PATTERN BASED METADATA REPOSITORY: TOWARD HIGH QUALITY DATA STANDARDS

11th October 2016, Barcelona
Alan Cantrell

- Julius Kusserow
- Julie James
- Deb Copeland
- Natraj Patro
- Isabelle de Zegher
<table>
<thead>
<tr>
<th>Introduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking the Line</td>
</tr>
<tr>
<td>Introducing PAREXEL Concepts</td>
</tr>
<tr>
<td>Introducing Patterns</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
</tbody>
</table>
INTRODUCTION
THE PROBLEM WE WANT TO SOLVE

• Many companies have deployed – or are in the process of deploying – a metadata repository (MDR) to manage their data standards.

• In most organisations, data standards are maintained in silos:
  • Data collection /CDASH standards are maintained within Data Management,
  • SDTM is governed by the clinical programmers
  • ADaM is managed by the statistical programmers.

• These different groups collaborate to maintain mapping between the different standards,
  • This remains a challenging process across separate groups with disjointed governance processes.
  • Mapping between the standards remains an “art”, based on manual interpretation and experience from the programmers.
INTRODUCTION
PAREXEL® END-TO-END DATA STANDARDS....

FULL TRACEABILITY

Metadata Repository

- eCRF Library
- SCE / Global Macro Library
INTRODUCTION
.. WITH CONCEPTS AT THE CORE

META DATA REPOSITORY (MDR)

MANAGEMENT OF DATA STANDARDS
(CDISC standards enriched with semantics)

CDASH SDTM ADaM

eCRF/Forms Definition

Form Usage Library
EDC IVRS ePRO Lab

Macro Library
Statistical Computing Environment
INTRODUCTION AND USAGE IN CONTEXT OF STUDY

MANAGEMENT OF DATA STANDARDS
(CDISC standards enriched with semantics)

MANAGEMENT OF STUDY INSTANCE METADATA
(Metadata driven collection form build)

META DATA REPOSITORY (MDR)

CDASH SDTM ADaM
eCRF/Forms Definition

Form Usage Library
EDC IVRS ePRO Lab

Macro Library
Statistical Computing Environment

Protocol Optimization
Data Surveillance
Clinical Data Warehouse

© 2016 PAREXEL INTERNATIONAL CORP. / CONFIDENTIAL
TABLE OF CONTENT

• Introduction

• Walking the Line

• Introducing PAREXEL Concepts

• Introducing Patterns

• Conclusion
We work in a linear way...

**BUILD**
We need to build a database before we can enter data

**COLLECT**
Data needs to be entered before we can produce tabulation data (SDTM)

**ANALYZE**
Tabulation data needs to be available before the analysis datasets are created

What would happen if we didn’t need to wait?
How can we change our process to avoid waiting?
WALKING THE LINE - WHAT IS NEEDED

We need an end-to-end (E2E) approach to standardize information.

When a data collection form is designed we can know how it will impact the tabulation and analysis data.

If we know the structure of the data collection forms, and how that is connected to the structure of tabulation and analysis data, we can reduce the waiting time.

Clinical Research Concepts and Patterns will help us to realize this.
TABLE OF CONTENT

• Introduction

• Walking the Line

• Introducing PAREXEL Concepts

• Introducing Patterns

• Conclusion
INTRODUCING PAREXEL CONCEPTS

DEFINITION

• To explore this approach at PAREXEL, we looked at
  
  • a specific module of data: adverse events (AE)
  
  • and a specific example of Severity

• A module is composed of 3 levels
  
  • Data Collection (from for EDC and other data collection instruments )
  
  • Tabulation Data (some from data collection, some derived, some assigned)
  
  • Analysis Data
INTRODUCING PAREXEL CONCEPTS
EXAMPLE: SEVERITY CONCEPT

• Certain aspects overlap:
  • Data Collection: AESEV (SDTM)
  • Tabulation Data: AESEV (SDTM)
  • Analysis Data: AESEV, AESEVN, ASEV and ASEVN (ADaM)

• If an element exists in 2 more places, it is (or should be) the same “thing”.

After identifying the different unique “things” present in the AE module, we created a grouping of them in a “concept”.
INTRODUCTION TO PAREXEL CONCEPTS
CONCEPT & DOMAIN DEFINITION VARIABLE

• Benefits
  • Semantic consistency in definition
  • Data Lineage
  • Consistency & quality in study execution:
    – if decision is NOT to collect severity in the CRF, we would not want the analysis severity variables in ADaM dataset.
    – Without a concept approach, we would not identify this before ADaM programmer realizes that he does not have the needed data

• Concerns
  • Need to counsel the variables
  • Render explicit these implicit relationships across variables.

Instead of having all variables in the concept, we could have a single variable which represents each grouping. We reduced the list of variables to a single element, called “Domain Definition Variable”.
INTRODUCING PAREXEL CONCEPTS
CONCEPT & DOMAIN DEFINITION VARIABLE

A “Domain Definition Variable”
• links the variables across all the components
• reduces the number of variables required to represent the collection, tabulation and analysis
CONCEPT & DOMAIN DEFINITION VARIABLE LINKING VARIABLES

Adverse Event Domain Definition

How to populate
• AESEV in AE SDTM domain
• ASEVN in AE ADaM
CONCEPT & DOMAIN DEFINITION VARIABLE RECREATING LINEAR MAPPING

Connections between the domain definition variable and the data collection, tabulation and analysis variables, recreate linear mapping.

Adverse Event Domain Definition

AE Domain Definition
SEV  Severity/Intensity

CDASH / Data Collection
AESEV  Severity

SDTM / Data Tabulation
AESEV  Severity/Intensity

ADaM / Data Analysis
AESEV  Severity/Intensity
AESEVN  Severity/Intensity (N)
ASEV  Analysis Severity/Intensity
ASEVN  Analysis Severity/Intensity (N)
SEVGRy  Pooled Severity Group y
SEVGRyN  Pooled Severity Group y (N)
How do we further build on these links: the patterns
TABLE OF CONTENT

• Introduction

• Walking the Line

• Introducing PAREXEL Concepts

• Introducing Patterns

• Conclusion
INTRODUCING PATTERNS
SCRUTINIZING VARIABLES

• We have seen the creation of a domain definition variable SEV.
• The variable is not just a character string datatype or numeric datatype.
• It has to have the capability of representing multiple datatypes related to the different variables linked to this domain definition variable.

• Data collection and data tabulation variables AESEV is governed by controlled terminology.

• Data analysis variables
  • AESEVN is a numeric representation of the AESEV variable.
  • ASEV and ASEVN is an imputed variable pair representing AESEV (e.g. in case of missing severity values in data collection).
  • SEVGRy and SEVGRyN can represent a pooled grouping of AE severity in the analysis based on ASEV values.
INTRODUCING PATTERNS
CONSISTENT “BEHAVIOR” ACROSS VARIABLES

• Coded values are linked together in a logical way.

<table>
<thead>
<tr>
<th>Collection / Tabulation</th>
<th>Analysis</th>
<th>Ordinal</th>
<th>Analysis Group</th>
<th>Codelist</th>
<th>Codelist Code</th>
<th>NCI Preferred Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILD</td>
<td>Mild</td>
<td>1</td>
<td>Mild/Moderate - 1</td>
<td>C66769</td>
<td>C41338</td>
<td>Mild Adverse Event</td>
</tr>
<tr>
<td>MODERATE</td>
<td>Moderate</td>
<td>2</td>
<td>Mild/Moderate - 1</td>
<td>C66769</td>
<td>C41339</td>
<td>Moderate Adverse Event</td>
</tr>
<tr>
<td>SEVERE</td>
<td>Severe</td>
<td>3</td>
<td>Severe - 2</td>
<td>C66769</td>
<td>C41340</td>
<td>Severe Adverse Event</td>
</tr>
<tr>
<td>&lt;null&gt;</td>
<td>Severe</td>
<td>3</td>
<td>Severe - 2</td>
<td>C66769</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• This logic can be defined
• building on the code list definition and some SAS logic
• for imputing the analysis value and defining the grouping.

• Example: “MODERATE”

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Value</th>
<th>Variables in ADaM</th>
</tr>
</thead>
<tbody>
<tr>
<td>.codelist().displayName()</td>
<td>MODERATE</td>
<td>AESEV</td>
</tr>
<tr>
<td>.codelist().value()</td>
<td>2</td>
<td>AESEVN</td>
</tr>
<tr>
<td>.codelist().imputed().displayName()</td>
<td>Moderate</td>
<td>ASEV</td>
</tr>
<tr>
<td>.codelist().imputed().value()</td>
<td>2</td>
<td>ASEVN</td>
</tr>
<tr>
<td>.codelist().group().displayName()</td>
<td>Mild/Moderate</td>
<td>SEVGR1</td>
</tr>
<tr>
<td>.codelist().group().value()</td>
<td>1</td>
<td>SEVGR1N</td>
</tr>
</tbody>
</table>
INTRODUCING PATTERNS
INTRODUCING COMPLEX DATA TYPE (CDT)

- Data structure storing the logic linking all of the values for collection, tabulation and analysis
- Enable representing multiple data types from the different values.
INTRODUCING PATTERNS
INTRODUCING COMPLEX DATA TYPE (CDT)

Adverse Event
Domain Definition

AE Domain Definition
SEV  Severity/Intensity

CDT connected with domain definition variable ‘SEV’ = a single variable with a complex datatype
CDT connected with domain definition variable ‘SEV’ = a single variable with a complex datatype to represent collection, tabulation & analysis data.
INTRODUCING PATTERNS
REUSING COMPLEX DATA TYPE (CDT): PATTERNS

• This complex datatype can be reused

• It is not unique to “Severity”

• But it can be used in the aspects of collection, tabulation and analysis
  • for “Causality” and “Toxicity Grade” in AE
  • and in many other variables across domains.
TABLE OF CONTENT

• Introduction

• Walking the Line

• Introducing PAREXEL Concepts

• Introducing Patterns

• Benefits and Conclusion
# BENEFITS AND CONCLUSION

## INDEPENDENTLY MAINTAINED STANDARDS

**Manual Mapping**

- **Flexibility**
  - Stepped Approach
  - Mapping Inconsistencies
  - Workload to Generate High Quality E2E Lineage

## CONCEPT LINKED STANDARDS

**Reusable Templates**

- Consistency
- Increased Quality
- Increased Efficiency

- Definitions of Patterns
- Process Change
- MDR Tool Requirement
BENEFITS AND CONCLUSION

Through the use of patterns, the PAREXEL® Clinical MDR can support you through the development and implementation of high quality E2E data standards.

Patterns facilitate the management of standards

DEVELOPED BY EXPERTS

SPECIFICATIONS AND DEVELOPMENT

TARGETING TRACEABILITY

Why not come and learn more about the PAREXEL® Clinical MDR at our booth and see PATTERNS in action!
THANK YOU

Alan Cantrell
Senior Manager, Statistical Programming
PAREXEL International
Alan.Cantrell@PAREXEL.com