

# The Use of Open Source and Open Standards in Web Content Management

Thomas Ferris Nicolaisen  
Department of Informatics  
University of Oslo  
Norway  
thomafni@ifi.uio.no

A. Bechina Arntzen  
ØSIR,  
College University i Hedmark  
Rena  
Norway  
aurillaa@hihm.no

## **Abstract –**

*The paper outlines how the key requirements of Web Content Management Systems (WCMS) can be met by including the use of open standards, and how these tools, their vendors and users can benefit from development in open source environments using open standards.*

**Keywords:** web content management, open source, open standards

## **1 Background**

The last few decades have seen revolution after revolution within information technology and telecommunications. The rise of the Internet, the success of the World Wide Web, the availability of personal computers and server performance, more recently the circulation of mobile devices and the distribution of broadband Internet are all trends of the new technological infrastructure which supports the new currency of digital data and information.

As storage space has grown, and network bandwidth has widened, the mass of digital information has exploded, both internally on intranets, and on the Internet. Users of the Internet have been most significantly affected by the increase in e-mail traffic and the amount of documents and pages available on the World Wide Web. This increase is driven and supported by processes of Web Content Management [Goodwin, 2002].

The value of information is only equal to that of its use. To use information, it must be found, recovered, formatted and presented. Information which is stored but never used is worthless. Digital information is enabled by the use of Information Systems. Before one can define the particular kind of Information System referred to as the Web Content Management System, one needs to define *content* itself, and separate it from *data* and *information*.

## **2 Web Content Management**

Data, information, content and knowledge are four ambiguous concepts which are regularly applied in Information Systems. If allowed to delimit the definition to digital representation, we leave out the definition of knowledge for now, focusing on the other three. These terms have various meanings, and are potential candidates for extensive ontological discussion. To avoid confusion, the meanings of these terms as used in this paper are defined as follows:

### **2.1 Data**

Data is seen as the basic unit of digital representation which can be used to construct information and content with more value for the consumer. Data is raw and granular. It does not inherently have any meaning as its meta-data is not self-contained.

Data is a set of symbols, ranging from a numeral value to a string of words, or even a large series of encoded symbols that compose a binary value representing sound or picture. One often mentions data processing, feeding data as input to a program or algorithm, the output being new data, information or content. Imagine calculating the mean of a hundred numerical values into one number. Data has been processed, but no meaning has been added. Had the value been wrapped with the context that this is the average temperature for the last three months, it could have been considered information.

### **2.2 Information**

One definition of information is *data with meaning*. The same information can be conveyed with different sets of data. Pieces of data combined with meta-data form a package of meaning that can be conveyed. In the first chapter of his Content Management Bible, Bob Boiko

(Boiko, 2005) includes all *the common forms of recorded communication*. Liz Orna is referenced with describing it as *knowledge transformed into a transportable format, visible or audible*. It appears information can be a primitive form of higher knowledge, or a more advanced composition of raw data. So where does content fit in?

## 2.3 Content

This is perhaps the vaguest term which we must define. Ideas include :

- Information put to use [Boiko, 2005]
- Information with human meaning and context
- Information shaped for an intended consumer
- Information with a purpose<sup>1</sup>

The definition used in this paper is streamlined for how content can be handled by an Information System. A collection or subset of information intended for a given audience or non-human consumer with a context of location, period and situation. To put it another way, content is information composite, ordered, built and delivered.

## 2.4 Content Management

Now that the definitions are in place, the segment of Information Systems known as Content Management Systems can be defined. Note that in the industry of content management, the use of the term is largely undetermined.

Content management means different things for different actors. The most basic life cycle of content is production and consumption. For the producer, the processes of content management include creation, formatting, structuring and integration of content. For the consumer, it includes search, export, and display. The sum of these processes makes out content management. A content management system (CMS) is a suite of tools designed to assist and support these processes.

A simpler and more practical perspective is to say that a CMS is a piece of software responsible for taking care of all the digital documents and files in an organization. The functionality of such a system includes document repository control, the company's digital library.

Such a system is potentially quite an advanced implementation depending on whether it includes features like access control, product management, content versioning, import/export, workflow and search functionality.

## 2.5 Web Content Management Systems

As pointed out earlier, the explosion of digital information has been most significant on the World Wide Web. To manage this mass of online content and use, a new breed of information systems has evolved; the Web Content Management System (WCMS). The responsibility of such a system is similar to that of the CMS, only it is delimited to content which consumption is done on the WEB by way of the World Wide Web.

It seems to be a natural course to have the WCMS integrated with the CMS. Parts of the content which should be exposed on the Web already exist somewhere in the CMS, perhaps on the intranet or on a central file server. In reality, many organizations do not have a proper CMS, or they are not pleased with the one they have.

## 3 Open standards

A standard is an agreement of two or more parties regarding a product, specification or other. Standards used by web applications are mostly guarded by the Internet Engineering Taskforce (IETF), the World Wide Web Consortium (W3C), Institute of Electrical and Electronics Engineers (IEEE) and International Telecommunications Union (ITU). Examples of successful standards are hypertext markup language (HTML), hypertext transfer protocol (HTTP) and resource description framework (RDF).

System developers can choose either to use existing standards or invent their own. Sometimes not having to follow a standard is easier and quicker than having to fulfill the specification's every need for quirky details, but along the network externalities [ref ant] in the system where other systems interconnect, open standards must be followed. This applies to the technology used for **transport** (all applications connected to the World Wide Web must be accessible through HTTP) and **storage** (a web page must output format in HTML, pure text or a standardized binary format like Bitmap pictures or even Macromedia's Flash), some of which are subjectively recognized de facto standards [ref standards].

Proprietary standards can be open (Adobe's PDF format and Macromedia's Flash file format) or closed (Microsoft

---

<sup>1</sup> By way of (Boiko, 2005), this is the now disbanded ContentWatch organization's definition.

Office Word documents and Powerpoint presentations). A proprietary standard can only be changed by its owner. You can make software that reads both open and closed standards, but figuring out how the closed standard is built up internally can be quite difficult, and often not cost-effective. Some standard-owners distribute their internals for a licensed fee.

A WCMS will naturally output its content through HTML on a web site. Internally, however, the implementation may store the content in a home-grown format, for example a relational database with a streamlined scheme following no standard whatsoever (except the standard of SQL itself). As long as the company uses the WCMS the way it was built to be used, the inside workings of the content repository is not important. The problem arises when the company either wishes to change the output or use of the content, or to replace the WCMS all together. In most organization, this does eventually happen. Requirements change.

How will the content be exported from the old WCMS and imported into the new one? Manually copying the HTML code from each web page will no doubt be a very tiresome effort. Another alternative is to read content directly from the relational database with an exporter-application. If the WCMS has not supplied one, developing this application could be a large task. And then an application would have to be developed for importing the content into the new WCMS.

The best solution would be if the storage of both WCMS-es utilized a standard content repository, so the content of the old system could simply be dragged-and-dropped into the new one. Unfortunately, today there exists almost as many different content repository implementations as there are content management system vendors.

## Open Source

Simply put, open source software refers to programs whose source code is made available for use or modification<sup>2</sup>. This means that open source software is in fact free to acquire [Walli, 2005].

A lot of people find this hard to believe, and many presume that such software is produced on a volunteer basis, and therefore lacks quality, security and consistency. This is true for a lot of smaller open source projects, but larger companies do in fact develop open source software on an economically feasible business model<sup>3</sup>.

---

<sup>2</sup> <http://www.free-soft.org>

<sup>3</sup>

[http://www.opensource.org/advocacy/case\\_for\\_business.php](http://www.opensource.org/advocacy/case_for_business.php)

The revenue can be generated by offering support, customization and plug-ins. Larger software companies like IBM and Sun have for the last years been funding, as well as founding, open source projects to ensure that their ideas and standards are established throughout the open software community.

This paper will not delve further into the principles and ideas of the open source movement. The interests of WCMS users lie in the TCO of the software and its content. It is important to remember that most open source material comes without guarantees and warranty unless support is bought from the vendor/developer, and this is where the cost of “free” software lies.

## 4 Proprietary software

Why does proprietary software use less open standards than open source software?

Because **they do not have to**. Having a closed set of developers invalidates the need to use extra-common standards. While some of the developers most likely know several applicable open standards that can be used in the project, including these might increase the learning curve for other developers. An open source project does perhaps not have the collaborative luxuries of an office with a crew of geographically concentrated people. The learning curve of joining in on the project must be overcome with explicit documentation, and specifications of the standards involved are a great place to start.

Proprietary software is not necessarily **shared** [Hanseth, 2002]. The software is made, and put to use. Libraries (software components) of an open source project, on the other hand, might have to be used by other open source projects, and thus they need standards to enforce possible interaction between projects. Proprietary software seldom has to cross borders between companies, with the exception of retailed software like operating systems, tool suites and computer games. Microsoft uses a multitude of proprietary standards to enable other vendors to produce software for the Windows platform, examples are DirectX<sup>4</sup> for graphics and MFC for desktop applications.

**It takes away the software's edge.** Using open standards gives the world a window into the code and its workings. This makes it possible for others to use or exploit functionality or the storage of the software directly instead of using the intended client software. An example is if Microsoft used an open document standard for storing Word-files, it would be easy for competing software vendors to develop Word-like tools, increasing the pressure

---

<sup>4</sup> <http://www.microsoft.com/windows/directx/default.aspx>

on Microsoft's developers. Exemplifying a transport standard, WebDAV<sup>5</sup>, enabling this protocol on a proprietary content management system would give other software access to the content, and as proprietary CMS-es go, this is not always the desired result. An open source project would only welcome such competition, as it would verify the worth of the project's effort, as well as increase the user base of the standard and further establish a community around it.

**Obscurity means security.** Or does it? As well as the previous paragraph reasons to guarantee that the software vendor keeps as much of the customer's money as possible, it also gives hackers a harder time getting into your system. The security holes that can not be seen can not be exploited. However, this is a false sense of security, but it has acquired mythical status having end-users believing that open source means insecure, while it actually means well-tested and security hole-less software<sup>6</sup>.

ActiveX is a Microsoft specification allowing powerful functionality in Internet Explorer that can be activated across the net, but the protocol has been heavily exploited by malware, viruses and worms to such a degree that many technical administrators have disabled this feature on company computers<sup>7</sup>.

Note that even though Microsoft and their Office products are frequently used as examples of a proprietary software, they are not the big bad wolf regarding use of open standards. Such advanced software can not always suffice for the bureaucratic democracy and slow development of open standards. Microsoft is more and more embracing the use of open standards like WebDAV and SOAP<sup>8</sup> in their newest software. In fact the next version of the Office suite will use zipped XML-files for storage, like OpenOffice has been doing for several years<sup>9</sup>.

## 5 Advantages of Having WCMS in the OS World

To summarize, the advantages of enabling an open source system utilizing open standards are as follows.

**Exposure.** Web content must be available to as many visitors as possible, regardless of browser and operating system. There was much controversy regarding Google's

newest service of video search, which earlier was only possible on the Windows operating system. Now Google uses Macromedia Flash technology, independent of operating system, to display their video content. Open standards make it easier for browser windows to handle different formats of content.

**Extendability.** The content must be available for third party software and plug-ins. As a WCMS has an infinite set of requirements which no single software company can hope to satisfy by itself. Open source software can be indefinitely modified to suit requirements, as long as there are resources for such development.

**Portability.** This goes for all kinds of server-side software. Different customers rely on different operating systems for their servers and to maximize the segment of the customer base, the software should be built on open standards to ensure platform independence.

**Lock-in.** Or rather the improbability of it. To avoid locking the organization to the current data repository, the WCMS should use open standards for storage and transport. A WCMS quickly builds up a huge amount of content, and being locked to a single vendor could prove to be a gold mine, for the vendor, that is.

**Customization.** If the customer wishes to alter the system and change functionality or looks of the website, can these changes be done by any qualified developer, or does the vendor have to be payed in full? With off-the-shelf software, perhaps no customization is offered or allowed at all.

**Reusability.** Both content and functionality should be reusable in new systems. The customer might have bought expensive plug-ins and built an excessive amount of well-structured content. Does the new WCMS store meta-data in the same way and search it the same way as the old one? Can the plug-ins function in the new environment, alternatively be ported over? The chances of this happening is larger if the WCMS has standardized plug-ins, such as portlets or web services.

Finally there is the low cost-of-entry to be considered. Many WCM efforts are tied to a low-budget process. Larger WCMS vendor might not be willing to audition their tools for small or medium sized customers free of charge, much less deploy for testing purposes.

## 4 References

[1] Goodwin, S., Vidgen, R. 2002, "Content, content, everywhere... time to stop and think? The process of web content management", *Computer & Control Engineering Journal*, April 2002, p. 66-70

---

5 <http://www.webdav.org>

6 <http://www.dwheeler.com/secure-programs/Secure-Programs-HOWTO/open-source-security.html>

7 <http://vx.netlux.org/lib/aas05.html>

8 <http://www.w3.org/TR/soap>

9 <http://www.microsoft.com/whdc/device/print/metro.mspx>

[2] Boiko, B. 2005, *Content Management Bible, 2nd Edition*, Wiley Publishing, Inc.

[3] Walli, S. 2005, "Free and Open Source Licenses, Software Development and Distribution"  
[[http://www.optaros.com/pdf/wp\\_SWalli\\_FOSS.pdf](http://www.optaros.com/pdf/wp_SWalli_FOSS.pdf)]

[4] Hanseth, O. 2002, "From systems and tools to networks and infrastructures - from design to cultivation. Towards a theory of ICT solutions and its design methodology implications".  
[[http://heim.ifi.uio.no/~oleha/Publications/ib\\_ISR\\_3rd\\_resubm2.html](http://heim.ifi.uio.no/~oleha/Publications/ib_ISR_3rd_resubm2.html)]