Mapping a Veeva SDS specification to ODM-XML

PHUSE EU Connect 2024

Strasbourg Convention Centre

Jørgen Mangor Iversen

11 November 2024

The views and opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the official stance or opinions of Veeva Systems.
The information presented is based on personal expertise and experience.

The views and opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the official stance or opinions of Novo Nordisk.
The information presented is based on personal expertise and experience.

*No artificial intelligence was harmed in the making of this presentation*

Agenda

BACKGROUND

Strategy

- The Veeva Study Design Specification (SDS) document is the main (only?) way of exporting CRF metadata from Veeva EDC

- ODM-XML is the main (only?) way of importing CRF metadata into (Open) Study Builder

- Hence the desire to map an SDS to an ODM-XML

- ODM-XML is a long term CDISC standard

  - Version 1.3.2 has been stable for many years

  - Version 2.0 was released in 2023, yet to be adopted by vendors and authorities
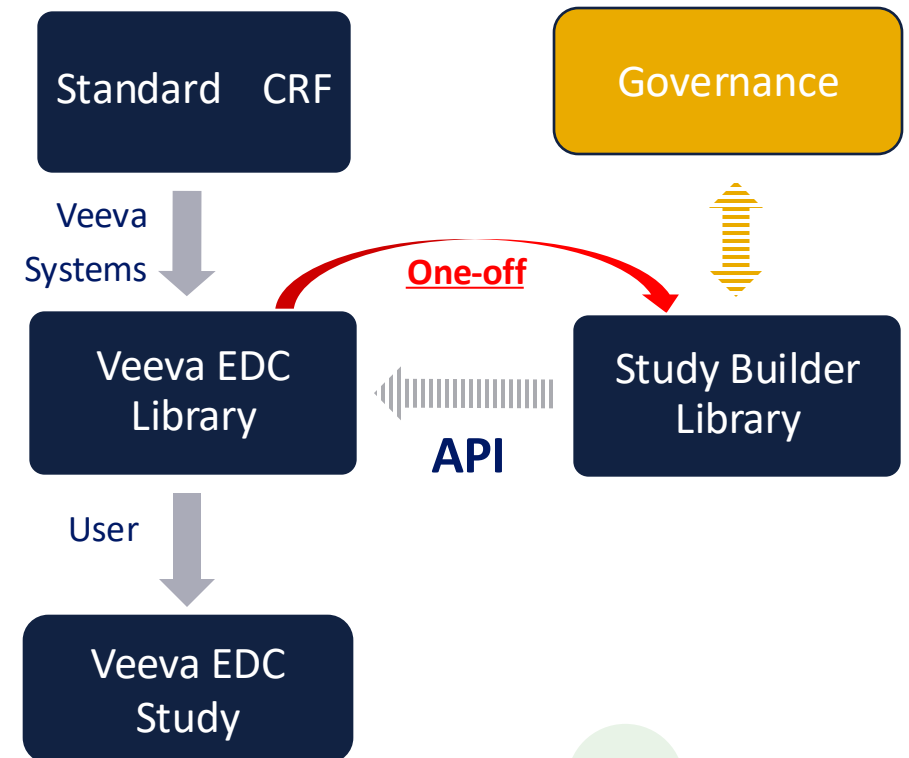
BACKGROUND

What is Veeva SDS

- Excel file containing many sheets exported from Veeva EDC

- Small subset of sheets contains CRF information

  - Summary

  - Schedule - Tree

  - Form Definitions

  - Code lists

  - Unit Code lists

  - Rules contains edit checks only indirectly in scope

As of: 2024-03-13T13:28:44	
Summary	
Vault	Novo Nordisk - CDMS - Development
Study	NN Veeva Standards Library_DEV1
Casebook Version Number	1
Study Build Version	1
Version	1.1
Description	
Change Reason	
External ID	InitialVersion
Specification Version	
Document Number	

BACKGROUND

What is ODM-XML

- XML file format defined by CDISC

- Formally defined as a superset of both CRF definitions and data definitions (define.xml)

- Contains in its pure form all metadata for CRF's

- Can contain much more including

  - Visit structure

  - User information

  - Reference data

  - Clinical data

  - Digital signatures

```xml
<?xml version="1.0" encoding="UTF-8"?>
<ODM FileOID         = "T_egALL_FORMS"
     FileType        = "Snapshot"
     CreationDateTime = "2024-04-18T16:14:06"
     AsOfDateTime    = "2024-04-09T11:28:11"
     ODMVersion      = "1.3.2"
     Originator      = "Novo Nordisk"
     SourceSystem    = "SAS Enterprise Guide: sds2odm.sas"
     xmlns:xml       = "http://www.w3.org/XML/1998/namespace"
     xmlns           = "http://www.cdisc.org/ns/odm/v1.3">
<Study OID="S_evALL_FORMS">
  <GlobalVariables>
    <StudyName>NN_Veeva_Standards_Library_DEV1</StudyName>
    <StudyDescription>NN Veeva Standards Library_DEV1</StudyDescription>
    <ProtocolName>NN_Veeva_Standards_Library_DEV1</ProtocolName>
  </GlobalVariables>
  <BasicDefinitions>
    <MeasurementUnit OID="Unit 1" Name="DOSE STEP">
      <Symbol>
        <TranslatedText>Unit 1</TranslatedText>
      </Symbol>
    </MeasurementUnit>
    <MeasurementUnit OID="mmol/L" Name="MMOL_L">
      <Symbol>
        <TranslatedText>mmol/L</TranslatedText>
      </Symbol>
    </MeasurementUnit>
  </BasicDefinitions>
  <MetaDataVersion OID="NN_Veeva_Standards_Library_DEV1" Name="NN Veeva
Standards Library_DEV1">
```

Global data

- **Summary** sheet contains good candidates for high-level identifiers (**StudyName**, **MetaDataVersion OID**)

- **Schedule – Tree** sheet contains possible candidates for other high-level identifiers (**FileOID**, **StudyOID**)

- The whole CRF structure is found in a single sheet

  - Relationships between **Forms**, **Sections**, and **Questions** must be deduced across rows

  - Mandatory sections are not explicitly marked

Event Group Name ▼	Event Name ▼
egALL_FORMS	
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS
egALL_FORMS	evALL_FORMS

Order Numbers

- Order numbers are not explicitly present

  - The ordering of everything relies only on the sequence of rows

  - Be careful when sorting data!

  - **Forms**, **Sections**, **Questions** are interleaved in the same sheet

  - ODM-XML has no requirements to **OrderNumber** beyond ordering, but humans like to start over for each form and section

- Row band colors are not machine readable by Excel importer used

Form Name	Item Group Name	Item Name
MH		
MH	MH	
MH	MH	MHCAT
MH	MH	MHTERM_ALLERGY
MH	MH	MHTERM_ALLEROTH
MH	MH	MHTERM_BREAST
MH	MH	MHTERM_BREASOTH
MH	MH	MHTREM_CVDP
MH	MH	MHTERM_CVDPOTH
MH	MH	MHTERM_CAS
MH	MH	MHTERM_CASOTH
MH	MH	MHTERM_STROKE
MH	MH	MHTERM_STROKOTH
MH	MH	MHTERM_PERI
MH	MH	MHTERM_PERIOTH
MH	MH	MHTERM_AF
MH	MH	MHTERM_AFOTH
MH	MH	MHTERM_CVD
MH	MH	MHTERM_CVDOTH
MH	MH	MHTERM_DIAB
MH	MH	MHTERM_DIABOTH
MH	MH	MHTERM_DYSLI

Identifiers

- Veeva SDS uses **Name** as the primary (visible) key

- SDS **External ID** a candidate for ODM-XML **OID**s

  - Independent of SDS references

  - Fewer restrictions on content (periods)

- Alignment of keys (**External ID** vs. **OID**) is essential

  - Matching existing contents when importing

  - Enabling software to compare CRFs

  - Mapping between SDS and ODM-XML

Form Name ▼	Item Group Name ▼	Item Name ▼	External ID ▼
MH			**MH**
MH	**MH**		**MH**
MH	MH	MHDIAG	MHDIAG
MH	MH	MHALRG	MHALRG
MH	MH	MHASPC_OTH	MHASPC_OTH
MH	MH	MHBREAST	MHBREAST
MH	MH	MHBRST_OTH	MHBRST_OTH
MH	MH	MHCVP	MHCVP
MH	MH	MHCHD	MHCHD
MH	MH	MHSTEN	MHSTEN
MH	MH	MHMI	MHMI
MH	MH	MHSTRK	MHSTRK
MH	MH	MHPAD	MHPAD

More on Identifiers

- StudyBuilder is very particular regarding key's values for matching previous content

  - MeasurementUnit **OID** must match internal values exactly, always! New units cannot be created when importing an ODM-XML

  - CodeList **OID** must match internal values exactly when referring existing code lists

  - CodeListItem must match an **OID** and a **Name** in the osb name space exactly when referring existing terms

- Consequently, any ODM-XML for import must have prior knowledge of StudyBuilder's preexisting contents

- When **Name** and **External ID** are set by Veeva SDS, a mapping must occur to translate them into StudyBuilder **OID**s

Novo Nordisk®

Data types

- Most data types are straight forward
  - Everything numeric having decimals are floating point
  - Units are integer (except floating point rule)
  - Code lists suffer from insufficient definition
    - Inspect values as complete list of values:
      - If any alphabetic characters, they are text
      - Else they are integer (except floating point rule)
- SDS has a label data type, supported by StudyBuilder as a comment
- Otherwise, minor change of case and name

SDS Data type	ODM Data type
Has decimals	float
Is a unit	integer (float if dec.)
Is a code list	Inspect terms for alphabetic characters
Boolean	boolean
Date	date
Date/Time	datetime
Number	integer
Text	text or string
Form Link	text or string
Label	comment

PRIMARY MAPPINGS

Mappings

SDS: Form Definitions	ODM: FormDef
Form Name	@OID
Form Label[1]	@Name
Repeats	@Repeating
Form Short Label[1]	@osb:instruction
Hover Help[1]	@osb:sponsorinstruction
Description[1]	<Description>

SDS: Form Definitions	ODM: ItemGroupDef
Item Group Name	@OID
Item Group Label[1]	@Name
IG Rep	@Repeating
Description[1]	<Description>

SDS: Form Definitions	ODM: ItemDef
Item Name	@OID
Item Name	@Name
Data Type	@DataType[2]
Length	@Length[2]
Decimal	@SignificantDigits[2]
Label[1]	<Question>
Codelist	<CodeListRef>[2]
Unit Codelist	<MeasurementUnitRef>[2]
Hint Label[1]	<Description>
Hover Help[1]	@osb:instruction
Description[1]	@osb:sponsorinstruction

1) Requires HTML encoding due to contents not allowed in XML

2) Requires additional processing

Mappings

SDS: Form Definitions	ODM: ItemGroupRef
Item Group Name	@ItemGroupOID
Display Format = 'Form'	@Mandatory[2]
[calculated]	@OrderNumber[2]

SDS: Form Definitions	ODM: ItemRef
Item Name	@ItemOID
Required	@Mandatory
[calculated]	@OrderNumber[2]

2) Requires additional processing

Mappings

SDS: Schedule - Tree	ODM: Study
Event Group Name	@FileOID
datetime()	@CreationDateTime
=Summary'!A1	@AsOfDateTime
Event Name	@Study_OID
=Summary'!B4	<GlobalVariables>

SDS: Codelists	ODM: CodeList
[calculated] or Name	@OID[2]
[calculated]	@Name[2]
[calculated from ItemDef]	@DataType[2]
Name	@SASFormatName
Description[1]	<Description>

SDS: Unit Codelists	ODM: MeasurementUnit
[calculated]	@OID[2]
Name	@Name
Choice Label[1]	<Symbol>
"1.0"	@Version

SDS: Codelists	ODM: CodeListItem
Choice Code[1]	@CodedValue
[calculated][2]	@OrderNumber
Choice Label[1]	<Decode>

1) Requires HTML encoding due to contents not allowed in XML

2) Requires additional processing

Vendor specific data

- Most of the SDS content is vendor specific beyond the CRF

- The XML definition allows for vendor specific name spaces

- The ODM definition allows for **Alias** tags having vendor specific **Content** attributes

- Both are equally valid options to carry vendor specific data

- Both are safe to omit without violating the ODM-XML validity

Name Space

<any_tag veeva:example="data value">

Alias

<Alias Context="Veeva"> Name="data value"/>

Edit checks

- Edit checks are specified on the **Rules** sheet

  - Formulated in a Veeva specific language, Veeva Clinical Query Language (CQL)

  - The **Rules** contents seems to be machine generated

  - A syntax for this language must exist or can easily be inferred

  - A cross compiler can be written to convert CQL to and from any rule definition language i.e., yaml

```
#define AMERICAN_INDIAN @Form.DM.AMERICAN_INDIAN
#define ASIAN @Form.DM.ASIAN
#define BLACK @Form.DM.BLACK
#define NATIE_HAWAIIAN @Form.DM.NATIE_HAWAIIAN
#define WHITE @Form.DM.WHITE
#define NOT_REPORTED @Form.DM.NOT_REPORTED


AMERICAN_INDIAN != true
&&
ASIAN != true
&&
BLACK != true
&&
NATIE_HAWAIIAN != true
&&
WHITE != true
&&
WHITE != true
&&
NOT_REPORTED != true
```

Data errors

- SDS allows for characters that ODM-XML does not

  - Spaces in identifiers to be **OID**s

  - Double quotes inside ordinary text (html encode)

  - Text quoting 'Unknown' instead of ASCII quoting 'Unknown' (html encode)

  - Binary characters (00x-1Fx)

- This may not be a generic issue

Examples
OID="NN Veeva Standards Library_DEV1"
Report time as "?" if diary response is 'Unknown' for time of episode.
Peripheral artery stenosis (≥50% stenosis)

ODM omissions

- Date and time for creation

- Explicit Mandatory flag for sections

- Explicit order numbers

- Explicit data types for questions having code lists

- Explicit data types for questions having units

- Language indication for questions and descriptions

ODM	SDS
CreationDateTime	Date time of program run or extract from file system
OrderNumber	Infer from row ordering
ItemGroupRef Mandatory	Assume Display Format = Form
Data type	Infer from decimals or code list values
xml:lang	Not present

Conclusion

- Mapping a Veeva SDS document to a CDISC ODM-XML document is doable

- Mapping allows for conversion between the two, both ways

- Transformations between the two are not substantial

- Discrepancies between the two are bridgeable

- Be aware of the smaller details regarding relationships, ordering, and data types

- Vendor specific data can be implemented as either XML name space or ODM extensions such as **Alias**

- Implementation of ODM-XML version 2.0 offers a golden opportunity to align

Thanks