Paper SI11

Standards Compliance - An automated way to measure the Interoperability of Clinical studies

Rammprasad Ganapathy, Genentech, South San Francisco, CA, USA

ABSTRACT

The use of standards and consistent terminologies enables data interoperability for faster and easier integration into analyzable datasets for generating insights. Adherence to the sponsor's data standards & terminologies is critical, not only to ensure that data are FAIR but also to ensure data will meet global regulatory expectations. Hence the benefits of adherence to data standards are significant and multi-dimensional. The more the clinical studies are aligned to the sponsor's standards, the more the downstream processes (SDTM and ADaM) can be automated. Sponsors & study teams are looking for ways to implement the 'Interoperability' component of FAIR principles in clinical studies. Metadata Interoperability tool is an R-based solution that compares a study's metadata to MDR via an API and provides useful information such as study Interoperability metrics, highlight Non-standard CRFs and data collection fields, discrepancies in standards adoption & automation of SDTM mapping and Controlled terminology for standards used in the study.

INTRODUCTION

Our vast collection of diverse clinical research data establishes a critical link between our patients and our ability to develop the medicines they need. Together, these data provide opportunities for generating scientific insights far beyond what anyone group of researchers can generate alone, but only if data are Findable, Accessible, Interoperable, and Reusable (FAIR) and shared across our global organization. Data's value extends long beyond the lifespan of a single study. However, it is often treated as more of a means to answer a specific question rather than as a strategic asset.

Adherence to data standards & terminologies is critical for the primary use of data, that is, submission of data to Health Authorities. It is also important for the secondary use of data, that is, generating insights from data to answer key scientific questions. The use of data standards & terminologies in the clinical studies contribute to the I of FAIR enabling interoperability and allowing for a consistent approach to the design of clinical studies and also in combining of data from multiple studies. It also supports the creation of standard processes, programs, macros and systems to streamline workflows in data collection, tabulation, and analysis with the aim of getting medicines to patients faster.

One of the key underlying challenges for the study teams is the adherence to sponsor standards and building Interoperable Clinical studies. Inconsistent Data Collection standards leads to inconsistent study design and mapping to SDTM and usage of inconsistent Controlled terminologies. These inconsistences lead to less Interoperability of the studies that makes it challenging to pool data across studies, molecules & Indication. Thus, limiting the further reuse of the clinical data. This also impacts the automation of downstream activities like Data Tabulation and Data Analysis. The study teams have no automated way to answer the below questions while designing a clinical study.

- How do I quantify the Interoperability of my study?
- Have I used standards & terminologies in all applicable cases?
- Can I quickly identify the non-standards in my study?
- Are there any deviations in adopting standards?
- Is there a way to automate the generation of eSubmission deliverables for SDTM?

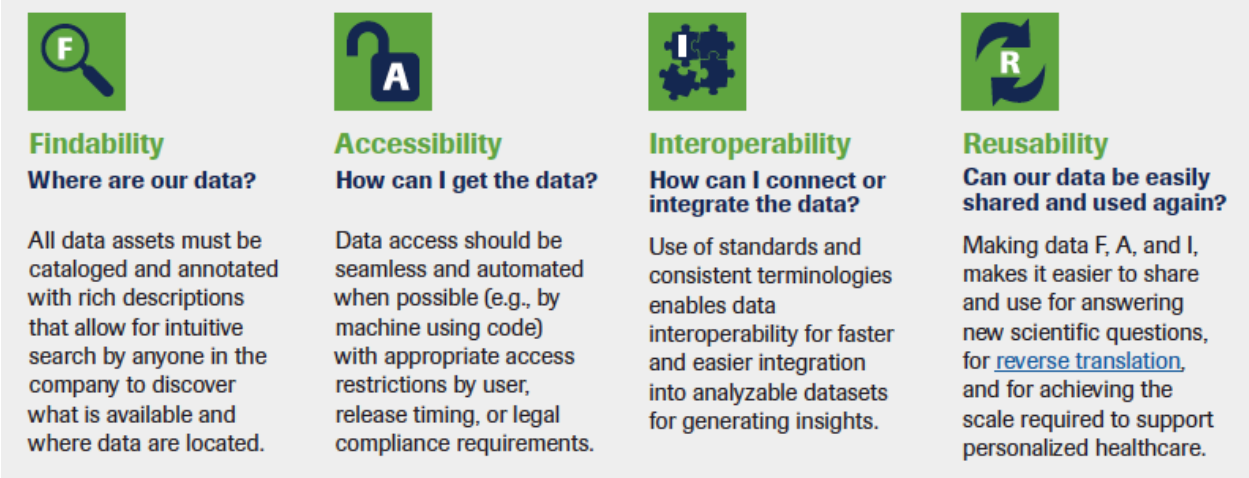
This paper explains a solution we have implemented to help the study teams answer these questions and measure the standards compliance in an automated way driven by the Metadata, sematic design and latest technologies. Metadata INTeroperabilty (MINT) R-Shiny App is an R-based solution that compares a study's metadata to the metadata available in MDR via an Application Programming Interface (API) and provides standards compliance score and other useful information to study teams.

FAIR OVERVIEW

The term FAIR originated during a 2014 workshop in Leiden, Netherlands with attendees from academia, industry, funding agencies, and scholarly publishers. The group shared an interest in overcoming data discovery and reuse obstacles and sought to "design and jointly endorse a concise and measurable set of principles," now known as the FAIR Principles. These principles describe a well-organized state of data (FAIR data) that enable it to be readily and widely used for generating scientific insights and driving more informed clinical decisions.

This international multi-stakeholder workshop led to the first formal publication of the FAIR principles[1] in 2016, with "the intent that these [Principles] may act as a guideline for those wishing to enhance the reusability of their data." The FAIR Guiding Principles underscore the importance of taking a proactive approach to proper data value management during the lifecycle for easier and faster reuse.

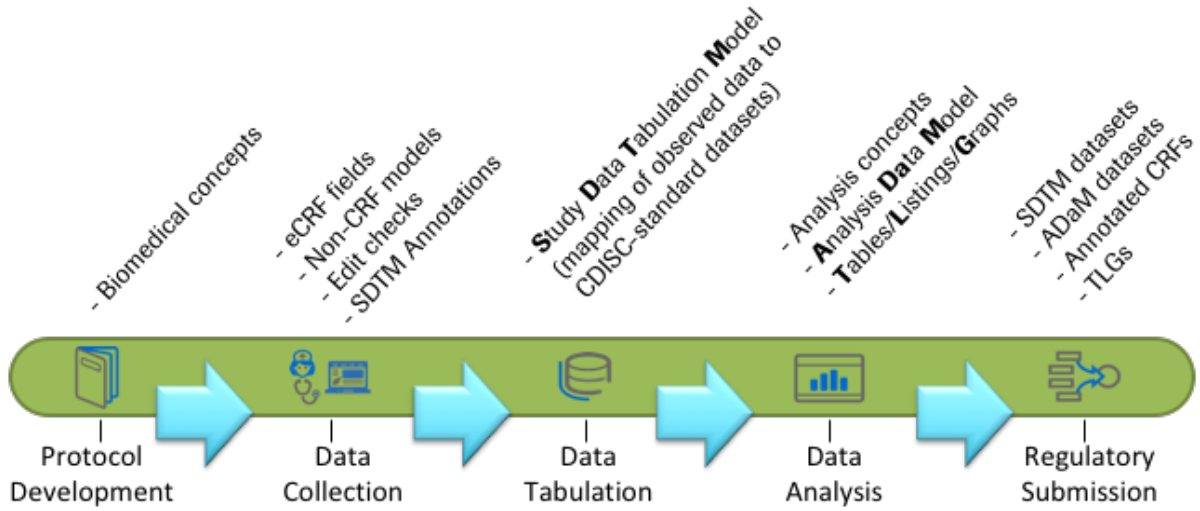
When fully adopted, the FAIR principles will significantly lower barriers which currently obstruct the rapid transformation of scientific data across multiple disciplines into meaningful and trustworthy insights. Since the first workshop, several collaborations (e.g. FORCE11[2] and the Pistoia Alliance[3]) have continued to play a significant role in the evolution of the FAIR Guiding Principles and their adoption.



Data Standards, adherence to standards and Terminologies is key to the 'Interoperability' aspect of the FAIR principles.

ROCHE GLOBAL DATA STANDARDS

Global data standards describe how our clinical trial data should be collected, tabulated, analyzed, and submitted to regulatory authorities. They include all the layers required for this description, from biomedical concept definitions to controlled terminology and metadata required to capture, tabulate, analyze and report data. They are fully aligned with CDISC submission data standards (SDTM and ADaM), extended to ensure consistency across our clinical trials. Our global data standards are housed in a metadata repository (MDR) which is GDSR. Our data standards span across multiple Therapeutic areas of interest and is governed across Roche. Usage of data standards and terminologies in clinical studies enable Interoperable study design and lead to various automation capabilities from data collection to data analysis and reporting.



FAIR MDR

From the Roche FAIR Data playbook: “FAIR data are best stored in a system that provides unique data asset IDs, tracking and management of current and historical versions, and facilities for structured annotation with metadata. Suitable systems must support an organized set of annotation tags and metadata that describe and characterize data and permit easy inspection and querying of the metadata by individuals and by other computing systems.”

The GDSR is Roche and Genentech’s MDR and is the single source of truth for clinical data standards, storing and making them available to users globally in the most business useful format. The data standards are stored in semantic graphs and the technology behind the GDSR is based on RDF, ontologies, and linked data principles aligned with Semantic Web (WC3) standards. Having common vocabularies for (meta) data assets ensures common understanding and facilitates integration and sharing for machines. Data standards described in the GDSR link to external datasets such as CDISC metadata and ontologies, enabling the exploration of connected vocabularies.

REST API

The GDSR makes content available through a public REST API. Users can programmatically access the data standards, catalog, publication information, and the archives, making automation of previously manual tasks possible.

PUBLICATION VERSIONING

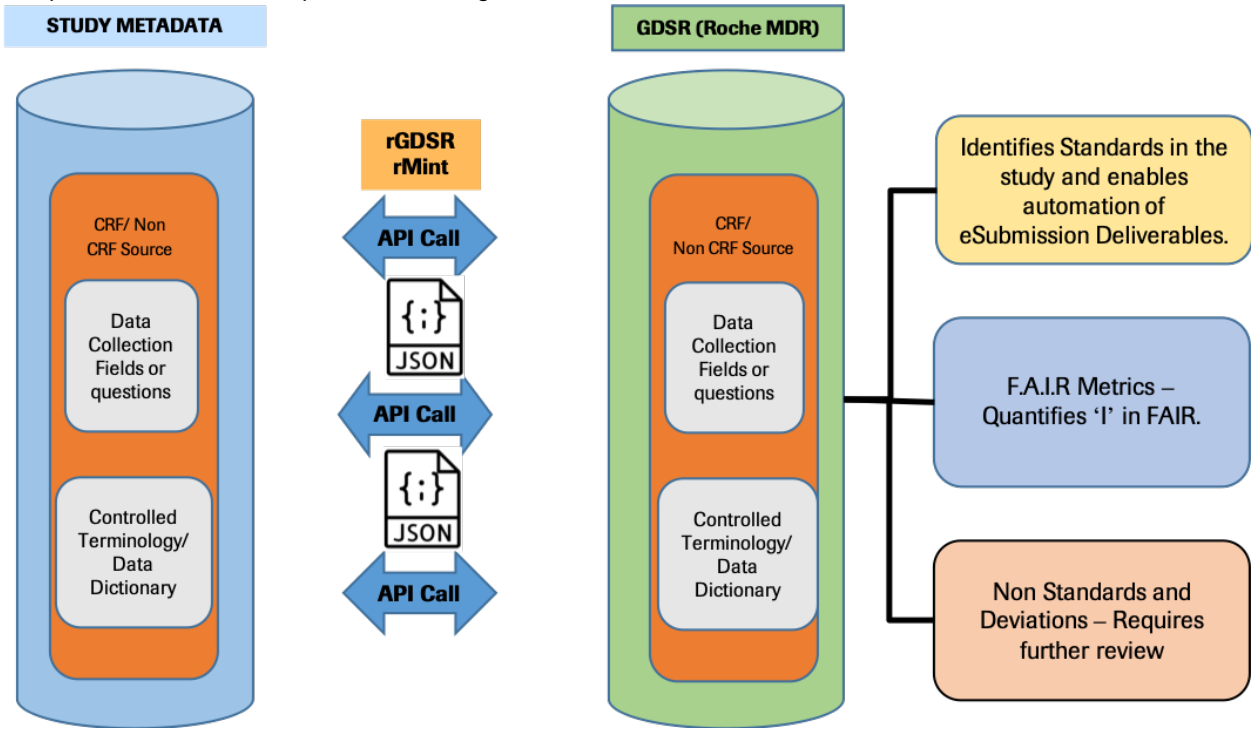
The GDSR uses publication versioning which allows users to access the archives of previous data standards versions. New, updated, and retired data standards are made available to users in a publication cycle and the previous publication is moved to an accessible archive.

METADATA INTEROPERABILITY APP

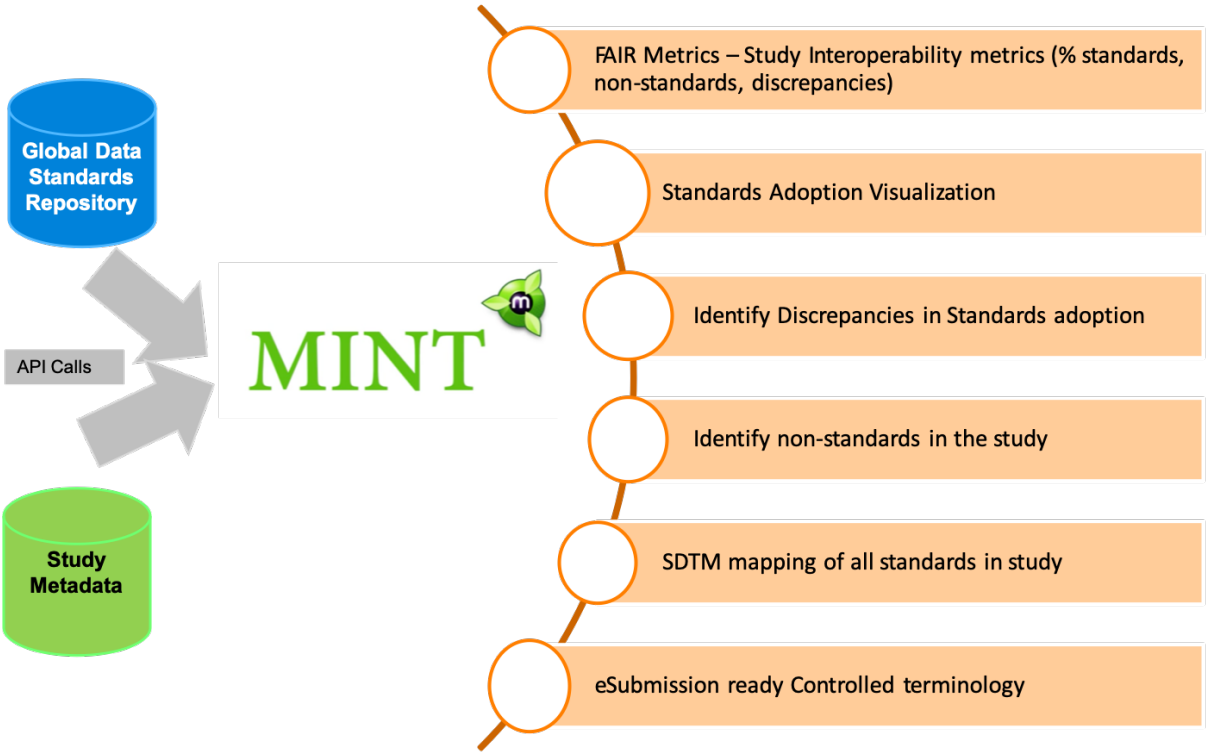
The Metadata INTeroperability App is R based solution that accesses the Metadata in GDSR via the REST APIs in real time to compare it with the study’s metadata either to a latest version or historic version of the standards. The app maximizes the use of data standards and terminologies in clinical studies by providing a snapshot of standards and terminologies used in the study design, deviation to standards and non-standards in the study.

APP ARCHITECTURE

The below picture explains the architecture of the app. At Roche, the study’s metadata lives in machine readable files. The CRF metadata is present in the Rave Architect loader file and the Non-CRF metadata lives in SAS datasets. The study’s metadata is then compared to the metadata in GDSR in real time based and they are compared based on the unique identifiers to generate the results.



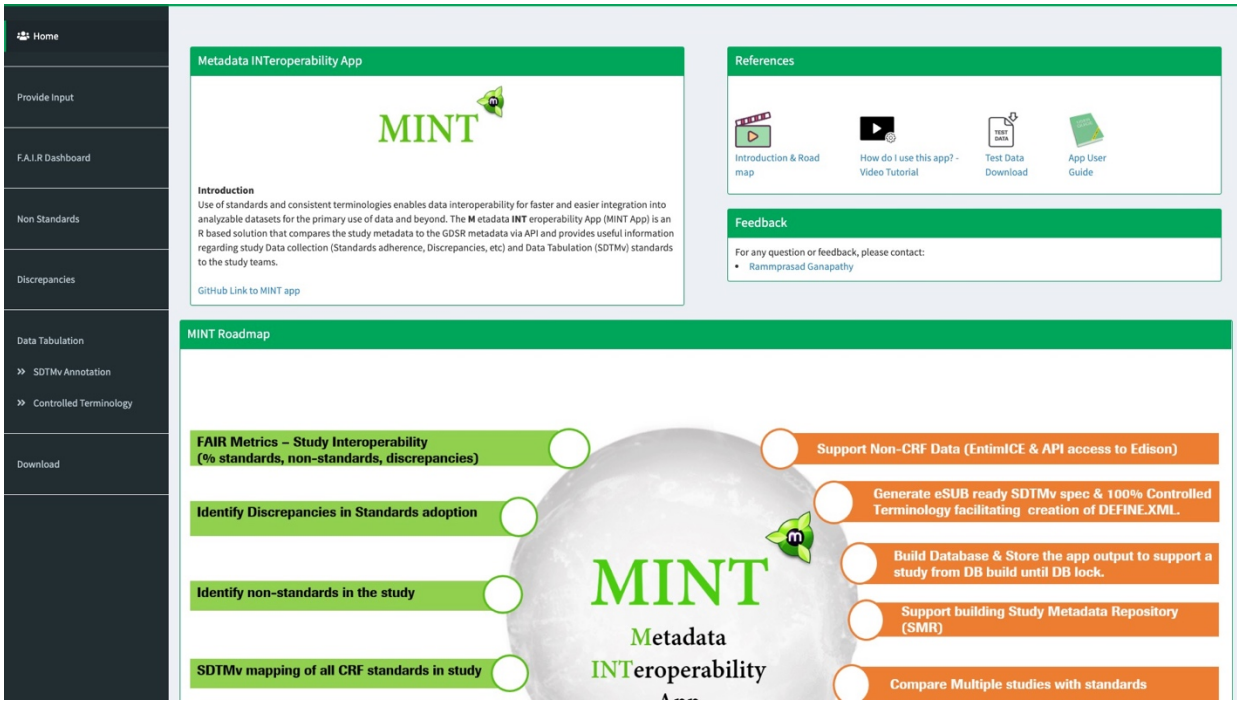
The results are categorized and presented to the stakeholders in an R-shiny app across multiple tabs and sub tabs. The users will have the ability to review the report in the App user interface or download the entire report in an excel format for offline review. The below picture represents the various functionalities of the app.



TECHNOLOGY - R PACKAGE AND SHINY MODULE

The app has been programmed using R and R shiny. To enable faster processing, the core logic of the app resides in the R-shiny packages and the user interface is built in R shiny using Shiny dashboard.

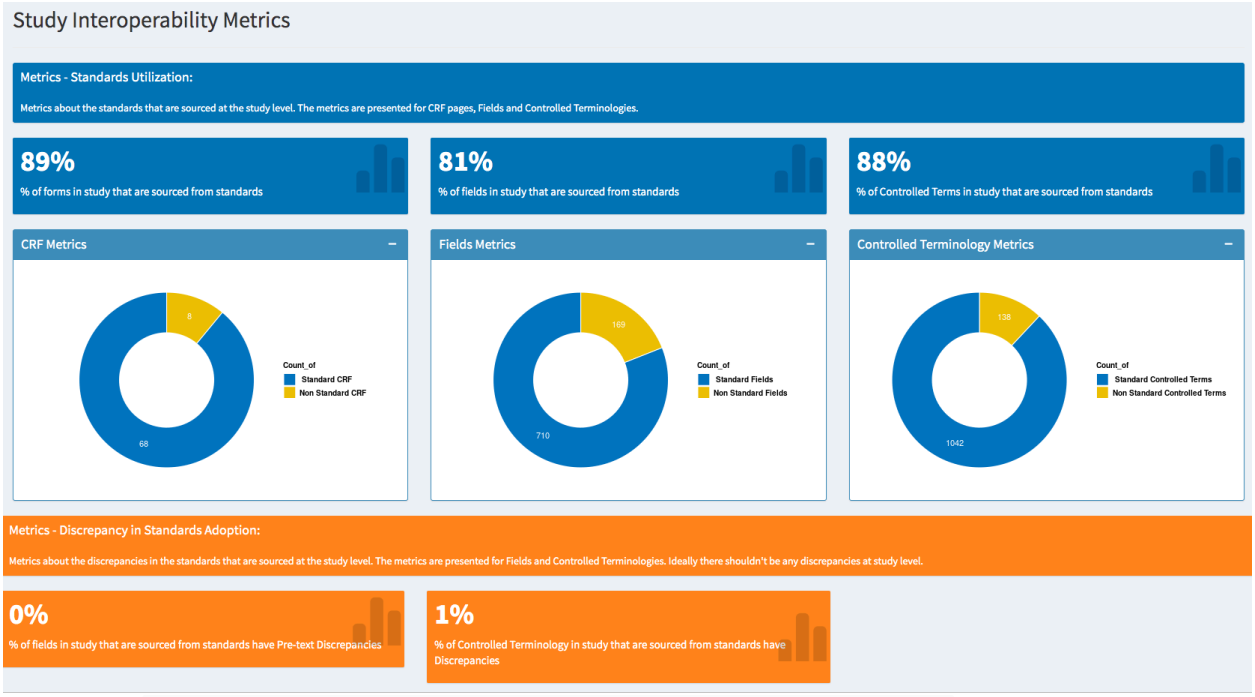
R packages are the fundamental units of reproducible R code. They include reusable R functions, the documentation that describes how to use them. We have leveraged this functionality and programmed to core logic of the Metadata INTeroperability app in various smaller functions that solves a particular problem. Some of the functions are made user facing where other programmers can install the package and use the functions in run time. As mentioned in the architecture diagram rGDSR and rMINT are the two home developed R packages that perform the code logic of the app. rGDSR parses the JSON files that is in nested tree like format to two-dimensional data frames for further processing by rMINT.



Shiny Dashboard (shinydashboard) makes it easier for the users to create R-shiny apps where the data has to be categorized and visualized in an organized fashion. We have leveraged this functionality to let the users easily use the app in a more organized manner. The picture above displaying the home page of the app with various tabs on the left.

HOW IT IS USED BY STUDY TEAMS

The app lets the study teams assess the standards compliance during various stages of a clinical study design and provides an opportunity to maximize the standards usage from GDSR. The visualization in the app provides the Interoperability metrics from data collection CRFs drilling down to the controlled terminologies used in the study design. Here is a snap shot of the Interoperability visualization from the app that provides the required details to the study teams on standards adherence and discrepancies in standards in the study.



The study teams use the app to

- Review the standards compliance and maximize standards adoption.
- Review the non-standards and make sure it adheres to the Roche design principles and also to make sure that the non-standards are mapped to SDTM following CDISC principles.
- Review and resolve the discrepancies in standards adoption before the study goes LIVE.
- Get the SDTM annotations of the standards in the study and use it to create SDTM aCRF and mapping specifications.

The versioning in GDSR lets user to choose the version of standards used for study design to produce accurate results. Essentially the users upload the study metadata to the app and choose the version of standards used for study design and review the results.

ENABLING AUTOMATION OF SDTM ESUBMISSION DELIVERABLES

The app enables the automation of SDTM eSubmission deliverables like SDTM aCRF. In future the app will enable automation of DEFINE.xml by automating the study SDTM specification and Controlled Terminology. If a study has 90% standards compliance, the app can automate 90% of SDTM eSubmission deliverables and study SDTM specifications.

USER EXPERIENCE

The app has been used more than 1500 times across Roche and Genentech since it was launched in Q3 last year. Among the first few studies used the app; the average Standards compliance percentage or the Study Interoperability percentage is 90%. This means that the study teams are able to automate a lot of activities in Data Collection and Data Tabulation. This also shows the quality of the Data Standards Roche has developed for many disease areas. The below section details the benefits the study teams got by adhering to the standards.

DATA COLLECTION

Study teams were able to copy 90% of the standard CRFs from the Rave Libraries to their studies in Rave. They were able to reuse 90% of edit checks along with pre-programmed Custom functions to their study. This means that the study teams have to concentrate only on the 10% of the CRF design in Rave and also the edit checks that are required as per the protocol and visit schedule.

DATA COLLECTION

Study teams were able to get SDTM mapping for 90% of the Standards CRFs from GDSR and the MINT app provides them in an excel format for further processing. The Study teams were able to use the template SAS programs and macros for the standard CRFs to create the SDTM datasets. Also, they were able to automate the e-Submission ready SDTM aCRF.

The study teams felt the app to be extremely useful in maximizing the standards adoption and thus get benefitted from all the downstream automation built around standards like Rave CRFs, Edit checks, Custom Functions, SAS programs to create SDTM datasets and automation tools to create SDTM annotated CRF.

CONCLUSION

Following FAIR principles and a Metadata driven approach has helped us create a solution that enables the creation of Interoperable clinical studies. Having the right infrastructure, well defined data standards, a FAIR MDR and efficient use of technology is key to the success in creation of a tool that automates the measure of Interoperability of Clinical studies and creation of high quality eSubmission documents. This idea could be extended to any EDC and MDR that meets the FAIR principles.

REFERENCES

1. https://www.nature.com/articles/sdata201618
2. https://www.force11.org/group/fairgroup
3. https://www.pistoiaalliance.org/
4. Roche FAIR Data playbook
5. GDSR FAIR – FAIR Data Standards in the GDSR (Global Data Standards Repository) Huw Mason, F Hoffmann - La Roche, Basel, Switzerland https://www.lexjansen.com/phuse/2019/pp/PP06.pdf
6. R packages by Hadley Wickham
7. Shiny Dashboard https://rstudio.github.io/shinydashboard/index.html

ACKNOWLEDGMENTS

The author wishes to thank our colleagues who contributed to various aspects of this initiative: The Data Standards & Governance team, the Early-Phase/Late-Phase Harmonization Team, the Global Data Standards Implementation Team, the Enhanced Data and Insights Sharing Reference Data Services, FAIR Playbook team and the Enhanced Data and Insights Sharing FAIR Metrics Team, Agile eSubmission team.

The author also wishes to thank Chris Price and Jonathan Chainey for reviewing the content of this paper.

CONTACT INFORMATION

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

Author Name – Rammprasad Ganapathy
Company – Genentech/Roche
Address – 1 DNA Way,
City / Postcode – South San Francisco. 64080
Email: ganapathy.rammprasad@gene.com