

Title: Extending InterMine to yeast, rat and zebrafish model organism databases**Acronyms**

GO	Gene Ontology	http://www.geneontology.org
GMOD	Generic Model Organism Database Consortium	http://www.gmod.org
Mine	One of the MOD-specific InterMine data warehouses developed in this proposal	
MOD	Model Organism Database	
modENCODE	The model organism ENCyclopedia Of DNA Elements project: project to identify all functional elements in the <i>Drosophila</i> and <i>C. elegans</i> genomes	http://www.modencode.org
QTL	Quantitative Trait Locus	
RGD	Rat Genome Database	http://rgd.mcw.edu/wg/
SGD	<i>Saccharomyces</i> Genome Database	http://www.yeastgenome.org/
SO	Sequence Ontology	http://www.sequenceontology.org
ZFIN	The Zebrafish Information Network	http://www.zfin.org

Specific Aims:**1. Apply InterMine infrastructure to generate Mines alongside each of the target Model Organism Databases (MODs).**

Custom parsers will be used to import data from each MOD's database into a 'Mine' made using the generic InterMine data warehouse software. This will make the data collection of each MOD accessible for flexible queries and bulk downloads. Powerful operations with lists, graphical and statistical summaries ('widgets'), and access to data via web services will be enabled. A common query interface will be provided to the different MODs and this will save users having to learn a new query interface for each MOD. The work in this specific aim will underpin the developments in specific aim 2 that will enable greater interoperability between the MODs.

2. Enable the different Mines to be interoperable and, thus, provide greater cross-organism integration

Through the use of InterMine, the Model Organism Databases (MODs) will operate a common software platform. Interoperability will be provided in two ways: lists of e.g. genes will be easily transferable between the MODs by users. This will make it straightforward to access data from other organisms and assist in refinement of the original list. Separately, a mechanism will be developed to allow high-value data from each Mine to be embedded as query results or widgets within other Mines, by developing an extension of InterMine web services. Such inter-operability will make it much easier to carry out cross-organism comparisons.

3. Configure the Mines, implement custom widgets and integrate tools

InterMine systems are highly configurable: each Mine will be tailored to include new pre-defined 'template' queries, widgets and public lists appropriate to the specific data available. Each Mine will be integrated with existing tools available at the MOD site and will be configured to reflect the appearance of the host MOD. Developments under specific aim 2 will enable the configuration of data exchange between the MODs and embedding of functionality provided by one MOD within the user interface of another.

4. Merge other external data sources into the MOD-specific Mines

InterMine is able to load data from several standard biological formats. We will use existing parsers, and develop new ones as necessary, in order to integrate additional types of data not yet included in the MOD. These new data will be available to query in the Mine alongside the core MOD data collection.

5. Develop code to a high quality, and make it freely available under an Open Source license in a well-documented form through the Generic Model Organism Database (GMOD) Consortium.

The best use of limited public funds are made when all aspects of software development projects are made freely available, and are constructed in such a way as to promote easy code re-use in the future. We are committed to such an Open Source development process and will maintain the high standards of coding and documentation that have been established for the InterMine and FlyMine projects.