PhUSE EU CONNECT 2018

SI10

Implementation of a Metadata-based Approach to Statistical Planning, Analysis & Reporting

By Frank Freischläger & Hanspeter Schnitzer
INDEX

Situation & Purpose
Clinical Study Metadata
Statistical Programming for Analysis
Statistical Reporting
Statistical Planning
Conclusion
SITUATION without METADATA

Traditional approach of statistical planning, analysis & reporting of clinical studies

Diagram:
- Protocol
- SAP
- TOC & Mock TFLs
- SP Specs
- Templates & Standards
- Clinical Study Report
- Programming
- TFLs
- Derived Data
- DD Specs
- Raw Data
- eCRF
PURPOSE

- Metadata approach to biostatistics
- Consider all aspects of processing
- Set up a system built on and compliant with standards
- Assure flexibility in support of clients and sponsors
- Let the system do the routine work
- Limit needs for programming to a minimum
- Allow statisticians and programmers to concentrate on advanced methods
- Enable respective shift in manpower
CLINICAL STUDY METADATA
**METADATA REPOSITORY**

Overview of data panels with content explicitly required per project

<table>
<thead>
<tr>
<th>General</th>
<th>Protocol</th>
<th>Deliverables</th>
<th>Statistics</th>
<th>Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Clients</td>
<td>• Documents</td>
<td>• Plans</td>
<td>• Populations</td>
<td>• LibDirSpecs</td>
</tr>
<tr>
<td>• Stakeholders</td>
<td>• IE Criteria</td>
<td>• Derived datasets</td>
<td>• Endpoints</td>
<td>• Output structures</td>
</tr>
<tr>
<td>• Projects</td>
<td>• Objectives</td>
<td>• Outputs</td>
<td>• Intercurrent events</td>
<td>• Report structures</td>
</tr>
<tr>
<td>• Studies</td>
<td>• Analyses</td>
<td>• Reports</td>
<td>• Estimands</td>
<td>• Plan structures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Evaluations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• References</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Epochs</td>
<td>• Sources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Visits</td>
<td>• Groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Disposition</td>
<td>• Variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Exposure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Page 6 // 7 Nov 2018
FLEXIBILITY in APPLYING STANDARDS

Approach with different layers

- Project-specific content is mandatory
- Other content may be defined on higher levels

Levels are dynamic
- Content is guaranteed
- Metadata entry or import
- Metadata edit checks are in place
- Independent double entry as an option
## METADATA REPOSITORY

Overview of data panels with content that may be given on any level

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Semantics</td>
<td>• InDataset Specs</td>
</tr>
<tr>
<td>• Tools</td>
<td>• Settings</td>
</tr>
<tr>
<td>• Templates</td>
<td>• Styles</td>
</tr>
<tr>
<td>• Method options</td>
<td>• Output options</td>
</tr>
<tr>
<td>• Derivations</td>
<td>• Report options</td>
</tr>
<tr>
<td>• Descriptors</td>
<td>• Plan options</td>
</tr>
<tr>
<td>• Time definitions</td>
<td></td>
</tr>
<tr>
<td>• Date imputations</td>
<td></td>
</tr>
</tbody>
</table>

- Resulting in project relevant metadata
- Processed and archived per production and delivery
WORKFLOW

From derived datasets to booklets of outputs

- Settings
- OutputOptions
- OutputControls
- OutputStructures

Style

Macro Library

Derived Data

Results Datasets

Output Datasets

Control Center

Macro-free source code

TFLs in Booklets

- For development programming
- With original SAS outputs
- With log checking summaries
## LAYOUTS and FORMATTING

### Example of PhUSE CSS Analyses and Code Sharing WG on demographics

#### Table 7.1.

<table>
<thead>
<tr>
<th>Demographic Parameter</th>
<th>PL (N=1000)</th>
<th>T1 (N=1500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n*</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>Female</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>Male</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>Missing</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>n*</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>Median</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>Q1, Q3</td>
<td>xx, xx</td>
<td>xx, xx</td>
</tr>
<tr>
<td>Min, Max</td>
<td>xx, xx</td>
<td>xx, xx</td>
</tr>
<tr>
<td>Missing</td>
<td>xx</td>
<td>xx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age Categories n (%)</th>
<th>PL (N=1000)</th>
<th>T1 (N=1500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;65</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>≥65 and &lt;75</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>≥75 and &lt;65</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>≥85</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
<tr>
<td>Missing</td>
<td>xx</td>
<td>xx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex n (%)</th>
<th>PL (N=1000)</th>
<th>T1 (N=1500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>xx (k.k)</td>
<td>xx (k.k)</td>
</tr>
</tbody>
</table>

---

**Label, n (%)**

**Label - n (%)**

**Label n (%)**

**Label, Unit**

**Label (Unit)**

**Label [Unit]**

**Missing category required?**

---

If 100%?

N (100.0)
N (100)

If 0?

0 (0.0)
0

-
DATA-TO-TEXT TECHNOLOGY

Utilizing methods of Natural Language Generation

- Developed by Ehud Reiter and others
- From structured data to text messages and documents
- Automatic creation of sentences and documents
- Content selection and surface realization in computational linguistics

- Examples include weather forecasts, automated journalism
- Published successful applications to healthcare data
Auto-generated report elements and post-processing

- Micro-writing and macro-writing
- Flexible post-processing
- Updateable with new clinical data
A total of 500 patients were screened, of whom 50 (10.0%) failed screening prior to randomization (see Table 1).

Most patients were female (62.2%) and white (54.4%).

Most [SUBJ_TXT] were [DEMO_GENDER_MOSTFRQ_LBL] ([DEMO_GENDER_MOSTFRQ_PCT] %) and [DEMO_RACE_MOSTFRQ_LBL] ([DEMO_RACE_MOSTFRQ_PCT] %).
STATISTICAL PLANNING
WORKFLOW

Auto-generated elements of a planning document, and post-processing

- Similar to report generation
- Increased complexity
- Sentences reusable for reporting
SENTENCES for a PLAN

Some examples: definition of TEAE & description of an estimand

An adverse event is treatment-emergent, if onset or deterioration of the event appear after the first study treatment intake and no later than \( TE_{LAGDEF} \) \( TE_{LAGUNIT} \) after the last study treatment intake.

**Population-level summary**

Measure: Difference in means between treatment conditions in the change from baseline to Week 24 in sBP in the targeted population regardless of intercurrent events.

Parameter: Specification of consideration of intercurrent events.
METADATA-BASED APPROACH

New approach of statistical planning, analysis & reporting of clinical studies

- Metadata Repository
  - Protocol
  - eCRF
  - Raw Data
  - DD Specs
  - Auto-generated Programming
  - Derived Data

- Control Center
  - Templates & Standards
  - SAP, Table and Mock Versions of TFLs, Statistical Programming Specifications

- Clinical Study Report
  - TFLs
  - Raw Data
  - DD Specs
  - Auto-generated Programming
  - Control Center
  - Metadata Repository
SUMMARY

- Extensive MDR
- Flexible & neural
- Productive & reliable

Limitations

- Requires careful metadata management
- Challenging in preparing coded links for sentences

OUTLOOK

- Applicable to quality documents like SOPs etc.
- Useful for service proposals

- Improve MDR frontend, backend and control center
- Consider OpenXML or Linked Data as alternatives
THANK YOU

Frank Freischläger
Estimondo GmbH
frank.freischlaeger@estimondo.com
www.estimondo.com

Hanspeter Schnitzer
Estimondo GmbH
hanspeter.schnitzer@estimondo.com
www.estimondo.com