BioRDF Breakout Update

Kei Cheung, Ph.D.

Yale Center for Medical Informatics

BioRDF Teleconference Call, November 9, 2009

Format

- Presentations
  - Introduction of microarray use case (Kei Cheung)
  - Overview of MAGE-TAB (Michael Miller)
  - vOID (Jun Zhao)
  - aTags (Matthias Samwald)
  - The above presentations can be downloaded from the BioRDF section of the HCLS F2F wiki page (http://esw.w3.org/topic/HCLSIG/Meetings/2009-11-02_F2F)
- Discussion
  - Participants: Kei Cheung, Lena Deus, Don Doherty, Yolander Gil, Peter Handler, Scott Marshall, Michael Miller, Eric Prud'hommeaux, Joshua Phillips, Matthias Samwald, Nigam Shah, Harold Solbrig, Jun Zhao, ...

Discussion

- What is the RDF structure
- Extension of SPARQL to empower data analysis
- Workflow and provenance
- Visualization
- How to integrate database and literature
- Integration with other types of data
- Inter-community collaboration
- Translational use cases

What should be the RDF structure?

- Experiments, Samples, Experimental conditions/factors/phenotypes (e.g., age, disease, behavior), Gene lists, Arrays/chips, Raw/processed data (e.g., CEL, GPR, gene matrix)
- Ontologies (NIFSTD, SWAN, OBI, MGED, Experimental Factors, ….)
- Linked data format (e.g., vOID)
- Semantic tagging of free text (e.g., aTags)

Extension of SPARQL

- Hierarchical queries
- Statistical analyses/tests
- Enrichment analysis

Workflow and provenance

- Taverna, Biomoby,
- bioconductor (R), matlab
- Genepattern (can be accessed from Java, matlab, R)
- Software vendors
- Grid computing

Visualization

- Examples: Cytoscape, Cluster TreeView

Integration between database & literature

- Genelists may be a place to start to tie literature and database together
- Possible use of atags?
- Can genelists associated with papers be referenced in mage-tab?

Inter-community Collaboration

- NCBO
- caBIG (caArray, caTissue, provenance, mage-tab – provenace graph, authentification, …)
- BioPAX
- SWAN

What other types of data can be integrated with microarray data

- Phenotype (OMIM), pathways (compare pathways),
- Different types of neuroscientists
- Image data (MRI) – ADNI (time series data, snp data)
- Gene ontology and pathways
- Entrez gene, genbank
- Atlases
- Drug (IUPHAR – receptor nomenclature and drug classification, LoDD)
- Model data?
- …

Translational use case

- Disease and phenotypes should be of interest to many neuroscientists

Acknowledgement

- Helen Parkinson (EBI)
- Maryann Martone (NIF)
- Nigam Shah (NCBO)
- HCLS
  - Susie Stephens, Scott Marshall, Eric Prud’hommeaux

Follow-up?