Content, content, everywhere... ...time to stop and think?

The process of web content management

by Steve Goodwin and Richard Vidgen

Most organisations have implemented a website. Many are now experiencing problems associated with maintaining the content of that website. We discuss what has led to web content management, what is driving it, and propose a WCM framework to help organisations to understand how to manage and control their websites better. Research involving the use of this framework is at an early stage, and we welcome any feedback.

All organisations that have a website (and that is nearly all organisations) must have a content management system, whether it be a manual process enacted by a single web manager or a decentralised system for producing and publishing web content supported by sophisticated technologies. WCM is primarily a process, not a product, and we define it as *'an organisational process, aided by software tools, for the management of heterogeneous content on the web, encompassing a life cycle that runs from creation to destruction'*.

As with many new IT trends, web content management (WCM) is in part a practical response to a pressing business problem—how to organise and manage large-scale websites—and in part a technology push on the part of software suppliers. Ovum in 2000 predicted that the market for content management tools would grow from the 1999 level of $475 million to $5·3 billion by 2004, while the Meta Group has predicted a market of $10 billion in 2004.

Early to market software suppliers, such as Interwoven (www.interwoven.com) and Vignette (www.vignette.com), are finding many others jumping on the WCM bandwagon. There is a plethora of companies providing WCM software, and the feeling of industry analysts in the first half of 2001 is that the WCM market is overcrowded and due for a shakeout as the industry leaders emerge over the next 12-18 months. Some of these offerings are extant software products that are being re-positioned (or possibly just re-branded) in the WCM domain. Some of the 'new' WCM products are rooted in document management, while others have developed from customer relationship management, e-commerce and software configuration management. In the crowded WCM marketplace there is some confusion about what constitutes WCM; many suppliers define it in terms of the product they sell—hence our definition above.

Our model is being developed to help organisations concentrate on the process of WCM when considering their options or evaluating products.

Issues in web content management

Many organisations have created a website and most have established some infrastructural support for their website, such as a web manager or a web services department. There has been an explosion of content on websites as the potential of the web for internal and external communication is recognised. For a website to 'live and breathe' it must be fed with new content and out-of-date content must be removed. Organisations therefore need to encourage the activities of content providers. However, increased activity in content generation has raised a number of issues, all of which can cause problems if not properly managed:
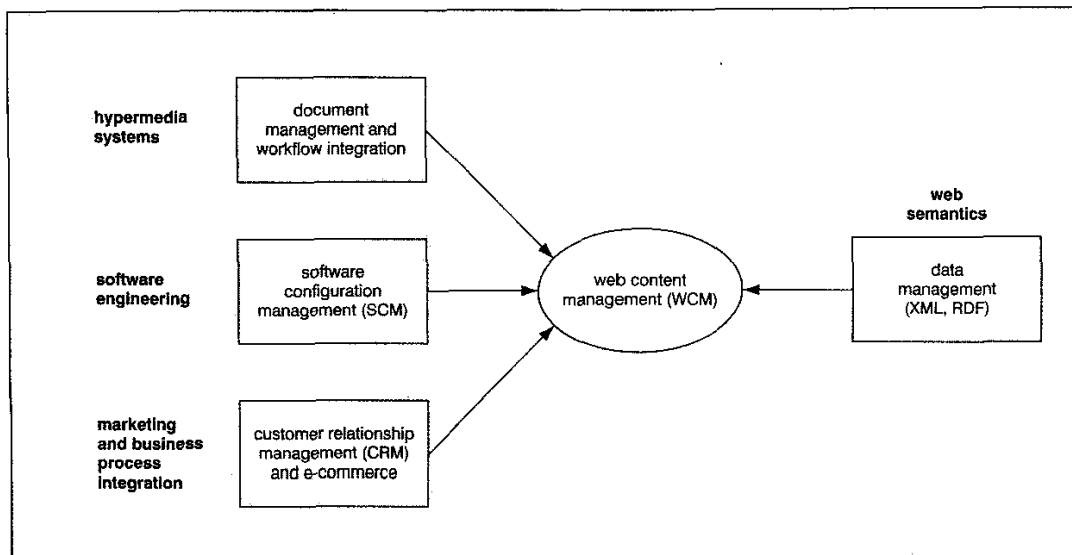
Fig. 1 Antecedents and enablers of web content management

- *Bottlenecks:* The web management function can become a bottleneck for content revision. Content arrives in different forms and has to be edited (usually manually) into a form suitable for publishing on the web. Funnelling content through a web manager resource can lead to delays in publishing on the web.
- *Consistency:* Where web editing is devolved to departments there can be inconsistencies in the look and feel of the site and variable quality of layout and content.
- *Navigation:* Where structure and content are not closely controlled, there is a danger that navigation and search capabilities will suffer. This is of major importance as, without these, it becomes hard for the user to find the required information, thus degrading the value of the entire intranet.
- *Data duplication:* In many cases, the content on the web is a copy of data held in a departmental or institutional system; changes to one system are manually replicated in the other systems. Ideally, data will not be stored redundantly in the organisation. There will be one source accessed by all business applications, whether internal or external. Where data needs to be copied then replication should be automated and controlled.
- *Content audit and control:* Unauthorised content may appear on the website. Material published on the web should be subject to a review and authorisation process to ensure that it is acceptable from a marketing and legal viewpoint. Procedures and controls need to be defined to manage the web publishing process.
- *Tracking:* To use content effectively it is necessary to know things about the content, such as who created it, when was it created and when was it last updated. The ability to track and reconstruct the changes that have

occurred to content is an important part of content management.
- *Business processes:* Content is often tied tightly to business processes. For example, the production of a market intelligence report is a complex business process, involving data collection, data analysis and the generation of commentaries and forecasts. Not only is the 'final' report published on the web, but also updates and revisions are likely to be needed on a regular basis. The business process and web content management need to be integrated, allowing content to be published internally for inspection and review and only released once it has been approved. Furthermore, the process itself may need to be redesigned to take account of differences between paper and web publishing.

Although many WCM packages are aimed at large organisations with tens or hundreds of thousands of pages, smaller organisations are also running into the problems caused by a lack of control over the content management process.

The genesis of web content management

WCM solutions must be capable of dealing with data with different degrees of structure. Some content will have a high degree of structure, such as employee records, and be amenable to being broken down into tables and stored in a traditional relational database system. Other content will be low in structure, such as a video clip of an Internet consultant talking about security issues. In between will be range of content, such as a health and safety manual or an in-house magazine, which will display greater or fewer degrees of structure.

A WCM must be capable of managing content in different media and with different degrees of structure, which suggests a combination of traditional data management and document (hypermedia) management technologies. Many e-commerce products have specialised WCM facilities to address the maintenance of product and customer data. There is also much to learn from software configuration management, which is concerned with aspects of change such as versioning and audit trails. From an enabling perspective, technologies that support semantic interoperability, such as XML (eXtensible Markup Language)[1] and RDF (Resource Description Framework)[2] provide a basis for sharing and the automated interchange of content between partners, i.e. humans are not the only users of Web sites.

Fig. 1 shows how the three antecedents of WCM—document management, customer relationship management and software configuration management—together with enabling technologies and web semantics can be combined to provide a platform for WCM systems. The antecedents also point to the range of disciplines that are encompassed by WCM. Research areas in hypermedia and document management, software engineering, marketing and business process design are all relevant, underpinned by data management and web semantics. This diversity of interest suggests that, although the idea of WCM is intuitively simple to grasp, the different emphases and permeable boundaries will make it a difficult area to tackle both in research and in practice.

A web content management systems framework

The integrated view of web content management developed above provides a basis for developing a framework that brings together a range of content management related issues and requirements (Fig. 2).

Content lifecycle

The core is the content lifecycle and a logically unified repository. The content lifecycle covers creation to destruction of content components. New content will arrive from a number of sources, including electronic documents (e.g. from word processors, such as MS Word), paper documents, templates (e.g. press releases, new product descriptions), web design tools (e.g. MS FrontPage, Macromedia DreamWeaver), and direct edit on the web into the repository. In many cases, the source will require some sort of *review*. This may be a review in terms of acceptable content, or it may be a review to determine the optimum place for the data within the structure of the system. The original data may require *storage* prior to publication and this may need to be continued after publication if it is published in a different form. With regard to *publication*, apart from the obvious requirement of making the content available, this should include:

- *Authentication*—which is concerned with identifying the user through a mechanism such as a user ID and password or biometrics.
- *Personalisation*—which relates to the ability to present different users with different views and different data depending on preferences, access profiles, role, previous accesses etc.
- *Transformation*—which is concerned with constructing (e.g. combining sub-components into new documents) and transforming content at the time of delivery.

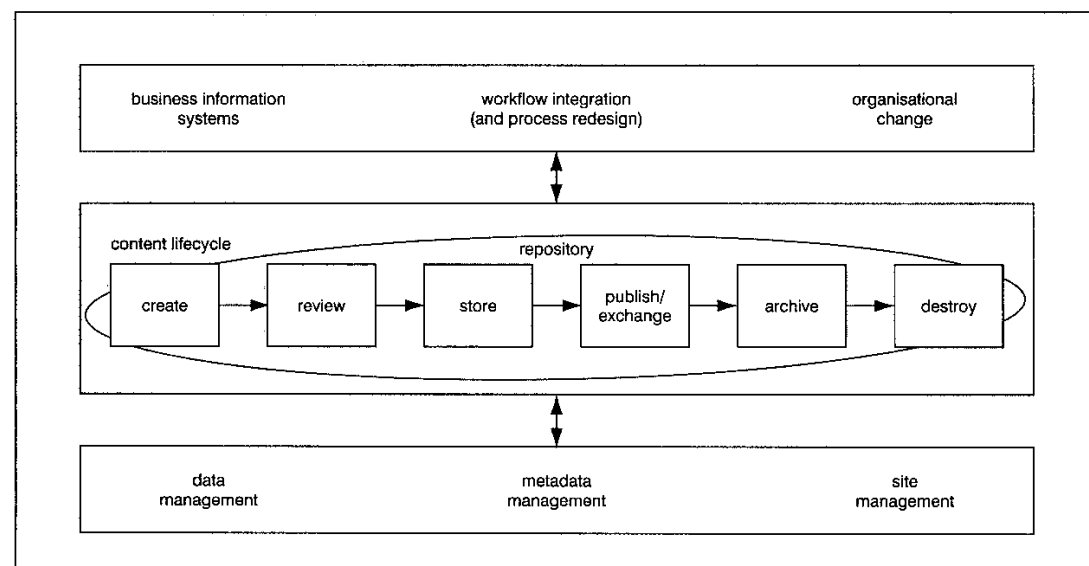There is often a requirement to *archive* the data. This

Fig. 2 A framework for web content management

could happen automatically after a given date, or it could be a manual process, with the archives stored online or offline. Finally, at some stage, there may be a need to remove (*destroy*) the content permanently. The repository is a collection of data stores that cater for components with more or less structure, including relational and object-oriented databases, document stores, file systems etc. The WCM must give seamless access to the content components regardless of where and how they are stored.

Organisational integration

Content is generated by and in support of business processes, which indicates a workflow and application perspective. To take advantage of the web effectively it is likely that some of the workflows will need to be redesigned, possibly to incorporate computer mediated communication and collaboration. This suggests that content management may go beyond routine automation and require, be constrained by, or initiate organisational change. As with any information technology, the consequences are likely to be unpredictable and are as much social as they are technical. The WCM core (content lifecycle and content component repository) must therefore be integrated with business processes and workflows and the existing business information systems that support those processes.

WCM process management

Management of the WCM process suggests a range of activities that need to be catered for, including data management, metadata management and site management.

Data management is essential if content is to be exchanged, reused, shared and searched intelligently. It is increasingly likely that web content will be marked up using XML and presentation will be defined using the style language XSL. To define and manage the XML data structures a modelling notation is required, such as the Unified Modelling Language (UML). If UML class diagrams are used to model the data structures (possibly represented in RDF), then XML documents can be thought of as physical implementations of an underlying and consistent enterprise class diagram. Connallen (www.rosearchitect.com) has shown how UML diagrams can be mapped into the hierarchical structure of XML documents. These hierarchies need to be shown to be consistent with the class diagram, i.e. that the UML schema can be derived by navigating through the class diagram. The combination of UML/RDF/XML is a promising approach to enterprise-wide data management.

Metadata is the information that needs to be stored about a data item for an agent to use it; for example: expiry date, source, revision history, title, keywords, date created, time to archive and version number. The metadata might also define how content can be combined to make new/virtual documents; for example, can this section be combined with that one? In what order? Is there a pre-requisite/co-requisite? 'Bursting' documents into subcomponents has great potential for data reuse but is going to be difficult to implement and control. Consequently, content management systems should be expected to have sophisticated metadata capabilities.

Site management is concerned with Web site design and structure. Content must be separate from style and be device-independent. It should be possible to change the look and feel of the site by changing a style sheet. It should be possible to add a new device, such as a personal digital assistant (PDA), without affecting the content structure.

The content lifecycle suggests that there will be a number of roles to be considered: content contributor, content publisher, content consumer, web designer, web developer, content manager, design/navigation manager. This list is not necessarily exhaustive and neither does it imply that, in all cases, all these roles will require separate individuals. However, the list does indicate the type of roles that will need to be considered in most cases in managing the WCM process.

Illustration: intranet development and content management

Our research is at an early stage and, although the model has been discussed with a number of organisations, as yet we have not had the opportunity to apply the model within an organisation that is restructuring its WCM process. However, in the example that follows, we have applied the model—after the event—to an intranet development programme that has been ongoing for some time. The company concerned is an autonomous division of a major multinational organisation.

Early in 2000, senior executives within the division decided to 'reinvent' the intranet and take it forward in a more proactive manner. Views tended to be short term: 'Let's just get something up and running' was a common comment. However, despite this short term-ism, a top-level structure was formulated with the agreement of all of the relevant departmental and business process managers. Furthermore, each of the managers agreed to take responsibility for the content of their areas.

The project started well. A prototype site was built and agreed and templates were provided to allow various groups to contribute content in a consistent fashion. However, despite advice to the contrary, further progress was poor and that can be explained through the following factors:

- there was little or no investment
- executive interest and attention drifted away
- there was no project champion
- there was little or no central control.

These are common issues in IS implementation and are well-documented in the IS literature.[3-6]

Some sections of the intranet, typically those with proactive owners or resident 'techies' who liked to play with web pages, expanded dynamically and at the end of 2000 were up-to-date and perceived to be useful. Other sections of the intranet had, after 12 months, no content whatsoever.

In terms of content management, the WCM process is manual, localised, and has little overall control. A common site look and feel and consistent navigation is achieved in part through web page templates, although this still requires content providers to be adept at using a web-authoring tool. Groups that happen to have web-authoring skills do quite well in terms of content, but frequently they rely on a web enthusiast whose actual duties are often very different. This can lead to a loss of activity in the areas of work that the content provider should be involved in and an under-recording of the time actually spent in generating and maintaining content. Groups without web-authoring skills have to rely on the very small central 'corporate' resource. This underlines the problems of the 'web manager bottleneck' mentioned above.

The intranet is providing value in localised areas, but has not matured into a corporate resource and first point of call for help and information.

The WCM framework as a diagnostic tool

In our case study, it is clear that there has been little thought about many of the areas of Fig. 2 in the organisation we considered. There has been almost no thought of business processes and workflows other than in the initial choice of top-level divisions of the intranet. It has been considered almost entirely as just another means of distributing information rather than as a possible tool to streamline and redesign processes.

There have been culture clashes between the decentralised tradition of some departments and the more centralised needs for a corporate intranet structure. In a recent paper, Rosabeth Moss Kanter[6] describes one of the ways to fail in moving your company onto the Internet, which is equally true in this intranet case:

Under the banner of decentralisation and business unit autonomy, reward each unit for its own performance, and offer no extra incentives to co-operate in cyberspace. Keep reminding divisions that they are separate businesses because they are different, and that's that.

Only in one or two small areas has there been an alignment between technology, structure and needs.

If we now look at the content lifecycle from Fig. 2 we again find discrepancies. First and foremost, there is no central strategy for, or control of, content. It is created randomly, rarely reviewed and published in an ad hoc manner. No one has thought about archiving or what happens to out-of-date data. Old items are rarely removed. There is no concept of data or metadata management other than a basic, free-text search engine. Site management is poor due to lack of skills, resources and experience.

Finally, we see little definition (or even understanding) of roles within the company. There are, by definition, a significant number of real or would-be content contributors as there are content consumers. However, the roles of content publisher, web designer, content manager and design/navigation manager are combined into a very small part of one person who has little understanding of, and who has had no training in, any of them.

Summary

Web content management (WCM) is a significant business issue for any organisation that maintains a website that has a large number of pages or has frequently changing content provided by multiple providers. WCM technologies are emerging with background in document management, customer relationship management and software configuration management. The WCM market is crowded and it is likely that there will be a fall-out of suppliers as the market matures and WCM concepts are clarified. This suggests that an investment in a WCM system today could be as risky as it was necessary. However, content management will become a more and more significant issue for organisations as they use the Internet to support more of their business activity. We believe the framework developed in this article to be a useful starting point for thinking about a WCM system strategy and, as such, we would welcome any feedback.

We accept that our case study is a double negative—'you didn't follow our model and you didn't have much success'. Despite this, we feel the model may be of use to organisations and we look forward to hearing from any readers who would like to help us with a positive case study.

References

1 VIDGEN, R., and GOODWIN, S.: 'XML: what is it good for?', *Computing & Control Engineering Journal*, June 2000, **11**, pp. 119-124
2 SWICK, R.: 'RDF: weaving the web of discovery', *Networker*, June 1999, pp. 21-25
3 FARBEY, B., LAND, F., et al.: 'How to assess your IT investment: a study of methods and practice' (Butterworth Heinemann, Oxford, 1993)
4 KOTTER, J. A. H.: 'Corporate culture and performance' (The Free Press, 1992)
5 THORP, J.: 'The information paradox' (McGraw-Hill, Montreal, 1999)
6 MOSS KANTER, R.: 'The ten deadly mistakes of wanna-dots', *Harvard Business Review*, 2001, **79**, (1) p. 91

Steve Goodwin is Managing Director of the Houndscroft Partnership. He also teaches and researches part-time at the University of Bath School of Management. He is an IEE Fellow; e-mail: sgoodwin@houndscroft.co.uk
Richard Vidgen is a Senior Lecturer in Information Systems at the University of Bath School of Management; e-mail: mnsrtv@management.bath.ac.uk.