

DisGeNET: a discovery platform to support translational research and drug discovery

Janet Piñero, Núria Queralt-Rosinach, Àlex Bravo, Ferran Sanz and Laura I. Furlong

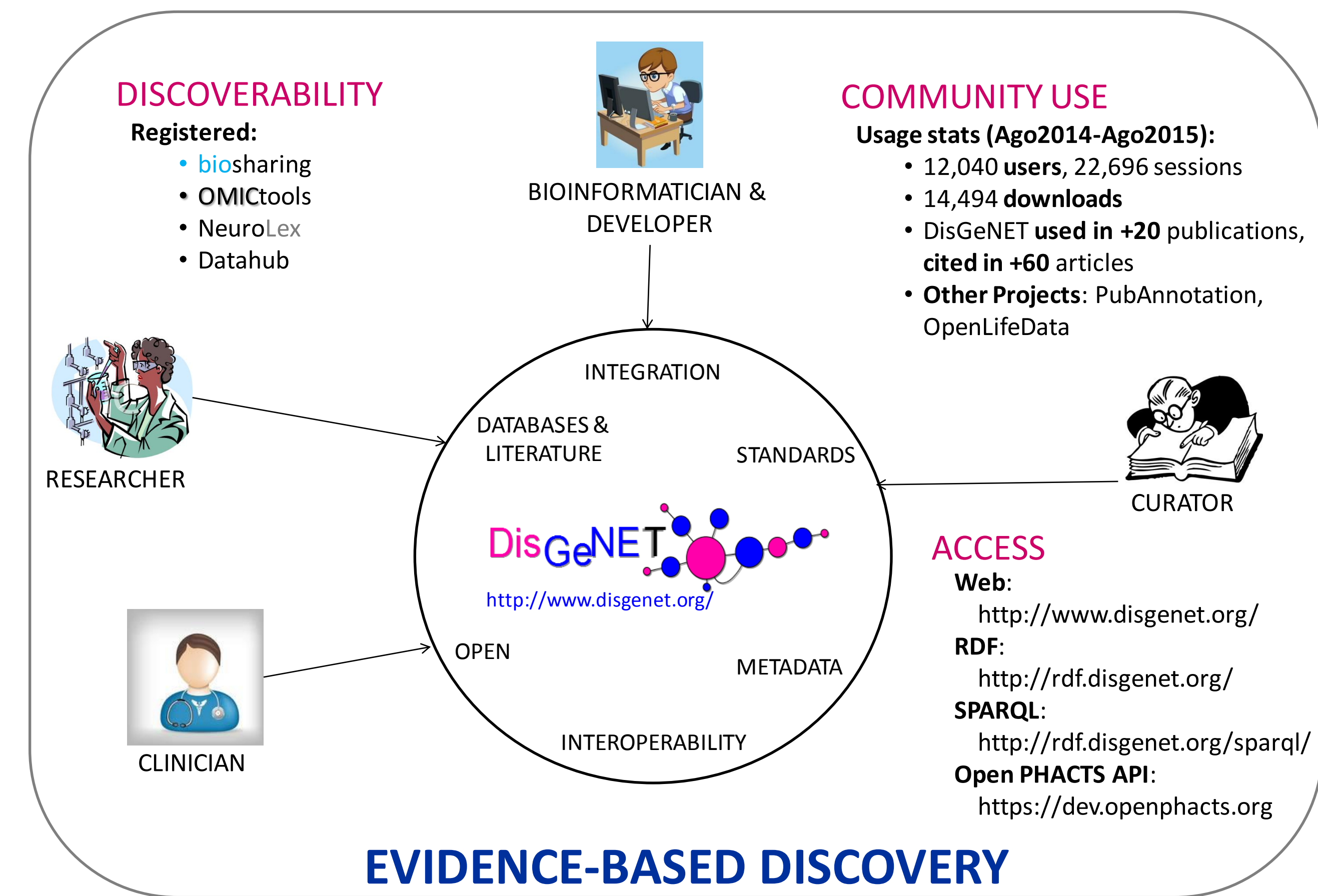
Integrative Biomedical Informatics Group, Research Programme on Biomedical Informatics; Hospital del Mar Medical Research Institute; Pompeu Fabra University

DisGeNET: Disease-Gene NETWORK of relations for discovery

Motivation: Better understanding of human gene component and disease mechanisms for translational research and drug discovery and development.

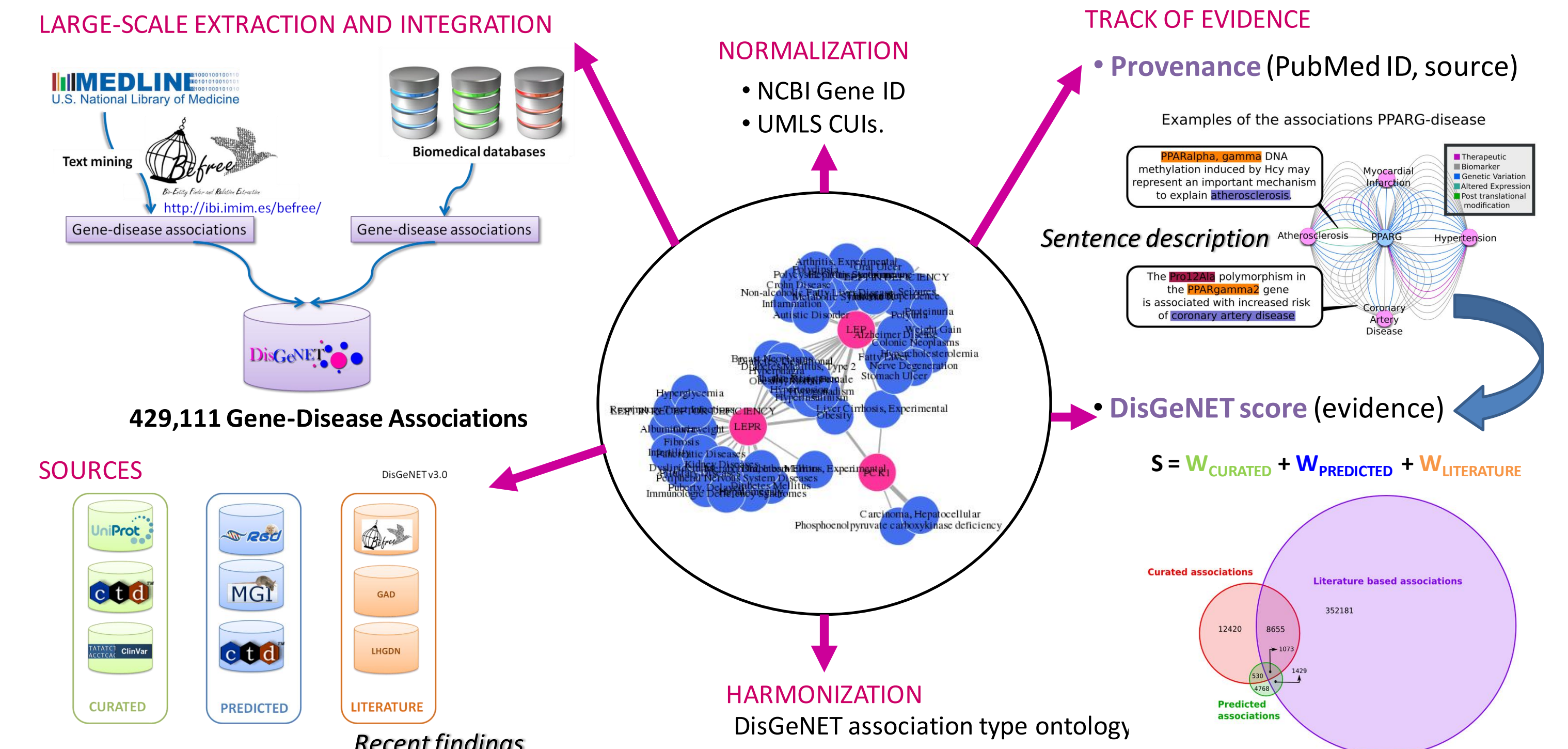
Challenge: One of the major current bottlenecks for knowledge discovery on the genetic component of diseases is that the information is fragmented. The vast amount of biomedical information about genotype-phenotype relations is distributed in several databases, represented and annotated using different data models, vocabularies and standards, and it is domain and technology-specific, which hampers their access, integration, analysis, and interpretation.

Approach: DisGeNET Discovery Platform¹ collects and integrates the available information on gene-disease associations (GDAs), covering the whole spectrum of human diseases, and using standards for their annotation and representation.

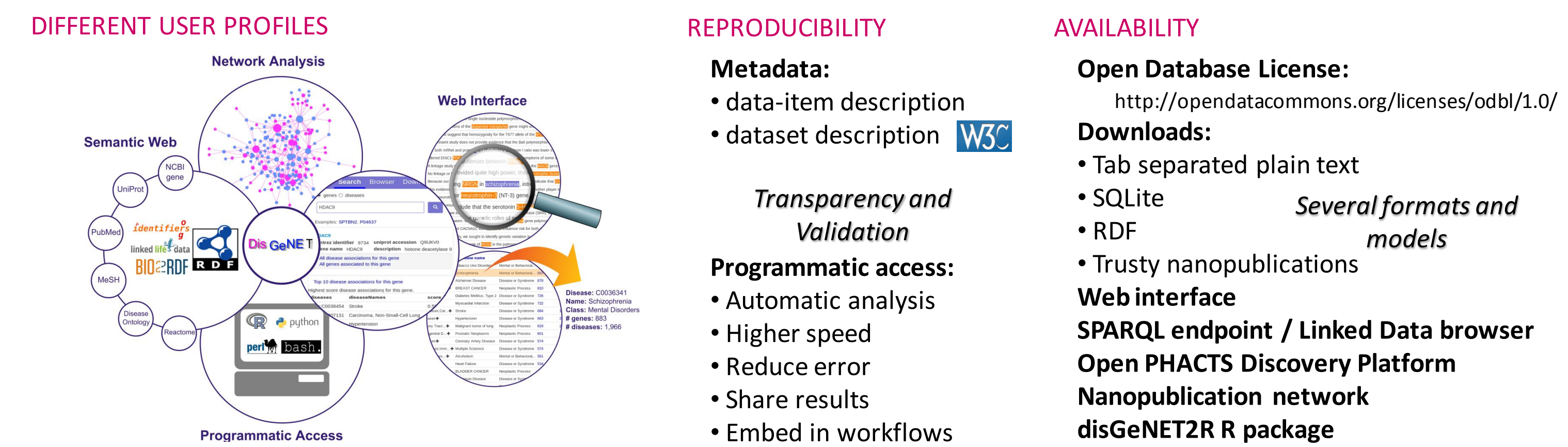


Implementation: The platform is composed of a knowledge base and a set of tools for data analysis and interpretation.

KNOWLEDGE BASE



TOOLS FOR EXPLORATION AND ANALYSIS



DisGeNET-RDF: a GDA Linked Open Data resource

DIGITAL PUBLICATION, SHARING AND LINKING

Present in the Semantic Web:

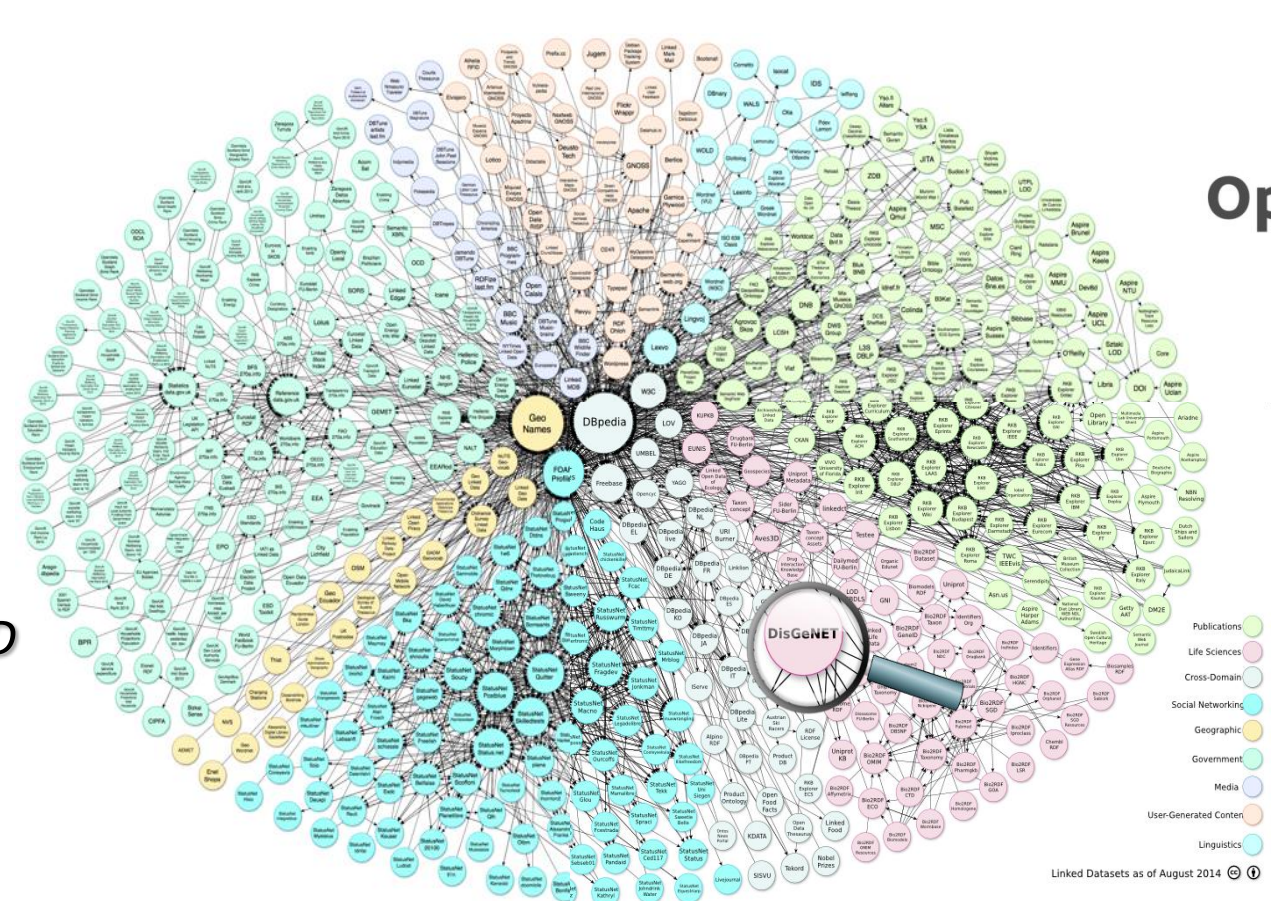
- URI/RDF/nanpublications
- Machine-processable
- Semantic integration
- Links to the Linked Open Data (LOD) cloud
- Data analysis across domains

4,962,315 RDF links to RDF datasets in the LOD



<https://datahub.io/dataset/disgenet>
(more statistics)

LOD cloud



<http://lod-cloud.net/>; Aug 2014

RDFIZATION



RDF

Open PHACTS

- Use Open PHACTS guidelines
- Dereferenceable URIs (primary or secondary)
- SIO

METADATA

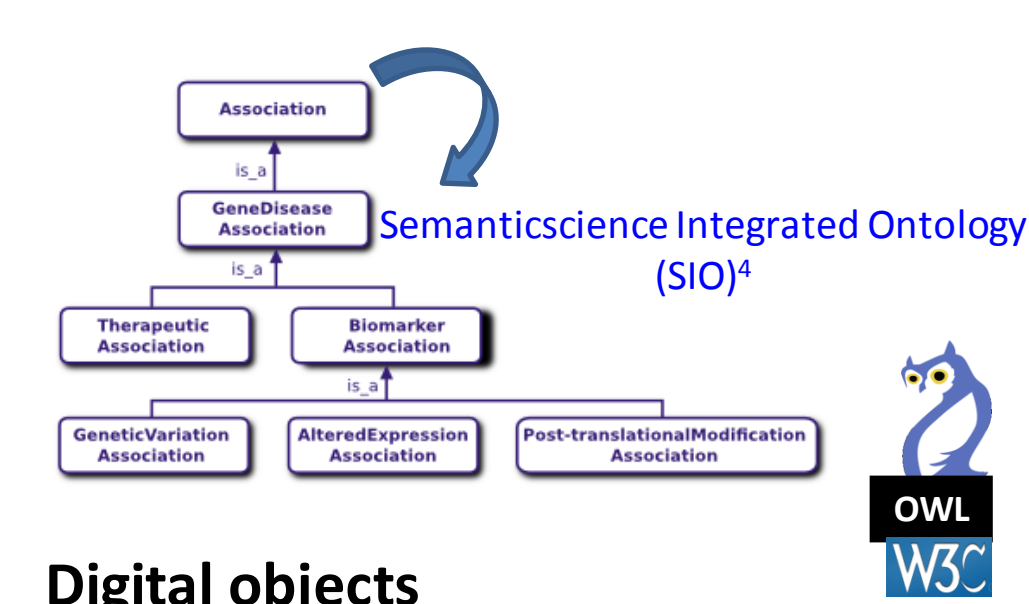
- Dataset (Open PHACTS + W3C)
- Linksets (Open PHACTS + W3C)

INTERLINKING

- BIO2RDF
- linked life data

STANDARDIZATION

DisGeNET association type ontology



Digital objects

- Normalized Identification Scheme
- <http://rdf.disgenet.org/resource/gda/> + ID

INTEROPERABILITY

COMMON IDs and ONTOLOGIES

- GENE: NCBI Gene ID, PANTHER Classification
- DISEASE: UMLS CUIs, MeSH Classification

SYNTACTIC

- RDF²
- Nanopublications³

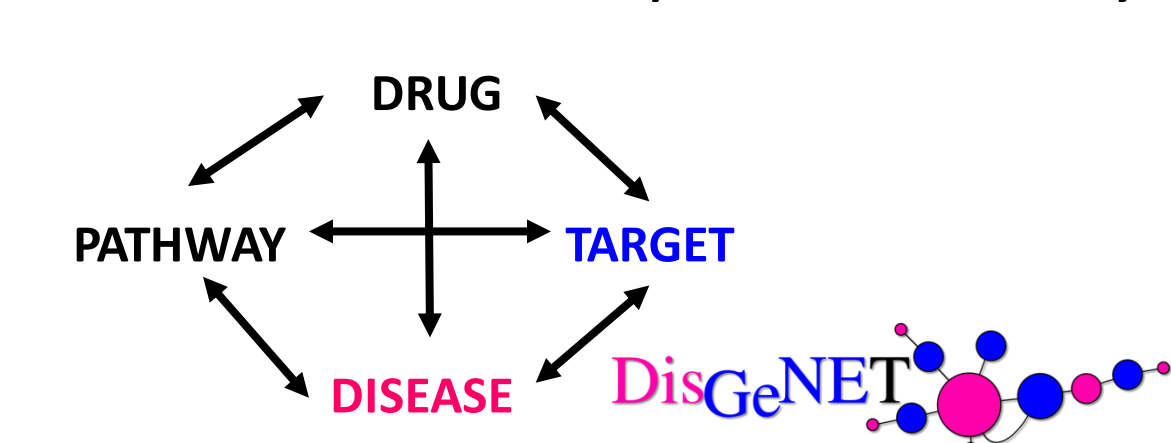
SEMANTIC

- 11 common ontologies in BioPortal

DisGeNET in the Open PHACTS Discovery Platform for drug discovery and development

DATA

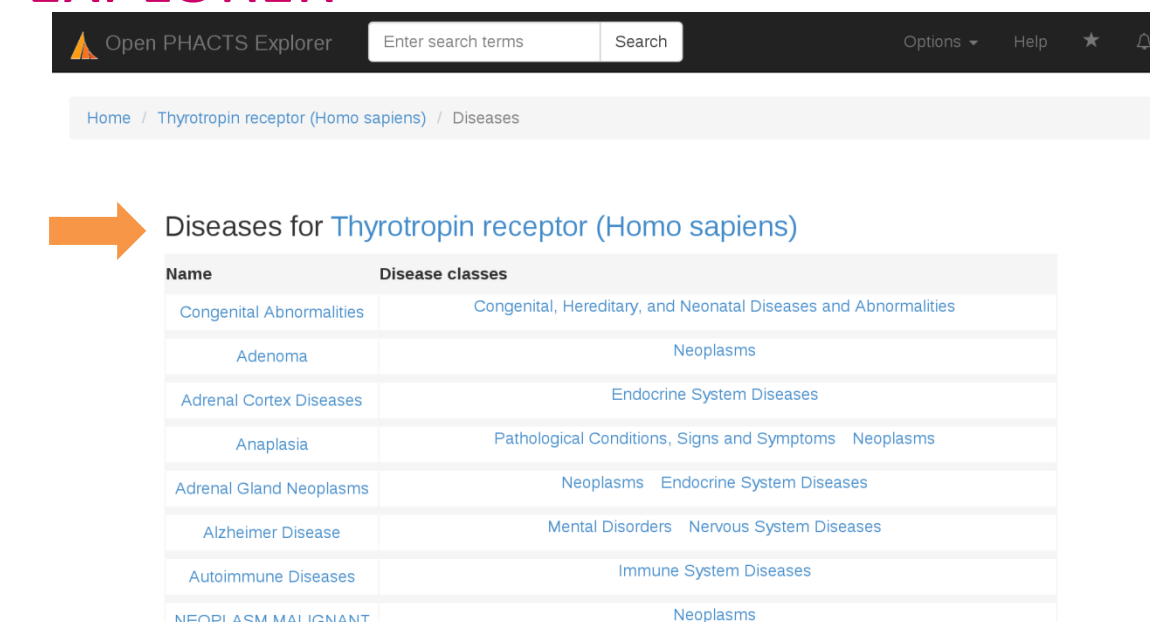
- Data providers
- Disease annotation in the Open PHACTS Discovery Platform⁵



- OMIM included
- > 20 000 000 of triples

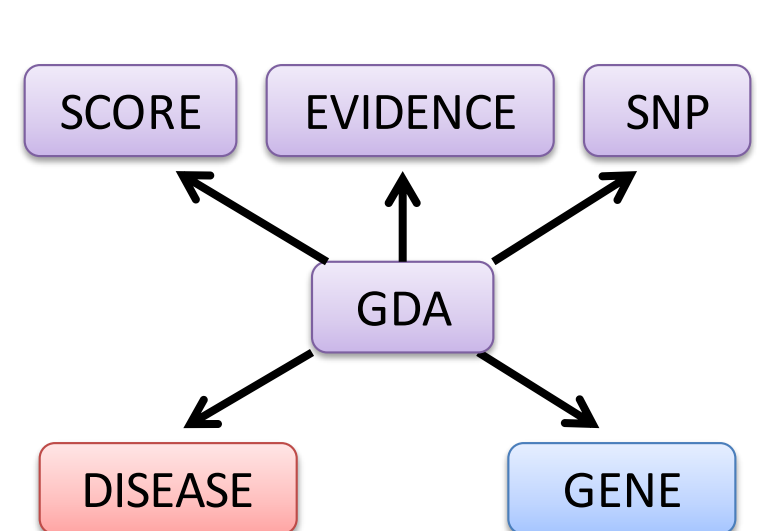
DISCOVERY

EXPLORER



RDF SCHEMA

- GDAs described by SIO



Gene-disease association as entity

METADATA

- Data item
- Dataset

<disease> <void:inDataset> <dgn-void:disease-dataset>

Summary Level Description

Version Level Description

Distribution Level Description

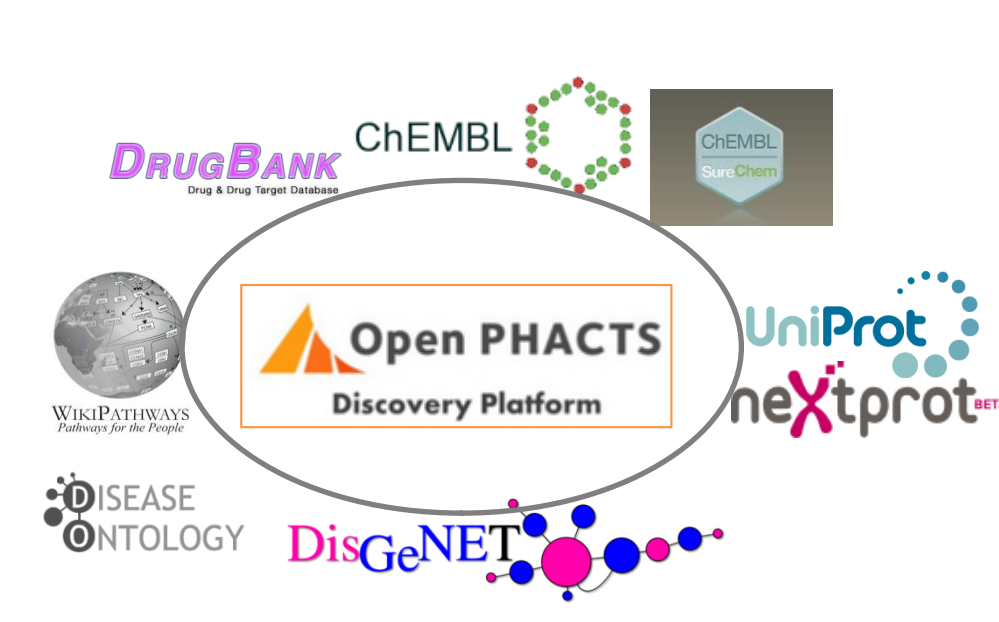
Version Number

Distribution Data

<http://rdf.disgenet.org/void-v3.0.0.ttl>

INTERLINKING

- Linksets providers
- > 70 000 number of linksets



FUTURE



NEW DATA:

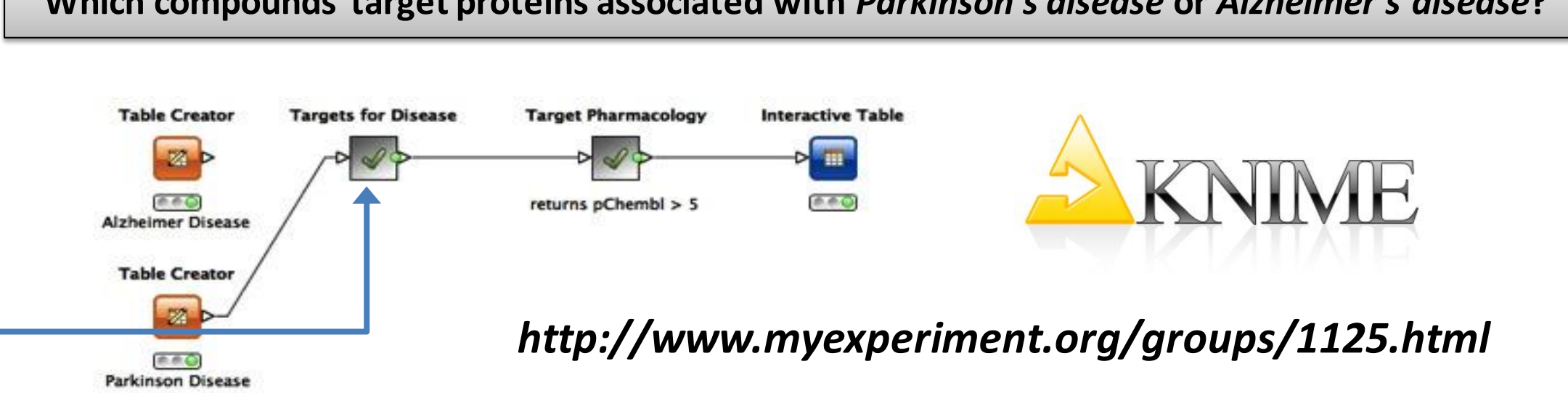
- Disease-phenotype associations (HPO)
- New use cases
- New API calls

Score:

- Add to API calls

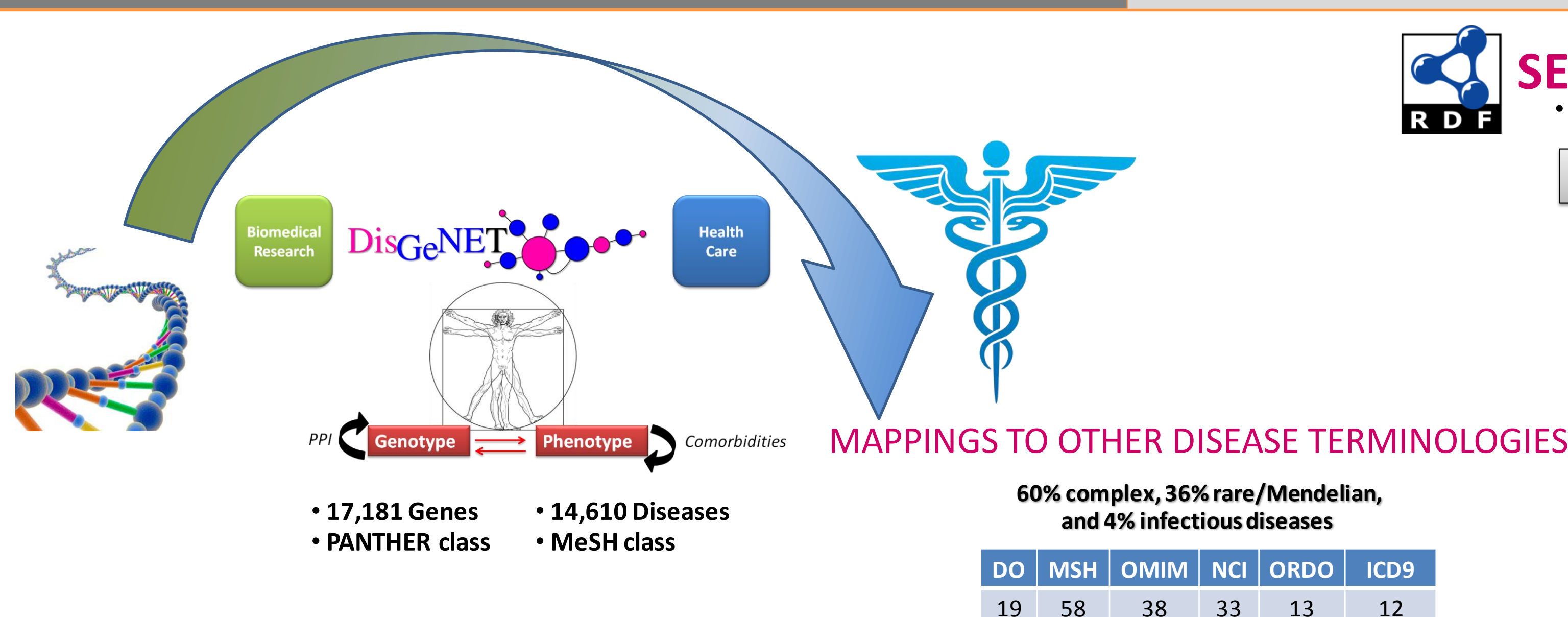
KNIME

Which compounds target proteins associated with Parkinson's disease or Alzheimer's disease?



<http://www.myexperiment.org/groups/1125.html>

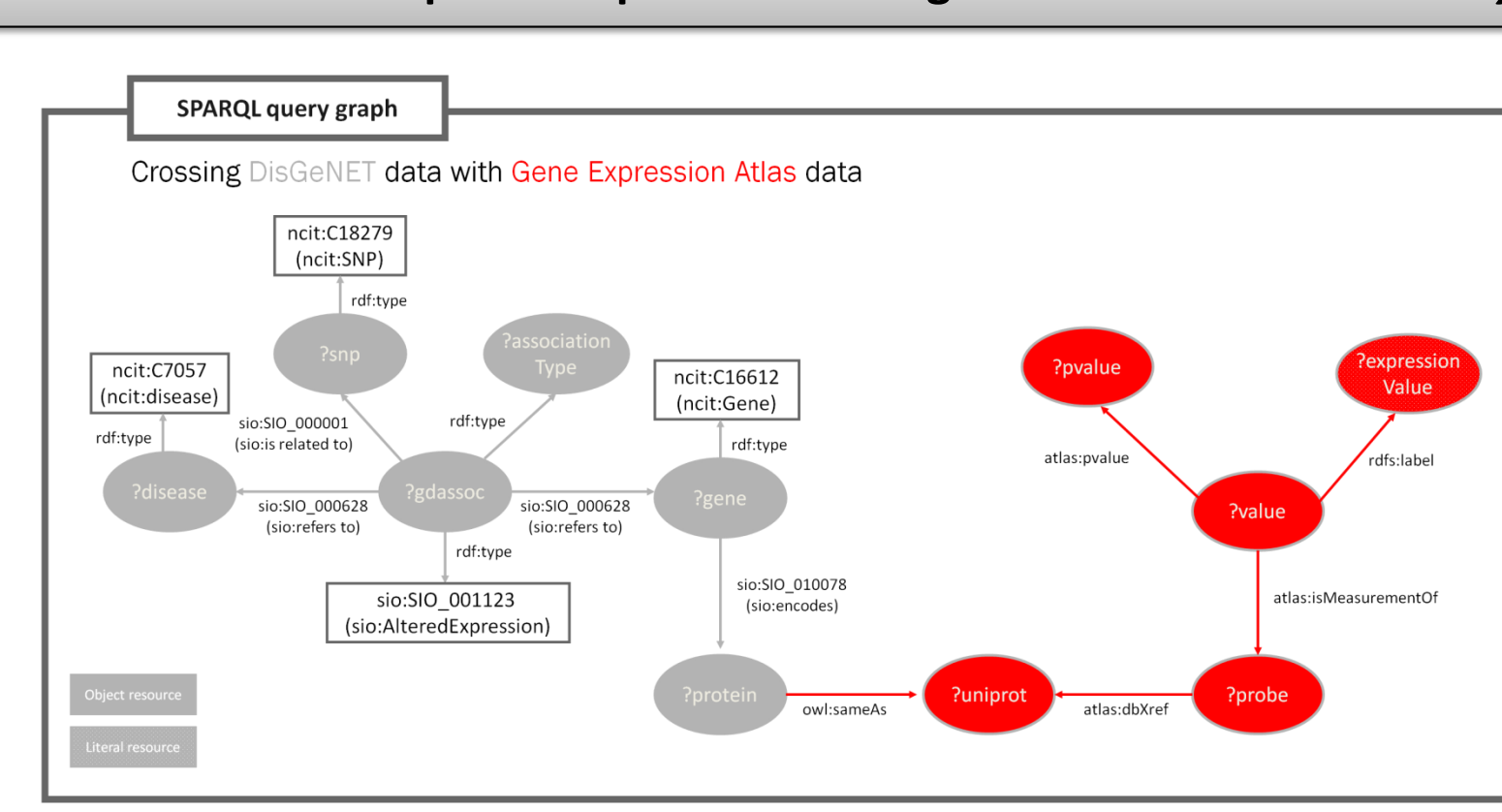
DisGeNET in the LOD cloud for translational research



SEMANTIC WEB

- Large-scale integration across domains

What is the tissue expression pattern of the genes associated to Obesity?



More @ <http://www.disgenet.org/web/DisGeNET/menu/rdf#sparql-queries-2>

- DisGeNET + external multidomain sources in LOD.

- It is interlinked to other biomedical databases to answer scientific questions that need the interrogation of cross-domain resources.

- It aims to support the development of bioinformatic Semantic Web applications to extract key knowledge on the molecular mechanisms of diseases.

References

1. Piñero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., ... Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database, 2015(0), bav028–bav028.
2. Queralt-Rosinach, N., Piñero, J., Bravo, A., Sanz, F. and Furlong, L. I. DisGeNET-RDF: harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases, 2015 (submitted).
3. Queralt-Rosinach, N., Kuhn, T., Chichester, C., Dumontier, M., Sanz, F., and Furlong, L. I., Publishing DisGeNET as Nanopublications. Semantic Web Journal, (to appear), 1–10, 2015.
4. Dumontier, M., Baker, C. J., Baran, J., Callahan, A., Chepelev, L., Cruz-Toledo, J., ... Hoehndorf, R. (2014). The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery. Journal of Biomedical Semantics, 5(1), 2014.
5. Gray, A. J. G., Groth, P., Loizou, A., Askjaer, S., Brenninkmeijer, C., Burger, K., ... Williams, A. J. (2014, January 1). Applying linked data approaches to pharmacology: Architectural decisions and implementation. Semantic Web. IOS Press. doi:10.3233/SW-2012-0088

Acknowledgements

The authors thank the Open PHACTS partners, Michel Dumontier and the OpenLink staff for their input, collaboration and help. Funding: We received support from ISCIII-FEDER (PI13/00082, CP10/00524), from the IMI-JU under grants agreements n° 115002 (eTOX), n° 115191 (Open PHACTS), n° 115372 (EMIF) and n° 115735 (iPE), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution, and the EU H2020 Programme 2014-2020 under grant agreements no. 634143 (MedBioinformatics) and no. 676559 (Elixir-Eccelerate). The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

