

TT01: Building a F.A.I.R. Foundation for Pharma

The PhUSE Project:

“Going Translational with Linked Data (GoTWLD)”

Co-Leads: Tim Williams (UCB)

Dr. Armando Oliva (Semantica LLC.)



**Working
Groups**

phuse.eu

What is F.A.I.R.?

- Many good resources available
 - This talk is not is not one
 - **go-fair.org**
 - “The FAIR Guiding Principles for scientific data management and stewardship”
<https://www.nature.com/articles/sdata201618>

FAIR Data Principles

Findability

- F1. globally unique, persistent id
- F2. rich metadata
- F3. searchable source
- F4. metadata specify data id

Accessibility

- A1. retrievable by id using standard protocol
 - A1.1 protocol open, free, universal
 - A1.2 protocol allows authentication
- A2 metadata available when data is not

Interoperability

- I1. formal, accessible, shared, broadly applicable language
- I2. uses FAIR vocabularies
- I3. qualified references to other data

Reusability

- R1. plurality of accurate and relevant attributes
 - R1.1 clear and accessible usage license
 - R1.2 provenance
 - R1.3 meets domain-relevant standards

How FAIR is Pharma?



phuse.eu

**Working
Groups**



Magic mirror on the wall...

**is
Pharma
F.A.I.R.**

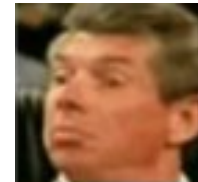
at all??



Introducing...

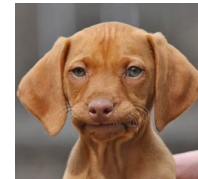
Mr. Tim's highly unscientific,
entirely biased...

Vince McMahon



/

Disappointed Dachshund



FAIRness Evaluation Scale

For an Actual Evaluation:

“Evaluating FAIR maturity through a scalable, automated, community-governed framework.”

- Nature, 20 Sept. 2019

<http://bit.ly/FAIRnessEval>

Linked Data = Foundation for FAIR

- 5-Star, Linked Open Data
 - Original vision for the Internet

5 Star Open Data Principles

Available



OL



phuse.eu

Working Groups

How Does Pharma Fare on FAIR?

Findability

- F1. globally unique, persistent id
- F2. rich metadata
- F3. searchable source
- F4. metadata specify data id

Accessibility

- A1. retrievable by id using standard protocol
 - A1.1 protocol open, free, universal
 - A1.2 protocol allows authentication
- A2 metadata available when data is not

Interoperability

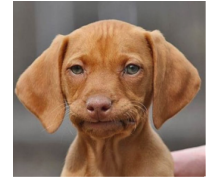
- I1. formal, accessible, shared, broadly applicable language
- I2. uses FAIR vocabularies
- I3. qualified references to other data

Reusability

- R1. plurality of accurate and relevant attributes
 - R1.1 clear and accessible usage license
 - R1.2 provenance
 - R1.3 meets domain-relevant standards

Findability

- F1. globally unique, persistent id



Human Study Subject

- PharmaCo
 - Study 1, Drug A
 - Study 2, Drug B
- DrugCo
 - Study 3, Drug b



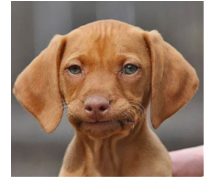
PhUSE Project: “Study URI”

Poster: “Linked Data URIs”

Merge Billy-Bob’s data from all studies.

Accessibility

- A2. Metadata available when data is not

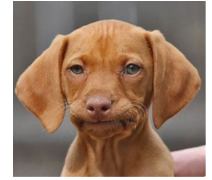


Data changes form during its journey from collection to regulatory submission.

Biostatisticians and Medical Writers do not have easy access to the metadata from data collection and transformation processes.

Interoperability

- I1. shared language for knowledge



Pharma Lacks:

- *Knowledge representation*
- *Language adoption*

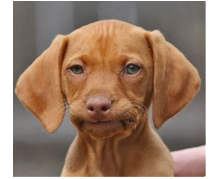
Linked Data / Knowledge Graph adoption is helping!

Core Challenge : Change of mindset

- **Currently:** Data modeled to industry standards for submission to regulatory authorities
- **Future:** Models of the process and the entities in the data.
- My ontology versus your ontology

Reusability

- R1.3 data meet domain-relevant community standards



Positive:

- We have standards!

Negative:

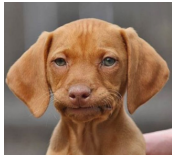
- We have standards...
 - Historically are row x column structure
 - Recent initiatives are not 5 Star Open / FAIR

Evaluate

Available



OL



API behind membership paywall

Structured



OL RE



JSON

Open



OL RE OF

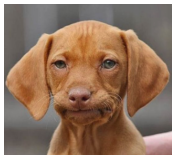


JSON. XML planned

URIs



OL RE OF URI

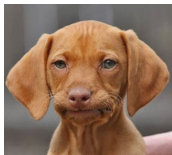


JSON, No RDF download (not even JSON-LD)

Linked



OL RE OF URI LD



Not possible with above restrictions...

A call for FAIRness

- Open Access (no paywall)
- Downloads as JSON-LD & RDF
- Supporting ontology, terminology, guidance as RDF



New membership model?

- Access *implementation expertise*
- Members influence model decisions?
- New opportunities?

Toward FAIR Pharma with PhUSE

Going Translational with Linked Data (GoTWLD)

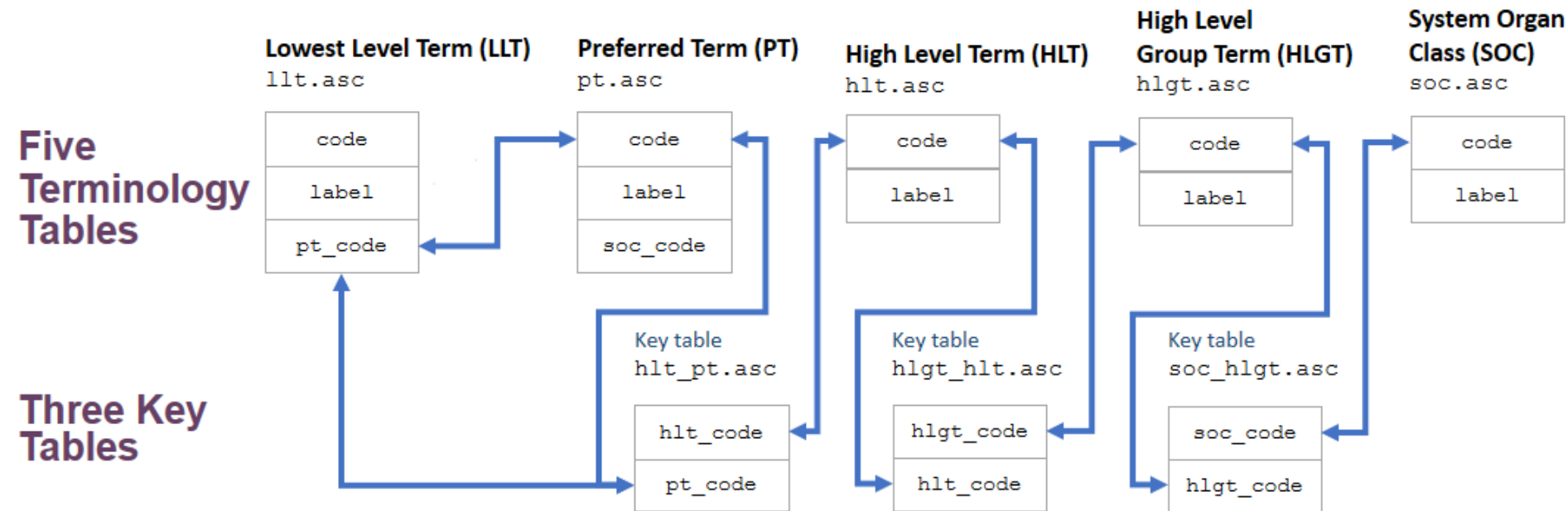
Three Sub-Projects:

1. MedDRA as RDF
2. Unique Identifiers for Pharma
3. SEND Conformance

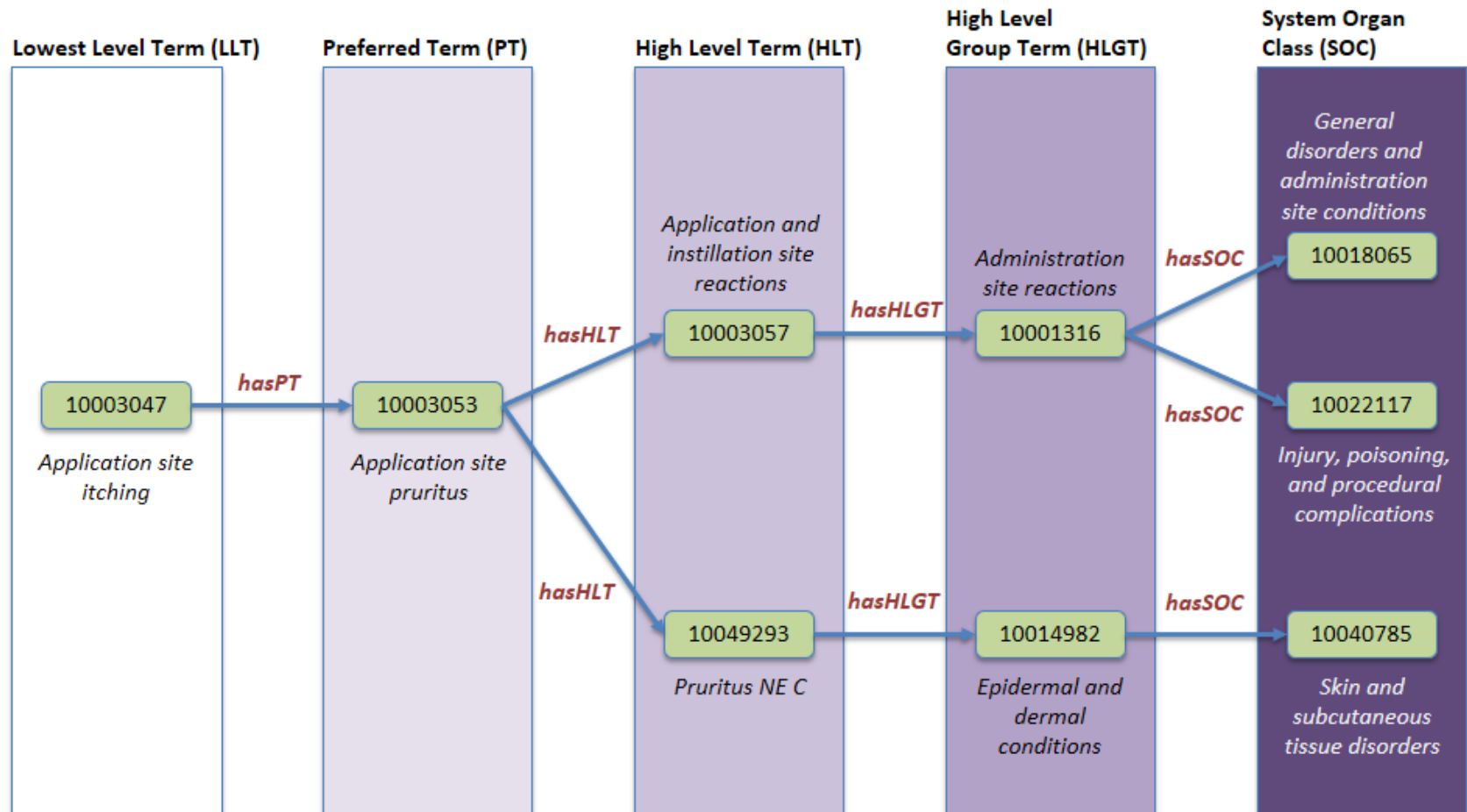


**Working
Groups**

1. MedDRA as RDF



1. MedDRA as RDF



1. MedDRA as RDF

GitHub: <https://github.com/phuse-org/MedDRAasRDF/>

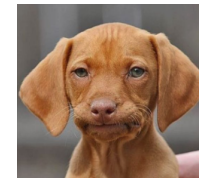
Ontology and conversion code available for download

Advantages

- Single File (no merging needed)
- Explicit, semantic relationships between terms
- Can include documentation and other metadata



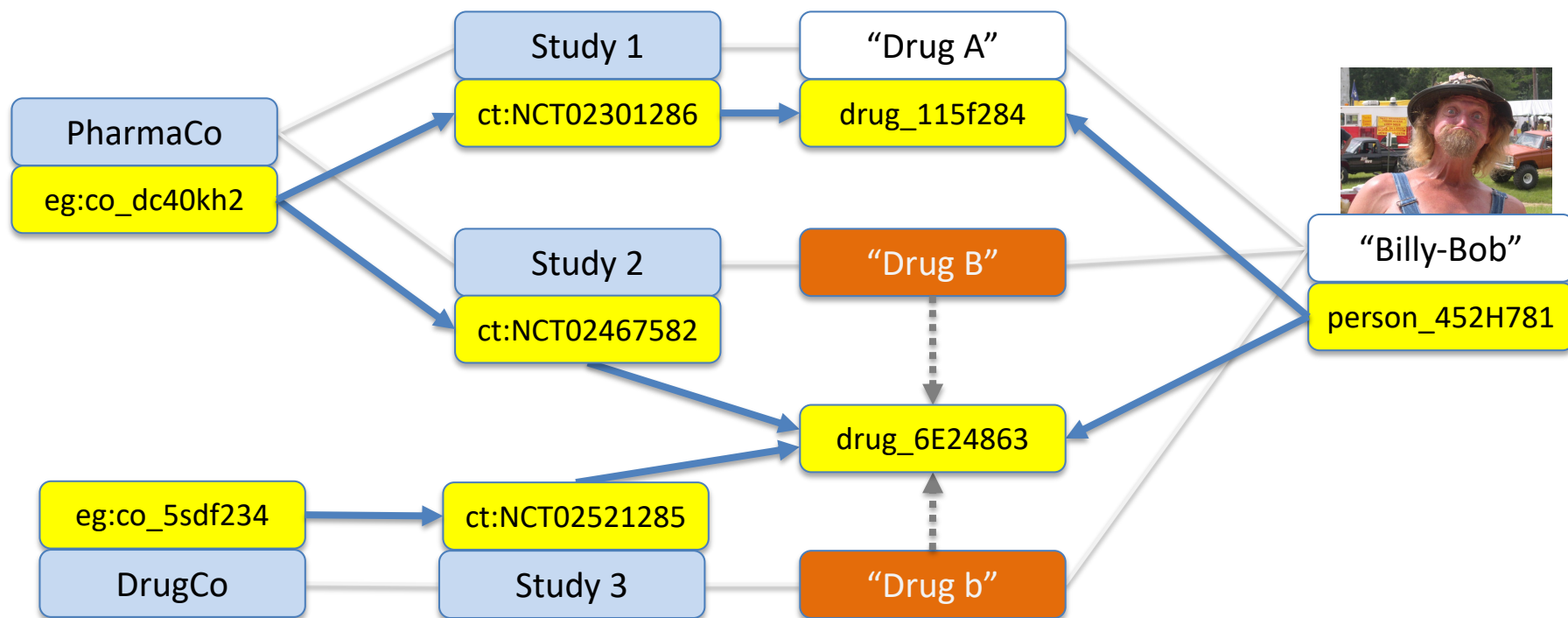
*MedDRA terminology files only
available under license, no RDF*



2. Unique Identifiers for Pharma

GitHub: <https://github.com/phuse-org/UIDPharma>

F1. globally unique, persistent id



Poster : Linked Data URIs

PP16: Linked Data URIs — Kerstin Forsberg

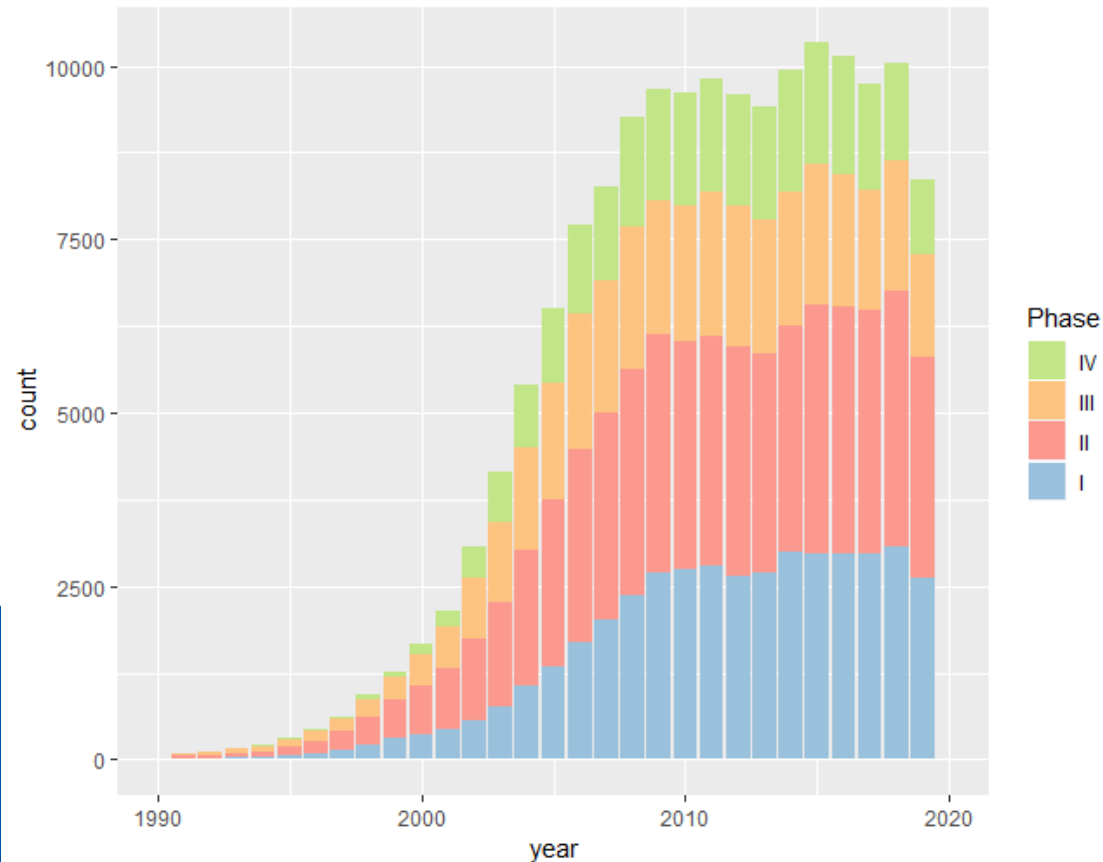
(AstraZeneca), Johannes Ulander (S-Cubed), Tim Williams (UCB BioSciences)

Discussion:

- How to construct Unique Identifiers?
- Who owns/curates them?
- How to encourage adoption?



Studies in Wikidata



Breaking: Decentralized Identifiers

Decentralized Identifiers (DIDs) v1.0

Core Data Model and Syntaxes

W3C First Public Working Draft 07 November 2019

<https://www.w3.org/TR/2019/WD-did-core-20191107/>

- Removes requirement for centralized repository and authority

3. SEND Conformance

SHApes Constraint Language (SHACL)

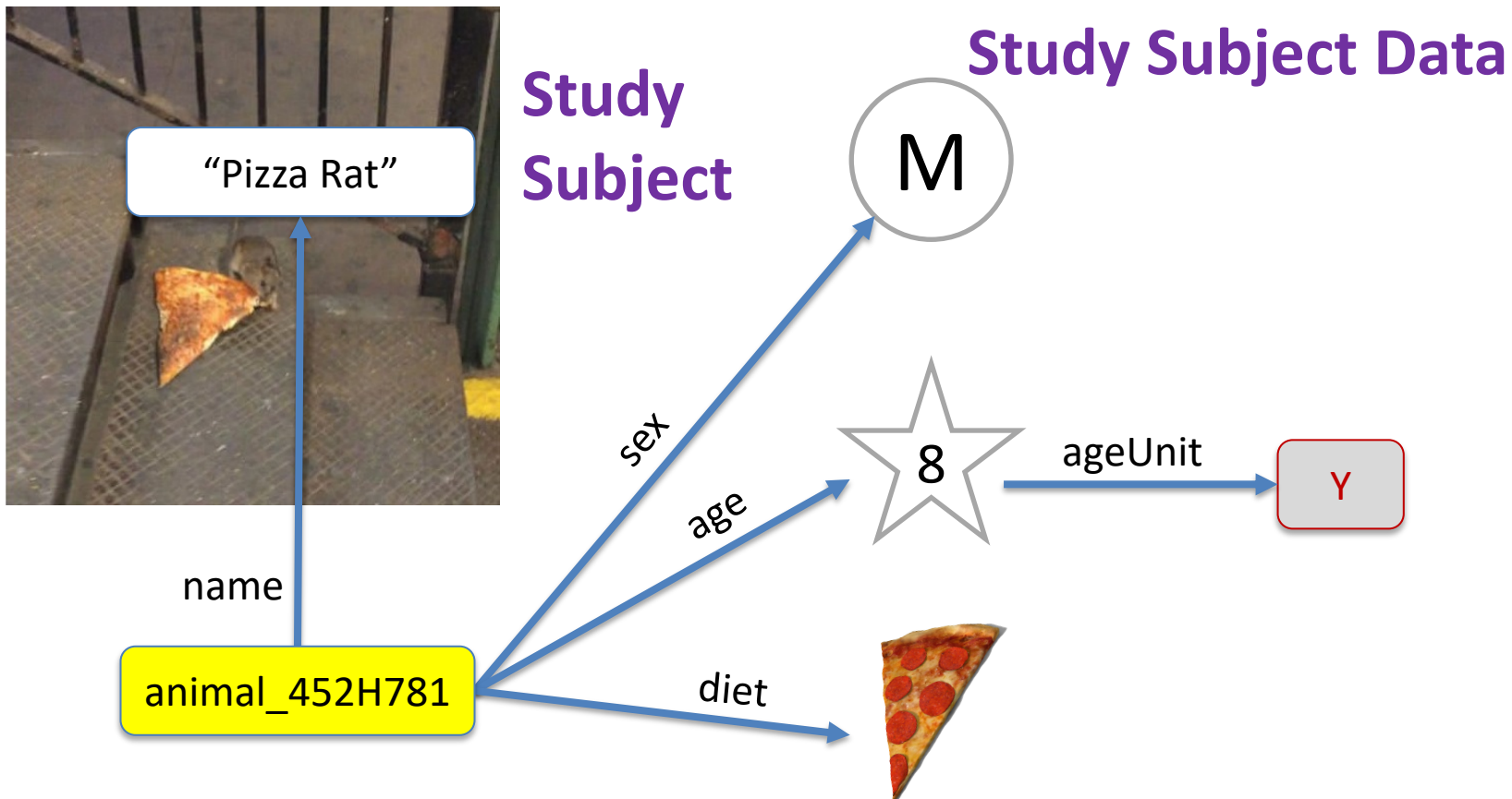


Study Subject Shape
(Validation Constraints)



Study Subject *Data*

3. SEND Conformance



3. SEND Conformance

"Pizza Rat"

Study Subject
Data

M

8

Y

SHACL

Study Subject Shape

M/F/U

SEX

8

AGE

AGEU

Wk

Violation!

Solid Chow CRF-1

DIET

Violation!

3. SEND Conformance

- GitHub :

<https://github.com/phuse-org/SENDConform/>

<https://phuse-org.github.io/SENDConform/>

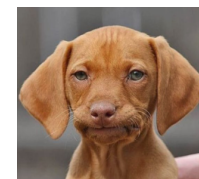
SHACL and Ontology available for download

Advantages

- Data validation becomes integral to the data
- Semantic relationships between data and checks



Getting started, much work to do!



Working
Groups

PhUSE US Connect 2020

“Rules in my Data! Data in my Rules! A Knowledge Graph Approach to SEND Conformance.”



Working
Groups

Related Papers at the this Conference

TT04: Deconstructing SDTM – Starting Again

TT06: Linked Data: Introduction and Deep Dive

ML02: Sub-population Detection Using Graph-based Machine Learning

DS01 : How Can We Ensure our Study Data is F.A.I.R?

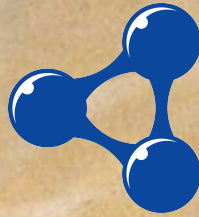
LI05: Statistical Programming in the Future...”
[FAIRness]



**Working
Groups**

Magic Mirror:

Pharma
is getting
F.A.I.R-er



after all.

We Need Your Help!

Join the PhUSE Emerging Trends & Technology Group!

Your knowledge of:

- MedDRA
- SDTM & ADaM
- SEND Data
- Clinical Trials data & process
- Data Validation & Submission Process
- R, Python
- Linked Data (RDF), SPARQL, OWL ...

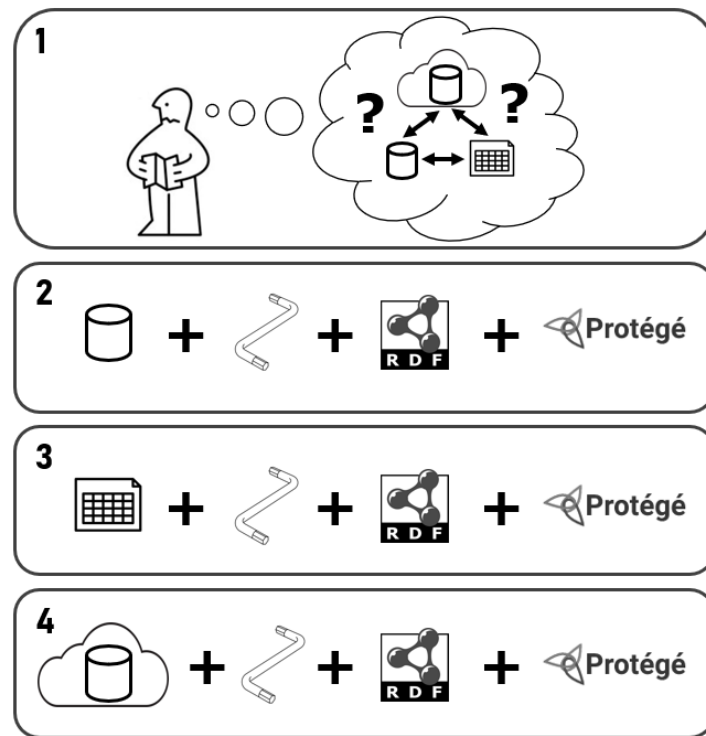
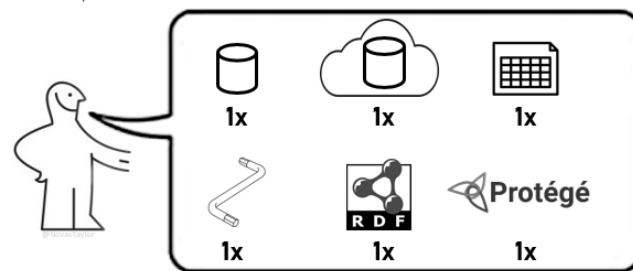
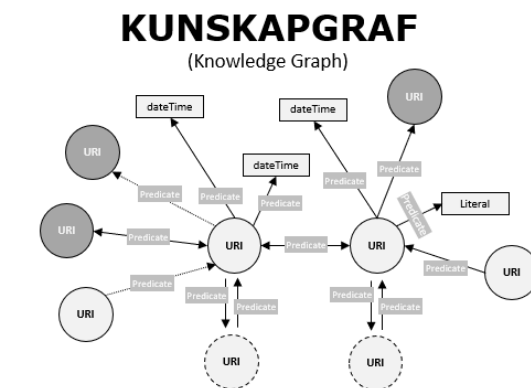


**Working
Groups**

Workshop: “Let’s Make a Knowledge Graph!”

*Observers
welcome!*

Tuesday
13:30–15:30



Thank you!

Email: tim.williams@phuse.eu

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

Twitter : @NovasTaylor