

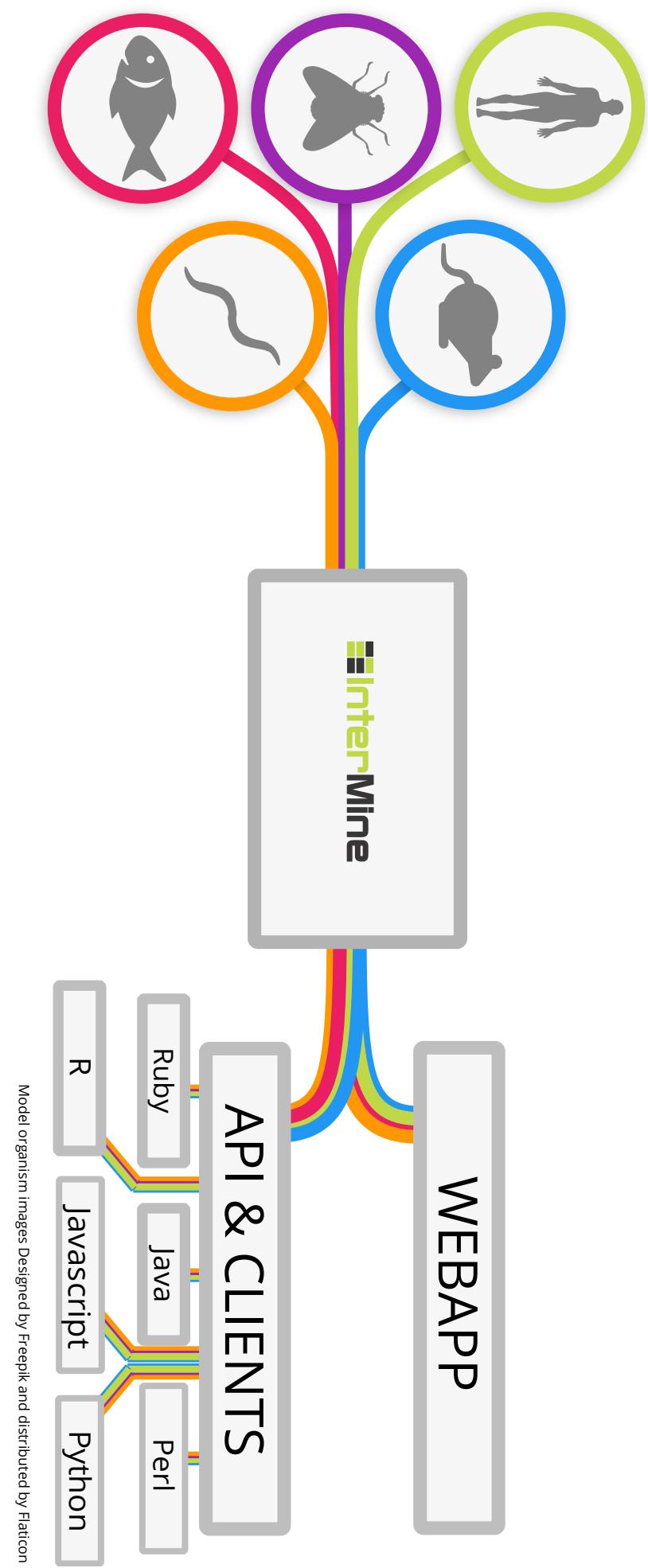
The FAIR data impl

D. Butano, J. Clark-Casey, S. Co
Department of Ge

Eindzicht

The FAIR principles aim to

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.



FullAccessibleInteroperableReusable

Accessible Interoperable Reusable

Science is generating data academia and industry that integrated and shared across. Hence, the FAIR guidelines traction in the life sciences.

But how do these affect them when developing the

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 identified using a communication protocol
 - A1.1.** the protocol is universally implemented
 - A1.2.** the protocol is authentication and access control procedure, where applicable
- A2.** metadata are registered or indexed in a searchable resource

when the data are

A

Generate persistent URIs for data

Current URLs change on every rebuild

HumanMine: Gene < www.humanmine.org/humanmine/report?doiId=1267467

HumanMine v5.0 2018 July An integrat

Gene : **PPARG** *Homo sapiens*

Name: peroxisome proliferator activated receptor gamma
Brief Description: peroxisome proliferator activated receptor gamma
Synonyms: PPARG2, PPARgamma, X90563, PPARG2, PPARgamma, OTTHUMG00000129764, CIN
Identifiers: 5468, ENSG00000132170, PPARG
Region: gene
Length: 183570
Location: 3:12287485-12471054
Cytlo location: 3p25.2

New URL

HumanMine: Gene < www.humanmine.org/humanmine/ensembl:ENSG00000132170

HumanMine v5.0 2018 July An integrat

Gene : **PPARG** *Homo sapiens*

Name: peroxisome proliferator activated receptor gamma
Brief Description: peroxisome proliferator activated receptor gamma
Synonyms: PPARG2, PPARgamma, X90563, PPARG2, PPARgamma, OTTHUMG00000129764, CIMT1, NM_005037, S
Identifiers: 5468, ENSG00000132170, PPARG
Region: gene
Length: 183570
Location: 3:12287485-12471054
Cytlo location: 3p25.2

127 Genes

25 Pathways

Reaction, KEGG

4 Diseases

OMIM

30 Mouse Alleles (MGD)

mouse alleles

SHARE

Quick Links:

Summary Function

Curated comments from UniProt

Type Comment

disease

Defects in PPARG can lead to type 2 insulin-resistant diabetes and hypertension. PPARG muta

HumanMine: Gene < www.humanmine.org/humanmine/report?doiId=1267467

HumanMine v5.0 2018 July An integrat

Gene : **PPARG** *Homo sapiens*

Name: peroxisome proliferator activated receptor gamma
Brief Description: peroxisome proliferator activated receptor gamma
Synonyms: PPARG2, PPARgamma, X90563, PPARG2, PPARgamma, OTTHUMG00000129764, CIMT1, NM_005037, S
Identifiers: 5468, ENSG00000132170, PPARG
Region: gene
Length: 183570
Location: 3:12287485-12471054
Cytlo location: 3p25.2

127 Genes

25 Pathways

Reaction, KEGG

4 Diseases

OMIM

30 Mouse Alleles (MGD)

mouse alleles

SHARE

Quick Links:

Summary Function

Curated comments from UniProt

Type Comment

disease

Defects in PPARG can lead to type 2 insulin-resistant diabetes and hypertension. PPARG muta

US
pre

25 Pathways
Reactome, KEGG

4 Diseases
OMIM

30 Mouse Alleles (MGJ)
mouse alleles

127 Gene Ontology

ge

Quick Links: [Summary](#) [Function](#) [Genomics](#)

Curated comments from UniProt
Type Comment
Reference in DDDRG can lead to human, mouse and ratMine. DDDRG mutations may have been

Using Identifiers.org we
can redirect to new URLs



<http://identifiers.org/humanmine/ensembl:ENSG00000132170>

HumanMine URI registered in Identifiers.org
[Prefix registered
in Identifiers.org](#)

Markup our web pages

Apply markup (standardized through *bioschemas*) to webpages. Search engines can then find it and give more information to users.

PPARG
Identifier
Name PPARG
Type Protein
Description peroxisom proliferator activated receptor gamma
Alternative names [PPARG2, 'X90563', 'PPARgamma']
From <http://www.humanmine.org/humanmine/ensembl:ENSG00000132170>

Results

Enter search term (e.g. 'data', 'registry', 'rrsH', 'lacZ lacY lacA')

PPARG

buzzbang

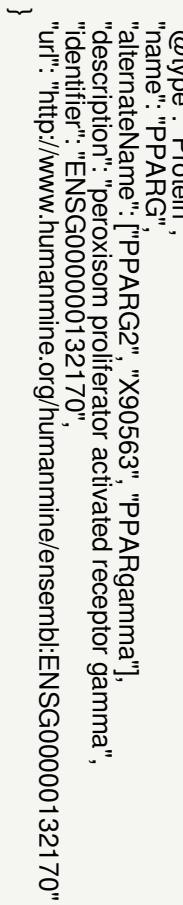


 Identifiers.org

 FAIRsharing.org
standards, databases, policies



"@context": "http://bioschemas.org",
"@type": "Protein",
"name": "PPARG",
"alternateName": ["PPARG2", "X90563", "PPARgamma"],
"description": "peroxisom proliferator activated receptor gamma",
"identifier": "ENSG00000132170",
"url": "http://www.humanmine.org/humanmine/ensembl:ENSG00000132170"



InterMine

Data principles and their implementation in InterMine

Intrino, M. Chadwick, J. Heimbach, R. Lyne, Y. Yehudi, J. Sutherland, University of Cambridge, Downing Site, Cambridge

Structured searches

Find and filter data

Disease > Name
CONTAINS diabetes

Organism > Short Name
= M.musculus

Search: Connect Us | Log In | Show Results

Help | about | cito | references | Edit Query

HumanMine v5.0 (July 2018) An integrated database of Homo sapiens genomic data

Home | Temples | Lists | OntoBuilder | Regions | Data Sources | API | MyMine

Human Disease → Human Gene + Orthologue Gene(s)

For a given human disease returns list of associated or implicated human genes and orthologues in other species (mouse and rat). Default mouse (disease source: OMIM)

Manage Columns | Manage Filters | Manage Relationships

Show 1 to 25 of 91 rows

Primary Identifier	Disease Name	Gene Symbol	Homologene Primary Identifier	Homologene Symbol
OMIM:175800	DIABETES INSIPIDUS, NEPHROGENIC, AUTOSOMAL	AQP2	MGI:109845	Aqp2
OMIM:175850	MATERNAL-ONSET DIABETES OF THE YOUNG, TYPE 1	HNF4A	MGI:109128	Hnf4a
OMIM:175851	MATERNAL-ONSET DIABETES OF THE YOUNG, TYPE 2	GCK	MGI:1270834	Gck
OMIM:175852	DIABETES MELLITUS, INSULIN-DEPENDENT 2	INS	MGI:96572	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	IGF1BP2	MGI:1893238	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	SLC30A8	MGI:242682	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	AKT2	MGI:104874	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	CGR	MGI:9572	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	GCK	MGI:1270834	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	GPD2	MGI:9578	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	HMGCR	MGI:6160	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	INSR	MGI:109128	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	PDX1	MGI:102851	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	IBST1	MGI:9454	

Trail Query

Human Disease → Human Gene + Orthologue Gene(s)

For a given human disease returns list of associated or implicated human genes and orthologues in other species (mouse and rat). Default mouse (disease source: OMIM)

Manage Columns | Manage Filters | Manage Relationships

Show 1 to 25 of 91 rows

Primary Identifier	Disease Name	Gene Symbol	Homologene Primary Identifier	Homologene Symbol
OMIM:175800	DIABETES INSIPIDUS, NEPHROGENIC, AUTOSOMAL	AQP2	MGI:109845	Aqp2
OMIM:175850	MATERNAL-ONSET DIABETES OF THE YOUNG, TYPE 1	HNF4A	MGI:109128	Hnf4a
OMIM:175851	MATERNAL-ONSET DIABETES OF THE YOUNG, TYPE 2	GCK	MGI:1270834	Gck
OMIM:175852	DIABETES MELLITUS, INSULIN-DEPENDENT 2	INS	MGI:96572	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	IGF1BP2	MGI:1893238	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	SLC30A8	MGI:242682	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	AKT2	MGI:104874	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	CGR	MGI:9572	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	GCK	MGI:1270834	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	GPD2	MGI:9578	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	HMGCR	MGI:6160	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	INSR	MGI:109128	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	PDX1	MGI:102851	
OMIM:175853	DIABETES MELLITUS, NONINSULIN-DEPENDENT	IBST1	MGI:9454	

Export results in JSON, XML, tabular, GFF3 and FASTA formats

Gene Ontology Enrichment

Number of Genes in this list not analysed in this widget: 2

Test Correction: Bonferroni | 0.05 | biological_process

Background population: Default | Change

View | Download | GO Term | P-Value | Matches

peptide hormone secretion [GO:0030072] | 9.056043e-19 | 20

hormone secretion [GO:0046679] | 2.5564347e-18 | 21

hormone transport [GO:0009914] | 5.443972e-18 | 21

insulin secretion [GO:0002073] | 1.814829e-15 | 17

regulation of peptide hormone secretion | 1.814625e-15 | 17

[GO:0090276] | 1.814625e-15 | 17

signal release [GO:0023061] | 2.071971e-15 | 21

regulation of hormone secretion | 4.564819e-15 | 18

CROSS-C
register
differer
around
the
world

FAIR Principles

Enrichment Statistics



o make data more Findable, Accessible, Interoperable and

a faster than ever before. Reliably storing and retrieving
nat datasets, produced at different times by different ins
ain without days of laborious manual effort.

es - to make data findable, accessible, interoperable an
's and beyond.

the practical design of software systems? In this poster
ne InterMine platform.

I

- are retrievable by their
 - a standarized
 - s protocol
 - ol is open, free, and
 - ementable
 - ol allows for an
 - and authorization
 - re necessary
 - are accessible even
- 1.** (meta)data use a formal, accessible
 - shared, and broadly applicable
 - language for knowledge representation
 - 2.** (meta)data use vocabularies that
 - follow FAIR principles
 - 3.** (meta)data include qualified
 - references to other (meta)data

Applying FAIR to InterMine

Describe data with a

InterMine is based on a generic relationships. These descriptions "chromosome", "located_on", etc supplement these with terms from Ontology (biological sequence



<https://identifiers.org/biosample>
<https://identifiers.org/ensembl>
<https://identifiers.org/uniprot>
<https://identifiers.org/humanmine>



generate persistent URIs

RegulatoryRegion

Protein



SG00000132170

Local ID minted by the
data resource provider

Link data resources

```
PREFIX owl: <http://  
PREFIX dc: <http://  
PREFIX ensembl: <h  
PREFIX humanmine:  
PREFIX sio: <http://  
PREFIX uniprot: <htt
```

s.org) to our data
more relevant results



sio:010035 (gene)
humanmine:ensembl:ENSG00000132170

dc:identifier

PPARG



amma

[ensembl:ENSG00000132170](#)

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



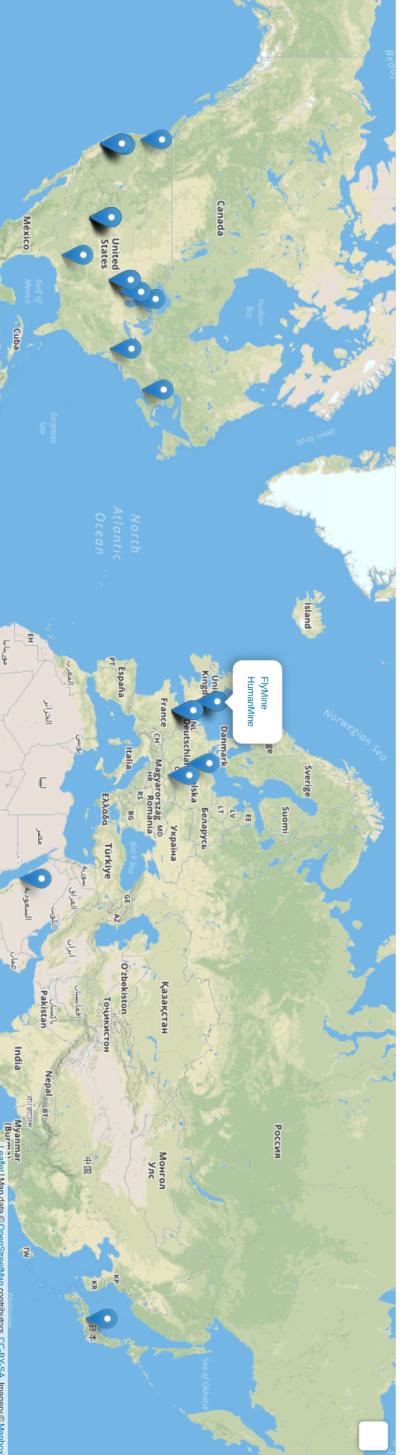
practical

J. M. C. Llivan and G. Micklem

CB1 3EH

area
dozens
of
interlinked
and

compatible biological data resources
the world, built with InterMine and
red at registry.intermine.org, covering
organisms and research targets.



nd Reusable.

it isn't enough - it's increasingly critical for institutions around the globe, can be found,

and reusable as shown below - are gaining

we will explore how we have interpreted

R

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear

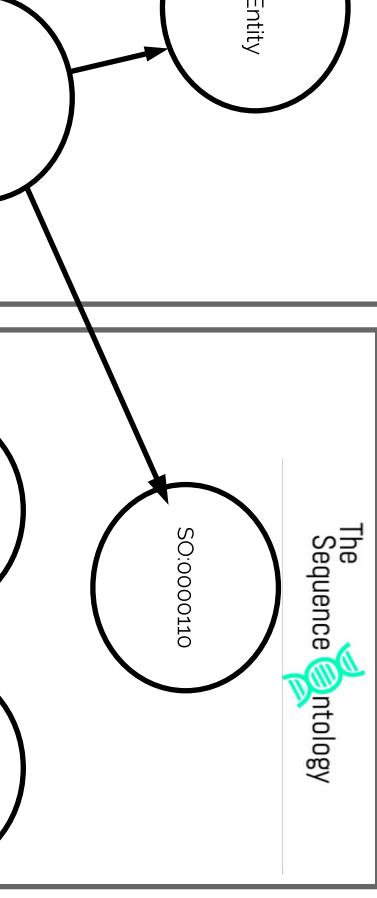
and accessible data usage licenses

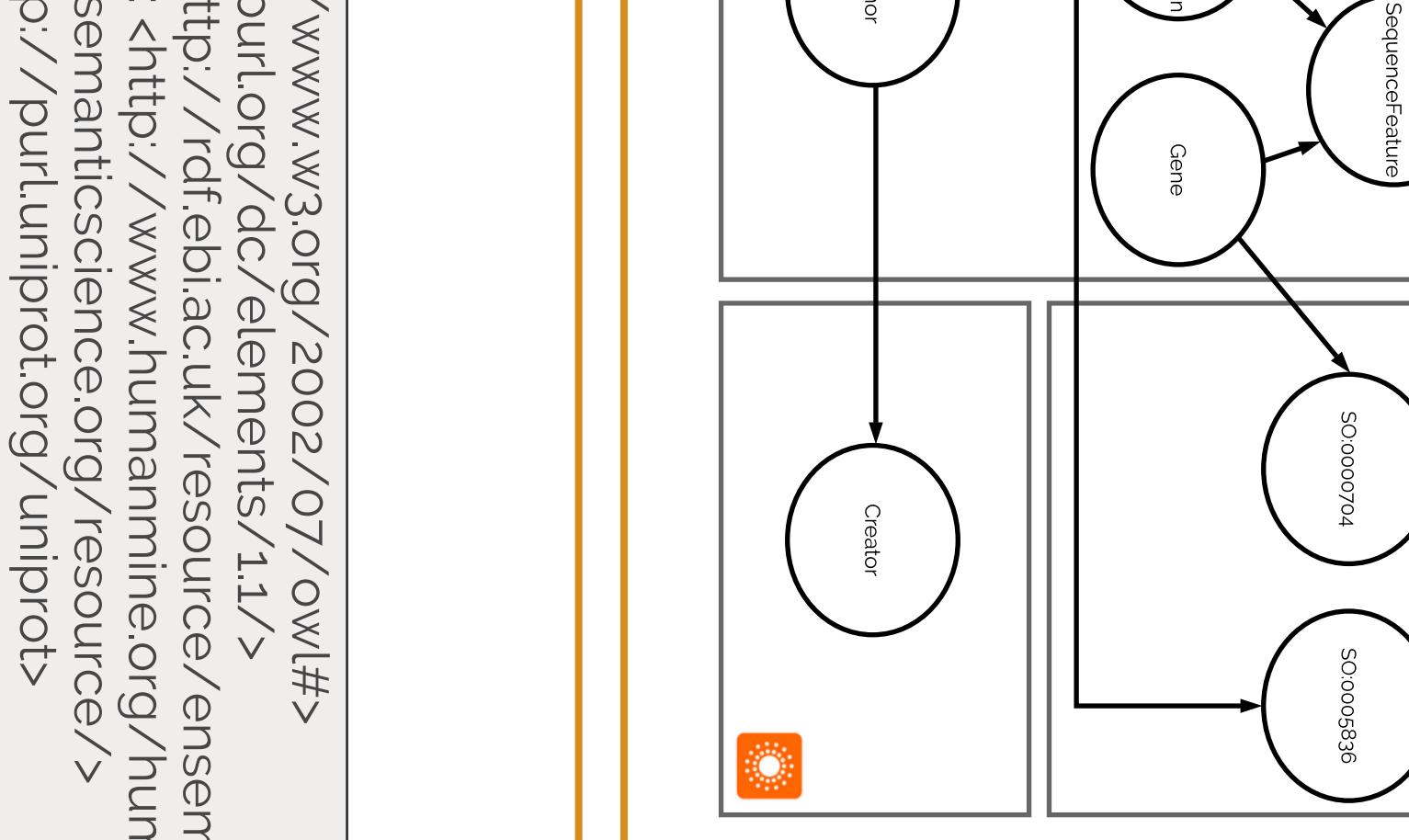
R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

ontologies

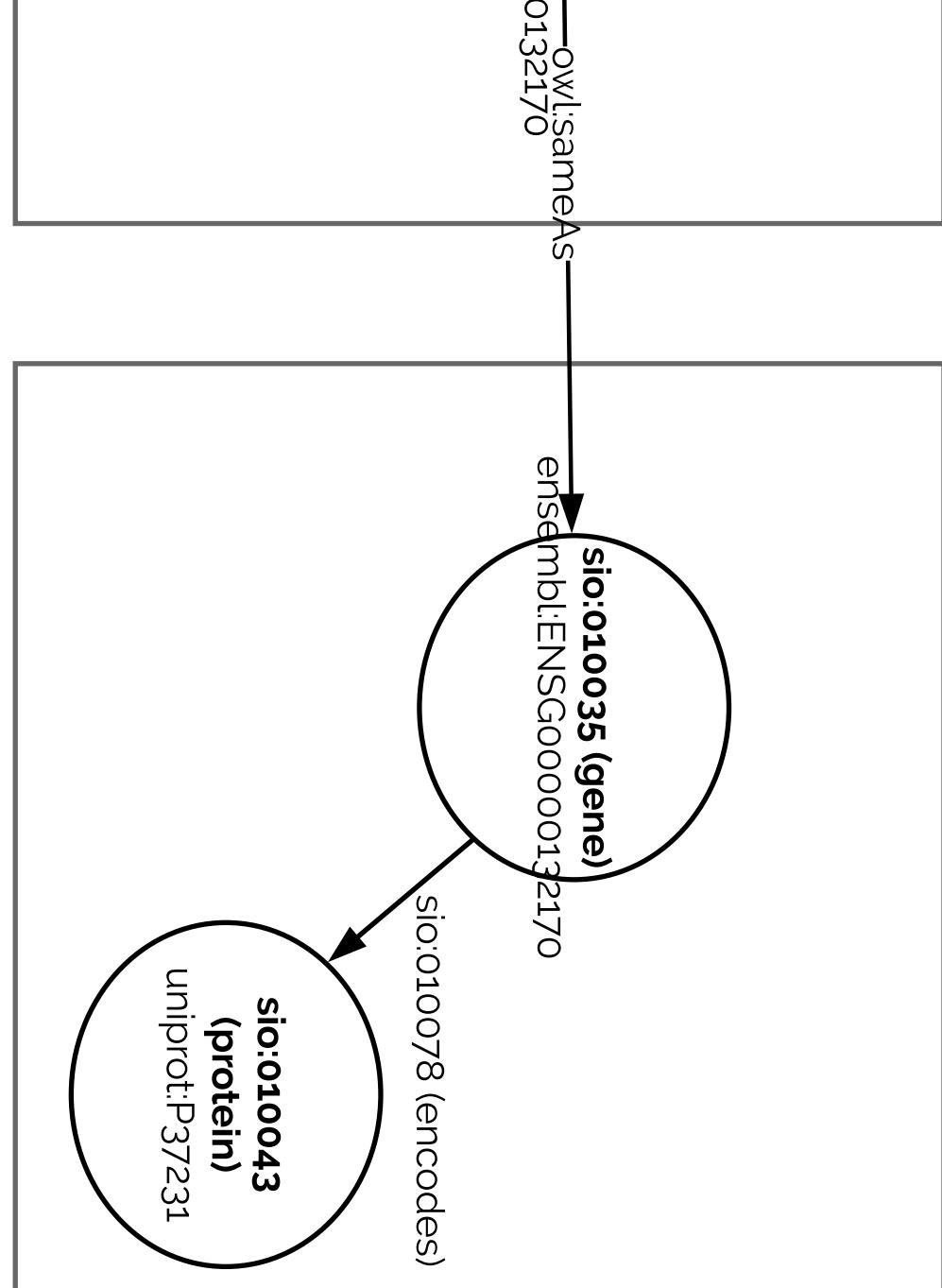
Our data model describing classes and their instances are currently simple labels ("gene", etc.). To improve interoperability we will draw from ontologies such as the Sequence Ontology (features), as used by other data resources.





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

External biological resources



This work was supported by the Wellcome Trust [099133, 208381]

