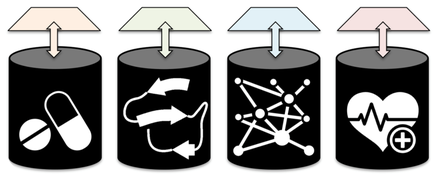
Mining the Life Sciences Linked Open Data Cloud: An Application to Discover Drug - Adverse Reaction Associations

STANFORD SCHOOL OF MEDICINE
Stanford University Medical Center

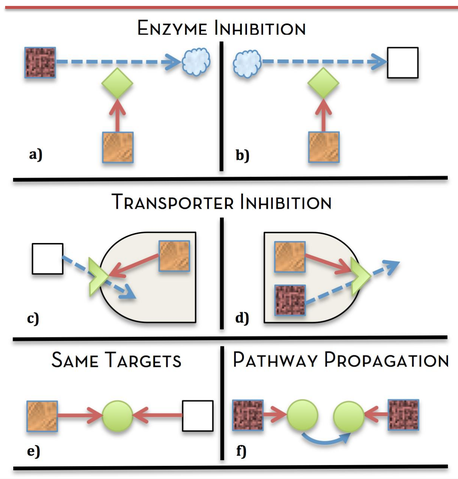
Maulik R. Kamdar (maulikrk@stanford.edu) and Mark A. Musen

Motivation



Distributed Data and Knowledge

- Formats (XML, CSV)
- Notations (Ensembl, HGNC)
- Schemas (molecular_weight, mol-wt)

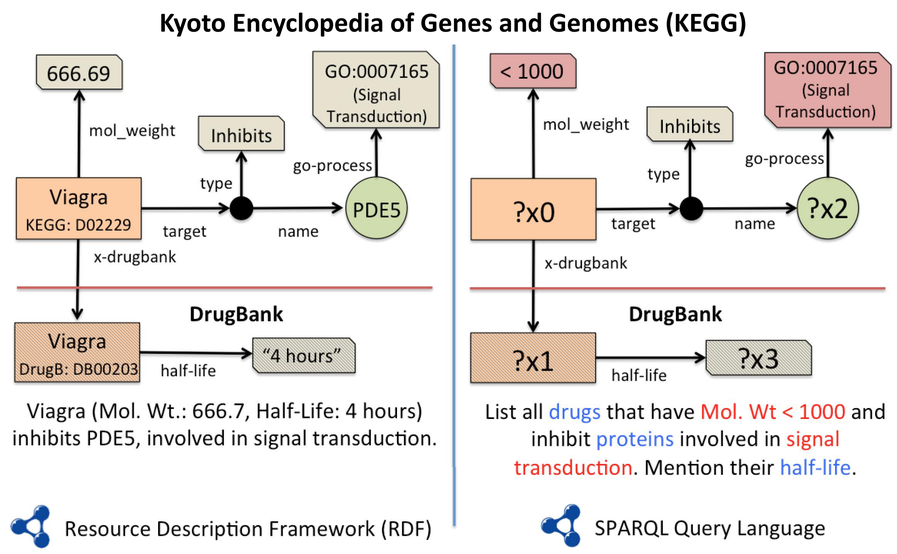


	Protein		Drug - Active
	Enzyme		Drug - Reduced Effect
	Transporter		Drug - Increased Toxicity
	Component		Drug - Inactive

Integrated Systems Pharmacology

- Discovery of novel drug-adverse reaction associations.
- Mechanism-based prediction of drug safety and toxicity.
- Exploration of the underlying biological mechanisms.

Life Sciences Linked Open Data (LSLOD) Cloud

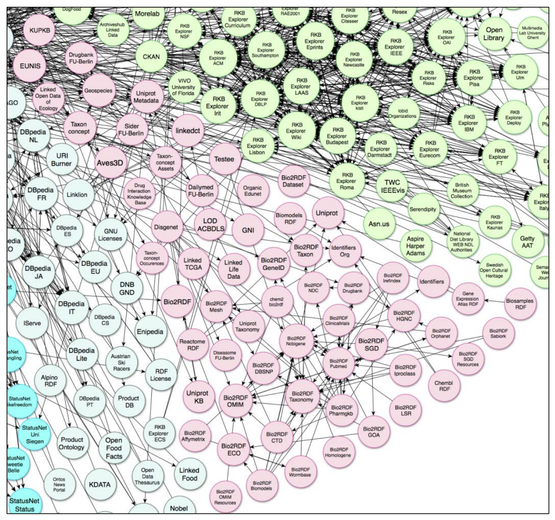


Viagra (Mol. Wt.: 666.7, Half-Life: 4 hours) inhibits PDE5, involved in signal transduction.

List all drugs that have Mol. Wt < 1000 and inhibit proteins involved in signal transduction. Mention their half-life.

Resource Description Framework (RDF)

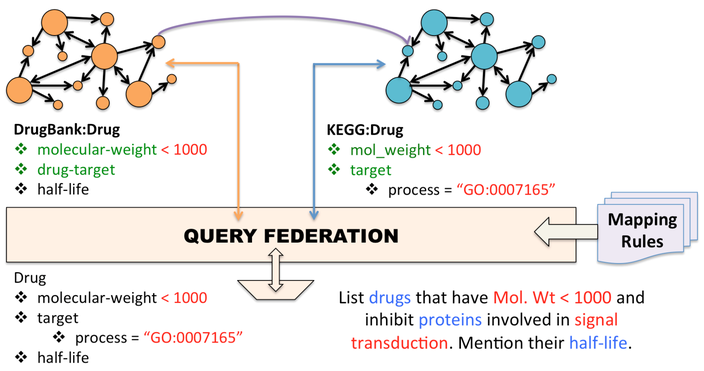
SPARQL Query Language



Reference: http://lod-cloud.net/

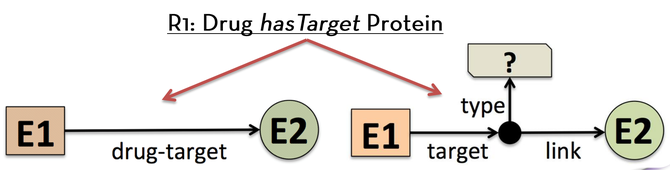
- **Life Sciences Linked Open Data Cloud:** 1T+ triples from 80+ biomedical sources.
- Develop Bio-mashups and Linked Biomedical Dataspaces facilitating *in silico* data discovery

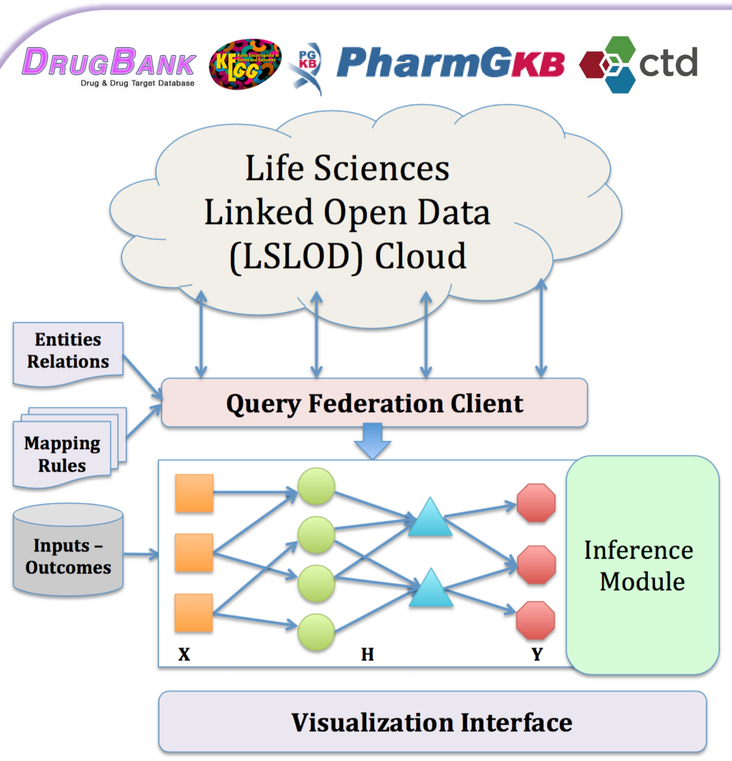
Query Federation



Entity Types		Relation Types	
E1	Drug	R1	Drug *hasTarget* Protein
E2	Protein	R2	Drug *hasEnzyme* Protein
E3	Pathway	R3	Drug *hasTransporter* Protein
E4	Phenotype	R4	Protein *isPresentIn* Pathway
		R5	Pathway *isImplicatedIn* Phenotype

Mapping Rules





Inputs-Outcomes Database

U.S. FDA Adverse Event Reporting System (FAERS) collects reports on the adverse reactions from multiple drug intake.

Inference Module

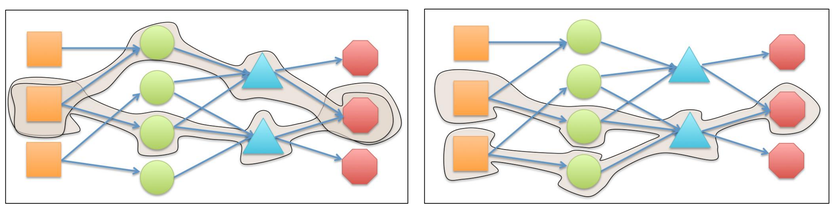
Reporting Ratio (Observed/Expected)

2x2 tables for drug-reaction and drug-drug-reaction associations

	ADR_k = Yes	ADR_k = No
$Drug_i$ = Yes	C_{11}	C_{10}
$Drug_i$ = No	C_{01}	C_{00}

	ADR_k = Yes	ADR_k = No
$Drug_i$ = Yes & $Drug_j$ = Yes	C_{11}	C_{10}
$Drug_i$ = No | $Drug_j$ = No	C_{01}	C_{00}

Enriching Short Subgraphs with Observed and Expected counts



- Compute shortest subgraphs between drugs and phenotypes.
- Enrich the subgraphs with actual and expected counts.
- Compute association scores

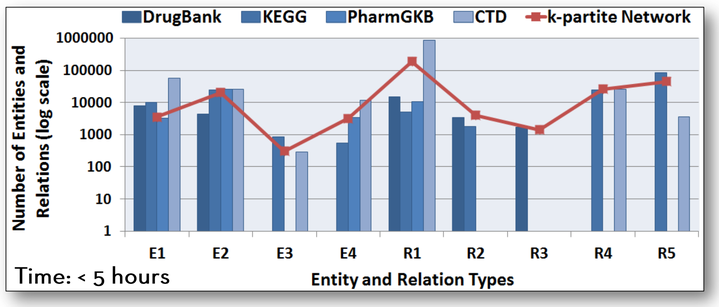
$$w(v_i, v_j) = 1 + \frac{p(v_i \rightarrow v_j \mid G(C_{obs}))}{p(v_i \rightarrow v_j \mid G(C_{\exp}))}$$

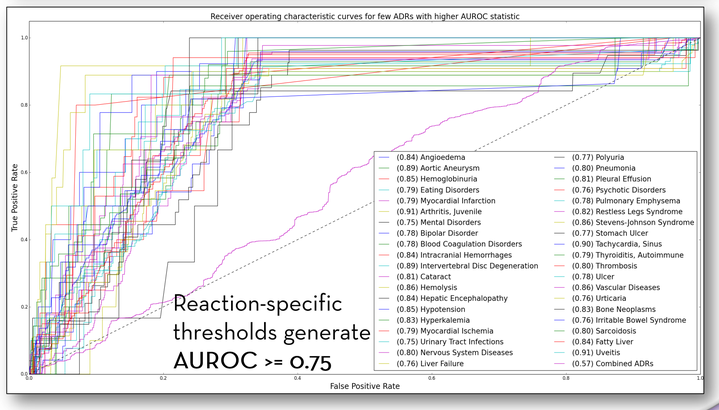
$$p(v_i \rightarrow v_j \mid G(C_*)) = \frac{\#(v_i \rightarrow v_j)}{\#(v_i \rightarrow \bullet)}$$

"Gold" Standard

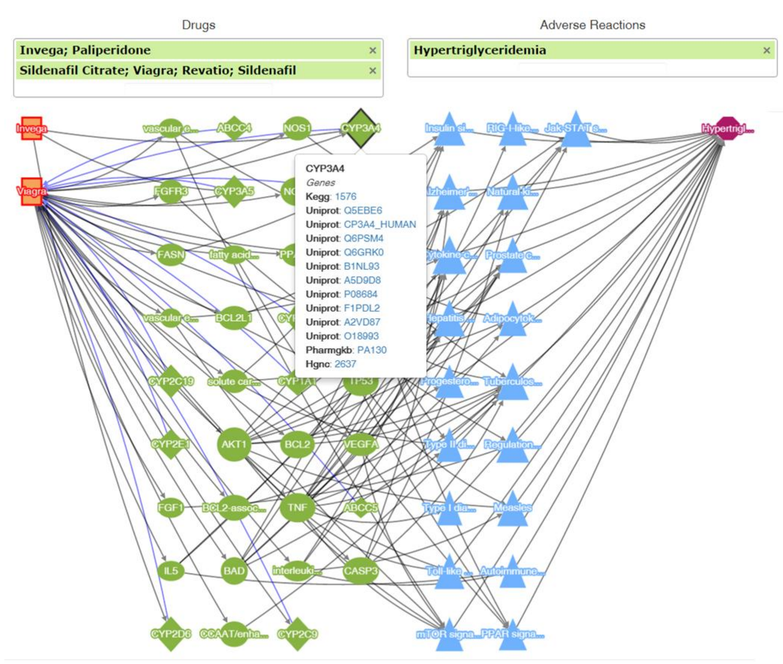
Manually curated from known drug-reaction and drug-drug-reaction associations listed under **MediSpan, Drugs.com** and **EU-ADR** datasets.

Preliminary Results





Drug - Adverse Reaction Visualizer



http://onto-apps.stanford.edu/phlegra

Conclusion

- An integrated systems pharmacology-based approach that queries the LSLOD cloud and generates a *k*-partite graph with drug, protein, pathway and phenotype entities.
- Proof-of-concept to discover drug-reaction and drug-drug-reaction associations, and explore the underlying biological mechanisms.
- Inference method can be repurposed using another Inputs-Outcomes data source.
- Similarly, different *k*-partite networks can easily be generated using the query federation client.
- Problems such as data **quality**, **reporting** biases, **indication** biases, patient **stratification**, interactions due to **multiple drugs** not dealt.
- We will like to validate discovered mechanisms using PubMed literature, as well as evaluate method using MGPS and BCPNN baselines.

Acknowledgments: Michel Dumontier (Bio2RDF), Erik van Mulligen (EU-ADR), Rainer Winnenburg (Drug-ADR)
This work is supported by grant U54-HG004028 from the US National Institutes of Health.