The Five Dimensions of Content Standardization™

Making Your Automation and Reuse Strategy a Success

Val Swisher
CEO
vals@contentrules.com

Regina Lynn Preciado
Sr. Director of Content Strategy Solutions
reginap@contentrules.com

Contents

Introduction

Across all industries, companies have reached a content breaking point. Traditional methods of document publishing no longer serve the best interests of the corporation, the content creators, or the customer. There is tremendous momentum behind efforts to modernize the way we create, store, manage, publish, and update content.

Everyone knows that copying and pasting (and tweaking) content from document to document is time-consuming and fraught with risk. To mitigate that risk, many industries are embracing the possibilities of structured content. These possibilities include (but are not limited to):

- Content reuse from a single source of truth
- Automated content assembly, formatting, and publishing
- Multi-channel publishing from a single source of truth
- Detailed audit trails from authoring to archiving for every piece of content
- Automated convergence of content and data

Companies across all verticals have made the decision to move to a structured content authoring environment. However, for structured content to deliver on its promises, the content must be standardized.

Content standards help ensure a successful integration of people, processes, and systems.

In this white paper, we describe what content standards are and why they are important. We introduce the Five Dimensions of Content Standardization™. We show how our content standardization framework makes content FAIR (findable, accessible, interoperable, and reusable). Finally, we suggest some next steps that you can take to prepare your content for the future.

Why Component-Based Structured Content Is Important

Component-based structured content is important because it creates opportunities for content reuse and automation. With component-based structured content, content is created in small, reusable "chunks" that are assembled to create deliverables such as knowledgebase articles and user guides.

Component-based structured content helps companies save time and money, produce multi-channel output, and scale.

Save Time and Money

In most companies, employees spend significant time (and money) searching for information. Once they find information, they spend yet more time determining whether what they have found is current.

In our experience, using standardized structured content enables companies to create content as much as five times faster than they do using unstructured content.

This time savings comes from many capabilities:

- Automate the creation, revision, and production of outputs
- Streamline the writing process so that authors focus on creating new and changed content rather than tweaking existing content
- Reduce review and revision cycles through content reuse rather than creating redundant content
- Send only new and changed information to translation, rather than sending an entire document (even if most of the document has not changed)

Multi-Channel Output

In a component-based structured content ecosystem, the content is separate from its format. For example, text exists as only text. It does not have any format associated with it. It is not "on a page" or "on the screen." At the point of publishing, the format is applied to the content.

This means that the same content can be used for multiple outputs. For example, a laptop screen, a tablet screen, and a phone screen can all use the same content. During the publishing process, the text is formatted for each type of device. Formatting can include font, font size, justification, and more.

Scale

Component-based structured content is scalable. Structured content helps companies realize their business goals: saving money and reducing time to market.

Using structured content, companies can:

- Automate content creation, revision, assembly, and delivery
- Create, update, manage, and translate content once and reuse it everywhere they need to say the same thing

Using standardized structured content, companies can deliver content in a multitude of formats. At the same time, companies can significantly reduce the time it takes to produce the output in all formats and languages.

What is 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, finalized, translated, managed, stored, and retired one and only one time.

Content reuse is not something you can start on an ad-hoc basis. You need up-front planning to determine which content to reuse, how to reuse it, and how to manage it within and across organizations. You also need a component-based structured content ecosystem to automate and facilitate content reuse.

One goal of creating reusable content is to make the reading experience seamless. The content consumer cannot tell that any piece of content originated at a different time, from a different author, in a different system, or as part of a different project.

To create a seamless reading experience, standardize all content.

FAIR Principles

A large team led by Mark Wilkinson developed The FAIR Guiding Principles, described in the publication Scientific Data on March 15, 2016. FAIR stands for Findability, Accessibility, Interoperability, and Reusability.

FAIR principles were developed to help guide the management and reuse of data. They are intended to support actions taken by both humans and machines.

Findable

To reuse data, you need to be able to find it. Findability is the principle that governs how to create and apply metadata.

Accessible

Once you can find the data you need, it needs to be accessible. The protocols for authorization need to be implemented in a clear and standardized manner.

Interoperable

The data needs to be able to interact with other data. Interoperability standards are important for storing, processing, and working with data in applications and other workflows.

Reusable

The end goal of the FAIR principles is to support data reusability. Standards and robust metadata are key to reusability.

FAIR Principles and Component-Based Structured Content

The very same FAIR principles developed to govern data can (and should) be applied to content. 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 anywhere that information is needed.

Content standards are key to successful structured content.

Content standards include:

- Preferred and forbidden words
- Style and grammar rules
- Voice and tone
- How sentences combine to create paragraphs
- How paragraphs combine to create sections or components
- How components combine to create the final output

Standards also define how content is written, stored, managed, published, and retired.

The Five Dimensions of Content Standardization

Content Rules developed the Five Dimensions of Content Standardization framework using knowledge gained over decades of helping companies evolve with component-based structured content.

The Five Dimensions of Content Standardization framework ensures that content is FAIR. It ensures that content is reusable, and that data and content can be integrated. When content is standardized in all five dimensions, processes from authoring to archiving can be automated, streamlined, and consistent. Content accuracy and quality increase. Time spent, costs, and risks are minimized.

The Five Dimensions of Content Standardization are:

- Output type
- Component
- Paragraph
- Sentence
- Word

Dimension One: Output Type

An output type is an assembly of content that is delivered as a complete unit. Some common output types include:

- Online help system
- Hardware installation guide
- Knowledgebase article
- Case study
- eBook
- Training course

A standardized output type provides a model to ensure that all requirements for a particular output are met. Output type content standards define:

- The type of content to include
- The order in which to include the content
- Which content is required and which content is optional
- Which content to reuse every time
- Which content to create new every time

The file format of your delivered content does not drive your output type standards. You can deliver an output type in many different file formats, such as a PDF, a web page, a spreadsheet, or a presentation. You may have certain design requirements for different file formats, such as whether to include an index in a PDF. Those requirements are part of formatting and design, not part of content standards. Output type standards define the requirements for content regardless of what format that content ultimately flows into.

The process of standardizing output types includes identifying opportunities to reuse content. You can then configure systems to provide the correct reused content component at the right time, automatically. For reuse that requires human intervention, you can train authors when to reuse which content.

Dimension Two: Component

A component is an independent unit of content that can be combined with other components to create an output.

Component standards include criteria such as:

- The type of component
- The order of the content within the component
- What content is mandatory and what content is optional
- Authoring guidelines for creating consistent content
- Boilerplate text where applicable

Some common components include:

- Introductory section in an overview
- A task in an installation guide
- A table for specifications in a data sheet

Headings and subheadings within a document emulate components. However, these components only become modular and reusable when used in conjunction with a component-based structured content management system. Components may also be more granular within a section to increase the potential for content reuse.

For example, you may have a situation where your content needs to include a summary component type and a detail component type for each item in the summary. Authors use the summary component to provide information that is intended for reuse, such documentation, training, and knowledgebase. Authors use the detail component type to provide more specific information that has fewer instances of reuse. Components that are unique to an output type may not be reused. Other components have high potential for reuse. Components used in multiple outputs should be single sourced. In other words, write the component one time and use it in every output you need to deliver. Developing single-sourced components according to standards makes content creation faster, improves content quality, and reduces the risk of introducing inaccuracies and errors.

All components need to follow your component standards, whether the component is reused twice or twenty times, or not at all. By standardizing the structure of components, you ensure that components written by different authors, different geographies, and different times of creation flow together in a cohesive, comprehensible way. You also ensure that a component viewed on its own contains all relevant information, in the expected structure.

Standards for a Global Impact

Component standards improve translation quality while reducing the cost and time of localizing and translating the content. Using components, you do not send the entire document or output to translation every time something changes. Instead, you send only those components that are new or have changed.

When translated content is reused, there is no associated translation cost as it's an exact match with the translation memory.

Dimension Three: Paragraph

The paragraph is the central building block of your content. Paragraph standards guide authors in how to provide information in accessible units. For example, scientific documentation has traditionally included many long, dense paragraphs that often combine data with narrative. These paragraphs are more difficult to read than short, crisp paragraphs that group information together both logically and visually.

Paragraph standards include criteria such as:

- Recommended maximum length
- Voice and tone guidelines
- Target reading level
- When to use lists or tables instead of paragraphs
- When to use notes, cautions, and warnings

Shorter paragraphs are easier to navigate and comprehend for people who have English as a second language or who are reading the content in translation. Everyone who has ever attempted to read complex materials on a mobile device will thank you for adopting shorter paragraphs and more frequent subheadings.

Paragraph standards help ensure that components fit together seamlessly when they are reused across different outputs. This unity is important to reuse content successfully.

Dimension Four: Sentence

Grammar and style rules govern how words combine into sentences.

Sentence standards are particularly important if you are going to reuse content. When multiple writers work on different components, sentence structures need to match so that you can ensure the readability and accuracy of the content.

Using Variables

A variable is a placeholder, such as <product name> or <model number>. A variable is resolved automatically throughout the content with the unique information (such as Acme Router) when the content is published. Using variables makes your content more reusable.

For example, you may have two models of the Acme Router, Model A and Model B. If the paragraph is the same except for the model number, use a variable as a placeholder within the text. When you publish the content, <model number> is replaced with either Model A or Model B, depending on which router you are publishing for.

Paragraph and sentence standards should include guidance for when and where to use variables. Your paragraph and sentence standards work together to ensure that the content makes sense.

Sentence standards include:

- Grammar
- Style

Grammar

Most of us learned about grammar back in elementary school (sometimes called grammar school). In technical terms, grammar is the study of how words are used in a sentence. Typical topics in the study of grammar include:

Parts of speech (also called word classes), such as:

- Nouns and pronouns
- Verbs
- Adjectives
- Adverbs
- Prepositions

- Conjunctions
- Negation, which includes:
  - Using the word not
  - Contractions such as isn't, can't, and couldn't
  - Other negating words such as never, nothing, and nobody

Sentence structure, which includes:

- Word order
- Dependent clauses
- Imperatives

Style

People often confuse style with grammar. Grammar defines the technically correct ways to use words according to a specific language. Style defines choices that affect the way sentences are understood. Style forms a bridge between words and grammar.

Style rules codify the intentional, specific, detailed decisions you make about how authors write on behalf of your company. Following a consistent set of style rules helps make components interchangeable and their assembly seamless. It also helps content created across different regions and countries sound the same, so that it can be reused easily.

Style assumes that you are choosing from grammatically correct options. Your grammar must be accurate for style rules to make sense.

There are hundreds of style rules, including decisions such as:

- Use of the Oxford (serial) comma
- Use of idioms, jargon, and colloquialisms
- Use of gender-neutral pronouns

Standards ensure that sentences make sense when data and words are combined. For automation to work, everyone needs to follow the rules.

Style Guides

Numerous standard style guides are available for you to use. Typically, a company selects at least one style guide from the various options and tries to enforce it across the enterprise. Style guides can also describe how to create citations for online and print content.

The Associated Press Stylebook

American journalists working with the Associated Press created the Associated Press Stylebook (often called the AP Style Guide). Even though the guide was written for journalists, over the years it has become one of the leading style guides for corporate communications and digital content.

Chicago Manual of Style

The Chicago Manual of Style (CMOS) is an American English style guide that has been around since the early 1900s. CMOS is one of the most widely used writing style guides available.

Microsoft Manual of Style

Microsoft publishes the Microsoft Manual of Style (MMoS). Although originally intended for corporate content creators, it has become a standard style guide for many technology companies.

Custom Style Guides

Most large companies use one of the standard writing style guides as a starting point for creating their own custom guide. In addition to topics that focus on sentences and parts of sentences, custom style guides often include:

- Allowed and disallowed terms
- Acronyms and company-specific words
- Legal boilerplate wording, including caution and warning notices
- Preferred systems of measurement to use (imperial/metric)
- Index usage and style
- Glossary usage and style

A writing style guide is sometimes incorporated into a larger, brand style guide. A brand style guide might also contain information about:

- Logo usage
- Brand mission
- Corporate colors and fonts
- Formatting
- Visuals
- Tone and voice

Dimension Five: Word

A word is the smallest standardized content unit. A collection of standardized words is called terminology. Standardize terminology so every piece of content you create uses the same word to mean the same thing. For example, many software companies standardize on terms such as "administrator" and "user" rather than "sysadmin" and "operator.

Terminology standards improve clarity. Readers do not have to wonder if a shift in word usage indicates a shift in meaning. Other reasons to manage terminology include:

- Enforce accuracy

- Improve readability

- Lower the cost of translation

- Lower risk

- Speed time to market

What Terms Should You Manage?

You cannot (and should not) manage every term in your content. Sometimes, deciding which terms to include and which terms to leave out can be difficult. We recommend that you focus on the following categories.

Terms with Legal / Regulatory Implications

When used incorrectly, terms with legal implications can open your company to a potential lawsuit. Correct use of such terms is crucial. Your legal team likely keeps close track of these terms, particularly if your company operates in a highly regulated industry.

Product and Brand Names

Strict adherence to product and brand names helps establish your product and brand in the marketplace. Inconsistent product and brand names make the content sloppy and confuse customers. Plus, you can accidentally create trademark or copyright infringement by straying from the registered name.

Neologisms and Portmanteaus

Manage all words that your company invents. When you use words that aren't in a dictionary, making sure everyone uses them in the same way becomes even more important. If your content needs to change quickly to keep up with the latest internet language, then you simply must manage those terms.

Prohibited Terms

Every company has a list of words it wants content creators to avoid. Keeping a list of prohibited terms is important. Equally important is selecting at least one preferred term for each term that you disallow. Telling people which terms they cannot use, without giving them an approved replacement, is annoying at best. At worst, you risk another prohibited term being selected in its place.

Terms That Promote Inclusivity—or Add Bias

Everyone has unconscious biases that emerge in our writing. The internet is inherently biased. Managing terminology helps to prevent these biases from making their way into your company's content.

Terms with Unusual Capitalization

Companies tend to get creative with capitalization. They might make proper nouns of everyday words to emphasize their importance. Or they might capitalize random letters as part of a branding attempt.

If you capitalize too many words, readers can get confused or start ignoring the words.

Enforcing Standards in Real Life

Documenting your content standards in a spreadsheet or document (or both) is not enough. Authors are not going to check your standards documentation every time they write. They certainly won't check the documentation after becoming familiar with the standards. Your authors can't and won't memorize your style guide – and they shouldn't have to.

There are software tools that help manage and enforce standards at scale. These content optimization applications plug in to your authoring environment. They provide real-time feedback to help authors use the right terms (and avoid prohibited terms) and follow the right spelling, grammar, style, and voice standards for content. Content Rules works with a number of these tool vendors. Contact us to learn more about their capabilities and if they are right for your organization.

What Not to Manage

When it comes to managing terms, steer clear of one big category:

Common words used in a common way.

There is no need to manage a term that is simply a word (or a word cluster) with nothing special about it.

Don't Wait for a New System

You do not need to wait for new tools before you develop standards.

It takes time to develop, test, and adopt standards. It also takes time for authors to learn to write to the standards.

You can begin standards development and change management right away. Authors can begin to create content according to standards and work "as if" you have a component-based structured content ecosystem. Your content quality will improve, and your teams will be poised to adopt new systems when the time comes.

About the Five Dimensions of Content Standardization™ Framework

The Five Dimensions of Content Standardization™ framework is the culmination of our decades of experience working in every industry including life sciences, financial services, high tech, and manufacturing. Our Five Dimensions framework enables our customers to:

- Automate content generation, assembly, formatting, and publishing
- Reuse content in any output type or file format
- Deploy artificial intelligence to gain actionable insights
- Increase translation quality while significantly reducing cost and turnaround time

The framework combines the foundational principles of component-based structured content with content optimization best practices such as global English and writing for reuse.

The FAIR Guiding Principles[1]:

- **To be Findable**
  - (meta)data are assigned a globally unique and persistent identifier
  - data are described with rich metadata
  - metadata clearly and explicitly include the identity of the data it describes
  - (meta)data are registered or indexed in a searchable resources

- **To be Accessible**
  - (meta)data are retrievable by their identifier using a standard communications protocol
    - the protocol is open, free, and universally implementable
    - the protocol allows for an authentication and authorization procedure, where necessary
  - metadata are accessible, even when the data are no longer available

- **To be Interoperable**
  - (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
  - (meta)data use vocabularies that follow FAIR principles
  - (meta)data include references to other metadata

- **To be Reusable**
  - meta(data) are richly described with a plurality of accurate and relevant attributes
    - (meta)data are released with a clear and accessible data usage license
    - (meta)data are associated with detailed provenance
    - (meta)data meet domain-relevant community standards

[1] https://www.go-fair.org/fair-principles/

Summary

Component-based structured content provides many capabilities:

- Content reuse from a single source of truth
- Automated content assembly, formatting, and publishing
- Multi-channel publishing from a single source of truth
- Detailed audit trail
- Automated convergence of content and data

To be successful using component-based structured content, you need to develop and implement standards. Content standards make content FAIR and help ensure the successful integration of people, processes, and systems.

The Content Rules Five Dimensions of Content Standardization framework addresses rules that govern:

- Output type
- Component
- Paragraph
- Sentence
- Word

When you clearly document and enforce standards across all five dimensions, you are rewarded with content that can be reused seamlessly. You also reduce time-to-market and cost, while increasing content quality in all languages.

Content Rules has helped scores of companies transform from document-based publishing to component-based, structured content management. We developed our Five Dimensions of Content Standardization Framework™ to help companies standardize content in all five dimensions to ensure that content is FAIR: findable, accessible, interoperable, and reusable. Contact us to see how we can help solve your complex content challenges.

Contact us to see how we can help you!