David Wild – Indiana University Bloomington & Data2Discovery Inc

Graduate program director, Data Science & Cheminformatics

IU IDSL-CCRG group researches semantic methods in drug discovery (and beyond)

Chem2Bio2RDF (2010)
SLAP (2012)

Data2Discovery provides semantic association finding capabilities for multiple markets, and is developing a phenotypic profiling toolset

http://idsl.soic.indiana.edu
http://d2discovery.com

Some observations

- The closer you get to people, the better the opportunities
  - $ Health Insurance > Pharma Marketing > Clinical Trails > Preclinical
  - Easier to demonstrate business value for new approaches
  - Easier to build decision making tools
  - No need to deal with the needs of those demanding scientists ☺
- Everyone has a data integration problem… still
  - Universal, across all industries – e.g. health insurance
  - All integration problems are semantic in nature
  - Moving from trying to solve all problems (data warehouse) to as-needed integration in as good-as-possible way
- Big data and Data Science are the big opportunity areas
  - Companies have well structured, internal data
  - Wild West of external structured, unstructured data of high value?
  - **OpenPHACTS is further ahead that we realize – working in the hardest market (business value / technical)**

A framework to maximize phenotypic assay impact

- Phenotypic Knowledge Networks
  - Opportunity for truly translational networks – bridging molecular, clinical and even post-market data
  - Public phenotypic data can seed a wealth of knowledge discovery
  - Progressive approach to data mapping
    - As simple as possible at first
    - Maximize use of existing data frameworks and ontologies – OpenPHACTS, BAO, ToxCast, AOP, …

- Phenotypic prediction and profiling tools
  - Maximize use of data from phenotypic knowledge networks
  - Prediction of missing links in knowledge networks
  - Target deconvolution
  - Virtual screening for phenotypic assays

OPDDR – a semantic sandbox for phenotypic assay data

- Open Phenotypic Drug Discovery Resource
- Supported by NIH and Eli Lilly https://tripod.nih.gov/pd2/
- Initial data consists of 2460 approved drugs supplied by NIH NCATS tested in Eli Lilly OIDD's 5 phenotypic panels (35 assays)
- More screens are being carried out in second phase of project
- New phase to develop a **Knowledge Network** around these assays, including **semantic mapping** of the compounds and assays and **missing link prediction** for profiling of assays and compounds against public data

Lilly

NIH National Center for Advancing Translational Sciences

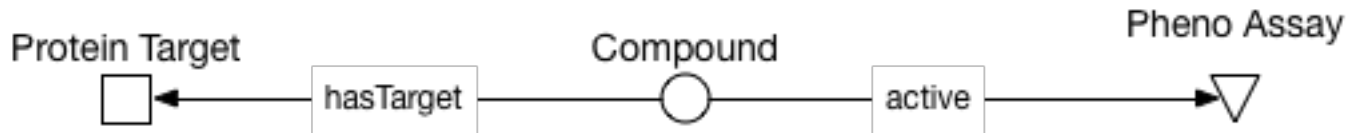
INDIANACTSI Clinical and Translational Sciences Institute

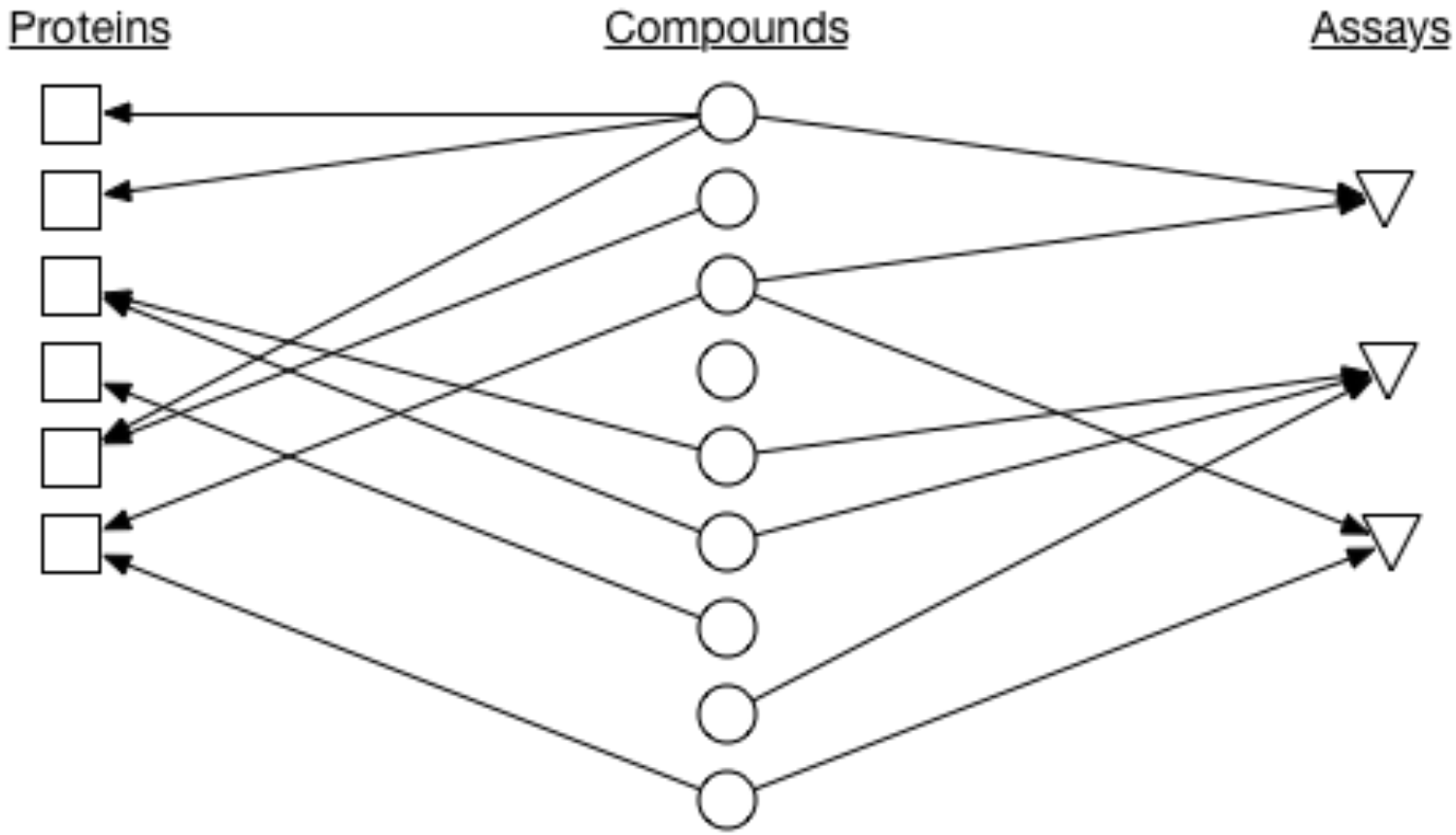
DATA2DISCOVERY INC

Phenotypic Profiling toolset under development

- Data2Discovery is developing an initial phenotypic profiling tool that will make association predictions for phenotypic assays
- Based on previous SLAP semantic link association prediction approached developed in the Indiana University IDSL
- Ability to quantitatively predict association between any two nodes in knowledge network using evidence paths that traverse networks – e.g. assay vs target; target vs disease; compound vs adverse event
- Ability to rank and profile – e.g. rank all human targets against an assay
- Partners include Eli Lilly, NIH NCATS; looking for partnerships to drive development of toolset

Simple target-phenotypic assay prediction

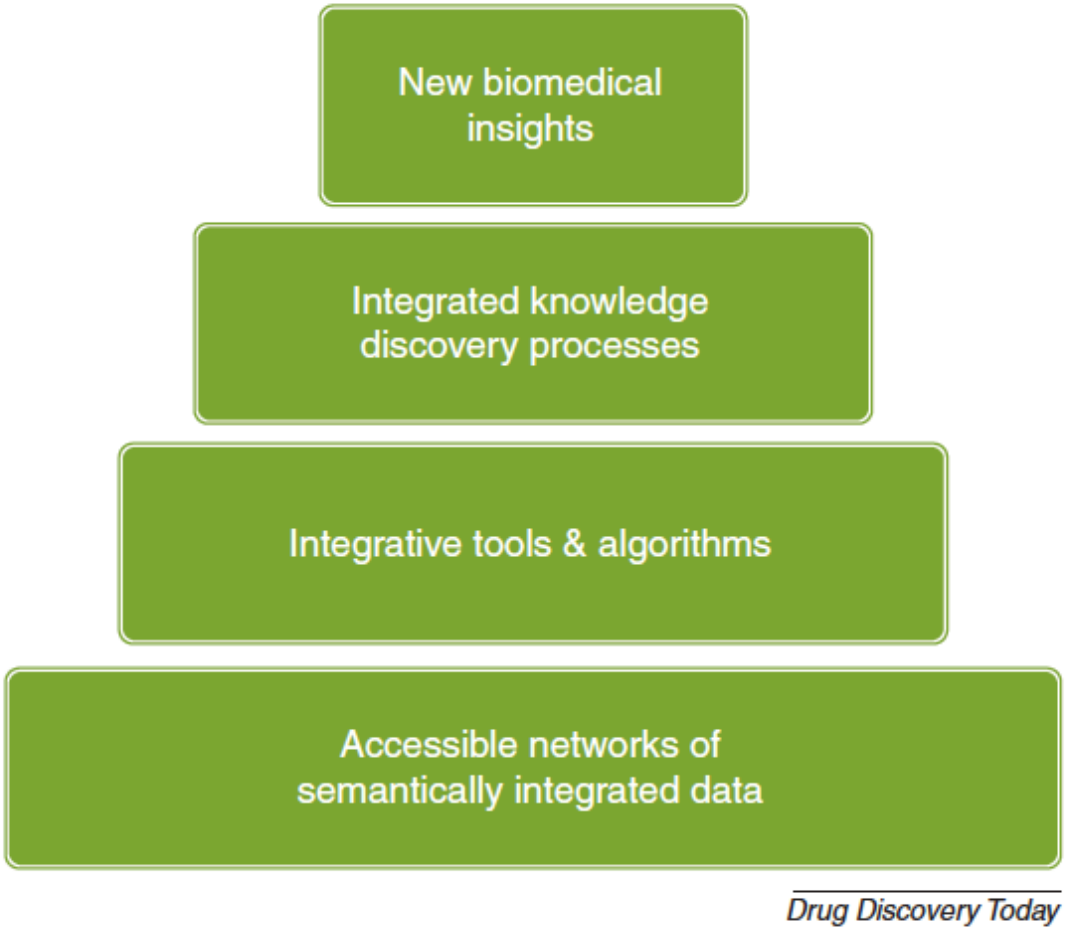




Use Cases go beyond early drug discovery

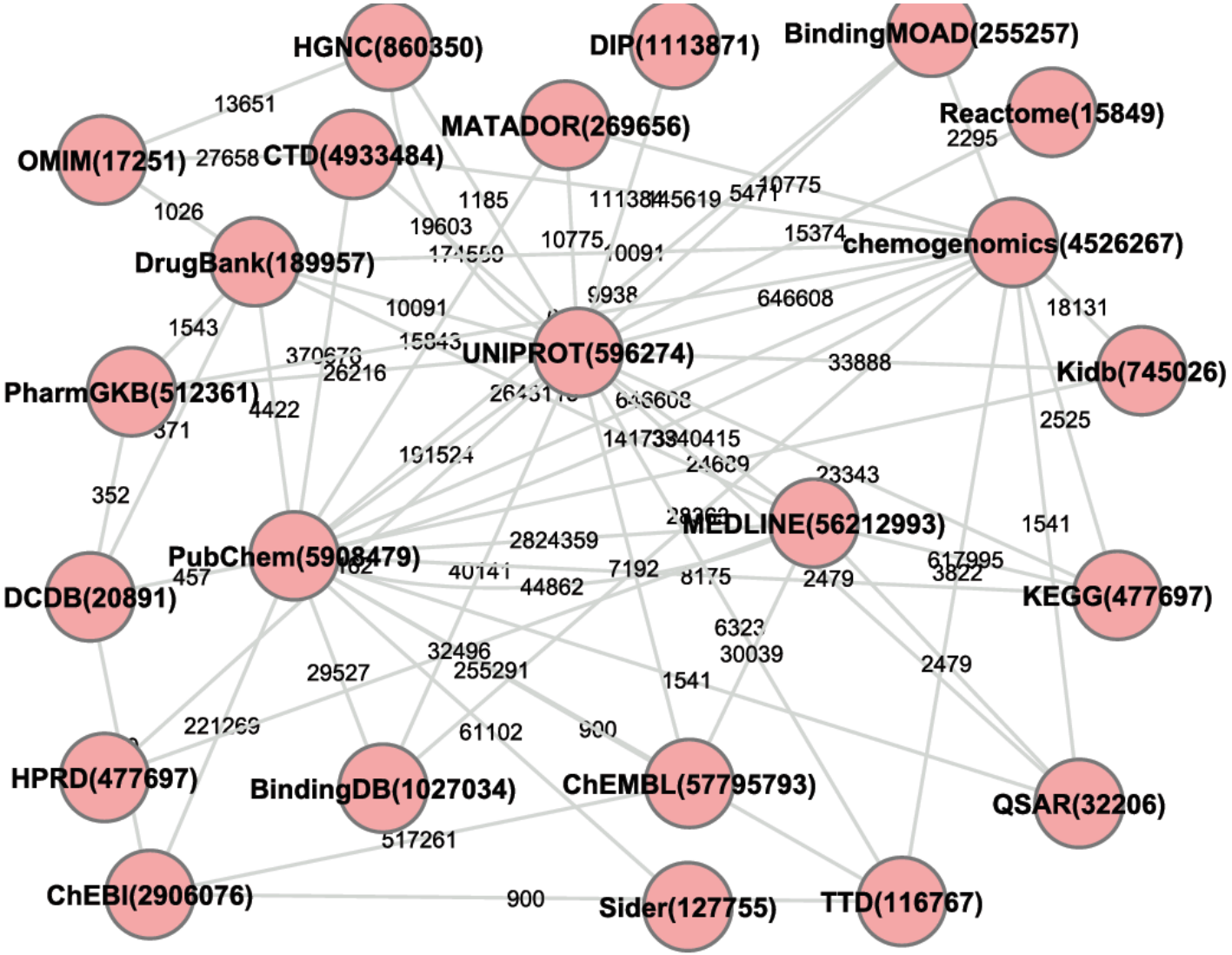
- Preclinical drug discovery (pharma, CRO)
  - Target deconvolution
  - Phenotypic virtual screening
  - Heatmap profile compounds against phenotypic assays / end points
  - Toxicology alerts
- Clinical trails (pharma, CRO)
  - Adverse event profiling using translational data
- Post market
  - Postmarket monitoring using translational data
  - Profile patients at molecular level based on drugs, diseases, phenotypes
- Environmental toxicology
  - Adverse Outcome Pathway prediction

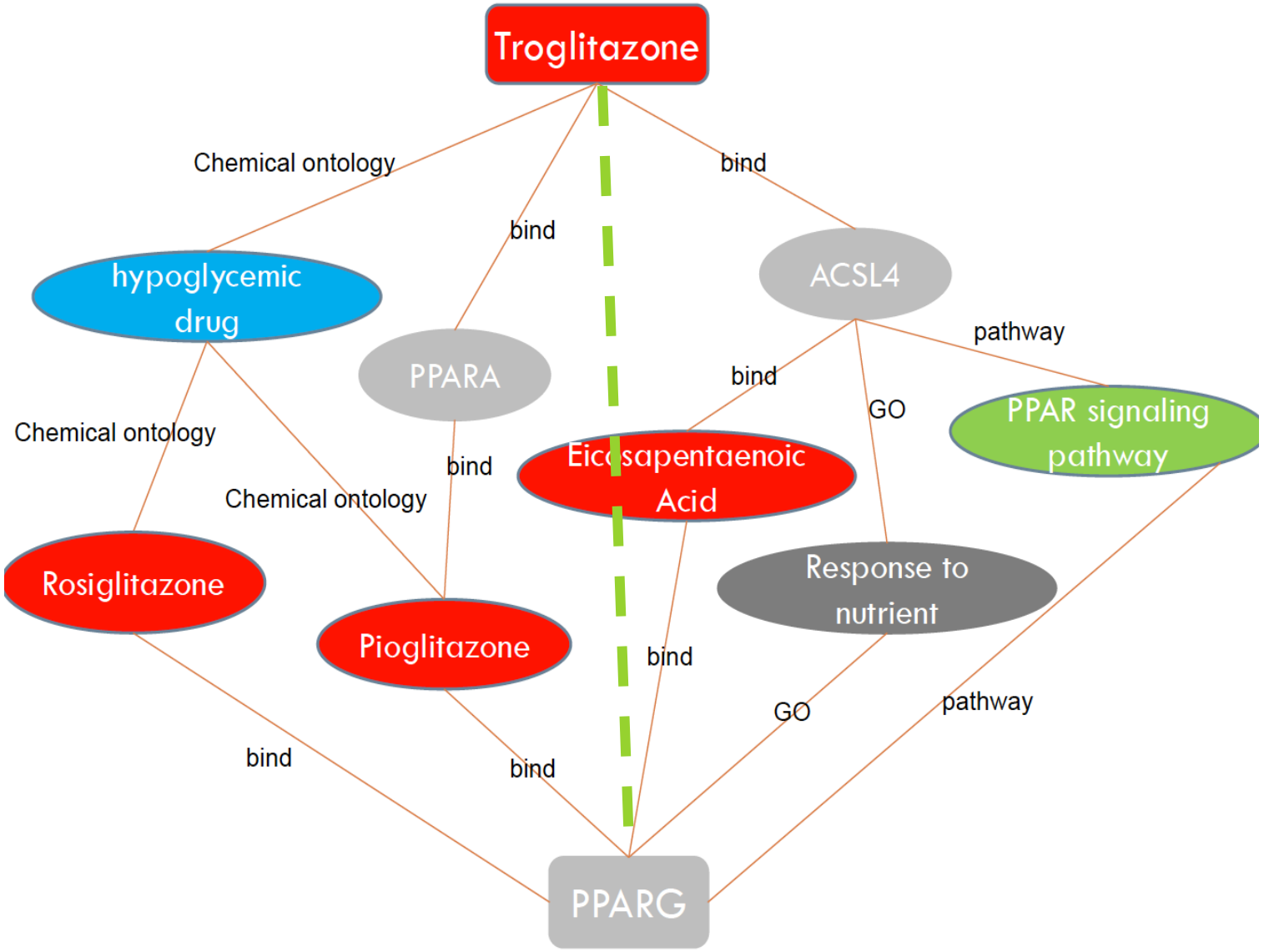
From semantic representation to scientific impact



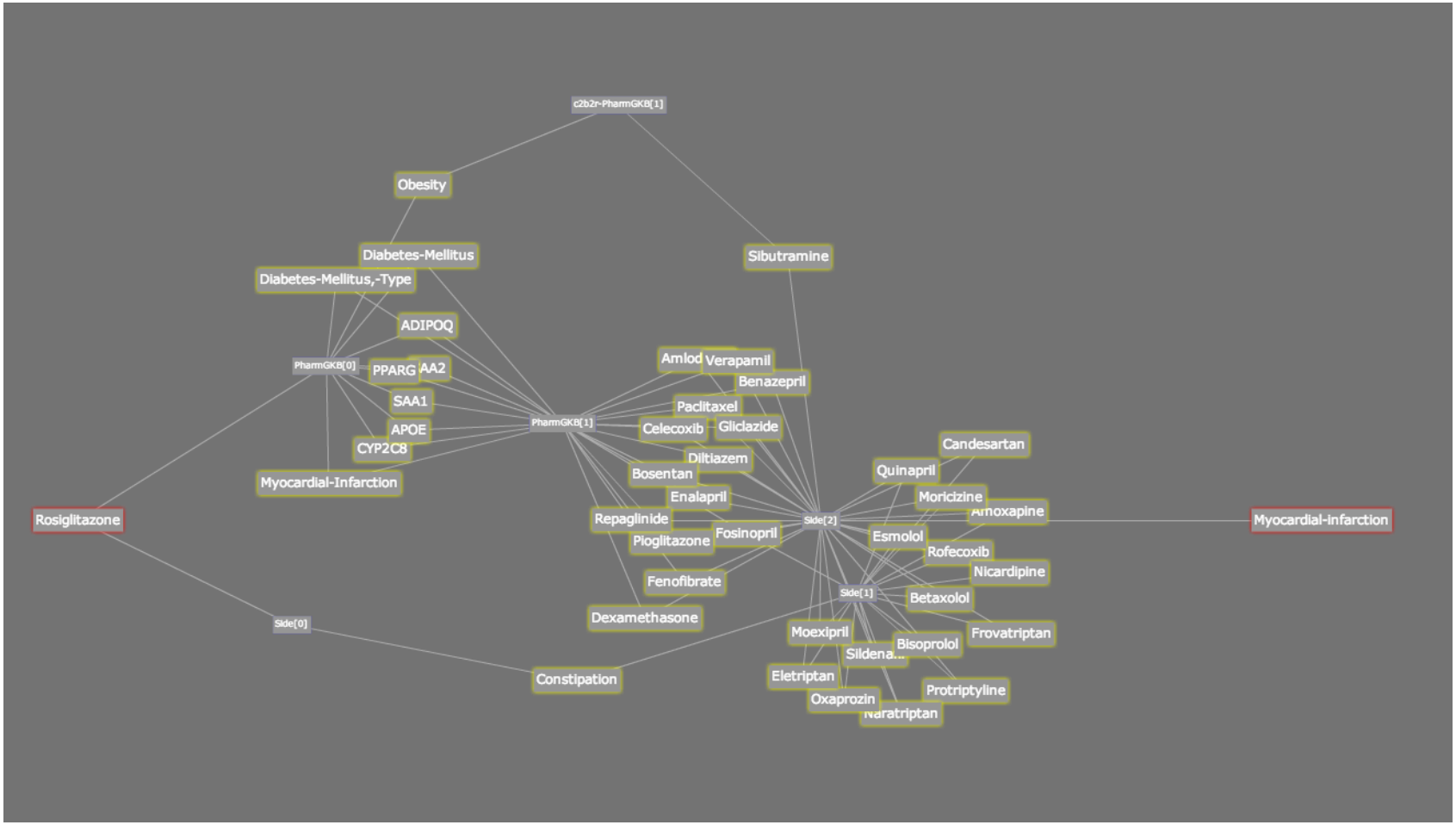
Wild, D.J., Ding, Y., Sheth, A.P., Harland, L., Gifford, E.M., Lajiness, M.S. **Systems Chemical Biology and the Semantic Web: what they mean for the future of drug discovery research**, *Drug Discovery Today*, 2012, 17, 469-474.

Chem2Bio2RDF




Troglitazone
Chemical ontology
bind
bind
hypoglycemic drug
ACSL4
PPARA
pathway
bind
GO
Chemical ontology
bind
PPAR signaling pathway
Eicosapentaenoic Acid
Chemical ontology
Rosiglitazone
Response to nutrient
Pioglitazone
bind
GO
pathway
bind
bind
PPARG

Association graph search – finding evidence paths



He, B., Tang, J., Ding, Y., Wang, H., Sun, Y., Shin, J.H., Chen, B., Moorthy, G., Qiu, J., Desai, P., Wild, D.J., **Mining relational paths in biomedical data** *PloS One*, 2011, e27506.

Knowledge Network Entity Graph