What is Structured Content?

Making the Case for XML

Val Swisher
Founder and CEO
Content Rules, Inc.
vals@contentrules.com

Regina Lynn Preciado
Senior Director of Content Strategy Solutions
Content Rules, Inc.
reginap@contentrules.com

content rules ®
the global content experts®

Introduction

For decades, companies have been quite successful using
tools such as Microsoft Word, unstructured FrameMaker, InDesign,
and Google Docs to create all sorts of content. There are large corporations that rely on these
tools to write everything from marketing briefs to API guides, internal procedures to training courses.

Unstructured authoring tools are simple to use. Intuitive, in fact. They get the job done and everyone
knows how to use them. Writers, reviewers, editors, and managers can all create and comment on
content when it is in a standard output from an unstructured authoring tool.

Why, then, are companies of all sizes making the move to structured content? They are moving to
structure because it:

- Saves time
- Saves money
- Provides the mechanism for content reuse and single sourcing
- Enables multichannel publishing from the same set of source files

In this ebook we demystify structured content and explain how you can get started extracting the
value from a structured content ecosystem.

Table of Contents

Workflow in an Unstructured Environment

In most writing groups, when someone creates content, it is almost always based on existing content. Every now and then new content is written from scratch. But most of the time, we create a document based on one or more documents that already exist.

Here's a simplified view of a typical workflow for creating documentation using Word:

1. Writer locates existing content that is as close as possible to what they need
2. Writer copies the very similar content from an existing Word document into a new Word document
3. Writer makes additions, deletions, and changes for new document
4. Reviewers make suggestions and corrections
5. Writer inputs changes (note: they do not normally go back to the source of the copy and paste to fix it, even if the change should be made to that source)
6. Final formats are applied to the document
7. Document is published

In the case of an update, even if only one section of the document needs review, the entire document must be checked out and opened. Writers and reviewers often get distracted by content in other sections. After all, who can resist making tweaks to content that already exists?

If the company needs a version of the document for a smartphone or a tablet, all the content must be reformatted for the device. The result is multiple variations of the same information, all needing to be updated when the time comes.

All this manual inefficiency during the content lifecycle can result in lower quality and is guaranteed to be more expensive. And that's without adding the complications of translation.

Content Drift Increases Risk

Manually copying and pasting content inevitably introduces content *drift*.

Content drift is what happens when writers copy, paste, and then tweak content. Most writers believe that revision, even if unnecessary, makes the content more fit for purpose. Unfortunately, this "freedom to tweak" results in redundant content that is almost - but not exactly - the same as its source. The content in the downstream document no longer matches the content in the upstream document.

The content drifts further and further apart with each iteration. Content drift creates inconsistencies in terminology, usage, and meaning that confuse readers and reviewers.

Content Reuse Solves the Problem

The solution to content drift is to content reuse. **Content reuse** is the practice of creating one piece of content and using that piece of content everywhere you need it. When you reuse content, each piece is created, approved, updated, translated, managed, stored, and retired one and only one time.

Copying and pasting content creates content redundancy. Multiple, redundant versions of the same information. The opposite of content redundancy is content reuse.

To reuse content, you must use structured content.
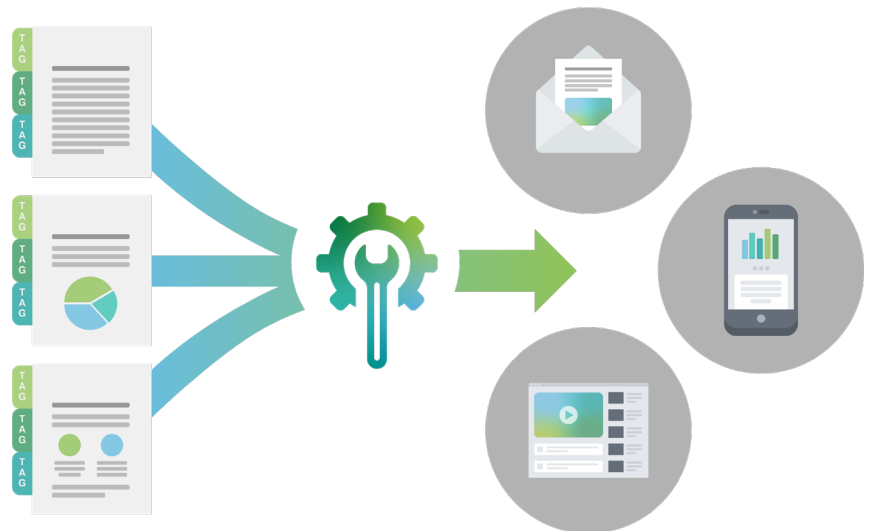
What Is Structured Content?

Structured content is content that is **modular**, **consistent**, and **reusable**.

Structured content is:

- Created as a set of building blocks of information (components)
- Created and stored with minimal formatting
- Tagged with metadata for findability
- Organized and stored in a centralized repository
- Publishable to a variety of formats

To produce final outputs such as documents or webpages, components are assembled into the required order and hierarchy. This assembly can be automated by a component content management system (CCMS). If needed, parts of the document can be pre-populated with reused content, boilerplate text, pre-formatted tables, and data retrieved from other sources. Authors can use their time to create valuable content rather than copying and pasting, formatting a Word document, or manually moving components into place.

In a structured environment, you author content in components. The lifecycle of creation, review, revision, and approval occurs outside of the process of final assembly. Reviewers and SMEs can provide input on just the components that are relevant to them, they can view the entire output, or both. Access is usually set up in the CCMS.

What is a Structured Content Ecosystem?

A structured content ecosystem contains all the elements that are needed to create, store, manage, and publish structured content.

In the diagram above, the left side of the CCMS are the inputs. To the right of the publishing engine are some possible outputs.

1.	**Components**	Components are small chunks of content. They are created according to their defined structures. Components are combined at the time of publishing to create an information product.
2.	**Metadata**	Components are tagged with metadata to help categorize and find content when it is needed.
3.	**Component Content Management System**	The CCMS is a database that contains the components. The components are organized using a taxonomy.
4.	**Localization**	Localized and translated components are often stored and managed with the source content.
5.	**Content Optimization**	Optimization includes terminology management, content optimization, and quality scores that are stored with the content.
6.	**Publishing Engine**	The publishing engine contains style sheets and transforms. They allow content to flow into the design and layout created for each information product.
7.	**Information Products**	Information products are the delivered outputs.

What Is a Structured Content Strategy?

A structured content strategy is a business plan for content. A structured content strategy defines the structures necessary to ensure that content is nimble, usable, reusable, findable, and relevant. It includes your information architecture, your governance model, and your business goals.

A structured content ecosystem starts with a structured content strategy. Without a strategy, the new processes and systems you implement will not be as successful. In fact, most of the failed attempts at adopting structured content in the past failed due to a lack of strategy, not a lack of technology.

A complete structured content strategy includes the following elements:

Content Models	A set of rules that define what content to include, under what conditions, in what order.
Structured Authoring Guidelines	Guidelines and examples to help authors create reusable, modular content according to standards.
Reuse Strategy	Rules and guidelines that define which content to reuse where and which mechanism to use to accomplish the reuse.
Taxonomy	An organizational structure that ensures authors and systems can find components to retrieve or reuse.
Metadata	A controlled set of tags that identifies content so that machines and people can process it appropriately.
Workflow	Status and transition indicators that streamline processes and provide an audit trail for all content.
Governance	Rules and guidelines for how people interact with the content now and in the future.

What Are Content Standards?

Content standards are the set of rules and guidelines that govern content.

Content standards enable you to make your content FAIR: findable, accessible, interoperable, and reusable. FAIR is a set of guidance principles developed by data standards bodies.

Like data, content must be FAIR so that we can view, exchange, and manage it. With FAIR data and content, we can build automation, connect applications, and ensure a successful flow of data from a single source of truth to wherever that information is needed.

The Five Dimensions of Content Standardization™ Framework

Content Rules developed the Five Dimensions of Content Standardization framework using knowledge gained over decades of helping companies transform to component-based structured content. The Five Dimensions of Content Standardization framework ensures that content is FAIR. It ensures that content can be reused everywhere it is needed, and that data and content can be integrated. When content is standardized across all five dimensions, processes from authoring to archiving can be automated, streamlined, and standardized. Content accuracy and quality increase. Risks are minimized.

The Five Dimensions of Content Standardization are:

- Output type
- Component
- Paragraph
- Sentence
- Word

To learn more about the framework and the importance of content standards, download our white paper [The Five Dimensions of Content Standardization™ Making Your Reuse and Automation Strategy a Success.](#)

Making the Case for XML

Component-based structured content requires a technology backbone that is simultaneously flexible enough to support global teams creating global content and rigorous enough to enforce standards for content and data across a wide array of outputs.

Structured content is more than just a document outline based on a Microsoft Word template. To enable content reuse and automation, the structured content ecosystem must be based on XML.

XML, or eXtensible Markup Language, is a text-based format for representing structured information. XML was derived from an older standard called SGML (ISO 8879). XML has become the standard for component-based structured content management and information exchange across all industries.

XML Makes Content FAIR

XML helps ensure that content is FAIR – findable, accessible, interoperable, and reusable. Originally created for data, the FAIR principles apply equally well to structured content.

Findable

XML uses semantic tags that provide meaningful identifiers for the content. These tags enable systems to filter, search, and retrieve content quickly and accurately.

Accessible

XML structures content consistently to improve usability for humans and machines. This structure enables business rules that automate many content creation, management, and publishing tasks.

Interoperable

XML is an industry standard that supports the exchange of information between systems.

Reusable

XML enables relationships between pieces of content, allowing you to retrieve and reuse content from a single source of truth. XML also provides tags that can be mapped to publishing transforms and style sheets so that the same piece of content can be formatted and published to any output format.

XML Is Our Secret

The most important thing about XML for authors is that they do not need to understand XML to create content.

Writers and other authors can create documentation, training materials, support and knowledgebase articles, and training courses in an XML-based system without ever seeing the XML itself.

One reason that structured content and content reuse have been so difficult for many companies to implement is the opposition to using XML-based tools.

The failure of XML in the past has little to do with the backend capabilities of XML tools. In fact, XML tools are far more advanced now than any other authoring suite on the market. What we did not have until recent years was an authoring environment that authors could use easily and effectively.

This situation has changed. We now have authoring environments that look and feel a lot like Microsoft Word or Google Docs. Several CCMS vendors have incorporated simple user interfaces directly in their platforms. Standalone authoring tools, such as Fonto/XML, are also available.

Using a built-in simple authoring tool, you can create content in a linear fashion. In other words, you can write a document and the system automatically componentizes the content in the background. Authors and reviewers can view the document the way they are accustomed to, with all components presented in the correct hierarchy. They can navigate the document in familiar ways such as search, table of contents, and navigation pane.

Authors can easily reuse components by inserting the content where it should go. The technical aspect of content reuse is taken care of behind the scenes.

You Need Structured Content

Over the years, companies have looked at several strategies to tackle their content-related challenges. There's talk of new tools and systems. There's talk of AI and automation. There's talk of process improvements and corporate reorganizations.

What people forget to talk about is the most basic starting point of all: the content. If you don't start with the content – if you build your content strategy plans based on the idea that content just appears from somewhere – then your digital transformation initiative is doomed to fail.

New tools and new roles for people are part of the solution. But without standardized, structured content, companies cannot achieve their business goals of saving money and reducing time to market.

The structured content solution involves a combination of process, people, and technology.

- **Process** – We need to change how a company "does" content. We need to implement a new ecosystem for how content is created, managed, and delivered. This ecosystem must be based on best practices for standardized structured content and a content strategy that is unique to each company's needs.

- **People** – We need people to create the vision, communicate the vision, and lead the company through the transformation from current state to future goal. We also need people to do the work of transformation. We may not need new people, but we need people to develop new skills and take on new roles.

- **Technology** – We need systems and integrations in place to provide automation and ensure quality, compliance, and efficiency throughout the content lifecycle.

With structured content, even large numbers of authors distributed across different teams and working from different locations can create content that fits together in a unified way.

The Benefits of Structured Content

Structured content offers tremendous benefit to the business.

It saves money, saves time, reduces risk, and produces higher quality content.

Structured content costs less to create, manage, review, publish, and translate.

Capability	Documents	Structured Content
Deliver as a document	✓	✓
Reuse a single chunk of information		✓
Maintain a single source of truth		✓
Update in one place to update all documents and outputs		✓
Fewer documents to manage		✓
Lower translation cost		✓
Use same content in different output formats		✓

Summary

Structured content is the only way to create content efficiently at scale. It's the only way to ensure data and content are interoperable. It's how you maintain a single source of truth while using content everywhere you need to use it.

Structured content also prepares content for future technologies such as artificial intelligence and natural language generation.

XML has been a challenge for companies in the past. The authoring tools were not mature enough to hide the complexity from authors. Other obstacles included the lack of experience connecting XML to other systems. Today, there are mature tools and experienced people to make these connections work.

In addition, the standards in data and content have matured. New authoring environments are comfortable and powerful. The Content Rules proprietary methodology of structured content strategy and the Five Dimensions Framework help companies make their transition to structured content a success.

Content Rules has decades of experience helping the world's largest and most innovative companies move to structured content. Contact us to see how we can help get you achieve your content goals.

Resources

- **5 Dimensions of Content Standardization**
- **Content Transformation: Breathe New Life into Legacy Content**
- **Discover more about content strategy**
- **Get more effective content with a content strategy**

content rules ®

the global content experts®

About Content Rules

We're content experts who have held the leading edge in our field since we opened for business in 1994.

The world's largest and most innovative companies trust us to develop their enterprise content strategies, transform their content ecosystems, optimize content for a worldwide audience, and develop effective content that gets results.

Content Rules has the knowledge and experience to deliver the highest quality content services every single time.

We are content experts.

We have held the leading edge in our field since 1994.

www.contentrules.com

Our services include proprietary methodologies owned by Content Rules, Inc. Please do not share or distribute information pertaining to these services without express written permission from an authorized representative of Content Rules, Inc. These proprietary methodologies include: (1) Five Dimensions of Content Standardization™ Framework, (2) The Rockley Method™ of Unified Content Strategy, (3) Content Rules Content Transformation and Migration™, (4) Content Rules Anatomy of Change Model™, (5) Content Rules Structured Content Authoring for Pharma™ Methodology.