

Guidelines for using subject metadata and controlled vocabularies for resource discovery

Sarah Eastwell and Christina Weeden, Department of Primary Industries and Fisheries, Queensland.

November 2004

Preface

In February 2004, Access Queensland approved and funded a project to develop guidelines for using controlled vocabularies to support the integrated delivery of information across the Queensland government.

Implementing subject metadata and controlled vocabularies is a specialist area that requires detailed and practical understanding. Everyone running an information management project must consider subject metadata and yet there is little practical advice available to government staff in this specialist area.

These guidelines aim to provide such advice. The advice goes beyond existing AGLS implementation guidelines but is related to them. Read these guidelines in conjunction with the AGLS implementation guidelines.

Acknowledgements

These guidelines have been produced with expert advice and feedback from:

The Search and Metadata Special Interest Group (SMSIG) of Queensland Government
National Library of Australia
National Archives of Australia
Distributed Systems Technology Centre (DSTC)
Department of Commerce, Office of Information and Communications Technology, New South Wales
State Library of Tasmania

Many references have been used in the writing of these guidelines. They are included in **Section 6 Appendix: Further reading**.

Contents

1	Introduction.....	4
1.1	Purpose.....	4
1.2	Audience	4
1.3	Scope.....	4
2	Subject metadata, controlled vocabularies and resource discovery	5
2.1	Subject metadata and resource discovery	5
2.2	Function metadata and recordkeeping	6
2.3	Other metadata elements.....	7
2.4	High quality content is essential	8
3	Determining controlled vocabulary specifications.....	8
3.1	What is the difference between a controlled vocabulary, thesaurus, taxonomy, classification scheme and ontology?.....	8
3.2	Building or using existing controlled vocabularies	9
3.3	Controlled vocabulary construction.....	10
3.4	Ongoing controlled vocabulary management and maintenance	11
3.5	Research and analysis	12
4	Examples.....	13
4.1	Site index and controlled vocabulary displays.....	14
4.2	‘More like this’	17
4.3	Navigation structures	19
4.4	Breadcrumb navigation.....	21
4.5	Defining topics for news, feature columns, subscription and other services....	24
4.6	Refined searching.....	28
4.7	Improved searching.....	30
4.8	Portals	33
5	Developing a business case	36
5.1	Benefits	36
5.2	Costs.....	37
5.3	Success factors	37
5.4	Change management.....	38
5.5	Models for subject metadata authoring.....	39
5.6	Technology	40
6	Appendix: Further reading.....	41

1 Introduction

1.1 Purpose

These guidelines provide practical advice on using subject metadata and controlled vocabularies to enhance resource discovery (information retrieval or finding information) on websites or databases with web-based interfaces.

The term *controlled vocabulary* in this document refers to any collection of words (often grouped by meaning) that are used for resource discovery. In resource discovery, *controlled vocabulary* is a broad term that can encompass *thesauri*, *taxonomies*, *classification schemes* and *ontologies*. For more information, see **Section 3.2 Controlled vocabulary types**.

These guidelines aim to improve decisions on information management by providing an overview of the relationships between subject metadata, controlled vocabularies, resources and resource discovery.

1.2 Audience

These guidelines are for Queensland government staff members who are responsible for implementing *Queensland Government Information Standard IS34 Metadata*, particularly those using the subject metadata element. This group includes staff members responsible for information resources (including datasets, recordkeeping and web-based information and services).

Readers should have a basic understanding of the role of metadata in resource discovery and be aware of *IS34* and the *AGLS Metadata Standard AS 5044*. (See **Appendix 6.1 Metadata implementation**)

Readers should also have a basic understanding of recordkeeping metadata requirements outlined in *Queensland Government Information Standard IS40 Recordkeeping*. (See **Appendix 6.14 Recordkeeping**)

1.3 Scope

These guidelines focus on using the **subject** metadata element and controlled vocabularies to improve resource discovery. This is not a comprehensive guide to implementing all metadata elements. These guidelines aim to provide detailed and practical advice on subject metadata beyond the level provided in existing AGLS implementation guidelines. Read these guidelines in conjunction with the AGLS implementation guidelines.

2 Subject metadata, controlled vocabularies and resource discovery

2.1 Subject metadata and resource discovery

What is subject metadata used for?

Subject metadata is used to assist resource discovery (information retrieval or finding information). We apply subject metadata to resources to describe the subject of the resource, or what the resource is about. Users can then discover or retrieve resources by subject. This makes the search results more useful and helps users organise their results. Use subject metadata to solve a specific and defined business problem. Only use it when the benefits are obvious .

What a controlled vocabulary does

A controlled vocabulary controls the terms used in the subject metadata element. The controlled vocabulary can then be used to generate subject-based navigation features on websites, and to improve searching.

You can apply more than one subject metadata element to a resource to achieve different discovery purposes. For example, a government agency web page might include subject metadata for discovery on the agency website, and a second subject metadata element for harvesting so the page appears in a subject portal.

Do I need to use subject metadata?

Any project that involves managing and discovering resources and implements metadata must involve making a decision about how subject metadata can be used. Make the decision by analysing the resources and considering what users need to discover. Consider whether users need to be able to retrieve resources by subject, and how this can be done in the retrieval interface. (See **Section 4 Examples** for suggestions.) The analysis will determine whether a controlled vocabulary is required and, if so, what its specifications are.

What do I do if I don't need subject metadata for my project?

If subject metadata (or function metadata – see below) does not play a central role in the resource discovery requirements for your project (at an agency or whole-of-government level), you might decide to make only minimal use of subject metadata. This level of use still complies with the *IS34* and *AGLS* standards.

Agencies may also decide to apply subject metadata to key resources only (such as website home pages) using a generic controlled vocabulary such as TAGS (Thesaurus of Australian Government Subjects). This makes the subject metadata available for use by others, but does not place a heavy burden on the metadata authors.

Even if you decide to make minimal use of subject metadata, you may still decide to apply other metadata elements such as Title and Description to all of your resources. The

Queensland Government AGLS Element Implementation Standard outlines the mandatory metadata requirements for Queensland Government. (See **Appendix 6.1 Metadata implementation**)

2.2 Function metadata and recordkeeping

How similar are subject and function metadata?

The subject and function metadata elements are similar in several ways. They can both be populated from a controlled vocabulary, and we use both of them to retrieve resources.

Subject and function metadata differ, however, in what else they do. Function metadata relates to the function that the resource plays in the organisation; subject metadata relates to the theme or topic of the resource. The National Archives website explains it well:

There is often confusion about the difference between function-based and subject-based classification. The easiest way to appreciate the difference is to think of functional classification as describing why a resource exists, while subject classification describes what the resource is about.

(From <http://www.naa.gov.au/recordkeeping/control/rkms/title.html>, 3.1 Scheme type)

For example, a media release about farming barramundi could be classified with both functional and subject metadata such as:

Function: *Community relations – Media relations – Media releases*

Subject: *Business and industry – Primary industries – Fisheries and aquaculture*

Place the controlled vocabulary terms in the subject metadata element if the controlled vocabulary you using is subject-based. Use the function metadata element if the controlled vocabulary is function-based.

Queensland Government Agencies must use *both* subject and function metadata to comply with IS34. (This differs from the *AGLS Metadata Standard AS5044* which requires Agencies to use at least one of either subject or function metadata.) (See **Appendix 6.1 Metadata implementation)**

How do I know whether to use the subject or the function metadata element?

Whether you choose to base the resource discovery for your project around the function metadata element, the subject metadata element, or on a combination of the two depends on your discovery requirements. The examples in Section 4 may help you determine your requirements.

For record-keeping purposes, the National Archives recommends using a function-based classification scheme and a function element. This is because recordkeeping retrieval is based on the need to be able to retrieve resources based on the business function of the agency. In a recordkeeping system, the controlled vocabulary may be linked to a

retention and disposal schedule (the length of time a record must be kept, and the date it may be deleted). The retention and disposal schedules are linked to the business function of the resource, which is why functional metadata is preferred for recordkeeping.

However, the recent trend toward decentralised recordkeeping means that organisations are tending to move away from function-based classification schemes towards subject-based classification schemes. Staff members tend to find subject-based classification easier to use because it is designed to reflect how they think about the organisation's records. The need for a practical classification scheme that meets the needs of the users is paramount. The article *Rolling out a record management system* is useful reading on this subject. (See **Appendix 6.14 Recordkeeping**)

What difference does using a function-based controlled vocabulary make in determining the requirements?

Whether you are considering a subject-based or function-based controlled vocabulary, you will undertake the same process to determine the requirements. These guidelines focus on subject-based controlled vocabularies, but you could extend the ideas to the use of function-based controlled vocabularies.

Never classify more than you need to. The Queensland State Archives guidelines for using the Keyword AAA thesaurus for recordkeeping says a good rule to remember is, 'only classify as much as is needed to locate and control the file/record.'

Is there a recommended function-based controlled vocabulary for Queensland government?

Although no particular function-based controlled vocabulary is mandatory, there is a whole-of-State government licence for both Keyword AAA and the Keyword for Councils controlled vocabularies. Agencies can use these as the basis for developing their own controlled vocabularies.

The Qkey controlled vocabulary, based on Keyword AAA, has been developed for use with Electronic Document Records Management Systems (EDRMS) in the Queensland government.

If you do want to know more about using a functional controlled vocabulary, help is available. The National Archives has guidelines for agencies that wish to develop a functional controlled vocabulary for recordkeeping purposes (see **Appendix 6.14 Recordkeeping**).

2.3 Other metadata elements

What other metadata elements can I use in combination with subject metadata?

Subject metadata is often used **in combination** with other metadata elements to enhance resource discovery. Which metadata elements you choose depends on the resources in question and the project's resource discovery requirements. For example, users may want

to retrieve media releases by date, or retrieve maps by location. These searches would involve the date and geographic location metadata elements. For the sake of simplicity, these guidelines focus only on subject metadata.

2.4 High quality content is essential

Will using metadata improve the quality of my information?

Metadata and content management systems are not solutions for poor content. No amount of effort in applying metadata and developing resource discovery systems will compensate for poor information. Content should be business driven and developed with the end user in mind. All publications require quality editorial input. Editorial policy and guidelines must be in place to control the quality of content.

How content is developed affects its quality. Agencies must develop content with the goal of integrated service delivery in mind. Follow a consultative process to develop content that has a whole-of-government perspective. This is an important step in the delivery of integrated services. This process should help agencies identify gaps in current information and resolve any conflicting advice on topics (see **Appendix 6.11 Creating high quality content**).

3 Determining controlled vocabulary specifications

3.1 What is the difference between a controlled vocabulary, thesaurus, taxonomy, classification scheme and ontology?

Thesauri, taxonomies, classification schemes and ontologies are all, in essence, controlled lists of words used as tools for resource discovery. In this document, the term *controlled vocabulary* is used as a generic term for all of these tools. Each of these tools is slightly different in their practical application but the distinction between them is blurred. A controlled vocabulary can be used for both organising information (such as creating a navigation system for a portal) and for aiding subject-based retrieval (such as improving search engine results and generating browse features). A summary of the main types of controlled vocabularies is given in Table 1.

Table 1. The main types of controlled vocabularies

Type	Important features
Thesaurus	<p>Terms relate to the subject of the resource</p> <p>Formally organised so that the <i>a priori</i> relationships between concepts (for example, 'broader' and 'narrower') are made explicit</p> <p>Defined in the ISO2788 and ANSI Z39.19 standards</p>
Taxonomy	<p>In its most specific sense <i>taxonomy</i> refers to a hierarchical structure of terms used in classification. The Linnean taxonomy used to classify living</p>

	<p>organisms is a wellknown example.</p> <p>The term <i>taxonomy</i> is commonly used in a similar way to <i>controlled vocabulary</i>, that is, for a wide variety of schemes that have varying degrees of conformity to standards and principles of thesaurus and classification scheme construction.</p>
Classification scheme	<p>A logical system for the arrangement of knowledge. For example, Dewey Decimal Classification used for shelving materials in libraries</p> <p><i>Faceted classification</i> refers to the ability to combine non-hierarchical subject concepts, for example, the subject facets for a bottle of wine might include region, grape variety and price.</p>
Function-based classification	<p>Terms relate to the purpose or ‘function’ that the resource plays in the organisation. Used mostly for recordkeeping.</p> <p>For example, Australian Government Interactive Functions Thesaurus (AGIFT) and Keyword AAA</p>
Ontology	<p>Contains terms, their definitions and the specification of relationships among terms.</p> <p>The term is borrowed from philosophy, where ontology refers to the study of the nature of being and existence.</p>

What is an encoding scheme?

The AGLS Metadata standard uses the term *encoding scheme* for any controlled list used to populate a metadata element. This may be a controlled vocabulary in the case of subject metadata, an ISO standard for date formats, and so on (see **Appendix 6.4 Controlled vocabulary types**).

3.2 Building or using existing controlled vocabularies

Should I use an existing controlled vocabulary or build a new one?

Building a new controlled vocabulary is difficult and time-consuming. Wherever possible, you should use an existing controlled vocabulary. You need to conduct thorough research and analysis to find an existing controlled vocabulary that meets your needs, or one that could be modified to meet your needs. Evaluate existing controlled vocabularies by, for example, checking whether the terms are specific enough or too broad, and whether the terms match the discovery needs of your users.

Using an existing controlled vocabulary also facilitates interoperability between metadata systems, which is a big advantage. For example, libraries that use the Library of Congress Subject Headings for their subject metadata make it possible for libraries to integrate their library catalogues (assuming both libraries are using a standard record format such as MARC). Libraries can also purchase cataloguing records for resources that have already been catalogued by another library.

I want interoperability so how can I find out what controlled vocabularies are already being used?

The National Archives maintains a register of controlled vocabularies used in AGLS metadata implementations. You should register any controlled vocabulary you use in your metadata implementation with the National Archives (see **Appendix 6.4 Controlled vocabulary types**). Check the register at the National Archives to see if there is a controlled vocabulary already in use that meets your needs.

What are some examples of existing controlled vocabularies?

Many controlled vocabularies already exist that have been designed to meet particular needs. A well-known example is the Library of Congress Subject Headings (LCSH). The LCSH contains about a quarter of a million terms and is growing at a rate of about 8000 terms a year. The LCSH was originally designed to be used to apply subject headings for the Library of Congress collection. It has now been extended to be applicable for all general libraries.

However, the LCSH is less useful for specialised fields; specialised controlled vocabularies exist for this purpose, for example, the Medical Subject Headings (MeSH). Medicine is a specialised area, and the MeSH aids the retrieval of medical information by associated professionals with specialised information needs and knowledge.

Where can I find out more about existing controlled vocabularies?

The Web has a lot of information about controlled vocabularies. A good starting point is the website maintained by Michael Middleton at the Queensland University of Technology (see **Appendix 6.4 Controlled vocabulary types**).

3.3 Controlled vocabulary construction

How do I find the right controlled vocabulary?

You may not be able to find an existing controlled vocabulary that suits your needs, so you will have to construct one. However, whether designing a new controlled vocabulary or choosing an existing one, be aware that sorting information into subject categories is always a subjective process that depends on the perspective of the individual creating and assigning the categories.

Here is a story to illustrate this point.

A teacher gave a group of primary school students the task of sorting a list of animals into categories. The list included domestic, wild and aquatic animals. One child placed dogs, cats and crocodiles into the same category. The teacher was perplexed and asked the student why this was a logical category. The student replied that they were all animals that carried their babies in their mouths. The student was from the Torres Strait Islands, and the student's life experiences meant that this was a logical grouping.

How can I tell if the controlled vocabulary is 'right'?

To create a controlled vocabulary you must examine the information resources that need organising and predict the possible ways that users might wish to retrieve the information. No controlled vocabulary is perfect and a controlled vocabulary is never unbiased.

The design of the controlled vocabulary will be influenced by these factors:

- nature and extent of the information that needs to be categorised
- emphasis required on different aspects of the information resources (the creation of a controlled vocabulary is always political.)
- needs of the users retrieving the information and the mental models they have of the information (which are influenced by their knowledge and prior experiences).

Develop your controlled vocabulary in consultation with users to ensure that the resulting resource discovery solutions meet user requirements.

How can I approach constructing or maintaining a controlled vocabulary?

Some approaches could include:

- Collect the search terms used by the website users, the words and phrases relevant to the website and construct a hierarchy and synonyms list from those terms.
- Allow users to nominate keywords in their metadata when they aren't able to find thesaurus terms, then review the frequency of user terms as a way of finding new entries for the thesaurus.

The readings in the **Appendix 6.3 Controlled vocabulary construction** give more advice on this subject.

What are the standards for controlled vocabulary design?

There are two standards for thesaurus design: *ISO 2788 Guide to the establishment and development of monolingual thesauri* and *ANSI / NISO Z39.19 – 2003 Guidelines for the construction, format and management of monolingual thesauri*. (See **Appendix 6.3 Controlled vocabulary construction**.)

3.4 Ongoing controlled vocabulary management and maintenance

What are the impacts of the controlled vocabulary changing over time?

Any controlled vocabulary that is used to populate subject metadata elements will need ongoing management and maintenance as the resources and discovery requirements change. For example, as noted, the LCSH list is growing at a rate of about 8000 terms per year to accommodate new information. You do need to be aware of the effects of any change. Changes to your controlled vocabulary will impact on the subject metadata for particular resources, and may affect the website architecture. Specialists should be responsible for any controlled vocabulary and changes should only be made with

appropriate consultation. The issue of sustainability /maintenance /staff resources must be considered.

How do I develop documentation for controlled vocabulary management?

Accompany any controlled vocabulary developed, modified or chosen with documentation on implementation and processes, such as how new terms will be added and how terms may be disposed or amended. The documentation should also explain how specific or general the controlled vocabulary terms should be.

Use controlled vocabulary management software such as TermTree to help you manage and maintain your controlled vocabulary (see **Appendix 6.8 Technology**).

3.5 Research and analysis

Ask yourself these questions to determine your controlled vocabulary specifications.

Resource discovery issues

- What are the possible uses for subject metadata in improving searching and generating browse features? (See **Section 4 Examples**.)
- What are the existing discovery features, and how might these be affected by the implementation of subject metadata?
- Are there any alternative or complementary solutions to help resource discovery other than just using subject metadata? (See **Section 4 Examples**.)
- How will other metadata elements be used and what role will they play in resource discovery?
- Are multiple controlled vocabularies required for different discovery needs?
- What type of controlled vocabulary is required?

Business issues

- How can discovery be enhanced at both an agency and /or whole-of-government or even national or portal level?
- What interoperability is required with other projects?
- Is there an existing controlled vocabulary that will meet or nearly meet defined specifications? Is it necessary to build a new controlled vocabulary?
- What model should be used for metadata authoring? Should it be centralised or decentralised? What about quality control?
- What are the costs and benefits of adding subject metadata, as well as the cost of not using accurate subject metadata?
- How will the metadata and controlled vocabulary be maintained?

(See also **Appendix 6.7 Queensland government ICT and IM strategy**)

4 Examples

Use the examples in this section to develop ideas about how you can use subject metadata and controlled vocabularies to enhance resource discovery.

You can use a single controlled vocabulary to generate navigation features and to enhance searching, depending on your resource discovery requirements. For example, the Department of Transport and Regional Services (DOTARS) website uses a controlled vocabulary to generate multiple discovery features, including:

- the main navigation structure for the website or portal. (**4.3 Navigation structures and 4.8 Portals**)
- a site index. (**4.2 Site index and controlled vocabulary displays**)
- allowing users to browse the controlled vocabulary terms. (**4.2 Site index and controlled vocabulary displays**)
- allowing users to search a single *facet* or subject area of the website (**4.6 Limiting searches**).

See <http://www.dotars.gov.au/index.htm>

Conduct further research into how you can use subject metadata

There are many ways to integrate controlled vocabularies into a web interface. The following examples are not exhaustive or definitive. The examples are intended to introduce you to some possibilities, but you will need to research the design and implementation possibilities further.

The examples are presented from simple features through to more complex. The examples are only illustrations and including them neither infers any endorsement nor indication of best practice.

Consider alternative approaches to using subject metadata

The examples include alternative approaches to using subject metadata to enhance resource discovery. Consider these approaches in case they provide a cheaper, simpler or more effective solution for your discovery requirements.

Note on the format of the examples

For illustration purposes, the examples in this section show the metadata encoded as HTML. Metadata can be encoded and stored in a number of ways. You must consider the metadata 'architecture', that is, how the metadata will be stored, gathered and queried, as part of your project. For more information, see NSW Metadata Guideline, Section 4 AGLS implementation. <http://www.oict.nsw.gov.au/content/2.3.34.AGLS.asp>

4.1 Site index and controlled vocabulary displays

Discovery feature

An index is an alphabetical list of entry points to content. A site index may also show the relationships between various subjects, thus leading users to more specific or related topics. A to Z lists provide a simple discovery option to support other discovery features.

For users who know what they are looking for, site indexes can provide fast, precise access. Indexes also provide a list of topics that is easy to scan rather than a hierarchy that is more difficult to navigate. Wherever possible, offer a single long index rather than dividing the index into separate pages.

An A to Z subject listing can be easily constructed from a hierarchical controlled vocabulary. Choosing a letter displays that part of the index.

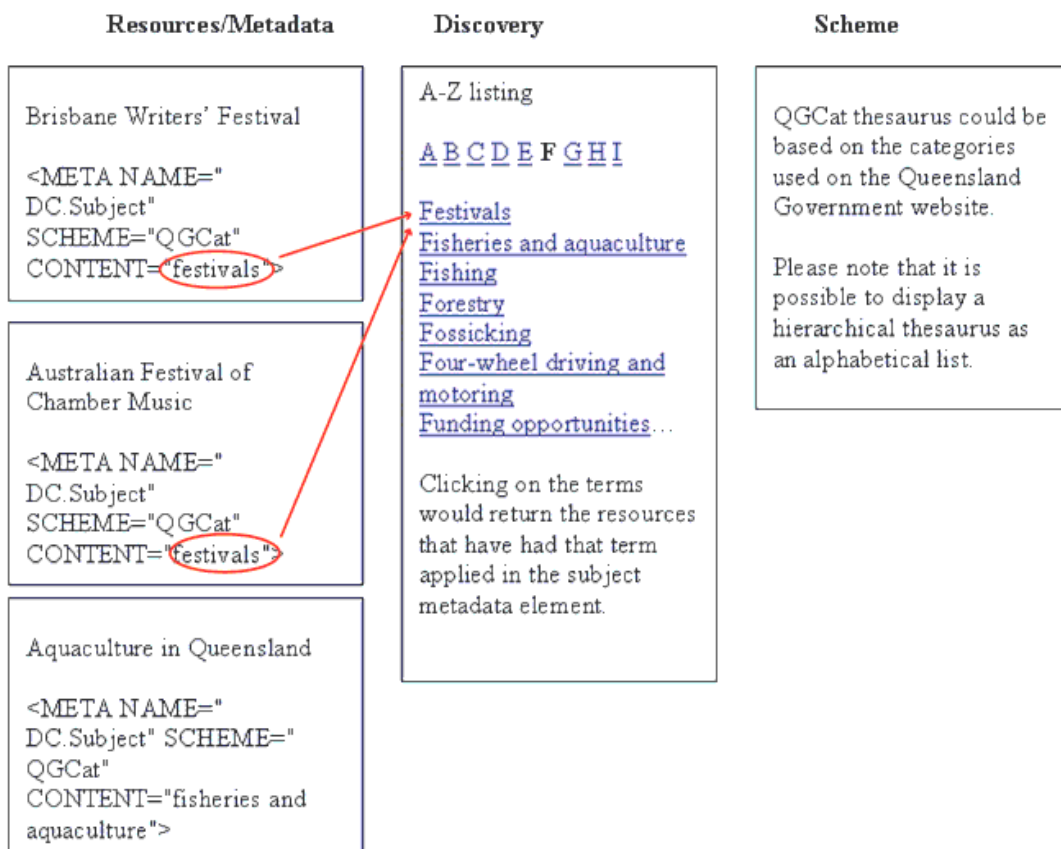


Figure 1

The following example illustrates a more sophisticated option for the controlled vocabulary display that makes use of broader and narrower terms (hierarchy information) and non-preferred terms ('use for').

ATRI – Australian Transport Index Thesaurus

Browse subject terms: [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#)...

Browse Term: P Term	Narrower Terms	Broader Terms
Paint		
Panel van... Use: Light commercial vehicle		
Parapet		Bridge deck
Paratransit	Car pooling	Public transport
Parent		
Park and ride		Transport facilities
Parking	No parking Off-street parking On-street parking Parking place Pay parking Service area	
Parking area... Use: Car park		
Parking garage... Use: Multi storey car park		
Parking infringement... Use: Offence		
Parking light		Vehicle lighting
Parking meter		
Parking place		Parking
Parking prohibited... Use: No parking		
Parking station... Use: Multi storey car park		
Partial cloverleaf... Use: Diamond interchange		

Figure 2

Controlled vocabulary considerations

A site index can be constructed from a controlled vocabulary used to populate the subject metadata element of resources. The controlled vocabulary may be displayed as an alphabetic listing and include 'use' or 'see' references from non-preferred terms and broader and narrower terms.

The terms must be specific enough so that there are not too many resources assigned to the term in order to limit results to a manageable set.

Websites that use this feature

Tasmania Online

<http://www.tas.gov.au/tasmaniaonline/SubjectIndex/>

Western Australia Online

<http://www.online.wa.gov.au/enhanced/atoz/>

Environmental Protection Agency

http://www.epa.qld.gov.au/site_information/site_index/

Alternative approaches

If a controlled vocabulary is not used, it is possible to create an index manually. However this process is time consuming and requires a large amount of maintenance. Software is available that can automatically create an index from the full text of the selected content, but for best results, the automatically created index must be checked and manually adjusted to make sense.

Searching the controlled vocabulary

You may wish to consider offering users the ability to search the controlled vocabulary. This allows the user to find a controlled vocabulary term and then to be able to find the resources grouped together under that term. This option is often offered in indexing and abstracting services. The advanced search function on the Australian agriculture and natural resources online (AANRO) database offers this option. Select the **keywords** option. See <http://www.aanro.net/page/advancedsearch.htm>

For more information read the article *Sitemaps and site indexes: What they are and why you should have them*. (See **Appendix 6.7 Website navigation design**)

4.2 'More like this'

Discovery feature

A 'more like this' feature helps users navigate from one document to other documents on the same subject matter. Any controlled vocabulary term used to describe a document can be displayed and linked in order to retrieve other documents using the same term. The controlled vocabulary terms may be displayed with search results, or included within the text of documents returned, perhaps in a footer area. The subject terms can be hyperlinked to retrieve other resources that include the term in their subject metadata. This feature is commonly used in library catalogues.

Resources/Metadata	Discovery
<p>Best friends, natural enemies</p> <p>Integrated Pest Management (IPM) of citrus has resulted in the reduction of pesticide usage by using natural and exotic biological control agents with no reduction in fruit quality...</p> <pre data-bbox="284 1160 823 1335"><META NAME="DC.Subject" SCHEME="LCSH" CONTENT="Agricultural pests -- Integrated control; Citrus fruits -- Diseases and pests; Pests -- Biological control"></pre>	<p>Best friends, natural enemies</p> <p>Integrated Pest Management (IPM) of citrus has resulted in the reduction of pesticide usage by using natural and exotic biological control agents with no reduction in fruit quality...</p> <p>Subject: Agricultural pests -- Integrated control Citrus fruits -- Diseases and pests. Pests -- Biological control.</p>

Figure 3

Websites that use this feature

University of Queensland library catalogue
<http://www.library.uq.edu.au/iad/mainmenu.html>

Brisbane City Council library catalogue
<http://www.brisbane.qld.gov.au/uhtbin/webcat/>

Controlled vocabulary considerations

A controlled vocabulary for this feature does not need to be strictly hierarchical. It may be an alphabetical listing of subject terms. The terms must be specific enough to return documents that are closely related in subject matter, but not so specific that there are only one or two items using a particular subject term.

This approach works best for a very large resource set. The controlled vocabulary will require ongoing management and quality assurance for applying terms to resources to ensure high quality results. A centralised model for applying metadata may be needed to ensure quality because applying detailed subject metadata is complex and requires significant training.

Alternative approaches

ISI cited reference related records

The ISI Web of Knowledge database provides a *related records* link that returns a list of articles that cite at least one of the articles cited by the original article. Articles that cite the largest number of articles in common with the original article are listed first.

The rationale behind this is explained by ISI:

The assumption behind related record searching is that articles whose reference lists include some of the same sources have a subject relationship, regardless of whether their titles, abstracts, or keywords contain the same terms. The more cited references two articles have in common, the closer this subject relationship is. Related Records are sorted from most relevant to least relevant.

<http://www.isinet.com/demos/webofscience/unique.html>

This functionality works extremely well, but is highly specific for citation databases.

Amazon.com 'similar items' links

Amazon.com generates links to similar items by analysing the selection patterns of previous customers. The link reads *Customers who bought this title also bought*.

See <http://amazon.com/>

4.3 Navigation structures

Discovery feature

Controlled vocabularies can be used to automatically generate navigation structures such as browse trees. The advantage of this approach is that it reduces the need for maintenance and hand-linking of resources. You can easily access items from more than one place in a hierarchy by assigning more than one metadata term.

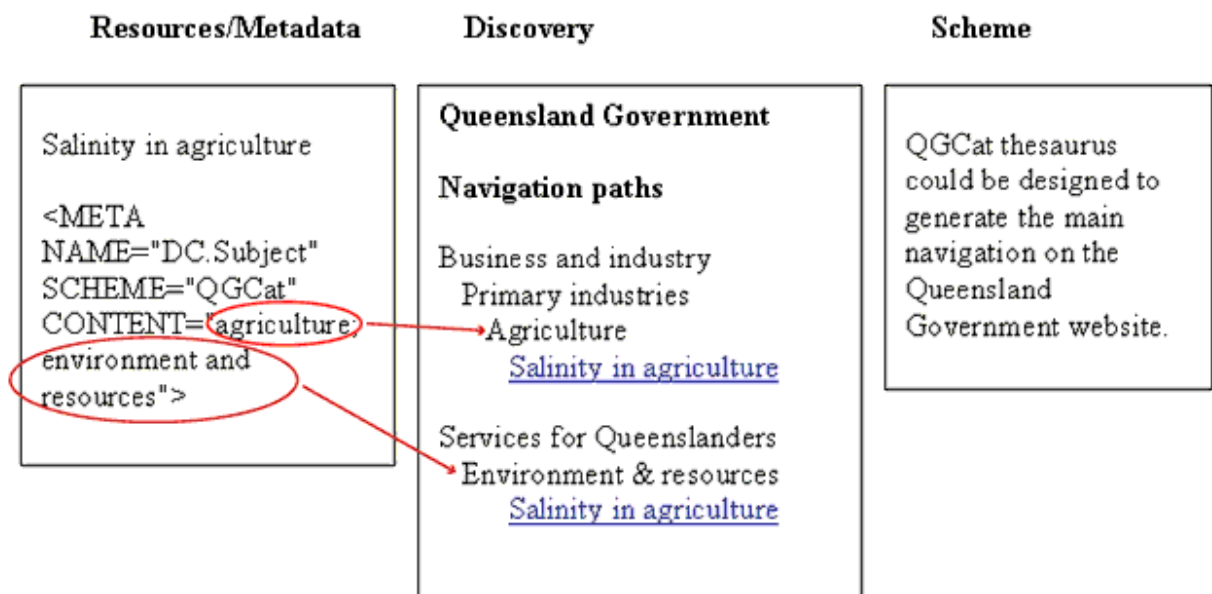


Figure 4

Content considerations

This approach is more likely to be effective with resources that can stand alone or tell a complete story. The information needs to be able to be read as a complete package without needing supporting or contextual information. News releases are good examples of this kind of content. In contrast, individual web pages often need to be read in the context of the website to make sense. You will probably find that automatic generation of menus works effectively only with higher-level web pages (home pages) rather than individual web pages. You need to apply editorial control to create effective navigation structures at lower levels. You must examine the content you are working with to decide how effective automatic generation will be.

Controlled vocabulary considerations

You need a hierarchical controlled vocabulary for this feature.

Technology considerations

DSTC's MetaSaurus is a tool that generates web page navigation from thesaurus files.
<http://www.dstc.edu.au/Products/index.html>

Alternative approaches

Often site structures are automatically generated from the folder structure of the website. Many content management systems offer this function. The disadvantage of this approach is that content cannot sit easily in more than one place in the hierarchy, and the metadata is not available for reuse in other features. Also, this approach does not allow information from across different websites to be integrated into a single website or portal.

4.4 Breadcrumb navigation

Discovery feature

In web design, a breadcrumb or breadcrumb trail is the part of the navigation that shows you where you are, as depicted in the Grimm Brothers' fairy tale *Hansel and Gretel*. Breadcrumb trails are often found near the top of web pages and define both the current location within the site hierarchy as well as primary pages above the current page. (From <http://webdesign.about.com/cs/webdesign/g/bldefbreadcrumb.htm>)

There are three main kinds of breadcrumb navigation:

1. Location breadcrumbs – a page **always** has the **same** breadcrumb, no matter how users get there
2. Path breadcrumbs – a page has **different** breadcrumbs, depending on how users get there
3. Attribute breadcrumbs – a page has **many** breadcrumbs, representing several possible locations

(For further explanation, see **Appendix 6.12 Website navigation design**.)

A controlled vocabulary can be used to generate location or attribute breadcrumbs by displaying the hierarchy for the subject metadata terms applied to a page.

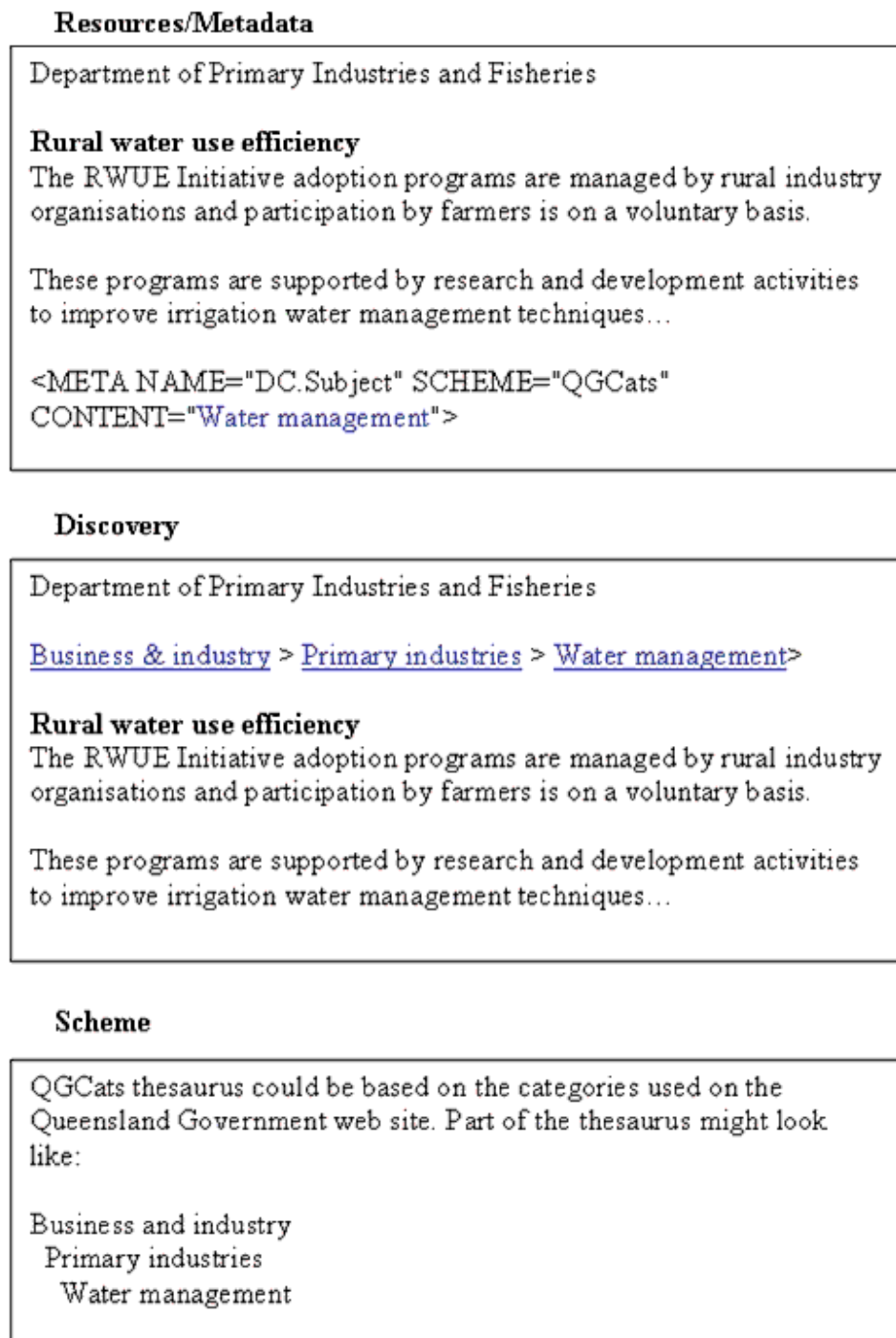


Figure 5

Controlled vocabulary considerations

A hierarchical controlled vocabulary can be used to generate breadcrumbs. You need to evaluate the usefulness of the breadcrumb navigation and the terms used in the controlled vocabulary. Breadcrumb navigation is not necessarily used to allow users to get *back*, but to allow users to explore related categories.

Alternative approaches

You can program path breadcrumbs to be generated automatically from the navigation path taken by the user as they browse the website. Breadcrumb navigation can also be generated from the folder structure of the website. Many content management systems offer this functionality.

Websites that use this feature

Some websites that use breadcrumb navigation include:

Ebay

<http://www.ebay.com/>

Yellow pages (browse categories feature)

<http://www.yellowpages.com.au/>

Amazon

<http://www.amazon.com/>

4.5 Defining topics for news, feature columns, subscription and other services

Discovery feature

Subject metadata can be used to stream information into subject areas to support services such as news subscription services, feature columns, events calendars and so on.

Here is an example. Subject metadata could be used to aggregate media releases from across agencies into subject areas, providing an integrated news service while allowing agencies to retain control of the publishing process. The media releases could then be divided into subject areas and emailed to subscribers, published as news feeds, or published as news feature columns on web pages.

Automatic generation based on metadata means that news or features columns can be updated regularly without the need for manual maintenance of links. Items may be tagged with more than one subject term, allowing content to be displayed in more than one subject area.

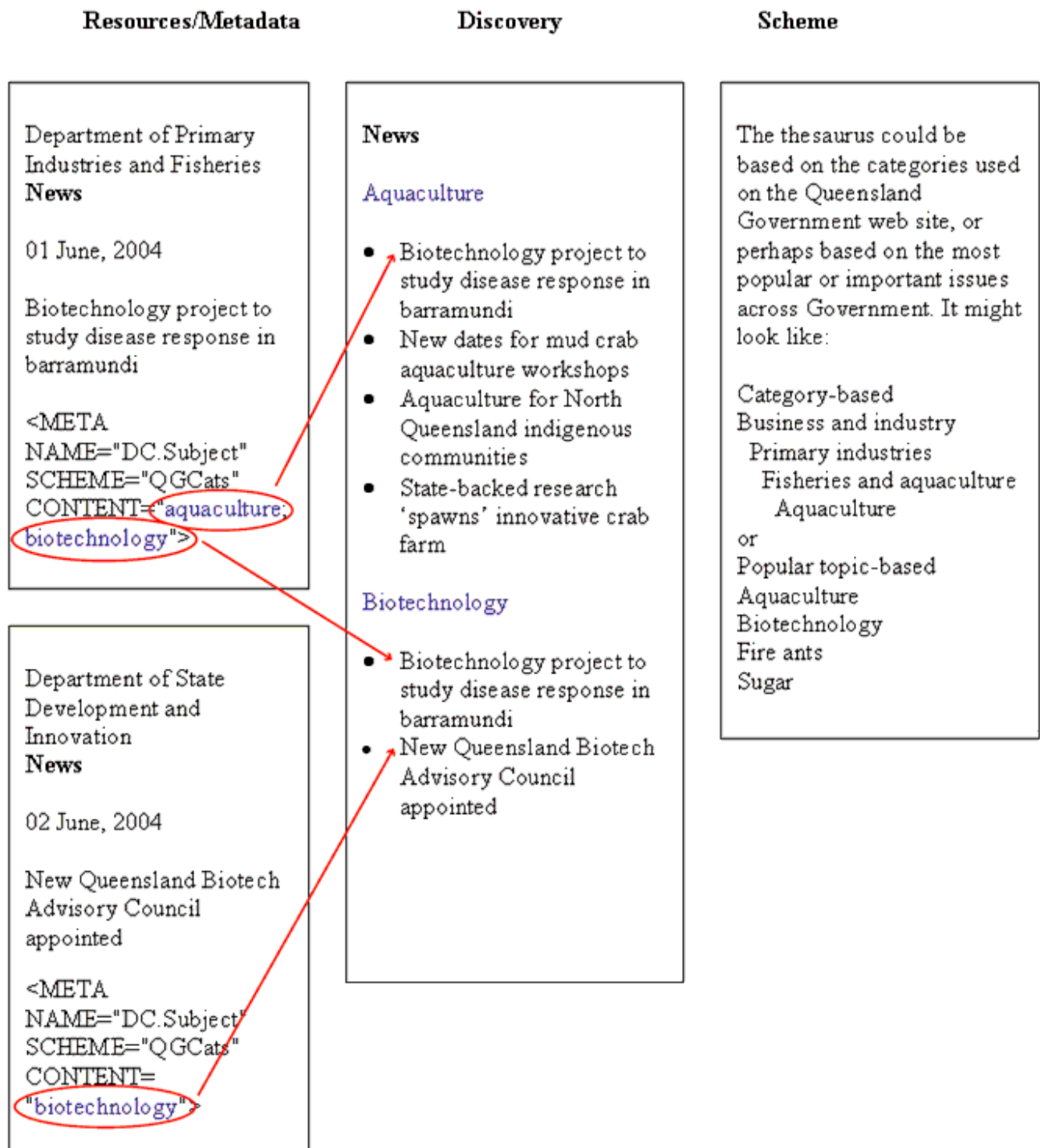


Figure 6

Content considerations

Aggregation and automated services are more likely to be effective with highly homogenous and stand-alone information such as media releases, than with more heterogeneous or context-specific information such as general website content.

Other metadata elements

Other metadata elements such as date, location, publisher and priority would need to be considered to provide the functionality required by the service. For example, date metadata may be used to specify start and expiry dates for display of items. Title and description metadata may be used for generating the links. Location metadata might be useful for sorting content by geographical regions.

Metadata authoring model

An appropriate model for metadata news releases could be a decentralised model in which agencies would be responsible for applying the metadata terms. The metadata could be stored with the news releases and harvested by software. This model is efficient for news releases because of the high volume of news releases published, the timely nature of the information, and the comparatively low impact of incorrect metadata. Centralised quality control and training would be necessary to ensure consistent high quality.

The most appropriate metadata authoring model would need to be considered on a case-by-case basis for each project.

Websites that use this feature

Education Network Australia (EdNA) news page

<http://www.edna.edu.au/edna/go/pid/2998>

Both the 'news' and 'recently added resources' columns are dynamically generated.

Technology considerations

You may wish to consider using RSS XML format for publishing news and other information for syndication because it is a popular Internet standard and there are many tools already written that support and use RSS. For an overview of RSS see:

Content feeds with RSS 2.0: Syndication goes mainstream

<http://www-106.ibm.com/developerworks/xml/library/x-rss20/>

Websites that use RSS include:

EdNA

<http://www.edna.edu.au/edna/go/pid/3153>

New Zealand Government

<http://www.govt.nz/en/news>

A standard for the publication of government news summaries (New Zealand)

<http://e-government.govt.nz/docs/rss%2Dv%2D1%2D0%2Dfinal/index.html>

Alternative approaches

An alternative approach would be to consider publishing news or subscription service notifications via an editor controlled process. In other words, it might be worth considering using the skills of a professional writer or editor to publish a regular subscription 'newsletter' or 'update'. This handcrafted approach is more time-consuming, but would ensure quality results and a more focused message.

You could also consider whether similar types of features might be able to be generated from other data, such as the most frequently viewed pages. For example, Amazon.com lists 'top sellers' as features on its home page.

See <http://amazon.com/>

4.6 Refined searching

Discovery feature

Subject metadata can be used to limit a search to a particular subject area. This technique will return more relevant results to the user.

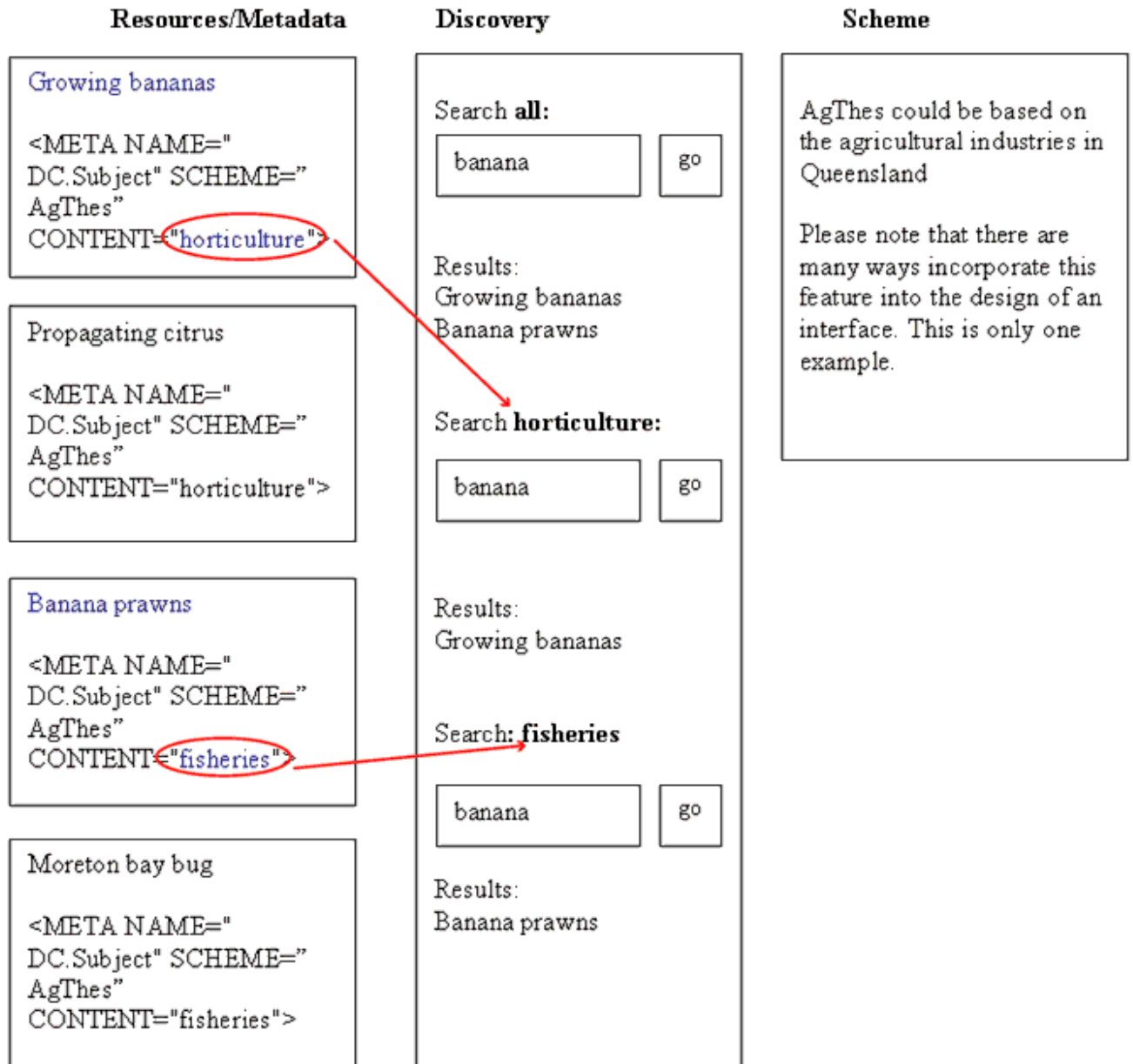


Figure 7

Other metadata elements

Please note that many metadata elements are commonly used to restrict the scope of a search, such as date, author and so on. You should consider using all metadata elements when designing the search interface. This example is concerned only with the use of subject metadata.

Websites that use this feature

Epicurious website

Users are able to search within a sub-category of the site. For example, users who browse to the recipes with chocolate as the main ingredient are presented with the option of searching only the chocolate recipes.

The advanced search for the Epicurious website allows users to restrict searches by a number of metadata subject facets (that is, subject areas). Users may limit the search by cuisine, special consideration, type of course, main ingredient, preparation method or occasion. It allows users to combine keyword searching with searching metadata.

http://eat.epicurious.com/recipes/enhanced_search/index.ssf/?/recipes/enhanced_search/index.html

Controlled vocabulary considerations

The terms assigned to documents need to be fairly broad. You could only achieve this by using the top levels of a hierarchical controlled vocabulary. Consider the number of possible terms in the context of the effectiveness of the search and the total number of documents in the collection.

Consider using other metadata elements such as date, author, publisher, and so on, when designing a useful interface.

Alternative approaches

Search engines can often be configured to restrict the search to a particular file folder on a website. If content is organised into folders according to subject, this approach might be sufficient. However, this solution does not offer as much flexibility as a metadata solution. The metadata approach allows more than one subject metadata term to be assigned to a document so the document can be found in multiple types of searches.

4.7 Improved searching

Discovery feature

Subject metadata can be used to enhance the full-text searching by suggesting alternative terms to users that may not appear in the text of the document, but that users might use to search for the document. For example, a user might use the search term *papayas*, but the documents include only the term *papaws*. The search engine could be configured to read the controlled vocabulary and suggest the preferred term to the user.

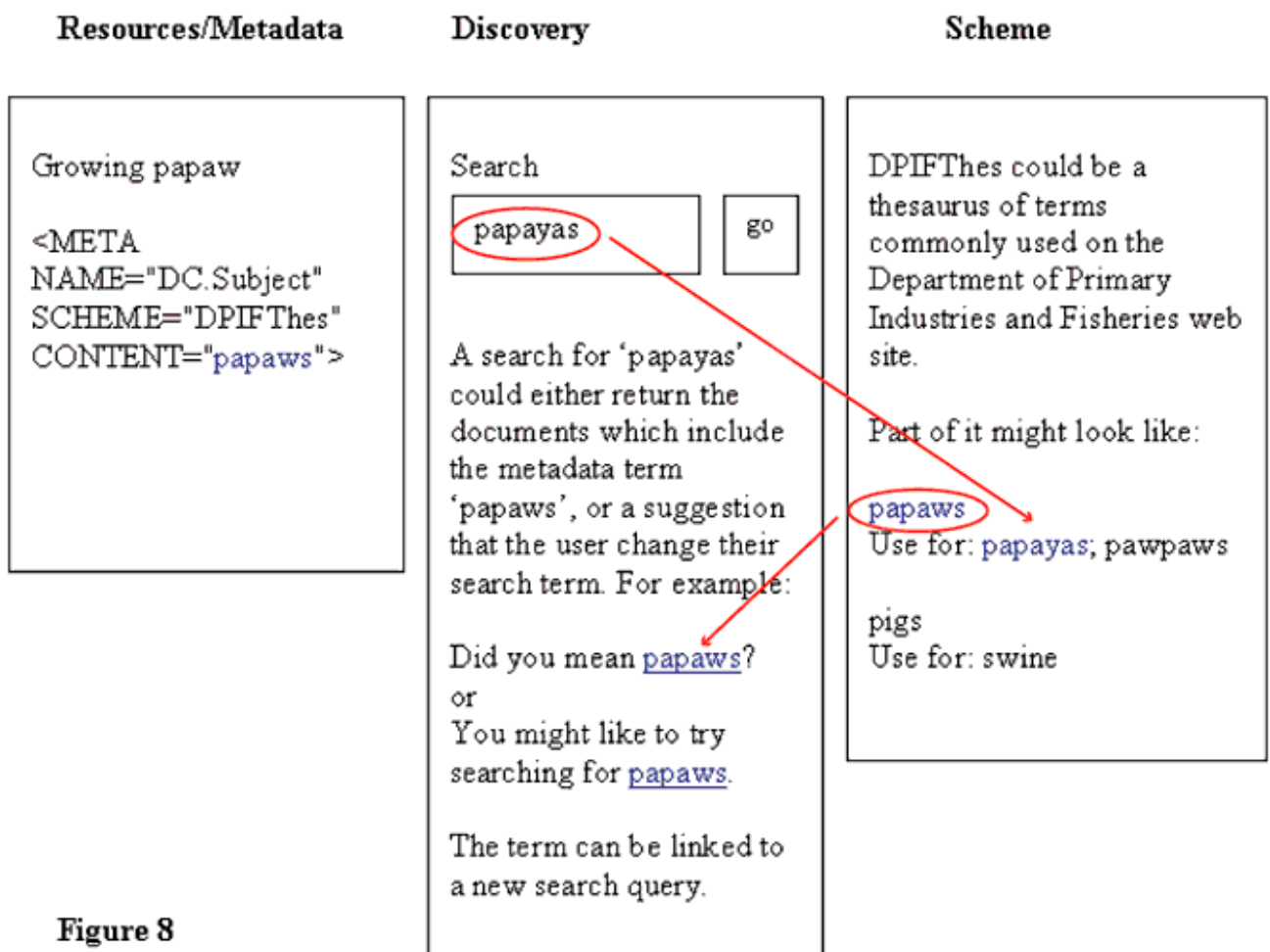


Figure 8

The controlled vocabulary could also be integrated into the search results in a more sophisticated manner by providing links to broader and narrower terms.

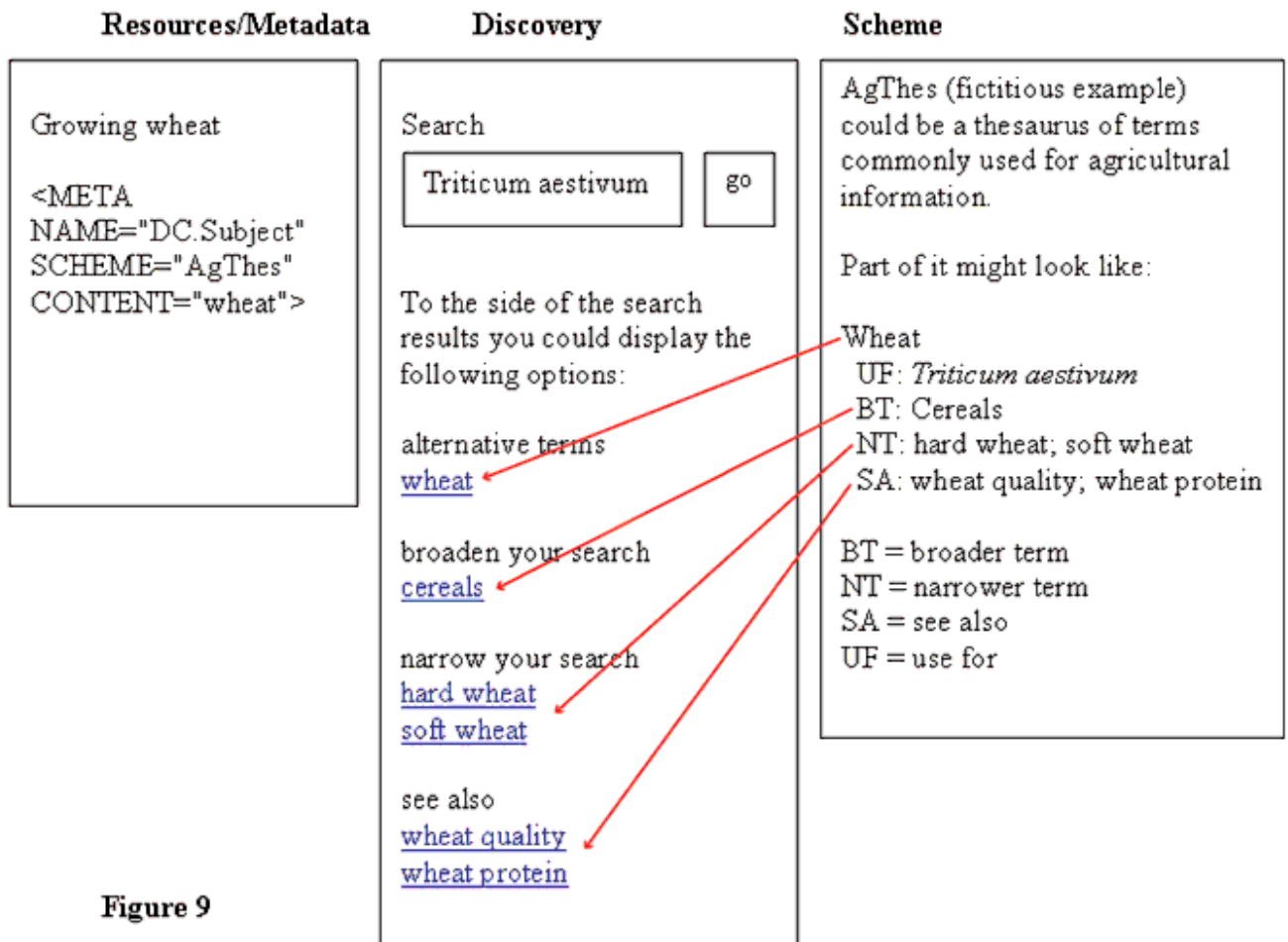


Figure 9

Controlled vocabulary considerations

The controlled vocabulary would need to include narrower, broader or 'preferred terms' or 'Use for' references.

Websites that use this feature

Department of Agriculture, Western Australia (Advanced search)
<http://www.agric.wa.gov.au/>

Alternative approaches

Free-text (uncontrolled) keywords may be sufficient for improving the search results. A controlled vocabulary may not be required. The keywords can be used to supplement the full-text indexing of the document. The search engine must be configured to include the metadata terms when the pages are indexed. Common misspellings can also be added to the metadata.

Subject metadata authors need to be given clear guidelines for the addition of keywords. Here are some examples.

- Add one to five keywords that relate directly to the content of your document.
- Separate keywords with a semicolon.
- You may wish to include alternative terms or spellings that people may try to use to find your document in addition to terms from the controlled vocabulary.

Make sure uncontrolled keywords undergo a central quality control process.

This approach would not have the benefit of being able to suggest broader or narrower terms.

Some search engines (such as Panoptic) can be configured to match search terms with particular documents. When users enter a term into the search engine, the administrator can specify the first item in the results list. This means that alternative terms can be matched to a particular document. This process is relatively resource intensive so would probably be carried out for only the most frequently used search terms (which can be identified by analysing the search logs).

Another approach is to use software that automatically generates broader and narrower term suggestions from the text of the documents. Guidebeam (<http://www.guidebeam.com/>) is an example of this type of software.

4.8 Portals

Description

Customer-focused portals provide access to resources located on the different participating sites. The main types of portal include those where the resources relate to a specific subject domain (for example, LawAccess Online <http://www.lawaccess.nsw.gov.au/>, Families portal <http://www.families.gov.au/>), particular audience type (for example, seniors) and government jurisdiction (for example, Queensland government, <http://www.qld.gov.au/>). Some portals combine these characteristics (for example, Tasmania Online, <http://www.tas.gov.au/>, provides Tasmanian coverage, audience-specific resources and cross-Tasmanian government agencies subject access).

The AGLS metadata elements that are particularly important to the delivery of browse lists and searches targeted to customer groups are Title, Subject, Description, Coverage, Function, Audience, Type and Availability. Participating sites must agree on how *portal-relevant* elements are populated and what metadata authoring model is implemented for effective discovery and consistent results display.

Diagram

The next diagram is modelled on the information available on the Australian government's Families portal (<http://www.families.gov.au/>). The diagram has been constructed to show how subject metadata and controlled vocabularies can be used to generate the portal automatically.

Resources

Criteria should be defined for the resources that are to be accessed via the portal.

Metadata element set

You need to identify the elements relevant to portal delivery. Establish the guidelines to populate these elements and any controlled vocabularies chosen or built.

The first occurrence of DC.Subject is for agency purposes and may be populated from the respective agency controlled vocabulary.

The second occurrence of DC.Subject must have at least one value and at least one of the other 'Fam' occurrences must have a value. Not all of the 'Fam' occurrences need values.

Controlled vocabularies

Each controlled vocabulary operates as the encoding scheme for a specific metadata element occurrence.

Discovery

A range of discovery features is available on the Families portal home page including a display of each 'Fam' controlled vocabulary, an A to Z listing of the Families controlled vocabulary and the site map showing all 'Fam' controlled vocabularies to two hierarchical levels.

Other discovery features are available from the home page. These rely on combining metadata elements.

Management

Portal development relies on a high level of collaboration and cooperation among participating organisations. Managing and maintaining portals incurs costs for the following aspects: quality controlling resource selection and metadata authoring, ongoing development and management of associated controlled vocabularies, maintaining relationships of participating organisations and monitoring and implementing relevant technology developments.

Alternative approaches

Project teams should consider whether the cost of establishing and maintaining a portal is justified in meeting business needs. Will relying on web search engines such as Google provide good enough results? Will configuring a search engine for particular purposes return good enough results? (For example, Queensland Government Website's Panoptic search engine trawls all Queensland Government agencies' websites)

Resources/Metadata

```
<META NAME="DC.Subject" SCHEME="AgencyThes" CONTENT="">
Portal meta tags
<META NAME="DC.Subject" SCHEME="Families" CONTENT="">
<META NAME="Fam1" SCHEME="Information" CONTENT="">
<META NAME="Fam2" SCHEME="ForChildren" CONTENT="">
<META NAME="Fam3" SCHEME="LifeEvents" CONTENT="">
<META NAME="Fam4" SCHEME="UsefulTools" CONTENT="">
<META NAME="HCLInclude" CONTENT="Y">
```

Discovery

Families		
Browse	Information	For children
Information
For children	Work	Living it
Life events	Youth	Homework..
Useful tools		
A – Z listing		
Site map	Useful tools	Life events

	Calculators	Starting a family
	Tax return	Starting school

Schemes

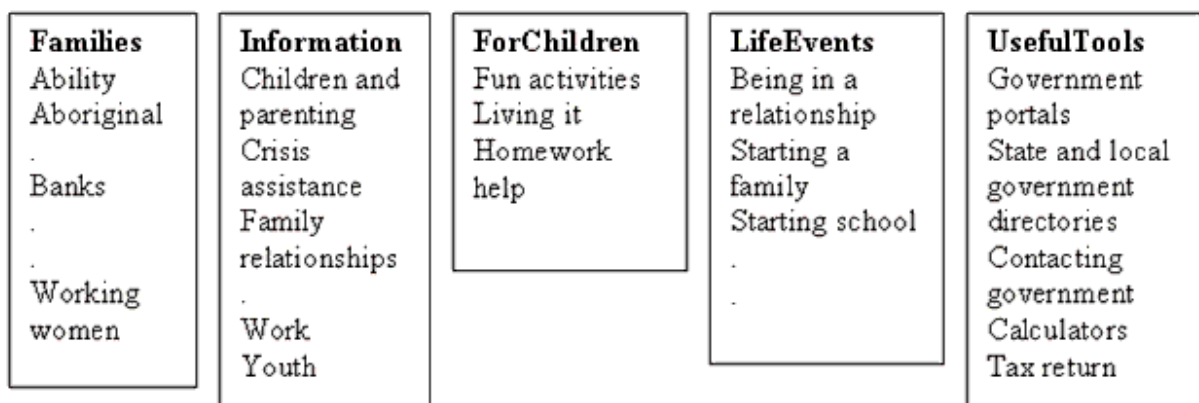


Figure 10

5 Developing a business case

A business case outlines the costs and benefits of different options. It is used to help make decisions about whether to allocate resources to a project. This section outlines some of the benefits and costs associated with implementing subject metadata. Subject metadata should be used to solve a specific and defined business problem. It should not be used unless clear benefits will arise from doing so.

5.1 Benefits

Document current situation

Before any new discovery solutions are implemented, document the pre-existing situation for resources/content, metadata implementation and discovery in order to provide a baseline from which to measure benefits.

Measure benefits

Measure the benefits of implementing subject metadata and controlled vocabularies. You can achieve this by identifying stakeholder needs, setting measurable objectives and designing well-crafted usability tests. Look at the McGovern Scorecard criteria and Jacob Nielsen user-testing techniques for useful ideas (see **Appendix 6.1 Business issues**).

Measurable benefits might include:

- additional navigation features to increase options for discovery
- enhanced search capability
- consistent resource discovery
- consistent implementation of *IS34*
- seamless discovery of resources regardless of type of resource and resource owner
- automation of agency website navigation and whole-of-government or subject-specific portals.

Measure benefits over the entire life of the controlled vocabulary

Benefits and costs should be applied to the whole of the life of any controlled vocabulary implementation, not just to the period of the project itself. Do not forget to assess maintenance costs, as well as the possibility of reducing future costs by re-using the controlled vocabulary for other purposes (see **Appendix 6.1 Business issues, NSW Metadata Guideline**).

5.2 Costs

What costs should I consider?

There are a several costs involved in implementing subject metadata and controlled vocabularies.

- **Personnel costs.** These may include a metadata specialist, a centralised tagging team and distributed metadata authors.
For example, the experience of Tasmania Online is that managing central extensive metadata authoring for approximately 5000 resources takes one full-time metadata specialist. An additional half-time specialist manages the associated strategic planning, policy and monitors the changing web environment for the Tasmanian Government. Metadata application is not a 'do it once' task because the task is undertaken in a continually changing environment. (pers. comm., Lloyd Sokvitne, Tasmania Online, 24 June 2004)
- **Training costs.** Include your ongoing training and communication requirements.
- **Technology costs.** Include content and metadata storage, authoring tools and reconfiguration of content management systems.
- **Resource costs.** These may be required for metadata changes to existing resources before managing metadata authoring for new resources.

Return on investment is difficult to quantify but the costs should be justified against benefits. Ask yourself this question, 'What is the cost of not using quality metadata?' The overhead for adding quality subject metadata to selected content should be worthwhile in terms of the investment in content and the access it provides to other content.

5.3 Success factors

What are the success factors?

A number of success factors are involved in the implementation of subject metadata and controlled vocabularies in Queensland government whether at agency or whole-of-government level. Some of the success factors are listed below.

- **Providing good resource discovery is not just a technology solution.** It involves attention to resources/content, metadata and discovery features.
- **Poor metadata is worse than no metadata!** Lack of ongoing controlled vocabulary management and maintenance will jeopardise the usefulness of the controlled vocabulary and impact resource discovery.
- **Employ staff with the right skill set.** Efficient and effective implementation of subject metadata will require a major effort requiring specialist resources.
- **Training is critical and must be ongoing.** Untrained staff members increase the likelihood of poor, inconsistent metadata.

- **High levels of collaboration, cooperation and commitment** are required for whole-of-government initiatives. Most agencies will need content and metadata reviews.
- **Good relationships and communication** provide seamless information delivery regardless of delivery channel. Call Centre and counter delivery staff leverage off web content and provide feedback on the relevance of web content.
- **Politics** influence information architecture on websites and thus influence any controlled vocabulary construction
- **Using controlled vocabularies** for automatically generating versus handcrafting discovery features on websites can decrease the likelihood of errors (such as broken links and spelling) and add to efficiency and effectiveness of web management.
- **A controlled vocabulary construction process** must take account of ‘top down, bottom up’ processes and whether it is best created automatically or through human effort, or a combination of both.
- **Agencies use different technologies** to deliver their websites. Seamless delivery via these technologies would be supported by consistent, quality metadata and appropriate content. Agencies should support interoperability.
- **The web environment is continually changing.** Topic maps, the Semantic Web and automatic categorisation are developments currently being researched. At present there are lots of promises but no tangible deliverables. These developments will complement metadata application but they are unlikely to take over in the short term.

5.4 Change management

Why should I plan for change management?

Training staff responsible for metadata management, authoring and quality control is essential to improving resource discovery. Subject metadata and controlled vocabulary implementation affects business areas at agency and whole-of-government levels, including:

- website management, especially content management and relationships across government
- nominating or recruiting metadata specialists and defining roles
- allocating resources for metadata authoring
- managing changes in workflows and practices
- training
- implementing technology.

(See **Appendix 6.1 Business issues, NSW Metadata Guideline.**)

5.5 Models for subject metadata authoring

Choose a metadata authoring model

There are two models for metadata authoring: centralised or distributed. The model you choose must take account of the quality of metadata, recruiting metadata specialists and providing initial and ongoing training in metadata application for all metadata authors. The model must also be appropriate to the business requirements for the project, and reflect the organisation's publishing model. You may choose to use a combination of the models for your project for different subject metadata occurrences. The models are summarised below.

Centralised metadata authoring

A centralised metadata authoring model uses metadata specialists to apply metadata and this model has the most likelihood of producing consistent, good quality metadata. You only need to train a select specialist group.

Distributed metadata authoring

A distributed metadata authoring model uses resource owners to apply metadata. This model presents a considerable risk to the consistency and quality of the metadata. The use of schema for the subject metadata tag helps to improve metadata quality through the use of a controlled vocabulary. Initial and ongoing training is critical to the effectiveness of this model, along with a central quality control process staffed by metadata specialists to check metadata. This model may be efficient and effective enough for general agency resources and for some whole-of-government initiatives, such as using subject metadata to present news releases in subject groupings. Quality audits of metadata entered using this model must be incorporated in workflow or site reviews.

What's so important about high-quality metadata and trained staff?

Applying **good** subject metadata is critical to resource discovery. Poor quality metadata may give worse results than no metadata!

Consistent use of controlled vocabulary terms is also important in achieving relevant, precise retrieval results. You should develop an indexing policy and procedures and ensure that the training teaches staff how to choose appropriate terms from any controlled vocabularies in use.

Consider integrating the controlled vocabulary into the metadata editor

A controlled vocabulary application may be integrated into an agency content management system so that a 'pick list of terms' can assist subject metadata application. One of the biggest stumbling blocks in correctly populating the subject element has been the lack of integration of controlled vocabularies within metadata editors.

5.6 Technology

Evaluate software

After the controlled vocabulary specifications and discovery solutions have been defined, you can evaluate the technology to support the metadata implementation. You may need to customise or build software to support your discovery needs if off-the-shelf software cannot meet requirements. If you are considering using alternatives to subject metadata then you should also evaluate alternative and complementary technologies. A starting point for investigating technology is given in **Appendix 6.1 Technology and controlled vocabulary construction**.

Subject metadata choices impact heavily on the possible search solutions used. The two should be considered together as your use of subject metadata may impact on your choice of software.

6 Appendix: Further reading

6.1 Metadata implementation

Standards

Queensland Government Information Standard IS34 Metadata. Retrieved 05 November, 2004, from <http://www.iie.qld.gov.au/site/informationstandards/current.asp>

Notes: These Guidelines outline the minimum standards relating to resource discovery with which Queensland Government agencies are expected to comply.

Queensland Government AGLS Element Implementation Standard. Retrieved 05 November, 2004, from

http://www.iie.qld.gov.au/site/informationstandards/toolbox/IS%2034%20v2_toolbox.asp

Notes: Outlines the mandatory metadata requirements for Queensland Government agencies.

AGLS Metadata Standard. Retrieved 05 November, 2004, from

http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html

Notes: The AGLS Metadata Standard is a set of 19 descriptive elements which government departments and agencies can use to improve the visibility and accessibility of their services and information over the Internet. It has been mandated for use by Commonwealth Government agencies.

Resources

AGLS Victoria: Metadata Implementation Manual. Retrieved 05 November, 2004, from <http://www.egov.vic.gov.au/Victoria/StrategiesPoliciesandReports/Reports/AGLS/aglsvic.htm>

Notes: The manual provides Victorian Government departmental and agency staff with practical advice and recommends approaches to the implementation and management of AGLS (Australia Government Locator Service) compliant metadata.

Australian Government Information Management Office (AGIMO). Better Practice Checklist: Use of Metadata for Web Resources. Retrieved 05 November, 2004, <http://www.agimo.gov.au/practice/delivery/checklists/metadata>

Notes: This checklist outlines a number of issues for consideration when managing metadata in an agency.

Meta Matters. Retrieved 05 November, 2004, from <http://metabrowser.dtdns.net/dcanz/>

Notes: Meta Matters is a collaboration between the National Library of Australia and Dublin Core ANZ group. Meta Matters was previously wholly developed and maintained by the NLA. Meta Matters is a 'put your money where your mouth is'

effort by the DC-ANZ group to show that the use of metadata in resource discovery is a fully working, viable method of resource discovery and dissemination.

National Library of Australia guidelines for subject metadata. Retrieved 05 November, 2004, from <http://www.nla.gov.au/guidelines/metaguide.html>
Notes: The Guidelines are intended to provide advice on determining the content for the Dublin Core/AGLS metadata elements that are most useful for resource discovery. The elements addressed are title, creator, subject, description, publisher, date, identifier and coverage.

NSW Metadata Guideline. Retrieved 05 November, 2004, from <http://www.oict.nsw.gov.au/content/2.3.34.AGLS.asp>
Notes: NSW has produced the AGLS Metadata Guideline to promote consistent implementation of the AGLS metadata standard by NSW agencies across Internet and non-Internet environments.

6.2 Using metadata

Moreville, P. (2002) Bottoms up: Designing complex, adaptive systems. New Architect December 2002 Retrieved 05 November, 2004, from <http://www.newarchitectmag.com/documents/s=7733/na1202b/index.html>
Notes: An excellent article which explains the complex processes involved in designing an effective metadata system for resource discovery.

NSW Office of Fair Trading, (2004) Metadata based search and browse functionality on the NSW Office of Fair Trading intranet: a case study. Retrieved 05 November, 2004, from <http://www.oict.nsw.gov.au/content/6.1.1.22.OFTINT.asp>
Notes: A good case study as it gives examples of using subject metadata in multiple ways. The intranet site uses metadata to weight search engine results, to offer pick-lists during search, broader and narrower term choices, a site index, and index pages.

Roberts, J. (2003) Describing services for a metadata-driven portal. Retrieved 05 November, 2004, from <http://www.e-government.govt.nz/docs/nzgls-for-portal/index.html>
Notes: This paper describes New Zealand e-government activities supporting the discovery of services through the use of Dublin Core-based New Zealand Government Locator Service (NZGLS) metadata. It notes the issues faced in collecting service metadata from agencies to populate a new whole-of-government portal. The paper then considers the adequacy of the metadata schema for service description, and identifies a difficulty in applying definitions

which refer to the content of the resource to a process-like resource such as a service.

Sokvitne, I. (2000). An evaluation of the effectiveness of current Dublin core metadata for retrieval. Retrieved 05 November, 2004, from

<http://www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF>

Notes: This article questions the value of the subject metadata element in information retrieval. It includes recommendations to test metadata searching against full-text retrieval. It discussed the need for a conceptual base and tools before applying subject metadata. It claims that quality metadata is required to improve information retrieval.

6.3 Controlled vocabulary construction

Standards

ANSI / NISO Z39.19 – 2003. Guidelines for the construction, format and management of monolingual thesauri. Retrieved 30 August, 2004, from

<http://www.niso.org/committees/MT-info.html>

ISO 2788:1986 Guidelines for the establishment and development of monolingual thesauri. *International Organization for Standardization*. Available for purchase from <http://www.iso.org/> Extracts available from

<http://www.collectionscanada.ca/iso/tc46sc9/standard/2788e.htm>

Resources

Aitchison, J., Bawden, D., Gilchrist, A. (2000) *Thesaurus Construction and Use: A Practical Manual*, 4th ed. ASLIB:London.

Maurer, D. (2003) Card-based classification evaluation, Boxes and arrows, Retrieved 05 November, 2004, from

http://www.boxesandarrows.com/archives/cardbased_classification_evaluation.php

Notes: For practical advice in testing classification systems.

Moreville, P. (2002) Bottoms up: Designing complex, adaptive systems. *New Architect* December 2002 Retrieved 05 November, 2004, from

<http://www.newarchitectmag.com/documents/s=7733/na1202b/index.html>

Notes: An excellent article which explains the complex processes involved in designing an effective metadata system for resource discovery.

Useit. Retrieved 05 November, 2004, from <http://useit.com/>

Notes: A good starting place for information on interface design.

W3C Web Ontology Language (OWL). Retrieved 05 November, 2004, from <http://www.w3.org/2004/OWL/>

Will, L. (2003) Comparison of thesaurus management software for PCs. Retrieved 05 November, 2004, from <http://www.willpower.demon.co.uk/thestabl.htm>

Will, L. (2004) Publications on thesaurus construction and use. Retrieved 05 November, 2004, from <http://www.willpower.demon.co.uk/thesbibl.htm>
Notes: Includes some references to facet analysis, taxonomies, ontologies, topic maps and related issues.

6.4 Controlled vocabulary types

Fast, K., Leise, F., & Steckel, M. (2002) All About Facets & Controlled Vocabularies, Boxes and Arrows. Retrieved 05 November, 2004, from http://www.boxesandarrows.com/archives/all_about_facets_controlled_vocabularies.php

Lider, B. & Mosiou, A. (2003) Building a metadata-based website. Retrieved 05 November, 2004, from http://www.boxesandarrows.com/archives/building_a_metadatabased_website.php

Louie, A.J., Maddox, E.L. & Washington, W. (2003) Using faceted classification to provide structure for information architecture. Paper presented at the ASIS&T 2003 Information Architecture Summit, Portland, Oregon, 21-23 March 2003. Retrieved 05 November, 2004, from http://depts.washington.edu/pettt/presentations/conf_2003/IASummit.pdf

Middleton, M. (2004) Controlled vocabularies. Retrieved 05 November, 2004, from http://sky.fit.qut.edu.au/~middletm/cont_voc.html

National Archives of Australia. Register of schemes used in AGLS metadata. Retrieved 05 November, 2004, from http://www.naa.gov.au/recordkeeping/gov_online/agls/register_schemes.html
Notes: In order to provide implementers of AGLS with information about the various schemes being used in AGLS metadata, the National Archives has compiled a register of the schemes being used. Schemes include encoding (structuring) schemes such as ISO8601, thesauri, and controlled vocabularies or lists.

6.5 Controlled vocabulary types examples to support information in table 1, Sec. 3.2

Dewey Decimal Classification. Retrieved 05 November, 2004, from <http://www.tnrplib.bc.ca/dewey.html>

Keyword AAA. Retrieved 05 November, 2004, from <http://www.records.nsw.gov.au/publicsector/rk/aaa/KeywordAAA.htm>

National Archives of Australia. AGIFT. Retrieved 05 November, 2004, from http://www.naa.gov.au/recordkeeping/gov_online/agift/summary.html

Scientific classification. Retrieved 05 November, 2004, from http://www.fact-index.com/s/sc/scientific_classification_1.html

TAGS, the Thesaurus of Australian Government Subjects. Retrieved 05 November, 2004, from <http://www.agimo.gov.au/services/tags>

6.6 Business issues

Jakob Nielsen (1994) Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. Retrieved 05 November, 2004, from http://www.useit.com/papers/guerrilla_hci.html

Jakob Nielsen (2000) Why You Only Need to Test With 5 Users, Jakob Nielsen's Alertbox, Retrieved 05 November, 2004, from <http://www.useit.com/alertbox/20000319.html>

McGovern Scorecard Retrieved 05 November, 2004, from <http://www.gerrymcgovern.com/scorecard.htm>

NSW Metadata Guideline. Retrieved 05 November, 2004, from <http://www.oict.nsw.gov.au/content/2.3.34.AGLS.asp>
Notes: NSW has produced the AGLS Metadata Guideline to promote consistent implementation of the AGLS metadata standard by NSW agencies across Internet and non-Internet environments.

6.7 Queensland government ICT and IM strategy

Agency information management strategies

Current information management (IM), knowledge management (KM) and information and communication technology (ICT) strategic plans should be consulted when planning controlled vocabulary developments so that duplication of effort is avoided and that developments are compatible and/or interoperable with existing systems.

Communication and Information Strategic Plan 1999 – 2004. Retrieved 05 November, 2004, from <http://www.iie.qld.gov.au/site/informationeconomy/stratplan.asp>
Notes: Sets out the Queensland Government's blueprint to achieve its vision of Queenslanders participating in the information age. The priority action area of e-government is the umbrella for controlled vocabulary development.

Information Economy, and Information and Communication Technology (ICT) Publications. Retrieved 05 November, 2004, from <http://www.iie.qld.gov.au/site/publications/comminfo/default.asp>

Information Standards and Guidelines Retrieved 05 November, 2004, from <http://www.iie.qld.gov.au/site/informationstandards/default.asp>
Notes: The Information Standards are the basis for achieving improved consistency and operations across Government. The standard of particular relevance to controlled vocabulary development is IS34 Metadata. Information is available on AGLS elements and populating the elements from NAA.

Internet gateway two year action plan prepared for Access Queensland Reference Group (2003) Retrieved 05 November, 2004, from http://ssq.govnet.qld.gov.au/Web/Resources/Folders/Web_Steering_Committee/Strategic%20directions/Gateway%202yr%20action%20plan.doc
Notes: Provides information on the direction for the Gateway. Agencies can use this information to enable alignment where possible.

Queensland Government Information Architecture (GIA) Mk1 (2001) Retrieved 05 November, 2004, from <http://www.iie.qld.gov.au/site/gia/about.asp>
Notes: The GIA is structured around a framework that partitions the communication and information resource environment. Within the GIA sub-architectures are developed in relation to information, application, technology and support services. These in turn govern a set of domains that provide a convenient mechanism to group and manage related activities and outputs of the architecture (the information standards). Controlled vocabulary development fits within the domain of Information Management / Information Portfolio.

Queensland Government Web Centre Resources. Retrieved 05 November, 2004, from http://ssq.govnet.qld.gov.au/web/resources/default.aspx?hFolderPath=Folders%5cContent_Management%5cImplementations%5cAccess+Queensland
Notes: Resources relevant to controlled vocabulary development include reports on syndication and aggregation, usability, CUE, etc.

Guidelines for using subject metadata and controlled vocabularies for resource discovery

6.8 Technology

Metasaurus. DSTC's Metasuite. Retrieved 05 November, 2004, from

<http://www.dstc.edu.au/Products/metaSuite/index.html>

National Archives of Australia. AGLS-compliant authoring tools. Retrieved 05 November, 2004, from

http://www.naa.gov.au/recordkeeping/gov_online/agls/tools.html

TermTree. Retrieved 05 November, 2004, from

<http://www.termtree.com.au/>

6.9 Alternative and complementary technologies

Guidebeam (Currently used on the Queensland Government website) Retrieved 05 November, 2004, from

<http://www.guidebeam.com/>

K-wise. Retrieved 05 November, 2004, from <http://www.k-wise.com/taxonomer.html>

Metatagger (Interwoven). Retrieved 05 November, 2004, from

http://www.interwoven.com/products/content_intelligence/index.html

Siderean Seamark. Retrieved 05 November, 2004, from

<http://www.siderean.com/products.jsp>

6.10 Automatic categorisation

Ainsbury, B. (2002) Cataloging's comeback: Classifying and organizing corporate documents. Online v.26 i2 p27(5)

Notes: An interesting article which has some good ideas like 'rule-based' categorisation, and pre-determined searching. Examples are given at the end of the article.

Kirk Lubbes, R. (2003) So you want to implement automatic categorization? Automatic categorization can be a powerful tool despite its limitations, but it is still important to test and evaluate before making a commitment to using it. (Records managers play important role in implementing automatic categorization) Information Management Journal v37 i2 p60(9)

Schewe, D. B. (2002) Classifying electronic documents: a new paradigm: the U.S. Department of Education set out to determine whether large volumes of electronic

data can be indexed cost-effectively. (LessonsLearned). Information Management Journal v36 i2 p54(5)

Notes: An example of automatic classification tools being used to categorise a large volume of information. This requires record managers to evaluate system results and train the system.

6.11 Creating high quality content

Contentious: News and musings on how we communicate in the online age. Retrieved 05 November, 2004 from <http://blog.contentious.com/archives/000186.html>

Editorial process in content management. Retrieved 05 November, 2004, from <http://www.nmpub.com/blog/archives/000098.html>

How to create an editorial process to publish web content. Retrieved 05 November, 2004, from <http://www.reachcustomersonline.com/content/2003/11/17/19.16.39/>

Jakob Nielsen's Alertbox, October 1, 2000: Content creation for average people. Retrieved 05 November, 2004, from <http://www.useit.com/alertbox/20001001.html>

New thinking: Content management strategies. Retrieved 05 November, 2004, from <http://www.gerrymcgovern.com/nt/class/cm.htm>

Why content management fails. Retrieved 05 November, 2004, from <http://www.adaptivepath.com/publications/essays/archives/000315.php>

6.12 Website navigation design

Boxes and Arrows (2004). Retrieved 05 November, 2004, from <http://www.boxesandarrows.com/>

Fox, C. (2003) Sitemaps and site indexes: What they are and why you should have them. Retrieved 05 November, 2003, from http://www.boxesandarrows.com/archives/sitemaps_and_site_indexes_what_they_are_and_why_you_should_have_them.php

Jacob Nielsen's Alertbox (2004). Retrieved 05 November, 2004, from <http://useit.com/>

Rosenfeld, L and Morville, P. (2002). Information Architecture for the World Wide Web, 2nd ed. Sebastopol: O'Reilly.

Three breadcrumbs overview. Retrieved 05 November, 2004, from <http://user-experience.org/uefiles/breadcrumbs/>

6.13 Conferences, training, associations

AGLS Training: Commonwealth Implementation of AGLS. National Archives of Australia. Retrieved 05 November, 2004, from http://www.naa.gov.au/recordkeeping/gov_online/agls/agls_training.html

Australian Society of Indexers. Retrieved 05 November, 2004, from <http://www.aussi.org/>
See the training page for training information.

Australian Library and Information Association. Retrieved 05 November, 2004, from <http://alia.org.au/>
The Australian Library and Information Association (ALIA) is the professional organisation for the Australian library and information services sector. It seeks to empower the profession in the development, promotion and delivery of quality library and information services to the nation, through leadership, advocacy and mutual support.

Master of Information Management (IT70), Queensland University of Technology (Faculty of Information Technology). [In 2004 this replaces the Graduate Diploma in Library and Information Studies (IT25)]. See the University Handbook, Retrieved 05 November, 2004, from <http://www.qut.edu.au/courses/>

Planning and building your taxonomy 18 – 19 October 2004, Sydney. Ark Group. Retrieved 05 November, 2004, from <http://www.ark-group.com/>

Records Management Association of Australasia. Retrieved 05 November, 2004, from <http://www.rmaa.com.au/>
The Records Management Association of Australasia (RMAA) offers a wide range of activities supporting the profession of records and information management. These activities include courses, conventions, seminars, specialised studies, projects, workshops, field trips, and publications.

Web Documents and Intranet Indexing Workshop
The University of New South Wales. Retrieved 05 November, 2004, from <http://cpd.sistm.unsw.edu.au/Indexing.html#IndexingWeb>
Indexing is all about access and websites and intranets are not useful unless they

are easily accessible. This workshop explores the skills of subject analysis, evaluation and annotation and their application to Web documents and websites. The concepts of metadata, including XML, and subject gateways are explored.

6.14 Recordkeeping

Developing a functions thesaurus (2003) National Archives of Australia, Retrieved 05 November, 2004, from http://www.naa.gov.au/recordkeeping/control/functions_thesaur/contents.html

Information Standard 40 Recordkeeping, Queensland Government, Retrieved 05 November, 2004, from <http://www.iie.qld.gov.au/site/informationstandards/current.asp>

Managing records of webpages and websites (2002) Queensland State Archives, Retrieved 05 November, 2004, from <http://www.archives.qld.gov.au/publications.asp#government>

Recordkeeping guidelines and tools, Queensland State Archives, Retrieved 05 November, 2004, from <http://www.archives.qld.gov.au/government/guidelines.asp>

Recordkeeping metadata standard for Commonwealth Agencies, National Archives of Australia. Retrieved 05 November, 2004, from <http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm>

Robertson, J. (2004) Rolling out a records management system. KM column, Step Two Designs. Retrieved 05 November, 2004, from <http://www.steptwo.com.au/index.html>

What is a "Functions thesaurus"? (2003) Willpower Information. Retrieved 05 November, 2004, from <http://www.willpower.demon.co.uk/functhes.htm>