Unlocking new efficiencies:

How structured content authoring is streamlining the production of clinical documents for the pharmaceutical industry

Mati Kargren[1], **John April**[2], **Gina Clark**[3], **Jonathan Mackinnon**[4], **Aliza Nathoo**[5], **Elizabeth Theron**[6]

[1] Parexel International Co., Ltd., Taipei, Taiwan
[2] Eli Lilly and Company, Indianapolis, USA
[3] Pfizer, Groton, USA
[4] Parexel International S.L., Madrid, Spain
[5] Hoffman-La Roche Limited, Mississauga, Canada
[6] Parexel International (South Africa) (Pty) Ltd, Bloemfontein, South Africa

doi: 10.56012/xafs6978

Correspondence to:
Mati Kargren
mati.kargren@parexel.com

Abstract

Current practice requires clinical and regulatory documents to be created and updated manually by medical writers throughout a product's development. Conventionally, document content is unstructured, with free-form text, figures, and tables that the medical writer can arrange in any configuration. By structuring and standardising clinical and regulatory content, the pharmaceutical industry can shift from a document-based to a content-based approach. This transition will require adopting structured content management tools and common structures, and standardising content. In tandem, medical writers must evolve their skillset and ways of working, primarily through planning and producing content and adopting structured content authoring practices to facilitate content creation and reuse. This article introduces structured content authoring and outlines how the medical writing role in the pharmaceutical industry may soon evolve.

The hidden value of structuring content

The burden of unstructured information

In clinical research, medical writers create and update clinical and regulatory documents at multiple points throughout a product's clinical development. Conventionally, document content is unstructured, with free-form text, figures, and tables that the medical writer can arrange in any configuration. In terms of structure, aside from high-level section headers defined in the table of contents, medical writers are free to organise content as they see fit – provided they fulfil content requirements described in the authoring guidance.

As medical writers often develop clinical and regulatory documents independently of one another, each document contains unstructured information that is created and organised differently. For example, if one medical writer prepares a briefing document while another medical writer prepares a clinical study protocol, similar information is created and managed independently. Ultimately, if the writers do not have a tool or process to ensure consistency between the two documents, then an additional step is needed before finalisation where the medical writers need to align content between the two documents to avoid discrepancies.

As medical writers often develop clinical and regulatory documents independently of one another, each document contains unstructured information that is created and organised differently.

Another limitation of unstructured information occurs during document revisions. As the information in each document is not linked, independently revising documents can result in changes to the information's meaning that leads to the same information in different documents becoming increasingly divergent over time. Resolving this "information drift" is inefficient as this requires repeated, and deliberate, consistency checks that can be especially burdensome for authors working on tight timelines.

Harmonising between-document information involves additional complexity in that if only one piece of information requires revision, then the entire document must be checked and updated. Unless sections not undergoing alignment are locked, checking an entire document introduces the risk that stakeholders will reconsider content in sections that do not require revision. This, in turn, can lead to further cases where information starts to diverge among documents.

What is structured content and how does it work?

Rather than creating the same content across separate documents, a structured content approach is based on the "create once, use often" principle where information is created once as a content component (Table 1) and reused often across multiple documents.[1] To enable this, teams define, create, manage, and archive individual content components using a centralised structured content management tool (Figure 1 and Table 1). The tool tags defined content components with metadata, which allows users to identify and retrieve components for a particular purpose or deliverable. Much in the same way that metadata fields (e.g., recruitment status, age groups, phase, or funder type) can aid study search and retrieval efforts using the US ClinGov register (https://clinicaltrials.gov), metadata-tagged content facilitates its reuse capability and allows writers to perform a more targeted search.

Once these structures and tools are in place, a medical writer can generate a deliverable using

Table 1. Table of definitions

Term	Definition
Content components	Individual content components (e.g., the study design, participant characteristics, study interventions, etc.) which are defined, created, managed, and archived in a centralised repository and with minimal formatting details for the purpose of reusing the components across clinical and regulatory documents
Content standards	The set of rules and guidelines that govern content, including how sentences are put together to make paragraphs, how paragraphs are put together to make sections or components, and how components are put together to generate a deliverable
Structured content authoring	The process and rules by which an author creates content using defined structures that can be easily reused, repurposed, and automated
Structured content management tool	A centralised, platform that allows for creation, management, and reuse of digital content

a structured content authoring (Table 1) approach by populating the document structure with content that is either created *de novo* or reused from the tool's content repository.

Using content standards to improve authoring

When considering content standards and reuse, inspiration can be drawn from data standards as for data to flow between systems it needs to have certain standards for reuse. The four guiding data principles were designed and jointly endorsed by a set of stakeholders representing academia, industry, funding agencies, and scholarly

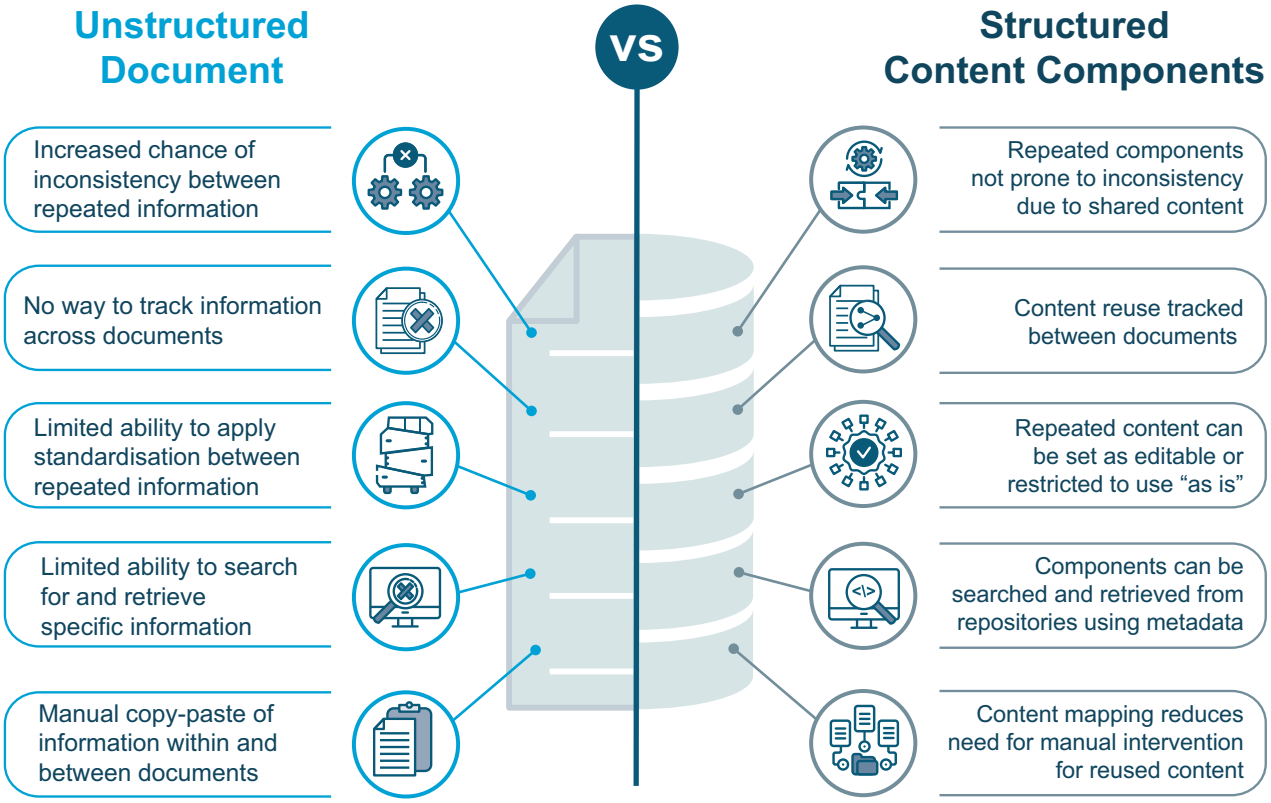

Figure 1. A comparison of information stored in unstructured documents vs. structured content components

publishers. These principles are commonly called the FAIR Data Principles: Findability, Accessibility, Interoperability, and Reusability.[2]

These same FAIR Data Principles can also be applied to structured content. By structuring content and adopting content standards (Table 1), the pharmaceutical industry can reduce document development time by enabling reuse of defined standard content. For example, TransCelerate Biopharma Inc.'s Clinical Content & Reuse Solutions include content libraries that provide select content standards that can be used in clinical study protocols and then reused in downstream documents.[3]

An example of structured content authoring

How an organisation defines content components will depend on how the organisation intends to use the content in the future. Our example of how structured content authoring can be applied to a clinical study starts with a study outline (Figure 2), a document that provides a high-level summary of the proposed clinical study. It includes sub-sections such as study design, overall rationale, study interventions, statistical methods, and so on. Using a structured content management tool, the individual subsections of a study outline can be defined as distinct content components. Once the team finalises the study outline, the medical writer can generate a draft protocol that is partially completed by automatically incorporating the study outline information based on a pre-specified content reuse plan and rules. In such a case, no manual manipulation is needed by the medical writer in copying and pasting content from one document to another. Furthermore, if the study development team has approved the study outline, the outline components can be locked for subsequent reviews, allowing for a faster and more targeted review process.

Similarly, when writing a clinical study report, the medical writers can populate the report shell by automatically pulling in relevant protocol components, e.g., background information, such as the study design, key participant characteristics, study interventions, and so on.

What does this mean for medical writers?

Medical writing skillset

As the clinical research landscape modernises and pharmaceutical companies deploy structured content management tools, the medical writer skillset will need to evolve to include content management principles. Medical writers will need to be trained on how to use structured content models, content standards, and content reuse authoring strategies in their everyday work. In turn, this will help medical writers concentrate on creating the unique *de novo* content rather than searching for or recreating content that has been developed elsewhere. Ultimately, the goal is to cut down on the manual intervention by medical writers in finding and transferring content. Furthermore, adopting automated content reuse will reduce errors and the need for consistency checks, thus allowing medical writers to focus on other tasks, such as interpreting clinical data or communicating with stakeholders.

The way we work

Much like a medical writer's skillset, the way in which medical writers work will also evolve to include new ways of approaching content development. Instead of developing information that resides in a single document, medical writers will need to work with content that is used throughout a product's development. This will require greater adherence to content standards as well as resisting editorial requests from stakeholders to rephrase content in a manner they prefer. Medical writers may be tasked with developing a protocol that contains content that will be used downstream, for example in a briefing document or a clinical study report. This content must be clear, easy to understand, and use agreed standards for terminology and acronyms. Similarly, the content must be devoid of any positional phrases, such as "see below" or "as mentioned above", or cross-references to other content sections that will not apply to downstream documents as these will be out of place in the location where the content is reused.

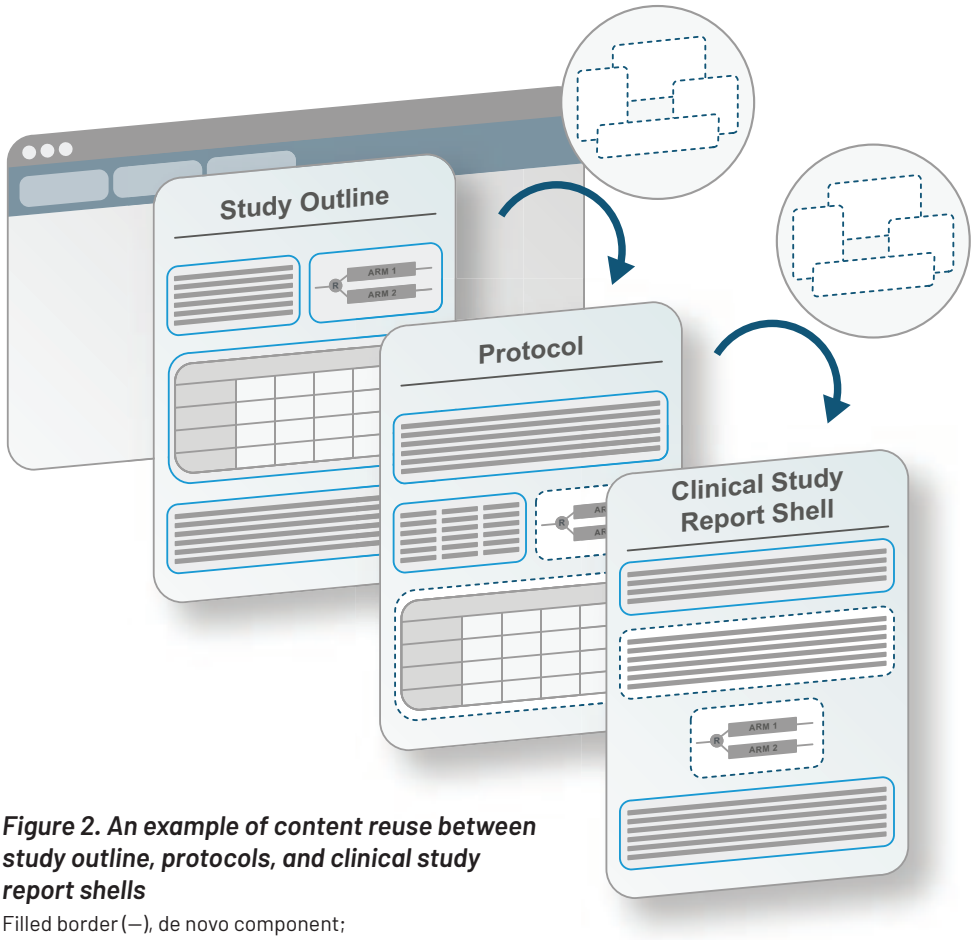

Figure 2. An example of content reuse between study outline, protocols, and clinical study report shells

Filled border (—), de novo component;
Dotted border (...), reused component.

Outlook

Technology has evolved to a point where the pharmaceutical industry has the capability to modernise the process of creating clinical and regulatory content. Technology-enabled content reuse allows organisations to facilitate the authoring process by using a "create once, use often" approach to develop clinical and regulatory documents.

> Technology enabled content reuse allows organisations to facilitate the authoring process by using a "create once, use often" approach to develop clinical and regulatory documents.

Although data standards are widespread in clinical development, structuring and standardising content is still in its infancy. Only some organisations have implemented initiatives to standardise content and increase content reuse. In the future, as structured content authoring matures, it is likely more cross-organisational coordination will emerge between pharmaceutical companies sponsoring clinical trials, contract research organisations, and regulatory authorities. To this end, the TransCelerate Biopharma Inc. Clinical Content & Reuse Solutions and more recently, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) M11 draft guideline (Clinical electronic structured harmonised protocol – CeSHarP) are important steps in this direction.[3,4,5] For the former, TransCelerate Biopharma Inc. has longstanding, publicly-available, technology-enabled templates and reusable library content. For the latter, the recent ICH M11 guideline provides a more global reference for the structure and technical properties of protocols, with the aim of enabling consistent and efficient exchange of protocol information between sponsors of clinical studies, investigational sites, independent review boards, regulators, ethics committees, and other related stakeholders.

At the organisational level, implementation of agreed standards such as the Clinical Content & Reuse Solutions and ICH M11 technical specifications will level the playing field, which will increase the likelihood of harmonisation of documents between stakeholders. In parallel, to effectively embed structured content authoring, organisations will need to implement authoring process changes, and content governance structures to adapt to content creation, reuse, and management practices in structured content management tools.

In addition, for true adoption success, time and resources must be allocated to ensure adequate training and support for users new to structured content management tools and content-based working practices. As organisations create, review, approve, revise, manage, archive, and, if needed, retire each piece of content individually, the requirement for users to manually perform these activities will diminish. Ultimately, structured content management tools

will facilitate the medical writer's ability to find, reuse, and repurpose content that will enable organisations to create content faster than developing unstructured content in individual documents.[6]

Conclusions

By structuring and standardising content, the pharmaceutical industry can shift from a document-based to a content-based approach for creating clinical and regulatory content. This transition will require adopting structured content management tools and common structures, and standardising content. Thus, medical writers must adopt content planning, structuring, and production practices to facilitate content creation and reuse. Ultimately, structured content will enable medical writers to save time by streamlining the writing process, allowing them to focus on tasks that require deliberate thinking, interpreting clinical data, and communicating with stakeholders.

Acknowledgements

The authors would like to thank Creative Services Manager Emma Grier, Senior Graphic Artists Terry Fountain and Sue Becker, and Creative Designer Amanda Moss, for their assistance designing and creating the figures.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employers or EMWA.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. Williams C, et al. The Future of Documents: Structured Content and Integrated Data. February 2, 2022. Available from: https://www.iqvia.com/-/media/iqvia/pdfs/library/white-papers/the-future-of-documents-structured-content-and-integrated-data.pdf. Accessed 10 March 2023.
2. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. [Erratum in: Sci Data. 2019;19;6.(1):6. [1]. Sci Data. 2016;3(1):160018. doi: 10.1038/sdata.2016.18.
3. TransCelerate Biopharma Inc. Clinical Content & Reuse Solutions; c2012–2023 [cited 2023 May 11]. Available from: https://www.transceleratebiopharmainc.com/assets/clinical-content-reuse-solutions/

4. DiCicco R. Clinical trials haven't harnessed the digital revolution – TransCelerate's latest initiative aims to change that. PharmaVoice. October 3, 2022 [cited 10 March 2023]. Available from: https://www.pharmavoice.com/news/clinical-trialsdigital-revolution-transcelerate-Rob-Dicicco/633088/.
5. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) M11 Clinical Electronic Structured Harmonised Protocol (CeSHarP) Draft version. Sept 27, 2022. TransCelerate BioPharma Inc. Digital Data Flow Solutions, [cited 10 Feb 2023]. Available from: https://www.transceleratebiopharmainc.com/initiatives/digital-data-flow/.
6. Swisher V, Preciado RL. The pharma content evolution: Content reuse and automation. Content Rules, Inc., 2022. Available from: https://contentrules.com/the-pharma-content-evolution-content-reuse-and-automation/



Author information

Mati Kargren is a Medical Writer at Parexel International and operational lead for structured content authoring implementation within Medical Writing Services. He is passionate about implementing new technological solutions and bringing awareness to the medical writing role in pharmaceutical product development.

John April is the Senior Director of Adaptive Content Strategy at Eli Lilly and Company. He is a structured content subject matter expert within Global Scientific Communications. He leads Lilly's Automated Content Generation programme aimed at automating document development processes.

Gina Clark is a Director at Pfizer where she is a Medical Writing Therapeutical Area Lead in Inflammation & Immunology. She has a deep understanding of the critical role medical writing plays in pharmaceutical product development and is committed to implementing innovative solutions and driving progress.



Jonathan Mackinnon is an Associate Director Medical Writing Services at Parexel International and project lead for structured content authoring implementation within Medical Writing Services. He also teaches protocol development and trial design at the London School of Hygiene and Tropical Medicine.

Aliza Nathoo is a Senior Director of Content Strategy in Product Development at Roche. In her role, she oversees the development, implementation, and scaling of structured content solutions, technologies and standards in service of various medical writing and documentation groups.

Elizabeth Theron is an Acting Manager Medical Writing Services at Parexel International and a structured content subject matter expert within Medical Writing Services. She also specialises in training and teaching medical writers in medical writing principles and processes.