

An open chemistry registry and mapping platform

Keywords: cheminformatics, software, chemical structures, graph theory, identifiers

Principle Investigator: Dr Egon Willighagen <<http://orcid.org/0000-0001-7542-0286>>

The Project: The Open PHACTS project (openphacts.org) is a European initiative, constituting a public-private partnership to enable easier, cheaper and faster drug discovery. The project is supported by the Open PHACTS Foundation and funded by contributions from several pharmaceutical companies. As part of Open PHACTS, a "Chemical Registration Service" was created to register chemicals of interest to the project, allowing compound linkage between data sets. A key concept is the support for "scientific lenses," which allows hierarchical mapping of chemical entities, including supporting characteristics such as charge state, tautomerism and stereochemistry. Open PHACTS aggregated various databases, including ChEMBL, ChEBI, DrugBank, UniProt, and WikiPathways. A new project builds on the Chemical Registration Service to establish an open chemistry registry and mapping service for general data set linkage. This expansion requires the support of multiple cheminformatics formats, the conversion and mapping of various identifiers, harmonized but configurable standardization, validation of the chemical structures, and the creation of new identifiers, to produce scientific lenses, or "link sets". The results will furthermore be repurposed for other knowledge bases.

Techniques: cheminformatics, Java, open science, version control

Literature:

Steinbeck C, et al. The Chemistry Development Kit (CDK): An open-source Java library for chemo-and bioinformatics. *J. Chem. Inform. Comput. Sci.* 2003 Mar 24;43(2):493-500. Cites: 464

Williams AJ, et al. Open PHACTS: semantic interoperability for drug discovery. *Drug Disc. Today.* 2012 Nov 30;17(21):1188-98. Cites: 118

Kutmon M, et al. WikiPathways: capturing the full diversity of pathway knowledge. *NAR.* 2015 Oct 19;44(D1):D488-94. Cites: 44

Kohonen P, et al. A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. *Nature Comm.* 2017;8. Cites: 0

Guha R, et al. The Blue Obelisk - interoperability in chemical informatics. *J. Chem. Inf. Model.* 2006 May 22;46(3):991-8. Cites: 298