

Moving to Define.xml V2.0

Davy Baele
Project Manager
Business & Decision Life Sciences

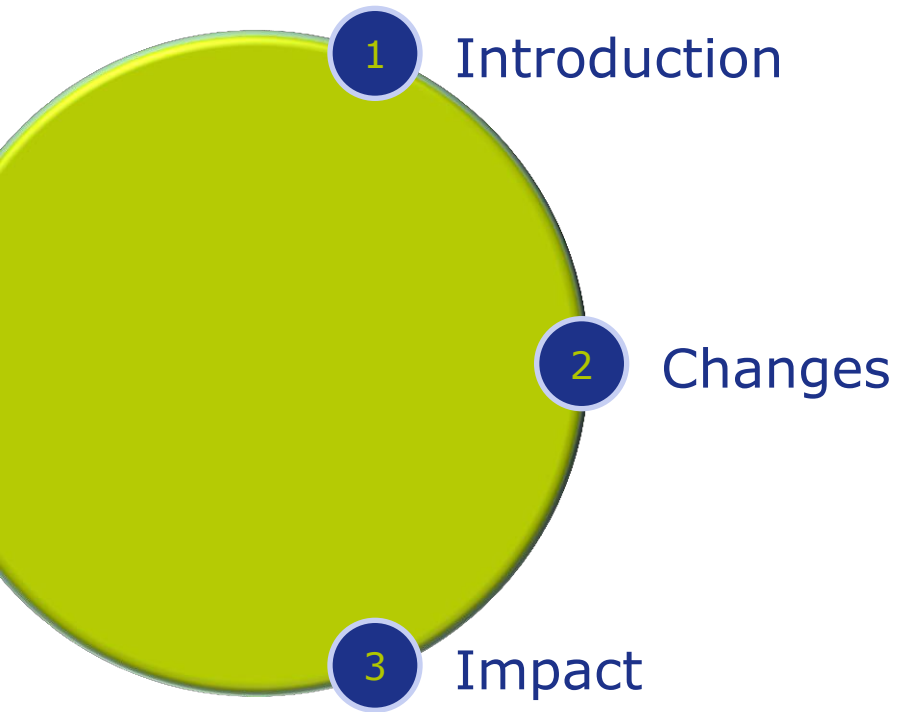
Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 485 65 32 99
davy.baele@businessdecision.com

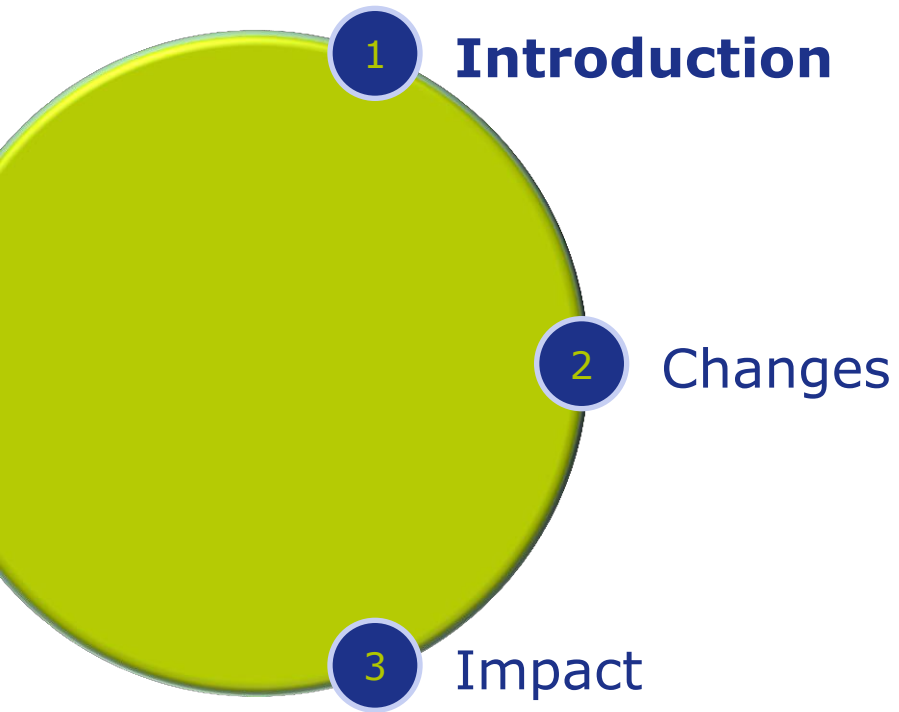
Sint-Lambertusstraat 141 Rue Saint-
Lambert
1200 Brussels

www.businessdecision-lifesciences.com



DRIVE YOUR PERFORMANCE





Introduction

- Feb 2005: CRT-DDS Final Version 1.0 released for implementation (Define.xml V1.0)
- Mar 2013: Define V2.0 posted on CDISC website
 - <http://www.cdisc.org/define-xml>

The screenshot shows the CDISC website header with navigation links: HOME, BLOG, SITE MAP, FAQ, CONTACT, SEARCH, PORTAL, and a Twitter icon. The CDISC logo is prominently displayed with the tagline "Strength Through Collaboration". Below the header is a navigation bar with links: ABOUT CDISC, STANDARDS & INNOVATIONS (highlighted), RESOURCES, NEWS, EDUCATION & EVENTS, MEMBERSHIP, and MEMBERS ONLY. The main content area features the heading "Define-XML" and the subheading "Define-XML 2.0 now available for use". The text below states: "The CDISC XML Technologies team is pleased to announce the release of the Define-XML 2.0 specification. Define-XML 2.0 is a major revision of the Define-XML standard for transmission of metadata for SDTM, SEND and ADaM datasets. The model has been extensively refined in response to implementation experience with Version 1.0, evolution of the SDTM, SEND and ADaM standards and best practices by SDTM and ADaM metadata experts." A section titled "Key changes to the model include:" is partially visible. On the right side of the page, there is a blue box with the text "Become a CDISC Member (Fee Based)" and a link "Become A Member". Below this is a "Member Log-in" section with input fields for "user name" and a password (represented by dots), a "Log In" button, and a link "Forgot my password".

CDISC Strength Through Collaboration

HOME BLOG SITE MAP FAQ CONTACT SEARCH PORTAL

ABOUT CDISC | STANDARDS & INNOVATIONS | RESOURCES | NEWS | EDUCATION & EVENTS | MEMBERSHIP | MEMBERS ONLY

Define-XML

Define-XML 2.0 now available for use

The CDISC XML Technologies team is pleased to announce the release of the Define-XML 2.0 specification. Define-XML 2.0 is a major revision of the Define-XML standard for transmission of metadata for SDTM, SEND and ADaM datasets. The model has been extensively refined in response to implementation experience with Version 1.0, evolution of the SDTM, SEND and ADaM standards and best practices by SDTM and ADaM metadata experts.

Key changes to the model include:

Become a CDISC Member (Fee Based)
[Become A Member](#)

Member Log-in

user name

.....

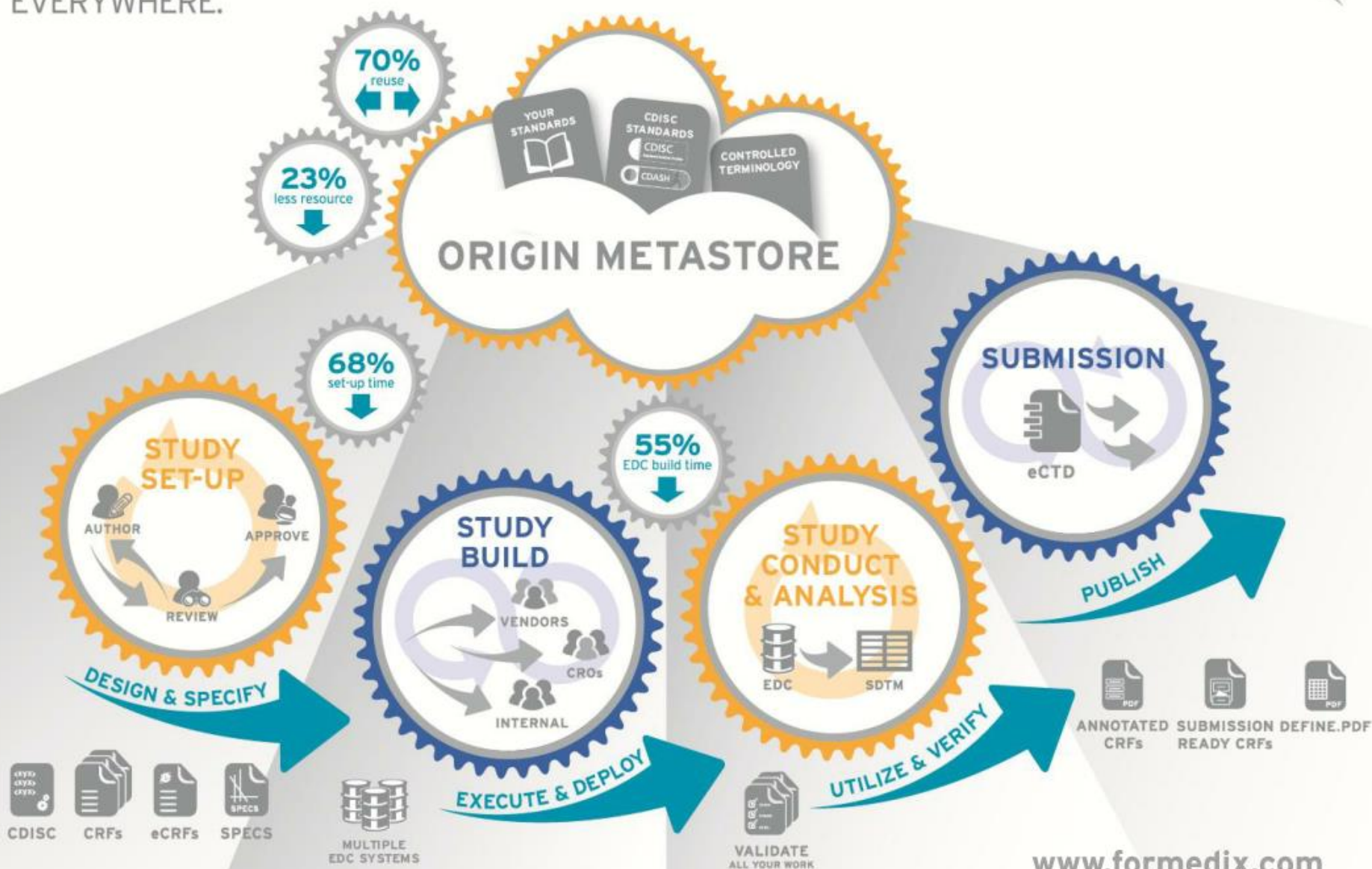
Log In

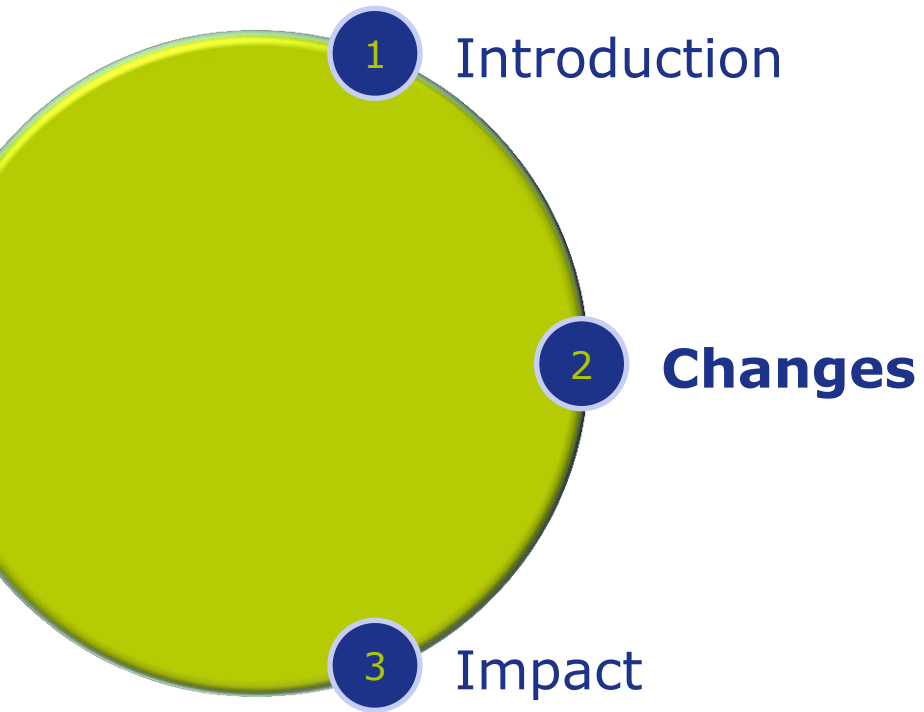
[Forgot my password](#)

Define.xml

- Accompanies FDA submissions
- Contains structured metadata on different levels
 - Dataset
 - Variable
 - Value Level Metadata
 - Controlled Terminology
 - Computational Algorithms
- Has much more potential than submissions only

YOUR CLINICAL
TRIALS AUTOMATED.
EVERYWHERE.





Changes - Overview

- Major changes
 - Documentation
 - SDTM and ADaM are covered with examples
 - XML Namespace
 - Split Datasets
 - Value Level Metadata
- Minor changes

Documentation - SDTM Example

- Example of SDTM define.xml and associated stylesheet

Date of document generation: 2013-03-03T17:04:44

Stylesheet version: 2013-03-04

Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt	
TE	Trial Elements	TRIAL DESIGN	One record per planned Element	Tabulation	STUDYID, ETCDCD	te.xpt	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt	
TS	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt	
TV	Trial Visits	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xpt	
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide
SE	Subject Elements	SPECIAL PURPOSE	One record per actual Element per subject	Tabulation	STUDYID, USUBJID, SESTDTC, SEENDTC, TAETORD, ETCDCD	se.xpt	

Documentation - ADaM Example

- Example of ADaM define.xml and associated stylesheet

Analysis Datasets for Study CDISC-Sample (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject-Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	one record per subject	Analysis	USUBJID	adsl.xpt	Screen Failures are excluded since they are not needed for this study analysis
ADQSADAS	ADAS-Cog Analysis	BASIC DATA STRUCTURE	One record per subject per parameter per analysis visit per analysis date	Analysis	USUBJID, PARAMCD, AVISIT, ADT	adqsadas.xpt	See referenced dataset creation program and Analysis Data Reviewer's Guide, Section 2.1 adqsadas.sas Analysis Data Reviewer's Guide

Go to the [top](#) of the define.xml

Subject-Level Analysis (ADSL) [Location: [adsl.xpt](#)]

Variable	Label	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	text	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier	text	11		Predecessor: DM.USUBJID
SUBJID	Subject Identifier for the Study	text	4		Predecessor: DM.SUBJID
SITEID	Study Site Identifier	text	3		Predecessor: DM.SITEID

Documentation - XML namespace

- Any XML included in a Define-XML document not described in the Define.xml V2.0 specification is considered an extension
- Vendor Extensions
 - Have no meaning with regards to the Define.xml standard
 - Can be ignored unless explicitly agreed upon their use
 - Vendors should be able to produce valid Define.xml files without extensions
- FDA submissions: Follow the standards

```
<!-- Dataset Definition (TA) -->
<ItemGroupDef OID="IG.TA"
  Domain="TA"
  Name="TA"
  Repeating="No"
  IsReferenceData="Yes"
  SASDatasetName="TA"
  Purpose="Tabulation"
  def:Structure="One record per planned Element per Arm"
  def:Class="TRIAL DESIGN"
  def:ArchiveLocationID="LF.TA">
  <Description>
    <TranslatedText xml:lang="en">Trial Arms</TranslatedText>
```

Documentation - Split Datasets

- Guidance on how to describe metadata for split domains
 - ItemGroupDef
 - Name = Split dataset name = « QSCG »
 - Domain = Parent domain = « QS »

QSCG	Questionnaire-QSCG (Questionnaires)	FINDINGS	One record per questionnaire per question per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSTESTCD, QSDTC, VISITNUM	qscg.xpt	QS is submitted as a split dataset. The split was done based on QSCAT as QSCG (CLINICAL GLOBAL IMPRESSIONS), QSCS (CORNELL SCALE FOR DEPRESSION INDEMENTIA) and QSMM (MINI MENTAL STATE EXAMINATION). See additional documentation in the Reviewer's Guide, Split Datasets Section. Reviewers Guide
QSCS	Questionnaire-QSCS (Questionnaires)	FINDINGS	One record per questionnaire per question per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSTESTCD, QSDTC, VISITNUM	qscs.xpt	QS is submitted as a split dataset. The split was done based on QSCAT as QSCG (CLINICAL GLOBAL IMPRESSIONS), QSCS (CORNELL SCALE FOR DEPRESSION INDEMENTIA) and QSMM (MINI MENTAL STATE EXAMINATION). See additional documentation in the Reviewer's Guide, Split Datasets Section. Reviewers Guide

Value Level Metadata – Define.xml V1.0

Value Level Metadata (ValueList.QS.QSCAT.CG.QSTESTCD)								
	Source Variable	Value	Label	Type	Controlled Terms or Format	Origin	Role	Comment
Value	CGI-I	CGIGLOB	Global Improvement	integer	CGIIMP X	CRF Page 17		
Question	Source Variable	Value	Label	Type	Controlled Terms or Format	Origin	Role	Comment
Variable	QSCAT	CGI-I	Clinical Global Impressions	text		QSM- CRF Page 13 , QSCS- CRF Page 14 , 15 , QSCG- CRF Page 17		
STUD	QSCAT	CSDD	Cornell Scale for Depression in Dementia	text		QSM- CRF Page 13 , QSCS- CRF Page 14 , 15 , QSCG- CRF Page 17		
DOMA	QSCAT	MMSE	Mini Mental State Examination	text		QSM- CRF Page 13 , QSCS- CRF Page 14 , 15 , QSCG- CRF Page 17		
USUB	QSCAT	MMSE	Mini Mental State Examination	text		QSM- CRF Page 13 , QSCS- CRF Page 14 , 15 , QSCG- CRF Page 17		
QSTEST	Question Name	text						
QSTEST	Question Name	text						
QSCAT	Category of Question	text						
QSORRES	Finding in Original Units	text						

Value Level Metadata – Define.xml V1.0

- Define.xml V1.0 provided limited accomodation for representing VLM
- Nesting was introduced to allow for additional structuring
- No electronic link between the VLM and the actual variable being described in it

Value Level Metadata – Define.xml V2.0

Vital Signs (VS) [Location: [vs.xpt](#)]

Value Level Metadata - VS [VSORRESU]							
Variable	Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	VSORRESU	VSTESTCD EQ HEIGHT (Height) AND COUNTRY IN ("CAN" , "MEX")	text	5	["cm" = "Centimeter"] < Unit (UH_MC) >	CRF Page 11	Join any Subject Level dataset with the Demographics dataset based on [IG.datasetname]IT.USUBJID = [IG.DM]IT.USUBJID, assuming 'IG.datasetname' is the OID of the ItemGroupDef that defines the subject-level dataset to be joined with the Demographics dataset. The data submitted only includes subjects in the USA since other sites did not enroll any subjects.
DOMAIN	VSORRESU	VSTESTCD EQ HEIGHT (Height) AND COUNTRY EQ USA	text	5	["IN" = "Inch"] < Unit (UH_NMC) >	CRF Page 11	Join any Subject Level dataset with the Demographics dataset based on [IG.datasetname]IT.USUBJID = [IG.DM]IT.USUBJID, assuming 'IG.datasetname' is the OID of the ItemGroupDef that defines the subject-level dataset to be joined with the Demographics dataset.
USL	VSORRESU	VSTESTCD EQ WEIGHT (Weight) AND COUNTRY IN ("CAN" , "MEX")	text	4	["kg" = "Kilogram"] < Unit (UW_MC) >	CRF Page 11	Join any Subject Level dataset with the Demographics dataset based on [IG.datasetname]IT.USUBJID = [IG.DM]IT.USUBJID, assuming 'IG.datasetname' is the OID of the ItemGroupDef that defines the subject-level dataset to be joined with the Demographics dataset. The data submitted only includes subjects in the USA since other sites did not enroll any subjects.
VSS	VSORRESU	VSTESTCD EQ WEIGHT (Weight) AND COUNTRY EQ USA	text	4	["LB" = "Pound"] < Unit (UW_NMC) >	CRF Page 11	Join any Subject Level dataset with the Demographics dataset based on [IG.datasetname]IT.USUBJID = [IG.DM]IT.USUBJID, assuming 'IG.datasetname' is the OID of the ItemGroupDef that defines the subject-level dataset to be joined with the Demographics dataset.
VST	VSORRESU	VSTESTCD EQ WEIGHT (Weight) AND COUNTRY EQ USA	text	4	["LB" = "Pound"] < Unit (UW_NMC) >	CRF Page 11	Join any Subject Level dataset with the Demographics dataset based on [IG.datasetname]IT.USUBJID = [IG.DM]IT.USUBJID, assuming 'IG.datasetname' is the OID of the ItemGroupDef that defines the subject-level dataset to be joined with the Demographics dataset.
VSP	VSORRESU	VSTESTCD EQ WEIGHT (Weight) AND COUNTRY EQ USA	text	4	["LB" = "Pound"] < Unit (UW_NMC) >	CRF Page 11	Join any Subject Level dataset with the Demographics dataset based on [IG.datasetname]IT.USUBJID = [IG.DM]IT.USUBJID, assuming 'IG.datasetname' is the OID of the ItemGroupDef that defines the subject-level dataset to be joined with the Demographics dataset.

Value Level Metadata – Define.xml V2.0

- Representation of VLM has changed
- VLM for a variable is now explicitly defined on the particular variable it applies to.
 - E.g., VLM on VSORRES to describe VSORRES
- VLM can unambiguously be defined on any variable

Value Level Metadata - VS [VSORRES]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
VSORRES	VSTESTCD EQ DIABP (Diastolic Blood Pressure)	integer	2		CRF Page 11	
VSORRES	VSTESTCD EQ FRMSIZE (Body Frame Size)	text	6	["LARGE", "MEDIUM", "SMALL"] <Size>	CRF Page 11	
VSORRES	VSTESTCD EQ HEIGHT (Height)	float	5.1		CRF Page 11	
VSORRES	VSTESTCD EQ PULSE (Pulse Rate)	integer	2		CRF Page 11	
VSORRES	VSTESTCD EQ SYSBP (Systolic Blood Pressure)	integer	3		CRF Page 11	
VSORRES	VSTESTCD EQ WEIGHT (Weight)	float	5.1		CRF Page 11	

Value Level Metadata - Visualisation

- The use of « Where clauses » allows for rendering VLM as
 - ValueLists: Listing of distinct values of one variable
 - Slices: View of a dataset for a specific condition (defined by Where-clause)

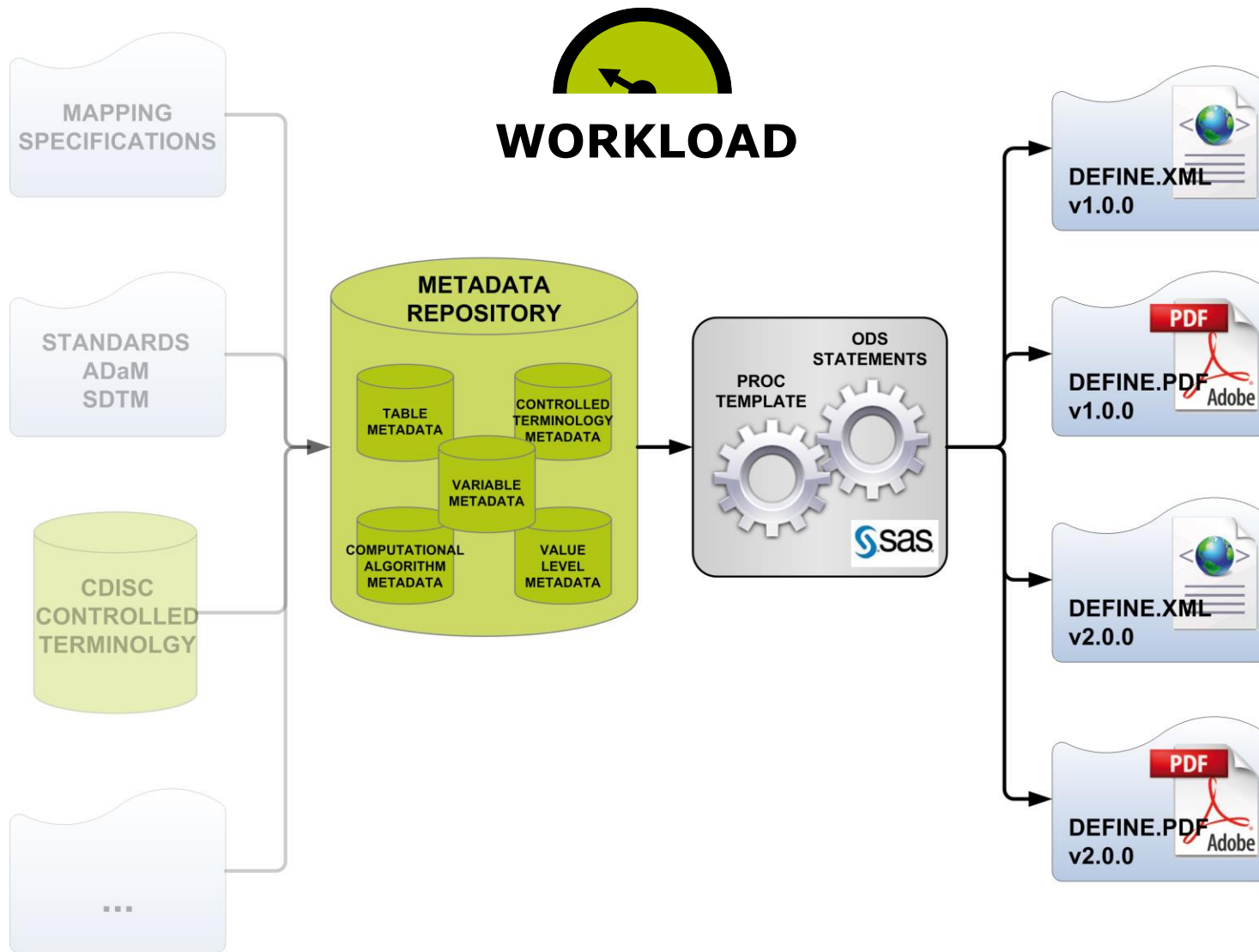
TEMP Slice: VS Domain Where VSTESTCD EQ 'TEMP'

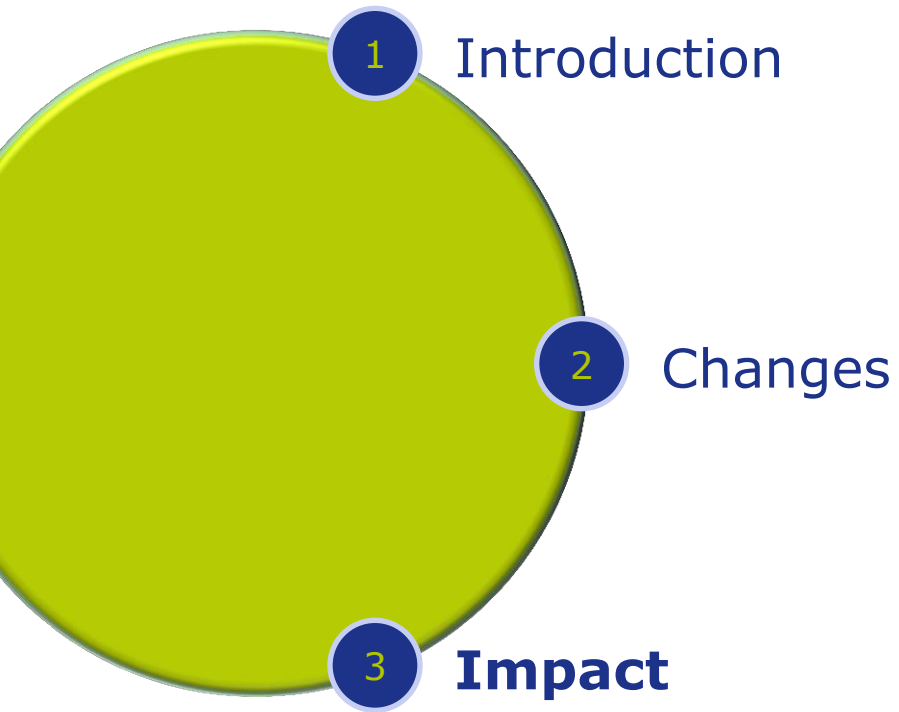
Name	Label	Type	Controlled Terms
STUDYID	Study Identifier	text	
DOMAIN	Domain Abbreviation	text	['VS']
USUBJID	Unique Subject Identifier	text	
VSSEQ	Sequence Number	float	
VSTESTCD	Vital Signs Test Short Name	text	Vital Signs Test Code (C66741)
VSTEST	Vital Signs Test Name	text	Vital Signs Test Name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

Minor Changes

- Deprecated components
 - Renamed to match with existing ODM components
 - Computational Algorithms
 - Labels
 - ...
- Comments
 - Comments can be added on dataset level
 - Comments are referenced
 - Comments can be re-used
- Origins
 - ADaM predecessors
 - Range of pages

Example: Generating a Define.xml





Impact

- Define.xml V2.0 is the new improved standard for submitting metadata
- Define.xml V2.0 is well documented
- Upgrade from Define.xml V1.0 to Define.xml V2.0 can be tricky



Thank you for your attention

Davy Baele
Project Manager
Business & Decision Life Sciences

Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 485 65 32 99
davy.baele@businessdecision.com

Sint-Lambertusstraat 141 Rue Saint-
Lambert
1200 Brussels

www.businessdecision-lifesciences.com



DRIVE YOUR PERFORMANCE