

Metadata Standards for Effective Resource Discovery and Records Archival

A presentation given at the 5th Annual Electronic Document and Records
Management Summit – Dockside, Sydney – 28-29 April 2004

Kerry Webb
Web Architect
ACT Information Management
ACT Chief Minister's Department
kerrya.webb@act.gov.au

2 Introduction

Some housekeeping announcements

- Statistics

There will be no meaningless statistics. At a recent conference, there was a claim that 80-90% of an organisation's knowledge is not known. How do they know – it's written on a piece of paper – a line worthy of an Eccles / Bluebottle exchange.

- Important points

Important points will be highlighted. There will be no music, but I will raise my voice slightly.

- Interaction

Please feel free to interrupt any time. I once spoke at a conference and a member of the audience disagreed strongly with something that I said about the use of metadata, but didn't have time at the end to explore this. Consequently, I received an adverse comment on the feedback form! I wouldn't want that to happen this time.

- It's about a journey!

Well, life is a journey. But, just as the act of planning is often more important than the eventual plan, so can the process of developing a standard be more useful than the resulting standard. And the journey doesn't finish with the publication of the standard ...

And finally:

It's what we did. Nothing that I say should be taken to imply that we're right and everyone else is wrong. We all have a whole swag of pressures on us. What I'm describing is what the ACT government has decided is in its best interests.

3 The question that we should always be asking ...

What are we trying to achieve?

(this is an important point, so I'll raise my voice)

4 The context of the exercise

The ACT Government is introducing an IDMS – as outlined by Robert Gotts yesterday

This has been going on for four years. Robert and I sat opposite each other and had many discussions about metadata and related matters and Rugby. The result is that the Brumbies continue to do well (last Saturday was an aberration!), and that both of us have a better appreciation of the other's project, and the implications of interoperability.

We have established a Territory Records Office with appropriate legislation

This arose in large part because of various occurrences in the last 15 years. First, when the ACT was granted self-government we were no longer covered by the Archives Act, and we realised that number of decisions taken over the last decade had not been recorded sufficiently. The Territory Records Act was passed in 2002, and ACT government agencies are now in the process of implementing it – an exercise involving some pain.

We continue to improve our Web publishing activities

There are around 100 ACT government websites, and growing. The resources on all of these have been defined as “records” and consequently will have to be managed in a record keeping system, as defined by the TRO.

5 IDMS Metadata

The difference between the approaches taken by Victoria and the ACT

Early in its planning for the IDMS, the ACT government decided to use the Victorian Electronic Records Strategy (VERS) standard – including its metadata requirements – and VERS compliance was a requirement in the RFT. Later it became clear that we were talking about different processes: VERS was interested more in archival issues than in business operations.

Implications for record keeping metadata

Accordingly, our emphasis on data management meant that we were looking at a more restricted metadata schema for the IDMS – which will eventually incorporate Web-based resources, which is what we're talking about here.

6 Metadata required by the Territory Records Act

Descriptive elements - more later!

- Date

This is the date when a record was “made”, or in the context of Web resources, last updated. This is relevant for access, for evidential reasons, and for context.

- Creator

This is the person who is responsible for taking an action relating to a record or file. This is important in understanding the record’s content, but also to determine whether that person had the authority to make the record or to carry out the business. It will not necessarily be the person who keys in the information.

- Title

This is the name given to a record or file. Remember that this should be sufficient to identify individual records and to bring them together from whatever system is used to store them. There will be more of this later.

- Identifier

This is an identifier or number or control symbol that is unique within the agency’s *recordkeeping system*. There’s potentially a difficulty here – if you use the URL as the identifier.

7 What does this mean for Web resources?

What are we trying to achieve?

- Metadata for resource discovery - Compliant with portal and search

A few issues here. The ACT government portal [<http://www.act.gov.au>] is driven by metadata; there is no content on the site, just a repository of links and metadata. The Subject metadata field is mapped to the information architecture on the site, so that when a user clicks on a node the subject(s) associated with that node are used to search the metadata in the repository and a set of results are returned.

Metadata is also used in the whole-of-government search facility that's provided on the portal site. Its main purpose is to provide important information in displaying search results, using Creator, Title, Description, Identifier etc.

Another use is potentially in Advanced Search, but our experience in this is in line with others' – people are reluctant to use it.

- Metadata for record keeping - Compliant with IDMS

We can't be definite about this just yet, but we try to ensure that what we specify is not incompatible with the IDMS requirements.

- Metadata for record keeping - Compliant with the TRA

This is more complex and will be address later in this presentation.

8 What does this mean for Web resources?

What are we trying to achieve?

- Not making it too hard for the PBWW!

This is probably the most important issue in the development of a metadata standard. If we accept that we need metadata and that the metadata has to be good to be useful, then we mustn't make it hard for people to create good metadata.

It's a (sad) fact that most Web workers don't leap out of bed each day and shout to the dawn: "I'm going to create great metadata today!" Most of them want to get on with publishing their Web stuff, and most of them see metadata creation as an unwelcome intrusion into their valuable time.

People's attitudes to metadata are strange. I've heard librarians ask what the fuss is all about – it's just cataloguing, isn't it? And who makes a fuss about cataloguing? Well, librarians do, that's who. They've been formulating cataloguing rules for centuries, they teach cataloguing at university and TAFE courses, they have special cataloguers sections in their professional organisations and they have conferences where they get to sing great songs about cataloguing. In which I've been proud to take part.

But it's the sheer professionalism of the discipline that's the difference. Professional cataloguers catalogue! There are no (or at least very few) full-time professional metadata creators. For the most part, we have to rely on unwilling Web workers to carry out this important task, and we have to educate them about the importance of this extra work that we're giving them.

And we also have to make the specifications clear and we have to have as many examples as we can create so that they don't have to spend too much time choosing their metadata.

9 The current status of metadata standards for Web resources

First there was Dublin Core

Not named after the home of Guinness, but after Dublin OH, home of OCLC.

Then there was AGLS

AGLS used to stand for the Australian Government Locator Service, now it's just a set of letters.

Now there's the ACT Government implementation

Different jurisdictions are free to implement AGLS in their own way – the ACT has just come late to the party. But there are difficulties in doing your own thing, even if it's only a little bit non-standard.

10 Dublin Core

The Dream

The Dublin Core metadata element set is a standard for cross-domain information resource description.

Simple to use, and general enough to be applied to resources in any discipline. It also helps if important organisations (eg search engine companies) take notice of it.

The Standard

Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights

What's mandatory, what's optional

Strictly speaking, no elements are mandatory. It really depends upon your implementation, and that's where application profiles come in ...

What does it mean in Australia

This is where it gets confusing. There is an international standard *ISO 15836:2003, Information and documentation - The Dublin Core metadata element set*. A Standards Australia committee (IT-19 – from the library world) is working towards cloning this as an Australian standard.

So, when this is published, an Australian user wanting to implement metadata will have some guidance, but ...

11 AGLS

Where it came from

The key year was 1997. A Commonwealth Government committee recommended that a government standard be developed based upon GILS, but a workshop held in December of that year recommended instead that it be based on Dublin Core.

Since then, it's gone on through a number of revisions, under the guidance of the NAA. Function, Availability, Audience and Mandate have been added to the basic Dublin Core elements. There has also been considerable work in developing schemes and controlled lists for audience, services etc.

The Standard (AS 5044-2002)

In 2002, Standards Australia committee IT/21 (from the record-keeping world) published the first Australian metadata standard, based on AGLS, which from that time has ceased to be the Australian Government Locator Service. AGLS is now just AGLS.

Mandatory

Creator, Date, Title, Availability / Identifier, Function / Subject, Publisher

Optional

Type, Audience, Contributor, Coverage, Description, Format, Language, Mandate, Relation, Rights, Source

My final comment on these standards is how they may be perceived by people coming new to the world of metadata. They will be presented with two slightly different standards, which is not a good idea. IMHO.

12 So ...

What are we trying to achieve?

- We're setting the framework for a system

We're trying to formulate up a metadata standard for Web-based resources that can:

- meet existing requirements for resource discovery (see the reference to the portal described above)
- interact with other systems (such as Governet ...) at some time in the future
- work (in general) with record keeping systems at some time in the future
- work with the IDMS at some time in the future

AND

- does not drive Web managers around the bend.

13 What Web resources should have metadata applied?

In a word ... everything

- For completeness and simplicity

The AGLS implementation notes say that you should be selective about what resources should have metadata (presumably to decrease the load on the workers).

But, in our implementation, we think that the basic metadata related to resource discovery (Title, Description, Creator/Publisher, Identifier) should be present for all resources (to assist in the display of results). And the metadata for record keeping (Title, Creator, date, Function, Identifier) must also appear for all resources.

So, to avoid Web workers having to weigh up how far they should go for any particular page or document, we simplify it: do 'em all!

14 Recordkeeping metadata – Function / Title / Subject

Identification of the record is what it's about

These elements should be enough to identify a particular record, although it may also be necessary to include Date.

15 Recordkeeping metadata – Creator

Different types of Creator

Internal creators – easy to identify, or are they? Remember it's the person responsible for the intellectual content. Not the Web editor, not the person who wrote the first draft, but who has the authority. And what about documents created by consultants?

External creators would include content obtained (with permission) from other sites, and public submissions.

More than one Creator?

We had some discussion about “multiple creators”. But, in the end, there has to be one. Just one.

Let's not identify them too closely

And a final thought on Creator – remember that the META tag will be (in most cases) “in clear”. You really don't want their e-mail addresses in the HTML. Or their names for that matter – which is why we use Position Title rather than Personal Name.

16 Recordkeeping metadata – Date

Which dates are important?

Created, of course, and Modified. Remember, we're trying to mimic what would be on a paper-based file. Valid From and To have little meaning in Record Keeping, although they do in Content Management (which is something that I haven't addressed in this presentation).

These are important: not just for indicating when the record was online, but also for who was responsible for it.

17 The end of the journey?

A Metadata Implementation Manual (now a Standard)

- Metadata for everything
- Mandatory: Creator, Date, Description, Title, Function, Subject, Identifier, Publisher
- Everything else is optional (or even discouraged)

And if you want to discuss anything about this:

kerrya.webb@act.gov.au

<http://www.alia.org.au/~kwebb/papers/>

Postscript

At the session where I gave this presentation, there was some concern about our “non-standard” approach. One person offered the option that if we’d taken this approach at Federation, all the states would be talking different languages. I can only suggest that he wasn’t listening to what I was saying.

The difference between the Commonwealth’s AGLS implementation and that of the ACT can be summarised as:

- The Commonwealth says do not apply metadata to all Web-based resources; the ACT says that metadata should be applied to everything.
- The Commonwealth says that Function or Subject must be applied to Web-based resources; the ACT says that both must be applied.
- The Commonwealth says that AGIFT may be used as a source of Function terms; the ACT directs that our own thesauri must be used.
- The Commonwealth makes no direction about which thesaurus should be used as a source of Subject terms; the ACT directs that our own subject thesaurus must be used.
- The Commonwealth says that Description is an optional element; the ACT says that it is mandatory.

If anything, the ACT standard is a stricter implementation of the AGLS standard than the Commonwealth version, and in the unlikely event that there is some future need for interoperability between ACT and other jurisdictions’ metadata (noting the imminent demise of the GOVERNMENT project) there will be no incompatibility at all.

Kerry Webb
12 May 2004