



Facilitating Data Integration For Regulatory Submissions

John R. Gerlach; SAS / CDISC Specialist

John C. Bowen; Independent Consultant



The Challenge

- Creating an Integrated (Harmonized) Collection of Clinical Data for Regulatory Submission
 - Labor Intensive
 - Error Prone
 - Modus Operandi – *Ad Hoc Programming*



The SAS Solution

- Reporting Tool to Evaluate Pair-wise Data Sets
 - Meta Data Level
 - Content Level
- Assumptions
 - Same Data Set Names
 - Same Variable Names
- Expandable

Meta Data Report

Comparison of the DM Data Set in the *Left* and *Right* Data Libraries (Metadata Level)

===== Left =====				===== Right =====		
Name	Type	Length	Label	Type	Length	Label
AGE	NUM	8	Age in AGEU at ...	NUM	8	Age in AGEU a t...
AGEU	CHAR	5	Age Units	CHAR	5	Age Units
ARM	CHAR	10	Description of ...	CHAR	10	Description of ...
ARMCD	CHAR	10	Planned Arm Code	CHAR	10	Planned Arm Code
BRTHDTC	CHAR	10	Date of Birth	CHAR	10	Date of Birth
* COUNTRY	CHAR	3	Country			
* DOMAIN	CHAR	2	Domain Abbreviation	CHAR	8	Domain Abbreviation
RACE	CHAR	10	Race	CHAR	10	Race
* RFENDTC	CHAR	20	Subject Reference End ...			
RFSTDTC	CHAR	20	Subject Reference Start	CHAR	20	Subject Reference Start ...
* SEX	CHAR	6	Sex	NUM	8	Sex
SITEID	CHAR	8	Study Site Identifier	CHAR	8	Study Site Identifier
STUDYID	CHAR	20	Study Identifier	CHAR	20	Study Identifier
* SUBJID	CHAR	10	Subject Identifier ...			
USUBJID	CHAR	15	Unique Subject ...	CHAR	15	Unique Subject Identifier

Content Level Report

Comparison of the AE Data Set in the *Left* and *Right* Data Libraries
(Content Level)

<i>Variable</i>	<i>Left</i>	<i>Right</i>
AESER	N Y	N Y
AEREL	< Null > N Y	DEFINITELY RELATED NOT RELATED POSSIBLY RELATED PROBABLY RELATED UNLIKELY RELATED



SAS Reporting Tool

- Base SAS
 - Macro Language
 - Data Step Programming
 - REPORT Procedure
- SQL with Dictionary Tables
 - TABLES
 - COLUMNS
- `%data_integrate(study101, study201, AE, HTML=N) ;`



Meta-Data Level Report Methodology

- Determine Both Data Sets Exist.
- Obtain Meta Data on Each Data Set.
- Perform Match-merge.
- Produce Report.

Meta Data Report

Comparison of the AE Data Set in the *Left* and *Right* Data Libraries (Metadata Level)

===== Left =====				===== Right =====			
Name	Type	Length	Label	Type	Length	Label	
AEACN	CHAR	100	Action Taken w..	CHAR	100	Action Taken with ...	
AEBODSYS	CHAR	100	Body System ..	CHAR	100	Body System or Organ Class	
AEDECOD	CHAR	100	Dictionary-Derived Term	CHAR	100	Dictionary-Derived Term	
AEENDTC	CHAR	20	End Date/Time of Adver..	CHAR	20	End Date/Time of Adverse ...	
AEENDY	NUM	8	Study Date of End of Event	NUM	8	Study Day of End of Event	
* AEENRF	CHAR	16	End Relative to Reference ...				
AEHLGT	CHAR	200	MedDRA Highest Level ...	CHAR	200	MedDRA Highest Level ...	
* AEOUT	CHAR	50	AE Outcome	CHAR	25	Outcome of Adverse Event	
* AEREL	CHAR	1	Causality	CHAR	20	Causality	
* AESDTH	CHAR	1	Results in Death				
AESSEQ	NUM	8	Sequence Number	NUM	8	Sequence Number	
AESER	CHAR	1	Serious Event	CHAR	1	Serious Event	
AESEV	CHAR	20	Severity	CHAR	20	Severity	
AESTDTC	CHAR	20	Start Date/Time of ...	CHAR	20	Start Date/Time of ...	
AESTDY	NUM	8	Study Day of Start of Event	NUM	8	Study Date of Start of Event	
* AETERM	CHAR	200	Reported Term for the ...	CHAR	100	Reported Term for the ...	
DOMAIN	CHAR	2	Domain Abbreviation	CHAR	2	Domain Abbreviation	
STUDYID	CHAR	20	Study Identifier	CHAR	20	Study Identifier	
USUBJID	CHAR	15	Unique Subject Identifier	CHAR	15	Unique Subject Identifier	

Meta Data Report

Comparison of the AE Data Set in the *Left* and *Right* Data Libraries (Metadata Level)

===== Left =====				===== Right =====		
Name	Type	Length	Label	Type	Length	Label
* AEENRF	CHAR	16	End Relative to Reference ...			
* AEOUT	CHAR	50	AE Outcome	CHAR	25	Outcome of Adverse Event
* AEREL	CHAR	1	Causality	CHAR	20	Causality
* AESDTH	CHAR	1	Results in Death			
* AETERM	CHAR	200	Reported Term for the ...	CHAR	100	Reported Term for the ...

Meta Data Report

- Assume Meta-data Report Indicates Perfect Match.
- Data Level - A Different Matter
 - Different Versions of MedDRA / WHO Codes
 - Variable Sex Having Value 'M' versus '1'

You Need BOTH Reports!



Content Level Report Methodology

- Identify Character variables, if any.
- For each Character variable -
 - Obtain unique values in the *Left* data set.
 - Determine data type of the respective variable in the *Right* data set. Why?



Content Level Report Methodology

- Obtain unique values in *Right* data set.
- Store as character values, regardless of data type.
- Combine *Left* and *Right* data sets keeping 30 observations.
- Assign the text '< Null >' for missing value.



Content Level Report Methodology

- Append data set representing the i th variable to the reporting data set.
- Produce the report.
- Do it again for Numeric Variables.

Data Integration Issue – AEOUT

Left Study

FATAL
RESOLVED
RESOLVED WITH SEQUELAE
UNKNOWN
UNRESOLVED

Right Study

FATAL
ONGOING
RESOLVED
RESOLVED WITH SEQUELAE

- Right side represents a subset of values.
- Active Study - “ONGOING” should change status by database lock.

Data Integration Issue – AEREL

Left Study

N
Y

Right Study

Definitely Related
Not Related
Possibly Related
Probably Related
Unlikely Related

- Dichotomous versus descriptive values.
- Unlikely Related & Not Related → N
- Other Values → Y



Data Integration Issue – AESDTH

- Manifested in Metadata report only.
- AESDTH variable exists in all studies, except one.
- However, AEOUT exists in the Domain.
- $\text{AESDTH} \leftarrow \text{Imputed from AEOUT (FATAL)}.$

Data Integration Issue – AESEV

Left Study

LIFE THREATENING
MILD
MODERATE
SEVERE

Right Study

<Null>
Mild
Moderate
Severe
Unknown

- Null and Unknown values may be an issue.
- Mixed case needs to be converted.

Data Integration Issue – ARMCD

Left Study

PROD_NAME
PLACEBO

Right Study

<Null>
DRUG_NAME
PLACEBO

- Embarrassing Null value for a Required variable.
- DRUG_NAME needs to be re-assigned to PROD_NAME.

Data Integration Issue – CMROUTE

Left Study

INTRAVENOUS

Right Study

I/V

IV

Intravenous

Intravenous Direct

Intravenous Injection

- Convert various forms of Intravenous.

Data Integration Issue – COUNTRY

Left Study

USA
ENG
ITA

Right Study

US

- ISO 3166 3-byte versus 2-byte.

Data Integration Issue – RFENDTC

Left Study

<Null>

2007-01-17

2007-01-23

2007-01-30

2007-01-31

Right Study

<Null>

2008-07-16T:00:00

2008-07-18T:00:00

2008-07-21T:00:00

2008-07-31T:00:00

- Null value acceptable for Screen failures only.
- Date / Time converted to ISO8601 Date only.

Data Integration Issue – SEX

Left Study

Right Study

M

<Null>

F

1

U

2

- Left study uses proposed CDISC Control Terminology.



Conclusion

- Data integration -- Part of the IT landscape.
 - ISS / ISE Submissions
 - Acquisitions (Differing Proprietary Standards)
- CDISC Standards -- No Guarantee for Harmonization Across Studies.
- Reporting Tool
 - Metadata Level
 - Content Level
- Standard Reports Promoting Good Communication.

Questions?

- John R. Gerlach
SAS / CDISC Specialist
jrgerlach@optonline.net
- John C. Bowen
Independent Consultant
jhnbn7@gmail.com