InterMine

2016 Updates

InterMine

Overview

- Started in 2003
- Open source biological data warehouse
- Components
  - Sophisticated build system
    - Libraries for common file formats, eg. GFF3, FASTA
    - Add your own data
  - Custom User interface
  - Public API, and clients
    - Access data easily
    - Java, JavaScript, Ruby, Perl and Python

UNIVERSITY OF CAMBRIDGE

InterMine

Lots of new InterMines

- PhytoMine - 50+ plant genomes and counting
- BovineMine - *Bos taurus*
- HymenopteraMine - Bees, Ants & Wasps
- SoyMine - Soybase soy bean data
- BeanMine, LegumeMine and PeanutMine - data from Legume Federation
- PlanMine - Planarian flatworms
- Wheat3BMine - Wheat chromosome 3B
- GrapeMine - Grapevine

InterMine

InterMine as a data sharing platform

- Named resource for the ELIXIR UK node
- Part of the NIH's BD2K Initiative
- Committed to FAIR data
  - Findable
  - Accessible
  - Interoperable
  - Reusable

UNIVERSITY OF CAMBRIDGE

InterMine

Semantic Web

- InterMine SPARQL end point at http://mo-ld.org
- Can generate RDF from any InterMine
- Pilot project by Michel Dumontier at Stanford
- Going to work on extending end point to all InterMines

InterMine

Database Benchmarking

- Postgres - was chosen +10 years ago. Still the best?
  - Custom ORM layer
  - Superfast queries, data loading is pain point
  - Largest InterMine is > 2.6 TB
- Working on benchmarking different database backends
  - Neo4j
  - Blazegraph
- Preliminary results:
  - Neo4j is performant for queries
  - Going to test Neo4j scaling next
- See our wordpress blog for overly detailed description of what we've done.

UNIVERSITY OF
CAMBRIDGE

InterMine

More projects

- InterMineR - R package to query InterMine
- Docker
  - Can deploy InterMine using Docker.
  - See docker-intermine project in GitHub
- Android app
  - Query any InterMine
  - Available at Google Play Store
- UI redesign
  - Early stages
  - ClojureScript

Contact Us!

- InterMine.org
- InterMine Google group
- Twitter @intermineorg
- Wordpress blog

InterMine