content rules®
the global content experts®

The Pharma Content Evolution:
Content Reuse and Automation

Val Swisher
CEO
vals@contentrules.com

Regina Lynn Preciado
Senior Content Strategist
reginap@contentrules.com

Contents

"Pharma is ready for a content evolution"

As an industry, pharmaceutical companies require a lot of content. So much content that pharma companies support an entire sub-industry of technology vendors that develop products specifically to help pharma companies manage their content. These products have mainly been developed to support traditional pharma processes.

For a long time, this arrangement has worked. Many pharma companies have gained efficiencies and improved the time, effort, and cost of creating so much content.

In recent years, pharma leadership has realized that it's not enough. Take the 2020 COVID-19 crisis as an example. The COVID-19 pandemic put tremendous pressure on drug-development organizations to accelerate product development far beyond anyone's wildest roadmaps.

While the entire world watched, pharma teams had to analyze the virus, develop potential vaccines, test the vaccines, and achieve emergency approval from regulatory bodies – condensing a years-long process into approximately one year.

"The fast-track part [of producing COVID-19 vaccines] were regulatory approvals, funding, data analysis and submission to the FDA (Food and Drug Administration).

Those are all paperwork items.

What was not fast-tracked was enrollment of patients, clinical follow-up of these patients, capturing the events which occurred and the follow-up. These trials were executed very well."[1]

- Dr. Andrew Badley, Infectious Diseases Physician At Mayo Clinic and Head Of Mayo Clinic's Covid-19 Research Task Force, Quoted In Mayo Clinic News Network Article December 2020

Health authorities and drug companies are quick to point out that they did not skimp on the drug development or on studies of its safety. Rather, to gain their "fast-track" approval – the Emergency Use Authorization – they cut back on the amount of "paperwork" involved.

In other words, they didn't produce quite as much content.

The entire industry also experienced a wake-up call about how much time and effort their legacy processes are costing them each day.

Unfortunately, adopting Microsoft Word, developing templates, and rigorous document control can only get them so far. Long monolithic documents in which every section is uniquely created (or copied and pasted from other documents and then tweaked) does not scale. Traditional processes prevent pharma companies from achieving their business goals: reducing time to market for new drugs and developing more products in the same time frame.

Using traditional document-based methods, pharma companies cannot:

- Automate content creation, revisions, assembly, or delivery
- Integrate data with content
- Create, update, manage, and translate content once and reuse it everywhere they need to say the same thing, in multiple geographies, submitting to multiple health regulatory agencies

Pharma is ready for a content evolution.

[1] https://newsnetwork.mayoclinic.org/discussion/will-fast-tracked-covid-19-vaccines-be-safe/

The Pharma Content Evolution

With standardized, structured content, pharma companies can still deliver the required documents in the required formats while significantly reducing the time it takes to produce those documents. They can configure digital data flow to bring data seamlessly from its sources into the content, eliminating the risk of copy-and-paste or other manual efforts. Content can be reused where needed, automatically. This capability frees up writers to focus on creating the unique content instead of rewriting content over and over and over again in different documents.

With standardized, structured content, pharma companies can still deliver the required documents in the required formats. But now, pharma companies can also **significantly reduce the time it takes to produce those documents**. In our experience, using standardized, structured content enables pharma companies to create content five times faster.

This time savings comes from many capabilities:

- Automate the flow of data from a source of truth into the content that presents that data
- Automate the creation, revision, and production of documents
- Streamline the writing process so that medical writers 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 reuse of exact same content rather than creating redundant content
- Prepare content for digital submission as some health authorities and regulatory bodies modernize their systems , in part by requiring submissions using XML

" It's a one-way street. We have to move in this direction, and there's no turning back. "

— a content transformation lead at one of our pharma customers

What Is Content Reuse?

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

Content reuse is how you create the dossier 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.

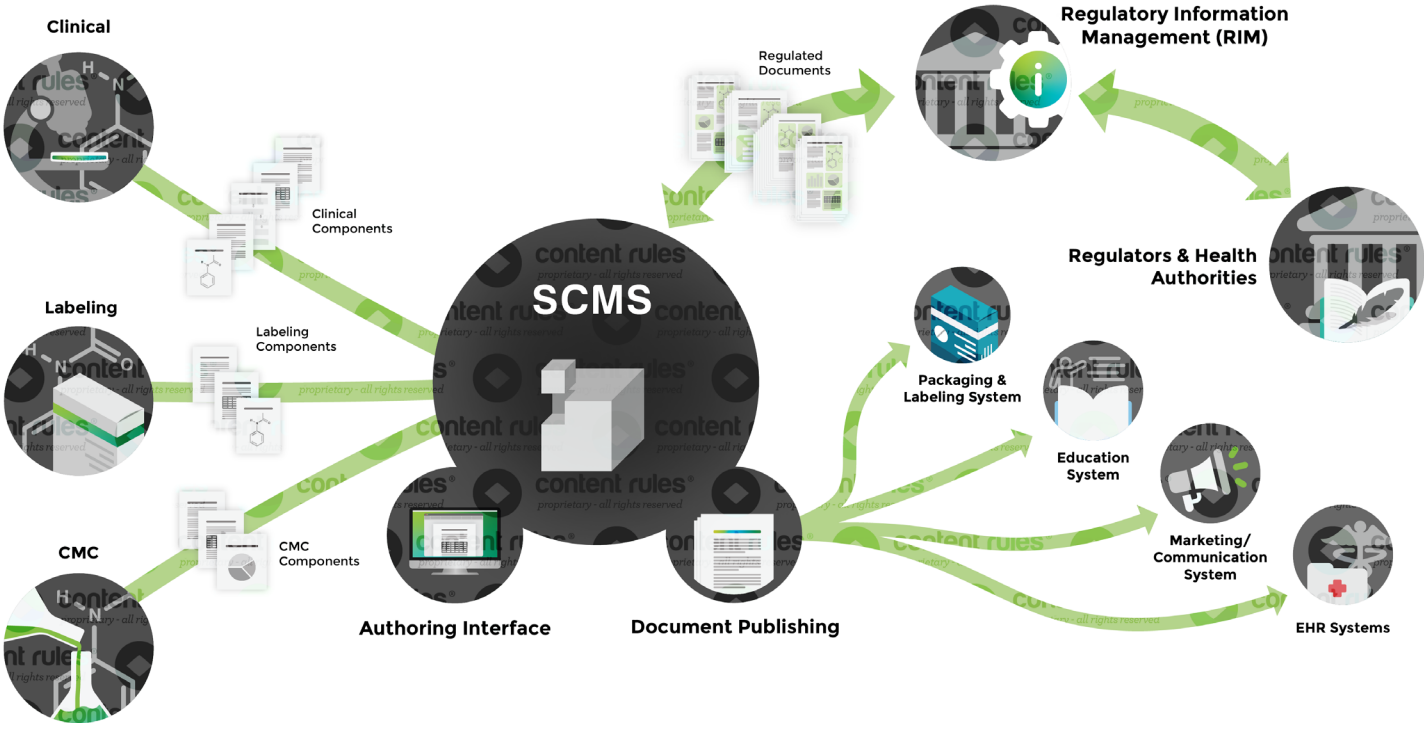
"The most effective way to leverage content reuse is to plan for reuse."

Content Reuse Across the Dossier

An analysis of any pharma company's dossier content typically shows a high level of **semantic similarity** along with a high level of **linguistic variation**. In other words, the content uses different words to say the same thing.

One reason pharma companies have not reused content in the past is that they didn't break the content down into small enough pieces. That meant that they couldn't reuse those pieces of content in the relevant sections of each document that required the content. This caused writers to endlessly copy, paste, and revise content as standalone documents while trying to meet aggressive deadlines.

When you look at dossier content in a standardized, structured content paradigm, you begin to realize that the dossier offers a plethora of reuse opportunities. Here are just a few examples from our pharma customers.



Clinical

In clinical reports, there is high reuse opportunity across Clinical Study Reports (CSRs) and the various summary documents, such as the Summary of Clinical Safety (SCS) and the Summary of Clinical Efficacy (SCE). The CSR contains content that traces back to the Protocol and the Statistical Analysis Plan (SAP).

While the data is different for every clinical trial, the language used to present that information does not need to differ from one document to another. It wastes time and adds cost when medical writers write new content to convey the same information when existing content could be reused instead. For example, there is no reason to write unique, custom content to describe the demographics of the study population. Standardized text descriptions for each demographic characteristic can be reused across all CSRs (and all other documentation of the study population, for that matter). Standardized text provides context for the data unique to that study.

Labeling

In labeling, content for all the regions around the world traces back to the Core Data Sheet (CDS), also known as the Company Core Data Sheet (CCDS). Some labeling content may also come from the Protocol, the clinical reports, and the Chemistry, Manufacturing and Controls (CMC) documentation.

A labeling reuse strategy accounts for regional differences in regulatory requirements, document structure, terminology, and so on. Every company's CDS is unique, but one thing they all share is the potential to structure the content in the CDS for optimum reuse to the USPI, SmPC, PIL, Medication Guide, and other labeling documents.

Standardizing the content in the CDS also goes a long way toward optimizing the labeling content for localization and translation.

Chemistry, Manufacturing, and Controls (CMC)

In CMC, reuse opportunities may relate to different views of the same data and to country-specific requirements. For example, batch reports often have a set of "core" information that is included for every country. For some countries, additional information must be provided to meet country-specific requirements.

Manual methods of creating this content – copy and paste, manually formatting tables, multiple reviews to validate the data – increase the risk of introducing errors into the content and waste time that could be better spent creating the unique narrative content.

Structuring this content to enable automated reuse of information reduces risk. It shortens the time required to create and review the document. Automating the reuse and assembly of core and country-specific information to produce the right documentation for the right country results in significant gains in reduced time, reduced cost, and greater accuracy.

Medical Education and Communications

Worldwide efforts to make pharma content more accessible to patients include the lay summary. The lay summary, which is required in Europe and encouraged in the United States, summarizes key information about a drug in consumer-friendly language. In fact, the European recommendations[1] say that the document should be understandable by a 12-year-old child.

Lay summaries have high potential for content reuse across other lay summaries, and across other patient education and communications content. Lay summaries also benefit from standardized, structured content to improve quality and reduce the time and costs of translation, as these documents must be made available in several languages.

[1] "EU Regulation 536/2014"

Content Reuse Directly Impacts the Business

These examples are just a few areas where strategic content reuse directly impacts pharma's real business of developing safe, effective drugs that improve people's lives.

It's time to replace risky, time-consuming processes with modern strategies and tools that are proven to help companies achieve their business goals.

It's time for a content revolution.

Replace Content Drift with Content Reuse

"Content drift" happens when writers copy, paste, and tweak content from one place to another…and then to another … and then to another. Even if the content continues to convey the same information, the language may get rewritten enough that it bears little resemblance to the original. Using traditional manual process, pharma companies do not have a single source of truth for content. Instead, they end up with redundant copies of the same content, each written and tweaked at different times. The more this content is copied and pasted, the further it drifts from its original.

Redundant content wastes time, costs money, and compromises quality in a number of ways.

It takes time to create, review, and manage the content. Reused content has already been reviewed and may not need to be reviewed again.

	Reused Content	Redundant Content
Time	Review and approve one time for a single source Update one time in a single source	Review and approve one time for every variation First, find every variation; then update every variation
Cost	Pay to translate the single source one time for each language	Pay to translate every variation for each language
Quality	Readers understand the meaning one time and recognize it each time they encounter the reused content	Readers must understand each variation and determine whether it has the same meaning as the other variations; requires extra work from the reader and may result in confusion

It takes more time and costs more money to have the content translated, which adds up quickly if you translate into multiple languages. Reused content is translated one time. Most companies save at least 10% on translation costs by switching to standardized, structured content.

Readers must work harder to extract the same meaning from content that uses different words to say the same thing. For example, a person reading a CSR is likely to encounter the same meaning in a summary or conclusion section as they did in the body of the document. Introducing a new way to say the same thing puts an unnecessary (and likely unwelcome) burden on the reader. This burden adds time and may even prevent the reader from understanding (or worse, approving) the content. Reused content is easier to understand.

Content drift makes it costly and cumbersome to keep the content in sync. If the content needs an update, writers must find everywhere that content was pasted and make the same or similar updates there. If the language was tweaked a lot, writers might not be able to find the content at all, and the updates won't be made. Reused content is easier and faster to update.

Content drift introduces significant risk of content becoming inaccurate. Risk mitigation requires humans and machines to continuously check and triple-check the content to ensure accuracy before it can be submitted. Reused content eliminates the risks of content drift.

Content drift wastes time, costs money, and compromises quality.

Content drift is extremely common across the dossier. Much of the time, copy, paste and tweak introduces unnecessary variations to the content. If the content was complete, accurate, and comprehensible the first time it was written, it should not be revised the next time the content needs to convey that same meaning. Likewise, if the revisions were necessary, then likely the original source should receive those revisions as well.

Content drift is inevitable in a document-based system. In a traditional setup, writers work from Word templates that provide authoring guidance and an outline. These Word templates do not provide standardized content or enforce consistency in terminology or syntax. They do not provide structures that support automation of data into the content. Nor do they optimize content for reuse. As a result, writers have a lot of freedom to write and rewrite however they prefer.

Unfortunately, this type of creative license does not move pharma closer to its business goal of producing a lot of content quickly, accurately, and affordably.

There is a better way.

“ We’re not creating art. We’re doing science. ”

— Structured content governance lead from one of our pharma customers

Content Reuse Glossary

Source content – the chunk of content that can be reused elsewhere

Reused content – the reused content in its destination

Granularity – the smallest piece of content that can be reused, such as a component, a paragraph, or a sentence

Verbatim reuse – reuse of content exactly as it is stored in the source, without any changes

Derivative reuse – content is reused and then edited; the system maintains a relationship between the original source and the revised content

What Is a Reuse Strategy?

You get the most ROI from content reuse when you approach it from a strategic perspective. A reuse strategy is a plan for which content you will reuse, how you will reuse it, and where you will reuse it.

Your reuse strategy identifies things such as:

- Which content is reused automatically
- Which content is reused manually
- What level of content can be reused
- Which mechanisms supply the reuse, such as variables and conditional logic to support “if/then” content reuse
- How reused content is managed
- Where content can be reused verbatim, without change from the original source
- Where content can be reused as a derivative, with some changes from the original source

While certain reuse best practices have been proven over time, there is no “one size fits all” for reuse strategies. One of our pharma customers defined a reuse strategy for clinical reports that allows for reuse at component and section level, but not at an individual paragraph or sentence.

As of this writing, another pharma customer is exploring the costs and benefits of a content reuse strategy at the paragraph level for global labeling. A more granular level of reuse can require more effort to manage. We’re helping them weigh the pros and cons to determine what’s right for them.

Regardless of how granular your level of reuse is, without a plan, your best efforts to scale content reuse will fail.

There's more to content reuse than breaking long documents into pieces and then putting them back together. To be successful reusing content, you need to create content standards.

Our reuse strategy design methodology incorporates standards for all five dimensions of content.

> Think of content standards the same way you think of electric standards or plumbing standards. You can buy any one of a thousand different kitchen faucets on the market, from any one of the various manufacturers, and it's going to work in your kitchen. It works because ANSI standards define the diameter of the pipe, the number of threads, the types of materials, and other allowable characteristics. You can "reuse" different faucets in any kitchen because underlying standards make sure everything fits together properly.

Five Dimensions of Reusable Content

One goal of creating reusable content is to make it 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. The content is simply the content, presented in a seamless, coherent way.

By "standardize," we mean creating a set of rules and guidelines for all content to meet.

Standardized content includes rules for every aspect of the content:

- Words that are preferred and words that are not allowed
- Style and grammar rules
- Voice, tone, and branding guidelines
- How sentences combine to create paragraphs
- How paragraphs combine to create sections or content components
- How components combine to create the final output

The five dimensions of reusable content, from largest to smallest, are output type, component, paragraph, sentence, and word. For content reuse to occur seamlessly, you must standardize the content across all five dimensions.

Output Type

Outputs are assemblies of content that get published and delivered, whether they are delivered internally to RIM teams or externally to health authorities, health care providers, or consumers. A standardized output type provides a framework for each output, such as the Protocol, CSR, USPI, technical document, lay summary, and so forth.

The process of standardizing output types includes identifying opportunities to reuse content. You can then configure systems to provide the right reused content component at the right time, automatically. For reuse that requires human intervention, you can train content creators when to reuse which content. For example, you can reuse content from the Safety section of the CSR into the CSR's Summary and Conclusion sections. You can also reuse it in the Summary of Clinical Safety (SCS) and the CSR Synopsis.

Regardless of the format you use to create, manage, and store your content, you can output your content to any format that you need. For example, you can create a Word version of a CSR or a PDF version of a label. You can publish a safety section to the web using HTML.

Component

The component is an independent unit of content that can be combined with other components to deliver an output. In your CDS, you might define separate components for Dosage and Administration and for Contraindication to facilitate reuse in the various labeling documents. In marketing communications, a component might include a product claim and its accompanying regulatory language.

Paragraph

The paragraph is the central building block of your content. Paragraph standards guide writers 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.

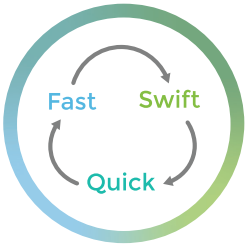
We're working with a pharma company to develop standards for shorter paragraphs and other ways to convey complex data clearly. Separate paragraphs provide narrative and interpretation. This logical grouping helps readers find and comprehend the data easily. Paragraph standards help ensure that the components fit together seamlessly when they are reused across different outputs. This unity is important to reuse content successfully.

Sentence

Grammar and style rules govern how words and data are combined into sentences. Sentence standards are particularly important for content that will be translated or read by non-native speakers of the source language.

Pharma content includes many opportunities to use variables as placeholders for data in standardized sentences. These placeholders are automatically replaced by the real data during the course of producing the content. Sentence standards ensure that sentences make sense when the data and words are combined. For automation to work, everyone needs to follow the rules.

Sentence standards also protect readers from shifts in grammar and style. Even if your medical writers are located all over the world and have different native languages, sentence standards enable them to write with a single voice. Readers don't need to do the mental work of adjusting for regional spelling, idioms, or even basic English conventions when encountering reused components.



Word

Words are the smallest standardized content unit. A collection of 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 "subject" or a "recipient" or a "volunteer."

Terminology standards protect reviewers, regulators, physicians, pharmacists, and consumers from needing to figure out whether a shift in terms indicates a shift in meaning.

Content Standards Enable Reuse

It is technologically possible to reuse content without standardizing it. Your structured content management system likely offers several mechanisms to support various types of content reuse.

However, without standardizing all five dimensions, content reuse can produce a terribly disjointed experience for the reader. Unstandardized content increases the risk of having health authorities reject the content or demand extensive revisions.

How to Create a Reuse Strategy

A reuse strategy is a plan for what content you reuse, how you reuse it, and how you manage it throughout its lifecycle.

This strategy defines:

1. **Granularity:** What is the smallest piece of content that can be reused?
2. **Matrix**: Where will the content be reused?
3. **Management**: Where will you store reused content?
4. **Governance:** Who can create, modify, or retire reused content?

1. Define Granularity

Granularity defines the smallest unit of content that can be reused. You can reuse content as small as a single character or as large as an entire document. The question, of course, is ... should you? And if so, when?

The smaller the granularity, the more reusable the content—and the more attention needed to manage it. The larger the granularity, the less reusable the content.

Your reuse strategy defines which types of content to reuse at each of the following levels of granularity:

- Section
- Component
- Element

Section Reuse

Section reuse is the practice of reusing an assembly of content in its entirety. An assembly of content usually contains several components. For example, you might reuse a "core" section of batch analyses reports across several country- and region-specific technical documents. Or you might reuse the safety subsection of the CSR Summary section into the CSR Synopsis.

With section reuse, you create and assemble the content one time. You can then lock that assembly so no one can add or remove content. Content creators can bring the entire section into their outputs without changing any content within the section. Systems can automatically include the section when content creators make a new output based on a standardized output type.

Component Reuse

A component is an independent unit of content that can stand on its own. It is a building block of content that can be used many times. Because it can stand on its own, it can be reused in many contexts without any change to the content.

Component reuse is the most common level of granularity for content reuse. With component reuse, individual components are reused exactly as they are. This level of reuse is why it is so important to define and create standards for each of your component types.

For example, you can create components to hold information about safety, efficacy, and study population. This information has high reuse opportunity across the core clinical reports and their various summary documents. This information also belongs in labeling and education content.

Structuring these components and developing standard to enable writers to create reusable building blocks streamlines the process of developing this content and enables automatic reuse everywhere that information is required.

Every reuse strategy we've ever worked on includes component reuse. Not all companies reuse sections or elements. But everyone we've worked with, without fail, reuses components.

Element Reuse

Element reuse happens at the paragraph, list, table, or sentence level. This level of reuse is often called fragment or snippet reuse.

Element reuse requires more planning and configuration than component or section reuse. You need to identify which elements can be reused and develop a management process for them.

2. Develop a Reuse Matrix

To identify your reuse opportunities at each granularity level, analyze your content and create a reuse matrix. The reuse matrix shows, at a glance, where reuse opportunities may exist across the body of content. It is not a fully detailed specification for configuring automated reuse in a structured content management system. Rather, it is a way to identify similar content based on an analysis of dozens of examples of each output type.

The following simplified example is for illustration purposes only.

	Protocol	Clinical Study Report	CSR Synopsis	Summary of Clinical Safety	USPI	SmPC
Clinical trial demographic info	X	X	X	X		
Most frequent adverse events		X	X	X	X	X
Method of administration	X		X		X	X

There are many possibilities for content reuse. Bear in mind that every new possibility comes with an associated management decision. The greater the variety of reuse types, the more effort you need to plan and manage the strategy. Also, the more reuse you build in – automated or manual -- the more important it is for all content creators to understand and abide by your reuse strategy.

3. Determine Reuse Management

Content reuse introduces some interesting possibilities into your content management operations.

Your reuse strategy should define where and how you store reusable content. Some structured content management systems use a folder structure. Others use a category or taxonomy structure. Some use both.

Various tools and services are available to identify reuse opportunities and create reuse libraries. These tools are quite valuable. Automating the process of locating reuse opportunities eliminates a lot of time and effort compared with manually searching for those opportunities. In addition, these tools provide reuse reports that offer insight into how your content creators are writing. You can use this information to help you standardize content across all five dimensions.

We advise against structuring your repository or reuse libraries by organizational silo. If you organize components and elements based on content rather than department, your reuse libraries can support any number of business reorganizations.

Any system that enables reuse will also track all the relationships across your content. You cannot lose content. Still, keeping it organized is a better option than making content creators hunt throughout a repository every time they need to find something.

4. Define Governance

Create a reuse governance strategy from the beginning. A governance plan can help you avoid the pitfalls that await those who leap into reuse with an entirely ad-hoc approach. We have seen companies spend a million dollars or more implementing a standardized, structured content ecosystem and not address governance. This almost always ends up as a disaster because the rules are not defined.

Your reuse governance strategy should define:

- Who can create or revise content that is reused across organizational silos?
- When does a unique component transition to a reusable component? What are the criteria and workflow for such a transition?
- Who owns reusable content: The team that created it, the team that uses it the most, a dedicated reuse manager, or someone else?

- When is reused content retired? Who can retire or archive it?
- What is the process for requesting changes to reused content?
- Who creates and maintains reuse mechanisms or automation?
- Who decides when to allow derivative reuse instead of verbatim?

We recommend answering the who questions with a role, rather than an individual or job title. Individuals change, and job titles often do not reflect what a person actually does on any given day.

Finally, determine where reused content is stored, who approves it, who maintains it, and any business rules for where the content can and cannot be reused.

Summary

Using standardized, structured content, pharma companies can significantly reduce the time it takes to produce content, from the dossier through education, marketing, and post-marketing communications. Pharma companies can automate systems to help ensure that the information is consistent across all documents in a submission, even if they are produced by different people at different times.

By standardizing content in all five dimensions, developing a reuse strategy, and implementing proven systems and processes to change how you "do" content, you can replace content drift with content reuse. You can speed content creation time by up to 400%. You can save at least 10% on your translation costs. And you can continue to scale content operations to meet the ever-increasing demands of the marketplace.

Content Rules has decades of experience helping life sciences companies achieve their business goal of speeding up "the paperwork" to get products to market faster without sacrificing safety or quality. A digital transformation of how you "do" content is not an undertaking to enter into lightly. But 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

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.

About the Authors

Val Swisher is the Founder and CEO of 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 is a senior content strategist with Content Rules. She helps companies transform how they organize, manage, and leverage content. Regina works with communicators in marketing, documentation, support, and training — sometimes all at once! Her clients include tier 1 companies in high-tech, life sciences, manufacturing, and financial services. She lives a dogspotting lifestyle.

Resources

- The Five Dimensions of Content Standardization (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