Open PHACTS

Milestone 3

First working release demonstrate use, associate partners enlisted and integrated

Prepared by UNIMAN, Janssen, HLU, VUA, ConnDisc, Pfizer, SIB, RSC, UNIVIE

Approved by Steering Committee

April 2012
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	**x**
For internal use only	

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 – BioSolveIT 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 proposal defines a major technical milestone at the end of Month 12, specifically a working version of the Open PHACTS system with the milestone verified by "Prototype delivered within the consortium and running". This document describes how this was achieved. This software was designated "Version 0.2" (or v0.2 for short) and this is used in the remainder of this document.
We have set up a process for adding Associated Partners and we successfully go along with enlisting and integrating them in the project.

2 Verification

Version 0.2 of the Open PHACTS Core platform and GUI was released within the consortium on March 30, 2012. The Milestone of delivery of a pilot version of the Open PHACTS platform and core GUI to the consortium was achieved. The system is now being used by the members of the Open PHACTS consortium, led by the user testing team and a process of gathering bugs, feedback and future needs is now in operation.
The process of adding associate partners was officially approved at the Steering Committee Meeting in Stevenage on February 7-8, 2012.

3 Prototype delivered within the consortium and running

3.1 Release Timeline

- Software coding ceased and the final package was delivered on the 2nd March 2012
- An acknowledgement was sent to the ExCo and PMU to notify the milestone had been met
- There then followed a period of technical and scientific testing until 16th March, followed by rapid bug-fixing before more wider release
- Full scientific testing and phased rollout began w/c 19th March, with email and webinars presenting the software
- A full release note guidance document was created and made available to all members of the consortium

3.2 Datasets in this release

- Enzyme: Necessary for the enzyme classifications
- ChEMBL v2: Essential SAR database
- Drugbank v3: Provides additional information on known drugs
- These are connected through ConceptWiki IDs (which contains Uniprot mappings) and ChemSpider Ids

3.3 Scientific Functionality

The following queries are possible:

- Query pharmacology data by target (via target name search)
- Query pharmacology data by compound (via compound name search)
- Query pharmacology data by target family (via Enzyme hierarchy)
- Structure searches to identify compounds in the system (which can then be used in pharmacology searches)
- Text search to identify targets/proteins

3.4 Functionality Details

- Text search now uses ConceptWiki. The interface has “Google-like” auto-complete functionality to aid entry. The identifier resolution service takes care of mapping the input concept to the required database identifiers in the system
- A more integrated view across targets is now available; protein identifiers in the different databases are now fully connected
- The system now enables pharmacology queries over the full ChEMBLdatabase
- Many interface enhancements and bug fixes, including pagination, table resizing and presentation enhancements (better labels etc.)
- More robust chemical structure searching via the ChemSpider API has now been implemented

3.5 Technical Achievements

- Sourced required data, converting to RDF or adjusting RDF as required (much of the existing RDF requiring modification to work correctly in Open PHACTS)
- Loaded approximately 70 million triples into the semantic data cache
- Dynamic identifier mapping is facilitated via the Identity Mapping Service using 7 linksets and providing 3,149,260 mappings that relate the core datasets. Additional non-critical mappings are ready to load once we are sure they will not affect system performance
- Created a real-time text to concept API using ConceptWiki and synonyms sourced from multiple databases, including validated synonyms from ChemSpider
- Data integration across the different databases is demonstrated,
- Chemistry normalisation and mapping through a ChemSpider pipeline is in place
- Created the Open PHACTS API which delivers results of optimised queries over the platform
- Demonstrated integration with “external” web services (ChemSpider) is a critical element of the system

- Created a pharmacology-oriented interface using real data, seeded by the Lundbeck LSP platform and querying the semantic data cache
- Deployed the GUI widget framework now in full use within the platform, and successfully used by multiple groups on the project

3.6 Known Limitations

- Currently the pharmacology data is ChEMBL02 source from the Chem2Bio2RDF project. This means that users cannot compare data to that at the ChEMBL website, which is currently version 13. We are planning to update the OPS platform to ChEMBL 13 as soon as possible.
- For target-based pharmacology, the user must currently select the specific species of protein (f.e. “mouse”) and will only return pharmacology data for that species. We are working on the next iteration of this, whereby data can be retrieved for all species and then filtered as required
- There will be duplicate rows present in the pharmacology data. These will be distinct ChEMBL records but often covering the same compound-target relationship. We will look for guidance from the STF as to how to de-duplicate once the system is available
- The user may be able to select targets or compounds that are not associated with any pharmacological data. These selections will return no results.
- Filtering, export and sorting of pharmacology records will only take place on the records currently loaded into the interface. We are working on a solution such that these operations can take place on the entire dataset

3.7 APIs available

These methods are documented at https://wiki.openphacts.org/index.php/Core_API.

3.8 Testing

It is critical that scientific software be tested before providing to users, who will expect that the results they obtain result from a tested system. While the v0.2 of the software was intended as an internal “demonstrator” release it still, nevertheless, required considerable testing. Within Open PHACTS our testing is split roughly into two separate elements – technical testing where we ensure the platform is robust and eliminate software bugs; and scientific testing, where we assess the scientific accuracy and coverage of the results obtained. In reality there is overlap between the two test data sets which is used to ensure technical algorithms are working correctly. Over the coming months WP5/6 will develop a barrage of scientific use-cases and tests that will be used to validate future versions of this platform. For the v0.2 release the tests covered the core components and major areas of functionality. Using the OPS GUI an initial set of queries were performed and graded into

pass/fail. The results of this testing are provided in document 1 in the Appendix. Each failure was analysed and changes were made to the system to remove the issue. Testing was also performed on components such as ChemSpider (Document 2) which provides much of the compound-structure based support inside OPS. Automated unit tests were also performed on components such as the IMS and Query Expander systems (Documents 3/4). Thus, a series of tests across the system confirmed it was operating correctly and the system could be released to project members.

3.9 Figures showing v0.2 in action

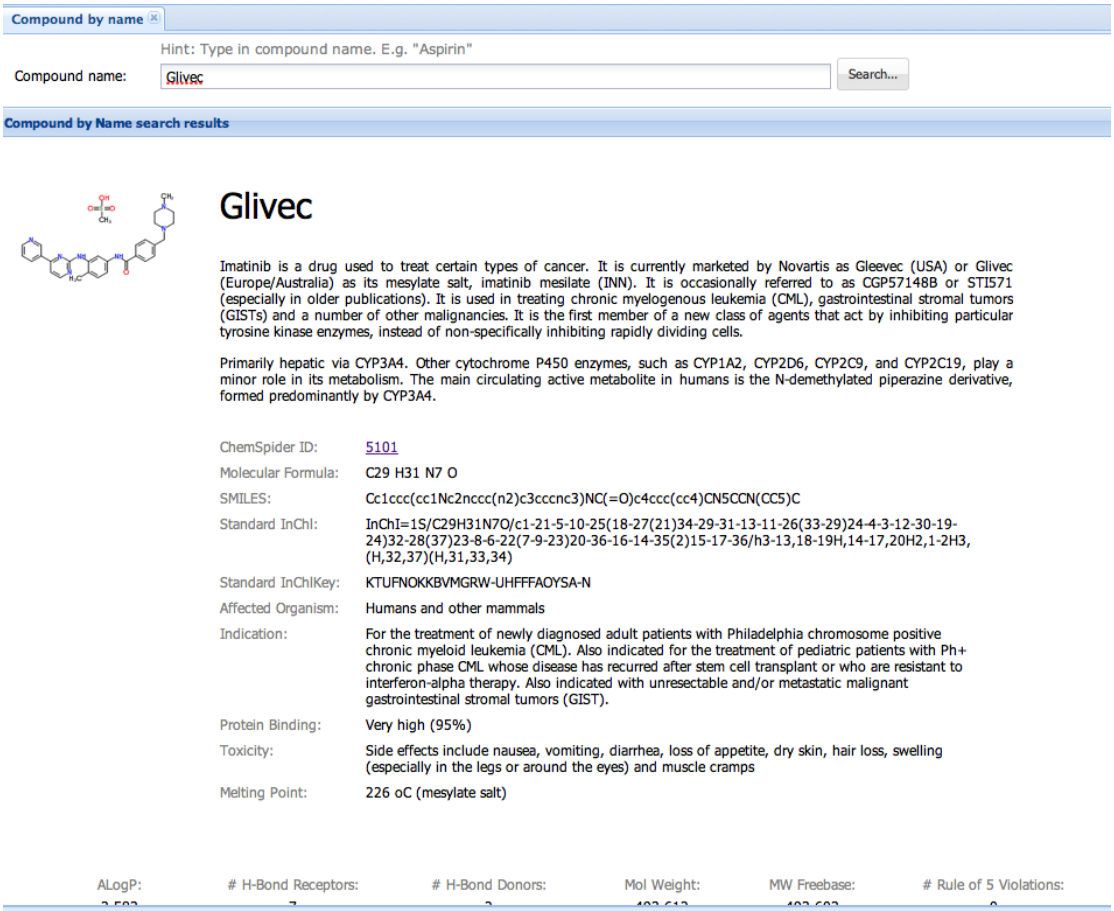


Figure 1: Compound Information. As the user types, ConceptWiki uses ChemSpider validated synonyms to map text entry to a Concept. This is then linked to a Chemspider ID which is used to retrieve data from both ChemSpider and the DrugBank database

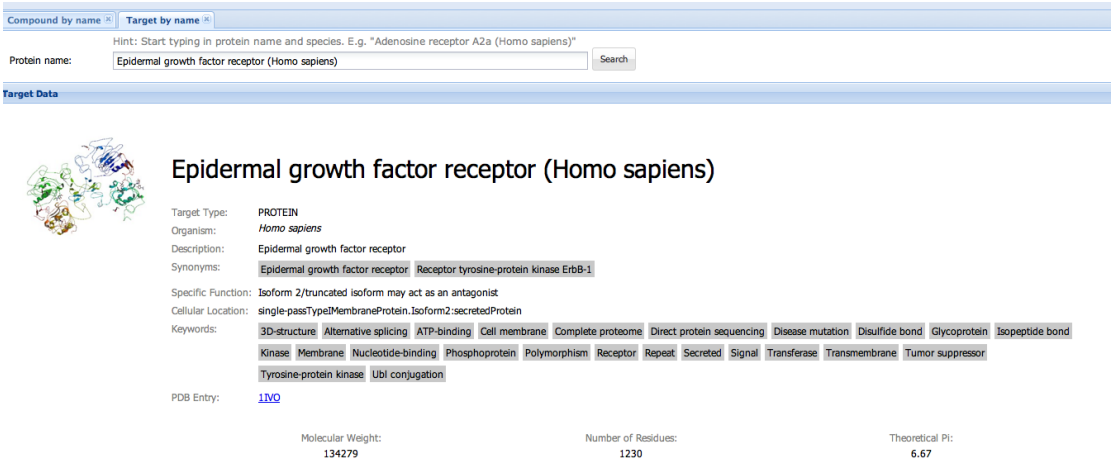


Figure 2: Target Information. As with compounds, ConceptWiki is used to translate entered text to an OPS Concept, using Uniprot as the main protein information system.

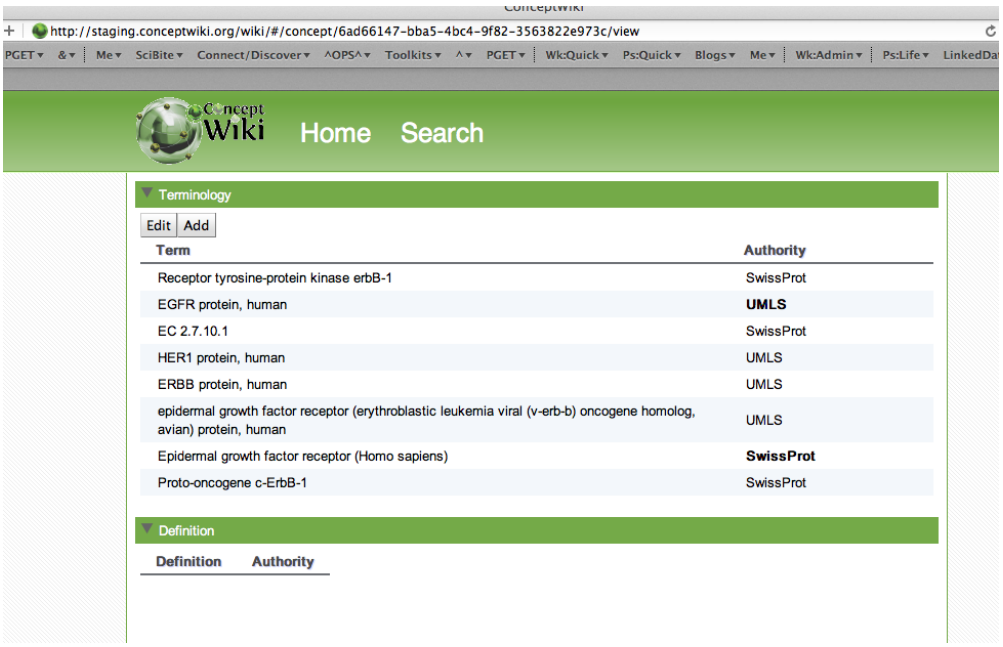


Figure 3: Compounds and targets can be community edited via ConceptWiki by clicking on the “Definition” link in the main OPS interface

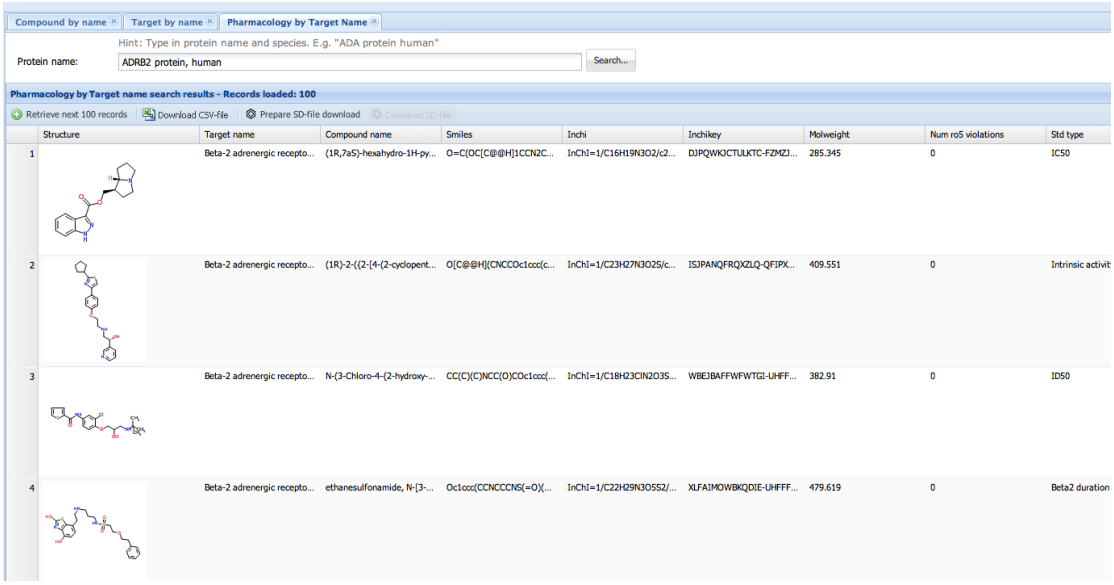


Figure 4: Pharmacology search for the B2-adrenergic receptor

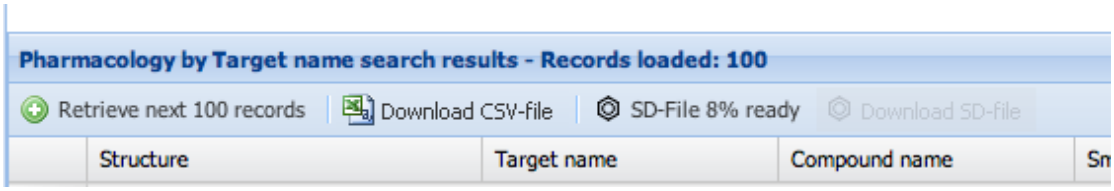
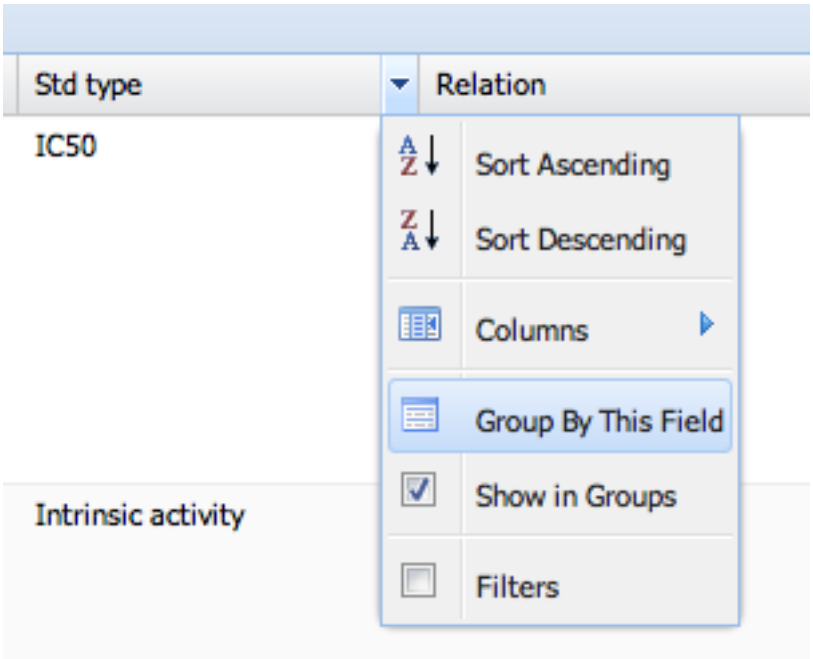


Figure 5: Pharmacology control bar. Here, SDFile export is progressing in the background and the “Download SD File” button will be activated once complete

A.



B.

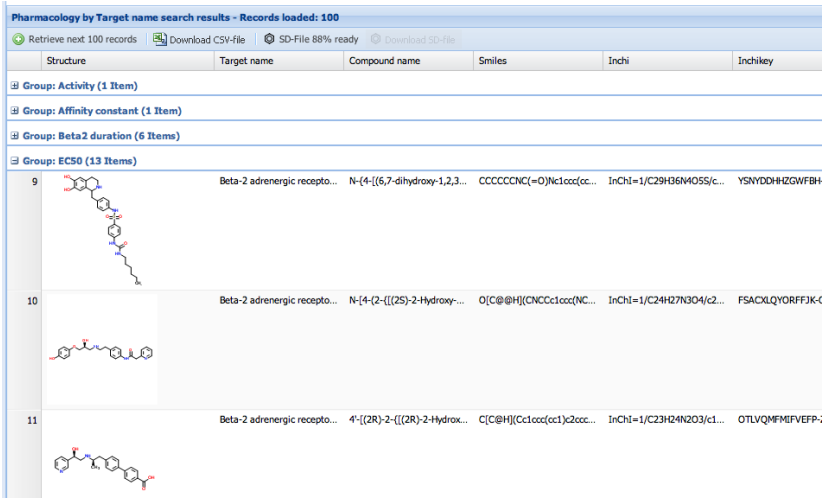


Figure 6 Grouping data records based on some column property.

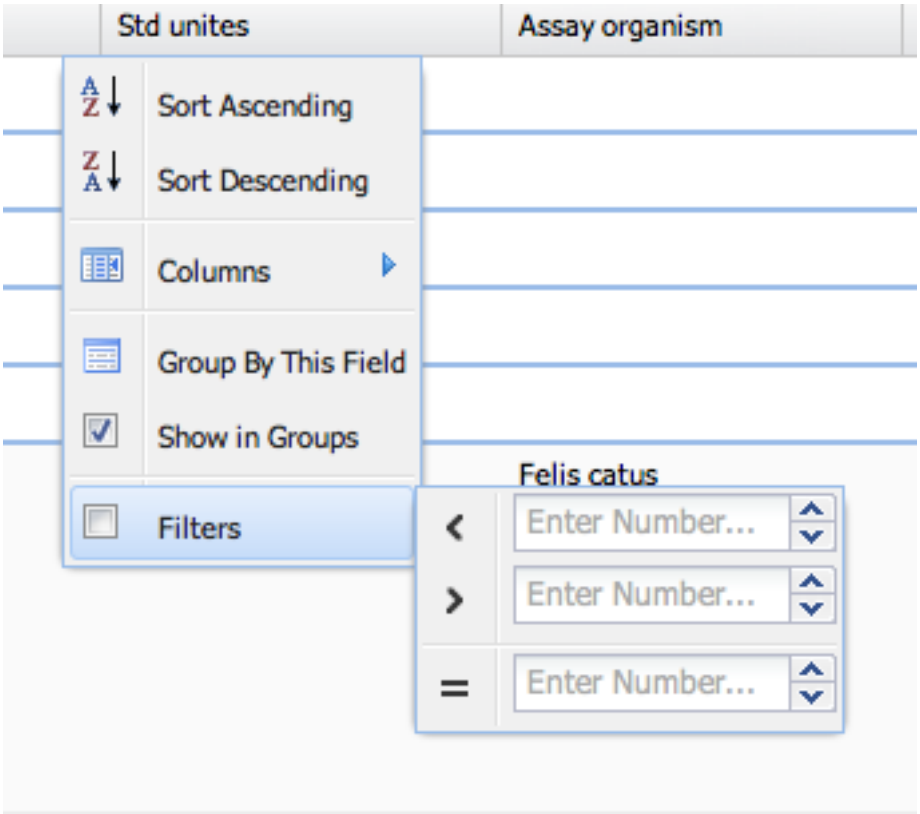


Figure 7: Filtering data based on column values

4 Associated partners enlisted and integrated

4.1 How to get involved in the project

Open systems need an engaged community, to grow, develop and sustain. We actively manage our partners, and the wider community. We term this the Open PHACTS Waiting Room, managed by a Gatekeeper (Bryn Williams-Jones)

- Our relationships with all partners are visible: what we are doing together and why
- Opportunities to engage and develop are open and are based on project needs

We hold regular community workshops and events

- Learn more about Open PHACTS and the Open Pharmacological Space
- Participate in new ideas and functions
- Engage in development of new use cases, help us answer new questions
- Contribute to development, and engage in plans for sustaining the Open Pharmacological Space

There are different degrees of involvement of an organization with the Open PHACTS project (shown in Figure 8):

- **Associated Partners:** We have a Memorandum of Understanding (MoU) ready for institutions to sign for mutual support and exchange of ideas, data or technology. Associated Partners will be the first to hear about the latest developments in the Open PHACTS project and will also have the opportunity to present ideas and use cases to the core Open PHACTS team.

- **Development Partnerships:** Once an institution is an Associated Partner and want to do some more specific development work together with us (e.g. develop APIs, new data, algorithms etc), they can enter a Development Partnership with the Open PHACTS project. This will give them greater access to the core of the project. Development Partnerships are focused on defined pieces of work of mutual interest and an agreed collaborative annex is added to the MoU.

- **Joining the consortium:** If an institution would like to become an integral part of the project, it might be worth to consider the option to join the Open PHACTS consortium.

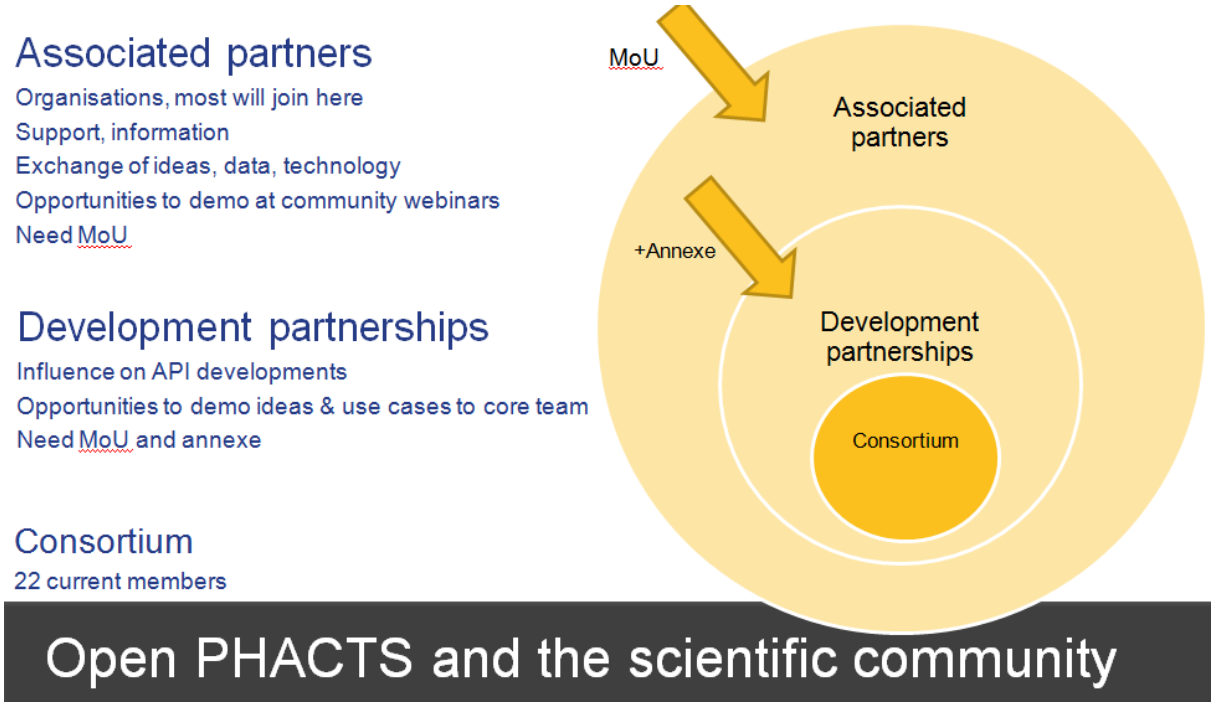


Figure 8: Different degrees of involvement of an organization with the Open PHACTS project

4.2 Process of adding Associated Partners

To become an Associated Partner a Memorandum of Understanding (MoU – document 6 in the Appendix) has to be signed. We agreed on the following process:

1. Internal sponsor proposes an Associated Partner by filling the "Associated Partner Details Form" (document 7 in the Appendix) including a short justification why it will be of benefit to the Open PHACTS project if the proposed institution becomes an Associated Partner. This will be the basis for deciding whether or not an Associated Partner proposal is accepted or not.
2. Discussion and decision in the Management Task Force (WPs 7-9).
3. Information about proposed Associated Partner will be sent out with the Friday PMU message. If there is no objection from the consortium within one week, we can sign the MoU.
4. Internal sponsor organizes the signature and logo of potential Associated Partner and sends it to the gatekeeper Bryn Williams-Jones.

4.3 List of Associated Partners

At the moment we have 4 Associated Partners enlisted:

- Aureus Sciences
- Code-N Computing
- GVK Bio
- Thomson Reuters

Further potential Associated Partners are in the process of being added at the moment.

The up-to-date list of our Associated Partners can be found on our website:
http://www.openphacts.org/index.php?option=com_content&view=article&id=64&Itemid=73

5 Appendix

Document 1: Initial GUI Analysis.

GUI Analysis.docx

Document 2: ChemSpider testing results

ChemSpider Web Services - Test results.

Document 3/4: IMS test results, see:
https://universityofvienna1.basecamphq.com/projects/6361646/file/117610170/ims-test-coveraga.tar.gz

https://github.com/openphacts/OpsPlatform/tree/master/ops-platform/larkc-plugins/plugin.imsSqarqlExpand/src/test/java/eu/ops/plugin

Document 5: LarkC core test results
For LarkC, stress testing was performed: 20 parallel threads which chose an API method at random and fired 20 requests without waiting for a response. This takes 94s to complete, and the platform responds as normal afterwards. Additional LarkC unit testing was performed and is reported in the code repository,
https://larkc.svn.sourceforge.net/svnroot/larkc/trunk/platform/src/main/test/eu/larkc/core/

Document 6: Memorandum of Understanding

OPS_MoU

Document 7: Associated Partner details form

OPS_AP details form