TAHO
Tasmanian Archive + Heritage Office

Information Management Advice 14 The value of recordkeeping metadata

Introduction

Have you identified a requirement to keep business information for long periods of time? If that information is in digital form, maintaining it for the ten, twenty or fifty years that is required by legislation or to meet business needs can be a difficult process.[1]

Metadata is a means of helping you to meet preservation requirements for digital information. Metadata can be used to ensure that agency information assets are maintained as part of your agency's digital continuity strategy. You can use metadata to flag your most valuable information requiring management through system changes. For example, you might use metadata to identify which documents stored on shared network drives need to be transferred into your recordkeeping system, or to identify what records need to be converted to a preservation file format to ensure continued access.[2]

Background

In the past, government agencies had records registries responsible for capturing business communications in paper files. This centralised system ensured a consistent approach to registering and classifying records (i.e. adding the recordkeeping metadata). The records registry also controlled storage and access to these files, and managed the disposal of these files at the end of the required retention period, so the information contained in the files remained largely unaltered over time. At the time of disposal, you could still glean much about the use, history and value of the records just by looking at the metadata written on the file or recorded in the register.

When desktop computers and individual email accounts arrived, business communication became an individual responsibility. Tasmanian government agencies have now been working with digital information for over 20 years. Over this period, agencies have accumulated massive amounts of information in multiple and disconnected locations. This has created inefficiencies, and increased compliance and security risks. It also means that staff are not able to find all of the information they need to perform their jobs, or are at risk of using information that is redundant or out-of-date to make decisions.

In the digital age, it is metadata that provides the critical context and controls to allow information to be accessed, used and re-used. Without appropriate and consistent application of metadata, digital information loses value, and will become inaccessible in a relatively short period of time. Your agency information management program should include recognising and harnessing the value of your metadata.

---

[1] Refer to *Advice 37 – Keeping digital records accessible* for practical strategies to avoid information loss
[2] *Guideline 19 – Digital Preservation Formats*

Department of Education
LINC Tasmania

Tasmania
Explore the possibilities

The value of metadata

"Metadata, and particularly metadata about records, which detail transactions and are routinely collected for business purposes, are the stuff of big data, data mining, semantic web relationship linking and other emerging analytic technologies coming down the line. Metadata is also at the core of the social concerns over privacy, the capacity of individuals to control their personal information and the acceptable limits of data linkage."[3]

Metadata is a very flexible and powerful tool. It can automate actions and be easily created, shared and reused. If information has metadata, we know what it is, we know where it is and we know how to use it.

Examples of metadata include:

- Controlled terms chosen from a drop down menu or pick list and entered into a pre-established cell or field.
- Sets of keywords or terms used to annotate or 'tag' a web page or blog post.
- Information that is entered manually as free text descriptions.
- Information derived automatically from software.

We are all familiar with metadata in one form or another. Metadata gives raw data its wider meaning. For example, the metadata associated with an email includes: the name of the sender and receiver, the date and time the email was sent and received, and the subject. Without this essential metadata, the purpose of the email will be unclear and its value as evidence will be non-existent. Separated from metadata, digital information quickly becomes impossible to find, authenticate, view or migrate to new technology platforms.

When information is recognised as a vital agency asset, with a recognisable and manageable value, risk, content and lifecycle, then it makes sense to look for ways to improve metadata management processes.

Why recordkeeping metadata in particular?

"What would happen if we took a Recordkeeping Analytics approach, envisioning recordkeeping metadata not as an overhead and a cost, but as an information asset from which real time insight into how both the business, and the business of recordkeeping…is performing? For this we need to enable the automated capture of [metadata] in machine processable forms at transactional layers, which can then be mined, meshed and augmented with…governance frameworks, and the insights gained able to be pumped back into business and recordkeeping processes and systems."[4]

Metadata is critical to support information management processes, and is an essential element of any system that manages State records. It is metadata that enables the creation, registration, classification, access, preservation and disposal of digital records.

In some respects, digital recordkeeping metadata is similar to the descriptive information that was captured on file covers and in file movement registers in paper-based recordkeeping systems. In the digital business

---

[3] Barbara Reed, Rethinking approaches to recordkeeping metadata, Recordkeeping Roundtable, 30 June 2014:
http://rkroundtable.org/2014/06/30/rethinking-approaches-to-recordkeeping-metadata/#_edn3
[4] Dr Joanne Evans, Reflections on the Promise and Pitfalls in Reinventing Recordkeeping Metadata, Recordkeeping Roundtable, 30 June 2014, http://rkroundtable.org/2014/06/30/reflections-on-the-promise-and-pitfalls-in-reinventing-recordkeeping-metadata/

environment, however, metadata must be actively collected and managed in order to find, understand, protect and maintain access to digital records over time. It is recordkeeping metadata that provides the evidentiary characteristics of the digital record.

Recordkeeping metadata assists agencies to identify, authenticate, describe and manage digital records in a systematic and consistent way through system migrations or government administrative change. Metadata also:

- Ensures that contextual information about business transactions is captured and linked to the relevant record.
- Supports searching of records by using controlled vocabularies and linking to business functions.
- Controls access to records by nominating, usually at the time of records creation, the security or legal status of records or any other caveats on their retention or use.
- Reduces the risk of unauthorised access to records.
- Supports management of digital storage costs and helps ensure that records are not kept beyond the period of their administrative use, by identifying disposal requirements.
- Supports the identification and management of vital records.
- Aids planning for data migration and other preservation needs by identifying, in standardised and accessible ways, the software and hardware dependencies of records.
- Provides a benchmark for measuring the quality of recordkeeping within and between public authorities for auditing and other purposes.
- Provides a key framework for specifying recordkeeping functionality when designing or procuring systems.

The processes for managing recordkeeping metadata described in *Guideline 5 – Recordkeeping Metadata* are intended to support agencies to identify and specify what information their systems (digital, paper or hybrid) need to capture about records. The Guideline sets a baseline for capturing, describing, using and managing metadata about records to ensure digital continuity. Standardised recordkeeping metadata allows information management staff to:

- Identify the record
- Understand the purpose and meaning of the record
- Understand the record structure and format
- Manage access and disposal
- Track use history, including recordkeeping and archiving processes
- Restrict unauthorised use
- Enable discovery, retrieval and delivery
- Authenticate the records
- Migrate or export the records

Recordkeeping metadata that is commonly implemented in digital recordkeeping systems includes:

- Identifiers (to identify an entity in the system such as a record or a record author),
- Dates (such as date registered or date destroyed), and
- Relationships and linkages from the record to other data about the business transactions that the record relates to.

In sophisticated digital recordkeeping systems (i.e. EDRMS), this metadata is captured automatically, in machine readable form, allowing for the automation of many recordkeeping tasks. However, managing metadata doesn't require implementation of a complex new system. Small agencies with paper-based, hybrid, or digital

recordkeeping systems can identify and implement metadata at the file level and manage the recordkeeping metadata schema in a simple spread-sheet. Larger agencies may choose to implement more sophisticated metadata registry systems to manage the agency's enterprise information architecture, controlled vocabulary and metadata schema.

When implementing metadata in an EDRMS, it is particularly important to check, before selecting a system, what metadata the system will capture, and compare it to the minimum metadata set in *Guideline 5 – Recordkeeping Metadata*.

Minimum recordkeeping metadata

Standardising your metadata throughout your agency will ensure that agency information assets will continue to meet business, accountability and user needs for the long term. A standard approach is particularly valuable when implementing new business information systems or applications. This is because capture of metadata can sometimes be limited by system features and configuration decisions. The best opportunities to improve the capture and management of metadata generally occur at times of system change or review.

Metadata requirements should primarily be considered at the following times:

- When specifying requirements for, and procuring, new systems
- When configuring new systems, or changing the configuration of existing systems
- When migrating systems, to ensure that important metadata is not lost
- When reviewing the recordkeeping functionality of existing systems.

Guideline 5 - Recordkeeping Metadata can be used to ensure that your EDRMS is compliant with the minimum recordkeeping metadata to will support digital continuity of your permanent and vital records, but it should also be used when reviewing other business systems to determine if they have sufficient recordkeeping functionality.

NOTE: Not all systems will have the ability to capture recordkeeping metadata if not this was not considered essential to the system functions. This means that capture of metadata may need to occur outside the system and be linked to the system. Implementing the requirements set-out in *Guideline 5 – Recordkeeping Metadata*, and the associated mapping template may assist with this process.

Identifying and mapping metadata

You can identify the recordkeeping metadata captured within a system by examining the system, reading system documentation, or researching agency data dictionaries and data models. Staff using the system will also be able to provide you with an understanding of the types of data they create and apply to records in their daily business operations.

In collaboration with system administrators, look for links between different elements and inheritance of metadata between different levels of aggregation. Documents and records are often grouped together as aggregations to help manage recordkeeping relationships. The representation of this aggregation is referred to as a file plan (commonly a hierarchy composed of three levels). In some systems, they are the same as the view presented to end users (for example, as folder structures on network drives, or as classification structures embedded into the presentation of folders in an EDRMS store), but it is worth noting that they do not have to be visually presented to users this way.

Depending on the system implementation, aggregations can exist in metadata only, and not necessarily be presented visually to the end users.

For example, are security restrictions applied to a work group, or to a case or project or job, rather than to a folder or a record? Will all documents in a folder inherit the access or retention and disposal metadata assigned at the folder level?

Descriptive metadata

This metadata is essential for access and disposal because it helps us understand how the record was used by the agency and other stakeholders. Descriptive metadata supplies the business context around the data, such as the title or name, owner or creator, and what business transactions it relates to. This metadata can also define the record type (e.g. file, letter, email, report, spread-sheet, etc.) This is especially useful metadata for records in business systems, and for unstructured records on shared drives which, when divorced from their context, are virtually meaningless.

Metadata can also be used to describe things other than records and information, such as people, workgroups, organisations, business transactions, business activities, business functions, laws, regulations, policies and business rules. You should determine whether metadata descriptions of any of these entities would be of use to your agency.

Process metadata

Process or workflow metadata must continue to be captured over time. This is metadata about specific processes that are performed, such as changes to the records content, structure and appearance. This metadata accumulates over the entire lifespan of the record. Version numbers, dates of significant actions, metadata about the use of the record, and changes to access rights and restrictions are all examples of process metadata. Any of the actions listed below that are performed on a record must be captured as metadata:

- Registration into the recordkeeping system
- Apply or change access rules
- Transfer of control/custody
- Disposal
- Migration

Capturing process metadata such as approval dates, revisions or amendments also builds trust and accountability into digital processes. Process metadata also supports document publishing and automated records disposal.

Few business systems are designed to capture sufficient process metadata. Some case management systems may keep process metadata in audit logs, but even these systems are unlikely to capture disposal and migration metadata. However, this functionality is standard in EDRMS, which is designed to deliver the security, access, version control and audit functionality required for true information management.

If you are unable to develop technical solutions to automatically capture process metadata, another approach may be to develop:

- Policy or procedures about capturing adequate metadata which explains specifically how they achieve this, or
- Training programs to show staff how to use existing recordkeeping tools, such as disposal authorities or thesauri, to ensure that they are entering adequate metadata in the correct data fields within the systems they use.

NOTE: Often process metadata, such as metadata about disposal or migration, can be applied and maintained at an aggregated folder level rather than to individual documents or records.

Technical metadata

This metadata is essential for ongoing storage and quality assurance management purposes. If it is a database, metadata provides technical information about the structure of the data, such as the name of the source table, the source table column name, and the data type (e.g. string, integer). If it is a digital image file, metadata includes the scan resolution, bit depth, file compression, etc. If it is a digital document, metadata includes encryption, fonts, word counts, etc. Technical metadata is usually managed by specialist IT staff such as database administrators.

Other metadata standards

In tandem with the minimum recordkeeping metadata set (*Guideline 5 – Recordkeeping Metadata*), consider implementing other metadata sets. This will ensure consistent description and classification of agency information assets. This will also assist with maintaining information through system migrations and administrative change, and improve interoperability of government data. Two common standards are the Dublin Core Metadata Schema[5] and the Australian Government Locator Service (AGLS) Metadata Standard[6].

Choosing what additional metadata to capture will depend on digital format and the business activity you need to support. For example:

- For a document, it would be important to capture the author, date created, title, version control, and business purpose and security classification.
- For an image, it would be important to capture the date the image was taken, photographer, location, client for whom it was taken and the copyright terms under which it can be reused.

Where an extension of elements to your metadata schemes is required to meet client or business needs, use the metadata extension methodology defined in whatever standards you are using.

Implementing metadata

Metadata management strategies work best when a collaborative approach is adopted by [RIM and IT] staff to ensure that metadata is applied to the right records and information, that it is done in sustainable and effective ways, and that opportunities to improve metadata are identified. [7]

Good metadata management processes ensure that information is:

- Accurate and can be trusted;
- Secure from unauthorised access, alteration and deletion;
- Can be found and understood; and is
- Kept for as long as required.
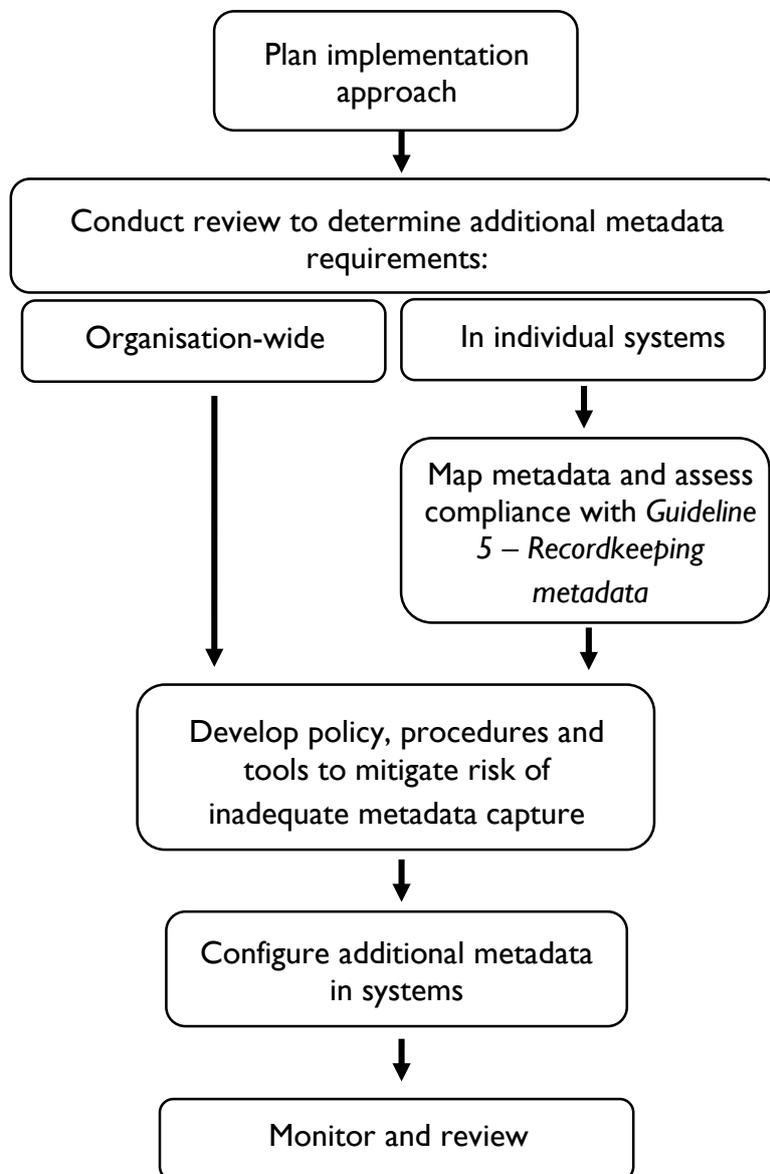
---

[5] http://dublincore.org/
[6] http://www.agls.gov.au/
[7] Metadata for digital continuity: a companion guideline to the Queensland Recordkeeping Metadata Standard, 2012: http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/MetadataDigitalContinuity.pdf

Management of metadata is an important element of your Records Management Program to ensure metadata capture, reusability, integrity and consistency. This includes implementing metadata standards and encoding schemes as key recordkeeping control tools for mitigation information risk. It is crucial that metadata is reviewed in all business systems, not just the agency EDRMS.

The diagram shows a possible implementation approach:

```
            ┌─────────────────────────┐
            │  Plan implementation    │
            │       approach          │
            └────────────┬────────────┘
                         │
                         ▼
      ┌──────────────────────────────────────┐
      │ Conduct review to determine additional│
      │        metadata requirements:         │
      └──────────────────────────────────────┘
      ┌──────────────────┐  ┌──────────────────┐
      │ Organisation-wide│  │In individual     │
      │                  │  │systems           │
      └────────┬─────────┘  └─────────┬────────┘
               │                      │
               │                      ▼
               │          ┌─────────────────────┐
               │          │Map metadata and     │
               │          │assess compliance    │
               │          │with Guideline       │
               │          │5 – Recordkeeping    │
               │          │metadata             │
               │          └──────────┬──────────┘
               │                      │
               ▼                      ▼
      ┌──────────────────────────────────────┐
      │ Develop policy, procedures and        │
      │ tools to mitigate risk of             │
      │ inadequate metadata capture           │
      └──────────────────┬───────────────────┘
                         │
                         ▼
      ┌──────────────────────────────────────┐
      │ Configure additional metadata         │
      │          in systems                   │
      └──────────────────┬───────────────────┘
                         │
                         ▼
      ┌──────────────────────────────────────┐
      │          Monitor and review           │
      └──────────────────────────────────────┘
```

Your implementation should consider prioritising systems based on both risk factors and continuous improvement. For example:

- Your initial assessment may be agency-wide and focus on addressing minimum metadata requirements, with further analysis scheduled to occur in conjunction with any planned future upgrades, or
- You conduct a review of the EDRMS first, followed by an assessment of other business systems.

Prioritising metadata

An important implementation consideration is identifying where additional metadata must be applied to enable the continued accessibility of digital information over time.

The table below lists some factors to consider in order to determine priorities for metadata management:

Factor	Explanation
Retention period	While the lifespan of a business system may be as little as 5 years, as a general rule of thumb closer attention should be paid to metadata for those systems that contain information with retention periods of 15 years or more. It is particularly critical for records of long term and permanent retention value.
Age of the information	Refers to the actual age of existing records, as distinct from their retention period. Older records, especially if they have not been well-managed in the past, may have a smaller 'window of opportunity' before access is lost or their authenticity is put at risk.
Continued accessibility	Records known to be at risk of becoming inaccessible, for example, records in rare proprietary formats or in legacy systems with little corporate knowledge of what they are and how to access and use them.
	Storage locations and environments also impact on this – offline records are at more risk of being 'forgotten about' and left in obsolete formats.
Duplication or redundancy of information	In some situations, similar information is duplicated in different records and different systems. The most unique information will be at highest risk, particularly if it is held in a legacy or siloed system, with no links to other systems. Where information is duplicated, the organisation will need to identify what is the authoritative version, and whether it is sufficiently protected.

Metadata in business systems

To capitalise on the value of your metadata, it has to be well specified and well managed, both immediately and in the long term. It should not be created on an ad hoc basis, and so:

- If your agency is developing and implementing a custom-built business system, you will need to design a comprehensive recordkeeping metadata set that supports your business needs.
- If you are implementing an off-the-shelf software package, you will need to identify and apply a metadata schema to support recordkeeping requirements for information in that system.
- If your agency takes up a service offering for a system that performs or supports business processes, you will need to work with the service provider to identify key metadata that maps to your requirements. Understanding and documenting the metadata within a service offering is important to managing the records and information through transition out of the service arrangement.

In these situations you need to make sure you have a standardised set of metadata elements that:

- Specify the particular metadata fields that will be used in the system
- Provide a definition for each of the fields, indicating what can and cannot be applied within them
- Identify the data that will be used to populate each field if metadata will be collected automatically from other systems etc.
- Identify the encoding schemes or pick lists that will be used to provide data values in the system if metadata collection cannot be automated.

TAHO has developed advice to assist you to identify, map and implement recordkeeping metadata in business systems. See *Advice 18 Managing records in business systems*.

Applying metadata appropriately

Once you have determined where the gaps in recordkeeping metadata are in your business systems, you will need to decide what level of aggregation to apply your recordkeeping metadata. For example, metadata can applied to:

- Individual documents and records (such as a title or unique identifier applied to an individual record)
- Groups of records (such as retention and disposal rules applied at the folder level in your EDRMS)
- Whole systems and information assets (such as standardised access rules applied to certain transactions in your personnel or finance database)

As well as being descriptive information about the information content of the record, metadata can provide information about the system the record was created by or captured in, the digital format, its relationship or links to other data in the system and the processes and events that have been carried out on the record.

Migrating metadata

You will need to ensure that relevant recordkeeping metadata remains persistently linked to digital records including through system migrations. Your metadata management strategy should also include consideration of version control and consider metadata maintenance if records are migrated to new systems.

Any permanent and long-term temporary records in a system will need to be managed after the system has been decommissioned. This means that if your export or migrate data from your system, you will also need to export or migrate the accompanying recordkeeping metadata.

Metadata can be embedded in a record, linked to the record or just associated with the record. If it is associated with the record (e.g. managed in a lookup list, spread-sheet or database), your metadata will require ongoing attention and management. Recordkeeping metadata will be essential to preserving and maintaining continuity of those records over time.

Disposal metadata

When records and information are destroyed in accordance with authorised retention and disposal schedules, some metadata must continue to be maintained.

In fact, it can be very difficult to separate the digital record from metadata in the system that created it. This makes carrying out disposal in the digital environment a vastly more complex process than in the paper environment.

If you are sentencing records in a controlled digital system such as an EDRMS, the process may be automated to assist agencies to save time, and avoid a processing backlog which would require additional manual resources to manage.

In some cases, it is necessary to wait until the file is closed, and inactive, to apply the retention period and correct disposal class. Case files are a good example of this, where disposal may depend on if approval was granted or declined, or some other action was completed or process activated.

Metadata is a record

It is also important to remember that metadata itself is a record. Recordkeeping metadata is a key record that attests to the integrity of digital records by documenting their context and management history. Metadata and the associated relationships (e.g. document level associated with folder metadata) must therefore be maintained be permanently or for the lifespan of the records.

Examples of the type of folder level metadata that needs to be retained includes:

- File title
- Dates of creation and closure or, if not closed, date of last item added.
- Disposal actions including the retention and disposal schedule number, disposal class number, disposal action and disposal trigger.
- Date of destruction or other disposal (for example, transfer to different agency or to TAHO).

In an EDRMS environment, individual document metadata that it may be important to retain includes:

- Title/name
- Author (individual or organisation)
- Addressee, as appropriate
- Date of creation, date sent or last action

Recordkeeping metadata needs to be retained in accordance with the Archives Act 1983 and all relevant disposal classes within approved retention and disposal schedules. When your permanent records are transferred to TAHO, you will be required to provide sufficient accompanying metadata for TAHO to verify each record's integrity, authenticity and usability.

NOTE: Your metadata schema and metadata mapping documentation may also need be transferred to TAHO, to assist future researchers to understand and work with your agency's digital information.

Further Reading

For Information Management Guidelines and Advice, visit the GISU website.

- Guideline 5 – Recordkeeping metadata (2015)
- Advice 18 – Managing records in business systems (2014)
- ISO 16175 (Part 1-3) Principles and functional requirements for records in electronic office environments[8]
- AS/NZS ISO 23081 Information and Documentation – Records management processes – Metadata for records[9]
- Dublin Core Metadata Schema[10]
- Australian Government Recordkeeping Metadata Standard Version 2.0 (AGRkMS)[11]

---

[8] Available from http://www.egovernment.tas.gov.au/services/standards_select_online
[9] Available from http://www.egovernment.tas.gov.au/services/standards_select_online
[10] http://dublincore.org/
[11] http://www.naa.gov.au/records-management/agency/create-capture-describe/describe/recordkeeping-metadata.aspx

Good practice example: University of Waterloo Electronic Recordkeeping Metadata Standard (2011)[12]

[12] https://uwaterloo.ca/records-management/university-waterloo-electronic-recordkeeping-metadata

Further Advice

For more detailed advice, please contact:

Government Information Strategy Unit
Tasmanian Archive and Heritage Office (TAHO)
91 Murray Street
HOBART TASMANIA 7000
Telephone: 03 6165 5581
Email: grk@education.tas.gov.au

Acknowledgements

- State Records NSW, Advice on Digitisation – Metadata requirements (2014)[13]
- State Records NSW, Recordkeeping Advice on Designing and Implementing Recordkeeping Systems (DIRKS) (2007)[14]
- Queensland State Archives, Public records brief – Understanding and applying metadata (2012) and Metadata for digital continuity - a companion guideline to the Queensland Recordkeeping Metadata Standard (2012) [15]
- State Records of SA, EDRMS Functional Specification Standard (Version 3.1, 2011) [16]
- National Archives of Australia (NAA), What is digital continuity?[17]

Information Security Classification

This document has been security classified using the Tasmanian Government Information Security classification standard as PUBLIC and will be managed according to the requirements of the Tasmanian Government Information Security Policy.

Document Development History
Build Status

Version	Date	Author	Reason	Sections
2.0	01-06-2015	Samara McIlroy	Major revision of all content to support new Guideline	All

---

[13] http://www.records.nsw.gov.au/recordkeeping/advice/digitisation/metadata-requirements
[14] http://www.records.nsw.gov.au/recordkeeping/advice/designing-implementing-and-managing-systems/dirks-manual/doing-a-dirks-project/specify-and-apply-recordkeeping-metadata
[15] http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/metadataintro.pdf and
http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/MetadataDigitalContinuity.pdf
[16] http://www.text.archives.sa.gov.au/management/standards.html
[17] http://naa.gov.au/records-management/agency/digital/digital-continuity/index.aspx

| 1.1 | 23-05-2011 | Caitlin Sutton | Reference to Australian Government Recordkeeping Metadata Standard (AGRkMS) | All |
| 1.0 | 13-07-2005 | | Initial release | All |

Amendments in this Release

Section Title	Section Number	Amendment Summary
	All	

Issued: July 2015

Ross Latham
State Archivist