

Challenges in Validation of ADaM data

Presented at PhUSE SDE
April 18, 2013

Topics

- What is data validation?
- Historical drivers for validation requirements
- Challenges with ADaM
 - Flexibility in structure and content
 - Metadata
 - Traceability
- What are users expectations?

What is a data validation?

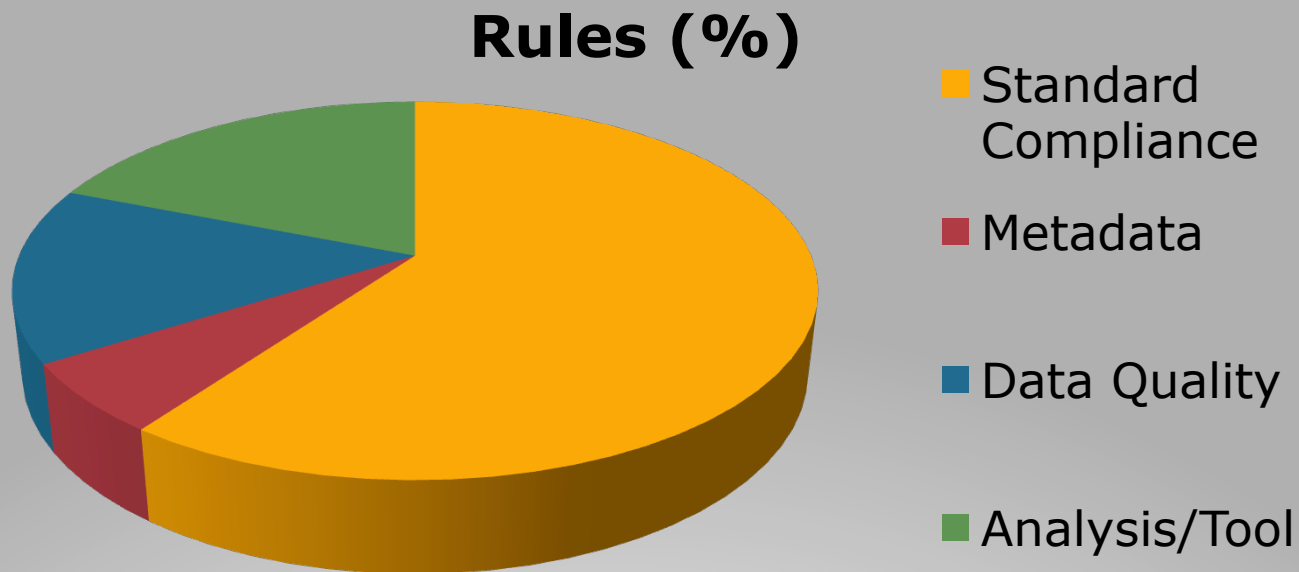
- A process to ensure that data is
 - compliant with standards
 - well described by metadata
 - of high quality
 - fit for intended use (analysis/tools)
- Standards compliance is a foundation for standard analysis and tools

FDA's Definition of Data Quality

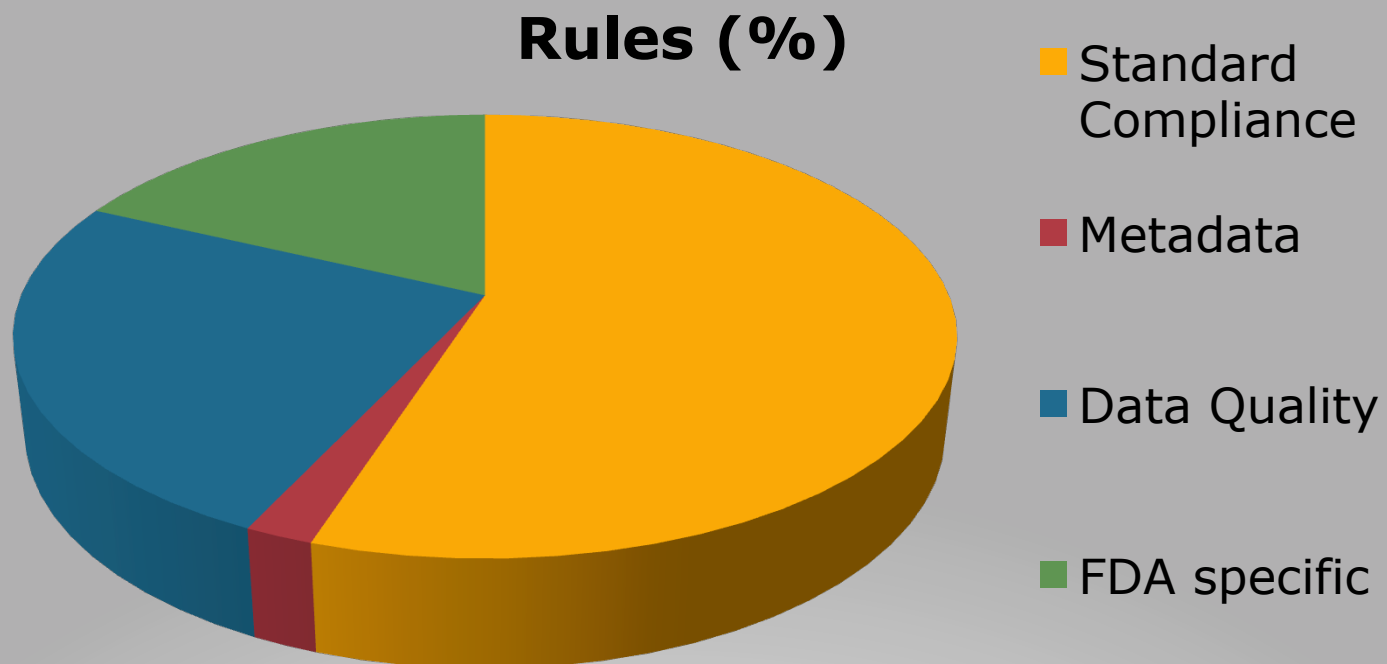
- High Quality Data is both *Compliant* and *Useful*
- *Compliant* means the data confirms to applicable data standards
- *Useful* means the ability of data to support the intended use

Validation is driven by intended use

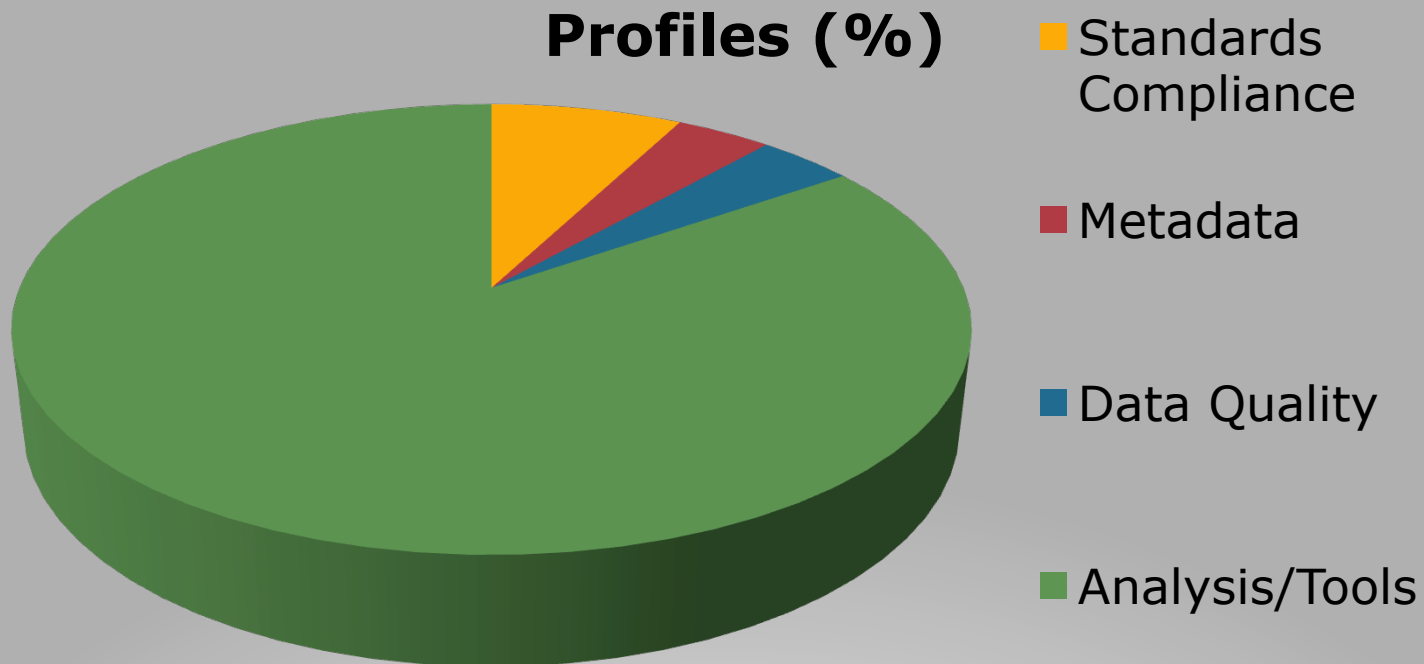
- First SDTM validation rules were driven by Janus data warehouse and WebSDM analysis requirements



OpenCDISC v1.3 rules for SDTM 3.1.2

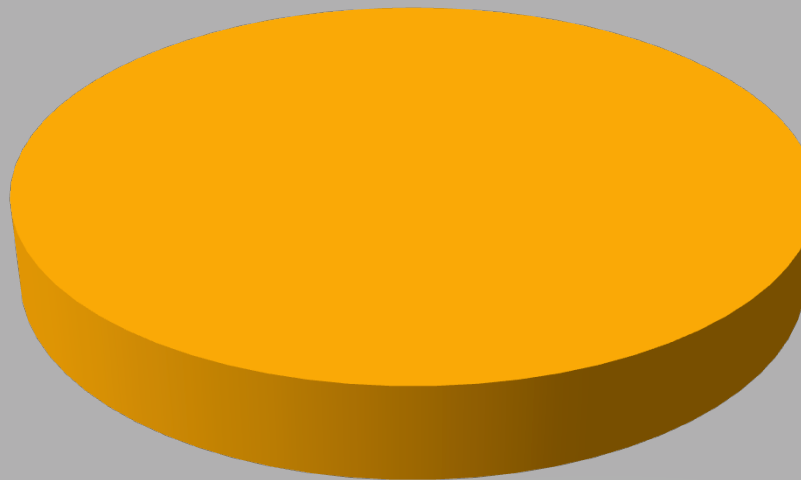


FDA DataFit profiles for SDTM 3.1.2



Current ADaM Validation Rules

Rules (%)



■ Standard Compliance

- Do we need validation rules for metadata, data quality, and analysis/tools?

Current analysis/reporting/storage tools

- SDTM
 - Standard SAS programs
 - JReview, JMP Clinical, Emperica, etc.
 - Janus, SDTM+/- data warehouses
 - ETL tools: SAS, Entimo, etc.
- ADaM
 - ???

Rigid vs. Flexible Structure

- Majority of SDTM tools produce simple standard analysis with assumption that SDTM data structure and content is the same
 - E.g., AE - ARM (ACTARM) and AEDECOD only
 - E.g., LB, VS, DM reports are also relatively simple
- Standard-based analysis/review tools are dependent on rigid structure and pre-defined content

Rigid vs. Flexible Structure

- Problems occur in cases where are not enough standardization
 - E.g., Disposition data: DSDECOD, DSCAT, DSSCAT.
 - Only Reason for Non Completion CT for DSDECOD
 - No CT for DSCAT and DSSCAT
 - Data Collection can be vary across studies
 - It's very risky to produce outputs based on DSDECOD variable only

What About “Standard” ADaM?

- The only “rigid” dataset is ADSL. As a result it is the only ADaM dataset which is widely used by tools
- ADaM BDS structure is highly flexible, so how do you create standard tools?
- Content of ADaM BDS datasets cannot simple be identified based on dataset names

Traceability

- How to ensure traceability?
- Is Define.xml mature enough to support automatic verification of traceability to source SDTM data?

Metadata

- If structure and content is not standardized, usage of metadata is critical for validation
- New features of Define.xml 2.0 specifically addresses ADaM needs and are key to better validation
- See Lex's presentation

Discussion/Questions

Max Kanevsky

mkanevsky@pinnacle21.net

Sergiy Sirichenko

ssirichenko@pinnacle21.net