

Metadata based search and browse functionality on the NSW Office of Fair Trading intranet: a case study

**Maxine Armitage
Office of Fair Trading
NSW Department of Commerce**

Summary

The NSW Office of Fair Trading launched its first intranet in June 2003. At the very beginning of the intranet project we recognised that unless users could find information easily the intranet would not succeed. We also understood that different people prefer to find information in different ways. To maximise the chances of searchers finding relevant information, and to provide flexibility in search options, we developed and implemented metadata driven search and browse functions. This case study describes the standards, tools and technology we used and how metadata was manipulated to retrieve information in a number of different ways.

Background

Before the creation of the intranet in June 2003 the NSW Office of Fair Trading accessed its corporate information from a network drive via the Windows Start Menu. The charter for the intranet was to migrate information from the network drive to the Teamsite content management system (CMS) and make it available through a web interface. We realised that there would be little business benefit to the Office if we simply recreated the old file structure on the CMS and failed to provide efficient methods of retrieving information – after all, that was the status quo.

We recognised that we needed standards and tools to allow us to describe, manipulate and retrieve information. As no standards and tools were operational in the Office at that time, or indeed operational in any other NSW Government agency that we were aware of, we had to identify and create our own.

As a NSW Government agency we wanted to ensure that we implemented NSW Government recommended standards where possible. We were aware of the importance of sharing information across government agencies, and indeed within our own agency.

The challenge was to define standards and to create tools to be implemented not just on the intranet but potentially across many sites. A subset of our standard is currently being applied to the Office of Fair Trading public website and the on-line Customer Assistance System and will be soon assessed for applicability as part of the records keeping metadata set.

Standards, tools & technology

Metadata standard

AGLS was adopted as the metadata standard. AGLS is a set of descriptive elements which government agencies can use to improve the visibility and

accessibility of their services and information in web environments. It has been mandated for use by Australian Commonwealth Government agencies and recommended for use by NSW Government agencies.

We adopted the following elements: 'creator', 'publisher', 'title', 'subject', 'description', 'language', 'date', 'type', 'format' and 'identifier'. Authority lists supporting the 'language', 'type' and 'format' elements were largely taken as prescribed in the NSW Office of Information and Communications Technology (OICT) draft AGLS standard. An authority list is a list of the terms that can be used to populate a given metadata element. The metadata set was checked for AGLS compliance by an external agency.

Corporate thesaurus

A corporate thesaurus of over 4000 terms was developed in-house for use behind the metadata Subject element. The thesaurus is hierarchical, i.e., it uses broader and narrower terms, as well as related and non-preferred terms. The Thesaurus was constructed in accordance with ISO2788 (Guidelines for the establishment and development of monolingual thesauri) and was checked for logical integrity by external experts.

Technology

The corporate thesaurus is constructed using Hierarch Version 2. We use Teamsite (v.5.5) as our content management system and to manage metadata authority lists including the thesaurus. Metadata is assigned using a simplified version of Metatagger. Searching and browsing is made available through Verity K2 V4.5.1.

Implementation

Metadata is assigned centrally by the Intranet Management Team.

Metadata is applied to all formats of electronic resources, not just to HTML pages. Metadata for HTML pages is embedded while metadata for other formats is exported to Verity in a "bulk insert file".

For all formats the URI and format metadata is automatically derived. For HTML pages the title is also automatically derived through the workflow process.

The metadata is manipulated in a number of different ways to allow individuals to find information in ways most suited to them. The following search and browse options are available:

1. The simple search

The simple search executes a search across the 'subject', 'title' and 'description' metadata fields as well as the text of HTML pages and attachments. If the search term has a synonymous term within the corporate thesaurus then this term will also be searched.

Search results are weighted primarily according to if and where a search term is found in the metadata. This forces the most relevant results to appear at the top of the list. As illustrated in Figure 1. the search results display the 'title', 'description', 'date', 'type' and 'format' metadata fields.

The left-hand sidebar in Figure 1 shows how the thesaurus is used to facilitate refined searching. If a search term matches a thesaurus term, or a synonym of a thesaurus term, then a hierarchy of broader, narrower and related terms will be displayed. Clicking on these terms executes another search.

Figure 1: The search results page



2. The advanced search

The advanced search allows the searcher to specify a subject, author, title, year and type. Any combination of these fields can be used in an advanced search. See Figure 2. Search results are displayed as shown in Figure 1.

Figure 2: The advanced search box

Subject: any terms ▼

Title: any terms ▼

Author: any terms ▼

Type: Select more than one type by holding down the CTRL button when making a selection.

Year: or from: to

3. The site index

Figure 3 shows the site index, an alphabetical list of corporate thesaurus terms that have been allocated to information resources. Additions and deletions to the list occur automatically with the updating of information resources to the intranet. Clicking on a term will automatically execute a search of that term and the results are displayed as in Figure 1.

Figure 3: The site index

Site Index

[A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z]

A

Aboriginal and Torres Strait Islander consumers	adoption leave
Aboriginal and Torres Strait Islander customer services	Advanced Reporting internet Enquiry System
Aboriginal and Torres Strait Islander customers	advisory bodies
Aboriginal and Torres Strait Islander Network (NSW)	AGLS
Aboriginal and Torres Strait Islander peoples	agreements
Aboriginal and Torres Strait Islander personnel	air travel
abstracts	annual reports
access and equity	archives
account codes	asset management
accounting	assets
	Australia Post
	Australian Securities Commission On Time
	award rates of pay

4. Indexes of forms and policies

By combining Teamsite functionality, AGLS metadata and administrative metadata we are able to generate index pages for types of information. We currently generate index pages for forms and for policies. Figure 4 below illustrates how the forms have been classified by Teamsite folders (Access & Equity, Administration, etc) and how they are displayed using 'title' and 'format' metadata. Resources can also be grouped under headings specified in the Teamsite index template and lists can be compiled from resources from a number of different folders.

An element in the administrative metadata is used to select those resources we want to index. An information resource can be indexed as more than one type if more than one type is allocated in the metadata.

Figure 4: Index page of forms

Forms

[Access & Equity](#)
[Administration](#)
[Compliance & Standards](#)
[Finance](#)
[Human Resources](#)
[Information Management & Technology](#)
[Internal Audit](#)
[Marketing](#)
[Ministerial Liaison](#)
[Policy & Strategy](#)
[Privacy and Personal Information Act](#)
[Records Management](#)

Access & Equity

- [Equal employment opportunity \(EEO\) data collection \[doc 75 kb\]](#)
- [International Women's Day scholarship : application form : Springboard Women's Development Program \[doc 76 kb\]](#)
- [Nomination for 2003 CLAS examination \[doc 29 kb\]](#)
- [Spokeswomen's quarterly report form \[doc 47 kb\]](#)

Administration

- [227 Elizabeth Street, Sydney car park delivery / booking form \[doc 80 kb\]](#)
- [Application for additional / replacement motor vehicle \[dot 23 kb\]](#)
- [Application for temporary access pass : Enterprise House \[dot 28 kb\]](#)
- [Application for use of departmental vehicle \[dot 44 kb\]](#)
- [Application for use of private motor car or motor cycle on official business \[dot 40 kb\]](#)
- [Lift booking request 227 Elizabeth Street, Sydney \[doc 89 kb\]](#)
- [Request for couriers or express delivery \[dot 33 kb\]](#)

Monitoring user searching and browsing success

Search reports are generated daily and cumulated on a monthly basis. These reports show the number of searches performed, the most popular search terms, search terms which resulted in success and those which resulted in failure. To date we only have reliable data for the last three months and are still looking at how to interpret these reports and how we can use them to develop intranet content and improve search functions.

During even this limited period, however the reports have been extremely valuable in identifying problems such as some searches not being executed quite as expected and nil results being given to search terms which have been copied and pasted into the search box.

Conclusion

It is still early days for the intranet and a true picture of usage has not yet emerged. We are confident, however, in saying that:

- Assigning metadata centrally, rather than decentralising the process, was the correct decision. It became obvious to us that two people assigning metadata frequently results in higher quality metadata than many people assigning metadata infrequently. The usefulness of the search and browsing functions depends totally on good metadata content, so no matter how good our standards and tools are, without proficient metadata assigners the project would have failed.
- Assigning metadata is cost effective given the many ways in which the metadata is used. Consider the time and effort required to manually maintain a site index or to maintain separate lists of forms and policies. Furthermore, content managed manually is rarely up to date. Our lists are up to date as soon as the content is published.
- By weighting search results according to where, and if, corporate thesaurus terms appear in certain metadata elements, and by searching on synonymous terms, we have been able to present the searcher with results intelligently ranked for relevance and with all information resources which might be useful to them. This is a good search strategy and it is working.
- The integration of 'broader' and 'narrower' terms in the search results (drawn from the corporate thesaurus) provides a powerful new way to assist users to find required information.

The NSW Office of Fair Trading has made a tangible commitment to good information management practices. Undoubtedly we will need to refine our standards, tools, technologies and methodologies in response to increased user expectation but we are confident that we have established a sound basis on which to manage information within the Office.

Author

Maxine Armitage
Information Coordinator
Office of Fair Trading
NSW Department of Commerce

Note

Maxine no longer works for this Department.
Her contact details are now:
Email: maxine@finditim.com.au
Phone: 0415178441

© Copyright 2004, NSW Office of Fair Trading. This article may be freely distributed as is. Permission must be gained before reprinting the article in any other format or media. All other rights are reserved.