cdisc

2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL

Clinical Data Reuse: Opportunities and Challenges

Presented by Sid Prabhu and Parag Limaye, Deloitte

Meet the Speakers

Sid Prabhu

Title: Manager

Organization: Deloitte

Sid has 5 years of R&D experience, specifically in clinical development, across global Pharmaceutical and Biotechnology organizations. He has deep experience in technology-led solution delivery including knowledge of CDISC SDTM/ADaM standards, medical data review business processes, and the governance setup for clinical data reuse. He has delivered multiple projects in the R&D domain spanning areas including strategy, operating model development, process optimization, clinical technology selection/implementation, and analytics.

Parag Limaye

Title: Senior Manager

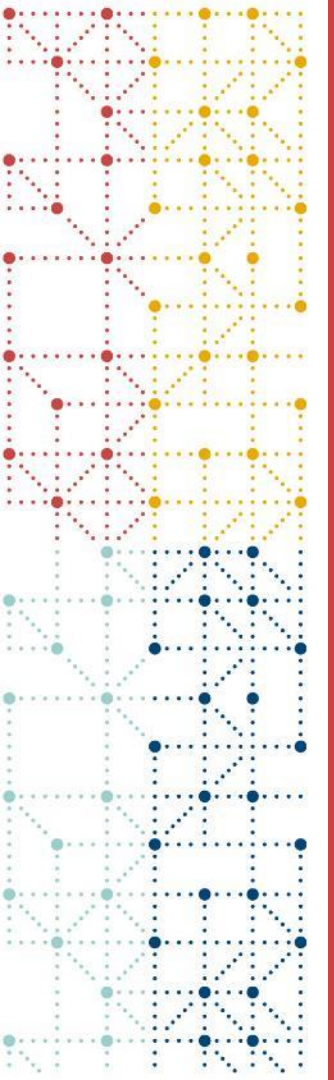
Organization: Deloitte

Parag is part of Deloitte's Life Science practice working closely with clients R&D functions. His primary emphasis has been on building innovative solutions for end-to-end clinical data management and advance analytics. He has led several large-scale implementations using agile methodology. In addition to his client role he also leads development of Deloitte hybrid assets in applications like automated data conversion, harmonization using advance analytics to generate insights like medical monitoring, safety signal detection, precision medicine etc..

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author(s) have no real or apparent conflicts of interest to report.*



Agenda

1. Building an R&D Organization of the Future
2. Approaching Data-Driven R&D: Challenges
3. Role of Standards in Data Reuse: Examples
4. Key Takeaways

Building an R&D Organization of the Future

The R&D Organisation of the future is data-driven

Changing industry and ecosystem trends mandate that R&D organisation capitalize on opportunities for data-driven research and development in order to remain competitive

Internal Aspirations

Cutting edge innovation in R&D Organisations today is tied to the ability to perform complex manipulations on a wide range of data and data types, which these organisations struggle to do.

Main challenges with the use of data

- Lack of cross-functional expertise to harness value from the data
- Siloed environments for data analytics
- Outdated data management tools that do not account for cross-study/cross-modality analysis
- No access to unified data for analysis
- Lack of standardized and efficient process to accessing data

External Trends & Dynamics

European Commission

The European Health Data Space (EHDS) initiative aims to facilitate **secondary use of data** for researchers, policy-makers and companies (European Commission, 2022)

FDA

The FDA Modernization Act 2.0 enables use of alternatives to animal testing when feasible, including **computer models** (JJ Han, 2023)

Insilico Medicine

Insilico's AI model application phase could **reduce time** in discovery phase **by > 1.5 years** (Insilico Medicine, 2022)

gsk GlaxoSmithKline 23andMe

GSK and 23andMe's partnership to reuse anonymised clinical data has already helped advance an immuno-oncology antibody **into clinical development** (GSK, 2018)

The Data-Driven R&D Organization Of The Future

Organisations can solve **their complex R&D challenges and contextualize other data modalities (such as RWD and omics)** by leveraging **their historic clinical data**. To do this successfully, an R&D organisation of the future **must bring together data standards, wrangling capabilities, analytics, and insight applications**

Approaching Data-Driven R&D: Challenges

Signature Issues in the Clinical Data Reuse

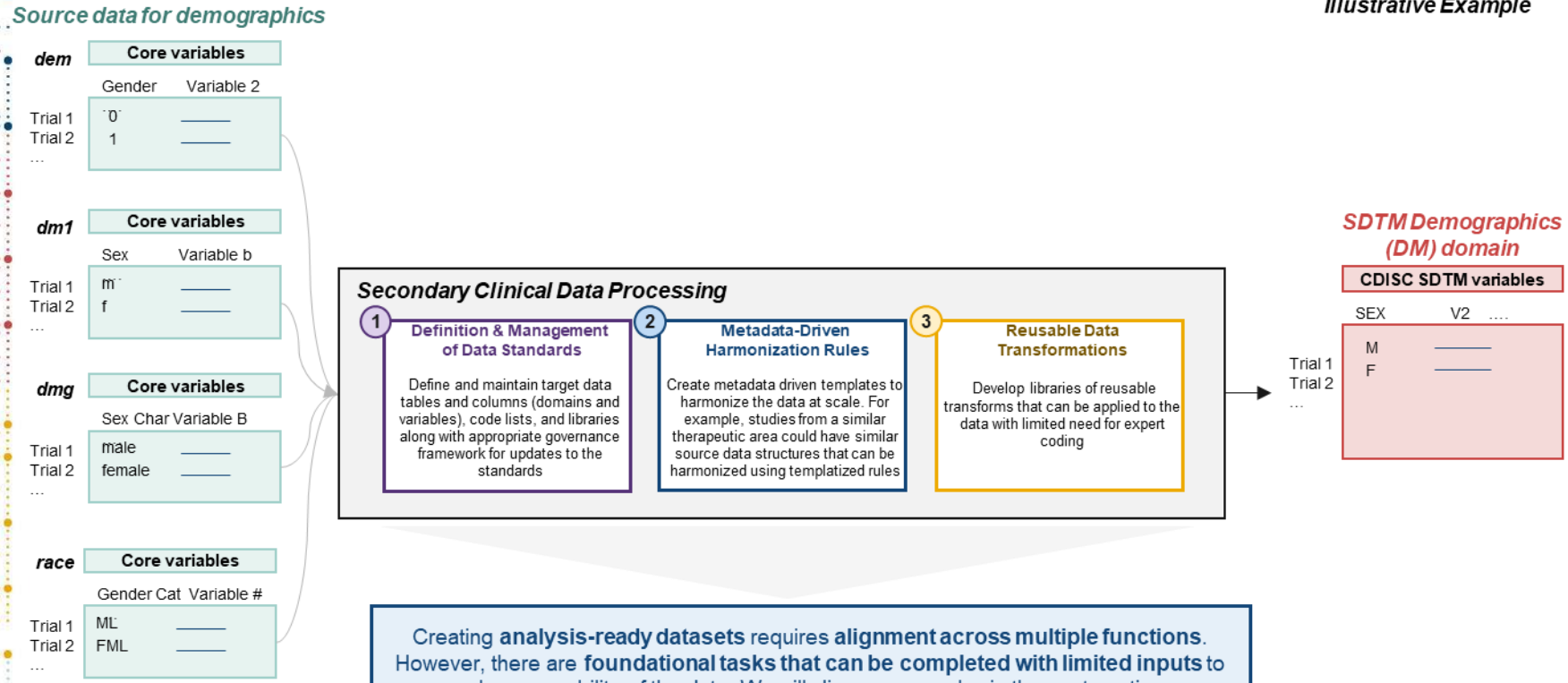
Operational, functional, & technical issues organisations face in clinical data reuse

Role of Standards in Data Reuse: Examples

Approach to Reusing Clinical Data

Overcoming the operational, functional, and technical challenges to reuse historic clinical data requires robust standards management, metadata driven harmonization rules, and reusable data transformations

Illustrative Example



Creating **analysis-ready datasets** requires **alignment across multiple functions**. However, there are **foundational tasks that can be completed with limited inputs** to enhance usability of the data. We will discuss examples in the next section

Opportunity A: Enhancing Quality of Historic Clinical Data

Studies that are 10-15 years old or may potentially have not been submitted to regulatory agencies for various reasons could have several quality issues that make the data unusable

1

Populating DTHFLG and DTHDTC

Incomplete DTHFLG and DTHDTC can be populated using the following logic:

1. *If the patient is shown to have died in the Disposition (DS) domain, but the date of death in the Demographics (DM) domain is null, then dm.DTHDTC should be populated with the disposition date corresponding to the death event.*
2. *If the patient is shown to have died in the Adverse Events (AE) domain, but the date of death in the Demographics (DM) domain is null, then dm.DTHDTC should be populated with the end date of the fatal event.*

2

Populating RFSTDTC and RFSXDTC

Incomplete reference date in demographic domain can be populated using following logic:

1. *If RFSTDTC is available at source keep the value as is else take the DSSTDTC value where DSDECOD is "informed consent date"*
2. *RFSXDTC is calculated by taking max of EXSTDTC where either EXODSE > 0 or EXDOSTXT is not null or EXTRT is Placebo or using EXSTAT*
3. *In cases where subject is in following 3 treatment arms ('SCREEN FAILURE', 'NOT ASSIGNED' or 'ASSIGNED, NOT TREATED'), then RFSTDTC should be set to null.*

3

Formatting Dates

In many instances, dates in the –DTC variables may not be formatted to the ISO standards or may be incomplete.
To enhance the usability of the data, the following can be enabled:

1. *Simple mid-point imputation of incomplete dates (for e.g., 02-2020 → 15-02-2020). The dates can be re-imputed by analysis teams once they have a hypothesis*
2. *Formatting of all dates to the ISO-8601 standard to ensure ease of comparison across time points in the study*

Use Cases Enabled

- Improved searchability of data for exploratory analysis
- Foundational quality improvements for trial design support
- Patient Population Safety Analytics across multiple studies

Opportunity B: Standardizing Findings Domains Across Clinical Repository

Findings domains like LB and QS have a wealth of information that can be leveraged for intelligent decision-making. However, limited standardisation within organizations makes a lot of this data unusable.

Source Data

Illustrative Example

STUDY	USUBJID	LBSPEC	LBTEST	LBTESTCD	LBORRES	LBORRESU
NCT001	NCT001_1	Urine	Alb_1	ALB_1	g/dL	4.4
NCT002	NCT001_2	Urine		ALB_U	g/dL	4.6
NCT003	NCT002_3	Serum	Alb_TEST		mg/dL	4500

Source Data Issues:

- *Missing Tests and Test Codes*
- *Inconsistent Test and Test Codes*
- *Internal Standards for Test and Test Codes*
- *Inconsistent Units*

Standards Repository

DOMAIN	TESTCD	SPEC	STRESU
LB	ALB	All	mg/dL
LB	ALB	All	mg/dL
LB	CREAT	Urine	g/L

DOMAIN	TEST	ORRES	STRES	CF
LB	All	g/dL	mg/dL	1000
LB	Bili	mg/L	umol/L	1.72

Units Repository:

- *Repository of standard units for a test and specimen*
- *Repository of conversion factors from source to target*

Codelist	Code	Synonyms
Lab Test	Albumin	ALB, Albumin
Lab Test Code	ALB	ALB, Albumin

Repository of Standard Tests and Test Codes*

*There must be consideration for sponsor-specific standards for exploratory labs

Output Data

Illustrative Example

STUDY	USUBJID	LBSPEC	LBTEST	LBTESTCD	LBORRES	LBORRESU	LBSTRES	LBSTTESU
NCT001	NCT001_1	Urine	Albumin	ALB	g/dL	4.4	4400	mg/dL
NCT002	NCT002_2	Urine	Albumin	ALB	g/dL	4.6	4600	mg/dL
NCT003	NCT003_1	Serum	Albumin	ALB	mg/dL	4500	4500	mg/dL

Findings domains consistently aligned:

- *Test and test codes standardized across studies allowing users to easily search and scan through data for cross-study analytics*
- *Units are standardized across all tests so that comparative analysis can be performed to study trends across patients in the study*

Use Cases Enabled

- Data Pooling for Publications and Poster Presentations
- Inclusion/Exclusion criteria refinement based on population response to drug
- Cataloging of important tests, test results, and baseline data availability across trials

Opportunity C: Derivation of Baseline and Associated Variables

The identification of baseline values for a test is dependent on the endpoint for the study, thereby leading to complexity when performing cross study analysis. Rederiving baseline based on exposure can alleviate this challenge.

Deriving –LOBXFL

SDTM Implementation Guide Definition: Last non-missing value prior to RFXSTDTC (Operationally derived)

STUDY	USUBJID	RFXSTDTC
NCT001	NCT001_1	02/23/2009
...	...	...
...	...	...

STUDY	USUBJID	LBTEST	LBDTC	LBSTRESN
NCT001	NCT001_1	Albumin	02/16/2009	2.6
NCT001	NCT001_1	Albumin	02/28/2009	3.9
NCT001	NCT001_1	Albumin	03/20/2009	3.8

Source Data Requirements

- *Derivation of reference start date by drug exposure for individual subjects*
- *Standardized dates for –DTC columns in Findings domains*

STUDY	USUBJID	LBTEST	LBDTC	LBORRES	LBLOBXFL	LBSTRESN_CHG_FROM_LOBX	LBSTRESN_PCHG_FROM_LOBX
NCT001	NCT001_1	Albumin	02/16/2009	2.6	Y	0	0%
NCT001	NCT001_1	Albumin	02/28/2009	3.9		1.3	33%
NCT001	NCT001_1	Albumin	03/20/2009	3.8		1.2	31%

Output Data

- *--LOBXFL derived for findings value captured prior to drug exposure*
- *Change from baseline, percent change for baseline derived for every subsequent reading after subject was exposed to trial drug*

Use Cases Enabled

- Exploratory Analytics to develop novel hypothesis for drug-repurposing
- Trial Design improvements based on learnings across multiple failed studies
- Patient Population Safety Analytics across multiple studies

Key Takeaways

Core Principles to Develop Next-Gen Data Management Platforms

Unlocking value from historic data while continuing execute on primary clinical data use cases requires holistic thinking and organizations need to deliver on fundamental values to be future-proof

CORE CAPABILITIES FOR R&D DATA MANAGEMENT

DATA PRODUCTS

Data assets, pools, and transformation tools that can be leveraged by multiple teams for exploration and decision-making

- TRANSPARENCY
- FLEXIBILITY
- REUSABILITY
- AUTOMATION
- CONFIGURABILITY
- WORKFLOWS

ANALYTICS PRODUCTS

Tools and algorithms that can be reused by teams to explore the data products and derive insights

- ACCESSIBILITY
- GOVERNANCE
- FLEXIBILITY

BUSINESS PRODUCTS

Governance and processes that allow teams to find and access the right data for the right use case

- SECURITY
- SEARCHABILITY
- STANDARDISATION

Key Takeaways

Creating value from the wealth of data that life-sciences organization collect and store requires leaders to bring together standards experts, engaged medical teams, and delivery specialists

1

STANDARDS EXPERTS

Experts on the clinical data standards can help curate datasets as per the needs of the analysis experts

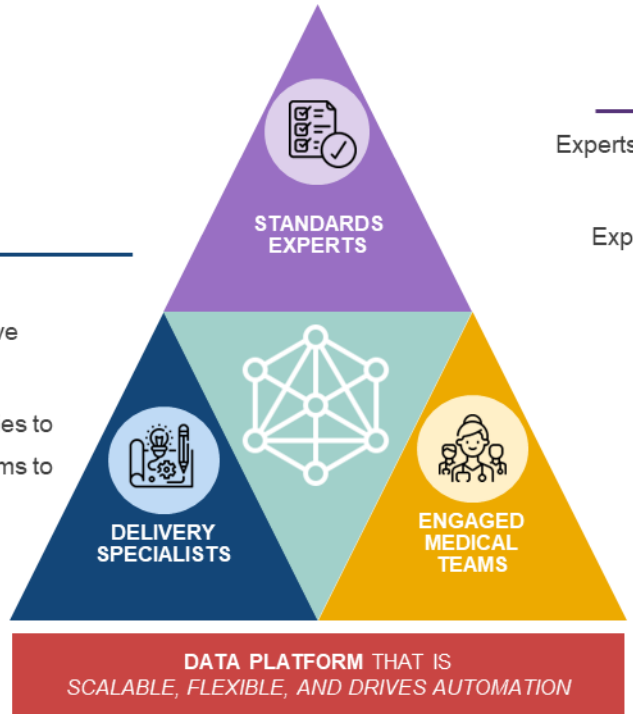
Experts on governance and anonymization of data can help expand the wider usability of the data

2

DELIVERY SPECIALISTS

Cross-functional domain specialists who can facilitate conversations with the team and drive outcomes.

Data Experts who have engineering capabilities to work with standards experts and medical teams to deliver data needed for analysis



3

ENGAGED MEDICAL TEAMS

Engaged scientific teams can develop hypothesis that can be tested using internal data assets

Medical experts who understand the disease areas and can identify trends that could support the novel hypothesis

Acknowledgements

Colin Terry

Deepak Kannangala

Ateeq Rehman

Rhea Varghese

Bhavesh Jain

Jonathan Dean

Rohan Khurana

Thank You!