

# InterMine

Open source data warehouse and web interface

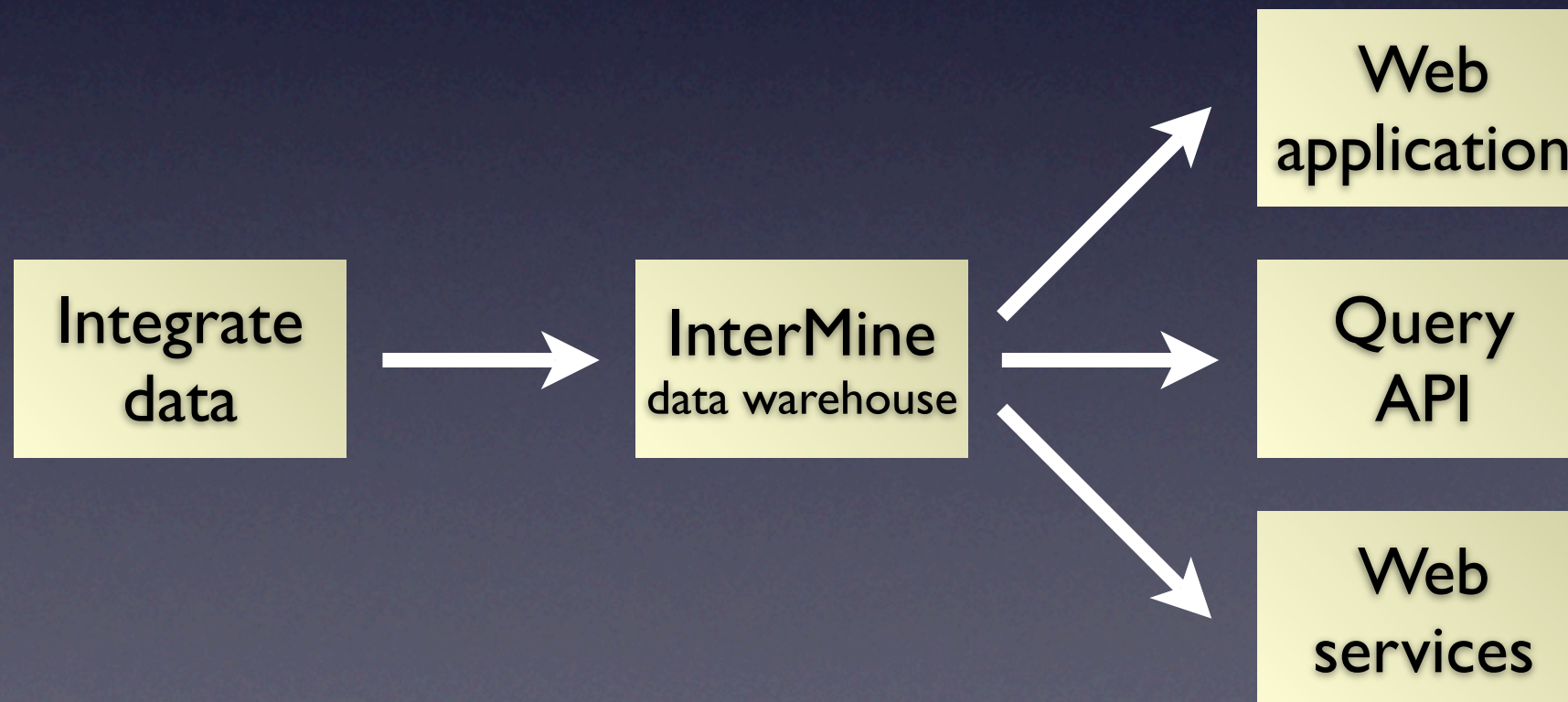
Richard Smith  
University of Cambridge, UK  
[richard@flymine.org](mailto:richard@flymine.org)

**[www.intermine.org](http://www.intermine.org)**



# Overview

- Query-optimised data warehouse system
- Java, object-based data model
- Free, open source (LGPL)
- Flexible querying

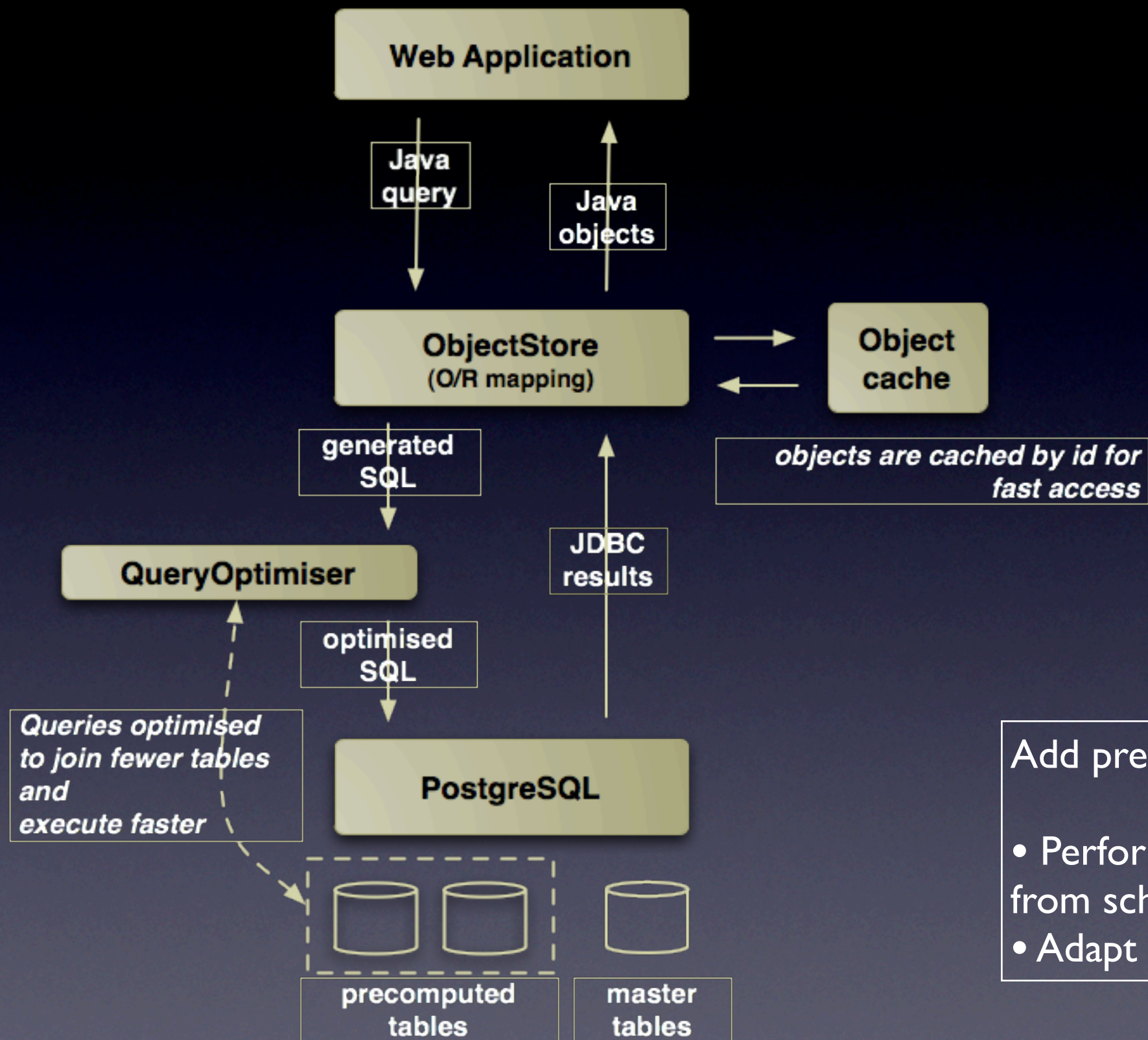




# Projects

- Five developers, since 2002
- FlyMine - [www.flymine.org](http://www.flymine.org)
  - 30+ data sources, *Drosophila* & *Anopheles*
- modENCODE - [www.modencode.org](http://www.modencode.org)
  - *C. elegans*/*D. melanogaster* high throughput, \$57M
- BOKU & IMP - Vienna
- MitoMiner - mitochondria
- MilkMine - milk proteins
- DeinoMine, Japan
- Genes to Cognition, Sanger Centre
- *Yeast, Rat, Zebrafish* model organism databases





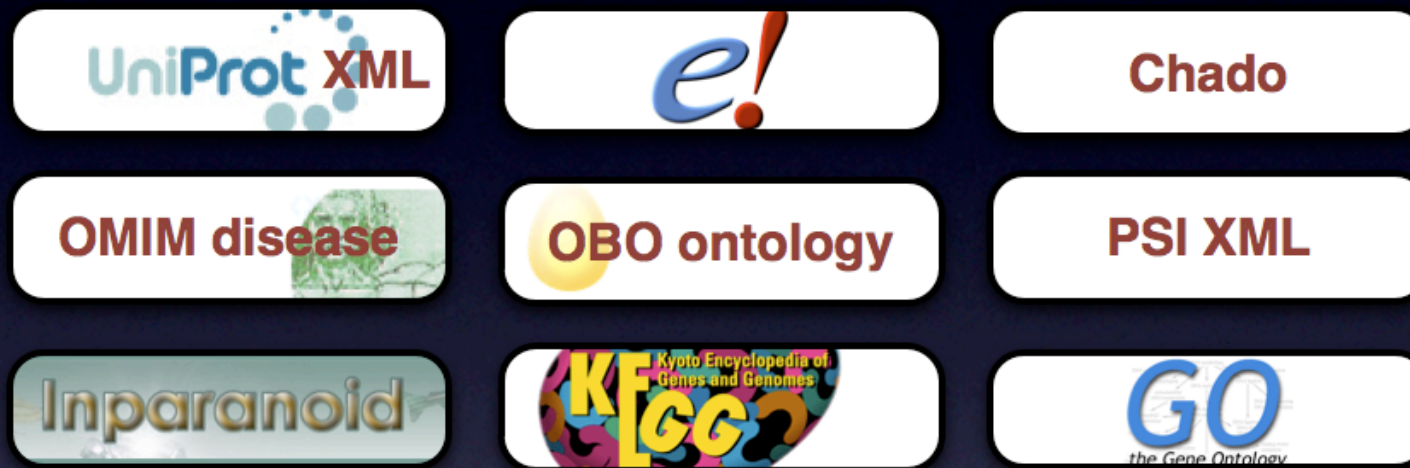
Add precomputed tables at any time:

- Performance optimisation separate from schema design
- Adapt performance to actual use



# Data Integration

*Existing data sources*



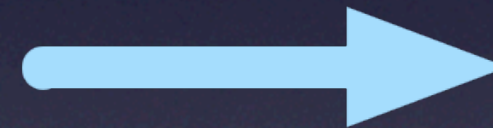
+

**Custom Data Sources**



*Java and Perl APIs*

*Configurable data integration*

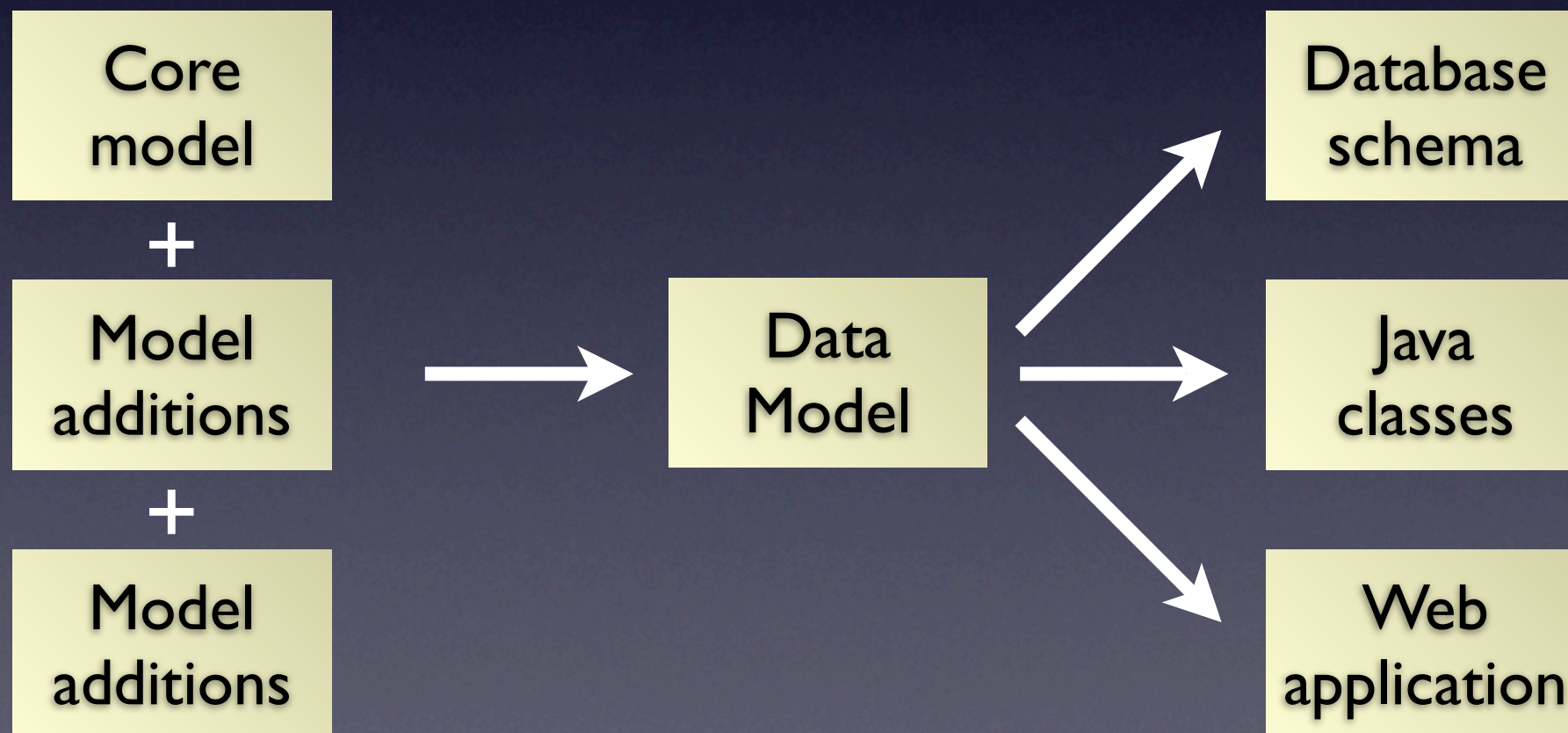


**InterMine  
data warehouse**



# Auto-generation

- Object model defined by XML file
- Low overhead to extending data model





# Custom data

- Any source can add to the data model:

```
<class name="Pathway" is-interface="true" >
  <attribute name="name" type="String"/>
  <collection name="genes" referenced-type="Gene"
    reverse-reference="pathways"/>
</class>
<class name="Gene" is-interface="true">
  <collection name="pathways" referenced-type="Pathway"
    reverse-reference="genes"/>
</class>
```



# Configure a new Mine

```
<project type="bio">
...
<sources>
  <source name="uniprot" type="uniprot" dump="true">
    <property name="uniprot.organisms"
      value="7227 6239" />
    <property name="src.data.dir"
      location="/data/uniprot" />
  </source>
  <source name="my-source" type="pathways">
    <property name="src.data.dir"
      location="/data/pathways" />
  </source>
  ...
</sources>
</project>
```



# Example Usage

## 1. Subject specific database:

- 'Slices' of data from repositories
- Data sets specific to focus
- e.g. Milk proteins, an organism, a disease

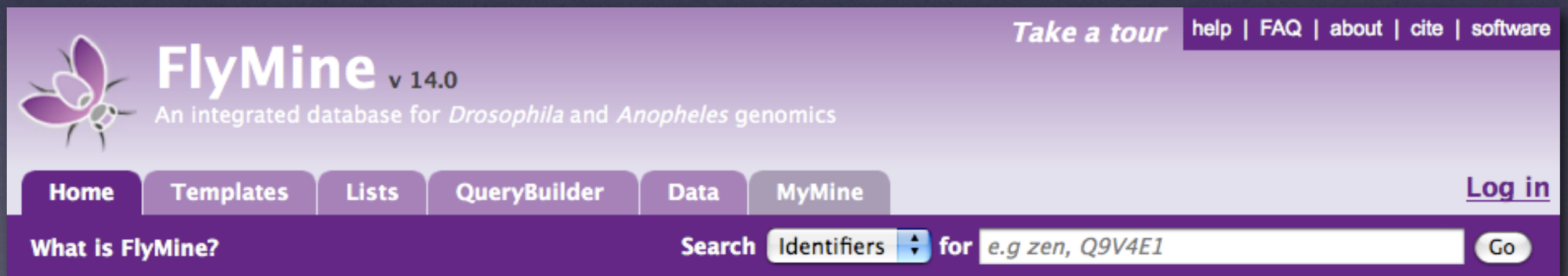
## 2. Present own data:

- Web interface for own data
- Add other sources to provide context



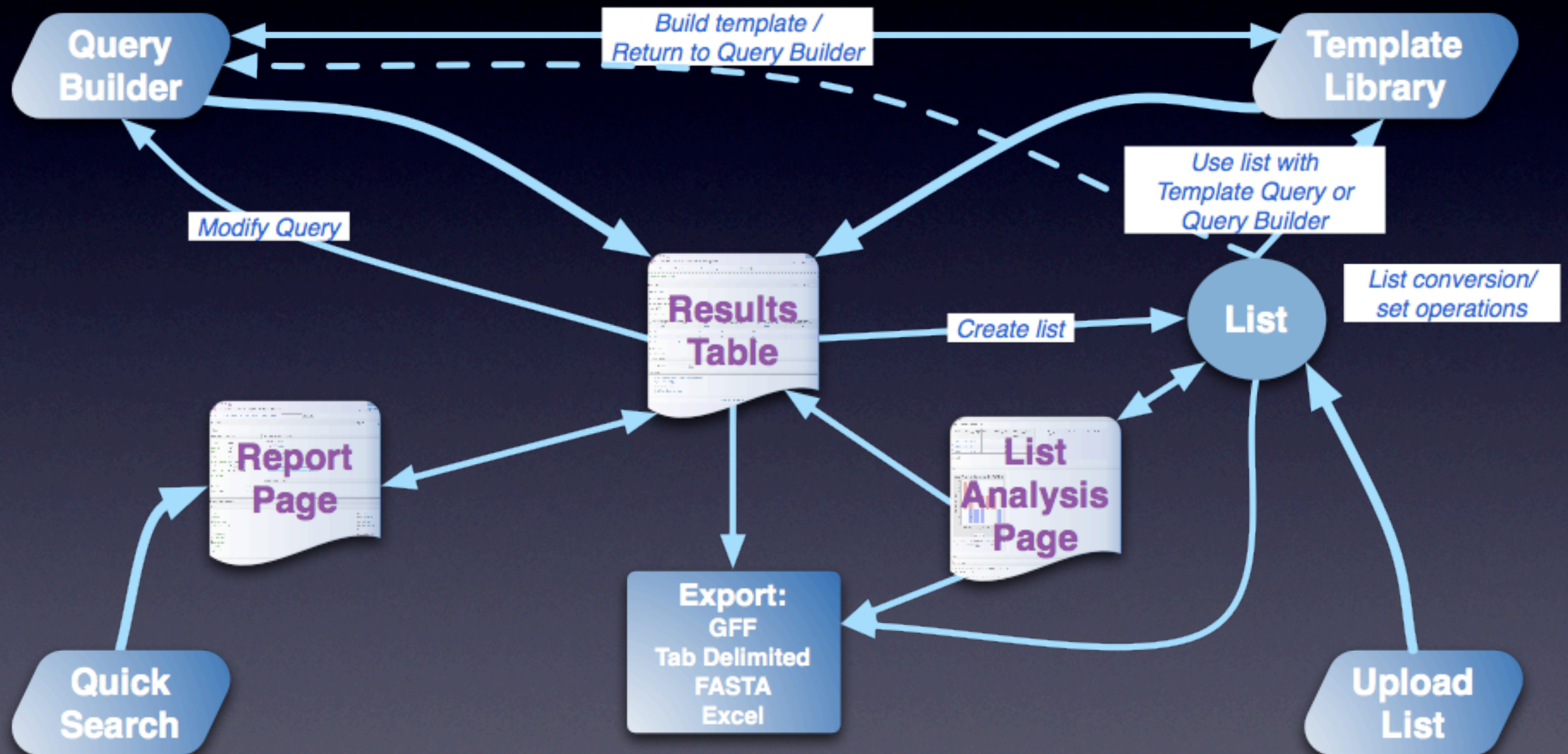
# Web Application

- Works for any data model
- Advanced functionality for bench biologist
- Highly configurable
- Configuration from within web interface





# Webapp Overview





# QueryBuilder

### Model browser ?

Browse through the classes and attributes. Click on [SUMMARY +](#) links to add summary of fields to the results table or on [SHOW +](#) links to add individual fields to the results. Use [CONSTRAIN +](#) links to constrain a value in the query.

- TFBindingSite ?
  - curated Boolean [SHOW +](#) [CONSTRAIN +](#)
  - identifier [SHOW +](#) [CONSTRAIN +](#)
  - length [SHOW +](#) [CONSTRAIN +](#)
  - name [SHOW +](#) [CONSTRAIN +](#)
  - [-] annotations Annotation collection [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] chromosome Chromosome ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] chromosomeLocation Location ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] evidence Evidence ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] factor Gene ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] gene Gene ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] motif Motif ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] objects Relation ? collection [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] organism Organism ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] overlappingFeatures LocatedSequenceFeature ? collection [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] relations SymmetricalRelation ? collection [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] sequence Sequence ? [SUMMARY +](#) [CONSTRAIN +](#)
  - [-] subjects Relation collection [SHOW +](#) [CONSTRAIN +](#)
  - [-] synonyms Synonym ? collection [SUMMARY +](#) [CONSTRAIN +](#)

### Constraints on the current query ?

Click on a class name below to view its fields

- TFBindingSite ?
- factor Gene ?
- IN demo bag ?
- chromosomeLocation Location ?
- object Chromosome ?
- evidence DataSet collection ?

### Constraint logic ?

one constraint

### Fields selected for output ?

#### Columns to Display

Use the [SHOW +](#) or [SUMMARY +](#) links to add fields to the results table. Click and drag the blue output boxes to choose the output column order.

TFBindingSite > factor > identifier [X](#)  
[EDIT](#)

TFBindingSite > factor > length [X](#)  
[EDIT](#)

TFBindingSite > factor > name [X](#)  
[EDIT](#)

TFBindingSite > factor > object > identifier [X](#)  
[EDIT](#)

TFBindingSite > chromosomeLocation > start [X](#)  
[SORT](#)

TFBindingSite > chromosomeLocation > end [X](#)  
[SORT](#)

TFBindingSite > gene > identifier [X](#)  
[SORT](#)

TFBindingSite > gene > symbol [X](#)  
[SORT](#)

#### Sort Results By Column

To sort the results by a specific field, click on [SORT +](#) in that field's blue box. Use the button in the purple box below to reverse the direction of the sort. Click [A-Z](#) to sort in ascending order. Click [Z-A](#) to sort the results in descending order.

TFBindingSite > factor > identifier  
[A-Z](#)

## Query summary

## Model browser

## Constraint editor

## Order output columns

## Set sort order






Demo



# Superuser

- Non-programmer can configure
- Public template queries
- Public lists
- Templates on report pages
- Tagging

	<div>Chromosome_Gene </div> <div> aspect:Genomics [x] im:public [x]</div> <div><input type="text" value="im:frontpage"/> <input type="button" value="Add"/> <input type="button" value="Done"/></div>	<div>Chromosome --&gt;</div> <div>All genes.</div>
---	---	--



# Query API

```
PathQuery q = new PathQuery(model);  
  
q.setView("Protein.primaryIdentifier",  
          Protein.genes.primaryIdentifier");  
  
q.addConstraint("Protein.proteinDomains.name",  
                Constraints.eq("Homeobox"));  
q.addConstraint("Protein.organism.shortName",  
                Constraints.eq("D. melanogaster"));
```



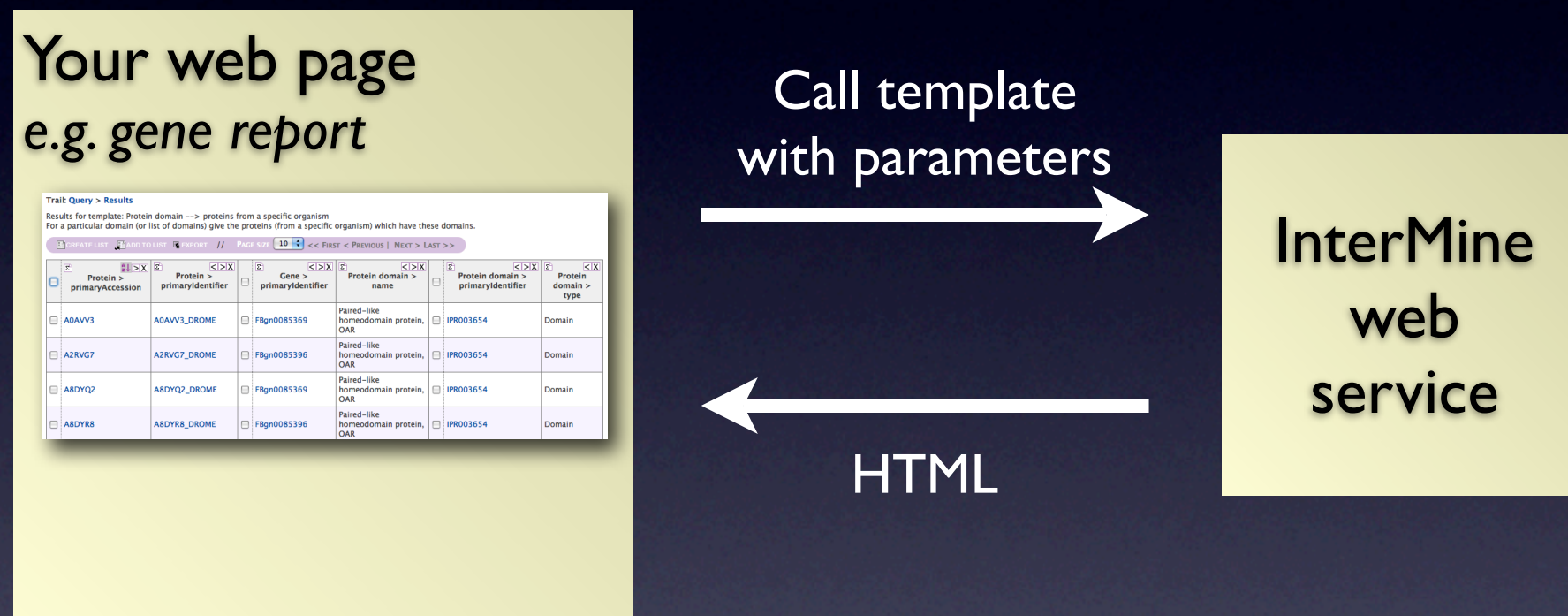
# Web Services

- RESTful web service
- Run queries or templates
  - export XML from web app or use query API
- Java client API
- *Perl client API coming soon*
- *Lists, widgets, logins, tags to be added*



# Embedding Templates

- Web service can return HTML



- ‘Embed this template’ link
- Saves remote site from integrating data
- *Widgets coming soon*



# InterMine Workshop

- 30-31 October (to be confirmed)
- Cambridge, UK
- Contact: [richard@flymine.org](mailto:richard@flymine.org)



# Acknowledgments

**Biologists** Hilde Jannsens, Rachel Lyne

**Developers** Richard Smith, Jakub Kulaviak, Julie Sullivan, Matthew Wakeling, Xavier Watkins

**Sys Admin** Dan Tomlinson

**modENCODE** Sergio Contrino, Kim Rutherford

**PI** Gos Micklem

**[www.intermine.org](http://www.intermine.org)**