PharmaSUG 2024 - Paper MM-447

Automating third party data transfer through digitized Electronic DTA Management

Vandita Tripathi, and Manas Saha, Tata Consultancy Services

ABSTRACT

CDM is a pivotal process in clinical trial, and not all data is directly captured into case report forms (CRFs), a large portion is also collected from external sources like third-party vendors, labs, wearables, devices, and external systems. This type of data is called “non-CRF data”. Maintaining Data integrity and quality of the non-CRF data has a critical influence on clinical data management process and study success.

To ingest non-CRF data from external third parties, Data Transfer Agreements (DTA) are signed between sponsor and vendor organizations. DTA defines the structure, frequency / timelines, and data definitions which enables transfer of non-CRF data from vendor to the study database. Often the DTA authoring, and review process takes months which is a significant challenge in the industry. Non CRF data received from vendors needs to be reconciled with CRF data which often takes lot of time due to associated quality and standard related issues with the non-CRF data.

An online, collaborative, highly automated, and real time DTA authoring solution along with integrated multi-step DTA review and approval system can significantly benefit the non-CRF data management landscape. The system can utilize predefined Master DTA templates, Data models and Code lists to automatically draft the DTA document. The auto generated DTA document can be viewed and edited in real time by multiple editors. The reviewers can quickly track the changes and accept / reject them with comments. Once the edits are reviewed and finalized, an approver can approve the DTA along with eSignature.

This can cut down the timelines from months to days by automating DTA drafting and utilizing real time collaboration. This could potentially turn out to be a paradigm shift for third party data management.

INTRODUCTION

At present, most of the clinical trial management systems have connectors or interfaces that enables them to seamlessly receive non-CRF data from external labs and vendors without manual intervention. However, most of these data is non-standardized, unstructured and to maintain data quality and integrity, both the vendor and the study sponsor needs to sign an agreement on the data transfer specifications.

This data transfer agreement or DTA is a pivotal artifact for the success of the trial. In the industry, the process for drafting and approval of this DTA is mostly manual, and often it takes long time to finalize the DTA. The purpose of this paper is to present a new approach for managing DTA and third-party data. This new approach utilizes template driven, AI powered automatic e-DTA document creation which then can be collaboratively edited and reviewed by multiple editors and reviewers in real time. Once the document is reviewed it can be digitally signed-off and in digital eDTA, rules pertaining to data quality conformance checks and reconciliation can be defined such that as soon as vendor data is received it gets automatically reconciled with eCRF data after meeting data quality and conformance requirement defined as per the specifications. Any failure to this process will send auto notification to all stakeholders for timely action and remediation.

In the following sections, this paper will highlight the current challenges the industry is facing regarding DTA management, as well as how the new solution can help to overcome the current challenges. This paper will also highlight the approach, architecture, and technology behind the new solution, and how the emerging technologies like artificial intelligence and generative AI has the potential to transform DTA and third-party data management in clinical trial industry.

CURRENT INDUSTRY APPROACH AND CHALLANGES

Non CRF data reconciliation involves identifying data that is non-conformant, which may include (but not limited to) missing data, out of the range data, duplicate data, or unrecognizable data. The data may also be irregular, or in non-standard format. Inaccuracies in the Non CRF data can greatly increase the risks in patient safety and study success. Which is why a Data transfer agreement (DTA) is vital to overcome these challenges.

However, in the current clinical trial industry, there is no industry wide DTA standard. The industry mostly uses custom DTA templates to draft the data transfer requirement specifications. And this DTA templates widely vary across sponsor to vendors.

The current process for drafting the DTA documents in the industry is mostly manual. Human editors with business knowledge are responsible for drafting the DTA, specifying each intricate details on data formats, data sets, data model fields, code lists, variables, validation rules and transfer frequencies. Manually drafting these details requires a lot of time and effort.

Once the document is drafted, the reviewers need to review the documents and send back the same to the editors in case any discrepancies is found. Once the cycle completes, the approvers need to digitally sign the DTA document to that the same can be put into effect.

The whole manual process of drafting the DTA document and the review and approval process usually takes several weeks to months to complete.

Having a lack of standardized DTA format across the industry, or even across the sponsor means for every new study, the DTA is drafted again with little or no reusability, resulting in a loss of manual effort, person dependency and time.

PROPOSED APPROACH

In the proposed approach, the new solution will help to overcome the existing challenges using the following key principles.

1. Define and Use standard Master e-DTA templates as the base for different source types, with placeholders for Data models, code list and other dynamic contents. For e.g.: eDTA standard for labs, biomarkers, ePRO etc
2. Using the master eDTA template, automatically generate the e-DTA document by populating the dynamic elements from metadata library or study protocol document. Leverage Machine learning to automatically select the most appropriate template and utilize AI methods like Named Entity recognition to automatically extract the dynamic variables from metadata library or protocol document.
3. Provide a real time collaborative editing system with change tracking, where the editors and reviewers can simultaneously make edits to the machine generated e-DTA documents to make corrections whereas necessary.
4. Provide a review and approval workflow with digital signature feature.
5. Once signed and approved, provide a rule editor for setting up real time data quality conformance and reconciliation rules with eCRF data

These key principles ensure that the turnaround time for e-DTA creation and third-party data management is quick and requires a significantly low manual effort compared to the existing process.

The following sections elaborates the key principles in detail.

THE PROPOSED PROCESS

In the proposed approach, an AI enabled e-DTA Document generation engine will automatically generate the DTA document. The engine will use the following approach:

1. A list of preset Master DTA templates to be maintained by Sponsor The e-DTA generation engine will automatically identify the most appropriate DTA template for the current study based on historical training data for similar past studies.
2. The template will have placeholders to fill in the dynamic variables like Code list, data model, data sets, transfer frequency and other custom objects.
3. The document generator will have access to either the Study Protocol or a Metadata Repository containing the study metadata. The document generator will use AI based Natural language processing techniques like Named entity recognition, Text Summarization, or Topic Modeling to extract the required dynamic information from the protocol document or from the metadata repository.
4. Once the required information is extracted, the document generator will fill in the placeholders with necessary information to prepare the auto-generated initial draft of the e-DTA document.

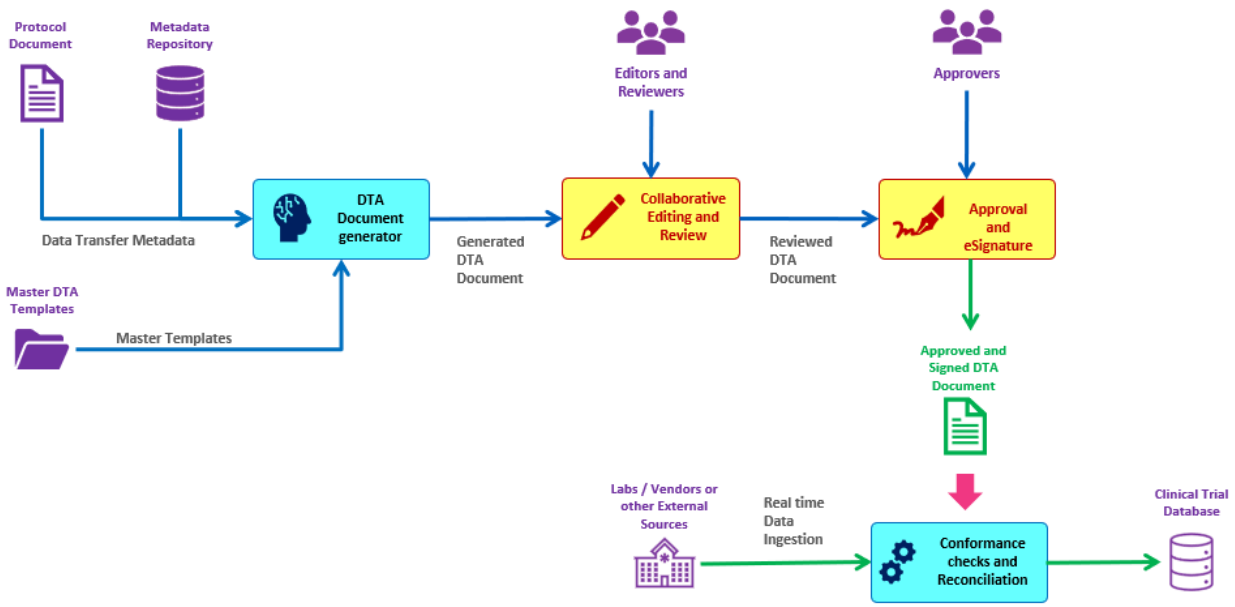

Figure 1. High level process flow

5. Once generated, the initial draft will be made available to the editors and reviewers who will use the real time collaborative editor to simultaneously make changes, track changes, review and add comment to the e-DTA document.
6. Once the editing and the review is done, the approvers can add digital signature to the e-DTA document and publish the same for implementation.

TECHNICAL ARCHITECTURE

The proposed architecture has the following ley components.

1. **Master DTA template library**: this is a collection of e-DTA templates. These will be prepared beforehand based on historical studies the sponsor has conducted in the past.
2. **Historical training data**: a machine learning training dataset based on which the template selector will be able to pick up the appropriate template for the current study.
3. **Protocol document**: This is source document from where an AI-based service will extract metadata to populate the e-DTA document. the metadata may include (but not limited to) code list, data set, data model, schedules, and field names.
4. **Metadata extractor**: this AI powered module will use natural language processing and information extraction methods like Named entity recognition, text summarization and topic modeling to extract required metadata from the protocol document. this metadata will be used to fill-in the placeholders in the chosen template to prepare the initial draft of the e-DTA document.
5. **Collaborative editor**: This is a WYSIWYG editor using which the editors and reviewers will be edit, comment and track changes in real time. Libraries like CK editor can be used to implement the same.
6. **PDF signing**: once the document is finalized, a PDF signing module will digitally sign and publish the final e-DTA document. Digital signature can be implemented using personal PCF digital signature certificate files, or by leveraging commercial PDF signing services like Adobe acrobat.

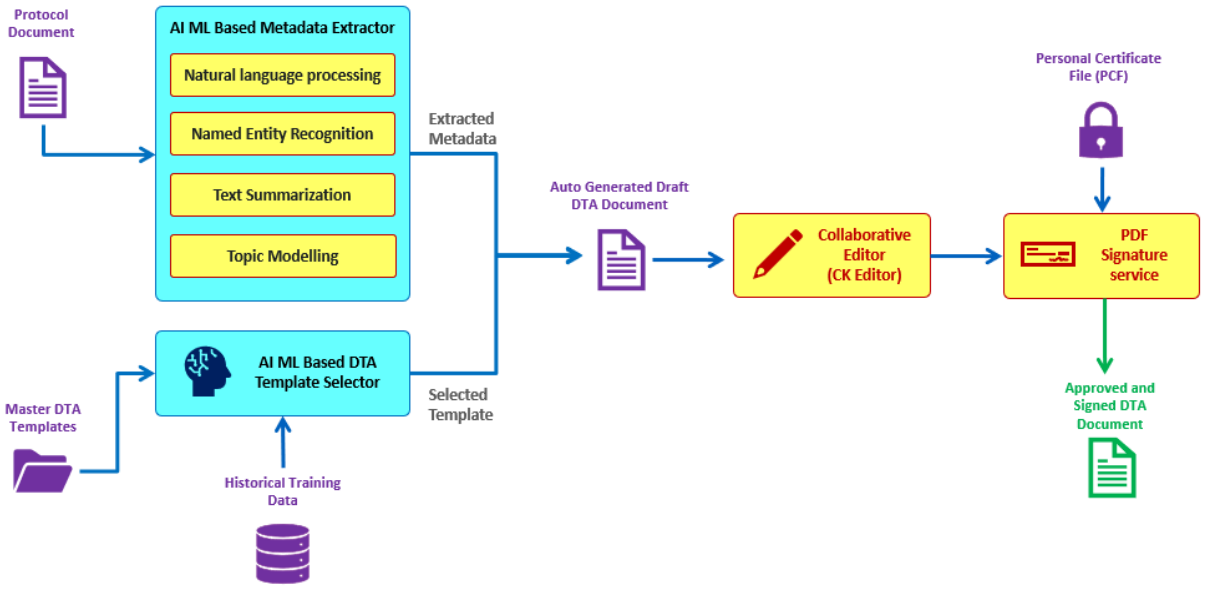

Figure 2. High level technical architecture for eDTA generation

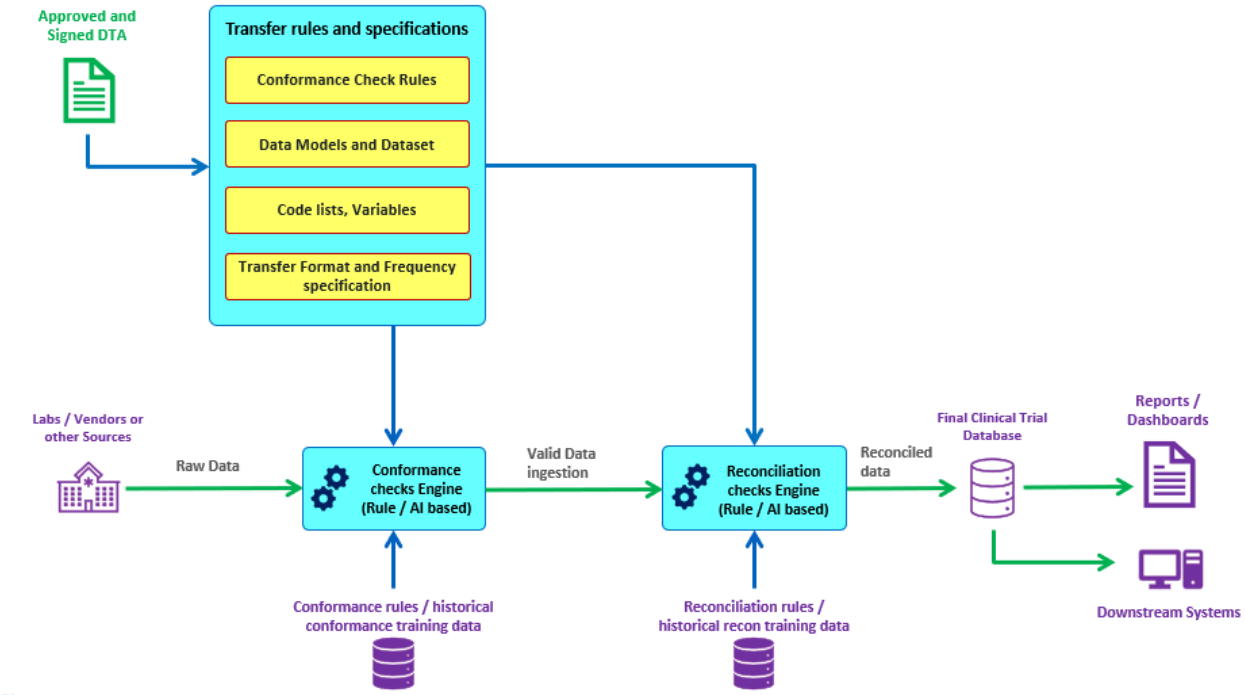

Figure 3. High level technical architecture for data ingestion with conformance check and reconciliation

CONCLUSION

The proposed e-DTA solution that can leverage AI-ML to extract metadata from protocol document and use standardized templates to automatically generate DTA documents, and provides a robust collaborative authoring, review, approval, and digital signature workflow can solve most of the problems the clinical trial industry is facing managing data transfer agreements.

This solution does not need the authors to spend weeks or months to meticulously draft each data transfer specification. On an average a custom DTA document authoring, review and approval takes weeks to months in existing clinical trial landscape. This proposed solution can cut down that turnaround time to less than a week by drastically reducing the time and manual effort needed to generate custom DTA documents which has the potential to transform the clinical trial landscape.

REFERENCES

Applied Clinical Trial. “Establishing Metrics and Standardization for Non-CRF Data in EDC.” July 20, 2021. https://www.appliedclinicaltrialsonline.com/view/establishing-metrics-and-standardization-for-non-crf-data-in-edc

CONTACT INFORMATION

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

Vandita Tripathi
Tata Consultancy Services
vandita.tripathi@tcs.com
https://www.linkedin.com/in/vandita-tripathi-3154a358/

Manas Saha
Tata Consultancy Services
saha.manas1@tcs.com
https://www.linkedin.com/in/sahamanas/