

TRANSFORMING CLINICAL TRIALS WITH LINKED DATA

PHUSE US CONNECT 2018

Raleigh, North Carolina
June, 2018

AUTHORS

Armando Oliva

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

Tim Williams

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

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



VS

Filter

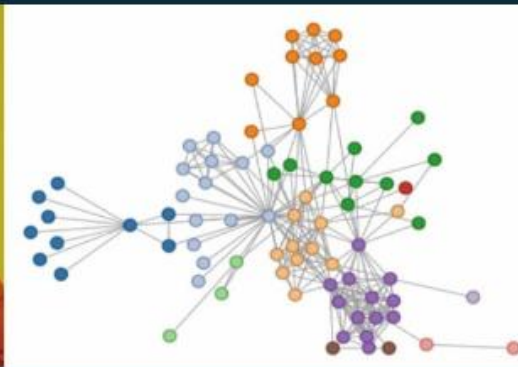
	studyid Study Identifier	domain	usubjid Unique Subject Identifier	vsseq
1	CDISCPLOT01	VS	01-701-1015	1
2	CDISCPLOT01	VS	01-701-1015	2
3	CDISCPLOT01	VS	01-701-1015	3
4	CDISCPLOT01	VS	01-701-1015	4
5	CDISCPLOT01	VS	01-701-1015	5

CDI

dm

Filter

	studyid Study Identifier	domain	usubjid	subjid Subject
8	CDI			
9	CDI			
10	CDI	1	CDISCPLOT01	DM 01-701-1015 1015
11	CDI	2	CDISCPLOT01	DM 01-701-1023 1023
12	CDI	3	CDISCPLOT01	DM 01-701-1028 1028
13	CDI	4	CDISCPLOT01	DM 01-701-1033 1033
14	CDI	5	CDISCPLOT01	DM 01-701-1034 1034
15	CDI	6	CDISCPLOT01	DM 01-701-1047 1047
16	CDI	7	CDISCPLOT01	DM 01-701-1057 1057

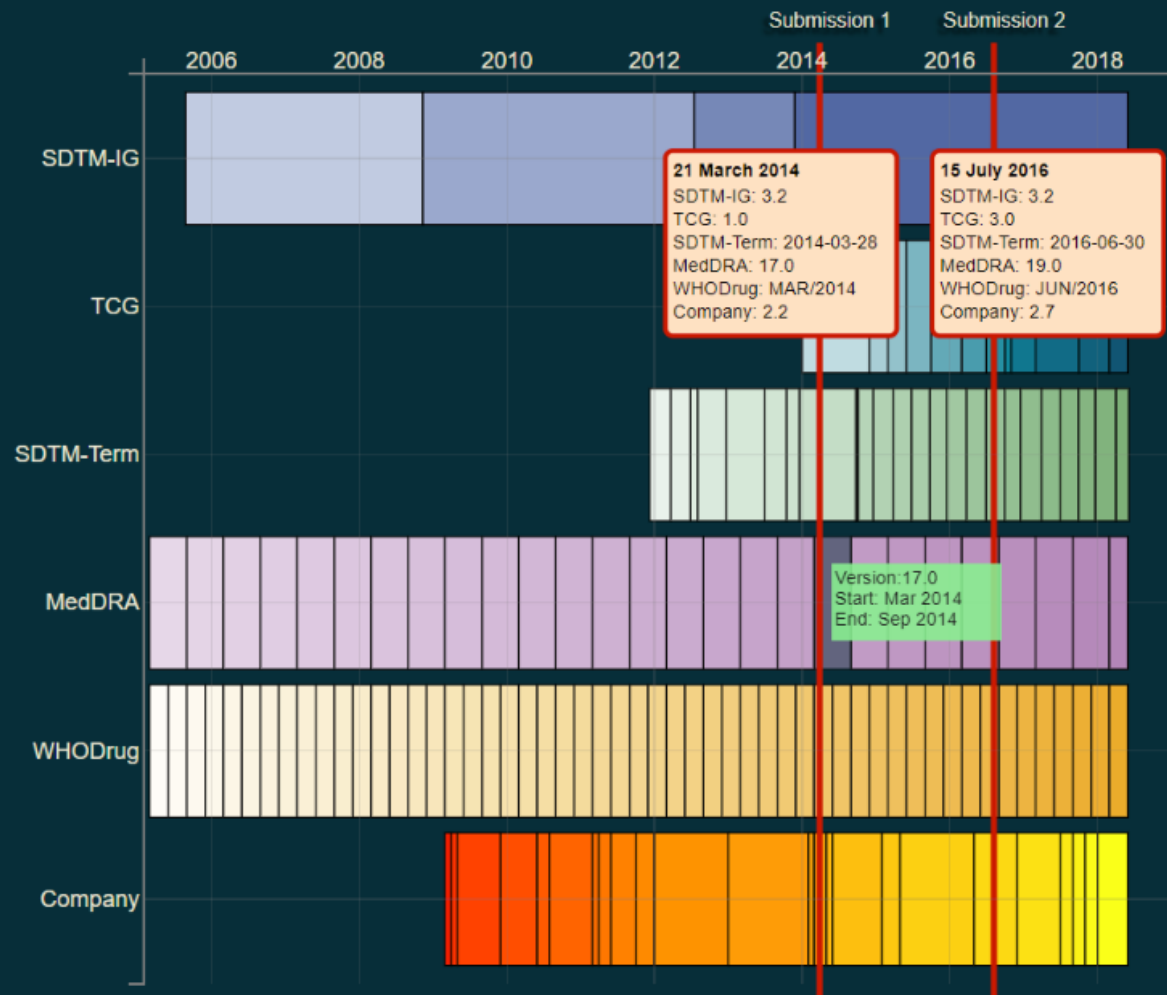




*underestimate

Standards Versioning

MULTIPLE ADVANCING STANDARDS



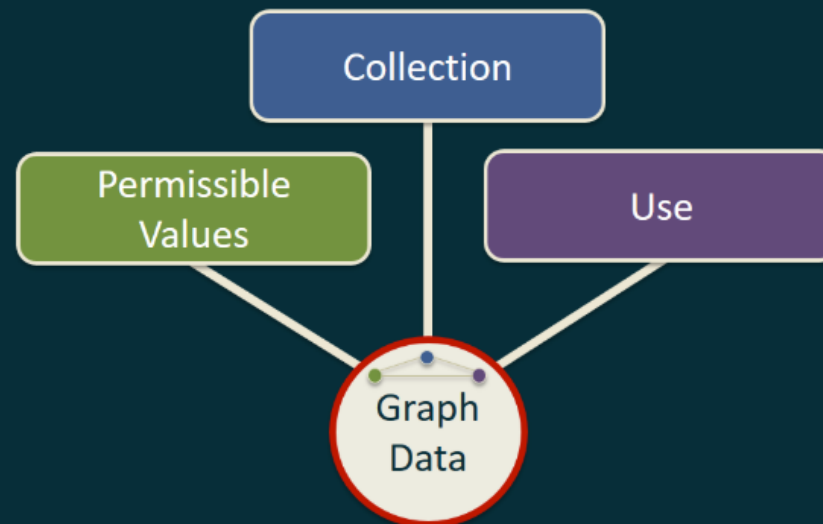
LEGEND

SDTM-IG	Study Data Tabulation Model (SDTM), Implementation Guide
TCG	Study Data Technical Conformance Guide
SDTM-TERM	SDTM terminology
MedDRA	Medical Dictionary for Regulatory Activities
WHODrug	World Health Organization Drug Dictionary
Company	Fictional company standard.

PROBLEM:
WHERE ARE THINGS *DEFINED*?

EXAMPLE: RACE

Permissible Values	SDTM Terminology (NCI) https://www.cancer.gov/.../cdisc
Collection	Collection of Race and Ethnicity Data in Clinical Trials (FDA) https://www.fda.gov/... .pdf
Use	SDTM-IG https://www.cdisc.org/.../sdtmig



PROBLEM:
WHERE ARE THINGS *STORED*?

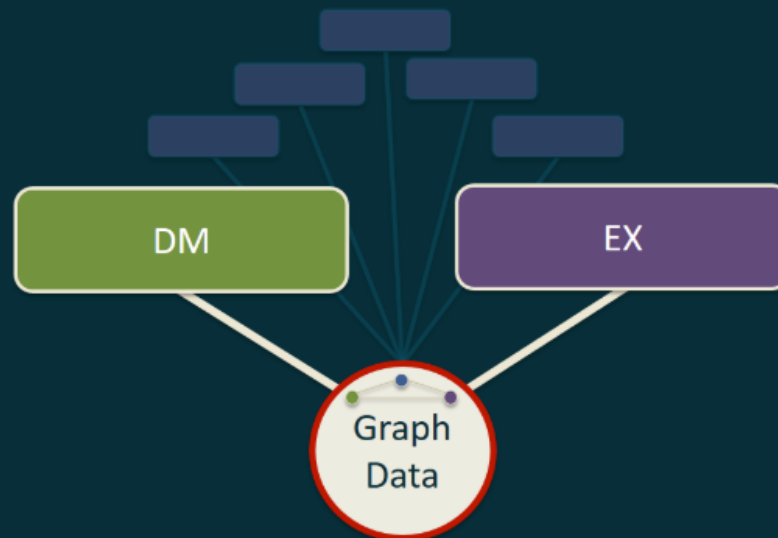
EXAMPLE: EXPOSURE

DM

usubjid	age	sex	race	rfxstdtc	rfxendtc
01-701-1015	63	F	WHITE	1/2/2014	7/2/2014
01-701-1023	64	M	WHITE	8/5/2012	9/1/2012
01-701-1028	71	M	WHITE	7/19/2013	1/14/2014

EX

usubjid	exseq	exstdtc	exendtc	exendy	exroute
01-701-1015	1	1/2/2014	1/16/2014	15	TRANSDERMAL
01-701-1015	2	1/17/2014	6/18/2014	168	TRANSDERMAL
01-701-1015	3	6/19/2014	7/2/2014	182	TRANSDERMAL



PROBLEM:
CAN DATA BE MERGED?
ARE THINGS THE SAME?

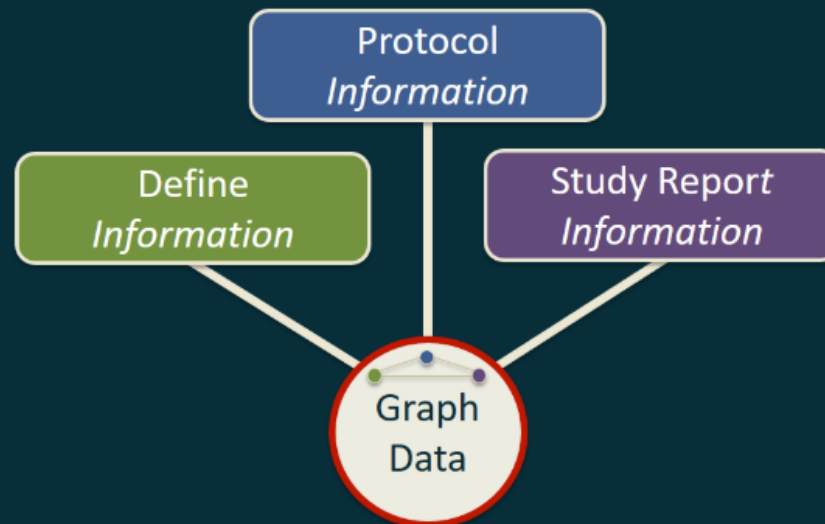
EXAMPLE: TREATMENT EMERGENT A.E.

Is **Treatment Emergent** defined identically
across multiple studies?

define.xml ?

protocol ?

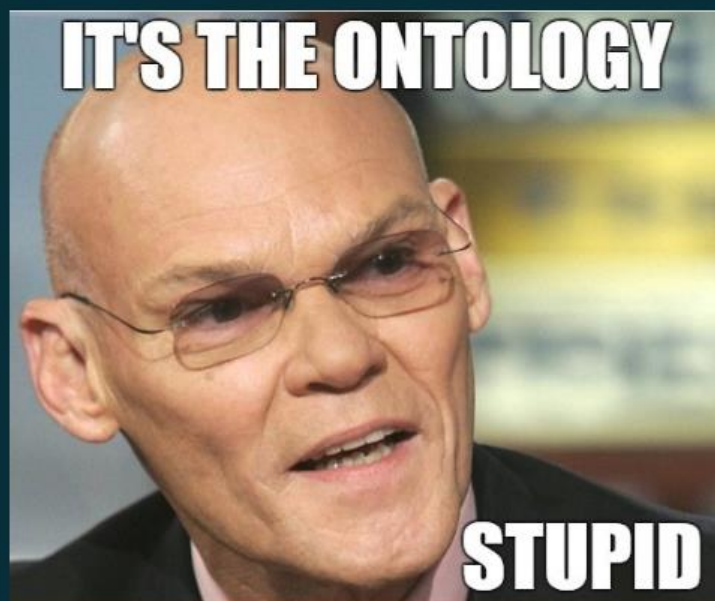
study report ?



HOW CAN LINKED DATA CAN SOLVE OUR 99* PROBLEMS?

*underestimate

GRAPH DATA MODEL BASED ON ONTOLOGIES

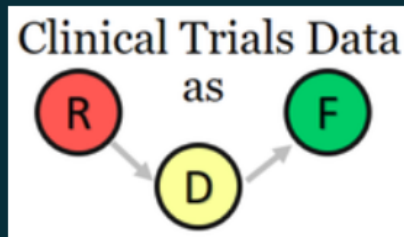


ONTOLOGY DESIGN PRINCIPLES : PART 1

- **Clinical trial = activities + outcomes**
 - Activities have Start Rules for when performed
- **Represent how data are generated and used**
 - Observe --> Assess --> Treat
- **Model the data, not the standard**
 - Meaningful classes and properties
 - Organized using relationships that make clinical sense
- **Standards agnostic**
 - Export to SDTM, ADaM, FHIR...

ONTOLOGY DESIGN PRINCIPLES (PART 2)

PHUSE PILOT PROJECT



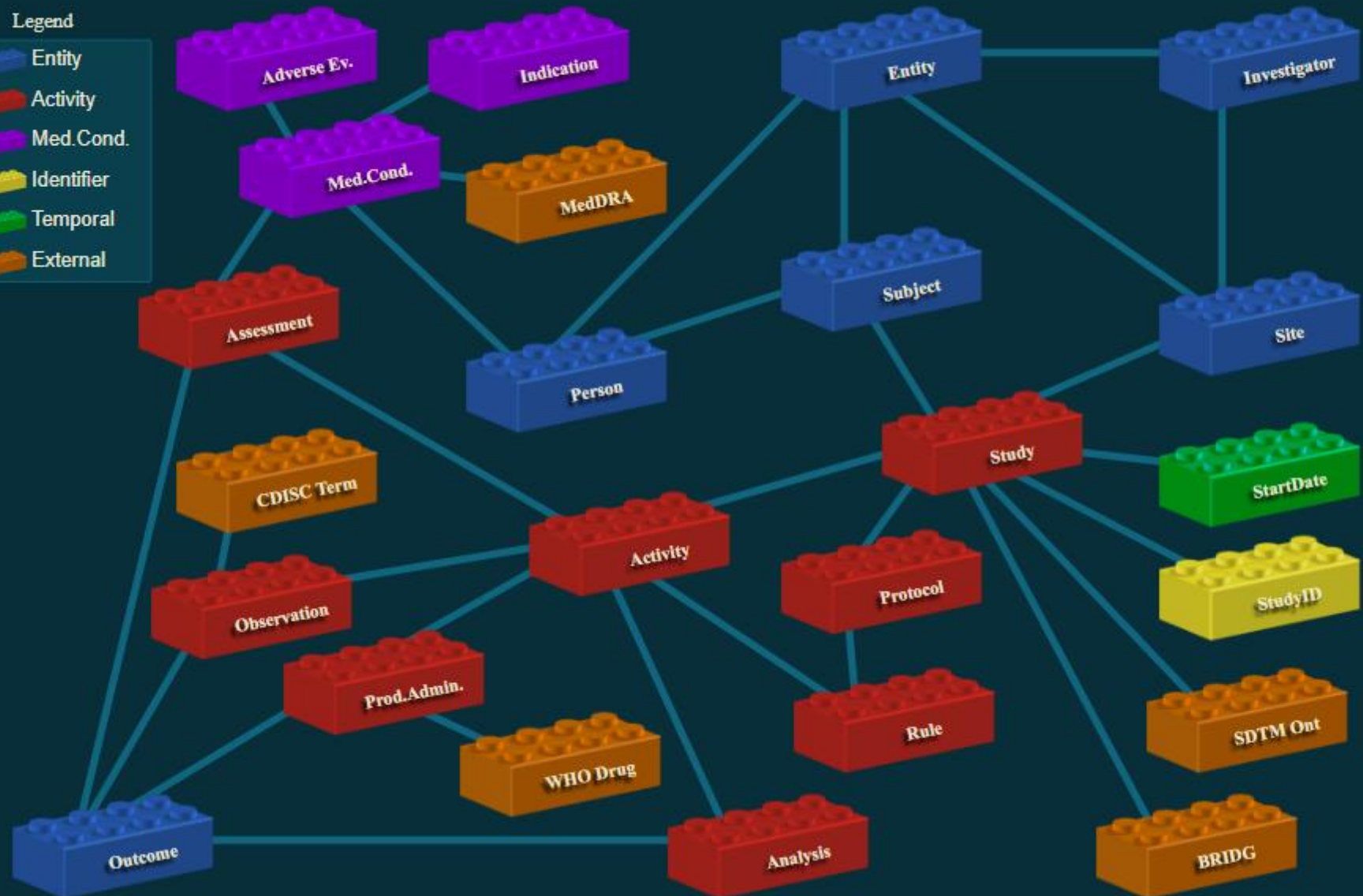
- **Build only what is needed**
 - To create DM, VS, EX
 - **Surprise: Define XML !**
- **Incremental Development**
 - Add more content later
 - Facilitated by Graph Data
- **Focus on:**
 - Flexibility
 - Extensibility
 - Reusability

Study Mini-Ontology

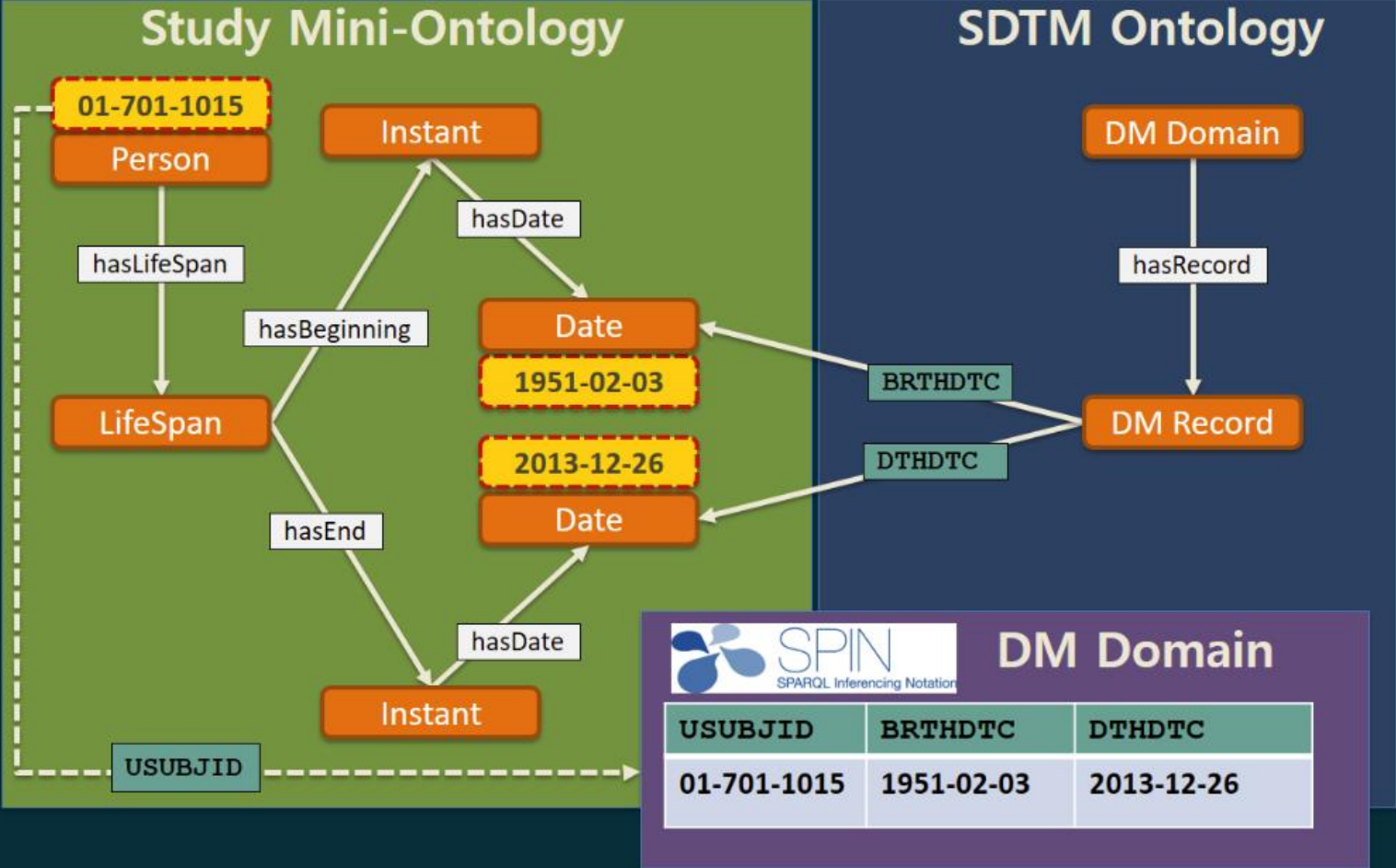
STUDY MINI-ONTOLOGY

Legend

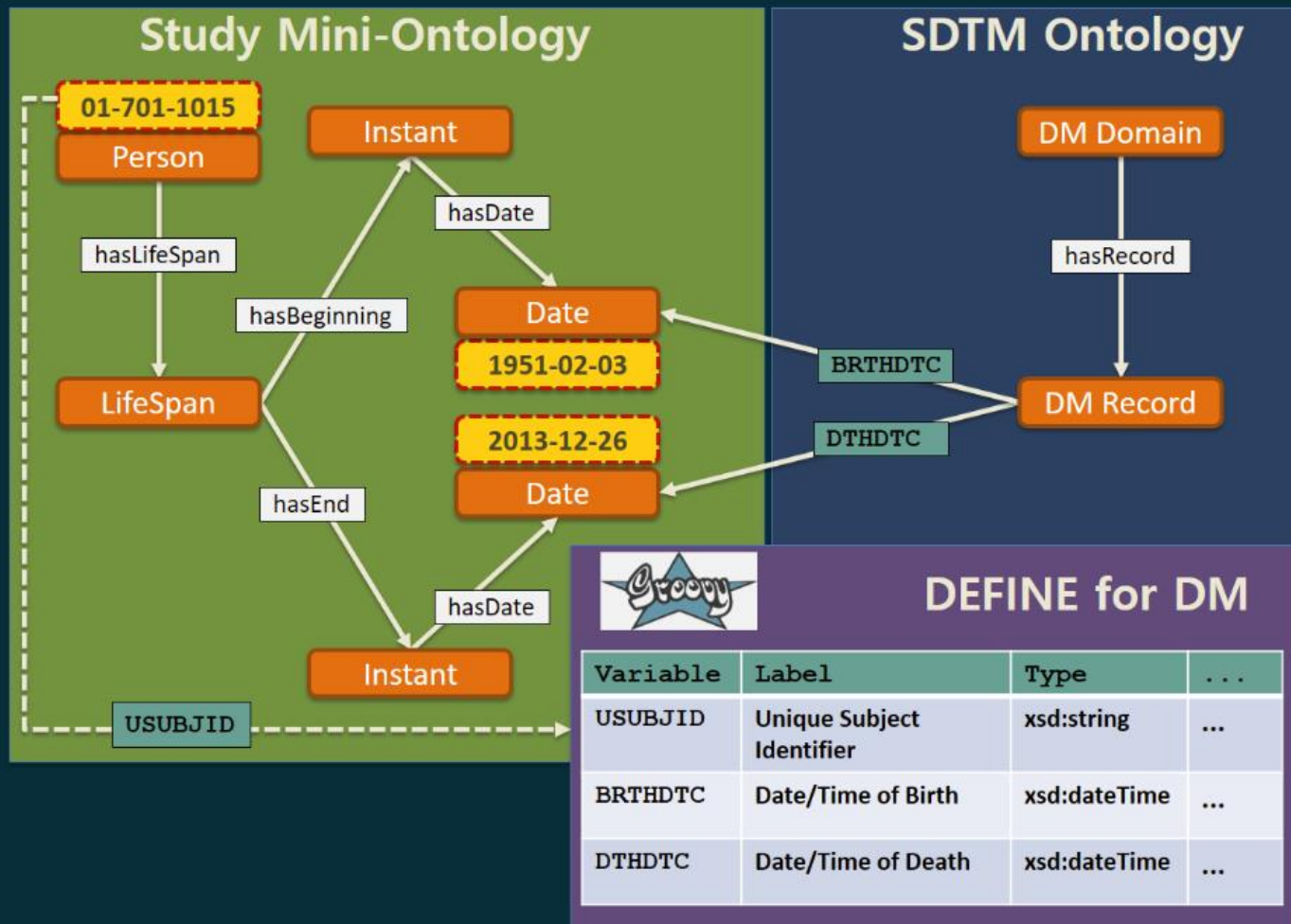
-  Entity
-  Activity
-  Med. Cond.
-  Identifier
-  Temporal
-  External



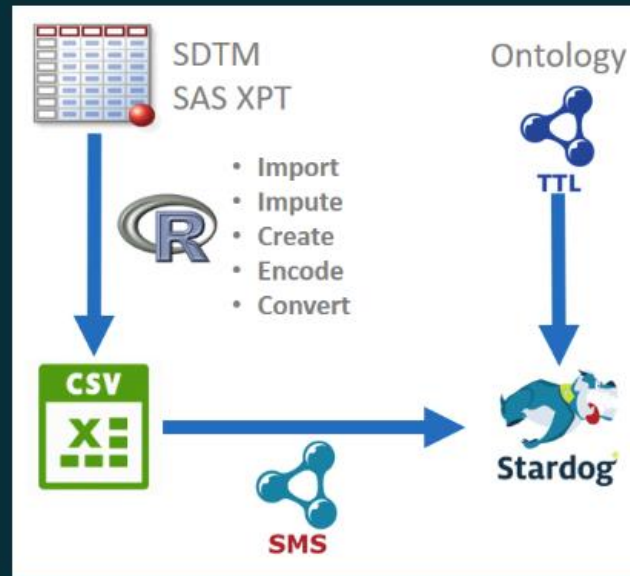
CREATE SDTM DOMAINS



CREATE DEFINE



DATA CONVERSION PART 1



Mapping process

SOURCE CSV

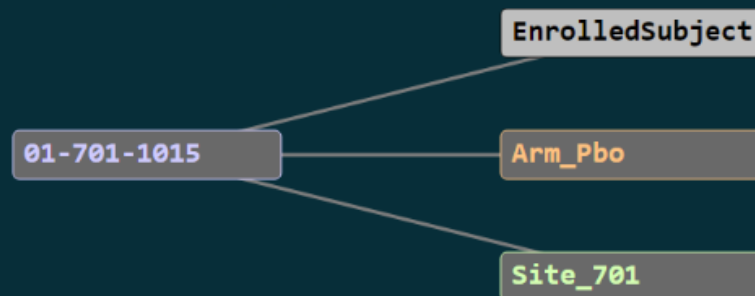
usubjid	actarmcd	siteid
01-701-1015	Pbo	701

SMS MAP

```
cdiscpilot01:{usubjid}  
  a study:EnrolledSubject;  
  study:actualArm cd01p:Arm_{actarmcd} ;  
  study:hasSite cd01p:Site_{siteid} .
```

RDF TRIPLES

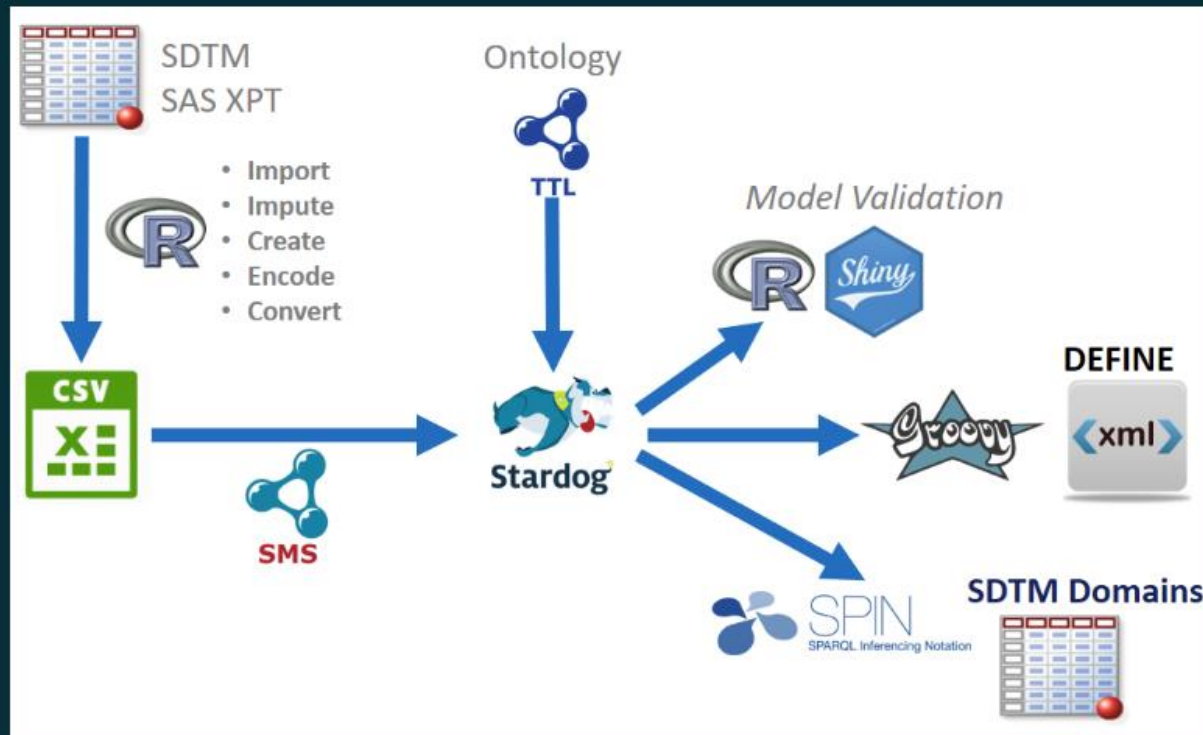
```
cdiscpilot01:01-701-1015  
  a study:EnrolledSubject;  
  study:actualArm cd01p:Arm_Pbo ;  
  study:hasSite cd01p:Site_701 .
```



STUDY CDISCPIL0T01, SUBJECT 1015
DM & VS : 200 TRIPLES



DATA CONVERSION PART 2



LEARN LINKED DATA

INTRODUCTION TO LINKED DATA INTERACTIVE, HANDS-ON WORKSHOP



4th-7th November
KAP Europa
Frankfurt, Germany

EU 2018

The Clinical Data
Science Conference



CONCLUSION



LINKED DATA CAN SOLVE OUR 99* PROBLEMS.

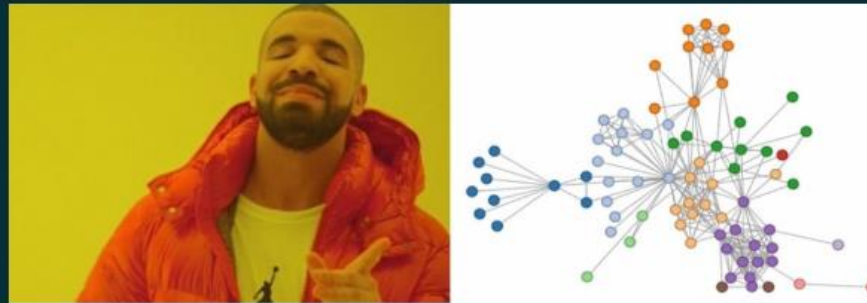
*underestimate

By modeling the entities within the Clinical Trial...

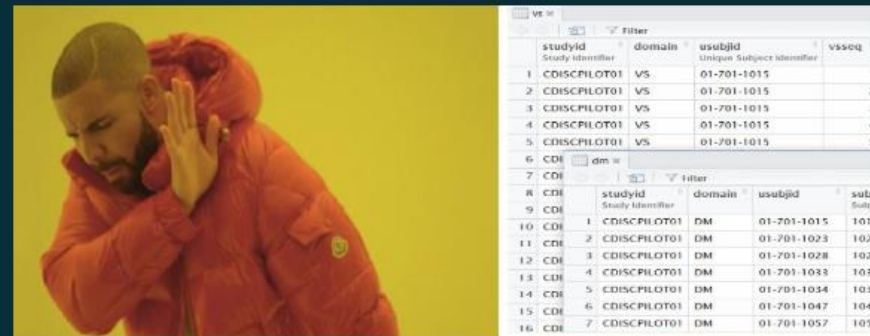
...we impact the entire data lifecycle.

OUR BIGGEST CHALLENGE

- IS NOT THE TECHNOLOGY



- IT IS TO STOP THINKING IN ROWS AND COLUMNS.



Thank you!

