

# **INTRODUCTION TO LINKED DATA AND GRAPH DATABASES**

**PHUSE CSS WORKSHOP**

**MARCH 19, 2017**

# WORKSHOP GOALS

## HANDS-ON EXPERIENCE WITH:

- Labeled Property Graphs (LPG)
- Resource Description Framework (RDF)

## PREPARE YOU FOR:

- CSS breakout sessions & projects
- Use Graph data in your own work
- ...*seeing graphs EVERYWHERE!*

# INSTRUCTOR AND ASSISTANTS

Tim Williams                      UCB BioSciences

Scott Bahlavooni                d-Wise

Ian Fleming                      d-Wise

Links to Workshop Scripts and Presentation PDF:  
<https://github.com/phuse-org/LinkedDataWorkshop>

# OUTLINE

- *Server Login*
- Introduction and Graph Overview
- Labeled Property Graph (LPG)
- Resource Description Framework (RDF)

*Time Permitting:*

- Federated Query
- Discussion

# OUTLINE

- **Server Login**
- Introduction and Graph Overview
- Labeled Property Graph (LPG)
- Resource Description Framework (RDF)
- Discussion

# SERVER LOGIN

## HANDOUT:

- Login Instructions.
- Exercises

*Assistance provided while the next section is covered.*

# OUTLINE

- Server Login
- **Introduction and Graph Overview**
- Labeled Property Graph (LPG)
- Resource Description Framework (RDF)

# WHY DATA AS A GRAPH?

## ONE EXAMPLE: SDTM DOMAINS



# SDTM DM DATA

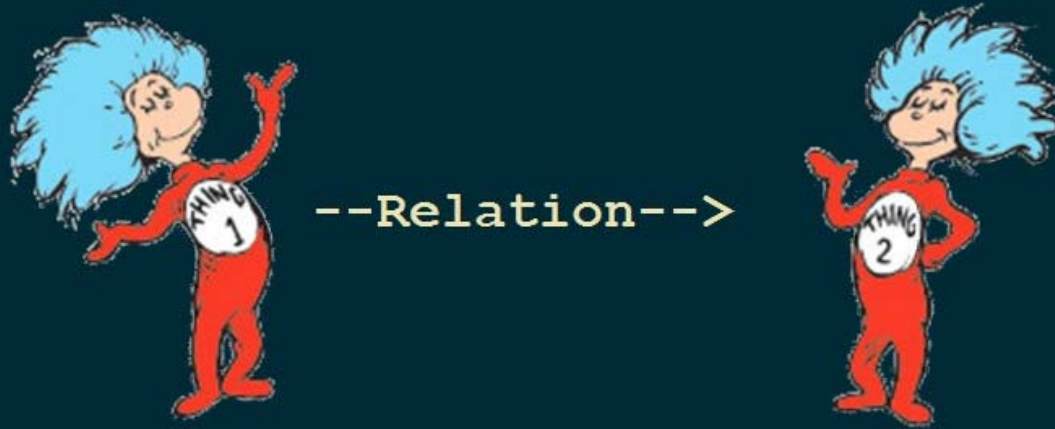
	A	B	C	D	E	O	P	Q	R	S	T	U	V	W	X
1		studyid	domain	usubjid	subjid	age	ageu	sex	race	ethnic	armcd	arm	actarmcd	actarm	country
2	1	CDISCPILLOT01	DM	01-701-1015	1015	63	YEARS	F	WHITE	HISPANIC OR LATINO	Pbo	Placebo	Pbo	Placebo	USA
3	2	CDISCPILLOT01	DM	01-701-1023	1023	64	YEARS	M	WHITE	HISPANIC OR LATINO	Pbo	Placebo	Pbo	Placebo	USA
4	3	CDISCPILLOT01	DM	01-701-1028	1028	71	YEARS	M	WHITE	NOT HISPANIC OR LAT	Xan_Hi	Xanomelir	Xan_Hi	Xanomelir	USA
5	4	CDISCPILLOT01	DM	01-701-1033	1033	74	YEARS	M	WHITE	NOT HISPANIC OR LAT	Xan_Lo	Xanomelir	Xan_Lo	Xanomelir	USA
6	5	CDISCPILLOT01	DM	01-701-1034	1034	77	YEARS	F	WHITE	NOT HISPANIC OR LAT	Xan_Hi	Xanomelir	Xan_Hi	Xanomelir	USA
7	6	CDISCPILLOT01	DM	01-701-1047	1047	85	YEARS	F	WHITE	NOT HISPANIC OR LAT	Pbo	Placebo	Pbo	Placebo	USA

What is wrong here?

- Inflexible, version specific row x column structure and format
- Mixture of concepts
- No integral metadata
- Data repetition

**HOW CAN GRAPH DATA FIX THIS??**

# DATA AS A GRAPH?



A Comparison

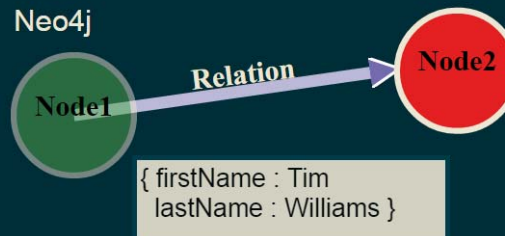
# SCREEN SHOT OF INTERACTIVE GRAPH

## RDF, NEO4J COMPARISON

RDF Triple



Neo4j



# OUTLINE

- Server Login
- Introduction and Graph Overview
- **Labeled Property Graph (LPG)**
- Resource Description Framework (RDF)

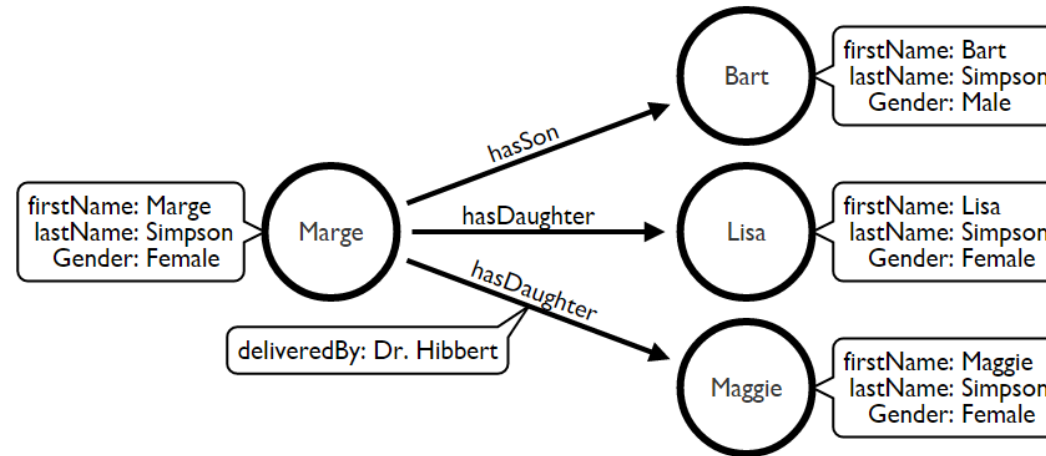
# **LABELED PROPERTY GRAPH (LPG)**

## **SIMPSONS FAMILY IN NEO4J**

WhiteBoard

# SCREEN SHOT OF WHITEBOARD MODEL

Simpsons Family. Marge and Children - for Neo4j



# NEO4J EXERCISES PART 1



## EXERCISE

### 1. Simpsons Family in Neo4j

# NEO4J: CREATE NODES AND RELATIONS

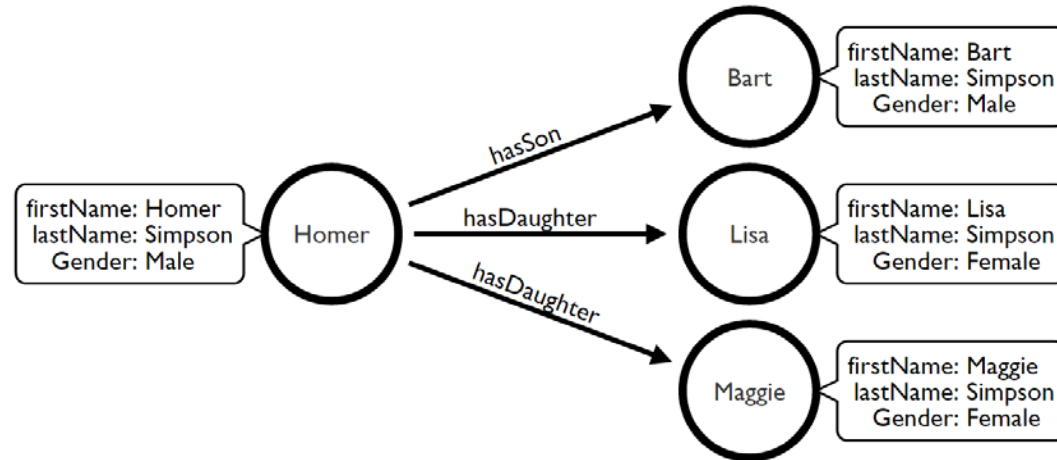
## ADD HOMER AND THE RELATIONSHIPS TO HIS CHILDREN

Homer WhiteBoard



# SCREEN SHOT OF WHITEBOARD MODEL

Simpsons Family. Homer and Children - for Neo4j



# NEO4J EXERCISES PART 2

## ADD HOMER AND RELATIONSHIPS TO HIS CHILDREN



### EXERCISE

#### 1.3 Create

# END OF NEO4J SECTION

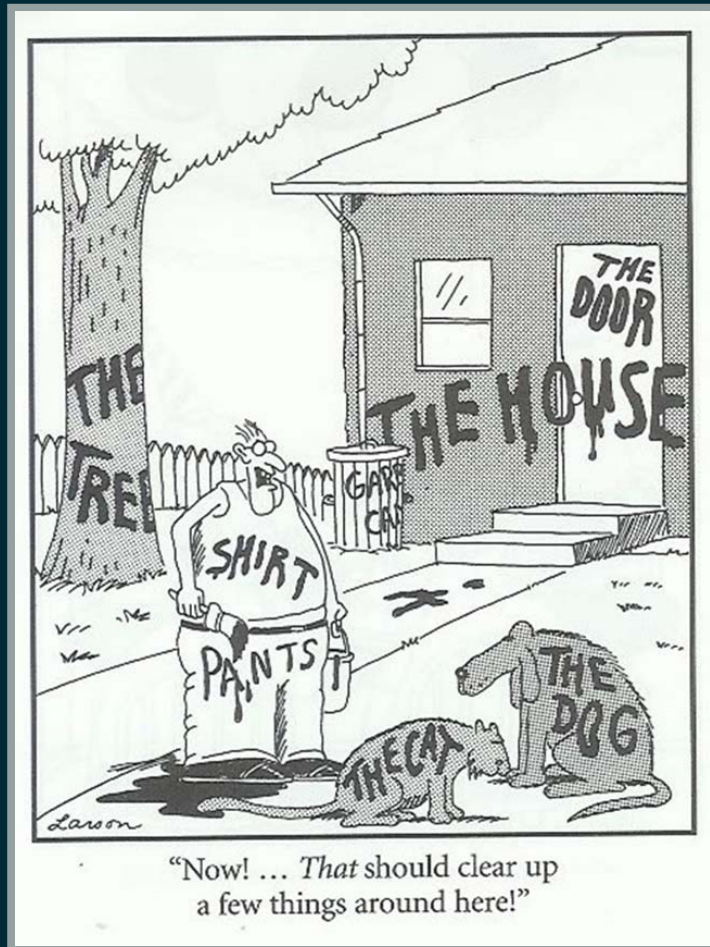
## IN THIS SECTION YOU:

- Explored a graph visually and using Cypher
- Created nodes and relations
- Queried nodes, relations, and the properties of each

# OUTLINE

- Server Login
- Introduction and Graph Overview
- Labeled Property Graph (LPG)
- **Resource Description Framework (RDF)**
- Discussion

# CORE RDF CONCEPTS



- All **things** have a **name**
- All names are unique and **addressable**
  - HTTP URI
- Things are **linked**
  - Directed graphs
- Links have **meaning**
  - Semantics

# TOOLS FOR CREATING AND EDITING RDF

Workshop:

- Protege Ontology Editor

Others:

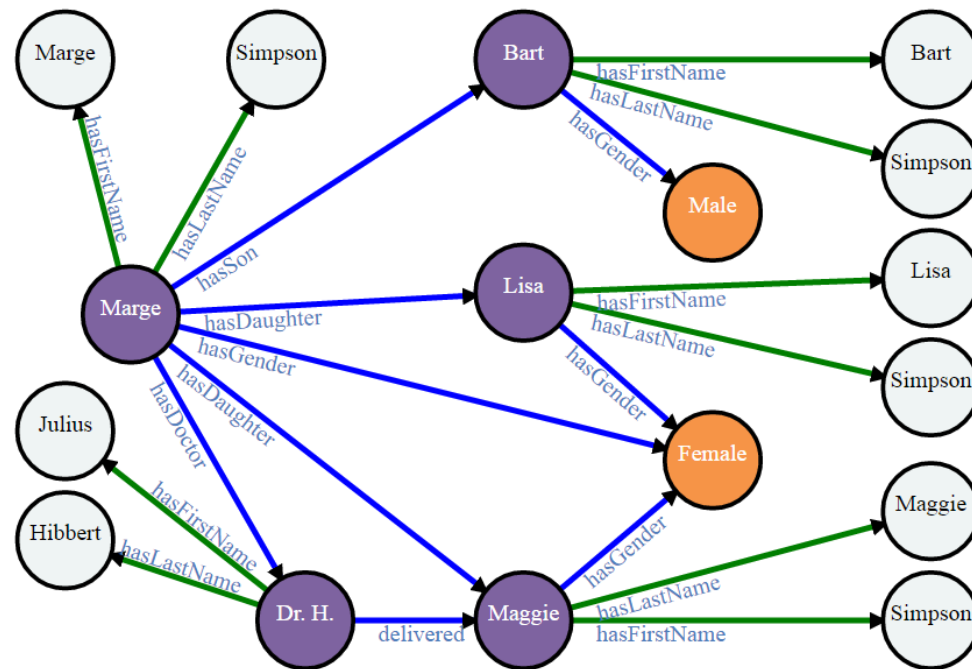
- SPARQL
- TopBraid, Ontorion Fluent Editor, Text Editor, R, Python, many others...

# SIMPSONS FAMILY AS RDF

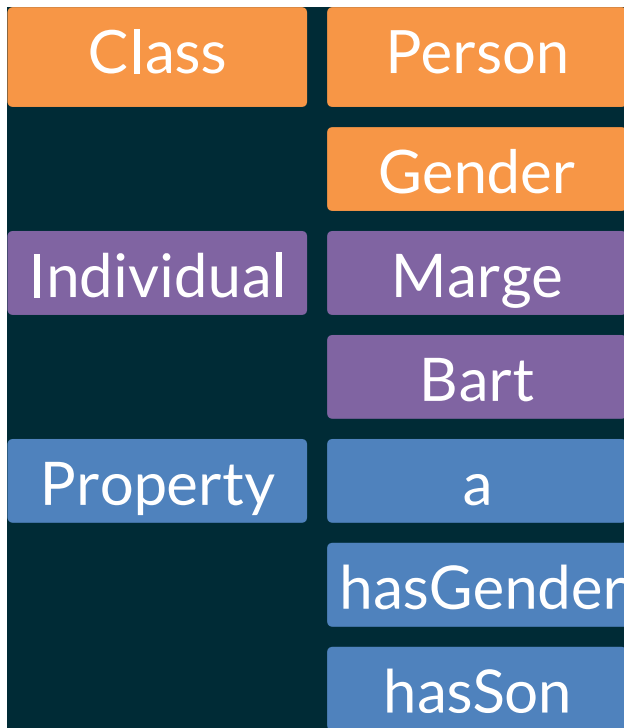
Data model

# SCREEN SHOT OF INTERACTIVE GRAPH

Simpsons Family: Marge and Children - for RDF







## AXIOMS



# RDF EXERCISES PART 1



## EXERCISE

### 2. Simpsons Family in RDF

## CYPHER: 101-LisaGender.cql

```
MATCH (a:Person)
WHERE a.firstName = "Lisa"
RETURN a.gender
```

## SPARQL: 201-LisaGender.rq

```
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT ?gender
WHERE {
    simpsons:Lisa simpsons:hasGender ?gender
}
```

## CYPHER: 102-MargeSon.cql

```
MATCH (pers1)-[:hasSon]-(pers2)
WHERE pers1.firstName='Marge'
RETURN pers2.firstName
```

## SPARQL: 202-MargeSon.rq

```
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT ?sonFirstName
WHERE {
    ?person simpsons:hasSon ?son ;
            simpsons:hasFirstName ?parentFirstName .
    ?son      simpsons:hasFirstName ?sonFirstName
    FILTER(regex(?parentFirstName, "^Marge"))
}
```

## CYPHER: 103-MaggieDoc.cql

```
MATCH (pers1)-[r:hasDaughter]-(pers2)
WHERE pers2.firstName='Maggie'
RETURN r.DeliveredBy as DeliveryDoctor
```

## SPARQL: 203-MaggieDoc.rq

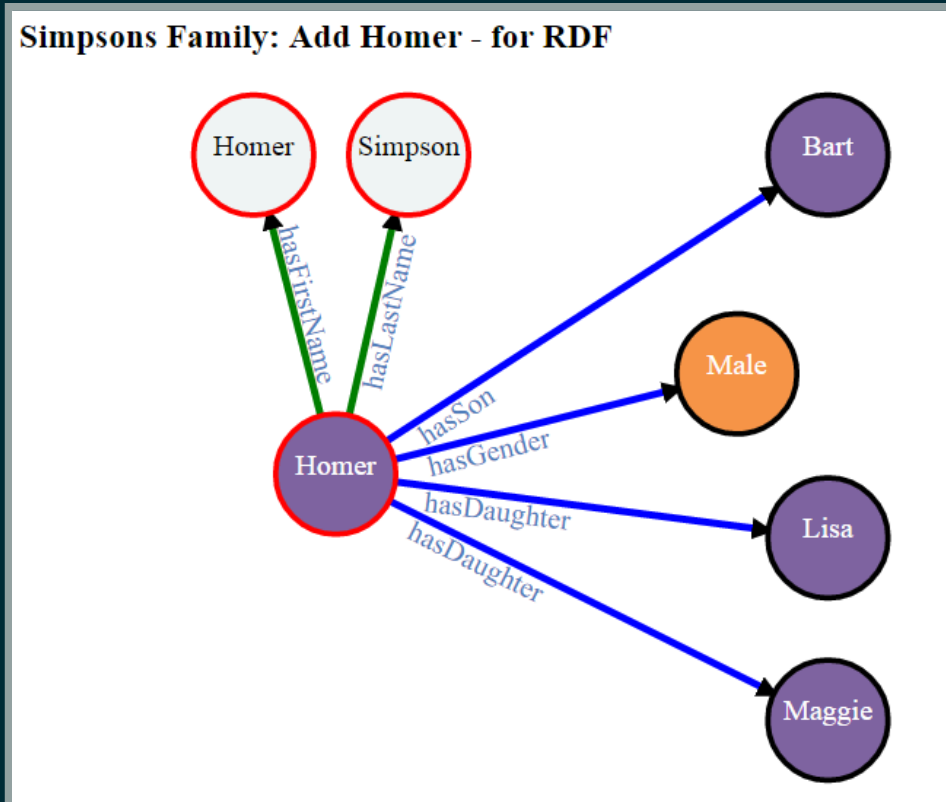
```
# NB: Protege uses SPARQL1.0 so CONCAT() not avail.
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT    ?title ?docFirstName  ?docLastName
WHERE {
    ?doc simpsons:delivered simpsons:Maggie ;
          simpsons:hasFirstName ?docFirstName_ ;
          simpsons:hasLastName ?docLastName_ .
    BIND( STR(?docFirstName_) AS ?docFirstName )
    BIND( STR(?docLastName_) AS ?docLastName )
    BIND("Dr.  " AS ?title)
}
```

# SIMPSONS FAMILY

## MODEL HOMER AND THE RELATIONSHIPS TO HIS CHILDREN

Homer WhiteBoard

# SCREEN SHOT OF INTERACTIVE GRAPH



# RDF EXERCISES PART 2

## ADD HOMER AND RELATIONSHIPS TO HIS CHILDREN



### EXERCISE

### 2.3 Create



## CYPHER: 107-HomerChildren.cql

```
MATCH (person1)-[r]->(person2)
WHERE (person1.firstName = 'Homer' AND
      (type(r) = 'hasSon' OR
       type(r) = 'hasDaughter'))
RETURN DISTINCT person2.firstName
```

## SPARQL: 204-HomerChildren.rq

```
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT ?child
WHERE
{
    {simpsons:Homer simpsons:hasDaughter ?child . }
  UNION
    {
        {simpsons:Homer simpsons:hasSon ?child . }
    }
}
```

## CYPHER: 107-HomerChildren.cql (modified)

```
MATCH (person1)-[r]->(person2)
WHERE (person1.firstName = 'Homer' AND
      (type(r) = 'hasSon' OR
       type(r) = 'hasDaughter'))
RETURN COUNT(person2)
```

## SPARQL: 205-HomerChildCount.rq

```
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT (COUNT(?child) AS ?count)
WHERE
{
    {simpsons:Homer simpsons:hasDaughter ?child . }
    UNION
    {
        {simpsons:Homer simpsons:hasSon ?child . }
    }
}
```

## CYPHER: 108-Parent.cql

```
MATCH (person1)-[r]->(person2)
WHERE (type(r) = "hasSon" OR
       type(r) = "hasDaughter")
RETURN DISTINCT person1.firstName
```

## SPARQL:206-Parent.rq

```
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT DISTINCT ?parent
WHERE
{
    {?parent simpsons:hasDaughter ?child .}
    UNION
    {
        {?parent simpsons:hasSon ?child . }
    }
}
```

## CYPHER: 110-LisaBrother-Name.cql

```
MATCH a = (brother)-[:hasSon]-(parent)-[:hasDaughter]
WHERE daughter.firstName='Lisa'
RETURN DISTINCT brother.firstName
```

## SPARQL: 207-LisaBrother.rq

```
PREFIX simpsons: <http://www.example.org/Simpsons#>
SELECT DISTINCT ?brother
WHERE
{
    ?parent simpsons:hasDaughter simpsons:Lisa ;
            simpsons:hasSon ?brother .
}
```

# EXTEND THE KNOWLEDGEBASE WITH OWL

Asserted triple:

Marge

hasSon

Bart

Add rules for Inferencing:

- *hasChild* as parent property of hasSon, hasDaughter
- *hasParent* as inverse of *hasChild*

Then use these new properties in queries!

# INFERRED TRIPLES

- Deduced by a reasoner applying rules on top of data
- Rules for relationships and data not in the original source!



# OWL

## Key Point:

- OWL rules usually applied on top of data.
- Exists separate from the instance data\*

# RDF EXERCISES PART 3



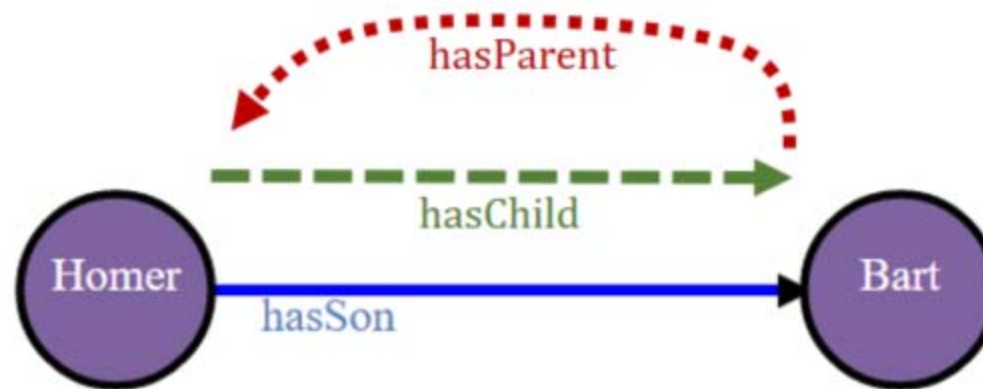
## EXERCISE

### 2.5 Extend with OWL



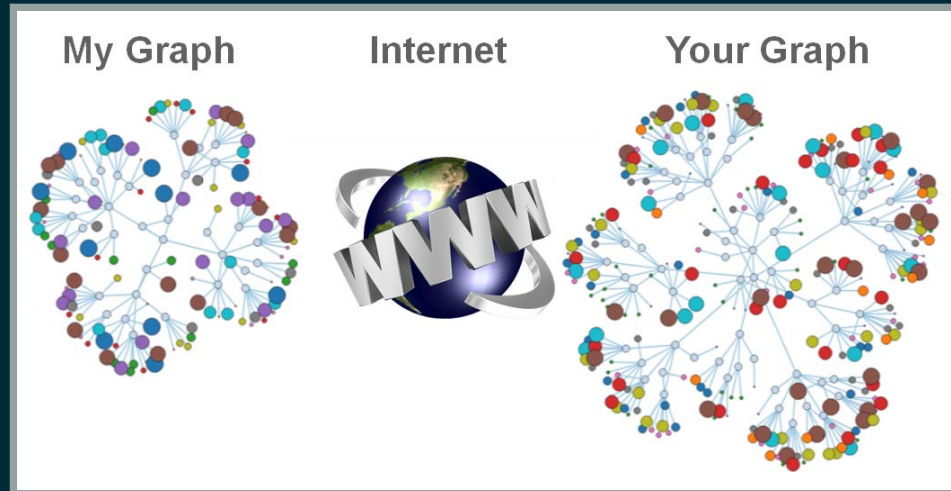
## Simpsons Family: Homer -- Bart

### Asserted and Inferred Relations



# FEDERATED QUERY

COMBINE DATA ACROSS (OFTEN REMOTE) GRAPHS!



# THE ELUSIVE FEDERATED QUERY



# MORE LIKELY SCENARIOS

- Download graph, upload to local triplestore
  - SDTM terminology, NCI thesaurus, Randomized Clinical Trials Ontology...
- Merge across graphs within your company using *Thing1-->owl:sameAs-->ThingOne*
  - Metadata repository
  - Clinical trials design
  - Company terminology/ontology
  - EHR... etc.

# FEDERATED QUERY OF DRUG DATA

Paper: "Consolidating drug data on a global scale using Linked Data" - Jovanovik & Trajanov

- 5-star Linked drug data from 23 countries
- Links to DrugBank and DBpedia

Paper is included with the files for the exercises.

# FEDERATED QUERY EXERCISE

EXECUTE A FEDERATED QUERY AT A SPARQL ENDPOINT



EXERCISE

## 3. Federated Query

# OUTLINE

- Server Login
- Introduction and Graph Overview
- Labeled Property Graph (LPG)
- Resource Description Framework (RDF)
- **Discussion**

# CHART THE PROS AND CONS BY CATEGORY

*Interactive Discussion*



# PLANS FOR CSS

## LINKATHON WITH EMERGING TECHNOLOGY'S "ALTERNATIVE TRANSPORT FORMAT" PROJECT.

- Monday afternoon, Tuesday morning.
- Workshop XPT versus RDF, LPG, XML, JSON-LD(?)  
...others ??

**KICKOFF OF LD&GD PROJECT: "SDTM DATA AS RDF",  
TUESDAY (1 HOUR)**

# CONTACT INFO

Tim Williams

tim.williams@PhUSE.eu



@NovasTaylor



<https://www.linkedin.com/in/timpwilliams/>

# RESOURCES

- Workshop materials, including the SPARQL and CYPHER scripts, plus PDF of this presentation:  
<https://github.com/phuse-org/LinkedDataWorkshop>

# NEO4J RESOURCES

Recommended Overview	<a href="https://neo4j.com/developer/graph-database/">https://neo4j.com/developer/graph-database/</a>
Overview of graph db and Neo4j [optional]	<a href="https://youtu.be/U8ZGVx1NmQg">https://youtu.be/U8ZGVx1NmQg</a> [45min]
Intro to Cypher [optional]	<a href="https://www.youtube.com/watch?v=1TSBXZMv6tc">https://www.youtube.com/watch?v=1TSBXZMv6tc</a> [49min]
Graph Modeling [optional]	<a href="https://www.youtube.com/watch?v=AaJS-DGBQX4">https://www.youtube.com/watch?v=AaJS-DGBQX4</a> [42min]
RDF in Neo4j [optional]	<a href="http://guides.neo4j.com/rdf-graphs/">http://guides.neo4j.com/rdf-graphs/</a>

# RDF AND PROTEGE RESOURCES

Introduction to Semantic Web	<a href="http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web">http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web</a>
What is Linked Data?	<a href="http://www.cambridgesemantics.com/semantic-university/what-linked-data">http://www.cambridgesemantics.com/semantic-university/what-linked-data</a>
Introduction to Linked data	<a href="http://www.cambridgesemantics.com/semantic-university/introduction-linked-data">http://www.cambridgesemantics.com/semantic-university/introduction-linked-data</a>
Protege Application	<a href="https://www.youtube.com/watch?v=8Nf2xf5akoM">https://www.youtube.com/watch?v=8Nf2xf5akoM</a>

*Thank you  
&  
Good night!*