content rules®
the global content experts®

Structured Content for Pharma

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

"Pharmaceutical companies share a common goal of getting drugs to market faster. To achieve this goal, they need to reduce the time it takes to create and deliver content."

According to The University of Texas at Austin[1], pharmaceutical companies lose an average of $1 Million per day when drug approvals are delayed. Several factors can contribute to approval delays. Some of these factors relate directly to content:

- **It takes so long to create content, the submission or application is delayed**
- **The submitted content is incomplete or inaccurate**
- **The submitted content is difficult to read, forcing health authorities to work harder and go slower to decipher it**
- **The submitted content is locked in long documents and any change triggers a new review of the entire document**
- **Content is not standardized beyond the minimum requirements to comply with the eCTD tree**

Compared to other industries, it takes pharma companies about **five times longer** than it should to create content. The content ecosystem used by most pharma companies is decades out of date. The tools are inadequate and hampered by legacy processes and traditions.

The result is a vast amount of content — much of it redundant — locked in monolithic documents where it is hard to find, hard to update, and hard to use.

Pharma content is extremely important to the business. In fact, without the content, there would be no business.

[1] https://news.utexas.edu/2020/06/19/old-drug-standards-delay-new-drug-approvals/

To paraphrase the old adage, "If it's not documented, it didn't happen." Regulators require detailed documentation to support every drug application. Content drives decision-making throughout the drug development, application, and approval process as much as data does.

Most pharma companies know that there's a better way to "do" content. Many are talking about structured content and what they could do if they implemented a structured content management system. However, the industry suffers from a lack of widespread knowledge around state-of-the-art content processes and technologies.

In this eBook, we help you understand:

- **How your current content ecosystem slows down time to market for drug products**
- **How a structured content ecosystem results in better content, faster**
- **Examples of content reuse in three content pipelines**
- **The importance of XML, CDISC, and content standards**
- **Next steps for digital content transformation**

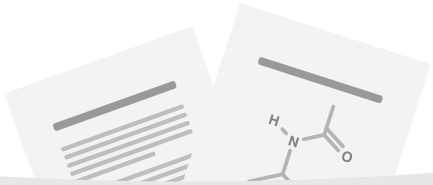
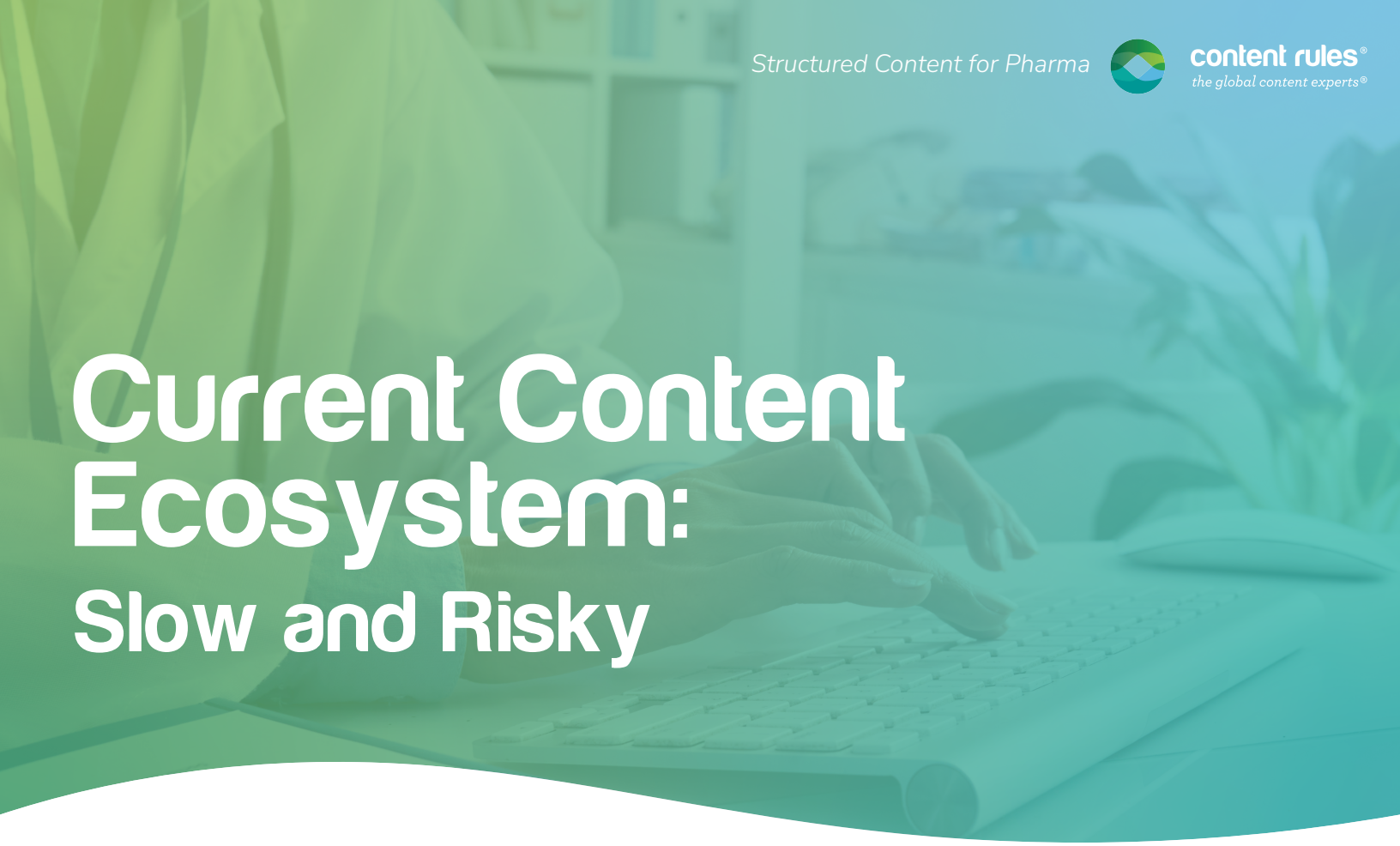


Table of Contents



Current Content Ecosystem:
Slow and Risky

The content ecosystem in use by most pharma companies is inadequate and outdated. It typically includes these tools:

- **Microsoft Word to create documents**
- **A document control system such as Sharepoint or Documentum**
- **A Regulatory Information Management (RIM) system used to consolidate, format, and submit documents to health authorities and regulators**
- **Systems to publish physical labels and packaging**
- **Files in many formats (including documents, spreadsheets, and presentations) to exchange information with internal and external partners, CROs, sites, and regulators**

In addition to old tools, the processes used to create and manage the content are complex and cumbersome.

Here's a simplified view of a typical workflow for a clinical summary report provided in Module 2 of the eCTD:

1. **Medical writer copies content from an existing Word document into a new Word document.**
2. **Writer makes additions, deletions, and changes for new document.**
3. **Data is copied from various sources and pasted into the document.**
4. **Content is copied and pasted into additional documents, often with minor tweaks such as verb tense or punctuation.**
5. **Writers and reviewers revise content, including all the redundant content created through copy, paste, and tweak.**

Even if only one section of the document needs review, the entire document must be checked out and opened. Writers and reviewers may get distracted by content in other sections. It may take several iterations before they can complete their reviews, especially when discrepancies or inconsistencies in the redundant content require someone to find and update the information in several documents.

The result is multiple variations of the same information, spread across dozens or hundreds of documents. It is almost impossible to find every variation when the documents need to be updated. The risk of inaccurate information making it into a submission and being sent back by regulators is high, wasting time and costing money.

The inefficiency of manually creating the same information multiple times during the content lifecycle results in lower quality and more expensive content that is fraught with risk.

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 the content. Most writers believe that revising the content makes it 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.

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

When the information needs to be revised, the updates must be made individually to each instance.

Making these updates is difficult and increases risk. First, every instance of the information must be found across the dossier and throughout the enterprise. Inconsistencies in terminology make searching multiple documents for the same information that much harder. Each instance must then be updated manually.

The risk that some instances of the information are never found and thus never updated can lead to queries and even inspections by regulatory bodies.

Example: Company Core Data Sheet

Let's look at the company Core Data Sheet (CDS) as an example.

The CDS includes information from multiple sources. These sources may include clinical trial data systems, clinical report documents, CMC data systems, CMC reports, and other product information stored in RIM and other systems. The CDS acts as a repository for reference information such as:

- **Dosing requirements**
- **Methods of administration**
- **Indications**
- **Packaging and storage requirements**
- **Incompatibilities**
- **Pharmacological properties**

The CDS is never published on its own. However, it contains content that is used in other documents, including:

- **SmPC**
- **USPI**
- **PIL**
- **Medications Guide**
- **Regional labels for global markets**
- **Packaging**

A lot of critical information goes into the CDS, and a lot of critical information is created from the CDS.

Many companies maintain their core data sheets in Microsoft Word files. Authors copy and paste content from source documents into the CDS. As soon as this happens, there are two versions of the information: one in the source document and the second in the CDS.

When labeling teams create the SmPC, USPI, and other documents based on information stored in the CDS, the same thing happens. Authors copy and paste content into their new documents.

Now, we have at least three versions of the same information:

If changes are made to the SmPC or USPI, those changes might also need to be made to the CDS. In fact, the change might need to be propagated across multiple internal documents and throughout the dossier.

As time goes on, iterations happen. Information is altered, corrected, and updated. Changes are made. Keeping all these documents in sync is a manual task. The more changes that are made, the more out of sync the information becomes.

Content Reuse Solves the Problem

Content reuse ensures that there are no discrepancies in the data and no mixed messages when one document needs to convey the same information as another document.

With content reuse, there is no copy and paste. You write and store the content one time. Then, you use that one version everywhere you need it.

The opposite of content redundancy is content reuse. And to have content reuse, you must have structured content.

Content Reuse and Automation

Structured content enables content reuse and automation throughout the content lifecycle.

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

Content automation harnesses the power of machines to do repetitive tasks related to content development, management, and publishing. Automation opportunities abound in pharma content. Here are some examples of automation that our customers have implemented:

1. **Populate reused content into new documents**
2. **Include/exclude content based on data values or other rules**
3. **Format content at time of publishing**
4. **Integrate data into content**
5. **Suggest content to reuse as authors write**

The pharma industry is rapidly moving toward automating submissions and data exchange between sponsors and regulators. Working groups such as ICH M11, TransCelerate Biopharma, and Accumulus Synergy are all pursuing technical solutions.

These technologies can only work well if the content they process is structured.

To learn more about how structured content helps pharma reduce risk and speed time to market, download our white paper: The Pharma Content Evolution: Content Reuse and Automation.

Structured Content Ecosystem:
Faster and Safer

The only way to reduce risk and produce pharma content at scale is to adopt a component-based structured content ecosystem.

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

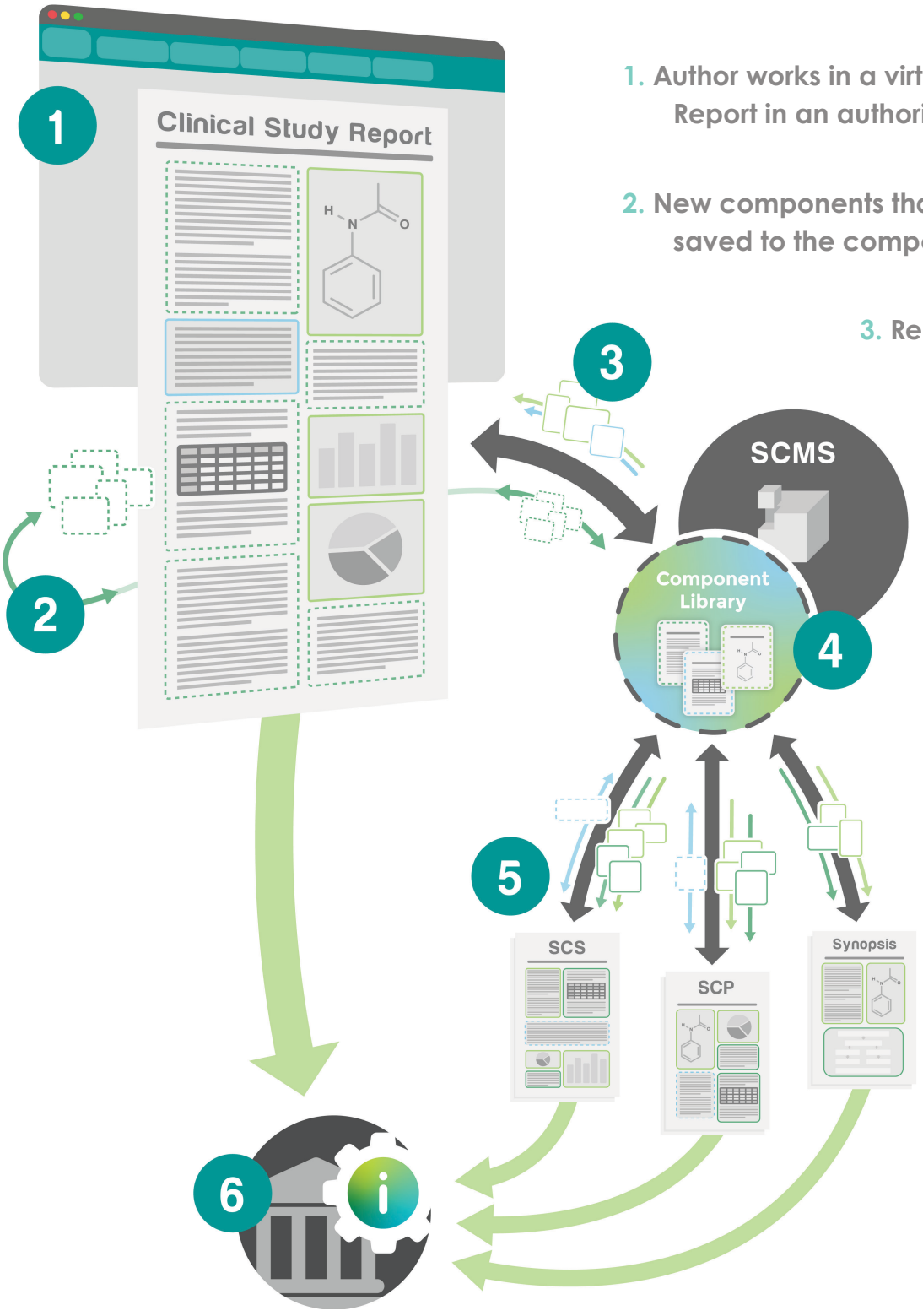
What Is Structured Content?

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**
- **Organized and stored in a centralized repository**
- **Publishable to a variety of formats**

To produce documents, components are assembled into the required order and hierarchy. The assembly can be automated by a structured content management system (SCMS). Much 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.

Structured Pharma Content - Illustrated



1. Author works in a virtual Clinical Study Report in an authoring interface

2. New components that are authored in-line are saved to the component library for later reuse

3. Reused content can be brought in directly from the component library. You don't need to re-write these standardized sections each time

4. SCMS stores, tracks, and automates single source of truth for content

5. Option to reuse components in other downstream content

6. Published outputs can be submitted to health authorities

Example: Clinical Content Pipeline

The eCTD Module 2 contains several summary documents. Much of the content in the summary reports originates with Protocols, Statistical Analysis Plans (SAPs), and Clinical Study Reports (CSRs). For example, the introductory sections of the CSR repeat a lot of the information first authored with the Protocol Synopsis, the Protocol, or even the Investigator Brochure. The results sections of the CSR provide ample opportunity for standardized text to provide context for data integration. This content is also reusable in the summary reports such as:

- **Summary of Clinical Safety (SCS)**
- **Summary of Clinical Efficacy (SCE)**
- **Summary of Biopharmacology (SBP)**
- **Summary of Clinical Pharmacology (SCP)**
- **CSR Synopsis**

Some of the content developed for the CSR and the summary reports is also pertinent to the company Core Data Sheet (CDS).

In a structured content ecosystem, authors may create the content as part of writing a CSR. Or authors may create the content in a series of stand-alone components that are later assembled and reused in the various reports. When and how content is created is typically part of a pharma organization's structured content strategy. The workflows you develop need to suit your organization and your teams.

Either way, you are authoring in components. Even though anyone can view the components in the context of whatever document they want to open, the lifecycle of create, review, revise, and approve can occur outside of the process of document assembly. Reviewers and SMEs can provide input on just the components that are relevant to them, or they can view entire documents, or both.

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 may not be successful. In fact, most of the failed attempts at adopting structured content in the past were due to a lack of strategy, not a lack of technology.

A complete structured content strategy includes the following elements:

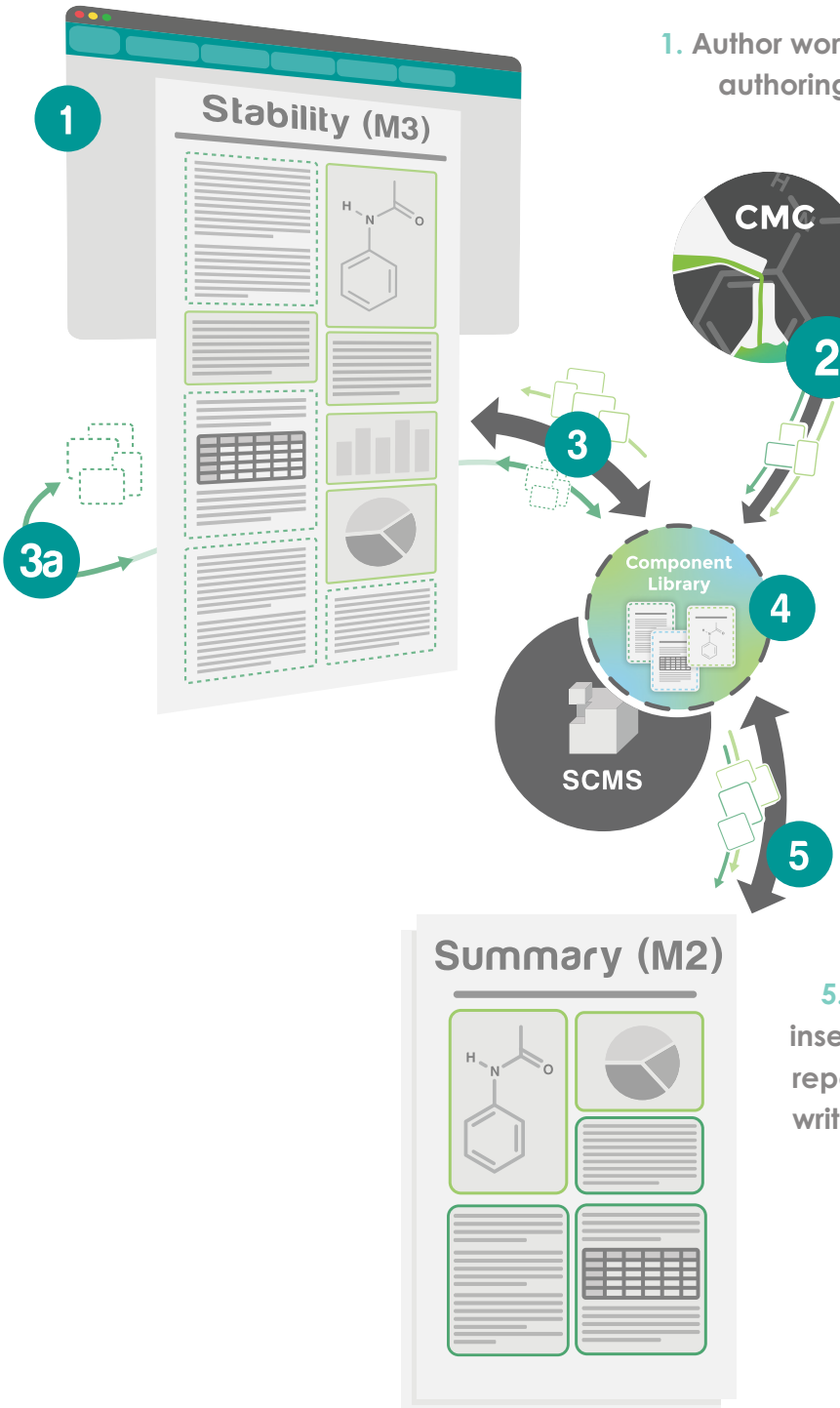
Content Models	A set of rules that define what content to include, under what conditions, in what order.
Structured Authoring Guidelines	Provides authors with guidelines and examples to help them create reusable, modular content according to standards.
Reuse Strategy	Rules and guidelines that define which content to reuse where and which mechanism to use to accomplish the reuse.
Reuse Map	A specification for configuring automated content 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 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.

Content Rules has a **proprietary methodology** for helping companies develop a structured content strategy that meets their goals.

www.contentrules.com info@contentrules.com

Example: Chemistry, Manufacturing, and Controls (CMC) Pipeline

CMC content includes information about drug manufacturing, GMP processes, and standard operating procedures. Some of this information is included in the eCTD for submission to health authorities.



1. Author works in a virtual Stability report in an authoring interface

2. Source databases provide data values and tables through integrations

3. Stability report includes reused content and data automatically

3a. New components that are authored in-line are saved to the component library for later reuse

4. SCMS stores, tracks, and automates single source of truth for content

5. Summary documents automatically insert reused content created with the full report; authors write only what needs to be written

Using a structured content strategy, CMC teams can create and organize content according to the standards defined in their content models and taxonomy. Each Batch Analysis component follows the same structure as every other Batch Analysis component.

Each SOP follows the same structure as every other SOP.

Every component of content is tagged with the relevant metadata to ensure that it can be found and accessed. The reuse map ensures that the SCMS automatically populates any planned content reuse. If the reuse strategy allows for manual reuse, authors may also reuse other content according to the structured authoring guidelines.

"A structured content ecosystem starts with a structured content strategy."

Once the content is componentized, tagged, and structured, other systems can use it. The summaries in Module 2 and full reports in Module 3 can automatically reuse relevant components. The CMC team can easily deliver SOPs to regulators who inspect manufacturing processes and documentation.

The most effective way to enable content reuse and automation at scale is to ensure that your content adheres to your content standards.

The 5 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 (findable, accessible, interoperable, reusable). 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:

1. **Output Type**
2. **Component**
3. **Paragraph**
4. **Sentence**
5. **Word**

For more information, download our white paper: **The Five Dimensions of Content Standardization™ for Pharma**

Example: Labeling Content Pipeline

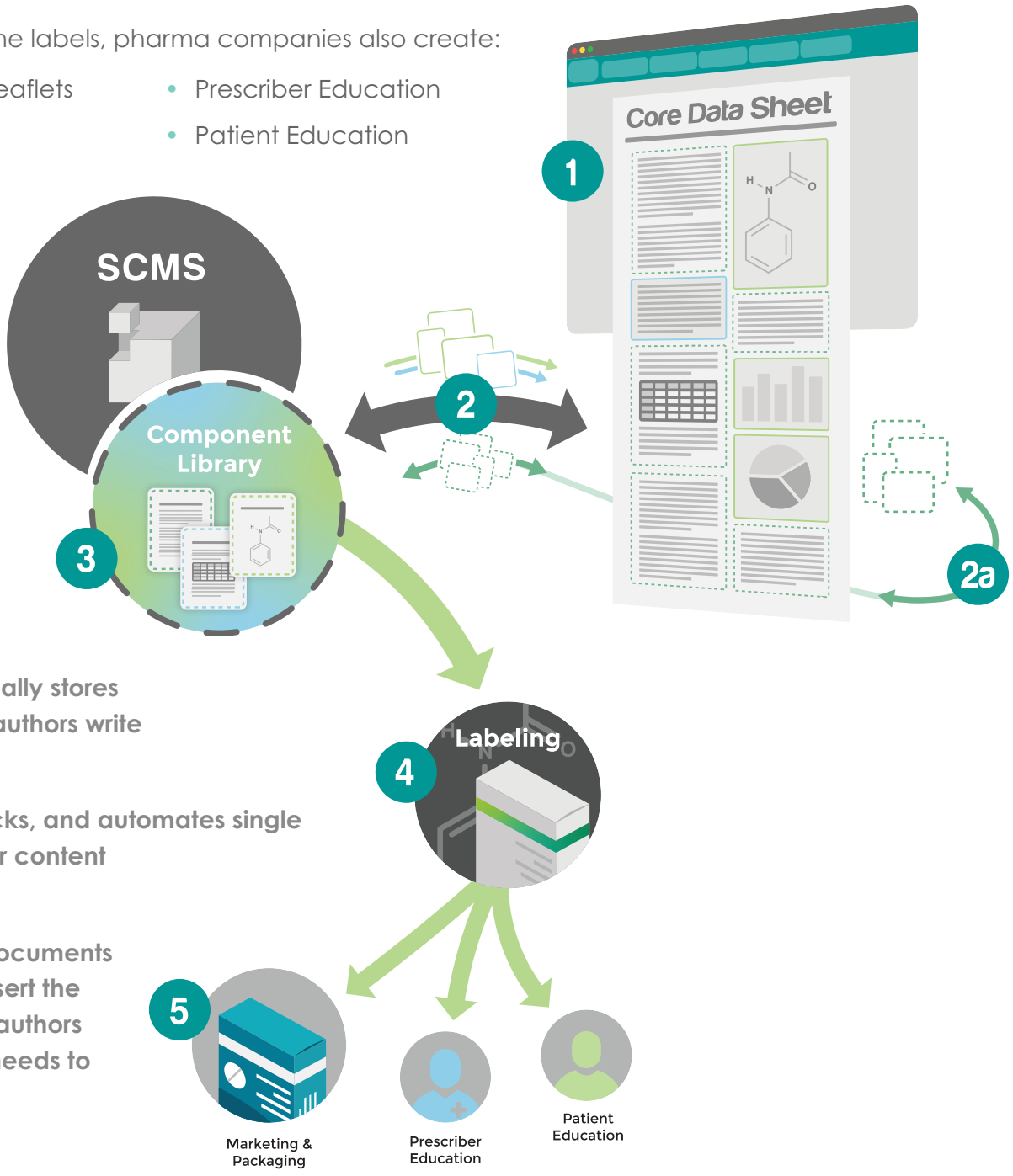
Every global region has different requirements and templates for what information to include in labels and where to place information on the label. The same information is used in additional documents where needed. Each label contains a combination of reused and unique content.

From the CDS and the labels, pharma companies also create:

- Packaging and Leaflets
- Marketing
- Prescriber Education
- Patient Education

1. **Author works in a virtual Core Data Sheet in an authoring interface**

2. **Virtual CDS includes reused content and metadata automatically**



2a. **SCMS automatically stores new content as authors write**

3. **SCMS stores, tracks, and automates single source of truth for content**

4. **Other labeling documents automatically insert the reused content; authors write only what needs to be written**

5. **Option to reuse components in downstream labeling, packaging, and education content**

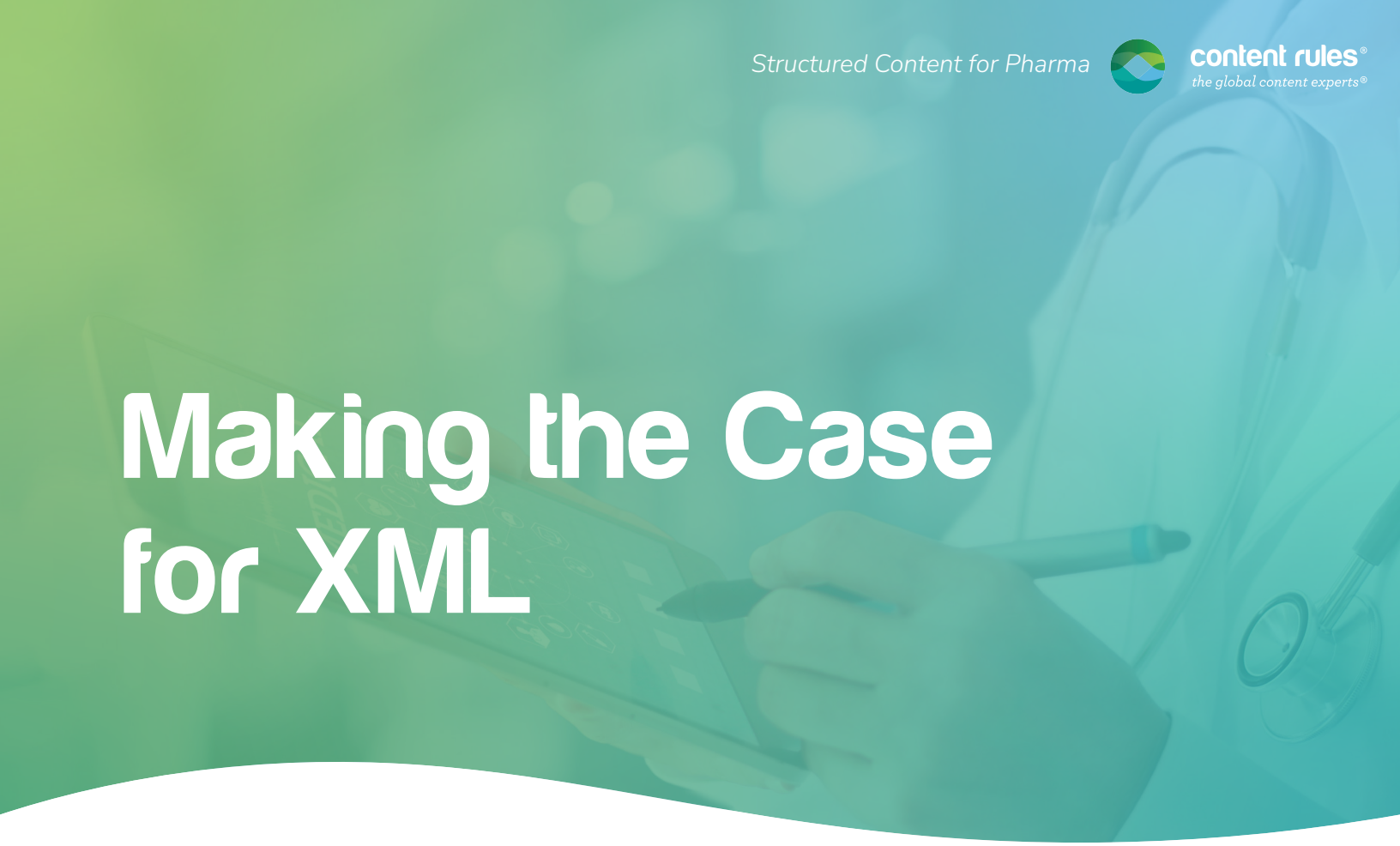
When content is created according to content standards, each piece of content can be reused in any of the outputs. Even though the content originates in different places and is written by different authors, standardized content fits together seamlessly. Content standards also ensure that when variable placeholders within sentences, paragraphs, and tables are replaced by data values, the surrounding text remains grammatical and meaningful.

Additionally, localized labels are much less expensive to produce. Each component is translated one time for each language, no matter how many outputs reuse that component. If a component is updated, only that component is sent out for translation. If the rest of the document did not change it does not need to be sent again, translated again, or reviewed again.

Consistency in words and sentences also improves the quality of each translation. Translators do not need to translate multiple words that all mean the same thing. They have fewer queries, resulting in fewer delays and less demand on subject-matter experts' time.

"Content standards also ensure that when variable placeholders within sentences, paragraphs, and tables are replaced by data values, and surrounding text remains grammatical and meaningful."

Any content that must be created *de novo* for a particular language or market can be created and inserted into the document where it belongs. These unique pieces of content do not change the common content or affect the single source of truth for the reused components.



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 the level of content reuse and automation that pharma companies require, 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, including life sciences.

XML is a robust and mature standard that is used successfully in every major industry and vertical.

XML Makes Content Fair

XML helps ensure that content is FAIR – findable, accessible, interoperable, and reusable. FAIR principles have been widely adopted for data. The FAIR principles apply equally well to structured content.

Findable

XML uses semantic tags that provide meaningful identifiers to 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 pharma authors is that they do not need to understand XML to create content.

Medical writers, safety analysts, and other authors can create clinical reports, CMC documentation, global labels, patient and professional education, and marketing collateral in an XML-based system without ever seeing the XML itself.

Why XML No Longer Fails in Pharma

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

The failure of XML in a pharma setting 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 an authoring environment that looks and feels a lot like Microsoft Word or Google Docs. The tool is called Fonto/XML. It is available for many component-based structured content management systems. Some big pharma companies have already adopted Fonto.

"XML is a robust and mature standard that is used successfully in almost every industry around the world."

Using Fonto, you can create content in a linear fashion. In other words, you can write a document and Fonto 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.

They can easily reuse components that have already been written simply by inserting the content where it should go. The technical aspect of content reuse is taken care of behind the scenes.

XML Standards

Pharma companies are well versed in the usefulness of data standards. They have led the charge toward standardizing the terminology and structure of data so that information can be exchanged between systems, companies, and regulators. One of the major organizations working on pharma data standards is CDISC, a working group that boasts more than 500 member companies.

"If we are going to solve the problems of time to market and content drift, we need to adopt modern authoring tools."

Like CDISC data standards, XML defines the core standards for content, including models, domains, and specifications for content representation. XML facilitates the sharing of structured content across different information systems.

You Need Structured Content

Pharmaceutical 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, pharma companies cannot achieve their business goals of reducing time to market and mitigating risk.

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

Process

We need to change how pharma "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 reduces risk, saves time, 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 pharmaceutical 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.

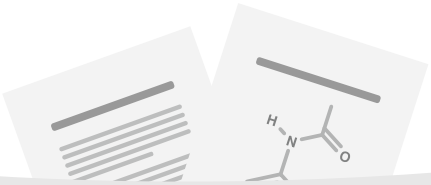
XML has been a challenge for pharma companies in the past. The authoring tools were not mature enough to hide the complexity from authors. Other obstacles included the lack of experience connecting XML to other systems. Today, there are mature tools and experienced people to make these connections work.

In addition, the standards in data and content have matured. Regulators are working with sponsors to develop XML standards to streamline submissions. The Fonto/XML authoring environment is comfortable and powerful. The Content Rules proprietary methodology of structured content strategy and the Five Dimensions Framework help pharma companies make their transition to structured content a success.

Pharma is poised to leapfrog the learning experiences encountered by other industries. Content Rules is poised to help get you there.

Resources

- 5 Dimensions of Content Standardization
- Content Rules Pharma Blog
- Content Transformation: Breathe New Life into Legacy Content
- Content Rules Services: Pharma and BioTech
- Content Rules Services: Medical Device
- The Pharma Content Evolution: Content Reuse and Automation



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

About the Authors

Val Swisher

Founder and CEO, Content Rules, Inc.

Val enjoys helping companies solve complex content problems. She is a well-known expert in content strategy, structured authoring, global content, content development, and terminology management. Val believes content should be easy to read, cost- effective to create and translate, and efficient to manage. When not working with customers or students, Val can be found sitting behind her sewing machine working on her latest quilt. She also makes a mean hummus.

Regina Lynn Preciado

Senior Director of Content Strategy Solutions, Content Rules, Inc.

Regina Lynn Preciado is the Senior Director of Content Strategy Solutions at Content Rules. She leads content strategy teams to help our customers adopt structured content successfully. Regina has helped organizations of all sizes make content work for people (instead of the other way around). She works with communicators in marketing, documentation, support, and training — sometimes all at once! Her clients include tier one companies in pharma, biotech, high-tech, manufacturing, and financial services. She lives a dogspotting lifestyle, and her professional goal is to work a picture of her dog into every presentation, in a seamless and logical way.