Linked Longitudinal Medical Record

Susie Stephens

Co-chair W3C Health Care & Life Science Interest Group

Outline

- Secondary Use of Health Care Data
- Introduction to the Semantic Web
- Translational Medicine Ontology

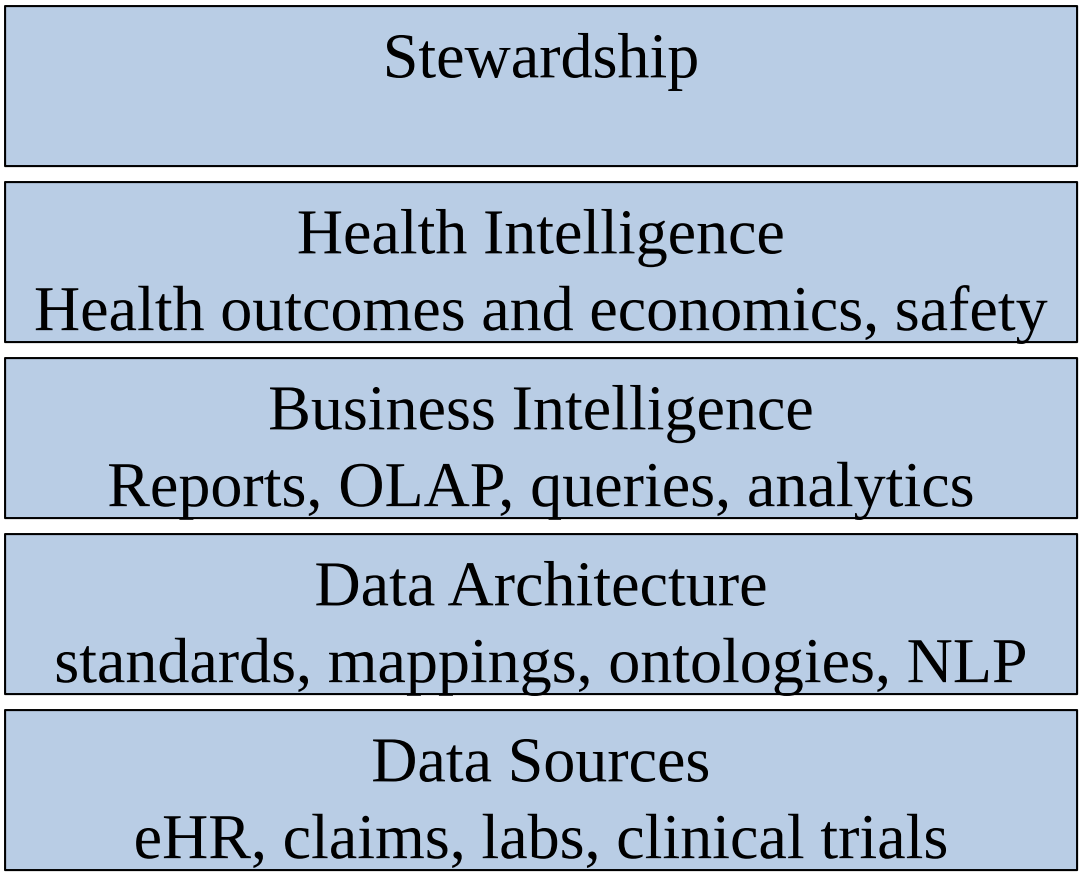
Pharma Uses for De-identified Data

- Improve understanding of disease
- Fine tune patient recruitment in clinical trials
- Examine treatment effectiveness, pharmacoeconomics and pharmacogenomics
- Track and understand drug safety
- Support formulary decisions
- Understand prescribing behavior

Hazards in Data Mining

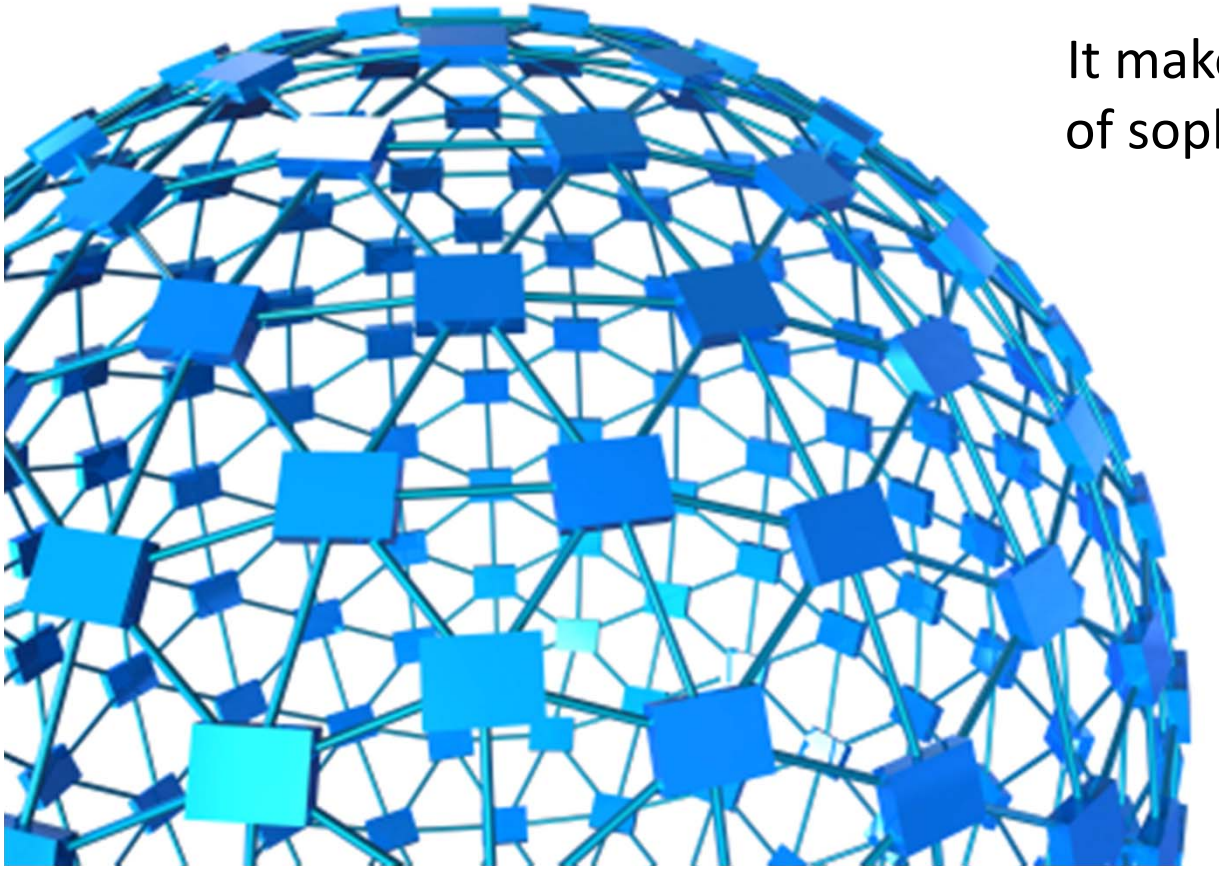
- Make enough comparisons and something will be significant
- Data is messy and in places inaccurate
- Difficult to know if someone is taking their medicine correctly
- Difficult to know if someone has died
- Lacking individuals with the necessary skill sets

Data Integration Framework



The Semantic Web is the New Global Web of Knowledge

It is about standards for publishing, sharing and querying knowledge drawn from diverse sources



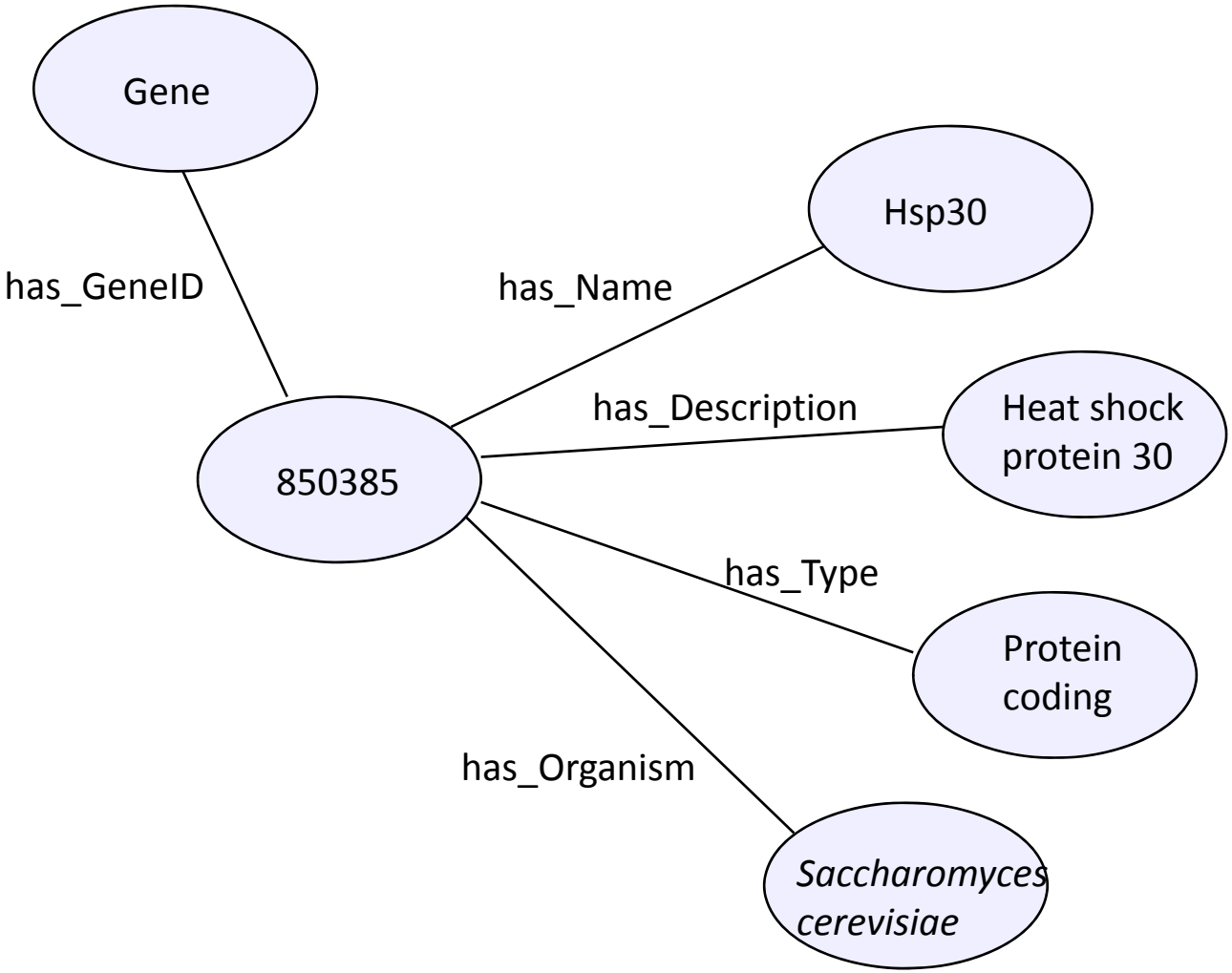
It makes possible the answering of sophisticated questions using background knowledge

W3C Semantic Web

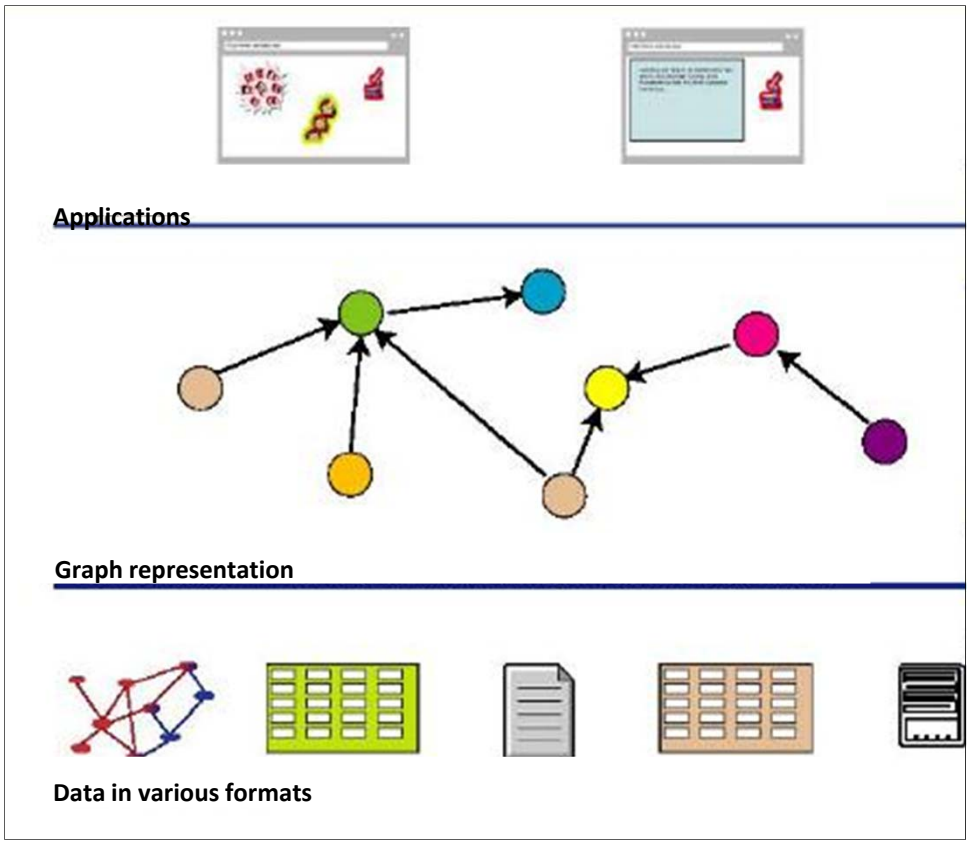
Semantic Web Standards

- RDF – store data as triples
- OWL – define systems of concepts called ontologies
- SPARQL – query data in RDF
- RIF – define rules
- GRDDL – transform data to RDF

RDF Triples


Gene
has_GeneID
850385
has_Name
Hsp30
has_Description
Heat shock protein 30
has_Type
Protein coding
has_Organism
Saccharomyces cerevisiae

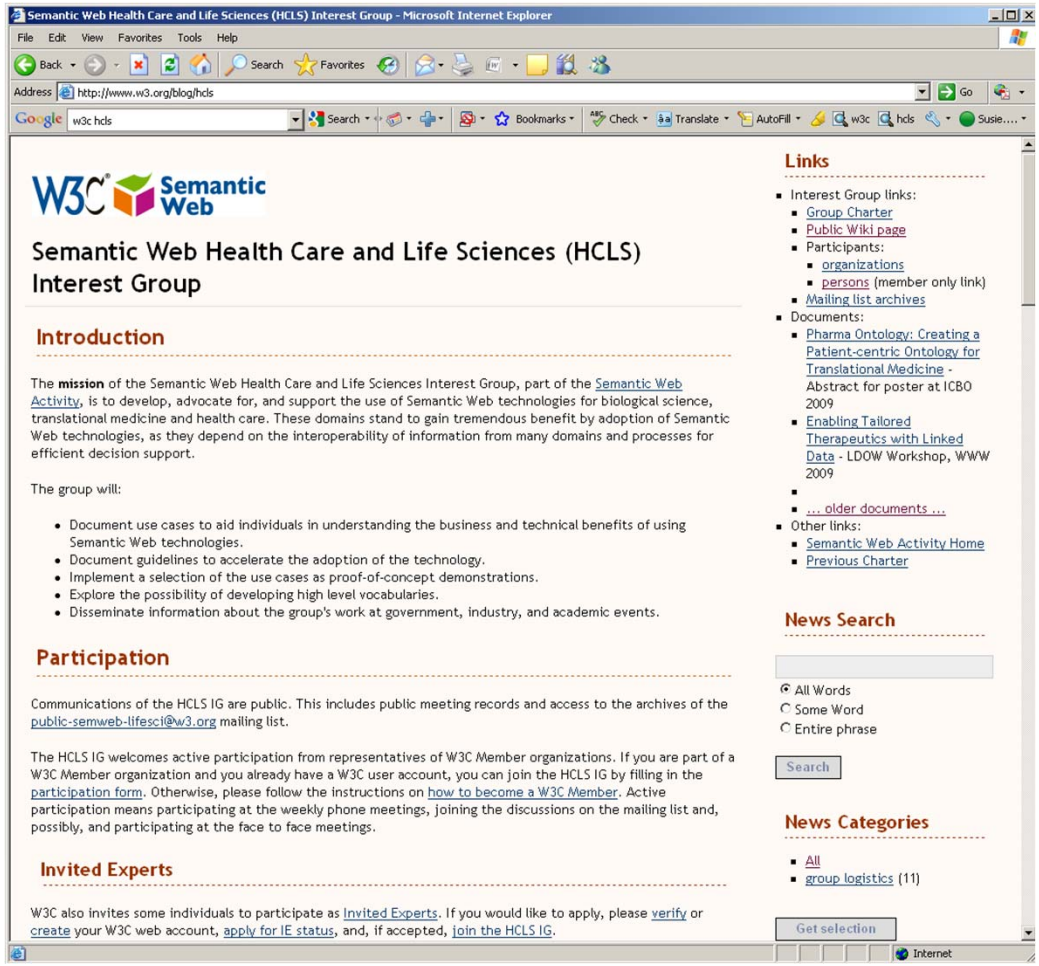
Creating a Web of Data



Advantages of Semantic Web

- Data will contain more meaning
  - Definitions, policies and permissions, explicit relationships, self describing (data = metadata)
- Multiple layers of abstraction
- Fine grained – ‘atomic’ data
- Easy to share and link
- Potential network effect
- Generic, dumber applications
- Greater level of future-proofing and reuse

W3C's HCLS Interest Group



Participants

- Bosse Andersson, AstraZeneca
- Colin Batchelor, RSC
- Olivier Bodenreider, NIH
- Tim Clark, HMS
- Christi Denney, Eli Lilly
- Christopher Domarew, Albany Medical Center
- Michel Dumontier, Carleton University
- Thomas Gambet, W3C
- Lee Harland, Pfizer
- Anja Jentzsch, Free University Berlin
- Vipul Kashyap, Cigna
- Peter Kos, HMS
- Julia Kozlovsky, AstraZeneca
- Timothy Lebo, RPI
- Joanne Luciano, RPI
- Scott Marshall, Leiden University Medical Center
- Jim McCusker, RPI
- Deborah McGuiness, RPI
- Jim McGurk, Daiichi Sankyo
- Chimezie Ogbuji, Cleveland Clinic
- Elgar Pichler, AstraZeneca
- Bob Powers, Predictive Medicine
- Eric Prud'hommeaux, W3C
- Matthias Samwald, DERI
- Lynn Schriml, University of Maryland
- Susie Stephens, Johnson & Johnson Pharmaceutical R&D
- Peter Tonellato, HMS
- Trish Whetzel, Stanford
- Jun Zhao, Oxford University

Personalized Medicine

- The end of the blockbuster era
- The right dose to the right patient at the right time
- Lowering cost of pharmaceutical development
- Improved patient response to medication

Data Challenges

- Patient data split across eHRs, clinical trial systems, genetic testing vendors, and longitudinal studies
- Drug information split across systems such as the Orange Book, DrugBank, ClinicalTrials.gov, DailyMed, SIDER, PharmGKB, formulary lists
- Disease information split across OMIM, GEO, commercial databases
- Different data representation approaches used by different communities
- No unifying schema to pull data together

The Translational Medicine Ontology

- Provides a global schema for the integration of patient, drug and disease centric data sets
- Establishes accurate mappings to relevant ontologies
- Translational Medicine knowledge base demo

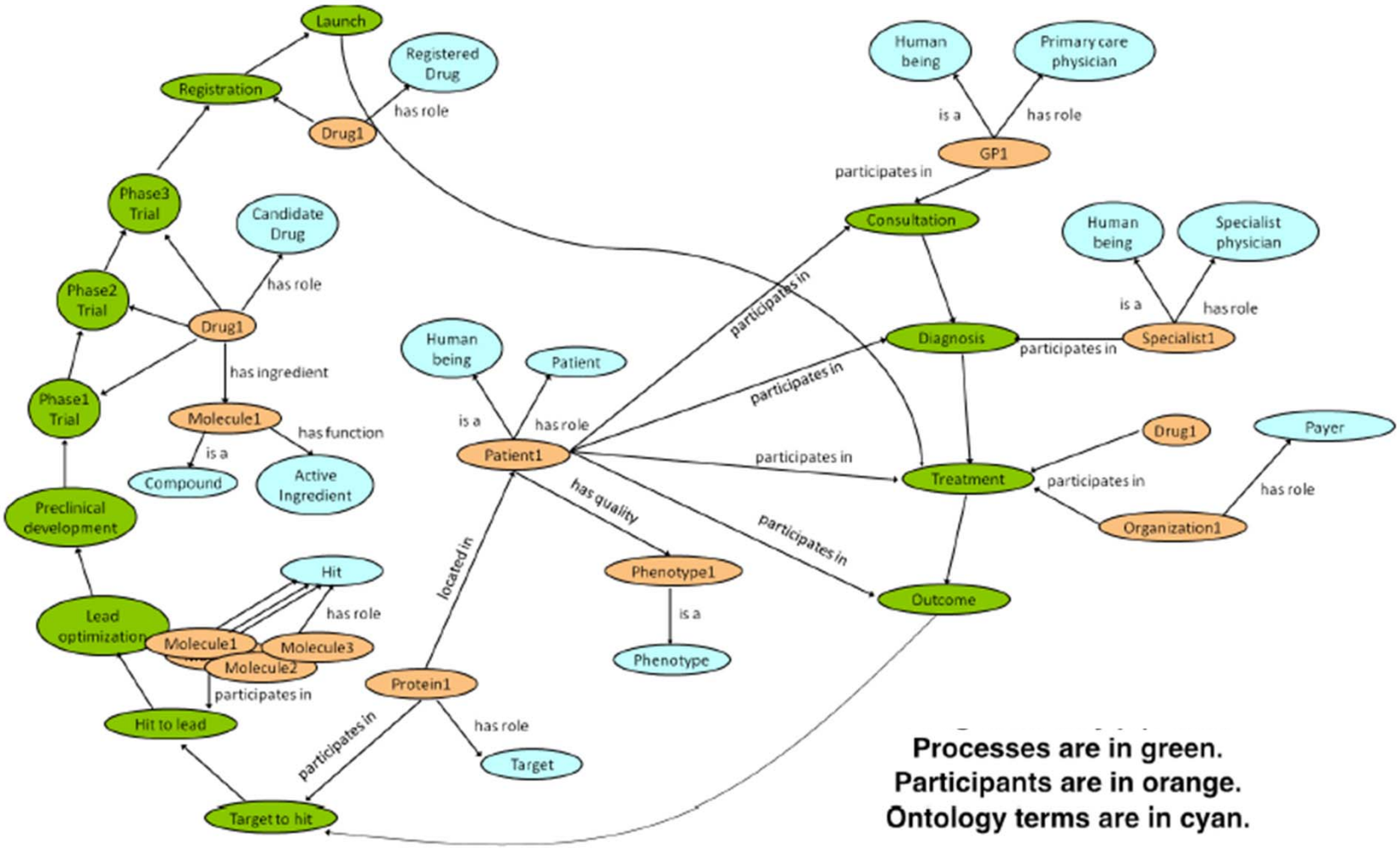
TMO Approach

- Undertake extensive user-focused requirements
- Identify key entities and establish their relations
- Extend the conceptualization specified by a foundational ontology
- Map linked data types to ontological entities
- Develop knowledge base containing ontology and mappings and data
- Demonstrate query answering over TMKB

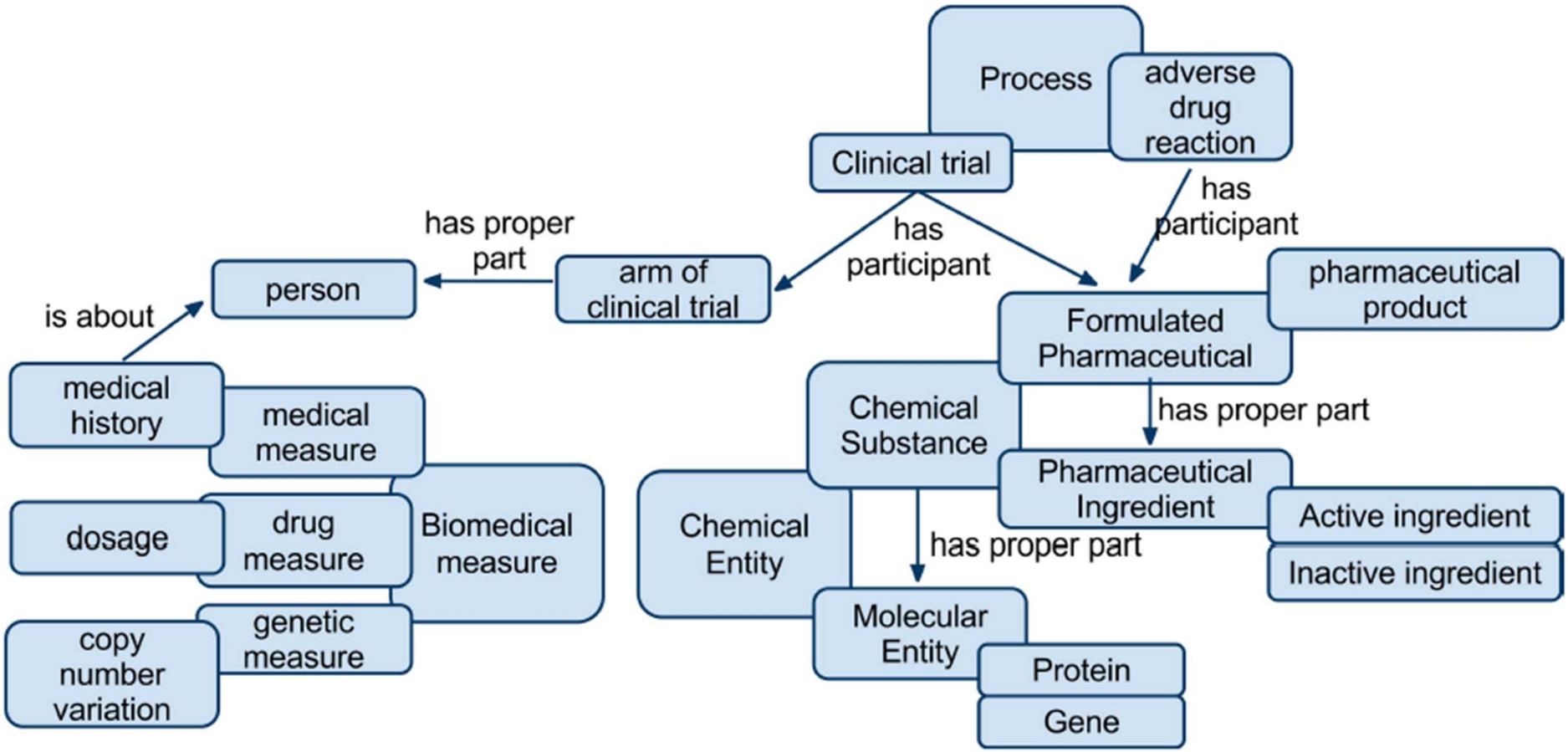
TMO Structure

- Extends BFO with 80 classes spanning material entities, processes, roles, informational entities
- Distinction among different kinds of material entities
  - molecular entities versus chemical substances
  - active ingredients versus pharmaceutical formulations
- Distinction among different kinds of informational entities
  - Genetic, medical, drug

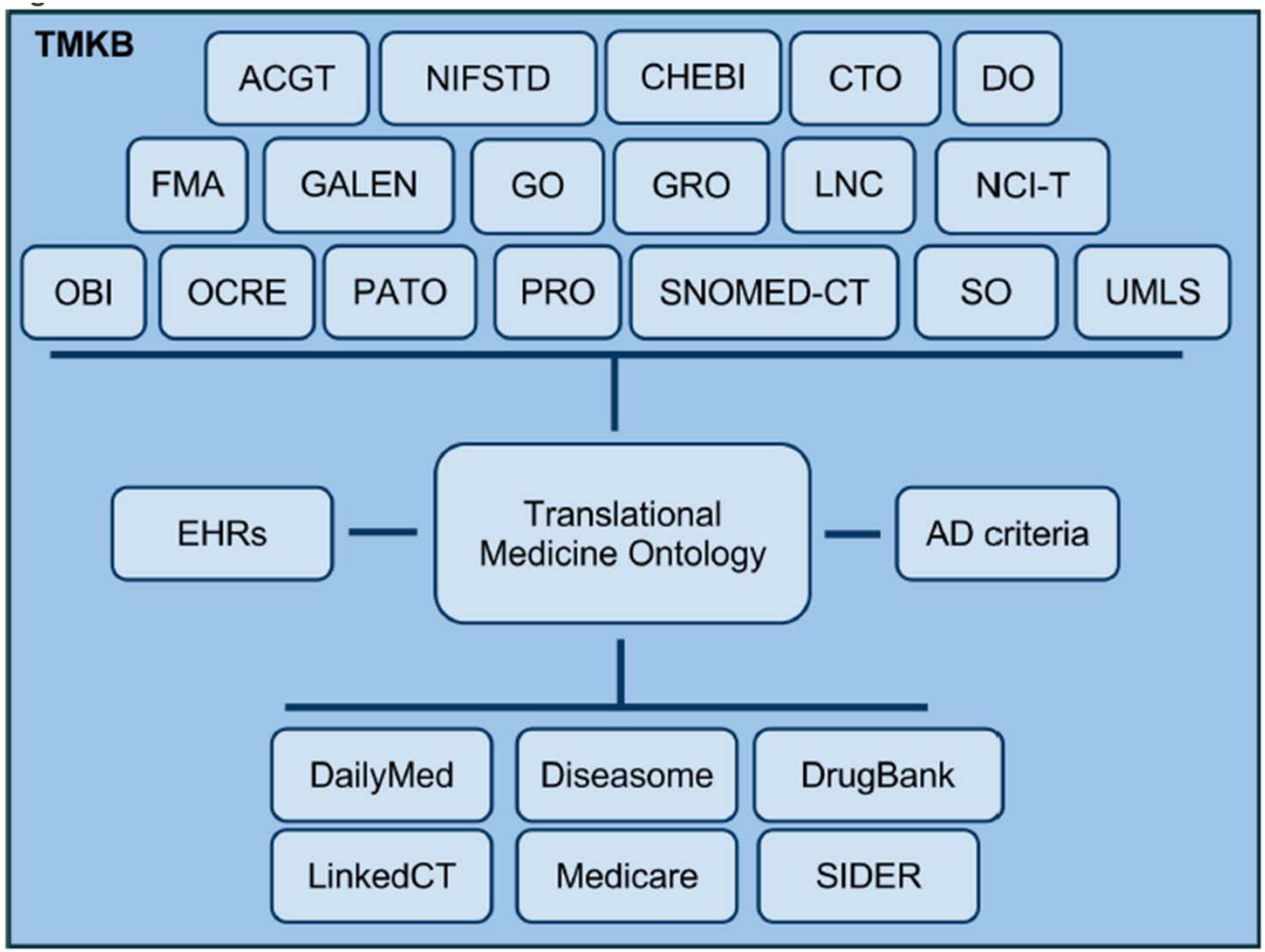
Scope of the TMO



TMO Structure



Translational Medicine KB



Linked Longitudinal Medical Record

Discovery Questions and Answers

What genes are associated with or implicated in AD?	Diseasome and PharmGKB indicate at least 97 genes have some association with AD.
Which SNPs may be potential AD biomarkers?	PharmGKB reveals 63 SNPs.
Which market drugs might potentially be repurposed for AD because they modulate AD implicated genes?	57 compounds or classes of compounds are used to treat 45 diseases, including AD, diabetes, obesity, and hyper/hypotension

Clinical Trials Questions and Answers

Since my patient is suffering from drug-induced side effects for AD treatment, can an AD clinical trial with a different mechanism of action be identified?	Of the 438 drugs linked to AD trials, only 58 are in active trials and only 2 (Doxorubicin and IL-2) have a documented mechanism of action. 78 AD-associated drugs have an established MOA.
Find AD patients without the APOE4 allele as these would be good candidates for the clinical trial involving Bapineuzumab?	Of the 4 patients with AD, only one does not carry the APOE4 allele, and may be a good candidate for the clinical trial.
What active trials are ongoing that would be a good fit for Patient 2?	58 Alzheimer trials, 2 mild cognitive impairment trials, 1 hypercholesterolaemia trial, 66 myocardial infarction trials, 46 anxiety trials, and 126 depression trials.

Physician Questions and Answers

What are the diagnostic criteria for AD?	There are 12 diagnostic inclusion criteria and 9 exclusion criteria
Does Medicare D cover Dopenezil?	Medicare D covers two brand name formulations of Donepezil: Aricept and Aricept ODT.
Have any AD patients been treated for other neurological conditions?	Patient 2 was found to suffer from AD and depression.

Insights from the Pilot

- Physicians have similar data requirements to pharma
- The data landscape for physicians is highly fragmented
- eHRs don’t capture a lot of data - but not all data should be stored in the eHR
- Pharmacogenomics rules need to be centralized
- Hundreds of vendors and platforms supporting different user communities
- Existing technology is typically very old
- Much data is not structured for analytical purposes
- Strong need for innovation

Summary

- W3C has strong interest in solutions for health care and life sciences
- Secondary use of data has significant value to many stakeholders
- New data architectures are needed
- New data mining and visualization capabilities are required
- Lack individuals with the necessary skills sets