

BREAKING THE MOLD: CLINICAL TRIALS DATA AS RDF

PHUSE 2017 ANNUAL CONFERENCE

Edinburgh, Scotland
October, 2017

AUTHORS

Tim Williams

UCB BioSciences
Raleigh, NC, USA
tim.williams@ucb.com

Armando Oliva

Semantica LLC
Fort Lauderdale, FL, USA
aoliva@semanticallc.com

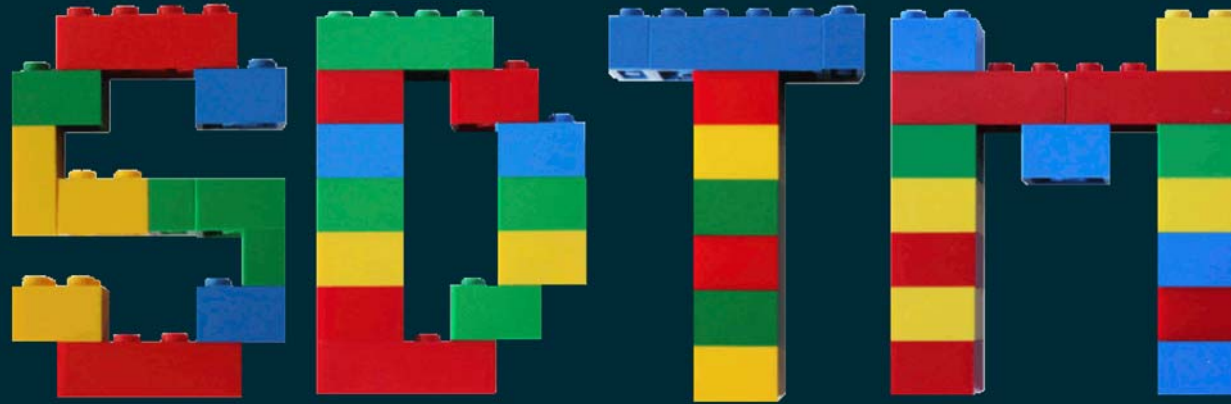
Views are those of the authors, not their employers, not PhUSE...

ALTERNATIVE TITLES

Linked Data, LEGO, and Classic Science Fiction Television!

Answer TV Trivia: Win  !

WHAT IS THE MOLD AND WHY DO WE WANT TO BREAK IT?



SDTM CAN BE IMPROVED

- Non-extensible, two dimensional data model and file format
- Data repetition, lacks true linkage to terminologies and codelists
- Lacks integral metadata
- ... etc. (see paper)
- **26% of CDER SDTM applications: at least 1 error**

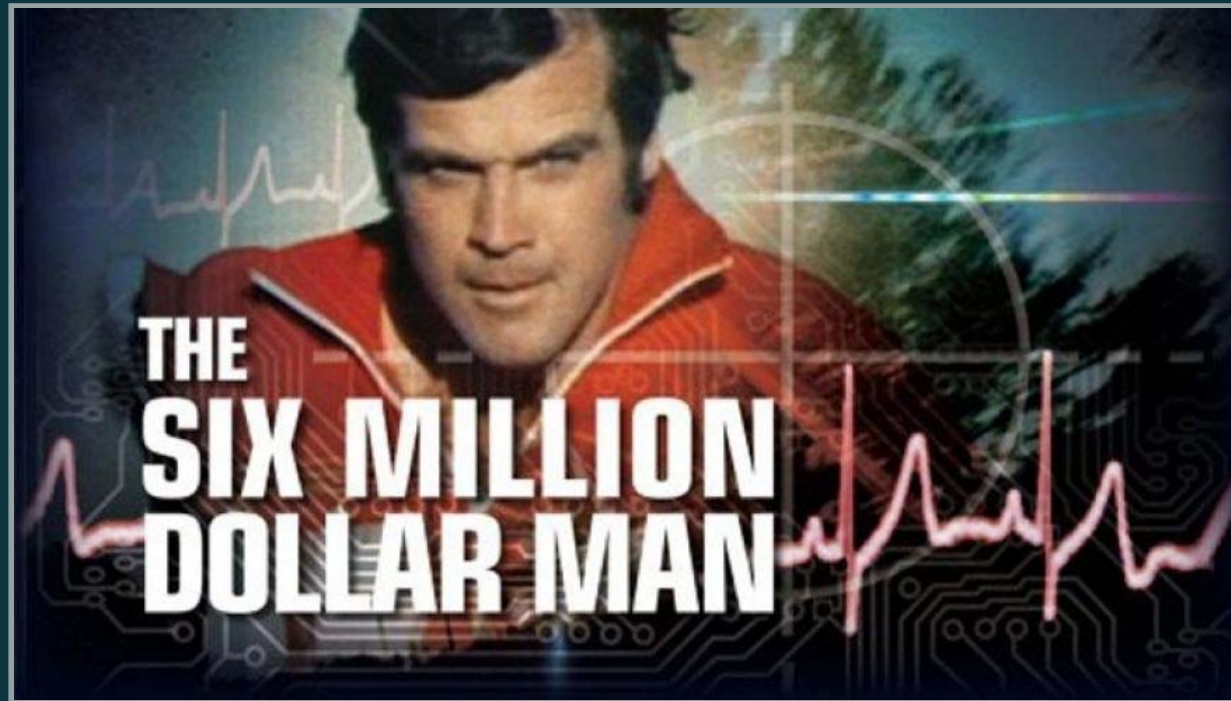
vs ✕

Filter

	studyid Study Identifier	domain	usubjid Unique Subject Identifier	vsseq		
1	CDISCPILLOT01	VS	01-701-1015	1		
2	CDISCPILLOT01	VS	01-701-1015	2		
3	CDISCPILLOT01	VS	01-701-1015	3		
4	CDISCPILLOT01	VS	01-701-1015	4		
5	CDISCPILLOT01	VS	01-701-1015	5		
6	CDI	dm ✕				
7	CDI	Filter				
	studyid Study Identifier	domain	usubjid	subjid Subject		
8	CDI					
9	CDI					
10	CDI	1	CDISCPILLOT01	DM	01-701-1015	1015
11	CDI	2	CDISCPILLOT01	DM	01-701-1023	1023
12	CDI	3	CDISCPILLOT01	DM	01-701-1028	1028
13	CDI	4	CDISCPILLOT01	DM	01-701-1033	1033
14	CDI	5	CDISCPILLOT01	DM	01-701-1034	1034
15	CDI	6	CDISCPILLOT01	DM	01-701-1047	1047
16	CDI	7	CDISCPILLOT01	DM	01-701-1057	1057

QUIZ: On what 1970's TV series did I base this quote?
"We can rebuild **SDTM**; we have the technology."

*"We can rebuild him; we have the technology."
The Six Million Dollar Man, 1973-1978*



FROM SDTM TO LINKED DATA? HOW...??

vs

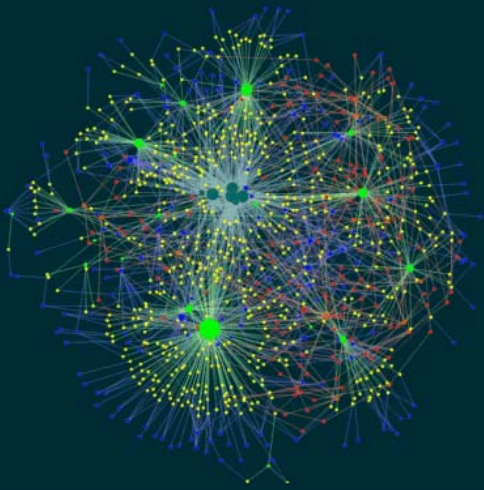
Filter

	studyid Study Identifier	domain	usubjid 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

dm

Filter

	studyid Study Identifier	domain	usubjid	subjid Subject		
8	CDI					
9	CDI					
10	CDI	1	CDISCPIL0T01	DM	01-701-1015	1015
11	CDI	2	CDISCPIL0T01	DM	01-701-1023	1023
12	CDI	3	CDISCPIL0T01	DM	01-701-1028	1028
13	CDI	4	CDISCPIL0T01	DM	01-701-1033	1033
14	CDI	5	CDISCPIL0T01	DM	01-701-1034	1034
15	CDI	6	CDISCPIL0T01	DM	01-701-1047	1047
16	CDI	7	CDISCPIL0T01	DM	01-701-1057	1057



POSSIBLE APPROACHES

1. SDTM *Data* --> Resource Description Framework (RDF)
2. SDTM *Model* as RDF --> create SDTM RDF Data.
3. **Model Clinical Trial Entities as RDF --> create SDTM RDF Data, (plus more!)**

THREE PIECES

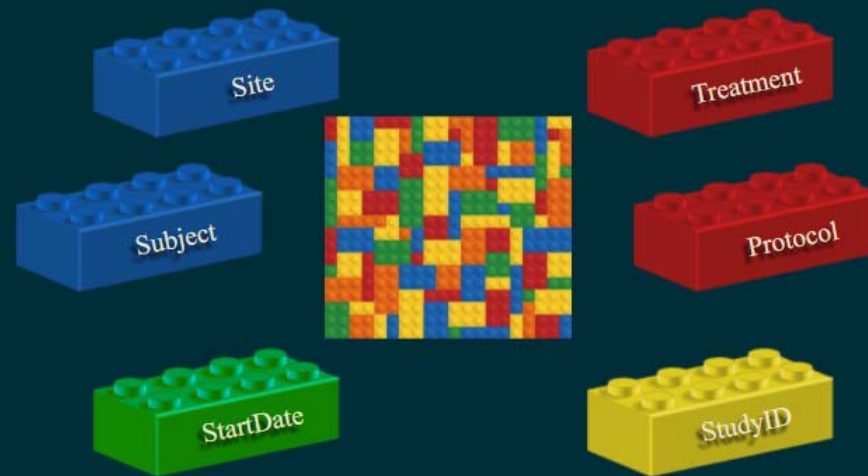
1. Atomic Components
 - Uniform Resource Identifiers (URIs)
 - Typed literals
2. Classification and Standardization
 - Ontologies
 - Terminologies
3. Rules
 - SPARQL Inferencing Notation (SPIN)

IT IS A LOT LIKE

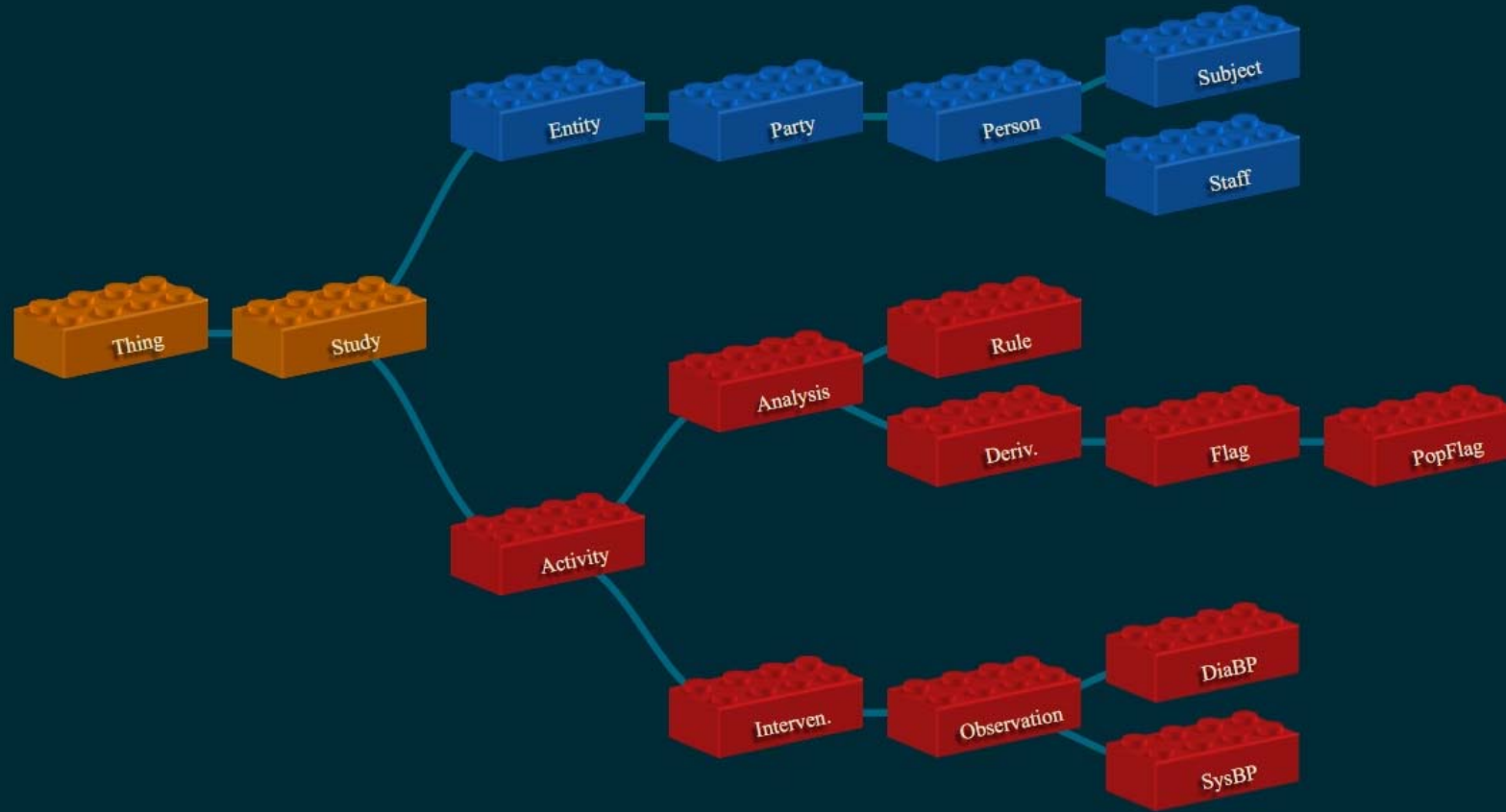


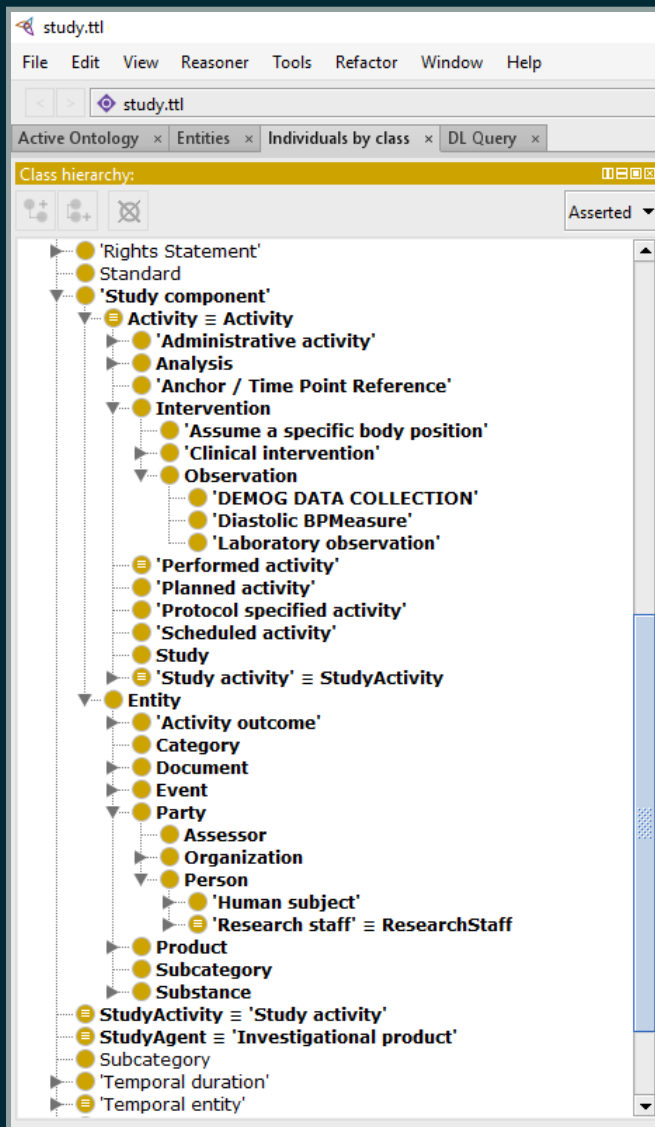
1. Atomic Components
2. Classification and Standardization
3. Rules

A STUDY HAS MANY COMPONENTS



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





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

RULES

SPARQL INFERRNCING NOTATION (SPIN)

Ages cannot be negative

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

QUIZ: On what 1960s TV series did I base this quote?

"I am not a number! I am a *rule-bound, classified, atomic component!*"

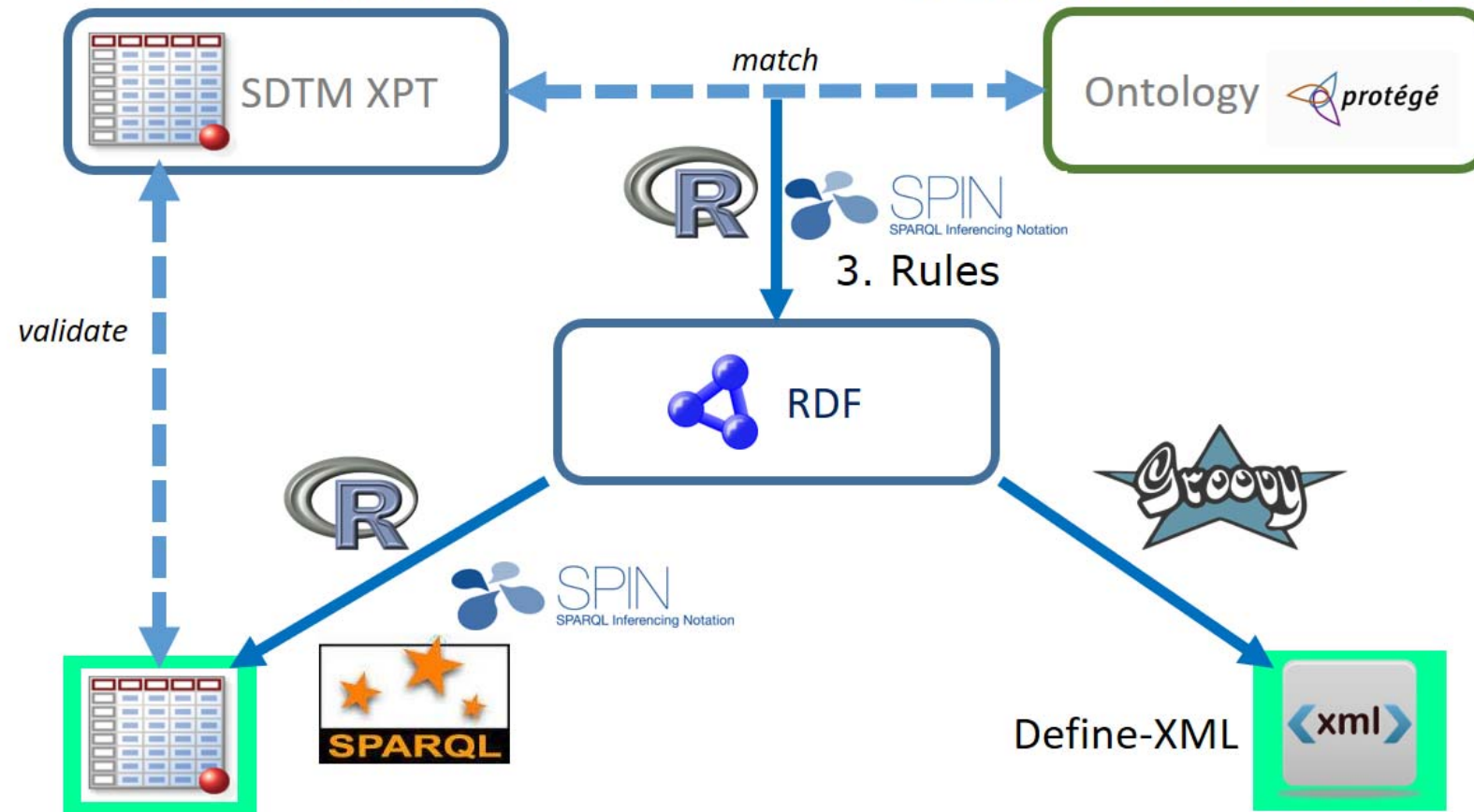
Hint:



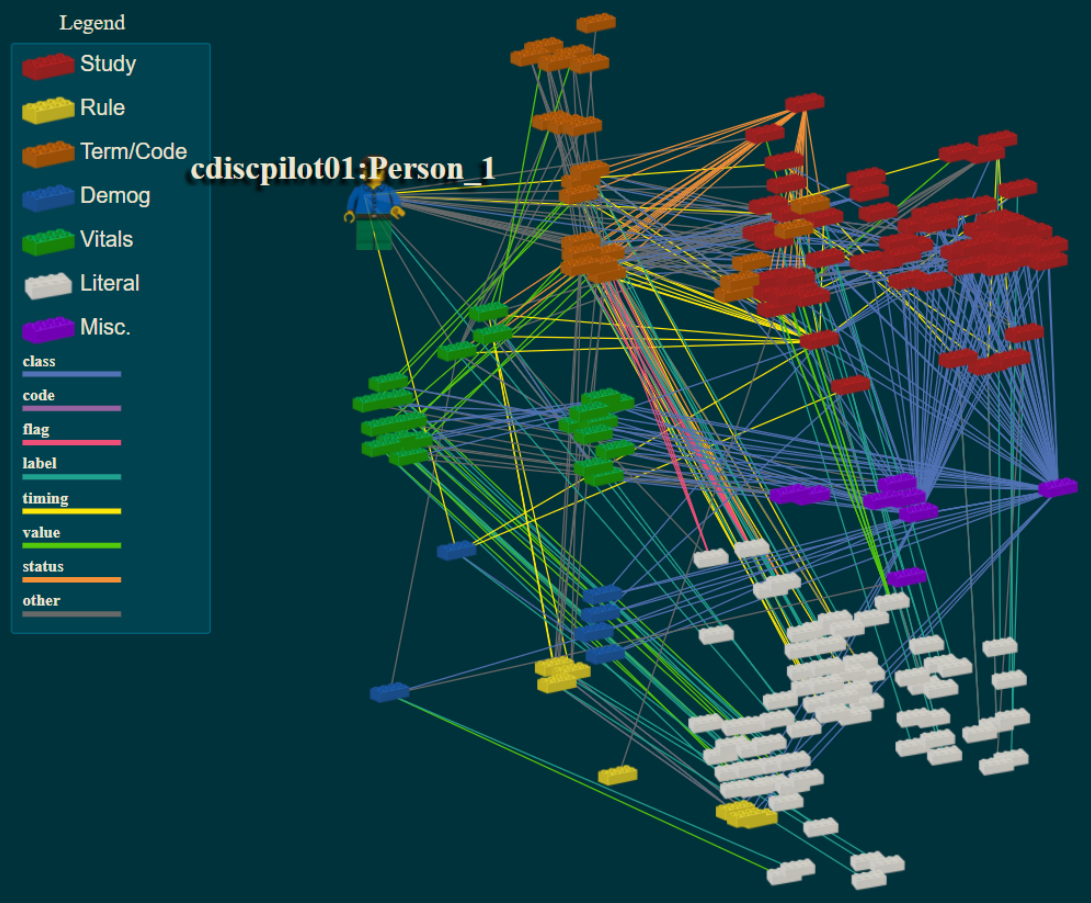
"I am not a number! I am a free man!"
The Prisoner, 1967

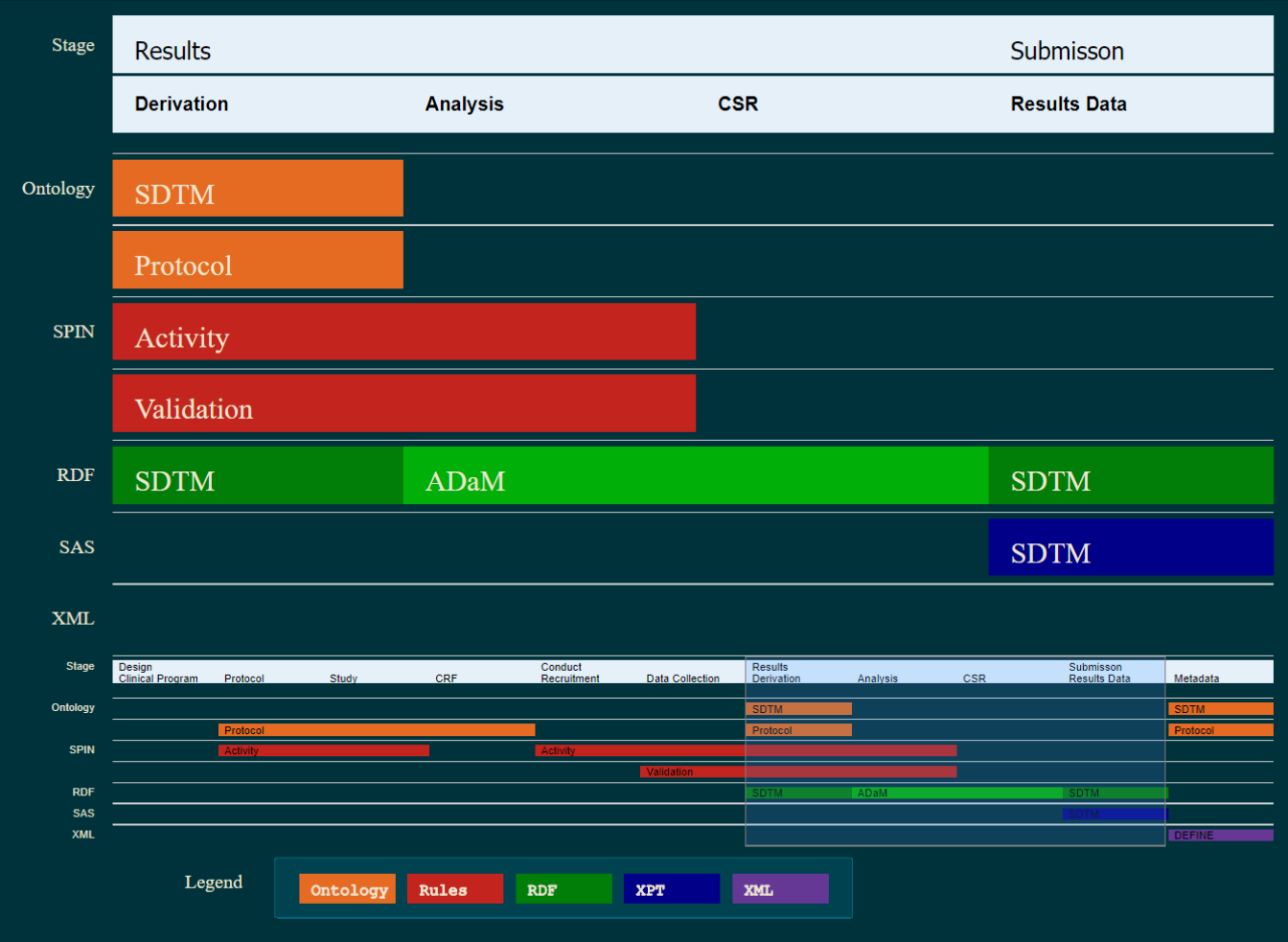
1. Atomic Components

2. Classification and Standards



PERSON 1 TRIPLES AND DESCENDANTS: DM, VS(SELECT)

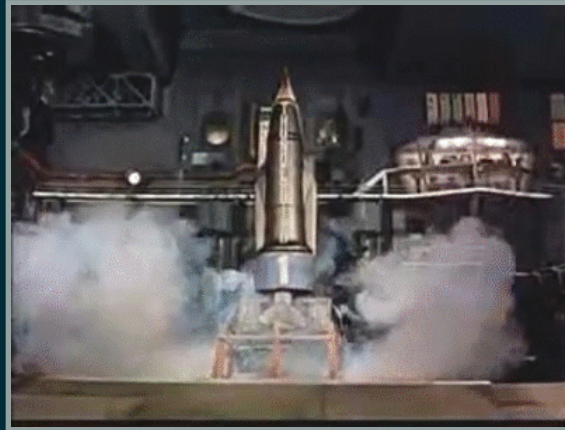




QUIZ: On what 1960s TV series did I base this quote?

"*New Data Models* are GO!!!"

Hint:



"Thunderbirds are GO!!!"
Thunderbirds are Go, 1965-1966

JOIN US!

PhUSE Project "**Clinical Trials Data as RDF (CTDasRDF)**"

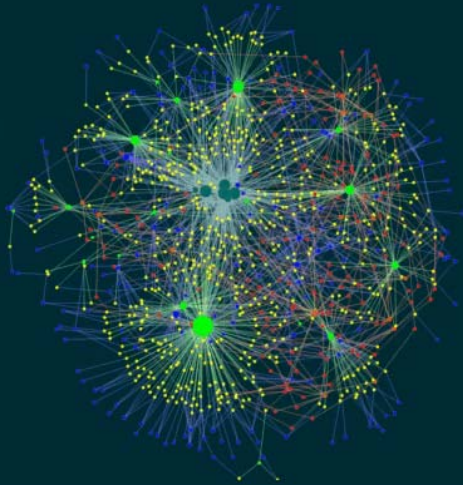
- Knowledge/Learn
 - Clinical Trial Lifecycle
 - Clinical Trial Data (SDTM)
 - Define XML
 - Linked Data RDF, Ontologies, SPARQL, SPIN
 - R, Data visualization, Apache Groovy...
 - Writing: White Paper

PROJECT TEAM

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

FROM SDTM TO GRAPH, AND BACK AGAIN!

	studyid	domain	usubjid	vsseq
	Study Identifier		Unique Subject Identifier	
1	CDISCPILOT01	VS	01-701-1015	1
2	CDISCPILOT01	VS	01-701-1015	2
3	CDISCPILOT01	VS	01-701-1015	3
4	CDISCPILOT01	VS	01-701-1015	4
5	CDISCPILOT01	VS	01-701-1015	5
6	CDI	dm		
7	CDI			
8	CDI	studyid	domain	usubjid
9	CDI	Study Identifier		Subject
10	CDI	1	CDISCPILOT01	DM
11	CDI	2	CDISCPILOT01	DM
12	CDI	3	CDISCPILOT01	DM
13	CDI	4	CDISCPILOT01	DM
14	CDI	5	CDISCPILOT01	DM
15	CDI	6	CDISCPILOT01	DM
16	CDI	7	CDISCPILOT01	DM



	studyid	domain	usubjid	vsseq
	Study Identifier		Unique Subject Identifier	
1	CDISCPILOT01	VS	01-701-1015	1
2	CDISCPILOT01	VS	01-701-1015	2
3	CDISCPILOT01	VS	01-701-1015	3
4	CDISCPILOT01	VS	01-701-1015	4
5	CDISCPILOT01	VS	01-701-1015	5
6	CDI	dm		
7	CDI			
8	CDI	studyid	domain	usubjid
9	CDI	Study Identifier		Subject
10	CDI	1	CDISCPILOT01	DM
11	CDI	2	CDISCPILOT01	DM
12	CDI	3	CDISCPILOT01	DM
13	CDI	4	CDISCPILOT01	DM
14	CDI	5	CDISCPILOT01	DM
15	CDI	6	CDISCPILOT01	DM
16	CDI	7	CDISCPILOT01	DM

Thank you!



"BREAKING THE MOLD": LINKS AND RESOURCES

Paper, Presentation PDF <https://github.com/phuse-org/CTDasRDF/tree/master/doc/Pubs/Annual2017/>

CTDasRDF Project <https://github.com/phuse-org/CTDasRDF>

RDF Primer: <http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>
Semantic University <http://www.cambridgesemantics.com/semantic-university>

SPARQL Query Language Learning SPARQL <http://www.learningsparql.com>
<https://www.w3.org/TR/rdf-sparql-query/>
<https://www.w3.org/TR/sparql11-query/>

SPIN <http://spinrdf.org/>

Protege Ontology Editor <http://protege.stanford.edu/>

D3JS (visualizations) <http://d3js.org>

RevealJS (slides) <http://lab.hakim.se/reveal-js>

IMAGE ATTRIBUTIONS

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



"The Six Million Dollar Man." American Broadcasting Company (ABC)



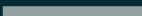
Tarantino, Q., et al. (2002). *Pulp fiction*. Miramax Home Entertainment.



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



"The Prisoner." British Broadcasting Corporation. GIF image retouches by @iamTalkyTina.



"Thunderbirds are GO!" British Broadcasting Corporation.