Component Content Management in Practice

Meeting the Demands of the Most Complex Content Applications

Bill Trippe
February 2005

X·HIVE

Executive Summary

As the market for content management technology continues to grow, so too do the ways in which organizations seek to use content management. What began as a market focused on web content management has grown to include document management, digital asset management, and records management. What has emerged along with this growth is the use of the umbrella term *Enterprise Content Management* (ECM) to describe a broad, enterprise-class platform of content management technology that can handle all kinds of content.

Yet, despite the best efforts of the ECM vendors to provide an all-encompassing platform for enterprises, specialized vendors continue to focus on challenging, specific content management applications. One specialized area for content management is the need for some organizations to manage large volumes of content that is used to support complex products before and after the products are sold. Examples include auto and truck manufacturers, airlines, and airplane manufacturers. In these cases, highly complex products are supported by lengthy, complex, and voluminous technical documentation, parts lists, troubleshooting guides, and maintenance manuals. Along with the sheer volume of information, this kind of application is marked by the requirement for the content to be kept up to date over many years, for it to be published in many physical formats, and for the content to be of the highest technical quality and accuracy.

Professionals who create and manage this kind of product support content understand these challenges well. These content professionals also understand that the nature of this content lends itself to significant automation. In many of these applications, there is a significant opportunity for organizations to develop processes and systems whereby core components of content can be managed in a way that they can be reused in many content products. Some recent research shows that as much as half of product support content is redundant and could be reused in this way. For a large organization, a systematic program of reuse could yield significant savings, efficiencies, and quality improvements over time.

While many content management systems do enable content elements to be reused in various ways, the kind of reuse application we are discussing here demands a specific kind of approach to content management. We introduce the term *Component Content Management System* (CCMS) in this paper to describe the kind of flexible management of components required. A CCMS must manage components at a fine granular level, in ways that allow the components to be easily used, reused, versioned, linked, assembled, and reassembled into different content products.

This white paper looks at the requirements of component content management in some industries and vertical markets. It compares the requirements of component content management with the capabilities of more general content management technologies, notably web content management and document management. It then looks at the technology behind CCMS in depth, and concludes with example applications where CCMS can have the most impact on an enterprise.

Component content management systems provide truly flexible reuse, with components at any point in the content hierarchy easily created, updated, managed, combined, recombined, and linked. But CCMS technology is also significant because it enables organizations to adopt reuse over time, because CCMS provides the tools for ongoing refactoring and conversion of content.

Introduction: What is Component Content?

Content management is a maturing technology, with installations of commercial, open source, and custom applications in thousands of organizations. Web content management alone is a substantial business, with a marketplace of products that range from low-end, personal publishing tools to enterprise platforms from companies such as EMC/Documentum, Interwoven, and Vignette.

In addition to growing as a business, content management has also expanded in technical and functional scope. While the earliest content management systems were often focused on Web content delivery, the term "content management" has broadened to include the management of all kinds of enterprise content. This includes compound documents, document images such as scans and facsimiles, electronic forms, email, and digital assets such as audio and video. Early technology for these different kinds of content tended to be specialized or dedicated applications, and many best of breed applications emerged for each category. More recently, though, as the scope of content management has expanded, the platforms for managing content have expanded in functionality as well. The past three years have seen acceleration in mergers and acquisitions, as the leading content management vendors have attempted to fill out their product portfolio with technologies that manage more and more content types. This led to the common use of a new, broader term, "enterprise content management," or ECM.

Enterprise Content Management

Web Content Management

Document Management

Digital Asset Management

Records Management
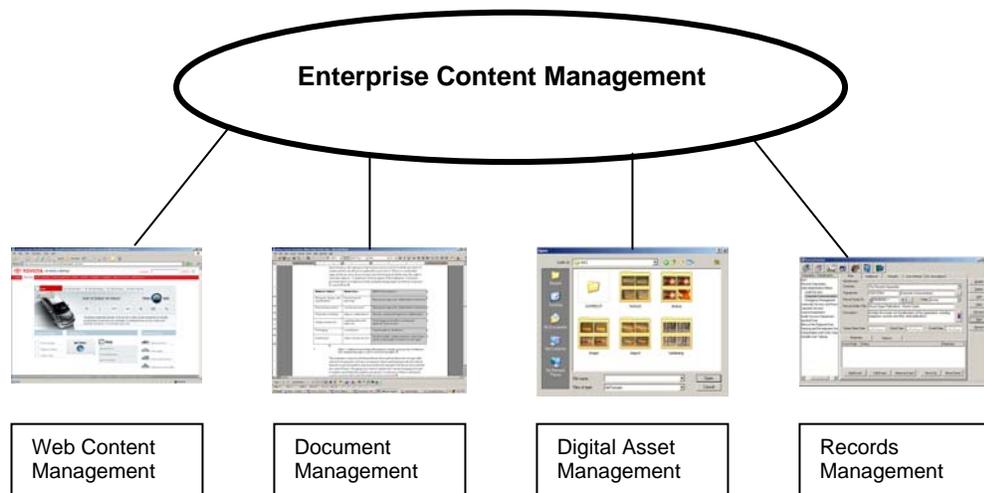
Figure 1. As leading content management vendors have expanded out their product portfolio with technologies that manage more and more content types, a new, broader term has become more common: enterprise content management, or ECM.

In addition to expanding the types of content that are now managed, ECM has also ushered in a recognition that content management is a peer application to other kinds of enterprise applications, such as enterprise resource planning (ERP) and customer relationship management (CRM). This has in turn led to an increasing focus on how content management systems meet the architectural requirements of larger enterprises. The major content

management vendors—Filenet, Documentum, and others—are now in a marketplace targeted by the major platform vendors like IBM, Oracle, and BEA.

At the same time that large enterprises are pushing ECM vendors to work with enterprise systems and databases, different requirements within content management continue to grow. For example, records management (RM) requirements are being driven by compliance and governance issues such as Sarbanes-Oxley and HIPAA, while digital asset management (DAM) is increasingly focused on managing the digital assets such as video, compound documents, and high-quality images used with product marketing and brand management. These expanding requirements have come about in recognition that, while content management technology must be open and well designed for integration, complex business processes need to be addressed.

Complex Products and Their Complex Content

One complex business process involves technical documentation, catalogs, and other kinds of content used in product support, where reusable content "components" can be managed in a way that complex content products can be assembled, reassembled, and published. This kind of application is especially critical in markets such as manufacturing, electronics, aviation and other transportation, and the military.

Complex products often require content necessary for a user to operate and maintain the product after it is sold that can include user manuals, maintenance manuals, service updates and advisories, and the voluminous content that is often produced in marketing the product before it is sold. For a complex manufactured platform such as an automobile, truck, or aircraft, it is not unusual for the product support content to number in the thousands of pages. Indeed, apocryphal or not, there's more than a grain of truth to the claim that Boeing aircraft documentation weighs more than the aircraft itself.

Figure 2. Applying CMS principles to repurposing of content provides strong ROI potential for an enterprise.

Complex products in fact do require equally complex and voluminous content, especially products like aircraft and trucks, which have long shelf lives and many systems, subsystems, and individual parts that require ongoing maintenance and support. A commercial aircraft, for example, can remain in service for thirty years and longer. The capital investment made by the aircraft buyers is fully realized when the aircraft is properly maintained by following a comprehensive, long-term—and content-intensive—maintenance program.

Indeed, in some industries, content-intensive processes are intrinsic to how the products are both developed and operated. Consider the pharmaceutical industry, where virtually every detail of a drug in development is documented and provided to the US Food and Drug Administration (FDA). Airlines provide detailed maintenance records to the Federal Aviation Administration, trucking companies to the US Department of Transportation, and power plant operators to the US Department of Energy. For reasons of operational efficiency, productivity, and regulation, many enterprises find themselves in the business of creating, maintaining, and archiving vast repositories of complex content.

Because many of these enterprises operate at a large scale, the content efforts themselves present opportunities for efficiencies. Most large-scale documentation efforts rely heavily on electronic distribution of their content for cost-savings, but even where "single source" publishing has been implemented successfully, the long-term cost savings may be minimal. Much of this content still needs to be produced on paper, and the requirement for electronic distribution itself can have significant cost associated with them.

Many enterprises have the volume of content that offers re-purposing opportunities. The practice of producing multiple media formats—print, online, CD-ROM, HTML—from a single-source is well-established, but enterprises face even greater volume of content across more and more versions, channels, markets, and languages. As the enterprise's requirements for multi-channel publishing expands, the enterprise must invest in platform architectures that can efficiently automate these processes.

Content as Components

Organizations are learning that the real savings may come in being able to actually *reuse* materials, and not just *repurpose* them from one format to another. For example, in an automotive application, several different manuals may well include several of the same procedures. Why not create the content in such a way that the reusable procedures can be created once, edited once, and stored once in a format that allows it to be *reused* in many different manuals and other content products?
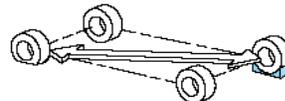

Figure 3. A procedure common to a wide range of a manufacturer's products could be a good candidate for efficient re-use. For example, instructions for or certain safety warnings about removing tires may be the same across an automobile product line. (Source: Ford Owners Manual, 2000 Taurus)

Enterprises that have solved the multi-channel distribution problem with single-source publishing also have often identified content reuse as a primary goal. Single-source publishing involves repurposing content into other formats, but when a content component is used in

multiple content types, it is called reuse. Examples of content reuse would include safety warnings common to many repair procedures, instructions for maintaining machinery that share particular parts or sub-systems, and step-by-step descriptions of processes in a software application as might be represented in the reference manual, online Help, user's manual, programming guide, and Web tutorial.

In fact, some recent research suggests the opportunity for reuse may be significant, especially in certain businesses. In developing a new product to analyze and mitigate this problem, the company Data Conversion Laboratory systematically looked at corporate and government content and concluded that typically 50% of product-related documentation is duplicated. This was true over a range of industries, including aerospace, pharmaceutical, and defense. In some of the documentation sets, the level of redundancy was dramatically *higher*, with one set of aerospace documentation having over 83% redundant information. Identifying the redundant pieces that make up the content can be thought of as one means of identifying content "components," and is central to the reuse strategy.

The key to the successful re-use of content is to manage it at a granular level. These grains of content—components—can be shared, reviewed, updated, or combined and compiled into different document aggregations and collections. Each component can be separately edited and re-used, and workflow processes enforced. Content components can have their own lifecycles and properties—version, owner, and approval—that support fine-grained reuse and the ability to track such usage.

In a later section, "Component Content Management in Action," we will discuss a complex application of reuse in the airline industry.

Broadly speaking, many CMSs manage content as components at some level, even though there are some—simpler document management platforms, for example—that do not go beyond storing documents as images or as unitary files. Typically, Web CMSs manage content as articles, heads, links, and other "chunks" associated with metadata within a database that enables these content components to be presented within a browser in various combinations. But the content structures supported by such CMSs are usually for simple, short documents, not for long, complex documents, and certainly not for long complex documents that have complex internal relationships among the components.

Do You Need a Component CMS?

Traditional content management systems focus primarily on managing content components in concert with managing the metadata associated with the components, where the metadata is used to efficiently retrieve the whole document or some component of the document.

Of course, the desire to reuse content and the ability to reuse it are two separate things. The reality of most enterprises is that the kind of content in question—product support documentation and related information—often resides in data formats that are not conducive to the kinds of information access and sharing that reuse dictates. For example, longer documents are often in proprietary and binary formats such as Microsoft Word and FrameMaker, while some of the technical data may reside in databases for applications such as enterprise resource planning. There is often not a straight line from the proprietary formats to a more general structure that would support reuse.

In some cases, enterprises that have gone to single-source publishing have also gone to a format-neutral data structure such as the eXtensible Markup Language (XML). In fact, some of the aforementioned industries, especially aviation and the military, were early adopters of Standard Generalized Markup Language (SGML), the predecessor to XML.

The MRU Question

In addition to providing a format-neutral data structure, XML encoding of the content also can support reuse by providing a ready mechanism for dividing the content into logical elements or components. For example, a repair manual for a transmission could be divided into the various tasks that can be performed, a parts catalog could be divided by part number, part description, and so on. Analysts and consultants have used terms like "minimum reusable unit" or MRU to describe the level at which content is logically stored for maximum efficiency. This MRU will differ from enterprise to enterprise, but the decision is largely driven by practical considerations such as how the content is written, edited, updated, distributed, and shared with partners.

Most content management systems try to solve the re-usability challenge by "shredding" or "chunking" documents into predefined components that are managed separately. For example, a hardware manufacturer might logically chunk its maintenance documents into components that handle a single task—removing and replacing a part. A software manufacturer might chunk its user manual into components that handle a single function—printing or deleting a file.

In practice, this reliance on chunking content into MRUs has both advantages and drawbacks. Done well, the organization can end up with a high-performance system for creating, updating, managing, and publishing its content. However, the requirement to determine an MRU upfront results in some tradeoffs—what is the best unit at which to edit the content? Are there applications where a different chunking level would be advantageous? Probably. Will new requirements emerge that would be better served by chunking the content in different pre-defined components? Almost certainly.

Breaking up documents—shredding—can make document components available as fine granules, but once the level is set and applied, changing the granularity of the content is difficult. In many of the systems that require this upfront MRU analysis and decision, the setup of the MRU level and the associated tools cannot be readily changed. In some cases, changing the shredding level requires the developer to export the content out of the database,

redesign the schema of the database, and then re-import the data, and perhaps only after transforming or conditioning the data to match the new database schema

In addition, pre-defining content components, or MRUs, may make it very difficult or inefficient to address any fragment of content within the chunk. Finally, the application will likely be integrated with and supported by numerous tools that have been developed to deal with content components at one level in the hierarchy and not another. Making these tools a fit with another content component level may be difficult or impossible without substantial redesign, redevelopment, and redeployment of the tools.

We see a significant difference between these systems for managing component content that rely on pre-defining components and a new breed of systems that handle XML-encoded content in a much more flexible and efficient manner. A key component in this new breed of systems is the ability to store and manage XML as *a native data type*, and not force the XML to be stored and managed in a system such as a relational database system. Indeed, much of the shredding and chunking behavior that we have been discussing is tied directly to the need of these RDBMS-backed systems to shred the XML data structures into the table and row structure best managed by relational data bases. This shredding is never a perfect fit— practical decisions need to be made—and the MRU decision is just one of them. Moreover, this shredding must happen each time the data goes into or out of the database, which can result in performance challenges.

Component content management software provides flexible, high performance access to content components by storing XML in its native format, by not requiring the shredding process to import and export the content, and by using standards-based tools for access to the XML, including standard APIs such as the Document Object Model (DOM) and query tools based on XQuery. The table below discusses some of the requirements of a CCMS in more detail.

Content Refactoring and Migration

So reuse presents a significant opportunity for organizations to create content more efficiently and with a better eye to detail, consistency, and quality control. Yet the reality of many organizations—especially the larger ones who could most benefit from reuse—is that there are large volumes of heterogeneous content that need to be digitized, converted, and normalized.

With technologies dependent on managing MRUs, organizations have to make some hard and fast decisions about how the content is going to be digitized, converted, and normalized prior to use. Organizations have traditionally invested a great deal of time and money in an up-front analysis at this point in the process—designing the data structures that they believe will support their content creation, management, and publishing needs.

In practice, however, organizations cannot really anticipate all of their ongoing needs. Especially in larger organizations, the content is too lengthy, complex, and variable to understand every detail—up front—of the needed data structures and tool. As organizations begin to work with the converted content, they begin to see requirements to transform the content—sometimes subtly and sometimes more radically—to make it usable and reusable.

However, if this content has been introduced to a system that requires an MRU design decision upfront, the process for now converting this content can be cumbersome, expensive, and time-consuming. In many cases, the content needs to be extracted from the system and converted, and then the content repository has to be reconfigured and in some cases redesigned to accommodate a different MRU structure. Depending on the technology, underlying tools that support the content are sometimes designed around the existing MRU structure. These too may need to be redesigned. Then the content needs to be imported back

into the repository for continued processing. Painfully, this process of exporting, converting, and redesigning the content data structures—and underlying repository and tools—will likely need to be repeated over the lifetime of the system.

A better approach would be to have flexible component content management technology that does not require the MRU decision up front—that allows for the content to be stored flexibly, with access to any node in the repository. Ideally, the CCMS would also provide tools for refactoring and converting the content in place. This way, the organization can incrementally convert and normalize content for inclusion in the CCMS, test it, work with it, and enhance it over time. Such an approach is much more manageable, and much more in tune with how organizations work with content.

Component Content Management in Depth	
Repository Functions	Roles-based access control for check-in, check-out
	Management of both content and metadata
	Reuse at any level
	Enable components to be freely moved around, with link integrity guaranteed
Linking, Link Integrity and Management	Components managed through all kinds of links, sources are managed, multiple links and link types are supported. Standards like XInclude and XLink are supported.
	Automated consistency checks
Content and Document History, Versioning	All content components versioned and complete history tracked
	Any individual component or collection of components (document) at any version can be reproduced
Workflow	Task Management
	Role Based
Content Lifecycle	Central storage for schemas
	"Impact analysis" for possible changes to schemas, content
	Instance modifications using style sheets
	Metadata can be handled, filtered at the same time
	Content can be enhanced, refactored over time within the repository
Conditional Publishing	Publish variant versions of content based on variables, schema constructs, abstract links
	Examples: publish based on dates, variables in the text (product name and numbers), varying schemas, varying stylesheets, varying based on the top-level component

	Images stored separately, and can be generated or manipulated at publish time
General Requirements	Scalability
	High performance
	Flexible and comprehensive security
	Maintainability
	Low total cost of ownership

There is tremendous potential for reuse in product support applications. As the research from Data Conversion Laboratory shows, organizations have an opportunity to gain real efficiencies in content development and deployment. We think that CCMS technology can be the driving technology for reuse.

Component Content Management in Action

Complex products like airplanes have voluminous, complex content that supports the products. This product support content is mission critical: if done well, this product support content can lead to increased up time, proper operation, increased safety, better compliance, and maximized investment in the aircraft itself. Boeing, Airbus, and other aircraft manufacturers deliver huge volumes of product support content to their customers, but the content management problem doesn't end there. The buyers of these aircraft—the airlines and cargo companies—are active users of this core content who then create substantial custom content to support their ongoing operations.

Airlines and their global travel partners serve thousands of destinations in hundreds of countries, often through multiple hub cities and maintenance facilities across the world. The leading airlines each employ many tens of thousands in fleet operations, and some operate fleets of aircraft from different manufacturers. For example, they may have dozens of Boeing aircraft with engines from GE, and several dozen more aircraft from Airbus with engines from GE and Rolls Royce. This need to incorporate third-party data within their operations adds to the content management challenge for airlines.

Airlines Content Challenge

Not only is aircraft maintenance content lengthy and voluminous, but also this product content is complex in a number of ways:

- ❑ The content is technical in nature, requiring expensive and expert creation and review.

- ❑ The content is technical in nature, so it is complex to publish (e.g., charts, tables, equations.)

- ❑ The content contains specialized formats (e.g., procedural task lists, cautions, warnings and notes, troubleshooting charts)

- ❑ The content is tied to regulation, so additional review, overhead, workflow, considerations must be undertaken

Great volumes of information come in from the airframe manufacturer (Boeing, Airbus, etc) and the engine manufacturer (GE, Rolls Royce). Information is constantly changing to reflect changes to the systems, subsystems, and individual parts and materials used, as well as to procedures because of changes in regulation and model configurations. For example, Boeing may be now advising airlines to perform a certain test or modification on a certain system or part based on something changing in the factory, or being changed because of regulation or safety concern from the FAA.

However, information as delivered by the manufacturers may or may not be suitable to be put to immediate use by the airline itself. The airline usually needs to take the manufacturer's information and develop more specific instructions based on many different factors, including:

- ❑ The airline's particular equipment and configurations

- ❑ The facilities and locales in which these procedures will take place
- ❑ The specific personnel who will perform the tasks and perform the inspections, either because of seniority, or levels of qualifications, or certifications required to perform certain tasks

What do these changes in the manufacturer's content look like at the airline level? Here are a couple of examples out of tens of thousands of possible iterations and conditions:

- ❑ An engine maintenance procedure done as part of a maintenance overhaul may presume that the entire engine compartment has already been exposed, whereas a specific maintenance procedure may begin with the steps the mechanic has to take in order to expose that particular part.

- ❑ An airline may have made a modification that is not yet reflected in the general maintenance information provided by the manufacturer.

- ❑ The manufacturer may have a modification left to the airline's discretion, which the airline has chosen not to implement, yet the general maintenance information already shows the modification.

Larger airlines have in-house engineers, mechanics, and technical writers to create this more custom content. Smaller airlines will do some of this work in-house, although there are also service companies that publish learning materials, charting and navigation documents, and related products for the aviation and maritime industries.

When an airline's in-house content departments create manuals, engineering overviews, and other maintenance and repair materials, they face a staggeringly high volume of complex content to customize. Fortunately, the content used in an airline's maintenance efforts contains much redundancy, making automation and re-use strategies attractive.

Automation and Reuse are Critical

A lot of the information airlines must manage is redundant—indeed, airline maintenance content has been shown to be as high as 80% or so redundancy. Many subtasks are done as part of many larger tasks (the same panels are removed; the same screws are tightened and loosened; the same lubricants, etc., are applied; the same cautions, warning, and notes are issued). Simply considering the high redundancy rates make reuse attractive, but when issues of quality, safety, and integration of different operations are also considered, reuse makes even more sense. If the technical writer, engineer, mechanic, inspector, manager, and the FAA have all signed off on a two-page procedure, for example, there is compelling cause to declare the procedure the one to use and re-use, until it is out of date or otherwise superceded.

Of course, there are many triggers that might prompt an airline to update or create a new procedure. There are the "upstream" revisions or additions from the manufacturers, but internal revisions in tasks can come from the maintenance workers themselves, on-site changes in procedures that are more effective than those provided by the manufacturers, or tasks specific to an aspect of customization undertaken by the airline. The key undertaking for airlines, when pursuing a re-use strategy, is to identify content components and documents that lend themselves to re-use. Here are some of the instances:

- ❑ Highly specific maintenance program tasks
- ❑ Manuals that describe the task at a level of detail suitable for mechanics in the field
- ❑ Self-contained tasks that can be accomplished in known times
- ❑ Tasks that typically represent work that can be accomplished in a single shift
- ❑ Different manuals for mechanics, cleaners, inspectors

❑ Documents that can be customized for personnel with different levels of experience, different certifications

❑ Maintenance tasks (and the corresponding documents) based on safety constraints that can be more stringent and complex for particular situations such as long-haul planes that fly over water
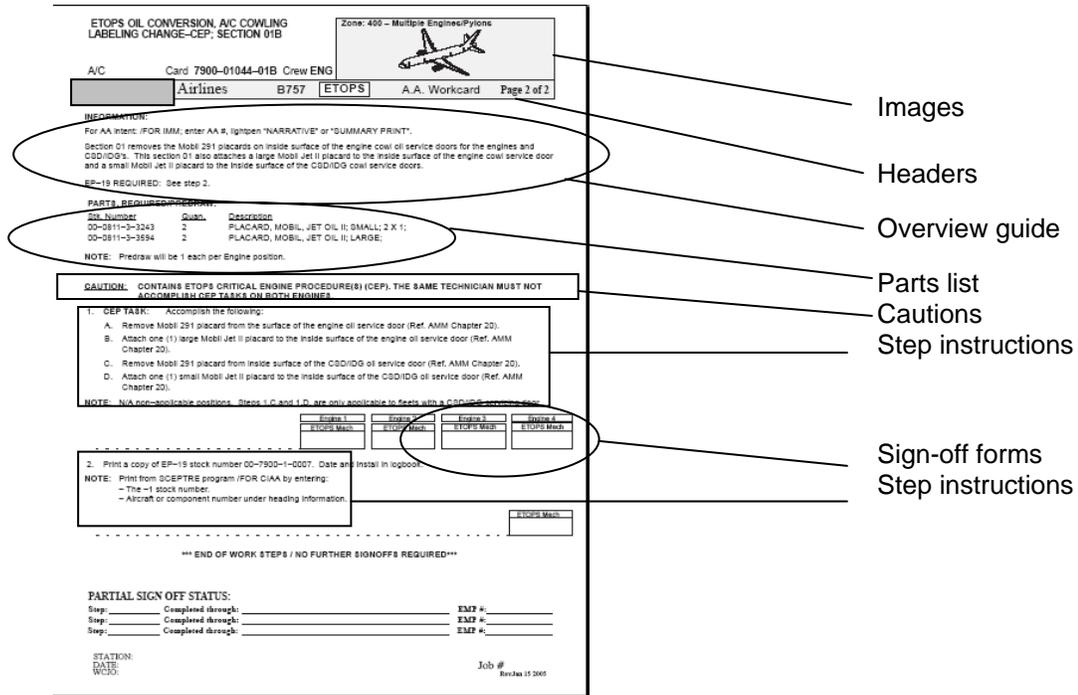

Figure 4. Much of the content that goes into airlines' maintenance instructions appears in many places— redundancy—and so can be re-used. Linking these content components into the document allows efficient re-use, especially when there is a high volume of documents.

Furthermore, the scale of operation for airlines is huge, further motivating the efforts to re-use content:

❑ Tens of thousands of procedures for a given aircraft

❑ Many thousands of procedures on the various components or major sub-systems that are part of the aircraft (e.g., engines)

❑ Source manuals from the manufacturer that are the content basis for the airline's own content can exceed 100,000 pages for each aircraft model, translating into many hundreds of thousands of work cards and ancillary material pages for the fleet

❑ Older aircraft almost entirely documented on paper, while newer aircraft are typically documented in SGML

CCMS and XML to the Rescue

How can XML and component content management address an airline's content management needs? A central repository of content facilitates repurposing the content into different formats, reuse of the content components, and electronic review of the documents. At many airlines, documents are often reviewed and approved or through paper reviews even face to face meetings.

Significantly, the hierarchical structuring of content that can be done with XML, and the management of the XML components, lends itself to the kind of information flow and

workflow in flight operations. Knowledge of required changes to a manual flows from the kind of outside information listed above, but currently many airlines have no formal linkage in the electronic documents; the linkages are maintained manually. Component management of reusable content modules would enable airlines to establish and maintain these links electronically, resulting in greater quality and accuracy.

Many of these update scenarios could leverage the intelligence of the repository to support the updates. For example, a change in one document could flag a need to change other documents; the inclusion of a new, related document could flag a need to review and possibly change existing related documents, and so on.

With the content managed at a component level, and with the ability to express rich links and triggers between the content components, a CCMS becomes a valuable platform for ongoing content management. Implemented correctly, a CCMS enables an organization such as an airline to have intelligent, actionable content that can support the broad variety of update tasks and processes that a complex maintenance organization faces.

Conclusions

There is a clear need for technology that best supports the unique challenges some companies face in creating, maintaining, and distributing the voluminous content that supports complex products in the field. Organizations with this challenge must look beyond single-source publishing to a more rigorous program of reusing their content across the many content products they produce. To enable this kind of reuse, organizations must look at encoding their content in XML and at adopting component content management systems that will most effectively leverage the XML-encoded content for reuse.

Is your organization in a position to benefit from component content management? You are if:

❑   You create voluminous product support content that needs to be repurposed into many different formats such as print, HTML, and Help.

❑   You create the kind of content that reuses components of content in many different content products.

❑   You create content that must be translated into multiple languages.

❑   You create content that supports conditional publishing, where variables within the content can be readily used to create variants of the published products.

Earlier content management systems provide some of necessary technology to truly support component content management and reuse. However, these earlier systems do not provide a robust enough platform to truly enable flexible reuse—where components at any point in the content hierarchy can be easily created, updated, managed, combined, recombined, and linked. Only when the underlying platform is flexible enough will organizations have mechanisms for ongoing, flexible use of their content over its long lifecycle.

One final significant point is the ability of component content management technology to enable organizations to adopt reuse over time. By allowing flexible access to all content without requiring the up-front MRU analysis and decision making—and by providing tools for ongoing refactoring and conversion of content—component content management technology enables organizations to adopt and increase reuse over time. This is a more manageable process for organizations and should result in more success, more efficiencies, and more return on investment over time.

Sponsoring Company Information

For more information, please contact:

X-Hive Corporation

HEADQUARTERS:
X-Hive Corporation B.V.
Aert van Nesstraat 45
3012 CA Rotterdam
The Netherlands

Phone: +31 10 281 80 80
Fax: +31 10 281 80 81

ON THE WEB:
www.x-hive.com