MIRIAM Registry and Identifiers.org

Robust and versatile identification in life sciences

Nicolas Le Novère, EMBL-EBI

Annu. Rev. Genomics Hum. Genet. 2001. 2:343–72

A New Approach to Decoding Life:
Systems Biology

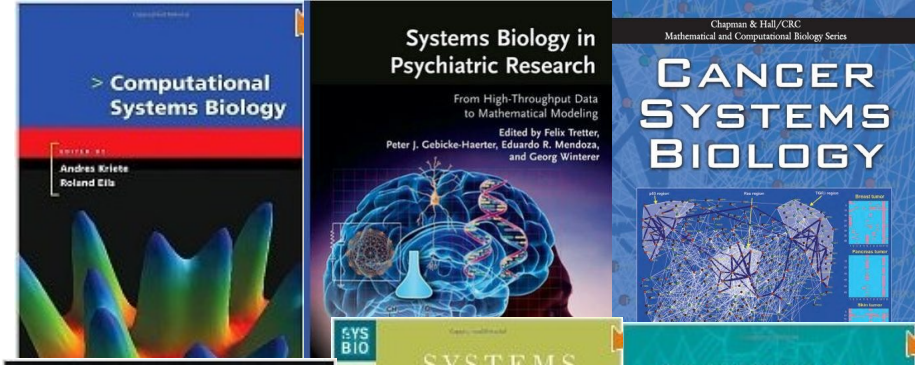
Trey Ideker[1,2], Timothy Galitski[1], and Leroy Hood[1,2,3,4,5]
Institute for Systems Biology[1], Seattle, Washington 98105; Departments of

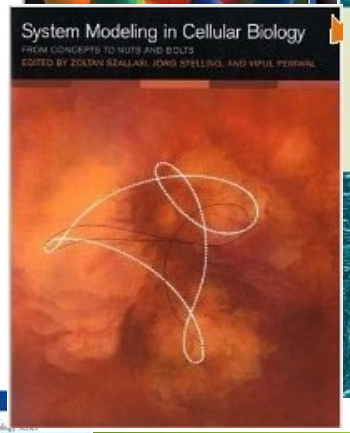
New Generation Computing, 18(2000)199-216
Ohmsha, Ltd. and Springer-Verlag

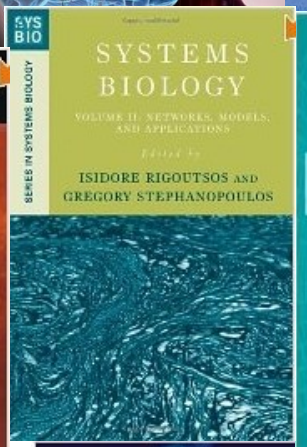
Invited Paper

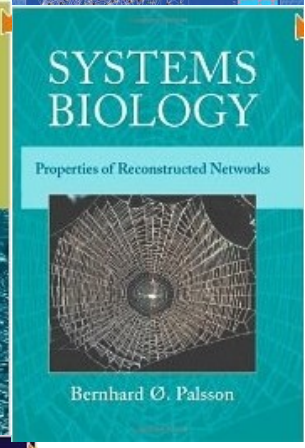
Perspectives on Systems Biology

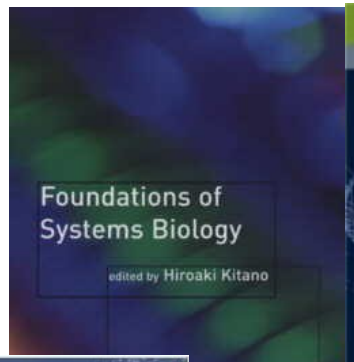
Hiroaki KITANO
Sony Computer Science Laboratories, Inc.

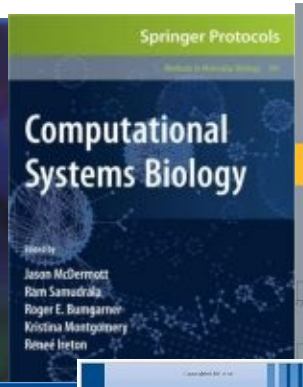


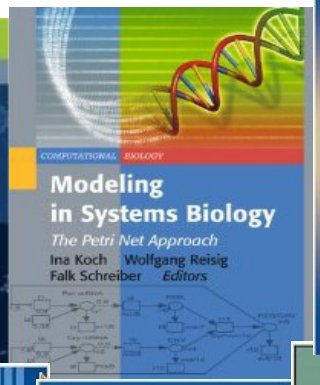


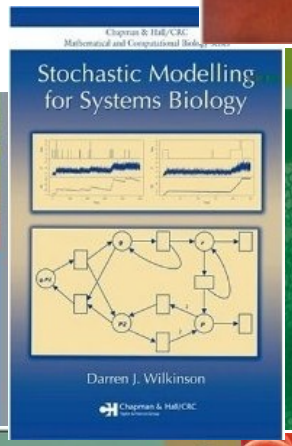


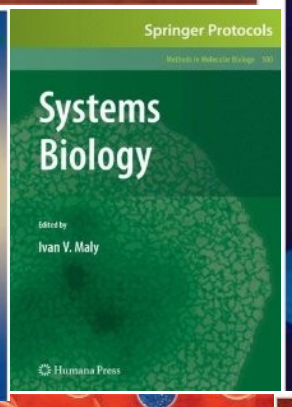


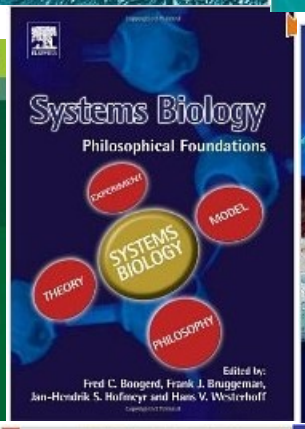


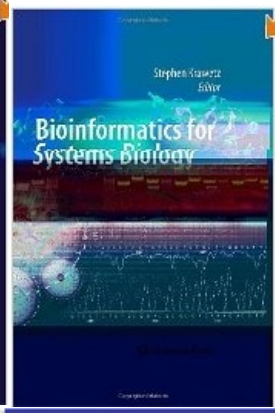


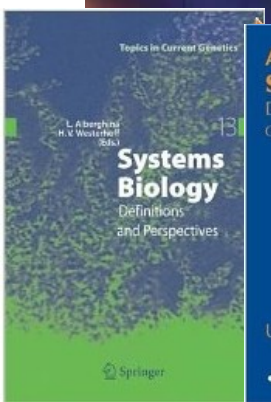


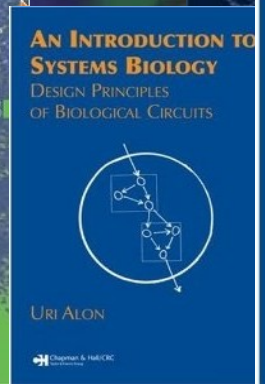


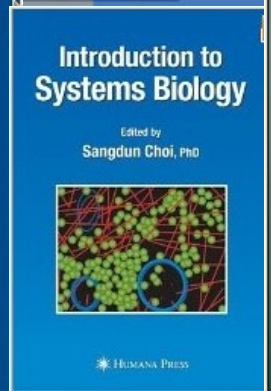


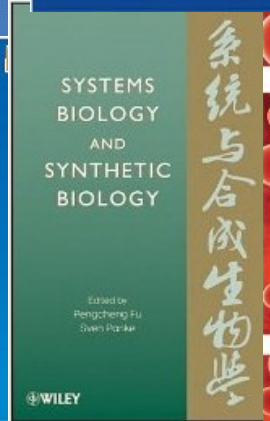


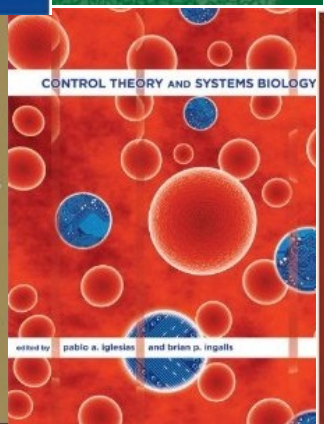


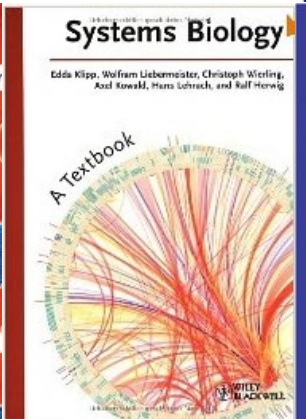


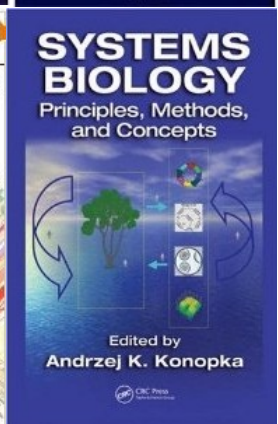






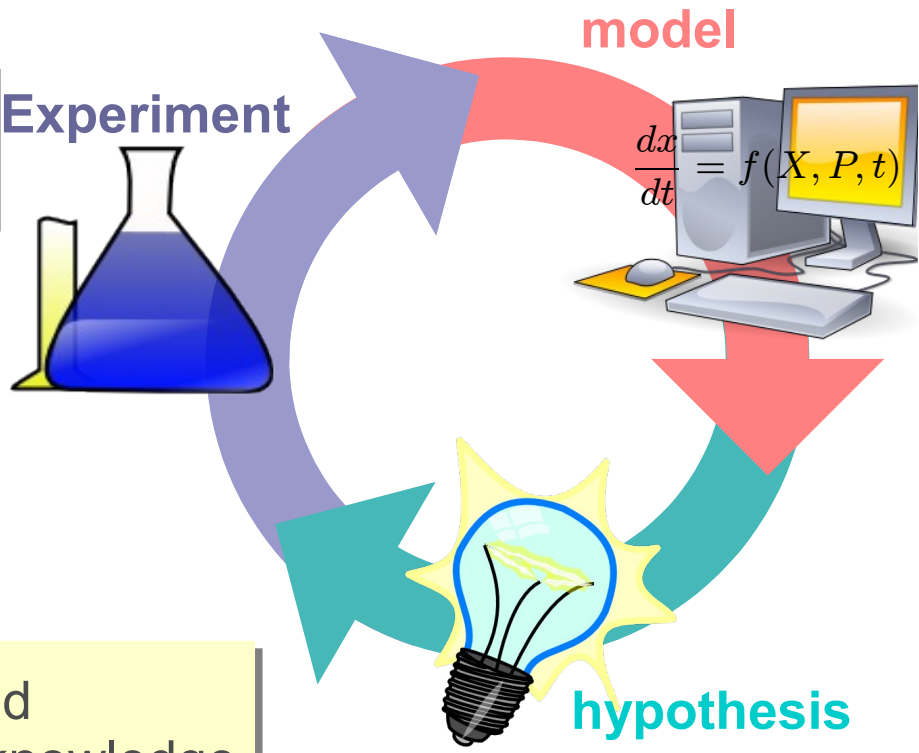


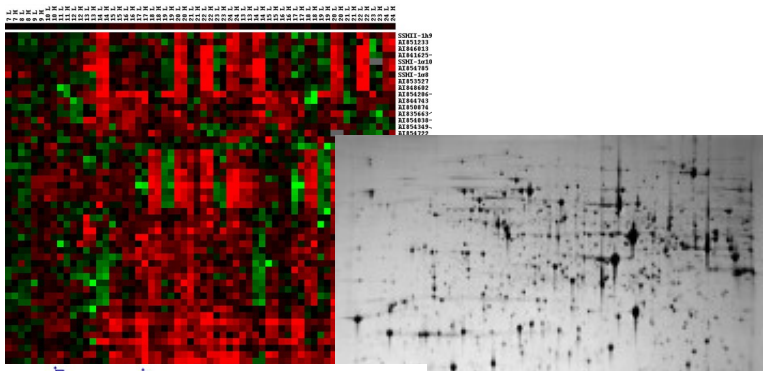


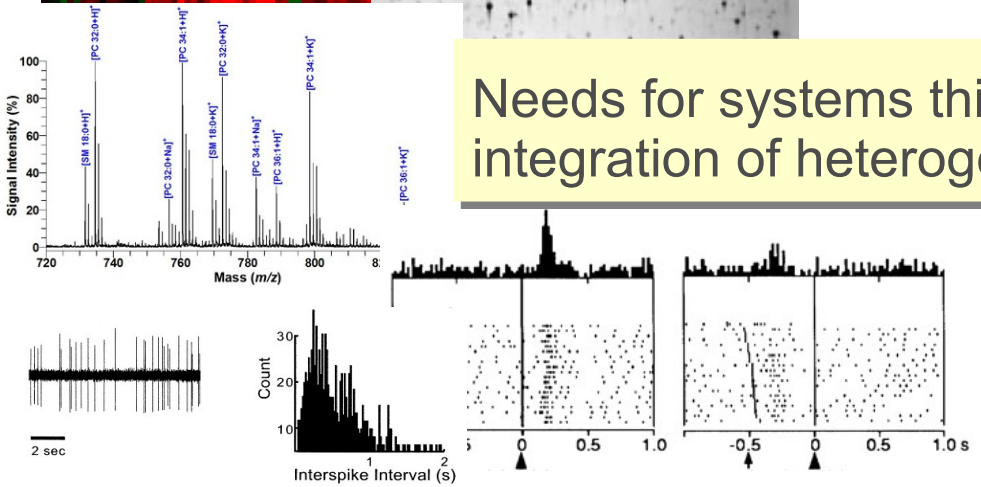


New way of doing biomedical research

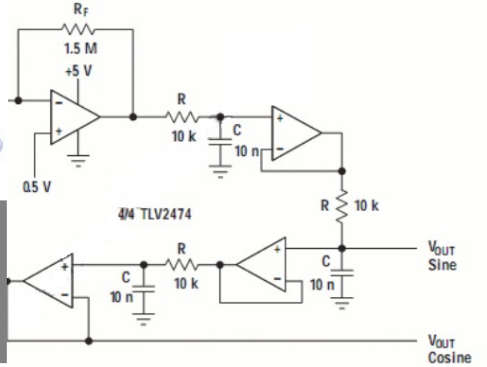
Needs for interplay between models and reality tests







Needs for systems thinking and integration of heterogeneous knowledge



Needs for cooperation and standardisation

Data annotation is essential for integrative biology

- Allows efficient search strategies
- Unambiguously identifies dataset components
  - Improves understanding the structure of the dataset
  - Eases comparison of different datasets
  - Eases the integration of datasets (of the same type)
- Add a semantic layer to the dataset
  - Improve understanding of the biology behind the dataset
  - Allow conversion and reuse of the dataset
  - Ease the integration of different types of datasets

Annotation heavily relies on cross-references

NP_012345

E-MEXP-1712

BIOMD0000000048

PMID:16333295

2018

978-1584885658

ENSG00000139618

WBGene00000001

CHEBI:36927

P62158

10.1038/nbt1156

000057272

REACT_1590

GO:0006915

EBI-2307691

0807.4956v1

Annotation relies on cross-references

E-MEXP-1712

BIOMD0000000048

NP_012345

PMID:16333295

2018

978-1584885658

but...

ENSG00000139618

WBGene00000001

CHEBI:36927

P62158

10.1038/nbt1156

000057272

REACT_1590

GO:0006915

EBI-2307691

0807.4956v1

Limitations of data collection specific identifiers

Limitations of data collection specific identifiers

- **ambiguous**

9606

In UniProt taxonomy
Homo sapiens

In PubMed
Kohlhausen K. *Offentl Gesundheitswes*. 1976 38:424-30.

In Biogrid
Arabidopsis thaliana senescence/dehydration-associated protein-related

In PubChem
3-Fluorotoluene

...

Limitations of data collection specific identifiers

- **ambiguous**

9606

In UniProt taxonomy
Homo sapiens

In PubMed
Kohlhausen K. *Offentl Gesundheitswes*. 1976 38:424-30

In UniBio
Arabidopsis thaliana senescence/dehydration-associated protein-related

In PubChem
3-Fluorotoluene

...

What is "9606"?

Limitations of data collection specific identifiers

- **ambiguous**

9606

In UniProt taxonomy
Homo sapiens

In BOLD taxonomy
Bombycilla cedrorum

In GRIN plant taxonomy
Catha edulis

Limitations of data collection specific identifiers

- **ambiguous**

9606

In UniProt taxonomy
Homo sapiens

In BOLD taxonomy
Bombycilla cedrorum

In GRIN plant taxonomy
Catha edulis

What is "taxonomy:9606"?

Limitations of annotation schemes

- Ambiguous
- Inconsistent

SWISS-PROT:P23222

Swiss-Prot:P23222

UniProt:P23222

UniProtKB:P23222

UniProtKB/Swiss-Prot:P23222

DBREF 1XJ4 A 151 269 UNP P23222

Limitations of annotation schemes

- Ambiguous
- Inconsistent

SWISS-PROT:P23222

Swiss-Prot:P23222

Are those identifying the same thing?

UniProt:P23222

UniProtKB:P23222

UniProtKB/Swiss-Prot:P23222

DBREF 1XJ4 A 151 269 UNP P23222

Limitations of annotation schemes

- **Ambiguous**
- **Inconsistent**
- **Not perennial**

http://www.ebi.ac.uk/citexplore/citationDetails.do?dataSource=MED&externalId=16333295

http://www.ncbi.nlm.nih.gov/pubmed/16333295

http://www.hubmed.org/display.cgi?uids=16333295

http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:16333295]

Limitations of annotation schemes

- **Ambiguous**
- **Inconsistent**
- **Not perennial**

http://www.ebi.ac.uk/citexplore/citationDetails.do?dataSource=MED&externalId=16333295

http://www.ncbi.nlm.nih.gov/pubmed/16333295

http://www.hubmed.org/display.cgi?uids=16333295

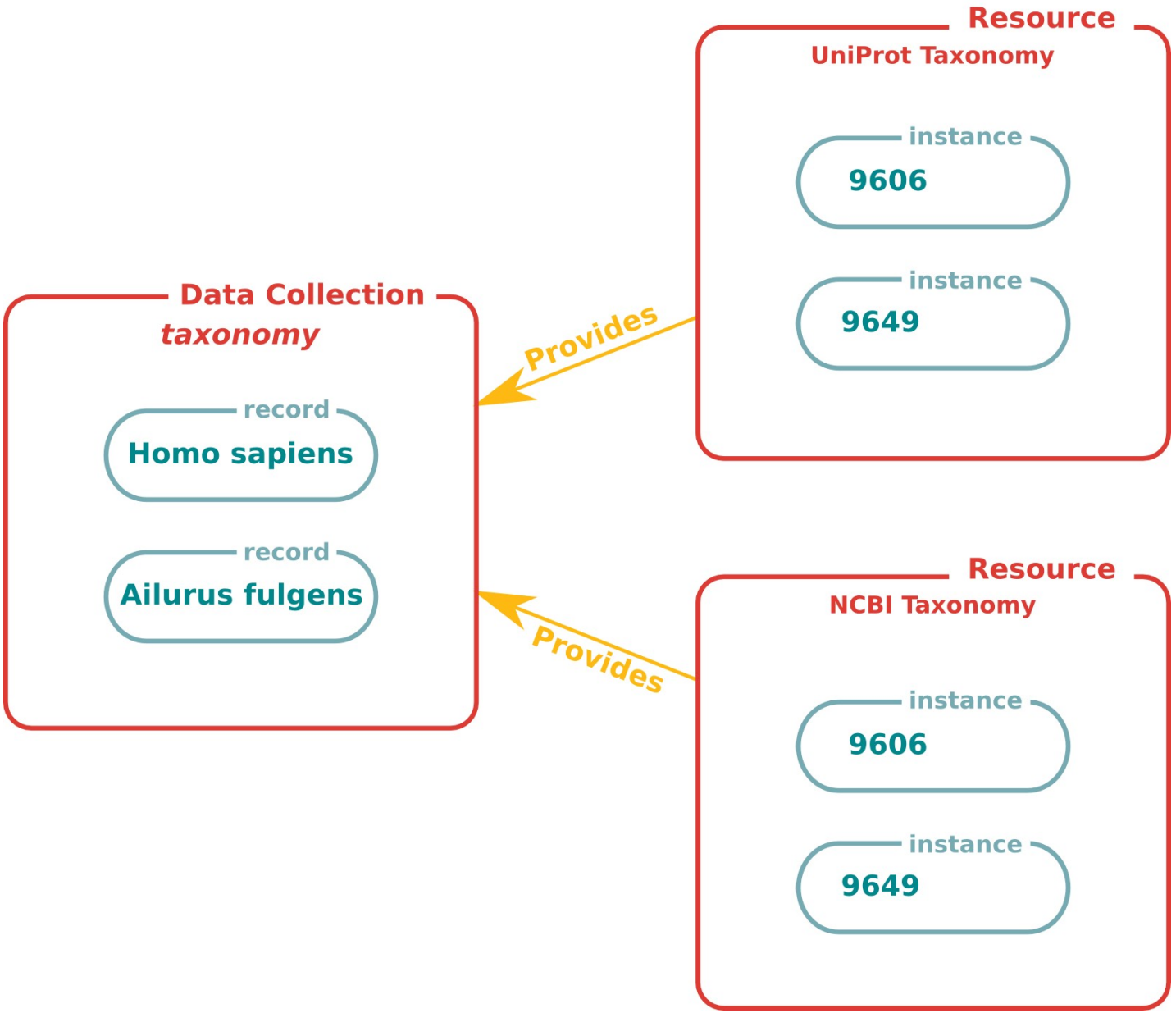
http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:16333295]

Will those still be valid in three months?

Other potential limitations

- Ambiguous
- Inconsistent
- Not perennial
- Location dependent
- Not standard compliant
- Non resolvable
- Non free to use

Naming conventions



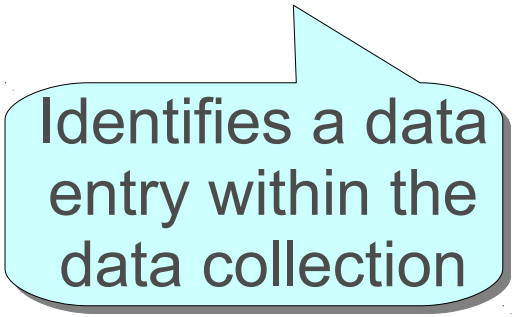
URIs, the standard way to encode identifiers

A **Uniform Resource Identifier** (URI) is a string of characters used to **identify** a source of information.

```
http://www.ebi.ac.uk/
http://en.wikipedia.org/wiki/Uniform_Resource_Identifier
https://www.ebi.ac.uk/chembldb/
ftp://public.ftp-servers.example.com/mydirectory/myfile.txt
ftp://ftp.uniprot.org/pub/databases/uniprot/uniref/
urn:ietf:rfc:2396
file:///home/username/presentation.pdf
```

Towards globally unique identifiers

Record identifier



provided by the data collection

unique within the data collection

format defined by the data collection

Towards globally unique identifiers

Collection namespace

Identifies a type of data

from a shared list of *namespaces*

Record identifier

Identifies a data entry within the data collection

provided by the data collection

unique within the data collection

format defined by the data collection

Towards globally unique identifiers

URI scheme

Identifies the type of identifier

Collection namespace

Identifies a type of data

from a shared list of *namespaces*

Record identifier

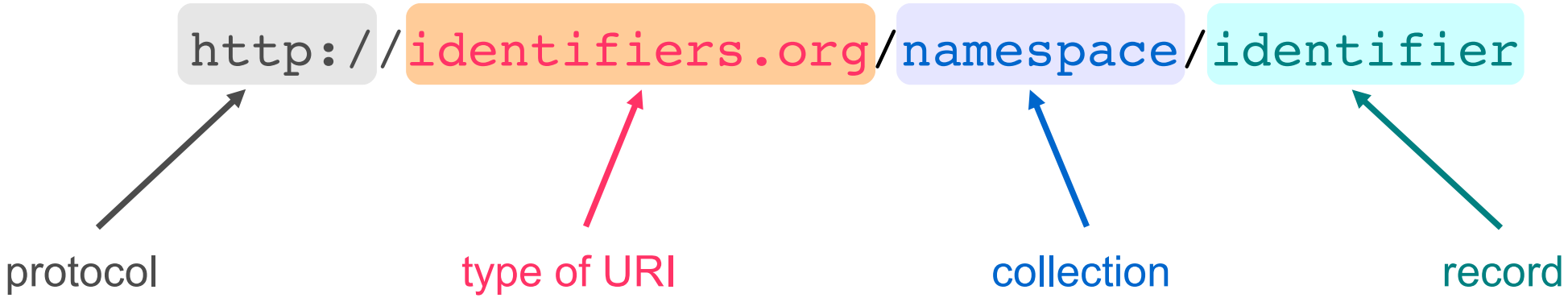
Identifies a data entry within the data collection

provided by the data collection

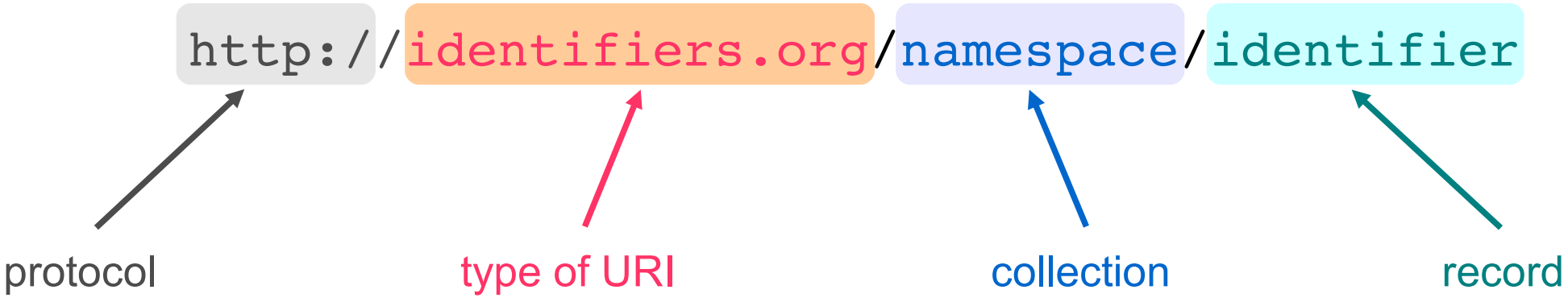
unique within the data collection

format defined by the data collection

Enters identifiers.org (aka new MIRIAM URIs)



Enters identifiers.org (aka new MIRIAM URIs)



http://identifiers.org/pubmed/22140103

http://identifiers.org/ec-code/1.1.1.1

http://identifiers.org/obo.go/GO:0000186

identifiers.org

Google Wikipedia PubMed Biomodels compneur calendar MyPage SBO Miriam TinyURL! PostCiteUL

identifiers.org

Home News Help Examples Registry About

Welcome to Identifiers.org!

Identifiers.org is a system providing resolvable persistent URIs used to identify data for the scientific community, with a current focus on the Life Sciences domain. The provision of a resolvable identifiers (URLs) fits well with the Semantic Web vision, and the Linked Data initiative.

Links

- About
- News
- Help
- Examples URIs
- MIRIAM Registry

http://identifiers.org/

Board of trustees

Identifiers.org is a community project which activities are overseen by the following board of trustees:

- Michel Dumontier (Carleton University, Ottawa, Canada - Bio2RDF, W3C HCLS)
- Michael Galperin (NCBI, U
- Pascale Gaudet (Swiss Ins
- Lee Harland (Connected D
- Michael Hucka (California
- Toshiaki Katayama (Univers
- Nicolas Le Novère (EMBL-EBI, Hinxton, UK - BioModels Database)
- Philippe Rocca-Serra (Oxford University, Oxford, UK - BioSharing)
- Mark Wilkinson (St. Paul's Hospital/UBC Vancouver, Canada - LSRN, SADI)

Juty N., Le Novère N., Laibe C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification *Nucleic Acids Research* (2012) 40: D580-D586

Contact

If you have any queries or experience any issues with this service, please contact: **biomodels-net-support [AT] lists.sf.net**

zotero

Welcome to Identifiers.org!

Identifiers.org is a system providing resolvable persistent URIs used to identify data for the scientific community, with a current focus on the Life Sciences domain. The provision of a resolvable identifiers (URLs) fits well with the Semantic Web vision, and the Linked Data initiative.

Links

- About
- News
- Help
- Examples URIs
- MIRIAM Registry

Board of trustees

Identifiers.org is a community project which activities are overseen by the following board of trustees:

- Michel Dumontier (Carleton University, Ottawa, Canada - Bio2RDF, W3C HCLS)
- Michael Galperin (NCBI, USA - NAR Database issue)
- Pascale Gaudet (Swiss Institute of Bioinformatics, Geneva, Switzerland - BioDBCore)
- Lee Harland (Connected Discovery, UK - Open PHACTS)
- Michael Hucka (California Institute of Technology, Pasadena, USA - SBML)
- Toshiaki Katayama (University of Tokyo, Japan - BioRuby, KEGG)
- Nicolas Le Novère (EMBL-EBI, Hinxton, UK - BioModels Database)
- Philippe Rocca-Serra (Oxford University, Oxford, UK - BioSharing)
- Mark Wilkinson (St. Paul's Hospital/UBC Vancouver, Canada - LSRN, SADI)

Contact

If you have any queries or experience any issues with this service, please contact: **biomodels-net-support [AT] lists.sf.net**

- Browse
- Search
- Tags
- Query services
- Submit new
- Export
- Curator Sign in

- Web Services
- Documents
  - MIRIAM Guidelines
  - FAQ
  - Documentation
  - Who's using MIRIAM?
  - Identification systems
  - News
  - BioModels.net Qualifiers

- MIRIAM on SourceForge

- Support
- Contact

MIRIAM Registry

MIRIAM Registry are a set of online services created in support of MIRIAM, a set of guidelines for the annotation and curation of computational models.

The core of *MIRIAM Registry* is a catalogue of data collections (corresponding to controlled vocabularies or databases), their URIs and the corresponding physical URLs or resources. Access to this data is made available via exports (XML) and Web Services (SOAP).

MIRIAM Registry is developed and maintained under the BioModels.net initiative, and are free for use by all.

Quick links

Browse

by data collection name
by tags

Web Services

services available
usage of the services
online demonstration

Search

generic search

Exports

XML

http://www.ebi.ac.uk/miriam/
http://identifiers.org/registry

Registry

MIRIAM Registry is compo[illegible]

Database

The core of the system is a MySQL database. It allows us to store the data collections (which can be controlled vocabularies or databases), their URIs and the corresponding physical URLs, and other details such as documentation and resource identifier patterns.

Each entry contains a diverse set of details about the data collection: official name and synonyms, root URI, pattern of identifiers, documentation, etc. Moreover, each data collection can be associated with several resources (or physical locations).

Web Services

List of data collections and their namespaces

Browse data collections: ***recently updated***

Recently updated | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Name	Namespace	Definition
EDAM Ontology	edam	EDAM is an ontology of general bioinformatics concepts, including topics, data types, formats, identifiers and operations. EDAM provides a controlled vocabulary for the description, in semantic terms, of things such as: web services (e.g. WSDL files), applications, tool collections and packages, work-benches and workflow software, databases and ontologies, XSD data schema and data objects, data syntax and file formats, web portals and pages, resource catalogues and documents (such as scientific publications).
LIPID MAPS	lipidmaps	The LIPID MAPS Lipid Classification System is comprised of eight lipid categories, each with its own subclassification hierarchy. All lipids in the LIPID MAPS Structure Database (LMSD) have been classified using this system and have been assigned LIPID MAPS ID's which reflects their position in the classification hierarchy.
Entrez Gene	entrez.gene	Entrez Gene is the NCBI's database for gene-specific information, focusing on completely sequenced genomes, those with an active research community to contribute gene-specific information, or those that are scheduled for intense sequence analysis.
CATH superfamily	cath.superfamily	The CATH database is a hierarchical domain classification of protein structures in the Protein Data Bank. Protein structures are classified using a combination of automated and manual procedures. There are four major levels in this hierarchy; Class (secondary structure classification, e.g. mostly alpha), Architecture (classification based on overall shape), Topology (fold family) and Homologous superfamily (protein domains which are thought to share a common ancestor). This colelction is concerned with superfamily classification.
Taxonomy	taxonomy	The taxonomy contains the relationships between all living forms for which nucleic acid or protein sequence have been determined.
Ontology of Physics for Biology	opb	The OPB is a reference ontology of classical physics as applied to the dynamics of biological systems. It is designed to encompass the multiple structural scales (multiscale atoms to organisms) and multiple physical domains (multidomain fluid dynamics, chemical kinetics, particle diffusion, etc.) that are encountered in the study and analysis of biological organisms.
Cell Cycle Ontology	cco	The Cell Cycle Ontology is an application ontology that captures and integrates detailed knowledge on the cell cycle process.
Human Disease Ontology	obo.do	The Disease Ontology has been developed as a standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts.
eggNOG	eggnog	eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) is a database of orthologous groups of genes. The orthologous groups are annotated with functional description lines (derived by identifying a common denominator for the genes based on their various annotations), with functional categories (i.e derived from the original COG/KOG categories).
Gramene genes	gramene.gene	Gramene is a comparative genome mapping database for grasses and crop plants. It combines a semi-automatically generated database of cereal genomic and expressed sequence tag sequences, genetic maps, map relations, quantitative trait loci (QTL), and publications, with a curated database of mutants (genes and alleles), molecular markers, and proteins. This datatype refers to genes in Gramene.
iRefWeb	irefweb	iRefWeb is an interface to a relational database containing the latest build of the interaction Reference Index (iRefIndex) which integrates protein interaction data from ten different interaction databases: BioGRID, BIND, CORUM, DIP, HPRD, INTACT, MINT, MPPI, MPACT and OPHID. In addition, iRefWeb associates interactions with the PubMed record from which they are derived.

Data collection: *Enzyme Nomenclature*

Overview | Categories | Miscellaneous

Overview of the data collection

Name		
Identifier		MIR:00000004
Name		Enzyme Nomenclature
Synonyms		Enzyme Classification
		EC code
		EC
Information		
Definition		The Enzyme Classification contains the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzyme-catalysed reactions.
Identifier pattern		^\d+\.-\.-\.-|\d+\.\d+\.-\.-|\d+\.\d+\.\d+\.-|\d+\.\d+\.\d+\.(n)?\d+$
URIs		
Namespace		ec-code
Root URL		http://identifiers.org/ec-code/
Root URN		urn:miriam:ec-code:
Physical Locations		
Resource MIR:00100002	**Access URL**	http://www.genome.jp/dbget-bin/www_bget?ec:**$id** [Example: 1.1.1.1]
	Website	http://www.genome.jp/dbget-bin/www_bfind?enzyme
	Description	KEGG Ligand Database for Enzyme Nomenclature
	Institution	Kyoto University Bioinformatics Center, Japan
Resource MIR:00100001	**Access URL**	http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=**$id** [Example: 1.1.1.1]
	Website	http://www.ebi.ac.uk/intenz/
	Description	IntEnZ (Integrated relational Enzyme database)
	Institution	European Bioinformatics Institute, United Kingdom
Resource MIR:00100003	**Access URL**	http://enzyme.expasy.org/EC/**$id** [Example: 1.1.1.1]
	Website	http://enzyme.expasy.org/
	Description	Enzyme nomenclature database, ExPASy (Expert Protein Analysis System)
	Institution	Swiss Institute of Bioinformatics, Switzerland
References		
http://www.chem.qmul.ac.uk/iubmb/enzyme/		
http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:10812475]		

Go back to the list of data collections

Suggest modifications to this data collection

Data collection: *Enzyme Nomenclature*

Overview | Categories | Miscellaneous

Overview of the data collection

Name	
Identifier	MIR:00000004
Name	Enzyme Nomenclature
Synonyms	Enzyme Classification
	EC code
	EC

Resources providing the collection

Information	
Definition	The Enzyme Classification contains the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzyme-catalysed reactions.
Identifier pattern	^\d+\.-\.-\.-|\d+\.\d+\.-\.-|\d+\.\d+\.\d+\.-|\d+\.\d+\.\d+\.(n)?\d+$

URIs	
Namespace	ec-code
Root URL	http://identifiers.org/ec-code/
Root URN	urn:miriam:ec-code:

Physical Locations		
Resource MIR:00100002	**Access URL**	http://www.genome.jp/dbget-bin/www_bget?ec:**$id** [Example: 1.1.1.1]
	Website	http://www.genome.jp/dbget-bin/www_bfind?enzyme
	Description	KEGG Ligand Database for Enzyme Nomenclature
	Institution	Kyoto University Bioinformatics Center, Japan
Resource MIR:00100001	**Access URL**	http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=**$id** [Example: 1.1.1.1]
	Website	http://www.ebi.ac.uk/intenz/
	Description	IntEnZ (Integrated relational Enzyme database)
	Institution	European Bioinformatics Institute, United Kingdom
Resource MIR:00100003	**Access URL**	http://enzyme.expasy.org/EC/**$id** [Example: 1.1.1.1]
	Website	http://enzyme.expasy.org/
	Description	Enzyme nomenclature database, ExPASy (Expert Protein Analysis System)
	Institution	Swiss Institute of Bioinformatics, Switzerland

References

http://www.chem.qmul.ac.uk/iubmb/enzyme/

http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:10812475]

Go back to the list of data collections

Suggest modifications to this data collection

Data collection: *Enzyme Nomenclature*

Overview | Categories | Miscellaneous

Overview of the data collection

Name		
Identifier		MIR:00000004
Name		Enzyme Nomenclature
Synonyms		Enzyme Classification
		EC code
		EC
Information		
Definition		The Enzyme Classification contains the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzyme-catalysed reactions.
Identifier pattern		^\d+\.-\.-\.-|\d+\.\d+\.-\.-|\d+\.\d+\.\d+\.-|\d+\.\d+\.\d+\.(n)?\d+$
URIs		
Namespace		ec-code
Root URL		http://identifiers.org/ec-code/
Root URN		urn:miriam:ec-code:
Physical Locations		
Resource	**Access URL**	http://www.genome.jp/dbget-bin/www_bget?ec:**$id** [Example: 1.1.1.1]
	Website	http://www.genome.jp/dbget-bin/www_bfind?enzyme
		…nclature
		…an
		…archEC&ec=**$id** [Example: 1.1.1.1]
MIR:00100001	**Description**	IntEnZ (Integrated relational Enzyme database)
	Institution	European Bioinformatics Institute, United Kingdom
Resource MIR:00100003	**Access URL**	http://enzyme.expasy.org/EC/**$id** [Example: 1.1.1.1]
	Website	http://enzyme.expasy.org/
	Description	Enzyme nomenclature database, ExPASy (Expert Protein Analysis System)
	Institution	Swiss Institute of Bioinformatics, Switzerland
References		
http://www.chem.qmul.ac.uk/iubmb/enzyme/		
http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:10812475]		

Suggest modifications to this data collection

Go back to the list of data collections

Suggest modifications to this data collection

- Browse
- Search
- Tags
- Query services
- Submit new
- Export
- Curator Sign in

- Web Services
- Documents
  - MIRIAM Guidelines
  - FAQ
  - Documentation
  - Who's using MIRIAM?
  - Identification systems
  - News
  - BioModels.net Qualifiers

- MIRIAM on SourceForge

- Support
- Contact

MIRIAM Registry

MIRIAM Registry are a set of online services created in support of MIRIAM, a set of guidelines for the annotation and curation of computational models.

The core of *MIRIAM Registry* is a catalogue of data collections (corresponding to controlled vocabularies or databases), their URIs and the corresponding physical URLs or resources. Access to this data is made available via exports (XML) and Web Services (SOAP).

MIRIAM Registry is developed and maintained under the BioModels.net initiative, and are free for use by all.

Quick links

Submit new

Browse

by data collection name
by tags

Web Services

services available
usage of the services
online demonstration

Search

generic search

Exports

XML

Registry

MIRIAM Registry is composed of four components: a database, some Web Services, a Java library and this web application.

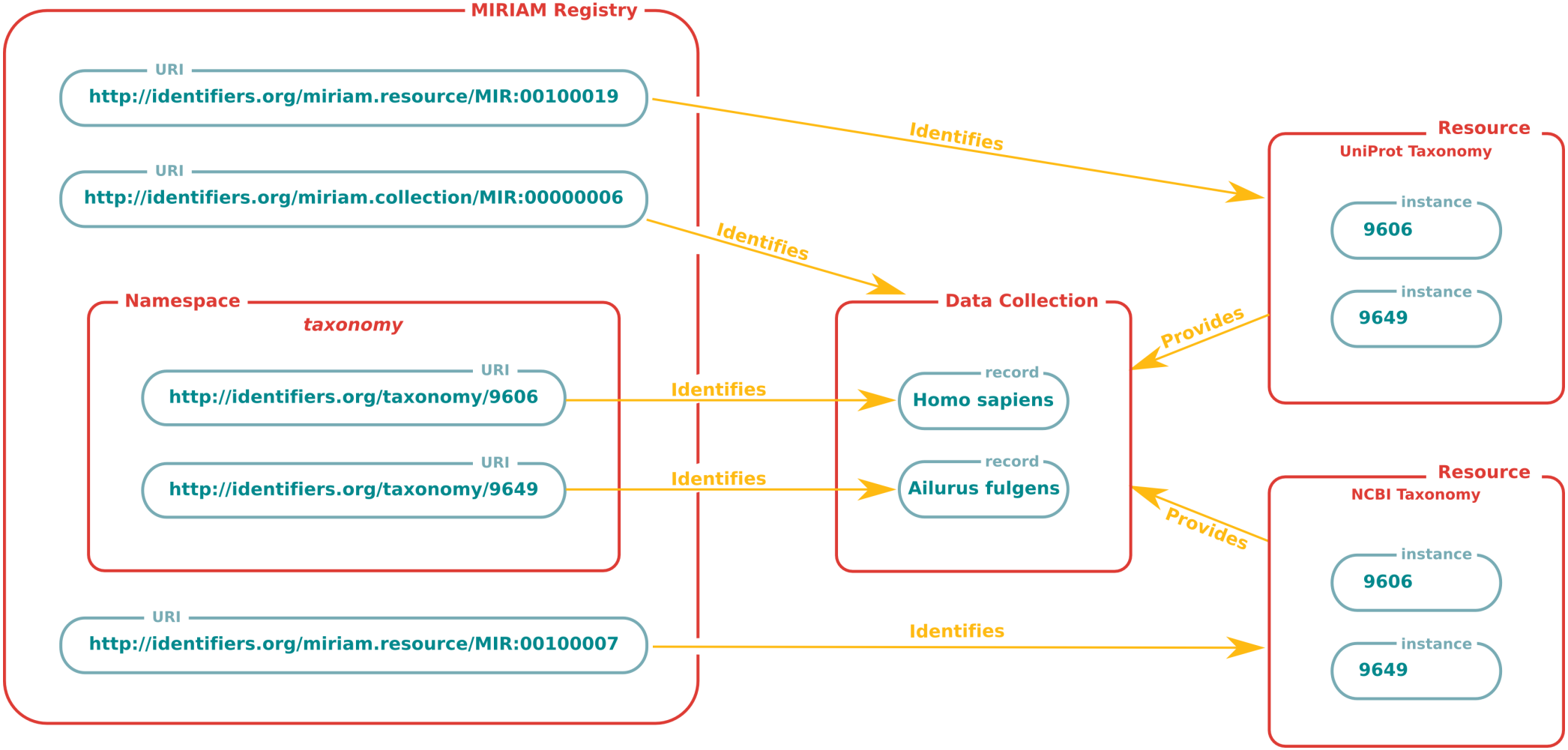
Database

The core of the system is a MySQL database. It allows us to store the data collections (which can be controlled vocabularies or databases), their URIs and the corresponding physical URLs, and other details such as documentation and resource identifier patterns.

Each entry contains a diverse set of details about the data collection: official name and synonyms, root URI, pattern of identifiers, documentation, etc. Moreover, each data collection can be associated with several resources (or physical locations).

Web Services

Everything has an identifier



How does that work?

To access a record:

`http://identifiers.org/ec-code/1.1.1.1`

To access a collection:

`http://identifiers.org/ec-code/`

(nb: synonymous to:
`http://identifiers.org/miriam.collection/MIR:0000004`)

http://identifiers.org/ec-code/1.1.1.1

3 physical locations (or resources) are available for accessing *1.1.1.1* (from Enzyme Nomenclature):

KEGG Ligand Database for Enzyme Nomenclature
Kyoto University Bioinformatics Center

Japan

(Uptime: 100%)

Enzyme nomenclature database, ExPASy (Expert Protein

Access to '1.1.1.1' via this resource (MIR:00100003)

Switzerland

(Uptime: 99%)

IntEnZ (Integrated relational Enzyme database)
European Bioinformatics Institute

United Kingdom

(Uptime: 100%)

http://enzyme.expasy.org/EC/1.1.1.1

ExPASy Bioinformatics Resource Portal

ENZYME

Home | Contact

Due to maintenance work, this ExPASy service will be unavailable from Sunday January 29th, 2012 to Wednesday February 1st, 2012.

ENZYME entry: EC 1.1.1.1

Accepted Name
Alcohol dehydrogenase.
Alternative Name(s)
Aldehyde reductase.
Reaction catalysed
An alcohol + NAD(+) <=> an aldehyde or ketone + NADH
Cofactor(s)
Zinc or iron.
Comment(s)

- Acts on primary or secondary alcohols or hemiacetals.
- The animal, but not the yeast, enzyme acts also on cyclic secondary alcohols.

Cross-references	
PROSITE	PDOC00058 ; PDOC00059 ; PDOC00060
BRENDA	1.1.1.1
EC2PDB	1.1.1.1
ExplorEnz	1.1.1.1
PRIAM enzyme-specific profiles	1.1.1.1
KEGG Ligand Database for Enzyme Nomenclature	1.1.1.1

More magic

```
http://identifiers.org/namespace/identifier?argument=value
```

More magic

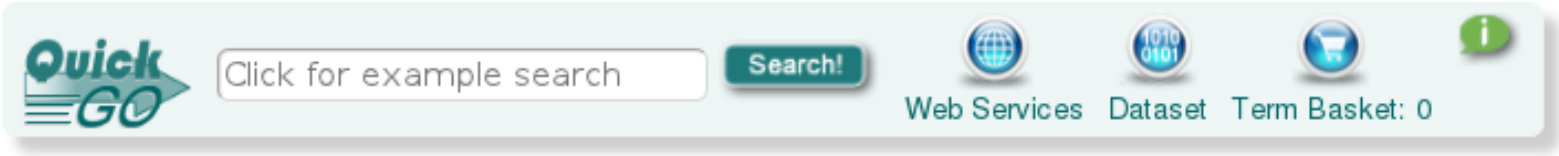
```
http://identifiers.org/namespace/identifier?argument=value
```

```
http://identifiers.org/obo.go/GO:0000186?format=rdfxml

http://identifiers.org/obo.go/GO:0000186?resource=MIR:00100012

http://identifiers.org/obo.go/GO:0000186?profile=most_reliable

http://identifiers.org/obo.go/GO:0000186?profile=MyProfile
```

```
- <rdf:RDF>
  - <rdf:Description rdf:about="http://identifiers.org/obo.go/GO:0000186">
      <!-- human readable description -->
    - <dcterms:title xml:lang="en-GB">
        Entity GO:0000186 provided by the data collection Gene Ontology (MIR:00000022).
      </dcterms:title>
    - <!--
        off
      -->
      <dcterms:URI>http://identifiers.org/obo.go/GO:0000186</dcterms:URI>
    - <!--
        identifier (as created and used by the data provider
      -->
      <dcterms:identifier>GO:0000186</dcterms:identifier>
    - <sio:SIO_000671>
      - <edam:EDAM_0002091>
          <sio:SIO_000300>GO:0000186</sio:SIO_000300>
        </edam:EDAM_0002091>
      </sio:SIO_000671>
      <!-- data collection -->
      <dcterms:source rdf:resource="http://identifiers.org/MIR:00000022"/>
      <!-- physical locations (resources) -->
    - <rdfs:seeAlso>
      - <rdf:Description rdf:about="http://www.informatics.jax.org/searches/GO.cgi?id=GO:0000186">
          <dcterms:format>application/xhtml+xml</dcterms:format>
          <dcterms:publisher rdf:resource="MIR:00100015"/>
        </rdf:Description>
      </rdfs:seeAlso>
    - <rdfs:seeAlso>
      - <rdf:Description rdf:about="http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0000186">
          <dcterms:format>application/xhtml+xml</dcterms:format>
          <dcterms:publisher rdf:resource="MIR:00100012"/>
        </rdf:Description>
```

http://identifiers.org/obo.go/GO:0000186?format=rdfxml

http://identifiers.org/obo.go/GO:0000186?resource=MIR:00100012



Term Information | Ancestor Chart | Child Terms | Protein Annotation | Co-occurring Terms | Change Log

ID	GO:0000186
Name	activation of MAPKK activity
Ontology	Biological Process
Definition	The initiation of the activity of the inactive enzyme MAP kinase kinase by phosphorylation by a MAPKKK.
Comment	
Secondary IDs	GO:0007255
GONUTS	GO:0000186 Wiki Page

GO Discussions | Synonyms | Taxon Constraints | Cross-references

The GO editorial group is intending to improve the information provided in this area of the GO.

If you would like to be involved in discussions regarding this development activity, please email the GO Consortium.

Please note that it is still appropriate to use this term for curation or analysis purposes.

Ontology Development Project


MGI
MGI has job openings for biologists and software engineers.
Keywords, Symbols, or IDs
Quick Search
About Help FAQ
Home Genes Phenotypes Expression Recombinases Function Pathways Strains / SNPs Orthology Tumors
Search Download More Resources Submit Data Find Mice (IMSR) Analysis Tools Contact Us
Gene Ontology Browser
Term Detail
GO term: activation of MAPKK activity
Synonym: activation of MAP kinase kinase activity
Synonym: activation of MAP/ERK kinase kinase
Synonym: activation of MAPKK activity during sporulation
Synonym: positive regulation of MAPKK activity
GO id: GO:0000186
Definition: The initiation of the activity of the inactive enzyme MAP kinase kinase by phosphorylation by a MAPKKK.
Number of paths to term: 111
denotes an 'is-a' relationship
denotes a 'part-of' relationship
denotes a 'regulates' relationship
denotes a 'positively-regulates' relationship
denotes a 'negatively-regulates' relationship
Gene_Ontology
biological_process
http://identifiers.org/obo.go/GO:0000186?profile=most_reliable
regulation of cellular metabolic process
regulation of cellular protein metabolic process
regulation of protein modification process
regulation of protein phosphorylation
regulation of protein kinase activity

What does using Identifiers.org bring

Flexible access to multiple resources serving the dataset

What does using Identifiers.org bring

Flexible access to multiple resources serving the dataset

No need to continuously maintain your own identifiers-URLs mapping

Private Vs public cross-references: hard-coding

```
<annotation>
    <ec-code>1.1.1.1</ec-code>
</annotation>
```

EC table

entry
1.1.1.1

Need to change the
data-model to add sources
=> users must redevelop software

```
<annotation>
    <ec-code>1.1.1.1</ec-code>
    <go>GO:0004022</go>
</annotation>
```

EC table

entry
1.1.1.1

GO table

entry
GO:0004022

Private Vs public cross-references: private list

```
<annotation>
    <dbXref
        database="ec-code"
        accession="1.1.1.1"/>
</annotation>
```

Xref table

database	accession
ec-code	1.1.1.1

Need to list the resources
and share the list with users

```
<annotation>
    <dbXref
        database="ec-code"
        Accession="1.1.1.1"/>
    <dbXref
        database="go"
        Accession="GO:0004022"/>
</annotation>
```

Xref table

database	accession
ec-code	1.1.1.1
go	GO:0004022

Private Vs public cross-references: public URIs

Xref table

URI
http://identifiers.org/ec-code/1.1.1.1

```
<annotation>
     <dbXref
URI="http://identifiers.org/ec-code/1.1.1.1"/>
</annotation>
```

Xref table

URI
http://identifiers.org/ec-code/1.1.1.1
http://identifiers.org/obo.go/GO:0004022

```
<annotation>
     <dbXref
URI="http://identifiers.org/ec-code/1.1.1.1"/>
     <dbXref
URI="http://identifiers.org/obo.go/GO:0004022"/>
</annotation>
```

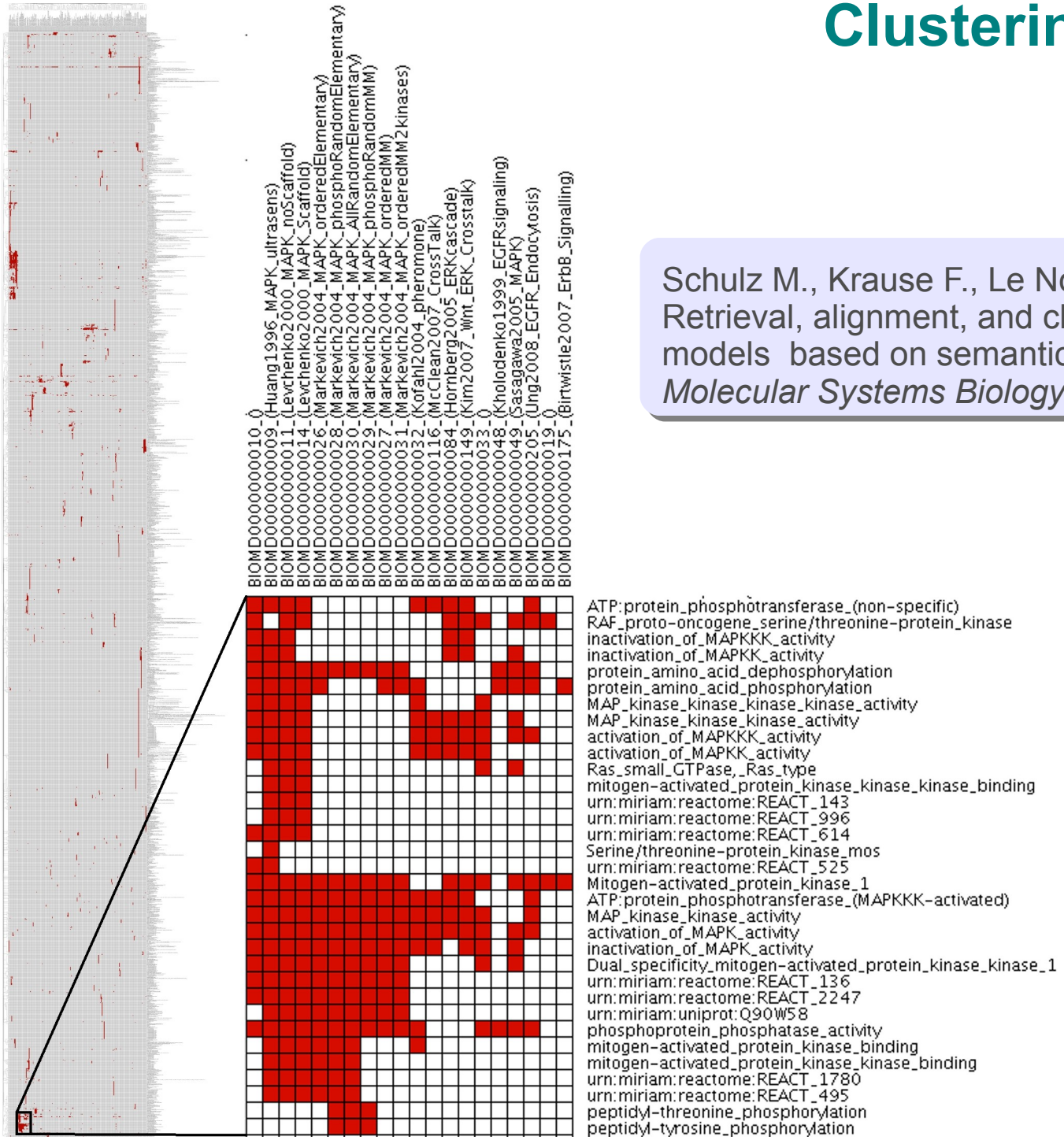
What does using Identifiers.org bring

Flexible access to multiple resources serving the dataset

No need to continuously maintain your own identifiers-URLs mapping

Using shared unique identifiers facilitate data comparison and integration

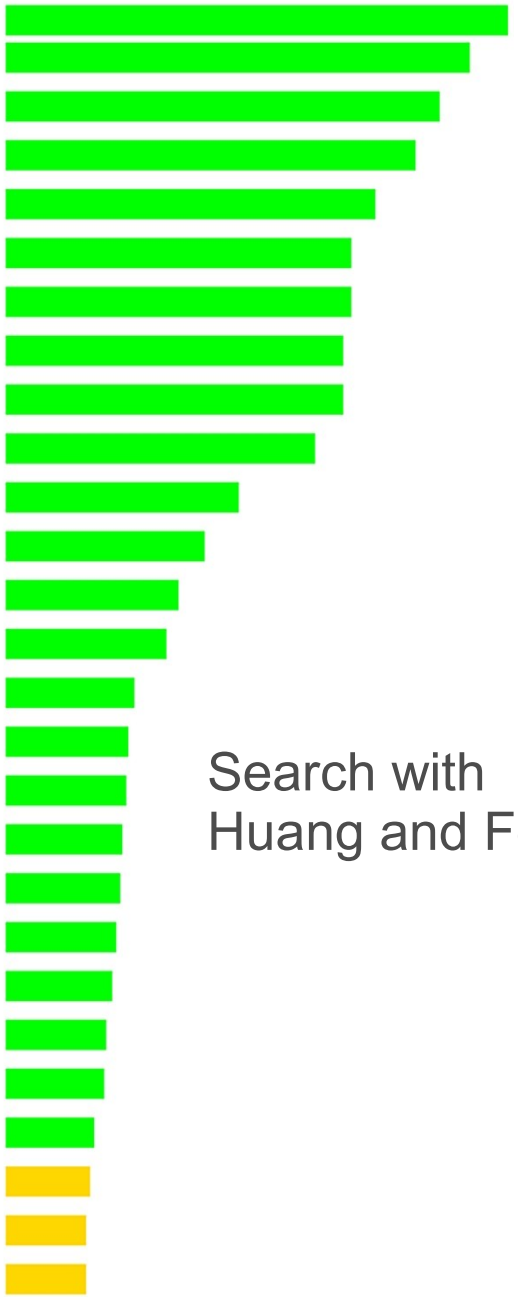
Clustering models (and data) based on metadata



Schulz M., Krause F., Le Novère N., Klipp E., Liebermeister W.
Retrieval, alignment, and clustering of computational models based on semantic annotations
Molecular Systems Biology (2011), 7: 512

Ranking and retrieval of models

Model	BioModel	Score	p-val num
Huang1996_MAPK_ultrasens	BIOMD0000000009	1.000	< 1/1000
Levchenko2000_MAPK_noScaffold	BIOMD0000000011	0.925	< 1/1000
Levchenko2000_MAPK_Scaffold	BIOMD0000000014	0.865	< 1/1000
Kholodenko2000_MAPK_feedback	BIOMD0000000010	0.816	< 1/1000
Markevich2004_MAPK_orderedElementary	BIOMD0000000026	0.737	< 1/1000
Markevich2004_MAPK_phosphoRandomElementary	BIOMD0000000028	0.690	< 1/1000
Markevich2004_MAPK_AllRandomElementary	BIOMD0000000030	0.690	< 1/1000
Markevich2004_MAPK_orderedMM	BIOMD0000000027	0.673	< 1/1000
Markevich2004_MAPK_orderedMM2kinases	BIOMD0000000031	0.673	< 1/1000
Markevich2004_MAPK_phosphoRandomMM	BIOMD0000000029	0.617	< 1/1000
Hornberg2005_ERKcascade	BIOMD0000000084	0.468	< 1/1000
McClean2007_CrossTalk	BIOMD0000000116	0.397	< 1/1000
Kofahl2004_pheromone	BIOMD0000000032	0.348	< 1/1000
Kim2007_Wnt_ERK_Crosstalk	BIOMD0000000149	0.323	< 1/1000
Brown2004_NGF_EGF_signaling	BIOMD0000000033	0.260	< 1/1000
Goldbeter1995_CircClock	BIOMD0000000016	0.244	< 1/1000
Sasagawa2005_MAPK	BIOMD0000000049	0.240	< 1/1000
Ung2008_EGFR_Endocytosis	BIOMD0000000205	0.234	< 1/1000
Leloup1999_CircClock	BIOMD0000000021	0.229	< 1/1000
Goldbeter1991_MinMitOscil_ExplInact	BIOMD0000000004	0.222	< 1/1000
Goldbeter1991_MinMitOscil	BIOMD0000000003	0.214	< 1/1000
Leloup1998_CircClock_LD	BIOMD0000000171	0.201	< 1/1000
Tyson1991_CellCycle_6var	BIOMD0000000005	0.199	< 1/1000
Veening2008_DegU_Regulation	BIOMD0000000240	0.180	< 1/1000
Neves2008_Cell_Shape	BIOMD0000000182	0.168	4.0e-03
Fisher2006_Ca_Oscillation_dpdnt_NFAT_dynamics	BIOMD0000000122	0.164	5.0e-03
Leloup2003_CircClock_DD	BIOMD0000000073	0.163	5.0e-03



Search with
Huang and Ferrell 1996

Data alignment and retrieval

Model	BioModel	Similarity	*P*-value	Overlap	*P*-value
Wolf2001 respiratory oscillations	BIOMD0000000090	0.207	<=1e-3	6	6.6e-09
Chassagnole2001_Threonine Synthesis	BIOMD0000000066	0.184	<=1e-3	4	1.5e-05
Curien2009_Aspartate_Metabolism	BIOMD0000000212	0.170	<=1e-3	5	3.6e-07
Curien2003_MetThr_synthesis	BIOMD0000000068	0.141	<=1e-3	2	1.0e-02
Proctor2007_ubiquitine	BIOMD0000000105	0.098	2.0e-03	1	1.4e-01
Curto1998_purineMetabol	BIOMD0000000015	0.063	1.1e-02	2	1.0e-02
Ibrahim2008_Spindle_Assembly_Checkpoint_dissociation	BIOMD0000000186	0.057	1.8e-02	0	1.0e+00
Ibrahim2008_Spindle_Assembly_Checkpoint_convey	BIOMD0000000187	0.057	1.8e-02	0	1.0e+00
Rodriguez-Caso2006_Polyamine_Metabolism	BIOMD0000000190	0.040	7.1e-02	1	1.4e-01
Nijhout2004_Folate_Cycle	BIOMD0000000213	0.032	1.1e-01	1	1.4e-01
Morrison1989_FolateCycle	BIOMD0000000018	0.030	1.3e-01	1	1.4e-01
Zatorsky2006_p53_Model3	BIOMD0000000154	0.023	2.5e-01	0	1.0e+00
Zatorsky2006_p53_Model6	BIOMD0000000155	0.023	2.5e-01	0	1.0e+00
Hunziker2010_p53_StressSpecificResponse	BIOMD0000000252	0.023	2.5e-01	0	1.0e+00
Zatorsky2006_p53_Model5	BIOMD0000000156	0.022	2.7e-01	0	1.0e+00
Zatorsky2006_p53_Model4	BIOMD0000000157	0.022	2.7e-01	0	1.0e+00
Zatorsky2006_p53_Model2	BIOMD0000000158	0.022	2.7e-01	0	1.0e+00
Zatorsky2006_p53_Model1	BIOMD0000000159	0.022	2.7e-01	0	1.0e+00
Proctor2008_p53_Mdm2_ATM	BIOMD0000000188	0.013	4.3e-01	0	1.0e+00
McClean2007_CrossTalk	BIOMD0000000116	0.012	4.7e-01	0	1.0e+00
Proctor2008_p53_Mdm2_ARF	BIOMD0000000189	0.012	4.9e-01	0	1.0e+00
Haberichter2007_cellcycle	BIOMD0000000109	0.011	5.0e-01	0	1.0e+00
Sasagawa2005_MAPK	BIOMD0000000049	0.006	5.5e-01	0	1.0e+00

Search with microarray experiment on yeast, showing coupled metabolic oscillations and burst of DNA replication

What does using Identifiers.org bring

Flexible access to multiple resources serving the dataset

No need to continuously maintain your own identifiers-URLs mapping

Using shared unique identifiers facilitate data comparison and integration

Help implementing the Linked and Open Data in a consistent manner

Use of identifiers.org in triples

Subject	Predicate	Object
EGFR	*located_in*	plasma membrane
P00533 (UniProt)	OBO_REL:0000008 (OBO Relation ontology)	GO:0005886 (Gene Ontology)

http://identifiers.org/uniprot/P00533

http://identifiers.org/obo.ro/OBO_REL:0000008

http://identifiers.org/obo.go/GO:0005886

Tools developing support for those URIs

- **Data resources**
  - BioModels Database (kinetic models)
  - Pathway Commons (BioPAX)
  - Physiome Model Repository (CellML)
  - SABIO-RK (reaction kinetics)
  - Yeast consensus model database
  - Human consensus model database
  - E-MeP (structural genomics)

- **Application software**
  - ARCADIA (graph editor)
  - BIOUML (modelling and simulation)
  - BridgeDb
  - CellDesigner
  - COPASI (Simulation)
  - Cpath
  - JSBML

- **Application software** (continued)
  - LibAnnotationSBML
  - LibSBML
  - PathText
  - Pathway Commons
  - SAINT (semantic annotation)
  - SBML2BioPAX
  - SBML2LaTeX
  - SBMLeditor (model editor)
  - SemanticSBML (annotation, merging, comparison, ...)
  - Snazer (network analysis, simulations)
  - Systems Biology Workbench (model design and simulation)
  - The Virtual Cell (simulation)
  - PSICQUIC
  - ...

This is what you store in your database

http://identifiers.org/ec-code/1.1.1.1

This is what you use in your import/export format

http://identifiers.org/ec-code/1.1.1.1

This is what you use in your webpage

http://identifiers.org/ec-code/1.1.1.1

This is what you paste in your web browser

http://identifiers.org/ec-code/1.1.1.1

This is what you quote in your publication

http://identifiers.org/ec-code/1.1.1.1

Nick Juty, Camille Laibe

The community of Computational Systems Biology

Identifiers.org advisory board

Michel Dumontier (W3C HCLS, Bio2RDF)

Michael Galperin (NCBI, USA - NAR Database issue)

Pascale Gaudet (BioDbCore)

Lee Harland (Connected Discovery, UK - Open PHACTS)

Michael Hucka (SBML)

Toshiaki Katayama (University of Tokyo, Japan - BioRuby, KEGG)

Nicolas Le Novère (MIRIAM Registry)

Philippe Rocca-Serra (BioSharing)

Mark Wilkinson (SADI, LSRN)

How is it different from LSIDs?

Collection identification is formatted consistently

Collection registry is public and open

Identifiers.org URIs are resolvable

urn:lsid:uniprot.org:uniprot:P12345

http://identifiers.org/uniprot/P12345

>>http://www.uniprot.org/uniprot/P12345

>>http://purl.uniprot.org/uniprot/P12345

How is it different from DOIs?

Identifiers.org re-use the datasets identifiers from the data provider

Identifiers.org URIs are free

Identifiers.org URIs can be resolved to alternative resources

Redirection of Identifiers.org is configurable

http://dx.doi.org/10.1093/nar/gkr1097

>>m.nr.oxfordjournals.org/content/40/D1/D580

http://identifiers.org/pubmed/22140103

>>http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:22140103]

http://www.ebi.ac.uk/citexplore/citationDetails.do?dataSource=MED&ex

>>http://www.ncbi.nlm.nih.gov/pubmed/22140103

>>http://www.hubmed.org/display.cgi?uids=22140103

How is it different from PURLs?

Identifiers.org URIs can be resolved to alternative resources

Identifiers.org URIs are unique. Everyone using the same to refer to a given piece of information

http://purl.uniprot.org/go/0016323

http://purl.org/obo/owl/GO#GO_0016323

How is it different from Sharednames?

Identifiers.org URIs can be resolved to alternative resources

Identifiers.org is (so far) limited to life sciences

Identifiers.org is what Sharednames aimed to be

The Linked and Open Data agenda

Linked Data

"The Semantic Web isn't just about putting data on the web. It is about **making links**, so that a person or machine can explore the web of data."

Tim Berners-Lee

- Use URIs as names for things
- Use **HTTP URIs** so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- Include links to other URIs. so that they can discover more things.

Advantages of Identifiers.org URIs

- Uniform syntax
  (URIs, as LSIDs, PURLs, DOIs)
- A unique URI per element of a collection
  (as DOIs, better than most PURLs)
- Reuse the dataset identifier of the provider
  (better than DOIs)
- Multiple dereferences
  (better than all systems)
- Curated registry to maintain the resolution
  (better than all systems)
- Multiply resolvable
  (better than DOIs and PURIs)
- Free, as in free beer (better than DOIs)