SDTM at the heart of Roche/Genentech's drive towards F.A.I.R. data

Presented by Rammprasad Ganapathy
Global Data Standards Manager, Data Standards & Governance, Biometrics, Product Development

F.Hoffmann-La Roche & Genentech Inc.

04.24.2019

Putting SDTM at the heart of Roche/Genentech's drive towards F.A.I.R. data

EDIS - Enhanced Data & Insights Sharing

An enterprise-wide initiative is ongoing at Roche/Genentech to make our internal data F.A.I.R* and CDISC's Study Data Tabulation Model (SDTM) is at the heart of this work. In this presentation, I will share insights into how we have applied SDTM as the target model for legacy data curation, as well its central role in the harmonization of data collection & tabulation standards for adoption across all of our early & late stage clinical trials.

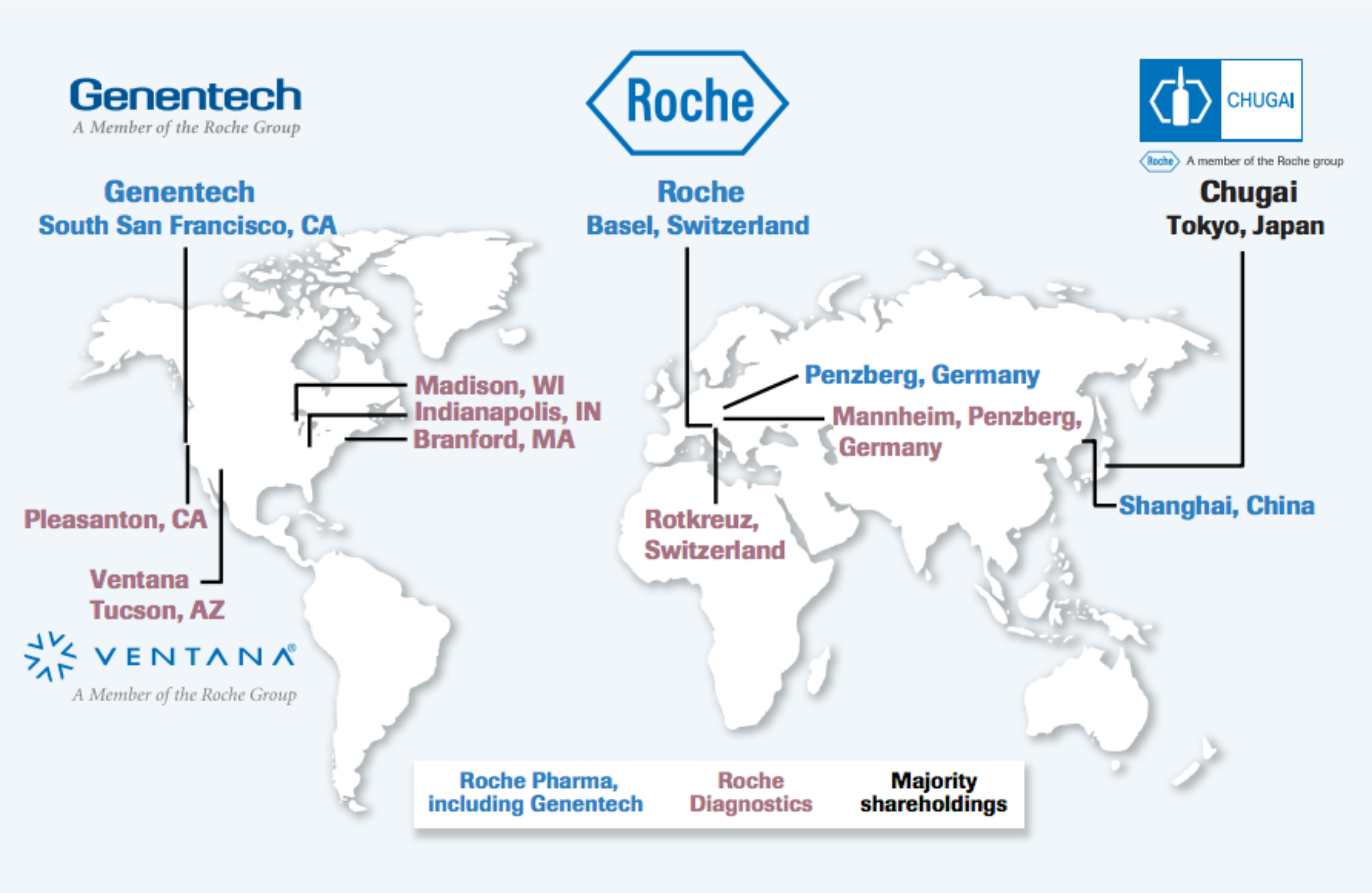
*FAIR - Findable, Accessible, Interoperable, Reusable

Agenda

1. Roche Landscape
2. Pharma Data Governance
3. F.A.I.R Introduction
4. Problem Statement
5. Retrospective View
6. Prospective View

Roche Landscape

The Roche Group

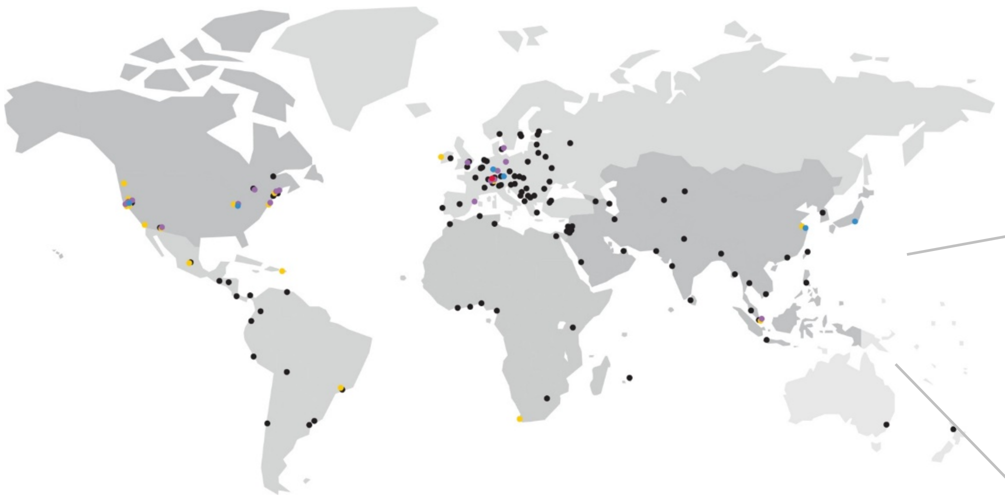


~90,000 employees worldwide

A global pioneer in personalized healthcare

We work across diagnostics and pharmaceuticals

Innovation is in our DNA



***30** R&D sites worldwide*

***26** Manufacturing sites worldwide*

***#1** R&D investor in healthcare*

*Among **top 10** R&D investors across industries*

🟥 Roche Group headquarters

🟦 Largest sites based on number of employees

🟪 Research and development sites in Pharmaceuticals and Diagnostic

🟨 Manufacturing sites in Pharmaceuticals and Diagnostics

⬛ Sales sites Pharmaceuticals and Diagnostics

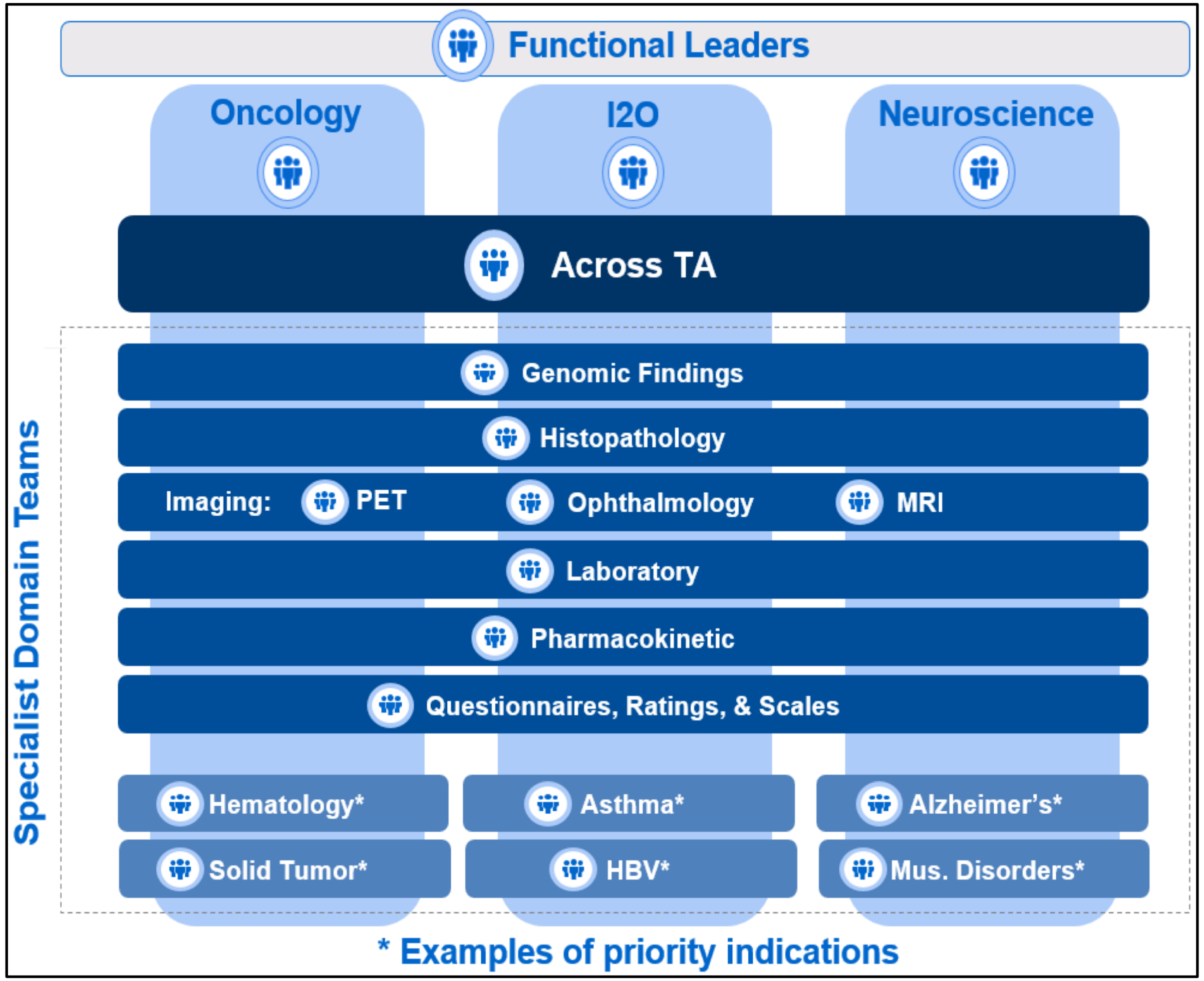
Roche Culture

- Highly de-centralized model -> diversity of thinking & approach.
- Collaborative environment with a shared sense of purpose.
- Encouraged to fail forward, learn, repeat -> succeed.
- Increased focus on Agility = Stability, Flexibility, & Speed.

Pharma Data Governance

Pharma Data Governance

Biomedical Domain



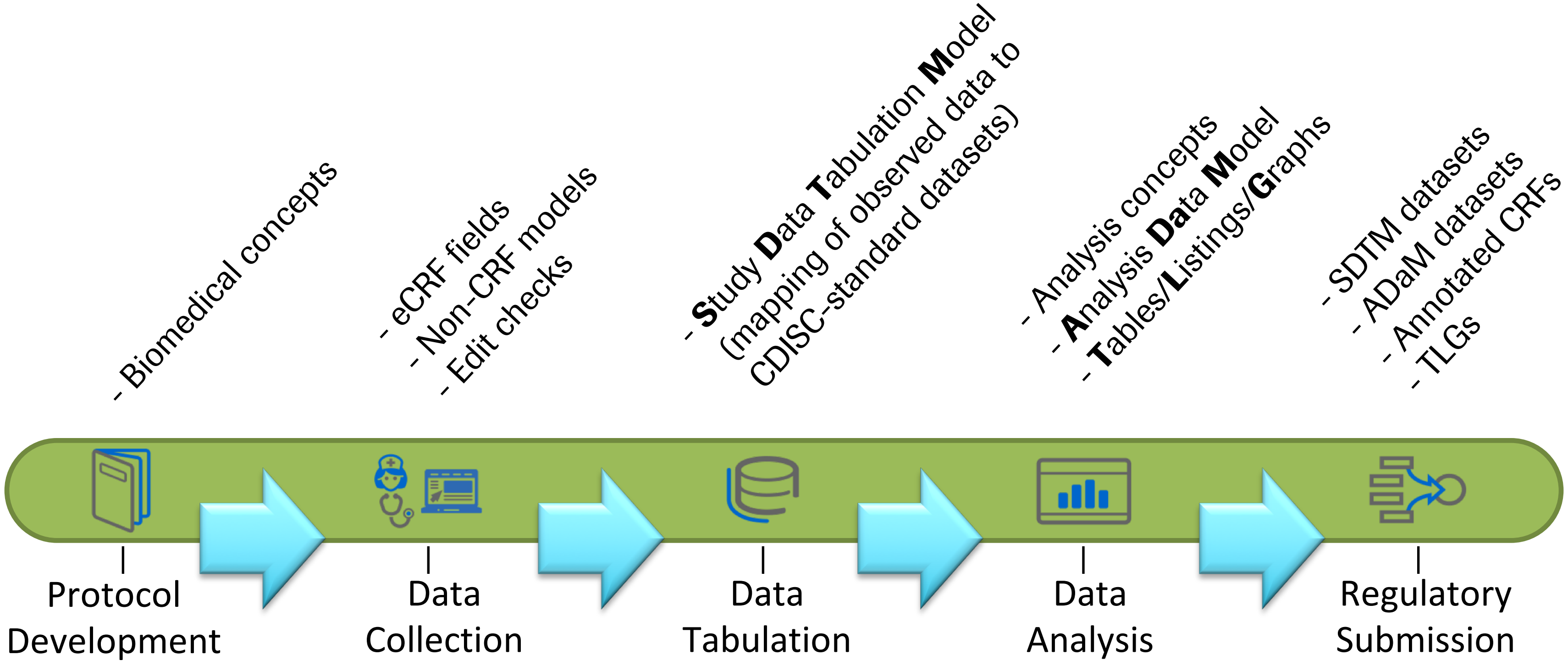
* Examples of priority indications

I2O: Immunology, Infectious Diseases, Ophthalmology

Roche Global Data Standards

What are they?

- Define how Roche clinical trial data should be collected, tabulated, analyzed, and submitted to regulatory authorities.
- Include all the ‘layers’ required, from biomedical concepts, metadata, and controlled terminology.
- Fully aligned with CDISC submission data standards (SDTM & ADaM), extended to ensure consistency across Roche clinical trials.



- Biomedical concepts

Protocol Development

- eCRF fields
- Non-CRF models
- Edit checks

Data Collection

- **S**tudy **D**ata **T**abulation **M**odel (mapping of observed data to CDISC-standard datasets)

Data Tabulation

- Analysis concepts
- **A**nalysis **D**ata **M**odel
- **T**ables/**L**istings/**G**raphs

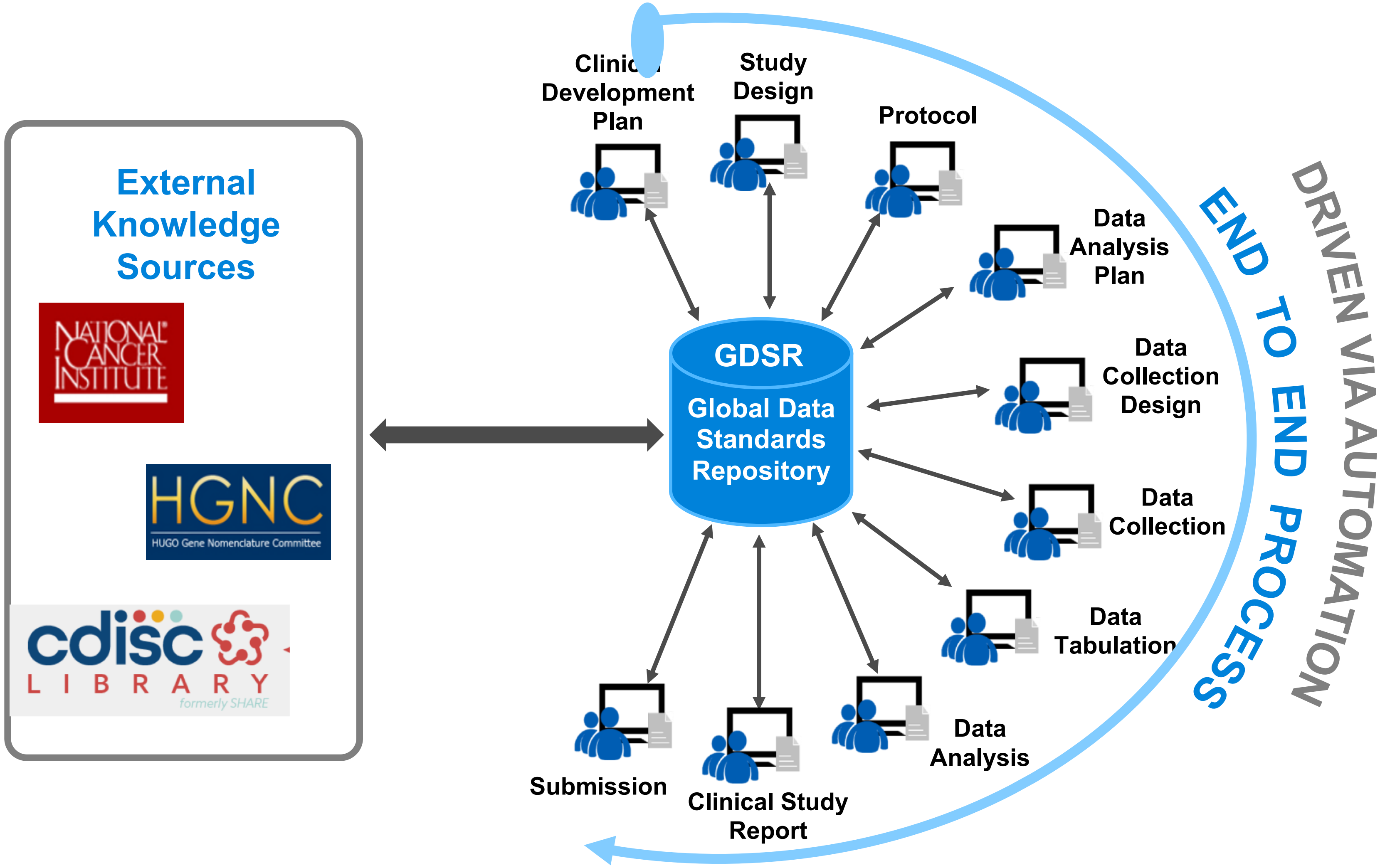
Data Analysis

- SDTM datasets
- ADaM datasets
- Annotated CRFs
- TLGs

Regulatory Submission

Technology

From Vision to Reality



External Knowledge Sources

GDSR

Global Data Standards Repository

Clinical Development Plan

Study Design

Protocol

Data Analysis Plan

Data Collection Design

Data Collection

Data Tabulation

Data Analysis

Clinical Study Report

Submission

END TO END PROCESS

DRIVEN VIA AUTOMATION

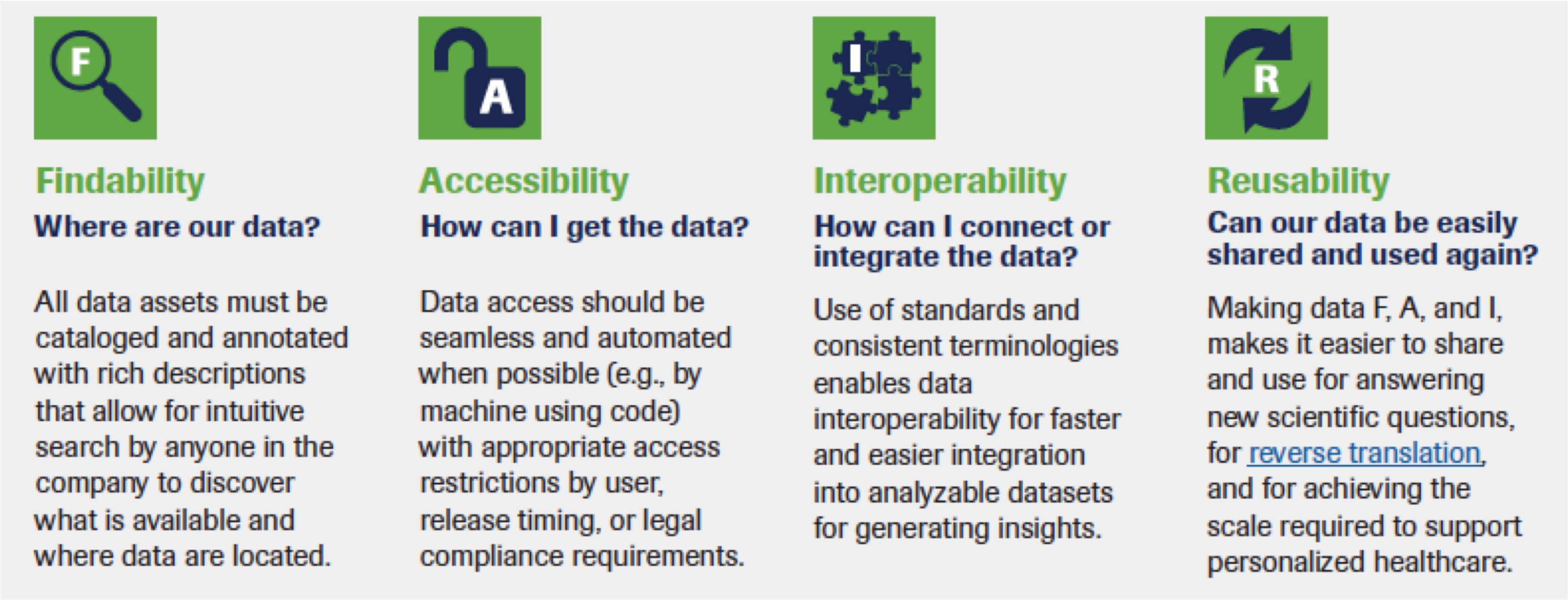
LET'S PLAY F.A.I.R.

EDIS

Enhanced Data & Insights Sharing

FAIR Overview

The term FAIR originated during a 2014 workshop in Leiden, Netherlands with attendees from academia, industry, funding agencies, and scholarly publishers.



Findability

Where are our data?

All data assets must be cataloged and annotated with rich descriptions that allow for intuitive search by anyone in the company to discover what is available and where data are located.

Accessibility

How can I get the data?

Data access should be seamless and automated when possible (e.g., by machine using code) with appropriate access restrictions by user, release timing, or legal compliance requirements.

Interoperability

How can I connect or integrate the data?

Use of standards and consistent terminologies enables data interoperability for faster and easier integration into analyzable datasets for generating insights.

Reusability

Can our data be easily shared and used again?

Making data F, A, and I, makes it easier to share and use for answering new scientific questions, for reverse translation, and for achieving the scale required to support personalized healthcare.

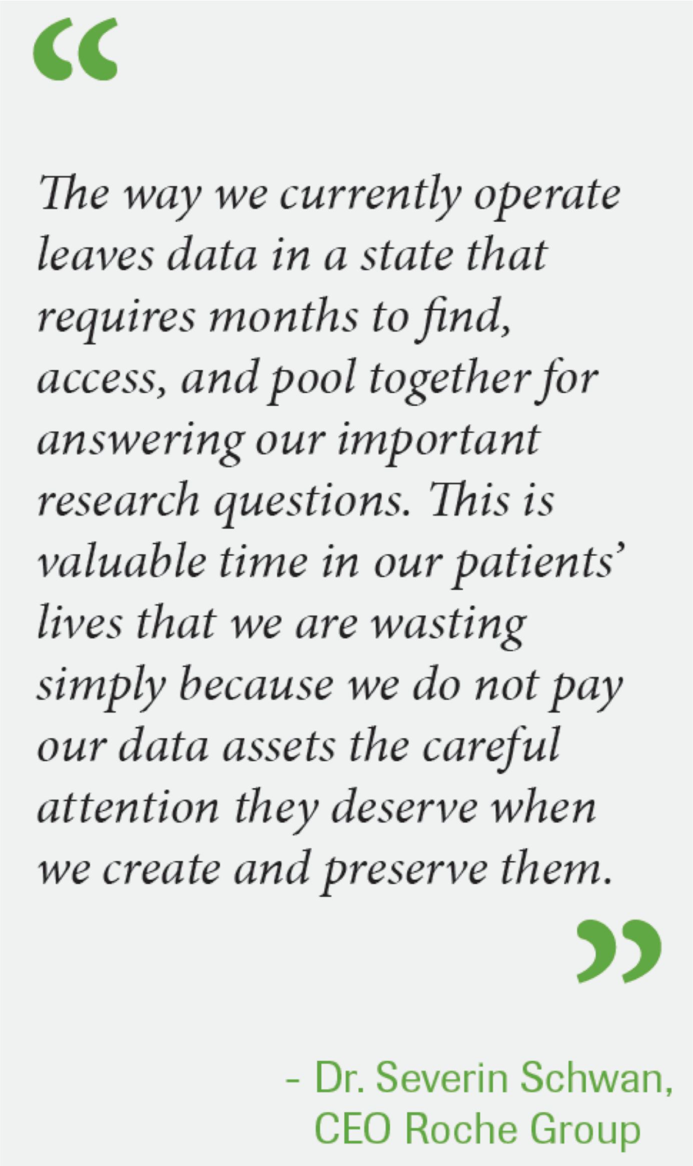
EDIS Problem Statement

Scientific Question:

“Across all Roche NSCLC studies, Do you have Whole Genome Sequencing data for patients with non-small cell lung cancer who had no history of smoking?”

“Give me one month to get back to you on that”

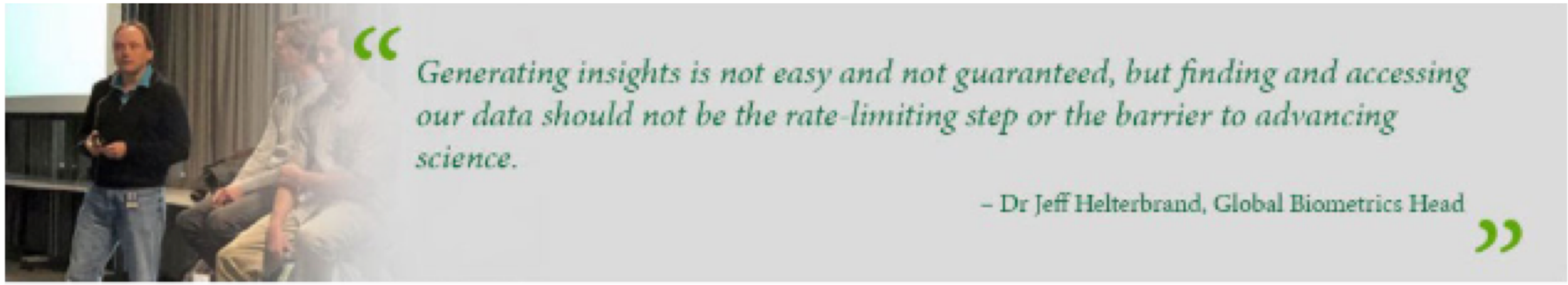
This type of information should be available at our finger tips.



> *The way we currently operate leaves data in a state that requires months to find, access, and pool together for answering our important research questions. This is valuable time in our patients’ lives that we are wasting simply because we do not pay our data assets the careful attention they deserve when we create and preserve them.*

- Dr. Severin Schwan,
CEO Roche Group

EDIS Problem Statement



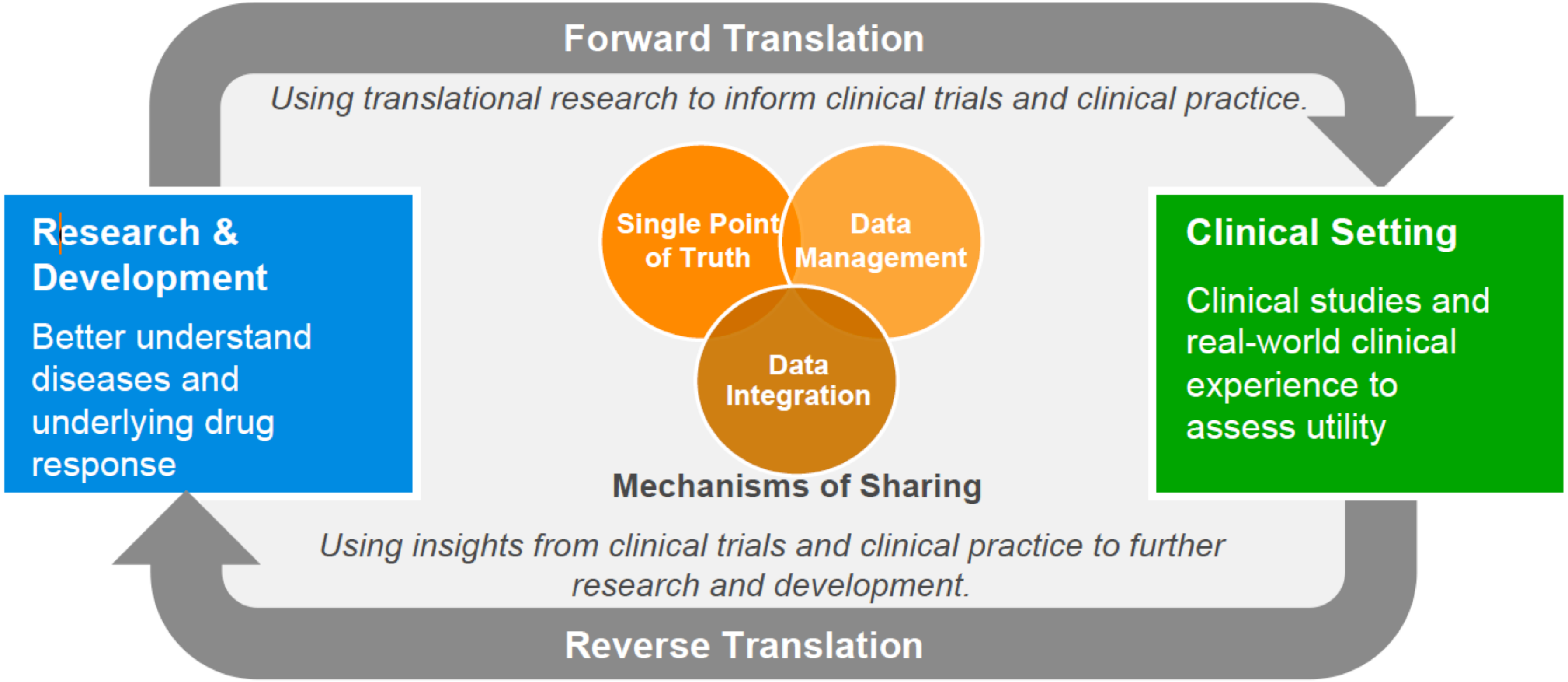
> *Generating insights is not easy and not guaranteed, but finding and accessing our data should not be the rate-limiting step or the barrier to advancing science.*
>
> – Dr Jeff Helterbrand, Global Biometrics Head

Difficult and time consuming to find, access, integrate and share our medical data, resulting in not maximizing the value of the expensive medical data we generate or Acquire

- Internal data assets – Clinical trials, genetics, Imaging, RNA, other biomarker
- Other data assets - Foundation Medicine, Flatiron, etc.
- External Data assets - TCGA, Genomics England, etc.

EDIS Vision

Accelerate reliable insights generation from data



Forward Translation

Using translational research to inform clinical trials and clinical practice.

Research & Development

Better understand diseases and underlying drug response

Single Point of Truth

Data Management

Data Integration

Mechanisms of Sharing

Clinical Setting

Clinical studies and real-world clinical experience to assess utility

Using insights from clinical trials and clinical practice to further research and development.

Reverse Translation

EDIS is not a system, it is the new way of working.

If we are investing to acquire data, we should invest to make good use of it.

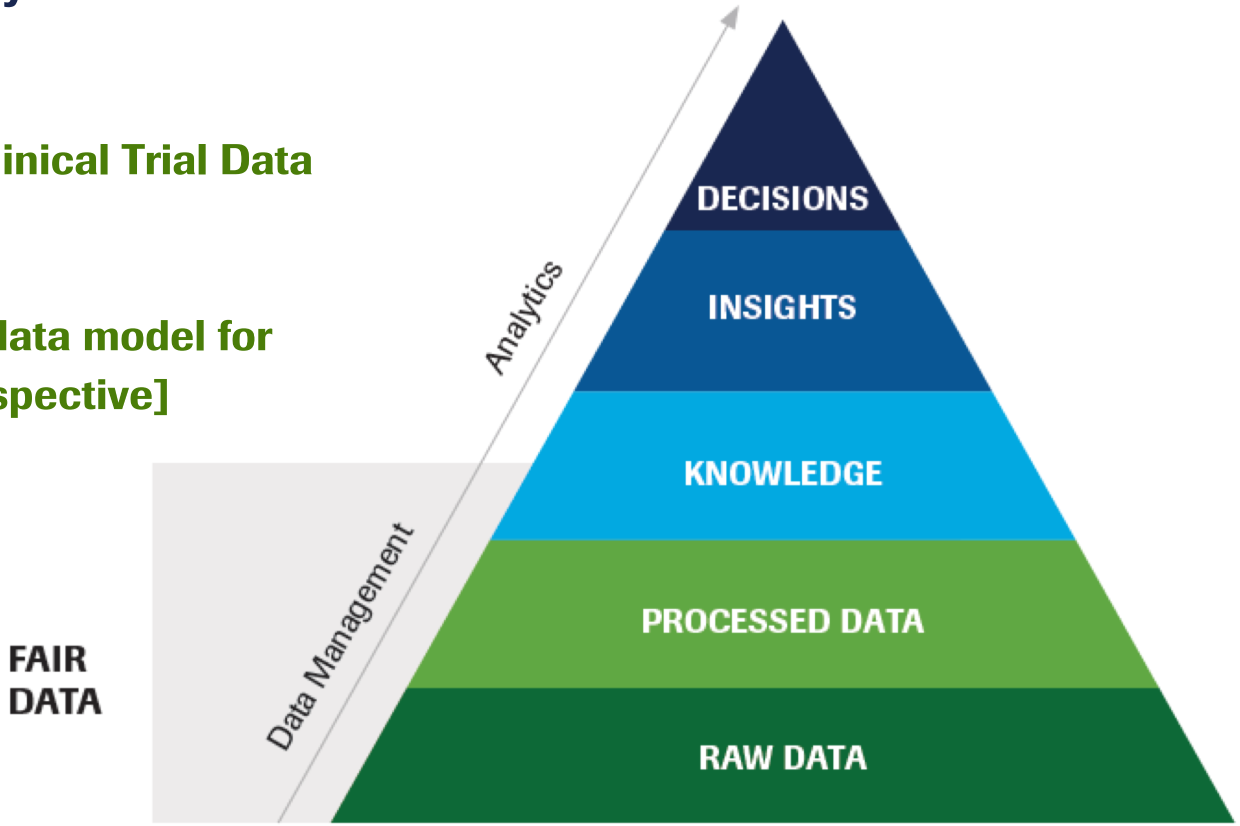
Our Strategy

EDIS is foundational to enabling transformation of data to insights

EDIS, with its focus on data capabilities, processes, and technologies for the first two layers, plays a foundational role.

Data later includes Diagnostics data, Clinical Trial Data and Real World Data.

CDISC SDTM is chosen as a preferred data model for clinical trial data [Prospective + Retrospective]



EDIS Program

Key Objectives – retrospective, prospective, and cultural

Maximize Value from Legacy Data

- ➢ Delivery of integrated datasets for specific use cases
- ➢ Retrospective curation and integration
- ➢ Define requirements for new infrastructure and tools

Drive FAIRification of Incoming Data

- ➢ F.A.I.R. data acquisition, processing, and management practices
- ➢ Infrastructure and tools including workflow and process automation

Support a Data Citizenship Culture

- ➢ Culture of F.A.I.R. and SHARED data
- ➢ Network of data ambassadors

Prospective View –
Data FAIRification of Incoming data

Goal – All incoming data will be F.A.I.R by 2020.

- The prospective part of the goal is ensuring that we have the **capabilities and processes in place** automatically from the beginning--from the get-go as the data comes in.

- If we follow these principles, then **we don't have to chase our tail every time** and go back retrospectively and do what has to be done.

- Basically, what this means is that the traditional work of a data collection, which has to be done to meet current needs for analysis & submission **(Primary use)** and, also needs to ensure that future use cases can still use your data. **(Secondary Use)**

Prospective View – Clinical Trial Data
Data FAIRification of Incoming data

Challenges:

1. Different collection standards across Early Phase studies & Late Phase studies.

2. Multiple Rave URLs that create SAS on demand data from EDC with different structures.

3. No harmonized process to create SDTM datasets across Roche.

4. No Metadata repository for Early Phase standards.

5. External vendors send data to Roche in different formats.

6. Gaps in tools to assess the studies on alignment to standards.

Prospective View – Standards Implementation

Why can't we all use the same standards, implement them the same way and release them at the same time?

- Launch of Global Data Standards Implementation team – 2018
- Proposal to harmonize standards across EP, LP and beyond

Proposal

*Further align Early Phase data collection standards with Late Phase standards to create **efficiencies downstream by** enabling the use of same programs, tools **and resources***

Challenges.. Leading to Innovation

Pre Harmonization	Post Harmonization
EP DT Standards not part of GDSR	GDSR (MDR) – SPoT for EP & LP DT standards.
Two sets of Template programs & Inconsistent process to create SDTM datasets between EP & LP	Harmonized Consistent process and Template programs/macros for EP & LP.
Template programs & macros for <30% Standard CRFs	200+ SDTM template SAS programs & macros.
Less efficient and more study specific updates.	Highly efficient – Smartly identify source and map it to Target SDTM.
No easy way to identify the study specific forms & fields that are not in GDSR (MDR).	R-Shiny App to compare the study metadata to GDSR (MDR) via API.

Philosophy & Approach

Global Data Standards Implementation (GDSIT) team – A cross-functional Data Standards Implementation (GDSIT) team was formed and tasked to implement the harmonized data collection & tabulation standards across EP & LP.

- Use best parts of EP & LP processes. DTS components were updated to include the best parts from EP.

- Data points collected are cleaned and tabulated the same way across EP & LP

- Create reusable tools and template programs so study teams can reuse them with ease.

SDTM Implementation Template SAS programs

How Reusable are the SDTM template SAS programs & macros?

Scenario	Reusability of Template programs
100% copy from the standards	100% Reusable. No updates required.
Only few data collection components are removed from the standard CRF while implementing in the study.	100% Reusable. No updates required.
New study specific data collection components are added to a standard CRF	Updated required to the template spec and the program. Will be highlighted via R-shiny app.
New study specific CRF	New program has to be created. This will be highlighted via R-Shiny app.

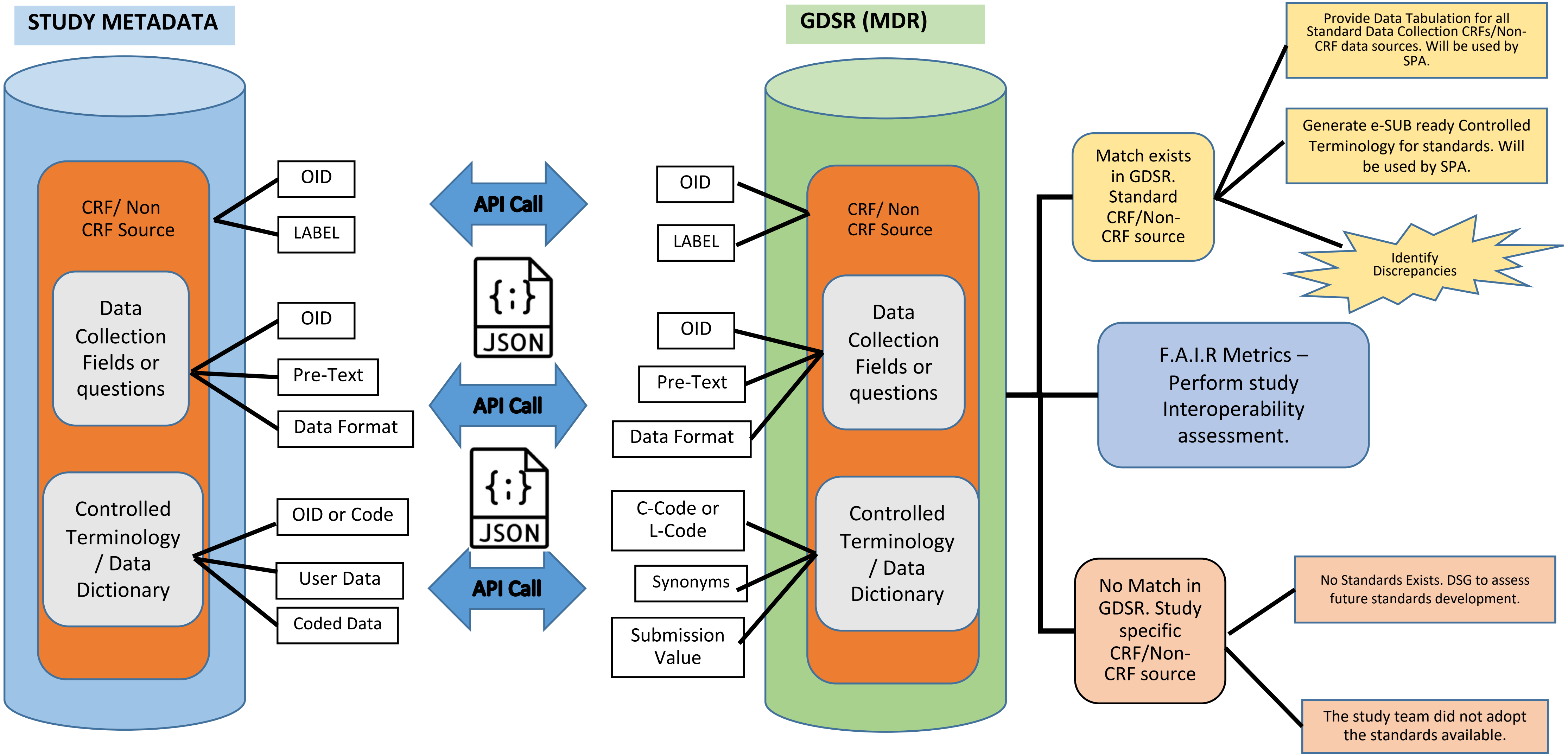
GDSR(MDR) API & Supporting Tools

GDSR(MDR) API Enhancements – New API/ Facets are programmed to provide the SDTM annotations for CRFs of all TA standards across EP & LP.

R- Shiny App - Leverage GDSR(MDR) API and Study metadata to provide the below to the study teams.

- **Identify Study specific forms & fields** – To help further discussion with the Biometrics & Data Standards Group.
- Provide **SDTM mapping of all standard forms & fields** in the study in one button click.
- Provide **Controlled Terminology for all standard NCI SDTM code lists** used in the study.
- Assess how close **the study is aligned to the standards**.
- Ensure **latest Lab (Local labs) & QRS metadata** is used in the study design.

R-shiny app architecture


STUDY METADATA
CRF/ Non CRF Source
Data Collection Fields or questions
Controlled Terminology / Data Dictionary
OID
LABEL
OID
Pre-Text
Data Format
OID or Code
User Data
Coded Data
API Call
{;}
JSON
API Call
{;}
JSON
API Call
GDSR (MDR)
OID
LABEL
OID
Pre-Text
Data Format
C-Code or L-Code
Synonyms
Submission Value
CRF/ Non CRF Source
Data Collection Fields or questions
Controlled Terminology / Data Dictionary
Match exists in GDSR. Standard CRF/Non-CRF source
Provide Data Tabulation for all Standard Data Collection CRFs/Non-CRF data sources. Will be used by SPA.
Generate e-SUB ready Controlled Terminology for standards. Will be used by SPA.
Identify Discrepancies
F.A.I.R Metrics – Perform study Interoperability assessment.
No Match in GDSR. Study specific CRF/Non-CRF source
No Standards Exists. DSG to assess future standards development.
The study team did not adopt the standards available.

Maximizing value of Legacy Clinical Trial Data –

Retrospective View

Maximizing value of Legacy Clinical Trial Data – Retrospective View

- Datasets targeted for specific Scientific use cases.

- The Use Cases are formed to re-use data collected from previous studies to explore different scientific questions with a goal to reveal new insights for R&D.

Challenges:

- Curation of the Legacy Data – Most time is spent is spent on finding the data and the associated Metadata.

- Mapping of Legacy data to SDTM IGv 3.2.

- It was challenging to map & curate the data that was collected ~8 to 10 years before to SDTM. Simply the data did not fit to the SDTM models.

- Not able to create a scalable model to create the mapped datasets.

Maximizing value of Legacy Clinical Trial Data – Retrospective View

Tools & Processes:

- Reusable R packages to map the source to target SDTM model.
- Use Machine Learning tools & techniques for legacy data curation.
- Concentrate only on priority domains that are required to answer key scientific questions along with the WGS or any other biomarker data.
- Create a subset of P21 checks that can work with the priority domains and focused on legacy data curation.
- Create R packages to perform the P21 checks to ensure the consistency of the mapped legacy data.

Our Story Continues…

Route 66
Pioneering Single Lane Highway

Autobahn
Information Superhighway

PHC: “the right drug for the right patient at the right time”
Same destination… but now with a faster and better path

And EDIS is foundational to transforming how we get there!

Scientific Question:

“Across all Roche NSCLC studies, Do you have Whole Genome Sequencing data for patients with non-small cell lung cancer who had no history of smoking?”

“I will get back to you in a day”

This is only possible if all the incoming clinical trial data is collected & tabulated in a harmonized fashion and also by adhering to the CDISC SDTM standards.

Thank you!!

Doing now what patients need next