



California
DEPARTMENT OF TECHNOLOGY

California Enterprise Architecture Framework

Enterprise Content Management (ECM) Reference Architecture (RA)

Version 1.0 Final
January 2, 2014



This Page is Intentionally Left Blank



TABLE OF CONTENTS

1	Introduction.....	1
1.1	Purpose	1
1.2	Limitations.....	2
1.3	Intended Users.....	2
1.4	Document Organization.....	2
1.5	Future Directions	2
2	Enterprise Content Management Overview	3
2.1	Definitions.....	3
2.2	Business Benefits	5
2.3	Usage Scenarios	6
2.4	Key Capabilities of an ECM Solution	6
2.4.1	<i>Content Lifecycle Capabilities</i>	<i>6</i>
2.4.2	<i>Collaboration Capabilities.....</i>	<i>7</i>
2.4.3	<i>Integration and Inter-Operability Capabilities.....</i>	<i>7</i>
2.5	Components of ECM Solutions	8
2.5.1	<i>ECM Capture Components</i>	<i>10</i>
2.5.2	<i>ECM Deliver Components.....</i>	<i>11</i>
2.5.3	<i>ECM Manage Components</i>	<i>12</i>
2.5.4	<i>ECM Store Components</i>	<i>13</i>
2.5.5	<i>ECM Preserve Components.....</i>	<i>14</i>
2.5.6	<i>ECM Development and Administration Components</i>	<i>14</i>
3	ECM Reference Architecture Description.....	15
3.1	ECM RA Conceptual View	15
4	Glossary	18
5	References	19
5.1	Federal and State Documents.....	19
5.2	Books and Papers.....	19
5.3	Web Sites	20
6	Document History	21



LIST OF FIGURES

<i>Figure 2-1 AIIM's ECM 101 Poster</i>	4
<i>Figure 2-2 AIIM's Five ECM Areas</i>	9
<i>Figure 2-3 Overview of the Main Groups of Components</i>	10
<i>Figure 2-4 ECM Capture Components</i>	11
<i>Figure 2-5 ECM Deliver Components</i>	12
<i>Figure 2-6 ECM Manage Components</i>	13
<i>Figure 2-7 ECM Content Repositories (Store and Preserve Components)</i>	13
<i>Figure 2-8 ECM Development and Administration Components</i>	14
<i>Figure 3-1 ECM Reference Architecture – Conceptual View</i>	16



LIST OF TABLES

Table 6-1 Document History 21



This Page is Intentionally Left Blank

1 Introduction

Enterprise Content Management (ECM) can be thought of as an outgrowth of a number of earlier attempts to deal with digital “content” – such as Document Management systems (“DM” - that focused on digital documents, or more precisely on management of files), or Content Management systems (“CM” or “WCM” that focused on managing content components, especially for websites and portals). Compared to the previous approaches, today’s ECM solutions can face huge volumes of information, which can grow – according to some estimates – as fast as 200% per year. The main delivery channel used to be thick clients (in DM), and more recently also web sites and portals, while in case of ECM the distribution and delivery of digital content can take place using any available channel, including Web Services and the ESB. ECM’s predecessors functioned within limited organizational scope, whereas ECM – as the name indicates – targets the whole enterprise, which entails dealing not only with technology, but also with organizations, processes, and people.

Given that it encompassed DM and CM systems, ECM can be framed as an evolutionary step with respect to previous ways of handling digital content. However, ECM has a fundamentally different focus compared to the traditional content management systems. The main goal of ECM is to capture, manage, store, preserve and deliver content *related to organizational processes*. ECM is not technology-driven as many traditional approaches have been, but rather is business processes-driven. Accordingly, ECM can be treated as an integrative approach for managing the entirety of an organization’s information assets, with strong links (and some similarities) to Business Intelligence (BI) and Enterprise Application Integration (EAI) domains.

Similarly, there have been significant changes with respect to technology between DM- and CM-based solutions on one hand, and ECM on the other hand. With the adoption of SOA and in SOA-driven perspective, ECM solutions are increasingly viewed as collections of cohesive *services* rather than a collection of content-related products or applications. Moreover, ECM solutions are increasingly considered for deployment in the Cloud, given its promise of increased usability and decreasing costs. All these factors shape the basic perspective on ECM in CEAF 2.0.

1.1 Purpose

The ECM Reference Architecture document provides guidelines and options for making architectural decisions when implementing ECM solutions.

The objectives for the document include the following:

- To introduce key terms and distinctions relevant for the topic
- To provide inputs for creating or evaluating architectures for ECM
- To identify building blocks (architectural layers, services, components) for integrating elements of an ECM solution
- To communicate the key architectural decisions relevant for creating or evaluating ECM solutions
- To communicate opportunities for solution and/or platform sharing at agency, cross-agency and/or state levels.

1.2 Limitations

The document focuses on ECM and related concepts at the enterprise architectural level in the context of CEAF 2.0 as relevant to the Reference Architecture. It does not include other areas relevant to planning and introducing an ECM solution, such as governance and policies, or legislative, regulatory or standards-related constraints and challenges that may apply to a given ECM solution. A practical implementation of an ECM solution has to include those aspects. Reference architecture, however, is concerned with how such aspects can be accommodated in general at the architectural level. This document is also not intended as an exhaustive introduction to ECM or related products available in the market today. Even though intention for the Reference Architecture is to provide examples of realization using specific products used in production in the State, this document is not intended as a product guide.

1.3 Intended Users

The primary intended users of this document are Enterprise Architecture practitioners and other architects that contribute to enterprise architecture. This broad group includes architects from other domains/disciplines such as Security, Application, Information, Business, Technology, Infrastructure, and Solution Architects. It is also beneficial to Managers, at senior or operational levels, who are involved with ECM or related areas, such as Service-Oriented Architecture, Cloud Computing, Identity and Access Management, and similar areas.

1.4 Document Organization

The ECM Reference Architecture documentation is organized as follows:

- Section “Enterprise Content Management Overview” provides background for the ECM RA by introducing descriptions and definitions of ECM, discusses the main usage scenario types found in ECM implementations, and identifies architectural components for respective usage scenarios.
- The section “ECM Reference Architecture Description” elaborates ECM RA using the Conceptual View (in the section “ECM RA Conceptual View”) which introduces the necessary capabilities for an ECM architecture and how they are supported by Architectural Building Blocks (ABBs)
- The section “Glossary” provides description of the terms and abbreviations used in the document
- The section “References” lists publications used for preparation of the document.

1.5 Future Directions

Future evolution of the document includes the following steps:

- Addition of an example or examples of existing realization of the ECM RA
- Identification and elaboration of solution sharing opportunities
- Formulation of implementation guidelines for ECM RA.

2 Enterprise Content Management Overview

This section provides a description of ECM, including clarification of key terms and concepts. It identifies ECM's intended business benefits and summarizes its main usage scenarios. A set of key capabilities of ECM solution are identified in this section and key components of the solution are described at a high level.

2.1 Definitions

Following the guidance from the Association for Information and Image Management (AIIM), CEAF adopts the following definition of ECM:

Enterprise Content Management (ECM) is the strategies, technologies, tools and methods used to capture, manage, store, preserve, and deliver content across an enterprise. It is important to emphasize that this content is related to and used by organizational processes, rather than simply content for its own sake.

According to Gartner, ECM refers to both a strategy to deal with all types of enterprise content and a set of software products for managing the entire life cycle of that content. At the most basic level, ECM strategies, processes, workflows, and tools allow for managing of an organization's information, regardless of the source of that information and the type of that information.

ECM solutions typically operate on unstructured (multimedia), semi-structured (e.g., emails, spreadsheets) information, although they can also allow management of structured content (such as data in relational databases or XML documents).

A complete ECM implementation typically involves many processes, tools and technologies with the common purpose of managing the lifecycle of digital content in support of the organizational processes. The areas covered by an ECM solution include the following:

- Creation of digital content using capturing, scanning (Production Imaging), or cross-format transformations
- Management of digital content, including creation, versioning, repurposing of the content, retention policies, and access control
- Document-centric collaboration
- Workflow and Document-centric Business Process Management
- Web Content Management
- Forms Management
- Digital Asset Management
- Records Management
- Search, Content Analytics, E-mail and Information Archiving, Packaged Application Integration

In a given context/organization, only some types of digital content can be relevant and all the capabilities offered by ECM may or may not be needed.

There are a number of ways of grouping the areas included in an ECM solution. The top level grouping provided by AIIM is shown in the following poster figure:

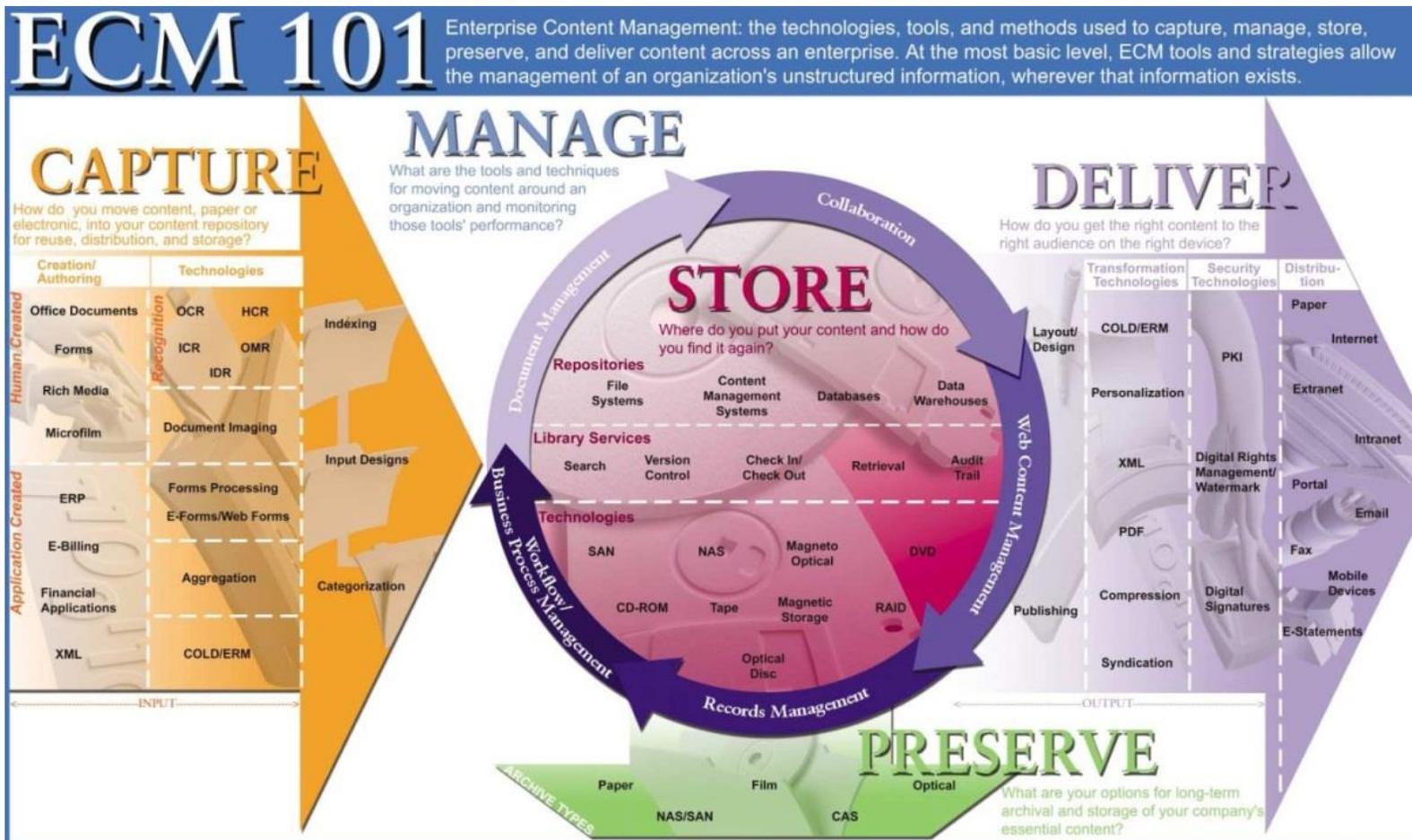


Figure 2-1 AIIM's ECM 101 Poster

According to AIIM, there are five key areas in an ECM, shown in the above figure:

- **Capture**, which is concerned with creation, importing of content, using physical or digital sources, and driven either by humans or applications
- **Manage**, which is concerned with lifecycle of the content and its movements in the organization
- **Store**, which deals with persisting and accessing the content
- **Preserve**, which deals with long-term storage and archival of relevant content
- **Deliver**, which is concerned with delivery of right content to the right recipients using right channel/device.

The above areas and main components in those areas are described later in the document.

2.2 Business Benefits

The most common drivers for adoption of ECM are as follows [9]:

- **Efficiency**: cost and time savings when managing various types of digital content; savings resulting from good availability and from avoiding unnecessary work (such as re-creation of the same content)
- **Availability**: making the available content easily navigable and searchable (including metadata-based querying); also, making it easy to distribute and re-use, while observing applicable regulatory and legal constraints
- **Traceability**: visibility of the lifecycle of specific content elements and ability to control the lifecycle of content to meet standards and regulations
- **Consistency**: unifying the content across an organization and standardizing the procedures of producing, reviewing, and auditing the content.

Case studies devoted to ECM [10] identified the business impacts of adoption of ECM:

- Direct cost savings in information processing operations and facilities
- Improved internal and external collaboration, involving knowledge creation and sharing through digital content in and among enterprises with commonly enacted practices
- Increased efficiency, effectiveness, and flexibility of knowledge work and business processes, including reuse of previously created content, metadata, templates, and navigation aids
- Improved reliability and quality of information content resulting in less errors in products and services
- Improved organizational memory recording the practice, history, and transactions of the enterprise
- Providing value-added or new customer services and products involving digital content;
- Improved modern and professional image of the enterprise in the eyes of its stakeholders
- Improved compliance with external regulations and standards, directly or indirectly governing the enterprise
- Obtaining capabilities to quickly develop targeted content management applications for emerging purposes.

2.3 Usage Scenarios

There are a number of types of usage scenarios for ECM, as follows:

- Content creation
- Content retrieval
- Content modification, potentially in a collaborative process
- Content delivery, potentially involving format transformation or transcoding, and using delivery channel appropriate to the content type.

2.4 Key Capabilities of an ECM Solution

The key capabilities of ECM can be grouped as follows:

- Content lifecycle-related capabilities
- Collaboration-related capabilities
- Integration- and inter-operability capabilities

The above groups of capabilities are presented in the subsections that follow.

2.4.1 Content Lifecycle Capabilities

ECM capabilities related to the lifecycle management of content include the following:

- Content creation, such as:
 - Using appropriate forms of capturing (such as scanning, recognition and form-processing in case of digital documents)
 - Transforming from one digital format into another
 - On-demand transactional content capture
 - Item classification, including extracting and processing applicable metadata and/or its contents
 - Analysis of composites into its parts and applicable decomposition (such as a web page into textual, visual, and audio parts)
- Routing and storing the content in appropriate type of store, including de-duplication
- Item retrieval using searches on metadata or contents indexes, including concurrent access and locking
- Access control and digital rights enforcement, where applicable
- Various forms of rendering of an item and applicable cross-format transformation
- Modifying the item, using tools appropriate to the item format, and
 - Item review and approval workflow
 - Content routing to workers, assigning work tasks and states
 - Item version control and creating audit trails
- Lifespan enforcement (e.g., enforcement of retention policies).

Given that the lifecycle of a given type of content can be managed using pre-defined workflows, additional ECM capabilities include:

- Graphical workflow and business process building
- Making standard activities (such as push to web site, electronic forms validation, security enforcement, etc.) within the workflow pre-defined and available

The capabilities listed above apply, in principle, to all kinds of digital content. However, some types of content do require specific capabilities. For example, records need to be managed using long-term retention of content through automation and policies (based on a records retention schedule), and with legal, regulatory and industry compliance (certified compliance with standards such as DoD Directive 5015.2-STD).

2.4.2 Collaboration Capabilities

ECM collaboration capabilities include the following:

- Concurrent access to existing content items
- Delivery of content items to tools supporting dissemination (such as blogs or wikis) or collaboration
- Collaboration when creating or changing a content item (potentially a composite item)
- Supporting reviewing of a changed content item and an approval workflow
- Creating change sets, audit logs

In government context, the main type of content subject to collaboration is digital documents (potentially with associated assets). However, collaboration on other type of contents (such as sound or video) may require specialized tools outside of the scope of an ECM; at the same time, interoperability capabilities of an ECM solution would be expected to make using such dedicated tools feasible.

2.4.3 Integration and Inter-Operability Capabilities

Integration capabilities of an ECM solution include the following:

- Integration with enterprise IdAM systems
- Integration with the EAI platforms
- Integration with standard infrastructure services, such as scheduling, logging, auditing services.

Interoperability (on-premises and in the cloud) capabilities of an ECM solution include the following:

- Interoperability with tools used to edit or modify a given type of item (such as digital document or audio file)
- Interoperability with Business Process/workflow execution platforms and tools
- Interoperability with format transformation tools and on-line channel optimization
- Interoperability with the business event propagation mechanism (e.g., ability to emit and to consume events)

2.5 Components of ECM Solutions

According to AIIM, there are five key areas in an ECM as already shown in the figure in the section “Definitions”:

- **Capture**, which is concerned with creation, importing of content, using physical or digital sources, and driven either by humans or applications
- **Manage**, which is concerned with lifecycle of the content and its movements in the organization
- **Store**, which deals with persisting and accessing the content
- **Preserve**, which deals with long-term storage and archival of relevant content
- **Deliver**, which is concerned with delivery of right content to the right recipients using right channel/device.

The following figure (from AIIM) portrays a flow-like relationship between the key ECM areas. Each of the five ECM areas is described in the subsections that follow.

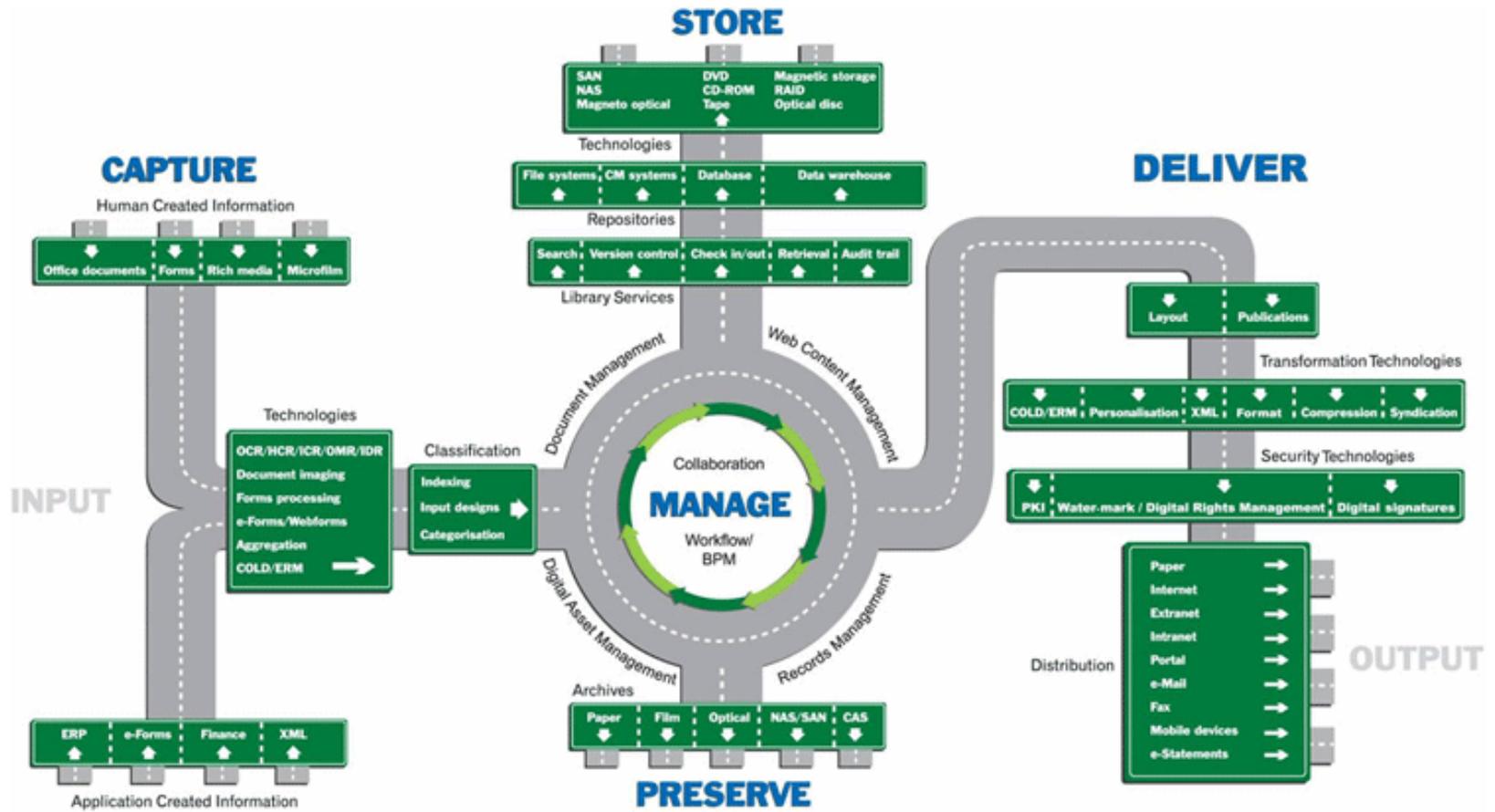


Figure 2-2 AIIM's Five ECM Areas

A simplified, layered version of the above figure is as follows:

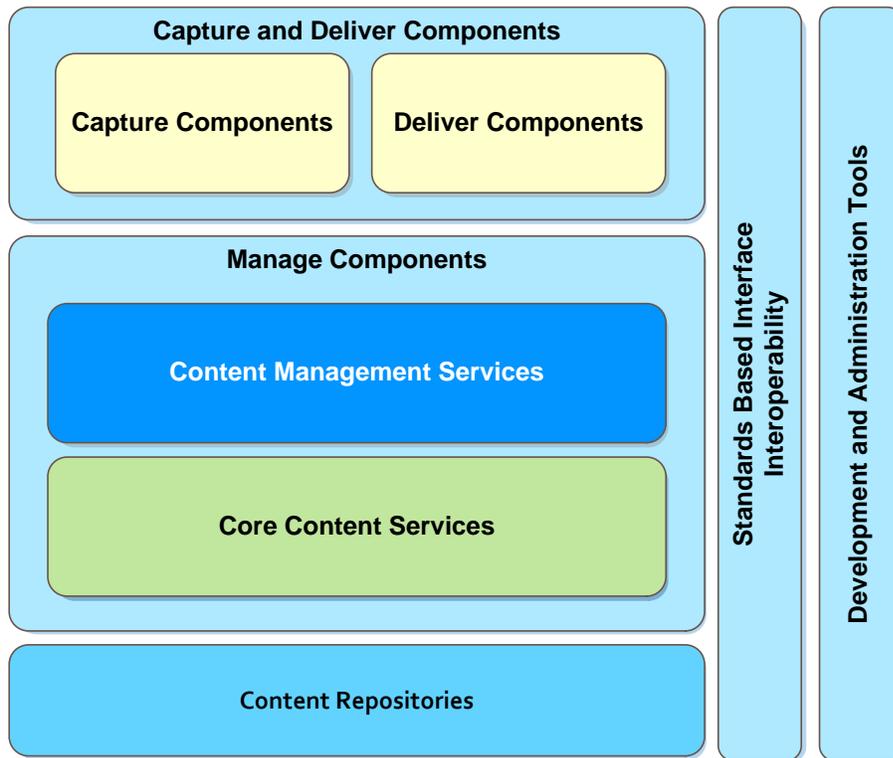


Figure 2-3 Overview of the Main Groups of Components

Each of the groups of components shown in the above diagram is described in the subsections that follow.

2.5.1 ECM Capture Components

The common characteristic of the “Capture” components is their basic function – acquiring digital content from digital and non-digital sources, regardless of whether the source is created by humans or whether it has been produced by a system or an application. The following figure provides an overview of ECM Capture components:

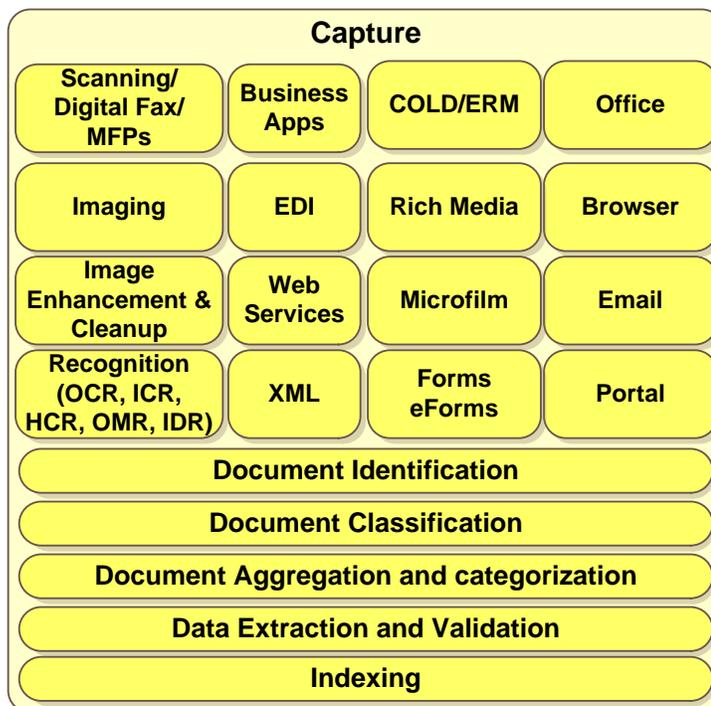


Figure 2-4 ECM Capture Components

ECM Capture components include the following:

- Document Scanning, Digital Faxes, Recognition Technologies (Optical Character Recognition (OCR), Handprint Character Recognition (HCR), Intelligent Character Recognition (ICR), Optical Mark Recognition (OMR), Barcode Recognition)
- Automatic or Semi-automatic Capture through EDI, XML, Business Applications, Forms Processing and other specialized applications
- Image Enhancement and Cleanup
- Document Identification, Data Extraction and Validation
- Computer Output to Laser Disc (COLD) for storage or archive
- Metadata creation and maintenance
- Indexing (metadata-based and content-based, where applicable).

2.5.2 ECM Deliver Components

The “Deliver” components of ECM present information/content in various forms and using various channels. The following figure shows the “Deliver” components in ECM:

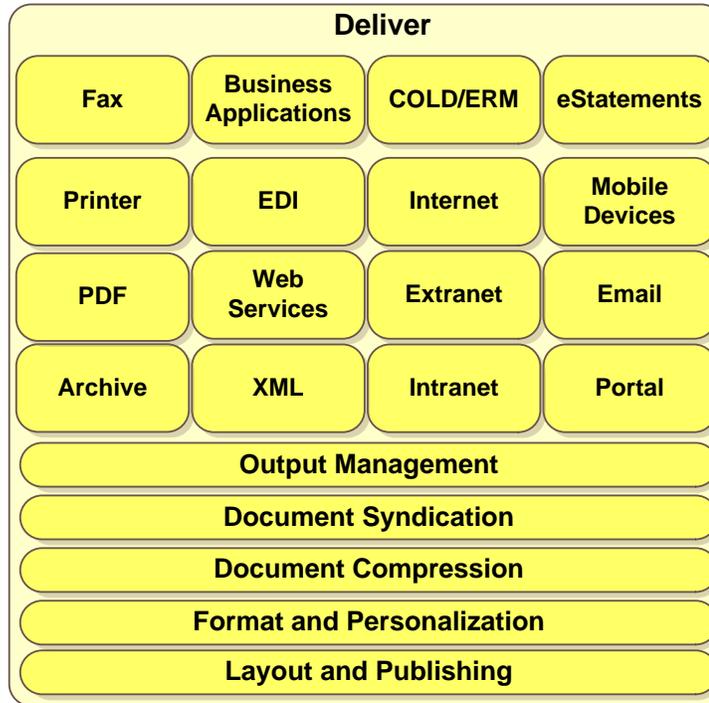


Figure 2-5 ECM Deliver Components

The Deliver components of ECM present information from the Manage, Store, and Preserve components:

- On-premise ECM, Software as a service (SaaS) ECM and Hybrid ECM
- Transformation Technologies
- COLD, Personalization, XML, PDF, XPS, Converters and Viewers, Compression, Syndication
- Security Technologies
- Electronic signatures, Public Key Infrastructure
- Digital rights management and watermarking
- Distribution and Output Management
- The Internet, Email, Fax, EDI, XML, Mobile, Data Media, Multimedia, Paper

2.5.3 ECM Manage Components

The following figure provides an overview of the “Manage” components in ECM:

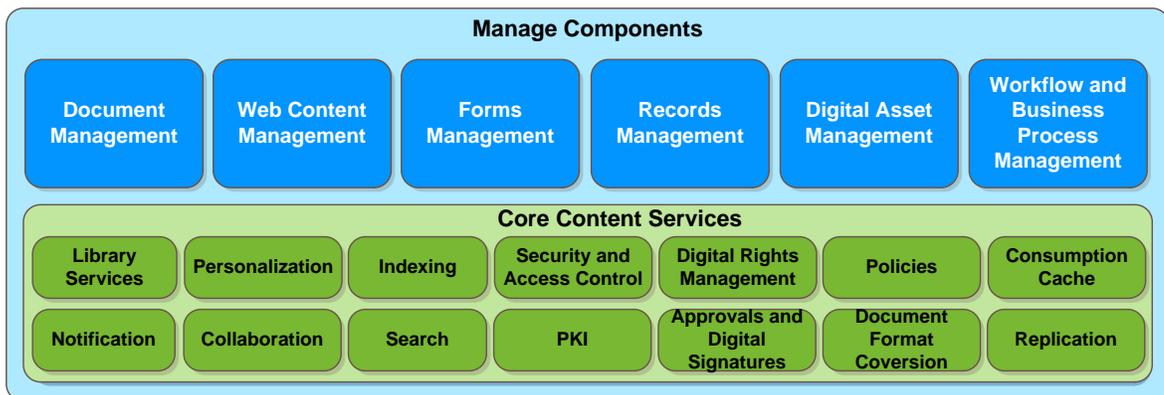


Figure 2-6 ECM Manage Components

ECM Manage components can be grouped into *Content Management Services* and *Core Content Services*. Content Management Services include the following components:

- Document Management (DM)
- Web Content Management (WCM)
- Forms Management
- Records Management
- Digital Asset Management
- Content-focused Business Process Management and Workflow

The Content Management Services are supported by a number of Core Content Services, including:

- Library services (check-in/ check-out etc.)
- Indexing and Search Services
- Digital Rights Management (DRM)
- Security and Access Control
- Document Format Conversion, etc.

2.5.4 ECM Store Components

The following figure shows the ECM Content Repositories used by the ECM “Store and Preserve” components:

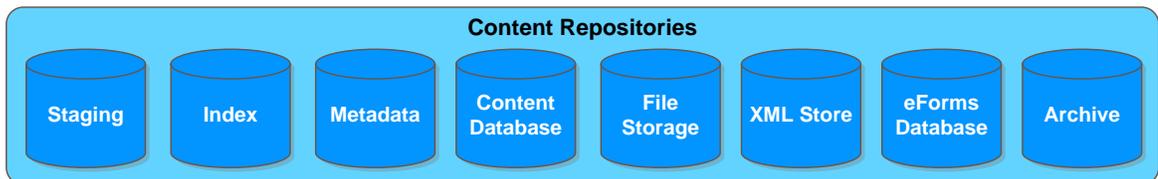


Figure 2-7 ECM Content Repositories (Store and Preserve Components)

ECM Store components include the following:

- Store components temporarily store information that isn't required, desired, or ready for long-term storage in Content Repositories:
 - File System(s)
 - Content Database (a COTS database or specialized storage used and managed by ECM system)
 - Other Databases and Data Warehouses
- Library Services, which are
 - Responsible for taking in and storing information from the Capture and Manage components
 - Responsible for working in concert with the Manage components' database to provide the necessary functions of search and retrieval
 - Responsible for version control and related check-out/check-in functions
 - Responsible for generating and persisting audit trails
- Storage Technologies, including SAN, NAS, optical storage and similar technologies.

2.5.5 ECM Preserve Components

ECM Preserve components include the following:

- Long-term safe storage and backup of static, unchanging information
- Accomplished by the Records Management features of an ECM system
- Preserve components contain special viewers, conversion and migration tools, and long term storage media
- Long-term preservation strategies, including migration between formats and emulation of formats.

2.5.6 ECM Development and Administration Components

The following figure shows administrative and development-related components of ECM:



Figure 2-8 ECM Development and Administration Components

ECM Development and Administration components include the following:

- Workflow/Business Process Design – it is highly desirable to automate ECM processes using workflow or business process engines. However, these solutions should be easily modifiable and extendable, including creation of new processes and modification of the existing ones.
- Taxonomy Management is responsible for design and maintenance of basic types and groups of content, which is a precondition for automating workflows and specifying business rules that apply to the content with respect to e.g. validation or distribution of content
- Forms and/or Template Design is required for managing automated form processing and, correspondingly, generation of digital output (not just text documents, but also e.g. web content)
- Content Authoring Tools are tools (such as editors for specific forms of digital content, and converters between digital formats) used typically by human users to create, transform, associate various components of the content in ECM.

3 ECM Reference Architecture Description

This section provides a focused description of ECM Reference Architecture (RA) using the Conceptual View which provides a summary of logical-level building blocks for ECM as presented in the Section 2 above.

3.1 ECM RA Conceptual View

The ECM RA Conceptual View brings together all major components of ECM solutions that have been already described in the section “Components of ECM Solutions”. The following figure represents a Conceptual View of ECM Reference Architecture.

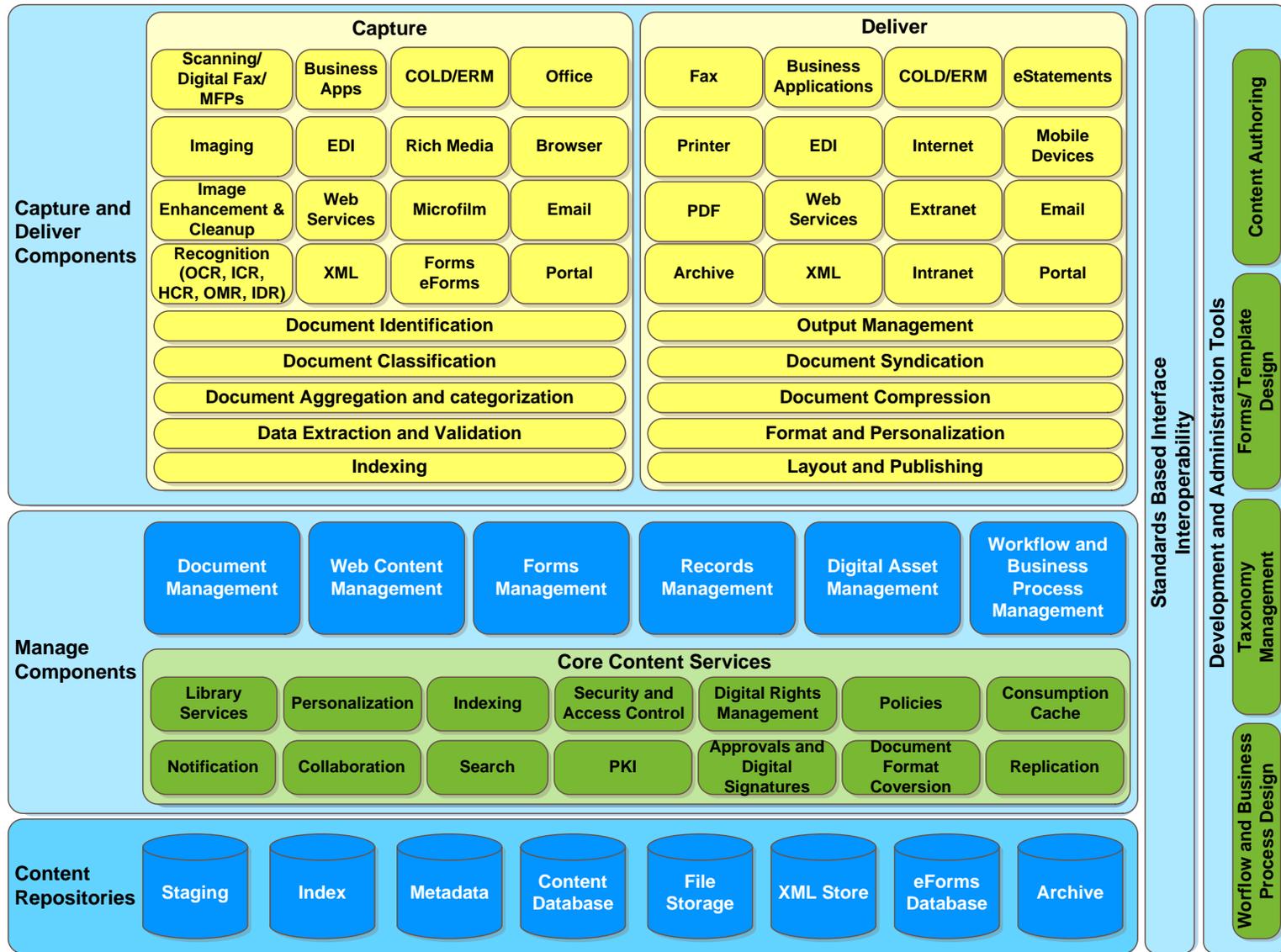


Figure 3-1 ECM Reference Architecture – Conceptual View

In the above figure, the following conventions are used:

- Groups of components laid out horizontally represent conceptual layers of key components in an ECM solution
- Groups of components laid out horizontally represent components that are auxiliary for an ECM solution (Development and Administration Tools) or which provide interfaces and interoperability for components in the horizontal layers.

The Conceptual View summarizes the presentation of ECM components in the section “Components of ECM Solutions” above. The following subsection presents a view of the ECM RA which involves services and application components.



4 Glossary

Application Architecture is an architecture that defines the major applications or service components needed to manage data and support business functions.

Document Management System is an application used to track and store electronic documents.

Enterprise Content Management the strategies, technologies, tools and methods used to capture, manage, store, preserve, and deliver content across an enterprise. [AIIS]

Content Management System is an application used to manage digital content, typically content needed to operate a web site or a portal.

Reference Architecture models the abstract architectural elements in the domain independent of the technologies, protocols, and products that are used to implement the domain.

Service Component is an actual application, program, or a subsystem providing implementation of a Service treated as a contract.

Web Content Management System – see Content Management.

5 References

5.1 Federal and State Documents

1. State of California, California State Information Technology Strategic Plan, November 2004
2. State of California, California Performance Review Report
3. Chief Information Officers Council, Federal Enterprise Architecture Framework, Version 1.1, September 1999
4. Chief Information Officers Council, A Practical Guide to Federal Enterprise Architecture, Version 1.0, February 2001
5. Federal Enterprise Architecture Program Management Office, The Business Reference Model, Version 2.0, June 2003
6. Federal Enterprise Architecture Program Management Office, The Service Component Reference Model, Version 1.0,
7. Federal Enterprise Architecture Program Management Office, The Technical Reference Model, Version 1.1, August 2003
8. Federal Enterprise Architecture Program Management Office, The Data Reference Model, Version 1.0, September 2004

5.2 Books and Papers

1. McNay, H.E., "Enterprise content management: an overview," Professional Communication Conference, 2002. IPCC 2002. Proceedings. IEEE International , vol., no., pp.396,402, 2002, doi: 10.1109/IPCC.2002.1049123
2. Päivärinta, T.; Munkvold, B.E., "Enterprise Content Management: An Integrated Perspective on Information Management," System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference
3. E Vasilescu, R., "Architectural model for a high performance content management system," Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for , vol., no., pp.1,6, 9-12 Nov. 2009
4. Chieu, T.C.; Liangzhao Zeng, "Service-Oriented Approach for Implementing an Extensible Content Management System," Congress on Services Part II, 2008. SERVICES-2. IEEE , vol., no., pp.96,103, 23-26 Sept. 2008, doi: 10.1109/SERVICES-2.2008.14
5. Kwok, K.H.S.; Chiu, D.K.W., "A Web services implementation framework for financial enterprise content management," System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on , vol., no., pp.10 pp.,, 5-8 Jan. 2004, doi: 10.1109/HICSS.2004.1265254
6. Schneider, S., "Requirements and Approaches for a Content Management Service," Automated solutions for Cross Media Content and Multi-channel Distribution, 2008. AXMEDIS '08. International Conference on , vol., no., pp.127,132, 17-19 Nov. 2008, doi: 10.1109/AXMEDIS.2008.31
7. O'Callaghan, Ramon and Smits, Martin, "A Strategy Development Process for Enterprise Content Management" (2005). *ECIS 2005 Proceedings*. Paper 148. <http://aisel.aisnet.org/ecis2005/148>
8. vom Brocke, Jan; Seidel, S.; Simons, A., "Bridging the Gap between Enterprise Content Management and Creativity: A Research Framework," *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* , vol., no., pp.1,10, 5-8 Jan. 2010, doi: 10.1109/HICSS.2010.86



5.3 Web Sites

1. California State Chief Information Officer, California Information Technology Council, Committees <http://www.cio.ca.gov/ITCouncil/Committees/Committees.html>
2. Gartner IT Glossary, <http://www.gartner.com/it-glossary>



6 Document History

Table 6-1 Document History

Release	Description	Date
Version 1.0 Draft	Initial creation	3/22/2013
Version 1.0 Second Draft	Revised based on internal comments	6/21/2013
Version 1.0 Final Draft	Addressed EAC comments	10/21/2013
Version 1.0 Final	Final version	01/02/2014