

Moving to structured documentation: migration stakes

Sara Chaoui, Margaux Chatelain, Julie Dezalay and Justine Gaillard

Abstract

Technical documentation can be written and delivered in two ways: it can be structured and will follow an XML information model; or it can be unstructured and cannot be validated against a model. Nowadays, companies tend to migrate their documentation from unstructured to structured content.

This article focuses on the stakes of moving from unstructured to structured content and explains the key aspects of this process. First, we will identify the reasons of the migration by studying different approaches in several companies. Secondly, the results and/or expected benefits of the migration will be discussed. Interviews of technical writers, documentation managers, and information architects will help us define what the advantages of structured documentation are. We will focus on the added value for the technical writer and for the documentation reader. Moreover, we will try to understand what the company gains from the migration. To conclude, we will draw five guidelines for a successful migration. Moving from unstructured to structured content is a challenging process and the best practices we propose are based on experience and know-how.

Keywords

Structured Documentation, Unstructured Documentation, XML Migration, XML, Information Model, Information Architecture, Writing Guidelines, CMS-Content, Management System, XML Migration Best practices, Added-value for Author, Added-value for User, Standardization of Documentation

Introduction

Structured documentation is a way to separate form from content while providing a model for the technical writer to follow and for the document to be validated against. According to Lori Fisher (2003), XML technology is the best way to reinforce this separation and thus enable the technical writers to focus on the content they produce. However, Fisher sustains that structured information architecture must also be provided. DITA and S1000D are both structured information architectures working with the XML technology. As Mike Cook states in his article (2012), S1000D is commonly used in companies documenting maintenance and operations in the aircraft or military domains. DITA is rather used to provide online documentation about “all [...] kinds of information”. The adaptability and usability of the DITA architecture model seem to prevail above all its characteristics. DITA has an information-type based structure – concept, task, reference – and this architecture is used worldwide. In his study, Keith Schengili-Roberts (2013) informs that 70 % of the companies using DITA are based in North America and 23 % in Europe. Half of them are specialized in computer software, information technologies and telecommunications. The author highlights the “staggering diversity [...] of the subjects” that it is possible to document with the DITA model. While establishing a list of the companies using the model, the author figured out that DITA is definitely a “trend”. According

to him, DITA is becoming “the way to do technical communication”.

Based on our observations and daily experience, this article examines the reasons why a company may want to migrate to structured documentation – particularly DITA – and what would be the added-value of such a process.

What justifies the migration process?

Migrating from unstructured to structured content can be motivated by strategic needs. For some companies, it is prompted by the need to use a technical documentation standard such as DITA, or to comply with the documentation practices of the company. Rizwan Virk (2010) explains how structured content provides significant advantages, such as “the simplification of the creation, storage, indexing, and publishing of documents”. This represents a benefit for companies, as it can reduce costs by improving efficiency in their documentation workflows.

Using XML-based content allows technical writers to meet specific or punctual needs, and offers a great potential for single sourcing and content reuse. For example, DITA provides an appropriate and flexible structure for the reuse of topics and lower-level elements, as well as a suitable reuse mechanism for links and references that are used in different maps or multiple parent topics. Single sourcing can be implemented through different output formats, or

through content reuse with filtered content, which is made easier thanks to the modular structure of the documentation. Structured content can help support frequent product releases by making document generation simpler and faster. It can also allow writers to implement scripts that extract content from the existing information system to automatically generate documentation.

Another reason to migrate to structured content is to improve content management. This can be done by implementing a Content Management System, which gives interesting opportunities to manage the lifecycle of the content. It provides a structure to implement reuse of content, use of metadata and taxonomies, and authoring workflows (Strategic Content, 2010). This also helps when technical writers in a team need to contribute content at the same time. Moreover, it also allows the implementation of content versioning, which is important to keep track of different files for different versions of a product. A Content Management System makes the process of versioning easier, more secure, and helps avoid loss of files.

Structured authoring is a way to improve the content in itself, through various methods. It provides architecture for creating content that is topic-oriented. The three main topic types in DITA are reference topics, concept topics, and task topics. These units of information help to dissociate the kind of information that is given. Writers can more easily review the structure of their documentation, and

standardize their content through guidelines and a specific tagging system, which consequently helps clarify the entirety of the documentation.

In an XML-based documentation system, the content is separated from the layout. Usually, a template is created with XSL, which is the type of style sheet used for XML documents, and this template is used for the formatting of topics. This way, writers can save up to 50% of their time by no longer having to spend time on formatting their documents (Rizwan Virk (2010)). It allows them to focus on their main competency, which is creating and structuring content. This is beneficial for the consistency of the documentation, from the point of view of both style and content. Separate style sheets ensure that the layout remains the same throughout the entire documentation set. This helps follow the guidelines of the company more thoroughly and avoid mistakes. More attention is paid to content, so that during a migration or the conception of new topics, writers can concentrate on their writing matching topic types as closely as possible. The compartmentalization of information types makes the documentation clearer and easier to tag and index.

What is the added value for the end user?

The use of structured documentation is also beneficial to users. Customer documentation is part of the assets of a company, and needs to be carefully

designed in order to ensure the best understanding and efficiency for end users. In the past, the design of a linear documentation that did not follow an information model led users to browse and review an important amount of data to find what they were looking for. The way the content was divided and organized was more difficult to control and manage. With structured documentation, content is modular and must be validated against a model. Modular information ensures that the type of data contained in a topic is clear to the user from the start. If users are looking for a procedure on a specific subject, they know they must find a task topic.

A structured model such as DITA provides users with a consistent way of displaying information and making content available. Technical writers must follow standardized guidelines and create their content according to the three main topic types: task, reference, and concept. There is a clear distinction between these three topic types. This helps ensure that users find the content they are looking for more easily, more rapidly, and it effectively improves documentation consistency.

The usability of the documentation is improved thanks to a XML-based content. The use of topic types, as well as tags, ensures a better management of the content. Elements and attributes are unambiguous and the level of granularity is better defined. Furthermore, the consistency of the content is better maintained, from a semantic and

structural point of view. The modular structure makes the documentation more flexible and easier to modify or update, which gives users a clearer vision of the product. It also supports the richness of the documentation, which “expresses the system’s capability to serve different types of readers” (Lyytikäinen and Tiitinen, 2001). Users are directed in their search of content according to their user profiles. This way, they are not lost among too much information that is irrelevant to them, and they can narrow down their search. Moreover, using a model that is unique to the whole documentation helps users to focus on the content, and not the formatting.

What is the added value for the technical writer?

For a technical writer, the added value of structured documentation is ensured. On the one hand, the profits of the migration are assessed a few months or sometimes a few years after the process. On the other hand, the advantages are quickly acknowledged. For example, using a template represents a huge novelty. A style guide with authoring rules may also be provided. Technical writers welcome this improvement and nowadays they work in a better collaborative environment with a unified writing style. Moreover, the validation of topics against a model rules out formatting issues. Thanks to DITA, technical writers gain new skills. Now they master information mapping, simplified English and XML. Reuse is another advantage of DITA. This capability provides other

benefits, such as authoring simplification and accuracy of information. Finally, technical writers admit that conditional text improves their productivity.

All these advantages and the added values that follow a migration enable the technical writers to be disciplined and to dedicate themselves to the content. To sum up the profits of structured documentation, a technical writer from Dassault Systèmes uses the following words: simplification, consistency, and productivity.

What is the added value for the company?

To convince a company to finance the implementation of a migration process, documentation managers need to insist on the business aspect and the ROI – Return On Investment – that can be expected at the end. The most common argument in favor of XML implementation is costs reduction. Indeed, moving to structured authoring can save a lot of money to companies that localize their content. The typical ROIs a company expects at the end of the migration are the following: better reuse opportunities for authors, which facilitate the documentation maintenance, and potential integration with translation management systems. By implementing single sourcing, authors can reuse content from specifications provided by subject matter experts. Furthermore, XML-based authoring increases the documentation quality through standardized writing styles, which directly impacts the reception of the documentation of the

company by the end user. The documentation is part of the global products delivered by a company, and it represents the showcase for user. Another expected ROI is an increase in efficiency for authors through optimized workflows. Of course, before making money, the company first needs to finance potential training for its technical authors.

Finally, the ROI predominantly depends on how skilled technical authors are, and how clean the unstructured documentation is. Trying to migrate inconsistent unstructured content into structured content is possible, but not without facing issues. In order to have the best possible ROI, all these parameters need to be taken into account.

Conclusion: five guidelines for a successful migration

Finally, migrating from unstructured to structured content allows technical writers and companies to stay competitive: the firsts gain the latest technical skills and the seconds benefit from a higher-skilled workforce that helps them saving money.

Since the migration process can sometimes be harder than expected, here are five guidelines for a successful migration.

1. **Communicate.** The first step consists in explaining the reasons why the company is getting involved in the migration process. The company must also inform technical writers what

- challenges they might face, and what tools they will use. Technical writers need to know all the process, methodologies, and useful tools to be prepared for the migration.
2. **Prepare your migration scope carefully.** If you try to convert in a hurry with no help and no prior experience, you are likely to fail. The migration process takes time, resources and money. Technical writers need to be trained.
 3. **Define your needs.** You must adopt a specific schema model (DTD) which defines the exact structure and rules that the XML content must comply with. Making this decision requires profound knowledge of XML. Indeed, do you want to use DITA, DocBook, or a customized DTD?
 4. **Consider content management.** To develop structured content, you need to think “Topic”. To manage your topics, you will need a Content Management System.
 5. **Be patient.** Migrating from unstructured to structured content takes time. Changing processes and methodologies also takes time. Of course, Rome was not built in a day, neither has a team ever migrated its unstructured documentation to structured content in a day.

References

Fisher, L., 2003, Moving from Single Sourcing to Reuse with XML DITA, *Best Practices*, volume 5, number 3, p. 63

Schengili-Roberts, K., 2013, Who is Using DITA and Where it is Being Used?, *Best Practices*, volume 15, issue 2, p. 29

Cook, M., 2012, S1000D or DITA - Which Should You Use?, *SDL Global Customer Experience Management*, <http://www.sdl.com/community/blog/details/9562/s1000d-or-dita-which-should-you-use>

Virk, R., 2010, Why Convert Documents into XML?, *CambridgeDocs*, www.cambridgedocs.com/resources/whitepapers/id35.htm

2010, Content Strategy In a CMS Project, *Strategic Content*, <http://strategiccontent.com/role-of-a-content-strategist-in-a-cms-project>

Lyytikäinen, V., Tiitinen, P., 2001, Usability of structured documentation - Methods for evaluation, in SGML project, University of Jyväskylä, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.22.3760&rep=rep1&type=pdf>

Abel, S., 2013, XML for Technical Writers – XML authoring coming to a desktop near you, *Writing Assistance, Inc*, <http://www.writingassist.com/resources/articles/xml-authoring-coming-to-a-desktop-near-you/>

Priestley, M., 2001, XML-based Technical Documentation Architecture, *Technology Review*, Volume 48, Number 3, <https://www.oasis-open.org/committees/download.php/6637/DITA-technologyreview.pdf>