The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1468-4527.htm

The IP's guide to the galaxy of portal planning: Part II. Content management

Rebecca H. Augustyniak, Dawn B. Aguero and Amy Finley
Clearinghouse for Applied Research and Public Service, Florida State University, Tallahassee, Florida, USA

Refereed article received
18 July 2005
Accepted for publication
31 July 2005

Abstract

Purpose – This article is the second in a four-part series that aims to illustrate the process involved in planning a portal and creating a portal definition document.

Design/methodology/approach – Based on a review of the academic and industry literature and using a case study, the authors share their experience in planning the Florida ExpertNet Research portal.

Findings – Portal development is a complex and costly endeavour that requires meticulous planning and design. As with any system development initiative, the time and energy spent in the planning process is reflected in the success or failure of the end product. Content management is the keystone of a portal. The components of a portal CM strategy include the content inventory and analysis, content acquisition, access structures, classifying content, content life cycle, CM software, and metadata standards.

Originality/value – This series will be useful to the information professional who is contemplating portal development and may be used as a model in developing a blueprint – the portal definition document. Whether the IP is considering doing development or outsourcing, it is important to understand the architectural requirements of a portal.

Keywords Content management, Portals, Project planning, Knowledge management, Information management

Paper type Case study

Introduction

This article is the second in a four-part series that illustrates the processes involved in planning a portal and creating a portal-definition document. The first article provided a general overview of portals and the specifics for defining a portal vision: the first component in a portal-definition document. This article will discuss the keystone of a portal (content management (CM). Incorporated in the series is a case study of the planning of the Florida ExpertNet Research Portal (Research Portal).

Portal CM is the post millennium challenge facing information professionals (IP) today. The success or failure of a portal is directly linked to the effectiveness of its CM strategy. Due to the explosion of available information via the web, the challenge for the IP who is developing a portal is not acquiring and storing voluminous amounts of content from various sources, but instead organising the information so that it can be delivered in a customised, personalised format. The goal of the portal is to provide end-users access to highly relevant data from a myriad of sources, simultaneously and in real-time.

Online Information Review
Vol. 29 No. 6, 2005
pp. 643-655

1468-4527
DOI 10.1108/14684520510638089

Since ancient times the librarian has had the challenging task of turning chaos into organisation. Throughout history, we have always been at the forefront in embracing and using new technologies to facilitate the improvement of information management (IM) and services. Taking terabytes of data and creating a classification and indexing system to turn chaos into organisation is what we do best. Unfortunately, a classification and indexing system alone will not provide a robust CM strategy.

Emery (2001) defines CM as the process and technologies associated with keeping information on a web site up-to-date and relevant. This involves the process for acquiring, storing, and classifying content so that it may be retrieved in an efficient and timely manner. CM "is a permanent ongoing process to control a never-ending stream of content – to file it, organise, and deliver it to the people who need it" (Melzer, 2005). The features of a CM system vary, but most include web-based publishing (content creation, review, and delivery), format management, revision control, archiving, indexing, search, and retrieval. A major misconception of a CM system is that it is just a piece of software. While software is an important component, by itself it does not create a CM system solution. Effective portal CM requires dedicated personnel who are highly skilled in IM.

This article will discuss and illustrate the components of a portal CM strategy, including the content inventory and analysis, content acquisition, access structures, classifying content, content life cycle, CM software, and metadata standards.

Content inventory and analysis

Providing the end-user with the ability to find the right information at the right time will ultimately determine the value of a portal. Therefore the portal will only provide as much value as the accessibility of its content. The successful retrieval of relevant content is the result of the CM strategy.

Identifying content-related issues in detail is the most labour intensive and critical step in defining the CM strategy best suited for your portal. The purpose of a content inventory is to determine what data exists, who owns it, where it is stored, and the current value. A content inventory will help identify gaps and shape your content plan as well as pinpoint duplicative data and data that is no longer of value.

According to Fraser (2001), there are three basic steps in the content inventory process. Table I defines the steps and purpose of each.

Fraser (2001) identifies categories of data to catalogue for each type of content identified:

- identification data (such as page title and URL);
- content data (which describes the page type and subject matter); and
- management data (which may include the content owner or producer);

She also suggests critical tasks that must be completed in preparation for the content inventory process:

- determining the type of information you will collect;
- creating data fields; and
- developing a database or spreadsheet application.

Steps	Description	Purpose
Survey	High-level review of content and sources	Aids in understanding the scope and nature of the material – type of content, topics it covers, types of sources, etc. Provides a clear understanding of the major pieces of content
Detailed audit	Comprehensive and in-depth inventory of each information type	Lists every content type by category; lists major attributes of the information that will eventually form part of the important metadata
Content map	Graphic representation of major content components grouped by user roles and business objectives	Provides visual, easy-to-understand overview of potential information architecture

Note: It is helpful to structure the survey as a miniature version of the detailed audit

Table I.
Content inventory process

Table II is a template organised by Fraser's categories that provides a detailed list of the data fields that should be captured or assigned for each item identified as part of the detailed audit.

After the inventory is complete, the analysis process can begin. An analysis helps establish content patterns and relationships. It also assists in identifying and understanding various types of content that need to be reconciled in the CM strategy. Fraser (2001) suggests as a means of mapping the content that one take all of the site's major content components, put them on a sticky note or index card, and cluster them according to user and business goals. Graphic charting and diagram software can be used for this activity as well to create a content map to use as a conceptual reference for architecture decisions. Such a map will facilitate building stronger relationships between content; identifying and eliminating duplications; and re-envisioning architecture with a view toward breaking out of content silos.

Content acquisition

Based on the content inventory and analysis, a content acquisition strategy can be developed. Content acquisition is the process of gathering information for the CM system. This process could be manual, automated, or a hybrid.

The content inventory and analysis will reveal content not available from internal and current sources. Additional sources must be identified to provide the missing content. These may include external sources or may need to be created. Each source type may require a different acquisition strategy.

The format of the content will affect the acquisition strategy as well. The content may be provided in a readily usable form (electronic file, metadata); otherwise the content may require extensive processing and restructuring.

Table III provides commonly used strategies for acquiring content. It is not uncommon to use a variety of acquisition strategies simultaneously.

Data field	Definition
Identification	
Audit ID	A numbering system assigned by the auditor to assist in referencing categories and content
Title	Title of the content item
Content	
Category type	The type of information defined by a category. A controlled list should be defined for consistency in assigning terms
Topics, keywords	Defining the content by using assigned "keywords" meta tag (see source code) or a controlled taxonomy
Content type	Describes how the content is formatted for distribution (forms, press release, white papers, database, spreadsheet)
File type	The type of file the content is stored in (MS Word, PDF, MS Excel, MS Access, etc.)
Management	
Content owner	Generally, the department or individual responsible for the creation and maintenance of the content
User type	Intended audience defined by a controlled list
Location	Where the content is located (URLs, shared directory, file drawer, hard drive of an individual's computer)
Update frequency	How often the content is scheduled for updating
Date created	When the content was created
Status	How the content is listed (currently available, in process, planned, not applicable, outdated, archived, etc.)
Publication target date	For the "in process" or planned content
Archive frequency	Frequency for archiving content
File size	File size in bytes
Notes	Additional information concerning the content that may be important

Table II.
Content inventory template

Strategy	Definition	Examples
Create	Developing content that is designed specifically for the portal	Publications Databases Applications
Purchase	Obtaining data for a fee	Subscription services (i.e. aggregate services, journals, newspapers, etc.) One-time purchase
Gather	Mining information available from various sources	Spider search and retrieval Federated search engine Manual search Push technology services

Table III.
Acquisition strategies

Unlike a business enterprise portal, where the majority of content is provided internally, external key partners will provide the core content for the research portal. This content will be high-quality, structured data as opposed to the mass of unstructured, disparate content typical of a large corporate portal. The primary strategy, therefore, is to create content for the portal.

Access structures

Access structures are the means of organising content in order to find it easily and reliably (Boiko, 2002). Access structures need to be considered from both the perspective of the end-user as well as that of the portal management team. A portal generally requires multiple access structures, depending on tasks and users.

Some examples of access structures are tables of content, keyword indexes, subject descriptors, and hyperlinks. Boiko (2002) identifies four types of access structures: hierarchies, indexes, cross-references, and sequences:

(1) A hierarchy is a "system of phrases that classifies and sub-classifies information" Boiko (2002). Hierarchies include tables of content, outlines, and taxonomies.

(2) An index derives from concepts that are extracted from the content and then translated into the language of the end-user. Indexes include keyword lists and synonym lists.

(3) A cross-referencing system relates one set of content to another. Cross-references include hyperlinks, glossaries, and citations.

(4) A sequence is a system of the logical flow and access to information in a particular order. Sequences include chapters, pages, and instructions.

Table IV is an abbreviated illustration of how access structures can be integrated into the portal planning process.

Metadata standards

One of the purposes of a portal is to provide one-stop access by seamlessly integrating content from disparate and external systems. In order to do this, there needs to be a common language and common rules. In the world of portals, extensible markup language (XML) is the language standard and Dublin Core (DC) is the standard set of rules.

	Access structures				
Purpose	Descriptors (taxonomy)	Content type (taxonomy)	Subject index	Site map (outline)	Navigational menu (outline)
End-user					
Allow registered users to filter content based on interests	X		X		
Allow precise and accurate searching	X		X		
Provide ability to browse		X	X	X	X
Portal management team					
Allow management to filter content based on tasks and role	X				
Allow precise and accurate searching	X		X		
Classify and index content	X	X	X		
Provide publishing control		X			
Manage content acquisition/ collection/creation	X		X	X	

Table IV. Access structures planning matrix

XML is "a standard for marking up data so as to clearly indicate its structure, generally in a way that indicates the meaning of different parts of it rather than how they will be displayed." For more information, see http://scripts.sil.org/cms/scripts/page.php?site_id = nrsi&item_id = glossary#xml

XML allows different applications to exchange data by using a common language. Many applications are currently using this mark up language to facilitate data transfer and usage. For more information about XML, see www.w3.org/XML/

The DC Metadata Element Set is a "standard set of about 15 elements (title, creator, subject, etc.), with optional qualifiers, used to structure descriptive records and facilitate information sharing. Originally intended for use in describing web-based resources, it is now used also for describing physical collections in museums, libraries, archives, and other repositories." For more information, see http://mic.imtc.gatech.edu/public_portal/pub_catglos.htm

It is created by a group of individuals from around the world, representing many disciplines.

DC is the web's version of a MARC record. It is a standard way of cataloguing primarily electronic information available on the web. It describes information about a given item and facilitates data transfer and reuse across systems. DC also aids in reliable search and retrieval by allowing the end-user to search by specific metatags. For more information about DC, see: http://dublincore.org/

Classifying content

"Classification (taxonomy, categorisation) is to content as mapping is to geography" (McGovern, 2001). Classification is an essential tool for quick and efficient navigation and search. One of the largest hurdles in developing an effective portal that provides customised data delivery is an effective classification strategy. Successful development of a portal requires the marriage of information management and information technology (IT). Although librarians have been classifying information for centuries, many in the IT world see customised data-delivery as a revolutionary concept. IT has traditionally approached data management through automated processes alone. Because quality indexing cannot be accomplished without some human mediation, there is no comprehensive automated software solution. However, there are various approaches to tackling this task with different outcomes. The basic dichotomy is between manual and automatic classification strategies.

When determining a classification strategy, you must determine what scheme you will use as well as how to apply it. Table V illustrates some standard strategies for developing and applying a classification scheme along with the advantages and disadvantages of each.

Many organisations choose auto-categorising software to take on the tasks of developing and applying a classification scheme as an efficient and low-cost means-to-an-end. These tools can screen a large number of items in multiple formats, and then with various approaches, automatically sort items into taxonomy-like categories and attach these terms as metatags.

An additional approach is to use taxonomy to classify items. A taxonomy is a means of categorising content. It is generally hierarchical, sometimes compared to a filing system. A basic taxonomy should have definitions and synonyms. A more

Table V. Choosing and maintaining a classification scheme

Classification strategies	Advantages	Disadvantages	Developing	Applying
Automated				
Full-text indexing	Quick and inexpensive to implement	No term equivalency Unable to browse by category Information overload	X	X
Auto-categorising – automatically sorts items into subject categories and attaches the terms as metatags	Quick to apply Assists in developing categories and a taxonomy Ability to browse by category Short turnaround between acquisition and categorisation of content	Can be costly Middle ground for accuracy and relevancy No perfect software solution currently available		
Manual				
Developing or adapting a taxonomy	Ability to browse by category and terms Highly specialised and customised	Costly and time-consuming Long-term commitment to maintain Requires expertise	X	X
Reviewing content and applying appropriate terms that describe the content	High accuracy and relevancy of classification Efficient and effective access	Time-consuming and costly Difficult to keep up with demand Long turnaround time between acquiring and classifying of content Requires expertise	X	
Hybrid				
Using an auto-categorisation tool to identify categories that is then manually reviewed for inclusion in the taxonomy	Efficient and effective access Ability to browse by category and terms More cost-effective and quicker than fully manual More accurate and relevant than fully automated	Less accurate than fully manual More costly and time-consuming than fully automated Requires expertise	X	
Using auto-categorisation of content with a manual review for low-confidence items	Efficient and effective access More cost-effective and quicker than fully manual More accurate and relevant than fully automated	More costly and time-consuming than fully automated Requires expertise	X	

detailed one will also have related terms and scope notes. As a package this would be called a thesaurus.

Developing and maintaining taxonomy is arduous and labour intensive. An estimate of $100,000 has been quoted for a typical taxonomy implementation. Adapting a currently existing taxonomy to meet portal business objectives is an often-used time-and cost-saving strategy.

Using taxonomy for indexing traditionally requires a person to assign terms to describe an item. This strategy provides more accurate results than an automated process. It greatly improves content filtering and search precision and efficiency. In the enterprise portal environment, taxonomy allows the user to navigate or browse and may be portrayed as the table of contents, content directory, or a collection of channels or folders.

Automated classification processes can be augmented by manual strategies to increase both accuracy of classification and the precision and efficiency of filtering. This hybrid approach may include:

- Manual review of items with low-confidence accuracy in classification;
- Content mapping; and
- Recommendation engines (an automated recommendation tool to assist manual classification).

Merging manual and automated strategies for both taxonomy development and content classification is the most commonly used approach today. Automated software products are improving and becoming less costly. An auto-categorisation tool could help identify categories to be manually reviewed for inclusion in the taxonomy. The revised taxonomy could then be automatically applied to the content. Any items that fail to meet a particular confidence level or that seem to require a new taxonomy term would then go to a human for review.

Case study

In relation to the case study presented in this series, Florida ExpertNet uses two subject taxonomies. One is a simple two-tiered hierarchical list of research terms developed by InfoEd International. InfoEd is a system similar to Florida ExpertNet that also classifies expertise. Although lacking in depth, it has great breadth in covering a wide area of academic topics. Since the scope of Florida ExpertNet is relatively narrow and budgetary constraints prohibited in-house development from scratch, this InfoEd taxonomy was a good fit.

A Florida ExpertNet Speaker's Bureau was launched in 2003. Because the research descriptor taxonomy was too detailed for this purpose, it was determined that a taxonomy with broader categories (subject headings) was needed. After reviewing existing taxonomies used for speakers' bureaus and not finding an appropriate match, a taxonomy was developed in-house.

Moving to an enterprise portal has required a reassessment of the classification strategy. The plan for the Research Portal greatly increases the scope of the content as well as the purpose of classification. A challenge in planning for the portal was deciding how to classify information within the expanded scope. Should very fine-grained research descriptors be used, or broad subject-headings, or both, or something else? The need for a unified taxonomy with multiple tiers was identified.

The following are the issues addressed in developing the CM strategy for the research portal:

- *Purpose of classification* – expanded beyond searching to include customised content filtering;
- *Content to be classified* – expanded to include publications, news, awards, patents, events, e-forums, blogs, chats, and links;
- *End-users* – expanded to include registered users, creating the need for sophisticated content filtering; and
- *Classification selection and implementation* – expanded because existing taxonomies will not serve as the classification tool for portal content; a unified multitiered taxonomy is needed, which is applicable to all systems within the portal.

Content life cycle

The content life cycle is the path of the data, from its creation or inception up through the time it is either archived or deleted. It is imperative to map out an efficient workflow schematic for the entire life cycle. Defining the content life cycle facilitates modelling the path of the content, thereby providing insight into patterns of usage and workflow, and allowing a better understanding of what system applications and processes will need to occur (Figure 1).

Life cycle considerations should be taken into account for the various types of content that will reside within a portal. Due to the variety of content types, life cycle processes may vary. This is a general guide for defining life cycle steps for the Research Portal. Each step may involve multiple business processes. Table VI displays information regarding steps within the content life-cycle, which was extrapolated from White (2004):

Content management software

In the portal industry, the term "content management" typically refers to a sophisticated software-based system. "A full-featured content management system takes content from inception to publication and does so in a way that provides for maximum content accessibility and reuse and easy, timely, accurate maintenance of the content base" (Warren, 2001). Until recently, CM system technology was in its infancy. Because of the overwhelming demand, software companies are rising to the challenge of providing the tools necessary to effectively gain control of voluminous amounts of information in a variety of formats and sources. A CM system facilitates the management of content within a framework of business processes, workflows, and business governance (White, 2004). An integrated CM system marries the content life cycle with all of the tasks associated with each step of the life cycle.

CM systems have made strides in the last few years. In 1998, we saw the foretelling of enterprise level CM systems "As we enter this new era of networked information and collabouration, leveraging content for competitive advantage will become a priority that requires an enterprise response" (Stear, 1998). We have seen the rise and fall of CM vendors. We have seen the consolidation of the industry, as mergers and acquisitions have taken place to streamline product and service availability. And, we are staying tuned as the advent of the enterprise suite comes into focus – solutions that provide significantly more than CM alone. Look for collaboration, administration, management

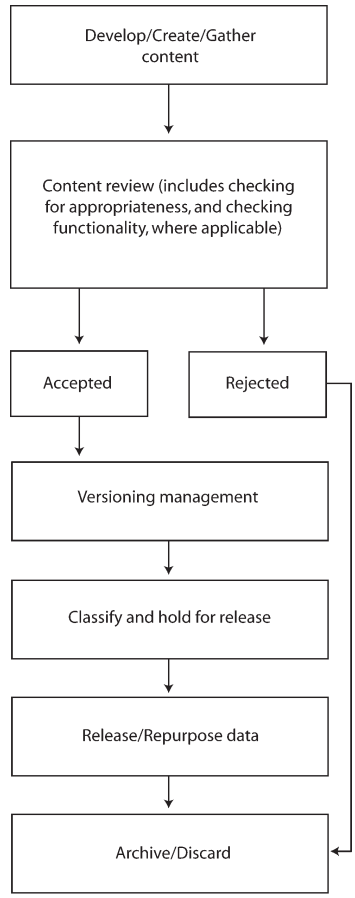

Figure 1.
Content life cycle model

support, taxonomies, metadata support, people finders, and federated search, wrapped together in a portal framework. "Live-in" versus "go-to" places are the order of the day (Feldman *et al.*, 2004). People want ways to manage content throughout its entire life cycle.

Reviewing and selecting a CM system requires developing specifications that clearly describe the business processes and expectations for how the CM system will facilitate the management of the content life cycle. Often portal software includes a CM system solution but rarely is it a robust enough application to support the content life cycle. Software specialised to CM is more likely a satisfactory solution.

White (2004) identifies the following key features, technology options, and cost of ownership considerations when evaluating CM system software solutions:

(1) *Key features*:

- Content creation through templates;
- Content review supported by workflow;
- Content versioning closely managed;

Steps	Description
Content creation	The inception and origination of the content, whether created by portal partners or administrators. This is content that will be distributed, *en masse*, through the portal
Content review	Includes review of the content for appropriateness, the quality assurance process of checking for adherence to editorial guidelines, and the testing of functionality. If the content is accepted, it moves along in the life cycle to the next step, content versioning. Otherwise, the content is rejected and either archived or discarded
Content versioning	Includes management of various versions of content, until the content is fully approved and ready to be released for public consumption
Content classified and held in repository	Content has been approved for appropriateness, completeness, and functionality, and is now ready to be classified to allow appropriate distribution, delivery, and search retrieval
Content released and repurposed for delivery to specific audiences	On release, the content is repurposed automatically by the system(s) to provide users with information in a value-added manner that is prespecified by content/portal administrators
Archive/discard	Content is archived either automatically by the system according to preset criteria, or manually in accordance with systematic guidelines. If the content is of no historical value, it may be discarded

Table VI.
Content life cycle steps

- Content tagging and holding in a repository;
- Content repurposed for delivery to specific audiences;
- Site design framework independent of content structure; and
- Comprehensive administrative functions.

(2) *Technology options*:

- Develop product internally or by contract: risk lack of long-term commitment and support;
- Open-source CM system products: favored by academic and not-for-profit due to low development costs; high cost of consulting support for implementation often not factored in;
- Commercial CM system products: when large base of users, high commitment to development, support, documentation, and training; and
- Corporate portal applications: currently weak compared to best commercial CM system.

(3) *Cost of ownership*:

- Price-per-server vs. price-per-user – cost of consulting, customising, and training range from one–four times the cost of software license for

commercial products; costs of implementation of internally developed or open-source may exceed implementation cost of commercial product.

Portal definition document

This article has provided the major issues and criteria surrounding the selection of a CM strategy for successful planning and development of a portal. The components of a portal CM strategy discussed and illustrated are the content inventory and analysis, content acquisition, access structures, classifying content, content life cycle, CM software, and metadata standards. The tools provided in this article will assist in creating the CM component of a portal definition document, which should include the following sections:

- Content inventory and analysis;
- Content acquisition strategies;
- Classification strategies; and
- CM software specifications.

What's next?

The next article in this series will discuss the issues and details of establishing a portal administrative framework, which outlines the management processes and roles of key stakeholders involved in portal development and maintenance.

References

Boiko, B. (2002), *Content Management Bible*, Wiley Publishing, New York, NY.

Emery, P. (2005), "Buzzword bingo: enterprise content management definitions and terminology", *AIIM E-Doc Magazine*, July/August, available at: www.edocmagazine.com/vault_articles.asp?ID=22005&vault=featu&header=e_features_header.g

Feldman, S., Caldwell, F., Byrne, T. and Pelz-Sharpe, A. (2004), "What's ahead for ECM?", *KM World*, Vol. 13 No. 3, pp. 20-2.

Fraser, J.C. (2001), *Taking a Content Inventory*, available at: www.webtechniques.com/archives/2001/10/fraser/

McGovern, G. (2001), "Web classification is essential", *New Thinking*, November 26, available at: www.gerrymcgovern.com/nt/2001/nt_2001_11_26_classify.htm

Melzer, J. (2005), "Enterprise information architecture in context", available at: www.jamesmelzer.com/bearings/archives/2005/05/enterprise_cont_1.html

Stear, E.B. (1998), "The content management strategy: don't go to work without it", *Online (Medford)*, Vol. 22 No. 3, pp. 87-9.

Warren, R. (2001), "Information architects and their central role in content management", *Bulletin of the American Society for Information Science and Technology*, Vol. 18 No. 1, available at: www.asis.org/Bulletin/Oct-01/warren.html

White, M. (2004), "Content management software: success factors for its selection and implementation", *Business Information Review*, Vol. 21 No. 1, pp. 36-43.

Further reading

Alsup, M. and Strong, K. (2004), "Enterprise content management: charting the path", *AIIM E-Doc Magazine*, Vol. 18 No. 3, pp. 18-27.

Choi, E. (2002), "Redesigning a large and complex website: how to begin and a method for success", *Proceedings of the 30th annual ACM SIGUCCS Conference on User Services*, available at: http://portal.acm.org/citation.cfm?id=588652

Huff, M. (2002), "Taxonomies and metadata for content management", available at: www.hkla.org/presentation/XML.ppt

Jaydip, M.R., Koong, K.S., Liu, L.C. and Yu, C.S. (2003), "An identification and classification of enterprise portal functions and features", *Industrial Management & Data Systems*, Vol. 103 No. 9, pp. 693-702.

Krozser, K. (2005), "The content inventory: roadmap to a successful CMS implementation", available at: www.alttags.org/archives/2005/04/05/46

LexisNexis (2003), "LexisNexis Premium content integration: capitalize on the opportunities now", *EContent and Information Today*, Special Supplement, June, pp. S8-S9.

Montague Institute (2002), "Ten taxonomy myths", *Montague Institute Review*, November, available at: www.montague.com/review/myths.html

Moore, A. (2004), "ECM in the eye of the beholder", *KM World*, Vol. 3 No. 3, pp. 3-22.

Nakano, R. (2002), *Web Content Management: A Collaborative Approach*, Addison-Wesley, Boston, MA.

Quirk, J. (2002), "A guide to defining an enterprise portal project", *Intranets*, January/February, available at: www.intranetstoday.com/Articles/?ArticleID=5050&ContextSubtypeID=56

Reidy, K. (2004), "Content in context: emerging portal standards facilitate content integration", *EContent*, Vol. 27 No. 5, pp. 36-40.

Stearns, S. (2004), "Automated classification tools to improve information discovery", available at: www.inmagic.com/news/press_room/white_papers/classification_request.html

Unitas Corporation (2001), "Managing unstructured content in enterprise information portals (EIPs): documents, taxonomies, content and knowledge management in EIPs", available at: www.unitas.com

Veen, J. (2002), "Doing a content inventory (or a mind-numbingly detailed odyssey through your web site)", available at: www.adaptivepath.com/publications/essays/archives/000040.php

Warner, A.J. (2002), "A taxonomy primer", *Lexonomy*, available at: www.lexonomy.com/publications/aTaxonomyPrimer.html

Weibel, S. and Miller, E. (2000), "An introduction to Dublin Core", available at: www.xml.com/pub/a/2000/10/25/dublincore/

Wodtke, C. (2002), "Mind your phraseology! Using controlled vocabularies to improve findability", *Digital Web Magazine*, August 13, available at: www.digital-web.com/articles/mind_your_phraseology