THE NATIONAL CHILDREN'S STUDY

HEALTH GROWTH ENVIRONMENT

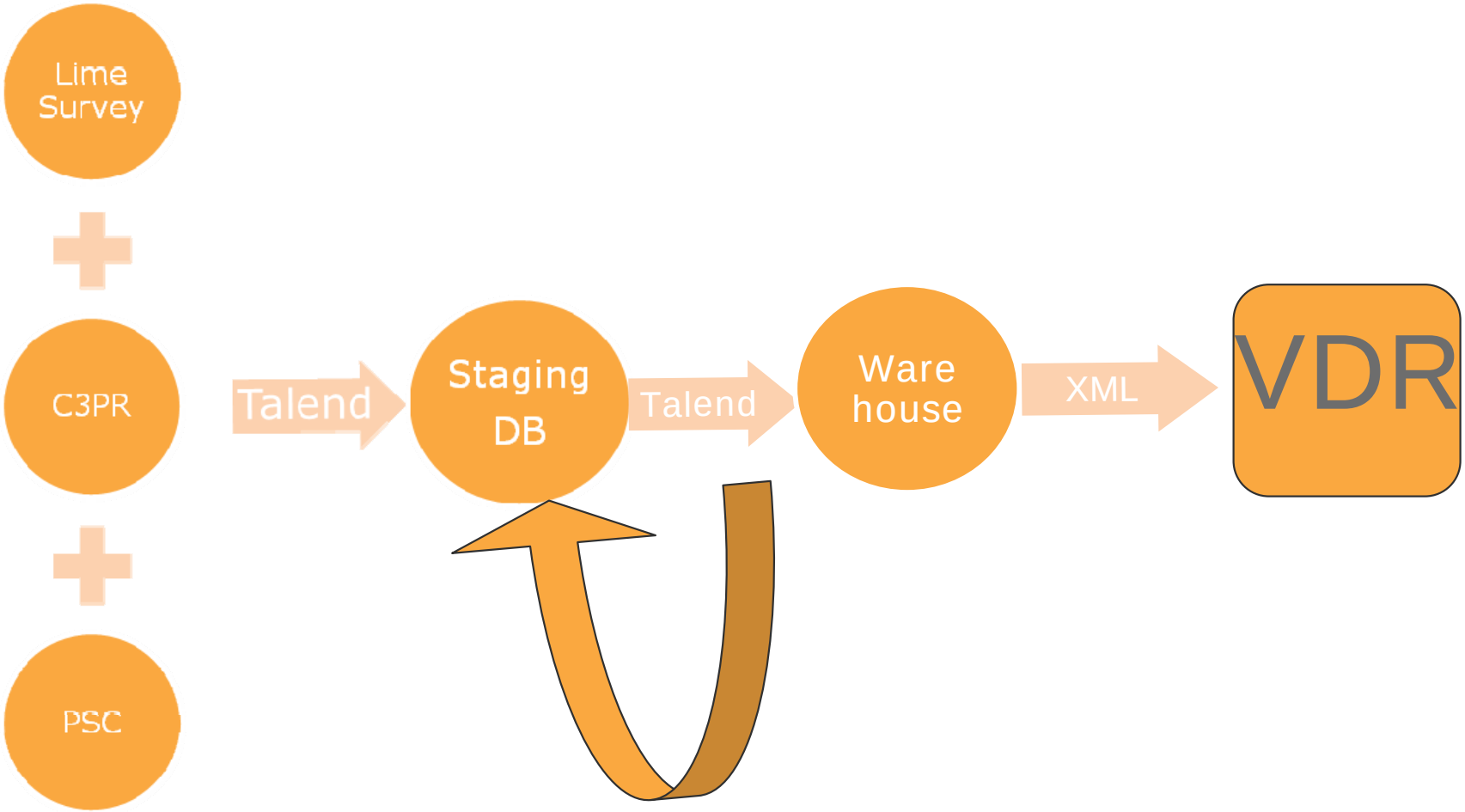
Metadata-Driven Survey and Study Development for the NCS

Umit Topaloglu Ph.D.
Arkansas Study Center
January 23, 2012

Content

- Current data collection and MDES mapping workflow at our study center
  - Challenges
- Common Data elements (CDE)
  - caDSR
- Meta Data Repository
  - OpenMDR
- A potential solution

VDR and Talend



tMap: Transformation

Data Collection with LimeSurvey

IMPROVING THE HEALTH OF AMERICA'S CHILDREN

THE NATIONAL CHILDREN'S STUDY

HEALTH GROWTH ENVIRONMENT

V2.0 - Telephone Pregnancy Screener (Am 52 - 04/15/2011)

0% 100%

PAGE 6

EL003.
What is your date of birth?

Jan 05 1986

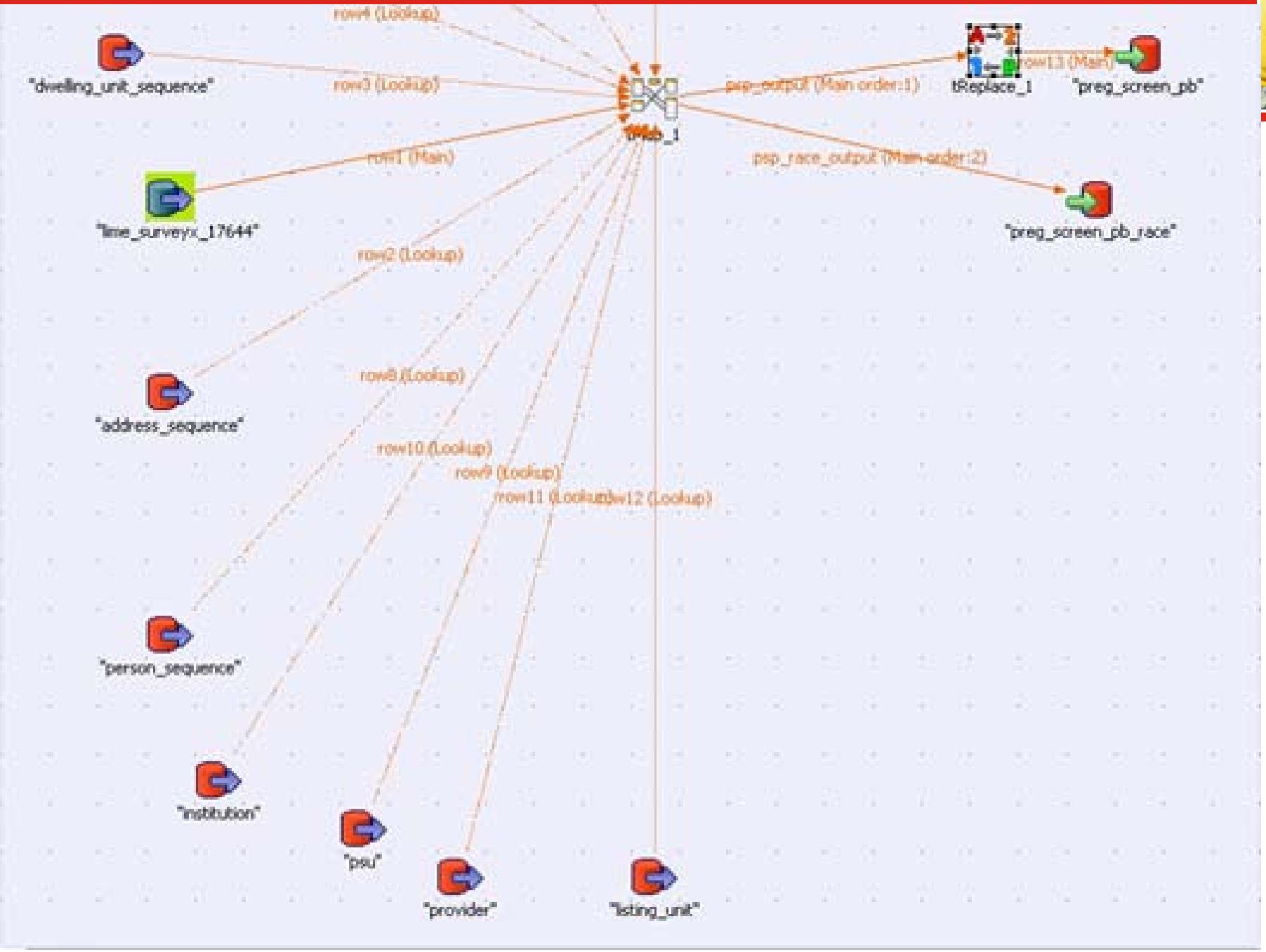
? INTERVIEWER INSTRUCTION: IF UNDER LOCAL AGE OF MAJORITY, NOT ELIGIBLE FOR THE STUDY, EVEN IF PREGNANT, SKIP TO (AGE_ELIG).

*
EL005
Which range best describes your age? Would you say...

- Less than 18
- 18-24
- 25-34
- 35-44
- 45-49
- 50-64
- 65 or older
- REFUSED
- DON'T KNOW

? INTERVIEWER INSTRUCTION: DESCRIBE HOW THE ANSWER TO THIS QUESTION DETERMINES HER ELIGIBILITY AND THAT ALL DATA ARE KEPT CONFIDENTIAL AND SECURE.

A Talend Job


row4 (Lookup)
"dwelling_unit_sequence"
row3 (Lookup)
psp_output (Main order:1)
tReplace_1
row13 (Main)
"preg_screen_pb"
tMap_1
row1 (Main)
psp_race_output (Main order:2)
"time_surveyx_17644"
"preg_screen_pb_race"
row2 (Lookup)
row8 (Lookup)
"address_sequence"
row10 (Lookup)
row9 (Lookup)
row11 (Lookup)
row12 (Lookup)
"person_sequence"
"institution"
"psu"
"provider"
"listing_unit"

Talend Open Studio - tMap - tMap_1

New version of instruments

New version of MDES

Schema editor | Expression editor

Column	Key	Type	N..	Date Pattern (Ctrl...	Length	Precision	Default	Comment
X17644X27X231	✓	String			5	0		
X17644X27X232		String	✓		21845	0		
X17644X28X240		String	✓		1	0		
X17644X28X233SNUM		String			21845	0		
X17644X28X233PRED		String	✓		21845	0		
X17644X28X233SNAM		String			21845	0		
X17644X28X233STYP		String			21845	0		
X17644X28X233POSD		String	✓		21845	0		
X17644X28X234		String	✓		21845	0		
X17644X28X235UTYP		String	✓		21845	0		
X17644X28X235UNUM		String	✓		21845	0		
X17644X28X236		String	✓		21845	0		
X17644X28X237		String	✓		5	0		
X17644X28X238		Double	✓		22	0		

Column	Key	Type	N..	Date Pattern (Ctrl...	Length	Precision	Default	Comment
phone		String			100	0		
r_phone_1		String			10	0		
time_stamp_2		String			19	0		
time_stamp_3		String	✓		19	0		
r_fname		String			30	0		
r_lname		String			30	0		
person_dob		String	✓		10	0		
age		String	✓		100	0		
age_range		String			100	0		
age_elig		String	✓		100	0		
time_stamp_4		String	✓		19	0		
address_id		String	✓		36	0		
address_1		String	✓		100	0		
address_2		String	✓		100	0		

Apply | Ok | Cancel

Expression for one column

Expression Builder

Expression

Wrap | Undo(Ctrl + Z) | Clear

```
Relational.ISNULL(row1.X17644X28X358) ||
row1.X17644X28X358 == "" || row1.X17644X28X358 == "null"
 ? row1.X17644X28X233SNUM + " " + row1.X17644X28X233PRED
+ " " + row1.X17644X28X233SNAM + " " +
row1.X17644X28X233STYP + " " + row1.X17644X28X233POSD :
row1.X17644X28X358.equals("N1") ? "-1" :
row1.X17644X28X358.equals("N2") ? "-2" : "-4"
```

+ - * / == < <= != >= > and or not ( )

Test

Test! | Clear

Var	Valu
row1.id	0
row1.submitdate	null
row1.lastpage	null
row1.startlangu...	null
row1.token	null
row1.datestamp	null
row1.startdate	null
row1.X17644X4...	null

Add | Remove

Categories

*All
*User Defined
DataOperation
Mathematical
Numeric
Relational
StringHandling
TalendDataGenerator
TalendDate
TalendString

Functions

Help

Please select a category and function.

Ok | Cancel

Batch job for all Talend jobs

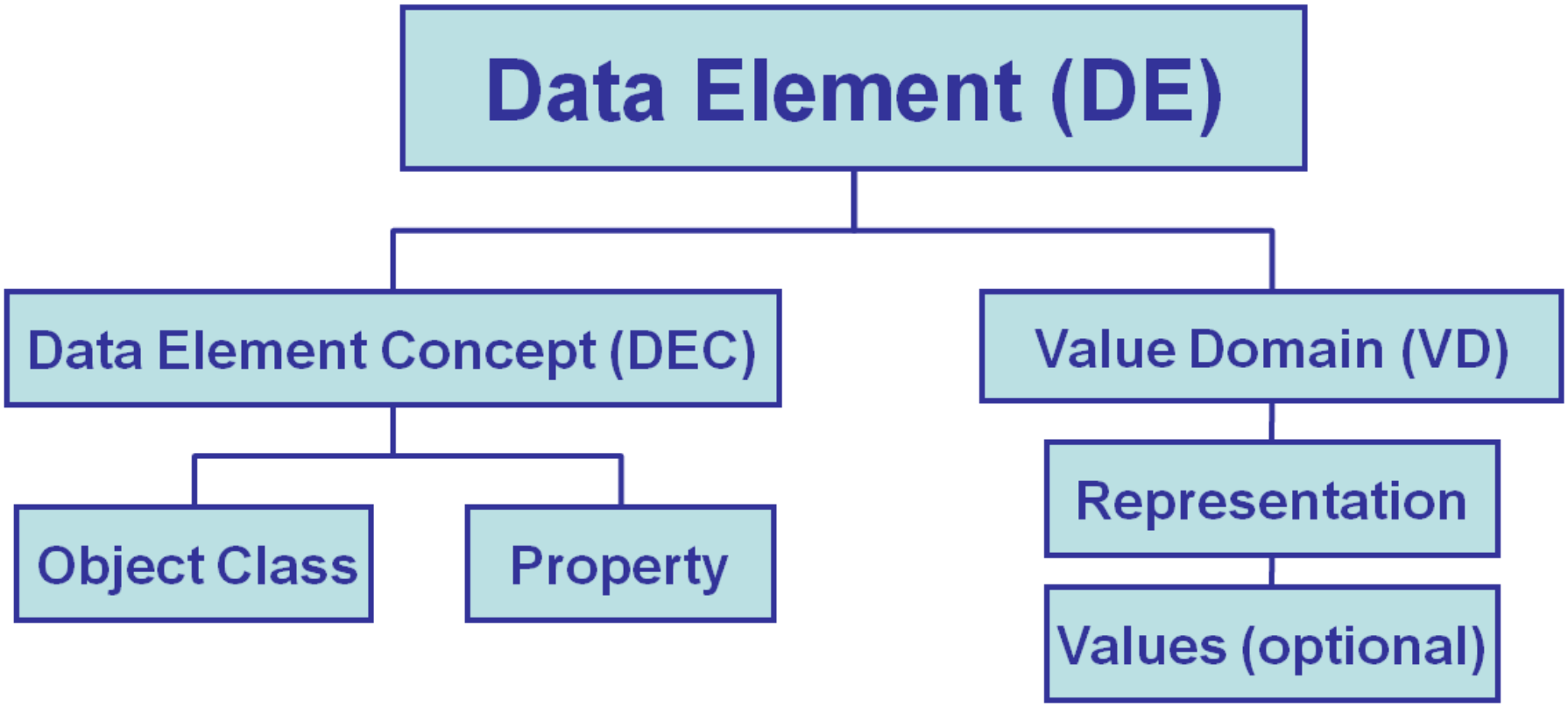
Challenges with the approach

- It is very time consuming to set up
  - 2 FTEs are constantly working for VDR submission
- It requires very tightly controlled version control.
  - Any change in data collection environment needs to be mapped manually to the VDR warehouse

Common Data Element (CDE)

- "is to standardize the collection of investigational data in order to facilitate comparison of results across studies and more effectively aggregate information into significant metadata results."
  - minimize the time requirement for data collection tools
  - Provides standardized, consistent data collection
  - Improve data quality
  - With semantic awareness data sharing is obtainable
  - Improve opportunities for meta-analysis and comparison of results from different studies
- Being used by the NCI, NINDS etc.

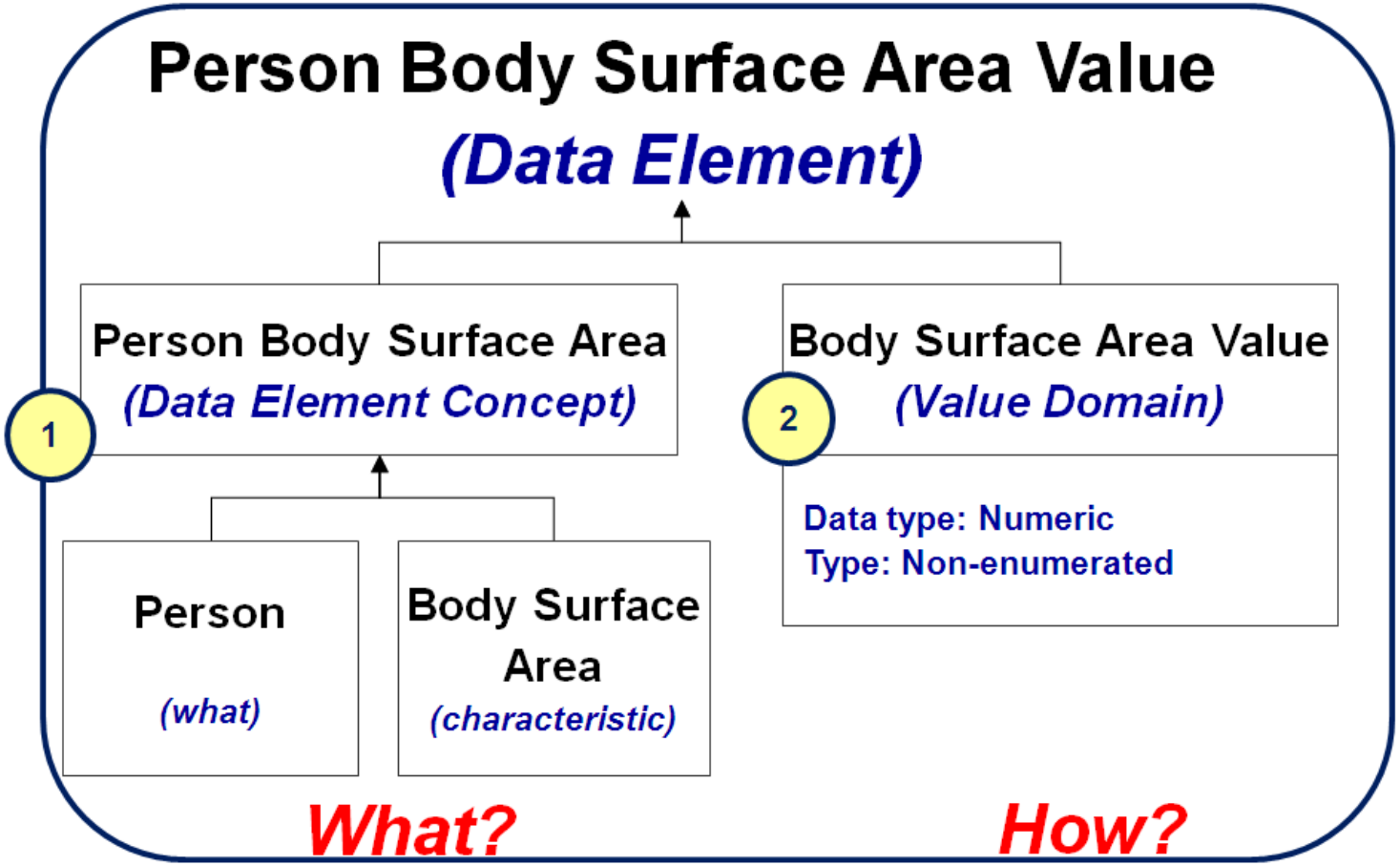
Data elements 101



Object Class + Property = Data Element Concept + Value Domain = Data Element

Example data element construction

- Data elements include:
  1. Conceptual entity (**THING**) about which data is being collected = **Data Element Concept**



2. Physical representation = **Value Domain**

caDSR

- caDSR (cancer Data Standards Repository) is a database and a set of APIs and tools to create, edit, control, deploy, and find common data elements (CDEs) for use by metadata consumers”
- https://cabig.nci.nih.gov/concepts/caDSR/

CDE Browser

National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov

caDSR

- PhenX
- PrincipalComponentsAnalysi
- ProteomicsLIMS
- protExpress
- RadLex
- Reactome Database Sharing
- RProteomics
- Sarcoma Database
- Seed
- SHP Microarray Analytical Ser
- SVM
- TARGET/CGCI
- Taverna-caGrid
- TCR
- The Cancer Genome Atlas
- Theradex
- Tissue Banks and Pathology
- Transcend
- Transcription Annotation Prio
- UAMS Clinical Research
  - UAMS New CDEs
- University of Michigan
- Utah

	Consent Ind-2	to registration and any study-related procedures being performed						
☐	Abdominal Incision Normal Healed Indicator	Abdomen normal healing?	caBIG			RELEASED	3060654	1.0
☐	Acute Pain Present Indicator	Sharp pains?	caBIG			RELEASED	3079313	1.0
☐	Anterior Bladder Peritoneum Culdoscopy Device Visual Indicator	Culdoscopy visualization of anterior bladder peritoneum	caBIG			RELEASED	3061290	1.0
☐	Anterior Bladder Peritoneum Laparoscopy Device Visual Indicator	Released by approval of UAMS team.	caBIG			RELEASED	3061490	1.0
☐	Appendix Culdoscopy Device Visual Indicator	Culdoscopy visualization of appendix	caBIG			RELEASED	3061295	1.0
☐	Appendix Laparoscopy Device Visual Indicator	Laparoscopy visualization of appendix	caBIG			RELEASED	3061501	1.0
☐	Breast Cancer Diagnosis	Breast cancer kind?	caBIG			RELEASED	3079088	1.0
☐	Breast Cancer Diagnosis Breast Laterality	Which breast was diagnosed with breast cancer?	caBIG			RELEASED	3111265	1.0
☐	Breast Cancer Diagnosis Indicator	Diagnosed with breast cancer?	caBIG			RELEASED	3079097	1.0
☐	Breast Cancer Family History Indicator	Family History Breast Cancer	caBIG			RELEASED	3079291	1.0
☐	Breast Implant Indicator	Breast implants?	caBIG			RELEASED	3079284	1.0
☐	Breast Lump Present Indicator	Breast lumps?	caBIG			RELEASED	3079319	1.0

Race DE

CDE Browser

Back Help

Data Element | Data Element Concept | Value Domain | Classifications | Usage | Data Element Derivation | Admin Info

Data Element Details

Public ID:	2513662
Version:	1.0
Long Name:	Participant Race Category
Short Name:	2513396v1.0:2015164v6.0
Preferred Question Text:	Race
Definition:	U.S. Office of Management & Budget (OMB) approved categories for racial origin.
Value Domain:	Race Category
Data Element Concept:	Participant Race
Context:	caBIG
Workflow Status:	RELEASED
Origin:	
Registration Status:	Qualified
Direct Link:	https://cdebrowser.nci.nih.gov/CDEBrowser/search?elementDetails=9&FirstTimer=0&PageId=ElementDetailsGroup&publicId=2513662&version=1.0

Reference Documents

Document Name	Document Type	Document Text	Context	URL
Race	Preferred Question Text	Race	caBIG	

Alternate Names and Definitions

Race value domain

Selected Data Element

Public ID:	2513662
Version:	1.0
Long Name:	Participant Race Category
Short Name:	2513396v1.0:2015164v6.0
Preferred Question Text:	Race
Definition:	U.S. Office of Management & Budget (OMB) approved categories for racial origin.
Workflow Status:	RELEASED

Value Domain Details

Public ID:	2015164
Version:	6.0
Long Name:	Race Category
Short Name:	RACE_CAT
Context Name:	CTEP
Definition:	racial origination based on OMB approved categories.
Workflow Status:	RELEASED
Datatype:	CHARACTER
Unit of Measure:	
Display Format:	
Maximum Length:	41
Minimum Length:	5
Decimal Place:	
High Value:	
Low Value:	
Value Domain Type:	Enumerated
Conceptual Domain Public ID:	2008552
Conceptual Domain Short Name:	PHYS_DESC_OF_INDS
Conceptual Domain Context Name:	CTEP
Conceptual Domain Version:	1.0
Origin:	

Race permissible values

Permissible Values

1 - 7 of 7

PV	PV Meaning	PV Meaning Concept Codes	PV Meaning Description	PV Begin Date	PV End Date	VM Public ID	VM Version
American Indian or Alaska Native	American Indian or Alaska Native	C41259	Denotes a person having origins in one of the indigenous peoples of North America, who lived on the continent prior to the European colonization. The term includes individuals belonging to a large number of tribes, states, and ethnic groups, many of them still enduring as communities.	2002-05-31		2572232	1.0
Asian	Asian	C41260	Denotes a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including Cambodia, China, India, Japan, Korea, Malaysia, Mongolia, Pakistan, the Philippine Islands, Thailand, and Vietnam.	2002-05-31		2572233	1.0
Black or African American	Black or African American	C16352	A term used in the United States to categorize a population group comprised of persons having origins in any of the black racial groups of Africa. Includes population subgroups (e.g., Kenyan, Nigerian, Haitian). The concept refers also to individuals who classify themselves as described.	2002-05-31		2572313	1.0
Native Hawaiian or other Pacific Islander	Native Hawaiian or Other Pacific Islander	C41219	Denotes a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. The term covers particularly people who identify themselves as part-Hawaiian, Native Hawaiian, Guamanian or Chamorro, Carolinian, Samoan, Chuukese (Trukese), Fijian, Kosraean, Melanesian, Micronesian, Northern Mariana Islander, Palauan, Papua New Guinean, Pohnpeian, Polynesian, Solomon Islander, Tahitian, Tokelauan, Tongan, Yapese, or Pacific Islander, not specified.	2002-05-31		2572235	1.0
Not Reported	Not Reported	C43234	Not provided or available.	2003-10-16		2572578	1.0
Unknown	Unknown	C17998	Not known, not observed, not recorded, or refused.	2002-02-11		2572577	1.0
White	White	C41261	Denotes person with European, Middle Eastern, or North African ancestral origin who identifies, or is identified, as White.	2002-05-31		2572236	1.0

Metada Data Repository -- OpenMDR

- OpenMDR is a suite of software that provides semantic metadata management capabilities. It has:
  - MDR Core,
  - MDR Query,
  - MDR Plugin, and
  - MDR Domain Model Generator.

MDR Core

- An ISO11179-compliant Metadata registry that stores semantic metadata for a domain or institution
- Capable of storing, versioning and maintaining semantic metadata in the form of Common Data Elements (CDEs)
- Provides a web front end for users to create, edit, deploy and re-use CDEs for new UML model development
- http://www.cagrid.org/display/MDR/Home

Creating a data element concept

openMDR

Welcome: admin || logout

MAINTENANCE | CONTENTS | SEARCH | REFERENCE DOCUMENTS | ANNOTATED MODELS | openMDR WIKI

Creating a New DataElementConcept

This form will allow you to create a new DataElementConcept in the metadata repository

Standard Administered Item Metadata

Preferred Name

Context *	CCTS
Name *	Patient Gender
Preferred	
Definition	Patient Gender
Language Identifier	US / eng
Source	

Data Element Concept Specific properties

Object Class URI *	cagrid.org_28421362-c5d7-4f07-8a6a-19239bced2c5_ Patient Change Relationship
Property URI *	cagrid.org_ce36df2a-2710-4051-9cea-b9a6b19a1f77_0 Gender Change Relationship
Conceptual Domain *	cagrid.org_bc780ebd-51f5-4fec-a6c2-beff525b7e7f_0.1 Gender Change Relationship

Store

Save | Clear | Return to Maintenance Menu

Preferred Name

Context *	CCTS
Name *	Gender
Preferred	●
Definition	please complete definitions...
Language Identifier	US / eng
Source	

Conceptual Domain

Conceptual Domain ID	cagrid.org_bc780ebd-51f5-4fec-a6c2-beff525b7e7f_0.1 (Change Relationship)
Conceptual Domain Name	Gender

Value Domain

Value Domain Data Type	string: XMLSchema
Value Domain Unit of Measure	select ...
Value Domain Format (E.g. MM-DD-YYYY)	
Value Domain Maximum Character Count	1
Value Domain Minimum Character Length	0
Value Domain High Value	
Value Domain Low Value	
Value Domain Decimal Place	

Possible Values	meaning	value
Permissible Value 1	Male	M
Permissible Value 2	Female	F

Store

Save Clear Return to Maintenance Menu

Creating a data element

openMDR

Welcome: admin || logout

MAINTENANCE | CONTENTS | SEARCH | REFERENCE DOCUMENTS | ANNOTATED MODELS | openMDR WIKI

Creating a New DataElement

This form will allow you to create a new DataElement in the metadata repository

Preferred Name

Context *	CCTS
Name *	Patient Gender
Preferred	
Definition	Patient Gender
Language Identifier	US / eng
Source	
Data Element Concept *	cagrid.org_c4738bce-4b53-4eba-b024-d3667342a57d_ Patient Gender [Change Relationship]
Value Domain *	cagrid.org_da251044-2fc3-4aa2-ba8b-a731166e2e19_ Gender [Change Relationship]
Example	please complete example...
Precision	

Store

Save | Clear | Return to Maintenance Menu

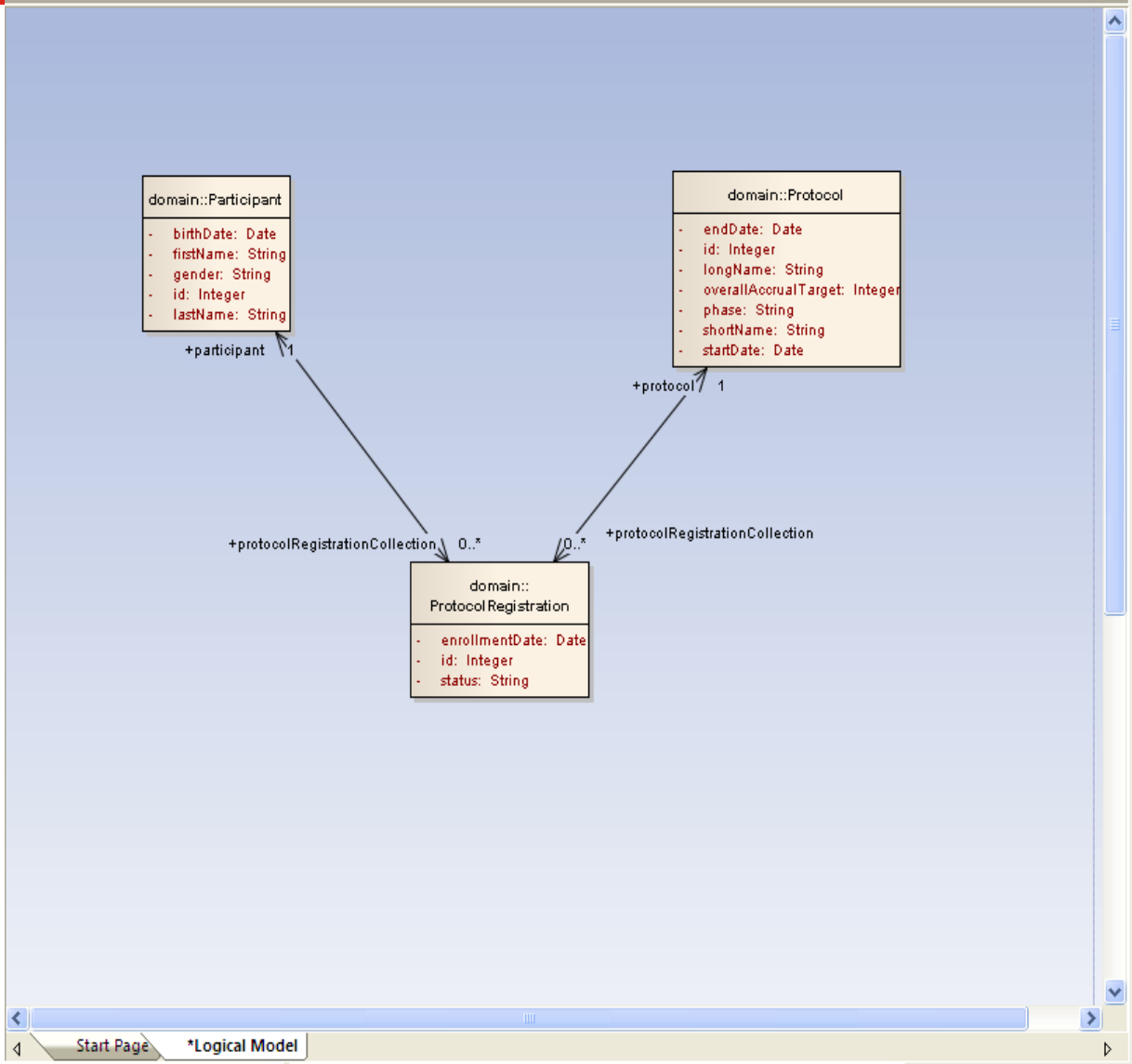
W3C XHTML 1.0 | W3C CSS

A Potential solution

1. Creation of a Meta Data Repository
   - OpenMDR is an existing open source
   - Creation of a context NCS
2. Use of Data Elements to define investigational data
   - caDSR houses more than 38K data elements
3. Create and annotate UML(Unified Modeling Language) models
   - Enterprise Architect has openMDR plugin
4. Study Centers can consume the online services
   - Or files

Creation of Logical Model

Logical Diagram: "Logical Model" created: 2/6/2009 9:55:06 AM modified: 2/6/2009 12:31:10 PM 100% 776 x 1000



Start Page *Logical Model

Project Browser

- Model
  - Logical View
    - Data Model
      - Diagrams
      - «table» PARTICIPANT
      - «table» PROTOCOL
      - «table» PROTOCOL_REGISTRATION
    - Logical Model
      - Diagrams
        - Logical Model
      - edu
        - osu
          - ccts
            - ctsatest
              - domain
                - Participant
                - birthDate
                - firstName
                - gender
                - id

Tagged Values

Data Model

Logical Diagram: "Data Model" created: 2/6/2009 10:10:05 AM modified: 2/6/2009 1:17:51 PM 100% 776 x 1000

Data Model::PARTICIPANT

«column»
*PK ID: NUMERIC(8)
BIRTH_DATE: DATE
FIRST_NAME: VARCHAR(50)
LAST_NAME: VARCHAR(50)
GENDER: VARCHAR(10)

«PK»
+ PK_PARTICIPANT(NUMERIC)

Data Model::PROTOCOL

«column»
*PK ID: NUMERIC(8)
LONG_NAME: VARCHAR(50)
SHORT_NAME: VARCHAR(50)
START_DATE: DATE
END_DATE: DATE
PHASE: VARCHAR(50)
OVERALL_ACCRUAL_TARGET: NUMERIC(8)

«PK»
+ PK_PROTOCOL(NUMERIC)

+PK_PARTICIPANT 1

+PK_PROTOCOL 1

(PARTICIPANT_ID = ID)
«FK»

(PROTOCOL_ID = ID)
«FK»

+FK_PROTOCOL_REGISTRATION_PARTICIPANT 0..*

0..* +FK_PROTOCOL_REGISTRATION_PROTOCOL

Data Model::PROTOCOL_REGISTRATION

«column»
*PK ID: NUMERIC(8)
FK PARTICIPANT_ID: NUMERIC(8)
FK PROTOCOL_ID: NUMERIC(8)
ENROLLMENT_DATE: DATE
STATUS: VARCHAR(50)

«FK»
+ FK_PROTOCOL_REGISTRATION_PARTICIPANT(NUMERIC)
+ FK_PROTOCOL_REGISTRATION_PROTOCOL(NUMERIC)

«PK»
+ PK_PROTOCOL_REGISTRATION(NUMERIC)

Start Page | *Logical Model | *Data Model

Project Browser

- Model
  - Logical View
    - Data Model
      - Diagrams
        - Data Model
      - «table» PARTICIPANT
      - «table» PROTOCOL
      - «table» PROTOCOL_REGISTRATION
    - Logical Model
      - Diagrams
        - Logical Model
      - edu
        - osu
          - ccts
            - ctsatest
              - domain
                - Participant
                - birthDate
                - firstName
                - gender

Tagged Values

Annotating data model with DE

OpenMDR Query Service

MDR Query

Resources

caDSR

Term | Participant Birth Date

Context: | Get List

Search | Query completed

Results

Participant Birth Date java.util.Date (Qualified:RELEASED)
Clinical Trial Participant Birth Date java.util.Date (-:DRAFT NEW)
Participant Birth Date java.util.Date (-:DRAFT NEW)
Animal Participant Birth Date java.util.Date (-:DRAFT NEW)
Abstract Participant Birth Date java.util.Date (-:DRAFT NEW)
Animal Participant Birth Date java.util.Date (-:DRAFT NEW)
Abstract Participant Birth Date java.util.Date (-:DRAFT NEW)

Result Panel- shows all matches

<< | >> | Annotate with CDE | Use

Details

Definition | Props/Values | ObjectClass/Props | Other

data-type	java.util.Date
units	

http://172.16.1.11:9090/wsrf/services/cagri | Connect | Reset

Select Resource to Query against

Select search type

Enter search term

Click Button to show all context's available

Click to Annotate with above selected CDE

Browse tabs for details

Enter Service URL and Click "Connect"

Data Model::PROTOCOL

«column»
*PK ID: NUMERIC(8)
LONG_NAME: VARCHAR(50)
SHORT_NAME: VARCHAR(50)
START_DATE: DATE
END_DATE: DATE
PHASE: VARCHAR(50)
OVERALL_ACCRUAL_TARGET: NUMERIC(8)

«PK»
+ PK_PROTOCOL(NUMERIC)

+PK_PROTOCOL 1

(PROTOCOL_ID = ID)
«FK»

+FK_PROTOCOL_REGISTRATION_PROTOCOL

Project Browser

- Data Model
  - Diagrams
    - Data Model
  - «table» PARTICIPANT
  - «table» PROTOCOL
  - «table» PROTOCOL_REGISTRATION
- Logical Model
  - Diagrams
    - Logical Model
  - edu
    - osu
      - ccts
        - ctsatest
          - domain
            - Participant
              - birthDate
              - firstName
              - gender
              - id
              - lastName

Tagged Values

birthDate (Attribute)

birthDate (Attribute)

birthDate

Exporting Model to XMI

Logical Diagram: "Data Model" created: 2/6/2009 10:10:05 AM modified: 2/8/2010 4:01:38 PM 100% 776 x 1000

Data Model::PARTICIPANT
«column»
*PK ID: NUMERIC(8)
BIRTH_DATE: DATE
FIRST_NAME: VARCHAR(50)
LAST_NAME: VARCHAR(50)
GENDER: VARCHAR(10)
«PK»
+ PK_PARTICIPANT(NUMERIC)

Data Model::PROTOCOL
«column»
*PK ID: NUMERIC(8)
LONG_NAME: VARCHAR(50)

+PK_PARTICIPANT 1

(PARTICIPANT_ID = ID)
«FK»

+FK_PROTOCOL_REGISTRATION_PARTICIPANT 0..*

Data Model::PROTOCO
«column»
*PK ID: NUMERIC(8)
FK PARTICIPANT_ID: NUMERI
FK PROTOCOL_ID: NUMERIC(8
ENROLLMENT_DATE: DAT
STATUS: VARCHAR(50)
«FK»
+ FK_PROTOCOL_REGISTRATION_PARTICIPANT(NUMERIC)
+ FK_PROTOCOL_REGISTRATION_PROTOCOL(NUMERIC)
«PK»
+ PK_PROTOCOL_REGISTRATION(NUMERIC)

Export Package to XMI

Root Package: Logical View

Filename: C:\Documents and Settings\Administrator\Desktop\EA\PlugIn_Test\Test.xmi

Stylesheet: (Optional stylesheet to post process XMI content)

General Options
- [x] Export Diagrams
- [ ] Export Alternate Images
- [x] Format XMI Output
- [x] Write Log file
- [ ] Use DTD
- [ ] Generate Diagram Images

Format:

For Export to Other Tools
- [x] Enable full EA Roundtrip

XMI Type: XMI 1.1

- [ ] Unisys/Rose Format
- [ ] Exclude EA Tagged Values

Warning: These options are for exporting EA model elements to other tools only.

View XMI | Export | Close | Help

Progress

Project Browser
- Model
  - Logical View
    - Data Model
    - Logical Model

ged Values
Logical View (Package)

Start Page | Logical Model | Data Model

Summary

- Program Office can create and annotate UML models using CDEs from caDSR
- Annotated UML models
  - can be used to export XMI filesmto be imported to the data collection tools.
  - Can generate VDR DB structure using Enterprise Architect, Altova UModel etc.
  - Relatively easy mapping and ETL since both use the same UML model
- NCS Study Sites can export semantically aware data/database strucuture and submit to VDR.

Questions

- Umit Topaloglu Ph.D.
  utopaloglu@uams.edu

- William Hogan MD. MS.
  Wrhogan@uams.edu