From SDTM to displays, through ADaM & Analyses Results Metadata, a flight on board METADATA Airlines

Omar SEFIANI - Stéphane BOUGET, Boehringer Ingelheim

DH13, PhUSE Barcelona 2016, October, 12th
Outline

• Background
• Metadata Driven Programming
• Metadata Repository
• SDTM mapping & derivations
• ADaM implementation
• Analysis Results Metadata (ARM)
• Data exportation / documentation
• Summary
Many changes can occur during the conduct of a project:
- standard updates leading to structural changes
- new scientific approaches
- new regulatory requests

Find a solution to:
- Reduce maintenance effort
- Increase readability
- Ensure consistency between similar implementations

=> One single program per area dealing with multiple trials simultaneously (centric approach)
• Metadata used for:
  - Generic SAS programs development
  - SAS Code generation
  - Multiple small macros implementation (elementary tasks)
  - Modular programming using KEYWORDS

• The concept presented is currently used in a multi-trial respiratory project
  - BI legacy data conversion to SDTM
  - ADaM transformation
  - Reporting program generation
Definition of algorithms

Identification in the code via meaningful KEYWORDS describing the functionality

- Flexibility
- Reduction of maintenance efforts
- Centralization and automatization

```
* All The Rules for Numerical and Character variables only;
if RULECB = 'FIRST' then do;
  if rowid=1 then &__varname.=r&___varname.;
  end;
else if RULECB = 'FIRST-NON-MISSING' then do;
  if missing(&__varname.) then &__varname.=r&___varname.;
  end;
else if RULECB = 'LAST' then do;
  if rc2=0 and has_nxtchk=0 then &__varname.=r&___varname.;
  end;
else if RULECB = 'LAST-NON-MISSING' then do;
  if not missing(r&___varname.) then &__varname.=r&___varname.;
  end;
else if RULECB = 'NONE' then do;
  &__varname.=r&___varname.;
  end;
```

<table>
<thead>
<tr>
<th>Name</th>
<th>ADAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Target ADaM</td>
</tr>
<tr>
<td>Type</td>
<td>text</td>
</tr>
<tr>
<td>Format</td>
<td>$8.</td>
</tr>
<tr>
<td>Case</td>
<td>UPPER</td>
</tr>
<tr>
<td>Context (Where condition)</td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td></td>
</tr>
</tbody>
</table>

```
RULECB03
Rule to apply for COMB03FL
text
$40.
UPPER
```

FIRST
FIRST
FIRST
FIRST
LAST
Metadata repository (1/4)

MS Excel or any MDR system

**SDTM / ADAM**
- One sheet per SDTM domain / ADAM dataset (structure)
- Differentiation of variable types (mapped or derived / predecessor or derived)
- Controlled Terminology embedded SDTM only
- Link derived variables to derivation methods keywords
- One sheet for all methods (link keyword to SAS macro)

**ARM**
TOC (Table of contents Generator):
- Unique display template
- Output (TLF)
- Statistics for main analyses
<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Variable</th>
<th>Label</th>
<th>Type</th>
<th>Length</th>
<th>Format</th>
<th>Class</th>
<th>Conversion Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM</td>
<td>STUDYID</td>
<td>Study Identifier</td>
<td>Text</td>
<td>20</td>
<td></td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td>DOMAIN</td>
<td>Domain Abbreviation</td>
<td>Text</td>
<td>8</td>
<td></td>
<td>Derived Level 0</td>
<td>RULE_DOMAIN</td>
</tr>
<tr>
<td>DM</td>
<td>USUBJID</td>
<td>Unique Subject Identifier</td>
<td>Text</td>
<td>20</td>
<td></td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td>SUBJID</td>
<td>Subject Identifier for the Study</td>
<td>Text</td>
<td>20</td>
<td></td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td>RFSTDTC</td>
<td>Subject Reference Start Date/Time</td>
<td>Datetime</td>
<td>20</td>
<td>ISO 8601</td>
<td>Derived Level 1</td>
<td>RULE_FIRST_DRUG_INTAKE</td>
</tr>
<tr>
<td>DM</td>
<td>RFENDTC</td>
<td>Subject Reference End Date/Time</td>
<td>Datetime</td>
<td>200</td>
<td>ISO 8601</td>
<td>Derived Level 1</td>
<td>RULE_TRIAL_COMPLETION</td>
</tr>
<tr>
<td>DM</td>
<td>RFXSTDTC</td>
<td>Date/Time of First Study Treatment</td>
<td>Datetime</td>
<td>20</td>
<td>ISO 8601</td>
<td>Derived Level 1</td>
<td>RULE_FIRST_DRUG_INTAKE</td>
</tr>
<tr>
<td>DM</td>
<td>RFXENDTC</td>
<td>Date/Time of Last Study Treatment</td>
<td>Datetime</td>
<td>20</td>
<td>ISO 8601</td>
<td>Derived Level 2</td>
<td>RULE_LAST_DRUG_INTAKE</td>
</tr>
</tbody>
</table>
## Metadata repository (3/4): ADaM setup dataset

<table>
<thead>
<tr>
<th>Name</th>
<th>ADSVAR</th>
<th>TSPARMCD</th>
<th>TSGPID</th>
<th>VALTYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Label</strong></td>
<td>Variable of ADTRIALS</td>
<td>Parameter Code of TS</td>
<td>Group ID</td>
<td>Conversion</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>text</td>
<td>text</td>
<td>text</td>
<td>text</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>$8.</td>
<td>$8.</td>
<td>$40.</td>
<td>$20.</td>
</tr>
<tr>
<td><strong>Case</strong></td>
<td>UPPER</td>
<td>UPPER</td>
<td>UPPER</td>
<td>UPPER</td>
</tr>
<tr>
<td><strong>Context (Where condition)</strong></td>
<td>DBLDT</td>
<td>DCUTDTC</td>
<td>DBLOCK</td>
<td>DATE</td>
</tr>
<tr>
<td></td>
<td>SNAPMED</td>
<td>MEDVER</td>
<td></td>
<td>CHAR4</td>
</tr>
<tr>
<td></td>
<td>SNAPWHO</td>
<td>WHOVER</td>
<td></td>
<td>CHAR6</td>
</tr>
<tr>
<td></td>
<td>TPHASELG</td>
<td>TPHASE</td>
<td></td>
<td>CHAR50</td>
</tr>
<tr>
<td></td>
<td>CNTRIES</td>
<td>FCNTRY</td>
<td></td>
<td>CHAR200</td>
</tr>
<tr>
<td></td>
<td>TITLE</td>
<td>TITLE</td>
<td></td>
<td>CHAR200</td>
</tr>
<tr>
<td></td>
<td>INTMODEL</td>
<td>INTMODEL</td>
<td></td>
<td>CHAR200</td>
</tr>
<tr>
<td></td>
<td>TBLIND</td>
<td>TBLIND</td>
<td></td>
<td>CHAR200</td>
</tr>
<tr>
<td></td>
<td>TCNTRL</td>
<td>TCNTRL</td>
<td></td>
<td>CHAR200</td>
</tr>
<tr>
<td></td>
<td>RANDOM</td>
<td>RANDOM</td>
<td></td>
<td>CHAR10</td>
</tr>
<tr>
<td></td>
<td>SEXPOP</td>
<td>SEXPOP</td>
<td></td>
<td>CHAR10</td>
</tr>
<tr>
<td></td>
<td>AGEMIN</td>
<td>AGEMIN</td>
<td></td>
<td>YEARS</td>
</tr>
<tr>
<td></td>
<td>AGEMAX</td>
<td>AGEMAX</td>
<td></td>
<td>YEARS</td>
</tr>
<tr>
<td></td>
<td>PLANENR</td>
<td>PLANSUB</td>
<td></td>
<td>INTEGER</td>
</tr>
<tr>
<td></td>
<td>FVFPATDT</td>
<td>SSTDTC</td>
<td></td>
<td>DATE</td>
</tr>
<tr>
<td></td>
<td>TSTOPDT</td>
<td>CTARDTC</td>
<td></td>
<td>DATE</td>
</tr>
<tr>
<td></td>
<td>EXTFL</td>
<td>EXTRL</td>
<td></td>
<td>CHAR1</td>
</tr>
<tr>
<td></td>
<td>PROJID</td>
<td>PROJNUM</td>
<td></td>
<td>CHAR10</td>
</tr>
<tr>
<td></td>
<td>WOTIME</td>
<td>REPDUR</td>
<td></td>
<td>DAYS</td>
</tr>
</tbody>
</table>

DH13 - PhUSE - Barcelona 2016
## Metadata repository (4/4): ARM

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Number</th>
<th>Section</th>
<th>Type</th>
<th>Title 1</th>
<th>Program</th>
<th>Outputno</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1</td>
<td>TRIAL PATIENTS</td>
<td>Table</td>
<td>Disposition of patients</td>
<td>disp.sas</td>
<td>1</td>
<td>EOT-DISP1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>EFFICACY EVALUATION</td>
<td>Table</td>
<td>Frequency [N (%)] of patients with adverse events - TS</td>
<td>ae.sas</td>
<td>1</td>
<td>XAE-1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>SAFETY EVALUATION</td>
<td>Table</td>
<td>Descriptive statistics and change from baseline over time - TS</td>
<td>vs.sas</td>
<td>1</td>
<td>EOT-DESC1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple exposure</th>
<th>Highest population</th>
<th>Treatment ADSL</th>
<th>ADSL replacement variables</th>
<th>Optionnal BDS endpoints</th>
<th>BDS where restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>flag</td>
<td>flag</td>
<td>Number Class name</td>
<td>used</td>
<td>ADV5_SYSBP[AVAL] / ADV5_SYSBP[CHG]</td>
<td>AVISIT = 52</td>
</tr>
<tr>
<td>01</td>
<td>TRTFL</td>
<td>1000</td>
<td>NONE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SDTM mapping and derivations (1/2)

**SDTM transformation engine:**
- All legacy data converted to SDTM on an ongoing basis
- Includes additional non-required information by SDTM (e.g. flags)

---

**Step 1 : SDTM mapping**
- One-to-one relationship between one legacy variable and one SDTM variable
- Obtain a bijection between one raw value and one SDTM value

**Step 2 : SDTM derivations**
- After all domains are created
- Facilitating the interaction between different domains
- Execution order needed
  - Example: EPOCH variable derived first in SE and then populated in other domains
SDTM mapping and derivations (2/2)

- Raw data preparation
- Domains
- Mapping
- Derivations

Legacy raw data → SAS → SDTM Domains raw data → Derivation → SDTM Domains final data

Structure check
Bijectivity check

Content check

EXCEL
METADATA SDTM transformation
Definition:
- Raw data preparation
- Domains
- Mapping
- Derivations

SAS

Exportation
XPT files

DH13 - PhUSE - Barcelona 2016
Metadata ADaMs

- Global datasets that contains attributes (study or patient level)
- Can be used by all subsequent ADaMs
- Automatically replicated to the main sheet as ‘Assigned’ variables to be extracted from the define.xml during the creation of the ADS.

ADaM dataset ADTARM in ADS plan

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Context (Where condition)</th>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Origin</th>
<th>Source / Derivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADTARM</td>
<td>&quot;ALL&quot;</td>
<td>STUDYID</td>
<td>Study identifier</td>
<td>Predecessor</td>
<td>TS.STUDYID</td>
</tr>
<tr>
<td>ADTARM</td>
<td>&quot;ALL&quot;</td>
<td>SNAPSHOT</td>
<td>Snapshot Name</td>
<td>Predecessor</td>
<td>TS.TSVAL [TSPARMCD = 'SNAPSHOT']</td>
</tr>
<tr>
<td>ADTARM</td>
<td>&quot;ALL&quot;</td>
<td>SNAPDT</td>
<td>Snapshot Date</td>
<td>Predecessor</td>
<td>TS.TSVAL [TSPARMCD = 'SNAPDT']</td>
</tr>
<tr>
<td>ADTARM</td>
<td>&quot;ALL&quot;</td>
<td>ARMCD</td>
<td>Planned Arm Code</td>
<td>Predecessor</td>
<td>TA.ARMCD</td>
</tr>
<tr>
<td>ADTARM</td>
<td>&quot;ALL&quot;</td>
<td>ARM</td>
<td>Description of Planned Arm</td>
<td>Predecessor</td>
<td>TA.ARMCD</td>
</tr>
<tr>
<td>ADTARM</td>
<td>&quot;ALL&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>B300T</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;E&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>B300T</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;F&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>P300T</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;DBLUE&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>DELUE</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;DRED&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>DRRED</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;SCRNFAIL&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>SCRNFAIL</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;SCRNONG&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>SCRNONG</td>
</tr>
<tr>
<td>ADTARM</td>
<td>ARMCD EQ &quot;NOTASSGN&quot;</td>
<td>GARMCD</td>
<td>Global Arm Code</td>
<td>Assigned</td>
<td>NOTASSGN</td>
</tr>
</tbody>
</table>
ADaM implementation (2/3)

Setup metadata datasets

- Contains algorithms identification
- Read via a generic macro
- Used to generate a SAS program creating the corresponding metadata dataset
- Used as setup dataset during the creation of an ADaM

```
data SETUP_SPTRIALS:
    attrib ADSVAR TSPARMCD TSGRPID VALTYPE
    label="Variable of ADTRIALS" format=$8.
    label="Parameter Code of TS" format=$8.
    label="Group ID" format=$40.
    label="Conversion" format=$20.
;
    call missing(ADSVAR,TSPARMCD,TSGRPID,VALTYPE);
delete;
run;
```

```
proc sql;
    insert into SETUP_SPTRIALS
    values (;
        'DDBDT',
        'DCUTDTC',
        'DBLOCK',
        'DATE'
    )
)
ADaM implementation (3/3)
ARMs used to:
- Automatically create the ARM section in the define.xml V2 for ADaMs
- Dynamically generate a part of SAS macro calls for displays

An output program generator is currently under development to automatically generate the different macro calls for:
- Data building (ADaM preparation)
- Analysis & reporting (calculation+output)
Table Of Content (metadata repository)

**TOC**
- Output description
- Statistical part of ARM

**GENERATOR**
- Data selection
- Macros parametrization

**SAS MACROS**
## Data exportation / documentation

<table>
<thead>
<tr>
<th>SDTM</th>
<th>ADAM</th>
<th>ARM</th>
</tr>
</thead>
</table>
| • Define.xml generated automatically at each execution  
• .XPTs generated at the same time  
• Pinnacle 21 checks performed  
• Bijectivity check  
• Possible data restriction for the export | • Define.xml main source of metadata for the ADaMs creation  
• .XPTs generated at the same time  
• Pinnacle 21 checks performed  
• Analysis Data Reviewer’s Guide contains partially derived information based on the ADS plan | • Present in the define.xml  
• Synchronized with the displays  
• Standardized results datasets exported for validation |
• Powerful approach to minimize risks of inconsistencies across different packages and studies - fits with a project centric approach (multiple trials)

• Development of programs in a generic manner:
  - high level of control needed (user ERRORs and WARNINGs)
  - custom code or hardcoding should be avoided
  - good level of algorithmic and use of complex technical solutions (eg. hashcode, extended attributes, arrays, doSubl, macros, …)

• Synergy between standardization and flexibility
  - more flexible with data diversity
  - reduction of implementation time and maintenance facility
  - improvement in productivity for the creation of similar outputs
  - facilitates delegation and simplify oversight
Thank you

Omar SEFIANI
Stéphane BOUGET
omar.sefiani@boehringer-ingelheim.com
stephane.bouget@boehringer-ingelheim.com