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!
External Computational Support!

Central Challenge: Overwhelming Complexity

Sets of Facts per Decision

- Proteomics and other effector molecules
- Functional Genetics: "OMICS"
- Structural Genetics: e.g. SNPs, haplotypes
- Decisions by Clinical Phenotype

Human Cognitive Capacity

VANDERBILT UNIVERSITY

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
<table>
<thead>
<tr>
<th>Method</th>
<th>Definition</th>
<th>Synonyms</th>
<th>Classification (isa)</th>
<th>Properties (has)</th>
<th>Other relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keywords</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dictionary</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controlled vocabulary</td>
<td>(X)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thesaurus</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxonomy</td>
<td>(X)</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parrot is a bird</td>
<td>Parrot has a beak</td>
<td></td>
</tr>
<tr>
<td>Ontology</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

You can search by a term’s properties
White matter

Corpus Callosum

Periventricular White Matter

Developmental Neurotoxin

has_toxicity

White matter

has_MRI_finding

PAH Mutation

Material basis

PKU

Autosomal Recessive Genetic Disease

Inherited Metabolic Disorder

Genetic Disease

PKU has material basis PAH Mutation

PAH Mutation is a PAH Mutation

PKU is_a Genetic Disease

PKU is_a Inherited Metabolic Disorder

PKU is_a Genetic Disease

PKU has finding Fair skin & hair

PKU has finding eczema

PKU has finding seizures

PKU has finding cognitive performance

PKU has finding Increased Blood Phe levels

PKU has finding Measured by Plasma A Acids

PKU has finding Measured by Gene Analysis

PKU has finding Phe Intake/Day

PKU has finding Sapropterin

PKU has finding Large Neutral Amino Acids

PKU has finding has_drug

PKU has finding has_diet_analyses

Relational?
Ontology Applications

Ontologies

- Information Integration
- Naming “Things”
- Natural Language Processing (NLP)
- Knowledgebase e.g., Foundational Model of Anatomy (FMA)
- Data Analysis e.g., Gene Ontology (GO)
- Data Exchange e.g., Biological Pathway Exchange (BioPAX)
Natural Language Understanding!

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.
Open Biomedical Ontologies (OBO) Foundry principles and framework.

<table>
<thead>
<tr>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>OCCURRENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRANULARITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Anatomical Entity (FMA, CARO)</td>
</tr>
<tr>
<td>CELL AND CELLULAR COMPONENT</td>
<td>Cell (CL)</td>
<td>Cellular Component (FMA, GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td>Molecular Function (GO)</td>
</tr>
</tbody>
</table>
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
All Common Data Elements (CDEs)
Modeling with Relations
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
ONSTR Statistics

Total number of classes: 1842
ONSTR native classes: 1100
Imported classes: 742
# BioPortal

[http://bioportal.bioontology.org/ontologies/49978](http://bioportal.bioontology.org/ontologies/49978)

National Center for Biomedical Ontology (NCBO), Stanford University

---

**Ontology for Newborn Screening Follow-up and Translational Research**

<table>
<thead>
<tr>
<th>Details</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontology ID</strong></td>
<td>3137</td>
</tr>
<tr>
<td><strong>Acronym</strong></td>
<td>OmniSTR</td>
</tr>
<tr>
<td><strong>Visibility</strong></td>
<td>Public</td>
</tr>
<tr>
<td><strong>BioPortal PURL</strong></td>
<td><a href="http://purl.bioontology.org/ontology/OMISTR">http://purl.bioontology.org/ontology/OMISTR</a></td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>OWL</td>
</tr>
<tr>
<td><strong>Categories</strong></td>
<td>Biological Process, Human Phenotype, Health, Development, Dysfunction</td>
</tr>
<tr>
<td><strong>Groups</strong></td>
<td>Rani Singh, Prabhu Shankar, Snezana Nikolic, <a href="mailto:snez.sn@gmail.com">snez.sn@gmail.com</a></td>
</tr>
<tr>
<td><strong>Homepage</strong></td>
<td><a href="http://omistr.googlecode.com/svn/docs/">http://omistr.googlecode.com/svn/docs/</a></td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Ontology for Newborn Screening Follow-up and Translational Research (OmniSTR) is an application ontology covering the domain of newborn screening, follow-up and translational research pertaining to patients diagnosed with inheritable and congenital diseases mainly identified through newborn dried blood spot screening. OmniSTR uses the basic Formal Ontology v2 (BFO2, v0.2012-07-20) as top-level ontology and extends the classes imported from BFO. Foundry ontologies and candidate ontologies. For latest release notes please see: <a href="http://omistr.googlecode.com/svn/tags/currentRelease/2013-03-20/">http://omistr.googlecode.com/svn/tags/currentRelease/2013-03-20/</a></td>
</tr>
</tbody>
</table>

**Reviews**

No reviews available.

**Versions**

<table>
<thead>
<tr>
<th>Version</th>
<th>Release Date</th>
<th>Upload Date</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>03/20/2013</td>
<td>03/22/2013</td>
<td>Ontology</td>
</tr>
<tr>
<td>0.1</td>
<td>08/30/2012</td>
<td>08/30/2012</td>
<td>Ontology</td>
</tr>
</tbody>
</table>
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
Tools already being developed…
Challenges

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

Levels of interoperability
Disorder
- PAH deficiency
  - BH4 deficiency
  - abnormal BH4

Lab finding
- Phenylalanine Level
  - datatype
  - units

Disorder
- Hyperphenylalaninemia
  - Phe level > 120 mmoles/L

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

Procedure
- Plasma Phenylalanine level test
- Heel Prick Dried Blood spot test
- CSF Phenylalanine level test
- Plasma Phenylalanine level transformation