sponsor requests ODM XML Mapper JavaScript ODM extensions define.xml iText HTML CSS XSL-FO validation

define.xml: A Crash Course

schema/XSD Xpath define.pdf metadata tables Oracle/database

Frank DiIorio
CodeCrafters, Inc.
Philadelphia PA

XML4Pharma define version 'x metadata interface XSL old school brute f (the other) define.pdf XMLPad metadata storage SAS Clinical Standards Toolkit CDISC standard version 'x

Remember define.pdf?

- Purpose: document deliverables
  - Datasets: description, structure, sort order
  - Variables: attributes, codes, derivation, *et al.*
- Created using:
  - Metadata, SAS macros
- Contents validated by:
  - Visual inspection
  - Programmatic checks of the metadata
- FDA now requests define.**xml**, aka CDSISC's "Case Report Tabulation Data Definition Specification"
- And *conceptually* it resembles define.pdf …

define.xml: Dataset-Level (transformed by XSL)

SDTM Datasets for Study CDISCPILOT01

Dataset	Description	Class	Structure	Purpose	Keys	Location
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCD	te.xpt
TI	Trial Inclusion/ Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM	tv.xpt
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
SE	Subject Elements	Special Purpose	One record per actual Element per subject	Tabulation	STUDYID, USUBJID, ETCD	se.xpt
SV	Subject Visits	Special Purpose	One record per actual visit per subject	Tabulation	STUDYID, USUBJID, VISITNUM	sv.xpt
CM	Concomitant	Interventions	One record per recorded medication occurrence or	Tabulation	STUDYID, USUBJID, CMTRT	cm.xpt

define.xml: Variable-Level (transformed by XSL)

Demographics (DM) dm.xpt

Variable	Label	Key	Type	Length	Code List / Controlled Terms	Origin	Role	Source/Derivation/Comments
STUDYID	Study Identifier		text	12		CRF Page 7	IDENTIFIER	
DOMAIN	Domain Abbreviation		text	2		Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier		text	11		Derived	IDENTIFIER	Concatenation of STUDYID, DM.SITEID and DM.SUBJID
SUBJID	Subject Identifier for the Study		text	4		CRF Page 7	TOPIC	
RFSTDTC	Subject Reference Start Date/Time		date	10	ISO8601	Derived	RECORD QUALIFIER	Date/time of first study drug treatment derived from EX
RFENDTC	Subject Reference End Date/Time		date	10	ISO8601	Derived	RECORD QUALIFIER	Date/time of last study drug treatment derived from EX
RFXSTDTC	Date/Time of First Study Treatment		datetime	20	ISO8601	Derived	RECORD QUALIFIER	RFXSTDTC=RFSTDTC
RFXENDTC	Date/Time of Last Study Treatment		datetime	20	ISO8601	Derived	RECORD QUALIFIER	RFXENDTC=RFENDTC
RFICDTC	Date/Time of Informed Consent		datetime	20	ISO8601	Derived	RECORD QUALIFIER	Date of informed consent was not entered in database (see annotated CRF)
RFPENDTC	Date/Time of End of Participation		datetime	20	ISO8601	Derived	RECORD QUALIFIER	DSSTDTC of last disposition event

define.xml: Similar, but …

- define.xml differs from define "classic":
  - Unlike a PDF, it is easily machine-readable
  - It follows a strictly defined format (schema)
  - It's "meatier" than define.pdf, requiring much richer metadata
  - Requires validation of
    - syntax
    - compliance with schema
- Clearly, we're dealing with something new and complex

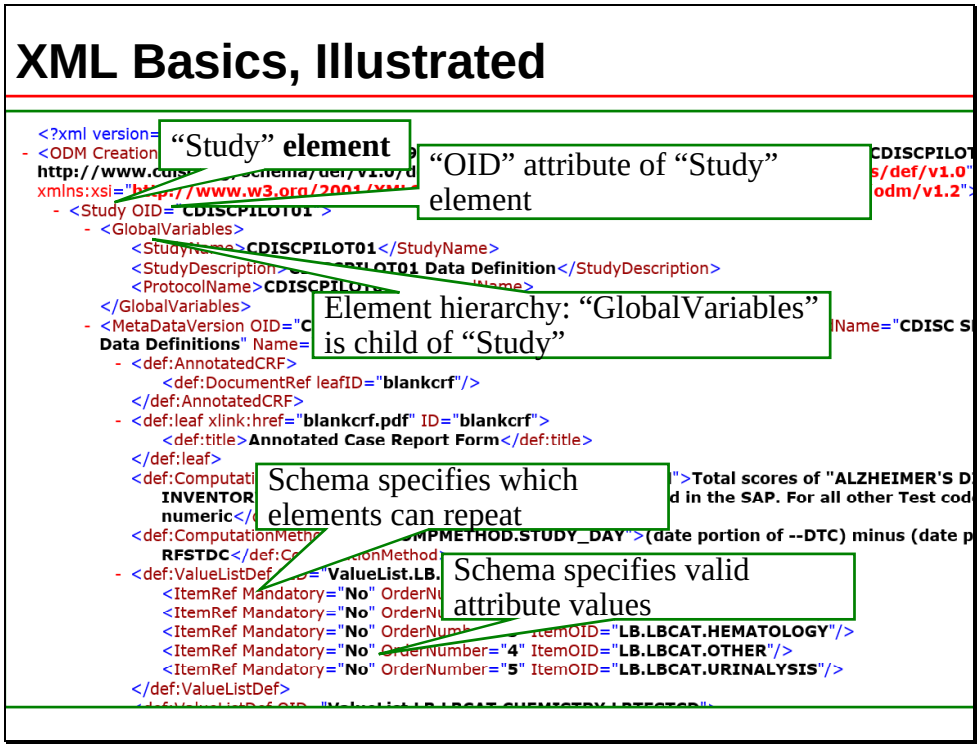
This Presentation …

- Briefly reviews XML basics
- Describes metadata needed to support construction of define.xml
- Presents one way to build the XML file
- Shows how to validate the file
- Discusses define.pdf (no, not *that* define.pdf!)
- Focuses on define Version 1 but identifies issues relevant to Version 2
- Is simply an *overview* of the file creation and validation process

XML Basics

- E**x**tensible **M**arkup **L**anguage: plain text with mark-up ("tags") similar in look & feel to HTML
- Content is user-defined, by schemas
- Files are collections of elements (aka "nodes"), each of which can have one or more attributes. Elements can be arranged in a hierarchy.
- Unlike HTML, emphasis is on data ***content***, not its ***display***
- XML is part of a "family" of specifications
  - XSL – transforms XML into another format
  - XPath – navigates within the document. Used by XSL.
  - XSD/Schema – defines rules for content and structure of an XML file

XML Basics, Illustrated



define.xml Basics

- define.xml must be valid from two perspectives:
  - Syntax
  - Content (compliance with schema)
- define schema/content
  - An extension of the CDISC Operational Data Model (ODM)
  - Schema controls *content*, not display
    - Rules for names, attributes, number of occurrences, order of nodes, etc.
    - A value can conform to the schema but still be wrong! (e.g., type is Integer but really should be Float)
  - Available at CDISC, OpenCDISC web sites
  - *Determining what goes where is, arguably, the hardest part of the file creation process*.

Node Order

Start of OpenCDISC XML file showing node order

```
<?xml version="1.0"?>
- <metadata>
  - <element name="ODM">
      <attribute name="FileType"/>
      <attribute name="FileOID"/>
      <attribute name="CreationDateTime"/>
      <attribute name="PriorFileOID"/>
      <attribute name="ODMVersion"/>
      <attribute name="Originator"/>
      <attribute name="SourceSystem"/>
      <attribute name="SourceSystemVersion"/>
      <attribute name="xsi:schemaLocation"/>
    - <element name="Study">
        <attribute name="OID"/>
      - <element name="GlobalVariables">
          <element name="StudyName"/>
          <element name="StudyDescription"/>
          <element name="ProtocolName"/>
        </element>
      - <element name="MetaDataVersion">
          <attribute name="OID"/>
          <attribute name="Name"/>
          <attribute name="Description"/>
          <attribute name="def:DefineVersion"/>
          <attribute name="def:StandardName"/>
          <attribute name="def:StandardVersion"/>
        - <element name="def:AnnotatedCRF">
          - <element name="def:DocumentRef">
              <attribute name="leafID"/>
```

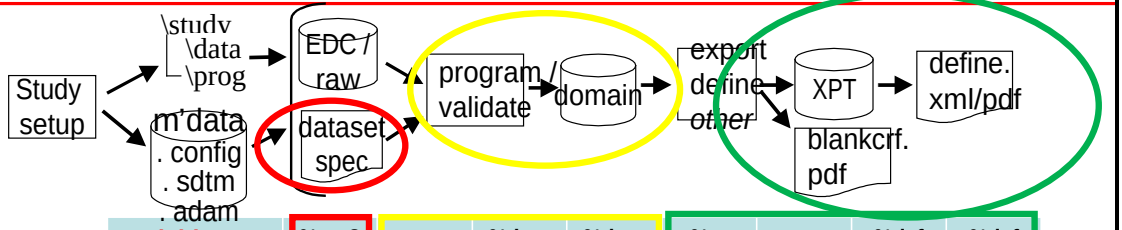
What You'll Need

- An XML Viewer/Editor (display ODM schema, define.xml, XSL) such as:
  - XMLpad
  - SAS XML Mapper
- Validator
  - OpenCDISC
  - SAS Clinical Standards Toolkit
  - XML4Pharma
  - *Can be supplemented with home-grown tools*
- Knowledge and patience
  - W3Schools.com, other sites/books

Between the Tags: Metadata

- Metadata
  - Drives the creation of the XML
  - And can also be used for various tasks throughout the project life cycle *(next slide)*
- Metadata tables can include:
  - Study-level: protocol name, standard name/version
  - Datasets: name, structure, key fields
  - Variables: attributes, controlled terminology usage, derivation/CRF source
  - Value: detail of variable values (test codes, etc.)
  - Comp. algorithms: extended and/or repeated derivations
  - Controlled terms: descriptions and values of coded/enumerated
  - Results: description of TFLs – name, content, source(s), etc. **(new in define v2)**

Metadata: Usage Throughout Study Life Cycle

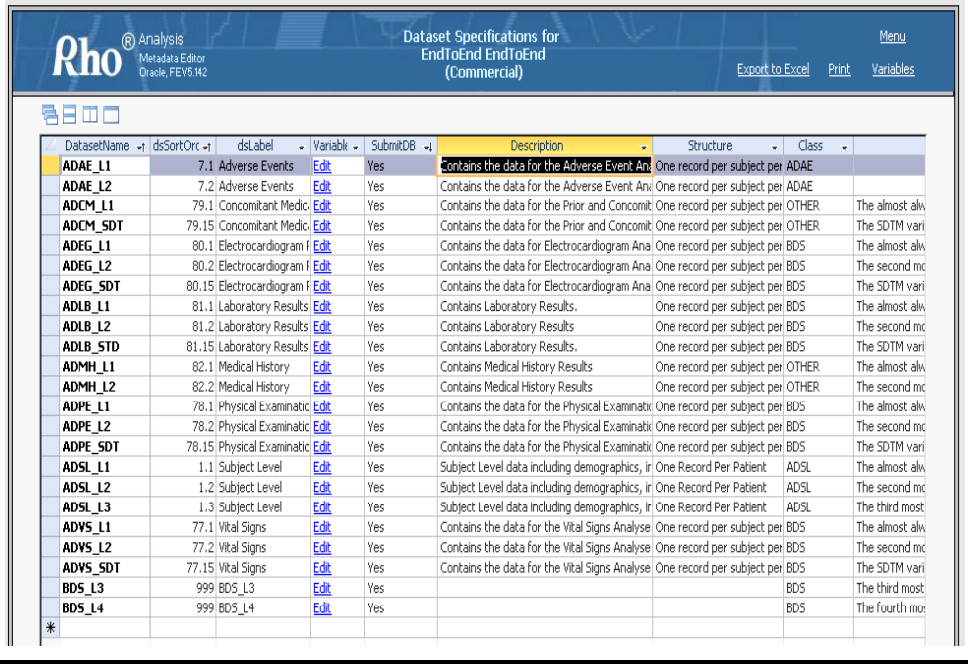


Variables Table	%cre8 Spec	%attrib	%dom Split	%dom Chk	%cr XFDF	%xpt	%def XML	%def PDF
domain	✓	✓	✓	✓	✓	✓	✓	
variable	✓	✓	✓	✓	✓	✓	✓	
type	✓	✓	✓	✓			✓	
length	✓	✓					✓	
label		✓	✓	✓			✓	
order	✓	✓	✓	✓		✓	✓	
definitionProg	✓							
definitionSub							✓	
use	✓	✓	✓	✓	✓	✓	✓	
crflocation					✓		✓	
core	✓		✓	✓			✓	

Metadata Issues

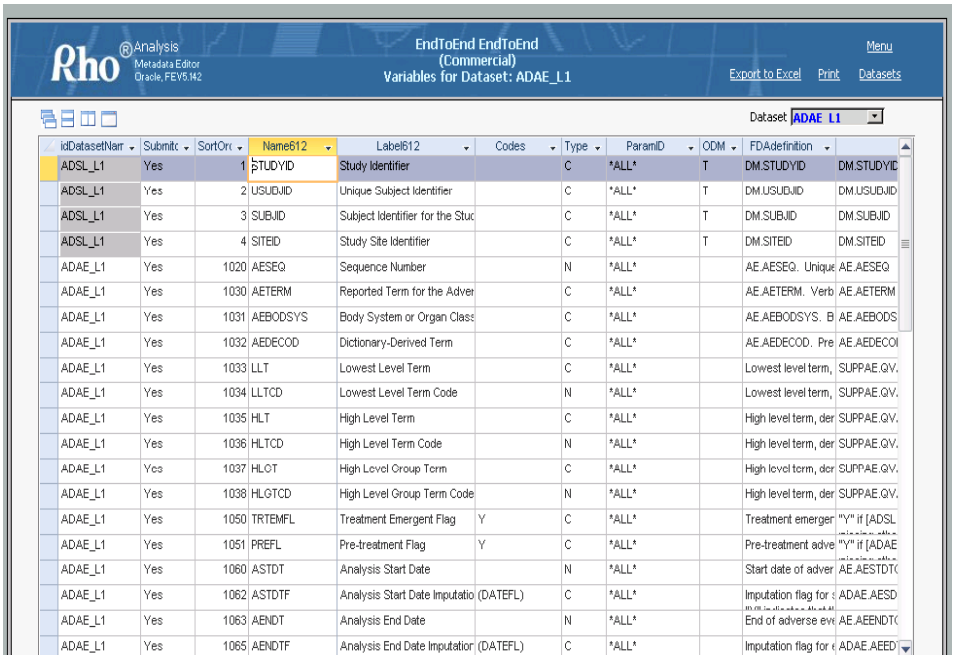
- Design
  - Ideally, maps (directly/views) to XML elements and attributes with a minimum of transformation
  - Should be sensitive to changes in standards:
    - define.xml
    - data (SDTM, ADaM)
- Storage
  - The metadata should be regarded as a valuable corporate asset.
  - **So don't store it in Excel!** Oracle or similar enterprise-level database is a far better choice (though more resource intensive).

Metadata Issues: Entry (Dataset-Level)



DatasetName	dsSortOrc	dsLabel	Variabl	SubmitDB	Description	Structure	Class	
ADAE_L1	7.1	Adverse Events	Edit	Yes	Contains the data for the Adverse Event Ani	One record per subject per	ADAE	
ADAE_L2	7.2	Adverse Events	Edit	Yes	Contains the data for the Adverse Event Ana	One record per subject per	ADAE	
ADCM_L1	79.1	Concomitant Medic	Edit	Yes	Contains the data for the Prior and Concomit	One record per subject per	OTHER	The almost alw
ADCM_SDT	79.15	Concomitant Medic	Edit	Yes	Contains the data for the Prior and Concomit	One record per subject per	OTHER	The SDTM vari
ADEG_L1	80.1	Electrocardiogram I	Edit	Yes	Contains the data for Electrocardiogram Ana	One record per subject per	BDS	The almost alw
ADEG_L2	80.2	Electrocardiogram I	Edit	Yes	Contains the data for Electrocardiogram Ana	One record per subject per	BDS	The second mo
ADEG_SDT	80.15	Electrocardiogram I	Edit	Yes	Contains the data for Electrocardiogram Ana	One record per subject per	BDS	The SDTM vari
ADLB_L1	81.1	Laboratory Results	Edit	Yes	Contains Laboratory Results.	One record per subject per	BDS	The almost alw
ADLB_L2	81.2	Laboratory Results	Edit	Yes	Contains Laboratory Results	One record per subject per	BDS	The second mo
ADLB_STD	81.15	Laboratory Results	Edit	Yes	Contains Laboratory Results.	One record per subject per	BDS	The SDTM vari
ADMH_L1	82.1	Medical History	Edit	Yes	Contains Medical History Results	One record per subject per	OTHER	The almost alw
ADMH_L2	82.2	Medical History	Edit	Yes	Contains Medical History Results	One record per subject per	OTHER	The second mo
ADPE_L1	78.1	Physical Examinatic	Edit	Yes	Contains the data for the Physical Examinati	One record per subject per	BDS	The almost alw
ADPE_L2	78.2	Physical Examinatic	Edit	Yes	Contains the data for the Physical Examinati	One record per subject per	BDS	The second mo
ADPE_SDT	78.15	Physical Examinatic	Edit	Yes	Contains the data for the Physical Examinati	One record per subject per	BDS	The SDTM vari
ADSL_L1	1.1	Subject Level	Edit	Yes	Subject Level data including demographics, ir	One Record Per Patient	ADSL	The almost alw
ADSL_L2	1.2	Subject Level	Edit	Yes	Subject Level data including demographics, ir	One Record Per Patient	ADSL	The second mo
ADSL_L3	1.3	Subject Level	Edit	Yes	Subject Level data including demographics, ir	One Record Per Patient	ADSL	The third most
ADVS_L1	77.1	Vital Signs	Edit	Yes	Contains the data for the Vital Signs Analyse	One record per subject per	BDS	The almost alw
ADVS_L2	77.2	Vital Signs	Edit	Yes	Contains the data for the Vital Signs Analyse	One record per subject per	BDS	The second mo
ADVS_SDT	77.15	Vital Signs	Edit	Yes	Contains the data for the Vital Signs Analyse	One record per subject per	BDS	The SDTM vari
BDS_L3	999	BDS_L3	Edit	Yes			BDS	The third most
BDS_L4	999	BDS_L4	Edit	Yes			BDS	The fourth mo

Metadata Issues: Entry (Variable-Level)



idDatasetNam	Submitc	SortOrc	Name612	Label612	Codes	Type	ParamID	ODM	FDAdefinition	
ADSL_L1	Yes	1	STUDYID	Study Identifier		C	*ALL*	T	DM.STUDYID	DM.STUDYID
ADSL_L1	Yes	2	USUBJID	Unique Subject Identifier		C	*ALL*	T	DM.USUBJID	DM.USUBJID
ADSL_L1	Yes	3	SUBJID	Subject Identifier for the Stuc		C	*ALL*	T	DM.SUBJID	DM.SUBJID
ADSL_L1	Yes	4	SITEID	Study Site Identifier		C	*ALL*	T	DM.SITEID	DM.SITEID
ADAE_L1	Yes	1020	AESEQ	Sequence Number		N	*ALL*		AE.AESEQ. Unique	AE.AESEQ
ADAE_L1	Yes	1030	AETERM	Reported Term for the Adver		C	*ALL*		AE.AETERM. Verb	AE.AETERM
ADAE_L1	Yes	1031	AEBODSYS	Body System or Organ Class		C	*ALL*		AE.AEBODSYS. B	AE.AEBODS
ADAE_L1	Yes	1032	AEDECOD	Dictionary-Derived Term		C	*ALL*		AE.AEDECOD. Pre	AE.AEDECOI
ADAE_L1	Yes	1033	LLT	Lowest Level Term		C	*ALL*		Lowest level term,	SUPPAE.QV.
ADAE_L1	Yes	1034	LLTCD	Lowest Level Term Code		N	*ALL*		Lowest level term,	SUPPAE.QV.
ADAE_L1	Yes	1035	HLT	High Level Term		C	*ALL*		High level term, der	SUPPAE.QV.
ADAE_L1	Yes	1036	HLTCD	High Level Term Code		N	*ALL*		High level term, der	SUPPAE.QV.
ADAE_L1	Yes	1037	HLGT	High Level Group Term		C	*ALL*		High level term, der	SUPPAE.QV.
ADAE_L1	Yes	1038	HLGTCD	High Level Group Term Code		N	*ALL*		High level term, der	SUPPAE.QV.
ADAE_L1	Yes	1050	TRTEMFL	Treatment Emergent Flag	Y	C	*ALL*		Treatment emerger	"Y" if [ADSL
ADAE_L1	Yes	1051	PREFL	Pre-treatment Flag	Y	C	*ALL*		Pre-treatment adve	"Y" if [ADAE
ADAE_L1	Yes	1060	ASTDT	Analysis Start Date		N	*ALL*		Start date of adver	AE.AESTDT(
ADAE_L1	Yes	1062	ASTDTF	Analysis Start Date Imputatio	(DATEFL)	C	*ALL*		Imputation flag for	ADAE.AESD
ADAE_L1	Yes	1063	AENDT	Analysis End Date		N	*ALL*		End of adverse eve	AE.AEENDT(
ADAE_L1	Yes	1065	AENDTF	Analysis End Date Imputation	(DATEFL)	C	*ALL*		Imputation flag for	ADAE.AEED

Building the XML

- Many ways to do this, among them
  - SAS Clinical Standards Toolkit
  - Brute force: Macros, DATA steps
    - Benefits: extreme flexibility with respect to order of dataset display, control of Comments content, selection of XSL, etc. Also, tool (macros) can perform XML validation, create ZIP file of deliverables
    - Drawbacks: lots of code; has to be responsive to changes in the standards

A Word About XSL

- Before writing your own XSL, consider …
- Different type of language: badly shaped learning curve (for most of us)
- Think about functionality to provide over and above CDISC-supplied files
  - Table sorting, printing
  - Additional navigation (next/previous table, etc.)
- Consider whether the sponsor will *accept* the XSL (ActiveX, JavaScript, security considerations)

Building (or not) the XSL

- XSL transforms XML into other formats (HTML is the most common) and makes the XML reader friendly.
- Since the define XML is in a predictable format, transformation of *any* file for *any* study can be done with a standard XSL file (the “XML Promise”)
- The XSL is identified by a reference in the XML:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="define.xsl"?>
```

- Your choice:
  - Use XSL found in the CDISC pilots
  - Write your own (as with define.XML: flexibility, at the cost of writing a lot of code)

Sample XSL from Early CDISC Pilot

```
<!-- ******
<!-- Code L
<!-- ******
<xsl:if
test="/odm:ODM/odm:Study/o
        CodeList[odm:CodeList
    <div id="decodelist">
    <xsl:for-each
select="/odm:ODM/odm:Study/odm:Met
                        CodeList[
     <fieldset>
        <xsl:attribute name="id">CL.<xsl:value-of
            select="@OID"/></xsl:attribute>
          <legend>Code List - <xsl:value-of
select="@Name"/>,
                                Reference Name
(<xsl:value-of
```

Element selection requires knowledge of XPath

Syntax resembles XML

Inclusion of "pure" HTML

The XSL can build HTML statements

Coding of XSL can dramatically affect transformation and readability of an XML file, as shown in next slides …

define.xml: Style Sheet 1

Annotated Case Report Form
- SDTM Datasets
- Value Level Metadata
- Controlled Terms
- External Dictionaries
- Computational Algorithms

SDTM Datasets for Study CDISCPILOT01

Dataset	Description	Class	Structure	Purpose	Keys	Location
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCD	te.xpt
TI	Trial Inclusion/ Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM	tv.xpt
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
SE	Subject Elements	Special Purpose	One record per actual Element per subject	Tabulation	STUDYID, USUBJID, ETCD	se.xpt
SV	Subject Visits	Special Purpose	One record per actual visit per subject	Tabulation	STUDYID, USUBJID, VISITNUM	sv.xpt
CM	Concomitant Medications	Interventions	One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC	cm.xpt
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC	ex.xpt
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AETERM, AESTDTC, AESEQ	ae.xpt
DS	Disposition	Events	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSDECOD, DSSTDTC	ds.xpt
MH	Medical History	Events	One record per medical history event per subject	Tabulation	STUDYID, USUBJID, MHTERM, MHSTDTC	mh.xpt
LB	Laboratory Tests Results	Findings	One record per analyte per planned time point number per time point reference per visit per subject	Tabulation	STUDYID, USUBJID, LBTESTCD, VISITNUM	lb.xpt
QS	Questionnaires	Findings	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSTESTCD, VISITNUM	qs.xpt
SC	Subject Characteristics	Findings	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD	sc.xpt
VS	Vital Signs	Findings	One record per vital sign measurement per time point per visit per subject	Tabulation	STUDYID, USUBJID, VSTESTCD, VISITNUM, VSTPTNUM	vs.xpt
RELREC	Related Records	Relationship	One record per related record, group of records or datasets	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, RELID, RELTYPE	relrec.xpt
SUPPAE	Supplemental Qualifiers for AE	Relationship	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	suppae.xpt
SUPPDM	Supplemental Qualifiers for DM	Relationship	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	suppdm.xpt

The difference is in the HTML created by the XSL, *not* in the XML itself!

define.xml: Style Sheet 2

Analysis Datasets for Study CDISC_Pilot

Dataset	Description	Structure	Purpose	Keys	Location
ADAE	Adverse Event Analysis	Analysis - one record per adverse event per subject	Analysis	USUBJID, AESEQ	adae.xpt
ADLBC	Chemistry Lab Analysis	Analysis - one record per lab test per visit per subject	Analysis	USUBJID, VISITNUM, LBTESTCD	adlbc.xpt
ADLBH	Hematology Lab Analysis	Analysis - one record per lab test per visit per subject	Analysis	USUBJID, VISITNUM, LBTESTCD	adlbh.xpt
ADLBHY	Hy's Law Lab Analysis	Analysis - one record per visit per subject	Analysis	USUBJID, VISITNUM	adlbhy.xpt
ADQSADAS	ADAS-Cog Analysis	Analysis - one record per parameter per analysis visit per subject	Analysis	USUBJID, VISITNUM, AVISITN	adqsadas.xpt
ADQSCIBC	CIBIC+ Analysis	Analysis - one record per parameter per analysis visit per subject	Analysis	USUBJID, VISITNUM, AVISITN	adqscibc.xpt
ADQSNPIX	NPIX Analysis	Analysis - Analysis dataset standards are still evolving, and most of the datasets from the original submission have been revised to follow that evolution. However, this dataset has not been revised, and it was decided that the dataset and associated metadata not be included in the published package.	Analysis	USUBJID	adqsnpix.xpt
ADSL	Demog. and Baseline Char. Analysis	Analysis - one record per subject	Analysis	USUBJID	adsl.xpt
ADTTE	AE Time To 1st Derm. Event Analysis	Analysis - one record per subject	Analysis	USUBJID	adtte.xpt
ADVS	Vital signs Analysis	Analysis - Analysis dataset standards are still evolving, and most of the datasets from the original submission have been revised to follow that evolution. However, this dataset has not been revised, and it was decided that the dataset and associated metadata not be included in the published package.	Analysis	USUBJID	advs.xpt

The difference is in the HTML created by the XSL, *not* in the XML itself!

Did We Get It Right? Validating the XML

- Recall define.pdf v. define.xml discussion: different, more stringent and definable validation requirements
- Ensures names/values, attributes, occurrences, order of nodes conform to the schema.
- But we *can't* validate that the data makes sense!
  - Var. length of 20 may be valid according to the schema, but if length in the dataset was >20, problem lies elsewhere
- Tools
  - OpenCDISC
  - SAS Clinical Standards Toolkit
  - XML4Pharma CDISC Define.xml Checker
  - Home-grown (specialized, client-requested checks)

Validation: OpenCDISC V1.3 Rules

http://www.opencdisc.org/projects/validator/cdisc-define.xml-1.0-valid

CDISC Define.xml 1.0 Validation Rules

The following is a listing of CDISC Define.xml version 1.0 validation rules implemented in OpenCDISC Validator. The rules are the result of work performed by OpenCDISC team and community.

Level of severity is arguable!

Category: <Any> Severity: <Any> APPLY

Rule ID	Message	Description	Category	Severity
DD0002	Missing or invalid namespace reference	In addition to default CDISC ODM namespace, the Define.xml must reference XML Schema Instance (xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"), XLink (xmlns:xlink="http://www.w3.org/1999/xlink"), CDISC Define (xmlns:def="http://www.cdisc.org/ns/def/v1.0") namespaces.	Structure	Error
DD0003	Required attribute is missing or empty	Required attributes must be included in Define.xml and cannot be empty.	Presence	Error
DD0004	Unknown attribute	The Define.xml must contain only attributes defined in the standard.	Metadata	Warning
DD0005	Attribute in wrong position within Define.xml	An attribute was found in the wrong position within Define.xml.	Metadata	Error
DD0006	Required element is missing or empty	Required elements must be included in Define.xml and cannot be empty.	Presence	Error
DD0007	Unknown element	The Define.xml must contain only elements defined in the standard.	Metadata	Warning
DD0008	Element in wrong position within Define.xml	An element was found in the wrong position within Define.xml.	Metadata	Error
DD0009	Element occurs more than once	Element occurs in Define.xml more than once, but it is limited to only one occurrence by the specification.	Consistency	Error
DD0010	ODM attribute should not be included in Define.xml	ODM attributes that are optional and are not explicitly specified in Define.xml standard should not be included.	Metadata	Warning

Validation: OpenCDISC Results (Summary)

Validation report has become part of our deliverables to the client. Inclusion of any item flagged as an Error or Warning must be explained.

File Edit View Insert Format Tools Data

Arial 10

B31

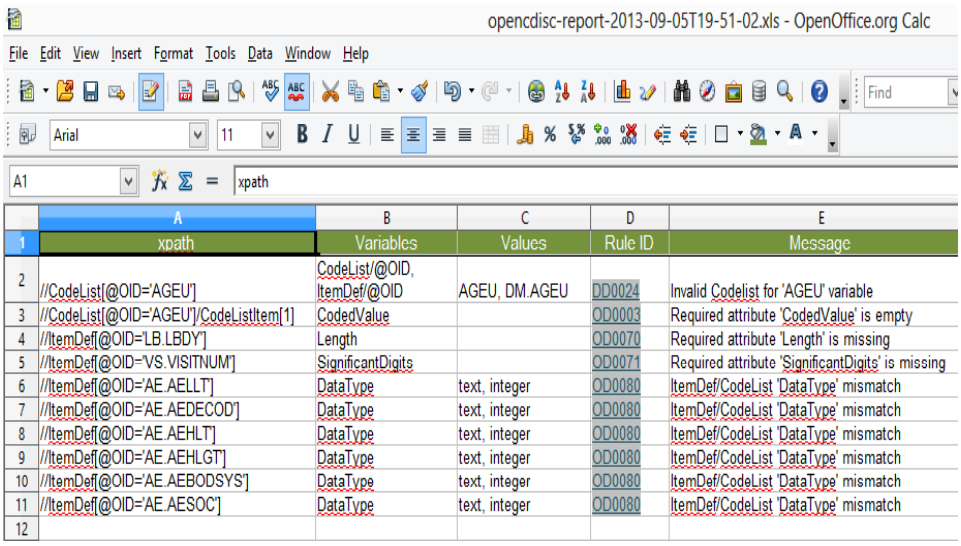
OpenCDISC Validator Report

Configuration: C:\OpenCDISC\OpenCDISC1.3\config\config-define-1.0.xml
Define.xml: C:\defineXML\examples\Pilot2SDTM-contaminated.xml
Generated: 2013-09-05T19:51:02

Issue Summary

Source	Rule ID	Message	Severity	Found
DEFINE				
	DD0024	Invalid Codelist for 'AGEU' variable	Error	1
	OD0003	Required attribute 'CodedValue' is empty	Error	1
	OD0070	Required attribute 'Length' is missing	Error	1
	OD0071	Required attribute 'SignificantDigits' is missing	Error	1
	OD0080	ItemDef/CodeList 'DataType' mismatch	Error	6

Validation: OpenCDISC Results (Detail)



	A	B	C	D	E
1	xpath	Variables	Values	Rule ID	Message
2	//CodeList[@OID='AGEU']	CodeList/@OID, ItemDef/@OID	AGEU, DM.AGEU	OD0024	Invalid Codelist for 'AGEU' variable
3	//CodeList[@OID='AGEU']/CodeListItem[1]	CodedValue		OD0003	Required attribute 'CodedValue' is empty
4	//ItemDef[@OID='LB.LBDY']	Length		OD0070	Required attribute 'Length' is missing
5	//ItemDef[@OID='VS.VISITNUM']	SignificantDigits		OD0071	Required attribute 'SignificantDigits' is missing
6	//ItemDef[@OID='AE.AELLT']	DataType	text, integer	OD0080	ItemDef/CodeList 'DataType' mismatch
7	//ItemDef[@OID='AE.AEDECOD']	DataType	text, integer	OD0080	ItemDef/CodeList 'DataType' mismatch
8	//ItemDef[@OID='AE.AEHLT']	DataType	text, integer	OD0080	ItemDef/CodeList 'DataType' mismatch
9	//ItemDef[@OID='AE.AEHLGT']	DataType	text, integer	OD0080	ItemDef/CodeList 'DataType' mismatch
10	//ItemDef[@OID='AE.AEBODSYS']	DataType	text, integer	OD0080	ItemDef/CodeList 'DataType' mismatch
11	//ItemDef[@OID='AE.AESOC']	DataType	text, integer	OD0080	ItemDef/CodeList 'DataType' mismatch
12					

You're Not Done Yet: define.pdf

- You mean define.**xml**
- No, define.**pdf** – a PDF rendering of the XML
- Why (oh why, oh why, …?)
- How
  - Read the XML with SAS XML maps, then use REPORT for the various pieces (Jansen paper)
  - iText open source library (Java)
  - XSL-FO (**F**ormatting **O**bjects) document description language
  - Our old friend, Brute Force *(next slide)*

define.pdf: Brute Force, No Finesse

defineXML.sas

```
data work.defpdf_value;
     set work.value;
     … write value-level XML …
```

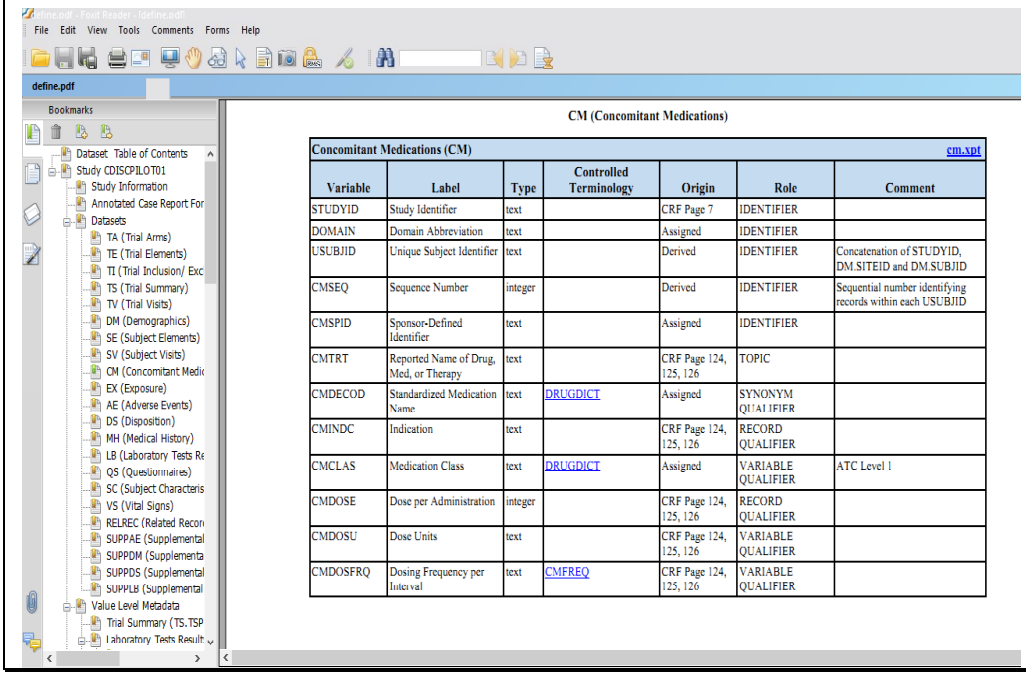
*defineXML**PDF**.sas*

```
… ODS PROCLABEL, other …
proc report data=work.defpdf_value;
```

Calling Program

```
%setup(project=study)
%defineXML(…parameters…)
%defineXMLPDF(…parameters…)
```

define.pdf: define.xml Transformed



CM (Concomitant Medications)

Concomitant Medications (CM) cm.xpt

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		CRF Page 7	IDENTIFIER	
DOMAIN	Domain Abbreviation	text		Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Derived	IDENTIFIER	Concatenation of STUDYID, DM.SITEID and DM.SUBJID
CMSEQ	Sequence Number	integer		Derived	IDENTIFIER	Sequential number identifying records within each USUBJID
CMSPID	Sponsor-Defined Identifier	text		Assigned	IDENTIFIER	
CMTRT	Reported Name of Drug, Med, or Therapy	text		CRF Page 124, 125, 126	TOPIC	
CMDECOD	Standardized Medication Name	text	DRUGDICT	Assigned	SYNONYM QUALIFIER	
CMINDC	Indication	text		CRF Page 124, 125, 126	RECORD QUALIFIER	
CMCLAS	Medication Class	text	DRUGDICT	Assigned	VARIABLE QUALIFIER	ATC Level 1
CMDOSE	Dose per Administration	integer		CRF Page 124, 125, 126	RECORD QUALIFIER	
CMDOSU	Dose Units	text		CRF Page 124, 125, 126	VARIABLE QUALIFIER	
CMDOSFRQ	Dosing Frequency per Interval	text	CMFREQ	CRF Page 124, 125, 126	VARIABLE QUALIFIER	

Closing Comments

- The process to create define.xml is more complex than define.pdf:
  - New technologies
  - More “moving partss” – metadata, XML, XSL, …
  - Stringent validation
- Keys:
  - Organizational commitment
  - Transparent access to robust metadata
  - Tools that facilitate flexible display (especially important to CROs)

Thank You!

Your comments are valued and encouraged:

frank@CodeCraftersInc.com