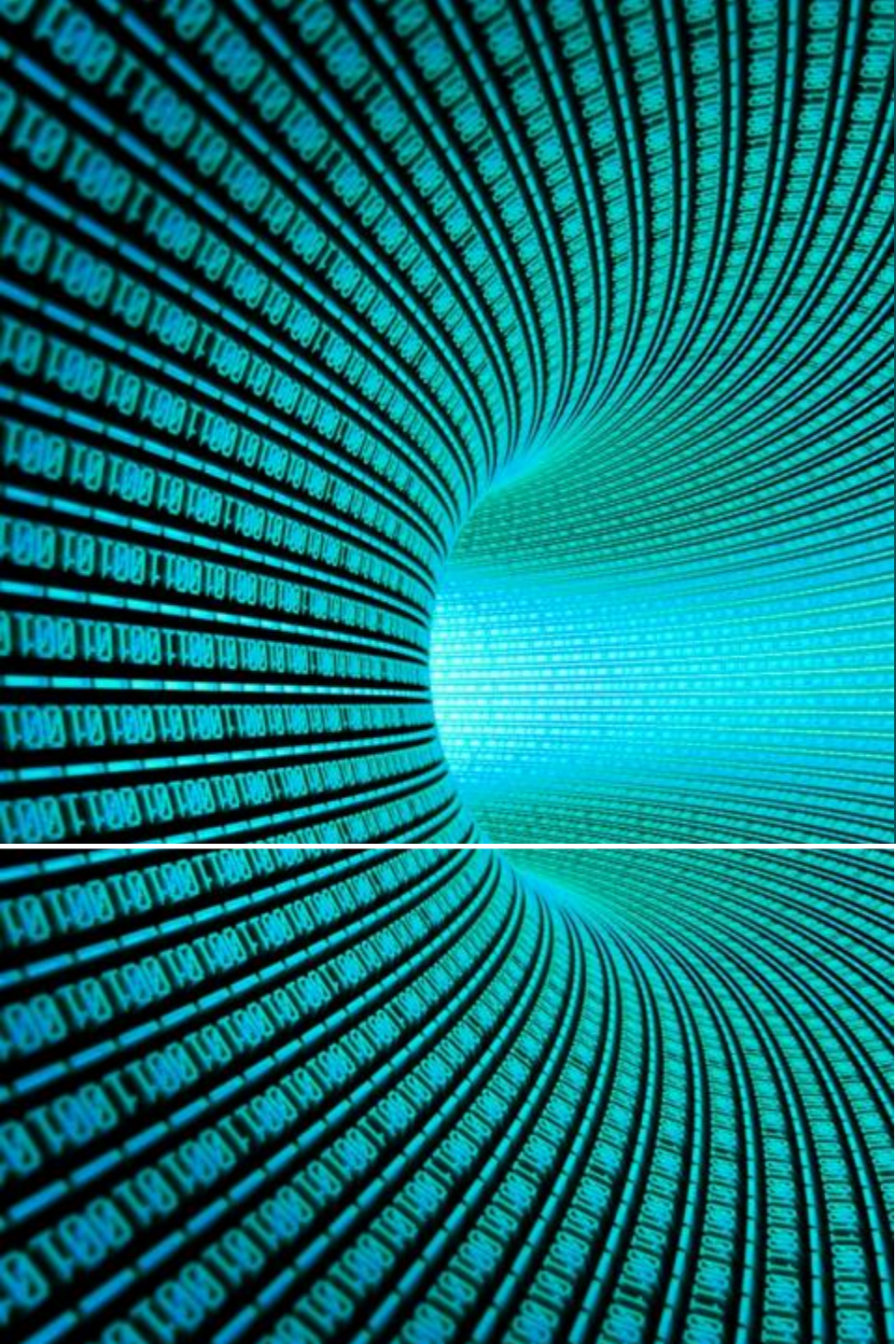




PATTERN BASED METADATA REPOSITORY: TOWARD HIGH QUALITY DATA STANDARDS

11th October 2016, Barcelona
Alan Cantrell

- Julius Kusserow
- Julie James
- Deb Copeland
- Natraj Patro
- Isabelle de Zegher

PAREXEL[®]
YOUR JOURNEY. OUR MISSION.[®]

TABLE OF CONTENT

- Introduction
- Walking the Line
- Introducing PAREXEL Concepts
- Introducing Patterns
- Conclusion

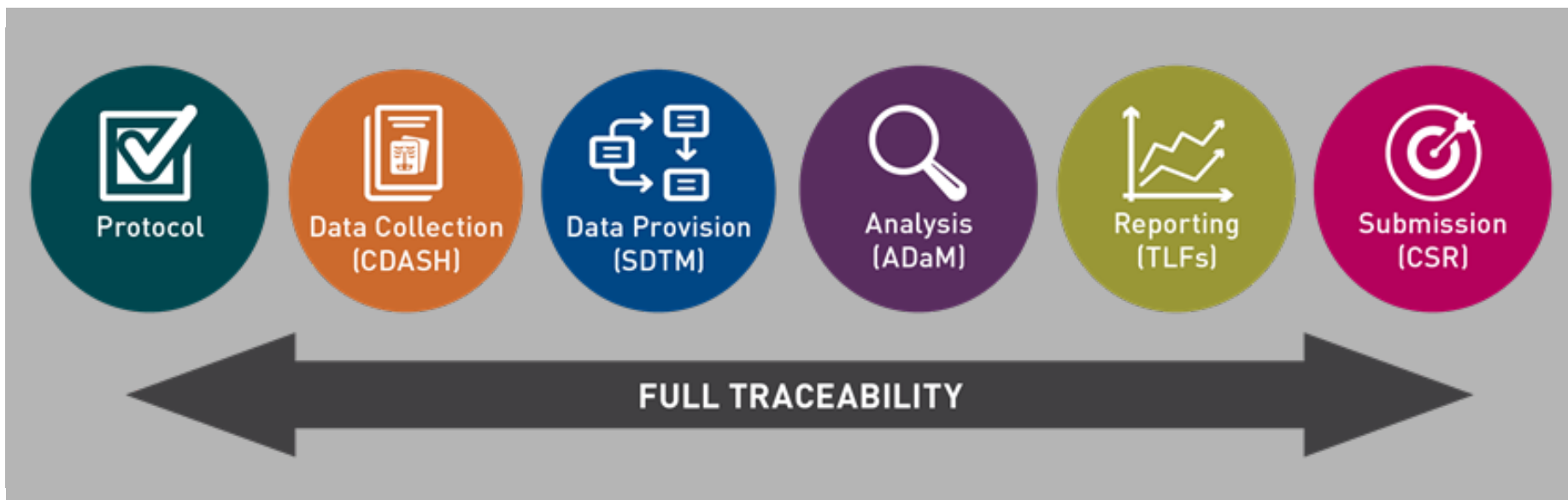
INTRODUCTION

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.

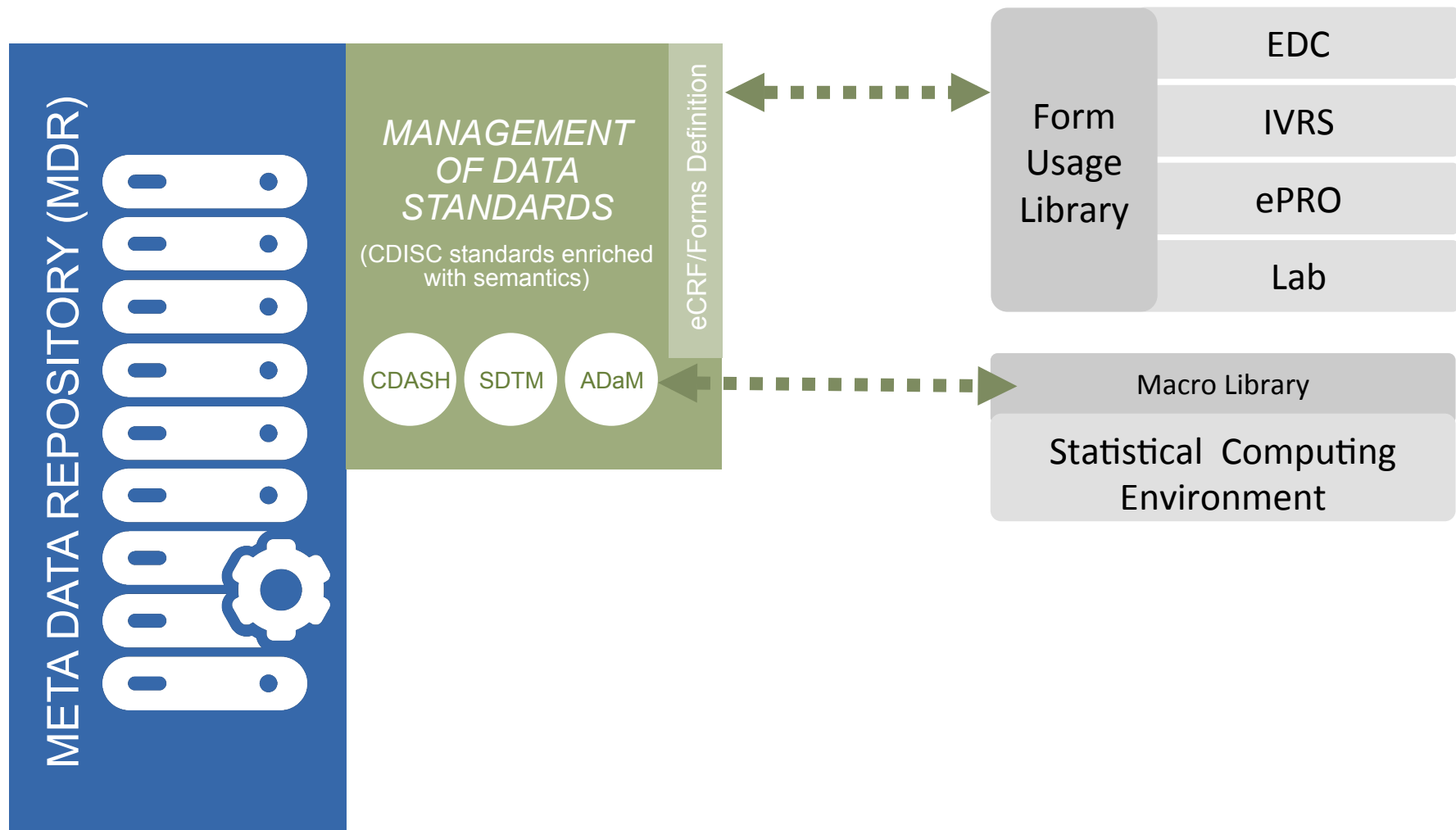
INTRODUCTION

PAREXEL® END-TO-END DATA STANDARDS....



INTRODUCTION

.. WITH CONCEPTS AT THE CORE



INTRODUCTION

.. AND USAGE IN CONTEXT OF STUDY

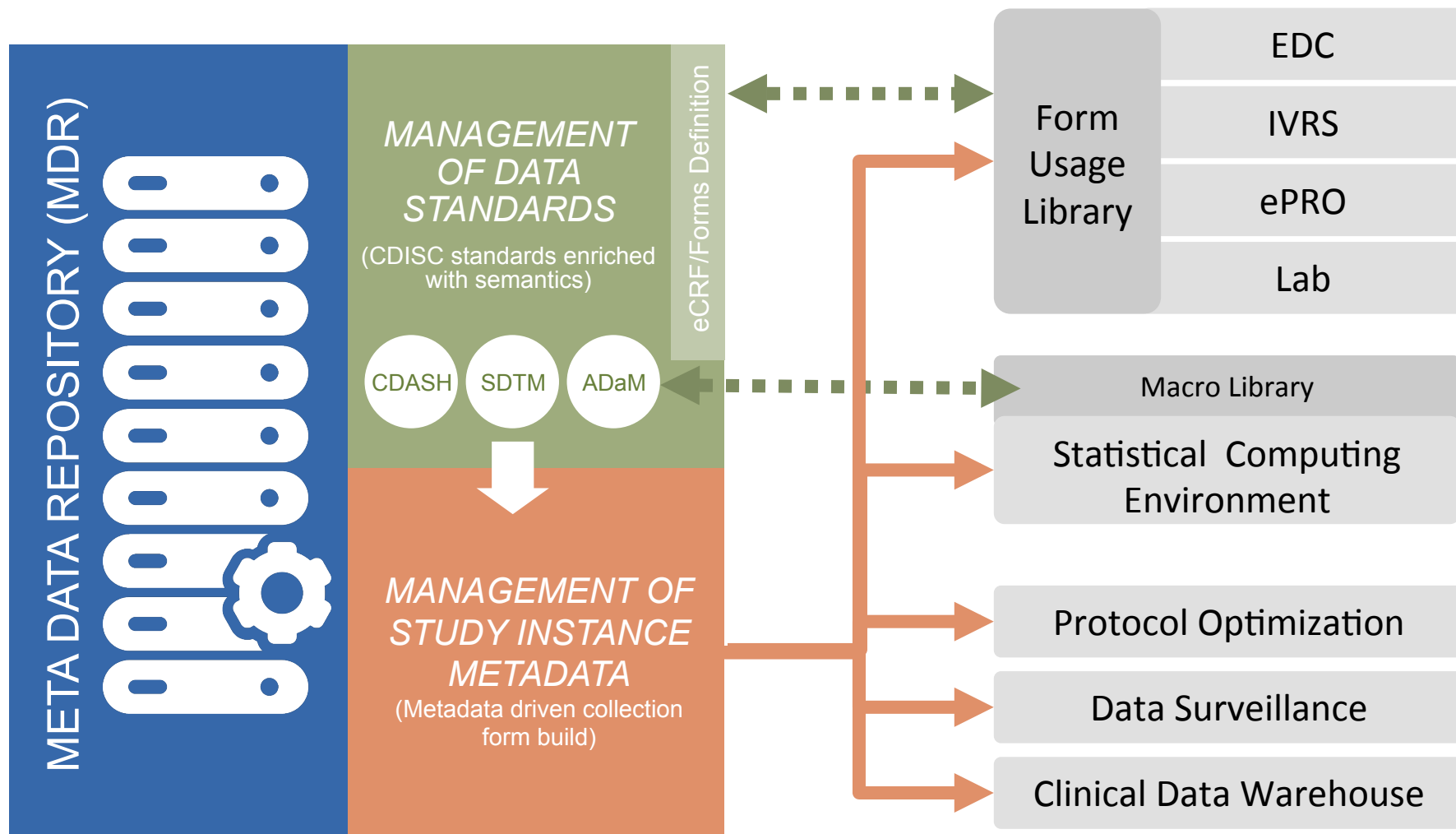


TABLE OF CONTENT

- Introduction
- Walking the Line
- Introducing PAREXEL Concepts
- Introducing Patterns
- Conclusion

WALKING THE LINE - WHAT IS HAPPENING TODAY

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

**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”.

CDASH / Data Collection
AESEV Severity

**Adverse Event
CDASH**

SDTM / Data Tabulation
AESEV Severity/Intensity

**Adverse Event
SDTM**

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)

**Adverse Event
ADaM**

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

**Adverse Event
Domain Definition**

CDASH / Data Collection
AESEV Severity

**Adverse Event
CDASH**



**Adverse Event
SDTM**

A “Domain Definition Variable”

- links the variables across all the components
- reduces the number of variables required to represent the collection, tabulation and analysis

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)

**Adverse Event
ADaM**

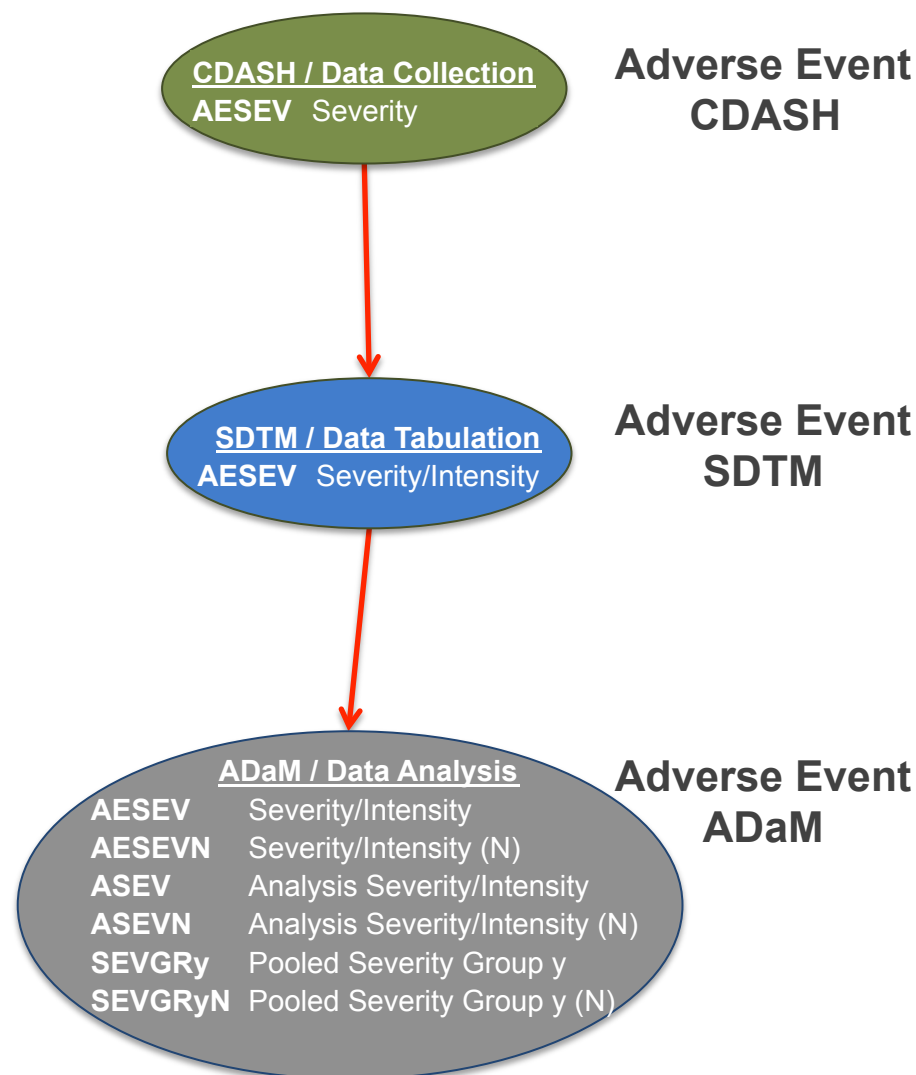
CONCEPT & DOMAIN DEFINITION VARIABLE LINKING VARIABLES

Adverse Event Domain Definition

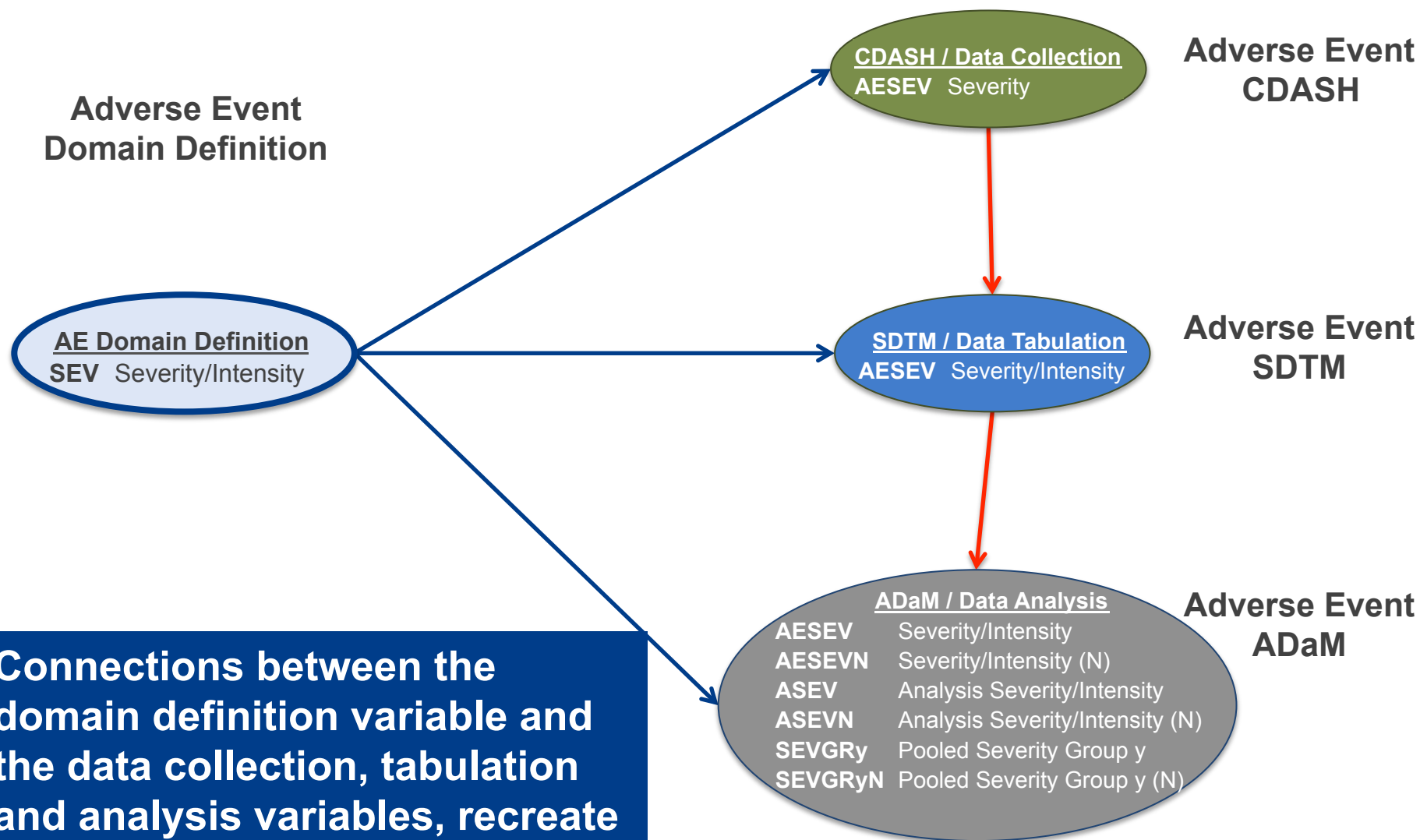


How to populate

- AESEV in AE SDTM domain
- ASEVN in AE ADaM



CONCEPT & DOMAIN DEFINITION VARIABLE RECREATING LINEAR MAPPING



CONCEPT & DOMAIN DEFINITION VARIABLE

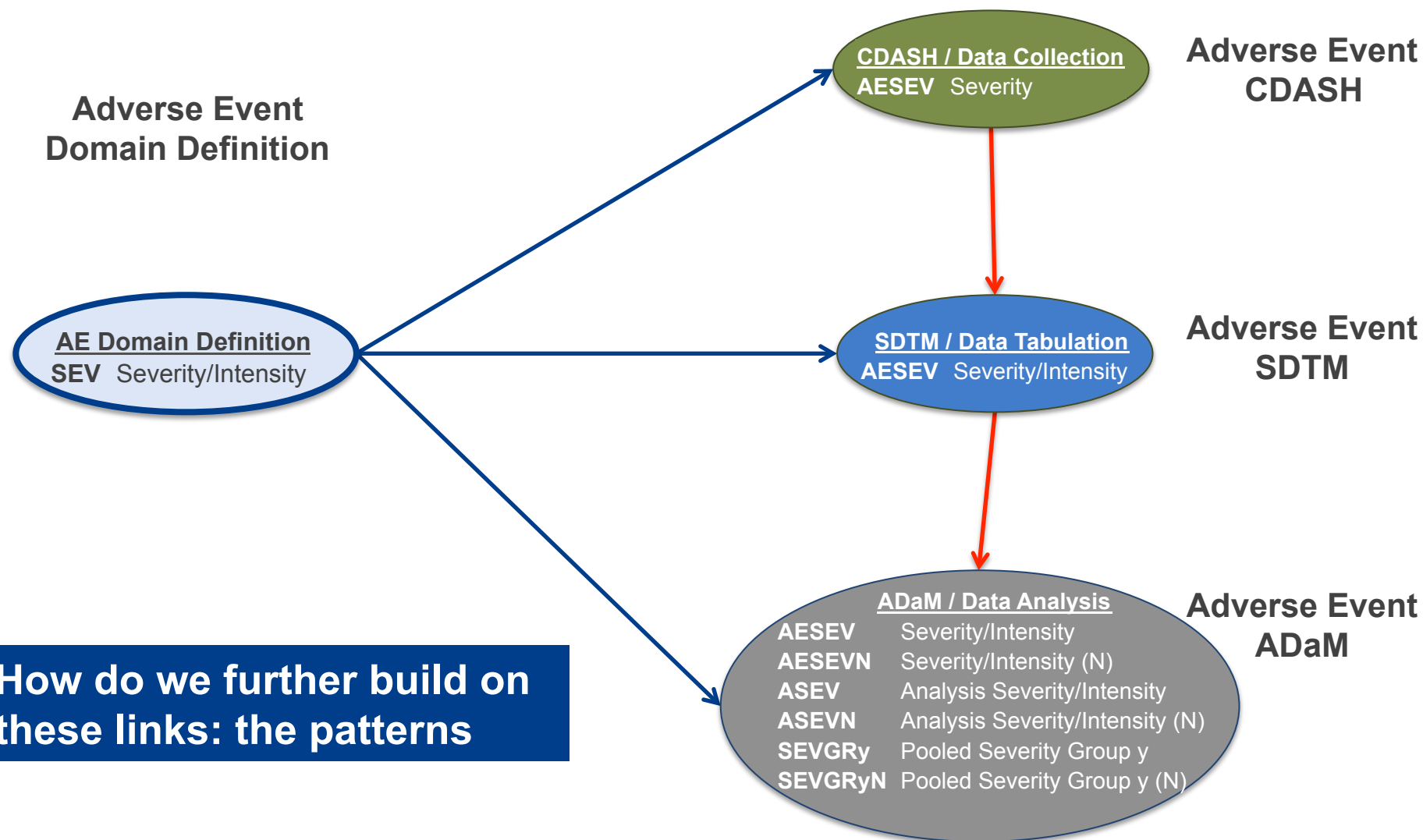


TABLE OF CONTENT

- 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.

Collection / Tabulation	Analysis	Ordinal	Analysis Group	Codelist	Codelist Code	NCI Preferred Term
MILD	Mild	1	Mild/Moderate - 1	C66769	C41338	Mild Adverse Event
MODERATE	Moderate	2	Mild/Moderate - 1	C66769	C41339	Moderate Adverse Event
SEVERE	Severe	3	Severe - 2	C66769	C41340	Severe Adverse Event
<null>	Severe	3	Severe - 2			

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

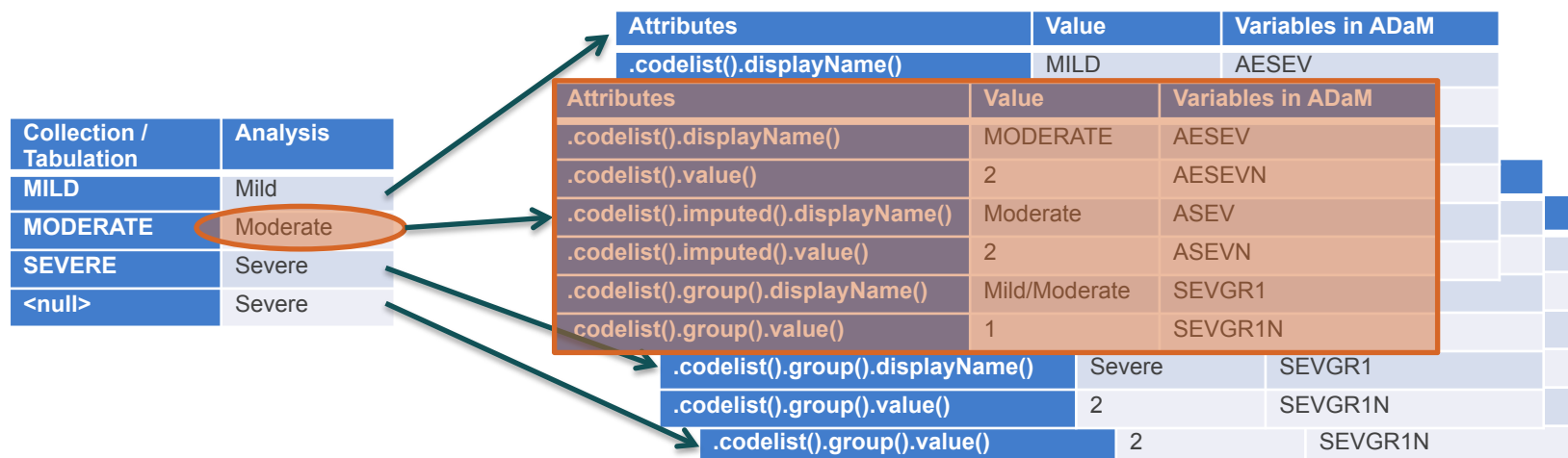
- Example: “MODERATE”

Attributes	Value	Variables in ADaM
.codelist().displayName()	MODERATE	AESEV
.codelist().value()	2	AESEVN
.codelist(). imputed ().displayName()	Moderate	ASEV
.codelist(). imputed ().value()	2	ASEVN
.codelist(). group ().displayName()	Mild/Moderate	SEVGR1
.codelist(). group ().value()	1	SEVGR1N

INTRODUCING PATTERNS

INTRODUCING COMPLEX DATA TYPE (CDT)

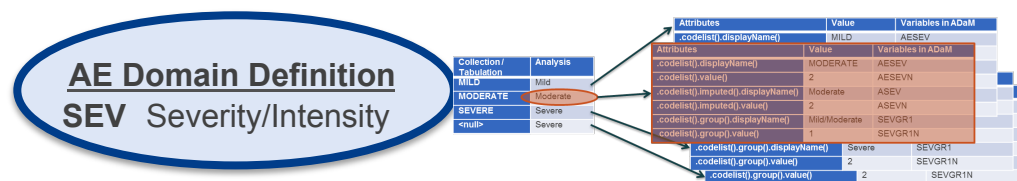
- Data structure storing the logic linking all of the values for collection, tabulation and analysis
- Enable representing multiple data types from the different values.



INTRODUCING PATTERNS

INTRODUCING COMPLEX DATA TYPE (CDT)

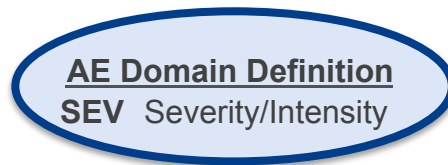
Adverse Event Domain Definition



CDT connected with domain definition variable 'SEV' = a single variable with a complex datatype

INTRODUCING PATTERNS COMPLEX DATA TYPE (CDT)

**Adverse Event
Domain Definition**



Collection / Tabulation	Analysis	Variables in ADaM
MILD	Mild	AESEV
MODERATE	Moderate	AESEVN
SEVERE	Severe	ASEV
Mild/Moderate	Severe	ASEVN
		SEVGR1
		SEVGR1N
		SEVGR2
		SEVGR2N
		SEVGR3
		SEVGR3N
		SEVGR4
		SEVGR4N
		SEVGR5
		SEVGR5N
		SEVGR6
		SEVGR6N
		SEVGR7
		SEVGR7N
		SEVGR8
		SEVGR8N
		SEVGR9
		SEVGR9N
		SEVGR10
		SEVGR10N

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.

CONCEPTS & PATTERNS

A HUB & SPOKE APPROACH

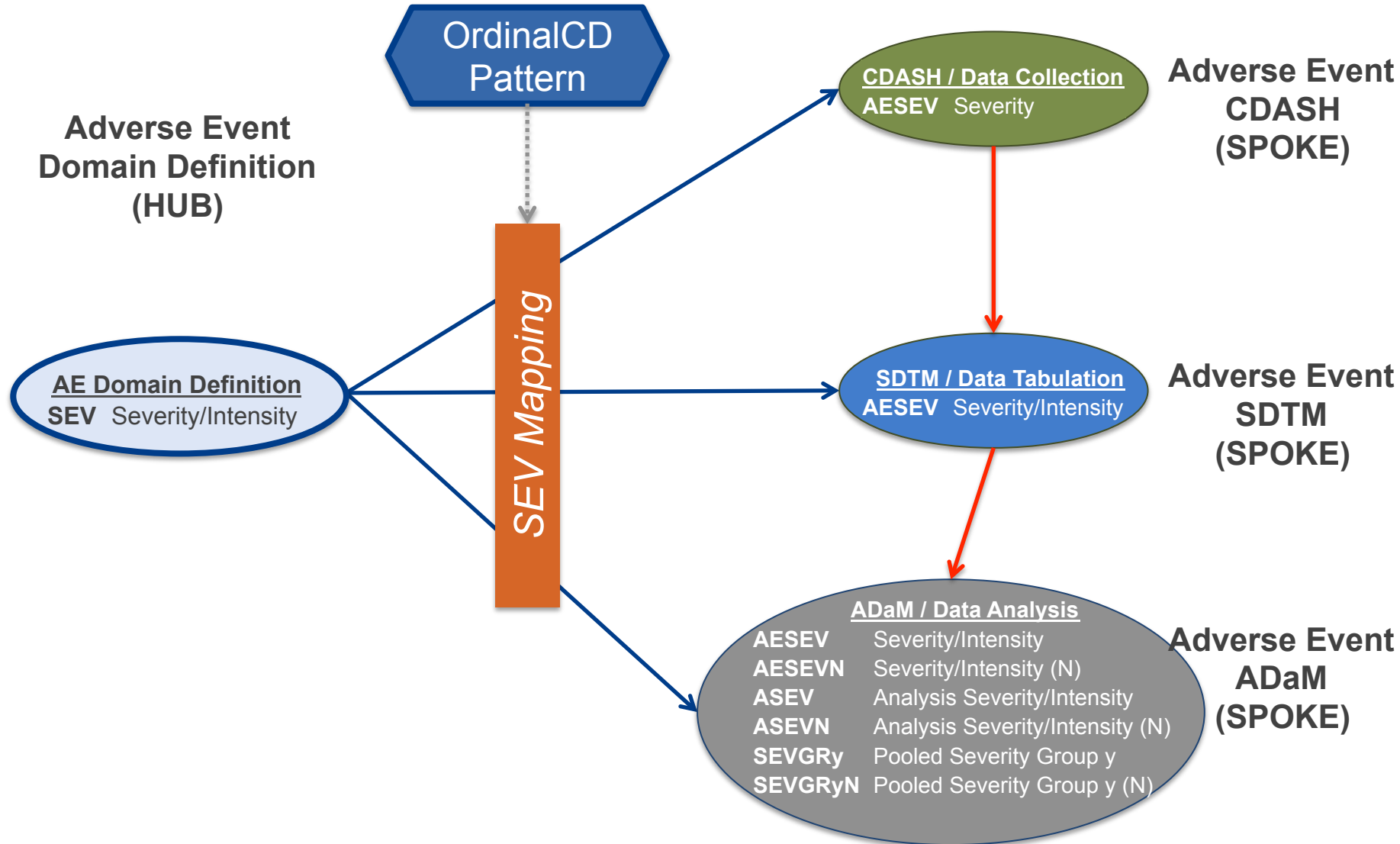


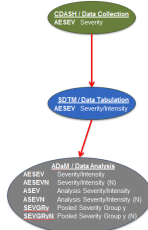
TABLE OF CONTENT

- **Introduction**
- **Walking the Line**
- **Introducing PAREXEL Concepts**
- **Introducing Patterns**
- **Benefits and Conclusion**

BENEFITS AND CONCLUSION

INDEPENDENTLY MAINTAINED STANDARDS

MANUAL MAPPING



FLEXIBILITY

STEPPED APPROACH

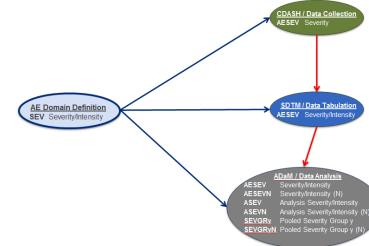
MAPPING INCONSISTENCIES

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*