

CONTENT MANAGEMENT SYSTEM IN ELECTRONIC ENVIRONMENT

Dr. M. Ananda Murugan

Deputy Librarian

Banaras Hindu University

Varanasi, Uttar Pradesh, India

Abstract

Digital content management system is a software system that provides preservation, organization and dissemination services for digital collections. By adapting the systems analysis process, its needs and developed content management system requirements for finding a suitable information system that addresses the increasing needs of digital content management. Dozens of commercial and open source candidates were examined to match against the requirements. This article provides detailed analysis of three major players (Greenstone, Fedora, and DSpace) in key areas of digital content management: preservation, metadata, access, and system. This paper describes the process used to analyze and evaluate potential candidates and includes results of analysis to illuminate the process.

Key words: Digital content, Content Management System, Information System

Introduction

Content management (CM) is the set of processes and technologies that support the collection, managing, and publishing of information in any form or medium. In recent times this information is typically referred to as content or, to be precise, digital content. Digital content may take the form of text (such as electronic documents), multimedia files (such as

audio or video files), or any other file type that follows a content lifecycle requiring management. Content management is the process of managing content from creation through editing, publishing and archiving to retirement. A content management system (CMS) usually refers to software that helps organizations collaboratively manage content. A CMS will provide a simple interface for end users to add or edit content. It may also have processes for assigning permissions (for who may add, edit or approve content), tracking changes, and versioning (the ability to roll back to a previous version of the content).

What is Content Management?

Businesses and organizations that have large amounts of information to provide to users need a method for guiding that information from creation through editing, approval, publishing and maintenance to archiving. This process is generally referred to as content management.

While the concept of content management has been around for a long time – newspapers have been using it for decades – it’s a relatively new term for most people. The rise of digital technology and the proliferation of Web sites have brought content management to the mainstream. All organizations, from large businesses with thousands of digital documents to small organizations with simple Web sites, find themselves having to deal with ever increasing amounts of information.

Definitions:

An **asset-based CMS** stores information as individual blocks or assets. Dynamic pages are then assembled by drawing from the pool of assets. Asset-based systems are often preferred for sites with large amounts of content or when content must be sorted and presented in multiple formats.

Content refers broadly to meaningful information made available for retrieval by users. It may take the form of text, images, audio or video. is a central database for storing content that is then managed by a CMS interface.

Content re-purposing is the ability to publish content from a single repository into multiple formats, for example, as a Web page, text document, etc.

Enterprise content management (ECM) is the process for managing all the information of an organization. This may include print documents, electronic documents and records, computer applications, images, multimedia files, Web site and email documents, etc.

In a **page-based CMS**, each page is a separate unit which can be modified individually. Page- based systems offer greater flexibility and may be less expensive than asset-based systems.

Web content management (WCM) is the process of managing digital content for Web sites.

The Process of Content Management:

Content management practices and goals vary by mission and by organizational governance structure. News organizations, e-commerce websites, and educational institutions all use content management, but in different ways. This leads to differences in terminology and in the names and number of steps in the process.

For example, an instance of digital content is created by one or more authors. Over time that content may be edited. One or more individuals may provide some editorial oversight thereby approving the content for publication. Publishing may take many forms. Publishing may be the act of pushing content out to others, or simply granting digital access rights to certain content to a particular person or group of persons. Later that content may be superseded by another form of content and thus retired or removed from use.

Content management is an inherently collaborative process. It often consists of the following basic roles and responsibilities:

- Creator - responsible for creating and editing content.
- Editor - responsible for tuning the content message and the style of delivery, including translation and localization.
- Publisher - responsible for releasing the content for use.
- Administrator - responsible for managing access permissions to folders and files, usually accomplished by assigning access rights to user groups or roles. Admins may also assist and support users in various ways.
- Consumer, viewer or guest- the person who reads or otherwise takes in content after it is published or shared.

A critical aspect of content management is the ability to manage versions of content as it evolves (see also version control). Authors and editors often need to restore older versions of edited products due to a process failure or an undesirable series of edits.

Another equally important aspect of content management involves the creation, maintenance, and application of review standards. Each member of the content creation and review process has a unique role and set of responsibilities in the development and/or publication of the content. Each review team member requires clear and concise review standards which must be maintained on an ongoing basis to ensure the long-term consistency and health of the knowledge base.

A content management system is a set of automated processes that may support the following features:

- Import and creation of documents and multimedia material.
- Identification of all key users and their roles.
- The ability to assign roles and responsibilities to different instances of content categories or types.

- Definition of workflow tasks often coupled with messaging so that content managers are alerted to changes in content.
- The ability to track and manage multiple versions of a single instance of content.
- The ability to publish the content to a repository to support access to the content. Increasingly, the repository is an inherent part of the system, and incorporates enterprise search and retrieval.

Content management systems take the following forms:

- a web content management system is software for web site management - which is often what is implicitly meant by this term
- The work of a newspaper editorial staff organization
- a workflow for article publication
- a document management system
- a single source content management system - where content is stored in chunks within a relational database

Content management expert Marc Feldman defines three primary content management governance structures: localized, centralized, and federated—each having its unique strengths and weaknesses.

Localized Governance:

By putting control in the hands of those closest to the content, the context experts, localized governance models empower and unleash creativity. These benefits come, however, at the cost of a partial-to-total loss of managerial control and oversight.

Centralized Governance:

When the levers of control are strongly centralized, content management systems are capable of delivering an exceptionally clear and unified brand message. Moreover, centralized content management governance structures allow for a large number of cost-savings opportunities in large enterprises, realized, for example, (1) the avoidance of

duplicated efforts in creating, editing, formatting, repurposing and archiving content, (2) through process management and the streamlining of all content related labor, and/or (3) through an orderly deployment or updating of the content management system.

Federated Governance:

Federated governance models potentially realize the benefits of both localized and centralized control while avoiding the weaknesses of both. While content management software systems are inherently structured to enable federated governance models, realizing these benefits can be difficult because it requires, for example, negotiating the boundaries of control with local managers and content creators. In the case of larger enterprises, in particular, the failure to fully implement or realize a federated governance structure equates to a failure to realize the full return-on-investment and cost-savings that content management systems enable.

Implementation:

Content management implementations must be able to manage content distributions and digital rights in content life cycle. Content management systems are usually involved with digital rights management in order to control user access and digital rights. In this step the read-only structures of digital rights management systems force some limitations on content management implementations as they do not allow the protected contents to be changed in their life cycle. Creation of new contents using the managed (protected) ones is also an issue which will get the protected contents out of management controlling systems. There are a few content management implementations covering all these issues.

How content management systems work:

A CMS will allow certain areas of your website to be changed, modified or updated. Generally most web site owners ask to do this whilst still maintaining the ‘brand integrity’ of the website, so the pages on the site are built using set templates that the CMS cannot alter. However, depending on the degree of content management required, and the skill set of the user (we always provide CMS training for site administrators), the CMS can let you manage the following:

- Words and copy
- Prices
- Images
- Documents and pdfs etc.,

Audio and video files All these ‘digital assets’ are stored by the CMS in a reusable fashion and can be managed by content authors through a software application.

Most content management systems provide for the addition of new template pages, once again drawn from a database of pre-created styles. A CMS should also provide for a ‘Publisher’ or administrator to sign off content prior to upload in cases where content creation may be more broadly distributed.

Some of the significant and important applications include, but not limited to, the following. These are primarily based on the experience and user feed back seen from literature and e-lists subscribed to by librarians/information professionals and hence cannot be claimed as comprehensive.

General Platforms/Applications

- i. Operating Systems: Linux, Free / Open BSD, Open Solaris...
- ii. Web Servers: Apache & Lots in Java (<http://java-source.net/open-source/web-servers>)

- iii. Web Server-side Scripting: PHP (Architecture Linux, Apache, MySQL, PHP (LAMP) Windows, Apache, MySQL, PHP (WAMP), eg. xampp (<http://www.apachefriends.org/en/xampp.html>))
- iv. Web Services: Apache Tomcat Web Container/Service, Apache Cocoon Content Framework/Service & Apache Ant Build Tool
- v. Programming Languages: Perl, PHP, Python
- vi. Database Management Systems: MySQL, PostgreSQL, mSQL
- vii. Applications: Apache, Tomcat, emacs, grep, MySQL, sendmail, ssh
- viii. Image processing: ImageMagick, tiffinfo/tiffdump
- ix. Server Log Analysis: Webalizer (<http://www.webalizer.org/>)

Library Specific Platforms/Applications

- i. Integrated Library Management Systems(ILMS):
KOHA (<http://www.koha.org/>)
Evergreen (<http://wiki.code4lib.org/index.php/Evergreen>)
Emilda (<http://wiki.code4lib.org/index.php/Emilda>)
OpenBiblio (<http://wiki.code4lib.org/index.php/OpenBiblio>)
phpMylibrary (<http://wiki.code4lib.org/index.php/PhpMyLibrary>)
NewGenLib (<http://www.verussolutions.biz/>)
- ii. Z39.50 Protocol for online search/retrieval (<http://www.loc.gov/z3950/>)
YAZ Z39.50 Client (<http://indexdata.com/yaz/>)
'Mercury' Z39.50 Client (<http://www.basedowinfosys.com/projects/mzc>)
- iii. MARC Parsers / Editors / Tools
MarcEdit <http://oregonstate.edu/~reaset/marcedit/html/index.php>
MARC.pm (Perl), MARC4J (Java)
- iv. Library Oriented Search Engines
Cheshire (<http://cheshire.berkeley.edu/>)
Pears
dbWiz (<http://researcher.sfu.ca/dbwi>)

- v. Portals MyLibrary, Wordpress
- vi. OAI service providers and data providers
PHP OAI Data Provider
- vii. Database Management Systems CDS/ISIS, Win/ISIS, GenISIS etc.
- viii. Serials Manager CUFTS (<http://researcher.sfu.ca/cufts>)
- ix. Citation Manager (from PKP, Simon Fraser University, Canada)
Bibliographic Management (<http://researcher.sfu.ca/cm>)
- x. Link Resolving
GODOT - Electronic (Online) Resources
Management(<http://researcher.sfu.ca/godot>)
- xi. OJS (Open Journal Publishing) <http://pkp.sfu.ca/ojs>
- xii. OCS (Open Conference workflow automation) <http://pkp.sfu.ca/ocs>
- xiii. Open URL 1.0 (<http://www.oclc.org/research/software/openurl/default.htm>)
- xiv. Open Digital Libraries
Greenstone, DSpace, Eprints, FEDORA etc.
- xv. Open Access Archives / Institutional Repositories
DSpace, Eprints, FEDORA, CDSWare, Greenstone etc.
- xvi. Open Archives Harvester
Harvester (<http://pkp.sfu.ca/harvester>)
- xvii. Learning Management Systems (LMS)
Moodle & Manhattan etc.
- xviii. Content Management Systems (CMS)
Joomla, Drupal, MediaWiki
- xix. XML Tools and Systems
Utilities (Xalan, Xerces, libxml, libxslt, saxon)
Editors (emacs / nxml-mode)
Database / Search Engines (Apache Xindice, Berkeley DB XML, eXist)
Publishing/Web Application Frameworks (AxKit & Cocoon)

This tutorial intends to focus on some of the most popular library requirements such as integrated library automation, digital library, open access archives (institutional repositories), open archives harvesting etc., and in particular, highlights the major features and functionalities of KOHA, Greenstone, Dspace and the PKP OAI Harvester.

Advantages of CMS:

The advantages of using a CMS to build a library website include:

- Separation of content, logic, and data
- Ability for multiple content providers (many staff members can edit the website, rather than just one expert)
- Easier or automatic integration with Web 2.0 tools (built in RSS for instance)
- Many cool add-ons that provide added functionality to the website (an Events Calendar for instance)

Conclusion:

The Challenge of effectively addressing these issues is complicated by the fact that most effective when the managers are quickly access the information they need or demand. Today most of the data information needed for managerial decisions making is stored in computer based files and databases. Implementation is more important than product. Remember that implementation – the actual deployment and use of the site – is more important than the CMS tool you choose.

References:

1. Boiko, Bob (2004). Content Management Bible. Wiley Publishers. p. 1176. ISBN 0-7645- 7371-3.
2. Rockley, Ann (2002). Managing Enterprise Content: A Unified Content Strategy. New Riders Press. p. 592. ISBN 0-7357-1306-5.
3. Hackos, JoAnn T. (2002). Content Management for Dynamic Web Delivery. Wiley Publishers. p. 432. ISBN 0-471-08586-3.
4. Glushko, Robert J.; Tim McGrath (2005). Document Engineering: Analyzing and Designing Documents for Business Informatics and Web Services. MIT Press. p. 728. ISBN 0-262-57245-1.
5. Yan Han, (2004). Digital content management: the search for a content management system, Library Hi Tech, Vol. 22 Iss: 4, pp.355 – 365
6. Sreekumar M.G. and Sunitha T. 2005 Essential Strategies and Skill Sets Towards Creating Digital Libraries Using Open Source Software. [Proceedings of NAACLIN 2005, DELNET, Bangalore, India].
7. IIMK Digital Library
8. Greenstone Support for South Asia
9. Witten, Ian H. 2003 Examples of Practical Digital Libraries : Collections Built Internationally Using Greenstone D-Lib Magazine 9 (3): 1-15.
10. Altman, M. 2001 Open Source Software for Libraries: from Greenstone to the Virtual Data Center and Beyond. IASSIST Quarterly. Winter : 1-7.