DReNIn: An RDF dataset for drug repositioning

Joseph Mullen

Interdisciplinary Computing and Complex BioSystems (ICOS) Research Group, School of Computing Science, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.

Abstract

Drug development is both increasing in cost whilst decreasing in productivity [1]. There is a general acceptance that the current paradigm of R&D needs to change. One alternative approach is that of drug repositioning, which focusses on the identification of novel uses for existing drugs. Many marketed examples of repositioned drugs have been identified through serendipitous or rational observations, highlighting the need for systematic methodologies. Systems approaches have the potential to enable the development of novel methods to understand the action of therapeutic compounds, but require an integrative approach to biological data [2]. Projects such as Open PHACTS integrate such data and present them in a common format, RDF, providing a holistic view of a drugs actions [3]. The Open PHACTS project has resulted in the improvement of many datasources, in content, format and accessibility; enabling other integration projects. In order to make informed drug repositioning inferences, however, data not currently captured in Open PHACTS, such as drug indications and clinical trial data may also prove beneficial. Here, we present DReNIn, an integrated RDF dataset for drug repositioning. DReNIn captures data describing drugs in relation to their effect on targets and diseases associated with *Homo sapiens*. Data in DReNIn is drawn from over 20 data sources, including pathways and multiple URIs from Open PHACTS. DReNIn has been developed as a platform to aid systematic data mining in the area of drug repositioning. Developed as a means of relating the different entities from a drug repositioning perspective, a high-level ontology, DReNIn_O, is also introduced. DReNIn_O semantically describes relationships between the 25 data types captured in DReNIn, including `Disease` (with child terms `Rare_Disease` and `Common_Disease`), `Drug_Molecule` (with child terms including `Small_Molecule`), `Biological_Molecule` (with child terms including `Protein` and `Gene`), and `Annotation` (with child terms including `Clinical_Trial`). DReNIn_O is used both during integration, as well as the querying of, DReNIn. We provide a SPARQL endpoint for querying DReNIn, which is accessible via a dedicated website found at www.drenin.ncl.ac.uk.

Keywords: Drug repositioning, Data integration, Data mining, RDF, SPARQL

[1] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, A. L. Schacht, How to improve r&d productivity: the pharmaceutical industry's grand challenge., Nature reviews. Drug discovery 9 (2010) 203–214.

[2] J. Mullen, S. Cockell, H. Tipney, P. Woollard, A. Wipat, Mining integrated semantic networks for drug repositioning opportunities, PeerJ 4:e1558 (2016).

[3] A. J. Williams, L. Harland, P. Groth, S. Pettifer, C. Chichester, E. L. Willighagen, C. T. Evelo, N. Blomberg, G. Ecker, C. Goble, B. Mons, Open phacts: semantic interoperability for drug discovery., Drug discovery today 17 (2012) 1188–1198.