

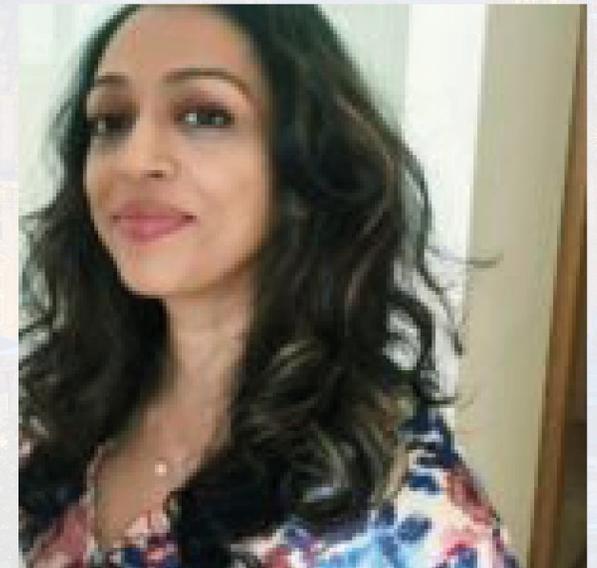


Parallel Session 2.1

Data Integration: How best to perform data management activities on data from multiple external sources

Presented by: Rashida Rampurawala

Manager – Study Data Management, GSK



+

•

○

How best to perform data management activities when data is collected from multiple external sources

Speaker:

Rashida Rampurawala: Manager – Clinical Data Management, GSK



DISCLAIMER + •

○

The content displayed in this presentation is the intellectual property of the presenter. You may not reuse, republish, or reprint such content without our written consent.

All information posted is merely for educational and informational purposes. It is not intended as a substitute for professional advice. Should you decide to act upon any information discussed, you do so at your own risk.

The views expressed are of the presenter and not of the organization the presenter is associated with.





Table of contents

- Introduction
- Defining Data Elements for integration
- Types of Data Integration
- Mapping of integrated data
- Data Integration – questions to ask
- Evaluating capabilities
- Challenges
- Cost and effort of data integration from multiple sources
- Risk mitigation
- Performing data validation activities on the integrated data
- Conclusion

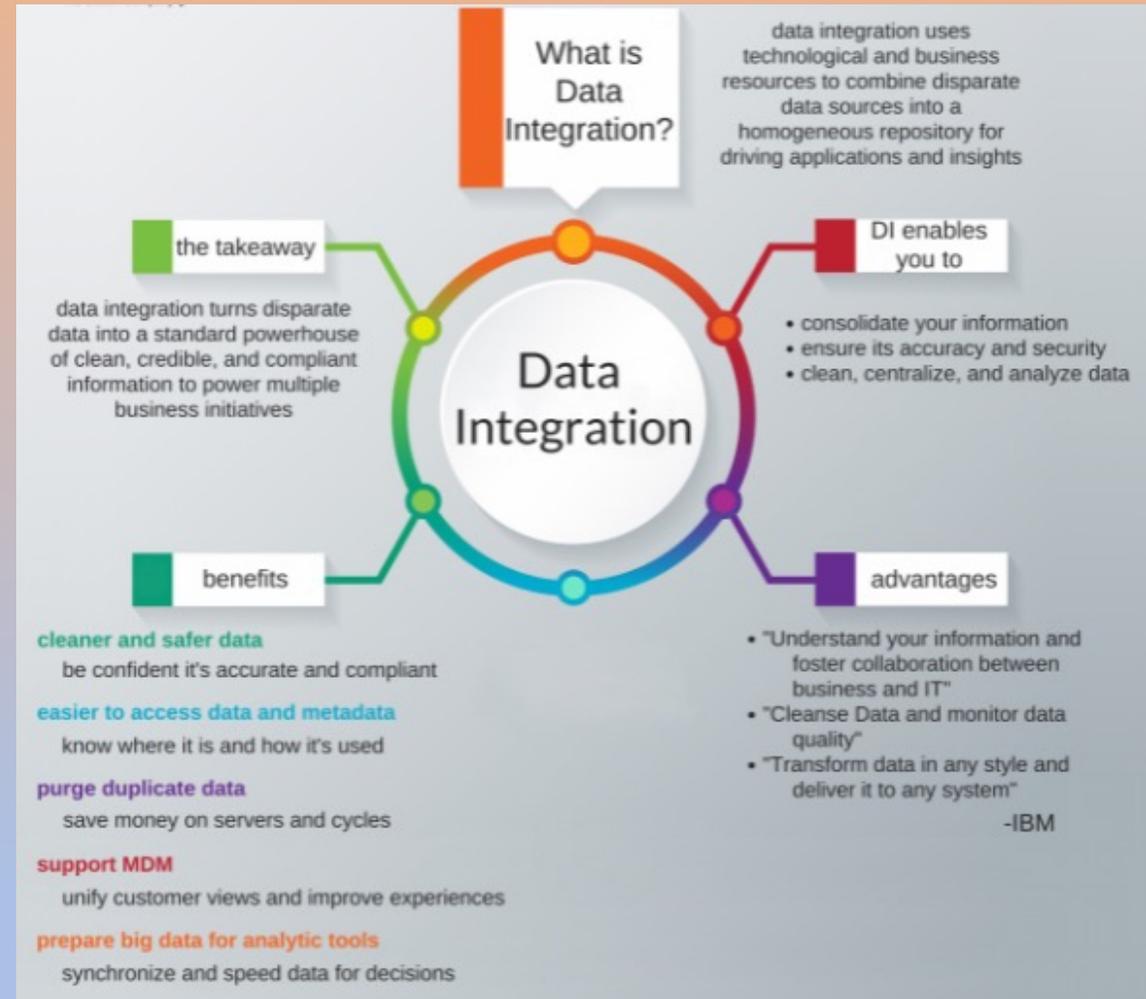


Introduction

Defining Integration: *The process of merging data from several sources into a single, cohesive perspective*

- Data integration has played an imperative role in the evolution of data management over the years. For the longest time, instream data collection on CRF was done by manual data entry and the external data came in as transfers.
- Industry has made a pragmatic shift towards data sciences/analytics and data integration had been a constitutive part of this shift. ***The need of the hour is to have a well-defined strategy on handling data coming instream into the database from multiple external sources via integration.***
- Data Integration is pivotal as it consolidates data in an organized manner within the database which accelerates efficient research analysis.
- Let's discuss about the critical points encircling the **“What, Why and How”** of handling integrated data that DM team needs to factor in.

Introduction



Defining Data Elements for integration

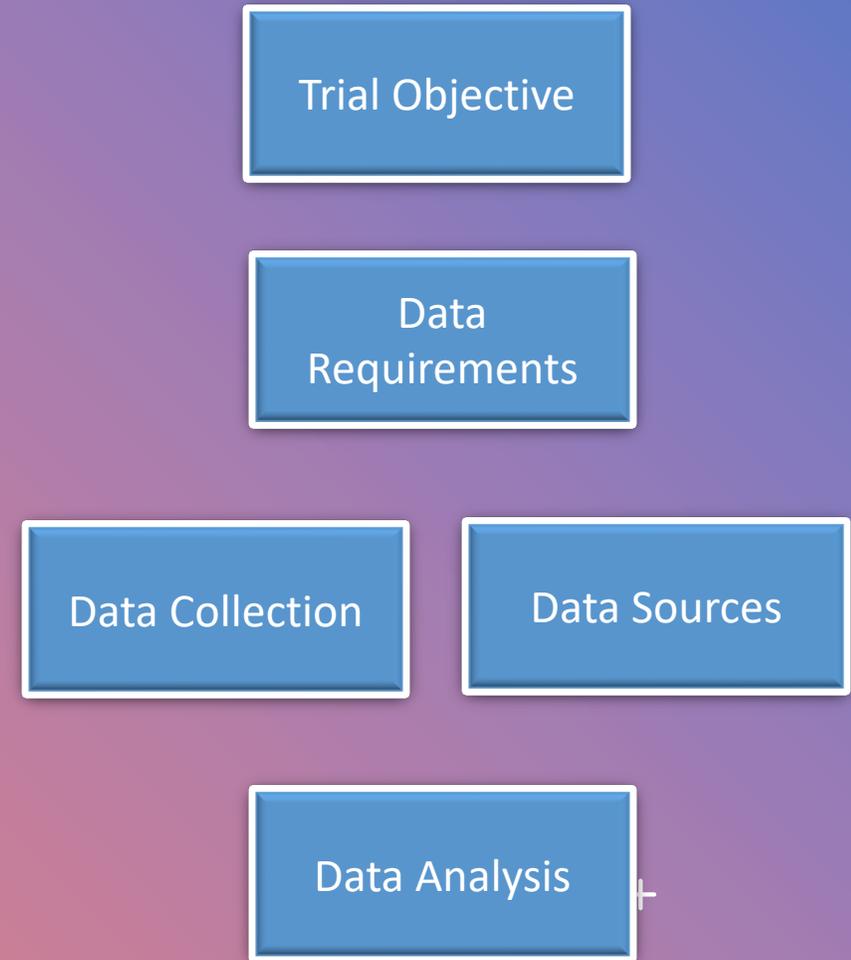
Defining data elements starts from the inception of the protocol

- Weigh out the “*pros and con’s*” , “*why and why not’s*” of what needs to be entered manually within the EDC tool and what can come in as integration.
- Factor’s to be considered:
 - *Type and Size of the clinical trial*
 - *No. of sites and regions involved*
 - *Type of assessments*
- In other words, we need to start working on “Data Modelling”

Defining Data Elements for Integration – *Data Model*



- Data model is a conceptual representation of ***data objects, their relationships and the rules*** that govern them.
- ***Organizing data description, data semantics, and data consistency*** requirements. Instead of focusing on what operations will be performed on data, the data model focuses on what data is required and how it should be organized.
- It begins by defining - ***how data will flow into and out of the database while building a database***. To meet the data flow needs, data model is used to specify the characteristics of the data formats, structures, and database handling functions.
- A comprehensive and optimized data model aids in the creation of a logical, streamlined database that ***eliminates redundancy, decreases storage needs, and allows for quick retrieval***.



Types of Data Integration

- Real-time analytics on streaming data is changing the way clinical trials are competing
- Some of the most common recent integration solutions seen:
 - *eCOA – electronic clinical outcome assessments*
 - *Safety data Integration*
 - *IRT integration and the likes (Randomization and kit assignment)*
 - *Lab data integration (LOINC Codes)*
 - *Integration from wearables (BOYD)*
 - *Electronic Health Records*
- Although stream analytics is still in its infancy, it's already proving to be a powerful tool for **“Monitoring”** , **“Analysing”** and **“Reacting to real-time events”** with a focus on quick detection and correction to ensure quality.
- ✓ A Data Manager needs to be more abreast with the external data , be mindful of what can be integrated and what cannot be integrated. The methodology and technology behind the integration (batch load or pick and push) what is the level of programming or dependency to be identified. Also how the eCRF build needs to be in sync with the integrated platform.

Mapping of integrated data – “Data Mapping”

- - **Data Mapping:** The process of matching fields from one database to another. It's the first step in making data transfer, data integration, and other data management activities more straightforward.
 - **Data must be homogenized:** Similar data is stored in different ways by different systems. As a result, in order to move and consolidate data for analysis or other tasks, a roadmap is required to ensure that the data arrives at its destination in a timely manner.. For ex , a source system's state field may display “Blood Pressure,” but the destination system may record it as “BP.”

Data integration from multiple sources require data mapping to be successful. Errors in data mapping can result in failure of integration and downstream issues

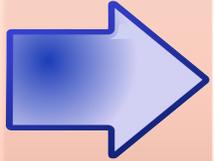
When data is moved from a source to a destination, data mapping bridges the gaps between two systems, or data models, ensuring that the data is accurate and useable.

+

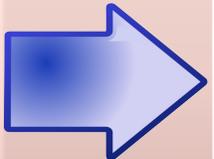
●

○

Mapping of integrated data



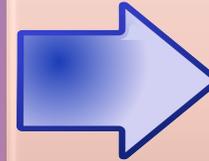
Define - Define the data to be moved, including the tables, fields within each table, and the field's format . The frequency of data transfer is also specified for data integrations.



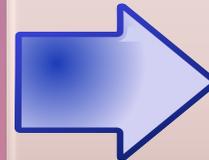
Transformation : The transformation formula or rule is coded if a field requires transformation.



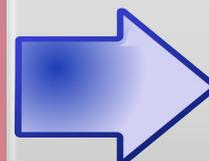
Push to Prod : Schedule an integration go-live event once the data transformation is confirmed to be performing as expected.



Map the Data : Match source fields to destination fields



User Acceptance Test: Run the transfer using a test system and sample data from the source to observe how it works and make any necessary modifications.



Maintenance and Updating : The data map is a live entity for continuing data integration that will require updates and adjustments as new data sources are added,.

Data integration – Questions to ask

It's critical to evaluate integration requirements at various stages of EDC deployment, especially considering the limitations in EDC functionality. Self-assessment questions for EDC preparation include the following:

Type of integration necessary during the early stages of EDC build

Type of integration needed to support conducting >25%, 50% etc. clinical studies and Total no. of data collection sources needed ?

Understand the options of in-house integration systems and prioritize it to achieve a certain degree of standardization at the same time evaluating the challenges of integrating it with the EDC as per the protocol requirement.

Does the value of collecting data electronically from the sites outweigh limited integration capabilities in the short terms or larger scheme of events?

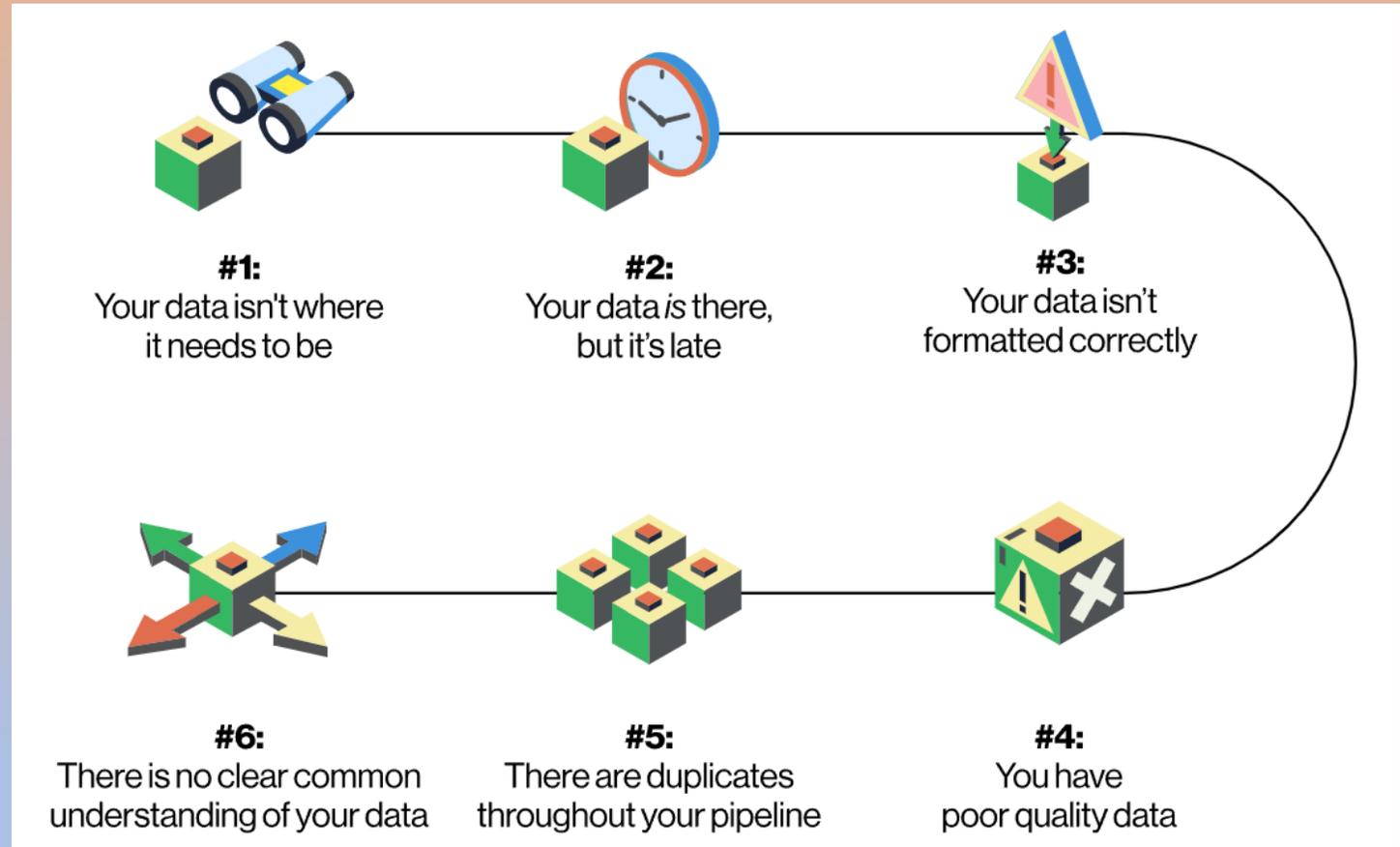
Will the existing legacy tools or EDC tools be phased out ?negating the need for integration ?

Is the integrated data in compliance with standards like SDTM mapping etc.?

Data integration – Evaluating Capabilities

- Give instances of other systems with which your system is integrated and assess how complex/sophisticated these interfaces are.
- Show and describe your EDC architecture's integration layer or integration features: ***what's built in to support integrations?***
- Examine the EDC product's built-in XML support.
- Take a look at the data structure of the EDC system you're thinking about. There have been instances where data structure has had a negative impact on integration.
- Determine the time and effort required to interface with other systems based on the EDC vendor's previous work.
- ***Review the data integration SOPs developed by the vendor.***

Data integration - Challenges



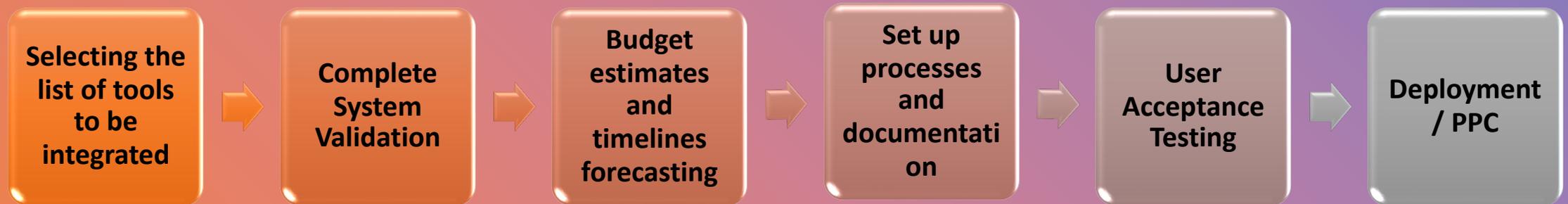
Data integration - Challenges

- Data Mapping issues between various systems with the EDC. The importance of comprehending metadata and underlying processes.
- Nature of the Data structure. Logically mapping heterogeneous data structure between source and target systems.
- Environment that is regulated
- Writing and maintaining integration programs takes more time and effort than introducing new EDC features.
- Understanding how source and target systems behave.
- Data formatting and validation checks are required prior to data loading
- The ability of an organization to assure data consistency may be hampered by possible information delivery latency
- Bringing diverse tools from several disciplines together to foster collaboration among team members.
- A large amount of data is associated with it which is coming from multi centers
- Changes after integration incur unanticipated costs. Issues with the infrastructure

Cost and effort of data integration from multiple sources

Cost - In any integration set up, cost is a major consideration. It is critical to precisely analyze the cost of data integration solution and maintenance before proceeding with data integration. An overly optimistic budget, especially in the situation of a budget gap and achieving more with less, can lead to unrealistic estimation.

Effort



+

Risk mitigation

The core risk areas that make up operations' landscape are constantly changing. One of them is integration, which is a regular source of deal-breaker problems that are challenging when implementing operational improvements.

System

- Include technical owners at all stages of the project and provide all stakeholders with clear and comprehensive information about the process, information, events, and requirements.
- Create unambiguous, stable test environments that replicate all of the process change's involved systems.
- Create complete integration testing scenarios that have been validated and authorized by all technical stakeholders.
- Run load testing on scenarios that are appropriate.

People

- People should be involved early in the process.
- Explain what's changing in a straightforward, simple, and transparent manner. If you're not sure what will change, be specific about the regions that will be affected.
- To ensure knowledge transfer and drive change selling, invest in building several layers of participation.
- Because training is never enough, make sure to develop appropriate training programs.

Process

- Check for consistency both upstream and downstream
- In each action, verify that all information is available at the appropriate time and that the source of that information is accurate and accessible.
- Ascertain that you have monitoring tools and metrics at the activity and process levels that can tell you if there are any exceptions and if performance is meeting your KPIs

Performing data validation activities on the integrated data

Validation to be performed by the Vendor before integrating data in the EDC

- Identify duplicates
- Verify and validate all fields to be integrated
- Validate the data structure before integration
- ***The DBL activities and documents related to DBL for Integrated data needs to be followed just as it is done for EDC data.***
- ***The locking is applicable at the vendor side and also the EDC side.***
- ***The checklist and timelines have to be completed and must be clear at the vendor side and need to be fed into the DM master timelines.***

Validation to be performed by the DM within EDC after integration

- Identify missing/incomplete data
- Timing/frequency of data validation
- In cases like ePRO integration where real time data is entered by the subjects the data validation comes with its limitations

Conclusion



- To summarize, numerous data integration can clearly simplify electronic information transmission while also decreasing the costs and complexities of developing interfaces across diverse systems, resulting in significant patient insights.
- Data Manager's role has evolved with the evolution of technology in the industry. DM is involved in the decision making, rolling out requirements and having a holistic outlook towards end-to-end data flow.
- Getting more involved in understanding the nuances of the data collection from multiple sources also is a transition of data management to data sciences.

