

Creating and using metadata

Discussion notes for presentation at the Community Web Workers Conference, Sydney, 8-9 November 2004

Prue Deacon, HealthInsite Editorial Team, Australian Government Department of Health and Ageing

prue.deacon@health.gov.au

phone: 02 6289 7505

HealthInsite background

- HealthInsite (www.healthinsite.gov.au) is a gateway website, linking to health information resources on Australian websites – 75 information partners around Australia are contributing nearly 12000 resources. There is a metadata record for each resource.
- Over half the partners are NGOs, many of them small support groups.

Introduction

- What I am going to tell you about metadata is from the viewpoint of a gateway site where metadata is essential.
- However, for small, independent sites metadata may not be necessary. I hope that what I say will help you to decide whether your site needs metadata.
- Metadata is quite expensive – not just its creation, but building it into your content management system, workflows and/or search facilities.

Metadata – what is it?

- Firstly, to repeat what I said earlier today, in web publishing we often refer to “resources”. A resource may be a fact sheet, a publication or information about a service or organisation. It may also be a group of items (where the entry point is the top level page) or a whole website (where the entry point is the home page).
- The metadata that I am going to discuss is catalogue type information about a resource – things like title, publisher, subject, date. These are called elements, but are equivalent to fields in a database.
- If your metadata is in a database or content management system (CMS), then you may have other fields to help control your resources – some of these might be considered as metadata too.
- To see an example of metadata, go to <http://www.mentalhealth.asn.au/resources/depression.htm> and have a look at the source code. The metadata is near the top.

What is it for?

- The prime purpose of metadata is to allow for enhanced search and navigation functionality.
- For a gateway site like HealthInsite, it is absolutely vital. For example, we link to hundreds of resources that mention asthma. Metadata helps us to clearly identify which ones have asthma as a major topic and give these a higher ranking if a user searches for asthma. Metadata also helps us to group resources into topic pages and this is the basis for our main navigation/browse

structure. So, for example, we can help users find resources on different aspects of asthma, such as asthma medicines.

- Some metadata elements (for example, language, date, audience) are useful for sorting search results or refining a search.
- Some metadata elements (for example title, description, publisher, date) are useful for display in search results.
- Finally, some metadata may help with site administration – for example, the date when a resource needs to be reviewed.

Metadata standards

- The Australian standard is AGLS. (AS 5044.1-2002 AGLS metadata element set. Part 1: Reference description. AS 5044.2-2002 AGLS metadata element set. Part 2: Usage guide.)
- AGLS is based on the international Dublin Core standard. When you see DC in metadata coding – for example, DC.Title – the DC stands for Dublin Core. These standards were developed by a wide user group, including IT developers and librarians.
- AGLS is very flexible, allowing organisations to refine the standards to meet their needs. This has a downside: two organisations may have incompatible metadata even though they both comply with AGLS. For example, HealthInsite tends to use the simpler options in AGLS; other government agencies have opted for more complexity.
- If you participate in a gateway site, then you are best to start with the minimum standard for the gateway.
- Then (and otherwise) you need to look at each element of AGLS and ask questions like: How could it be used? Will it help users to find particular resources on my site? Is it really needed? My recommendation is to take the simplest option that will satisfy your needs. You may find it worthwhile to develop your own schemes for some elements.

Can I get away without metadata?

- If you have a small, simple site and very little interaction with other sites, you may be able to avoid metadata.
- Metadata creation is labour intensive – you need to allow 10-20 minutes for each resource. So you may think it is quite expensive. However, I see metadata as the final stage of publishing a quality resource. When you look at the time taken to write a document, the extra effort for metadata is not much.
- In HealthInsite, we now allow our information partners to submit short metadata records – creator, publisher, title, description, date modified, language, format, identifier. We create the rest in our database.
- Even these short records have been difficult for some of our partners. However it is essential that partners maintain the records on their own sites and update them whenever they update resources. This allows the HealthInsite harvester software to retrieve the updated records from partner sites and use them to update the HealthInsite database.

Metadata is part of a system

- Metadata should not be implemented in isolation – it must be considered as part of an information retrieval system. There is little value in creating metadata if there is no system with the search functionality to use it.
- If you are contributing metadata to a gateway site, then you can leave this to the gateway developers.
- However, if it is your own site, then you need to work closely with the people who develop the search and navigation interfaces to get the most out of metadata. Note that, the simpler your metadata framework, the easier it will be to design search/navigation applications.

Metadata content

- Metadata should not be guesswork; it should match or reflect information which is accurately and clearly presented in the resource itself.
- In my opinion, trying to automate metadata content is risky and often leads to poor quality metadata. In HealthInsite we have found that this is particularly the case with dates. However, you may be able to use defaults for some elements (for example, publisher). Just make sure that your system allows you to check and change the defaults.
- It is important to have a good title and description – these elements often get displayed in search results and can entice (or discourage) a user to follow the link.
- There are various tools to support metadata creation – see the end of this paper.
- You do not need to create metadata for every page on your site – it is better to be selective.

How to store/display metadata

- The simplest method is to create and store the metadata record as a block of HTML code embedded in the source code of the resource. This can be updated whenever the resource is updated.
- However, if the site is very large, then it can be tedious to make changes affecting a lot of records.
- As sites move towards content management systems, then there is a tendency to store metadata as database fields – these can be automatically embedded into the source code for resources. Or there may be other ways to display it to gateway site harvesters. Syntax options are HTML, XHTML, XML, XML/RDF.
- It can get quite complex integrating metadata into a content management system. You need to work closely with your system developer to get what you want with the minimum of complexity.
- You need to establish a workflow so that updating is done accurately.

Where to get help

- DC-ANZ is a regional organisation for Dublin Core in Australia and New Zealand: <http://www.dc-anz.org>
- Their site has lots of links, including to the MetaMatters metadata resource site. This in turn links to the full AGLS standard and background information

on the National Archives of Australia site. (So you do not need to purchase the standard while you are just investigating.)

- The AGLS Working Group meets twice a year. As a member, I will be happy to provide advice or raise issues at the meetings. (See my contact details at the top of this paper.)
- The HealthInsite metadata specification is at <http://www.healthinsite.gov.au/metadata.cfm>
- I recommend getting librarians or indexers involved in the planning and implementation of your metadata system.