Yes, we can! Lessons from Using Linked Open Data (LOD) and Public Ontologies to Contextualize and Enrich Experimental Data

Erich A. Gombocz¹, Andrea Splendiani², Mark A. Musen³, Robert A. Stanley¹, Jason A. Eshleman¹

¹ IO Informatics, Inc., Berkeley, CA, USA ² IO Informatics, Inc., London, UK ³ Stanford Center for Biomedical Informatics Research (BMIR), Stanford, CA, USA
Correspondence: egombocz@io-informatics.com

Summary
Knowledgebases in RDF are growing, particularly as Linked Open Data (LOD) resources. However, alignment of public and experimental corporate standards, inconsistent namespace policies, and issues from use of internal, non-formal application ontologies are challenging. Resource reliability is often problematic, more and more “Open data” are closing for commercial use, and there are serious funding concerns related to government grantbacking. Despite this, we make the case that, yes, we can efficiently use LOD resources.

Mapping, Harmonization and Resource Alignment
1) Experimental data are imported using a mapper to create RDF
Toxicity studies on rats using Microarray GEP and LC/MS-MS metabolic profiling were performed in liver, brain, serum and urine, and statistical biomarker pre-selection was done at p<0.005, abs fc>5 (genes) and p<0.005, abs fc>2.5 (metabolites).

2) Formal public ontologies are selected, imported, aligned and linked via NCBO services

3) For classifiers, the integrated network is explored for similar perturbed genes and metabolites from multiple treatments
Public Resources: UniProt, Human and rat databases; Pathway (rat dataset); selective (imported from SPARQL endpoint query results) from Drugbank, Disease, GO NCBI Biosystems; and NCBI Gene.

Results & Discussion
- Public LOD resources were used to qualify statistical pharmacogenomic correlations with systems biological events to quality toxicity biomarkers.
- Contextualization and enrichment of internal experiments with public resources has proven feasible to associate metabolic and genetic responses with relationships to common pathway enzymes which are key to decode biological functions in toxicity.
- It needs to be recognized, however, that there are still open discussions on how to improve interoperability in-between certain LOD resources.
- Time and money saved has large socio-economic consequences. The society needs to acknowledge and support those efforts.

Table 1: Benzene-like toxicity marker with pharmacogenomic correlation and added biological function (18 genes, 2 metabolites, exported from LOD-enriched network)

See us in the Friday afternoon session at 3:00 PM to learn more!