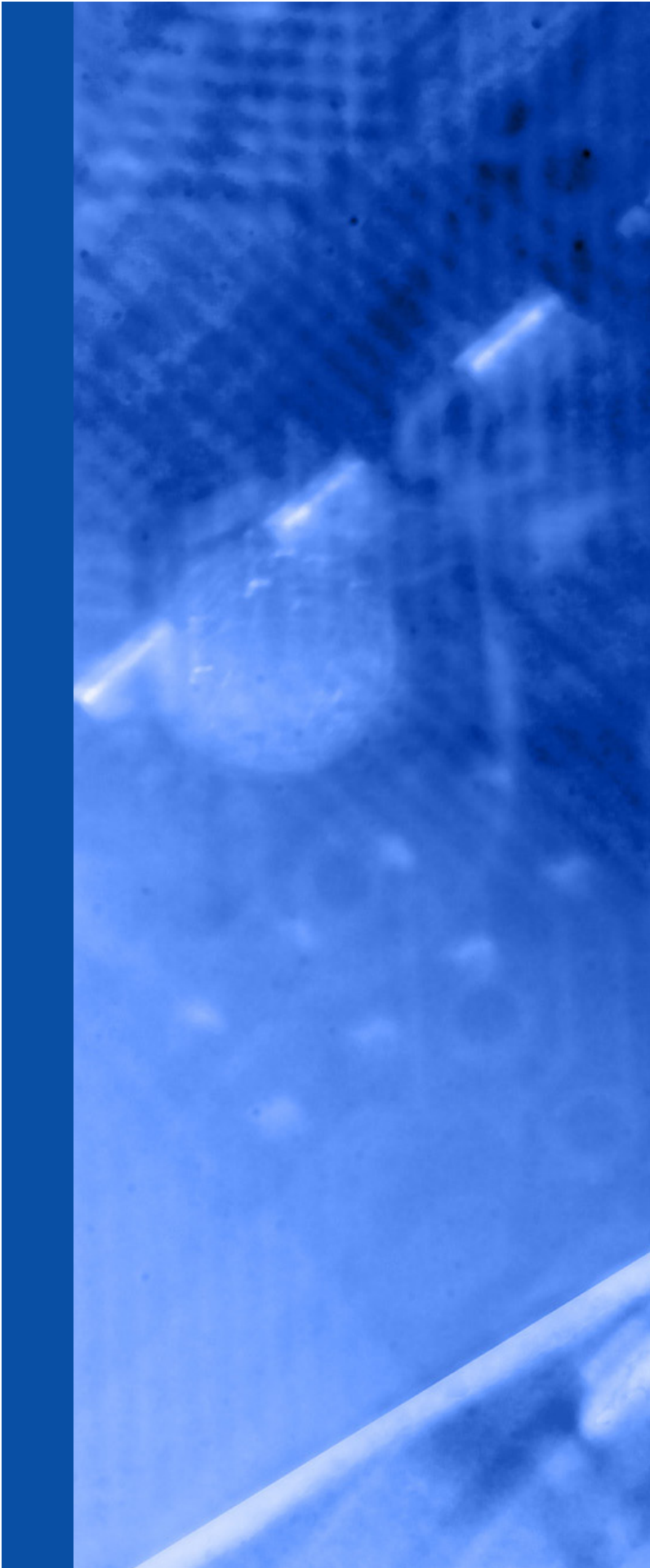




***A Case Study from Smart Content
in the Enterprise***

Single Source Publishing at NetApp

**Adopting an
Infrastructure for
Content Reuse**

August 2010

By Dale Waldt and Geoffrey Bock,
Senior Analysts

GILBANE GROUP
A DIVISION OF OUTSELL, INC.



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>

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.