Raising the bar for data standardization:

How to ensure your submission data supports automated review process at FDA and PMDA?

Sergiy Sirichenko

PhUSE SDE, Foster City, CA

February 22, 2018

PINNACLE 21

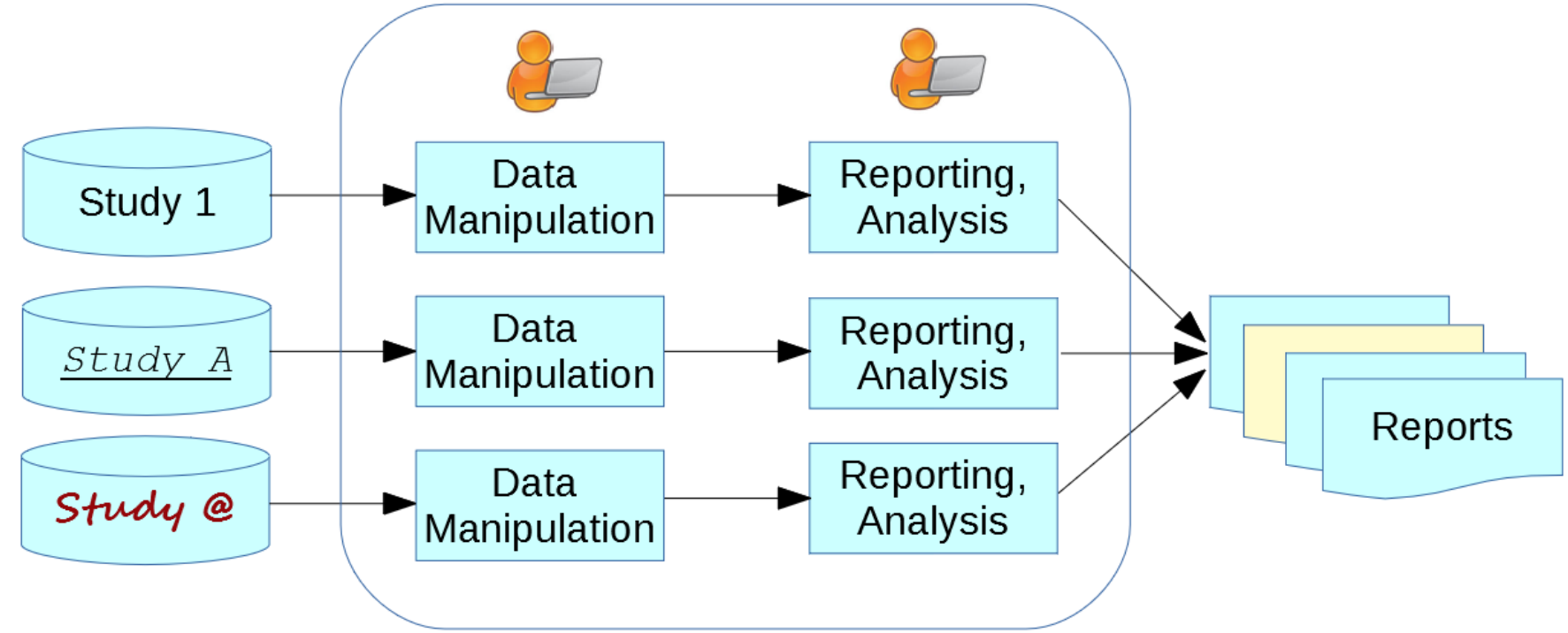
CDISC standards and their adaption

- 1997 – CDISC started as a Volunteer group
- 2005 – SDTM IG 3.1.1, Define-XML 1.0
- 2008 – OpenCDISC
- 2009 – SDTM IG 3.1.2, ADaM IG 1.0
- 2011 – SEND IG 3.0
- 2016 – Required by FDA for new studies
- 2016 (2022) – Required by PMDA

Benefits of standards

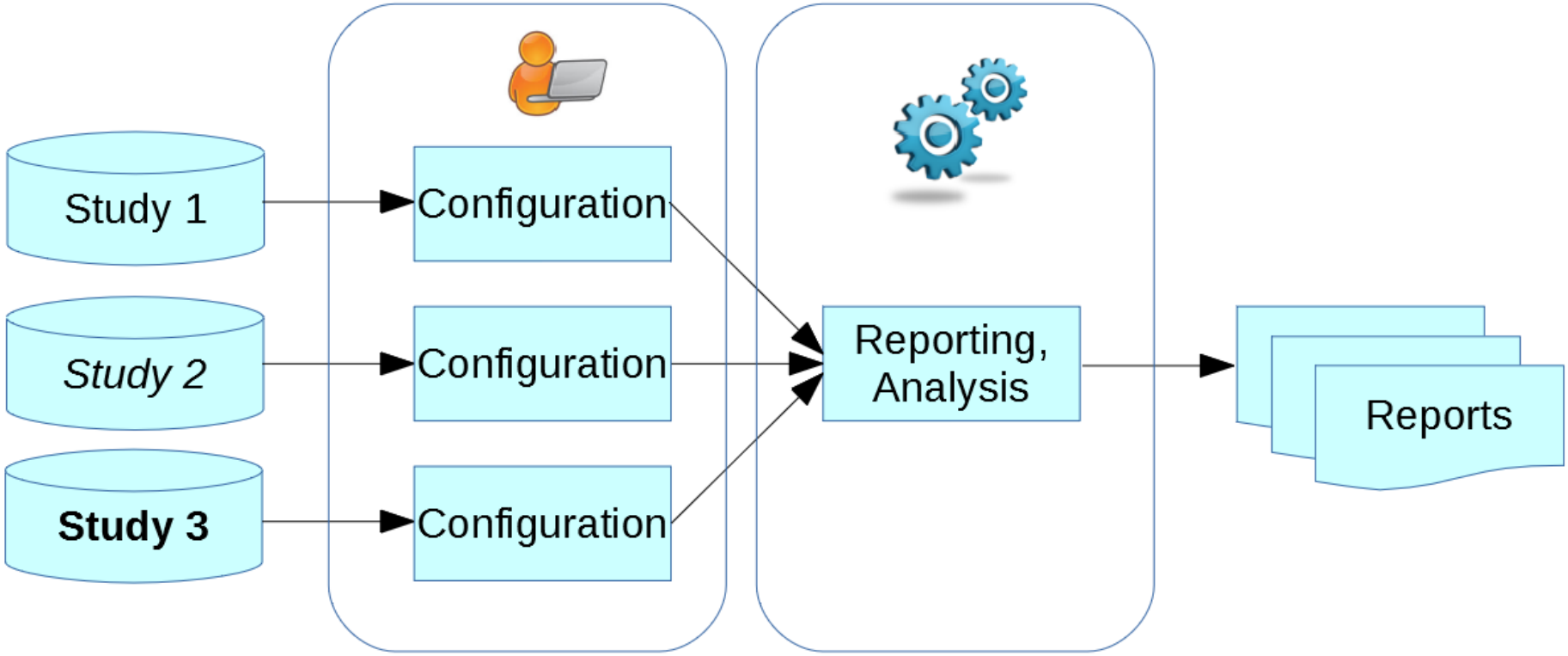
- Similarity across studies
- Ease of integration
- Re-use of code
- Standardized tools
- Automation
- …

Manual programming



- Input data is not standardized
- Data manipulation before reporting and analysis

Standardized tools

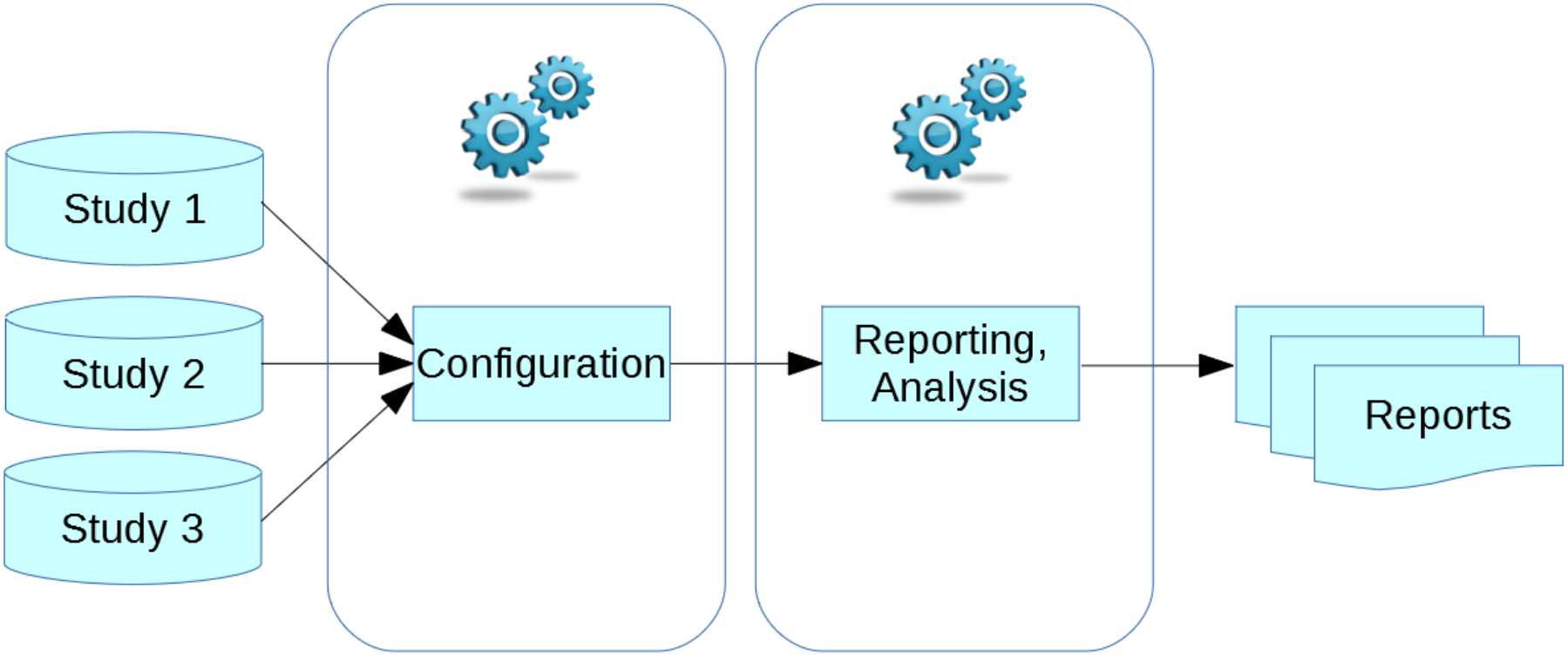


- Input data is partially standardized
- Manual configuration may be needed

Examples

- MAED
  - Subject, Treatment, Adverse Event term
  - One record per AE
- Liver Function Panel
  - Lab results, normal ranges
  - Default use of CDISC, but can be adjusted for non-standard structure and terms (SGOT instead of ASP)

Automated process



- Automated input for standardized tools
- High requirements for standardization

Automation relies on standards

- Location of files
- File names
- Structure
- Control Terminology
- Expected content
- Data quality

Example of wrong Normal Ranges

- Expected

LBTEST	LBSTRESN	LBSTNRHI	LBSTNRC
ALT	42	42	
AST	5	20	
BILI	28	38	

- Incorrect

LBTEST	LBSTRESN	LBSTNRHI	LBSTNRC
ALT	42		<42
AST	5		<20
BILI	28		<38

- Data is present, but it was implemented incorrectly
- Standard reporting tools will not produce results
- Data manipulation is needed to run analysis
  - LBSTNRHI is Num, LBSTNRC is Char variable

FDA definition for data quality

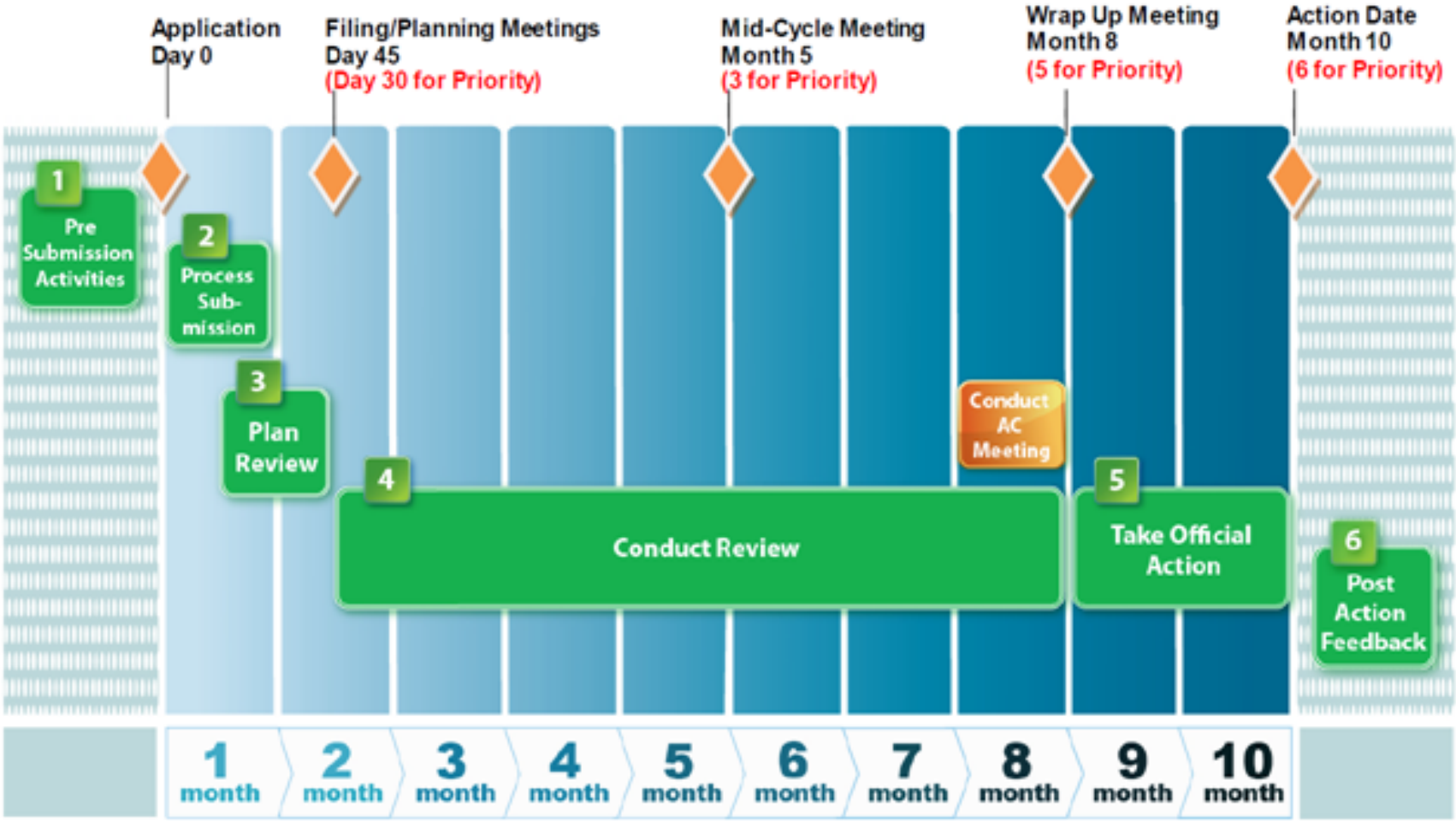
- *"Quality data would therefore be defined as data that sufficiently support conclusions and interpretations equivalent to those derived from error-free data"*

  Janet Woodcock, M.D., 1999

- *"Study data validation helps to ensure that the study data are compliant, useful, and will support meaningful review and analysis"*

  FDA TCG, 2017-10

Overview of the NDA/BLA Review Process and Major Steps for Completing the Review



Note: The timeline for review of NMEs/BLAs under PDUFA V's "Program" Review extends the *Conduct Review* Phase by two months.

See Appendix A for a timeline diagram for PDUFA V.

Increasing demand for data quality

- FDA goal is to make drugs available ASAP
- Manual review process has lower expectations on data quality
  - Often it's more quick for Reviewer to fix or work around issues rather than request fix from Sponsor
  - Data review time is limited
  - Data quality evaluation is difficult
- Automation relies on high quality data
  - Value of standard reporting and analysis for Reviewers increases demand for data quality
  - Standardized data allows early evaluation and enforcement of data quality

PMDA example

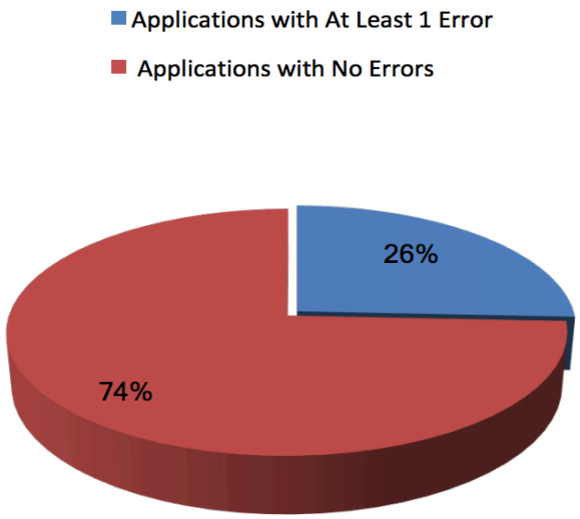
- Submission of data is few weeks before filing the application
- Consultation meetings
- Any Rejection issues (34 rules) must be fixed
- Reconciliation of validation results and Sponsor’s explanations

FDA has just started

- Current 4 FDA Rejection rules are not enough
- *"Trial Summary (TS) dataset must be included in every submission"*
  - Machine-readable metadata about study start date to decide is study must be in CDISC?
  - What about rules to ensure?
    - Presence of this info
    - Use of CDISC CT term for *Study Start Date*
      - *"SSTDTC"* - SDTM
      - *"STSTDTC"* - SEND
    - ISO 8601 format for date

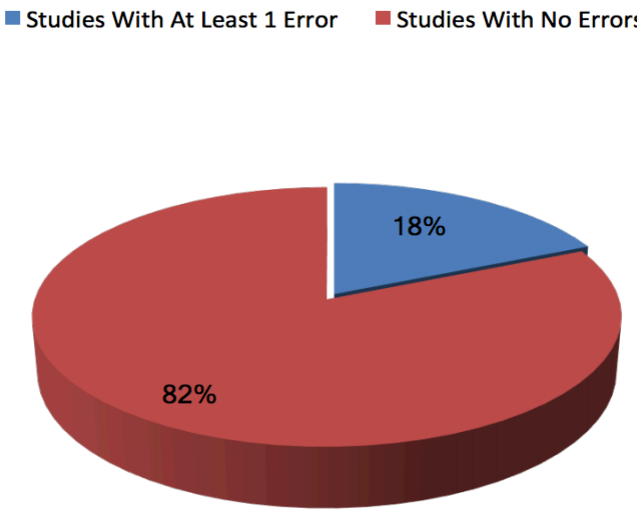
Errors by Application and Study

Potential Rejections



*70 Applications Assessed

Study Errors



*222 Studies Assessed

http://www.phusewiki.org/docs/2017_SDEs/Boston%20SDE/Technical%20Rejection%20Criteria%20for%20Study%20Data%20%E2%80%93%20Preliminary%20Findings.pdf

Automation of validation

- Pinnacle 21 (P21) Community – desktop tool
- P21 Enterprise – web-based application with centralized governance
- FDA DataFit and PMDA P21 Enterprise – P21 Enterprise with additional automation for FDA/PMDA environments based on
  - eCTD structure
  - study metadata

Automation logic

- Automation relies on standards
- Validation is performed according metadata provided by Sponsor (correct or incorrect)
- If study metadata is invalid or missing then some compensation actions are taken
  - For example, the latest versions of standards or dictionaries will be utilized instead

eCTD issues

- Data in incorrect folder is considered as missing
  - *90xxxx\0000\m5\datasets\study-1\tabulations\sdtm* – OK!
  - *90xxxx\0000\m5\datasets\study-1\sdtm* - incorrect
- Files with non-standard or inconsistent names
  - *define-study-1.xml* instead of *define.xml*
  - Define.xml has a reference to "*blankcrf.pdf*", but actual file is "*acrf.pdf*"
- When submitting additional or modified datasets, an update of define.xml for the whole study data is needed

SDTM version in define.xml

- P21C – manual input
- P21E – pre-configured by project manager
- DataFit – from define.xml file

```
def:StandardName="SDTM-IG"
def:StandardVersion="3.2">
```

- Example: study data utilized *SDTM IG 3.1.2*, but define.xml file has reference to *SDTM IG 3.1.3*
  - Unexpected PMDA Reject notification
  - Many false-positive and false-negative messages due to utilization of wrong configuration

https://www.cdisc.org/system/files/all/event/restricted/2017_Japan/Japan%202017_Session%206_Kominami.pdf

Invalid or missing MedDRA version

- DataFit uses define.xml file
  - *<ExternalCodeList Dictionary="MedDRA" Version="19.0"/>*
- If missing then the latest version of MedDRA is utilized
- Incorrectly implemented (e.g., like *Comment* instead of *Dictionary*) MedDRA version means missing info
- MedDRA version is utilized by some reporting tools
- Missing decimal points in version is PMDA Rejection criteria
  - "*19*" instead of "*19.0*"

https://www.cdisc.org/system/files/all/event/restricted/2017_Japan/Japan%202017_Session%206_Oda.pdf

Version of CDISC CT

- Today there is no machine-readable info about CDISC Control Terminology
  - Waiting for Define-XML v2.1
- DataFit utilizes the latest version of CDISC CT
  - 2011 – 104 Codelists
  - 2017 – 730 Codelists
  - Terms for NCI Code = C17988 are used inconsistently across codelists and versions:

Term	2011-07	2017-12
U	6	3
UNK	1	1
UNKNOWN	5	14

  - *U -> UNKNOWN* in SDTM CT 2014-03-28 for 3 non-extensible codelists ENRTPT, STENRF and STRTPT

Other examples Janus CTR

- None of Data Integrity checks are *Rejection* now
  - Missing subjects in DM -> fail to upload study data
- Upload process relies on *def:Class* in define.xml
- Trial Summary data is critical for CTR search

https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM442108.pdf

Summary

- The automation of regulatory review is a game changing event
- Standardized high quality study data is critical for automation
- Expect new Rejection rules and their enforcement from Regulatory Agencies

ssirichenko@pinnacle21.net