SEMANTIC WEB IN THE PHARMACEUTICAL INDUSTRY

A LANDSCAPE OF CHALLENGES AND OPPORTUNITIES

- Tim Williams

SWAT4HCLS

Antwerp, Belgium

2018-12-03

OUTLINE

1. Introduction
2. Data
3. PhUSE
4. The Way Forward
5. General Discussion

OUTLINE

1. Introduction

2. Data
3. PhUSE
4. The Way Forward
5. General Discussion

WHO AM I?

Interactive!

Questions and Discussion

WHO I AM

UCB BioSciences

- Statistical Systems Analyst
- Raleigh, North Carolina
- 🇨🇦, 🇺🇸

PhUSE

- Steering Committee: "PhUSE Computational Sciences Symposium"
- Co-lead : "Clinical Trials Data as RDF"*
- Co-lead : "Analysis Results Model (RDF Data Cubes)" (2016)
- Instructor: "Linked Data Hands-on Workshop"*

I ALSO LIKE #LINKEDDATA MEMES

Available
★
OL
Structured
★★
OL RE
Open
★★★
OL RE OF
URIs
★★★★
OL RE OF URI
Linked
★★★★★
OL RE OF URI LD

@NovasTaylor

WHO ARE YOU?

Hands up:

- Pharmaceuticals (any size pharma)
- Biotechnology (non-pharma)
- Health Care
- Research
- Other

WHO ARE YOU : SEM WEB ADOPTION?

Hands up if you are:

- Doing something (personally, professionally) with Semantic Web

Keep your hands up of you are using SW at work in:

- any way: Experiment, Prototype, Proof of Concept, Pilot, Production
- in a Validated Production Environment

OUTLINE

1. Introduction
2. Data
   - 2.1 Landscape
   - 2.2 Standards
3. PhUSE
4. The Way Forward
5. General Discussion

2.1 DATA *LANDSCAPE*

Non-clinical (Pre-clinical)

- Animal studies

Clinical

- Human Study Subjects

Phase	n	Description
0	~ 15	Safety
I	~ 20 - 80	Safety, Dosing
II	~ 100's	Safety, Treat Condition, Refine methods
III	~ 3,000	Efficacy, Double-blind. Comp. other treatments.
IV	1000's	Post-approval. Long term efficacy, safety...

DATA *SOURCES*

Traditional

- Case Report Forms (CRF)
- Electronic Data Capture (EDC)
- * Relational Database Management Systems (RDBMS)

  New

  - Wearables, Ingestibles, Devices
  - Social Media
  - Real World Evidence
    - See: openEHR - The 'open platform' Revolution
      Room A, 17:00-18:00

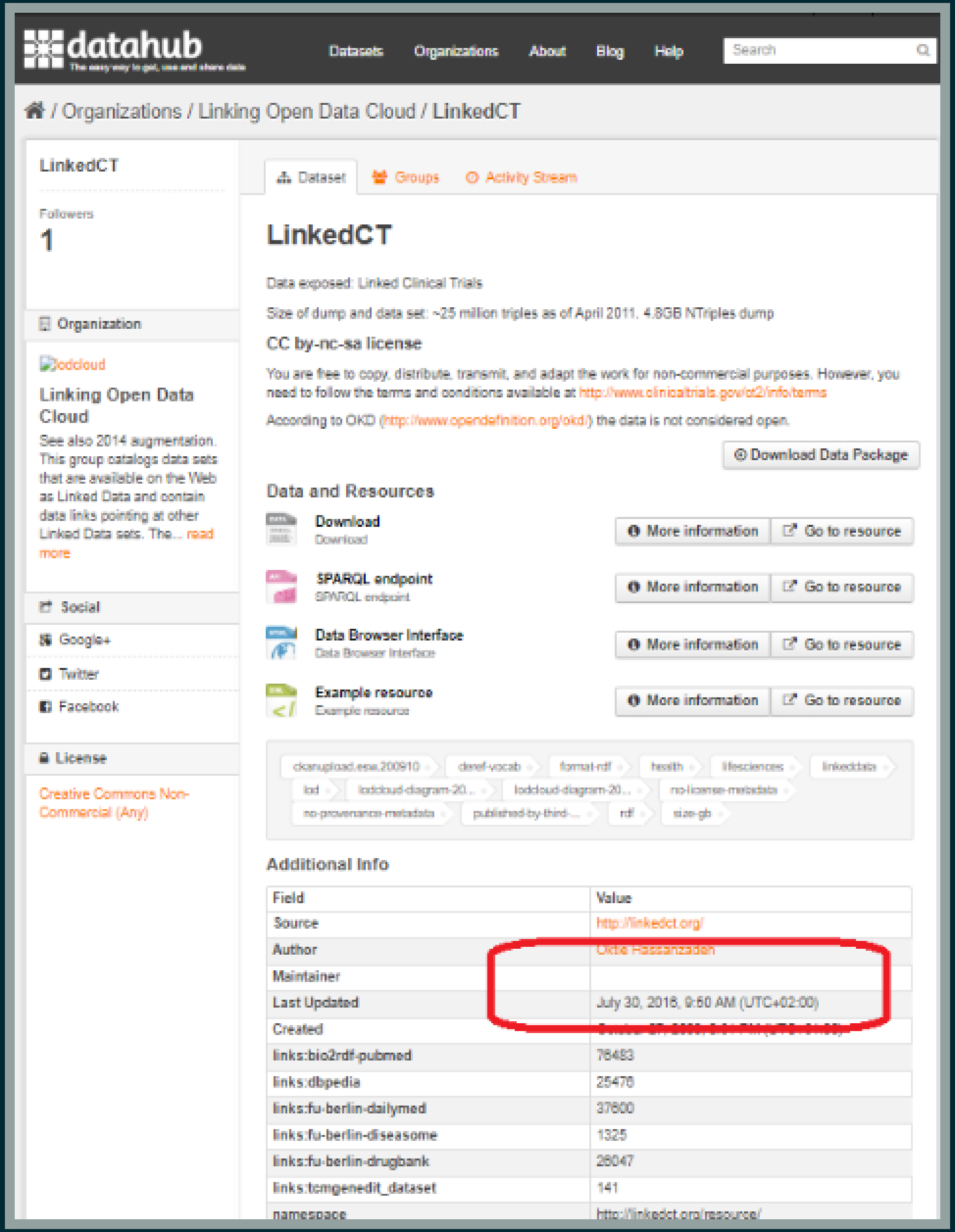
Other Data Sources?

DATA SOURCES (RDF)

RDF ENDPOINTS FOR LATE PHASE DATA?

https://old.datahub.io/dataset/linkedct

Your Experience?



/ Organizations / Linking Open Data Cloud / LinkedCT

LinkedCT

Followers
1

LinkedCT

Data exposed: Linked Clinical Trials

Size of dump and data set: ~25 million triples as of April 2011. 4.8GB NTriples dump

CC by-nc-sa license

You are free to copy, distribute, transmit, and adapt the work for non-commercial purposes. However, you need to follow the terms and conditions available at http://www.clinicaltrials.gov/ct2/info/terms

According to OKD (http://www.opendefinition.org/okd/) the data is not considered open.

Organization

lodcloud

Linking Open Data Cloud

See also 2014 augmentation. This group catalogs data sets that are available on the Web as Linked Data and contain data links pointing at other Linked Data sets. The... read more

Social

License

Creative Commons Non-Commercial (Any)

Data and Resources

- Download
  Download
- SPARQL endpoint
  SPARQL endpoint
- Data Browser Interface
  Data Browser Interface
- Example resource
  Example resource

Additional Info

Field	Value
Source	http://linkedct.org/
Author	Oktie Hassanzadeh
Maintainer	
Last Updated	July 30, 2016, 9:50 AM (UTC+02:00)
Created	[illegible]
links:bio2rdf-pubmed	76483
links:dbpedia	25476
links:fu-berlin-dailymed	37600
links:fu-berlin-diseasome	1325
links:fu-berlin-drugbank	26047
links:tcmgenedit_dataset	141
namespace	http://linkedct.org/resource/

2.2 STANDARDS

HL7® FHIR®

HEALTH LEVEL 7

> *"A set of international standards for transfer of clinical and administrative data between software applications used by various healthcare providers."*

2.2 STANDARDS

FAST HEALTHCARE INTEROPERABILITY RESOURCES

> *"A draft standard describing data formats and elements and an application programming interface for exchanging electronic health records. Created by Health Level Seven."*

FHIR as RDF

Who is using FHIR?

Who is using FHIR as RDF?

Who is attending:

HL7 FHIR and the Semantic Web

Harold Solbrig

Room A, 13:30

cdisc

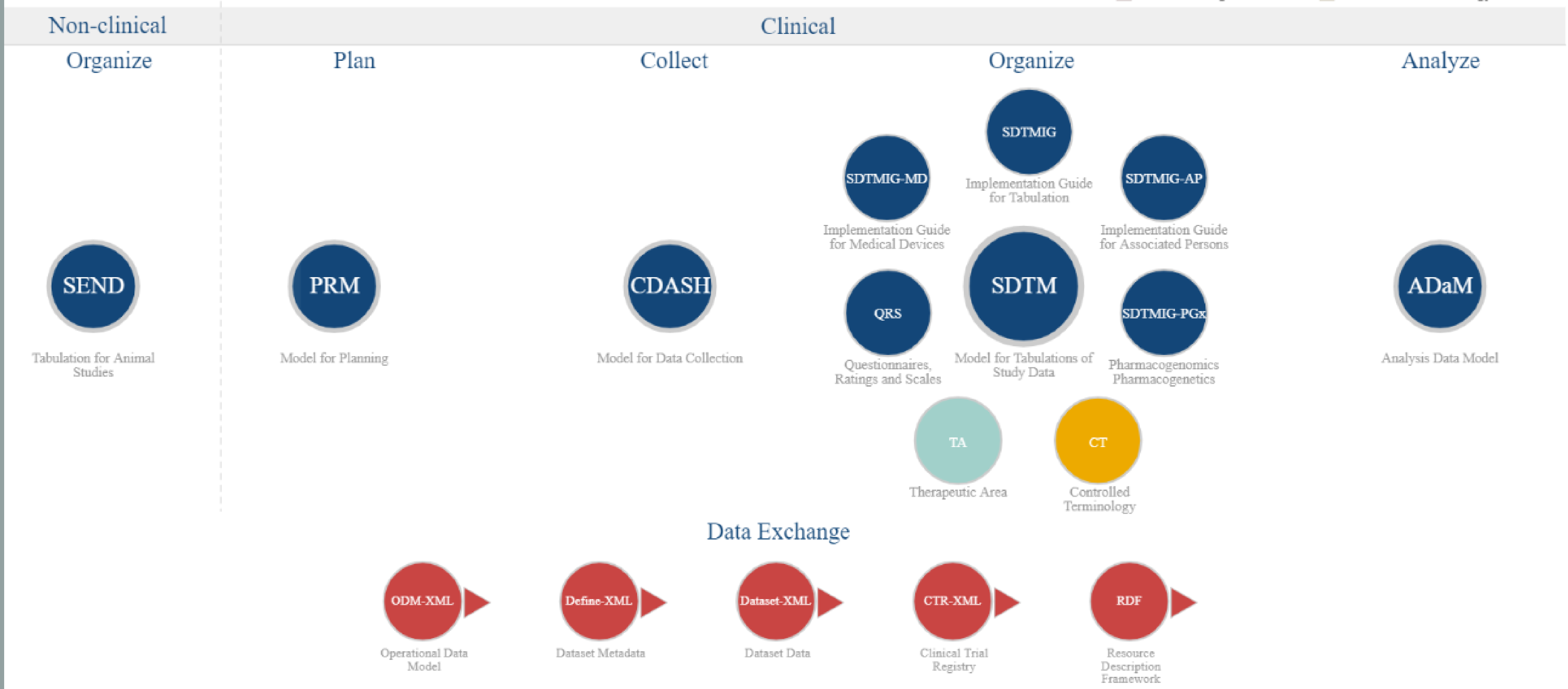
Clinical Data Interchange Standards Consortium

www.cdisc.org

Standards Overview

CDISC Standards in the Clinical Research Process

🟦 Foundational Standard
🟥 Data Exchange
🟩 Therapeutic Area
🟨 Controlled Terminology



Non-clinical

Organize

SEND
Tabulation for Animal Studies

Clinical

Plan

PRM
Model for Planning

Collect

CDASH
Model for Data Collection

Organize

SDTMIG
Implementation Guide for Tabulation

SDTMIG-MD
Implementation Guide for Medical Devices

SDTMIG-AP
Implementation Guide for Associated Persons

QRS
Questionnaires, Ratings and Scales

SDTM
Model for Tabulations of Study Data

SDTMIG-PGx
Pharmacogenomics Pharmacogenetics

TA
Therapeutic Area

CT
Controlled Terminology

Analyze

ADaM
Analysis Data Model

Data Exchange

ODM-XML
Operational Data Model

Define-XML
Dataset Metadata

Dataset-XML
Dataset Data

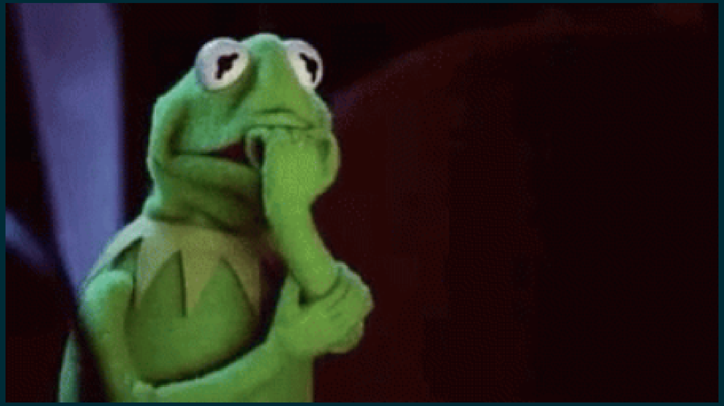
CTR-XML
Clinical Trial Registry

RDF
Resource Description Framework

Are you using CDISC as RDF?

Are you using CDISC?

If you are in Pharma and not using CDISC Standards, I am worried about you.



CDISC STANDARDS ARE A GOOD THING

BUT THERE ARE PROBLEMS

AND CHANGE IS NEEDED

SDTM DOMAINS

- Demographics (DM)
- Vital Signs (VS)
- Adverse Events (AE)
- ...

"23 defined domains within six broad categories." (SDTM 3.1)

PROBLEMS IN CDISC SDTM

> *"Domains represent discrete categories" - CDISC*

Reality Check: They do not.

- Example: Demographics Domain (DM)

  Also contains

  - Study ID
  - Treatment Arm Information (arm, coded value for arm)
  - Age units

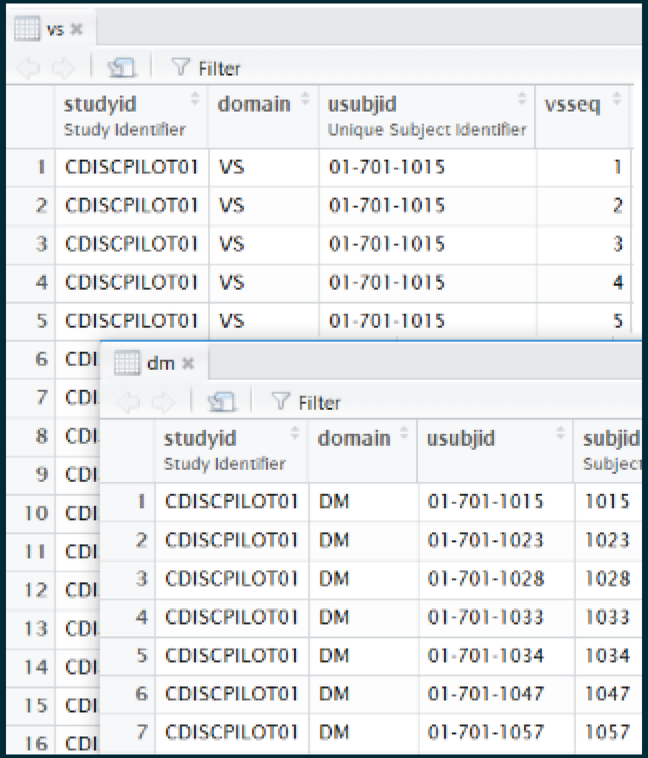
PROBLEMS IN CDISC SDTM

- Multiple approaches to represent Medical conditions
  - Medical History (MH)
  - Adverse Events (AE)
  - Clinical Events (CE)
- Multiple locations for same/similar information

  Death Information:
  - Demographics (DM)
  - Disposition (DS)
  - Adverse Events (AE)
- ...and more.

PROBLEMS IN CDISC SDTM

- Data Repetition
- Row-by-Column Structure



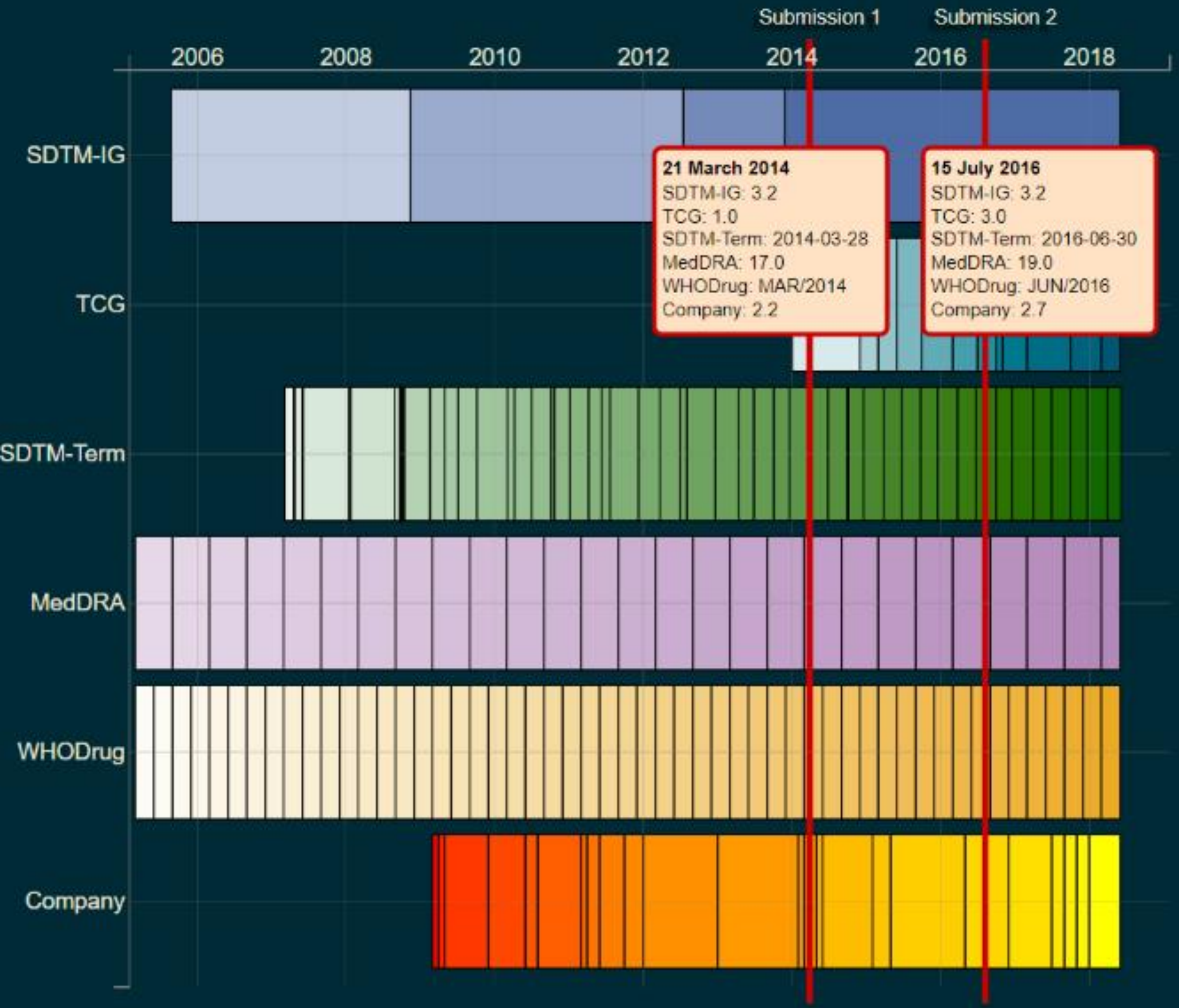
	studyid Study Identifier	domain	usubjid Unique Subject Identifier	vsseq
1	CDISCPILOT01	VS	01-701-1015	1
2	CDISCPILOT01	VS	01-701-1015	2
3	CDISCPILOT01	VS	01-701-1015	3
4	CDISCPILOT01	VS	01-701-1015	4
5	CDISCPILOT01	VS	01-701-1015	5
6	CDI			
7	CDI			
8	CDI			
9	CDI			
10	CDI			
11	CDI			
12	CDI			
13	CDI			
14	CDI			
15	CDI			
16	CDI			

	studyid Study Identifier	domain	usubjid	subjid Subject
1	CDISCPILOT01	DM	01-701-1015	1015
2	CDISCPILOT01	DM	01-701-1023	1023
3	CDISCPILOT01	DM	01-701-1028	1028
4	CDISCPILOT01	DM	01-701-1033	1033
5	CDISCPILOT01	DM	01-701-1034	1034
6	CDISCPILOT01	DM	01-701-1047	1047
7	CDISCPILOT01	DM	01-701-1057	1057

THE VERSIONING PROBLEM

- Standards Change over time
- Version-Conversion
  - Instance data is not version-independent

STANDARDS OVER TIME



21 March 2014
SDTM-IG: 3.2
TCG: 1.0
SDTM-Term: 2014-03-28
MedDRA: 17.0
WHODrug: MAR/2014
Company: 2.2

15 July 2016
SDTM-IG: 3.2
TCG: 3.0
SDTM-Term: 2016-06-30
MedDRA: 19.0
WHODrug: JUN/2016
Company: 2.7

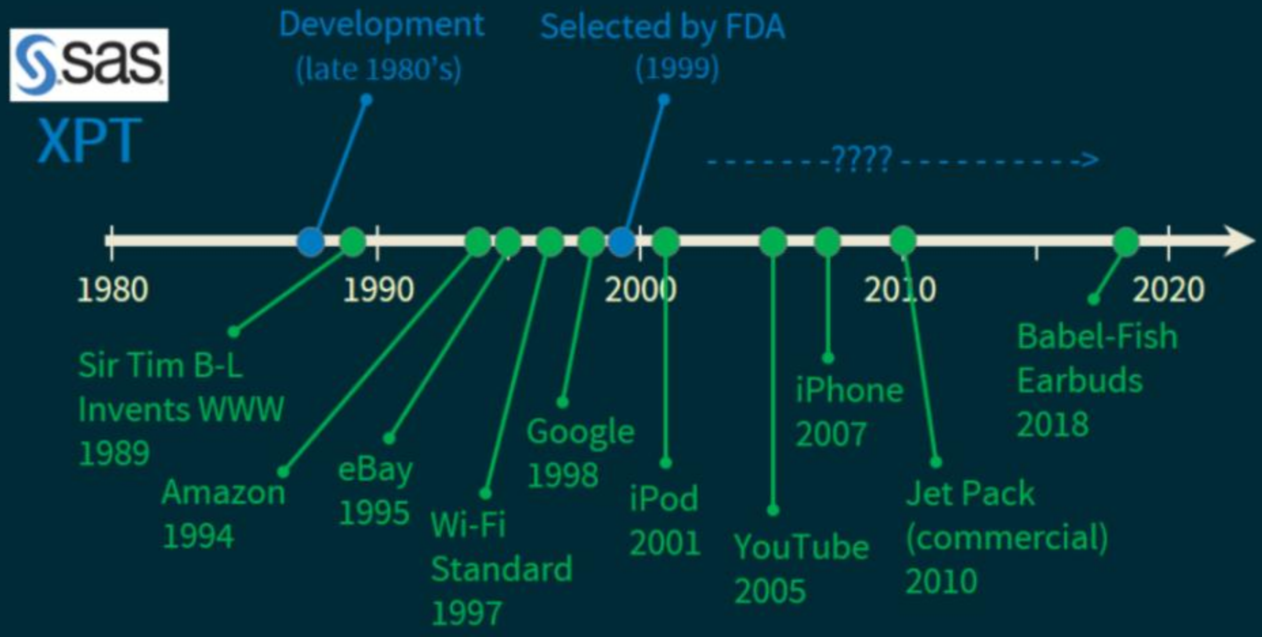
LEGEND

STDM-IG	Study Data Tabulation Model (SDTM), Implementation Guide
TCG*	Study Data Technical Conformance Guide
SDTM-TERM**	SDTM terminology
MedDRA	Medical Dictionary for Regulatory Activities
WHODrug	World Health Organization Drug Dictionary
Company	Fictional company standard.

*https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm#guides

**https://evs.nci.nih.gov/ftp1/CDISC/SDTM/Archive/

A TECHNOLOGY TIMELINE: XPT FORMAT FOR FDA SUBMISSIONS



IT GETS WORSE…

THE 30 YEAR-OLD XPT FORMAT FOR FILE TRANSFER…

IS BEING USED AS A STRUCTURE FOR DATA STORAGE

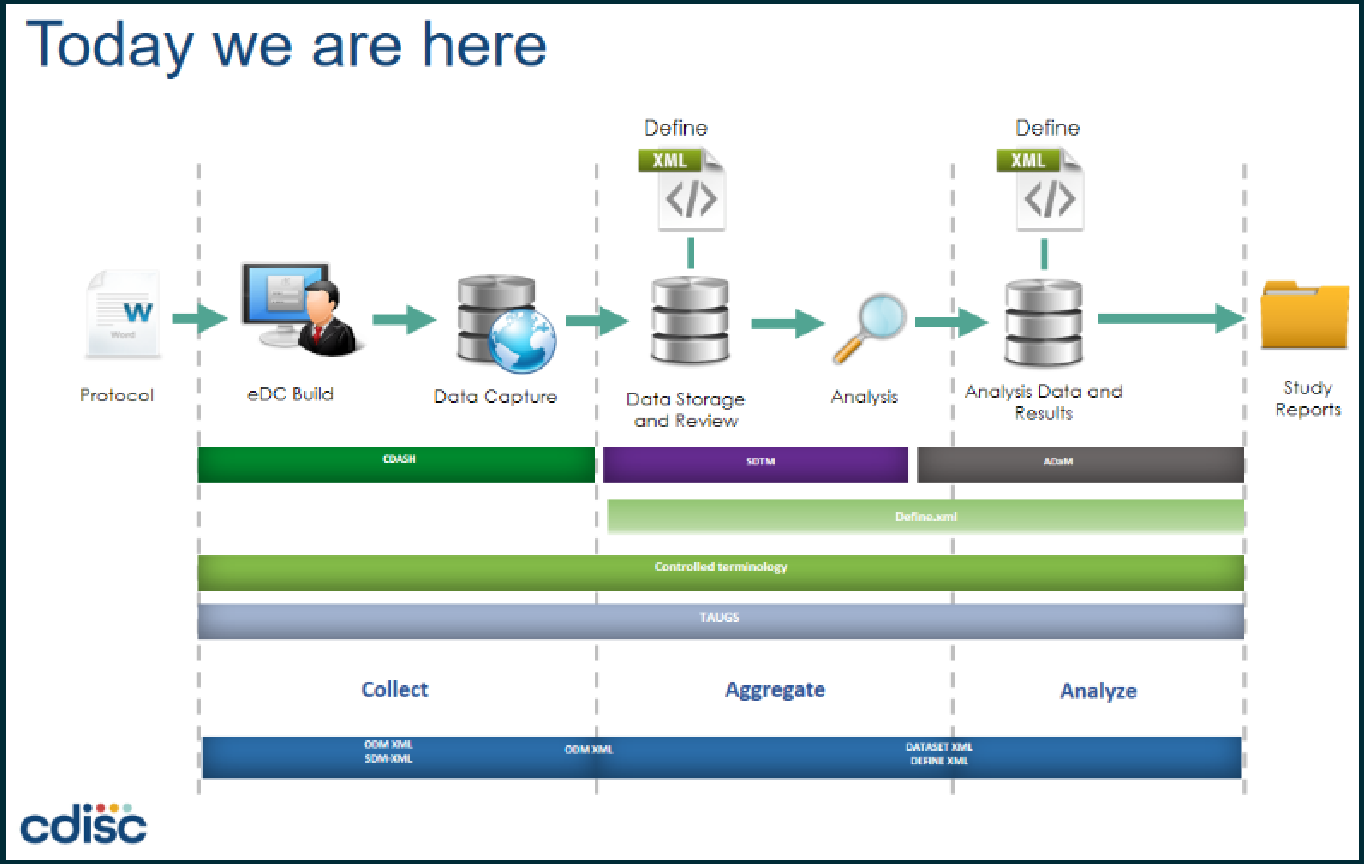
Have you converted data from one CDISC SDTM or ADaM version to another?

How can we replace XPT files?

CDISC IS TRYING TO CHANGE

CDISC PROOF OF CONCEPT

"Evolving our standards towards end to end automation"

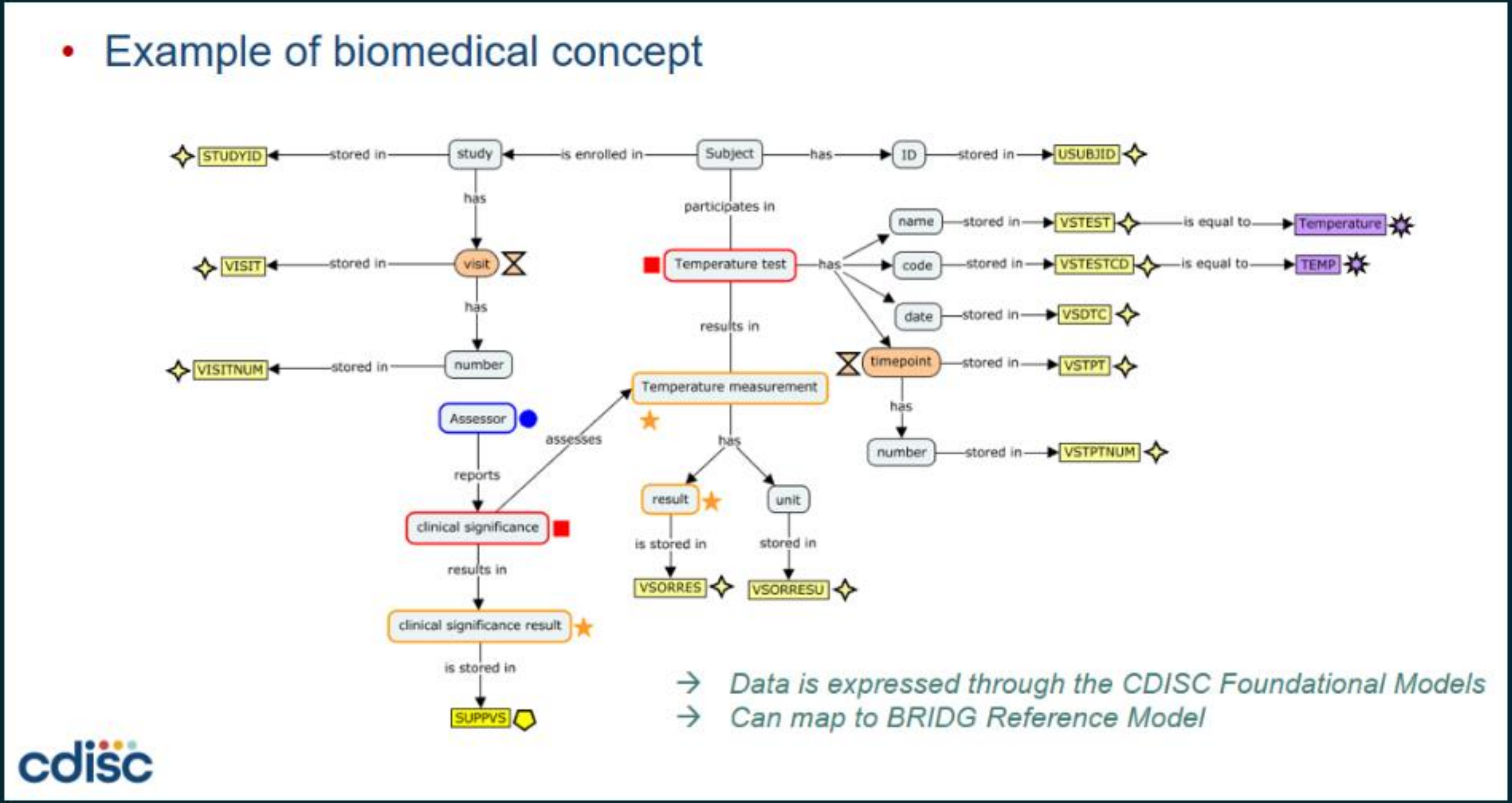


Today we are here

Image courtesy of P. van Reusel, Sam Hume. "CDISC Proof of Concept: Evolving our standards toward end to end auotmation."

CDISC PROOF OF CONCEPT

STANDARDS IN CONCEPT MAPS



- Example of biomedical concept

→ *Data is expressed through the CDISC Foundational Models*
→ *Can map to BRIDG Reference Model*

Image courtesy of P. van Reusel, Sam Hume. "CDISC Proof of Concept: Evolving our standards toward end to end auotmation."

Will CDISC Succeed?

OUTLINE

1. Introduction
2. Data

3. PhUSE

3.1 What is PhUSE?

3.2 PhUSE Linked Data Initiatives

3.3 CTDasRDF Project

- The Way Forward
- General Discussion

Pharmaceutical Users Software Exchange

Mission:

Provide an welcoming, **neutral platform** *for creating and sharing ideas, implementing data standards, processes, and tools, and exploring* ***innovative methodologies, techniques, and technologies****.*

phuse

Pharmaceutical Users Software Exchange

Working Groups Mission:

Provide an open, transparent, and ***collaborative*** *forum in an* ***non-competitive*** *environment in which Regulators, Life Science Companies, Technology Providers, SDOs, and Academia can address unmet computational science needs impacting product development and regulatory review as to improve human health*

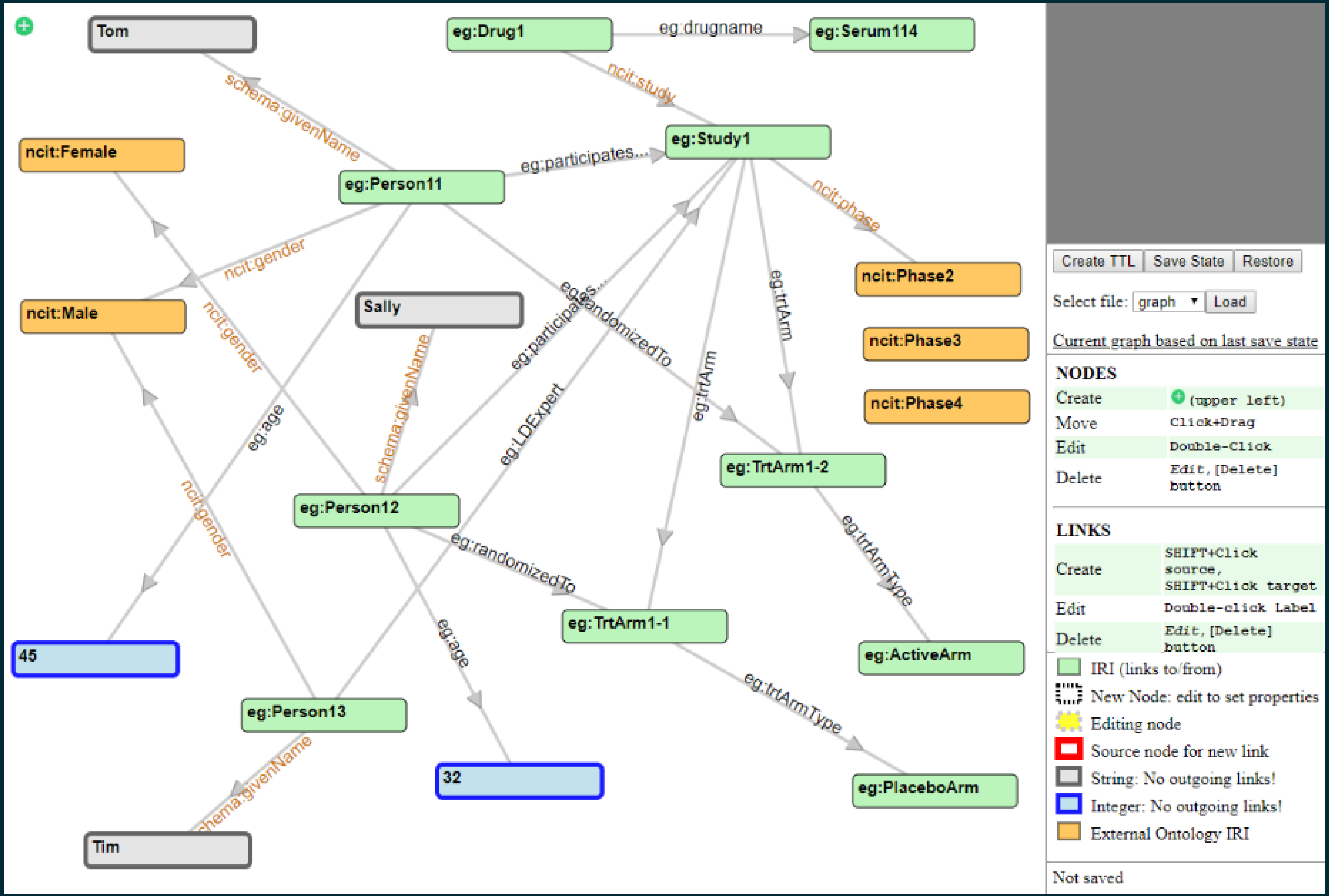
* - emphasis is mine

phuse

Pharmaceutical Users Software Exchange

- Membership: >8,700 spanning 30 countries
- Annual Conference: EUConnect, USConnect
- Single Day Events
- Computational Sciences Symposium (CSS)
  - A "working" conference

HANDS-ON WORKSHOP: GRAPH EDITOR



NODES

Create	⊕ (upper left)
Move	Click+Drag
Edit	Double-Click
Delete	Edit, [Delete] button

LINKS

Create	SHIFT+Click source, SHIFT+Click target
Edit	Double-click Label
Delete	Edit, [Delete] button

🟩 IRI (links to/from)

⬚ New Node: edit to set properties

🟨 Editing node

🟥 Source node for new link

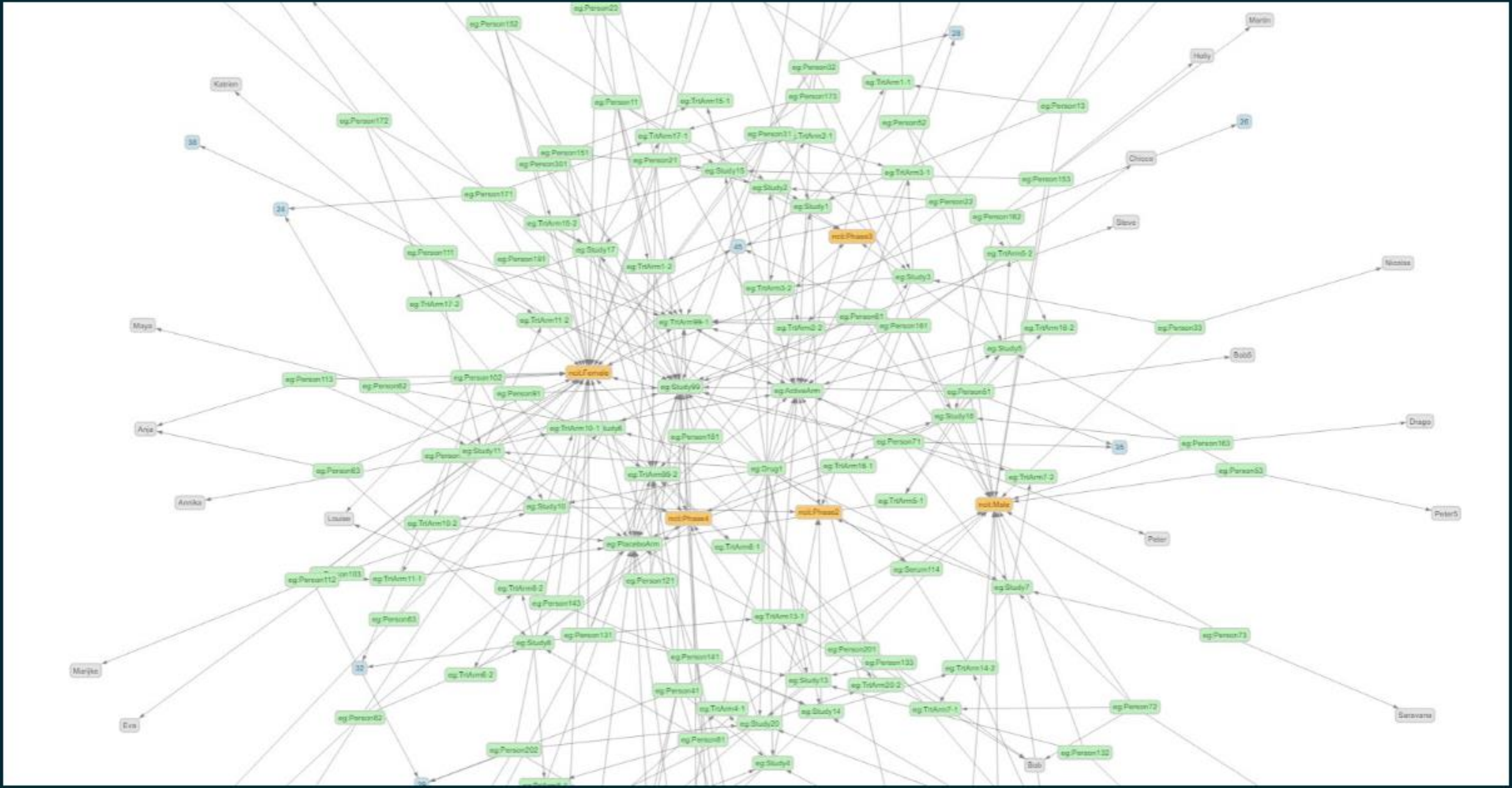
⬜ String: No outgoing links!

🟦 Integer: No outgoing links!

🟧 External Ontology IRI

Not saved

HANDS-ON WORKSHOP: 21 MERGED STUDIES



PHUSE SEMANTIC WEB (LINKED DATA) PROJECTS

Completed:

- CDISC Foundational Standards in RDF
- CDISC Conformance Checks (incomplete? Last update 2014?)
- Reusing Medical Summaries for Enabling Clinical Research [Demo, P.O.C]
- Analysis Results and Metadata (2016) [P.O.C]

PHUSE SEMANTIC WEB (LINKED DATA) PROJECTS

Past

- Regulatory Guidance in RDF (incomplete?)
- Clinical Program Design in RDF (incomplete?)
- CDISC Protocol Representation Model in RDF (on hold [indefinitely?])

Current

- Clinical Trials Data as RDF
- Understanding RDF/Linked Data for Nonclinical Use [NEW]

OBSERVATION:

CDISC AND PHUSE PROJECTS HAVE (MOSTLY) BEEN MODELING THE DATA STANDARDS

What is fundamental problem with this approach?

It does not model the clinical trial ***data***.

Proposal:

- Model the Clinical Trial *proess* and *instance data*
- Build the standards, data checks, etc. - *into that model*
- Instance data independent from Industry Standards
    - *Materialize instance data into a Standard*

OUTLINE

1. Introduction
2. Data
3. PhUSE

4. The Way Forward

4.1 Roofshot Manifesto

4.1.1 Roofshot: Study URI

4.1.2 Roofshot: SDTM Domains as RDF

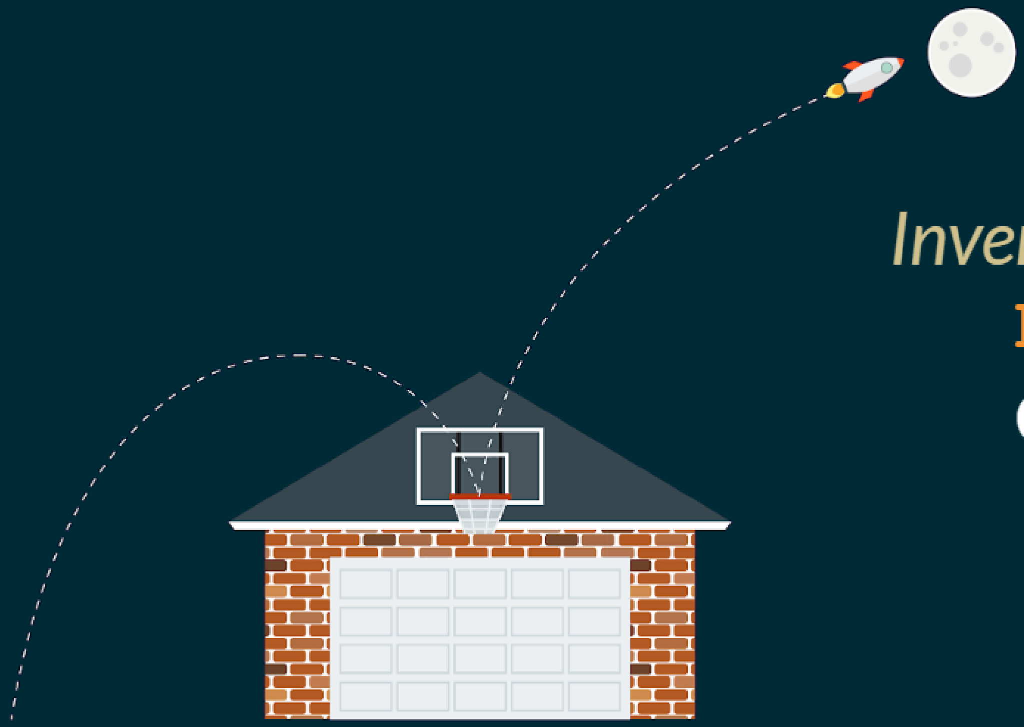
4.1.3 Roofshot: Open Source Onotology Development

5. General Discussion

R.O.I UNICORN

Image Attribution: https://bit.ly/2x0Hjmd

4.1 THE ROOFSHOT / MOONSHOT MANIFESTO



Moonshot

Invent & apply state-of-the-art

`Knowledge Graph`

Clinical trial lifecycle

Roofshot

Incremental impacts

- Study URI
- CTD (SDTM) as RDF
- Open Onotology Development

Concept & Image Attribution: https://rework.withgoogle.com/blog/the-roofshot-manifesto/

4.1.1 *ROOFSHOT:* STUDY URI AS AN INDUSTRY STANDARD (PROPOSAL)

> *"Study URI" - K. Forsberg, D. Goude. PhUSE EUConnect18.*

...and additional followup by J. Ulander (A3), T. Williams (UCB)

Why?

- Easy entrypoint for Pharma
- Familiar concept: NCT ID
    - CT.gov must first review and approve Protocol

STUDY URI COMPONENTS

https://data.pharma.abc/clinicaltrial/D3562C00096

1. Global Namespace
2. Resource type
3. Trial ID

Is anyone using a Study URI/IRI?

STUDY URI: GLOBAL NAMESPACE

https://data.pharma.abc/clinicaltrial/D3562C00096

- Company web URL
- URIs that de-reference: External/Internal

Discuss

STUDY URI: RESOURCE TYPE

https://data.pharma.abc/clinicaltrial/D3562C00096

- Easy? What else could it be called?
- Implications? Link to ontology?

Discuss?

STUDY URI: TRIAL ID

https://data.pharma.abc/clinicaltrial/D3562C00096

1. NCT ID available (ClinicalTrials.gov)
2. NCT ID not available: Unique Company ID (Company guidance)

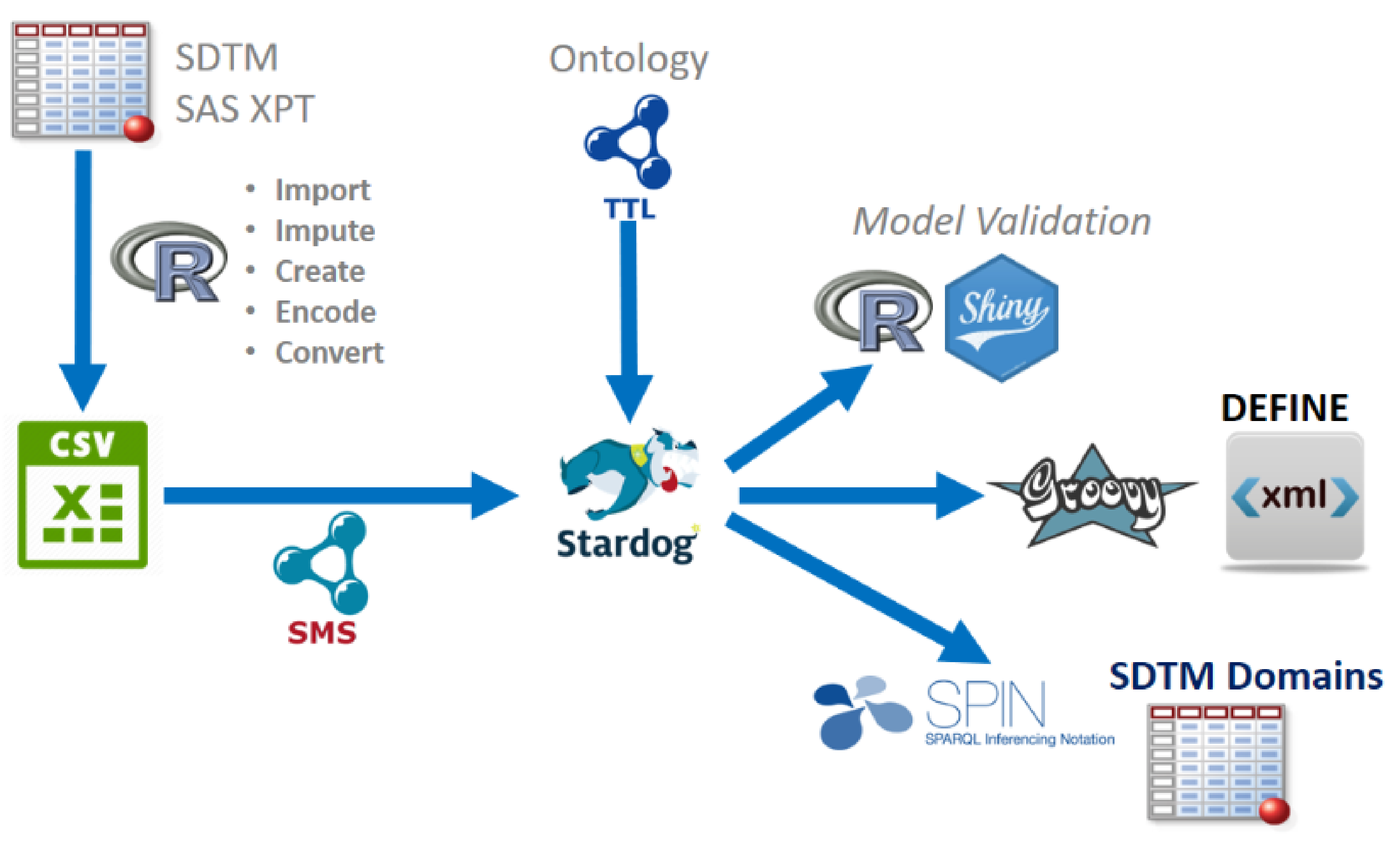
Discuss

STUDY URI: NEXT STEPS

- Review and comment at: https://github.com/phuse-org/LinkedDataEducation/blob/master/doc/StudyURI.md
- Invite comment from FDA, EMA, PMDA, CDISC... *You* !

4.1.2 *ROOFSHOT:* CTD (SDTM) AS RDF

PHUSE PROJECT: CLINICAL TRIALS DATA AS RDF



CTD AS RDF *PROJECT PHILOSOPHY*

DO NOT MODEL:

- Industry Standards

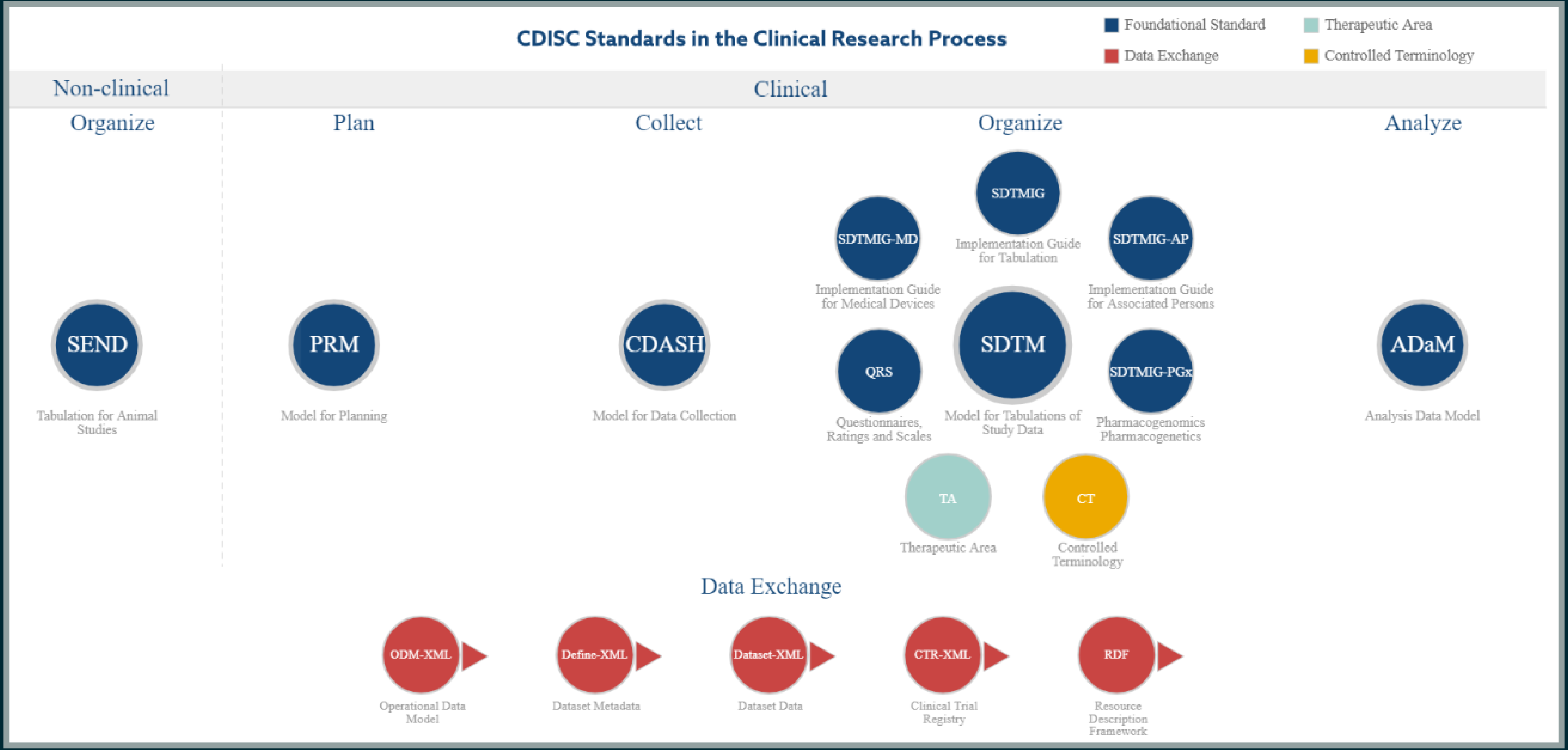
DO MODEL:

- Clinical trial process
- Data
- Rules

IMPLICATIONS "UP STREAM" FROM CLINICAL STUDIES

PHUSE PRE-CLINICAL JOINS CLINICAL RDF PROJECT!

Common Concepts: Pre-Clinical & Clinical Research



CDISC Standards in the Clinical Research Process

🟦 Foundational Standard
🟩 Therapeutic Area
🟥 Data Exchange
🟨 Controlled Terminology

Non-clinical

Organize

SEND
Tabulation for Animal Studies

Clinical

Plan

PRM
Model for Planning

Collect

CDASH
Model for Data Collection

Organize

SDTMIG
Implementation Guide for Tabulation

SDTMIG-MD
Implementation Guide for Medical Devices

SDTMIG-AP
Implementation Guide for Associated Persons

SDTM
Model for Tabulations of Study Data

QRS
Questionnaires, Ratings and Scales

SDTMIG-PGx
Pharmacogenomics Pharmacogenetics

TA
Therapeutic Area

CT
Controlled Terminology

Analyze

ADaM
Analysis Data Model

Data Exchange

ODM-XML
Operational Data Model

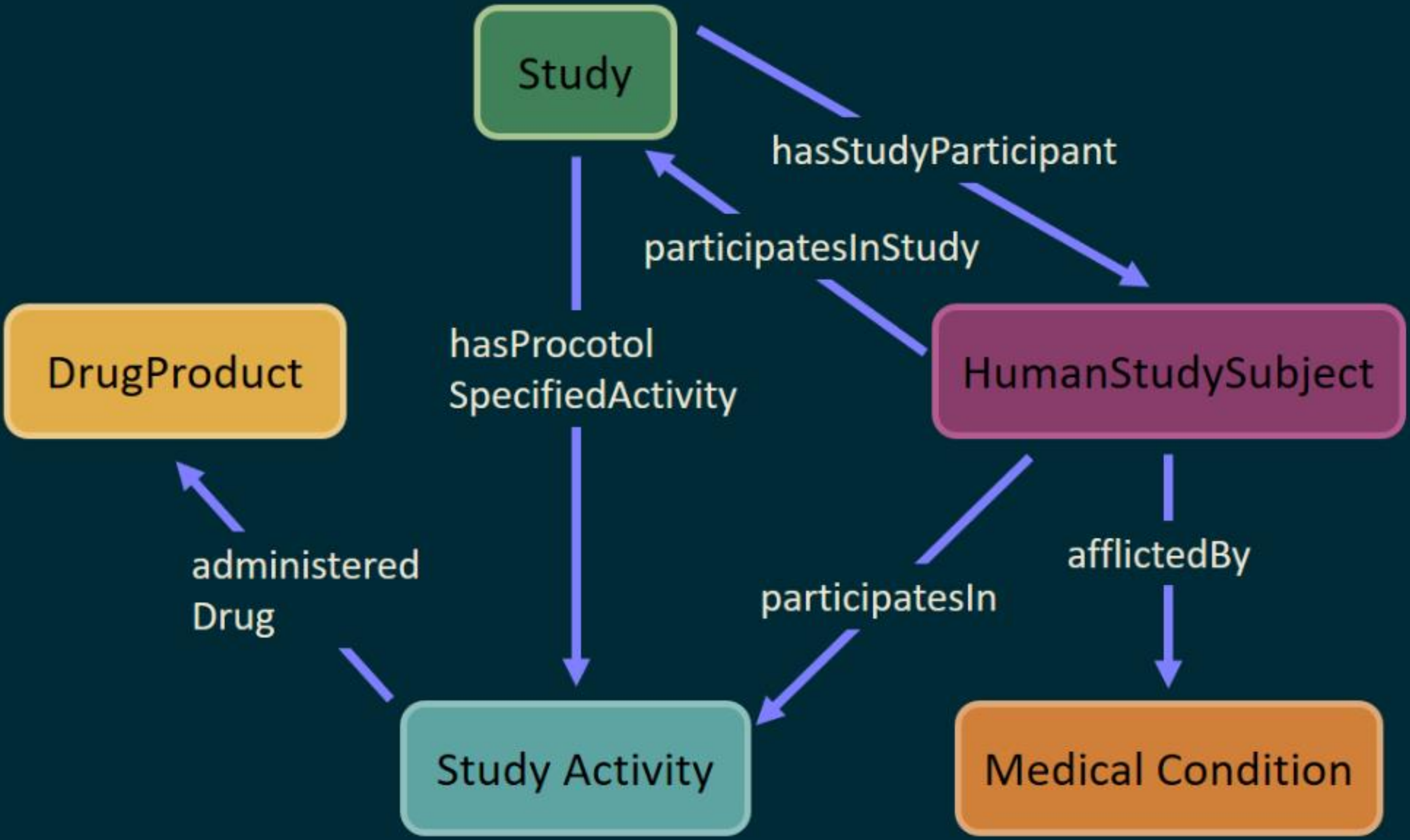
Define-XML
Dataset Metadata

Dataset-XML
Dataset Data

CTR-XML
Clinical Trial Registry

RDF
Resource Description Framework

CORE STUDY 'MINI' ONTOLOGY



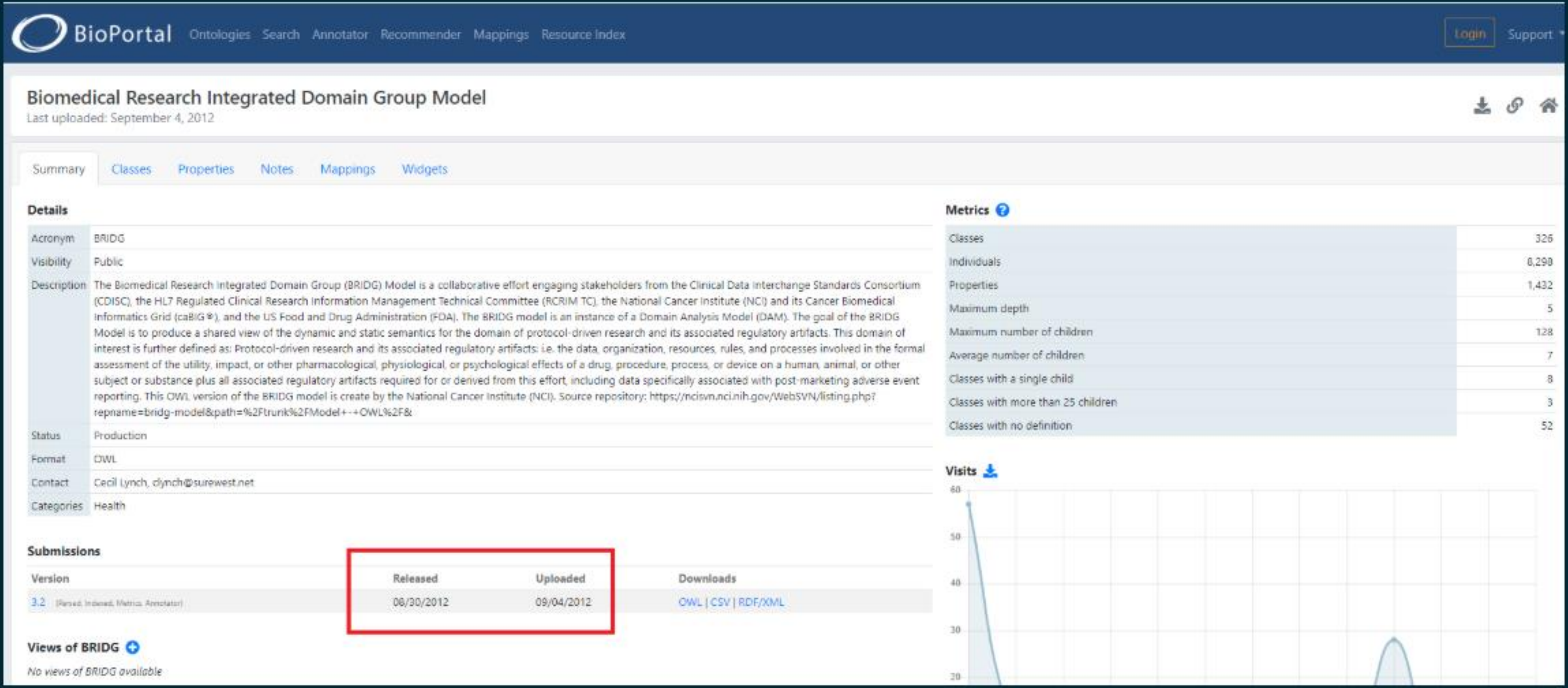
LEVERAGE EXISTING ONTOLOGIES

LEVERAGE EXISTING ONTOLOGIES

BIOMEDICAL RESEARCH INTEGRATED DOMAIN GROUP MODEL (BRIDG)

Collaboration:

- CDISC, HL7, NCI, caBIG, FDA
- OWL version from NCI
- Version 3.2 as RDF. Current is 5.x?



Biomedical Research Integrated Domain Group Model

Last uploaded: September 4, 2012

Details

Acronym	BRIDG
Visibility	Public
Description	The Biomedical Research Integrated Domain Group (BRIDG) Model is a collaborative effort engaging stakeholders from the Clinical Data Interchange Standards Consortium (CDISC), the HL7 Regulated Clinical Research Information Management Technical Committee (RCRIM TC), the National Cancer Institute (NCI) and its Cancer Biomedical Informatics Grid (caBIG®), and the US Food and Drug Administration (FDA). The BRIDG model is an instance of a Domain Analysis Model (DAM). The goal of the BRIDG Model is to produce a shared view of the dynamic and static semantics for the domain of protocol-driven research and its associated regulatory artifacts. This domain of interest is further defined as: Protocol-driven research and its associated regulatory artifacts: i.e. the data, organization, resources, rules, and processes involved in the formal assessment of the utility, impact, or other pharmacological, physiological, or psychological effects of a drug, procedure, process, or device on a human, animal, or other subject or substance plus all associated regulatory artifacts required for or derived from this effort, including data specifically associated with post-marketing adverse event reporting. This OWL version of the BRIDG model is create by the National Cancer Institute (NCI). Source repository: https://ncisvn.nci.nih.gov/WebSVN/listing.php?repname=bridg-model&path=%2Ftrunk%2FModel+-+OWL%2F&
Status	Production
Format	OWL
Contact	Cecil Lynch, clynch@surewest.net
Categories	Health

Submissions

Version	Released	Uploaded	Downloads
3.2 (Parsed, Indexed, Metrics, Annotator)	08/30/2012	09/04/2012	OWL | CSV | RDF/XML

Views of BRIDG

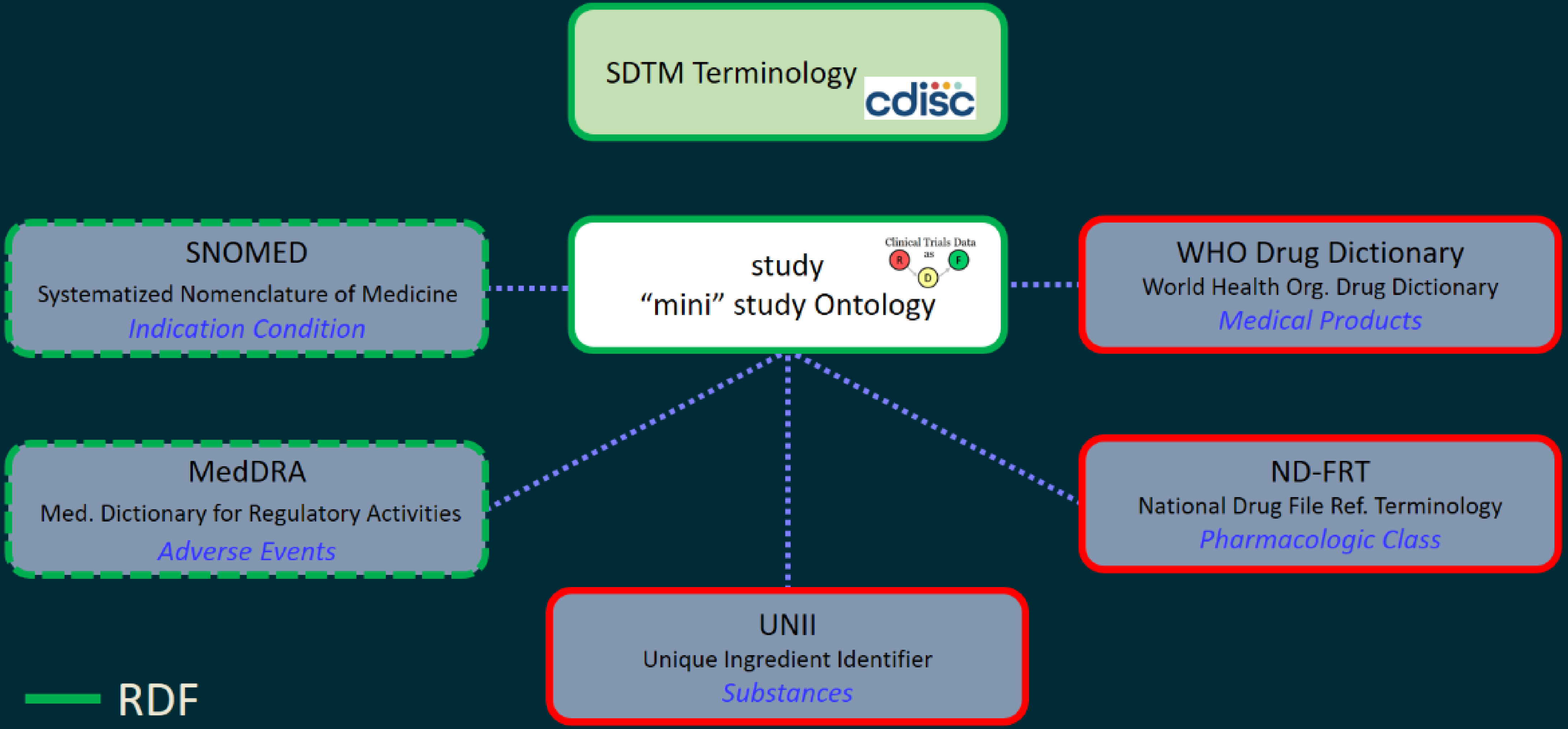
No views of BRIDG available

Metrics

Classes	326
Individuals	8,290
Properties	1,432
Maximum depth	5
Maximum number of children	128
Average number of children	7
Classes with a single child	8
Classes with more than 25 children	3
Classes with no definition	52

Visits

LEVERAGE EXISTING STANDARDS



SDTM Terminology

study
"mini" study Ontology

SNOMED
Systematized Nomenclature of Medicine
Indication Condition

MedDRA
Med. Dictionary for Regulatory Activities
Adverse Events

WHO Drug Dictionary
World Health Org. Drug Dictionary
Medical Products

ND-FRT
National Drug File Ref. Terminology
Pharmacologic Class

UNII
Unique Ingredient Identifier
Substances

🟩 RDF

🟥 non-RDF

4.1.3 *ROOFSHOT:* OPEN SOURCE ONTOLOGY DEVELOPMENT

Can an individual developer, project team, company, standards org., or regulatory org. create a solution for the industry?

> *"We cannot compete with centralized systems unless we collaborate."*
>
> *- Ruben Verborgh, Decentralizing the Semantic Web Through Incentivized Collaboration*

OPEN SOURCE MODEL FOR CLINICAL TRIAL ONTOLOGIES DEVELOPMENT

- Ontologies on GitHub?
- Cooperation in the pre-competitive space
  - PhUSE?
  - TransCelerate?
    - Common Protocol Template (not in RDF!)
  - CDISC?

Discuss

OPEN SOURCE ONTOLOGY

CHALLENGES

- Gate keeper?
- Will companies:
  - Participate?
  - Give back?
- Conflict resolution (approach, code)
- Volunteers

Is this feasible?

ONTOLOGY MAINTENANCE AND DISTRIBUTION

Will the Open Biological and Biomedical Ontology (OBO) approach work?

The OBO Foundry

Post-development curation?

ONTOLOGY MAINTENANCE AND DISTRIBUTION

Don't hide my OWL behind an API!

OUTLINE 5

1. Introduction
2. Data
3. PhUSE
4. The Way Forward

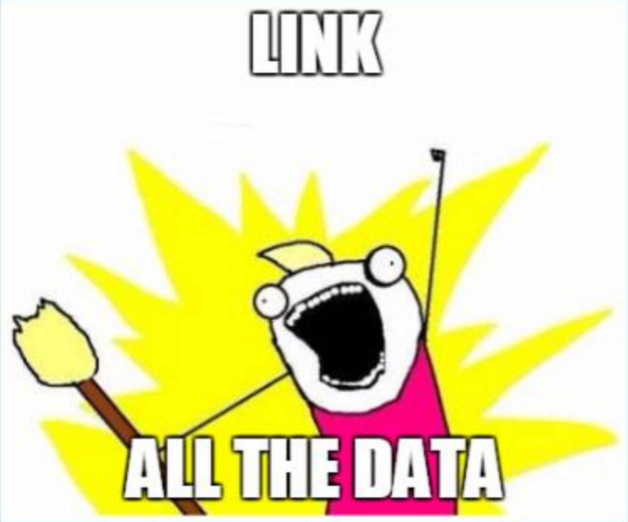
5. General Discussion

ADDITIONAL DISCUSSION POINTS

DISCUSSION:

How are we hindering our own progress?

- "High Priesthood"
- Too much emphasis on "*Linking all the things*?"



- Poor communication, translation to ROI?

What are we doing right?

DISCUSSION:

What are our main challenges in Pharma?

- Momentum of legacy technology
- Skill set, lack of knowledge
- Politics: Who owns innovation?
    - IT
    - Analytics
    - The "Business"

DISCUSSION:

Which is the best environment for SW in BioTech/Pharma?

- Startups, Small Pharma
- Mid-Sized
- Large Pharma

DISCUSSION:

What are you using for validation (& why?)

- SPIN
- SHEX
- SHACL
- OTHER?

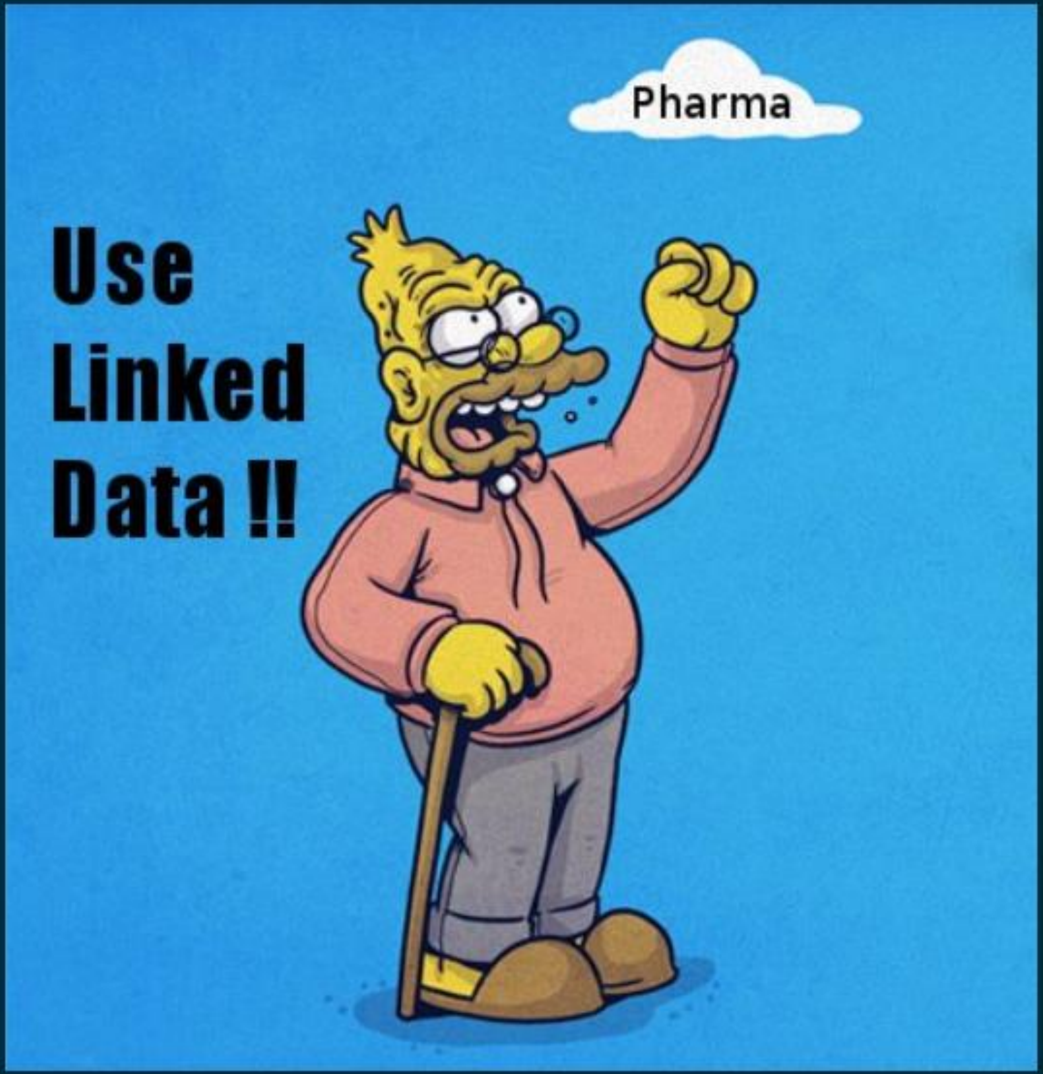
DISCUSSION:

What are you using for visualization?

- Commercial Applications
- Open Source Tools
- Bespoke
  - Python
  - Javascript (D3JS, other?)
  - R, RShiny
  - Other?

How do we make the Semantic Web happen in Pharma?

CONCLUSION



Thank you!