Linking Metadata from CDASH to ADaM

João Gonçalves
Agenda

- Missing Links
- Links Enable Automation
Metadata Links

- Content of standard variables often involves source variables
  - SDTM is created based on CDASH, calculations and rules
  - ADaM is created based on copied SDTM and calculations

- Metadata links are source variable metadata
Missing Links

ADaM

Character description of a grouping or pooling of the subject’s age for analysis purposes.

AGEGRy

DM.AGE AGE

DM.AGEU AGEU
Missing Links

SDTM

AGE

AGEU

ADaM

AGEGRy

Character description of a grouping or pooling of the subject’s age for analysis purposes.

AGE

AGEU

DM.AGE AGE

DM.AGEU AGEU
Missing Links

SDTM

AGE

SDTM

AGE

Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC.

Units associated with AGE.

ADaM

AGEGRy

AGE

AGEU

ADaM

AGE

AGEU

Restricted © Business & Decision Life Sciences 2016 All rights reserved.
Missing Links

CDASH

SDTM

ADaM

AGE

AGEU

RFSTDTC

BRTHDTC

AGEGRy

Reference Start Date/time for the subject in ISO 8601 format. Usually equivalent to date/time when subject was first exposed to study treatment.

Date/time of birth of the subject.
Missing Links

CDASH
VISDAT?
BRTHDAT
AGE
AGEU

SDTM
RFSTDTC
BRTHDTC
AGE
AGEU

ADaM
AGEEGRY
AGE
AGEU
## Missing Links

<table>
<thead>
<tr>
<th>CDASH</th>
<th>SDTM</th>
<th>ADaM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VISDAT?</td>
<td>Reference Start Date/time for the subject in ISO 8601 format. Usually equivalent to date/time when subject was first exposed to study treatment.</td>
<td>Character description of a grouping or pooling of the subject’s age for analysis purposes.</td>
</tr>
<tr>
<td>BRTHDAT</td>
<td>Date/time of birth of the subject.</td>
<td>DM.AGE AGE</td>
</tr>
<tr>
<td>AGE</td>
<td>Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC.</td>
<td>DM.AGEU AGEU</td>
</tr>
<tr>
<td>AGEU</td>
<td>Units associated with AGE.</td>
<td>AGEGRy</td>
</tr>
</tbody>
</table>

**AGEU**

- Character Description of a grouping or pooling of the subject’s age for analysis purposes.
### Metadata Links In CDISC Implementation Guides

#### Source variables are specified in the CDISC Notes
- **Textual description** of the calculations or source variable to be copied, or other business rules

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Type</th>
<th>Codelist/Controlled Terms</th>
<th>Core</th>
<th>CDISC Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>Age</td>
<td>Num</td>
<td></td>
<td>Req</td>
<td>DM.AGE. If analysis needs require a derived age that does not match DM.AGE, then AAGE must be added</td>
</tr>
<tr>
<td>AGEU</td>
<td>Age Units</td>
<td>Char</td>
<td>(AGEU)</td>
<td>Req</td>
<td>DM.AGEU</td>
</tr>
<tr>
<td>AGEGRy</td>
<td>Pooled Age Group y</td>
<td>Char</td>
<td></td>
<td>Perm</td>
<td>Character description of a grouping or pooling of the subject’s age for analysis purposes. For example, AGEGR1 might have values of “&lt;18”, “18-65”, and “&gt;65”; AGEGR2 might have values of “Less than 35 y old” and “At least 35 y old”.</td>
</tr>
</tbody>
</table>
Metadata Links In Define-XML

- **Define-XML format**
  - **Machine-readable** metadata description
  - Used for submissions but also to **describe standards**
    - eSHARE Downloads page
    - SHARE API

- **Define-XML predecessor links**
  - Defined as a **textual description** of the variable

- **Define-XML computational method links**
  - Defined as a **method** with a **textual description**, **linked** to the variable reference

- **Links are not machine-readable in Define-XML!**
Define-XML Predecessor Link

```
<ItemDef OID="IT.ADSL.AGEU" Name="AGEU" SASFieldName="AGEU"
DataType="text" Length="5">
  <Description>
    <TranslatedText xml:lang="en">Age Units</TranslatedText>
  </Description>
  <CodeListRef CodeListOID="CL.AGEU"/>
  <def:Origin Type="Predecessor">
    <Description>
      <TranslatedText xml:lang="en">DM.AGEU</TranslatedText>
    </Description>
  </def:Origin>
</ItemDef>
```

Name of the predecessor variable as text
Define-XML Computation Method Link

OID of the computational method defined

Definition of the computational method (with an OID)

```xml
<MethodDef OID="MT.ADSL.AGEGR1" Name="MT.ADSL.AGEGR1" Type="Computation">
  
  <Description>
    <TranslatedText xml:lang="en">AGEGR1 = "&lt;18" if AGE &lt;18.
AGEGR1 = "&gt;=18-&lt;25 " if AGE 18-24. AGEGR1 = "&gt;=25-&lt;65 " if AGE 25-64. AGEGR1 = "&gt;=65" if AGE &gt;=65.</TranslatedText>
  
  </Description>

</MethodDef>
```

Pseudo code / description of the method as text
Mapping Specifications

Industry practice to represent source variable metadata
- Semi-structured format, usually in Excel, organization-dependent
- Allows basic machine-readability
  - Metadata traceability checks via highlighting formulae

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>VARIABLE</th>
<th>LABEL</th>
<th>TYPE</th>
<th>LN</th>
<th>ROLE</th>
<th>CT</th>
<th>METHOD</th>
<th>ORIGIN</th>
<th>SOURCE DOMAIN</th>
<th>SOURCE VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>AGE</td>
<td>Age</td>
<td>num</td>
<td>8</td>
<td>Analysis</td>
<td></td>
<td></td>
<td>DM</td>
<td>DM</td>
<td>AGE</td>
</tr>
<tr>
<td>ADSL</td>
<td>AAGE</td>
<td>Analysis Age</td>
<td>num</td>
<td>8</td>
<td>Analysis</td>
<td>ADSL.AAGE</td>
<td>Derived</td>
<td>ADSL</td>
<td>AAGE</td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEU</td>
<td>Age Units</td>
<td>char</td>
<td>5</td>
<td>Analysis</td>
<td>AGEU</td>
<td></td>
<td>DM</td>
<td>DM</td>
<td>AGEU</td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEGR1</td>
<td>Age Group 1</td>
<td>char</td>
<td>20</td>
<td>Analysis</td>
<td>AGEGR1L</td>
<td>ADSL.AGEGR1</td>
<td>Derived</td>
<td>ADSL</td>
<td>AAGE</td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEGR2</td>
<td>Age Group 2</td>
<td>char</td>
<td>20</td>
<td>Analysis</td>
<td>AGEGR2L</td>
<td>ADSL.AGEGR2</td>
<td>Derived</td>
<td>ADSL</td>
<td>AAGE</td>
</tr>
</tbody>
</table>
Missing Links in Define-XML

- Links supported in Analysis Results definitions...
  - Each `AnalysisResult` references a `ParameterOID`, relies on a number of `AnalysisDatasets`, which contain `AnalysisVariables` and `WhereClauseRefs` necessary to calculate it.

- Could be similarly supported for ADaM to SDTM and SDTM to CDASH...

```xml
<ItemDef OID="IT.ADSL.AGEGR1" Name="AGEGR1" DataType="text" Length="20">
  <CodeListRef CodeListOID="CL.AGEGR1"/>
  <def:Origin Type="Derived"/>
  <ext:SourceDatasets>
    <def:DocumentRef leafID="LF.SDTM-Define"/>
    <ext:SourceDataset ItemGroupOID="IG.DM">
      <ext:AnalysisVariable ItemOID="IT.DM.AGE"/>
    </ext:SourceDataset>
  </ext:SourceDatasets>
</ItemDef>
```
Why machine-readable metadata links?

Automation!
Generate ADaM datasets and TFLs based on study metadata and standard Macros

- Selected 3 common tables, which require 3 ADaM datasets (ADSL, ADVS, ADAE)
- Goal: decouple study-specific metadata and code
ADaM Standard Macros

Re-usable across studies and within the same study
- Study-specific configuration moved to macro parameters

```bash
%MSDTM_AGEGRy(
    AGE_GROUP_NUMBER=1,
    AGE_GROUP='<18 years;>=18-<25 years;>=25-<65 years;',
    AGE_GROUP_LIST='<18;<25;<65;>65',
    AGE_GROUP_OTHER='>=65 years'
);

%MSDTM_AGEGRy(
    AGE_GROUP_NUMBER=2,
    AGE_GROUP='<41;>= 41 and < 65;',
    AGE_GROUP_LIST='<41;<65;',
    AGE_GROUP_OTHER='>= 65'
);
```
TFL Standard Macros

More complex but same principle for TFLs

%TABLE_DESC_STATS(
    STUDY='XX YY mg (STUDY_ID_001)',
    NAME='Table 14.1.5.1',
    DESCRIPTION='Demographics - Intention-To-Treat population',
    VARIABLES='ADSL.AGE;ADSL.AGEGR2;ADSL.SEX;ADSL.RACE;ADSL.HEIGHT;ADSL.WEIGHT;ADSL.BMI;',
    GROUPS='ADSL.ACTARMCD=A;ADSL.ACTARMCD=F;',
    GROUP_LABEL_VARIABLE='ADSL.ACTARM',
    WHERE='ADSL.ITTFL=Y',
    VARIABLE_FOOTNOTES='ADSL.WEIGHT;Weight at baseline (Week 1). If weight is missing at baseline, weight at screening is used;ADSL.BMI;Body Mass Index (BMI) (kg/m2) = Weight (kg) / [Height (m)]2;',
    FOOTNOTES='N = Number of subjects in specific group. n = Number of subjects. Calculation of percentages based on N.'
);
Auto-generation Proof of Concept – Step #2

User Interaction
Metadata
Data

Study Input Selector

Selected TFLs

Study-Specific Input

Library Metadata

Metadata-driven Script Generator

SDTM Datasets
ADaM Generator
ADaM Datasets
TFL Generator
TFLs

ADaM Standard Macros
TFL Standard Macros

Restricted © Business & Decision Life Sciences 2016 All rights reserved.
## Library-level Computational Methods

### Computational Methods that abstract study-specific input

- A computational method is implemented by a Standard Macro
- Macro generalization exercise identified study-specific input

<table>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
<th>Pseudo Code</th>
<th>Standard Macro</th>
<th>Study-Specific Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL.AAGE</td>
<td>Equals to ADSL.AGE</td>
<td>Equals to ADSL.AGE</td>
<td>MSDTM_AAGE</td>
<td></td>
</tr>
</tbody>
</table>
| ADSL.AGEGRy | Creation of the age groups. A subsequent number will be used as a suffix to distinguish between the different groupings. | Equals to %AGE_GROUP[n] when ADSL.AAGE < %AGE_GROUP_LIST[n], for the lowest n possible; Equals to %AGE_GROUP_OTHER if ADSL.AAGE > %AGE_GROUP_LIST[n], for the highest n possible | MSDTM_AGEGRy   | %AGE_GROUP_NUMBER - value for the y suffix  
%AGE_GROUP - list or array containing the values to be assigned to the AGEGRy variable, except the one for the highest age group  
%AGE_GROUP_LIST - list or array containing upper age limits in the same order as %AGE_GROUP  
%AGE_GROUP_OTHER - the value to be assigned to the AGEGRy variable for the highest age group |
Study Specification – Select TFLs

- TFLs required for the study are selected by the user

The Study Input Selector script uses Library Metadata to determine required study-specific input

<table>
<thead>
<tr>
<th>Reference</th>
<th>Param1</th>
<th>Param2</th>
<th>Param3</th>
<th>Param4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL.AGEGRy</td>
<td>%AGE_GROUP_NUMBER - value for the y suffix</td>
<td>%AGE_GROUP - list or array containing the values to be assigned to the AGEGRy variable, except the one for the highest age group</td>
<td>%AGE_GROUP_LIST - list or array containing upper age limits in the same order as %AGE_GROUP</td>
<td>%AGE_GROUP_OTHER - the value to be assigned to the AGEGRy variable for the highest age group</td>
</tr>
</tbody>
</table>
The user inputs the required (and possibly optional) study-specific input

<table>
<thead>
<tr>
<th>Reference</th>
<th>Param1</th>
<th>Param2</th>
<th>Param3</th>
<th>Param4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL.AGEGRy 1</td>
<td>&lt;18 years;&gt;=18-&lt;25 years;&gt;=25-&lt;65 years;'</td>
<td>'&lt;18;&lt;25;&lt;65;'</td>
<td>'&gt;65 years'</td>
<td></td>
</tr>
<tr>
<td>ADSL.AGEGRy 2</td>
<td>&lt; 41;= 41 and &lt; 65;'</td>
<td>'&lt;41;&lt;65;'</td>
<td>'&gt; 65'</td>
<td></td>
</tr>
</tbody>
</table>

The Metadata-driven Script Generator uses all the metadata to generate the SAS scripts that will process the data.
Auto-generation Execution

The SAS Scripts can then be run against the SDTM data …

... generating ADaM data and TFLs

<table>
<thead>
<tr>
<th>Table 14.1.5.1: Demographics - Intention-To-Treat population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age Group 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Height at Baseline (cm)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Conclusion

End-to-end metadata linking enables the new age of automation for clinical trials

- Metadata-driven data processing and reporting
  - Reduced manual work (and errors)
- Metadata-driven study specification
  - Specifying the required analyses renders the data that needs to be collected

Linking be easily implemented through Define-XML

- Already existing for Analysis Results Metadata
- Could be extended to link CDASH to SDTM and SDTM to ADaM
Thank you