What is Content Management?

Content Management Defined

To define content management, we need to define:

- What "content" is, and how it differs from traditional "data"
- What it means to "manage" content, and how this is different from traditional "data management"

What is Content?

The concept of structured vs. unstructured data…

Structured data fits neatly into well-defined buckets, and is the first area to have been attacked by I/T because of this simplicity

Structured data, however, is only the "tip of the iceberg" for the full set of critical information surrounding business applications; the remaining "unstructured" data, which does not fit so predictably into well-defined buckets, has become known as "content."

Examples

Business Process	"Structured" Data	"Unstructured" Data
Sales	Contact Information, Sales Pipeline and Projections	Sales Collateral, Cover Letters, Proposals, Contracts, RFPs
Marketing	Prospect Lists, Product Numbers and Prices	Brochures, Specifications, FAQs, Print Catalogs, Online Catalogs, Print Advertising, Web Banner Ads, Direct Mail Pieces, Customer Newsletters, Email Campaign Templates
Production	Bills of Materials, Inventory Levels	Engineering Drawings, Process Specifications, Technical Specifications
Customer Support	Customer Lists, Phone Logs, Contact History	FAQs, Customer Correspondence, Troubleshooting Guides, Cross-Sell Offers, Knowledge Base Articles
Purchasing	Vendor ID, Item Number, Price, Discount	Purchase Order Terms and Conditions, Vendor Contracts, Product Specifications, Vendor Catalogs, Bills of Lading
Human Resources	Employee Lists, Payroll and Benefits Information	Benefits Plans, Stock Option Plans, Employee Policies, Resumes, Performance Reviews,
Finance and Administration	General Ledger, Financial Projections	Corporate Annual Reports, SEC Filings, Board Minutes, Compliance Reporting, Accounting Policies

Many special kinds of content are associated with specific industries. For example:

- **Commercial Airlines / Airframe Manufacturers / Suppliers** (aircraft maintenance manuals, flight operations manuals)
- **Automotive** (dealer service manuals)
- **Process Manufacturing** (standard operating procedures, material safety data sheets)
- **Insurance** (insurance policies, insurance claims)
- **Public Utilities** (utility contracts, regulatory filings)
- **Banking and Financial Services** (loan applications, equity research, mutual fund reports)
- **Healthcare** (medical transcriptions, electronic patient records)
- **Telecommunications** (telecommunications practices, network and wiring diagrams)
- **Pharmaceutical** (new drug applications, FDA submissions, clinical research data)
- **Legal Publishing** (court records, legal briefs)
- **State and Local Government** (birth certificates, public health records, corporate filings, committee reports, legislative transcripts and calendars)

How is Content "Managed"?

Data management is relatively straightforward, whereas content typically goes through a much more complex author-edit-publish-update cycle.

As the following table illustrates, content management is significantly more complex than management of structured relational data:

Action	Data	Content
Create	Created automatically by applications or manually via a forms-based interface	Requires creative skills and often collaboration between multiple contributors
Review and Edit	If manual review is required, normally a quick double-check via a forms-based interface or audit report	Requires a complex iterative cycle in which multiple parties make comments and annotations that are factored into the next updated version
Link to Related Information	Through foreign keys and/or relational JOIN operations	Requires a combination of cross-referencing, hyperlinks, metadata, and "virtual document" parent-child relationships
Format and Deliver	Typically handled through standard reporting tools, Visual Basic interfaces or ASP/JSP tools on the Web	Requires complex formatting specifications and transformations between file formats, often with the need to support multiple renditions of the same content in different delivery formats
Update	Typically handled at either a field or record level in a well-defined application environment	Changes may occur at any level (e.g. a word or phrase vs. an entire chapter, etc.), requiring complex change management including both version control and more detailed revision control to track the specific items that were changed

Action	Data	Content
Index	Handled through a well-defined relational schema	Requires a combination of structured hierarchy (e.g. cabinet-folder structure) and flexible relational metadata, together with dynamic "where used" analysis for shared content components
Search and Retrieval	Typically handled though SQL queries using the defined relational schema	Often requires a complex combination of metadata, full text and structural elements, and sometimes even more exotic techniques such as Query-by-Image-Content

What Makes Content Management Difficult?

Bottom line: content is different than data, and content management presents a unique set of challenges beyond those encountered in data management. Some of these include:

- The inherent flexibility and unpredictability of content
- Lack of well-defined, industry-standard application infrastructure for handling content
- The need to integrate many products each designed with its own purpose and point of view
- Complex creation, update and change management cycles
- Complex reuse and repurposing issues
- Complex cross-referencing and indexing schemes
- Complex formatting and transformation requirements
- Complex search and retrieval issues

A Brief History of Content Management

Content has existed for at least 5,000 years, since the invention of written language. Formal content management, however, probably didn't begin until the founding of the Library of Alexandria in 150 B.C.

For at least the last 100 years, content has been playing a big role in business, in the form of brochures, catalogs, contracts, correspondence, invoices, purchase orders, billings and so forth. However, for most of this period "content management" has been mostly a matter of physical file folders and file cabinets.

As an I/T initiative, content management was virtually ignored until imaging technology in the 1980s made it possible to replace large-scale filing (such as in insurance claims processing) with more cost-effective electronic image repositories. These early applications evolved to include the concept of workflow, in which the images could be moved around electronically to the computer equivalent of "in" and "out" baskets.

Then, toward the end of the 1980s, we hit an era of severe "information overload" as electronic publishing suddenly accelerated content production at an exponential rate. The problem was compounded by personal computers, whose widespread use took away most of the ability for I/T to exert any centralized control.
As the 1990s dawned, personal computers were increasingly becoming linked by local area networks, and centralized client-server applications began to arise. With the realization that this provided a means to re-establish control over electronic content, the age of document management was born.

By the mid-1990s, the worlds of document management and imaging / workflow had begun to converge, and vendors of these products were beginning to experiment with a new phenomenon – the widespread adoption of the Internet.

By 1998, the Web had evolved from an interesting phenomenon to serious business, and was now composed of literally billions of individual Web pages. Suddenly "document management" began to go out of vogue, and "web content management" became the central focus.

The Web frenzy hit its crescendo in 1999, but with the dot.com and NASDAQ crash in the year 2000, attention has again turned to a more balanced combination of print and web-based content, together with increasing focus on wireless devices, audio clips, streaming video, and other new forms of electronic content. Also, while the rush to B2C e-commerce has slowed somewhat, there is now a renewed focus on automatically communicating electronic business content through XML-based B2B commerce networks.

Today many variations of content management exist, all based on the same principles but each with a slightly different purpose in mind:

Variation	Business Purpose	Example
Web Content Management	Ensure that complex Web site content is complete, up-to-date and properly linked	Managing all the content behind the Amazon.com Web site
Knowledge Management	Archive and index critical organizational knowledge so that less experienced employees can take advantage of it	Extensive knowledge base used by service technicians at a telecommunications company
Document Management	Manage complex document-based information so common elements can be reused, and documents can be dynamically assembled for publishing	Management of overlapping and constantly changing information in automobile user manuals, dealer service manuals, and technical specifications
Imaging Management	Replace costly and error-prone paper processing with electronic storage and workflows	Insurance claims processing
Digital Asset Management	Allow a mass of multi-media electronic content (photos, audio, video, etc.) to be stored in a way that the proper content can be easily found	Finding artwork for developing advertising creatives, archiving news video clips at CNN
Records Management	Ensuring that critical records are secure but accessible, and are deleted when they should be	Management of required documentation at a nuclear power plant

The Role of XML in Content Management

XML is a way to take documents that can be viewed by people, and turn them into data that can be processed by computers.

XML blurs the distinction between structured and unstructured data, allowing data items buried inside an unstructured document to be explicitly tagged.

XML plays at least three key roles in content management:

As a **source format for content publishing,** XML allows:

- Multiple presentation formats and/or media types (e.g. Web, paper, and wireless) to be supported from a single source repository.
- Smaller fragments of text, graphics, etc. to be managed as reusable information objects that can be dynamically assembled into final-form documents or Web pages.
- Direct integration of document-based data (e.g. part numbers, customer numbers, diagnostic codes, etc. embedded in text) with core business systems and databases.

As a **delivery format to the web**, XML allows:

- Straightforward aggregation of corporate information into enterprise information portals.
- Interchange of complex, application-specific data in business-to-business e-commerce applications (e.g. airline technical data, semiconductor data sheets, etc.).
- Rich presentation formats that can be individually tailored and controlled by the information consumer (using their own style sheets and Java applets).
- Client-side processing of web data that can be individually defined and controlled by the information consumer, and which can allow public web data to directly interface with local applications (e.g. automotive technical specifications that can "talk to" local dealership parts inventory systems).

As a **universal data interchange format**, XML allows:

- Flexible interchange of trading partner agreements and business "documents" across Web-based B2B partner trading exchanges (PTXs)
- Flexible interchange of document metadata across systems and to support federated searches.
- Integration of document metadata (e.g. customer who sent correspondence) with related relational data (e.g. customer database).
- Flexible means to specify input and output data containers in data-centric workflow.

How Content Management Fits Into Other IT Initiatives

eCommerce

- Web Site Content – managing the creation and update of all content used on the web site.
- Catalog Management – managing the creation and approval of content used in online product catalogs; automatic assembly of personalized catalogs based on matching customer preferences with content metadata; synchronization of online catalogs with printed catalogs.
- Enterprise Applications Integration (EAI) – providing a rich, neutral format to interchange business "documents" (sales orders, invoices, purchase orders, contracts, billings, etc.) between e-Commerce systems and other legacy enterprise applications.

Corporate Information

- Intranet Content – collecting from disparate sources, aggregating, and managing the corporation's digital assets to avoid duplicating previous work
- Enterprise Application Access – combining data from enterprise systems (ERP, CRM, etc.) with document templates to create information without the requirement for employees to know each enterprise application

Customer Relationship Management

Marketing Automation

- Campaign Management – managing the creation and approval of content used in direct mail campaigns; automatic assembly of personalized direct mail pieces based on customer profiles matched against content metadata.
- Personalized Customer Communications – managing the creation and approval of content components used in customer newsletters and other regular communications; automatic assembly of personalized documents out of standard components based on customer profiles matched against content metadata.

Sales Force Automation

- Proposal Generation / RFP Responses – managing the creation and approval of proposal boilerplate, templates, cover letters, product specifications, etc.; automatic assembly of proposals and RFP responses as virtual documents; reuse of document components for multiple purposes
- Personalized Sales Collateral – managing the creation and approval of sales collateral content; automatic assembly of personalized sales collateral out of standard components

Customer Service / Contact Center Automation

- Knowledge Base – managing the creation and approval of articles used in the agent knowledge base; replacing the weak knowledge bases used in most CRM packages with a more powerful knowledge server; automatically linking the agent knowledge base to personalized marketing and sales collateral so agents can see what the customer has received.
- Cross-sell / Up-sell Material – managing the creation and approval of content used in cross-sell / up-sell offers presented to agents in the call center; integrating content management with personalization software used to serve up the appropriate cross-sell / up-sell offer.
- Email Templates – managing the creation and approval of email templates and standard email boilerplate (especially important in financial services and regulated industries where email communications have significant legal ramifications and must be carefully controlled).

Supply Chain Management

- Trading Partner Agreements – managing supplier business agreements and ordering processes.
- Technical Specifications – managing the set of specifications used to drive supplier requirements.
- Technical Documentation – creating composite technical documentation out of the individual technical documentation provided by each supplier.