



PhUSE Annual Meeting, London 2014

**Metadata, Study data standards, Master data,
terminology, interoperability:**
Key concepts underlying compliance to FDA
guidance on electronic submission

Isabelle de Zegher, MD, MSc,
PAREXEL Informatics

Table of Content

- PhUSE CSS – Emerging Technology Working group
- Definitions
 - Metadata and study data standards
 - Master data
 - Controlled terminology
 - Interoperability
 - Pooling, aggregation, integration
- How can this help toward eSubmission
 - Lessons learned
 - Changing landscape: need for concept based MDR from data collection onwards

Table of Content

- PhUSE CSS – Emerging Technology Working group
- Definitions
 - Metadata and study data standards
 - Master data
 - Controlled terminology
 - Interoperability
 - Pooling, aggregation, integration
- How can this help toward eSubmission
 - Lessons learned
 - Changing landscape: need for concept based MDR from data collection onwards

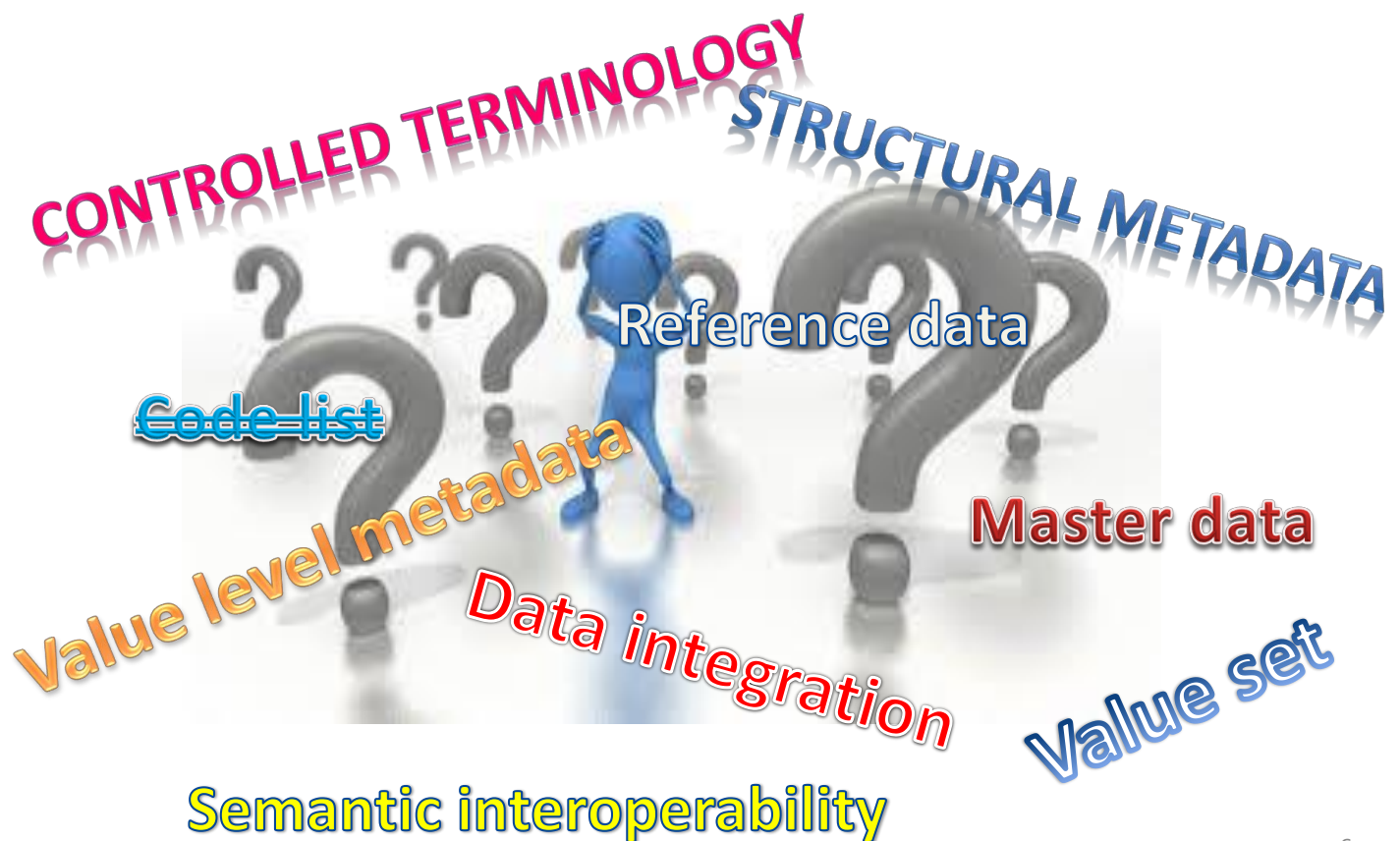
PhUSE CSS Emerging Technology

- FDA/PhUSE Computational Science Symposium (CSS) is a collaborative effort between industry and the FDA to work on implementation of data standard
- In 2013 we launch a new working group related to specific computational science topics, tools, technologies, and approaches.
 - semantic web applications (now in a dedicated working group)
 - **analysis metadata,**
 - cloud computing
 - (big data)

Table of Content

- PhUSE CSS – Emerging Technology Working group
- Definitions
 - Metadata and study data standards
 - Master data
 - Controlled terminology
 - Interoperability
 - Pooling, aggregation, integration
- How can this help toward eSubmission
 - Lessons learned
 - Changing landscape: need for concept based MDR from data collection onwards

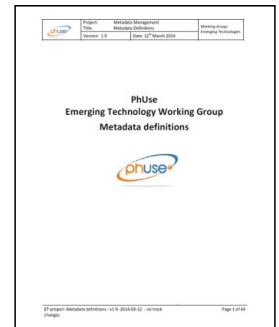
Metadata definition project:
the data standards soup....



Participants

- Mitra Rocca (co-chair) Mitra.rocca@fda.hhs.gov
- Marcelina Hungria (co-chair) mhungria@dicoregroup.com
- Yun Oldshue yun.oldshue@takeda.com
- Kenneth Stoltzfus kenneth.m.stoltzfus@accenture.com
- Julie James julie_james@bluewaveinformatics.co.uk
- Tim Church tim.church@torch.uk.net
- Gregory Steffens Gregory.steffens@novartis.com
- Praveen Garg Praveen.Garg@iconplc.com
- John Leveille jleveille@d-wise.com
- Aimee Basile abasile@celgene.com
- Sam Hume shume@cdisc.org

Definitions in scope



4.1 METADATA MANAGEMENT

- 4.1.1 *Metadata*
- 4.1.2 *Structural metadata*
- 4.1.3 *Descriptive metadata*
- 4.1.4 *Study Instance Metadata*
- 4.1.5 *Metadata repository*
- 4.1.6 *Metadata registry*
- 4.1.7 *Data element*
- 4.1.8 *Attribute*
- 4.1.9 *Class*
- 4.1.10 *Data type*
- 4.1.11 *Value level metadata*

4.2 CONTROLLED TERMINOLOGY, CODE SYSTEMS & VALUE SETS

- 4.2.1 *Controlled Terminology/controlled vocabulary*
- 4.2.2 *Code system*
- 4.2.3 *Dictionary*
- 4.2.4 *Concept*
- 4.2.5 *Code*
- 4.2.6 *Code list*
- 4.2.7 *Value set*

4.3 MASTER DATA MANAGEMENT

- 4.3.1 *Master Data*
- 4.3.2 *(Master) Reference Data*
- 4.3.3 *Master Data Management*

4.4 INTEROPERABILITY

- Categorization of Interoperability (by HL7)*
- 4.4.1 *Technical interoperability ("machine interoperability")*
- 4.4.2 *Semantic interoperability*
- 4.4.3 *Process Interoperability*

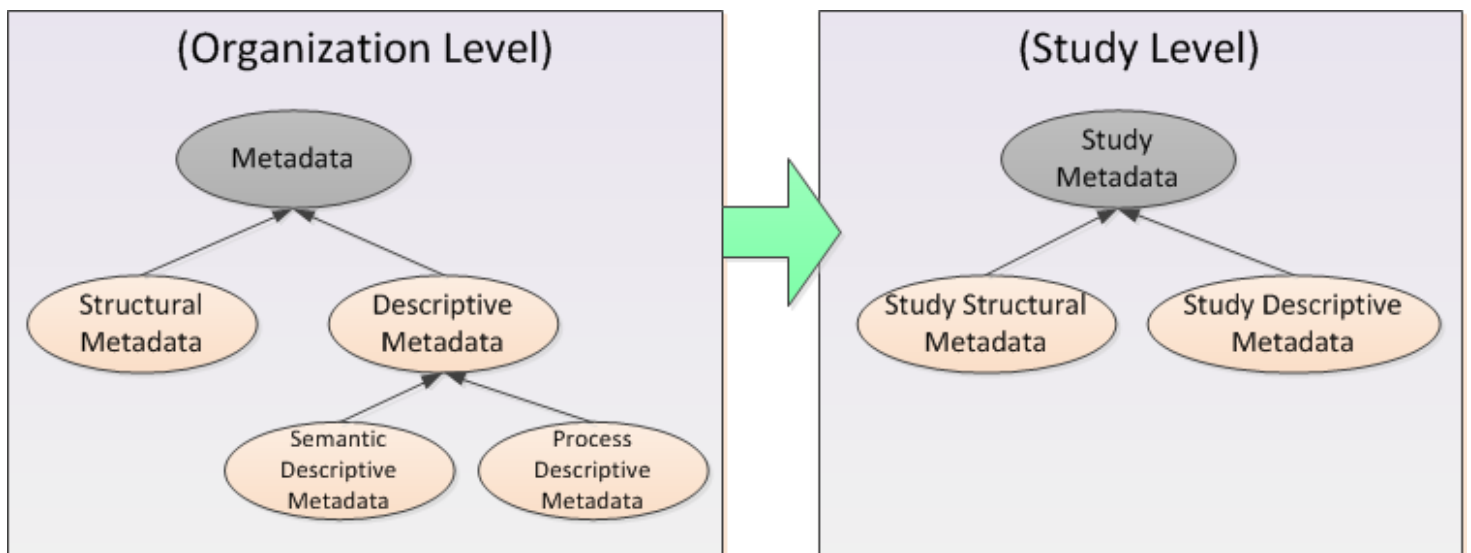
4.5 DATA AGGREGATION, INTEGRATION, POOLING

- 4.5.1 *Data pooling*
- 4.5.2 *Data integration*
- 4.5.3 *Data aggregation*

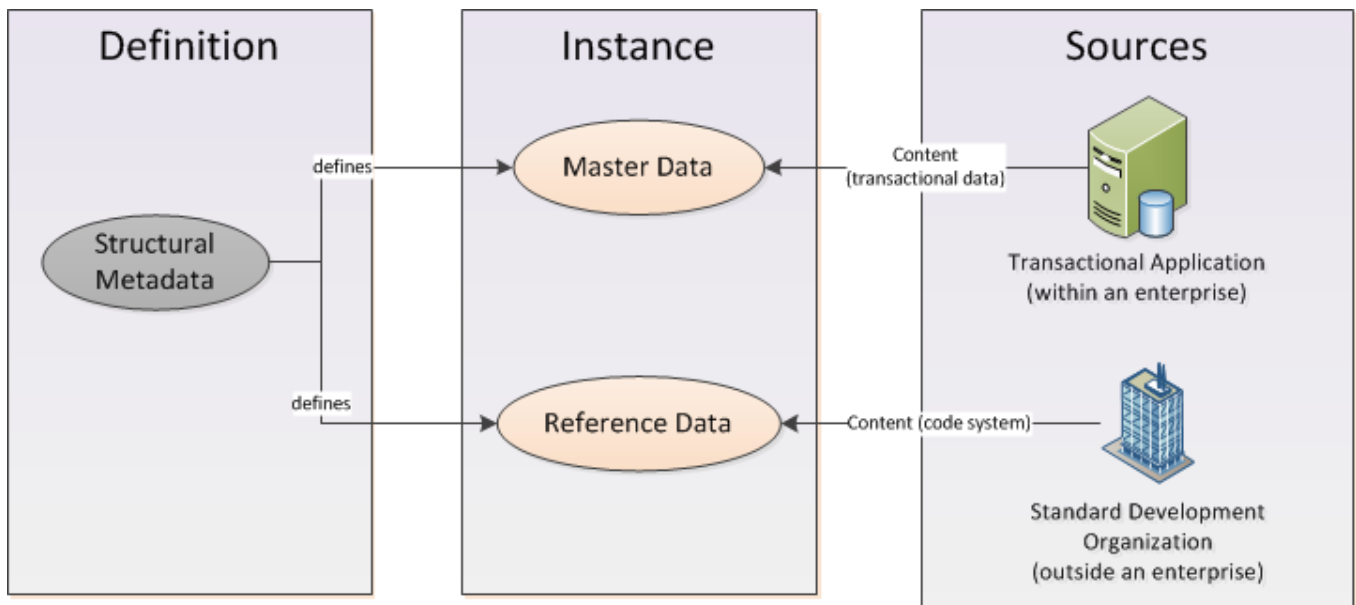
Approach

Synonym	Reference Data Management; MDM
Definition & source	[Gartner – Magic Quadrant for Master Data Management of Customer Data Solution] http://www.gartner.com/technology/reprints.do?id=1-1CK9UDO&ct=121019&st=sb MDM is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official, shared master data assets. [Source: Master Data Management] Master Data Management (MDM) is the collective application of governance, business processes, policies, standards and tools facilitate consistency in data definition. <i>The idea of Master Data focuses on providing unobstructed access to a consistent representation of shared information [Source: SAS White Paper on Supporting Your Information Strategy with a Phased Approach to Master Data Management]</i>
Description	Master Data Management (MDM) comprises of a set of processes and tools that consistently define and manage the master data and master reference data of an enterprise, which are fundamental to the company's business operations. MDM has the objective of providing processes & tools for collecting, aggregating, matching, consolidating, quality-assuring, persisting and distributing such data throughout an organization to ensure consistency and control in the ongoing maintenance and application use of this information. There are different models for master data management – the 2 main extremes are • Centralized model – where all data are managed within a central data store and pushed to the different applications within an organization. • Decentralized model (registry) where the master data are managed within each applications but then reconciled through a registry systems to federate.
Example	Specific products from vendors such as INFORMATICA, IBM, Software AG,...
Recommended definition	Set of processes and tools needed for the deployment of master data and master reference data within an organization.

Metadata

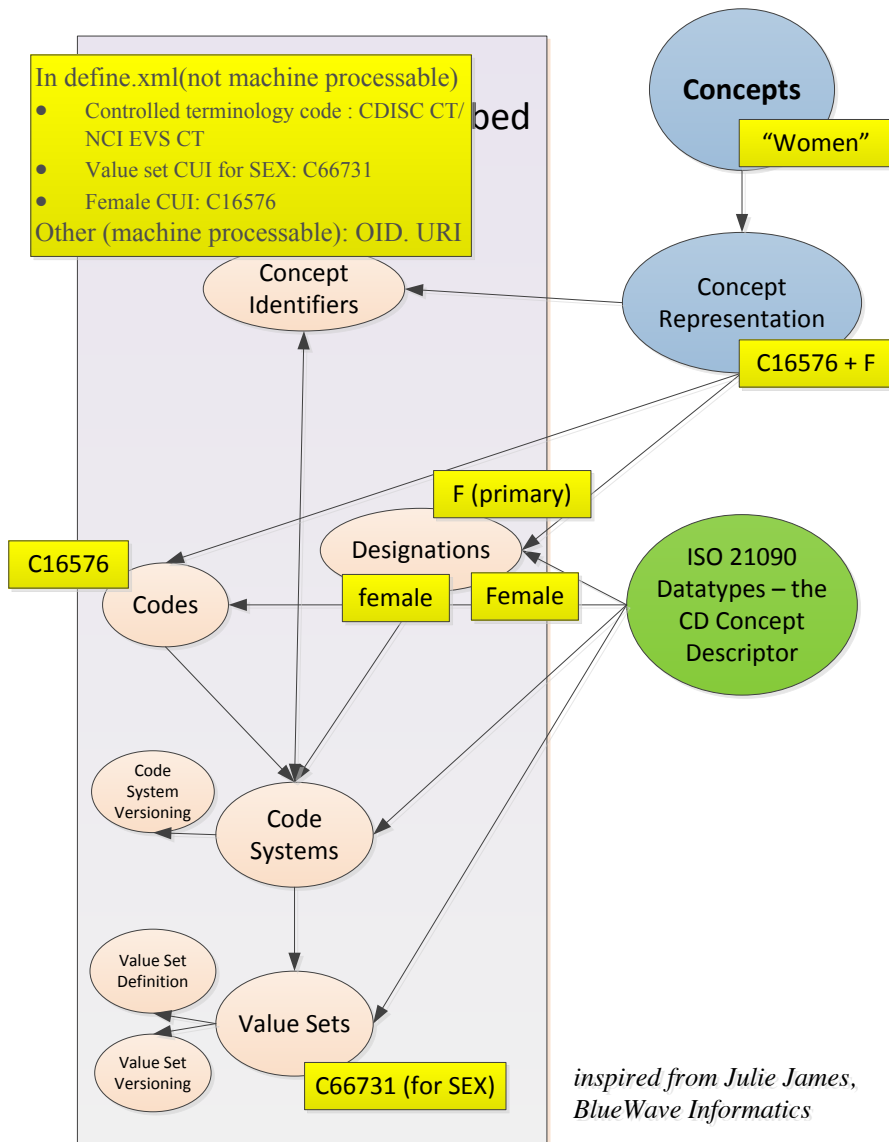


Master Data

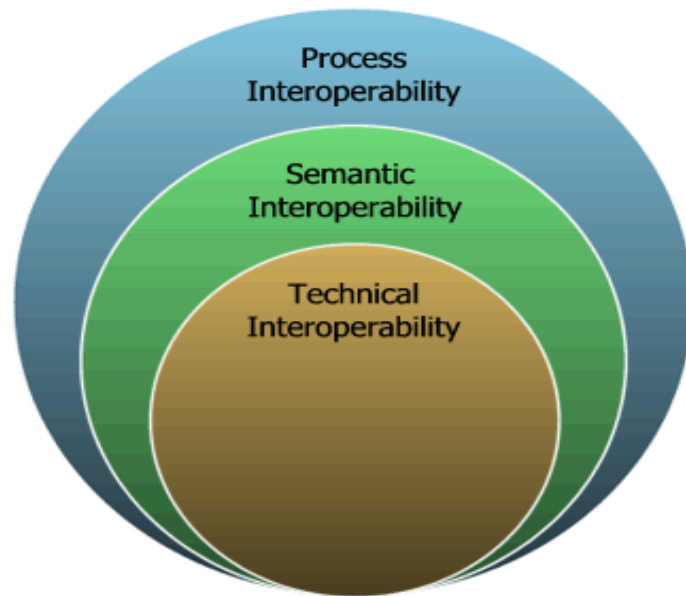


Controlled Terminology

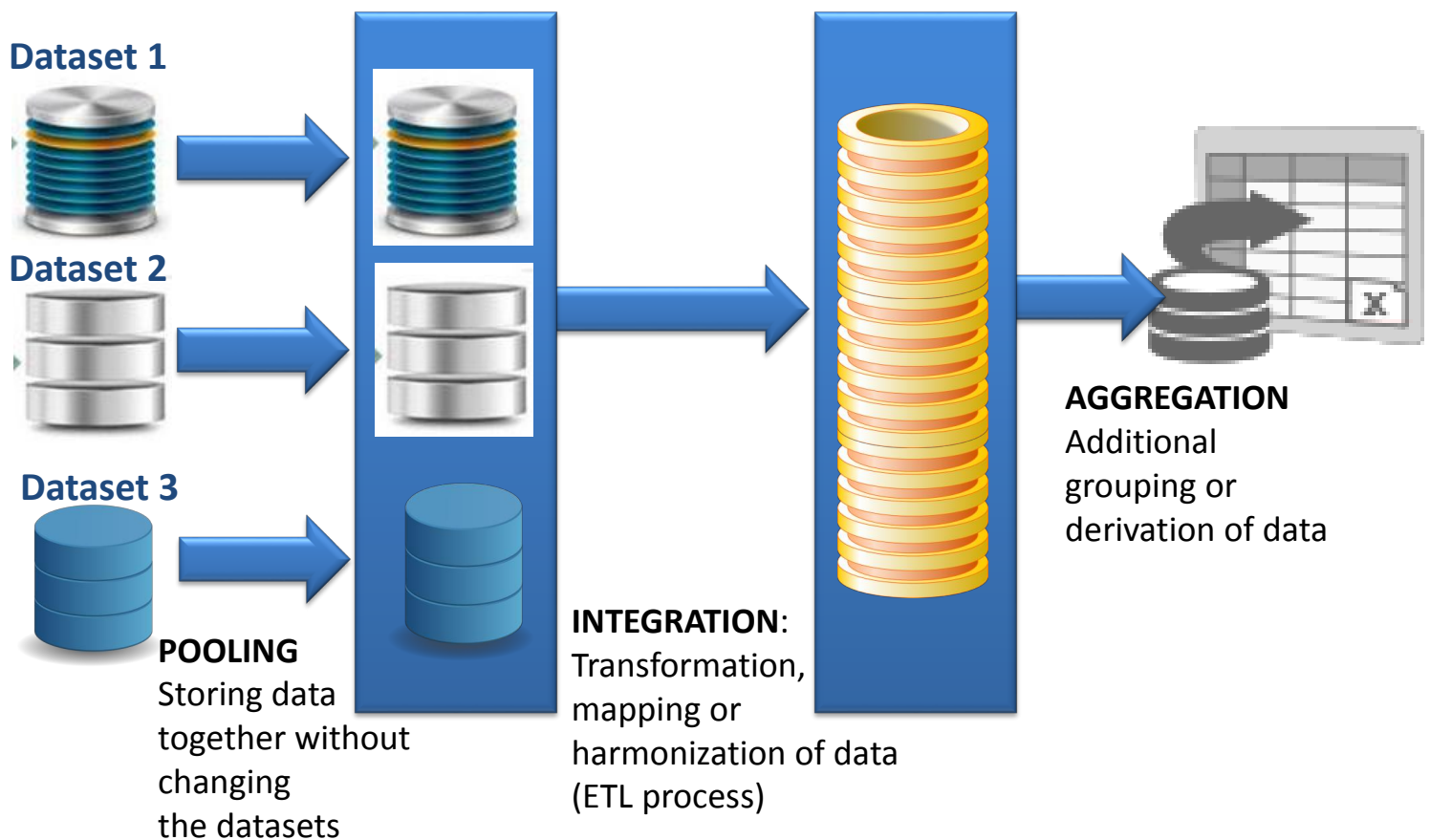
~~“Code list”~~
value set
with CDISC
example



Interoperability



Data Pooling, Integration, Aggregation



Next steps

- **Comments welcome**
- **Consolidation**

http://www.PhUSEwiki.org/wiki/index.php?title=Metadata_Management

Table of Content

- PhUSE CSS – Emerging Technology Working group
- Definitions
 - Metadata and study data standards
 - Master data
 - Controlled terminology
 - Interoperability
 - Pooling, aggregation, integration
- How can this help toward eSubmission
 - Lessons learned
 - Changing landscape: need for concept based MDR from data collection onwards

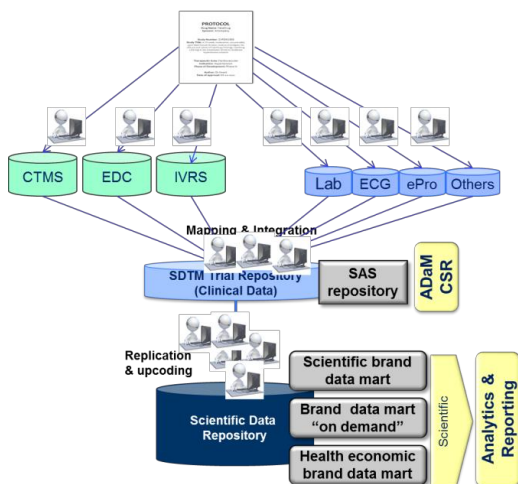
LESSONS LEARNED AT PAREXEL INFORMATICS

- **Efficient Data Integration and compliance to regulatory standards does not start after pooling (retroactive approach) ; it starts with the protocol (proactive approach)**
- **A proactive approach is based on 2 components**
 - Agreement on study “Master” Data (study ID, visit ID, ..) and related descriptive metadata
 - Definition of study structural metadata – aka study specific data standards – as a subset of the enterprise wide variables and value sets contained in a Metadata repository (MDR)
- **To be manageable, variables in an MDR need to be grouped in semantically meaningful “clinical research concepts” (CRC)**

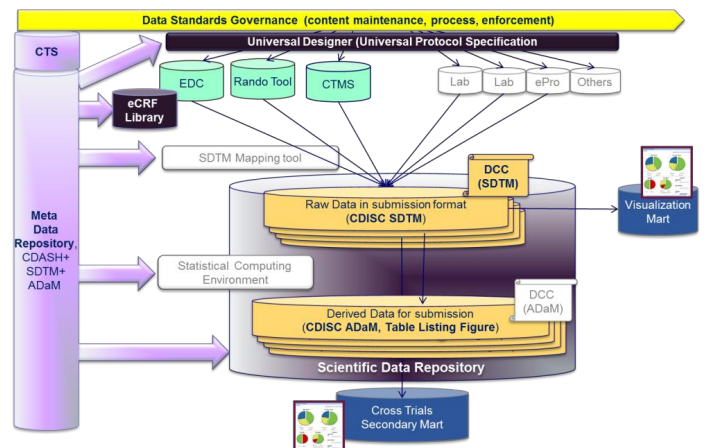
CHANGING LANDSCAPE :

Enforcing data standards from protocol onwards

Retro-active approach from paper protocol,



Pro-active approach with structural metadata



- Different interpretations of same protocol
- Limited standards
- Time to build integrated SDTM data sets

- One single interpretation of protocol
- Increased efficiency, consistency & quality through standards
- Reduced time for integration and secondary data use

CONCEPT BASED MDR :

Protocol is not about variables but about concepts

Period	Screening					
Week	-2	0	1	2	3	
Visit	1	2	3	4	5	
Type of contact	Visit	Visit	phone call	Visit	phone call	
Inclusion and Clinical evaluation:						
Informed Consent	X					
Patient Demography	X					
Medical and surgical history	X					
Physical Examination	X					
Vital Signs and Body Weight	X	X				
Previous and Concomitant Medication	X	X	X	X	X	X
Inclusion and Exclusion Criteria	X	X				
Randomization		X				
Central Laboratory Testing:						
HbA1c		X*				
Lipid Profile ¹ (in fasting state)		X*				
Biochemistry ¹		X*				
Serum Pregnancy Test		X*				
Treatment:						
Dispensation of insulin glargine or premixed insulin and corresponding training		X		X		
Dispensation of insulin glulisine if needed (see section 9.1.3) and corresponding training						
Accountability of Investigational Products		X		X		
Plasma Glucose Self-Monitoring:						
Dispensation of plasma glucose meter	X					
Dispensation and collection of diaries	X	X		X		
Training of the patient (PG meter, diary)	X	X				
Plasma glucose values of 3 consecutive days in the week before the visit				X		
7-point plasma glucose profile of 3 consecutive days in the week before the visit		X				
Insulin dose(s) of the day before the visit			X	X	X	X
Efficacy and Safety:						
Treatment Satisfaction Questionnaires	X					
Hypoglycemia						
AE / SAE						

Annotated eCRF for Patient Demography

Visit: Date of visit:

Demography

Date of birth: Age (years):

Sex:

Predominant race:

Ethnicity: Specify:

Source of subject referral: Specify:

If the subject is a screen failure, Click here

If you need to correct the Subject's initials, please enter correct initials and Click on the push button below:

Correct Subject initials: [Click here to correct initials](#)

The subject's data are considered clean:

?

?

SDTM data set (SAS)

(different variables names and different structures than eCRF)

CONCEPT BASED MDR: CDASH/SDTM can be organized by CRC

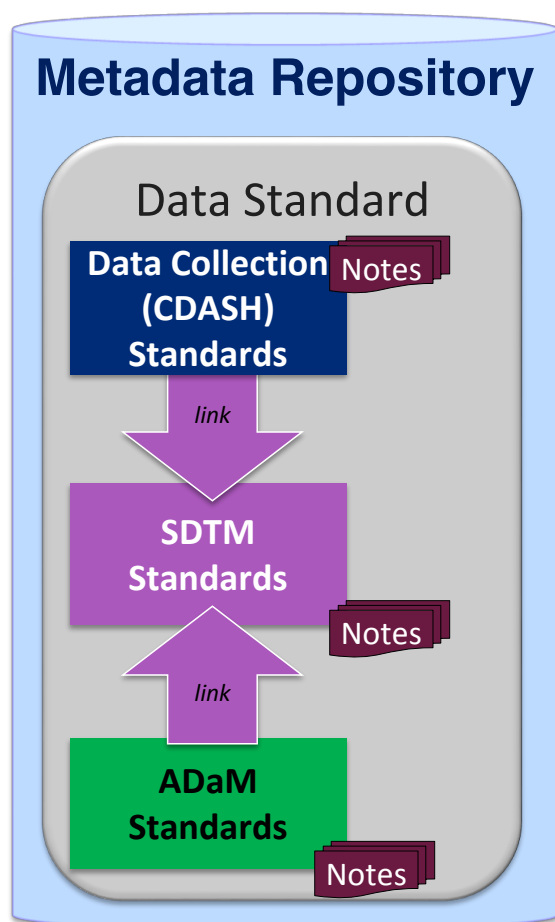
Period
Week
Visit
Type of contact
Inclusion and Clinical evaluation:
Informed Consent
Patient Demography
Medical and Surgical History
Physical Examination
Vital Signs and Body Weight
Previous and Concomitant Medication
Inclusion and Exclusion Criteria
Randomization
Central Laboratory Testing:
HbA1c
Lipid Profile ¹ (in fasting state)
Biochemistry ¹
Serum Pregnancy Test
Treatment:
Dispensation of insulin glargine or premixed insulin and corresponding training
Dispensation of insulin glulisine if needed (see section 9.1.3) and corresponding training
Accountability of Investigational Products
Plasma Glucose Self-Monitoring:
Dispensation of plasma glucose meter
Dispensation and collection of diaries
Training of the patient (PG meter, diary)
Plasma glucose values of 3 consecutive days in the week before the visit
7-point plasma glucose profile of 3 consecutive days in the week before the visit
Insulin dose(s) of the day before the visit
Efficacy and Safety:
Treatment Satisfaction Questionnaires
Hypoglycemia
AE / SAE

Concept	CDASH Question	CDASH Variable	SDTM Variable
Subject	What is the sex of the subject?	SEX	SEX
	What is the subject's date of birth?	BRTHDAT or BRTHYR BRTHMO BRTHDY	BRTHDTC
	What is the ethnicity of the subject?	ETHNIC	EHTNIC
	What is the race of the subject?	RACE	RACE
	What is the subject's age?	AGE	AGE
	What are the age	AGEU	AGEU

eCRF content description

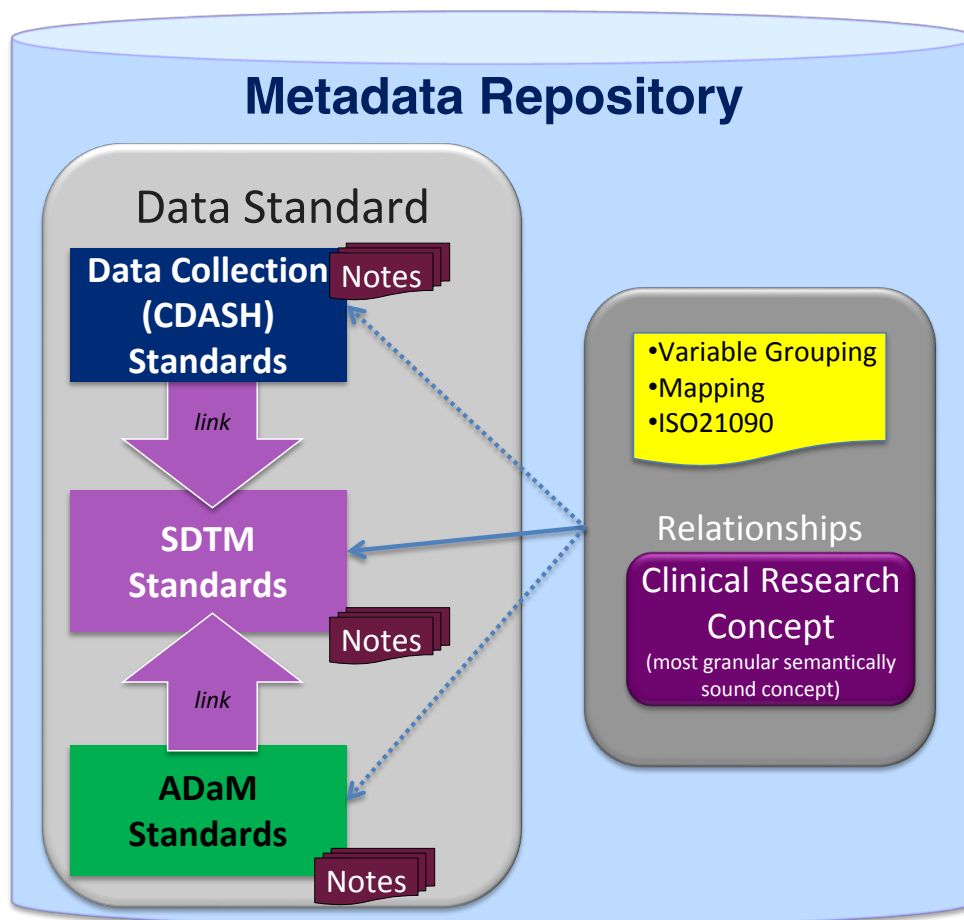
SDTM mapping

FROM A “VARIABLE” TO A “CONCEPT-BASED” MDR



- **Classical MDR**
- **We can expect more than 30.000 variables across all Therapeutic Areas**
- **How do we ensure ?**
 - semantic consistency when authoring
 - completeness when selecting variables

FROM A “VARIABLE” TO A “CONCEPT-BASED” MDR



CONCLUSIONS

- **Let us speak the same language**
- **We need to change the way we consider compliance to data standards and data integration:**
 - From a retro-active way (building define.xml at submission)
 - To a pro-active approach (study data standards defined at study set-up)
- **We need new tools to manage metadata: Concept based MDR**
 - Grouping variables into semantically meaningful concepts (following industry wide patterns)
 - Linking data collection (CDASH) to data submission (SDTM) variables
 - Linked with controlled terminology



Dr. Isabelle de Zegher
Senior Director, Clinical Information Management
PAREXEL Informatics
Parc des Collines
T +32 2 767 16 48
M +32 478 48 28 54

http://www.PhUSEwiki.org/wiki/index.php?title=Metadata_Management