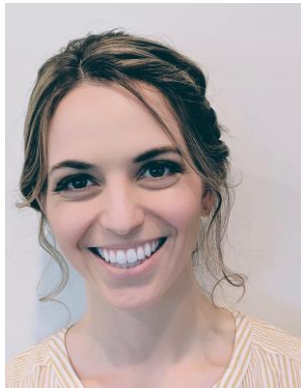
Knowledge Graphs: Changing How We Think About Data.

A Conversation with PHUSE, Stardog, and FDA DRSI

Presenters

Tim Williams

- Knowledge Graph Project Lead, PHUSE
- Lead Statistical Solutions Analyst, UCB Biosciences



Laura Firey

- Product Manager, Stardog

Outline

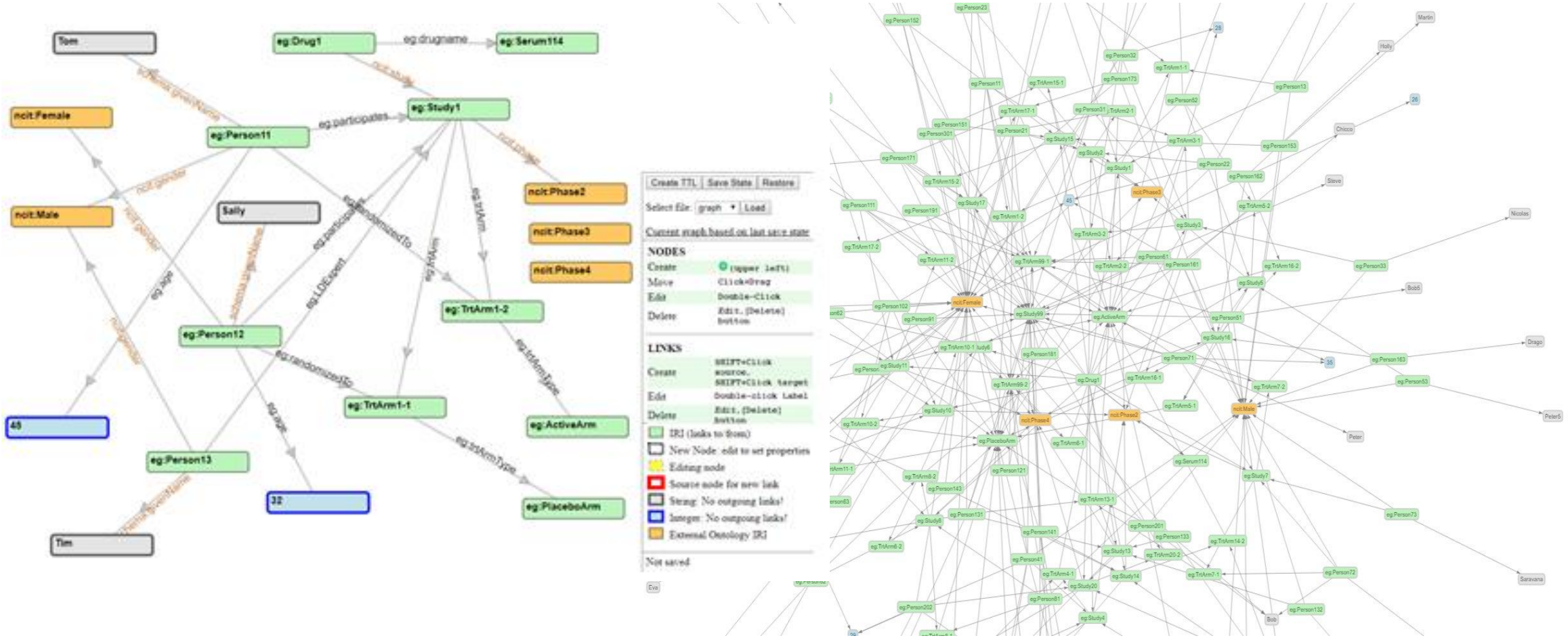
- Knowledge Graphs [~20 min]
- Cross Industry Perspective [~10min]
- Open Discussion [~30min]

Most important: “*A Conversation…*”

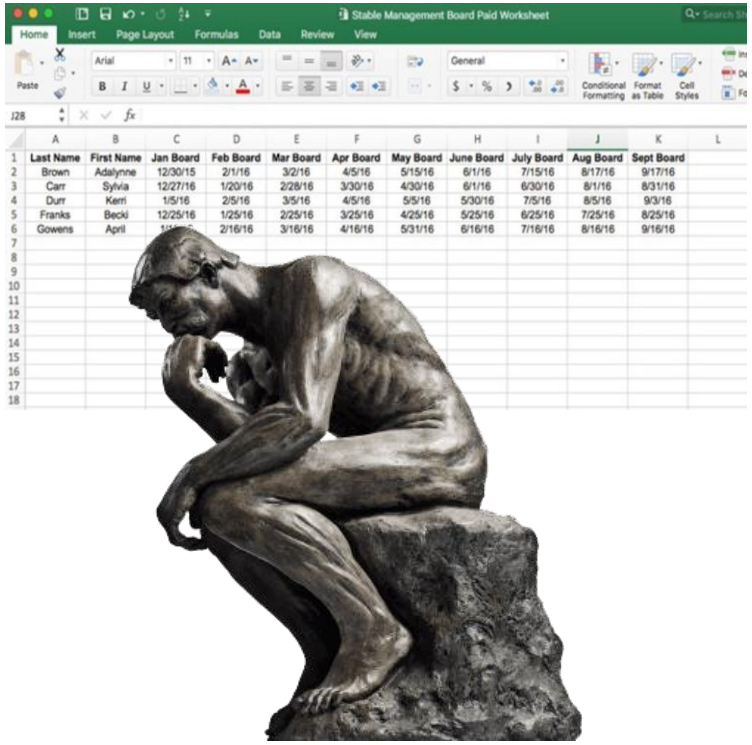
PHUSE Linked Data Projects

- CDISC Foundational Standards in RDF
- CDISC Conformance Checks
- Reusing Medical Summaries for Enabling Clinical Research
- Regulatory Guidance in RDF
- Clinical Program Design in RDF
- CDISC Protocol Representation Model in RDF
- **Analysis Results & Metadata**
  - *RDF Data Cubes for clinical trial results*
- **Clinical Trials Data as RDF**
  - *Study Data Tabulation Model as Linked Data*
- **Going Translational with Linked Data**
- **Study Data Validation and Submission Conformance**
  - Pre-clinical data + submission metadata

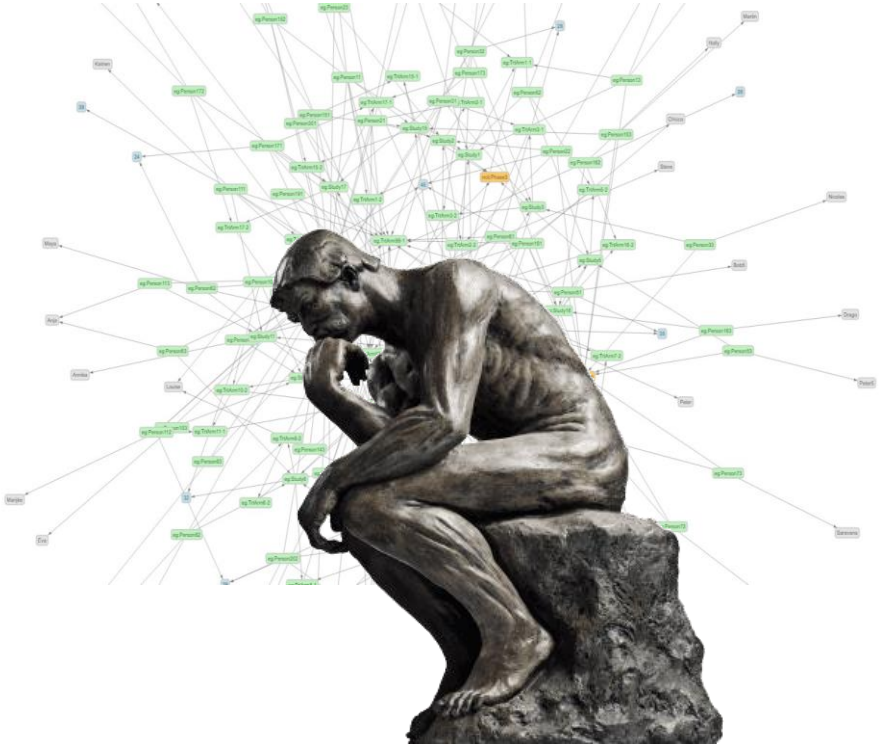
PHUSE Linked Data Workshop



Thinking About Data

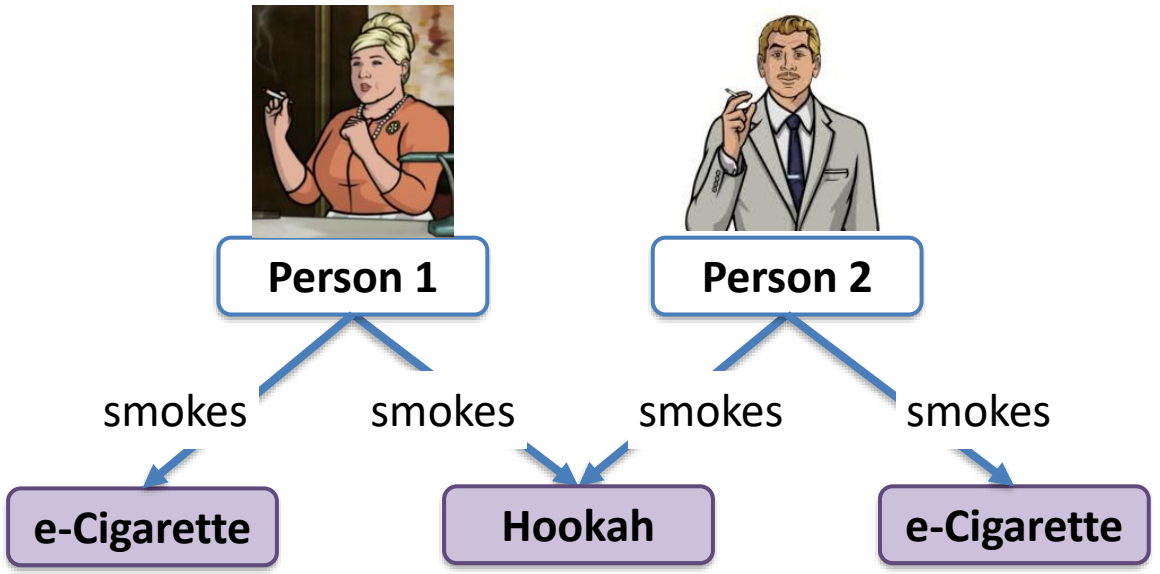


	A	B	C	D	E	F	G	H	I	J	K
1	Last Name	First Name	Jan Board	Feb Board	Mar Board	Apr Board	May Board	June Board	July Board	Aug Board	Sept Board
2	Brown	Adalynne	12/30/15	2/1/16	3/2/16	4/5/16	5/15/16	6/1/16	7/15/16	8/17/16	9/17/16
3	Carr	Sylvia	12/27/16	1/20/16	2/28/16	3/30/16	4/30/16	6/1/16	6/30/16	8/1/16	8/31/16
4	Durr	Kerri	1/5/16	2/5/16	3/5/16	4/5/16	5/5/16	5/30/16	7/5/16	8/5/16	9/3/16
5	Franks	Becki	12/25/16	1/25/16	2/25/16	3/25/16	4/25/16	5/25/16	6/25/16	7/25/16	8/25/16
6	Gowens	April		2/16/16	3/16/16	4/16/16	5/31/16	6/16/16	7/16/16	8/16/16	9/16/16
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											



Knowledge Graph Data-Centric Model

- Use-case neutral
- Real-world processes, entities, relationships



- **Knowledge Graph** as Resource Description Framework (RDF)

The whiteboard model you draw ***is*** the data.

Knowledge Graphs Facilitate Standards

- Shared Definitions and Understanding
  - What is a Tobacco Product?
  - Who is a Tobacco User?
- Coding of
  - Medical Conditions
  - Adverse Events
  - Products, Manufacturers
- Data Classification, Rules, Validation..
- Ontology Driven/Supported

Ontology-based Knowledge Graph Standards

What is an Ontology?

- **Dictionary**
  - **terms** and their definitions
- **Taxonomy**
  - **class** hierarchy
- **Thesaurus**
  - **relationships** between terms

A Data Engineer's Guide to Semantic Modelling
Ilaria Maresi (June 2020)
http://blog.thehyve.nl/news/ebook-semantic-model

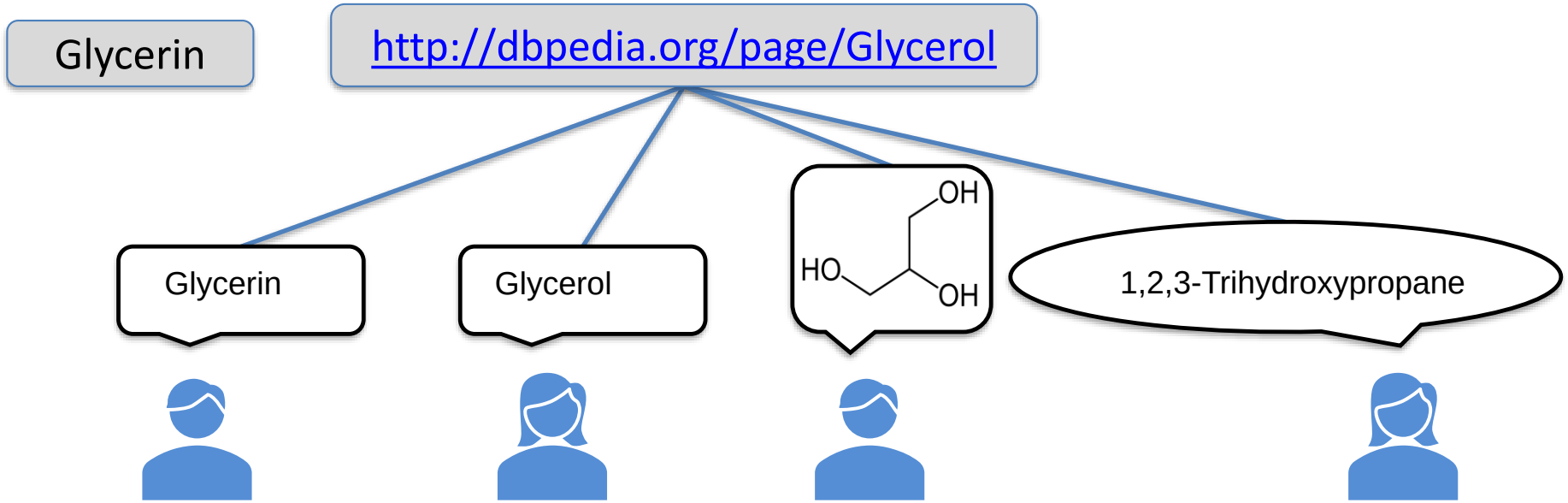
- Rules and Restrictions
  - Group membership, exclusions, types
  - Employ **reasoner**, ***infer*** values, relations

Standards: Identifier Examples

- Devices
- Products
- Manufacturers
- Ingredients

RDF

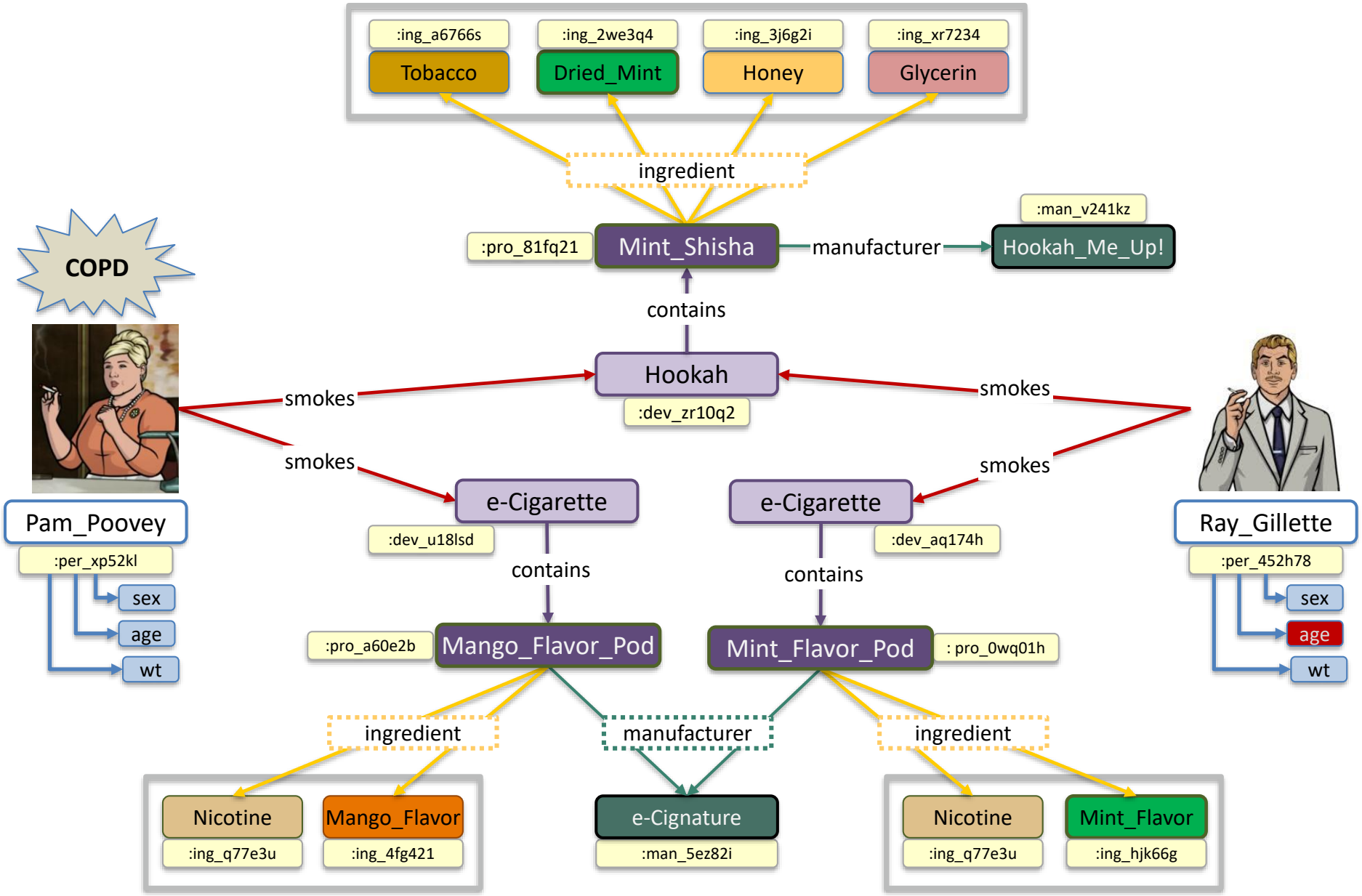
- Uniform Resource Identifier (URI)
- Internationalized Resource Identifier (IRI)



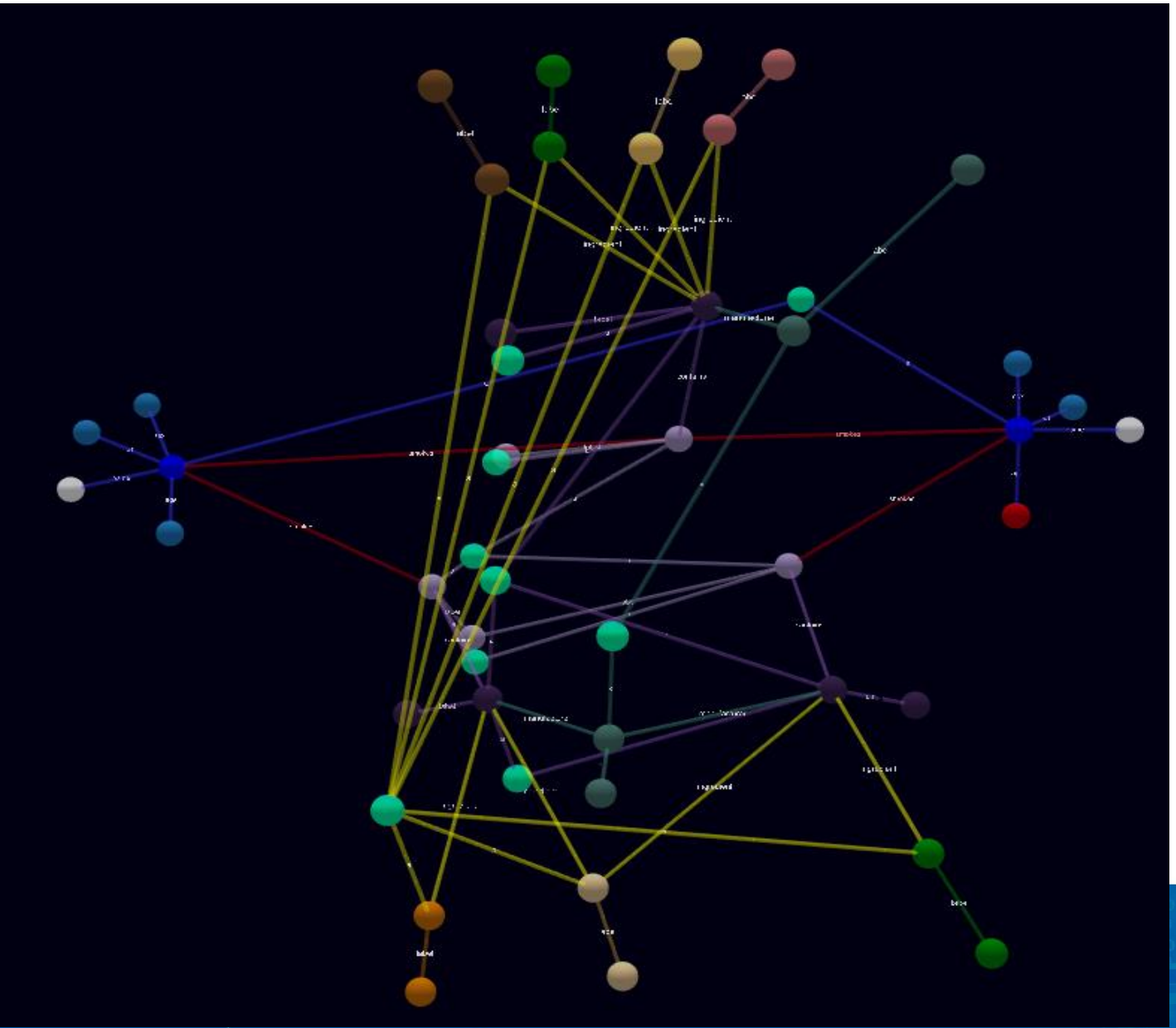
- Common terminology
- Link to other Knowledge Graphs

Knowledge Graph: Easy Answers to Complex Questions

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


:ing_a6766s
:ing_2we3q4
:ing_3j6g2i
:ing_xr7234
Tobacco
Dried_Mint
Honey
Glycerin
ingredient
:man_v241kz
:pro_81fq21
Mint_Shisha
manufacturer
Hookah_Me_Up!
COPD
contains
Hookah
:dev_zr10q2
smokes
smokes
smokes
smokes
e-Cigarette
e-Cigarette
:dev_u18lsd
:dev_aq174h
contains
contains
Pam_Poovey
:per_xp52kl
sex
age
wt
Ray_Gillette
:per_452h78
sex
age
wt
:pro_a60e2b
Mango_Flavor_Pod
Mint_Flavor_Pod
: pro_0wq01h
ingredient
manufacturer
ingredient
Nicotine
:ing_q77e3u
Mango_Flavor
:ing_4fg421
e-Cignature
:man_5ez82i
Nicotine
:ing_q77e3u
Mint_Flavor
:ing_hjk66g

Exposure Data



View the interactive visualization at:

https://bit.ly/PamAndRay

- Data error
- Demographics
- Device
- Entity Type
- Manufacturer
- Person
- Product

Ingredients

This seems complicated!

> "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

Data

```
:per_xp52kl :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_81fq21 .

:dev_aql74h skos:prefLabel "e-Cigarette"^^xsd:string ;
            a              :eCigarette ;
            a              :Device ;
            :contains      :pro_0wq01h .

:pro_81fq21 skos:prefLabel "Mint_Shisha"^^xsd:string ;
            a              :Product ;
            a              :TobaccoMix ;
            :manufacturer  :man_v241kz ;
            :ingredient    :ing_a6766s, :ing_2we3q4, :ing_3j6g2i, :ing_xr7234 .

:pro_a60e2b skos:prefLabel "Mango_Flavor_Pod"^^xsd:string ;
            a              :Product ;
            a              :FlavorPod ;
```

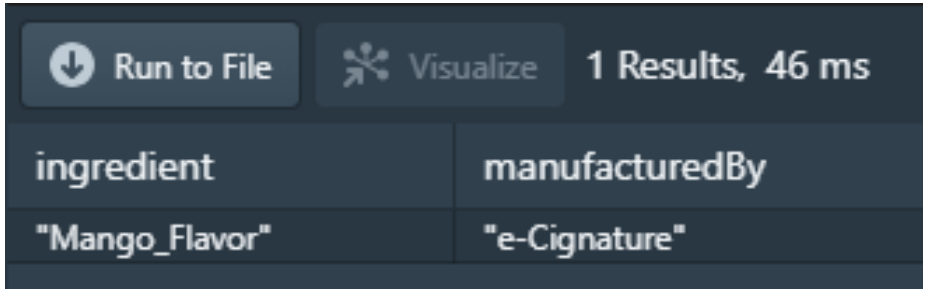
View the data file at:

https://bit.ly/PamAndRayTTL

Pam's Unique Exposure Ingredient?

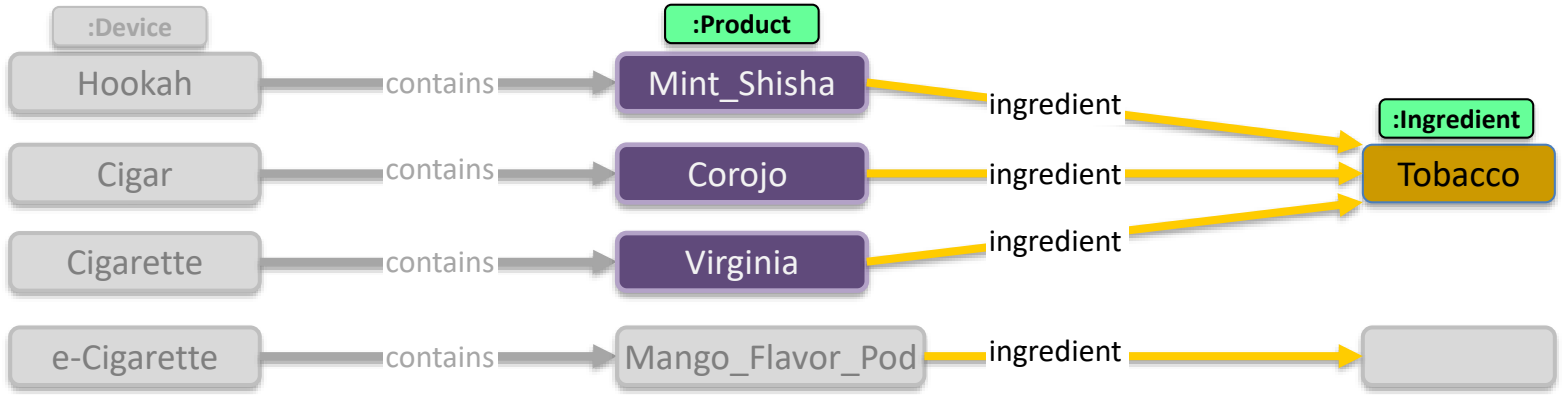
```
PREFIX  :     <http://example.org/Eg#>
PREFIX skos:  <http://www.w3.org/2004/02/skos/core#>

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))
}
```

PamUniqueIngred.rq

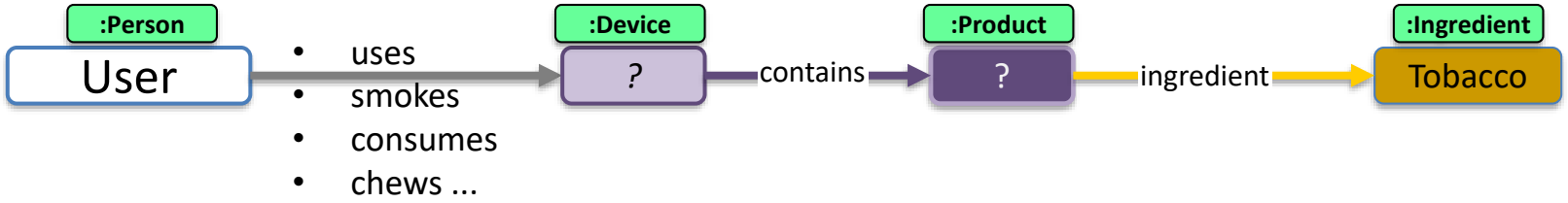


Ontology Revisited

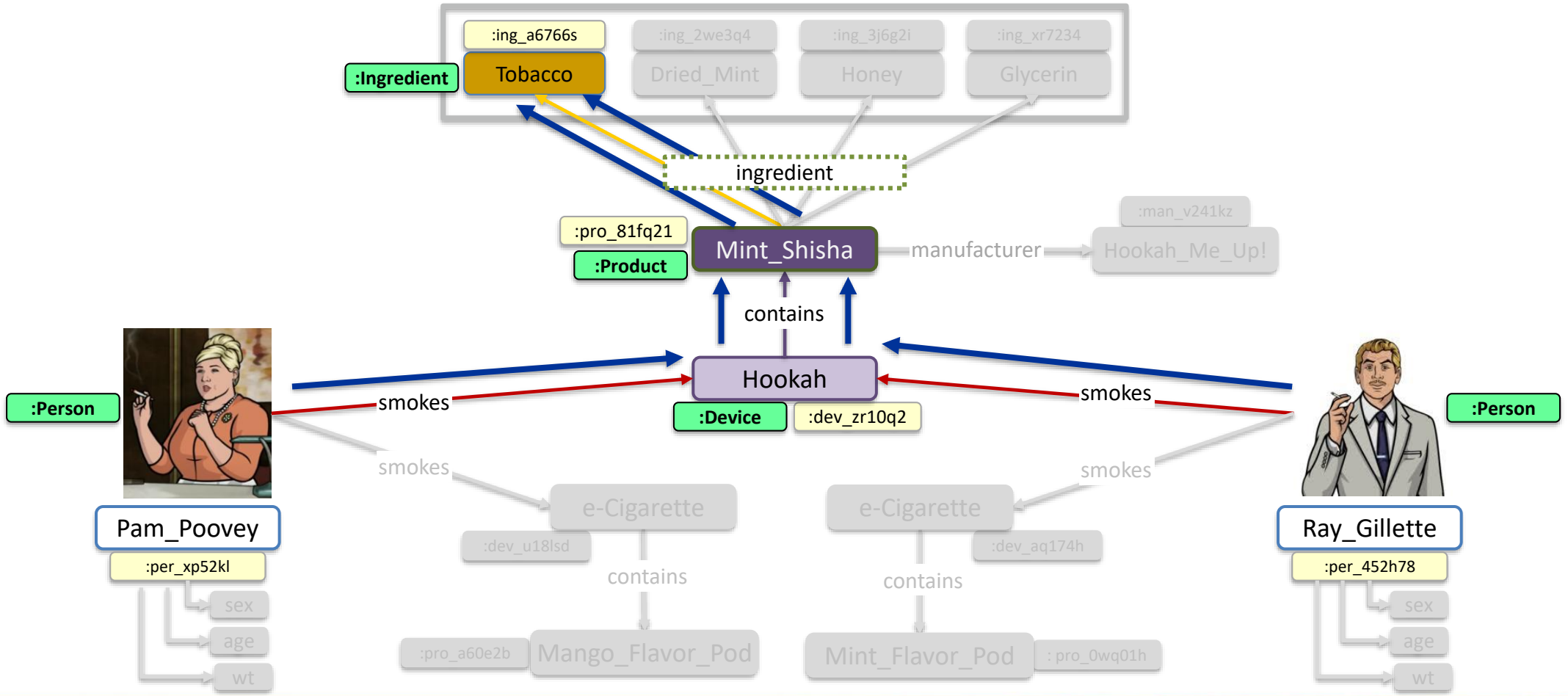
- What is a **TobaccoProduct**?



- Who is a **TobaccoUser** ?



Tobacco Smoker: The path between Person and Ingredient



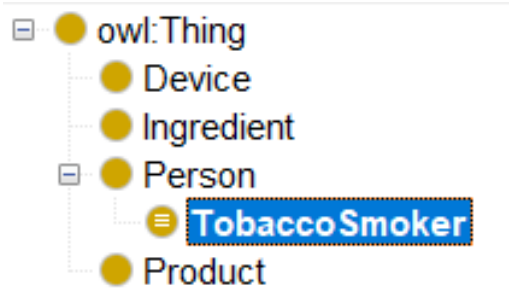
Query: Tobacco Users

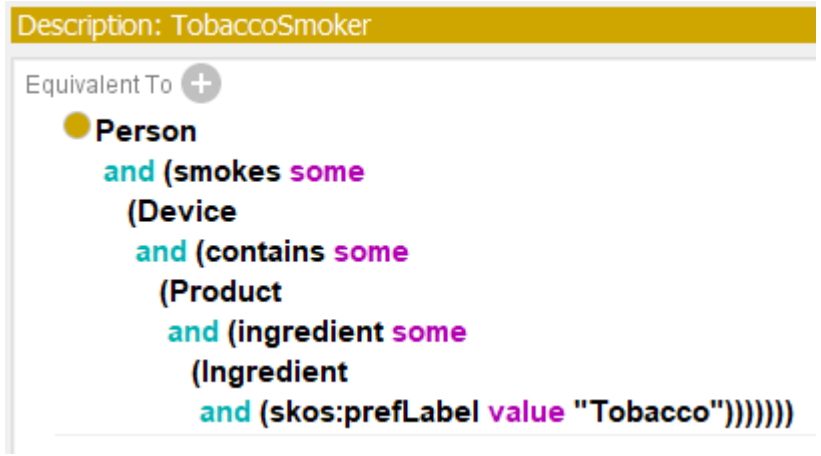
TobaccoUsers.rq

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

SELECT ?name ?ingredient

WHERE{
 ?pers      :name          ?name ;
            :smokes        ?device .
  ?device   :contains      ?product .
  ?product  :ingredient    ?ingred .
  ?ingred   skos:prefLabel ?ingredient .
  FILTER (REGEX(?ingredient, "Tobacco"))
}
```

name	ingredient
"Pam_Poovey"	"Tobacco"
"Ray_Gillette"	"Tobacco"

Tobacco Smoker: Ontology Definition & Query





TobaccoSmoker-Infer.rq

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

SELECT ?tobaccoSmoker
WHERE{
    ?person a       :TobaccoSmoker ;
            :name ?tobaccoSmoker .
}
```

tobaccoSmoker
"Ray_Gillette"
"Pam_Poovey"

Data Validation

Validating Ray's Demographics data

Data Has Shape. Validation has Shape.

SHApes **C**onstraint **L**anguage (SHACL)

“Person Shape”
(Validation Constraints)

Person Data

Validating Ray's Demographics

Ray Gillette

Demographics

sex: M

age: 43

wt: 165

Data Entry

M

403

165

SHACL

Person Shape

M/F/U

SEX

> 0
< 120

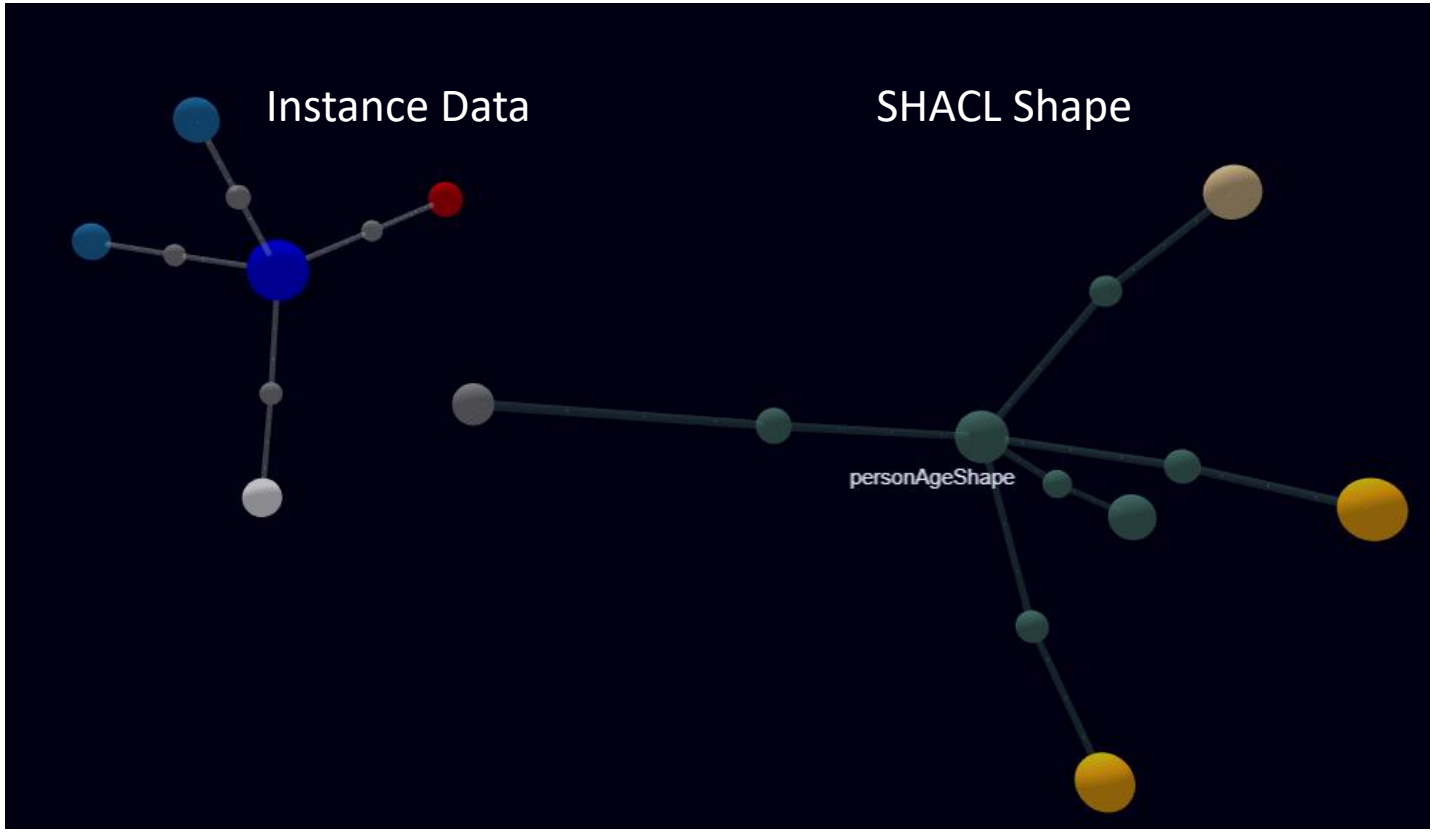
AGE

Violation!

> 0
< 800

WEIGHT

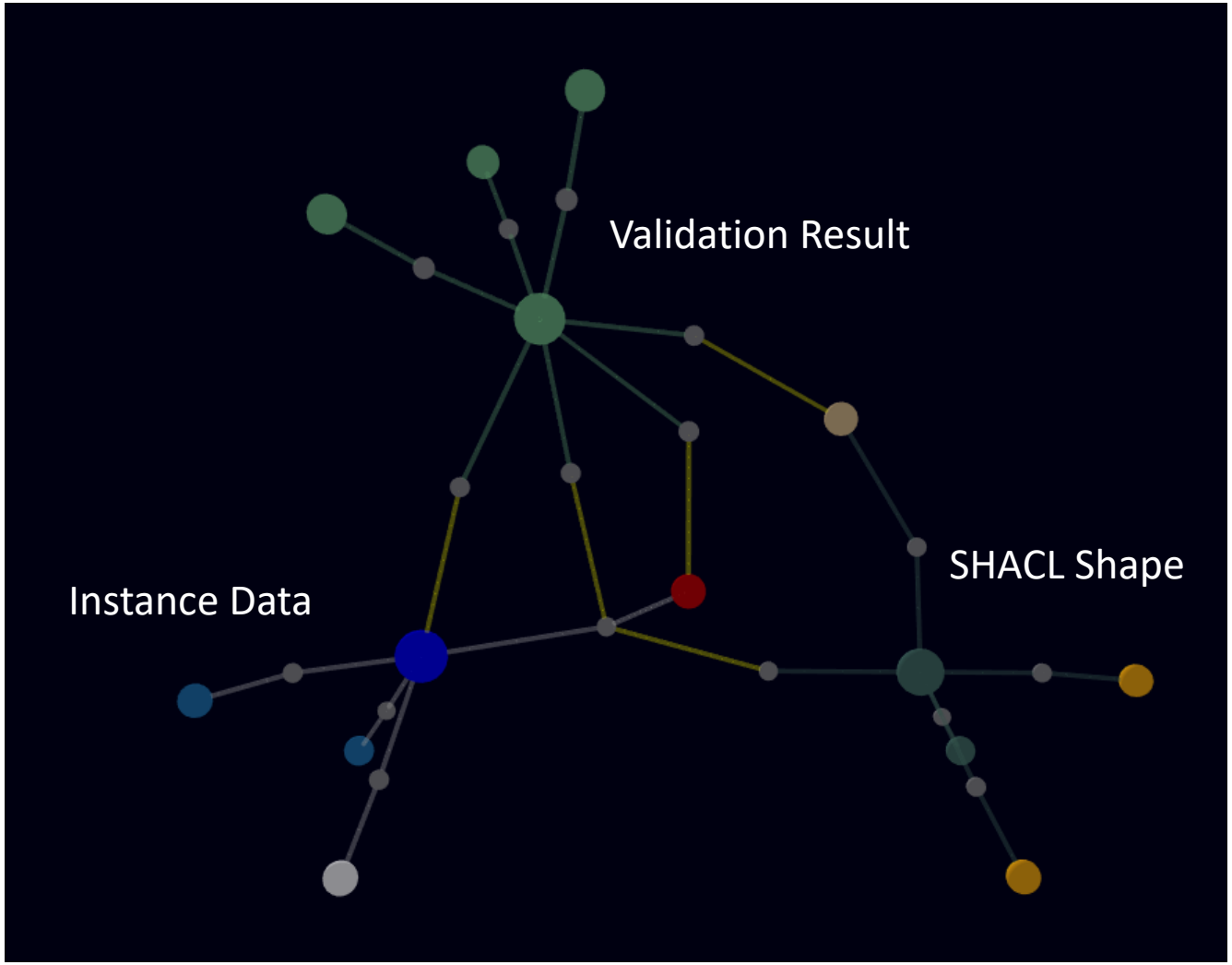
Validating Age with SHACL



View the interactive visualization at:

https://bit.ly/RayDemogAndSHACL

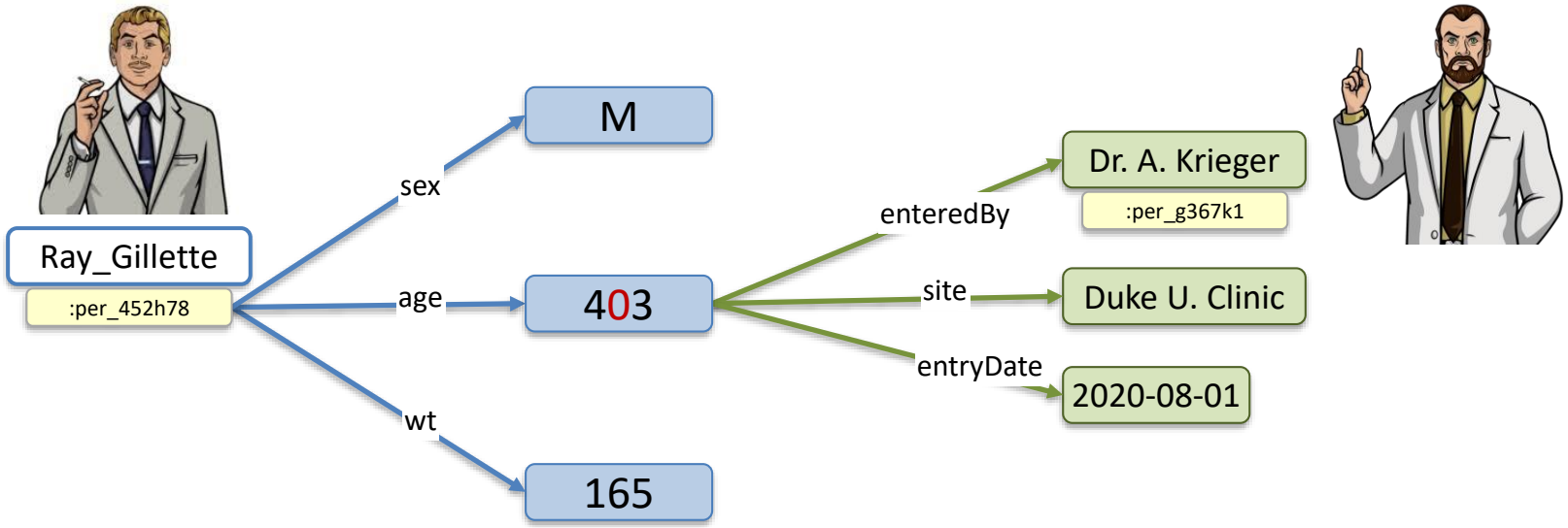
Demographics, Constraints, Report



View the interactive visualization at:

https://bit.ly/RaySHACLResult

Metadata is Part of the Graph



FAIR Data

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

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

- FAIR Implementation Project & Toolkit
  https://www.pistoiaalliance.org/projects/current-projects/fair-implementation/
- Ontologies Mapping
  https://www.pistoiaalliance.org/projects/current-projects/ontologies-mapping/

FAIR Data *is* Linked Data *is a* Knowledge Graph

Knowledge Graphs :

Now for a cross-industry perspective

Additional Reading

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 (3rd 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.*

https://bit.ly/PamAndRay
https://bit.ly/RayDemogAndSHACL
https://bit.ly/RaySHACLResult

Pam and Ray Exposure Data
https://bit.ly/PamAndRayTTL

Reference Slides

The Roofshot / Moonshot Manifesto

Moonshot

Invent and apply state-of-the-art

- **Enterprise and Industry Knowledge Graphs**
- **Across the Data Life Cycle**

Roofshot

Incremental impacts in production

Examples

1. Unique Identifiers
2. Validation Rules in SHACL
3. Open Ontology Development for your domain

Concept & Image Attribution: https://rework.withgoogle.com/blog/the-roofshot-manifesto/

The Stairway to the Stars Manifesto