Automation of ADaM Dataset Creation with a Retrospective, Prospective, and Pragmatic Process

Terek Peterson
Karin LaPann
Agenda

- Traceability Consideration for ADaM
- Retrospective Process
- Prospective Process
- Pragmatic Process
- Conclusions/Questions
• **ret·ro·spec·tive**

  • *adj.*
  • 1. Looking back on, contemplating, or directed to the past.
  • 2. Looking or directed backward.
  • 3. Applying to or influencing the past; retroactive.
Best Traceability Strategy

• Linear Method (Recommended)
  – Logical flow of the software development life cycle
  – Complete traceability from ADaM back to SDTM and CDASH source data
  – The CDASH standard is harmonized with SDTM and therefore assists in assuring end-to-end traceability
Tables First Philosophy

• Look from the Biostatistician’s viewpoint
  – Backwards from the End Goal (CSR)
  – Input at the beginning as to what is collected as described in the protocol
  – The focus goes into the SAP, where the tables are mocked up and final hypotheses will be tested.
Phases of Development

1. Annotate Standard TFL shell to ADaM
2. Create Standard ADaM specifications
   – Add Required and Conditional variables
3. Create Standard SDTM specifications
   – Add Required and Expected variables
4. Pull CDASH inspired EDC Forms
   – Apply any data points that are routinely collected back to the SDTM domains
      • E.g. Prompt questions
Annotate the TFLs to ADaM

<table>
<thead>
<tr>
<th>System Organ Class</th>
<th>Preferred Term</th>
<th>Statistic</th>
<th>Arm A (N=XX)</th>
<th>Arm B (N=XX)</th>
<th>Arm C (N=XX)</th>
<th>Total (N=XX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Adverse Event</td>
<td>n (%)</td>
<td></td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>System Organ Class 1</td>
<td>Preferred Term 1</td>
<td>n (%)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
</tr>
<tr>
<td>System Organ Class 1</td>
<td>Preferred Term 2</td>
<td>n (%)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
</tr>
<tr>
<td>System Organ Class 1</td>
<td>Preferred Term 3</td>
<td>n (%)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
</tr>
<tr>
<td>System Organ Class 1</td>
<td>Preferred Term 4</td>
<td>n (%)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
</tr>
<tr>
<td>System Organ Class 1</td>
<td>Preferred Term 5</td>
<td>n (%)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
</tr>
<tr>
<td>System Organ Class 1</td>
<td>Preferred Term 6</td>
<td>n (%)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
<td>xx (xx.x)</td>
</tr>
</tbody>
</table>

Note: Adverse events are coded using MedDRA version X.X. Only treatment-emergent adverse events are summarized. For each system organ class and preferred term, subjects are included only once, even if they experienced multiple events in that system organ class or preferred term.

Source: Listing xx.x.xx
Dataset: ADAAE, Program: xxxxxx.sas, Output: xxxx.rtf, Generated on: DDMMYYYY HHHMM, Page x

Programming note: Sort by decreasing frequency in the Total column for SOC and preferred term within SOC.
Null banner: No Adverse Events

* Note treatment emergent selection criteria is placed at the top of program

AOCCPFL = "Y"
AOCCSFL = "Y"
Outlines the ADaM Datasets and Variables Needed

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pages</th>
<th>Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>All</td>
<td>TRT01P, TRT01A, ITTFL, SAFFL</td>
</tr>
<tr>
<td>ADSL</td>
<td>1, 2, 3, 6</td>
<td>ENRLFL, RANDFL, SAFFL, ITTFL, PPROTFL, COUNTRY, SITEID, TRT01P, TRT01PN, COMPLFL, DSREASN, DSREAS, TRxxP, APxxSDT, APxxSDTM, APxxEDT, APxxEDTM, SEX, SEXN, RACE, RACEN, ETHNIC, ETHNICN, AGE</td>
</tr>
<tr>
<td>ADDV</td>
<td>4</td>
<td>DVCAT, DVTERM,</td>
</tr>
<tr>
<td>ADEX</td>
<td>5</td>
<td>APERIOD, APERIODC, AVISIT/AVISITN, (ACYCLE, ACYCLEN,) APERSDT, APEREDT,</td>
</tr>
<tr>
<td>ADVS</td>
<td>6</td>
<td>PARAM, AVAL, ABFL</td>
</tr>
<tr>
<td>ADMH</td>
<td>8, 9, 10,</td>
<td>ITTFL, TRTP, TRTPN, MHBODSYS, MHDECOD</td>
</tr>
<tr>
<td>ADMC</td>
<td>11, 12</td>
<td>TRTP, TRTPN, PREFL, CMDECOD, CONCFL</td>
</tr>
<tr>
<td>ADAE</td>
<td>13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36</td>
<td>SAFFL, TRTP, TRTPN, TRTA, TRATAN, AETERM, SEVGR1, TOXGGR1, ATOXGR, ATOXGRN, AREL, SEVGR1, AESDTH, AESER, ADISCFL, AEBODSYS AEDECOD, TRTEMFL, ASEVN, ASEV, ASEVN, AOCCPIFL, AOCCPFL, AEBODSYS, AESOC, AETERM, ASTDT, AENDT, APERIOD, AEACN, AEAOUT, ASTDT, AENDT, ASTDY, AENDY, AESDTH, RELGR1, AESDTH, AESLIFE, AESDISAB, AESHOSP, AESCONG, AESMIE, ADISCFL, AGE, RACE, ETHNIC, SEX, DEATHDT, DEATHFL</td>
</tr>
</tbody>
</table>

- See SDTM Domains needed: DM, AE, CM, VS, MH, DS, EX, DV....
• prospective
  • adj.
  • 1. Likely or expected to happen.
  • 2. Likely to become or be.
Traceability
- The ability to chronologically interrelate uniquely identifiable entities in a way that is verifiable

Transparency
- The assurance that data being reported are accurate and are coming from the official source

Confidence
- The state or quality of being certain
Transparency and Traceability

The ADaM Implementation Guide v1.0 states

Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately traceability in ADaM permits the understanding of the relationship between the analysis results, the analysis datasets, and the SDTM domains.
The system uses these fields in CDASH ending in DY, MO, YR and TIM to appropriately ‘push’ the data to SDTM ISO8601 format, and to the ADaM numeric format as per information stored in the GMDR. Due to standardization of naming conventions, the process is further automated for potential new date elements. The character date is kept for traceability from SDTM to ADaM.
**Example – Issue with Linear**

### Substance Usage - Tobacco (SU)

#### Substance Usage - Tobacco

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Date of Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Current or Former User</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Number of Smoking Years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Number Per Day</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conditional Question**

<table>
<thead>
<tr>
<th></th>
<th>List: NONE</th>
<th></th>
<th>List: SBSTUSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Value</td>
<td>Label</td>
<td>Value</td>
</tr>
<tr>
<td>[Blank]</td>
<td>NONE</td>
<td>[Blank]</td>
<td>CURRENTLY USES</td>
</tr>
<tr>
<td>None</td>
<td>NONE</td>
<td>Currently Uses</td>
<td>FORMER USER</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STUDYID</th>
<th>PATIENTID</th>
<th>SUDAT (Date of Assessment)</th>
<th>SUDSTOT (Number per day)</th>
<th>SUNONE (None)</th>
<th>SUDUR (Number of Smoking years)</th>
<th>SBSTUSE (current or former smoker)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRA101</td>
<td>2005</td>
<td>19JUL2013</td>
<td></td>
<td>NONE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRA101</td>
<td>2009</td>
<td>30JUL2013</td>
<td>3</td>
<td></td>
<td>8</td>
<td>FORMER USER</td>
</tr>
<tr>
<td>PRA101</td>
<td>3006</td>
<td>17JUN2013</td>
<td>5</td>
<td></td>
<td></td>
<td>CURRENTLY USES</td>
</tr>
</tbody>
</table>
TFLs Can Improve Collection

- To create the categories on the mock, need to combine the variables so that 'None' becomes 'Never Smoked'
- Table mock labels for the answers do not match the collection Controlled Terminology (CT)
- Solution: Data collected on the eCRF with the question having the three answers above rather than 'Former User' and 'Current User' and then match the text in the table mocks (Tables First!)
• **pragmatic**
  • *adj.*
  • 1. Dealing or concerned with facts or actual occurrences; practical.
  • 2. *Philosophy* Of or relating to pragmatism.
• The third element is a pragmatic one, where one decides what portion of the process can be automated by tools. This part requires extensive governance by the sponsor and the CRO. It needs:
  – Company-wide standards implementation
  – Sponsorship
  – Standard content alone is not a solution without a tool or tools
  – Jointly living the vision
ODM XML from EDC Has Link to Automation

• Using Alias or SDSVarName in ODM to push CDASH variables to SDTM

• Is it possible to further use Alias or SDSVarName in ODM to push SDTM variables to ADaM?
  – Useful for known SDTM variables that will not be modified
Use ODM Alias
– CDASH to SDTM

Multiple CDASH Variables merge to make one SDTM variable VSDTC

<ItemDef OID="..." Name="Date of vital signs measurement" DataType="date">
  <Question>...</Question>
  <Alias Context="SDTM" Value="date part of VSDTC"/>
  <Alias Context="CDASH" Value="VSDAT"/>
</ItemDef>

<ItemDef OID="..." Name="Time of vital signs measurement" DataType="time">
  <Question>...</Question>
  <Alias Context="SDTM" Value="time part of VSDTC"/>
  <Alias Context="CDASH" Value="VSTIM"/>
</ItemDef>
Use ODM Alias – SDTM to ADaM

Single SDTM Variable to make multiple ADaM variables

```xml
<ItemDef OID="..." Name="Date/Time of Measurements"
   DataType="text">
   <Question>...</Question>
   < Alias Context="ADaM" Value="VSDTC"/>
   < Alias Context="SDTM" Value="VSDTC"/>
</ItemDef>
<ItemDef OID="..." Name="Date/Time of Measurements"
   DataType="text">
   <Question>...</Question>
   < Alias Context="ADaM" Value="ADTM"/>
   < Alias Context="SDTM" Value="VSDTC"/>
</ItemDef>

** Warning: Syntax for illustration only**
To automate the creation of ADaM datasets no single approach will get the desired result.

- A retrospective, prospective, and pragmatic process must be used to automate the creation of ADaM datasets.

- Retrospectively, keeping the end in mind from the beginning with an integrated team approach will better influence systems earlier during data collection and SDTM conversion.
Final Thoughts

• Prospectively for the process to be automated we need to produce machine consumable metadata in the form of an MDR and other tools that are able to push metadata form one standard to the other in a linear process.
  – ODM XML does have features that allow for information system interoperability
Final Thoughts

• Organizations must be pragmatic in not taking on too much at one time
  – Clear commitment and sponsorship from the leadership in the company is essential with funding for tool implementation and governance
  – Company-wide standards implementation with longevity and perseverance over potentially many years to get to the point of automation
  – Giving more time to focus on the science and unique analyses
Contact information

Terek Peterson, MBA
Senior Director – Global Standards Strategies, Clinical Informatics, PRA International
direct: +1 215.444.8613 • 630 Dresher Rd • Horsham, Pennsylvania 19044 • USA •
petersonterek@praintl.com

Karin LaPann
Principal Clinical Programmer
PRA International
direct: +1 434.951.3436
lapannkarin@praintl.com
Thank you