

## Best Practices for Identifying High-Value Content in an Enterprise Information Management Initiative

Toby Bell, Debra Logan

Enterprises interested in content strategy should start with valuation — especially before migrating content toward new repositories. They should calculate the present cost of storing/managing a high percentage of low-value content and factor this against the future benefits of a comprehensive approach to governance.

### Key Findings

- Controlling the content under management can yield significant savings in storage and administration.
- Organizations should pare down to essential content before "standard" enterprise content management (ECM), business process management, basic content services (BCSs), or content-enabled vertical application (CEVA) implementation projects begin (see Notes 1 and 2).
- Compliance and legal discovery can be milestones toward a content strategy but shouldn't be the destination. Achieving more business value from content is the ultimate goal.
- Integrated content archiving platforms are an option to consider for e-mail and file-share archiving and management. These systems can help companies reduce content volume and therefore cost, in addition to making it easier to produce regulatory and legal content, separating high-value from low-value content.

### Recommendations

- Content has no value unless it is used within business processes. First, you have to determine which content your people and systems should use and which they shouldn't.
- Your company will need a combination of products and functionality to achieve full control over content and to manage it according to its purpose or value.

## ANALYSIS

---

**This document was revised on 3 June 2009. For more information, see the [Corrections page on gartner.com](#).**

Among the benefits of disruptive technology and economic turmoil are the conversations between IT and business leaders they can engender. The content management technical landscape is undergoing a disruptive change, with a move away from file shares and e-mail toward the broad implementation of basic content services. Archiving systems have proved themselves to be effective for the management of e-mail. Many of the e-mail archiving vendors are adding the ability to include other types of content in their repositories. The promised payback will be in lower storage costs, easier retrieval and quicker response to business needs. The economic turmoil we are now experiencing has focused organizations on cutting costs wherever they can find them. Enterprise architects, business analysts, records managers, infrastructure leads and representatives from legal departments are struggling with the complexities of redefining a content architecture and managing its content with more discipline, even as they add new e-discovery, taxonomy, search, e-mail management and collaboration tools.

The valuation benefits are easy to see from a forward-looking perspective. Although it takes evangelism, exhortation, training and support, enterprises are now making some headway with implementing basic content services, reducing their dependency on e-mail as a document management and workflow system and making processes more efficient. They are beginning to use BCSs to replace e-mail as the workflow and document management system of choice, which many use as a default way to organize work and documents, mainly because they had nothing else. According to a user wants and needs survey conducted in 2006, 40% of respondents indicated that they believed Microsoft was the stack vendor best placed to satisfy their organization's BCS needs in the coming 12 to 18 months, with 50% of the midsize businesses we spoke to saying that they are using Microsoft Office SharePoint. Other vendors have begun to respond to the SharePoint challenge, particularly Cisco, Google and IBM, which are taking on SharePoint by offering content and collaboration services in the cloud. In any case, this market activity mirrors users' desire to get their unstructured content under control and managed. See Note 1 for references to this research.

But before any of this happens, business users and their IT partners must first discuss the use and value of content. Given the state of content repositories today, which is the result of many years of neglect, any execution plan that does not triage the content by its business value will not achieve a satisfactory result. Moving these massive content stores without first having these conversations will create a different and even more expensive and less transparent system.

There are three reasons for this advice being timely in today's economic climate.

- First, the "storage is cheap" argument is not working in these recessionary times. Why pay for new storage when you may be able to cull the data you have and free storage you already own for ongoing business processes? At Pfizer, for example, document storage volumes were reduced using content valuation. In "Case Study: Pfizer Takes a Two-Pronged Approach to Content Management," Karen Shegda writes "Content migration efforts required the business to pare down content volumes. Though the human effort required was not trivial, the benefits were worth the effort. As part of this assessment, Pfizer found that some documents in the organization had not been touched in five years." Similar findings from other companies support this (see "What to Do About E-Discovery and Legacy Information Management in 2009").

- Second, legal and regulatory pressure is increasing, especially in the financial sector. More and more attention is being paid to privacy, security and transparency, both in the U.S. and Europe. Details of these regulatory changes and their anticipated effect on IT can be found in "Repair the U.S. Privacy Patchwork" and "Data Retention vs. Privacy in the U.S. and Europe." The Obama administration will accelerate this trend in the U.S. After one day in office, President Obama issued two sweeping presidential memoranda which called for an expansion of openness, transparency and collaboration across the federal government (see "Obama Takes Quick Steps to Ensure Open and Transparent Government"). The new administration has also begun the work of reforming financial regulations, which many U.S. commentators agree will be far-reaching. Gartner's contact with its own government clients has certainly reflected these ongoing and anticipated changes. As for specifics, the new Securities and Exchange Commission ruling, promulgating the use of Extensible Business Reporting Language (XBRL), is an example of the kind of impact that the reform of financial reporting will have on organizations and their IT operations (see "XBRL Mandate on Financial Reporting Will Improve Transparency").
- Finally, the so-called information overload problem has finally reached the executive suite. Again, the financial crisis is one of the reasons behind this increased level of attention to information management. Some of the early bankruptcies in the mortgage market have been blamed on the poor management of enterprise information (see Notes). Although this is better demonstrated through a risk example, it has become clear to upper-level management that the management of information is crucial to business value and success.

There are a number of things you might want to look for in individual unstructured information artifacts: named entities, such as customers, employees, suppliers; sensitive data; embedded dates; time and date stamps; and so on. The point is to try to sort the information into the categories we have called out before beginning a finer-grained analysis

## How to Value Information Assets

How does the enterprise decide what assets are most likely to result in economic growth for the company?

In "Strategic Information Management: The Information That Matters," Tina Nunno and Patrick Meehan have suggested four categories of information as a top-level classification. The authors suggest two criteria to determine which information is important: the **value of information**, which is related to the corporate goal it supports, and the **type of information**, or what it is related to. The value of information is mapped to four categories of interest: compliance, operation, growth and "no value" or nuisance. The type of information comprises four categories: employee, customer, product and financial. A useful exercise, and a simple one, might be to think about what types of systems map to what categories of value, and how this relates to the four entities of employee, customer, product and financial.

The key steps of your content valuation project should be (1) Understand what you have and categorize it; (2) Establish who "owns" content from a business perspective; (3) Evaluate content (that is, assign value, risk and cost); and (4) Review legacy content and decide what should be deleted or archived.

The first goal of your content valuation project will thus be to understand and categorize what you have at a very high level, and how information artifacts are being added. Although most commentators recommend ignoring the system of origin, or data type, in practice this is not possible and may not even be desirable. For content, think about application of origin, repository

and file type. Application of origin may be anything from Microsoft Word, Outlook and Excel to FrameMaker, XMetal or Quark. Repositories as such are file servers, mail servers, desktops, storage area networks, archiving applications and document management systems.

Start with simple rules that focus on metadata that is meaningful to human beings. Content creator, date of creation and the kind of system it is stored in, as well as the business process the content is used in, are all important. Try to distinguish "finals" from "drafts." Almost all content has a limited life span, so companies can eliminate waste by identifying content that hasn't been accessed over a period of time (say, a year) and deleting or archiving it. Early draft stages of documents may also be candidates for deletion. Content management systems can be configured to save the latest or final versions of documents and to either archive or purge the earlier drafts.

Legacy content management cleanup is a necessary exercise for most companies and can be done in parallel with the valuation effort. For existing content you should first create a list of file types that might contain what you are looking for (.doc, .ppt, .xls but not .exe), and also attempt to de-duplicate what you have. At the very least, you will need to do sampling of the existing stores, especially file servers.

Enterprise information architects should be involved in information governance decisions like content valuation, because an inventory of repositories, users and sponsors is necessary to plan effectively. It is likely that some new software — storage and content analytics, content migration, desktop search tools and records management among them — will need to be reviewed and managed as well. Longer-term, some metadata and policy management tools are likely to be added to the continuing content valuation project. But planners have found that the project doesn't begin with technology, but with people instead.

Recruit analysts from both business units and IT to calculate the cost, value and risk associated with the storage and provisioning of content. Consider IT measures that reflect the cost of storing and managing information back to the department or profit-and-loss center. In addition to company-specific criteria, the content analysts should consider:

- Audience size: Documents authored by technical professionals for the benefit of a large audience have higher value than casual content created for individual use.
- Authorship: Collaboratively generated content, with peer review and management approval, will have more value than documents created by individuals. Collaborative content includes contracts, proposals, presentations and so on, most likely already in repositories and network drives.
- Internal versus external: Content for customers, business partners, regulators or legal counsel demands greater consideration because it is generally more complete, accurate and clear than documents for internal consumption.
- Status: Content that's polished and ready for its audience should take precedence over content in draft. Although some business processes may require the storage of drafts, most often they get stored in repositories, e-mail and hard drives simply because the company lacks a process to eliminate them.
- Duplication: In any repository, there are bound to be large numbers of duplicates and drafts, which can be identified and, in most cases, eliminated.
- Quality: The quality of a piece of content is necessarily a subjective measurement, but to start, organizations should consider a few simple quality criteria. Quality is affected by age. Older documents are *usually* less useful than newer ones. Again, this is

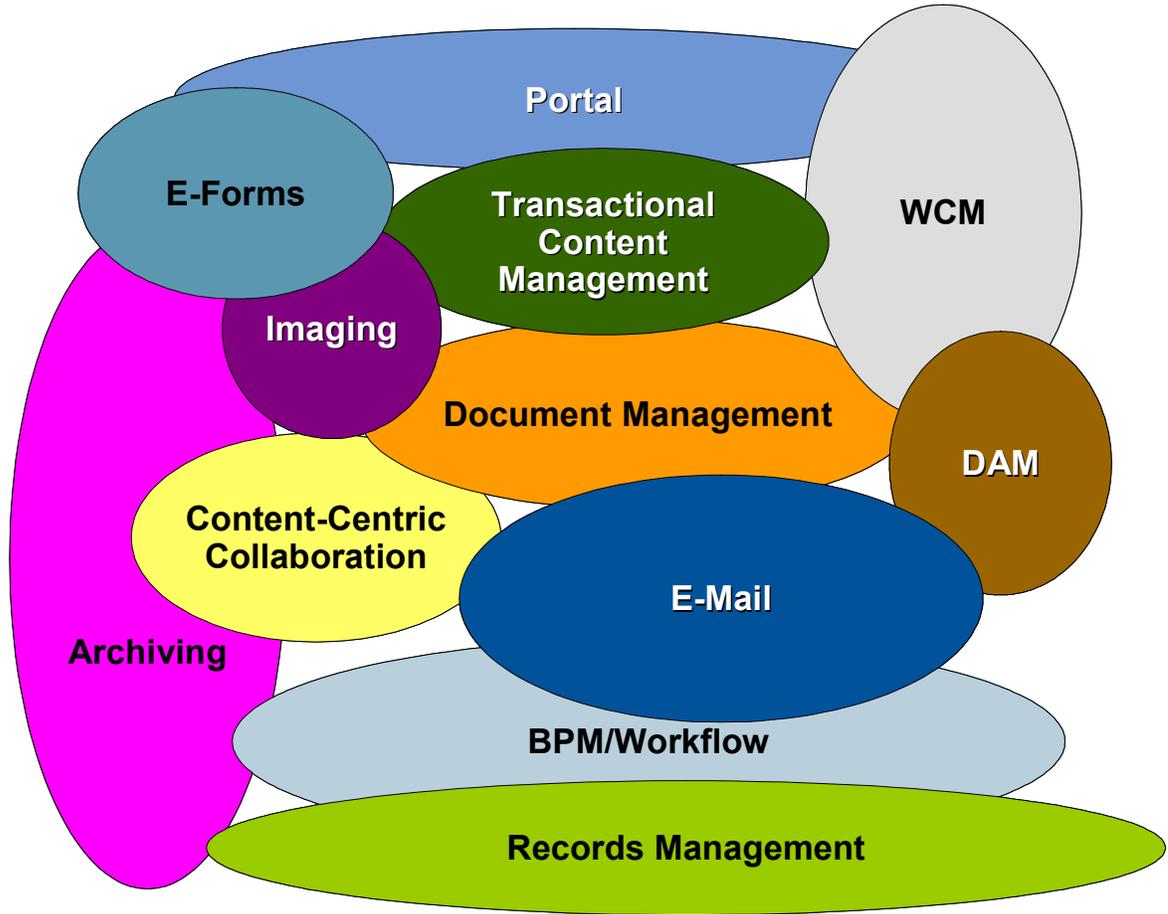
situationally dependent and can vary department by department, let alone business by business. Tracking access to a particular document over time is useful, but can't be done after the fact — that is, with an existing document. Finally, peer review is a good way to assess the quality of a final document. If a document has been peer reviewed and is the result of a process of review and editing, it will be of higher quality than other types of documents.

Using these criteria, the content analysts should scan the content stored in repositories, network drives and local hard disks, and eliminate what isn't needed for business continuity, compliance or reuse. Only after creating a content inventory and focusing on eliminating the toxic waste can the strategy of managing-in policies and procedures begin.

## **The Right Repository for the Job and the Repositories You Have**

One of the issues that prevent enterprises from undertaking content valuation exercises has to do with the variety and state of the various repositories they have. Figure 1 illustrates a subset of the possible repositories that you may be using. Enterprises undertaking content valuation exercises should map their repositories, by name, to the generic types in the diagram. Establish who "owns" the content from a business perspective and then begin the detailed valuation exercise. It may be that certain repositories will be little used, only serve a few users, or contain data that can be easily migrated into other repositories (archives) or eliminated altogether.

**Figure 1. Which Repository Will You Use?**



BPM = business process management, DAM = digital asset management, WCM = Web content management

Source: Gartner (May 2009)

There will always be multiple repositories and the number will increase, not decrease, in most instances, as users abandon one solution for another that they believe will "solve their problem." By solving their problem they mean that the repository does the work so that they don't have to. Unless and until users undertake the tasks involved in content valuation and in creating structures (taxonomies) and processes (publication workflow, life cycle management), content will continue to accumulate. Archiving e-mail and file-share content that has not been accessed for a certain period of time may be one way to think about the problem. We find, however, that clients are not keen to begin sweeping the masses of content that reside on shared drives into an archiving system, as then they are paying to store it twice, and everyone has the feeling that it is low-value, though usually no one can prove or disprove that with any data.

Using the techniques outlined in this report, you can overcome those hurdles, rationalize your repositories and your content, begin to save money and, most importantly, identify business value in the information that you have.

## RECOMMENDED READING

---

"IBM Plays Collaboration SaaS Card With LotusLive"

"Key Issues for SharePoint's Coexistence With ECM Systems Projects"

"IBM Retools Document Management Portfolio to Challenge Microsoft SharePoint"

"A Simple Method for Expressing Information Criticality and Classification"

### **Note 1**

#### **Basic Content Services**

BCSs provide document library services with complementary ad hoc imaging, basic Web publishing, document collaboration and document routing as their core functionalities. The key differences between BCSs and ECM are cost and ease of deployment. BCSs will serve the content management needs of knowledge workers who don't deal with mission-critical documents that are managed by entrenched ECM systems.

### **Note 2**

#### **Content-Enabled Vertical Applications**

CEVAs typically help to automate complex processes that previously required workers to sort through paper documents and other forms of content manually. CEVAs help reduce the cost of exception handling and optimize the rest of the work by applying better process controls and analytics. CEVAs are a mix of content, process and domain expertise, resulting in repeatable and reusable models and solutions. The term "CEVA" stems from earlier Gartner coverage of document-enabled vertical applications (DEVAs).

## REGIONAL HEADQUARTERS

---

### **Corporate Headquarters**

56 Top Gallant Road  
Stamford, CT 06902-7700  
U.S.A.  
+1 203 964 0096

### **European Headquarters**

Tamesis  
The Glanty  
Egham  
Surrey, TW20 9AW  
UNITED KINGDOM  
+44 1784 431611

### **Asia/Pacific Headquarters**

Gartner Australasia Pty. Ltd.  
Level 9, 141 Walker Street  
North Sydney  
New South Wales 2060  
AUSTRALIA  
+61 2 9459 4600

### **Japan Headquarters**

Gartner Japan Ltd.  
Aobadai Hills, 6F  
7-7, Aobadai, 4-chome  
Meguro-ku, Tokyo 153-0042  
JAPAN  
+81 3 3481 3670

### **Latin America Headquarters**

Gartner do Brazil  
Av. das Nações Unidas, 12551  
9º andar—World Trade Center  
04578-903—São Paulo SP  
BRAZIL  
+55 11 3443 1509