



Author	McKemmish, Sue; Cunningham, Adrian; Parer, Dagmar
Year	1998
Title	Metadata Mania: Use of Metadata for Electronic Recordkeeping and Online Resource Discovery
Publication Details	<p>Place, Interface and Cyberspace: Archives at the Edge, Proceedings of the 1998 Conference of the Australian Society of Archivists, Fremantle 6-8 August 1998.</p> <p>This version made available on Records Continuum Research Group Website</p> <ul style="list-style-type: none">• From 1998 to 2009 at http://www.sims.monash.edu.au/research/rcrg/publications/recordkeepingmetadata-sm01.html• From 2009 to 2015 at http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordkeepingmetadata-sm01.html

Copyright

This publication is protected by copyright. Copyright in the publication remains with the author. The Monash University Research Repository has a non-exclusive licence to publish and communicate this publication online.

[Monash University](#) > [InfoTech](#) > [Research](#) > [Groups](#) > [Rcrg](#) > [Publications](#)

Metadata Mania

Sue McKemmish, Adrian Cunningham and Dagmar Parer

INTRODUCTION

The title of this paper is the result of a hurried attempt to supply the conference organisers with a catchy title - hence the reliance on alliteration. A more considered and meaningful title might have been 'Metadata Demystified' or, if we wished to retain the alliteration, 'Metadata for the Masses'.

Over the next ninety minutes it is our intention to help you make sense of metadata. Along the way we shall explain the various types of metadata you are likely to encounter, untangle a complex array of acronyms and technical mumbo jumbo, explore the inter-relationships and interdependencies of a number of established and emerging metadata schemas and, most importantly, demonstrate what metadata means to us as recordkeeping professionals. Our analysis, which will be firmly grounded within the Australian recordkeeping tradition and recent developments in continuum-based thinking, will commence with a general overview of what metadata is and what it can do. Examples of various metadata schemas will be briefly examined in this opening section. From there we will look in more detail at two current Australian metadata initiatives, the Monash University-led SPIRT (Strategic Partnership with Industry – Research and Training) Research Project on Recordkeeping Metadata Standards and the National Archives-led Australian Government Locator Service (AGLS) initiative. We shall conclude with an exploration of the implications of these various metadata developments for archival descriptive practices. In particular, our conclusions will focus on the potential that now exists for the development of a national standard for the intellectual control of records which can be a happy marriage of traditional descriptive practices, continuum-based thinking and the very latest developments in metadata. At the end we intend to leave plenty of time for questions, discussion and audience interaction.

WHAT IS METADATA AND WHAT CAN IT DO?

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. For many librarians the emerging 'Dublin Core' standard *is* metadata.

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' was as precise a definition as anyone has so far been able to produce to explain the term. The imprecision of this definition has since allowed it to be applied to any computer-related descriptive information. Indeed, the deployment 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 term first started to appear in the archival literature about six years ago when people such as David Bearman began to urge archivists to use the self-documenting (or metadata generating) capabilities of electronic recordkeeping systems to both reduce their manual description backlogs and to help guarantee the recordness of electronic records.¹ Nevertheless, it seems to us that until recently many archivists concluded that metadata and emerging metadata-related initiatives in the broader information community were not their concern.

They could not have been more wrong.

If we think of metadata in its broadest and most flexible form, 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. Metadata is simply a new term for stuff that has been around for a very long time. With the advent of computers this stuff has been given a new name and very often looks a bit different. Traditional 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. Recordkeeping metadata is any type of data that helps us manage records and make sense of their data content. Indeed, Chris Hurley has argued that metadata which documents recordkeeping processes and contextual knowledge can be thought of as part of the record itself.²

But we are getting ahead of ourselves.

Recordkeeping metadata is but one of many types or uses of metadata. Other types of metadata include systems operating metadata (that is, the metadata that is necessary to make sense of a computer software platform), data management metadata, information management metadata and discovery and retrieval metadata (that is, data that assists in the location and delivery of data – discovery metadata enables you to learn about the existence of a resource, while retrieval metadata enables you to get your hands on it).

These categories of metadata are NOT mutually exclusive. Particular metadata schemas (or sets) and the sub-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.

The following examples of metadata standards and schemas all support discovery and/or retrieval by the general public as one of their principal objectives:

- MARC - MACHine Readable Cataloguing (standard format for the computerised exchange of library cataloguing records);
- AACR2 - Anglo-American Cataloguing Rules, 2nd Edition (international standard governing library cataloguing practices); ³
- DC - Dublin Core (a set of 15 core optional elements for the description of document-like information objects (DIOs) in networked electronic environments such as the World Wide Web - the schema has been developed over a series of international meetings convened by the US-based library automation consortium OCLC, The Online Computer Library Centre); ⁴
- GILS - Global Information Locator Service (originally developed under the name Government Information Locator Service with the intention of supporting the location of US Government resources via electronic networks such as the Internet); ⁵
- ISAD (G) - International Standard for Archival Description (General); ⁶
- MAD - UK Manual of Archival Description; ⁷
- RAD - Canadian Rules for Archival Description; ⁸
- APPM - Archives, Personal Papers and Manuscripts (US manual for cataloguing archival materials using the MARC standard); ⁹
- EAD - Encoded Archival Description (a standard means of marking up archival finding aids using SGML (Standard Generalised Mark-up Language) - originally developed by the Berkeley Finding Aids Project at the University of California, but since adopted by the Society of American Archivists and the Research Libraries Group); ¹⁰
- The Australia Series System; ¹¹ and
- AGLS - Australian Government Locator Service, more of which later.

Examples of recordkeeping metadata schemas include:

- Registry Systems;
- Series System;
- University of British Columbia templates for electronic records; ¹²
- University of Pittsburgh metadata specifications for evidence in electronic recordkeeping; ¹³
- ISAD (G), MAD, RAD, EAD, APPM;
- National Archives of Australia Recordkeeping Metadata Standard for Commonwealth Agencies (a draft standard that is still under development); and
- The standardisation framework that will emerge from the SPIRT Recordkeeping Metadata Project, more of which later.

Note well that some schemas fall into BOTH categories. Indeed, ALL recordkeeping metadata schemas support discovery and retrieval.

In terms of Frank Upward's four-dimensional records continuum model¹⁴, recordkeeping metadata schemas can be further sub-divided into:

- those that are designed to operate primarily in the first three dimensions but which can play a valuable role in the fourth or 'pluralising' dimension (The Registry System, the UBC templates and Pittsburgh specifications, and the draft National Archives of Australia Recordkeeping Metadata Standard);
- those that serve the purposes of traditional archival control and description and, as such sit within the fourth dimension (ISAD (G), MAD, RAD, APPM, EAD); and
- those that are intentionally designed to encompass the entire records continuum (the Series System, the outcomes of the SPIRT Project).

With the notable exception of the series system, archivists' efforts to date have been primarily concerned with the archival control and description of non-current records for discovery and retrieval purposes. The advent of electronic records and the growth of continuum-based thinking have highlighted the necessity for integrated regimes for metadata management through the entire records continuum. The old division between current and historical records that seemed convenient in the paper-based world is now not only inconvenient, it is an obstacle to durable recordkeeping. The advent of electronic recordkeeping systems and cyberspace demands new ways of achieving old objectives. Fortunately in Australia we are better positioned than most to implement such an integrated approach to recordkeeping metadata. This is because we have just such a model in place in most of our government archives and in many of our private organisational archives - namely the series system.

WHAT IS RECORDKEEPING METADATA?

The way in which recordkeeping metadata is defined in the SPIRT Recordkeeping Metadata Project has recently been explored in an article by Sue McKemmish and Dagmar Parer.¹⁵ It is related to evolving understandings of "description" in records continuum thinking. The term description is used in records continuum thinking to label a very broad concept. This concept is not set up as an alternative to traditional definitions of archival description. It does not focus on either the "front end" or the "back end" of the records life cycle. Rather it encompasses and extends traditional definitions with reference to the whole of the records continuum.

Description in the Records Continuum¹⁶

The concept of description in the continuum relates to a complex multi-layered recordkeeping function that is carried out through a series of parallel and iterative processes that capture and manage "recordkeeping metadata".

Recordkeeping metadata is also defined broadly to include all standardised information that identifies, authenticates, describes, manages and makes accessible documents created in the context of social and business activity. Recordkeeping metadata so defined has traditionally been captured and managed in both recordkeeping systems and archival control systems.

Description-related processes begin at or before records creation and continue throughout the lifespan of the records. Their primary aim is to provide the intellectual controls that enable reliable, authentic, meaningful and accessible records to be carried forward through time within and beyond organisational boundaries for as long as they are needed for the multiple purposes they serve.

This definition can be contrasted with the narrower traditional view of description as provided in *Keeping Archives*:

Description is the process of recording standardised information about the arrangement, contents and formats of the records [in archival custody] so that persons reading the descriptions will be able to determine whether or not the records are relevant to their research.¹⁷

The way description is conceptualised within continuum thinking enables exploration of the relationship between:

- registration, classification and other recordkeeping processes in current recordkeeping systems; and
- the documentation of recordkeeping systems and their contexts of creation and use in archival control systems

both historically and in our cyberspace future. It opens up questions about where and when recordkeeping

metadata might be captured and managed in electronic systems in distributed networks.

To give an idea of the scope of this definition of recordkeeping metadata, let's look at an historical case study developed by Sue McKemmish.

Professor Neumayer was a Bavarian scientist who visited Victoria in the 1850s. His business dealings with the colonial government relating to the building of the Melbourne Observatory are documented in the inwards and outwards correspondence of the Chief Secretary. The recordkeeping metadata linked to these letters is detailed below.

Metadata Associated with Documentation of Professor Neumayer's Dealings With the Colonial Government of Victoria Relating to the Melbourne Observatory¹⁸

- Chief Secretary's registers and indexes: provide registration, classification, content and context metadata, recordkeeping process metadata, use history, and retrieval metadata
- Inwards letters and outwards letter books: annotations on letters and dockets constitute metadata about immediate business context, specifically the business processes in which the documents were agents of action
- Physical form, ordering and juxtaposition: constitute structure metadata
- PROV (Public Record Office Victoria) series system: provides broader contextual and archiving process metadata, including metadata relating to contexts of creation and use, recordkeeping systems, and relationships between and amongst context and records entities, and metadata about access terms and conditions, conservation action and disposal status, and physical location
- PROV Summary Guide: includes metadata drawn from the series system and discovery/retrieval metadata

The article by McKemmish and Parer referenced above makes the following observations about this case study:

1. Traditionally some kinds of metadata, eg relating to records content, structure and aspects of their immediate business context, management and use, have been captured and managed in current recordkeeping systems. Other kinds of metadata, eg information about the broader contexts of recordkeeping and archival processes, have been captured and managed in archival control systems. Some metadata has been present in the physical form, ordering, juxtaposition and location of the records themselves.
2. The Chief Secretary's 19th century docketing system captured and managed metadata about records content and structure, and some context and recordkeeping process metadata. The associated registers and indexes ("control records") captured and managed more extensive metadata about business and recordkeeping processes, and the use of the record. The archival control system at the Public Record Office Victoria has captured and managed descriptive metadata about the Chief Secretary's recordkeeping system, its provenance and relationships to other records, as well as metadata about archival actions relating to the records in the system.
3. Much content and structure metadata in paper systems like the Chief Secretary's is captured and represented in the physical form of the documents themselves. Some context metadata is also captured and represented physically, eg by the physical placement of an inwards letter in a docket, the attachment of two pieces of related correspondence together, the physical ordering of folios, or the physical location of a records series in a registry or in the archives. In these systems physical location and custody carry contextual meaning. The associations thus made, eg between the documents that make up a record, between records of related transactions, or between records and their creator, reflect what today we would call logical associations. ¹⁹
4. In a paper world, as Chris Hurley has explored for us, a lot of broad contextual metadata is carried in the minds of users while the records remain in the organisation that created them. Like the records, users are located inside an organisation – the users know where they are and that defines the broader organisational context of the records for them. The contextual knowledge brought to the record by "insiders" includes information about organisational and functional provenance, the recordkeeping system itself and relationships between records. Physical ordering and location in a paper paradigm have also been partial evidence of the business process and its organisational context. Moreover requirements for the unique identification of records need only be satisfied within the local domain in which they are created.
5. When paper records move beyond the boundaries of the organisation or local domain in which they are created, then broader contextual metadata needs to be captured and the requirement for unique identification needs to be extended to satisfy the demands of a global domain. Typically such needs have come into play in the past when records are transferred to an archives repository (a global domain). If these

needs are not met, "outsiders" will not be able to uniquely identify, retrieve and understand the meanings of the records. ²⁰

6. The Australian series system has always had the capacity to document the broader contexts of recordkeeping both contemporaneously and historically. And the series system is able to deal with the intellectual control and management of records that will never be in the physical custody of the National Archives:

Scott's approach was to move away from describing records in the custody of an archival institution and arranged there into a single group for a single records creator, and to move towards describing the multiple interrelationships between numerous creators, and numerous series of records, wherever they may be: in the office(s) of creation, in the office of current control, or in the archives ... Scott's fundamental insight broke through not just the straight-jacket of the record group, but all the 'physicality' of archives upon which the record group and so many other approaches to archives are implicitly based. In this way, as is finally being acknowledged, Peter Scott is the founder of the post-custodial revolution in world archival thinking. Although he worked in a paper world, his insights are now especially relevant for archivists facing electronic records, where – just as in Scott's system – the physicality of the record has no importance compared to its multi-relational contexts of creation and contemporary use. ²¹

7. In cyberspace physical location may cease to carry meaning; physical boundaries break down; the distinction between insiders and outsiders based on physical location becomes less significant in relation to using records. Records in electronic networks may be managed from their creation in global rather than local domains. In electronic systems, in particular in distributed networks, it may be essential for much of the metadata that has been traditionally captured in archival control systems to be present in – or available to – current recordkeeping systems.
8. In order to uniquely identify, manage, retrieve and understand the meaning of records in the global domains of cyberspace, it becomes essential to:
- make what was before at least partially evident through physical formatting, ordering and location (custody) explicit in metadata captured in current recordkeeping systems, or knowledge bases linked to them
 - document fully the logical associations that derive from the role records play in business processes and their contexts
 - consider capturing and managing the broader contextual metadata, traditionally found in archival control systems, in current recordkeeping systems – or devising ways of linking more closely the metadata in archival control systems to current recordkeeping systems.

The SPIRT Recordkeeping Metadata Project has identified the following purposes of recordkeeping metadata.

Recordkeeping Metadata Purposes

- **Unique identification:** eg unique registration numbers, perhaps in combination with standardised titles or classification references
- **Authentication of records:** might include signatures, seals, authorisation stamps, folio numbering, annotations re registration, classification, copying or transmission processes, cross references between documents and entries in registers
- **Persistence of records content, structure and context** (involves fixing their content, ensuring that their structure can be re-presented, and maintaining sufficient context to preserve their meaning over time and beyond their context of creation): content metadata might include precis or index entries; structure metadata might be represented in formatting, physical placement, and ordering; context metadata might include the elements represented on the Transactionality and Identity Axes of the Records Continuum Model; or be represented in physical location and custody
- **Administering terms and conditions of access and disposal:** eg entries in movement register, access conditions or disposal status, disposal action, classification or declassification action
- **Tracking and documenting use history:** eg metadata re recordkeeping and archiving processes (I was viewed, copied, indexed, classified, disposed of by Jane on dd/mm/yy; a version of me that suppressed the names of intelligence officers was viewed by John on ...)
- **Enabling discovery, retrieval and delivery for authorised users:** eg index entries, classification numbers
- **Restricting unauthorised use:** eg security classifications
- **Enabling interoperability:** eg standardised metadata drawn from classification schemes or thesauri

Recordkeepers, records managers and archivists have always managed metadata for the recordkeeping purposes identified in the first seven dot points. However, they are only beginning to come to terms with the need in distributed networked environments to assure interoperability – so that records can be identifiable, searchable, retrievable, useable, available and restrictable through common user interfaces. It is this imperative that drives efforts to standardise recordkeeping metadata.

THE STRATEGIC PARTNERSHIP WITH INDUSTRY – RESEARCH & TRAINING (SPIRT) RECORDKEEPING METADATA PROJECT

The 1998 Strategic Partnership with Industry – Research & Training (SPIRT) Support Grant, 'Recordkeeping metadata standards for managing and accessing information resources in networked environments over time for government, social and cultural purposes', aims to provide a framework for standardising sets of recordkeeping metadata that can be attributed to records from their point of creation. The Project is jointly funded by the Australian Research Council and the industry partners (National Archives of Australia, Archives Authority of NSW, Queensland State Archives, Records Management Association of Australia, and the Australian Council of Archives).

The Project Team includes:

- Sue McKemish, Chief Investigator, Monash University
- Ann Pederson, Chief Investigator, University of New South Wales
- Steve Stuckey, Partner Chief Investigator, National Archives of Australia
- David Roberts, Archives Authority of NSW
- Lee McGregor, Queensland State Archives
- Dennis Wheeler, Records Management Association of Australia
- Gavan McCarthy, Associate Investigator, Australian Council of Archives and Australian Science Archives Project
- Glenda Acland, Research Consultant
- Luisa Moscato, Researcher, National Archives of Australia
- Kate Cumming, Australian Postgraduate Award (Industry) holder.

Consultation and communication strategies are currently being put in place, including:

- a Web site
- a network of stakeholders comprising
 - experts in records and archives and related areas
 - major clients and users of recordkeeping and archival services Australian and international researchers
 - software developers and vendors
 - the Australian and international recordkeeping community
 - the wider information and metadata community
- processes for consultation and validation.

The objectives and methodology of the SPIRT Recordkeeping Metadata Project are outlined in the following table.

[22](#)

SPIRT Project Objectives

- to codify, ie specify and standardise, the full range of recordkeeping metadata needed to manage records in electronic networked environments to meet current and future requirements for access to essential evidence
- to classify metadata elements according to their role in managing records to support decision making about what metadata to capture, and to assist in managing related risks (i.e. to enable people to make business cases about what level of functionality to build into their recordkeeping systems based on considerations like
 - how robust does this record need to be?
 - does it have to persist over long periods of time?
 - how sensitive are related terms and conditions re access and use?
 - how important is it to track and document its use?)
- to support interoperability with generic metadata standards, eg the Dublin Core and other sector-specific sets
- to support initiatives in relation to information locator systems, eg the Australian Government

Locator Service

Project Methodology

- Define functionality required with reference to national and international projects, e.g. Chris Hurley's work on the *Australian Common Practice Manual*, and the University of British Columbia and University of Pittsburgh projects²³
- Specify metadata sets captured in or associated with records in Australian recordkeeping and archival systems, eg registry systems, automated records and archival management systems, Australian series systems
- Analyse specifications to determine what functionality different types of metadata supports in relation to the dot points referenced above:
 - identification
 - authentication
 - persistence
 - administration of access and disposal terms and conditions
 - documentation of use history/RK processes
 - discovery, retrieval and delivery
 - restrictability
 - interoperability
- Identify matching elements, redundancies and gaps
- Specify additional metadata
- Develop standardised Australian recordkeeping metadata set with guidelines for use, including the use of schemes, e.g. the Keyword AAA Thesaurus, qualifiers, extensions and syntax
- Validate and develop strategies for promulgating recordkeeping metadata standards (possibly through the Standards Australia framework)
- Classify standardised set in terms of functionality
- Map set against selected generic and sector specific sets, eg AGLS, Dublin Core, archival descriptive standards

THE CONCEPTUAL BASIS OF AUSTRALIAN RECORDKEEPING METADATA RESEARCH IN RECORDS CONTINUUM THINKING

The frame of reference for Australian recordkeeping metadata research is records continuum thinking and practice as it has evolved in Australia over the last half century.

One of the keys to understanding the SPIRT Project's approach to what metadata needs to be captured, persistently linked to documentation of social and business activity, and managed through time and space, lies in the continuum view of records. In continuum thinking, they are seen not as "passive objects to be described retrospectively", but as agents of action, "active participants in business processes and technologies".²⁴ This way of envisaging records has implications for the wider information world of cyberspace:

Much of the initial thinking about documents on the Internet involved a translation of the paper paradigm. Paper minds see records and other information objects as passive things to be acted upon, rather than as active participants in business processes. In the networked environment and the newly emerging information paradigms, the document-like information object can itself become the agent of action. A simplistic passive notion of DIOs which sees them as existing only to provide and disseminate information will not further the requirements of organisations, government and individuals for information objects which can act as the transactors of business. The recordkeeping perspective links the dynamic world of business activity to the passive world of information resource.²⁵

Another key to the approach being taken in the project is found in the way description is conceptualised in continuum thinking, as discussed earlier in this paper, and further explored by McKemmish and Parer in the forthcoming issue of *Archives and Manuscripts* (May 1998).

If archival description is defined as the post-transfer process of establishing intellectual control over archival holdings by preparing descriptions of the records, then those descriptions essentially function as cataloguing records, surrogates whose primary purpose is to help researchers to find relevant records. In the continuum, archival description is instead envisaged as part of a complex series of recordkeeping processes, involving the attribution of authoritative metadata from the time of records creation. Such a view of archival description is radically different from that which informs most international initiatives to standardise archival descriptive metadata, just as the Australian series system represents a very different approach to the intellectual control of records than archival descriptive systems in other countries.

THE AUSTRALIAN GOVERNMENT LOCATOR SERVICE

Information locator systems provide knowledge structures for representing, identifying, locating and delivering information resources, including records.

The National Archives of Australia has been designated lead agency for the development and deployment of the Australian Government Locator Service (AGLS), an outcome of the work of the Information Management Steering Committee (IMSC) of the Office of Government Information Technology. This committee has proposed frameworks for government information policy and the deployment of technology into the 21st century. [26](#)

The objective of AGLS is to improve the visibility, accessibility and interoperability of government information and services through the provision of standardised Web-based resource descriptions which enable users to locate the information or service that they require. Deployment of AGLS metadata will enable Web-based search engines to do their job more efficiently. This in turn will help ensure that those who conduct Web searches are presented with meaningful and relevant results sets. The development of the AGLS represents a recognition by governments that, while the World Wide Web is a major means of communication and interaction with citizens, it is nevertheless a huge and chaotic information space which requires improved means of organisation and user assistance. As such, AGLS will be a key enabler which will encourage individuals and organisations to transact business electronically with government agencies at all three levels.

Related projects which are similarly aimed at making government resources accessible online include the development of a cross-jurisdictional Internet gateway to the services and information of Australian governments (Commonwealth, State and Local), and the identification of commercial search engines which have the ability to exploit AGLS metadata efficiently. [27](#)

At the time of the IMSC it was thought that an Australian Government Locator Service would be a variant of the U.S. Government Information Locator Service (GILS). Consequently, for much of its gestation period what is now known as AGLS was referred to as AUSGILS. However, late last year when a workshop of experts convened to develop the AUSGILS standard it was decided to abandon the GILS framework and instead base the metadata standard on the Dublin Core.

This decision was made in view of deployment difficulties that have been experienced by GILS, most of which relate to the overly complicated structure of GILS and the associated expense of GILS metadata creation. [28](#) Dublin Core, on the other hand, has simplicity of metadata creation and deployment as a primary objective. It is intended that most Dublin Core metadata can be created at the time of document creation either manually by the document author or automatically by the software platform used to create or publish the document.

The Dublin Core initiative aims to establish a generic metadata set to be applied to all DIOs on the Internet. [29](#) This core set is designed to be embedded or persistently linked to individual document-like information objects. Its primary objectives relate to information resource identification, discovery and interoperability, ie improving search capability in global networks.

Another reason for basing AGLS on the Dublin Core is that Dublin Core is building up an international momentum that will probably be unstoppable. It was considered vital for any Australian Government metadata standard to be compliant with the world's most commonly accepted and deployed generic resource discovery metadata standard. In essence, the game plan of AGLS has been to customise the Dublin Core in order to fulfill the vision of GILS. This is not a particularly inventive strategy, as Dublin Core and GILS are not all that dissimilar to begin with, notwithstanding the fact that Dublin Core is much simpler than GILS.

Though intentionally minimalist, the Dublin Core set is also designed to be "extensible". This means that each of its fifteen elements can be extended by adopting specialised sets of metadata elements to provide more information. For example, the basic subject element could be extended by using Library of Congress subject headings, provided these were standardised in such a way that they were Dublin Core compliant. An associated project is the development of the Warwick Framework, a container architecture that facilitates interoperability between different sector-specific metadata sets. [30](#) AGLS fits very comfortably within the Warwick Framework.

A key part of the AGLS is the promulgation of a standard set of metadata to be attributed to all Australian government documents made accessible in distributed networks. [31](#) The AGLS set adopted the fifteen Dublin Core elements and added two additional elements, functional descriptor and availability.

AGLS Elements

(Dublin Core plus 2*)		
Title	Contributor	Source
Creator	Date	Language
Subject	Type	Relation
Description	Format	Coverage
Publisher	Identifier	Rights
Functional Descriptor*		Availability*

The functional descriptor was considered an essential element in a set that will be attributed to information resources that comprise significant quantities of records. The *Keyword AAA Thesaurus*³² (a whole-of-government administrative function based thesaurus, developed by the Archives Authority of NSW, and being customised for use by Commonwealth Government agencies) and agency Functions Thesauri can be a source of descriptor terms for the AGLS functional descriptor metadata element. The availability element was added as the purposes of the AGLS stretch beyond document discovery (the primary focus of the Dublin Core) to encompass document delivery.

A related initiative which has been proposed with the intention of taking maximum advantage of the AGLS Function element, is a National Archives proposal to compile an Australian Governments Online Functions Thesaurus with natural language searching capabilities. The need for such a thesaurus, which would be incorporated into government Web access facilities such as jurisdictional entry points, will be explored by Marian Hoy and David Roberts in another session of this Conference.³³

An overriding objective of AGLS is to institute a government resource discovery metadata regime that is simple, flexible and dynamic. Although AGLS consists of 17 elements, only 7 of these are mandatory. Utilisation of the other 10 optional elements will be a business decision of individual agencies that will be made on the basis of a cost benefit analysis of the resources being described. We envisage that more effort will be invested in metadata creation for those resources which are considered to be of particular significance.

AGLS metadata is designed to be created at the point of document creation. Importantly, however, it can also be added to and improved as documents evolve or become more significant. Another example of the flexibility of AGLS metadata is that it can be linked to single items or to aggregations of resources.

AGLS metadata can be created automatically, either by customising the self-documenting capabilities of software applications such as records management systems or Web publishing packages or by specially designed metadata generating tools such as DSTC's 'Reggie'³⁴. AGLS metadata can also, of course, be created by human beings such as document authors, Web masters or, for really high quality 'value-added' metadata, by professional knowledge representation experts.

The Australian metadata community is actively exploring how the metadata specified in standardised sets such as AGLS can be cost-effectively created, stored, interrogated by search facilities and linked to the resources described. Associating metadata with information objects can occur by: embedding the metadata within an HTML document by means of 'metatagging'; by linking objects to separate metadata stores/repositories/databases; or by encapsulating the object with metadata. The Brisbane-based Cooperative Research Centre, the Distributed Systems Technology Centre (DSTC Pty Ltd), is especially prominent in a range of metadata-related research and development initiatives.³⁵ For example, through the National Library's MetaWeb project³⁶ and other initiatives, DSTC has been experimenting with techniques for deploying distributed metadata repositories using protocols such as Z39.50 and X.500 to facilitate metadata interrogation by Web-based search engines.

A significant recent development is the release of the eXtensible Markup Language (XML) standard for Web-based resource sharing. It is anticipated that XML, which combines the power and functionality of SGML with the simplicity of HTML, will become the dominant Web markup language in the next few years. The Dublin Core Community has been quick to recognise the significance and potential of XML, with Australia's DSTC Pty Ltd playing a leading role in the development of the Resource Description Framework (RDF), an XML-based framework for the deployment of resource description metadata over the Web.³⁷ The fact that AGLS is interoperable with the Dublin Core should enable it to take advantage of RDF-related innovations. In time it should become possible to create and store AGLS metadata using a dedicated XML Document Type Definition (DTD) which is compliant with the Resource Description Framework.

At the conception of the AGLS scheme it was recognised that a high proportion of information resources described or required online to support Internet based government services and transactions would be records.

That is, in many cases AGLS metadata would be assigned to government records.

It was also recognised that the prime purpose of assigning AGLS metadata, namely enabling resource discovery and resource retrieval by authorised users, is also one of the requirements of a recordkeeping system. Hence AGLS metadata assigned to records should theoretically be a subset of any standardised metadata set specified for recordkeeping purposes.

It is envisaged that the draft government recordkeeping metadata standard that is currently being developed by the National Archives will dovetail neatly with AGLS. As you would expect, there is a considerable degree of overlap between the two. National Archives project staff are working closely together to ensure maximum harmony and interoperability between the two standards, with the intention being that once full deployment is attained the two standards will co-exist within the one seamless, dynamic and overarching government metadata regime. These National Archives staff are also active participants in the SPIRT project, thus helping to ensure conceptual unity across the three closely related initiatives.

The SPIRT project aims to specify metadata for all of the functions required of recordkeeping systems, including the discovery and retrieval function. As the AGLS project preceded SPIRT, it was not possible for AGLS to be influenced by SPIRT findings. However it is important that, as the SPIRT initiative proceeds, it assesses AGLS to ensure that the metadata specifications for functional requirements common to AGLS and SPIRT are similarly represented.

Such compatibility would ensure that at document creation, AGLS metadata could be captured as part of the recordkeeping metadata capture process. The AGLS component could then be managed within a recordkeeping system and extracted for migration to a Web-accessible environment whenever information objects are to be made visible via the Internet. At this time the recordkeeping metadata could be supplemented by value-added resource discovery metadata (such as subject headings), such is the dynamic and flexible nature of AGLS.

The AGLS initiative recognises that agencies may wish to employ technology options other than recordkeeping systems for the creation and management of AGLS metadata. Nevertheless whatever technology option is chosen, the metadata requirements for AGLS and the resource discovery and retrieval components of the metadata specification arising out of SPIRT should be similar.

So, is AGLS really worth the effort?

Our answer would be, if it is worth publishing something on the Web, then it is worth linking it to some metadata to ensure people can find it. Of course the level of effort and investment that an agency invests in AGLS metadata creation and deployment will be a business decision of that agency. These decisions will be informed by considerations such as the priority an agency places on making its services and information visible and accessible to the community.

An issue which has the potential to make or break AGLS is the cost of metadata creation. It is self-evident that as long as metadata has to be created manually by human beings, not much of it will ever be created. A major challenge therefore is the development and deployment of automated metadata generating capabilities in Web publishing software, recordkeeping systems and other document management systems. As much metadata as possible needs to be generated by self-documenting systems, rather than by human beings. Through the G7 Government Online Community Australia is playing a leading role in developing an agreed statement of requirements in this area which can then be submitted to major commercial software vendors, many of whom have already indicated their willingness to cooperate. At the local level it is anticipated that, once the SPIRT recordkeeping metadata requirements are settled, a similar dialogue can be pursued with the vendors of commercial recordkeeping software platforms.

It is envisaged that the AGLS scheme will operate in a decentralised manner and that government agencies will assign AGLS metadata at aggregate and item/object level, manage that metadata, and make it available to Web based search engines for retrieval. An AGLS User Manual has been released and is available via the National Archives home page. ³⁸ A number of AGLS pilot projects are currently evaluating the ease with which the metadata can be created, captured, managed and migrated. The pilots will also collect data on the technologies employed by agencies to implement AGLS.

So, to sum up the AGLS. AGLS is a resource discovery metadata schema for governments which supports dynamic metadata creation at and from the point of document creation. As such, it provides a vital link between recordkeeping metadata (which should soon be standardised as a result of the SPIRT Project) and the generic and increasingly ubiquitous Dublin Core standard for Web-based resource discovery. In short, while AGLS should

be largely a subset of recordkeeping metadata, it is also a superset of the Dublin Core metadata standard.

WHAT DOES ALL THIS MEAN FOR ARCHIVAL DESCRIPTION?

It should be clear by now that the objectives traditionally served by 'archival description' should be but a subset of the broader set of objectives served by recordkeeping metadata. Just as AGLS is largely a subset of recordkeeping metadata, so archival descriptions for the most part represent another subset. Remember that our overriding aim as recordkeepers is to capture, preserve and provide access over time to reliable, authentic and meaningful evidence. We do this through the design and implementation of intellectual control/metadata regimes for both active and inactive recordkeeping systems. Continuum theory tells us that there should be as much integration as possible between the metadata regimes for active recordkeeping systems and inactive recordkeeping systems.

Does this mean that we should ditch traditional archival descriptive practices and start all over again using continuum-based recordkeeping metadata as our model?

The answer we believe is ... yes ... and no.

Yes, insofar as the advent of electronic recordkeeping is revolutionising much of what we do. For the most part, traditional practices will simply not work in this new environment. A new approach is not only desirable it is unavoidable.

No, insofar as we have much that is of value in our older control systems and it would be foolish to throw the baby out with the bath water. As we have already argued, the series system provides a good basis for a continuum-based approach to implementing recordkeeping metadata regimes. Moreover, while it is clearly desirable for professional recordkeepers to implement records control systems across the whole continuum, we know that this will not always be possible. A significant proportion of records will continue to be created in less regulated recordkeeping environments. These records will only come under the control of recordkeeping professionals after they have become inactive, for example in collecting archives. Traditional archival descriptive control practices will continue to be of value in these circumstances.

Once the SPIRT Project is completed, probably the major challenge facing us in this area will be how to integrate continuum-based recordkeeping metadata with traditional archival descriptive practices. While change is essential, we cannot lose sight of the fact that there is considerable investment in traditional systems of intellectual control. Retrofitting of existing descriptive systems would not only be prohibitively expensive, it would probably be impossible in any case. The aim should be to make the linkages between the new and the old as seamless and as interoperable as possible. If we bear in mind that new-fangled electronic metadata is really not all that different to the old fashioned metadata, this task of integration should not be impossible. Nevertheless, we should not underestimate the potential difficulties we shall face.

A major difficulty will be resolving the tension between the static documentary or bibliographic representations of non-current records that are the norm in traditional archival systems and the dynamic metadata regimes which mirror the processes of records creation and use, which is the model being pursued by the SPIRT Project. Another challenge will be integrating item-level metadata with recordkeeping system-level (or collection-level) metadata. Yet another challenge will be ensuring the user-friendliness of the control systems.

It is the linking of contextual metadata at different levels of description and in different dimensions of the records continuum that will perhaps provide the most significant challenge. One of the roles that traditional post-hoc archival description played was to link high-level contextual information to the item level records descriptions in order to provide external validation and meaning to the records being described. This will still need to occur in the brave new world of recordkeeping metadata. This linking of contextual (or ambient) metadata will need to occur whenever records are used outside of the recordkeeping domain in which they were created (eg. in cyberspace or in a collecting archives).

A PROPOSED NATIONAL STANDARD FOR THE INTELLECTUAL CONTROL OF RECORDS

With all of this in mind the ASA/ACA Joint Committee on Descriptive Standards (of which all three of us are members, and Dagmar Parer is Chair) has set itself the objective of devising a national standard for the intellectual control of records. This Standard, which we envisage will be a companion Standard to the Australian Standard for Records Management (AS4390), may perhaps be a container architecture that provides an integrated and theoretically consistent framework for the best practice application of a variety of compatible and interoperable descriptive metadata schemas.

Unlike other static post-hoc descriptive standards such as ISAD (G), the new Australian Standard should encompass the entire records continuum. Drawing on the results of the SPIRT Project, it should provide a standardised means of creating and dynamically maintaining recordkeeping metadata over time, both at item and aggregate level. It should provide guidance on the linking of record descriptions to function, activity, creator and recordkeeping process metadata.

As far as possible the new standard will endeavour to be compatible/interoperable with such existing international and national standards as ISAD (G), ISAAR (CPF), MARC and AGLS. Because we wish the Standard to be one that practitioners can both relate to and aspire towards, it should endeavour to be as compatible as possible with existing best practices of records description in this country, such as the series system. The conceptual framework of Chris Hurley's Australian Common Practice Manual (ACPM) should be of inestimable value in these efforts. At the end of the day, however, our overriding concern will be to devise a standard that is fully compatible with the existing AS4390 and with the spirit of the SPIRT Project. Compromises in other areas will almost certainly be necessary to achieve these twin ends.

The Standard should also provide guidance on the creation, management and presentation of the 'high level' or ambient metadata that is needed to manage and comprehend records in domains other than that in which the records originated. As Chris Hurley has said, this high-level metadata sits outside the records and provides external validation and meaning. ³⁹

One view that has been expressed within the ASA/ACA Descriptive Standards Committee is that the combination of local recordkeeping system discovery and retrieval metadata and 'high-level' metadata could together serve the purposes usually served by traditional archival descriptions. While this is an issue we still need to think through properly, the new Standard could provide a standardised means of constructing and presenting this combination of metadata. This component of the Standard should also be able to be deployed in situations where archivists have assumed control of records without having had any influence over the manner in which the recordkeeping metadata was created and maintained by the records creators.

Diagrams 1 and 2 present two different views of the conceptual universe that will be traversed by the proposed new Standard. These diagrams have been developed by Adrian Cunningham as 'discussion catalysts' rather than definitive statements. As such, they have generated discussion, rather than agreement amongst the authors of this paper.

1. Recordkeeping Metadata Sets and Sub-Sets

(Prepared by Adrian Cunningham for the ASA/ACA Descriptive Standards Committee)

2. The Wonderful World of Records and Metadata

(Prepared by Adrian Cunningham for the ASA/ACA Descriptive Standards Committee)

THE GLOBAL RECORDKEEPING DOMAIN

(I.E. THE 4TH DIMENSION OF THE RECORDS CONTINUUM)

We anticipate that it will take us about three years to produce the new Australian Standard. We crave your support for and input into this initiative which, while probably overdue, may nevertheless end up being so much the better for the delay.

In conclusion, we shall leave you with a mudmap which represents an attempt by Adrian Cunningham to illustrate the linkages between existing and emerging metadata and descriptive standards and to position the proposed new Australian Standard for the Intellectual Control of Records within this landscape. The mudmap is far from definitive. Indeed, it is not even fully accepted by the other authors of this paper. It is, however, an attempt to try to make sense of a potentially confusing environment. As such, the mudmap may provide a useful catalyst for the audience discussion and interaction that we hope will occupy the remainder of this conference session.

Australian Descriptive Standards Mudmap

(Prepared by Adrian Cunningham for the ASA/ACA Descriptive Standards Committee)

1. David Bearman, 'Documenting Documentation', *Archivaria*, No. 34, Summer 1992, pp. 33-49. For a useful overview paper on metadata for recordkeepers see Susan Myburgh, 'Metadata and its Meaning for Records Managers' *Records and Information*

- Management Report, Vol. 14, No. 5, May 1998.
2. Chris Hurley, 'The Making and Keeping of Finding Aids: (1) What Are Finding Aids For?' *Archives and Manuscripts*, Vol. 26, No. 1, May 1998 (forthcoming).
 3. *Anglo-American Cataloguing Rules*, 2nd ed., 1988 edition.
 4. See the OCLC site for more information on the Dublin Core: <http://purl.oclc.org>
 5. Eliot Christian, 'GILS What is it? Where's it going?' *D-Lib Magazine*, December 1996 at <http://www.dlib.org/dlib/december96/christian/christian.html>
 6. ISAD(G): General International Standard Archival Description, Secretariat of the ICA Ad Hoc Commission on Descriptive Standards, Ottawa, 1994.
 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. 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.
 10. See <http://sunsite.berkeley.edu/FindingAids/EAD/history.html#1995>
 11. 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, Vic., Ancora Press in association with Australian Archives, 1994, pp. 131-149.
 12. See the Website of the University of British Columbia research project, The Preservation of the Integrity of Electronic Records at <http://www.slais.ubc.ca/users/duranti/>
 13. See the Website of the University of Pittsburgh research project, The Functional Requirements for Evidence in Electronic Recordkeeping at <http://www.sis.pitt.edu/~nhprc/>
 14. Frank Upward, 'Structuring the Records Continuum Part One: Post-custodial principles and properties' *Archives and Manuscripts*, Vol. 24, No. 2, Nov. 1996, pp. 268-285.
 15. Sue McKemmish and Dagmar Parer, 'Towards Frameworks for Recordkeeping Metadata Standards', *Archives and Manuscripts*, Vol. 26, No. 1, May 1998 (forthcoming).
 16. This definition has been developed by the Records Continuum Research Group (Principal Researchers Sue McKemmish, Frank Upward, Barbara Reed and Livia Iacovino) at Monash University and provides a key frame of reference for the SPIRT Recordkeeping Metadata Project. See <http://www.sims.monash.edu.au/rcrg/>
 17. Judith Ellis (ed.), *Keeping Archives*, 2nd edition, W D Thorpe, Port Melbourne, 1993, Chapter 8, p.223.
 18. The 1857 correspondence between Neumayer and the Chief Secretary is captured and documented in:
VPRS 1189 Inwards Correspondence of the Chief Secretary's Office, Unit 744, 1857/8840, Letters re the Observatory
VPRS 1187 Register of Letters Sent, 1857
VPRS 1186 Inwards Correspondence Registers, 1857
VPRS 1411 Index to Letters Received, 1857
VPRS 1188 Index to Letters Sent, 1855-59
Public Record Office of Victoria Series System
Public Record Office of Victoria, Summary Guide, 1990
 19. As explored further in: Sue McKemmish, 'Are Records Ever Actual?' in Sue McKemmish and Michael Piggott, *The Records Continuum: Ian Maclean and Australian Archives first fifty years*, Ancora Press in association with Australian Archives, Clayton, 1994, pp.187-203.
 20. Chris Hurley explored these issues in the Masters teaching program at Monash University, especially through the distance education subject, LAR 5530 Managing the Records Continuum, developed by Sue McKemmish and Frank Upward, with critical inputs from Chris. See also his forthcoming article referenced in Endnote 2.
 21. Terry Cook, 'Archives in the post-custodial world: interaction of archival theory and practice since the publication of the Dutch Manual in 1898', paper delivered to the *XIII International Congress on Archives, Beijing, 1996*.
 22. The Project Objectives and Methodology are detailed in the Application and Research Plan for the Strategic Partnership with Industry – Research and Training (SPIRT) Support in 1998: Recordkeeping metadata standards for managing and accessing information resources in networked environments over time for government, commerce, social and cultural purposes. See <http://www.sims.monash.edu.au/rcrg/>
 23. University of British Columbia, 'The Preservation and Integrity of Electronic Records': <http://www.slais.ubc.ca/users/duranti/>
University of Pittsburgh, 'Functional Requirements for Evidence in Recordkeeping' project, Business Acceptable Communications model: <http://www.lis.pitt.edu/~nhprc/mea96.html>
 24. For further elaboration of this view, see Barbara Reed's article, 'Metadata: core record or core business', *Archives and Manuscripts*, Vol 25, No 2, Nov 1997, pp.218-41.
 25. 1999 Large Research Grant Application, 'Metadata architecture to support persistence of essential evidence of business, social and cultural activity in distributed networked environments', February 1998 (Chief Investigators: Barbara Reed and Sue McKemmish).
 26. See Management of Government Information as a National Strategic Resource: Report of the Information Management Steering Committee on Information Management in the Commonwealth Government, Canberra, Office of Government Information Technology, 1997. Also on the Web at <http://www.ogit.gov.au/publications/IMSC/imscrept.html>.
 27. 'Functional Requirements for a Whole-of-Australian-Government Search Architecture: a Report by the Search Engine Working Group (SEWG), December 1997', available at <http://www.nla.gov.au/oz/gov/sewg/>
 28. W.E. Moen and C.R. McClure, 'An Evaluation of the Federal Government's Implementation of the Government Information Locator Service (GILS) Final Report, June 30 1997' at <http://www.unt.edu/slis/research/gilseval/gilsdoc.htm>
 29. See the OCLC site for more information on the Dublin Core and Warwick framework: <http://purl.oclc.org/metadata>
 30. Carl Lagoze, 'The Warwick Framework: a container architecture for diverse sets of metadata' *D-Lib Magazine*, July/August 1996 at <http://www.dlib.org/dlib/july96/lagoze/lagoze.html>
 31. Australian Government Locator Service Implementation Plan at <http://www.ogit.gov.au/aglsindex.html>
 32. *Keyword AAA: A Thesaurus of General Terms*, Rev. Ed., The Archives Authority of New South Wales, Sydney, 1998.
 33. D. Roberts and M. Hoy, 'Understanding Government Bureauspeak: Natural Language Searching and Government Thesauri', available via the National Archives of Australia Web site at <http://www.naa.gov.au/>
 34. Reggie can be found at <http://metadata.net>
 35. See the DSTC site (<http://www.dstc.edu.au/>) for details of these projects, e.g. next generation middleware, flexible configurable workflow prototypes, and global information access.

36. See the MetaWeb Project home page at <http://purl.nla.gov.au/metaweb/home>
37. Eric Miller, 'An Introduction to the Resource Description Framework', *D-Lib Magazine*, May 1998 (<http://www.dlib.org/dlib/may98/miller/05miller.html>)
38. The Australian Government Locator Service (AGLS) Manual for Users at <http://www.dlib.org/dlib/may98/miller/05miller.html>
39. C. Hurley, 'Standards, Standardisation, and Documentation' in *Archives at the Centre: Proceedings of the Australian Society of Archivists Conference Alice Springs 24-25 May 1996*, Canberra, ASA, 1997, p. 64.

Final version, 4 August 1998

[Back to top](#)

Copyright © 2015 [Monash University](#) ABN 12 377 614 012 – [Caution](#) – [CRICOS Provider Number: 00008C](#) [Request an IT or web service](#)
Last updated: 23 November 2012 – Maintained by [eSolutions Service Desk](#) – [Privacy](#) – [Accessibility information](#) [service](#)