Open PHACTS

Milestone 2

Engineering structure, questions and requirements, governance in place. Full consortium meeting with lash-up discussion held

Prepared by UNIVIE, Pfizer, VUA, UNIMAN, RSC, LUMC
Approved by Steering Committee

October 2011
Version: 1.0

Project title: An open, integrated and sustainable chemistry, biology and pharmacology knowledge resource for drug discovery
Instrument: IMI JU
Contract no: 115191

Start date: 01 March 2011
Duration: 3 years

Nature of the Deliverable	
Report	**x**
Prototype	
Other	
Dissemination level	
Public dissemination level	
For internal use only	**x**

Definitions

- Partners of the Open PHACTS Consortium are referred to herein according to the following codes:

  Pfizer – Pfizer limited – **Coordinator**
  UNIVIE – Universität Wien – **Managing entity of IMI JU funding**
  DTU – Technical University of Denmark – DTU
  UHAM – University of Hamburg, Center for Bioinformatics
  BIT – BioSloveIT GmBH
  PSMAR – Consorci Mar Parc de Salut de Barcelona
  LUMC – Leiden University Medical Centre
  RSC – Royal Society of Chemistry
  VUA – Vrije Universiteit Amsterdam
  CNIO – Spanish National Cancer Research Centre
  UNIMAN – University of Manchester
  UM – University of Maastricht
  ACK – ACKnowledge
  USC – University of Santiago de Compostela
  UBO – Rheinische Friedrich-Wilhelms-Universität Bonn
  AZ – AstraZeneca
  GSK – GlaxoSmithKline
  Esteve – Laboratorios del Dr. Esteve, S.A.
  Novartis – Novartis
  ME – Merck Serono
  HLU – H. Lundbeck A/S
  E.Lilly – Eli Lilly

- **Grant Agreement**: The agreement signed between the beneficiaries and the IMI JU for the undertaking of the Open PHACTS project.

- **Project**: The sum of all activities carried out in the framework of the Grant Agreement.

- **Work plan**: Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried, out as specified in the Grant Agreement.

- **Consortium**: The Open PHACTS Consortium, composed of the above-mentioned legal entities.

- **Project Agreement**: Agreement concluded amongst Open PHACTS participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

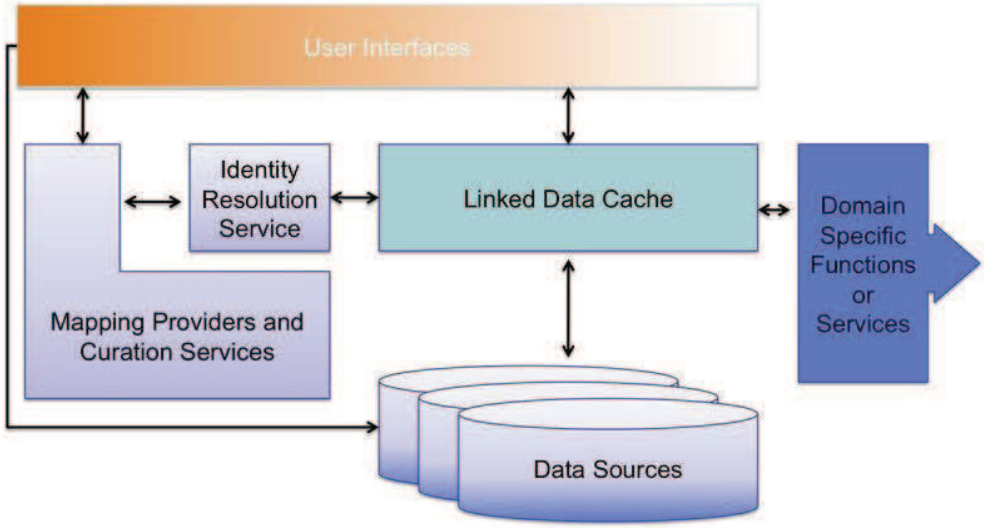
1. Introduction

The Open PHACTS project aims to develop an open access innovation platform, Open Pharmacological Space (OPS), utilizing a semantic web approach. The OPS system will support drug discovery projects by providing data, vocabularies and infrastructure open to the public domain. In order to guarantee broad applicability and acceptance by a wide user community, the Open PHACTS consortium chose a use case driven approach. Therefore, the Open PHACTS system needs to be created in several release cycles and the first functioning prototype – the first lash-up – is supposed to be demonstrated after 6 months at a Steering Committee meeting.

2. Milestone

The second milestone of the Open PHACTS project, due after month 6, presents the aim of the first 6 months of the project: to use existing infrastructure to “test the water” on how a semantic system might be used to address common pharmacology questions. The ultimate aim of this work was to build a system – the lash up – to present at the 2nd Steering Committee meeting in Hamburg on September 29th 2011. This was indeed achieved and the first version of a functioning Open PHACTS system was demonstrated and shown to answer key business questions as directed by the project leaders. Delivery of the system is available as a video presentation, at http://www.youtube.com/watch?v=DH_RInO4-P0 and also attached as power point presentation.

As directed by the project plan, the architecture was based on pre-existing components available via the consortia members. As the majority of this infrastructure has been described in previous deliverables it will not be regurgitated here. More information can be found on the wiki (https://wiki.openphacts.org/index.php/PlatformArchitecture) and in D3.2.3, the architecture. The summary below presents the first, working architecture plan which was used to define how the lash-up was devised and ultimately delivered a working prototype on time in the project.



3. Verification

Milestone 2 is linked to the 2nd Steering Committee meeting in Hamburg on September 29-30, 2011, where Paul Groth successfully presented the first lash-up – a functioning prototype for the Open PHACTS system.

4. Next Steps

The technical team will run a 'continous build' for the software, meaning that approximately every 2 weeks incremental updates will be released to the internal community. This allows to respond to changing priorities in a very rapid manner. However, the team will also deliver major releases, according to the OPS time-table. This development will be guided by the exitsting project deliverables as well as the direction from the drive cluster. The release schedule will follow the pattern below:

1. Pilot internal release: Dec 2011: To include additional pharmacology data, enhanced GUI, prototype developer/exemplar API.

2. First Internal Release: March 2012: To include full install instructions, address a new type of business question (driven by WP5/6, e.g. Pathway-compound relationships) and include additional data and GUI functionality. First prototype of automatic concept and identity resolution (IRS, Concept Wiki).

3. First Public Release: Sept 2012: To include a fully available system that external parties can access via the web and answer three distinct classes of business question covering more than 10-nanopublication data sources.

Open PHACTS
Open Pharmacological Space

6-month Lash-up

Technical Side

Paul Groth @pgroth
http://www.few.vu.nl/~pgroth

Open PHACTS
Open Pharmacological Space

Outline

- What is a lashup and why?
- Goals
- Results and Demo
- Feedback….
- Technical lessons learned and path forward

Open PHACTS
Open Pharmacological Space

First milestone: Deliver a working prototype of "Open Pharmacological Space" in 6 months using established technology

Characteristics:
- A working system (not a mockup)
- Use existing technologies present now in the consortium
- Rough

Why:
- Don't get bogged down in design dicussions
- No fear of change
- Have a test environment
- Understand where the gaps are in technology
- A shooting target

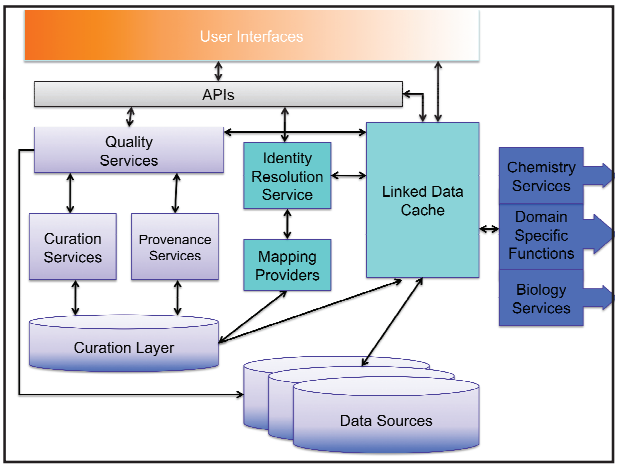
Open PHACTS
Open Pharmacological Space

6 Month Prototype Goals

1. Team building exercise
2. Answer two 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
3. Two data sources (one targets, one compounds)
4. Produce answers in "manual time"
5. Demo
6. Help set the direction on development

Open PHACTS
Open Pharmacological Space

RESULTS



Open PHACTS
Open Pharmacological Space

- Current Data Sources
  - Brenda
  - ENZYME DB
  - Kegg
  - PDSP
  - ChEMBL
  - ChemSpider

Open PHACTS
Open Pharmacological Space

Prioritised Research Questions

Number	sum	Nr of 1	Question
15	12	9	**All oxido,reductase 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

www.openphacts.org

Open PHACTS
Open Pharmacological Space

"the IC50 values and compounds are fully coincident between your automatic and our manual search. Also you *identify a compound we lost in the manual search* (Raloxifene) which value after doing a new manual search is correct. So, your software are the winner **(faster, more than 3 days for us!, and complete).**" – Mabel Loza

Open PHACTS
Open Pharmacological Space

Text Mined data from ScaiView

- Input

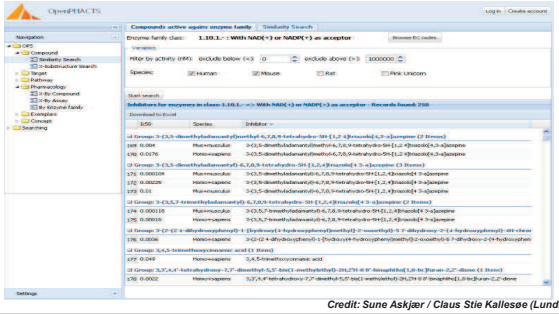
MAP files: References to external data sources (2,499,772 triples)
- SYN files: Dictionaries (7,719,711 triples)
- Sentence files: Sentence based full text (561,174,640 triples)

- Output
- PRT files: Hits (2,574,426,040 triples)
  - General information about matched entities
  - Synonyms
  - Positional information
  - Scoring information for machine learning based filtering

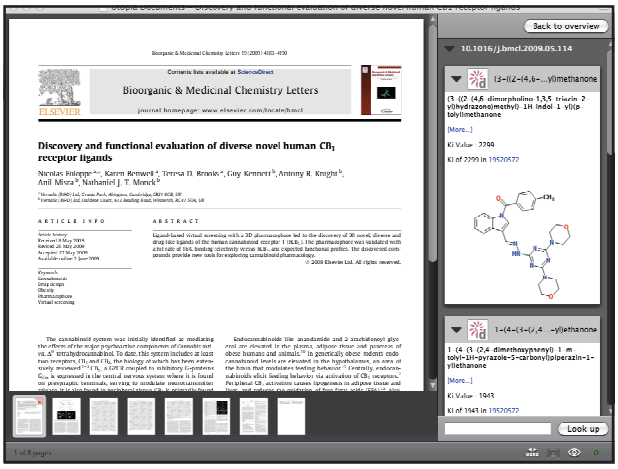
3.1 ^10 triples on disk: 221 GB

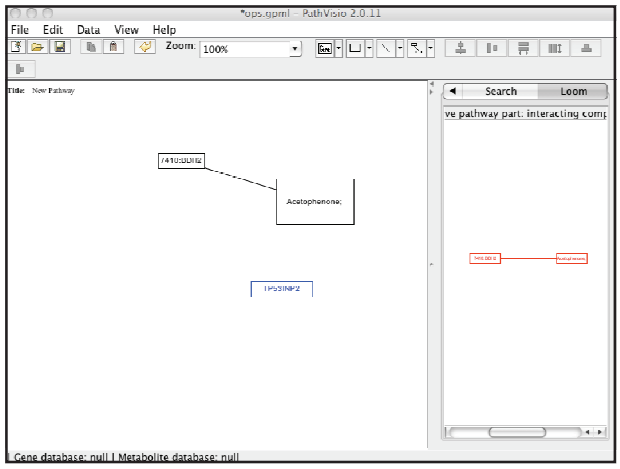
- Triple Stores (Insert 3,145,820,163 triples)
- Virtuoso Universal Server OpenSource (one core) over web GUI
  - Stops reacting at all
- Owlim-SE (8 cores) over API
  - Getting incredible slow after ~200 million triples
- 4store, Mulgara, … not tried due to missing distributed modus

Open PHACTS
Open Pharmacological Space



Credit: Sune Askjær / Claus Stie Kallesøe (Lundbeck)





Open PHACTS
Open Pharmacological Space

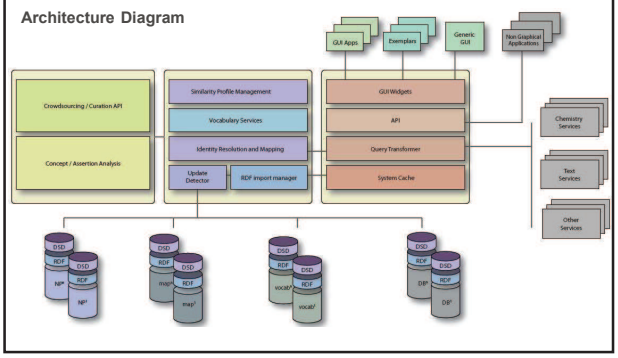
Technical Issues

- Major issue with scale in terms of storage
- Need simpler apis for GUI developers
- Clear processes for RDF exposure/integration
- Richer mechanisms for identity resolution

Open PHACTS
Open Pharmacological Space

TOWARDS SOLUTIONS

Open PHACTS
Open Pharmacological Space



Open PHACTS
Open Pharmacological Space

Storage

- Testing a triple enabled NOSQL + triples data management approach
- Multiple interface approach exposing to developers performance/expressiveness trade-offs

Open PHACTS
Open Pharmacological Space

RDF / Nanopublication Guidelines

- Input
  - Text mining schema (Bonn)
  - WikiPathway Schemas (Masstricht)
  - ConceptWiki (Leiden)
  - Nanopublications papers/discussion pieces (Leiden/VU)
- Workshop
- Draft Guidelines

Open PHACTS
Open Pharmacological Space

Vocabulary and Identity Management

- Performance Improvements on Concept Wiki
- Bridge DB improvements and integration
- New approaches for representing similarity
- Improved chemical resolver from ChemSpider

Open PHACTS
Open Pharmacological Space

Conclusions

- Strong technical basis
- Led to a comprehensive architecture
- New approaches still need to be tested
- Shows chemistry and target data working together
- **A lot done in 6 months with small developer team**