

Enhancing Clinical Study Report Authoring with Generative AI: Lessons Learned on TLF Formatting and Metadata Integration in Merck's iRAP Tool.

Madhusudhan Reddy Papasani, Merck & Co., Inc., Rahway, NJ, USA

Elisabeth Pyle, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

Merck is leveraging Generative AI to streamline clinical study report (CSR) authoring through the Intelligent Research Authoring Portal (iRAP) CSR tool. This innovative tool accelerates the creation of initial CSR drafts by integrating protocol details, Tables, Listings, and Figures (TLFs), and standardized templates, enabling medical writers to focus on data interpretation and clinical relevance. Pilot studies identified critical requirements for TLF formatting and the associated metadata within Data Meta (DM) files to ensure seamless incorporation of immutable TLFs into both the CSR body and appendix. In response, a comprehensive checklist was developed for Statistical Programmers to follow when delivering TLFs and DM files. This presentation will share detailed insights from the pilot phase and illustrate how adherence to the checklist improves compatibility with the iRAP tool.

INTRODUCTION

The pharmaceutical industry is undergoing a significant transformation driven by rapid advancements in Generative AI (GenAI) technologies. To remain competitive and meet the growing demands of an expanding pipeline, Merck is embracing these innovations to enhance operational efficiency and accelerate delivery timelines.

As part of this strategic initiative, Merck introduced the **Intelligent Research Authoring Portal (iRAP) CSR tool**, which leverages GenAI to streamline the creation of initial CSR drafts. By integrating protocol details, standardized templates, and immutable Tables, Listings, and Figures (TLFs), the tool enables medical writers to focus on data interpretation and clinical relevance rather than manual compilation.

Pilot studies revealed key requirements for TLF formatting and associated metadata within **Data Meta (DM) files** to ensure seamless CSR integration. A comprehensive checklist was developed to standardize TLF and DM file delivery, improving compatibility with the iRAP tool.

DATA META FILES FOR TLFS

Data Meta (DM) files accompany the delivery of TLFs in DOCX format and serve as the metadata backbone for CSR authoring. These files capture essential details such as TLF file names, titles, type (table or graph), layout, and linkage information, allowing medical writers to seamlessly incorporate immutable TLFs into the CSR without direct access to source files. By standardizing this metadata, DM files ensure compatibility with GenAI-driven tools like iRAP-CSR, streamline the authoring process, and support efficient integration of TLFs into both the CSR body and appendices.

Snapshot of DM File:

A	B	C	D	E	F	G
TLF File Name	TLF File Title	TLF Type	Section Number	TLF Order in Section	Layout	Microsoft Word Link Text
test0graph6.docx	Change From Baseline in CRQ Score (Averag Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
test0graph2.docx	Difference in HbA 1c (%) Mean Change frc Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
test1.docx	Product-Limit Survival Estimates; with Stand Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
test5.docx	Product-Limit Survival Estimates; with Stand Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
Items0on0page2.docx	Rainfall Plot; MK-0000 10 mg (N=21600) vs. Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
page0orientation0L.docx	Rainfall Plot; MK-0000 10 mg (N=21600) vs. Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
test0graph5.docx	Time to Onset of First Grade 3, 4, or 5 Adver Graph				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
gr0ich0efficacy0listing0002-a.docx	16.2.6a Listing of Subject Individual Efficacy Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
gr0an0summary0listing0001.docx	Allocation List by Investigator and Treatmer Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
gr0an0summary0listing00010range.docx	Allocation Number Range by Investigator; D Table				Landscape	
test0symbols.docx	Clinical Trial Exposure to MK-974 by Dose; E Table				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
gr0rmp0drug0exp0by0race0001.docx	Clinical Trial Exposure to MK-974 by Race Table				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
rmprace2.docx	Clinical Trial Exposure to MK-974 by Race Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
test0tblandscape.docx	Clinical Trial Exposure to MK-974 by Race; T Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
test0tblportrait.docx	Clinical Trial Exposure to MK-974 by Race; T Table				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
gr0disposition0001.docx	Disposition of Subjects; TC#1: DSTERM, subj Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
ECG.docx	MK-1602-002 Summary for Safety ECG Data Table				Portrait	INCLUDETEXT"\\\\bardsar-test\\standar
gr0disposition0001exc.docx	Reasons Subjects NOT Enrolled; TC#1: DSTE Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
gr0disposition20001.docx	Subject Disposition Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
medhist12.docx	Subject Medical History Conditions (Incident Table				Landscape	
gr0ae0inj0site0001.docx	Subjects With title text 1 Injection Site Adve Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
gr0ae0inj0site0001maxint.docx	Subjects With title text 1 Injection Site Adve Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar
gr0ae0inj0site0001maxsize.docx	Subjects With title text 1 Injection Site Eryth Table				Landscape	INCLUDETEXT"\\\\bardsar-test\\standar

GENERATIVE AI AND TLF STANDARDIZATION

CSRs are authored using the **iRAP CSR tool**, which employs GenAI to rapidly generate first drafts based on protocol details, TLFs, and standardized CSR templates. This initial draft is refined by medical writers and study teams, allowing them to focus on interpretation and clinical relevance.

To ensure successful integration of TLFs into the CSR body and appendix, a **TLF Standardization Checklist** was developed based on lessons learned from initial iRAP CSR tool pilot studies. The pilot phase revealed several barriers that hindered reliable and efficient automated CSR integration. Key issues included:

- **Inconsistent TLF formatting:** Variability in title structure, whitespace, paragraph handling, and document layout prevented the tool from reliably parsing and positioning tables.
- **Multiple DM files:** Delivery of more than one DM file across sequential handoffs caused parsing inconsistencies, as the iRAP CSR tool requires a single, cumulative DM file for accurate metadata mapping and TLF placement.
- **Non-standard image formats:** Use of EMF or WMF formats for figures led to rendering failures within the CSR body, highlighting the need for consistent use of PNG or JPG formats.
- **Improper folder organization:** Placing listings, PRO tables, and CSR-body TLFs in multiple directories reduced automation accuracy and complicated metadata linkage.
- **Sensitivity labels on files:** Sensitivity labels applied to TLF DOCX files (e.g., Public, Proprietary, Confidential) triggered import and parsing issues, requiring removal prior to delivery.

These findings guided the development of the standardized TLF and DM checklist.

TLF STANDARDIZATION CHECKLIST FOR IRAP CSR INTEGRATION

To maintain compliance with immutable table standards and ensure seamless integration into the iRAP CSR tool, the following requirements must be met:

1. DELIVERY FOLDER ORGANIZATION

- Store all finalized tables and figures—delivered as non-editable DOCX files for CSR use—in a single folder within the clinical production area (immutable-tbls).
- Ensure that every new TLF delivered over the course of the study is added to the same folder, including patient-reported outcome (PRO) tables.
- Only tables and figures designated for the CSR body (Sections 10–14) should appear in the immutable table folder in **DOCX format**.
- Listings designated for the CSR appendix Section 16 should be placed in the csr-appendices folder in **PDF format**. The iRAP CSR tool does not currently accommodate listings in Sections 10–14.

2. DM FILE FORMAT

- Create one comprehensive DM file named *dm.xlsx* (case-sensitive) that includes all TLFs for the CSR.
- Ensure file names in the DM file match those in the Delivery folder.
- Exclude listings designated for Section 16 and any temporary files (prefix ~\$).
- Exclude files that contain multiple tables in a single document.

3. CONTENT PROTECTION

- Apply content protection to all tables in DOCX files using the immutable table process.

4. SENSITIVITY LABEL REMOVAL

- Ensure TLF DOCX files do not contain sensitivity labels (e.g., Public, Proprietary, Confidential). Users must avoid applying labels when closing files.

5. IMAGE FORMAT

- Use **PNG or JPG** formats for figures in DOCX files. WMF and EMF formats are not compatible.
- Standard graphic macros default to PNG; if figures are created outside these macros, specify PNG or JPG explicitly.

6. TLF FORMATTING CHECKS

- Ensure TLF titles in DOCX files:
 - Do not have leading/trailing new lines or excessive white spaces.
 - Are contained within a single paragraph (including subtitles).
 - Citations should also be in a single paragraph.
- One leading whitespace in the title is acceptable.
- Avoid additional spaces or line breaks that could cause formatting issues.

CONCLUSION

Generative AI is transforming CSR authoring by automating initial draft creation and enabling teams to focus on interpretation and clinical insights. Merck's iRAP CSR tool, combined with standardized TLFs and Data Meta (DM) files, ensures seamless integration and efficiency. The TLF Standardization Checklist provides clear guidelines for formatting and metadata, reducing errors and supporting scalability. Adhering to these standards is essential for leveraging GenAI effectively and meeting the demands of an expanding portfolio.

ACKNOWLEDGMENTS:

The authors would like to thank BARDS Statistical Programming Leadership and Management at Merck & Co., Inc., Rahway, NJ, USA: Amy Gillespie, Mary Varughese.

CONTACT INFORMATION:

Your comments and questions are valued and encouraged. You may contact the corresponding author at:

Madhusudhan Reddy Papasani, Ph.D.
Principal Scientist
Biostatistics and Research Decision Sciences
Oncology Statistical Programming
Merck & Co., Inc., Rahway, NJ, USA
madhusudhan.papasani@merck.com