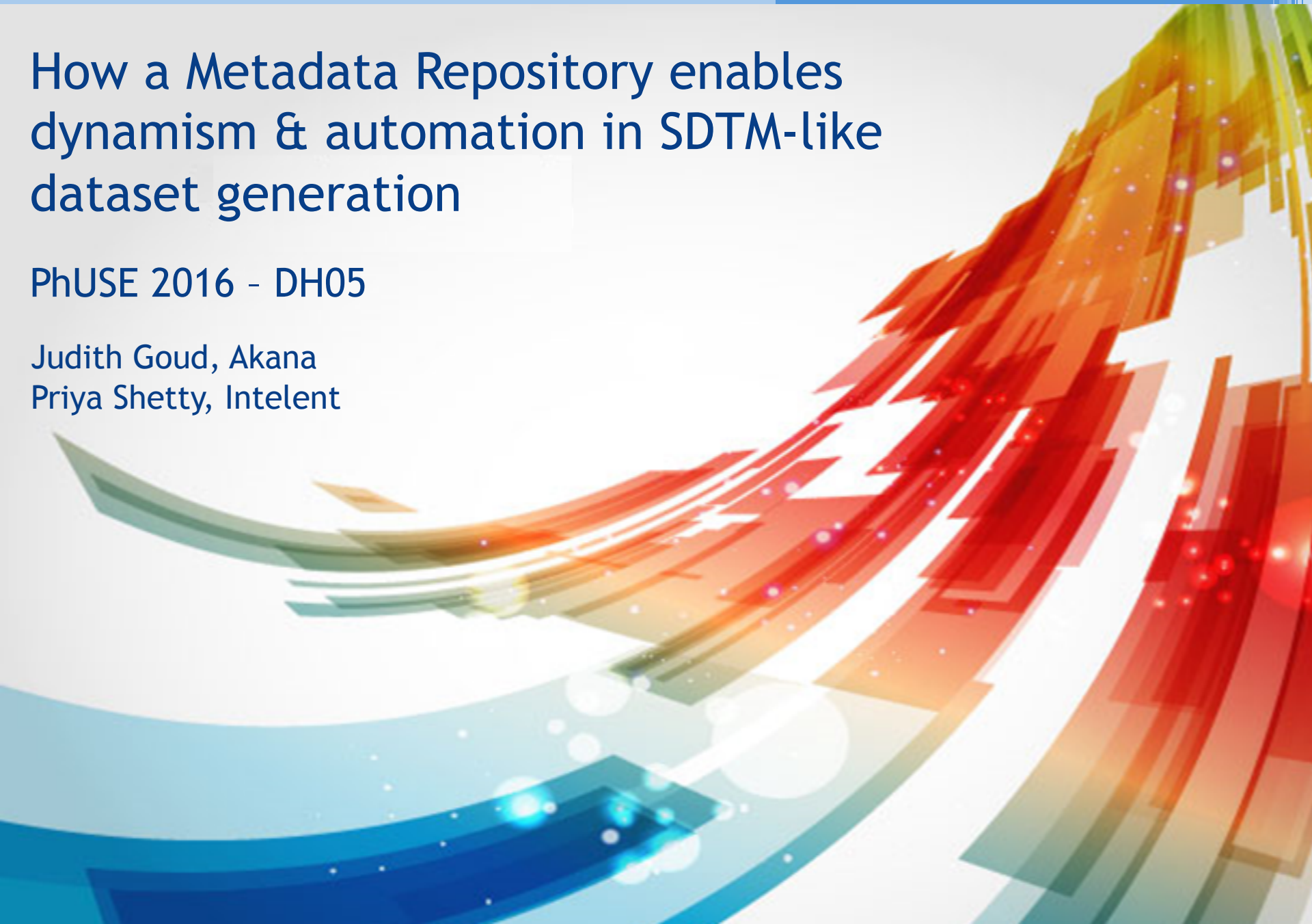


How a Metadata Repository enables dynamism & automation in SDTM-like dataset generation

PhUSE 2016 - DH05

Judith Goud, Akana
Priya Shetty, Intelent



Outline

- 
- Trends - Dynamism and automation in data integration
 - Business Processes - Traditional and Enhanced Approaches
 - Transformational Metadata Management
 - Metadata Driven Clinical Data Flow
 - Use Case Demo
 - Metadata Storage and Extraction
 - Dynamic Program Execution
 - Target Dataset Generation
 - Source to Target Transformations
 - Conclusion

Trends - Dynamism and Automation in Integration

Drivers

- Source, target and transformation metadata vary across studies
- Physical storage, data transfer mechanism and relevant data dictionaries may vary across studies
- Important to minimize code change, transformation pre-processing in ingestion layer
- Dynamic transformation rules needed
- Cost and time efficiency, adoption within enterprise

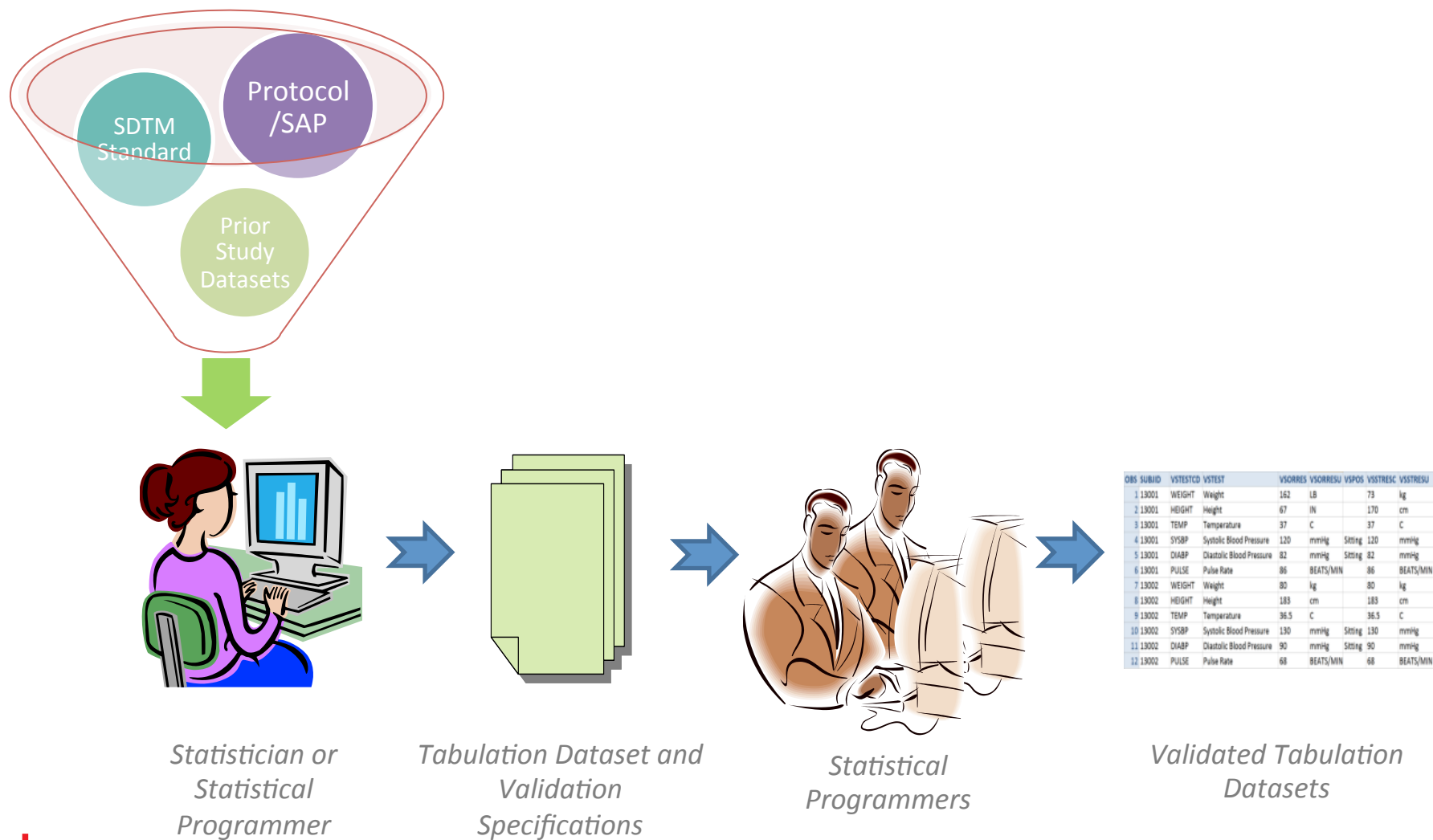
Dynamism

- Key is to minimize code change
 - Store structures in appropriate hierarchy and data lifecycle stages
 - Plan either by leveraging existing metadata or manufactured metadata
 - A robust user interface or means of configuration can address gaps

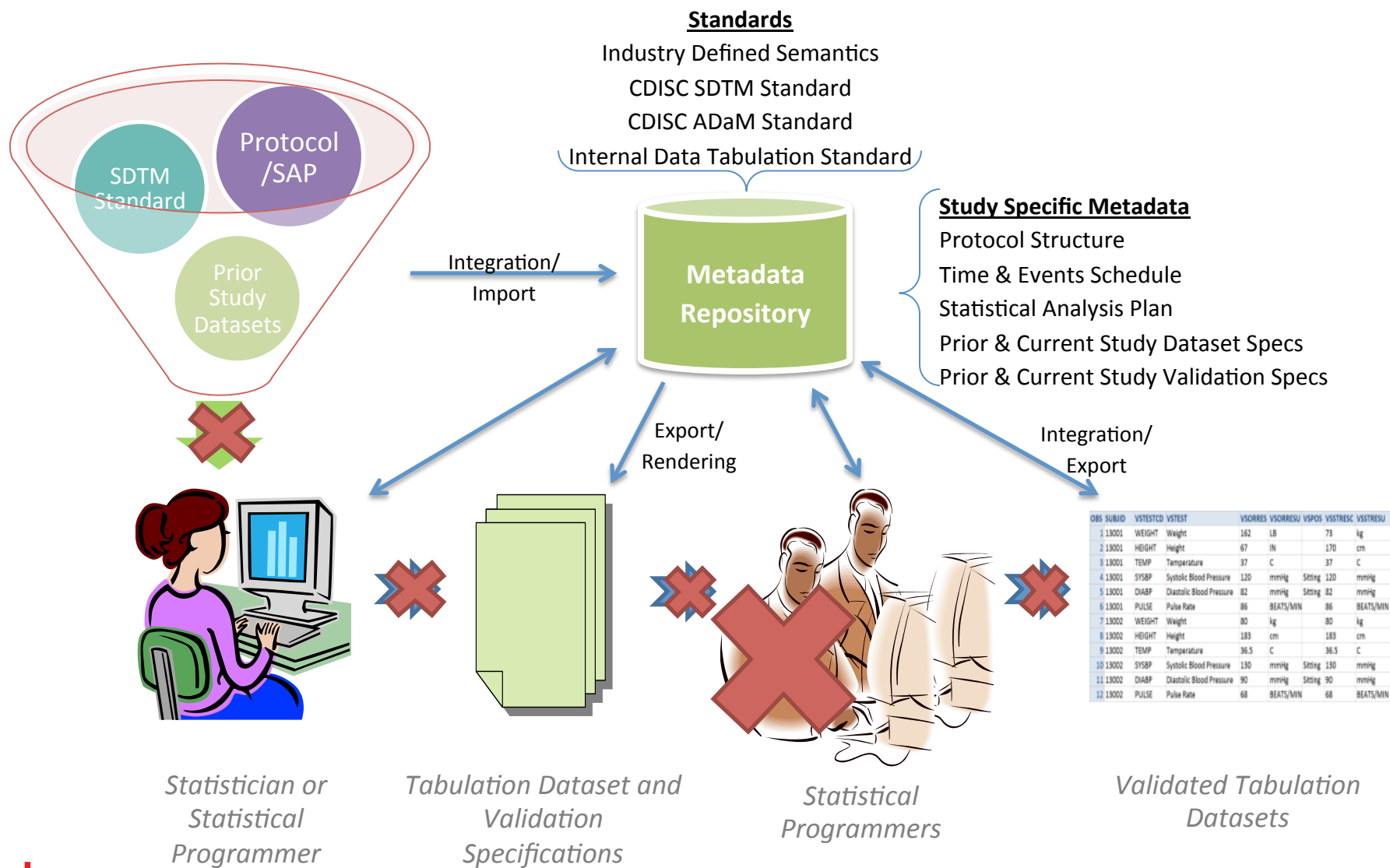
Automation

- Key is to automate data delivery in usable format with minimal manual intervention
 - High availability of data to points of analysis
 - From disparate sources: raw, integrated, cleansed/ not cleansed, reconciled, aggregated, etc.
 - Based on use cases: interim analysis, submission, central monitoring, etc.

Traditional Business Process



Enhanced Business Process



Transformational Metadata Management

Published Asset: AEENDTC ([Pilot SDTM] R 1)

Asset Type: Rule

Overview

Asset Context

Classifiers

Name	Value
Lifecycle Status	Approved
Asset Owners	support
Business Object Type	Transformation
Endorsement Level	Global
Clinical Lifecycle Stage	Data Tabulations

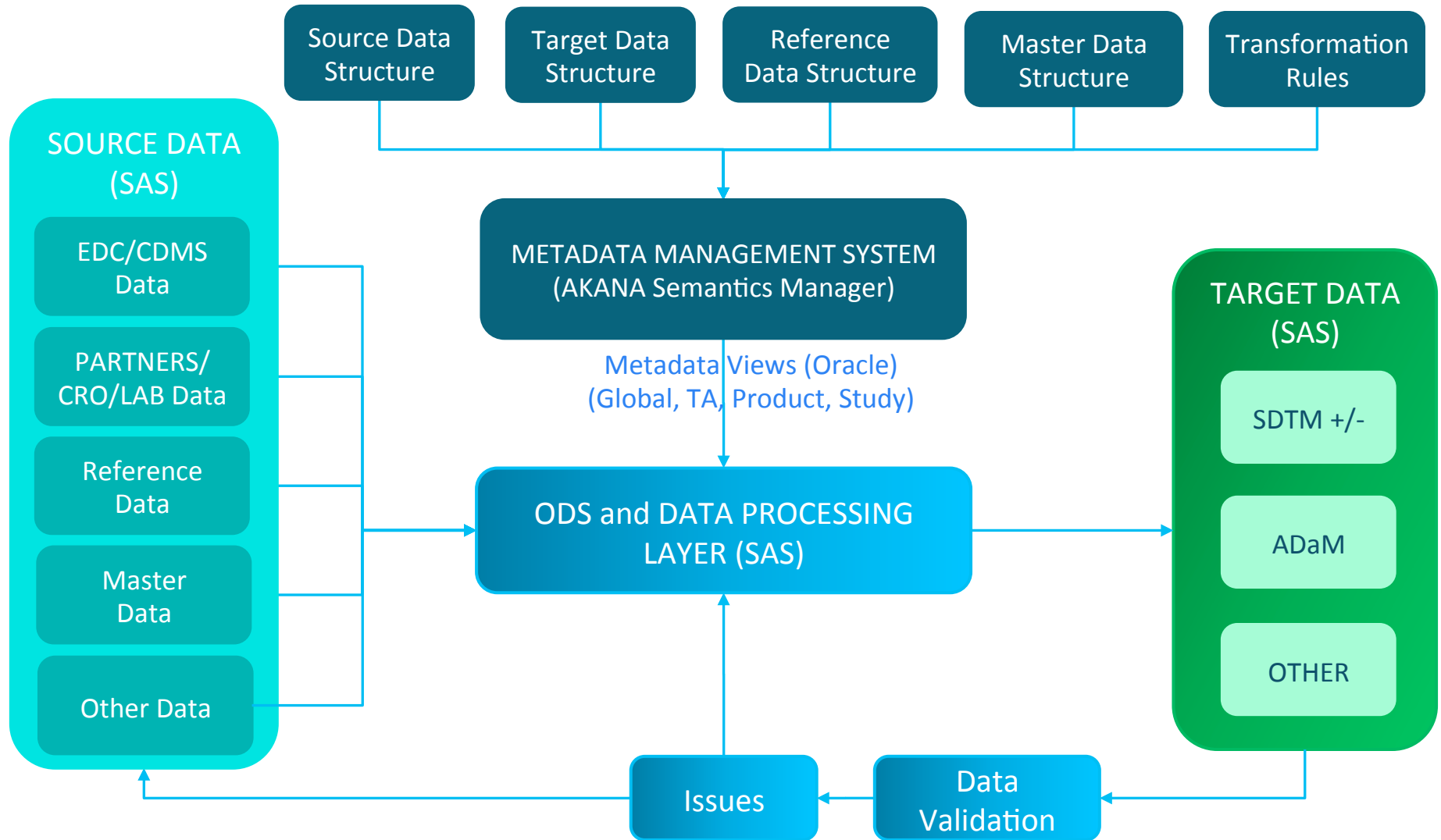
Artifacts

Name	Value
Programming Notes	Convert to ISO8601 Datetime format
SAS Derivation	PUT(AE.AEENDAT,'YYYY-MM-DD') 'T' PUT(AE.AEENTIM,'hh:mm:ss')

Relationships

Relationship	Asset Type	Asset Name
Encompassed By Rule	Rule	AE_ORDER (R 1)
Input Elements	Metadata Element	AEENDAT ([Pilot Raw] MDE 1)
Input Elements	Metadata Element	AEENTIM ([Pilot Raw] MDE 1)
Produces Asset	Metadata Element	AEENDTC ([Pilot SDTM] MDE 1)

Metadata Driven Clinical Data Flow



Source and Target Metadata extracted from Semantics Manager

DOMAIN	VARIABLE	VARORD	LABEL	TYPE	SORTORD	SOURCE_DOMAIN	SOURCE_ELEMENT	DERV_CODE	DERIVATION	EXEC_ORD	PARAMVAL	LENGTH	USED
Domain Prefix	Variable Name	Seq. For Order	Variable Label	Type	Sort order	SOURCE DOMAIN	SOURCE FIELD	TRANSFORMATION (SAS)	Derivation	Order of Execution	Derivation Parameters	Length	Used for study
AE	AEYN	1	AE Occurred?	Char		AE	AEYN		SUBSET	2	AEYN="Y"	1	Y
AE	STUDYID	2	Study Identifier	Char	1	AE	STUDYID		COPY	1		20	Y
AE	DOMAIN	3	Domain Abbreviation	Char					ASSIGN	3	"AE"	2	Y
AE	USUBJID	4	Unique Subject Identifier	Char	2	AE	STUDYID, SUBJID		CONCAT		"."	20	Y
AE	AESEQ	5	Sequence Number	Num		AE			SEQUENCE	99	Concatenate using Separator	8	Y
AE	AEGRPID	6	Group ID	Char		AE	AEGRPID		COPY	1		60	N
AE	AEREFID	7	Reference ID	Char								20	N
AE	AESPID	8	Sponsor-Defined Identifier	Char	5	AE	AESPID		COPY	1		10	Y
AE	AEPRESP	9	Pre-Specified Adverse Event	Char		AE	AEOCCUR	IF AEOCCUR IN("Y", "N") THEN AEPRESP="Y"	DERIVE	5		1	Y
AE	AEOCCUR	10	specific adverse event	Char		AE	AEOCCUR	IF AEOCCUR NOT IN("Y", "N") THEN AEOCCUR="N"	DERIVE	6		1	Y
AE	AETERM	11	Reported Term for the Adverse Event	Char	3	AE	AETERM		COPY	1		200	Y
AE	AEMODIFY	12	Modified Reported Term	Char			AEMODIFY		COPY	1		200	Y
AE	AELLT	13	Lowest Level Term	Char		AE,reflib.Meddra11_0	LLT_NAME		MERGE	14	AELLTCD#LLT_CODE	200	Y
AE	AELLTCD	14	Lowest Level Term Code	Num		AE	AELLTCD		COPY	1		8	Y
AE	AEDECOD	15	Dictionary-Derived Term	Char		AE,reflib.Meddra11_0	PT_NAME		MERGE	15	AELLTCD#LLT_CODE	200	Y
AE	AEPTCD	16	Preferred Term Code	Num		AE,reflib.Meddra11_0	PT_CODE		MERGE	16	AELLTCD#LLT_CODE	8	Y
AE	AEHLT	17	High Level Term	Char		AE,reflib.Meddra11_0	HLT_NAME		MERGE	17	AELLTCD#LLT_CODE	200	Y
AE	AEHLTCD	18	High Level Term Code	Num		AE,reflib.Meddra11_0	HLT_CODE		MERGE	18	AELLTCD#LLT_CODE	8	Y
AE	AEHLGT	19	High Level Group Term	Char		AE,reflib.Meddra11_0	HLGT_NAME		MERGE	19	AELLTCD#LLT_CODE	200	Y
AE	AEHLGTC	20	High Level Group Term Code	Num		AE,reflib.Meddra11_0	HLGT_CODE		MERGE	20	AELLTCD#LLT_CODE	8	Y

Structural and Transformational Metadata extracted from SM drives dynamic program for generating target datasets

Derivation Types

Derivation order of execution

Dynamic Program Execution to generate Target Datasets

SAS System Viewer - [sdtm_plus.sas]

File Edit View Window Help

Program : sdtm.sas
Purpose : Create SDTM datasets from source using metadata
Project : AKANA Mapping
Programmer : XXXXXXXX
Creation Date : 28-SEP-2015
Parameters :
Comments :
Modifications :

%include "F:\Project1\util_setup.sas";

*** Call derivation engine to derive SDTM+ datasets;
%derivationengine(domain=ae);



Execute Dynamic Program as Batch or Individual Mode

SASView

Batch Submit with SAS 9.2

Open

Open with SAS 9.2

Print

Print with SAS

macros

File Edit View Favorites Tools Help

Back Search Folders

Address F:\CDODATA\Akana\util\macros

Folders	Name	Size	Type	Date Modified
Akana	assign.sas	2 KB	SAS Program	9/11/2015 4:04 AM
adam	concatvar.sas	2 KB	SAS Program	9/23/2015 3:44 AM
sdtm	copyvar.sas	3 KB	SAS Program	9/11/2015 4:35 AM
sdtmplus	DerivationEngine.sas	4 KB	SAS Program	9/26/2015 3:44 AM
sdtmplus_pgm	derive.sas	2 KB	SAS Program	9/11/2015 4:09 AM
source	dtt.sas	2 KB	SAS Program	9/23/2015 2:55 AM
util	merge.sas	3 KB	SAS Program	9/26/2015 2:37 AM
global_lib	sequence.sas	2 KB	SAS Program	9/26/2015 2:44 AM
macros	studyday.sas	3 KB	SAS Program	9/23/2015 2:54 AM
others	subset.sas	2 KB	SAS Program	9/11/2015 3:53 AM
reference				

SAS Macros corresponding to SM transformational metadata

Metadata driven Source to Target Transformation Examples

ae.sas7bdat

	AEYN	STUDYID	SUBJID	USUBJID	VISIT	AESPID	AETERM	AEMODIFY	AELLTCD	AEPT	AESOC	AESTDAT	AESTTIM	AEENDAT	AEENTIM	AEONGO	AETOXGR	A
1	N	ABC	1001	ABC01001	ADVERSE E													
2	Y	ABC	1002	ABC01002	ADVERSE E 1		Headache		10019211	Headache	Nervous s	2-Jan-04	12:12	2-Jan-04	17:34		GRADE 1	MOD
3	Y	ABC	1002	ABC01002	ADVERSE E 2		Vomit	Vomiting	10047700	Vomiting	Gastroint	4-Feb-04				Y	GRADE 2	SEV
4	Y	ABC	1002	ABC01002	ADVERSE E 3		Anxiety		10002865	Anxiety	Psychiatr	4-Feb-04				Y	GRADE 2	SEV
5	Y	ABC	2001	ABC02001	ADVERSE E 1		Headache		10019211	Headache	Nervous s	2-Jan-04	12:12	2-Jan-04	17:34		GRADE 1	MOD
6	Y	ABC	2001	ABC02001	ADVERSE E 2		Vomit	Vomiting	10047700	Vomiting	Gastroint	4-Feb-04				Y	GRADE 2	SEV
7	Y	ABC	2001	ABC02001	ADVERSE E 3		Alveolitis		10000001	Alveoliti	Respirato	4-Apr-04				Y	GRADE 2	SEV

Concatenate using Separator

MedDRA merge

ISO date conversion

ae.sas7bdat

	AEYN	DOMAIN	USUBJID	AESQ	AESPID	AETERM	AEMODIFY	AELLT	AELLTCD	AEDECD	AEHLT	AEHLTCD	AEBODSYS	AEBODSYCD	AESV	AESTDTC	AEENDT
1	Y	AE	ABC-1002	1 3		Anxiety		Anxiety	10002865	Anxiety	Anxiety s	10002869	Psychiatr	10037175	SEVERE	2004-02-04	
2	Y	AE	ABC-1002	2 1		Headache		Headache	10019211	Headache	Headaches	10019233	Nervous s	10029205	MODERATE	2004-01-02T12:12	2004-01
3	Y	AE	ABC-1002	3 2		Vomit	Vomiting	Vomiting	10047700	Vomiting	Nausea an	10028817	Gastroint	10017947	SEVERE	2004-02-04	
4	Y	AE	ABC-2001	1 3		Alveolitis		"Ventilat	10000001	Alveoliti	Allergic	10027654	Immune sy	10021408	SEVERE	2004-04-04	
5	Y	AE	ABC-2001	2 3		Alveolitis		"Ventilat	10000001	Alveoliti	Lower res	10024972	Respirato	10038738	SEVERE	2004-04-04	
6	Y	AE	ABC-2001	3 1		Headache		Headache	10019211	Headache	Headaches	10019233	Nervous s	10029205	MODERATE	2004-01-02T12:12	2004-01
7	Y	AE	ABC-2001	4 2		Vomit	Vomiting	Vomiting	10047700	Vomiting	Nausea an	10028817	Gastroint	10017947	SEVERE	2004-04-04	

Update transformation metadata in SM (Add a hyphen) for USUBJID rule – automated update in target data

Merge with MedDRA dictionary populating required submission variables in target

Source Date Update reflected ISO 8601 format in Target

Conclusion

The use of a centralized repository to manage and govern metadata and standards in a flexible source agnostic data integration platform provides the following advantages:

- High availability of data (source, integrated, and standardized) facilitates rapid insights and accelerated timelines for submissions
- Dynamic and automated generation of datasets across studies with minimal customization necessary results in cost and time savings
- Accelerated response to metadata update drivers like new versions of standards and/or new regulations facilitate regulatory compliance
- Reusability of standard algorithms result in enhanced efficiencies for validation
- Dynamic and automated processes driven by metadata limit the need of change control
- Full traceability for audit tracking provides controlled standards management

Q&A and Contact



Judith Goud

Senior Solution Architect
BioPharma Europe
6-9 The Square, Stockley Park,
Uxbridge, UB11 1 FW, UK

judith.goud@akana.com

Mobile: +31 6 1241 7026

www.akana.com



Priya Shetty

Director
500 Alexander Park,
Princeton, NJ 08540
USA

priya.shetty@intelent.com

Mobile: +1 646 637 8583

www.intelent.com