

# Powering Translational Medicine with Semantic Web technologies

**Michel Dumontier, Ph.D.**

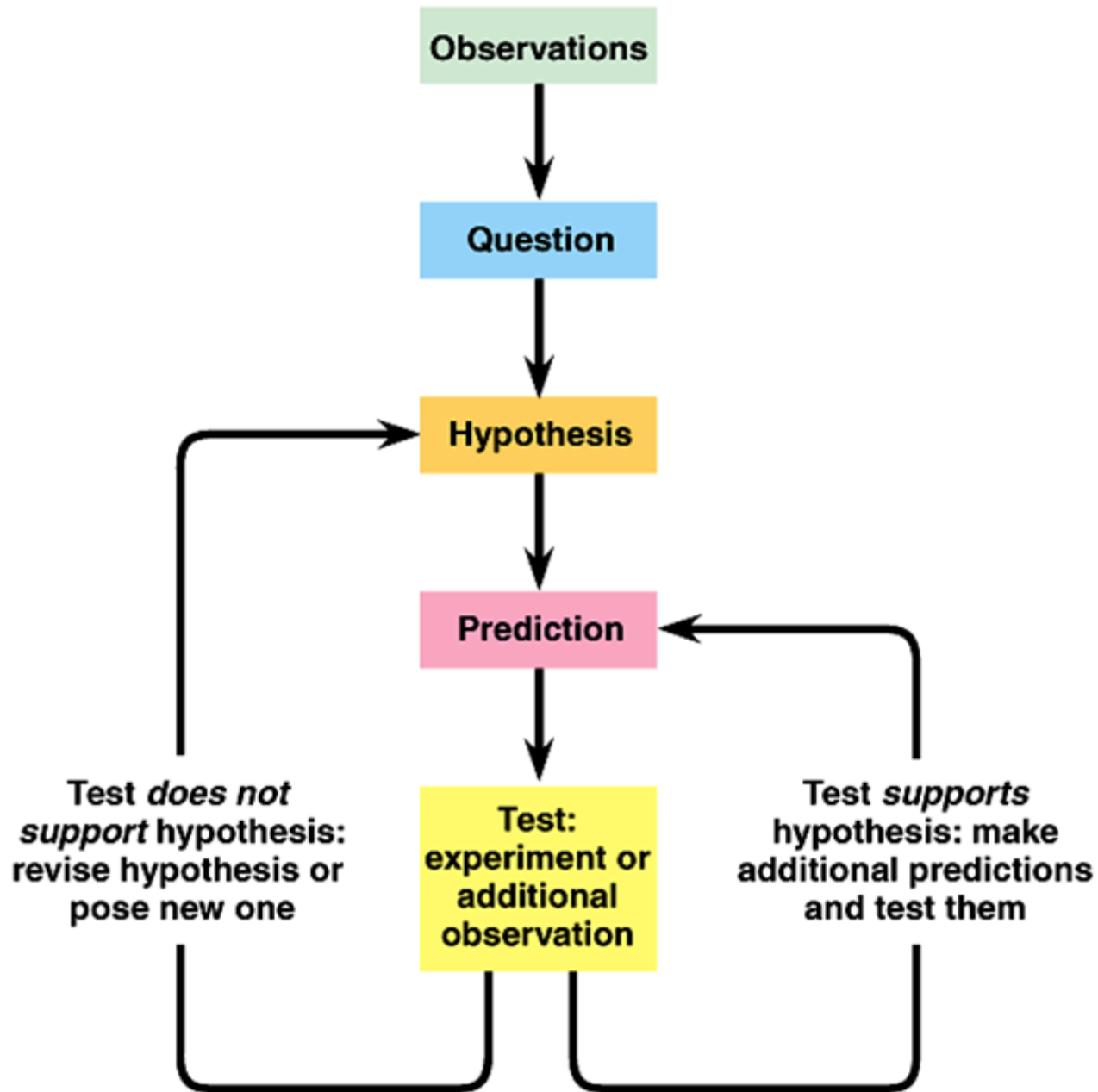
**Associate Professor of Medicine (Biomedical Informatics)  
Stanford University**



A photograph of a person's legs and feet walking on a sandy beach. The person is walking away from the camera towards the ocean. The sand is light brown and shows several footprints. The ocean waves are white and foamy, crashing onto the shore. The sky is not visible.

# Outline

- **reproducible science**
- **linked data for the life sciences**
- **the semantic clinical data warehouse**
- **integrated translational research**





**Scientists need to find evidence to support/refute a hypothesis which is, surprisingly, increasingly challenging with more data**

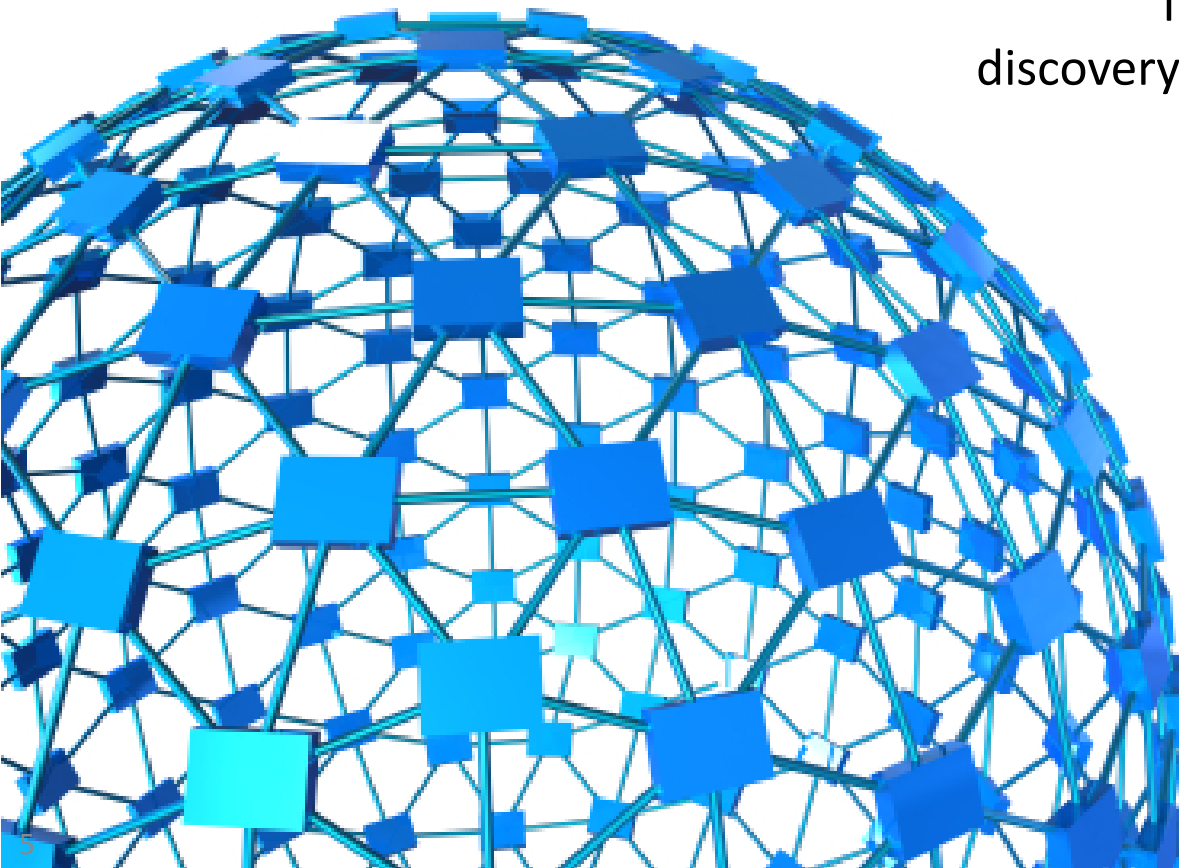
*need to know where to look,  
understand the nature  
and structure of data  
and how to process it*



# The Semantic Web is the new global **web of knowledge**

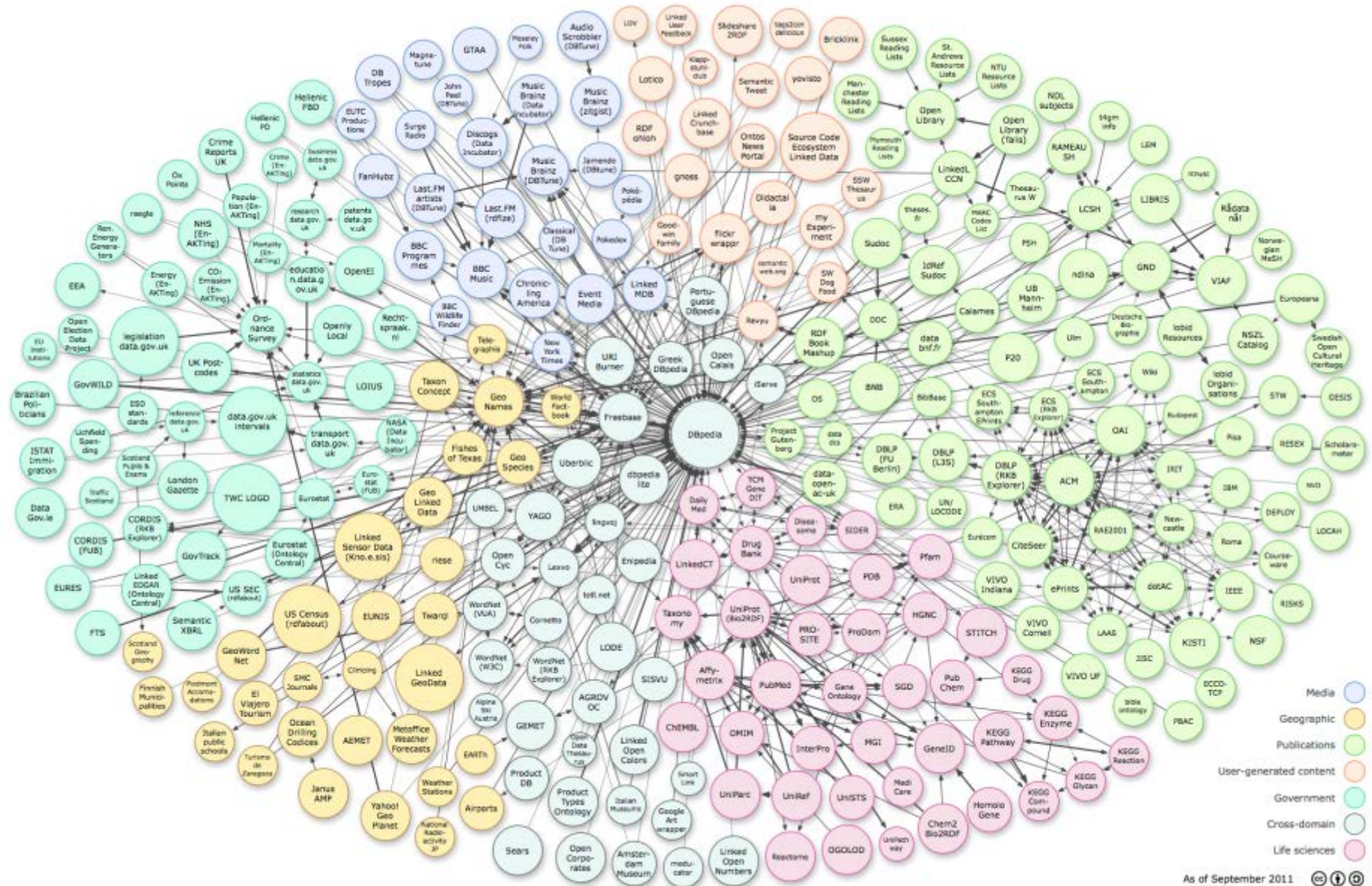
It involves standards for publishing, sharing and querying  
**facts, expert knowledge and services**

It is a scalable approach to the  
discovery of *independently formulated*  
and *distributed* knowledge





# We are building a massive network of linked open data

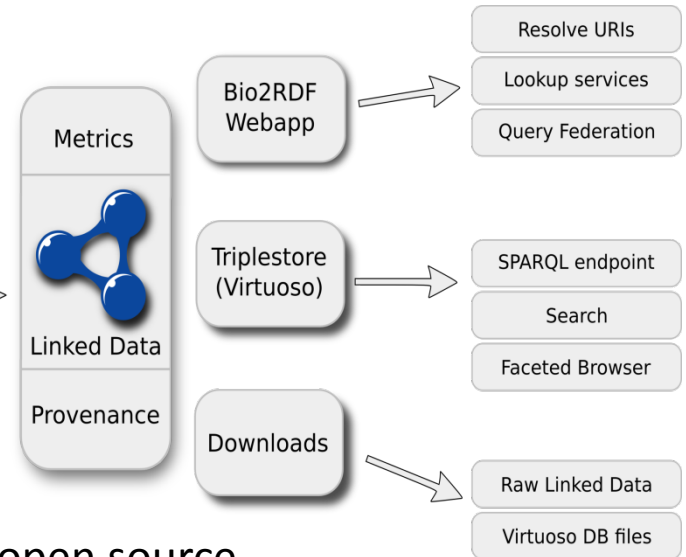


## Linked Data for the Life Sciences

A diagram showing four data formats represented by colored shapes: a blue square labeled 'TAB', a yellow square labeled 'CSV', a green square labeled 'XML', and a red cylinder labeled 'DB'.

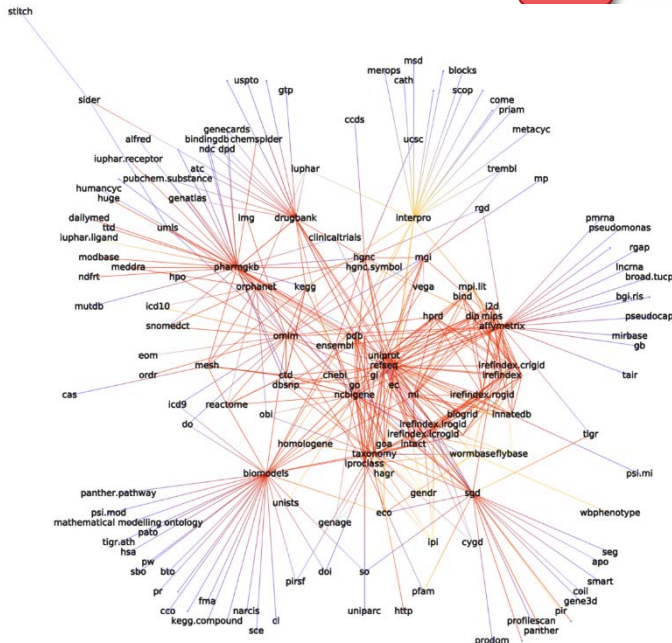


## Conversion Scripts



- Free and open source
- Leverages Semantic Web standards
- **10B+** interlinked statements from **30+** conventional and high value datasets
- Partnerships with EBI, SIB, NCBI, DBCLS, NCBO, OpenPHACTS, and many others

Alison Callahan, Jose Cruz-Toledo, Peter Ansell, Michel Dumontier:  
Bio2RDF Release 2: Improved Coverage, Interoperability and  
Provenance of Life Science Linked Data. ESWC 2013: 200-212





HyQue is the Hypothesis query and evaluation system

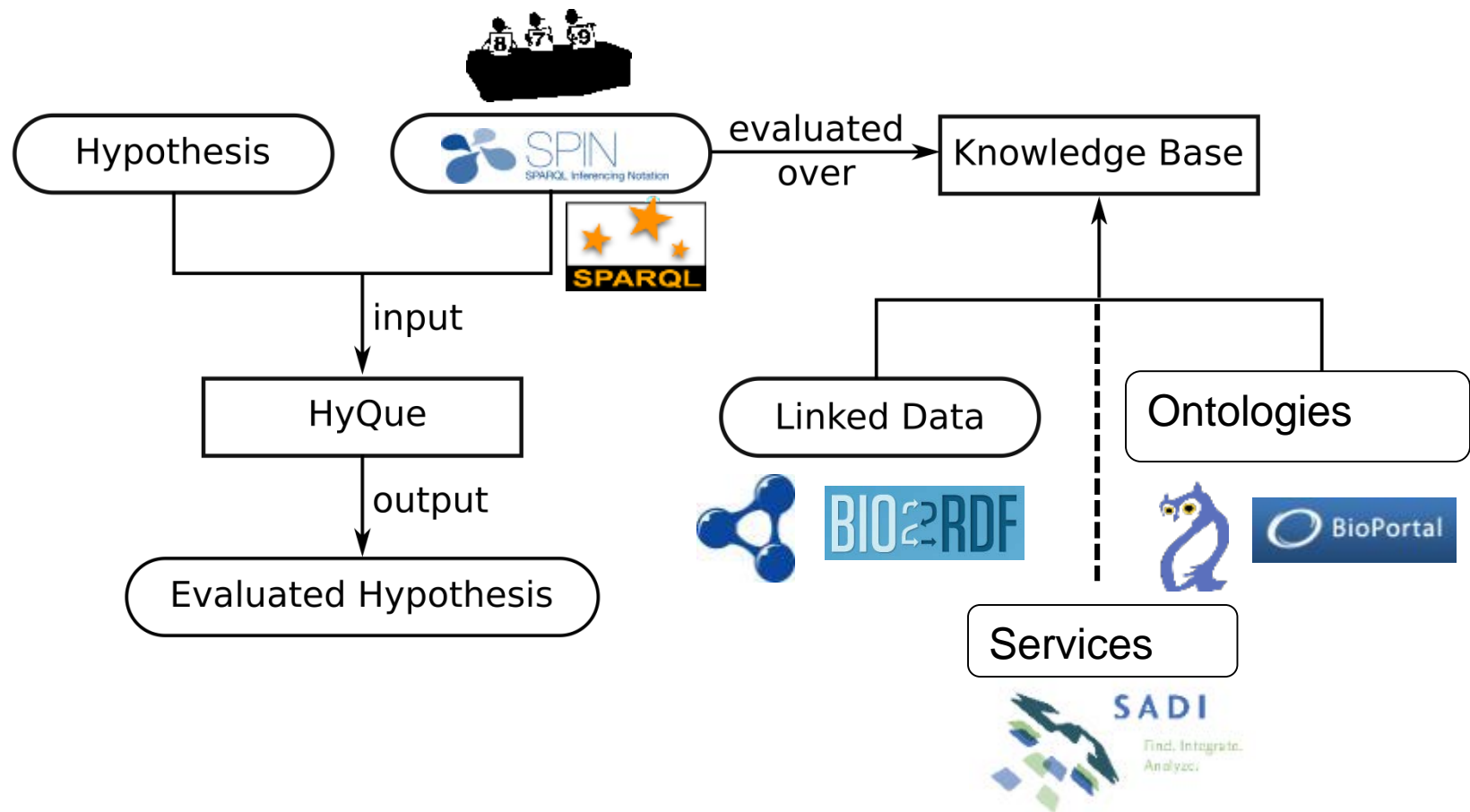
- A platform for **knowledge discovery**
- Facilitates **hypothesis formulation** and **evaluation**
- Leverages **Semantic Web technologies** to provide access to facts, expert knowledge and web services
- Pervasive **Provenance**
- Reproducible evaluation against **positive** and **negative findings**
- Transparent **evidence weighting**

HyQue: evaluating hypotheses using Semantic Web technologies. J Biomed Semantics. 2011 May 17;2 Suppl 2:S3.

Evaluating scientific hypotheses using the SPARQL Inferencing Notation. Extended Semantic Web Conference (ESWC 2012). Heraklion, Crete. May 27-31, 2012.



# HyQue is a Semantic Web Application that uses RDF, OWL, SPARQL, SPIN, and SADI

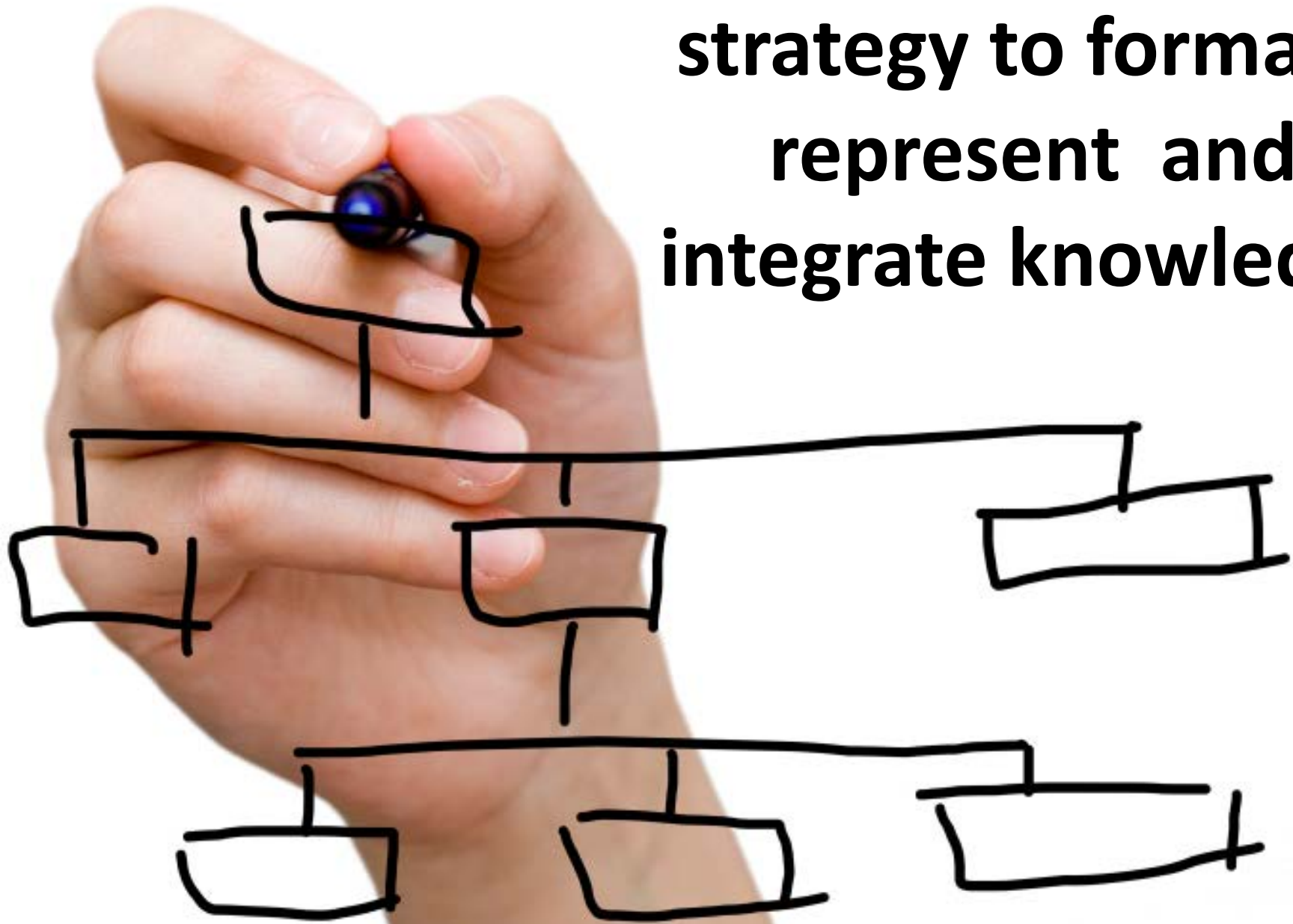


# Translational Research



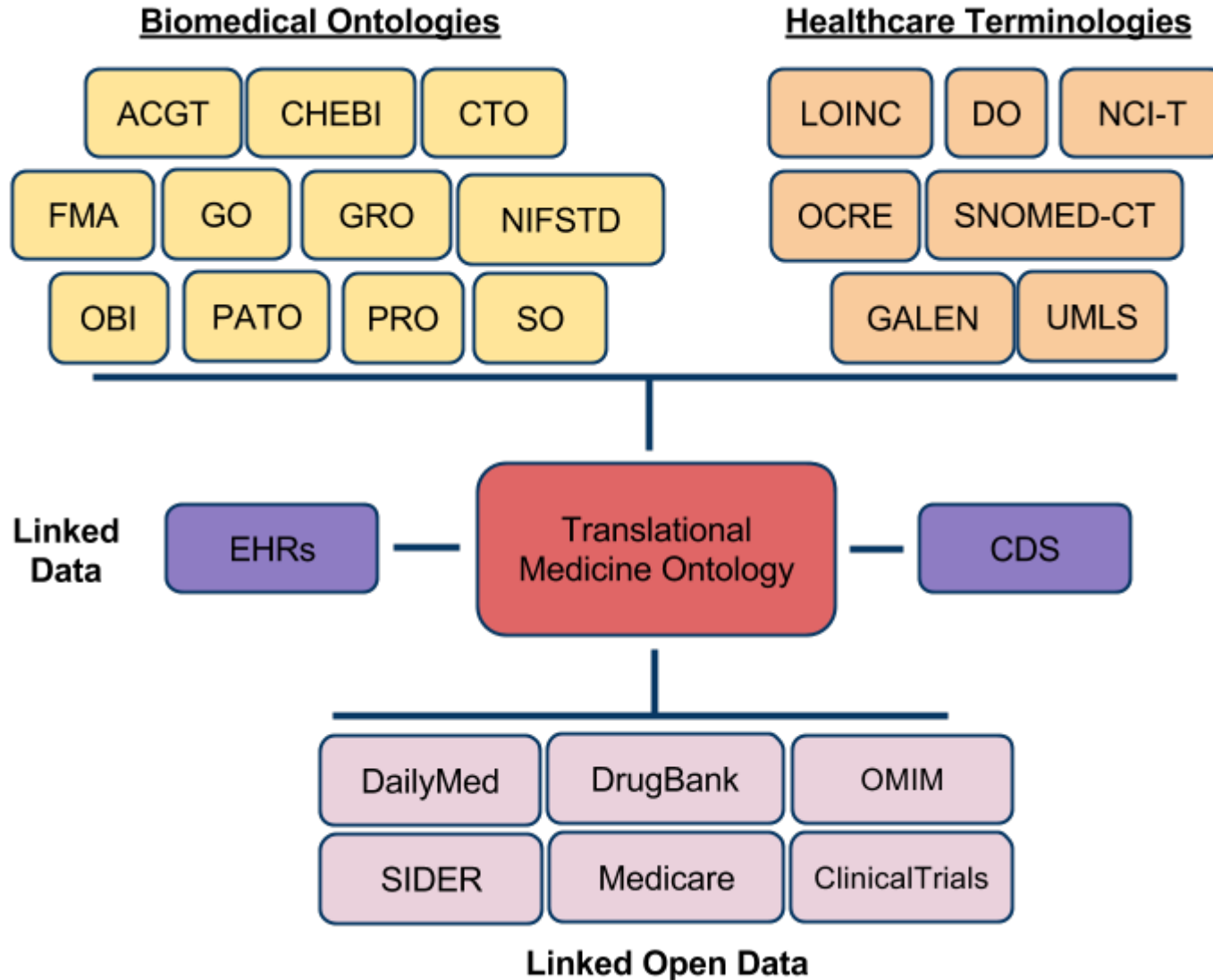
Using a *Semantic* Clinical Data Warehouse

**ontology as a  
strategy to formally  
represent and  
integrate knowledge**





# Semantic data integration through ontological mappings



# Applications in biomedical and clinical research

## Pharmaceutical Research

- Which existing marketed drugs might potentially be **re-purposed** for AD because they are known to modulate genes that are implicated in the disease?
  - *57 compounds or classes of compounds that are used to treat 45 diseases, including AD, hyper/hypotension, diabetes and obesity*

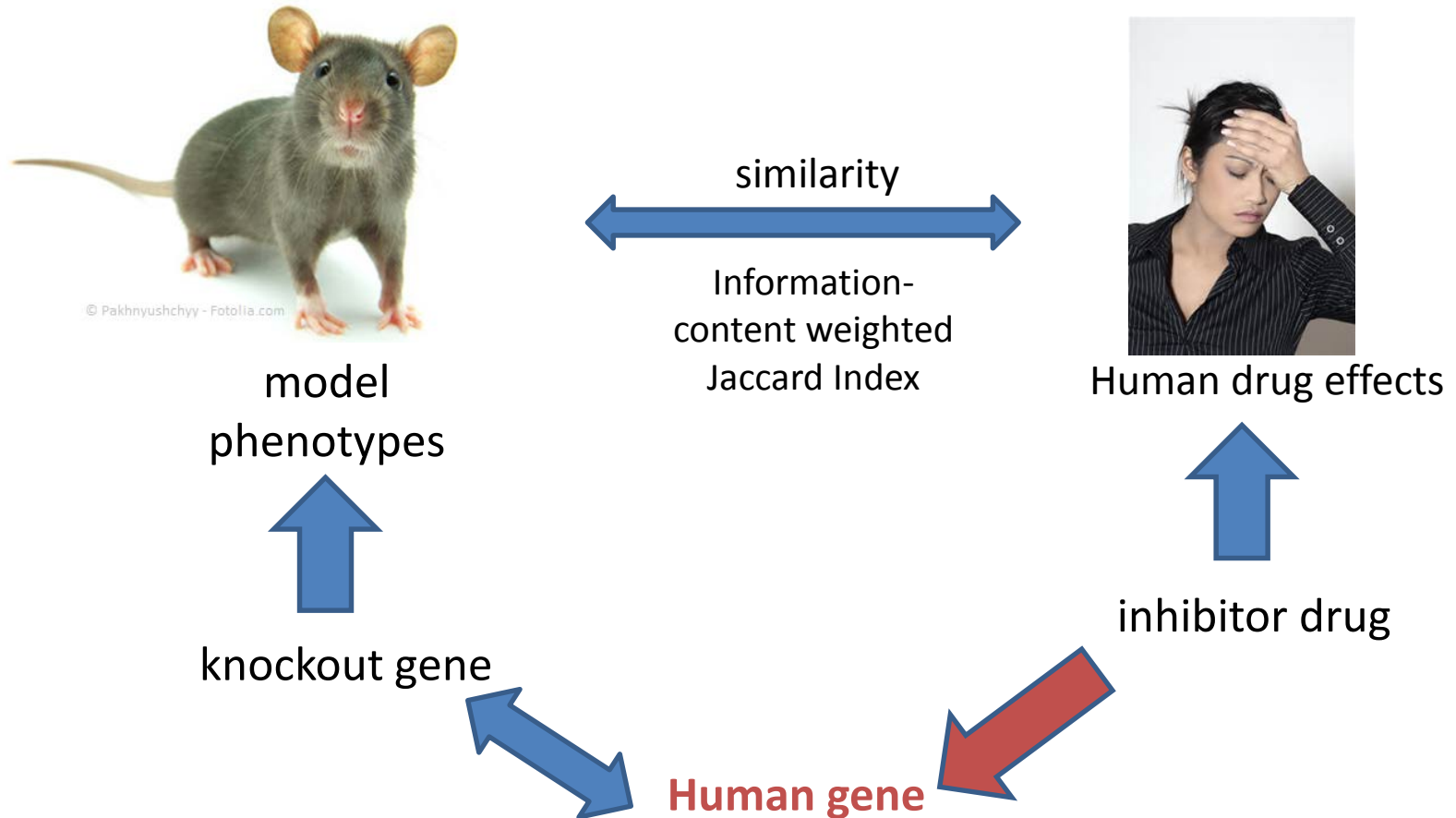
## Clinical research

- Identify an AD clinical trial for a drug with a **different mechanism of action** (MOA) than the drug that the patient is currently taking
  - *Of the 438 drugs linked to AD trials, only 58 are in active trials and only 2 (Doxorubicin and IL-2) have a documented MOA. 78 AD-associated drugs have an established MOA.*

## Health care

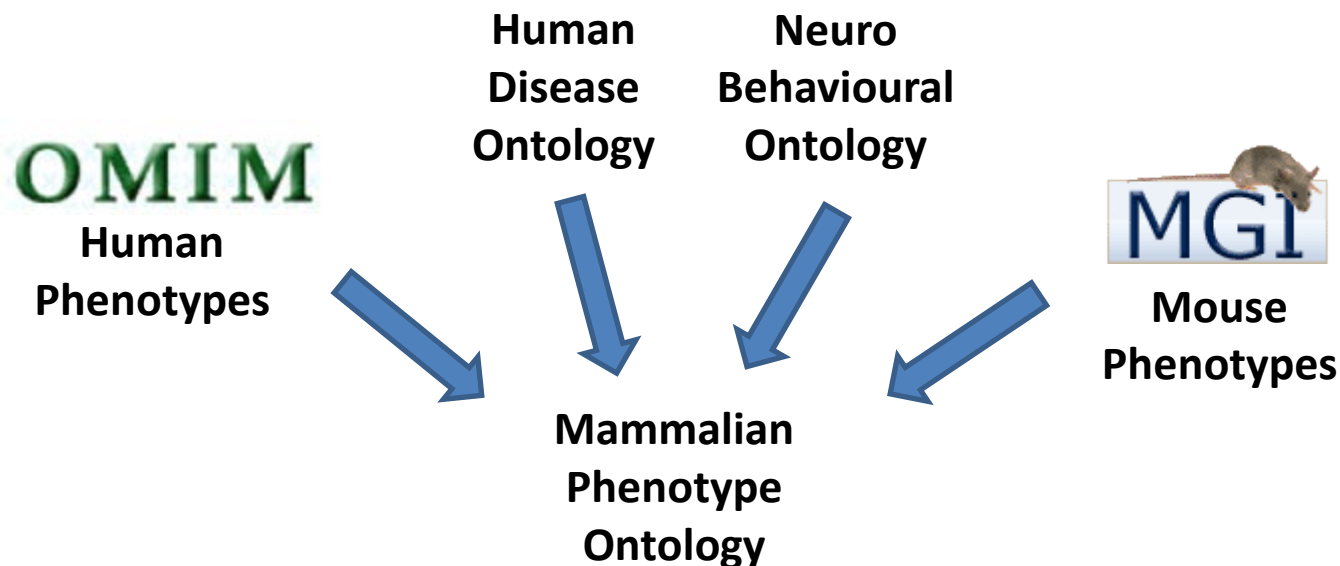
- Have any of my AD patients been treated for other neurological conditions as this might impact their diagnosis?
  - *Patient 2 is also being treated for depression.*

# Translational Research: Identifying human drug targets with animal model phenotypes





# Terminological Interoperability



PhenomeNet

PhenomeDrug



Drug effects

(mappings from UMLS to DO, NBO, MP)

# Terminological Interoperability *means* learning something new when you put them together.

**human** 'overriding aorta [HP:0002623]' EquivalentTo:

'phenotype of' some ('has part' some ('aorta [FMA:3734]' and 'overlaps with' some 'membranous part of interventricular septum [FMA:7135]')

**mouse** 'overriding aorta [MP:0000273]' EquivalentTo:

'phenotype of' some ('has part' some ('aorta [MA:0000062]' and 'overlaps with' some 'membranous interventricular septum [MA:0002939]')

**Uberon super-anatomy ontology provides inter-species mappings**

'aorta [FMA:3734]' EquivalentTo: 'aorta [MA:0002939]'

'membranous part of interventricular septum [FMA:3734]' EquivalentTo: 'membranous interventricular septum [MA:0000062]'

Thus, 'overriding aorta [HP:0002623]' EquivalentTo: 'overriding aorta [MP:0000273]'

# Summary

- A growing number of life science datasets are being made available as RDF Linked Data
  - easier to reuse these data than ever before
- We and others are building semantic clinical data warehouses to mine patient data for translational research
  - A standardized RDF representation will facilitate multi-site data exchange and data mining
- We must accept the emergence of multiple terminological and data sources, and learn how to make the most of it.



Let's use RDF to make it easier to do the work  
that really needs to be done.





dumontierlab.com

michel.dumontier@stanford.edu

Website: <http://dumontierlab.com>

*Presentations:* <http://slideshare.com/micheldumontier>