Business & Decision Life Sciences

Moving to Define.xml v2.0.0 for CDISC FSUG

Anne-Sophie Bekx / 22 May 2014

Business & Decision Life Sciences

Business & Decision Life Sciences

Introduction – History of Define.xml

- February 2005: CRT-DDS Final Version 1.0 released for implementation (**Define.xml v1.0**)
- March 2013**: Define.xml v2.0** posted on CDISC website
  - http://www.cdisc.org/define-xml

Define.xml

- Accompanies FDA submissions

- Contains structured metadata on different levels
  - Dataset
  - Variable
  - Value Level Metadata
  - Controlled Terminology
  - Computational Algorithms

Business & Decision Life Sciences

Dataset	Description	Class	Structure	Purpose	Keys	Location
CM	Concomitant Medications	Interventions	One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC	..\Tabulations\cm.xpt
EX	Exposure	Interventions	One record per constant dosing	Tabulation	STUDYID, USUBJID,	..\Tabulations\ex.xpt
AE	Adverse Events					

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Protocol, CRF	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
DOMAIN	Domain Abbr					
USUBJID	Unique Subjec					

Value Level Metadata (ValueList.VS.VSTESTCD)

Source Variable	Value	Label	Type	Controlled Terminology	Origin	Comment
VSTESTCD	BMI	BODY MASS INDEX	integer		Derived	See Computational Method: VS.BMI
VSTESTCD	DIABP	DIASTOLIC BLOOD PRESSURE	integer		CRF	
VSTESTCD	HEIGHT	HEIGHT	integer		CRF	
VSTESTCD	PULSE					
VSTESTCD	SYSBP					
VSTESTCD	WEIGHT					

POSITION, Reference Name (Codelist.POSITION)

Code Value	Code Text
STANDING	STANDING
SUPINE	SUPINE

SEX, Reference Name (Codelist.SEX)

Code Value	Code Text
F	Female
M	
U	

Computational Algorithms (VS.BMI)

Reference Name	Computation Method
VS.BMI	Equals to WEIGHT DIVIDED BY HEIGHT SQUARED

Computational Algorithms (VS.VSDY)

Reference Name	Computation Method
VS.VSDY	Equals to (VS.VSDTC - DM.RFSTDTC) + 1 IF VSDTC is on or after RFSTDTC Equals to (VS.VSDTC - DM.RFSTDTC) IF VSDTC precedes RFSTDTC

Business & Decision Life Sciences

Define.xml

- **Define.xml** has greater potential than to be used for submissions only

Business & Decision Life Sciences

Business & Decision Life Sciences

Changes - Overview

- **Major changes**
  - Documentation
    - SDTM and ADaM are covered with examples
  - The way in which we handle Value Level Metadata
    - Value Level Metatdata can be attached to any variable
  - XML Namespace
    - Now aligned with ODM 1.3.1
  - Comments
    - can be added on dataset level
    - are referenced
    - can be re-used
  - Better support for split domains
- **Minor changes**
  - Deprecated Components
  - Origins

Business & Decision Life Sciences

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, ETCD	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, ETCD	se.xpt	

Business & Decision Life Sciences

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

Business & Decision Life Sciences

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>
```

Business & Decision Life Sciences

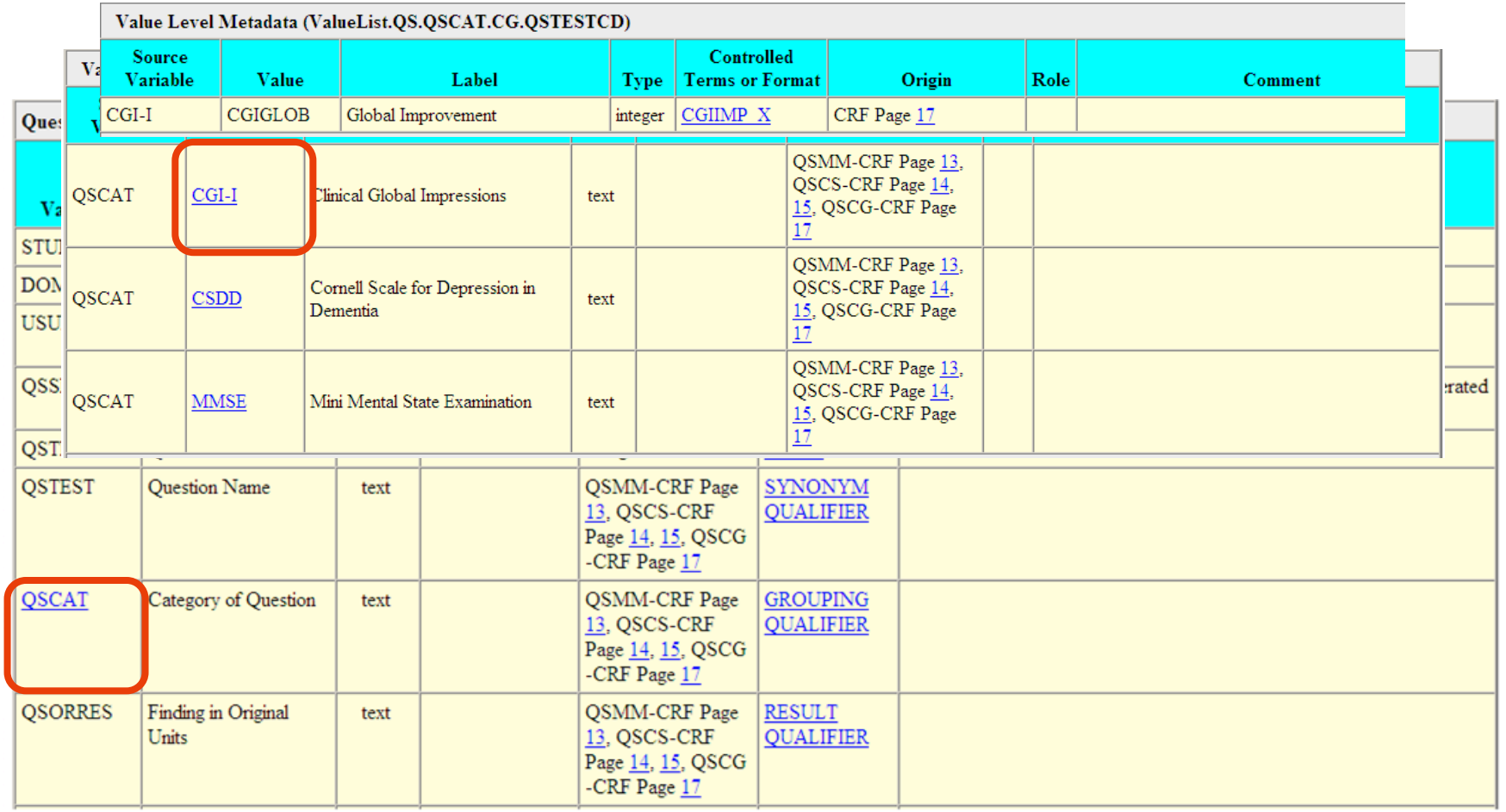
Split Datasets

- 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

Business & Decision Life Sciences

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
CGI-I	CGIGLOB	Global Improvement	integer	CGIIMP_X	CRF Page 17		
QSCAT	CGI-I	Clinical Global Impressions	text		QSMM-CRF Page 13, QSCS-CRF Page 14, 15, QSCG-CRF Page 17		
QSCAT	CSDD	Cornell Scale for Depression in Dementia	text		QSMM-CRF Page 13, QSCS-CRF Page 14, 15, QSCG-CRF Page 17		
QSCAT	MMSE	Mini Mental State Examination	text		QSMM-CRF Page 13, QSCS-CRF Page 14, 15, QSCG-CRF Page 17		

QSTEST	Question Name	text		QSMM-CRF Page 13, QSCS-CRF Page 14, 15, QSCG-CRF Page 17	SYNONYM QUALIFIER	
QSCAT	Category of Question	text		QSMM-CRF Page 13, QSCS-CRF Page 14, 15, QSCG-CRF Page 17	GROUPING QUALIFIER	
QSORRES	Finding in Original Units	text		QSMM-CRF Page 13, QSCS-CRF Page 14, 15, QSCG-CRF Page 17	RESULT QUALIFIER	

Business & Decision Life Sciences

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

Business & Decision Life Sciences

Value Level Metadata – Define.xml V2.0

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

Variable
STUDYID
DOMAIN
USU
VSS
VST
VST
VSP
VSORRES
VSORRESU

Value Level Metadata - VS [VSORRESU]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
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.
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.
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.
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.

Business & Decision Life Sciences

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	

Business & Decision Life Sciences

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)

Business & Decision Life Sciences

Comments

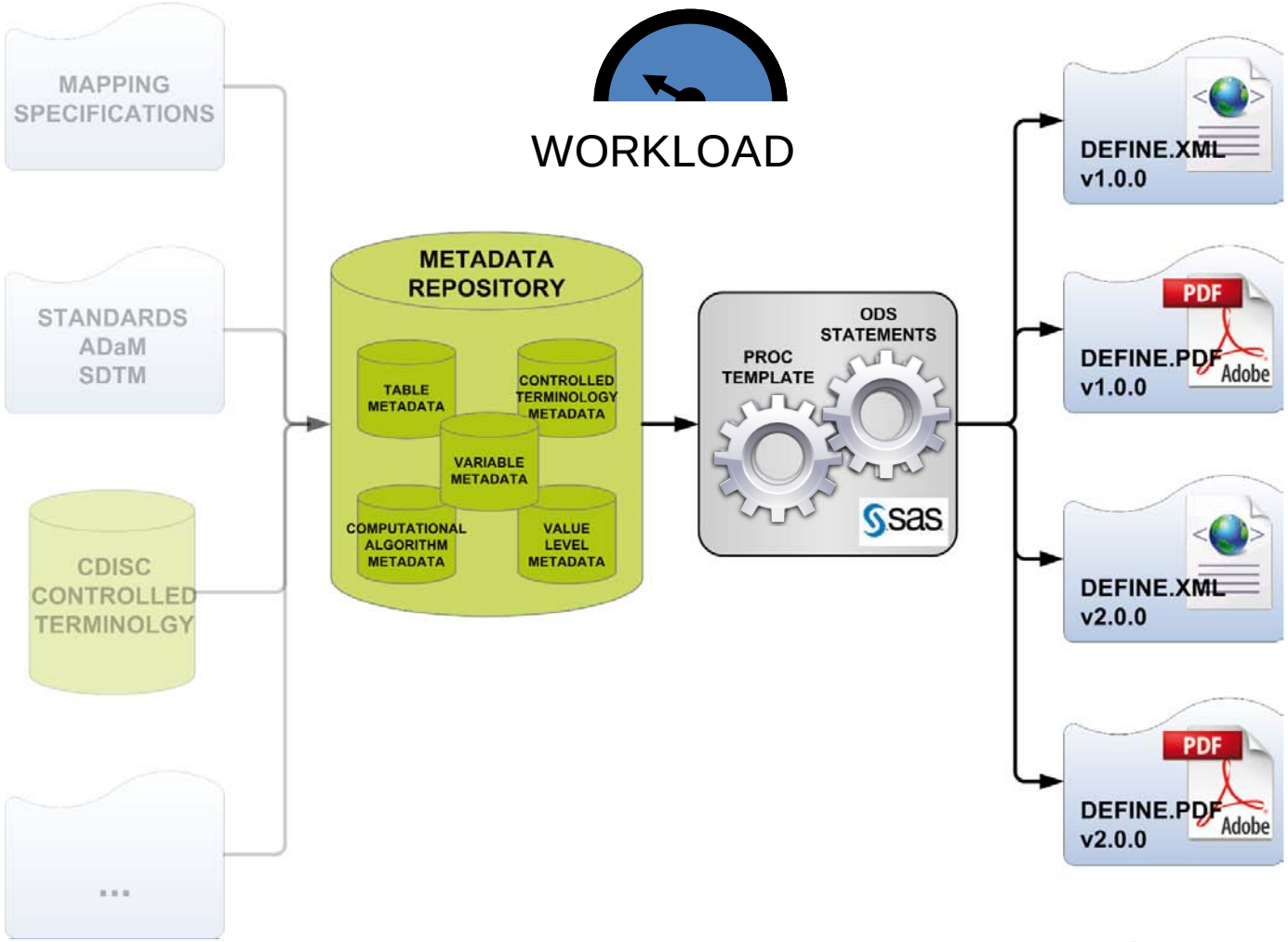
- Definition of comments now at:
  - dataset level
  - variable level
  - value level
- Referenced and re-usable

Minor Changes

- Deprecated components
  - Renamed to match with existing ODM components
    - Computational Algorithms
    - Labels
    - …
- Origins
  - ADaM predecessors
  - Range of pages

Business & Decision Life Sciences

Example: Generating a Define.xml



Business & Decision Life Sciences

Business & Decision Life Sciences

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

Business & Decision Life Sciences

Thank you

Brussels, Belgium, 22 May 2014

Business & Decision Life Sciences

Business & Decision
Life Sciences

Anne-Sophie Bekx | Line Manager Data Standards | anne-sophie.bekx@businessdecision.com

Business & Decision Life Sciences
Sint-Lambertusstraat - 141 rue Saint-Lambert
B-1200 Brussels
T: +32 2 774 11 00
F: +32 2 774 11 99
lifesciences@businessdecision.com
http://www.businessdecision-lifesciences.com/