SAS Drug Development as a platform governing data standards specification and clinical data exchange between pharmaceutical organizations

12th October 2011, , Brighton, UK.

Tina Apers
Business & Decision Life Sciences

Tel +32 2 774 11 00
Fax +32 2 774 11 99

info.ls-be@businessdecision.com

Sint-Lambertusstraat 141
1200 Brussels

www.businessdecision-lifesciences.com

Mark Lambrecht
SAS

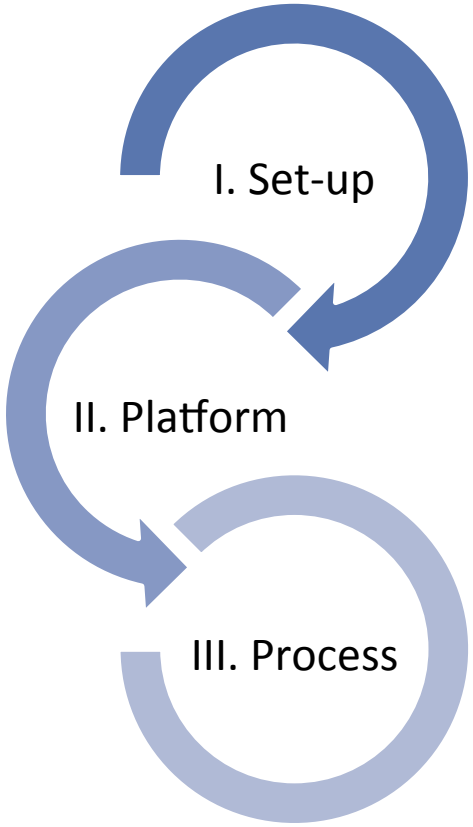
Tel +32 2 766 07 00
Fax +32 2 766 07 77

info@sbx.sas.com

Hertenbergstraat 6
3080 Tervuren

www.sas.com/industry/pharma

Contents



Business & Decision Life Sciences

sas THE POWER TO KNOW.

I. Challenges to apply data standardization during trial

- Applying CDISC standards **before** trial start (as opposed to data conversion before submission)
- Managing different organisational **standards** (CDASH, SDTM, EDC-extensions)
- Managing **study-specific extensions**
- **Validation** and **comparison** of incoming clinical data/ metadata against upfront defined study specifications and global standards
- **Overseeing** all standards-changing activities at any given time with different ongoing trials
- **Technology environment**, that can support activities, run SAS programs and integrate different standard libraries in a secure, controlled, compliant and integrated manner

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Set-up : summary

- Definition of different libraries as central areas for standards-exchange process
- Set-up of study specifications (pick and select) process in SAS
- Exchange of study specifications with external partners such as CRO's
- SAS Drug Development as robust platform for enabling process
  - Central auditable repository
  - SAS program execution
  - Workflow capabilities

Business & Decision Life Sciences

sas THE POWER TO KNOW.

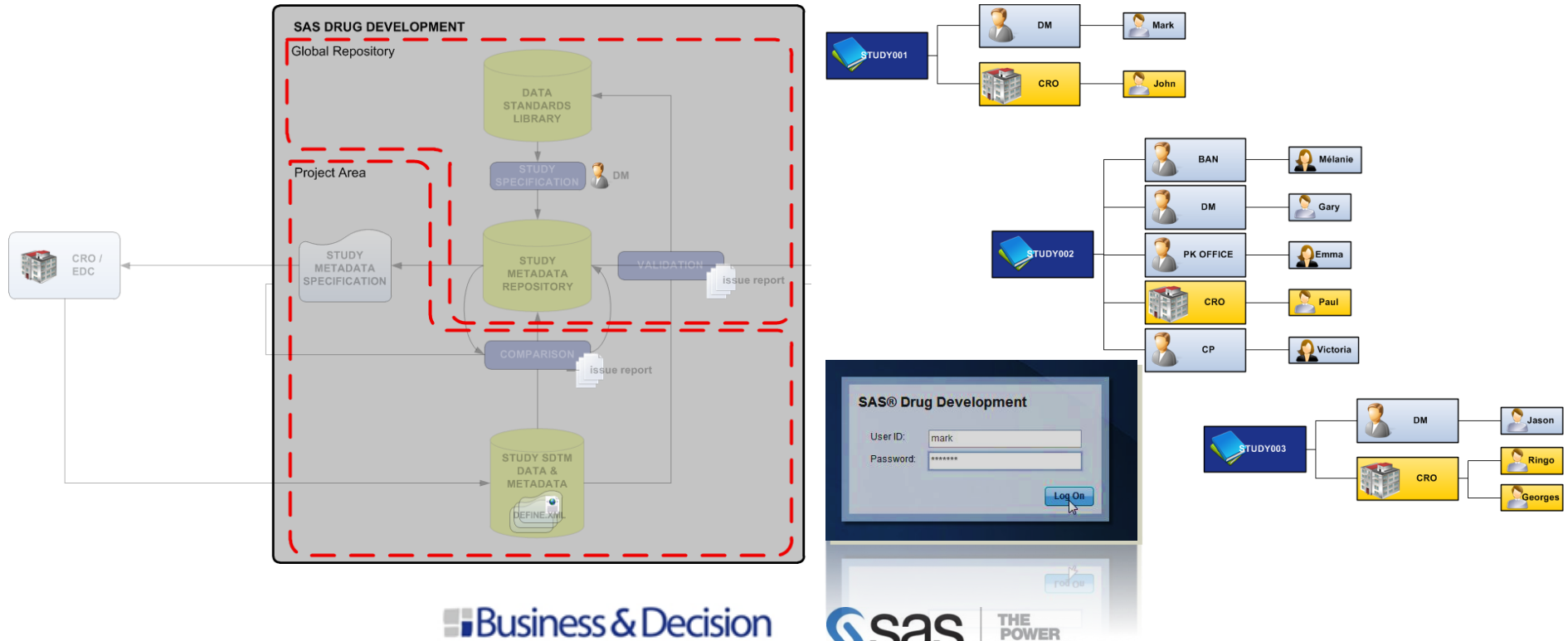
II. Platform

Business & Decision Life Sciences

sas THE POWER TO KNOW.

SAS Drug Development

- Secure repository to store clinical information in a regulatory-compliant manner

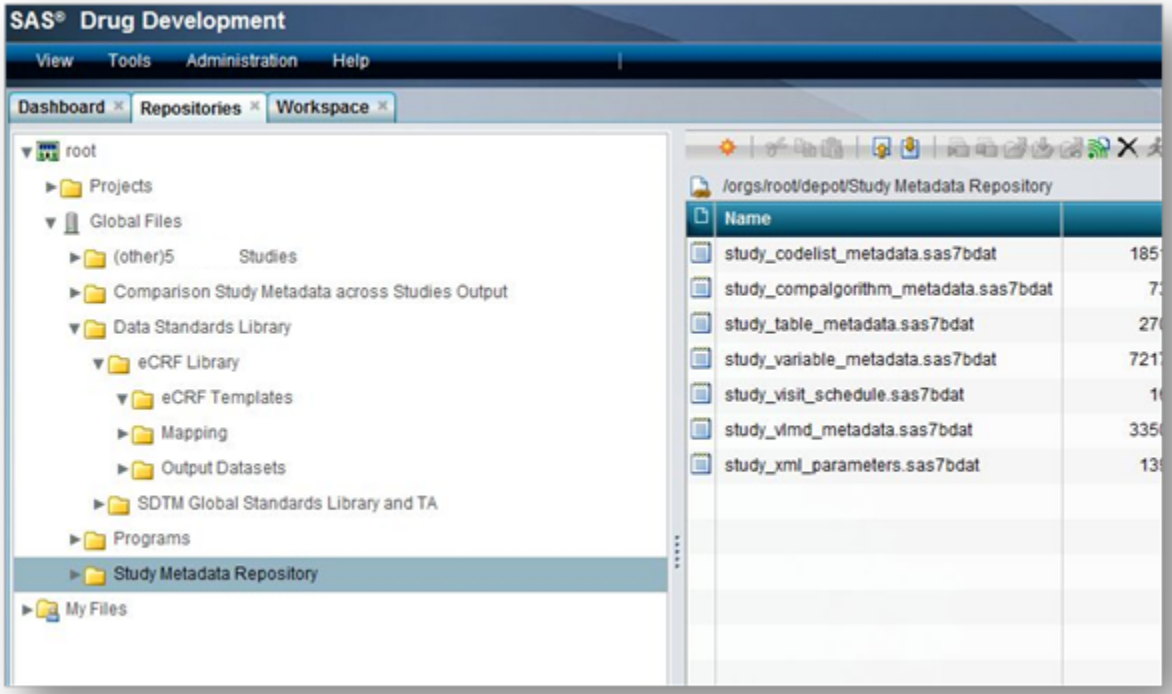


Business & Decision Life Sciences

sas THE POWER TO KNOW.

SAS Drug Development

- Secure repository to store clinical information in a regulatory-compliant manner



Business & Decision Life Sciences

sas THE POWER TO KNOW.

SAS Drug Development

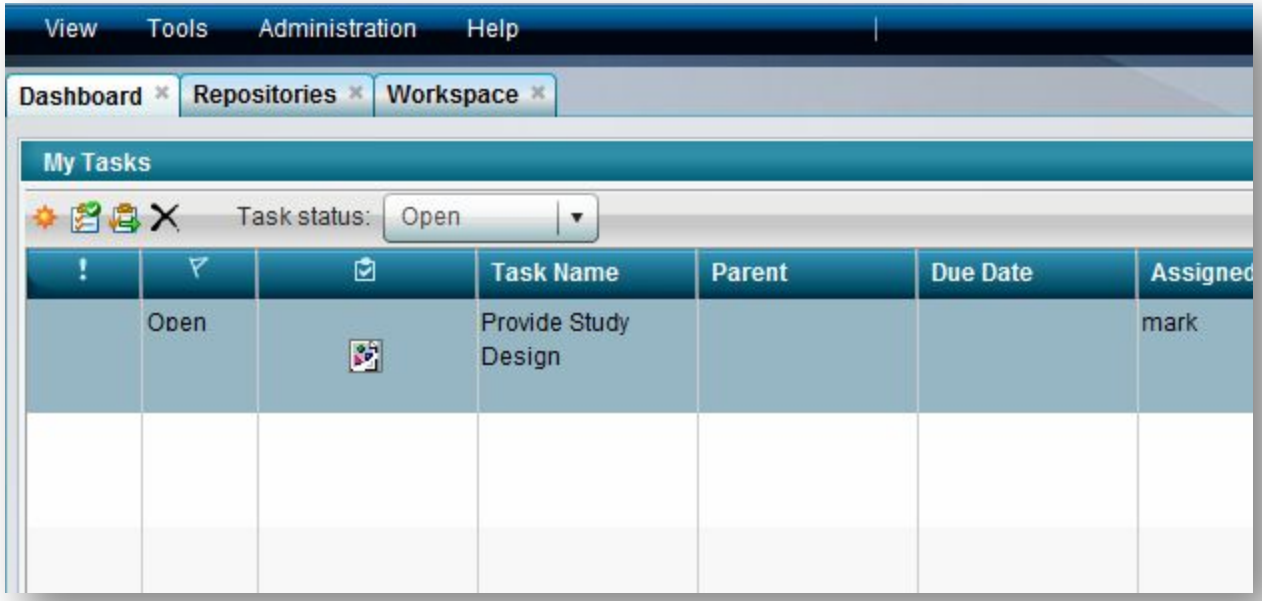
- Accessible using web browser and has integrated SAS execution engine
  - SAS programming
  - CDISC standards

Business & Decision Life Sciences

sas THE POWER TO KNOW.

SAS Drug Development

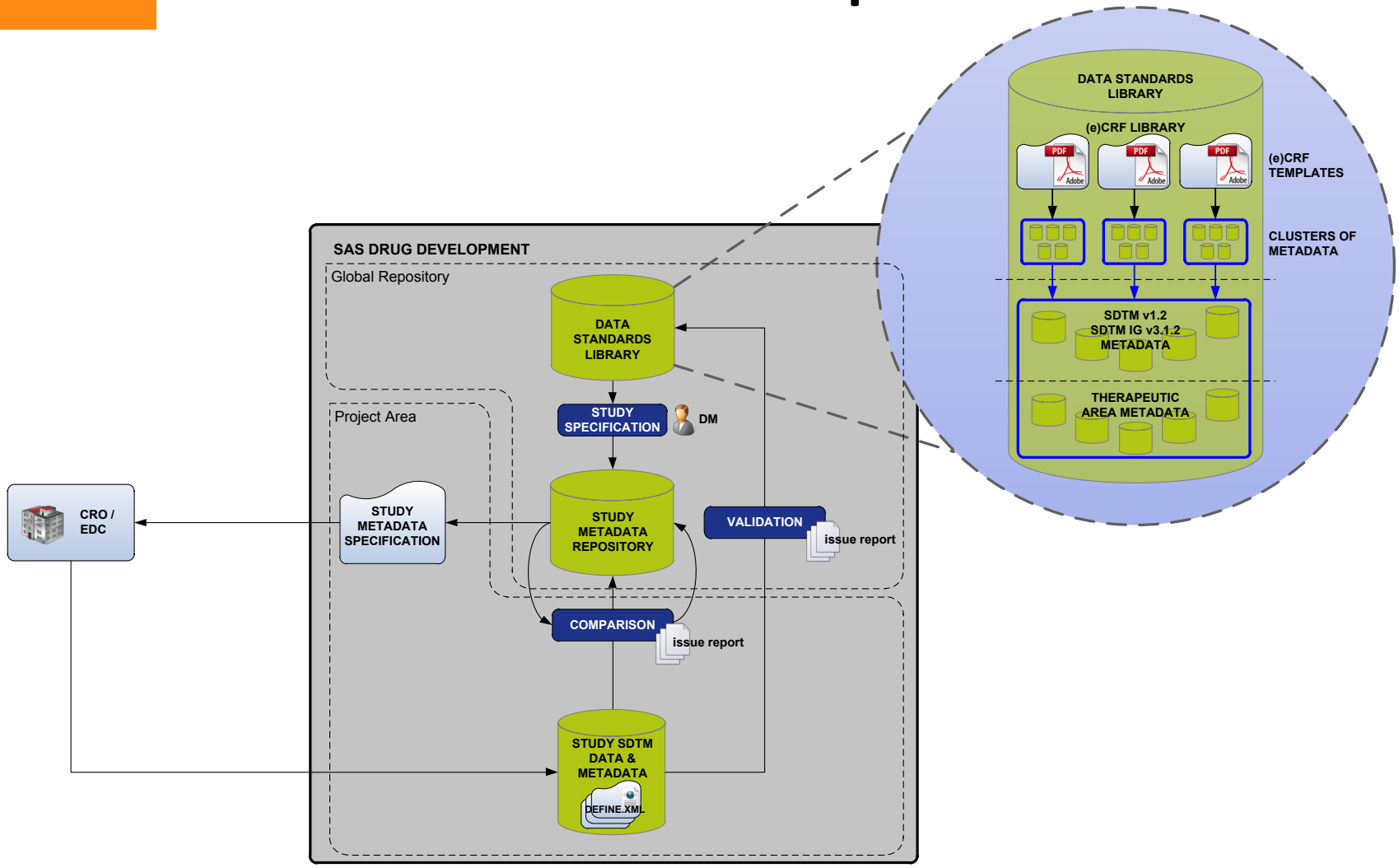
- Rigidly controlled yet flexible business workflows : e.g. upload of new clinical data to SDD



Business & Decision Life Sciences

sas THE POWER TO KNOW.

III. Process explanation



Business & Decision Life Sciences

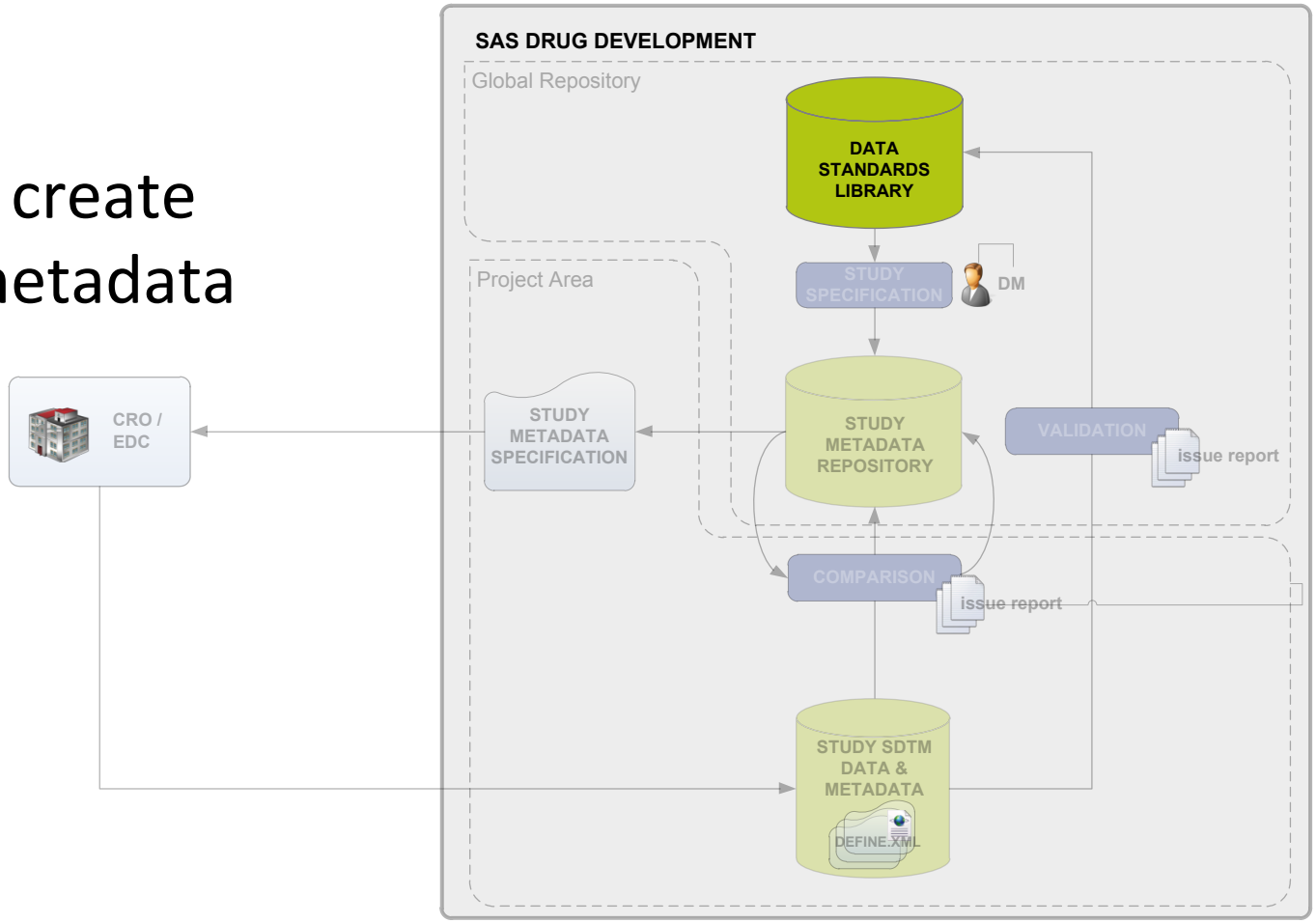
sas THE POWER TO KNOW.

4 different steps

1. Data Standards Library
2. Study Metadata Repository
3. Generating Study Metadata Specifications and Reports
4. Verifying incoming study data and metadata

1. Data Standards Library

- used to create study metadata

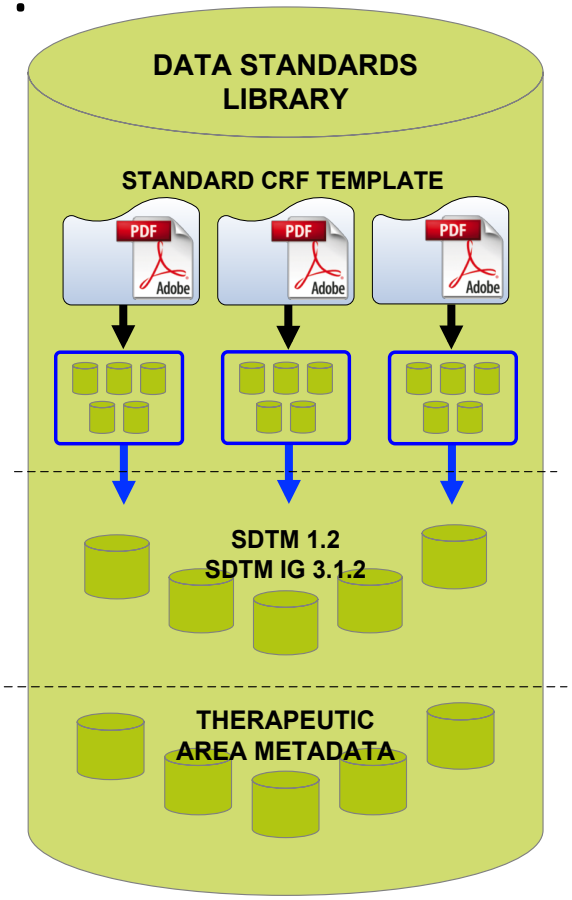


Business & Decision Life Sciences

sas THE POWER TO KNOW.

Data Standards Library (cont.)

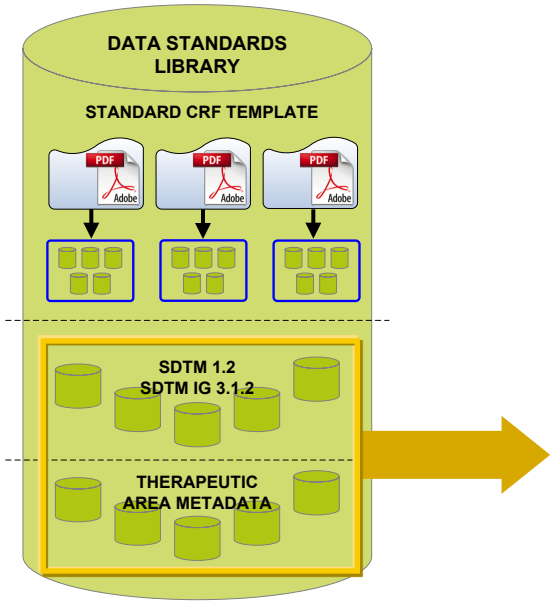
- The Data Standards Library contains :
  - **Standard CRF Templates:**
    - using CDASH metadata
    - with clustered SDTM metadata
    - annotated for CDASH
    - annotated for SDTM
  - **Metadata Definitions :**
    - SDTM Standards
    - Therapeutic Area Standards



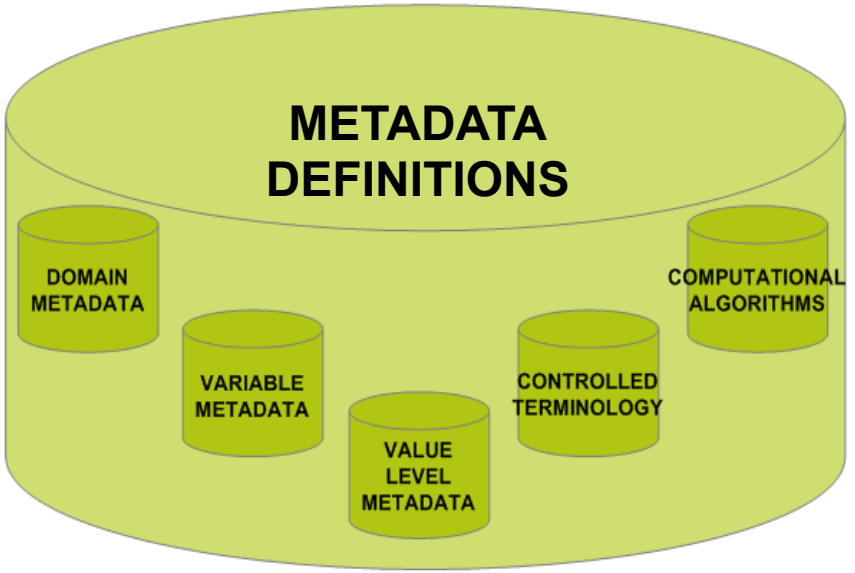
Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata definition



- 5 levels of metadata
- SDTM IG 3.1.2 and TA metadata stored in the same physical tables



Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Table Metadata

Dataset Name
Dataset Label
Class
Structure
Purpose
Keys

Variable Metadata

Variable Name
Variable Label
Type
Controlled Terminology
Origin
Role
Comment

Value Level Metadata

Source Variable
Value
Label
Type
Controlled Terminology
Origin
Comment

Controlled Terminology

Code Value
Code Text

Computational Algorithms

Reference Name
Computation Method

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Table Metadata

- Dataset Name
- Dataset Label
- Class
- Structure
- Purpose
- Keys

Variable Metadata

- Variable Name
- Variable Label
- Type
- Controlled Terminology
- Origin
- Role
- Comment

Value Level Metadata

- Source Variable
- Value
- Label
- Type
- Controlled Terminology
- Origin
- Comment

Controlled Terminology

- Code Value
- Code Text

Computational Algorithms

- Reference Name
- Computation Method

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Table Metadata

- Dataset Name
- Dataset Label
- Class
- Structure
- Purpose
- Keys

	dataset	label		structure	purpose	keys	flag
1	AE	Adverse Events	...nts	One record per a...	Tabulation	STUDYID, USU...	GL
2	CE	Clinical Events	...nts	One record per e...	Tabulation	STUDYID, USU...	GL
3	CM	Concomitant Medications	...rventions	One record per re...	Tabulation	STUDYID, USU...	GL
4	CO	Comments	...cial Purpose ...	One record per c...	Tabulation	STUDYID, USU...	GL
5	DA	Drug Accountability	...ings	One record per dr...	Tabulation	STUDYID, USU...	GL
6	DM	Demographics	Special Purpose ...	One record per s...	Tabulation	STUDYID, USU...	GL
7	DS	Disposition	Events	One record per di...	Tabulation	STUDYID, USU...	GL
8	DV	Protocol Deviations	Events	One record per pr...	Tabulation	STUDYID, USU...	GL
9	EG	ECG Test Results	Findings	One record per E...	Tabulation	STUDYID, USU...	GL
10	EX	Exposure	Interventions	One record per c...	Tabulation	STUDYID, USU...	GL
11	FA	Findings About Events or Interventions	Findings	One record per fi...	Tabulation	STUDYID, USU...	GL
12	GT	GenoType	Findings	One record per s...	Tabulation	STUDYID, USU...	TA
13	IE	Inclusion/Exclusion Criteria Not Met	Findings	One record per in...	Tabulation	STUDYID, USU...	GL

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 interval per subject	Tabulation	STUDYID, USUBJID, EXSTDTC, EXENDTC, EXDOSE, EXTRT	..\Tabulations\ex.xpt
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AETERM, AESTDTC	..\Tabulations\ae.xpt

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Table Metadata

- Dataset Name
- Dataset Label
- Class
- Structure
- Purpose
- Keys

Variable Metadata

- Variable Name
- Variable Label
- Type
- Controlled Terminology
- Origin
- Role
- Comment

Value Level Metadata

- Source Variable
- Value
- Label
- Type
- Controlled Terminology
- Origin
- Comment

Controlled Terminology

- Code Value
- Code Text

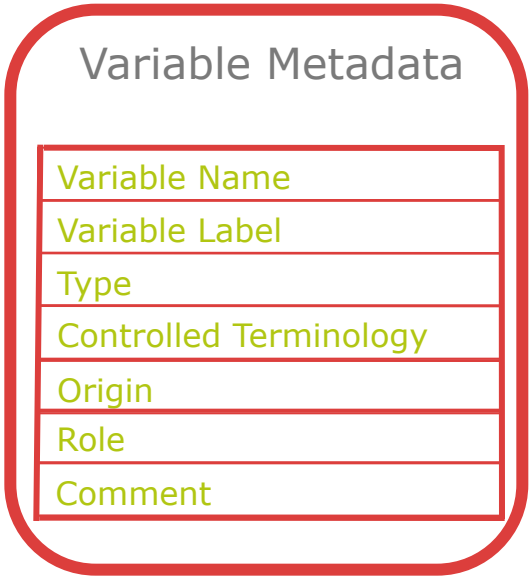
Computational Algorithms

- Reference Name
- Computation Method

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions



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 Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Protocol	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZZ' the 6-digit patient code

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Table Metadata

- Dataset Name
- Dataset Label
- Class
- Structure
- Purpose
- Keys

Variable Metadata

- Variable Name
- Variable Label
- Type
- Controlled Terminology
- Origin
- Role
- Comment

Value Level Metadata

- Source Variable
- Value
- Label
- Type
- Controlled Terminology
- Origin
- Comment

Controlled Terminology

- Code Value
- Code Text

Computational Algorithms

- Reference Name
- Computation Method

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Value Level Metadata

- Source Variable
- Value
- Label
- Type
- Controlled Terminology
- Origin
- Comment

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	PULSE RATE	integer		CRF	
VSTESTCD	SYSBP	SYSTOLIC BLOOD PRESSURE	integer		CRF	
VSTESTCD	WEIGHT	WEIGHT	integer		CRF	

Metadata Definitions

Table Metadata

- Dataset Name
- Dataset Label
- Class
- Structure
- Purpose
- Keys

Variable Metadata

- Variable Name
- Variable Label
- Type
- Controlled Terminology
- Origin
- Role
- Comment

Value Level Metadata

- Source Variable
- Value
- Label
- Type
- Controlled Terminology
- Origin
- Comment

Controlled Terminology

- Code Value
- Code Text

Computational Algorithms

- Reference Name
- Computation Method

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

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	Male
U	Unknown

Controlled Terminology

Code Value
Code Text

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

Table Metadata

- Dataset Name
- Dataset Label
- Class
- Structure
- Purpose
- Keys

Variable Metadata

- Variable Name
- Variable Label
- Type
- Controlled Terminology
- Origin
- Role
- Comment

Value Level Metadata

- Source Variable
- Value
- Label
- Type
- Controlled Terminology
- Origin
- Comment

Controlled Terminology

- Code Value
- Code Text

Computational Algorithms

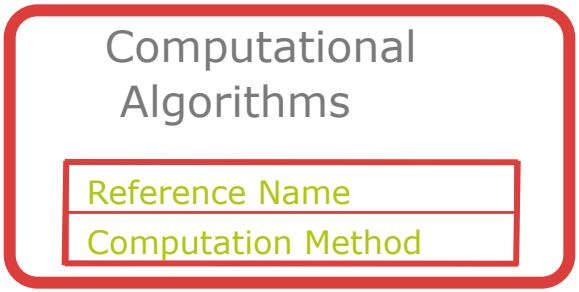
- Reference Name
- Computation Method

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Metadata Definitions

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

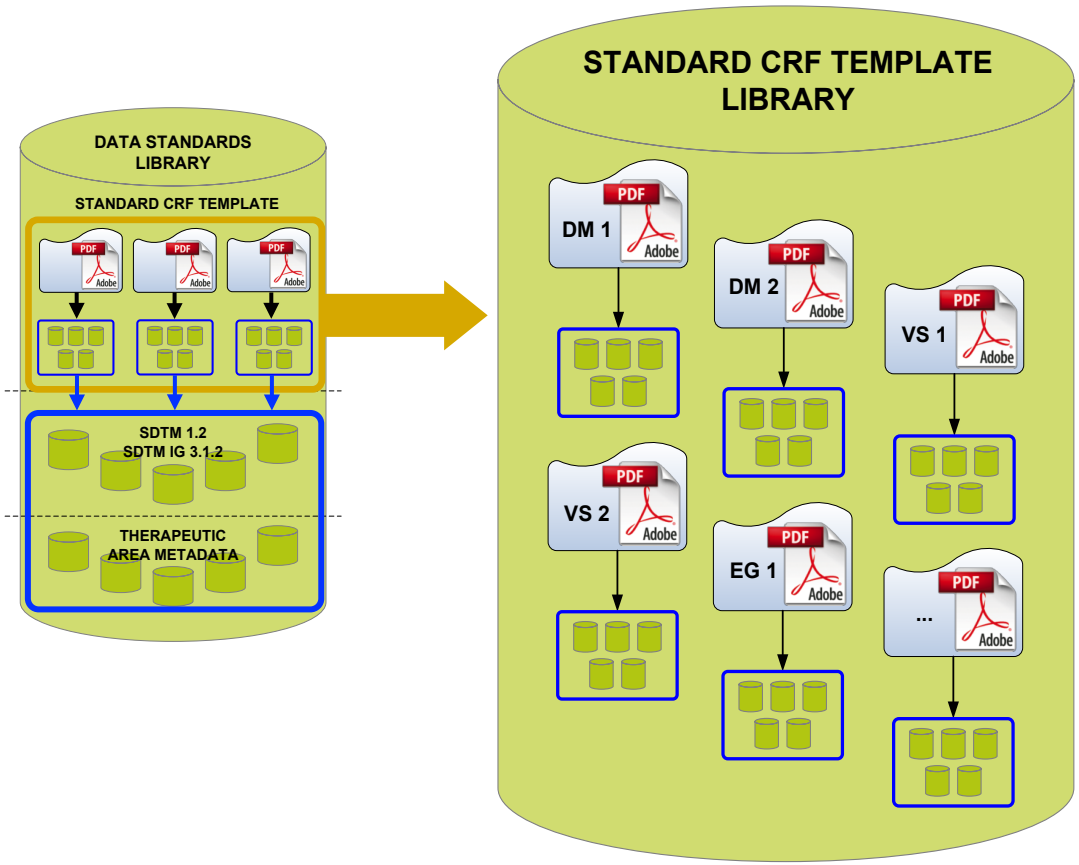


Business & Decision Life Sciences

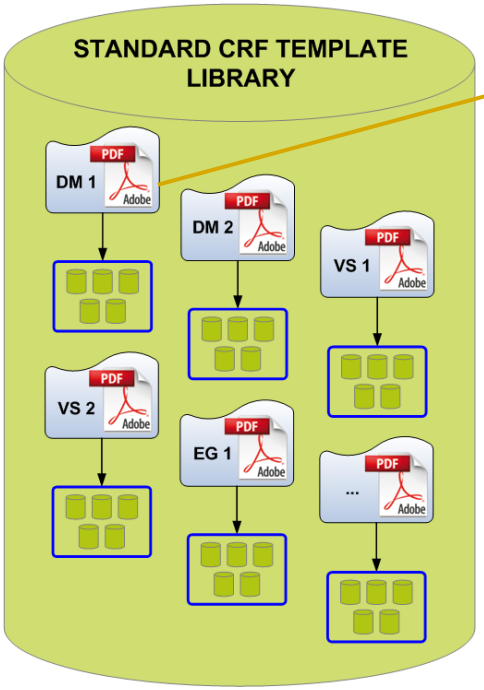
sas THE POWER TO KNOW.

Standard CRF Templates

- **Standard CRF Template Library :**
  - using CDASH metadata
  - with clustered SDTM metadata
  - annotated for CDASH
  - annotated for SDTM



Standard CRF Templates - CDASH



DMDAT

Date of the visit dd mmm yy

Demographics

Demographic data

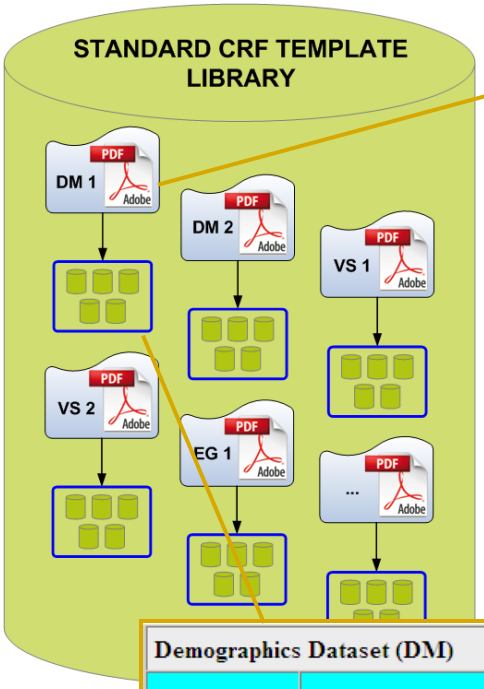
Date of birth dd mmm yyyy BRTHDAT

Race
RACE
- White
- Black or African American
- Asian

Sex
SEX
- Male
- Female

- CRF template using CDASH annotations

Standard CRF Templates - SDTM



Date of the visit dd mmm yy

Demographics

Demographic data

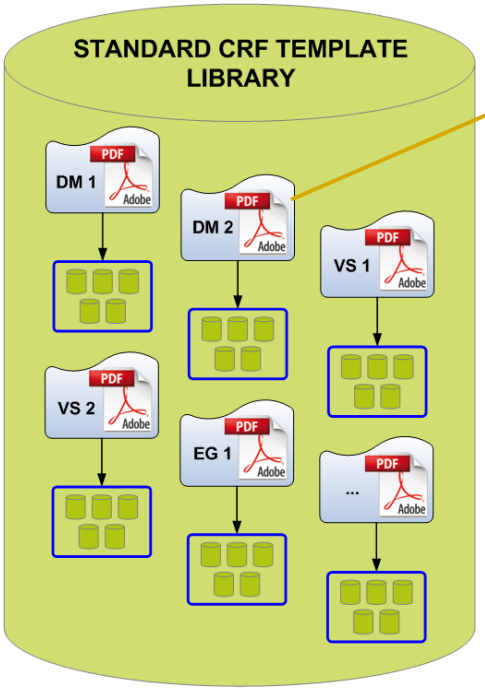
Date of birth dd mmm yyyy — BRTHDTC

Race — RACE
- White
- Black or African American
- Asian

Sex — SEX
- Male
- Female

Demographics Dataset (DM)						..\Tabulations\dm.xpt
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Origin**	**Role**	**Comment**
BRTHDTC	Date/Time of Birth	datetime	ISO 8601	CRF	RECORD QUALIFIER	Date/time of birth of the subject.
SEX	Sex	text	SEX	CRF	RECORD QUALIFIER	Gender of the subject.
RACE	Race	text	RACE	CRF	RECORD QUALIFIER	Race of the subject.

Standard CRF Templates - CDASH



DMDAT

Date of the visit dd mmm yy

Demographics

Demographic data

Date of birth BRTHDAT dd mmm yyyy

Race RACE

- White
- Black or African American
- Asian

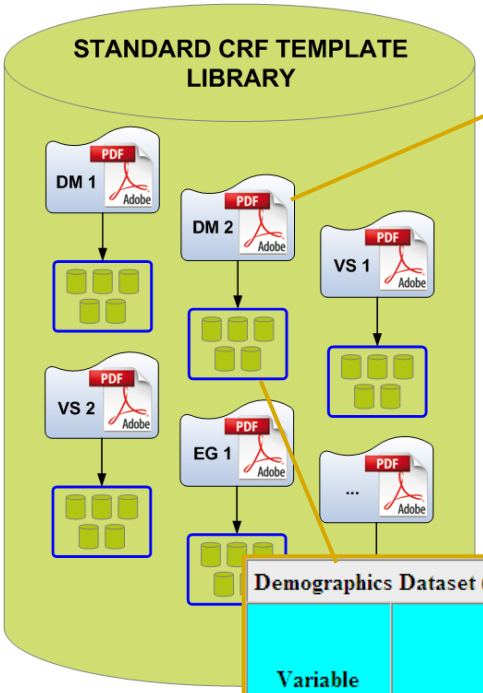
Sex SEX

- Male
- Female

Age ..... years AGEU AGE

- CRF template using CDASH annotations

Standard CRF Templates - SDTM



Date of the visit dd mmm yy

Demographics

Demographic data

Date of birth dd mmm yyyy **BRTHDTC**

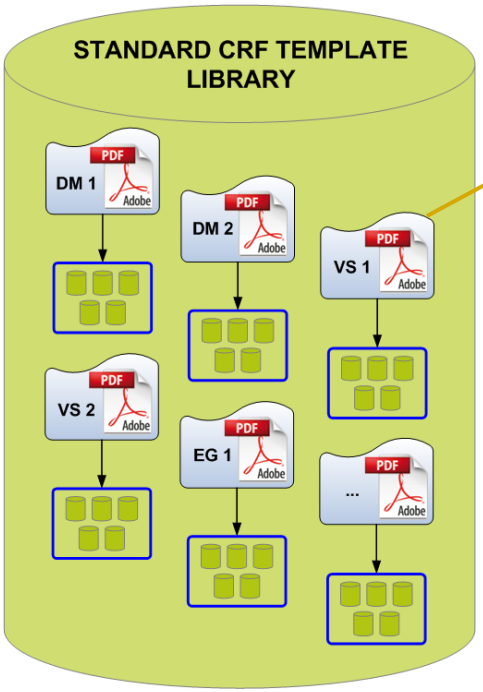
Race **RACE**
- White
- Black or African American
- Asian

Sex **SEX**
- Male
- Female

Age ..... years **AGEU** **AGE**

Demographics Dataset (DM)						..\Tabulations\dm.xpt
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Origin**	**Role**	**Comment**
BRTHDTC	Date/Time of Birth	datetime	ISO 8601	CRF	RECORD QUALIFIER	Date/time of birth of the subject.
AGE	Age	integer		Derived, CRF	RECORD QUALIFIER	Age expressed in AGEU. Derived from RFSTDTC and consent date. See Computational Method: DM.AGE
AGEU	Age Units	text	AGEU	Assigned	VARIABLE QUALIFIER	Units associated with AGE.
SEX	Sex	text	SEX	CRF	RECORD QUALIFIER	Gender of the subject.
RACE	Race	text	RACE	CRF	RECORD QUALIFIER	Race of the subject.

Standard CRF Templates - CDASH



Vital Signs

Vital signs

VSDAT — Date and time performed — dd mmm yy — VSTIM — hh mm — (24 h time clock)

VSTEST Weight — VSORRESU — VSORRES — kg

VSTEST Height — VSORRES — cm — VSORRESU

VSTEST Pulse — Supine* — VSORRES — beats/min — VSORRESU

Standing* — beats/min

VSTEST Blood pressure — VSPOS — Supine* — VSORRES — systolic/diastolic (mmHg) — VSORRESU

Standing* — systolic/diastolic (mmHg)

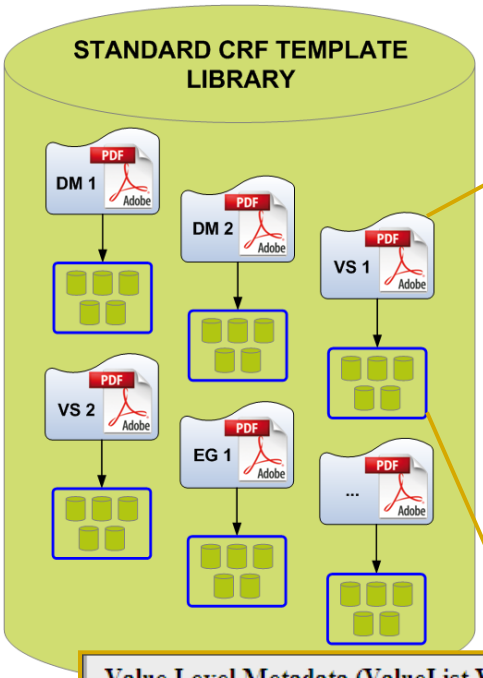
* (supine: after ≥ 5 min. rest; standing: after ≥ 1 min. standing)

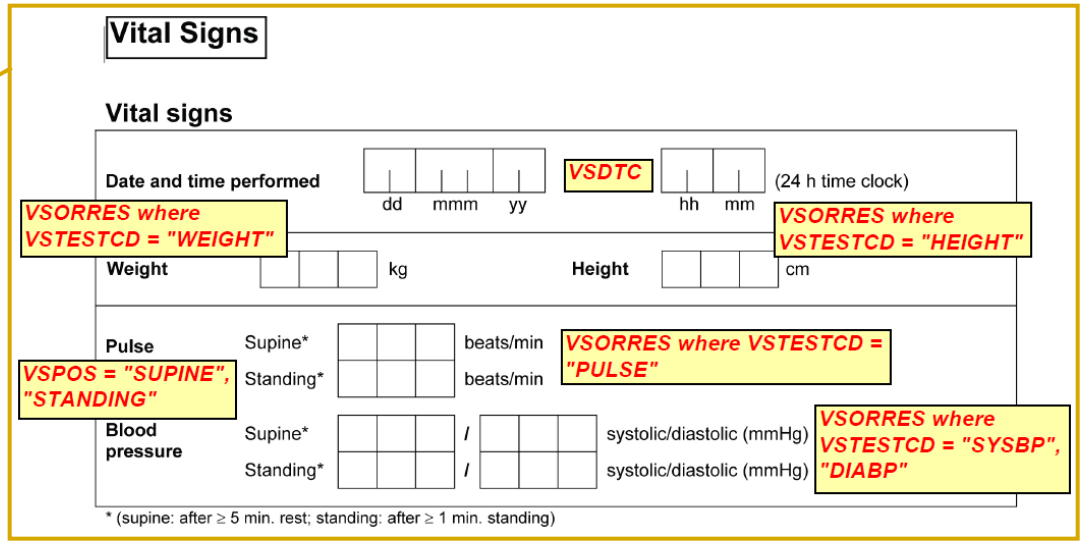
- CRF template using CDASH annotations

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Standard CRF Templates - SDTM

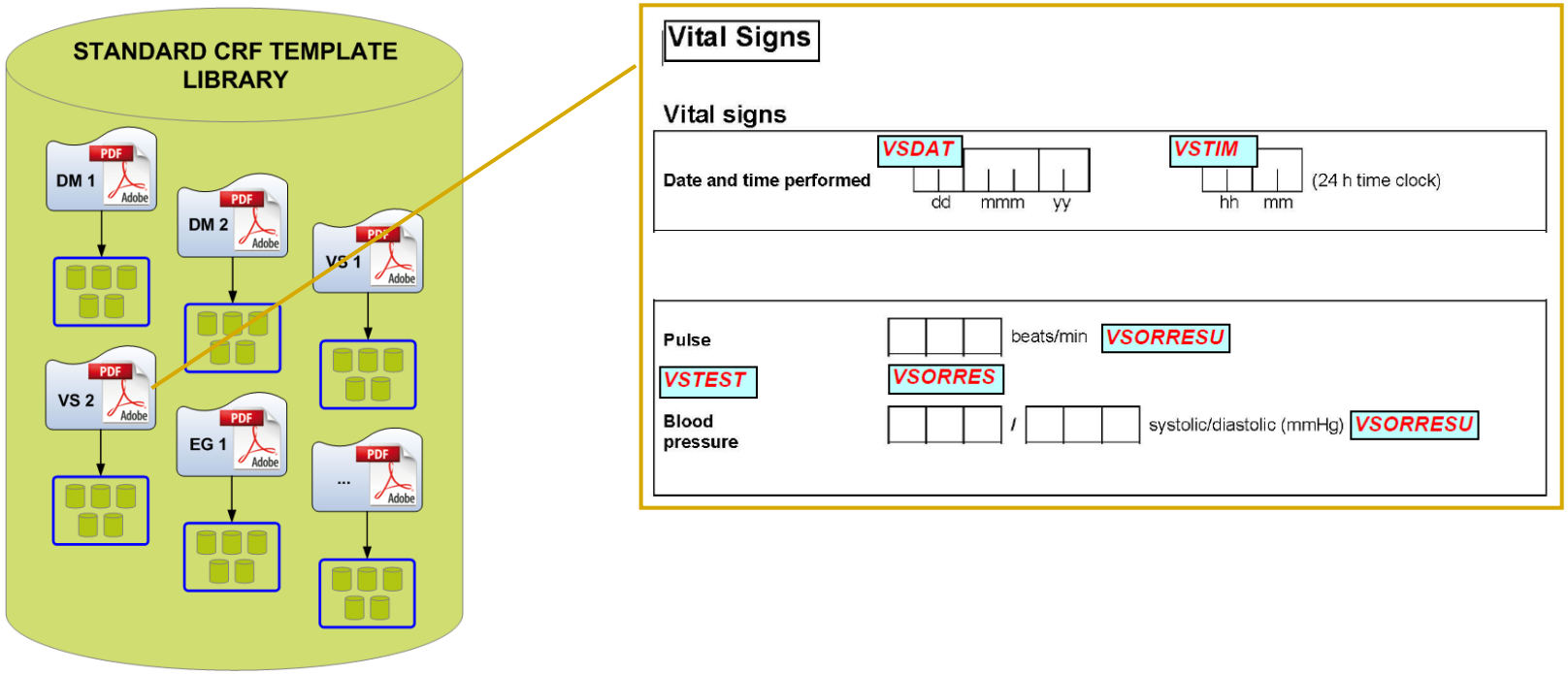




Value Level Metadata (ValueList.VS.VSTESTCD)

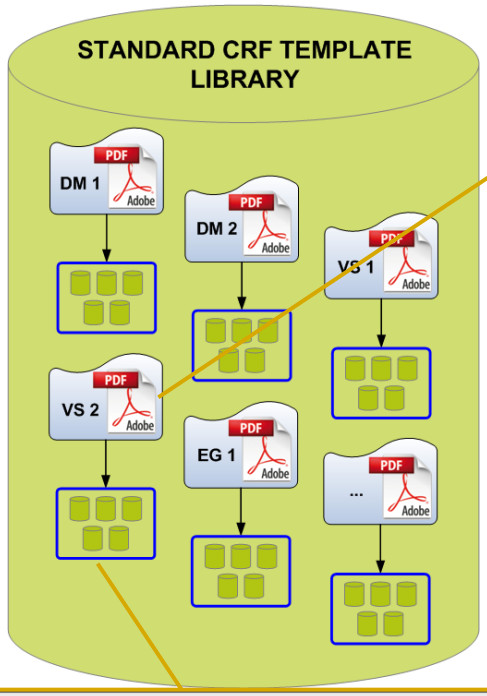
Source Variable	Value	Label	Type	Controlled Terminology	Origin	Role	Comment
VSTESTCD	DIABP	DIASTOLIC BLOOD PRESSURE	integer		CRF		
VSTESTCD	HEIGHT	HEIGHT	integer		CRF		
VSTESTCD	PULSE	PULSE RATE	integer		CRF		
VSTESTCD	SYSBP	SYSTOLIC BLOOD PRESSURE	integer		CRF		
VSTESTCD	WEIGHT	WEIGHT	integer		CRF		

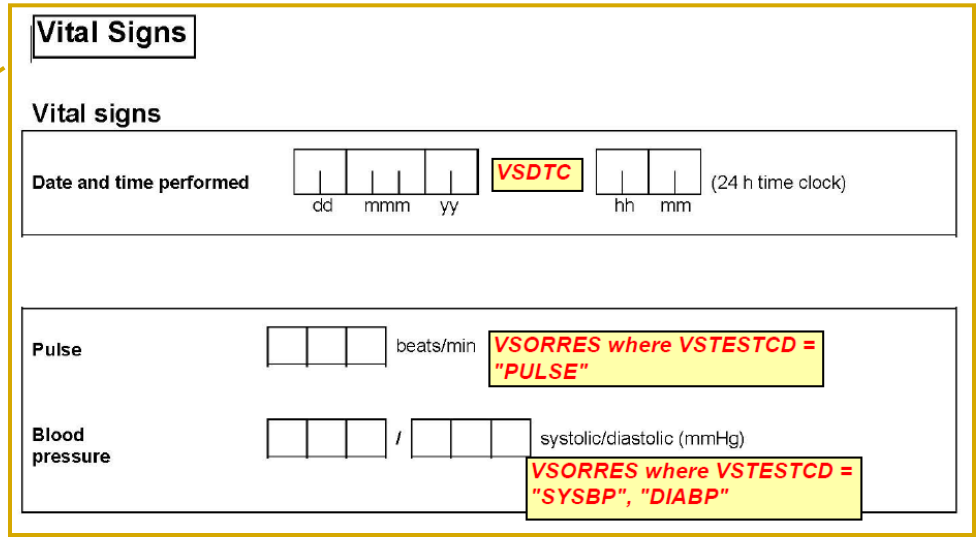
Standard CRF Templates - CDASH



- CRF template using CDASH annotations

Standard CRF Templates - SDTM

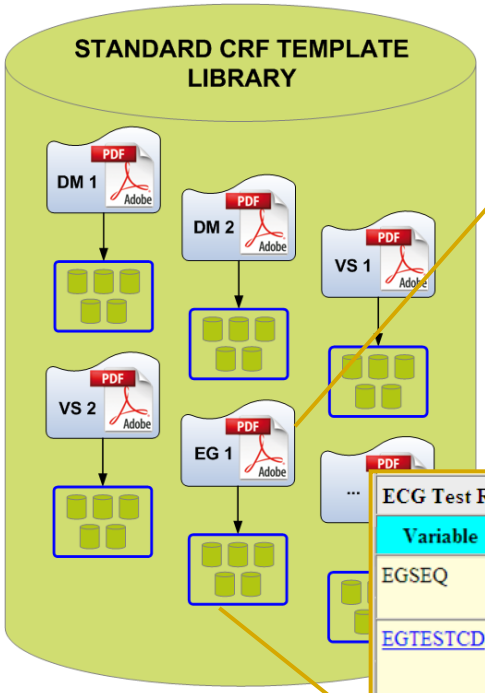




Value Level Metadata (ValueList.VS.VSTESTCD)

Source Variable	Value	Label	Type	Controlled Terminology	Origin	Role	Comment
VSTESTCD	DIABP	DIASTOLIC BLOOD PRESSURE	integer		CRF		
VSTESTCD	PULSE	PULSE RATE	integer		CRF		
VSTESTCD	SYSBP	SYSTOLIC BLOOD PRESSURE	integer		CRF		

Standard CRF Templates - SDTM



GENERAL DATA TRANSFER AGREEMENT

TYPE OF DATA: ECG

Version 1.1

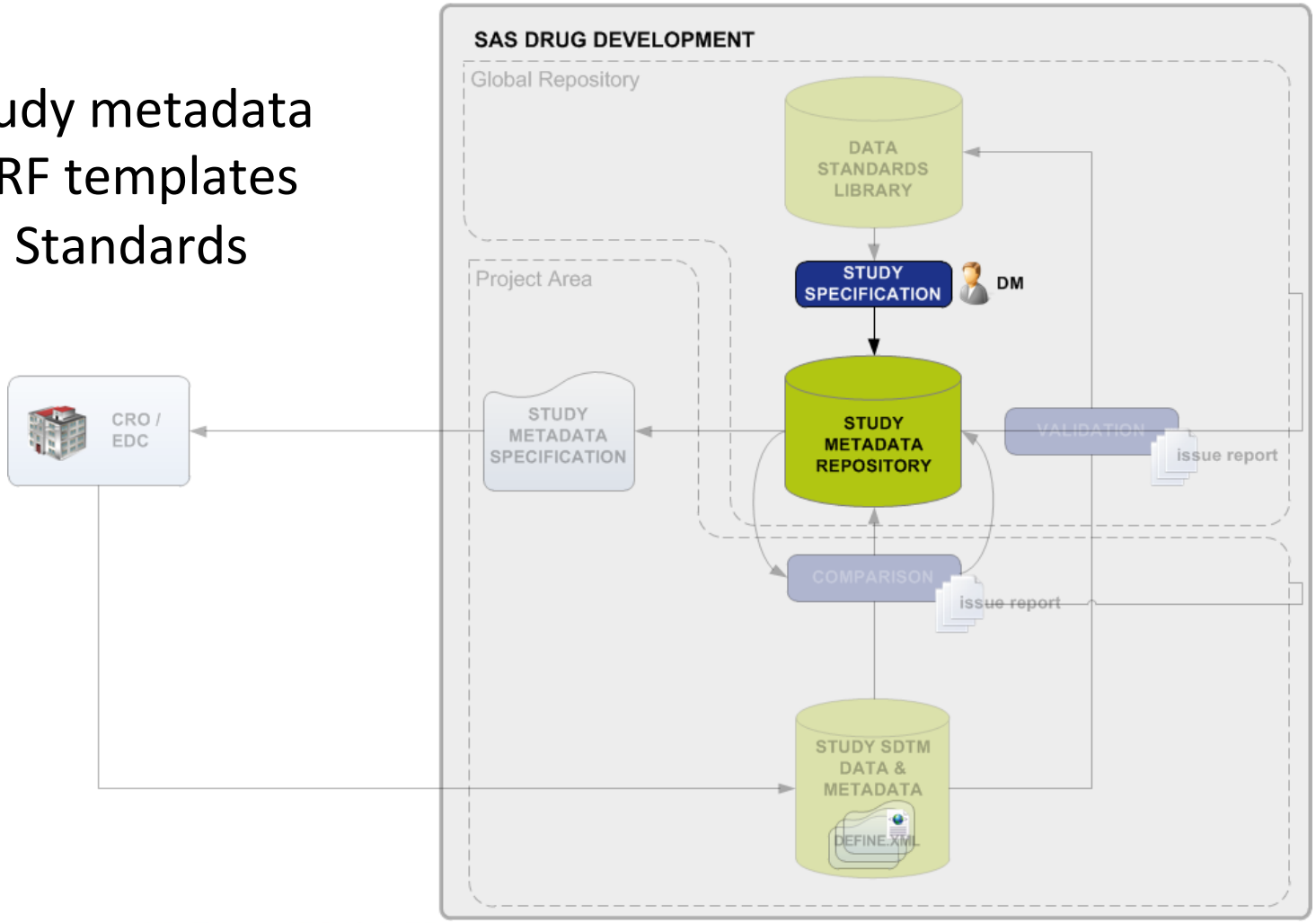
Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Core	Length	CDISC Notes
EGTEST	ECG Test or Examination Name	Char		Req	40	Verbatim name of the test or examination used to obtain the measurement or finding. The value in EGTEST cannot be longer than 40 characters. Examples: Summary (Mean) PR Duration, Summary (Mean) QT Duration
EGCAT	Category for ECG	Char		Perm	40	Used to categorize ECG observations across subjects. Examples: MEASUREMENT, FINDING, INTERVAL
EGSCAT	Subcategory for ECG	Char		Perm	40	A further categorization of the ECG.
EGPOS	ECG Position of Subject	Char	POSITION	Perm	20	Position of the subject during a measurement or examination. Examples: SUPINE, STANDING, SITTING.
EGORRES	Result or Finding in Original Units	Char		Exp	200	Result of the ECG measurement or finding as originally received or collected. Examples of expected values are 62 or 0.151 when the result is an interval or measurement, or "ATRIAL FIBRILLATION" or "QT PROLONGATION" when the result is a finding.
EGORRESU	Original Units	Char	UNIT	Perm	20	Original units in which the data were collected. The unit for EGORRES. Examples: sec or msec.
EGSTRESC	Character Result/Finding in Std Format	Char		Exp	200	Contains the result value for all findings, copied or derived from EGORRES in a standard format or standard units.

ECG Test Results Dataset (EG) — ..\..\tabulations\sdtm\eg.xpt

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
EGSEQ	Sequence Number	integer		Derived	IDENTIFIER	
EGTESTCD	ECG Test or Examination Short Name	text		Assigned	TOPIC	
EGTEST	ECG Test or Examination Name	text	EGTEST	Derived	SYNONYM QUALIFIER	
EGORRES	Result or Finding in Original Units	text		eDT, CRF	RESULT QUALIFIER	-ORRES value can be empty if associated with a populated --STAT and/or -REASND variable or if associated/linked with record(s) in the corresponding SUPPQUAL or the CO domain. All other observations with empty -ORRES fields are removed from the domain.

2. Study Metadata Repository

Generating study metadata by selecting CRF templates from the Data Standards Library



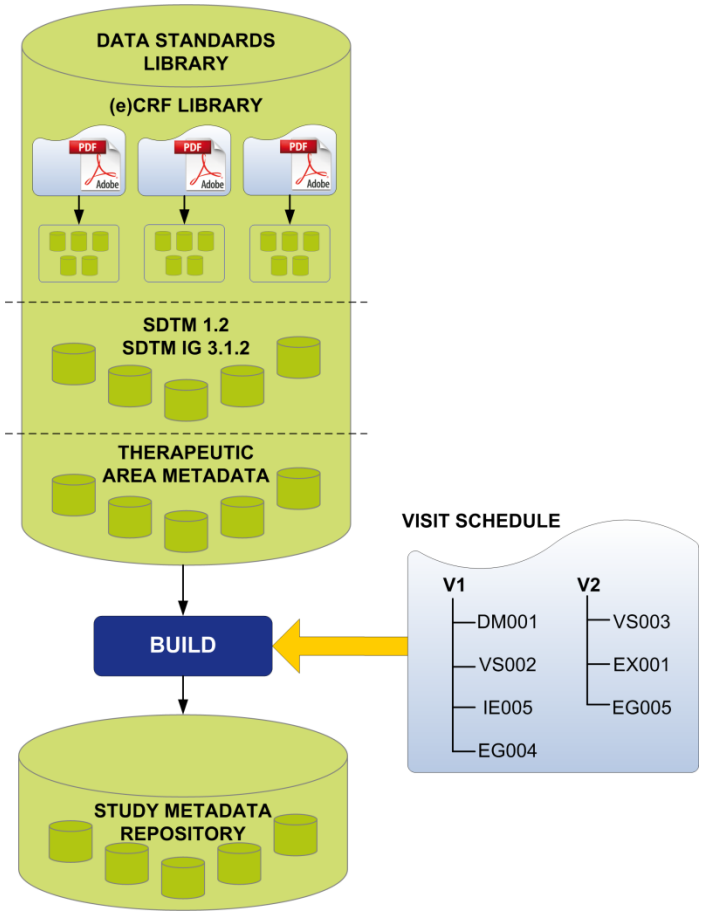
Business & Decision Life Sciences

sas THE POWER TO KNOW.

Generation of Study Specifications

Generation of study metadata requires :

- visit schedule and CRF templates
- study parameters

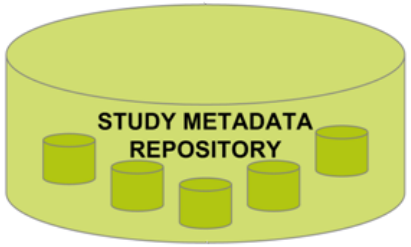


Business & Decision Life Sciences

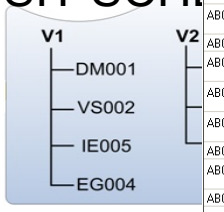
sas THE POWER TO KNOW.

Study Metadata Repository

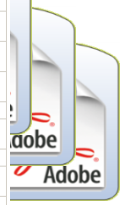
- **The Study Metadata Repository contains :**
  - 5 levels of metadata in physical tables
  - Visit schedule and selected CRF templates
  - Study specifications/parameters :
    - Define Values: Study name, MedDRA & WHODrug version, metadata versions, …
    - Trial Summary Values: Age Group, Trial Indication, Planned Number of Subjects, …
  - Possibility to add study spec…ons



VISIT SCH…TES



STUDYID	PAR_NAME	PAR_VALUE	PAR_ID	PAR_DESCRIPTION
ABC-0001	StudyOID	ABC-0001	Header	Study Identifier(Requried Internal ODM key used to uniquely identify the entity)
ABC-0001	StudyName	ABC-0001	Header	Study Name (Part of the Globalvariables in ODM markup (required section))
ABC-0001	StudyDescription	ABC-0001	Header	Sudy title (Part of the Globalvariables in ODM markup (required section))
ABC-0001	ProtocolName	ABC-0001	Header	Name of the Protocol (Part of the Globalvariables in ODM markup (required section))
ABC-0001	ExternalDictionary	MedDRA	Footer	Name of the External Dictionary (Optional)
ABC-0001	DictionaryVersion	13.0	Footer	Version of the External Dictionary (Optional)
ABC-0001	DictionaryOID	MEDDRA	Footer	Unique Identifier for an External Dictionary (Optional)
ABC-0001	DictionaryDesc	ADVERSE EVENT DICTIONARY	Footer	Description of External Dictionary
ABC-0001	ExternalDictionary	WHO Drug	Footer	Name of the External Dictionary (Optional)
ABC-0001	DictionaryVersion	WHO-DD 10:MAR	Footer	Version of the External Dictionary (Optional)
ABC-0001	DictionaryOID	WHODRUG	Footer	Unique Identifier for an External Dictionary (Optional)
ABC-0001	DictionaryDesc	DRUG DICTIONARY	Footer	Description of External Dictionary
ABC-0001	DictionaryOID	ISO 8601	Footer	Unique Identifier for an External Dictionary (Optional)
ABC-0001	DictionaryDesc	ISO 8601	Footer	Description of External Dictionary
ABC-0001	DictionaryOID	ISO 3166	Footer	Unique Identifier for an External Dictionary (Optional)
ABC-0001	DictionaryDesc	ISO 3166	Footer	Description of External Dictionary

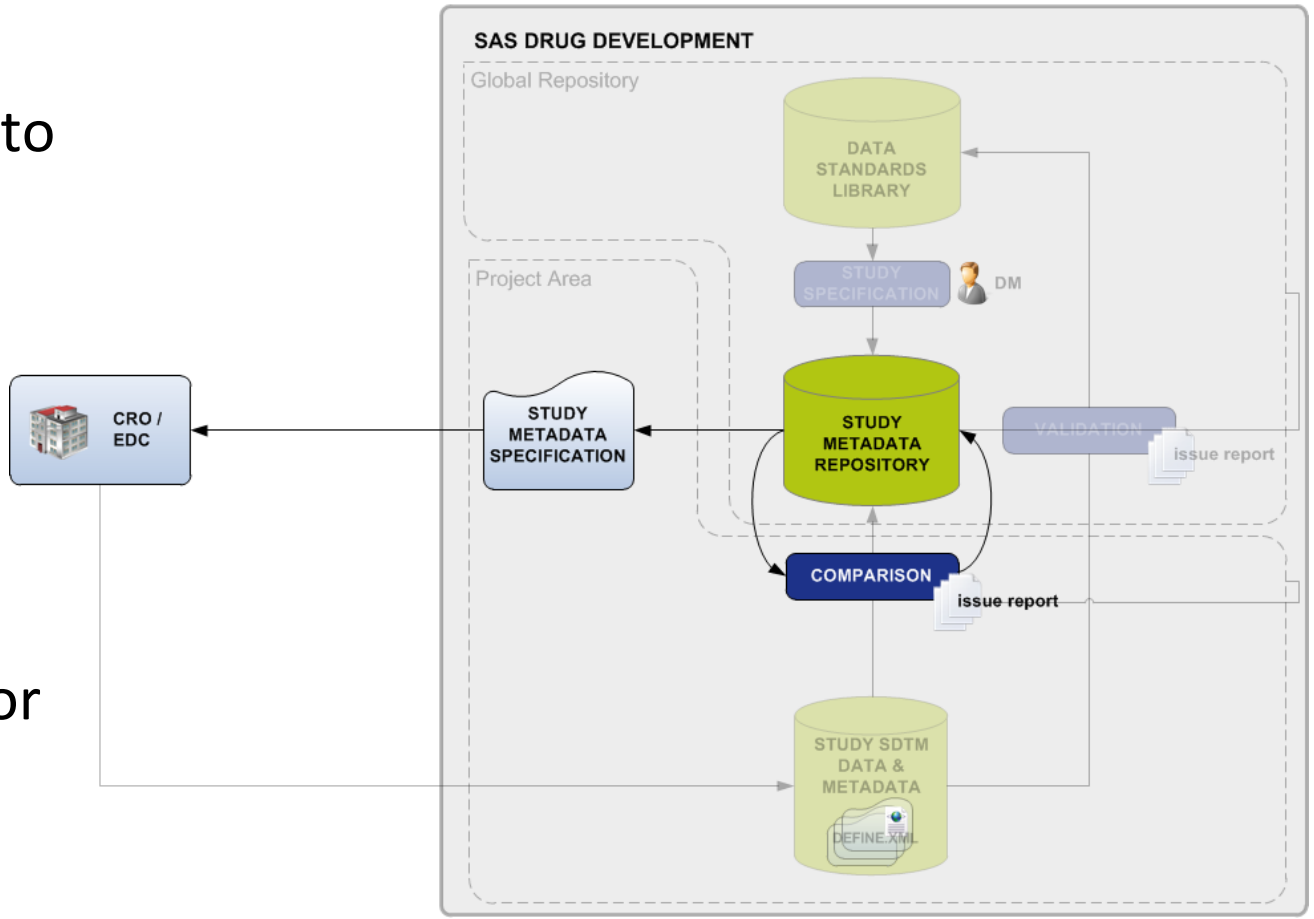


Business & Decision Life Sciences

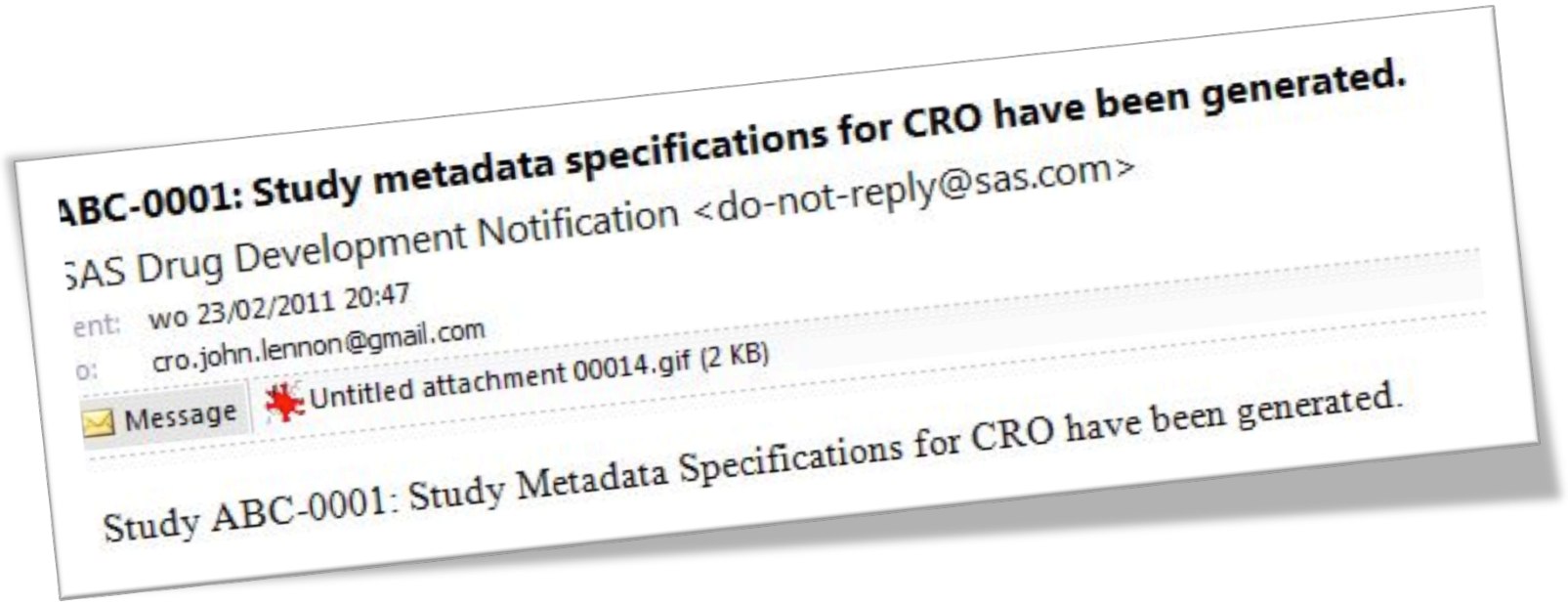
sas THE POWER TO KNOW.

3. Generating the Study Specifications and Reports

- Study Metadata Repository is used to measure the consistency of metadata
- Study metadata is sent to external or internal partners for database build



Study Metadata Specifications and Reports



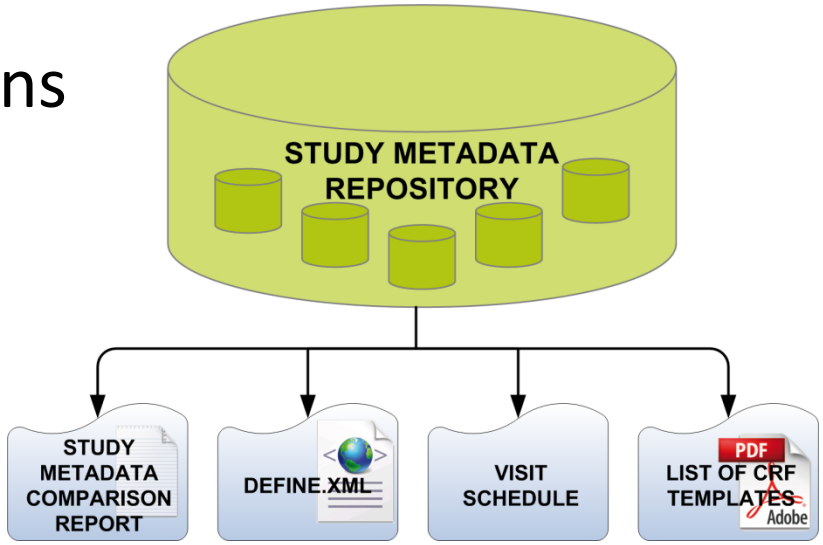
Business & Decision Life Sciences

sas THE POWER TO KNOW.

Study Metadata Specifications and Reports

- Comparison reports
  1. study metadata vs. Data Standards Library
  2. study metadata across the Study Metadata Repository

- Study metadata specifications
  1. 5 levels of metadata
  2. define.xml
  3. list of CRF templates
  4. visit schedule

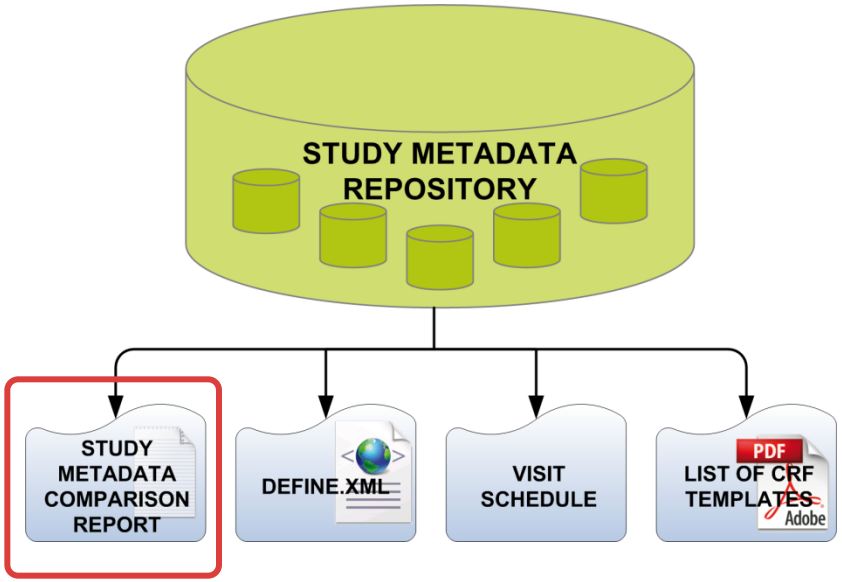


Business & Decision Life Sciences

sas THE POWER TO KNOW.

Study Metadata Comparison Reports

- Comparison reports
  a) To compare study metadata against the Data Standards Library
  b) To measure metadata consistency between studies



Comparison done across all the levels of metadata existing in the Study Metadata Repository

Business & Decision Life Sciences

sas THE POWER TO KNOW.

a) Comparison with the Data Standards Library

- 3 different statuses :
  - *Ok* : metadata present in Data Standards Library
  - *Study Specific* : metadata not present in the Data Standards Library
  - *Difference* : difference in the metadata value

Status	Dataset	Variable	*Length*	A1000-0000	Global Standards
Difference	CM	USUBJID	40	X	
Difference	CM	USUBJID	20		X
Study specific	CM	VISIT	40	X	
Study specific	CM	VISITNUM	8	X	
Ok	CO	CODTC	20	X	X

Business & Decision Life Sciences

sas THE POWER TO KNOW.

b) Comparison across studies

- 2 different statuses:
  - Ok : no metadata conflict
  - Difference : difference in metadata value between studies

Status	Dataset	*Label*	A1000_0000	A1000_0001	A1000_0002
Ok	AE	Adverse Events	X	X	X

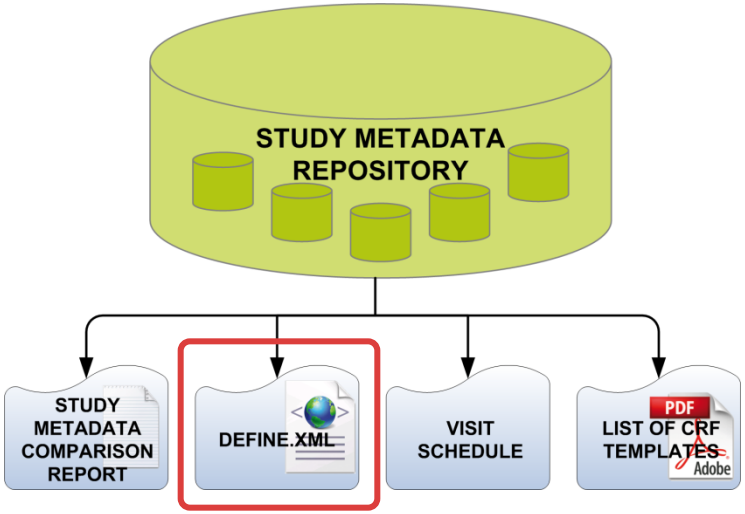
Status	Reference	*Computational algorithm*	A1000_0000	A1000_0001	A1000_0002
Difference	DM.AGE	Equal to (DS.DSSTDTC-DM.BRTHDTC)/365.25 where DSTERM = 'INFORMED CONSENT OBTAINED'		X	
Difference	DM.AGE	Equals to (DS.DSSTDTC-DM.BRTHDTC)/365.25 where DSTERM = 'INFORMED CONSENT OBTAINED'	X		X

Difference	DM	Demographic		X	
Ok	DS	Disposition	X	X	X

Study Specifications for Data Collection Area

Study specifications

1. 5 tables, one for each level of metadata
2. Define.xml



Demographics Dataset (DM)						..\Tabulations\dm.xpt
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text		Protocol, CRF	IDENTIFIER	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	Two-character abbreviation for the domain most relevant to the observation.
USUBJID	Unique Subject Identifier	text		Derived	IDENTIFIER	Unique subject identifier within the submission. This is the concatenation of STUDYID, SITEID and SUBJID.
SUBJID	Subject Identifier for the Study	text		CRF	TOPIC	Unique subject identifier within the study.

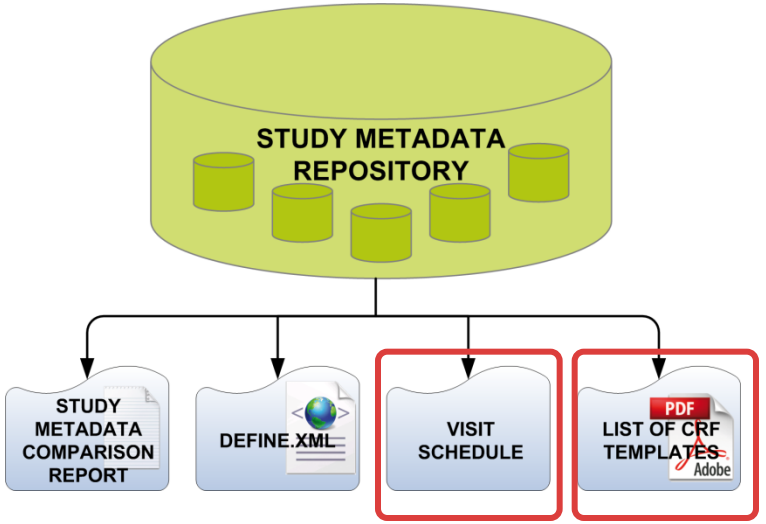
Business & Decision Life Sciences

sas THE POWER TO KNOW.

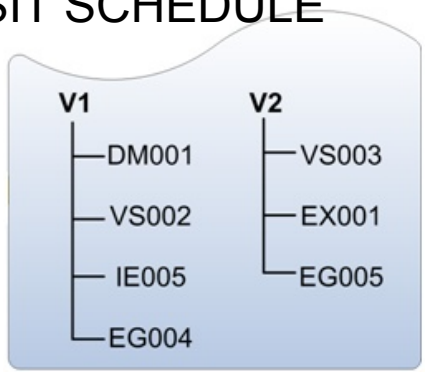
Study Specifications for Data Collection Area

Study metadata specifications

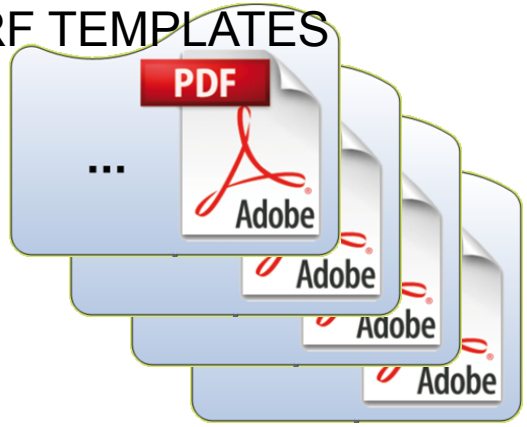
3. Visit schedule
4. CRF templates



VISIT SCHEDULE



CRF TEMPLATES



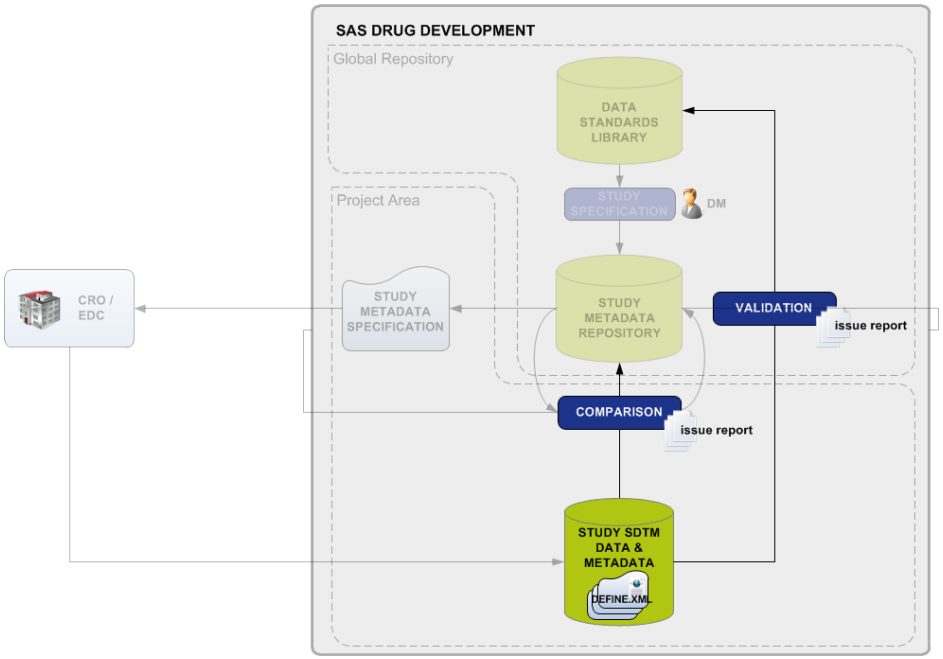
Business & Decision Life Sciences

sas THE POWER TO KNOW.

4. Incoming data and metadata verification

The data and metadata are verified by a two-step process :

- Comparison against the Study Metadata Library
- Validation against the Data Standards Library



Business & Decision Life Sciences

sas THE POWER TO KNOW.

Comparison of study metadata

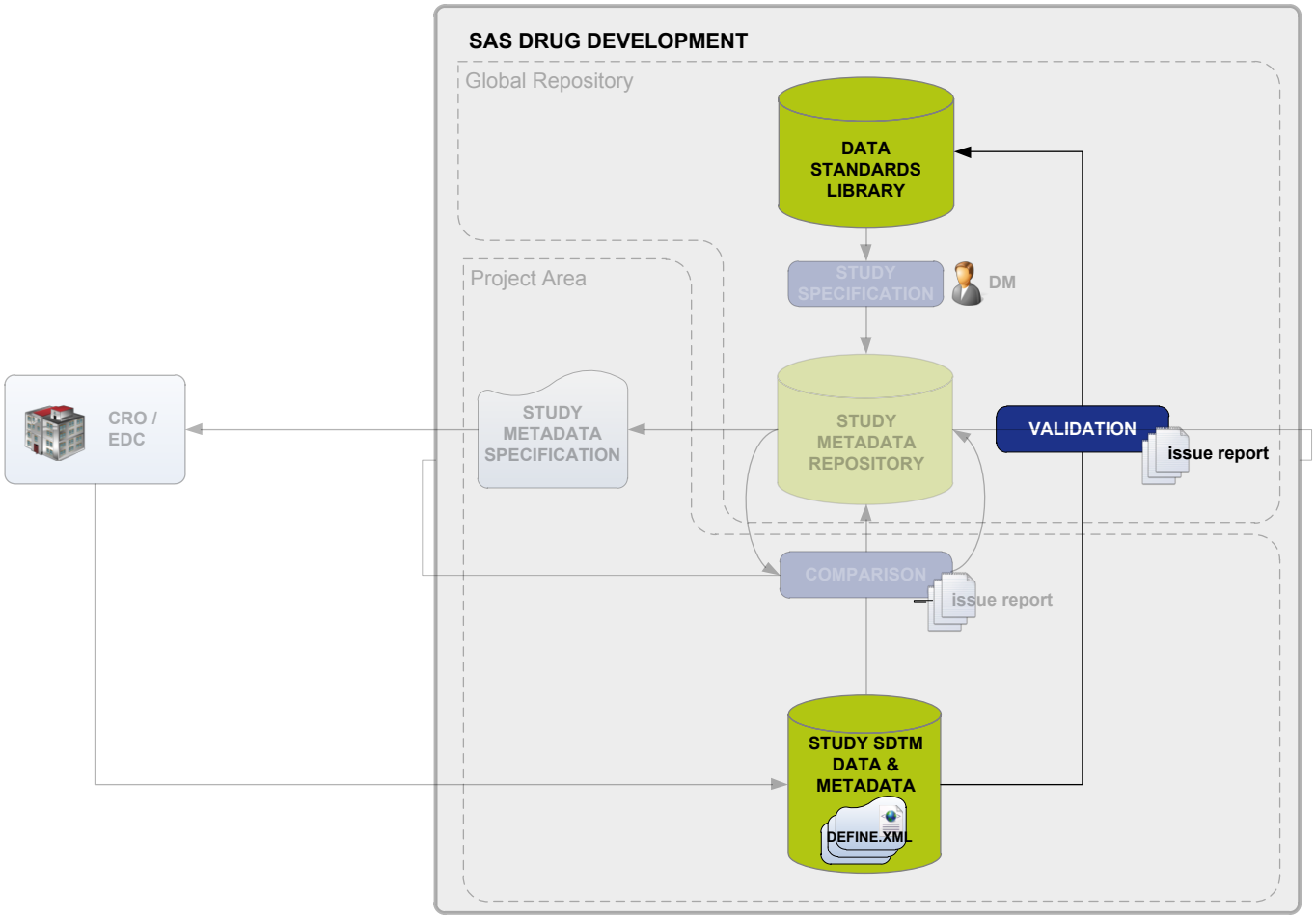
- **4 different statuses :**
  - Ok : no conflict in metadata
  - Difference : difference in the metadata value
  - Missing : metadata not in study data
  - Addition : metadata not in study specifications

Status	Dataset	Variable	*Label*	Specifications	Study
Difference	AE	AEACN	Action Taken with Study Trtment		X
Difference	AE	AEACN	Action Taken with Study Treatment	X	
Ok	AE	AEBODSYS	Body System or Organ Class	X	X
Missing	AE	AECAT	Category for Adverse Event	X	
Addition	AE	AECOM	Comment about Adverse Event		X
Ok	AE	AECONTRT	Concomitant or Additional Trtmnt Given	X	X

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Validation of study data and metadata



Business & Decision Life Sciences

sas THE POWER TO KNOW.

Validation - Checks

SDTM Checks compare the incoming data/metadata against the Data Standards Library

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR5000	All	Identifies domain table that has zero rows and therefore contains no data.	Medium	Null tables are normally not submitted
IR5001	All	Identifies a null (empty) value found in a column where (Standard) Core attribute is 'Req'	High	Required variables may not be null
IR5002	All	Identifies records where the value for a date doesn't conform to the ISO 8601 standard.	Medium	
IR5003	All	Identifies records where the value in the Domain Abbreviation column doesn't match the name of Domain	Low	
IR5004	All	Identifies records where non-unique values for Sequence Number variable exist within a subject	High	
IR5005	EG, LB, QS, VS	Identifies subjects where there are no records with a value of 'Y' in the baseline flag variable (Baseline Flag), excluding Arm Code='SCRNFAIL' or 'NOTASSGN'	Low	Check of Arm Code value is case-insensitive
IR5006	EG, LB, QS, VS	Identifies Measurement, Test or Examination Short Name values where standard units value (Standard Units) is not consistent across all records	Medium	
IR5007	AE	Identifies records where the value for the Preferred Term could not be found in the specified version of the MedDRA dictionary	Medium	
IR5008	AE	Identifies records where Serious Event='Y' but none of Involves Cancer, Congenital Anomaly or Birth Defect, Persist or Signif Disability/Incapacity, Results in Death, Requires or Prolongs Hospitalization, Is Life Threatening, Other Medically Important Serious Event, or Occurred with Overdose equals 'Y'	Medium	
IR5010	All	Identifies records where the value for Visit Number is formatted to more than three decimal places	Medium	
IR5011	DM, TA	Identifies records that violate the condition [If Arm Code='SCRNFAIL' then Description of Arm must equal 'Screen Failure', and vice versa]	Low	Checks of Arm Code and Arm values are case-insensitive
IR5100	Timing	Identifies records that violate the condition [(Study Day of Start of Observation less than or equal to Study Day of End of Observation)], limited to records where [Study Day of Start of Observation is not null and Study Day of End of Observation is not null]	High	When time component is not provided, 00:00:00 is used for Start and 23:59:59 is used for End

Business & Decision Life Sciences

sas THE POWER TO KNOW.

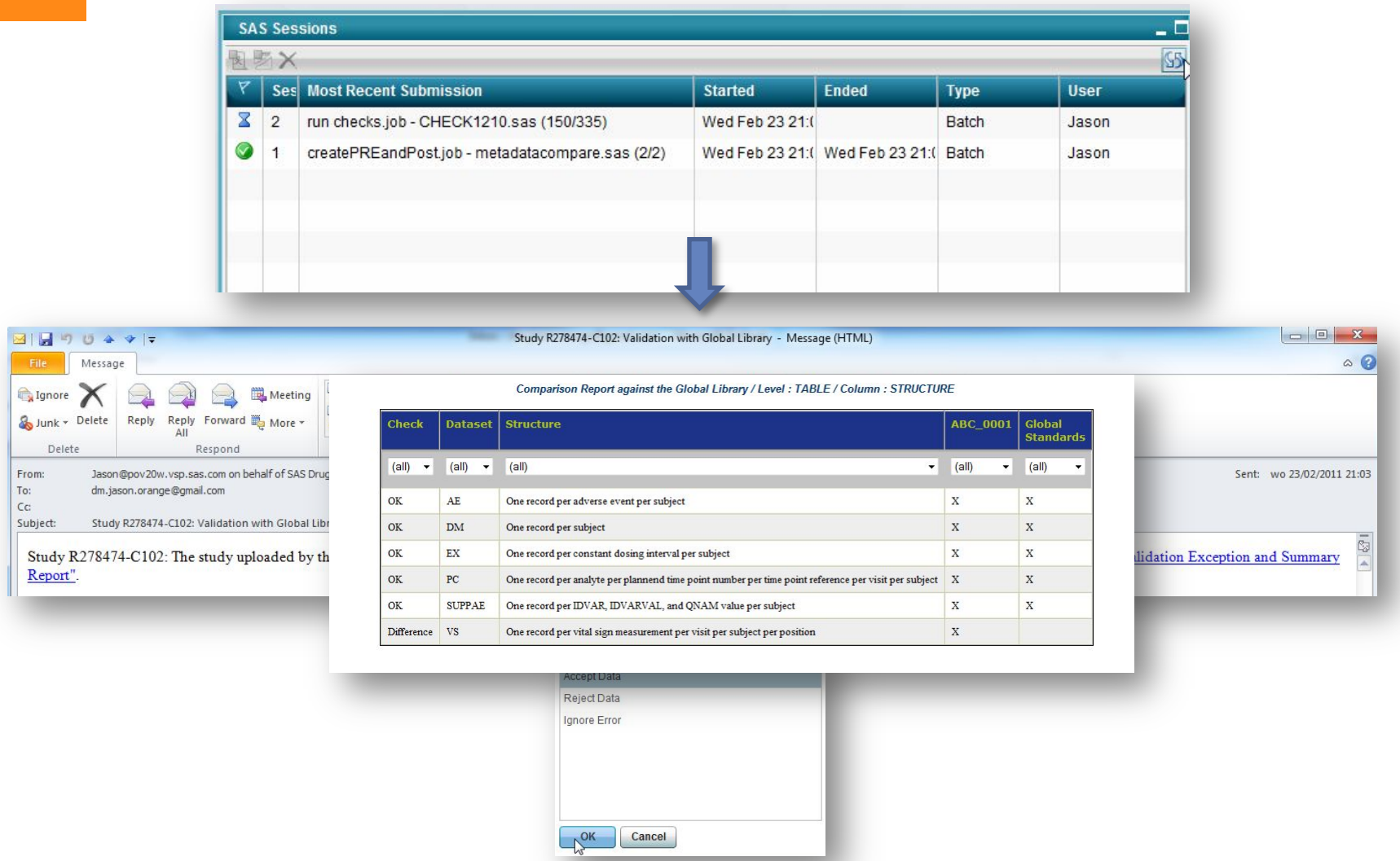
Validation - Reports

CHECKID	ERRORMSG	SEVERITY	STUDYID	DOMAIN	USUBJID	KEY1CD	KEY1	VAR1CD	VAR1	VAR2CD	VAR2
1002	The STANDARD UNIT is not consistent per TEST or EXAMINATION.	High	1000-0000	LB				LBTESTCD	LIPASE	LBSTRESU	%
1002	The STANDARD UNIT is not consistent per TEST or EXAMINATION.	High	1000-0000	LB				LBTESTCD	LIPASE	LBSTRESU	mmol/L
1002	The STANDARD UNIT is not consistent per TEST or EXAMINATION.	High	1000-0000	LB				LBTESTCD	ALKP	LBSTRESU	U
1002	The STANDARD UNIT is not consistent per TEST or EXAMINATION.	High	1000-0000	LB				LBTESTCD	ALKP	LBSTRESU	U/L
1002	The STANDARD UNIT is not consistent per TEST or EXAMINATION.	High	1000-0000	LB				LBTESTCD	LIPASE	LBSTRESU	IU/L
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	AE	1000-0000-00019	AESEQ	1	AESTDTC	1998-12-01T17:00	AEENDTC	1998-12-01T16:30
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	AE	1000-0000-00037	AESEQ	1	AESTDTC	1998-12-15T07:40	AEENDTC	1998-11-25T08:00
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	SE	1000-0000-00004	SESEQ	2	SESTDTC	1998-11-17T08:06:00	SEENDTC	1998-11-17T07:07:00
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	SE	1000-0000-00009	SESEQ	2	SESTDTC	1998-11-24T08:00:00	SEENDTC	1997-11-24T08:01:00
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	SV	1000-0000-00022			SVSTDTC	1998-12-03	SVENDTC	1998-12-02
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	SV	1000-0000-00010			SVSTDTC	1998-11-24	SVENDTC	1998-11-22
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	SV	1000-0000-00014			SVSTDTC	1998-11-16	SVENDTC	1998-10-16
1005	START DATE/TIME OF OBSERVATION falls after the END DATE/TIME OF OBSERVATION.	High	1000-0000	SV	1000-0000-00035			SVSTDTC	1998-12-17	SVENDTC	1997-12-17
1007	The LENGTH of the --TEST variable is more than 40 characters.	High	1000-0000	LB	1000-0000-00001	LBSEQ	35	LBTEST	ERY. MEAN CORPUSCULAR HB CONCENTRATION MEAS	LENGTH	43
1007	The LENGTH of the --TEST variable is more than 40 characters.	High	1000-0000	LB	1000-0000-00030	LBSEQ	54	LBTEST	PARTIAL THROMBOPLASTINE TIME MEASUREMENTS	LENGTH	41
1008	The --TESTCD starts with a number.	High	1000-0000	LB	1000-0000-00001	LBSEQ	3	LBTESTCD	2AMYL		
1009	An ORIGINAL RESULT is completed but the STANDARD RESULT (CHARACTERISTIC) is missing.	High	1000-0000	IE	1000-0000-00006	IESEQ	1	IEORRES	N	IESTRESC	
1009	An ORIGINAL RESULT is completed but the STANDARD RESULT (CHARACTERISTIC) is missing.	High	1000-0000	SC	1000-0000-00031	SCSEQ	1	SCORRES	R-K	SCSTRESC	
1009	An ORIGINAL RESULT is completed but the STANDARD RESULT (CHARACTERISTIC) is missing.	High	1000-0000	SC	1000-0000-00014	SCSEQ	1	SCORRES	W-B	SCSTRESC	
1009	An ORIGINAL RESULT is completed but the STANDARD RESULT (CHARACTERISTIC) is missing.	High	1000-0000	SC	1000-0000-00022	SCSEQ	1	SCORRES	M-U	SCSTRESC	

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Workflow governing validation



Business & Decision Life Sciences

sas THE POWER TO KNOW.

Summary

An innovative set-up is described here that consists of

- upfront clinical study data specifications in SAS Drug Development® (SDD) using CDISC standard data models.
- automation of the exchange of specifications to their internal and external partners.
- a central data standards library and a metadata repository, centrally accessible in SDD
- the processes that govern exchange of study specifications and the subsequent execution of validation and verification SAS programs are enabled by automated and fully secured workflows

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Best practices and advantages

- Best practices
  - Use (CDISC) data standards throughout the data management flow
  - Generate high quality clinical standards data adhering to study specifications and organisational data standards
  - Use SAS Drug Development as a data exchange and computing platform to strictly control and support this process with the required flexibility
- Advantages
  - Reduce the time for data manager to monitor quality of data
  - Enable collaboration with external partners and groups

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Business & Decision Life Sciences

sas THE POWER TO KNOW.

Business & Decision Life Sciences

Tel +32 2 774 11 00
Fax +32 2 774 11 99

info.ls-be@businessdecision.com

Sint-Lambertusstraat 141
1200 Brussels
Belgium

www.businessdecision-lifesciences.com

SAS

Tel +32 2 766 07 00
Fax +32 2 766 07 77

info@sbx.sas.com

Hertenbergstraat 6
3080 Tervuren
Belgium

www.sas.com/industry/pharma