LINKED DATA FOR CLINICAL TRIALS HANDS-ON WORKSHOP

PHUSE CSS 2018

Silver Spring MD

04 March, 2018

INSTRUCTORS

Tim Williams

Principal Statistical Systems Analyst
UCB BioSciences
tim.williams@PhUSE.eu

Johannes Ulander

Principal Consultant
S-Cubed ApS
ju@s-cubed.dk

Assisted by:

- AJ Cook
- Stephen Nowell

Workshop Files, Presentation PDF:

https://github.com/phuse-org/LinkedDataWorkshop/CSS2018

(for later)

OUTLINE

0. (very brief) Introduction to Linked Data
1. Create a Study Graph
2. Query Graph Data
3. Ontology and Inference
4. Merge Studies
5. Linked Data in the Real World

PREPARATION

- Your laptop [Power up!]
- Copy of:
  1. Exercises
  2. Graph Editor Introduction
  3. Info sheet
  4. SPARQL reference
- Log in to Cloud Server

OUTLINE

0. Introduction to Linked Data

1. Create a Study Graph
2. Query Graph Data
3. Ontology and Inference
4. Merge Studies
5. Linked Data in the Real World

DIFFERENT TYPES OF LINKED DATA

Property Graph

Apache TinkerPop

neo4j

Resource Description Framework (**RDF**)

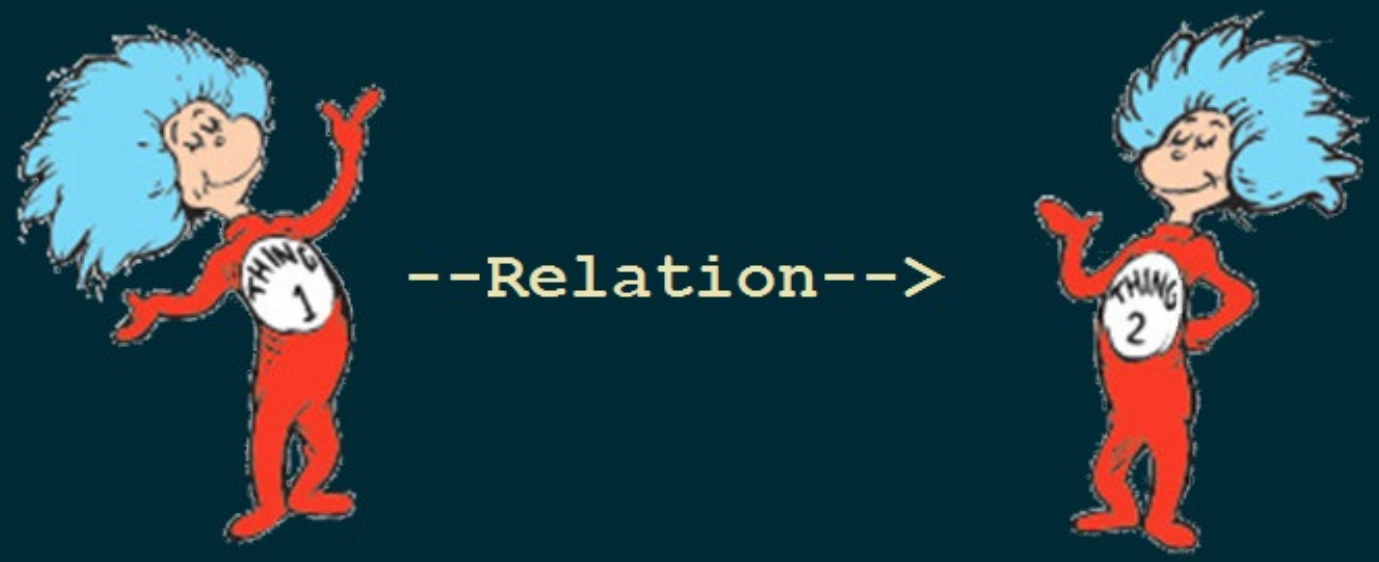
Semantic Web

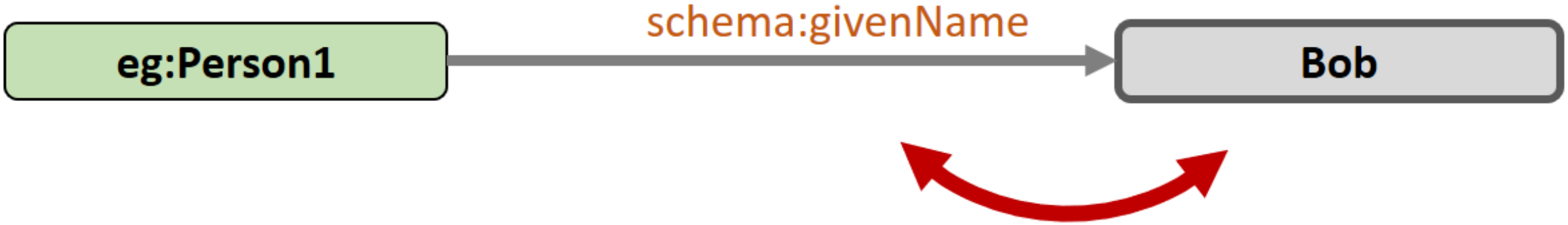
R D F

SPARQL

Stardog

DATA AS A GRAPH?





Identifier	Key=Value Pair
Person 1	givenName='Bob'

" Person1 *has given name* 'Bob' "

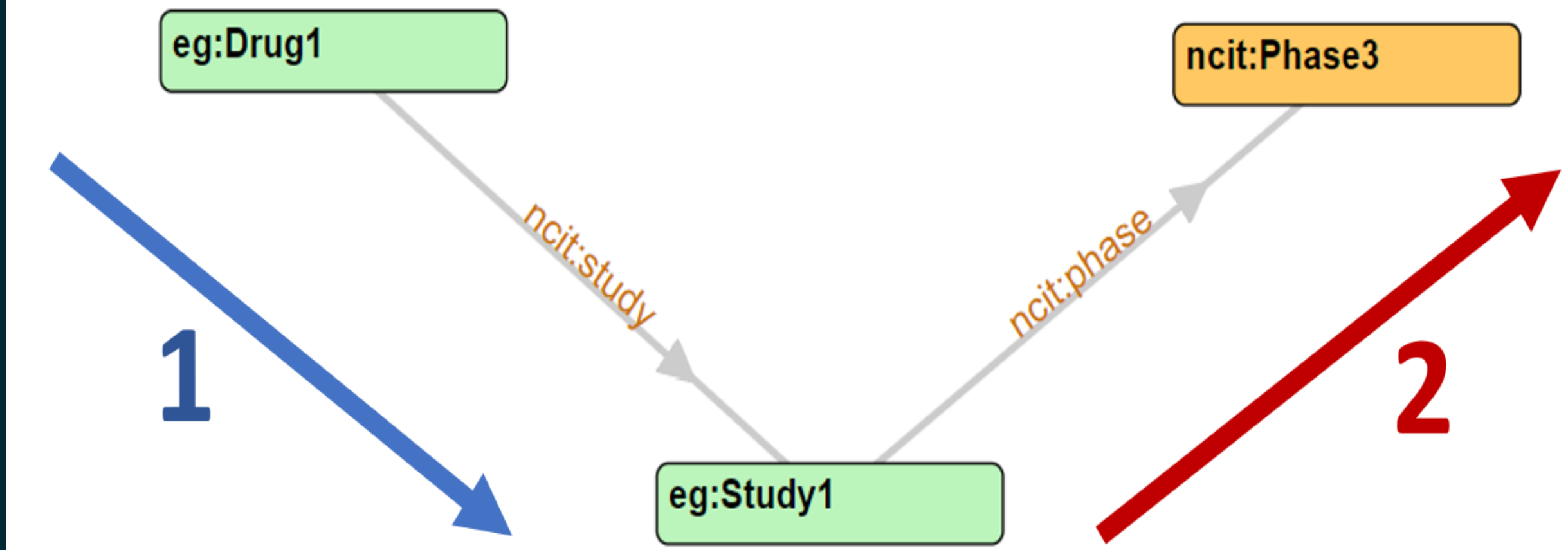
Identifier

Drug 1

Key=Value Pair

study=Study1

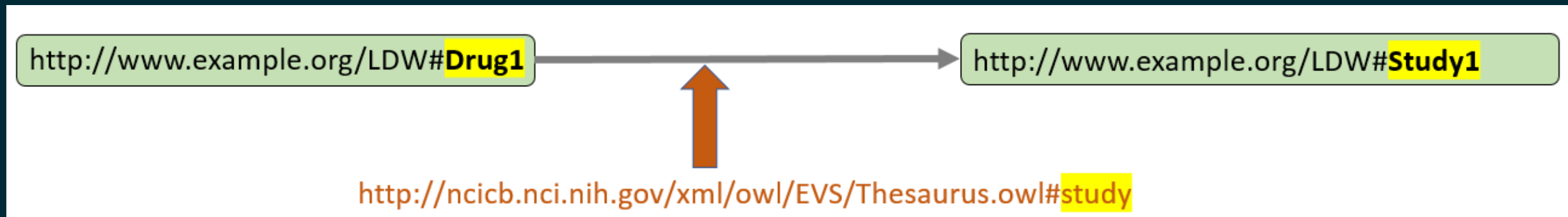
“ Drug1 *has study* Study 1”


eg:Drug1
ncit:Phase3
ncit:study
ncit:phase
1
2
eg:Study1

"THINGS" NEED UNIQUE IDENTIFIERS

IRI: INTERNATIONALIZED RESOURCE IDENTIFIER

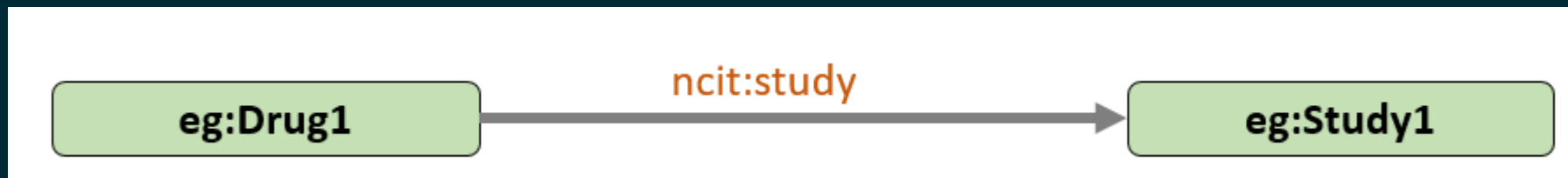
- Unique Identifier
- Uses HTTP://xx.xx.xx/***XXXX***



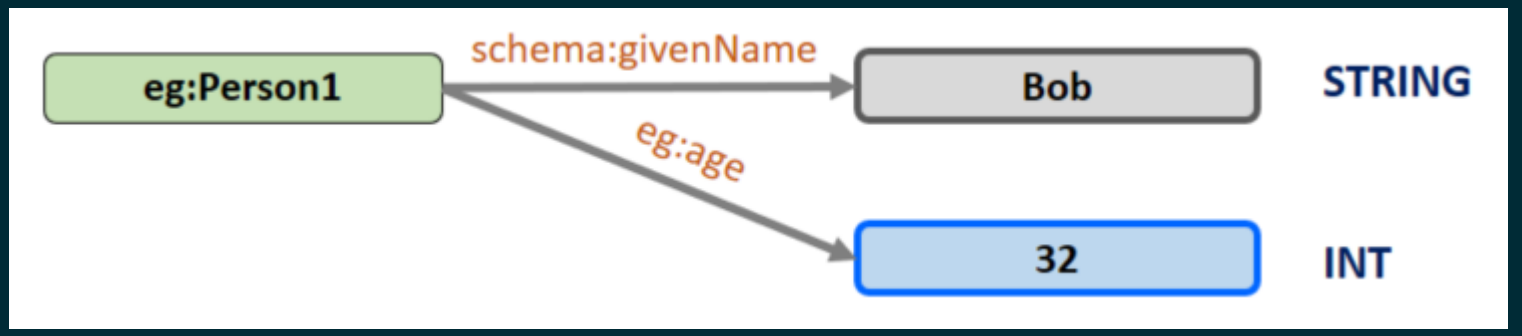
WORKSHOP PREFIXES

- Prefixes shorten IRIs for readability

```
@prefix eg: <http://example.org/LDWorkshop#> .
@prefix ncit: <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#> .
@prefix schema: <http://schema.org/> .
```



LITERALS

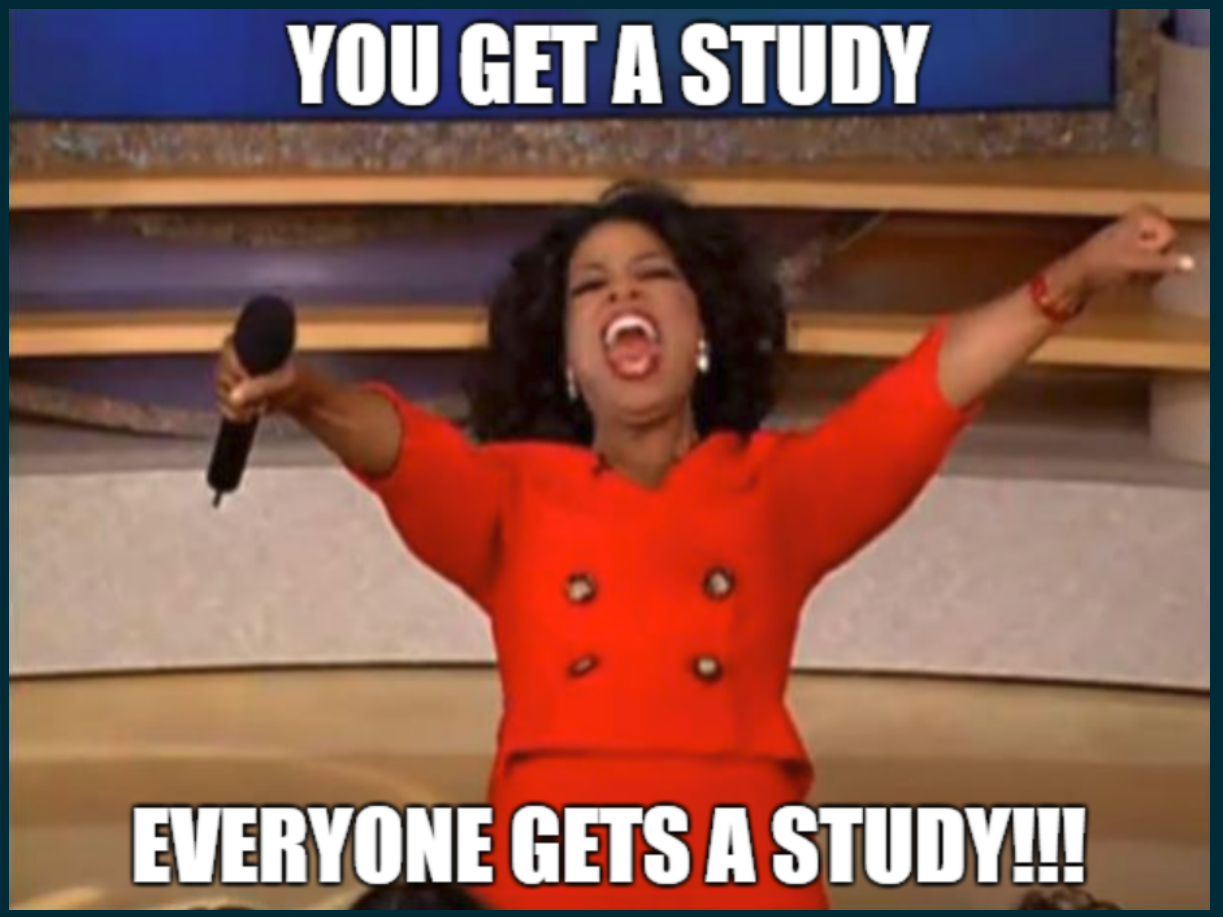


- ***string***
- number
  - ***integer (INT)***
- date

No links *from* a literal

OUTLINE

0. Introduction to Linked Data
1. **Create a Study Graph**
2. Query Graph Data
3. Ontology and Inference
4. Merge Studies
5. Linked Data in the Real World


YOU GET A STUDY
EVERYONE GETS A STUDY!!!

INTRODUCTION TO THE GRAPH EDITOR

Reference: .../doc/**Graph Editor Introduction.pdf**

CASE

NODES

eg:Study<n>

eg:TrtArm<n>-<x>

eg:Person<n><y>

eg:PlaceboArm

eg:LDExpert

ncit:Phase<z>

ncit:Male

ncit:Female

LINKS

eg:age

eg:LDExpert

eg:participatesIn

eg:randomizedTo

eg:trtArm

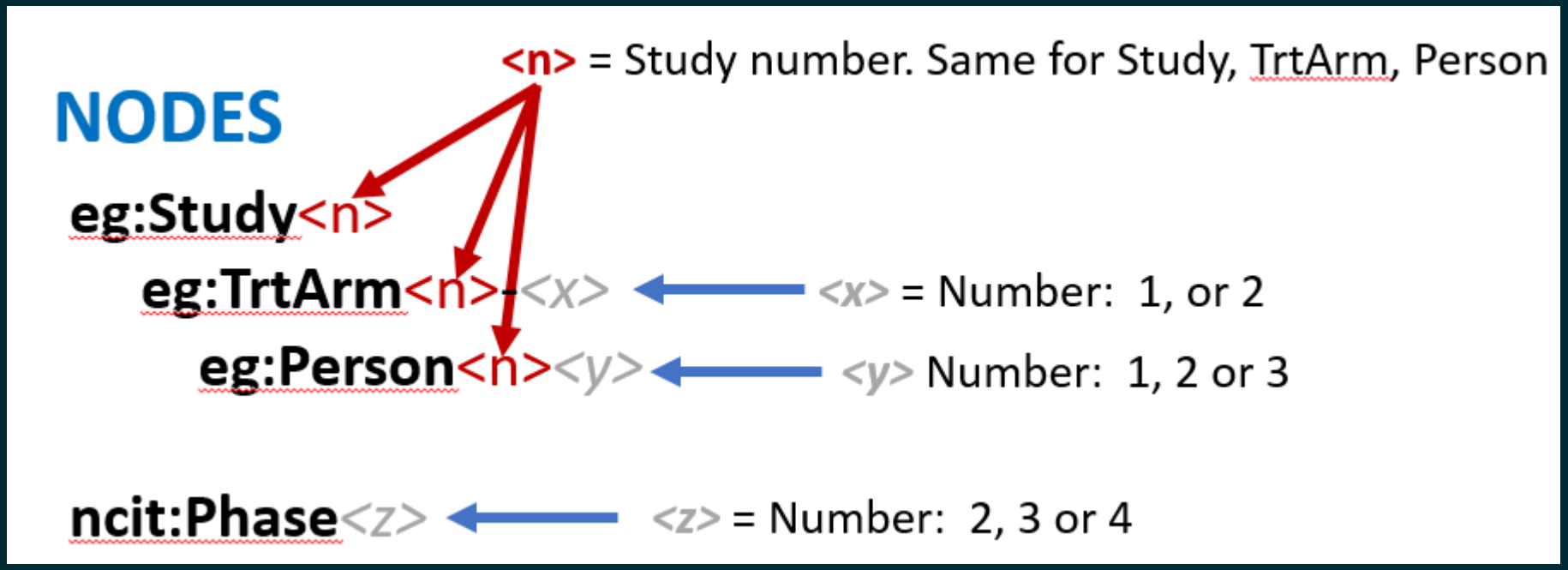
eg:trtArmType

ncit:gender

ncit:phase

ncit:study

NUMBERING


<n> = Study number. Same for Study, TrtArm, Person
NODES
eg:Study<n>
eg:TrtArm<n>-<x>
<x> = Number: 1, or 2
eg:Person<n><y>
<y> Number: 1, 2 or 3
ncit:Phase<z>
<z> = Number: 2, 3 or 4

EXERCISE

1. Create a Study Graph

OUTLINE

0. Introduction to Linked Data
1. Create a Study Graph
2. **Query Graph Data**
3. Ontology and Inference
4. Merge Studies
5. Linked Data in the Real World

EXERCISE

2. Query Graph Data

OUTLINE

0. Introduction to Linked Data
1. Create a Study Graph
2. Query Graph Data
3. **Ontology and Inference**
4. Merge Studies
5. Linked Data in the Real World

Ontologies, Reasoners & Inference

Ontology

A vocabulary of things and how they relate to each other

- ...just more nodes and links
- Tools: Protege, TopBraid

Reasoner

An *engine* that applies the ontology to the graph and ***infers*** values and relationships not in your original data.

THINK ABOUT THAT AGAIN:

Ontologies and Reasoning create **values** and **relations** ***not in your original data!***

StudyOntology.TTL

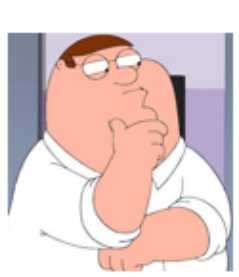
A SUBSET OF THE STUDY ONTOLOGY FILE

```
eg:drugProductName rdf:type owl:ObjectProperty ;
                   rdfs:domain eg:DrugProduct ;
                   rdfs:range eg:DrugProductName .

eg:participatesIn rdf:type owl:ObjectProperty ;
                  rdfs:domain eg:HumanStudySubject ;
                  rdfs:range ncit:Study .

eg:trtArm rdf:type owl:ObjectProperty ;
          rdfs:domain ncit:Study ;
          rdfs:range eg:TrtArm .

eg:trtArmType rdf:type owl:ObjectProperty ;
              rdfs:domain eg:TrtArm ;
              rdfs:range eg:TrtArmType .

#################################################################
#    Classes
#################################################################

###  Products
eg:Product rdf:type owl:Class .

eg:DrugProduct rdf:type owl:Class ;
               rdfs:subClassOf eg:Product .

### Product Name
eg:ProductName rdf:type owl:Class .

eg:DrugProductName rdf:type owl:Class ;
                   rdfs:subClassOf eg:ProductName .

### Study
ncit:Study rdf:type owl:Class .


### Treatment Arms, Types
eg:TrtArm rdf:type owl:Class .

eg:TrtArmType rdf:type owl:Class .

### Person and Types of Persons
schema:Person rdf:type owl:Class .

eg:HumanStudySubject rdf:type owl:Class ;
                     rdfs:subClassOf schema:Person .
```

1.

Reasoner sees relation:

```
eg:participatesIn rdf:type owl:ObjectProperty ;
                  rdfs:domain eg:HumanStudySubject ;
                  rdfs:range ncit:Study .
```

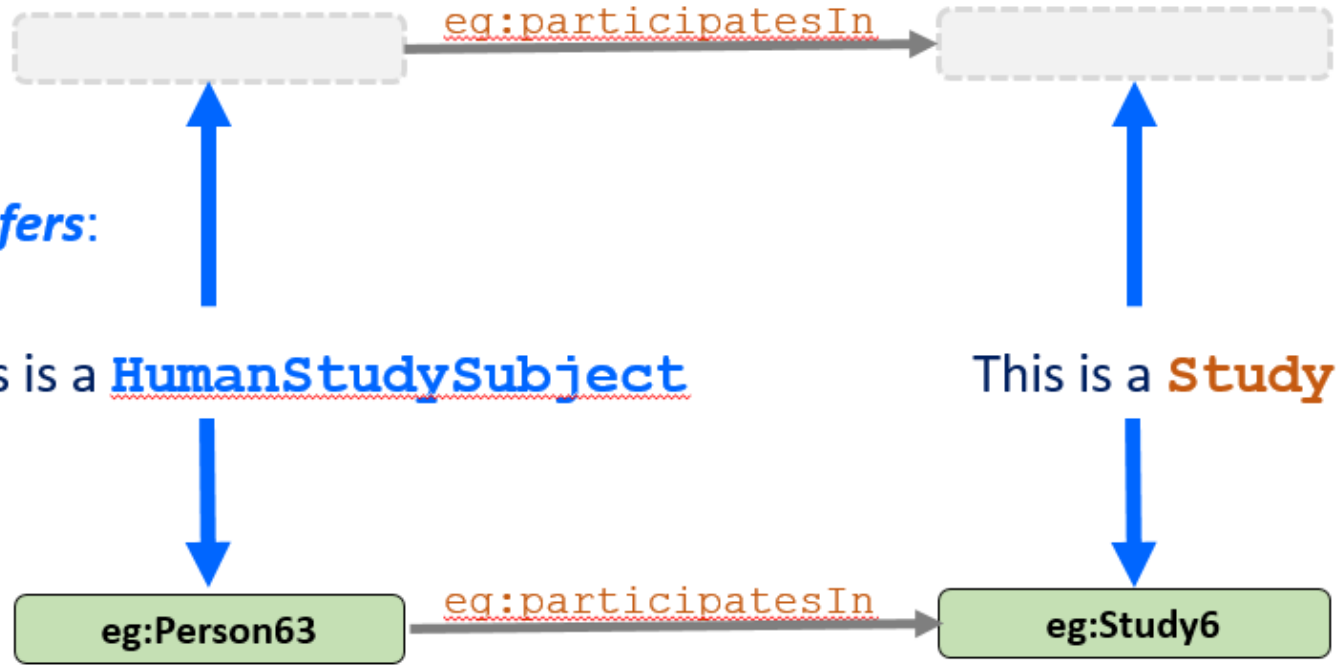


2.

Reasoner ***Infers***:

3.

Therefore:



1.

Reasoner sees Ontology
has classes and subclasses

```
eg:HumanStudySubject rdf:type owl:Class ;
                     rdfs:subClassOf schema:Person .
```

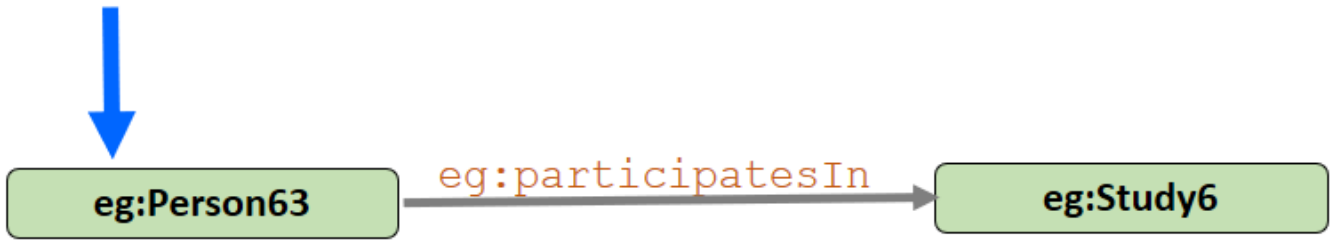
2.

- Thing
  - Person
    - HumanStudySubject

3.

Therefore:

This is a **HumanStudySubject**, **Person**, and **Thing**



EXERCISE

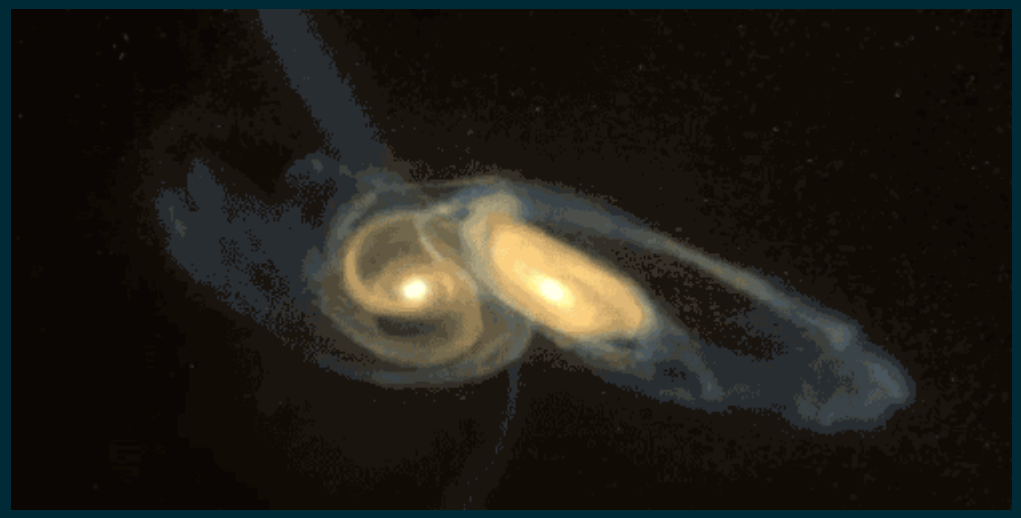
3. Ontology and Inference

OUTLINE

0. Introduction to Linked Data
1. Create a Study Graph
2. Query Graph Data
3. Ontology and Inference
4. **Merge Studies**
5. Linked Data in the Real World

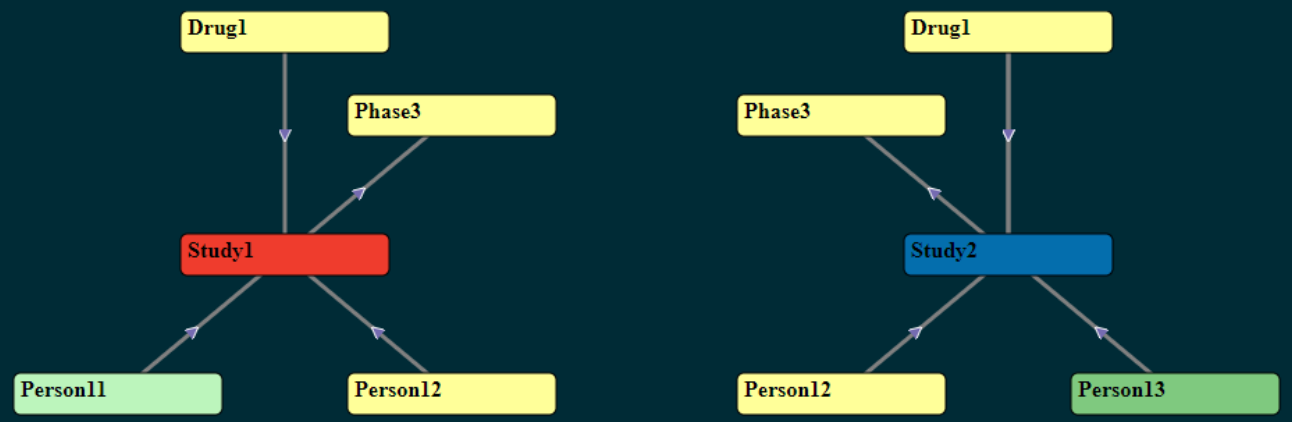
When IRIs are the **same**, merging is ***automagic!***

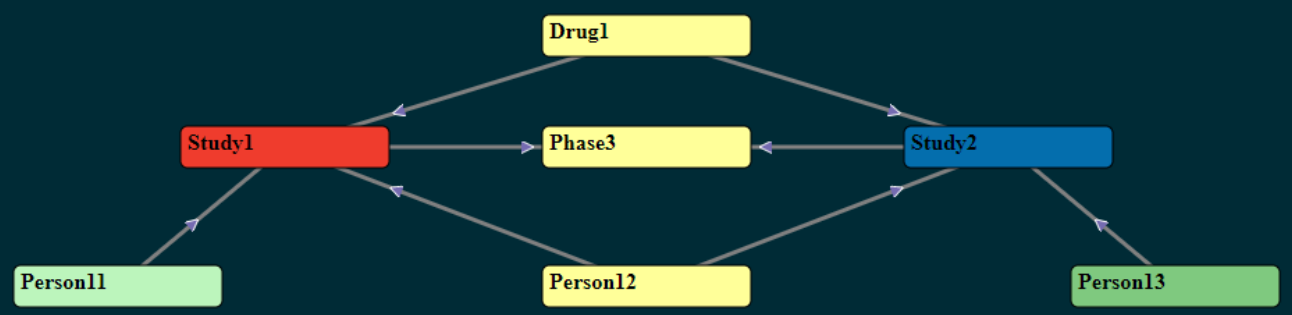
WITH RDF, MERGING BE LIKE:

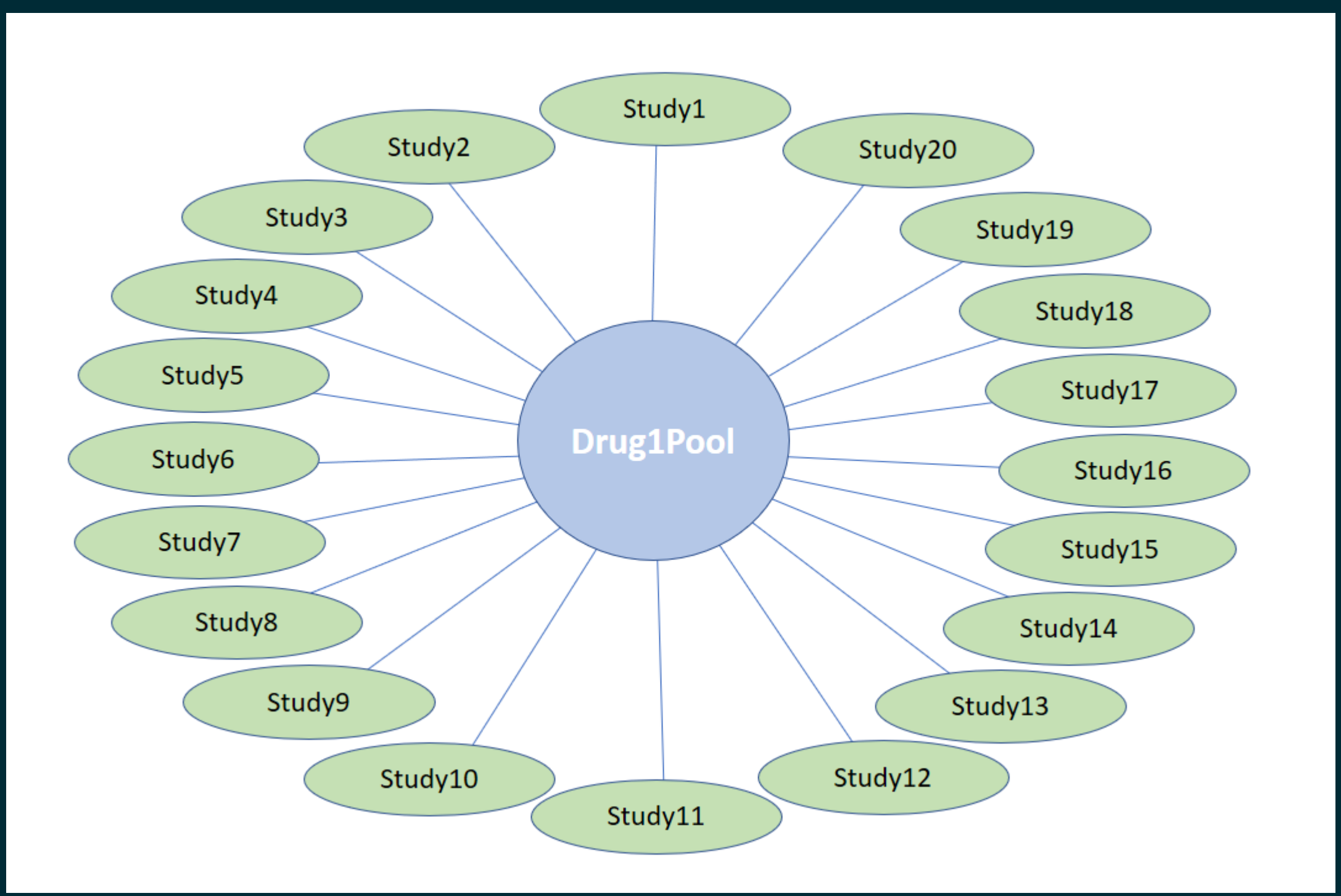


What? How?

GRAPH MERGE


Drug1
Phase3
Study1
Person11
Person12
Drug1
Phase3
Study2
Person12
Person13


Drug1
Study1
Phase3
Study2
Person11
Person12
Person13


Study1
Study2
Study3
Study4
Study5
Study6
Study7
Study8
Study9
Study10
Study11
Study12
Study13
Study14
Study15
Study16
Study17
Study18
Study19
Study20
Drug1Pool

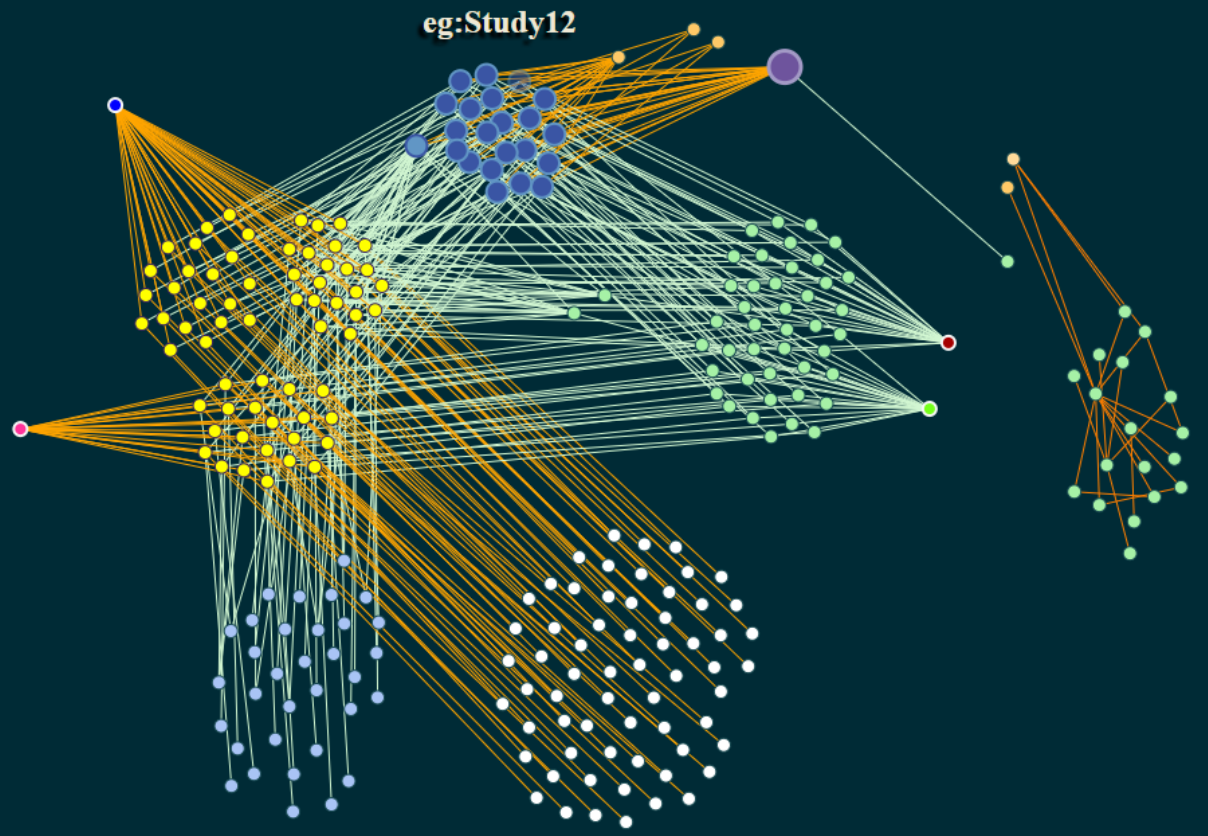
EXERCISE

4. Merge Studies

DRUG 1 POOL

Network Graph

DRUG 1 DATA POOL



OUTLINE

0. Introduction to Linked Data
1. Create a Study Graph
2. Query Graph Data
3. Ontology and Inference
4. Merge Studies
5. **Linked Data in the Real World**

BUT FIRST: ACKNOWLEDGEMENTS

- PhUSE - server costs
- **Stephen Bamford**
- **Chris Decker** - server cloning
- **Lauren, Wendy, Jane, Tora** and the entire PhUSE admin team
- **Stardog Union** - Triplestore, configuration, support.
- ...everyone else I forgot to mention
- And: **YOU**!

BUT SECOND: RESOURCES

- Workshop materials, including the Graph Editor, SPARQL scripts, PDF of this presentation: https://github.com/phuse-org/LinkedDataWorkshop/CSS2018

RESOURCES

Introduction to Semantic Web	https://www-stage.cambridgesemantics.com/semantic-university/introduction-semantic-web
What is Linked Data?	https://www-stage.cambridgesemantics.com/semantic-university/what-linked-data
Introduction to Linked data	https://www-stage.cambridgesemantics.com/semantic-university/introduction-linked-data

LINKED DATA IN THE REAL WORLD

StarDog Union

EXTRA SLIDES

The LOD cloud

SDTM CAN BE IMPROVED

vs

Filter

	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			

dm

Filter

	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

- Non-extensible, two dimensional
- Data repetition
- Terminology, codes not linked
- *Version Conversion*: time, $£€
- Does not model study entities
- Lacks integral metadata

  ... and more.
- **26% of CDER SDTM applications: at least 1 error**

OPPORTUNITES

- Merge data from diverse sources (no silos!)
- Data integration across the life cycle
- Integral Metadata
- High quality submissions data
- New ways to explore and analyze
- Build a foundation for AI and ML
- [!! YOUR IMAGINATION !!]