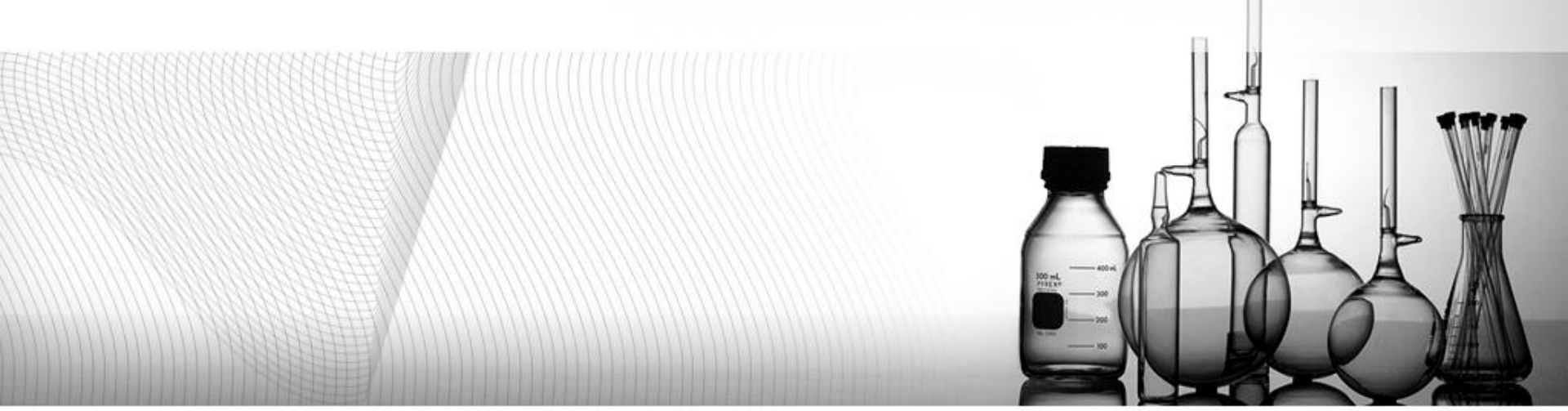


Open PHACTS
Open Pharmacological Space

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?

Open PHACTS
Open Pharmacological Space

PROJECT BACKGROUND

Pre-competitive Informatics:

Pharma are all accessing, processing, storing & re-processing external research data

Literature
Patents
PubChem
Genbank

Databases

Downloads

Data Integration

Data Analysis

Firewalled Databases

X Repeat @ each company

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

Open PHACTS
Open Pharmacological Space

http://openphacts.org
pmu@openphacts.org
@Open_PHACTS

Open PHACTS Project Partners

www.openphacts.org

Pfizer Limited – Coordinator
Universität Wien – Managing entity
Technical University of Denmark
University of Hamburg, Center for Bioinformatics
BioSolveIT GmBH
Consorci Mar Parc de Salut de Barcelona
Leiden University Medical Centre
Royal Society of Chemistry
Vrije Universiteit Amsterdam

Spanish National Cancer Research Centre
University of Manchester
Maastricht University
Aqnowledge
University of Santiago de Compostela
Rheinische Friedrich-Wilhelms-Universität Bonn
AstraZeneca
GlaxoSmithKline
Esteve

Novartis
Merck Serono
H. Lundbeck A/S
Eli Lilly
Netherlands Bioinformatics Centre
Swiss Institute of Bioinformatics
ConnectedDiscovery
EMBL-European Bioinformatics Institute
Janssen
OpenLink

Pfizer
universität wien
RSC Advancing the Chemical Sciences
130 jaar VU
AstraZeneca
gsk GlaxoSmithKline
DTU
UHH Universität Hamburg
cnio
MANCHESTER 1824
ESTEVE closer to you
NOVARTIS
Maastricht University
AQnowledge Semantics for Science
MERCK
Lundbeck
BioSolve
Parc de Salut MAR
USC
universitätbonn
Lilly
LUMC
janssen
EMBL-EBI
connecteddiscovery
nbic
OPENLINK SOFTWARE Making Technology Work For You
SIB Swiss Institute of Bioinformatics

Open PHACTS
Open Pharmacological Space

Using the Power of Open PHACTS, London, 22-23 April 2013

"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"

ChEMBL | DrugBank | Gene Ontology | Wikipathways | GeneGo

ChEBI | UniProt | UMLS | GVKBio

ConceptWiki | ChemSpider | TrialTrove | TR Integrity

Business Question Driven Approach

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 for a given list of compounds defined by their chemical structure (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? 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.
44	13	8	Give me all active compounds on a given target with the relevant assay data
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

… 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

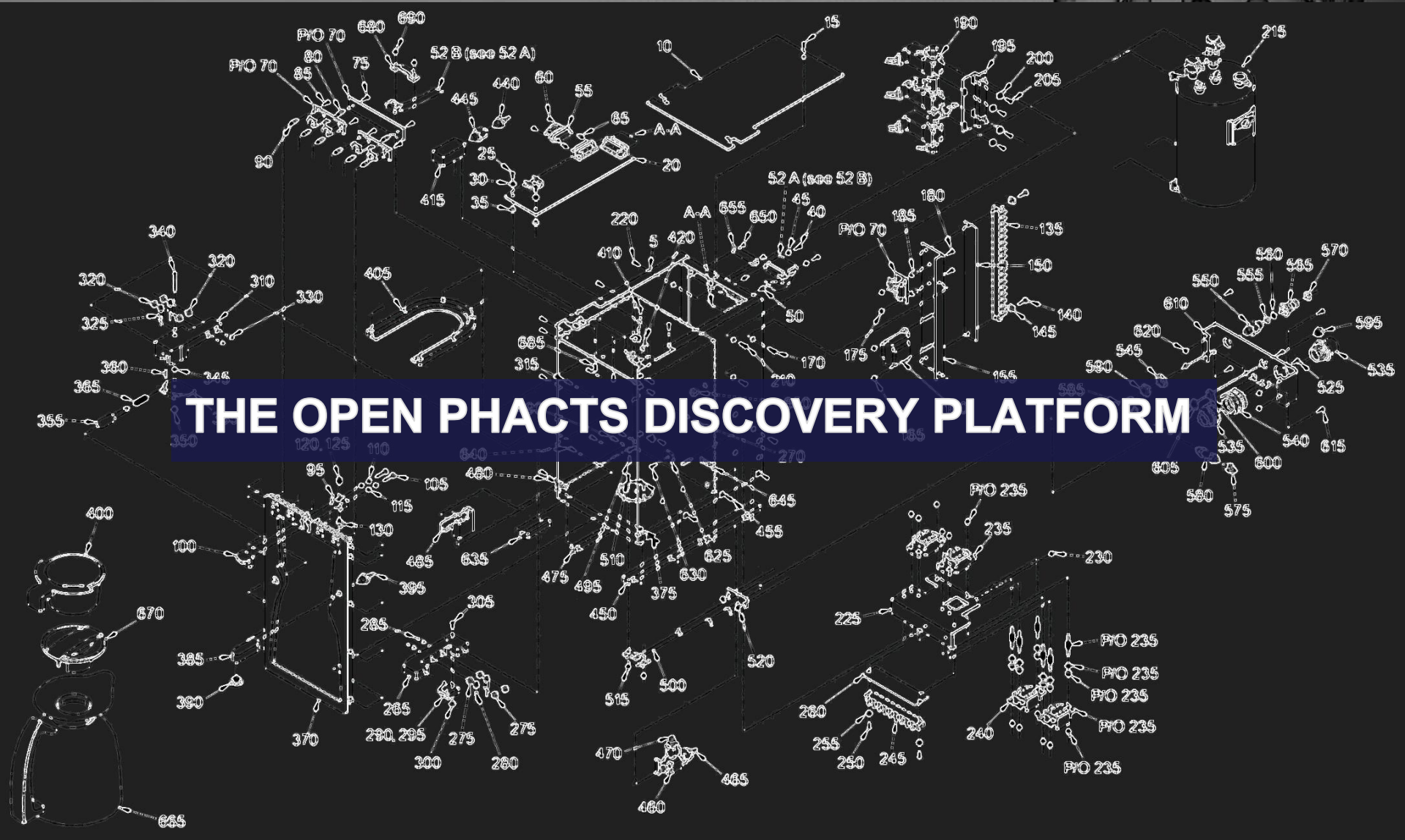
- Mapping & Populating
- Architecture
- Interfaces & Services

Inputs

- Vocabularies & Identifiers (URIs)
- Content Structured & Unstructured

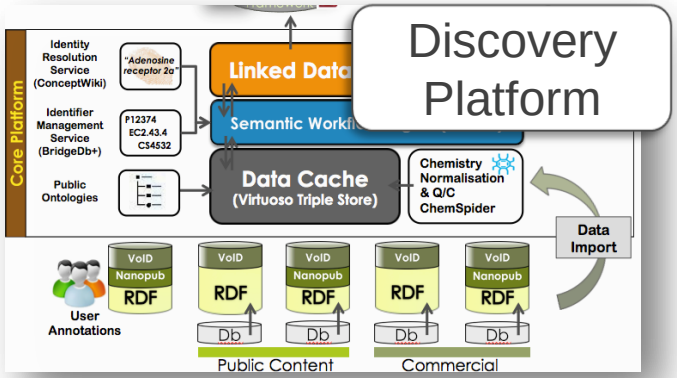
Open PHACTS

Open Pharmacological Space



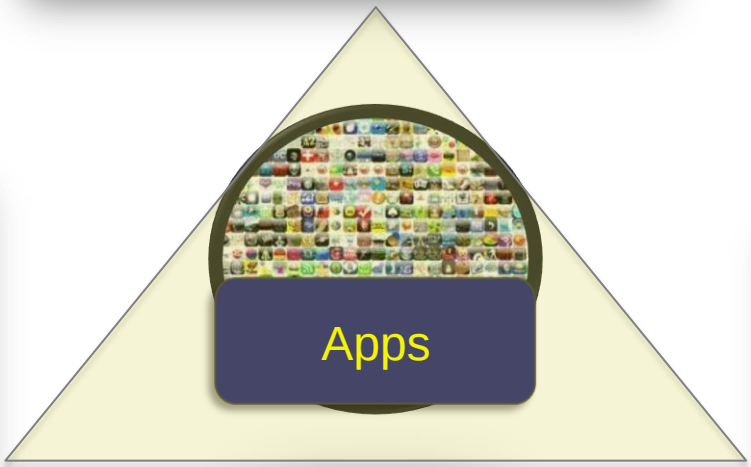
THE OPEN PHACTS DISCOVERY PLATFORM

Open PHACTS Elements



API

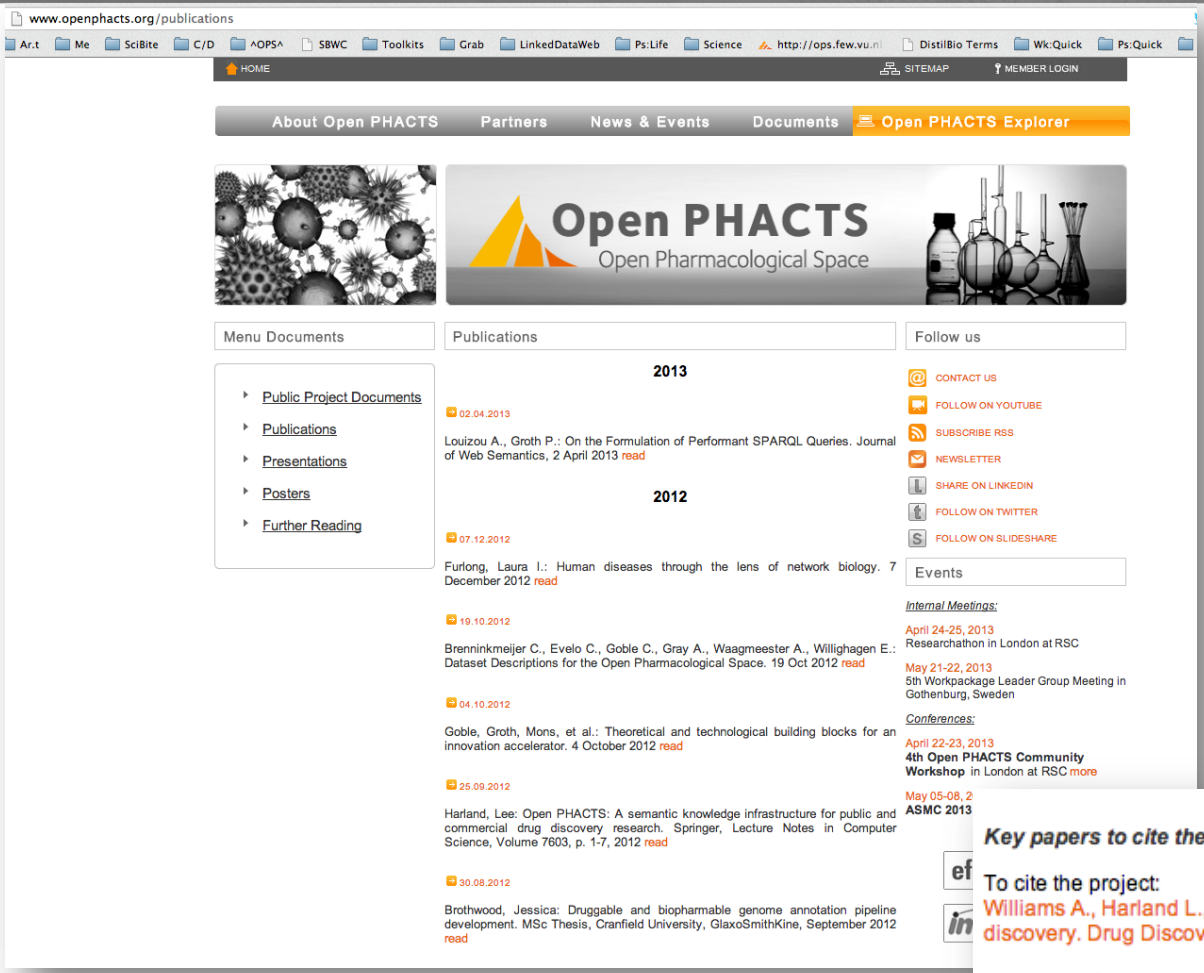
```
w_uri .
quiv_target .
et_name ;
        void:inDataset <http://www.conceptwiki.
?equiv_target dc:title ?target_name;
        ops:target_organism ?target_organism ;
        ops:targetOfAssay ?equiv_assay ;
        void:inDataset <http://data.kasabi.com/
ops:targetOfAssay owl:inverseOf chembl:hasTarge
?equiv_assay chembl:organism ?assay_organism ;
        chembl:hasDescription ?assay_descriptio
        ops:assayOfActivity ?activity_uri .
ops:assayOfActivity owl:inverseOf chembl:onAssa
?activity_uri chembl:type ?std_type ;
```



Standards

Open PHACTS
Open Pharmacological Space

Using the Power of Open PHACTS, London, 22-23 April 2013



Key papers to cite the project

To cite the project:
Williams A., Harland L., Groth P,. et al.: Open PHACTS: Semantic interoperability for drug discovery. Drug Discovery Today, June 06, 2012

To cite the Open PHACTS Discovery Platform:
Gray A., Groth P., Loizou A., et al.: Applying Linked Data Approaches to Pharmacology: Architectural Decisions and Implementation. Semantic Web Journal, April 18, 2012

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

You!

Core Platform

Identity Resolution Service

"Adenosine receptor 2a"

Identifier Management Service

P12374
EC2.43.4
CS4532

Linked Data API (RDF/XML, TTL, JSON)

Semantic Workflow Engine

Data Cache
(Virtuoso Triple Store)

Domain Specific Services

Chemistry Registration Normalisation & Q/C

Indexing

Public Ontologies

VoID
RDF
Db

VoID
Nanopub
RDF
Db

Public Content

VoID
RDF
Db

VoID
Nanopub
RDF
Db

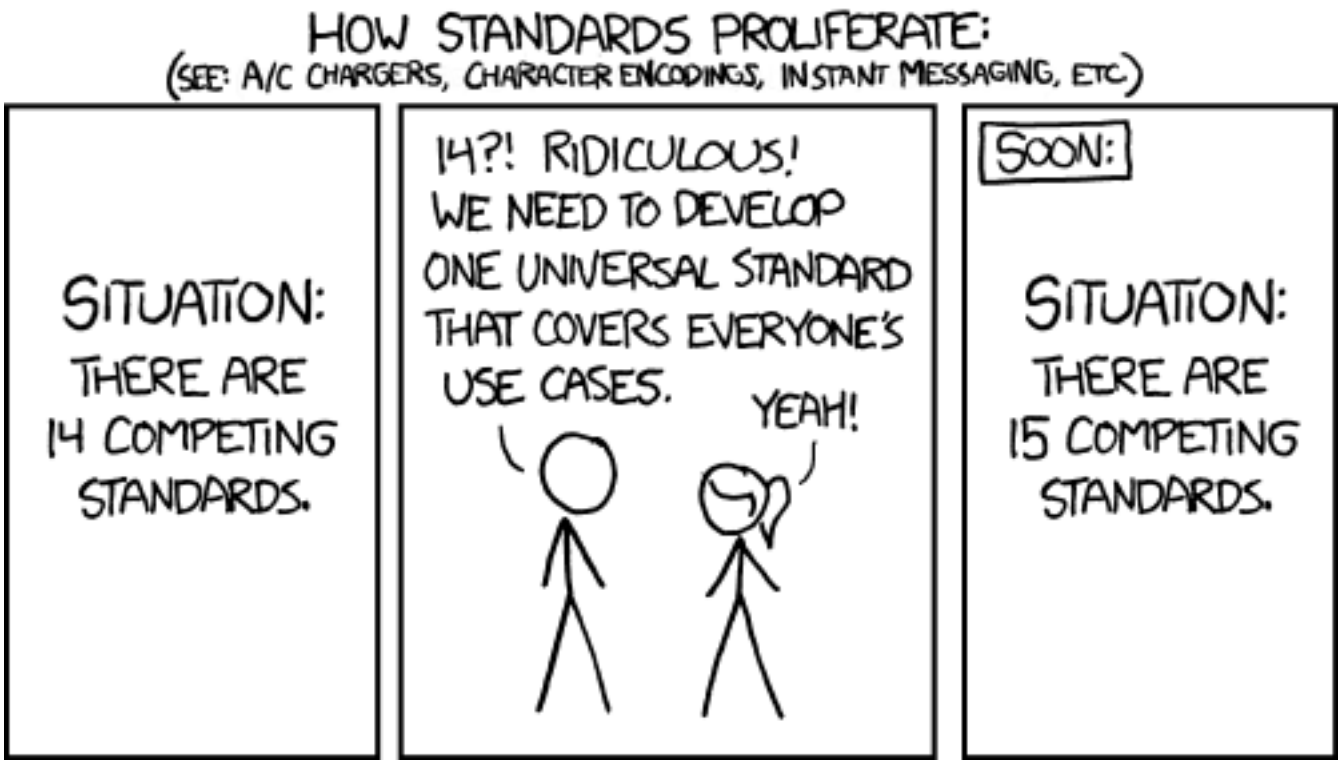
Commercial

VoID
Nanopub
RDF

User Annotations

Open PHACTS
Open Pharmacological Space

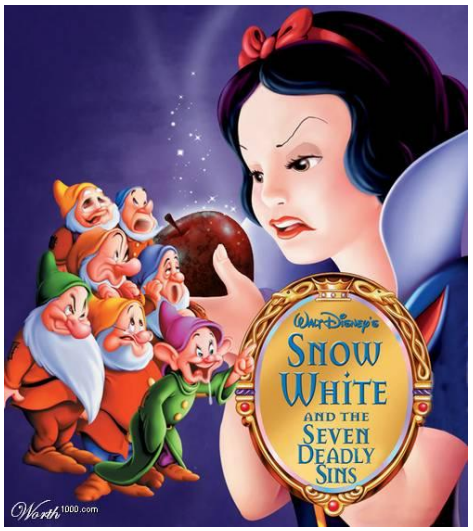
WHAT WE THINK ABOUT



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

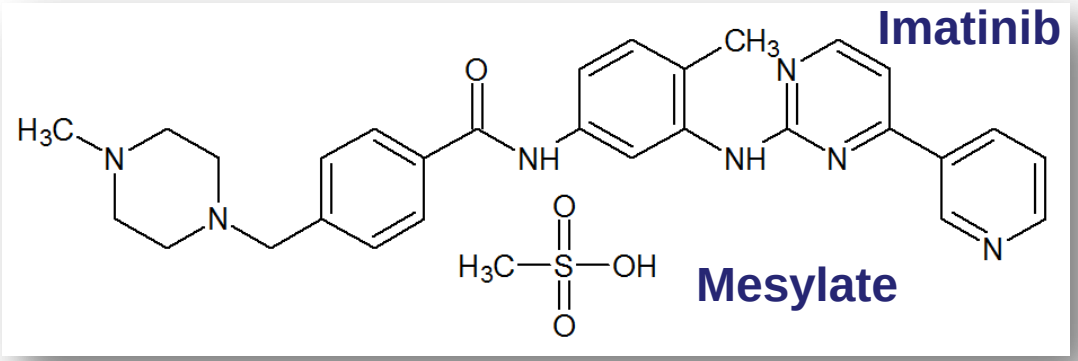
Professor Carole Goble The University of Manchester, UK
The myGrid project, OMII-UK

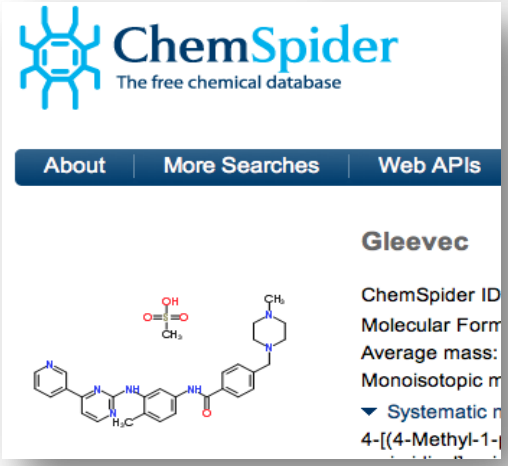
Andy Law's Third Law

“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.

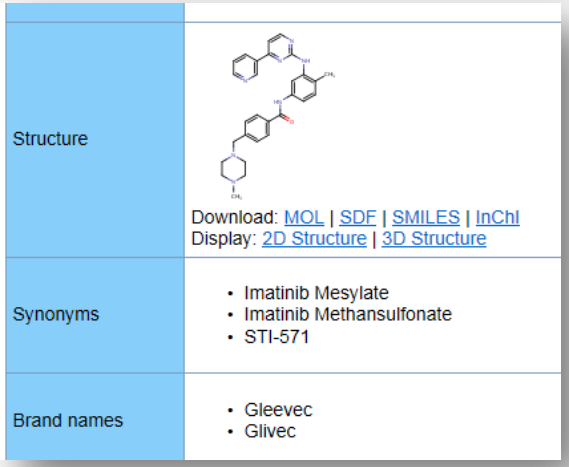
http://bioinformatics.roslin.ac.uk/lawslaws.html

Let the IMS take the strain….

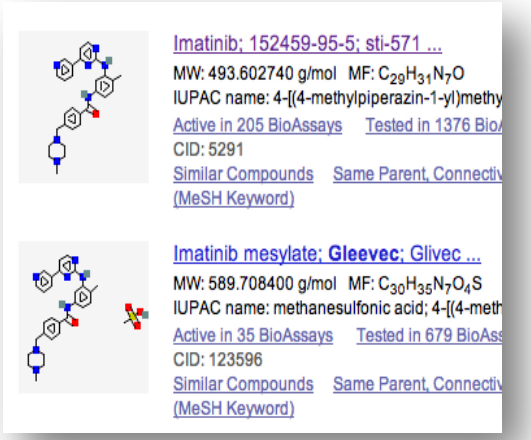




ChemSpider



Drugbank



PubChem

Its easy to integrate, difficult to integrate well:

Dynamic Equality

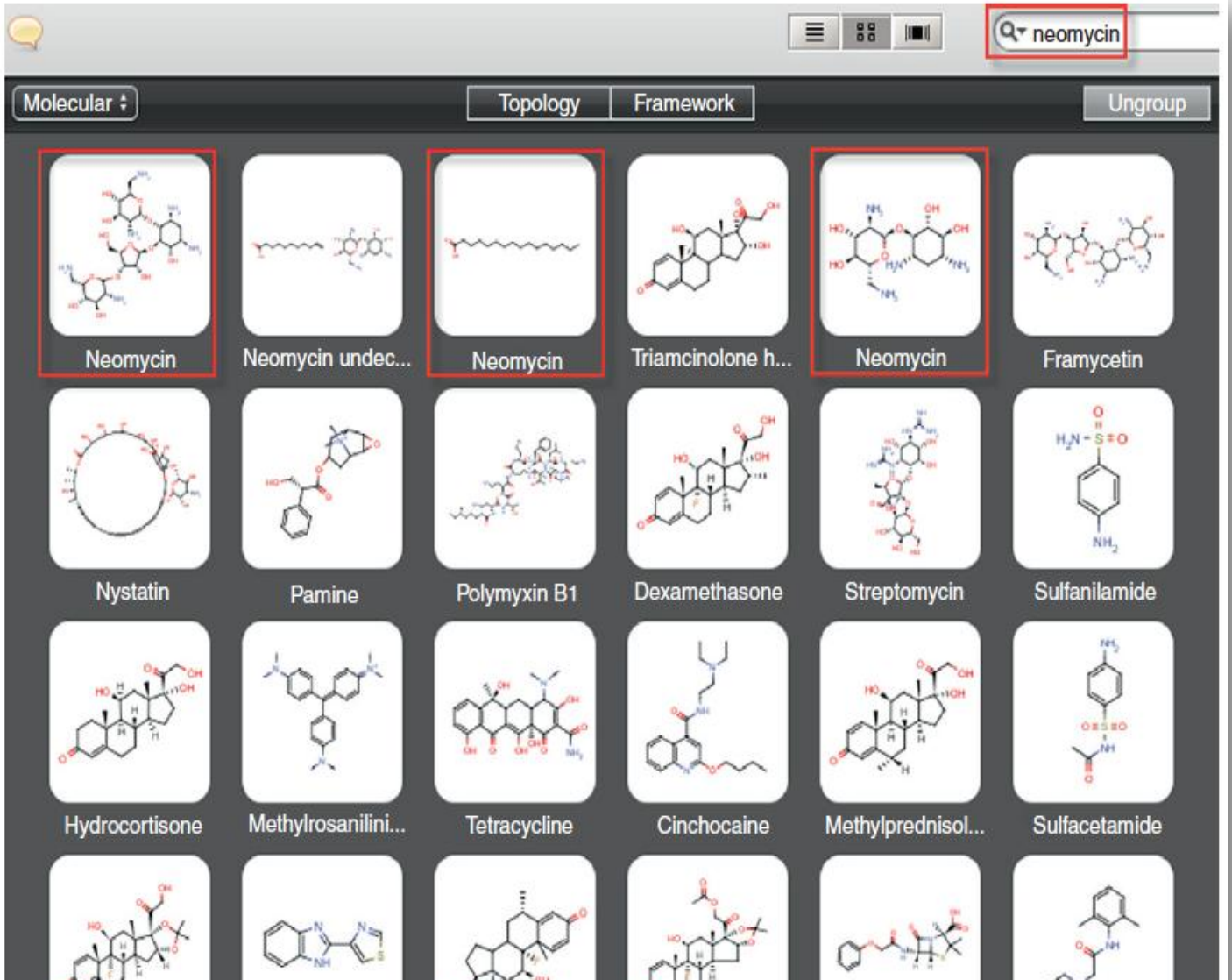
Strict

Relaxed

Analysing

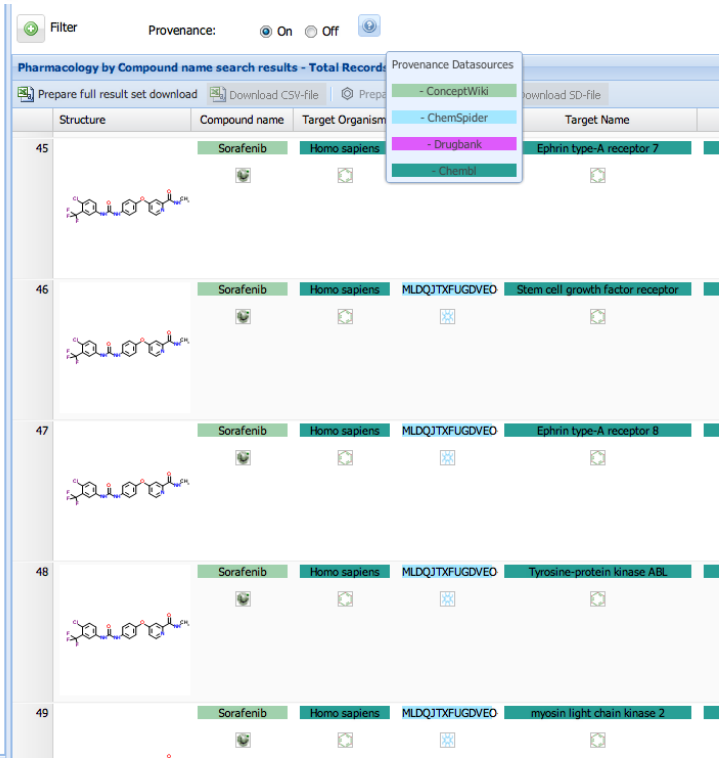
Browsing

Open PHACTS
Open Pharmacological Space



“Provenance Everywhere”


```
▼<result format="linked-data-api" version="0.2" href="http://ops.few.vu.nl/compound?u
 4ed1d7bdf9d5">
  ▼<primaryTopic href="http://www.conceptwiki.org/concept/38932552-111f-4a4e-
     <prefLabel xml:lang="en">Sorafenib</prefLabel>
    ▼<exactMatch>
       <item href="http://www.conceptwiki.org/concept/38932552-111f-4a4e-a46a-
      ▼<item href="http://rdf.chemspider.com/187440">
          <ro5_violations datatype="double">0.0</ro5_violations>
          <psa datatype="double">9.235E-18</psa>
          <logp datatype="double">4.818</logp>
          <hbd datatype="double">3.0</hbd>
          <hba datatype="double">7.0</hba>
          <inDataset href="http://www.chemspider.com"/>
          <smiles>0</smiles>
          <inchikey>MLDQJTXFUGDVEO-UHFFFAOYSA-N</inchikey>
          <inchi>0</inchi>
        </item>
      ▼<item href="http://data.kasabi.com/dataset/chembl-rdf/chemblid/CHEMBL13
          <inDataset href="http://data.kasabi.com/dataset/chembl-rdf"/>
          <rtb datatype="int">6</rtb>
          <mw_freebase datatype="double">464.825</mw_freebase>
          <molform>C21H16ClF3N4O3</molform>
          <full_mwt datatype="double">464.825</full_mwt>
        </item>
      ▼<item href="http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB(
          <toxicity>0</toxicity>
          <proteinBinding>99.5%</proteinBinding>
          <description>0</description>
          <biotransformation>0</biotransformation>
          <inDataset href="http://linkedlifedata.com/resource/drugbank"/>
        </item>
      </exactMatch>
      <inDataset href="http://www.conceptwiki.org"/>
  </primaryTopic>
  <extendedMetadataVersion href="http://ops.few.vu.nl/compound?uri=http%3A%2
  4ed1d7bdf9d5&_metadata=all%2Cviews%2Cformats%2Cexecution%2Cbindings%2Csite
  <definition href="http://ops.few.vu.nl/api-config"/>
</result>
```

<inDataset href=“http://rdf.chemspider.com/void.rdf#chemSpiderDataset” />

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