National Council on Orientalist Library Resources

More than a catalogue

The Hybrid Oriental Library and the Content Management System

Introduction

The Management Challenge of the Distributed Hybrid Oriental Library

Libraries are services that provide organized access to the intellectual record, wherever it resides, whether in physical space or in scattered digital information spaces. The "hybrid " library of the future will be a managed combination of physical and virtual collections and information resources.[1]

The recent RSLP funded projects for Oriental Libraries and digitization projects funded by the New Opportunities Fund have accelerated the development of hybrid resources in the Oriental field and it seems appropriate to consider the long term future of these resources as we enter the era of the post-millennial Web.

The pervasive nature of the Web means that it has become the preferred vehicle for content delivery and the future for Oriental library resources is likely to include cross-institutional management of a number of Web-based electronic resources that might include:-

- Databases

- Images

- Links to other websites

- Geographical Information Systems

As funding opportunities encourage oriental libraries to deliver more of their resources in electronic format to a far wider user group than the traditional academic researcher, the websites we create are going to become increasingly complex to manage. Increasingly oriental librarians will become involved in the following activities:-

- **Authoring:** creating content

- **Metadata tagging:** describing content

- **Editing:** changing and updating

- **Collaboration:** several people editing content together

- **Security:** stopping the wrong people from manipulating content

- **Versioning:** keeping track of how content has changed

---

[1] Carr, R. *Towards the hybrid library: the national perspective in the UK.* Presentation to the MALIBU Conference, King's College London, 26 March 2001 [Web page last accessed 11/11/02] http://www.bodley.ox.ac.uk/librarian/malibu2001/malibu2001.htm

- **Personalization:** re-using the same set of contents so that they are displayed differently to different visitors

In an ideal "distributed hybrid library" environment the user would be presented with a seamless interface in a number of languages/scripts that allowed complex searches across different electronic resources. This interface would also be able to offer information in formats accessible to the visually impaired. On the management side the ideal "distributed hybrid library" would offer "self-service authoring" facilities to allow the relevant specialists in participating institutions to update entries for which they have responsibility, without having to have specialist web-authoring or design skills. Authored content would be stored in repository with version control so that conflicts between multiple authors would not arise and previous versions could be found and restored if required.

Commercial and public sector institutions are increasingly turning to the CMS, or the Content Management System, as a way of managing complex multi-author websites. The following paper explores some of the issues surrounding the CMS as a way forward for distributed electronic oriental library resources.

LMS & CMS

Library Management Systems and Content Management Systems

The HE/FE library sector has so far been slow to embrace the CMS[2], perhaps because the traditional role of librarians has been in bibliographical control rather than website content creation and management.

Libraries have invested heavily in Library Management Systems and some question that Content Management Systems have anything extra to offer. The LMS and the CMS do indeed have features in common from the point of view of workflow. In an LMS, acquisitions staff use Web-based templates to create records for new stock. Cataloguers add value to the bibliographic records (with less experienced staff being subject to an approval process). Metadata is recorded automatically detailing who entered/modified the content and when. A subset of the record is then re-used in several ways, such as a page on a Web OPAC, as a paper recall notice or an e-mail recall notice.

The principle difference between an LMS and a CMS is, therefore, in the way that a CMS facilitates website content creation and management and has the ability to bring together diverse electronic resources (which may include an LMS) 'under one roof.'

It may be that future LMS developments will include feature sets to cater for the demands of website content creation and management. It seems more likely, however, that the growing demand for products based on open architectures, fuelled by the technical standards demanded by bodies such as the New Opportunities Fund, will lead to the right hooks and sockets being provided so that interoperability between systems can be achieved. Library Management Systems of the future may well be integrated with Content Management Systems, alongside portals, student reading lists, distance learning programs, online archives and digitized image banks.

---

[2] For an example of a distributed hybrid library under the control of a CMS see the Forced Migration Online website at http://www.forcedmigration.org [website last accessed 11/11/02]

The CMS & Institutional IT Services

The interface between the subject specialists & IT professionals

The pre-millennial Web is characterized by highly manual approaches to maintenance, where those responsible for content have to pass material to a Webmaster (or Web team) before it can be published on the Web. The Webmaster is usually an IT specialist with many other roles, which may include server support as well as installation and maintenance of staff workstations.

The resulting situation was summarized by Paul Browning and Mike Lowndes in their article *Content Management Systems: who needs them?*

> Institutions are struggling to maintain their Web sites. Out of date material, poor control over design and navigation, a lack of authority control and the constriction of the Webmaster (or even Web Team) bottleneck will be familiar to many in the HE/FE sector.[3]

The experience of those Oriental RSLP projects that were involved in Web content creation was often one of frustration as project officers, initially recruited to deal purely with content issues were forced to gain high levels of technical expertise in order to try and overcome the Webmaster bottleneck.

Despite the pressures on institutional IT teams regarding Web management, the Browning and Lowndes article makes the following observation.

> We detect resistance within traditional campus computing services and MIS departments to the concept of the CMS (and indeed the need for an automated, post-millennial Web). Fundamentally a CMS devolves control over content to the owners of that content (rather than the technician), and then scales without increasing management overheads.[4]

This resistance, possibly due to fears about job security and/or reluctance to surrender the "creative" part of Web management, may mean that the CMS is slow to be adopted in HE/FE. The future for Oriental Web resources created through RSLP funding may be that of hard pressed oriental librarians trying to fight the manual maintenance system and the Web master bottleneck in order to get necessary updates implemented.

---

[3] Browning, P & Lowndes, M. Content Management Systems: who needs them? *Ariadne* Issue 30, Dec 2001. http://www.ariadne.ac.uk/issue30/techwatch/intro.html [website last accessed 11/11/02]
[4] ibid.

CMS features

Functionality requirements for the complex, multi-author, dynamic website

Alice Grant in her paper on Content Management Systems for the NOF-digitise programme describes a CMS as a database (though in fact some Content Management Systems store data in XML files rather than databases) : -

> which organises and provides access to all types of digital content - files containing images, graphics, animation, sound, video or text. It contains information about these files (known as 'digital assets'), and may also contain links to the files themselves in order to allow them to be located or accessed individually. A content management system is usually used to manage digital assets during the development of a digital resource, such as a website or multimedia production. It might be used by staff digitising images, authors and editors, or those responsible for the management of the content development process (content managers). Content management systems range from very basic databases, to sophisticated tailor-made applications. These more complex systems can be integrated with the eventual digital resource in order to enable access to digital assets and to allow regular updating.[5]

A CMS is a 'concept' rather than a product, the key goal of which is to increase the integration and automation of the processes that support efficient and effective Internet delivery. At present the boundaries of CMS space are blurred and substantial overlaps exist with document management systems, source control systems, and virtual/managed learning environments. Browning and Lowndes[6] identify seven broad approaches taken by CMS developers

- **Document Management Systems:** software designed to manage storage and internal publication of 'corporate' information

- **Electronic news/publishing:** online publishing of magazines and news websites, and electronic discussion groups

- **Source/versioning management control:** process control/source control among groups of contributors

- **'Middleware on steroids':** tools for dynamic web site creation from file system and database assets

- **Web content management frameworks:** 'second generation' tools built from the ground up for dynamic website creation management

---

[5] Grant, A. *Content management systems*. Nof-digitise Technical Advisory Service report
http://www.ukoln.ac.uk/nof/support/help/papers/cms.htm [website last accessed 12/11/02]
[6] Browning, P & Lowndes, M. Content Management Systems: who needs them? *Ariadne* Issue 30, Dec 2001.
http://www.ariadne.ac.uk/issue30/techwatch/intro.html [website last accessed 11/11/02]
[6] http://mkdoc.com/news/laptopchallenge/ [website last accessed 14/11/02]

- **XML processors:** a 'third generation' of CMS products appearing based upon XML technology

They suggest that the diverse nature of the HE/FE sector website (and the same would hold true for a hybrid library site) should rule out consideration of products with a specific focus on e-commerce or those which consist of a single database as content repository for all applications. Products with a document management provenance should also be avoided as they are likely to be set up for the institutional intranet rather than the outward facing Web.

The CMS market is still immature, nevertheless, some consensus is beginning to emerge amongst public sector institutions as to core functions that should be required of any Content Management System.

Browning and Lowndes identify three core features of a CMS[7]

- **Versioning:** so that groups of individuals can work safely on a document and also recall older versions

- **Workflow:** so that content goes through an assessment, review or quality assurance process.

- **Integration:** so that content can be stored in a manageable way, separate from web site design 'templates' and then delivered as web pages or re-used in different web pages and different document types.

The CMS and non-roman scripts

Can a CMS handle Oriental languages?

Many Content Management Systems on the market employ UTF-8, which is an ASCII-preserving encoding method for Unicode (ISO 10646), the Universal Character Set (UCS). The UCS encodes most of the world's writing systems in a single character set, allowing you to mix languages and scripts within a document without needing any tricks for switching character sets. Unfortunately, some commonly used web browsers cannot handle UTF-8. At a recent NOF workshop on the challenges of designing web resources in non-roman scripts some of the shortcomings of even the most recent browsers were exposed when demonstrating multilingual websites. Internet Explorer, for instance cannot view pages which contain a mix of Java Script and UTF-8. Technical experts at the NOF Community Languages workshop drew attention to open-source browsers such as Mozilla[8], which often provide better UTF-8 support than more commonly used commercial browsers.

Laptop challenge[9] is an example of a UTF-8 multilingual site generated from a CMS called MkDoc[10]. Content editors were able to use keyboard mapping create and edit documents

---

[7] Browning, P & Lowndes, M. Content Management Systems: who needs them? *Ariadne* Issue 30, Dec 2001. http://www.ariadne.ac.uk/issue30/techwatch/intro.html [website last accessed 11/11/02]

[8] http://www.mozilla.org/mozorg.html [website last accessed 15/11/02]

[9] http://mkdoc.com/news/laptopchallenge/ [website last accessed 14/11/02]

[10] http://mkdoc.com/ MkDoc is a low cost CMS, at approx. £5,000 plus VAT [website last accessed 14/11/02]

in Bengali, Hordhac (Somali), Panjabi and Gujarati. In this particular application of the CMS the workflow was basic, without review or quality assurance because the aim of the event was to mount material on the web 'live' while the conference was actually in progress. Participants created non-roman script material in the multi-lingual editor Uniedit[11] using the editor's keyboard mapping facility. Information was entered into templates supplied by the MkDoc CMS, which then published them to the web.

There are also potential problems in dealing with Oriental scripts if a CMS has a component database such as MySQL. Such databases are not yet Unicode compliant and so would not be able handle non-roman scripts in Unicode without considerable ingenuity on the part of the programmers.

The CMS Market Place

Buy or build?

The CMS market at the moment consists of a large number of competing products ranging in price from Zope[12], which can be downloaded free, to big market players such as Vignette[13] at around £500,000. In an immature market the risks of buying from a vendor which does not survive or a product that turns out to be a technical cul-de-sac are considerable. Sticking to Content Management Systems that rely on open source applications can reduce the financial risks by creating content that does not have to be extensively re-engineered should the vendor cease trading.

The broad range of content contained in a hybrid library is likely to be handled better by the 'buy and build' solution, a product built on open source applications, with a certain amount of post-purchase customization. The fact that the NOF-digitise initiative is encouraging public sector heritage bodies to use Content Management Systems should create a body of relevant 'buy and build' case studies that could help in understanding the current market place. There is also a high quality, relatively low-volume e-mail discussion list devoted to Content Management Systems called cms-list[14], though contributions to this list can be extremely technical.

---

[11]Uniedit was the editor used for script input. Information about this editor can be found at http://www.humancomp.org/uniintro.htm. Unipad is another Unicode editor that can handle multiple scripts and can be downloaded from www.unipad.org/ [websites last accessed 14/11/02]

[12] http://www.zope.org/ [website last accessed 15/11/02]

[13] http://www.vignette.com/ [website last accessed 15/11/02]

[14] Details can be found at http://cms-list.org/ [website last accessed 15/11/02]

The Future

The Institutional/Distributed Resource Tension

Despite the resistance in traditional University IT departments noted earlier, it is likely that HE/FE establishments will eventually turn to the CMS for their websites. The dilemma for the oriental library is similar to that faced with the Library Management System. A large complex institution will place the need for a CMS that complies with the technical standards necessary for the management of oriental scripts extremely low on its list of priorities.

Opportunities to purchase Content Management Systems, which more closely match the needs of the oriental library are likely to come with externally funded projects. The NOF funded Digital Shikshapatri, for instance, may be sharing Oxford ArchDigital's CMS[15] with the RSLP/British LibraryCo-operation and Partnership Programme funded South Asia Through Official Eyes project, so that some of the Geographical Information System and datasets developed in the most recent phase of SATOE can be re-purposed for the Digital Shikshapatri site.

The advantage of the project driven CMS is that it is inevitably a closer fit to oriental library needs than a large, expensive institutional CMS. The disadvantage of the project driven CMS is that when large institutional Content Management Systems come to be installed there may be considerable reluctance on the part of institutional IT departments to host more than one Content Management System on their servers. The future of the project driven CMS may therefore be on a server outside the host institution. This means that there are ongoing costs after the project funding has ceased, whereas internally hosted resources can to a large extent be absorbed within the general IT overheads.

Conclusions

The CMS is here to stay

It seems that although the Content Management System market is relatively new, the CMS is here to stay. Oriental librarians have to prepare themselves for a future in which the functionality of the CMS will, like the Library Management System, become an area for negotiation within the larger institutional framework and any successful project bid to create an electronic resource will have to include some CMS provision.

Despite the potential pitfalls, the CMS brings us one step closer to the "one stop shop" for the user, where the interface does the hard work of integrating diverse electronic resources. It also offers the future in which subject librarians can escape the webmaster bottleneck to create and edit content in templates that will automatically be delivered in the institutional house style.

---

[15] http://www.oxarchdigital.com/ [website last accessed 18/11/02]