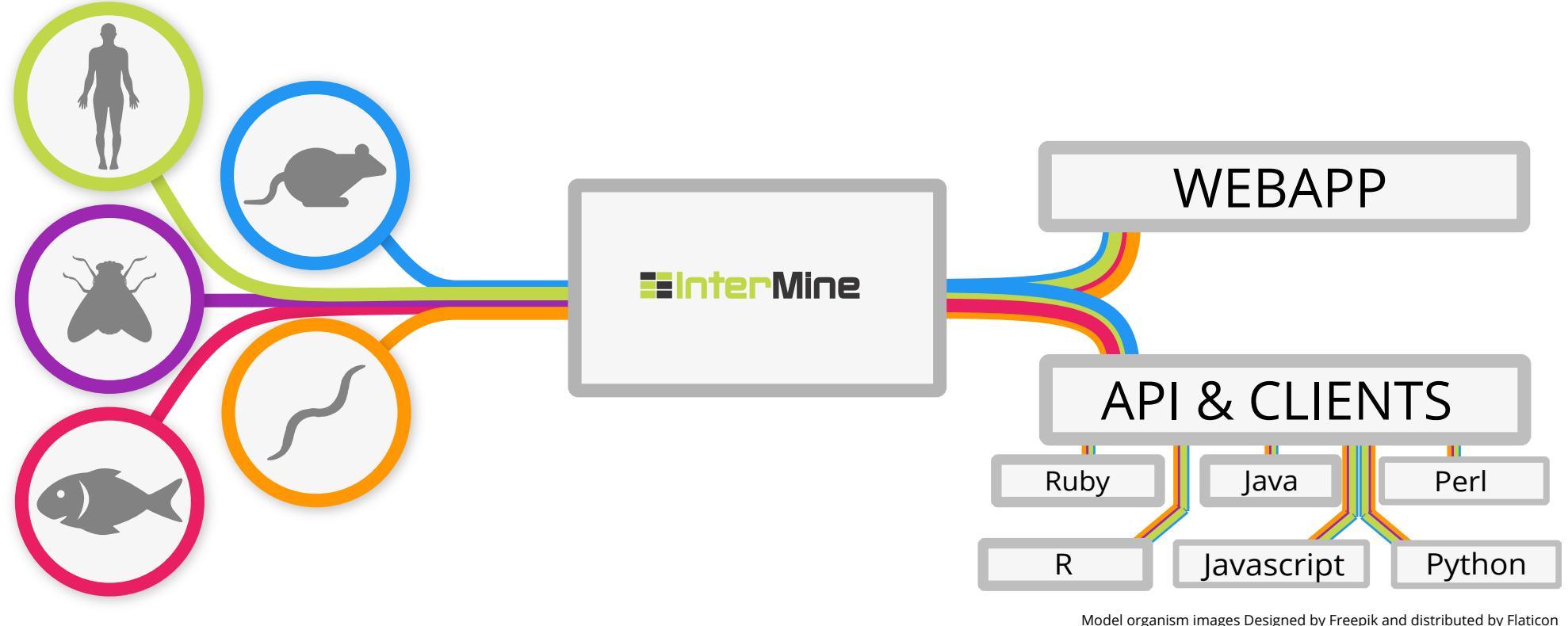


The FAIR data principles and their practical implementation in InterMine

D. Butano, J. Clark-Casey, S. Contrino, M. Chadwick, J. Heimbach, R. Lyne, Y. Yehudi, J. Sullivan and G. Micklem
Department of Genetics, University of Cambridge, Downing Site, Cambridge CB1 3EH

InterMine is an open source system to integrate many commonly used biological data sources and formats. It provides data through a web interface, comprehensive RESTful web services and client APIs for many programming languages.



Findable Accessible Interoperable Reusable

F

- F1.** (meta)data are assigned a globally unique and persistent identifier
- F2.** data are described with rich metadata
- F3.** metadata clearly and explicitly include the identifier of the data it describes
- F4.** (meta)data are registered or indexed in a searchable resource

A

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1.** the protocol is open, free, and universally implementable
 - A1.2.** the protocol allows for an authentication and authorization procedure, where necessary
 - A1.2.** metadata are accessible, even when the data are no longer available

I

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
 - I2.** (meta)data use vocabularies that follow FAIR principles
 - I3.** (meta)data include qualified references to other (meta)data

R

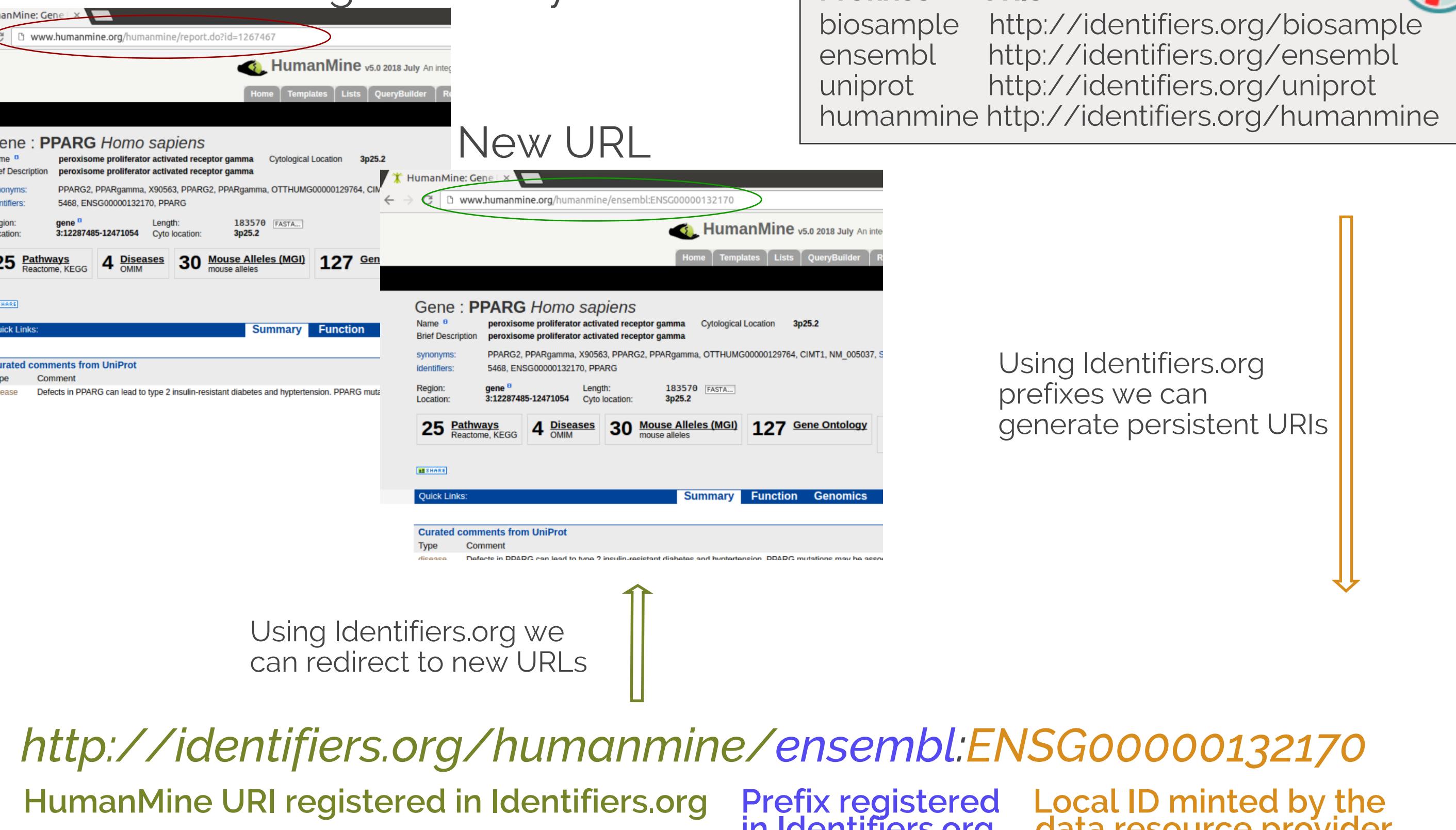
- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1.** (meta)data are released with a clear and accessible data usage licenses
 - R1.2.** (meta)data are associated with detailed provenance
 - R1.3.** (meta)data mmet domain-relevant community standards

The FAIR Guiding Principles for scientific data management and stewardship
Scientific Data volume 3, Article number: 160018 (2016)

Applying FAIR to InterMine

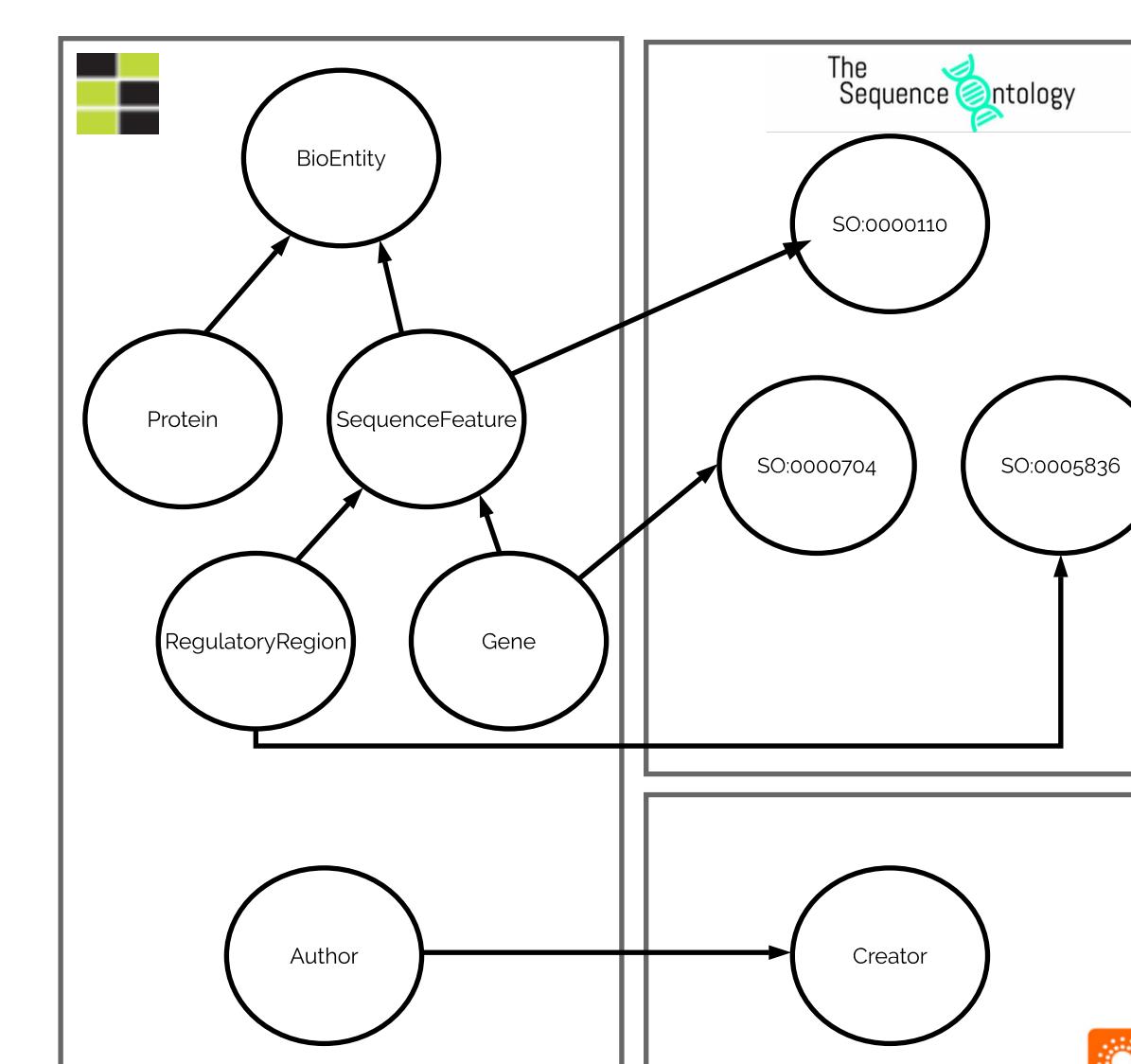
Generate persistent URIs for data

Current URLs change on every rebuild



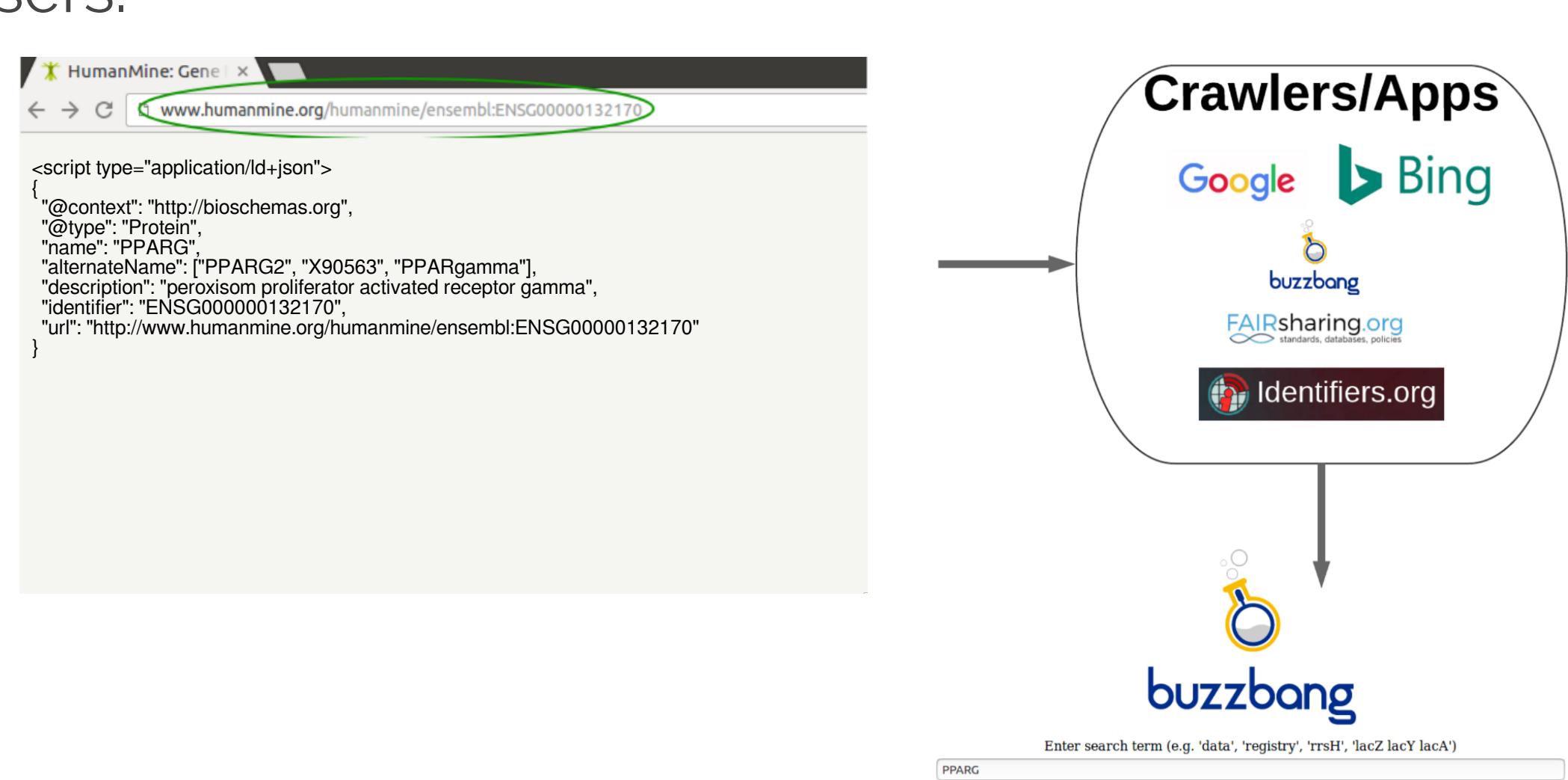
Describe data with ontologies

InterMine is based on a generic data model describing classes and their relationships. These descriptions are currently simple labels ("gene", "chromosome", "located_on", etc.). To improve interoperability we will supplement these with terms from ontologies such as the Sequence Ontology (biological sequence features), as used by other data resources.



Markup our web pages

Apply markup (standardized through bioschemas.org) to our data web pages. Search engines can then find it and give more relevant results to users.



Link data resources

```
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX ensembl:<http://rdf.ebi.ac.uk/resource/ensembl/>
PREFIX humanmine:<http://www.humanmine.org/humanmine/>
PREFIX sio:<http://semanticscience.org/resource/>
PREFIX uniprot:<http://purl.uniprot.org/uniprot>
```

External biological resources

