

Abstract

Model organisms such as budding yeast provide a common platform to interrogate and understand cellular and physiological processes. Knowledge about model organisms, whether generated during the course of scientific investigations, or extracted from published articles, are integrated and made available by model organism databases (MODs) such as the Saccharomyces Genome Database (SGD). SGD and many MODs use **InterMine**, a system for integrating, analysing, and republishing biological data from multiple sources that also enables data-driven bioinformatic analyses through a web user interface and programmatic web services.

Here, we developed a cloud-ready dockerized platform that uses **Semantic Web technologies** to transform and make available model organism data in a manner that makes it easier to discover, explore, and query. First, we developed a pipeline to extract, transform, and load a Linked Data representation of the InterMine store. Second, we use **Docker** to package both software and data for local or remote deployment. Third, we built a lightweight dashboard that packages together existing and SPARQL-aware applications to search, browse, explore, and query the InterMine-based data. Our work extends the InterMine platform, and supports new query functionality across InterMine installations and the network of open Linked Data.

Specific Aims

1. To improve interoperability of MOD data by publishing Linked Data
2. To enable and demonstrate federated queries between MOD data and the network of Linked Data
3. To package our software and data for easier local and cloud-based deployment

Availability

Website: <http://mo-lid.org>

Tutorial : <https://github.com/mo-lid/mold-dock>

- Only 5 commands to deploy a Linked-MOD !

Docker (container engine) to build and deploy the MOLD infrastructure

<https://hub.docker.com/u/mold>

Microservices architecture for reusability and extensibility :
Web application, API and Virtuoso images

- Cloud-Ready - tested on Amazon EC2



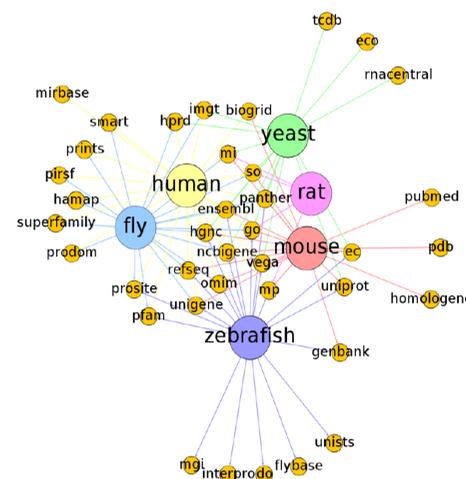
OPEN LINK SOFTWARE

Displaying Type where:
7s1 has any Attribute with Value "hexokinase" Drop.

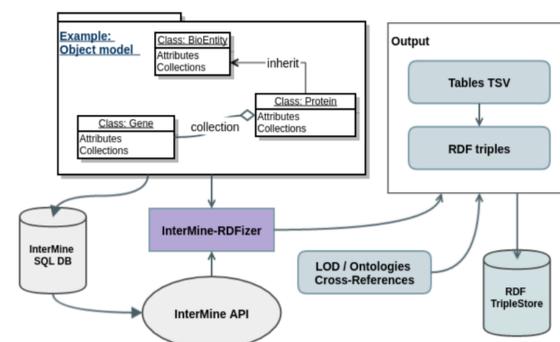
view query as SPARQL Facet permalinks

7s1 instanceOf Type	Count
Title">http://mo-lid.org/mine_type>Title	916
http://mo-lid.org/rel...humanmine_BioEntity	554
http://mo-lid.org/rel...humanmine_SequenceFeature	505
http://mo-lid.org/rel...humanmine_Region	505
http://mo-lid.org/resource/humanmine_Even	484
http://mo-lid.org/mine_type/AbstractText	380
http://mo-lid.org/mine_type/Description	322
http://mo-lid.org/mine_type/Citation	224
http://mo-lid.org/rel...humanmine_BioEntity	125
http://mo-lid.org/mine_type/Summary	121
http://mo-lid.org/rel...Ontology/TermSynonym	110

Powerful Search. Users can use keywords to initiate a faceted search across the MODs and filter to find associations of interest.



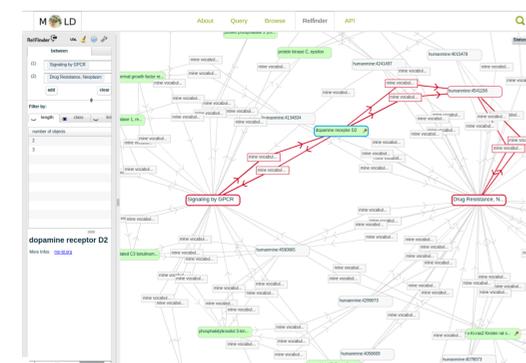
Network Visualization. Automatically generated links drawn from the ontology and cross-references tables. Missing links indicates *incorrect* use of the InterMine database.



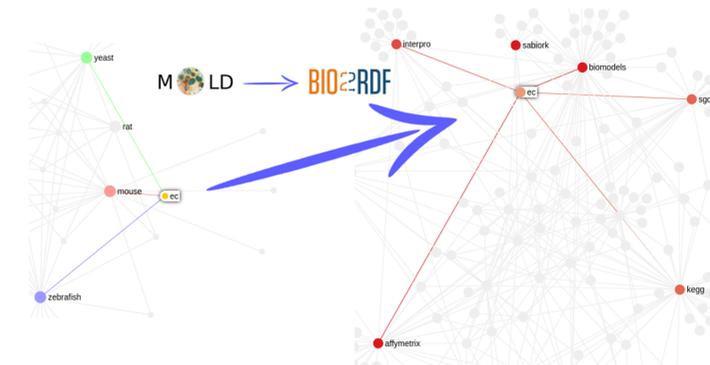
The InterMine RDFizer queries the InterMine API with XML description of the object model. It then upconverts the tabular results into triples (RDF). It merges resources with the same primary keys and Links data with external datasets. Finally, the data is loaded into a RDF-compatible database (triple store).

References

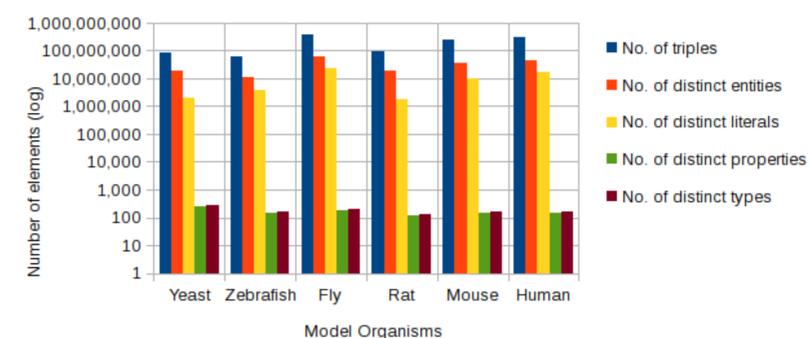
M Deraspe, K Karra, G Binkley, J Sullivan, G Micklem, J Corbeil, M Cherry, M Dumontier. Semantic research platform for Model Organisms. Proceedings of the 1st International Workshop on Biomedical Data Integration and Discovery (BMDID 2016). Kobe, Japan.



Hypothesis Exploration. The RelFinder tool can be used by researchers to find connections between entities of interest.



Knowledge Exploration. Users can more easily follow links or query from MOD data to the wider network of biomedical data contained within the BIO2RDF project.



MO-LD project statistics, by MOD. Conversion of the 6 MODs from their InterMine stores yielded a total of 1.16B triples, 192m distinct entities, 56m literals, 1081 types, and 977 properties.

Acknowledgements

Funding for this work was provided as an administrative supplement to the Genomic Database for the Yeast Saccharomyces (U41HG000131521).