

Content Management Using Rational Unified Process

Part 1: Content Management Defined

Introduction

This paper presents an overview of content management, particularly as it relates to delivering content on the Web. The goal is to introduce the reader to the broad scope of content management and the associated challenges. The paper then examines why Rational Unified Process (RUP) best addresses those challenges by providing an efficient framework and proven methods.

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 the content to the Web site does not ensure that it will reach the proper constituencies or meet their information needs.

The content must be constantly updated if it is to have value. It needs to be processed, to ensure that it is ready for delivery to customers. The content delivery needs to be designed so that the user can locate what they need in a minimum number of steps. The content that is provided to the user should be matched to the needs of their business role — technical documents to technical users, and so on.

The collection and processing of content intended for a site is a more complex and time-sensitive process than many legacy document management systems are designed to handle. A content management application may require integration with existing document management systems, or may require a separate workflow of its own.

So, while the concept of content management is not new, implementation of a content management project is still a challenge for which there are more products than processes.

Content management, therefore, represents a collection of tools and methods that are used together to collect, process, and deliver content of diverse types. There are many variations on this theme.

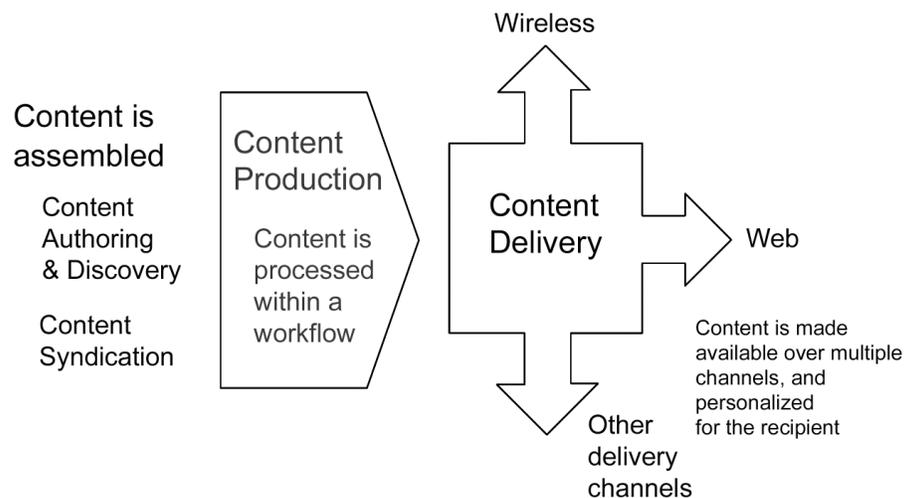


Figure 1 - Content Management Process Overview

Content can be aggregated from a variety of sources. Some content exists in a structured form such as a database. However, most of what needs to be delivered is unstructured content such as articles, white papers, and marketing brochures. 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 an essential chore is managing the workflow of activities related to unstructured content. Within that workflow are varieties of tasks (such as content conditioning) that require human input. The following table provides a few examples of tasks associated with unstructured content.

Content conditioning	The preparation for content entering a content management system that includes assigning keywords to facilitate categorization, searches and personalization.
Versioning documents with multiple authors	A document management issue with unstructured content is that there may be multiple authors, or an editing process, that requires that more than one person may be making changes to a content item. Changes need to be made without losing the work of any of the contributors, which involves checking documents out and in. As well, it is desirable that it be a simple procedure to roll a content item back to an earlier version, if the need arises.
Integration of code and content	Unstructured content items may be associated with particular templates for delivery, or may need to contain or reference code (such as style sheets) that affects the delivery or behavior of the content.

Table 1 – Examples of common activities associated with unstructured content

Content is delivered using software applications that are able to meet (sometimes anticipate) the needs of various constituencies. Content may be delivered by means such as wireless devices, as well as the Web. There is an increasing need to provide the information in multiple human languages, as well.

Content management, then, entails everything from identifying the need to provide some type of content, through the creation, processing, and delivery of the content, which may entail the translation of the content both for different delivery channels and for human languages. Due to the rate at which content grows and the way that our use of the information changes as we interact with it, parts of the solution, such as categorization and user interface, will probably evolve over the life of the project, leading to ongoing iterations.

In order to better understand the activities of any complex process, it is useful to group the activities. Within RUP, groups of activities within a discipline are called workflow details. 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 tools for dynamic content generation, such as templates
- Content production, which addresses the workflow of tasks used to manage the content through its lifecycle
- Content delivery, which addresses getting personalized content to the user over multiple channels

This paper examines each of these workflow details, describes the function of each, provides general guidelines, defines related terms, and provides lists of basic activities with corresponding types of artifacts. There are further examples of application of these workflow details, using the simplified case study that is a companion to this document. In addition, it introduces the application of RUP as an appropriate framework for content management projects.

This is Part 1 of a series on content management and RUP. Part 1 identifies the challenges of content management and presents a high-level approach to understanding those challenges within the workflow details of content management. The goal is for the reader to recognize the scope of content management projects and to understand why RUP is an appropriate framework for implementing these projects. Part 1 concludes with a view of content management in the context of the RUP phases.

Part 2 explains how to apply RUP as a framework for content management and examines the content management workflow details in detail. It provides guidance for requirement definition, identifies tasks and roles, and introduces content management artifacts for RUP.

It is assumed that the audience for this series of papers has some experience with RUP and an interest in the discipline and requirements of content management.

RUP and Content Management

Why should Rational Unified Process (RUP) be used as the framework for content management projects?

Content management entails far more than buying or building an application that provides content workflow. Made up of database, production workflow, content delivery, and Web applications, the development of a content management solution is a group of related software projects, making it a perfect candidate an established software-development framework such as RUP.

Why RUP should be used for Content Management Projects

<p>Content management projects are software projects</p>	<p>A content management project has much in common with any other software development project, and therefore much of what needs to be addressed is within RUP's existing disciplines. This precludes the need to start from scratch. In addition, there is a wealth of related guidelines within existing RUP roadmaps such as the e-Business Solution roadmap.</p>
<p>Content management projects are complex</p>	<p>There is a broad range of roles — business, technical, and creative — that are typically involved in a content management project. It is important that the framework for managing such a project be able to encompass the diverse needs of those constituencies. RUP helps strengthen the communication among the various roles by providing artifacts that they share as they perform the activities</p>

	within their respective workflow details. In addition to clarifying the vision and requirements, the artifacts help identify risks so that they can be addressed.
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. RUP has demonstrated success as such a process.

Table 2 - Reasons for using RUP for content management

Content Management Defined

This section provides an overview of how content management has evolved and what is encompassed within a content management solution.

The Evolution of Content Management

The introduction of HTML as a standard document format that allows for the free mixing of content and formatting within a document gave birth to the Web, which is now the preferred channel for delivering content.¹ The Web as a communication medium has increased both the demand for content and the need to manage that content so that it can be delivered in a timely and meaningful way.

The largely unstructured content that is delivered via the Web — content like this white paper, marketing brochures, articles — cannot be decomposed to a lower level. While structured content (such as a parts catalogue) has similar analysis requirements, the need for elaborate meta tagging or content conditioning is greatly reduced because structured content is already scrubbed down to its lowest level.

Web content delivery has shifted the focus from traditional database and transaction systems to the ability to rapidly process large amounts of unstructured content. Instead of two newspapers per day, there is now hourly real-time production turnaround.

¹ For information on the history of HTML and the Web, visit <http://www.w3.org/MarkUp/ - historical>. There are additional links at that location to other sites containing information on the evolution of the Web.

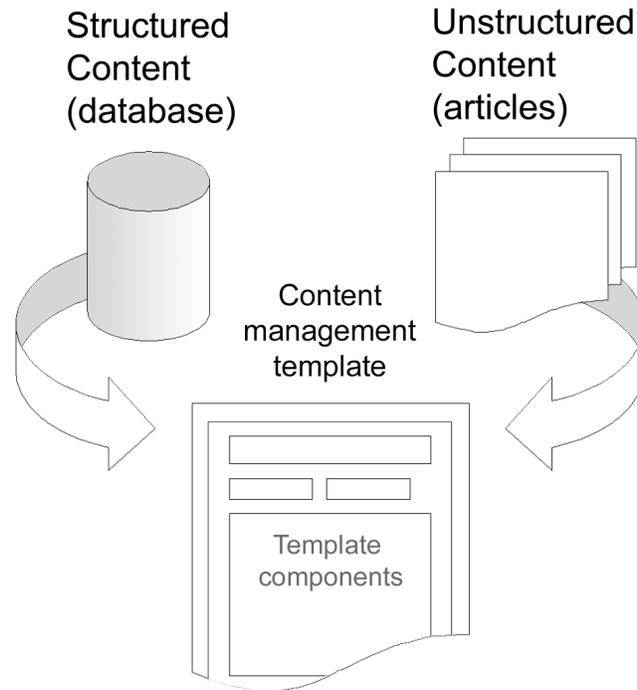


Figure 2 - Dynamic content delivery based upon templates

The introduction of additional delivery channels such as wireless devices further complicates the delivery of unstructured content, in that all of the elements or components of the content piece may not be compatible with the delivery medium.

The value of working through the roadmap presented in this series of white papers, regardless of the specific content of your site, is that every website has a mix of structured and unstructured content that requires management. As observed earlier within this paper, some of the greatest challenges entail managing unstructured content types.

The Scope of Content Management Projects

As has been stated previously in this paper, content management is much more than simply creating content or specifying that it be published to the Web. Content management includes activities such as:

- Assembling diverse content types from multiple sources
- Using templates to make the content available through a variety of channels, including wireless devices, e-mail, FAX, as well as the standard browser, while at the same time avoiding the need to create or manage redundant copies of the information in each format
- Using workflow to administer the content throughout its life-cycle
- Creating categories for cataloging the content
- Tagging content appropriately so that users can easily find what they need, having it provided to them based upon who they are, or having it recommended because of other content that they have accessed
- Identifying relationships and linking one piece of content to another

- Automatically notifying users when content changes or when there are special business opportunities that relate to that content
- Integrating content management with other parts of an information system, such as knowledge management and document management

In summary, content management is a broad undertaking that generates and dispenses content of various types, delivers that content using applications that are able to meet (sometimes anticipate) the needs of various constituencies using a variety of channels for delivery, and uses workflow to administer the content.

The next section divides the activities necessary for a content management project into three workflow details. As the section discusses each workflow detail, it provides examples of the challenges that might crop up for each topic, using a simplified case study.

The Content Management Discipline

As previously mentioned, in order to better understand the activities of any complex process, it is useful to group the activities into workflow details. The workflow details of content management are:

- Content development
- Content production
- Content delivery

This section describes each of the content management workflow details, and it presents typical challenges encountered in a content management project, using the Acme scenario.

Content Development Defined

Content development consists of the framework of tools, languages, file formats, and guidelines involved in creating or editing the content, including content delivery means such as templates. Content development defines much of the architecture of a content management solution. It can be broken into content authoring, development tools, development guidelines, and content structure.

Concept	Related activities and concepts	Types of artifacts produced
Content authoring	Writing text	Articles, white papers, brochures
	Developing multimedia content	Sound files, multimedia files
	Developing dynamic content delivery templates	Code for dynamic delivery of Web content, as well as for other channels, such as wireless

	Addressing legal issues such as intellectual property rights	Identified risks of publishing intellectual property to the Web. Legal agreements for users of the site with verbiage stating site policies for use of content.
	Internationalizing versions of the site for other languages and locales	Translated content. Message catalogues in other languages,
Defining development tools and guidelines	Selecting development platforms	Standards for site development, including supported code development platforms, databases and mark-up languages. An environment for testing and developing using any of the selected or proposed platforms.
	Internationalizing development	Confirmation of support for international character sets within the software specifications. Multilingual versions of software development tools.
	Selection of the development process	Adoption of the process, and the tools for implementing it.
Defining content structure	Defining XML strategy	XML Document Type Definition and related schemas
	Define database requirements	Database schema and requirements for database application
	Define data conversion requirements	Data conversion strategy

Table 3 - Content development

Content authoring

In addition to writing text, content authoring activities includes recording audio or generating other multimedia content types, creating templates for dynamic content generation, 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 time formats, sort orders, and numeric representations.

Defining development tools and guidelines

The environments within which various content types are created, such as mark-up languages, scripting or programming languages, and development platforms all fall within development tools. Development tools also address internationalization concerns such as application, programming language, and database support for international character sets. 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 Unified Modeling Language (UML) and RUP.

Defining content structure

Content structure is the definition of data structures. XML document type definition (DTD) or XML standards are addressed within this area, as are database schemas and data conversions. The content delivery solution often depends upon this architecture, making content structure design one of the highest priorities for content management.

Content Production Defined

Content production is the processing of the content in preparation for publishing it. Topics covered include using workflow to process content, conditioning content, administering content, aggregating content and syndicating content.

Concept	Related activities and concepts	Types of artifacts produced
Using workflow to process content	Content routing	Business analysis of how content is currently handled, including all aspects of creation, delivery and archiving. Reference within the Vision document to the content routing application and content production workflow. Rules for assigning content items to the users responsible for working on them (based upon use cases), as well as methods for manual assignment within the

		<p>content workflow application.</p> <p>Appropriate features within the workflow application to meet the content routing requirements.</p>
	Task management	<p>List of discrete tasks within the proposed content workflow, with descriptions of what they entail and what constitutes completion of each.</p> <p>Use cases describing each task.</p>
	State management	<p>List of the various states through which a content item might pass, criteria for those states and the tasks that relate to those state changes.</p>
	Role-based security planning	<p>A completed security plan containing a list of the roles within the security structure, with the rights and constraints for each role as they relate to both content items and system administration.</p>
Conditioning content	Categorizing content	<p>A categorization schema, with associated category keywords and an application (within the content production workflow) that will associate the keywords with the content items to support searches.</p>

	Describing content	Features within the workflow application for capturing metadata for each content item describing the subject matter, the format of the content and any other information that is part of the content conditioning schema.
Content administration	System administration of the workflow	System administration features that manage the creation and characteristics of users and groups, as well as the configuration of the workflow and the rules that govern it.
	Administration of content items	Content production features within the application that manage the state of content and provide for standard content management operations.
Aggregating content	Gathering content from various sources for use on the site	<p>Descriptions (within the software development plan) of all the sources from which content is to be gathered, in what format it is expected to be at the source, how it is to be transferred and in what format it is to be for delivery.</p> <p>Architectural plans to support communication between the various systems.</p> <p>Security plans for those instances where content will be coming from locations outside the organization, such as partner companies.</p>

	Content rationalization	Software development plans for the collection of various types of content from multiple sources and presenting them as a cohesive whole. Supporting templates for those situations requiring content rationalization.
	Content conversion	Template components for translating the format of the data, such as converting results from a database query to HTML.
Syndicating content	Subscription management	A schedule for when updates are to be available. Considerations within the security plan for who can receive updates.
	Digital goods delivery	Strategies within the software development plan addressing moving different types of content from its source to its destination.

Table 4 - Content production

Using workflow to process content

Workflow is the routing of content items to the users responsible for working on them. The routing should be rule-based, with the option to manually assign responsibility when required. Workflow is concerned with providing the information and tools required to support each step of the cycle. It needs to be broken into discrete tasks, which are then assigned using rules or administrative procedures.

It is advisable to implement role-based security measures to limit who has access to various content items and the changes that 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

Personalization describes the methods of matching content to the needs of specific groups of users. Conditioning is adding information to or about a content item to facilitate that matching. Conditioning uses the categories that are established for personalization and fine-tuning searches. This is accomplished using category keywords (terms used for matching search criteria about the content) and metadata (information that describes the content). Ideally, multiple category keywords are supported, and there is a means of rating the suitability or importance of content to its categories. In this way, an article describing installing a network card for 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.

It is a requirement for conditioning that there be an accepted taxonomy, or descriptive nomenclature, to describe the categories. It is not uncommon to have more than one way of describing the same thing, even within the same organization. The taxonomy is used to create a categorization hierarchy or tree. The hierarchy can be used as a basis for a method of navigation. Since content can be associated with more than category, it should be possible to locate it by navigating from different directions.

Documents can carry categorization metadata within them for the automation of content categorization. Examples of document formats that support this are HTML and XML. In addition, a content management system can provide categorization for document types that do not have a way of carrying their own metadata.

Conditioning is included within the content workflow.

Content administration

Content administration encompasses the administration of both the workflow and the content items. The state of the content needs to be managed. Examples of states include incomplete, requiring additional editing, ready for publication or requiring the approval of the legal department. Once content has been processed by a task within the workflow, the state of that content might automatically change. After the content is published or rejected, it needs to be deleted, archived, or marked for expiration. A summary of content administration needs includes:

- Submitting content items for publication
- Locking content items to prevent access by more than one writer at a time
- Versioning content as a method of tracking changes by version number
- Archiving content, which is maintaining copies of content items
- Deleting content or replacing it with content that is more current
- Approving content items for publishing and sending them to the content delivery system

Aggregating content

Organizations keep content in many places and in many forms. As well, they often receive or subscribe to content from external sources, such as partner companies. Aggregation is the assembling of content from a variety of sources, which may be both internal and external. It is unlikely that assembled content will all be in the same format. Therefore, two key concepts that fall within aggregation are:

- Content rationalization, which is the collection of various types of content and presenting them as a cohesive whole. An example is 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.
- Conversion, which is translating the format of the data. An example is converting information that results from a database query to HTML for display in a browser.

Syndicating content

Syndication is providing the most current content to other sites, or to browsers that have been configured to receive scheduled updates for that site. It is becoming increasingly common for companies to share information on their products and services with other companies for inclusion on their websites. In addition, there are client subscription technologies, which 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 is the scheduling of updates and the security considerations of syndication: who can receive the updated content and how frequently the update should be performed.
- Digital goods delivery, which addresses the fact that not all content is simple HTML, and that a variety of means may exist to move the content from its source to its destination.

Content Delivery Defined

Content delivery describes the parts of the content management solution which have the greatest public visibility: the website and other delivery channels. In addition, content delivery describes the methods used to match content to the user, and to address the inherent overhead of dynamic page delivery. Personalization is the technique of delivering content that matches content to information discovered or recorded about the user. Since the performance of the site is an aspect of delivery that can be affected when content is provided dynamically, caching is included as a technique with content delivery as well.

Application needs, which are dictated by the content delivery requirements, largely determine the standards established within content development. In fact, decisions made about the final delivery of the content drive most of the content management project.

Concept	Related activities and concepts	Types of artifacts produced
Personalizing content for users	Creating a personal experience for users	<p>Plans for gathering information about users explicitly or implicitly.</p> <p>Defined roles for users that play a part in determining the most appropriate content for delivery.</p> <p>A content categorization schema, which has been applied to the content items so that all of the proper content can be submitted to a particular role.</p>
Delivering content over multiple channels	Making content available to various delivery media, including Web browsers, personal digital assistants (PDA) and wireless devices such as pagers and cell phones	<p>The Vision of the project needs to specify the delivery channels that will be supported.</p> <p>Templates that appropriately format the content for the medium over which the user is requesting it.</p>
Caching content	Keeping frequently used elements in memory or on disk to improve performance	<p>A content delivery application that uses caching.</p> <p>Methods for identifying items that are to be cached to the content delivery server.</p>

Table 5 - Content delivery

Personalizing content for users

Creating a personal experience for visitors to a site based on information gathered from them, whether by invitation and submission (explicit), or based upon observation of their behavior (implicit), falls within the set of techniques called personalization. Again, there is a relationship between content delivery and other content management, since the conditioning of the content (metadata and categorization) that is addressed within content production is based upon the personalization design.

Delivering content over multiple channels

A channel is a method of content delivery, the default method being the Web. In addition to Web browser access, methods might include: e-mail, pager, cell phone, and Personal Digital Assistant (PDA). Wireless Application Protocol (WAP) is an example of technology that is growing rapidly in some parts of the globe. The increased availability of broadband access implies a growth in technologies such as streaming media that depend upon high bandwidth for effective use.

Caching

Caching describes techniques used to keep recently or frequently used pages in memory or on disk to improve performance. The dynamic page generation commonly employed in content delivery increases overhead. 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 RUP's Phases

Previous sections of this paper examined content development, content production, and content delivery. The sections briefly described the work that each entailed and provided a conceptual vignette, using the Acme company example.

The remaining sections of the paper look at applying RUP to content management. These sections examine the workflow details of content management in the context of the four phases of RUP, assuming that the reader has a reasonable familiarity with RUP.

The phases of RUP are:

- Inception
- Elaboration
- Construction
- Transition

By examining the content management process by phase, the flow of a content management project will emerge. Evaluation criteria are quoted from RUP, where applicable. This is an incomplete list of criteria for assessing the phases of a project using RUP. For brevity, the emphasis is on those criteria that are more specific to a content management project.

Inception

The overriding goal of the Inception phase is to achieve concurrence on the lifecycle objectives for the project. It is important to define the vision of the project and record it. For content management the vision typically focuses on content delivery, with the some interest in content production driven by those responsible for the content workflow.

Content development involves establishing the project's software scope and boundary conditions, including an operational vision and acceptance criteria: what is intended to be in the product and what is not.

Architecture decisions can be facilitated by exhibiting, and maybe demonstrating, at least one candidate architecture against some of the primary scenarios.

Investigating the needs of the content production system may include preparing the supporting environment for the project.

Content delivery and content production will depend upon defining the critical use cases of those systems. The primary scenarios of operation will drive the major design trade-offs.

Content delivery has the highest visibility within content management projects. Stakeholder concurrence on the scope of content delivery is critical, as is agreement as to what it will take to accomplish the desired results.

There needs to be agreement that the right set of requirements has been captured and that there is a shared understanding of these requirements.

Lifecycle Objectives Milestone	
Related activities and concepts	Types of artifacts produced
Developing a vision for the completed site	Vision document describing the content delivery interface and content production workflow.
Addressing legal issues such as intellectual property rights	Identified risks of publishing intellectual property to the Web.
Selecting development platforms	Standards for site development, including proposed code development platforms, databases and mark-up languages.
Creating a personal experience for users	Plans for gathering information about users explicitly or implicitly. Defined roles for users that play a part in determining the most appropriate content for delivery.
Content routing	Business analysis of how content is currently handled, including all aspects of creation, delivery and archiving.
Task management	List of discrete tasks within the proposed content workflow, with descriptions of what they entail and what constitutes completion of each.
State management	List of the various states through which a content item might pass, criteria for those states and the tasks that relate to those state changes.

Making content available to various delivery media	The Vision of the project needs to specify the delivery channels that will be supported.
--	--

Table 6 - Lifecycle Objectives Milestone

Elaboration

The goal of the Elaboration phase is to baseline the architecture of the system to provide a stable basis for the bulk of the design and implementation effort in the construction phase. It should also be demonstrated that the baselined architecture will support the requirements of the system at a reasonable cost and in a reasonable time.

The stability of the architecture is evaluated through one or more architectural prototypes. An example of useful prototypes is the development of templates for dynamic content delivery.

The architecture evolves out of a consideration of the most significant requirements (those that have a great impact on the architecture of the system) and an assessment of risk. Content delivery, especially where personalization and channels are included, will include some of the greatest risks, due to its complexity.

In order to achieve these primary objectives, it is necessary to set up the supporting environment for the project. This includes creating a development case and guidelines as well as setting up tools. The initial setup of a content management workflow application, if one has been purchased or is being given a trial, occurs during this phase. In order to allow prototyping of templates, at least one development Web server will need to 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.

Lifecycle Architecture Milestone	
Related activities and concepts	Types of artifacts produced
Selecting development platforms	Standards documents for site development, including supported code development platforms, databases and mark-up languages. An environment for testing and developing using any of the selected or proposed platforms.
Selection of the development process	Adoption of the process, and the tools for implementing it.
Internationalizing development	Confirmation of support for international character sets within the software specifications. Multilingual versions of software development tools.

Creating a personal experience for users	A content categorization schema, which has been applied to the content items so that all of the proper content can be submitted to a particular role.
Gathering content from various sources for use on the site	<p>Descriptions (within the software development plan) of all the sources from which content is to be gathered, in what format it is expected to be at the source, how it is to be transferred and in what format it is to be for delivery.</p> <p>Architectural plans to support communication between the various systems.</p> <p>Security plans for those instances where content will be coming from locations outside the organization, such as partner companies.</p>
Categorizing content	A categorization schema, with associated category keywords and an application (within the content production workflow) that will associate the keywords with the content items to support searches.
Content rationalization	Software development plans for the collection of various types of content from multiple sources and presenting them as a cohesive whole.
Subscription management	<p>A schedule for when updates are to be available.</p> <p>Considerations within the security plan for who can receive updates.</p>
Defining XML strategy	XML Document Type Definition and related schemas
Define database requirements	Database schema and requirements for database application
Define data conversion requirements	Data conversion strategy

Content routing	Rules for assigning content items to the users responsible for working on them, as well as methods for manual assignment within the content workflow application. Use cases for all possible types of content routing.
Role-based security planning	A completed security plan containing a list of the roles within the security structure, with the rights and constraints for each role as they relate to both content items and system administration.
Digital goods delivery	Strategies within the software development plan addressing moving different types of content from its source to its destination.

Table 7 - Lifecycle Architecture Milestone

Construction

The focus of the Construction phase is to clarify the remaining requirements and complete the development of the system based upon the baselined architecture.

To accelerate the development activities, it is desirable to achieve some degree of parallelism in the work of development teams. The proper use of content management templates, and in particular the reuse of components within templates, can greatly reduce content development time.

The flow of content from content production to content delivery must be stable by the end of this phase.

The content delivery application can assemble content dynamically and deliver it to the channels that, in turn, pass it on to the users by the end of this phase.

The progress of the organizational transformation must be far enough along for the content production workflow to become part of the routine of the associated workers before this project can be handed off to the transition team.

Initial Operational Capability Milestone	
Related activities and concepts	Types of artifacts produced
Writing text	Articles, white papers, brochures
Developing multimedia content	Sound files, multimedia files

Developing dynamic content delivery templates	Code for dynamic delivery of Web content, as well as for other channels, such as wireless
Content routing	Appropriate features within the workflow application to meet the content routing requirements.
System administration of the workflow	System administration features that manage the creation and characteristics of users and groups, as well as the configuration of the workflow and the rules that govern it.
Administration of content items	Content production features within the application that manage the state of content and provide for standard content management operations.
Describing content	Methods within the workflow application for capturing metadata for each content item describing the subject matter, the format of the content and any other information that is part of the content conditioning schema.
Content rationalization	Supporting templates for those situations requiring content rationalization.
Content conversion	Template components for translating the format of the data, such as converting results from a database query to HTML for display in a browser.
Making content available to various delivery media	Templates that appropriately format the content for the medium over which the user is requesting it.
Define data conversion requirements	Data is converted for this project. Ongoing data conversion requirements are successfully addressed.

Internationalizing versions of the site for other languages and locales	Translated content. Message catalogues in other languages,
---	---

Table 8 - Initial Operational Capability Milestone

Transition

The focus of the Transition Phase is to ensure that site is available for its end users. The Transition Phase can span several iterations, and includes testing the site in preparation for release, and making minor adjustments based on user feedback. At this point in the lifecycle, user feedback should focus mainly on fine-tuning. Examples would include the effectiveness of personalization and ease of use. All the major structural issues should have been worked out much earlier in the project lifecycle.

As mentioned earlier in this article, content management projects are software projects. Part 2 of this series will provide greater detail on transition phase tasks. However, when applying RUP to content management, the transition phase tasks are essentially the same as they would be for any other RUP-managed project. Apart from the usual project management issues, the primary evaluation criteria for the transition phase involve the answers to this question: Are the users satisfied?

Product Release Milestone	
Related activities and concepts	Types of artifacts produced
Developing dynamic content delivery templates	Tested templates that reliably deliver dynamic content to all supported channels
Addressing legal issues such as intellectual property rights	Legal agreements for users of the site with verbiage stating site policies for use of content.
Internationalizing versions of the site for other languages and locales	An ongoing strategy for translating content.
Keeping frequently used elements in memory or on disk to improve performance	A content delivery application that uses caching. Methods for identifying items that are to be cached to the content delivery server.

Table 9 - Product Release Milestone

Summary

Content management is a topic with both breadth and depth. This paper has presented the three workflow details of content management:

- Content Development
- Content Production
- Content Delivery

For each of these areas, we have provided definitions, objectives and concerns.

Content management, then, entails everything from defining the need to provide some type of content, through the creation and delivery of the content, including the creation, configuration and administration of the delivery mechanism, and the translation of the content for different delivery channels.

RUP provides a proven method and multi-disciplined approach to identifying and assigning activities and responsibilities, as the project is iteratively delivered. In addition, RUP captures many best practices in a form that can be tailored for a wide range of projects and organizations.

Part 2 in this series examines using RUP for content management in detail, identifies specific activities within the content management workflow details, along with associated role assignments and artifacts.

Appendix

Glossary

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.
Broadband	A transmission medium capable of supporting a wide range of frequencies, typically from audio up to video frequencies.
Caching	In content management, caching refers to 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	Classifying a collection of items into a distribution of groups, as classes, orders, families, etc., according to some common relations or affinities.
Category	A specifically defined division in a system of classification
Category keywords	A small set of words 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.

Term	Definition
Channel	In communications, the term channel refers to a communications path between two computers or devices. It can refer to the physical medium (the wires) or to a set of properties that distinguishes one channel from another.
Conditioning content	The preparation for content entering a content management system that includes assigning keywords to facilitate categorization, searches and personalization. An additional conditioning technique is to maintain 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	Generally, the subject matter of a written work such as an article or technical paper. Since media and its formats are constantly evolving, the word "content" is a generalization that replaces other more format specific terms, such as Web page, that are dependant upon a particular method of delivery.
Content collection	A group of content items that share something in common such as physical location or subject matter.
Content delivery	Content delivery describes the methods by which content is made available to users. This includes the personalization of the content.
Content development	Content development consists of the framework of tools, languages, file formats, and guidelines involved in creating or editing the content, including content delivery means such as templates.
Content discovery	The location or recognition of content as such, within existing information systems.
Content management	Content management represents a collection of tools that are used together to collect, process, and deliver content of diverse types.
Content production	Content production is the processing of the content in preparation for publishing it. It includes workflow, conditioning, content administration, aggregation, and syndication.
Content rationalization	Content rationalization is the collection of various types of content and presenting them as a cohesive whole.
Discipline	A branch of knowledge. A system of rules of conduct or method of practice. A sequence of activities performed in a business that produces a result of observable value to role.

Term	Definition
Document management	Software used to check documents out and in, to prevent multiple editors from changing the content at the same time
Document Type Definition.	The definition of a document type in SGML or XML, consisting of a set of mark-up tags and their interpretation.
DTD	See "Document Type Definition"
Dynamic	In the context of content management delivery, information that is created at the time the user requests it. Dynamic information changes over time so that each time users view it, they see different content. Contrast with static.
Internationalization	Preparing the architecture and system design to support processes in a global environment. Included aspects are browser support, character sets, and message catalogs for common error or response messages.
Keywords	See "category keywords"
Knowledge management	A distributed hypermedia system for managing knowledge in organizations.
Localization	The process of making a content item fit to regional taste and expectation. This relates to user interface design, locale dependant features such as the way that dates or times are displayed, and the actual translation of content.
Message catalogue	A set of common HTTP and application messages that would be encountered as someone navigates a site. Most Web servers come with a set of common HTTP messages, such as the ubiquitous 404, which indicates that something that has been requested cannot be found. In addition, applications that are integrated into a site will require a set of messages to address malfunctions, such as missing data from a required field. Internationalization requires that the message catalogue be translated for each supported human language.
Metadata	Data about data. In content management, meta-data is definitional data that provides information about or documentation of other data managed within an application or environment. For example, metadata would document data about data elements or attributes, (name, size, data type, etc) and data about records or data structures (length, fields, columns, etc) and data about data (where it is located, how it is associated, ownership, etc.).
Meta-tagging	The process of adding or assigning metadata to a piece of content.
PDA	See "Personal Digital Assistant"

Term	Definition
Personal Digital Assistant.	A handheld device that can combine computing, telephone/fax, and networking features. A typical PDA can function 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	Personalization describes creating a personal experience for visitors to a site based on information gathered from them, whether by invitation and submission (explicit), or based upon observation of their behavior (implicit).
Static	Static content does not change unless it is edited or replaced at the Web site. Contrast with dynamic.
Streaming media	A technique for transferring data such that it can be processed as a steady and continuous stream. With streaming, 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 all format specifications that store documents as text files. This includes HTML and XML.
Taxonomy	Division into ordered groups or categories.
Template	Within content management, a pre-defined structure used to deliver content. A template will usually be made up of component pieces, each of which provides either static content, (such as logos and site navigation), or dynamic content that is usually derived from the request for the page.
Unstructured content	Content such as this white paper, marketing brochures, and articles. While such content can have metadata associated with it, it is not inherently part of a data structure. Contrast with structured content.
WAP	See "Wireless Application Protocol"
Wireless Application Protocol	Wireless Application Protocol is a secure specification that allows users to access information instantly via handheld wireless devices such as mobile phones, pagers, two-way radios, smartphones and communicators.

Term	Definition
Workflow	In content management, the sequence of activities performed in a business that produces, processes and delivers the content from its entry into the system to it expires and is deleted or archived.
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.

Copyright 2001 Rational Software Corporation. All rights reserved.