

A Simple Interface for Metadata

Magnus Mengelbier
Director

Topics

- Introduction
- Metadata
- Data Sets and Tables
- Interfaces
- Summary

Metadata

- Everyone uses metadata

- Facebook
- LinkedIn
- Skype
- E-mail
- Shopping



Built on metadata

- Analysis and statistical programming uses metadata

- Protocol
- SAP
- Data Set Specifications
- Table Shells



- Microsoft Word
- Microsoft Excel
- PDF

Define Metadata

- Data about data

I am 42 years old.

- Information about data
- Information about analysis

Metadata: Data Sets

- Name
- Label
- Sort order
- Keys (identifies a single observation)
- Ordered list of variables

PROC CONTENTS

- | | |
|----------|-------------------|
| ▪ Name | ▪ Code list |
| ▪ Label | ▪ Source |
| ▪ Type | ▪ Derivation rule |
| ▪ Length | |
| ▪ Format | |

Metadata: Tables

- | | | |
|---|---|--|
| <ul style="list-style-type: none"> ▪ Title(s) ▪ Population(s) ▪ Treatment arms ▪ Summary information ▪ Footnote(s) ▪ Source ▪ Status ▪ Version(s) |  | <ul style="list-style-type: none"> ▪ Variables ▪ Time points ▪ Statistics |
|---|---|--|

Sources for Building a Table

XXX-nnn Example Compound Table 15. Demographics and Baseline Characteristics All Subjects Safety Efficacy					
Demographics and Baseline Characteristics	Treatment A N = nn	Treatment B N = nn	Treatment C N = nn	Treatment D N = nn	Total N = n
Gender					
Male	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)
Female	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)
Age (years)					
N	xx	xx	xx	xx	xx
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx.x	xx.x	xx.x	xx.x	xx.x
Min - Max	xx - xx	xx - xx	xx - xx	xx - xx	xx - xx
Race					
White	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)
Black	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)
Hispanic	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)	xx (xxx%)
Source: Listing ...					
X:\...\my_program.sas DRAFT mengelbier 15OCT2012 16:42					

Building a Table: Summary Section

Age (years)	
N	xx
Mean (SD)	xx.x (xx.xx)
Median	xx.x
Min - Max	xx - xx

- Section Label
- Statistics
 - N
 - Mean (SD)
 - Median
 - Min - Max
- Composite “Statistics” Pattern
 - Mean (SD) : Mean , Standard Deviation
 - Min – Max : Minimum , Maximum
- Source
 - Data Set
 - Variable(s)
- Formats
- Treatment Groups

Building a Table: Summary Section

Race	
White	xx (xxx%)
Black	xx (xxx%)
Hispanic	xx (xxx%)
Asian	xx (xxx%)
Other	xx (xxx%)

- Section Label
- Category values
 - All values
 - Selected / Excluded
 - Consolidated
- Composite “Statistics” Pattern
- Source
 - Data Set
 - Variable(s)
- Formats
- Treatment Groups

Readout

- Data sets and outputs (TLGs)
 - Consistency within deliverables
 - No consistency across deliverables

- Interface features
 - “Flexible” environment
 - Informal conventions
 - Manual administration
 - Difference between editing and programming

Metadata: A Simple Interface

H42			
	E	F	G
2	t_dm	study-compound	XXX-nnn Example Compound
3	t_dm	table-number	15
4	t_dm	table-title	Demographics and baseline characteristics
5	t_dm	population	All Subjects
6	t_dm	population	Safety
7	t_dm	population	Efficacy
8	t_dm	treatment-group	trt_a
9	t_dm	treatment-group-label	Treatment A
10	t_dm	treatment-group	trt_b
11	t_dm	treatment-group-label	Treatment B
12	t_dm	treatment-group	trt_c
13	t_dm	treatment-group-label	Treatment C
14	t_dm	treatment-group	trt_d
15	t_dm	treatment-group-label	Treatment D
16	t_dm	treatment-group-total	right
17	t_dm	treatment-group-total-label	Total
18	t_dm	table-body-title	Demographics and baseline characteristics
19	t_dm	section-label	Age (years)
20	t_dm	section-variable	ADDM.AGE
21	t_dm	section-variable-unit	ADDM.AGEU
22	t_dm	section-statistics	N
23	t_dm	section-statistics-label	N
24	t_dm	section-statistics-format	8
25	t_dm	section-statistics-justify	decimal
26	t_dm	section-statistics-select	_ALL_
27	t_dm	section-statistics-exclude	_NONE_
28	t_dm	section-statistics	MEAN_SD
29	t_dm	section-statistics-label	Mean (SD)
30	t_dm	section-statistics-format	null
31	t_dm	section-statistics-justify	space
32	t_dm	section-statistics-select	_ALL_
33	t_dm	section-statistics-exclude	_NONE_
34	t_dm	section-statistics	MEAN
35	t_dm	section-statistics-label	_COMPOSITE_
36	t_dm	section-statistics-format	8.1
37	t_dm	section-statistics-justify	null
38	t_dm	section-statistics-select	null
39	t_dm	section-statistics-exclude	null
40	t_dm	section-statistics	SD
41	t_dm	section-statistics-label	_COMPOSITE_
42	t_dm	section-statistics-format	8.2
43	t_dm	section-statistics-justify	null

Define a Data Set

Meta-data set					
http://atlas.se.limellogic.local/browse/example14/spec/dm					
Example 14 Specification					
Status	Draft	Created	21 Mar 2012 08:34	Last Comment	Reordered variables
Assigned	Magnus Mengelbier	Last Modified	02 Oct 2012 16:03		
Domain	DM -- Demography				
Sort	STUDYID USUBJID				
Keys	STUDYID USUBJID				
Variable	Label	Type	Length	Format	
STUDYID	Study Identifier	Char	25		
DOMAIN	Domain Abbreviation	Char	10	"DM"	
USUBJID	Unique Subject Identifier	Char	35		
SUBJID	Subject Identifier for the Study	Char	35		
RFSTDTC	Subject Reference Start Date/Time	Char	35	ISO 8601	
RFENDTC	Subject Reference End Date/Time	Char	35	ISO 8601	
SITEID	Study Site Identifier	Char	20		
INVID	Investigator Identifier	Char	20		
INVNAM	Investigator Name	Char	200		
BRTHTDC	Date/Time of Birth	Char	35	ISO 8601	
AGE	Age in #AGEU at RFSTDTC	Num			
AGEU	Age Units	Char	15		

- Name
- Label
- Sort order
- Keys
- Ordered list of va

- Name
- Label
- Sort order
- Keys
- Ordered list of variables

Drawing a Table

Meta-data table			
← → ↺ http://atlas.se.limellogic.local/browse/example14/spec/t_dm		oracle11g - How to i... MindSalt Login references beskattning vid före... SAS BEGINNINGS - ... sport Iliya Polihronov: Hig...	
Example 14	Specification		
Status	Draft	Created	15 May 2012 22:12
Assigned	Magnus Mengelbier	Last Modified	15 May 2012 22:12
Table t_dm -- <i>Demographics and Baseline Characteristics</i> Populations <i>All Subjects Safety Efficacy</i> Treatment Groups <i>All Groups Total</i>			
XXX-nnn Example Compound Table 15. Demographics and Baseline Characteristics All Subjects Safety Efficacy			
Demographics and Baseline Characteristics		Treatment A N = nn	Treatment B N = nn
Gender			
Male		xx (xxx%)	xx (xxx%)
Female		xx (xxx%)	xx (xxx%)
Age (years)			
N		xx	xx
Mean (SD)		xx.x (xx.x)	xx.x (xx.x)
Median		xx.x	xx.x
Min - Max		xx - xx	xx - xx
Race			
White		xx (xxx%)	xx (xxx%)
Black		xx (xxx%)	xx (xxx%)
Hispanic		xx (xxx%)	xx (xxx%)
Asian		xx (xxx%)	xx (xxx%)
Other		xx (xxx%)	xx (xxx%)

- Title(s)
- Popul
- Treatr
- Summ
- Footn
- Sourc
- Status
- Versio

- Title(s)
- Population(s)
- Treatment arms
- Summary information
- Footnote(s)
- Source
- Status
- Version(s)

Programming with Metadata

```
%rules_metadata( meta = dm, output = work.meta_dm );  
  
%rules_metadata( meta = t_dm, output = work.meta_t_dm );
```

```
%rules_make_dataset( meta = dm, output = sdtms.dm );  
  
%rules_make_dataset( meta = addm, output = adams.addm );
```

```
filename tables "my tables folder";  
  
%rules_make_table( meta = t_dm, output = tables );
```

Automation

... that proc that creates your m5 deliverables, including the define.xml, intext tables, hyperlinks, all our TLGs – validated - , your SDTMs, ADaMs, etc. It even does a nice QA.

- Greg Fagan

```
filename t_prg "my programs folder for tables";

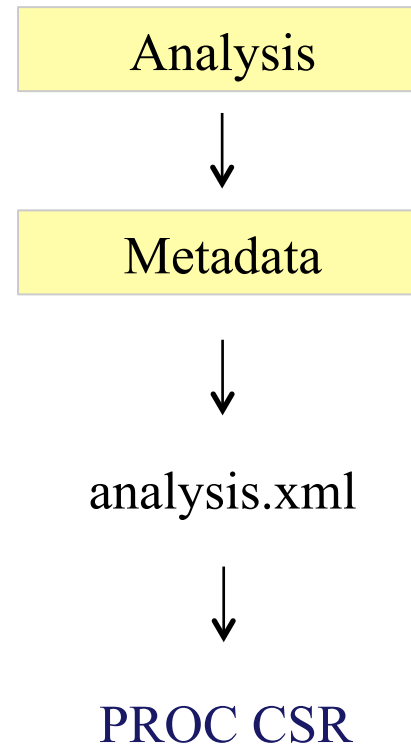
%rules_make_programs( type = tables, output = t_prg );
```

```
filename a_prg "my analysis programs folder";

%rules_make_analysis( type = safety, output = a_prg );
```

Summary

- Metadata is complex
- Simple interfaces
- Single “simple” interface does not exist
- Complexity is internal
- Automation



Questions & Comments

- Introduction
- Metadata
- Data Sets and Tables
- Interfaces
- Summary

Magnus Mengelbier

Director

Limellogic

mmr@limellogic.com

+46 708 110 876

+44 208 144 5701

THE END