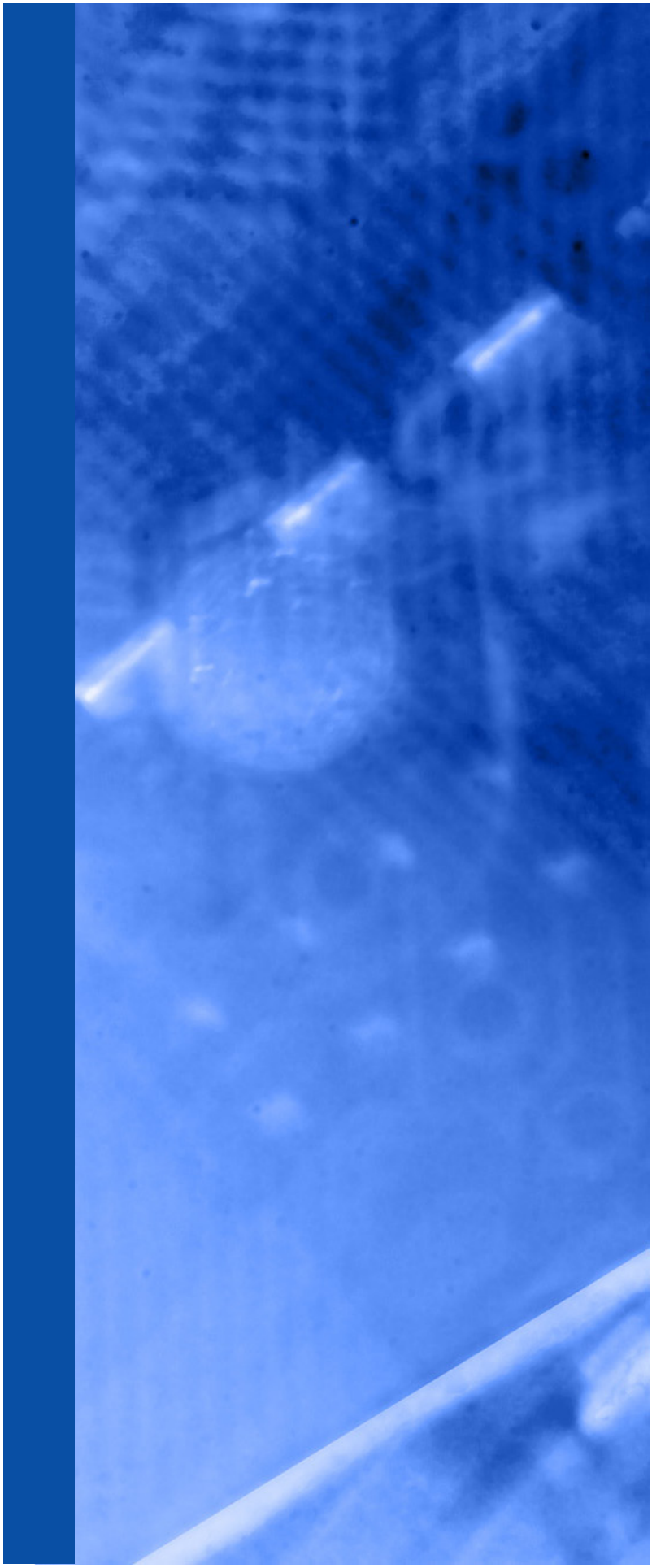

Research Report

Smart Content in the Enterprise

How Next Generation XML Applications Deliver New Value to Multiple Stakeholders

August 2010

by Geoffrey Bock, Dale Waldt, and Mary Laplante

GILBANE GROUP
A DIVISION OF OUTSELL, INC.

The Gilbane Group gratefully acknowledges the support of the sponsors of the research informing this report. This work would not have been possible without them. Please see the Sponsor Acknowledgement section for descriptions of providers.

Gilbane Group
A Division of Outsell, Inc.

763 Massachusetts Avenue
Cambridge, MA 02139 USA

Tel: 617.497.9443
Fax: 617.497.5256

info@gilbane.com
http://gilbane.com

Table of Contents

List of Figures and Tables

Executive Summary

XML applications have long proven their significant value — reducing costs, growing revenue, expediting business processes, mitigating risk, improving customer service, and increasing customer satisfaction. XML-based solutions driven by reusable, componentized content have brought true innovation to production processes, partnership enablement, and customer interactions.

For all the benefits, however, managers of successful XML implementations have struggled with attempts to bring XML content and applications out of their documentation departments and into their larger enterprises. Even the most articulate champions find it challenging to communicate XML value propositions to broader audiences. Yet at the same time, companies are beginning to understand how to capture metrics to make business cases, such as cost savings from reuse. Technologies for XML content processes have matured and are widely available at a variety of price points and delivery models. Service providers have developed knowledge and process expertise that can offload much of the work that distracts enterprise users from focusing on their core businesses.

But for all this market readiness, so much XML content value remains untapped. Why is this so? Why do XML applications that have performed well at the departmental level fail to be adopted in other areas of the organization where they can also deliver business benefits? What does it take to break out of the XML application silo? What is the magic formula for an enterprise business case that captures and keeps the attention of senior management?

To answer these fundamental questions and issues related to them, analysts in the XML practice at the Gilbane Group (a division of Outsell, Inc.) set out to investigate the current and emerging landscape of XML applications, looking for insights into where the true obstacles to broader enterprise adoption are rooted. The primary objective of our research was to find out where the common points of failure have been, and why. More importantly, where there have been successes with moving beyond an initial application, we wanted to know why, how, and to what benefit. Finally, we wanted to understand how that experience could be abstracted and universalized so that other companies could learn from it and start mining the value of XML content and applications throughout the enterprise.

Our primary research comprised extended conversations with business and technology leaders within organizations that have deployed at least one XML application in new and innovative ways. Through a series of open-ended questions, we discussed business objectives, development activities, and outcomes. (We are publishing the results of these conversations as a series of case studies that accompany this report.) We combined the knowledge and insight from this research with our industry and implementation expertise as well as our deep knowledge of the XML markets and technologies, drawing on our roles as industry analysts.

The results of our inquiry are reported in this study on *Smart Content in the Enterprise: How the Next Generation XML Applications Deliver New Value to Multiple Stakeholders*.

The research and analysis led us to make the case for thinking about XML content in a new and fundamentally different way, giving rise to the need for the next generation of XML applications, capabilities, and competencies. Organizations that begin to understand this will be at the forefront of leveraging content to deliver new value to new stakeholders, including customers.

We cannot thank our study participants enough for the time and effort they contributed to our research. They represent the very best of XML application leadership—talented, passionate, and dedicated to improving practices in their areas of expertise. We are grateful that they chose to share their stories with us and, more importantly, with our readers. We also extend sincere appreciation to our sponsors, who made this important market education possible.

Study Highlights

Drawing on our research, analysis, and experience, this study finds:

- ***A shift from an inward- to outward-facing view of content practices and infrastructures.*** Key drivers for broader deployments lie in outward-facing customer impact rather than inward-facing operational efficiencies and cost reductions. The connection between content, customers, and business goals and objectives is being articulated in compelling ways by the organizations that are leading the way.
- ***An emphasis on content utility***. The shift to customer impact gives rise to keen attention to the usefulness of content to the user at a specific point in performing a task, first identified as *content utility* by Gilbane Group's 2009 study on global product content.[1]
- ***Delivery as the starting point of application design.*** Recognizing the value of content utility turns the traditional view of the content life cycle on its head. In the past, we planned and viewed our content processes from creation through delivery. Leading organizations now start by considering the best possible experience with *consuming* that content and working backwards, building their content management and authoring processes to support the end-goal instead of the other way around.
- ***Enrichment as the new content competency.*** The emphasis on content utility creates new demand for tools, processes, and skills for enriching content with embedded metadata, represented as XML tags and tag sets, throughout the content life cycle, not just at the point of creation. Developing capabilities

[1] *Multilingual Product Content: Transforming Traditional Practices into Global Content Value Chains*, July 2009.

in content enrichment will create competitive advantage for companies who do it well.

- ***Social content and computing within content ecosystems.*** Content development is no longer "owned" by a single organization. Collaboration around enriched content is the defining characteristic of applications that have successfully jumped departmental walls and flowed into the enterprise. As a result, social content and computing are becoming part of the landscape for the next generation of XML applications.
- ***Big impact from modest content enrichment.*** Flowing XML content into the broader enterprise does not always require a major initiative. Enriching information with just a small amount of metadata can deliver compelling benefits. As initial experiments begin to pay off, leading practitioners are learning to walk the line between under- and over-investing in content resources.

Why "Smart" Content?

In our earliest interviews, we noted that when companies succeeded with XML deployment beyond a single application, adoption was very often driven *by the content itself*. The remainder of our research supported this observation. The pattern was intriguing. Companies have been creating tagged, structured content for decades. Why was *this* class of XML content succeeding in attracting broader attention within the enterprise? Clearly there was something fundamentally different about it.

Conventional XML market wisdom has largely held that broader enterprise adoption would typically be imposed, or driven as a "push" force based on an initial success in a department. That there are obstacles on this path from department to enterprise was, in fact, a central premise of this very research. What we discovered, however, is that in many cases, adoption is *not a push from an initial success in a department, but a pull based on content value.* Customers are rapidly coming to expect accessing just the essential information that answers questions and solves their problems.

The XML content that is driving contemporary enterprise applications is

- Granular at the appropriate level
- Semantically rich
- Useful across applications
- Meaningful for collaborative interaction

In this study, we have opted to use the term "smart content" to define this class of content. Smart content is a natural evolution of XML structured content, delivering richer, value-added functionality (as illustrated and discussed later in the section on *New Value Propositions for Content*).

In our research, we see that when companies invest in making their content smart, the value of the content begins to speak for and prove itself. Barriers to adoption outside of the initial application begin to erode, or at least decline to the point where they become more manageable. Considering how content may be used for multiple purposes, as well

as designing a data model and creation and enrichment processes to meet these goals, is key to creating smart content. This opens the door to thinking about XML content in new and innovative ways, spurring adoption throughout the enterprise.

A Word About DITA

The Darwin Information Typing Architecture (DITA), the OASIS standard[2] for XML architecture for designing, writing, managing and publishing information, figures prominently in the research conducted for this report. Readers will it find referenced throughout, particularly in the customer case studies. DITA is perhaps the most common example of smart content as it is being deployed today, although it is not the only XML schema for creating content that is useful throughout the enterprise.

Within the context of this report, readers are encouraged to think about smart content in its broader sense, and DITA as one instance of that class. In the coming years, other XML standards may capture other aspects of the enterprise market as DITA has. In fact, several industry groups and other organizations have created standards for their stakeholder communities that embrace these smart content concepts, though these may not be as widely adopted or understood as DITA. Building content practices grounded in the principles of smart content, as described in this study, will create a path to competitive advantage, no matter which XML application-level standard suits the business needs of the organization.

Using This Report

Smart Content in the Enterprise is designed for two primary audiences:

- ***For managers of XML applications*** who see a broader opportunity for XML within their organizations. The study will help them learn from the success of others, communicate value, build business cases, and constitute cross-functional teams. The report will also be useful for managers who are in the process of implementing XML and want to ensure that they are designing and building their applications to align squarely with contemporary user and content requirements. We refer to this audience as "operational champions," a term we coined in Gilbane's 2008 content globalization research.
- ***For suppliers of technologies and services***. The study is meant to help them develop offerings that alleviate pain points and address issues and challenges, and to market and position their products in ways to make their value propositions clear to buyers. We refer to this audience as "product champions."

Our report comprises an analysis of the research, supplemented by a collection of case studies supporting it. The report is organized five sections:

[2] http://dita.xml.org/

- This executive summary
- *New Value Propositions for Content*
- *Smart Content in Action*
- *Smart Content Technology Landscape*
- *Conclusions and Gilbane Perspectives*

Following the analysis are two appendices: *Developing a Smart Content Roadmap* and *A Guide to Smart Content Concepts*. These supplemental materials present the action-oriented fundamentals and are meant as educational background for readers who are not currently XML practitioners, and operational or product champions. The report concludes with a series of profiles of organizations whose experiences illustrate the real and potential benefits of smart content when it is more broadly deployed within the enterprise.

We realize that operational and product champions (who have designed, implemented, and managed XML and SGML applications over many years) may be familiar with some of the insights presented in this report. But we also appreciate their frustration that executives and managers in other functional areas have not recognized the XML content opportunity.

This research and Gilbane's experience working with global enterprises clearly show that attitudes towards and appreciation for content are finally changing, opening windows of opportunity for XML-savvy champions to play a role in moving their applications and experience beyond their existing implementations. We hope that this report and the success stories included therein will help drive conversations that accelerate the adoption of smart content applications and the realization of benefits that go with them.

New Value Propositions for Content

Investments in any new technology or practice only deliver value when they are aligned with the larger business goals and objectives of the organization. Our study on smart content therefore starts with a business perspective. What external factors are having impact today on XML-aware content strategies, practices, and infrastructures? Our research surfaced three common influences:

1. ***The enterprise as network.*** Global businesses are now completely network-based. In less than two decades, the web has fundamentally transformed business operations, from internal process management to transactions with customers and communications with internal and external stakeholders.
2. ***Demand for immediacy and interactivity.*** Instant and always-on connectivity is not only an expectation but also a requirement for maintaining competitive advantage. Connectivity is clearly related to the second influence on content practices: an expectation of immediate or near-immediate response and high-levels of interactivity in the engagement.
3. ***Pervasive emphasis on global customer satisfaction.*** Customer satisfaction, the third significant influence, has always been a primary lever of business success. Since the economic downturn of 2009, however, Gilbane has seen unprecedented executive and operational commitment to delivering customer satisfaction through positive experience. When new revenues are hard to come by, companies put renewed efforts into keeping current customers.

   This attitude towards and focus on customer satisfaction is now becoming institutionalized — and so is having a profound impact across *all* enterprise functions. Customer satisfaction through positive experience is no longer limited to the managers and staff in customer support organizations. It extends to all the ways that companies touch and engage with their customers.

It is at the nexus of these influences — connectivity, immediacy and interactivity, and customer satisfaction and experience — that new value propositions for content take shape.

Connecting Content with Business Levers

Content is a fundamental opportunity to engage customers, deliver satisfying experiences, and drive revenues. This insight will not surprise operational champions who have long recognized the value of XML or the product champions who are driving technology innovations. Its significance lies in the window of opportunity that it opens and in its implications for enterprise content practices — primary topics of this report.

Within forward-thinking companies, we finally see that executives and managers are beginning to understand the business prospects. Content is turning the corner from cost center to profit contribution. And in this way, new value is delivered to new

stakeholders throughout the business ecosystem, including partners, customers, prospects, and global workforces.

As organizations begin to understand how content can deliver new value to new stakeholders, they are shifting their perspectives on their content practices and infrastructures from an inward- to an outward-facing view. It is no longer enough to focus only on internal needs and operational efficiencies. The usefulness of content to its ultimate end-consumer is becoming the primary consideration for investments in practices and infrastructure (people, process, and technology). It is about what happens *out there* rather than *in here* that matters most today.

These trends are distilled in two key findings of our research:

- At the operational level, a shift from an *inward-facing* view of content practices and infrastructures to an *outward-facing view*. Organizations are realizing that their content applications must be driven by external requirements rather than (or at least in addition to) internal needs.
- At the content level, an emphasis on *content utility*. If customer satisfaction is a primary goal, then content must enable satisfying interactions. Organizations recognize this and are building practices and infrastructures based on content utility — the delivery content that is useful at specific points in a business process.

To make these assertions less theory and more practical, consider the content-centric business applications implemented within organizations profiled in this report:

- Search optimization
- Partner management
- Customer self-service
- Community-based content development and support
- Mission-based marketing
- Public policy development and communication

In each case, a need or desire for content utility drives the organization to approach the content application from an external perspective. How will the end-consumers of the content use the information provided to accomplish a business task? How does the organization go about publishing that content and the delivering the experience that goes with it? What are the implications for the content itself?

By looking across our research and our experience working with enterprise users, we can identify a particular set of characteristics associated with the content that drives these applications. That content is:

- Granular for flexible use
- Rich in descriptive information about the individual components
- Useful across multiple applications
- Meaningful for collaborative interaction

The first three are defining characteristics of all XML applications; we find, however, that granularity, metadata, and application interoperability take on new meaning for content that is designed with content utility first and foremost in mind. The fourth defining characteristic of smart content is emerging based on social use and value of XML content, which figures prominently in the next generation of XML applications. The rise of collaborative interaction around XML content is another key finding of our research and an important milestone in the evolution of XML applications, as explored later in this report.

Characterizing Smart Content

We believe that these four characteristics come together in the next phase in the evolution of XML content — what we call smart content. When we tag content with extensive semantic and/or formatting information, we make it "smart" enough for applications and systems to use the content in interesting, innovative, and often unexpected ways. Defining various uses of content leads to robust data models and enrichment processes that support multiple purposes for the content, allowing it to meet a broader range of business objectives. Organizing, searching, processing, discovery, and presentation are greatly improved, which in turn increases the underlying value of the information customers access and use.

Figure 1 charts the industry's progression from unstructured content to structured content, and finally to smart content, in terms of automated processes, business value, and the resulting customer experience.

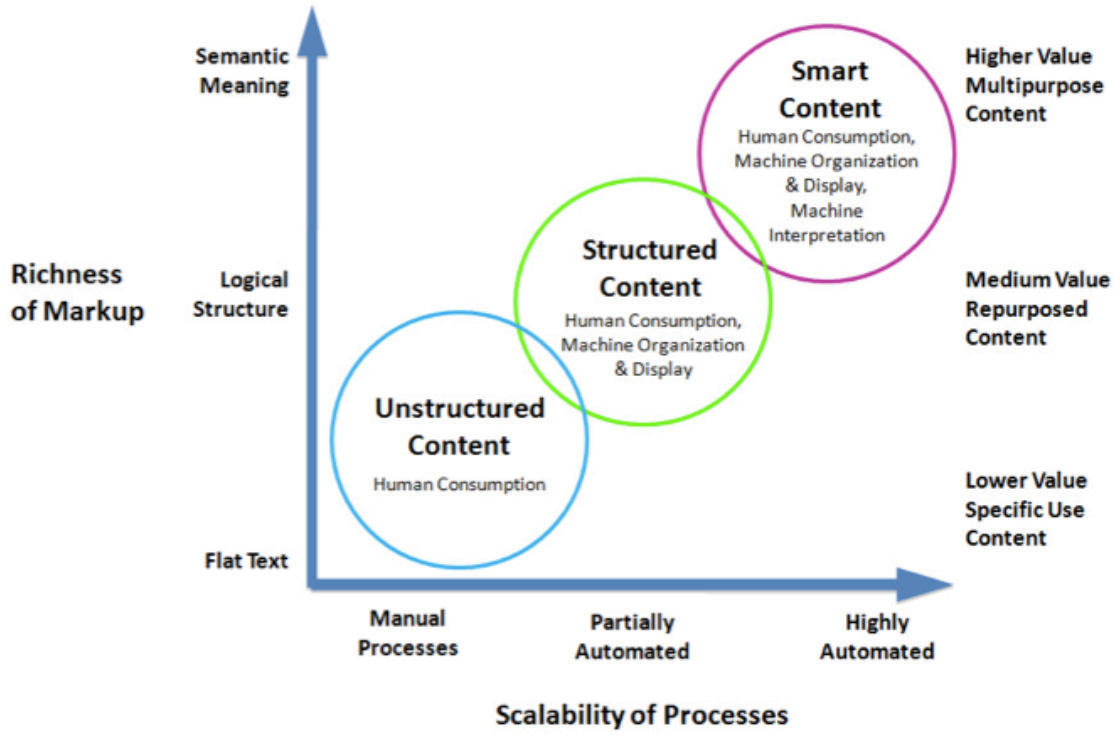

Figure 1. Evolutionary capabilities for content

Figure 1 illustrates the range of properties and capabilities of content managed with various tagging strategies, indicating a steady progression of markup richness, process scalability, and content value. Richness of markup is an important technology driver and a key to making content ever more "intelligent."

Smart content extends the core capabilities of structured content in important and innovative directions. It includes a set of semantic tags and metadata that describe the meaning and relationships among the underlying information resources. Defined by XML, these tags encapsulate a variety of factors such as metadata and attributes applied to structured elements, semantically named elements that reference predefined taxonomies, and/or presentation elements for high-fidelity rendering.

However the information is enriched and tagged, the results are available to both humans and applications to process. Smart content enables highly reusable content components and automated dynamic document assembly. Smart content enhances precision and recall for information retrieval by tagging content with detailed clues about what it is all about, where it comes from, and how it is related to other content components. In short, smart content helps to develop robust and valuable content ecosystems with highly reprocessable information.

The Business Value of "Smart"

Beyond a Fixed Structure

Many currently deployed XML applications fall into the middle circle in Figure 1 above. Structured content includes well-formed XML documents, compliant to a schema, or even the content managed within relational databases (RDBMs). There is limited intelligence embedded in the content, such as boundaries of elements (or fields) being clearly demarcated, and element names (or metadata) that mean something to both humans and systems that consume the information. Automatic processing of structured content includes reorganizing content, generating navigation aids and tables of contents, segmenting content into components, rendering it for print or online display, and other predefined processes streamlined by the structured content data models in use. Systems can augment human capabilities and help with the processing within predefined domains.

Structured content is designed to produce predefined results within a defined context. For example, content initially designed to be published as a printed or electronic document can be repurposed for additional formats including viewing in a laptop web browser or on a mobile device. Smart content, by comparison and by design, contains additional intelligence to support multiple purposes to produce results in multiple contexts, often enabled by mash-ups or other automated processes — as directory listings, for example, that are automatically organized by locations and geospatial codes.

From Repurposing to Multipurposing Content

Many organizations have deployed single source structured publishing solutions to achieve efficiencies in their production processes, to repurpose content for new delivery channels, and to address new market opportunities. So why is structured publishing and content repurposing not adequate to meet the needs of organizations developing smart content solutions?

When we repurpose content, we reorganize or reformat it for a use other than the original purpose it was design to satisfy, such as transforming a book (published as a linear document) into a collection of standalone web articles. In a smart content solution, we consider the customer experience and the business value for the content upfront, and build a data model and content production process that meet these multiple purposes from the onset.

Multi-purposing can provide significant advantages over repurposing. Many of the organizations we researched recognize that they can tie their content development process to other business processes more effectively with the right granularity and enrichment. Multi-purposed content can be used for more than documentation by organizing it for both publication and, for example, loading into systems to configure manufacturing equipment. Training material can be more easily used to guide a user through a complex task such as troubleshooting software or hardware. Directory content can be integrated with geospatial information to produce location specific lists of services. In short, multipurpose content is more versatile, flexible, and interoperable that classic structured content designed to be repurposed.

Focusing on the Customer Experience

As our research shows, smart content opens the door to using information in innovative ways throughout an enterprise. Successful organizations that deploy XML content beyond an initial application recognize the need to recast and redefine their content processes to focus on the overall customer experience. They are no longer repurposing content, as a way of transforming information from one use to another, but rather multi-purposing content for multiple business solutions, right from the start.

While many organizations see the value of XML content in their broader enterprise and are even willing to invest in smart content, they fail to recognize that most departmental XML applications are built on a legacy of linear publishing and self-contained documents. This model no longer delivers sufficient value and, in the digital age, is not sustainable.

We are now seeing the emergence of the next generation of XML applications, based on sets of content components. Successful organizations invest the time and resources to define and maintain their core content components, and apply them to multiple business applications. These organizations are able to manage content components on their own, as discrete objects, rather than as parts of linear publishing processes. Successful organizations are able to focus on identifying the content customers want and expect—and then determine how to organize, store, manage, and deliver it.

The new business proposition for content is all about delivering the information that at minimum meets — and wherever possible exceeds — customer expectations. To make content smart, it is essential to optimize the business value of content components by investing in the information architectural activities required to add relevant metadata. We need to recognize the power of enrichment, together with the capabilities of content component management.

Smart Content in Action

Our research confirms that smart content is current practice, driving familiar business functions like customer support and partner management. The scenarios that follow are hardly hypothetical examples about the application of content components. Rather, they are derived from interviews that we conducted as part of the research for this report and are composite examples of situations based on our case studies.

- ***Search optimization.*** A software company wants to make sure that essential support information is easily findable by search engines over the web. Lengthy technical manuals with detailed tables of contents made this impossible. The company decomposed the information into discrete content components, the pithy snippets that customers find useful. These components are tagged using an extensive set of support-related index terms, designed to enhance findability by third-party search engines. Instead of searching sequentially within individual documents, customers seeking support information can search across entire libraries, and even narrow their search using topics and metadata to improve the precision of their results.
- ***Partner management.*** A semiconductor design organization needs to manage the flow of complex technical information to multiple supply chain partners. Rather than publishing a series of predefined manuals on various topics, the organization tags content by the detailed attributes of various technologies, and then dynamically assembles unique publications for each partner, containing just the information each needs. The organization automatically creates slightly different versions for each recipient without a lot of manual reworking of the content.
- ***Customer self-service.*** With customers expecting ever more detailed technical support information, a computer manufacturer needs to expand the scope of content produced for its rapidly growing family of products. The company adopts a modular approach to its technical documentation rather than publishing manuals for each product. The firm proceeds to identify discrete content components on a wide range of technical topics, tags components by product attributes, solutions, and other customer-centric capabilities, and then dynamically delivers them as needed to address customer concerns.

In each example, the outcomes entail content utility enabled by just-in-time information delivery – ensuring that customers, business partners, employees, and other stakeholders can easily find the information they need, just at the point when they are going to make a decision or take action. The information is sometimes assembled and rendered as dynamic documents, and in other cases is simply delivered over the web or to a mobile application as part of an interactive experience.

Designing Smart Content

We believe that these examples highlight a design model for smart content. Too often organizations create and publish content, without considering how customers are going to use the information to meet their business objectives. The emphasis on content

utility—ensuring positive customer experience with content that is useful to the user at a specific point in performing a task — demands that we turn the traditional view of the content lifecycle on its head.

Rather than taking an inward-facing approach and simply focusing on authoring and publishing information, leading practitioners are starting from an outward-facing perspective. Tagging is based on predefined criteria for delivery and consumption, rather than for purposes of internal efficiencies. Content is not only created and managed, but also enriched to optimize the customer experience.

Here are the basic steps for designing smart content.

- First, we need to understand all the ways the content is going to be consumed – what customers want and expect, and what they need to know, the limits and capabilities of each delivery channel. Along the way, it is important to capture the business drivers – how content adds value to the organization that is making the investment, and enhances the quality of the customer experience.
- Next, we need to plan for enrichment. We need to identify the categories and attributes that capture the model, and define the criteria for segmenting content into its component parts. Often we incorporate the indexing criteria and tags from a predefined taxonomy, a schema, or other externally defined sets of categories. Sometimes we can enrich content in context, by capturing predefined states when the content is created, modified, assembled, or displayed.
- At the point of content creation, we need to embed tags and associate metadata within the content components. In some cases, we need to provide content creators with the tools to tag and organize information, using predefined categories and metadata. Sometimes we tag and categorize content by inferring the context in which it is created and managed. Along the way, we transform authoring, editing, and production processes. Authors and editors are able to develop and publish content in entirely new ways, and expand their reach to meet the needs of new audiences.

Table 1 summarizes the design activities for our three application examples. With appropriately defined tags and tag sets, we can enrich the content to enhance the quality of the end-user experience.

Use Case Scenario	Content Delivery	Content Enrichment	Content Creation
Search Optimization	• Improve search precision and navigation by delivering just content components that answer questions (rather than lengthy document).	• Search categories, derived from pre-defined taxonomies or controlled vocabularies. • Standardized tag sets make search categories visible to web crawlers.	• Tagging contents by relevant terms. • Validation to ensure completeness of content to meet search requirements.
Partner Management	• Customized content delivered for each partner. • XML tagged files for end-to-end delivery.	• Standardized tag sets (DITA) with specializations. • Tagging maps to partner entitlements.	• Authoring templates and workflows meet partner content provisioning and timing requirements. • Flexible content organization allows reorganization for specific partner needs.
Customer Self-service	• Targeted content delivery through dynamic publishing. • Content tagged for intelligent retrieval by third party and web-wide search engines.	• Content elements classified and mapped to specific user / delivery types. • Content roles validated for compliance with policy and regulatory requirements.	• Topic-oriented writing enables collaboration and easy updating. • Writers specialize on topic areas of expertise. • Content created and marked up sufficiently for use in multiple delivery formats / uses.

Table 1: Design activities for smart content scenarios

Enriching Content with Tags and Metadata

There is a common thread running through these three examples. All three organizations encounter problems with their conventional (or linear) publishing processes when delivering electronic information that meets their customers' expectations. Tried and true procedures that initially made sense for print-based linear documents no longer work effectively when delivering interactive content over the web.

All three organizations proceed to develop tools, techniques, policies, and procedures for enriching content and making it useful. All three are proceeding to break through the confines of single-use information silos, and make the information widely accessible across networked environments.

Steps for Enriching Content

In fact, all three organizations are in the midst of transforming their publishing processes in significant ways. These organizations are transitioning from conventional linear process to newer approaches that enable innovative enrichment and delivery capabilities. Each organization is focusing on developing an environment with a high level of content granularity, and where contributors can enrich content by adding tags and metadata to the information as they create and revise it.

Sometimes the enrichment is explicit. For instance, when contributors complete the fields of a form or file content in a content management system, database, or file system, they need to add the subject matter, product related information or other descriptive categories. Sometimes the enrichment is implicit – and is derived from the context in which contributors are operating. For example, contributors may use a predefined template that functions as a style sheet within an editor. This template automatically componentizes and embeds both formatting and semantic tags as part of an editing activity.

When publishing smart content, these organizations perform several important steps where they are able to enrich content in depth.

- Enrichment begins with componentizing the information and defining the granular content components.
- Next these organizations define the content categories. These are the tags and tag sets that define how the content is used, and how components are related to one another. Wherever possible, leveraging industry standard tag sets (such as DITA) helps to accelerate the design process, and also to generalize the categories beyond single applications.
- Then each organization develops its own processes to assign the metadata and tags to the components.

We further describe the underlying technical concepts for structuring smart content in Appendix B, "A Guide to Smart Content Concepts."

The end result, however, is a large set of granular content components that are enriched based on a predefined XML schema. Keeping track of all these components is often a chore. Thus, with increasing frequency, organizations are beginning to rely on the capabilities of a component content management system to manage the content components within the repository, and to structure the organization and application of relevant metadata.

Enriching Content for Smart Publishing

How can we capture the lessons learned from these three organizations? When we produce smart content, there is still a publishing process. But, our capability to enrich content takes on an added significance. No longer are we focusing simply on a linear publishing process where we are first creating and then delivering content.

There is a third dimension, as shown in Figure 2. Organizations must come to terms with how they enrich content as they create and deliver content. We need to take these additional editorial activities to take into account.

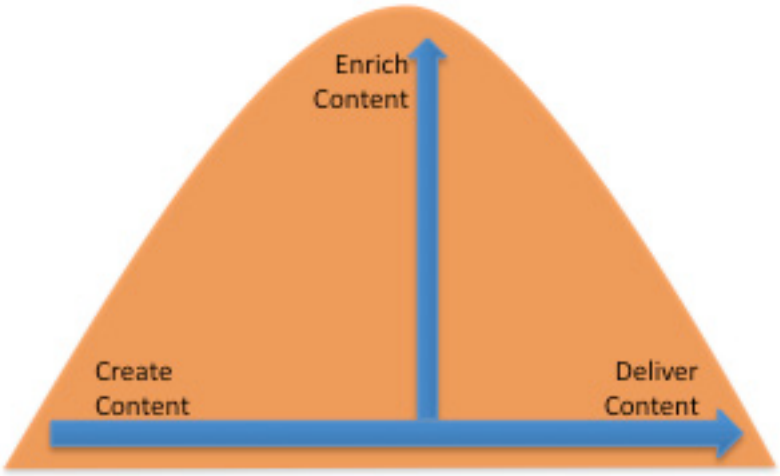

Figure 2. Enrichment activities as integral to the publishing process

Of course, content enrichment does not occur in a vacuum. A publishing process is still required to create and manage content. It is essential to model content delivery by how customers utilize the information to solve business problems.

Enrichment is our challenge. Thus it is important to make the transition:

- From conventional approaches to publishing, where there are limited opportunities for enrichment
- To smart content publishing, where we are producing componentized content with extensively defined tag sets and automatically embedded semantics

There is an additional factor to consider – the contributor's experience. To the maximum extent possible, it is essential to hide the complexity of content enrichment from the contributors who author and edit content on a day-to-day basis. Let's now turn to the range of technologies available to facilitate enrichment -- what we term the smart content application landscape.

Smart Content Application Landscape

As with all business applications, an opportunity for a smart content solution often starts with conversations and the business case — conversations among the stakeholders with a vested interest in content utility, and the business case to secure funding for related investments in practice and infrastructure. Developing a shared understanding of where the organization is today and where it needs to go is essential to these processes. A very useful approach is to map current capabilities and future requirements on a technology landscape — a high-level view of the various technologies that are relevant to a specific class of business applications.

Figure 3 illustrates the technology landscape for smart content. It reflects the evolutionary nature of XML applications, as described throughout this report.

- At the center are fundamental XML technologies for creating modular content, managing it as discrete chunks (with or without a formal content management system), and publishing it in an organized fashion. These are the basic technologies for "one source, one output" applications.
- In the middle ring are the technologies that enable single-sourcing—the reuse of content components for multiple outputs. They include true component content management, multi-channel delivery, and structured editing for casual contributors and workflow approval.
- The outermost ring includes the technologies for smart content applications, as described below.

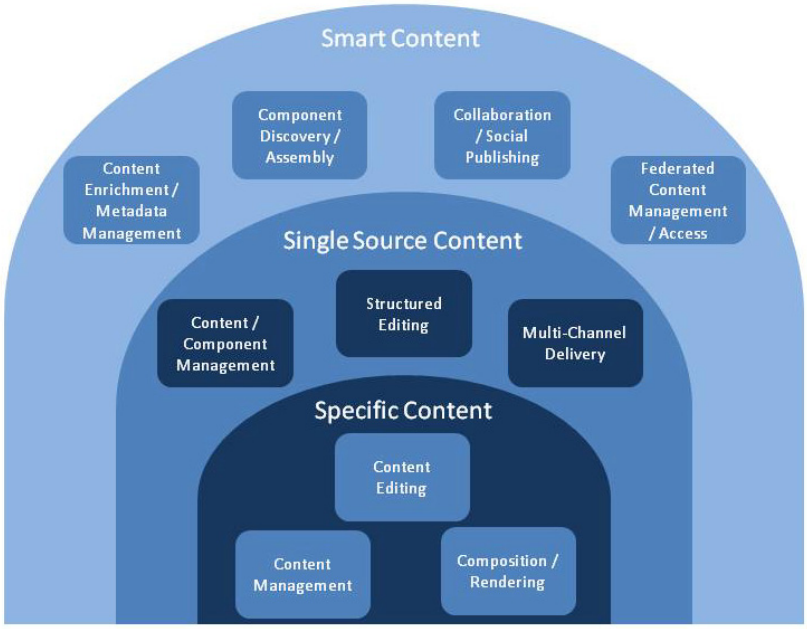

Figure 3. Technology landscape for smart content

Operational champions can create a customized view of the smart content landscape by identifying the XML technologies already in place and those that will be required for a smart content application. Such high-level gap analysis can be extremely useful as a means of engaging stakeholders (particularly those in IT) and laying the groundwork for the business case. Creating a visual picture of current state and the requirements for greater content utility (and the related business benefits) makes the opportunity less theoretical, more practical.

Smart Content Capabilities

Smart content solutions rely on structured editing, component management, and multi-channel delivery as foundational capabilities, augmented with content enrichment, topic component assembly, and social publishing capabilities across a distributed network. What capabilities do these enabling technologies bring to smart content applications? It is important to identify how content components are defined, as well as how they are tagged and enriched by predefined sets of metadata.

Content Enrichment / Metadata Management. Once a descriptive metadata taxonomy is created or adopted, its use for content enrichment will depend on tools for analyzing and/or applying the metadata. These can be manual dialogs, automated scripts and crawlers, or a combination of approaches. Automated scripts can be created to interrogate the content to determine what it is about and to extract key information for use as metadata. Automated tools are efficient and scalable, but generally do not apply metadata with the same accuracy as manual processes. Manual processes, while ensuring better enrichment, are labor intensive and not scalable for large volumes of content. A combination of manual and automated processes and tools is the most likely approach in a smart content environment. Taxonomies may be extensible over time and can require administrative tools for editorial control and term management.

Component Discovery / Assembly. Once data has been enriched, tools for searching and selecting content based on the enrichment criteria will enable more precise discovery and access. Search mechanisms can use metadata to improve search results compared to full text searching. Information architects and organizers of content can use smart searching to discover what content exists, and what still needs to be developed to proactively manage and curate the content. These same discovery and searching capabilities can be used to automatically create delivery maps and dynamically assemble content organized using them.

Distributed Collaboration / Social Publishing. Componentized information lends itself to a more granular update and maintenance process, enabling several users to simultaneously access topics that may appear in a single deliverable form. Subject matter experts, both remote and local, may be included in review and content creation processes at key steps. Users of the information may want to "self-organize" the content of greatest interest to them, and even augment or comment upon specific topics. A distributed social publishing capability will enable a wide range of contributors to participate in the creation, review and updating of content in new ways.

Federated Content Management / Access. Smart content solutions can integrate content without duplicating it in multiple places, rather accessing it across the network in the original storage repository. This federated content approach requires the repositories to have integration capabilities to access content stored on remote systems, platforms, and environments. A federated system architecture will rely on interoperability standards (such as CMIS), system agnostic expressions of data models (such as XML Schemas), and a global network infrastructure (such as the Internet).

These capabilities address a broad range of business activities and, therefore, fulfill more business requirements than single-source content solutions.

A Functional View of Smart Content Applications

How do the capabilities in the landscape fit together in a functional application view?

Managing content as components remains the defining characteristic of all XML applications, from first generation to today's emerging smart content applications.

An emerging best practice is the implementation of a purpose-built component content management system – a content management system specifically developed to manage large numbers of content components as discrete elements. But it is certainly the case that many organizations begin by managing components with other tools and technologies, from file sharing to source code control systems.

Regardless of technology, the fundamentals of XML content management stay the same – robust validation and management of content in a generic format suitable for transformation to a variety of delivery formats. In smart content applications, these enabling technologies are not replaced. Deep experience with them has evolved into additional significant organizational value by allowing them to be used in new ways. It is the new tools for smart content that extend these applications to deliver new content value to an every growing range of stakeholders.

As shown in Figure 4, a shared repository is at the core of smart content applications. This enables multiple contributors with different skills and expertise to easily collaborate on writing and tagging content as well as managing tag sets. The environment supports dynamic content delivery in print, over the web, and directly to various applications. Content components are enriched with both structure and semantic tags.

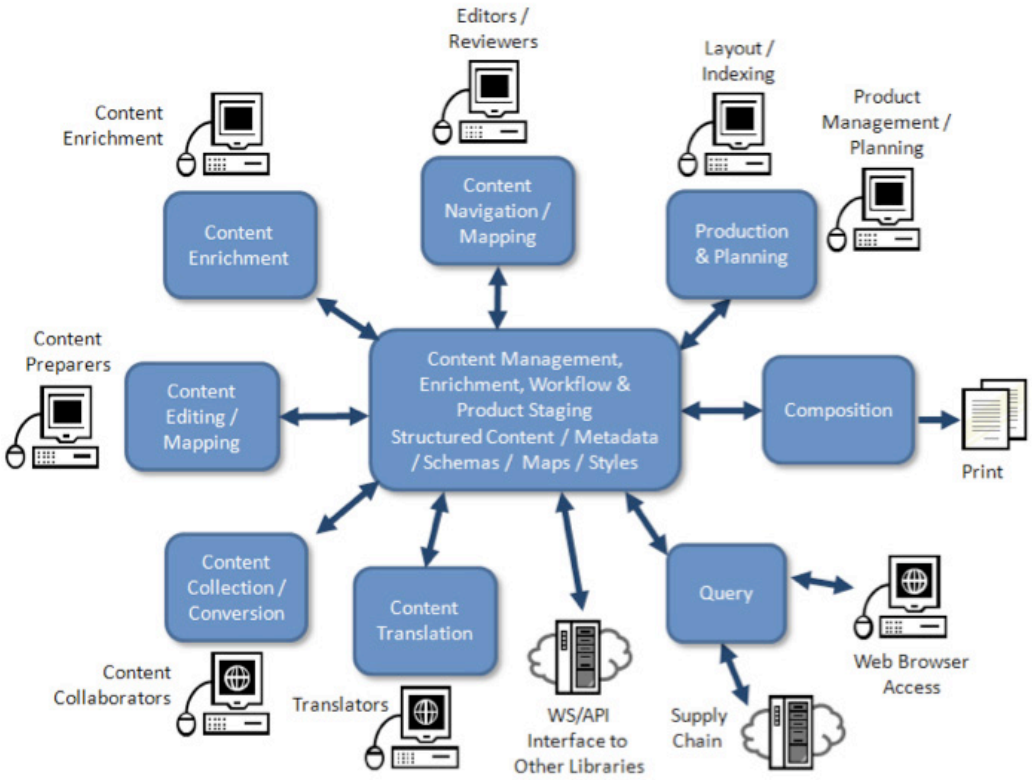

Figure 4. Shared repositories as the functional model

Content contributors can have several different roles. In some cases, these participants may be new to content-centric processes; they might not have served as content reviewers in the past, for example. The shared repository manages access rights and permissions. It contains all the content components, schemas, style-sheets, document maps, metadata, and other descriptive information used to publish information in multiple output formats and make it widely accessible to external applications over the web.

Content preparation may require people and processes beyond the scope of a conventional documentation department, such as translation capabilities for producing multilingual content. Also, the shared repository can enable controlled access by members of an extended enterprise – third party contributors and information supply chain partners -- through a web browser or API.

A variety of different tools can be used to access and edit content.

- Documentation specialists may use robust XML structured editors to create and update content.

- Casual users may use a browser-based simple editing tool to review content and make quick updates.
- External users may access content through queries or dynamic assembly mechanisms that deliver the most current information or a controlled subset of the content based on their rights.
- Outside contractors for indexing or other tasks may be given controlled access as well.

In short, access to the shared repository, and the tools that support it, can be configured to provide the appropriate functionality for each processing step.

In addition, smart content publishing can leverage the capabilities of a shared repository to include core workflow capabilities. Workflow processes can be designed to support a distributed team in the creation, review, enrichment, and publishing of content. Some customers and stakeholders for this system may work remotely, while others are part of an internal documentation or editorial team. Work may progress along a set of steps, each of which may spawn automated conversion, transformation, rendering, or other processes.

Social Publishing Processes

Smart content development does not rely exclusively on end-to-end publishing processes – where authors and editors produce content for delivery to customers. Some very interesting applications are evolving that use innovative content aggregation and crowd sourcing techniques to involve groups of content contributors to publish information on the web on specific topics. These social publishing processes aggregate content from disparate sources while also providing a framework for groups of editors to curate the content and organize it for dynamic delivery over the web. While one of the leading examples of social publishing is Wikipedia, there are many others across the web.

In our research we explored a content aggregation site that collects and organizes information from government agencies, commercial businesses, and a wide variety of third-party resources. As the information is collected, content creators use tools to add classification information to create enriched content and deliver it to targeted customers. Social publishing borrows from crowd sourcing, allowing targeted customers and other stakeholders to add to this content, in the form of comments about the services delivered and recommendations about how to accomplish tasks. The site owners retain the editorial control to reorganize and curate the contributed content. The information collection grows and is refined over time.

Social publishing can be applied to traditionally published content delivered via the web as well as to content aggregations, mash ups, and other types of content delivered on the Internet. For example, a software vendor may rely on its customers for practical trouble-shooting tips. Similarly, customers and fans can enrich restaurant listings and hotel information with reviews and ratings.

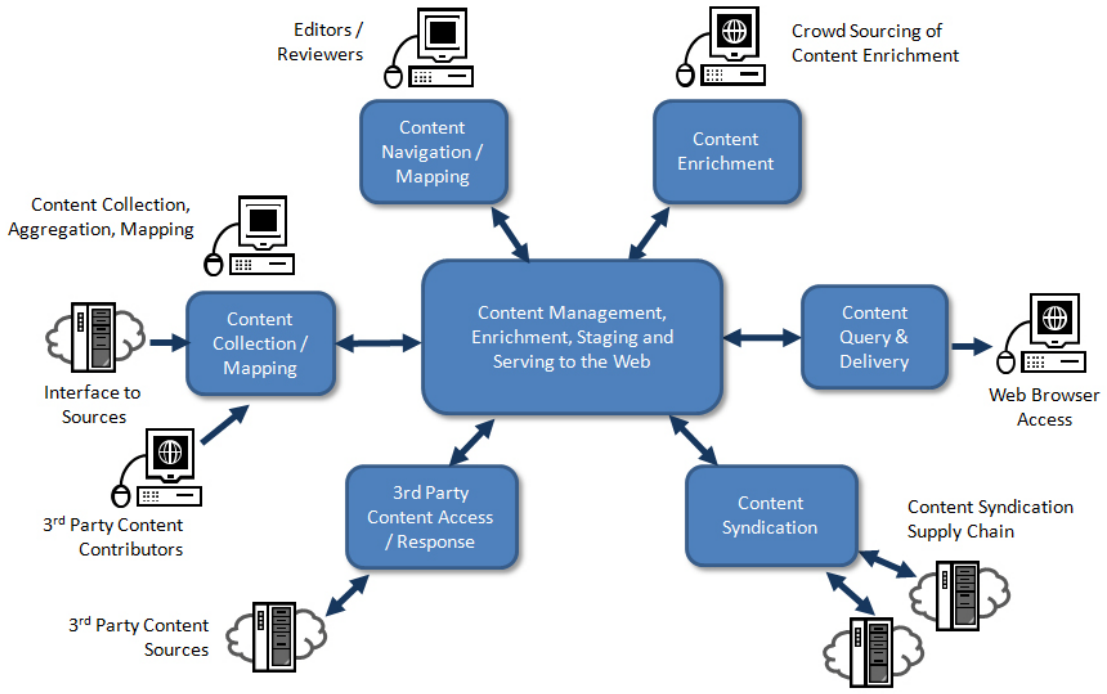

Figure 5. A functional view of social content processes

The overview illustrated in Figure 5 shows how smart content might be delivered to other consumption platforms through syndication. In an interesting twist, the original creators of the information, such as a government agency, can actually be consumers of the enriched version of the content they originally provided. The figure also shows how third-party content sources can be incorporated into a real time query/ response process to enrich and integrate content on the fly at request time. For instance, once a request for certain information is received, the server could incorporate additional content (such as geographical coordinates) and use the combined result set to formulate the response back to the requestor.

In short, with the technical infrastructure to support social publishing in place, content contributors can easily maintain content as modular topics that are accessed across a distributed environment. Crowd sourcing is emerging as a powerful addition to the overall publishing process for organizations that want to leverage content, developed by experts, customers, and other stakeholders, to collaboratively solve business problems. Crowd sourcing exploits the capabilities of richly defined metadata, which in turn enable next generation XML applications to deliver added value to multiple stakeholders.

Conclusion and Gilbane Perspectives

Evolving Content Capabilities

How can organizations implement business applications that make content smart? What is it going to take to embed added intelligence into information that enterprises distribute over the web? As described in our case studies, there is a steady evolution of content development capabilities.

Smart content requires granular content components and tagging – innovative capabilities that leading organizations are in the midst of incorporating into their content environments. But equally important is the *evolution*, rather than the *revolution*, of content development. Smart content leverages the capabilities of single source content, which in turn exploits core capabilities of specific content.

Table 2 characterizes key elements of this evolution, focusing on content delivery, content enrichment, and content creation. It is important to be able to identify how things change.

Publishing Type	Content Delivery	Content Enrichment	Content Creation
Specific Content	• Unstructured content for human consumption • Publish self-contained documents, stored and managed as discrete files	• File-level properties and metadata • Document-level tables of contents and other predefined references and indices	• Ad hoc formatting or style sheets lock content into predefined formats • Content reuse (e.g. from print to web) is semi-automated or performed manually
Single Source Content	• Structured content suitable for machine organization and rendering • Automatically publish content to multiple formats • Generate navigation aids from content structure	• Generic tags used to generate multiple output formats • May include metadata elements to support processing	• Content chunked into components based on hierarchies of delivery formats • Content tagged to accommodate predefined document models
Smart Content	• Smart content suitable for human and machine interpretation • Content components include semantic tags that describe content and how customer groups might use it • Metadata drives automated processes and dynamic assembly of content into views	• Semantic tags based on or enriched with controlled vocabulary, taxonomies, and/or standardized schemas • Semantic tags often referenced through publicly available web resources and services	• Content chunked into components based on an information architecture • Content tagged either explicitly, through fields within templates, or implicitly, in the context of routine business activities

Table 2. Core capabilities of different content technologies

As highlighted in many of our case studies, organizations leverage existing investments in terms of managing documents, content components, tag sets, and metadata. Once they have identified the components, they can further define the metadata over time and in light of experience, to continually make the content smarter and more applicable to different business situations.

More useful content is more valuable content for business operations. It is important to focus on four factors—content granularity, content enrichment, content interoperability, and content collaboration.

Content Granularity

Content needs to be just granular enough to capture customer expectations. Organizations need to componentize their content to a level that their target audience will find appropriate and useful. It is essential to identify and tag the atomic elements of a content collection.

Design insight and human judgment are required to determine the elemental components. In fact, designing for smart content publishing is not unlike book design skills in an earlier era, required for publishing physical artifacts. Excessive granularity creates more effort and complexity to reassemble into deliverables than is justified by the business requirements. Inadequate granularity requires special purpose application development, often coupled with manual efforts, to repurpose the content to meet the business drivers.

Content Enrichment

Enrichment should be just smart enough to meet business needs. Certainly there are many technical ways to enriching content. Organizations need to be able to describe and model how components are interrelated to one another. They need to be able to define tag names and values in a systematic fashion.

Along the way, many organizations will rely on externally defined authoritative resources. For content enrichment to succeed, it is going to become increasingly important to exploit and leverage relevant industry standards. Excessive enrichment requires more effort to maintain than is justified by the business drivers. Inadequate enrichment leaves ambiguities in the content that may require human intervention to overcome.

Content Interoperability

Richly tagged content becomes increasingly interoperable content, useful for automatically exchanging information among various applications over a network. For instance, directory listings can be combined with geospatial information to produce maps showing services in a specific area. Parameters published in the technical documentation may also be loaded directly into automated manufacturing and production systems to configure them.

Organizations should consider requirements for increasing the interoperability of their content. Any complexity or detail in the data model should be designed to meet a specific business requirement to justify the additional effort to add that detail. With XML tag sets accessible to third parties in standardized formats, smart content may be used in surprising and new ways by external applications in mash-ups, aggregation sites, and other situations.

Content Collaboration

With smart content, there is an ever increasing role for distributed collaboration tools. Organizations can capture the insights from a broad range of content contributors across their information supply chain. Enabling customers and stakeholders beyond core documentation teams to collaborate in the creation, updating, and maintenance of content can improve quality and timeliness.

Tools for collaboration are evolving and can work within client editors or across the network in a thin client browser, enabling both local and remote contributors to participate in content preparation. Collaboration requires some thoughtful governance

with policies and configuration settings to incorporate these stakeholders into the process. In short, smart content and collaboration tools expand the publishing process throughout the enterprise by putting the right content in the right hands for a particular task.

Managing the Organizational Change

The Role for an Information Architect

Moving toward a smart content creation and management environment can be done all at once or gradually in phases. Deciding to migrate slowly through a sequential and deliberative process, or making a radical break from prior activities, requires an understanding the organization's business drivers and its capacity for managing change. We find both approaches in use in our case studies, as operational champions move beyond departmental-level implementation of XML applications to those that have broad appeal within an enterprise environment.

These champions focus on understanding their organizations' business requirements and constraints. They proceed with a deliberate plan, based on an information architecture as well as the capability to exploit XML tag sets and metadata.

Time and again we have encountered a new type of content development professional working within an organization – an information architect who focuses on producing the customer experience and who develops the management and delivery plans for enhancing content utility. This person has a deep understanding of what customers need and expect, and has the task of organizing content for satisfying experiences. The information architect engages other stakeholders within the documentation team and across the organization, to chart the end-to-end processes for content development and distribution.

Evolutionary or Radical Change

Whether to promote evolutionary or radical change, there are two schools of thought. The success, of course, depends on the overall management environment and insights into organizational culture.

- Many organizations evolve gradually, first adopting XML and related tools to gain the benefits of single source publishing. Then, once an organization is familiar with managing XML tagged content components, teams can further enrich content around semantic topics, adopt powerful end-user oriented editing tools, and add the processes and systems for multi-purposed content delivery. We can identify the roadmap for technology evolution – beginning with specific content technologies, then adopting single-source content capabilities, and finally smart content tools and environments.
- Other organization decide to adopt smart content capabilities with a new delivery system, often as an effort to make a radical break from the past or to "leap frog" into a future environment. These organizations begin with an understanding of content componentization and an approach to enriching

content. Their approach to component content management is often linked to business value – focusing on delivering a new type of customer experience.

It is important to note that enriching content and making it 'smart' does not always have to be done immediately, when creating and editing the information. In fact, a phased strategy, starting simple and enriching further in light of business activities, may front load benefits while allowing the organization to manage the cost and complexity of enhancing content as time allows. Validation and other quality assurance tools can also be added to provide feedback to the author and allows richly tagged data to be created quickly and accurately.

The bottom line is that moving content to a robust XML data model creates a foundation for subsequent streamlining. It also provides useful information and "hooks" into data to support further enhancements such as componentization, enriching the data, and automating processes. In short, a foundation of XML data is the first step toward a smart content solution.

Business Prospects for Smart Content

Smart content, in short, is tagged with metadata where the information can be dynamically delivered through interactive environments to mold customer experiences. Smart content is delivered from the "outside in," based on an understanding of the content utility -- the ways in which multiple stakeholders use the information to solve business problems. Smart content augments customer insights through ongoing enrichment, continuously capturing metadata about how the content is applied and used.

Smart content is an emerging trend over the web today. Our case studies snapshot how different organizations are leveraging current XML capabilities and are also prospecting for future business opportunities. We anticipate that on the horizon are extensive pools of content components, readily identified by their XML tag sets and able to be dynamically accessed and delivered to solve business problems. The future depends on applications and frameworks being able to manage these semantics (defined as XML tag sets that embed concepts as metadata) at an enterprise scale.

To get started, organizations need to become familiar with managing XML tagged content. As a first step, it is important to componentized content and launch a content enrichment initiative. Along the way, an organization needs to adopt the relevant tools and technologies – ones that make sense in terms of its technology infrastructure. From a business perspective, the appropriate level of investment in managing content at a component level goes a long way towards delivering an intuitive, smart, and satisfying customer experience.

Appendix A: Developing a Roadmap to Smart Content

Creating smart content requires new approaches, processes, and tools. Understanding the requirements for content delivery helps clarify the content creation and enrichment requirements. And by understanding all three, you can gauge your readiness to begin producing smart content.

To assess your organizational readiness for smart content, you need to evaluate your current capabilities and begin to plan for improvements. Here is a checklist of general questions to consider.

- Have you identified business drivers where smart content will provide benefits and advantages? Have you defined the challenges for migrating toward a smart content solution? Have you identified the resources and steps needed to develop a smart content solution?
- Do you have a data model in use, or planned, that will support content modularization, enrichment, and assembly? Do you have the skills and tools needed to enrich content? Can you provide the tags and metadata necessary for optimizing search and navigation, assembling content components on the fly, or reorganizing content for diverse audiences?
- How do you categorize your content? Have you established a taxonomy (or taxonomies) of key categories that are important to your customers and stakeholders? Do you have access to one or more standardized taxonomies – often produced by industry and/or professional groups -- that organize terms in a way that your customers will readily understand?
- Do your processes and deadlines support the additional content enrichment steps needed to provide the semantics necessary to support your delivery requirements? Are there automated methods or external audiences that can augment your resources to enrich your content?

These questions are designed to help you establish an overall baseline for your organizational readiness. You then need to focus on issues in five areas -- defining content components, managing semantics, distributed authoring and collaboration, information governance, and expanding beyond an initial implementation. Let us consider how each area contributes to your overall roadmap.

Defining Content Components

Content that is managed as stand-alone components can be easily assembled into a variety of deliverable forms. Traditional content is more difficult to reorganize in this way. Components can also have lower-level portions identified as variable content that can be turned on and off for different delivery needs (e.g., DITA uses CONREFs (content references) for this type of variable content). It is essential to componentize content around business needs and to develop a formal data model that describes the schema for your content components. You will also need to implement the appropriate tools and processes to work with this schema.

Content components can be enriched with specific classification and property metadata in greater detail than traditional document or document-oriented content. While it may require more sophisticated processing tools, this powerful component paradigm gives organizations much more flexibility and efficiency in managing their information.

Management of content in modules increases the number of objects being managed. It helps to rely on a content management system and workflow tools to improve the overall management and storage of content components. Structured editing tools (such as DITA-aware editors for DITA-tagged content) can simplify the creation and update of content components. Content management, assembly, and rendering tools are also needed to collect and transform components into page displays and published documents. These tools may need to work automatically to assemble deliverable products on the fly for timeliness and custom delivery as well.

Unlike the sequential activities required to produce lengthy documents, componentization also allows authoring and editing to be broken up and performed in parallel. Of course, components will require someone to develop the build lists and maps needed for assembling the information into deliverables, but this also provides an opportunity to create more meaningfully organized assembled deliverables, even to curate the information with more control.

- How does your content lend itself to componentization?
- Will your authors and editors have to extensively re-write and/or restructure your existing information resources to define stand-alone content components? Who will do the work?
- Does your organization need the type of flexibility that components and variable content provide to streamline processes for custom delivery, audience-specific organization, and reuse that components can deliver?
- Does your business case justify the investment in migrating to a smart content solution all at once, or should you plan on a more gradual evolutionary approach to upgrade your systems, data, and processes?
- Do you have the content management and workflow management tools and skills to manage your information as components? How robust are your content and workflow management requirements if you move toward componentization?
- Can your content development team benefit from parallel work processes?
- Can your content use a more flexible and meaningful organization, better targeting of metadata, and curating topics to improve its usefulness and resulting customer satisfaction?

A Little Bit of Semantics Goes a Long Way

As we have seen in the discussion above and the related case studies, enriching content with even a small amount of descriptive information can enable streamlined processes and powerful delivery options. A single element or field used to capture classification values can transform a static web site organized as a document into a library of information accessible for specific audiences. Several elements can provide additional

functionality and value. A little bit of metadata, such as including product line and documentation type classifications can help search and navigation, as well as planning and tracking product development. Enriching content takes some effort. Be sure to focus first on the business drivers as you plan how to enrich your content to produce the desired deliverables and customer experiences.

By organizing content in stand-alone components rather than a sequence of chapters or documents, content can be reassembled into many forms. To determine the appropriate granularity and structure of the content, it is important to understand how the information is consumed and is best organized to solve specific problems.

Bring new stakeholders into the design process – people who understand how the target content consumer will benefit from improved content experiences.

Including topical information and metadata, and organizing content into clearly identified topics has a powerful affect on processing efficiency. By defining topics and other elements with predefined tags, you can substantially improve your ability to make your content visible to search engines, and to optimize the precision and recall of search results.

Smart content can be enriched with a combination of automated and manual processes. But, overinvesting in the enrichment of your content may quickly have diminishing returns and create cumbersome or untimely processes. Authors and editors need to be able to complete the task of enriching the information within their schedules and skill sets.

- Do your delivery requirements include topical searching and/or assembly that might be enhanced with relevant metadata?
- Can processes be automated with the inclusion of key information about each component, the state of completion it is in, or other categories related to the content or context?
- Can alternative processes or resources be used to enrich your content?
- Do you need subject matter experts to define the enrichment criteria?
- Does your organization (or community) have tagging and markup standards, as well as the classification and indexing skills, that allow metadata to be applied consistently and accurately?

Distributed Authoring and Collaboration

Componentization has an additional benefit of allowing work to be spread out among a broader team and to be done in parallel to improve update cycles. As described in a number of the case studies that accompany this report, subject matter experts, such as engineers, customer support personnel, and field technicians, often have the ability to contribute directly to creation of content in a documentation process. This direct involvement on the part of subject matter experts often improves the accuracy, usability, and even timeliness of the content.

At the same time, organizations face the challenge of providing a tool that is simple and intuitive enough not to require extensive training on the part of this expanded team of content contributors. As we describe in some of our case studies, simple editing and classification tools and interfaces are being developed that do not require extensive knowledge in structured editing tools or even the underlying XML markup. The benefit is that these subject matter experts can make changes to the content themselves, without relying on interpretations by documentation specialists (followed by added review cycles) to incorporate updates.

Many organizations are required to have their content reviewed and verified by outside specialists or regulators. A simple tool that provides access to distributed structured content, and allows editing, commenting, and other forms of collaboration may streamline the regulatory review process.

- Could your organization leverage the experience and skills of a broader range of employees, partners, and other resources outside the documentation teams that work on your content? Can these resources benefit from simplified, yet structured tools that allow them easily to contribute to the content being developed?
- Can your organization leverage the knowledge and efforts of an expanded audience of contributors? Can support technicians, instructors, and other consumers of your content provide feedback, and even perform enrichment steps as part of your overall content preparation and management processes?
- Does your organization have regulatory review requirements that could be facilitated with distributed collaboration tools? Will enriching your content improve the accuracy of your content in support of the regulatory review process?

New Roles and Enhanced Governance

Componentization and content enrichment require more planning and collaboration than traditional document creation processes. With smart content components, authors will most likely work as teams creating and managing a collection of related topical information. Content contributors need to have an awareness about how content is organized and indexed.

This collaboration requires guidelines for a consistent writing style, shared formatting templates, standardized language templates, and consistent application of content enrichment. Often the authoring teams are distributed across multiple locations, and based in different countries with different customs. Implementers of effective smart content solutions create governance bodies to oversee the development and compliance with these standards.

- Does your organization manage multi-departmental governance processes well? Does it need to develop policies and processes for managing shared responsibilities that transcend organizational boundaries?
- Does your organization have established editorial guidelines and style procedures for accommodating how content is used in multiple contexts? Are

these adhered to and is the net benefit to the customer widely understood by members of the broad content creation and review teams?

- Can your editing, workflow, and content management tools provide assurances, checks and feedback to enforce the governance of these standards and policies?

Appendix B: A Guide to Smart Content Concepts

Smart content and the systems that produce it depend on three key concepts.

- Componentization allows content reuse in ways not possible with content published only in conventional, linear, documents.
- Content enrichment enables robust process automation and search optimization.
- Dynamic publishing allows content to be delivered in the right form at the right time.

Let's summarize the capabilities of each of these concepts of a smart content solution.

Componentization

Content Components

Componentization entails chunking content into discrete building blocks of information that can be combined in one or more ways to produce a wide range of output products. It is easy to think of an encyclopedia as a set of independent content components that have been assembled into chapters and volumes. It may be harder to envision other information sources in this way, but technical documentation, marketing content, news articles, and other types of content are increasingly being created, managed, and assembled into print and electronic formats from a library of smart content components.

Chunking and managing content as components provides several benefits, including:

- Easing reuse of content between documents, departments, and organizations
- Enabling incremental updating of content to reduce publishing, indexing, translation, and other process costs and schedules
- Enabling dynamic and delivery of components into different specific configurations for different audiences
- Improving search accuracy by providing discrete information objects with specific metadata rather than large documents covering a broader range of topics
- Enabling specific portions of content to be reviewed and enriched by team members and others best suited for that specific task
- Managing the localization and translation of discrete units of content into multiple languages, efficiently and cost-effectively

Content components, if defined and organized correctly, can translate into lower costs and schedules, improved customer service, and even addressing new business opportunities not previously feasible without componentization.

Componentization is at the heart of all of the case studies in our report. For IBM as a semiconductor manufacturer, managing components offers more flexibility. Content components are assembled in slightly different versions for each of their manufacturing partners to allow them to see only the content that they are authorized to see. Even so, IBM is able to reduce the amount of duplicated content and the associated writing, reviewing, and publishing effort. The net result is a system that enables expansion of partner programs, improved audience use and satisfaction, and optimized revenue growth.

In the various case studies that focus on software documentation, organizing the content as components allows the documentation updates to be synchronized with the updates of the software modules they support, improving customer access to timely information.

For government organizations and related stakeholders, componentization enables a highly distributed authoring process with more concurrent updating and automated assembly. This allows content to be prepared rapidly, to be easily and quickly assembled and updated, and to be automatically published in multiple formats.

Traditionally, many types of information were created as complete documents since, ultimately, they would most likely be printed as such. Over the years, as electronic delivery has increased, so has the realization that storing information as documents was constraining to organizations needing to deliver in electronic forms such as the web.

Mapping Components

Information broken into individual objects or topics can easily be reassembled into the document form, or several wildly different document organizations as needed. As shown in Figure 6, a component library can be used to produce two different types of manuals.

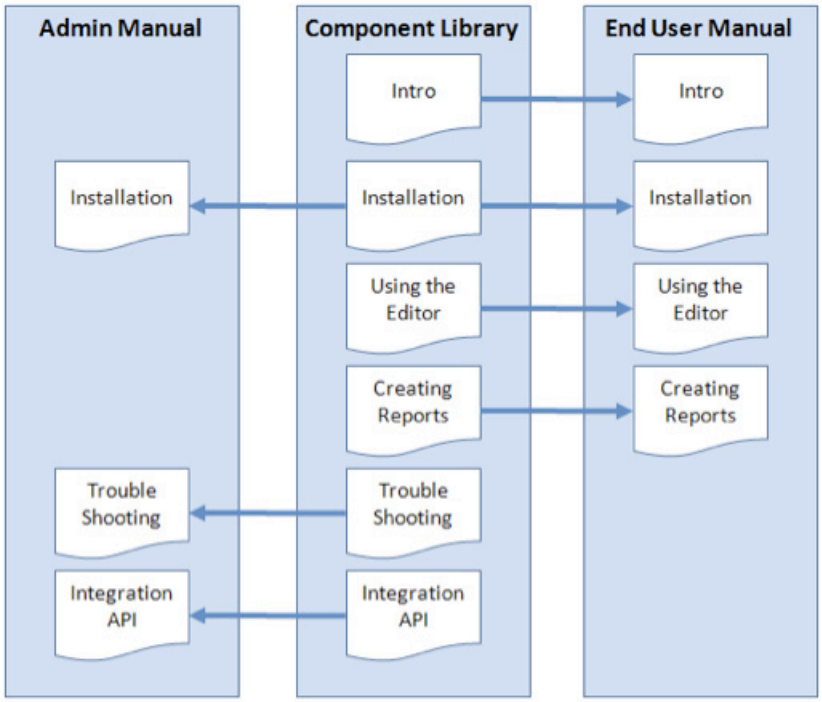

Figure 6. Component reuse for different views

Stand-alone content components can be assembled into different document "views" intended for different audiences, where some information may apply to one or more views.

Many organizations have segmented their documents into individual topics or components in order to gain these efficiencies. But simply chunking content is rarely sufficient. Each component needs to be designed to support content delivery and consumption requirements, and, therefore requires them to be easily identified with unique names or IDs, and with additional descriptive metadata and properties. Appropriate chunking of content in this way also enables more accurate application of metadata since the components are standalone topics rather than a mix of concepts usually included in more narrative document oriented content.

When content is managed as components, it is essential to have a mechanism for assembling these components into deliverables such as documents, help files, web sites, and other delivery forms. Business requirements drive the organization of the delivered content. The assembly process relies on a map of the assembled content components used to guide the automated assembly and delivery of the content.

For example, DITA-based systems use a separate document called the DITA map that contains references to all the components in the sequence and hierarchy of a specific delivery format. A different DITA map can reorganize the same content components in a significantly different form. Componentized content will most likely require an assembly mapping method to create the various required views and achieve the intended benefits.

Other tools have similar mapping capabilities stored as configuration files or build lists. When someone accesses a web page, it may contain a reference to the map and automatically pull in all the components according to the rules specified in the map. Similarly, composition systems can automatically assemble and render a set of content components according to a content map. Manual cutting and pasting of content for reuse should be relegated to environments with extremely low volume or complexity.

Content Components Require New Ways of Writing

One example of writing differently for components is the standard style for resolving acronyms, where the first occurrence of a term would be spelled out with the acronym in parenthesis, and all subsequent occurrences would simply only show the acronym. In a book made up of a series of topics, we only have to spell out the term in the first section. But in a group of stand-alone (and self-contained) topics, we may need to spell out the term at the beginning of each topic.

Also, some of the text in each stand alone content component may need adjusting, either to remove or reword cross references, or to add introductory information previously appearing in a different section of the document. Similarly, in a document, we can make references to an "earlier" or "later" section. But with components that may be organized into several different structures, we may not know if the referenced section is earlier or later and should avoid that type of terminology in references.

Components that are created and edited by different people may inadvertently end up with slightly different writing styles or "voices." Successfully editing components requires writing style guidelines and policies to provide a consistent voice when they are assembled into a deliverable document.

In many of the case studies for our report, we found that the content development teams formed governance committees to develop guidelines for ensuring a consistent writing style and voice. In some organizations, documenters have shifted away from a single author developing an entire document to teams that collaborate on creating content components, together with style guides designed to keep the content consistent. The other organizations we interviewed created similar governance teams and processes.

Content Enrichment

A key difference between a structured document and a smart document is the degree of descriptive information provided along with the actual content. There are several ways to enrich content including:

- Using descriptive tagging focusing on what the information element is and not what it looks like or how it should be processed
- Including metadata, such as topical keywords and identifiers, to aid in processing and searching the content
- Using additional maps (e.g., DITA map) or look-up tables to drive content integration and assembly for producing multiple "views" into the data

Descriptive Content Models and Element Names

A powerful form of content enrichment is the use of descriptive tags and attribute names that tell what the information is, not what it looks like or how to process it. HTML is a commonly understood markup system that works well in formatting content in a browser. But HTML is not very good at describing the intended use for each information object.

For instance, when a software manual is marked up in HTML, the sections for "Installation Instructions" and "Trouble Shooting" may both have a <h2> tag applied to them. This format-oriented generic markup does not help to identify the intended use of each section and is not sufficient to include/exclude that topic from a specific audience view. More information is needed to tell the two sections apart.

There are several approaches to enriching content.

- A program can parse the words and phrases in the content to distinguish among identically tagged elements. While becoming more proficient over time, language-based processing is often ambiguous and prone to errors.
- Adding additional markup with specific key words to the HTML may help differentiate the two sections (e.g., RDF attributes)
- Using an <installation> element for one section and <troubleshooting> for the other would distinguish the two, albeit creating a set of tags that are not as widely understood as HTML markup.

A powerful method for clearly identifying the role of the content is to use a standard data model, such as a schema with predefined elements and tag names that are widely understood and supported in software.

For instance, DITA defines four basic component types, <topic>, <task>, <concept>, and <reference>, each with its own set of subelements. And, therefore, a system processing DITA content may be able to process a <task> into an interactive checklist used to track steps taken in a procedure much more easily than generic HTML.

```
<?xml version='1.0'?>
<task id="installstorage">
  <title>Installing a hard drive</title>
  <shortdesc>Open the CPU case and insert the drive.</shortdesc>
  <prolog>
    <metadata>
      <audience type="administrator"/>
      <keywords>
        <indexterm>hard drive</indexterm>
      </keywords>
    </metadata>
  </prolog>
  <taskbody>
    <steps>
      <step>
        <cmd>Unscrew the cover and remove it.</cmd>
        <stepresult>The drive bay is exposed.</stepresult>
      </step>
      <step>
        <cmd>Insert the drive into the hard drive bay.</cmd>
        <info>Do not force it into place. Adjust the angle until
          it slips in easily.</info>
      </step>
    </steps>
  </taskbody>
  <related-links>
    <link href="installmemory.dita"/>
  </related-links>
</task>
```

Figure 7. Sample of descriptive tag names

Figure 7 shows richly tagged DITA content using element names that are easy to understand and interpret (by both humans and computers). Other specific, standardized data models are used in various industry verticals and domains as well. This descriptive markup removes some of the ambiguity in the data and can be used for many processes, from formatting for print or the Web, to generating a "wizard" that guides a user through each step.

Without rich, descriptive markup, a manual process would most likely be needed to transform this content into an interactive dialogue. Also, the richness of the markup helps identify that this is a task and not a reference or other type of topic, and that the task is identified with an audience element to be directed at system administrators, not end users. These clues would not be apparent in HTML.

DITA is just one example of a standardized data model and vocabulary used for marking up content. Pharmaceutical content preparers may use one of the standards for describing their types of content (e.g., SPL, PIM, etc.), while the financial industry may use a different standard (e.g., XBRL). DITA tends to be useful to a broad range of vertical industries and content types and is widely supported by software vendors.

Several of our case studies describe systems that use the DITA document model and tag names. It is already understood by most vendor processing tools and is very descriptive and easy for users to understand. An organization that has adopted DITA does not require any specializations and is able to use DITA elements as defined in the standard.

At IBM, the standardized element names allow some content to serve as both documentation and as a configuration file that is loaded into the manufacturing equipment to configure the processing. Previously this information had to be rekeyed instead of loaded, which introduced a higher risk of keying errors and required more set up time.

Metadata and Taxonomies

For the purposes of this discussion, we will refer to any additional information that supports and improves identifying, processing, selecting, or searching content as metadata. Even the properties of flat files, such as date updated, author, file type, can be useful in processing identifying and information components, and would qualify as metadata in this context. It is also common to include some metadata in the actual text of the information, albeit sequestered in a markup designed to identify these elements.

Figure 8 shows a small XML document with a block of metadata included in the tagged markup:

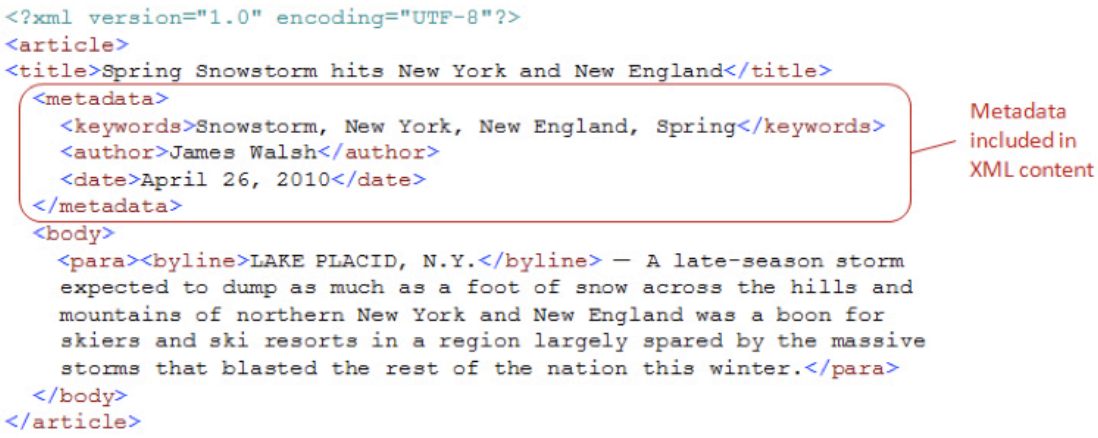

Figure 8. Sample metadata element

Of course, metadata can also be stored outside of the content in a content management system or other storage areas, similar to file properties in a file directory. In Figure 9, we see a hypothetical dialog used to assign metadata values to content components being managed in content management system.

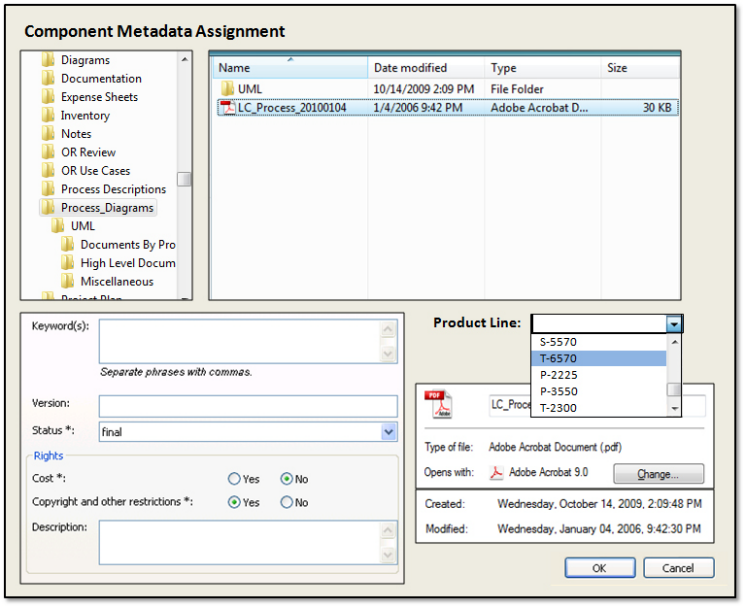

Figure 9. Assigning metadata explicitly

Many organizations create controlled vocabularies and taxonomies for use in categorization of content. Systems that use these standardized taxonomies are able to create enriched content that is more easily understood by other systems and tools that also utilize these standards. Even so, many organizations have created their own taxonomies for products, departments, information types, and other metadata that can support automated processing, assembly, and search optimization.

Enrichment Through Crowd Sourcing

Crowd sourcing, or letting the users of the information add to it and enrich it, is common on the web. For instance, restaurant review sites allow the users of the information to add their comments and reviews. The same goes for electronic products, hotels, music, and many other areas. Why not other types of content produced by businesses and organizations that address the needs of specific audiences of all types?

First, a small internal production team working on aggregating many content sources may not be able to perform even a fraction of the enrichment without adding significant internal resources and schedule delays. A large audience of users enriching the content may be much more scalable and timely. Also, the audience of users has an inherent interest in the content being correctly enriched to make it more usable to themselves and other users.

Secondly, crowd sourcing may remove obstacles to enriching content. For instance, government agencies may produce the raw information, but may not be allowed to add qualitative reviews or comments without being seen as partial to one source over another. Third parties may provide a fairer assessment of the services or good listed in the content to be enriched.

Lastly, crowd sourcing enables the content to be enhanced continually and prevents it from going stale. Increasingly, web users have come to expect content to be refreshed

frequently, to be integrated in a variety of ways, and to be interactive with commenting and other features.

Crowd sourcing has some challenges, such as managing and editing comments to prevent misinformation and malicious comments. Even so, there are approaches that address and minimize these challenges. A common technique is to require the content enrichers not be anonymous, by requiring user ids and profiles before commenting. This encourages better behavior and can allow the content provider to follow up on comments, while still allowing only the first name or user id to be displayed along with the comment as needed. In more critical situations, the crowd sourced enrichment comments and metadata can be edited by the content aggregator (of course this requires editing resources).

Crowd sourcing offers promise and has great potential to shape the way we deliver information going forward. It breaks from some traditional publishing concepts and limitations, and adds others that leverage the power and ubiquity of the web. The results may be sophisticated smart content applications that cannot be delivered in other ways.

Dynamic Publishing

Dynamic Content Assembly

Dynamic publishing is often used to describe processes where content is assembled "on the fly" at run-time, rather than being organized into a document in advance. This can provide significant benefits in terms of timeliness of content and custom content delivery. In addition, dynamic publishing may include concepts such as variable content for different audiences, real-time updates of content, and content repurposing.

Dynamically assembled content processes can deliver the most current version of all content components when accessed by a user or produced as a hard-copy (printed) document. This is similar to the way some web content is delivered, where some portions of a web page query data and return a result set that is then sorted by date, topic, or other parameters. Each time content is accessed, a more current view may be delivered if it is frequently updated. The content selected and delivered may be based on predefined topics, user's rights to various types of information, or other variables. The end result is intended to be a more relevant set of topics for a specific audience.

Simple forms of dynamic content assembly have been in use on the web for some time, but more robust capabilities are now possible with the use of a content management system or other content technologies. These tools and systems may use metadata or other selection criteria to dynamically assemble the result sets.

Componentized content, such as DITA topics, work well in a dynamic assembly environment. DITA topics define the stand-alone building blocks for certain types of information, and are usually enriched with metadata. It is the combination of stand-alone data objects, metadata, and selection criteria, and a robust assembly engine that provides the greatest degree of control for dynamic publishing.

As described in several of our case studies, a dynamic publishing system allows authors to work independently on content components which are managed in a central repository. End users for this information also access the shared repository, via a web browser, to view the published results, which are automatically organized into subject-oriented topics. This powerful automated approach has reduced considerably the amount of time and effort required to transform a collection of stand-alone objects into a sophisticated library of information.

Content Variations and Customization

Another nuance to dynamic content delivery is the use of variable data. In DITA, these data variables are called Content References, or CONREFs. A CONREF is an area within a DITA topic that may have two or more versions, each one intended for a specific audience or use. They include information about the intended audience used to select it at the appropriate time. Other systems and approaches may use variables buried into content, similar to mail merge functions in word processors.

CONREFs allow whole paragraphs, or even larger chunks of content, to be managed as variable content within a topic. CONREF is a reuse mechanism that allows content to be pulled in from external sources. CONREF and conditional processing can be combined to incorporate (or aggregate) different content for different audiences. However, conditional processing also works directly on content within a single topic by marking up elements that apply to specific contexts so that they can be filtered out for other contexts.

The combination of dynamic assembly and variable content provides a very strong mechanism for customizing content for specific audiences. Variable content and dynamic assembly enable special configurations of content to be produced automatically and avoid a lot of manual work to achieve the same result.

Figure 10 below shows how variable content within content components might be turned on and off for various audiences. When the three topics are accessed or assembled, the specific content variable could be displayed for each version of the delivered content -- one for Macintosh users and another for Windows users.

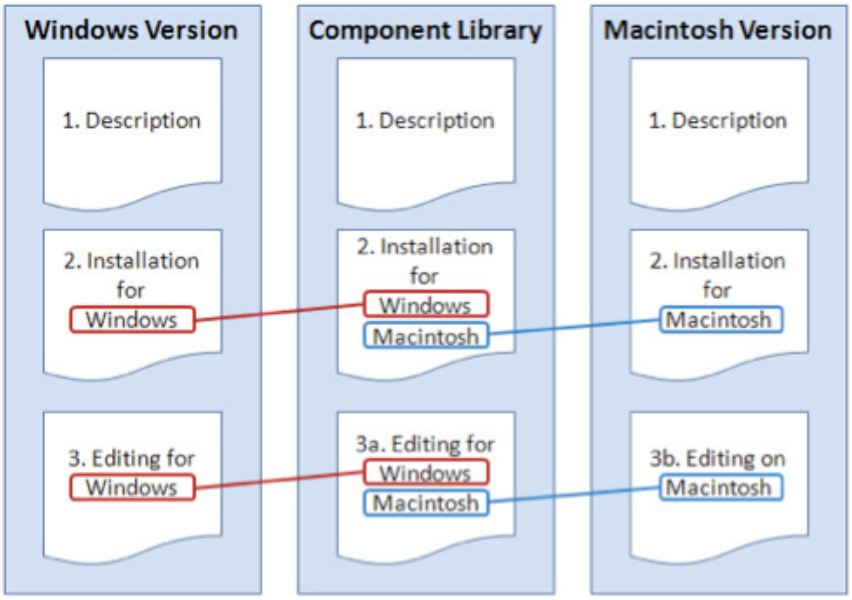

Figure 10. Content variables and versions

It is true that two different topic components can be produced for each component that has a variable in it above. Content variables are useful when only small differences appear in a component that is mostly identical. If two individual components are created instead, there would be more editing, proofreading, and other work needed than in this approach where one component is created with a small amount of variable content.

Smart Content In Practice

Our insights and analysis are only part of our research agenda. We also want to capture the view from the trenches – what we term smart content in practice.

As we described at the beginning of this report, we set out to take a careful look at what is actually happening within leading edge organizations. We wanted to focus on how practitioners are transforming the concepts of component content management into operational projects that deliver business results. Key to our research, we interviewed experienced XML practitioners to learn how they addressed the challenges of managing and delivering high-value content in today's ever more digital business environment. From this effort we gleaned insights into the industry's pioneering experience with smart content.

We are grateful to our study sponsors. Their willingness to connect us with their customers was key to the success of our research project. We have benefited enormously from listening to stories from operational and product champions about the experiences they are creating for their customers and stakeholders.

Why these stories and to whom did we talk? We asked our sponsors to introduce us to one or two of their key customers -- organizations currently using their products and solutions to develop XML applications. We then interviewed both the technical and the business leads responsible for these projects. Depending on the organization and the project scope, our case studies are based on conversations with anywhere from one to five practitioners.

When we conducted the interviews, we asked practitioners a consistent set of questions about:

- The types of content they are publishing and managing
- The target audiences they are trying to reach
- The tools, systems, and platforms they are using to manage XML tagged content
- The business goals and outcomes driving their adoption of XML tagged content
- The ways in which their organizations are able to change and improve their content delivery activities
- The challenges of achieving wide scale adoption of XML tagged content across their organizations

As we analyzed the interviews, we identified the themes for deploying XML tagged content within a business context and for harnessing various application capabilities. Once we drafted the case studies, we fact checked them with the interviewees to produce these published results. Of course, we realize that our case studies are only snapshots about these organizations and their uses of content technologies.

As part of our investigation, we focused on identifying the best practices within the business context – what practitioners needed to do to transform content development

activities and deliver identifiable value to their organizations. We are publishing these case studies as a series so that you, our readers, can easily identify the common themes and practices among them.

In this section, we include six case studies that highlight how XML practitioners are developing and deploying smart content applications for business results. These are the stories that are now ready for publication. But this is only the beginning. We are also working on several additional case studies and will publish them as an update to this report before the end of the calendar year.

Stay tuned and join in. We believe we have identified an essential market trend. We expect that there are going to be continuing conversations about smart content in the enterprise for many years to come. We welcome your contributions.

Optimizing the Customer Experience: Facing the Challenge of the Web

Founded in 1989, Citrix is a leading provider of virtual computing, cloud computing, and networking solutions, delivering on demand IT services to customers ranging from small and medium-sized businesses to large enterprises. Like other software firms, the company develops and delivers product documentation and training materials in print and electronic forms. Technical content includes product documentation, user manuals, administration manuals, and "read me" documents. Training content includes printed and online courseware materials.

Over the years, Citrix developed extensive sets of technical documents and training materials using conventional editing, formatting, and production tools. A single Citrix product may have more than 30 individual documents, all produced electronically and stored as PDF files. With worldwide markets, documentation sets are frequently translated into seven or eight languages. The company also invests in developing its own courseware and produces upwards of 2,400 pages of training materials annually.

By the mid-2000's, Citrix realized that its customers expected to find technical information over the web. However, the company's tried and true documentation and training practices lagged market demands and customer expectations.

While many of the technical manuals and training materials were posted online as self-contained documents, customers had difficulty navigating through the collections and finding answers to their questions. Moreover, once they published content for print distribution, editors and courseware developers were manually transforming and repurposing content for delivery over the web – adding extra time and expense to production processes. Citrix lacked automated capabilities to reuse content.

A Modular Solution for Component Delivery

Adopting DITA and an XML Editor

Citrix needed to streamline the production of its technical documentation and training materials and make the content useful to customers over the web. To gain production efficiencies and foster reuse of content throughout the organization, Citrix decided to componentize its content into self-contained chunks, and to adopt DITA as the XML tagging standard for documentation-related content.

Both the technical documentation and training teams have adopted the XMetaL XML editor from JustSystems for creation and editing of content topics. However, the two teams are managing components in different ways.

- Content for technical documentation is stored in a shared repository. XMetaL is tightly integrated with the repository using customized adapters.
- Training materials are managed with a revision control system that maintains and updates flat files within a host file system. (The training team is exploring the adoption of a content management system.)

Both management environments provide sufficient flexibility in how the content is created and updated. Both approaches use XMetaL's validation capabilities to ensure properly structured DITA content. Both manage the content in discrete, topic-based chunks instead of complete documents or chapters. In addition to XML validation, style guides, naming conventions, and extensive templates guide the writing and tagging of content. PowerShell scripts (a scripting utility that comes with Microsoft Windows 7) provide additional feedback to authors on style inconsistency and other errors.

The result is a consistent and well-structured set of content components that can be assembled into various "documents" and delivery formats. Citrix also is considering additional feedback and content verification tools and techniques to further streamline its content creation processes. For instance, on the technical documentation side, authors can use XMetaL Reviewer to compare versions of content to increase collaboration and provide additional feedback during the authoring process. The modularity of the content components allows specific topics to be reviewed by the appropriate subject matter expert.

Implementing a Component-based Solution

Citrix began working in a DITA-based system in 2007. Creating and managing content as components required new processes, review steps and tools, and better standards for content consistency.

For the training materials, for example, courseware developers no longer had to cut and paste content from print versions to create online versions of content. For technical documentation, editors had to create the processes used to produce a new HTML version of content delivered in addition to the PDFs delivered previously. Since work is shared across authors, rather than having a single author create an entire document, style guides and rules governing content collaboration and review also had to be developed.

For the technical documentation, migration of content was done manually at first, beginning with the existing structure of the documentation sets. Once the rules for mapping from one form of markup to the newer DITA markup were better understood, the documentation group began to use some automated conversion tools to create the DITA components. Even then, all migrated data required extensive review and verification to ensure that it was converted correctly. Over a period of three years, the various document sets were migrated to the new system and processes.

Enriching Content to Enable New Products

For the training materials, Citrix took a workshop approach that provided training to new authors while converting content from legacy tools to DITA. Citrix took the

opportunity of moving to DITA to refresh the look and feel of its classroom training guides. For online learning content, Citrix first converted the DITA content to standard HTML tags. Then a secondary pass was used to wrap the HTML content in SCORM (Shareable Content Object Reference Model, a collection of standards and specifications for Web-based e-learning), and add metadata and other content types and components common to SCORM training materials (e.g., rich media, navigation structures, etc.).

The enriched smart content enabled products to be assembled based on topics and other metadata, help content creators discover existing topics instead of rewriting duplicate content, and create a variety of versions to support specific customer use cases. These activities are more streamlined in a smart content approach than the previous, largely manual processes, and have been instrumental in new opportunities and growth.

Changing Roles and Working Smarter

Citrix found that moving to a DITA-based solution is really part of a bigger shift in the way the company operates and delivers content. Both within the documentation and training groups, roles have changed. Rather than assign a single publication to an author, teams work more collaboratively as workgroups. Processes had to be revised to reflect this collaboration and to define style guidelines to keep materials consistent with each other. Scheduling work is easier since smaller updates are required. Citrix has even formed an Information Architects Council that addresses design and process standardization and to facilitate the use of best practices. In general, less redundant work of assembling pages is required, and users can focus more on quality and enhancements to the data.

The Consequences of Componentization: Improved Management and Delivery

Previously, technical documentation had been made available only as stand-alone PDF documents. A user would have to access several documents and search each one individually to see if it contained the topic or information he or she was looking for. In the new approach, a user can search across a group of topical information and have individual topics returned relating to the search criteria, eliminating the redundant searching of the PDF files.

In addition, since content is updated in smaller topic chunks, it is kept more current and is refreshed more frequently. This allows Citrix to keep the content synchronized with software updates more easily, and to plan the update and delivery the content on a timely basis.

For training materials, the focus is more on operational efficiency. End users do not experience much of a difference in how the materials are delivered. But the processes used to produce these materials have been streamlined considerably. Training materials can be kept current with the software updates more easily because they are updated in small topical chunks.

Also, during the migration to a smart content solution, Citrix discovered that componentized content can lead to new revenue opportunities, including shorter versions of training that can be given away free and used to generate new training sales leads.

Other groups within Citrix are now watching the migration to DITA-based publishing and the successes the documentation and training teams have demonstrated, and are considering adopting similar processes and tools. As the structured editing and component content management tools mature and become simpler to use, expanding the use of DITA will be much easier.

Smart Content Insights

Citrix has adopted XML in various ways through the organization. Authors and editors in the technical documentation group, as well as courseware developers in the training group, can now focus more on the quality of their content rather than on processes associated with page and document assembly.

Content reuse is now a reality. Teams have the content management capabilities to easily produce multiple language variants of the same product documentation sets. Content can be easily shared between teams and used in its native format.

Finally, well-structured and well-tagged content can now be easily indexed and retrieved by web-wide search engines. Citrix customers can easily find authoritative content, produced by the company, when searching on Google.

Content Delivery: DITA components provide flexibility for organizing content, reducing the time it takes to assemble multiple versions of the same content topics for different audiences. This allows new products to be created to meet a broader set of user requirements, such as variations and versions of training materials and technical documentation. Also, modularity and the use metadata allows improvements in search results and ease of searching across collections of documents.

Content Enrichment: Citrix has found that content enrichment, such as adding descriptive DITA tags and other information, enables the content modules to be assembled automatically, eliminating manual repurposing of the content. Also, the enriched modular content has optimized searching and improves the customer experience.

Content Creation: Using a DITA-aware editor and managing content in a more granular data model is easier and more timely than starting with a print-oriented formatting editor and later reworking the content for Web delivery. Validation and other quality assurance tools provide feedback to the author and allow richly tagged data to be created very quickly and accurately.

Documenting Semiconductor Devices at IBM: Addressing Design Complexity

Following the logic of Moore's Law, the IBM Microelectronics' Semiconductor Research and Development Center (SRDC) designs ever more sophisticated semiconductor technologies. As a major supplier in a global marketplace, IBM works closely with multiple Alliance Partners to co-develop technologies and bring new devices to market.

By the mid-2000's, SRDC leadership realized that IBM's then current procedures for documenting semiconductor design guidelines lagged its innovations in technologies and business processes. IBM needed to address the growing complexity of its technical design manuals.

Information Flows for Technical Documentation

Semiconductor technology developers specify the parameters for electronic components that are etched into semiconductor wafers. Developers pass their design specifications to technical writers who produce the design manuals for a technology. The manuals contain tables of values, rules, diagrams, and textual descriptions about the technology. Hardware devices are created as specializations of various components within the technologies, defined by specific values and rules.

Semiconductor design engineers, in turn, use the information contained in these manuals to constrain their chip designs to meet the requirements of the fabrication processes. To assemble, layout, and validate an entire chip design, these engineers need a set of rules for their devices, together with the technical descriptions about how they perform under varying test conditions.

With the ever-increasing densities of semiconductor technologies, design manuals were growing in scope and page counts. Delivered as self-contained, PDF-formatted documents – more than 500 pages in length, containing 60 to 70 percent tabular data, and often updated every couple of weeks -- they were rapidly becoming unmanageable. IBM needed to develop new methods for technical documentation without simply adding writers and increasing development costs.

Information Flows for Alliance Partners

In addition, Alliance Partners contract with IBM to fabricate specific devices and are entitled to access only certain sets of technical information. SRDC controlled the information flows by producing a base manual for each technology, and then a series of addendum manuals for developing specific devices.

Design engineers from partner firms needed to know when to replace the rules and values in the base manual with those in an addendum. Looking to the future, with the

ever-increasing complexity of technical information and the additional page counts, SRDC managers and technical leaders realized that this fixed-content approach to documentation was unsustainable.

Content Components for Documenting Devices

Mirroring Chip Designs

SRDC leadership adopted a component-oriented approach to documenting devices that mirrored the chips themselves. A semiconductor technology encompasses many electronic components, documented by associated content components.

Rather than continuing to publish multiple manuals about a technology, SRDC leadership decided to:

- Decompose the semiconductor design documentation into its component parts.
- Map the electronic components to the associated content components.
- Dynamically assemble and publish relevant content components to document specific devices.

Henceforth, Alliance Partners would only receive the information to which they were entitled, with the appropriate tables of values and rules interleaved into a single manual. IBM would structure the technical documentation to recognize different technology components and the needs of various Alliance Partners.

Furthermore, to address the consequences of Moore's Law, SRDC leadership decided to change the way developers and writers would work together to author content. Rather than passing their design specifications to technical writers, technology developers would use an editing tool to record the information in a predefined, structured fashion. The technical writers and documentation specialists in turn would be responsible for the integrity of the overall manual and collaborate closely with the developers on assembling the content components.

Finally, the technical information needed to be consumed both by people and by automated processes. The manuals would be published not only as PDF-formatted documents but also electronically as XML-tagged information, accessible either as files or served to design engineers at Alliance Partners through web services.

Over a period of 18 months, SRDC leadership proceeded to develop and implement

- An information architecture for documenting semiconductor components.
- An enterprise content management (ECM) infrastructure to support its component-oriented approach to technical documentation and delivery.

SRDC leadership relied on DITA (Darwin Information Typing Architecture) to define the information architecture, and IBM Filenet P8 to provide the ECM platform for managing it.

Structuring Content Components with DITA

SRDC leadership uses DITA to define the content components for technical publishing. Every electronic component in a technology has an associated content component, defined as a DITA topic.

Each device has a DITA map, which is a list of all the content components that need to be published in a manual. The technical manual for a device can be dynamically assembled from the collection of content components on demand. IBM can thus publish technical manuals for its various Alliance Partners, documenting just the devices within a technology that they are entitited to receive.

Managing Content Components

However, by adopting DITA, IBM faced an exponential growth in the number of content components it needed to manage. SRDC managers adopted IBM FileNet P8 to manage the content components, as shown in Illustration 1.

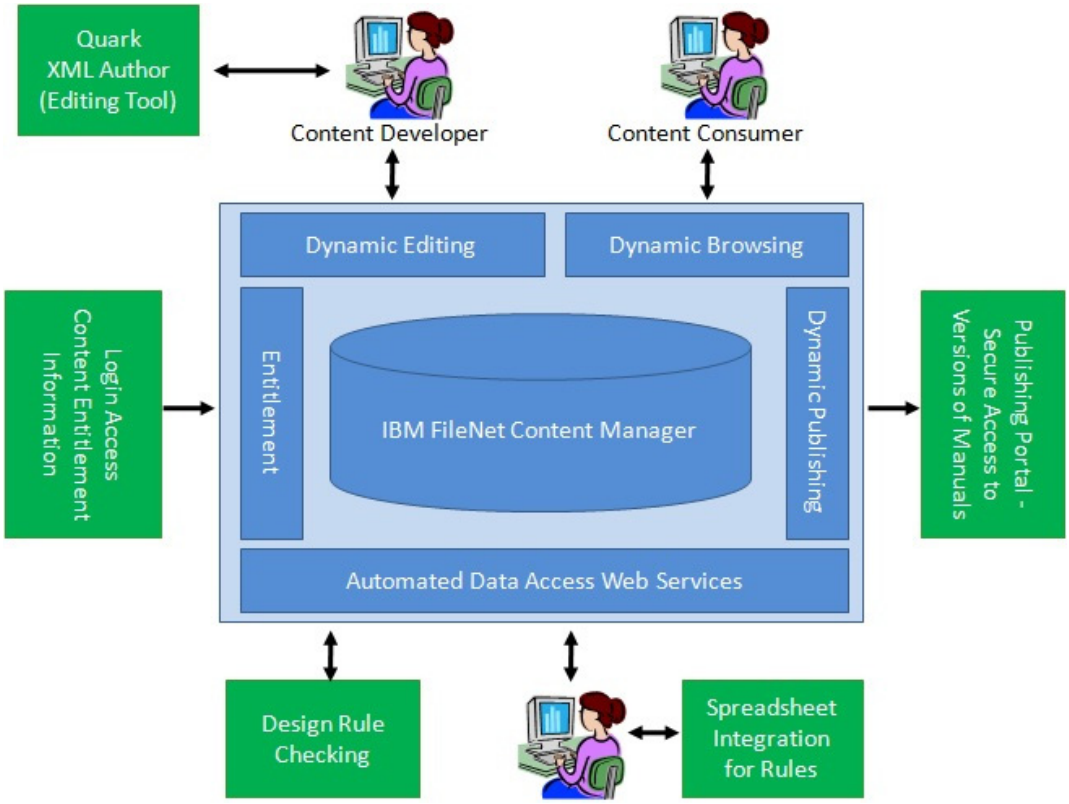

Figure 11. IBM FileNet P8 as shared repository

This platform provides native support for DITA, including:

- A metadata model within the object store that is DITA-aware.
- DITA schemas that are part of the object store.
- A specialized DITA classifier that extracts properties and DITA relationships when DITA files are ingested into the repository.

- A rendering engine integrated with the DITA Open Toolkit.

FileNet P8 manages the end-to-end publishing process and supports core ECM functions including security, search, component lifecycles, review and approval, and centralized publishing.

Authoring Content Components

Finally, IBM modified the authoring processes by adopting Quark XML Author (QXA), an easy-to-use XML authoring tool from Quark. Semiconductor technology developers began to use this tool to document technology components on their own, without having to directly involve technical writers for making individual changes.

Developers use a template to key-in content. (In the near future, they will be able to forward their additions or changes to a workflow queue for others to review.) The QXA template adds the appropriate XML tags to DITA content components. Technical writers no longer have to rekey requests from developers. Writers and documentation specialists are better able to manage the authoring processes and keep up with the rapid growth in semiconductor technologies being documented.

How DITA Fuels Growth

Making the Transition to Dynamic Publishing

The technical documentation group at SRDC is in the midst of making the transition to DITA and dynamic publishing. For the past two years, documentation specialists have been migrating current semiconductor design manuals to DITA. This has required some cleansing and adjusting of the technical information to manage it as content components.

By January 2010, more than 20 semiconductor technology engineers are actively involved in authoring content. The overall goal is to spread this capability to many more engineers and designers within the SRDC during the calendar year, and provide them with the capabilities to update content on their own. IBM recognizes that this will require strong management control, due to the complexity of the technical information.

Reducing Bottlenecks

With the adoption of DITA and FileNet P8, there are fewer bottlenecks. Designers create and edit the technical information on their own, without having to communicate the rules and values to technical writers. For instance, a designer with appropriate access rights can check out three columns of a table for editing, make the changes in a familiar desktop editing environment (using QXA), and then submit the updated columns to the ECM repository for subsequent review and approval.

The document owners for device manuals can now focus on the modular content design. The new documentation processes allows them to focus on value added tasks, such as ensuring that the right content components are assembled into a published manual, rather than the manual drudgery of rekeying information.

Publishing with a Laser Focus

The end-result is the promise of publishing semiconductor design information not with a fire hose, but rather with a laser focus. Each device within a technology is now documented on its own.

With the move to dynamic publishing, Alliance Partners receive all of the information they need to create the design of a device as a self-contained manual. Semiconductor design engineers at partner firms no longer have to compile the relevant information on their own, checking information in the base manual with additional rules and values in the customized addendums. Rather each device has its own design manual, published both as a PDF document and as an XML tagged file.

Thus there is also the option of end-to-end publishing. Design enablement engineers no longer have to interpret the rules and values contained in the various tables to manually create and re-create the tools used by circuit designers. Rather, they can automatically capture the rules and values for the DITA tagged content components, and use them to automatically generate these tools. This is an important step for automating the creation of design enablement tools and infrastructure to accommodate Moore's Law and fabricate ever more dense devices.

Finally, there is the business outcome. Rather than simply cutting costs and optimizing the documentation processes, IBM can also accelerate its semiconductor research and development efforts, and make them more profitable. IBM can now slice and dice its technologies in more fine-grained ways and add more Alliance Partners, capable of producing unique devices. In short, IBM can better leverage its research and development investments in semiconductor technologies.

Beyond the Initial Deployment

This case study illustrates how an enterprise-scale platform can spread beyond its initial deployment. There are three enabling factors:

- An information architecture.
- A scalable infrastructure.
- A business justification.

An eighteen-month upfront design and development period was required to chunk existing technical manuals into their component parts, to define the DITA tags, and to map content components into predefined technical manuals. As an ECM platform, IBM FileNet P8 provided the management framework and content storage infrastructure. Quark QXA adds end user tools that help to transform authoring and editing activities.

The end result was (and is) a documentation process that can scale up to accommodate additional Alliance Partners. IBM can substantially enhance its documentation processes as a supplier of semiconductor technologies to the chip design industry.

Smart Content Insights

As a solution to the growing complexity of semiconductor design manuals, DITA works. Facing the consequences of Moore's Law, IBM maps the electronic components within a technology to an associated set of content components. DITA standardizes the definitions of the content components and the maps for publishing customized manuals.

But the adoption of DITA raises an important issue: how to scale the publishing environment. Rather than producing discrete manuals containing hundreds of pages, IBM needs to manage a very large number of content components and dynamically publish them as needed. IBM's approach is to rely on an ECM platform, FileNet P8.

Content Delivery: By defining content delivery requirements up front, IBM's publishing strategy is aligned with its business strategy. IBM dynamically publishes manuals for Alliance Partners and provides them with just the design information they are entitled to receive. It is essential to match the granularity of content components with the granularity of semiconductor devices.

In addition, the information within a manual is no longer captured by a predefined format, and published only as a PDF document. IBM also publishes XML-tagged files where the content components can be automatically incorporated into external tools, such as those required by semiconductor design engineers for the design of microchips. It is now feasible to rely on DITA as a standard for end-to-end publishing in machine-readable formats.

Content Enrichment: IBM relies on DITA to enrich the content components through a series of specializations – beginning with the definitions of "topic," "task," "concept," and "reference." There are a series of predefined XML tags within resulting sets of hierarchally-defined components.

Enrichment begins with an upfront design. IBM benefits from an explicit and detailed information architecture. Once the content is enriched through DITA, IBM is then able to exploit its investment by introducing an ECM platform that can effectively manage a very large number of granular content components in a systematic manner.

Content Creation: By adopting DITA and focusing on content components, IBM is able to transform its authoring and editing processes. With access to an easy-to-use editing tool, semiconductor designers key in DITA tagged content that flows directly into the technical manuals. Technical writers and editors are responsible for the overall form and substance of the manuals, and managing the overall processes.

It is essential to rely on the capabilities of the underlying ECM platform for structuring the flow of content and managing the very large number of content components. It is also essential to provide end users with easy-to-use editing tools when authoring DITA tagged content.

Towards Smart Publishing at IBM: Facing the Challenges of Technical Documentation

Continuing Industry Leadership

IBM is no stranger to the challenges of publishing technical documentation and spearheading the adoption of industry standards to solve them. In the 1960s, as businesses began to depend on computers to perform routine tasks, the company pioneered the development of the Generalized Markup Language (GML) for formatting text with then available technologies. Twenty years later, as the initial generations of distributed computing environments took shape, IBM contributed personnel and resources to define the Standard Generalized Markup Language (SGML), and helped to make it an industry standard for producing electronic documents. Then came the web and the promise of providing just the right information at the right time to solve business problems.

To help exploit the power of the web to deliver targeted information about specific subjects, IBM introduced the Darwin Information Typing Architecture (DITA) in March 2001, and worked with other groups to shepherd its adoption by the Organization for the Advancement of Structured Information Standards (OASIS) as an industry standard in June 2005. DITA defines content components as XML-tagged topics, and describes the maps for collecting and publishing this tagged content in both human-readable and machine-consumable formats.

Furthermore, DITA specifies an extensible content architecture for creating new types of topics and maps by developing specializations of existing types. Designed initially for technical documentation, DITA contributes key elements to the definition of a modern content infrastructure, suitable for web-wide access and distribution, including componentization, reuse and repurposing, content-independent navigation, and targeted content types. From our perspective, DITA is an example of smart content in operation – content components tagged with extensive metadata where the information can be dynamically delivered through various interactive environments to mold customer experiences.

Business Drivers for a DITA Solution

But a standard is only one step towards a solution. As IBM customers increasingly rely on the web for doing business in the digital age, they need online access to targeted information for solving problems. Customers are no longer satisfied with paging through sections of lengthy documents, published either electronically or in print. They expect a seamless experience with relevant content immediately at hand.

IBM, in turn, has had to respond by paying greater attention to how customers use its technical information and by developing new methods for managing content. With more than 1,500 authors creating over one billion words a year and supporting translations into more than 40 natural languages, IBM faces a large-scale problem to satisfy customer expectations.

To overcome a series of product content silos and improve the effectiveness of technical information delivered to customers, IBM has developed a component content management platform, Information Development Content Management System (IDCMS). It is designed to support dynamic information delivery and improve the automation and integration of tools and processes across IBM's multiple hardware, software, and services organizations. As an enterprise solution, IDCMS seeks to reduce the costs of developing technical information and enable business agility by providing a platform that supports content reuse, drives dynamic publishing, and helps to automate end-to-end translation processes.

Delivery As the Starting Point for Application Design

Managing Content Components

DITA changes the scale and scope of technical documentation tasks. Once published as lengthy books and manuals, technical information now needs to be decomposed into granular content components. Authors, editors, and solution designers need to respond accordingly, by developing the writing processes, supported by the appropriate technical solutions, for managing content components in a systematic manner. There is a transition from a linear document to an interactive online paradigm. For information designers and implementers, it is all a matter of finding the right level of granularity and enrichment.

Let's consider the experience of an early adopter of IDCMS and DITA. A team of writers, editors, and an information architect continue to produce the technical information for IBM's Customer Information Control System (CICS), a product first brought to market in 1968. CICS is still widely used in large organizations for online transaction management and connectivity for mission-critical applications. CICS documentation has steadily evolved over the life of the product, most recently from 49 discrete technical manuals (published in print and as electronic documents) to over 35,000 DITA topics, organized into 40 different information units and now distributed online through the CICS Information Center.[3]

Adopting DITA

The CICS documentation group started the move to DITA in early 2004. Making the transition from book-oriented content (tagged as SGML) to DITA-defined content

[3] See http://publib.boulder.ibm.com/infocenter/cicsts/v4r1/index.jsp

components was an iterative process. While DITA standardized the object definitions and tag names for componentizing content, human judgment was still required to tailor information for online distribution. Beyond simply developing and publishing the technical information, it was important to define how customers would use it, and then enrich the content components with relevant tags to enhance usability.

When evolving from a linear document paradigm to customer-centric content components, the CICS documentation group faced a series of design challenges. It needed to:

- **Define the right level of granularity.** Componentizing content needed to be based on meaning, rather than predefined formats. While SGML formatting tags served as a guide, such typographical elements as headings, paragraphs, and lists needed to be grouped into logical components or discrete topics.
- **Identify discrete topics.** With a book-orientation and continuous text, there are often mixed topics. For example, a set of problem-solving steps in a published manual frequently combined a list of steps with an explanation about why they were important. Customers were familiar with scanning printed pages to pick out just the information they needed. For component-oriented delivery, information needed to be focused on discrete topics – one that contains the step-by-step instructions and another, referenced by a link, the explanation about the whys and wherefores. When working online, customers have fixed (and short) attention spans, so it is essential to provide focused and targeted information.
- **Clarify context.** When published within a book, words themselves were often ambiguous and derived meaning from their location or context. For instance, a section about "Recommendations" in an installation manual had an implicit context. When published online, the context needed to be explicitly identified and tagged as "Installation Recommendations."

Moreover, with the emphasis on content delivery, components needed to be enriched with summary information and other appropriate metadata. For instance, each component included a "short description" field that encapsulated the entire topic in a single sentence. This short description field could then be part of the customer experience for content delivery – including the phrases displayed via online hovers and the tagged information made accessible to search engine web crawlers.

Benefits and Implications

The CICS team has realized substantial benefits by adopting DITA, including reusing content through content references (tagged by such criteria as product names and versions), richer tagging, and additional flexibility with organizing and categorizing content. When using IDCMS (described below) team members can easily identify relationships among content components (including cross references and external references to graphics), capabilities that are useful for determining broken links and orphaned items. Thus the team relies on this DITA-aware content management system to facilitate effective content delivery.

In short, the CICS team has faced a substantial set of tasks to redesign how it produces and delivers technical information. The end result is well-tagged (and richly described) content components, defined at an appropriate level of granularity. This team's experiences are hardly unique. Other IBM software and hardware product groups are facing similar issues when componentizing technical information and making the transition from linear documents (designed for book publishing) to interactive online resources.

At IBM, adopting DITA has been a multi-year, iterative development effort. It is essential to design and add tags for content delivery, and to characterize how customers use content for business purposes. Content enrichment becomes an integral part of the information development process.

A DITA-Aware Platform

Moreover, enriched content must also be maintained. Managing DITA-defined content components on an enterprise-wide scale requires a comprehensive content infrastructure. This is where IDCMS fits in.

Based on IBM's FileNet P8 enterprise content management (ECM) system, IDCMS provides an XML-enabled content component management platform. IDCMS features unique content component handling and interpretation capabilities, in addition to familiar (and expected) content management, access control, and reporting capabilities. All information design and development activities are organized around a shared repository, which then builds, syndicates, and publishes content to a variety of customer-facing environments (see Figure 3).

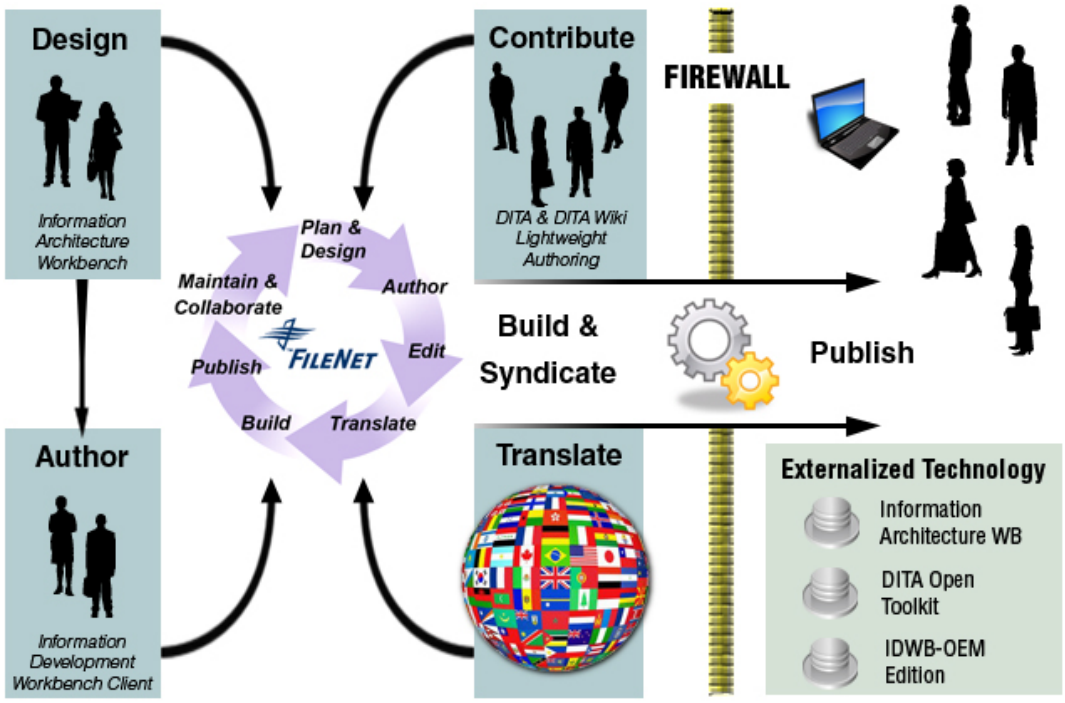

Figure 12. IDCMS relies on FileNet P8 as a shared respository for DITA content

Source: IBM

IDCMS is DITA aware. With the relational schema embedded within FileNet P8 optimized for processing high volumes of components, the platform readily understands how to ingest and manage DITA tagged content. IDCMS is integrated with several popular XML editing tools. Authors and editors use "smart forms" to automatically identify and add the DITA tags to content components. The editing tools hide the complexity of the underlying content components.

IDCMS provides native support for DITA components. As content components are ingested into the repository, IDCMS automatically interprets the DITA tags and metadata, and then categorizes and organizes the information within the repository. IDCMS thus populates the indexes that enable discovery, retrieval, and extraction processes. Rights management and workflow processes provide controls to manage creating, updating, and publishing content components to a variety of output formats and delivery channels. IDCMS also provides integration features to support federated data access across distributed workgroups.

Standardizing the Content Infrastructure

The Business Benefits of Information Typing

IDCMS thus demonstrates the importance of a component content management platform, based on a standardized content infrastructure. DITA provides the common currency for structuring content components; as a standard, it is the means to an end rather than an end in itself.

While DITA enables targeted content delivery, authors, editors, and information architects still need to define, design, and deliver the essential elements for the customer experience. With IDCMS, they can manage the content components and the associated metadata in a consistent fashion. IDCMS enables the auto handling of the metadata and the ability to identify the properties inside each type of topics. Some of the metadata can be related to the customer-centric (and customer displayable) characteristics, while other metadata relate to application calls or display parameters. With a component content management platform, contributors have fine-grain content management capabilities. They can easily maintain the parent-child relationships among components and readily identify container topics and dependencies. These are useful capabilities for molding the customer experience.

From Topics to Embedded Semantics

For IBM, the topic paradigm -- with an emphasis on tasks, concepts, and references -- represents a new and innovative way of thinking about delivering content. With DITA, the information architect designs how and what information should be created and organized for use by the customer. When adding or updating information, authors are freed from the limits of a book-orientation and from having to define where in a linear document to make the additions. Authors can concentrate on their subject areas and not be concerned about the overall organization and display of content. They can focus

on making the content comprehensive and accurate. IDCMS then automatically produces the updates and additions within the relevant DITA maps.

But DITA-defined content components are only as smart as the embedded tags. Currently DITA defines the topic structure – metadata is available at the topic level for such factors as product version, release, and keywords. There are only limited capabilities to produce intuitive customer experiences where components are linked together based on their meaning. IBM is exploring a variety of approaches to rely on DITA for enriching components with extensive semantic metadata.

Embedding semantic metadata in DITA topics and DITA maps significantly enhances discovery, retrieval, and content reuse. When included within an information architecture, DITA-aware content repositories and DITA-aware content delivery systems can extract and use that metadata to improve content creation, content management, and content delivery. For instance, content can include important identification and taxonomy metadata such as subject classification, version/release level, product and feature identification -- all of which can be extracted and mined as a properties for search, retrieval, reuse, workflow automation, custom and personalized assembly, and delivery. Using DITA as the trusted source of semantic metadata ensures portability and system independence for the life of the content.

As a result, recognizing the use of structure and adding semantic metadata to content types will produce added intelligence and an increasingly "smarter" environment with more options for automated and semi-automated processing. It is important to leverage the capabilities of an underlying standard for defining and maintaining content components – in terms of this IBM example, DITA.

One of the issues designers are facing is how to define a set of semantic tags in the first place. Within IBM, top down taxonomies are difficult to maintain beyond the boundaries of product groups. Content enrichment needs to be based on understanding customer needs – for example, tagging content based on knowing what search terms customers are most likely to use. Different customer audiences use different terminology; it is important to enrich content by adding synonyms as metadata. Thus there are many opportunities for collaborative development of taxonomies and controlled vocabularies within task groups and business teams.

Smart Content Insights

With its investment in DITA as a standard and the adoption of IDCMS as a component content management platform, IBM is changing the paradigm for technical documentation from books to buckets. Instead of publishing linear documents, suitable for technical manuals, technical documentation teams produce content components that can be automatically mapped into multiple customer experiences.

IDCMS, in turn, provides the essential content infrastructure for delivering on the promise of DITA – the ability to maintain multi-purposed content to drive multiple customer experiences. IDCMS supports the management capabilities required for maintaining the granularity of content components and for mapping components to customer experiences. IDCMS illustrates how added intelligence (in the form of DITA

tags) can be incorporated into structured content tools. It is all a question of defining and managing the relevant tags and metadata.

Content Delivery. For IBM, the customer experience is the starting point. The information architect, a new role within IBM's information development organization, owns the customer experience. IDCMS is designed to make it easy for the information architect to manage the customer experience and to coordinate the work activities of authors and editors who create and update the content in the first place.

Content Enrichment. DITA standardizes the tag sets and maps for topic-oriented enrichment. DITA provides a flexible framework that enables a broad range of content types to be enriched and classified, and therefore referenced and assembled into specific maps. IDCMS provides the technical framework for content enrichment. DITA tags are maintained within the content components and used to automatically index information maintained by the IDCMS repository.

Content Creation. The key step is being able to manage and maintain granular content components in an easy-to-use yet systematic manner. IDCMS provides the underlying repository and is integrated with popular editing and content creation environments. Authors and editors use familiar – frequently WYSIWYG – editing environments, and rely on templates plus intelligent drop-down tagging environments to manage their interactions with the relevant tag sets. They can easily manage content components and links, and quickly identify problems such as components with missing links and orphans. By integrating with easy-to-use editing environments, IDCMS features the capabilities to guide the authors and editors on how to best apply DITA tags to optimize delivery for the customer experience.

Single Source Publishing at NetApp: Adopting an Infrastructure for Content Reuse

The Growing Need for Product Documentation

NetApp Inc. is a leading provider of storage and data-management solutions and hardware for a wide range of business systems and processes. With a diverse product line including data center and application storage solutions, the company delivers more than eighty new and enhanced product releases each year, all of which require updated documentation. Delivering high quality technical documentation is key to the company's business success. Doing so for such a large company is a daunting task, one that structured content is designed to address.

The Information Engineering group, a centralized documentation team, is responsible for producing documentation for multiple NetApp business units, covering all the company's products. Approximately 80 people in six locations in the United States, Canada, and India produce documentation on installation, configuration, administration, support, and troubleshooting, of these software and hardware products.

The documentation team has faced an all-too-familiar problem. There has been significant growth in the number and versions of products the company produces, requiring a significant growth in the amount of feature-related content and the need for specialized documentation types. Without the ability to grow headcount infinitely, the team has had to develop a new methodology to keep up with the growing demands of the company.

Toward a Modular Solution

Beyond Sequential Publishing Processes

Business pressures on Information Engineering led the NetApp management team, in late 2006, to the decision that the company needed a new approach to replace its redundant and lengthy documentation processes. The company needed to produce both PDF and HTML documents through a single, integrated process, and thus save time and money. NetApp management also wanted to reduce redundant content authoring and editing through a modular content reuse strategy.

At the time, the NetApp documentation team was using dedicated and separate tools for specific outputs, and storing the results as self-contained files within a file system. There were separate PDF files for electronic documents and print, and HTML files for

online delivery via the web. The content was organized as documents. A single writer was responsible for each title.

Even though writers and editors recognized the redundant efforts, the traditional tools and processes did not readily allow for easy content reuse between output types—for example, between online help and reference manuals. In many cases they rewrote and reedited the same information, with subtle differences in details due to variations among product lines and output formats for different documents.

Adopting DITA

Just as "interchangeable parts" revolutionized manufacturing, NetApp wanted to migrate to a more modular, consistently structured environment where the information contained within documents could be easily reused to reduce redundant content development. At the same time, the company wanted to ensure that customers would continue to receive the same level of detail and completeness of information that they had come to expect. The company needed to adopt more efficient processes while keeping the process changes opaque to customers, and visible only to the NetApp staff. This was no simple goal to achieve.

NetApp could envision how single source publishing and content reuse would improve the overall quality of the technical documentation produced. Writers would no longer have to rewrite and update existing content. Rather ,they could devote more time to developing new topics and identify new ways to reuse their existing topic content.

For the Information Engineering group, the solution entailed topic-oriented authoring where each topic was defined and tagged as a self-contained content component. NetApp selected DITA (Darwin Information Typing Architecture) as the predefined standard for structuring technical information and defining content components. NetApp compared DITA to other standards such as DocBook and realized that since DITA was more modular, it would achieve higher reuse and provide for more flexibility and modularity. For this reasons, and because DITA was beginning to be recognized as a standard among the technical documentation community, the Information Engineering group decided to adopt DITA.

To move to DITA, NetApp chose to have the team rewrite existing content directly in DITA, rather than do a massive automated conversion. This method would be the best way for the writers to learn the new topic-based methodology. Team members first reviewed their existing documentation sets, and then componentized the content into self-contained and meaningful parts. DITA provides the framework to map content to a modular structure and easily allows content to be associated with conceptual structures such as product components, features, and tasks. DITA provides the tags to structure and enrich the content components.

At first, the group did not have a content management system (CMS) and relied only on the file system for storing content. Team members found this method challenging to manage and maintain the thousands of versions of a large number of content components.

Managing Content Components

As the NetApp team moved into authoring topics, they realized that they needed a system to manage all the variations among components and relationships with one another. Such a system would have to be specifically designed to handle DITA and the challenges of technical writing. After reviewing options in the industry, writers and editors within the Information Engineering group adopted SDL Trisoft, a component content management (CCM) solution designed specifically for technical publications organizations. Team members also rely on the XMetaL Editor from Just Systems for XML editing, Antenna House for composition, and SDL Global Authoring Management for grammar and style validation.

The information resides within a shared SDL Trisoft repository stored natively as DITA content components, giving all team members access to the current content components (and prior versions if needed). SDL Trisoft functions as a hub that enables team members to centralize and coordinate their content creation, versioning, enrichment, and production processes. SDL Trisoft keeps track of how various versions of content components are related to particular product releases.

Changing Authoring and Editing Processes

A best practice process change that accompanied the transition to DITA is the development of specific areas of expertise. Instead of being responsible for a particular product publication, writers are now typically responsible for specific topic areas. Furthermore, a subset of writers is responsible for the overall validity of a complete manual. Writers creating a topic may not even be aware of where the content will appear eventually.

Since content is shared and reused across products and concurrent product releases, the team developed practices to work more collaboratively. This includes standardizing editorial workflows and ensuring consistency in writing style. Collaborative writing practices are particularly important as writing team members are often based in several locations around the world.

Maintaining the DITA Map

While authors write stand-alone topics, they can also see the tree structure of the map in which the content is reused. Thus writers can easily view the context where their content is going to be used. A writer will create versions of a specific topic to align with the software release schedule for that feature, keeping the software and supporting documentation more closely in line with each other. The SDL Trisoft Publication Manager application provides an overview of the document as a whole and enables writers see information such as the relevant topics and versions, including the workflow state in the update cycle.

Currently at NetApp, more than 1,400 DITA maps referencing roughly 18,000 discrete topics are used to create 1,350 individual documents (and several trial documents). The Information Engineering group has developed best practices to ensure content quality and consistency in a modular writing environment. Collaborative document planning

and deliberate design of the underlying information architecture is essential for creating compelling customer deliverables.

Benefits and Impact of the New Approach

With DITA, the Information Engineering group now manages the production of content in a methodology that mirrors software development, with its ability to maintain code modules in a source control system. This has the benefit of making it easier for Information Engineering to keep pace with software development releases.

With the success over the last few years, NetApp managers now understand the benefits of managing content components. Modular information source, coupled with SDL Trisoft use, enables content developers to be more agile in a product delivery schedule. Features and functionality can be marked and included or excluded late in a product-development lifecycle. The flexibility of easily recombining topics into different maps enables Information Engineering to more quickly release documents that match the expanding and new functionality of a product as it moves toward full release status.

Also, the new approach enables rapid prototyping of new or highly customized deliverables for special audiences that would not have been feasible under the legacy system. DITA's underlying flexibility enable the organization to meet faster product release schedules, develop more variations of product content, and tailor those deliverables to the interests and background of their customers.

New Writing Processes

NetApp has had several years of learning from its experience of writing content in modular form. Authors and editors have needed to adjust their writing styles.

Since they no longer create content in isolation, writers must adapt shared best practices for collaborative content development. DITA requires writers to deconstruct information into content types. For example, DITA separates task-oriented and procedural information from conceptual content. Writers also have to envision how the information will be used in one or more publications, and often write generically to support all uses.

Some writers have transitioned to the role of "document captains" and are responsible for organizing topics into publications. DITA maps are created for each product deliverable and product view.

Semantic Enrichment

DITA has also enabled the content to be prepared in a way that optimizes searching for specific content. Topical and other descriptive metadata from each topic can be added to the content components to enrich their definitions and thus improve the search precision of the HTML output.

Improved search results should allow more customers to successfully resolve issues without resorting to NetApp technical support. In the future, the Information Engineering group plans to focus more attention on semantic enrichment.

Expanded Uses for Technical Information

NetApp has not traditionally shared its product content with reseller partners. Now with the flexibility of creating special reseller versions, and with resellers themselves adopting DITA, the company is already benefitting from DITA as the *lingua franca* for its information supply chain partners. NetApp can provide its resellers with content in DITA and they can rapidly rebrand and reuse the content in their OEM solutions. Modular content enables business agility and provides a firm with the ability to rapidly respond to new opportunities.

The DITA implementation at NetApp continues to evolve as the Information Engineering group expands its understanding of the power of component content development. The Information Engineering group has formed a governance team to evaluate and communicate best practices for authoring and managing content components, based on DITA. The vision for the future is to dynamically deliver technical information based on customers' profiles, and to use a methodology for capturing user-generated content that can be subsequently incorporated into technical documentation.

Smart Content Insights

Single source publishing and content reuse, based on DITA-defined modules, provides the foundations for competitive advantage in an information intensive firm such as NetApp. But adopting DITA, while necessary, is not sufficient. It is also important to adopt a CCM system to manage all of the content components in a systematic manner. With the infrastructure for single source publishing in place, an organization has the agility to rapidly respond to new business opportunities.

Content Delivery: Moving from a document orientation approach to one that focuses on topics provides extensive flexibility in how the content is delivered. With componentized content delivery, a company can support rapid product release cycles and changing customer requirements. Metadata, combined with discrete topics, provides the building blocks and details needed to access and assemble specific content components, dynamically publish documents, and deliver a rich user experience. Web-wide search engines can detect the tagged information and use it to optimize search results.

Content Enrichment: Content enriched with topical metadata and managed in a component content management system provides an efficient means of assembling product and training content. DITA provides a predefined set of tags that are optimized for technical publishing and content reuse. Furthermore, descriptive metadata, including such factors as product line and documentation type, can help improve search and navigation, as well as planning and tracking product development.

Content Creation: The value of the modular DITA-based approach is evident to the content creators as well as others in the organization, especially engineers used to working with and assembling independently created software modules. Marketers also appreciate having access to specific information not buried in large publications.

NetApp is expanding the scope of content contributors to include engineering and others throughout the organization and to leverage the information they create for use in support documentation. Roles need to change when moving to a componentized content approach. Document captains are needed to facilitate collaboration and document consistency.

Symitar Solutions for Credit Union Management: Documenting Software Modules

Symitar™, a division of Jack Henry & Associates, Inc.®, is a leading provider of core processing software solutions for credit unions. Symitar's Episys® solution consists of a suite of software components, based on a modular design, that allows the system to be deployed in custom configurations that best meet each credit union's operational requirements.

Software modules provide key functionality for teller operations, reporting, online banking, system installation and setup, trouble-shooting, and many other processes used to operate and manage credit unions of all sizes. Modules encompass business functions that are readily understood by customers -- the staff members of credit unions using the solutions. Symitar produces documentation in HTML and PDF to support users of their products. The company also provides credit unions with training materials for tellers and other employees who use the Symitar system components.

The Support Department at Symitar creates and manages over 40,000 pages of documentation. The support and training documentation is delivered and consumed as PDF documents and in the HTML help format called CHM (for the .chm file extension). CHM is a compiled set of HTML pages with generated navigation aid (e.g., table of contents and linking). Symitar has developed and enhanced the CHM files delivered with custom features that integrate the help file information with training materials to improve the usefulness of the materials and the overall customer experience.

Toward a Modular Solution

Improving Customer Service

In 2005, Symitar management began looking for ways to improve customer service by providing better documentation. At the time, the company produced large, monolithic documents for each product and each type of guide. These documents were comprehensive and self-contained. They were designed to provide everything one needed to know about a particular area of the system. However, they were not organized around performing specific tasks.

Symitar decided to adopt a modular approach to its documentation and training materials, and restructure them into a set of modules that corresponded to the software modules (or subsystems) themselves. The company focused on content reuse at the

module level. The company sought to provide the documentation that corresponds to the system configuration deployed by each credit union.

Modeling and Prototyping with DITA

Symitar embarked on a transition toward modular, integrated content tied to the software development process. Technical Publications, a unit within the Support Department, began by prototyping several approaches to modeling and managing content. After considering various approaches to structuring XML tagged content, the group leaders decided to adopt DITA (Darwin Information Typing Architecture).

Organized around topics, concepts, tasks, and references, DITA provided sets of predefined content component objects and tags. Content was segmented into modules addressing specific system components and operations. Content components were then created or updated in parallel with system components. DITA could be readily adapted to Symitar's need to document predefined software modules. No specialization of DITA components was required.

Technical Publications at Symitar took time to prototype new ways to organize the content. Transitioning to the modular topics used in a DITA-based architecture required thoughtful reworking of the content into self-contained modules. Some rewording and style changes had to be made to the content along with chunking it into logical topical divisions.

DITA also required planning the navigation methods and links that the end users needed to consume the information. Technical Publications adopted a task-oriented approach to documenting software capabilities, and sought to improve the clarity and effectiveness of the writing.

Master Topic Areas

Over several years, Technical Publications restructured the product manuals into 90 master topic areas that cover the entire system functionality and administration. Some functionality is common to all configurations of the system (e.g., "logging in"). Documentation for all configurations of the system shares a single topic for these common features, while other topics may have several variations to support the different configurations.

This strategy emphasizes reuse of content and eliminates much duplication. Also, some topics are similar, but not identical, and may have slight variations for specific product lines. Strategic use of conditional references for subtle variations, termed CONREFs, allows a balance of customized output and minimal redundant preparation of content.

An XML Editor and a File System

Symitar has deployed JustSystems' XMetaL XML editor for creating and updating its content using the DITA data model without any specializations. XMetaL was selected because it is similar to the tools used previously to edit the content, and the learning

curve for content creators would be reduced significantly. XMetaL is also used to create and edit the DITA maps.

Symitar's content is currently stored as flat files in the file directory. The Information Architect and three technical writers have migrated certain document sets to DITA by focusing on document type and audience. They began with the user guides, and then migrated the programming guides, back office support guides, and troubleshooting guides.

Symitar has created a large collection of topics, with multiple views into the content for each deliverable product that may share content with other products. All new content for these document types are developed in DITA from scratch. Release notes are still created in the legacy system but will eventually be migrated to the new environment.

Improved Efficiency and Higher Quality

Content Reuse

By reorganizing product information into reusable content components, Symitar has been able to develop a well-integrated and accurate knowledge base, from which the bulk of its product documentation is now produced. The DITA-based system is both more efficient and produces higher quality support documentation.

Many content components are now reused in multiple product manuals. Some of these even include subtle variable information specific to product line or even jurisdiction in which the credit union operates. Even when managing flat files in a directory, Technical Publications is able to streamline editing processes and free up writers to do value-add content enhancement. Symitar has been able to eliminate some of the reformatting and cutting and pasting required for content reuse in the old approach.

Managing Content Components

Currently, document content is managed with a system of "outrageous spreadsheets," as described by Kathryn Showers, the Information Architect leading the DITA adoption efforts. She also leads initiatives to establish best practices and policies for creating and maintaining the content.

Symitar is considering adding a content management system in the future to further streamline processing, add workflow tracking, and improve content management and navigation. The CMS is expected to provide search tools to improve researching topics for inclusion in product guides. The CMS will also help authors and editors to tie content development activities more closely to the software development cycles they are designed to support.

Combining Product Documentation and Training

Ultimately, Symitar defines success for this content and the system in terms of reducing or eliminating technical support calls. The new approach has also brought the documentation and training teams much closer together. While it began as grass roots

evangelism, the DITA-based system has expanded into a leading example of best practices across the organization. In addition to the technical publications and training groups, other groups within Symitar are looking at expanding their approaches for producing software design and test documentation, and are expecting comparable benefits. Disparate groups are now able to more closely work together, and to better coordinate their business activities.

Smart Content Insights

Symitar has found that migrating to topic-oriented modular content helps to improve the quality of documentation and align results towards meeting business objectives. More is required than just defining and developing topic modules. Also essential is a thoughtful information architecture, writing guidelines, and a strategic view of how the information is going to be used.

With a well-defined approach to content components, one system can address the requirements of another. At Symitar, the success of DITA for product documentation leads to its adoption for training modules – and the merger of two groups into a more comprehensive one that focuses on customer information.

The combination of DITA, good governance policies, and an information architecture has better aligned various teams toward meeting the underlying business goals. Content reuse is the organizing principle driving the transformation of Symitar's product content practices.

Content Delivery: The granularity and structure of content depends on careful consideration of how the content is consumed and how it is best organized to solve a specific problem. Consumers of the Symitar support and training documentation are provided with enhanced content that is organized to teach or support specific tasks, improving their ability to perform their work without relying on technical support calls. The customers are more efficient and Symitar is able to reduce its technical support costs.

Content Enrichment: Symitar enriches content (defined by DITA topics) at the module level. Content components map to software modules. By including topical information and metadata, and organizing content into clearly identified topics, Symitar improves its documentation processes, and facilitates its ability to reuse content components in multiple manuals.

In addition, content enrichment is spreading. Other groups within Symitar are learning from the experiences of Technical Publications about the business benefits of DITA, and enriching content components with various kinds of semantic and syntactic information.

Content Creation: Writing for stand-alone modules requires consistency of voice and care in creating complete, useful information objects that can be combined into integrated guides and training materials. Collaboration across product content requires governance of processes and style. Management of content in modules increases the number of objects being managed and can quickly become difficult without the aid of a

content management system and workflow tools. Fortunately, with DITA-aware editing tools, some of this complexity is made manageable.

The Warrior Gateway and the Power of Social Publishing: Supporting the Military Community

Across the country, many government agencies, military programs, and local organizations offer medical, mental health, employment counseling, and educational services to the military community – soldiers, veterans, and their families. Yet for returning warriors seeking to reintegrate into local communities, making sense of all these services and finding the right ones over the web is often hard. There are few digital resources that not only aggregate content about disparate services, but also make the information useful to community members.

The Warrior Gateway (http://warriorgateway.org) is designed to fill this void. Seeking to support the military community members reentering civilian life after deployments in Iraq and Afghanistan, the Warrior Gateway collects, organizes, enriches, and redistributes content about a wide range of health, welfare, and veteran-related services. It also serves as a resource for employers wanting to hire veterans and for organizations seeking to engage with the community. It appeals to a digitally savvy audience, comfortable with the web and social media tools.

More than simply an online catalog of service providers' listings, organized by topics and locations, the Warrior Gateway restructures the content that government, military, and local organizations produce, and enriches it by adding veteran-related categories (e.g., Veteran's Administration offices and services, mental health and physical therapists, etc.). Furthermore, the Warrior Gateway adds a social dimension by encouraging contributions from veterans and family members. These "voices of veterans" include short comments about the quality of listed services, ratings, online reviews, and moderated forums, together with capabilities for organizing and tagging content.

The Warrior Gateway thus augments government, military, and local organizational efforts, by adding insights and opinions from the community participants to published information. For veterans and their families, the appeal of the Warrior Gateway is curated content from multiple sources, organized into a consistent set of online resources, and supplemented with advice, commentary, and categories from communities of contributors.

Aggregating and Socializing Content

A Web Framework

From a technical perspective, the Warrior Gateway is a content aggregation and integration environment, based on a model-view-controller framework and hosted in the cloud, utilizing Amazon Web Services (EC2). Initially developed within two weeks by a four-person team, the Warrior Gateway is continuously enhanced through an agile development process that quickly incorporates audience feedback to add new features and functions.

The Warrior Gateway currently collects content from over 30,000 service providers. It is expanding its aggregation efforts and, by the end of 2010, expects to encompass over 50,000 providers. Government, military, and local organizations publish their content in many different ways, with varying degrees of accuracy and currency. The Warrior Gateway seeks to capture and normalize the essential elements of these listings, and provide links to the remote resources.

Essentially, the Warrior Gateway is a social publishing platform that combines the curated content from authoritative sources with the user-generated content and categories provided by soldiers, veterans, and family members. Initially, the Warrior Gateway deployed network-crawlers to crawl remote service providers' web sites, extract relevant content and links, and aggregate the results within its own repository. Depending on the granularity of the content published on remote sites, the Warrior Gateway captures varying levels of unstructured and semi-structured information. With access to granular content and self-describing XML tags, the crawlers can automatically retrieve detailed information that providers publish to describe their services.

Once stored within the Warrior Gateway, the content is reorganized and reclassified to provide the veterans' perspective about areas of interest and importance. Soldiers, veterans, service providers, and others who want to contribute to the Warrior Gateway can add comments and ratings to refine the information on services listed in the directory. Volunteers working with Warrior Gateway can scrub the data and add new classifications when necessary. Service providers can claim their profile and improve their own data details.

With contributions from multiple stakeholders, this form of social publishing allows data to be enriched over time without requiring a large internal staff to add the extra information. It also recognizes the natural interest of veterans and service providers to make the information as accurate and complete as needed to meet their needs.

The Smart Content Infrastructure

The Warrior Gateway relies on the MarkLogic Server to deliver the technical capabilities for its content infrastructure. Specifically, the MarkLogic Server provides a flexible database that can easily manage large volumes of disparate content sources, including structured content (e.g., HTML, XML, etc.) as well as common unstructured formats (e.g., PDF, word processing documents, spread sheets, etc.). The MarkLogic Server

powers the agile development processes. It provides search and analytics functionality to easily navigate through large content collections that are distributed over the web. Systematic indexing of these disparate sources together with XQuery-based tools provide end users with easy access to the content, organized and accessed in consistent and useful ways.

The MarkLogic Server also supports geospatial index information, enabling content to be classified and delivered according to location. Geospatial information drives innovative delivery options, such as displaying services geographically on maps, to help veterans locate services and enable agencies to plan coverage.

Finally, the MarkLogic Server is used to select and transform the content stored within the repository into specific views and formats, which are then syndicated to external web-based resources. The Warrior Gateway, with its sophisticated content management, indexing, and transformation capabilities, can serve as a master resource for information about veteran-related services, and can easily redistribute content to third-party environments.

As a result, service providers only need to update their information once within the Warrior Gateway, and can then syndicate it to additional veteran-related sites with no extra effort. Various organizations can rely on the Warrior Gateway to store and provide access to their own content, enriched with categories, comments, and ratings that the community members generate. By redistributing directory ratings and other information from the Warrior Gateway to their own networks of stakeholders, service providers can thus leverage the enriched content without duplicating development efforts.

Improving Consistency and Accuracy

By design, the content is managed in the Warrior Gateway's central repository, rather than simply linked from remote sites, as shown in Figure 1. This improves consistency and accuracy, but also maintains references and links to the originating sites. This approach also presents challenging logistics for capturing and enhancing the collected information.

Warrior Gateway initially used automated load processes and content crawlers to automatically aggregate and enrich content. Now the information can be uploaded by service providers or entered directly using dialogs or software provided by the service. Content can also be syndicated from other sites and search queries that feed information to the Warrior Gateway.

The Warrior Gateway provides a social publishing platform that captures contributions from throughout the military community. Much of the work of providing, organizing, and enriching the information is done at the grassroots level by veterans and service providers, rather than by the internal staff. The user dialogs allows *ad hoc* updating and classification of specific directory entries over time as the information becomes available.

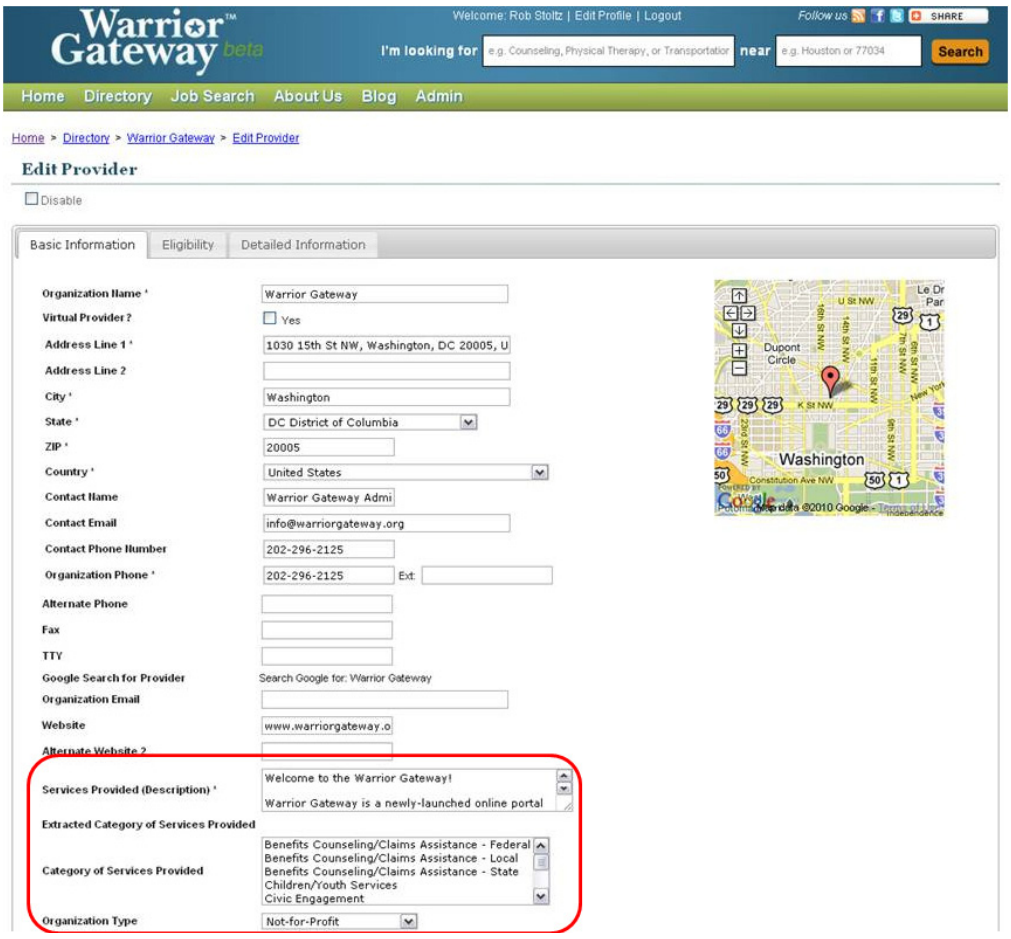

Figure 13: Social publishing web dialog

The Consequences of Crowd Sourcing

Aggregating and organizing content in a consistent way and adding a social dimension to the information makes it easy for users to find what they want and need. Augmenting the content with metadata using easy-to-use tools enables targeted searches by topic or location, as well as innovative mashups with mapping and other third-party resources. Adding user comments with social media tools helps retiring warriors evaluate the services described and make informed decisions on the services they seek.

The end results are more useful and valuable information resources than would otherwise be available. There are several additional benefits enabled through the use of crowd sourcing.

First, there is the power of network connections. Only a small team is needed to support the core platform and provide the framework that facilitates content aggregation and social publishing. Service providers join the effort and make their information available to veterans and their families. By providing the tools and services to enrich content contributed by others, the Warrior Gateway manages web-wide resources for spreading the word and amplifying the contributions of local organizations.

Second, crowd sourcing removes many of the obstacles to enriching content. For instance, government agencies may produce the raw information, while third parties add assessments and evaluations of the listings. The end results are more informative online contributions than the information produced by any single source.

Third, crowd sourcing enables the content to be continually enhanced. Increasingly, web users have come to expect content to be refreshed frequently, to be integrated in a variety of ways, and to be interactive with commenting and other features. A large audience of contributors enriching the content helps to keep it timely and relevant.

Finally, once content is collected and enhanced by multiple contributors, new uses for the information and new services for the user are possible. Warrior Gateway can help identify current gaps in services based on the needs of the military community located in a particular geographic area. The site can help reveal areas that are underserviced or in need of a particular service or specialty, as well as those areas that may be over served based upon the community needs and population size.

For example, using metadata that identifies all the services from a single region, the Warrior Gateway can produce a map showing where all the services are located in a veteran's area. Or, the site can list certain types of services being sought, such as physical therapists, psychologists, or even day care and lawn mowing. The result is a rich, flexible, current, and accurate resource for soldiers, veterans, their families, and other stakeholders within the military community.

Smart Content Insights

The Warrior Gateway represents an exciting and innovative approach to collecting large amounts of disparate information, in a scalable and feasible fashion and within a limited budget. The Warrior Gateway also illustrates how social publishing enables content contributors and various service providers participate in enriching and classifying content to improve its relevance for soldiers, veterans, and their families. Lastly, the enriched content can be easily repurposed (or multi-purposed) and/or queried on other sites and even combined with additional content and applications in mashups, therefore making Warrior Gateway content even more valuable and accessible to the military community it serves.

Content Delivery: Service providers can have their information delivered in a highly targeted manner, by geospatial coordinates, service types, or through other types of enrichment techniques. Content that includes rich metadata can be sorted, organized, and searched more intelligently, repurposed and combined in ways that unstructured or more generic content cannot. At the Warrior Gateway, searching becomes much more accurate and likely to return relevant results. Powerful query tools and rich smart content enable flexible delivery and syndication features.

Content Enrichment: The Warrior Gateway contains not only listings of services, but also descriptive metadata that is used to provide powerful and accurate delivery to the returning warriors. Content can be enriched using a combination of automated and user generated methods including social media tools without requiring an internal team to add to and classify content.

Content Creation: Tools such as web forms and dialogs ease the collection and enhancement of content to create a robust data source. Automated collection and syndication tools ease timely and high-volume data gathering. A robust repository that can manage large volumes of heterogeneous information provides the platform for aggregating and integrating smart content.

Sponsor Acknowledgement

The Gilbane Group appreciates the contribution of content for this section from our study sponsors.

IBM (Gold Sponsor)

IBM Enterprise Content Management helps companies make better decisions faster by managing content, optimizing associated business processes and enabling compliance through an integrated information infrastructure by enabling companies to integrate content with processes to add value and transform their business; streamlining and optimizing complex processes to improve the flow of work throughout the global enterprise; and by delivering an integrated, open platform that can be globally deployed and that provides interoperability with the widest selection of IT systems, thereby reducing costs and improving efficiency. http://ibm.com

JustSystems (Gold Sponsor)

JustSystems is a leading global software provider with three decades of successful innovation in office productivity, information management, and consumer and enterprise software. With over 2,500 customers worldwide, the company is continuing a global expansion strategy based on its XMetaL product line. JustSystems is one of the 2010 KMWorld 100 Companies that Matter in Knowledge Management and a member of the 2009 Software Magazine Software 500 ranking. Major strategic partnerships include IBM, Siemens and EMC. For more information, please visit http://justsystems.com.

MarkLogic

MarkLogic is revolutionizing the way organizations leverage information. The company's flagship product is a purpose-built database for unstructured information. Customers in industries including media, government and financial services use MarkLogic to develop and deploy information applications at a fraction of the time and cost as compared to conventional technologies such as relational databases and search engines. MarkLogic is headquartered in Silicon Valley with field offices in Austin, Boston, Frankfurt, London, New York, and Washington DC. The company is privately held with investors Sequoia Capital and Tenaya Capital. For more information, to download a trial version, or to read the award-winning Kellblog, written by MarkLogic CEO Dave Kellogg, go to http://marklogic.com.

MindTouch

MindTouch is built on the belief that enterprise software must be scalable, agile and extensible. Our product, MindTouch 2010, is the killer app for strategic content, transforms the way organizations author, discover and curate their content to achieve measurable results with customers, partners and colleagues. Our open source project, MindTouch Core, is used by over 18 million people and is supported by one of world's most active communities. Founded in 2005, MindTouch is headquartered in San Diego, California and is privately held. Many of the world's most respected brands rely on MindTouch. Our more than 1,000 customers include NASA, SAIC, Booz Allen, Microsoft, Cisco, Washington Post, Viacom, the New York Times, AXA, Timberland and HCA. http://mindtouch.com

Ovitas

Founded in 2004 and headquartered in Burlington, Massachusetts, Ovitas is an employee owned network of companies currently incorporated in the US, Hungary and Norway. The Ovitas International Network allows each organization to provide ddditional services, solutions, and expertise to our customers, regardless of location. Ovitas provides consultancy, expert design, development, and deployment of content lifecycle solutions. We build solutions to fit our clients' needs, using our own Ovitas Workflow Portal, Ovitas Publishing Bridge, and proven software products. These highly configurable tools for structured and unstructured content management, workflow, search and retrieval, content integration, and packaging let us build cost-effective solutions targeted at your highest priority business issues. Our market focus is on organizations which are staged for the next level of content management solutions. Our primary client base is organizations whose major product is content produced as a main revenue source, or content produced in support and delivery of products. http://ovitas.com

Quark

Revolutionizing Publishing. Again.™

Quark® Inc. is a leading provider of publishing software solutions for professional designers, large organizations, as well as small and mid-sized businesses in more than 160 countries. Two decades ago, our flagship product — QuarkXPress — changed the course of traditional publishing. For more than 25 years, Quark has built on its knowledge and experience in design and publishing to provide software solutions that support collaborative workflows and automated publishing across multiple channels. Today, Quark is revolutionizing publishing again by setting new standards in XML-based publishing across print, the Web, and digital media. Denver-based Quark Inc. is privately held. http://quark.com

SDL

SDL is the leader in Global Information Management solutions thatprovide increased business agility to enterprises by accelerating the delivery of high-quality multilingual content to global markets. The company's integrated Web Content Management, eCommerce, Structured Content and Language Technologies, combined with its Language Services, drive down the cost of content creation, management, translation and publishing. SDL solutions increase conversion ratios and customer satisfaction through targeted information that reaches multiple audiences around the world through different channels. SDL's Structured Content Technologies division is the worldwide leader in Component Content Management (CCM) and Dynamic Publishing software. Leveraging XML standards such as DITA and S1000D, the division's suite of products empower global companies to efficiently create, share, manage and publish technical information that is up-to-date and tailored to the interests of their global customers. Global industry leaders who rely on SDL include ABN-Amro, Bosch, Canon, CNH, FICO, Hewlett-Packard, KLM, Microsoft, NetApp, Philips, SAP, Sony and Virgin Atlantic. http://sdl.com