The Open PHACTS Project: Progress and Future Sustainability

Lee Harland & Bryn Williams-Jones
Open PHACTS / ConnectedDiscovery

Tom Plasterer
AstraZeneca/Open PHACTS Rep
Fundamental issue:

- There is a *lot* of science outside your walls
- It’s a chaotic space
- Scientists want to find information quickly and easily
- Often they just “can’t get there” (or don’t even know where “there” is)
- And you have to manage it all (or not)
Pre-competitive Informatics:
Pharma are all accessing, processing, storing & re-processing external research data

Repeat @ each company

Lowering industry firewalls: pre-competitive informatics in drug discovery
Nature Reviews Drug Discovery (2009) 8, 701-708 doi:10.1038/nrd2944
The Innovative Medicines Initiative

- EC funded public-private partnership for pharmaceutical research
- Focus on key problems
  - Efficacy, Safety, Education & Training, Knowledge Management

The Open PHACTS Project

- Create a semantic integration hub ("Open Pharmacological Space")…
- Delivering services to support on-going drug discovery programs in pharma and public domain
- Not just another project; Leading academics in semantics, pharmacology and informatics, driven by solid industry business requirements
- 23 academic partners, 8 pharmaceutical companies, 3 biotechs
- Work split into clusters:
  - Technical Build
  - Scientific Drive
  - Community & Sustainability
“What is the selectivity profile of known p38 inhibitors?”

“Let me compare MW, logP and PSA for known oxidoreductase inhibitors”

“Find me compounds that inhibit targets in NFkB pathway assayed in only functional assays with a potency <1 μM”
Number | sum | Nr of 1 | Question
---|---|---|---
15 | 12 | 9 | All oxidoreductase inhibitors active <100nM in both human and mouse
18 | 14 | 8 | Given compound X, what is its predicted secondary pharmacology? What are the on and off, target safety concerns for a compound? What is the evidence and how reliable is that evidence (journal impact factor, KOL) for findings associated with a compound?
24 | 13 | 8 | Given a target find me all actives against that target. Find/predict polypharmacology of actives. Determine ADMET profile of actives.
32 | 13 | 8 | For a given interaction profile, give me compounds similar to it.
37 | 13 | 8 | The current Factor Xa lead series is characterised by substructure X. Retrieve all bioactivity data in serine protease assays for molecules that contain substructure X.
38 | 13 | 8 | Retrieve all experimental and clinical data (with options to match stereochemistry or not).
41 | 13 | 8 | A project is considering Protein Kinase C Alpha (PRKCA) as a target. What are all the compounds known to modulate the target directly? i.e. return all actives at the level of the target family (i.e. PKC). Give me all active compounds on the target.
44 | 13 | 8 | Give me all active compounds on the target.
46 | 13 | 8 | Give me the compound(s) which hit most specifically the multiple targets in a given pathway (disease).
59 | 14 | 8 | Identify all known protein-protein interaction inhibitors.
“Provenance Everywhere”
THE OPEN PHACTS DISCOVERY PLATFORM
### Present Content

#### Statistics of Datasets Loaded into Open PHACTS Version 1.3

<table>
<thead>
<tr>
<th>Source</th>
<th>Version</th>
<th>Supplier</th>
<th>Downloaded</th>
<th>Initial Records</th>
<th>Triples</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chembl</td>
<td>Chembl 16 RDF</td>
<td>EBI</td>
<td>25 June 2013</td>
<td>1,247,403 (~1,236,686 compounds, 9844 targets, 6243 target components, 873 protein classes)</td>
<td>304,420,681</td>
<td>77</td>
</tr>
<tr>
<td>DrugBank</td>
<td>Aug 2008</td>
<td>Bio2Rdf (www4.wiwi.fu-berlin.de)</td>
<td>08 Aug 2012</td>
<td>19,628 (~14,000 drugs, 5000 targets)</td>
<td>517,584</td>
<td>74</td>
</tr>
<tr>
<td>SwissProt, UniParc, UniRef</td>
<td>2013_06</td>
<td>SIB</td>
<td>2013_06</td>
<td></td>
<td>533,394,147</td>
<td>82</td>
</tr>
<tr>
<td>ENZYME</td>
<td>2013_07</td>
<td>SIB</td>
<td>2013_07</td>
<td>6,187</td>
<td>47,661</td>
<td>2</td>
</tr>
<tr>
<td>ChEBI</td>
<td>Release 104</td>
<td>EBI</td>
<td>19 June 2013</td>
<td>40,575</td>
<td>40,575</td>
<td>2</td>
</tr>
<tr>
<td>GeneOntology</td>
<td>Jan 21, 2013</td>
<td>GO</td>
<td>21 Jan 2013</td>
<td>38,137</td>
<td>1,265,273</td>
<td>26</td>
</tr>
<tr>
<td>GOA</td>
<td>2013</td>
<td>GO</td>
<td>09 Sept 2013</td>
<td>various species</td>
<td>23,489,501</td>
<td>15</td>
</tr>
<tr>
<td>WikiPathways</td>
<td>v0.7_1_20130710</td>
<td>Maastricht</td>
<td>10 July 2013</td>
<td>946</td>
<td>1,449,981</td>
<td>34</td>
</tr>
<tr>
<td>ChemSpider</td>
<td>Open PHACTS Chemistry Registry (OCRS)</td>
<td>Nov 11, 2013</td>
<td></td>
<td></td>
<td>tbc</td>
<td></td>
</tr>
<tr>
<td>ConceptWiki</td>
<td>version 1.3</td>
<td>NBIC</td>
<td>09 Sept 2013</td>
<td>2,828,966</td>
<td>3,739,884</td>
<td>1</td>
</tr>
</tbody>
</table>
Data Licensing Solution

Chose John Wilbanks as consultant

A framework built around STANDARD well-understood Creative Commons licences – and how they interoperate

Deal with the problems by:

- Interoperable licences
- Appropriate terms
- Declare expectations to users and data publishers
- One size won’t fit all requirements
Its easy to integrate, difficult to integrate well:
What Is Gleevec?

Imatinib Mesylate

**ChemSpider**
- About
- More Searches
- Web APIs

**Gleevec**
- ChemSpider ID: 5281
- Molecular Formula: C_{22}H_{25}N_{5}O
- Average mass: 463.5 g/mol
- Monocisotopic mass: 463.5 g/mol
- Systematic name: 4-{4-Methyl[1-[(4-methyl-1-[[4-[4-(3-Pyridinyl)]pyridin-2-yl]]imidazol-2-ylmethoxy]-phenoxy]methoxy]phenyl}imidazole

**Drugbank**
- Imatinib
- Mesylate
- Gleevec
- Glivec

**PubChem**
- Imatinib: 5281
- Mesylate: 571
- Gleevec: 7557
- Glivec: 7557

Active in 36 BioAssays
Tested in 673 BioAssays
Dynamic Equality

LinkSet#1 {
  chemspider:gleevec hasParent imatinib ...
  drugbank:gleevec exactMatch imatinib ...
}

chemspider:gleevec

drugbank:gleevec
Open PHACTS API

<table>
<thead>
<tr>
<th>Service</th>
<th>Endpoint</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Structure Exact Search</td>
<td>/structure/exact</td>
<td>GET</td>
</tr>
<tr>
<td>InchiKey to URL</td>
<td>/structure</td>
<td>GET</td>
</tr>
<tr>
<td>Inchi to URL</td>
<td>/structure</td>
<td>GET</td>
</tr>
<tr>
<td>Chemical Structure Similarity Search</td>
<td>/structure/similarity</td>
<td>GET</td>
</tr>
<tr>
<td>SMILES to URL</td>
<td>/structure</td>
<td>GET</td>
</tr>
<tr>
<td>Chemical Structure Substructure Search</td>
<td>/structure/substructure</td>
<td>GET</td>
</tr>
<tr>
<td>Get concept description</td>
<td>/getConceptDescription</td>
<td>GET</td>
</tr>
<tr>
<td>Map free text to a concept URL based on semantic tag</td>
<td>/search/byTag</td>
<td>GET</td>
</tr>
<tr>
<td>Map URL</td>
<td>/mapURL</td>
<td>GET</td>
</tr>
<tr>
<td>Map free text to a concept URL</td>
<td>/search/freetext</td>
<td>GET</td>
</tr>
<tr>
<td>Get ChEBI Ontology Class Members</td>
<td>/compound/chebi/members</td>
<td>GET</td>
</tr>
<tr>
<td>Get ChEBI Ontology Root Classes</td>
<td>/compound/chebi/root</td>
<td>GET</td>
</tr>
<tr>
<td>Get ChEBI Ontology Class</td>
<td>/compound/chebi/node</td>
<td>GET</td>
</tr>
<tr>
<td>ChEBI Class Pharmacology Count</td>
<td>/compound/chebi/pharmacology/count</td>
<td>GET</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>app_id</td>
<td></td>
<td>Your access application id</td>
</tr>
<tr>
<td>app_key</td>
<td></td>
<td>Your access application key</td>
</tr>
<tr>
<td>searchOptions.Molecule</td>
<td>(required)</td>
<td>A SMILES string. E.g. C(=O)(O)c1ccccc1C(=O)O</td>
</tr>
<tr>
<td>searchOptions.SimilarityType</td>
<td></td>
<td>0: Tanimoto ; 1: Tversky ; 2: Euclidian</td>
</tr>
<tr>
<td>searchOptions.Threshold</td>
<td></td>
<td>Double &lt;= 1.0</td>
</tr>
<tr>
<td>commonOptions.Complexity</td>
<td></td>
<td>(Not supported at the moment) 0: Any ; 1: Single ; 2: Multi</td>
</tr>
<tr>
<td>commonOptions.Isotopic</td>
<td></td>
<td>(Not supported at the moment) 0: Any ; 1: Labeled ; 2: NotLabeled</td>
</tr>
<tr>
<td>commonOptions.HasSpectra</td>
<td></td>
<td>(Not supported at the moment) Boolean</td>
</tr>
<tr>
<td>commonOptions.HasPatents</td>
<td></td>
<td>(Not supported at the moment) Boolean</td>
</tr>
<tr>
<td>resultOptions.Limit</td>
<td></td>
<td>Integer. Search limit. Specify how many results return back during the search. Default value: -1</td>
</tr>
<tr>
<td>resultOptions.Start</td>
<td></td>
<td>Integer. Return results starting the index. Default value: 0</td>
</tr>
<tr>
<td>resultOptions.Length</td>
<td></td>
<td>Integer. How many results should be returned starting from Start index. Default value: -1.</td>
</tr>
</tbody>
</table>
APPS
Mitogen-activated protein kinase 14 (Homo sapiens)

<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Target Name</th>
<th>Target Organism</th>
<th>Assay Organism</th>
<th>Assay Description</th>
<th>Activity Type</th>
<th>Relation</th>
<th>Value</th>
<th>Units</th>
<th>Mol Weight</th>
<th>SPEILS</th>
<th>InChI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2-(Furan-2-yl)-7H-pyrrolo[5,4-c]quinoline-5-carboxamide</td>
<td>Adenosine receptor A3 (Homo sapiens)</td>
<td>Homo sapiens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-[(2-Furan-2-yl)-7H-pyrrolo[5,4-c]quinoline-5-carboxamide]</td>
<td>Adenosine receptor A3 (Homo sapiens)</td>
<td>Homo sapiens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-[(2-Furan-2-yl)-7H-pyrrolo[5,4-c]quinoline-5-carboxamide]</td>
<td>Adenosine receptor A3 (Homo sapiens)</td>
<td>Homo sapiens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-[(2-Furan-2-yl)-7H-pyrrolo[5,4-c]quinoline-5-carboxamide]</td>
<td>Adenosine receptor A3 (Homo sapiens)</td>
<td>Homo sapiens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[http://explorer.openphacts.org](http://explorer.openphacts.org)
<table>
<thead>
<tr>
<th>Compound</th>
<th>Target</th>
<th>Assay</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2'-Hydroxy-4-bromochalcone</td>
<td>Homo sapiens</td>
<td>Induction of quinone reductase activity in human MCF7 cells</td>
<td>Activity = 303.151, C1=CC=C(O(=C1)O)=C=O/C=C=O=C=C=C=C=O/C=C=O</td>
</tr>
<tr>
<td>2'-hydroxychalcone</td>
<td>Homo sapiens</td>
<td>Induction of quinone reductase activity in human MCF7 cells</td>
<td>Activity = 224.266, C1=CC=C(C=C=C=C(=C(=C)C)=C(C=C)=O/C2=C=O</td>
</tr>
<tr>
<td>(2E)-1-(2-hydroxyphenyl)-3-(pyridin-2-yl)prop-2-en-1-one</td>
<td>Homo sapiens</td>
<td>Induction of quinone reductase activity in human MCF7 cells</td>
<td>Activity = 225.243, C1=CC=C(C(=C1)O)=C=O/C=C=O=C=C=C=C=O/N2/0</td>
</tr>
</tbody>
</table>
## Advanced analytics

<table>
<thead>
<tr>
<th>Application</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChemBioNavigator</td>
<td>Navigating at the interface of chemical and biological data with sorting and plotting options</td>
</tr>
<tr>
<td>TargetDossier</td>
<td>Interconnecting Open PHACTS with multiple target centric services. Exploring target similarity using diverse criteria</td>
</tr>
<tr>
<td>PharmaTrek</td>
<td>Interactive Polypharmacology space of experimental annotations</td>
</tr>
<tr>
<td>UTOPIA</td>
<td>Semantic enrichment of scientific PDFs</td>
</tr>
</tbody>
</table>

## Predictions

<table>
<thead>
<tr>
<th>Application</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARFIELD</td>
<td>Prediction of target pharmacology based on the Similar Ensemble Approach</td>
</tr>
<tr>
<td>eTOX connector</td>
<td>Automatic extraction of data for building predictive toxicology models in eTOX project</td>
</tr>
</tbody>
</table>
Uptake at AstraZeneca: a Use Case

Applying BioAssay Ontology to facilitate HTS analysis

Linda Zander Balderud
Ola Engkvist

Chemistry Innovation Centre, Discovery Sciences
AstraZeneca
Assay Informatics project
Benefits in Adopting BioAssay Ontology (BAO)

• Common language for assay annotation
• Improved project success analyses based on assay technologies
• Better understand the impact of technology artifacts like frequent hitters
• Assay design and screening cascade support during assay development in early projects
• Improved capability to perform combined data mining of internal and public data

FLIPR Tetra High Throughput Cellular Screening System (from Molecular Devices)
The BioAssay Ontology (BAO)
Computational Science, University of Miami, USA

Domain:
• Assay design
• Assay format
• Detection technology
• Meta target
• Endpoint
• Perturbagen

BioAssay Ontology imports:
• **NCBI taxonomy** - organism names and IDs
• **Uniprot** - protein target names and IDs
• **Unit Ontology** - concentration and time unit terms
• **Ontology of Biomedical Investigation** – descriptions of biological assays
• **Gene Ontology** - biological processes
• **Cell Line Ontology** - cell line names
• **CL** – cell types
• **UBERON** – anatomical entities
• **PATO** – cell phenotype
• **SAR connect** – target classifications
Migration to BAO
Annotation of HTS assays

Manual annotation of protocols

**HTS assay: reporter gene assay**
- Assay method: reporter gene method: beta lactamase induction
- Detection technology: FRET
- Bioassay: beta lactamase assay

  - *Assay kit: LiveBLAzer FRET - B/G Loading Kit*
  - *Wavelength: ex 405 em 460, 535*

- Biological process
- Disease

**HTS assay: FLIPR**
- Assay method: molecular redistribution determination assay
- Detection technology: fluorescence intensity
- Bioassay: calcium redistribution assay

  - *Assay kit: Fluo-8 No Wash Calcium Assay Kit*
  - *Wavelength: ex 480 em 530*

- Biological process
- Disease

Over 900 PubChem assays have been annotated by the BioAssay Ontology team
412 in-house HTS assays since 2005 have been annotated according to the BioAssay Ontology. The assay design and technology of the annotated assays were analyzed together with 239 primary assays from PubChem. The analyzed PubChem assays are biochemical assays, assays detected by luminescence and/or assays using GPCR targets.

From the annotated assays, 515 assays were using human targets and combined 311 different human targets were represented in the study.

15 of the in-house targets were also screened in at least one PubChem assay. Eight of these were GPCR targets.
Assay Development Support

Detection Technology of AZ and PubChem Biochemical Assays
Assay Development Support

Assay design of in-house and PubChem GPCR HTS

One explanation for the low usage of cAMP redistribution method among the annotated PubChem assays could be that no class B GPCRs have been screened.
Sustaining The Project

The Open PHACTS Foundation
Kick-Starting Sustainability

Open PHACTS

Industry
Grants
Collaboration
API Users

Apps

API
The Open PHACTS Foundation

OPF is a not-for-profit membership organisation, supporting the Open PHACTS Discovery Platform:
A sustainable, open, vibrant and interoperable information infrastructure for applied life science research and development.

To reduce the barriers to drug discovery in industry, academia and for small businesses, the Open PHACTS Discovery Platform provides tools and services to interact with multiple integrated and publicly available data sources. To integrate this data, extensive cross-referencing of scientific concepts is needed across all databases.

The Open PHACTS Foundation ensures the sustainability of the Open PHACTS Discovery Platform infrastructure and acts as a hub for relevant scientific research and development.