

A need for prescriptive define.xml

PhUSE SDE Foster City, CA

Sergiy Sirichenko, Max Kanevsky
December 2, 2015

Abbreviations

- › CT – CDISC Control Terminology
- › EDC – Electronic Data Capture
- › MDR – Metadata Repository
- › SDRG – Study Data Reviewers Guide
- › VLM – Value Level Metadata

Highlights

- › Current problems with define.xml
- › What is metadata?
- › Use it or lose it
- › Benefits of prescriptive approach
- › Implementation considerations

Major problems in define.xml

- › Usage of outdated Define.XML v1.0
- › Lack of expertise
- › Missing study specific metadata (a major purpose of define.xml)

Outdated Define.XML v1.0

- › 10+ years old
- › The first versions are never perfect
- › Value Level metadata cannot be clearly specified
 - › No reference to applied variable
 - › Multiple conditions are not available
- › Lack of structural consistency
 - › Origin and other related attributes
 - › Codelists

Lack of expertise

- › Technical compliance issues
 - › ISO8601 format is defined as a Codelist
 - › Broken XML structure
- › Confusion about correct Origins
 - › *Assigned or Derived?*

Missing study specific metadata

- › Study specific information is crucial for reviewers
- › However in most submission packages it's missing
- › Value of define.xml, SDRG, aCRFs is to explain what is unique in this particular study

Missing codelists

- › Codelists are limited to variables which are assigned to standard CT
- › Commonly missing study specific codelists for variables
 - › Category (--CAT), Subcategory (--SCAT)
 - › EXTRT, ARMCD, --TESTCD/--TEST, QNAM, TPT
 - › RDOMAIN in CO and RELREC domains
 - › XXTOX, ...

Merged codelists

- › Due to confusion between Standard CT Codelist and study Variable Codelist
- › Example:
 - › Define.xml has one codelist (UNIT) assigned to all --DOSU, --VAMTU, --ORRESU, --STRESU variables
 - › This codelist includes all unique terms across all study “units” variables and have 450 items, while for example EXDOSU variable is populated with one “*mg*” term only
 - › A reference to 450-terms codelist is not relevant

What is define.xml Codelist?

- › Define.xml Codelist describes data collection process and should be limited to all terms used for data collection of specific data element (a particular Variable or Value Level)
 - › For example, LBSTRESU, EGORRESU, EXDOSU usually have separate codelists based on the same (UNIT) standard CT
- › If data is collected as a free text, then Codelist may be not applicable
 - › Common example is CMDOSU, CMDOSFRQ, CELOC, etc.

Missing terms in codelist

- › Term is present in data
 - › SD0037 check
 - › Programming error
 - › Due to misspelling, leading space characters, etc.
 - › Due to missing Decoded value for some items
 - › CodeList vs. EnumaretedItem
- › Codelist was populated based on collected data, but some options from CRF page are not included
 - › Example: Only race “*WHITE*” is collected, while 6 options are present on CRF

Why do we have problems with study metadata?

- › There are some obvious challenges
 - › Lack of expertise
 - › Lack of good tools
- › Education and technology can help
- › But what about missing study specific metadata?
 - › Tools and expertise will be not enough

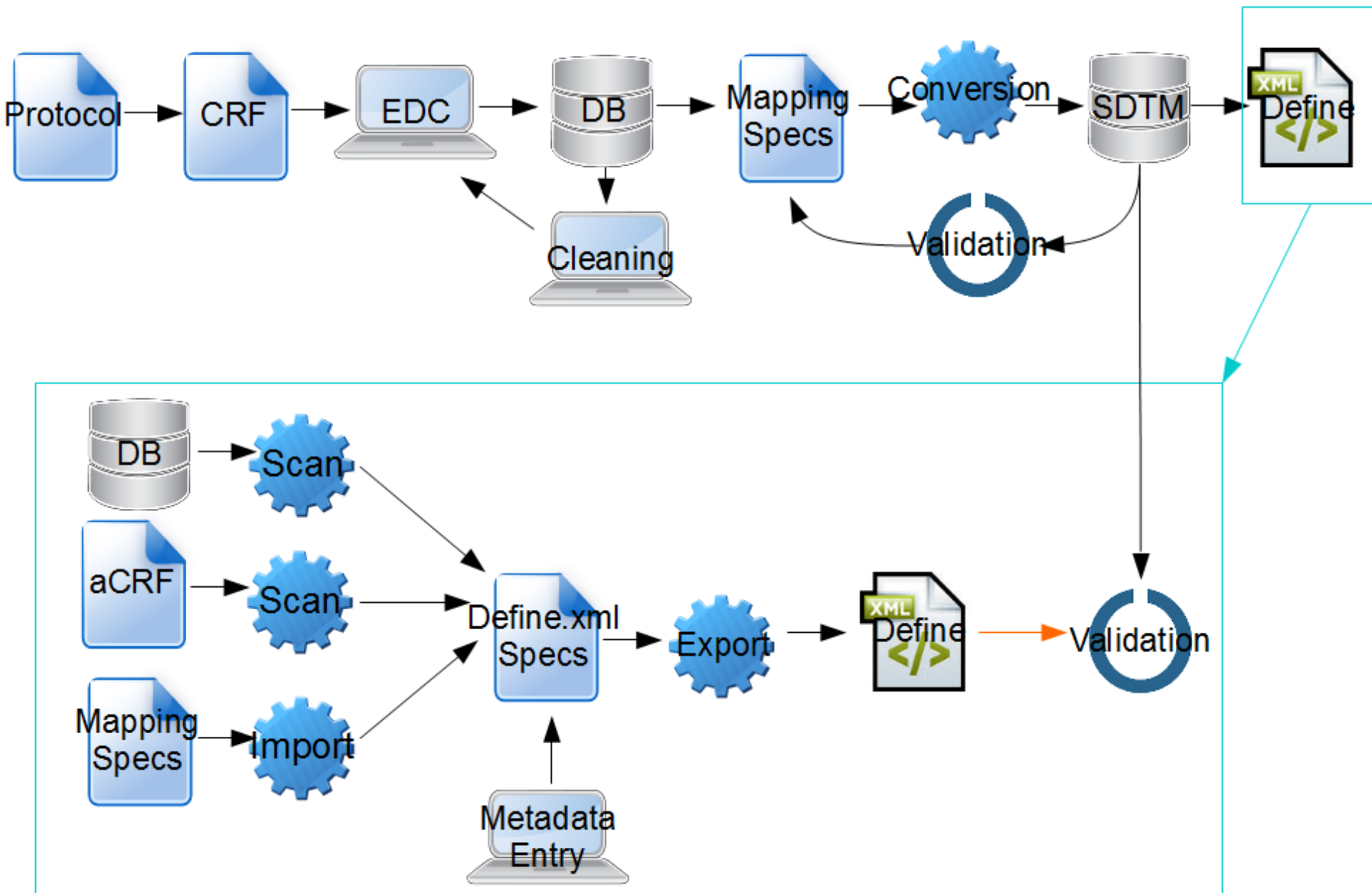
What is Metadata?

- › Common definition: *“data about data”*
- › *“physical data and knowledge-containing info about business, tech processes, and data , used by corporation”*
 - › Physical data: in software and other media
 - › Knowledge: in employees and various media
- › *“**knowledge** about business, data, ...”*

Process for descriptive define.xml

- › Most commonly used approach
- › Based on finalized data
- › For and before regulatory submission
- › Advantages
 - › Can be easily automated
 - › No need to worry until submission

Process for descriptive define.xml



Use it or lose it principle

- › Study metadata is usually stored as
 - › Non-machine readable excel specs with free text entry format
 - › Different specs from multiple systems
 - › People knowledge
- › **Define.xml is not used internally** and created only as a “*required*” file for submission
 - › Considered as extra burden
 - › Focus on populating a file rather than its high quality content

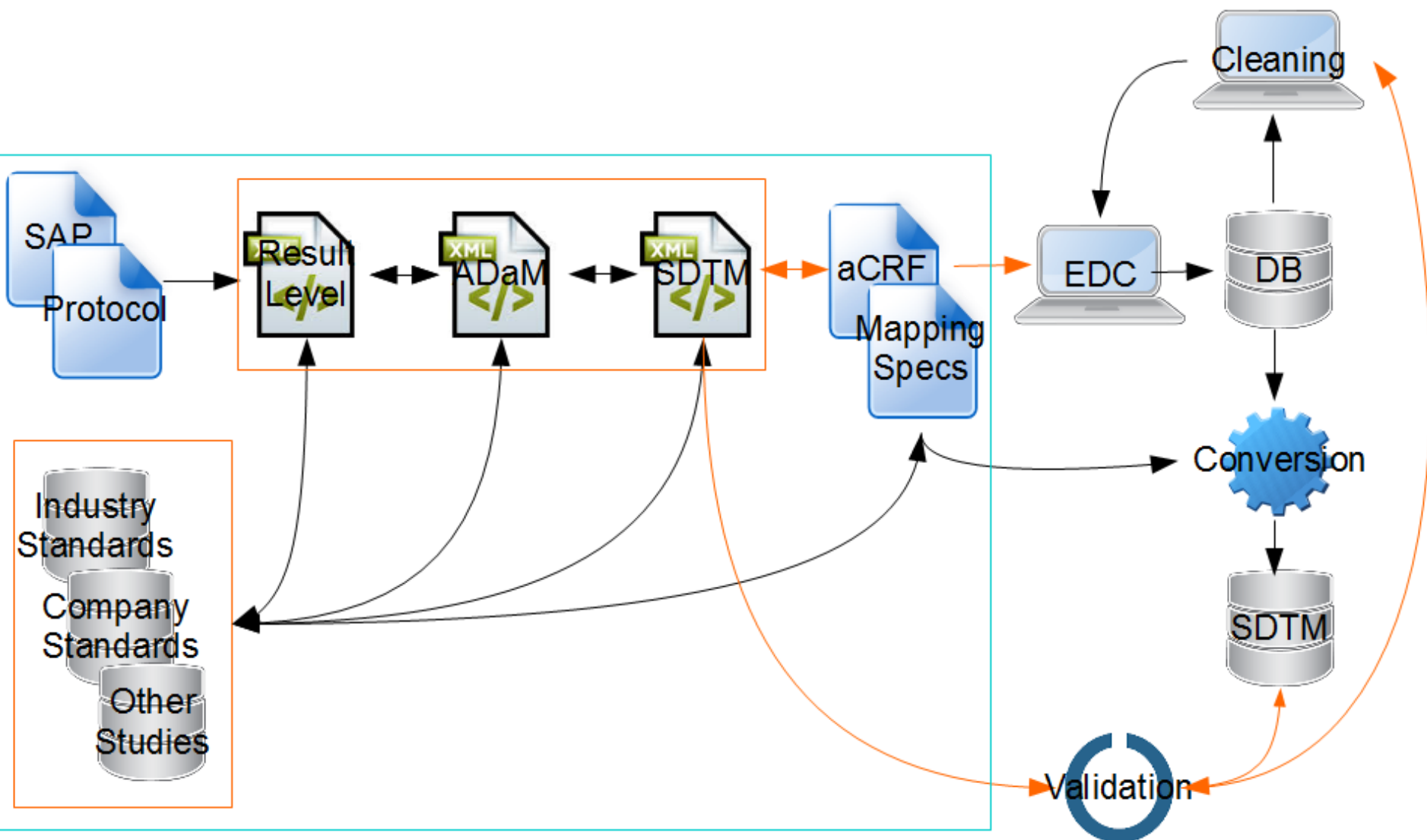
Reviewer vs. Programmer

- › FDA reviewer is limited to submitted study metadata
 - › Define.xml, aCRF, Reviewer's Guide
- › Company programmer uses different sources of study knowledge
 - › EDC specs
 - › Excel specs for mapping and programming
 - › Emails and discussions with team members
- › Descriptive Define.xml is not used in internal processes

Process for prescriptive define.xml

- › Study metadata is defined in advance
- › Based on Protocol and SAP
- › Utilizing industry and company standards
- › “*Begin with the end in mind*” approach
- › Prescriptive define.xml is used as study data spec and input for study specific data validation

Process for prescriptive define.xml



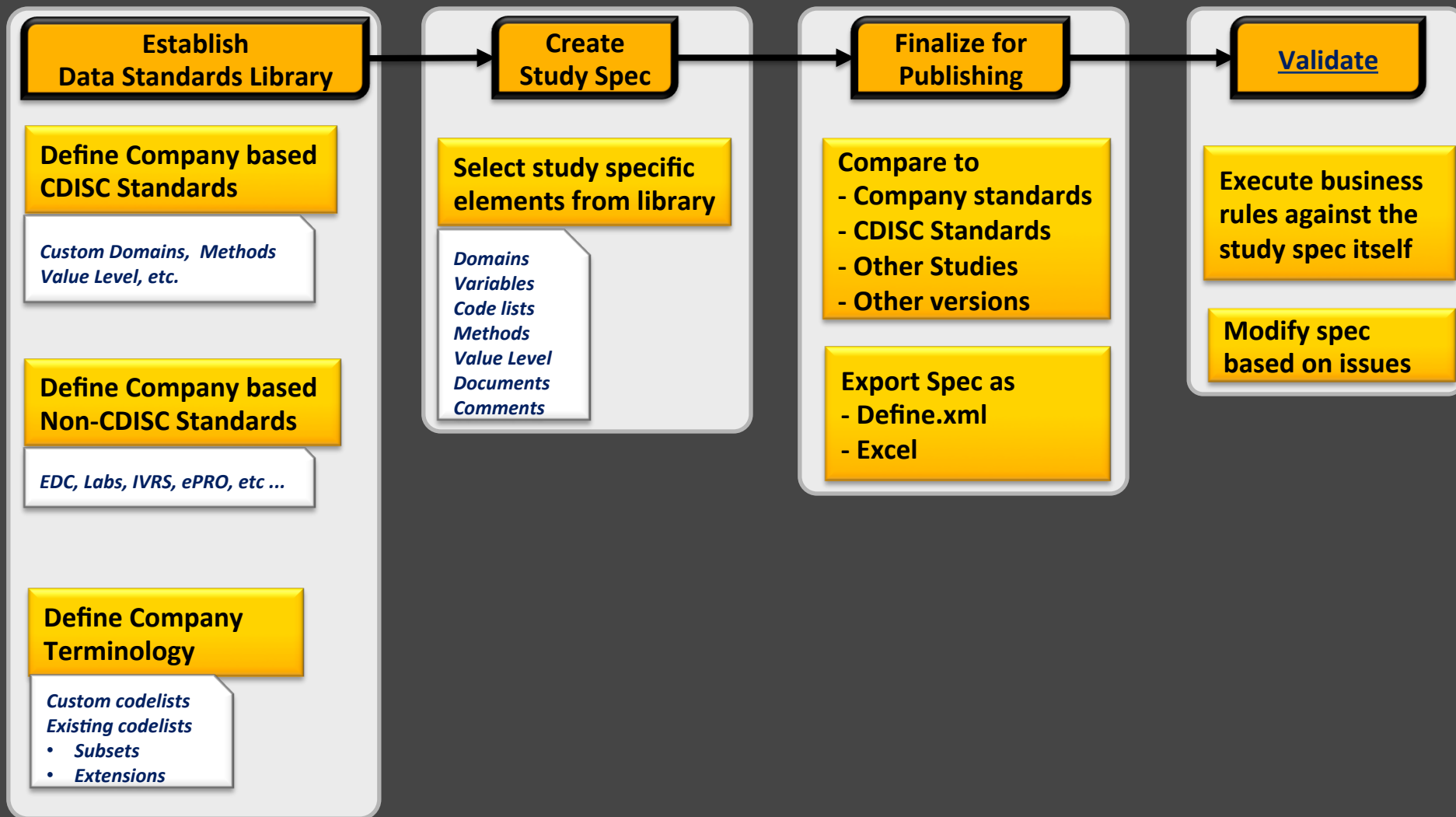
Benefits of prescriptive define.xml

- › Done in advance
 - › Avoids “*too late to fix*” problems
- › Early planning allows implementation oriented to regulatory expectations rather than driven by already collected data with potential issues

Benefits of prescriptive define.xml (2)

- › Utilization of company standards and best practices
 - › Controlled by company rather than the result of inconsistent implementations by different teams and vendors
- › Prescriptive define.xml is used as study data spec and input for study specific data validation
- › Utilization of define.xml in data processing increases the quality of study metadata

Standards Management



Company standardized metadata

- › Define.XML (v2.0) is a good, simple and well structural metadata standard
- › If you use metadata for managing internal processes, why not base it on Define.XML?
 - › Industry standard
 - › Simple, logical, structural, process oriented
 - › Can be easily extended for particular needs
 - › *Define.XML+*
 - › Easy/automatic import of company specs into define.xml

Technical requirements

- › Prescriptive approach needs
 - › Metadata repository
 - › Standards management
 - › Simple user interface
 - › Version control

What is wrong with Excel?

- › Common practice is a usage of Excel for specs
- › Free text entry does not allow entry control
 - › Misspelling
 - › Inconsistency within specs and across studies
- › Single user (personal) document rather than a controlled multiuser environment
- › Missing features
 - › Versioning
 - › Comparison
 - › On-line validation / edit checks
 - › Centralized

Summary

- › Descriptive define.xml is not used in data processing, so it's unlikely to result in high quality
- › Utilization of prescriptive define.xml in data processing increases quality of study metadata
- › Prescriptive metadata requires new process and special tools for standards management
- › Prescriptive define.xml can serve as a backbone for company data standardization process including standards management, programming specifications and validation

Contact info:

Sergiy Sirichenko
ssirichenko@pinnacle21.net