PhUSE SDE Meeting, NY 2015

PhUSE Metadata Management Project

Metadata, Study data standards, Master data, terminology and interoperability definitions

Mitra Rocca, FDA

Marcelina Hungria, DIcore Group

Table of Content

- PhUSE CSS – Emerging Technology (ET) Working group
- Metadata Management Project
  - Metadata Definitions
  - Metadata Implementation
- Definitions
  - Metadata and study data standards
  - Master data
  - Controlled terminology
  - Interoperability
  - Pooling, aggregation, integration
- Lessons learned

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 standards
- In 2013 a new working group was established focusing on the following emerging technologies:
  - semantic technology (now in a dedicated working group)
  - **Metadata management**
  - Cloud computing
  - Big data
- The ET WG has re-organized in 2014

Metadata Management Project Goals

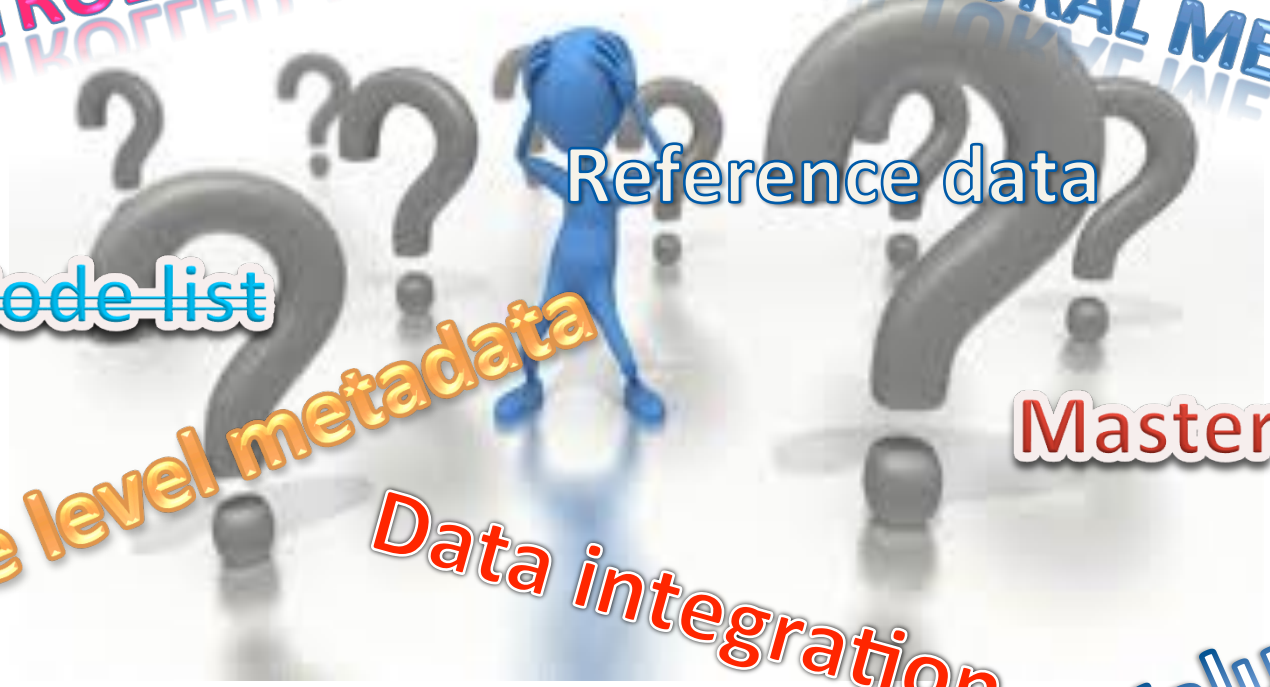
Changing landscape: need for concept based Metadata Repository (MDR) from protocol to data submission

Project Team Deliverables

- **Metadata Definitions Document**
  - http://www.PhUSEwiki.org/wiki/index.php?title=Metadata_Management

- **Comments to FDA Guidances**
  - **Submitted to the FDA docket (by the May-2014 deadline)**

Metadata Definitions Soup ….



CONTROLLED TERMINOLOGY

STRUCTURAL METADATA

Reference data

~~Code list~~

Value level metadata

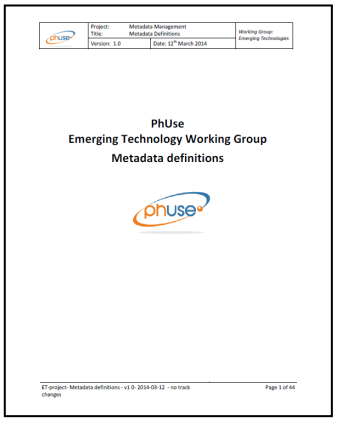
Master data

Data integration

Value set

Semantic interoperability

Metadata Definitions


PhUse
Emerging Technology Working Group
Metadata definitions

1 METADATA MANAGEMENT
- *1.1 Metadata*
- *1.2 Structural metadata*
- *1.3 Descriptive metadata*
- *1.4 Study Instance Metadata*
- *1.5 Metadata repository*
- *1.6 Metadata registry*
- *1.7 Data element*
- *1.8 Attribute*
- *1.9 Class*
- *1.10 Data type*
- *1.11 Value level metadata*

2 CONTROLLED TERMINOLOGY, CODE SYSTEMS & VALUE SETS
- *2.1 Controlled Terminology/controlled vocabulary*
- *2.2 Code system*
- *2.3 Dictionary*
- *2.4 Concept*
- *2.5 Code*
- *2.6 Code list*
- *2.7 Value set*

3 MASTER DATA MANAGEMENT
- *3.1 Master Data*
- *3.2 (Master) Reference Data*
- *3.3 Master Data Management*

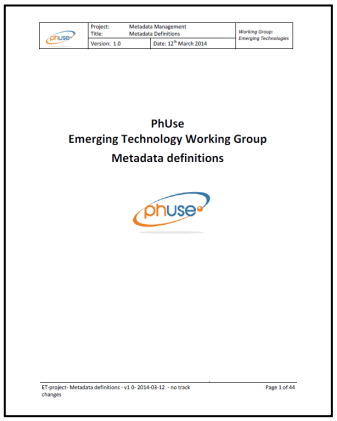
4 INTEROPERABILITY
- *Categorization of Interoperability (by HL7)*
- *4.1 Technical interoperability ("machine interoperability")*
- *4.2 Semantic interoperability*
- *4.3 Process Interoperability*

5 DATA AGGREGATION, INTEGRATION, POOLING
- *5.1 Data pooling*
- *5.2 Data integration*
- *5.3 Data aggregation*

PhUSE SDE Meeting, NY 2015

Approach
Definitions
Lessons Learned

Metadata Definitions


PhUse
Emerging Technology Working Group
Metadata definitions

1 METADATA MANAGEMENT
- *1.1 Metadata*
- *1.2 Structural metadata*
- *1.3 Descriptive metadata*
- *1.4 Study Instance Metadata*
- *1.5 Metadata repository*
- *1.6 Metadata registry*
- *1.7 Data element*
- *1.8 Attribute*
- *1.9 Class*
- *1.10 Data type*
- *1.11 Value level metadata*

2 CONTROLLED TERMINOLOGY, CODE SYSTEMS & VALUE SETS
- *2.1 Controlled Terminology/controlled vocabulary*
- *2.2 Code system*
- *2.3 Dictionary*
- *2.4 Concept*
- *2.5 Code*
- *2.6 Code list*
- *2.7 Value set*

3 MASTER DATA MANAGEMENT
- *3.1 Master Data*
- *3.2 (Master) Reference Data*
- *3.3 Master Data Management*

4 INTEROPERABILITY
- *Categorization of Interoperability (by HL7)*
- *4.1 Technical interoperability ("machine interoperability")*
- *4.2 Semantic interoperability*
- *4.3 Process Interoperability*

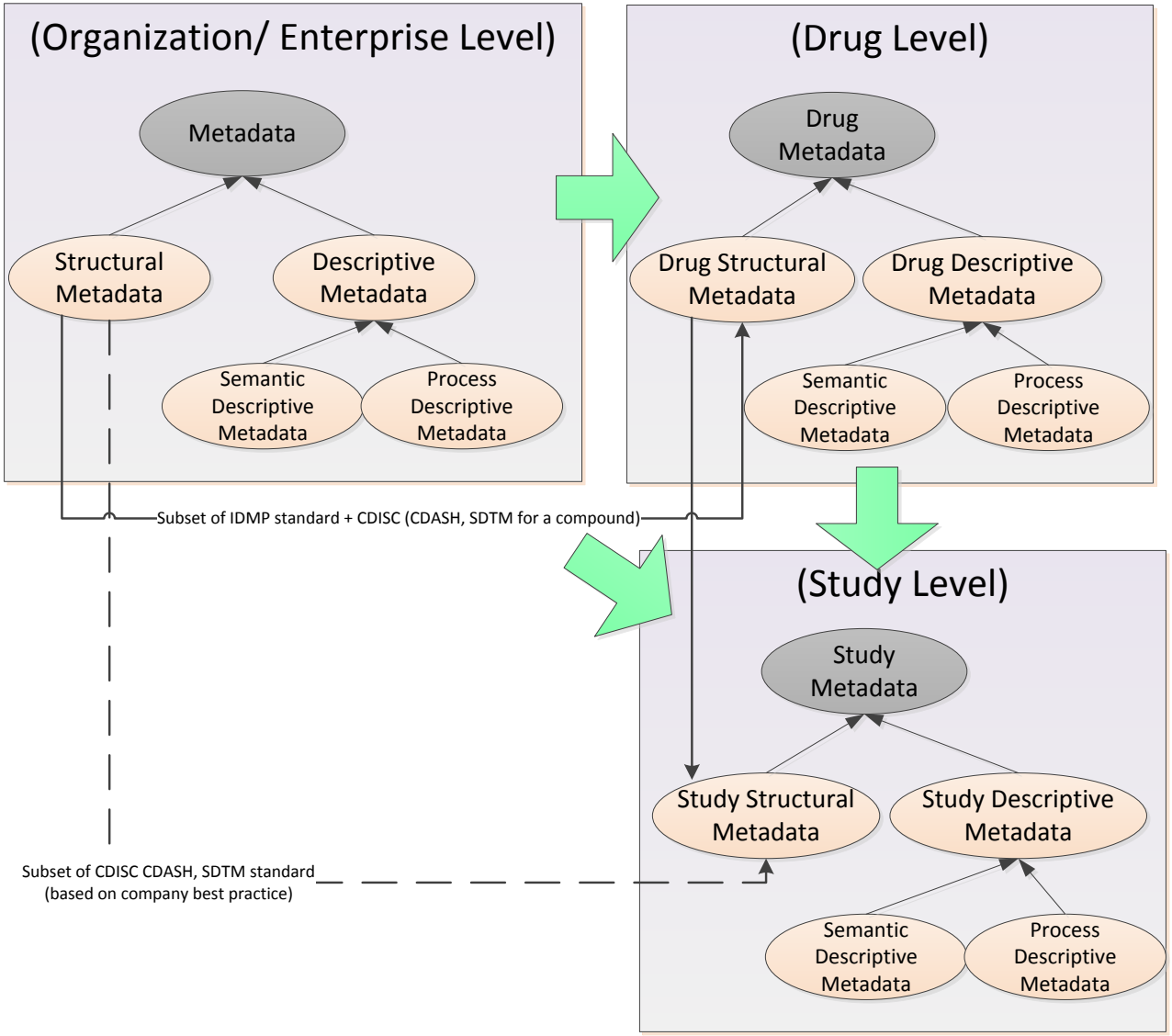
5 DATA AGGREGATION, INTEGRATION, POOLING
- *5.1 Data pooling*
- *5.2 Data integration*
- *5.3 Data aggregation*

Approach

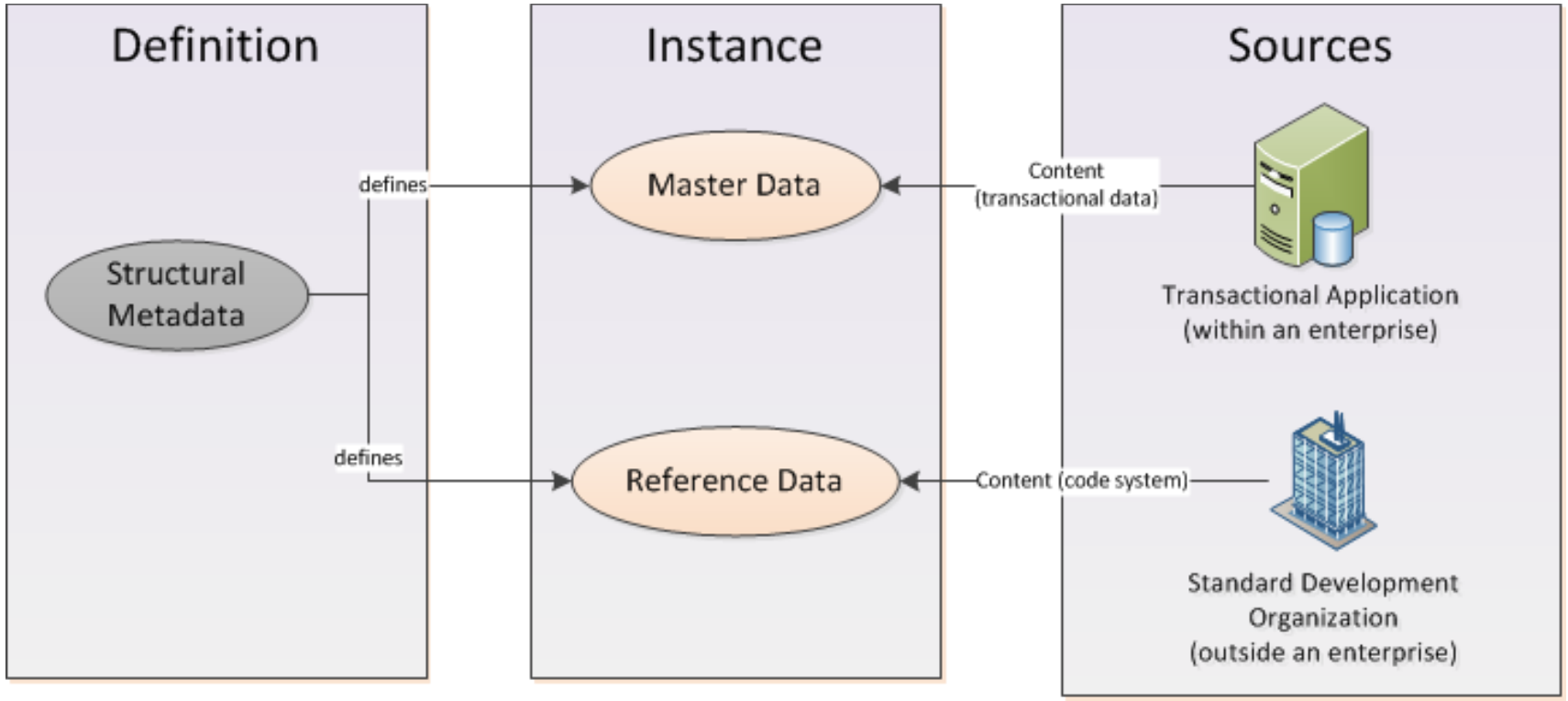
Master Data Management

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. *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
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.

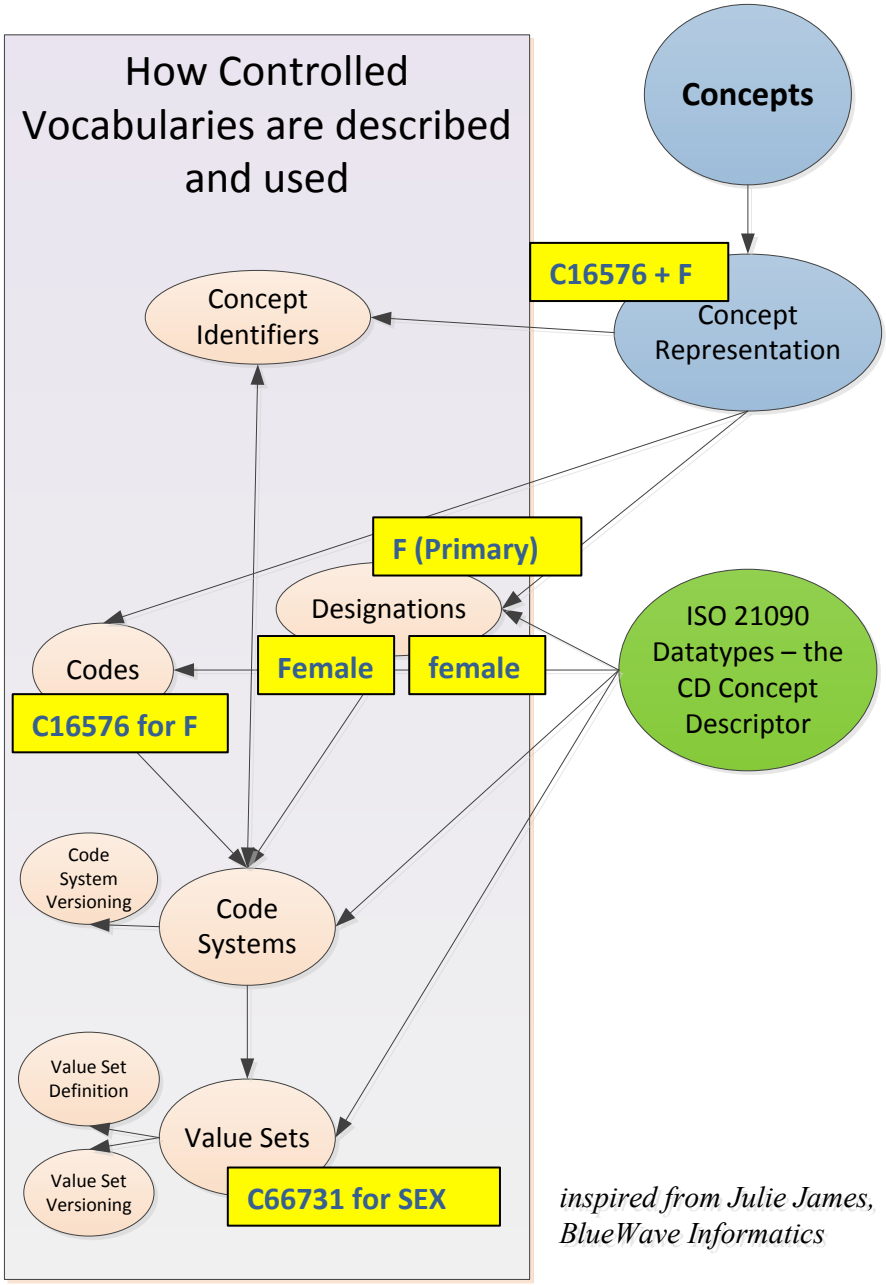
Metadata


(Organization/ Enterprise Level)
Metadata
Structural Metadata
Descriptive Metadata
Semantic Descriptive Metadata
Process Descriptive Metadata
(Drug Level)
Drug Metadata
Drug Structural Metadata
Drug Descriptive Metadata
Semantic Descriptive Metadata
Process Descriptive Metadata
Subset of IDMP standard + CDISC (CDASH, SDTM for a compound)
(Study Level)
Study Metadata
Study Structural Metadata
Study Descriptive Metadata
Semantic Descriptive Metadata
Process Descriptive Metadata
Subset of CDISC CDASH, SDTM standard
(based on company best practice)

Master Data


Definition
Instance
Sources
defines
Master Data
Content
(transactional data)
Structural
Metadata
Transactional Application
(within an enterprise)
defines
Reference Data
Content (code system)
Standard Development
Organization
(outside an enterprise)

Controlled Terminology



inspired from Julie James, BlueWave Informatics

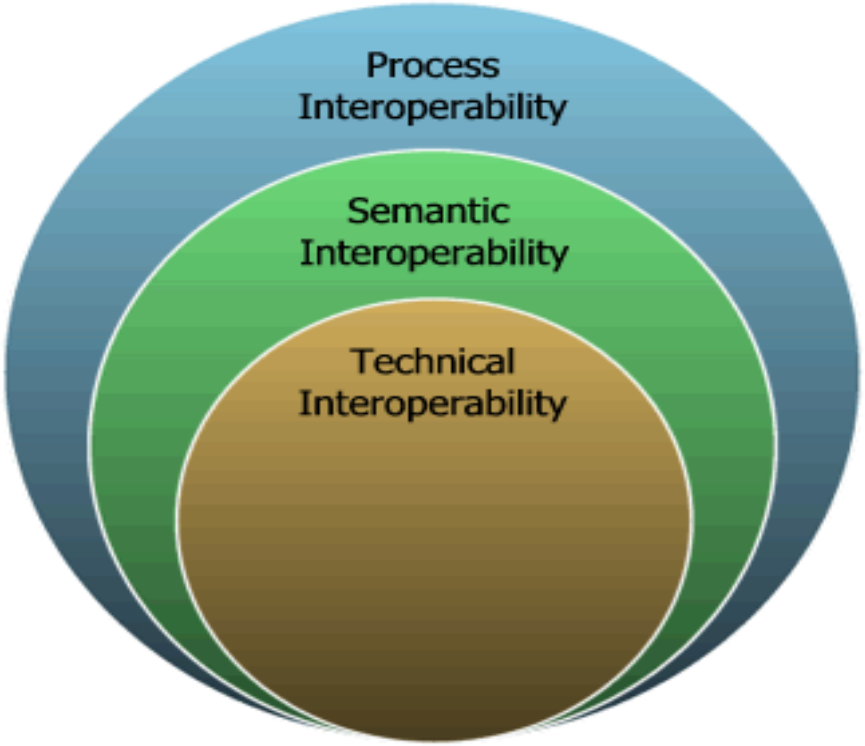
In define.xml (machine processable):

- **Code System (CodeList Context): nciExtCodeID (not directly processable – URI instead)**
- **Value Set (CodeList) CUI for SEX: C66731**
- **Code CUI for Designation F (Female): C16576**

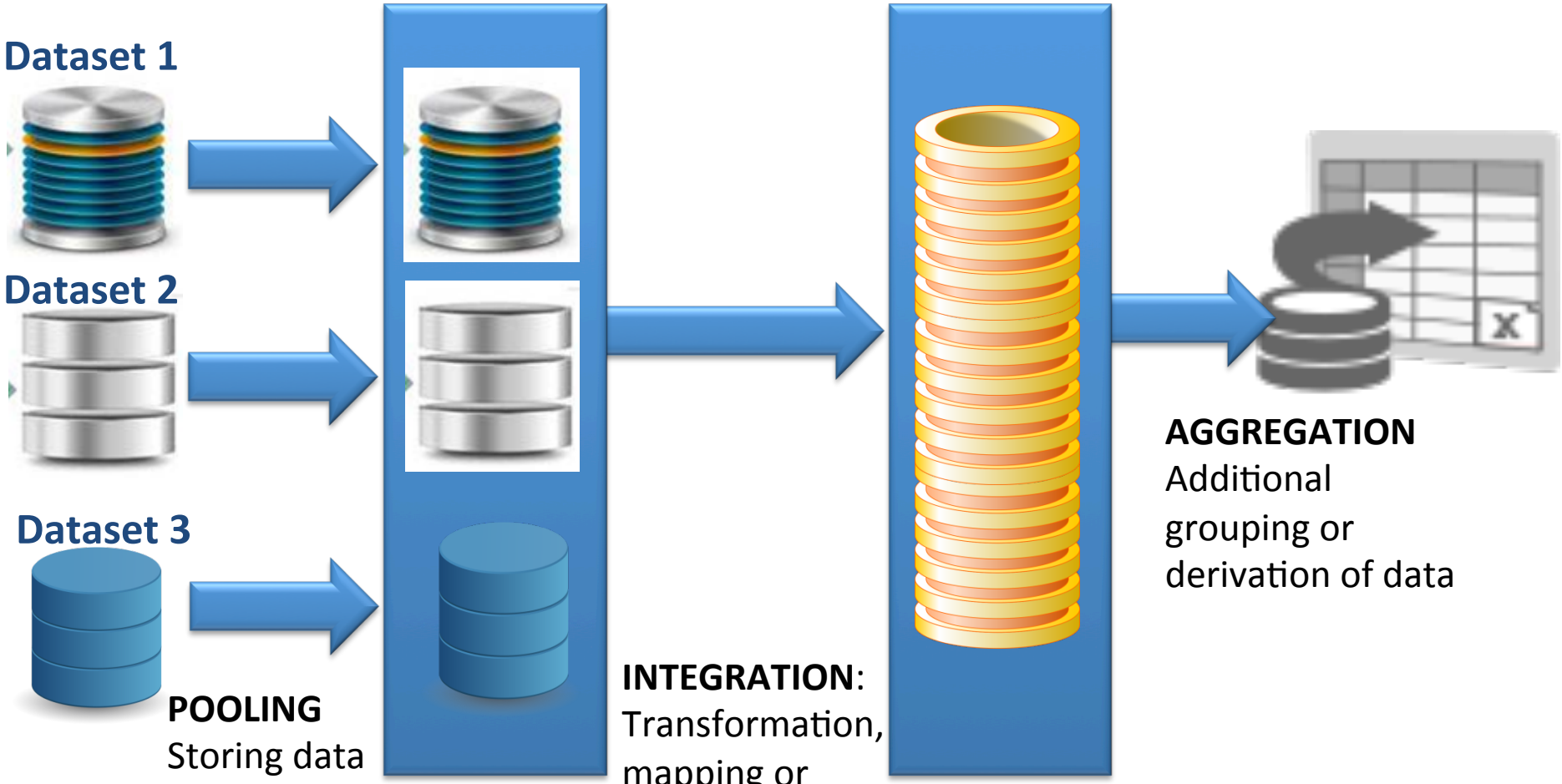
"~~Codelist~~"

Value Set & Code with CDISC example

Interoperability



Data Pooling, Integration, Aggregation



Lessons learned (Compiled from different team members)

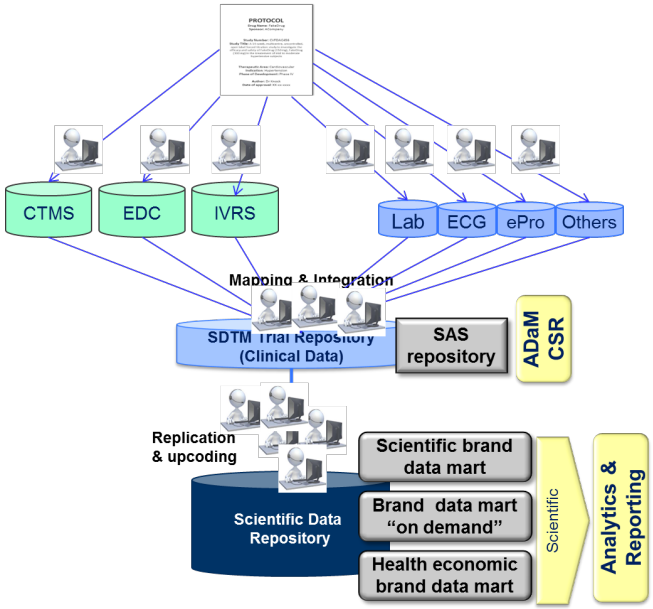
- **Efficient Data Integration and compliance to regulatory standards does not start after pooling (retroactive approach); it starts with the protocol (proactive approach)**

- **A proactive approach is based on two components:**
  - Definition of Master Data (Drug Products, Studies, Sites, Investigators, ..) and associated 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)**

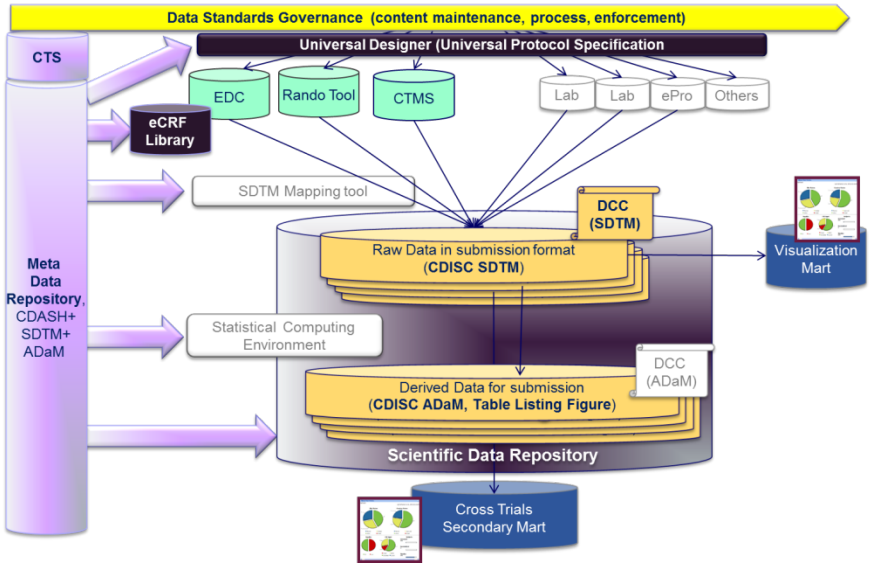
CHANGING LANDSCAPE :
Enforcing data standards from protocol onwards

Retro-active approach from paper protocol,



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

Pro-active approach with structural metadata

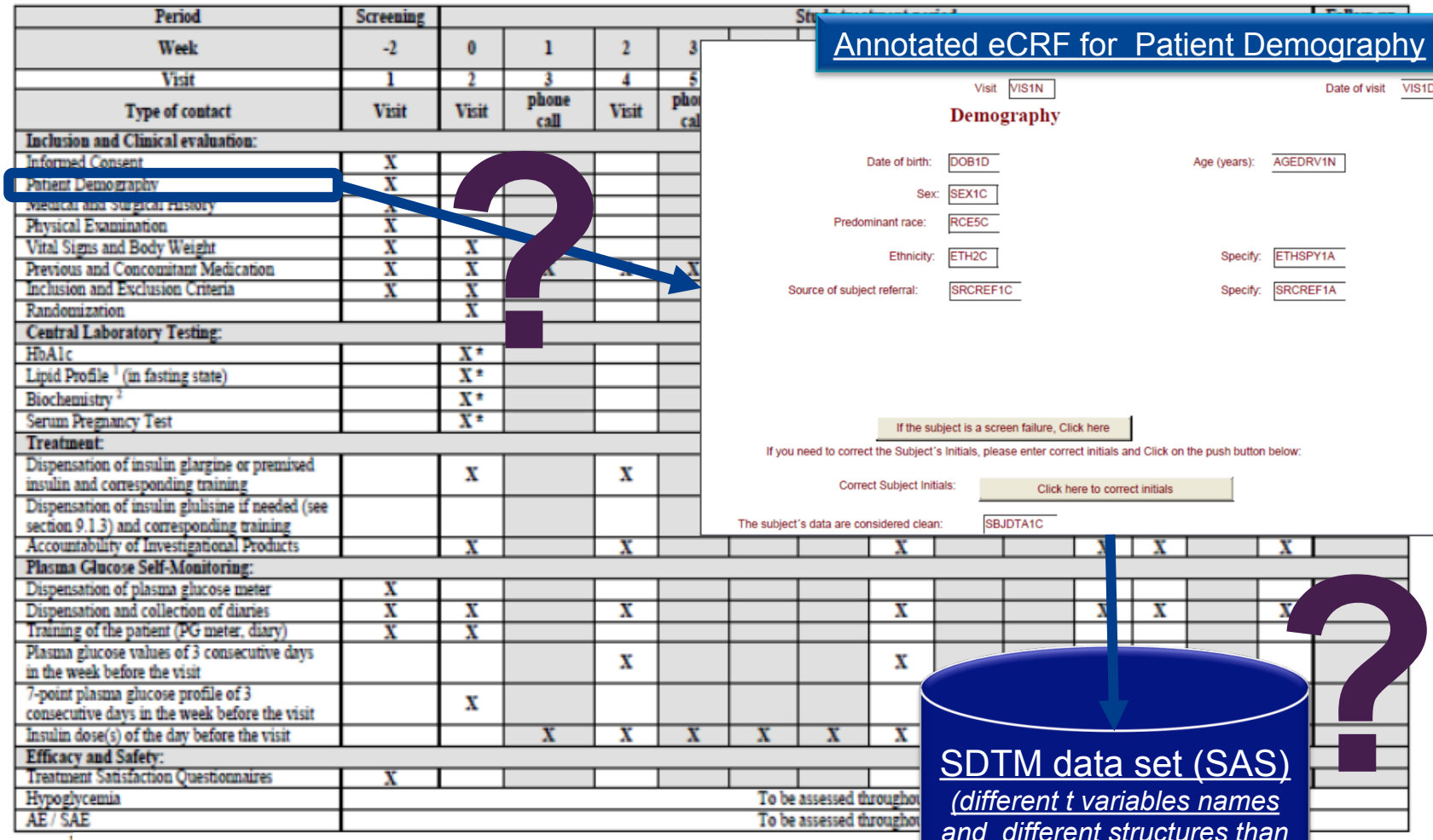


- One single interpretation of protocol
- Increased efficiency, consistency & quality through standards
- Reduced time for integration and secondary data use

Courtesy of Isabell de Zegher

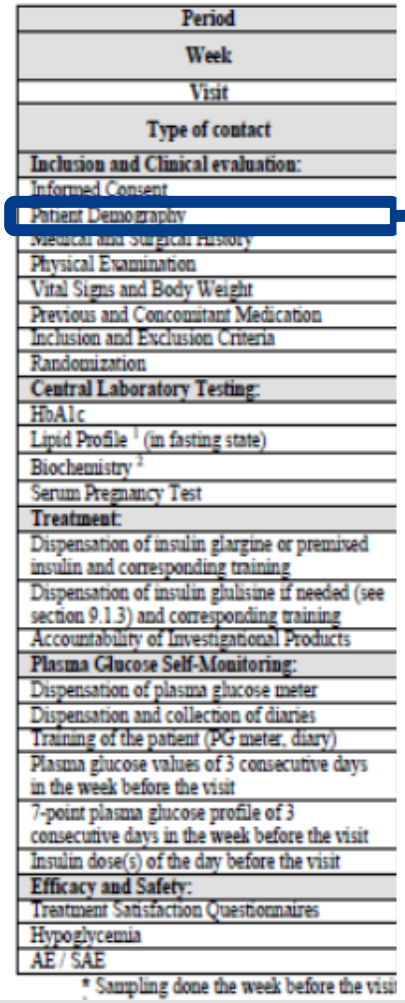
PAREXEL

CONCEPT BASED MDR :
Protocol is not about variables but about concepts



Courtesy of Isabell de Zegher

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 [2]
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

* Sampling done the week before the visi

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 units used?	AGEU	AGEU

eCRF content description

SDTM mapping

Courtesy of Isabell de Zegher

Conclusions

- Let us speak the same language

- We need to change the way we consider compliance to data standards and data integration:
  - From a retroactive way (building define.xml at submission)
  - To a proactive approach (study data standards defined at study setup)

- We need new tools to manage metadata: Concept based MDR
  - Grouping variables into semantically meaningful concepts (following industry wide patterns)
  - Linking data sources (e.g, CDASH based collection) to data submission (SDTM) variables
  - Linking with controlled terminology
  - With capabilities to handle standards versioning

Metadata Definition Project Participants

Isabelle de Zegher (co-chair)	Parexel
Mitra Rocca (co-chair)	FDA
Marcelina Hungria (co-chair)	DiCore Group
Julie James	BlueWave Informatics
Tim Church	Torch
Yun Oldshue	Takeda
Praveen Garg	ICON
Kenneth Stoltzfus	Accenture
Gregory Steffens	Novartis
John Leveille	d-Wise
Aimee Basile	Celgene
Sam Hume	CDISC

PhUSE SDE Meeting, NY 2015

Mitra Rocca

Senior Medical Informatician
Office of Translational Sciences
CDER, FDA
Mitra.rocca@fda.hhs.gov

Marcelina Hungria

Clinical Data Standards & Integration
Consultant / Owner
DIcore Group, LLC
mhungria@dicoregroup.com