

## Mobile applications driven by Open PHACTS semantic web technology

Christine Chichester<sup>1</sup>✉, Lee Harald<sup>2</sup>, Tim Harder<sup>3</sup>

<sup>1</sup>Swiss Institute of Bioinformatics, CALIPHO group, University of Geneva, Geneva, Switzerland

<sup>2</sup>Connected Discovery Ltd., London, United Kingdom

<sup>3</sup>Center for Bioinformatics, University of Hamburg, Hamburg, Germany

Received 1 August 2013; Accepted 8 August 2013; Published 14 October 2013

**Competing interests:** the authors have declared that no competing interests exist.

### Abstract

The Open PHACTS project is a large-scale public-private partnership funded under the European Innovative Medicines Initiative. The aim of the project is to create a stable infrastructure that combines diverse biomedical databases and enables scientists to answer complex questions of relevance to drug discovery and human health. The data integration is achieved through semantic web technologies, with data modeled in RDF and served through a high performance triple store. The main point of interaction with the system is an API, designed to be “developer friendly” by using familiar web technologies such as REST services and JSON. The vision is the Open PHACTS API becomes the foundation of an “application ecosystem”, enabling anyone to create applications targeting different use-cases within this domain. Here, we present two examples of the use of the Open PHACTS API to create mobile scientific applications. The ChemBioNavigator is a tool that allows researchers to analyze and triage sets of bioactive small molecules. Physicochemical properties are used to separate the molecules into different areas of a graph, upon which biological information from the Open PHACTS database is overlaid. This visualization allows users to understand the relationship between these data and select the molecules with the best overall combinations of these properties. The second example is the iPharm application which is designed to address a common situation where a researcher needs a high-level overview on a drug or target. The user simply opens up the application, enters a drug name and is presented with key clinical and molecular data summarizing the action of that drug. An important aspect of both applications is that while they are based on semantic technologies, the user interfaces are specifically designed to look like “normal” applications. We demonstrate the combination of semantic data with natural user interfaces provides a powerful mechanism to address scientific data navigation.

### Motivation and Objectives

The Open PHACTS project is a large-scale public-private partnership funded under the European Innovative Medicines Initiative (IMI). The aim of the project is to create a stable infrastructure that combines diverse biomedical databases and enables scientists to answer complex questions of relevance to drug discovery and human health. The data integration is achieved through semantic web technologies, with data modeled in RDF and served through a high performance triple store. The main point of interaction with the system is an API, designed to be “developer friendly” by using familiar web technologies such as REST services and JSON. The vision is the Open PHACTS API becomes the foundation of an “application ecosystem”, enabling anyone to create applications targeting different use-cases within this domain. Here, we present two examples of the use of the Open PHACTS API to create mobile scientific applications. The ChemBioNavigator is a tool that allows researchers to analyze and triage sets of bioactive small molecules. Physicochemical properties are used to separate the molecules into different areas of a graph, upon which biological information from the Open PHACTS database is overlaid.

This visualization allows users to understand the relationship between these data and select the molecules with the best overall combinations of these properties. The second example is the iPharm application which is designed to address a common situation where a researcher needs a high-level overview on a drug or target. The user simply opens up the application, enters a drug name and is presented with key clinical and molecular data summarizing the action of that drug. An important aspect of both applications is that while they are based on semantic technologies, the user interfaces are specifically designed to look like “normal” applications. We demonstrate the combination of semantic data with natural user interfaces provides a powerful mechanism to address scientific data navigation.

### Methods

The mobile applications presented in this work are based on the data retrieved from the Open PHACTS platform, which is essentially a semantic data integration platform. This semantic approach to data integration has been pioneered by other research efforts in the biomedical domain (Belleau *et al.*, 2008; Chen *et al.*, 2010; Hardy *et al.*, 2010; Hassanzadeh *et al.*, 2009), as

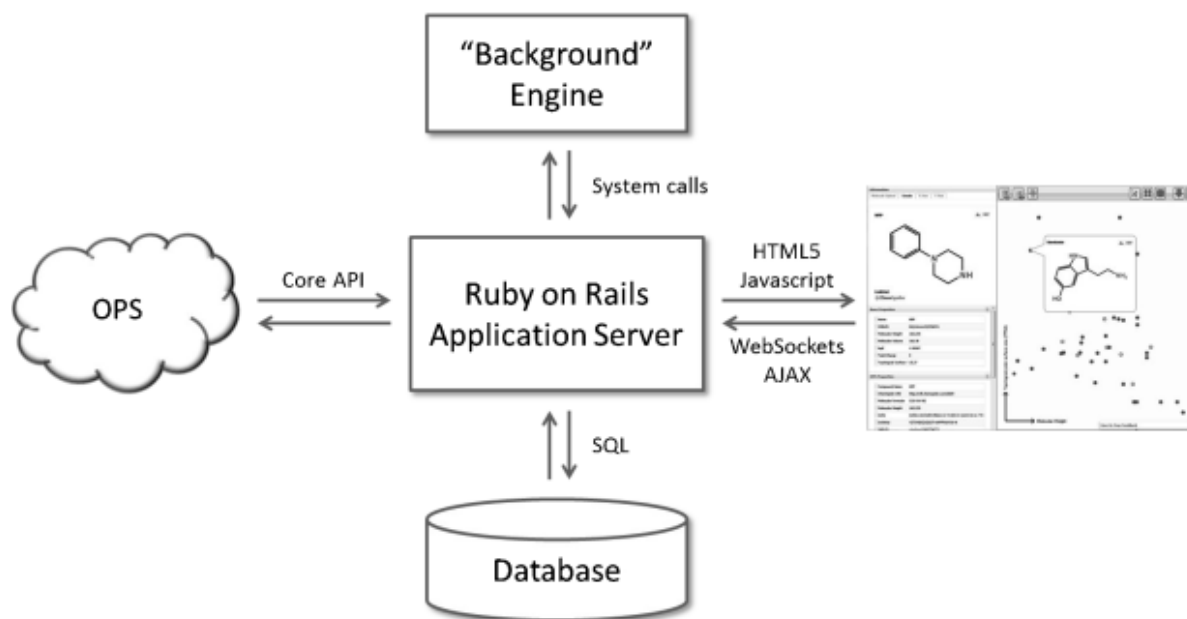


Figure 1. Schematic overview of the CBN technical architecture (Stierand *et al.*, 2012). The central application server connects the different parts of the system and handles the client communication.

well as by [LinkedLifeData](http://linkedlifedata.com/)<sup>1</sup> and the [World Wide Web Consortium Semantic Web Health Care and Life Sciences Interest Group](http://www.w3.org/blog/hcls/)<sup>2</sup>. The platform provides a semantically integrated view of the data by exploiting an identity mapping service (IMS) to construct appropriate responses based on the contextual aspects of each query. For instance, will users differentiate to the same level of granularity? *i.e.* do they want results for both genes and proteins. The platform also contributes to performance by caching the semantic representations of frequently used datasets. Additionally, an important aspect of the platform is the ability to integrate datasets and services; for example, the integration of the ConceptWiki for resolving textual queries and the possibility for community curation.

Although the behind the scenes semantic technology is required to provide the data, very important to creating the mobile applications is the rich Open PHACTS API. This interface provides REST-style based interfaces for common queries. These queries are defined in agile cycles working with application and user interface developers. By driving API development from the user per-

spective, Open PHACTS reduces the complexity of data integration by focusing on those core semantic types and properties that are necessary for the end-users. A key difficulty in many linked data solutions today has been the threat to performance caused by difficult or challenging queries and it is not yet clear in many current prototypic applications how to deal with complexities in the representation of facts in RDF. By adopting a well-defined API, the SPARQL queries can be optimised to ensure that such performance difficulties do not occur. Finally, the API provides rich information and access to the provenance of the results it returns. This includes the data sources from which those results are provided including the ability to track licensing information. Currently, the platforms includes datasets from: [DrugBank](http://www.drugbank.ca/)<sup>3</sup>, [ChEMBL](https://www.ebi.ac.uk/chembl/)<sup>4</sup>, [SwissProt/UniProt](http://www.uniprot.org/)<sup>5</sup>, [ChEBI](http://www.ebi.ac.uk/chebi/)<sup>6</sup>, [Gene Ontology](http://www.geneontology.org/)<sup>7</sup>, [GOA](http://www.ebi.ac.uk/GOA/)<sup>8</sup>, [Wikipathways](http://wikipathways.org/)<sup>9</sup>, [ChemSpider](http://www.chemspider.com/)<sup>10</sup> and [ConceptWiki](http://ops.conceptwiki.org/)<sup>11</sup>.

<sup>3</sup> <http://www.drugbank.ca/>

<sup>4</sup> <https://www.ebi.ac.uk/chembl/>

<sup>5</sup> <http://www.uniprot.org/>

<sup>6</sup> <http://www.ebi.ac.uk/chebi/>

<sup>7</sup> <http://www.geneontology.org/>

<sup>8</sup> <http://www.ebi.ac.uk/GOA/>

<sup>9</sup> <http://wikipathways.org/>

<sup>10</sup> <http://www.chemspider.com/>

<sup>11</sup> <http://ops.conceptwiki.org/>

<sup>1</sup> <http://linkedlifedata.com/>

<sup>2</sup> <http://www.w3.org/blog/hcls/>

Technically, the main challenges for the mobile applications using Open PHACTS is to design an architecture that is able to handle remote querying of the platform and at the same time provide an intuitive and responsive user interface. Further, the applications should be flexible enough to be able to adapt to possible future request such as additional data sources like proprietary algorithms or databases. Specifically for the ChemBioNavigator (CBN) example, the architecture is designed around a Ruby on Rails web-application server, connecting the different parts of the application (Figure 1) and modern web technologies such as HTML5, JavaScript and AJAX/JSON.

## Results and Discussion

Delivering an API for building applications based on the Open PHACTS platform has resulted in much knowledge into the technical challenges concerned with the large amounts of data, the diverse and often non-standard formats along with the use of different biological identifiers in most data sources and what is needed for ease of adoption for developers. Clearly, these results: developing a deeper understanding of the application of Semantic Web technologies to the life sciences, the integration and mapping of disparate data types and sources, the influence of data quality, the crowd-sourcing opportunities and as well as the processes and approaches necessary to deliver a groundbreaking technology platform to application developers are results that will help refine and augment the Open PHACTS platform going forward.

Applying modern web technologies in the development of mobile applications allows the seamless integration across different platforms

and even device types as shown with CBN and iPharm. These technologies accommodate the growing amount of tablet computers and other mobile devices for easy access to the necessary web services and enable the average bench scientist to easily explore large amounts of data, augmenting their research results with new information. The mobile applications can thus make the information accessible wherever needed, at the office computer or the tablet in a meeting or at the lab bench.

## Acknowledgements

This work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115191 for Open PHACTS, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. We greatly appreciate the efforts from all the partners in the entire Open PHACTS consortium.

## References

- Belleau F, *et al.* (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Inform.* **41**, 706–716. doi:10.1016/j.jbi.2008.03.004.
- Chen B, *et al.* (2010) Chem2Bio2RDF: A semantic framework for linking and datamining chemogenomic and systems chemical biology data. *BMC Bioinformatics* **11**, 255. doi:10.1186/1471-2105-11-255.
- Hardy B, *et al.* (2010) Collaborative development of predictive toxicology applications. *J. Cheminform.* **2**, 7. doi:10.1186/1758-2946-2-7.
- Hassanzadeh O, *et al.* (2009) LinkedCT: A linked data space for clinical trials. <http://arxiv.org/abs/0908.0567>
- Stierand K, Harder T, *et al.* (2012) The Internet as Scientific Knowledge Base: Navigating the Chem-Bio Space. *Molecular Informatics* Special Issue: Open Innovation in Drug Discovery **31** (8), 543-546. doi:10.1002/minf.201200037.