a better way

From Enterprise Content Management to Effective Content Management

by Seth Gottlieb

The vision of enterprise content management (ECM) as a single system to manage all content became popular in 2000 and 2001. In theory, having a single, centralized system to organize and manage all of an organization's content into an accessible and navigable electronic library is very attractive. Companies must manage a wide array of assets, including internal how-to documents to support business processes, internal and external correspondence, marketing literature, and product documentation. Organizations are buried in digital content, leaving people scrambling to find the right information when they need it.

Both vendors and customers enthusiastically embraced the ECM vision of transforming content from a chaotic mess to an organized and powerful asset. Recently, however, this vision has been the focus of substantial criticism from customers and industry analysts alike. In practice, implementing one system to manage all electronic content within an enterprise has turned out to be impractical, risky, and expensive. As a result, customers are turning to solutions with less lofty claims and more reasonable costs. Increased attention to middle market vendors and open source software is evidence of this trend.

WHAT IS ECM?

Despite its widespread use, there is very little consensus on the meaning of the term "enterprise content management." Seasoned CM professionals typically describe ECM as a set of best practices and tools to manage documents. These practices address authoring, formats, workflow, storage, publication, and archiving. Steve Manning of the Rockley Group lists ECM, along with Web content management (WCM), digital asset management (DAM), and learning content management (LCM) as distinct categories of content management [4]. The Association for Information and Image Management (AIIM) defines ECM as "technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes" [1]. In other words, the "enterprise" refers to the subject of the content rather than all content contained within an enterprise.

In 2000, ECM became an industry buzzword to describe a class of commercial software that could "manage all of your content types for all of your business challenges" [6]. How did ECM go from being standards and practices for managing content to a solution to manage all content? There were a number of market forces at work.

First, companies found themselves swimming in documents but starving for information and not sharing across departmental boundaries. Information is power, but it is useless when scattered across employee hard drives and e-mail in-boxes across the enterprise.

CONTENT MANAGEMENT DEFINITIONS

Web content management (WCM) refers to technologies that manage content on a Web site. WCM systems tend to be "page" or "article" focused and use a publishing paradigm of workflow and approvals. During the Internet bubble, WCM attracted much attention in the CM space.

Records management (RM) refers to technologies and processes used to store records (such as e-mail and digitized copies of paper documents) over time.

Document management (DM) refers to technologies used to store documents. Typically these systems support basic file storage plus versioning, access control, and basic metadata capture.

Digital asset management (DAM) systems establish a library of reusable digital assets with a focus on image files (such as logos) and media files (audio and video).

The large WCM companies, such as Vignette and divine, had reached the apex of their market and looked at enterprise content as an attractive sector to pursue. The large document management companies, such as Documentum and Interwoven, retaliated by trying to break into the WCM space. The two camps launched a feature checklist war. To be recognized by industry analysts as a player in the ECM sector, nearly every content management system (CMS) vendor claimed it could holistically solve the crisis of content growing out of control. This content crisis resonated with customers who were willing to suspend their disbelief in the hope that ECM really was the magic bullet to solve it.

THE RACE TO ECM

- **December 2001:** Documentum acquires Bulldog for DAM.
- **April 2002:** FileNet acquires eGrail for WCM.
- **October 2002:** Documentum acquires eRoom for collaboration.
- **December 2002:** Vignette acquires Epicentric for portal capabilities.
- **June 2003:** Interwoven acquires MediaBin for digital RM.
- **August 2003:** Open Text acquires Gauss for WCM.
- **August 2003:** Interwoven acquires iManage for DM.
- **September 2003:** Vignette acquires Intraspect for collaboration and knowledge management.
- **February 2004:** Vignette acquires Tower for DM.
- **August 2004:** Interwoven acquires Software Intelligence's RM systems.
- **August 2004:** Open Text acquires Artesia Technologies for DAM.

THE "ERP OF CONTENT"

From a marketing perspective, ECM is an excellent term because it connotes "enterprise class" and creates an up-market brand. Targeted CMS products that were worthy of consideration by large companies were effectively excluded because they didn't carry the ECM label. ECM became the "ERP of content." ECM vendors asked customers to standardize on an architecture or suite of products and, in return, promised efficiency, control, and better intelligence. To fulfill this promise, ECM vendors extended their products or acquired other products to deliver functionality in what came to be defined as the three areas of ECM: WCM, DM, and DAM.

But there are several inherent flaws in the vision of becoming the ERP of content. First, ERP is focused on automating and supporting certain well-defined and/or highly regulated business processes (accounting, human resources management, manufacturing, etc.), and there is standardization on the structure of the data, the organizational roles of its users, and how it is used. While the business environment is starting to change in regard to regulating the management of content, content management is still less defined. "Content" is an abstract term that spans a diverse set of assets, users, and uses. Everyone within an enterprise is responsible for producing or consuming some sort of content. Users do not think of all these assets as being variants of "content" to be handled in a uniform way. Why should the system? Managing content is often inseparable from the business processes and organizations that the content supports. Consequently, content management is a specific, rather than general, enterprise problem. Sharing information across departments is an enterprise problem, but ECM is not necessarily the solution. While executing an ERP project is a formidable challenge,[1] ECM projects could be considered even riskier because of the diversity of requirements, stakeholders, and goals and the comparative difficulty of measuring the benefits. It is difficult to align the goals of project sponsors from across the enterprise, and as a result, ECM projects are hard to sell.

PROJECT SCOPE AND TIME TO BENEFIT

When ECM tries to universally solve many distinct CM problems across the enterprise, the number of stakeholders becomes overwhelming and ECM projects grow large and complicated. Requirements are difficult to collect across disparate user groups with diverse processes. Analysis takes time. Each group must make compromises that it would not need to make if the system were built just for its use.

The result is that the business waits a long time to receive a CMS that

[1] A survey of CEOs found that 65% believe that ERP systems could be harmful [2].

disappoints in the end. After the system is deployed, enhancement requests are often ignored as they are prioritized and reconciled with conflicting requests. Users no longer feel ownership of the system and look for workarounds that subvert the intent of a unified system. One survey found that more than a quarter of CMS owners were so disappointed with their systems that they were contemplating scrapping them and building custom systems in their place [3].

FROM "ONE THROAT TO CHOKE" TO "ONE NOOSE TO HANG IN"

For some buyers, a major attraction of a universal CMS or suite of projects sold by a single vendor is risk reduction. The components are proven to work together; and if they do not, there is only one phone call to make. However, many customers seeking this security have had a frustrating experience. The major ECM vendors have been going through a period of rapid change both in strategy and in technology. The pace of mergers and acquisitions (M&As) has outstripped their ability to integrate their components. For example, Vignette customers have been forced through difficult upgrades as the product architecture shifted from Tcl to ASP to Java, and the acquisition of Intraspect brought Jython into the mix. With one major product release, Vignette customers were asked to repurchase their licenses. Customers of FatWire's Content Server had to endure a harrowing journey of their product being passed from vendor to vendor through acquisitions and bankruptcies. At one point, a judge overseeing a liquidation auction forced Content Server, along with what is now SilkRoad's Eprise products, into a bundle with a hosting business.

This rocky history of M&As and technology change has reduced the benefit of "one throat to choke" to a risk of "one noose to hang in." An organization's content is too valuable to be controlled by a single vendor. Despite some vendor backpedaling, most analysts expect more M&A activity in the future. The content *will* outlive the vendor. Managing content with a single product that necessarily has a fixed lifespan is irresponsible. Customers would do best to hedge their bets and stagger their investments in CM solutions.

Analysts today are skeptical of ECM vendors' ability to integrate their acquisitions. In the opening keynote panel at the November 2004 *Gilbane Conference on Content Management Technologies*, IDC's Joshua Duhl began the discussion with the provocative statement "ECM is a myth," and a lively debate resulted. Most of the panelists concurred that they were not aware of an enterprise-wide ECM implementation that successfully managed all content types. Furthermore, several commentators asserted that it would be years before the ECM vendors could integrate the technologies they had acquired. Interestingly, in the April 2005 Gilbane conference opening panel, one analyst who had defended ECM in the 2004 debate made the point that there is no one-size-fits-all solution in content management. Two of the major themes of the spring Gilbane conference were specialization and content-centric applications.

Since then, ECM vendors have been toning down their ECM visions as well. FatWire has repositioned its Content Server product, an ECM aspirant back in 2001, as a WCM tool. New ECM entrant Oracle calls its product "ECM Light" and "ECM for the rest of us" by adding versioning, simple workflow, metadata capture, and improved authentication to the file system paradigm to create an easy-to-use, inexpensive document management solution. Vignette's marketing site now centers on specific business problems such as Web site and brand management, employee intranets, and customer self-service Web sites.

TIME TO RETHINK

While the major ECM vendors were focused on offering more breadth, many forward-thinking customers started to look for specialized and usable CM solutions and ways to share information between them. They found usability a key success factor in their CM initiatives because their internal users found ways to work around unusable systems.

Two of the most interesting trends in content management are open source and hosted solutions (such as Atomz and CrownPeak). Middle market products that focus on a particular discipline within content management (such as WCM, DM, or

DAM), which had been languishing in the shadows of the major ECM vendors, are also getting more attention because of their fitness to purpose. Targeted solutions tend to be cheaper and have the potential for greater usability if they are designed to solve a specific problem.

A HEALTHIER APPROACH

ECM is a strategy to eliminate content stovepipes, which reduce efficiency within an organization. However, if such stovepipes are interconnected with open APIs and standards, they become a "federated" application infrastructure that achieves the same benefits without the large initial investment and risk of consolidation. There are many public examples of third-party applications that integrate content from disparate sources to provide a more holistic view of information. For example, consider all the shopping aggregator sites that provide the shopper a unified experience of product catalogs from multiple sources. Note that most of the information sources were not originally designed to have their content used in this way, but open standards and architectures (at first, simple HTTP and now full APIs) make it possible for existing content to have new value.

A number of strategies and tactics can help an organization establish a federated content architecture that integrates content across the enterprise without forcing users into one infeasible solution. With the federated approach, managers can use best-of-breed technologies as they emerge with less disruption to the organization. The following are some key concepts and technologies to consider.

Metadata, Taxonomy, and Structured Data Capture

No matter what the technology, any successful organization-wide CM initiative (ECM or otherwise) requires a holistic view of the content that the organization intends to manage. One practice that helps develop this perspective is defining a taxonomy to classify content. By having a common language to organize and describe content, it is easier for consumers of content, be they users or other systems, to find what they need. Creating a taxonomy is tedious work that requires input from across the business. However, it is essential to have a holistic view of content even if it exists in different systems. This process often has two positive side effects for managers: it maps out organizational knowledge and enables a better understanding of business processes.

In addition to taxonomy, the organization should capture other metadata that describes content, how it should be displayed, and to whom. Many existing CMSs can be configured to enforce the same metadata collection requirements and publish this metadata along with the content. The Dublin Core is a good place to start.[2] Additional metadata may enable more sophisticated content reuse and integration, and metadata can be captured at the entire asset and subelement level. However, extensive metadata capture requirements will make the system more difficult to use and less likely to be adopted. As always, organizations must strike the right balance.

The more structured the content, the more potential there is for reuse. Storage formats that separate content structure from layout and break the asset down into granular chunks can be used by different systems to be published in a variety of formats. Using open formats for storage (such as MP3 over Real Media) will increase the likelihood that other systems will be able to read the content assets.

Enterprise Search and Portals

The most common and straightforward technology for aggregating content from disparate sources is enterprise search. After all, search made the Web — a potentially chaotic universe of independent publishers — an amazingly useful resource. The quality of the result that a search tool can provide is greatly enhanced by metadata that accompanies the content. Search-based content integration is most effective in environments with simple authorization rules and network security for viewing content. If the organization uses an application security approach instead, IT

[2]The Dublin Core is a standard set of metadata that describes the content, origin, and purpose of a resource. Examples of Dublin Core elements are title, coverage (the topics that the resource covers), source, date, audience, publisher, and language. The Dublin Core is managed and promoted by the Dublin Core Metadata Initiative (http://dublincore.org) and is widely used in systems that manage and distribute content.

professionals will have to manage authorization logic needs within two systems: the search system and the source system.

Portal technologies can deliver more fine-grained access control and tighter content integration. Through standards-based architectures, such as the emerging Web Services for Remote Portlets (WSRP) and Java Specification Request (JSR) 168: Portlet Specification, it is possible to embed both content and application logic (such as authorization rules) in aggregated pages. The portal architecture has become mainstream in large enterprises, and the infrastructure is highly leveragable for content integration strategies.

Emerging Publishing Standards That Enable Information Sharing Between Systems

Beyond portals, there are several other emerging standards to help Web-based applications make their content accessible to other applications. XML-based publishing formats such as Rich Site Summary (RSS), Atom, and Resource Description Framework (RDF) may change the CM technology landscape from monolithic architectures to networks of content applications that share information. Most of these standards-based technologies can be retrofitted onto existing applications. If these trends take root, and they appear to be doing so, the advantage of having a centralized universal CMS will diminish rapidly. Of course, the quality and standardization of the metadata capture and the openness and structure of the storage format will drive the degree to which content can be shared and reused.

API-Level Integration

There are numerous examples of companies that are integrating their CMSs at the application level to develop a comprehensive content infrastructure. An excellent example of an application-level integration strategy is financial services company Wachovia and its Content Access Services system [5]. Through acquisition of several large financial services companies, Wachovia inherited a heterogeneous group of document and digital records management systems, including IBM ImagePlus, FileNet, and Mobius. Wachovia's data on customers was scattered across several systems, while its customer support representatives needed a unified view to serve the customer. Rather than try to migrate all the data into one system and retire the rest, Wachovia developed an abstraction layer that provides a common interface accessible to desktop call center support tools. This architecture minimizes migration costs, leverages existing investments in content technologies, and enables Wachovia to select best-of-breed solutions and embrace new technologies that meet the requirements of specific business groups.

Web services are also becoming an important standard in integrating content technologies. SOAP interfaces can be built on existing systems to expose various functionality. Many open source and commercial CMSs natively support SOAP APIs. Through the use of Web services standards like Web Services Flow Language (WSFL), it is easier to coordinate different systems to support complex business processes.

Open Repositories

In July 2002, JSR 170 was submitted to the Java Community Process (JCP) for a standard Java Content Repository (JCR). This standard creates a content repository that different CMS products can use to store content much in the same way that a SQL-compliant database is used by applications for saving data. The specification is led by Day Software (a CMS vendor), and many of the major ECM vendors (Vignette, Documentum) and companies that are now getting into the ECM market (Oracle and IBM) participated in the expert group. JSR 170 is currently in the proposed final draft stage and is expected to be finalized in the next couple of months.

Once final, companies such as IBM are expected to offer JSR 170–compliant repositories. Jackrabbit, an Apache Software Foundation project to build a JCR reference implementation, is progressing nicely, and Day Software's CMS and several open source CMS projects (Magnolia, Apache Lenya, Graffito) are integrating with it. In addition, Day has just released a beta version of its commercial JCR.

The existence of a common content repository with a standards-based API could change the face of ECM from an application-centric architecture to a data-centric architecture in which multiple applications can read from the same repository or applications can access multiple repositories. The JCR would reduce vendor lock-in and help companies rapidly develop content-centric applications that solve specific business needs.

Web-based Distributed Authoring and Versioning (WebDAV), a standard defining the HTTP extensions that enable reading and writing files from a repository, is already widely used in commercial and open source software. When used with XML-based content, WebDAV can be a powerful shared content repository, accessible from a number of clients and applications.

CONCLUSION

In this age of information and the knowledge worker, companies that cannot make more effective use of content will be operating at a distinct disadvantage. The risk of information loss, redundancy, and misinformation necessitates a sound content strategy for the enterprise. However, the ECM vision of one system to manage all content is not the only solution, nor is it the best solution, because it introduces costs and risks that are not worth the benefit. Full and successful ECM implementations are rare, if any exist at all.

The term "ECM" will continue to be used to describe best practices within the processes of managing content (capture, manage, store, deliver, preserve), but the assertion that one system can do it all is rapidly losing credibility. The trend is toward a more distributed, rather than consolidated, approach of targeted content applications that are made interoperable by open standards and technologies.

REFERENCES

1. "About Enterprise Content Management (ECM)." AIIM Web site (www.aiim.org/about-ecm.asp).

2. Cliffe, Sarah. "ERP Implementation: How to Avoid $100 Million Write-Offs." *Harvard Business Review*, Vol. 77, No. 1, January-February 1999, pp. 16-17.

3. Howard, Jim. "'ECM' — Don't Buy It." *CMS Watch*, 26 March 2003 (www.cmswatch.com/Features/OpinionWatch/FeaturedOpinion/?feature_id=87).

4. Manning, Steve. "What's the Best Content Management System? It Depends ... " *The Rockley Report*, Vol. 1, No. 1, March 2004 (www.rockley.com/TheRockleyReport/V1I1/Tools%20and%20Technology.htm).

5. Walter, Mark. *Wachovia's CAS: Harnessing the Value of Multiple Content Repositories Across a Large Enterprise*. Content Technology Works report, Gilbane, 2004 (www.gilbane.com/case_studies_pdf/CTW_Wachovia_Final.pdf).

6. West, Kelley, Brook Foust, and Christine Klima. "Understanding the ECM Market." *AIIM E-DOC Magazine*, November/December 2002 (www.edocmagazine.com/vault_articles.asp?ID=25375).

Seth Gottlieb leads the Content Management and Collaboration Practice at Optaros, Inc. Optaros is a consulting and systems integration firm based in Cambridge, Massachusetts, USA, that helps organizations lower their technology costs by building more flexible and higher-quality systems based on open source software. With 10 years of IT and technology experience in both software and professional services, Mr. Gottlieb has helped numerous companies improve the efficiency and effectiveness of their CM and publishing processes.

Mr. Gottlieb can be reached at Optaros, Inc., 155 Second Street, Cambridge, MA 02142, USA. Tel: + 1 617 225 2455; E-mail: sgottlieb@optaros.com; Web site: www.optaros.com.