

# Answering drug-discovery research questions with Open PHACTS and KNIME

Daniela Digles and Gerhard F. Ecker

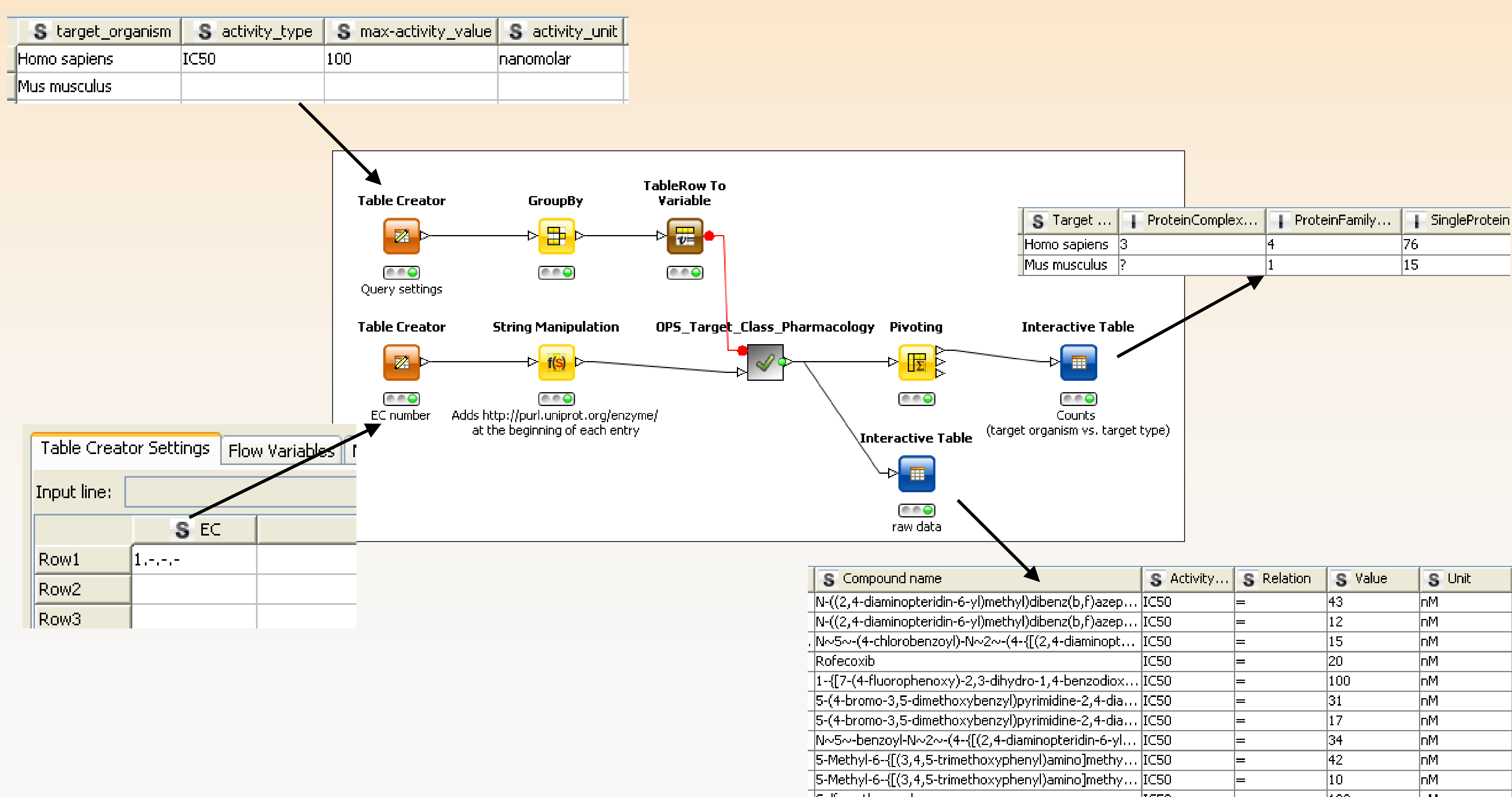
University of Vienna, Department of Medicinal Chemistry, Althanstraße 14, 1090, Vienna

## Introduction

A recent paper by Azzaoui et al. [1] describes scientific competency questions which are relevant for research in the drug-discovery process. While many of those questions can be answered with public databases, the integration of these data sources is a challenging task that is tackled by the Open PHACTS project [2]. The data from the project can be accessed with web tools, such as the Open PHACTS Explorer ([www.openphacts.org/explorer](http://www.openphacts.org/explorer)) or by using a programming interface (available from [dev.openphacts.org](http://dev.openphacts.org)). To answer questions requiring several different queries in an automated way without needing programming knowledge, workflow tools such as KNIME or Pipeline Pilot are a valuable resource. Here, tasks (such as reading a file or performing a calculation) are added as an icon (called a node), which are connected by arrows. Once a workflow is created, it can easily be reused for new data.

## The workflow

The target of interest of the first research question are all oxidoreductases. This includes all proteins which belong to the enzyme class 1. The Open PHACTS system (OPS) has a special query to return activity values for a whole class. The resulting workflow is shown below:



## Aim of the work

- Which of the proposed questions can be answered using Open PHACTS and workflow tools?
- Shown example: the implementation of the first ranked question (Give me all oxidoreductase inhibitors active <100 nM in human and mouse).

## Methods

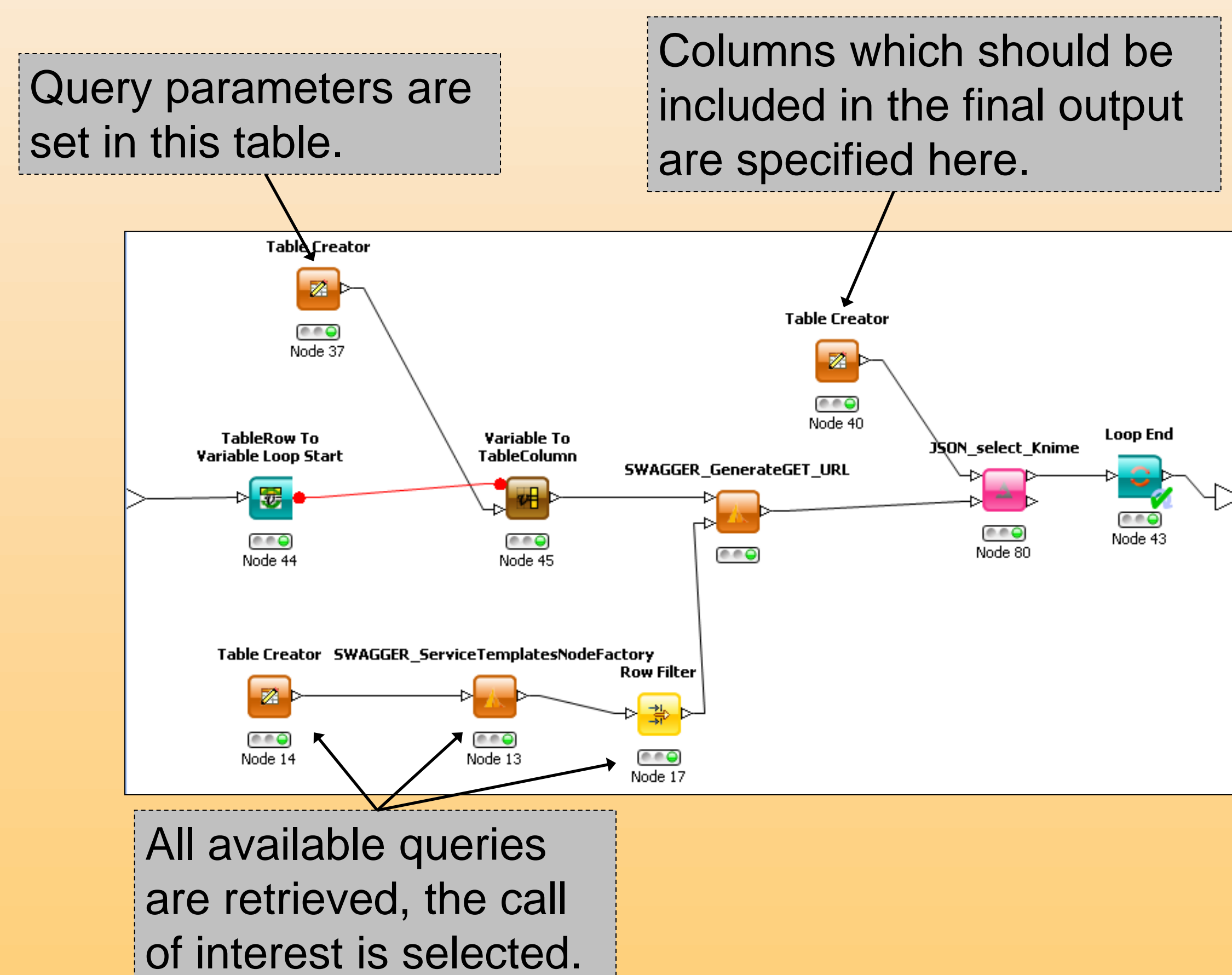
Workflows were created with KNIME.

Open PHACTS nodes developed by Ronald Siebes (VU University Amsterdam), which are available from <https://github.com/openphacts/OPS-Knime>, can be used to implement queries from Open PHACTS.



Explanation: The EC number of interest can be easily specified in the first table of the workflow (e.g. 1.-.-.- for the oxidoreductases). In a second table, query parameters can be specified. Several target organisms can be set here, as well as limits for the wanted activity type. Then the workflow automatically performs the necessary steps and searches for data in OPS.

## General implementation of API calls with KNIME nodes



## Results

On principle, retrieving the pharmacology data for all oxidoreductases can be performed easily with the Open PHACTS Explorer, or the programming interface. However, workflow tools allow to easily work with the retrieved data (e.g. to perform statistical analysis). For the shown research question, pharmacology data for 91 single protein targets, 5 protein families and 3 protein complexes were retrieved. A closer look at the retrieved data, however, shows that some proteins were erroneously classified as human. Therefore, additional filtering or curation steps will be necessary.

## Conclusions

Although some additional datasets (e.g. patent data) are still required to solve all of the proposed research questions, several can already be answered with Open PHACTS. As shown for the example of the oxidoreductases, workflow tools provide the possibility of combining several steps which are necessary to retrieve the final results.

## Acknowledgement:

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. [115191], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and in-kind contribution of EFPIA companies.

## References:

- [1] Azzaoui K, Jacoby E, Senger S et al. (2013) *Drug Discovery Today* 18: 843 – 852.
- [2] Williams A, Harland L, Groth P et al. (2012) *Drug Discovery Today* 17: 1188 – 1198.