

Evaluation of Open Source Content Management System: A Comparative Study

Sanjeev K Sunny

Abstract

There are many web-authoring software like FrontPage, Dreamweaver, etc which have been used to develop and maintain the websites. But gradually, it became difficult to maintain and update the websites because of their very dynamic nature and a variety of file formats. Content Management Systems(CMS) evolved as an alternative to such web-authoring tools. A content management system offers a way to manage large amounts of web-based information that escapes the burden of coding all of the information into each page in HTML by hand. Although the rapid proliferation of commercially available content management systems makes it easier to find a vendor that might have right solution, the cost to purchase a commercial CMS application is very high. As an alternative to commercial CMS, open source solutions are a tantalizing option. Though, there are a huge number of open source CMS available in public domain, it is difficult to decide what solution will work best for a particular organization (or library). The author explain on this paper the functional requirements, for novice users, of a CMS and to evaluate a few most popular open source CMSs.

Keywords: Content Management System, Comparison of CMS, Open Source Software

1. What is Open Source Software

That software which satisfies the Open Source Definition given by Open Source Initiative (<http://www.opensource.org/docs/definition.php>) to determine whether or not a software license can be considered open source, is called open source software.

2. What is Content Management

In simple terms, content management can be defined as a process of creating, collecting, organizing, categorizing and structuring information resources of any type or format so that they can be saved, retrieved, published, updated and re-purposed in any way desirable (Yu, Holly, 2004). Content Management encompasses a broad spectrum of areas such as document management (DM), knowledge management (KM), records management (RM), electronic content management (ECM), web content management (WCM), etc. In general, the term content management is used in connection with web pages that can be maintained by a browser.

Content management gained importance during the explosion of Web sites in mid 1990s. Initially, many institutions used to provide information resources to their users through their static websites (HTML websites). But as the size of the site increased, they started feeling difficulty in managing and updating all the information resources (contents) in that way

What is Content.

In the context of digital environment, it could include:

- ◆ Text
- ◆ Links
- ◆ Graphics
- ◆ Pictures
- ◆ Sounds
- ◆ Videos
- ◆ Data

Among all the above types of contents, the textual and data contents are in abundance in most of the organizations. Organizations may deal with a wide range of textual and data contents. This can be categorized as:

- ◆ simple pages
- ◆ complex pages, with specific layout and presentation
- ◆ dynamic information sourced from databases, etc
- ◆ training materials
- ◆ online manuals (policy & procedures, HR, etc)
- ◆ general business documents
- ◆ thousands of pages in total
- ◆ extensive linking between pages

3. What is Content Management System (CMS)

Content Management System (CMS) contains the terms content and management (administration) that imprecisely refer only to a system that administers content. Such a system could be a blackboard and a piece of chalk, or it could be something like Wikipedia (the free online encyclopedia at <http://www.wikipedia.org>), or an online auction house such as eBay (<http://www.ebay.com>). In all these cases, content is administered; at times even for a large number of participants as in the case of the last two examples. These participants play a major role with the CMS, on one hand as the administrators, and on the other hand as users (Graf, 2006).

A CMS is a software system used for content management. It offers a way to manage large amount of web-based information that escapes the burden of coding all the information into each page in HTML by hand. A CMS takes content from inception to publication and does so in a way that provides for maximum content accessibility and reuse and easy, timely and accurate maintenance of the content base (Warren, 2001). The idea behind a CMS is to make these files available on Intranet, as well as over the web.

4. Evaluation of CMS

Selecting and implementing a content management system (CMS) is one of the largest IT projects tackled by many organizations. There is no 'one size fits all' solution: no two organizations have the same requirements. Therefore, there is no single best list of requirements for a content management system. So, there is need of identifying the requirements of the organization before starting actual evaluation of the CMS. In the process of requirements identification, all the stakeholders - users, authors, etc. - should be involved.

A general checklist of requirements can be prepared, which can be used by any type of organization, small or large; profit-making or non-profit making, etc. Since there will be a huge list of requirements for a CMS, structured investigation methods should be used to ensure that the list of requirements is both manageable and sufficient. Any organization can use this checklist for their purpose as it is or after modifying some of the requirements depending upon their needs.

The requirements list can be structured in following sections:

1. Product Overview

Basic information about the product such as technology used, status of the product, total number of installations, etc. are necessary before choosing it.

2. Creation & Publishing

This section deals with the process of web content delivery including content, creation, approval, publishing and quality control.

3. Content Presentation

This section describes the presentation of content and accessibility features of the CMS.

4. Content Retrieval

This section deals with the retrieval features in terms of metadata, navigation and Searching.

5. Technical

This section summarizes the technical aspects of the CMS product including compliance, security, reporting and monitoring.

6. Interdependencies

This section deals with interdependencies with internal and external systems, to ensure seamless planning, implementation and management.

7. Built-in Applications

As per the requirements of today's web i.e. Web 2.0, many applications are required to be integrated with the CMS.

8. Support and Training

This section deals with ongoing product technical support as well as staff training and support.

4.1 Product Overview

This section can be further categorized in sub-sections of requirements as follows:

◆ Basic Information

Here, following facts can be checked: Product name, Company/Organization name, Company/Organization Web Page, Product Web Page, Company description, Product description, License, etc.

◆ Technology

Here, the technology used to develop the software and required for installation and running of the product should be checked. For example:

Operating System	:	Linux, Windows, etc.
Web Server	:	Apache, IIS, etc.
Programming language	:	Java, PHP, etc.
DBMS	:	Oracle, MySQL, PostgreSQL, etc.

Here, one fact is very important regarding the required software- whether they are proprietary or open source software. For example Oracle is proprietary and MySQL is open source DBMS.

◆ Status

Status means the year of introduction of the product, versions of the software, current versions of the software, frequency of updating, number of downloads, number of installations, active developer website, etc.

◆ Installation

The time and skills required for typical installation is checked with the available human resources in the organization.

4.2 Creation & Publishing

This section explains the main functionality of CMS. A good CMS should ensure that the process of content creation is easy. Sufficient authoring tools should be integrated with the CMS to ensure that all web content has appropriate approval, etc. This section can be further divided into following sub-sections of requirements:

◆ Content Creation/Authoring

One of the reasons for an organization to choose CMS is to try to avoid the bottlenecks caused by having only a few people skilled in the art of HTML. The expectation of a CMS is that it will allow staff with little or no HTML authoring knowledge to add content to the site. This is the functionality required by the authors (content creators) using the CMS. Without an effective authoring process a CMS can't succeed.

◆ Content Review

This is critical when managing the creation of content by multiple authors. This functionality ensures that all the content has appropriate approval and content is accurate and appropriate. The aim of this process is to enable a piece of content to be created, and then automatically forwarded to another person for review and /or for approval to publish. This process is known as the workflow process. The workflow process involves Workflow approval, Workflow Management and Notification.

◆ Content Publishing and Repurposing

This functionality is meant to streamline the publishing process and facilitate online review and approval. The publishing engine of the CMS takes the content, stored in the repository, and generates the final pages. There are two approaches to the way in which content is served up into the browser: the first of these serves static pages and so only requires the use of a web server. This is also referred to as static rendering. The second approach is through the dynamic creation of pages. The system selects as appropriate display template and then assembles the elements of the page from tagged components in the repository before presenting the page in the browser.

◆ Content Version Management

Every time a piece of content is checked out of the repository even the most minor of changes (e.g. a spelling mistake) will result in a new version number being created. A long document that is being created by a number of authors can therefore quickly build up a very long list of versions, and identifying intermediate versions can be difficult. Version management enables a website or an Intranet to be rolled back to a specific date and time.

◆ Content Metadata Tagging

Metadata is at the heart of CMS and the way that is implemented in a CMS has a very significant impact on the extent to which content can be found and reused. Each piece of content need to have metadata attached before it is added to the repository. In the context of CMS, there are following broad categories of metadata:

- i) Content metadata provides a way of identifying documents that may contain relevant information. This is usually what most people think about when the word metadata is mentioned.
- ii) Descriptive metadata enables the type of document to be identified. In this way a search can be limited to web content, streaming video, etc.
- iii) Administrative metadata identifies the relationship of the document to the business context. These metadata include, for example, the person and department owning the document, the date when the document is checked for relevancy, etc.

4.3 Presentation

The purpose of this section is to ensure that information and services are presented in a standard style and they present a professional image by maintaining control over the look and feel of the site. This section can be further divided into following sub-sections of requirements:

◆ **Templates**

Templates are used to separate the design from the content. It may be predefined, and/or customizable to control the visual presentation of content items.

◆ **Customization and Personalization**

The information being provided by the CMS can be tailored to a particular user. For this, CMS uses logging/tracking tools. Similarly specific information can be provided to specific users through these tools. Besides the information, the look of the page can also be personalized.

◆ **Multilingual Support**

For the globalization of the website, multilingual support is indispensable. CMS should provide the content and information in Language other than English also.

4.4 Content Retrieval

The purpose of this section is to ensure that the users are able to quickly and successfully find what they are looking for with minimum effort. This functionality can be achieved via metadata, via search and via navigation.

◆ **Metadata Generation**

As stated earlier, metadata helps in finding information from the repository. Metadata can be generated automatically at the time of creation of the content, or can be assigned by the author or any other person even after the creation.

◆ **Search Features**

The search function may be a part of the CMS or it may be a third party engine. It is very necessary to provide comprehensive search facility across the website, multiple sites or sub sites to support content retrieval. For this CMS may use automatic indexing, keyword generation, and full-text indexing. It may support the use of Boolean operators and proximity operators, etc.

◆ **Navigation Structure**

The site structure helps in navigation. Contents may be organized in sections and categories. The depth of categorization has an impact on the retrieval of information through navigation method.

4.5 Technical

The purpose of this section is to ensure that the CMS complies with industry standards, to provide support for the development, testing and deployment, security, etc. This section warrants checking

the requirements under following sub-headings:

◆ **Compliance**

CMS should comply with the industry standards so that the system is robust and integrates with existing platforms and applications. It should support standard based development tools (e.g. XML, XSLT and other open standards). It should also support different browsers and other productivity applications.

◆ **Reliability and Performance**

CMS should provide a reliable and stable web site with a proven capability. It is very important to check the number of users the CMS can handle at a time. Response time i.e. speed of operation is very important.

◆ **Audit trail**

The CMS systems should have this feature. It should keep track of all the changes, which occurs in /or with the content, after approval for publishing until the content items are removed from the CMS.

◆ **Content Aggregation and Syndication**

To syndicate the site is to make the Content Items from the Web site assigned to the Front page Component available for syndication via a distributable file. The Syndication Component is used to publish a Newsfeed of site's Public Content Items, which have been published to the Front Page in the Front-end of CMS. It should support for both inbound and outbound syndication.

◆ **Content Migration**

Content Migration is exporting the existing web content along with structure and content links into a new CMS with minimum effort and reworking. At the same time the associated content records (metadata) can also be exported.

◆ **Security**

CMS provides role based security (based on roles as defined in the CMS) across all templates, contents, services and repositories. It should provide ability to override the once set security.

◆ **Content Repository**

CMS should provide support for a content repository of a wide range of common content types: structured content (database data, MS- Excel spreadsheets, etc.), documents (MS- Word files, PDF, RDF, HTML, etc.), presentations and media (MS-PowerPoint files, image files, sound files, video files and other rich media files), etc.

◆ **Reporting and Monitoring**

The CMS must provide an extensive range of reports, for both users and administrators such as total logins, content browsing, downloads, etc. Ideally, the system should pro-actively report on any issues that arise. There should be provision for customized reporting also.

4.6 Interdependencies

The purpose of this section is to ensure that the CMS is capable of exchanging information with other organizations and ensures seamless exchange of information across internal systems as well as external systems, and integrate with existing process and infrastructure. This section warrants the study of requirements under following headings:

◆ Interfacing with External Systems

This functionality is mean to exchange information with other organizations (sites), and enable being part of a portal environment with other online service providers. To achieve the above goal the CMS should provide an Application Programming Interface (API) and architecture to connect and interact with external systems. Ability to integrate with external search engine is also desirable.

◆ Interfacing with Internal Systems

The ability to integrate with Lightweight Directory Access Protocol (LDAP) directories is desirable for CMS, in particular information from underlying directories that may be used for workflow routing and notification. Similarly ability to integrate with other products such as MS-Office, Lotus Notes, etc is also desirable.

4.7 Built-in Applications

To cope up with the requirements of today's web i.e. Web 2.0, many applications are required to be integrated with the CMS. Such as: Blog, Chat, Wiki, RSS, etc. Besides, Discussion Forum, FAQ management, Data Entry, etc. are other applications which may be desired with the CMS. Now a day many finance related applications are also available to be integrated with the CMS such as Inventory Management, Pluggable Payments, Pluggable Tax, Shopping Cart, etc.

4.8 Support and Training

This section is meant to check whether the CMS is supported by manuals, tutorial and training (free as well as commercial), developer community, online help, public forum, public mailing list, third-party developers.

5. Findings

There is a huge list of open source CMS on the websites such as <http://www.cmsmatrix.org> and <http://www.opensourcecms.com>. Choosing the right candidates for evaluation was not an easy task. Four most popular CMS, according to these sites, were chosen for evaluation. The findings are described in the following sections:

1. Drupal

"Drupal is software that allows an individual or a community of users to easily publish, manage and organize a great variety of content on a website. Tens of thousands of people and organizations have used Drupal to set up scores of different kinds of web sites. Drupal is open source software licensed

under the GPL, and is maintained and developed by a community of thousands of users and developers". (Source: Official Website)

"Tens of thousands of people and organizations have used Drupal to set up scores of different kinds of web sites, including community web portals and discussion sites, corporate web sites/intranet portals, personal web sites, aficionado sites, e-commerce applications, resource directories, etc." (Source: www.cmsmatrix.org)

Drupal includes features to enable content management systems, blogs, collaborative authoring environments, forums, newsletters, picture galleries, and file uploads and download and much more. Drupal is open source software licensed under the GPL, and is maintained and developed by a community of thousands of users and developers. It is relatively youngster compared to other CMS. Session Management is built into the core functions, which is more than other CMSs provided. Drupal has scalability, or ease of growing a Web site from small set of users to an enterprise level. It provides unlimited depth of categorization of content. Drupal has very good feature of using Thesauri and Taxonomy for assigning keywords to the content.

There is literally nothing you can't do with this CMS once you get into it. Weather you want a blogging platform, a static site, a fully interactive community site or something in between, Drupal can do it. Many basic site configurations can be done with the core set of modules, but if that can't do what you want there is almost certainly a contributed module or modules which can do it for you. The only serious deficiency in Drupal is that is the lack of a good WYSWYG editor.

5.2 Mambo

"Mambo is a full-featured content management system that can be used for everything from simple websites to complex corporate applications. Continue reading for a detailed feature list". (Source: Official Website)

"Mambo Open Source is one of the finest Open Source Content Management Systems available today. The default installation of Mambo is easy to set up and easy to maintain. The set up utility uses a 4 step wizard interface that allows you to install the entire system without the need of advanced technical knowledge. Once installed, the system includes a variety of templates you can choose and a large number of functionalities that are ready to go. Content can be added, edited and manipulated without having to know HTML, XML or DHTML — just enter your content via a friendly editor and click publish". (Source: www.cmsmatrix.org)

Mambo is very easy to install and it has an attractive, easy-to-use administrative interface and the WYSIWYG editor is very promising. The list of add-ons developed for Mambo is very huge like Joomla! but Session management and categorization are limited like many CMS other than Drupal.

5.3 Typo3

"TYPO3 is a free Open Source content management system for enterprise purposes on the web and in intranets. It offers full flexibility and expendability while featuring an accomplished set of ready-made interfaces, functions and modules". (Source: Official Website)

"TYPO3 is an enterprise-level open source content management system released under the GPL. It runs on more than 122,000 servers worldwide. The application has been translated into 43 languages and is actively being developed in a community of over 27,000 users in 60 countries. Some of its users include BASF, DaimlerChrysler, EDS, Konika-Minolta, Government of Québec Volkswagen, UNESCO, as well as numerous universities, government agencies and non-profit organizations." (Source: www.cmsmatrix.org)

Typo3 is big. Big application. Big Community. Big adoption. Big list of extended features. But the templating system seemed very complex, compared to other solutions. It takes a few hours to find out how to create a page. But once you get used to it, it has the best template system out of all of them (including the well known Mambo and Joomla). The administrative interface felt awkward and old, especially compared to Mambo and Joomla! Typo3 is not good for a project which needs to be finished next week; this could lead to a very frustrating experience. You'll need a lot of time to build your first sites, no doubt about that, but if your searching for a platform you can use in all your projects and still is very easy to handle and fully adjustable (once you understood how it works!), then Typo3 is what you're searching for.

5.4 Joomla!

"Joomla is one of the most popular and commonly used CMS today, the software is released under GNU General public license, so anybody can use the software and make modifications to cater for their own use". ."(Source: www.cmsmatrix.org)

The administrative interface is very attractive and easy-to-use. Joomla! comes with a very powerful WYSIWYG editor, so content creation is very easy. Templating system is very simple and non-technical users can easily use it. Basically there is nothing you can't do with Joomla; everything is available however you need to do little research to get what you want more than 1600 modules. But it has a highly limited taxonomy. Everything either has to be not categorized at all or in a section and subsection. The only work around for this is to edit the category descriptions so you can have at least ONE page in a section but no subsection. It has 3 types of modules, which is highly confusing for the novice users.

References

1. Bramscher, Paul F. and Butler, John T. LibData to LibCMS: One library's evolutionary pathway to a content management system. *Library Hi Tech*, 2006, 24(1). p.14-28.
2. Drupal 5 Getting Started. Available at http://drupal.org/files/getting-started_2.pdf (Accessed on 16/12/2007).

3. Drupal: Installing Modules Available at <http://drupal.org/node/176044> (Accessed on 20/12/2007).
4. Goodwin, Susan et al. CMS/CMS: content management system/change management strategies. *Library Hi Tech*, 2006, 24(1), p. 54-60.
5. Graf, Hagen. *Building Websites with Joomla!*. Birmingham-Mumbai: Packt Publishing, 2006.
6. Guenther, Kim. Content management systems as "silver bullets". *Online*, 2006, 30(4), p.54-55.
7. <http://www.cmsmatrix.org>
8. <http://www.opensourcecms.com>
9. Huttenlock, Terry L.; Beaird, Jeff W.; Fordham, Ronald W. Untangling a tangled web: a case study in choosing and implementing a CMS. *Library Hi Tech*, 2006, 24(1), p.61-68.
10. Joomla! Administrative Manual. Available at <http://help.joomla.org/content/view/15/280> (Accessed on 25/12/2007).
11. Joomla! Installation Manual. Available at <http://help.joomla.org/content/category/48/268/302> (Accessed on 25/12/2007).
12. Kmetz, Tom and Baily, Ray. Migrating a library's website to a commercial CMS within a campus-wide implementation. *Library Hi Tech*, 2006, 24(1), p.102.
13. Mambo Administrative Manual. Available at http://help.mamboserver.com/index.php?option=com_content&task=category§ionid=16&id=101&Itemid=121 (Accessed on 25/12/2007).
14. Mambo Installation Manual. Available at http://help.mamboserver.com/index.php?option=com_content&task=category§ionid=15&id=99&Itemid=70 (Accessed on 25/12/2007).
15. Salazar, Ed. Content management for the virtual library. *Information Technology and Libraries*, 2006, 25(3), p.170-175.
16. Seadle, Michael. Content Management Systems. *Library Hi Tech*, 2006, 24(1), p. 5.
17. Steve White and Andy Wallace. Joomla! User Manual. Available at <http://help.joomla.org> (Accessed on 25/12/2007).
18. The Drupal Cookbook (for beginners) Available at <http://drupal.org/handbook/customization/tutorials/beginners-cookbook> (Accessed on 20/12/2007).
19. Typo3 Tutorial. Available at http://typo3.org/documentation/document-library/tutorials/doc_tut_quickstart/0.1.0/view/ (Accessed on 25/12/2007).

About Author

Mr. Sanjeev K Sunny, ADIS Student, Documentation Research and Training Centre ,Indian Statistical Institute, Bangalore-560059.

E-mail: sanjeev@drtc.isibang.ac.in; sanjeev.sunny@live.com