

CONTENT MANAGEMENT: AN OVERVIEW

by

Jambhekar Ashok *

* Librarian & Head, Indian Institute of Management, Ahmedabad – 380 015

Help and Contribution of Mr. Shivpal Singh Kushwah, Sci./ Tech Officer-I, INFLIBNET, UGC Ahmedabad, and Mr M Paul Pandian, Librarian, Indian Institute of Management, Indore is gradually acknowledged.

0 Introduction

Before discussing content management, it is necessary to define the content. The content word is dependent on its context. Content is made up of items, elements, and things that have no meaning without a well understood context. The simplest meaning of content is information put to use. Information is put to use when it is packaged and presented (published) for a specific purpose. More often than not, content is not a single “piece” of information, but a conglomeration of pieces of information put together to form a cohesive whole. Books, newspaper etc. all have content. The Web is no different because the sites are also made of articles, indexes, graphic etc properly organized and presented.

The Web explosion brought with it the proliferation of published content and the need for content management. It has been around as long as content. It was basic and manual for a very long time. In the pre-web years content was not the concern of technologists but that of publishing industry. The traditional goal of managing content was to get it published to paper. Authoring of content, the editorial process and the design, production and marketing of a publication was done with care. However, the movement of content through this process and the potential value of it beyond just publication did not receive attention. For a long time technology did not matter for publishing industry but in the recent years it has become a key for its survival. Technology has brought about a significant change. It changed the publishing industry, business of content and the importance of content management. It began with the content being published on diskettes, CD-ROMs and eventually the Web. The value was in the content. The content management, therefore, has become important.

1 Vocabulary of Content Management

✍️ **Workflow:** The heart of most WCM applications, workflow allows you to assign a sequence of interdependent tasks, and assign individuals or roles for each task. Workflows can be assigned for different types of *content* or for different Web

sites—Internet or Intranet—or Web sites tied to specific departments that participate in Web site development.

✍️✍️**Auditing or Audit Trail:** Many WCM applications supporting workflows also support auditing. Auditing is the ability to look at a sequence of activities within a workflow to determine who, when, and why *content* was changed.

✍️✍️**Staging:** Many large Web sites choose to “stage” Web pages or *content* before they are pushed out to production. Often “staging” is done as part of the workflow where the last task will be to stage the *content* on a staging server for review before placing it onto the production server.

✍️✍️**Roll-back or Versioning:** The ability to “roll-back” a page or a part of the Web site to a previous version. This has great implications for organizations that are highly regulated like health care, where the organization may need to show what was contained on the site in previous versions.

✍️✍️**Check-in/Check-out:** A feature designed to facilitate collaborative participation on a single document without participants overwriting one another.

✍️✍️**Behavior Tracking:** The ability to observe a user of your site and capture click-through or click-stream data to be stored in a user profile. Often, this information is tied to a cookie file for repeat visitors so that more targeted information is provided, e.g. ads, news, etc.

✍️✍️**User Profile:** Information stored in a profile database that reflects the users preferences for *content*. A profile may also store historical information, such as past purchases.

✍️✍️**Business Rules:** Business rules are tied to behavior tracking and the creation of users profiles. Rules are created and assigned to specific profiles that determine the types of *content* they might receive and the format of that *content*. So, a fictitious example from my own health care environment might set the following conditions: consumers who type in the search phrase “acupuncture” will be shown a banner ad for an upcoming seminar on alternative therapies.

✍️✍️**Personalization:** The ability to match *content* to predefined profiles. Profiles are either defined by the end-user through an entry form or assumed based upon behavior tracking and applied business rules.

1 What is Content Management

Management is the process of organization, planning, command, coordination and control to achieve defined objective. Content management is effectively collecting, managing, and making information available in targeted publications. In other words it is a discipline that involves the collection, management, and publication of content with clearly defined rules, methods, documented workflows, and applicable tools and techniques with effective publishing system.

According to Forrester Research “Content Management is a combination of well-defined roles, formal processes, and supporting systems architecture that helps organizations contribute, collaborate on and control page element such as text, graphics, multimedia, and applets”.

As mentioned above the Content Management is effectively collecting, managing, and making information available in targeted publications, these three aspects mean :

- ☞ In Collection, information is either created or acquired.
- ☞ Converted to a master format (such as XML) and segmented into discrete chunks called content components. Components are metadata containers for the contents that make it easier to organize, store, and retrieve the information.
- ☞ Content is managed within a repository that consists of database records and/or files containing content components, plus administrative data (i.e. the system’s users).
- ☞ To make content available, the content management system publishes to targeted publications such as Web sites, printable documents, and email newsletters.

2. Core Concepts behind Content Management

Content management concepts include the following:

- ?? Understanding content domain, from which all of the structural decisions flow.
- ?? The notice on content components, which allow content processes (collection, management and publication) to be automated.
- ?? Target publications, which are the end result of any content system.
- ?? A Metatorial? framework, which unites all of the content into a single system of meta information.

☞ **Content domain:** It is the scope or range of information that is intended to be captured, managed and published. There is direct relationship between the content domain and the goals of CMS. CMS should have a concise domain statement so that it can be known what is and is not of interest.

☞ **Content components:** The content can then be broken up into its component pieces. Components divide information into convenient and manageable chunks. They share some common attributes, such as format or length. A component should have meaning in and of itself, without needing the context of other components to make it meaningful.

These are small reusable pieces of content that can be linked together to achieve larger results.

✍✍ **Target Publications:** The viewers or readers of content do not care about content components. They do not care how these components are collected and managed. They care about the presentation of the information it wants, it wants to see what it is used to seeing – a normal publication, like a book, magazine, or web site. Publishing is simply releasing information that was previously being developed. All publications should have the following aspects to them:

📁📁 Publication Purpose

📁📁 Publication Publishers (Management, Editorial, Technical, Creative and architectural Staff).

📁📁 Publication Authorship

📁📁 Publication audiences

📁📁 Publication format

📁📁 Publication structure

Editorial and Metatorial Frame work.

This is governed by three types of rules:

✍✍ Correctness Rules like rules on punctuation, word usage etc.

✍✍ Communication Rules that assure that contents projects a specific image, and targets a specific audience, these concern the voice of the content (active, passive, first person etc.) as well as other stylistic rules.

✍✍ Consistency Rules – Once defined a term is always used the same way.

In dynamic publications, like web sites, additional rules are needed to assure that the publication stays organized. These are defined as the metatorial framework. It is a system of meta information which means information about the information you have created. Meta information about an article might be its publication data, author, its target audience etc. An editorial framework provides rules for creating content, a Metatorial framework provides rules for tagging previously created content with meta information.

3. Why Content Management

A content management system helps organize and automate collection, management, and publishing processes. CMS is needed because

?? For processing large amount of information effectively.

?? Information is changing too quickly only systematic process can provide solution.

- ?? For publishing more than one publication from a single base of content.
- ?? To make the design of a publication static to maintain uniformity and time saving.
- ?? And making content dynamic, versatile, and powerful.

4. Steps for the Content Management Process

Possible approaches are:

- 1) Create and document manual procedures that all participants understand and follow.
- 2) Manage the sequential nature of the process, or portions of the process, by selecting a software product that allows you to develop powerful and flexible workflow solutions.
- 3) Select special purpose tools that integrate into the entire process to help the team perform the most important tasks.
- 4) Select an all-in-one content management system that truly fits your needs for flexibility, overall functionality, technology choices, and future growth.

5. Ten Principles of Content Management

- Content must be stored only once.*
- Content must be stored separately from tagging.*
- Content granularity must be available at any level.*
- Content management capability must be enhanced through the use of metadata and system extensibility.*
- Content and publication load, edit and extract must be independent operations.*
- Content objects must have power.*
- Content management should be an enterprise solution.*
- Content views must support the users' needs.*
- Content management should start, not end, with the DTD.*
- Content management and workflow integration should be seamless.*

6. Content Management System

A content management system helps organize and automate your collection, management, and publishing process and needed when this process becomes too complex to manage manually. The need for a content management system can be assessed by the amount of content, the amount of change in the content, and the number of publications intended for creation. A library, archive, or museum management or cataloguing system, a picture library system, a word processing or other text file containing lists of digital resources, a presentation file, PowerPoint file, or a Multimedia application is not a content management system. Content management

systems do not contain information about the presentation of the digital content (end-user interface, navigation, design, or layout).

7. Essential Components of a Content Management System:

Following are the components:

- 1) Content Collection system
- 2) The Administration/Management system
- 3) The workflow system
- 4) *The publishing system*

1) **Content Collection system** is the tools, procedures, and staff that are employed to gather content, and provide editorial and metatorial processing. Content collection system can be sub-grouped into the following groups based on activities and nature of work performed.

?? **Authoring**. This component of the system allows the Authors to fit their content into the structures of a target publication. Metatorial framework is important for use of the content. So authors should use meta information framework within their content. Graphic artists, videotape production crews, photographers, technical writers, advertising writers, application developers, Web page developers, lawyers, human resource personnel, marketers, or anyone else that produces original material for the Web site perform this role.

?? **Aggregating**. This component of the system is a process of format conversion and subsequent editorial and metatorial processing. The editorial processing serves to segment and tag the content for inclusion in the repository. The aggregation is easier if the original content is editorially (its style and elementation) and metatorially (its componentization and the meta information that has been entered) closer to the content management system's framework.

?? **Converting**. This is the process of changing the elementation scheme (i.e., the tagging structure) of the content. In this process the structural as well as the format related codes must be handled. Conversion includes transforming Word documents to formatted HTML text, modifying bitmapped images to load faster on the Web, changing image formats, and modifying database elements. Web developers use templates, layouts, themes, and other methods to convert text into uniformly formatted Web pages.

?? **Editorial/Metatorial Services**. These are used for the purpose of formatting, voice, and style. Metatorial services are used to fit each new component into a system of structures and connections. Editors use style guide and Metators use a metatorial guide. Like the style guide, the metatorial guide details the meta information system and gives direction on how to fit new components into it. Both the guides are at the core of content collection.

2) Management System the repository of all content and meta information, as well as the processes and tools employed to access and manage the collected content and meta information.

Functions of Repositories:

?? **Store content.** Basically it is a kind of repository system that may be one or a set of databases of various kinds. It can include the file system and network resources of the host computer. If the repository is distributed among databases, one database is often in a master position, organizing the information in the others. The system must be able to store *Textual content, Components, Binaries and file-based data, Meta information.*

?? **Select content.** The system must allow access and selection of content from within itself.

?? **Manage content.** The system must facilitate the following management tasks:

- *Security*, including read and write access permissions for components
- *User maintenance* that interfaces to system user management resources
- *Content statusing* and tracking for staging publications, workflow triggers, and maintenance operations
- *Transaction logging and rollback* of major changes in individual databases or to the repository as a whole
- *Bulk automated processes* that run periodically against subsets of the repository
- *Input/output processes* that load in and push out information

?? **Connect to other systems.** The system must be able to communicate over the network with a variety of clients. Ideally, the repository should be able to communicate with LAN-based Web browsers, Internet-based Web browsers, and LAN- or internet-based non-Web client applications

3) Workflow is the tools, procedures, and staff employed to assure that the entire process of collection, storage, and publication runs effectively and efficiently, according to well-defined timelines and actions. It sets and administers the chain of events around collecting, repositing, and publishing.

A successful workflow system should:

??**Extend over the entire process.** Every step of the process, from authoring through final deployment of each publication, should be able to be modeled and tracked within the same system.

??**Represent** all of the significant parts of the process including:

- Staff members
- Standard processes
- Standard tools and their functions
- Time and data flow with a variety of transitions and charting representations

Represent any number of small cycles within larger cycles, with some sort of drill down to the appropriate level of detail.

Have a **visual interface** that shows cycles and players in the process graphically.

?? **Make Meta information in the repository available.** The workflow system should not have to store its own staff members, content types, outlines, and other meta information. It should be able to read the data that is stored in the repository, and make it available when appropriate in its dialogs and selection screens.

?? **Provide a conduit to the repository for bottom up meta information.** Whether or not the workflow system stores meta information, its screens will be a natural place for staff to enter meta information. Data such as author, status, and type are naturally entered in workflow screens. This data must be able to be transmitted into the repository from the workflow system.

Publishing

Content publishing describes the process by which content is drawn out of the repository and formatted into Web sites and other publications. To be flexible enough to produce a wide range of publications, the publishing system must include:

?? **Design.** This is really "step zero" in the process. Design is part of the origination phase of the content management process and will vary depending on the type of content. In this phase, the content is defined that will be published or installed on the Web site.

?? **Publication templates.** These templates draw content into the appropriate context for each particular publication.

?? **Runtime dependency resolution.** When content is added to the repository it cannot be determined where and when it will be used in a publication. Therefore, the publication system must be able to read and resolve content links when the publication is being produced. For example, if component A has a link to component B in the repository, but component B is not being published, then A's link must be suppressed by the publication system to avoid a bad link in the publication.

?? **File and directory creation.** The publication system must be able to create the appropriate file and directory set for the target publication. Additionally, the

system must have some mechanism for deploying the built publication to its final storage location.

?? **A full programming language** is required to manage the complexity in transforming content in the repository into a publication. The language should provide complete access to the repository databases and files, access to external objects and libraries, and all of the standard variable types and control structures of major programming languages.

I. Principle of CMS

Principle provides the basic guidelines to system developer, managers, and product and services developers. Especially when we talk about library and information community or nonprofit service community the guiding principle could be the five laws given by great Dr S R Ranganathan with the reinterpretation in the light of CMS. In the present context, the "Five Principles of Intelligent Content Management" described by Dan Sullivan are relevant..

In terms of content management five laws of Dr. Ranganathan can be reinterpreted in the following manner.

- ☞☞ Contents are for the use. (Books are for use.)
- ☞☞ Every User should get his/her content. (Books are for all; or every reader his books.)
- ☞☞ Every Content should get its client. (Every book its reader.)
- ☞☞ Save the time of the user/client. (end user). (Save the time of the reader.)
- ☞☞ Content System is a growing organism. (A library is a growing organism)

8. Five principles postulated by Dan Sullivan are as follows:

- ☞☞ Principle 1: Make Metadata King
- ☞☞ Principle 2: Know The User
- ☞☞ Principle 3: Control Access To Content
- ☞☞ Principle 4: Support Rich Searching
- ☞☞ Principle 5: Keep Content Timely, Automatically

Make Metadata King:

Metadata describes the essential aspects of text, such as main topics, author, language, publication, and revision dates, to improve the precision of full text and keyword searching. It is also helpful for classifying and routing content, purging expired texts, and determining the need for additional processing, such as translation. Managing content metadata involves extracting it from text, and then storing it. Storage must support a number of access and retrieval methods. Metadata can be stored in two ways: manage it separately from documents in a relational database or directly within the

document. Quality control and document ranking is another use for metadata. Metadata can also be extended to include automatically generated summaries and clustering data.

Know the user:

Profile creation is an important key to improving the precision and recall of information retrieval. Profiles are explicit representations of interests, and use the same representation schemes as the metadata describing the contents of documents.

Kuflik and Shoval have identified six different kinds of profiles.

- ⚡⚡ **User-created profiles** are the easiest to implement but put the burden on the user to create and maintain them.
- ⚡⚡ **System-generated profiles** analyze word frequencies in relevant documents to identify patterns indicative of interesting texts.
- ⚡⚡ **System-plus user-created profiles** start with an automatically generated profile that the user can subsequently adjust.
- ⚡⚡ **Neural-net profiles** are trained using interesting texts provided by the user to output relevancy rankings for other texts.
- ⚡⚡ **Stereotype models** of interests shared by a large group of users can provide the basis for building individualized profiles.
- ⚡⚡ **Rule-based filtering** implements explicit *if-then* rules to categorize content.

Control Access to Content:

This is required to ensure sharing within the bounds of well-defined security business rules. Content can be grouped into three broad access control areas:

Open-access information, license-restricted information, and privileged information.

Access to license – restricted information is defined by agreements with content providers. User authentication or IP address verification can be used to control access in this situation.

In privileged information access is granted on a need-to-know basis.

Support Rich Searching :

Keyword searching has been the most commonly used. However, to reach a higher percentage of precision and recall, it is a better approach to combine three techniques: Keyword searching, clustering, and visualization. With relatively high precision and recall, keyword searches can produce a large number of hits. Clustering addresses this problem. It is sufficient to say that the hierarchical clusters (Process of building a tree structure) provide a familiar taxonomy – like structure that let users navigate from broad collection of topics to more narrowly focussed texts. Clustering effectively groups documents based upon content.

There is another technique known as scatter/gather algorithm. It lets the user scatter documents into groups, gather a subset of these groups and then rescatter them to form new groups.

Keep Content timely, automatically:

Some aspects of Content Management should be automated to keep pace with the available supply of potentially useful content. We can use harvesters, crawlers, and file retrieval programmes to gather documents for inclusion in the content repository. These programmes are themselves driven by metadata about which sites to search and which directories or document management system to scan for relevant content. Only metadata about documents and indexing detail need to be stored in the portal or document warehouse, and the document themselves can be retrieved on an as-needed basis.

9. Features of CMS

Summary of features of the CMS are enclosed as Appendix 1

10. Skills required for Content Management

The following skills are essential for content management work:

- ? **Analytical Skills** – to bring together a diverse set of requirements, rationalize information from multiple sources, find the common patterns in requirements and content, and implement structures that will support a variety of needs.
- ? **Design Skills** – both creative and structural, to layout the various pieces of the system from the user input forms to the administrative structures, to the publication templates.
- ? **Technical Skills** – to be able at least to conceptualize the overall system and to interact with software developers and system administrators.
- ? **Organizational Skills** – to manage the inputs to the system and to develop methods for ensuring that the content and other parts of the system are designed for manageability.
- ? **Management Skills** – to work with others to design, develop, deploy and keep the system up and running.

11. Evaluation of Content Management System

Evaluation should consider: Selection, Creation, Management, and Delivery. There are many content types, from simple text and graphics to video and audio files. Modern Content management software makes operating your web site or e-publishing systems much easier if you pick up the right system for your work or task. A successful CMS must integrate content from various sources and customize it for the user or site, deliver

it to the users in their desired format or device compatible format whether that device is a desktop PC or mobile handset.

Before selecting CMS, selector should observe and evaluate the following:

- ✍✍ Over all goal and the strategy of organization
- ✍✍ Specific requirement of functionality
- ✍✍ Content management environment (including people and process)

Selection criteria approach must be based on the above points and should consider the following issues:

Creation

1. Does it support task-based workflow? e.g. Writing -editing-reviewing-designing.
- 2. Does it support collaborative work environment? Means more than one people can work at the same time**
3. Can non-technical person operate the system without much training?
4. Can participants contribute in both what you see is what you get, and source code view e.g. HTML

Management

- 1 Can content be separated from the format {template supporting}
- 2 Can third party tool be integrated with the system
- 3 Can metadata be managed and applied
- 4 Are both staging and production environments supported by system.
- 5 Does the tool support time-release content and removal of dated content
- 6 Does it support different kinds of content i.e. text, database, image etc.
- 7 Does it support multiple file format system? jpg, gif, xml, html etc
- 8 Does it support site management utilities i.e. automatic user tracking, link tracking, visited IP tracking etc
- 9 How many users, size of content does it support
- 10 Does it support access control environment

Delivery

- 1 Can content be repurposed to a variety of templates and modalities e.g. handhelds

- 2 Is user grouping, user targeting, content personalization possible at the time of delivery?
- 3 Can it support maintaining of users profile?
- 4 Does it support email list or chat integration?

12. Indian Scenario:

In India there is a lot of scope for the librarians to become content manager by taking advantage offered by the technology for publishing and distributing content. Today the community at large wants web based information sources due the location, time, accuracy, and speed advantage. Gradually the users are turning away from the conventional library systems. The need is to provide information on their desktop.

Publishing industry, personal publishing, research organizations, and government are engaging themselves in publishing e- content.

We have a rich heritage resource and rare and special literature, which are the correct candidates for the digitisation. Some of them are listed below:

- ☞☞ Indian heritage history
- ☞☞ Related to educations
- ☞☞ Related to public information
- ☞☞ Government Circulars
- ☞☞ Manuscripts
- ☞☞ Indian theses
- ☞☞ Content in Indian Languages

It is satisfying to note that already such efforts have been made for example the digitisation of theses undertaken by University of Mysore under NISSAT.

13. Conclusion:

CMS products are available, as well as developers who can build a custom CMS. This discipline is gradually growing, and with its growth easy-to-use solutions are likely to emerge. Comprehensiveness and flexibility required to deliver the appropriate publications are very important conditions for choosing the system. Electronic publications – the web has given tremendous boost to the growth of content management. In the competitive world and age of knowledge economy, it is critical for communicating on a large scale. Content Management System has enabled to control information. Now information can be delivered in a manner that produces a richer, more timely, more targeted experience for audiences, and a rational, cost-effective process for the publisher.

References:

1. BOIKO (BOB). Understanding Content Management. Bulletin of the American Society for Information Science and Technology; October/November, 2001.
2. Content Management.
<http://www.intelligentkm.com/feature/010507/feat1.shtml>
3. CMS Feature List.
<http://www.bris.ac.uk/ISC/cms/features.htm>
4. CVETAN (MICHAEL P). The Commandments of Content Management in a Database.
5. DAN (SULLIVAN). Five Principles of Intelligent Content Management.
http://www.intelligententerprise.com/010831/413feat1_1.shtml
6. GUNTHER (KIM). What is a Web Content Management Solution. Online; July/August, 2001.
7. JIM (REYNOLDS) and ARMINDER (KAUR). Content Management. Microsoft Enterprise Services White Paper E-Commerce Technical Readiness
<http://www.microsoft.com/technet/treeview/default.asp?url=/technet/itsolutions/e-commerce/maintain/operate/contmgmt.asp>
8. PAUL (FREDRICK). Choosing the Right Content Management System. P C World's; July 2001. p 142-146.
9. RANDALL (NEIL). Content Management. PC Magazine; Feb 1, 2001.
10. RANGANATHAN (S R). The Five Laws of Library Science 1931. Bangalore: Sarada Ranganathan Endowment for Library Science; New Delhi: Exclusive distributors, UBS, 1988, 1931.
11. WARREH (RITA). Information Architects and Their Central Role in Content Management Bulletin of the American Society for Information Science and Technology October/November, 2001

