

Open PHACTS

Milestone 8

Version 4 of the Open PHACTS Discovery Platform released

Prepared by Nick Lynch (OPF)

Approved by OPF, NBIC, GSK, Lilly, CD, Pfizer, LUMC, USC

May 2015
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: 60 months

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

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	2 / 10

Definitions

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

- 1 - **GSK** – GlaxoSmithKline – **Coordinator**
- 2 - **UNIVIE** – Universität Wien – **Managing Entity of IMI JU funding**
- 3 - **DTU** – Technical University of Denmark
- 4 - **UHAM** – University of Hamburg, Center for Bioinformatics
- 5 - **BIT** – BioSolveIT GmbH
- 6 - **PSMAR** – Consorci Mar Parc de Salut de Barcelona
- 601 - **FIMIM** – Fundacio Institut Mar d'Investigacions Mediques
- 602 - **UPF** – Universitat Pompeu Fabra
- 7 - **LUMC** – Leiden University Medical Centre
- 8 - **RSC** – Royal Society of Chemistry
- 801 - **RSCWW** – RSC World Wide Ltd
- 9 - **VUA** – Stichting VU-VUMC
- 10 - **CNIO** – Centro Nacional de Investigaciones Oncológicas
- 11 - **UNIMAN** – University of Manchester
- 12 - **UM** – Universiteit Maastricht
- 13 - **ACK** – ACKnowledge
- 14 - **USC** – Universidade de Santiago de Compostela
- 15 - **UBO** – Rheinische Friedrich-Wilhelms-Universität Bonn
- 16 - **AZ** – AstraZeneca AB
- 17 - **Pfizer** – Pfizer Limited
- 18 - **Esteve** – Laboratorios del Dr. Esteve, S.A.
- 19 - **Novartis** – Novartis Pharma AG
- 20 - **ME** – Merck
- 21 - **HLU** – H. Lundbeck A/S
- 22 - **Lilly** – Eli Lilly and Company Limited
- 23 - **NBIC** – Stichting Netherlands Bioinformatics Centre
- 24 - **SIB** – Swiss Institute of Bioinformatics
- 25 - **CD** – ConnectedDiscovery
- 26 - **EMBL-EBI** – European Molecular Biology Laboratory
- 27 - **Janssen** – Janssen Pharmaceutica NV
- 28 - **OGI** – OpenLink Group Ltd
- 29 - **OPF** – The Open PHACTS Foundation
- 30 - **ALM** – Laboratorios Almirall S.A.
- 31 - **SciBite** – SciBite Limited

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	3 / 10

1 Introduction

Version 4 (internal name v1.5) of the Open PHACTS Discovery Platform was officially released to the community on 20 May 2015 using a variety of channels that the project has developed over time including announcements:

- on our project website¹
- on the website of the Open PHACTS Foundation²
- on Twitter
- via newsletter³

Furthermore, our Associated Partners were personally invited by email to try out the new version. We have prepared a full technical and user-focused release documentation on our [support portal](#).

The Open PHACTS Discovery Platform continues to run 24/7 and is used by Academia as well as EFPIA & other Biotechs/Pharma. The new version can be tried out at <https://dev.openphacts.org>.

We are pleased to report a number of updates to the Open PHACTS Discovery Platform since the previous report and this report will provide a summary of those updates.

The project has also generated a number of other updates on specific aspects of the platform during this reporting period including Explorer and Ontology updates.

2 Chemistry Challenges & Data Refresh

This period of the project has needed to deal with the absence of the RSC development and data team due to negotiations with IMI on the project funding.

Early on during the release we decided to decouple the release from any chemistry changes and carry on with our other plans. We needed to review some of the changes in progress to minimize the impact of not having this RSC development team.

What this has meant for the release is that the chemistry registration system (OCRS) and the database have not been updated during this release. Although we now have ChEMBL 20 in the LDC, there are no new structures and their IDs in the OCRS from the last update of the system. This impacts about 1% of the total dataset as it covers new chemistry structures added since the last ChEMBL refresh and those that have been reassigned as well.

We felt the benefit of having ChEMBL 20 data heavily outweighed any disadvantages in not having the structural data or delaying the release. With our resources focused on other aspects the project did not prioritise alternates to the OCRS at this stage and during May, the

¹ <http://www.openphacts.org/news-and-events/news-archive/2015/388-open-phacts-api-version-1-5-is-here>

² <http://www.openphactsfoundation.org/open-phacts-api-version-1-5-is-here/>

³ <http://us7.campaign-archive2.com/?u=7cb502eb5648a0323d1dbe2a1&id=4a2852dab8&e=d0eebf90cd>

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	4 / 10

RSC development team has rejoined the project to help us plan for future releases. We will be able to update on the plans during subsequent reports.

3 Maintaining the System

Since the last report on our overall platform in September 2014, there have been several additions to support the growth of users of the Open PHACTS environment. Supporting the existing user base is a key part of the services and the delivery over this period.

A [support portal](#) has been introduced and plays a key role in Open PHACTS engaging with its user community:

- Allowing us to share best practice on the platform
- Supporting the submission of questions/issues from our users and presenting answers back to the users
- Wider information on the more technical aspects of the projects
- Allowing users to create their own comments and pages

We know there is more work to do to create an effective and sustainability user community and this is part of ongoing work.

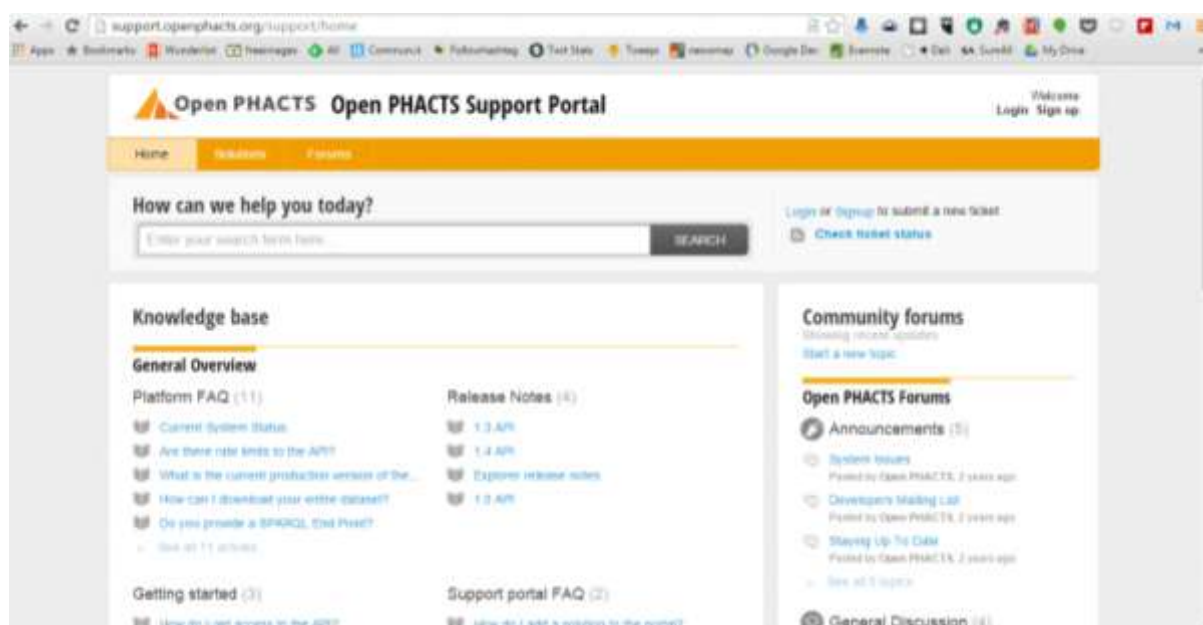


Figure 1: The Open PHACTS Support Portal

4 Data

As mentioned previously a large part of the effort for this release has been the data refresh since a period of time has passed since the last update. A careful plan was created to work through each of the data sets to investigate how much they had changed since last time and ensure that the mappings remained the same.

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	5 / 10

A key of our future strategy is to support easier and quicker refresh of the data and this will also support the SureChEMBL integration that is planned for later in 2015.

The work to develop the Virtual Machine (VM) version has benefited from the analysis for version 1.5.

The following data sources are integrated within version 1.5 of the Open PHACTS system:

Datasets	Update Status
Uniprot	(manually curated entries only) from 28 Jan 2015, release 2015_1
Chembl	from 18 Feb 2015, ChEMBL 20
DrugBank	19 Feb 2015, version 4.1
SwissProt	02 Feb 2015, release 2015_1
ENZYME	02 Feb 2015, release 2015_1
ChEBI	04 Mar 2015 ChEBI, Release 125
GeneOntology	04 Mar 2015
Gene Ontology Annotations GOA	17 Feb 2015
WikiPathways	20 Mar 2015, v20150312
ChemSpider/ACD Labs	OCRS Not updated
ConceptWiki	Not updated
DisGeNet	31 Mar 2015, v2.1.0
FDA Adverse Events (FAERS)	09 Jul 2012
NextProt (Tissues)	Not updated

4.1 Core System

As part of the release a number of new features, bugs and issues were addressed through the system including:

- Various SPARQL optimizations have been done to improve API calls and results.
- The IMS has been enriched with additional patterns for various datasets.
- Quality assurance comparing the API results to the native data sources was done to assure the same content.

4.2 Data Refresh & Volumes

For this release the following updates to the data and Entities have been made. We are not expecting the volume increases from 1.4 to make a difference to performance but we will be monitoring this over the next period.

Our ENSO plans for SureChem will place larger impact on volume of structures so this will be part of the planning and analysis for the autumn release and design.

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
	IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0 6 / 10

Dataset	Triples	Predicates
http://purl.uniprot.org	979928936	122
http://www.openphacts.org/goa	879448347	15
http://www.ebi.ac.uk/chembl	445732880	146
http://www.nextprot.org	249403405	24
http://rdf.imim.es	15011136	77
http://aers.data2semantics.org/	13557070	40
http://www.wikipathways.org	6110015	68
http://www.conceptwiki.org	4331760	4
http://www.openphacts.org/bio2rdf/drugbank	4028767	104
http://www.geneontology.org	1366494	45
http://www.ebi.ac.uk/chebi	1012056	22
http://purl.uniprot.org/enzyme	61467	13
http://www.nextprot.org/caloha	14552	24

Table 1

Entities	Number
Compound	1565746
Target	547357
Pathway	1448
Diseas	8162
Tissue	817

Table 2

Identifier mapping service:

- 31,382,458 Mappings
- 186 Mapping Sets
- 50 Source Data Sources
- 4 Predicates
- 50 Target Data Sources

5 Open PHACTS API & Access Points

The last 6 months, since the release of 1.4 have seen a number of changes and updates to the API.

As part of our plans for the 1.5 release we decided that version 1.3 needed to be decommissioned in March 2015 to allow for transition to 1.5 and remove this deprecated API. We planned this decommission process and publicity by analyzing the remaining users.

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	7 / 10

- Explorer 2.0 was released to highlight the 1.4 API
- Explorer 1.0 was then decommissioned since it was still using API 1.3
 - An alert was placed on the Explorer 1.0 tool to alert users to the changes
- Communication including newsletter and direct emails to those systems using 1.3 were sent out during February and March
- Further reminders were sent out prior to the final switch off at end of March 2015

As with any API, our desire is to help users get the most value from the API and we will continue to support those Apps in moving from an older version of the API. We do see a lag in the adoption of API and we will have more data on this during 2015 in terms of the spread of usage of the APIs.

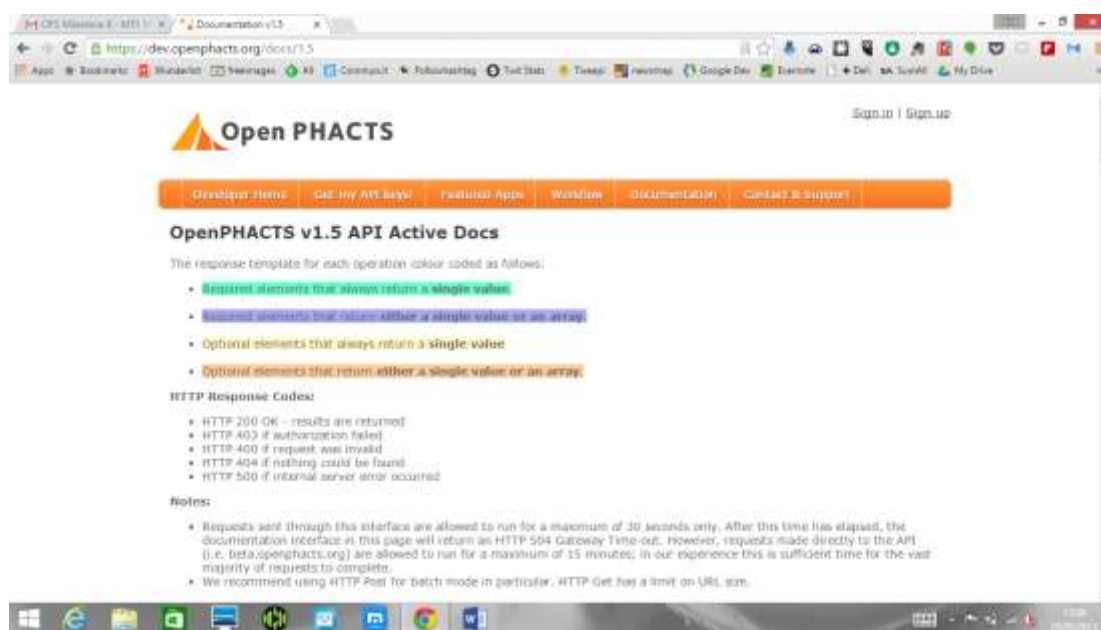


Figure 2 API 1.5 Documentation

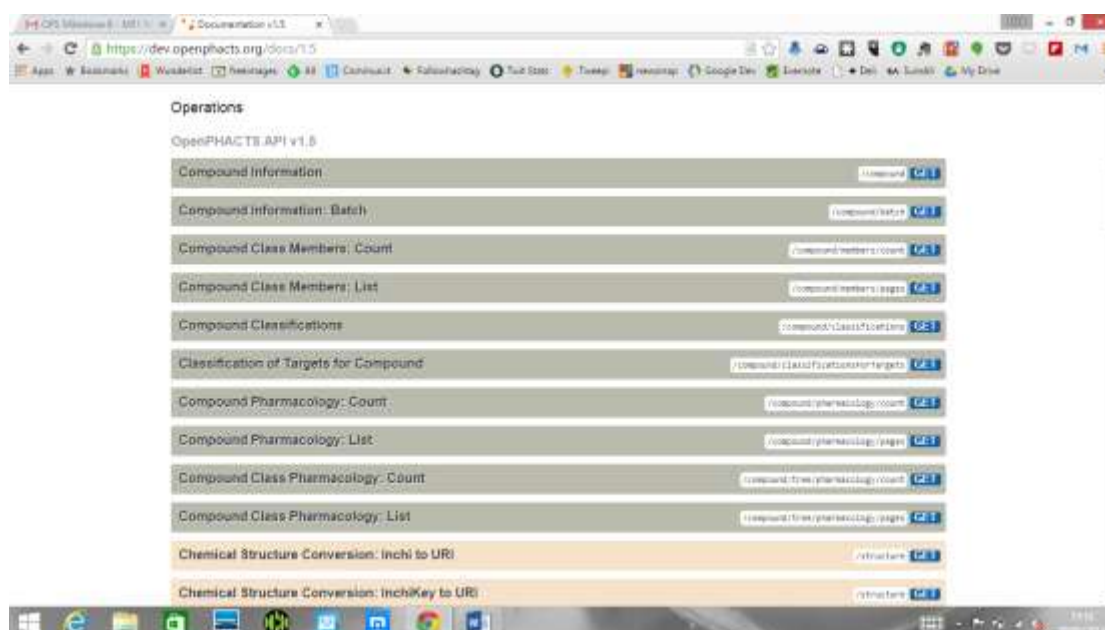


Figure 3 Further API details

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	8 / 10

6 The Open PHACTS Explorer

As mentioned in other reports and above, Open PHACTS released a new version of the Explorer in February 2015 to replace Explorer 1.0.

This version was a completely new design since the previous versions and we expect this platform to evolve, due to its flexible design, for the remainder of the project rather than the need to be re-designed again. We also hope to offer the components that Explorer is now based on to the wider community through the Bio.js community.

Equally the use of Explorer is only one part of the ways that users interact with Open PHACTS data and we expect the project to continue to encourage other exploitation tools to use the API and to develop new visualization tools that support new business questions.

7 Workflow Access

A key part of scientific informatics is the use of workflow tools such as Knime, Pipeline Pilot and also Taverna to allow users routes of chaining a series of processes together to support the answering of key questions. A very successful workshop was run at the end of May to support the knowledge sharing and best practice around workflows and we thank the contributors and trainers who made this possible.

Recent work on the Swagger documentation in Open PHACTS released in 1.5 has enabled easier generation of the workflow nodes in Knime and Pipeline Pilot.



Figure 4 Example nodes for the Pipeline Pilot

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	9 / 10

8 App Ecosystem

A key theme of the project from the beginning has been the enablement of groups to use the API to then build their own Apps and workflows on top of Open PHACTS data and expertise. This is seen in the range of Apps that currently use the API & data and a snapshot of some of the current tools are listed below.

We are continuing to work with these groups to understand their business questions to see how that develops in future API calls or ways of delivering the data.

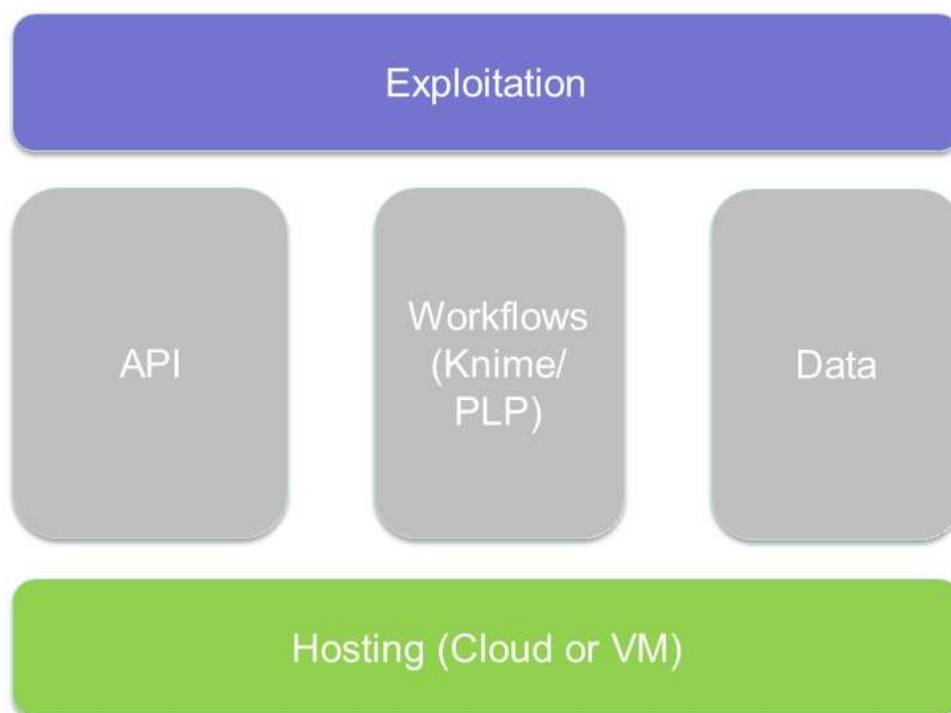


Figure 5

The figure displays eight application cards arranged in a 2x4 grid. Each card includes a logo, a brief description of the application's function, and a button to explore the app.

- Open PHACTS:** Browse and search the data within the Open PHACTS Discovery Platform. Developed by the University of Manchester and University of Vienna. Button: Open PHACTS Explorer >
- SciBites:** Connects the latest news and events in Pharma and Biotech directly to Open PHACTS pharmacology data. Developed by SciBite Limited. Button: SciBites >
- GARField:** Intuitively predicts target pharmacology based on the Similar Ensemble Approach. Developed by the Technical University of Denmark. Button: GARField >
- collector:** Extracts data to build QSAR predictive models with data from the eTOX project. Developed by PSMAR as part of the eTOX project. Button: collector >
- ChemBioNavigator:** Visualise the chemical and biological space of a molecule group. Developed by the University of Hamburg and BioSolveIT GmbH. Button: ChemBioNavigator >
- PHARMATREK:** Navigate pharmacological space in a flexible and interactive way. Developed by the Consorci Mar Parc de Salut de Barcelona (PSMAR). Button: PharmaTrek >
- Target Dossier:** Interconnecting Open PHACTS data with multiple target centric services. Developed by the Spanish National Cancer Research Centre (CNIO). Button: Target Dossier >
- utopia:** Allows the semantic enrichment of scientific articles in PDF format. Developed by the University of Manchester. Button: Utopia >

Figure 6

Open PHACTS	Milestone: Version 4 of the Open PHACTS Discovery Platform released	Milestone 8	
IMI - 115191	Author: Nick Lynch (OPF)	Version: 1.0	10 / 10

9 The Future

We are already planning the next release of Open PHACTS for later in the summer 2015 and the key aspects of this will be the chemistry refresh from the RSC and also further updates to the IMS Linksets.

This current release which focused significantly on a complete data refresh after a period without a release has allowed us to re-engage with our community and also to plan for remainder of ENSO and into 2016.

This release of the platform continues to provide a stable base upon which to build a cutting edge system that will address real issues in drug discovery today.