LET'S MAKE A KNOWLEDGE GRAPH!
A HANDS-ON, INTERACTIVE, LINKED DATA WORKSHOP

PHUSE EUCONNECT 2019

Amsterdam
2019-11-12

INSTRUCTORS

Tim Williams
Statistical Solutions Lead
UCB
tim.williams@PhUSE.eu

Johannes Ulander
Principal Consultant, S-Cubed
CDISC Subject Matter Expert
Authorised CDISC SDTM Instructor
ju@s-cubed.dk

PREPARATION

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

Workshop Files, Presentation PDF:

https://github.com/phuse-org/LinkedDataWorkshop/EUConnect19

(for later)

OUTLINE

0. What is a Knowledge Graph?
1. Create Your Study Graph
2. Query Your Graph
3. Extend to Other Graphs (Federated Query)
4. Ontology and Inference
5. Merge Studies
6. Discussion

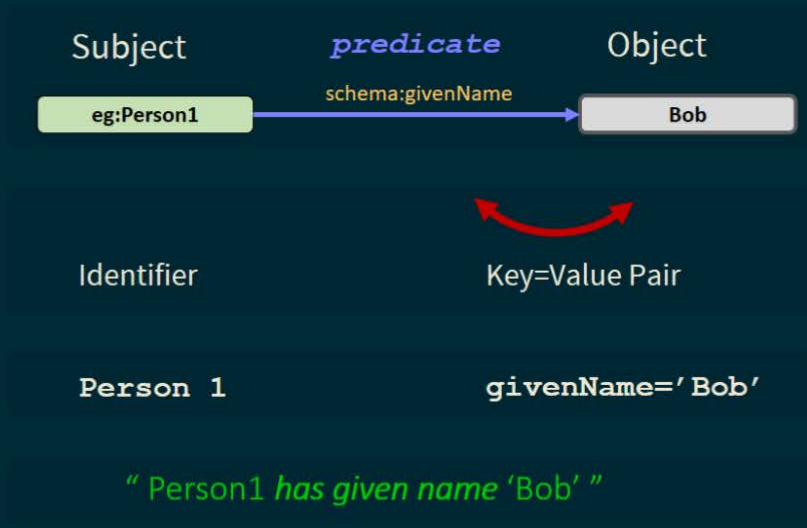
WHAT IS A KNOWLEDGE GRAPH?

- An interconnected network of information consisting of meaningful relationships that are understandable by both people and computers.
- Built on Linked Data

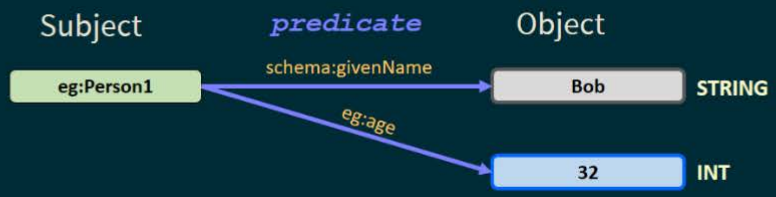
WHAT IS LINKED DATA?

- Data that has meaningful (semantic) relationships
- Resource Description Framework (RDF)

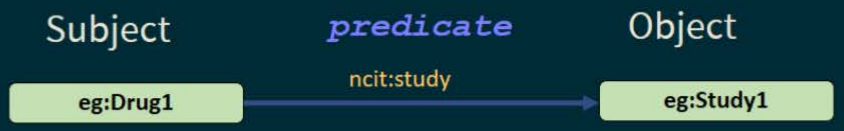
RDF TRIPLE DESCRIBING PERSON 1



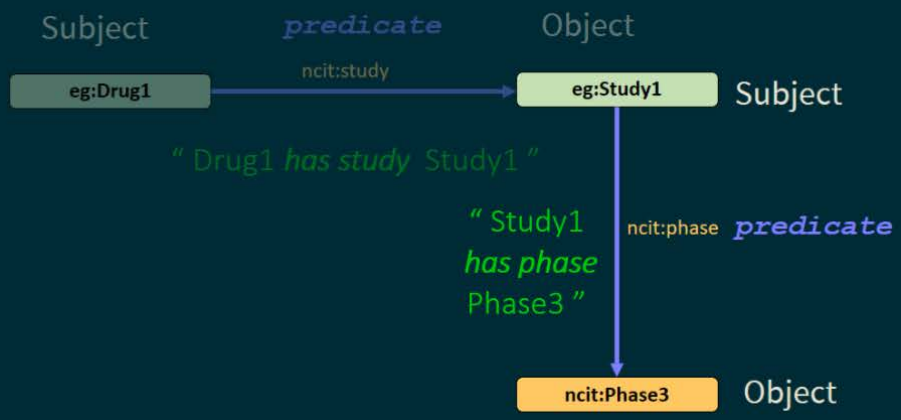
PERSON 1 NAME AND AGE



" Person1 *has given name* 'Bob', age 32 "



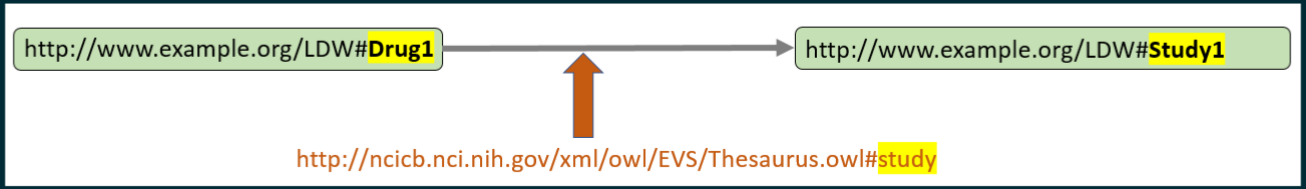
" Drug1 *has study* Study1 "


Subject
predicate
Object
ncit:study
eg:Drug1
eg:Study1
Subject
" Drug1 has study Study1 "
" Study1
has phase
Phase3 "
ncit:phase
predicate
ncit:Phase3
Object

"THINGS" NEED UNIQUE IDENTIFIERS

IRI: INTERNATIONALIZED RESOURCE IDENTIFIER

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

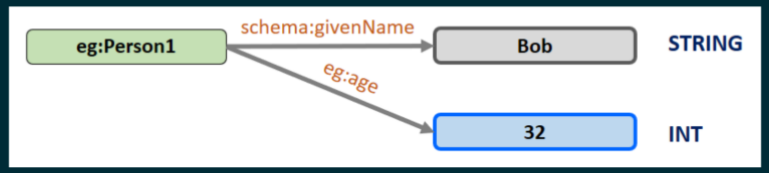


WORKSHOP PREFIXES

- 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

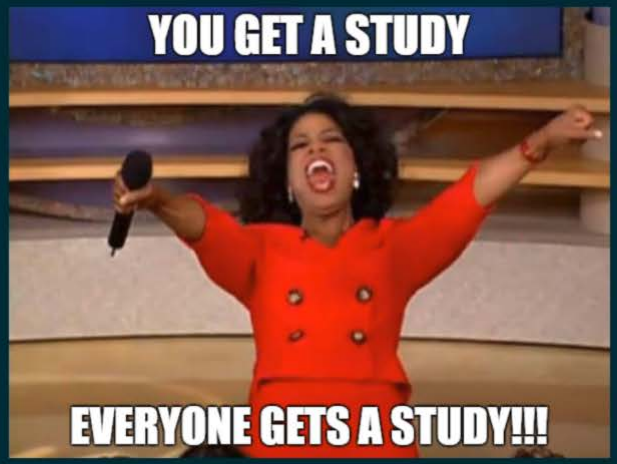


- ***string***
- number
  - ***integer (INT)***
- date

No links *from* a literal

OUTLINE

0. What is a Knowledge Graph?
1. **Create Your Study Graph**
2. **Query Your Graph**
3. Extend to Other Graphs (Federated Query)
4. Ontology and Inference
5. Merge Studies
6. Discussion


YOU GET A STUDY
EVERYONE GETS A STUDY!!!

INTRODUCTION TO THE GRAPH EDITOR

See your handout

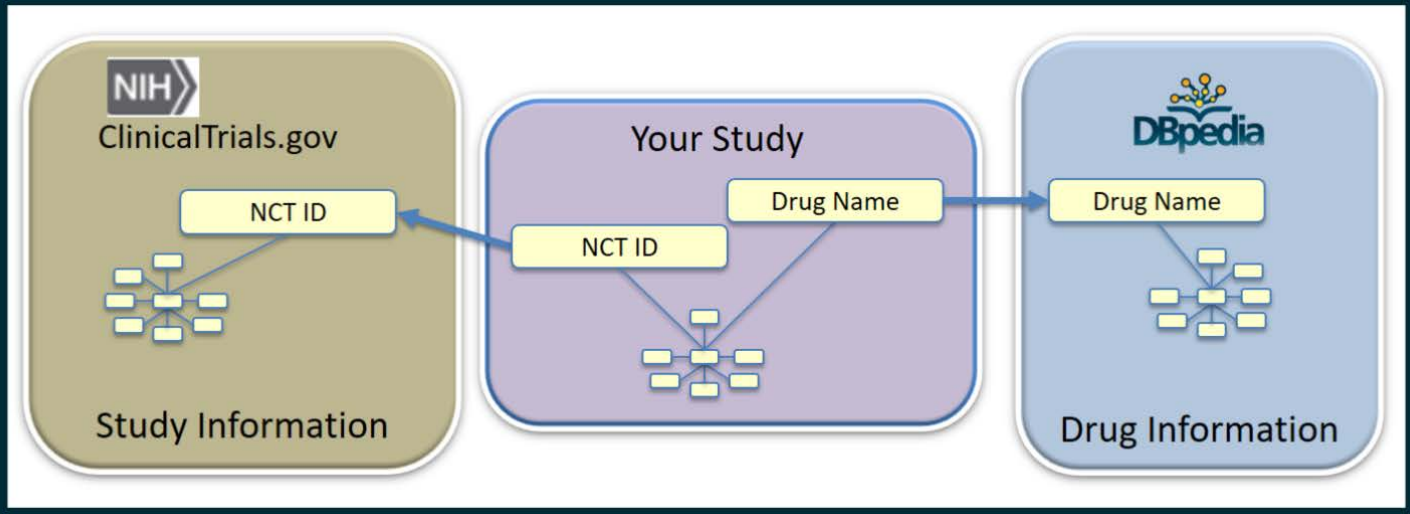
Reference: .../doc/**Graph Editor Introduction.pdf**

EXERCISE

1. Create Your Study Graph
2. Query Your Graph

OUTLINE

0. What is a Knowledge Graph?
1. Create Your Study Graph
2. Query Your Graph
3. **Extend to Other Graphs (Federated Query)**
4. Ontology and Inference
5. Merge Studies
6. Discussion


NIH
ClinicalTrials.gov
NCT ID
Study Information
Your Study
Drug Name
NCT ID
DBpedia
Drug Name
Drug Information

EXERCISE

3. Extend to Other Graphs (Federated Query)

OUTLINE

0. What is a Knowledge Graph?
1. Create Your Study Graph
2. Query Your Graph
3. Extend to Other Graphs (Federated Query)
4. **Ontology and Inference**
5. Merge Studies
6. Discussion

Ontology and 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

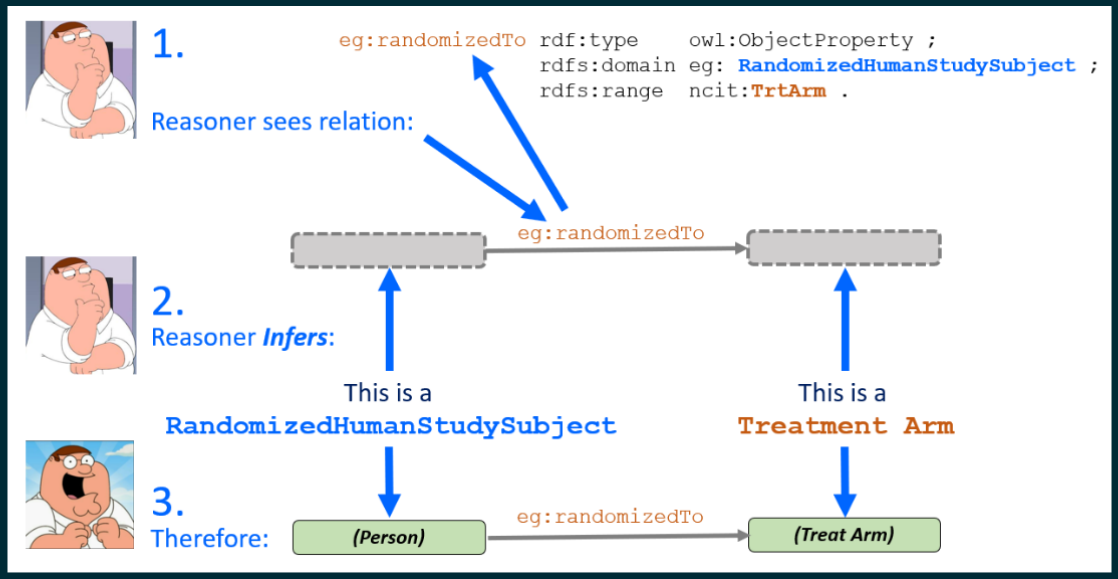
A SUBSET OF THE STUDY ONTOLOGY FILE

```
eg:randomizedTo rdf:type owl:ObjectProperty ;
                rdfs:domain eg:RandomizedHumanStudySubject ;
                rdfs:range eg:TrtArm .
###################################################################
#    Classes
###################################################################

schema:Person rdf:type owl:Class .

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

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


1.
eg:randomizedTo rdf:type owl:ObjectProperty ;
rdfs:domain eg: RandomizedHumanStudySubject ;
rdfs:range ncit:TrtArm .
Reasoner sees relation:
eg:randomizedTo
2.
Reasoner Infers:
This is a
RandomizedHumanStudySubject
This is a
Treatment Arm
3.
Therefore:
(Person)
eg:randomizedTo
(Treat Arm)

1.

Reasoner classes and subclasses

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

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

2.

- Thing
  - Person
    - HumanStudySubject
      - RandomizedHumanStudySubject

3.

Therefore:

(Person) —eg:randomizedTo→ (Treat Arm)

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

EXERCISE

4. Ontology and Inference

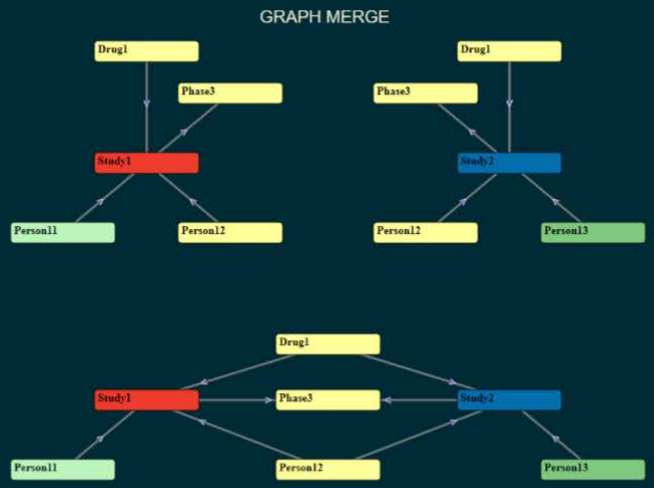
OUTLINE

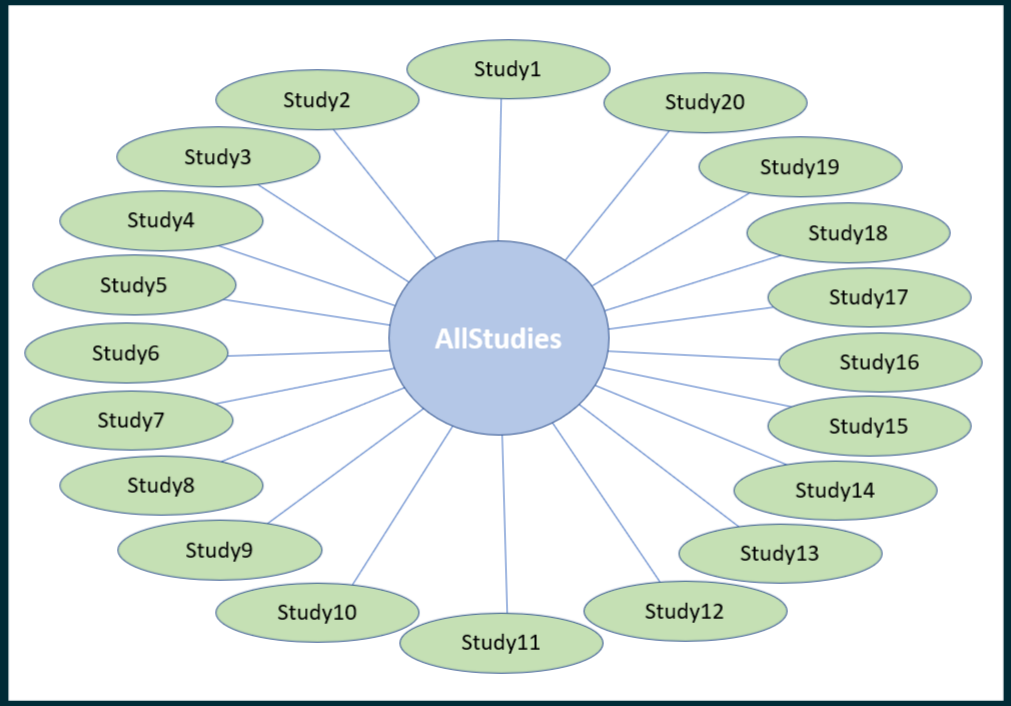
0. What is a Knowledge Graph?
1. Create Your Study Graph
2. Query Your Graph
3. Extend to Other Graphs (Federated Query)
4. Ontology and Inference
5. **Merge Studies**
6. Discussion

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

WITH RDF, MERGING BE LIKE:

What? How?




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

EXERCISE

5. Merge Studies

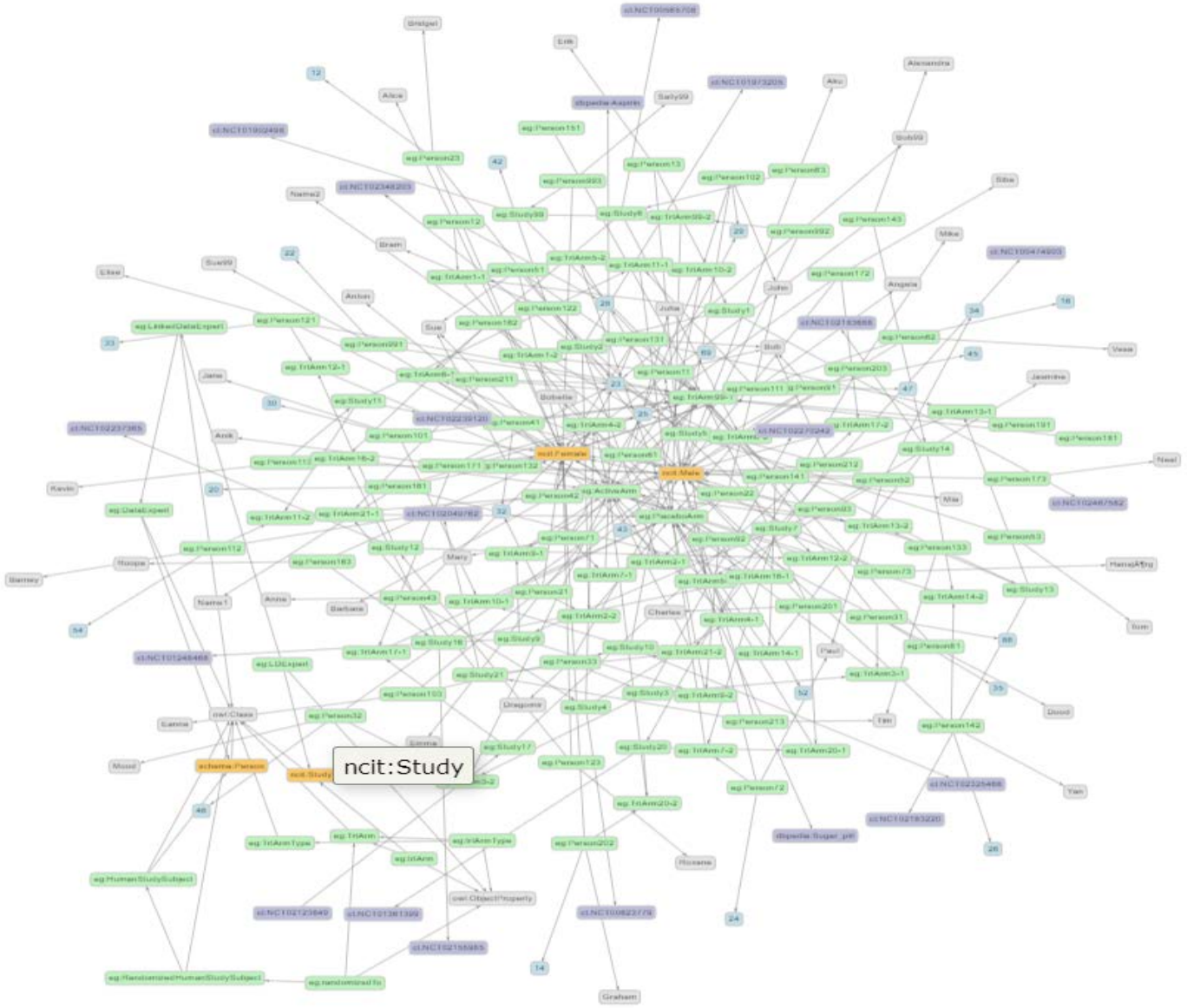
ALLSTUDIES DATA POOL

BONUS!

Visualize your Data Pool.

`AllStudiesPoolVis.R`

Data Pool – All Workshop Studies



OUTLINE

0. What is a Knowledge Graph?
1. Create Your Study Graph
2. Query Your Graph
3. Extend to Other Graphs (Federated Query)
4. Ontology and Inference
5. Merge Studies

6. Discussion...after final words

ACKNOWLEDGEMENTS

- YOU!
- PhUSE
  - Lauren - Prep Webinars and coordination
  - PhUSE Admin Team
- Stardog Union
  - Servers, Stardog Triplestore, Stardog Studio
  - John Bresnahan (Stardog) - server cloning

RESOURCES

Stardog Union

- fetch.stardog.com/phuse/
- www.stardog.com

RESOURCES

- Workshop materials, including the Graph Editor, SPARQL scripts, PDF of this presentation:
  https://github.com/phuse-org/LinkedDataWorkshop/EUConnect19

Join one of our PhUSE Linked Data Projects:

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

RESOURCES

Linked Data Introduction
https://www.youtube.com/watch?v=4x_xzT5eF5Q
SPARQL in 11 Minutes
https://www.youtube.com/watch?v=FvGndkpa4K0

DISCUSSION