

Managing digital records: 3

Contents

3.	Use recordkeeping metadata for digital recordkeeping	1
3.1	What is the value of metadata?	1
3.2	Create metadata schemas for your business systems	2
3.3	Consider key points when developing metadata schemas	3
3.4	Consider key points when implementing metadata schemas	4
3.5	Facilitate preservation with metadata	5

3. Use recordkeeping metadata for digital recordkeeping

3.1 What is the value of metadata?

Recordkeeping metadata is a key business and recordkeeping tool. Good metadata makes a record findable, useable and attests to a record's integrity.

Good metadata enables good records management but, seen more broadly, the effective application of metadata can contribute to better information accessibility, organisational efficiency and greater accountability in business operations.

Time and effort should be invested in creating high quality metadata in your business systems. All efforts applied in creating good metadata will be repaid in accessible, useable data that is much easier to manage long term. The inverse is also true: incomplete or poor quality metadata, or ad hoc metadata that is created without due consideration of business requirements will require costly remediation actions in the long term and will limit the useability, integrity, manageability and effectiveness of data throughout the lifespan of your systems.

Tip: Assess your business needs before you implement your metadata

One large organisation bought an off-the-shelf software package and did not configure it to capture all metadata elements that were necessary to its business. As a result specific start and end dates were not captured for all records registered in the system. Some time later it was realised that as a consequence of the lack of these elements, it was not possible to run reports based on date criteria, nor to automate business activities based on date information. This had a significant impact on business efficiency. The organisation obtained a quote from a contractor about the cost of retrospectively applying the necessary metadata. Due to the vast numbers of records involved, the quote received was more than \$5,500,000.

It is critical to get your metadata right. You need to properly assess all your business needs that can be met by system metadata prior to metadata implementation. Not doing so could literally cost you millions and impact on business efficiency.

The importance of developing good metadata cannot be overestimated. Remember too that bad or ineffectual metadata can be worse than no metadata because it represents wasted effort and results in daily annoyances and inefficiencies.

3.2 Create metadata schemas for your business systems

To be a key organisational information resource, metadata has to be well specified and well managed, both immediately and in the long term. It cannot be created 'on the fly' or on an ad hoc basis.

If you are developing and implementing a custom business system, you will need to determine a comprehensive metadata schema that meets specific business needs. If you are implementing a software package, you will need to create a schema that identifies and clearly defines the specific elements of that system that should be used for your particular business purposes.

In both of these situations you need to make sure you have a standardised set of metadata elements. To do this you need to develop specific metadata element sets known as schemas which:

- specify the particular metadata fields that will be used in the system,
- provide a definition for each of the fields, indicating what can and what cannot be applied within them, and
- identify the encoding schemes or 'picklists' that are going to be used to provide data values in the system.

Example: What happens if you don't standardise your metadata?

One organisation implemented a records management software package. It did not recommend standard elements for its staff to use when registering records. Nor did it define elements for staff or explain what each element should and should not be used for. The software package it installed had over 500 possible data fields.

When it came time to migrate to the next generation of the software, IT staff confronted significant problems. Staff members had used over 80 data fields in the system when registering records. There was no standardisation in the data fields used – for example, some staff had used the 'Creator' field whereas others had used the 'Author' field to capture information about a record's creator. Some used 'Provenance', some used 'Creating organisation' to capture information about the controlling agency. Multiple options had been used for all commonly used metadata values.

When it came to migrating the system, simple metadata mapping and translations between the systems were impossible. A lot of data cleansing was necessary, multiple metadata mappings were required and ultimately it transpired that a lot of the metadata could not be migrated because the costs of translating multiple fields and values into specific fields in the target system were considered too excessive. Migration between systems was therefore very time consuming and complex and significantly exceeded the allocated budget. It also had ongoing implications for the authenticity and useability of the records which could potentially have legal and cost implications for the organisation.

Example: What happens if you do standardise your metadata?

Another organisation implemented a records management software package. It issued procedures to staff specifying the specific elements in the package that should be used to register records. It developed business rules for the system

that identified the specific relationships, access rules and disposal requirements that could be applied in the system. It also employed a range of encoding schemes that identified the specific values that could be used in a range of fields.

When it came time to migrate to the next generation of the software, IT staff had clearly defined documents that outlined all the fields used in the system and rules that defined how these fields were used. Regular monitoring by records staff had ensured that staff complied with these rules. Therefore data in the system was consistent and well defined.

When it came to migrating the system, relatively simple one to one mappings were easily achieved for all record types. There was minimal data cleansing and all data was successfully migrated from one system to another. Because of the straightforward nature of the migration, system administrators had time to concentrate on configuring the new system, implementing business rules and making improvements to system functionality and performance. Migration came in on budget and resulted in actual business improvement for the organisation.

3.3 Consider key points when developing metadata schemas

1. *Metadata must support business.* The metadata schemas you develop need to be based on specific knowledge of the business that needs to be performed within the system. Talk to staff and perform system assessments so that you know what is currently done and so that you can gain an awareness of how things can be done better. Is, for example, information retrieval an issue? What metadata fields could be included to address this problem? Do standard reports need to be run? What data elements are necessary to generate these? What system structure and functionality that can be enabled by metadata is required?
2. *Metadata must support recordkeeping requirements, as well as business requirements.* Incorporating the mandatory identifier, title, date, creator, business and format elements that are required by the State Records' *Standard on Digital Recordkeeping*, as well as the elements that document the management processes you perform on your records is critical to sustaining their immediate and long term authenticity and integrity. Long term, it is the aggregation of this form of metadata that will allow your records to be accepted as meaningful and accountable evidence. Capturing and keeping each of these forms of metadata is also necessary to provide you with the information you need to develop the migration and other preservation strategies that will sustain your records in the long term.
3. *Metadata schemas may not be uniform across your organisation.* This is because business needs may differ from section to section. Remember to support individual business requirements wherever possible through the inclusion of specific metadata fields that, for example, drive particular workflows or incorporate a current client specific naming schema, etc. Each specific schema you develop will need to be documented.
4. *Metadata can be scalable and can be applied to a range of different things.* Metadata can be used describe records at different levels of aggregation. For example, metadata can be used to describe:
 - individual records at the document level (for example, you should apply titles and unique identifiers to each individual record in your system)

- groups of records (for example, you may want to apply disposal rules at the file level to facilitate the management of this process)
- systems (for example, you may want to apply access rules metadata to the whole personnel records database, to standardise access control across the system)

You will need to determine which levels of metadata can best meet the business needs of your organisation. Metadata can also be used to describe things other than records, such as people, workgroups, organisations, business transactions, activities and functions, as well as mandates such as laws, regulations, policies and business rules. You should determine whether metadata descriptions of any of these entities would be of use to your organisation.

5. *Develop good encoding schemes or 'picklists'*. Encoding schemes provide a controlled list of all the acceptable values that can be applied within a certain metadata field. Use of encoding schemes can be critical for enabling interoperability between systems. They also promote standardisation, consistency and accuracy, they can make it easier for staff to automatically apply metadata values, they can facilitate metadata reuse in other business areas and, in the longer term, they can contribute to system sustainability. Good picklists can add so much value to your system and dramatically improve its useability so a good deal of time and effort should be spent on their development. Encouraging automation through picklists, such as automation of disposal information through the application of controlled titling values, is also valuable functionality that should be implemented where possible.

3.4 Consider key points when implementing metadata schemas

1. *Plan for metadata automation, inheritance and reuse*. Automating of as much metadata capture as possible makes life easier for staff in your organisation, enables more consistent metadata application and can also save time and money. You can also improve efficiency if metadata created for one business purpose can be shared with other applications for other business purposes. For example, flexible interfaces could be established between records systems, desktop document authoring applications, web content management systems, human resource management systems, work flow systems and/or business databases to share rather than recreate metadata. Metadata reuse in this way saves money and can dramatically improve how business is performed in your organisation. Mapping and standardising your metadata is critical to enabling this type of functionality. ¹ Fluid metadata implementations of this type need to be fully documented to help manage the migration and long term preservation of these systems.
2. *Metadata must start to be captured from the moment a record is created*. More metadata must also be added to the record when it is used in different processes or as different management actions are performed upon it. No metadata should be deleted during this process, only new metadata added. This is necessary to ensure ongoing the authenticity, integrity and reliability of records

¹ For more information about the reuse of metadata, see Monash University, *Create once, use many times: the clever use of metadata in eGovernment and eBusiness recordkeeping processes in networked environments: final report*, viewed July 2008, <<http://www.infotech.monash.edu.au/research/groups/rcrg/crkm/docs/rpt-final.doc>>

3. *Metadata itself is a key record.* Through time it will attest to a record's integrity by documenting its context and management history. Metadata must therefore be maintained. For the lifespan of the record to which it relates, metadata must be kept and kept in context. That is, it must be persistently linked with the digital records and aggregations of digital records to which it relates, including when they are transferred out of their original creating environment and through subsequent migrations. If the record it relates to is destroyed (in accordance with an authorised retention and disposal authority) the metadata must continue to be maintained. Retention and disposal authorities relevant to your organisation will outline the specific retention requirements that apply to the different forms of metadata maintained in your organisation. General Retention and Disposal Authority – *Administrative records* (GA28) contains retention requirements for some metadata.²
4. *Systems are only as good as the people who use them.* You could spend significant amounts of time and money developing standardised metadata schema but if people do not implement these as intended, all the efforts involved in their development will be wasted. Spend time and money explaining the roles of the different metadata fields in your business systems and explaining how the particular encoding schemes that provide values for these fields operate. This type of work may be difficult and repetitive but it is a critical component of the effective operation of your systems and also for the long term sustainability and integrity of the systems.

Tip: Metadata is crucial for records stored on removable storage media, such as CD and DVD-R

To facilitate accessibility and long term preservation, removable storage media, if used, must be well described with metadata. A label should make clear what is on a particular disk. The disk itself should also contain all metadata necessary to explain the content, context and management of all the records contained on it.

Further assistance:

For further guidance on developing and implementing recordkeeping metadata schemas for your organisation see *Strategies for documenting government business: the DIRKS manual* at http://www.records.nsw.gov.au/recordkeeping/dirks-metadata_1849.asp.

3.5 Facilitate preservation with metadata

Within a well specified system, recordkeeping metadata can be used as a tool to proactively plan and then perform migration or other preservation activities.

This metadata can take the form of:

- technological dependence metadata

² See for example the retention requirements outlined in State Records Authority of NSW, General Retention and Disposal Authority - *Administrative records* (GA 28, 12.9.1 and 12.9.2) , viewed December 2008, http://www.records.nsw.gov.au/recordkeeping/introduction_12727.asp

- disposal metadata, or
- migration metadata.

Technological dependence metadata

To facilitate their management, all records at the item or aggregate level should be tagged with metadata that documents their technological dependencies. For example, all records created in a specific software format should have the name of that format appended to them. The version number should also be recorded, if applicable.

Where possible, systems should be designed to capture this metadata automatically, deriving it from the systems and applications initially used to create the record. The metadata that needs to be recorded does not need to be extensive. It just needs to provide a meaningful and concise representation of the hardware and software dependencies of the record. For long term preservation, it can also be critical to record the operating system and the necessary peripherals that support different types of records. You need to think about the specific format dependencies of each of your record types and determine the metadata that will need to be recorded about each of these.

In many circumstances it may be appropriate for this metadata to be applied to each record individually. The metadata can then be used to audit records and determine which need to be migrated or it can be used to automatically trigger the migration of specified format types. It will also allow usage of different formats to be tracked and preservation strategies planned for the diversity of record formats used within an organisation.

Where all records in a business system have the same dependencies, this metadata can be applied at the aggregate or system level if preferred.

Tip: Remember contextual metadata

While contextual and descriptive metadata may not serve obvious preservation purposes, these forms of metadata are vital for maintaining the ability to use and understand a record through time. This ultimately is critical for ensuring long term accessibility.

Disposal metadata

Migration and other forms of preservation actions are frequently costly to perform. Therefore preservation actions should only be applied to the records that require them.

To ensure that only appropriate records are subject to preservation actions, regular disposal operations should be conducted. Metadata can play a key role in:

- flagging the disposal conditions that apply to each record
- automating disposal operations where appropriate
- ensuring that disposal operations are carried out in a timely and accountable manner, and
- documenting the performance of disposal operations.

Migration metadata

Driving migration

Metadata can be used to drive and document preservation actions. For example, in appropriately designed systems, metadata can be used to identify all records with specific format dependencies, specify an appropriate migration pathway and migrate these records to a nominated format.

Documenting migration

Metadata can also be used to document and describe the migration activities performed upon records. You may want to tag each individual item that has been migrated with a description of the type of migration it was subject to.

Alternatively, you may find it preferable to describe migrations at the aggregate level by specifying that a group of records was migrated on a particular date according to the specified migration strategy.

Whichever option is selected for documenting migration, it is important to note that the documentation itself is a record of the preservation action. It becomes an important component of the record's management history that can be used to attest to its authenticity and appropriate management.

Remember too that each preservation action performed on a record needs to be documented. Therefore if a record undergoes multiple migrations, it will accrue numerous instances of preservation metadata to describe each of these operations.

Documenting negative ramifications of migration

In certain circumstances, migration actions may not go as planned. In these instances metadata can be used to document how migration has negatively affected certain files that have been damaged or corrupted by the process.

© State of New South Wales through the State Records Authority, February 2009. This work may be freely reproduced and distributed for most purposes, however some restrictions apply. See the copyright notice on www.records.nsw.gov.au or contact State Records.

ISBN: 978-0-9805148-9-6