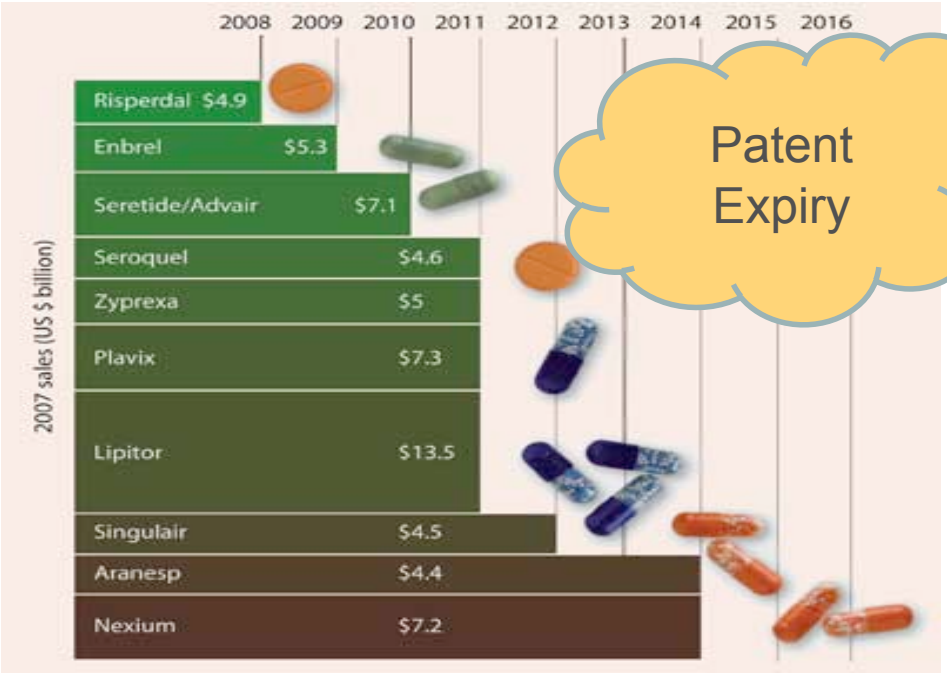
Open PHACTS
Linked Open Data for Drug Discovery

Herman van Vlijmen – Janssen Pharmaceutica
Orri Erling – OpenLink Software

Linked Open Data congres, Hilversum, 25 June 2014

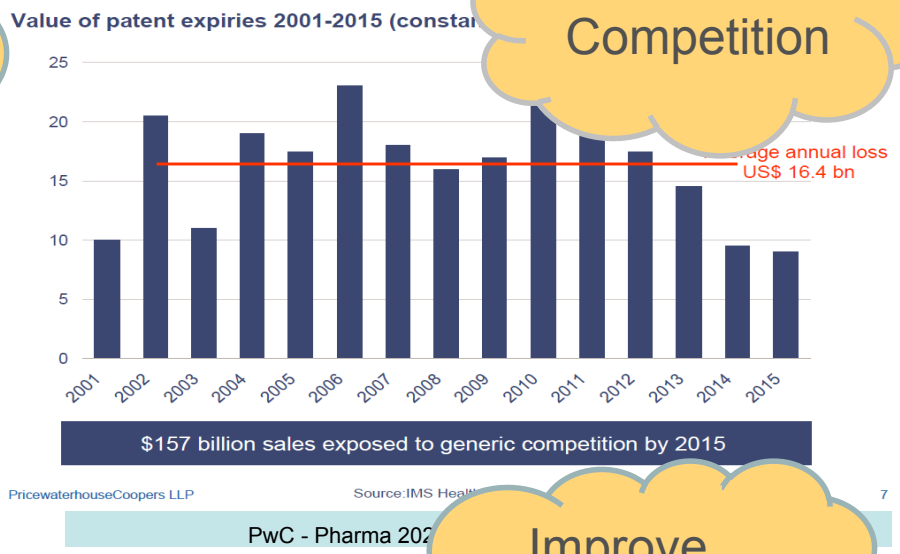
The Pharma Industry
Challenges in a Changing Landscape

Open PHACTS
Open Pharmacological Space



Patent Expiry

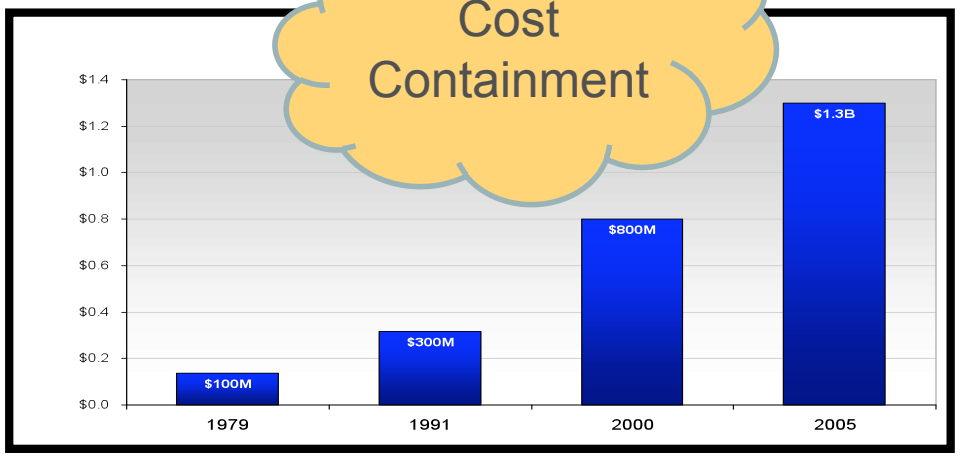
Resulting in a sales reven



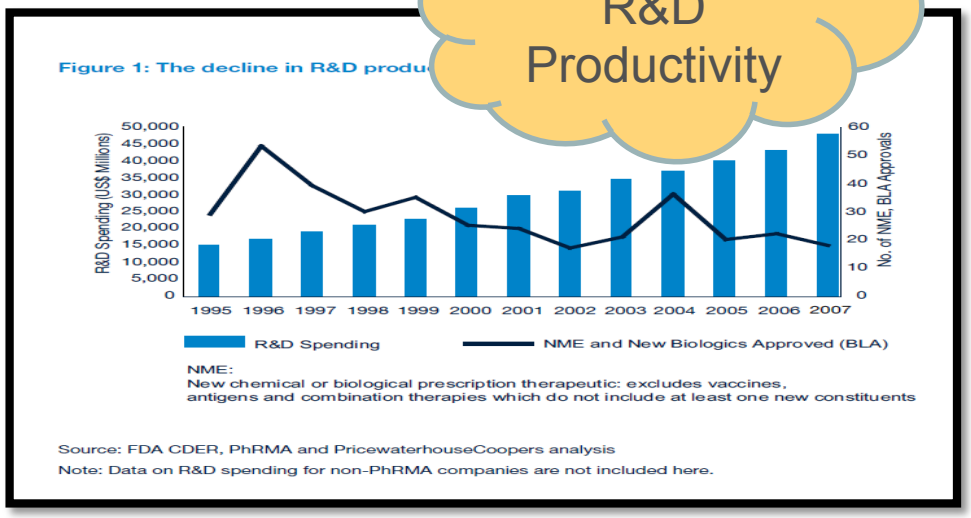
Generic Competition

PwC - Pharma 202

http://www.rsc.org/chemistryworld/Is
PharmaRefocusesOnThePatent



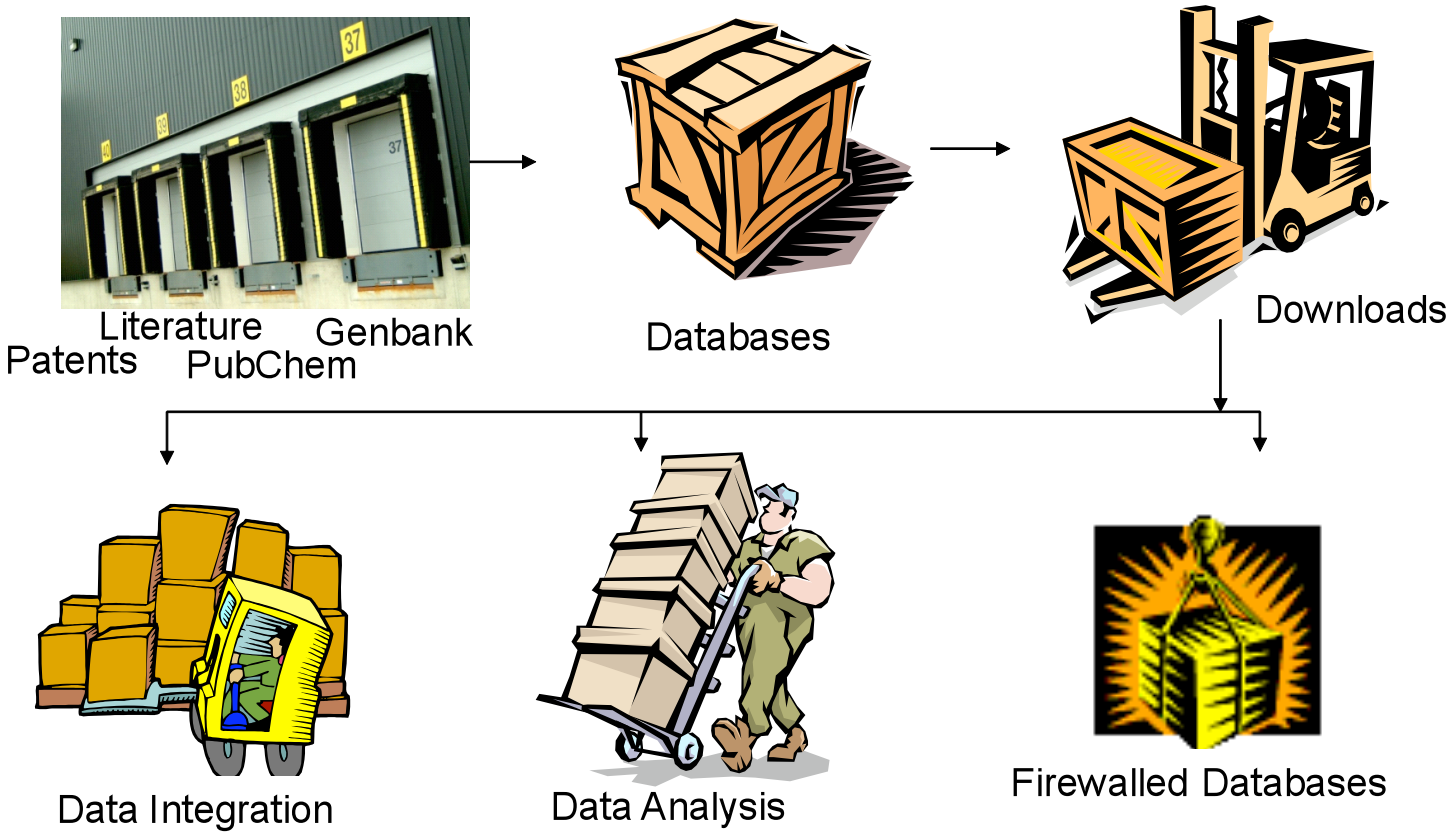
Cost Containment



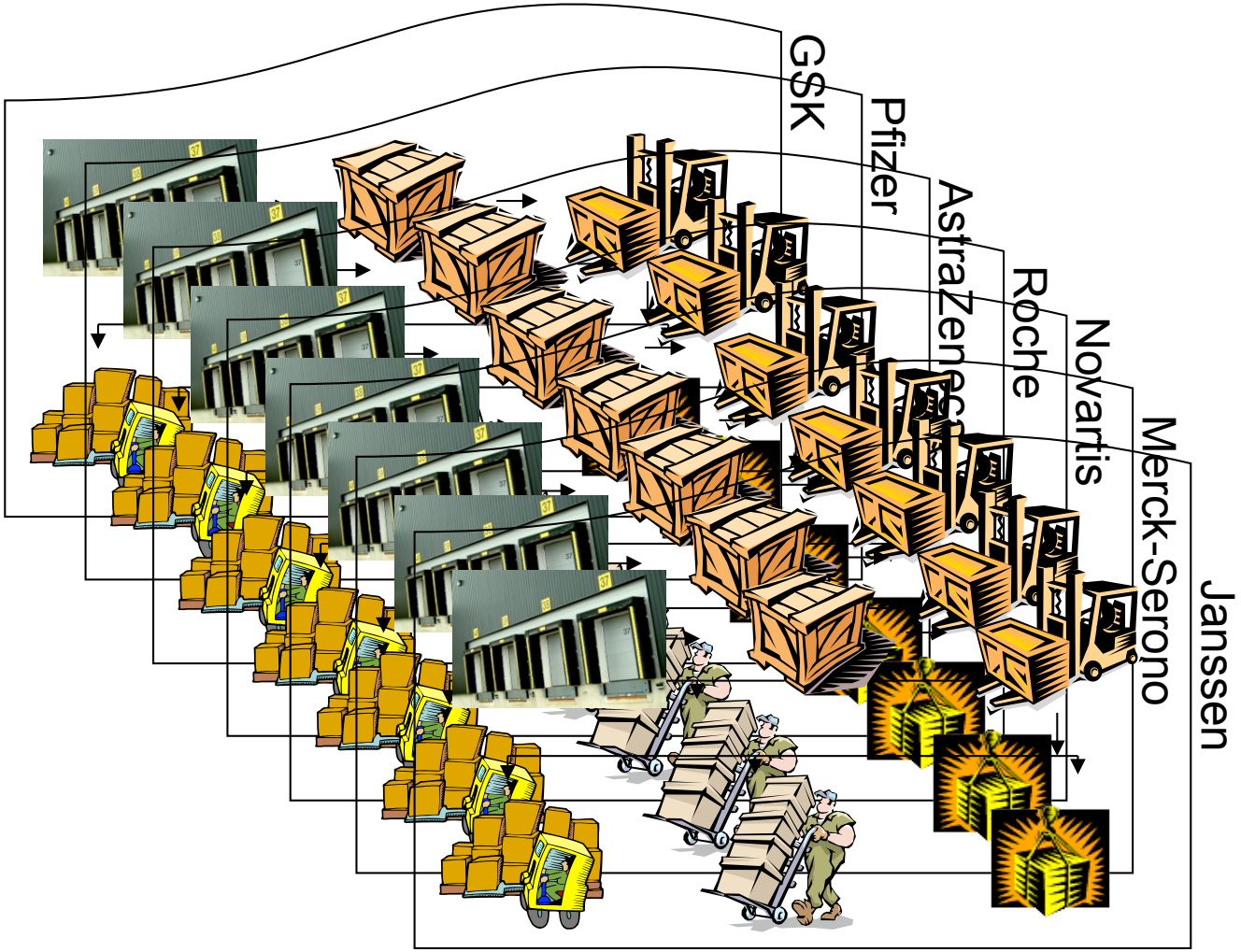
Improve R&D Productivity

Public Domain Drug Discovery Data:

Pharma are accessing, processing, storing & re-processing



We are all doing this many times……

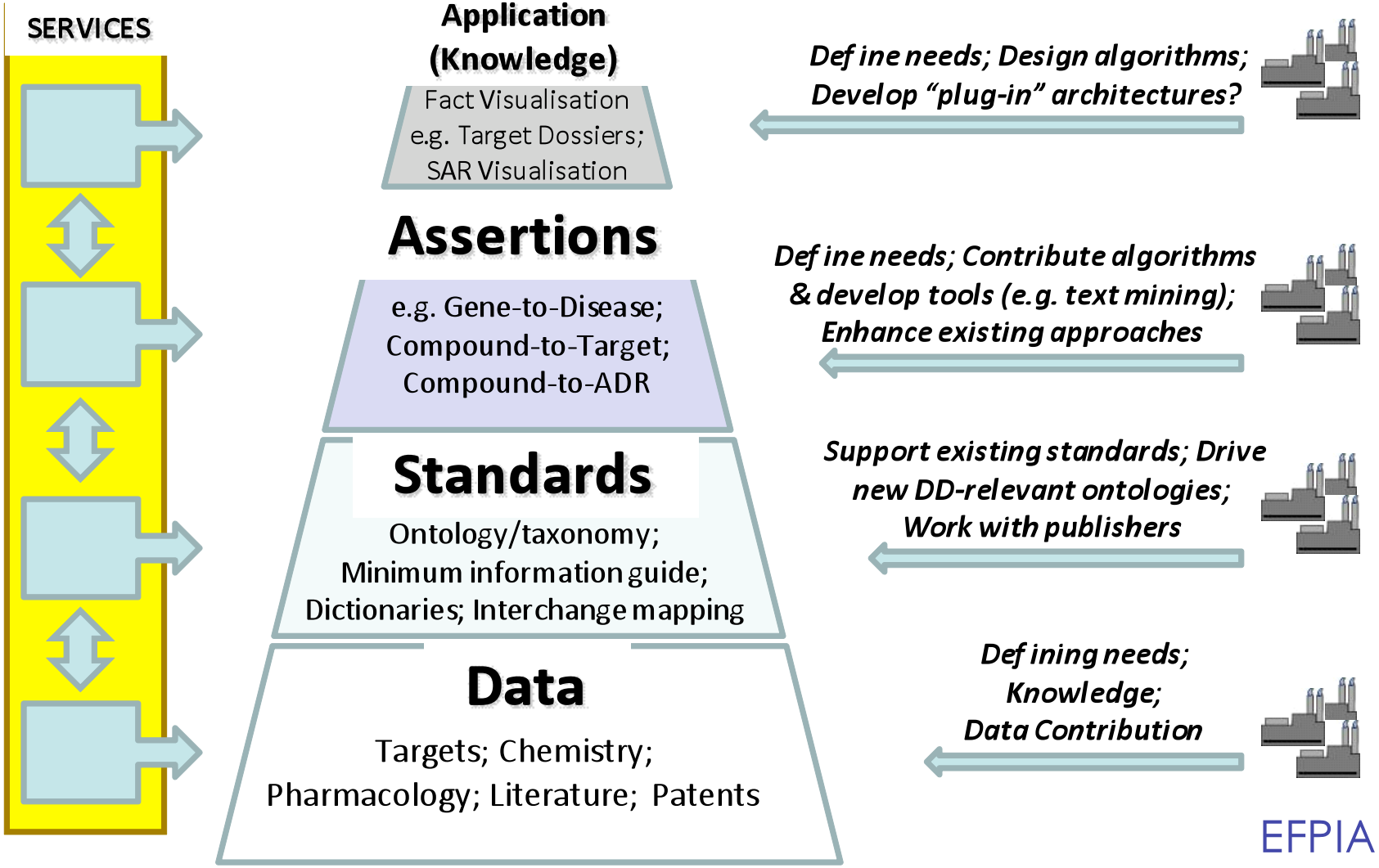


The Open PHACTS Project

- 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
- 16 academic partners, 8 pharmaceutical companies, 4 biotechs
- Work split into clusters:
  - Technical Build: Create the technology
  - Scientific Drive: Develop use cases and exemplar applications
  - Community & Sustainability: Engage community and build the future

OPS Components

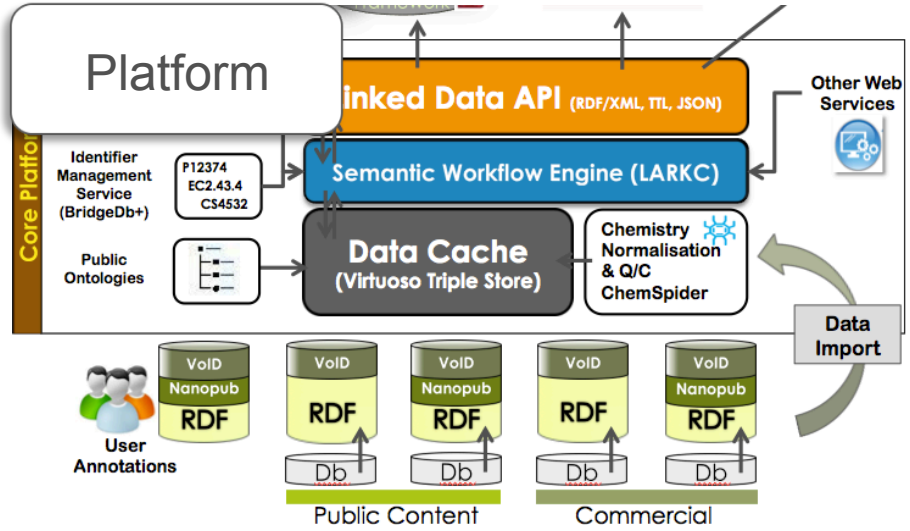
Open PHACTS
Open Pharmacological Space



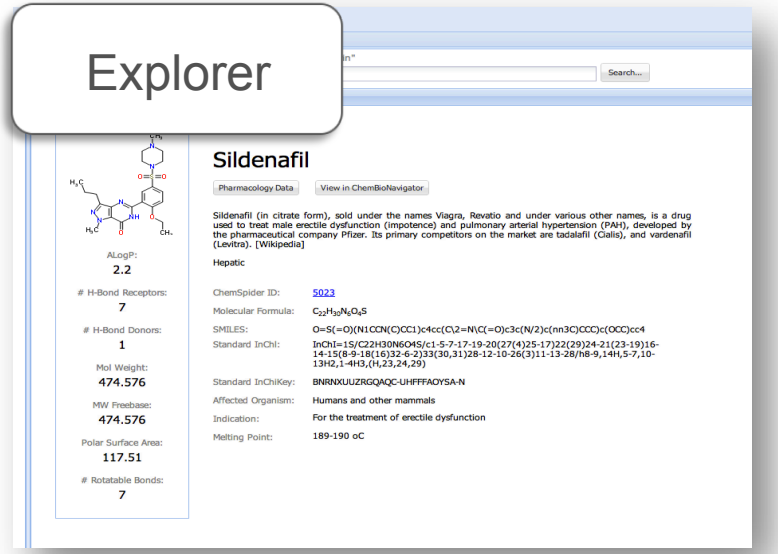
Open PHACTS

Open Pharmacological Space

Platform



Explorer



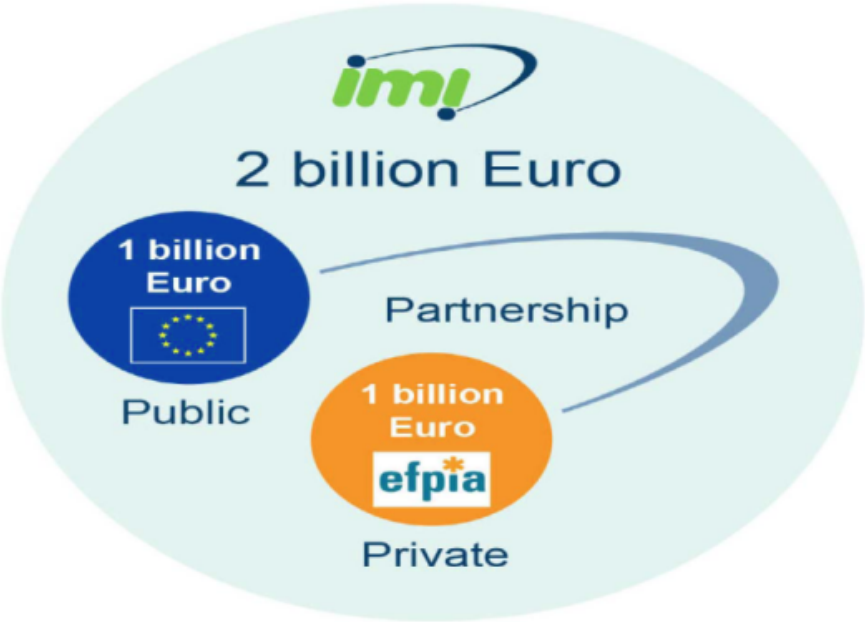
Apps

API

```
?ops_item skos:exactMatch ?cw
?ops_item skos:exactMatch ?eq
?cw_uri skos:prefLabel ?targe
        void:inDataset <http:
?equiv_target dc:title ?targe
        ops:target_organism ?
        ops:targetOfAssay ?eq
        void:inDataset <http:
ops:targetOfAssay owl:inverse
?equiv_assay chembl:organism
        chembl:hasDescription
                                y ?
                                verseOf chembl:onAssa
                                ?std_type ;
```

Standards

IMI: The Innovative Medicines Initiative



- Biggest public-private partnership in area of medicine
- Collaboration between European Commission and European Federation of Pharmaceutical Industries and Associations (EFPIA)
- Promotion of medical innovation in Europe
- Tackle key bottlenecks
- Recognises “in kind” contributions
- Focus on key problems
  - Efficacy, Safety, Education & Training, Knowledge Management

Project Partners

Open PHACTS
Open Pharmacological Space

Universität Wien
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
Netherlands Bioinformatics Centre
Swiss Institute of Bioinformatics
ConnectedDiscovery
EMBL-European Bioinformatics Institute
OpenLink Software
Pfizer
Novartis
Merck Serono
H. Lundbeck A/S
Eli Lilly
Janssen
AstraZeneca
GlaxoSmithKline
Esteve

Associate Partners

Open PHACTS
Open Pharmacological Space

orphanet

Global Health Equity Foundation
research · advocacy · capacity building

VIVO connect share discover

eagle genomics

Sage BIONETWORKS

mobileflow

Concept Web Alliance

Numedicus

ACD/Labs

MGH 1811

codeN

Genetta SOFT

thehyve

CERTARA
Implementing Translational Science

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

INDIANA UNIVERSITY

IRB BARCELONA
INSTITUTE FOR RESEARCH IN BIOMEDICINE

ih Ian Harrow Consulting

AUREUS SCIENCES

GVK BIO
Accelerating Research

Sequenomics

accelrys

Erasmus MC

Parc Científic de Barcelona
B·KC

IUPHAR DATABASE
International Union of Basic and Clinical Pharmacology

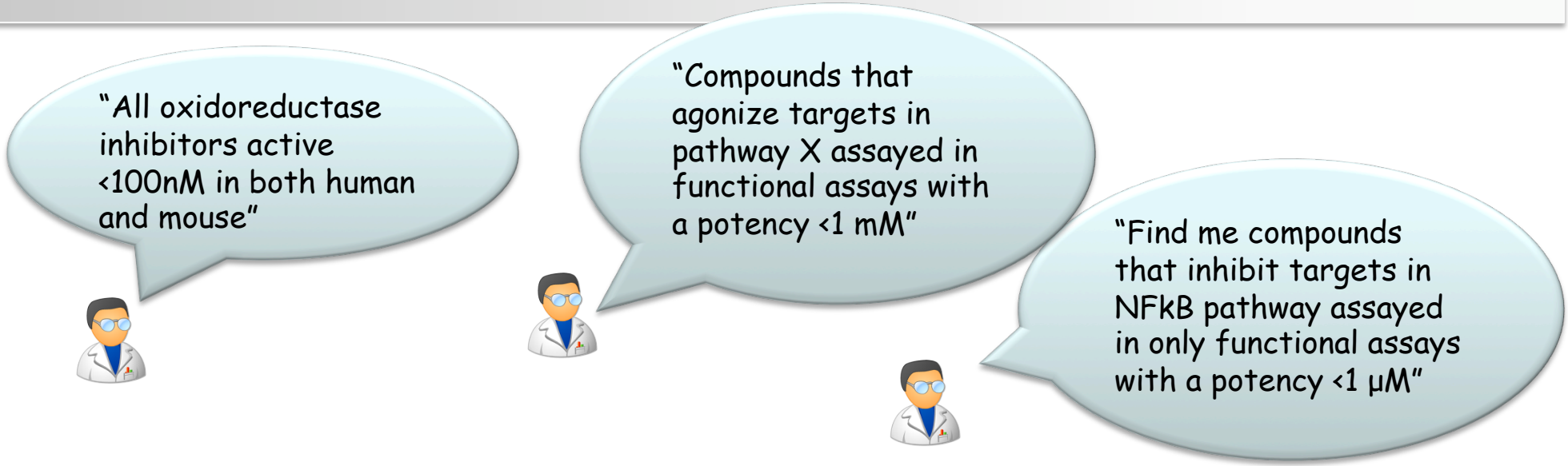
THOMSON REUTERS

A use-case driven approach, focussed on delivery for the real world

- Main architecture, technical implementation and primary capabilities driven by a set of **prioritised research questions**
- Based on the main research questions define **prioritised data sources**
- Develop **three Exemplars** to demonstrate the capabilites of the Open PHACTS System and to define interfaces and input/output standards

What do we need?

Open PHACTS
Open Pharmacological Space



ChEMBL | DrugBank | Gene Ontology | Wikipathways | GeneGo

ChEBI | Uniprot | UMLS | GVKBio

ConceptWiki | ChemSpider | TrialTrove | TR Integrity

The Open PHACTS infrastructure can support many different domains & questions

TABLE 1

The top 20 research questions

Question number	Question
Cluster I	
Q1	Give me all oxidoreductase inhibitors active <100 nM in human and mouse
Q2	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?
Q3	Given a target, find me all actives against that target. Find/predict polypharmacology of actives. Determine ADMET profile of actives
Q4	For a given interaction profile – give me similar compounds
Q5	The current Factor Xa lead series is characterized by substructure X. Retrieve all bioactivity data in serine protease assays for molecules that contain substructure X
Q6	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 could modulate the target directly? I.e. return all compounds active in assays where the resolution is at least at the level of the target family (i.e. PKC) from structured assay databases and the literature
Q7	Give me all active compounds on a given target with the relevant assay data
Q8	Identify all known protein–protein interaction inhibitors
Q9	For a given compound, give me the interaction profile with targets
Q10	For a given compound, summarize all 'similar compounds' and their activities
Q11	Retrieve all experimental and clinical data for a given list of compounds defined by their chemical structure (with options to match stereochemistry or not)
Cluster II	
Q12	For my given compound, which targets have been patented in the context of Alzheimer's disease?
Q13	Which ligands have been described for a particular target associated with transthyretin-related amyloidosis, what is their affinity for that target and how far are they advanced into preclinical/clinical phases, with links to publications/patents describing these interactions?
Q14	Target druggability: compounds directed against target X have been tested in which indications? Which new targets have appeared recently in the patent literature for a disease? Has the target been screened against in AZ before? What information on *in vitro* or *in vivo* screens has already been performed on a compound?
Q15	Which chemical series have been shown to be active against target X? Which new targets have been associated with disease Y? Which companies are working on target X or disease Y?
Q16	Which compounds are known to be activators of targets that relate to Parkinson's disease or Alzheimer's disease
Q17	For my specific target, which active compounds have been reported in the literature? What is also known about upstream and downstream targets?
Q18	Compounds that agonize targets in pathway X assayed in only functional assays with a potency <1 μM
Q19	Give me the compound(s) that hit most specifically the multiple targets in a given pathway (disease)
Q20	For a given disease/indication, give me all targets in the pathway and all active compounds hitting them

Data Associations

Open PHACTS
Open Pharmacological Space

Drug Discovery Today • Volume 18, Numbers 17/18 • September 2013

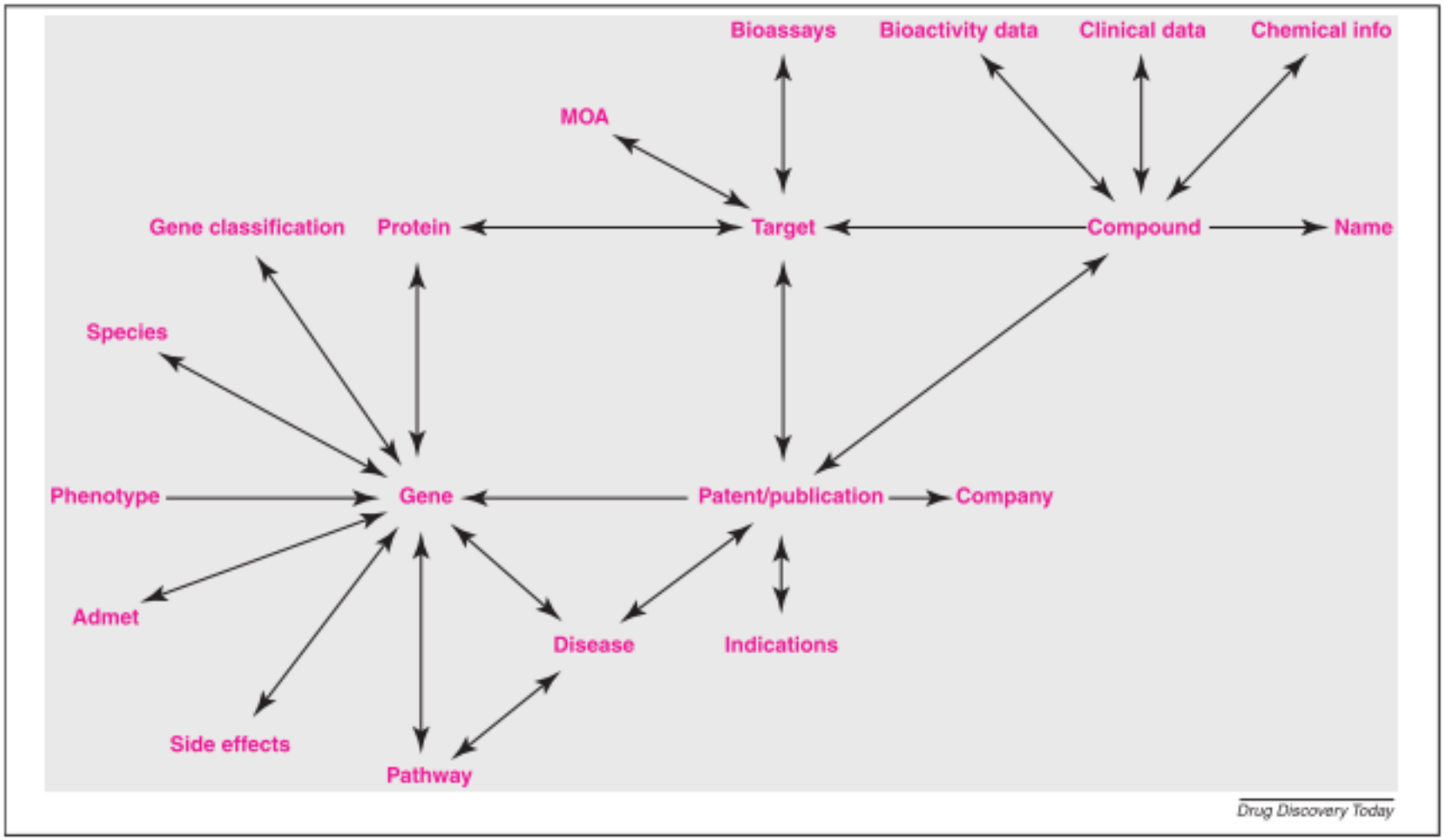
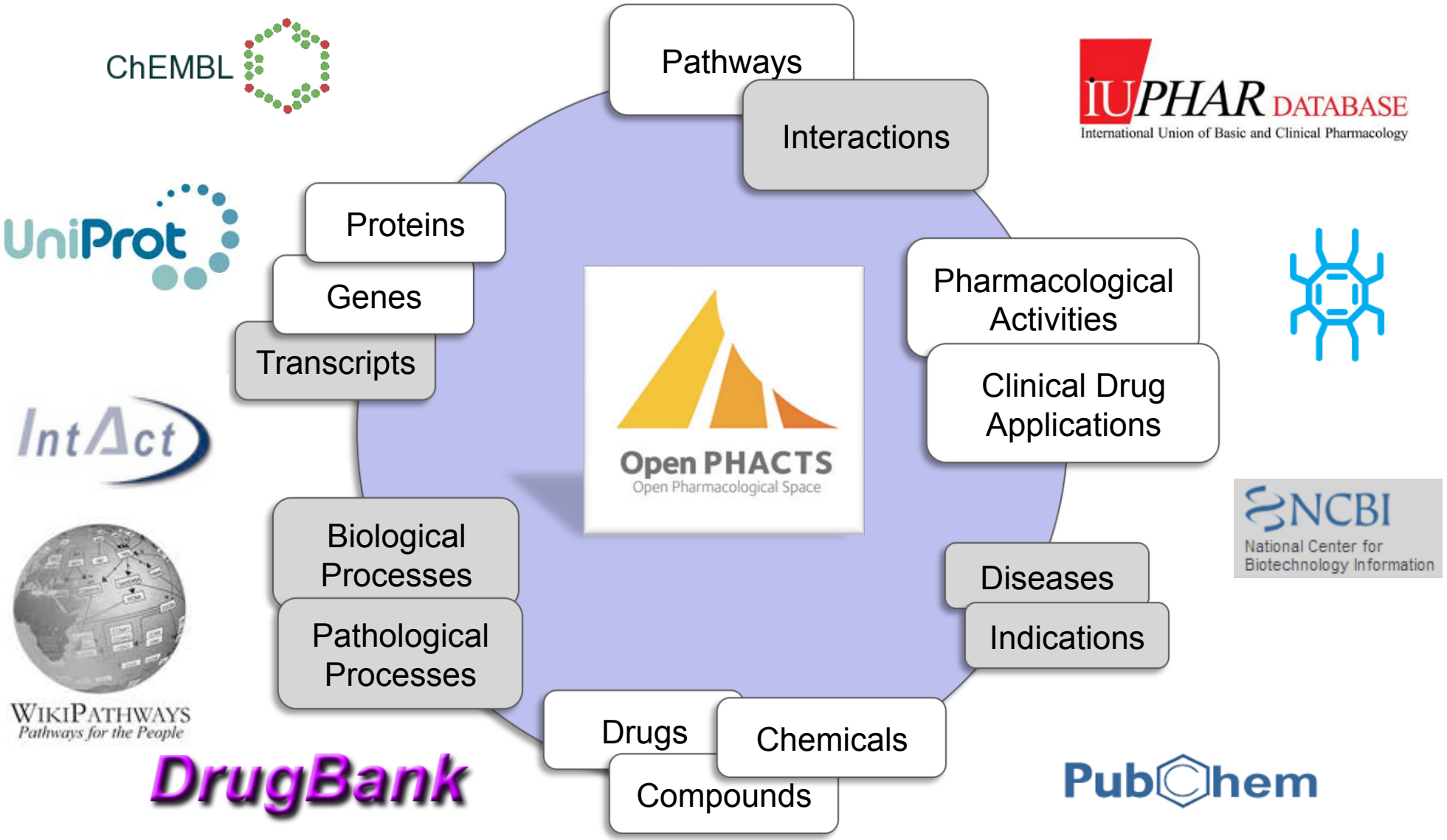


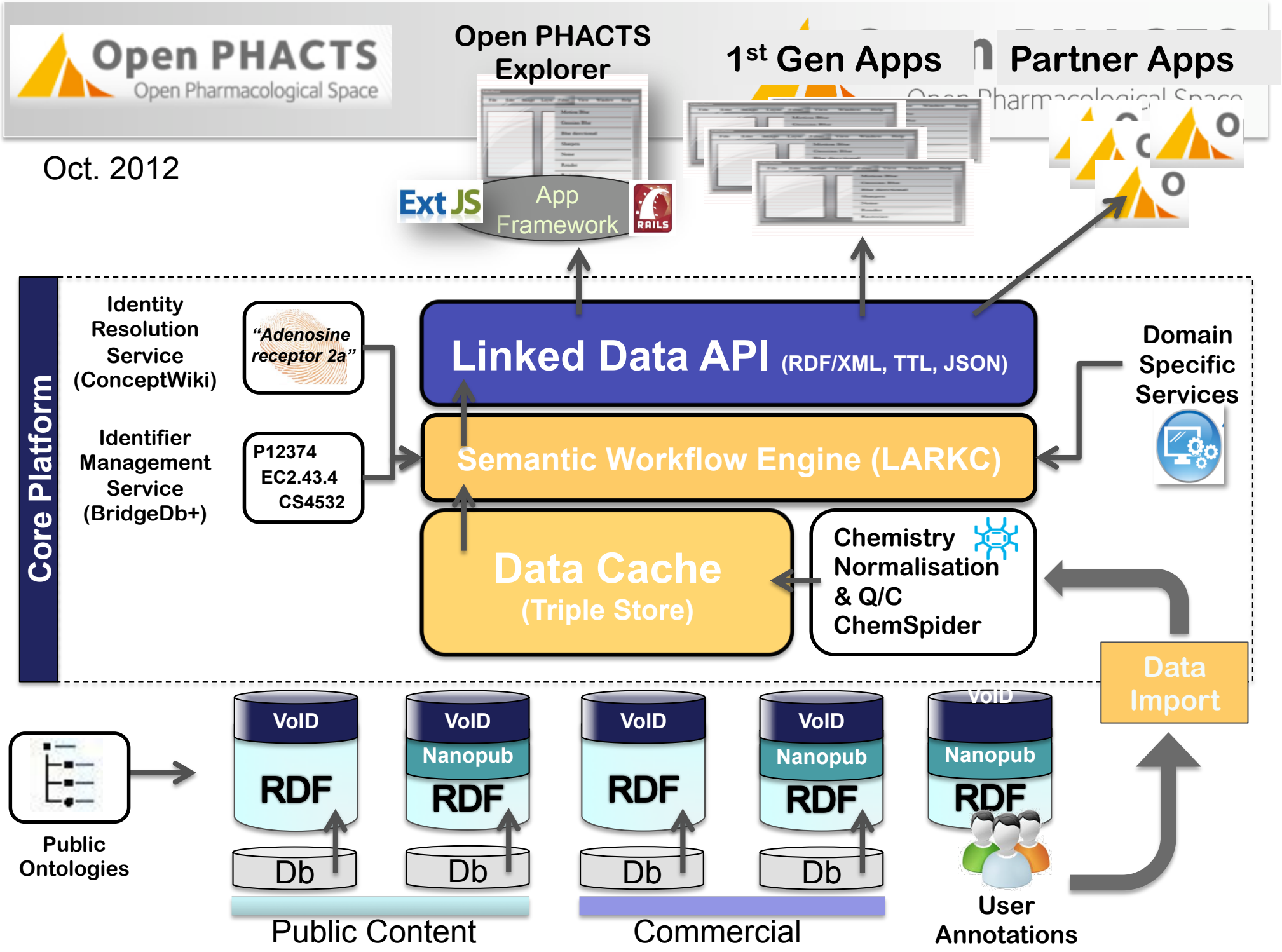
FIGURE 2

Network of data associations needed to answer the top-ranked scientific competency questions. The network reflects a cartoon that summarizes the data associations that are needed to target the top 20 research questions.

Public Data Sources

Open PHACTS
Open Pharmacological Space




Open PHACTS
Open Pharmacological Space
Open PHACTS Explorer
1st Gen Apps
Partner Apps
Oct. 2012
Ext JS
App Framework
RAILS
Core Platform
Identity Resolution Service (ConceptWiki)
"Adenosine receptor 2a"
Identifier Management Service (BridgeDb+)
P12374
EC2.43.4
CS4532
Linked Data API (RDF/XML, TTL, JSON)
Semantic Workflow Engine (LARKC)
Data Cache
(Triple Store)
Domain Specific Services
Chemistry Normalisation & Q/C ChemSpider
Data Import
Public Ontologies
VoID
Nanopub
RDF
Db
Public Content
Commercial
User Annotations

Statistics of Datasets Loaded into Open PHACTS Version 1.3

Source	Version	Supplier	Downloaded	Initial Records	Triples	Properties
Chembl	Chembl 16 RDF	EBI	25 June 2013	1,247,403 (~1,236,686 compounds, 9844 targets, 6243 target components, 873 protein classes)	304,420,681	77
DrugBank	Aug 2008	Bio2Rdf (www4.wiwiss.fu-berlin.de)	08 Aug 2012	19,628(~14,000 drugs, 5000 targets)	517,584	74
SwissProt, UniParc, UniRef	2013_06	SIB	2013_06		533,394,147	82
ENZYME	2013_07	SIB	2013_07	6,187	47,661	2
ChEBI	Release 104	EBI	19 June 2013	40,575	40,575	2
GeneOntology	Jan 21, 2013	GO	21 Jan 2013	38,137	1,265,273	26
GOA	2013	GO	09 Sept 2013	various species	23,489,501	15
WikiPathways	v0.?1_20130710	Maastricht	10 July 2013	946	1,449,981	34
ChemSpider		Open PHACTS Chemistry Registry (OCRS)	Nov 11, 2013		tbc	
ConceptWiki	version 1.3	NBIC	09 Sept 2013	2,828,966	3,739,884	1

Example of vocabulary/ontology challenge

Open PHACTS
Open Pharmacological Space

Quantitative Data Challenges

STANDARD_TYPE	UNIT_COU
AC50	7
Activity	421
EC50	39
IC50	46
ID50	42
Ki	23
Log IC50	4
Log Ki	7
Potency	11
log IC50	0

>5000 types

--	STANDARD_UNITS	COUNT(*)
IC50	nM	829448
IC50	ug.mL-1	41000
IC50		38521
IC50	ug/ml	2038
IC50	ug ml-1	509
IC50	mg kg-1	295
IC50	molar ratio	178
IC50	ug	117
IC50	%	113
IC50	uM well-1	52
IC50	p.p.m.	51
IC50	ppm	36
IC50	uM-1	25
IC50	nM kg-1	25
IC50	milliequivalent	22
IC50	kJ m-2	20

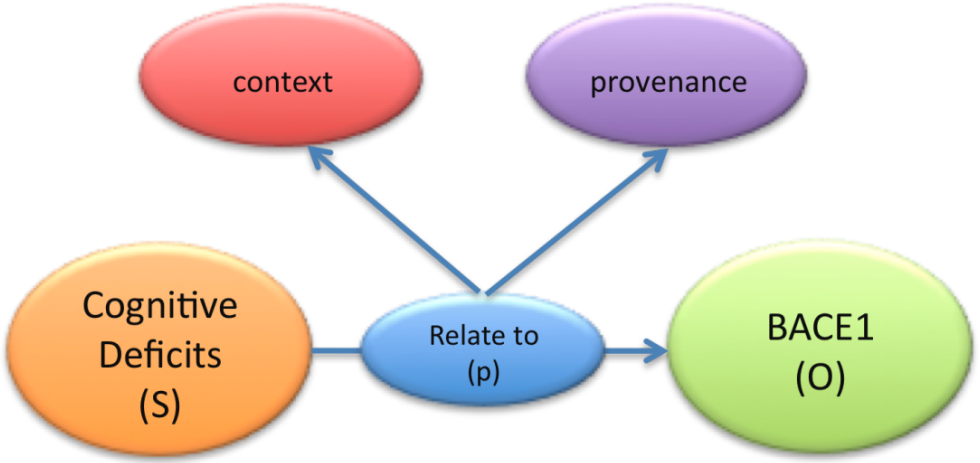
~ 100 units

Implemented using the Quantities, Dimension, Units, Types Ontology (http://www.qudt.org/)

Concept: nanopublications for provenance

Open PHACTS
Open Pharmacological Space

Nanopublications – Capturing scientific information in the Triple Store



nature genetics

nature.com ▸ journal home ▸ archive ▸ issue ▸ commentary ▸ full text

NATURE GENETICS | COMMENTARY

The value of data

Barend Mons, Herman van Haagen, Christine Chichester, Peter-Bram 't Hoen, Johan T den Dunnen, Gertjan van Ommen, Erik van Mulligen, Bharat Singh, Rob Hooft, Marco Roos, Joel Hammond, Bruce Kiesel, Belinda Giardine, Jan Velterop, Paul Groth & Erik Schultes

Affiliations | Contributions | Corresponding author

Nature Genetics **43**, 281–283 (2011) | doi:10.1038/ng0411-281

Published online 29 March 2011

Nano-Publication in the e-science era

Barend Mons[1,2,3] and Jan Velterop[1,2],

[1] Concept Web Alliance, [2] Netherlands BioInformatics Centre, [3] Leiden University Medical Center.

barend.mons@nbic.nl, velterop@conceptweballiance.org

The Anatomy of a Nano-publication

Paul Groth
VU University Amsterdam
De Boelelaan 1081a
1081 HV Amsterdam,
The Netherlands
pgroth@few.vu.nl

Andrew Gibson
University of Amsterdam
Nieuwe Achtergracht 166, C-712
1018 WV Amsterdam
The Netherlands
a.p.gibson@uva.nl

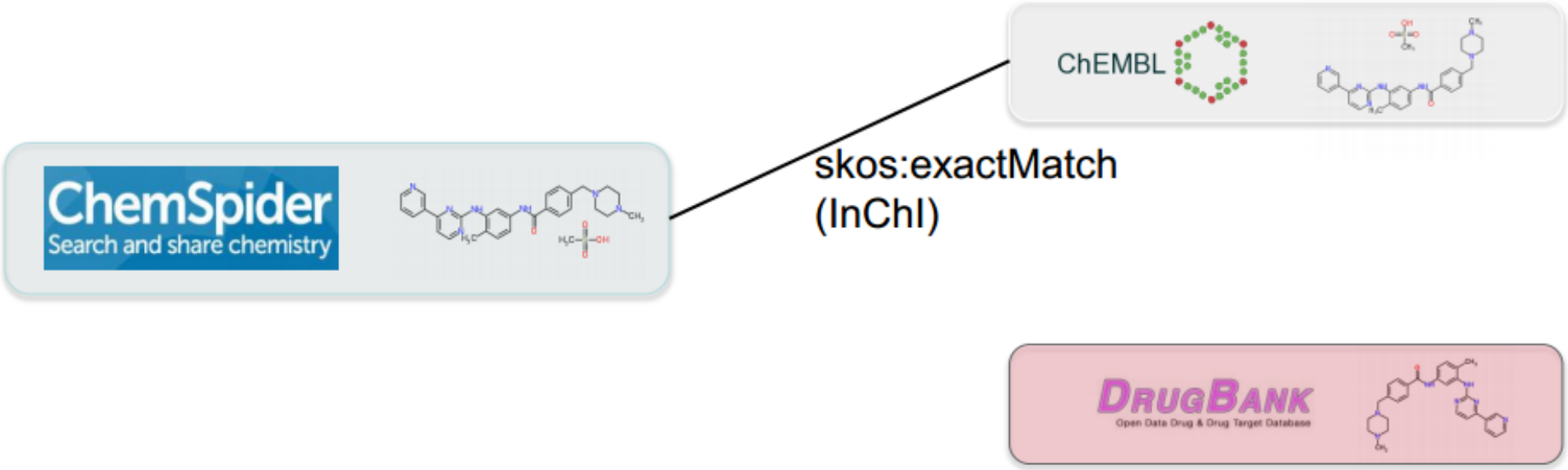
Johannes Velterop
Concept Web Alliance / NBIC
9 Benfleet Close
Cobham, Surrey, KT11 2NR
United Kingdom
jan.velterop@nbic.nl

ABSTRACT

Newer standards like RDFa also facilitate this and integrate with

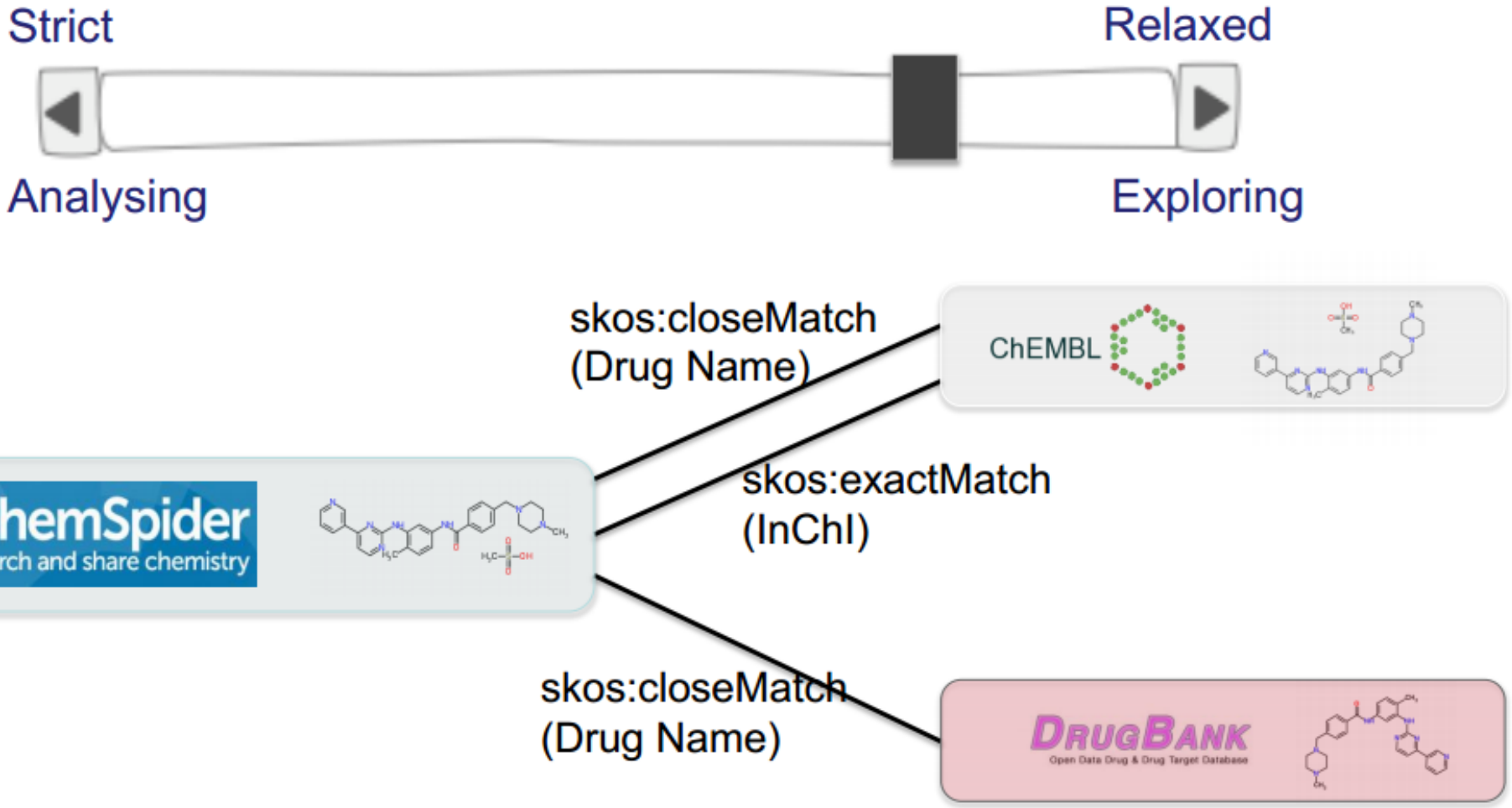
Concept: Scientific lenses

Open PHACTS
Open Pharmacological Space



Concept: Scientific lenses

Open PHACTS
Open Pharmacological Space



Example applications

Open PHACTS
Open Pharmacological Space

Open PHACTS

Browse and search the data within the Open PHACTS Discovery Platform.

Developed by the University of Manchester and University of Vienna

ChemBioNavigator

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

Navigate pharmacological space in a flexible and interactive way.

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

SciBite

Connects the latest news and events in Pharma and Biotech directly to pharmacology data within the Open PHACTS platform.

Developed by SciBite Limited

utopia

Allows the semantic enrichment of scientific articles in PDF format.

Developed by the University of Manchester

GARField

Intuitive predicts target pharmacology based on the Similar Ensemble Approach.

Developed by the Technical University of Denmark

collector

Extracts data to build QSAR predictive models with data from the eTOX project.

Developed by PSMAR as part of the eTOX project

accelrys Pipeline Pilot

A repository of useful Pipeline Pilot components and workflows has been developed.

Open PHACTS - Pipeline Pilot Community

KNIME

A KNIME repository of components and workflows has been developed.

Open PHACTS - KNIME Community

Excel

Queries the Open PHACTS API from Microsoft's Excel spreadsheet software.

Developed by the University of Vienna

AQnowledge Semantics for Science

Identifies significant entities in scientific text, and provides links to Open PHACTS Explorer.

Developed by AQnowledge

he

Helium for Excel Community Edition contains three functions that use the Open PHACTS API.

Developed by Ceiba Solutions

More complex use cases: Relating two distant concepts

Open PHACTS
Open Pharmacological Space

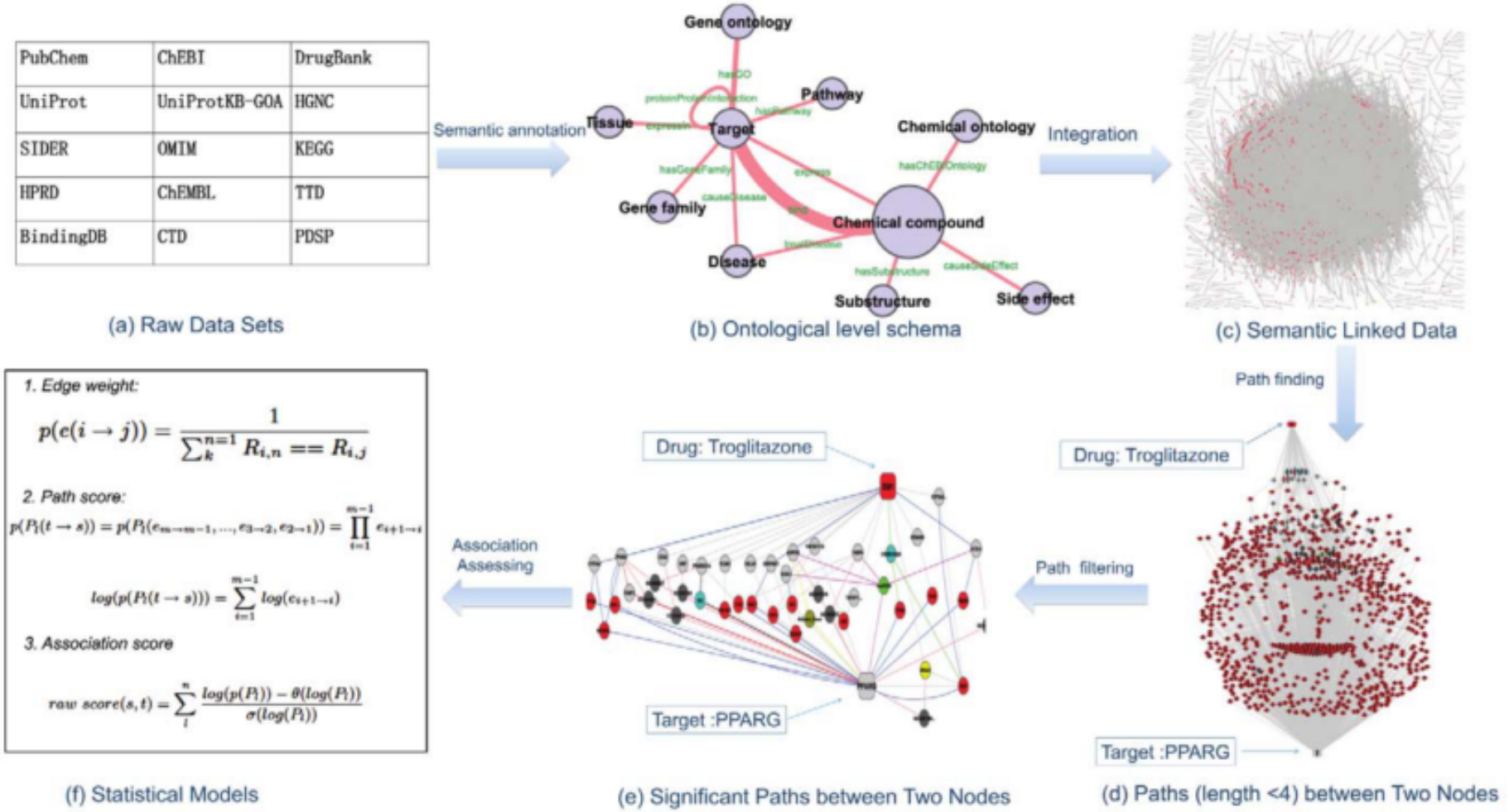


Figure 1. SLAP pipeline. An ontology is used to annotate public data sets and integrate them into a semantic linked network. Two nodes are linked by one or more number of paths, but only a small number of significant paths are kept for association estimation. The path significance and drug target associations are assessed by statistical models derived from random samples.
doi:10.1371/journal.pcbi.1002574.g001

Chen et at. PLoS Comp Biol 2012

Developments

Open PHACTS
Open Pharmacological Space

- Continue improving the system: features, performance, API calls, etc
- Expand implementation of data sources based on new set of scientific use cases – Project received 2 years additional funding
- Development and improvement of new and existing applications that use the OpenPHACS API
- Set up organizational model to continue maintenance and development after IMI funding

Acknowledgments

The OpenPHACTS consortium

Play!
API: https://dev.openphacts.org/