How Database set up and validation impacts “*Data visualization*”

DM’s need of Data Visualization

From DM’s perspective ***“Data visualization”*** comes into picture while performing :

1. Data Transfer
2. Regulatory Submissions
3. Data Trends and Risks (RBM)

***Futuristic and Holistic approach in order to understand Data Visualization**.

Why is data visualization important?

- Using Tables/Charts, graphs/Figures and Listings to visualize large amounts of complex data is easier than pouring over spreadsheets or reports.

- Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making needed adjustments.

- Data visualization can also:
  - Identify areas that need attention or improvement.
  - Clarify which factors influence a certain pattern or trend.
  - Help you understand which Data to place where.

DM approach to reduce the Data Issues

Data Formats

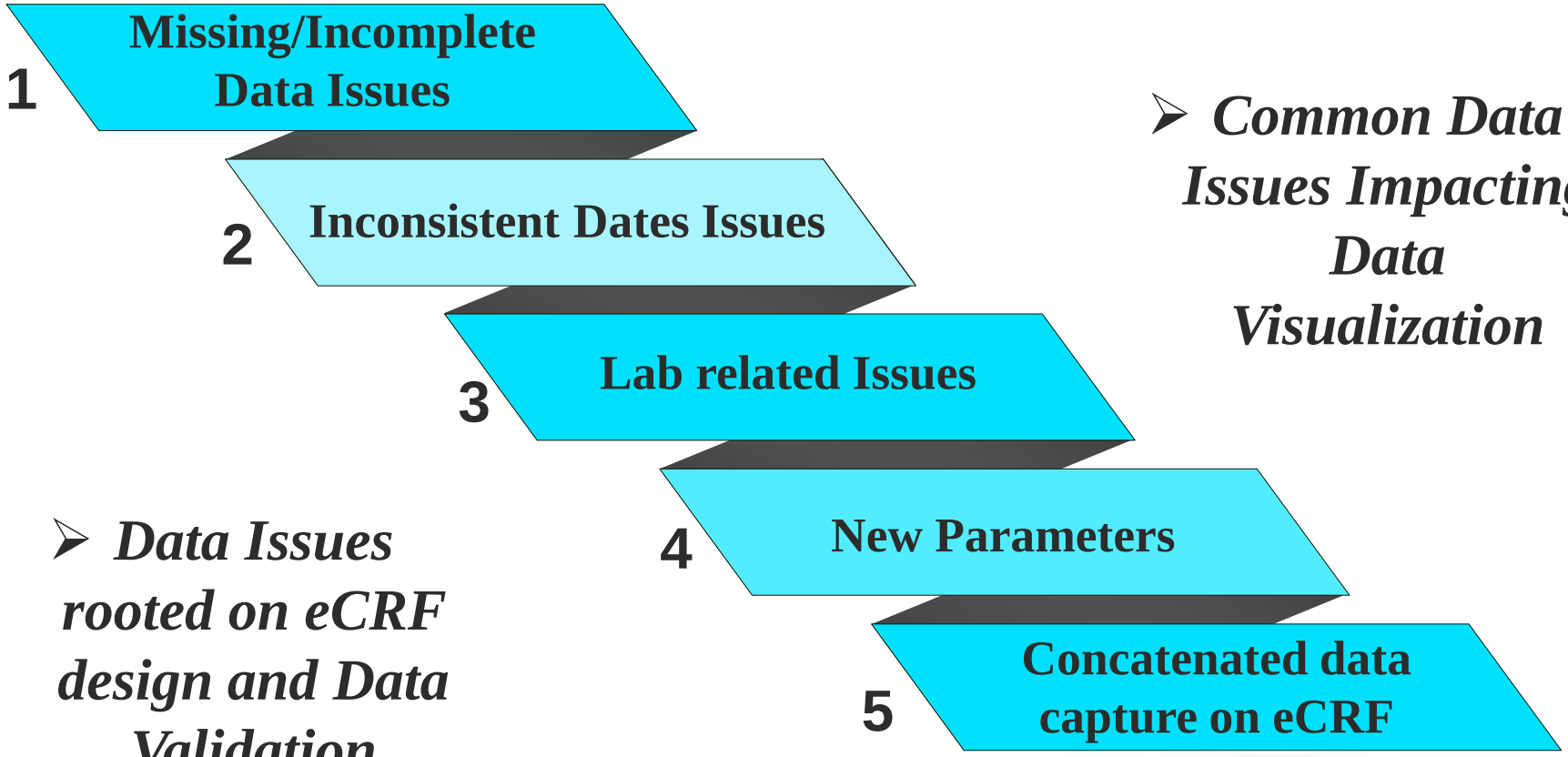
Completeness of Data

Efficient CRF Design

Standardization of Data

Tables Figures Listings

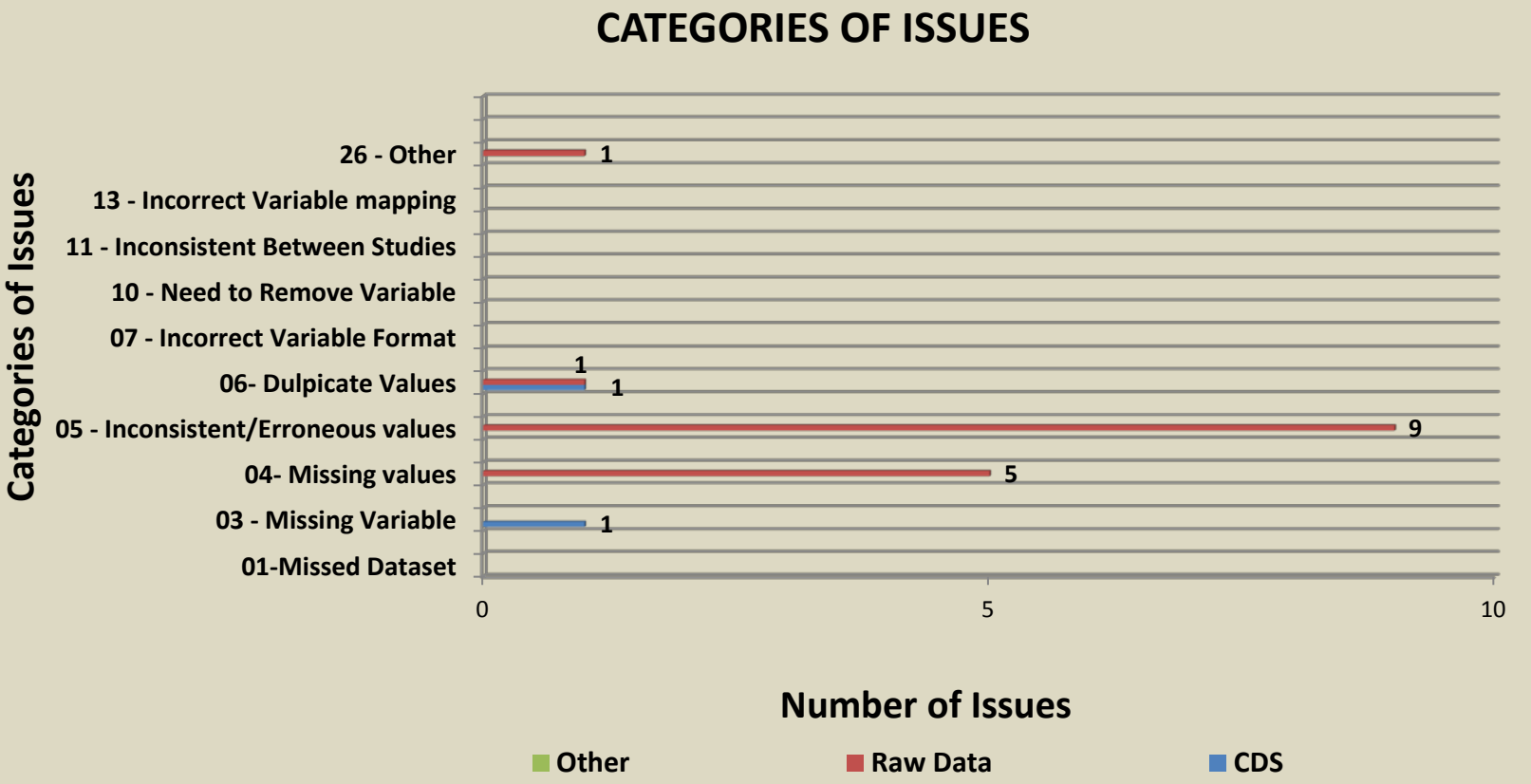
Various Data issues



- ***Common Data Issues Impacting Data Visualization***

- ***Data Issues rooted on eCRF design and Data Validation***

Data Issue Trends

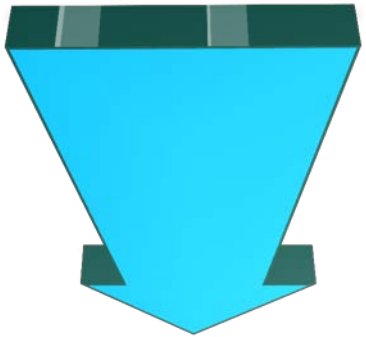


DM approach to reduce the Data Issues

Efficient eCRF Design

- Structured Data Collection
- Data Formats
- List of Complete / Partial Dates
- Dependent/ Corresponding forms
- Standardization of Data

Database Build



Raw Datasets

Optimal Edit Checks

- Dynamics / Custom Functions
- System Checks for Sites
- System Checks for DM
- Checks on Dates/Lab Data
- Restricting Non-conformant Data

Data Collection Schema

➢ Requirement Analysis

The initial part of the requirement analysis step involves the development of a detailed plan for data collection, analysis, and reporting.

- *What data should be collected?*
- *What are the key questions to be answered?*
- *How will the data be analyzed?*
- *What reports will be used?*
- *What formats will be used?*
- *How often will a new report be generated?*

Data Collection Schema

- **Logical design**
  - *A conceptual data model diagram, that shows all the data and their relationships*
  - *Ultimately, design of the study database also may be restricted by the need to be compatible with other databases and some other internal guidelines*

- **Physical design**
  - *This step involves selection of indexes , partitioning, and the clustering of data. The purpose of physical design is to optimize performance as closely as possible.*

Coding Schema

- Coding is the process of converting questionnaire data into meaningful categories to facilitate analysis. You need to think about your coding scheme at the beginning of the study and, wherever possible, build it into your CRF.

- Coding schema should be carefully documented and followed precisely. Related study documentation must contain the following information:

  - Database space to be reserved for each attribute and format

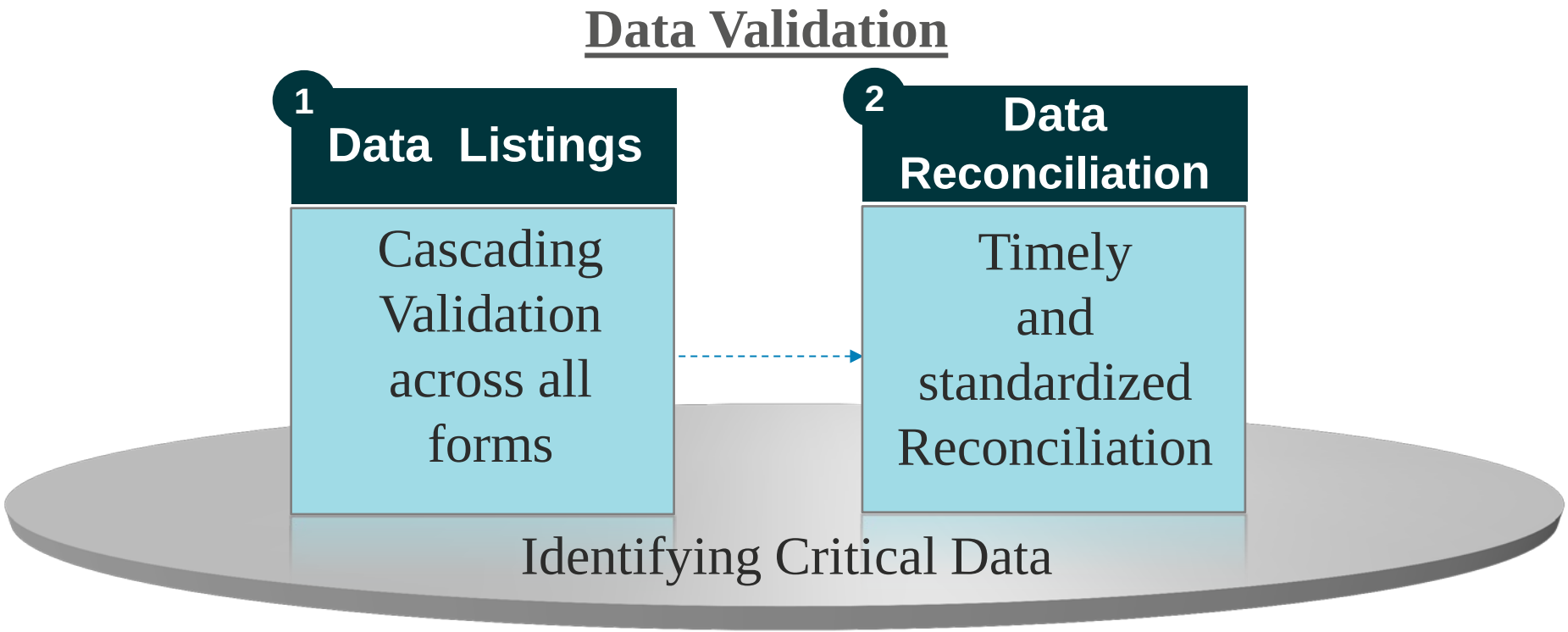
  - Full description (decoding) of every code used in the study

  - Mapping of elements to items in a CRF . This information belongs to database specifications if the data is collected via CRFs, or data transfer specifications if data is collected via IVRS or transferred from outside vendor.

Optimal Approach towards updating DB

- To strategically allow *“Not Done”*, *“Not Applicable”* values for data not available from sites.
- To clearly Map the study specific parameters rather than using generic values.
- In Database set up allow all the acceptable units and restricts the values which don’t fall under the study specifics.
- To have checks that cross checks EVERY dependent data variable and CRF.

DM approach to reduce the Data Issues



Listings and Reconciliations

- Edit checks and Dynamics have limitations to restrict or validate data.
- The above gaps are covered through Listings review and Reconciliation across forms and external vendor.
- To improvise SAS listings based on Data trends / data issues spotted during first dry run
- To update Lab reconciliation pertaining to Data issues on lab modules.

Smarter approach in listing specification

- Data listing programmed in SAS and output in MS Excel where in all patient data from different study domains is not only presented together but is also interleaved chronologically by date. This kind of listing offers a more powerful tool for data review.

- Some of the advantages compared to the traditional approach are:
  - *Don't need to sift through different output files to get an assessment of patients' status.*
  - *Can get a full picture of what information is collected by study sites at each of the patient visits.*
  - *Rows of data from different domains (say AE and Study Drug Exposure) can be put together using data filter feature of MS Excel to identify relational issues in the data.*

- This kind of all-in-one listing in MS Excel has the potential to reduce the number of traditional data review and patient profile outputs. This translates to savings in time, effort and reduce the data issues.

Jack of All Listings

One can review the various study domains or data fields collected at a certain visit against assessments that were required in the schedule of observations and procedures (also called Table of Events) in the study protocol.

	A	B	C	D	E	F	G	H	I
1	Patient	Dataset/Domain	Field Label	Date	Visit Label	Visit	_1	_2	_3
2	1005	Demographics	Date of Birth	4-Apr-53	BASELINE	1	Gender: Male	Race: White, Non-Hispanic and Non-Latino	Race, other:
3	1005	Pre-Treatment Diagnoses Signs & Symptoms	Start Date	7-Nov-08	BASELINE	1	SOC Text: Vascular disorders	PT Text: Arteriosclerosis	Continuing: Check
4	1005	Cancer History	Date of Primary Diagnosis	18-Nov-08	BASELINE	1	Stage at Primary Diagnosis: IV	Anatomic Site of Primary Diagnosis: OrganY	Other, specify:
5	1005	Prior and Concomitant Medications	Start Date	2-Dec-08	MEDS	901	Preferred Drug Name: LIDOCAINE	Continuing:	Indication: ANAESTESIA FOR BRONCHOSCOPY
6	1005	Prior and Concomitant Medications	Stop Date	2-Dec-08	MEDS	901	Preferred Drug Name: LIDOCAINE	Continuing:	Indication: ANAESTESIA FOR BRONCHOSCOPY
7	1005	Pre-Treatment Diagnoses Signs & Symptoms	Start Date	2-Dec-08	BASELINE	1	SOC Text: Investigations	PT Text: Bronchoscopy	Continuing:
8	1005	Subject Enrollment Notification	Informed Consent Date	23-Dec-08	BASELINE	1	Patient Initials: UUU	Which treatment regimen assigned?: ArmX	Informed Consent Date: 23DEC2008
9	1005	ECG	Date of ECG	25-Dec-08	BASELINE	1	Date of ECG: 25DEC2008		
10	1005	Lesion Identification & Evaluation	Date of Scan, PE, etc.	25-Dec-08	BASELINE	1	Sum of Longest Diameters (cm):	Lesion Type: NTLE	Lesion Number: 1
11	1005	Lesion Identification & Evaluation	Date of Scan, PE, etc.	25-Dec-08	BASELINE	1	Sum of Longest Diameters (cm): 26.1	Lesion Type: TARG	Lesion Number: 1
12	1005	Lesion Identification & Evaluation	Date of Scan, PE, etc.	25-Dec-08	BASELINE	1	Sum of Longest Diameters (cm): 26.1	Lesion Type: TARG	Lesion Number: 2
13	1005	Height, Weight, BSA Assessment	Date of BSA Assessment	26-Dec-08	BASELINE	1	Date of Height, Weight, BSA Assessment: 2008	Height (cm): 187	Height (cm)-EXC:
14	1005	ECOG Zubrod Performance Status	Date of Assessment	26-Dec-08	BASELINE	1	ECOG Score: 0	Date of Assessment: 26DEC2008	
15	1005	Physician Assessment Sensory Neuropathy	Date of Assessment	26-Dec-08	BASELINE	1	No Sensory Neuropathy: Check	PAPN Score:	Date of Assessment: 26DEC2008
16	1005	Physical Exam	Date of Physical Exam	26-Dec-08	BASELINE	1	Date of Physical Exam: 26DEC2008		
17	1005	Pat. Self Asmnt. Peripheral Neuropathy	Date of Assessment	26-Dec-08	BASELINE	1	FACT-Taxane Question Text: I AM BOTHERED BY THE WAY MY HANDS OR NAILS LOOK.	FACT-Taxane Question Response: Not at All	Date of Assessment: 26DEC2008

Example : AE and Study Drug Administration

Data is filtered to keep only Study Drug Dosing and Adverse Events for column B (Dataset/Domain). Below snapshot shows how the onset of AEs occurs with respect to study drug administration. Interleaving of AEs with Dosing as in above image might give more perspective than if Dosing and AE records are presented in separate outputs or separate sections within patient profile.

	A	B	C	D	E	F	G	H	I	J
1	Patient	Dataset/Domain	Field Lab	Da	Visit Labe	V	_1	_2	_3	_4
34	1007	Study Drug Dosing	Dose Date	29-Dec-08	C1D1	2	Dose Date: 29DEC2008	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 869
35	1007	Study Drug Dosing	Dose Date	29-Dec-08	C1D1	2	Dose Date: 29DEC2008	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
57	1007	Study Drug Dosing	Dose Date	5-Jan-09	C1D8	3	Dose Date: 05JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2):	Total Calculated Dose (mg):
58	1007	Adverse Events	Start Date	10-Jan-09	AES	902	SOC Text: Skin and subcutaneous tissue disorders	PT Text: Alopecia	Grade: 2	Outcome (Check one): Ongoing
59	1007	Study Drug Dosing	Dose Date	13-Jan-09	C1D15	4	Dose Date: 13JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
64	1007	Study Drug Dosing	Dose Date	20-Jan-09	C2D1	5	Dose Date: 20JAN2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 900
65	1007	Study Drug Dosing	Dose Date	20-Jan-09	C2D1	5	Dose Date: 20JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
85	1007	Study Drug Dosing	Dose Date	26-Jan-09	C2D8	6	Dose Date: 26JAN2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
86	1007	Study Drug Dosing	Dose Date	2-Feb-09	C2D15	7	Dose Date: 02FEB2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
87	1007	Adverse Events	Start Date	6-Feb-09	AES	902	SOC Text: Investigations	PT Text: Weight increased	Grade: 1	Outcome (Check one): Resolved
100	1007	Adverse Events	Start Date	9-Feb-09	AES	902	SOC Text: Blood and lymphatic system disorders	PT Text: Neutropenia	Grade: 1	Outcome (Check one): Resolved
101	1007	Adverse Events	Stop Date	12-Feb-09	AES	902	SOC Text: Blood and lymphatic system disorders	PT Text: Neutropenia	Grade: 1	Outcome (Check one): Resolved
104	1007	Study Drug Dosing	Dose Date	25-Feb-09	C3D1	8	Dose Date: 25FEB2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 859
105	1007	Study Drug Dosing	Dose Date	25-Feb-09	C3D1	8	Dose Date: 25FEB2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
125	1007	Study Drug Dosing	Dose Date	3-Mar-09	C3D8	9	Dose Date: 03MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
126	1007	Study Drug Dosing	Dose Date	10-Mar-09	C3D15	10	Dose Date: 10MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
147	1007	Study Drug Dosing	Dose Date	17-Mar-09	C4D1	11	Dose Date: 17MAR2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 851
148	1007	Study Drug Dosing	Dose Date	17-Mar-09	C4D1	11	Dose Date: 17MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
152	1007	Study Drug Dosing	Dose Date	24-Mar-09	C4D8	12	Dose Date: 24MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
153	1007	Study Drug Dosing	Dose Date	30-Mar-09	C4D15	13	Dose Date: 30MAR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
169	1007	Study Drug Dosing	Dose Date	7-Apr-09	C5D1	14	Dose Date: 07APR2009	Dosing Drug: Study Drug 2	Dose (mg/m^2):	Total Calculated Dose (mg): 883
170	1007	Study Drug Dosing	Dose Date	7-Apr-09	C5D1	14	Dose Date: 07APR2009	Dosing Drug: Study Drug 1	Dose (mg/m^2): 100	Total Calculated Dose (mg): 200
191	1007	Adverse Events	Start Date	9-Apr-09	AES	902	SOC Text: Nervous system disorders	PT Text: Hypoaesthesia	Grade: 1	Outcome (Check one): Ongoing at time of Death

DM’s collaboration with Biostatistics

- UAT Extract instead of the PROD - Dirty Data for the initial dry run
- DM to review SDTM specification and performing SDTM dataset QC
- Providing a list of New parameters or exception pertaining to a study
- Coding Mismatch Review
- Identifying Data Issue Trends to circle back and make relevant updates to the Database through CPPC

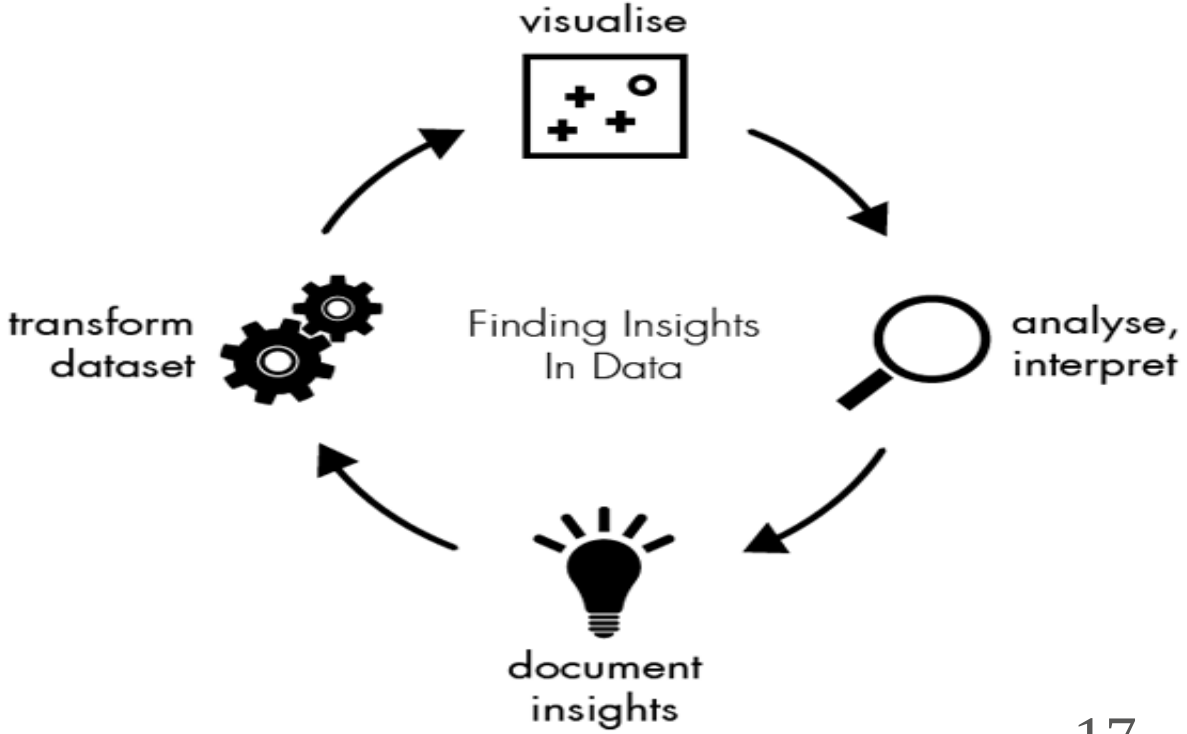
Process flow - Datasets to TFL's

In order to make the finding of insights in data more effective

What does this mean in the context of the data?

Are there any interesting patterns?

What can I see in this image? Is it what I expected?



Thank you

Rashida Rampurawala
Senior Clinical Data Lead
(Study Start Up SME)
Email - Rashida.Rampurawala@iconplc.com

iconplc.com