PharmaSUG 2026 - Paper PO-406

Smarter, Faster, Better: GenAI-Driven Authoring for Data Reviewer's Guides

Christina E. Scienski, Pfizer Inc.,
Christine Rossin, Pfizer Inc.

ABSTRACT

Leverage generative AI (GenAI) to streamline the creation of Data Reviewer's Guide (DRG) for regulatory submissions (e.g., FDA, PMDA, NMPA) by intelligently re-using existing regulatory documents such as statistical analysis plans (SAP), Protocols, annotated case report forms (aCRFs), and internal metadata files.

- Streamline writing process by creating uniform documentation across all therapeutic areas and assets.
- Establish consistency in content and quality.
- Improves authoring efficiency, i.e., resources and time.

Once documents are ingested and intake questions are completed, the user initiates the authoring process. The system generates draft content from source documents, which the user can review and apply. The document can then be saved as a first draft for further review, refinement, and finalization.

INTRODUCTION

The clinical trial industry faces increasing complexity and volume in regulatory submissions, driving the need for innovative documentation solutions. This paper will outline a framework utilizing Generative Artificial Intelligence (GenAI) to automate the creation of Data Reviewers Guides (DRGs), which provide crucial context for CDISC-compliant datasets submitted to regulatory agencies. Gen AI employs Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG) to generate content from reference documents, including study protocols, annotated case report forms, and statistical analysis plans. The paper demonstrates the use of generative artificial intelligence (GenAI), to ingest existing regulatory document templates and internal metadata to populate key sections of Data Reviewer's Guide for regulatory submissions (e.g., FDA, PMDA, NMPA).

OBJECTIVE

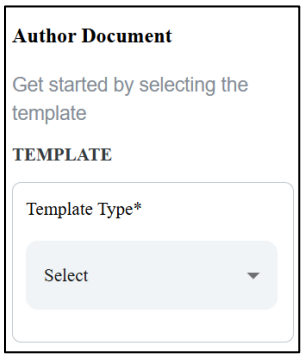
The objective is to create a regulatory document that must comply with the highest standards of accuracy, completeness, and regulatory intent. The development and business team worked together to create technical specifications that would result in uniform, standardized DRGs across all therapeutic areas and decrease processing time and manual effort. .

METHODS

- Leverage GenAI to extract protocol number, title, version, and design from protocol
- Leverage GenAI to extract page numbers and field names for "NOT SUBMITTED" annotations from aCRF
- Generate texts and tables for DRG based on extracted data points using GenAI
- Human-in-the-loop final review / update

PROCESS

The user begins the process of creating the DRG by opening Microsoft Word blank document and selecting the GenAI document generation add-in that is incorporated into the application. Once this is selected, the user will select the Template Type of DRG from a drop-down box. Once DRG is selected, the user will select the document subtype of cSDRG or ADRG.



This selection will prompt the user to provide ingestible documents that will populate the body of the Data Reviewer's Guides. For each of the DRG templates, the user must enter the version date of the template used, Sponsor name, Protocol number, and Protocol title. This information is displayed at the title page of the reviewer's guide.

If cSDRG is selected, there are four documents to upload for ingestion:

- ✓ Protocol (.docx format)
- ✓ Exported Pinnacle 21 cSDRG document (.docx format)
- ✓ Final Annotated CRF (acrf.pdf)
- ✓ Upstream data management metadata file for Inclusion/Exclusion database build (.xlsx format)

As the documents are uploaded, the user must provide metadata prior to ingestion. The metadata includes the version number and the protocol number. The annotated case report form (aCRF) serves to identify "Not Submitted" variables for Section 3.3 Annotated CRFs, of the cSDRG under the header, Explanation of data fields [Not Submitted]. The auto-generated content saves time for the user who manually populates the section. The ingested protocol will auto-generate Section 2 of the cSDRG with the title, protocol versions, and protocol design. The Pinnacle 21 exported version of the cSDRG will provide the protocol specific framework exported from the data validation, including the issue summary. Finally, the inclusion/exclusion criteria excel file will serve to populate Appendix 1: Inclusion/Exclusion Criteria.

If ADRG is selected, there are three documents to upload for ingestion:

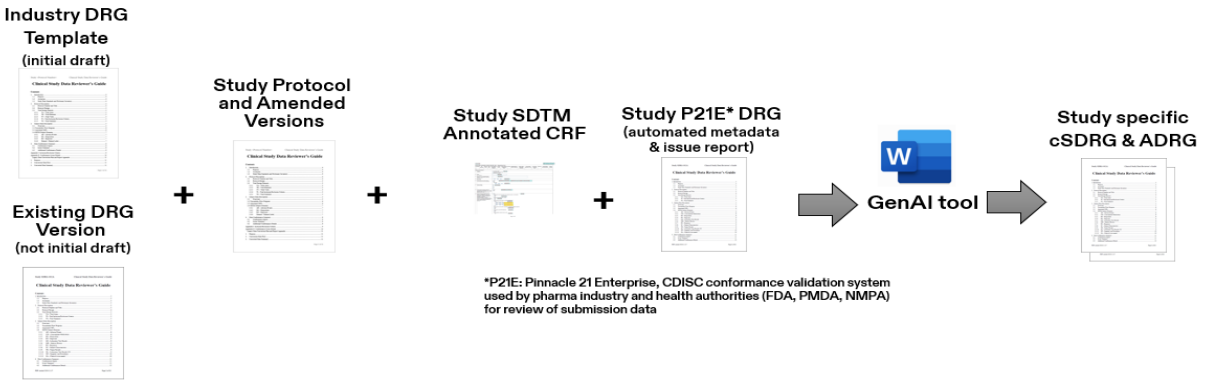
- ✓ Protocol (.docx format)
- ✓ Exported Pinnacle 21 ADRG document (.docx format)
- ✓ Statistical Analysis Plan (SAP) (docx format)

In addition to the above documents for ADRG development, the user must provide a system level path and user id where the submittable programs are stored. This will facilitate the section of the ADRG for programs.

There are two user prompts that will ask the user if there are intermediate datasets or if there are imputation rules. When "Yes" is selected for the prompts, the section of the ADRG will auto-generate content as "Yes" for Section 4.3 Intermediate Datasets, Section 3.5 Imputation/Derivation Methods.

The ingested protocol will auto-generate Section 2 Protocol Description, with the title, protocol versions, and protocol design just as it does with the cSDRG.

PROCESS MAP



RESULTS

When the documentation is ingested, the GenAI add-in will create a document where sections of the DRG are auto-updated or auto-generated. The user must go to the generated content and review and apply or dismiss the content from being added. Auto-generated content will need the user to click on “Apply” button to accept that content. Version history is available when the user creates a Version 2 or more of the document. The user can review the previous content and the date of entry via the version history. Auto-updated content is generated and applied without user intervention.

The user can also generate specified sections of the document by selecting the sections of the document that need to be updated. Auto-updated content will be applied to the sections and the user can dismiss the flag of the auto-updated content.

When one of the source documents is revised, the user does not need to start the documentation process again. The user can simply upload the latest version of the document and apply the amendment information to the DRG while creating an updated version of the DRG.

CONCLUSION

The manual nature of the legacy approach is laborious, error prone, and leads to significant QC findings. By leveraging the latest in machine learning and LLM approaches, DRG is able to reuse content from sources such as the clinical protocol, SAP, annotated CRF metadata and other clinical trial specifications to automate content. Additionally, the capability allows for Q&A functionality to provide the right information at the right time, reducing colleagues’ efforts to find the information they need to complete key tasks.

The impact of GenAI-assisted documentation is immediate and measurable, producing well-structured Data Reviewer’s Guides (DRGs) with the core sections already populated. By eliminating repetitive cut and paste actions and constant editing, the approach ensures consistent and standardized language. It is important to note that GenAI does not replace human expertise; reviewers remain essential to verify accuracy, refine content, and approve the final document. Pfizer’s approach significantly accelerates the authoring process while enhancing consistency and overall quality across Data Reviewer’s Guides.

REFERENCES

Deliverables – WORKING GROUPS – PHUSE Advance Hub

Analysis Data Reviewer's Guide (ADRG) Package - WORKING GROUPS - PHUSE Advance Hub

Clinical Study Data Reviewer's Guide (cSDRG) Package - WORKING GROUPS - PHUSE Advance Hub

FDA Study Data Technical Conformance Guide reference in Guidance for Industry Providing Regulatory Submissions in Electronic Format – Standardized Study Data

Study Data Technical Conformance Guide Technical Specifications Document

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Christina E. Scienski
Christina.scienski@pfizer.com

Christine Rossin
Christine.rossin@pfizer.com