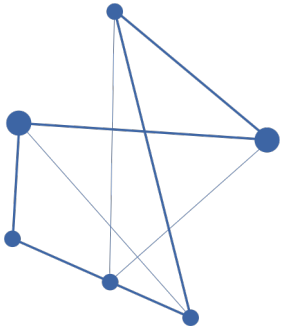




DEFINE-XML REVIEW:

Failing to Plan is Planning to Fail



**Your 1st choice
of global FSP partner**

**Presenter:
Robin Mann**

November 12, 2019



The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of Employer.

The author(s) have no real or apparent conflicts of interest to report.



Define-XML: TOC of Submission Package



Required Standard for Data Submission



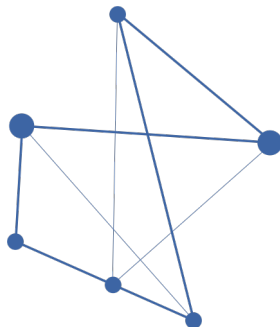
Gap Analysis: Why are we Still Struggling With Define-XML?



14 Steps to Clean Define-XML



Conclusion



Bestselling Author

Colorful Cover

Reputed Publishing
House

Beautiful Synopsis
on the Back



Analogy:

Big Publishing House	→	Pharma Company
Renowned Reputed Authors	→	Clinical Trials Team
Customer/End user	→	Regulatory Authority
Well Publicized Book	→	Clinical Trials Data
Table of Contents	→	Define-XML

ADaM-IG 1.0

Date of Define-XML document generation: 2016-02-16T15:11:22-05:00

Stylesheet version: 2016-02-11

- Analysis Data Reviewer's Guide
- Clinical Study Report
- Statistical Analysis Plan
- ▶ Analysis Results Metadata
- ▶ Analysis Datasets
- ▶ Parameter Value Level Metadata
- ▶ Controlled Terminology
- ▶ Analysis Derivations
- ▶ Comments

Standard	ADaM-IG 1.0
Study Name	CDISC-Sample
Study Description	CDISC-Sample Data Definition
Protocol Name	CDISC-Sample
Metadata Name	Data Definitions for CDISC-Sample, ADaM-IG 1.0
Metadata Description	Data Definitions for CDISC-Sample, ADaM-IG 1.0

Analysis Results Metadata (Summary) for Study CDISC-Sample

[Table 14-3.01](#) Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)

[Dose response analysis for ADAS-Cog changes from baseline](#)

[Pairwise comparisons to placebo for ADAS-Cog changes from baseline](#)

[Table 14-5.02](#) Incidence of Treatment Emergent Serious Adverse Events by Treatment Group

[Incidence of Treatment Emergent Serious Adverse Events by Treatment Group](#)

Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14-3.01

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
Analysis Result	Dose response analysis for ADAS-Cog changes from baseline
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)

Table of Contents

Page

Acknowledgements	i
Abstract	ii
Table of Contents	iii
List of Tables	iv
List of Figures	v
Nomenclatures	vi
1. Introduction	1
2. Literature Survey	2
3. Chapter ONE: Introduction Preference	4
4. Chapter TWO: The Concept of Intermodel Transport	14
4.1 Transport	24
4.2 Various Types of Ocean Transport	26
4.3 The Intermodel Concept	30
5. Chapter THREE: The Development of Intermodel Transport	40
5.1 Containerization	45
5.2 The Role of Ports	50
5.2.1 The Changing Port Function	55
5.2.2 The Sea-Land Interface	60
5.2.3 Load Centers	65
5.2.3.1 The Far East	70
5.2.3.2 The U.S.A.	75
5.2.3.3 Europe	80
5.2.3.4 Other Areas	85
6. Conclusions and Recommendations	90
Bibliography	95
Appendices	96
Appendix 1: Physical Trade Data	97
Appendix 2: Cost Data	98
Appendix 3: Regression Analysis	99



Required Standard for data submission to FDA (U.S) and PMDA (Japan)



Clean and Clear Define-XML is the **responsibility** of the submission team for speedy review

Common Issues in 2016 (% of applications)

Non-descriptive values used for race (Other, Missing, <null>, etc.)

→ 83%

Applicant extended by 10% or more when the codelist was extensible

→ 67%

Separate UNIT codelist not used for each variable

→ 77%

Trial Summary Domain does not follow standards

→ 54%

Define.xml v2.0 not used

→ 73%

Potential clinically relevant duplicate records exist

→ 43%

RFPENDTC not populated according to SDTM guidance

→ 70%

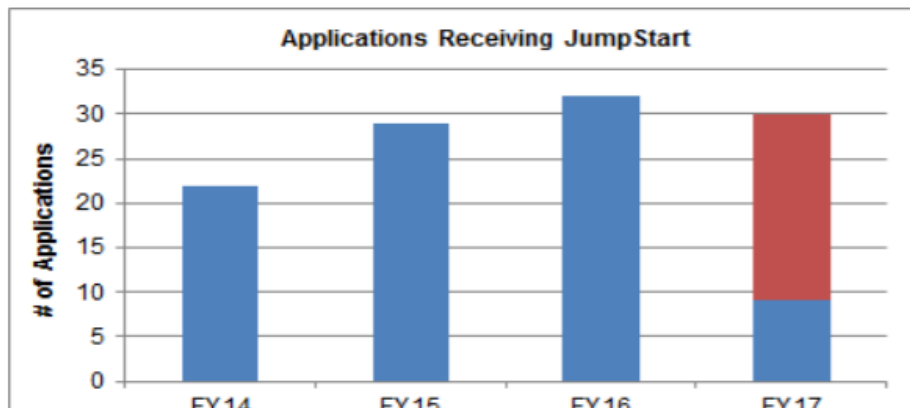
Potential clinically relevant issues with controlled terminology (AE, LB & VS)

→ 43%

No information provided for why subjects failed screening (should be in IE domain, >80%)

→ 42%

JumpStart has analyzed 92 applications since FY14



Data Package Suggestions from JumpStart Findings

Define File

- Use Define.xml v2.0
- Include detailed description of data elements, for example:
 - Code lists that describe categories, subcategories, and reference time-points
 - Detailed and reproducible computational algorithms for all derived variables
 - Applicable value level metadata and description of SUPPQUAL domains
 - Explanations of sponsor-defined identifiers (i.e. – SPID, GRPID, etc)
- Provide separate unit code lists for each domain

Study Data Reviewer’s Guide (SDRG)

- Provide SDRG for each data package with each section populated
- Fix all possible issues identified by FDA Validation Rules and include clear and detailed explanation for all “non-fixable” issues
- Provide Data Flow diagram that shows traceability between data capture, storage, and creation of datasets

14 Steps to Clean Define-XML

Note: In addition to Pinnacle 21 Checks

Step	
1	Document Folder Check
2	Metadata Level Information Check
3	Dataset Level Information Check
4	Individual Datasets / Variable Level
5	Value Level Metadata Check
6	External Dictionaries Check
7	Control Terminology / Codelists
8	Computational Algorithms / Analysis Derivations
9	Computational Algorithms that reference Internal Datasets
10	Analysis Results Metadata Check
11	ORIGIN Check
12	Annotated CRF
13	Incorrect aCRF Links
14	Miscellaneous

- File name? – Define-XML
- File is opening without issues.



- Reviewer's guide (cSDRG.pdf/ADRG.pdf) and aCRF.pdf (SDTM)
- Transport files (SAS Datasets) named correctly
- Broken hyperlinks

- Standard version (SDTM or ADaM), Study name, Study description, Protocol name.
- Pinnacle 21 validator run on Define-XML - Acceptable messages explained in the cSDRG /ADRG.
- Pinnacle 21 validator run on Datasets - Acceptable messages explained in the cSDRG/ADRG.

- All Domains(.xpt) datasets present and are not empty.
- Dataset labels as as per IGs and not missing in Define-XML.
- Domain label matches with the actual Dataset (xpt) label.
- All attributes (structure, class or keys) are present.
- Order of domains in Define-XML and cSDRG/ADRG is correct and matching.

- Variables and attributes consistent with the content of Datasets (xpt).
- Metadata is Study Specific and not general definition from Implementation guides.
- Variable Data types are as per Define-XML section 4.2.1 “*Data type considerations of the CDISC Define-xml-2.0- specifications.pdf*”.
- No Raw/Internal data references in derivations or comments.
- Blindly copied derivations and comments from the Metadata specifications.

- SDTM/ADaM Datasets has its corresponding value level metadata?
- Length attribute is present when Data Type is integer, float, or text.
- The Significant Digits attribute is present when Data Type is float.

The MedDRA version used in the coding of the adverse events needs to be specified in every application.

- Drug dictionary name and version is present and match with reviewers guides.
- Dictionary versions in Define-XML matching with cSDRG and ADRG.

- Code list contains all the planned CRF options and not just those present in the data.
- No one UNIT code split into UNIT(EGRESU) , UNIT (LBRESU) etc.)
- Codelist title includes the code list code – e.g. “AGE UNIT [CL.AGEU, c66781]” .
- If codelist is extended “ * Extended Value” is present.
- Missing codelists for study specific data elements.

- Computational algorithms are clear, well defined, are consistent with specifications and are in plain English language.
- Document correctly and clearly describes all origins and derivations.

- “DERIVED” variables have clear and detailed description of computational
- Computational algorithms does not contain references to internal Datasets or variables which are not submitted in the data package.
- Create an additional pdf document named complex algorithms and document the complex algorithms there or simply document the complex algorithms in reviewers guide.

- Hyperlinks added as applicable and are working.
- Outputs selected for ARM are present.
- Analysis result is consistent with output titles.
- Analysis Parameters are listed when input dataset is BDS.
- Analysis variable is consistent with programming statements.

Was the data collected on CRF, Derived, or received from Laboratory?

- If collected on CRF, then on what pages? annotated on CRF?
- If derived, then with what method? Detailed derivation algorithm is in Define-XML?

If variable has more than one ORIGIN value:

The column ORIGIN should be blank in the “Tabulation Dataset” section and described in the Value Level Metadata.

- Missing or incorrect annotations
- Annotations not created with pdf Annotation feature, but highlighted text or pdf form fields
- Annotations reference internal database fields instead of SDTM variables.

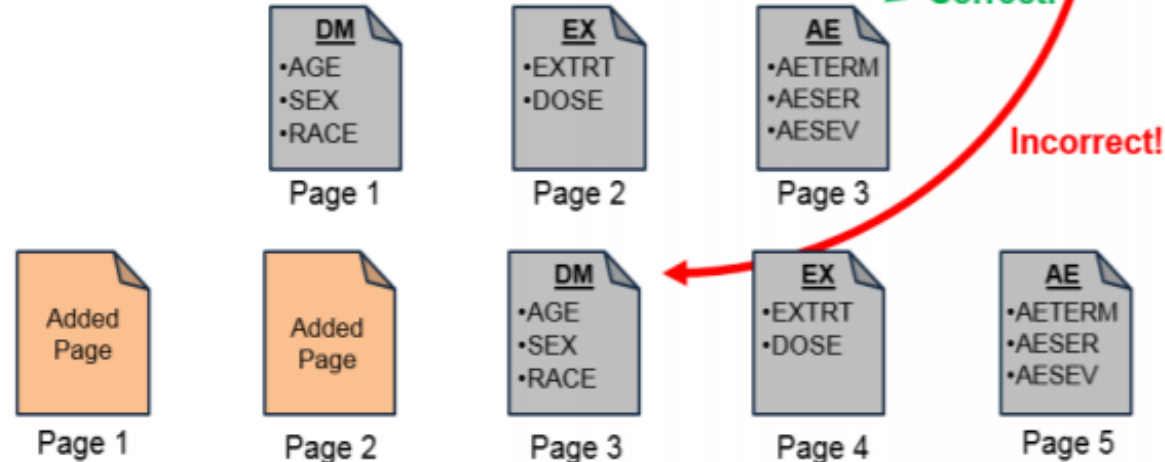
Step 13: Incorrect ACRF Links

Incorrect aCRF link

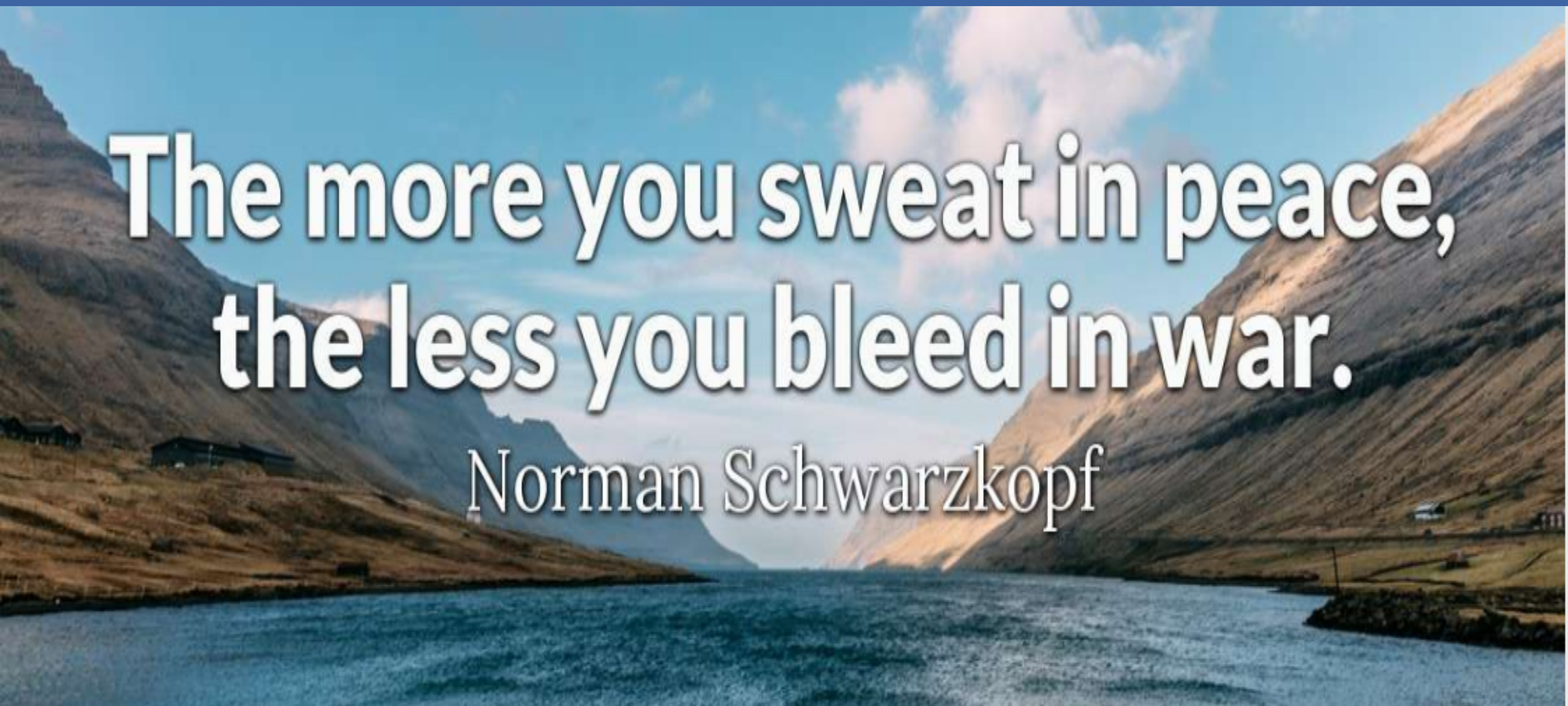
Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Type	Length	Origin
AETERM	Reported Term for the Adverse Event	text	67	CRF Page 3

Example aCRF:



- Derivation at multiple places or too long in Define-XML
- Are baseline flags populated in SDTM?
- Separate pdf file for large derivations that require formatting.

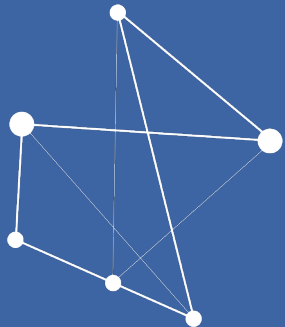


**The more you sweat in peace,
the less you bleed in war.**

Norman Schwarzkopf



an IQVIA business



Your 1st choice
of global FSP partner

robin.mann@gcesolutions.com
parveen.k12@geninvo.com