

# DISRUPTING CLINICAL TRIALS WITH LINKED DATA

PHUSE SINGLE DAY EVENT

Durham, NC  
9 November, 2017

# CONTENT BY

Tim Williams

CTDasRDF Project Co-Lead  
UCB BioSciences  
[tim.williams@ucb.com](mailto:tim.williams@ucb.com)

Armando Oliva

CTDasRDF Project Co-Lead  
Semantica LLC  
[aoliva@semanticallc.com](mailto:aoliva@semanticallc.com)

*Views are those of the authors, not their employers, not PhUSE...*

*"Oh dear! What is he going to say this year?"*

LAST YEAR: "THE ANGRY CANADIAN" 

"Any fool can criticize, condemn and complain and most fools do."  
- Benjamin Franklin

"I criticize by creation, not by finding fault."  
- Marcus Tullius Cicero

Not criticism: *opportunitites for improvement.*

# SDTM CAN BE IMPROVED

The image shows two screenshots of SDTM datasets. The top screenshot, titled 'vs', displays a table with columns: studyid (Study Identifier), domain, usubjid (Unique Subject Identifier), and vsseq. It contains 5 rows of data for study CDISCPIL01, domain VS, and subject 01-701-1015. The bottom screenshot, titled 'dm', displays a table with columns: studyid (Study Identifier), domain, usubjid, and subjid (Subject). It contains 7 rows of data for study CDISCPIL01, domain DM, and various subjects.

|   | studyid<br>Study Identifier | domain | usubjid<br>Unique Subject Identifier | vsseq |
|---|-----------------------------|--------|--------------------------------------|-------|
| 1 | CDISCPIL01                  | VS     | 01-701-1015                          | 1     |
| 2 | CDISCPIL01                  | VS     | 01-701-1015                          | 2     |
| 3 | CDISCPIL01                  | VS     | 01-701-1015                          | 3     |
| 4 | CDISCPIL01                  | VS     | 01-701-1015                          | 4     |
| 5 | CDISCPIL01                  | VS     | 01-701-1015                          | 5     |

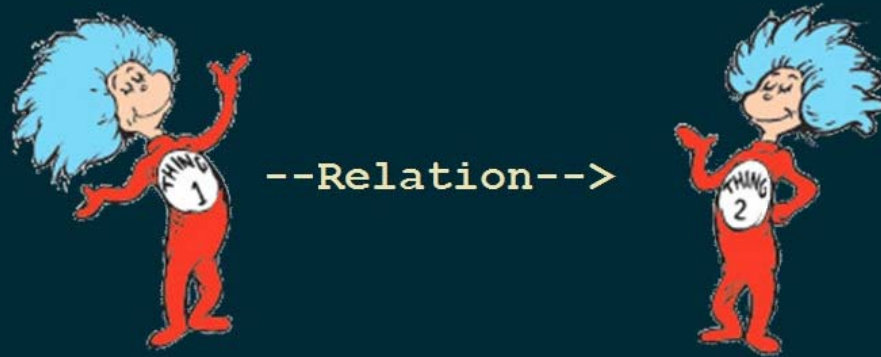
|   | studyid<br>Study Identifier | domain | usubjid     | subjid<br>Subject |
|---|-----------------------------|--------|-------------|-------------------|
| 1 | CDISCPIL01                  | DM     | 01-701-1015 | 1015              |
| 2 | CDISCPIL01                  | DM     | 01-701-1023 | 1023              |
| 3 | CDISCPIL01                  | DM     | 01-701-1028 | 1028              |
| 4 | CDISCPIL01                  | DM     | 01-701-1033 | 1033              |
| 5 | CDISCPIL01                  | DM     | 01-701-1034 | 1034              |
| 6 | CDISCPIL01                  | DM     | 01-701-1047 | 1047              |
| 7 | CDISCPIL01                  | DM     | 01-701-1057 | 1057              |

- Non-extensible, two dimensional
- Data repetition
- Terminology, codes not linked
- *Version Conversion*: time, **\$£€**
- Does not model trial entities
- Lacks integral metadata
- ... and more.
- **26% of CDER SDTM applications: at least 1 error**

# IMPROVE SDTM (AND BEYOND) WITH DATA AS A GRAPH

Resource Description Framework (RDF)

# WHAT IS LINKED DATA AS RDF?



...show me the details!

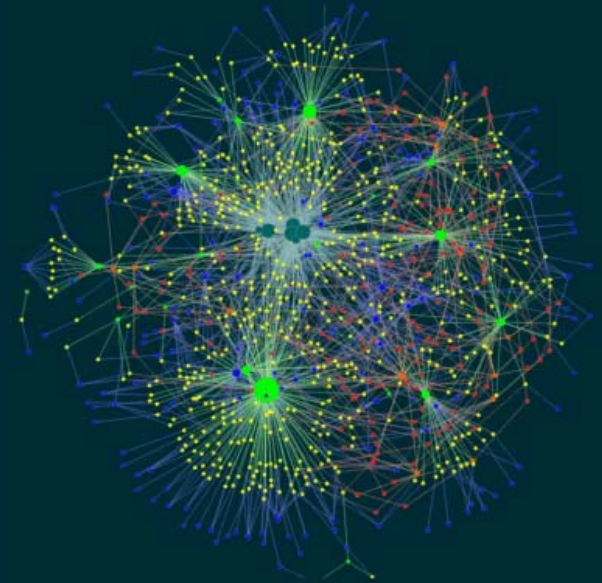
# RDF NODES AND RELATIONS





# SDTM TO GRAPH. HOW?

|           |                             |                             |                                      |                              |
|-----------|-----------------------------|-----------------------------|--------------------------------------|------------------------------|
| dm * vs * |                             |                             |                                      |                              |
| Filter    |                             |                             |                                      |                              |
|           | studyid<br>Study Identifier | domain                      | usubjid<br>Unique Subject Identifier | vsseq                        |
| 1         | CDISCPIL0T01                | VS                          | 01-701-1015                          | 1                            |
| 2         | CDISCPIL0T01                | VS                          | 01-701-1015                          | 2                            |
| 3         | CDISCPIL0T01                | VS                          | 01-701-1015                          | 3                            |
| 4         | CDISCPIL0T01                | VS                          | 01-701-1015                          | 4                            |
| 5         | CDISCPIL0T01                | VS                          | 01-701-1015                          | 5                            |
| 6         | CDI                         | dm * vs *                   |                                      |                              |
| 7         | CDI                         | Filter                      |                                      |                              |
| 8         | CDI                         | studyid<br>Study Identifier | domain                               | usubjid                      |
| 9         | CDI                         |                             |                                      | subjid<br>Subject Identifier |
| 10        | CDI                         | 1 CDISCPIL0T01              | DM                                   | 01-701-1015                  |
| 11        | CDI                         | 2 CDISCPIL0T01              | DM                                   | 01-701-1023                  |
| 12        | CDI                         | 3 CDISCPIL0T01              | DM                                   | 01-701-1028                  |
| 13        | CDI                         | 4 CDISCPIL0T01              | DM                                   | 01-701-1033                  |
| 14        | CDI                         | 5 CDISCPIL0T01              | DM                                   | 01-701-1034                  |
| 15        | CDI                         | 6 CDISCPIL0T01              | DM                                   | 01-701-1047                  |
| 16        | CDI                         | 7 CDISCPIL0T01              | DM                                   | 01-701-1057                  |



## BY DISRUPTION!

# DISRUPTING CLINICAL TRIALS WITH LINKED DATA



*Disruptors* cause damage by exciting the molecular bonds of targets to such great extents that those bonds are weakened, and/or broken, by the energies emitted, which often manifest as explosive force.

- Wikipedia

# DISRUPTING CLINICAL TRIALS WITH LINKED DATA



*Linked Data Disruptors* break traditional data structures and standards into their atomic components, allowing reassembly into future-proof, semantic data.

- Some guy who likes Linked Data

# APPROACHES TO DISRUPTION

1. Results Data --> Graph Data (RDF)
2. CDISC Models --> CDISC Graph --> Graph Data
3. **Clinical Trial Entities as Graph --> Graph Data**      PhUSE project:  
Clinical Trial Data as RDF

# THREE ASPECTS

## 1. Atomic Components



- Uniform Resource Identifiers (URIs)
- Typed literals

## 2. Classification and Standardization

- Ontologies
- Terminologies

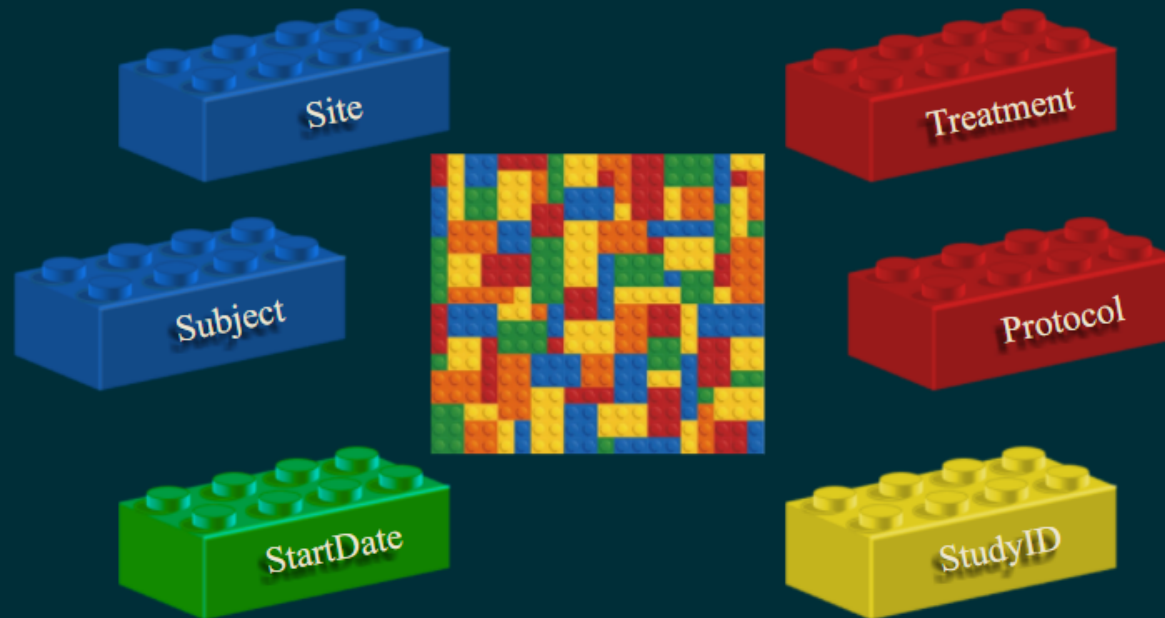
## 3. Rules

- SPARQL Inferencing Notation (SPIN)

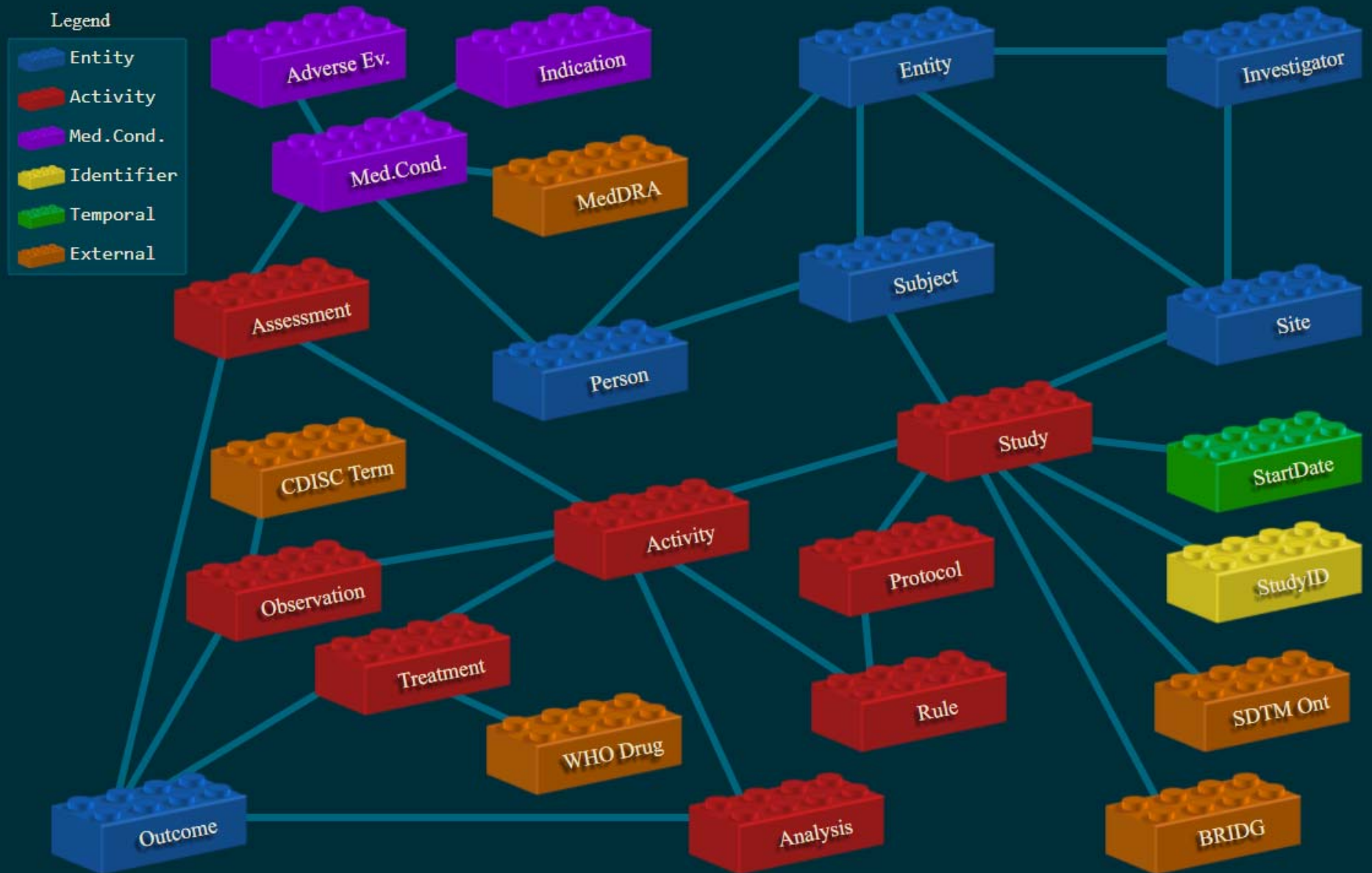
IT IS A LOT LIKE



# A STUDY HAS MANY COMPONENTS



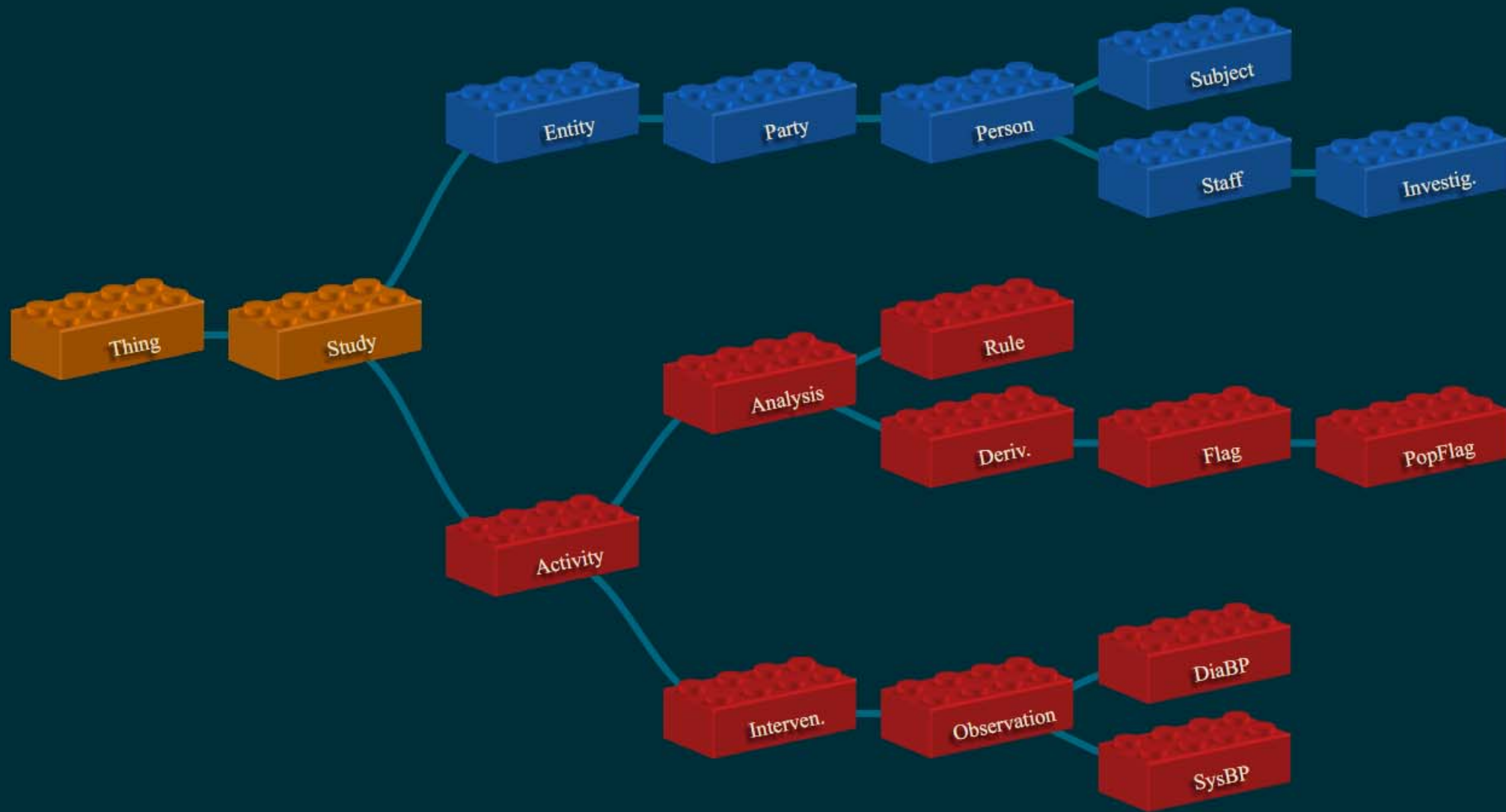
# A WEB OF STUDY DATA



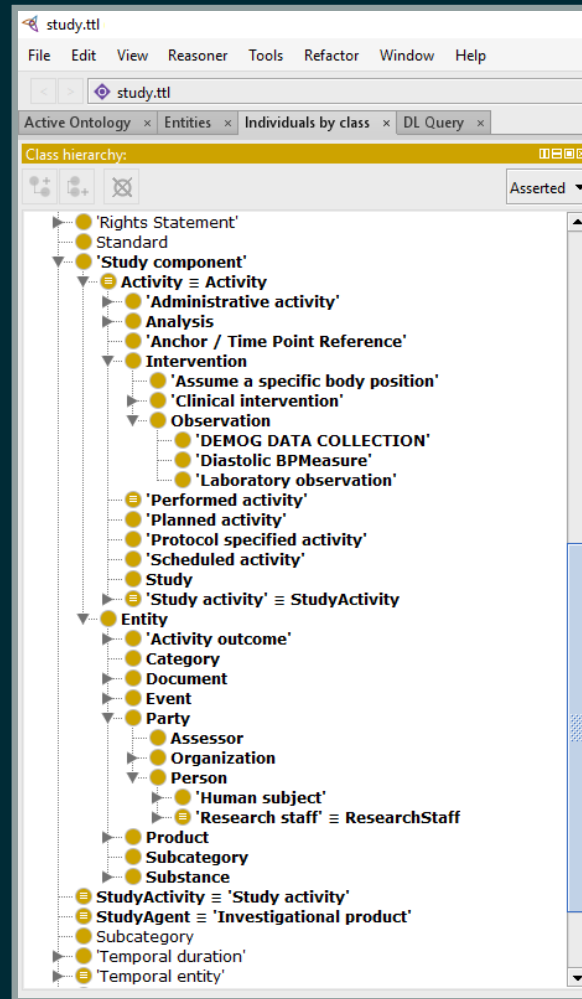


1. Atomic Components
2. Classification and Standardization
3. Rules

# CLASSIFICATION (ONTOLOGY)



# STUDY ONTOLOGY IN PROTEGE



1. Atomic Components
2. Classification and Standardization
3. Rules

# RULES

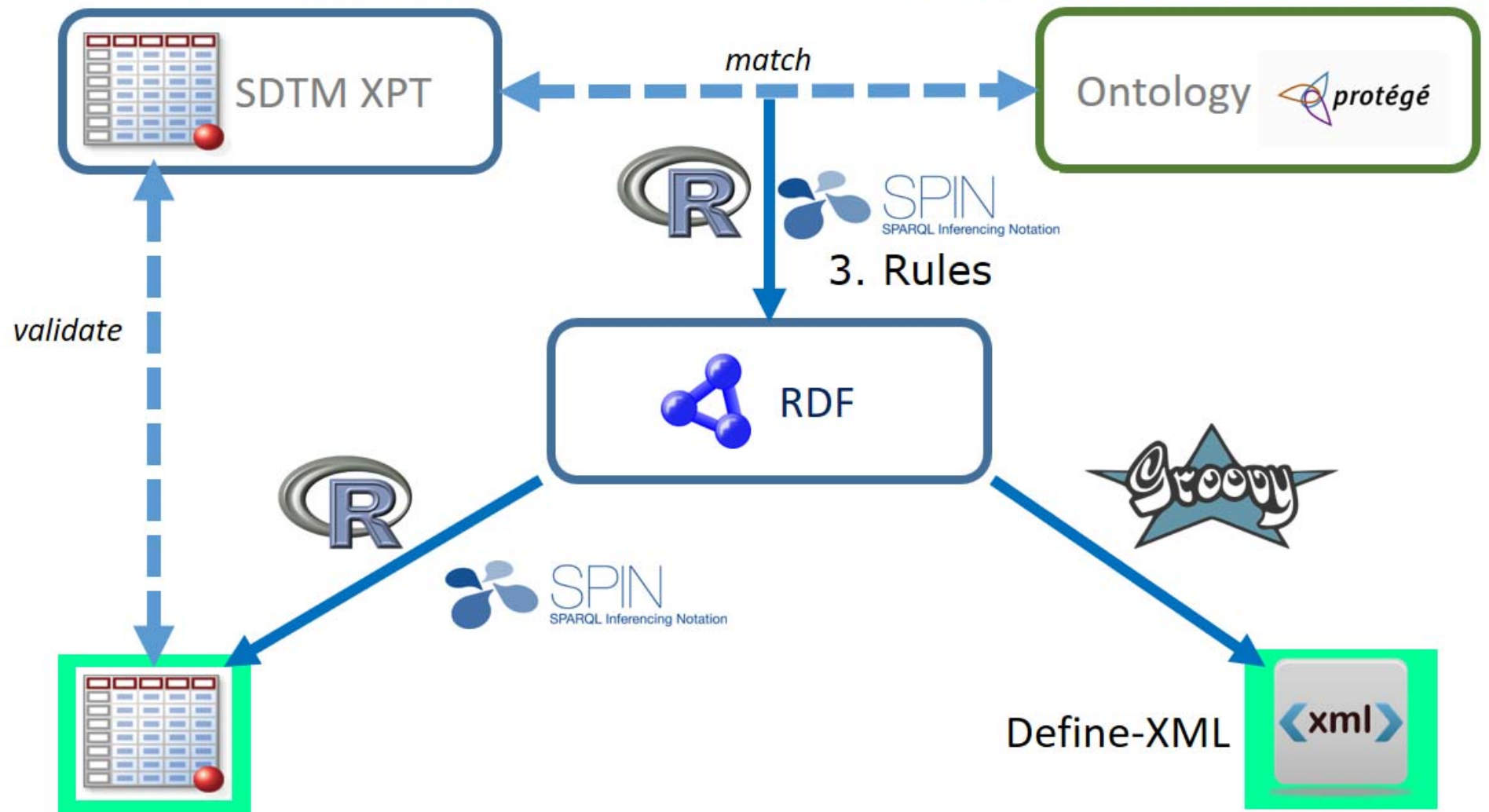
## SPARQL INFERENCE NOTATION (SPIN)

Ages cannot be negative

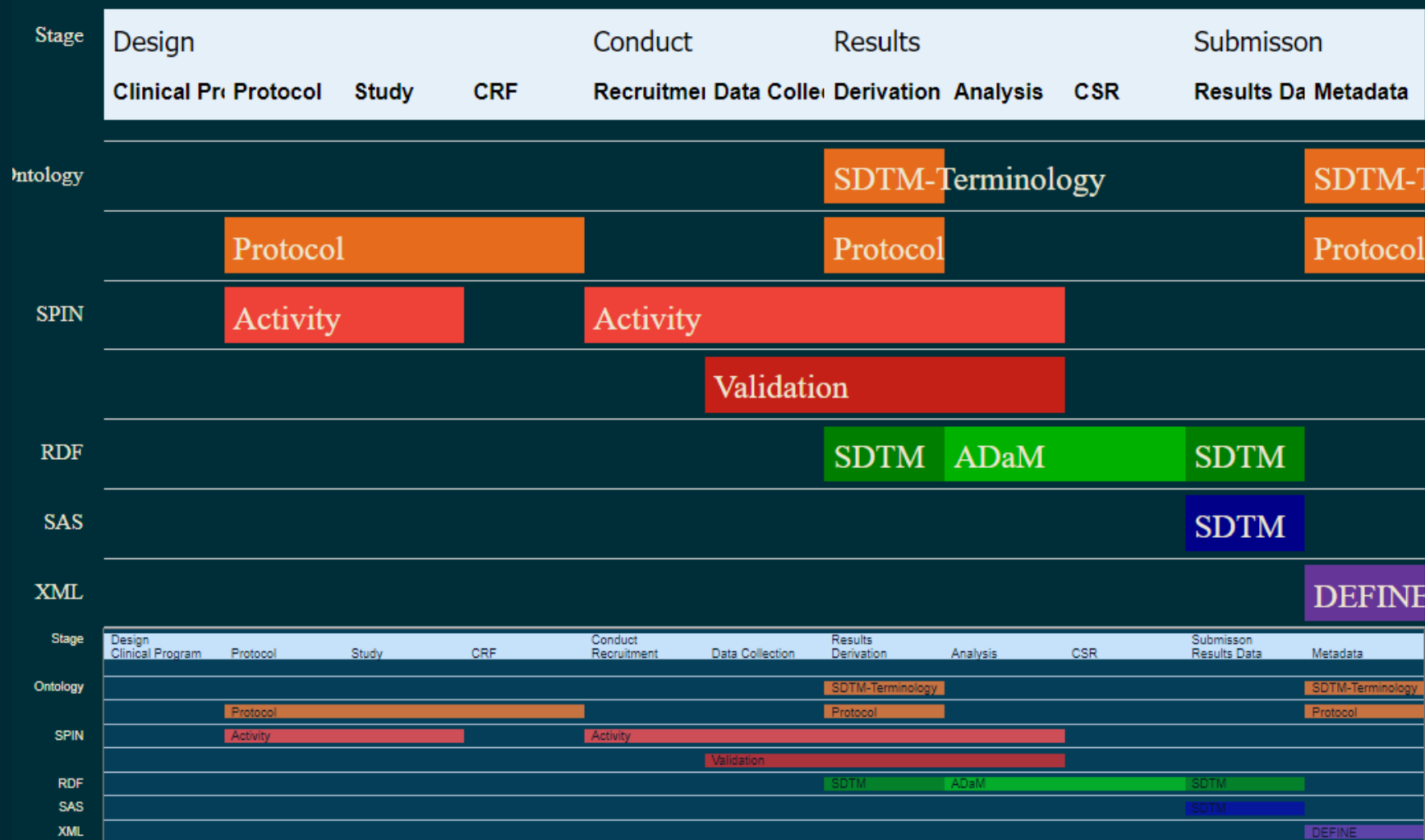
```
spin:constraint
ASK WHERE {
  ?this a study:AgeOutcome .
  ?this code:hasValue ?age .
  FILTER ( ?age < 0 ) .
}
```

## 1. Atomic Components

## 2. Classification and Standards



# CTDasRDF: Impact across the Clinical Trials Lifecycle





# CDISC *GETS IT*

- CDISC Standards as RDF
  - <https://www.cdisc.org/standards/transport/rdf>
- Project: CDISC & Transcelerate
  - CDISC SHARE 2.0 RDF Model and REST API



# INTRODUCTION TO LINKED DATA HANDS-ON WORKSHOP

White-board concepts, convert to graph data. Query, visualize.

- PhUSE CSS, Silver Spring MD, March 2018
- PhUSE EU Connect, Frankfurt, November 2018

*Enrollment limited!*

# *PHUSE PROJECT*

## **"Clinical Trials Data as RDF (CTDasRDF)"**

- Knowledge/Learn:
  - Clinical Trial Lifecycle
  - Clinical Trial Data (SDTM)
  - Define XML
  - Linked Data RDF, Ontologies, SPARQL, SPIN
  - R, Data visualization, Apache Groovy...
  - Writing: White Paper

# *PROJECT TEAM*

- Meet every 2 weeks
- White Paper, Ontology, R scripts for data conversion...
- Project Deliverables: CSS Conference, March 2018
- Final Results: US Annual Connect Conference, June 2018

# *Thank you!*



# RESOURCES

|                         |                                                                                                                                                                                                                                                                                         |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CTDasRDF Project        | <a href="https://github.com/phuse-org/CTDasRDF">https://github.com/phuse-org/CTDasRDF</a>                                                                                                                                                                                               |
| RDF                     | Primer <a href="http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/">http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/</a><br>Semantic University <a href="http://www.cambridgesemantics.com/semantic-university">http://www.cambridgesemantics.com/semantic-university</a>       |
| SPARQL Query Language   | Learning SPARQL <a href="http://www.learningsparql.com">http://www.learningsparql.com</a><br><a href="https://www.w3.org/TR/rdf-sparql-query/">https://www.w3.org/TR/rdf-sparql-query/</a><br><a href="https://www.w3.org/TR/sparql11-query/">https://www.w3.org/TR/sparql11-query/</a> |
| SPIN                    | <a href="http://spinrdf.org/">http://spinrdf.org/</a>                                                                                                                                                                                                                                   |
| Protege Ontology Editor | <a href="http://protege.stanford.edu/">http://protege.stanford.edu/</a>                                                                                                                                                                                                                 |
| D3JS (visualizations)   | <a href="http://d3js.org">http://d3js.org</a>                                                                                                                                                                                                                                           |
| RevealJS (slides)       | <a href="http://lab.hakim.se/reveal-js">http://lab.hakim.se/reveal-js</a>                                                                                                                                                                                                               |

Contact: [tim.williams@phuse.eu](mailto:tim.williams@phuse.eu)

# IMAGE ATTRIBUTIONS

Images not listed below are freely available under the [Creative Commons license](#) or are attributed to the author.



Tarantino, Q., et al. (2002). *Pulp fiction*. Miramax Home Entertainment.



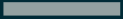
Image and use of LEGO is a trademark of the LEGO Group of companies which does not sponsor, authorize or endorse this presentation.



Romulan Disruptor Rifle.  
<http://bit.ly/2iBjAWI>



Adapted from images by Dr. Seuss



Altered from Public Domain as per  
<https://en.wikipedia.org/wiki/File:Unclesamwantyou.jpg>