TODO:

PhUSE CSS

Linked Data for Clinical Trials

An Interactive Hands-on Workshop

EXERCISES

Version 3.0

04 March 2018

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

Contents

[Introduction 2](#_Toc504054497)

[Materials 2](#_Toc504054498)

[Symbols and Conventions 2](#_Toc504054499)

[Server Login 3](#_Toc504054500)

[Exercises 4](#_Toc504054501)

[1. Create a Graph in the Graph Editor 4](#_Toc504054502)

[1.1 Open the Graph Editor 4](#_Toc504054503)

[1.2 Add a Clinical Trial and Treatment Arms 4](#_Toc504054504)

[1.1 Add a PERSONs to the Trial and a Treatment 6](#_Toc504054505)

[1.2 Add another person to the Trial 6](#_Toc504054506)

[1.3 Add additional concepts and entities 6](#_Toc504054507)

[1.4 Export to TTL 6](#_Toc504054508)

[2. Upload to Triplestore 6](#_Toc504054509)

[3. Explore the Data 7](#_Toc504054510)

[3.1 Visualize your graph 7](#_Toc504054511)

[3.2 Query your graph 7](#_Toc504054512)

[3.2.1 Who enrolled in your trial? 7](#_Toc504054513)

[3.2.2 How many patients received which treatments? 7](#_Toc504054514)

[3.3 Linking to Other Graphs 7](#_Toc504054515)

[3.3.1 CDISC Terminology 7](#_Toc504054516)

[3.3.2 Leveraging an Ontology 7](#_Toc504054517)

[3.4 Federated Query\* (\*assumes can be configured for workshop) 7](#_Toc504054518)

[3.4.1 What trials were conducted for Drug1? 7](#_Toc504054519)

[3.4.2 How many patients participated in Drug1 trials? 7](#_Toc504054520)

[3.4.3 What is the gender breakdown for all Drug1 trials? 7](#_Toc504054521)

[3.4.4 Identify patients that participated in multiple trials. 7](#_Toc504054522)

[3.4.5 etc. 7](#_Toc504054523)

[Appendix 1: Course Resources 9](#_Toc504054524)

# Introduction

Instructions in this document are specific to the PhUSE cloud server environment, Graph Editor application and workshop methods. The exercises represent one of many possible approaches to the material and make no claim to be best or recommended method. The instructor will complete the exercises during the session, along with the attendees.

Your feedback is welcomed and encouraged. Please send your comments to: **tim.williams@PhUSE.eu**

## Materials

* Laptop with Remote Desktop capability
* Printed copies of the following will be supplied by the instructor:
  + Exercises (this document)
  + Info Sheet
  + Graph Editor Reference

## Symbols and Conventions

Symbols provide important information:

|  |  |
| --- | --- |
|  | Stop and wait. Do not proceed to the next section until instructed. |
| **!** | Cautions and warnings. Failure to follow these steps may lead to unanticipated results and problems. |
| IdeaIcon_clean_20mm | Helpful tips and advice. |
| Image result for information icon | Additional information. |

# Server Login

Login to the cloud server using Remote Desktop. Instructions assume Windows 10 operating system.

1. In the search box on the taskbar on your laptop, type **remote desk,** then select **Remote Desktop Connection** from the Programs section.﻿
2. Click **Show Options** if needed to show the fields for **Computer:** and **User name:**

The instructor will provide the password and an IP Address for the Computer: field.

**Computer:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_

(See Info Sheet)

**User name:** phuseldw

**Password:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Provided by Instructor)

1. Entering values in the **Computer:** and **User name:** fields, then click **Connect**.
2. When prompted, enter the password supplied by the instructor and click **OK**.

|  |  |  |
| --- | --- | --- |
|  | Stop here and wait for the instructor.  Presentation follows |  |

# Exercises

# Create a Graph in the Graph Editor

# Open the Graph Editor

On the cloud server:

1. Open Google Chrome.
2. Click on "GraphEditor" on the shortcuts bar. 
3. The Graph Editor opens and shows two IRI nodes: One for **Drug1** and one for the drug's name, **Serum114.**



Ask for assistance if you do not see these nodes.

# Add a Clinical Trial, Trial Phase, and Treatment Arms

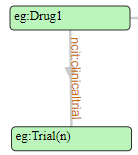
|  |  |
| --- | --- |
| **!** | When you see values like Trial**(n)**, Phase**(n)**, Treatment**(n)**, etc. you must change your graph to values you find in the Info Sheet or create on your own. The values used in your graph become : Trial**1**, Phase**3**, Treatment**7-1**, etc. |

Create a node that represents unique clinical **Trial** conducted for **Drug1** and the Treatments for that trial. Reference the Info Sheet for values where needed.

1. Create a **Trial** node with these values:

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | *Info Sheet: x.x: [Trial1]* | Image result for information icon*See Info Sheet for value* |
| **Type:** | IRI |  |



1. Create the relation ***from*** **Drug1** ***to*** the **Trial** node:

Link values:

|  |  |
| --- | --- |
| **Prefix:** | ncit |
| **Label:** | clinicaltrial |

1. Create a Trial **Phase** node. You may choose any of these phases: **Phase2**, **Phase3**, **Phase4.**

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | ncit |  |
| **Label:** | *Choose a value:* | Image result for information icon*You choose a value for phase:* **Phase2, Phase3, Phase4**  *NO SPACE in the name!* |
| **Type:** | IRI |  |

1. Create the link ***from*** the **Trial** ***to*** the Trial **Phase**.

Link values:

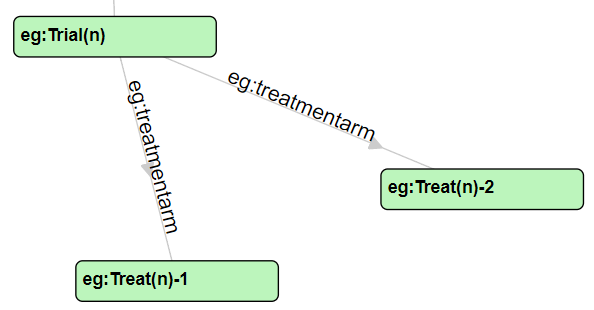
|  |  |
| --- | --- |
| **Prefix:** | ncit |
| **Label:** | phase |

1. Create two **Treatment** nodes for the **Trial**.

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | *Info Sheet: x.x, x.x [Treat1-1,Treat1-2]* | Image result for information icon*See Info Sheet for values.* |
| **Type:** | IRI |  |

Ensure you have two treatment nodes in your graph.



1. Create the relations ***from*** the **Trial** ***to*** the two **Treat**ment nodes.

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | treatmentarm | Image result for information icon*Label is all lowercase.* |

Link values:

There should now be one treatmentarm link to each of the **treatment** nodes.

1. Create two "treatment type" nodes to represent **Placebo** and **Active** treatments

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | Placebo |  |
| **Type:** | IRI |  |

Node values:

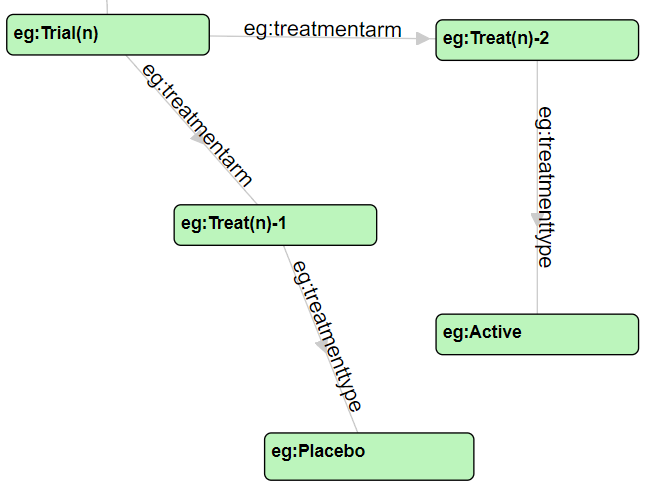
|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | Active |  |
| **Type:** | IRI |  |

1. Now join each **Treat**ment node (created in step x) to one of the Treatment Type nodes using the following relationship. It does not matter which is the A**ctive** or P**lacebo**. You will create two edges:

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | treatmenttype | Image result for information icon*Label is all lowercase.* |

The **Trial** and **Treat**ment portion of your graph should look similar to:



|  |  |
| --- | --- |
| IdeaIcon_clean_20mm | Click  to create a copy in case anything goes wrong in the next steps. |

# Add Persons and Treatments

1. Add three **Person** nodes to the graph:

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | *Info Sheet: x.x, x.x , x.x [Person11, Person12, Person13]* | Image result for information icon*See Info Sheet for values.* |
| **Type:** | IRI |  |

1. For two of the **Person** Nodes: Create ***enrolledin*** relations ***from*** the **Person** ***to*** the **Trial** node.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | enrolledin | Image result for information icon*Label is all lowercase.* |

1. For the remaining **Person** node: Create a ***primaryinvestigator*** link ***from*** the **Person** node ***to*** the **Trial** node.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | primaryinvestigator | Image result for information icon*Label is all lowercase.* |

1. Assign one of the two **Person** nodes ***enrolledin*** the Trial to one of the **Treat**ments by creating the ***treatment*** link ***from*** that **Person** ***to*** the **Treat**ment.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | treatment | Image result for information icon*Label is all lowercase.* |

1. Repeat for the second person node ***enrolledin*** the Trial by creating a link ***from*** the **Person** node ***to*** the other **Treat**ment.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | treatment | Image result for information icon*Label is all lowercase.* |

Now both **Person**s ***enrolledin*** the **Trial** are assigned to **Treat**ments. The **Person** with ***primaryinvestigator*** relationship is not assigned to a **Treat**ment.

# Add Gender to Persons

1. Create a node for **Male** and an node for **Female .**

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | Male |  |
| **Type:** | IRI |  |

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | Female |  |
| **Type:** | IRI |  |

2. Create links ***from***all 3 **Person** nodes ***to*** a Gender value using the ***gender*** relation.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg |  |
| **Label:** | gender | Image result for information icon*Label is all lowercase.* |

|  |  |
| --- | --- |
| IdeaIcon_clean_20mm | Click  to save a copy of your graph. |

# Demographics for Person Nodes

A. Create a **Name** for each **Person** in the graph:

1. Create **Name** nodes as strings:

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** |  | Image result for information iconPrefix is not applicable to literals. You may ignore this field. |
| **Label:** | *<Any name you want>* |  |
| **Type:** | STRING |  |

2. Link ***from*** a **Person** ***to*** a **Name** use the ***name*** relation.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | schema | Image result for information icon Many other prefixes could be used here. |
| **Label:** | name |  |

B. Add an **Age** (in years) for each **Person.**

1. Create **Age** nodes as INT (integers):

Node values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** |  | Image result for information iconPrefix is not applicable to literals. |
| **Label:** | *<Any age integer value you want>* |  |
| **Type:** | INT |  |

2. Link ***from***a **Person** ***to*** an age using the ***age*** relation.

Link values:

|  |  |  |
| --- | --- | --- |
| **Prefix:** | eg | Image result for information icon eg is used because age calculation may be specific to this graph. |
| **Label:** | age |  |

|  |  |
| --- | --- |
| IdeaIcon_clean_20mm | Click  to save a copy of your graph. |

**Optional Activity**

Time permitting, you may create these additional nodes while you wait for the class to catch up. Do not add too many nodes or your whiteboard may become difficult to read.

C. Add additional Person nodes and assign them to a Treatment.

D. Add additional treatments, or add details to the existing treatments.

|  |  |  |
| --- | --- | --- |
|  | Stop here and wait for the instructor. |  |

# Export to TTL

1. **As for the instructor orassistant to review your graph prior to taking the next steps. Click**  **if you make any changes.**
2. Click on  to export the graph to TTL format.
3. The graph will be saved to your downloads folder as **WhiteBoardTriples.TTL :**

**C:\Users\phuseldw\downloads\WhiteBoardTriples.TTL**

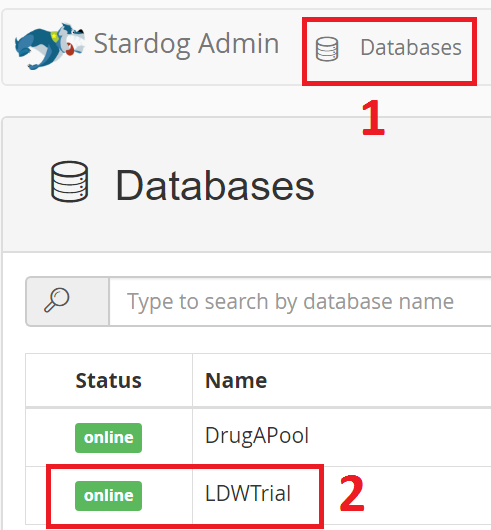
|  |  |
| --- | --- |
| **!** | Leave the graph open in Chrome. You will refer to it in the next section when creating queries. Ask for assistance if you accidentally close or refresh the Graph Editor window. |

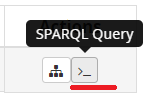
1. Use Windows explorer to navigate to the TTL file.
2. Double click on the .TTL file to open it.
3. Follow along with the instructor's explanation of the TTL file content.
4. Find the various ***nodes*** and ***links*** you created in your graph.

|  |  |  |
| --- | --- | --- |
|  | Stop here and wait for the instructor. |  |

# Upload to Triplestore

1. Open a new window in Google Chrome, keeping your graph editor window open.
2. Click the Stardog shortcut on the bookmarks bar.
3. Click "Databases" on the top menu
4. Find the database named **LDWTrial .** It is currently empty.



1. On the far right of the row named Trials, click on the query icon 



1. Select **Data | Add** from the Stardog menu:
2. Navigate to the TTL file you created from the GraphEditor and load it into Stardog.

C:\Users\phuseldw\Downloads\**WhiteBoardTriples.TTL**

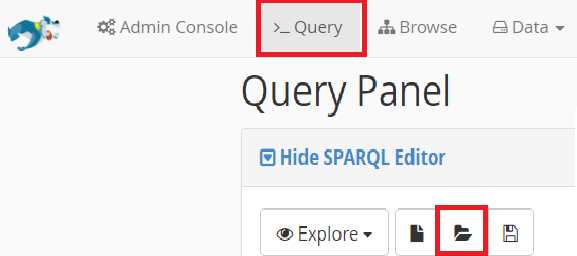
|  |  |
| --- | --- |
| **!** | Multiple copies of the TTL file will be present if you exported to TTL more than one time.: WhiteBoardTriples (1).TTL, WhiteBoardTriples (2).TTL, etc. Choose the most recent file with the largest (n) number in the name. |

# Query your graph

# Queries without Reasoning

# Show all triples

Query the graph to find all Subject-Predicate-Object relations in your graph.

1. In Stardog, click on the Query item in the top menu bar, then the Open Folder icon in the Query panel.
2. Navigate to the file C:\LinkedDataWorkshop\**scripts\SPARQL\100-AllTriples .rq** . Double-click the file to load it into the query window.
3. With the query loaded in to the window, click the  button. Scroll down to review the results.
4. Compare this view to what you expected after viewing the TTL file.

# How many people *enrolledin* the trial?

1. Review the graph you created. You want a count of the nodes that have an *enrolledin* relationship.
2. Selecting the number (COUNT) of those relationships is easy when you know SPARQL syntax:

SELECT (COUNT (?person) AS ?numP)

Solution: 101-EnrolledIn.rq

# How many people have each type of treatment (Active, Placebo)?

Adapt your query from x.x.x (above) to count the number of Person's that have each treatment.

Hint: <how to do the BY for each>

ANSWER: xxx-NumEachTreatType.rq

|  |  |  |
| --- | --- | --- |
|  | Stop here and wait for the instructor.  Presentation on Ontologies and Reasoning |  |

# Ontology

Steps to Load and view the ontology for this graph using Protege.

***<<ROUGH DRAFT FOLLOWS>***

# Review the Ontology

1. Double-click on the file ...<PATH>/data/TrialOnotology.TTL to open into a text editor.
2. Follow along with the instructor's explanation of the file.

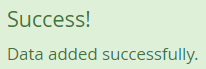
# Add the Ontology to the LDWTrial database

Select Data | Add



Navigate to this OWL file in this location and load it into the database:

C:\LinkedDataWorkshop\**scripts\LDWTrial.TTL**

You should see the message: 

Turn on the reasoner.

# Queries with Reasoning.

What Types of things (classes) are in our data?

Explore the data without reasoning:

Ensure that reasoning is turned off:

Select Explore | Class Hierarchy

A SPARQL query appears in the window. Execute the query and view the results. Not much is returned, only the fact that Classes and Properties appear in the data.



Turn on the reasoner.

Execute the query again. Note HumanStudySubject and Investigator have the ***supertype*** **Person**. Reasoning allows you to use these entities to query your data, even though your data did not contain these terms.

x.x Who are the HumanStudySubjects in your trial?

Instead of querying on the enrolledin predicate, you can now query using the inferred entity HumanStudySubject.

Load and execute the query xxx-HumanStudySubject.rq.

x.x Investigators?

Change the HumanStudySubject query to instead query for investigators.

Answer: xxxx-Investigators.rq

X.X All people associated with the trial?

Write a query to find all people associated with the trial.

Hint: HumanStudySubject and Investigators are subclasses of what superclass?

|  |  |  |
| --- | --- | --- |
|  | Stop here and wait for the instructor.  Presentation on Merging Data |  |

<PRESENTATION HERE: Merging Graphs, and Federated Queries>

# Merging data from all trials

Instructor creates a single database of all merged trials, live demo.

Then each attendee:

Obtain the data pool using the service query.

Select the Drug1Pool database.

Execute the query to obtain the pool from the instructor.

# How many triples are in the pool?

xxx-PoolTripleCount.rq

# What Trials were conducted for Drug1?

xxx-PoolListTrials.rq

# How many of each Phase?

# How many patients participated in Drug1 trials?

# What is the gender breakdown for all Drug1 trials?

# Identify patients that participated in multiple trials.

Congratulations! You've made to the

---- END OF THE EXERCISES ----

|  |  |  |
| --- | --- | --- |
|  | Stop here and wait for the instructor.  Presentation follows |  |

# Appendix 1: Course Resources

* [Introduction to Semantic Web](http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web)

http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web

* [What is Linked Data?](http://www.cambridgesemantics.com/semantic-university/what-linked-data)

http://www.cambridgesemantics.com/semantic-university/what-linked-data

* [Introduction to Linked data](http://www.cambridgesemantics.com/semantic-university/introduction-linked-data)

http://www.cambridgesemantics.com/semantic-university/introduction-linked-data

* [Learning SPARQL](http://www.learningsparql.com/)

http://www.learningsparql.com/

-------------------------------------------------------------------------------------------------------

[Hands-on Workshop Github Repository](https://github.com/phuse-org/LinkedDataWorkshop/tree/master/Annual2017-EU)

Course content is available at:

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