

ELIXIR: Applying FAIR to InterMine



InterMine is a platform to **integrate** and **access** life sciences **data**, providing a user-friendly web interface and RESTful web services. InterMine comes with a core data model for common biological entities and loaders for popular data sources and file types; different deployments can extend these components to publish any type of data.

We present here the **practical results** of our experience of applying **FAIR data principles** in InterMine.

Generating persistent URLs for web pages

InterMine already has unique URLs to identify the report pages but they are not persistent. To achieve URL persistency we have generated new URLs.

<http://humanmine.org/humanmine/gene:5468>



Generating persistent URIs for data

In order to generate RDF representing the data integrated in InterMine instances, we need to generate persistent URIs for the data.

We recommend any InterMine instance which create new entities to adopt Identifiers.org as a Permanent URL (PURL) provider, by registering itself as a data collection.

<http://identifiers.org/legumemine/geneticmarker:118M3>

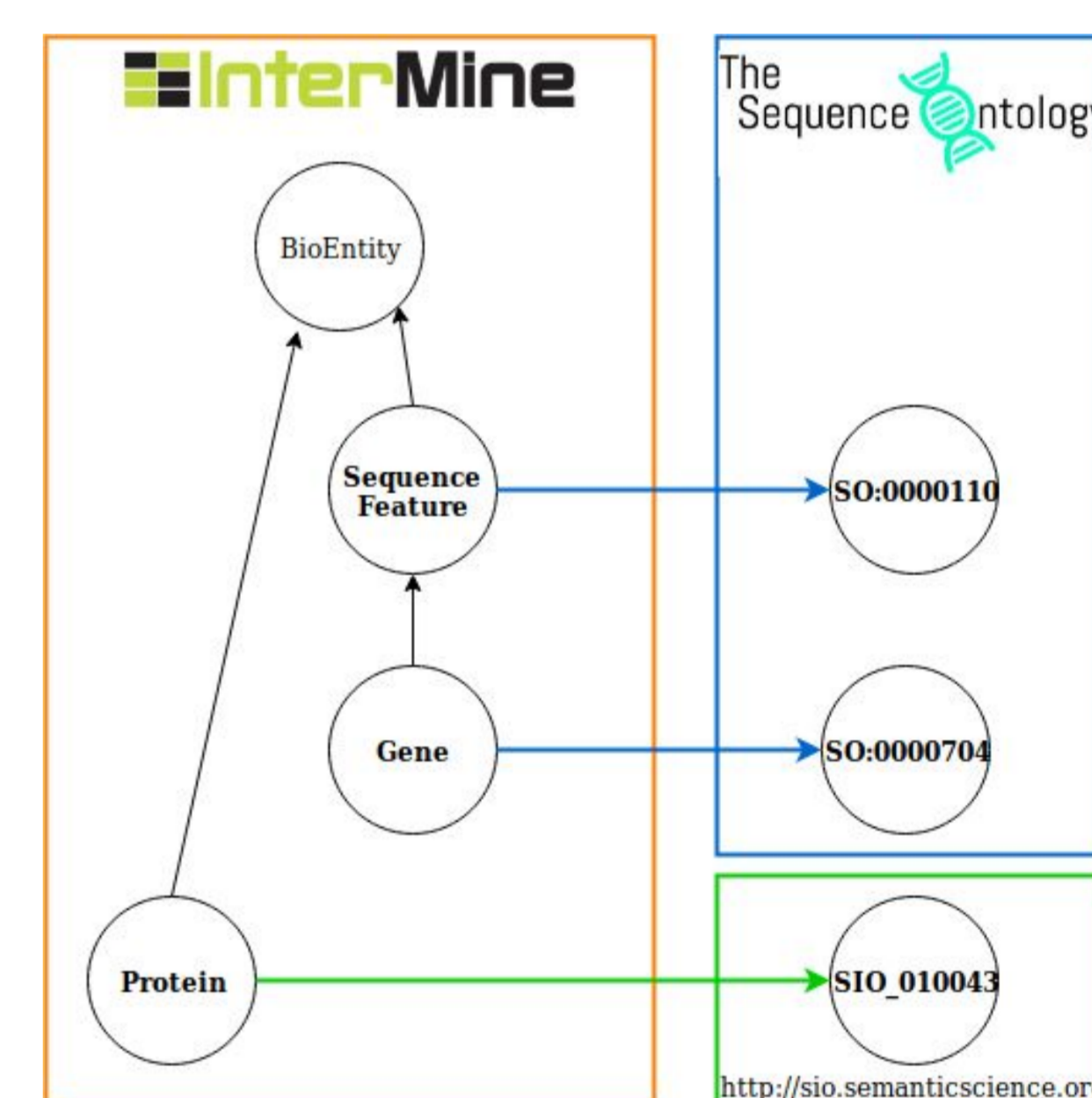


<https://mines.legumeinfo.org/legumemine/geneticmarker:118M3>

Describing data with ontologies

The InterMine system is based on a core data model, described in an XML file, which defines classes (the entities in the model) and the relationships between them.

InterMine already automatically **applied terms from the Sequence Ontology** to its data model. To improve data interoperability we have now **added more ontologies** to its core data model and provided InterMine instance administrators with the ability to apply any ontologies to their data model extension.



Marking up web pages

We have applied structured data in JSON-LD format to InterMine web pages, using Bioschemas.org types and profiles.

DataCatalog	InterMine instance home page
DataSet	Report Page for entities with type DataSet

Publishing data licences

InterMine has update its model to include the licences that govern the data sets that have been integrated.

Data Set Name	Data Set Licence	Data Set Version
BioGRID interaction data set	https://opensource.org/licenses/MIT	3.4
IntAct molecular interactions	https://www.apache.org/licenses/LICENSE-2.0	4.2.12
Orthologue and paralogue predictions	http://www.gnu.org/licenses/gpl.txt	13.1
Reactome pathways data set	https://creativecommons.org/publicdomain/zero/1.0/	0
Swiss-Prot data set	http://creativecommons.org/licenses/by/4.0/	2018_05
TrEMBL data set	http://creativecommons.org/licenses/by/4.0/	2018_05
UniProt keywords data set	http://creativecommons.org/licenses/by/4.0/	0

Contact

www.intermine.org
info@intermine.org
github.com/intermine
twitter.com/intermineorg

