DReNIn: an RDF dataset for drug repositioning

j.mullen@ncl.ac.uk

Joseph Mullen[1] , Simon J Cockell[2], [redacted][3] & Anil Wipat[1]

[1]ICOS, School of Computing Science and [2]Bioinformatics Support Unit, Newcastle University, Newcastle upon Tyne, UK

[3][redacted]

Introduction

Drug development is both increasing in cost whilst decreasing in productivity [1]. One complementary approach is that of drug repositioning. Systems approaches have the potential to enable the development of novel methods for drug repositioning but require an integrative approach to biological data [2]. The Open PHACTS project has resulted in the improvement of many datasources, in content, format and accessibility; enabling other integration projects [3]. To make informed drug repositioning inferences 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 developed for drug repositioning.

DReNIn Ontology

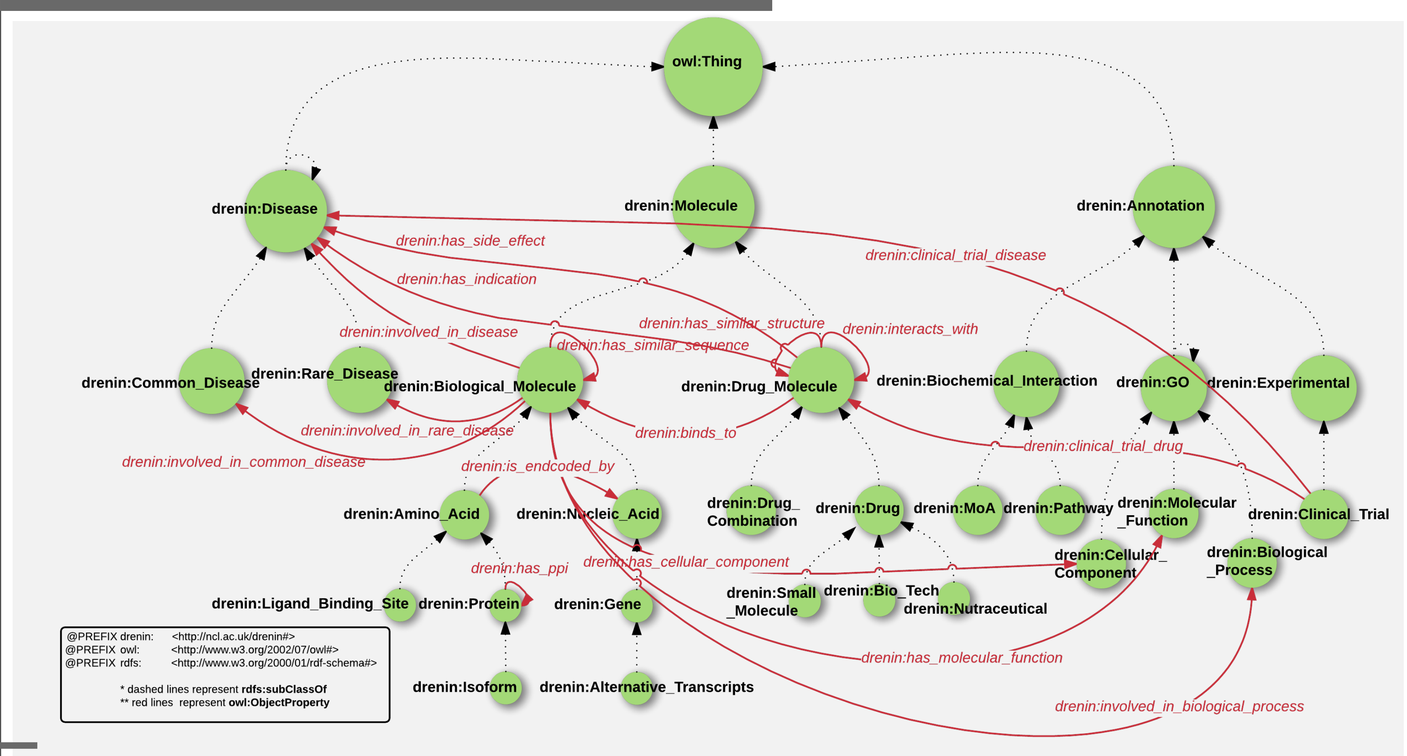


Fig 1. Visualisation of the classes included in the DReNIn ontology. The ontology contains 25 classes, representing the data types to be included in DReNIn, with relation types captured as 17 object properties. Available for download from: **http://bitbucket.org/ncl-intbio/drenin_ontology**

Integration Strategy

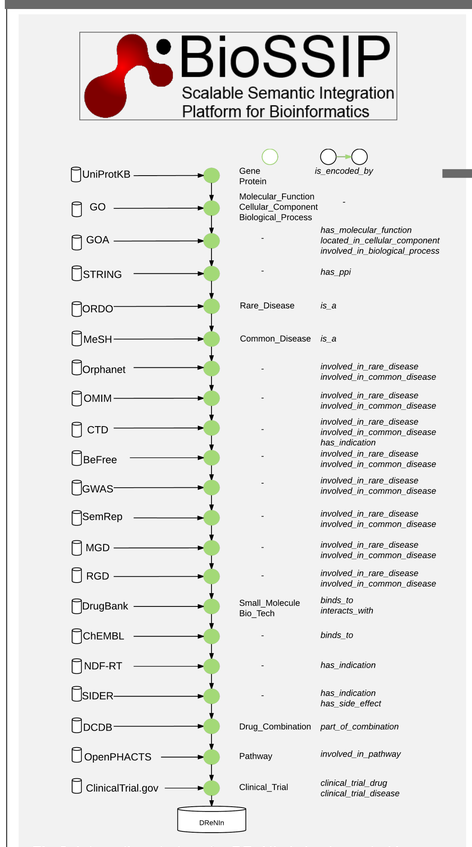


Fig 2. Integration strategy for DReNIn is implemented in BioSSIP. Makes use of all datasets listed as well as mappings resources (UniChem, PharmGKB and Metathesaurus). Green circles represent parsers which integrate the two input sources.

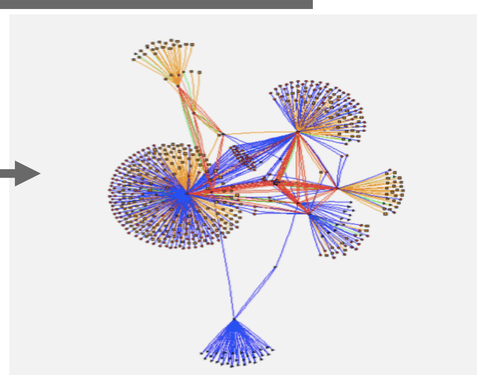


Fig 3. After integration a graph-based representation is produced. In this network the types of vertices and edges are defined by the DReNIn Ontology.

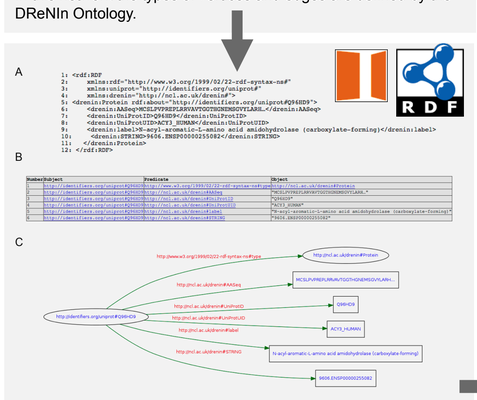
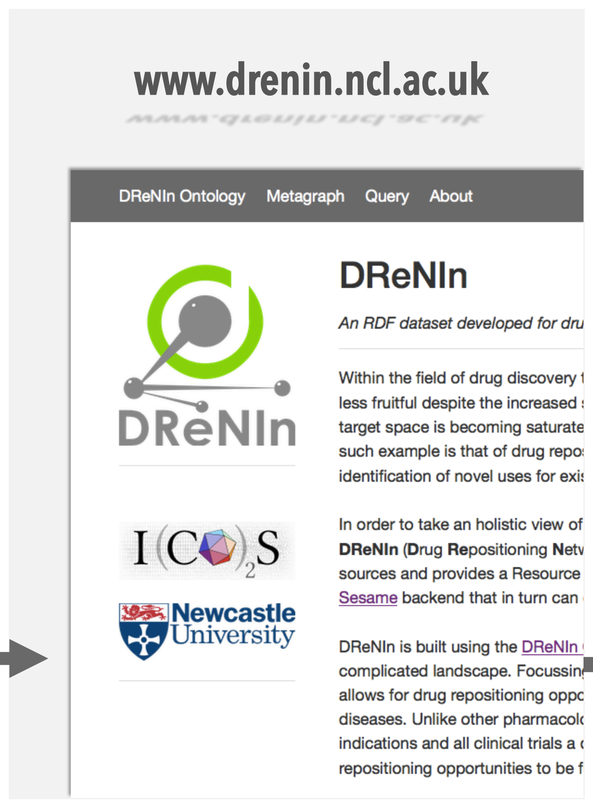


Fig 4. Graph is exported to RDF and stored in a Sesame triplestore. **A** an example XML serialised RDF document for a single protein in DReNIn. **B** triples from A in the form subject-predicate-object. **C** graph representation of RDF showing edge to every property value. *Note:* Images were produced using W3C RDF validation service



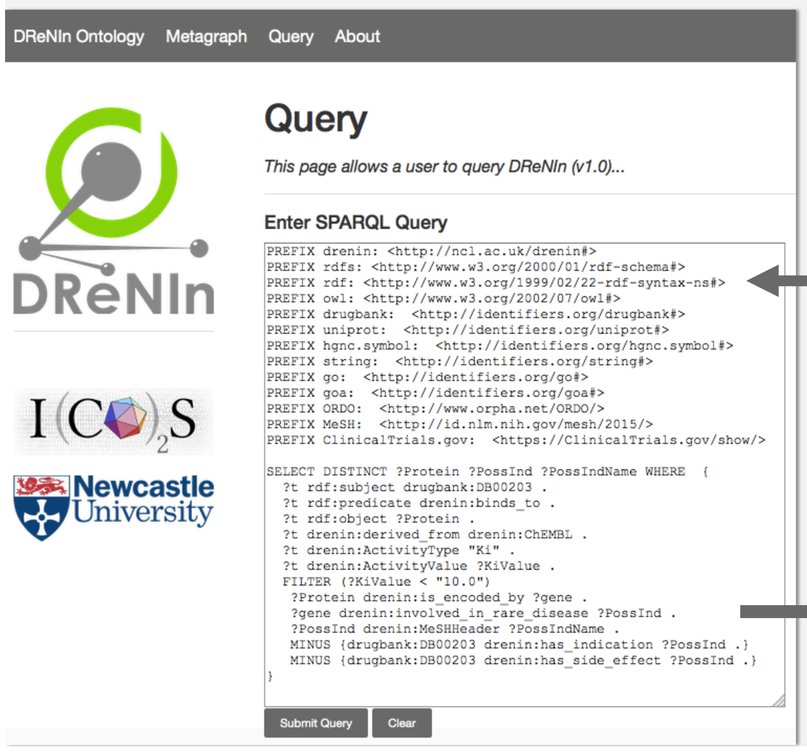
SPARQL endpoint

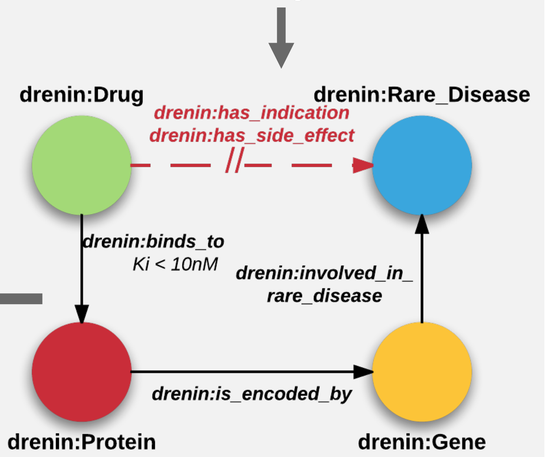
Although DReNin is made up of over 8.5 million triples it provides a simplified view of a complex world and can be queried directly using a SPARQL endpoint

What kind of questions can we ask DReNIn?

- **Indications** allows for areas of unmet need to be identified prior to any research efforts as well as the systematic validation of large, automated prediction tasks
- **Clinical trials** data allows for the novelty and validity of predicted indications to be investigated rapidly
- **Rare diseases** are captured in such a manner that allows them to be easily separated from Common disease

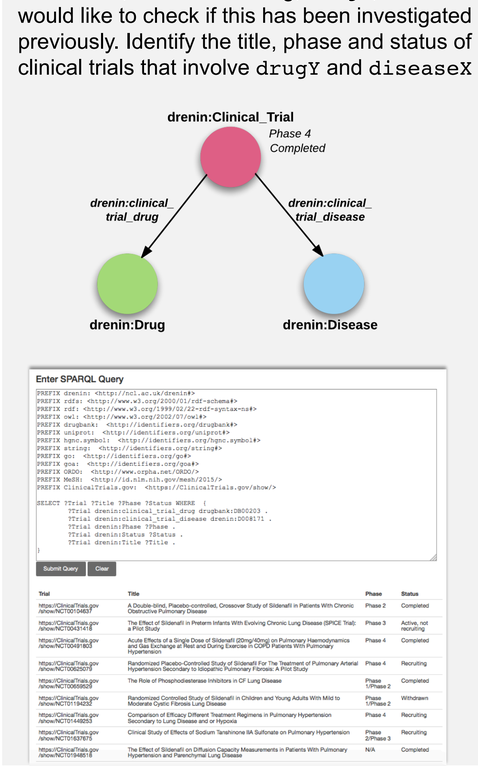
Example 1: I am interested in inferring novel uses for `drugX`. Identify all targets that `drugX` potently binds to and the rare diseases associated to these targets. Do not include diseases `drugX` is marketed to treat or are known side-effects of the drug



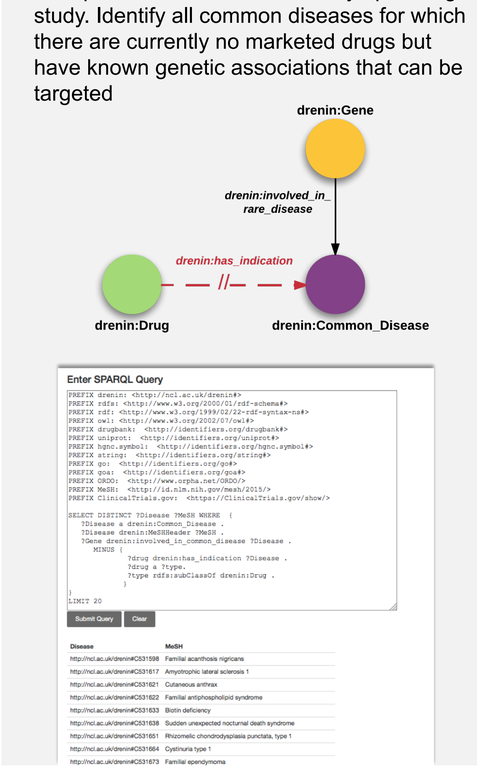


Protein	PossInd	PossIndName
http://identifiers.org/uniprot#O76074	http://www.orpha.net/ORDO/Orphanet_167848	Cardiomyopathies
http://identifiers.org/uniprot#O76074	http://www.orpha.net/ORDO/Orphanet_182095	Lung Diseases, Interstitial
http://identifiers.org/uniprot#O76074	http://www.orpha.net/ORDO/Orphanet_232	Anemia, Sickle Cell
http://identifiers.org/uniprot#O76074	http://www.orpha.net/ORDO/Orphanet_586	Cystic Fibrosis
http://identifiers.org/uniprot#O76074	http://www.orpha.net/ORDO/Orphanet_98845	Hodgkin Disease

Example 2: I have a potential indication, `diseaseX`, for a known drug, `drugY`, and I would like to check if this has been investigated previously. Identify the title, phase and status of clinical trials that involve `drugY` and `diseaseX`



Example 3: I would like to identify a therapeutic area of interest for my upcoming study. Identify all common diseases for which there are currently no marketed drugs but have known genetic associations that can be targeted



References

[1] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lind- borg SR, Schacht AL, How to improve R&D productivity: the pharmaceutical industry's grand challenge., *Nature reviews. Drug discovery* 9 (2010) 203–214.

[2] Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, Evelo CT, Blomberg N, Ecker G, Goble C, Mons B. (2012) Open PHACTS: semantic interoperability for drug discovery., *Drug discovery today* 17 1188–1198.

[3] Mullen J, Cockell SJ, Tipney H, Woollard PM, Wipat A. (2016) "Mining integrated semantic networks for drug repositioning opportunities". *PeerJ* 4:e1558 https://doi.org/10.7717/peerj.1558

EPSRC
Engineering and Physical Sciences Research Council