



USING DEFINE.XML FROM STUDY DESIGN TO SUBMISSION

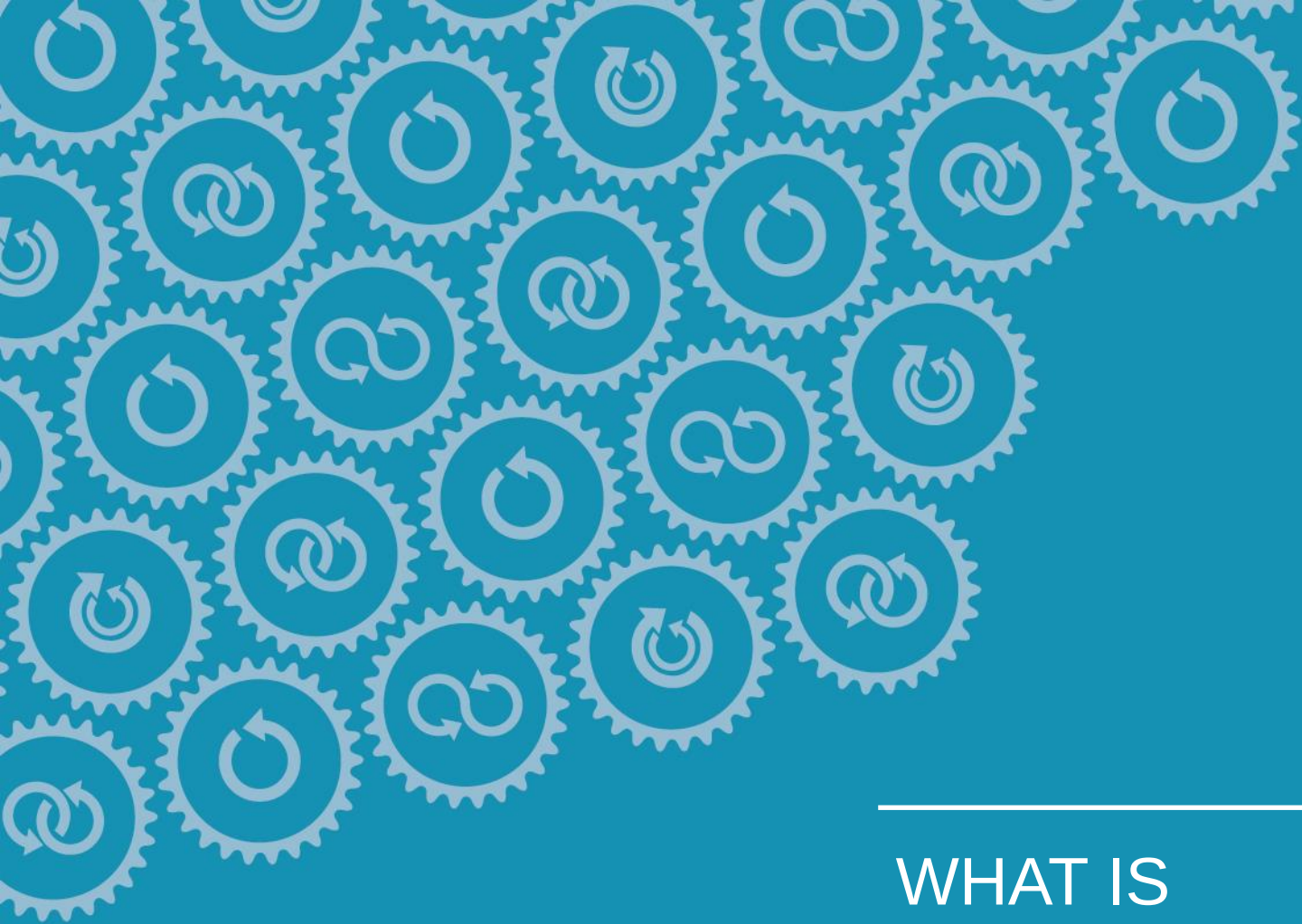
Mark Wheeldon
PhUSE SDE London
May 19, 2015

YOUR CLINICAL
TRIALS AUTOMATED.
EVERYWHERE.

- Introduction
- What is Define-XML?
- Your Clinical Trials Automated. Everywhere.
- How can I Produce Define?
- Dataset Libraries Stored with Define
- Streamlining Study Set-up with Define-XML
- Study Conduct & Analysis Optimized
- Biggest Change in Define-XML 2.0
- Conclusions

- Vendor neutral using CDISC since 2000
- Optimize End-to-End Clinical Trial Process
- Formedix Origin™ - Universal Design Tools
 - Origin Study Modeller and Origin Submission Modeller
- Formedix Transform™ - Universal Execution
 - Data Acquisition – 8 EDCs, ePRO, IVRS
 - Conversion
 - Publication – Annotated CRF / Define.pdf
 - Multi platform & multi vendor
 - Create one library work with all. Train in one work in all
- Formedix Consult™ Consultancy Services
 - Planning, preparation, implementation, multi-EDC execution
 - Data Standards Director has won 2 awards from CDISC





WHAT IS DEFINE-XML?

WHAT IS DEFINE-XML?

WHAT PEOPLE
THINK IT IS

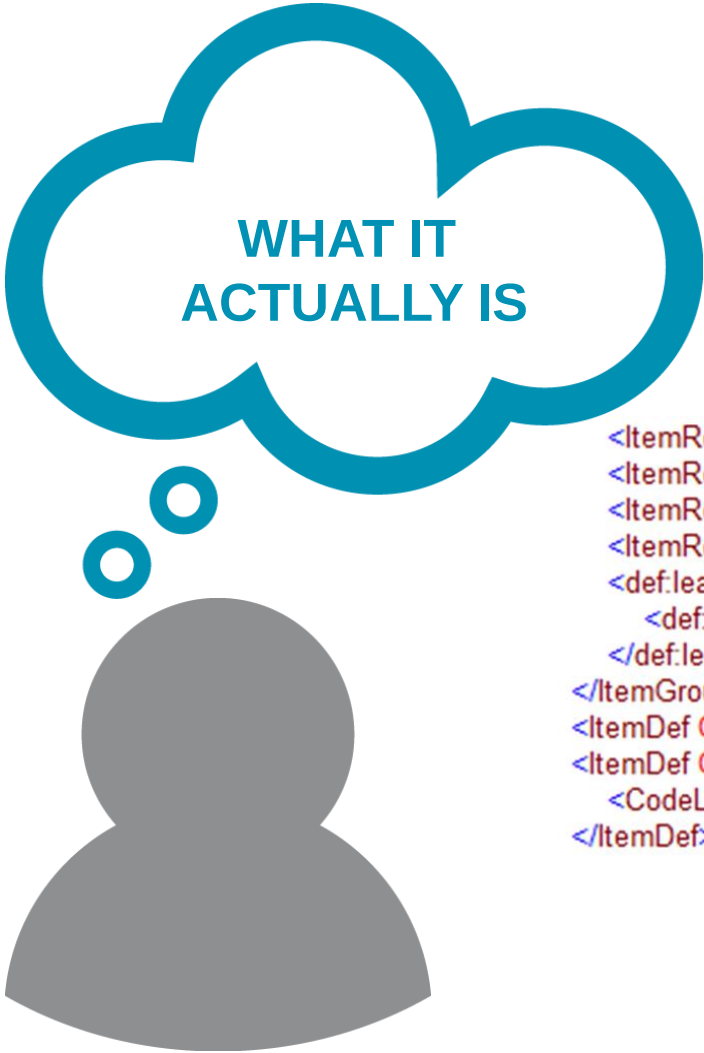


Study: TBTC/USPHS Study 02

Data Definitions: CDISC SDTM

DM (Demographics)

DM (Demographics) -- SPECIAL PURPOSE						dm.xml
Name	Label	Type	Controlled Terms	Origin	Role	Comment
STUDYID	Study Identifier	text	["TBTC022"]	Assigned	Identifier	
DOMAIN	Domain Abbreviation	text	["DM"]	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assigned	Identifier	Protocol ID concatenated with subject ID
SUBJID	Subject Identifier for the Study	text		CRF Page 3	Topic	Site concatenated with subject ID for this study
RFSTDTC	Subject Reference Start Date/Time	text	ISO8601	Assigned	Record Qualifier	Date of first dose of study medication
RFENDTC	Subject Reference End Date/Time	text	ISO8601	Assigned	Record Qualifier	Date of last observation on this study
SITEID	Study Site Identifier	text		CRF Page 1	Record Qualifier	
AGE	Age	integer		Derived	Record Qualifier	Age = integer of (informed consent date - birth date + 1)/365.25
AGEU	Age Units	text	["YEARS"]	Assigned	Record Qualifier	
SEX	Sex	text	["F", "M"]	CRF Page 5	Record Qualifier	
RACE	Race	text	RACE	CRF Page 5	Record Qualifier	The collected term Hispanic was assumed to be an ethnicity not a race. The Race for these subjects is coded as Unknown (Hispanic).
ETHNIC	Ethnicity	text	["HISPANIC", "NON-HISPANIC"]	CRF Page 5	Record Qualifier	Ethnicity was assumed to be non-hispanic for all collected categories of race not identified as Hispanic.



WHAT IT
ACTUALLY IS

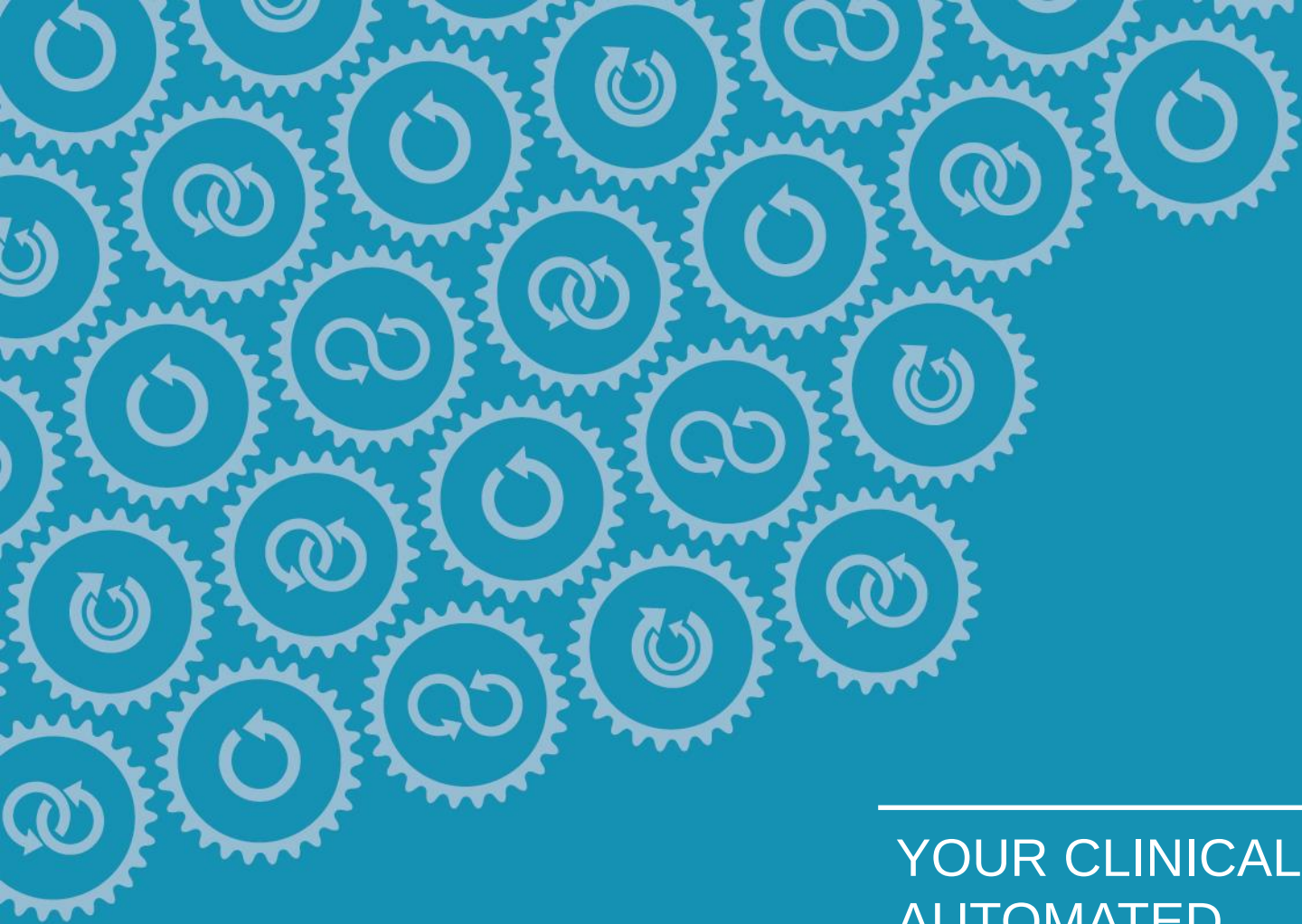
```
<ItemRef ItemOID="SUPPDM.QLABEL" Mandatory="Yes" OrderNumber="7" Role="QUALIF
<ItemRef ItemOID="SUPPDM.QVAL" Mandatory="Yes" OrderNumber="8" Role="QUALIFIE
<ItemRef ItemOID="SUPPDM.QORIG" Mandatory="Yes" OrderNumber="9" Role="QUALIFIE
<ItemRef ItemOID="SUPPDM.QEVAL" Mandatory="No" OrderNumber="10" Role="QUALIFIE
<def:leaf xlink:href="suppdm.xpt" ID="ARCHIVELOCATION.SUPPDM">
  <def:title>suppdm.xpt</def:title>
</def:leaf>
</ItemGroupDef>
<ItemDef OID="DM.STUDYID" Name="STUDYID" DataType="text" Length="200" Origin="CRF"
<ItemDef OID="DM.DOMAIN" Name="DOMAIN" DataType="text" Length="2" Origin="Derived"
  <CodeListRef CodeListOID="CL.DM"/>
</ItemDef>
```



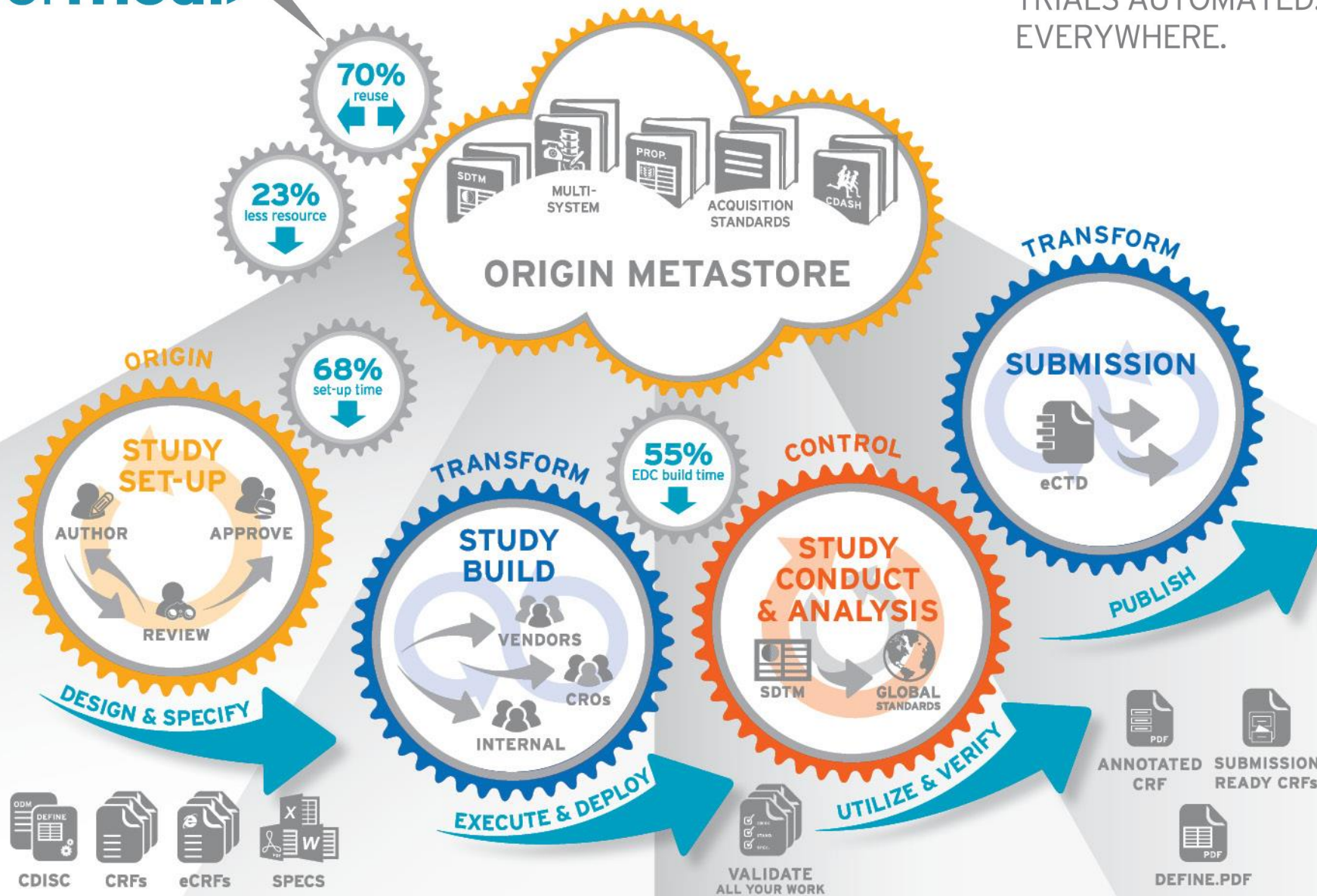
**THIS IS WHO
NEEDS IT...**

- 2005: "FDA Study Data Specifications" recommended using Define-XML for submissions
- 2011: "CDER Data Standards Common Issues Document" stated that "a properly functioning Define.xml file is an important part of the submission of standardized electronic datasets and should not be considered optional."

- Study
 - Name, Description, Link to Annotated CRF, Link to Supplemental Docs etc.
- Domains
 - Name, Description, Link to dataset etc.
- Variables
 - Name, Label, Data type, Origin, Role, Comment etc.
- Derivations
- Controlled Terminology
- SDTM Concepts
 - Split Domains, Supplemental Qualifiers, Value Level Metadata

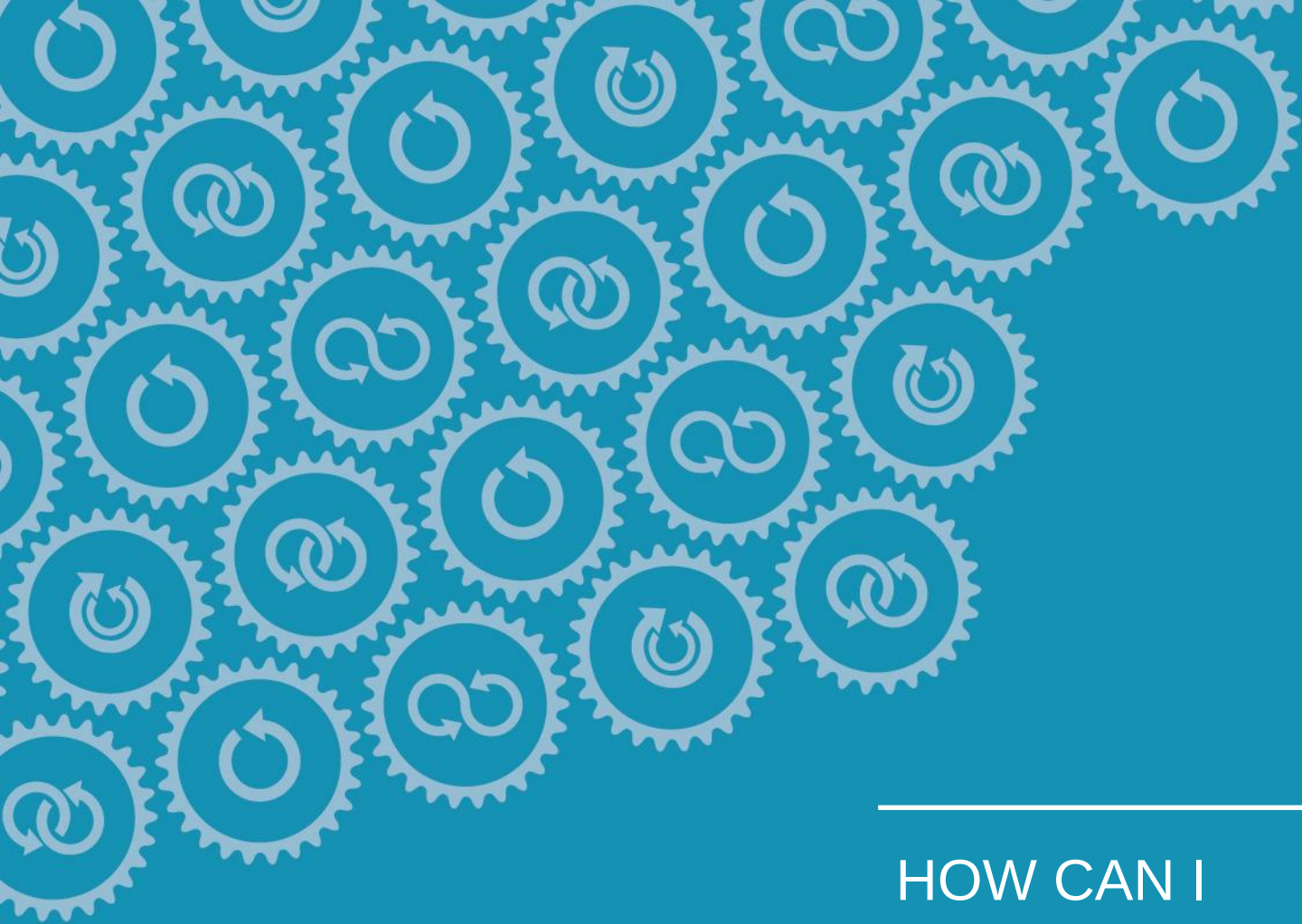


YOUR CLINICAL TRIALS
AUTOMATED.
EVERYWHERE.



- Libraries and Templates
 - Re-use of all dataset designs across end-to-end
 - Proprietary EDC, SDTM, ADaM
 - Standards Governance
 - Mappings and Presentation
- Study Set-up
 - Protocol to Submission Process
 - Dataset designs without the ambiguity
- Study Build
 - Automated configuration of ETL
 - Automated configuration of EDC exporters

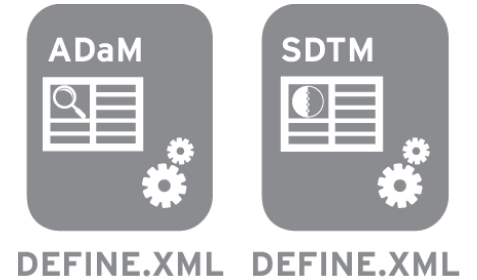
- Study Conduct and Analysis
 - Metadata driven dataset conversions
 - Automated clinical data repository load
 - Automated validation
 - Study vs. Standards
 - “As specified” vs. “As delivered”
- Submission
 - Define.pdf and Define.html from Define.xml
 - TOC, bookmarking and hyperlinks automatically
- Who said Define was only about Submissions?



HOW CAN I
PRODUCE DEFINE?



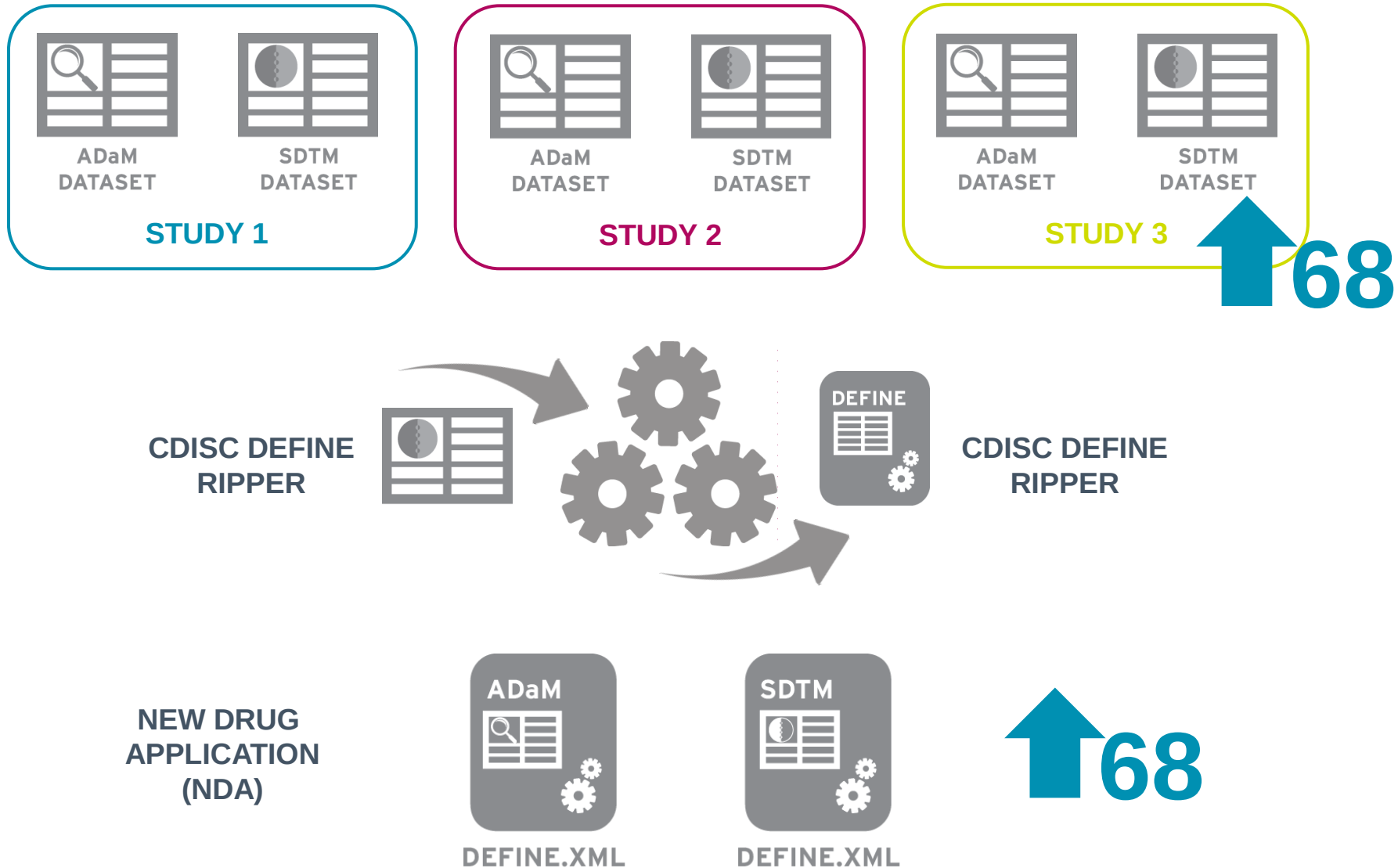
TEMPLATES



DATASET
DESIGN TOOL



POSTHUMOUS DEFINE CREATION



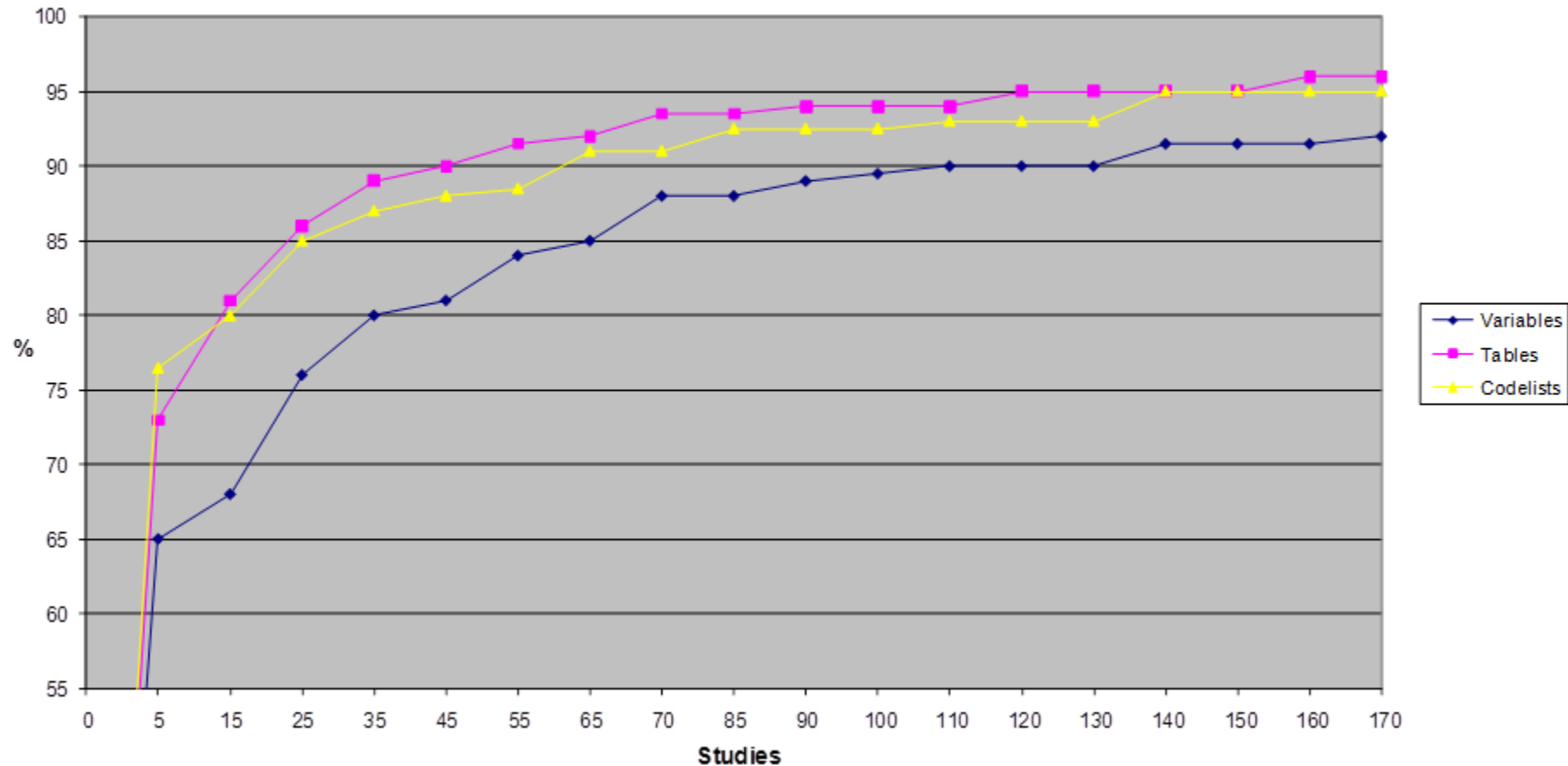


DATASET LIBRARIES
STORED WITH DEFINE

- Every dataset design you require
 - From EDC extract through SDTM to ADaM
- For SDTM
 - “Common to All” = SDTM-IG in define
 - “Therapeutic Specific” = SDTM superset (GOC)
 - Split domains lead to better metadata management
 - Multiple QS domains with multiple value lists
- So much more than dataset designs
 - Mappings
 - Presentation and named destinations
- Converters get you there quickly ...
 - Bye Bye Excel ...

LIBRARIES REDUCE DESIGN EFFORT & TIME

GDD Reuse Rate



Source: Schering Dublin May 2003



- Quality is built in up-front not tested in
- CDISC Libraries/Metadata Repositories
 - Vendor neutral and portable
- Re-use of all dataset design components
 - Across end-to-end clinical trial
 - Across multiple platforms
 - Across multiple vendors & systems
- Design once
 - Only test new content or changed content
 - Standardized EDC, target SDTM datasets & mappings



STREAMLINING STUDY SET-UP WITH DEFINE

- Protocol Content to CRFs
- “Common to All”
 - CDASH Forms and SDTM Implementation Guide (IG)
- New Content?
 - Look to General Observation Classes (GOC)
 - Finding, Intervention or Event?
 - Gives you Identifier, Topic, Qualifier & Timing variables
- Create Form
 - Backfill question text
- SDTM target
 - SDTM Annotations and Mappings

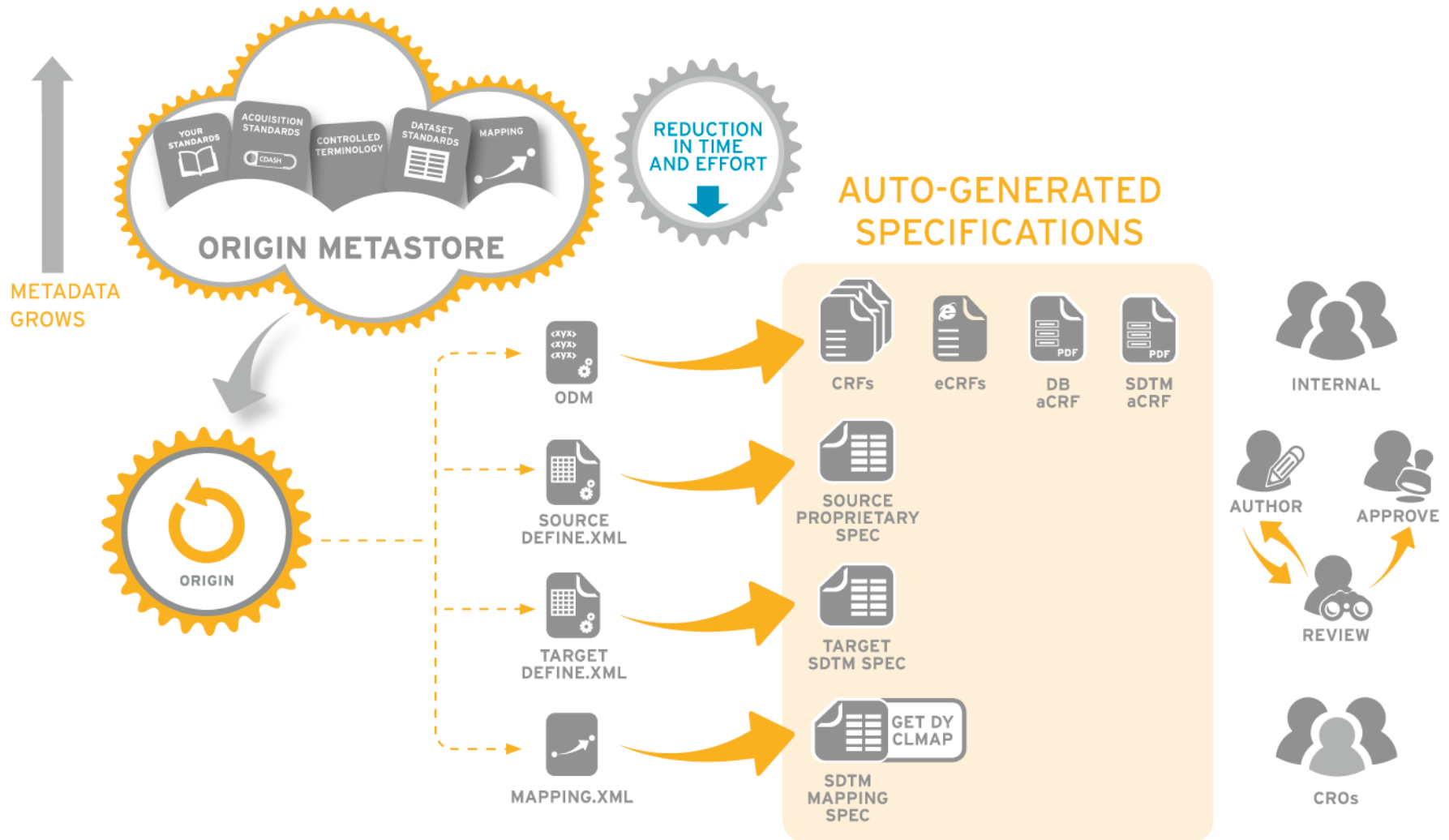
STUDY SET-UP

THE EASY WAY WITH DEFINE



- Human Readable Specification
- Been using this way since 2006
 - With CROs, Partners & Clinical Data Warehouses
- Eliminates ambiguity & variance
 - Datasets, variables, length, orders
 - Value lists & supplemental qualifiers
 - Controlled Terms (CT) & recently version of CT
 - Variance between versions of SDTM
- Exploit Define.xml extensions
 - Reason for variables use, Pseudo-code for mappings
- User Defined Templates





Study: Demo

Data Definitions: SDTM-IG - Demo 1.0

FT (Pulmonary Function Tests)

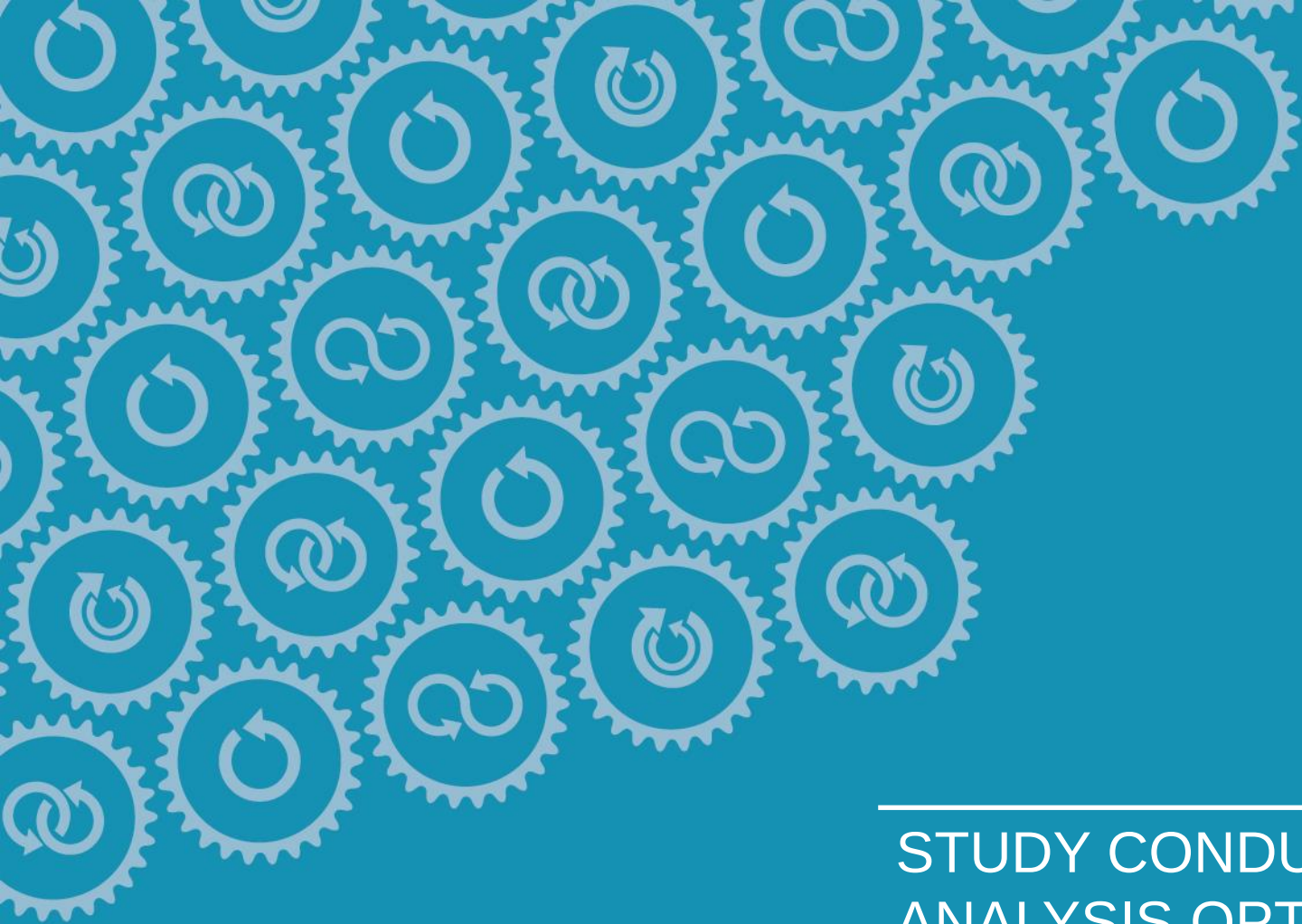
FT, Pulmonary Fu
H-V mapping mea
ValueList ItemDef
Name
La
STUDYID
DOMAIN
USUBJID
FTSEQ
USUBJID
FTSEQ
FTTESTCD
FTTEST
FTPOS
FTORRES
FTORRESU
FTSTRESC
FTSTRESN
FTSTRESU
FTSTAT
FTREASND
FTBLFL

FT (Pulmonary Function Tests) -- Findings						
Name	Label	Type	Controlled Terms	Origin	Role	Comment
STUDYID	Study Identifier	text		CRF Page 1	Identifier	
DOMAIN	Domain Abbreviation	text	["FT"]	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		CRF Page 1	Identifier	
FTSEQ	Sequence Number	float		Derived	Identifier	
FTTESTCD	Pulmonary Function Test Short Name	text	FTTESTCD	Assigned	Topic	
FTTEST	Pulmonary Function Test	text	FTTEST	Assigned	Synonym Qualifier	
FTPOS	Position of Subject During Observation	text	Position	CRF Pages 2, 3, 4, 5, 6, 7	Record Qualifier	
FTORRES	Result or Finding in Original Units	text		CRF Pages 2, 3, 4, 5, 6, 7	Result Qualifier	
FTORRESU	Original Units	text	Unit	CRF Pages 2, 3, 4, 5, 6, 7	Variable Qualifier	
FTSTRESC	Character Result/Finding in Std Format	text		CRF Pages 2, 3, 4, 5, 6, 7	Result Qualifier	
FTSTRESN	Numeric Result/Finding in Standard Units	float		CRF Pages 2, 3, 4, 5, 6, 7	Result Qualifier	
FTSTRESU	Standard Units	text	Unit	CRF Pages 2, 3, 4, 5, 6, 7	Variable Qualifier	
FTSTAT	Completion Status	text	["NOT DONE"]	CRF Pages 2, 3, 4, 5, 6, 7	Record Qualifier	
FTREASND	Reason Not Done	text		CRF Pages 2, 3, 4, 5, 6, 7	Record Qualifier	
FTBLFL	Baseline Flag	text	["N", "NA", "U", "Y"]	Derived	Record Qualifier	

given in "STAT

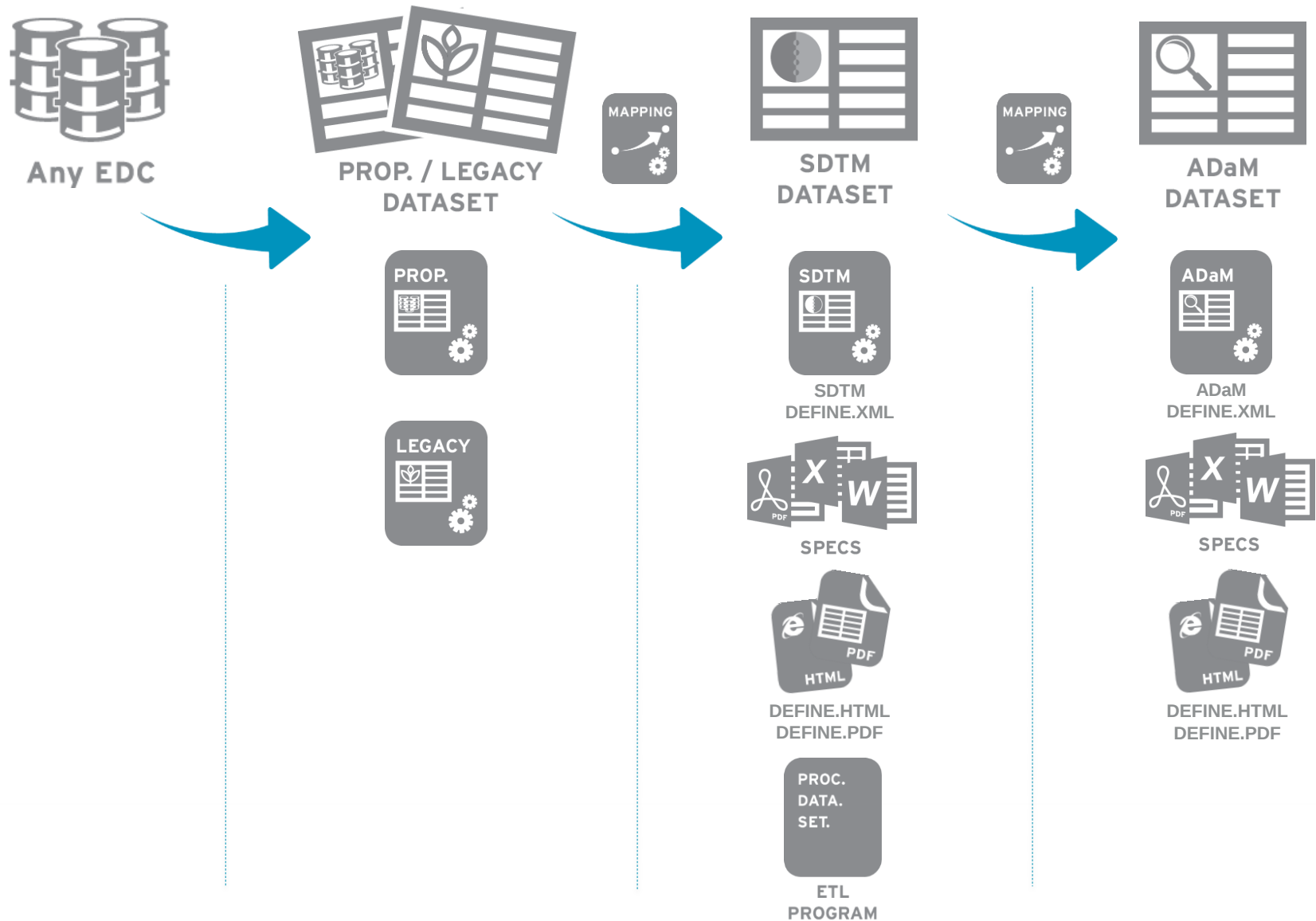
Page 3 of 28

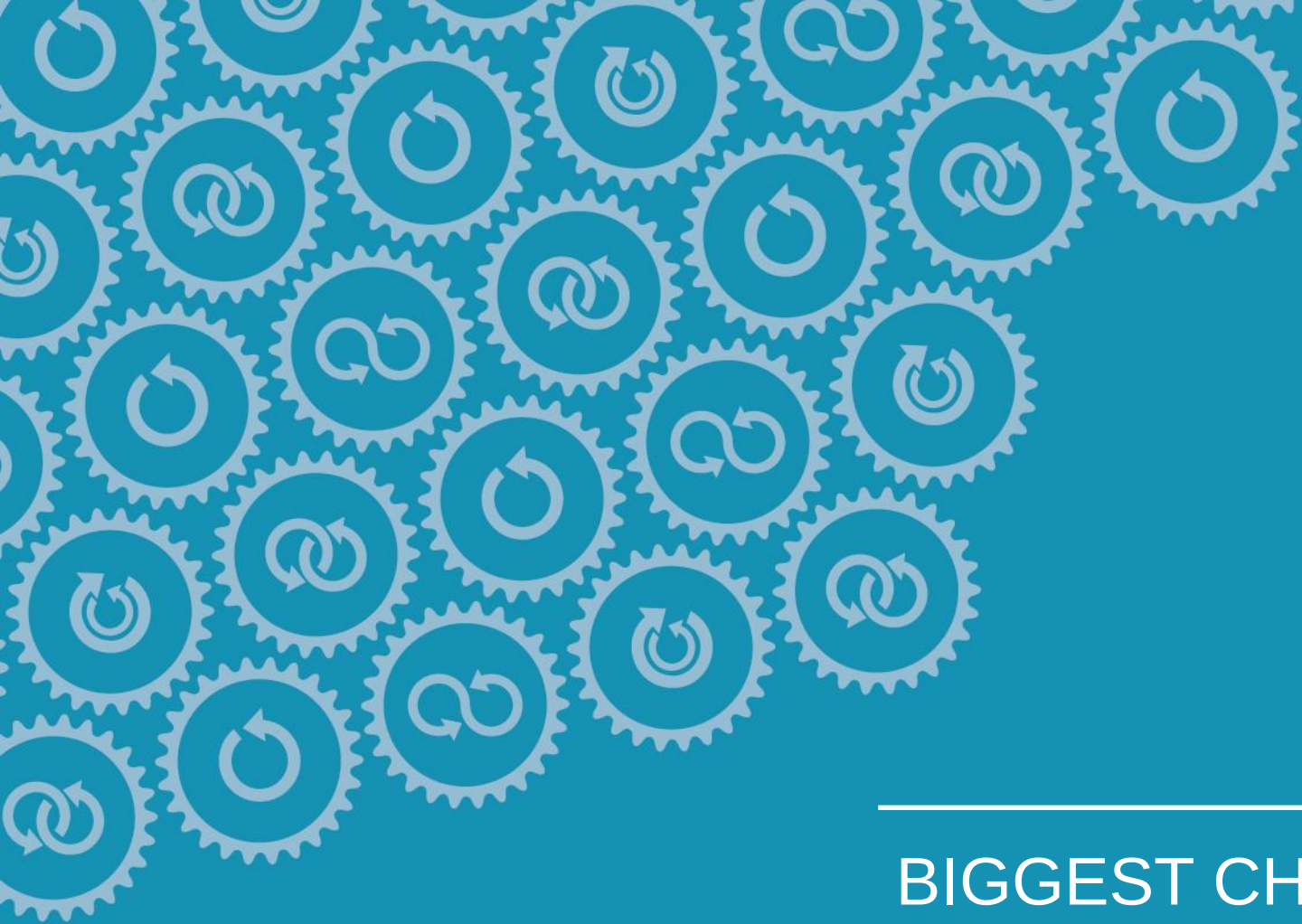
specify



STUDY CONDUCT &
ANALYSIS OPTIMIZED
WITH DEFINE

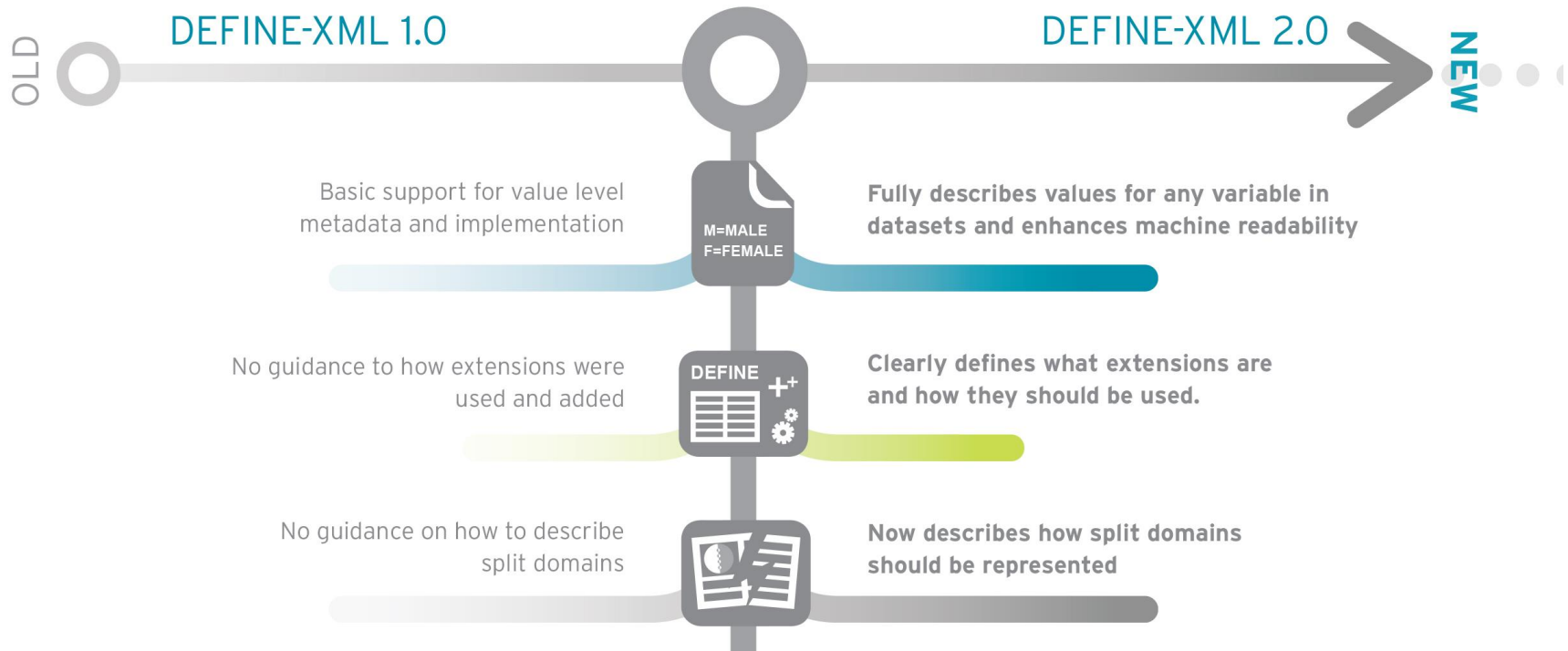
DEFINE DRIVEN DATA TRANSFORMATIONS



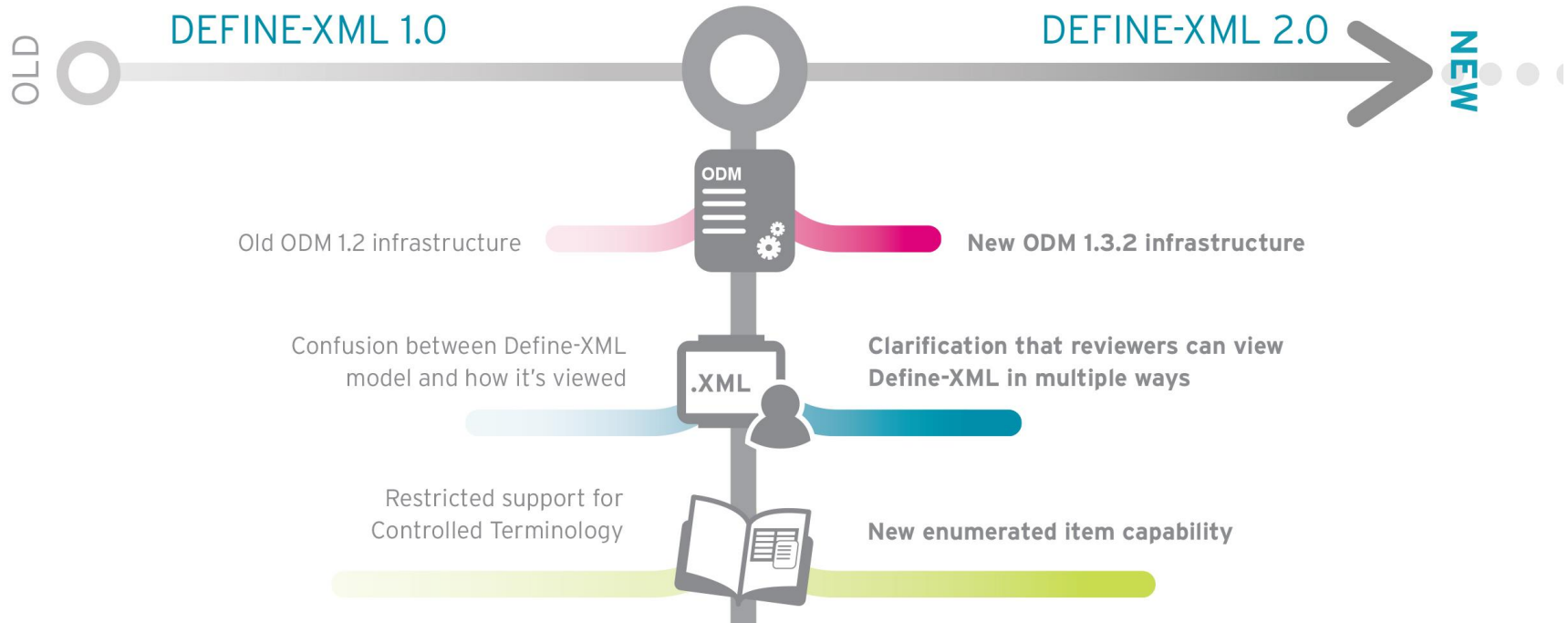


BIGGEST CHANGE IN DEFINE-XML 2.0

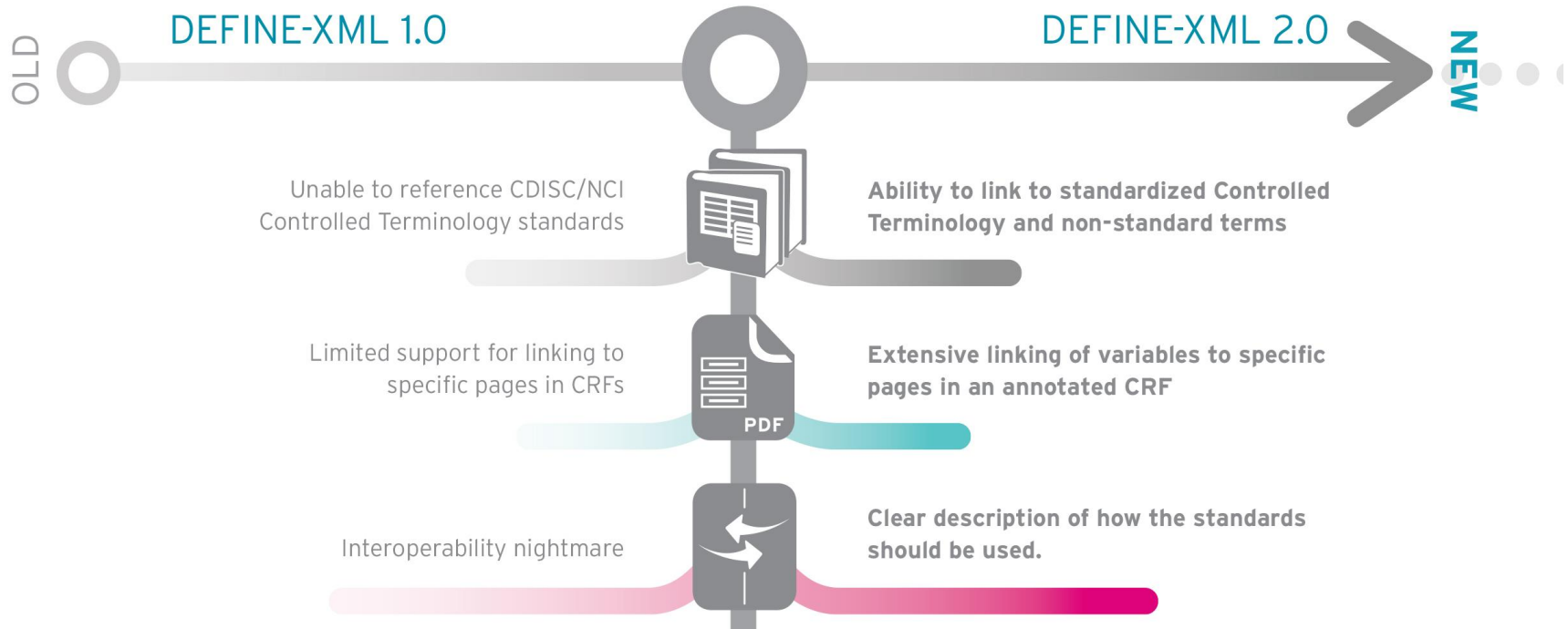
WHAT'S NEW IN DEFINE 2.0



WHAT'S NEW IN DEFINE 2.0



WHAT'S NEW IN DEFINE 2.0



VITAL SIGNS DATASET – DEFINE-XML 1.0

CDISC SDTM 3.1.2 Domain Definitions

Dataset	Description	Structure	Purpose	Keys	Location
VS	Vital Signs	One record per vital sign measurement per time point per visit per subject	Tabulation	STUDYID, DOMAIN, USUBJID, VSSEQ	vs.xpt

VS, Vital Signs, FINDINGS, CDISC SDTM 3.1.2, One record per vital sign measurement per time point per visit per subject, [vs.xpt](#)

Name	Label	Type	Controlled Terms	Origin	Role	Comment
STUDYID	Study Identifier	text			Identifier	
DOMAIN	Domain Abbreviation	text	["VS"]		Identifier	
USUBJID	Unique Subject Identifier	text			Identifier	
VSSEQ	Sequence Number	float			Identifier	
VSTESTCD VL1	Vital Signs Test Short Name	text	Vital Signs Test Code (C66741)		Topic	
VSTEST	Vital Signs Test Name	text	Vital Signs Test Name (C67153)		Synonym Qualifier	
VSORRES	Result or Finding in Original Units	text			Result Qualifier	
VSORRESU	Original Units	text	Units for Vital Signs Results (C66770)		Variable Qualifier	
VSSTRESC	Character Result/Finding in Std Format	text			Result Qualifier	
VSSTRESN	Numeric Result/Finding in Standard Units	float			Result Qualifier	
VSSTRESU	Standard Units	text	Units for Vital Signs Results (C66770)		Variable Qualifier	
VSBLFL	Baseline Flag	text	No Yes Response (C66742)		Record Qualifier	
VISITNUM	Visit Number	float			Timing	
VSDTC	Date/Time of Measurements	datetime	ISO 8601 (Dates/Times)		Timing	

Value Lists

Value Level Metadata for Vital Signs (VS) Domain [VL1](#)

Source Variable	Value	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSTESTCD	TEMP	Temperature	float				
VSTESTCD	HR	Heart Rate	integer				

VITAL SIGNS DATASET – DEFINE-XML 2.0

VS, Vital Signs, Findings, CDISC SDTM 3.1.2, One record per vital sign measurement per time point per visit per subject

Name	Label	Type	Controlled Terms	Origin	Role	Comment
STUDYID	Study Identifier	text			Identifier	
DOMAIN	Domain Abbreviation	text	["VS"]		Identifier	
USUBJID	Unique Subject Identifier	text			Identifier	
VSSEQ	Sequence Number	float			Identifier	
VSTESTCD	Vital Signs Test Short Name	text	Vital Signs Test Code (C66741)		Topic	
VSTEST	Vital Signs Test Name	text	Vital Signs Test Name (C67153)		Synonym Qualifier	
VSORRES	Result or Finding in Original Units	text			Result Qualifier	
VSORRESU	Original Units	text	Units for Vital Signs Results (C66770)		Variable Qualifier	
VSSTRESC	Character Result/Finding in Std Format	text			Result Qualifier	
VSSTRESN	Numeric Result/Finding in Standard Units	float			Result Qualifier	
VSSTRESU	Standard Units	text	Units for Vital Signs Results (C66770)		Variable Qualifier	
VSBLFL	Baseline Flag	text	No Yes Response (C66742)		Record Qualifier	
VISITNUM	Visit Number	float			Timing	
VSDTC	Date/Time of Measurements	datetime	ISO 8601 (Dates/Times)		Timing	

Value Lists

Value List for Vital Signs (VS) Domain ^{VL1}

Source Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSORRES	Temperature	float				WHERE VSTESTCD EQ TEMP and COUNTRY IN UK, AUSTRALIA WHERE VSTESTCD EQ TEMP and COUNTRY IN USA
VSORRES	Heart Rate	integer				WHERE VSTESTCD EQ HR

Value List for Vital Signs (VS) Domain ^{VL2}

Source Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSORRESU	Temperature	text	["C"]			WHERE VSTESTCD EQ TEMP and COUNTRY IN UK, AUSTRALIA
VSORRESU	Temperature	text	["F"]			WHERE VSTESTCD EQ TEMP and COUNTRY IN USA
VSORRESU	Heart Rate	text	["BPM"]			WHERE VSTESTCD EQ HR

DEFINE-XML 2.0 WHERE CLAUSES AND WHAT THEY MEAN TO YOU



Value Lists

Value List for Vital Signs (VS) Domain ^{VL1}

Source Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSORRES	Temperature	float				WHERE VSTESTCD EQ TEMP and COUNTRY IN UK, AUSTRALIA WHERE VSTESTCD EQ TEMP and COUNTRY IN USA
VSORRES	Heart Rate	integer				WHERE VSTESTCD EQ HR

Value List for Vital Signs (VS) Domain ^{VL2}

Source Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSORRESU	Temperature	text	["C"]			WHERE VSTESTCD EQ TEMP and COUNTRY IN UK, AUSTRALIA
VSORRESU	Temperature	text	["F"]			WHERE VSTESTCD EQ TEMP and COUNTRY IN USA
VSORRESU	Heart Rate	text	["BPM"]			WHERE VSTESTCD EQ HR

DEFINE-XML 2.0 WHERE CLAUSES AND WHAT THEY MEAN TO YOU



Value Lists

Value List for Vital Signs (VS) Domain VL1

Source Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSORRES	Temperature	float				WHERE VSTESTCD EQ TEMP and COUNTRY IN UK, AUSTRALIA WHERE VSTESTCD EQ TEMP and COUNTRY IN USA
VSORRES	Heart Rate	integer				WHERE VSTESTCD EQ HR

Value List for Vital Signs (VS) Domain VL2

Source Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSORRESU	Temperature	text	["C"]			WHERE VSTESTCD EQ TEMP and COUNTRY IN UK, AUSTRALIA
VSORRESU	Temperature	text	["F"]			WHERE VSTESTCD EQ TEMP and COUNTRY IN USA
VSORRESU	Heart Rate	text	["BPM"]			WHERE VSTESTCD EQ HR

01-0002	8	HR	Heart Rate	68	BPM	68	68	BPM
01-0002	9	TEMP	Temperature	97.8	F	36.56	36.56	C
01-0002	10	TEMP	Temperature	97.8	F	36.56	36.56	C
01-0002	11	TEMP	Temperature	97.3	F	36.28	36.28	C

Define-XML 2.0 allows you to
define where each value applies

- Vendor Neutral Portable Dataset Libraries
 - Proprietary, SDTM, ADaM, Other ...
- Auto-generate Study Set-up Specifications
- Streamline Downstream Data Transformations
- Machine Validate Study Deliverables/Standards
- Define-XML 2.0
 - Allows Value Level Metadata for all Variables
 - Is more robust and machine readable
 - Is more interoperable than ever before
- Who said Define was only about Submissions?



YOUR CLINICAL TRIALS
AUTOMATED. EVERYWHERE.

www.formedix.com

formedix US

15 New England Executive Park, Burlington MA, USA, 01803
t: +1 781 685 4995 f: +1 781 685 4601

formedix UK

The Connect Building, 59 Bath Street, Glasgow G2 2DH
t: +44 141 353 4010 f: +44 141 258 4135