Open PHACTS
Open Pharmacological Space

Pharmacological information discovery with Open PHACTS

www.openphacts.org

- **The Innovative Medicines Initiative**
  - EC funded public-private partnership for pharmaceutical research

- Focus on key problems
  - Efficacy
  - Safety
  - Education & Training
  - **Knowledge Management**

Open PHACTS
Open Pharmacological Space

- **22 partners, 8 pharmaceutical companies, 3 biotechs, RSC**
- **36 months project length and 8 months in**

Pfizer– Coordinator
(Bryn Williams-Jones)
Universität Wien – Managing entity of IMI JU funding (Gerhard Ecker)
Technical University of Denmark
(Sören Brunak)
University of Hamburg, Center for Bioinformatics
(Mattias Rarey)
BioSolveIT GmBH
(Christian Lemmen)
Consorci Mar Parc de Salut de Barcelona
(Ferran Sanz)
Leiden University Medical Centre
(Barend Mons)

Royal Society of Chemistry
(Richard Kidd, Antony Williams)
Vrije Universiteit Amsterdam
(Paul Groth, Frank van Harmelen)
Spanish National Cancer Research Centre
(Alfonso Valencia)
University of Manchester
(Carole Goble, Steve Pettifer)
Maastricht University
(Chris Evelo)
AQnowledge
(Jan Velterop)
University of Santiago de Compostela
(Mabel Loza)
Rheinische Friedrich-Wilhelms-Universität Bonn
(Martin Hofmann-Apitius)

AstraZeneca
(Niklas Blomberg)
GlaxoSmithKline
(Andrew Leach)
Esteve
(Leo Salgado)
Novartis
(Edgar Jacoby)
Merck Serono
(Thomas Grombacher)
H. Lundbeck A/S
(Askjaer Sune)
Eli Lilly
(Hans Constandt)

CTO: Lee Harland

Open PHACTS
Open Pharmacological Space

Information Tombs...

- Internal and external
- Built to manage content
- Built to meet primary use-case
- Tailored indexes, GUIs
- **Unique language & metadata**
- Poor interoperability/integration
- Proliferation of Powerpoint, documents, excel, etc.
- Many suppliers of systems and content in a single workflow
  - ...

Public Domain Drug Discovery Data

Pharma are accessing, processing, storing & re-processing

Literature
Genbank
Patents
PubChem

Databases

Downloads

Data Integration

Data Analysis

Firewalled Databases

Open PHACTS
Open Pharmacological Space

Open PHACTS: an infrastructure project

- Develop / apply a set of robust **standards**…
- Implementing the standards in a **semantic integration platform** *("Open Pharmacological Space")…*
- Delivering **services** to support on-going drug discovery programs in pharma and public domain
- **Mix ideal with the pragmatic**. Build open that can accommodate non-open components in the real world.

Guiding principle is open access, open usage, open source
- Key to standards adoption -

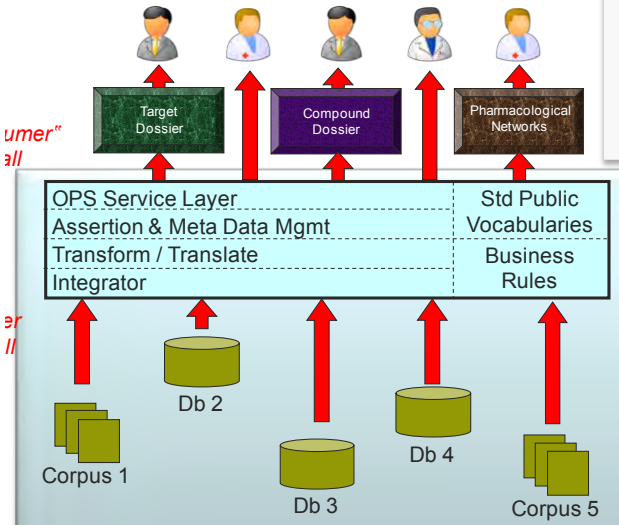
Open PHACTS
Open Pharmacological Space

Major Work Streams

Build: OPS service layer and resource integration “commons”

Drive: Development of exemplars & applications

Sustain: Community engagement and long-term sustainability



Work Stream 2: Exemplar Drug Discovery Informatics tools
Develop exemplar services to test OPS Service Layer
Target Dossier (Data Integration)
Pharmacological Network Navigator (Data Visualisation)
Compound Dossier (Data Analysis)

Work Stream 1: Open Pharmacological Space (OPS) Service Layer

Standardised software layer to allow public DD resource integration

- Define standards and construct OPS service layer
- Develop interface (API) for data access, integration and analysis
- Develop secure access models

Existing Drug Discovery (DD) Resource Integration

Open PHACTS
Open Pharmacological Space

OPS Services

- **Integrate data** on target expression, biological pathways and pharmacology to identify the most productive points for therapeutic intervention
- **Investigate** the *in vitro* pharmacology and mode-of-action of novel targets to help develop screening assays for drug discovery
- **Compare** molecular interaction profiles to assess potential off-target effects and safety pharmacology
- **Analyse** chemical motifs against biological effects to deconvolute high content biology assays

OPS Semantic Fabric – Linked Data to Linked Knowledge

- Publishing information as linked statements that are sufficiently well described that the information can be automatically linked, brokered and processed.

Define, index & link across ….

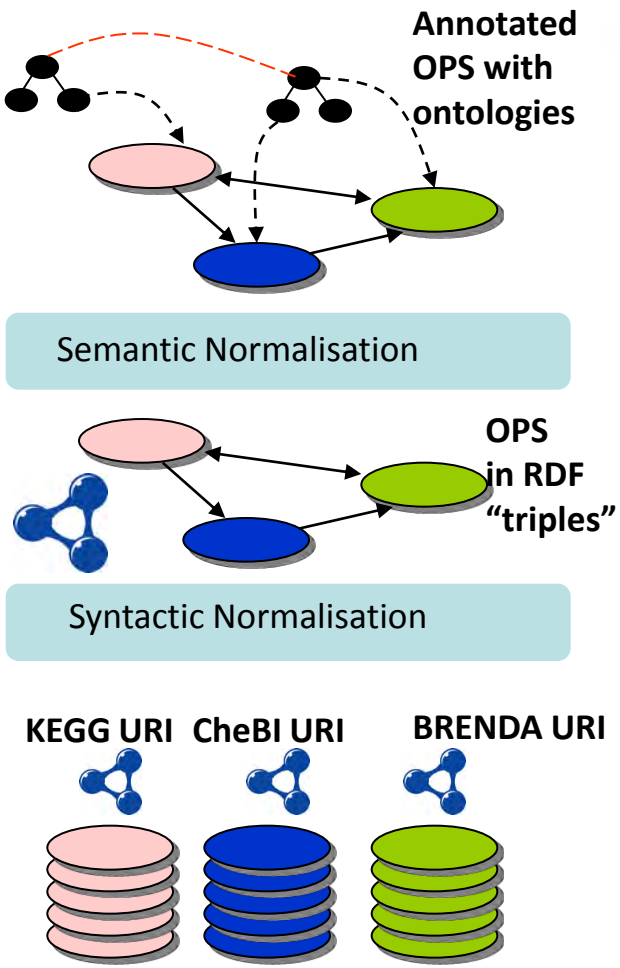
- **Common data identity and entity, data structure, data meaning**

Open PHACTS
Open Pharmacological Space

The Semantic Web & Linked Data

Harmonising data sets
Id & Syntactic & Semantic

Concept vocabulary services
Identity resolution services
Concept mapping services
Mapping identifier services



Linked Knowledge

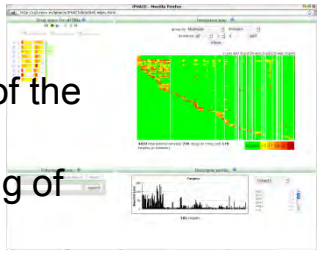
Linked Data

Open Data

Open PHACTS
Open Pharmacological Space

A use case driven approach

Polypharmacology browser map coverage of the chemo-biological space for polypharmacological profiling of small molecules



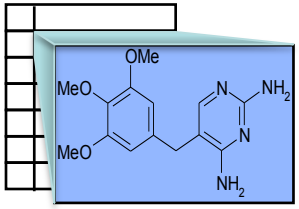
Developers (Builders)

Prioritised research questions

Prioritised data sources

Exemplars

Chem/bio space navigator of sets of pharmacologically annotated small molecules, by chemical substructures, pharmacophores, biological activities



End users (Drivers)

Target dossiers about targets, incorporating related information on sequences, structures, pathways, diseases and small molecules

Open PHACTS
Open Pharmacological Space

A use case driven approach

Developers
(Builders)

Prioritised research questions

Prioritised data sources

Bench mark Pilots

End users
(Drivers)

Fusion/aggregation of data from different domains to improve predictions of drug-transporter interactions

Combination of physicochemical data & data from transporter interaction for prediction of blood-brain barrier permeation and tissue distribution

Target validation work-bench: *in silico* target validation studies

Example Research questions

- Give all compounds with IC50 < xxx for target Y in species W and Z plus assay data
- What substructures are associated with readout X (target, pathway, disease, …)
- Give all experimental and clinical data for compound X
- Give all targets for compound X or a compound with a similarity > y%

- 73 questions identified across consortium

Open PHACTS
Open Pharmacological Space

Prioritised Research Questions Analysis

- **Prevalent Concepts**
  - Compound
  - Bioassay
  - Target
  - Pathway
  - Disease
- **Prevalent data relationships**
  - Compound – target
  - Compound – bioassay
  - Bioassay – target
  - Compound – target – mode of action
  - Target – target classification
  - Target – pathway
  - Target – disease
  - Pathway - disease
- **Required cheminformatics functionality**
  - Chemical substructure searching
  - Chemical similarity searching
- **Required bioinformatics functionality**
  - Sequence and similarity searching
  - Bioprofile similarity searching

Open PHACTS
Open Pharmacological Space

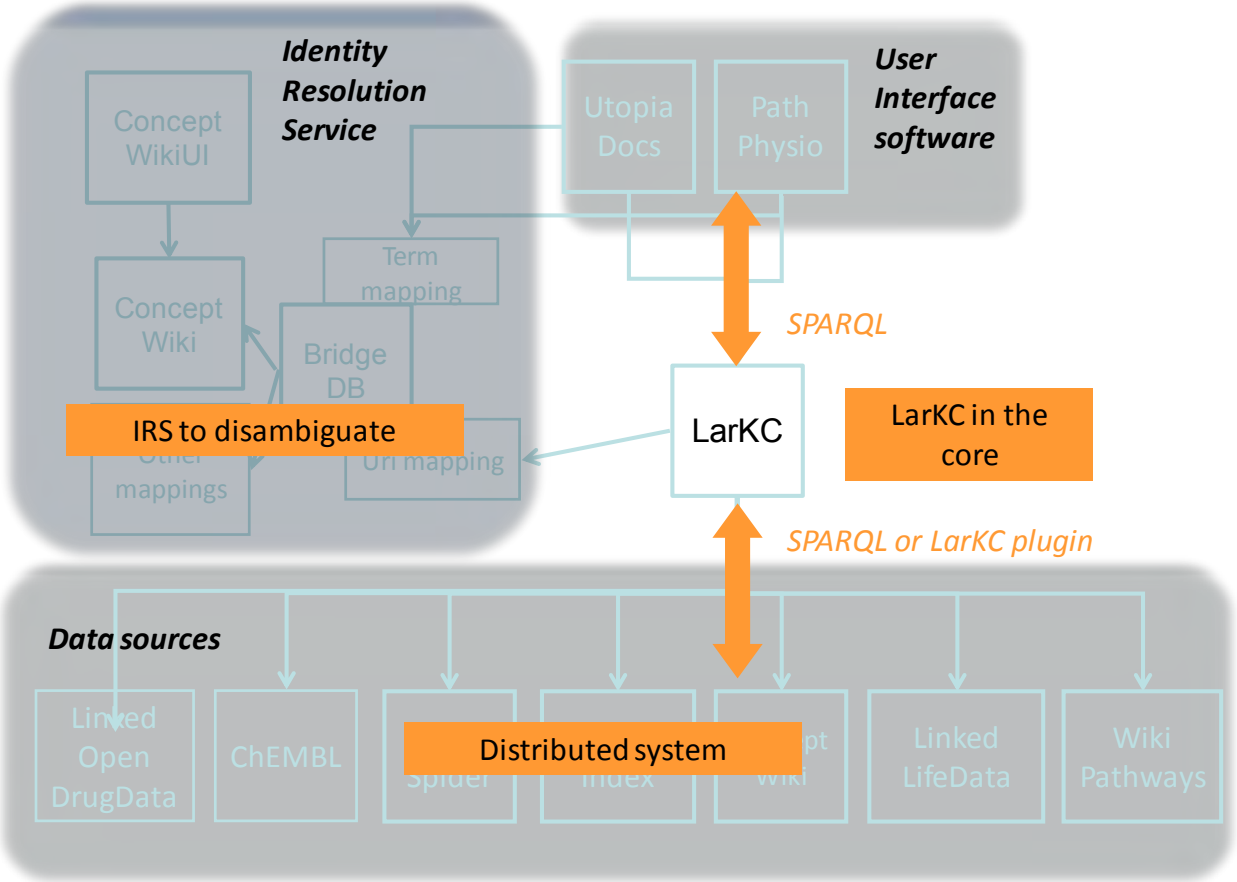
Selection of prioritised data sources

- **Chemistry**
  - ChEMBL
  - DrugBank
  - ChEBI
  - PubChem
  - ChemSpider
  - Human Metabolome DB
  - Wombat (commercial)
- **Ontologies**
  - AmiGo (The Gene Ontology)
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
  - OBI (The Ontology for Biomedical Investigations)
  - Bioassay Ontology
  - EFO (Experimental Factor Ontology)
- **Biology**
  - EntrezGene
  - HGNC
  - Uniprot
  - Interpro
  - SCOP
  - Wikipathways
  - OMIM
  - IUPHAR

Agile Development: 6 month "lash up"

- Produce a working "lash up" system
- Constrained to technologies in consortium + a few data sources
- Focused on 2 prioritized research questions (Q15 and Q30)
  - **Q 15: All oxidoreductase inhibitors active <100nM in both human and mouse**
  - **Q 30: For a given compound [clozapine], give me the interaction profile with [human or mouse] targets**
- Minimum requirements: two data sources (one targets, one compounds) and able to produce answers in "manual time".
  - Brenda, KEGG, PDSP, ChEMBL, ChEBI, ENZYME DB, Chem2Bio2RDF

Build a lash up



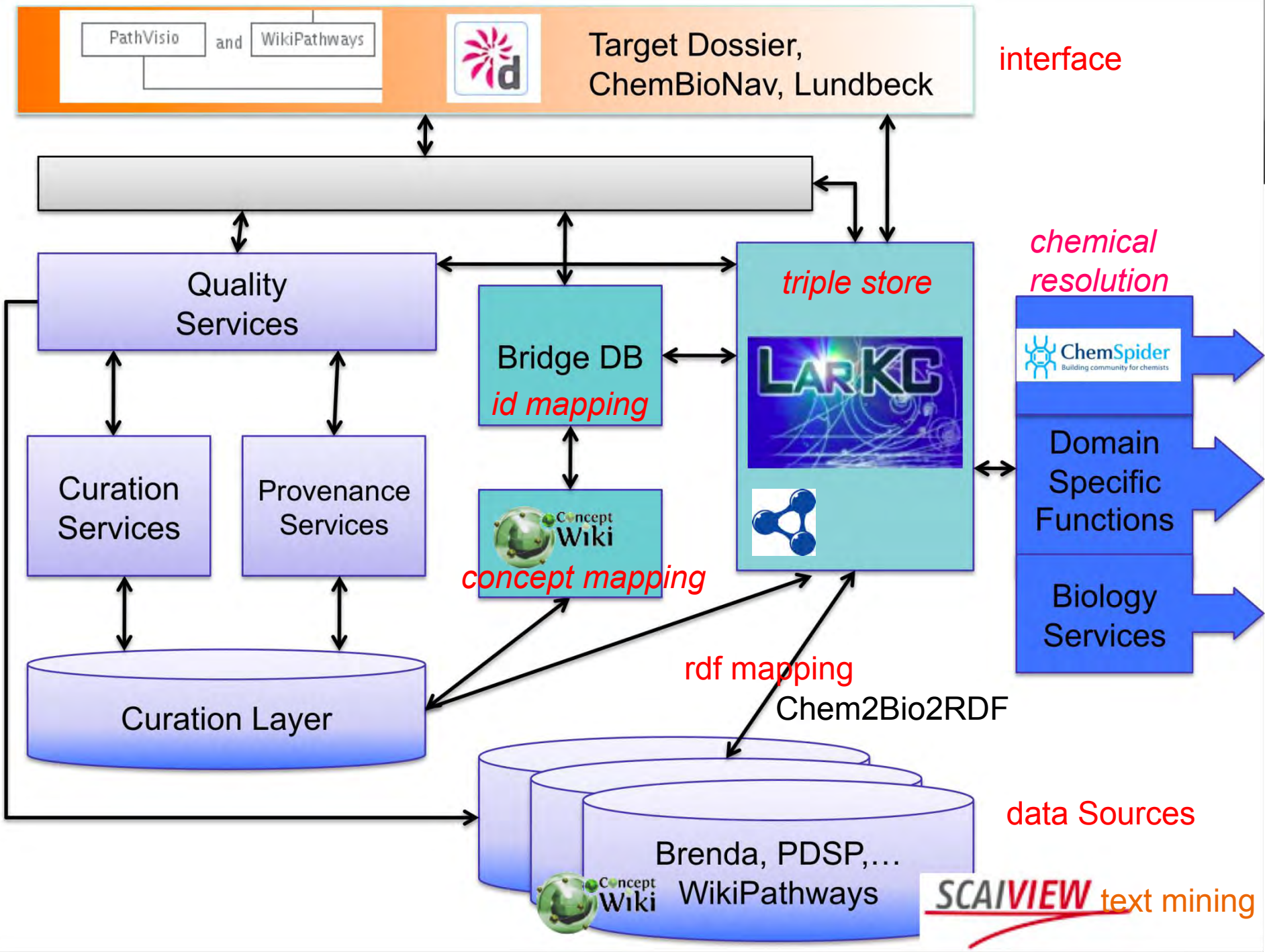
Outcomes of exercise:

- Team building
- Performance / scalability analysis
- Does it provide an adequate answer to the questions 15 and 30?
- Demo for users (drive group) to recalibrate build tasks in order to better respond to user requirements

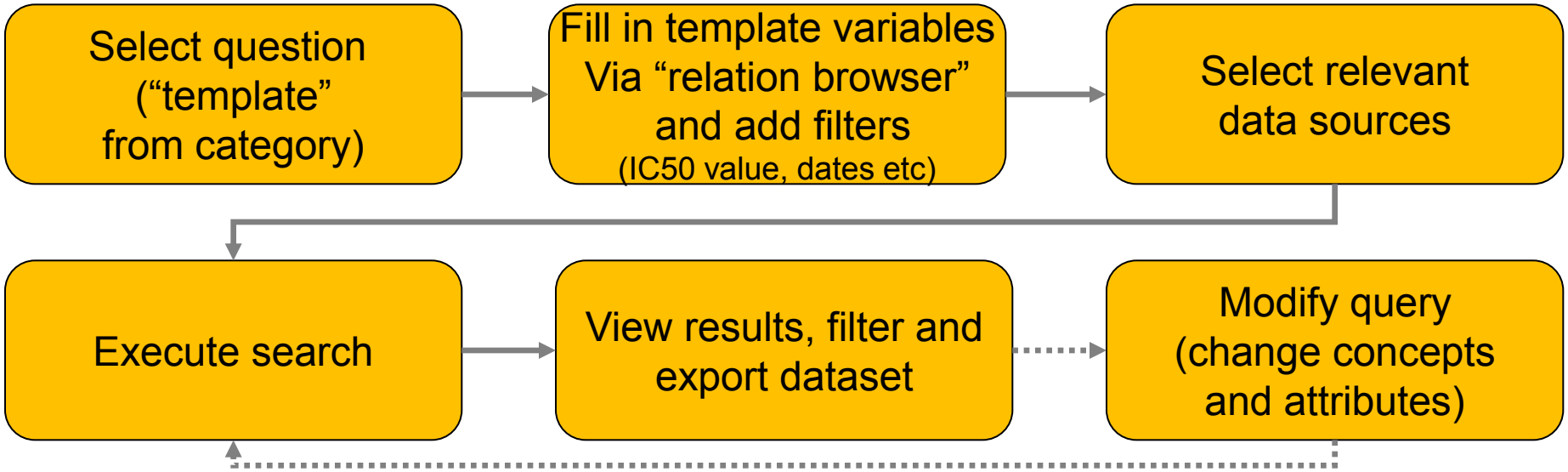
Build a lash up

Outcomes of exercise:

- Team building
- Performance / scalability analysis
- Does it provide an adequate answer to the questions 15 and 30?
- Demo for users (drive group) to recalibrate build tasks in order to better respond to user requirements


PathVisio
and
WikiPathways
Target Dossier,
ChemBioNav, Lundbeck
interface
Quality
Services
Bridge DB
id mapping
triple store
chemical
resolution
ChemSpider
Building community for chemists
Curation
Services
Provenance
Services
Concept
Wiki
concept mapping
Domain
Specific
Functions
Biology
Services
rdf mapping
Chem2Bio2RDF
Curation Layer
Brenda, PDSP,…
WikiPathways
data Sources
SCAIVIEW text mining

GUI - User suggestions for workflow



Open PHACTS
Open Pharmacological Space

“Lash up” Facts and Figures

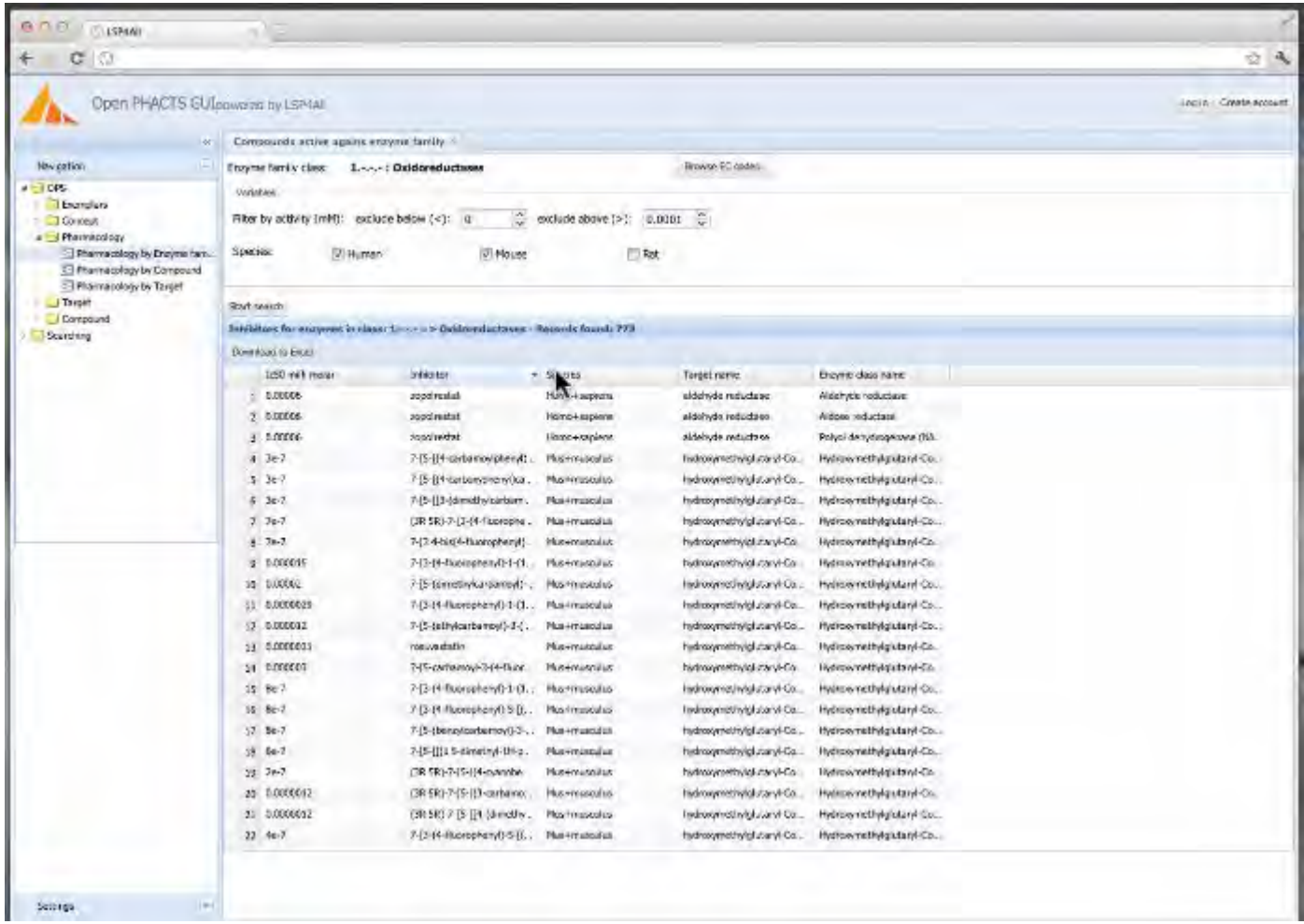
Total Number of RDF Triples: 62,054,627

Data Source	Number of RDF triples	Triple Producer
BRENDA	345,935	In house
PDSP	719,991	In house
ENZYME DB	39,540	UniProt
KEGG	247,292	Chem2Bio2RDF
ChEMBL	57,795,793	Chem2Bio2RDF
ChEBI	2,906,076	Chem2Bio2RDF

ScaiView (text mined triples) acquired - 4 billion triples
LinkedLifeData - 5 billion triples
Chem2Bio2RDF - 83 million triples
ConceptWiki - 300 million triples
Bio2RDF - 30 billion triples

Open PHACTS
Open Pharmacological Space

Lash Up Demo

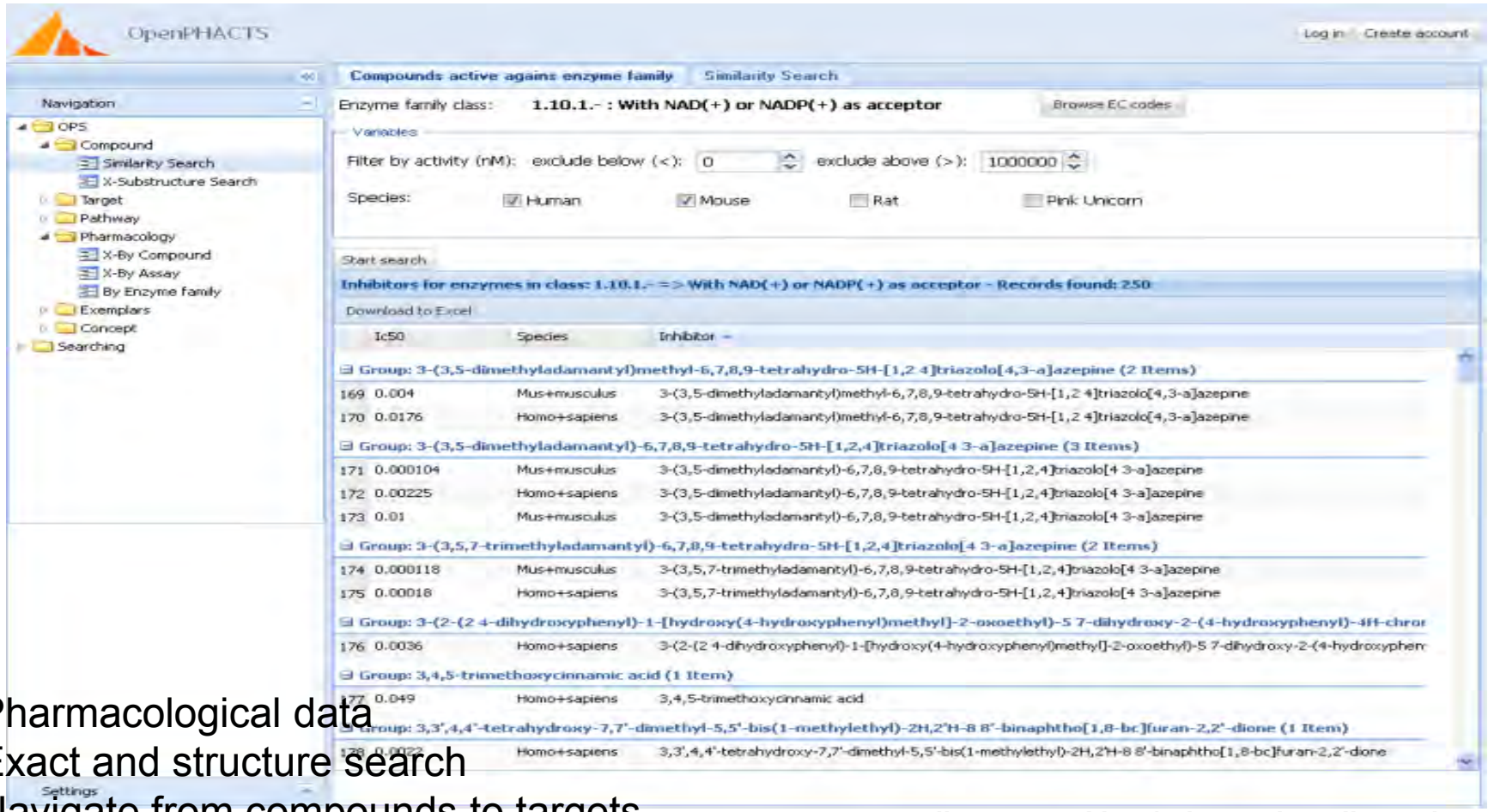


http://www.youtube.com/OpenPHACTS

Open PHACTS
Open Pharmacological Space

LSP4All (Lundbeck) Generic Interface search by enzyme family

Q15: All oxidoreductase inhibitors active <100nMolars in both human & mouse



Pharmacological data
Exact and structure search
Navigate from compounds to targets

Credit: Sune Askjær / Claus Stie Kallesøe (Lundbeck)

UTOPIA Documents (U Manchester)

Bioorganic & Medicinal Chemistry Letters 19 (2009) 4183–4190

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Discovery and functional evaluation of diverse novel human CB_1 receptor ligands

Nicolas Foloppe [a,*], Karen Benwell [a], Teresa D. Brooks [a], Guy Kennett [b], Antony R. Knight [b], Anil Misra [b], Nathaniel J. T. Monck [b]

[a] Vernalis (R&D) Ltd, Granta Park, Abington, Cambridge, CB21 6GB, UK
[b] Vernalis (R&D) Ltd, Oakdene Court, 613 Reading Road, Winnersh, RG41 5UA, UK

ARTICLE INFO

Article history:
Received 8 May 2009
Revised 25 May 2009
Accepted 27 May 2009
Available online 2 June 2009

Keywords:
Cannabinoids
Drug design
Obesity
Pharmacophore
Virtual screening

ABSTRACT

Ligand-based virtual screening with a 3D pharmacophore led to the discovery of 30 novel, diverse and drug-like ligands of the human cannabinoid receptor 1 (hCB_1). The pharmacophore was validated with a hit rate of 16%, binding selectivity versus hCB_2, and expected functional profiles. The discovered compounds provide new tools for exploring cannabinoid pharmacology.

The cannabinoid system was initially identified as mediating the effects of the major psychoactive components of *Cannabis sativa*, Δ^9-tetrahydrocannabinol. To date, this system includes at least two receptors, CB_1 and CB_2, the biology of which has been extensively reviewed.[1–3] CB_1, a GPCR coupled to inhibitory G-proteins $G_{i/o}$ is expressed in the central nervous system where it is found on presynaptic terminals, serving to modulate neurotransmitter release. It is also found in peripheral tissue. CB_1 is primarily found

Endocannabinoids like anandamide and 2-arachidonoyl glycerol are elevated in the plasma, adipose tissue and pancreas of obese humans and animals.[30] In genetically obese rodents endocannabinoid levels are elevated in the hypothalamus, an area of the brain that modulates feeding behavior.[31] Centrally, endocannabinoids elicit feeding behavior via activation of CB_1 receptors.[7] Peripheral CB_1 activation causes lipogenesis in adipose tissue and liver, and reduces the oxidation of free fatty acids (FFA).[32] Also

Back to overview

10.1016/j.bmcl.2009.05.114

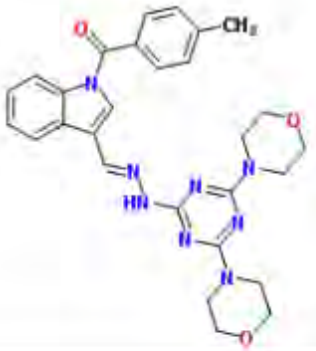
(3-((2-(4,6-...yl)methanone

(3-((2-(4,6-dimorpholino-1,3,5-triazin-2-yl)hydrazono)methyl)-1H-indol-1-yl)(p-tolyl)methanone

[More...]

Ki Value : 2299

Ki of 2299 in 19520572



1-(4-(3-(2,4...-yl)ethanone

1-(4-(3-(2,4-dimethoxyphenyl)-1-m-tolyl-1H-pyrazole-5-carbonyl)piperazin-1-yl)ethanone

[More...]

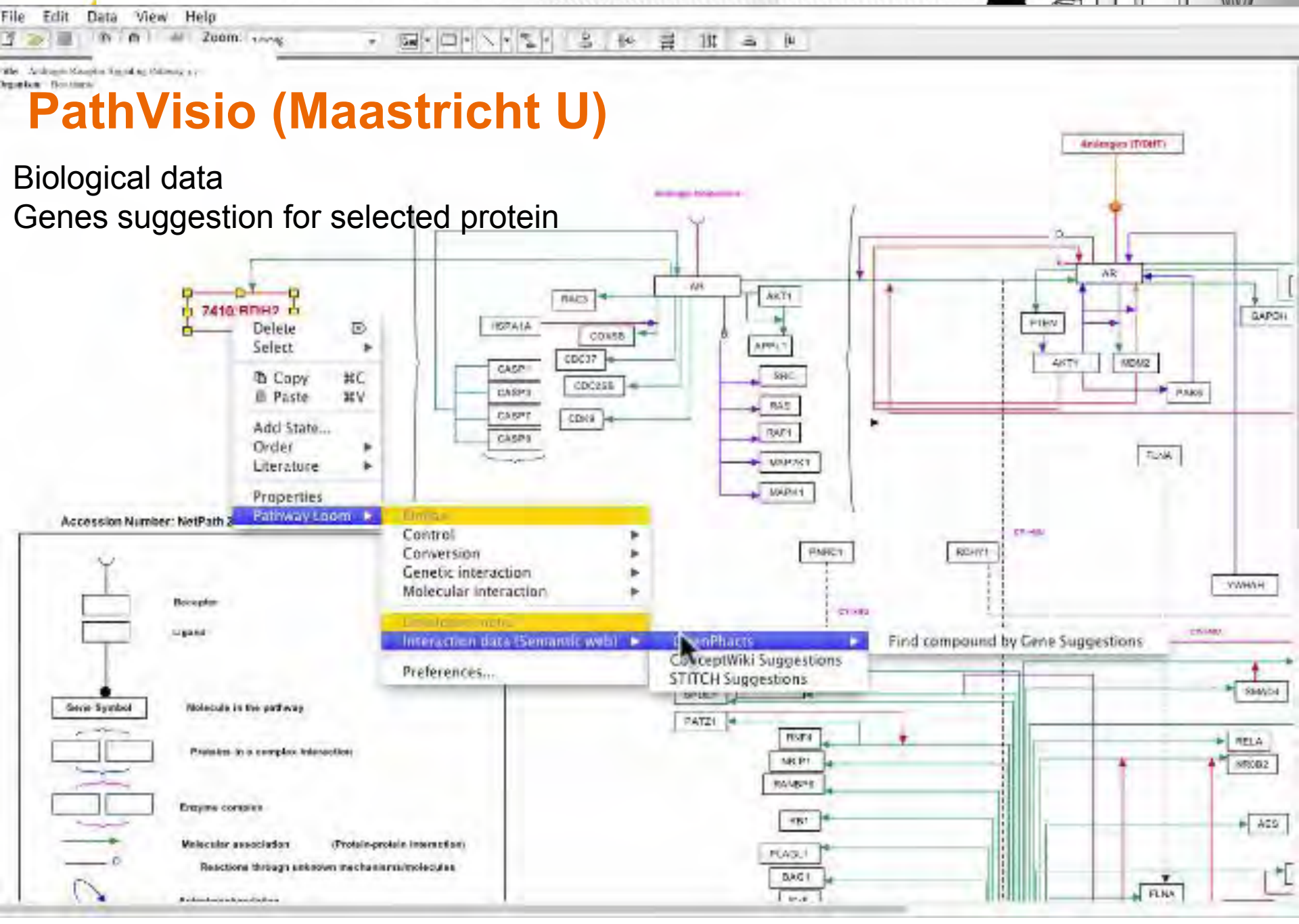
Ki Value : 1943

Ki of 1943 in 19520572

Look up

PathVisio (Maastricht U)

Biological data
Genes suggestion for selected protein



Open PHACTS
Open Pharmacological Space

"Lash Up" Sanity Check

Q15: All oxidoreductase inhibitors active <100nM in both human and mouse

- IC50 values and compounds fully coincident between the automatic and manual search.
- "Lash up" identified a compound lost in the manual search (Raloxifene) which value after doing a new manual search was correct.
- Manual search took 3 days (Mabel Loza's team @USC)
- Automated search took milliseconds.

Representing Text Mining Results for Structured Pharmacological Queries
http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/PostersDemos/iswc11pd_submission_19.pdf

Open PHACTS
Open Pharmacological Space

Onwards and Upwards

- Connection between developers and users
  - Solidify interfaces for exemplar developers
  - Review lash up for technology, content and exemplars
- Architecture
- Services: e.g. entity identification and resolution and representing similarity, ORCID, DataCite
- Models: RDF / Nanopublication model spec and guidelines
- Tender documents for commercial storage providers

Prototype

- March 2012: Internal Prototype Delivery
- September 2012: Release 1st Prototype

Early community adoption

- Talking to major partners to take part in the project

Open PHACTS
Open Pharmacological Space

Nanopublications
Capturing scientific information in the Triple Store

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

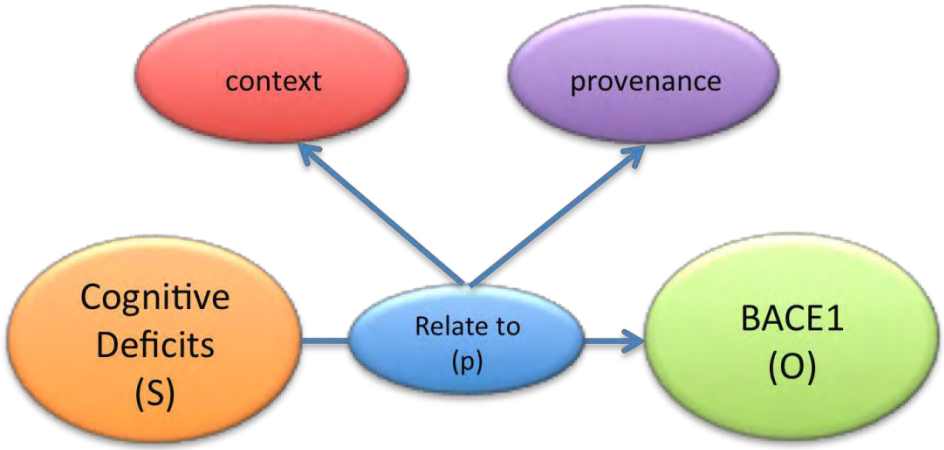
NATURE GENETICS | COMMENTARY

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 Medic
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

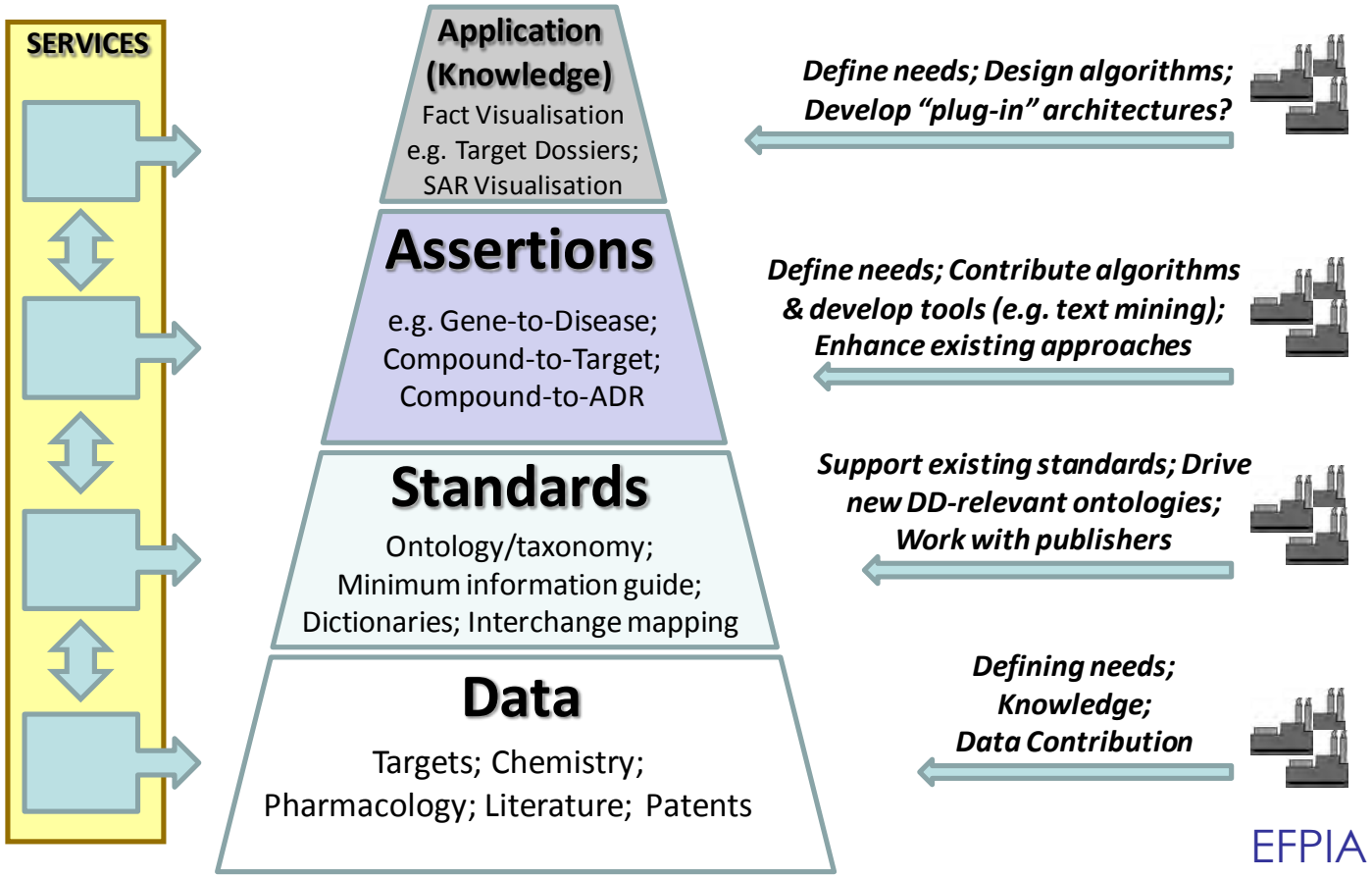
Johannes Velterop
Concept Web Alliance / NBIC
9 Benfleet Close
Cobham, Surrey, KT11 2NR
United Kingdom
jan.velterop@nbic.nl

ABSTRACT

Newer standards like RDFa also facilitate this and integrate with

Open PHACTS
Open Pharmacological Space

OPS Component Stack



Open PHACTS
Open Pharmacological Space

Open Flavours

- **OPS Open** - open access to all
- **OPS Consortia** - data sets licensed just to the consortia
- **OPS Academia** - fully open to academia.
- **“My OPS”**

- **Open Source**
- **Open Access Infrastructure**. GUI and back-end platform, online at openphacts.org or download both + data for local setup.
- **Open Services**: for example, RSC services.
- **Open Data + Private Data**: licensing fun for all the family.
- **Commercial providers:** abstract service interface to swap in commercial and open source platforms

Open PHACTS
Open Pharmacological Space

OPS Community Workshops

Focus on different aspects of drug discovery, the technology used, **data sharing, sustainability, licensing** and practical applications.

- 1st Volendam (near Amsterdam) September 19-20, 2011
  - Joint with GEN2PHEN
- Solving Bottlenecks in Data Sharing in the Life Sciences

- 2nd Location TBD
  April 16-17, 2012

Open PHACTS
Open Pharmacological Space

Pistoia Alliance

- Pre-competitive Pharma industry body that works to develop and promote standards for software interoperability
- Active in many areas of Pharma software pipeline, including “Information Ecosystem”.
- Pistoia project “**Semantic Enrichment of Scientific Literature (SESL)**” directly related to Open PHACTS http://www.pistoia-sesl.org
- Open PHACTS seen as important to the Pistoia mission

Pistoia Alliance

Open PHACTS
Open Pharmacological Space

Academia-Commercial Venture

- **Focus**
  - One area - **pharmacology**
  - “Production Level” software
  - Currency/Updates & Licensing key
  - Semantic Pragmatics: everyday use by scientists not informaticians
- **Future**
  - An **infrastructure** that can be built upon, to provide a stable foundation for further pre-competitive informatics collaboration
  - Sustainability

Developers
(Builders)

End users
(Drivers)

Open PHACTS
Open Pharmacological Space

Outside interest?

- **Infrastructure**
  - Reduce delivery and platform costs
  - RDF delivery to customer
  - Data and database preservation
- **Sustainability**
  - By number of users
  - By number of data providers
  - By service providers
  - To maintain after project ends

Open PHACTS
Open Pharmacological Space

http://www.openphacts.org

slideshare.com/open_phacts

@Open_Phacts

Thanks to Carole Goble, Lee Harland, Antony Williams