

# EU Connect 2021

The Clinical Data Science Conference

15th–19th November

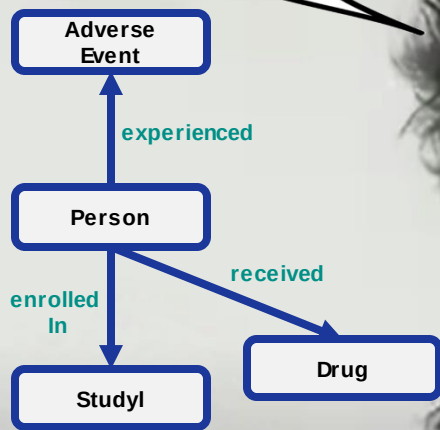
## Knowledge Graph Visualization Information Solution or Ball of Confusion?

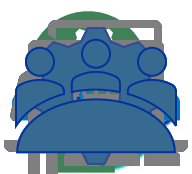
# Visualizing Knowledge Graphs

**Information  
Solution?**

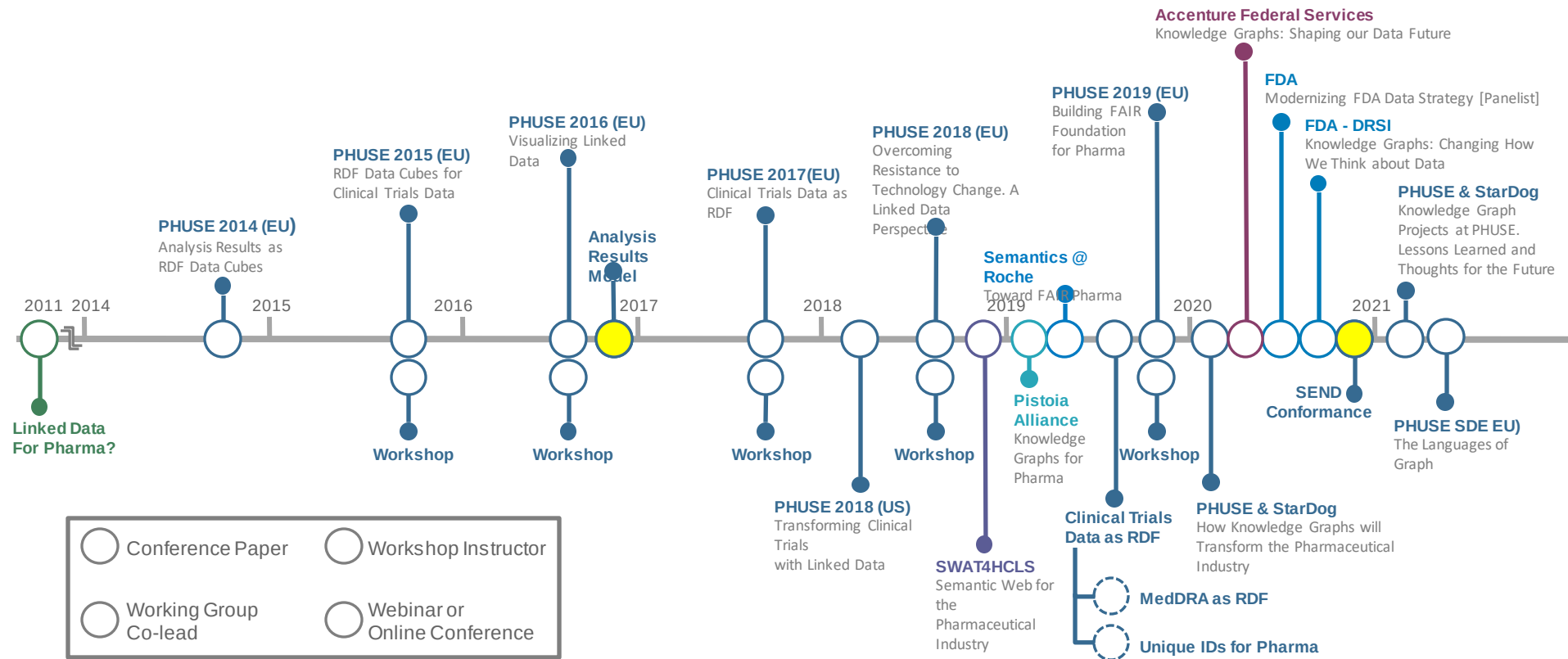
**OR**

**Ball of  
Confusion!**





# Knowledge Graphs: My Personal Timeline



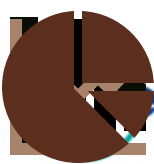


# This talk is:

- A basic introduction to Graph Data and Knowledge Graphs
  - Additional learning materials in [Appendix 1: Resources and References](#).
- About how visualization can both *help* and *hinder* understanding
- The Future for Knowledge Graphs in the pharmaceutical industry.

## This talk is not:

- An overview of **Knowledge Graph Visualization Tools**
  - Select examples presented in the [Appendix 2: Applications and Examples](#)
- An introduction to Graph Databases and related concepts.

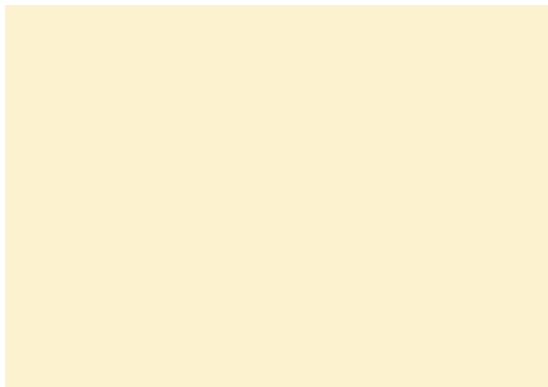


# The Treachery of Semantics. What is a Graph?



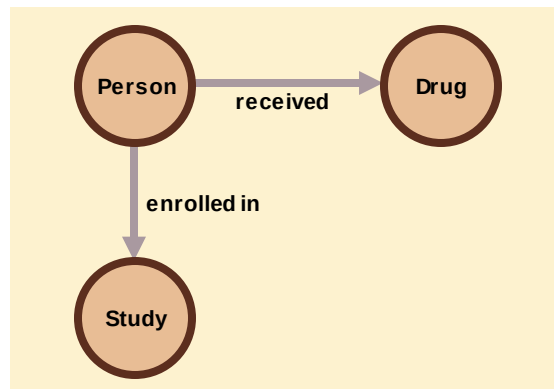
*This is not a pipe.*

- "...it's just a representation [of a pipe]"  
René Magritte
- Source: The Treachery of Images  
[https://en.wikipedia.org/wiki/The\\_Treachery\\_of\\_Images](https://en.wikipedia.org/wiki/The_Treachery_of_Images)



*This is not a graph*

- These are charts, plots, visualizations...
- It's just a representation of data.



*This is a graph.*

- Discrete mathematics
- A set of vertices and edges (nodes and links)



# The Treachery of Semantics. What is a Graph?

“...some mathematicians might claim they own [the word *graph*] in reference to connection and network graphics, - that is, ‘graph theory’..”

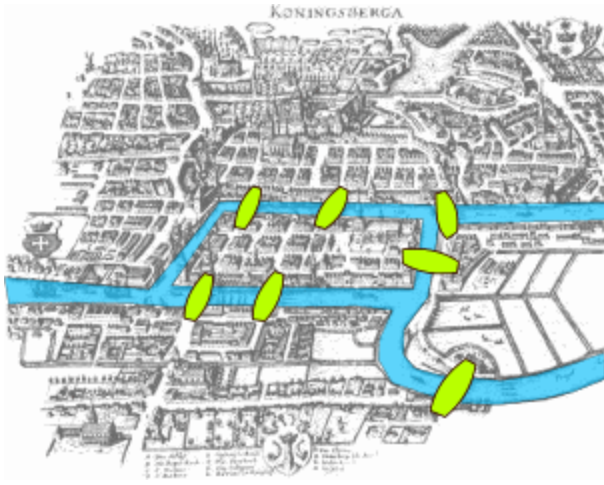
“I’m not willing to unleash a turf war, so I’ll just say that you are both right...”  
- Alberto Cairo, “TheTruthful Art”

“When graph data is displayed using network and force network *graphs*, use of ‘graph’ is appropriate for these visualizations that illustrate the connected nature of graph data.”

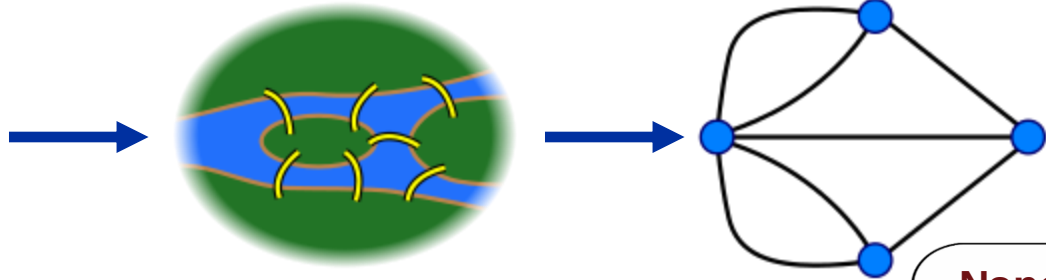
- Tim Williams , Graph Nerd  
“Interactive Visualization of Linked Data”, PHUSE 2016

# The Birth of Graph Theory

Leonard Euler, 1736 “Can you walk through the city and cross each bridge once, and only once?”



The Seven Bridges of Königsberg



Nope!

Euler proved mathematically:





# What is a Knowledge Graph?

- Semantic graph data that is interpretable by both machines and humans
- Use an Ontologies
  - Dictionary – terms and definitions
  - Taxonomy – class hierarchy
  - Thesaurus - relationships between terms
  - Rules and Restrictions - Group membership, exclusions, types
- Use a reasoner to infer values and relationships not in the original data

Free e-book

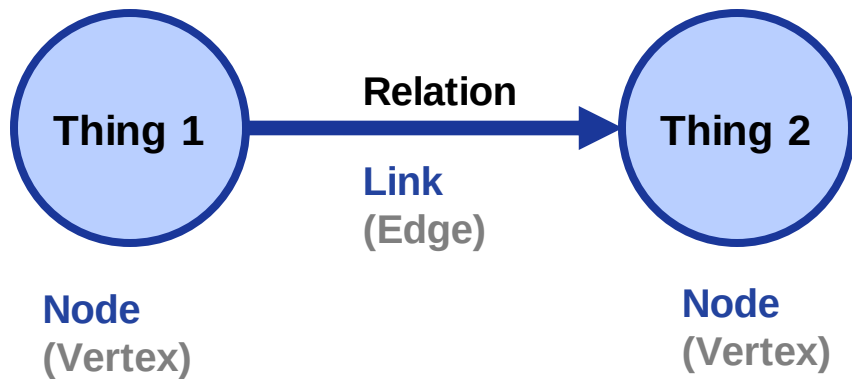
[A Data Engineer's Guide to Semantic Modelling](#)

- Ilaria Maresi (June 2020)

<http://blog.thehyve.nl/news/ebook-semantic-model>



# Graph Data is Elegantly Simple





# Different Types of Graphs

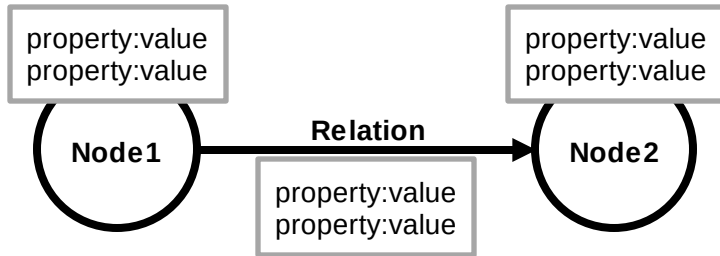
## Resource Description Framework (RDF)



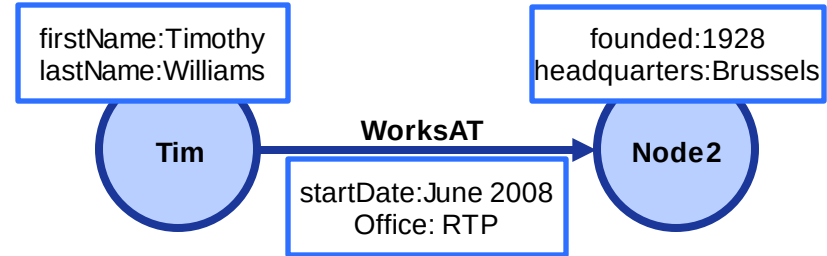
Example:



## Labeled Property Graph



Example: neo4j



# Which Type of Graph Should I Use?

*We will not have  
this fight today.*





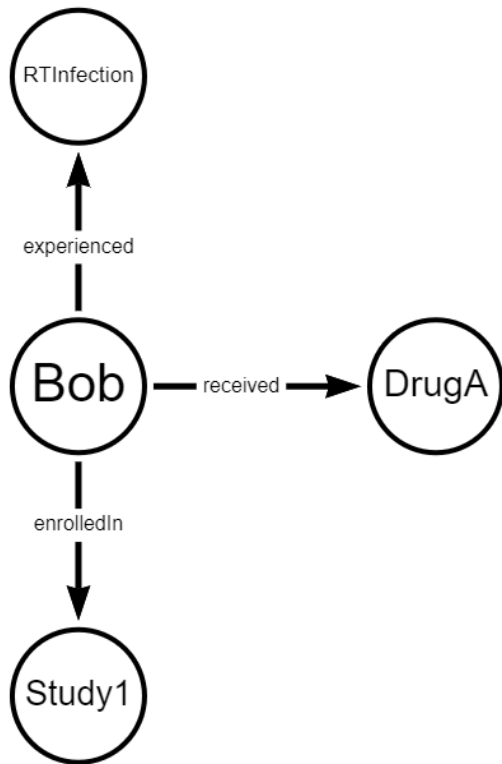
# Graph Whiteboard Application

---

<https://arrows.app/>



# The (Whiteboard) Graph



The Whiteboard *IS* the data

The graph is the *data*,  
The data is the *graph*.

**THE GRAPH  
IS THE  
DATA**

**THE DATA  
IS THE  
GRAPH**

**THE GRAPH  
IS THE  
DATA**

**THE DATA  
IS THE  
GRAPH**

**THE GRAPH  
IS THE  
DATA**



**THE MEDIUM  
IS THE  
MESSAGE**

**THE MEDIUM  
IS THE  
MESSAGE**

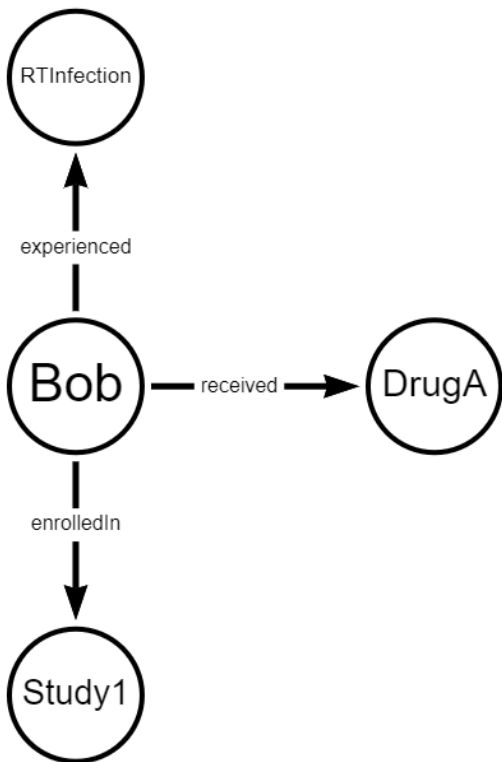
**THE MEDIUM  
IS THE  
MESSAGE**

**THE MEDIUM  
IS THE  
MESSAGE**





# The Graph



# IS The Data



***RDF** Graph data in abbreviated  
**Terse Triple Language (TTL)***

```
:Bob      :enrolledIn  :Study1 ;  
:received  :DrugA      ;  
:experienced :RTInfection.
```



# Read and Display the Data

```
sparql <-'  
  SELECT ?from ?to ?relation  
  WHERE {  
    ?from ?relation ?to .  
  }'  
'
```

Program: BobStudy1.R

from                      relation                      to

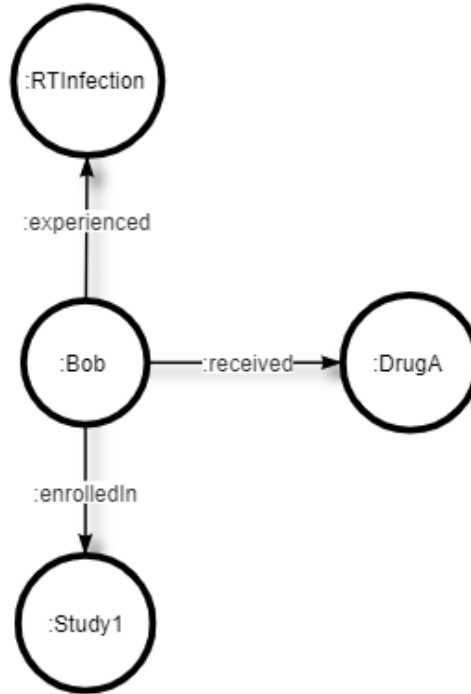
↓                                      ↓                                      ↓

:Bob    :enrolledIn    :Study1 ;  
         :received       :DrugA ;  
         :experienced :RTInfection .



# Screen Capture from R Program

Select by id ▼

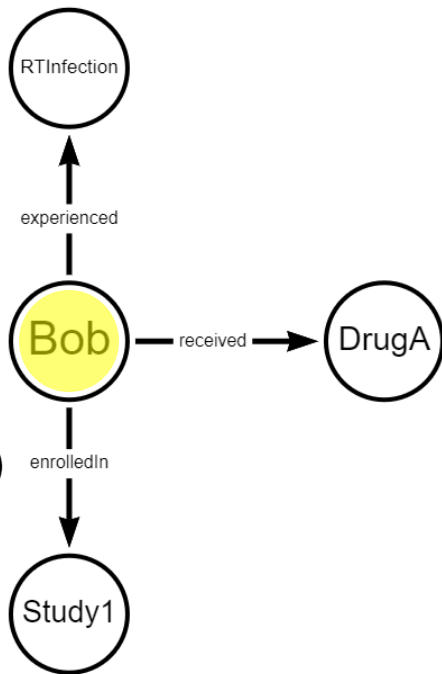




# The Graph



# IS The Data



Information  
Solution!

*The whiteboard model is the data model*

Graph data in abbreviated  
*Terse Triple Language (TTL)*

```
Bob :enrolledIn :Study1 ;  
    :received   :DrugA ;  
    :experienced :RTInfection .
```

## IS The Graph

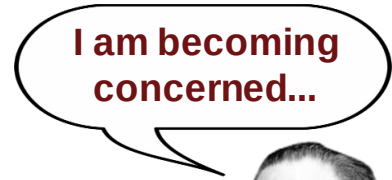
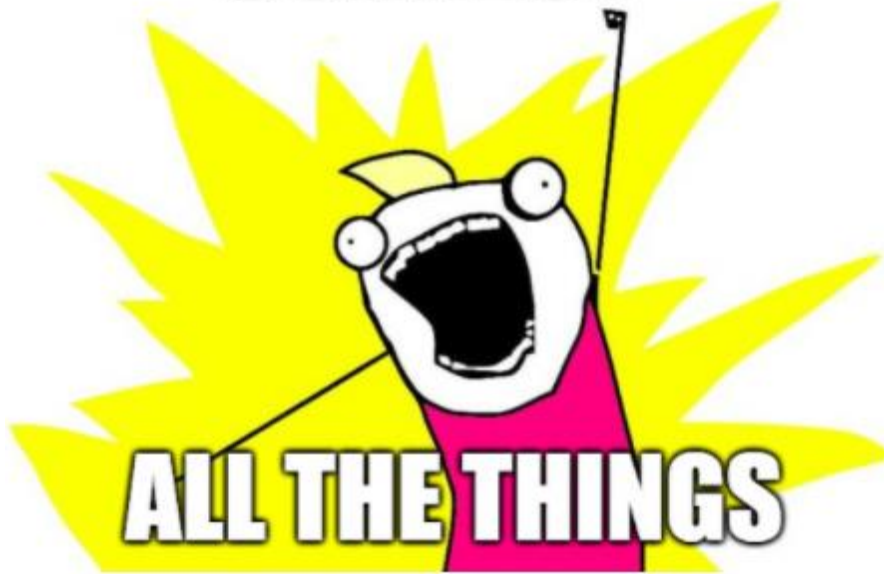


Agreed!



# When the Medium Obscures the Message

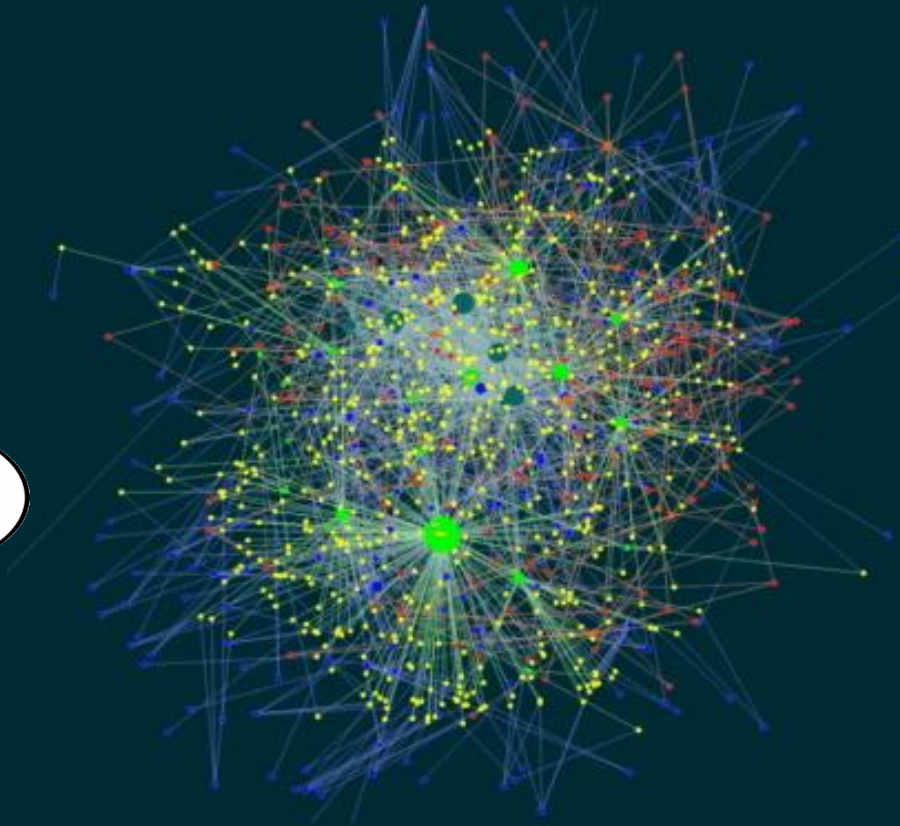
CONNECT and visualize



# PHUSE Conference 2016

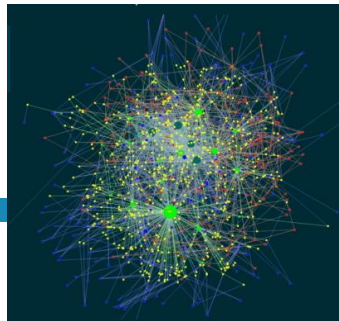
Attendees, Presentations,  
Workshops

**Ball of  
Confusion!**





# Medium Subverts Message



- “Knowledge Graphs are just pretty, interactive visualizations”
  - Visualization as tool for understanding complex data and the connections within it. The plot is not the message!
- “From the plot, I see Knowledge Graphs are just connected data (I can do that in relational systems)”
  - Knowledge Graphs are ***much*** more:
    - Reasoners infer new values and relationships
    - Built-in: Unique identifiers, integral metadata
    - Semantic Data Mesh, Virtual Graphs, Foundation for Machine Learning, AI...
    - Can be more intuitive than relational systems

# 7,563 main characters in the Expanded Star Wars Universe

- Kirell Benzi <https://bit.ly/3j95RkV>

Is your data ready for this?  
Do you have the **tools** and  
**skills** to extract information?

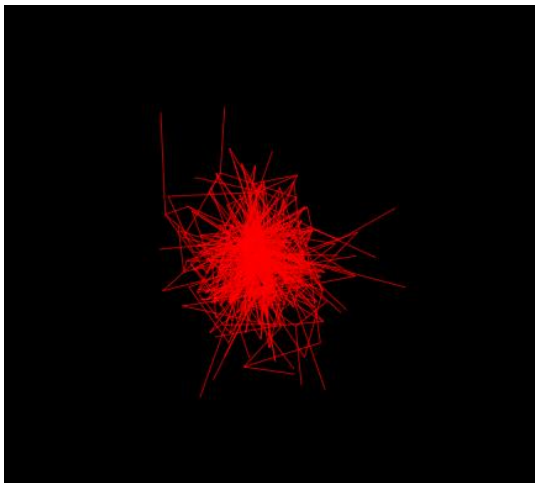
Ball of  
Confusion!

Where is Baby  
Yoda?

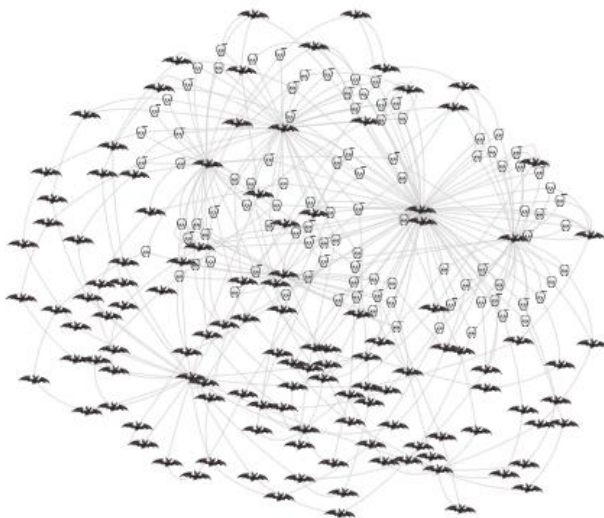


<https://bit.ly/KGVisFun>

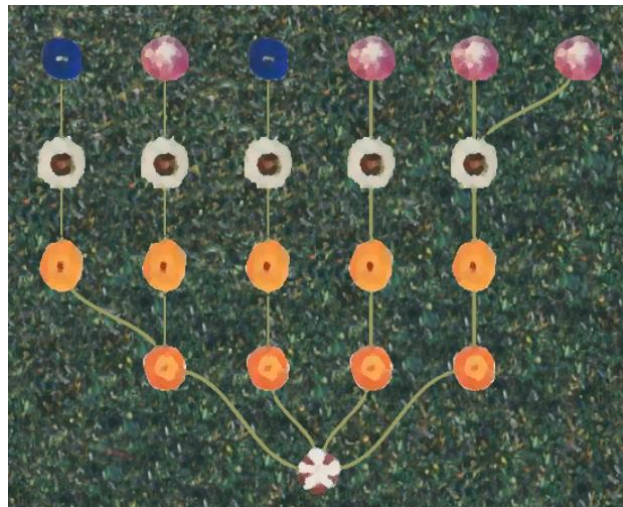
## ANGER



## FEAR



## WATERCOLOURS

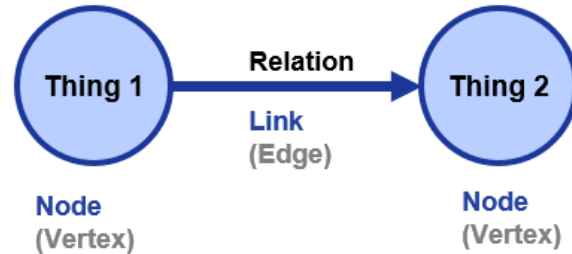




# What happened to this?



## Graph Data is Elegantly Simple



“People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated.”

- attributed to Dan Brickley and Libby Miller  
in “Validating RDF Data”





# Example

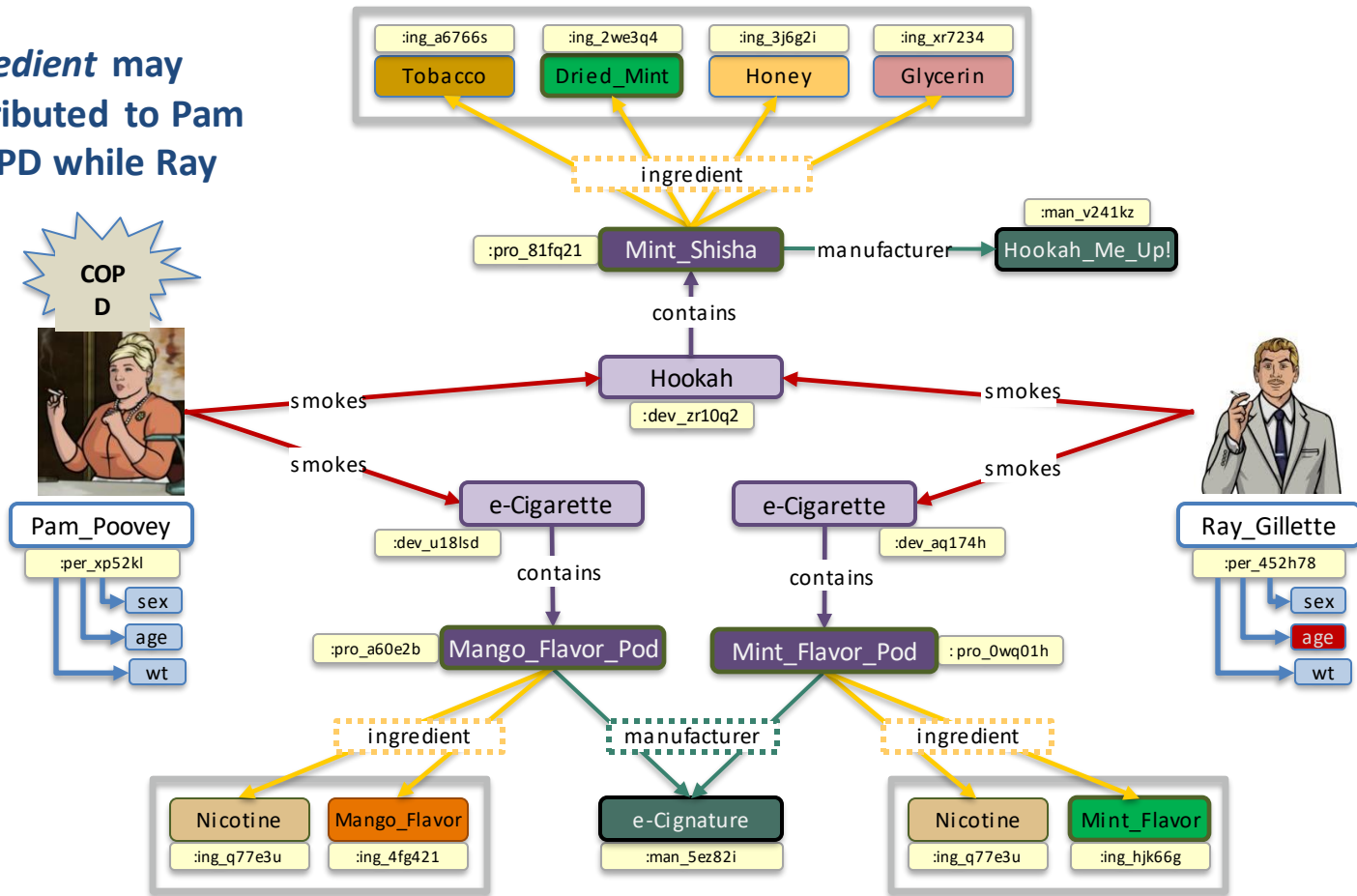
## **“Knowledge Graphs: Changing How We Think About Data”**

Presented to:

FDA Department of Regulatory Science & Informatics (DRSI)

- August, 2020

What *ingredient* may have contributed to Pam having COPD while Ray does not?



# Exposure as Graph Data

View the data file at:

<https://bit.ly/PamAndRayTTL>

```
:per_xp52k1 :name "Pam_Poovey"^^xsd:string ;
            a      :Person ;
            :age    "38"^^xsd:integer ;
            :wt     "205"^^xsd:integer ;
            :sex    :F ;
            :smokes :dev_ul8lsd, :dev_zrl0q2 .

:per_452h78 :name "Ray_Gillette"^^xsd:string ;
            a      :Person ;
            :age    "403"^^xsd:integer ;
            :wt     "165"^^xsd:integer ;
            :sex    :M ;
            :smokes :dev_aql74h, :dev_zrl0q2 .

:dev_ul8lsd skos:prefLabel "e-Cigarette"^^xsd:string ;
            a      :eCigarette ;
            a      :Device ;
            :contains :pro_a60e2b .

:dev_zrl0q2 skos:prefLabel "Hookah"^^xsd:string ;
            a      :HookahPipe ;
            a      :Device ;
            :contains :pro_8lfq2l .

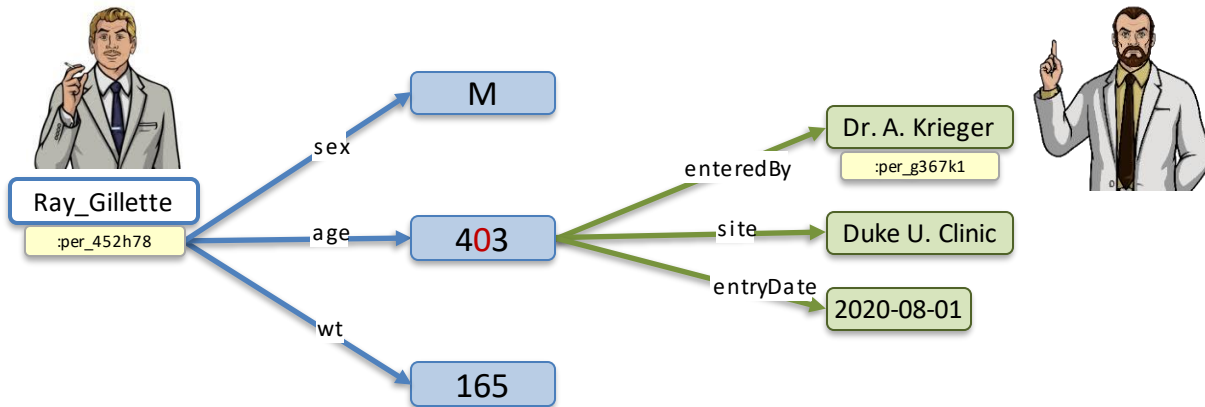
:dev_aql74h skos:prefLabel "e-Cigarette"^^xsd:string ;
            a      :eCigarette ;
            a      :Device ;
            :contains :pro_0wq0lh .

:pro_8lfq2l skos:prefLabel "Mint_Shisha"^^xsd:string ;
            a      :Product ;
            a      :TobaccoMix ;
            :manufacturer :man_v24lkz ;
            :ingredient   :ing_a6766s, :ing_2we3q4, :ing_3j6g2i, :ing_xr7234 .

:pro_a60e2b skos:prefLabel "Mango Flavor Pod"^^xsd:string ;
            a      :Product ;
            a      :FlavorPod ;
```



# Metadata is Also Part of the Graph



**What does this look like as a graph?**



**Ball of  
Confusion!!**



**That's what the  
world is today.**

**Hey  
hey!!!**



***This is the actual data!***

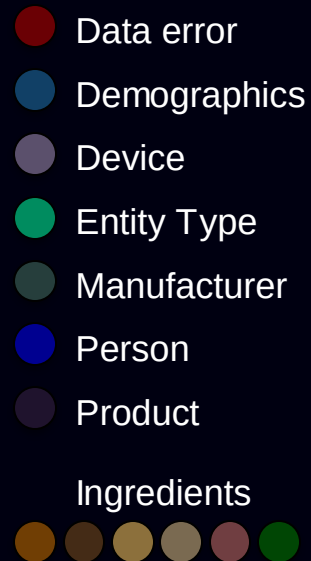
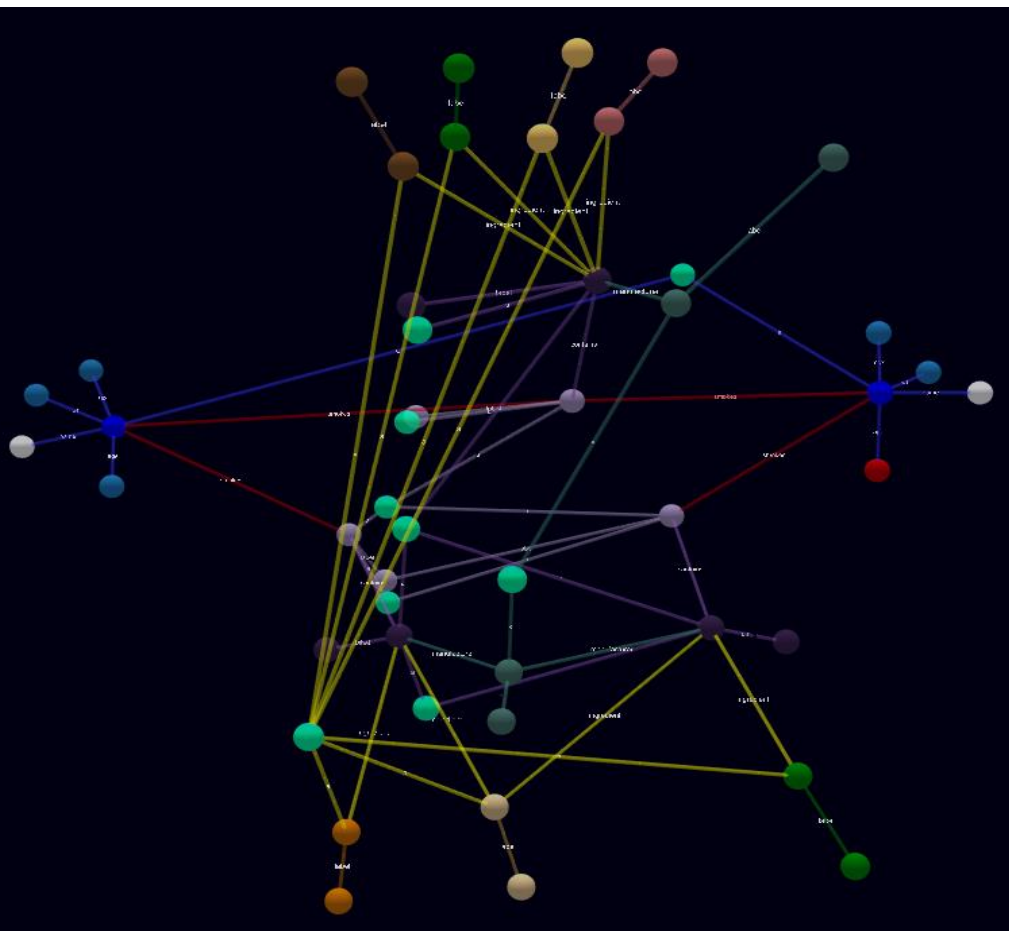


Music Video segment: <https://www.youtube.com/watch?v=-ALRLZQf42s&t=196s>

# 3-D Force Graph

View the interactive  
visualization at:

<https://bit.ly/PamAndRay>



# The Graph is The Data. So Query it!

PamUniqueIngred.rq

```
PREFIX : <http://example.org/Eg#>
```

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
SELECT ?ingredient ?manufacturedBy
```

```
WHERE{
```

```
# Pam : Person 1 Exposure
```

```
?person1      :smokes      ?smokeDevice ;
```

```
|||           :name        ?personName .
```

```
?smokeDevice  :contains    ?mixture .
```

```
?mixture      :ingredient   ?ingred ;
```

```
|||           :manufacturer ?man .
```

```
?ingred       skos:prefLabel ?ingredient .
```

```
?man          skos:prefLabel ?manufacturedBy .
```

```
FILTER( regex(?personName , "Pam"))
```

```
# Ray : Person 2 Exposure
```

```
OPTIONAL{
```

```
?person2      :smokes      ?smokeDevice2 ;
```

```
|||           :name        ?personName2 .
```

```
?smokeDevice2 :contains    ?mixture2 .
```

```
?mixture2     :ingredient   ?ingred2 .
```

```
?ingred2      skos:prefLabel ?ingredient.
```

```
FILTER( regex(?personName2 , "Ray"))
```

```
# Ingredients that are common to both will be BOUND to ingredient2
```

```
FILTER (?ingred = ?ingred2)
```

```
}
```

```
# Keep only those that are not in the ingredient 2 set, i.e. not bound
```

```
# as an ingredient for person2
```

```
FILTER(! BOUND(?ingred2))
```

```
}
```

What is Pam's unique exposure ingredient?

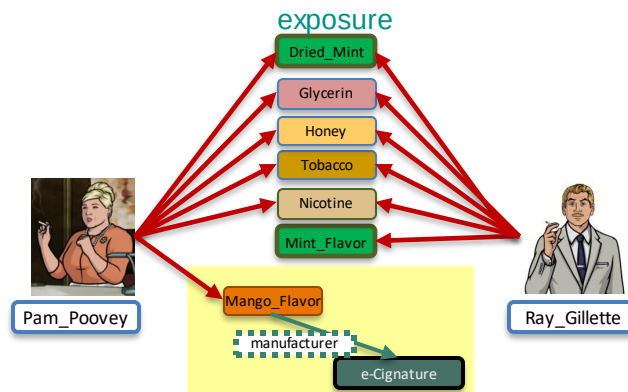
Query Result

 Run to File

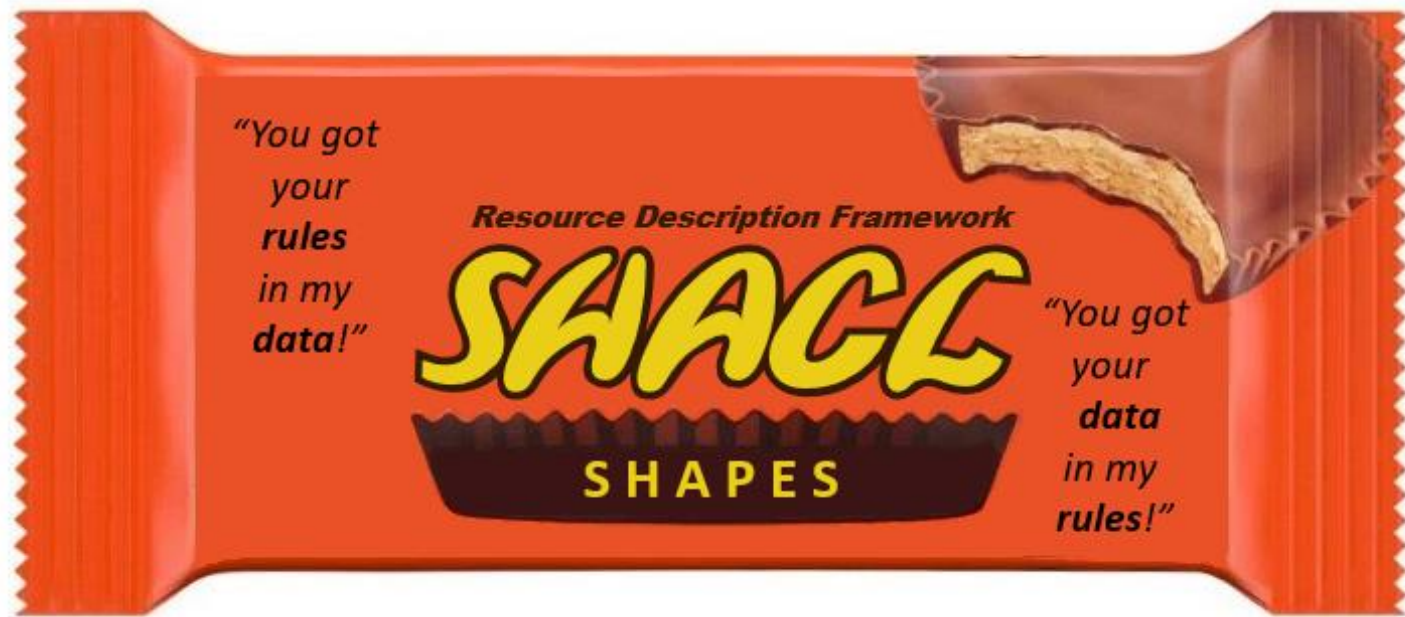
 Visualize

1 Results, 46 ms

| ingredient     | manufacturedBy |
|----------------|----------------|
| "Mango_Flavor" | "e-Cigarette"  |



# Visualizing Graph Data Validation





# SHApes Constraint Language (SHACL)

Data is a Graph  
Data has Shape



**“Person Shape”**  
(Validation Constraints)

Validation is a Graph  
Validation has Shape



**Person Data**

# Validating Ray's Demographics



Ray Gillette

**Demographics**

sex: M

age: 43

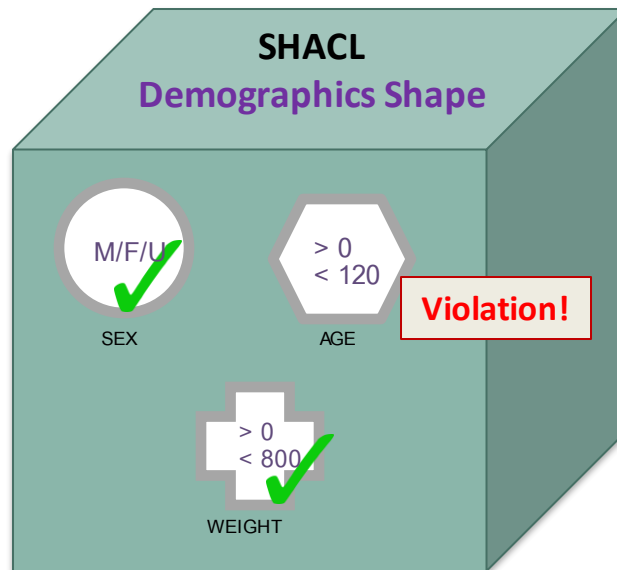
wt: 165

**Data Entry**

M

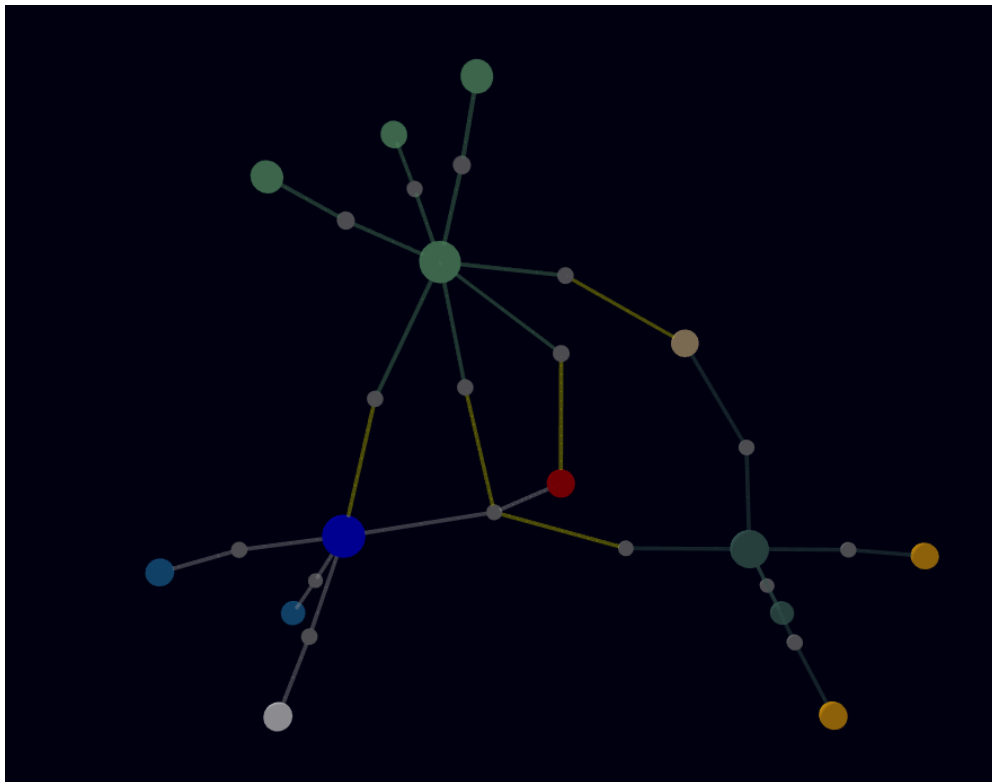
403

165





# Demographics, Constraints, Report

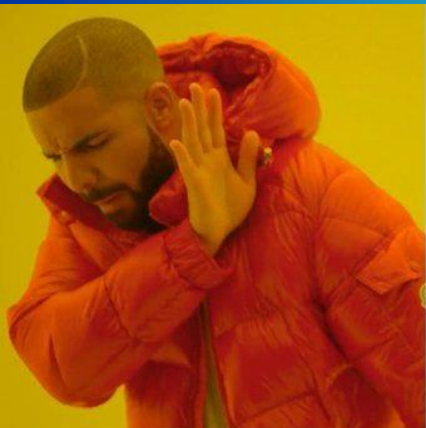


View the interactive visualization at:

<https://bit.ly/RaySHACLResult>



# Validation Report – Triples to Tables



```
[
  a sh:ValidationReport ;
  sh:conforms false ;
  sh:result [
    a sh:ValidationResult ;
    sh:resultSeverity sh:Warning ;
    sh:sourceShape <https://www.example.org/foo/fdoc#keyword-minShape> ;
    sh:sourceConstraintComponent sh:MinCountConstraintComponent ;
    sh:focusNode <https://www.example.org/foo/fdoc#FILE_f93ba176> ;
    sh:resultPath <https://www.example.org/foo/fdoc#keyword> ;
    sh:resultMessage "File lacks keyword. At least one required."
  ] , [
    a sh:ValidationResult ;
    sh:resultSeverity sh:Violation ;
    sh:sourceShape <https://www.example.org/foo/fdoc#Status-ValueSetShape> ;
    sh:sourceConstraintComponent sh:InConstraintComponent ;
    sh:focusNode <https://www.example.org/foo/fdoc#FILE_ee3ba2c0> ;
```

## SHACL Validation Report is Triples



### Validation Report Summary

Report: [/data/rdf/SHACL FileDoc Report.TTL](#)

| File ↑                                          | Severity ↑       | Message ↑                                  |
|-------------------------------------------------|------------------|--------------------------------------------|
| <a href="#">Link</a> SDTMIGSPec-MH-EDGES.rq     | <b>Violation</b> | File Status is not in approved list.       |
| <a href="#">Link</a> SDTMSpecInstanceNoMatch.rq | <b>Warning</b>   | File lacks keyword. At least one required. |

- Query
- Reformat
- Present using RMarkdown



# Strategies for Avoiding the Ball of Confusion



# Subsetting + Traditional Displays

*Subset to the desired Study Graph*

## People Working on a Study

Main About

Product

☒ Drug1

Study

Study1

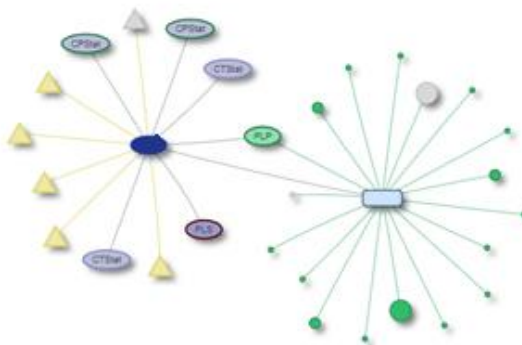
Study2

Study3

Study4

### Study: Study1

Statisticians, Leads, Code Committers



## People

| Role   | Status                               | Person   |
|--------|--------------------------------------|----------|
| CPStat | <span style="color: green;">●</span> | John Doe |
| CPStat | <span style="color: green;">●</span> | Jane Doe |
| CTStat | <span style="color: green;">●</span> | Dane Joe |
| CTComm | <span style="color: green;">●</span> |          |

*Select a Study.*

- Show staff, roles, experience
- Click-through to staff contact information on company Intranet

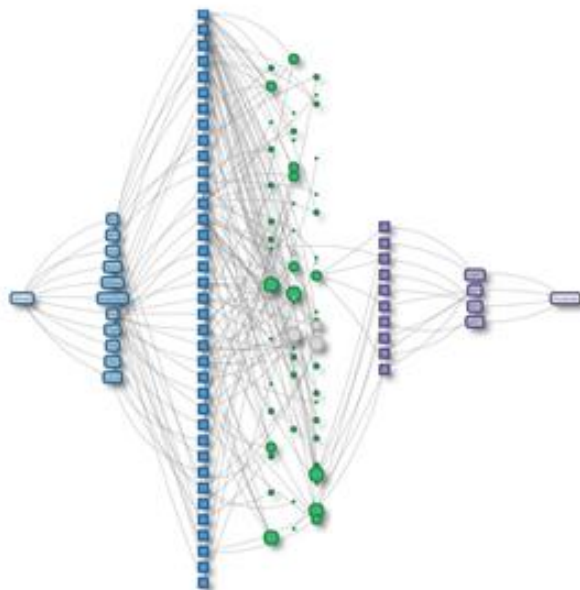
*Size, shape and colour of nodes for different types and frequency of entities*

*Table of Values includes live links to more information*

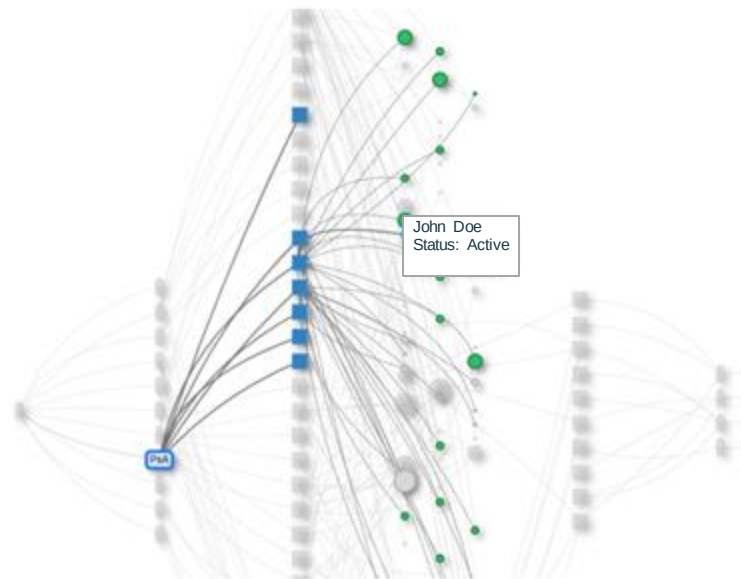
- Identify staff with the most programming experience for an indication
- Click-through to contact information



- *Interactive Display:*
- *Connection highlighting*
- *Information with mouseover,*
- *Click-through to more information.*



Find Programmers with Indication Experience





# Collapsible/Expandable Trees

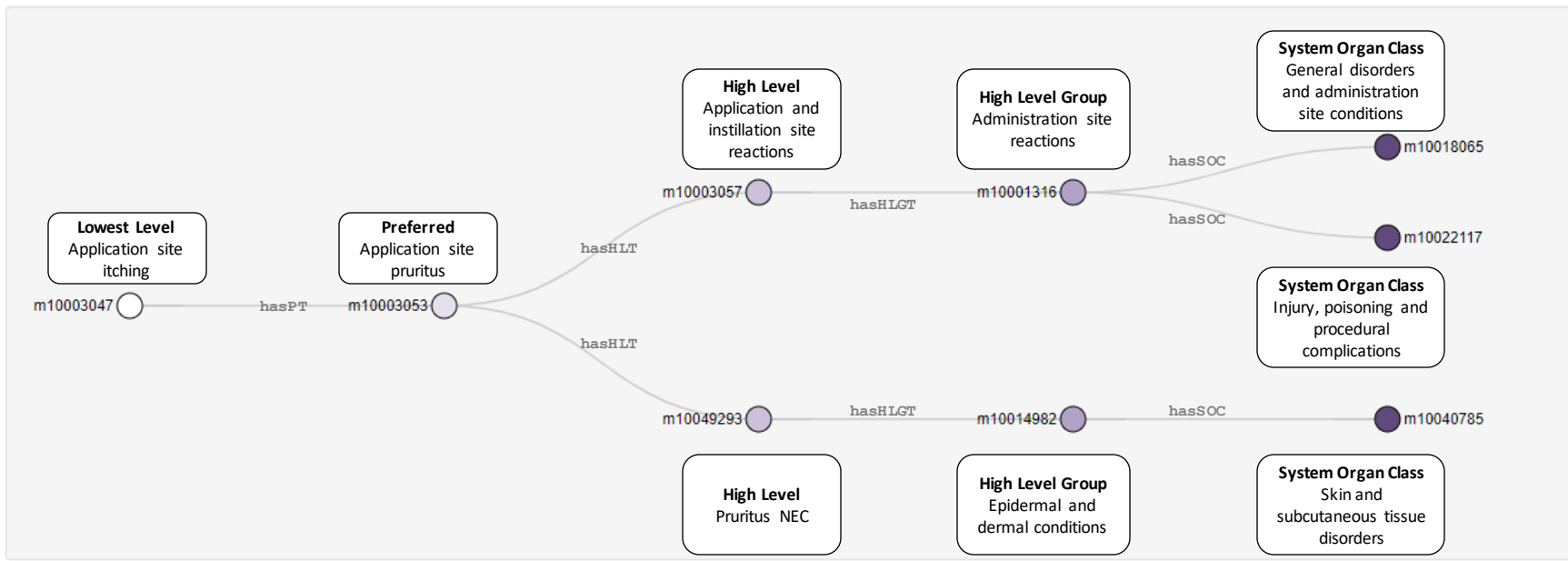
MedDRA Tracing from Lowest Level Term (LLT) to System Organ Class (SOC)

LLT :

Application Site Itching

*Subset to desired term*

*Interactive Display: Collapsible nodes, information with mouseover, click-through to more information.*

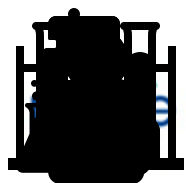




# A “Simplify” Summary



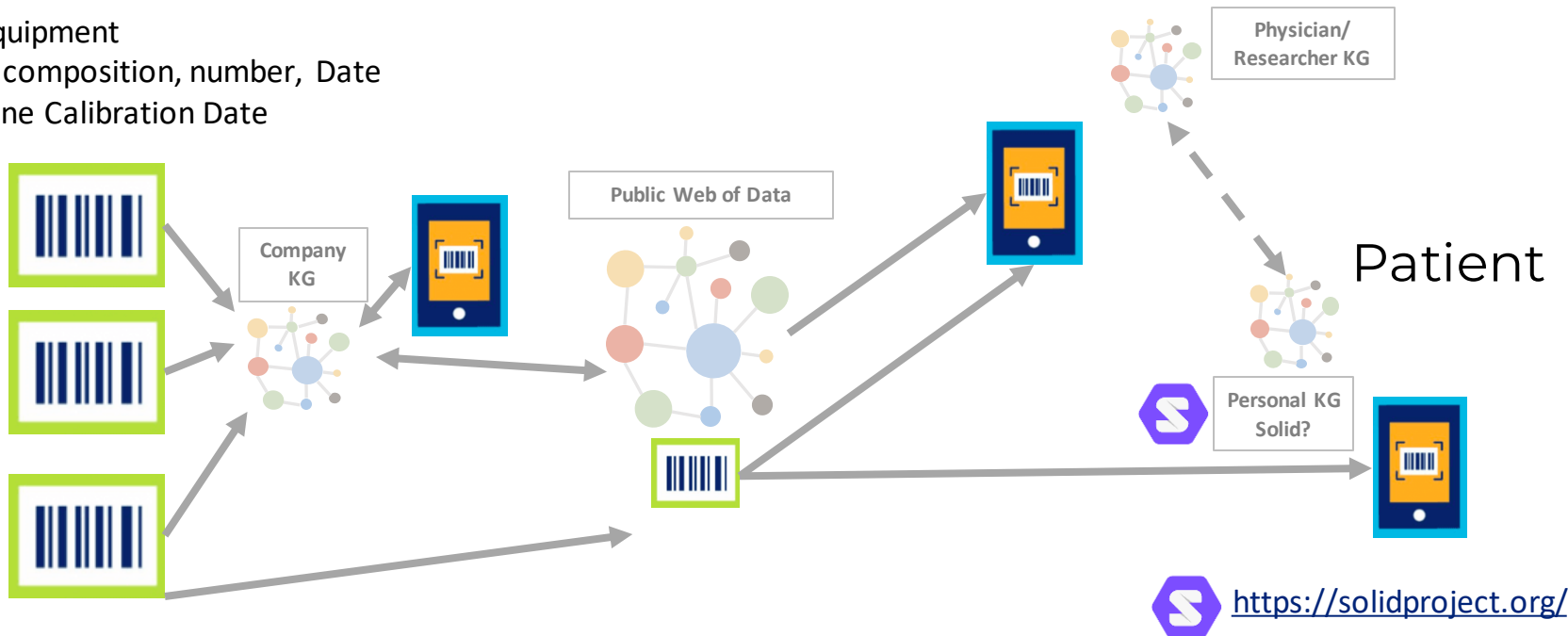
- Subset the graph to the nodes and relations of interest
- Clustering
- Path highlighting on selection
- Consistent Colours and Legends
- Use Interactivity
  - Collapsing / Expanding groups of nodes
  - Mouseover to expose underlying information of interest
  - Click-through to underlying data and sub-graphs
- Augment with familiar displays (Tables, listing, traditional plots)
- Familiar displays, but with values as URLs that lead to a path of data discovery



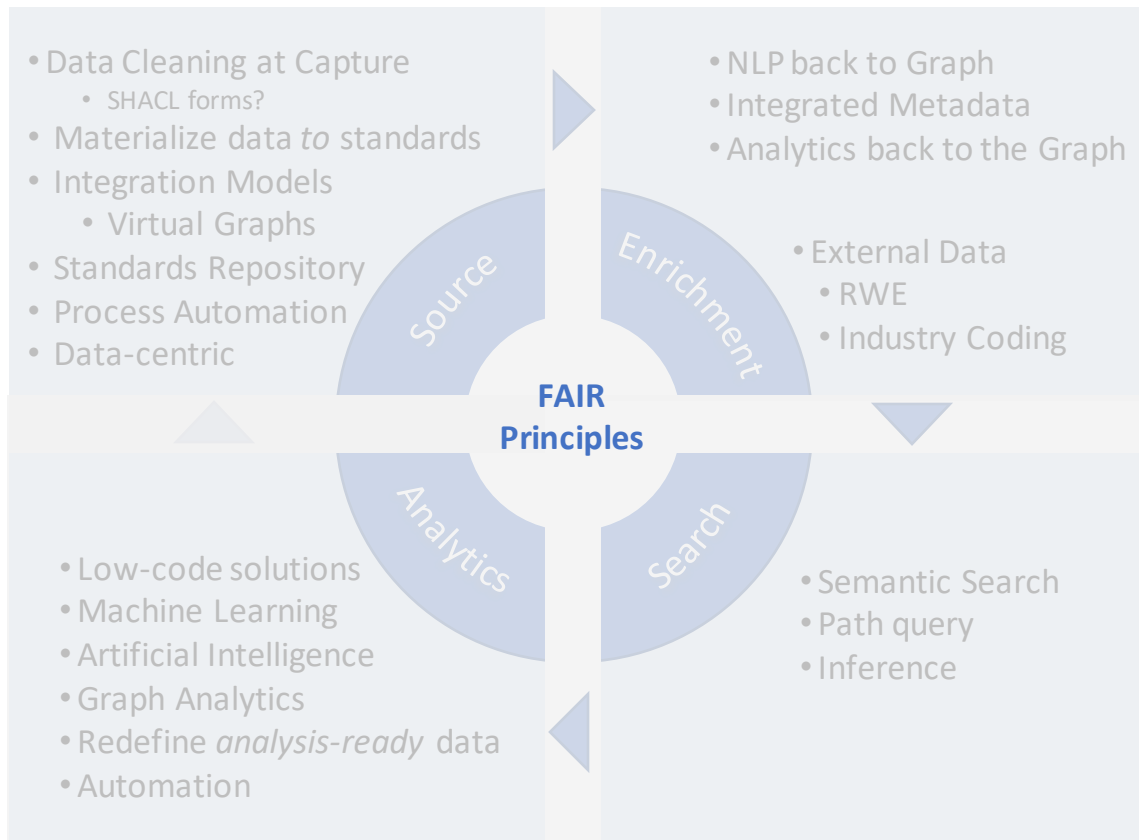
# Connecting the Physical to the Web of Data

## Lab / Pharma Co

- Lab equipment
- Batch composition, number, Date
- Machine Calibration Date



# Knowledge Graphs in Pharma



Adapted from:

<https://www.slideshare.net/AlanMorrison/datacentric-business-transformation-using-knowledge-graphs>

<https://www.go-fair.org/fair-principles/>



# Conclusion

## Visualizing Knowledge Graphs

- Informative
- Connections, pathways, similarities, clustering
- *Follow-your-nose* data exploration
- The white board (graph) *is the data*.

**“Rap on,  
brother, rap on!”**

The medium is  
the **massage**



1967



# Appendix 1 : Resources & References

## General / Introductory

- [A Data Engineer's Guide to Semantic Modeling](#) - Maresi . *Free e-book download.*
- [The Data Centric Revolution](#) - McComb

## Technical

- [Semantic Web for the Working Ontologist](#) (3<sup>rd</sup> Ed, 2020) – Hendler, Gandon, Allemang
- [Demystifying OWL for the Enterprise](#) – Uschold, Ding, Groth
- [Validating RDF Data](#) – Gayo, Prud'hommeaux, Boneva . *Comparison of SHEX and SHACL.*
- [Learning SPARQL](#) - DuCharme . *Learn RDF by querying the data.*
- [3D-force-graph](#) *Network graph visualizations in this presentation.*

*Network graph visualizations in this presentation.*

- [3D-force-graph](#)

Pam and Ray Exposure Data

<https://bit.ly/PamAndRayTTL>

<https://bit.ly/PamAndRay>

<https://bit.ly/RayDemogAndSHACL>

<https://bit.ly/RaySHACLResult>



## Appendix 2 : Applications and Examples

Examples of graph software and visualizations presented by Ellie Young and Paco Nathan in their talk: “The Case For Knowledge Graphs” @ *The Knowledge Graph Conference*

*Many more examples are available online.*



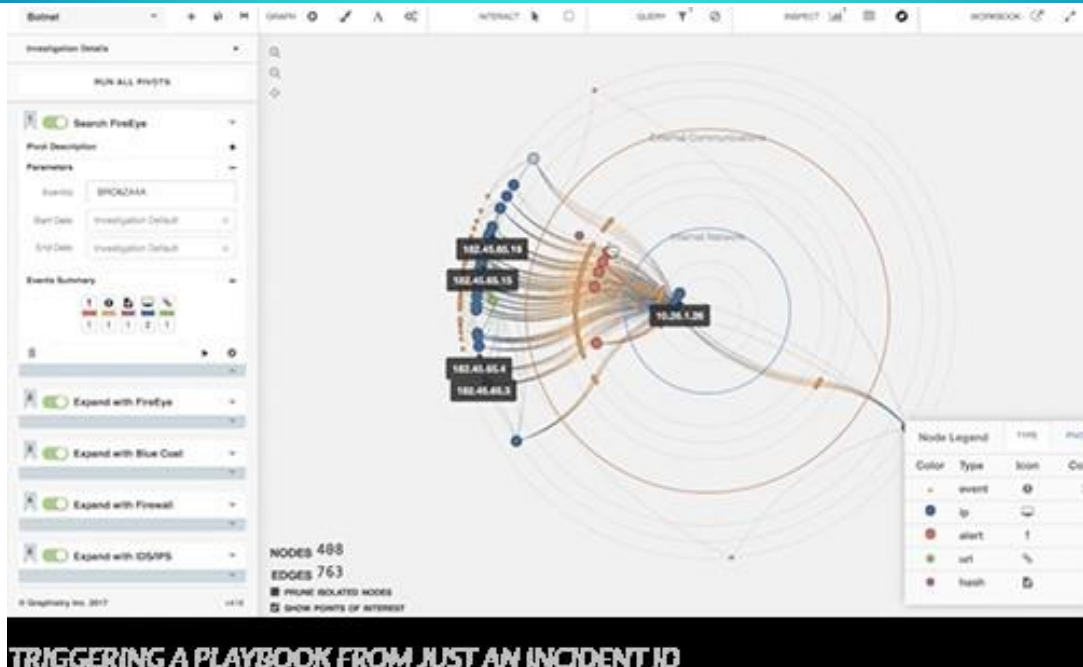
# Election 2020 Interference



- The number of Facebook and Twitter postings with misinformation intended to influence the election and countries from which the postings originated. Credit: [Atlantic Council](#).



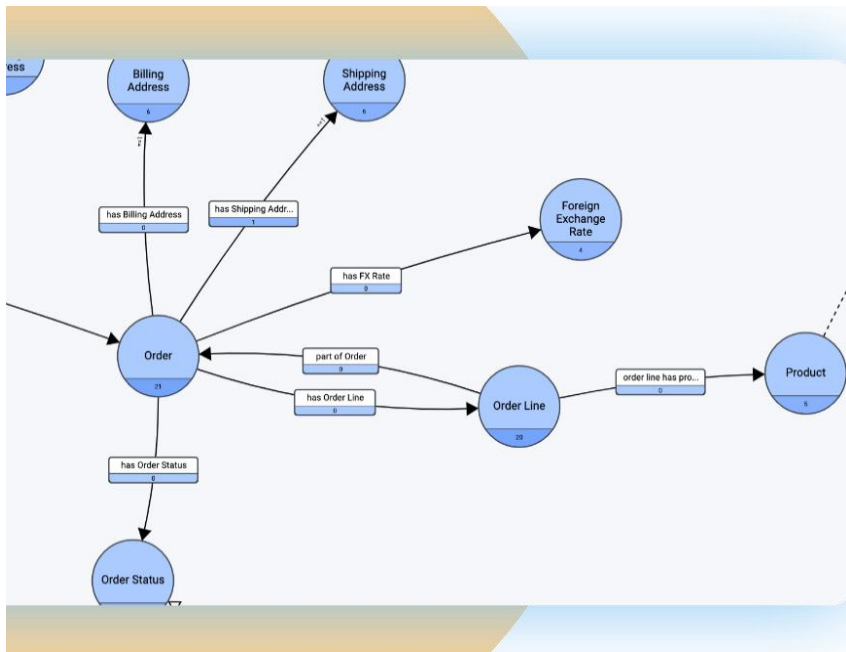
# Security and Incident Management



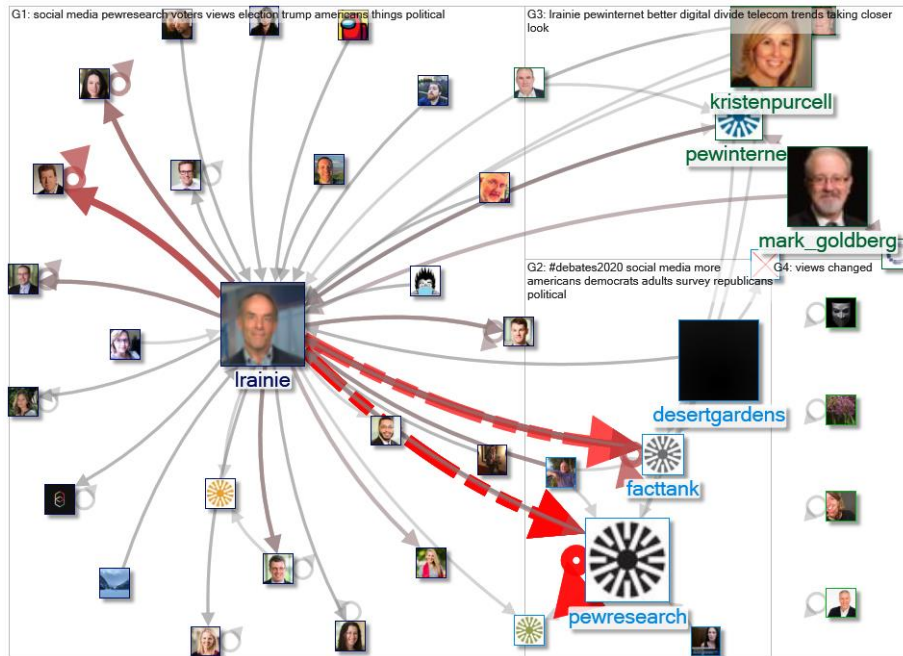
- [Graphistry](#)
- Security Incident and Event Management ([SIEM](#)) showing showing event progression, linkages, scope, root cause and more. Credit: Leo Meyerovich



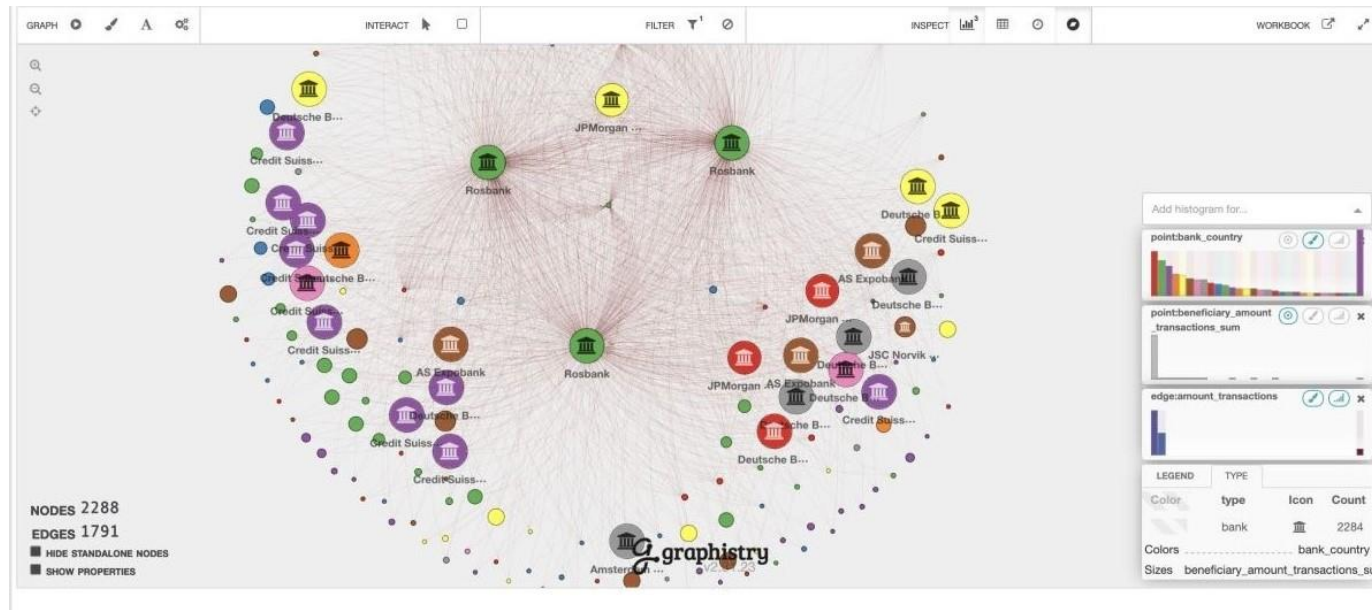
# Graph Creation & Management with Grafo



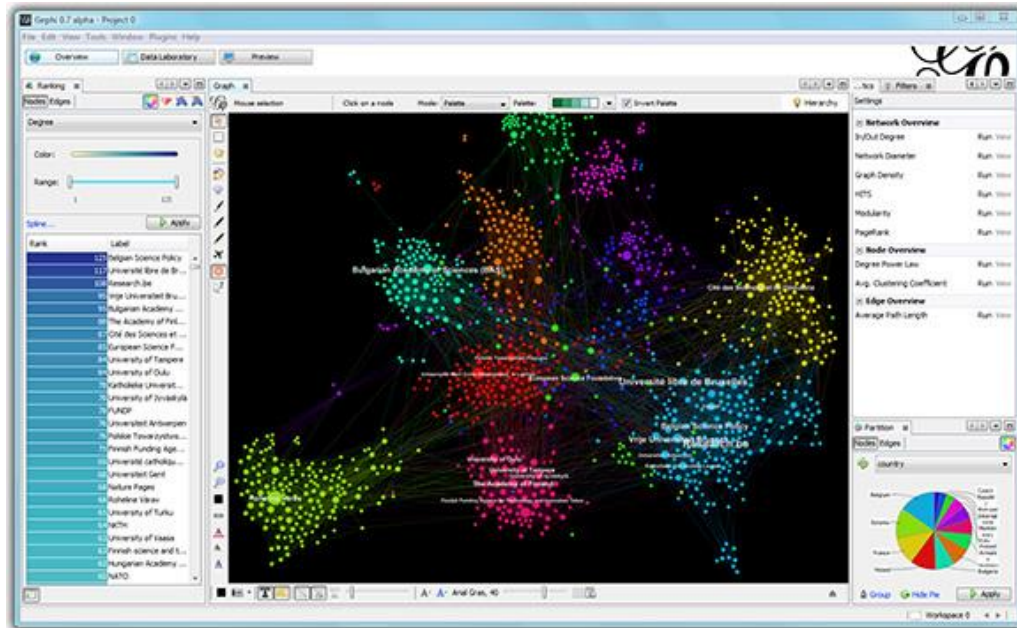
- [Grafo](#), a product that simplifies creation of knowledge graphs.
- Credit: Juan Sequeda



- A graph in [NodeXL](#) showing a network of 42 Twitter users whose tweets over a two-week period contained "Irainie", or who were replied to or mentioned in those tweets. Credit: Marc Smith



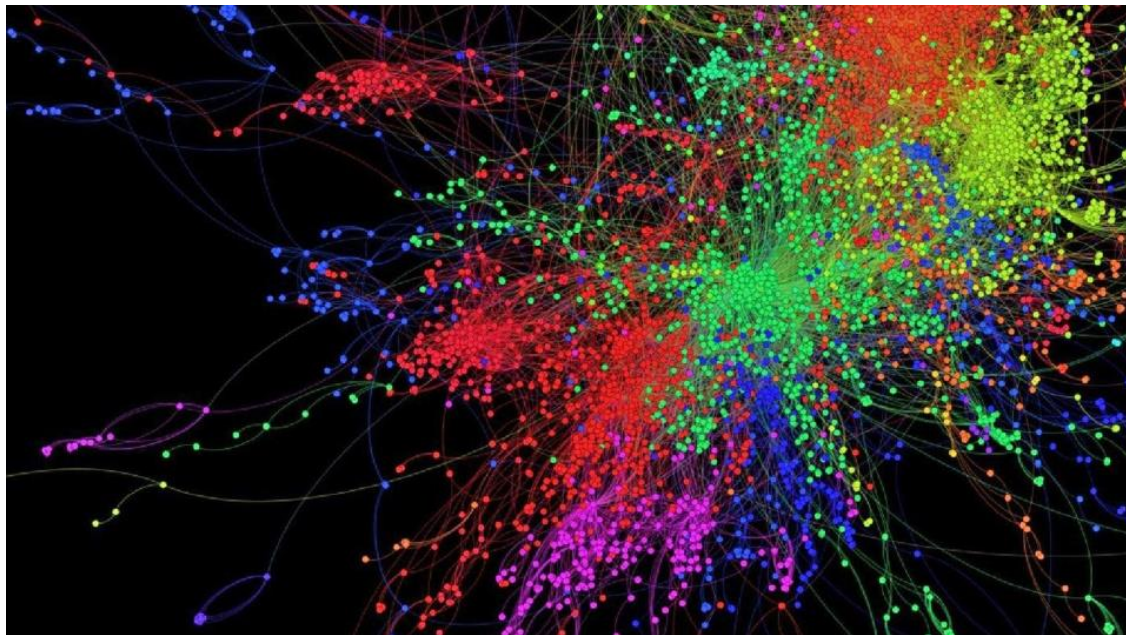
- A graph created using with Graphistry tracking Suspicious Activity Reports of transactions between banks for purposes of anti-money laundering (AML). Credit: Leo Meyerovich



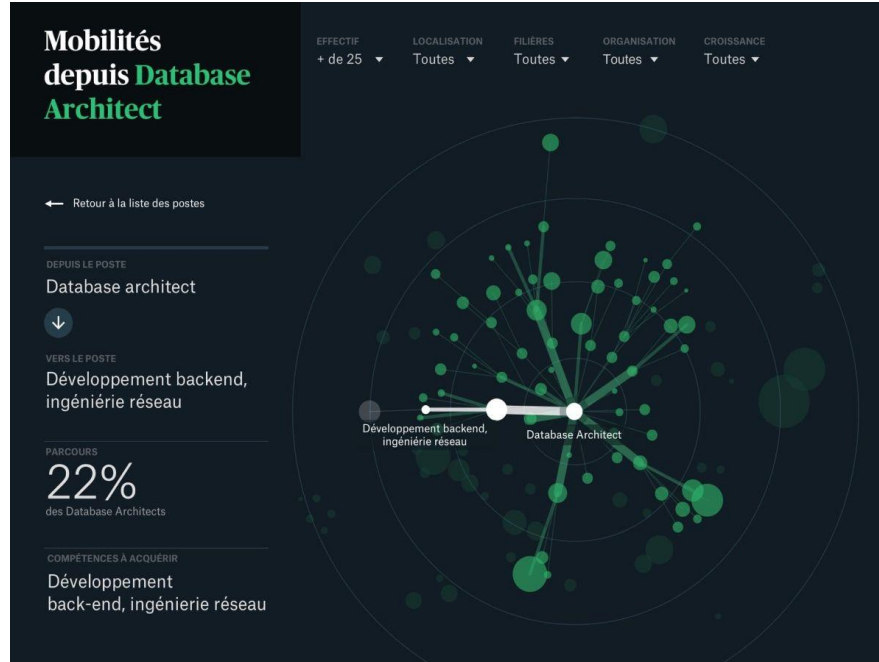
- Open source [Gephi](#) creates graphs to complement to traditional statistical analysis. This graph shows connections between various European universities. Credit: Mathieu Bastian.



# The Expanded Star Wars Universe



- Connections between the 7,563 main characters in the [Expanded Star Wars Universe](#). Credit: [Kirell Benzi](#)



- [Dataveyes](#)
- Visualizing the organization, its skill sets and career pathways. This image shows career options at a company and how it is possible to progress from one job into other. Credit: Caroline Goulard

# Appendix 3: Image Attribution



The Simpsons. Matt Groening , Fox Network.



The Lego Group



Archer. Adam Reed , FX Productions



Ōkiku naru Ko (Television program)



Dr. M. MacLuhan



Love and Rockets (band). Daniel Ash.



Napoleon Dynamite. Fox Searchlight Pictures