national and international archives⁹ and selected Australian government agencies¹⁰ to gather information about their responses to these business problems. One-hour interviews were scheduled, and separate questions were developed for archival authorities and agencies. Each interview was recorded and detailed notes were written up, with overall results collated in spreadsheets.¹¹

The following sections provide a brief overview of the results with a focus on a few key themes:

Government agencies

Size, range, and nature of datasets

The agencies interviewed can be broadly divided between those creating and managing research and scientific datasets, such as BOM and Geoscience Australia, and those managing datasets containing the personal information of Australian citizens and which relate to rights and entitlements, such as the Department of Social Services, the Australian Taxation Office (ATO), and Services Australia. Some agencies straddle both categories, for example, ABS and the Department of Agriculture, Fisheries and Forestry.

The agencies can be further categorized by those managing extremely large datasets in the petabyte size such as BOM and those like ABS managing hundreds or sometimes thousands of individual datasets that together amount to a very large quantity of data.

Records authorities

All of the interviewed agencies had retention and disposal schedules/authorities, referred to as RAs, though they varied widely in currency. Some agencies had detailed DIRKS-era¹² RAs, while others had streamlined 'rolled-up' RAs characterized by a smaller number of 'bucket' classes. A common theme from the interviews was that RAs were difficult to interpret and apply to data and datasets, partly because current records and information managers were not involved in their development. Most said that their RAs required updating either because they were too old and used outdated terminology or because they had significant gaps in coverage. A few felt that their RAs adequately covered their historical datasets but did not cover some current datasets. A number of agencies were already working on updating their RAs; however, it was a slow process because of limited resources and the need for wide stakeholder engagement.

Public access

Generally, the agencies creating and managing research and scientific datasets are already providing public access to their data, for example, from their websites like BOM and the ABS, from third-party providers like some Geoscience datasets, or from public data archives. These research datasets tend to be heavily used by the public.

The agencies creating and managing the private data of Australian citizens do not provide public access to this information though they may provide access to summary data for social research purposes. They are required to release private data to individuals as a result of freedom of information requests.

Distributed custody arrangements

All of the interviewed agencies expressed the need for distributed custody arrangements because of the size and complexity of their datasets and ongoing business needs that required retention of custody. Some agencies felt that distributed custody arrangements were appropriate due to the complexities involved in appraisal and disposal (including transfer) of massive datasets containing sensitive personal information (although it was understood that sensitivity does not exempt an agency from transferring records to the National Archives).

However, though accepting the need for distributed custody, all agencies expressed concerns about requirements that may be imposed under such agreements, in particular:

- The potential for significant additional costs, for example, the requirement for multiple redundant storage options
- Physical inspections, which some agencies said would not be possible for security reasons, and which in any case are not appropriate for digital records
- Access requirements under the Archives Act, especially for personal sensitive information
- Difficulties determining a custodian for shared data, i.e. data that is shared between multiple agencies or shared with in joint venture arrangements
- Any legal requirements such as insurance would need to be approved by legal teams and senior management.

On the other hand, some of the agencies said that many of the requirements for long-term preservation and access are already in place as part of normal data management and data protection practice.

Government archival authorities

Interviews were conducted with other national archives and a number of local archival jurisdictions both national and international (e.g. the Landesarchiv Baden-Württemberg).

Regulatory environment

The level and degree of information management regulation and compliance can be broadly split between countries whose legal systems are based on common law or civil (codified) law. In civil law countries, which include much of mainland Europe, many of the archival authorities' issue regulations or orders that have a relatively high level of agency compliance. The archival authorities often issue or endorse detailed functional requirements for business systems managing records and issue regulations requiring compliance with technical standards, including technical requirements for transfer.¹³ Common law countries, such as United Kingdom, the United States, Canada, Australia, and New Zealand, may issue standards and guidance but tend not to enforce them, and agency compliance varies widely. In effect, in these countries, agencies self-regulate and there is a high degree of latitude in interpreting and applying guidance and standards.

In European countries, government agencies managing large-big datasets tend to be more aware of their records and information management responsibilities regarding data. This is partly because of stricter regulatory regimes, which mandate technical requirements for

systems and transfer but also because they have considerable experience over many years in database preservation and transfer. Most of the key database preservation research projects have been European, such as the development of the Software Independent Archiving of Relational Databases (SIARD) format at the Swiss Federal Archives and the work of the European Union-funded E-Ark Project. Nevertheless, the interviews did indicate that even in Europe it can be difficult to find fully compliant agency business systems. One European archival authority said that they treat transfers as a snapshot of data at a point in time. They cannot guarantee the accuracy of the data as they cannot be responsible for agency information management practices, for example, if archival value data was overwritten as part of a thinning process.

Retention and disposal schedules

Most of the archival authorities said that retention and disposal schedules were required by law and should cover all information, including databases and datasets. Functions-based disposal schedules are commonly used, but most archival authorities said that disposal schedules generally did not adequately cover data and datasets. Most archival authorities reported large disparities in coverage between disposal schedules, for example, some take a ‘big bucket’ approach, while others are more granular. Often a permanent value business system or dataset was a single line in a disposal schedule although it may contain temporary or nonarchival information.

There was also broad agreement among archival authorities that retention and disposal schedules invariably do not help to determine the archival ‘record’ to be transferred from a database or business system to the archival authority. The Estonian National Archives adopted a macro-appraisal approach to determine the value of databases across government. For each archival value system, they conduct a high-level appraisal of the data within the business system, e.g. system files and views can be discarded. Other European archives have well-established transfer regimes and, as mandated in regulation, determine what is to be transferred when the business system is being developed. A common problem identified in the interviews is that records and information officers tend to take a narrow view of the record and often do not consider data and datasets as records.

Distributed custody arrangements

Under distributed custody arrangements a body other than the archival authority retains custody of archival value records, while control and ultimate responsibility for the records rests with the archival authority. While distributed custody arrangements are common for analogue records,¹⁴ few if any have been established for digital records. The broad view across the interviewed archival authorities was that distributed custody arrangements for large-big datasets were desirable and that a practical and implementable management regime overseen by the archival authority was a necessary component of it.

Reflection

Two key findings of the project are that, for common law countries like Australia, (1) agencies are retaining custody of archival-value digital records that are eligible for transfer to the archival authority and (2) archives do not have distributed custody arrangements in place for those records.

The first point is well known. For archival authorities, the so-called digital deluge has been just around the corner for a couple of decades now, but so far the flood still hasn’t eventuated. At the NAA, the vast bulk of the digital records received from agencies are from temporary agencies such as Royal Commissions or Commissions of Inquiry, closed agencies, or records

for which there is no inheriting agency as a result of machinery of government change. If we expected regular transfers from standards-based systems like Electronic Document and Records Management Systems to become the norm, we were mistaken. In 2023, we should be receiving archival value records created in 2008 or earlier, but presumably they are still in the custody of agencies being managed in a recordkeeping environment that can only be described as post-custodial by default. The reasons for the lack of transfers are doubtless multifaceted and complex, and a recent Australasian Digital Records Initiative (ADRI) project investigated barriers to digital transfers in government jurisdictions in Australia and New Zealand. The report setting out the findings of the project will be published in late 2024.

The second point also requires explanation. As we've seen, the need for distributed custody arrangements for digital records was recognized as early as the mid-1990s, and the National Archives of Canada developed and published model terms and conditions for distributed custody arrangements. But even in Canada, it appears distributed custody arrangements for digital records have not been pursued.

One reason for the absence of distributed custody agreements for digital records was the development of digital preservation systems in the 2000s. Post-custodialism and distributed custody became influential in the 1990s because archival authorities did not have the infrastructure and systems to manage and preserve digital records. However, by the early 2000s, digital preservation standards and workflows began to be published and soon afterward, software solutions that implemented them became available. The National Archives abandoned distributed custody in 2000 when it embarked on a project, called Agency to Researcher, tasked with developing an in-house digital preservation program.¹⁵ Public Record Office Victoria's (PROV) Victorian Electronic Records Strategy (VERS) appeared in 1996 and was the basis of its digital archive and digital preservation standards. A partnership between the UK National Archives and software company Tessella produced Safety Deposit Box in 2003, which was to become Preservica. By the second decade of the 2000s, there were many commercial and open-source digital preservation systems to choose from. There was nothing preventing archival authorities taking custody of digital records – all they had to do was wait for the records to arrive.

Another reason for the absence of distributed custody agreements is their complexity. They are legal instruments and therefore enforceable with penalties for noncompliance (typically, immediate transfer to the archive). The legal nature of the agreements means that finalizing them can be time-consuming process involving legal teams scrutinizing every provision. Agencies may be encouraged to enter distributed custody agreements if a more streamlined model was adopted, for example, a generic set of provisions and requirements that an agency could opt into. The proposed streamlined approach should not impose significant legal barriers for agencies, and as much as possible, the conditions should not impose any significant extra costs on the agency.

A third reason is due to the continuing lack of clear rules for care of the records for which distributed custody arrangements are required. For analog records, these special rules usually refer to storage and conservation standards. For digital records, well-known digital preservation maturity models such as the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation¹⁶ and the DPC Rapid Assessment Model (RAM)¹⁷ could be repurposed as a set of conditions within a distributed custody agreement.

Conclusion

In a 2017 article, Mpho Ngoepe argues that the South African National Archives (SANA) is unconsciously following a post-custodial approach to the preservation of digital records

because SANA does not have the infrastructure to support the transfer, management, and preservation of digital records.¹⁸ Consequently, at the time of publication, almost no agencies had transferred records to the archive. Records remained in the custody of the agency, but the concern, naturally, was that records were being lost.

Although, in contrast, almost all the archives interviewed for the Big Data Project did have the infrastructure and systems to accept transfers, most government archives are still not receiving them in a regular, scheduled way; you could say that, like South Africa, we're following a post-custodial approach by default.

This article argues that big data collections are prime candidates for distributed custody arrangements (as found in theoretical discussions dating back at least as far as the 1990s) as they are high value, have ongoing business use, and come with technical and financial barriers to their transfer into the custody of the archive. However, determining which components of these collections are for permanent retention as national archives and then establishing the special rules for big data collections are not necessarily easy undertakings. These special rules – the terms and conditions – are what we need to develop to ensure appropriate management and control of these distributed collections, without imposing unreasonable costs and unhelpful complexity.

Notes on contributor

Dr. James Doig has worked at the NAA for more than 20 years. In that time, he has worked in many roles in collection management, including digital preservation, transfer, description, and collection review. He has presented regularly at conferences such as Australian Society of Archivists (ASA) and Records and Information Management Practitioners Alliance (RIMPA) and has published articles in *Archives & Manuscripts*, *American Archivist*, and *Script & Print*. He is on the Research and Practice subcommittee of the Digital Preservation Coalition. He has a PhD in medieval history from Swansea University.

Acknowledgements

I would like to thank Rowena Loo for her comments on a draft of this article.

Notes

1. For this and what follows see Simon Davis, *Looking Back to the Future: 30 Years of Keeping Electronic Records in the National Archives of Australia*, National Archives of Australia, Canberra, 2004.
2. The e-permanence logo was a lower case 'e' engraved into a stone tablet, mimicking a cuneiform clay tablet. The blurb read: 'The e-permanence symbol represents the new standard in recordkeeping developed by the National Archives for use by all Commonwealth Government agencies. Though it applies to all forms of records, the new recordkeeping standard is particularly suited to deal with the challenges presented by the new electronic environment which has engendered an elusive and transitory quality to the Government's information assets'.
3. A revised version, *DIRKS: A Strategic Approach to Managing Business Information*, was released in 2001.
4. Heavily influenced by the work of David Bearman, see in particular Margaret Hedstrom and David Bearman, 'Reinventing Archives for Electronic Records: Alternative Service Delivery Options', in Margaret Hedstrom (ed.), *Electronic Records Management Program Strategies*, Archives and Museum Informatics, Pittsburgh, 1993, pp. 82–98. Cf. Adrian Cunningham's criticisms in 'Some Functions Are More Equal than Others: The Development of a Macroappraisal Strategy for the National Archives of Australia', *Archival Science*, vol. 5, 2005, pp. 163–84.
5. Terry Cook, 'Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era', *Archives and Manuscripts*, vol. 22, 1994, pp. 300–28. For the arguments of custodialists and post-custodialists during this period see Don Boadle, 'Reinventing the Archive in a Virtual Environment: Australians and the Non-Custodial Management of Electronic Records', *Australian Academic & Research Libraries*, vol. 35, no. 3, 2004, pp. 242–52, and Alistair G. Tough, 'The Post-Custodial/Pro-Custodial Argument from a Records Management Perspective', *Journal of the Society of Archivists*, vol. 25, no. 1, 2004, pp. 19–26.

6. Stephen Ellis and Steve Stuckey, 'Australian Archives' Approach to Preserving Long-Term Access to the Commonwealth's Electronic Records', in Stephen Yorke (ed.), *Playing for Keeps: The Proceedings of an Electronic Records Management Conference, Hosted by the Australian Archives, Canberra, 8–10 November 1994*, Australian Archives, Canberra, 1995, p. 128, https://web.archive.org/web/20050308142922/http://ourhistory.naa.gov.au/library/playing_for_keeps.html.
7. Terry Cook, 'Leaving Archival Electronic Records in Institutions: Policy and Monitoring Arrangements at the National Archives of Canada', *Archives and Museum Informatics*, vol. 9, no. 2, 1995, pp. 141–9.
8. *Ibid.*, p. 149.
9. National jurisdictions: US National Archives and Records Administration (NARA); National Archives of Finland; Archives New Zealand; Public Records Office, Northern Ireland; National Archives of Estonia; Danish National Archives; Swedish National Archives (Riksarkivet); UK National Archives (TNA). Local jurisdictions: Landesarchiv Baden-Württemberg; Public Records Office Victoria (PROV); State Records NSW; Queensland State Archives.
10. Australian Bureau of Meteorology (BOM); Australian Bureau of Statistics (ABS); Geoscience Australia; Commonwealth Scientific and Industrial Research Organisation (CSIRO); Department of Social Services; Department of Agriculture, Fisheries and Forestry; Australian Taxation Office (ATO); Services Australia.
11. Each recording was deleted after the interview was written up.
12. The DIRKS Manual was used in the Commonwealth government from 2000 to 2007.
13. For example, Norway has NOARK 5: <https://www.arkivverket.no/forvaltning-og-utvikling/noark-standarden/noark5-standarden>. The European DLM Forum has published modular requirements for records systems, MoReq2010.
14. CAARA Policy 15: <https://www.caara.org.au/index.php/policy-statements/models-for-the-distributed-custody-and-management-of-government-archival-records/>.
15. In 2000, the NAA released *Custody Policy for Commonwealth Records*, which signalled an in-principle undertaking to accept custody of all digital records appraised as having archival value, regardless of format. On the Agency to Researcher project see David Pearson and James Doig, 'Tales from "The disK Files": Lessons Learnt from a Data Recovery Project in 2003–2006 at the National Archives of Australia', *The American Archivist*, vol. 85, no. 2, 2022, pp. 361–2.
16. <https://ndsa.org/publications/levels-of-digital-preservation/>.
17. <https://www.dpconline.org/digipres/dpc-ram>.
18. Mpho Ngoepe, 'Archival Orthodoxy of Post-Custodial Realities for Digital Records in South Africa', *Archives and Manuscripts*, vol. 45, no. 1, 2017, p. 34, doi: 10.1080/01576895.2016.1277361.