



**Cloud deployment of FAIR resources:  
Success stories from the EOSC-LIFE RI  
community**



## **PombeMine: A FAIR workflow-enabled PomBase in the cloud**

Intermine, Dept. of Genetics, University of Cambridge

Daniela Butano

Rachel Lyne

Gos Micklem

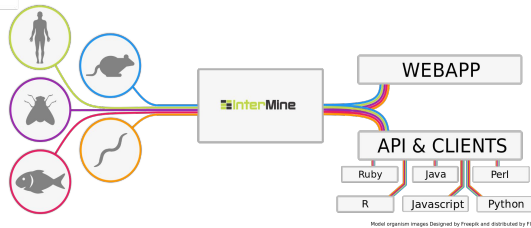
PomBase Dept. of Biochemistry, University of Cambridge

Kim Rutherford

Valerie Wood



Platform to integrate and access life sciences data



More than 30 instances around the world

AltAnnoMine	altannomine	AltAnnoMine integrates many types of data for all the Model Organisms Fly, Worm, Zebrafish, Yeast, etc...
BMAP	brmapmine	BrascanMap Alignment Project
BioMine	biomine	Common bean (Phaseolus vulgaris) and lima bean (Phaseolus linaoid) data from the Legume Informatics System
BioMine	bioinform	An integrated data warehouse for the Bacteria Genome Database
ChIMine	chimbic	An integrated database for Chlamydia genomes and OMG calls
ChickpMine	chickpmine	Chickpeas data from the Legume Information System
CovidMine	covidmine	InterMine for integrated COVID-19 genomic and geographic distribution data
CropMine	cropmine	Croppeas data from the Legume Information System
FruitMine	fruitmine	An integrated Omics data warehouse for FRL, Spodoptera frugiperda
FlyMine	flymine	An integrated database for Drosophila genomics
GrapeMine	grapevine	An integrated database for grapevine data
HumanMine	humanmine	An integrated database of Human genome genomic data
HymenopteraMine	hymenoptera	An integrated data warehouse for the Hymenoptera Genome Database
IndigoMine	indgomine	INDIGO enables the integration of annotations for the exploration and analysis of newly sequenced micro...
JostiteichMine	jostiteichmine	Asynchronome event genomic data from the Legume Information System
LegumeMine	legumemine	Genomic and metabolite data for various legumes from the Legume Information System
LocustMine	locustmine	An integrated Omics data warehouse for Locust, Locusta migratoria
LegumeMine	legumemine	Wheat and narrow-leaf lupin data from the Legume Information System
MediMine	medmine	Metablogi functional and Metablogi pathway data from the Legume Information System
ModMine	modmine	A data warehouse for the modENCOD project
MouseMine	mousemine	MouseMine is a powerful new system for online access to mouse data from MGI (Mouse Genome Informatics)
OrbitMine	orbitmine	Quercus colour annotation database

Loaders for common formats & custom parser

Sophisticated query builder

Templates (predefined queries)

Client libraries to generate code from queries

Public InterMine databases <http://registry.intermine.org/>



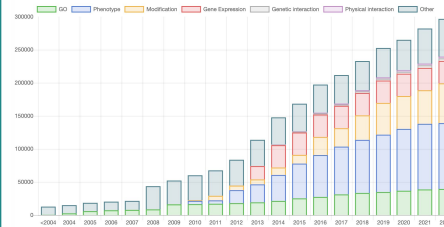


# PomBase (Model Organism Database)

MOD: integrate molecular data for fission yeast (*S. pombe*)

The screenshot shows the PomBase website interface. At the top, it says 'PomBase The scientific resource for fission yeast'. Below that, there's a search bar and navigation links for 'Search', 'Submit', 'Genome Status', and 'Downloads'. A 'Welcome to PomBase' section follows, describing the database as a comprehensive resource for the fission yeast *Schizosaccharomyces pombe*. It lists features like 'Advanced search', 'Genome browser', 'Overview: Function • Process • Component', 'Disease association', and 'Phenotype'. There are also links for 'Example pages: Gene • Term • Publication'. At the bottom, there are three featured articles: 'Research spotlight: Lara-Ramirez et al., 2022', 'Recent community curation', and 'Explore PomBase: Disease associations for fission yeast proteins'.

> 300,000 ontologised statements



Active community, advanced community curation

The screenshot shows the Canto v1682 Community curation interface. It features a search bar at the top with the Canto logo and the text 'Canto v1682 Community curation'. Below the search bar, there's a large heading 'Curate a paper for PomBase' and a sub-heading 'Start curating using a PubMed ID:'. There is a text input field for entering a PubMed ID and a 'Find' button.

.....the only major MOD without a Mine





## InterMine Build, Deployment and Configuration

- Model extended for new data types
- Data mapped to existing InterMine core model
- Set up cloud contract and instance with CSC at EBI
- Configuration of interface, tools and templates
- Troubleshooting iteratively to populate the model fully
  - Includes InterMine bug fixes, improved PomBase allele curation, improved data modelling





## PomBase Data Loaded

- Sequence and features
- Gene Ontology, annotation extensions
- Phenotypes, alleles, conditions, penetrance
- Protein families and domains
- Orthologs
- Disease associations
- Genetic and physical Interactions





## Impact of PombeMine for RIs and EOSC

### PomBase (Elixir Node Service) Users

- Web services and client libraries to enable fission yeast data to be used in analytic workflows
- PomBase Gene lists can be sent to InterMine
- New ways to query data (alleles and annotation extensions)
- Access to enrichment tools
- Potential elevated interoperability with other ELIXIR resources (e.g FlyMine)

### InterMine (Elixir Interoperability Resource)

- To test cloud Software-as-a-Service (SaaS) model for other potential cloud InterMine instances around the world.
- Scalable solution to manage increasing volumes of data.





## Future work and sustainability

- Finalise GO annotation extension query configuration
- Add modification data
- Migrate to a permanent URL
- Automate Update pipeline
- Publish manuscript (in progress)





## Experience of working in EOSC-LIFE and technical teams

- WP2 (Tool interoperability, workflows, standardisation and containerisation):  
Jean-Karim Heriche, Antonio Rosato Initial conversations and advice on docker
- WP7(Cloud deployment support and resource):  
Justin Clark-Casey advised on the Allocation Resource Process  
Alvaros Gonzales from CSC (short training session, prompt support)
- Workflow: Jean-Marie Burel (as a PoP created a workflow/notebook connecting humanmine and IDR, <https://workflowhub.eu/workflows/242>)





## PombeMine in use

CURRENT URL <http://pombemine.rahtiapp.fi/pombemine>

PERMANENT URL (next week)

<https://www.pombase.org/pombemine>

RECORDED DEMO

[https://www.youtube.com/watch?v=1t\\_AUjr6xQ8&ab\\_channel=ValerieWood](https://www.youtube.com/watch?v=1t_AUjr6xQ8&ab_channel=ValerieWood)





 <b>BBMRI-ERIC</b> <sup>®</sup>	 <b>eatris</b>	 <b>ECRIN</b> EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK
 <b>elixir</b>	 <b>EMBRC</b> EUROPEAN MARINE BIOLOGICAL RESOURCE CENTRE	 <b>EMPHASIS</b>
 <b>erinha</b> European Research Infrastructure on Highly Pathogenic Agents	 <b>euopenscreen</b>	 <b>EURO BIOIMAGING</b>
 <b>INFRAFONTIER</b> mouse disease models	 <b>instruct ERIC</b>	 <b>ISBE</b> Infrastructure for Systems Biology Europe
	 <b>MIRRI</b> MICROBIAL RESOURCE RESEARCH INFRASTRUCTURE	

## Acknowledgements





## Spare slides

