



REFLECTION

Reimagining the Commonwealth Record Series System

Carey Garvie and James Doig

National Archives of Australia, Canberra, Australia

Abstract

In the 1960s, Peter Scott proposed a new way of controlling records at the National Archives of Australia that became known as the Commonwealth Record Series (CRS) system. Acknowledging the ever-changing nature of governments, the CRS focused on the Series as the central entity for controlling records allowing connection to multiple Agents (creators/controllers). What constitutes a record though has always been open for discussion and has become potentially more ephemeral in the digital realm. This paper looks at recent work undertaken at the National Archives to reimagine the underlying data model of the CRS system to allow for more flexibility in capturing digital records.

Keywords: *Archival control; Series system; Digital records*

**... you should not let the functionality of existing mechanisms drive your decisions
about how you should describe and arrange digital content**

Digital Preservation Coalition
Novice to Know-How: Providing Access to Preserved Digital Content
May 2021

**The farther backward you can look
The farther forward you are likely to see**

Winston Churchill

These quotes sum up rather nicely both the challenges the National Archives of Australia has been trying to address over the past few years in revising the National Archives' archival control model and the approach we took in developing it. This reflection article will range backwards and forwards in time to give a flavour of the challenges faced and also demonstrate that in many respects these are not new challenges; in fact, in one form or another they have been around since the Australian Series System was developed.

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

The genesis of the work was the creation in 2017 of a short-term Branch within the National Archives called the Digital Archives Taskforce (DAT). The aim of the DAT was to accelerate digital transition at the National Archives by reviewing, developing requirements and improving or replacing current systems and processes, particularly in relation to the management of digital records. Our digital archive, including an in-house developed digital preservation software platform, had been operational since 2007 and RecordSearch, our modular archival management system, for much longer. The key modules of RecordSearch, including Search and Retrieve (the internal and public catalogue), Describe Records and Provenance (the module for intellectual control of records) and Transfer, Location and Lending (the module for physical control of records), had been introduced at different times between 1997 and 2001.

One of the areas of work identified by the DAT was to review our archival control model and develop an improved metadata schema for records. What started out as a seemingly straightforward task morphed into something much larger. Whilst there was universal agreement within the National Archives that our existing archival control model, the Commonwealth Record Series (CRS) system, was sound, it became apparent that our implementation of the record Item in RecordSearch was problematic. In particular, it became clear that the current management of items was compromising our ability to control the different representations of records that we receive either in transfers, for example, complex digital objects, or generated internally through digitisation and migration processes. This made developing a metadata schema for digital records difficult for a variety of reasons.

Some context and a historical interlude

To better understand this, let us take a step back and look at the Series system as originally conceived by Peter Scott, Ian Maclean and others. Figure 1 shows the original CRS system as drawn by Peter Scott in 1969.¹ The diagram is interesting for a number of reasons, for example in demonstrating the totality of his vision for an archival control system. However, it clearly shows the CRS system as a web of interconnected relationships with the record series at its centre, in much the same way that medieval maps show Jerusalem at the centre of the world.

In fact, the first thing that leapt out at us was the similarity with linked data visualisations, which, in turn, reminded us of the recently released Records in Context Conceptual Model and its call for the use of graph technologies to underpin archival description to ‘enable unbounded representation of networks of interconnected data objects as well as real world objects (represented by data)’.² Certainly, we felt that the ability to record and manage complex relationships between record Items would solve many of the problems staff were experiencing in trying to manage complex records that consist of various related parts, for example:

- the various physical representations of the same intellectual content, as is common with photographic records, audio-visual records and digital records;
- digital records that were linked with other records in the business system in which they were created and managed;
- digital records that must interact with other records or digital objects in order to be accessed or understood;
- instances of multiple parallel provenance, where the same records may exist in different series created by different agencies.³

Other standards like PREMIS, the international digital preservation metadata standard and its concepts of Intellectual Entity and Representations were useful here in both helping us to understand the problem and in developing a solution.

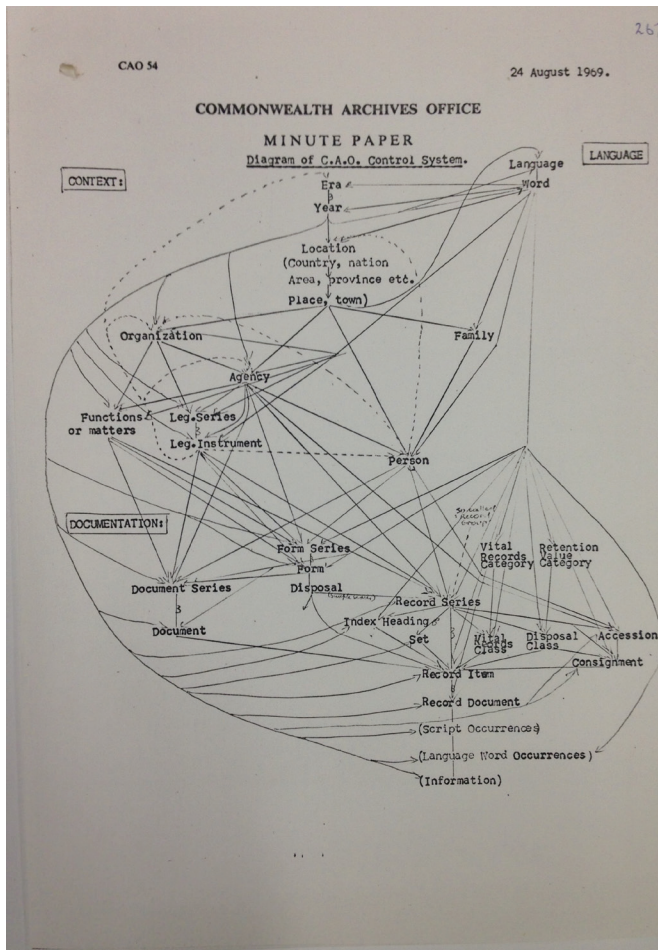


Figure 1. Diagram of the C.A.O. control system. National Archives of Australia, A750, 1967/19: Development of Context Control System, fol. 267.

We noted that these problems relate almost exclusively to the item entity. In the existing CRS data model, the other entities have well established relationships; for example, series has ‘related series’, ‘controlling series’, ‘previous series’, ‘subsequent series’, and ‘series controlled’. In fact, the relationships for series, agency and organisation were fundamental to the development of the CRS system. They were recorded in the paper-based finding aids (the Australian National Register of Archives and Documentation⁴) that were used from the 1960s to the introduction of the first computer systems at the National Archives in the mid-1980s. The first volume of the massive four volume tender documentation for a computer system issued in 1984 diagrammatically illustrates the relationships required in the system (see figure 2. Intellectual control was implemented in the Records Information Service, or RINSE, one of the three main applications developed).

The focus on the agency and series entities is to be expected, as the essential features of the CRS system are that the series is the basis of archival control and description, and that time-bound series relationships with the provenance entities are the basis of managing and recording administrative change over time. Item relationships were not a priority – in the analogue world, the need for piece or sub-item relationships was rare; in any case, item relationships

Figure 8.5
RECORD SERIES RECORD RELATIONSHIPS

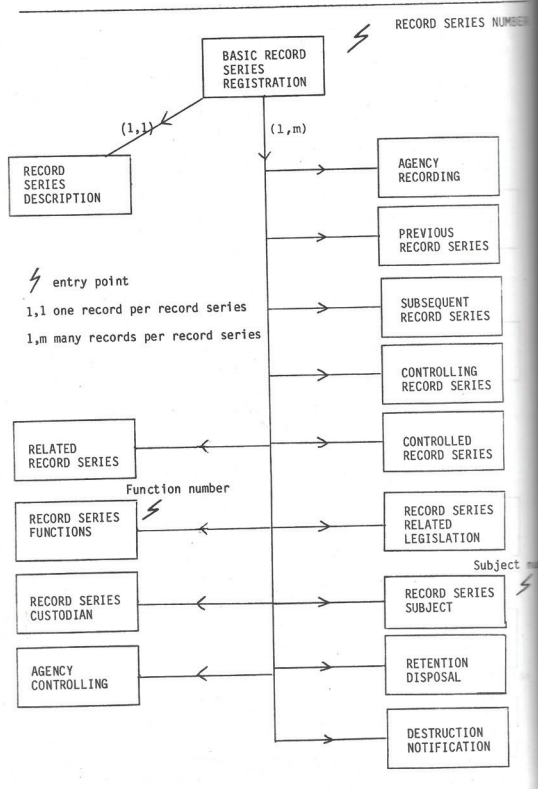


Figure 2. Diagram of record series record relationships (*Request for tender [volume 1] for the supply of a computer system for the Australian archives* [Canberra, ACT: Department of Administrative Services, 1984], page 8–44).

such as previous and later papers were written on the covers of paper files! In their exhaustive 1993 review of the CRS system, Russell Kelly and Mark Wagland observed: ‘There has always been the scope within the CRS system to include information at the piece level in inventories of items. This has been done on rare occasions. Because of the marginal nature of this information level, a comparative table for piece level information has not been prepared’.⁵

The last quarter of a century has seen a revolution in access to information brought about by the World Wide Web, globalisation, digitisation and changing research paradigms, and as a result, the expectations of users have forever changed. Users expect more and more granular levels of description and discovery, and these expectations are increasingly being realised through the trend for digital collections to be treated as big data sets that can be mined using computational techniques.⁶ At the same time, the management and preservation needs of complex digital records require relationships to be established with, for example, multiple aggregations of related records reflecting recordkeeping structures, dependent digital files like software or system files, or artefacts that provide meaning

and context like data dictionaries or a simple readme file in plain text. These developments have focused attention on the record item and the types of relationships that can exist within and between items.

When, in 1995, the Systems Integration and Redevelopment Project was conceived to integrate different applications and automate information retrieval and data capture, three-item relationships were built into the resultant system, RecordSearch: parent item-sub item; source item-copy item; and an odd relationship responding to a very 1990s issue – managing records from multiple series that have been copied for preservation or other reasons to physical carriers such as microfilm, photo albums, tape, compact disks, etc. This was the aggregate series-aggregate item-constituent item relationship. It is a relationship that has been rarely used for the purpose for which it was developed; it has been decoupled from the aggregate series concept and has been applied inconsistently over time. Similarly, the parent item-sub item relationship has been reinterpreted over time and implemented in many different ways. It was originally conceived as managing parts of items that were physically removed and stored elsewhere, for example, for preservation, security or other reasons. It has become used for most parent–child relationships, to manage aggregations of records and their component parts regardless of whether the component parts are physically separated. In effect, over the years, the aggregate item-constituent item and parent item-sub item have been used interchangeably to manage all types of hierarchical parent–child relationships. Description decisions were increasingly being driven by the capabilities of the archival management system, RecordSearch, and not CRS policy.

This notion of ‘CRS policy’ leads us down another interesting historical path. We have outlined in a potted way the development of the National Archives’ archival control systems, from paper registers to the first, unintegrated computer systems, to the current integrated computer system, RecordSearch. But where does CRS policy reside?

The CRS manual: another potted history

Peter Scott has said that he regretted not being involved in the development of a CRS Manual before he left the National Archives in 1989. The first CRS Manual was developed in the mid-1980s and was conceived as forming the single source of truth, as different practices had developed in the different state and territory offices. Originally, the CRS Manual was estimated as consisting of two volumes, and a draft of the manual was completed by July 1985.⁷ A report on the proposed format of the CRS Manual set out its purpose: ‘The manual will serve as a guide for Archives officers to the systems of intellectual control operating within the Australian Archives...At another level the manual will assist in the establishment of a consistent standard of documentation throughout the Archives...This standardization of format has become more essential as the Archives moves towards the implementation of the ADP [i.e., Automated Data Processing] System’.⁸ However, issuing the manual was delayed until the introduction of ADP as the computer system would result in ‘major changes to some procedures and it is hoped to incorporate these changes before issuing the manual’.⁹ In 1987, with the introduction of the computer applications RINSE, ANGAM II and the Physical Control System (PCS), a much expanded 13 volume ‘CRS Manual’ was conceived, which included volumes with detailed procedures for each major functional activity or process (e.g., a volume for RINSE, volumes with procedures for Series registrations, Agency registrations, administrative change and so on). Notwithstanding the completion of the CRS Manual and the release of a third edition in 1990, variations in key CRS definitions, for example, the major CRS entities, had arisen, which was causing confusion and resulting in inconsistent practice. One of the recommendations of the 1993 review of the CRS system was that ‘the definitions of terms used in the CRS System be set by and controlled from a central point within the Archives’.¹⁰ As a result, a completely revised, definitive edition of the CRS Manual appeared in 1997. With the

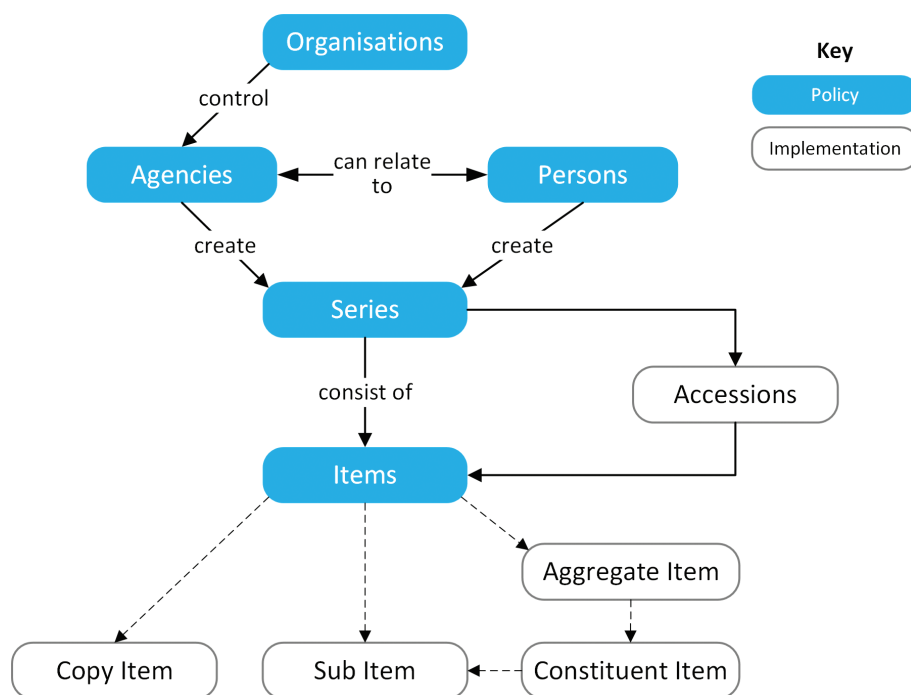


Figure 3. RecordSearch data model showing item relationships. Figure adapted from ‘Basic structure of the CRS system’, National Archives of Australia, The CRS Manual, October 2004.

introduction of RecordSearch, a new edition of the CRS Manual was released in 1999. Whilst it retained much of the language and definitions of the 1997 edition, it codified and defined the item relationship concepts that were introduced in the Identification (i.e. intellectual control) module of RecordSearch. The last major review of the CRS Manual was undertaken in 2004 and did not result in significant revisions of the 1999 edition.

The analysis of the development of the CRS Manual over time illustrates a couple of key points: first, the close, symbiotic relationship between the CRS Manual and the National Archives’ database systems for archival control, and second, the ongoing view that the manual must be a definitive and exhaustive account of descriptive practice. This can be compared to debates about the value of ‘black letter’ versus principles-based legislation. Black letter legislation attempts to cover every possible example and, therefore, is designed to be easy to interpret and make judgements, but it requires frequent amendment to accommodate every new situation, whereas principles-based legislation requires interpretation by Judges, which can lead to some idiosyncratic decisions, but it does not require constant amendment. Certainly, there is a strong case to redesign the CRS Manual and pull out its component parts: policy, procedures, data dictionary and system business rules, so that it remains flexible, responsive and relevant.

So, to summarise the situation at the commencement of the Archival Control Model (ACM) project:

- RecordSearch is a bespoke in-house database designed in the 1990s, primarily for paper-based records.
- It uses a 1:1 data model where intellectual and technical metadata are captured together to describe the Item (see Figure 3 for the RecordSearch data model).

To deal with access and preservation requirements, Item sub-types were introduced which

- were given strict definitions and constrained connections producing hierarchies as opposed to relationships,
- were increasingly being used beyond their original intent to deal with more complex configurations of records,
- were inconsistently applied for the same situations resulting in confusing and misleading item descriptions hampering access,
- did not have the flexibility and extensibility to deal with aggregations of records beyond three levels, and
- only documented analogue records ignoring the attached digital surrogates which are considered an important asset and part of the collection.

Archival Control Model project outcomes

The ACM project brought together subject matter experts from across the organisation a series of workshops, and sprints were run to investigate and design a potential solution. Whilst the principal focus was the Record entity the project also reviewed the implementation of other Entities to see how they could be updated to meet the complex challenges of the digital environment.

A valuable exercise to improve the understanding of recent developments in archival description and the emerging technology landscape were to invite Adrian Cunningham, a member of the ICA's Experts Group on Archival Description, to deliver a presentation on the Records in Context Conceptual Model. Some of the recent trends he identified included:

- Reimagining description in relation to new and emerging communications technologies and avenues for online sharing/exchange of descriptions;
- Trend towards separating the components of description (started in Australia with the series system, continued with International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition, etc.);
- Need for description to support multiple modes of access, plus renderings of descriptions for different audiences via different channels;

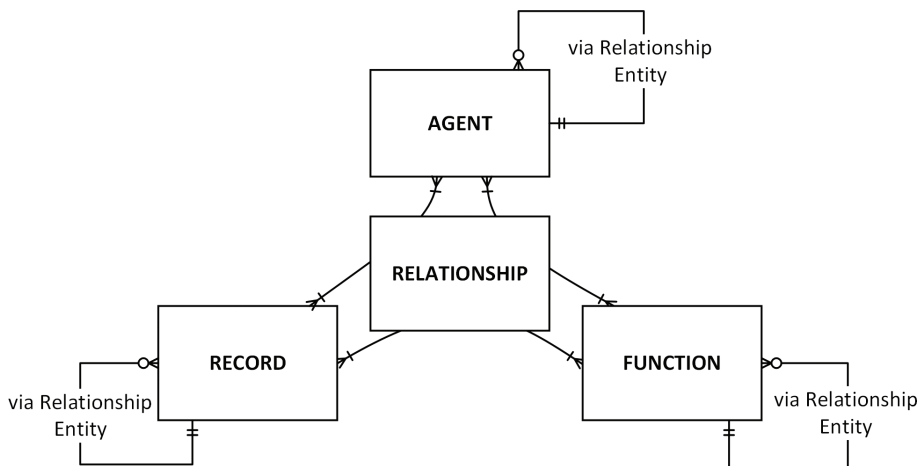


Figure 4. ACM data model. National Archives of Australia, Archival Control Model, 2 August 2019.

- Interoperability and automated reuse of descriptive metadata, for example, Linked Open Data;
- Convergence of archival description with metadata for recordkeeping (ISO 23081, Australian Government Recordkeeping Metadata Standard [AGRkMS], etc.).

Ultimately, the ACM project team settled on four entity types: Agent, Record, Function and Relationship with recommendations made to look at future implementation of Mandate and potentially Event, which is currently absorbed into Relationship (see Figure 4). For each entity, we created updated definitions and types and rules of application.

The project team also proposed adopting the use of relationship statements like ‘has part’ to create linkages, rather than the existing strictly defined relationship types. It also proposed adopting the concept of Intellectual Entities and Representations as described by PREMIS to manage digital surrogates.

These were based on concepts taken from the existing CRS system, AGRkMS, the Commonwealth’s implementation of AS/NZS 5478, and PREMIS the international standard for digital preservation metadata.

The goal was both to return to the original vision of the CRS system and move towards a linked data approach. Key changes to the existing CRS data model include:

- A simplification of the Record entity from Series, Item, Sub item, Aggregate and Constituent to Intellectual Units and Representations. An Intellectual Unit may be either a Series or Record unit whilst a Representation may be either original or created. This allows us to link multiple representations to the same intellectual content and identify and manage the primary archival representation including those created through the digitise and dispose policy where the original analogue record is destroyed and the digital surrogate is the primary archival representation, from which other representations can be made.
- The Agent entity has been changed from Commonwealth (i.e. federal government) Organisation, Agency or Person to Organisation, Person and Other allowing for the capture of technologies (software and hardware) and non-commonwealth entities involved with the creation of records.
- The movement of functions between agencies is a core part of the reasoning behind the CRS system. Previously, we had captured functions in a static thesaurus linked to the Agent, thus the series (in fact, migrated to RecordSearch from the earlier RINSE application). The new model promotes function to a more dynamic entity mapped to the core business functions of Agents as captured in Record Authorities.
- Finally, whilst relationships existed previously, they had become hidden as specific properties of other entities. The new model now requires that all links between entities occur through the Relationship entity. The new relationship entity captures both intellectual and technical relationships in semantic terms such as ‘has part’ rather than archival terms like constituent or aggregate item. This has been done to move us closer to being able to institute a linked data approach as Peter Scott originally envisioned.

Implementation challenges

The National Archives is currently in the process of upgrading our archival management systems including our digital archive. There are several challenges that we face in implementing the ACM data model and schema.

Whilst modern digital preservation systems align to the concepts in PREMIS, our RecordSearch catalogue database, developed over two decades ago, was developed on analogue principles where an item is a single representation such as a paper file. As such, it holds both the intellectual and physical/technical metadata at the same level.

Given that relationships form the core of the CRS system, to operate effectively they need to be automated as much as possible particularly given the scale of the collection, which is estimated at 40 million records with around 15 million described at Item level. This figure does not include the aforementioned digital surrogates and the born digital records that have not been described as well as we would like.

Existing archival processes are also a challenge, for example, the incremental partial release of records through the access examination process. Managing multiple digital access versions was not envisaged when RecordSearch was developed. In the analogue world, redactions and masks are generally contained within the original paper record; thus, the identifier does not change, only the access status of the record. In the digital world, new digital objects are created that require their own management and hence require their own unique identifier. These need to be clearly distinguishable from the unredacted master to prevent inappropriate release.

The key lesson that we have learnt is to see our data model as a living thing that will need to be regularly reviewed and updated to continue to meet the challenges ahead. Over time, we have confused system implementation with policy, and our descriptive practices have been driven to a large extent by the systems that implement the CRS data model, schema and descriptive rules. We have also tended to impose an analogue view onto digital records, resulting in a rich source of data being effectively hidden. We hope that the updated archival control model will assist us in reassessing our approach to records in all forms and improve access for our users.

Notes on contributors

James Doig is an Assistant Director, Digital Archives Innovation and Research at the National Archives of Australia with over 20 years of experience in digital preservation and digital archiving.

Carey Garvie is the current Manager of Digital Preservation and previous Project Manager of the Archival Control Model redevelopment project within the Digital Archives Taskforce.

Notes

1. National Archives of Australia, A750, 1967/19: Development of Context Control System, fol. 267.
2. International Council on Archives, 'Records in Context: A Conceptual Model for Archival Description', consultation draft v0.1 September 2016. Version 0.2 was published in July 2021, by the ICA Experts Group on Archival Description.
3. Chris Hurley, 'Parallel Provenance', *Archives and Manuscripts*, vol. 33, nos. 1 & 2, 2005, in particular 2, pp. 68, 80–81.
4. The registers had procedures for completion, for example 'Record Series Registration Sheet: Notes on Completion' dated 18 April 1971, with a hand-written note by Peter Scott at the top of the first page that the document was based on a draft of 22 August 1966, see A750, 1967/15: Development of Documentation Control System, ff. 302-290. Relationships described in the Note for Completion are Agency Recording (fols. 302-301), Previous record series – diachronic and synchronic (fols. 294-293), Related, controlling or controlled series (fols. 293-292), agency controlling (fols. 292-291). For the 1966 draft procedures see A750, 1966/88: Registration Activities (series, agencies, etc.) Procedure (fols. 50-42).
5. R Kelly and M Wagland, 'CRS Review Report', Australian Archives, 1993, unpublished internal report. Their 262 page review formed the basis of their important paper on the CRS System, Kelly and Wagland, 'The Series System – A Revolution in Archival Control', in S McKemmish and M Piggott (eds.), *The Records Continuum: Ian Maclean and Australian Archives First Fifty Years*, Australian Archives, 1994, Clayton, Victoria: Ancora in association with Australian Archives. Clayton, Victoria: Ancora in association with Australian Archives.
6. Luciana Duranti, 'From Digital Diplomats to Digital Records Forensics' *Archivaria*, vol. 68 (2009), pp.39–66; Frederick B. Cohen, 'Digital Diplomats and Forensics: Going Forward on a Global Basis' *Records Management Journal*, vol. 25 (2015), pp.21–44; Michael Moss, David Thomas, and Tim Gollins, 'The Reconfiguration of the Archive as Data to be Mined' *Archivaria*, vol. 86 (2018), pp.118–151; Devon R. Mordell, 'Critical Questions for Archives as (Big) Data' *Archivaria*, vol. 87, pp. 140-161.

7. National Archives of Australia, A750, 1984/233: Development of CRS System Manual – Admin Structures and Analysis Section, ‘CRS System Manual’, 1985 internal report.
8. A750, 1984/233, ‘CRS System Manual’, 1985 internal report, fol. 40.
9. A750, 1984/233, minute dated 16 July 1985, fol. 50.
10. R Kelly and M Wagland, ‘CRS Review Report’, Australian Archives, 1993, unpublished internal report, Recommendation 1, p. 35. Also see A750, 94/1072: CRS Manual – Revision – Overview, fols. 62-60.