Hype Cycle for Content Management, 2008


Content management has evolved to include diverse content sources, content types and content-focused processes. This is a divergence from the exclusive focus on enterprise-specific resources, documents and document-centric applications. The technologies in this year's Hype Cycle illustrate this.
# TABLE OF CONTENTS

Analysis ............................................................................................................................................. 4  
  What You Need to Know ........................................................................................................... 4  
  The Hype Cycle ....................................................................................................................... 4  
  The Priority Matrix .................................................................................................................. 6  
  Off The Hype Cycle ............................................................................................................... 6  
  On the Rise .................................................................................................................................. 7  
    Enterprise Metadata Taxonomy and Ontology Management ............................................ 7  
    Information-Centric Infrastructures .................................................................................... 8  
    Video Search ......................................................................................................................... 10  
    Federated Search ................................................................................................................. 10  
    SaaS — Enterprise Content Management ........................................................................ 11  
  At the Peak .............................................................................................................................. 12  
    DITA-Aware Editors .............................................................................................................. 12  
    Mashup Applications ............................................................................................................. 12  
    E-Discovery Software .......................................................................................................... 14  
    Social Search ......................................................................................................................... 15  
    User-Generated Content (UGC) .......................................................................................... 16  
    Open-Source Content Management .................................................................................... 17  
    Basic Content Services ........................................................................................................ 18  
    Content-Enabled Vertical Applications ............................................................................... 18  
  Sliding Into the Trough ............................................................................................................ 20  
    JSR 170/283 ............................................................................................................................ 20  
    Content Integration .............................................................................................................. 20  
    Content Analytics .................................................................................................................. 21  
    ISO-Standard XML Formats For Office Productivity ....................................................... 22  
    XML-Enabled Database and Content Management Systems ............................................. 23  
    E-Mail Archiving .................................................................................................................... 25  
    Folksonomies/Social Tagging ............................................................................................... 26  
    Wikis ......................................................................................................................................... 27  
  Climbing the Slope .................................................................................................................... 28  
    E-Forms .................................................................................................................................. 28  
    Records Management ........................................................................................................... 29  
    Enterprise Content Management Suites .......................................................................... 30  
Appendixes .................................................................................................................................. 32  
  Hype Cycle Phases, Benefit Ratings and Maturity Levels ..................................................... 34  

Recommended Reading ............................................................................................................... 35

# LIST OF TABLES

Table 1. Hype Cycle Phases ........................................................................................................... 34  
Table 2. Benefit Ratings ............................................................................................................... 34  
Table 3. Maturity Levels .............................................................................................................. 35
LIST OF FIGURES

Figure 1. Hype Cycle for Content Management, 2008 ................................................................. 5
Figure 2. Priority Matrix for Content Management, 2008 .......................................................... 6
Figure 3. Hype Cycle for Content Management, 2007 ................................................................. 32
ANALYSIS

What You Need to Know

Many enterprises turn to content technologies to drive productivity around information. Content management applications collectively cover the entire content life cycle: from content concept through creation, to presentation, to retrieval, to archiving and disposal. They govern one of an enterprise’s most valuable corporate assets — its content.

This report answers the key question: What key market trends are changing the landscape of content management? It captures the maturation of technologies that were just concepts a few years ago — such as open source content management — and highlights technologies like e-discovery and content analytics software that are beginning to emerge in different enterprises. They play a key role in attaining such corporate goals as reducing costs and improving interactions in the enterprise ecosystem (individuals, departments, divisions and business units). Various technologies provide the content backbone essential to interacting with and supporting suppliers, partners, subcontractors, regulatory agencies and customers. These issues are of primary concern to business planners and IT architects.

This Hype Cycle report also covers mashups and user-generated content, unique content entities created by external organization sources. Their place in the enterprise is not so obvious. Analyses of newer content types and sources, such as video and social search, are also covered.

The Hype Cycle

Content management covers a range of technologies that drive business process productivity through effective use and collaboration around content. It includes enterprise content management (ECM) suites and basic content service (BCS). It has expanded to include search and content environments accessed through the Web. Increasingly, content management functions will be offered through emerging business models such as software as a service (SaaS).

Content integration is one of the bigger challenges; companies demand that their content be tied to mission-critical business processes, such as enterprise resource planning (ERP), customer relationship management (CRM) and HR systems. Content-enabled vertical applications (CEVAs) are the industry view of these technologies.

Organizations must integrate both structured and unstructured content, and they must do so by linking management applications (for example, content management with relational databases) as well as identifying content (with Extensible Markup Language [XML], for example) in ways that enable all applications to access it.

Web 2.0 content applications (such as folksonomies and blogs) have significant representation in the Hype Cycle.

Leading-edge organizations are strategizing on how to integrate and how to plan for future flexibility in this area by examining our research on information-centric infrastructure. This describes the alignment of technologies and functional capabilities that support consistent and seamless enterprisewide information capture, persistence, transformation and delivery.
Figure 1. Hype Cycle for Content Management, 2008

Source: Gartner (July 2008)
The Priority Matrix

Many of the technologies that are on the content management Hype Cycle are two to five years from maturity. However, the interrelationships among these technologies are the most challenging to understand. Assess your in-house products and how those products leverage different technologies. For example, how do content management technologies handle XML granularity? Can Web content (such as wikis or folksonomies) be managed as are traditional content structures? Enterprises must define the paths they will take through this technology to understand what they need from their vendors. By assessing the technologies and the R&D investment potential vendors are making, enterprises can understand vendor road maps and how they will affect their goals.

Figure 2. Priority Matrix for Content Management, 2008

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Years to Mainstream Adoption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 2 years</td>
</tr>
<tr>
<td>Transformational</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Mashup Applications</td>
<td></td>
</tr>
<tr>
<td>Content-Enabled Vertical Applications</td>
<td></td>
</tr>
<tr>
<td>E-Forms</td>
<td></td>
</tr>
<tr>
<td>ISO-Standard XML Formats For Office Productivity</td>
<td></td>
</tr>
<tr>
<td>XML-Enabled Database and Content Management Systems</td>
<td></td>
</tr>
<tr>
<td>E-Discovery Software</td>
<td></td>
</tr>
<tr>
<td>E-Mail Archiving</td>
<td></td>
</tr>
<tr>
<td>Enterprise Metadata</td>
<td></td>
</tr>
<tr>
<td>Taxonomy and Ontology Management</td>
<td></td>
</tr>
<tr>
<td>Source: Gartner (July 2008)</td>
<td></td>
</tr>
</tbody>
</table>

Off The Hype Cycle

- XML Rendering.
- Web-Based Distributed Authoring and Versioning.
On the Rise

Enterprise Metadata Taxonomy and Ontology Management

Analysis By: Mark Beyer

Definition: Ontology and taxonomy management is the ability to compare the context of one use case for an information asset with another use case. The tools for ontology and taxonomy management are still at a very early stage of development and the definition of the business need continues to specify expected features or functionality. The relative analysis of multiple ontologies and taxonomies recognizes that multiple, simultaneous metadata descriptions can exist for each information asset and it proceeds to reconcile them. The various metadata sources include business process modeling, enterprise information integration (EII), extraction, transformation and loading (ETL) and metadata repository technologies. Ontology management allows business analysts to leverage the value of these assets more effectively, while promoting improved understanding across business units and IT management personnel. Enterprise metadata management (EMM) addresses the problem of information assets created by different processes, defined by different business terms and interpreted through disparate semantics to produce competing taxonomies and ontologies.

Position and Adoption Speed Justification: Significant interest in metadata management has developed because of the emergence of service-oriented architecture (SOA), composite application deployment and other loosely coupled architecture concepts that need to address information management rights, business process flows and business context that were formerly deployed within discrete applications. When context is dependent on composite workflows instead of discrete processing tracks, the presence of multiple use cases and their resulting contextual definitions of both processes and information assets is guaranteed. One high business value use of metadata is found in the ability to justify and identify how decisions were made, based on information available at a given time. The new demand for metadata that describes end-user interpretations of "fact" will force the introduction of annotation metadata in daily workflows — for example, business process metadata.

From a traditional data perspective, most metadata management functionality is a feature of existing metadata tools limited to model extension with end-user defined, descriptive fields of interest and metadata versioning with no workflow or administrative enforcement beyond the development team. With the arrival of SOAs and the active use of metadata in registries built to support dynamic service virtualization, it will become imperative that the business becomes involved in linking the business process management workflows with information management workflows. This will force the development of new metadata management tools with user interface extensions to support multiple roles, adding a radically different interface layer for the business user. This will be especially true when reconciling information assets that are shared.
among multiple business processes and documenting any governance program that supports this activity. Through 2011, the number of registries/repositories that support federation will continue to grow and organizations will federate their metadata to receive big benefits, such as end-to-end traceability. However, enterprise-level metadata management is not a current priority for most organizations and, while tool use will increase over time, the culture change necessary (recognizing that metadata is a definition of perspective in how data is used) will take place very slowly.

User Advice: Identify data management and integration tools that include metadata repository management interfaces, supporting metadata model extensions.

Identify data management and integration tools that expose metadata repositories via application programming interfaces (APIs) and service calls, rather than through metadata import/export functionality only. Registries and repositories with federation capabilities extend this, with the ability to store links to data in other repositories using the APIs of those foreign repositories.

Acclimatize business personnel to their role in creating information assets and the importance of metadata as a precursor to introducing these practices.

Initiate a data administration task to capture various business ontologies of integrated information resources, with the understanding that ontology evolves continuously.

Business Impact: Enterprise metadata taxonomy and ontology management will bring tighter integration between business process changes and IT system changes. Using federation capabilities will also enable impact assessment and traceability. Business units and users will be able to relay their concerns more effectively regarding use of information assets throughout the organization. It will also enable better assessment by business analysts of the risks and benefits that accrue in the business regarding maintenance and security of information assets.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: Mondeca; Pragmatic Solutions

Information-Centric Infrastructures

Analysis By: David Newman

Definition: The information-centric infrastructure is the technology building block within an organization's enterprise information management program. It represents the principles, models and requirements necessary to flexibly share and exchange information assets, as expressed within the enterprise information architecture. This infrastructure's architectural style is modular, distributed and loosely coupled. It features the alignment of metadata, semantics, standards and content formats to achieve the consistent and seamless capture, persistence, transformation and delivery of content across the enterprise, independently of how it is stored in source systems or packages. It thus breaks the tight binding of content to applications — a critical first step in increasingly service-oriented, context-aware and cloud-based computing environments.

Position and Adoption Speed Justification: Organizations are moving away from architecture styles of tightly coupled systems and self-contained applications toward modular software components and reusable services. During this transition, old issues are suddenly exposed (such as poor data quality, lack of metadata transparency, conflicting master data and inconsistent business intelligence or analytics). Although many of the technologies required to address such
barriers are scattered today throughout the organization, they will eventually be consolidated into information-centric infrastructures that deliver "information as a service" to support a variety of enterprise needs.

Speeding the adoption of information-centric infrastructures are emerging computing styles, such as service-oriented architecture (SOA), software as a service (SaaS), cloud computing and context delivery architectures (CoDAs). Such styles require the delivery of information as a service, which will be achieved by bringing together related information management technologies and capabilities in data cleansing, transformation and integration, metadata management, business intelligence, content consolidation, semantic reconciliation, and master data management. Speed of adoption will also be driven by the need for a standard platform to normalize the chaos across different content types and to deliver content convergence (combining structured and unstructured content) to meet a variety of needs, free from application-specific proprietary protocols.

Business process improvement initiatives will also speed the adoption of information-centric infrastructures to support efficiency and agility drives, requiring the assembly of reusable data services, as well as content-driven mashups, next-generation business intelligence and executive dashboarding business process solutions.

Adoption of information-centric infrastructures is nascent. Business needs (particularly to meet agility and flexibility requirements) and technology trends (particularly for loose coupling, late binding and alternative delivery models) will continue to spur interest and adoption. Yet, significant barriers remain before information-centric infrastructures gain wider acceptance. These include cultural and funding models — particularly among IT organizations — in which the application "container" remains the solution to nearly every business process need.

**User Advice:** First, identify how project teams use different systems, formats and standards to exchange similar types of information (for example, master data). Look for common ways to locate, normalize and extract meaning so that content can be flexibly shared, independent of the application. Use existing system analysis and design patterns as starting points to develop common information models, which will then be used by different processing components and system entities. Exploit emerging standards (such as XML) or data and metadata interchanges to create a common component library of metadata objects, thereby enabling reuse.

**Business Impact:** The information infrastructure brings balance to many application-driven environments, because it "normalizes" the chaos caused by having different and diverse standards, formats and protocols.

It supports a variety of business process needs by: 1) identifying, abstracting and rationalizing commonalities across content; 2) applying semantics for information exchange and interoperability; and 3) implementing metadata management for discovery, reuse and repurpose.

**Benefit Rating:** Transformational

**Market Penetration:** Less than 1% of target audience

**Maturity:** Emerging

**Recommended Reading:** "Key Issues for an Information Infrastructure Project, 2008"
"Key Issues for Delivering an Enterprise Metadata Management Project, 2008"
"Key Issues for Information-Centric Infrastructures, 2007"
"Key Issues for Enterprise Information Management Initiatives, 2008"
Video Search

Analysis By: Whit Andrews

Definition: The ability to search within a collection of videos, owned or uncontrolled, is an inevitable flare in the content market as expectations sparked by YouTube and LiveLeak enter the enterprise. Video search will incorporate elements of social networking, social tagging, metadata extraction and application, video and audio transcription, and conventional enterprise search to make it possible to find videos using ordinary enterprise search.

Position and Adoption Speed Justification: Expectations driven by YouTube are intriguing consumers as they see the possibility for improved searchability in rich media. Ultimately, enterprise search will subsume video search as simply another format, just as it has with audio and graphical media.

User Advice: Only enterprises with the greatest ambition for video in their operations should invest in video-specific search capabilities. Others will more likely turn to application service providers (expect a variety of formal and informal enterprise editions of video hosting from Web conferencing vendors and others) or wait for video search as an element of enterprise search.

Business Impact: Making video easier to locate will boost the use of non-textual elements in training and communications in the enterprise.

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: Autonomy; ClearStory Systems; Coveo; EveryZing; Sonic Foundry

Federated Search

Analysis By: Whit Andrews

Definition: Passing a user's search to one or more other search engines, retrieving the results and serving them back to the user is federated search. Sometimes federated search interleaves the results so that they are compared against each other; sometimes they are just presented next to each other as alternatives.

Position and Adoption Speed Justification: Federated search is growing in visibility as enterprises realize they have multiple search technologies in place and would benefit from a means of combining results from more than one search engine. The challenge, however, is significant and as enterprises realize that, they recoil from the difficulty and the fact that embarking on such an effort today represents a commitment to futurist and uncertain thinking. Problems include the difficulty of juxtaposing radically different formats of content (video vs. text), sources of content (an internal e-mail vs. an external news story), and the inclusion of data with extreme time sensitivity (such as CRM records). Federated search demands a means of effective disambiguation that is only now being developed by many vendors.
User Advice: Federated search that exploits conversational results strategies, such as guided navigation, is an achievable goal today. Interleaving results is profoundly challenging and difficult to accomplish, because downstream search engines, to which the search engines send their queries, are likely to have varying relevancy models that do not make it easy to assemble results.

Business Impact: Improved access to information for workers and customers.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Autonomy; EMC; Endeca; Expert System; Microsoft; MuseGlobal; Open Text; Progress Software; Vivisimo

Recommended Reading: "Enterprise Search: Back to the Giants"

SaaS — Enterprise Content Management

Analysis By: Karen Shegda; Mark Gilbert

Definition: Software as a service (SaaS) for enterprise content management (ECM) is an emerging business model that reduces the upfront costs, complexity and resource requirements normally associated with purchasing and implementing ECM solutions. SaaS is an evolution of the application service provider (ASP) or hosted model. SaaS is about providers building out a specific application, which they run in a multi-tenancy environment. They then bring it to market with a configurable front end and a subscription-based business model.

Position and Adoption Speed Justification: In the broader content management (CM) market, hosted content and document management are gaining popularity, especially among midsize organizations and departments in larger organizations. Often, these organizations are trying to reduce their IT overhead and focus on core competencies. Although demand for SaaS — and for hosted solutions generally — is increasing, these offerings will remain a relatively small part of the overall ECM market through 2010.

User Advice: Users considering SaaS or hosted solutions should weigh their requirements — and the cost of those requirements — carefully. The pricing models for hosted solutions vary: some charge low monthly subscription fees, others price by the volume of documents hosted. Enterprises should consider SaaS approaches if they are midmarket companies, as many offerings are tailored to the needs of small and midsize businesses. For larger organizations, SaaS may make sense as a tactical or short-term option while they evaluate vendors or prepare for a procurement.

Business Impact: CM SaaS is more attractive in the early stages of a market, for more tactical choices, or where more standardized functionality is acceptable because there is little need for customization.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Microsoft; salesforce.com; SpringCM; Xerox

Recommended Reading: "When to Evaluate SaaS as a Viable ECM Model for Your Company"
At the Peak

DITA-Aware Editors

Analysis By: Rita E. Knox

Definition: Darwin Information Typing Architecture (DITA) is an IBM-developed, OASIS-approved XML Document Type Definition (DTD) to define modular content components ("topics"). DITA-aware editors provide structured DITA-defined environments for content editing. They limit content categories to DITA topic definitions. Existing specializations (topic types) can be used for new topic creation (new specializations). These editors also create "topic maps" for topic aggregations to communicate instances of specific content, such as documents or Web pages.

Position and Adoption Speed Justification: Enterprises that have traditionally used XML-aware publishing applications are already replacing those with DITA processes. The generic DITA topic models are easier to understand and implement, and don't require industry-specific customization, but may include them. Adoption will expand vertically (such as for insurance letters, financial investment offerings and corporate filings) and horizontally (such as HR benefits descriptions) to support all content-centric communications. It will simplify those communications and ensure corporate and regulatory compliance.

User Advice: Validate your enterprise needs using the DITA Quotient (http://dita.xml.org/wiki/what-is-your-dita-quotient-dq). If the assessment validates the need, enterprises should consider investing in DITA as a replacement for current publishing approaches. All content-related applications — content creation, management, formatting and output — should be examined. High value/risk content — such as medical equipment technical manuals, teacher's instruction modules and plans, and aircraft maintenance manuals — should be DITA-based to ensure consistency, compliance and completeness.

Business Impact: DITA can accelerate content creation, increase content reusability, reduce creation time and production costs, and support highly distributed process responsibilities. All XML-aware applications have been or are in the process of being extended to include DITA. It also opens opportunities for many new applications, such as automated content validation.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Adobe Systems; InvisionResearch; Justsystems/XMetaL; Oxygen; PTC (Arbortext); Stylus Studio

Recommended Reading: "DITA-Aware Content Processing: Simpler, Cheaper, Faster"

Mashup Applications

Analysis By: Anthony Bradley; David Gootzit

Definition: A "mashup" is a lightweight, tactical presentation layer integration of multisourced applications or content into a single, browser-compatible offering. It is a lightweight variant of the older notion of a composite application ("composite app") and the heavier service-oriented architecture orchestration approach to composite applications. In the usual use of the term, composite apps are built on internal facing enterprise platforms and are not necessarily Web-based. A mashup is a Web-based composite application that leverages systems (data, business logic and presentation) to create a new capability.
Enterprises are beginning to transform mashups from Web entertainment to enterprise relevance. Even in the enterprise, mashups leverage data and services from Web sites, such as Google Maps, craigslist, eBay, Amazon.com and others. Because mashups leverage content and logic from other Web sites and Web applications, they're lightweight in implementation and are built with a minimal amount of code (which can be client-side JavaScript or server-side scripting languages, such as PHP or Python). These are not fixed requirements, but they reflect the original implementation of the mashup concept in Web 2.0 startup companies, which typically do not use enterprise-oriented platforms, such as Java or .NET.

Mashups exploit lightweight mechanisms, such as representational state transfer-based application programming interfaces (APIs), to Web services, as well as Ajax "snippets" and "widgets" (see "Adopting Ajax Means Choosing From Four Levels of Ajax Technology"). Mashups aren't intended to be strategic, systematically built, industrial-strength enterprise applications; rather, they're created quickly or opportunistically to meet a focused tactical need.

Mashups generally are personalized to fulfill personal productivity needs rather than the requirements of a long-standing corporate role. The cultural context of mashups involves the confluence of many innovations: Web APIs, lightweight client-side scripting, delivery of content via Really Simple Syndication (RSS), wikis, Ajax, social networking and the explosion of Web-based communities. For a long time, the closest thing to mashup creation tools for "civilians" (for example, users who do not write code) was an RSS feed reader or podcasting client, which enabled them to "mash" content from more than one site. This situation has improved, with more-powerful tools such as Yahoo Pipes, Microsoft Popfly and Google Mashup Editor.

**Position and Adoption Speed Justification:** Mashups are driven by the Web culture. There are thousands of mashups on the Web, often built by nonprofessional programmers. For example, HousingMaps.com combines data from Google Maps with apartment rental information from craigslist to create a new application that shows the location of available apartments in a given city — all accomplished without the direct participation of Google or craigslist staffs.

In addition, Web sites such as Facebook, Pageflakes, Netvibes and iGoogle have enabled thousands of non-technical users to build mashups using prebuilt mashable assets (for example, gadgets and widgets). Web APIs and mashups are available in an online directory at ProgrammableWeb (http://www.programmableweb.com). In keeping with the consumerization trend, the popularity of mashups on the Web is driving enterprise interest and experimentation in how mashups deliver value inside the firewall. Although this Web mashup movement is significant, it's importance should not be overstated to enterprises, because the vast majority of Web mashups is more for entertainment value rather than business value. Greater business value and the realization of mashup hype is growing as more enterprise implementations mature.

**User Advice:** Mashups quickly integrate content or functions from multiple sources and present easily understandable items of interest. The trade-off is faster time to market and reduced development costs over application robustness and longevity. An enterprise mashup environment is only as valuable as the quantity and quality of the mashable assets available for users and professional or nonprofessional developers to assemble into mashup applications. The role of IT developers is to provide this high-quality repository of mashable assets.

Mashups are not without risks. Because mashups combine data and logic from multiple sources, they're vulnerable to failures in any one of these sources. There are also risks and concerns regarding the use of intellectual property and the longevity of provider relationships. Mashups will continue to experience steady growth through 2009. In revenue, the strongest growth will come from the "enterprise mashup" variant as vendors of enterprise software, platforms and middleware adopt this Web 2.0 concept and adapt it to enterprise needs.
**Business Impact:** Mashups can deliver significant application agility by bridging the gap between how quickly the business situation changes and how slowly enterprise applications evolve. Mashups provide power users and knowledge workers with the means to quickly build software solutions without the direct involvement of developers. Mashups also extend the reach of the applications group and provide greater developer productivity through a potentially highly leveraged repository of mashable assets.

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Emerging

**Sample Vendors:** Connoteate; Denodo Technologies; Google; IBM; JackBe; Kapow Technologies; Lixto Software; Magic Software Enterprises; Microsoft; Netvibes; Nexaweb Technologies; Oracle; Pageflakes; RSSBus; Serena Software; WorkLight; Yahoo

**Recommended Reading:** "Reference Architecture for Enterprise 'Mashups''

"Key Issues for Enterprise 'Mashup' Practices, Technologies and Products, 2008"

**E-Discovery Software**

**Analysis By:** Debra Logan; John Bace

**Definition:** Electronic discovery (or e-discovery) software facilitates the collection, processing and review of large amounts of electronically-stored information (ESI) within an enterprise to meet the mandates imposed by common law requirements for discovery. These demands may be due to civil or criminal litigation, regulatory oversight, or administrative proceedings. An independent group of consultants, legal scholars and vendors have created and put into the public domain an "E-Discovery Reference Model" ([EDRM] www.edrm.net) that maps traditional common-law discovery into a six-step, nine-process framework for technology. There are hundreds of vendors that have products that fit within the EDRM framework that do everything from policy management and search and analysis, to production and presentation. When Gartner focuses on the e-discovery software market, we are concentrating on the technology providers who work at the nexus of where IT and the legal staff meet: the preservation and collection of relevant ESI from the technologist point of view, and the search, review, and analysis of its content for the legal professional.

**Position and Adoption Speed Justification:** End users are increasingly doing legal discovery in-house. End-user companies are most interested in the information management, identification, preservation and collection phases of the EDRM. Because of the number of vendors in the market and the remaining uncertainty around what the U.S. courts are expecting companies to do, firms remain confused about what the best practices are should they become involved in litigation.

**User Advice:** The move to acquire e-discovery software is driven by efforts to reduce risk and drive cost efficiencies.

Evaluate products that can aid the identification, preservation and collection of potential evidence. The second most important area around e-discovery software is the ability of these tools to create, communicate, enforce and document compliance with litigation hold orders.

Finally, these tools can provide the in-house legal staff with the technology to query custodian repositories in early-case assessment to determine things like cause, motive and action.
Business Impact: Major enterprises undergo dozens, or even hundreds, of investigations per year, resulting in high costs for specialized litigation support companies and outside law firms. Software that supports the ability to conduct and manage discovery activities in-house not only saves money, but also enables enterprises to have higher levels of control over investigations.

The market is just emerging, with point products that handle part of the process being the norm in the market space. There are hundreds of vendors that claim e-discovery functionality in some form or another. The tools have emerged from several adjacent and related areas, like forensic investigations, records and document management, e-mail archiving and search and information access. There is also a large, stand-alone review and analytics market, which focuses on providing review and analysis tools for legal personnel. Although the market has already begun to consolidate around a set of tools to handle information management, identification, collection, preservation and processing, we do not consider there to be complete end-to-end e-discovery suites on the market. Aspects of the problem remain difficult, particularly those relating to information access and finding relevant data in the masses of content that most enterprises have.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Anacomp; Attenex; Autonomy; AXS-One; Catalyst; CA; Clearwell; CommVault; EMC; FTI Ringtail; Guidance Software; IBM; Kazeon; Mimosa Systems; Oracle; PSS Systems; Recommind; Seagate Technology; StoredIQ; Symantec; ZyLAB

Recommended Reading: "MarketScope for E-Discovery and Litigation Support Vendors"

Social Search

Analysis By: Whit Andrews

Definition: Social search uses elements of user behavior, implicit and explicit, to improve results for searches. It also enables users to more-effectively disambiguate results from their queries. Examples include such steps as saving searches to shared folders, tagging of searches or documents to express what they are about for other users, and the use of implicit indicators of value such as saving documents as shared bookmarks or printing documents for later use.

Position and Adoption Speed Justification: Social search is an element of Internet search behaviors and provides significant value to Web users. Consumerist expectations will drive it quickly into the enterprise where it will augment, but not transform, search capabilities. The ability to use social search at Web sites such as Amazon (where tagging and search recommendations are used) will drive users to demand it.

User Advice: Include elements of social search in your projects, but do not expect it to dramatically improve results. Disambiguation will benefit from social search use as users scrutinize colleagues' and fellow searchers' results to establish meanings and relationships.

Business Impact: Enterprise workers and customers will get better results faster through the use of social search.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent
User-Generated Content (UGC)

Definition: User-generated content (UGC) is content created by individuals who operate quasi-independent of the core authoring, editing and approval processes of a particular content management instance. The content distinguishes itself from that contributed by the traditional employed or contracted author role primarily in that the source does not arise from individuals who are under direct control of the content manager or who are formal registered users of the internal workings of the content management instance. The usual governance is therefore not always applied to the published content. Instead, these individuals can be visitors to a Web site or to a content management system internal to a given organization. Their content is contributed using a variety of techniques from traditional content upload to the application of any of the newer Web 2.0 technologies. Examples include blogs, wikis (such as Wikipedia), rich media sites (such as Flickr and YouTube), podcasting, e-commerce sites (such as eBay), community (such as Facebook and MySpace) and forum postings. This is a significant trend in Web content management and internal document management as "users" of a Web site or employees of an enterprise want to have their say.

Position and Adoption Speed Justification: A number of different factors will contribute to UGC quickly becoming a mature technology. Pervasive Web use, content tailoring to any topic or interest, growing sophistication of all user groups, the "legitimacy" of Web content as a trusted information source, and corporate decisions to encourage employees' Web presence rather than discourage or contain them, have fueled the pervasive use of this communication source. The general availability of inexpensive or free multimedia creation tools have made it practical for individual users to become prodigious publishers. The memorable 2006 Time cover "You" as "Person of the Year" highlights the changing role of visitors to content contributors. The positioning of UGC on the Hype Cycle therefore reflect two opposing forces. These comprise the high consumer and employee interest in adopting UGC, and the struggle currently evidenced in enterprises to capture, monetize and optimally control this content.

User Advice: Whether sanctioned or unofficially generated, UGC is difficult to suppress and can benefit companies in ways that official communications cannot. Enterprise-sanctioned employee blogs or supported user groups cannot only benefit public relations, but can also be an extremely useful source of feedback and perception. Enterprises should have a policy that guides employee-generated UGC; the policy need not be too restrictive, but should provide guidelines for what is and isn't acceptable. UGC can also be completely beyond enterprise control; monitor the Web for references that point to you. Enterprises should strongly consider facilitating UGC on their Web sites to develop better relationships with their segmented target audiences. This feedback should be captured and analyzed to help guide contracted authors to produce more effective and contextualized content for those target audiences.

Business Impact: User content ratings and comments (such as on issues, movies and consumer products) influence behavior as well as create topic metadata. This information not only determines Web content location (such as tops of lists and front page), but also determines how a discussed topic will fare. UGC is published almost instantaneously, without process or editing oversight, and often beyond enterprise control. This can have positive and negative effects alike.
**Benefit Rating:** Moderate

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

**Sample Vendors:** Alfresco; Digg; Facebook; Flickr; IBM; Microsoft; MySpace; salesforce.com; YouTube

**Open-Source Content Management**

**Analysis By:** Mick MacComascaigh

**Definition:** The "open-source software model" describes a set of characteristics and properties for developing, delivering and supporting software. Open source is licensed software in which the source code is made available to users so that they are enabled with the freedom to modify it for their own purposes and, with very limited restrictions, redistribute original and derived works as they see fit. "Open-source content management" pertains to the application of this model, primarily to document management, workflow, records management, imaging, Web content management and document-centric collaboration.

**Position and Adoption Speed Justification:** An expanding group of vendors is focusing on open-source-based enterprise content management (ECM) technologies. Open source, while attractive, still remains an immature area for ECM. Such tools are, and will remain, primarily of interest to enterprises with the technical skills and resources to adapt the products to fit their requirements, and to counter the risk associated with the perceived lack of a trusted entity behind the product. The number of possibilities on the market are also diffusing what would otherwise be a focused, concentrated development of a particular platform. To this end, open-source ECM is not enjoying the rate of adoption of its "LAMP" counterparts (Linux, Apache, MySQL and PHP). Whereas interest in open-source ECM is increasing, only low to modest levels of adoption has been observed in the market. A market driver for this interest is certainly the cost savings for the licenses, or having to pay only maintenance costs. However, uncertainty over the necessary quantity and availability of the required services to complete a program of work has given rise to the limited adoption.

**User Advice:** Technically oriented users who are interested in taking on a project will be attracted. The majority of users will continue to turn to traditional offerings. The adoption of open-source ECM will depend in part on the culture of your organization with regard to change, because there are likely to be shorter periods between required updates. This replacement of the more-traditional, longer software life cycles will therefore require a very different strategy when it comes to change management. Both intra- and inter-organizational collaboration is recommended and required for best results to ensure optimal reuse and sharing of components already developed. It is through this mechanism that the additional costs for development can be partially offset. Organizations should, however, not underestimate the cost of deploying and supporting open-source software. When calculating the total cost of ownership as part of their selection process, they should also consider their intended deployment road map and compare the costs accordingly.

**Business Impact:** Open-source content management has ignited modest interest among buyers overall. However, the perceived uncertainties surrounding the model in terms of support and code quality continues to slow adoption. Most open-source content management tools are primarily of interest to enterprises with the technical skills and resources to adapt the products to fit their requirements.

**Benefit Rating:** Low
Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Alfresco; Bricolage; Magnolia; OpenCms

Recommended Reading: "Decision Framework: Use These Criteria to Determine Whether Open Source Should Be a Part of Your ECM Strategy"

Basic Content Services

Analysis By: Karen Shegda; Mark Gilbert

Definition: Basic content services (BCS) are easy-to-use document management applications that provide document library (check-in and check-out, version control and document-level security) services with basic Web publishing, document collaboration and document routing as core functions. Their key characteristics are low license cost per seat and ease of deployment and integration with desktop platforms. BCS tools typically address workgroup or team requirements for collaborative document management.

Position and Adoption Speed Justification: Basic content services represent one end of the content management spectrum. Organizations are looking to BCS to provide the basic capability that most of their users need to replace the use of shared file drives, manage individual and team documents for increased productivity and collaboration, and reduce the number of duplicate documents and versions being stored and routed via e-mail. Microsoft and IBM have begun delivering these basic content-for-the-masses capabilities as part of their infrastructure, tying them to their portals and desktop applications, thus accelerating adoption.

User Advice: Enterprises that have skilled workers should deploy basic content services to lighten the load on their e-mail systems and improve collaboration among those who create routine documents. Some companies think that a search tool may solve their problems, but basic content services provide more functionality in the creation and management stages of the document life cycle. Consider content management as a spectrum of capabilities, with basic content services at one end for basic and horizontal capabilities, and full enterprise content management at the other for process-centric applications.

Business Impact: Basic content services will become a common element of the future workplace, enabling companies to improve productivity by reducing the time employees spend filing and looking for documents. Basic content services serve the content management needs of knowledge workers who don't deal with mission-critical documents.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: IBM; Microsoft; Xerox; Xythos

Recommended Reading: "Magic Quadrant for Enterprise Content Management, 2007"
"SharePoint for Content Management Revisited"

Content-Enabled Vertical Applications

Analysis By: Toby Bell; Mark Gilbert
**Definition:** Content-enabled vertical applications (CEVAs) typically help to automate complex and structured processes that previously required workers to sort manually through paper documents and other forms of content. CEVAs help reduce the costs of exception handling and optimize the rest of the work by applying better process control and analytics. CEVAs are a mix of content, process and domain expertise, resulting in repeatable/reusable models and solutions.

**Position and Adoption Speed Justification:** Companies are working further to link more of their content to industry-specific, human-centric processes such as insurance claims-handling or by supporting research on new drug development through packaged process templates and repeatable solutions. This approach usually means building or modifying CEVAs on top of enterprise content management (ECM) environments. Evaluating CEVAs can be difficult because they are available from several different classes of vendor, including ECM, business process management (BPM), system integration and stand-alone CEVA vendors. In addition, CEVAs are technically complex. They work because they have deep roots in a company's IT infrastructure and incorporate deep knowledge about the industry in which they are deployed. The market is beginning to form around CEVAs and demand for solutions is driven by line of business (LOB) buyers.

**User Advice:** Companies should pay particular attention to what ECM vendors are doing with CEVAs. ECM vendors have made the most progress in assembling the complete package of capabilities needed for CEVAs, but the BPM market has also begun to provide packaged templates for some content-enabled processes. Whether through partnerships, acquisitions or internal development, ECM vendors have begun incorporating BPM functions. They offer integration services and have started specializing in certain industries. CEVAs are present in a wide range of industries, but different ECM vendors have expertise in different industries. Organizations should ensure they know the industries in which their ECM vendor is competent to offer CEVAs. Not every provider has the competence to deliver effective CEVAs, so companies must scrutinize CEVA references before committing to an implementation. Watch for "solutions as a service" to become a more common delivery model in the next few years as contracts, case management, claims and patient records hybrids emerge.

**Business Impact:** CEVAs enable companies to use established applications for new purposes, and to respond better to workers' and customers' information needs. They can also replace custom code with reusable and flexible frameworks on top of existing infrastructure investments. But implementing CEVAs is complicated. It involves new types of information-centric infrastructures, content-specific technologies, integration skills, domain expertise, best practices and methodologies. Key considerations for driving CEVAs include:

- The flexibility and ease of use of the BPM technology, either embedded in ECM or available by itself.
- The number of vendors delivering bundled frameworks and templates that range in completeness from 25% to 75%.
- The influence of the business buyer looking for near-term cost recovery, proven return on investment, and generally making less use of system integration, IT services and support for LOB applications. The perception is that the ratio of services to software costs is too high.
- A better relationship between structured application data and more granular XML-based content objects.
- BPM and ECM vendors moving quickly to differentiate themselves from basic content service providers, stack players and competitors by gaining specific industry expertise.
The pace at which organizations move toward adopting a common information-centric infrastructure, which enables the business to flexibly share and exchange all types of content as part of a program to achieve enterprise information management.

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

**Sample Vendors:** Adobe; EMC Documentum; Global 360; Hyland Software; IBM; Interwoven; Metastorm; Open Text; Pegasystems; Vignette

**Recommended Reading:** "Use CEVAs to Generate Value From Your Content"

"Evaluate Content-Enabled Vertical Applications on Six Levels"

"Findings From the 'Content Management' Research Meeting: The Race to Basic Content Services"

### Sliding Into the Trough

**JSR 170/283**

**Analysis By:** Mark Gilbert

**Definition:** Java Specification Request (JSR) 170 and 283 are content repositories for Java application programming interfaces (APIs) that make content management (CM) repositories more open to integration. This should enable CM systems to access and manage content across multiple repositories using a standard connector specification.

**Position and Adoption Speed Justification:** JSR 170 was finalized at the end of 2004, approved in May 2005 and is now being engineered into commercial software products. JSR 283 is the evolution of 170, and is designed to simplify the repository architecture and reduce integration costs.

**User Advice:** Ask your CM vendors what their specific plans are for supporting JSR 170/283. Some vendors are hesitant to open up their repositories, which they will do by supporting JSR 170/283. A lack of support for these standards may force clients to choose which vendors they continue to do business with.

**Business Impact:** The potential benefits of JSR 170/283 are real, and enterprises with multiple CM repositories should embrace them.

**Benefit Rating:** Moderate

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

**Sample Vendors:** Day; EMC; IBM; Interwoven; Open Text; Oracle; Vignette

**Recommended Reading:** "Learn the Difference Between Portals and Content Integration"

### Content Integration

**Analysis By:** Toby Bell
Definition: Content integration is access to unstructured content (such as images, documents and video) dispersed throughout the enterprise (for example, in applications and databases). Content from multiple sources is consolidated as a single view (for example, a repository or Web site). Content integration tools may sit above repositories (for example, as data-integration middleware), or above workflow and business process management systems to provide a unified interface with work items. Some packaged connectors provide bidirectional integration, while others provide integrated information through federated search.

Position and Adoption Speed Justification: Most enterprises have multiple repositories. Customized interfaces and commercial connectors/adaptors dominate. Many enterprise content management (ECM) suites use connectors from IBM, Day or Vital Path. But the days of custom connectors may be limited, partly by industry-sponsored interoperability (interoperable ECM [IECM]), Web services or other integration options, such as the Java Specification Request (JSR) 170/283 standard. Many enterprises are also looking at leveraging portals, XML databases and federated search as options for consolidating frequently used content at different levels of abstraction.

User Advice: Enterprises should look beyond JSR 170/283 and Web Distributed Authoring and Versioning to integration architectures from vendors (including IBM/Venetica, Oracle/Context Media, Day, Vignette and Vital Path) and third-party offerings (for example, from Vamosa). Most system integration partners also have toolkits to integrate products they support with multiple repositories and business applications. So, if your vendor or services provider can't make its technology "play well with others," it may be cause for concern.

Business Impact: Content integration provides greater interoperability between a company's content and its content-centric processes, and supports both governance and cost reduction. Classic use cases include legal discovery and federating content across newly merged or acquired companies. But the biggest potential impact of content integration is relating it to the semantic layer over structured data in application databases, as well as linking a horizontal business information supply chain to vertical processes and end-user context. Content integration is key to controlling costs, improving availability, providing a single view of the customer, and relating content and data objects via metadata management.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Day; IBM; Oracle; Vamosa; Vignette; Vital Path

Recommended Reading: "An Information-Centric Infrastructure: The 'XML Pump'"

Content Analytics

Analysis By: Rita E. Knox

Definition: Content analytics "mines" and processes information to derive answers to specific questions. It may consist of a single function or a series of functions that pass results in a sequence from one operation to the next. It does not consist solely of new information access components, but it does depend on newer approaches to how processing steps are serialized. It can process non-textual content (such as photographs and speech), as well as textual content (for example, photograph captions, blogs and news sites). Its most important feature is the serialization (sequencing) of analyses to create layered interpretations of "raw" content or search results. Examples of such analytics results are: identifying high-priority clients, interpreting competitors' activities and understanding consumer responses to a new product.
**Position and Adoption Speed Justification:** Content analytics is not new, but the ability to serialize components to enable complex analyses is a significant advance. Since mid-2007, understanding how to apply these analyses to enterprise work processes has increased dramatically. Users are actively seeking ways to automate processes that have surpassed human capabilities and to mine sources for information that can improve client satisfaction, reduce product time to market and increase profits. The need to access and analyze "secondary" sources (surveys, customer calls) has increased dramatically. Enterprises view it as underexploited and underappreciated information.

Other capabilities are pointing to integrations with other applications. Content analytics can provide precursor analysis to search applications (for example, by creating metadata — a "country" or "company" tag — appropriate to the analysis that has been performed) or post-process search results (for example, metadata created by the content analytics function can be added to the search index). Content analytics augments business intelligence (BI) analyses by adding structure to unstructured content.

**User Advice:** Enterprises should employ content analytics to replace time-consuming and complex human analyses. They should identify the analytics most able to simplify and demystify complex business processes and analyses. Users should identify vendors with specific products that meet their requirements and review case studies to understand how other customers have exploited these technologies.

**Business Impact:** Content analytics is used to support a broad range of functions. It can: identify high-priority clients, product problems, customer sentiment and service problems; analyze competitors’ activities and consumers’ responses to a new product; support security and law enforcement operations by analyzing photographs, and detect fraud by analyzing complex behavioral patterns. Increasingly, it replaces difficult and time-consuming human analyses with automation, often making previously impossible tasks tractable. And complex results are represented as visualizations easily understood by people.

**Benefit Rating:** Moderate

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

**Sample Vendors:** Attensity; Clarabridge; ClearForest; Dow Jones Factiva; Endeca; IBM; IxReveal; PowerReviews; Stratify; Trampoline Systems

**Recommended Reading:** "Use Content Analytics to Extract Meaning"
"Key Issues for Content Analytics, 2007"
"Integrating BI and Content Analytics Gives Better Results Than Using Them Separately"

**ISO-Standard XML Formats For Office Productivity**

**Analysis By:** Michael Silver; Rita E. Knox

**Definition:** ISO-standard XML formats for office productivity include ISO/IEC 26300, known as OpenDocument (ODF), and ISO/IEC DIS 29500, known as Office Open XML. Both formats are designed for exchanging data between office productivity products such as OpenOffice.org and its distributions (which introduced ODF) and Microsoft Office (which introduced OOXML when it shipped Office 2007) with other applications. Both are XML-defined schemas to describe content formats for text, spreadsheet, presentation, graphics (such as scalable vector graphics) and other content types, to support exchange among different applications.
Position and Adoption Speed Justification: OpenOffice.org is not as widely used as Microsoft Office applications or MS Office's document formats (.doc, .ppt and .xls), which are widely used for content creation. However, OpenOffice.org has been increasing in stature, especially in governments and by organizations using older versions of MS Office that are hoping to save some money by moving away from Microsoft Office. OpenDocument is based on the OpenOffice.org productivity suite, and continues to be developed by the OASIS working group. In May 2008, Microsoft announced that it would officially support ODF in future versions of its products.

Although Microsoft Office is in broad use, OOXML is still a relatively new format, and most organizations still use Microsoft's legacy binary document formats. Office 2007, when used with OOXML, enables new features. Documents saved in OOXML formats can be significantly smaller than when saved in the legacy binary formats, thus saving bandwidth and storage.

Interest from organizations looking to use the formats for data exchange with other applications has been relatively low; they're not seeing more general uses for an XML format, which might make it important to use one format over the other. Both sides have been lobbying heavily to make their format the standard for the future although, by 2009, the Microsoft and OpenOffice.org-based products will support both formats.

User Advice: Users should recognize that they will eventually be saving their office document data in an XML-based format. Realize that XML-based formats are more important for exchanging data between office productivity applications and other applications than for exchanging data between office productivity products from different vendors. Although the standard formats will be well-supported, visual fidelity will still be problematic when opening an OOXML document in OO.o or an ODF document in Microsoft Office. This is because XML only specifies content, not output representations. The file data representations will be standardized, but a compatible and consistently-defined rendering model (a mapping from the XML markup to presentation) is not part of either standard document format.

Select the format needed based on the office productivity products being used or the products that are exchanging data. In the short term, the Microsoft legacy binary formats will be the most widely supported formats.

Business Impact: These XML-based standards enable exchange among complying office productivity suites, but more importantly, can be used for exchanging data between applications written to do so. XML-defined document formats are becoming increasingly important as content applications move to Web provisioning and multimedia/multimodal representation increases in importance.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Apple; Corel; Google; IBM (Workplace); Microsoft; Novell; Sun

Recommended Reading: "Use XML to Structure and Manage Content"

XML-Enabled Database and Content Management Systems

Analysis By: Rita E. Knox; Donald Feinberg

Definition: These are any form of database management system (DBMS) or content management (CM) system to which native Extensible Markup Language (XML) has been added
as a data type. In the early stages of XML-enabled DBMS development, the XML object was stored in a binary large object or shredded by removing element tags and storing the content alone in columns with the element tag as the column name. Although these options still exist (and are being used), the key difference today is that the XML object is stored as a binary tree. XML-enabled DBMSs cover all of these techniques. In 2008, structured and XML-tagged content are brought together to exist side by side in traditional DBMS applications (IBM, Oracle, Sybase).

XML-aware content management uses embedded XML markup (tags) to identify and manage content components (as small as a word or phrase). Most CM vendors have supported XML for some time, but the granularity has been much coarser (for example, at the paragraph level). The level of granularity with these XML-enabled systems matches whatever levels are identified by the markup.

**Position and Adoption Speed Justification:** Most DBMS vendors have implemented XML as an extension to their DBMS. It is another example of a disruptive technology being absorbed into the more well established DBMSs and greatly enhancing their capabilities. Now that the application independent software vendors (ISVs) (such as Document Systems and Nextance) have begun to use these native implementations with the more standard DBMS engines, there is an increased growth in the use of the XML data types and XML-enabled DBMSs. Furthermore, due to compliance issues (such as Sarbanes-Oxley Act [SOX] and Basel II), and the need to store and access XML in data warehouses, the use of XML-enabled DBMSs is accelerating.

The recognized value of granular content (such as text) components is growing. Darwin Information Typing Architecture (DITA) makes the value of granularity and content reusability obvious. With DITA's growing publicity and enterprise adoption the need to extend XML content management to highly-granular components is necessary. Examples in the manufacturing and regulated industries are growing. These repositories will continue to manage larger aggregations, such as sections and complete documents.

**User Advice:** Begin using the XML capabilities of the DBMS engines to include XML data in databases, facilitating management and security in a single DBMS. It is now reasonable to store XML documents and granular content components in a data warehouse for the purposes of text mining and analytics. As document management and CM application vendors begin to support the major XML-enabled DBMSs, formulate a strategy for migration to those DBMSs to leverage the benefits of fewer DBMSs and CM systems, including lower total cost of ownership, simplified governance and easier management (as opposed to multiple DBMSs and file systems), improved security, high availability, and backup and recovery.

Enterprises that plan to create and reuse content components should adopt XML-aware content management if they haven’t already. This is especially important for required documentation in different industries. Such adoption will be simpler if a CM vendor already has this capability. Enterprises should ask vendors for their plans; if it's not on their road map, consider alternative long-term solutions. This is especially important for required documentation in different industries such as healthcare, aerospace, automotive and financial services.

**Business Impact:** Enterprises (especially regulated industries) that manage reusable content components have reported 90% cost savings for updated documentation, and publishing times that shrink from months to weeks. The quality and consistency of content are also reported to improve.

The ability to store XML data in a standard DBMS enables XML to have the same data persistence, reliability, security and stability as data stored in a central data storage facility. With the increasing volumes of data from SOX and other compliance initiatives (requiring documents to be stored), as well as most documents being available in XML format, XML-enabled DBMS usage should grow dramatically. The ability to search XML efficiently and to use XML for text
analytics is allowing businesses to use this information in data warehousing and other applications. It also is less expensive to store all content in a common database than to have one for structured, another for less-structured data, and one for less-structured data.

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

**Sample Vendors:** Astoria Software; EMC Documentum; IBM; In.vision Research; Ixiasoft; Oracle; Siberlogic; X-Hive; XyEnterprise

**Recommended Reading:** "Technology Overview for XML-Defined Content Processes"

"DITA-Aware Content Processing: Simpler, Cheaper, Faster"

"Case Study: Lessons Learned From Implementing a Darwin Information Typing Architecture Publishing System"

E-Mail Archiving

**Analysis By:** Carolyn DiCenzo; Kenneth Chin

**Definition:** Leading e-mail archiving solutions provide tools for capturing all, or selected, e-mail messages in a central repository for efficient storage and access. These tools can also prune messages from the active e-mail system data stores, leaving stubs that redirect users to the archive for viewing or retrieval. This activity is designed to keep the active message system data stores lean, to improve application performance and reduce recovery times. The need for users to maintain personal stores is eliminated and established stores can be migrated to the archive. Quotas can be used to trigger archiving or can be eliminated totally, with the stubbing being driven by a policy that usually is based on age and/or size of the message and related attachments. Legal discovery and content retention and search are used to meet discovery and compliance requirements. They also provide a way to export data for use with special purpose or more robust e-discovery tools. Tools for sampling and reviewing messages are available with many e-mail archiving products, in response to requirements specifically for the regulated portion of the financial industry. With the increase in the use of laptops for mobile workers, it is becoming increasingly important to provide users with the option to have a copy of their portion of the archive on their local disk.

**Position and Adoption Speed Justification:** The number of vendors offering e-mail archiving solutions continues to increase, with most offering archiving and mailbox management solutions appropriate for the markets that they target. The ability to put messages on legal hold has improved, but largely remain untested because many users have yet to set a time frame for deletion or, if one was set, refuse to approve the deletion. As more features within installed products are tested, or as company archiving requirements change, some companies will look to replace their initial product with another. A choice of products is available for all the major mail systems, and providers of mail solutions are continuing to provide better links within their mail systems for archiving vendors to leverage. Companies with large volumes of e-mail and long retention periods are putting stress on the scalability and reliability of current solutions, requiring improved index methods and, in some cases, major architectural changes. The need to support more than just e-mail and attachments, such as documents or presentations, is also causing vendors to redesign their offerings. Enterprises that bought tactical solutions or overbought and are unable to deal with the complexity are preparing to change to more-appropriate products. An increasing number of companies have implemented e-mail archiving for some part of their mail users. For those that have implemented archiving for all users, the opportunity still exists to sell
more capabilities to existing customers to extend the archive access to users or to provide more robust legal discovery support.

**User Advice:** As requirements to search and recover old e-mail messages grow, and in the face of increased demand for higher user quotas as users struggle to keep up with increased numbers of messages and larger messages, companies should implement an e-mail archiving solution now. Consolidating all archived messages into regional repositories or a centralized repository will support a quick response to inquiry and will enable a quick implementation of the official retention policy when it is ready. Migrating personal stores to the archive should be part of the deployment of an archive system. Look for archiving solutions that provide the additional option to archive files and other content types that also may be needed in the future.

**Business Impact:** E-mail archiving improves e-mail application performance, delivers improved service to users, and enables timely response to legal discovery and business requests for historical information. Archived data can be stored on less expensive storage technology, with the opportunity to even take some of the data offline or to delete it. Removing old data to an archive will also reduce backup and recovery time.

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

**Sample Vendors:** Atempo; Autonomy; AXS-One; Barracuda Networks; C2C Systems; CA; CommVault; EMC; Estorion; GFI; GFT inboxx; GWAVA; HP; H&S Software; IBM; Jatheon Technologies; MessageSolution; Messaging Architects; Mimosa Systems; Open Text; Quest Software; Sherpa Software; Symantec; Waterford Technologies; ZL Technologies

**Recommended Reading:** "What Is E-Mail Active Archiving?"

"E-Mail Active Archiving Market Update, 2008"

"Magic Quadrant for E-Mail Active Archiving, 2008"

**Folksonomies/Social Tagging**

**Analysis By:** Whit Andrews; Rita E. Knox

**Definition:** "Folksonomies" (the combination of "folks" and "taxonomies") are created with social tagging or user-defined metadata. End users (laypersons) don't explicitly create folksonomies; rather, they result from bottom-up categorization using individually developed or selected keywords. There's rarely a prescribed purpose, but a folksonomy evolves when many users create or store content at a particular site and then identify what they think the content is about. "Tag clouds" pinpoint various identifiers and the frequency of use on a folksonomy site. This type of grass-roots community classification (similar to other social networking approaches, such as blogs and wikis) is subject to vandalism and manipulation; nevertheless, it's a fine example of collective intelligence.

Don't confuse folksonomies with taxonomies. Folksonomies structure content via user tags; taxonomies are classifications defined by methods that don't necessarily include user-generated tags and rarely are viewed as similar.

**Position and Adoption Speed Justification:** Folksonomies are no longer simply consumer applications; rather, they're increasingly being used as corporate communication tools and as locations where users collect their bookmarks. Public folksonomies include Amazon.com (customer recommendation), del.icio.us (bookmark management), doof (social gaming), FaceTag
(people search), Furl (Web site management), Flickr (photo management), Technorati (blog search), 43 Things (goal sharing), CiteULike (academic papers), craigslist (online communities), LinkedIn (job networking) and Connotea (reference management). Many knowledge workers are using folksonomies to support their work.

**User Advice:** By definition, folksonomies fall outside formal manipulation or control. They may prove useful in acquiring and sharing corporate knowledge; formally, however, they can't be expected to do so. Setting up examples on a corporate site is one way to explore their usefulness. Folksonomies aren't a commercial technology, but they've been proposed as an approach to acquiring and documenting corporate knowledge. All companies are interested in successful methods of documenting such information, so they should investigate folksonomies as a potentially useful match to their corporate personalities.

**Business Impact:** As a form of personal knowledge management or personal productivity tools, folksonomies may make individuals more productive and may translate into enterprise productivity, but the causal relationship will be difficult to prove and the impact difficult to measure. However, folksonomies *could* feed improved search capabilities or be used as information services to support market intelligence (for example, who tags our products on Amazon.com and how?).

**Benefit Rating:** Moderate

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Adolescent

**Sample Vendors:** Amazon.com; craigslist; CiteULike; Connotea; del.icio.us; doof; Flickr; Furl; LinkedIn; MySpace; Technorati

**Recommended Reading:** "Use Content Analytics to Extract Meaning"

"Finding and Exploiting Value in Semantic Technologies on the Web"

"What Does Web 2.0 Mean to Government?"

**Wikis**

**Analysis By:** Nikos Drakos; Whit Andrews

**Definition:** A wiki is a simple collaborative system for creating and maintaining hyperlinked collections of Web pages. A wiki usually enables users to add or change pages "in context" without having to worry about where and how the content is physically stored. A wiki is, by design, an information synthesis tool that simplifies the modification and reorganization of existing information and encourages what is often referred to as "wiki gardening." This is the process of incrementally editing a wiki space to preserve continuity, make additional connections and links, and generally to clean it up. Key enabling functions are:

User-friendly "click to edit" features for any wiki page that invites users to create, link, edit or reorganize the information they see on the screen, without having to understand the physical file organization.

The ability to track changes, to compare different versions and to revert to a previous version makes it easier to make changes in the knowledge that any mistakes can be reversed easily.

Static Web addresses to any wiki page or even to any component of a wiki page (such as a paragraph). This means it's easier to organize related ideas using Web links, not only inside a
wiki but also between a wiki and other, Web-addressable sources of information (for example, other wikis, the intranet or the public Internet).

**Position and Adoption Speed Justification:** Wikis are now available from commercial vendors, in addition to many open-source products, and are becoming available from established enterprise vendors. Wikipedia is the best-known publicly available wiki, and it has raised awareness to the point where users often demand wiki-style collaboration support from their IT departments. Although the wiki functionality is still at the core of many products, it is beginning to defy clean categorization as it expands into adjacent areas either through plug-ins and extensions or by enhancements to the core product that typically includes blogs, discussions, user profiles and tagging.

**User Advice:** Understand that there are advantages to using a wiki over a conventional repository-style collaboration system, and be receptive to users who will argue that a wiki will improve team collaboration. Now is the time to move from pilots to more systematic deployments, as appropriate, and to evaluate the suitability of wikis in different collaboration scenarios, if this hasn't been done already.

**Business Impact:** Wiki-style information creation and sharing has significant advantages over traditional collaboration environments in terms of improved transparency, usability and information reuse. Like the public Web, it encourages the creation of a "web of interrelated information," where it is just as easy to create internal cross-references as it is to reference external resources. Within businesses, wikis are used as informal repositories for maintaining technical documentation, client communication, issue tracking, e-learning and training, general information sharing and knowledge management. They are also used to support communities of practice or communities of interest, product development and idea exploration.

**Benefit Rating:** Moderate

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

**Sample Vendors:** Atlassian; GroveSite; IBM; Jive Software; MediaWiki; Microsoft; MindTouch; Socialtext; TWiki

**Recommended Reading:** "A Blend of Wikis, Blogs and Discussion Forums Can Support Collaborative Business Activities"

**Climbing the Slope**

**E-Forms**

**Analysis By:** Toby Bell; Rita E. Knox

**Definition:** Electronic forms (e-forms) provide a user interface to data and services typically offered on the Web. E-forms enable users to interact with enterprise applications and back-end systems linked to them. Web applications, e-government and e-commerce solutions have sparked the demand for better Web forms that support richer and more dynamic interactions than are possible with existing HTML forms. New e-form applications include XML foundations, multiple data callouts, field-level validation and embedded process logic contained within a secure and often portable format.

**Position and Adoption Speed Justification:** Most external and internal horizontal and vertical processes have a forms component. E-forms can yield significant savings over their paper-based counterparts in design, development, production, inventory, delivery, data quality and integration
activities. In particular, e-forms can reduce the time and effort associated with paper-intensive processes, offering the potential of significant benefits. The further development of reusable solutions and standards for e-forms will accelerate adoption. Some major implementations and government support are recent stimuli for growth.

**User Advice:** Retail, insurance, and government organizations and agencies should develop a strategy for e-forms, and revamp their Web sites and back-end systems to introduce e-forms as quickly as possible, if only to minimize the costs of rekeying data. As e-forms supplement business process management and enterprise content management technologies, organizations should consider using them as a "lead in" to process improvement, replacing costly paper-based approaches and, in some cases, entire graphical user interfaces.

Convey a sense of urgency ("revamp as soon as possible") in many situations for many technologies. Gather information such as what vertical you are in (your insurance competitors are doing this, you should be too) or the amount of data that needs to be collected over the Web (online retailers). This will indicate who exactly should do this sooner rather than later.

**Business Impact:** E-forms offer data integration, a crisp user experience, process efficiencies and cost-avoidance potential. They can test the value of merging structured and unstructured content processing in a fairly low-cost effort. The environment is "user friendly," incorporates structured and transactional data, validates them and presents an easy to read "document" as a result.

The savings from eliminating production, delivery and data re-keying activities can offset much of the investment required to implement e-forms. Further benefits arise from simplified administration and better data quality through input validation logic and end-user updates on pre-filled fields. E-forms, by merit of the reduction in paper alone, warrant consideration by enterprises with a "green" leaning.

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Mature mainstream

**Sample Vendors:** Adobe Systems; Autonomy Cardiff; Cerenade; Formatta; IBM-FileNet; IBM (PureEdge); Microsoft

**Recommended Reading:** "Forms Technology to Transform Customer Communication"

**Records Management**

**Analysis By:** Kenneth Chin; Debra Logan

**Definition:** Records management (RM) technologies enable organizations to enforce the policies and rules for the retention and disposition of content required for documenting business transactions, and automate the management of their record-retention policies. These technologies, implemented with well-formulated and consistently enforced RM strategy and policies, form an essential part of organization-wide life cycle management of records. Records management principles and technology apply to any media, including non-electronic media.

**Position and Adoption Speed Justification:** RM was introduced 10 years ago, but focused mainly on paper documents and, until recently, lacked impetus for adoption. Legislation and regulations are driving organizations to meet similar compliance requirements for electronic content. RM has become a core component of enterprise content management (ECM) product suites and is being deployed as part of an overall ECM strategy. Integrated content archiving
platforms are also including varying levels of retention management capabilities. Major drivers in
the U.S. federal government include the eGovernment Act of 2002, the Office of Management
and Budget Circular A-123 (Sarbanes-Oxley-type requirements for financial management and
access controls) and the President's Management Agenda initiative (e-records management),
which address correspondence management, e-records management and permanent records
archiving. More recently, the Federal Rules of Civil Procedures has driven the demand for RM,
and we expect RM will move from departmental deployment to organization-wide
implementations.

User Advice: All public companies and government sector organizations should implement and
maintain a comprehensive RM program, supported by appropriate technology and processes.
Your records management program should be part of a broader enterprise content management
strategy. Records management programs are necessary if organizations want to manage the
amount of content they have over time, because a life cycle approach ensures that content is
managed, archived, stored and deleted as appropriate.

Business Impact: Global companies will adopt RM to comply with regulations, mitigate the risk
of litigation and liability, and meet legal discovery requirements. Many government organizations
worldwide are also implementing RM solutions. In most cases, this is being driven by compliance
with local RM legislation and regulation.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: ASG; Autonomy; CA; EMC; IBM; Iron Mountain; Objective; Open Text; Oracle;
Tower Software; Vignette; ZyLAB

Recommended Reading: "Toolkit: Seven Key Criteria for Evaluating Records Management
Solutions"

"MarketScope for Records Management, 2007"

Enterprise Content Management Suites

Analysis By: Mark Gilbert; Karen Shegda

Definition: Enterprise content management (ECM) suites combine document management, Web
content management, records management, document-centric collaboration, workflow and
document imaging components into a comprehensive suite provided by a single vendor. Content
management (CM) really constitutes a spectrum of capabilities, with basic content services (BCS)
tools at one end for basic and horizontal capabilities and ECM suites at the other for process-
centric applications.

Position and Adoption Speed Justification: The technologies that compose ECM suites are
mature as stand-alone technologies, but as tightly integrated, comprehensive suites from one
vendor, they are not yet mature. Some ECM vendors have done a better job of incorporating
acquired technology components into their suite. Others are still selling them as stand-alone
products or have bundled them very loosely. As enterprises realize that they must have an ECM
strategy, more enterprises will investigate vendors that have truly integrated the ECM
components. Enterprises that need high scalability and robust functionality may require an ECM
suite.
**User Advice:** Enterprises should inventory their content management applications and platforms. Those that mainly have content stored in file servers have an opportunity to start anew. However, the majority of enterprises have multiple ECM components, often from multiple vendors. Pressure for reducing complexity and cost of CM systems and trends toward standardization are creating conditions for significant change in the ECM market. IT leaders should aim to reduce the number of vendors and view CM as an infrastructure that will support multiple business applications. The first step is deciding which content-centric applications are critical to keep, potentially for the long term and which, if any, of the existing vendors can address your needs with an ECM suite. Next, enterprises should attempt to integrate those repositories where it makes sense and develop exit strategies for the content applications that are not considered critical or that may be addressed by a suite vendor. Enterprises should also evaluate how ECM efforts connect to their information-centric infrastructure and support business needs to share and exchange all content (both structured and unstructured) as part of a program for enterprise information management.

**Business Impact:** ECM suites support many content-enabled vertical applications (CEVAs) and some horizontal uses, such as regulatory compliance; mergers and acquisitions, and deal rooms; intranet, extranet and public Web site publishing; plant and facilities management; marketing; e-commerce; and employee, customer and investor relations.

**Benefit Rating:** Moderate

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

**Sample Vendors:** EMC; IBM; Interwoven; Open Text; Oracle; Vignette

**Recommended Reading:** "Magic Quadrant for Enterprise Content Management, 2007"

"Predicts 2007: Information Infrastructure: Content Matters"
Appendixes

Figure 3. Hype Cycle for Content Management, 2007
# Hype Cycle Phases, Benefit Ratings and Maturity Levels

## Table 1. Hype Cycle Phases

<table>
<thead>
<tr>
<th>Phase</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology Trigger</td>
<td>A breakthrough, public demonstration, product launch or other event generates significant press and industry interest.</td>
</tr>
<tr>
<td>Peak of Inflated Expectations</td>
<td>During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the technology is pushed to its limits. The only enterprises making money are conference organizers and magazine publishers.</td>
</tr>
<tr>
<td>Trough of Disillusionment</td>
<td>Because the technology does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.</td>
</tr>
<tr>
<td>Slope of Enlightenment</td>
<td>Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the technology's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.</td>
</tr>
<tr>
<td>Plateau of Productivity</td>
<td>The real-world benefits of the technology are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.</td>
</tr>
<tr>
<td>Years to Mainstream Adoption</td>
<td>The time required for the technology to reach the Plateau of Productivity.</td>
</tr>
</tbody>
</table>

Source: Gartner (March 2008)

## Table 2. Benefit Ratings

<table>
<thead>
<tr>
<th>Benefit Rating</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformational</td>
<td>Enables new ways of doing business across industries that will result in major shifts in industry dynamics</td>
</tr>
<tr>
<td>High</td>
<td>Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise</td>
</tr>
<tr>
<td>Moderate</td>
<td>Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise</td>
</tr>
<tr>
<td>Benefit Rating</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>Low</td>
<td>Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings</td>
</tr>
</tbody>
</table>

Source: Gartner (March 2008)

Table 3. Maturity Levels

<table>
<thead>
<tr>
<th>Maturity Level</th>
<th>Status</th>
<th>Products/Vendors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryonic</td>
<td>• In labs</td>
<td>• None</td>
</tr>
<tr>
<td>Emerging</td>
<td>• Commercialization by vendors</td>
<td>• First generation</td>
</tr>
<tr>
<td></td>
<td>• Pilots and deployments by industry leaders</td>
<td>• High price</td>
</tr>
<tr>
<td></td>
<td>• Uptake beyond early adopters</td>
<td>• Much customization</td>
</tr>
<tr>
<td>Adolescent</td>
<td>• Maturing technology capabilities and process understanding</td>
<td>• Second generation</td>
</tr>
<tr>
<td></td>
<td>• Uptake beyond early adopters</td>
<td>• Less customization</td>
</tr>
<tr>
<td>Early mainstream</td>
<td>• Proven technology</td>
<td>• Third generation</td>
</tr>
<tr>
<td></td>
<td>• Vendors, technology and adoption rapidly evolving</td>
<td>• More out of box</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Methodologies</td>
</tr>
<tr>
<td>Mature mainstream</td>
<td>• Robust technology</td>
<td>• Several dominant vendors</td>
</tr>
<tr>
<td></td>
<td>• Not much evolution in vendors or technology</td>
<td></td>
</tr>
<tr>
<td>Legacy</td>
<td>• Not appropriate for new developments</td>
<td>• Maintenance revenue focus</td>
</tr>
<tr>
<td></td>
<td>• Cost of migration constrains replacement</td>
<td></td>
</tr>
<tr>
<td>Obsolete</td>
<td>• Rarely used</td>
<td>• Used/resale market only</td>
</tr>
</tbody>
</table>

Source: Gartner (March 2008)

RECOMMENDED READING

"Cool Vendors in Content Management, 2007"

"Magic Quadrant for Enterprise Content Management, 2007"

"Understanding Gartner's Hype Cycles, 2008"

This research is part of a set of related research pieces. See "Gartner's Hype Cycle Special Report for 2008" for an overview.
REGIONAL HEADQUARTERS

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

European Headquarters
Tamesis
The Gianty
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
9° andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509