Implementing FAIR identifiers in InterMine with Identifiers.org

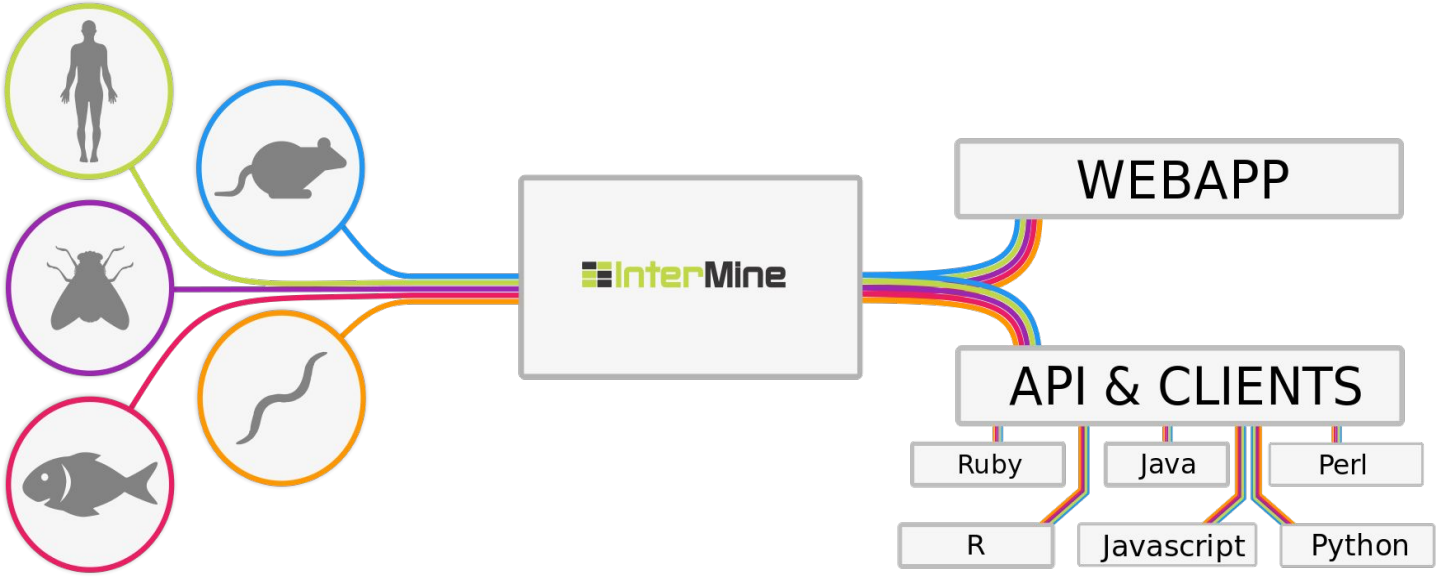
Daniela Butano, Software Engineer - daniela@intermine.org
J.Clark-Casey, S.Contrino, J.Heimbach, R.Lyne, J.Sullivan,Y.Yehudi, and G.Micklem

Building a FAIR Bioinformatics environment
October 22-24, 2018, Genoa

UNIVERSITY OF CAMBRIDGE

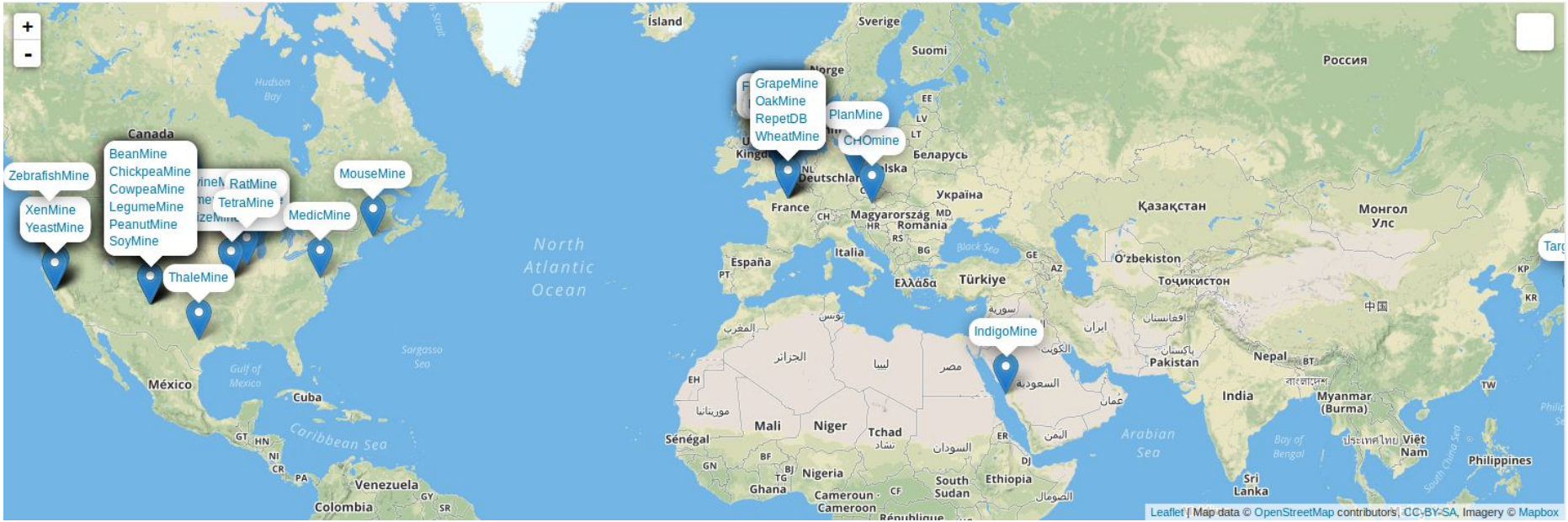
InterMine

InterMine Overview



Model organism images Designed by Freepik and distributed by Fla

InterMine around the world



UNIVERSITY OF CAMBRIDGE

InterMine

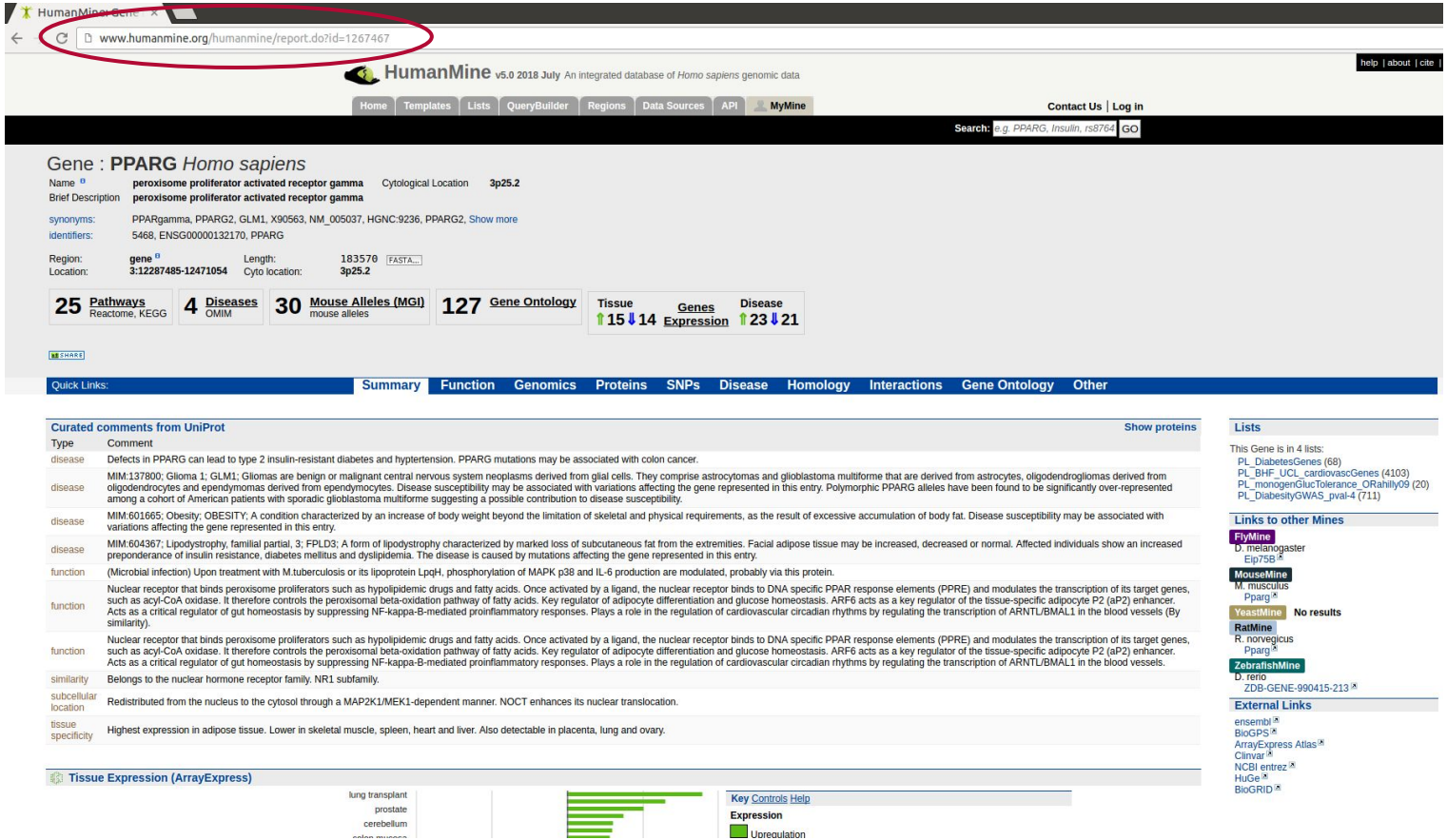
URI: Uniform Resource Identifiers

From FAIR principles:
F1: (meta)data are assigned a globally unique and persistent identifier
A1: (meta)data are retrievable by their identifier using a standardized communication protocol

Assigning URIs to entities:

```
http://purl.uniprot.org/uniprot/P12883
http://identifiers.org/biosample/SAMN00808411
http://identifiers.org/ncbigene/5468
```

Current InterMine URIs change at every build



UNIVERSITY OF CAMBRIDGE

InterMine

New FAIR identifiers for InterMine

Identified two possible solutions

- ✓ No InterMine generated IDs
- ✓ Based on external IDs: LUI = Local Unique Identifier issued by the data source provider

1. Using Identifiers.org prefixes: humanmine.org/humanmine/**ncbigene:5468**
2. Using InterMine core model class names: humanmine.org/humanmine/**gene**:**5468**

Schema A: using Identifiers.org prefixes and LUI

humanmine.org/humanmine/**ncbigene:5468**

- ✓ URI is guaranteed unique
- ✓ No embedded type
- ✓ Automatic links to other sources

Prefixes in Identifiers.org
biosample
ensembl
kegg.pathway
ncbigene
pubmed
reactome
uniprot

Schema B: using core model class names and LUI

humanmine.org/humanmine/**gene**:**5468**

- ✓ No configuration required
- ✓ Human-readable
- ✓ Straightforward URI resolution

InterMine core.xml

```
<class name="BioEntity" is-interface="true" extends="Annotatable">
  <attribute name="secondaryIdentifier" type="java.lang.String"/>
  <attribute name="symbol" type="java.lang.String"/>
  <attribute name="name" type="java.lang.String"/>
....
</class>
<class name="SequenceFeature" extends="BioEntity" is-interface="true">
  <attribute name="score" type="java.lang.Double"/>
  <attribute name="length" type="java.lang.Integer"/>
....
</class>
<class name="Gene" extends="SequenceFeature" is-interface="true">
  <attribute name="description" type="java.lang.String"/>
  <attribute name="briefDescription" type="java.lang.String"/>
.....
</class>
```

UNIVERSITY OF CAMBRIDGE

InterMine

Schema B 

It requires less human intervention

Identifiers.org as Permanent URL provider

Registering HumanMine in Identifiers.org with URI *http://identifiers.org/humanmine*

http://identifiers.org/humanmine:protein:D2KUA6

http://humanmine.org/humanmine/protein:D2KUA6

InterMine Team

Gos Micklem Lab

- Daniela Butano
- Sergio Contrino
- Julie Sullivan
- Yo Yehudi
- Rachel Lyne

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

UNIVERSITY OF CAMBRIDGE

InterMine

Thank you