



Public Release of SureChEMBL data in the Open PHACTS Discovery Platform

11 May 2016

Cambridge, UK –The Open PHACTS Foundation is proud to announce the official public release of SureChEMBL data in the Open PHACTS Discovery Platform. This marks the first time that SureChEMBL data has been made available to query programmatically via an API.

SureChEMBL (<https://www.surechembl.org/>) is a patent chemistry resource made freely available by the ChEMBL group at EMBL-EBI. SureChEMBL uses a live and automated cloud-based pipeline that combines full-text and image mining to extract chemical annotations from patent documents, convert them to compounds, and make them readily searchable and publicly available.

The latest release of the Open PHACTS Discovery Platform includes data extracted from approximately 3.6 million full-text life-science-related patents, published between 1975 and 2015 by the EPO, WIPO and USPTO authorities. As well as SureChEMBL chemistry annotations, the Open PHACTS dataset also includes annotations of biological entities such as genes and diseases.

A custom data model was developed to represent SureChEMBL patent annotations in RDF, linking biological and chemical entities to relevant patents and to other entities. Brand new Open PHACTS API calls allow researchers to query this data alongside other Open PHACTS datasets, giving researchers programmatic access to SureChEMBL data for the first time.

To mitigate the fact that patents are an inherently obfuscated and thus noisy data source, the Open PHACTS and ChEMBL teams have worked with SciBite to develop and validate algorithms to assess the relevance of each annotated entity within a particular patent, enabling ranking and filtering to reduce noise and spurious results.

This addition to the Open PHACTS Discovery Platform represents a powerful new resource for researchers in drug discovery and other areas of life science.

About

The Open PHACTS Foundation

The Open PHACTS Foundation is a not-for-profit membership organisation, established to ensure the sustainability of the Open PHACTS Discovery Platform and other outcomes of the Open PHACTS project. The Open PHACTS Discovery Platform was built to answer key scientific questions for applied life sciences research and development, and to meet the needs of drug discovery scientists. The data integrated by Open PHACTS currently contains over three billion triples, and the diversity and size of data sets are growing rapidly.

The Open PHACTS Foundation is a registered charity in the United Kingdom.

<http://www.openphactsfoundation.org/>

The Open PHACTS Project

Open PHACTS began as an Innovative Medicines Initiative (IMI) knowledge management project. Running from March 2011 to February 2016, the project received support from the IMI Joint Undertaking under grant agreement n° 115191.

Together the 30-member project consortium developed the Open PHACTS Discovery Platform, and laid the foundations for a strong community of researchers from academia, industry, and smaller enterprises, committed to sustaining and building on the project's achievements.

<http://www.openphactsfoundation.org/about/project-history/>

The Innovative Medicines Initiative (IMI)

The Innovative Medicines Initiative is Europe's largest public-private initiative. Resources are composed of a financial contribution from the EU Seventh Framework Programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in-kind contribution.

<http://www.imi.europa.eu/>