Applying Semantic and Network Methods in AOP Knowledge Discovery

David Wild, Ph.D
Associate Professor, Indiana University School of Informatics & Computing
Founder & CEO, Data2Discovery Inc

Adverse Outcome Pathways: From Research to Regulation
NIH Workshop, Bethesda MD, September 3-5 2014

Email djwild@indiana.edu | Web http://djwild.info
Purpose of this presentation and discussion

A view of what is possible when we bring together the emerging science of AOP’s, and state of the art in the computational techniques of data science, semantic technologies and network science.

For technical details, see SOT presentation at http://djwild.info

The Usborne Book of the Future, 1979
Semantic technologies and AOP’s – a new opportunity

- Our understanding of the effects of chemicals on our body is moving from a reductionist approach to a system, network approach

- The impacts of a chemical on the body are complex
  - Multiple targets, pathways
  - Indirect cascade effects
  - Phenotype and genotype dependent

- Semantic technologies fit this model well, as a way to handle big, complex, networked data sets from multiple sources
  - Applications in drug discovery, safety and chemical toxicity
New “big” data approaches going mainstream in science

• NoSQL
  • Good for large amounts of simple or unstructured data
  • Very lightweight data structures e.g. tagging

• Semantic technologies
  • Good for large amounts of complex data
  • Represents data as networks rather than tables
  • Highly flexible in incorporating and linking many different kinds of data
  • Ontologies apply meaning to the data and relationships
  • Identified by Gartner as one of the top technology trends impacting information infrastructure in 2013: http://www.gartner.com/newsroom/id/2359715
  • Now heavily used internally Google, Facebook, etc
  • Increasingly applied in scientific domains
Value proposition

• Semantic and network technologies could aid researchers in building AOP’s and knowledge around AOP’s
  • Predicting associations between compounds, targets and end points
  • Testing hypothesis
  • “Auto suggestion” of AOP associations

• Semantic and network technologies could help us apply established AOP’s in problems like toxicity prediction
  • Profiling compounds across toxic end-points using computational representations of AOP’s
Example – Liver Fibrosis

Source: AOP Wiki
### Isoniazid

<table>
<thead>
<tr>
<th>PubChem CID</th>
<th>Structure</th>
<th>Drug Name</th>
<th>SimilarityRelated Diseases</th>
<th>ATC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3757</td>
<td><img src="image1" alt="Isoniazid" /></td>
<td>Isoniazid</td>
<td>Tuberculosis</td>
<td>J04AC01</td>
</tr>
<tr>
<td>3198</td>
<td><img src="image2" alt="Econazole" /></td>
<td>Econazole</td>
<td>0.915</td>
<td>D01AC03, D01AF05</td>
</tr>
<tr>
<td>5362440</td>
<td><img src="image3" alt="Indinavir" /></td>
<td>Indinavir</td>
<td>0.894</td>
<td>HIV, J05AE02</td>
</tr>
<tr>
<td>5362497</td>
<td><img src="image4" alt="Rifapentine" /></td>
<td>Rifapentine</td>
<td>0.885</td>
<td>Tuberculosis, J04AG06</td>
</tr>
<tr>
<td>5342</td>
<td><img src="image5" alt="Sulfapyrazine" /></td>
<td>Sulfapyrazine</td>
<td>0.873</td>
<td>M04AB02</td>
</tr>
<tr>
<td>4169</td>
<td><img src="image6" alt="Miconazole" /></td>
<td>Miconazole</td>
<td>0.841</td>
<td>A01AB03, A07AC01, D01AC02, G01AF01, J02AB01, S02AA13</td>
</tr>
<tr>
<td>12560</td>
<td><img src="image7" alt="Erythromycin" /></td>
<td>Erythromycin</td>
<td>0.819</td>
<td>J05AF02, J05FA01, S01AA17</td>
</tr>
<tr>
<td>392622</td>
<td><img src="image8" alt="Ritonavir" /></td>
<td>Ritonavir</td>
<td>0.809</td>
<td>HIV, Viral infection, J05AF03</td>
</tr>
<tr>
<td>2955</td>
<td><img src="image9" alt="Dapsone" /></td>
<td>Dapsone</td>
<td>0.806</td>
<td>Leprosy, J04BA02</td>
</tr>
<tr>
<td>5261104</td>
<td><img src="image10" alt="Perindopril" /></td>
<td>Perindopril</td>
<td>0.8</td>
<td>Hypoparathyroidism, A11CC07</td>
</tr>
<tr>
<td>4060</td>
<td><img src="image11" alt="Mechanotoin" /></td>
<td>Mechanotoin</td>
<td>0.793</td>
<td>Epilepsy, N03AB04</td>
</tr>
<tr>
<td>55283</td>
<td><img src="image12" alt="Itraconazole" /></td>
<td>Itraconazole</td>
<td>0.783</td>
<td>J02AC02</td>
</tr>
<tr>
<td>5472</td>
<td><img src="image13" alt="Ticlopidine" /></td>
<td>Ticlopidine</td>
<td>0.768</td>
<td>Stroke, B01AC05</td>
</tr>
</tbody>
</table>

**Target** | **p value** | **score** | **type** | **chemohub** |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C9</td>
<td>3.0e-04</td>
<td>597.54</td>
<td>predicted</td>
<td>see paths</td>
</tr>
<tr>
<td>CYP2A6</td>
<td>0.002</td>
<td>266.26</td>
<td>predicted</td>
<td>see paths</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>0.0027</td>
<td>228.29</td>
<td>predicted</td>
<td>see paths</td>
</tr>
<tr>
<td>HMOX1</td>
<td>0.0029</td>
<td>219.85</td>
<td>predicted</td>
<td>see paths</td>
</tr>
<tr>
<td>CYP2B6</td>
<td>0.0034</td>
<td>202.05</td>
<td>predicted</td>
<td>see paths</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.0038</td>
<td>190.21</td>
<td>approved interaction</td>
<td>see paths</td>
</tr>
<tr>
<td>CYP17A1</td>
<td>0.0091</td>
<td>118.54</td>
<td>predicted</td>
<td>see paths</td>
</tr>
<tr>
<td>GSTA2</td>
<td>0.0432</td>
<td>43.14</td>
<td>approved expression</td>
<td>see paths</td>
</tr>
<tr>
<td>HMOX2</td>
<td>0.0744</td>
<td>28.28</td>
<td>approved expression</td>
<td>see paths</td>
</tr>
<tr>
<td>ABAT</td>
<td>0.1524</td>
<td>14.75</td>
<td>approved interaction</td>
<td>see paths</td>
</tr>
<tr>
<td>HPRT1</td>
<td>0.1983</td>
<td>11.16</td>
<td>approved expression</td>
<td>see paths</td>
</tr>
<tr>
<td>PLOD2</td>
<td>0.4368</td>
<td>3.81</td>
<td>approved expression</td>
<td>see paths</td>
</tr>
</tbody>
</table>
There exists a set of drugs with known MI side effect, that interact with a certain subset of genetic targets that Rosiglitazone also interacts with.

Apolipoprotein E increases LDL cholesterol.

# Application – Profiling Adverse Events

<table>
<thead>
<tr>
<th></th>
<th>hERG</th>
<th>Atrial Fib</th>
<th>Long QT</th>
<th>Tachycardia</th>
<th>Bradycardia</th>
<th>Cardiotoxicity</th>
<th>Cardiac failure</th>
<th>Cardiomyopath.</th>
<th>Myocard Inf.</th>
<th>Tachycardia</th>
<th>Cardiotoxicity</th>
<th>Long QT</th>
<th>Atrial Fib</th>
<th>hERG</th>
<th>Long QT</th>
<th>Tachycardia</th>
<th>Cardiotoxicity</th>
<th>Bradycardia</th>
<th>Tachycardia</th>
<th>Cardiotoxicity</th>
<th>Tachycardia</th>
<th>Atrial Fib</th>
</tr>
</thead>
<tbody>
<tr>
<td>A10366245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A10366585</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A49585949</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A48480494</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspirin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGGREGATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Why is semantic data powerful?

- Breaking down data and domain silos
  - Chemistry – biology – toxicology – adverse event - endpoint
  - Molecular – patient
  - Public – commercial – proprietary

- Easy to repurpose existing and harvest new data
  - RDF format is standard
  - Separation of the data from the structure of the data

- Semantic networks -> biological networks
  - Systems chemical biology / network biology
  - Move away from naïve drug/target or target/endpoint
  - Hugely powerful algorithms in networking community
  - Prediction, hypothesis testing, interpretation
Proposal: Semantic Toolkit for AOPs

- New biomedical insights
- Integrated knowledge discovery processes
- Integrative tools & algorithms
- Accessible networks of semantically integrated data

Semantic Toolset for AOP’s

Compound-endpoint prediction and association finding

Comprehensive RDF triple store and ontologies for AOP’s
Comprehensive semantic store for AOPs

- Contains all public data of relevance, from compound to organism. As a start...
  - **OnTop**: PubMed, GO, KEGG, MeSH, NCI, UniProt, Entrez Gene, NCBO, CTD, ACToR, ToxRefDB, ToxMiner, ToxCat
  - **Chem2Bio2RDF/Chem2Bio2OWL**: 52 public datasets relating to compounds, genes, pathways, diseases and side effects
  - Other relevant sets – e.g. FDAERS, social media
- Ontologically mapped to concepts in AOPWiki
- SPARQL endpoint for searching

Compound-Endpoint prediction & association finding

• Predicting compound-endpoint associations with SLAP
  • Modified version of current compound-target algorithm
  • Association score and p-value

• Automatic generation of preliminary AOP networks
  • Using SLAP significant subnetwork between compound and endpoint
  • “starting point” for understanding potential AOPs

• Generation of literature supported association networks
  • More open-ended association finding and visualization

• Random-Walk methods
  • Most recent research at IU
Semantic Toolset for AOP’s

- Open toolset made available in association with AOP KB and integrated with other tools
- AOP prediction, exploration, hypothesis testing & application
Summary

- Semantic technologies becoming mainstream for big / complex data problems; increasing applications in science
- IU and EPA have demonstrated applicability of semantic technologies in chemical, biological data and for AOP’s
- AOP’s map particularly well onto the semantic approach
- Huge potential is realized when network and predictive algorithms are applied – the “semantic stack”
- Direct opportunity to engage semantic technologies in the emerging AOP KB / AOPWiki projects