

Archiving conference report



Kerry Webb

kerry.webb@alianet.alia.org.au

*... cute animations
went out with
the millennium
celebrations, and
selecting a domain
name with an obscure
commercial title is not
really best practice.*

Last November, I was one of the fortunate group at the National Library's 'Archiving Web Resources' conference. It was a stimulating few days, with a broad range of local and international experts and a diverse collection of attendees. We were treated to some great presentations on best practice and good planning from a notable band of speakers. Thanks to a couple of months' hard work from Diana Dack, the conference report is now online at <http://www.nla.gov.au/webarchiving/>. It's a good summary of what was discussed at the conference, and by extension a fair exposition on the state of play in an area that most people don't give enough attention to.

Hey there, good lookin'

It's a good sign when you go to a website about design ... and it's well designed. That was my first impression of 'Designing Libraries — Library Buildings Online' at <http://www.designinglibraries.org.uk/>. Its main purpose is to provide a database of descriptions and images of recent UK public library building projects — although they are also interested in receiving submissions from other countries. The database is small, but growing, and features a number of useful links, including one to the US AIA/ALA Library Buildings Award site.

Bigger isn't better

The Australian Government Information Management Office (AGIMO) — the entity formerly known as NOIE and now speaking to you from the Department of Finance and Administration, thus completing the circular route that it took from its roots in the mid-1990s — contracted CSIRO last year to survey Australian Government websites. The results warrant careful analysis, showing that the size of the average resource was 61 KB, with the average size of a Word document being 147 KB and the average PDF a sizeable 284 KB. These are much higher than you'd find on the web in general (where resources average 15 KB) and of the US Government at 17.7 KB. The implications are that downloading information from government websites is going to take longer than you'd expect — especially as broadband access isn't as widespread as in some other countries.

Trim those files

So, how do we address this situation? Breaking documents into smaller pieces is one way, but if that's not feasible, you could try to optimise the size of

your PDFs, at least. Andy King, in Web Pro News, sets out a few ideas in his article at <http://www.webpronews.com/ebusiness/seo/wpn-4-20050401HowtoOptimizePDFFilesforWebSites.html>.

Science alerts

Users of the US Government's science.gov site can now get personalised e-mail alerts about their areas of interest. The site has been around for a couple of years now, with a pretty good search facility over the government's science databases, but this new feature lets the user specify what they're interested in and then enter their e-mail address so that a weekly report matching their needs can be mailed to them. The site is at <http://www.science.gov/>.

What a friend we have in LC

The Library of Congress has announced the publication of the final report for the AMeGA (Automatic Metadata Generation Applications) project. AMeGA had three main goals: to evaluate the current automatic metadata generation applications; to survey metadata professionals to find out which aspects of metadata generation are most amenable to automation; and to report on recommended functionalities for automatic metadata generation applications. The final report, which you'll find at the project website at <http://ils.unc.edu/mrc/amega.htm>, is worth reading (although at 1.5 MB it's bigger than even most Australian Government publications) but could be summarised as 'there's still a lot of work to do'.

And from our other friends, in the UK

The e-Government Unit's taxonomy working group has produced a bibliography on Taxonomies and Thesauri with specific emphasis on public sector applications (which is an appropriate market, as we don't see much interest from the private sector). It's a very good resource, with links to general background information, individual topics such as thesauri and controlled vocabularies, what various countries are doing, and software for thesaurus management. See it at http://www.govtalk.gov.uk/schemasstandards/metadata_document.asp?docnum=940.

and more on taxonomies

There's a bit of a discussion on the role of taxonomies in an article at the e-Content Institute by Patricia Beelby and Marcel Roy, in which they examine their roles in document management systems, and especially now that 'everyone' is creating documents

Softlink web based library management software.
Makes life a browse!

and knowledge in an organisation and it's harder to find resources in the absence of some standard way of describing the context of a document. In many cases, a search engine alone can't do the job. Decide for yourself at http://www.econtentinstitute.org/issues/ISarticle.asp?is=161566&story_id=63142111836&issue=03012005&PC/.

Children, don't try this at work

Obfuscations (<http://www.obfuscations.com>) is a fun site that shows you how to replace simple words with ones that take up a lot more space. Now, I can't think of many uses for this (apart from columnists on a deadline who need to find that extra 750 words), but you can get a lot of enjoyment from it. 'Blat blat/Onyx ruminant/Do you possess any cylindrical fibers of keratin'. Now, that's poetry!

and at the other end of the erudition scale

The Urban Dictionary can give you an idea of what young people (and those who strive to be) are talking about. It relies a lot on contributions (like the Wikipedia) but you'll usually find that any bogus definitions are debunked quickly. So, in its own way, it has a sort of warrant. It's also a good guide to trends in slang, like the discussion on 'phat', pace Ian Thorpe. Bookmark it at <http://www.urbandictionary.com/>.

Metacrawlers reconsidered

It used to be that metacrawlers were quite popular, until the big guys became really big. But there are still some differences in the indexes of the big players like Google and Yahoo! that should make you think again about using sites like Dogpile and (yes) metacrawler. In his article at <http://www.isedb.com/news/article/1151/>, Robert McCourty makes a good case for looking beyond the usual suspects. He points out that, because of different ranking algorithms, you'll probably find that something that doesn't make it into the top 100 at Google may well be much higher in the list of search results at one of the other popular engines, so by amalgamating the list of results, you get a better chance of finding what you need.

Good news from Macquarie Street

The State Library of New South Wales has initiated a project to put a huge collection of Australian culture material online, with its plans to make available the entire contents of its 'DNA of Australian Culture' collection in digitised format on its website over several years. So far, you can see a copy of Joseph Banks' 1768-1771 *Endeavour* journal, architect Joern Utzon's original sketches of the Sydney Opera House and early black and white photographs from around the nation. There are a couple of strange com-

ments though in the announcement. One was that a Telstra spokesman described the service as one that would allow Australian children to do their homework on the bus to school by accessing the collection via their mobile phones — a comment no doubt made with a view to ramping up the company's stock price ahead of the forthcoming sale or possibly a submission of the silliest techno-quote of the new century, you take your choice. The other was the suggestion that the site represented world's best practice in the area. Well, cute animations went out with the millennium celebrations, and selecting a domain name with an obscure commercial title is not really best practice. But apart from those quibbles, it's a very good start. Set yourself for exploration at <http://www.atmitchell.com>.

Do you see?

One of the strange features of the World Wide Web is that there's a lot of it that you can't see, or that it's quite difficult to find. The Invisible Web Directory site at <http://www.invisible-web.net/> is a companion to the book *The Invisible Web: finding hidden internet resources search engines can't see*. It includes a directory of some very useful resources on the Invisible Web, starting with eighteen major categories (from Art and Architecture to Transportation), each of which is further subdivided. The actual entries feature a short description, a link to a search facility for that site and (where appropriate) a link to a related resource.

... and further on sight

One of the more intriguing areas of usability research is in tracking where the subject's eye moves when they're visiting a web page. Web designers need to take note of the results of this sort of testing to make sure that they have the most important elements in the right place. The article from Search Engine Watch at <http://searchenginewatch.com/searchday/article.php/349086/> sets out some of the lessons learned from eye tracking studies. One of the companies offering such facilities is Eyetools, where Greg Edwards has a weblog called Eyetools Research (<http://blog.eyetools.net/>) where he describes the research and conclusions drawn from looking at a wide variety of websites. In one example, they inserted gibberish into a client's 'visual dead zone', but the results were noticed by only 1 in 25 people. You can benefit from this research by checking out their report of a homepage redesign using Eyetools eye tracking (4 pages, 905 K, images of a site before/after, plus Eyetools heatmaps) at http://www.eyetools.com/casestudy/CaseStudy-Web_Site_Redesign_20050303.pdf. ■

Web designers need to take note of the results of this sort of testing to make sure that they have the most important elements in the right place.



To find out more – www.softlink.com.au

or call 1800 777 037