content rules®
the global content experts®

The Five Dimensions of Content Standardization™

Making Your Automation and Reuse Strategy a Success

Val Swisher
CEO
vals@contentrules.com

Regina Lynn Preciado
Senior Content Strategist
reginap@contentrules.com

www.contentrules.com | info@contentrules.com

Contents

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

Introduction

The pharmaceutical industry has reached a tipping point. There is widespread agreement that traditional methods of document publishing no longer serve the best interests of pharma companies, regulators, health care providers, or patients. There is tremendous momentum behind the 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, the industry is 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
- Digital information exchange internally and externally with sites, CROs, regulators, and partners
- Detailed audit trails from authoring to archiving for every piece of content
- Automated convergence of content and data

Pharma companies already understand data standards. Everyone understands that for data to flow from one system to another, that data must comply with standards. The same principles apply to content. 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.

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 pharma industry is intimately familiar with the importance of standards for data. The data standards organization CDISC (CDISC.org) has more than 500 member organizations[1] working to develop standards for different types of data. Some of the resulting data standards have been adopted into requirements by regulators.

People understand that to make data exchange possible, the data must adhere to a defined structure based on standards. To build applications that allow us to visualize data and gain insights from the numbers, that data must be standardized.

The same principles apply to content. Content standards are key to successful structured content.

Content standards include:

- Words that are preferred and words that are not allowed
- Style and grammar rules
- Voice and tone (such as scientific paper or lay summary)
- How sentences combine to create paragraphs (and where the data comes in)
- How paragraphs combine to create sections or content components
- How components combine to create the final output

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

[1] https://www.cdisc.org/membership/our-members

The Role of Standards in Content Reuse

One goal of creating reusable content is to make the reading experience seamless. By seamless we mean that 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, all content must be standardized.

"In our experience, using standardized structured content provides the opportunity for pharma companies to create content as much as five times faster than they do today."

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 how you create the contents of the eCTD in less time, at a reduced cost. It's how you scale operations to provide the vast amount of interrelated content that must accompany every molecule, at each stage of that molecule's lifecycle.

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.

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 clinical reports and labels.

Component-based structured content enables companies to reduce risk and to scale drug development.

Risk

There is enormous cost in how much time people spend searching for information within a pharma company. Once they find information, they spend yet more time determining whether what they have found is current.

This problem becomes much worse when an inspection is required and all content must be frozen in place, wherever and however that content is stored. If the inspector:

- Finds multiple copies of the same content but no definitive source of truth
- Sees that content has changed but has no history of when, why, or who changed it
- Cannot find the required content at all

the cost is immeasurable.

At best, the pharma company loses time that they could have been marketing the drug product. At worst, people do not get the medicine they need to improve their quality of life.

Scale

Component-based structured content is scalable. Structured content helps pharma companies realize their business goals: reducing risk, reducing time to market for new drugs, and developing more products in the same time frame.

Using structured content, pharma companies can:

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

Using standardized structured content, pharma companies can deliver the required documents in the required formats. At the same time, pharma companies can significantly reduce the time it takes to produce those documents.

In our experience, using standardized structured content provides the opportunity for pharma companies to create content as much as five times faster than they do today.

This time savings comes from many capabilities:

- Automate the creation, revision, and production of documents
- Streamline the writing process so that authors focus on creating unique content that requires human intervention (such as interpretations of clinical trial results), rather than waste time writing "wrappers" for data or pasting tables into Word documents
- Reduce review and revision cycles through content reuse rather than creating redundant content
- Prepare content for digital submission as some health authorities and regulatory bodies modernize their systems
- Automate the flow of data from a source of truth into the content that narrates that data

A Future Beyond Documents

We believe that over time the need for documents will diminish. Content components will be delivered directly to regulators. Reviewers will assess content in smaller pieces that relate directly to their area of responsibility.

Data will be exchanged directly without being buried in long documents full of related data and content that is not immediately relevant. Systems will automatically compile relevant information into an assembly at the request of the reviewer, clinician, or patient, rather than the pharma company having to assemble and publish documents first.

That said, documents are likely to be with us for several years to come. Component-based structured authoring enables you to create documents faster and to assemble, format, publish, and even deliver them automatically.

The Five Dimensions of Content Standardization™

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 in 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

Dimension One: Output Type

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

- A Clinical Study Report
- An SmPC
- A PBRER
- A clinical trial summary on ClinicalTrials.gov
- A P.8.2 Post-Approval Stability Protocol and Stability Commitment

A standardized output type provides a model to ensure that all requirements for a particular output are met. For example, the standardized output type for a UPSI defines rules for what content to include in the Highlights and what content to include in the main body of the label.

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 *de novo* 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.

Today, the output types included with a submission are typically delivered as PDF documents organized by the XML structure defined by the eCTD. In fact, the eCTD itself is one giant output type. Clinical trial summaries for ClinicalTrials.gov are delivered to the public as web pages. They are based on an underlying XML structure. Those summaries may never be a "document" at all.

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.

When a Document Is Not a Document

Pharma content is usually delivered in documents. These documents may be delivered in PDF format, Word format, a spreadsheet, or even a PowerPoint file.

Increasingly, pharma companies are delivering content in mobile- and web-friendly formats without need of a document. This content is delivered as a set of components that can be viewed on their own or organized into a flow.

Let's look at a recent example of delivering clinical content without a document: ClinicalTrials.gov.

The ClinicalTrials.gov website provides a collection of clinical trial summaries. The public can search the database for upcoming trials, for trials that are recruiting, and for completed trials that have results. Users can search and filter by various criteria. They can then view each trial summary as a web page.

Rather than page through a document looking for specific information, users can skim an overview and use a navigation menu to get closer to what they want to know. They can then click through to see more details in a particular section.

The content is provided in a series of components and assembled into the ClinicalTrials.gov output type. There is no PDF or Word document involved.

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 in pharma content include:

- Adverse Event
- Therapeutic Indication
- Mechanism of Action
- Stability Data
- Analysis Set

Pharmaceutical content traditionally uses the concept of components by providing headings and subheadings within a document. However, these components only become modular and reusable when they are used in conjunction with a component-based structured content management system.

Components may also be defined at a more granular level within a section to increase the potential for content reuse.

For example, the Adverse Events section of a clinical study report can be structured to include a ***summary*** component type and a ***detail*** component type for each adverse event.

Authors use the summary component to provide information that is intended for reuse, such as into the clinical summary reports in the eCTD Module 2 and into the company core data sheet for reuse in labeling. Authors use the detail component type to provide data-heavy information that has fewer instances of reuse. The detailed narrative and exact data about that adverse event may only be appropriate for the clinical study report for that particular trial. Or it may have several opportunities for reuse, depending on your reuse strategy.

The words *summary* and *detail* do not appear in the final output. They are not headings. Rather, they are internal guidelines that help authors write modular components in a consistent structure that facilitates content reuse.

Some components are unique to an output type and do not get reused. Other components have high potential for reuse. Components that are used in multiple documents and other outputs should be single-sourced. Single-source means that you create one component that you can reference into any output that needs it. You do not maintain duplicates or copy/paste the component. It is literally one file that the output can point to wherever that information is needed.

"Single-sourced components developed according to standards make content creation faster, improve content quality, and reduce 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 that are mixed and matched from different authors, different geographies, and different times of creation all flow together in a cohesive, comprehensible way. You also ensure that when a component is viewed on its own, it contains all relevant information, in the expected structure.

Standards for a Global Impact

Component standards improve translation quality while reducing the cost and time involved 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 because the source content is an exact match within the translation memory.

Translation memory is a database that stores approved translations so that the translations can be reused. When a sentence or phrase is translated, the original (source) language and the translation are stored as a pair in 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, clinical reports have traditionally included many long, dense paragraphs that 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, particularly for integrating data into the content
- When to use notes

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. And 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. These standards enable content reuse from the SmPC to the patient information leaflet and the lay summary, and from the body of the USPI into the Highlights section. This unity is important to reuse content successfully.

Bringing Data and Content Together

To automate the flow of data into the content, the system must be able to identify what data to put where. We do this identification using variables in the content. Variables act as placeholders. The variables connect to a source that provides the data. The data replaces the variable as the content is developed.

To take a simple example, here is a sentence that contains a variable:

- The mean age was <AgeMeanTreatment> years in the treatment group and <AgeMeanControl> years in the control group.

Here is the same sentence with the data filled in:

- The mean age was 35 years in the treatment group and 37 years in the control group.

Some sentences may contain several variables, for information such as:

- ID or name of the study group or arm
- Numbers and percentages for various results
- Name of substance or excipient

Variables are part of your data integration strategy. They enable content reuse by separating the unique information (data) from the standardized sentence that describes that information. This separation makes the sentence reusable in many contexts. It also enables you to translate the sentence into other languages one time, rather than paying for a new translation every time the data changes or it is copied and pasted from one document to another.

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, and that the data and content appear together where required.

Dimension Four: Sentence

Grammar and style rules govern how words and data are combined into sentences. Sentence standards are particularly important for ensuring that data flows logically into the content. Some content requires that the data be placed in the middle of a sentence, using variables as placeholders until the data is available. Other content separates the introductory sentence (or narrative paragraph) from the list or table that provides the data.

Sentence standards are 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.

Sentence standards include:

- Grammar
- Style

Grammar

Most of us learned about grammar back in elementary school (sometimes referred to as 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 encompasses 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 your content components interchangeable and their assembly seamless. It also helps content that is 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 the Oxford (serial) comma
- Avoid idioms, jargon, and colloquialisms
- Use gender-neutral pronouns wherever possible

Standards ensure that sentences make sense when the 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 will select at least one style guide from the various options and try to enforce it across the enterprise. Style guides can also include how to create citations for online and print content.

The Associated Press Stylebook

The Associated Press Stylebook (often referred to as the AP Style Guide) is an English language style guide that was originally created by American journalists working with the Associated Press. Even though the guide was written for journalists, over the years it has become one of the leading style guides for medical writing. As of February 2022, the AP Style Guide is in its 55th edition.

AMA Manual of Style: A Guide for Authors and Editors

The AMA Manual of Style is the guide used by the Journal of the American Medical Association (JAMA). As of February 2022, the AMA Style Guide is in its 11th edition.

NLM Style Guide for Authors, Editors, and Publishers

The National Library of Medicine style guide is written by the International Committee of Medical Journal Editors (ICMJE). It is often used to form proper citations and is the basis for the format of PubMed and MEDLINE citations. The NLM style guide uses the ANSI/NISA Z39.29-2005 (R2010) Bibliographic References standard.

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, in a clinical trial, a "patient" can mean something different than a "participant" or a "volunteer."

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 regulatory compliance
- Enforce accuracy
- Improve readability
- Lower the cost of translating content
- 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

Using the wrong term increases the risk of health authorities requiring revisions to the content, delaying approvals. For example, a mistake as simple as using "participant" instead of "patient" can change the meaning of a sentence. Using the wrong term can cause the data to appear inaccurate.

When used incorrectly, terms with legal implications can open your company to a potential lawsuit. Correct use of such terms is crucial.

For example, the ISO IDMP standards provide standardized definitions for the identification and description of medicinal products for human use. European and North American health authorities now require pharma data and content to adhere to IDMP.

Product and Brand Names

Strict adherence to product and brand names mitigates risk and ensures comprehension for reviewers and users of the content. For example, some products have different brand names in different countries. Using the wrong name in a piece of content can not only cause confusion for readers, it can delay regulatory review or create legal challenges, depending on when and where the wrong term is used.

Prohibited Terms

Every company has a list of words it wants authors to avoid. A term may be prohibited for a variety of reasons:

- Legal and regulatory
- Clarity
- Consistency
- Comprehension
- Brand
- Translation

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 term for replacement, is annoying at best. At worst, you risk another prohibited term being selected in its place.

Terms Used for Data Exchange

One advantage that pharma companies have over many other industries is that terminology standards already exist. Often, terminology is defined as part of data standards with the intention of facilitating data exchange within and across systems. These same terms are generally also appropriate for use in narrative content.

For example, health authorities in Europe, the United Kingdom, France, the U.S., and Japan have adopted the International Nonproprietary Names (INN) standard[1], developed by the World Health Organization. The INN standard provides a unique name for a pharmaceutical substance or active ingredient that is globally recognized and is public property.

CDISC maintains several controlled terminology[2] sets as part of several data standards. The organization also provides a glossary to harmonize terms and definitions across the industry. The National Cancer Institute hosts these terminology and glossary resources, which are available for download at https://datascience.cancer.gov/resources/cancer-vocabulary/cdisc-terminology.

[1] https://www.who.int/teams/health-product-and-policy-standards/inn
[2] https://www.cdisc.org/standards/terminology/controlled-terminology

The MedDRA[1] terminology standard developed by ICH is available in more than a dozen languages. MedDRA includes terms related to pharmaceuticals, vaccines, and drug-device combination products intended for human use.

What Not to Manage

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

Common words that are 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.

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 especially won't check the documentation after they've become familiar with the standards and believe they have memorized everything.

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 authors to help authors use the right terms (and avoid the 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.

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.

[1] https://www.ich.org/page/meddra

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
- Exchange information directly with partners, CROs, sites, and regulatory agencies without locking it into documents first
- Increase translation quality while significantly reducing cost and turnaround time

The framework combines the foundational principles of component-based structured content with industry standard FAIR guidelines.

Summary

The pharmaceutical industry is embracing the possibilities of component-based structured content. Component-based structured content provides many capabilities:

- Content reuse from a single source of truth
- Automated content assembly, formatting, and publishing
- Digital information exchange
- 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 risk, time-to-market, and cost, while increasing content quality in all languages.

As one of our pharma customers says, "It's a one-way street. We have to move in this direction and there's no turning back."

About Content Rules

Content Rules is the only end-to-end service provider with experience in pharmaceutical content. We have helped numerous 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. We can help you solve your complex content challenges. ***Contact us*** to see how we can help you!

Proprietary Methodologies

Our services include proprietary methodologies owned or licensed by 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™

Resources

- The Pharma Content Evolution: Content Reuse and Automation (white paper)
- Content Rules Pharma Blog (articles)
- Content Transformation: Breathe New Life into Legacy Content (eBook)
- Content Rules Services: Pharma and Biotech
- Content Rules Services: Medical Device