

Content Management Using the Rational Unified Process®

Michael McIntosh

Rational Software White Paper

TP 164, 9/01

Table of Contents

Introduction	1
Content Management Overview	1
The Challenge of Unstructured Content	2
The Scope of a Content Management Project	3
The RUP and Content Management	3
The Content Management Discipline	4
Content Development Defined	4
Content Production Defined	5
Content Delivery Defined	8
Mapping Content Management to the RUP's Phases	9
Inception Phase	10
Elaboration Phase	10
Construction Phase	12
Transition Phase	13
Summary	13
Content Management Case Study	13
About Acme	13
Content Development Applied: Establishing the Framework	14
Content Production Applied: Pulling Together the Content	15
Content Delivery Applied: Serving the Content	16
Getting Real	16
Content Management Glossary	17
Related Resources	19
About the Author	19

Introduction

Content management refers to a variety of tools and methods that are used together to collect, process, and deliver content of diverse types. The scope of content management is broad, and its challenges are many. This white paper presents an overview of content management, particularly as it relates to delivering content on the Web, and examines why the Rational Unified Process (RUP) is the best framework for content management projects.

This paper installment takes a high-level view, identifying the challenges of content management and the many activities it entails, and examining where those activities fit into the RUP. (Some previous familiarity and experience with the RUP is assumed.)

At the end of this white paper is a simplified case study that provides a context for examples (through an imaginary company named Acme) and a glossary of terms related to content management. You may prefer to wait until you've read the entire article before consulting the case study.

Content Management Overview

Organizations increasingly turn to the Web as a means of distributing information, communicating with their customers, and seeking to differentiate themselves from their competition. However, simply uploading content to a Web site doesn't ensure that it will reach the proper audience or meet their information needs. The content must constantly be updated if it is to have ongoing value. It needs to be processed to ensure that it's ready for delivery to customers, and the delivery has to be designed such that users can locate what they need in a minimum number of steps. The content provided to each user should be matched to the needs of that user's business role; for example, technical documents to technical users.

Collecting and processing content intended for a Web site is a more complex and time-sensitive process than many legacy document management systems are designed to handle. A content management system may require integration with existing document management systems or may require its own separate workflow. There are many factors to consider.

So, while the *concept* of content management is not new, the *implementation* of a content management project is still a challenge, one for which there are more products than processes. Figure 1 shows an overview of content management from the process perspective.

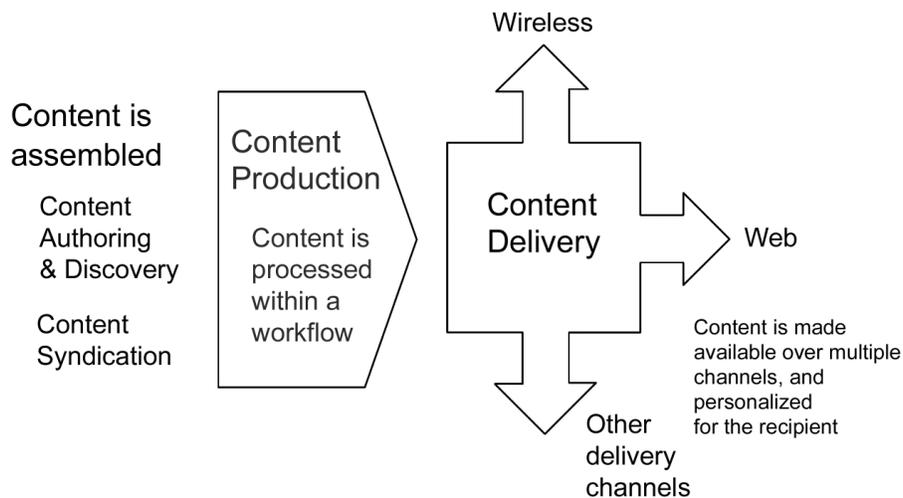


Figure 1: Content management process overview

Later in this paper we'll look more closely at the content management aspects noted in Figure 1, but two key high-level points are worth mentioning here: the challenge presented by unstructured content, and the broad scope of a typical content management project.

The Challenge of Unstructured Content

The content management process starts with aggregating content from a variety of sources. Typically, some content exists in a **structured** form such as a database, but most of what needs to be delivered is **unstructured** content, such as articles, white papers, and marketing literature. This is especially true as a result of the Web’s emergence as a delivery vehicle. (For more about the evolution of content management on the Web, see the section on that topic in the article [“Content Categorization: An Orientation to Vignette and Content Management.”](#))

Organizing and maintaining collections of unstructured content that continually change is one of the chief challenges of content management. A cursory examination of many content management products reveals that managing the workflow of activities related to unstructured content is an essential chore. The tasks within that workflow (some of which require human input) include:

- **Conditioning content** — This important early step involves preparing the content for entry into the content management system, including assigning keywords to facilitate categorization, searches, and personalization.
- **“Versioning” documents that have multiple contributors** — A document management issue related to unstructured content is that a content item may have multiple authors, or the editing process may require more than one person to be able to change a content item. So that additions and changes can be made without losing the work of any of the contributors, documents must be checked out and in. In addition, it’s desirable to have a simple procedure for rolling a content item back to an earlier version.
- **Integrating code and content** — Unstructured content items may be associated with particular templates for delivery (as may structured content; see Figure 2), or may need to contain or reference code (such as style sheets) that affects the delivery or behavior of the content.

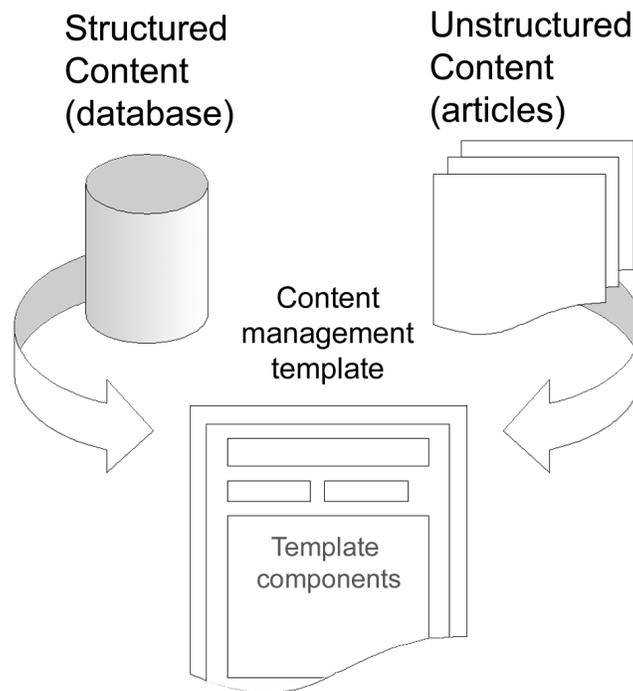


Figure 2: Templates for dynamic content delivery

The Scope of a Content Management Project

Content management, then, is much more than simply creating content and specifying that it's to be published to the Web. The many activities a content management project entails include:

- assembling diverse types of content from multiple sources
- using **templates** to make the content available through a variety of channels (including wireless devices, e-mail, and fax as well as the Web), at the same time avoiding the need to create or manage redundant copies of the information in multiple formats
- providing content in multiple human languages
- using **workflow** (a predefined sequence of events) to administer the content throughout its lifecycle
- creating categories for cataloging the content
- tagging content appropriately so that users can easily find what they need, either by having it provided to them based on who they are or by having it recommended because of other content they've accessed
- identifying relationships and linking one piece of content to another
- automatically notifying users when content changes or when there are special business opportunities related to new or updated content
- integrating with other parts of an information system, such as knowledge management and document management

To summarize, content management is a broad undertaking that involves generating content of various types, delivering that content using applications that can meet (or even anticipate) the needs of various constituencies through a variety of delivery channels, and using workflow to administer the content.

The RUP and Content Management

Content management involves far more than buying or building a single application that manages content production workflow (which we'll call the **workflow application** in this paper). A content management solution is made up of database, workflow, content delivery, and Web applications, developed as a group of related software projects. This makes it a perfect candidate for an established software development framework such as the RUP. Reasons for using the RUP as the framework for content management projects include the following:

- **Content management projects are software projects.** Content management projects have a lot in common with other software development projects, and therefore much of what they need to address falls within the RUP's existing disciplines. Using the RUP precludes the need to start from scratch. There's also a wealth of related guidelines within existing RUP roadmaps (such as the e-business Solution roadmap).
- **Content management projects are complex.** A content management project typically involves a broad range of roles—business, technical, and creative—and it's important that the framework for such a project be able to encompass the diverse needs of those constituencies. The RUP helps strengthen communication among the various roles by identifying artifacts for them to share as they perform their activities. In addition to clarifying the vision and requirements of the project, these artifacts help identify risks that can then be addressed. The RUP is also well suited to handle the fact that parts of the content management solution (such as categorization and user interface) will most likely evolve over the life of the project, leading to ongoing iterations.
- **The RUP is a proven method with an installed base of over 100,000.** The need for methods that have been demonstrated to work continues to be rediscovered. Heroic effort does not scale, nor is it always successful. A configurable process that provides a common point of reference for the diverse roles and disciplines has been shown to offer the greatest success. The RUP has demonstrated success as such a process.

To better understand the activities of any complex process, it's useful to group the activities. In RUP terms, content management is a **discipline**, and groups of related activities within a discipline are called **workflow details**. Consistent with the process overview shown in [Figure 1](#), the workflow details of the content management discipline are:

- **Content development**, which addresses the standards, practices, and structure that govern the development of the content management solution, as well as the tools for dynamic content delivery (such as templates).
- **Content production**, which addresses the workflow of tasks that manage the content through its lifecycle.
- **Content delivery**, which addresses getting personalized content to users over multiple channels.

In the remainder of this paper, we'll look more closely at what activities and concepts (and related artifacts) make up these workflow details, and how they map to the RUP's phases.

The Content Management Discipline

This section elaborates on the function of each of the content management discipline's workflow details (content development, content production, and content delivery), providing general guidelines, defining related terms, and listing basic activities along with corresponding types of artifacts. It presents typical challenges that might crop up in each area, and refers to the Acme case study as an example; for an introduction to Acme and its content management-related problems see the Content Management Case Study at the end of this paper.

Content Development Defined

Content development consists of establishing the framework of tools, languages, file formats, and guidelines that will be involved in creating or editing the content, as well as the definition of the content delivery means (such as templates). This workflow detail, which determines much of the architecture of the content management solution, can be divided into the areas listed in Table 1.

Table 1: Content development workflow detail

	Activities and concepts	Artifacts produced
Content authoring	Writing text	Articles, white papers, marketing literature
	Content discovery	A list of possible content collections where existing content can be found
	Developing multimedia content	Sound files, multimedia files
	Developing content delivery templates	Code for delivery of dynamic content over the Web or other channels (such as wireless) Tested templates that reliably deliver dynamic content to all supported channels
	Addressing legal issues	Identified risks of publishing intellectual property to the Web Legal agreements stating site policies for use of content
	Internationalizing versions of the site for other languages and locales	Translated content and message catalogs
Defining development tools and guidelines	Selecting development platforms	Standards for site development, including supported code development platforms, databases, and markup languages An environment for testing and development using any supported platform
	Internationalizing development	Support for international character sets within the Software Requirements Specification or Supplementary Specifications Multilingual versions of software development tools
	Selecting the development process	Adoption of the process and the tools for implementing it

	Activities and concepts	Artifacts produced
Defining content structure	Defining an XML strategy	XML DTDs and related schemas
	Defining database requirements	A database schema and database application requirements
	Defining data conversion requirements	A data conversion strategy

The following subsections elaborate on content development, and examples of applying it to Acme’s scenario are provided in the corresponding sections of the case study.

Content Authoring

In addition to writing text, content authoring activities include recording audio or generating other multimedia content types, creating templates for dynamic content delivery, and identifying or managing intellectual property issues such as copyrights or patents. In an international setting, translation of text to other languages falls in this area, as does the localization of data and time formats, sort order, and numeric representation.

Defining Development Tools and Guidelines

The area of development tools includes defining the environments within which various content types can be created, including markup languages, scripting or programming languages, and development platforms. It also addresses internationalization concerns, such as support for international character sets at the application, programming language, and database levels. Decisions regarding development tools are very significant to the application architecture.

Development guidelines describe the use of methods or modeling to define processes for development. This can include, of course, the use of the Unified Modeling Language (UML) and the RUP.

Defining Content Structure

The definition of data structures falls here, including XML document type definitions (DTDs) or XML standards, as well as database schemas and data conversions. The content delivery solution often depends on this architecture, making content structure design one of the highest priorities for content management.

Content Production Defined

Content production refers to the processing of the content in preparation for publishing it. Table 2 lists the areas covered by this workflow detail.

Table 2: Content production workflow detail

	Related activities and concepts	Artifacts produced
Using workflow to process content	Content routing	<p>Analysis of how content is currently handled, including creation, delivery, and archiving</p> <p>Specification within the Vision document of the content routing in the workflow application</p> <p>Rules for assigning content items to those responsible for working on them, as well as features within the workflow application to manually assign responsibility</p> <p>Appropriate features within the workflow application to meet content routing requirements</p> <p>Use cases for all routing options</p>
	Task management	<p>A list of discrete tasks within the content production workflow, describing what they entail and what constitutes completion of each</p> <p>Use cases describing the various content management tasks</p>
	State management	<p>A list of states through which a content item might pass, criteria for the states, and tasks related to state changes</p>
	Role-based security planning	<p>A security plan listing the roles within the security structure and, for each role, the rights and constraints related to both content items and system administration</p>
Conditioning content	Categorizing content	<p>A categorization schema, with associated category keywords, and features within the workflow application to associate keywords with content items</p>
	Describing content	<p>Features within the workflow application to capture metadata for each content item, describing the subject, format, and other content conditioning information</p>
Content administration	System administration of the workflow	<p>System administration features for managing the creation and characteristics of users and groups, as well as the configuration of the workflow application and the rules governing it</p>
	Administration of content items	<p>Features within the workflow application to manage the state of content and provide for standard content management operations</p>
Aggregating content	Gathering content from various sources	<p>Descriptions within the Software Development Plan of the content sources and the expected source format, transfer method, and delivery format for each</p> <p>Architectural plans to support communication among the various systems</p> <p>Consideration within the security plan for the receipt of content from outside locations</p>
	Content rationalization	<p>Strategies within the Software Development Plan for collecting various types of content and presenting them as a cohesive whole</p> <p>Supporting templates for situations requiring content rationalization</p>
	Content conversion	<p>Template components for translating the data format, such as converting database query results to HTML</p>

	Related activities and concepts	Artifacts produced
Syndicating content	Subscription management	A schedule for when updates are to be available, and considerations within the security plan for who can receive updates
	Digital goods delivery	Strategies within the Software Development Plan for moving different types of content from a source to a destination

The following subsections elaborate on content production and examples are provided in the corresponding sections of the case study.

Using Workflow to Process Content

The content production workflow focuses on routing content items to the people responsible for working on them. The routing should be rule-based (where the rules are based on use cases), with the option of manually assigning responsibility when required. The workflow is also concerned with providing the information and tools required to support each step of the content management project’s lifecycle. It’s divided into discrete tasks, which are then assigned using rules or administrative procedures.

It’s advisable to implement role-based security measures to limit who has access to various content items and the changes they can make to those items. As in any security infrastructure, the use of groups facilitates granting and denying access to both the content and the administration of the workflow.

Conditioning Content

Conditioning content means adding information to or about a content item to facilitate searching and personalization (matching content to the needs of specific groups of users). It uses categories that are established for these purposes — specifically, using **category keywords** (terms used for matching search criteria about the content) and **metadata** (information that describes the content). Attaching metadata to content is also called **meta-tagging**, referring to the tags used to include information about a document in the head of an HTML document.

To describe the categories, there must be an accepted **taxonomy** (a division into ordered groups) and corresponding descriptive nomenclature. The taxonomy is used to create a category hierarchy, or tree, that can serve as a basis for a navigation scheme. It’s not uncommon to have more than one way of describing the same thing, even within the same organization. Content associated with more than category should be able to be located by navigating from different directions.

Ideally, there’s a means of rating the suitability or importance of content to its categories. An article describing how to install a network card, for example, might be associated with both installing hardware and installing drivers; a search from either direction would produce a hit. If the article discusses hardware configuration in depth and barely touches on drivers, a rating system could weight the categorization accordingly.

To enable content categorization to be automated, documents can carry metadata within them. Examples of document formats that support this are HTML and XML. In addition, a content management system can provide categorization for document types that don’t have a way of carrying their own metadata.

Content Administration

Content administration encompasses the administration of both the workflow and the content items themselves. The needs associated with content administration include:

- submitting content items for publication
- locking content items to prevent access by more than one contributor at a time
- versioning content (assigning version numbers) as a method of tracking changes
- rejecting or approving content items for publication and sending approved content to the content delivery system
- archiving content (maintaining copies of content items)
- deleting content (or marking it for expiration) or replacing it with content that’s more current

One aspect of the content that’s important to manage is its state. Examples of states include “incomplete,” “requires additional editing,” “ready for publication,” and “requires approval of the Legal department.” Once content has been processed by a task within the workflow, the state of that content might automatically change.

Aggregating Content

Organizations keep content in many places and in many forms, and often receive or subscribe to content from external sources as well; for example, from partner companies. **Aggregation** is the assembly of content from a variety of sources, whether internal or external. Since it’s unlikely that assembled content will all be in the same format, the following fall within aggregation:

- **Content rationalization**, which means collecting various types of content from multiple sources and presenting them as a cohesive whole; for example, bringing together a database of parts and prices, and merging it on demand with Web page descriptions and pictures from another source, using a template.
- **Content conversion**, translating the format of the data; for example, converting information resulting from a database query to HTML for display in a browser.

Syndicating Content

Syndication means providing the most current content to browsers that have been configured to receive scheduled updates for the site, or providing content to other Web sites. It’s increasingly common for companies to share information about their products and services with other companies for inclusion on their sites. In addition, client subscription technologies allow users to have their browsers automatically download the latest version of a page from a site.

Technologies that fall within syndication include:

- **Subscription management**, which involves the scheduling of updates and the security considerations of syndication (who can receive the content and how frequently updates should be performed).
- **Digital goods delivery**, which addresses the fact that not all content is simple HTML, and that the Development Plan needs to include multiple strategies for moving content from a source to a destination.

Content Delivery Defined

Content delivery describes those parts of the content management solution that have the greatest public visibility: the Web site and other delivery channels. As shown in Table 3, it refers to the methods by which content is made available to users, including personalization (matching content to information that has been discovered or recorded about the user) as well as actual delivery. Since the performance of the site is an aspect of delivery that can be affected when content is provided dynamically, multiple-server solutions and caching are also included in the area of content delivery.

Table 3: Content delivery workflow detail

	Related activities and concepts	Artifacts produced
Personalizing content for users	Creating a personal experience for users	Description within the Vision document of personalization goals for the site Plans for gathering information about users explicitly or implicitly Defined roles for users, which play a part in determining the most appropriate content for delivery A categorization schema applied to the content items so that the proper content can be submitted to a particular role
Delivering content over multiple channels	Making content available to various delivery media	Description within the Vision document of goals for multi-channel delivery Templates that format the content for the medium over which the user requests it

	Related activities and concepts	Artifacts produced
Performance and fault tolerance	Multiple-server solutions	Description within the Vision document of performance and fault tolerance goals A check of compatibility with platforms documented in the Software Requirements Specification or Supplementary Specifications Methods for addressing personalization issues with multiple Web servers Methods for synchronizing dynamic content across multiple servers
	Caching	Description within the Vision document of performance goals A caching strategy, including methods for identifying items to be cached to the content delivery server

The content delivery requirements dictate application needs, which largely determine the standards established during content development. In fact, decisions made about the final delivery of the content drive most aspects of the content management project.

An elaboration of content delivery follows, and the corresponding sections of the case study give examples.

Personalizing Content for Users

Personalization means the creation of a personal experience for visitors to a site based on information gathered from them, whether by invitation and submission (explicit) or based on observation of their behavior (implicit). Again, there’s a relationship between content delivery and other content management aspects, since the conditioning of the content (metadata and categorization) addressed within content production is based on the personalization design.

Delivering Content over Multiple Channels

A **channel** is a method of content delivery, the default method being the Web. Other methods might include e-mail, pager, cell phone, and personal digital assistant (PDA). Wireless Application Protocol (WAP) is an example of technology that’s growing rapidly in some parts of the world. The increased availability of broadband access implies a growth in technologies, such as streaming media, that depend on high bandwidth for effective use.

Performance and Fault Tolerance

Given the complexity of content delivery applications, both performance and high availability are key concerns. **Fault tolerance** means the ability of a system to continue to operate despite the failure of some part of the system. The use of redundant servers provides fault tolerance, as well as performance gains when all servers are operating optimally. There are issues with server redundancy in content delivery systems that provide dynamic content, especially where personalization is concerned, and these issues need to be addressed in application and site design.

Another method of improving performance is **caching**; that is, keeping recently or frequently used pages in memory or on disk. The dynamic page generation commonly employed in content delivery increases overhead, and caching is a strategy for making the static parts of a page template readily available, reducing the time it takes to build and deliver a page.

Mapping Content Management to the RUP’s Phases

In this section, we’ll examine the workflow details of content management in the context of each of the four phases of the RUP: inception, elaboration, construction, and transition. The activities and concepts, and the resulting artifacts, listed previously in Tables 1 through 3, for each of the workflow details are mapped to the RUP’s phases—a one-to-one mapping except for a few cases (marked in *italic*) where work on a particular artifact is split between two phases. As a result, the flow of a content management project emerges.

Where applicable, evaluation criteria from the RUP are presented here, but these don’t include all of the criteria for assessing the phases of a project using the RUP. The emphasis is on those criteria that are most applicable to a content management project.

Inception Phase

The goal of the inception phase is to achieve concurrence on the lifecycle objectives for the project. Table 4 shows which aspects of a content management project come into play during this phase. Primarily, it's important at this time to define and record the vision of the project in the Vision document. For content management, the vision typically focuses on content delivery, with some interest in content production driven by those responsible for the content production workflow. Further additions to the Vision document are made during the elaboration phase.

Table 4: Lifecycle objective milestone

Activities and concepts	Artifacts produced
Content discovery	A list of possible content collections where existing content can be found
Addressing legal issues	Identified risks of publishing intellectual property to the Web
Selecting development platforms	Standards for site development, including supported code development platforms, databases, and markup languages (<i>exploration</i>)
Creating a personal experience for users	Description within the Vision document of personalization goals for the site Plans for gathering information about users explicitly or implicitly Defined roles for users, which play a part in determining the most appropriate content for delivery
Content routing	Analysis of how content is currently handled, including creation, delivery, and archiving Specification within the Vision document of the content routing in the workflow application
Task management	A list of discrete tasks within the proposed content production workflow, describing what they entail and what constitutes completion of each Use cases describing the various content management tasks (<i>start</i>)
State management	A list of states through which a content item might pass, criteria for those states, and tasks related to state changes
Making content available to various delivery media	Description within the Vision document of goals for multi-channel delivery

Related evaluation criteria as applied to a content management project are:

- **Concurrence on scope definition** — Content development involves establishing the project's software scope and boundary conditions, including an operational vision and acceptance criteria: what's intended to be in the product and what's not. In addition, concurrence on the scope of content delivery is critical (since content delivery has the highest visibility within content management projects), as is agreement on what it will take to accomplish the desired results.
- **Requirements understanding** — Content delivery and content production will depend on defining the critical use cases of those systems. The primary scenarios of operation will determine the major design trade-offs to be made. Exhibiting (and maybe demonstrating) at least one candidate architecture against some of the primary scenarios can facilitate architecture decisions. In addition, investigating the needs of the content production system may include preparing the supporting environment for the project.

Elaboration Phase

The goal of the elaboration phase is to baseline the architecture of the system, providing a stable basis for the design and implementation efforts during the construction phase (see Table 5). It should be demonstrated that the baseline architecture will support the requirements of the system at a reasonable cost and in a reasonable amount of time.

Table 5: Lifecycle architecture milestone

Activities and concepts	Artifacts produced
Selecting development platforms	Standards for site development, including supported code development platforms, databases, and markup languages (<i>selection</i>) An environment for testing and development using any supported platform
Selecting the development process	Adoption of the process and the tools for implementing it
Multiple-server solutions	Description within the Vision document of performance and fault tolerance goals A check of compatibility with platforms documented in the Software Requirements Specification or Supplementary Specifications
Caching	Description within the Vision document of performance goals
Internationalizing development	Support for international character sets within the Software Requirements Specification or Supplementary Specifications Multilingual versions of software development tools
Creating a personal experience for users	A categorization schema applied to the content items so that the proper content can be submitted to a particular role
Gathering content from various sources	Descriptions within the Software Development Plan of the content sources and the expected source format, transfer method, and delivery format for each Architectural plans to support communication among the various systems Consideration within the security plan for the receipt of content from outside locations
Categorizing content	A categorization schema, with associated category keywords, and features within the workflow application to associate keywords with content items
Content rationalization	Strategies within the Software Development Plan for collecting various types of content and presenting them as a cohesive whole
Subscription management	A schedule for when updates are to be available, and considerations within the security plan for who can receive updates
Defining an XML strategy	XML DTDs and related schemas
Defining database requirements	A database schema and database application requirements
Defining data conversion requirements	A data conversion strategy
Content routing	Rules for assigning content items to those responsible for working on them, as well as features within the workflow application to manually assign responsibility Use cases for all routing options
Task management	Use cases describing the various content management tasks (<i>complete</i>)

Activities and concepts	Artifacts produced
Role-based security planning	A security plan listing the roles within the security structure and, for each role, the rights and constraints related to both content items and system administration
Digital goods delivery	Strategies within the Software Development Plan for moving different types of content from a source to a destination

Critical during the elaboration phase is the need to set up the supporting environment for the project, including creating a Development Case and Guidelines as well as setting up tools. The initial setup of a workflow application, if one has been purchased or is being given a trial, occurs during this phase. To allow prototyping of templates, at least one development Web server needs to be put in place. If the automation of the content delivery system is incomplete, content that has passed through the test workflow can be manually copied to the server for testing.

The criteria worth noting for this phase are:

Stability of the architecture — This is evaluated through one or more architectural prototypes; for example, templates for dynamic content delivery.

Addressing the major risks — The architecture evolves out of a consideration of the most significant requirements (those that have a big impact on the architecture of the system) and an assessment of risk. Because of its complexity, content delivery will present some of the greatest risks.

Construction Phase

The focus of the construction phase is to clarify the remaining requirements and complete the development of the system based on the baseline architecture (see Table 6).

Table 6: Initial operational capability milestone

Activities and concepts	Artifacts produced
Writing text	Articles, white papers, marketing literature
Developing multimedia content	Sound files, multimedia files
Developing content delivery templates	Code for delivery of dynamic content over the Web or other channels (such as wireless)
Content routing	Appropriate features within the workflow application to meet content routing requirements
System administration of the workflow	System administration features for managing the creation and characteristics of users and groups, as well as the configuration of the workflow application and the rules governing it
Administration of content items	Features within the workflow application to manage the state of content and provide for standard content management operations
Describing content	Features within the workflow application to capture metadata for each content item, describing the subject, format, and other content conditioning information
Content rationalization	Supporting templates for situations requiring content rationalization
Content conversion	Template components for translating the data format, such as converting database query results to HTML
Multiple-server solutions	Methods for addressing personalization issues with multiple Web servers Methods for synchronizing dynamic content across multiple servers
Making content available to various delivery media	Templates that format the content for the medium over which the user requests it
Internationalizing versions of the site for other languages and locales	Translated content and message catalogs

To accelerate the development activities that take place during the construction phase, it's desirable to achieve some degree of parallelism in the work of different development teams. The proper use of content management templates, and in particular the reuse of components within templates, can help achieve this goal.

The primary criterion for this phase is whether the product release is stable and mature enough for deployment. The flow of content from content production to content delivery needs to be stable by the end of this phase. The content delivery application must be able to assemble content dynamically and deliver it to the channels that pass it on to users. Before this project can be handed off to the transition team in the next phase, the progress of the organizational transformation must be far enough along that the content production workflow has become part of the routine of the team members involved.

Transition Phase

The focus of the transition phase is to ensure that the site is available for end users (see Table 7). This phase can span several iterations, and it includes testing the site and making minor adjustments based on user feedback. At this point in the project lifecycle, user feedback should focus mainly on fine-tuning; for example, on the effectiveness of personalization and ease of use. All major structural issues should have been worked out much earlier in the lifecycle.

Table 7: Product release milestone

Activities and concepts	Artifacts produced
Developing content delivery templates	Tested templates that reliably deliver dynamic content to all supported channels
Addressing legal issues	Legal agreements stating site policies for use of content
Caching	A caching strategy, including methods for identifying items to be cached to the content delivery server

When the RUP is applied to content management, the transition phase tasks are essentially the same as they would be for any other RUP-managed software project. The primary evaluation criterion for this phase is whether the users are satisfied.

Summary

Content management covers a lot of territory, from defining the need to provide some type of content, to creating and delivering the content (including creating, configuring, and administering the delivery mechanism and translating the content for different delivery channels). This article has presented the three workflow details of content management — content development, content production, and content delivery — along with related definitions, objectives, and concerns, and examples in the form of a case study.

We've also looked at some reasons why the best framework for content management projects is the Rational Unified Process—a proven method and multidisciplinary approach to identifying and assigning activities and responsibilities as a project is iteratively delivered—and at how content management activities and concepts map to the RUP's phases.

Content Management Case Study

This case study is based on an imaginary company (named Acme) with typical content management problems. It presents a simplified example that illustrates the broad scope of a content management project and its related challenges. An introduction to Acme and the challenges it faces is followed by sections describing how this company applies the three workflow details—content development; content production, and content delivery — that make up the content management discipline.

About Acme

Acme produces a line of modems and network cards. Like many companies, they've always strived to make resources such as software drivers, patches, problem resolutions, and information documents available as soon as possible. The content comes from various groups within Acme—in fact, they've discovered that they've been duplicating efforts internally—and multiple Web sites have evolved to deliver the content to customers and distributors.

Although Acme has good products and a dedicated work force, the current public perception of the company is that they're having customer satisfaction problems. Users complain that they have difficulty finding what they need because of the multiplicity of servers. Acme seeks to improve the quality of their customer service by improving the delivery of information

and by better matching the information to the needs of the user. For the project at hand, they want to integrate the customer support site with their main Web site.

Customer Support Challenges

Acme's Customer Support department has an "in the trenches" attitude and a strong desire to meet customer needs. They're concerned about how long it takes support and service staff to get up to speed on Acme's extensive product line. To address this, Customer Support has built a searchable repository of issues and resolutions to help solve problems without duplicating effort.

The customer support server requires that users log in to register an issue, giving it a rating for level of urgency. In addition, users can search the solution database for information that has been approved for customer access. The user interface is a bit clunky, and the site is far from glamorous, but it does provide online access to support "tickets", and access to the solution database for more experienced users.

Acme plans to move the customer support application to the main site. Users will still have to log in, but placing all supporting documentation on one site will make it easier to manage. In addition, Customer Support would like the results of database searches to dynamically list other information documents that apply to the problem at hand, including those that have been recently added or updated. They'd also like to have a method for cataloging content that will facilitate cross-referencing between documents. Links to other documents would be built automatically, based on information entered whenever new content is added to the site.

Issues with the Existing Main Web Site

The Marketing department manages the content of Acme's main Web site, the primary function of which is to communicate the features and benefits of Acme's products. When service or technical content is added to this site, there are problems with organizing and presenting the information in a way that meets the needs of the diverse user communities. Many users refuse to continue to use the site because they're unable to find answers to their questions. Technical users complain that a search is as likely to bring up a product overview as an in-depth article or a solution to a problem, and non-technical users are similarly dissatisfied with the search results.

Acme would like to have everyone use a single Web site to access any type of information on any of their products and they want to match the content to the needs of the user. They also want to let each user know when there's a change to any part of the site that affects that person's areas of interest; for instance, to let customers know when a new driver is released for a product they've purchased.

Content Development Applied: Establishing the Framework

Content development is mainly concerned with identifying and setting standards, choosing the platforms that will be the basis for the content management solution, designing the content structures, and setting guidelines for development.

Getting Started

Realizing that they're going to be integrating content from different systems, Acme assigns a group to learn what platforms will need to work together, and to discover the best tools for integration and development.

Acme also reviews their procedures for publishing intellectual property to the Web and, with that, the nondisclosure agreements that they have with customers and partners.

In addition, they proceed with a business analysis of existing content development practices. The information gathered will affect content production as well.

Integrating Applications

The existing customer support application, a Web front-end to a database application, is to be integrated with the main Web site. Most aspects of this application integration project are the same as they would be in any Web integration project. Content structure is a key concern in such a project. This includes recording or designing the schema of any databases that will be used, managing any data conversions that need to be performed, and discovering or designing any XML document type definitions that will be part of the solution.

Creating and Adapting Content for the New Site

The assumption at Acme is that the Technical Publications group will be responsible for generating the new information pieces, white papers, and other new content, as well as making existing content available for electronic publication.

The impact of a new system on an organization requires the same careful planning that goes into the software itself. For example, Acme will need to determine whether Technical Publications is sufficiently staffed for this additional work. The group's current practices (which are based on less frequent revisions and on a print environment) will need to be examined to determine whether they'll meet the demands of the faster-paced Web environment. The time, training, and effort required to implement the content management system should be part of the plan from the start.

Early on, Acme needs to identify the environments within which content or applications will be created or where they already exist. Legacy systems are important to consider; for instance, document management tools Acme is already using might be integrated into the workflow of the site. The decisions made in this area will significantly affect the application architecture of the solution.

Seeking to both use existing content and plan for new content raises questions about the construction of the content:

- What content would be best addressed by creating templates that dynamically present the information in an appropriate format? Information being pulled from a database is an example of content that's best formatted by a template.
- What parts of the Web pages would remain the same, and which would change? Determining this will help with the planning for content delivery caching.

The relationship between content development and content delivery becomes very clear when the above points are considered. There must be a completed content delivery design so that content templates can be developed properly.

In addition to the standard Web interface, other content delivery channels (such as wireless) may be used. Using different technologies for delivery may mean that the content is best generated in, or translated to, a neutral format that can be content management interpreted for multiple channels. This is one reason for the popularity of XML in content management solutions.

Content Production Applied: Pulling Together the Content

Content production involves processing the content in preparation for publishing it.

Conditioning Content

To provide for better-qualified searches, all unstructured content (such as articles and white papers) will have category keywords assigned to them. Acme begins by clearly separating marketing and sales materials from technical documents, and categorizing the technical documents within their respective disciplines.

Acme will need to develop procedures for defining which categories are most appropriate for the content being classified. These procedures will need to be integrated into the categorization workflow.

The taxonomy, especially for the technical side, will grow as the categorization of the content takes place. If a new category is added after content has already been classified, a plan will need to be formulated to locate conditioned content that should be included in that category.

One of the chief benefits of conditioning is that a search can return the most appropriate content first. To accomplish this, Acme will develop a categorization design that allows for rating the suitability of the content in relation to categories.

As part of site management, reports on content use will be gathered. If the reports show that some content is not being accessed, a process will need to be in place to determine if that content is no longer meaningful or if it should be put in a more suitable category.

Choosing a Workflow Application

Acme is examining a number of applications for managing the content production workflow. They want to be sure that the application they choose will meet their needs now and in the future.

When choosing or building such an application, it's necessary to identify which features will be required to meet the needs of the business. The needs and corresponding features that Acme is considering include the following:

- Many organizations have document states other than "completed" or "incomplete", such as "internal use only" or "available to partners but not to customers". Being able to configure states and associate rules with them is part of a complete workflow solution.
- Some workflow applications require each user to log in and check their assignments. However, a more automatic mechanism, such as e-mail notification, would deliver workflow assignments much more effectively.

Collecting Content

Acme will assemble content from a variety of sources, both internal and external. Some of the content is contained in existing databases, but most is unstructured content. In addition, the site will deliver diagnostic tools and software patches.

Other vendors have approached Acme with requests to place information and advertising for their products and services on Acme's site. Because this would add value and encourage additional traffic to their site, Acme is determining what would be necessary to make this happen. Their content management solution will need to provide for some means of automated content collection.

As content is collected, the system should automatically record certain information about it that will help track the content within the workflow. Having content carry its own metadata, if possible, can automate some of the content conditioning.

Because Acme plans to subscribe to content from other vendors, its system should have a means of scheduling the file transfer for times of low traffic, to minimize the impact on system performance. In addition, security measures need to be taken in partnership with the parties providing the content. Content aggregation requires a thorough security architecture plan.

Providing Content

Acme plans to supply content to other vendor sites. The content management solution must, therefore, provide for syndicating content. Again, corresponding security planning is critical.

In addition, there are individual users who will want to know when new content has been added or when some portion of the site changes. Automated client subscription, notifying users when elements of the site have changed, should place the corresponding links within the notification. The system should support managing notification at either a folder or file level.

Content Delivery Applied: Serving the Content

Content delivery is concerned with getting personalized content to users over multiple channels.

Providing Role-Based Content

A key goal for Acme's new Web site is that the most appropriate content be delivered to each visitor. Users should no longer have any reason to complain about searches yielding too many hits that were unrelated to what they wanted. Along these lines, the company has already planned to condition the content to refine searches and match them to users' needs (or roles).

In a setting like Acme's, the most practical type of personalization is explicit: the users define their profiles, typically by completing an online form. Acme will have to determine what characteristics define a role, which will consequently be used to match content to the user.

Reaching People Where They Are

Acme plans to allow customers to request notification of changes or additions to the Web site. The default method will be e-mail. In addition, distributors of Acme's products have requested that wireless notification be a feature of the improved site. Therefore, Acme is examining the requirements for using wireless media for notification of updates, as well as for special offers. It's clear that the standard HTML Web pages will not work over the wireless devices; however, wireless media can deliver much of the same text, and the goal is to eliminate as much redundant entry and content generation as possible.

Considering Performance and Fault Tolerance

The plan is that as traffic to Acme's site increases, additional servers will be added to handle and balance the load, and the site will remain operational if one server should fail. Therefore, the applications that are part of the content delivery solution must support a multiple-server scenario.

As is typical of content management applications, templates will be used to place the dynamic content in an easily readable format. Recurring elements within the templates will be requested from the application server with each use, by default. Since caching reusable elements will reduce the amount of time required to build a page in response to a user request, Acme plans to include caching as a site feature.

Getting Real

This case study has presented an imaginary content management project with typical needs. The breadth of content management solutions is such that this case study was kept simple, highlighting the major activities within each of the content management discipline's workflow details: content development, content production, and content delivery.

Note that the activities highlighted here are very general in nature, and would likely be more specific in a real-life content management project. Adding even just a single condition, such as making the Web site bilingual, could greatly increase the complexity of the project.

Content Management Glossary

This glossary lists only those terms specifically related to content management. For RUP-related terms, see the RUP product itself or the book [*The Rational Unified Process: An Introduction*](#).

Term	Definition
Aggregation	A composition technique for building a new object from one or more existing objects that support some or all of the new object's required interfaces. In content management, this term refers to building content from more than one source, which may call for supporting multiple interfaces and formats.
Broadband	A transmission medium capable of supporting a wide range of frequencies, typically from audio up to video frequencies.
Caching	In content management, storing frequently reused elements of pages in folders on the Web server, precluding the need to regenerate or request them when building a page that includes those elements.
Categorization	The classification of a collection of items into a distribution of groups, as classes, orders, families, and so on, according to some common relations or affinities.
Category	A specifically defined division in a system of classification.
Category keywords	A small set of terms designed to convey the subject of content, such as a technical article. Some content management solutions specify a fixed set of keywords from which those for a particular content item should be chosen.
Channel	In communications, a path of communication between two computers or devices. This term can refer to the physical medium (the wires) or to a set of properties that distinguishes one device from another.
Conditioning content	Preparing content for entry into a content management system, including assigning keywords to facilitate categorization, searches, and personalization, and maintaining metadata that describes the content. The combination of categorization and metadata allows more precise specification when searching for content than does simple indexing of the incidence of words within the content.
Content	A written work, such as an article, white paper, or even multimedia material, delivered to a recipient in any of a number of ways. Since formats and delivery channels are constantly evolving, this term is a generalization that replaces other more format-specific terms (such as "Web page") that depend upon a particular method of delivery.
Content collection	A group of content items that have something in common, such as physical location or subject matter.
Content delivery	The methods by which content is made available to users, including the personalization of the content.
Content development	Establishing the framework of tools, languages, file formats, and guidelines involved in creating or editing content, including content delivery means such as templates.
Content discovery	Locating or recognizing content within existing information systems.
Content Management (CM)	A variety of tools and methods that are used together to collect, process, and deliver content of diverse types.
Content production	Processing content in preparation for publishing it, including workflow, conditioning, content administration, aggregation, and syndication.
Content rationalization	Collecting various types of content from multiples sources and presenting them as a cohesive whole.

Term	Definition
Document management	Software used to (among other things) check documents out and in, to prevent multiple contributors from changing the content at the same time.
Document Type Definition (DTD)	The definition of a document type in SGML or XML, consisting of a set of markup tags and their interpretation.
DTD	See document type definition .
Dynamic content	Content that's created at the time the user requests it. Dynamic content changes over time, so that each time users view it, they see different content. Contrast with static content .
Internationalization	The preparation of architecture and system design to support processes worldwide, including browser support, support of diverse character sets, and message catalog translation for each supported human language.
Keywords	See category keywords .
Knowledge management	A distributed hypermedia system for managing knowledge in organizations.
Localization	The process of making a content item fit regional tastes and expectations, including user interface design, locale-dependent features (such as date and time format), and the actual translation of content.
Message catalog	A set of common HTTP and application messages that could be encountered as someone navigates a site. Most Web servers come with a set of common HTTP messages, such as the ubiquitous message 404 indicating that something requested couldn't be found. In addition, applications that are integrated into a site will require a set of messages to address malfunctions, such as data missing from a required field.
Metadata	In content management, data that provides information about or documentation of other data managed within an application or environment. Metadata could document data about data elements or attributes (such as name, size, and data type), about data records or structures (such as length, fields, and columns), or about data itself (such as where it's located, how it's associated with other data, and ownership).
Meta-tagging	The process of attaching metadata to a piece of content, either by adding it as part of the content or by assigning it in a separate database; so-called because of the tags used to include information about a document in the head of an HTML document.
PDA	See personal digital assistant .
Personal Digital Assistant (PDA)	A handheld device that typically combines computing, telephone/fax, and networking features, functioning as a cellular phone, fax sender, and personal organizer. Unlike portable computers, most PDAs began as pen-based, using a stylus rather than a keyboard for input.
Personalization	The creation of a personal experience for visitors to a site based on information gathered from them, whether by invitation and submission (explicit) or based on observation of their behavior (implicit).
Static content	Content that doesn't change unless it's edited or replaced at the Web site. Contrast with dynamic content .
Streaming media	A technique for transferring data such that it can be processed as a steady, continuous stream. The client browser or plug-in can start displaying the data before the entire file has been transmitted.
Structured content	Content in forms (such as a database) that are already part of a related whole and have an inherent ability to support queries or sorts. Contrast with unstructured content .
Syndication	The process of supplying content to others for use on their Web sites.
Tag	A command inserted in a document that specifies information about the document or a portion of the document (such as formatting). Tags are used by format specifications such as HTML and XHTML that store documents as text files.
Taxonomy	Division into ordered groups or categories.

Term	Definition
Template	In content management, a predefined structure used to deliver content. A template is usually made up of component pieces, each of which provides either static content (such as logos and site navigation tools) or dynamic content derived from the request for the page.
Unstructured content	Content such as articles, white papers, and marketing literature. Unstructured content is not inherently part of a data structure (although it can have metadata associated with it). Contrast with structured content .
WAP	See Wireless Application Protocol .
Wireless Application Protocol (WAP)	A secure specification allowing users to access information instantly using handheld wireless devices such as mobile phones, pagers, two-way radios, smart phones, and communicators.
Workflow	In content management, the sequence of activities a business performs to produce, process, and deliver the content from its entry into the system until it expires and is deleted or archived.
Workflow application	In content management, the content production workflow application — that is, the application used to administer content and otherwise manage the workflow related to content production.
XML	Short for Extensible Markup Language, a specification developed by the W3C. XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

Related Resources

[“Content Categorization: An Orientation to Vignette and Content Management”](#) by Damien Kennedy (Rational Developer Network article). In addition to discussing the evolution of content management on the Web, this article describes a hierarchy of categories for classifying the many aspects of content management.

[W3C’s HTML home page](#). Here you’ll find information on the history of HTML and the Web, as well as links to other sites containing information on the evolution of the Web.

[The Rational Unified Process: An Introduction, Second Edition](#) by Philippe Kruchten (Addison-Wesley, 2000).

About the Author

Michael McIntosh is a freelance writer, project methodologist, and Web consultant. Since 1987, he has trained and consulted in project management, Web development, and computer networking. Formerly with Vignette Corporation, he was instrumental in the development of Vignette’s Solution Methods (VSM) 3.0. He lives in Austin, Texas, with his wife Julie and son Willie. Michael can be reached [via e-mail](#).

Rational®

the software development company

Dual Headquarters:

Rational Software
18880 Homestead Road
Cupertino, CA 95014
Tel: (408) 863-9900

Rational Software
20 Maguire Road
Lexington, MA 02421
Tel: (781) 676-2400

Toll-free: (800) 728-1212

E-mail: info@rational.com

Web: www.rational.com

International Locations: www.rational.com/worldwide

Rational, the Rational logo, and Rational Unified Process are registered trademarks of Rational Software Corporation in the United States and/or other countries. Microsoft, Microsoft Windows, Microsoft Visual Studio, Microsoft Word, Microsoft Project, Visual C++, and Visual Basic are trademarks or registered trademarks of Microsoft Corporation. All other names used for identification purposes only and are trademarks or registered trademarks of their respective companies. ALL RIGHTS RESERVED. Made in the U.S.A.

© Copyright 2002 Rational Software Corporation.
Subject to change without notice.