ONSTR: Ontology for Newborn Screening Follow-up and Translational Research

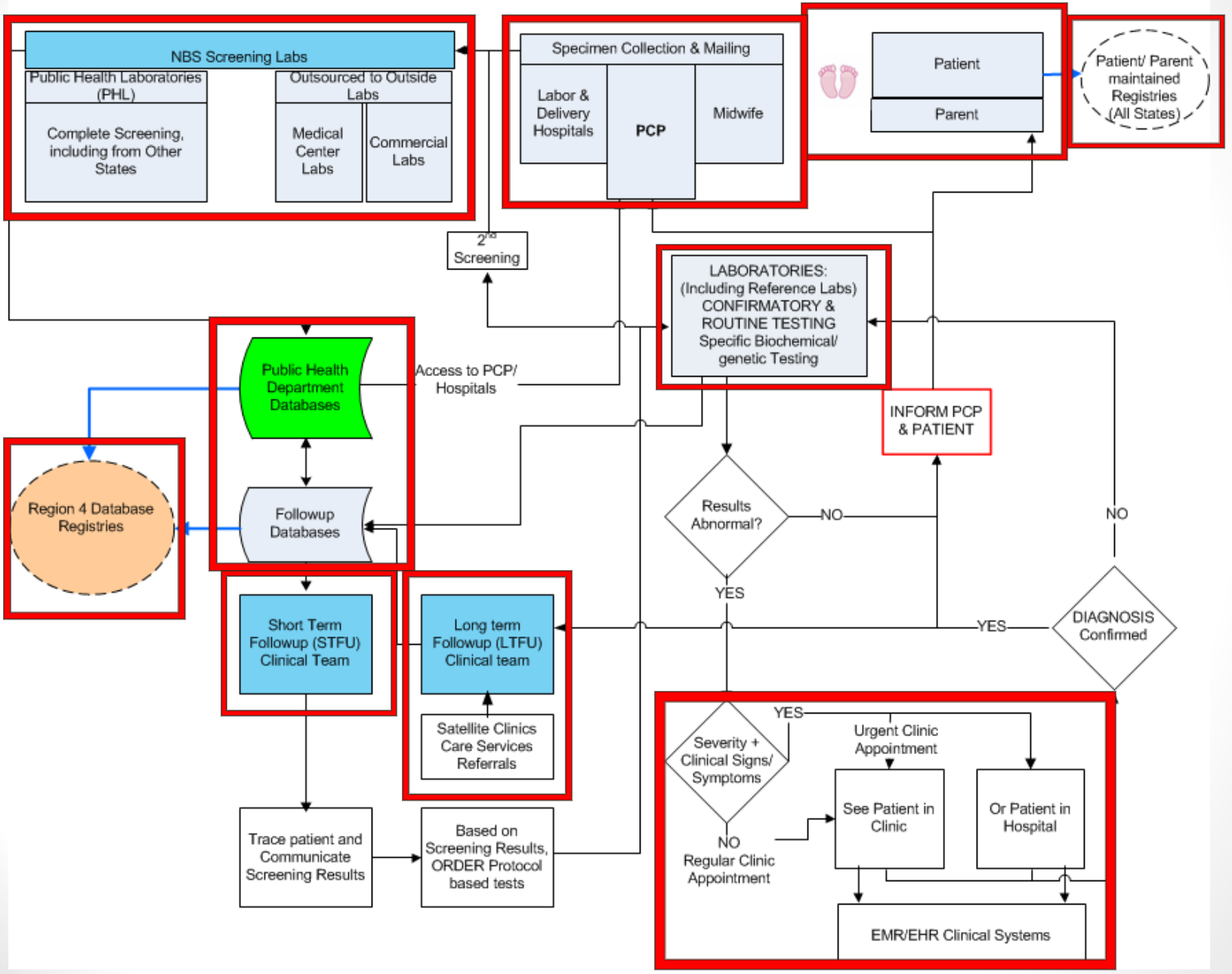
2013 Joint Meeting of the Newborn Screening and Genetic Testing Symposium
Thursday, May 9th
Atlanta

Prabhu Shankar MD, MS
Snežana Nikolić, MA,
Sivaram Arabandi, MD, MS
Shamakant Navathe
Kunal Malhotra
Rani H. Singh PhD, RD

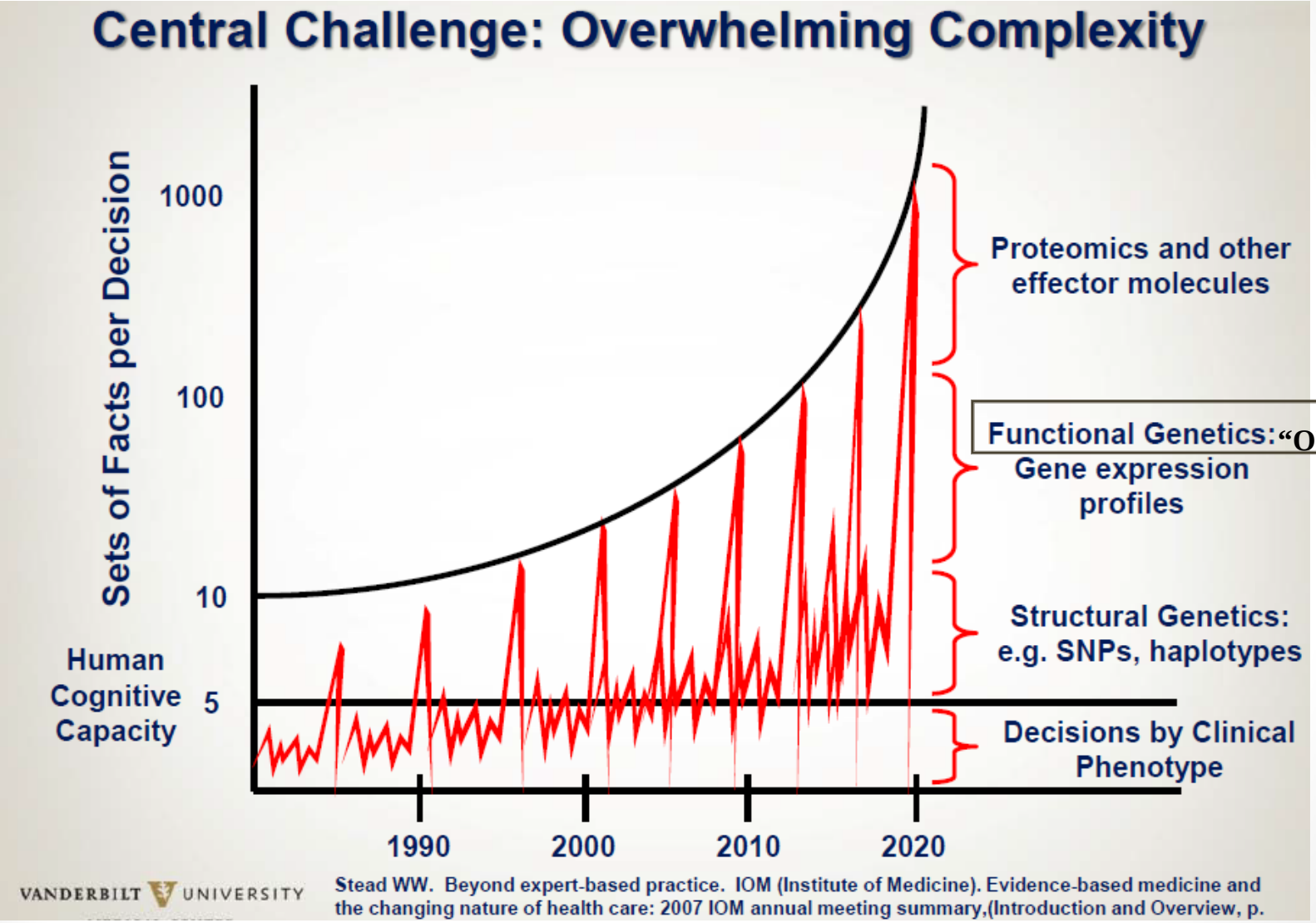
Objective

- NBS and follow-up workflows and data complexity
- Ontology and Semantic Web technologies
- ONSTR
- Newborn Screening Follow-up Data Integration Collaborative (NBSDC)
- Semantic Web technology success stories in healthcare – if time permits!

High Level View Of NBS and Follow-up System Workflow



External Computational Support!

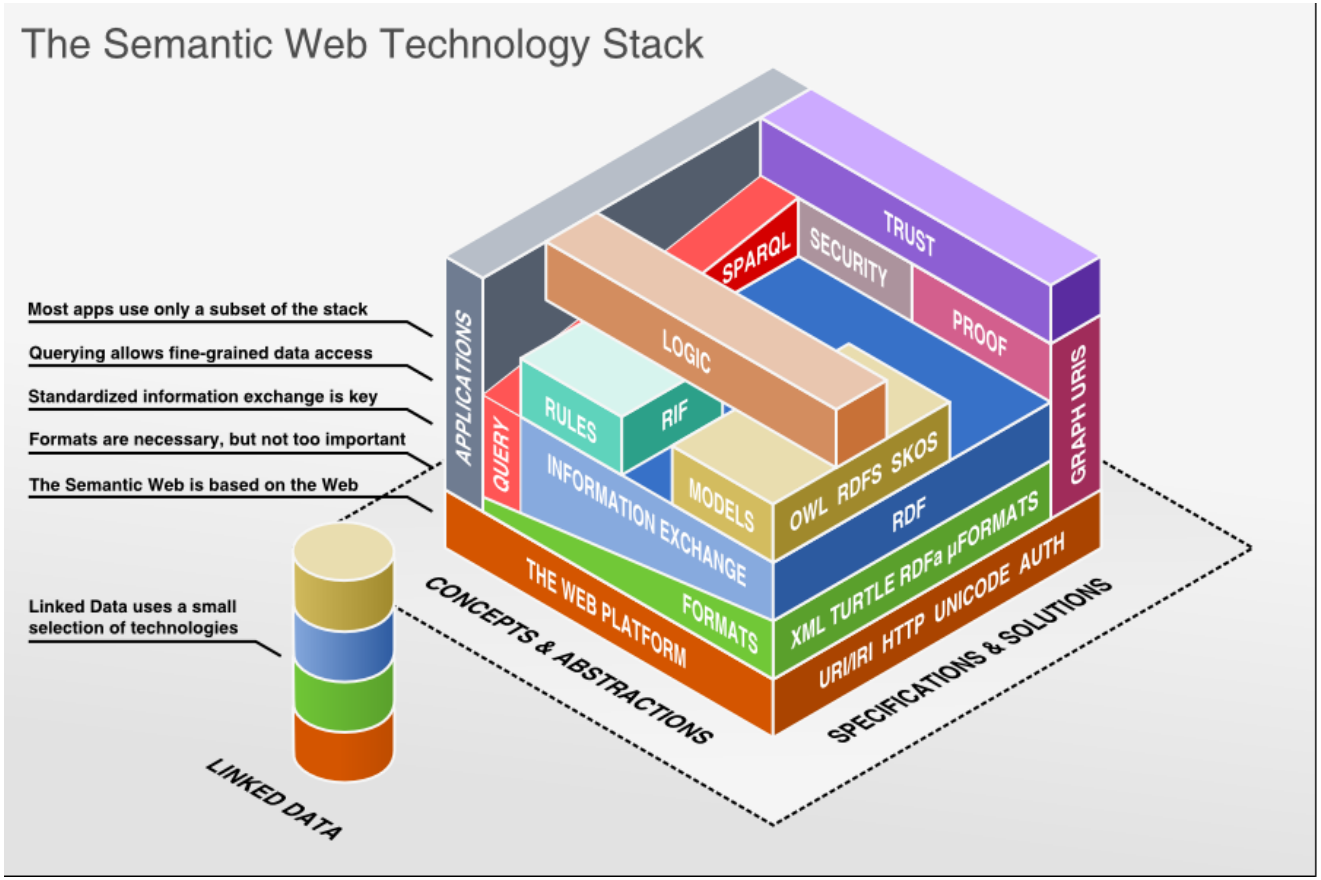


In Summary NBS Data is:

- **Geographically distributed** (data silos!)
- **Intersects clinical as well as many biomedical domains,** e.g., biochemistry, pathways, metabolomics, genomics, proteomics, pharmacogenomics
- **Various formats – structural, schematic and semantic variability**
- **Of rare diseases!**

Semantic Web Technologies

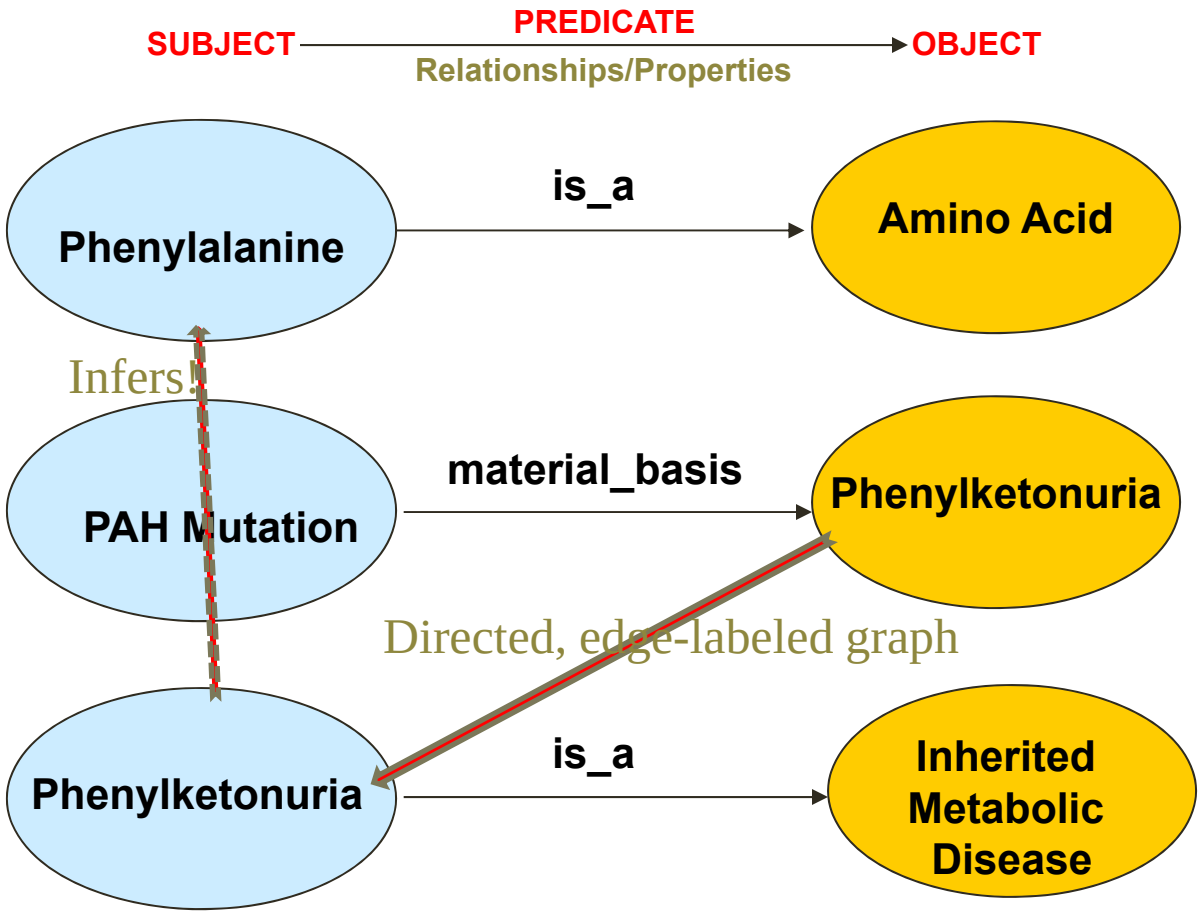
…technology stack to support a "Web of data"



"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries." World Wide Web Consortium (W3C)

Triples:

Resource Description Framework (RDF)



Simple, Dynamic, Extensible, Interoperable

RDF Schema (RDFS), Web Ontology Language → 'Ontology'

What is ontology?

1. A branch of philosophy, studying categories and types of beings existing in the universe.

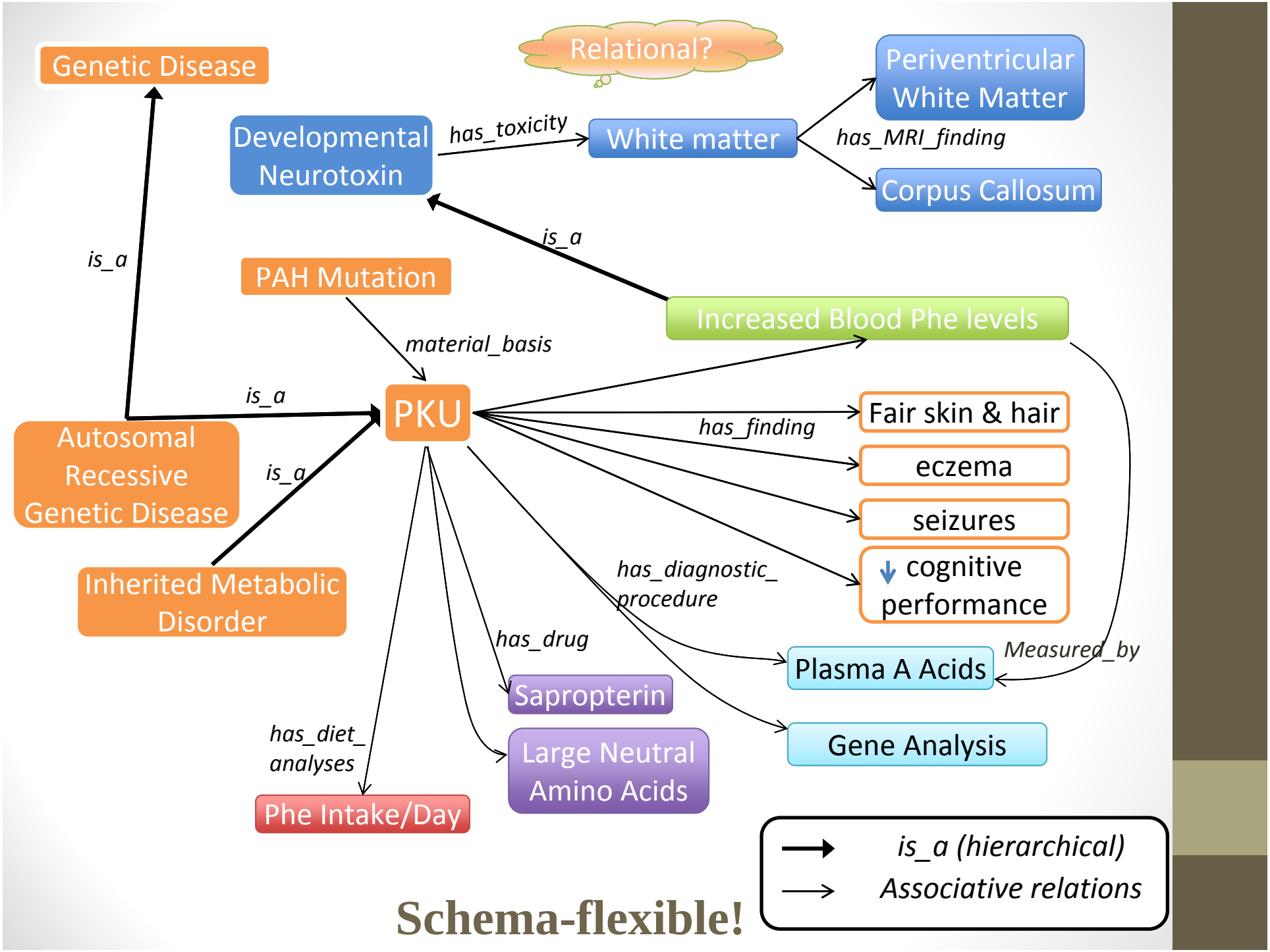
2. In Informatics, **explicit formal specifications** of the terms in the domain and relationships among them.

- Consensus based
- Associated with documentation and definitions
- **Expressed in formal logic** to support automated reasoning
- Interpretable by humans and computers

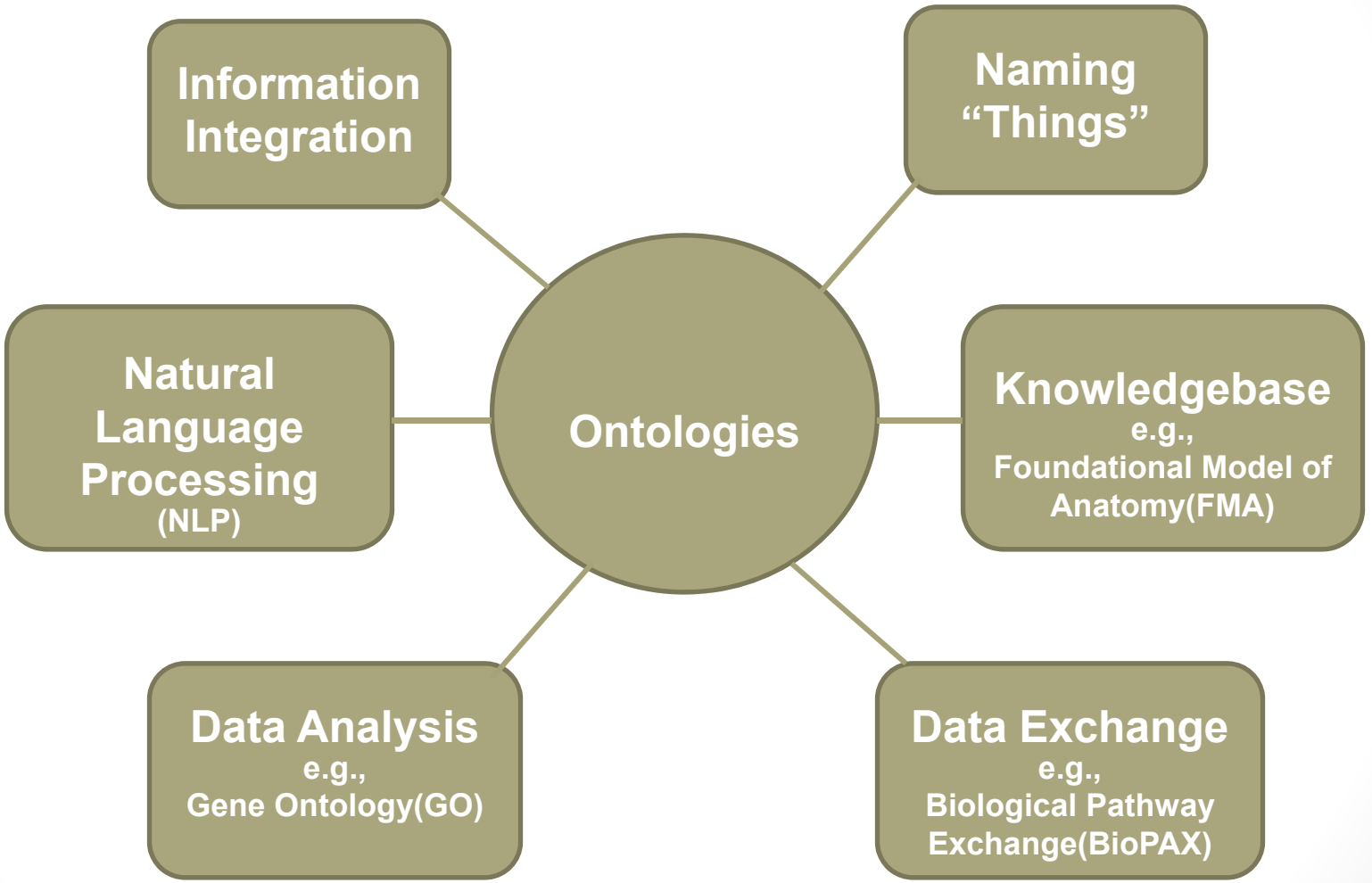
Semantic Methods and Characteristics

Deanna Pennington, LTER DataBits, Spring 2006

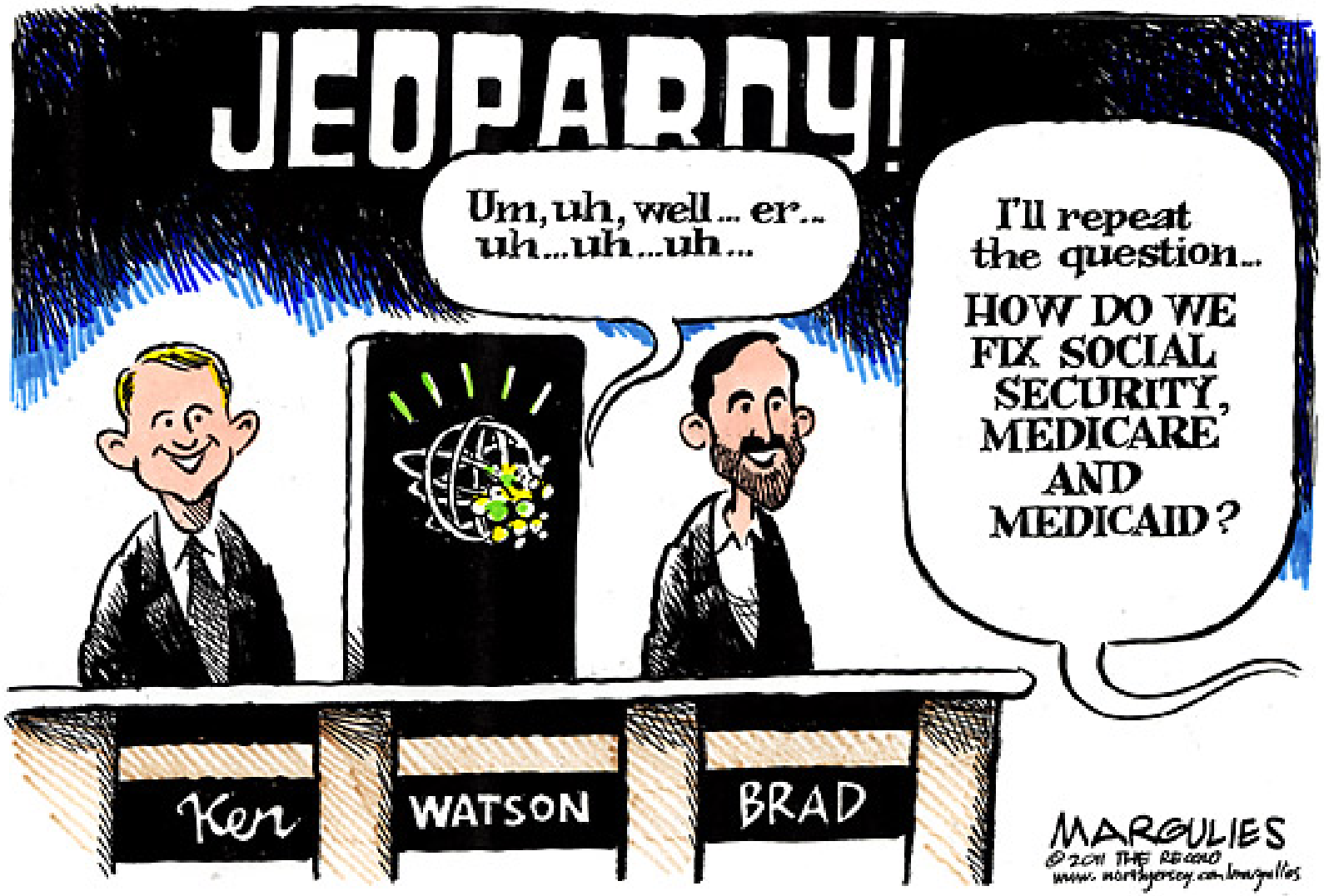
Method	Definition	Synonyms	Classification (isa)	Properties (has)	Other relations
Keywords					
Dictionary	X				
Controlled vocabulary	(X)	X			
Thesaurus	X	X			
Taxonomy	(X)	X	X Parrot is a bird	(X) Parrot has a beak	X
Ontology	X	X	X	**X** You can search by a term's properties	X


Genetic Disease
Relational?
Periventricular White Matter
Developmental Neurotoxin
has_toxicity
White matter
has_MRI_finding
Corpus Callosum
is_a
is_a
PAH Mutation
Increased Blood Phe levels
material_basis
is_a
PKU
Fair skin & hair
has_finding
Autosomal Recessive Genetic Disease
is_a
eczema
seizures
Inherited Metabolic Disorder
has_diagnostic_ procedure
cognitive performance
has_drug
Measured_by
Plasma A Acids
Sapropterin
has_diet_ analyses
Large Neutral Amino Acids
Gene Analysis
Phe Intake/Day
is_a (hierarchical)
Associative relations
Schema-flexible!

Ontology Applications



Natural Language Understanding!



http://sc6.blogspot.com/2011/02/cartoon-of-week_20.html

ONSTR: Ontology for Newborn Screening Follow-up and Translational Research

What is ONSTR?

An **application ontology** representing the processes, entities and knowledge in the Newborn Screening and follow-up system (Domain):

- Newborn screening Dried Blood Spot (NDBS) covering Inherited Metabolic Diseases (**IMDs**).
- **Genetic basis** of IMDs.
- Positive tested cases follow-up practice including: **medical/clinical confirmatory testing (biochemical and molecular).**
- **Medical and nutritional treatment** (dietary analysis monitoring)
- **Outcomes**, e.g., **physical** and cognitive **growth and development evaluation.**
- Research related to IMDs and NBS.

Why are we building ONSTR?

- To provide basis for **standardization of data** annotation in NBS domain.
- To provide **knowledge base** for integrating, aggregating and reasoning over data collected from different NBS sources.
- To **develop tools** for knowledge and data sharing to be used by greater IMD/NBS community.

Not Alone……

Open Biomedical Ontologies (OBO) Foundry principles and framework.

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)	Organ Function (FMP, CPRO)	Phenotypic Quality (PaTO)	**Biological Process** (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	**Cellular Component** (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)		**Molecular Function** (GO)		Molecular Process (GO)

ONSTR building process

1. Use case Definition

'of all the diagnosis confirmed patients who were new born screening positive, between 2005-2010, matching age and matching mutation (R408W), did good nutritional management VS Kuvan + Nutritional management had better outcome with regards to MRI White matter changes at five years?'.

2. Identification of key entities and relationships holding between these entities

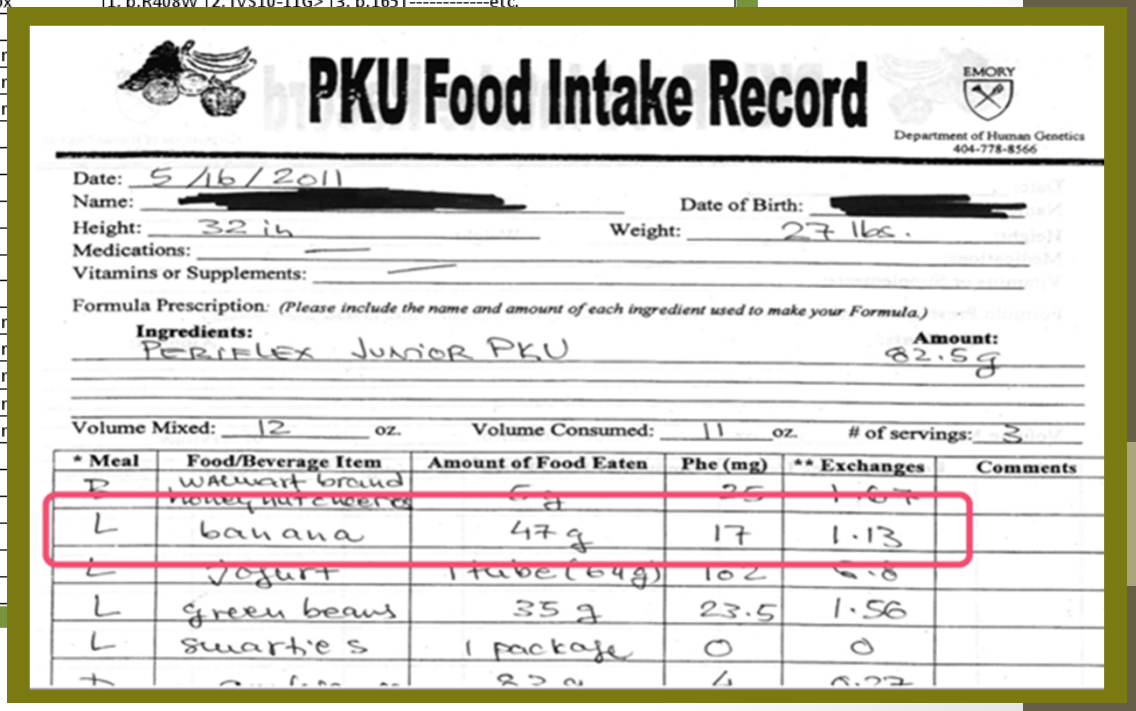
Methodology:

- Top Down and Bottom Up
- Survey of relevant literature
- Identifying the common data elements (CDEs)
- Follow OBO Foundry best practices

Top Down and Bottom Up

Common Data Elements (CDEs)

317	Understanding about blood Phe monitoring demonstrated	yesno	1, Yes | 2, No
318	Phenylketonuria		
319	**Data Label**	**Data Type**	**Choices**
320	Plasma Amino Acids	dropdown	1, Not done | 2, Abnormal | 3, Normal | 4, Non-diagnostic | 5, Unknown
321	Urinary Pteridine Profile	dropdown	1, Not done | 2, Abnormal | 3, Normal | 4, Unknown
322	Biopterin	text	
323	Pteridin	text	
324	Neoptrin	text	
325	Red Blood Cell (RBC) Dihydropteridine reductase (DHPR) activity	text	
326	Genetic testing performed since last visit	radio	1, Yes | 2, No | 3, Unknown
327	If yes, genetic mutation detected	Drop down	1, Phenex-100 |2,Phenex |3,Pro----------etc.
328	**MEDICAL FOOD PRESCRIPTION**		
329	Medical Food Prescribed	radio	1, Yes | 2, No | 3, Unknown
330	Name of the Medical Food prescribed	checkbox	1, p.R408W |2, IVS10-11G> |3, p.165T----------etc.
331	Intact (Natural) Protein Prescribed	radio	
332	Grams of Protein/day Prescribed from Intact (Natural) Protein	text/Nur	
333	Milligrams of Phe/day Prescribed from Intact (Natural) Protein	text/Nur	
334	Number of Phe exchanges/day	text/Nur	
335	Large neutral amino acids	radio	
336	Tyrosine supplementation	radio	
337	Pharmacological therapy (Kuvan)	radio	
338	Alternative pathway (PEG-PAL)	radio	
339	**DIET ANALYSIS**		
340	Type of Diet Record	radio	
341	Diet Analysis Results:		
342	Milligrams of Phe intake/day = mG/Day	text/Nur	
343	Milligrams of Tyrosine intake/day = mG/Day	text/Nur	
344	Grams of Protein intake from MEDICAL Foods/Day = G/Day	text/Nur	
345	Overall Total Protein intake/Day = G/Day	text/Nur	
346	Overall Total Energy intake/Day = G/Day	text/Nur	
347	**DIET/PRESCRIPTION ADHERENCE AND COMPLIANCE**		
348	Drinking Medical Food?	radio	
349	If yes, full or less than full	radio	
350	Consuming higher Phe than prescribed?	radio	
351	Consuming MEATS and DAIRY PRODUCTS?	radio	
352	Consuming MODIFIED LOW PROTEIN FOODS?	radio	



PKU Food Intake Record

EMORY
Department of Human Genetics
404-778-8566

Date: 5/16/2011
Name: [redacted] Date of Birth: [redacted]
Height: 32 in Weight: 27 lbs.
Medications: —
Vitamins or Supplements: —

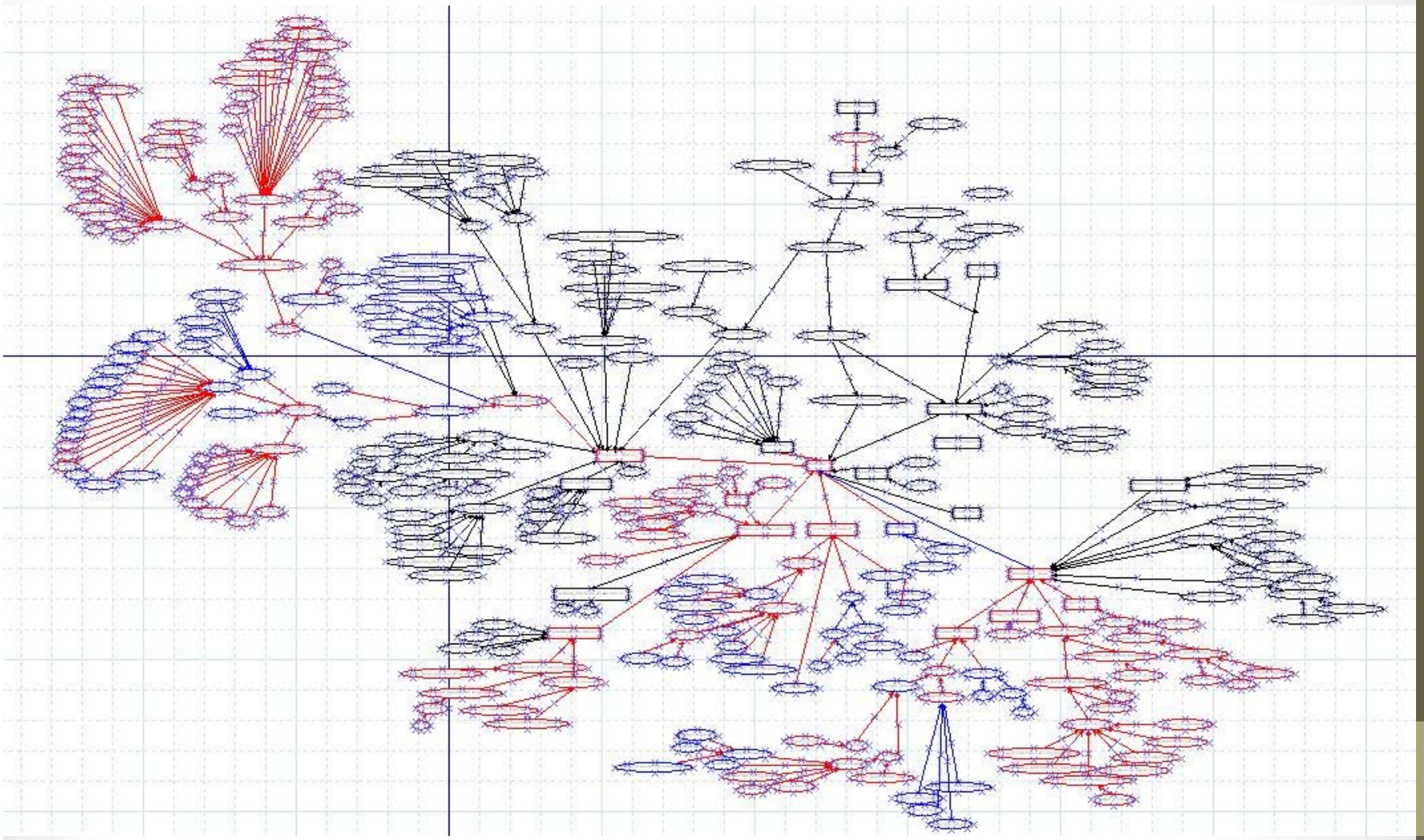
Formula Prescription: *(Please include the name and amount of each ingredient used to make your Formula.)*

Ingredients: PERIFLEX JUNIOR PKU Amount: 82.5 g

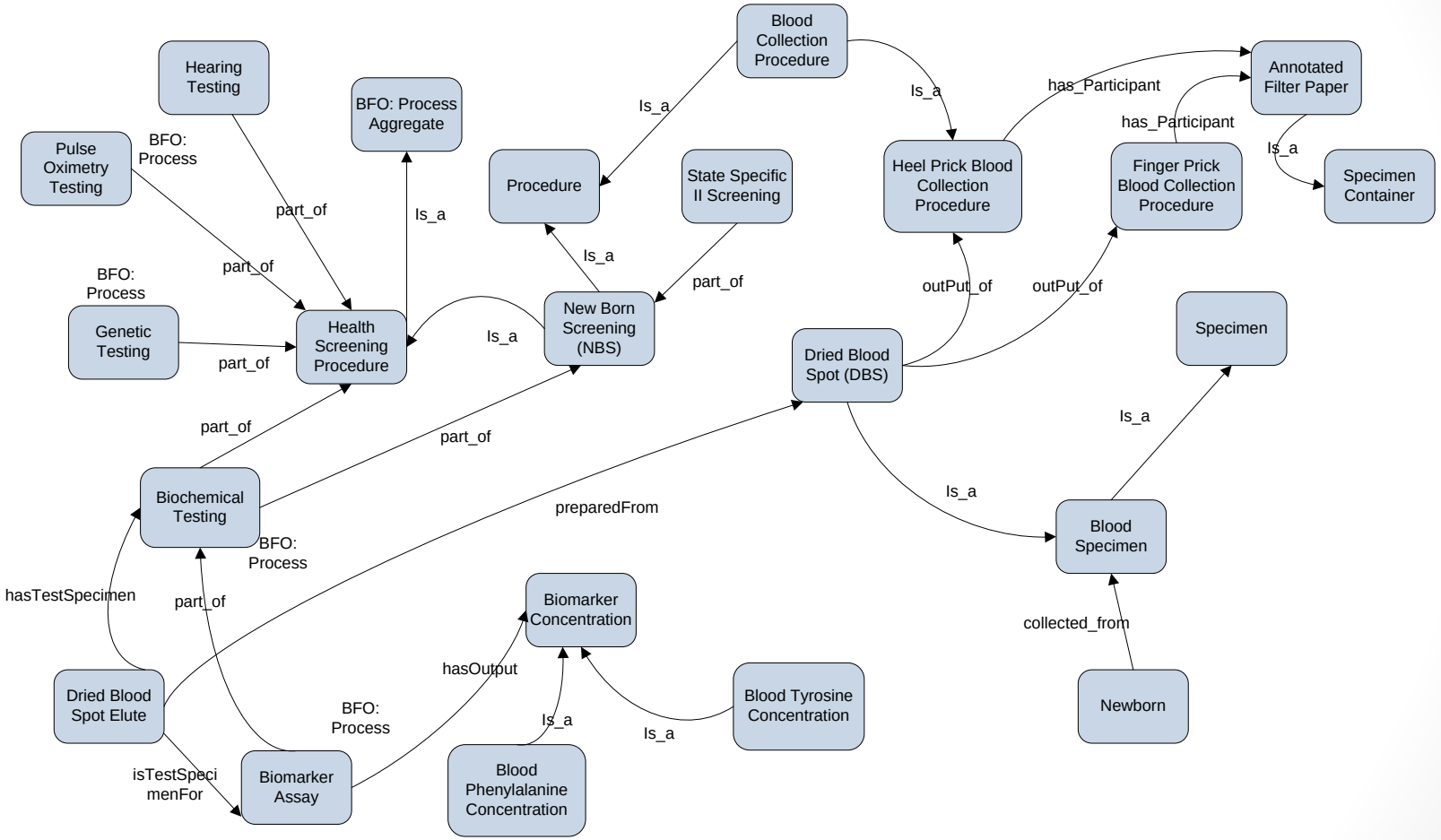
Volume Mixed: 12 oz. Volume Consumed: 11 oz. # of servings: 3

* Meal	Food/Beverage Item	Amount of Food Eaten	Phe (mg)	** Exchanges	Comments
B	Walmart brand Honey Nut Cheerios	6 g	25	1.67	
L	banana	47 g	17	1.13	
L	yogurt	1 tube (64g)	102	6.8	
L	green beans	35 g	23.5	1.56	
L	Smartie s	1 package	0	0	

All Common Data Elements (CDEs)



Modeling with Relations



ONSTR building process contd.

3. Ontology coding
   - ONSTR is formally encoded as a **RDF/XML serialization of OWL2 (W3C semantic Web standards)**

4. Ontology integration
   - Mappings between ONSTR and other relevant ontologies/vocabularies (Future work).

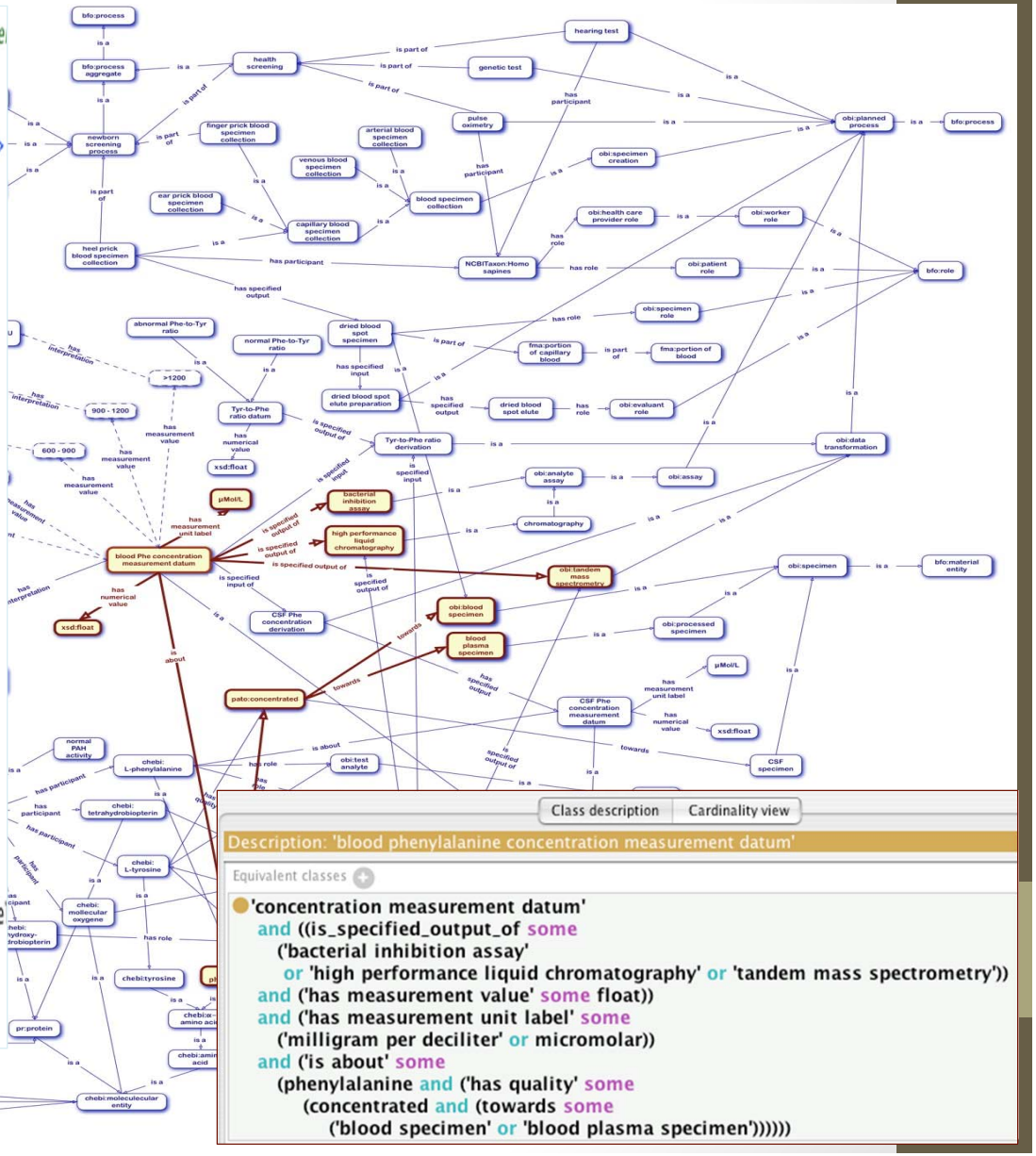
5. Ontology evaluation
   - In progress, concomitant with ONSTR development.

6. Ontology documentation
   - Available on the ONSTR project page: http://code.google.com/p/onstr/source/docs)

ONSTR graph and Logical Definitions

```
<!-- http://onstr.googlecode.com/svn/trunk/ONSTR.owl#ONSTR_2672569 --> <!-- head circumference-for-age pe

<owl:Class rdf:about="&ONSTR;ONSTR_2672569">
    <rdfs:label rdf:datatype="&xsd;string">head circumference-for-age percentile derivation</rdfs:label>
    <rdfs:subClassOf rdf:resource="&ONSTR;ONSTR_7375219"/>
    <owl:equivalentClass>
        <owl:Class>
            <owl:intersectionOf rdf:parseType="Collection">
                <rdf:Description rdf:about="&ONSTR;ONSTR_7375219"/>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&obo;OBI_0000293"/>
                    <owl:someValuesFrom rdf:resource="&ONSTR;ONSTR_5773210"/>
                </owl:Restriction>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&obo;OBI_0000293"/>
                    <owl:someValuesFrom rdf:resource="&ONSTR;ONSTR_3215569"/>
                </owl:Restriction>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&obo;OBI_0000293"/>
                    <owl:someValuesFrom rdf:resource="&ONSTR;ONSTR_2150908"/>
                </owl:Restriction>
            </owl:intersectionOf>
        </owl:Class>
    </owl:equivalentClass>
    <skos:definition rdf:datatype="&xsd;string">Head circumference-for-age percentile derivation is deve
    <rdfs:comment rdf:datatype="&xsd;string">Definition source: ONSTR curators.</rdfs:comment>
    <skos:editorialNote rdf:datatype="&xsd;string">Curation status: Incomplete</skos:editorialNote>
</owl:Class>
```



ONSTR Statistics

Total number of classes: 1842

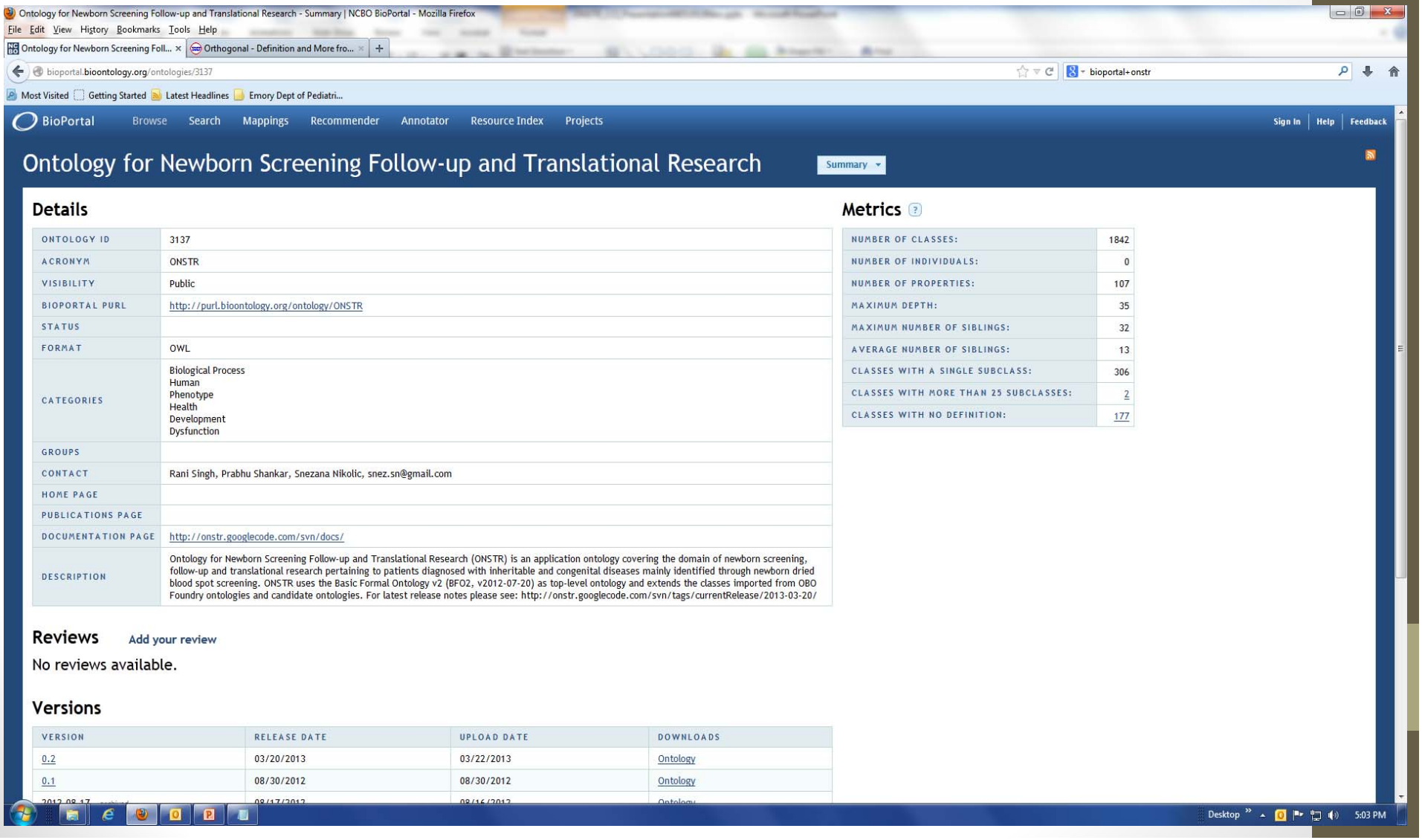
ONSTR native classes: 1100

Imported classes: 742

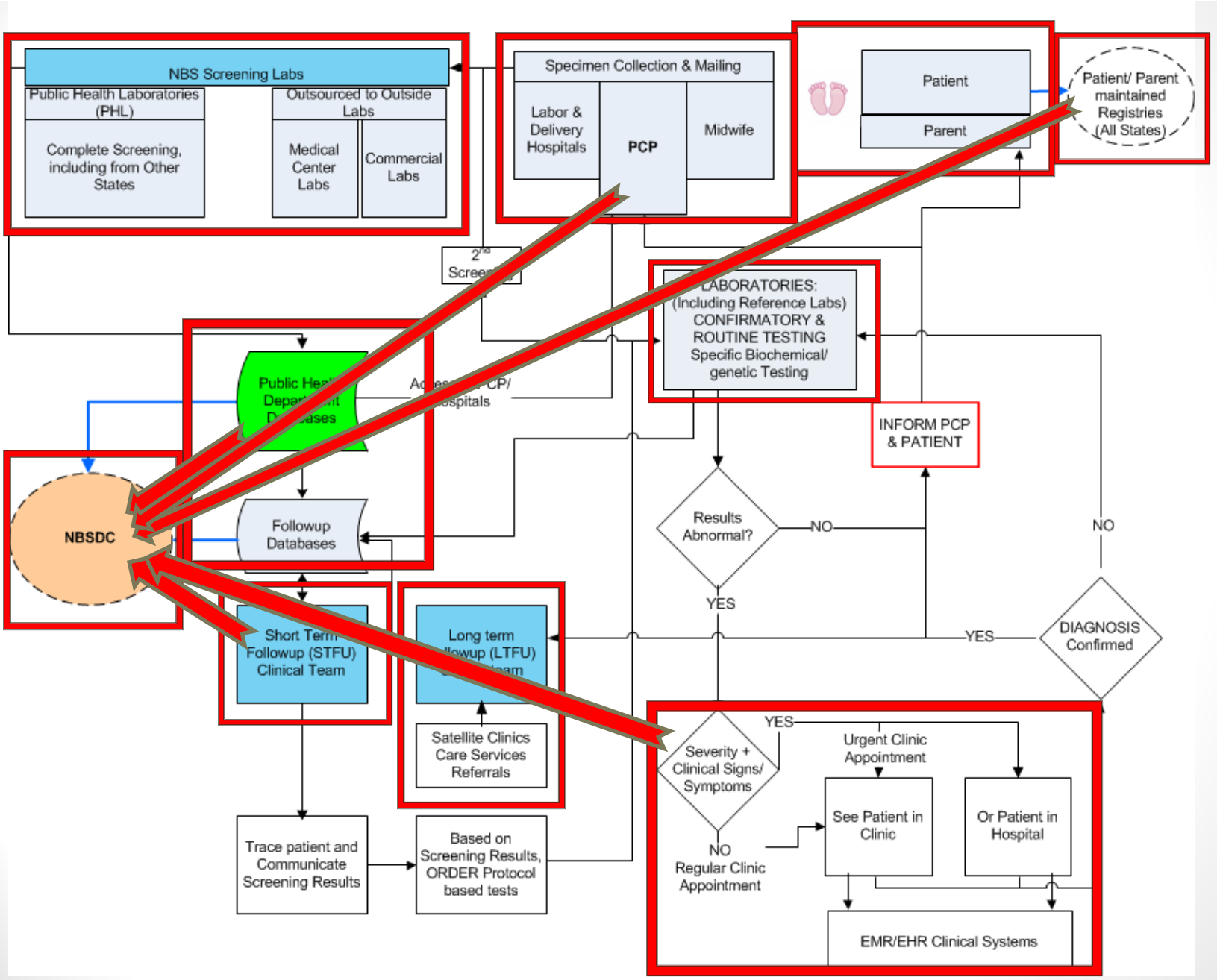
BioPortal

http://bioportal.bioontology.org/ontologies/49978

National Center for Biomedical Ontology (NCBO), Stanford University



Newborn Screening Follow-up Data Integration Collaborative (NBSDC)



Acknowledgements:

Funded by:

- **2009-10 HSI Seed Grant**, a Clinical Outcomes Research and Public Health (CORPH) Pilot Grants Program, jointly supported by **Georgia Tech and Children's Healthcare of Atlanta** and

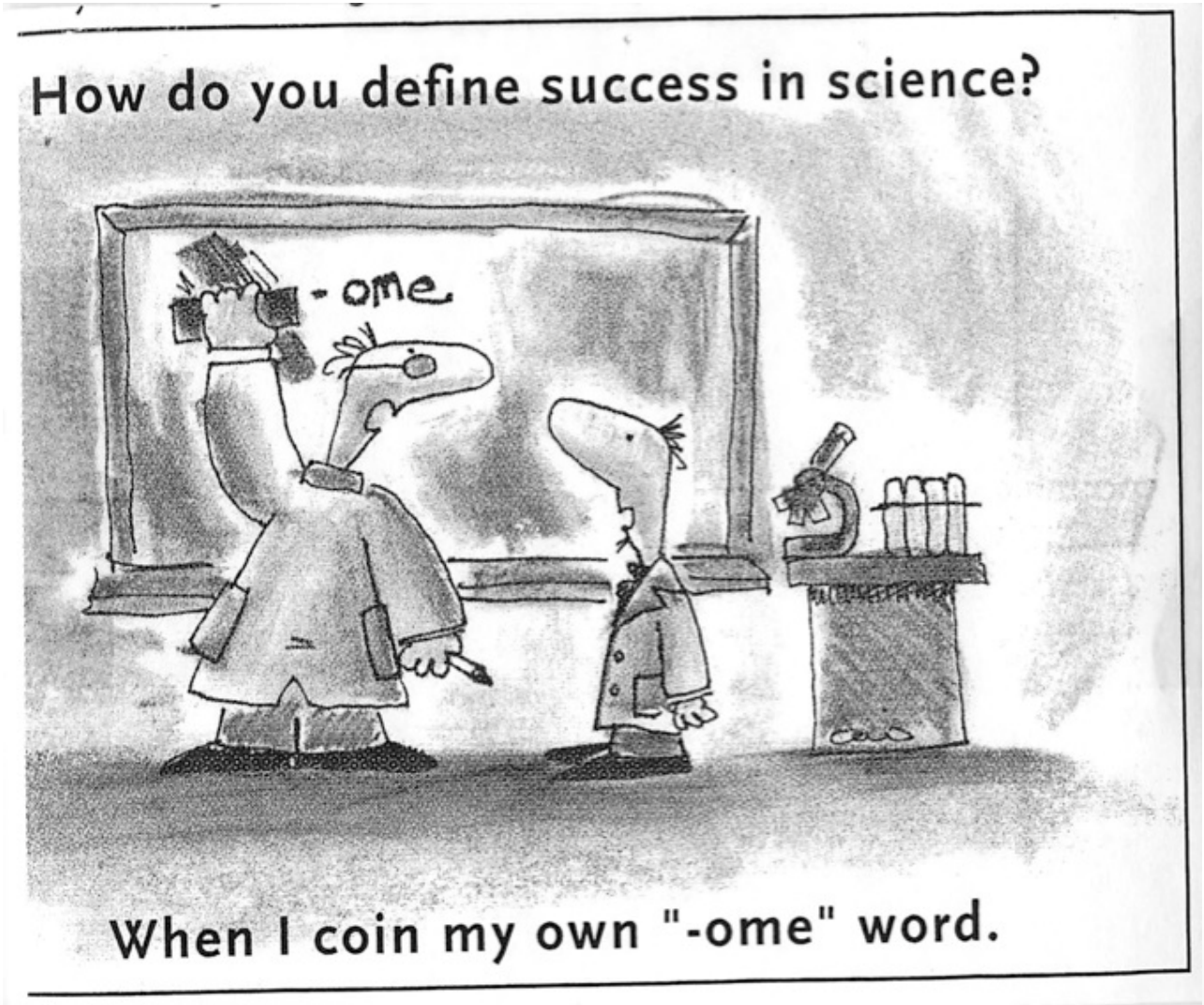
- The **Southeast NBS & Genetics Collaborative (SERC)** Grant from the Maternal and Child Health Bureau, **HRSA Grant U22MC10979.**

Special thanks to:

- **Dr. Barry Smith**, National Center for Ontological Research (NCOR), University @ Buffalo, Buffalo.

Thank You

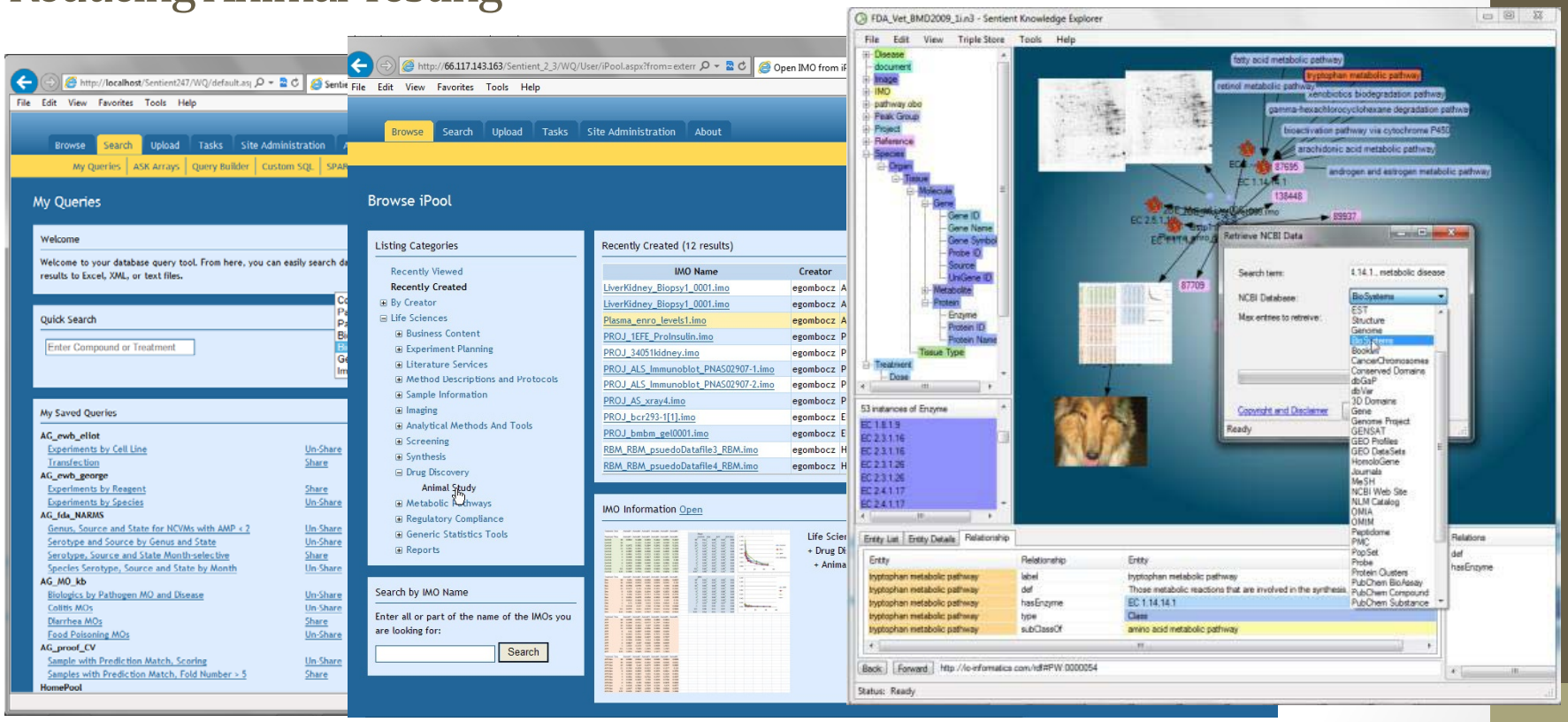
Questions: PRSHANK@emory.edu



Semantic technologies in action….

Cross-Species Biomarkers

Reducing Animal Testing



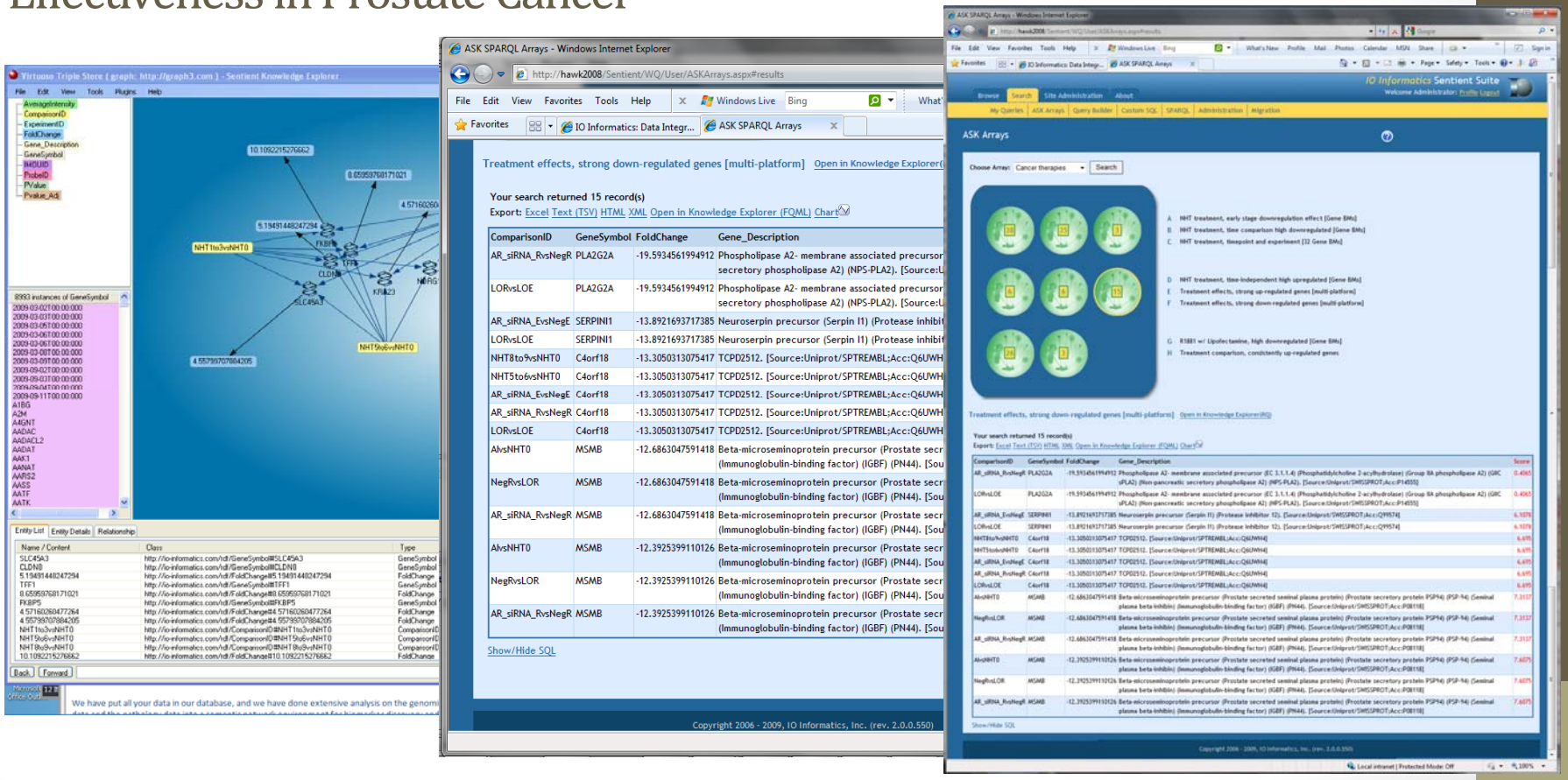
Result: Semantic integration (large animals to small animals to cell culture) to discover cross-species biomarkers applicable to human adverse events and diseases

Courtesy: Erich Gombocz, VP & CSO, IO Informatics, Inc.

Semantic technologies in action….

Combination Treatment

Effectiveness in Prostate Cancer

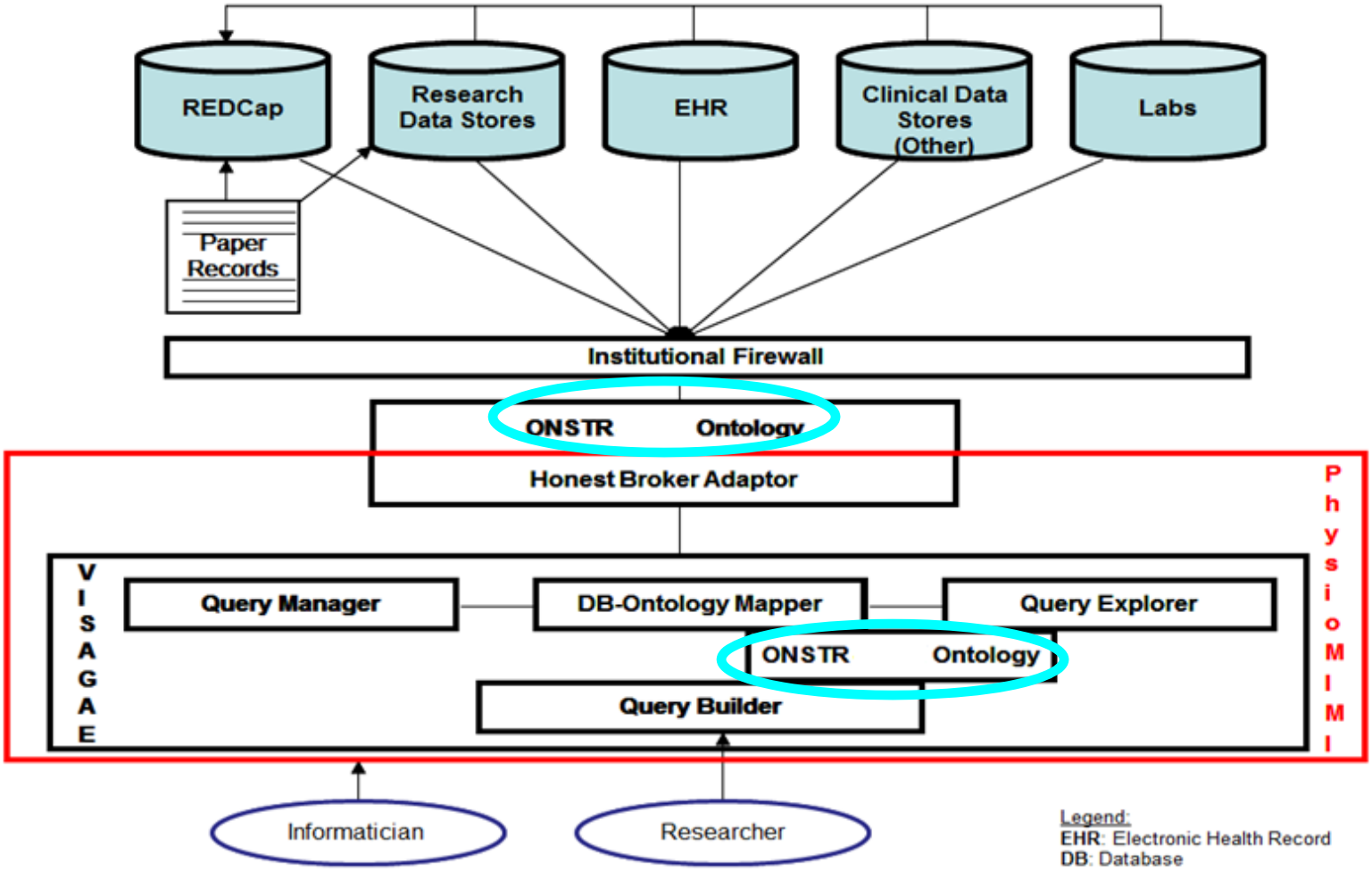


Result: effectiveness comparison of different combination treatments based on multi-platform genomic and proteomic marker profiles and patient match

Semantically Integrated Biological Networks Are Leading to Actionable Knowledge

INFORMATICS

Tools already being developed…

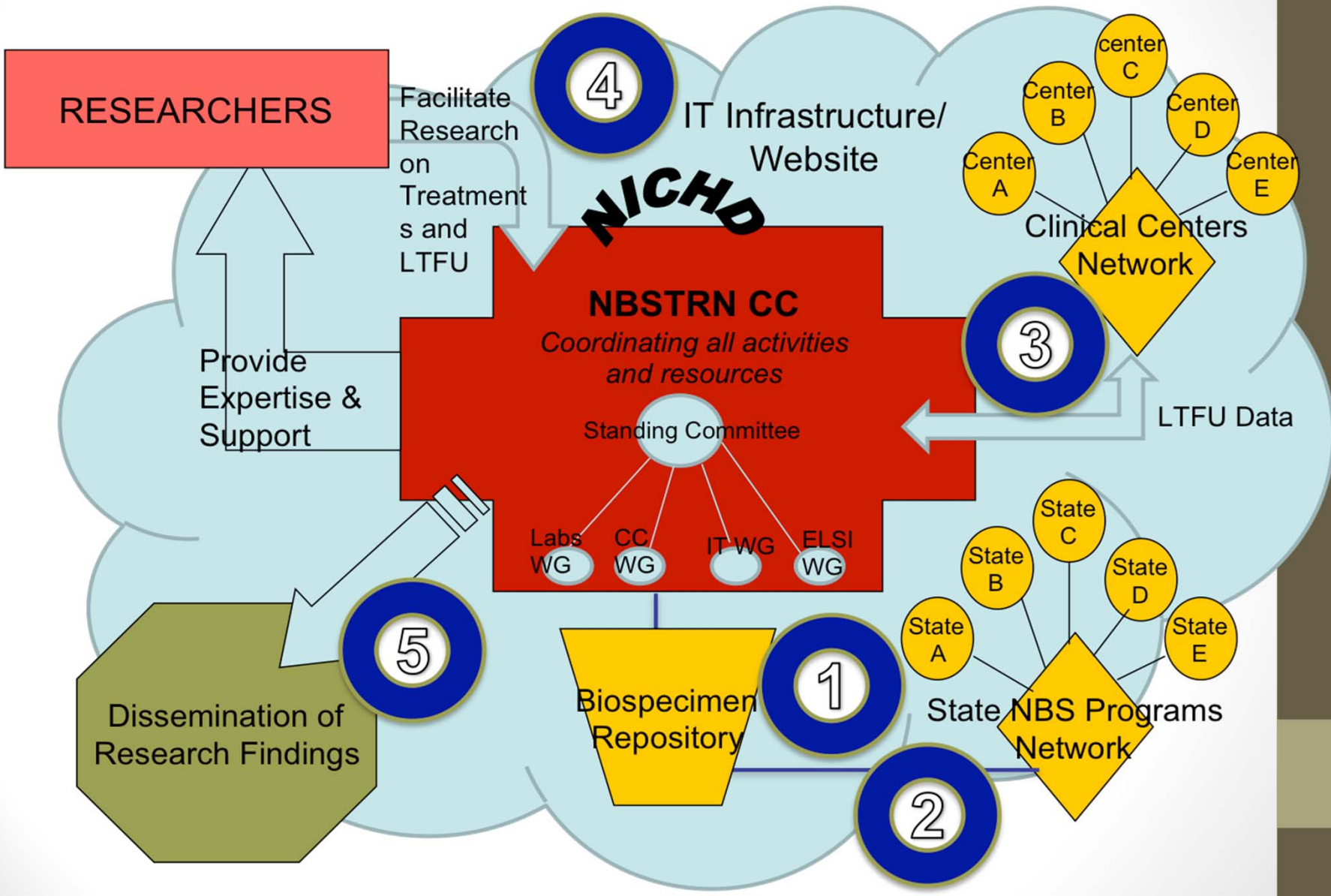


Legend:
EHR: Electronic Health Record
DB: Database
ONSTR: Ontology for Newborn Screening and Translational Research
PKU: Phenylketonuria
VISAGE: Visual Aggregator and Explorer
Physio-MIMI: Multi-Modality Multi-Resource Environment for Physiological and Clinical Research
REDCap: Research Electronic Data Capture

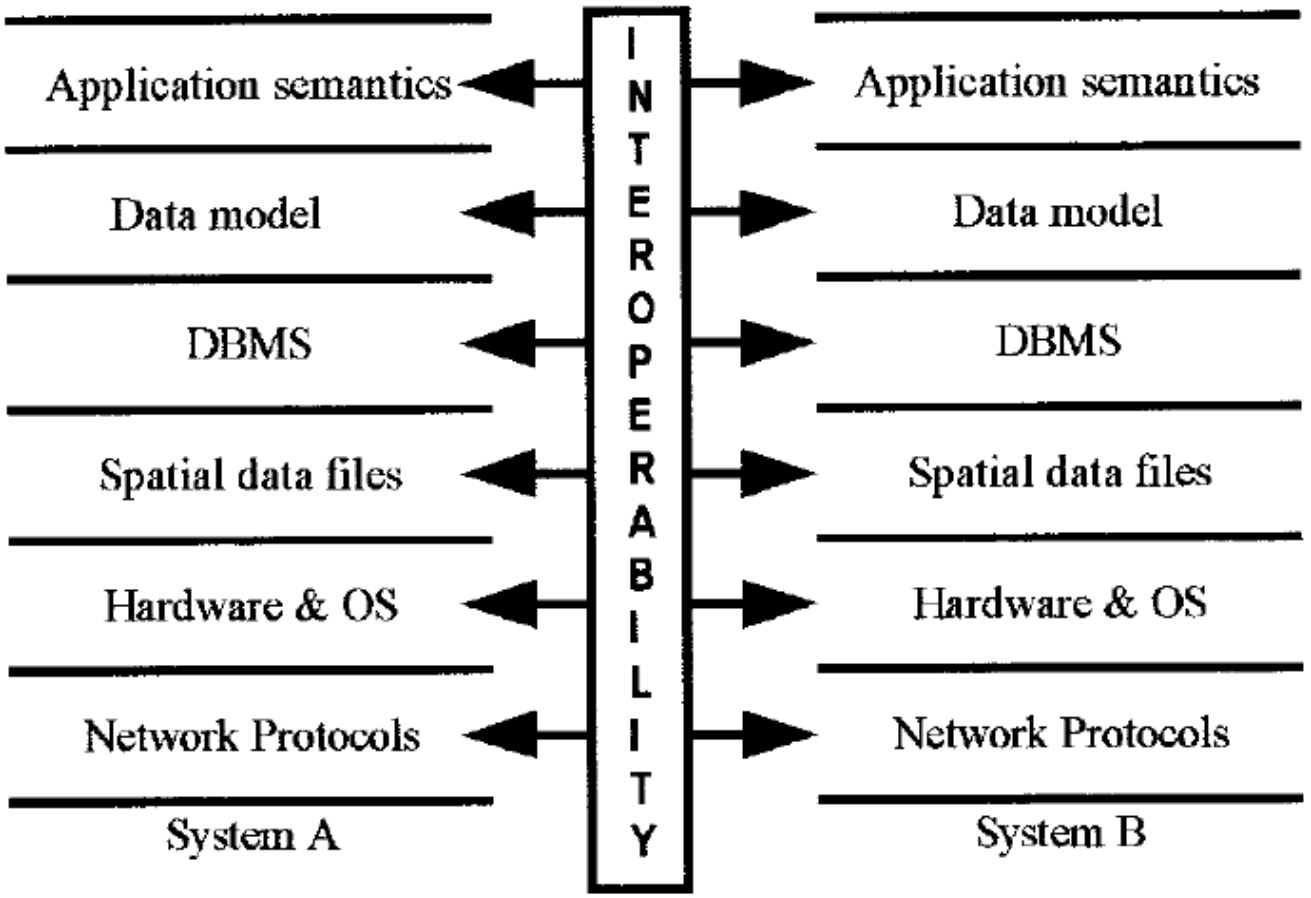
Challenges

- Time consuming
- Domain knowledge, Multi-disciplinary
- Computing Capacity to process Graphs
- Skilled personnel
- Funding
- Issues with data sharing
  - Buy in
  - Policy
  - HIPAA

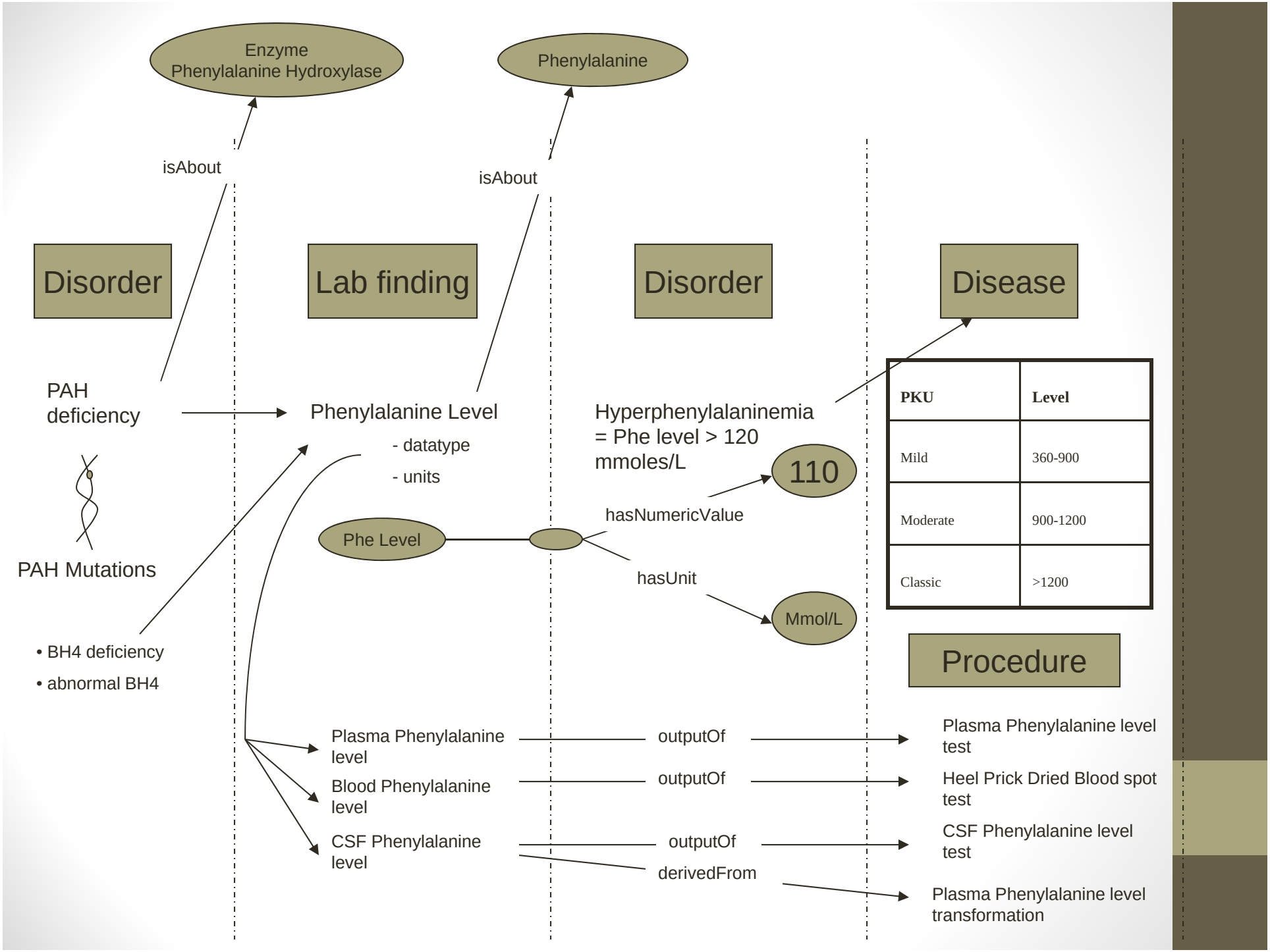
NBSTRN


RESEARCHERS
Facilitate Research on Treatments and LTFU
4
IT Infrastructure/ Website
NICHD
NBSTRN CC
Coordinating all activities and resources
Standing Committee
Labs WG
CC WG
IT WG
ELSI WG
Provide Expertise & Support
center C
Center B
Center D
Center A
Center E
Clinical Centers Network
3
LTFU Data
State C
State B
State D
State A
State E
State NBS Programs Network
5
Dissemination of Research Findings
Biospecimen Repository
1
2

Interoperability

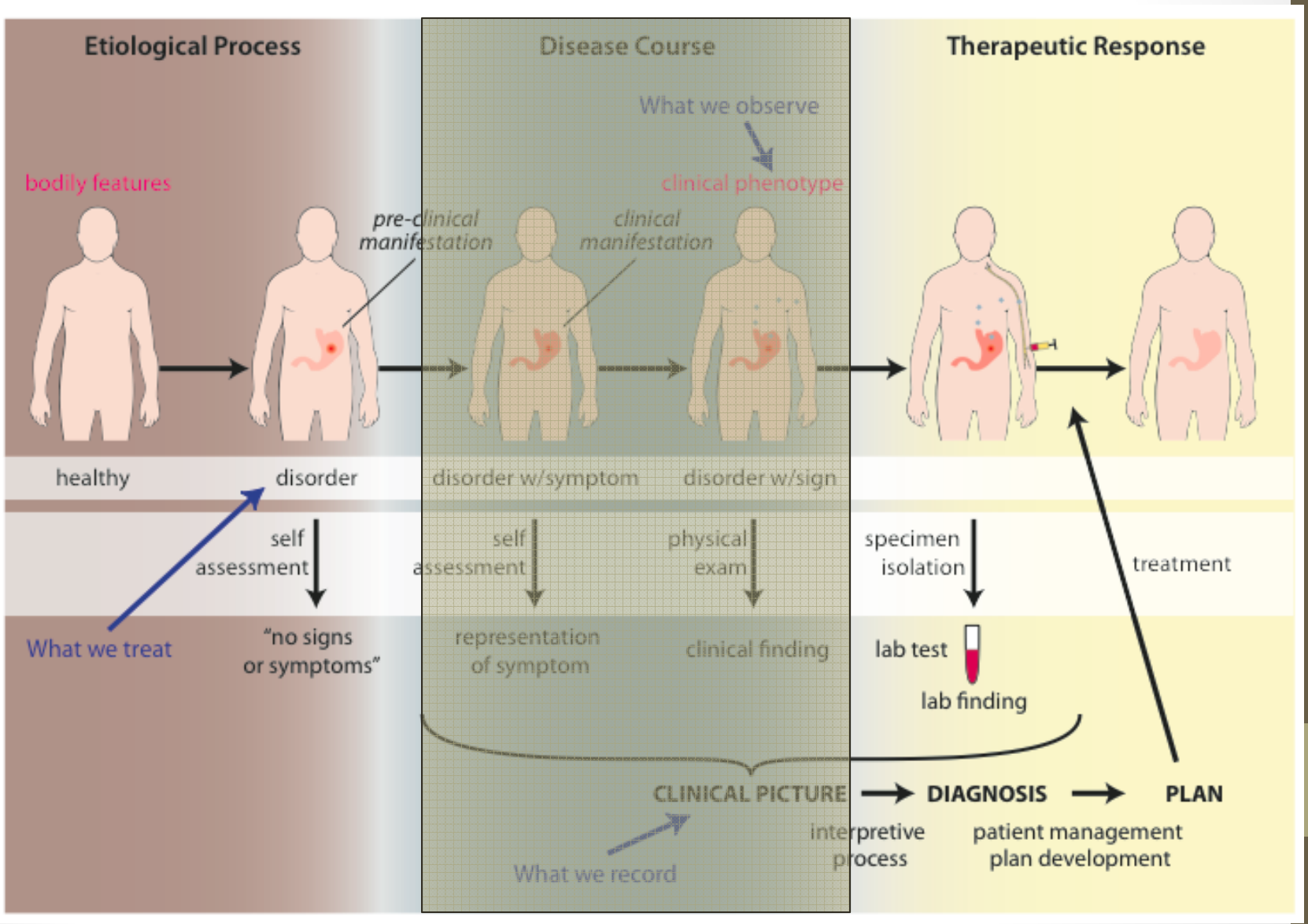


Levels of interoperability



PKU	Level
Mild	360-900
Moderate	900-1200
Classic	>1200

Big Picture


Etiological Process
Disease Course
Therapeutic Response
What we observe
bodily features
clinical phenotype
pre-clinical manifestation
clinical manifestation
healthy
disorder
disorder w/symptom
disorder w/sign
self assessment
self assessment
physical exam
specimen isolation
treatment
What we treat
"no signs or symptoms"
representation of symptom
clinical finding
lab test
lab finding
CLINICAL PICTURE
DIAGNOSIS
PLAN
What we record
interpretive process
patient management plan development