Why you need structured content

BLUESTREAM
CONTENT SOLUTIONS

info@bluestream.com
www.bluestream.com

WHITE PAPER

What is structured content and should you be using it?

Types of Documents/Content

There are two different types of documents/content: structured and unstructured, but what does this mean?

Unstructured Documents/Content

Unstructured documents/content refers to content created using standard office tools such as Microsoft Word or Google Docs. Google Docs and Microsoft Word users might follow corporate style guidelines and apply 'Paragraph Styles', but most don't. Even when 'Paragraph Styles' have been used, these documents are still classed as unstructured because the styling only relates to appearance; the content cannot be easily shared. The content in unstructured documents is of very little use to other systems. Also, publications must be updated manually, often with lots of cutting and pasting.

Structured Documents/Content

Structured documents/content refers to content created with re-use in mind. Structured content is created as XML, using authoring applications such as Oxygen Author or structured Adobe FrameMaker.

The benefits of structured content/XML

Structured content can be harder to create but brings considerable benefits, including:

- Omni-channel publishing — the same content published instantly to multiple formats; HTML, PDF, XHTML, HTML5, WebHelp etc.
- Future proof — XML is not dependent upon any single application; humans can also read and understand it.
- Easy to perform global updates.
- Lower cost of localization.
- Easier to find content.
- Easier to share information.

©2024 Bluestream Content Solutions

BLUESTREAM
CONTENT SOLUTIONS

Why you need structured content

info@bluestream.com
www.bluestream.com

Omni-channel publishing/COPE

With XML, because the content is independent of the formatting COPE (Create Once Publish Everywhere) becomes a reality — the same content published to multiple formats; HTML, PDF, XHTML, HTML5, WebHelp, etc.

Future proof

XML has long been the choice for documenting products with long lifespans because it is not dependent upon any single authoring application.

A ship or train, for example, can be expected to be in use for twenty-plus years; can you open and read a document created in Word Perfect twenty years ago?

XML documents can even be printed as flat text files and read and understood by humans.

Easy to perform global updates.

With a modular XML system such as DITA, a single topic, for example, a warning or safety notice, might be referenced by multiple publications. Once the source topic is updated, all publications using that topic can be updated simultaneously — there is no risk of updates being missed because of a manual process. This helps ensure compliance when guidelines are changed.

Lower cost of localization.

Because DITA topics are re-used, unchanged content can be reused across translations. This can reduce localization work and costs; savings depend on the content and translation workflow.

Easier to find content.

Structured content is easy to index and DITA makes use of metadata, this can also be used when searching for content.

Easier to share information

Because XML is not 'tied' to a single application, it is much easier to share content across systems.

When should you consider using structured content?

Structured content isn't always the best approach, but there are certain areas/types of publications where it can offer compelling advantages; these include:

- Regulatory requirements — it is much easier to control documentation in a structured content workflow. This includes the addition of metadata and ensuring compliance with guidelines.
- Large documents — MS Word was not designed for the production of large complex documents; the numbering can be particularly problematic. Structured content is much better at supporting long and complex documents.
- Need for localized versions — you need a means/method to manage the process if you need localized content. But, again, this is something you get with structured content, and with DITA, there is an excellent opportunity to also cost-save.
- Content re-use — re-using content via 'cut and paste' is inherently dangerous, especially within a regulatory environment. In addition, documents built up in this way are hard to maintain when changes occur. Structured content supports controlled re-use.
- Long product lifespan — where products have a long lifespan such as a train or ship, you need to ensure the documentation will be accessible throughout that period. Unfortunately, there are many examples of content no longer accessi-ble because formats have been discarded.
- Need to publish to multiple locations — if you need to publish to multiple channels, the most straightforward approach is to start with structured content.

©2024 Bluestream Content Solutions

BLUESTREAM
CONTENT SOLUTIONS

Why you need structured content

info@bluestream.com
www.bluestream.com

Introduction to XML

XML isn't a specific language but a set of rules governing the syntax of invented vocabularies. The invention of XML came out of a need to describe the content.

Word processors and desktop publishers focus on the formatting of content. When you create new content in these tools, you do so as a part of the layout and formatting process. With XML, you describe the content you are entering, such as a paragraph, a chapter, a book, an article, a caption, or whatever.

XML provides a standard syntax for creating vocabularies to describe your content but does not specify the actual grammar or appearance. XML Schemas or DTDs (Document Type Definitions) are used to identify the exact labels and grammar of a particular XML vocabulary.

While you can invent your own XML vocabulary, doing so means that you will have to customize editing tools to understand your content.

Many XML users adopt a shared standard/common vocabulary instead of creating a vocabulary from scratch. There are standards to represent almost any type of data, whether recipes, musical scores, articles, chapters, books, or anything else. If a community exists around my particular XML standard, we can share tools and techniques that reduce the effort required to deploy content solutions.

Lastly, XML separates content from appearance; XML tags identify what content is rather than how content should look. As a result, a single XML document can be simultaneously published in multiple formats.

Benefits of a shared standard/common vocabulary

Having a common vocabulary means that users can share information, tools, and code to handle the content. For example, if you use a DITA-based format, several editing tools can be used.

Tools used to process the content can also be shared. For example, DITA includes the code and stylesheets needed to create PDF, HTML, and other output formats, and the community is constantly evolving. New formats may appear, and other DITA-based solutions can use the existing tools to support the new format without modifying their processes.

For DITA, the community provides the DITA Open Toolkit. This toolkit includes a variety of transformations that can take DITA content and render it in HTML, PDF, and other formats. It also provides an extensible architecture. If you have a customized version of DITA, you can create a plugin that can enable DITA solutions to handle the specific requirements of your customizations.

Toolkit plugins can be used to configure editing tools, extend the rules of DITA, or modify the included stylesheets used to render content so that they can account for a most specific vocabulary adapted from the base DITA stylesheets.

Any DITA tool can process content even if it is based on proprietary extensions because all of those proprietary extensions are mapped to more generic DITA structures.

So if you use a DITA-based vocabulary that defines a 'chapter,' systems that do not understand 'chapter' can always treat the encoded content as a more generic 'topic.'

So, while XML is a set of rules for creating a particular language to encode your content, DITA is a specific language that was designed to be able to be extended to more specific uses that still share a common grammar.

For DITA, the community provides the DITA Open Toolkit.

This toolkit includes a variety of transformations that can take DITA content and render it in HTML, PDF, and other formats. It also provides an extensible architecture.

If you have a customized version of DITA, you can create a plugin that can enable DITA solutions to handle the specific requirements of your customizations.

Toolkit plugins can be used to configure editing tools, extend the rules of DITA, or modify the included stylesheets used to render content so that they can account for a most specific vocabulary adapted from the base DITA stylesheets.

©2024 Bluestream Content Solutions

info@bluestream.com
www.bluestream.com

More about DITA

DITA stands for Darwin Information Typing Architecture. It is an XML-based open standard initially developed at IBM and maintained by OASIS. DITA is used to create and reuse topic-oriented technical content.

To add an explanation to the name, Darwin because DITA uses the principles of inheritance and specialization pioneered by the naturalist Charles Darwin. Information/ Typing and Architecture are self-explanatory.

DITA is based on the concept of topics. A topic is a unit of information that can be read in isolation or inserted into a larger document. To join together topics, DITA uses the concept of a map file. A map file is simply an XML file that acts as a table of contents linking a series of topic files, in DITA it is called a DITAmap.

The term 'topic' is generic. DITA allows, however, the generic topic to be adapted to represent more specific structures. The basic DITA specification includes Concept, Task and, Reference. These content units are more specific versions of the generic topic. They can be handled with special rules if you want. But if you don't have special rules, they can also be treated more generically as topics.

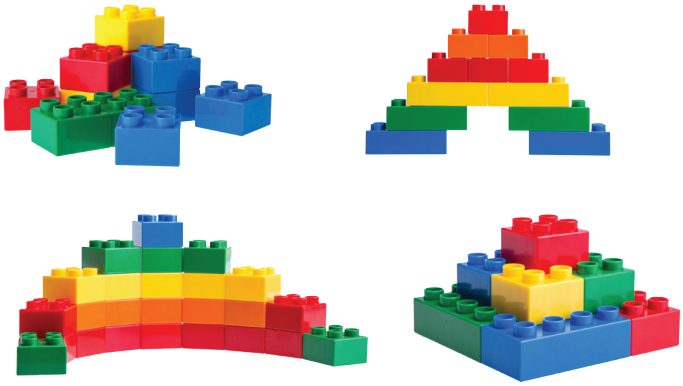
Re-use

DITA differs from other standards in that it uses a topic-based approach to authoring; each topic should be self-contained in that it makes sense on its own. These topics fall into three established categories:

- Concept – provides an overview of what something does.
- Task – provides information on how to do something.
- Reference – provides information on how to check something.

Once created, the topics are assembled for a particular publication using a "DITAmap," that defines their order. The modular authoring approach and self-contained nature of topics enable them to be easily re-used across multiple publications.

A good way of thinking about DITA is to compare 'topics' to Lego or other building block toys.



If each Lego block represents a topic, they can be assembled to make differnet structures.

This is particularly useful when companies produce numerous products, which share components as aspects of one manual are easily incorporated into others.

Reusing topics can reduce authoring and translation work across related manuals. Reusing a given percentage of topics does not establish an equal percentage reduction in total localization costs; savings depend on the content and workflow.

Output

DITA content can be published with the open-source DITA Open Toolkit. Core outputs include PDF, HTML5, XHTML, HTML Help, Eclipse Help and Markdown. Additional formats are available through publishing plug-ins.

Structured content requires managing

If you are to reap the benefits from structured content, it must be properly managed. Using an open-source software code repository system is one option, but it is easier to use a CCMS (Component Content Management System). There are reasons both for and against each approach.

©2024 Bluestream Content Solutions

Why you need structured content

info@bluestream.com
www.bluestream.com

CCMS Pros

- Predictability — features are documented and can generally be demonstrated; pricing is consistent although sometimes high.
- Robust — packed with features and built to a commercial standard.
- Cost to implement — implementation is usually more straightforward since the system is already built and there are no customizations.

CCMS Cons

- Dated technology — some CMSS are based on older technology. The vendors have invested a great deal of time and money over many years as the product has evolved. This makes it expensive to replace the underlying technology due to the related work required.
- 'Handcuffed' to one provider — supported only by the company that developed the CCMS or by limited implementation partners — if you're unhappy with service, support, flexibility, there are no other options.
- Costs – ongoing costs associated with proprietary systems are typically higher than those based upon open standards.
- Lack of customization and flexibility — 'what you see is what you get' — the system is not created for your unique needs, but generalized to meet the needs of all their clients. If you want to do something the system doesn't do? It can be challenging to get new features developed.

Note

Bluestream XDocs DITA CCMS is very different, it has been developed with customization and flexibility in mind. As a result it uses a modular design giving customers the opportuinity to select only the compoents they require.

Futhermore, extensive APIs can connect every part of the system to 3rd party applications. Lastly, Bluestream XDocs DITA CCMS has also been developed in a way that let's Customers configure the system in the way that exactly meets their own requirements.

Open Source System Pros

- Customizations — although based around established core components, an open-source solutions can be built and customized to your specific needs, both in the initial setup and in the future; as those needs change, so can the system
- Flexibility — often built to be easily integrated with other technologies and systems
- Widely Supported — open source technology is supported by a community rather than a single company; as a result, solutions can generally be found. This support is reduced when the system has been heavily customised.
- In–house development — systems based around open source technologies can be further enhanced by internal development teams.

Open Source System Cons

- Upfront cost — since they are highly customized, there is more upfront effort to get it off the ground. Often, the costs to modify an open-source solution to properly support all the functionality required can out-weigh the cost of an 'off-the-shelf' CCMS.
- Support — open-source systems developed by in-house teams rely heavily on those people for maintenance and support. As a result, they are rarely correctly documented, and should the in-house resource move on to a new company, who will fix them when problems arise?

Other areas that can benefit from structured content

Many companies are now accumulating vast amounts of data in a format that can be easily re-used and automatically output in a multitude of formats, but where should they look to take further advantage of DITA?

Support Teams

Responding quickly and effectively to customer issues is more critical now than ever, but Customer expectations are also evolving.

©2024 Bluestream Content Solutions

Why you need structured content

info@bluestream.com
www.bluestream.com

Today everyone expects information to be more accessible; if you have a problem you search Google (other search engines are available), and irrespective of the device you are using (PC, Smartphone, Tablet etc.), fractions of a second later, you receive pages of results – these pages are full of links to multiple sites offering answers in various ways; blogs, wikis, videos, message boards, and forums.

This is how users now expect to receive data; it is no longer sufficient to just deliver a PDF of a manual and expect someone to trawl through it to find the relevant section.

But how do you migrate both technical and non-technical information from static documents into dynamic Topic-based content that can be delivered instantly to any platform, providing engineers, prospects, and customers tailor-made answers?

One obvious solution is to start leveraging your DITA content — this is already broken down into Topics, and with the right tools, these can be selected by either the customer or support team member via wizards to build dynamic documents published in the desired way. Video/links can be embedded in electronic documents to allow the recipients to access additional instructions if required.

Training

So much training is delivered around PowerPoint, yet Trainers are creating their own slides, often manually copying content from technical manuals – WHY? If the content is in DITA, it is possible to generate slides automatically; DITA can also create training reference material.

Marketing

DITA is not a solution for everything, but marketing departments in hi-tech manufacturing companies can exploit DITA content to ensure technical specifications are kept up-to-date on brochures and utilize the flexibility of XML to multi-channel publish material.

Service Engineers

Engineers in the field can spend a long time looking through long PDFs, or even paper manuals.

Publishing machine/installation-specific documentation accessible on any device/platform is the obvious solution. This documentation, especially IPCs (Illustrated Parts Catalogs), can also be directly linked to parts-fulfillment systems.

About Bluestream XDocs DITA CCMS

XDocs DITA CCMS is the leading single vendor solution for organizations looking to improve and enhance their information workflow. With a unique rapid-to-deploy methodology coupled with unrivaled functionality, **XDocs DITA CCMS** is inspiring teams of authors to create powerful user-focused content that is flexibly delivered to any device or platform.

©2024 Bluestream Content Solutions