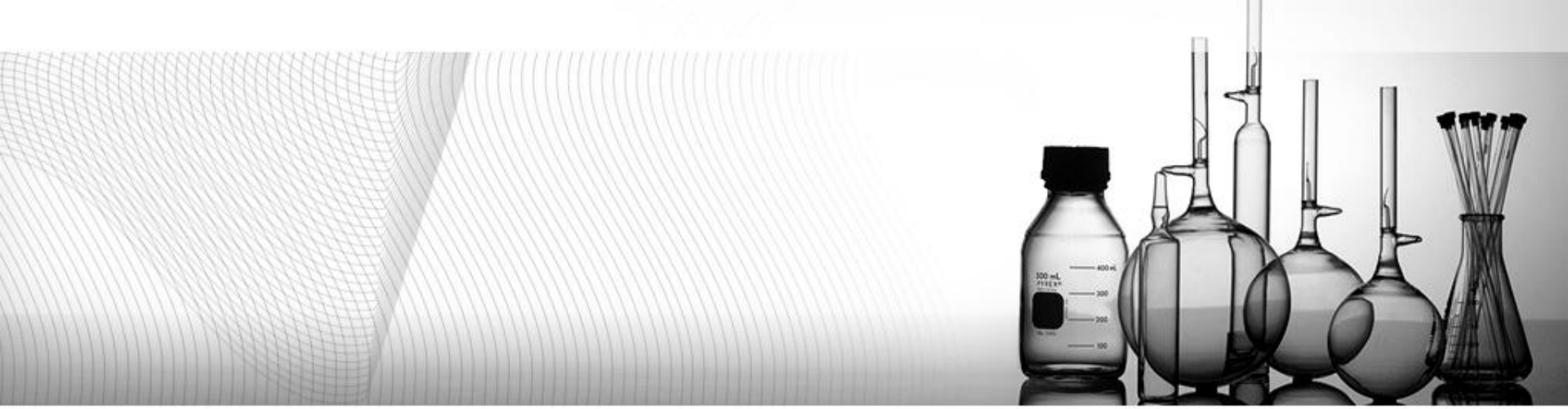


Open PHACTS
Open Pharmacological Space

Open PHACTS: a precompetitive infrastructure for pharmacological research

Bryn Williams-Jones

Open PHACTS
Open Pharmacological Space

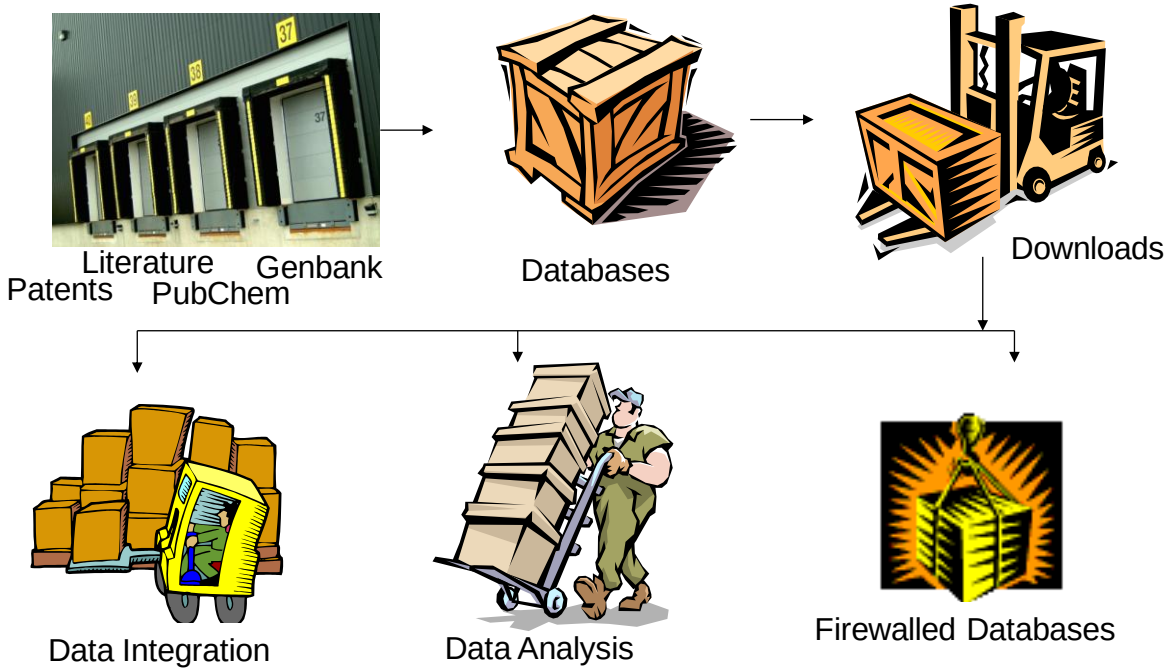
Fundamental issue:

- 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)

Open PHACTS
Open Pharmacological Space

Pre-competitive Informatics:

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



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

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")…* to start with, moving to broader biomedical topics later
- *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
- >120 people. Delivered production system, live, useful (and being used) within 18 months
- Strong, active participation from pharma companies – not passengers

Open PHACTS
Open Pharmacological Space

Open PHACTS Project Partners

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

pmu@openphacts.org
@Open_PHACTS
Open PHACTS

Open PHACTS

Open Pharmacological Space

"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

Open PHACTS
Open Pharmacological Space

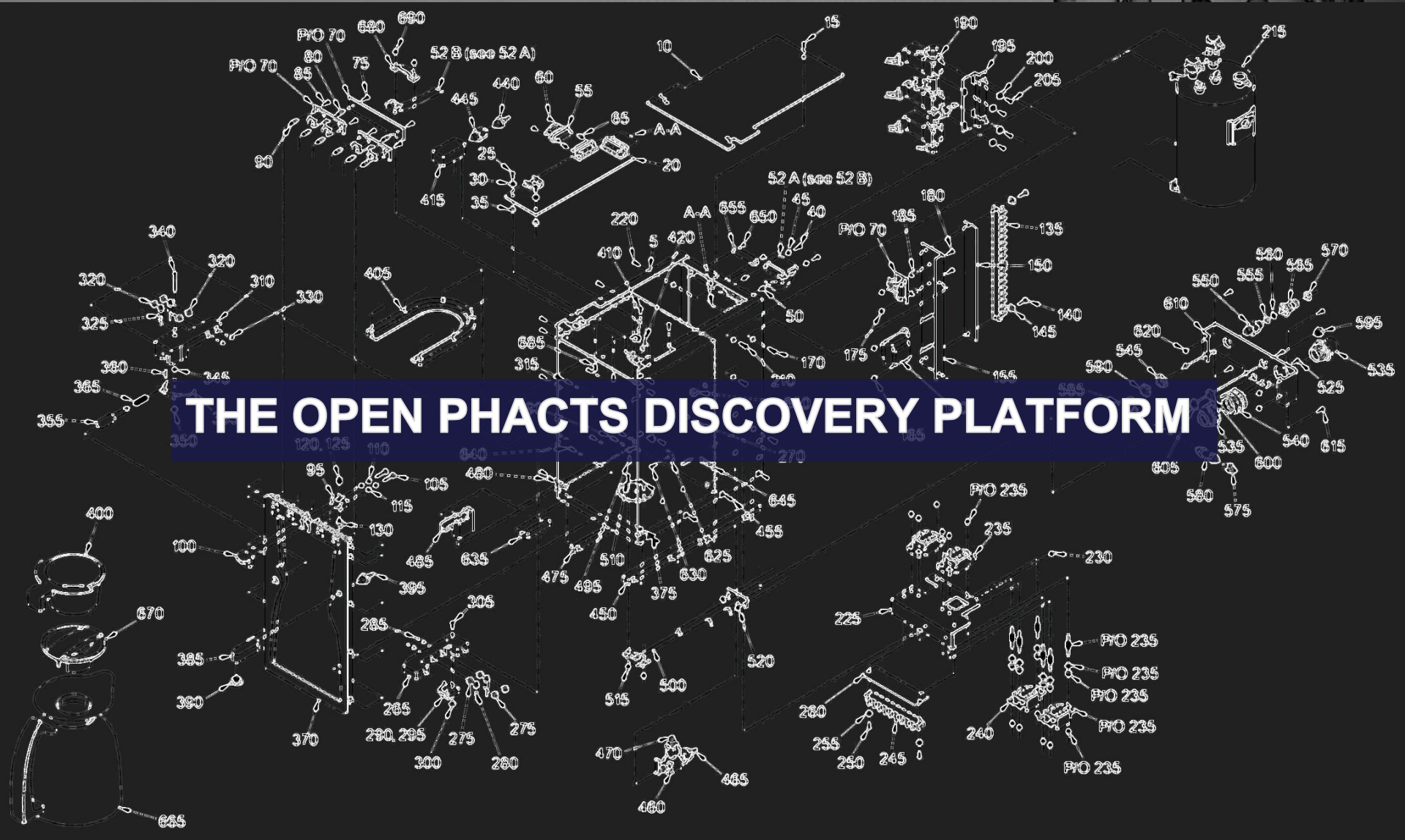
Business Question Driven Approach

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

Drug Discovery Today…paper in press

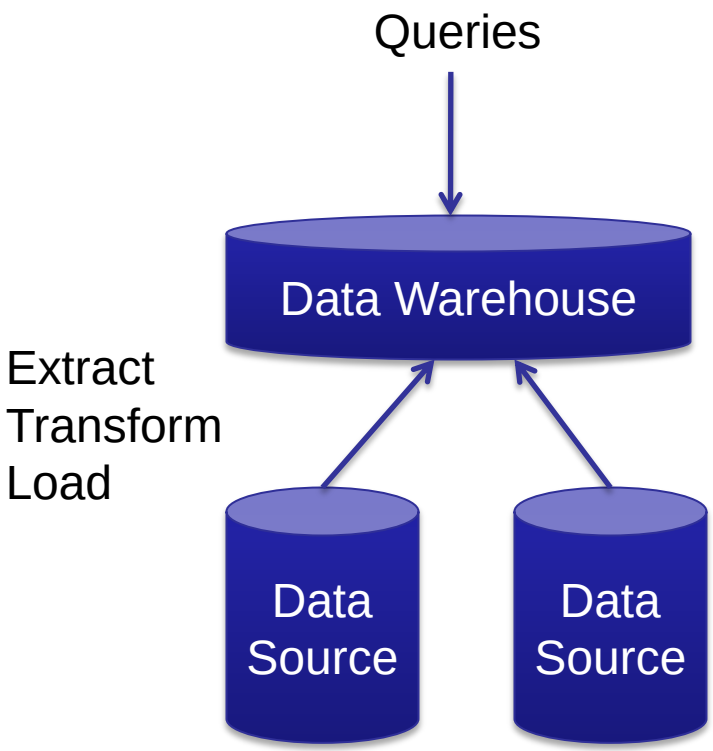
Open PHACTS

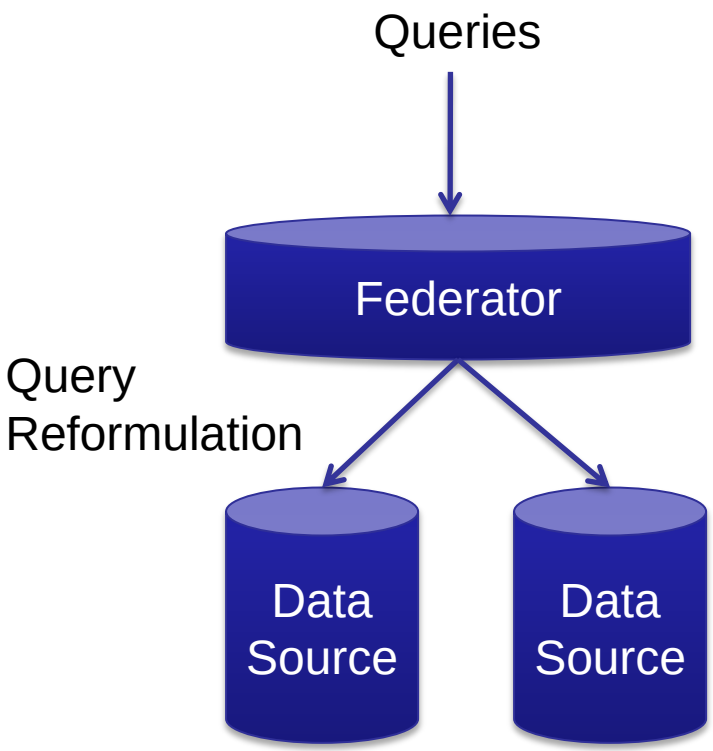
Open Pharmacological Space

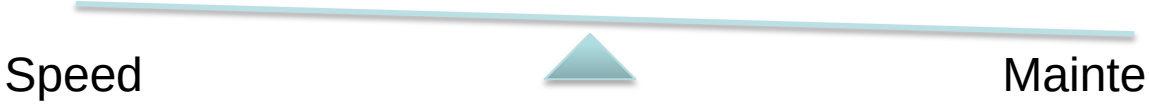


THE OPEN PHACTS DISCOVERY PLATFORM

Data Integration Approaches







Open PHACTS
Open Pharmacological Space

The Open PHACTS Approach

- **A Hybrid Model**
  - Cache data locally that requires computing over. “Cache” rather than “Warehouse” – obliterate & rebuild at will (think Google)
  - Bring in ancillary properties by wiring in web-services
  - Both provide an opportunity for secure data (see later)

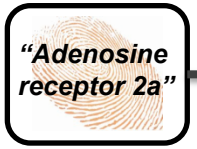
- **Use Semantic Technology**
  - “Schema Free” means you don’t need to change your warehouse when the data changes (as just happened for ChEMBL)
  - Open standards increase opportunities (not tied to any particular vendor) and shared, interoperable data models (code public & internal data to the same abstract standard)

Open PHACTS
Open Pharmacological Space

Apps

Core Platform

Identity Resolution Service



Identifier Management Service

P12374
EC2.43.4
CS4532

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

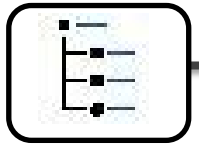
Semantic Workflow Engine

Data Cache
(Virtuoso Triple Store)

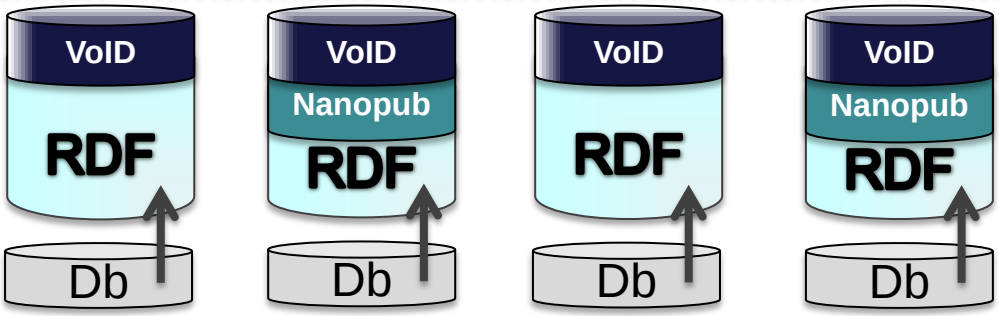
Domain Specific Services

Chemistry
Registration
Normalisation
& Q/C

Indexing



Public Ontologies



Public Content

Commercial

User Annotations

Open PHACTS
Open Pharmacological Space

Present Content - Pharmacology

Source	Initial Records	Triples	Properties
Chembl	1,149,792 ~1,091,462 cmpds ~8845 targets	146,079,194	17 cmpds 13 targets
DrugBank	19,628 ~14,000 drugs ~5000 targets	517,584	74
UniProt	536,789	156,569,764	78
ENZYME	6,187	73,838	2
ChEBI	35,584	905,189	
GO/GOA	38,137	24,574,774	42
ChemSpider/ACD	1,194,437	161,336,857	22 ACD, 4 CS
ConceptWiki	2,828,966	3,739,884	1

> 1 billion triples

Open PHACTS
Open Pharmacological Space

Infrastructure

Hardware (development)

- 2 x Intel Xeon E5-2640 - 384 GB DDR3 1333MHz RAM - 1.5 TB SSD - 3TB 7200rpm

Triple Store

- Virtuoso 7 column store
- Shown to scale to > 100 billion triples
- Project aiming for 30-50 billion mark

Network

- AMX-IS
- Extensive memcache

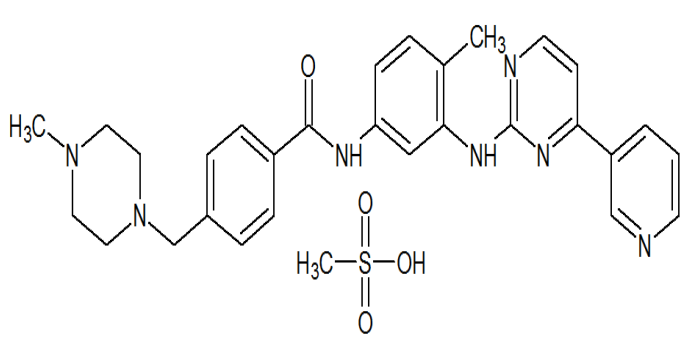
Semantic Workflow Engine

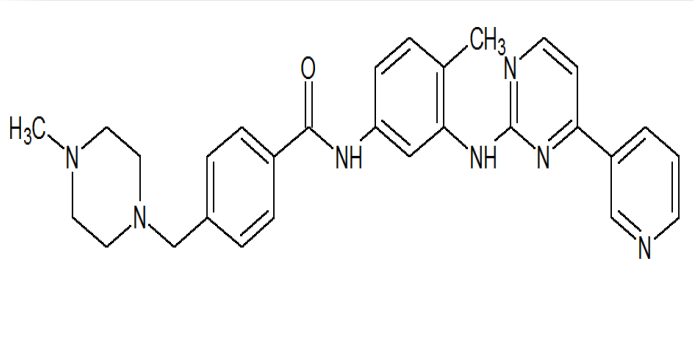
Data Cache
(Virtuoso Triple Store)

OPENLINK SOFTWARE
Making Technology Work For You®

Open PHACTS
Open Pharmacological Space

Are These Two Molecules The Same(*)





*Really: Is it sensible to combine data associated with these two molecules?

Open PHACTS
Open Pharmacological Space

Chemistry Registration

- Existing chemistry registration system uses standard ChemSpider deposition system: includes low-level structure validation and manual curation service by RSC staff.

- New Registration System in Development
  - Utilizes ChemSpider Validation and Standardization platform including collapsing tautomers
  - Utilizes FDA rule set as basis for standardization (GSK lead)
  - Will generate Open PHACTS identifier (OPS ID)

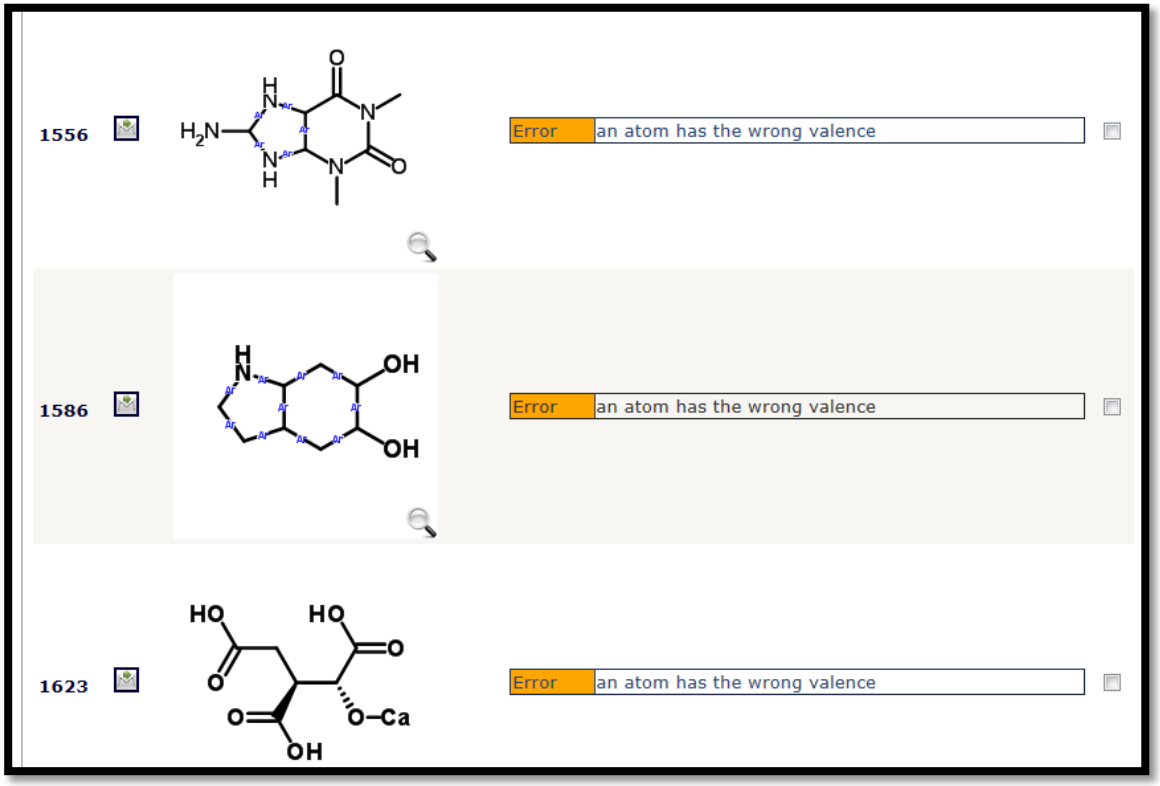
Chemistry Registration Normalisation & Q/C

ChemSpider
The free chemical database

Open PHACTS
Open Pharmacological Space

RSC | Advancing the Chemical Sciences

Quality Assurance

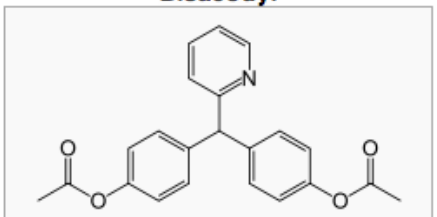


ChemSpider Validation & Standardization Platform
http://bit.ly/NZF5VB

Open PHACTS

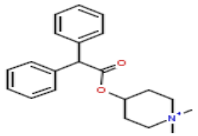
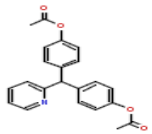
Open Pharmacological Space

Bisacodyl



Systematic (IUPAC) name

(pyridin-2-ylmethanediyl)dibenzene-4,1-diyl diacetate
OR
4,4'-(pyridin-2-ylmethylene)bis(4,1-phenylene) diacetate

Compound	Synonym
 CHEMBL168067	Bisacodyl
 CHEMBL942	Bisacodyl

Concept Wiki

Home Search

Terminology

Save Add Click on any of the labels to start editting

Term	Authority
Milnacipran	**Community, ChemSpider**
Enter A Term Here	Community
(1R,2S)-2-(Aminomethyl)-N,N-diethyl-1-phenylcyclopropanecarboxamide	Community, ChemSpider
cis-(±)-2-(Aminomethyl)-N,N-diethyl-1-phenylcyclopropanecarboxamide	Community, ChemSpider
cyclopropanecarboxamide, 2-(aminomethyl)-N,N-diethyl-1-phenyl-, (1R,2S)-	Community, ChemSpider
92623-85-3	Community, ChemSpider

Open PHACTS
Open Pharmacological Space

Developer Centric API

Open PHACTS

leeharland Dashboard | Account | Logout

Developer Home | Want help? | Documentation | Get my API keys! | Featured Apps | Search

OpenPHACTS API Active Docs

The response template for each operation colour coded as follows:

- Required elements that always return a **single value**.
- Required elements that return **either a single value or an array**.
- Optional elements that always return a **single value**
- Optional elements that return **either a single value or an array**.

Operations

OpenPHACTS API

Chemical Structure Exact Search	/structure/exact	GET
InchiKey to URL	/structure	GET
Inchi to URL	/structure	GET
Chemical Structure Similarity Search	/structure/similarity	GET
SMILES to URL	/structure	GET
Chemical Structure Substructure Search	/structure/substructure	GET
Get concept description	/getConceptDescription	GET
Map free text to a concept URL based on semantic tag	/search/byTag	GET
Map free text to a concept URL	/search/freetext	GET

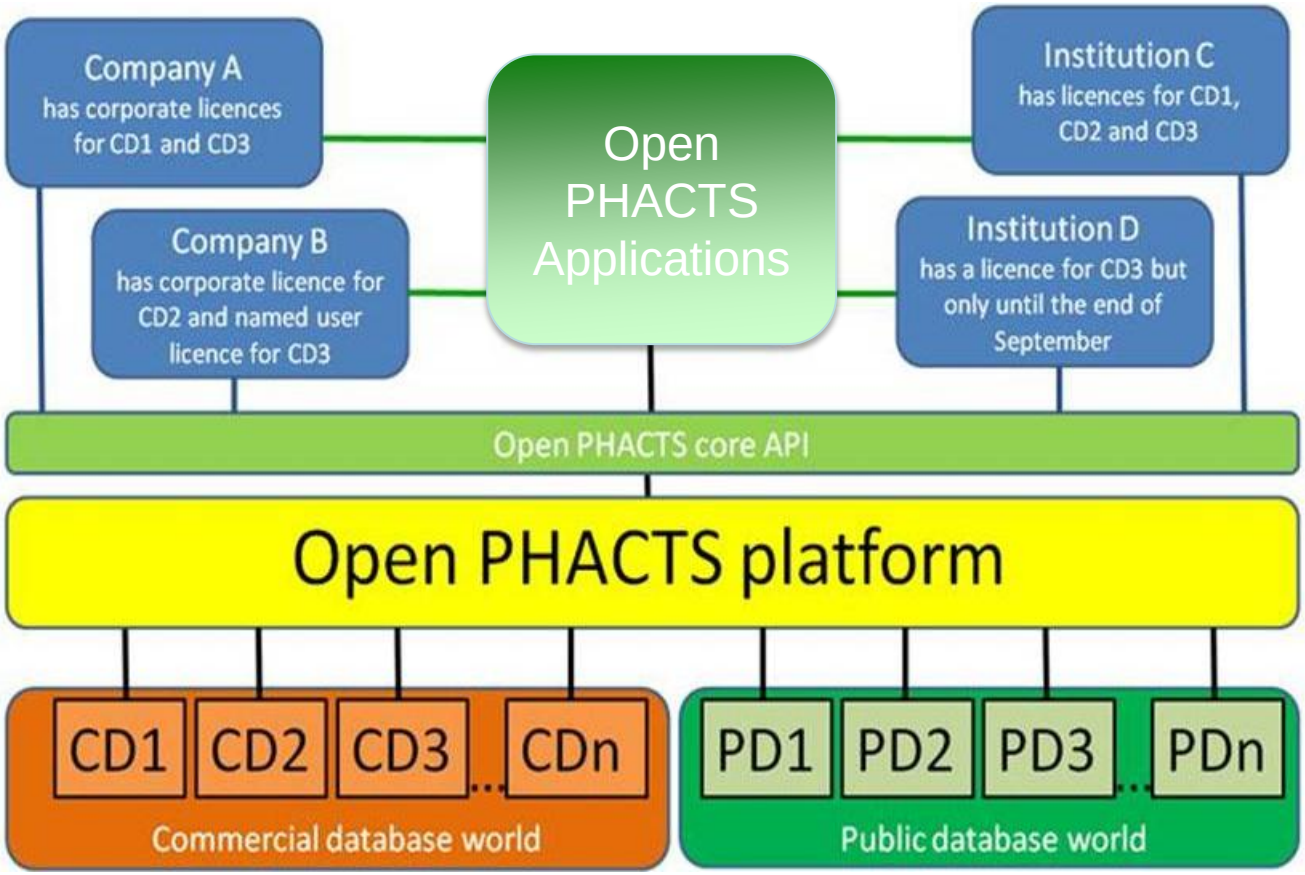
dev.openphacts.org

Open PHACTS
Open Pharmacological Space

Benefits Of the Open PHACTS API

- **Access to a wide range of interconnected data** – easily jump between pharmacology, chemistry, disease, pathways and other databases without having to perform complex mapping operations
- **Query by data type, not by data source** (“Protein Information” not “Uniprot Information”)
- **API queries that seamlessly connect data** (for instance the Pharmacology query draws data from Chembl, ChemSpider, ConceptWiki and Drugbank)
- **Strong chemistry representation** – all chemicals reprocessed via Open PHACTS chemical registry to ensure consistency across databases
- **Built using open community standards**, not an ad-hoc solution. Developed in conjuction with 8 major pharma (so your app will speak their language!)
- **Simple, flexible data-joining** (join compound data ignoring salt forms, join protein data ignoring species)
- **Provenance everywhere** – every single data point tagged with source, version, author, etc
- **Nanopublication-enabled**. Access to a rich dataset of established and emerging biomedical “assertions”
- **Community-curation** tools to enhance and correct content
- Access to **a rich application network** (many different App builders)
- **Toolkits** to support many different languages, workflow engines and user applications
- **Developer-friendly** JSON/XML methods. **Consistent API** for multiple services
- **Seamless data upgrades.** We manage updates so you don’t have to
- **Professionally Hosted** (Continually Monitored)
- **Private and secure**, suitable for confidential analyses
- **Neutral Party. Active & still growing** through a unique public-private partnership

Commercial Data Pilot (aka Authentication)



Open PHACTS
Open Pharmacological Space

Creating A Biomedical "App Store" .. How far have we come?

Open PHACTS

Open Pharmacological Space

The Open PHACTS Foundation

Open PHACTS | The Foundation | Services | Membership | Apps | Contact

The Application Ecosystem / *Which applications use the Open PHACTS API and how do I build my own?*

All applications
Browser Advanced Analytics Predictions Workflow Integration

Open PHACTS

The Open PHACTS Explorer allows browsing of the Open PHACTS Discovery Platform in an intuitive and interactive manner.

Developed by the University of Manchester and University of Vienna

ChemBioNavigator

The ChemBioNavigator allows the user to visualise the chemical and biological space of a molecule group in a chemically-aware manner.

Developed by the University of Hamburg and BioSolveIT GmbH

PHARMATREK

PharmaTrek allows the navigation of pharmacological space in a flexible and interactive way.

Developed by the Consorci Mar Parc de Salut de Barcelona (PSMAR)

Target Dossier

Interconnecting Open PHACTS with multiple target centric services. Exploring target similarity using diverse criteria.

Developed by the Spanish National Cancer Research Centre (CNIO)

utopia

Utopia Documents semantic enriches scientific articles, allowing the user to experience the convenience and reliability of a PDF with the flexibility and power of the web.

Developed by the University of Manchester

GARField

The Graph-Activity-Relationship visualisation Field (GARField) tool allows the intuitive prediction of target pharmacology based on the Similar Ensemble Approach.

Developed by the Technical University of Denmark

collector

The Collector uses allows the extraction of data to build QSAR predictive models with data from the eTox project.

Developed by the Consorci Mar Parc de Salut de Barcelona (PSMAR) as part of the eTox project

accelrys® Pipeline Pilot

Accelrys' Pipeline Pilot workflow tool can be used to query the Open PHACTS API. A repository of useful Pipeline Pilot components and workflows has been developed.

Open PHACTS - Pipeline Pilot Community

KNIME

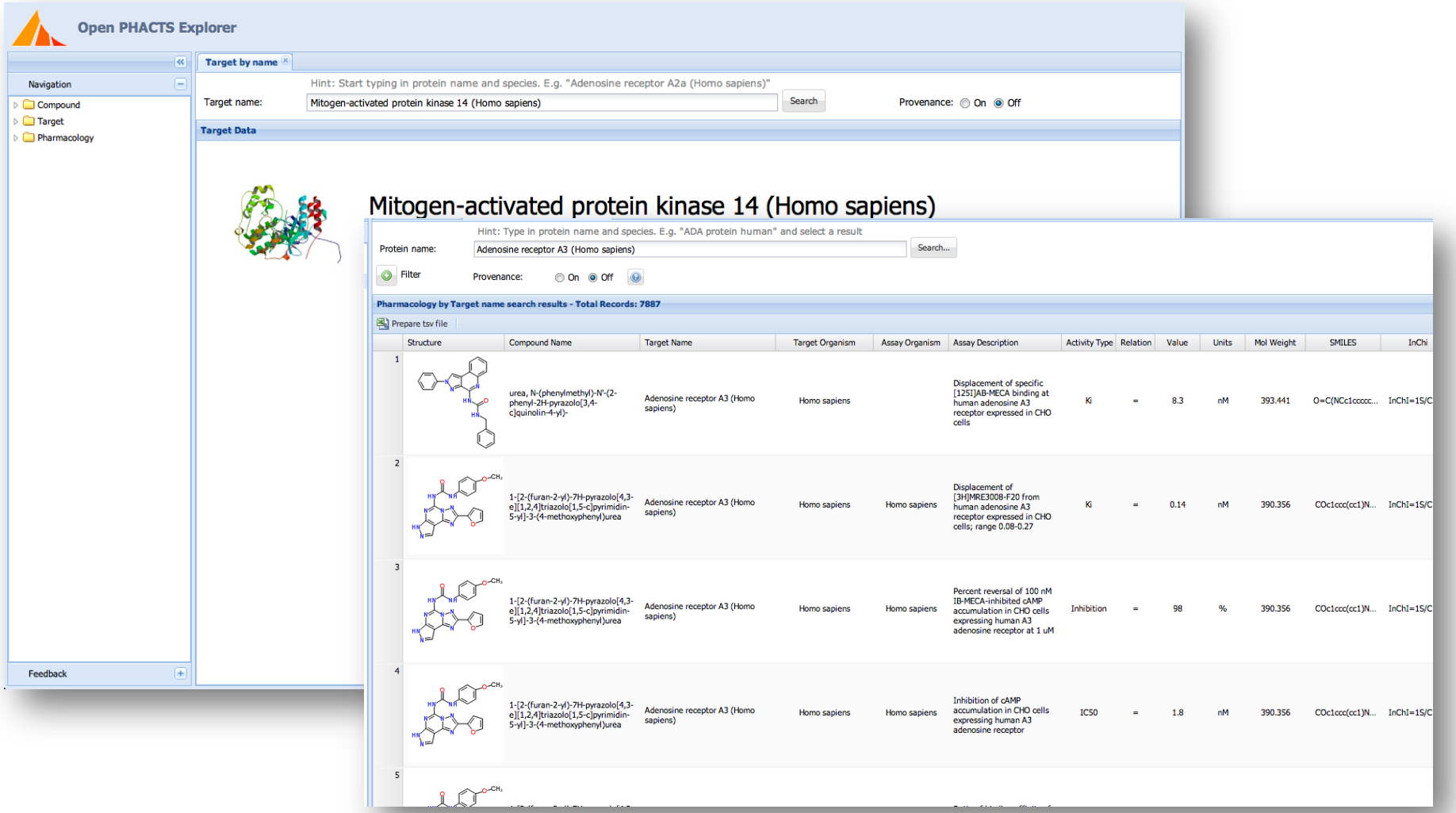
KNIME's workflow tool can be used to interact with the Open PHACTS API. A repository of useful KNIME components and workflows has been developed.

Open PHACTS - KNIME Community

Excel

Open PHACTS

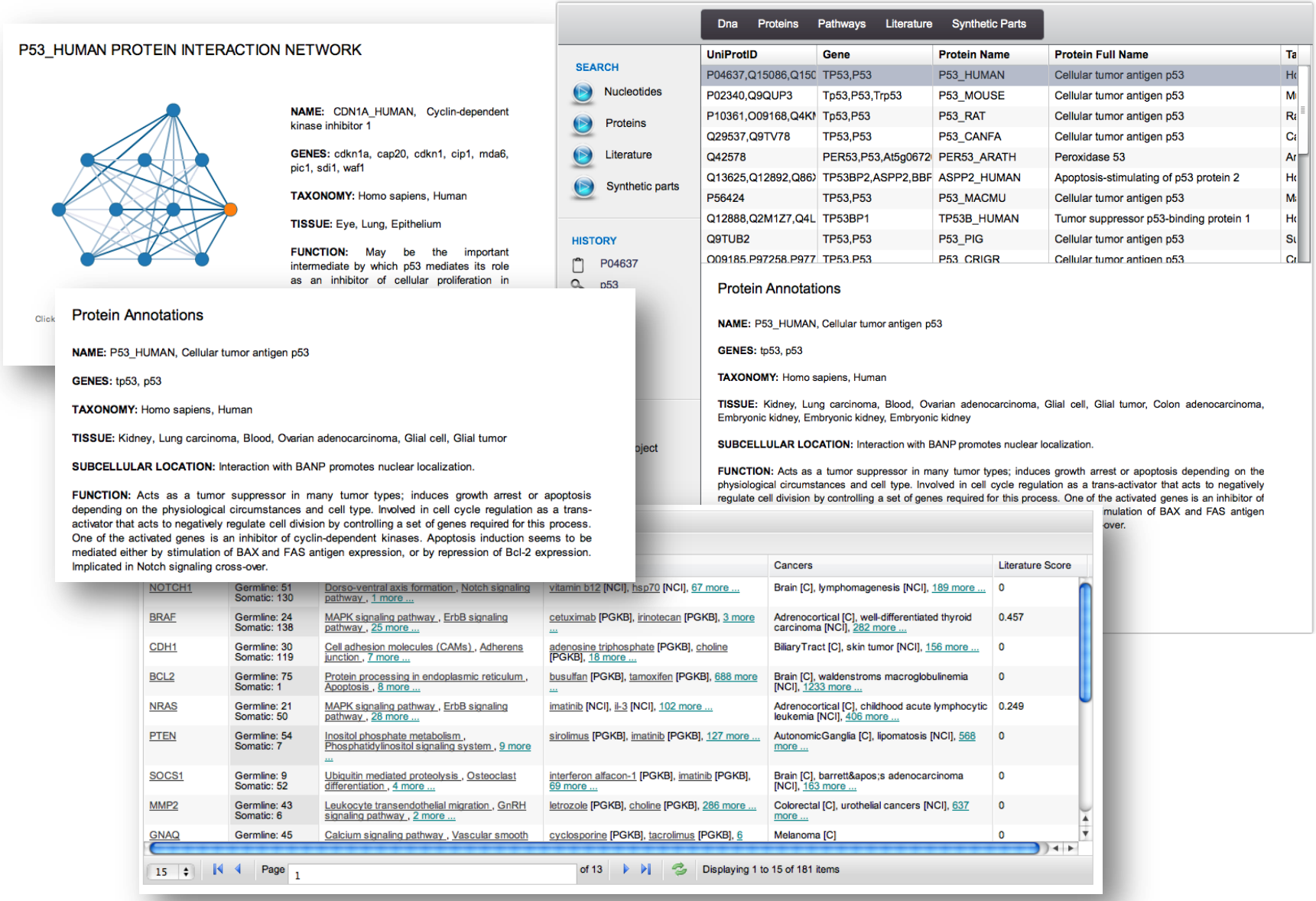
Open Pharmacological Space



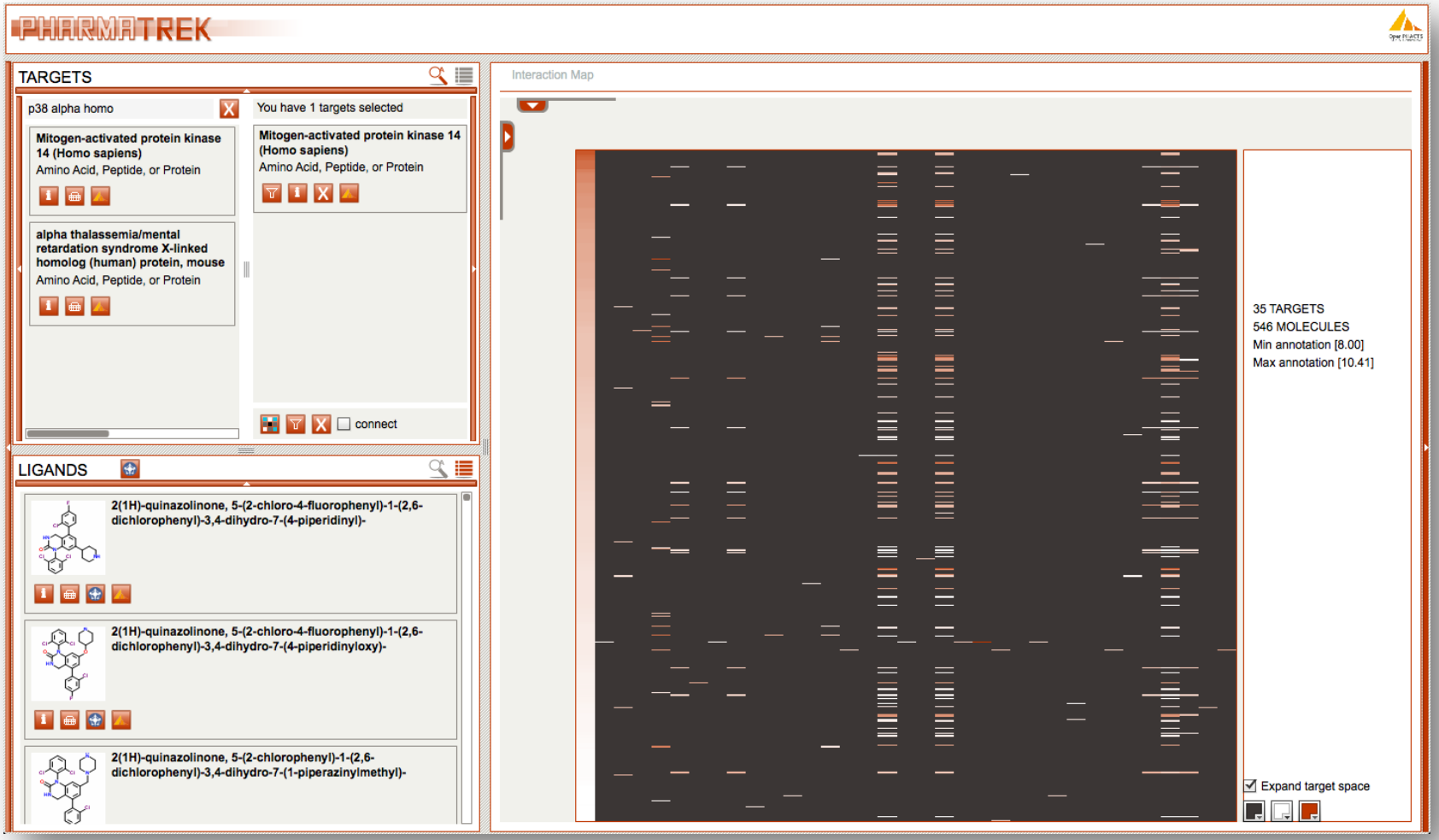
explorer.openphacts.org

Target dossier (CNIO)

Front-end framework to visualize biological data

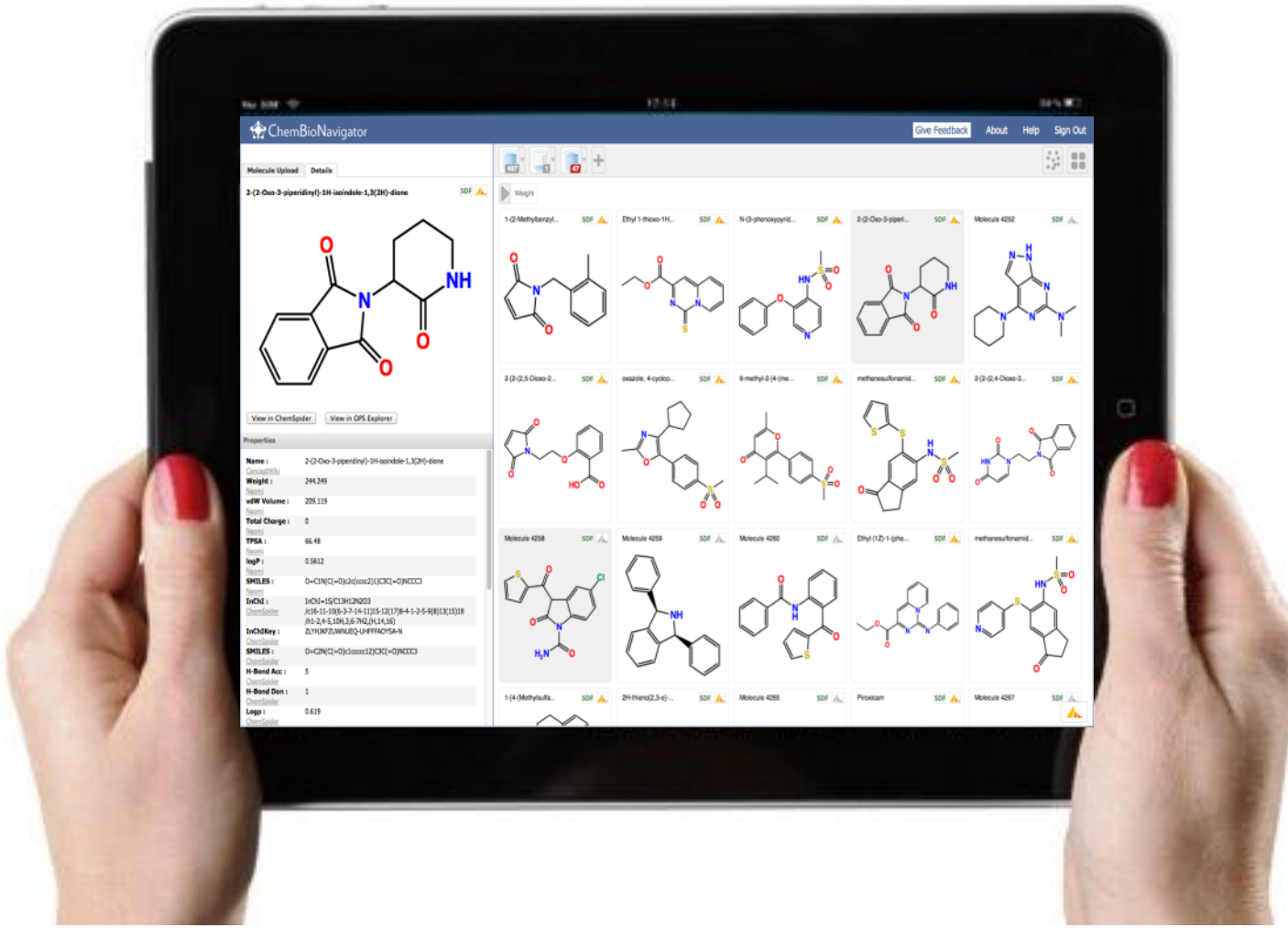


Open PHACTS
Open Pharmacological Space



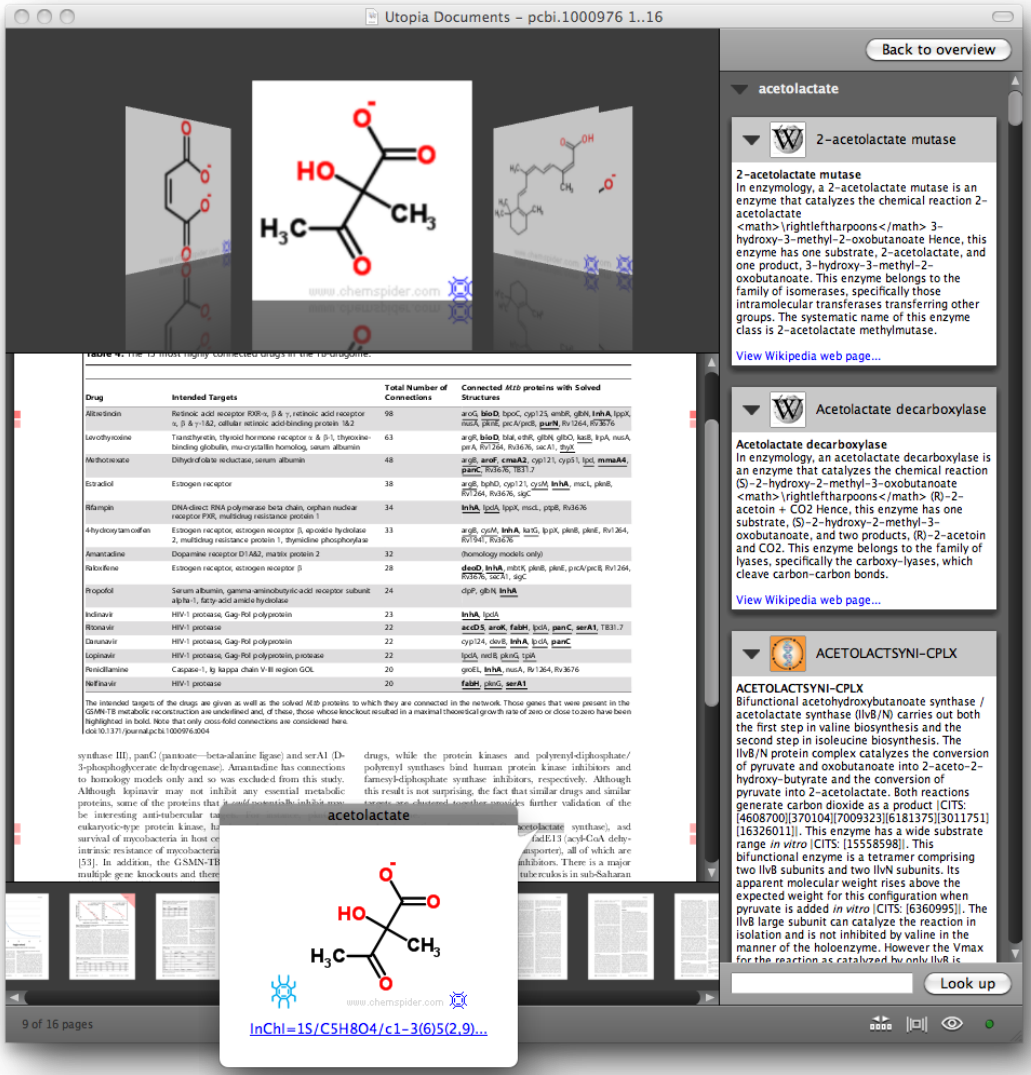
Open PHACTS

Open Pharmacological Space



Open PHACTS

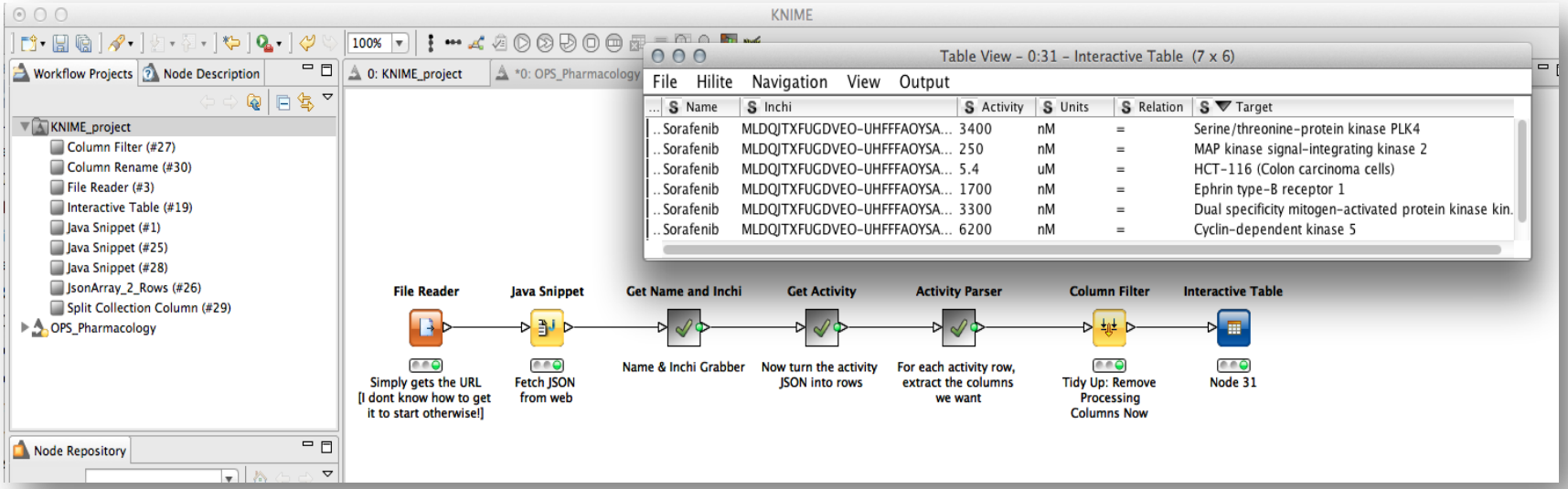
Open Pharmacological Space

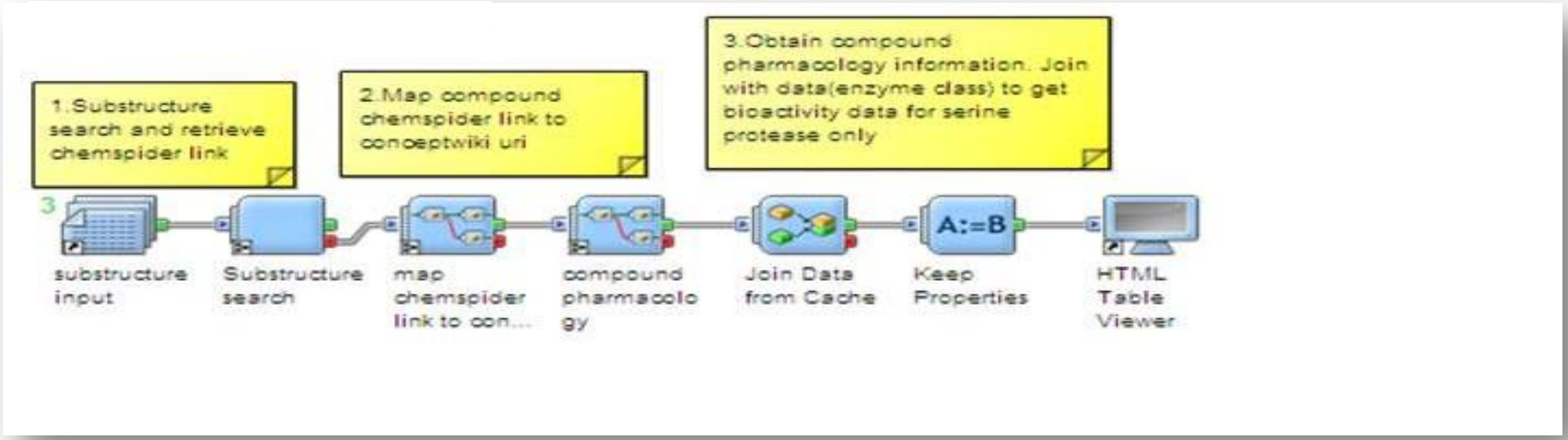


Open PHACTS
Open Pharmacological Space

Open PHACTS

Open Pharmacological Space





Open PHACTS
Open Pharmacological Space

Links

- Home Page: http://openphacts.org
- Papers/Publications: http://openphacts.org/publications http://openphacts.org/posters
- Developer API: http://dev.openphacts.org
- Explorer: http://explorer.openphacts.org
- GSK/Pharmatrek in use video: http://www.youtube.com/watch?v=nXLg8VXLREk
- iPhone app video: http://www.youtube.com/watch?v=0aGB6YqtuQ0
- Accelrys Community Open PHACTS group: https://community.accelrys.com/groups/openphacts

Open PHACTS
Open Pharmacological Space

A Precompetitive Knowledge Framework

Pharma Needs

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

Integration

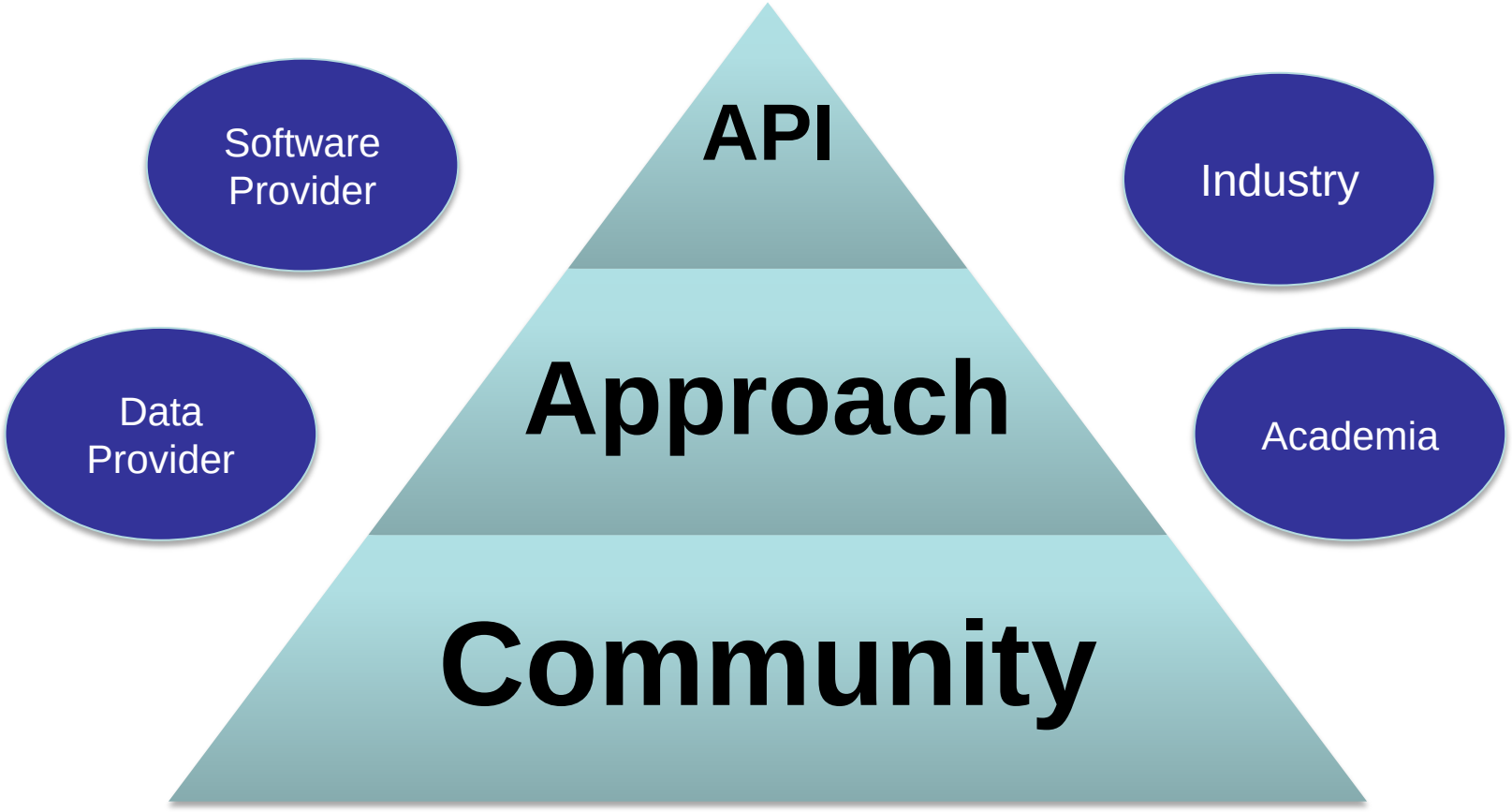
- Mapping & Populating
- Architecture
- Interfaces & Services

Inputs

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

Open PHACTS
Open Pharmacological Space

The Ecosystem is ….



Open PHACTS
Open Pharmacological Space

The Open PHACTS community ecosystem

Pfizer
universität wien
DTU
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG
BioSolveIT
Parc de Salut MAR
LUMC
Leids Universitair Medisch Centrum
RSC Advancing the Chemical Sciences
VU University Amsterdam
cnio
The University of Manchester
Maastricht University
AQnowledge
Semantics for Science
USC
universitätbonn
AstraZeneca
gsk GlaxoSmithKline
ESTEVE
NOVARTIS
MERCK
Lundbeck
Lilly
nbic
SIB Swiss Institute of Bioinformatics
connectediscovery
EMBL-EBI
janssen
OPENLINK SOFTWARE Making Technology Work For You

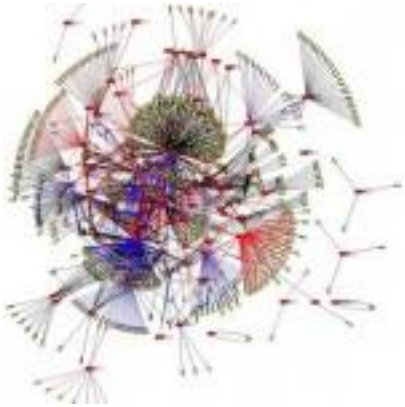
accelrys
ACD/Labs
AUREUS SCIENCES
bel
CERTARA
BIO2RDF
CODDES
codeN
Concept Web Alliance
eagle
ENTAGEN
ACCELERATING INSIGHT
Erasmus MC
Genetta SOFT
Global Health Equity Foundation
GVK BIO
Ian Harrow Consulting
IOEA consult
inte:ligand
SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington
IRB BARCELONA
IUPHAR-DB
The IUPHAR Database of Receptors and Channels
JGU
JOHANNES GUTENBERG UNIVERSITÄT MAINZ
LDBC
MGH 1811
mobileflow
THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY
Numedicus
ontoforce
orphanet
UB Parc Científic de Barcelona
PHORTOS CONSULTANTS
PRESTWICK CHEMICAL
The medicinal chemistry company
Sage BIONETWORKS
The Scripps Research Institute
SEQUENOMICS
syngenta
thehyve
THOMSON REUTERS
TWINFORMATICS
VIVO
connect • share • discover

Sustaining Impact

Open PHACTS
Open Pharmacological Space

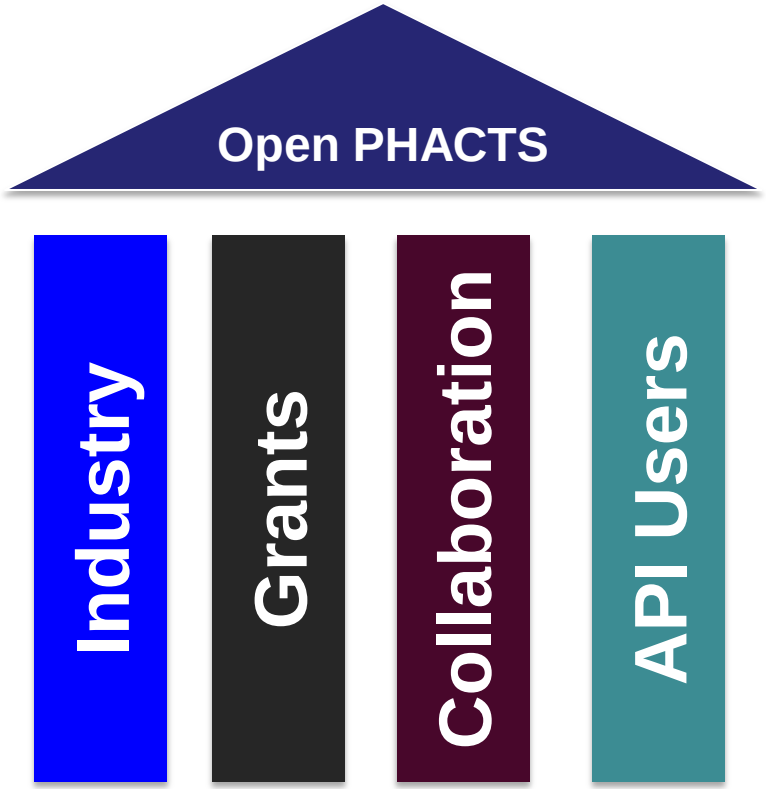
- "Software is free like puppies are free - they both need money for maintenance"

- …and more resource for future development



Open PHACTS
Open Pharmacological Space

Kick-Starting Sustainability



Open PHACTS
Open Pharmacological Space

Becoming part of the Open PHACTS Foundation

A UK-based not-for-profit member owned company

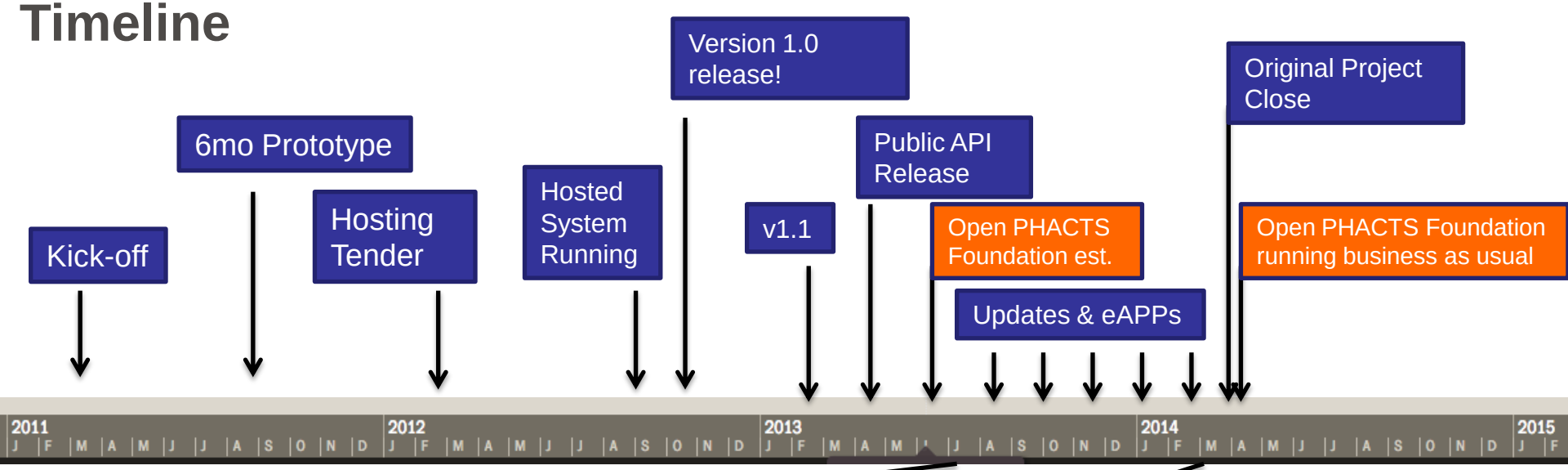
Members

- membership offers early access to platform updates and releases
- the opportunity to steer research and development directions
- receive technical support
- work with the ecosystem of developers and semantic data integrators around Open PHACTS

- tiered membership
- familiar business and governance model

Open PHACTS
Open Pharmacological Space

Timeline



- Pathways
- Ontology-based queries
- Nanopublications (incl. publisher data)
- Human genetics & disease
- Human drug data (e.g. adverse events)
- Commercial Data Pilot results

Beyond 2013

- Internal data integration
- Full commercial data implementation
- Advanced analytics
- Translational data
- Other IMI integration?
- …..

Conclusions

- There is a lot of public data out there. To deal with it you must:
  - Talk to the providers (EBI, NCBI, NIH, NBIC, UofM, Publishers, SMEs)
  - Identify the use cases
  - Promote data standards
  - Physically integrate the data
  - Manage the nightmare of different identifiers
  - Manage the complexity of equality
  - Maintain the data
  - Identify quality issues, have a plan to address them
  - Develop apps, build scientific success stories

Open PHACTS provides a cost-effective way to accomplish greater impact of public (and beyond) scientific data by sharing this burden across industry

Open PHACTS Project Partners

Open PHACTS
Open Pharmacological Space

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

pmu@openphacts.org

@Open_PHACTS

Open PHACTS