

In today's presentation I am going to look at some new techniques for deploying classification schemes that you may want to consider when implementing recordkeeping systems in your Councils or organisations - a shift in thinking about classification to take advantage of new technologies.

We will explore the opportunities that are available to us in the digital world.

CONTEXT

We have been managing digital objects for a while now but many of our classification techniques were simply automated paper based processes. By that I mean we have been simply managing our digital objects, in the same way as we managed paper objects.

But digital recordkeeping demands a more sophisticated approach, particularly in regards classification as there is greater flexibility of implementation options in systems these days which

actively exploit the use of metadata in multiple ways and end user tagging.

A great deal of work has been done on classification in recent times in particular around the publishing of the Standards Australia Records Classification Handbook.

My colleagues, from Recordkeeping Innovation, Barbara Reed and Cathy Ashton have been heavily involved in the development of the classification handbook and IT 21 – Standards Australia Committee for Records Management which:

- Developed AS 4390 the first records management standard in the world which became the basis for ISO 15489
- Development of metadata standards – which has been internationalised, and also

- Development of the work processes analysis standard for records which was internationalised as an International Technical report

WHY DO RECORDKEEPERS CLASSIFY?

Note: Ask audience “why do recordkeepers classify?”

There are very good reasons why recordkeepers classify. The purpose served by classification based on business activities has been well established for many years. It's a cornerstone of best practice recordkeeping.

Fundamentally the reason why we use classification is to link the records we manage with the business action (note: write on board) without this link, records aren't contextualised and it becomes difficult to make assertions about their status as evidence of action.

What is changing are the ways we develop and apply our enterprise classification schemes. This is driven largely by

emerging technologies and opportunities to access and use digital records in clever ways. With smart configurations we can deliver what our users want, removing from them the burden of complex recordkeeping and rules based systems.

So the role of business classification is to ensure that the links to business, and how records are grouped together, reflect what was done and to enable the re-telling of the business action story.

This is known as context. And in the recordkeepers world, the four pillars of context are:

Authenticity –ensuring that we can make assertions that the record was part of business, at a specific time and place.

Systemisation – reducing and controlling the unmanaged or ad hoc changes to the records systems so that the links to business process are reliable, can be automated and are sustainable across time.

Usability – ensuring that the “story” can be retold, so that a user can understand the original sequence of events, but can also see and re-use the records in multiple sequences and enable us to see the sequencing of records that precede and succeed individual transactions.

Relationships –ensuring that there are visible explanations to how and why the record was created. In the hardcopy world this was overt and much easier – by its placement on a file, by knowing who had had custody or ownership of the record. But in a virtual world, where there is no “physical” thing or “one location”, this is achieved by preserving the relationships between the content and the metadata that describes and records these characteristics.

WHY WE CLASSIFY? METHODS

But the urge to classify, group and present information is not the sole provenance of recordkeepers – nor does it have to be a method for arranging information at the expense of all others. It

has to be present sure, but it can be in the background, embedded in the system design.

The urge to sort and arrange things to make the world understandable or to achieve a goal is not unique to recordkeepers.

- a shop manager arranging their products by type, price, category or specials,
- the librarian coding and key wording books by subjects for retrieval.
- the botanist that arranges their specimens by genus, species, or family. These schemes are very formal but others may be ad hoc or made up such as the tagging in the family photo collection or arranging by date.

In an organisation the need to classify is not unique to recordkeepers – there's a lot to share with organisation designers and enterprise architects for example. Our methods of arranging

information don't have to impose a record centric view at the expense of all others.

We have to find the common ground with our related information, knowledge and content management professionals and if we can't there is the potential to create risks for the recordkeepers:

1. We don't always utilise to our fullest extent the techniques and innovations pioneered and adopted by our colleagues in the related information management fields.
2. We don't always see the linkages to other professionals integral to recordkeeping and recordkeeping work, but which are not necessarily called records managers or archivists – and I'm thinking here of the enterprise architects, business processes analysts.
3. We will have no relevance in the digital world, where our users have a higher degree of expectation. We will become the "compliance" people, the "disposal people".

And that would be a huge shame. Because the recordkeeping profession, with its disciplinary heritage of sustainability, authenticity, longevity, and preservation, has a lot to offer in a digital world, because structure in this world is now virtual and the linkages between information, business and systems can be very fragile.

And there are many ways of presenting information to our users, like this word cloud on the slide here - which is illustrating relevance. The word cloud is a strong visual indicator and distinguisher of the relevance and topics of groups of search results. We should be embracing and learning how to use new techniques to enhance how we deliver retrieval results and information to our users.

MYTHS

So I would like to dispel a few myths and separate the mechanisms that we created for the paper-based world from how we can operate in a digital world.

A place to file

This is a paper-based paradigm. In the digital world, structure is virtual. While many of the early tools of the electronic document and records management systems replicated this **place to file** thinking – in terms of file plans and tree structures, it should only have been considered a transition, rather than maintaining it for as long as we have to keep us in our comfort zones. In the digital world, it doesn't really matter where in a file an object is; as long as we have context and how the records are grouped reflect what was done.

Attribute not structure

Classification is a metadata attribute, not a structure. It's only when we reposition classification as an attribute that it can be applied to multiple levels of aggregation, not just containers of documents. And this is necessary if we're are to deliver customised views of information, suited to individual needs that our users in a My Space, Facebook, Web 2.0 world demand, and

which can actively exploit our need for multiple classificatory purposes, such as searching, end user tagging and retrieval.

Search and retrieval:

In the past records classifications were sold to users as a solution to searching and retrieval and they failed. Records classification has never been about search, but enabling the explicit identification of business context. Search and retrieval is best met through even the simplest search applications that are able to distinguish data according to a range of metadata attributes.

Using terms that describe and control the subject, case names and topics establish access points to facilitate the retrieval of records and are based on content. They can be derived from a preferred organisational vocabulary, industry standards and encoding schemes used in business to describe information.

It is important to ensure that content labels are consistent and that different words are not used to describe the same topics, e.g. dog v canine, or car v motor vehicle.

Disposal

Disposal is not the reason we classify, but it is a very useful by-product. Designing classification schemes to meet disposal requirements is just the disposal tail wagging the classification dog. While in the paper world there was a need to divest organisations of voluminous and costly records, in the digital world, work process analysis, which allows more specific representations of business, supported by automatic sequencing, makes attribution of disposal much easier.

Hierarchical

We've all probably been scarred by the three level file plans of functions, activities and subjects/ transactions at some stage in our recordkeeping lives. In the digital world, these constraints have been rendered meaningless. Structure is virtual and is dependent on the needs of the business. Who is concerned about the number of containers on top of transactions? By

conceiving classification as an attribute rather than a structure we are liberating it from fixed levels of hierarchy.

Functional Classification is passé

Analysis of business has been at the core of recordkeeping for centuries. There are very sound reasons for sticking with that approach because we **need** to link our records to the business actions that created them. What needs to change is how we deploy and implement classification – hiding it from the end user, customising it for discrete sets of business, embedding it within the system design. Many other parts of an organisation adopt a functional approach, then place their own disciplinary gloss above it, for example, the enterprise architect with his model of business so that IT investment can be rationalised. So functional classification is not passé - lets implement classification a bit more cleverly.

NEW TECHNIQUES

When developing a classification scheme for your organisation the development method and purpose for its use, will influence the look and feel of the scheme. We can all relate to the simplest form of classification based on a single hierarchy. It has few additional term attributes and linkages and incorporates links to topics and subjects within a simple hierarchy.

Additional information, such as topic descriptions, specify prompts and segmentation instructions can be applied to the simple scheme to enable application within software. The specify by prompts provide additional guidance to users on how to segment groups of records that rely on folders to classify and aggregate records. So in construction some of the things you need to think about:

- how will the term be used - as part of workflow, linked to roles or document types.
- What are the informational, subject or indexing points.

NEW TECHNIQUES

This is a more detailed scheme that incorporates a description of work processes and transactions and can be used to define the configuration of systems, as an input into information architecture or to assist with implementation of a recordkeeping system.

This detailed model demonstrates how the scheme can be used to link the record classification with disposal mandates, business owners, access and security. The model shows how the scheme can be used to identify and expose transactions so that the scheme can re-group records to tell the story within a workflow rather than as folders or documents aggregated into containers.

This allows for more implementation options, for example:

Note: walk through the slide with the audience

- Work can be linked together to form a continuous record of activity using relationships.

- The meaning of terms used is clearly understood by those involved – they are clear action statements.
- Work can be distributed, linked and tracked to the appropriate people.
- Work processes, the resulting records and user permissions can all be linked
- The scheme can be used in business systems, EDRMS, collaboration based systems and CMS.
- Additional metadata, topics and indexing points, can be applied to enhance the findability of the records and information.

This type of model has significant applications in the digital world in the configuration options that are possible for business and in the way that we present views and information to our users.

OPPORTUNITIES IN THE DIGITAL WORLD

Digital records provide opportunities to think laterally, overcoming the **physical constraints** associated with managing paper files.

Paper records forced a single view of information, hence the focus on hierarchical folders. The current practice of using file plans presents functional classification schemes that are hard to follow and use, because the sequence of records is disturbed. And our users were very often unimpressed with these – often they were not very meaningful to the end user.

Digital environments support multiple views of records, so **location is not fixed** by who 'owns' the record – instead of physical duplication based on where the user sits in the organisation, different sequences can be created based on roles and access rules. This is a more sophisticated understanding of how records travel through organisations, enabling transactions, where there may be many sequences of actions from the same record. Instead of duplication, we need automated processes, to use workflows and develop the associated access rules.

The digital world provides an opportunity for designers and implementers of classification to free themselves from the constraints of the paper world, no longer restricted to a single view of information but supporting multiple views of records.

Note: discuss with audience the different views that different users require to support different uses of records.

Purchase Order example - finance view and business unit/department view.

Business units can search for all their purchase orders, retrieve them and view them according to particular suppliers, or by date or by product or a combination.

This is different to the way that Finance need to view them.

If we look at the Infringement Notice - the compliance area has a process that the Infringement moves through but there may also be the need to view that infringement in relation to others for a particular property or person. So the property view is required.

In the recent paper-past, when the volume of records being produced increased, it forced recordkeepers into designing schemes largely for folder or files of records, as it was no longer feasible to describe, index and link individual documents, as was common in the early registry systems of large organisation.

However, with capacity to inherit and capture metadata automatically created by even the simplest document and record creation applications, the burden to manually record this information was eased, and the need to concentrate all indexing points within the single title or free-text of record eradicated. This means better data structuring, creating a range of lists or pre-populated fields, reducing data entry.

OPPORTUNITIES IN THE DIGITAL WORLD

So the increased feasibility of reverting to the past practice of item description, and the automatic generation and capture of metadata, creates numerous opportunities, which were not possible in a paper environment, such as:

- The capacity to customise and present **personal views** of information, which the “my space” generation of users now come to expect from their information systems.
- Assigning multiple classification systems to a single digital record, project classifications, geographical systems, names. These richer digital associations enable multiple perspectives of one object, including the ad-hoc, the personal and folksonomies, for a richer, more intuitive user experience that better reflects their needs and expectations.
- This ability to associate a single digital record, to multiple sequences, people, organisations, eliminates the practice of physical duplication. These digital associations enable multiple perspectives of one object.
- And combined with the extremely clever and ever increasing capabilities of our **search**, our users are no longer required to navigate and browse, that which reflects a paper based paradigm.

These emerging capabilities require clever metadata design and presentations, use of customisable data capture fields, the ability to change metadata field names for different types of records, template management for automated data capture and pre-populated drop down lists – anything to reduce data entry by users. In effect we are taking the free out of freetext.

These opportunities within the digital world remove the pressure from recordkeeping classification, or any method of classification for that matter, to be all things to all users.

Semantic-Aware Applications. New applications are emerging that are bringing the promise of the semantic web into practice without the need to add additional layers of tags, identifiers, or other top-down methods of defining context. Tools that can simply gather the context in which information is couched, and then use that context to extract embedded meaning are providing rich new ways of finding and aggregating content. At the same time, other tools are allowing context to be easily modified, shaped, and

redefined as information flows are combined. There is a great deal of work going on in this area, and we can expect to see significant advances in the coming years.

ELEVEN - TROVE

The National Library of Australia has developed a search engine that provides access to more than 90 million items about Australian and Australians.

The website is searching across metadata mostly from cultural heritage organisations and there is over a thousand organisations that have been providing data. TROVE displays links to items as opposed to a Google search which provides a list of websites for search results.

The National Library of Australia Trove website is split into eight searchable categories:

- Books, journals, magazines, articles
- Pictures and photos

- Australian newspapers (1803 – 1954)
- Diaries, letters, archives
- Maps
- Music, sound and video
- Archived websites (1996 – now)
- People and organisations

AND provides a single point of access to the resources of the deep web and facilitates access to a significantly greater range of resources from major sources.

This slide displays an example of a search on Nellie Melba that then brings together a range of resources derived from the eight categories.

So the importance of metadata is not only key in the creation, management and preservation of electronic records as well as their use as archives, memory and knowledge.

SUMMARY

So in summary...

Shift in thinking from hardcopy to digital models to take advantage of new technologies

- From classification structure to classification as an attribute
- From aggregations of items in folders to items
- Classification embedded in workflow and also
- Sites, business systems and collaboration spaces

Virtualisation” makes the functional approach even more valid there is

- No longer “either/or” – could it be this term or could it be that term
- Not guidelines on vocabulary control but clear statements about business with
- Links to the actors and agents involved in processes and

- Relative to business need, risk and objectives

“Harnessing the crowd” by...

- Enabling customised end user views and personalisation of workspaces.
- Using multiple classificatory schemas and folksonomies and tags, and.
- Encoding schemes deployed as individual metadata elements and values

Thank you