



Helping the Define.xml User

Dave Iberson-Hurst
Assero Ltd & A3 Informatics

PhUSE Conference, Edinburgh
October 2017

Abstract

The FDA often comment at industry gatherings on the quality of define.xml files received as part of submissions; either they are of low quality or are missing! This presentation will look at tooling designed to help users build better define.xml files.

When creating a define.xml file there can be a lack of source information: just an annotated CRF, possibly an existing define file from a previous study or there might be a full electronic study design. The presentation will discuss in detail:

- How graph technologies can help;
- How using a semantic metadata repository can aid the process in particular with Value Level Metadata;
- How the tool can “learn” as define files are created to aid future work and guide and assist a user;
- Allow for easy upgrades from one version of define.xml to another;

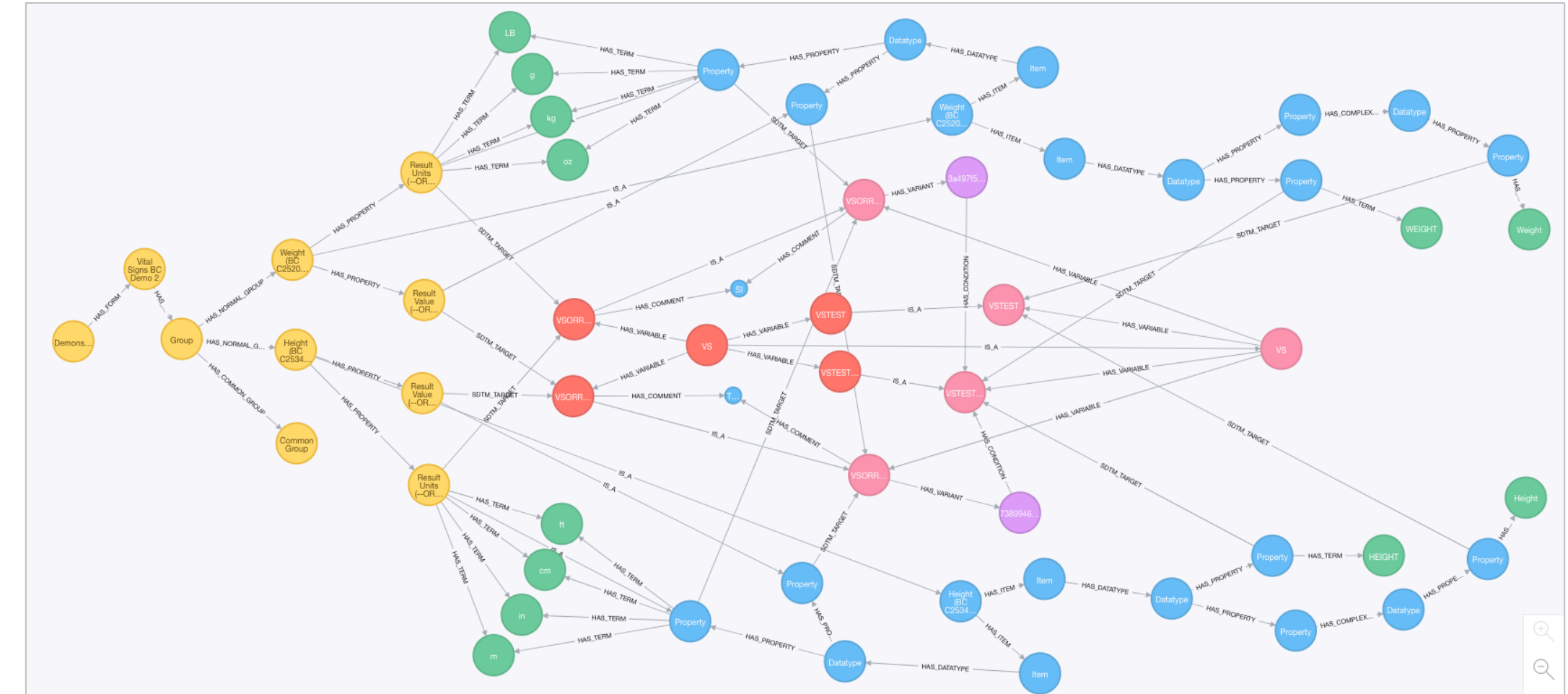
All focused around easing the user’s task and creating higher-quality define.xml files.



A Good Start

Start with an apology

- Paper written a few weeks ago.
- Have since realised that better to do it the proper way
- See the CSS Working group SDTM as RD Working Group



PhUSE 2017

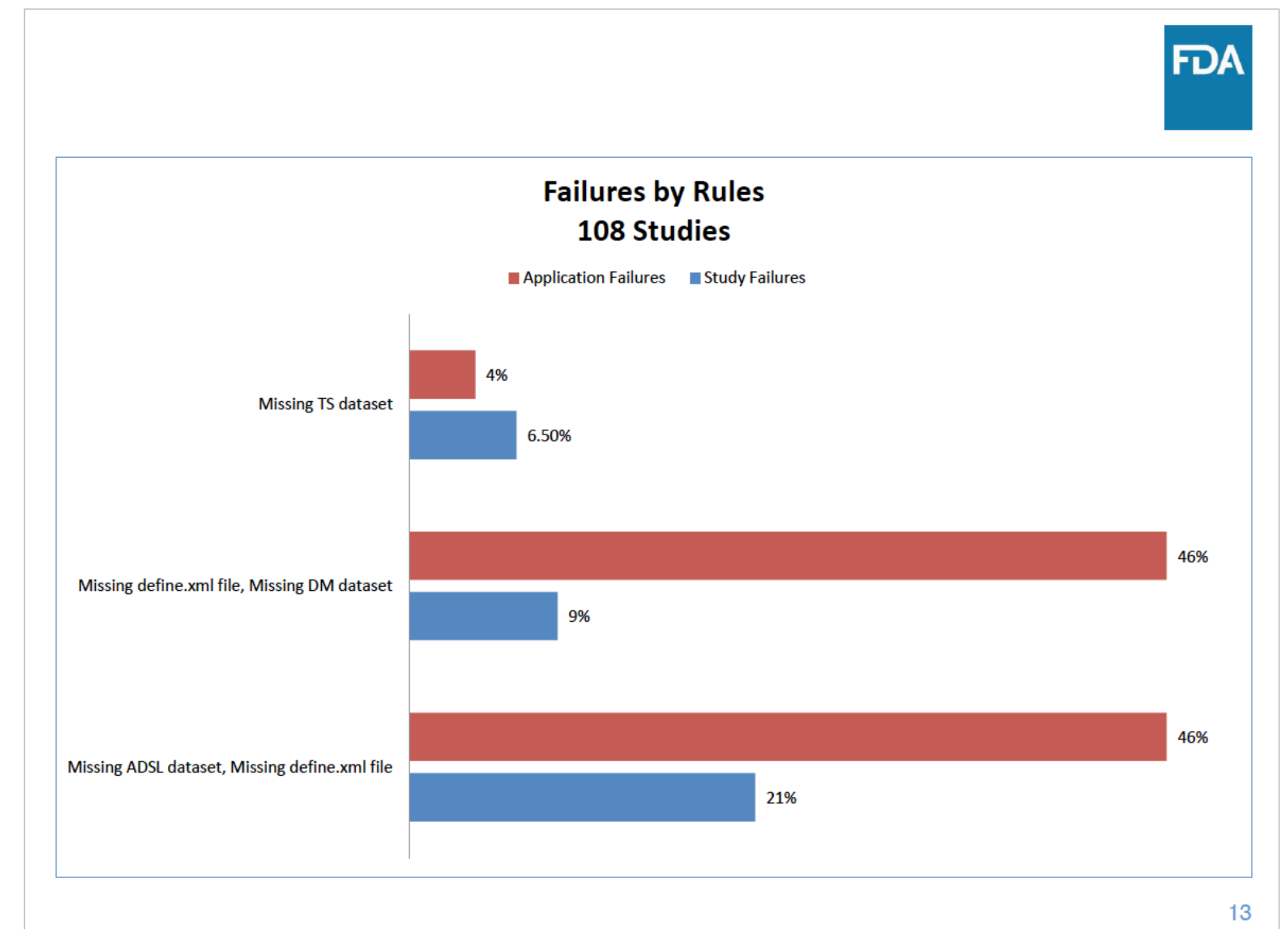
THE CORRECT GRAPH?

It is worth stepping back at this point and ask a hard question. Is this, or any define.xml graph, the right graph? Should we be modeling define.xml? The answer short-term is yes, we have issues with define.xml today, these need to be resolved as this is the reality of our work today. Longer term, the answer is no. With better metadata earlier in the life-cycle it is apparent that define.xml could be dispensed with altogether. The work presented in this paper is about solving today's issue rather than tomorrow's.

Drivers

Why look at this problem

- Quality of define.xml files
- Ease of construction of those files
- Should not need to understand the structure of the XML or the define standard
- We are describing a study
- This work has a SDTM focus



13



The Dark and Distant Past (2005)



SDTM, ODM and Define.XML



- For the wine connoisseur, it is all about the wine.
- The bottle and the glass are means of transporting the wine from the vine yard to the table.
- The wine label provides important information about the wine, the contents of the bottle.



SDTM, ODM and Define.XML

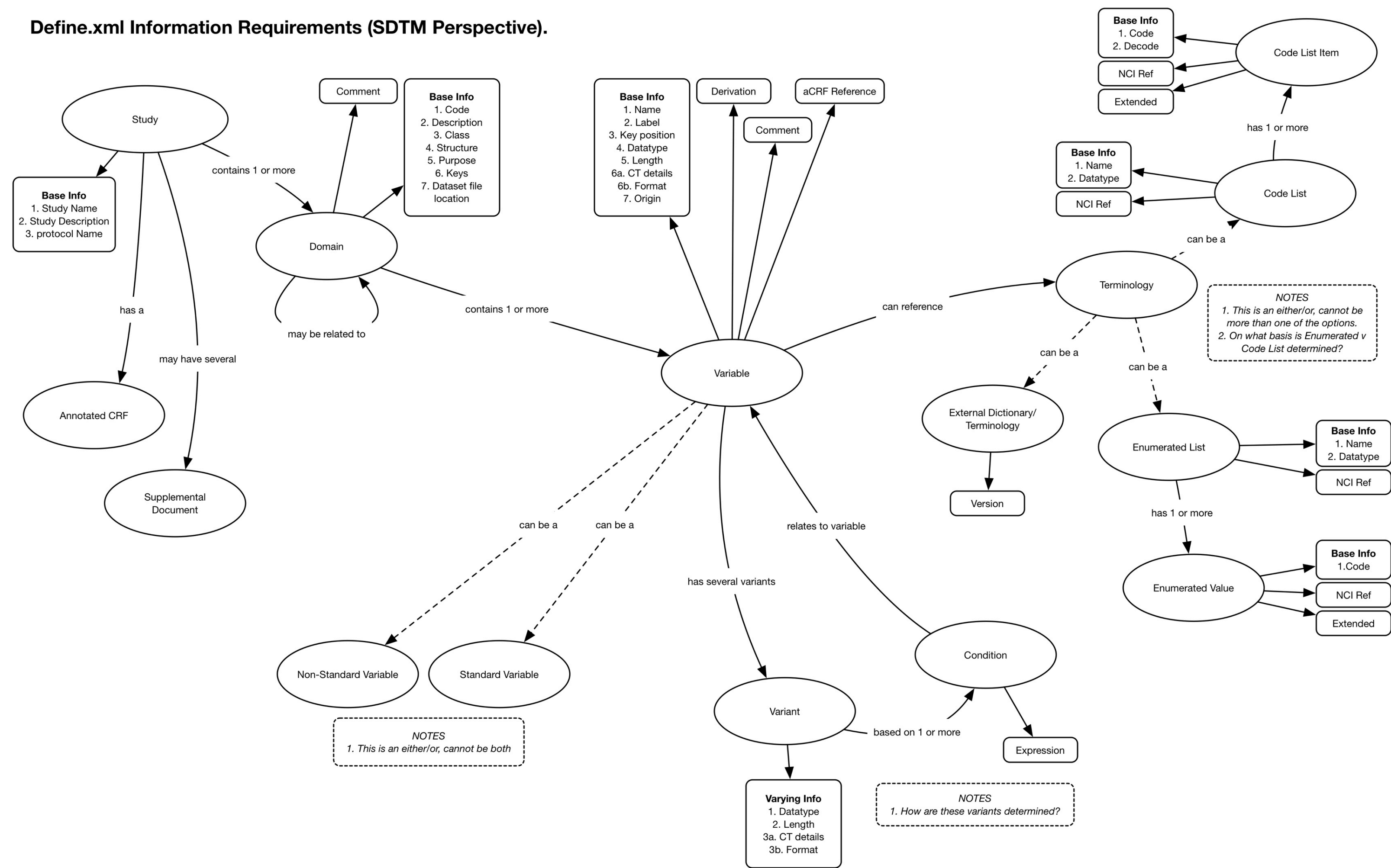


- SDTM is our wine, the important content.
- ODM is the bottle, containing the SDTM data. Currently SAS transport files are used
- Define.XML is our label, telling us what SDTM we have. A machine-readable format that replaces Define.PDF for transmitting submission metadata.

- We are describing a study, the SDTM content
- Old slides from 2005 and the CDISC Paris Interchange, the days of define.xml V1-0-0

Logical model

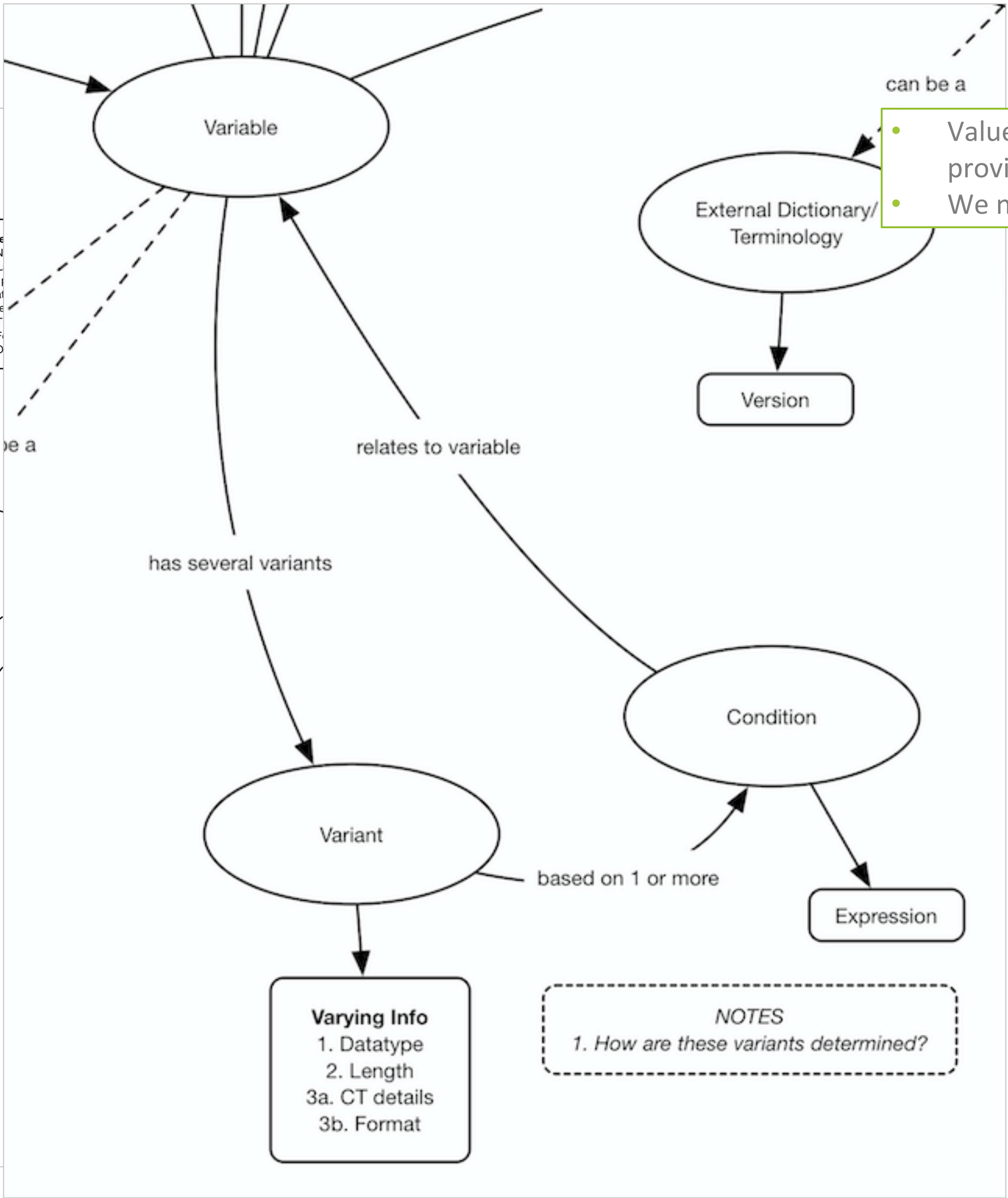
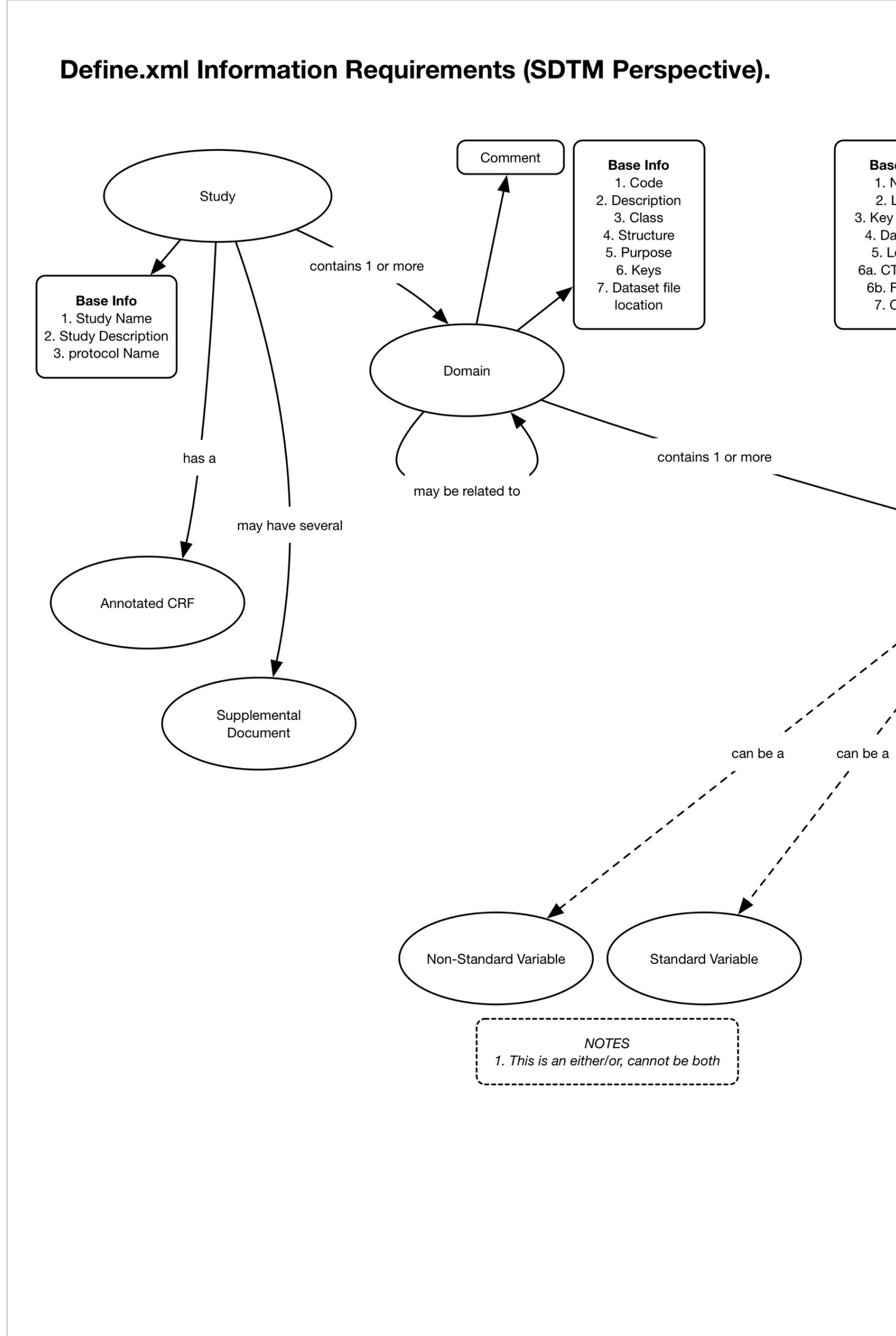
Define.xml Information Requirements (SDTM Perspective).



© Assero Limited, 2016

- A logical view of a study and the contents of a define file.
- This is an SDTM perspective and not an ADaM perspective.
- Larger version in the paper

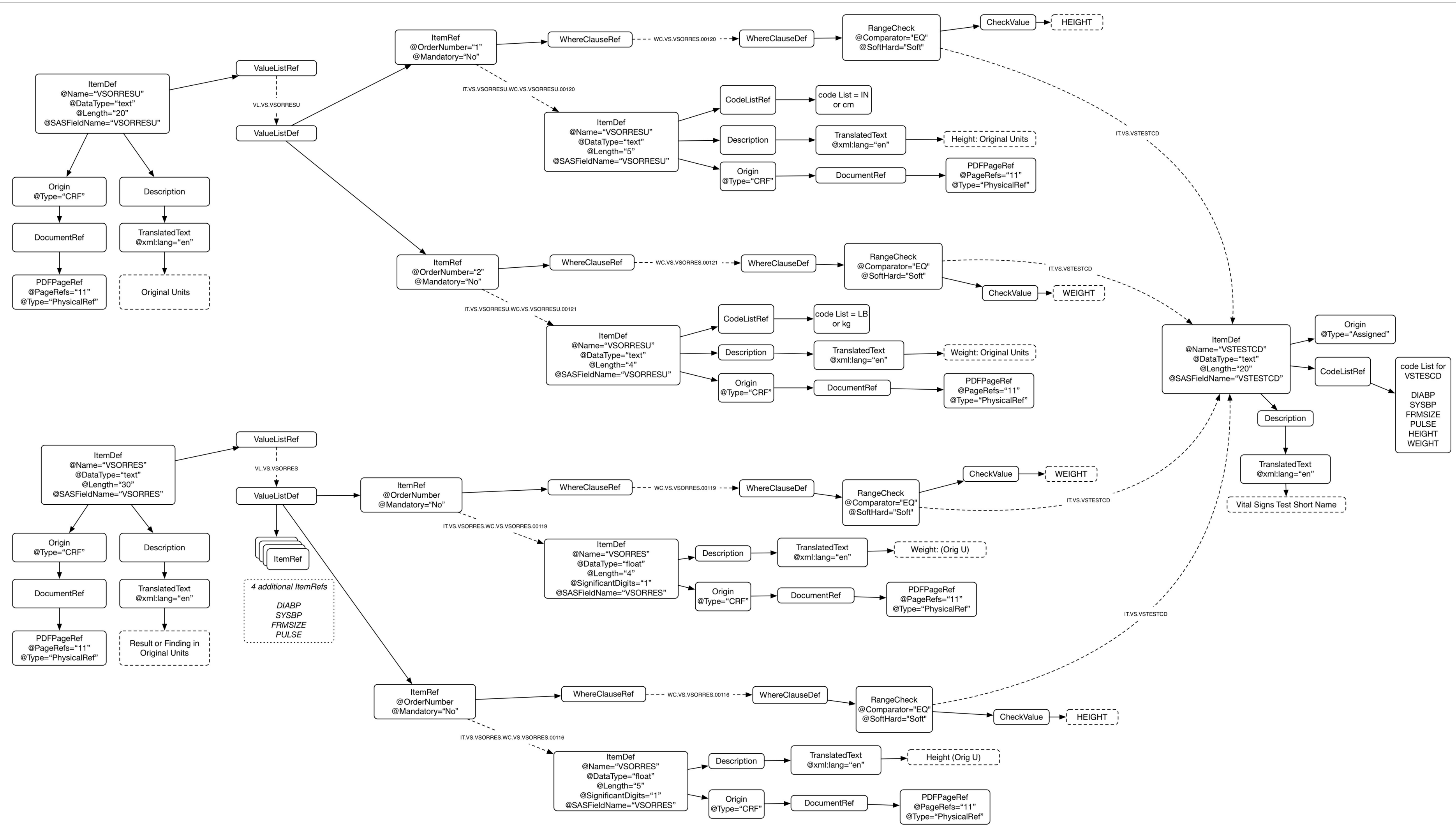
Logical model



- Value Level Metadata always seems to provides challenges
- We need to make this easy for users.

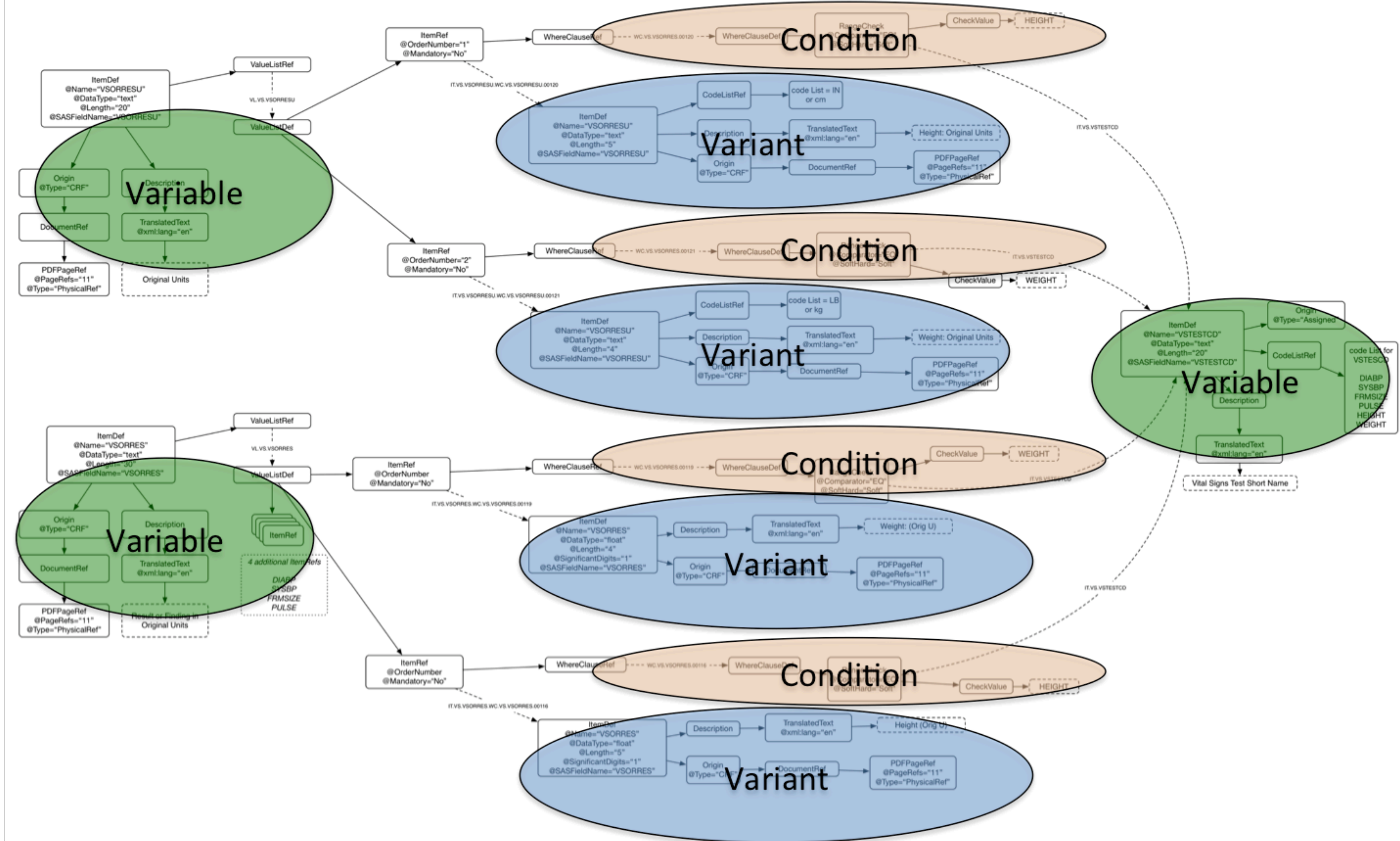


Some XML Logic



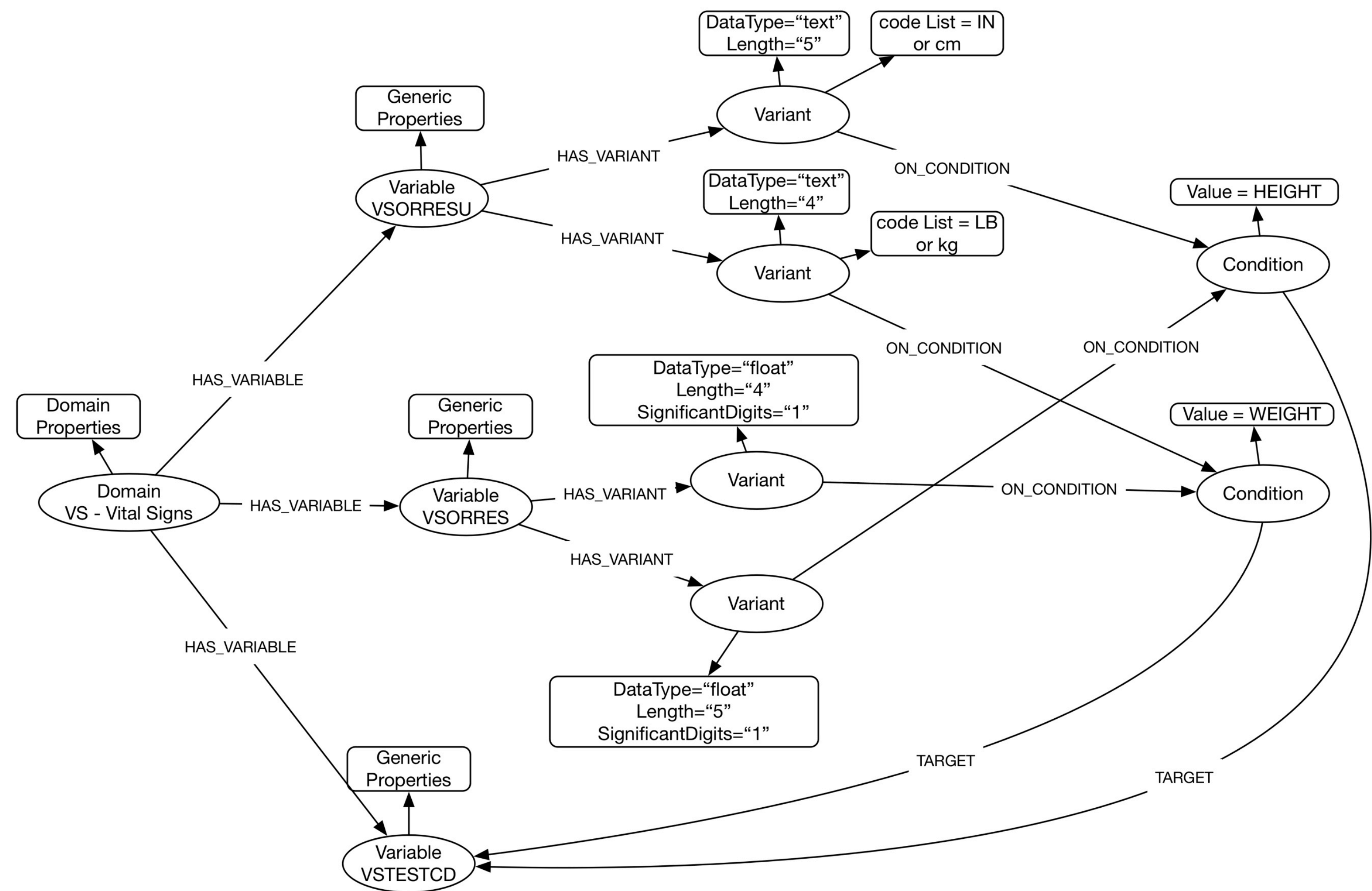
- Larger version in the paper
- Note that comments and derivations not shown to help with keeping it manageable

Back to a Logical View



- Can overlay the logical construction onto the XML

A Graph

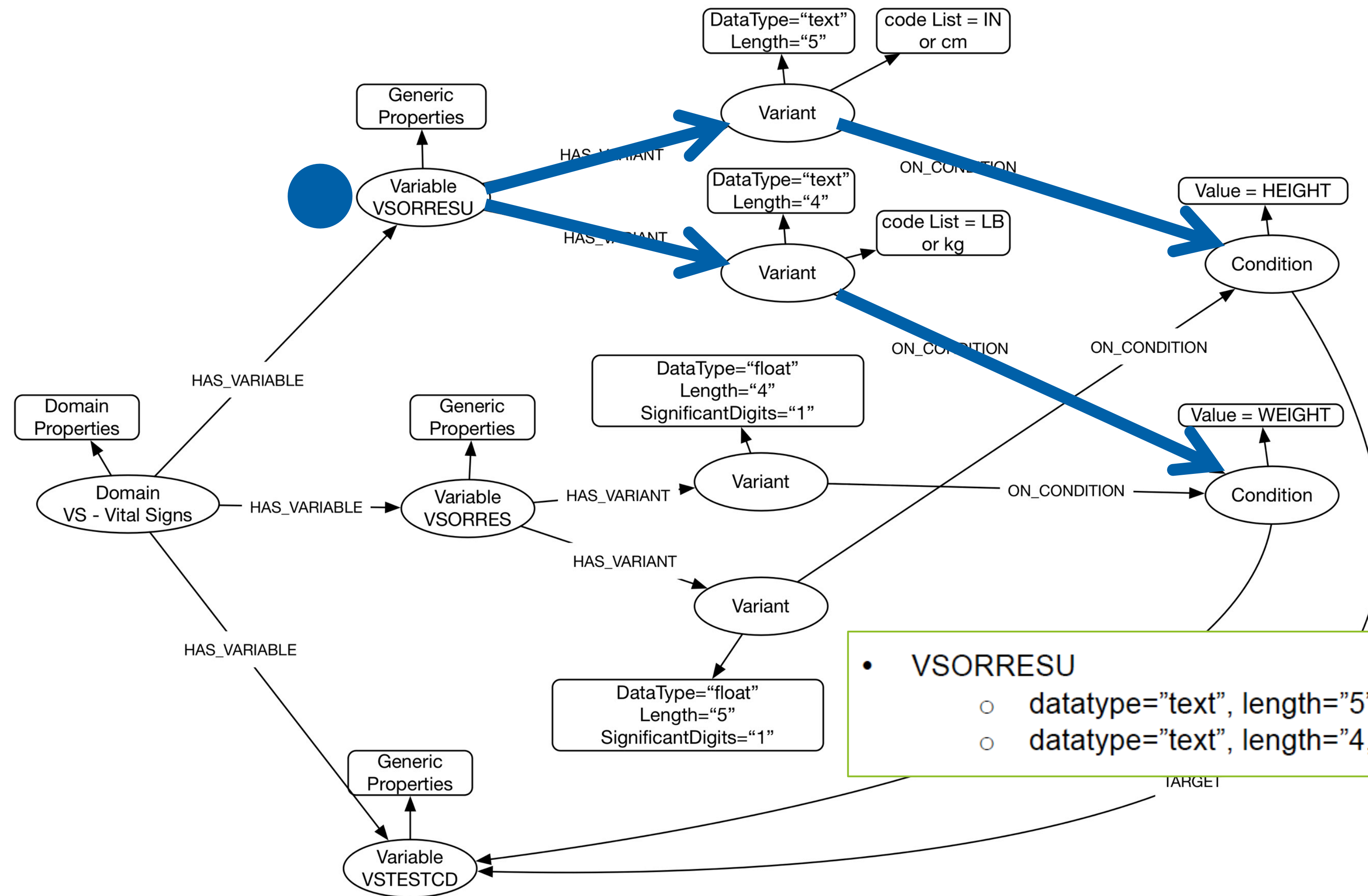


- Turn the previous into a graph
- We start to see the non-rectangular nature of our information
- LB domain example with multiple conditions in the paper



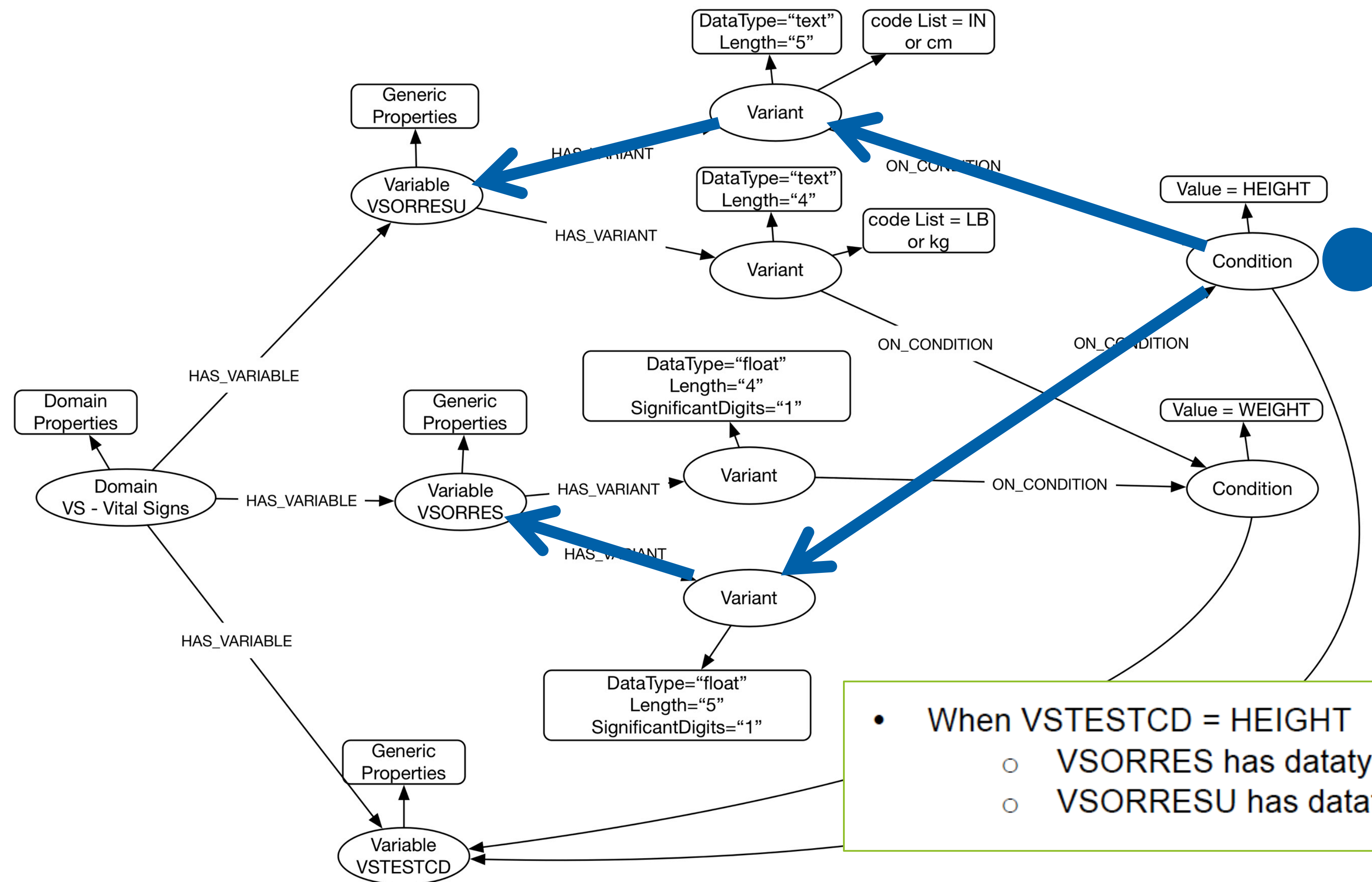
Walk the Graph

- We can walk the graph
- Start at any point and navigate the relationships



- VSORRESU
 - datatype="text", length="5", code list="IN" or "cm" when VSTESTCD = HEIGHT
 - datatype="text", length="4", code list="LB" or "kg" when VSTESTCD = WEIGHT

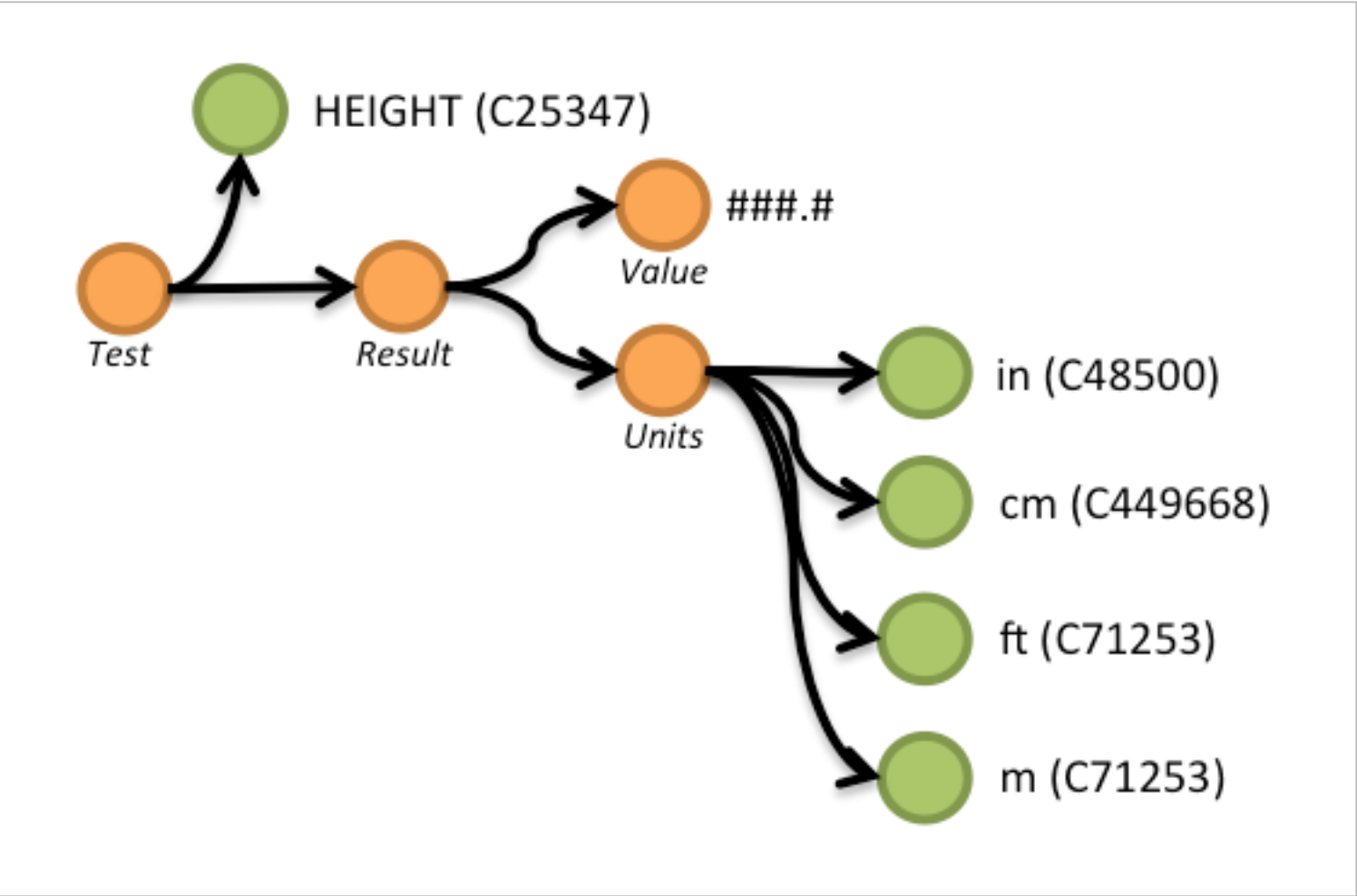
Walk the Graph



- Or in other directions or start points

- When VSTESTCD = HEIGHT
 - VSORRES has datatype="float", length="5", significant digits="1"
 - VSORRESU has datatype="text", length="5", code list="IN" or "cm"

Source of Metadata



- Better structure and relationships, not just individual variables 'floating'. A collection of variables with logical and consistent structure.
- Note the definition of value level metadata
- BCs derived from concept maps as seen in TA User Guides
- Can exist independent of SDTM & CDASH

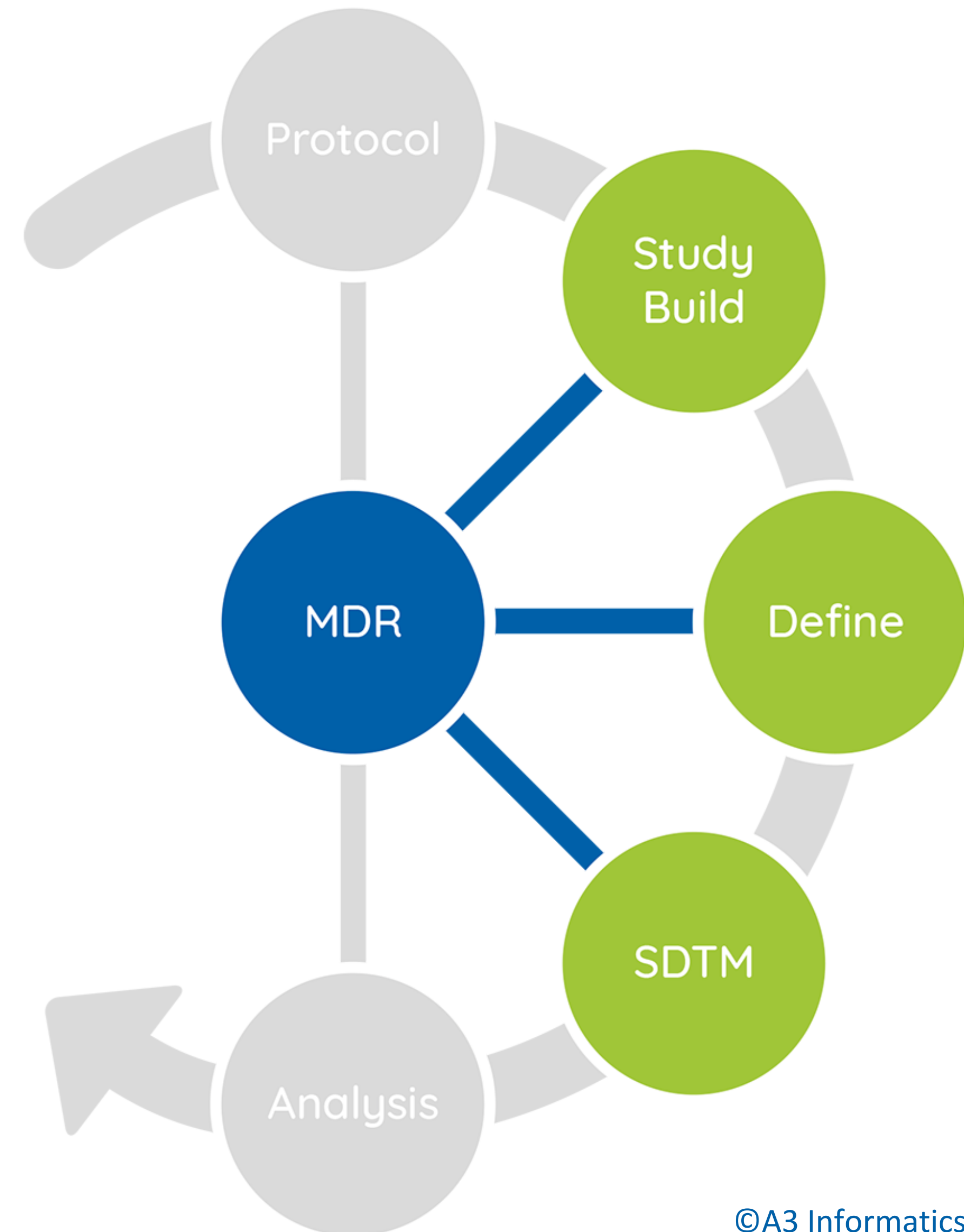
Date	Identifier	Submission Value
2014-03-28	C48500	IN
2014-06-27	↓	↓
2014-09-24	↓	↓
2014-10-06	↓	↓
2014-12-16	↓	IN in
2015-03-27	↓	↓

- Screen shot from the MDR. Noted the change in submission value so I checked the changes.
- December 2014 the submission value for Inch changed from 'IN' -> 'in'



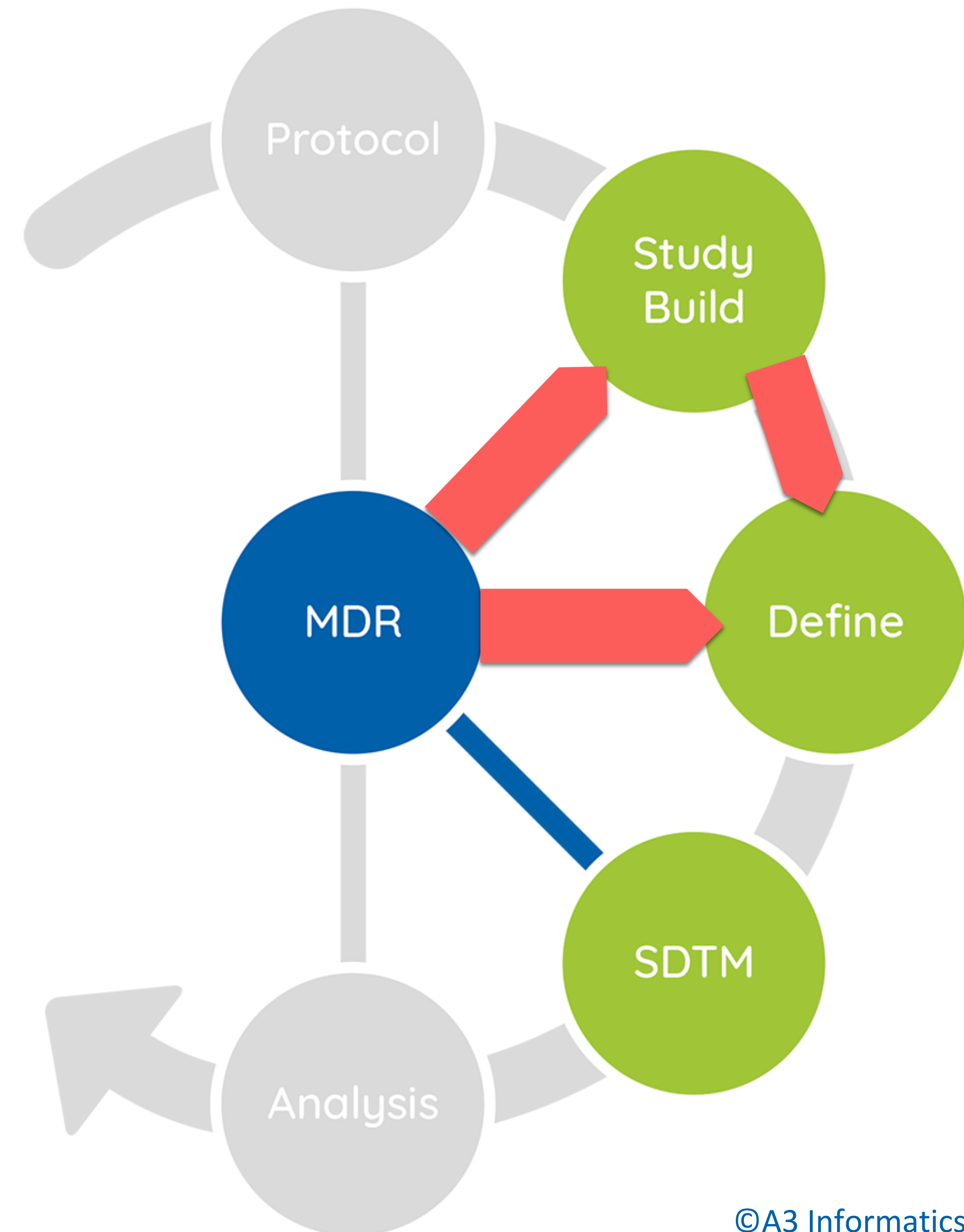
Co-operating Tools

- A central MDR containing definitions
 - Terminology
 - Biomedical Concepts
 - Forms
 - SDTM
- Study Build tool to define a study
- Define.xml tool to take study definition and build the define.xml file maximising automation where possible



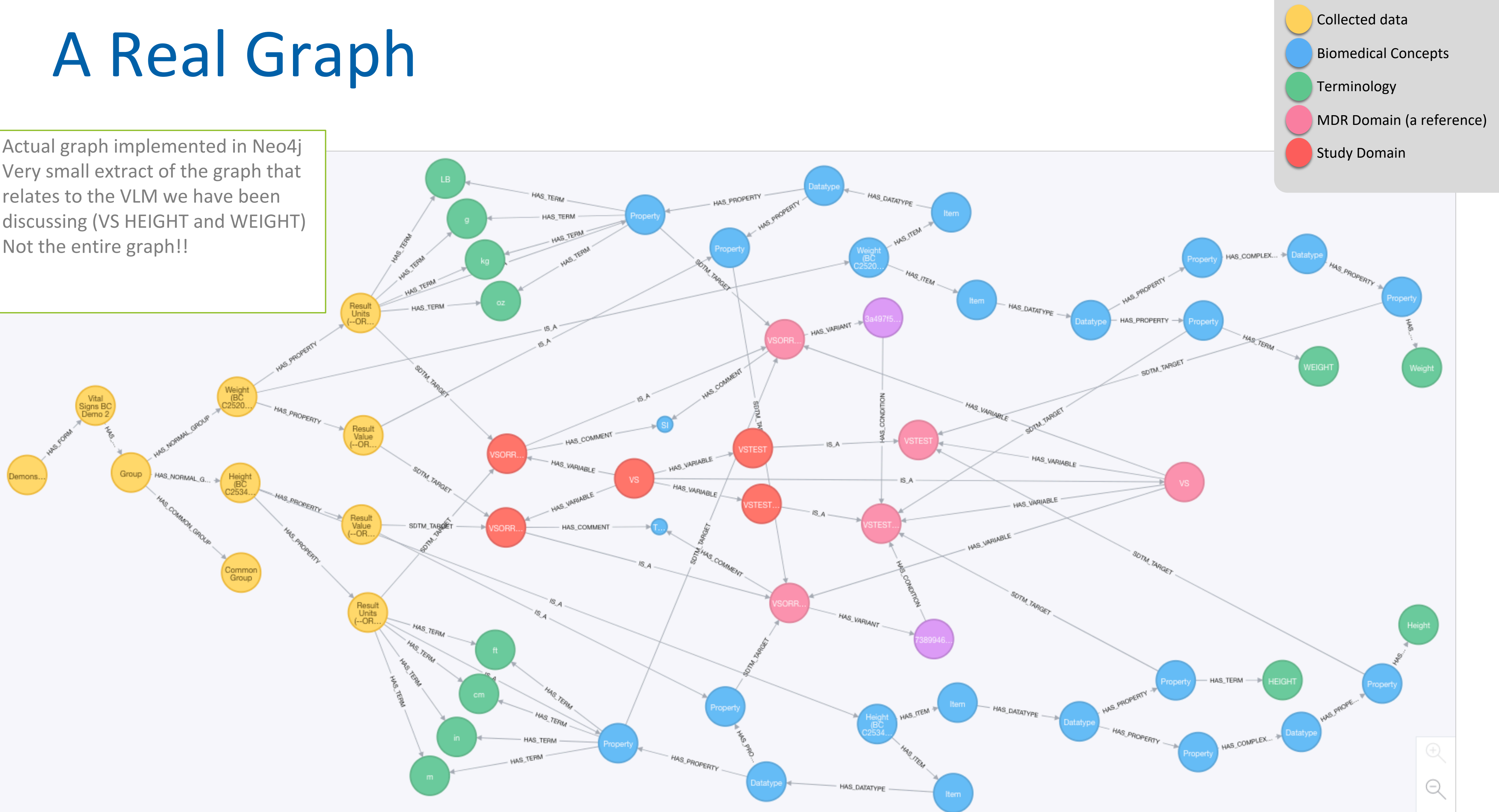
Application Programming Interface

- Tools use APIs to communicate
- Study Build sources definitions from the MDR
- Define tool sources definitions from the Study Build tool and the MDR

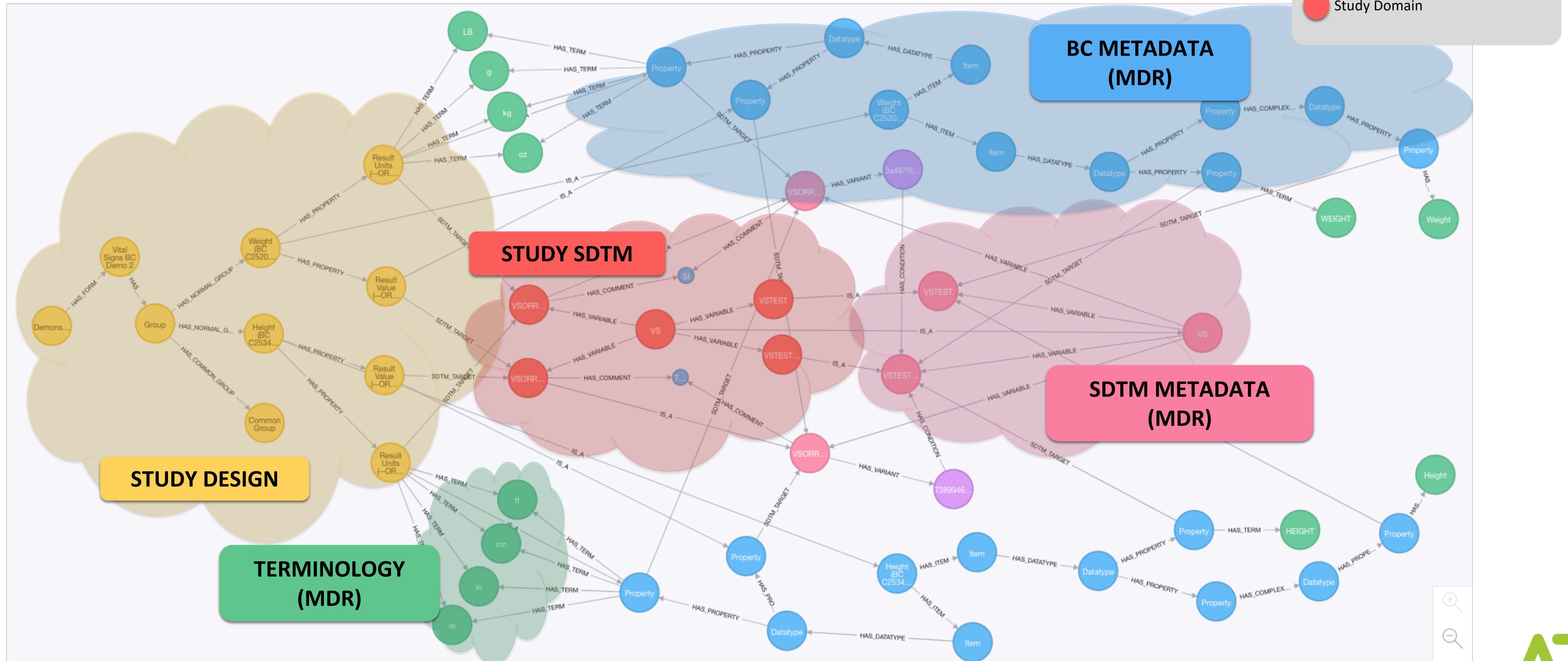


A Real Graph

- Actual graph implemented in Neo4j
- Very small extract of the graph that relates to the VLM we have been discussing (VS HEIGHT and WEIGHT)
- Not the entire graph!!

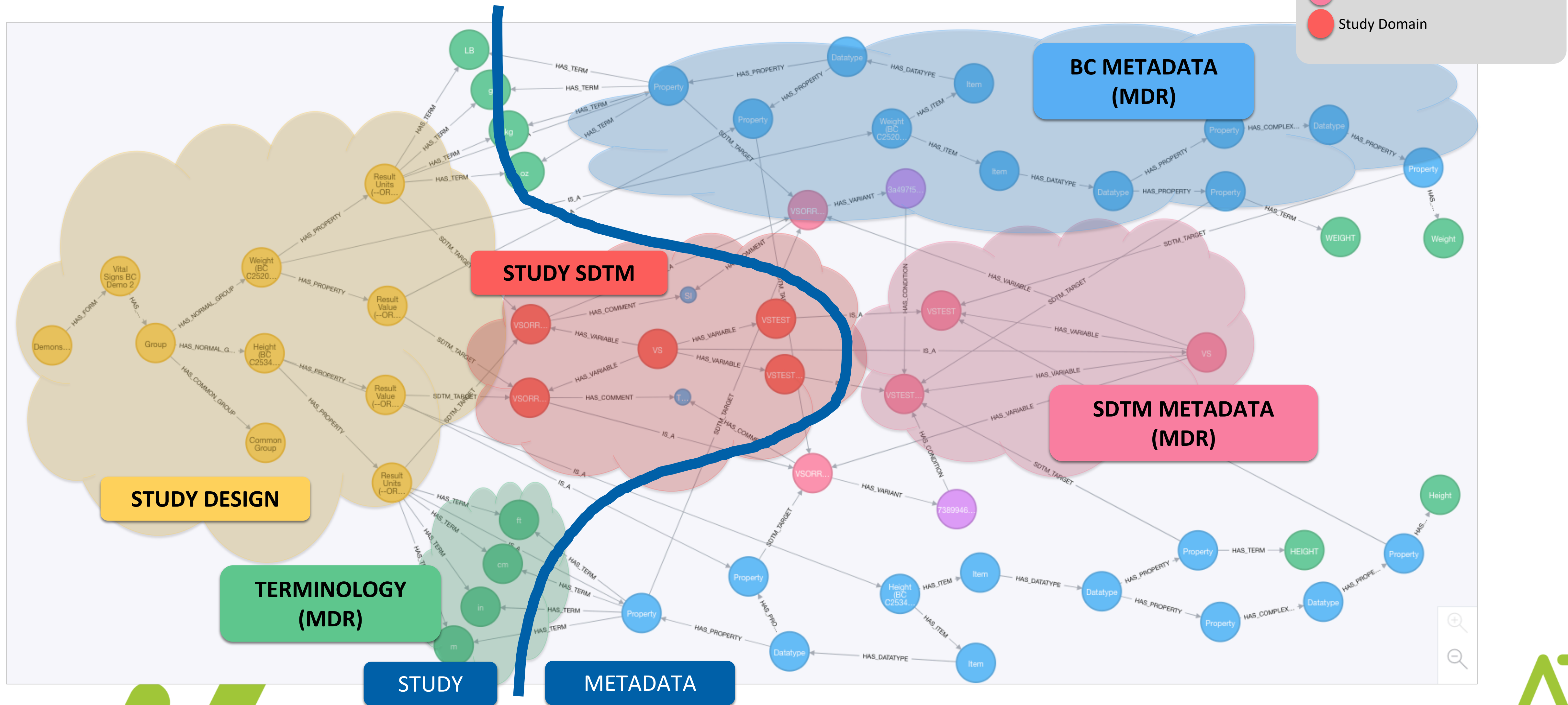


Views of the World



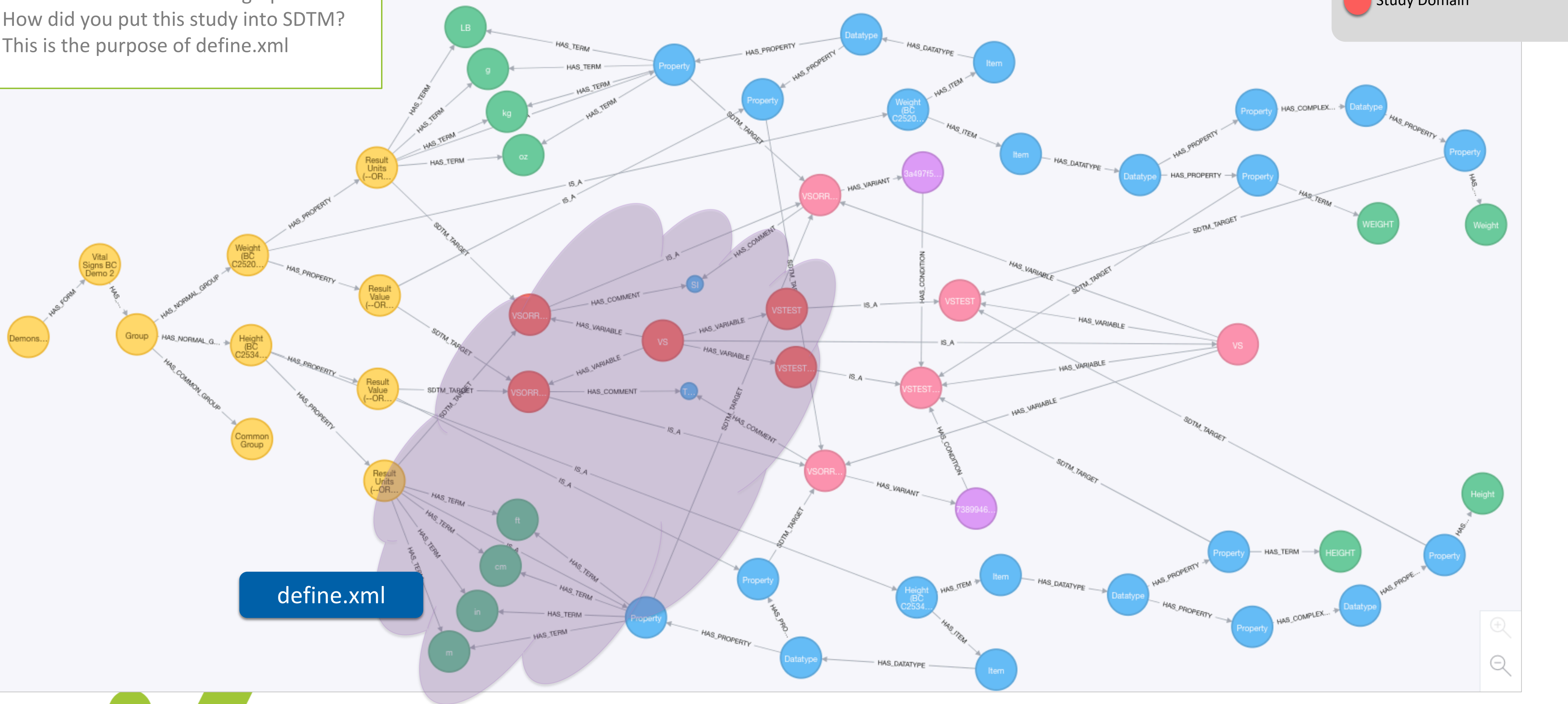
Views of the World

- Collected data
- Biomedical Concepts
- Terminology
- MDR Domain (a reference)
- Study Domain

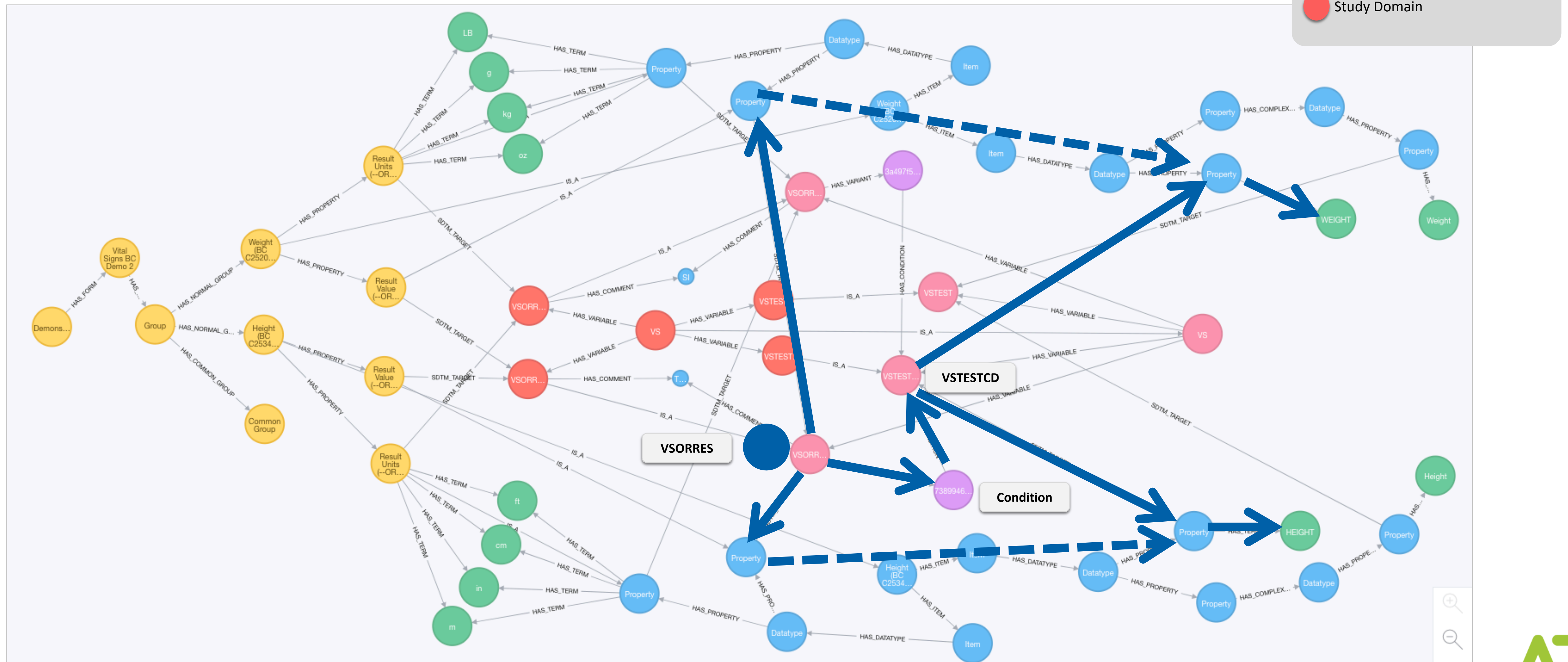


Views of the World

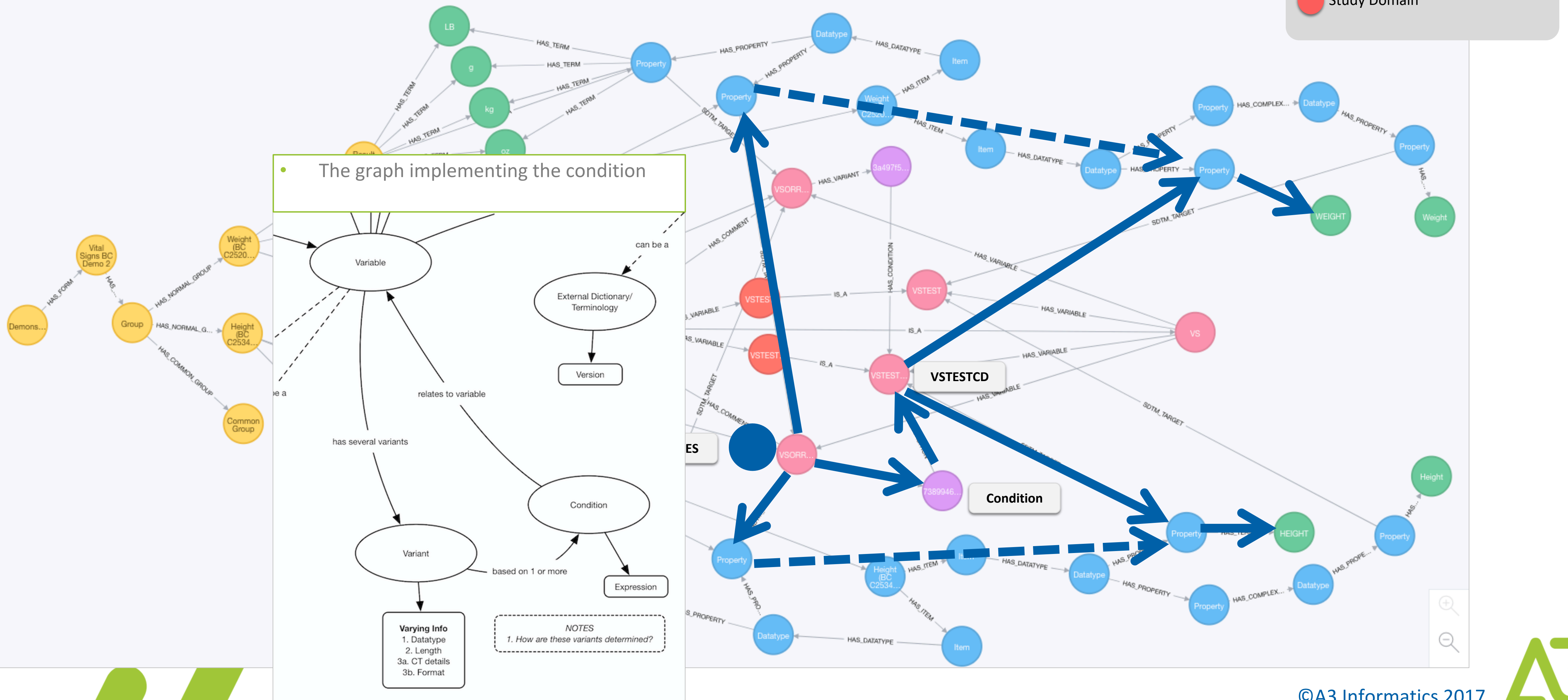
- Define.xml is a view of this graph
- How did you put this study into SDTM?
- This is the purpose of define.xml



Walk the Graph



Walk the Graph



User Views

Value Level Metadata		
Format	Terms	When
string 20	ft	VSTESTCD = HEIGHT
	m	
	in	
	cm	
string 20	oz	VSTESTCD = WEIGHT
	g	
	LB	
	kg	

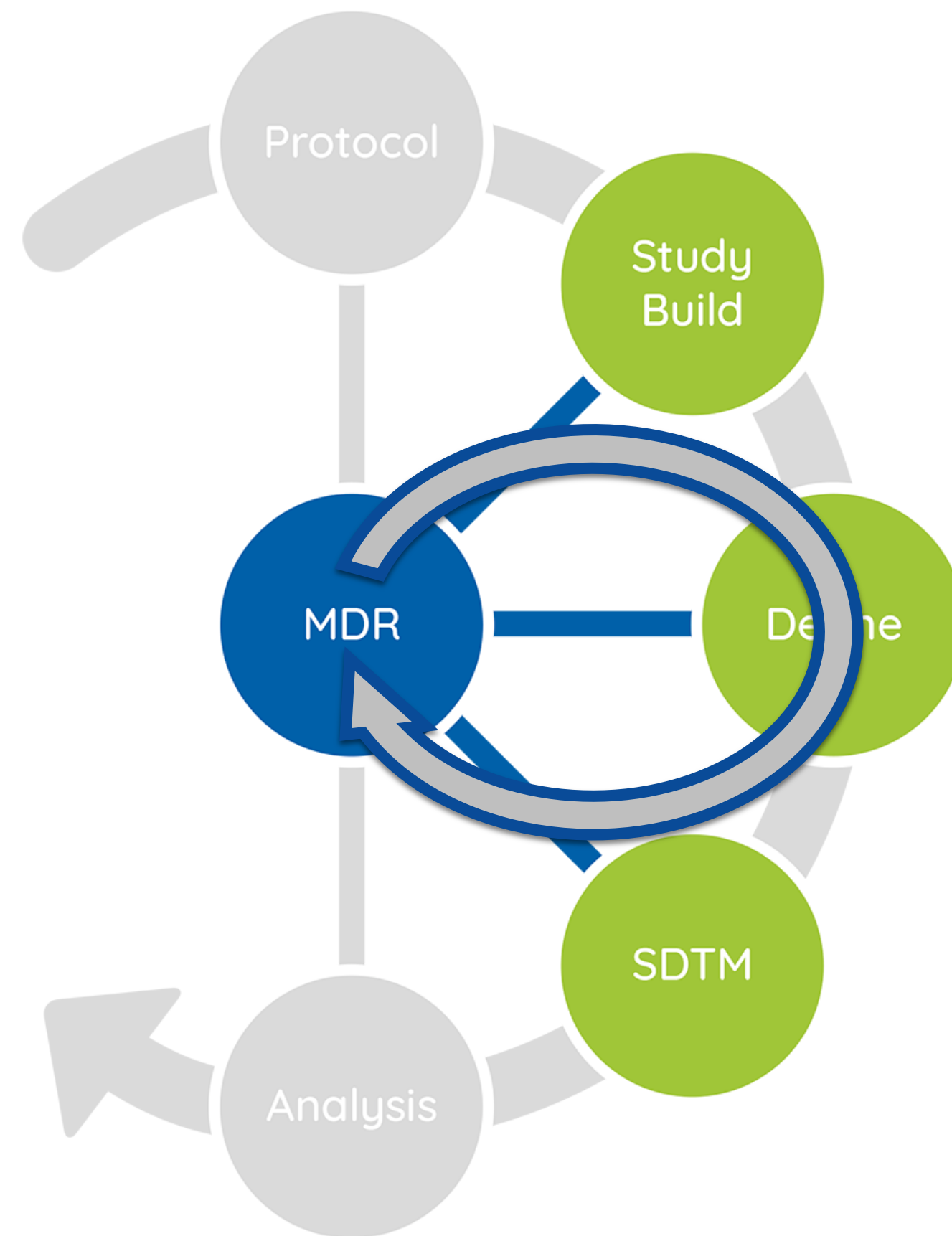
Terminology				
Identifier	Submission	Preferred Term	Synonym(s)	Definition
C71253	ft	International Foot	Foot	A unit of length defined by the U.S. National Bureau of Standards as 30.48 centimeters. It is equal to 0.3048 meter, 12 inches, or to approximately 0.999998 survey foot.(NCI)
C41139	m	Meter	Meter	A meter is defined as the length of the path traveled by light in a vacuum during a time interval of 1/299 792 458 of a second and is equal to 1.093 61 yards.(NCI)
C48500	in	Inch	Inch	A traditional unit of length equal to 2.54 centimeters. (NCI)
C49668	cm	Centimeter	Centimeter	A basic unit of length equal to one hundredth of a meter or approximately 0.393700787 inch.
C48519	oz	Ounce	Ounce	A unit of mass, the avoirdupois ounce is equal to 1/16 pound, or 28.3495 grams, or 0.911 457 troy ounce.(NCI)
C48155	g	Gram	Gram	A metric unit of mass equal to one one thousandth of a kilogram. (NCI)
C48531	LB	Pound	Pound	A traditional unit of mass. By international agreement, one avoirdupois pound is equal to exactly 0.453 592 37 kilogram, 16 ounces, or 1.215 28 troy pounds. (NCI)
C28252	kg	Kilogram	Kilogram	The basic SI unit of mass. It is defined as the mass of an international prototype in the form of a platinum-iridium cylinder kept at Sevres in France. It is the only basic unit still defined in terms of a material object, and also the only one with a prefix [kilo] already in place. A kilogram is equal to 1,000 grams and 2.204 622 6 pounds. (NCI)

- Sample presentation of the Value Level Metadata
- Compared to the standard stylesheet presentation
- Same information, different presentation.

Value Level Metadata - VS [VSORRESU]						
Variable	Where	Type	Length or Display Format	Controlled Terms or Format	Origin	Derivation / Comment
VSORRESU	VSTESTCD = "HEIGHT" (Height)	text	5	Unit (UH) • "IN" = "Inch" • "cm" = "Centimeter"	[CRF] Annotated Case Report Form 11	
VSORRESU	VSTESTCD = "WEIGHT" (Weight)	text	4	Unit (UW) • "LB" = "Pound" • "kg" = "Kilogram"	[CRF] Annotated Case Report Form 11	



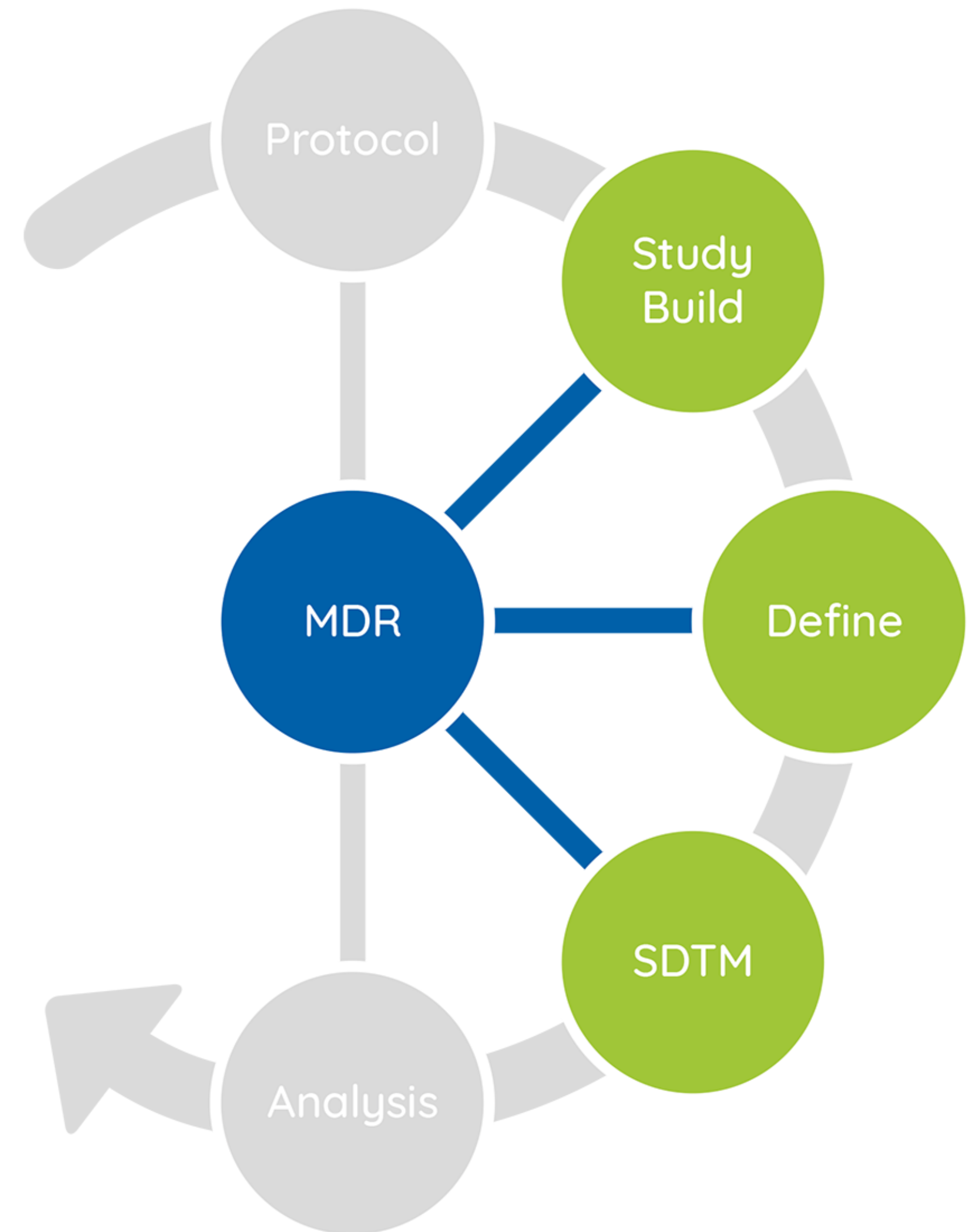
Aid the User



- Use the MDR to assist the user when study build information is lacking
- Also can be used to check define.xml terminology subsets are sensible
- When information has been created why waste it?
- Possible for the metadata created during the define.xml process can be feed back to the MDR to allow for better studies to be built in the future. Build a feedback loop.

Summary

- Graphs aids tool construction
- Our world is not rectangular
- High quality definitions can assist in constructing better define.xml files
- While designed for define.xml the tool is more, actually detailed study metadata, could be used for other purposes. Those other uses are just different queries.
- And the data? ...





Contact Information

Email:

dih@A3Informatics.com

More information at:

www.A3Informatics.com

