Research Summary: Enterprise Content Management

January 4, 2007
Prepared by: Robert Woolley, Chief Technologist and Strategic Planner
David Fletcher, Chief Technology Officer

Introduction
This research report was prepared in response to a request for information on enterprise document management systems which are now most generally marketed and packaged as a component of an enterprise content management (ECM) solution.

The Enterprise Content Management Association (AIIM) has defined ECM as “the technologies, tools, and methods used to capture, manage, store, preserve, and deliver information, content and documents related to organizational processes. ECM tools and strategies allow the management of an organization’s unstructured information, wherever that information exists.”

The document management component of ECM generally focuses on managing unstructured content so that it is more easily managed and accessible to enterprise resource users. Document management system solutions provide capabilities for:

- capture and storing digital images of documents in single or multiple integrated repositories;
- integration with business process management;
- access to documents with appropriate controls; and,
- retention, archiving, and storing of documents.

Scope
This research review is primarily concerned with enterprise class systems for ECM and related document management and imaging subsystems that are available commercially and from open source vendors.

---

Research Summary

There are three major sources available that look at ECM and related disciplines such as document management and imaging that provide a comprehensive view of key vendors and issues in this market’s space.

- Gartner provides a Magic Quadrant for Enterprise Content Management\(^2\) that identifies key vendors and their performance from a magic quadrant perspective.
- Forrester has released a Forrester Wave report on Content Centric Applications\(^3\) that looks at key ECM vendors using the Wave perspective.
- AIIM, the ECM association, has released a comprehensive Guide to ECM Purchasing\(^4\) that provides excellent background information on selecting ECM systems and identifies all major ECM vendors and solution providers in a provider matrix.

Numerous other white papers and reports are also available from major ECM vendors that provide additional background research and information.

Gartner has characterized the ECM marketplace as being in a state of rapid change as basic content and document management solutions migrate from historical silo implementations to becoming a core component of IT infrastructure. Gartner has identified ECM suites as encompassing all of the following core components:

- **Document Management** for check-in/check-out, version control, security, and standard document library services.
- **Web Content Management** for managing dynamic content and user interaction.
- **Records Management** for archiving and automation of retention and compliance policies and for ensuring regulatory compliance.
- **Document Imaging** for capturing, transforming, and managing paper documents.
- **Document Centric Collaboration** for sharing documents and supporting project teams.
- **Workflow** for supporting business processes, task assignments, and creating audit trails.

Additional features are appearing that support document archive and retrieval systems, electronic forms content capture, digital asset management, and e-mail archiving for compliance and discovery.

Gartner provides a magic quadrant illustrated in Figure 1 that looks at vendors from the perspectives of completeness of vision and ability to execute. Each vendor has been assigned a placement in the quadrant based upon Gartner’s view of the vendor position in the ECM marketplace. The Gartner analysis does not consider any open source vendors,

---


and included vendors had to have at least $10 million dollars in ECM license and support revenue to be included.

Key vendors identified as industry leaders from both perspectives included EMC Documentum, IBM, Open Text, and Stellent. Vignette, Interwoven, and Hyland also were identified as key players that were evolving rapidly toward best-in-class providers of ECM products.

The Magic Quadrant approach, while one of the most widely used in the industry, is of most value when its results are compared to actual user experience with ECM products. Results in these kinds of scenarios may identify different sets of vendors that are the strongest alternatives for State government.

Forrester, by contrast, has looked at this same market area from three different perspectives: transactional, business content, and persuasive. A Forrester Wave chart on transactional content centric applications is presented in Figure 2.
Transactional ECM applications support people-centric processes by aligning content resources with ongoing business processes. These types of applications may capture computer based data, digital images, and e-mails and integrate them with business process management (BPM) and workflow tools.

Business content ECM applications, as illustrated in Figure 3, include applications that focus on providing content information in support of employee collaboration and decision making.
The third category of ECM capabilities that were analyzed by Forrester (illustrated in Figure 4) included persuasive content applications that drive content to better support the end user customer experience by making information broadly available irrespective of location or format.

Forrester concludes that these three areas represent the most in-demand areas of content management required by line of business users. Forrester provided additional detailed analysis that looks at vendors in each of these areas from an assessment of their current offering, strategy, and market presence perspectives. Based on this type of analysis, the key vendors in a leadership position were as follows:

**Transactional Content Application ECM Vendors**
FileNet, IBM, and Hyland all do an excellent job with high performance document imaging and workflow integration. EMC Documentum, Vignette, and Open Text also offer strong document imaging and management capabilities.

**Business Content Centric Application ECM Vendors**
Hummingbird, Interwoven, and Open Text are clearly the leaders in this area with strong document and records management capabilities. EMC Documentum, IBM, and Vignette also offer strong capabilities in this area.

**Persuasive Content Application ECM Vendors**
FatWire, Tridion, and Interwoven are the market leaders in this area of functionality. Most other vendors have only limited capabilities.

**Architecture Implications**
A comprehensive ECM architecture is complex and has many different components. Doculabs has developed useful reference architecture (see Appendix B) for a complete ECM environment. This architecture is illustrated in the *AIIM Guide to ECM Purchasing* and has been widely referenced in many industry sources. The most common technology components of the reference architecture are illustrated in Figure 5.
Implementation of ECM using service oriented architecture (SOA) is a key emerging ECM trend and should be part of the deployment evaluation criteria for building ECM infrastructure. Similarly the identification of key infrastructure components that can be leveraged by specialized imaging applications appears to be an important issue for long term infrastructure development.

These basic components offer opportunities to build an ECM infrastructure that can be leveraged to support agency specific applications for ECM, such as document management or electronic forms. Many of the other components already exist, but the ECM architecture brings them together in a content management context.

**Commercial Alternatives**

Vendor decisions are of course based upon clearly defined requirements. The emphasis of the document management solution needed by the State at this point focuses most heavily on transactional and content based requirements for capturing, managing, and retrieving many different kinds of documents representing unstructured content. Based upon this perspective, the key commercial vendors most likely to meet the States requirements include FileNet, IBM, Hyland, Documentum, Open Text, and possibly Interwoven. Hummingbird is also a possibility, but lacks strength from a transactional perspective.

**Open Source Alternatives**

There are two principle enterprise ECM solutions available from the open source community that are consistent with the overall technical architecture of the State and that have the apparent capacity and capability to perform as enterprise solutions:

---

• Nuxeo (<http://www.nuxeo.com/en>) claims to be the leading open source ECM software platform. The product offers a comprehensive group of components to create, process, manage, publish, and archive content, and has a fairly extensive user base.

• Alfresco (<http://www.alfresco.com>) offers open source Enterprise Content Management (ECM), including document management, collaboration, records management, knowledge management, Web content management, and imaging. The company leverages best-of-breed open source technologies such as Spring, Hibernate, Lucene, and standards such as JSR-168, JSR-170, Web Services, and Java Server Faces. The product is supported by a team with substantial experience, including the co-founder of Documentum, and a number of former Interwoven employees.

Both of these platforms appear to be comprehensive and scalable, and should be evaluated as possible enterprise solutions. Support and documentation for both products is available and seems to be mature enough to be usable by the State.

**Summary and Conclusion**

Understanding the business problem that ECM is trying to solve is of primary importance. There are four key areas that need to be considered before making a solution decision:

- How do agencies create, capture, and acquire content? Are there substantial differences that may make a single solution difficult to implement?
- What controls and methods exist for managing and controlling the flow of information created by agencies?
- How will this information be stored and archived to ensure preservation and reuse?
- How will information captured by an ECM component be shared, delivered, and output to users of information and communication assets?

A significant trend identified by Gartner is the development of basic content management solutions, such as digital imaging, to become part of the IT infrastructure using major software stack providers such as IBM, Microsoft, Oracle, APD, etc. This is causing the vendor landscape to migrate toward content-enabled vertical applications (CEVAs) that address specific content management requirements of disciplines. This model is causing substantial consolidation in the ECM marketplace and may suggest a set of assumptions as to which ECM services are part of a common infrastructure and which services may need to be specific to agency requirements.
Even if the State is focusing solely on document management as an ECM component, implementation is not without issues and challenges. Doculabs\(^6\) has identified the following issues that need to be addressed for a successful implementation:

**Lack of a Repeatable Implementation Approach**
“Organizations struggle with creating consistent deployment processes, often because ECM capabilities are “one part” infrastructure and “one part” application.”

**Poor Adoption Rates**
“Adoption rates are only 40% to 60% of what was initially projected. Poor adoption is not due to technology, but rather to a high level of complexity and internal dynamics.”

**User Dissatisfaction with IT**
“Consistently, department heads complain that IT does not understand their business objectives, processes, or pains.”

**Over-customization**
“IT over-customizes the product, increasing both initial development and ongoing maintenance costs beyond what is acceptable to users.”

**Poor Problem Diagnosis**
“Organizations experience trouble diagnosing shortcomings related to stalled ECM implementations.”

**Metrics and Measurement**
“Organizations are ineffective at measuring the impact ECM has had on business metrics, thus unable to garner follow-on investment.”

**Funding Allocations for Enterprise Deployment**
“Rolling out multiple ECM applications across the organization becomes a chaotic, unpredictable budgeting exercise that fails to demonstrate tangible cost savings and benefits”

To the extent that the State anticipates and mitigates these issues, and most importantly provides adequate project management support, an ECM implementation can be successful and meet original business requirements.

---

References
AIIM, Enterprise Content Management: The AIIM Guide to Purchasing, 2007


Appendix A.

ECM Definitions and Terminology
Definitions for this report have been selected from the *AIIM Guide to ECS Purchasing, 2007*. This list is not comprehensive and includes some terminology not used in this document. A more complete listing of ECM terminology can be found in *ANSI/AIIM TR 2—Glossary of Document Technologies*.

**Aggregation**
The process of combining data inputs from different creation and authoring tools and other systems.

**Audit Trails**
Log of changes for accountability.

**Categorization**
Organizing documents, Web pages, and other content into logical groupings based on their content.

**CD-ROM (Compact Disc Read Only Memory)**
Optical disc that is created by a mastering process and used for distributing read-only information.

**Check In/Out**
Ensures that only one person can work on a document at any time.

**Collaboration**
Tools (collaborative authoring, video conferencing, shared whiteboards, etc.) that allow multiple users to work on the same content in a common environment.

**Compression**
Technique used to reduce the number of bits in a digital image file. JPEG and GIF are two examples.

**Content Addressed Storage (CAS)**
Storage methodology designed for rapid access to fixed content.

**Content Management System**
The capability to manage and track the location of, and relationships among, content within a repository.

**Data Warehouse**
Central repository for all, or most, of an organization’s structured data.

**Digital Rights Management**
Enables secure distribution, and disables illegal distribution, of paid content over the Web.
Disposition
An action taken after a record is no longer needed for current business. Archiving and destruction are two possible actions.

Distributed Capture
Strategy for getting documents into the business process in decentralized locations across a company rather than sending all documents to a central location for scanning.

DoD 5015.2-STD
Department of Defense is the standard for evaluating electronic records management applications used within the DoD. It has been endorsed by the National Archives and Records Administrations. Many records management products use this standard.

Document Imaging
Process of capturing, storing, and retrieving documents regardless of original format, using micrographics and/or electronic imaging (scanning, OCR, ICR, etc.).

Document Management
Software that controls and organizes documents throughout an enterprise. Incorporates document and content capture, workflow, document repositories, output systems, and information retrieval systems.

E-Forms/Web Forms
Forms designed, managed, and processed completely in an electronic environment.

File System
The way in which files are named and where they are placed logically for storage and retrieval, most commonly in a hierarchical (tree) structure.

Forms Processing
The ability for software to accept scanned forms and extract data from the boxes and lines to populate databases. Software usually includes the ability to drop out the form so that recognition accuracy improves. Intelligent Document Recognition automatically identifies document types from the layout and structure of the document.

Full-Text Indexing and Search
All words in a document are indexed, enabling the document to be retrieved by words or phrases within the document.

HCR (Handprint Character Recognition)
OCR technology designed to turn images of hand printed characters into ASCII code.

ICR (Intelligent Character Recognition)
An advanced form of OCR technology that may include capabilities such as learning fonts during processing or using context to strengthen probabilities of correct recognition or that can recognize hand printed characters.
Indexing
Identification of specific attributes of a document or database record to facilitate retrieval.

Input Designs
Templates used to enable authors to more easily enter content into a system, typically customized, based on the type and format of content to be entered.

JPEG (Joint Photographic Experts Group)
Image compression format for storing color photos and images. There are multiple JPEG formats.

OCR (Optical Character Recognition)
Technique by which images of characters can be machine-identified, then converted into codes that can be processed by a computer.

OMR (Optical Mark Recognition)
Detects presence, or absence, of marks in defined areas; used for processing questionnaires, standardized tests, etc.

PDF (Portable Document Format)
Format developed by Adobe Systems for document publication.

Records Management
Enables an enterprise to assign a specific life cycle to individual pieces of corporate information from creation, receipt, maintenance, and use to the ultimate disposition of records. A record is not necessarily the same as a document. All documents are potential records, but not vice versa. A record is essential for the business; documents are containers of “working information.” Records are documents with evidentiary value.

Records Retention
The process of determining how long an organization needs to keep its records, taking into account the operational business needs, legal, and/or regulatory requirements.

Repositories
Part of a Document Management system with specific functionality to control the check-in/out of material, version control, and look-up against defined attributes.

Retrieval
Procedure for searching for and extracting database records or content that allows movement of data between computer systems.

Scalability
The ability of a system to expand capacity and number of users.
**Service-Oriented Architecture (SOA)**
Strategy for loosely-coupling applications together without the need to customize links between the applications to do so.

**Syndication**
Supply of content for reuse and integration with other material, often through a paid subscription.

**Taxonomy**
Way to structure and categorize content. Usually hierarchical, categories (nodes) in the hierarchy progress from general to specific. Each subsequent node is a subset of the higher level node. There are three basic types of hierarchical taxonomies: subject, business-unit, and functional.

**TIFF (Tag Image File Format)**
Widely used image file structure that consists of a series of headers or tags, plus the image data. There are many choices among the tabs used, such as type of image, compression used, resolution, colors planes, bit sequences, and annotations.

**Transformation**
Changing content from one format to the needed delivery format.

**Version Control**
Procedures to identify the authorship and the sequence of different versions of a document.

**XML (eXtensible Markup Language)**
An established standard, based on the Standard Generalized Markup Language, designed to facilitate document construction from standard data items. Also used as a generic data exchange mechanism.

**Web Content Management**
A technology that addresses the content creation, review, approval, and publishing processes of Web-based content.

**Workflow/BPM (Business Process Management)**
Automation of business processes, in whole or in part, where documents, information, or tasks are passed from one participant to another for action, according to a set of rules. A business process is a logically related set of workflows, work steps, and tasks that provide a product or service to customers. BPM is a mix of process management/workflow with application integration technology.
Appendix B. ECM Reference Architecture (Part 1)
Appendix B. ECM Reference Architecture (Part 2)

The components illustrated in this reference architecture comprise an overview of the major elements of a comprehensive ECM infrastructure. The intent of this reference architecture is to view ECM as IT infrastructure that provides support for key ECM components like Document Management and Imaging as they are needed by the enterprise.

Characteristics of the ECM Reference Architecture

Doculabs7 has described the characteristics of the ECM reference architecture as:

- **Service-oriented** – “Allows applications to be broken into services that can be accessed by other applications and systems to create powerful composite applications based on the functionality available in applications across the enterprise.”
- **Event-driven** – “Provides a fundamental mechanism to capture key changes in business needs and technical implementation. These changes can then be used to effect instantaneous changes to business processes and the underlying systems that support them.”
- **Aligned with lifecycle support processes** – “Organizations are constantly designing, deploying, managing, and re-evaluating their applications. Until now, the process of making decisions on design, development, and optimization has not been based on empirical evidence and real data about application usage patterns and business model behavior. Going forward, architectures must provide the ability to collect, disseminate, and use this information to help organizations make better decisions.”
- **Able to support assembly and integration** – “Once applications are segmented into smaller functional units, the ability to assemble these components into applications is critical. In the past, writing code was the only way to achieve the goal. Today, process management technology achieves the same goals while reducing reliance on costly code development.”
- **Able to leverage existing applications and infrastructure** – “As organizations look for different ways to minimize unnecessary technology spending, they are looking for ways to re-use existing technology. For most organizations, existing infrastructure, systems, and applications are home to the core data and functions that drive the business forward day to day. These systems must be leveraged to provide maximum benefit.

“A sound service-oriented reference architecture embodies each of the architectural constructs described above. That is, it is event-driven, aligned with lifecycle support processes, able to support assembly and integration, and able to leverage existing applications and infrastructure.”