PhUSE CSS

Introduction to Linked Data

Hands-on Workshop

EXERCISES

Version 3.0

March 2018

DRAFT TO BE UPDATED FROM 2017 ANNUAL

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

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

Contents

[Introduction 2](#_Toc493085067)

[Required Materials 2](#_Toc493085068)

[Symbols 2](#_Toc493085069)

[Server Login 3](#_Toc493085070)

[Exercises 4](#_Toc493085071)

[1. Neo4j 4](#_Toc493085072)

[1.1 Diagram the Data Model 4](#_Toc493085073)

[1.2 Transfer Diagram to Spreadsheet 6](#_Toc493085074)

[1.3 Upload to Neo4j 8](#_Toc493085075)

[1.4 Query and Visualize 9](#_Toc493085076)

[1.4.1 Who enrolled in Study 1? 9](#_Toc493085077)

[1.4.2 What treatment did Bob receive? (Graph result) 10](#_Toc493085078)

[1.4.3 What treatment did Bob receive? (Value result) 10](#_Toc493085079)

[1.4.4 Graph Exploration 10](#_Toc493085080)

[2. Resource Description Framework (RDF) 11](#_Toc493085081)

[2.1 RDF Model Spreadsheet to the RDF Diagram 11](#_Toc493085082)

[2.2 Assign ObjectType 12](#_Toc493085083)

[2.3 Create RDF (TTL) File 13](#_Toc493085084)

[2.4 Query and Visualize 14](#_Toc493085085)

[2.4.1 Who enrolled in Study 1? 15](#_Toc493085086)

[2.4.2 What treatment did Bob receive? (Triples + Graph result) 16](#_Toc493085087)

[2.4.3 What treatment did Bob receive? (Value result) 16](#_Toc493085088)

[2.4.4 Graph Exploration 16](#_Toc493085089)

[Demonstrations 17](#_Toc493085090)

[Appendix 1: Error Messages and Resolutions 18](#_Toc493085091)

[A1.1 Neo4jFromSpreadsheet.R : Error Messages and Resolutions 18](#_Toc493085092)

[A1.2. RDFFromSpreadsheet.R : Error Messages and Resolutions 18](#_Toc493085093)

[A1.3 Unprotect .XLSX Spreadsheet in Libre Office 19](#_Toc493085094)

[Appendix 2: Course Resources 19](#_Toc493085095)

# Introduction

Instructions in this document are specific to the PhUSE cloud server environment 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:
  + LDWorkshopExercises.pdf (this document)
  + LDWorkshopInfoSheet.pdf
  + GraphEditorSheet.pdf (may not be needed if all info can be displayed within the app)

## Symbols and Conventions

SUBJECT and OBJECT URIs are written in [color] uppercase. Object literals are in mixed case.

Relations are written in lowercase green. These case conventions should be followed when creating these items in the Graph Editor.

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

(IP Address provided by Instructor)

**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

# Graph Editor

# Prepare

An initial model containing a small number of entities in a clinical trial is provided as starting point. You will add additional nodes and relations to the diagram before translating it into RDF Linked Data

1. **Sign in to the cloud server** as described in the [Server Login](#_Server_Login) section.
2. Double click the shortcut to the Graph Editor <ADD SHORTCUT Screenshot>
3. Observe the two nodes at the top of the editor that show a STUDY URI and a label for the Study as a String Literal.
4. Observe the nodes at the bottom left of the editor that represent gender codes (F,M) and their corresponding NCI codes.

**Guidelines for adding nodes and relations**

Follow these guidelines when creating the nodes, p:v pairs, and relations. Mandatory requirements are shown in bold.

|  |  |
| --- | --- |
| Image result for information icon | Many of these recommendations serve to simplify the exercises and do not apply outside of the workshop. |

**Nodes**

|  |  |
| --- | --- |
| **Guideline** | **Examples/Explanation** |
| URIs (link to other nodes or can link) |  |
| Use **UPPERCASE letters** | PERSON1, not Person1 or person1.  Exercise requirement only. |
| **No spaces or special characters ($"%&^!~…etc.) in the names** | PERSON2, not "Person 2" |
| Short node names | PERSON2 not "Person 2 in the Clinical Study" |
| Follow naming conventions already used in the diagram | PERSON2, not Person\_2, not Patient2 |

**Node Type**

Each node must be assigned a **node type**. For simplicity, this exercise uses only three:

|  |  |
| --- | --- |
| **uri** | Object nodes that link to other nodes in your existing model, or represent things that *could* link to other nodes. These Objects would become Subjects in another relation, or *could* become Subjects if more data was available. |
| **string** | Character values with no outbound link.  Examples: "Bob", "Protocol for Study 123", "Male", "11-Oct-2017", etc. Dates can be represented as strings for this exercise. |
| **int** | Integer values with no outbound link. Example: 32 |

**Relations (Links, Edges)**

|  |  |
| --- | --- |
| **Guideline** | **Examples/Explanation** |
| Use **lowercase letters** | treatmentarm, not TREATMENTARM or treatmentArm. Helps to differentiate from Nodes. |
| **No spaces or special characters ($"%&^!~, etc.)** | Special characters may have unanticipated effects in the conversion scripts. |
| Relationships have *direction* and *describe* how the nodes are related. Pick names that define the relation concisely. |  |
|  |  |

# Add a STUDY

Create your own, unique study for PRODUCT1. The study URI name is found in the LDWorkshopInfoSheet.pdf.

1. xxxx
2. xxxxxx
3. Add a node the designates your new STUDY node as rdf:type STUDY

a. xxxx

b. xxxxxx

1. Connect your new **STUDY** node to the existing **PRODUCT1** node.

a. xxxx

b. xxxxxx

# Add Treatments to the STUDY

Create two treatments for the study.

1. xxxx
2. xxxxxx
3. Connect your new **STUDY** node to the existing **PRODUCT1** node.

a. xxxx

b. xxxxxx

# Add a PERSON1 to the STUDY and a Treatment

1. xxxx
2. xxxxxx
3. Connect your new PERSON(n) node new Subject them with a gender relation to either the M or F node.

a. xxxx

b. xxxxxx

# Add another person to the STUDY

Add a new, unique person to the study.

1. xxxx
2. xxxxxx
3. Connect your new PERSON(n) node new Subject them with a gender relation to either the M or F node.

a. xxxx

b. xxxxxx

# Add additional concepts and entities

Add other entities relevant to the study....[ADD guidance]

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

# Export to TTL

[steps to export. finding the file, etc.]

# Upload to Triplestore

[steps to upload]

# Explore the Data

# Visualize your graph

[Steps to create a visual of the graph. Using R as pre previous workshop?]

# Query your graph

[steps to be added]

# Who enrolled in your study?

[steps to be added]

# How many patients received which treatments?

[steps to be added]

# Linking to Other Graphs

# CDISC Terminology

# Leveraging an Ontology

# Federated Query\* (\*assumes can be configured for workshop)

# What studies were conducted for PRODUCT1?

# How many patients participated in PRODUCT1 studies?

# What is the gender breakdown for all PRODUCT1 studies?

# Identify patients that participated in multiple studies.

# etc.

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/Annual2017-EU