Introduction To Open PHACTS

Lee Harland
ConnectedDiscovery / Pfizer

Using the Power of Open PHACTS
Open PHACTS 4th Community Workshop, London, 22-23 April 2013
Introduction

- Aim of the next 2 days
- Brief background to the project
- What the “Open PHACTS Discovery Platform” actually is
- Things we worry about (so you don’t have to)

(Tomorrow)

- The “Ecosystem”, how we see it working
Today Is Launch Day!

#opslaunch
Over the next two days we will discuss:

- Open PHACTS as a unique opportunity, creating information standards, unified access points and a “data market place” across pharma companies (customers)
- What the system can offer to aid your business and what you can deliver to your users
- **Day 1**: Details on the system, the data content, the API and some use cases
- **Day 2**: User, business and strategic views. Getting your input!

At the end of the workshop, we would like to leave with:

- All participants feeling they are “up to speed” on what the Open PHACTS discovery platform is and what it offers
- Feedback on where you see the opportunities and any barriers
- Specific actions, how can we get Open PHACTS working for you?
PROJECT BACKGROUND
Pre-competitive Informatics:
Pharma are all accessing, processing, storing & re-processing external research data

Lowering industry firewalls: pre-competitive informatics in drug discovery
Nature Reviews Drug Discovery (2009) 8, 701-708 doi:10.1038/nrd2944
“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 \(<1 \mu M\)”
## Business Question Driven Approach

<table>
<thead>
<tr>
<th>Number</th>
<th>sum</th>
<th>Nr of</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>12</td>
<td>9</td>
<td>All oxidoreductase inhibitors active &lt;100nM in both human and mouse</td>
</tr>
<tr>
<td>18</td>
<td>14</td>
<td>8</td>
<td>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?</td>
</tr>
<tr>
<td>24</td>
<td>13</td>
<td>8</td>
<td>Given a target find me all actives against that target. Find/predict polypharmacology of actives. Determine ADMET profile of actives.</td>
</tr>
<tr>
<td>32</td>
<td>13</td>
<td>8</td>
<td>For a given interaction profile, give me compounds similar to it.</td>
</tr>
<tr>
<td>37</td>
<td>13</td>
<td>8</td>
<td>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.</td>
</tr>
<tr>
<td>38</td>
<td>13</td>
<td>8</td>
<td>Retrieve all experimental and clinical data for a given list of compounds defined by their chemical structure (with options to match stereochemistry or not).</td>
</tr>
<tr>
<td>41</td>
<td>13</td>
<td>8</td>
<td>A project is considering Protein Kinase C Alpha (PRKCA) as a target. What are all the compounds known to modulate the target directly? What are the compounds that may modulate the target directly? i.e. return all cmpds active in assays where the resolution is at least at the level of the target family (i.e. PKC) both from structured assay databases and the literature.</td>
</tr>
<tr>
<td>44</td>
<td>13</td>
<td>8</td>
<td>Give me all active compounds on a given target with the relevant assay data</td>
</tr>
<tr>
<td>46</td>
<td>13</td>
<td>8</td>
<td>Give me the compound(s) which hit most specifically the multiple targets in a given pathway (disease)</td>
</tr>
<tr>
<td>59</td>
<td>14</td>
<td>8</td>
<td>Identify all known protein-protein interaction inhibitors</td>
</tr>
</tbody>
</table>

... paper coming very soon in DDT
A Precompetitive Knowledge Framework

Pharma Needs
- Sustainability
- Stability
- Security
- Management / Governance
- Data Mining Services / Algorithms
- Community KD Innovation

Integration
- Mapping & Populating
- Architecture
- Interfaces & Services

Inputs
- Vocabularies & Identifiers (URIs)
- Content Structured & Unstructured
Open PHACTS Elements
Why a Semantic Technology approach?

- Different formats, different structures, different vocabularies, different concepts, different meaning
- Data should be structured (not ASCII)
- Structure should be data-oriented (not HTML)
- Meaning of data should be clear (not XML)
- Reusable mappings between data are needed (not XSLT)
- Avoid extensive schema rewriting (not data warehouses)
- Data should have standard APIs (not Flickr)
- Synergise with many public efforts
WHAT WE THINK ABOUT
How standards proliferate:
(see: A/C chargers, character encodings, instant messaging, etc.)

Situation: There are 14 competing standards.

14?! Ridiculous! We need to develop one universal standard that covers everyone’s use cases.

Yeah!

Soon:

Situation: There are 15 competing standards.

http://imgs.xkcd.com/comics/standards.png
Adoption Of Standards

- Basic Semantic web standards
  - SPARQL 1.1, RDF(S), SKOS
- Dataset descriptions
  - Vocabulary of Interlinked Datasets (VoID)
  - VoID linkset descriptions
- QUDT Quantities, Units, Dimensions and Types
- Provenance
  - W3C PROV, PAV, Nanopublications
- BioPortal, ConceptWiki, ChEMBL, identifiers.org, Uniprot, ChemSpider
The Seven Deadly Sins of Bioinformatics

Professor Carole Goble The University of Manchester, UK
The myGrid project, OMII-UK

Andy Law's Third Law

“The number of unique identifiers assigned to an individual is never less than the number of Institutions involved in the study”... and is frequently many, many more.

http://bioinformatics.roslin.ac.uk/lawslaws.html
Let the IMS take the strain....
It's easy to integrate, difficult to integrate well:

Type a compound name:

glee

- Gleevec
- Gleevec
Dynamic Equality

Strict

Analysing

Relaxed

Browsing
"Provenance Everywhere"
To Conclude:

- Welcome!
- Open PHACTS is much more than software, it’s a whole new approach (but they software is pretty cool too!)
- There’s a lot of detail in the technology we won’t present, but we’re happy to talk about (esp. on day 2)
- We want to enhance your business – its in both our interests

Keeping up to date

- Email us at pmu@openphacts.org to be added to the ops-dev-public mailing list
- Follow @Open_PHACTS on twitter for service info etc