

Abstract

Toward Semantic Interoperability of Public Available Data Sources

Thomas Exner, Douglas Connect GmbH, Basel, CH thomas.exner@douglasconnect.com

Nowadays combining data from different sources is a time-consuming endeavor. First, the data must be searched using multiple interfaces, downloaded in the format provided by the database and then curated and harmonized to identify corresponding entries like compound name, concentration and finally biological effect. We will present a new concept of application programming interfaces (APIs) for data sources that can be added to already existing data warehouses or to wrap simple files and hold the data in tabular formats that are more accessible to both humans and computers. Using examples of reference sources like ToxRefDB, ToxCast and TG-GATES, we will demonstrate how these APIs can be used to generate tools for searching and browsing across all sources and how the selected data can then directly be integrated in analysis and modelling services using scripting languages like Python and R or workflow managers like KNIME. We will also show the first approaches to make these API semantically rich. Data schemas, i.e. descriptions of the data format easily understandable to humans and computers will be used to annotate important features of the datasets like compounds under investigation, concentrations and exposure times.

This work was developed within OpenRiskNet (Grant Agreement 731075), a project funded by the European Commission under Horizon 2020 Programme.

OpenTox USA 2017

<http://www.opentox.net/events/opentox-usa-2017/semantic-interoperability-of-public-available-data>