Enhancing Organism Based Disease Knowledge Using Biological Taxonomy, and Environmental Ontologies

Ken Baclawski
Northeastern University
Neil Sarkar
Marine Biological Laboratory

Research Issues

- Biomedical knowledge relevant to the study of infectious diseases is currently in a variety of heterogeneous data sources
  - Citation databases
  - Health reports
  - Molecular databases
- Understanding infectious diseases requires
  - Environmental and geo-location
  - Biodiversity and biomedical resources

Disease Knowledge Sources

- Research Literature Citation Indexes
  - Medline of the US National Library of Medicine
  - Agricola of the US National Agricultural Library
- Health Reports
  - Global Outbreak Alert and Response Network (GOARN) of the World Health Organization
  - Program for Monitoring Emerging Diseases (ProMED) of the International Society for Infectious Diseases

Biodiversity Sources

- Biodiversity Heritage Library
- Global Biodiversity Information Facility (GBIF) hosted by the University of Copenhagen
- Encyclopedia of Life
- Many others…

Some Background Ontologies

- NCBI Taxonomy of the US National Center for Biotechnology Information
  - Alpha taxonomy associated with molecular data (GenBank)
- Environmental ontology (EnvO)
  - Emerging Open Biomedical Ontology (OBO) of biological habitats
- Geo-location instance hierarchy (Gaz)
  - Emerging OBO instance hierarchy of geo-locations

Example of integration of disease knowledge, genetic information, biodiversity information and geographical information

Geographic distribution of hantavirus disease outbreaks (boxes) and genetic samples (helices)

Geographic distribution of biodiversity information for the two most common US deer mouse species

OOR Hosted Ontology

- Union of Biological Taxonomy (uBiota)
- Derived from these sources:
  - NCBI Taxonomy
  - Species2000
  - Integrated Taxonomic Information System
- Only Considers Linnaean Ranks
  - Kingdom (8); Phylum (140); Class (324); Order (1464); Family (8801); Genus (148,459); Species (1,451,748)

Developer Requirements

- **Must have** the ability to browse and query small segments of an ontology.
- **Good to have** the ability to dynamically curate and suggest changes via the user community.
- **Ideally**, **it can** be used to navigate across inferred information that is associated with a small set of terms and that comes from many ontologies.

End User Requirements

- **Must have**
  - Ability to efficiently navigate multiple hierarchies
  - Consistency across multiple ontologies
- **Good to have**
  - Ability to provide live feedback
  - Allow annotating relationships or propose new terms
- **Ideally, it can**
  - Support scientific hypothesis testing