

Identifiers.org / MIRIAM Registry *identifiers* annotation and cross-referencing framework

European Bioinformatics Institute, an Outstation of the European Molecular Biology Laboratory

Nick Juty, Nicolas Le Novère, Camille Laibe

Introduction

More than ever, the era of *data integration* has highlighted the key requirement to reference specific data in an unambiguous and perennial way, in order to enable community-level sharing, development, exchange and reuse of information. In the field of Systems Biology, which is concerned with creating quantitative models of biological processes, these requirements have directly led to the creation of the Minimal Information Required In the Annotation of Models (MIRIAM, http://biomodels.net/miriam/) [1]. MIRIAM provides a specific set of guidelines that can be implemented within any structured modelling format.

We have developed the MIRIAM Registry (http://www.ebi.ac.uk/miriam/) [2] to support one requirement of the MIRIAM guidelines: the annotation of a model by identifying all its components. This is achieved by means of Uniform Resource Identifiers (URIs).

Identifiers.org is built above the MIRIAM Registry. Created to facilitate the sharing of knowledge beyond the Systems Biology community, Identifiers.org allows the creation of persistent and directly resolvable identifiers, in Uniform Resource Locator (URL) form.

Cross-references Annotations

Annotations are essential for data identification and semantics:

Characteristics of a useful identifier for cross-references:

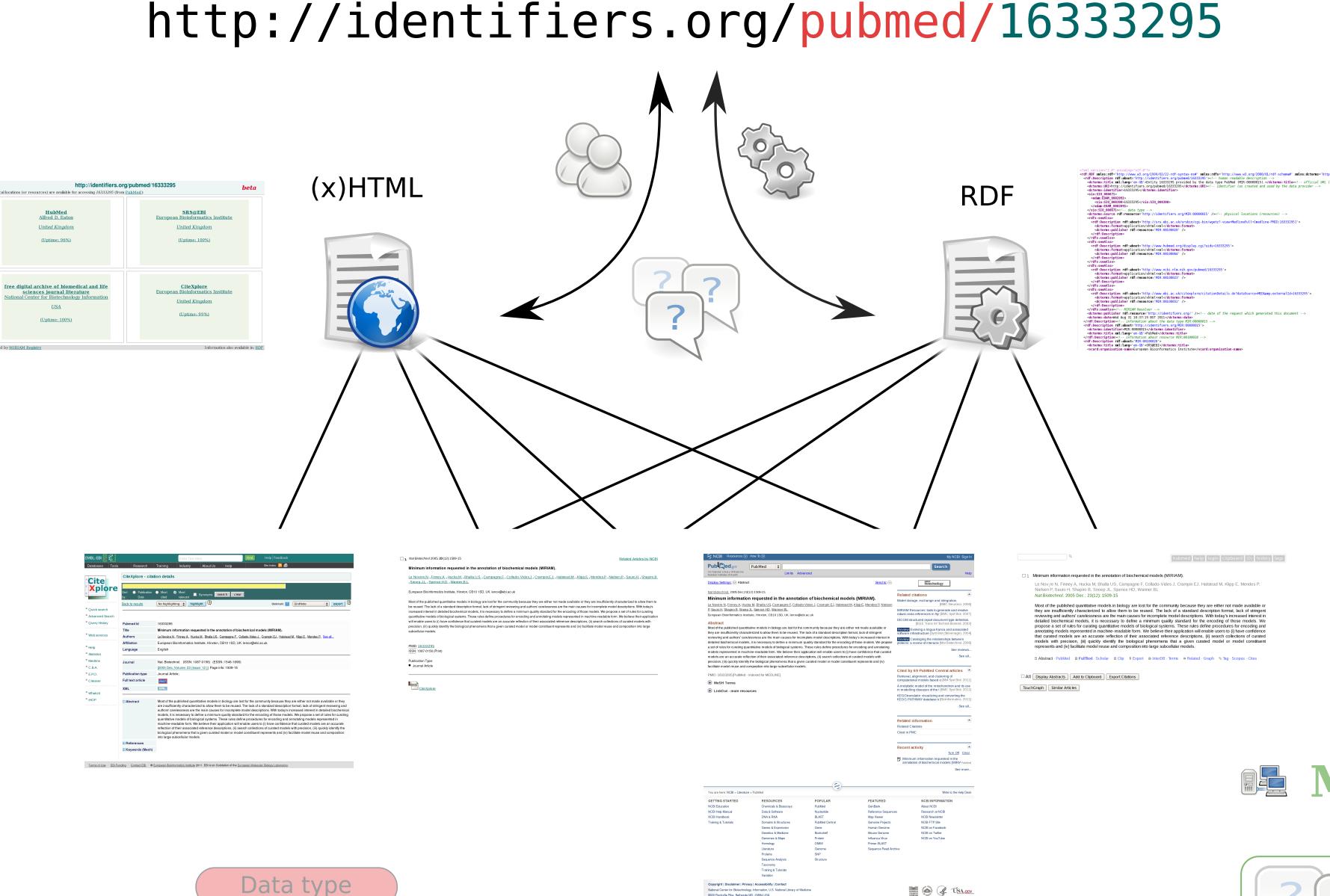
SBML URI example



- data understanding and reuse
- data comparison
- data integration

- unique and unambiguous
- perennial
- standard compliant
- resolvable
- free of use

Identifiers.org architecture



E **MIRIAM URIS**

Human calmodulin: P62158 in UniProt urn:miriam:uniprot:P62158 http://identifiers.org/uniprot/P62158 Alcohol dehydrogenase: 1.1.1.1 in EC code urn:miriam:ec-code:1.1.1.1 http://identifiers.org/ec-code/1.1.1.1 MAPKK activation: GO:0000186 in Gene Ontology urn:miriam:obo.go:G0%3A0000186 http://identifiers.org/obo.go/G0:0000186

Resolving services

The MIRIAM Registry is a set of on-line services centered around a catalogue of data types. Data types can be ontologies, such as Gene Ontology, or primary data resources available via the Web, such as UniProt and PubMed. Each of these is uniquely identified within the MIRIAM database, and information stored regarding the corresponding physical URLs (data resources) through which their associated data can be accessed.

There are two ways to create and resolve annotations arising from data types registered in the database:

i) Web services are available to generate a MIRIAM URI from a data type name and dataset identifier. Resolution services that generate physical locations (URLs) for MIRIAM URIs are provided as SOAP and REST Web Services. ii) Resolvable URLs using the **Identifiers.org** framework. These URLs resolve directly to an intermediate location, providing previews for the total set of all resources where that information may be resolved.

Collado-Vides,J., Crampin,E.J., Halstead.M., Klipp,E., Mendes,P., Nielsen,P., Sauro,H., Shapiro,B.	US National Library of Medicine National Institutes of Health	Limits Adv	anced	
Collado - Vilessi, , Claimphiless, , Palskadum, , Nipple, , Memesie, , Meseller, , Saulori, , Shaphole.	Display Settings; V Abstract			
l.ac.uk	Nat Biotechnol. 2005 Dec;23(1	2):1509-15.		
	Minimum information	requested in the annotati	on of biochemical m	
mmunity because they are either not made available or they are insufficiently characterized to allow them to reviewing and authors' carelessness are the main causes for incomplete model descriptions. With today's effine a minimum quality standard for the encoding of those models. We propose a set of rules for curating	Le Novère N, Finney A, Hucka P. Sauro H, Shapiro B, Snoep J	M, Bhalla US, Campagne F, Collado	-Vides J, Crampin EJ, Halste	
sine a minimum quality samata to in the incoming of more models. We propose a set of meson characteristic res for encoding and annotating models represented in machine-readable form. We believe their application curate reflection of their associated reference descriptions, (ii) search collections of curated models with	European Bioinformatics Institu Abstract	ute, Hinxton, CB10 1SD, UK. lenov@e	bi.ac.uk	
uralled model or model constituent represents and (iv) facilitate model reuse and composition into large	Also track. Most of the published quantitative models in biology are lost for the community because they are they are insufficiently characterized to allow them to be reused. The lack of a standard decipity reviewing and authors cleacelessness and the main causes for incomplete model descriptions. We detailed toochemical models, it is necessary to define a minimum quality standard store theiron a set of housis for cuanting quantitative models of biological systems. These rules define procedur models represented in machine-readable form. We believe their application will enable users to models represented in machine-readable form. We believe their applications, (iii) quarkly users to a set of neglicity defined in the standard believe their applications and is and clear clearly models in a accurate meticicity of their associated reference descriptions, (ii) search collection precision, (iii) quarkly identify the biological phenoment that a given curated model or model co facilitate model reuse and composition in the substellular models.			
	PMID: 16333295 [PubMed - ir	idexed for MEDLINE]		
	MeSH Terms			
	🖶 LinkOut - more resou	rces		
	You are here: NCBI > Literature	> PubMed	3	
	GETTING STARTED	RESOURCES	POPULAR	
	NCBI Education	Chemicals & Bioassays	PubMed	
	NCBI Help Manual	Data & Software	Nucleotide	
	NCBI Handbook Training & Tutorials	DNA & RNA Domains & Structures	BLAST PubMed Central	
	rraning & Tutoriais	Genes & Expression	Gene	
		Genetics & Medicine	Bookshelf	
		Genomes & Mans	Protein	

tract			Send to: 🗸		biotechnology]
;23(12):1509	15.			Related cita	tions	
tion requ	ested in the annotation o	f biochemical mode	is (MIRIAM).		tions , exchange and integr	
lucka M. Bh	alla US, Campagne F, Collado-Vides	J, Crampin EJ, Halstead M	I, Klipp E, Mendes P, Nielsen	mouer siorage,	[BMC]	auon. Veurosci. 2006]
	ice HD, Wanner BL				irces: tools to generate	
Institute, Hin	xton, CB10 1SD, UK. lenov@ebi.ac.ul	k			ferences in Sy: [BMC	
				DICOM structu	ured report document t EEE Trans Inf Technol	ype definition. Biomed. 2003]
	idels in biology are lost for the comm to allow them to be reused. The lack			Review Evolv software infras	ring a lingua franca ar tructure [Syst Biol (Ste	d associated venage). 2004]
	are the main causes for incomplete			Review Catal	loging the relationship	s between
	cessary to define a minimum quality				iew of interactic [Mol Bi	
	models of biological systems. These dable form. We believe their applicat					See reviews
	their associated reference descriptio					See all
	ogical phenomena that a given cura	ted model or model constitu	ent represents and (iv)			
id compositio	on into large subcellular models.			Cited by 69	PubMed Central	articles 🕒
ed - indexed	for MEDLINE]				nment, and clustering	
				computational	models based o (Mol	Syst Biol. 2011]
sources				in modelling d	odel of the mitochondi iseases of the t [BMC	Syst Biol. 2011]
					or: visualizing and cor /AY database tr [Bicinf	
						See all
				Related info	rmation	
				Related Citatio	ms	
				Cited in PMC		
				December 201		
				Recent activ		<u>urn Off</u> <u>Clear</u>
				🔲 Minimum i	information requested	
					of biochemical model	
						See more
ature > PubM	ed				Write	to the Help Desk
	RESOURCES	POPULAR	FEATURED		NCBI INFORMATIO	DN .
	Chemicals & Bioassays	PubMed	GenBank		About NCBI	
	Data & Software	Nucleotide	Reference Sequence	25	Research at NCBI	
	DNA & RNA	BLAST	Map Viewer		NCBI Newsletter	
	Domains & Structures	PubMed Central	Genome Projects		NCBI FTP Site	
	Genes & Expression Genetics & Medicine	Gene Bookshelf	Human Genome Mouse Genome		NCBI on Facebook NCBI on Twitter	
	Generacis & Medicine Genomes & Maps	Protein	Influenza Virus		NCBI on YouTube	
	Homology	OMIM	Primer-BLAST		1001000	
	Literature	Genome	Sequence Read Arcl	nive		

NCBI Education	Chemicals & Bioassays	PubMe
NCBI Help Manual	Data & Software	Nucleo
NCBI Handbook	DNA & RNA	BLAST
Training & Tutorials	Domains & Structures	PubMe
	Genes & Expression	Gene
	Genetics & Medicine	Booksh
	Genomes & Maps	Protein
	Homology	OMIM
	Literature	Genom
	Proteins	SNP
	Sequence Analysis	Structu
	Taxonomy	
	Training & Tutorials	
	Variation	
Copyright Disclaimer Privacy Ad	cessibility Contact	
National Center for Biotechnology Info	mation, U.S. National Library of Medicine	



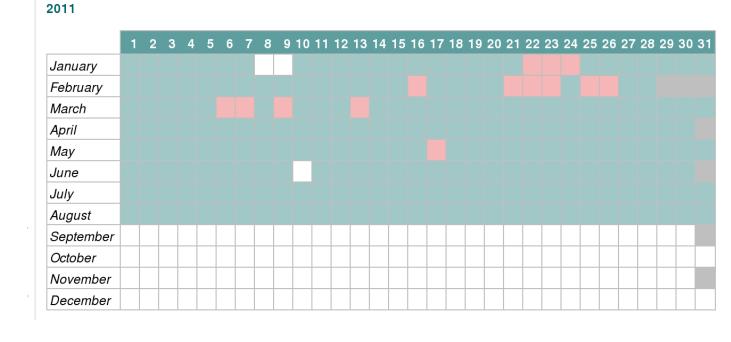
General	Tags	Example Usage	Web Services		
General inform	ation abou	it the data type			
			Name		
Identifier			MIR:0000022		
Name Gi			Gene Ontology		
Synonyms			GO		
			URIS		
Namespace			obo.go		
Deprecated	http://www.geneontology.org/				
			Information		
Definition			The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism.		
Identifier Pati	tern		^GO:\d/7\\$		

Identifier Pattern		^GO:\d{7}\$		
Physical Locations				
	Access URL	http://www.ebi.ac.uk/ego/GTerm?id= \$id [Example: GO:0006915 6]		
Resource	Website	http://www.ebi.ac.uk/ego/		
MIR:00100012	Description	QuickGO (Gene Ontology browser)		
	Institution	European Bioinformatics Institute, United Kingdom		
	Access URL	http://www.informatics.jax.org/searches/GO.cgi?id=\$id [Example: GO:0006915 聲]		
Resource	Website	http://www.informatics.jax.org/searches/GO_form.shtml		
MIR:00100015	Description	GO Browser		
	Institution	The Jackson Laboratory, USA		
	Access URL	http://www.godatabase.org/cgi-bin/amigo/go.cgi?view=details&query=\$1d [Example: GO:0006915 🚱]		
<u>Resource</u>	Website	http://www.godatabase.org/cgi-bin/amigo/go.cgi		
MIR:00100013	Description	AmiGO		
	Institution	The Gene Ontology Consortium, USA		
	Access URL	http://bioportal.bioontology.org/ontologies/1070?p=terms&conceptid=\$id [Example: GO:0006915 2]		
Resource	Website	http://bioportal.bioontology.org/		
MIR:00100237	Description	GO through BioPortal		
	Institution	National Center for Biomedical Ontology, Stanford, USA		
Deprecated Physical Locations				
References				

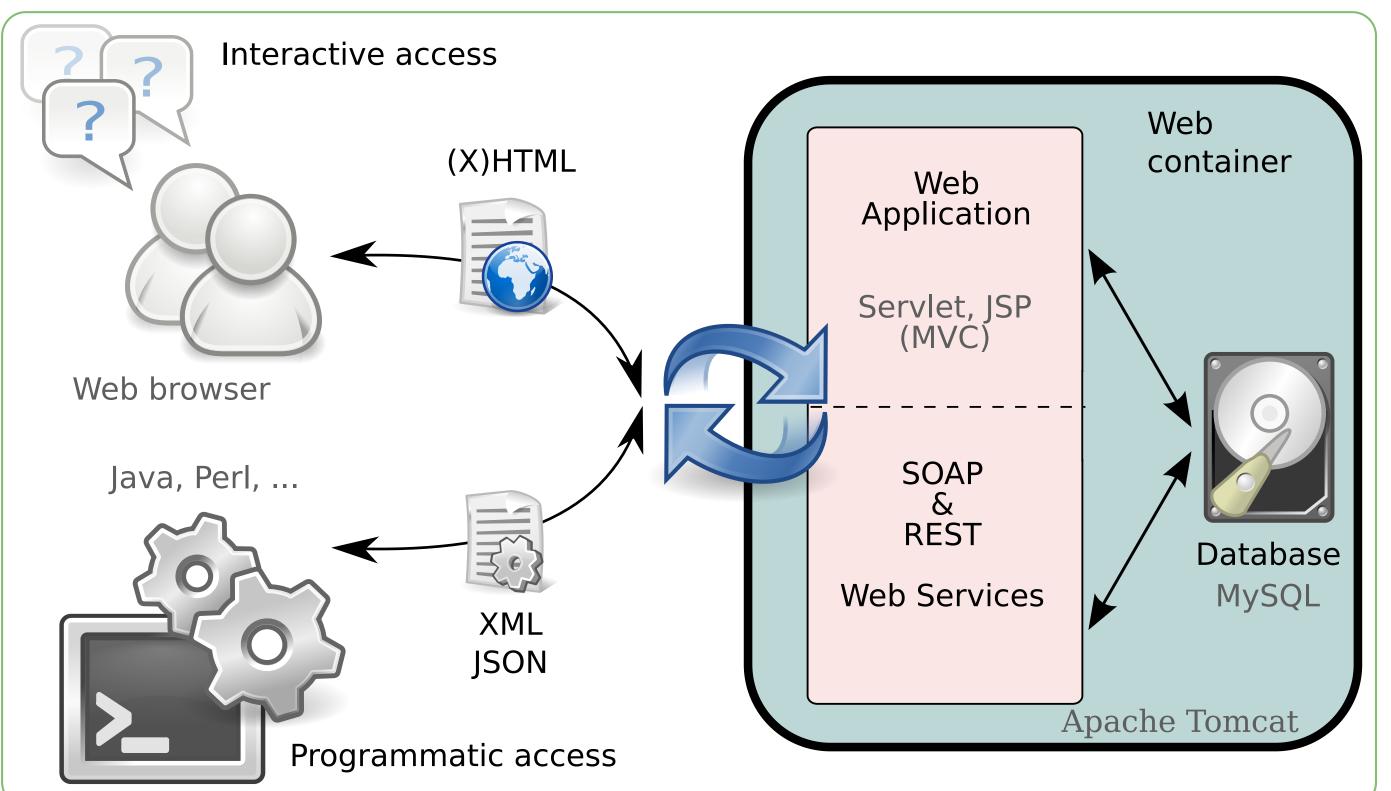
Resources health checks

Health histor

Full record of the health checks performed on this resource



MIRIAM Registry architecture



Summary

- MIRIAM annotations are widely accepted, and are being used and supported by various communities such as BioModels.net [3], SBML, CellML, BioPAX and the Proteomics Standards Initiative (PSI).

- In an effort to enable their use by the Semantic Web and Linked Data communities, we have provided an additional, parallel system of identification that provides resolvable URLs, and furthermore allows access to the MIRIAM Registry information in *Resource Description Framework* (RDF) format.

- MIRIAM URIS facilitate the identification, exchange, and integration of data in the Life Sciences.

References

[1] N. Le Novère, A. Finney, M. Hucka, U. Bhalla, F. Campagne, J. Collado-Vides, E.J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J.L. Snoep, H.D. Spence, and B.L. Wanner. Minimum Information Requested In the Annotation of biochemical Models (MIRIAM). Nature Biotechnology, 23(12):1509–1515, 2005. [2] C. Laibe and N. Le Novère. MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology. BMC Systems Biology, 1: 58, 2007. [3] Le Novère N. BioModels.net, tools and resources to support Computational Systems Biology. Proceedings of the 4th Workshop on Computation of Biochemical Pathways and Genetic Networks. Logos, Berlin, pp. 69-74, 2005.



EMBL - European Bioinformatics Institute Wellcome Trust Genome Campus



MIRIAM Registry and Identifiers.org gratefully acknowledge funding by:

- European Life-science Infrastructure for Biological Information Biotechnology and Biological Sciences Research Council
National Institute of General Medical Sciences - European Molecular Biology Laboratory