What is Content Management?

By Rita Warren | March 2003

There's no denying it. Content plays a big role in what we deal with in business every day. In fact, business people have dealt with content for decades. Oddly enough, the term "content management" only surfaced in the past few years, and its origins come not from the people who work with content every day—writers and editors—but from the software world. More often than not, content management has been defined as a product—a content management system (CMS). But, in reality, content management is not just a software technology, it's a discipline. It's a way of doing business.

One of the first things to do when figuring out content management is to have a clear definition of what is—and what is not—content. Following this understanding, you must then look at content management as a discipline, separate from the "system." And, finally, armed with a clear picture of what content is, and what is involved in the discipline of content management, you can apply what you've learned to understanding the "system"—a CMS.

What is Content?

One of the most elusive parts of understanding content management is defining the word "content." While there are many definitions, the one that works best for me is:

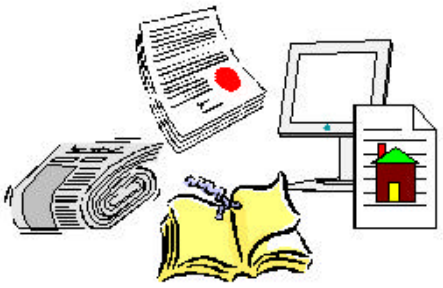
Content = information that is published

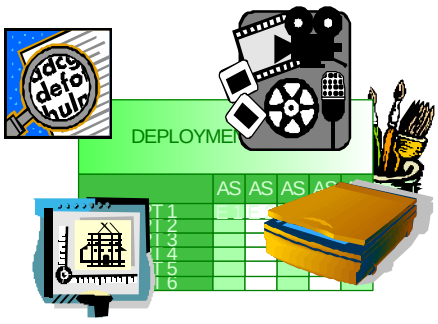
The term "publish" means "to make public." This definition helps to differentiate between what is content and what is not. For example, this document you're reading now is published, meaning that it is intended to be distributed or made available for a wider audience to view.

When you think about "publishing," depending on your background and experience, you might think of publications as strictly printed materials. A magazine is published. A book is published. With the advent of the computer and the Web, publishing has taken on an expanded meaning. In today's world, you might publish a document to a public file server, or publish to a site on the World Wide Web, or publish content via e-mail as a newsletter.

Some of the types of publications we encounter in business are:

- Word documents
- Spreadsheets
- Intranet pages
- Presentations
- E-mail newsletters or promotions
- Web pages
- Newsletters
- Catalogs
- Movies, TV programs, training videos





Publications can be made up of one or more types of *content*:

- Text
- Graphics
- Video
- Audio

While there are no hard-and-fast rules about what is and what is not content, it's important to not fall into the trap of thinking about *all* digital information as content. I've known of companies that consider the information in their customer database (CRM system), their employee database, or their product database to be content. I wouldn't consider any of these types of information to be content in the strictest sense. I'd call them *data*.

Data often creeps into the definition of what content management encompasses. While data plays an important role in content management and in creating publications, it is not—in-and-of-itself—content. As illustrated in the table below, there are subtle differences between content and data.

Content vs. Data

Content	Data
Examples: article, product description, picture, TV commercial, news story	***Examples:*** customer record, financial statement, product attribute set
Is the result of a creative effort	Is a result of an analytical effort
Can reflect either facts or opinions	Only reflects facts
Can be "free-form"	Always follows a set format or structure
Is primarily used to communicate	Is primarily used to track, calculate, or analyze
Is often structured in an outline or hierarchy	Is typically stored in rows & columns (relational db)
Often depends upon style formatting for readability	Doesn't rely on style formatting (bold, italic, etc.)
Requires editorial review	Requires quality assurance but not editorial review
Has an audience	Has a recipient or recipients

Keep in mind is that data is frequently published in publications. From this standpoint, it's easy to think that data is content. For example, a product part number in a catalog is a piece of data that is published along with the picture and description of the product. The main difference is that a part number is very data-like. It is factual. It follows a set structure (e.g. 4 digits and 2 letters), and its primary use is for tracking the item. Part numbers are easy to store in a cell in a database.

Besides these aspects, the tools and processes needed to store and manipulate content and data are different. Content, such as an article, is often structured in a hierarchical fashion and relies upon specific formatting such as size of headings, boldface, italic, and paragraph spacing for readability. Content, being the product of a creative process, also typically involves editorial review cycles for spelling, grammar, presentation, and meaning, whereas data simply needs a quality assurance check.

In keeping with the definition of publishing meaning "to make public," content and data differ albeit subtly in their intended viewers. A piece of content, such as a *product information sheet* is intended for defined audience of many. While a personal e-mail message between friends discussing the evening's plans is not data, it's not content either; it's a message. Similarly, something like your bank statement doesn't quite fit as content either, since it's intended for a specific recipient—not for public consumption. Again there are no hard-and-fast rules, but hopefully these examples will help you to define for your own purposes what does and does not constitute content.

Why Manage Content?

As I've said, content management isn't new; it's just that with the new technology available and the overabundance of content on networks and Internet, content management has now come front and center.

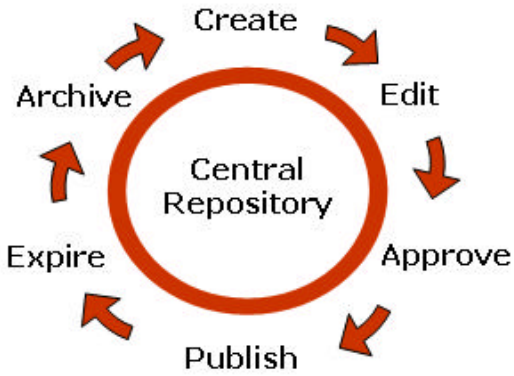
The reasons why organizations must now *manage* content are that:

- **People often share the same content.** Think of a network directory on your corporate network. One of the main reasons you save a file onto the network is so that other people besides you can use it. Whether it is a report, a specification, a layout, or a draft of a document, others will want to read it, revise it, copy all or part of it, distribute it, collaborate on it, or use it in any number of ways.
- **Different publications often share the same content.** Content "reuse" is the buzzword of the day. And it should be. I'm sure that you have run into the situation where the same basic content exists in several places, gets edited by different people at different times, and you end up with no clear idea of which is the "real" content. Maintaining multiple versions of what is *supposed* to be the same content not only wastes time but it also exposes the company to all kinds of risks—like when the reader gets bad information because the wrong version is published.
- **People need to find information.** The volume of content available today is simply overwhelming. The more content that people have access to, the more difficult it becomes to sift through it and find what they need. To find content, it needs to be organized, but—guess what—content doesn't organize itself. *People* organize content. New technologies are emerging that can organize content for you, but today it requires a human brain. Keep in mind that making content easy to find requires both automated processes and a people processes.
- **Content changes.** Look at any of the types of content that a workgroup creates and you'll be able to see that content follows a lifecycle. It is created by the original author, edited, approved, and published. While some content's useful life is practically infinite—like that in a classic novel or historical archives—much of the content created in business has a limited useful life. A piece of news may become irrelevant in as quick as a day, while other content may be relevant for years. Most content in business eventually becomes obsolete and needs to expire, be archived, or be destroyed (deleted).

The bottom line is that there are two main reasons for managing content well: efficiency and effectiveness. Now that it is easier for people to share content across networks or across the world, we need more efficient ways to find it and use it in publications. With this increase in information volume and accessibility, we need to make sure that the content that is out there is well written, well targeted to the audience, accurate, and current. This is what content management helps us do.

The Content Lifecycle

A useful way of thinking about content and why it needs to be managed is to look at the phases that content goes through from start to finish—the content lifecycle.



Whether a graphic, a video, audio, or a document, all content originates somewhere. It is created. Today in business, more than 70% of documents have more than one contributor. For example, once created, the piece of content then is edited or revised by a second person. Often approval happens by yet another person, and publishing by yet a fourth person. Think about your organization or your individual group. No doubt there are people in your group that perform some or all of these tasks.

Where the process often breaks down—and limits the effectiveness of the content—is in the expiring and archiving of content. File servers with thousands of mystery files, and Web sites with "old news" are all the result of poor or missing systems for managing the content. On a practical everyday level, the two main reasons that the discipline of content management exists are basically this:

- To make it easier for people find the content they need
- To make it easier for people to publish content to their audiences

The Discipline of Content Management

Now that you have an idea of what *content* is—information that is published—I'll delve into the fundamentals of what *content management* is. Management is exactly what it says—to manage. I like the second definition of "manage" from *Merriam-Webster*:

manage: to treat with care

When you think of content as one of the most valuable assets an organization owns, this is precisely what we'd want to do with content—treat it with care!

I like to refer to content management as a *discipline* because that's what it requires—discipline. Fortunately, with the miracle of modern technology, much of the discipline can be enforced through software—but not all of it. The most succinct definition I've been able to come up with for content management is this:

Content Management = The infrastructure for creating and maintaining content throughout its lifecycle

Not very glamorous, but, as one person once said, "Content management is like plumbing. It's not the most exciting thing in the world, but I sure wouldn't want to live in a house without it!"

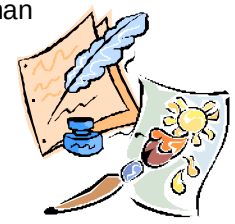
Are you living without it? Take a look at the following facets of content management. Do you have all of these pieces and processes in your organization? How well are they working? Is there room for improvement?

The multiple facets of content management as a discipline:

- **Content Planning.** While still not well ingrained in everyone working with content management systems, true content management starts *before* content gets created. It starts with a planning phase. Content planning involves defining what types of content will be created and when, as well as what types of publications will be published and when. The most successful publishing organizations develop content and publication calendars that define these things, as well as clear rules of who owns each piece, how the content will be distributed, and how the organization will measure the success of the content and the publications.

- **Guidelines & Standards.** Traditional publishing has a long history of working with guidelines and standards. The typical company style guide or publication style guide are old hat to publishers of print documents. Web guidelines and standards were slower in coming, but are now fairly common among Web publishing groups—especially those dealing with lots of Web pages and multiple content authors. Guidelines can be of various types, ranging to editorial style (what spelling, capitalization, and punctuation to use), to formatting (what size and treatment to use for text, what line spacing, where to place images on a page) and technical (what resolution to save images, what code standards for HTML in Web pages, or which settings to choose when saving to PDF.)

- **Content Authoring.** The term "authoring" in this context is used in a wider sense than usual. Authoring is creating any type of original content. It can be a writer writing, but also a graphic artist creating an image, or a videographer making a movie. Content management involves defining what tools authors will use to create content, keeping in mind that different types of authors and different types of content require different tools. For short, highly structured types of content, like a product description or a news item, authors may use Web-based forms. For more free-form content, like a chapter in a book, authors may use a word processing application. Artists and designers, of course, will use tools like Photoshop®, Illustrator®, Quark®, InDesign®, or Flash®. Content management systems often integrate with these authoring tools to allow content contributors to submit content to the CMS directly from their preferred desktop application.

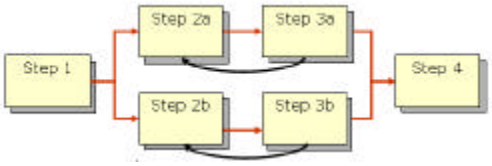


- **Content Acquisition.** Not all content that you store or use is authored by your own organization. Content that comes from somewhere else is called "acquired" content. It may be content that your suppliers or partners provide, or content that is purchased from companies that are content aggregators (known as "syndicated" content). As part of your content management infrastructure, you need to establish guidelines and standards for acquiring content, as well as systems and processes for converting, storing, using, and tracking the use of acquired content.

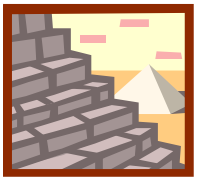
- **Editorial Workflows.** As we've said, content is unique from other types of information (like data) in its need to go through an editorial workflow process. A basic editorial workflow might involve a few simple steps like this:

  Step 1: Author | Step 2: Edit | Step 3: Approve | Step 4: Publish

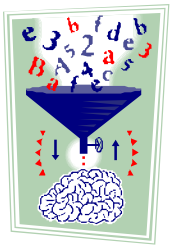
  Editorial workflows in business publishing groups are typically much more involved than the above example, requiring multiple review and approval steps for technical accuracy and legal appropriateness. Plus, in publications that include both text and images (or even online videos and animations) there are "parallel" workflows where different components follow separate paths and then come together on the page for final approval. One of the major tasks when implementing content management is defining all the editorial workflows for all of the different kinds of content, including what the steps and tasks are, who performs them, and how long each should take.



- **System Workflows.** Editorial workflows are not the only kind of workflow on the block, although many organizations don't put as much thought into system workflows. A workflow is basically a business processes. Other business processes that must occur when handling large volumes of changing content including workflows for testing Web templates, backing up the database, synchronizing the list of users with the network directory, and many others. Fortunately, many types the system workflows can be automated. But they all need to be carefully thought out and designed into the CMS configuration.



- **Content Organization and Storage.** A term you will encounter over and over in content management is "repository." The repository is the place where the content is stored and managed. It typically is a database, multiple databases, or a combination of databases and file servers. Two of the key objectives of having this central repository are to reduce the amount of redundancy by keeping only one "master" version of the content, and to organize the content so that it can be easily located and retrieved when needed. Of course, to store content in a database, you first have to organize it. Analyzing the content and defining how it should be stored in the repository is one of the most time-consuming undertakings in a content management initiative. A document that describes the organization and categorization of content is typically called a *taxonomy*.



- **Metadata.** I've talked about data being different than content, but also emphasized that data plays a key role in content management. The type of data that is most used in content management—for both finding and publishing content—is called *metadata*. Metadata literally means "information about the information." For example, a piece of metadata about an article is its author. A piece of metadata about a product would be its color or size. Metadata is what makes it easier to find information, and it is what allows you to publish content in an automated fashion. A primary task in content management is to define the "metadata model"—basically a listing of all the attributes about your content types that you want to use in order to track them, find them, or use them.

- **Search**. As mentioned previously, one of the main reasons why we manage content is so that we can find it when we need it. Search plays a dual role in a content management system. You may want to be able to find content in the repository itself (typically something someone internal to the organization would to do.) Or,

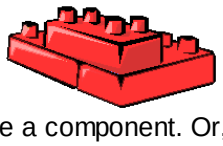
you may want to search for content within a publication (something someone external would typically do). Search can work in two ways: keyword search or full-text search. Keywords are part of your *metadata.* The person searching using keyword search will get the results they want as long as the keyword(s) they enter are part of the metadata for the document. If the search is set up for full-text search, you can retrieve all pieces of content that contain the search word or words if they appear *anywhere* in the text of the document.

- **Security.** While storing content in a central location is a great way to make it accessible to all, some content is either not appropriate or not applicable to all of the people in your organization. Part of any well-designed content management system is a security system that lets an administrator set permissions about who gets to create, view, change, or delete content. The security system is often integrated with the corporate network so that when a user is added or deleted, these changes are reflected in the CMS as well.

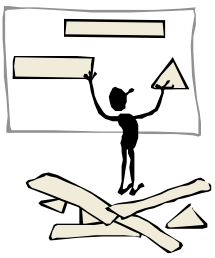
- **Publication Project Management.** Producing a publication of any notable size or complexity requires some level of project management. In particular, it is important to know who is working on what part of the publication, when that piece is due, and where that piece of content is in the process. Metadata, again, plays a key role in managing the publication production process. For example a piece of content stored in the repository will have metadata for *owner*, *status*, and *due date*. You can then generate reports that a project manager can use to keep track of the schedule. It is also possible to track the amount of time each person spends on each workflow step and thus measure productivity and compare against a budget.

- **Version Control.** One of the driving forces that started content management as a discipline is keeping track of the different versions of electronic files. How many times have you done a "save as" so you can keep a previous version of a document you're creating…just in case? How many times has your computer crashed before you saved the document that you'd worked on for hours? What happens when someone posts new content on the Web site that's just plain wrong, and you quickly have to get it off and get the old stuff back on before too many people see it? Finding which of many documents with various names is the most current isn't always easy. These are the needs for version control that have prompted organizations to come up with file naming conventions, to create elaborate backup systems, and to use the "track changes" features applications. An essential part of content management is to maintain the integrity of the content. With a CMS, this is done automatically with version control features that ensure that every iteration of a piece of content is saved so that it can be viewer or restored again later if needed.

- **Content Components.** Content components are the building blocks of publications. A component is a piece of content that can "stand on its own." For example, an *article* is a component; *basic product information* can be component; an *image* can be a component. Content doesn't just come in components. Often you have to create the content according to a set structure in order for it to become a component. Or, you need to take existing publications, and break them into their constituent components to make them reusable. Components are made up of *elements*, smaller pieces that make up the component, kind of like an atom is made of electrons and neutrons, and a nucleus. For example, an *article* component might be made up of a *title*, *subtitle*, *author*, *publish date*, *body text*—and some additional metadata, like *create date*, *owner*, *audience*, *technical level*, etc.. Because each component element is stored separately in the repository, content that is componentized can be easily reused across pages and publications. It's important to note that not all content lends itself well to componentization. Some content is better left as a whole document or a whole publication.



- **Templates and Content Formatting.** A limiting factor in dealing with large volumes of content is the amount of time it takes to lay out and format the content—especially when you want to use the exact same content in multiple pages or multiple publications—each with a different look and feel. The concept of templates is not new, but the real power of templates started to be



realized with the advent of dynamically generated Web sites—sites that draw their content from a database. This dynamic publishing is made possible through *templates*, standard page layouts with placeholders for the content. Templates are what bring economies of scale to publishing by automating the population of pages. But, publications created using templates have some limitations—the content must follow a set of guidelines and standards so that publication turns out right. The trick is to create templates in such a way that provides for a certain level of flexibility in the design, without turning layout back into a completely manual process.

- **Publishing Content.** One of the two main results of good content management (the other being making content easier to find) is making content easier to publish. Publishing means getting the content out there, making it accessible to the audiences. In the past, publishing typically meant *printing* on paper. Publishing today may mean putting a document on a file server and sending out an e-mail link to it. It may be creating a link to it on an intranet page. It could be copying the HTML files and associated images to the Web server. It could be sending it out as an e-mail "blast" to your customer list. Content management systems usually—but not always—include some mechanism for publishing content to its ultimate destination. One of the most exciting capabilities of content management is that it enables us to match content with customer data and publish "personalized" content to audience segments or to individuals.

- **Expiring Content.** An often forgotten aspect of content management is expiring content that is no longer accurate, relevant, or appropriate. Without any sort of content management, content in electronic form can be around virtually forever. Again, this is where guidelines come into play. For each type of content—or sometimes each specific document or content component—you need to determine at what point it should be reviewed and/or automatically expired. When you expire the content, you must also decide if you keep it (in an archive) or discard it. As the volume of content grows. The need for expiring old content becomes more and more critical. Old content just gets in the way of finding the new content.

If you look back at all of these facets of content management, you can see why content management is a discipline and not just a system. Just about all of the areas described above can be achieved without technology. In fact, many of these functions and pieces are handled today through manual processes—even in the largest of publishing organizations. But, it is with technology, a CMS, that the two biggest "bangs for the buck" of content management—efficiency and effectiveness—deliver significant returns.

Content Management Systems

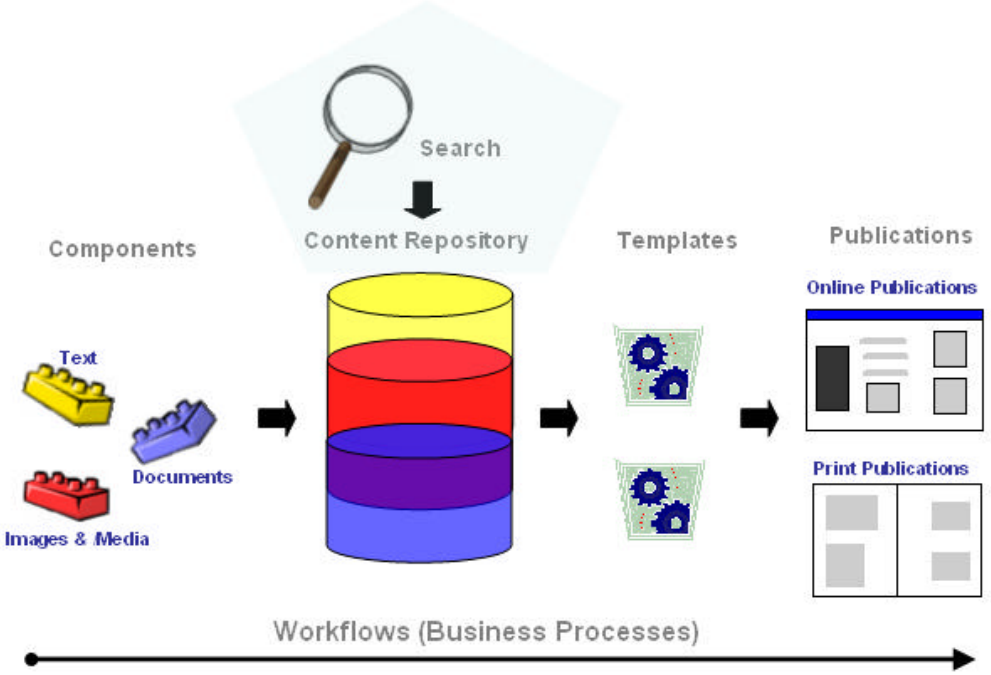
If content management is infrastructure for creating and maintaining content throughout its lifecycle, then:

CMS = The systems infrastructure for creating and maintaining content throughout its lifecycle

Software products called "content management systems" arrived on the scene in about 1995, so they belong to a relatively new software category. CMS software came about largely because of the need to manage content destined for the Web. As the software matured, the developers and marketers began to recognize that content management is not just a Web-related issue. However, even today, most of the CMS software products on the market focus primarily on managing content for the Web, with only a handful of products that deal with files destined for print.

For the most part, any full-featured content management system can handle all of the facets discussed above. In a nutshell, a good content management system allows you to efficiently create, acquire, store, access, publish, and retire content. The following illustration gives you a conceptual picture of what a content management system looks like:

Conceptual View of a Content Management System (CMS)

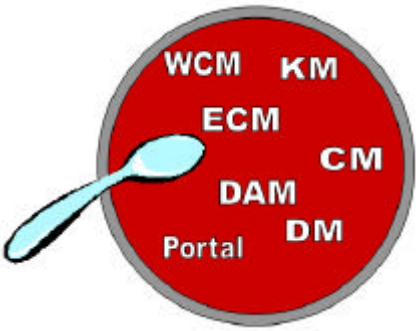


The idea is that you start with content—either by creating or acquiring it—and save it into a content repository. When the content is in the repository—with the proper metadata—you can search and **find** the content you need. Content moves through its lifecycle by way of workflows. You can also use templates to **publish** the content out to different formats including online formats and print formats.

Back in 1995, there was only one term: content management system (CMS). But, back then, without a clear definition of what *content* is, and what *content management* is, vendors of just about any kind of software product that covered at least a few of the facets, considered themselves to be offering a CMS. Functionality of the software ranged anywhere from managing documents, to managing programming code, to publishing Web sites, to creating HTML for Web sites, to storing images in a database.

Fortunately, over the past few years, a much more savvy group of people recognized that there are lots of different kinds of content management software products. I call them "flavors."

The "Flavors" of Content Management Software



The other way I describe all the different types of software products on the market today is "acronym soup"—since they all seem to go by acronyms:

- **CM = Content Management.** This term originally applied only managing and publishing Web content, but now it is the "catch-all" phrase to describe any system that manages any type of content.
- **WCM = Web Content Management.** WCM is the newer—and more appropriate—name for a system that is specifically designed to help you collect, store, and publish content to the Web. WCM systems are very good at allowing you break your content up into components and elements so that you can reuse content across multiple pages and multiple *online* publications.
- **DM = Document Management.** A more established discipline that has been around years, document management involves the organization, tracking, and search and retrieval of files (documents) on a file server. Metadata plays a key role in document management. Document management systems do not,

however have capability to *publish* content; their main purpose is to help people find information in the form of whole documents.

- **DAM = Digital Asset Management.** This is a relatively new type of system that is specifically designed to store, organize, and help people find images and media files. Digital assets have some unique qualities that require specialized software. For example, you will want to store multiple versions of the same image in different formats, but ensure that these images are linked to each other. You may have an original image that was created in Photoshop, but then save it to different versions—high resolution for print, and low resolution for Web. DAM systems also typically automatically generate thumbnails (small versions of the image so that you can easily browse and find them). Some of the newer DAM systems also offer dynamic image rendering capabilities that allow you to create the size and format of the image needed in your publications on the fly.
- **KM = Knowledge Management.** KM has been a popular term throughout the 1990s both in terms of software systems and as a discipline. Knowledge management refers to standards, tools, and processes for ensuring that the knowledge (the stuff inside people's heads) in an organization is documented and shared. Considering that the knowledge of a company's employees is a valuable corporate asset, it makes sense to ensure that this knowledge is not isolated and that it is not lost when the person leaves. KM generally deals with internal content (not customer-facing content), and the knowledge—once transformed into content— is frequently made accessible through an intranet site or a portal.
- **Portals.** The one flavor of content management without an acronym, a portal simply means a "door." It is a doorway into content or software applications. Technically, a portal is not content management at all. It's really a potential front end to a content management system—a way for people to browse the content repository, to view their content-related task list, or to view the status of content in a project.
- **ECM = Enterprise Content Management.** The newest term in the bunch is ECM. This is an umbrella term that encompasses all of the types of content management described above…and more. ECM strives to meet the lofty goal of managing all of the content across the entire enterprise.

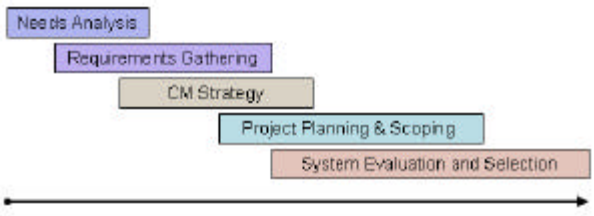
One of the most important aspects of selecting a content management software product is knowing precisely what the problem is that you want to solve. As you can see, different products specialize in different types of content. The products also vary in functionality. Some have very powerful and flexible workflow systems. Others may specialize in dynamically serving personalized content to e-commerce Web sites. While the vision of implementing a software system to help manage all of the content across your enterprise is an excellent one in the long term, it is best to first define one or more projects of limited scope and phase content management into your enterprise gradually over a period of months—or even years.

How a Content Management Initiative Works

Many a content management initiative begins by selecting a software package, but, as I've emphasized, the right place to start is to first understand the underlying problems. You need to identify a) what kind of content you're dealing with, and b) the problem you're trying to solve (in other words, what you want to *do* with your content).

On an enterprise scale, a content management initiative comprises several well-defined projects but is based on an overall strategy. A typical content management project might look something like this:

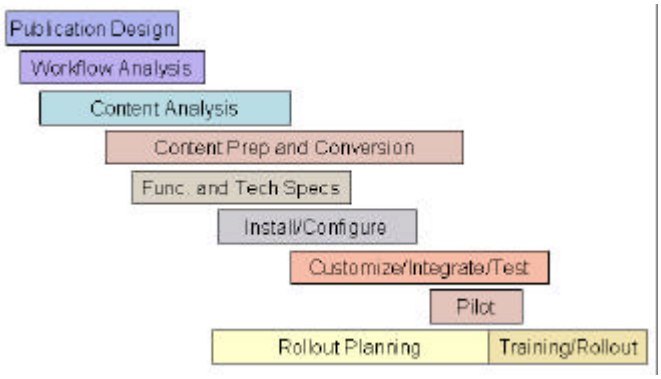
Phase 1: CMS Planning



In the initial phase, you start by identifying the problem you're trying to solve. One way to fully understand validate the issues is to perform a high-level needs analysis, which is typically done by canvassing key stakeholders from a number of different groups in your organization. Once you've established a general idea of what the needs are (e.g. the need for a better way for employees to find graphical images, or the need to update our Web site more frequently), you can take on a more detailed requirements gathering effort.

The requirements process will likely uncover other needs. The requirements also help you start define clear goals, and start planning the project—defining the scope, resources, and budget. It also sets the stage for selecting the software product or products that meet the defined requirements.

Phase II: CMS Implementation



There's a saying, "it's a lot easier to get somewhere if you know where you're going." This is especially true when implementing a content management system for publishing content. Start by designing the end product—the publications—you'll have a much easier time making sure that you collect all of the information about the content that you will eventually need—publication design.

Business process plays an important part in any content management system. This is where the workflow analysis comes into play—identifying the various roles within your organization that interact with content—seeing how they currently work, and defining ways for them to work more efficiently.

Content analysis involves identifying all of the *types* of content you will manage in your system. You might call these *component* types. For example, your content base might be made up of *articles*, *news items*, *tips*, *advertisements*, *graphics*, *video clips*, and so on. Once you know the content types, you can start to break them down into their constituent elements, and define what metadata is needed to track and publish the content.

Similar to implementing any other kind of enterprise application, you will write specifications, you will install and configure the software, customize, integrate, and test it. It's always a good idea to identify a "pilot" project—a mini-implementation that is limited in size and scope. A pilot project lets you test out your process and systems with less risk. Once the pilot is running smoothly, you can expand your rollout to the larger scope and train your users and let them start using the system in their daily work.

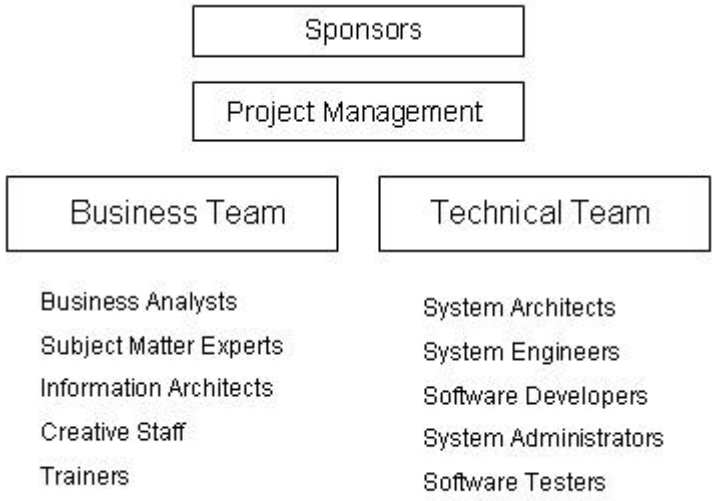
The implementation phase is where the rubber hits the road. A successful CMS implementation, like any enterprise initiative, requires careful planning and constant monitoring. If the upfront requirements were done correctly, however, you have a much higher chance for success. Likewise, having the right team with appropriate experience and knowledge is key to success.

The Content Management Team

Because content management initiatives almost always involve more than a single group within an organization, one of the most critical pieces of a content management initiative is to have one or more sponsors. A sponsor is typically at an executive level and sets the vision for the outcomes of the initiative.

Strong project management is a key success factor in any endeavor that involves multiple groups and a multi-disciplinary team made up of both business and technical resources.

A CM implementation team typically looks something like this:



This team may be made up of all internal resources or, more likely, a combination of consultants, third-party integrators, software vendors' professional services teams, and contractors. The team comes together for the planning and implementation, stages. Once implemented, some of the team members move on to other things, but still a core team stays to run the system and ensure that the content is well managed.

The End Goal—Well -managed content

The definition of success of a content management initiative is that producing content is more efficient and the content itself is more effective than it was before. When done right, content management can deliver significant returns (ROI). Many content management initiatives have paid for themselves in a matter of months. While not easy, and still a maturing discipline, content management is here to stay. And the time organizations take now to learn the discipline and develop the systems to manage content is well worth the effort and the investment.

About the Author
Rita Warren is president and principal consultant of ZiaContent Inc., a Seattle-based firm specializing in content management planning and implementation. E-Mail: rita@ziacontent.com.
Web Site: www.ziacontent.com.