PRACTICAL IMPLEMENTATION OF DEFINE.XML

Mark Wheeldon, Formedix

CDISC UK Network

June 7, 2016

YOUR CLINICAL
TRIALS AUTOMATED.
EVERYWHERE.

AGENDA

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

FORMEDIX INTRODUCTION

FOUNDED IN 2000

PRODUCTS & SERVICES

CDISC ODM Certified

CDISC Registered Solutions Provider

BASED ON CDISC

MULTI EDC INTEGRATION

VENDOR NEUTRAL

medidata

Mi-Co

nextrials our technology your success

OmniComm eClinical Solutions for Life™

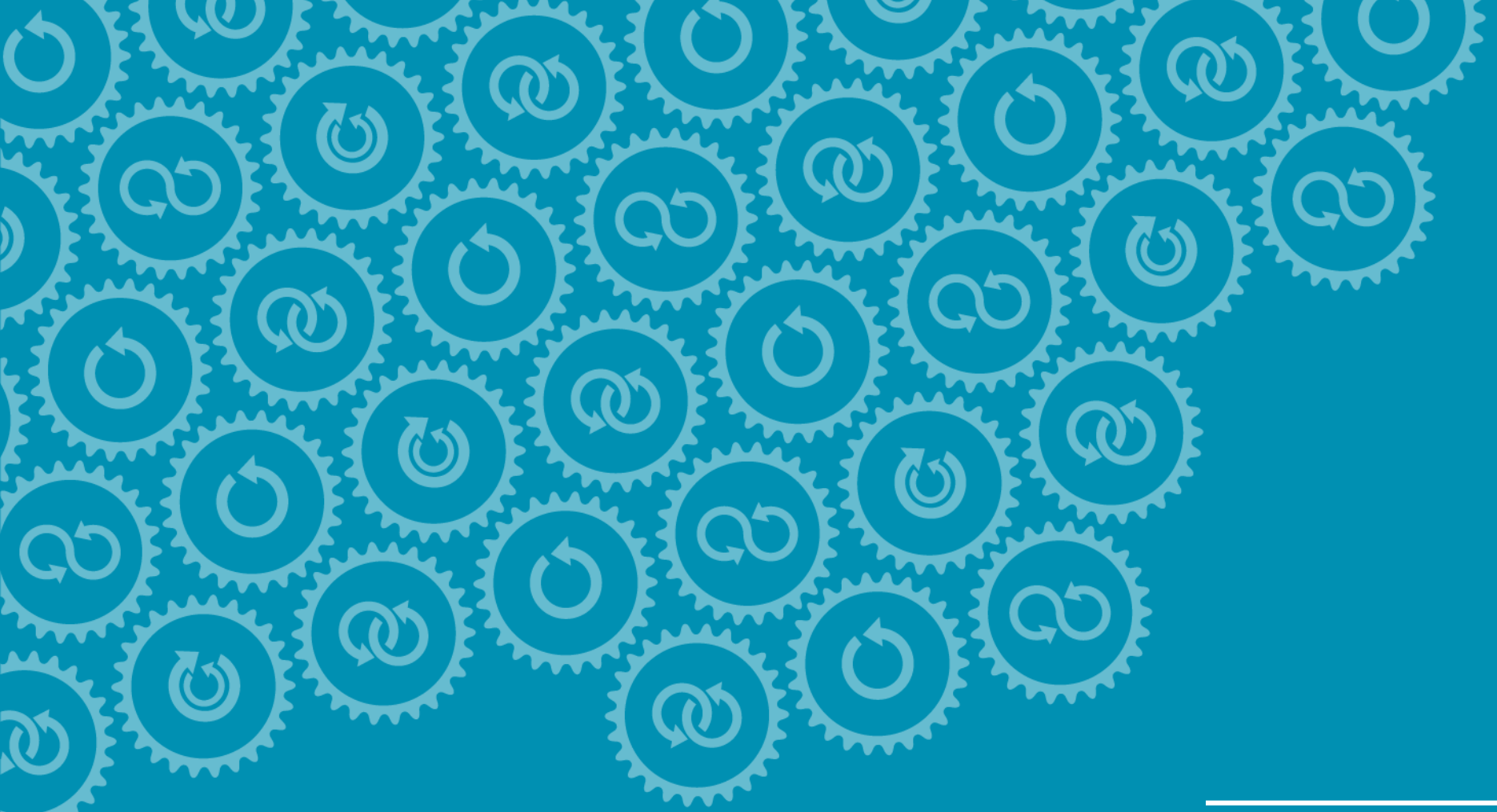
OpenClinica Open Source for Clinical Research

ORACLE

perceptive PARTNER PROGRAM

clinical

PUBLICATION CAPABILITIES



WHAT IS DEFINE-XML?

WHAT IS DEFINE-XML?

Study: TBTC/USPHS Stu[illegible]22 Data D[illegible]initions: CDISC SDTM

DM (Demographics)

DM (Demographics) -- SPECIAL PURPOS[illegible]						dm.xpt
Name	Label	Type	Controlled Terms	Origin	Role	Comment
STUDYID	Study Identifier	text	["TBTC022"]	Assigned	[illegible]entifier	
DOMAIN	Domain Abbreviation	text	["[illegible]"]	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assign[illegible]	Identifier	Protocol ID concatenated with subject ID
SUBJID	Subject Identifier for the Study	text		[illegible]F 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		C[illegible] 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	[illegible]able Qua[illegible]	
SEX	Sex	[illegible]xt	["F", "M"]	CRF Page 5	Record Qualifier	
RACE	Race	text	RACE	CRF Page 5	Record Qualifier	The collected term Hispanic was [illegible]med to be an ethnicity not a race[illegible]s Race for these subjects is coded a[illegible]nknown (Hispanic).
ETHNIC	Ethnicit[illegible]	text	["HISPANIC", "NON-HISPANIC"]	CRF Page 5	Record Qualifier	Ethnicity w[illegible]ssumed to be non-hispanic for al[illegible]llected categories of race n[illegible]identified as Hispanic.

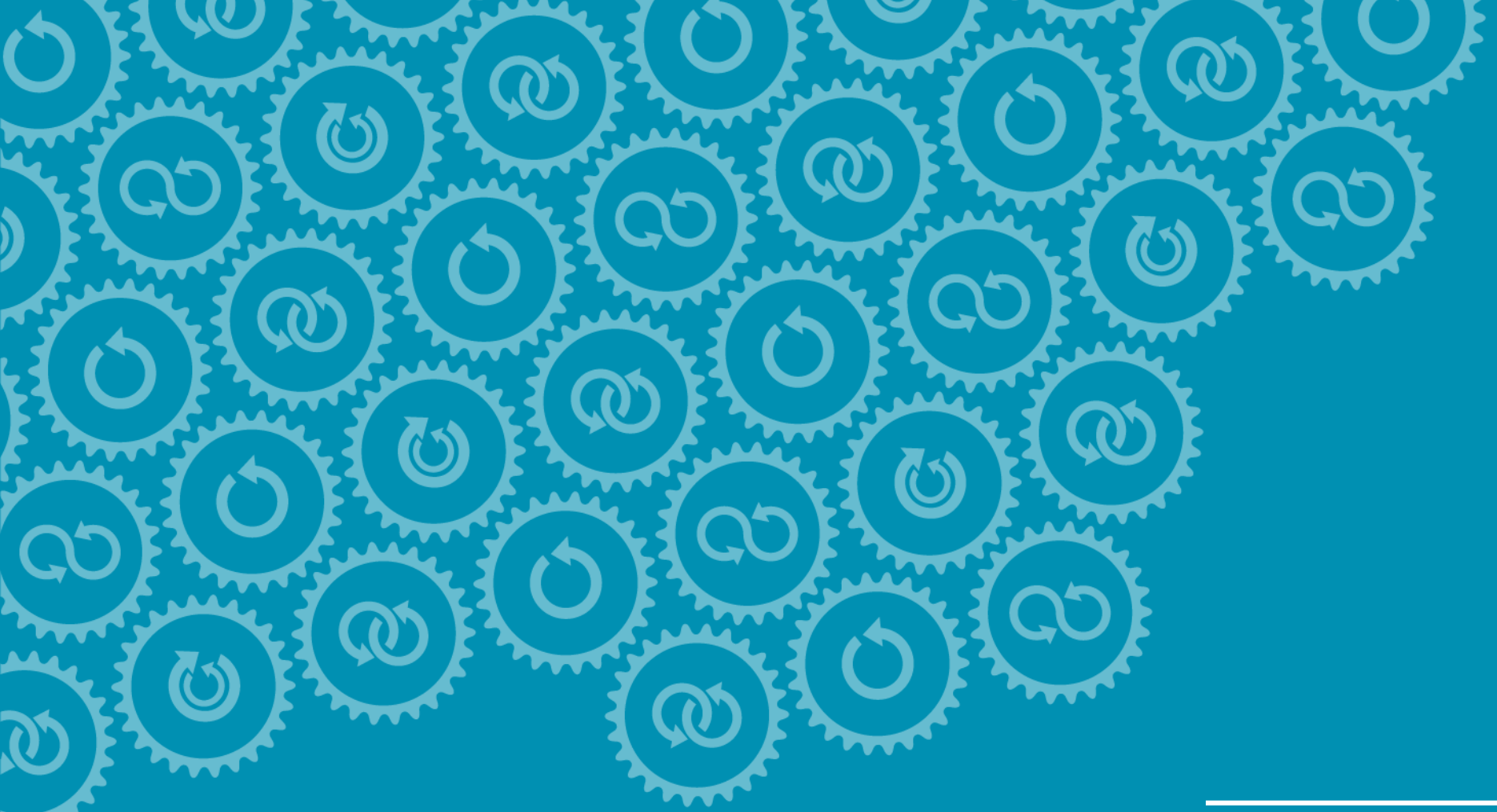
Page 5 of 89

WHAT IS DEFINE-XML?


```
    <ItemRef ItemOID="SUPPDM.QLABEL" Mandatory="Yes" OrderNumber="7" Role="QUALIF
    <ItemRef ItemOID="SUPPDM.QVAL" Mandatory="Yes" OrderNumber="8" Role="QUALIFIEF
    <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>
```

DEFINE-XML ALLOWS DESCRIPTION OF…

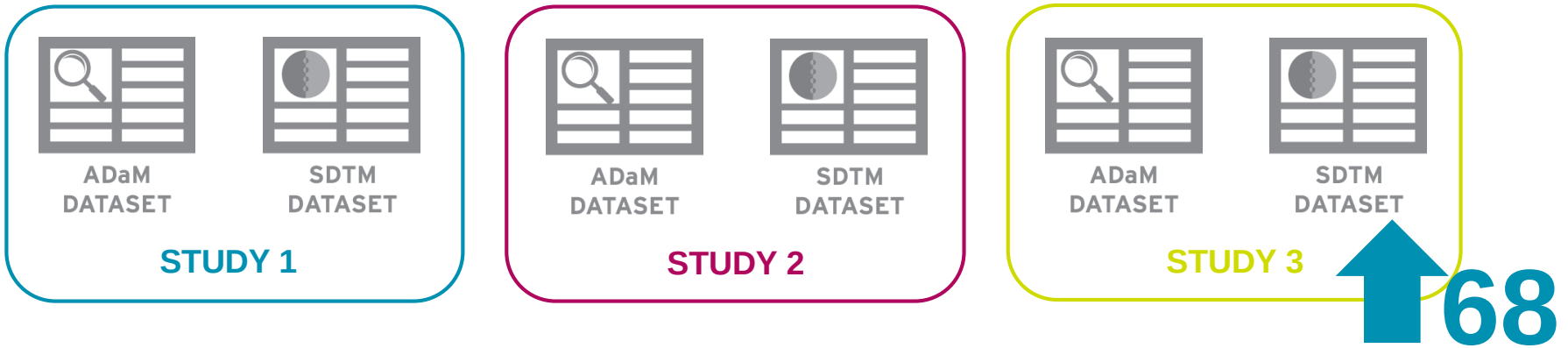
- Study
  - Name, Description, Link to Annotated CRF, Link to Supplemental Docs etc.
- Domains
  - Name, Description, Link to dataset etc.
- Variables
  - Name, Label, Data type, Origin, Role, Comment etc.
- Derivations
- Controlled Terminology
- SDTM Concepts
  - Split Domains, Supplemental Qualifiers,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

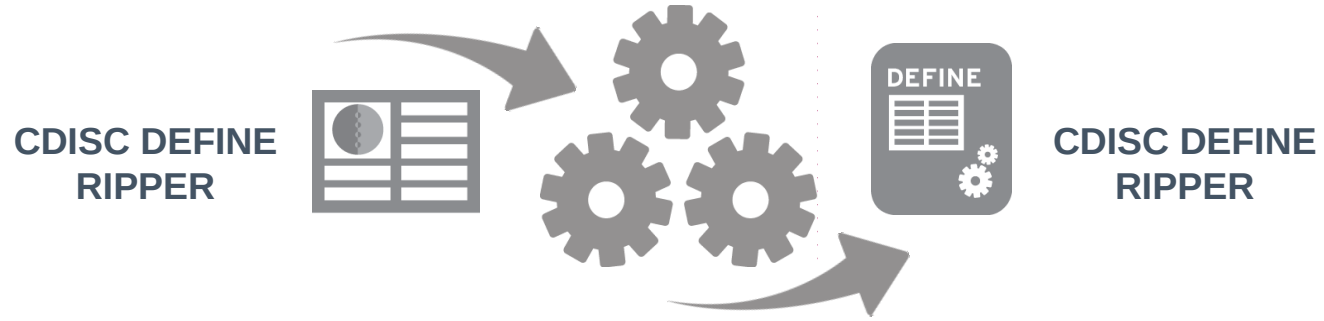


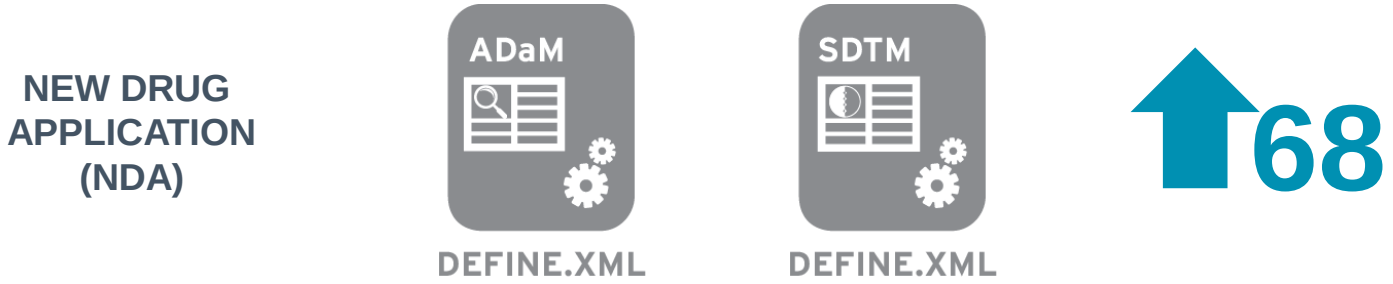
HOW CAN I PRODUCE DEFINE?

POSTHUMOUS DEFINE CREATION

formedix







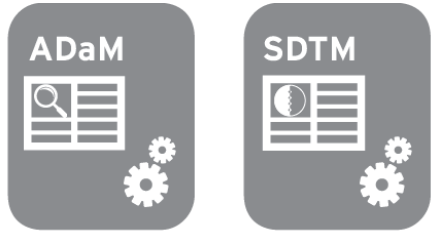
SCALING FOR LEGACY DEFINE.XML CREATION

formedix

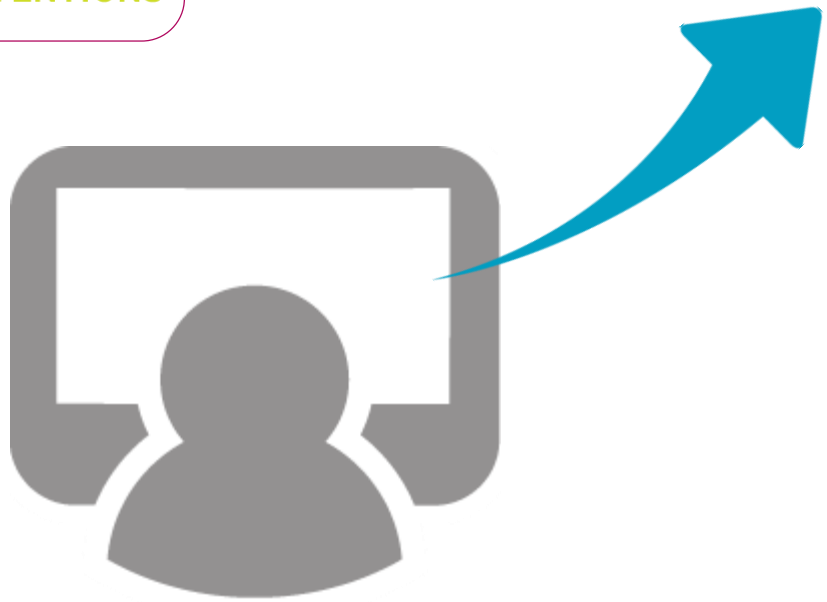
- SDTM Define.xml & ADaM Define.xml submission deliverables
- Commonest scenario
  - Piles of SDTM and ADaM Datasets
  - Define.xml is a “submission afterthought”
- Define.xml must be
  - Structurally and content compliant and match the data provided
- Strategies – specialist define editor, SAS or dataset ripper
- Dataset ripper & smart validator
  - Like CD to MP3
  - Template driven – any conformant dataset to CDISC define.xml

UTOPIA

TEMPLATES



DEFINE.XML DEFINE.XML



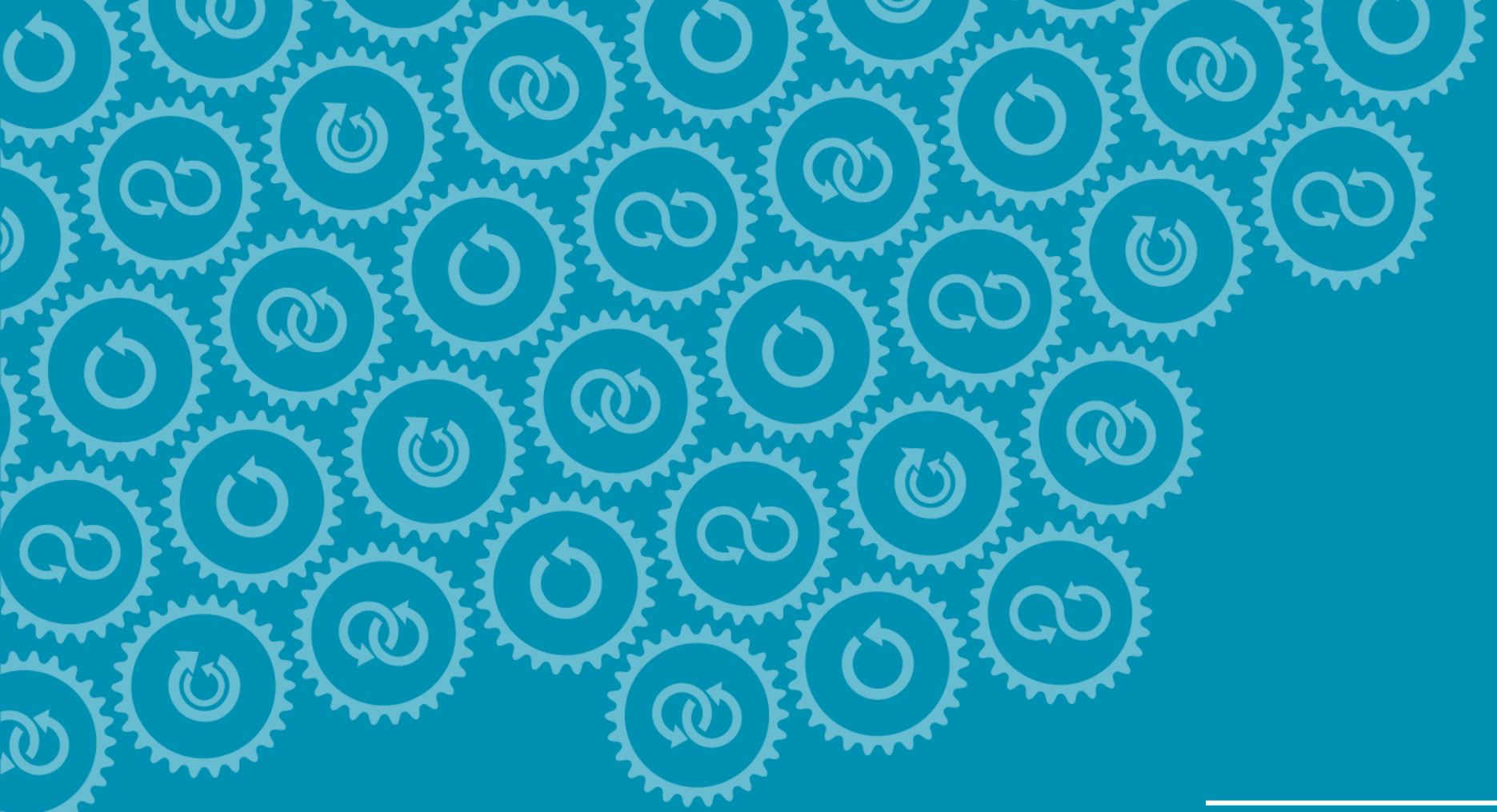
DATASET
DESIGN TOOL

DEFINE CREATION AND APPROVAL PROCESS

formedix

PROCESS STEP	OR 1 SAS.XPT or Excel source Create Define in Authoring tools	2 Approval of Define specification	3 Automated Comparison and Report Creation	4 Validate Define	5 Validate SDTM, ADaM, SEND and CT content
DELIVERABLES	Define dataset specification	Approved Define dataset specification	"Gap" report	Validation report in Excel format	Validation report in Excel format
SOFTWARE USED	Converters Excel to Define SAS to Define Dedicated Dataset Design Tools		Comparison Tools	Validators Define Structure	Validators Content, Controlled Terms, SDTM Version, Your Standards
OPTIMIZATION POSSIBILITIES	Dataset design reuse in Define.xml		Automated comparison between Define.xmls and/or between dataset data and Define.xml	Automated validation against Define rules and templates	Automated validation against Content rules and templates

QC / Validation processes
Define.xml and reviewers guides delivered

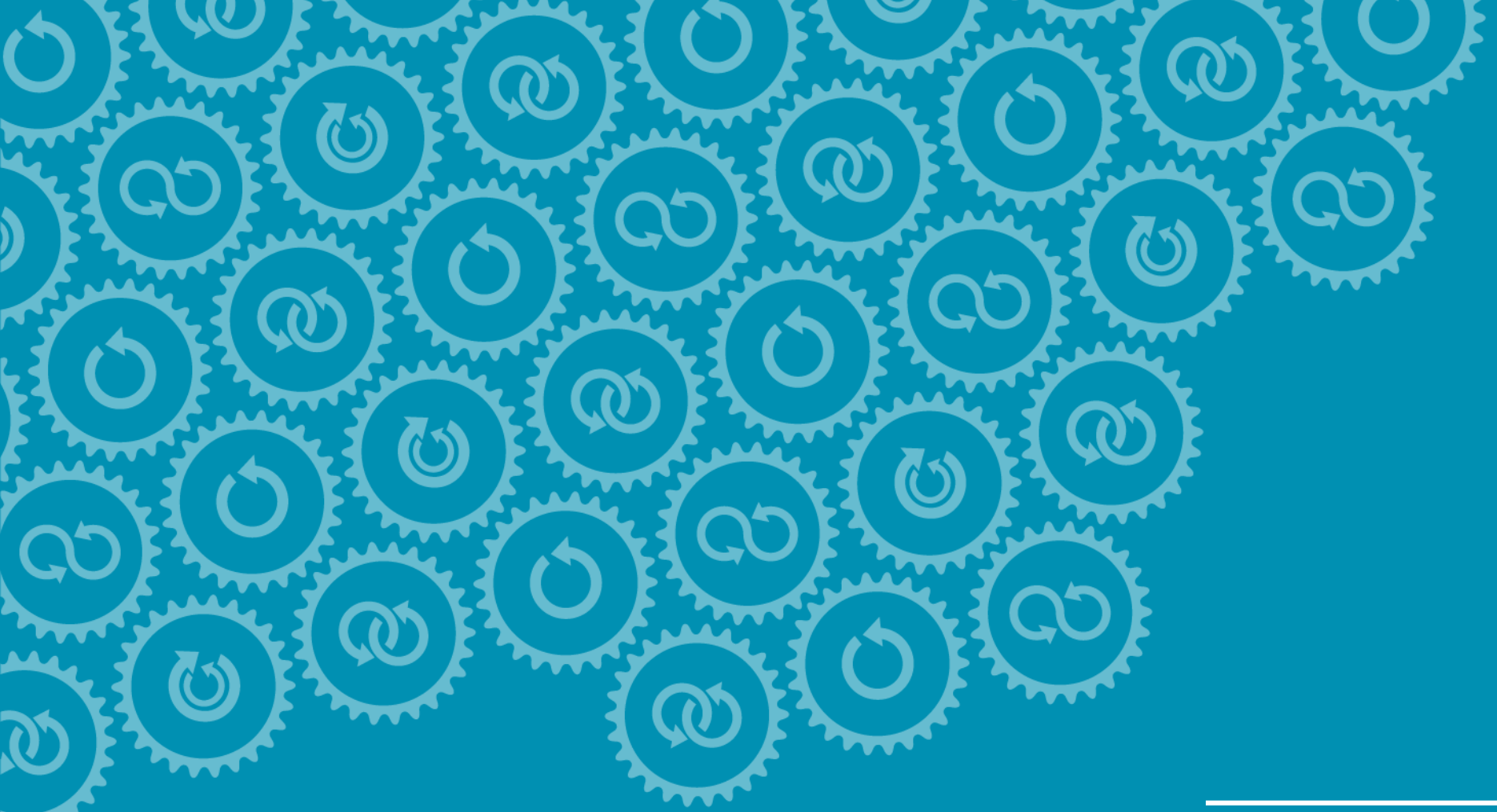


YOUR CLINICAL TRIALS AUTOMATED. EVERYWHERE.

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/data transformations
  - Automated configuration of EDC exporters

- Study Conduct and Analysis
  - Execution of Metadata driven dataset conversions
  - Automated clinical data repository load
  - Automated comparison & validation
    - Define structure conformance
    - Content Conformance with External Standards
    - Compare Study vs. Standards
    - Compare “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?



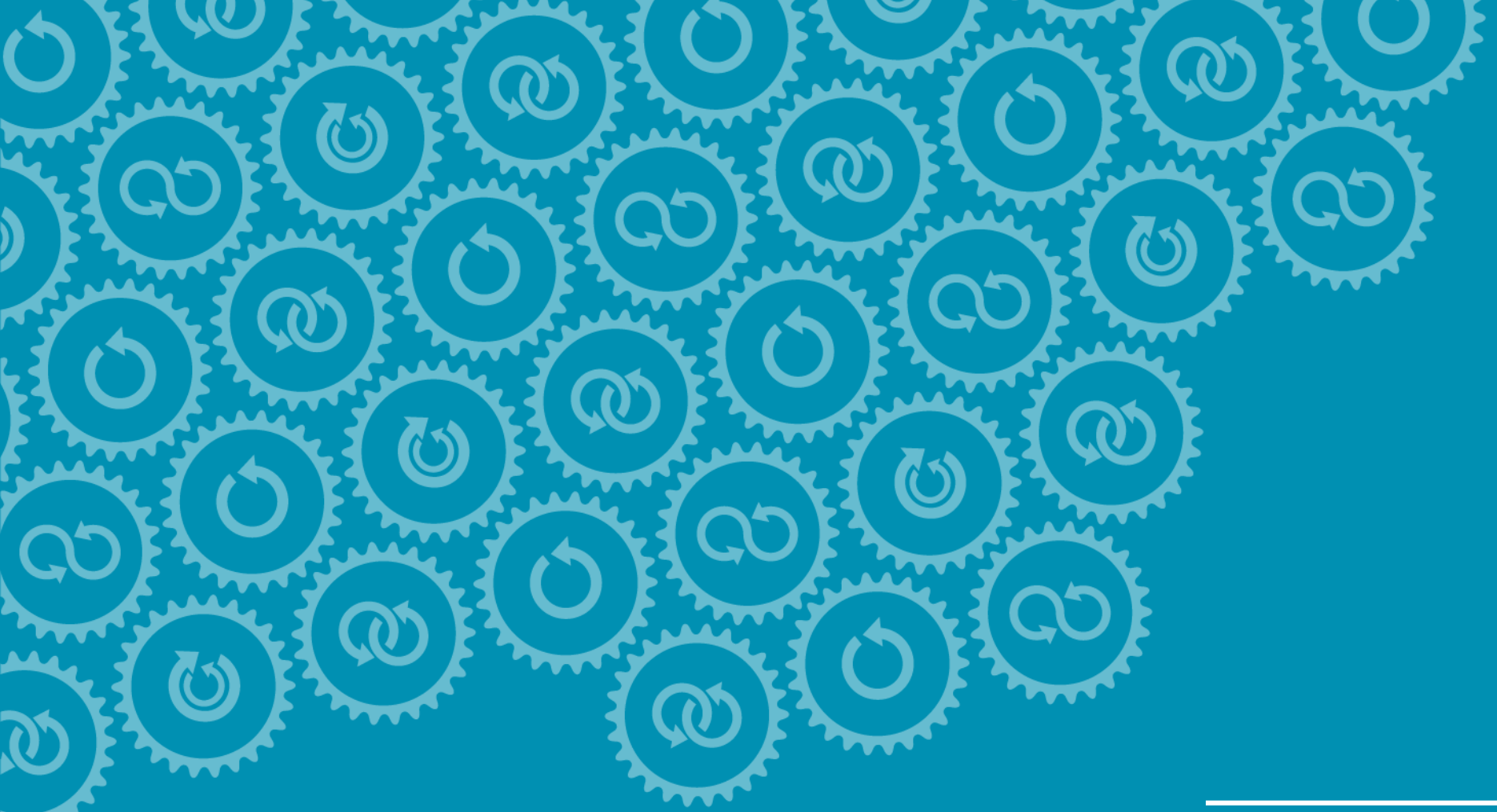
DATASET LIBRARIES STORED WITH DEFINE

WHAT DOES A DEFINE LIBRARY CONTAIN?

- 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

Source: Schering Dublin May 2003



STREAMLINING STUDY SET-UP WITH DEFINE

PROTOCOL TO SUBMISSION
DEFINE AS A FORM DESIGNER

- 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

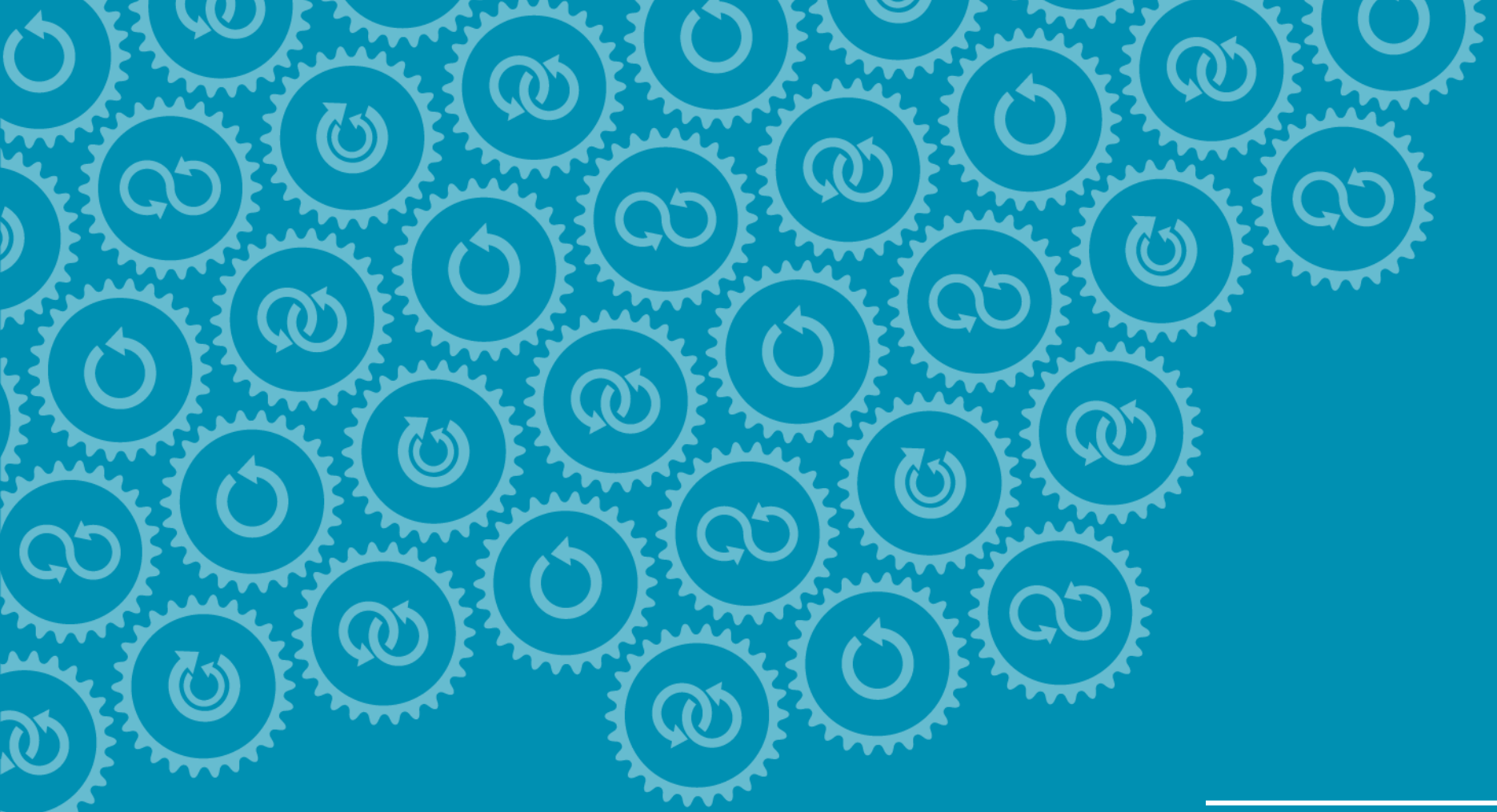
- Human Readable Specification
- Been using this way since 2006
  - With CROs, Partners & Clinical Data Warehouses
- Eliminates ambiguity & variance
  - Datasets, variables, length, orders
  - Value lists & supplemental qualifiers
  - Controlled Terms (NCI and your own)
  - Variance between versions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

MULTIPLE SPECIFICATIONS AUTOGENERATED

FT (Pulmonary Function Tests)

FT (Pulmonary Function Tests) -- Findings						ft.xpt
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	

specify



LEGACY DATASET CONVERSIONS

LEGACY DATASET CONVERSION
MANUAL AND ERROR PRONE

- Source to Target conversion
  - Source could be Lab, EDC, ECG and patient diaries
  - Multiple vendors = multiple output formats/datasets
- Specification issues
  - Hard to visualize source, target and mappings
  - Mappings require specialist knowledge across functional groups
  - Misunderstandings lead to re-programming= $$$s
- Conversion coding
  - Double code in SAS
  - QC check that same end result has been reached
  - In multi-study submissions hope some internal standards were used

MAPPING TYPES

- Domain Mappings
  - Horizontal to Horizontal
  - Horizontal to Vertical
  - Filters
  - Joins
- Variable Mappings
  - Hardcoded strings
  - Choices – If, Then, Else
  - Function – concatenate, date conversions
  - Lookups – search values in other datasets

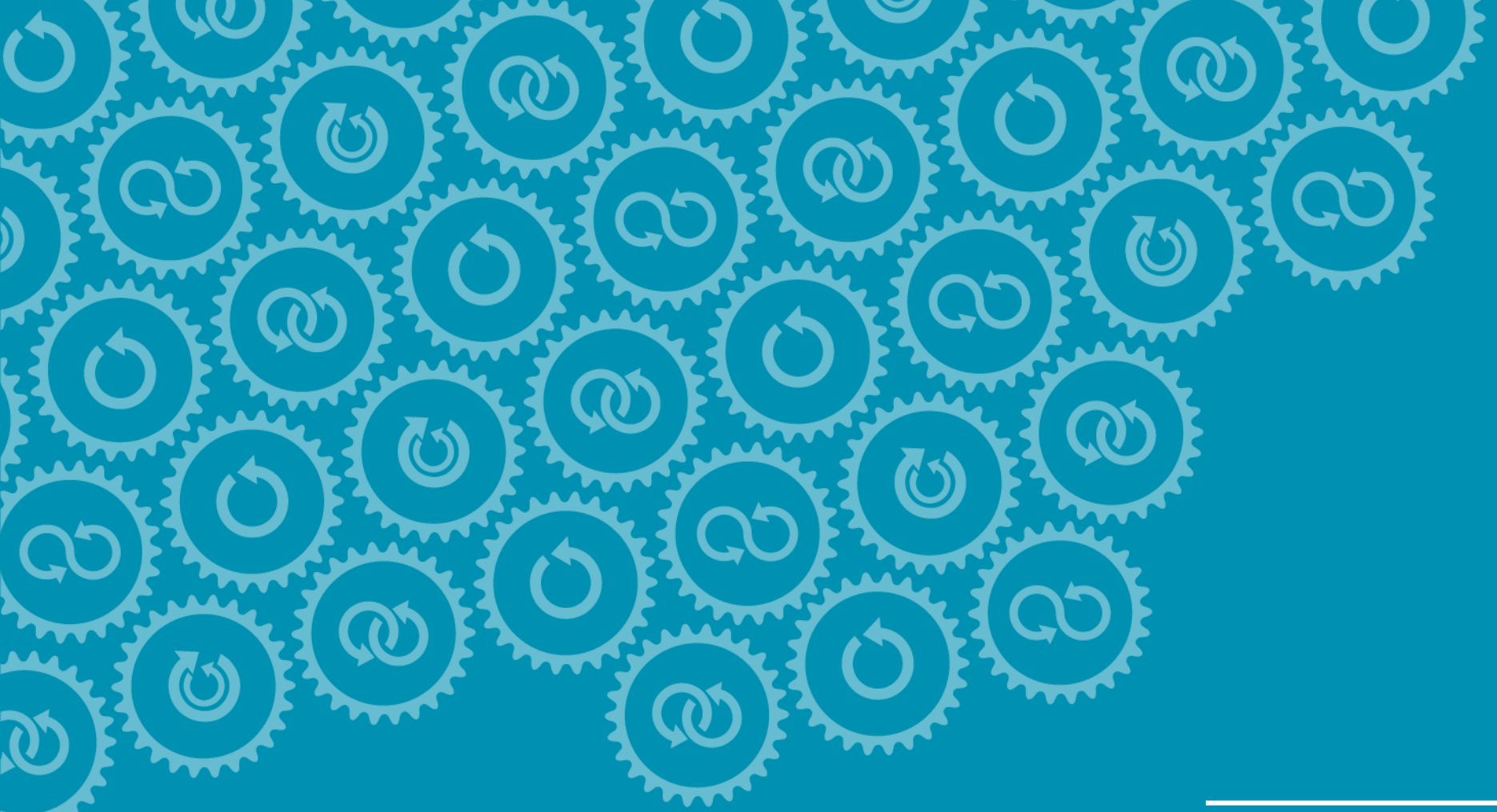
CRF – SDTM DATA MAPPING HORIZONTAL TO VERTICAL

COLLECTION DB

USUBJID	TEMP	TEMP_U	PULSE	PULSE_U
1001	38.0	°C	82	BPM
1002	38.5	°C	79	BPM

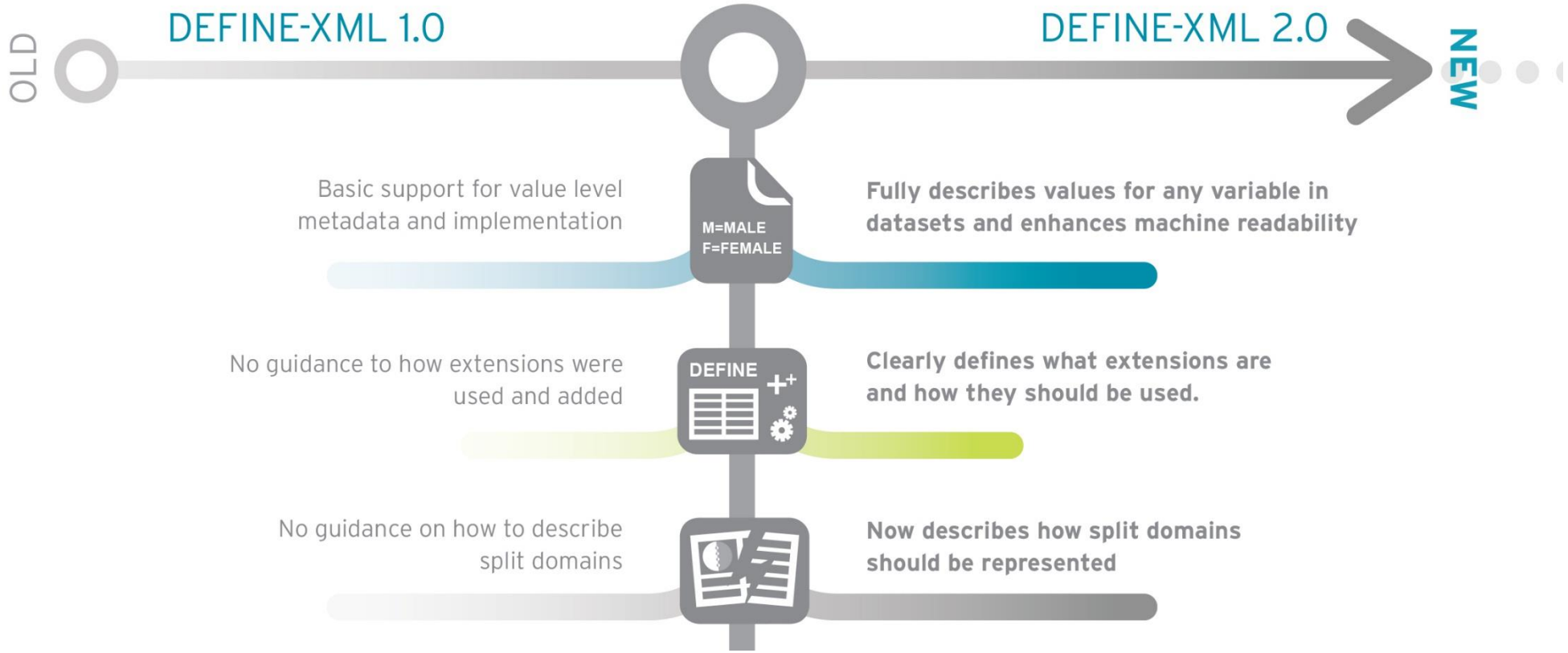
SDTM DATASET

USUBJID	VSTESTCD	VSORRES	VSORRESU
1001	TEMP	38.0	°C
1001	PULSE	82	BPM
1002	TEMP	38.5	°C
1002	PULSE	79	BPM

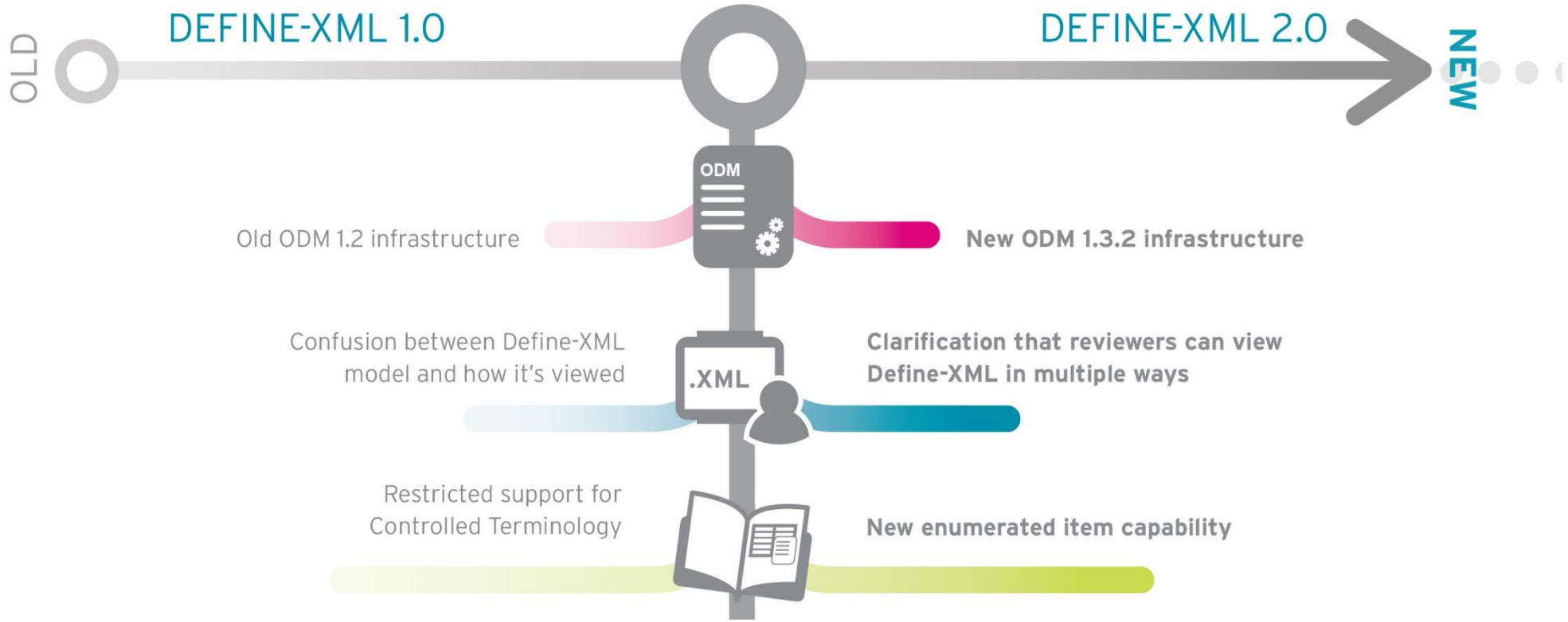


BIGGEST CHANGE IN DEFINE-XML 2.0

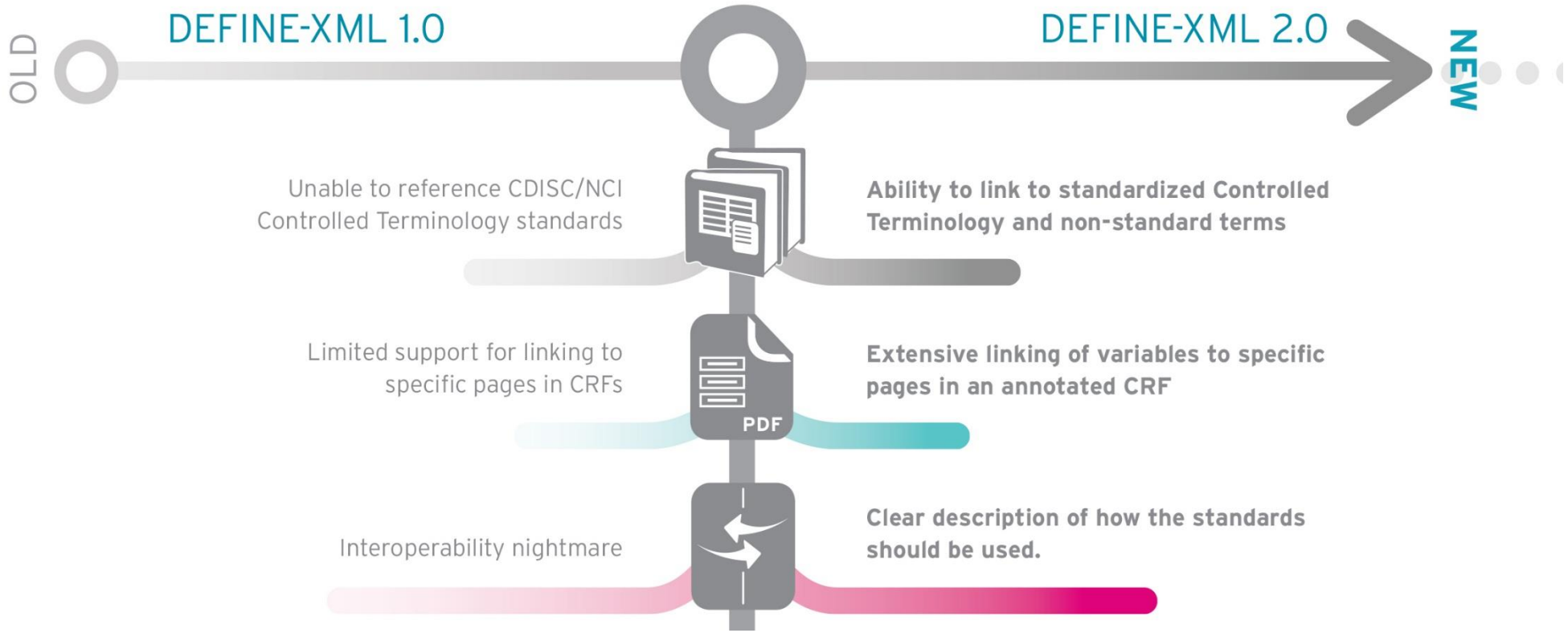
WHAT'S NEW IN DEFINE 2.0



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

formedix

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*

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	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

CONCLUSIONS

- 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?

formedix

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

END-TO-END AUTOMATED

formedix

STUDY SETUP

DATA ACQUISITION DESIGN

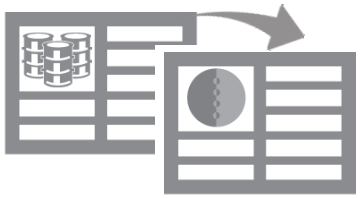
PROPRIETARY DATASET DESIGN

SDTM DATASET DESIGN

STUDY BUILD

MULTI EDC BUILD

STUDY CONDUCT & ANALYSIS



CDISC MODELS

PROPRIETARY DEFINE.XML

SDTM DEFINE.XML

OUTPUTS

SPECS

SPECS

SPECS

PROP DEFINE.HTML
DEFINE.PDF

SDTM DEFINE.HTML
DEFINE.PDF

medidata
Technology Partner

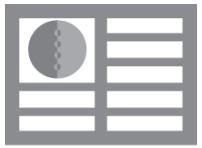
nextrials
our technology your success

OmniComm
eClinical Solutions for Life

OpenClinica
Open Source for Clinical Research

ORACLE

perceptive
PARTNER PROGRAM



SOURCE

TARGET

www.formedix.com