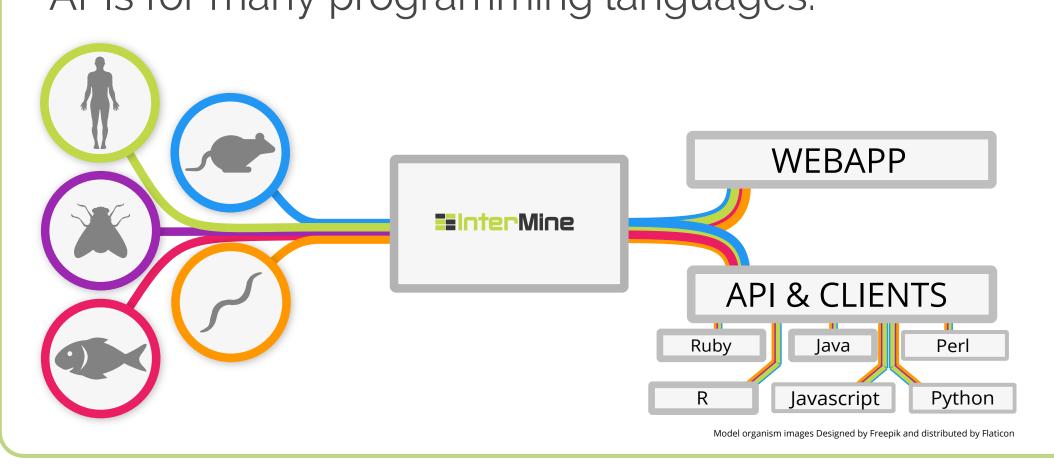
The FAIR data principles and their practical implementation in InterMine

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.



InterMine Structured searches Find and filter data Export results in JSON, XML, tabular, GFF3 and FASTA **Enrichement Statistics**

There dozens interlinked and cross-compatible biological data around the world, built with InterMine registered at registry.intermine.org.

	Name	Description	Organisms
2-	BMAP	Brassicales Map Alignment Project	A. linifolium, C. amplexicaulis, C. hispanica, C. maritima, C. rubella, C.
A.	BeanMine	A mine with common bean data from the Legume Info tripal.chado database	A. ipaensis, A. duranensis, A. thaliana, C. arietinum, G. max, M. truncat
G	BovineMine	An integrated data warehouse for the Bovine Genome Database	B. taurus, C. hircus, O. aries
	CHOmine	An integrated database for Cricetulus griseus and CHO cells	C. griseus
	ChickpeaMine	A mine with chickpea data (both desi and kabuli varieties) from the Legume Information Systems (LIS) tripal.chado database	A. ipaensis, A. duranensis, A. thaliana, C. arietinum desi, C. arietinum
30	CowpeaMine	A mine containing both cowpea genetic and genomic data, courtesy UC-Riverside	A. duranensis, A. ipaensis, C. arietinum (desi), C. arietinum (kabuli), C.
4	FlyMine	An integrated database for Drosophila genomics	D. melanogaster
	GrapeMine	An integrated database for grapevine data	V. vinifera
•	HumanMine	HumanMine integrates many types of data for Homo sapiens and Mus musculus	H. sapiens
4	HymenopteraMine	An integrated data warehouse for the Hymenoptera Genome Database	A. dorsata, A. echinatior, A. florea, A. mellifera, B. impatiens, B. terres
	IndigoMine	INDIGO enables the integration of annotations for the exploration and analysis of newly sequenced microbial genomes.	Archae
	LegumeMine	Multi-organism mine integrates data from legume species: string bean, soybean, peanut and barrel medic	A. duranensis, A. ipaensis, C arietinum, G. max, M. truncatula, P. vulg
	MaizeMine	An integrated data warehouse for MaizeGDB	
	MedicMine	MedicMine integrates many types of data for Medicago truncatula. You can run flexible queries, export results and analyse lists of genes.	A. thaliana, M. truncatula, M. truncatula f. tricycla, M. truncatula var.
	MitoMiner	MitoMiner is an integrated web resource of mitochondrial localisation evidence and phenotype data for mammals, zebrafish and yeasts.	D. rerio, H. sapiens, M. musculus, R. norvegicus, S. cerevisiae, S. pom
	ModMine	A data warehouse for the modENCODE project	
Ĉ	MouseMine	MouseMine is a powerful new system for online access to mouse data from MGI (Mouse Genome Informatics). For more than 20 years, MGI has pro	M. musculus
	OakMine	Quercus robur annotation database	Q. robur
27	PeanutMine	A mine with peanut data from Legume Information Systems (LIS) PeanutBase	A. duranensis, A. ipaensis, A. thaliana, G. max, M. truncatula, P. vulga
2-	PhytoMine	An InterMine interface to data from Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, which	Viridiplantae
	PlanMine	An integrated web resource of data & tools to mine Planarian biology	D. lacteum, P. nigra, P. tenuis, P. torva, S. mediterranea, S. mediterranea
	RatMine	RatMine integrates many types of data for Rattus Norvegicus, Homo Sapiens, Mus Musculus and other organisms. You can run flexible queries, exp	R. norvegicus
	RepetDB	Repeats intermine database	Arabidopsis lyrata subsp. lyrata, Arabidopsis thaliana, Arabis alpina,
<i></i>	SoyMine	A Soybean mine instance with genomic data from the Legume Information Systems (LIS) multi-species chado database and genetic data pulled fro	A. ipaensis, A. duranensis, A. thaliana, C. arietinum, G. max, M. trunc
	TetraMine	Advanced search and retrieval of genome data for Tetrahymena thermophila using Intermine	T. thermophila
Ø	ThaleMine	ThaleMine enables you to analyze Arabidopsis thaliana genes, proteins, gene expression, protein-protein interactions, orthologs, and more.	A. thaliana, H. sapiens, S. cerevisiae
	WheatMine	Wheat IWGSC RefSeq v1.0	T. aestivum
=	WormMine	Intermine data mining platform for C. elegans and related nematodes	C. elegans
	XenMine	Search and retrieve integrated Xenopus tropicalis and Xenopus laevis data	C. elegans, D. melanogaster, R. norvegicus, S. cerevisiae, X. Iaevis, X.
10	YeastMine	Search and retrieve S. cerevisiae data, populated by SGD and powered by InterMine	S. cerevisiae
-	ZebrafishMine	ZebrafishMine is powered by the InterMine data warehouse system, and integrates biological data sets from multiple sources. It currently includes u.	D. rerio

FAIR Principles

The FAIR principles aim to make data more Findable, Accessible, Interoperable and Reusable.

Findable

Accessible

Interoperable

F1. (meta)data are assigned a globally

F2. data are described with rich metadata

F4. (meta)data are registered or indexed

unique and persistent identifier

F3. metadata clearly and explicitly

include the identifier of the data it

in a searchable resource

Reusable

describes

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

A1.2. metadata are accessible, even

when the data are no longer available

procedure, where necessary

integrated and shared again without days of laborious manual effort.

will explore how we've interpreted them when developing the InterMine platform.

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

Science is generating data faster than ever before. Reliably storing and retrieving it isn't enough - it's increasingly critical for

academia and industry that datasets, produced at different times by different institutions around the globe, can be found,

Hence, the FAIR guidelines - to make data findable, accessible, interoperable and reusable as shown below - are gaining

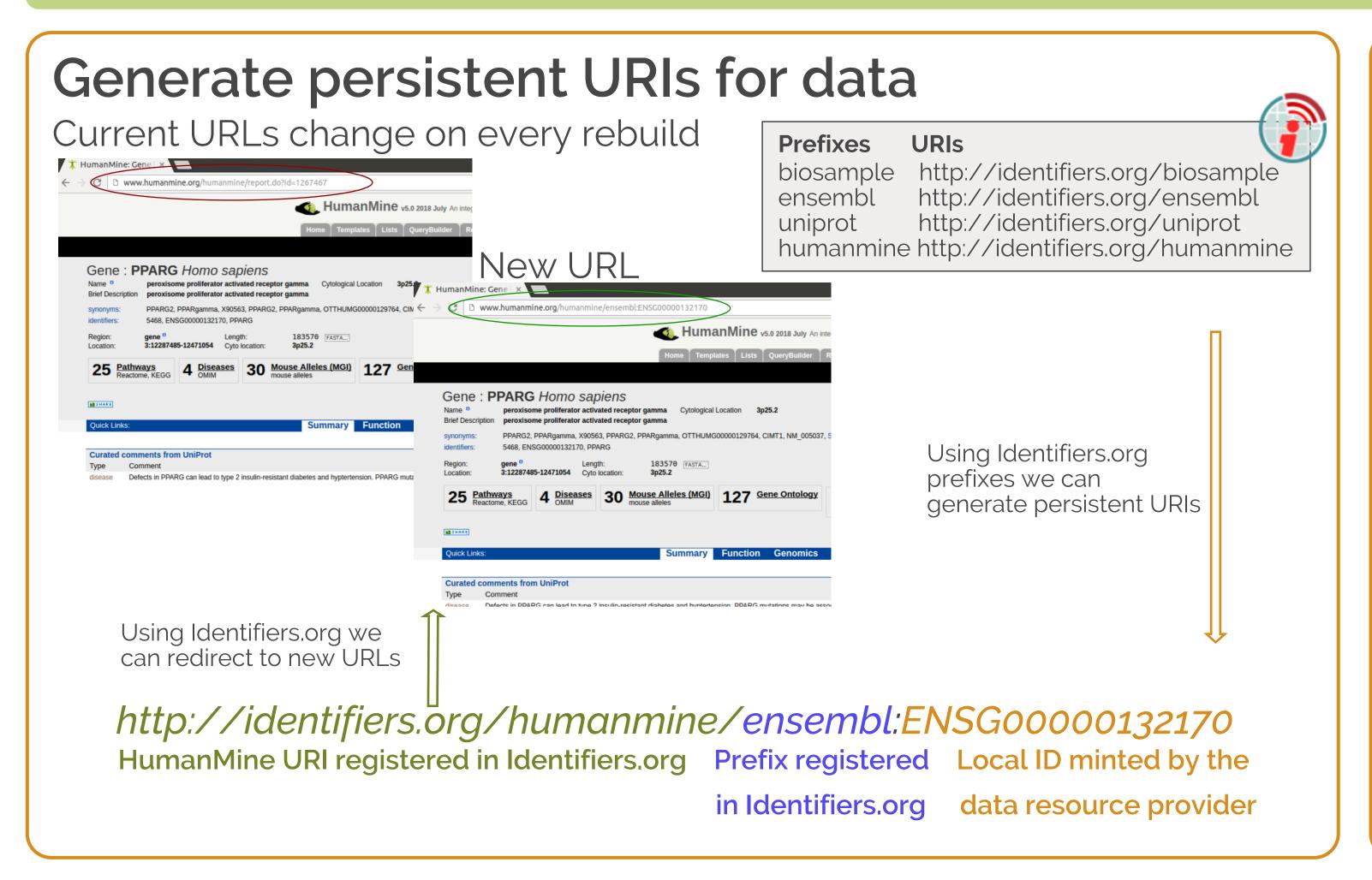
traction in the life sciences and beyond. But how do these affect the practical design of software systems? In this poster we

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 R.1.2. (meta)data are associated with

detailed provenance R.1.3. (meta)data mmet domain-relevant community standards

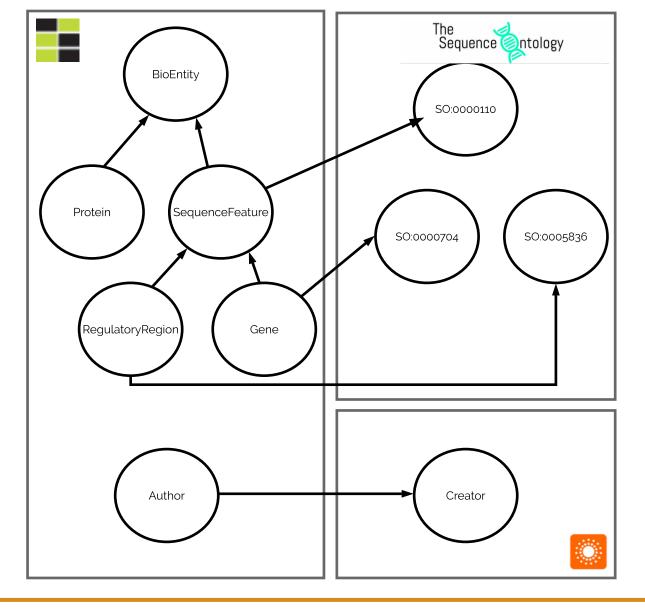
The FAIR Guiding Principles for scientific data management and stewardship Scientific Data volume 3, Article number: 160018 (2016)

Applying FAIR to InterMine



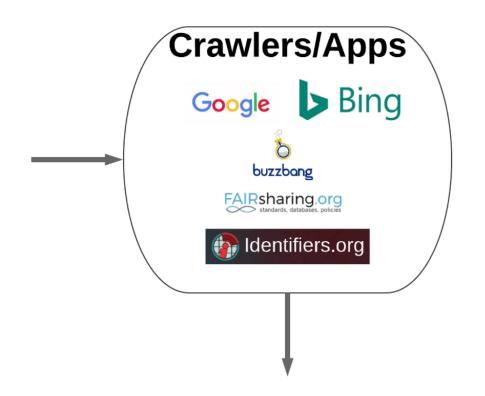
Describe data with ontologies

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) and Dublin Core (library terms), as used by other data resources.



Markup our web pages

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



Link data resources

PREFIX 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

