https://doi.org/10.2352/issn.2168-3204.2020.1.0.101

Reimagining the Archival Control Model of the National Archives of Australia for the digital age.

Carey Garvie; National Archives of Australia; Canberra, Australia.

Abstract

In the 1960's Peter J Scott and colleagues at the now National Archives of Australia developed a new way of documenting records known as the Australian 'Series' System. Adopted by public records institutions in Australia and New Zealand, and selectively around the world, this approach forms the basis of the National Archives Commonwealth Record Series (CRS) system. In 2018 following views expressed that digital records pose a serious challenge to traditional ways of contextualization it was decided to review the CRS system in this respect. This paper looks at the process of that review and the eventual development of an enhanced model merging concepts from PREMIS with the CRS to enable a more flexible approach of documenting records in all forms.

Introduction

In the 1960's Peter Scott and colleagues at the now National Archives of Australia developed a new way of documenting records commonly known as the Australian 'Series' System. Rejecting the concept of single hierarchical descriptions it separated information about record creators from the records themselves allowing for the documentation of multiple provenance (1). This approach forms the basis of the Commonwealth Record Series (CRS) system that has been used by the National Archives since that time.

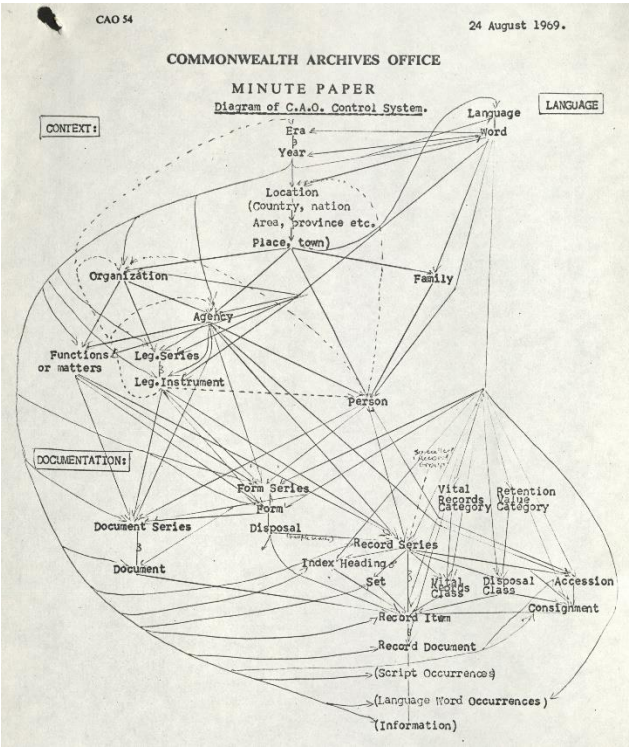

***Figure 1**. Peter Scott's original diagram of the CRS System*

In 2018 following views expressed internally and externally that digital records pose a serious challenge to traditional ways of contextualizing it was decided to review the CRS system to see if it could withstand this challenge. Of particular concern was that the CRS metadata model may not be flexible enough to deal with the complex relationships that can exist between and within records, for example an email with attachments, the levels of control within an Electronic Document and Records Management System (eg. Trim Box, File, Document and attachments) or semi to unstructured computer filing systems. In moving towards more granular levels of control and access, we recognized a need for more effective ways of describing complex relationships between records in both the analogue and digital domains.

The initial review began with a comparison of the CRS system against a variety of models used in the Archive and Library sector both nationally and internationally to see if an alternative approach provided a solution. Models reviewed included the International Council on Archives Records in Context – Conceptual Model (RiC-CM) (2), BIBFRAME (Bibliographic Framework) (3), PREMIS: 'PREservation Metadata: Implementation Strategies'(4), The National Archives UK 'digital cataloguing practices'(5) and the Public Record Office of Victoria's Archival Control Model (6) along with our own internal model used for documenting audio-visual records.

From this it was concluded that while there were many approaches there was no perfect solution. What was evident is that more approaches were moving towards entity relationship models and linked data. As the underlying principles of the CRS system align with this approach it was decided to continue with enhancing the metadata schema to enable more effective management of digital records. While this seemed straight forward we immediately hit a problem with our concept of an Item and its implementation within our cataloguing systems.

Identifying the Problem

As noted earlier the CRS system utilizes a relational model between Agents and Records allowing for multiple connections. In respect to Records there are only two types Series and Items. At least that was the intent as shown in figure 2. With implementation of the CRS in the Archives' custom collection management system *RecordSearch* operational needs had forced false hierarchies into the system as the organization tried to keep pace with the often rapid changes in the digital sphere and the desire of the public for ever more granular levels of description and access. Item sub types were introduced to manage access and preservation requirements. More importantly these new sub types (Aggregate, Constituent, Sub and Copy):

- 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.
- Did not have the flexibility and extensibility to deal with aggregations of records beyond three levels.
- Only documented analogue records ignoring the attached digital surrogates.

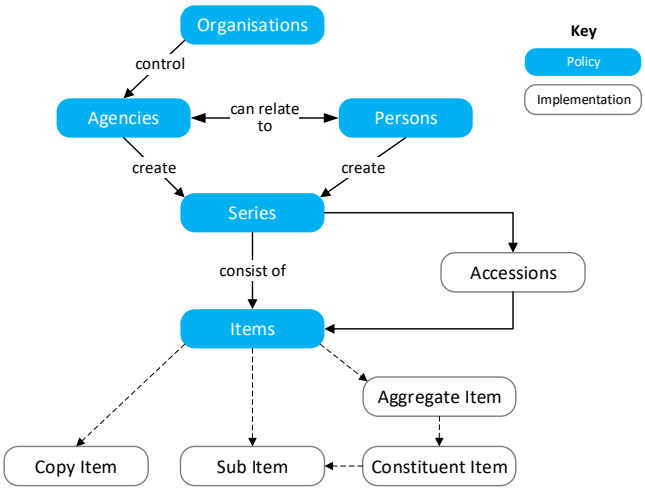

***Figure 2**. CRS System Policy and Implementation*

A simple example of this issue can be seen in the documentation of Series B941 *Black and white general prints, alphabetical series*. A series of B&W prints created by the Trade Publicity Directorate (CA 3358) these were originally housed in groups in yellow envelopes each of which was described as a single Item. Rehoused into archival albums and digitized by Preservation each print now became an item with multiple digital surrogates attached (TIFF and two JPEGs). In the approach shown below for CATTLE BEEF/GRAZING/MURRAY GREY/5 the Album is described as an aggregate item, the first print (original item description) was converted to a constituent item and the remaining prints and folios were attached to this as sub items. The digital surrogates are not documented in any system but are instead stored in different locations and the JPEGs attached to the record through the search engine.

This is just one of many examples where the original intent of aggregates and constituent item types have been repurposed to deal with records that have been rehoused and/or digitized. Investigation showed that for the same situation offices had taken alternative approaches depending on their interpretation of the description guidelines. More disconcertingly while these situations were arising to deal with the creation of digital surrogates the digital records themselves were not being documented, were all being given the same identifier as the original record and were not being effectively managed in respect to preservation policies. It is estimated that the National Archives has over 120 million digital surrogates with the figure growing on a daily basis through preservation and access digitization programs.

To complicate matters the audio-visual component of the collection is being managed in a separate commercial digital asset management system. While this system provides effective documentation and management of digital records it takes an entirely different descriptive approach for Items that does not align to the one in RecordSearch. As a result audio-visual records while better managed are misleadingly described in the search engine leading to increasing issues for access.

In respect to born digital records previous efforts to arrange and describe EDRMS and computer folder systems had also led to a mix of approaches. In some cases all of the documents contained in an EDRMS File were given the same title and identifier as that file. In other cases documents contained in computer folders may or may not have been described separately to the folder. Acknowledging that these approaches were unsustainable and did not accurately reflect the records it was decided to re-imagine the Record entity for the digital age.

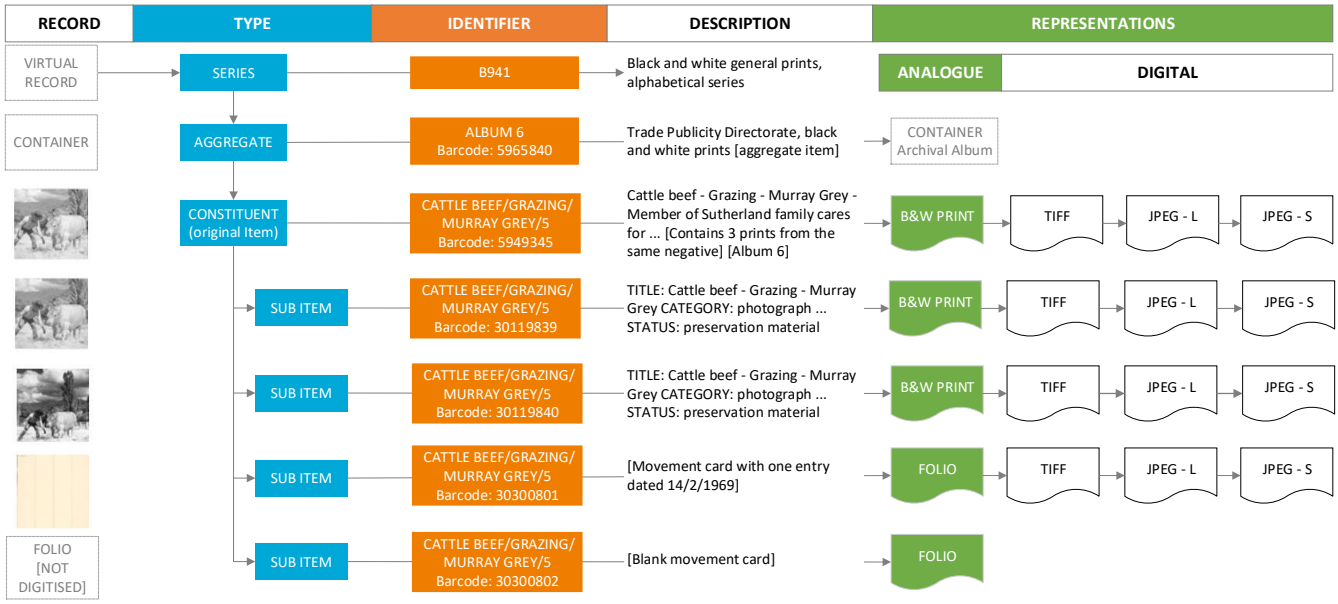

***Figure 3**. B941: CATTLE BEEF/GRAZING/MURRAY GREY/5*

Designing a Solution

Bringing together Subject Matter Experts (SMEs) from across the organization a series of workshops and sprints were run to investigate and design a potential solution. While the principal focus was the Record entity the project also reviewed implementation of the other Entities to see how they could be updated to meet the challenges of the digital age.

With respect to the Record entity three options were put forward to the SMEs for consideration.

- **Option 1:** Continue to use the existing Item sub types and expand as required to meet the increasing levels of aggregations. Review current definitions and guidelines for use to provide more consistency. Document each digital surrogate as an Item separate to the Analogue source but within the same Series.
- **Option 2:** Remove the use of Item sub types going forward. Retain the Record types: Series, Item and use relationship statements like 'contains', 'has part' to create linkages. Adopt the concept of Intellectual Entities and Representations as described by PREMIS to manage digital surrogates.
- **Option 3:** Remove the use of record types: Series, Item going forward. Use relationship statements like 'contains', 'has part' to create linkages between records. Adopt the concept of Intellectual Entities and Representations as described by PREMIS to manage digital surrogates.

It was agreed by the group to discount option 1 as it was too close to the current system which was showing considerable signs of strain already. A small sprint team was formed to test options 2 & 3 as possible solutions. A selection of examples were chosen for testing including an existing digitized series, the series containing our own administrative records from our EDRMS that had been transferred in 2012, a staff members personal computer files and our own cataloguing database RecordSearch.

Testing showed that both options were viable. In particular staff liked the Intellectual Entity and Representation approach as it allowed for the capture of Items that did not have a one to one relationship to a physical or digital unit for example a computer folder, EDRMS File Box or File. While Series have always been seen as intellectual constructs with multiple connections Items are more traditionally seen as discrete physical or digital units. Focusing on relationships and separating intellectual metadata from technical metadata was seen as a more flexible approach to managing both digital and analogue records in their many aggregations. This included allowing for records to be described in multiple ways for access purposes that still respected the original record but enhanced access.

That said while both options were seen as viable the majority of the group believed that there was still a need to distinguish the Series as a type in its own right with particular regard to the need for backwards compatibility. It was agreed that adopting option 2 now

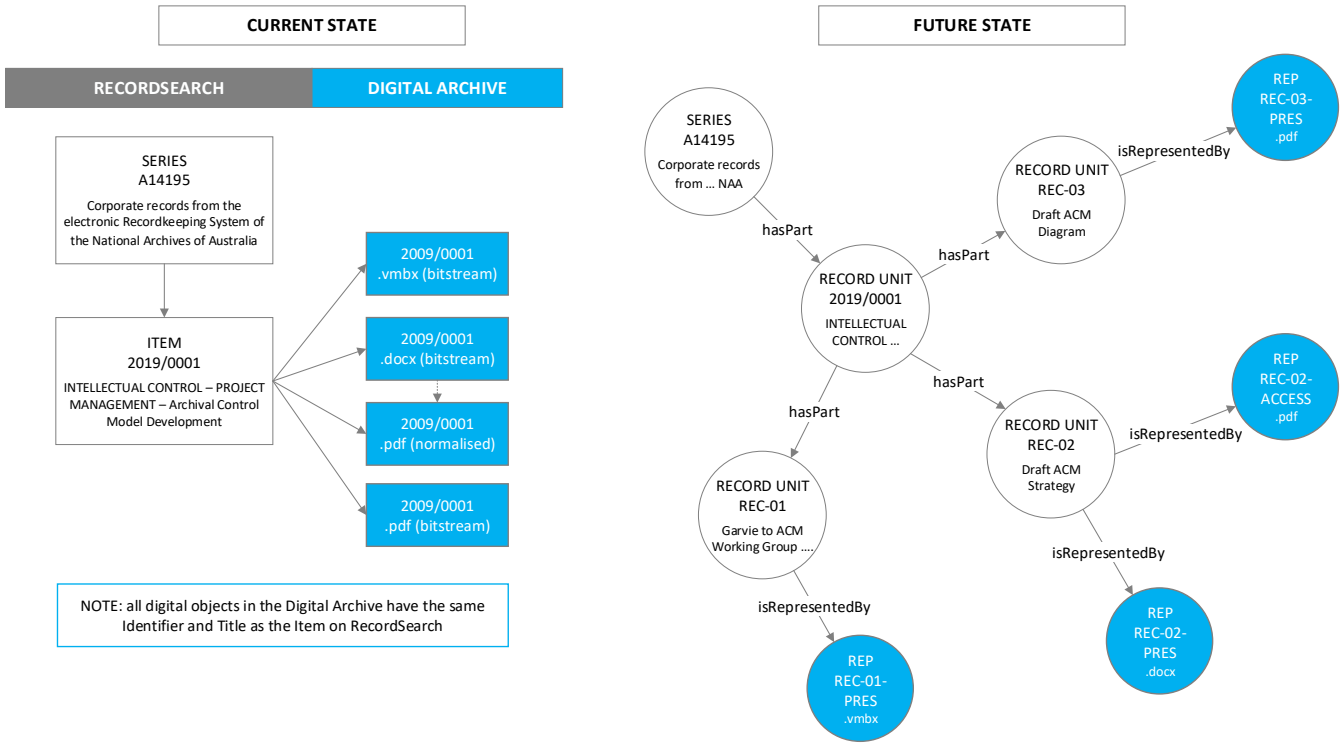

***Figure 4**. Mapping of current state to proposed future state for Series A14195*

allowed room to transition to option 3 in the future. The resulting model for the Record entity can be seen in figure 5.

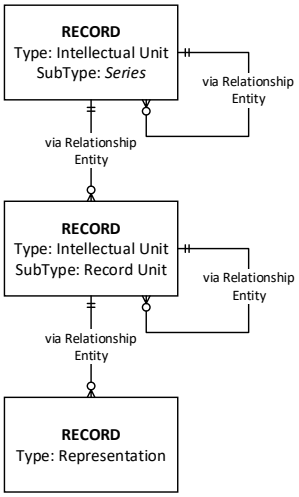

Figure 5. Archival Control Model Record entity

A New Model

As noted earlier while the project began in response to issues observed with implementation of the Item the project had a broader objective of reassessing the entire data model. While the original CRS system as shown in figure 1 was designed to capture a vast array of relationships as shown in figure 2 implementation had focused on only a select few: Organization, Agency, Person, Series and Item. It became evident in re-imagining the Record entity that relationships were central to how the CRS system operates.

In the 1990's the National Archives began working with colleagues from across the information management sector to develop a Recordkeeping Metadata Standard for Commonwealth Agencies. Evolving over time the most recent version published in 2015, the *Australian Government Recordkeeping Metadata Standard version 2.2 (AGRkMS) (8),* adopts a multiple entity approach: Record, Agent, Business, Mandate and Relationship. Drawing on this approach and the use of Intellectual Entities and Representations from PREMIS a new data model emerged. The updated Archival Control Model is comprised of four entities: Record, Agent, Function and Relationship.

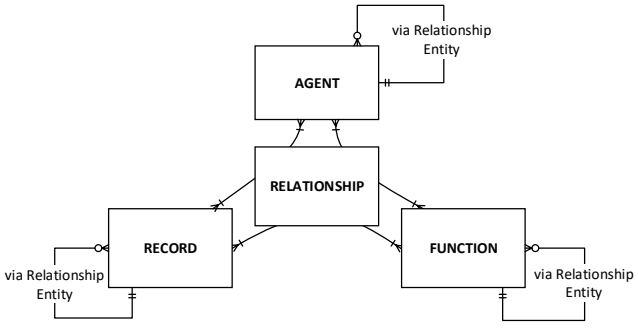

***Figure 6**. National Archives of Australia – Archival Control Model*

The key changes instituted were:

- Changing the **Record** entity from Series and Item to Intellectual Units and Representations. Intellectual Units now incorporate Series and Record Units (the re-conceptualized Item) and Representations can either be original (primary archival record) or created.
- Broadening the **Agent** entity to cover non-commonwealth and technology agents. The new Agent categories are now Organization, Person and Other.
- Converting **Functions**, originally hidden in a static Thesaurus, to dynamic entities based on the core business functions of Agents as captured in Record Authorities.
- Establishing a **Relationship** entity. While relationships existed previously they had become hidden as specific properties of other entities. It is now mandatory 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 linked data as Peter Scott originally envisioned.

Implementation

The new Archival Control Model was approved by the Executive Board in August 2019. An updated metadata schema has also been developed merging the CRS schema with components of AGRkMS and PREMIS. In late 2019 the National Archives began a pilot to test new digital preservation and archival management systems to replace existing in house technology that is now over 20 years old. As part of this pilot test trials for implementing the new model and schema are proposed.

To date evaluation has occurred on the data models of the two systems to determine how much configuration work would be required to implement the proposed Archival Control Model. It has always been acknowledged that compromises will need to be made during implementation based on the capabilities of the systems being used. While the digital preservation system aligns well to PREMIS the major challenge has been finding an archival management (cataloguing) system that aligns to the CRS. Some of the key challenges observed include:

- A preference for defined hierarchies over relationships. Most systems are designed for the US & European markets which utilize the Fonds system of description as outlined in *ISAD(G): General International Standard Archival Description*.
- Automating relationships. 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.
- Operating at scale. The National Archives has an estimated 40 million records with around 15 million described online. This figure does not include the afore mentioned 120 million digital surrogates and the born digital records that have not been described at document/object level.
- Managing the peculiarities of our legislation. Under the *Archives Act 1983* [10] we are responsible for declassifying records. As records are open after 20 years from date of creation it is not uncommon for them to be redacted several times during their life in response to changes in sensitivities

and release of previously closed period information. This requires us to manage a variety of versions of the record for both preservation and access.

Future Challenges

As we continue to test implementation options there are two key challenges that we face. The first is managing user access. We need to ensure the decisions we make enhance rather than impede user access. One of the options we have been looking at is implementing the new Archival Control Model in the Discovery interface. This would allow us to minimize changes to existing systems while we decide when and how they will be replaced, provide a more consistent view of our entire collection and allow us to be more flexible in addressing the second challenge of transitioning to the new model.

Given that we have over twenty years' worth of Items described using the existing model, and are continuing to create more daily, we will need to make some practical decisions in respect to future reconfiguration work. This will become important when we need to either migrate existing data into a new system or reconfigure our existing system to the new model. Work still needs to be undertaken to determine which records can be easily reconfigured to the new model, the amount of work required to do this and when in the data migration phase this work will need to occur.

The key lesson that we have learnt is to see our data model as a living model 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 by the former. 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.

References

[1] Cunningham, A, Millar, L & Reed, B, Peter J Scott and the Australian 'Series' System: its origins, features, rationale, impact and continuing relevance, International Congress on Archives, Brisbane, Australia, 23 August 2012.

[2] International Council on Archives, Records in Context: A conceptual model for archival description, consultation draft v0.1, September 2016.

[3] Library of Congress, Overview of the BIBFRAME 2.0 Model, April 21, 2016.

[4] PREMIS Editorial Committee, PREMIS Data Dictionary for Preservation Metadata version 3.0, revised November 2015.

[5] The National Archives (UK), Digital cataloguing practices at The National Archives, March 2017.

[6] Public Record Office of Victoria, Archival Control Model, 30 August 2017.

[7] National Archives of Australia, The CRS Manual: Registration and description standards for the Commonwealth Record Series (CRS) System, October 2004.

[8] National Archives of Australia, Australian Government Recordkeeping Metadata Standard version 2.2 (AGRkMS), June 2015.

[9] International Council on Archives, ISAD(G): General International Standard Archival Description – Second edition, 2000.

[10] Australian Government, Archives Act 1983, No.79, 1983, Compilation No. 42, 25 April 2019.

Author Biography

Carey Garvie received her BAppSc in Conservation of Cultural Materials from the University of Canberra (2003) and is currently completing a Masters in Information Studies (Data Management) at Charles Sturt University. Carey currently works for the National Archives of Australia in the Digital Archiving unit focusing on Data Governance and Analytics. Prior to this she was responsible for managing preservation and digitization services.