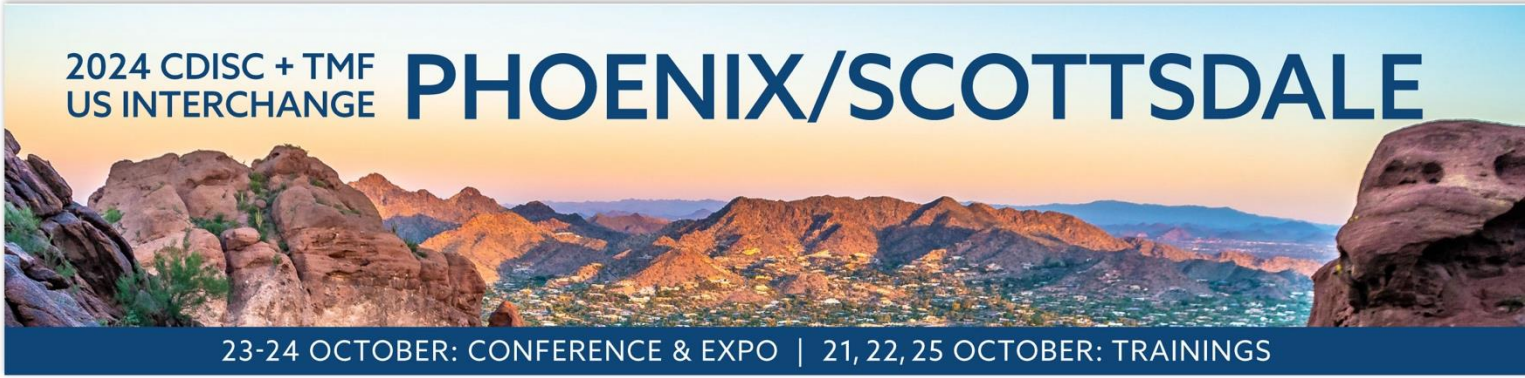


Analyzing the Shift: From CDISC Define-xml 2.0 to 2.1

Presented by Dinesh Kumar, Manager, Clinical Biometrics, Zifo Scientific Informatics

Meet the Speaker

Dinesh Kumar

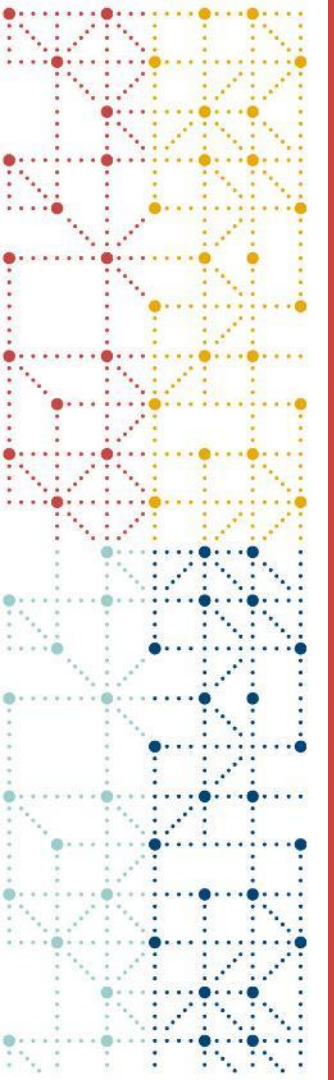
Title: Manager

Organization: Zifo Scientific Informatics

Dinesh is a seasoned Consultant with over 9 years of experience in both the clinical and research & development domains. As the Manager of the US Clinical Team, he oversees the successful execution of numerous projects for analysis and regulatory submission while simultaneously driving business development initiatives. His expertise spans a wide range of clinical operations, enabling him to deliver exceptional results and build strong partnerships. Dinesh is instrumental in initiating various analyses of Real-World Data (RWD) to optimize clinical operations and drive operational efficiency.

Disclaimer and Disclosures

- *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 CDISC.*



Agenda

1. Introduction
2. Notable Enhancements from V2.0
3. Challenges Faced
4. Conclusion

Throwback

Applicable for
- SDTM IG 3.1.2 and higher
- ADaM IG 1.0 and higher

Introduction

- The FDA and PMDA mandate Define-XML for each study to detail the datasets, variables, controlled terms, and other specified metadata utilized.

Discussing Define-XML 2.1:

- New attributes and elements
- Challenges faced

1. References to versions of content standards

Define xml 2.0

def: StandardName attribute

Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt	

Define xml 2.1

Included Standards and Standard elements

Standards for Study CDISC01_1

Standard	Type	Status	Documentation
SDTMIG 3.1.2	IG	Final	The CDISC01 study was modeled on a very old SDTMIG and no attempt was done yet to upversion it to a newer SDTMIG
SDTMIG 3.3	IG	Final	As an example, the CDISC01 study was adjusted to include a new Domain available in SDTM IG 3.3
SDTMIG-MD 1.0	IG	Final	As an example, the CDISC01 study was adjusted to include a new Domain available in SDTMIG-MD 1.0. The XS Domain is expected to reference the device used with variable SPDEVID.
CDISC/NCI SDTM 2011-12-09	CT	Final	Assuming the CT was not upversioned for this study
CDISC/NCI SDTM 2015-12-18	CT	Final	The CT version applicable for the new Domain is the 2015-12-18 version
CDISC/NCI DEFINE-XML 2023-06-30	CT	Final	This is the CDISC CT Package 54 (2023-06-30) associated to the CDISC Define-XML Specification Version 2.1.6.

1. References to versions of content standards (Cont'd)

Standards for Study CDISC-Sample			
Standard	**Type**	**Status**	**Documentation**
ADaMIG 1.1	IG	Final	
CDISC/NCI ADaM 2017-09-29	CT	Final	ADaM specific CT is applicable for a few variables only.
CDISC/NCI SDTM 2018-06-29	CT	Final	SDTM CT is applicable for variables copied from SDTM to ADaM.
CDISC/NCI DEFINE-XML 2023-06-30	CT	Final	This is the CDISC CT Package 54 (2023-06-30) associated to the CDISC Define-XML Specification Version 2.1.6.

Status attribute:

1. Allowed values – Draft, Provisional, Final
2. If no value is provided, the assumption is that the publishing status is **Final.**

1. References to versions of content standards (Cont'd)

Multiple SDTM Guidelines or any other therapeutic guides are used

- *For Integration study purpose*
- *Multiple standards – (SDTM-MD, SDTM-PGx)*
- *Any domains from latest SDTM versions*

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
TS [SDTMIG 3.1.2]	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ		ts.xpt
MB [SDTMIG 3.3]	Microbiology Specimen	FINDINGS	One record per microbiology specimen finding per visit per subject	Tabulation	STUDYID, USUBJID, MBTESTCD, VISITNUM, MBDTC		mb.xpt
DM [SDTMIG 3.4]	Demographics	SPECIAL	One record per subject	Tabulation	STUDYID, USUBJID	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide [section2.1]	dm.xpt
EC [SDTMIG 3.2]	Exposure as Collected	INTERVENTIONS	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, ECSTDTC, ECENDTC, ECTRT, ECDOSE		ec.xpt

Addition of new column "Standard" (Recommended by P21) in Datasets Sheet.

2. Dataset definitions

Datasets							
Dataset	**Description**	**Class - SubClass**	**Structure**	**Purpose**	**Keys**	**Documentation**	**Location**
ADSL [ADaMIG 1.1]	Subject-Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	one record per subject	Analysis	STUDYID, USUBJID	Screen Failures are excluded since they are not needed for this study analysis. See referenced dataset creation program and ADRG adsl.sas Analysis Data Reviewer's Guide [6]	adsl.xpt
ADQSADAS [ADaMIG 1.1]	ADAS-Cog Analysis	BASIC DATA STRUCTURE	One record per subject per parameter per analysis visit per analysis date	Analysis	STUDYID, USUBJID, PARAMCD, AVISIT, ADT	See referenced dataset creation program and ADRG adqsadas.sas Analysis Data Reviewer's Guide [Section2.1]	adqsadas.xpt
ADAE [ADaMIG 1.1]	Adverse Events Analysis Dataset	OCCURRENCE DATA STRUCTURE - ADVERSE EVENT	one record per subject per adverse event	Analysis	STUDYID, USUBJID, AETERM, ASTDT, AESEQ	See SAS program adae.sas	adae.xpt

*Addition of new column "**Subclass**" (Recommended by P21) in **Datasets** Sheet.*

3. Controlled Terminology (Dataset Variable Definitions)

Standards for Study CDISC01_1			
Standard	**Type**	**Status**	**Documentation**
SDTMIG 3.1.2	IG	Final	The CDISC01 study was modeled on a very old SDTMIG and no attempt was done yet to upversion it to a newer SDTMIG
SDTMIG 3.3	IG	Final	As an example, the CDISC01 study was adjusted to include a new Domain available in SDTM IG 3.3
SDTMIG-MD 1.0	IG	Final	As an example, the CDISC01 study was adjusted to include a new Domain available in SDTMIG-MD 1.0. The XS Domain is expected to reference the device used with variable SPDEVID.
CDISC/NCI SDTM 2011-12-09	CT	Final	Assuming the CT was not upversioned for this study
CDISC/NCI SDTM 2015-12-18	CT	Final	The CT version applicable for the new Domain is the 2015-12-18 version
CDISC/NCI DEFINE-XML 2023-06-30	CT	Final	This is the CDISC CT Package 54 (2023-06-30) associated to the CDISC Define-XML Specification Version 2.1.6.

*Addition of new column "**Terminology**" (Recommended by P21) in **Codelists** Sheet.*

Domain Abbreviation (DM) [*C66734*] [CDISC/NCI SDTM 2011-12-09]	
Permitted Value (Code)	**Display Value (Decode)**
DM [*C49572*]	Demographics
Domain Abbreviation (EC) [*C66734*] [CDISC/NCI SDTM 2015-12-18] The Domain codelist was updated when SDTMIG 3.2 became production. Referencing a newer CT version that includes the revised codelist	
Permitted Value (Code)	**Display Value (Decode)**
EC [*C49587*]	Exposure as Collected

Description text for codelist can be added

4. Origin (Dataset Variables Definitions)

- Addition of two key attributes: **Type** & **Source**
- Managed and published with CDISC Controlled Terminology in non-extensible codelists
- For SEND Datasets – only "Type" attribute is used

Addition of new column "Source" (Recommended by P21) in Variables and Value level Sheet.

Deprecation

- Values "CRF" and "eDT" are removed from Origin

Addition

- Values "Collected" and "Not Available" are added into Origin
- New Source values
  - Subject
  - Investigator
  - Vendor
  - Sponsor

Used in Legacy datasets (not provided by Organization)

4. Considerations with Origin - Clinical Studies

Define xml 2.0	Define xml 2.1		Use cases
	Type	Source	
CRF	Collected	Subject / Investigator	*Subject case*: EHR, ePRO in Registry, RWE studies *Investigator case*: Information collected directly from CRF
eDT		Vendor	Laboratory & ECG assessments results, Pharmacokinetic concentrations and its parameters
Derived	Derived	Vendor	Scores derived based on the results from external Lab vendors (Eg: Cumulative Pain score)
		Sponsor	Baseline, Study day derivations For all ADaM Variables
Assigned	Assigned	Vendor	Confirmatory analyses done by third party based on result values (Eg: Confirmation of tumor based on the response)
		Sponsor	Predefined Codelist values (eg: --TESTCD, DOMAIN) For all ADaM Variables
Protocol	Protocol	Sponsor	Methods used for lab assessments & Specimen collection information, the mode of administration planned for study drug
Predecessor	Predecessor		SDTM Variables retention in ADaM Domains

4. Considerations with Origin (cont'd)

Sample SDTM Submission

Define.xml 2.0

Demographics (DM) [Location: dm.xpt]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["DM" = "Demographics"] <Domain Abbreviation (DM)>	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID
SUBJID	Subject Identifier for the Study		text	6		CRF Page 3	

Define.xml 2.1

DM (Demographics) – [SDTMIG 3.3]

Related Supplemental Qualifiers Dataset: SUPPDM (Supplemental Qualifiers for DM)

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier	text	7		Protocol (Source: Sponsor)
DOMAIN	Domain Abbreviation	text	2	Domain Abbreviation (DM) • "DM" = "Demographics"	Assigned (Source: Sponsor)
USUBJID	Unique Subject Identifier	text	14		Derived (Source: Sponsor) Concatenation of STUDYID and SUBJID
SUBJID	Subject Identifier for the Study	text	6		Collected (Source: Investigator) Annotated CRF [3]

4. Considerations with Origin (cont'd)

Sample ADaM Submission

Define.xml 2.0

AGE	Age	integer	8		Predecessor: DM.AGE
AGEGR1	Pooled Age Group 1	text	5	["<65", "65-80", ">80"] <AGEGR1>	Derived: Character variable derived from ADSL.AGEGR1N
AGEGR1N	Pooled Age Group 1 (N)	integer	8	["1" = "<65", "2" = "65-80", "3" = ">80"] <AGEGR1N>	Assigned: AGEGR1 = 1 if AGE <65. AGEGR1 = 2 if AGE 65-80. AGEGR1 = 3 if AGE >80.
AGEU	Age Units	text	5	["YEARS"] <Age Unit>	Predecessor: DM.AGEU

Define.xml 2.1

AGE	Age	integer	2		Predecessor: DM.AGE
AGEGR1	Pooled Age Group 1	text	5	Age Group • "<65" • "65-80" • ">80"	Derived (Source: Sponsor) Grouping of AGE into <65, 65-80, and >80
AGEGR1N	Pooled Age Group 1 (N)	integer	2	Age Group (N) • 1 = "<65" • 2 = "65-80" • 3 = ">80"	Assigned (Source: Sponsor) Numeric code for AGEGR1
AGEU	Age Units	text	5	Age Unit • "YEARS"	Predecessor: DM.AGEU

5. What if data is not collected, which is intended to collect?

- Applicable for SDTM Submissions
- If a submission dataset/variable within a dataset has no values collected till the end of the trial, “[No Data]” will be captured in “Datasets” sheet

Dataset level:

SUPPRP [SDTMIG 3.3] **[No Data]**	Supplemental Qualifiers for Reproductive System Findings (Reproductive System Findings)	RELATIONSHIP	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM		

Variable or Value level:

XSSTRESU **[No Data]**	Standard Units	text	9	Units for S Findings Results • "g/dL" = "g/dL" • "mg/dL" = "mg/dL"	Assigned (Source: Sponsor) Planned Numeric tests were not performed.

Advantages of Define v2.1 over Define v2.0

1. References to versions of content standards
2. Controlled terminology

1. Dataset definitions

1. References to versions of content standards
2. Dataset definitions
3. Controlled terminology
4. Origin
5. Handling domains with no data

6. Additional Updates

- Alias Element allows for SAS variables or dataset names longer than 8 characters – to support legacy datasets submission.

- The **def:Context** attribute on the ODM root element can have values such as
  - Submission – Intended for a regulatory submission
  - Other - Not intended for a regulatory submission

Date of document generation: 2014-04-26T11:14:29-04:00

Stylesheet version: 2013-12-12

Date/Time of Define-XML document generation: 2023-07-05T14:00:00

Define-XML version: 2.1.6

Define-XML Context: Other

Stylesheet version: 2019-02-11

Date/Time of Define-XML document generation: 2023-07-05T14:00:00

Define-XML version: 2.1.6

Define-XML Context: Submission

Stylesheet version: 2019-02-11

Common issues faced in Define.xml 2.1 Implementation

- *Pinnacle validation of datasets with define 2.1 doesn't identify classes of supplementary domains as well as custom domains*

- *The ADaM-specific Analysis Results Metadata (ARM) have not yet been integrated into the main Define-XML 2.1. As a result, users must still reference the ARM standard within the ODM element when working with Define-XML 2.1.*

- *The error **"Attribute 'def' is not allowed in element 'ItemRef'"** in Pinnacle occurs when def:HasNoData is applied at the value level for blank results, which is acceptable per Define XML 2.1. However, adding comments to the value-level ItemDef did not resolve the validation issue.*

Enhancements of Define v2.1 relative to Define v2.0

- With Define v2.1, the connection between standards and datasets, as well as controlled terminologies and codelists, is enhanced, simplifying the validation process for reviewers.
- The inclusion of Subclasses in v2.1 helps in organizing complex datasets more efficiently.
- Defining the origin of data elements reduces the time needed to understand data sources.
- These improvements collectively enhance the efficiency and accuracy of the review process, making it simpler for reviewers to validate data.

Conclusion

- *Define.xml version 2.1 supports advanced automation features, allowing for more efficient data validation and error checking.*
- *Certain fields, like the HasNoData flag, are best populated once the data is finalized.*
- *Other fields, such as SubClass, should be filled in early to ensure accurate validation results.*
- *The right resources and a clear grasp of the changes can lead to a smooth transition.*

Thank You!