



CSS 2024

The Computational Science Symposium

3-5 June
Silver Spring, MD



phuse.global



Plenary Overview

Plenary Session with Updates from Health Authorities

- **Data Submission in Europe – An EMA Update**
Eftychia Eirini Psarelli, *EMA* & Peter Van Meer, *Medicines Evaluation Board*
- **A Risk Based Framework for AI Use in Drug Development**
Tala Fakhouri, *FDA*
- **CSS Poster Awards**
Jean-Marc Ferran, *Qualiance*, *Data Transparency Working Group Lead & CSS Poster Chair*



The Computational
Science Symposium

CSS 2024

3–5 June
Silver Spring, MD



Poster Award Winners

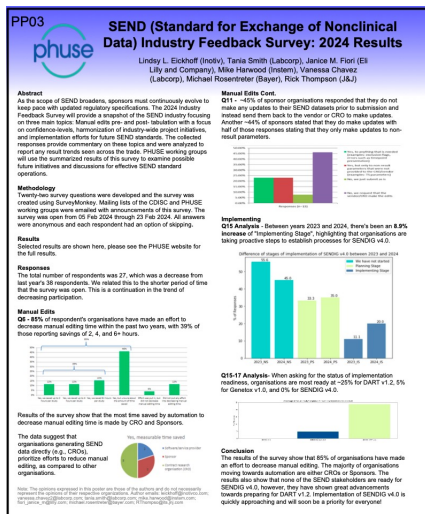




Collaboration

PP03 –SEND (Standard for Exchange of Nonclinical Data) Industry Feedback Survey: 2024 Results

Vanessa Chavez, Labcorp
Lindsay Eickhoff, Inotiv





Industry

PP18 – Bridging the Gap – Enhancing TFL Generation with CDISC ARS Model and Community TFL Library

Bhavin Busa, Clymb Clinical



PP18

CLYMB CLINICAL

Bridging the Gap - Enhancing Analysis Results Generation with CDISC ARS Model and Community TFL Designer

Author: Bhavin Busa, Principal & Co-Founder, Clymb Clinical, bhavin@clymbclinical.com

Current State: The construction of TFL mock-up shells and analysis results metadata in our industry faces challenges stemming from variability, lack of standards, and manual processes.

Analysis Results Generation: Top 5 pain points

- Complexity of generating TFLs from source code
- Regulator writes the code to generate analysis, downstream (various data sources or reusing the code)
- Too much variability across studies, design assets, and templates
- Static results with no or limited linking facility (e.g. to Protocol, SAP, ADaM data)
- There is no industry standards for analysis results

Survey results from 253 participants conducted during the TFL Designer Design Thinking Workshop, 13th Sep 2022, highlight the pressing need for standardized, machine-readable TFL display shells and analysis results metadata to bolster efficiency and collaboration.

Survey Results from Design Thinking Workshop

Challenge	Frequency	Impact
Complexity of generating TFLs from source code	High	Time-consuming and error-prone
Regulator writes the code to generate analysis, downstream (various data sources or reusing the code)	Medium	Lack of transparency and reproducibility
Too much variability across studies, design assets, and templates	High	Inconsistent and hard to maintain
Static results with no or limited linking facility (e.g. to Protocol, SAP, ADaM data)	Medium	Limited flexibility and scalability
There is no industry standards for analysis results	High	Lack of interoperability and collaboration

CDISC Analysis Results Standard (ARS)¹: The CDISC ARS Model and accompanying user guide have been developed to provide a logical model that fully described analysis results and associated metadata to support automation, consistency, reuse, and traceability of results data.

TFL Designer (Community, free version)²: This platform automates the creation of TFL shells and provides machine-readable metadata, which can then be seamlessly ingested for downstream automation in the programming of the TFLs.

Enhancing Analysis Results Generation:

To address the current limitations of analysis results and associated metadata reporting, we are proposing a new workflow that shifts the focus from retrospective reporting to prospective planning.

ARS Model is complete! How do I operationalize it and generate analysis results metadata prospectively?

ARS Model Representation using CMAP

Design, Specify, Build and Execute: In our vision for the future state, we anticipate the availability of open-source or community tools, such as the TFL Designer Community, in the industry.

Analysis Results Workflow w/ TFL Designer

Development Vision & Conclusions:

Enhancing CDISC ARS, along with a TFL design tool and workflows (including use of Analysis Results Data), will bridge the existing gap in analysis results generation. We are confident that this will foster an open-source community in line with the CDISC Open-Source Alliance (OSA) & PHUSEpharmaverse vision, facilitating the creation of shareable TFL templates, metadata and codes (in both SAS and R languages) by the community.

TFL Designer (Community, free version)²: This platform automates the creation of TFL shells and provides machine-readable metadata, which can then be seamlessly ingested for downstream automation in the programming of the TFLs.

¹ TFL Designer Community <https://tfl-designer.org>

² ARS Scheme <https://cdisc.org/standards/ars/scheme>



FDA

PP23 – A Framework Enabling LLMs into Regulatory Environment for Transparency and trustworthiness and its Application to Drug Labelling Document

Shraddha Thakkar, FDA



PP23 A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document

Leihong Wu*, Joshua Xu*, Shraddha Thakkar*, Magnus Gray*, Yanyan Qu*, Dongying Li*, Weida Tong*

*Office of Computational Sciences (OCS), Office of Translational Sciences (OTS), Center for Drug Evaluation and Research (CDER), US FDA

* Division of Bioinformatics and Biostatistics (SDB), National Center for Toxicological Research (NCTR), US FDA.

Introduction

Regulatory agencies consistently deal with extensive document reviews, ranging from product submissions to both internal and external communication. Large Language Models (LLMs) like ChatGPT can be invaluable tools for these tasks. However, several potential drawbacks exist: model biases, lack of transparency, and proprietary and private information. This document will be the need for customized solutions for specific review needs and transparency and traceability of the model's output, under the need for a local and customized solution imperative.

To tackle these challenges, we formulated a framework named AiFDALabel for FDA drug labeling documents that is a crucial resource in the FDA drug review process. AiFDALabel operates within a secure IT environment and comprises two key modules: a semantic search (Model 1) and a GAN-based generation module (Model 2). The Model 1 is built on word embeddings to enable comprehensive semantic queries within labeling documents. The Model 2 utilizes a trained LLM to generate responses based on information from Model 1. In the output, our framework enabled model LLMs to perform comparably to ChatGPT by providing a comprehensive response suitable for regulatory application.

To conclude, through AiFDALabel, we have showcased a pathway that harnesses LLMs to support agency operations within a secure environment, offering tailored services for the needs of regulatory agencies.

Framework to facilitate LLMs application in regulatory setting

To address these challenges, we developed a framework named AiFDALabel. The framework is designed to enhance the customization and transparency of LLMs. The proposed framework was implemented in a secure, secure IT environment, to meet the specialized security needs of many regulatory documents.



Case Study

	INTERNAL	EXTERNAL	CONFIDENTIAL	REGULATORY
ADVERSE	Yes	Yes	Yes	Yes
ADVERSE	Yes	Yes	Yes	Yes
ADVERSE	Yes	Yes	Yes	Yes
ADVERSE	Yes	Yes	Yes	Yes
ADVERSE	Yes	Yes	Yes	Yes
Average	15.00%	15.00%	15.00%	15.00%
Standard	0.00%	0.00%	0.00%	0.00%

Table 2: Here we asked AiFDALabel the following question: "Does [Drug Name] contain any cardiovascular adverse events or risks?" where the [Drug Name] is from the list of 100 drugs. AiFDALabel returned answers from the top 3 responses, and the drug confidence and the number of words in the answer was "Yes". The averaged performance is summarized in Table 2. We also compared the AiFDALabel response to the ChatGPT result. As shown in Table 2, AiFDALabel outperformed ChatGPT by overall accuracy as 77.00% vs 40%, where ChatGPT was better at generating and AiFDALabel was better at handling.

Summarizing multiple answers into one response using an additional summary LLM



Improving search relevance by module optimization

Query	Search Type	Relevance	Latency
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15
ADVERSE	Internal	0.15	0.15

Table 3: This report presents the results of the search relevance test based on whether the query could be generated from the current labeling documents. We developed the framework to improve the search relevance by using the search on three different types of entities: generic name, product name, and application number. For example, the query used in the test was "Is the adverse reaction of [Drug Name]?" or "Is the adverse reaction of [Drug Name]?" The relevance of the query was measured by the percentage of the query that was returned by the AiFDALabel database query of the designated entities. The result was compared to the result of the search of the database by the search engine. The overall relevance was measured by the ratio of the result of the search engine to the result of the database query. The result of the search engine was used to calculate the relevance of the search engine.

Objectives and Challenges

- Objective: Develop a framework to enable LLMs in regulatory applications
- Challenges:
 - Regulatory agencies need functionalities to support internal or external communication
 - Customized functionalities specific review requirements must be developed for the regulatory agencies
 - Transparency and accountability
 - Regulatory agencies need transparent chain-of-thought and traceable evidence from the relevant data sources
 - Customize data processing capabilities to support the regulatory agencies
 - Data security
 - Regulatory agencies operating and control be shared with public LLMs
 - Some test data and regulatory research are not publicly accessible due to confidentiality or privacy considerations

Proposed Solutions

- A proof-of-concept study aimed at developing a framework to address these needs
- Developed a secure framework and demonstration on FDA Drug Labeling documents
- A GAN-based generation module to generate responses to support review queries, and search results linked back to the original documents
- A GAN-based generation module to generate responses to support review queries, and search results linked back to the original documents
- Compare performance of the framework against ChatGPT

Conclusion

In conclusion, while AiFDALabel framework presents substantial advancements in the analysis of drug labeling documents, it continuously highlights the need for further research and development. The requirement of maintaining the confidentiality of sensitive data, optimizing system performance, and understanding the extent of review coverage are ongoing challenges. Integrating AI within the pharmaceutical domain. Future work should continue to focus on these areas, with the ultimate goal of optimizing the AiFDALabel system to best support the vital work of FDA scientists and reviewers.

Acknowledgments

We would like to thank the CDER Knowledge Base for feedback for our use within the FDA. This research project was funded and supported by NCTR, FDA (project ID: 2023FDA-00000001).



Thank you to all the presenters!



Poster Judges:

- Anthony Chow (CDISC)
- Dale LeSueur (Regeneron)
- Jesse Anderson (FDA)
- Setondji Dega (FDA)

PHUSE Project Coordinators

- Lauren White
- Alexandra Pearce

EU 2024

10-13 November
Strasbourg Convention Centre

Iconic
Strasbourg
Building



Strasbourg Cathedral



US 2025

16–19 March

Renaissance Orlando at SeaWorld Florida

phuse.global



Thank You

We look forward to seeing
you at the CSS next year!