

Annex I

Description of Work

Acronym: **Open PHACTS**

Full title:

An open, integrated and sustainable chemistry, biology and pharmacology knowledge resource for drug discovery

IMI Call topic: 2009 - 2, Open Pharmacological Space - OPS

Name of the coordinating person: **Stefan Senger (GSK)**

List of participants:

| Participant No. | Participant organisation name | Participant organisation short name |
|---------------------------------------|---|--|
| 1 (Coordinator) | GlaxoSmithKline Research and Development Ltd | GSK |
| 2 (Managing entity of IMI JU funding) | Universität Wien | UNIVIE |
| 3 | Technical University of Denmark | DTU |
| 4 | Universität Hamburg, Center for Bioinformatics | UHAM |
| 5 | BioSolveIT GmbH | BIT |
| 6 | Consorti Mar Parc de Salut de Barcelona | PSMAR |
| 601 | Fundacio Institut Mar d'Investigacions Mediques | FIMIM |
| 602 | Universitat Pompeu Fabra | UPF |
| 7 | Leiden University Medical Centre | LUMC |
| 8 | Royal Society of Chemistry | RSC |
| 801 | RSC World Wide Ltd | RSCWW |
| 9 | Stichting VU-VUMC | VUA |
| 10 | Centro Nacional de Investigaciones Oncológicas | CNIO |
| 11 | University of Manchester | UNIMAN |
| 12 | Universiteit Maastricht | UM |
| 13 | ACKnowledge (terminated on 17.09.2014) | ACKnowledge |
| 14 | Universidade de Santiago de Compostela | USC |
| 15 | Rheinische Friedrich-Wilhelms-Universität Bonn | UBO |
| 16 | AstraZeneca AB | AZ |
| 17 | Pfizer Limited | Pfizer |
| 18 | Laboratorios del Dr. Esteve, S.A. | Esteve |
| 19 | Novartis Pharma AG | Novartis |
| 20 | Merck | ME |
| 21 | H. Lundbeck A/S | HLU |
| 22 | Eli Lilly and Company Limited | Lilly |
| 23 | Stichting Netherlands Bioinformatics Centre | NBIC |
| 24 | Swiss Institute of Bioinformatics | SIB |
| 25 | Connected Discovery Ltd | ConnDisc |
| 26 | European Molecular Biology Laboratory | EMBL-EBI |
| 27 | Janssen Pharmaceutica NV | Janssen |
| 28 | OpenLink Group Ltd | OGL |
| 29 | Open PHACTS Foundation | OPF |
| 30 | Laboratorios Almirall S.A. | ALM |
| 31 | SciBite Limited | SciBite |

Table of Contents

| | |
|---|---|
| Table of Contents | 3 |
| 1. Executive summary | 4 |
| 2: Scientific case | 5 |
| 2.1 Concept and objectives | 6 |
| 2.1.1 Rationale..... | 6 |
| 2.1.2 Success factors | 6 |
| 2.1.3 General approach and objectives..... | 9 |
| 2.2 Progress beyond the state-of-the-art..... | 9 |
| 2.3 Potential impact of project results | 10 |
| 2.4 Intellectual Property Principles in the Open Pharmacological Space | 11 |
| 3. Project plan | 13 |
| 3.1. Overall project plan..... | 13 |
| 3.2 Timing chart..... | 19 |
| 3.3 Work description | 20 |
| Table 3.3 a: Work package list | 21 |
| Table 3.3 b: Work package description..... | 22 |
| Table 3.3 c: Deliverables list..... | Fehler! Textmarke nicht definiert. |
| Table 3.3 d: Staff efforts..... | Fehler! Textmarke nicht definiert. |
| Table 3.3 e: List of milestones | Fehler! Textmarke nicht definiert. |
| 4. Implementation | 75 |
| 4.1 Governance structure and management procedures..... | 75 |
| 4.2 Individual participants | Fehler! Textmarke nicht definiert. |
| 4.3 Consortium as a whole | 80 |
| 4.4 Resource allocation and budget | Fehler! Textmarke nicht definiert. |
| 4.4.1: Description of the use of resources..... | Fehler! Textmarke nicht definiert. |
| 5. Ethical Issues | 83 |
| 5.1 Ethical issues table..... | 84 |
| Annex 1 – Challenges in building Semantically integrated concepts for drug discovery | Fehler! Textmarke nicht definiert. |
| Annex 2 – Memoranda of Understanding with Consortia from call topic 7 and call topic 9 . | Fehler! Textmarke nicht definiert. |
| A 3.2 Draft for Memorandum of Understanding with Associated Partners . | Fehler! Textmarke nicht definiert. |
| Addendum – No 1 | Fehler! Textmarke nicht definiert. |
| Addendum – No 2 | Fehler! Textmarke nicht definiert. |

1. Executive summary

Drug discovery is data-hungry and all major pharmaceutical companies maintain extensive in-house instances of public biomedical and chemical data alongside internal data. Analysis and hypothesis generation for drug-discovery projects requires careful assembly, overlay and comparison of data from many sources, requiring shared identifiers and common semantics. For example, expression profiles need to be overlaid with gene and pathway identifiers and reports on compounds *in vitro* and *in vivo* pharmacology. Utility of data-driven research goes from virtual screening, HTS analysis, via target fishing and secondary pharmacology to biomarker identification. Alignment and integration of internal and public data and information sources is a significant effort and the process is repeated across companies, institutes and academic laboratories. This represents a significant waste and an opportunity cost.

To address these challenges, the Open PHACTS project will develop an open source, open standards and open access innovation platform, Open Pharmacological Space (OPS), via a semantic web approach. OPS will comprise data, vocabularies and infrastructure needed to accelerate drug-oriented research. This semantic integration hub will address key bottlenecks in small molecule drug discovery: disparate information sources, lack of standards and shared concept identifiers, guided by well defined research questions assembled from participating drug discovery teams.

Workflows for data capture, processing, interoperability, visualization, and chemogenomics will be developed to create a comprehensive Systems Chemical Biology Analysis Network. Security issues around proprietary data, shared via the Open PHACTS Discovery Platform and accessible for safe querying and reasoning will be properly addressed with expert trusted parties.

The Open PHACTS consortium comprises 20 European core academic and SME partners, with leading experts in the fields of data mining, annotation, small molecule data storage and manipulation, target related bioinformatics, RDF-type information handling, massive *in silico* reasoning and chemical biology. The 10 EFPIA members of Open PHACTS will contribute data sets, internal tools, expertise in drug discovery and software engineering as well as programming capacity to the project.

2: Scientific case

In fact, there is not much of a *specific scientific case* to make for this project. Rather, we argue that this project is a *prerequisite to enable effective scientific discovery* in the pharmacological space for decades to come.

The proposed Open PHACTS Discovery Platform will be a **widely accessible, open innovation platform** that will serve EFPIA partners as well as biotech and academic drug-discovery for one of their core business activities: knowledge discovery and verification. In that sense it also serves other IMI projects and is an enabling project throughout the pharmaceutical industry and beyond. We actually argue at several points in this proposal that the project should be viewed and managed much more as an **enabling technology project** than as a classical research project. This view has repercussions for budget allocation to technical governance beyond project management and infrastructure, i.e. by appointing a 'Central' Chief Technical Officer (CTO).

The Open PHACTS project will develop largely **Open Source** and **Open Data services**, but will also allow for **secure querying and reasoning environments** and the 'plug-in' of proprietary data sources and analysis services. Open PHACTS will provide a comprehensive framework of chemical, biological and pharmacological information, confederated from vast, distributed and variable data and information sources. This semantically enriched and **fully interoperable platform** will deliver information on small molecules and their pharmacological profiles, including pharmacokinetics and ADMET data as well as on biological targets, pathways, and diseases. Practicing drug-discovery scientists in both the public and the private sector will turn to Open PHACTS services to:

- integrate data on target expression, biological pathways, pharmacology, and diseases 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 programmes
- compare molecular interaction profiles to assess potential off-target effects and safety pharmacology
- analyze chemical motifs against biological effects to deconvolute high content biology assays

Thus, the core success-measure for the Open PHACTS Discovery Platform is that EFPIA and academic users embrace the framework as an indispensable semantic web environment for their day-to-day knowledge discovery business.

In the spirit of **Open Innovation** we have also defined the concept of **Open PHACTS Associated Partners**. These are partners that are not or cannot be funded via the IMI scheme, but have essential expertise, content or tools to augment what the core Open PHACTS consortium partners can deliver. The actual list of 'associated partners' is listed on our web site www.openphacts.org and many of them actively collaborate in different contexts with Open PHACTS core project partners.

2.1 Concept and objectives

Open PHACTS Vision:

Create an open knowledge infrastructure enabling semantic integration of chemical and biological data to support drug discovery

2.1.1 Rationale

The Open PHACTS proposal effectively addresses the central problem statement in the KM call topic 8 (IMI_Call_2009_8, Open Pharmacological Space-OPS) that it is currently *'difficult to effectively utilize public domain resources for the support of drug discovery research (in industry or academia)'*

The consortium proposes a **highly innovative, semantic web enabled approach** to the 'development of a set of open access standardized tools that could enhance existing databases to allow comprehensive integration of information on small molecules and their targets'.

The final aim is to 'improve the accessibility of public domain drug discovery resources by interoperability with existing public domain systems'.

The rationale for the Open PHACTS semantic web approach is that classical data warehousing methods are no longer scalable to the size, spread and complexity of life-science data-sets, information resources and data analysis. The size and complexity of this project to deliver an **integrated framework on top of federated resources** has led us to organize the work into a set of specific work packages that essentially deliver the specific building blocks of the Open PHACTS Discovery Platform. While each of these building-blocks will have lasting value on their own, the delivery of a **coherent Open PHACTS framework** is ensured by adopting a working-to-working method of iterative, full-project, vertical builds to **deliver answers to increasingly complex research questions**.

Recognizing that the power of standards lies in their widespread adoption the core framework is built on the principles of Open Source, Open Access and Open Data. We firmly believe that the only **long-term sustainable model** for a scientific system of this nature is full openness around the core semantic components, vocabularies and interfaces. However, this does not preclude the delivery of **proprietary content** through this semantic integration hub, nor does it preclude value-added closed-source or **commercial services** delivered on top of this system.

We have chosen to use a federated approach to implement the Open PHACTS Discovery Platform. We recognize that this approach brings with it a number of specific challenges, which we address in specific Work Packages. This makes the Open PHACTS proposal quite different in style compared to more 'vertical' IMI projects.

2.1.2 Success factors

The key success factor will be that both pharmaceutical partners and academia will be able to use the Open PHACTS Discovery Platform to **enhance their knowledge discovery process**.

More specifically:

- The ability of the platform to provide **answers to drug discovery related questions** and to be an integrated part of the drug-discovery workflow.
- **Measurable use of the system** beyond the original Open PHACRS partners, not only for direct knowledge discovery but also as a framework to build and deliver a wide range of services.
- The Open PHACTS Discovery Platform should grow to acceptable levels of **quality, performance, usability and completeness** and should be easily accessible and extendable within the scope of the project as a reliable enterprise system.

- Early **community adoption through the engagement of 'associated partners'** within and outside the European Union, among which are **major data and infrastructure providers** such as the National Centre for Biomedical Ontology (NCBO in the USA) and networked initiatives such as Sage Bionetworks and the World Wide Web Consortium (W3C).
- **Commercial information supply chain companies** deliver data through and build services on top of the Open PHACTS Discovery Platform.

Glossary of terms and definitions

| Term | Definition |
|----------------------------|--|
| Open PHACTS | The project and the consortium proposing it |
| OPS | Open Pharmacological Space, the semantic integration hub delivered by the Open PHACTS project |
| Open PHACTS Commons | The content that is made interoperable in Open PHACTS |
| Open PHACTS Infrastructure | The software and services to manage the Open PHACTS Commons |
| Core partners | Partners (academic) funded through IMI or contributing (EFPIA) |
| Associated Partners | (AP) partners that work closely with Open PHACTS |
| Concept | A 'unit of thought' or reference (principally unambiguous) |
| Symbol | Terms, Identifiers, URI's or any other token referring to a concept |
| UUID | Universally Unique Identifier (opaque stable symbol for concept) |
| Vocabulary | A simple collection of symbols referring to concepts |
| Thesaurus | A simple, mostly linguistic, hierarchical organized vocabulary |
| Ontology | A formal, explicit representation of knowledge by a set of concepts |
| IRS | Identity Resolution Service, cross mapping symbols to concepts |
| RDF | Resource Description Framework: standard model for information exchange on the Web |
| Triple | A RDF statement that essentially comprises an 'Assertion' |
| OWL | An RDF-mapped language for representing ontologies. |
| SKOS | An RDF-mapped language for representing thesauri |
| S+P+O | Most assertions have the basis form Subject+Predicate+Object |
| Provenance | Assertions need to be traceable and citable, by provenance metadata |
| Nanopublication | An assertion captured in RDF + its provenance |
| Evidence Factor | A computed value based on status and repetitive finding of nanopublications with the same basic S+P+O |
| Cardinal assertion | A 'collapsed' form of a unique S+P+O+ Evidence Factor |
| Reasoning | Inferencing over many cardinal assertions to discover knowledge |
| Concept Web | The full collection of loosely interconnected nanopublications |
| Agile development | A proven software engineering method with short development and feed back cycles guided by business questions or use-cases |
| 4+1 | A specific form of application/user driven software development |
| SPARQL | query language for RDF, |
| REST | REpresentational State Transfer (REST), is a key design idiom for web enabled systems |
| RESTful | RESTful applications maximize the use of the pre-existing, well-defined interface and other built-in capabilities |

Open PHACTS terminology:

- **Component 1: Open PHACTS Discovery Platform**
This is the ‘core’ of the infrastructure that is accessible via the *Open PHACTS API* (component 2).
- **Component 2: Open PHACTS API**
API that is used by applications to ‘communicate’ with the *Open PHACTS Discovery Platform*.
- **Component 3: Open PHACTS Explorer**
The web-based GUI that has two main purposes:
a) to provide basic query features for users. More ‘elaborate query and visualization needs’ will be satisfied by *Applications (Apps)* (in due course),
b) to provide a means to expose/surface some of the ‘clever things’ that the *Open PHACTS Discovery Platform* allows users to do.
- **Component 4: Example Applications (eApps)**
Applications are a key element of the Open PHACTS infrastructure since they will hopefully allow users to take full advantage of the capabilities of the *Open PHACTS Discovery Platform* in a variety of ways. To demonstrate this *eApps* are being built by members of the Open PHACTS consortium.
- **Component 5: Applications (Apps)**
Applications that third parties will be able to built using the *Open PHACTS API*.

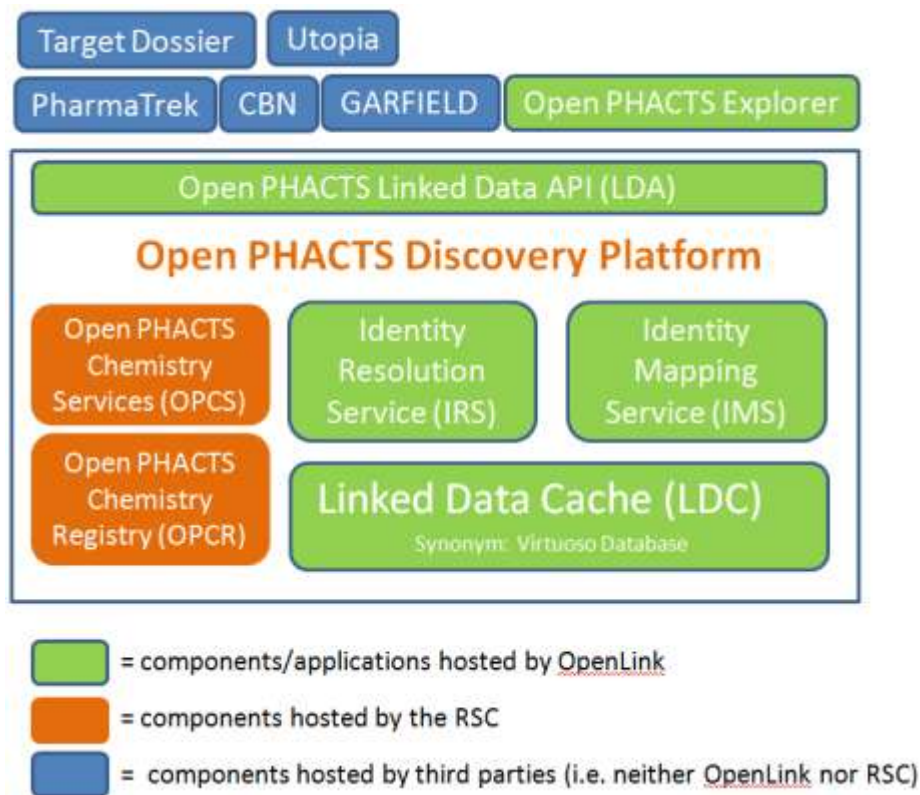


Figure 1: The Open PHACTS Discovery Ecosystem

2.1.3 General approach and objectives

Open PHACTS aims to **integrate a relevant and continuously expanding subset of distributed heterogeneous data sources** into one 'virtual resource' via the creation of a Semantic Interoperability layer. This Interoperability backbone will form the core infrastructure of the Open PHACTS Discovery Platform, comprising (*inter alia*):

- Dynamic updates to track the expanding set of relevant target and drug data sources
- Access/link to internationally accepted applied vocabularies, ontologies, standards and formatting tools.
- An open graphical query interface with orthogonal navigation between data types
- An accessible RDF-triple store (with SPARQL-endpoint) providing the Open PHACTS Discovery Platform as “linked data”
- Open RESTful web service interfaces to all data
- Supported and documented APIs backed by developer training options
- Hosting of Open PHACTS example applications and services
- A validated security model for querying with proprietary data
- Close to 24/7 operability and proper SLA and back up strategies
- Long term sustainability aligned with common EU infrastructure via ELIXIR

2.2 Progress beyond the state-of-the-art

In order to succeed and **gain widespread community adoption**, the Open PHACTS Discovery Platform will need to go beyond the current state of the art of tools in this domain. The transition from classical, hypothesis driven research towards systems approaches where high throughput data from genomics, transcriptomics, proteomics and metabolomics are combined with High Throughput Screening (HTS) toxicology and pharmacological data requires **rigorous new methodology**.

One key challenge is that **current data sources are largely incompatible** with massive computational approaches and the vast majority of drug-discovery sources cannot easily interoperate. Recently developed data and text mining approaches, improved data capture standards, and leveraging **semantic web technology** open a **first-time-opportunity to achieve interoperability** through the semantic harmonization of data in key data sources *a posteriori*. A large and influential consortium involving academic as well as industrial drug-discovery partners collaborating on Open PHACTS is likely to increasingly drive researchers around the globe to capture and distribute data and information in a semantically interoperable and computer readable format, as their data will 'connect' and 'mean' more from the onset. **A priori data interoperability** 'at the source' is therefore a desired long-term effect of our distributed approach.

In the general field of molecular biology, the 2014 Nucleic Acids Research Database Issue includes descriptions of 58 new molecular biology databases and recent updates to 123 databases previously featured in NAR or other journals. Studies in the scope of ELIXIR have shown that out of 531 databases surveyed, 63 were either not online anymore or had not been updated since 2005 and, for a further 78, the update status was unclear. This shows that **database quality is a serious issue** to address in the context of Open PHACTS. More importantly, less than 10% of the biomolecular resources surveyed indicated that they had multi-annual funding secured. The data resource landscape is therefore very fragile and **Open PHACTS can play an important role** in capturing the most important assertional content globally in a **stable, interoperable and sustainable format**.

The Open PHACTS knowledge creation and management concept includes the following main **innovative approaches**:

- Free-text, table, image, molecular sequence and structured information will be extracted and encoded in Resource Description Framework (RDF) assertions enriched with provenance data to form the basic building block of interoperability in the Open PHACTS Discovery Platform: nanopublications.
- These assertions will be mined and constructed in the form of three Universally Unique Identifiers (UUID's) using high performance concept mapping engines (e.g., ProMiner, Peregrine), terminology rewriting rules, and lexical and contextual statistics for term disambiguation.
- Disambiguation and quality control of captured assertions will be facilitated by an Identity Resolution Server (IRS) using, for example, the ConceptWiki (www.conceptwiki.org) equipped with synonym mapping via, for example, BridgeDB
- In order to reduce redundancy of identical assertions mined from all included resources, 'cardinal assertions' will be created with full consideration of the repetition-value (citation, curation) of the multiple supporting nanopublications, with an open calculation of 'evidence level' based on provenance data of these individual nanopublications.

However, the primary objective of Open PHACTS is not to develop novel technology. The ultimate goal is to establish the Open Pharmacological Space as an integration hub with pervasive impact throughout public and private drug discovery research. Hence, strong emphasis will be placed on the **quality control** of data, vocabularies and assertions to ensure the high level of data quality necessary to fuel real-world applications such as QSAR analyses, generation of pharmacological profiles, and prediction of toxicities via collaboration with eTOX. In contrast to many existing free resources in the drug discovery domain, Open PHACTS will provide **high quality curated data**, with the option for community annotations. For instance, RSC-ChemSpider is a crowdsourcing platform for the deposition and curation of community data, which presently supports structures, alphanumeric text, analytical data, images and multimedia. The **ChemSpider database** currently contains almost 30 million unique chemical structures linked out to over 500 data sources including government databases, other public compound databases, as well as **repositories** from **chemical vendors**, **patents** and **publications**.

Innovative query and visualization tools for the linked chemo-biological space will be facilitated and demonstrated by developing **web-based environments** for representation and annotation of **textual information** (e.g. assays, pharmacology, mode-of-action), **biological database information** (sequence, pathways, dose-response data), and **chemical information** (including chemical structure queries). Furthermore, the project will research suitable models for remote linking and querying with proprietary data.

The Open PHACTS project will be one of the first international attempts to create a **reliable and scalable system**, a common product beyond collective prototyping. Open PHACTS aims to deliver sustainable, reliable web based environment through proven agile software engineering models. Hence, a significant part of the funding will be reserved for **professional software engineering** and to support the progressive **integration of existing resources from organisations not belonging to this project consortium**. Thus, Open PHACTS will be an important step in the direction of full support for next generation drug discovery needs.

2.3 Potential impact of project results

A successful Open PHACTS framework would spark a revolution in data sharing and collective intelligence for biomedical and drug-discovery in the public and private sectors. Open PHACTS would also **effectively drive standards** from a practical and user driven perspective, which would in turn

enable new data to be deposited, connected and interpreted with a much higher grade of effectiveness by all parties concerned.

The 'Open Innovation' principles of this project are crucial to **enable the exploitation of the widest possible collections of data** in a phase of our biological understanding of complex health and disease questions that could potentially make all known biological systems a potential target for pharmacological intervention.

Typical **Open Source** (viral spread) and **Open Access/Open Data** effects will ensure optimal dissemination of the power of Open PHACTS into the scientific and health community at large.

This project will therefore have a major **impact on other IMI projects** and far beyond as an enabling technology and content platform that will accelerate knowledge discovery. Open PHACTS, although focused on drug discovery related questions itself, will be **expanded by associated partners** to other fields of biology and biochemistry. New insights gained in these other fields will in turn be invaluable for deeper insights in biological systems for intervention and thus also for generation of new lead compounds.

We therefore emphasize that - if successful and sustainable - this project is crucial to the IMI core mission and is likely to significantly contribute to more successful and cost-effective development of drugs and vaccines in human and animal health, as well as in nutrition and personal genomics.

Although this project does not aim at a specific disease, many patient organisations, such as those federated in our associated partner organisations Orphanet, EURORDIS and NORD will contribute knowledge, annotation and community building efforts to Open PHACTS.

Turning data into knowledge is the cornerstone of successful drug discovery but, as the previous unstable public resources that drove the Pharmaceutical Industry to develop proprietary infrastructure have matured, at the same time the industry recognises that data issues facing us today are largely pre-competitive. A future driven by the open sharing of data, tools, services and workflows benefits the whole scientific community. Hence, the industrial partners in Open PHACTS have formulated the following core industry expectations – the industry business-case:

- The development and implementation of core entity **standards** in each domain will allow content integration to enable novel approaches and *in silico* research into previously unreachable areas
- Core **data query services** of high quality with stable performance and secure access will allow companies to shut down internal systems.
- The creation of a **vibrant drug discovery community** with significant exchange of ideas and experience across industry and public research will benefit pre-competitive activities such as the identification and validation of novel targets, understanding cell signalling and regulatory systems, research into novel high content assay formats and methods to address previously intractable target-classes such as protein-protein interactions.

2.4 Intellectual Property Principles in the Open Pharmacological Space

Intellectual Property (IP) is the principle currency of drug discovery. However, owned and derived IP does not automatically lead to the creation of commercial value. This truism is one of the founding principles of Open Innovation and a key concept informing the Open PHACTS strategy. Open innovation assumes a flexible business model in which new product innovation originates from both internal and external ideas. As they are applied in the Open PHACTS project, the principles of Open Innovation are intended to support the development of IP by innovative use of data rather than ownership of data. In order to achieve this, the Open PHACTS project seeks to construct an open and unrestrictive operating environment for the user community, to enable **Open PHACTS users** to make innovative use of data.

The Open PHACTS project will be executed fully within the framework of the IMI IP policy (see full details at http://imi.europa.eu/intellectual-property_en.html). This policy defines IP in three broad categories: **Background IP** (IP that exists before the project), **Foreground IP** (IP that is generated in the course of the project) and **Sideground IP** (Intellectual property created in the duration of the project, but which is not considered as part of the project objectives). The implications of each of these forms of IP for the Open PHACTS project, is considered below:

Background IP

Open PHACTS aims to integrate and make accessible a broad range of biological and chemical data (Background data). In all cases the pre-existing, Background IP and other legal restrictions around background data will remain unaltered. Open PHACTS will simply provide access to data on pre-existing terms. Where data is specifically included in the project, all background IP and legal restrictions shall be identified in the Project Agreement. Each Participant shall remain the exclusive owner of its Background IP. EFPIA members have agreed that data, software, tools, algorithms contributed to the project will be provided either on unlimited license or IP rights will be waived completely.

The data, software, tools, algorithm contributions will be defined by the needs identified in WP6 business questions. The EFPIA members are committed to the release of data around standard drug discovery concepts (e.g. CACO2 flux, standard profiling, high content biology, permeability, PK/PD) and these will be provided in a focused and defined mechanism to ensure that the availability of high quality drug-discovery relevant data in the public domain is augmented.

Foreground IP

It is the clear intention of the Open PHACTS consortium that the product of the Open PHACTS project (i.e. standards, services, tools, and infrastructure) will be freely available to use and modify. Unless otherwise agreed in the project agreement, all Foreground IP generated will be waived or provided free on unlimited license. The Open PHACTS members may agree exception to this policy in the project agreement, to allow the development of commercial services, e.g. secure data access.

Sideground IP

As the Open PHACTS project seeks to integrate diverse resources in a highly inclusive manner, provision for Sideground IP on those developments outside the scope of the project, should not be prejudicial to the openness of the overall project. On the contrary this should strengthen the reach and impact of the project. Indeed, IP generated from the use of data within Open PHACTS would be classified as Sideground IP and should be beneficial to the widespread adoption of Open PHACTS.

3. Project plan

3.1. Overall project plan

The Open PHACTS project is a radically innovative approach to combining data from multiple and currently siloed resources of chemical, biological and pharmacological information into a unified framework called the Open Pharmacological Space (OPS). This **Open Innovation** framework can be used to answer core business questions of the pharmaceutical industry such as target identification and validation, explore interaction profiles of compounds and targets, reposition existing drugs to new therapeutic areas, mitigate toxic interactions, *et cetera*. The Open PHACTS Discovery Platform will be open to all other life scientists in academia and SMEs thus helping to build a vibrant **drug discovery community**.

The Open PHACTS consortium has a strong background in '**formal semantics**', meaning in this context that the *meaning* of data and information is revealed as *an intrinsic part of its representation*. Moreover, the Open PHACTS comprises leading **text-mining**, **computational biology** and **cheminformatics** teams that together with the EFPIA partners expects to contribute significantly with expertise, **curation** resources and **data depositions**.

The general approach is to build a fully interoperable Open PHACTS Discovery Platform with:

- Semantically interlinked and **interoperable content** to answer key drug discovery questions.
- Coordinated core Open PHACTS **infrastructure and services** for managing and curating the content.
- **Ingest services** harvesting public and proprietary resources, including patents and literature, for populating the Open PHACTS Discovery Platform.
- **Digest services**, which include generic issues, such as Identity Resolution, redundancy checks, metadata/provenance, annotation, addition and data cleansing.
- **Interpretation services** that use the Open PHACTS commons for data integration, semantic-driven browsing and reasoning, complex analytical queries, and visual exploration to drive novel case-studies within public and private drug-discovery teams.

The Open PHACTS project is grounded in **drug-discovery research** and the fundamental component of the overall project plan is a strong **focus on the usage and usability** of the Open PHACTS Discovery Platform by practising teams. The potential scope of the integrated data in Open PHACTS is vast, however, to guide analysis and planning of experiments the ability of the data-model and architecture to support relevant queries and the quality of the linked data is key for success. Hence, the project will start with an intense research and **prioritisation of drug-discovery research questions** and the mapping of these questions to the most critical data sources. We envisage that within the first year pioneering users will turn to Open PHACTS to answer questions such as:

- Give me all oxidoreductase inhibitors active in human and mouse with an IC₅₀ value < 100 nM
- The current Factor Xa lead series is characterised by a bromonaphthyl substructure. Retrieve all bioactivity data in serine protease assays for molecules that contain a bromonaphthyl substructure.
- Compile a Kv channel opener set. Retrieve all bioactivity data for molecules that have been reported as 'openers' in Kv channel assays.
- What pathway/networks have been targeted by current therapies for inflammatory diseases and which toxicities have been reported

Over the course of the Open PHACTS project the Open PHACTS Discovery Platform will be increasingly capable and through the efforts and assets of core and associated partners will **support**

complex queries, analysis and visualisation. Clearly, outlining the road-map of research questions addressed will be a critical early achievement.

The stakeholders in the public and private sectors will not only act as suppliers of the Open PHACTS content, but also as curators and exploiters of the Open PHACTS content and as contributors of associated services modules to the Open PHACTS infrastructure. This will ensure a coherent system with a focus on the knowledge management issues facing practising drug discovery teams.

The Open PHACTS consortium will involve **core partners** as well as **associated partners** in the development of new and improved modules of the Open PHACTS Discovery Platform itself. To keep this complicated and ambitious project manageable we distinguish for convenience:

- The **Open PHACTS Commons** – which is the interoperable content and,
- The **Open PHACTS Infrastructure** – which are the services, algorithms and software

Together they form the integrated **Open PHACTS Discovery Platform**, and interoperability will be secured both at the content and at the infrastructure level. These two areas of development require very different measures and approaches. The stakeholders (core partners and associated partners) will contribute and exploit on a collaboration spectrum, ranging from completely closed (private data, private access, private curation) to completely open. The commons will thus need to be a blend of variably secured and accessible content with the accompanying infrastructure to support this.

The Open PHACTS Commons will be constructed as a **structured, semantically linked data and concept web** using Semantic Web technologies enhanced by data quality attribution, provenance and security capabilities. The foundations of such an approach have been laid by the Open PHACTS core consortium members and our strategic associated partners (e.g. Chem2Bio2RDF, Bio2RDF, NCBO and Sage Bionetworks).

The Open PHACTS Commons and the Open PHACTS Infrastructure that manages it will not only be used to pose and answer current research questions more efficiently; it will also offer the option to answer questions that could not be posed or answered before: for instance to make inferences over implicit pharmacological assertions to reveal new insights and to drive a new generation of multi-data discovery services.

The technical and scientific approaches go hand-in-hand and are represented as two work streams (Figure 2).

- **Work stream 1** builds the Open PHACTS Discovery Platform which covers the core services, the ingest services and the digest services.
- **Work stream 2** builds the drug discovery services that exploit, and in turn contribute to, the commons managed by the infrastructure of Work stream 1.

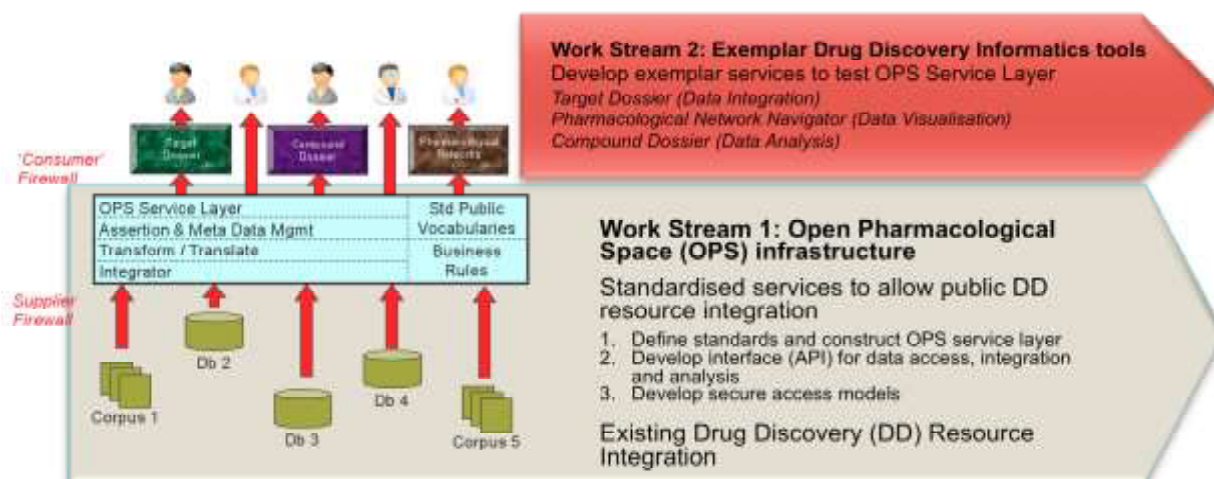


Figure 2: The two major Open PHACTS work streams in summary

Table 1 gives the mapping of work packages to work streams. All work packages have some activity and responsibilities in both work streams. This reflects the strongly interwoven approach of Open PHACTS where the discovery services and their needs drive the development of the technology and the content, while technical possibilities may in turn drive the development of highly innovative discovery approaches.

| Work Packages | Work Stream 1 | Work Stream 2 |
|--|---------------|---------------|
| WP1: IRS and Vocabulary Services | XX | X |
| WP2: General GUI and API's | X | XX |
| WP3: Architecture | XX | XX |
| WP4: Linked Data and population of Open PHACTS | XX | X |
| WP5: Specific Discovery Services | X | XX |
| WP6: Milestone Demonstrators | X | XX |
| WP7: Central engineering and Sustainability | X | X |
| WP8: Community and partnerships | X | X |
| WP9: Management and coordination | X | X |

Table 1: The mapping of work packages to work streams

Overview of prerequisites and how WP's address these in general:

The Open PHACTS Discovery Platform (commons and infrastructure) effectively builds a '**Concept Web**' of pharmacological information. For such a **semantic web of linked data** we need:

- **Common identifiers** to link data and **common vocabularies** to build **concept maps** across data.

- A **data model** for structuring, linking and organizing data. This data model captures assertions on data. These assertions are organized using (a) the **domain semantics** drawn from ontologies and information models of the community and (b) an Open PHACTS **nanopublication model** with provenance, annotation, attribution, citation and quality.
- A specialized, international team headed by a CTO to govern the **technical management** and **professional software engineering** as well as the content capture, hosting, quality assurance and streaming.
- A set of focused, **goal oriented services** to build the Open PHACTS Discovery Platform around (user driven approach).
- A **long-term sustainability strategy**, as this is not a regular research project with a finite end goal.

The design and construction of the Open PHACTS Discovery Platform and selected example applications is organized into 6 major RTD Work Packages and 3 Work Packages that deal with the managerial, the social and the sustainability aspects of the Open PHACTS Discovery Platform and community.

Work Stream 1:

WP1: (Vocabularies *et al.*) Addressing the necessary infrastructure required to create and maintain vocabularies and ontologies required for the Open PHACTS Discovery Platform; the establishment of an Identity Resolution Service to provide stable, high-quality cross referencing between different identifiers representing identical concepts; and interaction with the wider community to connect Open PHACTS vocabularies with future public standards development (with WP8). This will be performed in close collaboration with the DDMoRE (call topic 7) and EHR4CR (call topic 9) consortia

WP2: (GUIs *et al* – the Open PHACTs Explorer.) Addressing the challenge of designing and developing interfaces for semantic-driven browsing, defining complex analytical queries, and visual exploration of Open PHACTS Commons. This also includes 'software-internal' interfaces.

WP4: (Linked Data *et al.*) Addressing the services for populating, structuring, integrating, querying and basic reasoning over the Open PHACTS Commons. **Billions of assertions** of both quantitative and qualitative data will be integrated into one cohesive information model and linked data cache

The **connective tissue** between the WPs in Work Streams 1 and 2 is formed by a special engineering Work Package. Crucially the two top-level work streams cannot and must not diverge, and must be firmly coupled throughout the entire lifetime of the project from the very start. The Open PHACTS Discovery Platform (Work Packages 1, 2, 4) must be developed against a concrete set of applications represented by the Open PHACTS Explorer as **example applications** (Work Packages 2, 5, 6), governed by a central dedicated Work Package dealing with technical management of this complex interplay. The agile, **user-driven approach** of developing working-to-working releases is clearly seen in the top-level Gantt-chart of this project; infrastructure development, population with data and exploitation must proceed in parallel throughout the whole project life-time and beyond.

WP3: (Architecture *et al.*- the Open PHACTS API) A strong **technical management team** that designs and implements the technical design protocols, messages, interfaces, plug-in frameworks and governance that is needed to organize and maintain the services. Because of the special character of this proposal, namely with the aim to deliver a professional enabling technology platform, not a scientific, finite discovery, we have included this separate Work Package, in addition to the classical scientific management tasks of the project, which are covered in WP9.

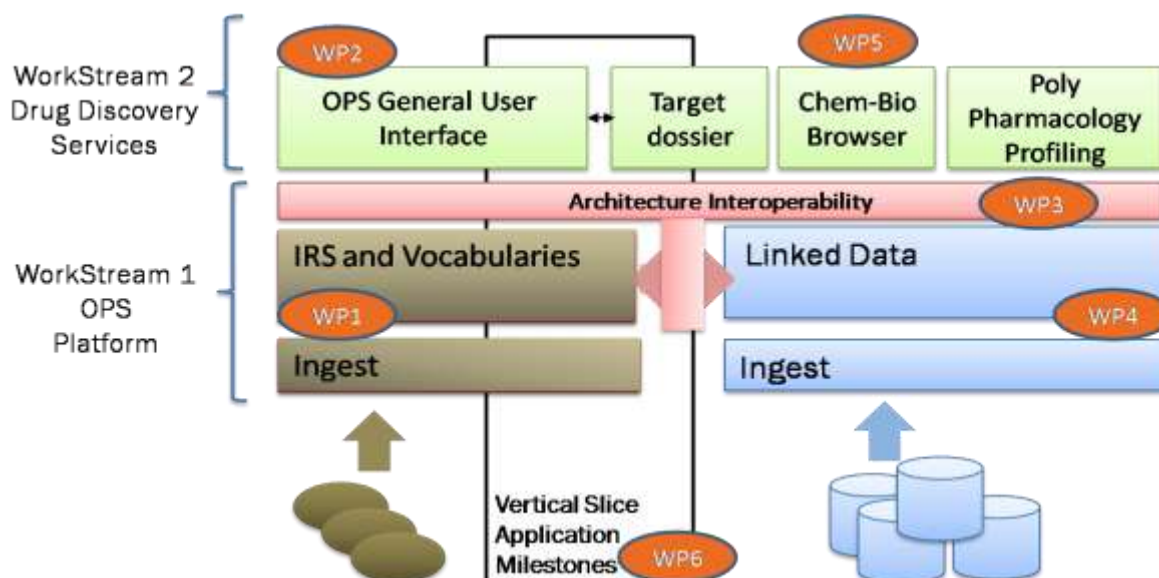


Figure 3: The mapping of the work packages to the major elements of Open PHACTS

Work stream 2:

The Drug Discovery services are organized into two groups:

WP5 (Core services *et al.*- example applications) Addressing the ability of the Open PHACTS Discovery Platform to deliver **practical answers** to drug discovery research questions by facilitating powerful and user-friendly access to the integrated chemical and biological resources. Example applications will include a target dossier, chem-bio browser and poly-pharmacology browser.

WP6: (research use cases) To ensure the delivery and alignment of **key functionality** which will allow exploitation of the Open PHACTS Discovery Platform in order to answer key research questions important to both industry and academia in application domains and exemplar services tailored for specific research questions. Three case studies dealing with prediction of drug/transporter interaction, blood-brain barrier permeation and tissue distribution, and target validation will be conducted

We can think of the Discovery Services as “applications” of the Open PHACTS Discovery Platform. The purpose of the Open PHACTS Discovery Platform is to **serve applications**. Without applications we do not know the appropriate content of the vocabularies and the commons, the appropriate functionality of the services and the acceptable operational (non-functional) capabilities of the services.

We therefore organize the project to be completely **Application-Driven**. The design, development and deployment of the Open PHACTS Discovery Platform is organized around “vertical slices” through the services: just enough content and just enough capability to support applications to answer priority research questions defined by the industry and academic members involved in drug discovery research. These applications start simple and become increasingly more sophisticated through **Agile Development Cycles** (see WP3), incorporating user feedback, more services, richer capabilities of those services and more content. The milestones of the project are based on these vertical slices.

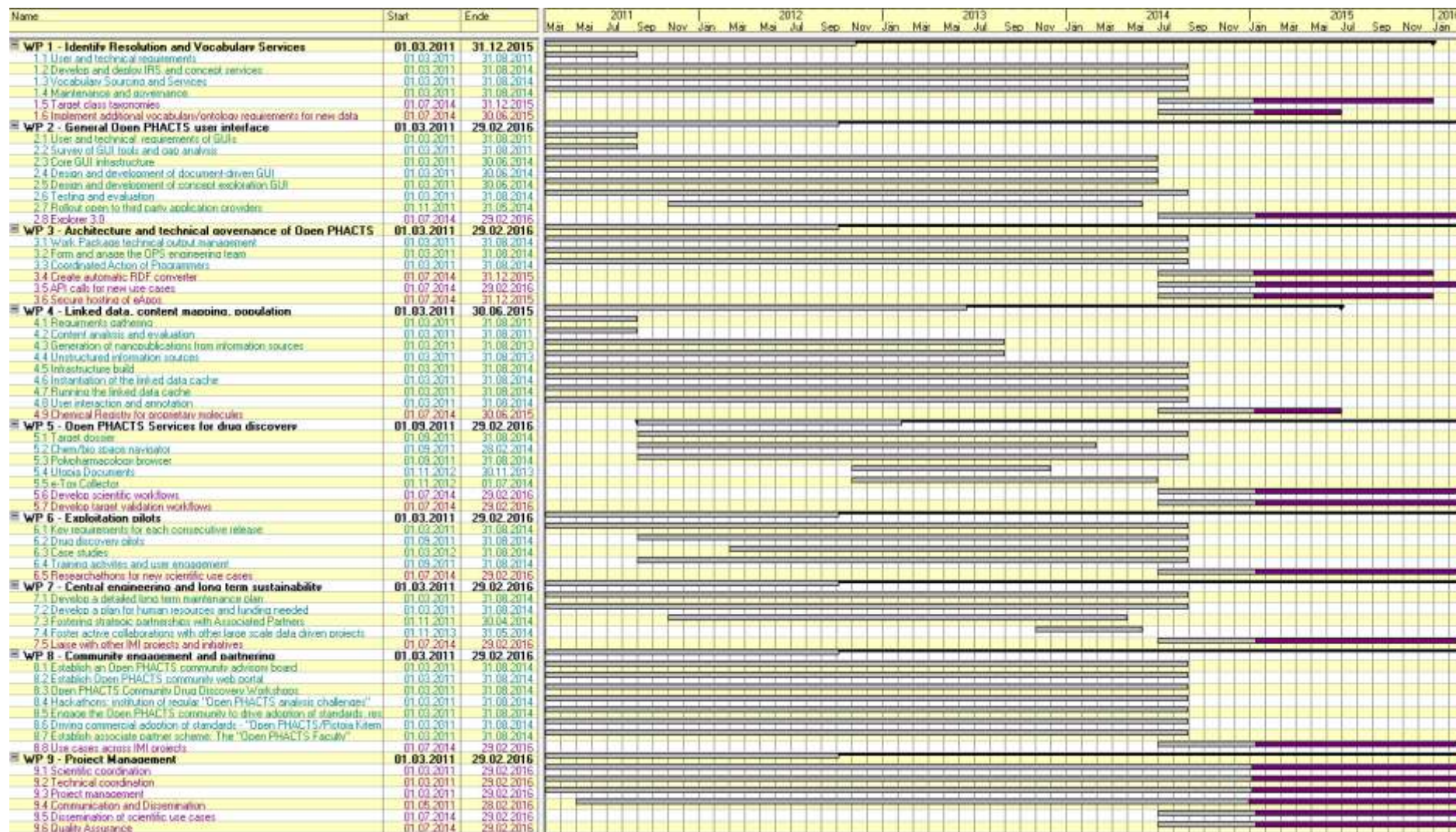
By being application driven, the utility and potential of the Open PHACTS Discovery Platform can be revealed early on in the project, scrutinized and steered. Our scientific (and technical) users and partners can be brought on board earlier, so we can build **community engagement** more readily. Our technical team (WP3) can get concrete requirements and build to realistic needs to rally around and ensure that the services interoperate to achieve a specific goal.

The “builds” of the Open PHACTS Discovery Platform for the applications can be organized into phases that are milestones for the projects. The first will be as much a fact-finding, team building, service stressing exercise. Subsequent milestone builds are intended to be **incremental “working to working”** releases that become richer and more capable.

WP7: (Sustainability *et al.*) The system needs to be operational way beyond the finite time period covered by this IMI project. This 'sustainability' involves many elements, ranging from architectural, infrastructure related, hosting/services related, human resources and funding. These aspects will be addressed properly and professionally from the onset of the project to ensure that once successfully completed, Open PHACTS is sustainable, expandable and well managed, so that it can serve the scientific and pharmaceutical communities far **beyond the projects life time**.

WP8: (Community *et al.*) As stated earlier, although the Open PHACTS consortium covers a broad range of expertise, much additional value is available by collaboration with our 'associated partners'. Some crucial partners (NCBO, Onto Text, CrossRef and others) are already aligned with us, either through the Concept Web Alliance or through specific Memoranda of Understanding. A dedicated WP is in charge of actively and continuously identifying, approaching, engaging and maintaining relationships with an expanding group of Open PHACTS associated partners and the wider drug discovery community.

3.2 Timing chart



3.3 Work description

Open PHACTS is an ambitious and challenging project with many interdependent work-packages. Details of tasks and deliverables are found in the specific work packages outlined below and a more **in-depth description of the semantic integration**, concept mapping and approach to provenance tracking and assertional compression can be found in **Annex 1**.

The key objective of this project is to build the Open Pharmacological Space infrastructure, populate the Open PHACTS Discovery Platform with semantically linked data and deliver this to on-going drug discovery projects. A key challenge for this project will be to ensure full interoperability and ability of the framework to solve research questions and we have chosen to respond to this challenge by adopting a user-driven agile software development approach where the development of this project proceeds through a series of “vertical builds”, i.e. full Open PHACTS releases with gradually more interlinked data and complex functionality. The Open PHACTS road map is developed by the academic and industrial drug discovery teams in WP6 to guide the technical requirements gathering in WP 1, 2 and 4.

A second challenge is to ensure that the underlying concept maps and architecture is solid, performing, and capable of responding to the needs of users, that vocabularies and data-sources of good quality and correctly mapped and that **data provenance** (the “**evidence**” for an assertion) is transparent and traceable. This challenge is met by developing a series of core services for drug research (WP5), deploy a series of specific research projects (WP6) as well as application within ongoing drug-discovery projects (EFPIA partners). The needs and feedback from the drug-hunting teams will **drive services and GUI capabilities** but also guide the development and quality control of vocabularies and steer the selection and linking of data. It will also drive decisions on concept maps, data-models and architecture (e.g. how to handle chemical structures, quantitative data from dose-response experiments and representations of results from high-content biology and *in-vivo* experiments)

Standards are only as good as their **implementation** and as useful as their **adoption**. A third challenge for this project is to ensure involvement of the life-science community beyond the Open PHACTS partners. A related issue is the long-term sustainability of the Open PHACTS framework and establishment of security models for handling proprietary data and queries. The Open PHACTS project approaches this with a dedicated effort in community building by establishing the concept of **associated partners**, a series of open, community-building, workshops hosted by EFPIA partners and activities to drive the academic and commercial adoption of standards (WP8). Options for long term sustainability, links to **European life-science infrastructures** and establishing appropriate **security models** are developed in WP7.

Finally, the overall management and scientific coordination of this large, distributed project is outlined in WP9. We recognize the need for strong, professional project management in addition to the scientific steering and governance, dissemination and IPR activities. An additional challenge in a large, distributed, software development project is to ensure the technical governance, interoperability and architectural coherence. This is addressed in the Open PHACTS project through a strong technical and architectural steer (WP3). We will develop a coordinated engineering approach with clearly articulated roles and responsibilities for participating parties as outlined in WP3. In addition we will appoint a “Chief Technology Officer” (CTO) to take overall lead and responsibility for the technical aspects of this project and a “Chief Information Officer” (CIO) to have the same role for content. These are key tasks for the research activities and critical to the successful development of the Open Pharmacological Space, hence we separate this into a separate work-package to ensure visibility, authority and governance.

Table 3.3 a: Work package list

| WP No. | Work package title | Type of activity | Lead part. No. | PMs[§] | Start month | End month |
|---------------|--|-------------------------|--------------------------------------|------------------------|--------------------|------------------|
| 1 | Identity Resolution and Vocabulary Services | RTD | 11, 27 UNIMAN, Janssen | 239,7 | 1 | 58 |
| 2 | General Open PHACTS User Interface | RTD | 11, 21 UNIMAN, GSK | 183,9 | 1 | 60 |
| 3 | Architecture and Technical Governance of Open PHACTS | RTD | 9, 17/25 VUA, Pfizer/ConnDisc | 197,8 | 1 | 60 |
| 4 | Linked Data, Content Mapping, Population | RTD | 8, 23 RSC, SIB | 390,8 | 1 | 52 |
| 5 | Core specific services for drug discovery questions | RTD | 6, 16 PSMAR, AZ | 583,5 | 6 | 60 |
| 6 | Exploitation pilots | RTD | 2, 19 UNIVIE, Janssen | 231,7 | 1 | 60 |
| 7 | Central Engineering and Long term sustainability | OTHER | 7, 17/25 LUMC, Pfizer/ConnDisc | 114,9 | 1 | 60 |
| 8 | Community Engagement and Partnering | OTHER | 1, 8 GSK, RSC | 108,9 | 1 | 60 |
| 9 | Governance, Project Management and Dissemination | MGMT | 1, 2 GSK, UNIVIE | 168,6 | 1 | 60 |

[§] Person months

Table 3.3 b: Work package description

| | | | | | | | | | | | |
|-----------------------------|---|--------|------------------------------|-------|------|-----|----------|----------|---------|---------|------|
| Work package number | 1 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Identity Resolution and Vocabulary Services | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 1 | 2 | 4 | 6 | 7 | 8 | 9 | 11 | 12 | 14 | 15 |
| Participant short name* | GSK | UNIVIE | UHAM | PSMAR | LUMC | RSC | VUA | UNIMAN | UM | USC | UBO |
| Person-months | 0,5 | 1 | 1 | 7,5 | 12 | 9 | 26 | 40 | 39 | 7.5 | 17,5 |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | F | F |
| Work package number | 1 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Identity Resolution and Vocabulary Services | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 16 | 17 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 31 | |
| Participant short name* | AZ | Pfizer | HLU | Lilly | NBIC | SIB | ConnDisc | EMBL-EBI | Janssen | SciBite | |
| Person-months | 22 | 4,6 | 4 | 0.5 | 22,5 | 4,9 | 0,3 | 16,7 | 3 | 0,2 | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | IK | IK | IK | IK | F | F | F | F | I | F | |

Preamble

In the scope of Work Stream 1: WP4 defines, extracts and manages the nanopublications that make up the Open PHACTS Commons by using vocabularies to describe, index, link, query and reason over these nanopublications. WP2 uses vocabularies and ontologies that process the nanopublications in the Open PHACTS Commons and **render them to machines and humans** for the final purpose of Open PHACTS; Knowledge discovery

WP1 manages the domain vocabularies that emerge from text mining the literature (papers, patents) and the vocabularies that are defined by the community as controlled vocabularies and used by **curators of datasets** (such as the OBO family of ontologies). It manages the mappings between these vocabularies and the design of curated ontologies. It also manages the vocabularies for provenance, discourse and policy needed to define and operate the nanopublications in the Open PHACTS Commons as defined in WP4.

The WP distinguishes between two important notions of vocabularies:

1. the **symbols** used to denote and resolve concepts (denotations)
2. the **organisation and structuring** of the concepts using ontologies and thesauri

The Open PHACTS vocabularies and their associated services fall into the following categories:

An **Identity Resolution Service (IRS)** to look up concepts by their denotations and map to and between the vocabularies by their denotations

- This will provide the editable dictionaries of identifiers, names, synonyms & basic taxonomic relationships required to integrate and to link the Open PHACTS Commons and to create a

Concept Map of the Open PHACTS Commons. The IRS will be based on the ConceptWiki developed by NBIC and BridgeDB by UM for general identifier resolution mappings. The IRS will contain an active identity resolution and mapping toolbox, allowing different symbols from distributed systems to be cross mapped in order to ensure full interoperability. The IRS will work with other community identity services and resources.

- Vocabularies and vocabulary services for creating, incorporating and extending domain vocabularies of the Open PHACTS Discovery Platform.
 - This will provide access to the vocabularies in the field, source, ingest and convert third party vocabularies, **synchronise** local extensions with **community updates**, feed updates into community extensions, support the development and merging of vocabularies when needed and the support creation of new concepts by community crowd-sourcing, expert-sourcing, text mining and bulk-updates. These will be based on a number of systems including (a) community ontologies from NCBO and the OBO Foundry and minimum information models from the MIBBI group (b) expert vocabulary creation and editing tools using ontologies (e.g. Protege-OWL, SKOSedit, OBOEdit), (c) crowd-source concept creation tools (e.g. ConceptWiki) and (d) text mining (e.g. UBO, LUMC, EMC and CNIO). The consortium members and partners are the authors of these tools and experts with track records in creating and managing vocabularies.
- Vocabularies and vocabulary services for creating, incorporating and extending nanopublication vocabularies of the Open PHACTS Discovery Platform.
 - We will build upon emerging vocabularies for provenance (OPM, VoiD), discourse (SWAN), communities (FOAF, SIOC), versioning (Momento) and aggregation (OAI-ORE).
- Vocabulary services for querying, browsing and reasoning over the Open PHACTS Commons
 - This will provide the richer services that use the structures, constraints and relationships of ontologies to drive the user interfaces in WP2, the querying, tagging and text-mining in WP4, rich applications in WP5 and pilot capability in WP6. These services will be based on established services such as LarKC, BioPortal and ConceptWiki and new services identified in the project.
- Vocabulary governance and policies for the Open PHACTS Discovery Platform.
 - The IRS, supported by the Vocabulary services, will deliver the services and governance policies to enable "**gold standard**" vocabulary creation, vocabulary evolution, quality, separations between community and authority contributions, and third party ingest/takedown.

Objectives

To enable community engagement will be key. A workshop will be held to specifically engage key stakeholders and the wider community (see WP8). Beyond this, efforts to build a wider Open PHACTS community will continue to drive the adoption of Open PHACTS standards, such as those for nanopublications (see www.nanopub.org) and vocabularies

There are a range of consumers for the services including:

1. **people and machines adopting different roles** – populating, curating, validation, data mapping and information extraction; and
2. machines who will use IRS as a service.

The Specific Objectives of this WP are:

O.1.1: Identify, define and deliver the IRS and Vocabulary Services

O.1.2: Identify, define and deliver the content of the IRS and Vocabulary Services

O.1.3: Define the governance for the content and change management of the IRS and Vocabulary Services

O.1.4: Run the IRS and Vocabulary Services (With WP3)

O.1.5: Promote adoption of IRS and Vocabulary Services beyond Open PHACTS (with WP8)

Description of work

As an industrial and academic community we suffer from a significant hidden problem and cost associated with how we describe (refer to) concepts in biomedical science (diseases, targets, drugs, compounds, bio-processes, assays, toxicological endpoints, pathologies, institutes, reagents, etc). The lack of generally adopted naming standards (vocabularies, taxonomies and ontologies), especially in pre-clinical research, seriously compromises our ability to:

- Discover new knowledge and develop predictive approaches (for example drug efficacy and safety).
- Collaborate within and across research domains, thereby blocking information sharing.

The BioSharing group (www.biosharing.org) distinguishes three kinds of metadata for data:

- **Minimum information models** that *prescribe* properties needed for capturing and organizing different classes of data.
- **Vocabularies** used to *describe* the **properties** of data and the **relationships** between data.
- Formats that define the **syntax** for structuring data.

These types of metadata are interrelated and have various organisations that drive and coordinate community activity. Minimum information models are promoted by the MIBBI.org; vocabularies by the OBO foundry and NCBO and formats by ad-hoc groups such as EMBOSS, ISA-TAB and FuGE. In the context of the Semantic Web approach of Open PHACTS, formats to express information, such as **RDF** and **OWL** are also an integral part of the system.

Vocabularies define agreed, common and controlled terms that are used to describe data so that it can be shared. Each concept in Open PHACTS has one or (usually more) vocabulary terms and the concept has a **Universally Unique Identifier** (UUID) and all terms, tokens, Identifiers etc. that denote that particular concept are called 'symbols' from here on. Symbols (Vocabulary terms) denote concepts.

Vocabularies range from simple keyword collections (such as UniProt keywords) through simple thesauri (such as MeSH terms) to terms organised into rich models of knowledge called ontologies. An ontology is a formal, explicit representation of the knowledge by formalising the meaningful relationships between a set of **concepts** within a domain, the properties of each concept describing various features and attributes of the concept, as well as restrictions on the properties that define the relationships between concepts.

The taxonomic structure of an ontology means that symbols, and the data described by those symbols, are properly classified and indexed. Simple thesauri-like ontologies use simple terms to *denote* concepts and organise them into indexing and synonym structures - SKOS is an RDF-mapped language for representing thesauri. Richer ontologies *define* concepts and use reasoning over those definitions to infer classifications and entailments - OWL is an RDF-mapped language for representing rich ontologies. One of the major challenges addressed by Open PHACTS is to allow comprehensive and **correct cross-mapping of symbols to concepts**, thus **removing ambiguity** and introducing **interoperability** between environments using different symbols for the same concept.

In the Open PHACTS Discovery Platform, ontologies are used to structure vocabularies and act as controlled vocabularies for data annotation, linking and mapping data and driving text mining. The ontology classification taxonomies are used as mechanisms for indexing, querying and reasoning and driving rich visualisation tools. Rich constraints on properties and allowable relationships are used for reasoning to infer inconsistencies and new assertions.

These vocabularies are used throughout the Open PHACTS framework. For example the vocabularies will serve as controlled vocabularies for concept tagging (WP4). The selection and development of Open PHACTS vocabularies will be guided by the research questions and capability road-map delivered in WP6.

Since 2005, the systems biology community has developed a system to provide shared URIs to identify concepts used in annotations (MIRIAM identifiers). These URIs can be resolved through the domain identifiers.org into various resources serving the corresponding information using MIRIAM Registry. Identifiers.org is a community project that is replacing many of the former URI resolving systems such as LSRN, Bio2RDF, Sharednames. Several of the goals of the Open PHACTS project will be reached only with the existence of shared stable URIs. These will permit unique identification of the subjects and objects of RDF triples, and therefore data integration. The registry and <http://identifiers.org> will allow the users of Open PHACTS technology to access the various data from alternate providers, for instance based on reliability or geographical proximity. Associated services, such as identifier validation, will help improve the quality of Open PHACTS content. Finally, MIRIAM URIs will make Open PHACTS and DDMoRe infrastructure directly interoperable. We expect the Open PHACTS' project requirements to lead to an expansion of the registry and growth of the Identifiers.org user community.

To ensure maximum community impact, collaborations with the consortia of call topic 7 and call topic 9 will be established. Respective memoranda of understanding are given in Annex 2

| | | | | | | | | | | | |
|-----------------------------|------------------------------------|--------|------------------------------|-------|------|-----|----------|---------|-----------------------|------|-----|
| Work package number | 2 | | Start date or starting event | | | | | | | M1 | |
| Work package title | General Open PHACTS user interface | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 2 | 4 | 5 | 6 | 7 | 8 | 11 | 12 | 13 | 14 | 15 |
| Participant short name* | UNIVIE | UHAM | BIT | PSMAR | LUMC | RSC | UNIMAN | UM | ACK (till 17.09.2014) | USC | UBO |
| Person-months | 3,5 | 3,5 | 6 | 2,6 | 11 | 9 | 33 | 4 | 59,7 | 14,5 | 6 |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | F | F |
| Work package number | 2 | | Start date or starting event | | | | | | | M1 | |
| Work package title | General Open PHACTS user interface | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 1 | 18 | 21 | 22 | 23 | 24 | 25 | 27 | 31 | | |
| Participant short name* | GSK | Esteve | HLU | Lilly | NBIC | SIB | ConnDisc | Janssen | SciBite | | |
| Person-months | 7,2 | 6 | 3,5 | 4 | 6 | 0,5 | 1,3 | 2,5 | 0,1 | | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | | |
| Funding claimed (F/ IK / N) | IK | IK | IK | IK | F | F | F | IK | IK | | |

Preamble

This work package addresses the mechanisms of communication between the Open PHACTS Discovery Platform and its users. It includes the analysis, design, documentation and implementation, testing and refinement of a Graphical User Interface (GUI – Open PHACTS Explorer) for interaction with human users. The work package also influences the design of the system's Application Programming Interface (API) (WP3) by ensuring that where possible they are suitable for real-time interaction and thus the creation of future GUIs and 'mashups'. The design of such interfaces is of primary importance to the success of this project: from the perspective of the majority of users (especially novice or casual users), the graphical interface *is* the system. Similarly the uptake, impact and sustainability of the Open PHACTS Discovery Platform by future integration with 3rd party sources and tools rely on a well-designed lightweight API. Thus, rather than assuming that suitable interfaces will emerge merely as a by-product of building the Open PHACTS Discovery Platform (as experience strongly suggests this is not the case), an entire work package is dedicated to ensuring that user-facing components are built in a consistent and coherent manner, based on current best-practices in Human-Computer Interaction.

The objectives and tasks of this work package are deeply intertwined with those of WP1, WP3, WP4 and WP6 and are driven by the research questions. The GUI (Open PHACTS Explorer) that will be developed here form the prime user-facing interface and the manifestation of the requirements of those work packages. The ontologies for scientific discourse, provenance etc., and the tools for disambiguation of the concepts that constitute nanopublications are managed by WP4. WP3, which deals with architectural matters, sets out the primary general computer-facing interfaces.

Objectives

The general objective of this work package is to produce a suite of interactive interfaces via which users can access the Open PHACTS Discovery Platform. We intend to support two starting-points for exploration:

abstract concepts - in which the user has in mind a particular question and wishes to access the knowledge held in the Open PHACTS Discovery Platform with regard to that topic (e.g. "which enzyme inhibitors are associated with the treatment of bowel cancer"); and concrete artifacts (for example Web Pages or scientific articles), in which concepts can be identified that are then explored within the Open PHACTS Discovery Platform. The interactive components of Open PHACTS fall into three main categories: concept-driven exploration; artifact-driven exploration; and general interaction. Informed by the scientific questions and the requirements of all stakeholders, specific objectives are thus:

O.2.1. To develop the system's core interface, responsible for exposing the contents of the Open PHACTS Discovery Platform and the available services in a variety of ways (e.g. browser, mobile devices, downloadable for offline analysis).

O.2.2 To develop the system's Concept Exploration interface, which provides a mechanism by which users can navigate and visualize the system's nanopublications and their relationships.

O.2.3 To develop the system's Document Exploration interface, enabling users to access the Open PHACTS Discovery Platform from the starting-point of online and electronic documents and articles.

O.2.4 To provide a gateway to external data sources and services, including both primary data sources and 'peri-scientific' ones (i.e. information on resources needed in the process of creation and discovery of further knowledge).

O.2.5 To provide GUI components suitable for incorporation into other online systems via 'mashups'.

O.2.6 To drive requirements for Open PHACTS content and services.

O.2.7 To deliver a sustainable foundation for the long-term provision of these interfaces (WP7).

Description of work

For many users, the Open PHACTS Explorer will be the primary means of interaction with the Open PHACTS Discovery Platform and its content. As well as being a vital component for users to access the system without the need for specialist knowledge or tools, presentation of content in a coherent means provides the consortium with a mechanism to validate the underlying information model and controlled vocabularies, and thus iteratively drive the requirements developed in WP1 and WP4.

Appropriate aspects of the general GUI (Open PHACTS Explorer) should be exposed in a manner that can be incorporated as a service component into other tools (e.g. those developed within WP5 as well as external tools). The GUI will reuse as many prior tools as possible, and will be developed in a continuous and iterative manner, following the pilots defined in WP6. As part of this work package, we will revisit and refresh the initial reports and surveys on the state interfaces to linked data systems in the field of biology / biochemistry.

A suite of interfaces will be built to be used with different modes of access to the Open PHACTS Discovery Platform. These will cover both concept- and document-driven exploration via a variety of devices (e.g. desktop and mobile). For each, a phased development is foreseen with regular releases incorporating the results of rounds of testing, evaluation and feedback, following the Agile approach as described in WP3. Software engineers and designers working in interfaces will be an integral part of the ET (TTF in WP9). As

example applications may also be seen as user interface, there is some overlap with WP5. However, in WP the technical issues are dealt with, whereas in WP5 the scientific ones are described.

T.2.1 User and technical requirements of GUIs.

The core GUI (Open PHACTS Explorer) enables the contents of the Open PHACTS Discovery Platform to be searched, used and 'interrogated' for the purpose of knowledge discovery and reasoning (deduction).

The GUI will also be made to function as a 'funnel' to add knowledge to the Open PHACTS Discovery Platform by enabling users to contribute concepts and assertions that are not yet recognised in the text being looked at, and presumed still missing in the Open PHACTS information store. The GUI will combine this with the functionality for core concepts and cardinal assertions in (scientific) text to be indexed/recognised and made interactive to produce a menu of links to appropriate further knowledge, such as literature, databases, experts, and search options.

To this end, a set of interfaces is to be created with the functionality to:

- enable assertions in the central triple store (Open PHACTS content) and their connections and hierarchies (nanopublications > cardinal assertions) to be read by humans in natural language;
- enable the community at large to contribute, review, validate, curate, and comment on nanopublications (providing the intuitive interfaces for the objectives of WP4 to be realised);
- enable natural language assertions to be recognised by users in the text and rendered into machine-readable format;
- render human-readable cardinal assertions in (a growing number of) languages other than English;
- map nanopublications to cardinal assertions

Points of view specific to drug discovery pharmacologists, chemists, and molecular biologists will be incorporated. Such points of view will be obtained via personal meetings and wider surveys and involve partners and users into the in-process testing activities as well as their practical execution. The data collected in the aforementioned surveys will be systematically analysed in order to transform them into concrete and usable feedback for the development of the interfaces (Requirements document, deliverable D.2.1), in regard of usability (Usefulness of the tools, deliverable D.2.2) and of user satisfaction with the different versions (Series of versions satisfaction, deliverable D.2.3 coordinated with deliverables from WP6 pilots). Thus, based on academic and industrial partners' and users' detailed requirements, plans for subsequent phases will continuously be updated to take account of such requirements.

Prioritization of requirements for the Open PHACTS Explorer is to be based on analyses of replies to periodical requirements surveys and validation of prioritization with all IMI-Open PHACTS partners.

Contributing Partners: UHAM, BIT, PSMAR, LUMC, NBIC, RSC, UNIMAN, ACK (till 17.09.2014), USC, UBO, GSK, Lilly, Esteve, Janssen

T.2.2 Survey of GUI tools and gap analysis

In parallel with Task 2.1, a survey will be conducted to identify existing 3rd party components that could contribute to the construction of the Open PHACTS Explorer. This survey will cover: mechanisms of intercepting / analysing online content (e.g. via proxies, portals, relays or browser plug-in architectures), document interrogation mechanisms (e.g. PDF, html, word processor documents), and visualisation/exploration tools (e.g. Ontogator, Similie, BioGPS as well as tools from the biochemical domain). The survey will be analysed against the project's requirements to identify gaps where specific components will need development.

Contributing Partners: UNIVIE, UHAM, BIT, LUMC, UNIMAN, ACK (till 17.09.2014), UBO, GSK, Lilly, HLU, Esteve

T.2.3 Core GUI infrastructure

This task encapsulates the development of the core graphical interface to the Open PHACTS Discovery Platform, and essentially provides the 'portal infrastructure' necessary to sustain the other GUI development tasks.

Contributing Partners: UNIVIE, LUMC, NBIC, SIB, UNIMAN, UM, ACK (till 17.09.2014), UBO, Lilly, HLU, Esteve

T.2.4 Design and development of Document-driven GUI

This task provides a graphical interface via which concepts in documents (online and PDF) can be associated with Open PHACTS nanopublications. The interface will provide mechanisms for exploration, curation, contribution, and visualisation of nanopublications directly from within the act of reading scientific articles online or in PDF form.

Contributing Partners: LUMC, NBIC, RSC, UNIMAN, UBO, ACK (till 17.09.2014), USC, Lilly, Esteve

T.2.5 Design and development of concept exploration GUI

This task develops mechanisms for exploring the nanopublications within the Open PHACTS Discovery Platform, providing faceted mechanisms for identifying concepts and their relationships, and for displaying these in human-readable form. As with Task 2.4, mechanisms will be provided for community contribution and curation of nanopublications.

Contributing Partners: LUMC, NBIC, RSC, UNIMAN, UBO, ACK (till 17.09.2014), USC, Lilly, Esteve, Janssen

T.2.6 Testing and evaluation

Closely associated with Tasks 2.3, 2.4 and 2.5, rollout and running of the interfaces will progressively be broader in line with their development phases. A key part of this task will be the rounds of testing, evaluation and feedback. This specific task includes testing aimed at defining methods and logistics for the participation of drug discovery experts (industrial and academic) in the testing and feedback of the usability and ergonomics of the interface.

Contributing Partners: LUMC, NBIC, RSC, UNIMAN, ACK (till 17.09.2014), USC, Lilly, Esteve, GSK

T.2.6.1 Requirements gathering document generation and internal review

A requirements-gathering document will be generated together with the other WP2 partners. It will be distributed to all Open PHACTS partners for review and validation.

T. 2.6.2 Feedback on Usability A selection of at least 10 organisations from both the pharma industry and academia, and belonging to several EU countries (they will be chosen taking into account their market representativeness and expertise in the field) will be invited to contribute input to the requirements gathering document. They will interact with the consortium through ad-hoc meetings and surveys.

T.2.7 Rollout open to third party application providers (WP8)

A key part of Task 2.7 will be to engage with the providers of scientific information (authors and researchers as well as STM publishers), to encourage participation. This will be done by formulating specifications that would ensure proper citeability for nanopublications, so that every nanopublication, if extracted from a published peer-reviewed article, effectively becomes a citation to that article, providing a strong incentive for publishers to participate. In addition, interfaces will be created that allow the 'exposure' of nanopublications in conjunction with their source (narrative) texts. This task will be done in conjunction with WP8

Contributing Partners: LUMC, NBIC, SIB, UNIMAN, ACK (till 17.09.2014), USC, Lilly, Esteve

T.2.8 Explorer 3.0

Develop, test and implement the next generation GUI which includes query possibilities for different target classes, pathway data and diseases. This will increase the capabilities for running complex scientific use cases. It will also provide enhanced download capabilities and interlink with priority Open PHACTS eApps.

Contributing Partners: UNIMAN, UNIVIE, SciBite, GSK, Esteve

Task Summary

Tasks 2.3, 2.4 and 2.5 deal with the practical execution of the user interface. They are iterative in nature and will be done concurrently to a degree, allowing for intermediate analysis of results and production of the corresponding report.

The internal survey coordinated by the project partners related to the needs of the drug discovery world will be used as a first source of feedback. In this task, the methods and the logistics will be defined for the participation of experts (industrial and academic) in the testing activities as appropriate to an inter-organizational setting.

Frequent technical meetings will be undertaken in order to follow-up on the development within a strategy of rapid prototyping intended for obtaining early feedback from users.

An internal survey coordinated by the project partners related to the drug discovery world needs will be used as first source of needs.

These tasks will be coordinated with WP6 in order to incorporate the interface in the successive development of the Open PHACTS pilots. All technical developments will be coordinated through WP3

| | | | | | | | | | | | |
|-----------------------------|--|--------|------------------------------|------|-------|----------|---------|------|--------|---------|----|
| Work package number | 3 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Architecture and technical governance of Open PHACTS | | | | | | | | | | |
| Activity Type * | RTD | | | | | | | | | | |
| Participant number | 1 | 2 | 4 | 5 | 6 | 7 | 8 | 9 | 11 | 12 | 16 |
| Participant short name* | GSK | UNIVIE | UHAM | BIT | PSMAR | LUMC | RSC | VUA | UNIMAN | UM | AZ |
| Person-months | 0,5 | 3 | 1,4 | 6 | 2,9 | 27,4 | 4,5 | 40,5 | 12 | 10 | 3 |
| Other resources (YES/NO) * | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | F | IK |
| Work package number | 3 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Architecture and technical governance of Open PHACTS | | | | | | | | | | |
| Activity Type * | RTD | | | | | | | | | | |
| Participant number | 17 | 20 | 22 | 23 | 24 | 26 | 27 | 28 | 30 | 31 | |
| Participant short name* | Pfizer | ME | Lilly | NBIC | SIB | EMBL-EBI | Janssen | OGL | ALM | SciBite | |
| Person-months | 18,8 | 8 | 4,5 | 9,3 | 6,6 | 17 | 1 | 8 | 2,5 | 10,9 | |
| Other resources (YES/NO) * | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | IK | IK | IK | F | F | F | IK | F | IK | F | |

Preamble

This WP governs the activities and approaches Open PHACTS has in place to ensure proven practices to enable a professional **Open Source software development environment**, including building a mixed distributed and central team as an absolute requirement for an international software development project.

One of the most frequently reported traps to fall into is to expect that academic and industrial groups, united around a common goal such as building a widely used and reliable knowledge discovery platform will spontaneously move from cyclic prototyping to the delivery of a **working system**.

The design and creation of a **reliable and scalable system** is typically beyond the output of the scientific process. To date there are hardly any 'bioinformatics' environments meeting the criteria defined for the Open PHACTS Discovery Platform earlier.

As Open PHACTS is a radically different approach to combining data and information from multiple and currently siloed resources and will be open 'external contributors and to the widest constituency of the pharmacology and biomedical research communities, including public and private enterprises, significant managerial challenges are intrinsic to this project. Bringing the vast expertise, prototypic tools, databases and practices of over 20 distributed partners together into a unified and reliable Open PHACTS Discovery Platform system is already a major challenge and requires this specific WP to secure progression and **delivery beyond prototyping**.

As described in more detail in the general description of work we distinguish:

- the Open PHACTS Commons – which is the content and
- the Open PHACTS Infrastructure – which are the services and software

Together they form the integrated Open PHACTS Discovery Platform.

The stakeholders will contribute and exploit on a collaboration spectrum ranging from completely closed (private data, private access, private curation) to completely open. The commons will thus need to be a blend of variably secured and accessible content with the accompanying infrastructure to support this.

The major technical innovation (and challenge) is that the Open PHACTS commons will be constructed as a structured, semantic linked data web using Semantic Web technologies enhanced by data quality attribution, provenance and security capabilities. The foundations of such an approach has been laid by the Open PHACTS consortium membership (e.g. LarKC, ConceptWiki, ConceptWebAlliance, Protégé-OWL, LinkedLifeData) and our strategic partners (e.g. Chem2Bio2RDF, Bio2RDF, NCBO, Sage Bionetworks etc).

The working architecture figure is split into a series of services. The services are the focus of various work packages (see Figure 4 below). WP3 focuses on the coordination of the Open PHACTS technical development and technical services, data format standards and the Open Pharmacological Space as a platform for plugging in third party source services, third party clients and alternative third party core services (such as alternative reasoners, alternative data stores etc).

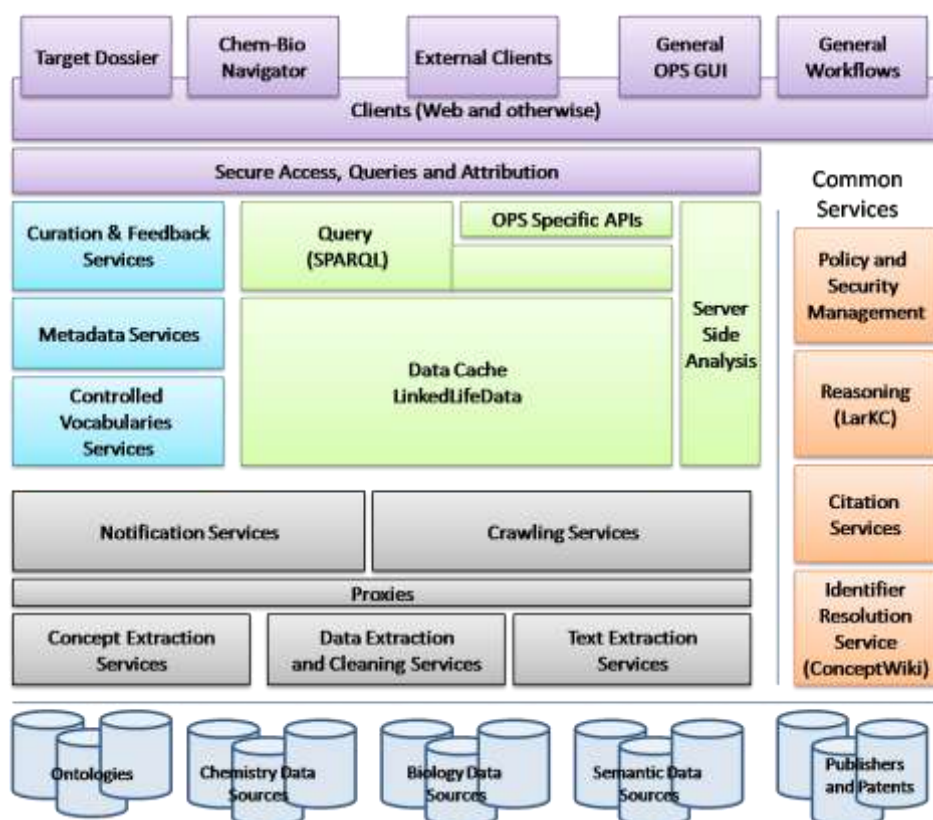


Figure 4: The broad architecture envisioned for the Open Pharmacological Space

Open PHACTS Methodology

As the Open PHACTS Discovery Platform is developed, populated and maintained by 'the community' and thus in a distributed fashion, with its own challenges, the Open PHACTS consortium will adopt a software engineering model for mixed academic and 'industrial' software development. Such a mixed model has been successfully implemented in the context of several Open PHACTS partner institutions such as the Netherlands Bioinformatics Centre (NBIC), OMII-UK, RSC/ChemSpider and the Swiss Institute for Bioinformatics (SIB). The most elaborate approach as deemed to be necessary here, implements an

organisation in three layers: work package management, coordination through an engineering team, and scientific programmers that perform most of the prototypic programming.

Objectives

O.3.1. To form, manage and monitor the technical management groups in Open PHACTS and build a sustainable technical core group to continue beyond the project phase.

O.3.2. To build and manage the professional software and content management environment to make the Open PHACTS Discovery Platform a production environment.

Tasks

T.3.1 Work Package technical output management

Each of the *work packages* (WP) is coordinated by a pair of contributors: a *Principle Investigator* (PI) and a *Technical Project Leader* (TPL). The PI is the strategic scientific lead for the WP, the TPL is responsible for the architectural and technical aspects of the collaboration. The PI and the TPL tasks call for different personalities and are therefore difficult to combine in one person. The PI is a person with a long term vision that can oversee the full project, comprising the WP. In this proposal, the PIs are typically the WP leaders. The TPL is a professionally trained senior software engineer (typically a CTO type function in the WP leaders' group) and translates the long term WP goals into practical small steps and organizes the logistics to get these steps done, one by one. For more details about the tasks of the TPL and PI, see below.

Contributing Partners: UNIVIE, PSMA, LUMC, NBIC, SIB, RSC, UNIMAN, UM, VUA, Pfizer, Lilly, AZ

T.3.2 Form and Manage the Open PHACTS Engineering Team (TTF-WP9)

In the overall organization the TPLs of all WPs are represented in an *engineering team* (ET), led by a full-time *Chief Technology Officer* (CTO), who is effectively the chair of the Technical Task Force (TTF, see overall management in WP9). The ET has two primary tasks: to facilitate the integration of the products of the WPs and to develop and integrate supporting software.

To facilitate the integration, each PL will report in ET meetings about the progress in their WP. Frequent progress reports will make it possible for the TPLs to take early note of possible areas of overlap or oversight in the entire project, so that the impact of such events can be minimized.

The ET will also have a Chief Information Officer (CIO), who is primarily responsible for governing the implementation of 'content decisions' made in conjunction with the PIs of WP1 (the vocabularies and IRS), WP2 (interfacing), WP4 (the data and information) and WP5 and 6 (ensuring that vocabularies and content support the research questions defined).

Software development in the ET focuses on *enhancing* reliability and usability of software developed elsewhere, either in the individual WPs or originating from outside of the project. This engineering work will be focused on the coverage of unit tests and documentation and on user interfaces, APIs and interoperability between components (with WP2 in the PI-lead). Performance may be an issue as well in some cases. The motto of the software engineering work in the ET will be “*no thinking beyond this point*”, indicating that this work is not meant to find new data or solutions to problems, but to implement existing prototypic solutions from the WP's into a central and reliable Open PHACTS Discovery Platform. Work methods will be based on Agile/Scrum and the 4+1 methodologies (see below).

Contributing Partners: PSMA, LUMC, NBIC, SIB, UNIMAN, VUA, Pfizer, AZ, OGL

T.3.3 Secure coordinated action of dispersed Scientific Programmers in different Work Packages

The bulk of the software engineering for Open PHACTS will be carried out by scientific programmers dedicated to a particular WP or even at associated partner (AP) institutions, not directly accountable nor governed by the Open PHACTS Executive Committee. Each software engineer is located on location in one of the collaborating groups. Continuous communication between the scientific programmers and frequent

meetings and joined coding sessions (Open PHACTS Hackathons) must assure that the 'early' efforts in the collaborating groups stay aligned. A major role of the CTO and the CIO of the ET is to manage the alignment of the developments, not only in terms of technical approach and interoperability, but also in 'timing'. For instance, the vocabularies and IRS (WP1) should grow in close coordination with the content covered in the Open PHACTS commons (WP4) and both should be just right to support the services developed in WP2, 5 and 6. This alignment is a major challenge in the Open PHACTS project, but a centralised ET as proposed here has proven to be effective in such circumstances.

The scientific programmers are positioned on location so that they can each solicit guidance from the practical users of the technology in their WP/AP. The work between the scientific programmers will therefore be distributed in such a way that optimal use can be made of the experience in each of the hosting groups. However, to mitigate the inherent risks of distributed scientific programming, such as undue duplication of effort, divergence in approach and standards, becoming stuck in eternal prototyping and more, regular meetings will be organized by the Project Leaders to discuss the progress and requirements within and across WP's/AP's.

APs can 'second' programmers or content specialists to WPs to ensure an open and inclusive approach to realize Open PHACTS. When becoming an AP, such partners agree to a minimal set of criteria, including the acceptance of the final decisive role of the CTO, the CIO and finally the EC of Open PHACTS for inclusion of data, information or software modules into the Open PHACTS Discovery Platform. Once a module of the Open PHACTS Discovery Platform (content or software) meets certain criteria (to be developed as an early deliverable in the project) it will enter the ET for further development under direct responsibility of the CTO and the CIO and the relevant TPLs, while the PIs and the scientific programmers involved will assume the role of consultants to the ET.

Contributing Partners: PSMAR, LUMC, NBIC, SIB, VUA, Pfizer, AZ

T.3.4 Create automatic RDF converter and loader

Inclusion of new data sources such as those for target validation and personalized medicine, as well as connection of commercial sources and EFPIA in house data to the Open PHACTS Discovery Platform would strongly benefit from an automatic RDF converter that loads data directly into the platform. This tool will follow the guidelines developed by Open PHACTS and allow reproducible conversion of data sets to Open PHACTS compliant RDF.

Contributing Partners: VUA, UM, ME, Janssen

T.3.5 API calls for new use cases

One of the main tasks is to explore new scientific opportunities via concrete new scientific use cases. The recent link of a patent database to ChEMBL will allow the inclusion of patent data into the Open PHACTS triple store. However, this requires to tag genes and database identifiers in patents and to develop respective API calls, driven by scientific use cases. This and the implementation of new functionalities such as strong links to workflow engines require new API calls. These will be provided, tested, and then deployed at the Open Link Platform.

Contributing Partners: EBI, SciBite, VUA, UM, OGL, ME, UNIVIE

T.3.6 Secure hosting of eApps

The full potential of the eApps can only be exploited once they are hosted in a secure environment. This requires deploying them at the Open Link server and developing a secure access and data safety policy. Priority eApps for this implementation include PharmaTrek and ChemBioNavigator

Contributing Partners: OGL, PSMAR, BIT, UHAM, GSK, Lilly, ALM

Task distribution between TPL and PI

Every WP is led by a pair of closely collaborating people. The distribution of the leadership is made such that all the work with a long horizon is done by a scientific PI, and senior scientist from multiple partners. All the leadership work with a shorter horizon is performed by a TPL, a senior software engineer (also called *software integration engineer*). In particular,

The TPL:

- reports about the progress in the WP to colleagues in the ET
- consults the scientific programmers about possible technologies
- meets other TPL's to discuss gaps or overlap in efforts
- has weekly one-on-one calls with each of the scientific programmers in her/his WP to discuss progress
- has weekly one-on-one calls with the PI to discuss progress
- calls meetings as required
- structures the software development process
- makes sure each of the scientific programmers can work efficiently
- Has weekly one-on-one calls with the CTO and the CIO of the ET

The PI(s):

- Make sure the long term goals are clear and observed
- Keep a close eye on the technology produced from the eye of the user.

The ET CTO:

- Is finally responsible for the technical outcome of the distributed development process and for the time-alignment of software developments, hosting capacity and other performance related issues
- Meets weekly with the CIO to ensure the alignment and synchronization of the content in the Open PHACTS Discovery Platform with the vocabulary development and the services delivered for alpha testing
- Calls and chairs the various ET meetings
- Is involved in interviews and securing expertise balance of the growing software engineering group
- Reports to the EC of Open PHACTS on technical progress

The ET CIO:

- Is finally responsible for the content in the Open PHACTS Commons
- Meets weekly with the CTO to ensure the alignment and synchronization of the content in the Open PHACTS Discovery Platform with the vocabulary development and the services delivered for alpha testing
- Calls and chairs the ET meetings when purely related to content acquisition, quality control issues, curation and streaming.
- Is involved in interviews and securing expertise balance of the growing software engineering group
- Reports to the EC of Open PHACTS on content linking progress

Agile/Scrum methodologies in Software Development

Traditional development processes often result in late results, frequently are delivered over budget, and sometimes do not even deliver what was desired. We will avoid these problems by implementing Agile methodologies for our software development.

Traditional development processes, in hardware as well as software development, work with so-called stage-gates. Such processes first spend time getting a complete and detailed picture of the requirements of

the final tools, and then implement these in large chunks interleaved with go/no go decisions. The process of collecting the functional specifications is very hard and time consuming, because it has to be completed while the product is completely imaginary. In the most extreme installment, all functionality becomes available at once at the end of the project and no interaction about the interpretation of the functional specifications takes place with the prospective users during the technical implementation. In a worst case scenario this can result in the release of a product that is obsolete before it is complete; other projects may be late or over budget because of specifications that prove to be less important than originally imagined. This lack of iteration between prospective users and architects/engineers is arguably the most common reason for the graveyard of failed software projects.

In contrast: **Agile** software development methodologies aim to make the process much more flexible. With Agile, rather than spending a lot of time specifying the end result in great detail, the end result is only roughly specified at the beginning and refined along the way (hence the leading role of WP5 and 6). The process of working towards this goal is split up in very small tasks, that each will result in a tangible result that can immediately be shown to the users and will solicit feedback that helps to clarify the end goals. A team working with Agile software development works in short installments called "sprints" that typically last a few weeks. At the beginning of each sprint the developers decide together how much work can be handled. The frequent repetition of the sprint process means that this estimation becomes quite accurate over time. Short sprints ensure quick feedback to the people that ask for the new features. Very important in Agile methodologies is that each deliverable is always "thoroughly completed": the process does *not allow* cutting corners like delaying the creation of automated tests or making only preliminary documentation. All code is subjected to thorough automatic testing using standardized quality requirements. Focusing attention allows agile teams to show a better than average throughput, while delivering better than average code reliability.

Implementing Agile methodologies also means that the development process itself is monitored continuously, and subjected to continuous improvements that make it most suited to the people and the tasks at hand.

It will be obvious that combining Agile development with a widely distributed consortium setting poses specific additional managerial challenges, but several Open PHACTS partners have unique experience in this field.

The approach to develop a system such as the Open PHACTS Discovery Platform driven by concrete user scenarios is a well established one called 4+1 in software architecture (Figure 5) (http://en.wikipedia.org/wiki/4%2B1_Architectural_View_Model). The 4 views are used to describe the system from the viewpoint of different stakeholders, such as end-users, developers and project managers. The 4 views of the model are logical, development, process and physical view. Use cases or scenarios drive the architecture serving as the "plus one" view. In Open PHACTS the work stream 2 drug discovery services serve as the "plus one" view. The CTO and the CIO are finally responsible for securing that all elements are in sync.

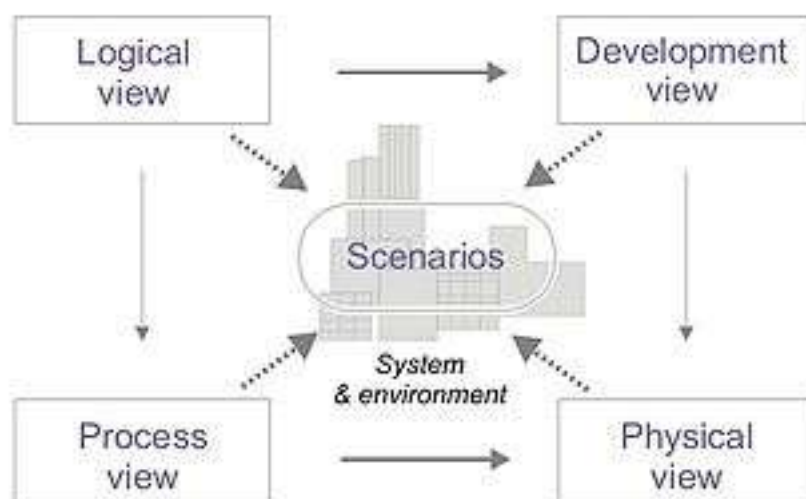


Figure 5: 4+1 Architecture

| | | | | | | | | | | | |
|-----------------------------|--|--------|------------------------------|---------|--------|--------|------|-------|------|--------|----------|
| Work package number | 4 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Linked data, content mapping, population | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 11 | 12 |
| Participant short name* | GSK | UNIVIE | DTU | UHAM | BIT | PSMAR | LUMC | RSC | VUA | UNIMAN | UM |
| Person-months | 12,7 | 3 | 15 | 1,5 | 2 | 13,8 | 11,5 | 62 | 43 | 19,2 | 35 |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | F | F |
| Work package number | 4 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Linked data, content mapping, population | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 13 | 14 | 15 | 16 | 17 | 18 | 21 | 22 | 23 | 24 | 25 |
| Participant short name* | ACK (till 17.09.2014) | USC | UBO | AZ | Pfizer | Esteve | HLU | Lilly | NBIC | SIB | EMBL-EBI |
| Person-months | 3,8 | 13 | 45 | 10 | 8,2 | 4 | 7 | 6,5 | 4,6 | 30,5 | 18 |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | F | F | F | IK | IK | IK | IK | IK | F | F | F |
| Work package number | 4 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Linked data, content mapping, population | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 27 | 28 | 30 | 31 | | | | | | | |
| Participant short name* | Janssen | OGL | ALM | SciBite | | | | | | | |
| Person-months | 3,5 | 17 | 0,9 | 0,1 | | | | | | | |
| Other resources (YES/NO) | NO | NO | NO | NO | | | | | | | |
| Funding claimed (F/ IK / N) | IK | F | IK | F | | | | | | | |

Preamble

This work package is focused on the delivery of the format and the content of the Open PHACTS Discovery Platform.

The Open PHACTS Discovery Platform will shift in the quality and scale of data integration to facilitate vast improvements in mining and analyzing data and information across both public and private resources. It therefore builds upon the architecture and services described in work packages 1, 2 and 3 and will be closely integrated with the core services described in WP5. As described in the general proposal, a general principle underpinning the tasks defined in each technical work package is the emphasis on delivery against a series of prioritized research questions posed by drug discovery scientists from a number of pharmaceutical companies. Further details on the prioritization process are given in WP5 and WP6. As the Open PHACTS community grows, additional research questions will be sought as outlined in WP8.

Objectives

The objectives of this work package can be broadly divided into three categories: 1) the analysis of the research questions and mapping them against a set of key data sources and production of an initial model for the Open PHACTS data (tasks 4.1 and 4.2) 2) the definition and establishment of the linked data cache (tasks 4.4, 4.5 and 4.6) and 3) addressing potential gaps and errors in the data using a variety of approaches such as data- and text mining, "nanopublications", curation and crowd sourcing of review and curation.

Specific objectives

O.4.1 To create a stable, high-performance content infrastructure to support Open PHACTS objectives by providing integrated access across multiple drug discovery and human protein-centric data sets

O.4.2 Deliver a technical foundation for the long term provision of the Open PHACTS data service during and after the completion of the IMI project

O.4.3 Facilitate additional data incorporation and provide clear avenues for extended community participation through the publication of Open PHACTS integration standards

O.4.4 Engage the broader scientific community to provide expert-assertions and curation/validation of content through Open PHACTS-based tools (strong link to WP8)

Description of the work

Tasks

T.4.1 Requirements gathering

T.4.1.1 Define the overall content requirements by cross-referencing EFPIA-prioritised research questions against existing data sources

T.4.1.2 Develop a coherent map of entities and assertional knowledge required to enable questions to be answered

T.4.1.3 Develop judgement criteria and expectations defining the overall quality and resolution of the selected data sources.

Based on the research questions supplied (WP5 and 6), we need to identify the available sources of information for inclusion into the Open PHACTS Discovery Platform. Side-by-side analysis of the content and questions should identify the entities that are involved in addressing the questions. Typical examples of entities that are to be represented and integrated in this system are genes, compounds, tissues, cells, disease etc. However, it will be necessary to consider a finer resolution: genes versus proteins versus transcripts, compounds versus drugs versus formulations etc., to ensure that current data are mapped to the correct granularity needed to achieve the overall objective. Where concepts/vocabularies or granularity levels are missing these will be defined and implemented in close collaboration with WP1. In addition to the various maps of entity types required, relationships between them will be mapped (e.g. bioactivity, protein interaction etc, aka "types" of database) together with contextual information (for example, species, cell type used in an experiment, concentrations etc). Coupled with an examination of the data quality within these various databases a decision matrix will be formed that will allow us to judge each content source for its applicability for incorporation; for example, an activity database that does not capture species or target properly may be unsuitable for Open PHACTS without further work by the database owner.

Contributing Partners: UHAM, BIT, PSMAR, LUMC, VUA, UNIMAN, USC, UBO, Pfizer, GSK, Lilly, Esteve, AZ, EMBL-EBI, Janssen, SIB

T.4.2 Content Analysis & Evaluation

T.4.2.1 Deliver a comprehensive review of the content sources based on the output from Task 4.1 including scoring against identified quality metrics

T.4.2.2 Produce an initial information model for the Open PHACTS data by analysing the core datasets with respect to the research questions to be answered. This model will provide guidance in terms of which vocabularies and representation schemes will drive the final Open PHACTS content recommendation.

T.4.2.3 Develop clear value propositions for the data owners including the provision of incentives to engage with the Open PHACTS project (with WP8), specifically demonstrated by inclusion of neXtProt data.

T.4.2.4 Engage with EFPIA partner members to obtain a full listing of available internal pharmaceutical resource data sets which could be consumed by Open PHACTS (with WP5 and 6)

T.4.2.5 Perform a gap analysis to identify and prioritise content for further analysis and possible inclusion in Task 4.3.

A rigorous set of criteria will be defined against which potential data sources can be judged and potential sources will be assessed. A preliminary list of potential key sources has been defined by members of the consortium and EFPIA members and this will be the starting point for analysis. Each source will be assessed not only in terms of entity mapping but also other factors such as data availability, service expectations and licensing details from the data owners. This list of core datasets, together with the research questions and the entity mapping will drive a self-consistent set of vocabularies and representation schemes that will feed directly into the Open PHACTS Commons (previously defined in the General section of this document). It is likely that some of these core data sources do not currently meet all of the requirements and a key part of T.4.2 will be to engage with the providers to encourage participation. This task will be done in conjunction with WP8 and might involve, for example, the provision of appropriate incentives to participate. Such incentives may include citation tracking methods described in WP2. In addition to existing public data sources it is anticipated that participating EFPIA companies will provide datasets that meet the general inclusion criteria and so mechanisms will need to be developed to capture such data as part of the Open PHACTS Commons. An important final deliverable from this task will be a gap analysis that will ensure the identification of missing data as a prioritised input to task 4.3.

Contributing Partners: UNIVIE, DTU, PSMAR, LUMC, RSC, UM, USC, UBO, Pfizer, GSK, HLU, Esteve, AZ, SIB, Janssen

T.4.3 Generation of Nanopublications from Information Sources

The objectives and deliverables of this task focus on the generation of nanopublications (for definition of this term see Annex 1) from both structured and unstructured information sources. The bias of the project is initially on integrating existing structured data sources and then filling in gaps between the resulting structured data LDC and the research questions as outlined in tasks 4.1 and 4.2.

Structured Information Sources (Public and Commercial)

The objectives and deliverables of this task focus on nanopublication generation from structured information sources.

T.4.3.1 Engage and consult with EFPIA partners and other domain members regarding existing practices and in-house tools utilized to perform data merging and mapping of data between public and commercial databases. Consortium member resources include ChemSpider (RSC), WOMBAT (DTU) and National Spanish and EU-OPENSOURCE libraries, assays and data (USC), HumLoc (DTU), RICORDO (EMBL-EBI) and ChEMBL (EMBL-EBI)

T.4.3.2 Based on the Content Evaluation Task 4.2 develop processes and procedures to merge appropriate slices or subsets of primary data source content into the LDC. Define a priority order for structured data source inclusion to be addressed during each time period

T.4.3.3 Define a pipeline for content extraction to nanopublications using existing and improved processes delivered by both consortium and EFPIA members. Expand the vocabulary services for post-translational modifications and subcellular location for nanopublications.

T.4.3.4 Define how database hosts can best format their database content for consumption by the pipeline. As a specific example, the SIB will develop and implement a pipeline for importing neXtProt posttranslational modification and subcellular locations into the LDC.

T.4.3.5 Define acceptable processes for linking to commercial databases (one partner, DTU, is a supplier of a commercial database and will be engaged to define acceptable processes)

T.4.3.6 Develop and implement a provenance strategy integrated with respect to nanopublications for content from original data sources.

The final goal of task 4.3 is to establish processes to integrate structured public domain and commercial databases to feed new nanopublications into the Open PHACTS Commons. A number of the partners involved have a proven track record in the field of data integration, chemical structure and biological data handling and information extraction, semantic text analysis in the life sciences. For example, RSC already hosts a public domain chemistry database of millions of chemical entities with associated identifiers, both validated and non-validated, and this will be one of the foundation datasets used as the basis for assertional mapping.

The priority data sources for extraction will be identified in task 4.2. Public domain and commercial chemical resources of interest to this project are generally available as either structure-metadata data sets (chemical structure collections with associated data) or as identifier-metadata sets (chemicals identified only by chemical identifiers with associated data). Biological sources of data contain text-based assertions between diseases, proteins, targets, drugs etc and mapping between chemical and biological space will be required. Processes will be established to merge and integrate structure-based databases using the chemical structures as the primary keys. A chemical structure database with associated identifiers will be used to map onto the concepts and facilitate the various sub-projects that will be pursued during the Open PHACTS project. The chemical repository will be built on the existing RSC-ChemSpider platform. The majority of chemicals have between one and many chemical identifiers associated with each chemical compound. The relationship between an individual chemical entity and its associated label(s) will be established using existing curated data sources and additionally by crowd sourced assertions (task 4.7, in conjunction with WP2 and WP8). The vocabularies governed by WP1 will be used in the context of WP4 for concept mapping.

Contributing Partners: PSMAR, LUMC, NBIC, RSC, VUA, UNIMAN, UM, USC, UBO, Pfizer, GSK, HLU, Esteve, AZ, SIB, EMBL-EBI, Janssen

T4.4 Unstructured Information Sources

T.4.4 The objectives and deliverables of this task focus on triple extraction from unstructured information sources.

Based on the Content Evaluation Task 4.2, list the unavailable information in the currently existing data sets that would be of value to research the research questions. Identify which unavailable information exists in text sources. Explicitly define the targets of extraction to be addressed during each time period.

T.4.4.1 Define a pipeline for text extraction to nanopublications using existing and improved text mining algorithms.

T.4.4.2 Integrate the pipeline with infrastructure leveraging components such as the IRS, Metadata Services and Citation Services.

T.4.4.3 Develop and implement a provenance strategy integrated with respect to nanopublications for automatically extracted content.

T.4.4.4 Interact with one or more journal publishers to obtain exemplar data slices and mine to generate nanopublications for the LDC. Publisher(s) to provide a report regarding the (citation) impact of provision of data for text-mining. As necessary, define potential business models with publishers for provision of nanopublications (collaboration with CrossRef).

T.4.4.5 Define how content authors can best format their textual content for consumption by the pipeline.

The main goal of task 4.4 is to establish a generic, robust, validated, distributed, extensible and easy-to-operate triple mining service from unstructured information sources which feeds new nanopublications into the Open PHACTS Linked Data Cache (OPS LDC), which is the 'under the hood' part of the Open PHACTS Commons. The partners UBO, NBIC, and CNIO have a proven track record in the field of information extraction and semantic text analysis in the life sciences. The consortium has taken part in public evaluation studies (BioCreAtivE, TREC Chemistry Track) and brings in technology developed by Open PHACTS and associated partners in EC- funded public sector projects such as CALBC and EU-ADR. All relevant and proven components will be brought together in the existing open framework for information extraction and management UIMA (<http://uima.apache.org/>). WP3 and particularly the TTF will assist in license and IP issues related to background technology brought in by associated partners.

Contributing Partners: PSMAR, LUMC, NBIC, RSC, VUA, UNIMAN, UM, USC, UBO, Pfizer, GSK, HLU, Esteve, AZ, EMBL-EBI

T.4.5 Infrastructure Build

The general architecture of the Open PHACTS Discovery Platform will have many content related aspects. The partners in WP4 will contribute key insights and requirements to the TTF (WP3) in this respect.

T.4.5.1 Infrastructure Design part 1: Help the TTF to create an architecture road map for the Open PHACTS LDC and associated services. This will include the information model and review of existing standards for managing the assertional content and describe how these will be used and augmented as required.

T.4.5.2 Infrastructure Design part 2: Define data integration methodology, citing expectations from WP1 and the data providers. Identify any hurdles limiting data integration and devise mechanisms to address these. Provide clear Open PHACTS interoperability standards to both the pharmaceutical and the wider scientific community that enable integration

T.4.5.3 Deployment strategy part 1: Define test and validation criteria, as well as appropriate metrics to test (performance, scalability etc). Identify and clarify all consumer expectations (e.g. number of concurrent users the Open PHACTS Discovery Platform should handle with no performance impact). OpenLink Software will define a list of scalability and correctness tests for hosting of the Open PHACTS Discovery Platform, based on the data and applications produced by the project at each stage, spanning the duration of the project.

T.4.5.4 Deployment strategy part 2: Identify the hardware requirements and physical provider of the computing environment required to run the Open PHACTS linked data cache. Within this analysis, clearly describe the anticipated needs foreseen for future development of the system

T.4.5.5 Create and communicate a deployment strategy that incorporates pharma-internalisation and security considerations. Provide mechanisms for both query and analysis submission that align with the confidentiality requirements of EFPIA member partners

The Open PHACTS Linked Data Cache is the central mediator between the external resources that Open PHACTS will integrate and the services that sit on top of that infrastructure. It provides an important buffer between the services and the underlying data ensuring a consistent data model as well high-performance and secure operation. More specifically, the LDC is a semantically-enabled data store providing a means of querying over integrated data sets on a transient basis. This data store is augmented by several services:

- Notification and crawling services allowing the infrastructure to proactively retrieve known external datasets so users can be informed when external data sets are updated.
- Reasoning service provides both the query engine and extraction services the capability to infer or derive new connections between data already residing within the cache leveraging the semantics of nanopublications. This capability is essential for realizing business cases such as constructing target dossiers and connecting compounds and diseases.
- Server Side Analysis provides a mechanism to allow complex analyses to run on site with the LDC ensuring high performance execution.
- Policy and Security Management services allow systems to ensure that their analysis and queries are run both securely and within a given users requirements. For example, specifying whether queries must remain within the LDC or can be sent to external data sources.

Thus, the final aim of Task 4.5 is to provide both a design for the cache and an implementation strategy for its rollout. Both the design and strategy must meet a number of key goals, namely, scalability, security, and standards compliance. There are several desiderata that are core to the Open PHACTS approach that will be considered when designing the Linked Data Cache:

- To ensure that the LDC is an integration point and not a permanent data repository, the cache should be able to recreate its entire contents using only external data sources and Open PHACTS services
- The LDC should leverage already existing semantic technologies. For example, building on large scale triple stores such as OWLIM and associated reasoners like the LarKC platform
- The LDC must provide clear interfaces to outside parties to enable easy integration of new data sources. Examples include integration to EU-OPENSOURCE.

As part of Task 4.5.1, it is necessary to have a clear information model and assertional content/nanopublication model as a foundation for development. These models will reflect the data source selection in Task 4.2, the annotation model defined in 4.7 and the vocabularies specified in WP1. The information model will build on existing standards and where no suitable standards exist the task will define the needs, resources and time lines required to develop these standards. The information and nanopublication model will provide the locus around which a data integration methodology will be specified. Such a methodology will define exactly how a data source can be integrated with the proposed platform based on a set of widely deployed community standards. In particular, the methodology should leverage existing Web standards. A key part of both the nanopublication model and the associated integration methodology is the ability to track the provenance or source of every nanopublication. This will allow services to assign evidence or quality levels to nanopublications. The integration methodology will specify the approach to defining and generating such evidence levels. Such evidence level production will require batch (e.g. weekly) reasoning to provide such quality measures.

The design of the infrastructure along with the requirements defined in Task 4.2, will impose certain metrics upon the software and the hardware environment it runs on. In Task 4.5.2, clear test and validation criteria will be defined along with specific requirements. We envision that this will require a sophisticated environment to ensure quality software that can manage the billions of nanopublications involved.

A graph level security model will be implemented, with data owners controlling access to their graphs. The data can be granted to specific roles in the LDC database. The roles and grants are managed via a SQL interface.

Contributing Partners: LUMC, NBIC, SIB, RSC, VUA, UM, UBO, Pfizer, Esteve, AZ, EMBL-EBI, OGL

T.4.6 Instantiation of the Linked Data Cache

T.4.6.1 Implement the Open PHACTS Linked Data Cache, using processes and approaches derived from the work defined above, to produce a stable and high performance framework for the broader Open PHACTS program connecting with the IRS in Task 4.1. The Virtuoso RDF store is tested to function with all Open PHACTS applications and datasets.

T.4.6.2 Deliver quality control systems to ensure that data can be tracked back to its original source and quality control can be maintained. The source data will be identified and a script will be implemented for reconstituting the LDC from source data. Metrics of data relevance will be defined, for example citation scores for nanopublications. Processes for database resident computation of these metrics will be implemented. A set of test scripts will be created for verifying correct loading and function of queries and applications. These are applied when either the published content or before a new version of the LDC software enters user testing or production. This set of tests can also measure the performance of the Open PHACTS Discovery Platform with different equipment and server configuration, e.g. single server vs. cluster.

T.4.6.3 Deliver human interfaces for workflow management, data integrity checking, data exploration, modification and general administration

T.4.6.4 Deliver application programming, query and analysis interface to the Open PHACTS linked data cache (with WP2). Ensure that the core application programming interface functions fulfil the needs of WP5 and WP6. Programming interfaces will be tested and new REST or SQL interfaces may be offered for supporting the Open PHACTS content production pipeline and applications.

T.4.6.5 Deliver a data source integration framework for external data sources both in terms of iterative crawling and notification. An incremental data reloading facility will be tested, allowing a dataset to be replaced while online, without a need for full delete plus reload. Users may have notification needs expressed as queries. The queries and their previous results will be stored in the Open PHACTS Discovery Platform and the queries will be re-evaluated after content refresh and users will be notified if the results have changed.

T.4.6.6 Iteratively integrate new data sources following the data integration methodology defined in Task 4.5.1 and make use of the prioritized list of data sources given in Task 4.2. A set of reports will be generated for tracking vocabulary utilization and possibly checking data integrity. The criteria to check for will be specified by the data owners.

T.4.6.7 Apply the data integration standards defined in Task 1.4 and information model in Task 4.5.1 to ensure all data within the Open PHACTS framework is compliant with a set of defined vocabulary standards

T.4.6.8 Advise on service-level agreements (SLAs) regarding how to respond to user feedback regarding data corruption, quality issues (WP3)

T.4.6.9 Develop application-driven query optimizations. Open PHACTS expects to perform complex analytics on large data volumes, where obtaining the right query plan is essential for performance. This guarantees that the queries produced by users or applications are properly optimized for best performance.

T.4.6.10 Custom development - This task expands the inference capabilities in Virtuoso so as to allow using an arbitrary predicate in the meaning of owl:sameAs and to scope such "sameAs" triples to

specified sets of graphs. Other alterations to database resident inference capabilities may be introduced as the project requires and the budget permits.

Task 4.6 consists of developing the LDC and services and then ingesting and making available new data sources. This development activity will follow the Deployment Strategy defined WP3. The task will follow the Open PHACTS phases incrementally adding new functionality and ingesting new data sources throughout the project lifetime. The aim will always be to have a fully functional version of the Linked Data Cache available at all times following the Open PHACTS working to working principle. Development will adhere to best practices including functional and unit testing as well as live deployment.

Importantly the development will feed back into Task 4.5 and the overall design and requirements ensuring that design can adjust to experience. Beyond ensuring a high-performance Linked Data Cache, this task must concentrate on developing high-quality well documented integration points. The first integration point is the data source integration framework, which will be used by both outside parties and Open PHACTS members to connect to the infrastructure. The second integration point will be the application programming, query and analysis interface, which is essential for supporting the core drug discovery services. Integration points will reflect extensive interaction with these external developers.

Contributing Partners: LUMC, RSC, VUA, UBO, Pfizer, Lilly, HLU, Esteve, AZ, OGL

T.4.7 Running the Linked Data Cache

Running a Linked Data Cache including provenance and redundancy features is a novel challenge and therefore this WP (and effectively all relevant LarKC and W3C consortium partners), will have a key role in advising and assisting the WP3 TTF and ET in implementation issues such as:

- Develop and deliver to WP7 a comprehensive maintenance plan for both the cache and the content
- Define the points of accountability, request access points (e.g. a “helpdesk” function), SLAs, downtime procedures and so on to WP3.
- Develop comprehensive change management procedures for both system maintenance and content changes. Clearly define the decision-making process for accepting such changes (controlled through governance via WP3).
- Develop communication mechanisms to inform consumers of changes and perceived impacts (with WP8)
- Develop a communication plan for roll out of the system to both the project members and the wider community (with WP8).
- Produce an expert-level user guide to the infrastructure, with exemplar queries and access mechanisms to allow consumers to explore utility (with WP5,6,8)
- Develop backup and failsafe approaches. For example, maintaining two concurrent production systems to provide backup when the main system down for maintenance (WP3)
- Develop automated reporting mechanisms. Examples include 1) a stats dashboard to look at server load etc. 2) a public stats board to allow users to see current numbers of entities, schemas, usage etc. and 3) an administration dashboard (with WP3)
- The Open PHACTS Discovery Platform will be hosted in two copies, with one copy reflecting approved content and tested applications and a second copy under development, serving as a staging area. When the staging area is considered mature in terms of both content and software, it is copied over the production version. This will capture user generated content stored in the production version, e.g. user account settings, security, notifications etc. The staging area will then be used as a development system for new content and/or software functionality.

- Data statistics, e.g. VOID graphs will be generated following load of new data. A web page summarizing the system status will be made available to users. Virtuoso has a query logging capability which will be enabled by default, providing a full record of all executed queries, complete with query plans and extensive performance statistics. This will serve for performance diagnostics as well as a basis for a periodic utilization report. The exact content of the utilization report will be specified as a part of this task.
- A period backup facility with streaming of backups into external cloud storage will be implemented. The backups will be periodically restored at the remote cloud location to verify their integrity.

Contributing Partners: RSC, VUA, Pfizer, Lilly, Esteve, AZ, OGL

T.4.8 User interaction/annotation

T.4.8.1 Research and evaluate curation and annotation models and provide an optimal strategy for implementation into the Open PHACTS Discovery Platform. Specifically, engage EFPIA partners and other domain members (e.g. EBI, PDB and others) and review their existing internal tools and approaches for examining data quality. Utilize best-in-class approaches to assist in system design

T.4.8.2 Implement a curation/annotation system that allows the community to provide machine-consumable feedback across the Open PHACTS data. As a specific application of the annotation system, an appropriate interface for describing posttranslational modifications and subcellular location will be developed. Identify the needs of security, privacy and legal issues and ensure incorporation into the system (with WP2)

T.4.8.3 Manually populate the nanopublication store using the annotation system with drug discovery protein centric data sets. Develop mechanisms to assess human annotation content and investigate quality metrics for annotations entered into the system.

Existing data in public data sources are known to be of variable quality. With regard to this Open PHACTS project errors in data feeding the LDC will commonly include the miss-association of entities (compounds, sequences, etc.) with data, annotations and assertions. The application of text-mining approaches as outlined in Task 4.3.2 will also contribute imperfect data as a result of the applied technologies. Ongoing efforts in terms of automated and manual quality control during the process of data aggregation to form the LDC will be necessary. The majority of primary data sources feeding the LDC do not allow edits or modifications to the original data. It is therefore necessary to allow users of the Open PHACTS Discovery Platform to incorporate annotations/assertions to the LDC.

Quality control systems will be developed in parallel with the efforts of Task 4.3. Standard processes for loading data into the Open PHACTS Discovery Platform will be developed and will include approaches for entity validation (for example, checking structures with chemistry rules, validating consistency between entities and identifiers using look-up tables and software tools, etc.). Standard operating procedures (SOPs) will be defined and the outcome of different approaches will be reported. Recognition systems in the form of micro-attribution and citation will be developed to encourage and acknowledge the contributions of Open PHACTS users who annotate the LDC. The result will be a citation index for the nanopublications contained within the LDC and will be managed through WP8.

Contributing Partners: UNIVIE, PSMAR, LUMC, RSC, UNIMAN, UM, ACK (till 17.09.2014), USC, UBO, Pfizer, GSK, Lilly, Esteve, AZ, Janssen, SIB

T.4.9. Chemistry registry for proprietary molecules

The chemistry registry service is run via ChemSpider. However, this requires calling a web-service run by RSC, which is impossible for molecules from commercial sources and EFPIA in house compounds. This

task will explore and develop solutions for allowing the integration of commercial and confidential molecules into the Open PHACTS Discovery Platform

Contributing Partners: RSC, UNIMAN, OGL, SciBite, GSK, Janssen

| | | | | | | | | | | | |
|-----------------------------|---|----------|------------------------------|-------|-----|----------|---------|------|---------|----|--------|
| Work package number | 5 | | Start date or starting event | | | | | | | M6 | |
| Work package title | Open PHACTS Services for drug discovery | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 1 | 2 | 3 | 4 | 5 | 6 | 9 | 10 | 14 | 16 | 17 |
| Participant short name* | GSK | UNIVIE | DTU | UHAM | BIT | PSMAR | VUA | CNIO | USC | AZ | Pfizer |
| Person-months | 17,2 | 30,2 | 65 | 96 | 80 | 96,8 | 8 | 82 | 14 | 30 | 4,6 |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | IK | IK |
| Work package number | 5 | | Start date or starting event | | | | | | | M6 | |
| Work package title | Open PHACTS Services for drug discovery | | | | | | | | | | |
| Activity Type* | RTD | | | | | | | | | | |
| Participant number | 18 | 19 | 20 | 22 | 24 | 25 | 27 | 30 | 31 | | |
| Participant short name* | Esteve | Novartis | ME | Lilly | SIB | ConnDisc | Janssen | ALM | SciBite | | |
| Person-months | 6,6 | 9 | 21 | 4 | 2 | 3 | 11 | 3 | 0,1 | | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | | |
| Funding claimed (F/ IK / N) | IK | IK | IK | IK | F | F | IK | IK | F | | |

Preamble

The ability of the Open PHACTS framework to deliver practical answers to drug discovery research questions by facilitating a powerful and user-friendly access to the integrated chemical and biological resources is a key challenge for this project. On the other hand, the implementation of third-party services on top of the Open PHACTS Discovery Platform will be a key success measure.

There is room for the development of proprietary exemplar services taking data feeds from the Open PHACTS Discovery Platform. This will be explored in the development of work stream 2 (WS2). We also expect to focus funding in WS2 on academic partners who wish to develop innovative exemplar services (example applications – eApps) that further enhance the public drug discovery toolset. A key feature of these latter examples will be the provision of facile expert-level analyses to non-expert scientists, and drive the discovery of new potential targets, or new uses for drugs.

Clearly the widespread adoption of Open PHACTS standards will define the success of this project and the implementation of third-party services on top of the Open PHACTS Discovery Platform will be a key success-measure. This includes commercial offerings and we envisage several business models. For example, a publishing house could offer high quality assertions as a value added service within the Open PHACTS framework, high-quality reasoning, visualization and query-builders interacting with the Open PHACTS Discovery Platform could be developed by start-ups or SMEs and the Open PHACTS framework could be used to semantically enrich documents and literature to subscribers.

Objectives

Develop a set of core services to answer drug discovery research questions for public and private drug discovery research and demonstrate the usage and utility of the Open PHACTS Discovery Platform.

Specific objectives

O.5.1 Develop a “Chem-Bio Navigator” to facilitate the querying and visualization of sets of biologically annotated small molecules, on the basis of chemical substructures, pharmacophores, biological activities, etc.

O.5.2 Deliver comprehensive *in silico* dossiers about targets showing pharmacological interest, incorporating related information on sequences, structures, pathways, diseases and small molecules.

O.5.3 Deliver a “Polypharmacology browser” to map the coverage of the chemo-biological space, allowing the detection of significant gaps and facilitate the polypharmacological profiling of small molecules.

O.5.4 Deliver a 'Scientific Article Reader' capable of identifying concepts from the Open PHACTS Discovery Platform, and linking these to their source data. Additionally, the tool will allow the recovery of small molecules and 'Markush-like' structures (i.e. combinations of 'skeleton' molecules with wildcard components, and substituents and their associated properties in associated tables) from the scientific literature, for deposition as nanopublications back into the Open PHACTS Discovery Platform.

O.5.5 Develop tools that allow the automatic extraction of annotated series of compounds, suitable for carrying out Structure-Activity Studies, that could be integrated with *in silico* prediction systems.

Description of work

In line with the fundamental principles outlined in WP3 the development of these services will proceed through a series of increasingly capable releases following the release plan outlined in WP6. The services will dynamically access and integrate data from the underlying core sources through the Open PHACTS semantic layer and deliver this in a research context to answer increasingly complex drug-discovery questions.

Tasks

T.5.1 Target Dossier

Building on the Open PHACTS Discovery Platform the target dossier will build a comprehensive view of pharmacologically relevant targets to answer questions regarding druggability, tissue expression profiles and implications in pathways, disease states and physiological mechanisms. Additionally the Target Dossier will incorporate views on assays, small molecule interactions as well as other therapeutic modalities with the objective to provide a decision support platform for target selection and progression.

The basic target annotation workflow will be complemented with additional methods to provide information of specific interest for the target proteins regarding binding and interaction sites, including the annotation of known binding sites in protein structures and the corresponding ligands. The relation between proteins and pathways will be extracted from open metabolic pathways following standards in the field (BioPAX, Biological Pathway Exchange format (<http://www.biopax.org/>) used by different pathway databases such as Reactome (<http://www.reactome.org/>) or BioCyc (<http://biocyc.org/>).

To complement the annotated pathways and to increase the level of the annotations for target proteins outside canonical pathways we will incorporate pathway extension methods based on the information provided by curated protein interaction networks (see for example Baudot *et al.*, 2010 <http://www.infobiotics.org/pathexpand>, or <http://wikipathways/pathvisio/cytoscape>) .

The annotations of the targets with the associated drugs will be mined from the available databases in the Open PHACTS data store (ChEMBL, PubChem and the Comparative Toxicogenomics Database (CTD) (Wiegiers *et al.* 2009, Davis *et al.*, 2009 Mattingly *et al.*, 2006), DrugBank (Wishart *et al.* 2006)

and others) (WP4). To ensure the delivery of an intuitive system for use by experimental biologists and chemists they will be engaged to garner their views in the development of the corpus of texts containing semantic annotations of the entities of interest and for the internal checking and validation of these annotations taking into account that for the evaluation of those data not only direct interactions will be extracted, but indirect associations between targets and compounds from bio-assay activity data will be used as well.

The drug – target interaction will be made as explicit as possible (physical binding, measured affinities, off-rates, etc.). The expression of the target transcript(s) as well as protein(s) in healthy tissue and in disease states will be extracted from public gene expression (GEO, ArrayExpress) and protein expression (ProteinAtlas) repositories (WP4). We aim to integrate gene/protein expression differences in healthy and disease tissue across different expression experiments in order to derive target expression - disease associations.

Contributing Partners: UNIVIE, DTU, PSMAR, CNIO, USC, Pfizer, GSK, Esteve, Novartis, AZ, ME, Janssen

T.5.2 Chem/bio space navigator

The aim of the Chem-Bio-Navigator (CBN) is to provide a compound-centric view to the Open PHACTS data resource. This will address the manifold research questions that arise within small molecule drug-discovery research, for example:

- For a given compound, show all 'similar compounds' and their biological activities
- Show all oxido-reductase inhibitors active at <100nM in both human and mouse
- Compile a Kv channel opener set. Retrieve all bioactivity data for these compounds.
- Compound A is a P2X4 antagonist. Retrieve all bioactivity data in P2X electrophysiology full-curve assays for molecules that have a similarity > 0.8 to compound A.
- Retrieve all experimental data for molecules that have been tested in Factor Xa and Thrombin assays grouped by the response and assay type.

All of these questions have in common that molecules must be retrieved from the query engine and should be reported back to the user in a user-friendly display in the context of chemical and biological data. The development of the CBN will proceed in an iterative process guided by on-going user-feedback and the research questions developed in WP6.

The central technology components of the CBN can be split into a web-based front-end and visualization and chemical compute engines. The web-based front-end will be developed within a flexible framework for the comprehensive and intuitive navigation and visualization in high-dimensional data spaces (WP2). The CBN will also allow for the calculation and display of key chemical property descriptors. It will offer the capabilities of query creation, including chemistry queries, and will be linked to the scalar (quantitative) data in the Open PHACTS linked data store. The query creation engine will be based on the IRS and vocabularies developed in WP1 and build on all data and information resources in Open PHACTS (WP4). Moreover, the CBN will be seamlessly connected to the general Open PHACTS GUI (Open PHACTS Explorer, WP2).

Contributing Partners: UNIVIE, DTU, UHAM, BIT, PSMAR, Pfizer, GSK, Esteve, Novartis, AZ, ME, Janssen

T.5.3 Polypharmacology browser

The central tenet of rational drug design, i.e., one drug selectively interacts with one target, leading to therapeutic alterations for one disease, is increasingly challenged by the unprecedented volumes of high

quality data documenting experimental bioactivities of drugs and their metabolites on multiple targets, i.e., polypharmacology. Two thirds of the FDA approved drugs are currently believed to interact with 2 targets or more, half of them interacting with more than 7 targets.

Comprehensive strategies to facilitate the polypharmacology profiling of small molecules will be developed on top of the linked data store (WP4). Chemical-target association information, i.e. targets that may play a role in modulating drug responses will be also integrated in the polypharmacology views to display high-quality relational data comprising chemical structures, bioactivities and target information.

We will also extend chemical-target annotations with *in silico* predictions, starting with features computed from chemical structures. Diverse sets of descriptors such as physicochemical descriptors, circular fingerprints, pharmacophore (2D and 3D) features and the similarity ensemble approach (SEA) will be implemented and comparatively evaluated in order to assess and enrich the profile of a chemical.

Finally, a functional link between the small molecule chemical space and the protein space associated with disease (defined as a Systems Chemical Disease Network) will emerge as a new way to address drug action across multiple levels of complexity such as molecular and cellular levels to tissue and biological systems levels. Via protein-protein interaction (PPI) networks, we can better understand the properties of biological systems with respect to the link between drug action and disease susceptibility genes, thus contribute to the development of systems pharmacology.

These types of approaches can improve the *in silico* evaluation of approved drugs for repurposing, as well as our ability to select new chemicals based on many-target estimates, including estimates related to anti-targets and adverse drug events. Such systemic evaluation may be critical for the identification of additional targets that may play a role in modulating drug responses, which could lead to new therapeutic options in drug discovery.

To ensure the development of an intuitive system both experimental biologists and chemists will be involved as outlined in task 5.1. The survey results of the research questions and sources will also be incorporated into the development of the system.

Contributing Partners: UNIVIE, DTU, UHAM, PSMAR, CNIO, USC, Pfizer, GSK, Esteve, Novartis, AZ, ME, Janssen

T.5.4 Utopia Documents

As the pace of publication accelerates, the need for computational support to identify and analyse this information grows relentlessly. In spite of the proliferation of biological databases and ontologies, much of the knowledge in the life sciences remains buried in prose and figures in the scientific literature, and at present it can only be recovered through the manual labour of experts. Currently, individual scientists regularly annotate, extract data from and organize their personal collection of papers for their own purposes, and in the process apply intuition and specialist knowledge that far surpasses that of any automated techniques. These insights however, are invisible to the broader community, locked in unstructured notes, spreadsheets, or collections of PDFs that are only accessible or interpretable by their creator. The cost of sharing these micro-insights or fragments of extracted data in an organised way is currently prohibitive. Thus the community is currently squandering valuable resource, effort and knowledge by repeating tedious and error prone tasks such as manually re-keying data from tables or re-drawing molecules from diagrams. By creating what is essentially a crowd-sourcing platform that makes performing these daily activities easier in itself, this task aims to resurrect the knowledge currently buried in the literature to provide a shared resource that can contribute to hypothesis generation, the creation of systematic reviews and which itself enhances and makes more efficient the experience of reading the biological literature.

For example, a biochemist is interested in how a range of small molecules interact with a key cellular protein to regulate its activity and in particular building structure activity relationship (SAR) models that

allow him to predict the relationship between chemical structure and inhibitory activity on that protein. He reads articles with Utopia Documents and finds an experiment reporting interesting data for a novel molecule that appears to regulate the protein he's interested in. A previous reader has extracted the table of data describing the interaction between small molecule x and protein y. He is able to take this validated data table and add it to his repository of SAR data that he will later use in his analysis. As he reads other papers in the same area, he finds other molecule structures (either drawn explicitly or represented in Markush-like form) and accompanying SAR tables, this time not already extracted. He highlights the molecule's structure and table using Utopia Documents, adjusts a few elements and adds it to his corpus, extracting both the molecular structures and the accompanying biological data. He continues to read papers, extracting such data to build a rich dataset for his molecular modelling work. A side-effect of this activity is that his work is available for others to use.

Contributing Partners: UNIMAN, UNIVIE, AZ

T.5.5 eTox Collector

A common task in drug development is the compilation of series of compounds, aiming to understand what are the properties of the compounds that determine their biological behaviour. Such studies, usually called Structure-Activity Relationships (SAR) are applied at different steps of the drug development, from the lead optimization (in order to optimize potency) to pre-clinical drug safety studies (in order to avoid adverse effects). Irrespectively of the biological property of the compound under investigation, the quality of the results depends critically on the series that is used for the analysis; the collection of compounds must be as large as possible, contain drug-like structures that cover a relevant section of the chemical space, and the distribution of the studied biological property must be balanced.

The growing availability of databases of annotated compounds allows the compilation of such collections from diverse open sources, even if the task is difficult, tedious and prone to errors. For this reason, we will develop a specific tool for carrying out this task, (Collector) exploiting the vast amount of target-ligand data of the Open PHACTS framework (WP4). Moreover, the development of this tool will allow establishing connections with the IMI project eTOX, which aims to develop *in silico* tools for the prediction of *in vivo* toxicity endpoints of drug candidates. By focusing on the needs of a concrete project, we will have the opportunity to develop our tool in close contact with the end users, producing software that meets their real requirements, making also optimal use of the resources in two different IMI projects.

Collector software will be designed to extract series of annotated compounds in a semi-automatic manner, populating a local database with updated series of compounds, which will be filtered according to user-defined criteria. Collector will be integrated with other WP5 tools like CBN, for visualizing the extracted series, but also it will be tightly integrated with the eTOX prediction systems that will be used as model of Collector exploitation in real pharmaceutical environments.

Contributing Partners: UNIVIE, DTU, UHAM, PSMAR, CNIO, USC, Pfizer, GSK, Esteve, Novartis, AZ, ME, Janssen

T.5.6 Develop scientific workflows

Workflow engines such as Pipeline Pilot, KNIME, and Taverna are powerful tools for performing complex tasks. To fully exploit the combination of the Open PHACTS Discovery Platform and workflows requires both the development of new nodes as well as the implementation of new API calls. These developments will be driven by new scientific use cases created in researchathons.

Contributing Partners: VUA, UNIVIE, PSMAR, SIB, AZ, GSK, Esteve, ME, Lilly, Janssen, ALM

T.5.7 Develop target validation workflows

In this task we will identify queries that combine our different databases in order to support the linkage of a target to a disease or potential side-effects. The Open PHACTS Discovery Platform is in a unique position to deliver this type of analysis direct to scientists via a simple API or workflow. We will provide examples and pre-packaged workflows that execute these queries for a target or disease of interest.

Contributing Partners: SciBite, UNIVIE, VUA, PSMAR, SIB, CD, GSK, Esteve, Lilly, Janssen, ALM

| | | | | | | | | | | | | |
|-----------------------------|---------------------|--------|------------------------------|---------|----------|-------|-------|------|-----------|----------|---------|--|
| Work package number | 6 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Exploitation pilots | | | | | | | | | | | |
| Activity Type* | RDG | | | | | | | | | | | |
| Participant number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 11 | 12 | |
| Participant short name* | GSK | UNIVIE | DTU | UHAM | BIT | PSMAR | LUMC | RSC | VUA | UNIMAN | UM | |
| Person-months | 7,5 | 58 | 18 | 8,2 | 6,7 | 16,4 | 3 | 27,5 | 2 | 24 | 2 | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | F | F | |
| Work package number | 6 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Exploitation pilots | | | | | | | | | | | |
| Activity Type* | RDG | | | | | | | | | | | |
| Participant number | 14 | 16 | 17 | 18 | 19 | 20 | 22 | 24 | 25 | 26 | 27 | |
| Participant short name* | USC | AZ | Pfizer | Esteve | Novartis | ME | Lilly | SIB | Conn Disc | EMBL-EBI | Janssen | |
| Person-months | 20 | 6 | 4,6 | 1 | 6 | 1 | 3,5 | 3 | 3 | 6 | 3 | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | F | IK | IK | IK | IK | IK | IK | F | F | F | IK | |
| Work package number | 6 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Exploitation pilots | | | | | | | | | | | |
| Activity Type* | RDG | | | | | | | | | | | |
| Participant number | 28 | 29 | 30 | 31 | | | | | | | | |
| Participant short name* | OGL | OPF | ALM | SciBite | | | | | | | | |
| Person-months | 1,5 | 0,5 | 1,2 | 0,1 | | | | | | | | |
| Other resources (YES/NO) | NO | NO | NO | NO | | | | | | | | |
| Funding claimed (F/ IK / N) | F | F | IK | F | | | | | | | | |

Preamble

The Open PHACTS Discovery Platform is a large complex system built on distributed resources and utilising data provided by numerous sources. The objective of this work package is to ensure the delivery and alignment of key functionality, which will allow exploitation of the Open PHACTS Discovery Platform in order to answer key research questions important to both industry and academia. The design, development and deployment of the Open PHACTS Discovery Platform is organised around “vertical slices” through the application services. These will initially be rudimentary in nature start simply and become increasingly more sophisticated, incorporating more services, richer capabilities of those services and more content. The milestones of the project are based on these vertical slices.

By being exemplary application driven the utility and potential of the Open PHACTS Discovery Platform will be revealed early, continuously scrutinized and steered as appropriate. Scientific (and technical) users and partners can be brought on board as early as possible thus raising awareness in the scientific community and encouraging academic groups, publishers, and both public and commercial databases as well as SMEs

to provide or integrate their content and applications and to explore the capabilities of the Open PHACTS Discovery Platform. This will not only foster community engagement (WP8), but also provide user feedback. As a result the technical team (WP3) will be able to obtain concrete requirements and define the system around realistic needs. Furthermore, this approach will also ensure that the services interoperate to achieve specific goals.

Objectives

The Open PHACTS project will be delivered through a series of 6-monthly demonstrators centered around key research issues for drug discovery. This process is application and user driven and will start simply and become increasingly more sophisticated, incorporating more services, richer capabilities of those services and more content. This will also guide the selection and development of vocabularies, the linking of priority data-sources and mappings to the knowledge model. Finally, a complete end-to-end system will be delivered that will have the capability of answering the research questions posed by drug discovery scientists.

The objectives of this work-package is

6.1 Develop and refine a series of drug-discovery questions to guide the technical requirements, prioritize the development of vocabularies and steer the linking of data-sources.

6.2 Coordinate a series of documented research questions to articulate a road-map for Open PHACTS capability through the release cycles and plan for documentation and community engagement

6.3 Deliver a series of case-studies reporting innovative usage of linked drug-discovery data to demonstrate the usage and impact of Open PHACTS.

Description of work

Tasks

T.6.1 Key requirements for each consecutive release

The objective of this task is to articulate the key requirement for each consecutive release in details through an on-going iterative process with milestone review of the full capabilities of the release at regular intervals. Based on an on-going review of drug-discovery issues this task will articulate the use-cases that drive the development of a fully integrated Open PHACTS Discovery Platform. Use cases will be defined via an internal survey (surveys, consensus meetings, etc.) on research questions, which will be categorised into a series of specific streams, such as, e.g. target dossier, compound/target relationship, or compound profiling. These use-cases will guide priorities and selection of the underlying technology, the data and vocabularies for the incremental delivery of pilot services.

This task will consist of the definition of the methods and logistics for the participation of project partners and users into the in-process testing activities as well as their practical execution. The data collected will be systematically analysed in order to transform them into feedback for the drug discovery pilots improvement. As a result, based on the on-going exploitation of the Open PHACTS framework by academic and industrial partners and associate partners detailed requirements and plans for the next phase will continuously be updated and translated into action plans and scientific programmes for the Open PHACTS workshops. The output will be in the form of short reports or white-papers for the Open PHACTS programme.

Contributing Partners: UNIVIE, UHAM, PSMAR, RSC, USC, Pfizer, GSK, Lilly, Novartis, AZ, ME, EMBL-EBI, Janssen

T.6.2 Drug discovery pilots

This task will deliver a set of Drug Discovery Pilots (DDPs) including handbooks and tutorials on top of the Open PHACTS framework. This will help to define the necessary technical requirements for practical research use. Furthermore, the drug discovery pilot scenarios are the main vehicles to attract the scientific community and to motivate academic labs to use the system. This will also encourage academic groups as well as content providers to link their in house data for use by the Open PHACTS community. DDPs will be built incrementally, starting by delivering simple browser functionalities, target dossiers and compound similarity searches. Upon progress of the project, the pilots will become increasingly complex, thus allowing more holistic research questions to be addressed such as: *“list all compounds reported to induce cholestasis and provide their activity values at transporters expressed in the liver”* and *“identify all compounds linked to increased oxygen consumption in rats and have reported nanomolar activity towards a GPCR expressed in the brain”*. DDPs will be released internally 6 months prior to public release for extensive expert alpha testing, quality control and bug fixing. Public release will include manuals and tutorials and will be accompanied by strong dissemination activities (WP8) utilising for instance the manifold capabilities of the European Federation for Medicinal Chemistry and EFPIA. 'External' users will also be asked for feedback on the quality of the services as well as the quality of the data provided by Open PHACTS. The following general time line is proposed:

Phase 1 (6 months) will start with a “lash up” integration of available services for WP1 WP2 and WP4 and will be followed by first use of the Open PHACTS Discovery Platform and the Open PHACTS Explorer. This comprises simple browser functionality, a primitive dossier and queries for a few research questions.

Phase 2: (12 months): First generation of services to demonstrate vertical integration of the Open PHACTS framework and architecture. Identification of performance issues/bottlenecks to address for optimal user experience.

Phase 3: (18 months): First release of a targeted exemplar Open PHACTS Discovery Platform to the wider public to demonstrate power of interlinked drug discovery data and services. Delivery of first pilot studies to challenge the Open PHACTS framework and guide M18-36 deliveries.

Phase 4: (24 months): The Open PHACTS framework and services demonstrate ability to seamlessly move between domains.

Phase 5: (42 months): A fully functional Open PHACTS Discovery Platform including secure hosting and reasoning for EFPIA companies.

Contributing Partners: UNIVIE, DTU, UHAM, BIT, PSMAR, LUMC, RSC, UNIMAN, USC, Pfizer, Lilly, Novartis, AZ, ME, EMBL-EBI, Janssen

T.6.3 Case Studies

Defined case studies on research questions will challenge the system through in-depth *in silico* drug-discovery analysis/studies executed on public data to demonstrate value and to guide the underlying development. This task will start with the first internal release, thus providing sufficient time to execute small to medium research projects. Research projects will be defined with strong input from EFPIA companies and start with public data retrieved from the Open PHACTS Discovery Platform. Based on the experience of partners involved in the consortium the following case studies are defined:

T6.3.1 Fusion/aggregation of data from different domains to improve predictions of drug-transporter interactions.

Active efflux of compounds mediated by ATP-driven transporters (ABC-transporter) has been shown to influence drug absorption, -distribution, -elimination and has also been linked to the phenomenon of multiple drug resistance in tumour therapy. Both qualitative and quantitative data on interactions of compounds with ABC transporter will be retrieved from the Open PHACTS Discovery Platform and analyzed. Transporters targeted are those relevant for tumour drug resistance and for drug/drug interactions in the liver. These include at least ABCB1, ABCB11, ABCC1, ABCC2 and ABCG2. Qualitative data on substrate properties (yes/no) will be used to derive classification models for prediction of substrate properties of drug candidates. Methods applied include support vector machines, random forest classification, rule fit and counter propagation networks. Quantitative data on drug/transporter interaction (e.g. IC50 values) will be used to establish QSAR and pharmacophore models for individual transporter. Validation of models will be performed by EFPIA companies on their in house data sets. As UNIVIE is also a partner at eTox, this case study also will enable interaction with the eTox data layer for prediction of off-target effects. This case study will challenge the Open PHACTS Discovery Platform especially with respect to quality of qualitative and quantitative data.

T.6.3.2 Combining physicochemical data and data on transporter interaction for prediction of blood-brain barrier permeation and tissue distribution.

Penetrating barriers such as the intestine and the blood-brain barrier as well as negotiating excretory epithelia such as the kidneys represents a prerequisite for proper ADME properties of drug candidates. Within the past decade the importance of active uptake and -efflux transporters for absorption, distribution and excretion of drugs has been increasingly recognised in the community. Although numerous *in silico* models for predicting blood-brain barrier permeation are known, most of them are based on passive diffusion models only. Only a very few take also drug/transporter interactions into account. Currently more than 400 transporters are known as potential interaction partners for drug candidates. Within this task we will retrieve data on the expression levels of selected uptake and efflux transporters at distinct barriers, such as the intestine and the blood brain barrier and kidneys. Subsequently compound interaction profiles will be created and serve as a basis for *in silico* models for interaction. These profiles will be combined with physicochemical data such as solubility, logP values and membrane permeability data in order to create models for prediction of tissue uptake. Within congeneric series of compounds QSAR and pharmacophore models will also be established to retrieve information on the influence of distinct substructures on absorption, excretion, distribution and BBB permeation. This task will challenge the capabilities of the Open PHACTS Discovery Platform to retrieve quantitative physicochemical data as well as tissue specific profiles. In addition, the above tissue- and anatomy-related site data for transporter interaction will be visualized, together with other types of relevant nanopublication information, onto a human disease map. This map will provide a graphical overview of potential interactions between transporters relevant to drug pharmacokinetics and disease states.

T.6.3.3 Target validation work-bench: In-silico target validation studies

Validation of targets and translation of pre-clinical research to predict the clinical efficacy is a major unsolved issue for drug-discovery research. *In-silico* analysis to validate targets and create and prioritize hypotheses for experimental studies to translate to human holds great promise.

A target validation unit will be used as model, with its three units, bioinformatics, *in vitro* validation and *in vivo* validation linked to pathophysiology. A real experimental process in the work-bench from a selected target will be incorporated, including the cascade and all the data generated.

An *in silico* case study comparison with clinical results will be built for each one of the experimental procedures in target validation: identification/fictionalization, feasibility-tools, localization, function modulation, *in vitro* validation and *in vivo* validation. Methods applied include homology studies, polymorphisms, ESTs, frequency, arrays, transgenic mice, iRNA, protein and pathway interactions, immunohistochemistry. Furthermore, this case study will look for new ways of aggregation of data to

support the development and analysis of increasingly complex and predictive high-throughput screening assays.

Contributing Partners: UNIVIE, DTU, PSMAR, RSC, UNIMAN, USC, Pfizer, GSK, Novartis, AZ, ME, EMBL-EBI, Janssen

T.6.4 Training activities and user engagement

Open PHACTS aims to establish standards, semantic integration and services for public and private drug discovery. Hence, one of the major success-criteria for this project is the widespread adoption and creation of a community of users and associated partners that contribute to content (e.g. via crowdsourcing and depositions) as well as additional services by leveraging the Open PHACTS semantic integration hub. Dissemination, engagement and training activities are a key task in building this community as well as a pre-requisite for effective partnering.

Training and engagement activities will take many different forms. The main core training will be based on one-day training workshops at EFPIA-company premises to promote formal and informal knowledge exchange between academia and industry. Another pillar is the organisation of user group meetings. At these meetings case studies will be discussed and new developments will be presented and hands-on training session will be organized.

The project will also use social media, moderate Open PHACTS forum/discussion groups, in different subjects of drug discovery linked to one day training-discussion activities (e.g. LinkedIn groups or VIVO/ORCID collaboration), as well as webinars and video streams to deliver training and outreach. These activities will also be linked to European teaching & training activities, such as the EUROPIN PhD programme in Pharmacoinformatics (www.europin.at).

Contributing Partners: UNIVIE, PSMAR, RSC, USC, Pfizer, GSK, Lilly, Novartis, AZ, ME, EMBL-EBI, Janssen

T.6.5 Researchathons for new scientific use cases

The concept of a researchathon, i.e. bringing scientists from academia and EFPIA together for jointly formulating and solving new scientific use cases, proved very successful for implementing API calls for the pathway functionality. This instrument will be developed further with respect to a strategic instrument for exploring new innovative scientific ground related to interoperability of life science data.

Contributing Partners: USC, UNIVIE, UHAM, BIT, PSMAR, RSC, VUA, UNIMAN, UM, SIB, CD, OGL, GSK, OPF, Esteve, Lilly, Janssen, ALM, SciBite

| | | | | | | | | | | | | |
|-----------------------------|--|--------|------------------------------|---------|------|-----|-----|----|--------|--------|-------|--|
| Work package number | 7 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Central engineering and long term sustainability | | | | | | | | | | | |
| Activity Type* | Other | | | | | | | | | | | |
| Participant number | 1 | 2 | 5 | 6 | 7 | 8 | 12 | 16 | 17 | 18 | 22 | |
| Participant short name* | GSK | UNIVIE | BIT | PSMAR | LUMC | RSC | UM | AZ | Pfizer | Esteve | Lilly | |
| Person-months | 2,8 | 17,5 | 3 | 4,4 | 19,3 | 17 | 6,5 | 2 | 2,1 | 1 | 4 | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | IK | IK | IK | IK | |
| Work package number | 7 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Central engineering and long term sustainability | | | | | | | | | | | |
| Activity Type* | Other | | | | | | | | | | | |
| Participant number | 23 | 24 | 25 | 27 | 28 | 29 | 30 | | | | | |
| Participant short name* | NBIC | SIB | ConnDisc | Janssen | OGL | OPF | ALM | | | | | |
| Person-months | 3 | 10,3 | 6,4 | 1 | 6,6 | 7,7 | 0,3 | | | | | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | | | | | |
| Funding claimed (F/ IK / N) | F | F | IK | IK | F | F | IK | | | | | |

Preamble

Open PHACTS is not just another research project with a defined ending. It is supposed to develop and deliver a crucial, intensively used service. It is thus imperative that Open PHACTS as a Public Private Partnership service, once established, will be *stable*, high *performance*, user *friendly* and *sustainable*.

The EFPIA partners of Open PHACTS will not be the only participants basing some of their core research and business on the system. It is expected, and encouraged, that other organizations will also develop services layered on top of the Open PHACTS Discovery Platform. Actually it will be a significant measure of success **when other PPP's and businesses build useful services and applications** using the Open PHACTS Discovery Platform and data.

Technically this means that the Open PHACTS system should be developed according to best practices in a mixed team where scientists, scientific programmers as well as professional software engineers collaborate to design and develop a *user friendly*, properly *distributed*, *scalable* and *backed up* system that is online and available as close to 24/7 as possible, and able to handle *millions of interactions* on a daily basis. WP7 will co-develop plans and strategies with the technical work packages, but specifically also with WP3 to ensure that long term sustainability of Open PHACTS is a continuous concern of all technical activities during and beyond the project.

Organizationally this means that *stable and professional organizations* should develop and host the infrastructure and that the services be safeguarded by *trusted parties*, both from a *private and public sector* point of view. Clearly, 'stable organizations' also means that they should be (or become) stable in terms of sustained funding. Organizational aspects are closely linked to WP9 (Governance, project management and dissemination) and WP8 (Community engagement and partnering), but here we focus on specific organizational and management needs to secure sustainability of the system and the running of its long term operations.

Objectives

This Work Package governs the development and implementation of a long-term sustainability plan and infrastructure for the services connected and developed in the project. Sustainability of the Open PHACTS Discovery Platform is obviously strongly related to the overall technical reliability, user friendliness and scalability. The purely technical aspects of sustainability are covered in WP3, but in this WP we deal specifically with the human resources, social and funding aspects of sustainability. As development of an Open Pharmacological Space without a solid sustainability plan and approach beyond the duration of the Open PHACTS project would be an unwise investment, a dedicated work package is proposed to oversee the further development and the execution of these plans during the run time of the project. The key deliverable of this work package is an architecture and an organization that is scalable and sustainable beyond the scope and the time of the project.

O7.1. To design a dedicated strategy and approach to ensure sustainable technical operation of Open PHACTS beyond the run time of the project (in close collaboration with WP3)

O7.2. To identify, initiate and foster strategic relationships with associated partners (link to WP8), both as strategic partners in execution and for long term financial support.

Description of work

The **purely technical aspects** of sustainability are handled in WP3

Organizational stability and support

The ‘classical’ infrastructure providers should be involved in Open PHACTS wherever possible (examples include RSC, EBI, SIB etc.). However, new ways to provide infrastructure and support should also be fully exploited. This includes incorporation of both **GRID**- and **CLOUD**-based approaches as well as other federated approaches. The sheer **magnitude of the datasets**, and especially those that will be derived from the Open PHACTS project (> 200 billion triples as an example) require widespread, **distributed computing possibilities** and not a single centralized database or classical data warehousing approach. The new trend is one whereby the analysis ‘agent’ algorithms ‘visit the data’ rather than demand the central collection of data. Data ‘**exposure**’ is the new approach that is replacing classical data integration/federation. Therefore, *by definition*, a range of infrastructure providers should be involved, governed and funded by many different resources. This is a guarantee for maximum risk mitigation in terms of disasters and funding instability. Collaboration with specialized international organizations will be actively pursued. Some of these include ORCID, VIVO, CROSSREF, EBI, LARKC, NUGO, SAGE bionetworks etc. (full list in WP 8)

In terms of *financial sustainability*, it is always a major risk to rely entirely on one funding stream. It is therefore proposed to start working on long-term sustainability of the service delivered by the Open PHACTS project from several funding *sources* from *day one*. (this is an executive task in sustainability WP). These sources will include:

- **Annual private funding:** contributions of end users including pharma and others and also by private enterprises building commercial services around the Open PHACTS core and data.
- **Annual public funding:** contributions from public sector institutions, funders, governments and PPPs
- **Grant based funding:** data sharing policy/plan and budget lines should be included in all data generating grants. These policies and plans should be developed and a dedicated team of experts should be assigned to this issue as part of Open PHACTS.

The Open PHACTS Foundation as the successor organization of the Open PHACTS project will actively engage in these funding streams and especially coordinate grant proposals for other IMI calls as well as e.g.

Horizon 2020 projects. Furthermore, it will alliance with other data intensive IMI projects, such as ELF, and related ESFRI projects.

In Europe, a very important program to align with is **ESFRI**. Participants in ELIXIR and all other life sciences oriented ESFRI projects should be informed and involved in Open PHACTS from the inception on. Also DG INFSO (in charge of ‘digital libraries and data exchange’) should be informed and involved in funding the long term sustainability of Open PHACTS. In that light it is crucial to have some showcase examples early on that show that Open PHACTS is useful beyond the pharmaceutical industry alone (WP5 and 6).

The relationship between Open PHACTS and ELIXIR

There is considerable overlap between the country affiliation of the Open PHACTS participants and the current or prospective ELIXIR nodes and supporting countries. Relative to this ‘national’ perspective, Pan-European Open PHACTS partners such as LarkC, BBMRI, NuGO and Orphanet have many multilateral connections, of which there are many with EBI and/or SIB.

As stated in the formal ELIXIR communication, it is to be expected that several of the ‘existing’ central or networked infrastructures for bioinformatics (for example, NBIC, SIB, Karolinska, Fraunhofer and TDU) will play an important role in the future construction and sustainability of the ELIXIR plan.

It should be clear from this intricate collaboration structure that the sustainability of the Open PHACTS approach can be guaranteed by the gradual building of capacity and increasing interoperability in the developing national nodes of ELIXIR as well as in the Hub, which is the EBI.

The multiple funding model group also should work hard to secure the growing consensus (e.g. formal recommendations from Sage Bionetworks, HGVS and NSF (under preparation)) that data generating grants should contain a peer reviewed plan and budget line for data capture (in standard-compliant format), storage (in sustainable systems) sharing (in interoperable format) and ‘exposure to analysis’. The national and international ‘infrastructure providers’ (such as EBI, NCBI, RSC, but also Amazon and YAHOO types) could collect some of that-widely based and therefore least prone to major fluctuations-subsidy stream to augment structural annual funding by private and public partners. Within the first year of the project, a detailed plan for both the technical and organizational sustainability of the project should be delivered by the Open PHACTS consortium ‘sustainability working group’.

Tasks

T.7.1 Develop a detailed long term maintenance plan for Open PHACTS from a technical perspective. This includes detailed documentation of the technical expertise and financial resources required for operating and expanding Open PHACTS past the term of the project. It is recognized that this requires much more than maintenance and that means have to be found for supporting continued innovation as user demands evolve.

Contributing partners: UNIVIE, BIT, PSMAR, LUMC, NBIC, SIB, RSC, UM, Pfizer, Lilly, ConnDisc, OGL, GSK, OPF

T.7.2 Develop a plan for human resources and funding needed to ensure long term sustainability of Open PHACTS

Contributing partners: UNIVIE, BIT, PSMAR, LUMC, NBIC, SIB, RSC, Pfizer, Lilly, ConnDisc, GSK, OPF

T.7.3 Foster strategic partnerships with the associated partners defined and engaged in WP8, and with partners that can secure long term financial sustainability of Open PHACTS

Contributing partners: UNIVIE, PSMAR, LUMC, NBIC, RSC, UM, Pfizer, GSK, Lilly, Janssen, ConnDisc, OPF

T.7.4 Foster active collaborations with other large scale data driven projects and get engaged in Horizon 2020 initiatives to ensure the sustainability of the Open PHACTS Discovery Platform.

Contributing partners: OPF, UNIVIE, RSC, ConnDisc, PSMAR, GSK, Lilly, Janssen

T.7.5 Liaise with other IMI projects and initiatives

There are a lot of EFPIA projects running or in the early phase of starting, which produce data which are of interest for an Open Pharmacological Space. We will extend our current concept of associated and development partners by building strategic alliances with these projects. This task will also develop a concept for the Open PHACTS Foundation becoming a central data hub for IMI projects, which will contribute to the long term sustainability of Open PHACTS.

Contributing Partners: CD, UNIVIE, PSMAR, LUMC, SIB, OPF, OGL, AZ, GSK, Esteve, Lilly, Janssen, ALM

| | | | | | | | | | | | |
|-----------------------------|-------------------------------------|--------|------------------------------|------|----------|---------|-----------------------|-----|----|--------|----------|
| Work package number | 8 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Community engagement and partnering | | | | | | | | | | |
| Activity Type* | Other | | | | | | | | | | |
| Participant number | 1 | 2 | 6 | 7 | 8 | 12 | 13 | 14 | 16 | 17 | 19 |
| Participant short name* | GSK | UNIVIE | PSMAR | LUMC | RSC | UM | ACK (till 17.09.2014) | USC | AZ | Pfizer | Novartis |
| Person-months | 4,5 | 16,5 | 7,5 | 6,5 | 19 | 5 | 4,8 | 9,5 | 9 | 2,1 | 3 |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | IK | IK | IK |
| Work package number | 8 | | Start date or starting event | | | | | | | M1 | |
| Work package title | Community engagement and partnering | | | | | | | | | | |
| Activity Type* | Other | | | | | | | | | | |
| Participant number | 21 | 22 | 23 | 24 | 25 | 27 | 29 | 30 | | | |
| Participant short name* | HLU | Lilly | NBIC | SIB | ConnDisc | Janssen | OPF | ALM | | | |
| Person-months | 2 | 6 | 1 | 4,5 | 5,2 | 1 | 1,5 | 0,3 | | | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | | | |
| Funding claimed (F/ IK / N) | IK | IK | F | F | IK | IK | F | IK | | | |

Preamble

Principles of an Open PHACTS community:

The Open Pharmacological Space was conceived by EFPIA members in recognition of the shared reliance between public and private drug discovery enterprise. The project is based on a merger of the principles of Open Innovation, Open Source and Open Access. In general the EFPIA objective is to build an open platform (tools and infrastructure) for drug discovery that encourages and supports innovation in a way that does not hinder the freedom of operation of other members of the drug discovery community. Following these general principles, both EFPIA and Open PHACTS recognize the need to achieve the following objectives for Open PHACTS:

1. **Open data:** The use of data not “ownership” of data is the force for innovation.
2. **Open standards:** Data needs to be openly structured, using common semantics and standards.
3. **Open infrastructure:** Databases, semantic layer & services will be Open Source.
4. **Open community:** Building a non-exclusive Open PHACTS community will ensure Open PHACTS adoption.
5. **Free use:** the Open PHACTS Explorer system will be entirely free or at least provided with unlimited license. The interoperable data will be free under the Creative Commons License or equivalent to ensure maximized use and uptake.

Accommodating commercial data and services within an open innovation model

Beyond the immediate open principles of the Open PHACTS project, there is also a recognition of the need to engage with existing commercial information systems that may contribute to Open PHACTS:

- a) Open PHACTS may use existing services which may themselves not be Open Source (e.g. ChemSpider, which is run off Microsoft SQL Server) This approach does not conflict with the ‘open service’ principle built on open standards.
- b) Others outside the core-funded project (associated partners and third parties) would be free and encouraged to build whatever extensions they like into the Open PHACTS Discovery Platform which do not necessarily have to be open source, as long as the service standards are adhered to. This will enable access to existing subscription text content in Open PHACTS in some form; probably through extracted assertions. By making this possible, Open PHACTS could extend its reach outside just the public data providers and become the *de facto* data delivery format for all published information. This should provide an additional return to EFPIA companies, increase the long term sustainability of Open PHACTS and align very closely to Pistoia objectives (with similar principles to the Pistoia SESL project).
- c) The RSC will complete a review of business models to determine which data can be shared on a commercial or open basis. RSC will provide access to a series of appropriate journals of interest to the Open PHACTS consortium for the purpose of text-mining, generation of triples and deposition into the Linked Data Cache. RSC will monitor resulting uptake and use of this information and issue a report regarding the results of this exercise and discuss the results with other publishers. The intention is to establish an appropriate model which will allow additional publishers to be engaged in populating the LDC without threatening their business models.

The need for a wider “Open PHACTS Community”

If the Open PHACTS project is to become sustainable into the future, it is critical to engage a wide community of researchers and information providers in Europe and world-wide. This needs to be done early and needs to be broader than the core Open PHACTS consortium to ensure that the benefits are maximized for the community. The degree of community engagement will be the key to long term success or failure of the Open PHACTS Discovery Platform.

We envisage a broader “Open PHACTS Community”, gathered around open innovation principles with real tangible benefits for participants, including:

1. Knowledge transfer through regular meetings, workshops and networking.
 - a. EFPIA hosted twice-yearly open Open PHACTS workshops
 - b. Open PHACTS 'Hackathons'
2. Direct contact (and potential collaborations) with EFPIA companies.
3. **Stakeholder Influence** on the direction of the IMI project.
4. The Open PHACTS “**kitemark**” concept.
 - a. Tools & research based on Open PHACTS data and standards will improve the value proposition of a product or research grant application
 - b. active review and curation can be **awarded and exposed**.
5. Unrestricted terms of collaboration on Open Innovation principles
 - e.g. The IP for the core Open PHACTS Discovery Platform will be open but there will be no preclusion on 3rd parties building value-added services on top later if they have customers

Objectives

O.8.1: General Objective: to build an Open PHACTS community

By building an Open PHACTS community early and extending "ownership" beyond the IMI funded participants we are effectively creating associated partners who will act as *Open PHACTS ambassadors*. This concept will be critical to ensure the buy-in of the wider drug discovery community, as we cannot rely upon a “build it and they will come” approach. The concept of the Open PHACTS community will also be important to ensure the engagement of non-European researchers, which will also be critical to long-term adoption of Open PHACTS.

O.8.1.1: Community = More Data and Tools

The great ambition of the project tensioned against the modest resources means that the focus of the project lies on building the core Open PHACTS commons and infrastructure. For this reason alone it is essential to partner with a wide range of players in the community and encourage them to contribute content, collaboratively curate and build services.

O.8.1.2: Community = Community Annotation (Crowdsourcing)

In addition, the success of the project will be associated with the engagement of the constituencies of the stakeholders in the process of crowdsourcing the community annotation of the “semantic layer”. Specific activities will be planned with a dedicated budget to ensure the engagement of the partners mentioned below and systematic efforts to create and sustain the relations with the non-IMI funded associated partners will be undertaken as part of the overall sustainability plan. (with WP2, WP4)

O.8.1.3: Community = Providing a roadmap for the future

Engaging all providers of drug discovery information allows the Open PHACTS project to define an information delivery roadmap that provides return for all participants – both consumers of information and providers of information. This also allows business models to be discussed and examined as part of the project to enable understanding and buy-in from existing public and non-public resources.

O.8.1.4: A strong community = a clear value proposition for engagement of resource owners

To maximize resource integration, Open PHACTS will need buy-in from (non-funded) resource owners (e.g. PubChem, ChEMBL, etc.) and screening centers. A clear value proposition will need to be apparent for the data owners to provide incentives to engage with the Open PHACTS process (this work to be done in conjunction with WP4). If a strong Open PHACTS community exists then resource owners will gain direct benefit and exposure for their resource by membership of the community.

O.8.1.5: Community = Utilizing partner’s own networks

Some detailed in appendix. We have a range of partners in the project, which have their own contacts to different sub-communities. It will demonstrate a coordinated and compelling approach if drug discovery researchers and providers are approached in several different ways from different source.

O.8.2: Ambassador Projects = Demonstrable use of Open PHACTS for Drug Discovery

Although we are not funding actual drug *discovery* with this IMI grant, It will be critical to maximize early opportunities for public domain and industry use of the Open PHACTS Discovery Platform to demonstrate value. This would be important for further grant applications etc, to ensure the Open PHACTS legacy. Some exemplars may be addressed in WP6. Active collaborations with ongoing sister projects in IMI will also benefit early adoption of Open PHACTS.

O.8.3 The Development of an Open PHACTS Waiting Room

Such that we might better manage the interaction of Open PHACTS with the wider communities identified above, and synergise with the activities in Work Package 7, we will develop the concept of the Open PHACTS Waiting Room. The Open PHACTS Waiting Room is a managed process by which data/technology/content providers and/or individuals are able to engage with Open PHACTS, sign MoUs and necessary work annexes and engage with and understand the aims and deliverable of the Open Pharmacological Space. The Open PHACTS Waiting Room will be managed by a single accountable person, The Gatekeeper, with overall responsibility and accountability for the management of this engagement process. In many ways this mechanism is analogous to the role of the CTO in the technical work packages. The Gatekeeper will work closely with leaders of work packages 7 and 8, the executive committee, as well as the Open PHACTS Foundation, to collectively ensure oversight and delivery of the community and sustain activities.

Description of work

T.8.1 Establish an Open PHACTS community advisory board (CAB)

The Open PHACTS project and the wider Open PHACTS community would benefit from a non-executive Strategic Community Advisory board (CAB) constituted of selected IMI-funded and non-funded partners. Inclusion of non-funded members would create an opportunity to engage with the wider community, including worldwide researchers. The CAB would meet occasionally and act in a purely advisory role. The CAB could take on the responsibility for organizing the Open PHACTS workshop programs.

Contributing partners: UNIVIE, PSMAR, LUMC, SIB, RSC, UM, ACK (till 17.09.2014), USC, Pfizer, GSK, Lilly, HLU, Novartis, AZ, Janssen, ConnDisc

T.8.2 Establish Open PHACTS community web portal

A web portal will be established at the start of the project to provide BLOG style updates on project progress, publicity and links to project resources, WIKI and training materials etc. This would be an important resource to support Open PHACTS outreach activities and probably needs to be online from inception.

Contributing partners: UNIVIE, PSMAR, LUMC, SIB, RSC, UM, USC, Pfizer, GSK, Lilly, HLU, Novartis, AZ, ConnDisc, OPF

T.8.3 Open PHACTS Community Drug Discovery Workshops

A key vehicle for building the Open PHACTS community and encouraging wider engagement would be a series of Open PHACTS-focused workshops/seminars, with high quality speakers and science content, focusing on different aspects of drug discovery. These would be twice yearly, one day events hosted by EFPIA partners. The workshops could focus on key areas of relevance to the Open PHACTS project and would also allow for specific brainstorming of issues in breakout sessions. Some workshops might be by invitation only if specific focus is required. The workshops would also be a launch platform for Open PHACTS 'Hackathons' and Open PHACTS analysis challenges. High quality speakers and science would be a requirement to ensure good attendance and maximise exposure to the Open PHACTS project and systems.

Workshops would be hosted by EFPIA partners and would be funded at the preferred level of the EFPIA host. This could range from provision of conference facilities only, to full funding of facilities, catering and speaker honorariums.

Contributing partners: UNIVIE, PSMAR, LUMC, RSC, UM, USC, Pfizer, GSK, Lilly, Novartis, AZ, Janssen, ConnDisc

T.8.4 Hackathons: institution of regular “Open PHACTS analysis challenges”

Community analysis competitions along the lines of the long running CASP structure prediction competition (<http://en.wikipedia.org/wiki/CASP>). This could take the form of a competitive in-silico prediction phase (e.g. predicted polypharmacology of a range of compounds), ahead of an EFPIA contribution of laboratory determined data (e.g. lab-determined polypharmacology data). This would be used to address the accuracy of predicted results. Results would be published, with the possibility of a prize for the winning group. Possible subject areas for OPS analysis challenges:

- Polypharmacology prediction for a given compound set
- Druggable genome annotation
- Orphan disease/unmet medical need challenge

Contributing partners: UNIVIE, PSMAR, NBIC, USC, GSK, Lilly, Novartis, AZ, Janssen

T.8.5 Engage the Open PHACTS community to drive adoption of standards, resource curation & peer review

A rigorous set of criteria will be defined against which potential data sources can be judged and potential sources will be assessed. A preliminary list of potential key sources (see WP4) has already been defined by members of the consortium and EFPIA members and this will be the starting point for analysis. Each source will be assessed not only in terms of entity mapping but also other factors such as data availability, service expectations and licensing details from the data owners. This set of core datasets, together with the business questions and the entity mapping will drive a self-consistent set of vocabularies and representation schemes that will feed directly into the Linked Data Cache (LDC) (previously defined in the Architecture and Methodology section of this document). It is likely that some of these core data sources do not currently meet all of the requirements and a key part of task 4.2 will be to engage with the providers to encourage participation. This task might involve, for example, the provision of appropriate incentives to participate. Such incentives may include citation tracking methods described in WP2.

This is a critical task in WP8, which will be dependent on the mutuality and good will established in the wider Open PHACTS community

Contributing partners: PSMAR, RSC, USC, AZ, ConnDisc, OPF

T.8.6 Driving commercial adoption of standards - “Open PHACTS/Pistoia Kitemark”

Commercial and not-for-profit organizations may be interested in a formal seal of approval on their systems to show compliance with Open PHACTS standards. This would demonstrate to funders that a system has long-term utility. As goals in this area are shared with Pistoia, this could be implemented in partnership with the Pistoia Alliance. Although commercial organizations would be expected to provide some Open PHACTS sponsorship in return for an Open PHACTS Kitemark, the primary benefits of this scheme would be wider adoption of standards, rather than a revenue stream.

Contributing partners: UNIVIE, LUMC, RSC, GSK, Pfizer, AZ, ConnDisc

T.8.7 Establish associate partner scheme: The “Open PHACTS Faculty”

The Open PHACTS community is intended to be entirely open, with unrestricted membership, however it will also be important to identify key stakeholders and specifically invite participation (e.g. Resource Amendment 12, 15 January 2015

Owners, KOLs, Related Organizations, e.g. Pistoia). Associate partners would provide direct input into the project and some members would serve in an advisory role. Partners could be published on the Open PHACTS portal and would act as a broader “Open PHACTS Faculty”

Contributing partners: UNIVIE, PSMAR, LUMC, RSC, UM, ACK (till 17.09.2014), USC, Pfizer, GSK, Lilly, HLU, Novartis, AZ, OPF

A list of confirmed associated partners is provided at www.openphacts.org.

T.8.8 Use cases across IMI projects

The Open PHACTS project is increasingly mentioned in IMI calls and we have memoranda of understanding with several running IMI projects. Within this task we will expand these collaborations to the development of joint scientific use cases, thus utilizing synergies and opening new opportunities. Obvious projects to be approached include, among others, eTOX, ELF, Webae.

Contributing Partners: PSMAR, UNIVIE, CD, LUMC, USC, AZ, GSK, Lilly, Janssen, ALM

| | | | | | | | | | | | | |
|-----------------------------|--------------------|-----------------------|------------------------------|------|-----|---------|------|-------|------|------|----------|--|
| Work package number | 9 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Project Management | | | | | | | | | | | |
| Activity Type * | MGT | | | | | | | | | | | |
| Participant number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| Participant short name* | GSK | UNIVIE | DTU | UHAM | BIT | PSMAR | LUMC | RSC | VUA | CNIO | UNIMAN | |
| Person-months | 7,6 | 78 | 1.4 | 2.4 | 2.5 | 5.5 | 5 | 6.2 | 3.4 | 1.4 | 5 | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | IK | F | F | F | F | F | F | F | F | F | F | |
| Work package number | 9 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Project Management | | | | | | | | | | | |
| Activity Type * | MGT | | | | | | | | | | | |
| Participant number | 12 | 13 | 14 | 15 | 16 | 17 | 21 | 22 | 23 | 24 | 25 | |
| Participant short name* | UM | ACK (till 17.09.2014) | USC | UBO | AZ | Pfizer | HLU | Lilly | NBIC | SIB | ConnDisc | |
| Person-months | 2.4 | 1.4 | 2.4 | 1.4 | 9 | 11,3 | 1.5 | 5 | 2 | 3 | 4,2 | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | |
| Funding claimed (F/ IK / N) | F | F | F | F | IK | IK | IK | IK | F | F | IK | |
| Work package number | 9 | | Start date or starting event | | | | | | | | M1 | |
| Work package title | Project Management | | | | | | | | | | | |
| Activity Type * | MGT | | | | | | | | | | | |
| Participant number | 26 | 27 | 28 | 29 | 30 | 31 | | | | | | |
| Participant short name* | EMBL-EBI | Janssen | OGL | OPF | ALM | SciBite | | | | | | |
| Person-months | 1 | 1 | 1,5 | 0,8 | 0,3 | 2 | | | | | | |
| Other resources (YES/NO) | NO | NO | NO | NO | NO | NO | | | | | | |
| Funding claimed (F/ IK / N) | F | IK | F | F | IK | F | | | | | | |

Preamble

The governance structure of the Open Pharmacological Space is constructed to enable the development and needs of a large-scale distributed project, whilst remaining agile, adaptable and flexible. Successful development of the Open Pharmacological Space as a keystone of the public scientific computational resources will require a focus on communication between partners and work packages to ensure deliverables are timely and fit-for-purpose. The structure also needs to allow the large number of partners to be engaged in governance and delivery, balance potentially competing EFPIA and academic needs, whilst supporting the needs of the target ‘customers’ in the scientific user community.

Objectives

Work package 9 covers the overall management of the project, with responsibility for the scientific and technical co-ordination, as well as for communication, dissemination of results and future

hosting and development. This work package will thus comprise leadership, organizational activities, coordination of all technical and scientific activities and internal and external communication and will be

lead by the Project Coordinator and the Managing Entity of the IMI JU funding in close collaboration with the work package leaders and the technical and scientific task force. In particular, monitoring of milestones, on time submission of deliverables and of pilot exemplar services will be ensured. Furthermore, this WP will also take care that feedback received from the scientific community (link to WP6 and WP8) will be considered and implemented into the following release of the exemplar services.

The main objective of this work package is to ensure a firm and dynamic scientific and technical coordination of the project. This involves both leadership and coordination. Open PHACTS covers several different scientific disciplines, such as semantic web approaches, data-, text-, and picture mining, model generation, software engineering, and last but not least application to medicinal chemistry research questions. Combining scientific and technical knowledge, vision and collaborative work, we will monitor and direct the activities in an optimal way. This will mainly be ensured by a deep connection between scientific and technical coordination and project management. In addition, both a scientific task force and a Scientific Advisory Board will be implemented. These will guarantee a permanent push for top scientific level paired with high end, robust, and visionary technical implementation.

The project management aims to ensure that the project is appropriately managed and the work is implemented according to the plan. It will monitor the progress in all work packages and ensure that specific results are delivered in time, fulfill proper quality criteria, and are obtained within the budget assigned to this task. The management will follow a rolling working plan policy to ensure enough flexibility for adaption of tasks to avoid delays and to combat risks identified. It will further support the partners in financial and administrative purposes and in providing reports. Furthermore, the project management will take responsibility for a proper execution of the grant agreement and the project agreement and will take care on IPR issues. Last but not least it will coordinate all dissemination activities and will act as primary contact to the IMI JU and other bodies.

A key element in an interoperability project with a relatively high number of participant organisations is to establish a solid management structure that ensures the coordinated execution of the project, both at the whole consortium level and at the level of the working teams, hence avoiding the risk of running a series of disconnected mini-projects. In particular, the management structure will consist of the following components:

Executive Committee (EC): An operational body comprising project leaders and their deputies both from EFPIA and the applicant consortium (voting rights), as well as a project manager, the chairs of the scientific and technical task force, and the chair of the Scientific Advisory Board.

Steering Committee (SC): The ultimate decision making body with representatives of all IMI funded partners and EFPIA companies signing the grant agreement.

Associated Partners (AP): A group of organizations linked to Open PHACTS through contribution of data, tools and services, but without funding. They are invited to the Steering Committee meetings, but without voting rights.

Scientific Advisory Board (SAB): An external body of experts in computational life sciences giving advice to the Executive Committee on the development and dissemination of Open PHACTS.

Project Management Unit (PMU): A management team composed of a Project Manager and a project management office, which is located at the managing entity of the IMI JU funding (University of Vienna).

Technical Task force (TTF): A core group, which monitors, guides, and coordinates the technical implementation of the Open Pharmacological Space (effectively WP3)

Scientific Task Force (STF): A core group, which comprises scientists from all core disciplines, related to Open PHACTS (semantic approaches, bioinformatics, cheminformatics, medicinal chemistry)

Work Package Leaders Group (WPLG): Comprises all work package leaders responsible for the proper execution of WPs 1-9.

All these bodies will operate in a structured, interconnected way to achieve dynamic scientific and technical coordination of the Open PHACTS project as well as top level day to day project management. Finally, it will ensure the delivery and dissemination of high quality reports and products.

Description of work

T.9.1 Scientific Coordination

This task will focus on the scientific project leadership and its proper coordination. Due to the scientific complexity of the project and the manifold disciplines involved this is crucial for the success of Open PHACTS. The scientific coordination will thus not only rely on an Executive Committee composed of 2 scientists from EFPIA companies and two from the applicant consortium (all of them coming from different scientific disciplines) and on the work package leader group, but also encompass a scientific task force and an external scientific advisory board. The latter two bodies ensure that scientific progress in the field is constantly monitored and immediately implemented into the respective tasks and work packages. The main aim of this task is thus to foster regular contacts between these groups and to coordinate their communication. It will further include coordination and promotion of contacts and relationships to other initiatives in the area, especially those funded by IMI.

Close coordination with technical activities as outlined in task 9.2 and the project management (task 9.3) is an absolute must for a success of task 9.1. Thus, the chair of the Technical Task Force (the CTO of WP3) will be a permanent observer to the Executive Committee and take part on all its meetings and teleconferences. This ensures a regular discussion of scientific and technical issues in the executing bodies and ensures a proper awareness of latest developments. This will be complemented by the Steering Committee meetings, which will be the main forum for scientific exchange in its broadest way. As they will include also the associated partners this forum is expected to host at least 50 top scientists from all disciplines, including also experts from US and Asia. This forum will be the world leading group in the field and ensure a broad and immediate implementation of tools and standards implemented within this project.

Contributing partners: Pfizer, UNIVIE, DTU, UHAM, BIT, PSMAR, LUMC, NBIC, SIB, RSC, VUA, CNIO, UNIMAN, UM, ACK (till 17.09.2014), USC, UBO, AZ, GSK, Esteve, Novartis, ME, HLU, Lilly, EMBL-EBI, ConnDisc, Janssen, OGL, OPF, ALM, SciBite

T.9.2 Technical Coordination

The design and creation of a reliable system running close to 24/7, thus enabling the implementation of business services around it is normally beyond the output of the scientific process. Thus, besides scientific coordination, this project will also enforce a technical coordination. In order to ensure that the long-term scientific and technical goals of Open PHACTS are implemented in a concrete, step by step way leading to professional software, a technical task force (TTF) will be implemented. The TTF is a core group with overall responsibility for the technical implementation and overall architectural design and integrity of the Open PHACTS Discovery Platform. The task force will monitor technical development, prototyping, beta-testing, work-to-working implementation and release of the overall Open PHACTS architecture, platform and the exemplar services. The TTF chair will effectively operate as the Chief Technology Officer (CTO) of Open PHACTS (see WP3 for detailed description)

Each of the work packages will be monitored and technically supervised by dedicated technical project leaders (TPL), responsible for the tactical technical aspects of the work packages. The TPLs are assigned to and assists in translating the long-term goals of the WP into practical small steps to be implemented.

This will help to facilitate the technical integration of the WP's and to guide the development of all the software and services outlined in the proposal. Regular reporting on the technical progress in the WPs will identify possible areas of overlap and help to minimize the risk of lack of interoperability.

This task will also develop, implement and monitor procedures for a continuous communication between the scientific programmers. Organisation of frequent meetings and joined coding sessions (Open PHACTS Hackathons) will assure that the programming efforts in the individual WPs stay aligned. The technical coordination will also ensure that the scientific programmers stay in are in permanent close contact with practical users so that they can solicit guidance from the end users of the OIPS system and the exploitation pilots developed.

Participating partners: Pfizer, UNIVIE, DTU, UHAM, BIT, PSMAR, LUMC, RSC, VUA, UNIMAN, UM, ACK (till 17.09.2014), AZ, GSK, Novartis, HLU, ConnDisc, OGL, SciBite, ALM, Janssen, EMBL-EBI, SIB

T.9.3 Project Management

The project management is devoted to cooperate with and provide support to the scientific and technical coordination, the executive committee and the steering committee as well as to the scientific and the technical task force. This task will include:

- Support to task and WP leaders in day-to-day management, decision making and conflict resolution
- Work plan control, assurance of in time delivery of deliverables and implementation of corrective actions
- Setup and management of quality control procedures on deliverables and dissemination material
- Setup and maintenance of tools for efficient communication and cooperative work among all partners
- Support to the Executive Committee, and the scientific and technical task force in decision making
- Support to organization of meetings, production of minutes, implementation of corrective actions into the work plan and follow up of all action points decided
- Promotion of synergy and efficiency throughout the consortium
- Implementation of the grant agreement and the project agreement and amendments thereof
- Communication with the IMI office
- Management of the relationship to associated partners and to external partners

Furthermore, this task will also encompass the specific reporting activities agreed upon in the grant agreement. This includes guidance and support to the WP leader in appropriate reporting, development and distribution of standard forms, as well as in time collecting of deliverables to ensure a proper quality control before submission to the IMI office. All these periodic reporting activities (once a year) will be facilitated via a collaborative web-based system, which will provide a daily actual overview on all project activities.

A major activity of the project management will also be devoted to management of the IMI JU funding. This includes distribution of the funding, cost control and justification, management of the overall budget and proper reporting to the IMI office. Cost control will be tightly linked to deliverables, reports, and quality control mechanisms to ensure a proper use of the budget allocated to each task. Budget assignment will be flexible to allow dynamic and rapid reaction of new scientific and technical developments and to help to use the money in the most efficient way.

Finally, this task will deal with risk management. This comprises the identification, analysis, assessment and monitoring of risks affecting the project or its results as well as the development and monitoring of risk management procedures aiming at mitigating the threats and utilizing the opportunities. Following a bottom up approach, risks will be identified by the work package leaders (upon notification by task leaders) and reported to the Executive committee. Assessment and quantification of the risks reported will be performed combining their probability and impact, each quantified in a scale of 1-10. Multiplying the two variables will give a risk factor ranging from 1 to 100, which will be used for prioritizing the risks. In a next step each risk will be characterized (scientific or technical, owner, etc.) and actions to affect probability and/or impact in order to avoid the risk (mitigation plan) will be defined. In addition, procedures will be defined for incorporation of risks that effectively happened and affected the project into the work plan.

Contributing partners: GSK, AZ, UNIVIE, LUMC,

T.9.4 Communication and Dissemination

This task will focus on the development and execution of a dissemination and communication plan for Open PHACTS. Extensive use of the Open PHACTS Discovery Platform will be a direct measure of success of the Open PHACTS project. Thus, a structured communication plan for rising awareness in the community, promoting the system, and gaining feedback is of utmost importance. This plan will be based on the following main actions:

- Definition of the communication objectives, such as raising awareness in the community, promoting usage of the Open PHACTS Discovery Platform, encouraging academic groups to provide their data, encourage academic groups and SMEs to develop business around the Open PHACTS Discovery Platform, ...
- Definition of the target audience, e.g. academic groups in Europe, US and Asia, SMEs, other EU-funded initiatives and consortia, ...
- Definition of the actions to be undertaken, such as writing articles in high impact journals, establishing a newsletter, organizing sessions at international meetings, ...
- Definition of the tools used, such as web-page, newsletter, articles, blogs, interviews, flyer, ...

Upon approval of the plan by the Steering Committee (D9.2, **Month 6**), the respective tools will be developed and implemented. A major partner in the dissemination and communication of Open PHACTS will be the European Federation for Medicinal Chemistry (EFMC) with its network of more than 6.500 medicinal chemists, regular meetings, short courses, and the e-newsletter MedChemWatch. Gerhard Ecker, the coordinator of the academic consortium, served as president of EFMC 2009 – 2011 and in this capacity certified the commitment of the EFMC to promoting the Open PHACTS project and encouraging the medicinal chemistry community to provide data and tools to the Open PHACTS Discovery Platform.

Subsequently, the dissemination activities as defined will be undertaken by the consortium. Due to the technical nature of the project, this will not only include reports, articles, talks at conferences, and other scientific activities, but also encompass the delivery and maintenance of exemplar pilot services as developed in WPs 2, 5 and 6 and respective training material and -events. Appropriate review processes and beta-testing activities for the pilot services are defined in WPs 3 and 6 and will be strictly monitored to ensure high quality products serving the community in its need for mining the huge amount of public available information available. Top quality documentation and support will also ensure that the Open PHACTS Discovery Platform is utilized in creating business and high quality services attractive for EFPIA companies.

Participating partners: Pfizer, UNIVIE, DTU, UHAM, BIT, PSMAR, LUMC, RSC, VUA, CNIO,

UNIMAN, UM, ACK (till 17.09.2014), USC, UBO, AZ, GSK, Esteve, Novartis, ME, HLU, Lilly, NBIC, ConnDisc, EMBL-EBI, SIB, Janssen, OGL, OPF, ALM, SciBite

T.9.5 Dissemination of scientific use cases

One of the major success factors of the Open PHACTS project was the immediate adoption by a broader community. This was also due to the presence of Open PHACTS scientists at international conferences, both as speaker and as exhibitors. However, this put quite some pressure on the travel budgets of individual partners. Therefore, we reserved a budget of € 20.000 for giving travel grants to young scientists presenting scientific use cases and innovative technological solutions connected to the Open PHACTs project.

Contributing Partners: OPF, UNIVIE, UHAM, BIT, PSMA, LUMC, RSC, VUA, UNIMAN, UM, USC, SIB, CD, EMBL-EBI, OGL, SciBite, GSK, Lilly, Janssen, ALM

T.9.6 Quality Assurance

Quality of the Open PHACTS Discovery Platform as well as the eApps developed is of crucial importance for the success and sustainability of Open PHACTS. Quality and usability will be assured by a permanent monitoring of the tools by all partners, especially by those from EFPIA. In addition, controlled pre-releases to a selected group of users combined with feedback gathering will ensure a stringent monitoring of the quality.

Contributing Partners: GSK, UNIVIE, DTU, UHAM, BIT, PSMA, LUMC, RSC, VUA, CNIO, UNIMAN, UM, ACK (till 17.09.2014), USC, UBO, AZ, Pfizer, Esteve, Novartis, ME, HLU, Lilly, NBIC, ConnDisc, EMBL-EBI, SIB, Janssen, OGL, OPF, ALM, SciBite

4. Implementation

4.1 Governance structure and management procedures

The governance structure of the Open Pharmacological Space is constructed to enable the development and needs of a large-scale distributed project, whilst remaining agile, adaptable and flexible. Successful development of the Open Pharmacological Space as a keystone of the public scientific computational resources will require a focus on communication between partners and work packages to ensure deliverables are timely and fit-for-purpose. The structure also needs to allow the large number of partners to be engaged in governance and delivery, balance potentially competing EFPIA and academic needs, whilst supporting the needs of the target ‘customers’ in the scientific user community.

The consortium consists of 30 European core partners with top expertise in all elements addressed in this call topic. Many of these partners have associated partners with current collaborations. These ‘associated partners’ - both in Europe and in the USA - together ensure that whatever the core partners develop, whether it is scientific approaches, methodologies, best practices and semantic standards will be followed and adopted by the whole community. Through its ‘secondary circle of partner institutions’, the consortium influences open standard setting and best practices in partner institutions world-wide. Upon decision by the Steering Committee, associated partners will be invited to join the Associated Partners Group of Open PHACTS, which will participate as observers to the steering committee meetings. These partners include major data provider (e.g. EPO, IUPHAR), biomedical research institutions, biobanks (BBMRI), semantic web associations (e.g. W3C, HCLS), not for profit screening centers (EU-OPENSOURCE), and small to medium sized enterprises (e.g. Knewco, Quertle). As interoperability of data and information is a global issue by definition, the associated partners are not restricted to Europe alone. Through the existing federated approach of the Concept Web Alliance (CWA), the National Center for Biomedical Ontology (NCBO), hosting and curating all OBO ontologies, the World Wide Web Consortium (W3C) and the LarKC consortium will ensure that also data sources based in North America and elsewhere will be in sync with Open PHACTS (e.g. PISTOIA, Sage Bionetworks, Bio2RDF, Chem2Bio2RDF)

The project is based on a breakdown into work packages and tasks which are executed in parallel. This implies an intensive and structured communication as well as a strict monitoring of all milestones and deliverables. Furthermore, as a successful implementation of the the Open PHACTS Discovery Platform is crucial for attracting the scientific community to provide data and tools, this project requires both a strong scientific and technical leadership as well as intense dissemination and community communication activities. Thus, the governance structure and management of Open PHACTS reflects these unique needs and extends in its concept beyond standard project management architectures. In particular, it covers the following objectives:

- Fulfilment of the work plan
- Implementation of the project agreement
- Fulfilment and proper execution of the grant agreement
- Monitoring of the achievement of milestones and deliverables and follow up activities in case of delays
- Monitoring the quality and efficiency of the project activities and the products delivered
- Coordination of the communication amongst partners
- Conflict management and – resolution
- Communication with the scientific community and dissemination of project results
- Liaison with related activities in the US and Asia

A key element in an interoperability project with a relatively high number of participant organisations is to establish a solid management structure that ensures the coordinated execution of the project, both at the whole consortium level and at the level of the working teams, hence avoiding the risk of running a series of disconnected mini-projects. This management structure will also make sure that synergistic connections with external related initiatives (such as IMI eTox) as well as with winning consortia for other IMI call topics related to knowledge management (call topics 7 and 9, see Annex 2) is established and maintained during the project.

The size of the consortium, the highly ambitious aims, the diverse profiles of the participating institutions as well as the complex scientific and technical questions approached renders Open PHACTS a very challenging project both from the management and the scientific point of view. In addition, the management structure needs to balance and harmonize the needs of both the academic as well as the industry partners. In order to achieve these aims a project management and governance structure is proposed which allows tight links between the day to day management on the one side and the scientific, technical and community engagement activities on the other side. In particular, the management structure will consist of the following components (Figure 6):

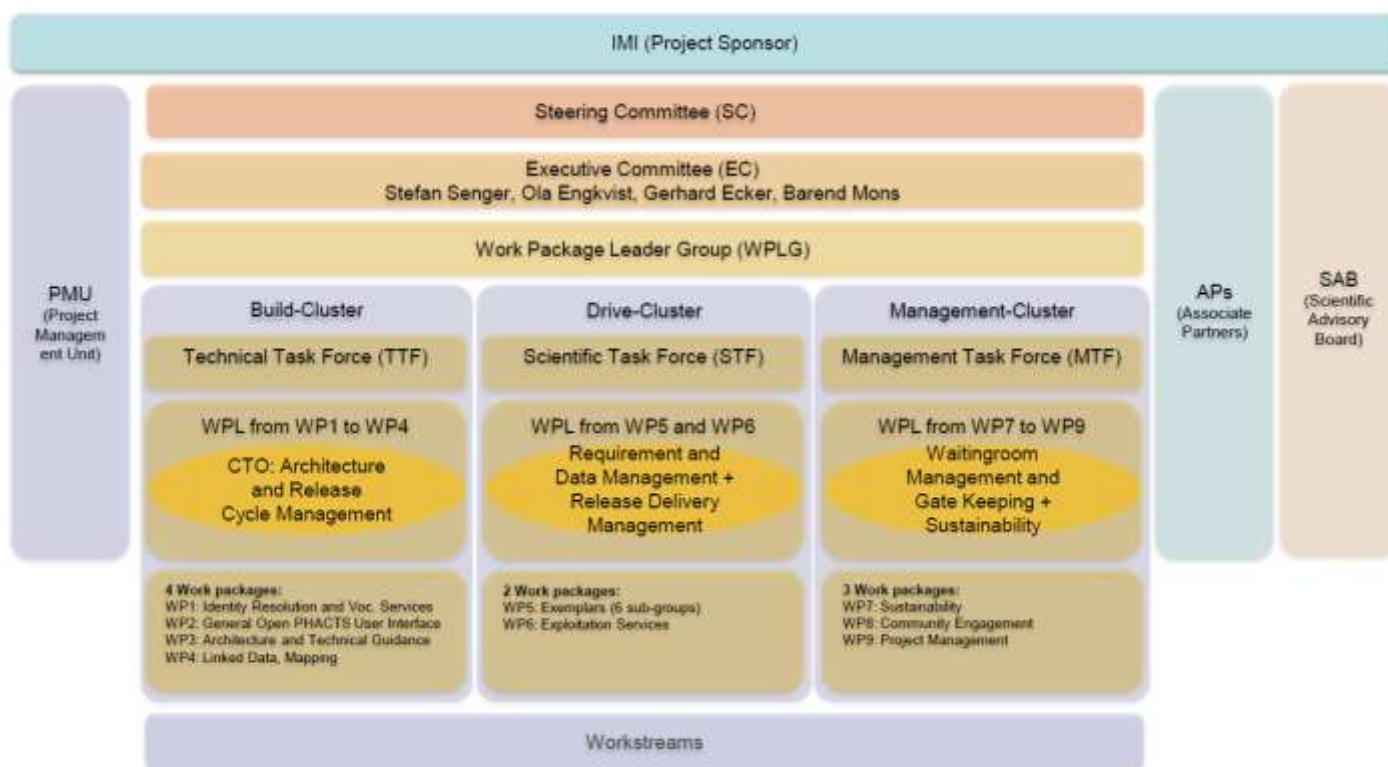


Figure 6: Management and governance of the Open PHACTS project

Executive Committee (EC): An operational body comprising project leaders and their deputies both from EFPIA and the applicant consortium (voting rights), as well as a project manager, the chairs of the scientific and technical task force, and the chair of the Scientific Advisory Board.

Steering Committee (SC): The ultimate decision making body with representatives of all IMI funded partners and EFPIA companies signing the grant agreement.

Associated Partners (AP): A group of organizations linked to Open PHACTS through contribution of data, tools and services, but without funding. They are invited to the Steering Committee meetings, but without voting rights.

Scientific Advisory Board (SAB): An external body of experts in computational life sciences giving advice to the Executive Committee on the development and dissemination of Open PHACTS.

Project Management Unit (PMU): A management team composed of a Project Manager (PM) and a project management office, which is located at the managing entity of the IMI JU funding (University of Vienna).

Technical Task force (TTF): A core group, which monitors, guides, and coordinates the technical implementation of the Open PHACTS Discovery Platform. The TTF is chaired by the Chief Technological Officer (CTO)

Scientific Task Force (STF): A core group, which comprises scientists from all core disciplines, related to Open PHACTS (semantic approaches, bioinformatics, cheminformatics, medicinal chemistry)

Work Package Leaders Group (WPLG): Comprises all work package leaders responsible for the proper execution of WPs 1-9.

Work Packages (WP): A work Package is comprised of all partners contributing to the work defined for this WP. A work package might be structured into tasks, which are the smallest unit of work defined in this proposal.

Management of the Open PHACTS project will be performed by the **EXECUTIVE COMMITTEE (EC)**. The EC is an operational body responsible for the tactical day to day running of the Open PHACTS project, with decision powers on technical development (advised by the technical task force), budget assignments within work packages, and overall ownership of the communication and dissemination function. It will review and prepare all reports for approval by the Steering Committee (SC) and may propose changes to the work plan and to the composition of the consortium to the SC. The EC will constitute the EFPIA co-ordinator (Bryn Williams-Jones, Pfizer) and deputy (Ola Engkvist, AZ), the academic co-ordinator (Gerhard Ecker, UNIVIE) and deputy (Barend Mons, LUMC) as voting members. In the event of a 50:50 vote, the EFPIA co-ordinator as chair will have an additional casting vote. Also attending the EC in a non-voting role are the Project Manager, the chairs of the technical and scientific task forces, and the chair of the Scientific Advisory Board. The EC may also call specific or all WP leaders for a meeting if a wider consensus. It is expected that the EC will meet every 4 weeks in a teleconference and 2 times a year in a one day face-to-face meeting.

The **STEERING COMMITTEE (SC)** will be made up with one representative from all funded partners and EFPIA. The SC will have ultimate decision-making responsibility in matters affecting overall project strategy, composition of the consortium, appointment of associated partners, and budget allocation between work packages. It will approve project deliverables and amendments to the work plan. The SC will be jointly chaired by the EFPIA and applicant co-ordinators, and 75% of partners represent a quorum. Voting rights within the steering committee are as follows: each partner has one vote and decisions are made with a 2/3 majority. With a composition of 10 EFPIA companies, 15 academic groups, 4 SMEs and 1 learned society/publisher, this ensures that EFPIA companies cannot be overruled by the applicant consortium. It is expected that the steering committee will meet twice a year, and be the ‘anchor’ meeting for satellite WP and other functional Open PHACTS meetings.

The **ASSOCIATED PARTNERS (AP)** are organisations engaged in the Open PHACTS project through contributing data, tools or services but not receiving IMI JU funding. Their participation in all major meetings of Open PHACTS ensures the active participation of key player in the field. These associated partners - both in Europe and in the USA - together ensure that whatever the core partners develop, whether it is scientific approaches, methodologies, best practices and semantic standards will be followed and adopted by a community currently comprising over 200 institutions world-

wide. Conditional on their level of commitment and upon decision of the SC, associate partners will be invited to participate in SC meetings as observer.

The **SCIENTIFIC ADVISORY BOARD** (SAB) will be an external body of experts in the computational science domain, and will represent the global scientific user community with special focus of representation of US and Asia. There will be a balance of EFPIA and Academic advisors, and their role will be to advise on development and dissemination of the Open PHACTS Discovery Platform. The SAB chair will attend SC meetings, and is an ad hoc non-voting member of the EC. The SAB will be appointed at the kick-off meeting.

The **PROJECT MANAGEMENT UNIT** (PMU) will be comprised of a project manager (PM) and a management office and will be affiliated at the University of Vienna. The University of Vienna will be the managing entity of the IMI JU funding, with responsibility for financial tracking and reporting, project secretariat, and ‘back office’ support. It will provide support to the coordinators and all committees in the day-to-day management organize the meetings and will facilitate communication among partners. It will also maintain the contacts to the IMI office and provide legal support for IPR issues. Finally, it will maintain the web-page of the project. The Project Manager (PM) will manage the project with respect to checking timelines and deliverables, prepare the EC and SC meetings and taking a leading role in the management of the online workspace.

The **TECHNICAL TASK FORCE** (TTF,) is a core group with overall responsibility for the technical implementation and overall architectural design and integrity of the Open PHACTS Discovery Platform. The core of the task force will consist of 6-8 members split equally between EFPIA and academia. These members need to have a senior software engineering level in their respective partner organisation. The TTF will monitor technical development, prototyping, beta-testing, work-to-working implementation and release of the overall Open PHACTS architecture, platform and the exemplar services. The TTF chair will effectively operate as the Chief Technology Officer (CTO) of Open PHACTS and will be appointed by the EC after consultation of the TTF. The CTO will attend the EC to represent technical components of all WPs, and offer technical advice to voting EC members.

The **SCIENTIFIC TASK FORCE** (STF) is a core group of 6-8 members split equally between EFPIA and academia with the aim to coordinate the broad scientific initiatives ranging from data mining, annotation, small molecule data storage and manipulation, target related bioinformatics, pathway annotation, protein structure analysis, massive daily *in silico* reasoning and meta-analysis, chemical biology, cheminformatics and medicinal chemistry. The STF chair will be elected by STF members, will attend the EC to represent scientific components of all WPs, and offer advice to voting EC members. The STF will have regular telephone conferences and organize meetings devoted to special tasks, such as prioritisation of research questions as well as the researchathons..

The **WORK PACKAGE LEADERS GROUP** (WPLG) is composed of all work package leaders and will be responsible for tactical issues relating to interdependence between work packages. This group will be empowered to take tactical decisions relating to work package delivery, and will pass up issues to EC for decision if resolution cannot be achieved in the resources held by the work package leaders. This will also be a communication forum for work packages leaders to advertise progress and to request focused support for particular needs. EC members (if not already present as WP leaders) will attend ad hoc. The WPLG will meet 2 times per year and will be chaired by the EFPIA coordinator. The Work Package Leaders will intensively coordinate with the TTF for technical implementation of the prototypic outputs of their Work Package

WORK PACKAGES (WPs) will be comprised of all partners contributing to the WP. Each work package is jointly coordinated by a scientist from an EFPIA company and by one from the applicant

consortium. These work package leader have the responsibility for the day-to-day management and coordination of activities of their work package. They will produce the deliverables and reports, identify risks and report regularly the progress of the WP to the SC and EC. The work package members will meet according to their own schedule, and in person at least twice a year around the SC meetings. Issues that cannot be resolved within the WP are referred to the EC or WPLG as appropriate.

The Engineering set up of the Open PHACTS Discovery Platform

The design and creation of a reliable system is normally beyond the output of the scientific process. So far there are hardly any 'bioinformatics' environments that meet the criteria defined for the Open PHACTS Discovery Platform earlier. Thus it is proposed to implement a software development model for mixed academic and 'industrial' software development. This encompasses an organisation in three layers: work package management, coordination through an engineering team, and scientific programmers that perform most of the prototypic programming.

Work-package management

In addition to the two WP leaders (one from EFPIA, one from academia), each of the work packages is monitored and technically supervised by a technical project leader (TPL). This TPL is responsible for the tactical technical aspects of the work packages he/she is assigned to and assists in translating the long term goals of the WP into practical small steps to be implemented. Naturally the technical project leaders will form the core of the Technical Task Force, which will be chaired by a chief technical officer. The TTF has two primary goals: to facilitate the technical integration of the WP's and to guide the development of all the software outlined in the respective WPs, especially the GUI and the exploitation pilots.

To facilitate the integration, each TPL will regularly report about the progress in their WPs. Frequent progress reports will make it possible to take early note of possible areas of overlap or oversight in the entire project, so that the impact of such risks can be minimized. Software development will focus on enhancing reliability and usability of software developed either in the WP's or originating from outside of the project. This engineering work will be focused on the coverage of unit tests and documentation and on user interfaces, API's and interoperability between components.

Scientific programmers

The bulk of the software engineering is carried out by scientific programmers that are dedicated to a particular WP. Continuous communication between the scientific programmers and frequent meetings and joined coding sessions (Open PHACTS Hackathons) assure that the efforts in the WP stay aligned. The scientific programmers are in close contact with practical users so that they can solicit guidance from the practical users of the technology they develop. The work between the scientific programmers will therefore be distributed in such a way that optimal use can be made of the experience in each of the hosting groups. Regularly, the group leaders will meet to discuss the progress and requirements.

Communication and Conflict Management

The management and governance structure has been designed to promote smooth and dynamic collaboration between the Project partners. As Open PHACTS is not only a scientifically challenging project, but also a technical one, strong emphasis will be put on both communication inside and outside the consortium as well as on conflict resolution procedures, which are key aspects in a distributed project of these characteristics.

With respect to **communication**, a communication plan (D 9.2) will be established, making extensive use of **electronic resources**. A password-protected internet structure will be set up to support management activities, communication and exchange of information between partners and knowledge management and training. This will include online tools for progress reporting, different mailing lists (whole consortium, task forces, WP members, etc.), as well as an repository of all publications funded by Open PHACTS. Technical cooperation between participants will also be facilitated by electronic platforms as needed. At least twice a year **face-to-face meetings** will take place which will also include the associated partners to encourage a maximum level of scientific and technical exchange among all the partners. Partners will also be encouraged to meet within their work package, task, or even bilateral for the implementation of the work plan.

With respect to **conflict resolution**, a **bottom-up approach** will be taken. Conflicts between cooperating partners in any given activity will at first instance be solved by the respective work package leaders. If this is not possible (or in case of a conflict of interest), the issue will be discussed in the executive committee, and if needed, referred to the steering committee. Especially in case of **reassignment of roles** or of **budget** a **voting** might be necessary. Respective rules for this will be established in the project agreement. The coordinators will assist in conflict resolution at all levels and provide an objective analysis of the situation. They will use bilateral contacts, negotiation and mediation to solve the issues, always aiming to reach consensus. In cases where **legal action** is needed, the project managing unit (UNIVIE) will seek the required **authorization from the steering committee** and act accordingly in agreement with the legal documents regulating the development of the Open PHACTS project.

4.3 Consortium as a whole

The Open PHACTS consortium is ideally positioned to develop an open pharmacological space. It comprises **30 organisations** (10 EFPIA companies, 15 academic groups, 4 SMEs and 1 learned society/publisher), with top scientific and technical expertise in all elements addressed in this call topic. In case of partner PSMAR, Fundació IMIM (FIMIM) and UPF will provide resources to PSMAR and thus be third parties of PSMAR. RSC Worldwide Ltd will be a third party of RSC. The consortium unite for the first time leading European institutions dealing with ontologies and standards, reasoning, workflows, API and web service development, major database developers and owners, standards in data capture, information exchange and integration, data and text mining, annotation, curation and quality control in the chemical and biological fields, the development of tools for target characterization, information navigation and visualization techniques, as well as QSAR and integrated model generation and partners with 9 major European pharmaceutical companies.

Furthermore, consortium members are connected to or directly involved in (i) ChemProt, which contains more than 600,000 compounds annotated to almost 6000 proteins, covering over 2 million chemical-protein annotations (**DTU**), (ii) the ConceptWiki, which currently contains 13.5 million concepts extracted from the Unified Medical Language System (UMLS), UniProt-SwissProt and PubMed, (**LUMC/NBIC**) (iii) the Linked Life Data Store, a public RDF warehouse which semantically integrates more than 20 popular biomedical data sources and currently contains over 4 billion triples connecting almost 6 million RDF resources (**VUA/LarKC**), (iv) the gene and protein index which is part of SCAIView comprising Medline, SwissProt, NCBI Gene, NCBI dbSNP, MeSH, DrugBank (**UBO**), and (v) the biocreative metasever and related resources that currently operate a set of 12 distributed text mining servers dedicated to the normalization of protein names and the identification of protein interactions (**LUMC/UBO**).

Many of the academic groups are also centrally involved in existing consortia of collaborating partners, such as Pistoia alliance, ESFRI projects or Networks of Excellence and global alliances

dealing with interoperability and Semantic Web developments. The Open PHACTS consortium also will invite selected organisations/institutions to become **associated partners**. The participation of these ‘associated partners’ - both in Europe and in the USA - ensures that whatever will be developed in the framework of the Open PHACTS project, whether it is scientific approaches, methodologies, best practices and semantic standards, will be followed and adopted by a community currently comprising over 200 institutions world-wide.

Through its ‘secondary circle of partner institutions’, the consortium influences **open standard setting** and **best practices** in partner institutions world-wide. These partners include biobanks (BBMRI), semantic web associations (W3C, HCLS) and small to medium sized enterprises (Knewco, Quertle, Molecular Networks, Lhasa, Ontotext). Core partners also participate in multiple ESFRI projects, such as ELIXIR, BBMRI, EATRIS, EU-OPENSSCREEN, and ECRIN.

As interoperability of data and information is a global issue by definition, the associated partners are not restricted to Europe alone. Through the existing federated approach of the Concept Web Alliance, the National Center for Biomedical Ontology (NCBO), hosting and curating all OBO ontologies, the World Wide Web Consortium (W3C) and the LarKC consortium will ensure that also data sources based in North America and elsewhere will be in sync with the Open PHACTS consortium. The strong links to two major learned societies, the European Federation of Medicinal Chemistry and the Royal Society of Chemistry ensure maximum **awareness** and **dissemination** of the results throughout Europe.

The role of the pharmaceutical companies in providing data and tools as well as in driving the exemplar pilot services via prioritized research questions is crucial for the broad acceptance and the success of the Open PHACTS Discovery Platform.

All the partners have singular characteristics which are extremely valuable for the Open PHACTS consortium. As outline in the following table, tasks and contributions are well distributed among the whole consortium and each work package is covered by 19 – 28 partners. This ensures that the success of a distinct WP does not solely rely on the performance of a single partner.

Table 4.3a: Distribution of contributions among the work packages

| NAME | WP 1 | WP 2 | WP 3 | WP 4 | WP 5 | WP 6 | WP 7 | WP 8 | WP 9 |
|------------------|------|------|------|------|------|------|------|------|------|
| GSK | x | x | x | M | M | x | x | WPL | WPL |
| UNIVIE | x | x | x | x | M | WPL | xM | M | WPL |
| DTU | | | | M | M | M | | | x |
| UHAM | x | x | x | x | M | x | | | x |
| BIT | | x | x | x | M | x | x | | x |
| PSMAR | x | x | x | M | WPL | M | x | x | x |
| LUMC | M | M | M | M | | x | WPL | x | x |
| RSC | x | x | x | WPL | | M | M | WPL | x |
| VUA | M | | WPL | M | x | x | | | x |
| CNIO | | | | | M | | | | x |
| UNIMAN | WPL | WPL | M | M | | M | | | x |
| UM | M | x | x | M | | x | x | x | x |
| ACK ¹ | | M | | x | | | | x | x |
| USC | x | M | | M | M | M | | x | x |
| UBO | M | x | | M | | | | | x |
| AZ | M | | x | x | WPL | x | x | x | x |
| Pfizer | x | | WPL | x | x | x | WPL | x | x |
| Esteve | | x | | x | x | x | x | | |
| Novartis | | | | | x | x | | x | |
| ME | | | x | | M | x | | | |
| HLU | x | WPL | | x | | | | x | x |
| Lilly | x | x | x | x | x | x | x | x | x |
| NBIC | M | x | x | x | | | x | x | x |
| SIB | x | x | x | WPL | x | x | x | x | x |
| ConnDisc | x | x | WPL | | x | x | WPL | x | x |
| EML-EBI | M | | M | M | | x | | | x |
| Janssen | WPL | x | x | x | M | WPL | x | x | x |
| OGI | | | x | M | | x | x | | x |
| OPF | | | | | | x | x | x | x |
| ALM | | | x | x | x | x | x | x | x |
| SciBite | x | x | M | x | x | x | | | x |

WPL: work package leader; **M:** major contribution; **x:** contribution

¹ till 17.09.2014

5. Ethical Issues

There are no direct or immediate ethical issues in the proposed approach to the development of open knowledge management. Open PHACTS will not deal with clinical information (e.g. patient medical records, personal information from health control groups, biobank resources) and consequently will not be affected by the specific restrictive regulations that apply to this type of information. The project does not propose to hold any person data which is not already in the public domain, and the partner institutions - although expected to be already adhering to national data protection legislation - will have this made clear in any Memorandum of Understanding for involvement in the project. Any personal data needed for carrying out the Open PHACTS project shall be processed in accordance with the Directive 95/46/EC, and any applicable national legislation.

As *Open PHACTS* will allow better *in silico* reasoning it may significantly reduce the fall out of compounds in later stages, and therefore it may have a profound effect on the reduction of the number of animals needed for testing.

5.1 Ethical issues table

| | YES | PAGE |
|---|-----|------|
| Research on Humans | | |
| Does the proposed research involve children? | | |
| Does the proposed research involve patients? | | |
| Does the proposed research involve patients or persons not able to give consent? | | |
| Does the proposed research involve adult healthy volunteers? | | |
| Does the proposed research involve Human Genetic Material? | | |
| Does the proposed research involve Human biological samples? | | |
| Does the proposed research involve Human data collection? | | |
| I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL | YES | |
| Research on Human embryo/foetus | | |
| Does the proposed research involve Human Embryos? | | |
| Does the proposed research involve Human Foetal Tissue / Cells? | | |
| Does the proposed research involve Human Embryonic Stem Cells? (hESCs) | | |
| Does the proposed research on hESCs involve cells in culture? | | |
| Does the proposed research on hESCs involve the derivation of cells from Embryos? | | |
| I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL | YES | |
| Privacy | | |
| Does the proposed research involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)? | | |
| Does the proposal involve tracking the location or observation of people? | | |
| I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL | YES | |
| Research on Animals | | |
| Does the proposed research involve research on animals? | | |
| Are those animals transgenic small laboratory animals? | | |
| Are those animals transgenic non-rodents? | | |
| Are those animals transgenic farm animals? | | |
| Are those animals cloned farm animals? | | |
| Are those animals non-human primates? | | |
| I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL | YES | |
| Research Involving Developing Countries | | |
| Does the proposed research involve the use of local resources (genetic, animal, plant etc) | | |
| Is the proposed research of benefit to local communities (e.g. capacity building, access to healthcare, education, etc)? | | |
| I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL | YES | |
| Dual Use | | |
| Research having direct military application | | |
| Research having the potential for terrorist abuse | | |
| I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL | YES | |