

# Recent Developments in Standards for Archival Description and Metadata

## Adrian Cunningham

National Archives of Australia

*Presented at the International Seminar on Archival Descriptive Standards, University of Toronto, March 2001*

- What is archival description?
- Standards for archival description
- The Australian 'Series' System
- What is metadata?
- What is recordkeeping metadata?
- Standards for electronic recordkeeping metadata
- Conclusion
- Notes

## What is archival description?

According to the *International Standard for Archival Description (General)* 2nd edition, the purpose of archival description is to:

'... identify and explain the context and content of archival material in order to promote its accessibility. This is achieved by creating accurate and appropriate representations and by organising them in accordance with predetermined models'<sup>(1)</sup>

Traditionally, archival description has referred to the process of establishing intellectual control over archival holdings following the transfer of records to archival custody. In other words, the production of surrogate descriptions 'whose primary purpose is to help researchers find relevant records and understand something of their purpose and origins'.<sup>(2)</sup>

Recently, however, the advent of electronic records has encouraged archivists to think more broadly about the scope and purpose of archival description. A significant addition to the introduction to the second edition of the recently published *International Standard for Archival Description (ISAD-G)* has this to say:

'Description-related processes may begin at or before records creation and continue throughout the life of the records. These processes make it possible to institute the intellectual controls necessary for reliable, authentic, meaningful and accessible descriptive records to be carried forward through time.

Specific elements of information about archival materials are recorded at every phase of their management (e.g., creation, appraisal, accessioning, conservation, arrangement) if the material is to be on the one hand securely

preserved and controlled, and on the other hand made accessible at the proper time to all who have a right to consult it. Archival description in the widest sense of the term covers every element of information no matter at what stage of management it is identified or established. At every stage the information about the material remains dynamic and may be subject to amendment in the light of further knowledge of its content or the context of its creation.<sup>(3)</sup>

I shall have more to say about this shift in thinking later. For the moment it is sufficient to think of archival description as a dynamic and iterative process that can commence at or before the time of records creation and continue for as long as the record exists and sometimes even after it ceases to exist.

Perhaps the most important concept to understand when discussing archival description is the concept of 'record'. Records are the core business of archivists and both the subject and the object of archival description. According to the Australian Records Management Standard (AS 4390) records are:

Recorded information, in any form including data in computer systems, created or received and maintained by an organisation or person in the transaction of business or the conduct of affairs and kept as evidence of such activity.<sup>(4)</sup>

Explicit in this definition is the notion that records have provenance and context – records are created, received or maintained by an organisation or person. Records are not entirely self-contained or disconnected objects or packages of information. The major feature that sets archival description apart from the world of library cataloguing or museum curating is the archival principle of *respect des fonds*. This principle translates to mean that records have to be controlled, described and understood in the context of their creation and use. For archivists the critical questions are - who created, received and used the records in the course of what activity?

Flowing on from this notion of provenance is the idea that records not only relate to people, organisations, functions and activities – they also relate to each other. Usually, when conducting activities people and organisations create more than one record. Collectively, these aggregations of records constitute evidence of the activity. The evidential nature of records is derived from the fact that they sit within inter-related aggregations of records. A feature of archival description has been a focus on describing aggregations of records created by a particular individual or organisation (the *fonds*), or that relate to a particular function or activity (the series).

A common technique used in archival description to reflect the organic nature of records aggregations is the use of multi-level description. This is the practice of producing archival descriptions that proceed from the general to the specific or, alternatively, using relational databases to construct descriptions at the fonds, series, file and/or item level and documenting the part/whole relationships that connect these different levels of description.

## **Standards for archival description**

Unlike libraries, which were compelled to adopt standards from an early stage in order to stop different libraries cataloguing the same book thousands of times over in thousands of different ways, archives have been much slower to adopt commonly agreed descriptive standards. When describing unique materials, unique descriptions and unique methods of description seemed far more acceptable.

Even today there are wide variations in archival descriptive practices from institution to institution and from country to country. Archivists are very often creatures of habit who become attached to their own idiosyncratic way of doing things and can be very reluctant to change their practices in the name of standardisation. Archivists may all subscribe to the theory of *respect des fonds*, but there are an infinite number of ways in which this theory can be applied in actual descriptive practice.

Nevertheless, over the past 10-15 years there has been a growing acceptance by archivists of the need for descriptive standards. There is a recognition that such standards help the end user by reducing the variety of descriptive systems that they need to master when conducting cross-institutional research. It is also recognised that the promotion and adoption of standards can encourage improved practices within archival institutions. However, it is the advent of computers and electronic networks that has given the major impetus to the development and adoption of descriptive standards.

When designing automated systems for providing intellectual control and access to records, systems designers like to work with pre-existing standards. Systems are built to satisfy functional requirements. If the functional requirements have not been articulated, then someone has to sit down and write them. Not only is this an expensive process, there is also the risk that if you are doing it on your own or in a small team you might get it wrong. The existence of commonly agreed standards defined and adopted by a professional community of experts saves people the time and effort of having to 'reinvent the wheel'. It also gives commercial software vendors the certainty that they can design a product that is likely to meet the needs of a multitude of customers, rather than just one customer. Of course, in a networked environment standards are absolutely essential if information is to be shared or exchanged between distributed computers or different networks.

Not entirely in jest, somebody once said that 'the nice thing about standards is that there are so many of them to choose from!' The field of archival descriptive standards is no exception to this observation. Internationally, there now exists a variety of archival descriptive standards to choose from. The diversity is partially a reflection of the fact that the different standards focus on different (though not mutually exclusive) aspects of the descriptive process, and partially a reflection of different national traditions and differing conceptual frameworks.

The peak international standard, first issued by the International Council on Archives in 1994 and now into its second edition, is **ISAD(G)** – the International Standard for Archival Description (General). This standard 'provides general guidance for the preparation of archival description. It is meant to be used in conjunction with existing national standards or as the basis for the development of national standards'.<sup>(5)</sup>

Beyond ISAD(G) are a veritable alphabet soup of other more specific descriptive standards:

**MARC-AMC** – a variant on the MARC (Machine Readable Cataloguing) standard for Archives and Manuscript Collections, used for producing archival descriptions for incorporation into library databases;

**APPM** – Archives, Personal Papers and Manuscripts, an Anglo-American Cataloguing Rules-type manual for the description of archival holdings, used mostly in the United States<sup>(6)</sup>;

**MAD** – Manual of Archival Description, a British manual for the production of standardised archival finding aids<sup>(7)</sup>;

**RAD** – Rules for Archival Description, the Canadian descriptive standards bible<sup>(8)</sup>;

**EAD** – Encoded Archival Description, a standardised set of semantic rules and syntax for the encoding in SGML (Standardised General Markup Language) or XML (eXtensible Markup Language) of archival description for Web-based access, searching and data exchange. Originally developed in the United States to facilitate access to and searching of traditional archival finding aids over the World Wide Web, but becoming increasingly popular internationally as a *lingua-franca* for the Web-based exchange of heterogeneous multi-level archival descriptions.<sup>(9)</sup>

### **The Australian 'Series' System**

Since the 1960s government archives in Australia have developed yet another approach to the description and intellectual control of records. The 'Series' System, as it has become known, while it is solidly grounded in archival theory, differs from other approaches to archival description in a number of significant ways.

Firstly, unlike other descriptive standards outlined above, the series system is used to describe both 'current' and 'historical' records. In other words, the system is custodially non-specific – records can be described by the series system well before they are transferred to archival custody. Indeed, it is often the case that records that will probably never be transferred to archival custody are described using the series system.

As such, the series system provides for a more dynamic approach to the intellectual control of records – an approach that foreshadowed the so-called post-custodial revolution that I alluded to earlier when I described significant changes to the new edition of ISAD(G).<sup>(10)</sup>

Another significant feature of the series system is that it accommodates the documentation of multiple provenance. One of the problems with traditional archival descriptive practices is that it is based on the assumption of unitary provenance – ie. that an aggregation of records can only have one creator. In reality, record aggregations can have many creators – a fact that is most evident in modern bureaucracies where regular changes to the machinery of government can see functions, and hence active records series, transferred from the control of one agency to that of another agency. Multiple provenance (of both the successive and simultaneous variety) can also be found in aggregations of family papers and in recordkeeping systems where the private records of an individual are intermixed with the records that individual accumulates in some official capacity acting on behalf of one or more organisations.

The series system copes with multiple provenance by abandoning the use of the *fonds* as the primary locus of intellectual control, adopting instead the series as the highest level of archival description. This approach permits a more sophisticated and accurate documentation of context through the creation of separate but linked descriptions of records, records creators and their functions and activities. It is worth emphasising however, that while the series system does not involve the production of *fonds*-level description of records, this in no way implies that the system constitutes a rejection of the principle of *respect des fonds*. On the contrary, proponents of the series system assert that their approach is a more accurate means of representing the true complexity of the *fonds*. At this point it is useful to highlight the distinction between the inputs to an archival descriptive system and the outputs of that system. While the series system does not require *fonds*-level descriptive inputs, it is certainly capable of generating *fonds*-level descriptive outputs whenever such outputs are needed.<sup>(11)</sup>

The Australian approach of producing separate, but linked descriptions of records and records creators is accommodated in the international descriptive standards arena by the combination of ISAD(G) and its companion standard, **ISAAR(CPF)** – The *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, which was published by the International Council on Archives in 1996.<sup>(12)</sup>

More recently, in March 2001, Yale University in cooperation with the Research Libraries Group, convened an international workshop in Toronto to develop a Document Type Definition which will enable the creation and online exchange of standardised XML encoded descriptions of archival provenance entities. To be called Encoded Archival Context (**EAC**), this new standard will be a

companion to Encoded Archival Description (EAD) in the same way that ISAAR(CPF) is a companion to ISAD(G).

## **What is metadata?**

When most of us first encountered the term metadata, we were probably repelled by yet another debasement of the English language by a bunch of barbarian techno-boffins. The fact that the term can very often mean quite different things to different people simply highlights its slippery and infuriatingly imprecise definition. If you talk to software programmers about metadata they will almost certainly be imagining something very different to a group of librarians discussing the same term.

The term metadata emerged in the IT community many years ago. In those days it referred solely to the data that was necessary to make sense of data stored in a computer system. The Greek prefix 'meta' is defined in the Oxford Dictionary as 'denoting position or condition behind, after, beyond or transcending'. The definition of metadata as 'data about data' is as precise a definition as most people are prepared to venture. The imprecision of this definition has since allowed it to be applied to any computer-related descriptive information. Indeed, use of the term has become so flexible that now it does not even have to be related to computer technology - any old data about data can now be metadata.

The main point I want to make here is that metadata is simply a new term for information that has been around for a very long time, but which now looks a bit different due to the advent of computer technology. A better more informative definition of metadata than 'data about data' that I would propose is:

*Structured information that describes and/or allows us to find, manage, control, understand or preserve other information over time.*

If we think of metadata in these terms, then archivists are metadata experts - it is just that we tend not to think in those terms about the work that we do and the things we produce. Traditional archival finding aids, index cards, file covers, file registers, the headers and footers of paper documents - all of these things contain metadata and all of them have their computerised equivalents that may or may not look different, but which nevertheless fulfill the same functions.

There are many different types of and uses for metadata. These include:

- systems operating metadata (that is, the metadata that is necessary to make sense of a computer software platform);
- data management metadata (eg. ISO 11179);
- information management metadata;
- recordkeeping metadata (more of which later);

- resource discovery metadata (eg. the 'Dublin Core' metadata standard for Web-based resource description and discovery)<sup>(13)</sup>;
- digital preservation metadata (eg: the Open Archival Information System or OAIS reference model)<sup>(14)</sup>; and
- rights management metadata (eg. the INDECS initiative)<sup>(15)</sup>.

These categories of metadata are NOT mutually exclusive. Particular metadata schemas (or sets) and the elements that make up those schemas can serve more than one purpose. Indeed, one of the sources of confusion about metadata is people's failure to realise that there is often a great deal of overlap and numerous inter-relationships between various metadata sets - hence the need for the so-called 'crosswalks' between various metadata standards which identify linkages between related elements in different metadata sets. This is not to say that particular metadata sets are redundant. On the contrary, metadata sets usually get developed in response to a particular set of well-defined requirements. It is just that many sets of requirements overlap with other related sets of requirements and, as a consequence, so do the metadata schemas.

### **What is recordkeeping metadata?**

At an international workshop on recordkeeping metadata held in The Netherlands in June 2000 the following definition of recordkeeping metadata was developed:

*Structured or semi-structured information which enables the creation, management and use of records through time and across domains. Recordkeeping metadata can identify, authenticate and contextualise records and the people, processes and systems that create manage and use them.*

Recordkeeping metadata helps us do the following:

- Unique identification of records;
- Authenticate records;
- Document and preserve the content, context and structure of records over time;
- Administer conditions of access and disposal;
- Track use history and recordkeeping processes;
- Facilitate interoperability;
- Restrict unauthorised use; and, most importantly, it helps
- Users find and understand records.

In other words, recordkeeping metadata is the means by which both records managers and archivists do their jobs. It can be found both in current records management systems and it can be found in archival management systems. Recordkeeping systems of whatever kind are metadata systems, because without metadata the systems simply cannot function.

The products of traditional post-hoc archival description are all metadata, but only a subset of the totality of recordkeeping metadata.

The current focus on recordkeeping metadata is helping to facilitate a trend towards convergence between the traditionally separate occupations of records managers and archivists by offering the possibility of integrating the previously separate intellectual control regimes for records management and archives.

This is in line with the recognition that, in the era of electronic records, archivists cannot continue to divorce themselves from the processes of records creation and recordkeeping system design. Put bluntly, unless electronic records are created and managed properly in well-designed systems that can guarantee the authenticity, reliability, durability, useability and accessibility of those records, archivists are not going to have many records that they can preserve for long-term use or that will be worth preserving for long-term use.

Moreover, the emergence of electronic systems provides us with the opportunity to create and capture metadata /description at the time of records creation that can be re-used for archival control purposes, whenever this becomes desirable. Although there will always be a place for the creation of value-added descriptions by archivists, much of the hard work of archival description can be reduced by recycling or reusing metadata that has already been created. For this to be possible we need a single seamless framework of interlocking standards for recordkeeping metadata and archival description.<sup>(16)</sup>

### **Standards for electronic recordkeeping metadata**

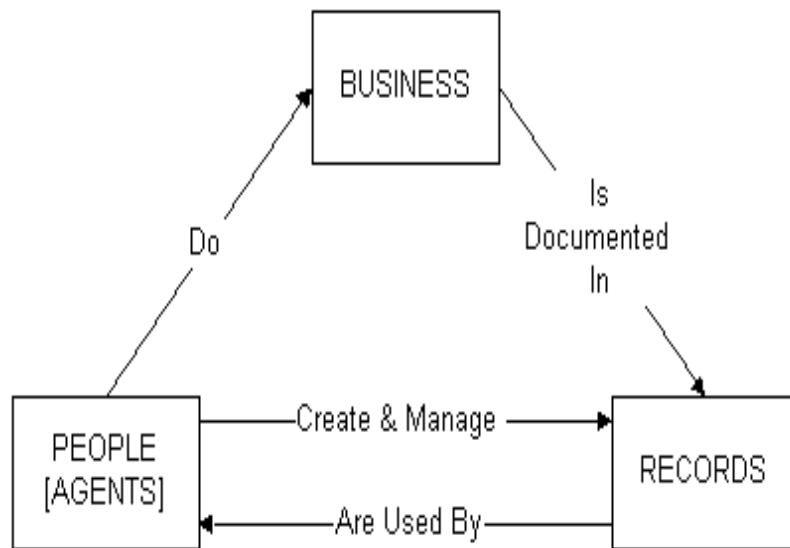
In recent years a number of initiatives have endeavoured to define criteria for the development of standards for recordkeeping metadata. At the University of Pittsburgh a research team led by David Bearman and Richard Cox has developed a set of 'Functional Requirements for Evidence in Electronic Recordkeeping' which is based on a concept of records as 'metadata encapsulated objects' and a set of metadata specifications for good recordkeeping.<sup>(17)</sup>

Another research project, led by Luciana Duranti at the University of British Columbia, has produced metadata templates for the 'protection of the integrity of electronic records'<sup>(18)</sup>, while the United States Department of Defense has issued a *Design Criteria Standard for Electronic Records Management Software Applications*, which incorporates metadata specifications.<sup>(19)</sup>

In Australia a cooperative research project led by Sue McKemmish at Monash University has developed a 'Recordkeeping metadata framework for managing and accessing information resources in networked environments over time for government, social and cultural purposes'. This project builds on the findings of the earlier projects mentioned above to produce a holistic framework for standardising recordkeeping metadata. Conceptually, the framework is grounded in Australian 'records continuum' and post-custodial' theory. It is also

firmly grounded in the same conceptual thinking that informs the Australian 'series' system for archival description. The Monash project framework proposes three entities about which metadata needs to be captured in recordkeeping systems for any 'recordkeeping event' or transaction that leads to the creation and capture of a record. The relationships between the three entities, people business and records, is illustrated in the following diagram<sup>(20)</sup>:

**Figure 1: The Business**



Having developed and promoted its recordkeeping metadata framework, the next step is to use that framework as the basis for developing a formal national standard for recordkeeping metadata under the auspices of Australia's national standards organisations, Standards Australia. Standards Australia has indicated that it is prepared to support this initiative and work is about to get under way on the project.

The National Archives of Australia was an active industry participant in this Monash University Research Project. Concurrent with the development of the Monash framework the National Archives developed and published a *Recordkeeping Metadata Standard for Commonwealth Agencies* Version 1.0.<sup>(21)</sup> This standard is consistent with, but more specific than the Monash framework. The aim of the National Archives standard is to define the records entity metadata that government agencies should capture in their agency recordkeeping systems. In other words, it aims to define that segment of the recordkeeping metadata universe that government agencies need to deal with when managing their current records. In order to facilitate access to current records via the World Wide Web, the National Archives standard is also consistent with, but greatly extends the Dublin Core metadata standard for online resource discovery.

The National Archives recordkeeping metadata standard consists of 20 descriptive elements, eight of which are mandatory, and a further 65 'sub-elements' or qualifiers that add richness and complexity to the 20 main elements. The intention is that selected metadata captured in agency recordkeeping systems can be imported into the National Archives archival control system (an implementation of the series system) whenever it is deemed to be an appropriate time to do so. At this time the agency-generated metadata can be supplemented by additional contextual metadata created by Archives staff in the course of their normal intellectual control/descriptive procedures.

As can be seen from these developments in Australia, the emphasis in the various recordkeeping metadata initiatives has been on integrating the systems for generating and managing metadata in order to make metadata creation and use/reuse as automated as possible. The aim here is to achieve records that are as 'self-documenting' as possible through careful systems design and implementation. A feature of these efforts has been a shift away from the traditional 'top-down' archival approach of describing aggregations of records to an approach that places greater emphasis on item-level control and description with links to virtual aggregations being created by virtue of the documentation of contextual relationships.

Having got this far, however, we feel that we are only scratching the surface of the possibilities and potential in this area. There remain many unanswered questions and potential research projects to pursue. For example, our understanding of the means by which contextual relationships can be identified and documented is still fairly rudimentary. An outstanding problem is that, while these recordkeeping metadata frameworks can be implemented in regulatable recordkeeping environments such as government agencies, there are a lot of unregulated recordkeeping environments that will prove more difficult to influence.

## **Conclusion**

The main purpose of this paper has been to highlight that there are other more dynamic archival strategies for managing and achieving intellectual control over records. If you want to do more than catalogue objects, if you want to capture, manage and provide access to reliable evidence of activities, then you might want to pursue some of the strategies I have outlined. It is my contention that the emerging recordkeeping metadata consensus is providing an 'event aware' as opposed to 'object oriented' approach to describing and managing records in context and over time.

Archivists used to see themselves as people who collected and provided access to artefacts – the documentary residue of society. Archivists in the future will continue to do those things but will also be much more actively involved in the processes of creating and using records throughout the entire records continuum. In the words of my Australian colleague Barbara Reed:

Records are not passive objects to be described retrospectively. Rather, they are agents of action, active participants in business activity that can only be described through a series of parallel and iterative processes.<sup>(22)</sup>

### Notes

(1) International Council on Archives, *International Standard for Archival Description (General)*, 2nd ed., Ottawa, 2000.

(2) Sue McKemmish, et al, 'Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema', *Archivaria*, no. 48, Fall 1999, p. 8.

(3) International Council of Archives, *op. cit.*

(4) Standards Australia, *Records Management*, AS 4390, Homebush, 1996.

(5) International Council of Archives, *op. cit.*

(6) Stephen Hensen, *Archives, Personal Papers and Manuscripts: a cataloguing manual for archival repositories, historical societies and manuscript libraries*, 2nd ed., Chicago, Society of American Archivists, 1989.

(7) Michael Cook and Margaret Proctor, *Manual of Archival Description*, 2nd ed., Aldershot, Society of Archivists, 1989.

(8) Bureau of Canadian Archivists, *Rules for Archival Description*, Ottawa, 1992.

(9) *EAD: Encoded Archival Description Application Guidelines Version 1.0*, prepared by the Encoded Archival Description Working Group of the Society of American Archivists, Chicago, 1999.

(10) Terry Cook, 'What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift', *Archivaria*, No. 43, Spring 1997, pp. 38-39.

(11) Adrian Cunningham, 'Dynamic descriptions: Australian strategies for the intellectual control of records and recordkeeping systems', in P.J. Horsman, F.C.J. Ketelaar and T.H. Horsman (eds), *Naar een nieuw paradigma in de archivistiek*, Gravenhage, 1999, pp. 23-44; P.J. Scott, 'The Record Group Concept: A Case for Abandonment', *American Archivist*, vol. 29, 1966, pp. 493-504; Mark Wagland and Russell Kelly, 'The Series System – A Revolution in Archival Control', in Sue McKemmish and Michael Piggott (eds), *The Records Continuum: Ian Maclean and Australian Archives first fifty years*, Clayton, Victoria, Ancora Press, 1994, pp. 131-149; and Chris Hurley, 'The Australian ('Series') System: An Exposition', in Sue McKemmish and Michael Piggott (eds), *The Records Continuum: Ian Maclean and Australian Archives first fifty years*, Clayton, Victoria, Ancora Press, 1994, pp. 150-172.

(12) International Council on Archives, *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, Ottawa, 1996.

(13) See the Web site for the Dublin Core Metadata Initiative at [http://www.purl.oclc.org/metadata/dublin\\_core/](http://www.purl.oclc.org/metadata/dublin_core/)

(14) Brian Lavoie, 'Meeting the challenges of digital preservation: The OAIS reference model', *OCLC Newsletter*, Jan/Feb 2000, pp. 26-30.

(15) David Bearman, et al, 'A Common Model to Support Interoperable Metadata: Progress Report on Reconciling Metadata Requirements from the Dublin Core and INDECS/DOOI Communities', *D-Lib Magazine*, Vol. 5, No. 1, Jan. 1999. Available at: <http://www.dlib.org/dlib/january99/bearman/01bearman.html>

(16) This is not to argue that archivists have never before recycled or reused metadata from records management systems – there is in fact a long tradition of this, especially in Australia. The problem, however, has been that the

separate standards and separate systems have made this more difficult than it should have been. With the emergence of new standards and new electronic systems we now have the opportunity to make this process as easy and as automatic as possible.

(17) The Web site of the 'Pittsburgh Project' can be found at: <http://www.sis.pitt.edu/~nhprc/prog1.html>

(18) The UBC Project and its outcomes are described at: <http://www.slais.ubc.ca/users/duranti/>

(19) The U.S. Defense Department specifications can be found at: <http://jitc-emh.army.mil/recmgt/dod50152.doc>

(20) McKemmish, et al, *op. cit.*; See also the project Web site at <http://www.sims.monash.edu.au/research/rcrq/>

(21) National Archives of Australia, *Recordkeeping Metadata Metadata Standard for Commonwealth Agencies* Version 1.0, May 1999. Available at: <http://www.naa.gov.au/recordkeeping.control/rkms/summary.htm>

(22) Barbara Reed, 'Metadata: Core Record or Core Business?', *Archives and Manuscripts*, Vol. 25, No. 2, Nov. 1997, pp. 218-241.

### **To quote this document**

CUNNIUGHAM, Adrian. <i>Recent developments in standards for archival description and metadata</i> . March 2001 [Online] Available in: <a href="http://www.enj.org">www.enj.org</a> . [Date of Access]
---

Taken of: <http://www.archivists.org.au/cds/cunningham.html>