

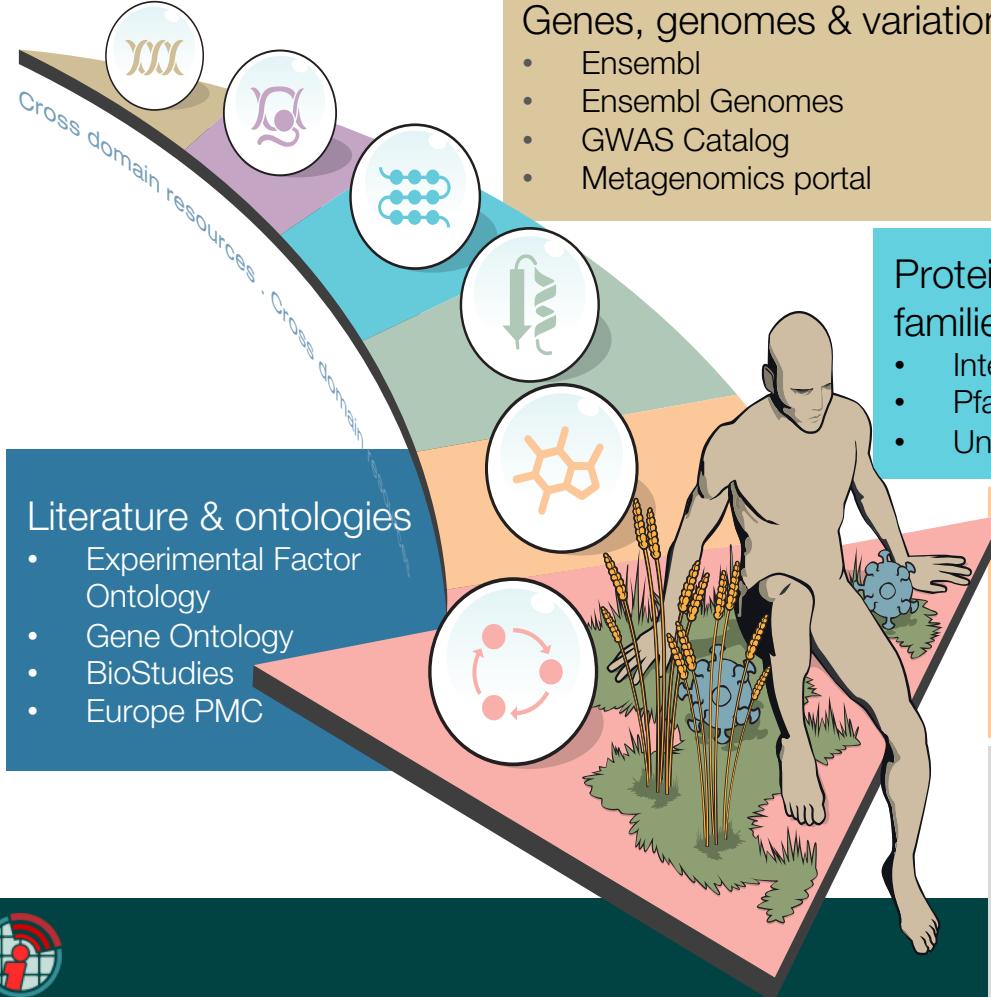


Identifiers.org

Compact Identifier resolution services

MSc. Manuel Bernal Llinares
EMBL-EBI

Data resources at EMBL-EBI



Data collections and identifiers in life sciences

Actionable identifiers embedded in URLs

<https://www.ebi.ac.uk/pdbe/entry/pdb/2gc4>

<http://www.wormbase.org/db/gene/gene?name=WBGene00000001;class=Gene>

<http://www.ebi.ac.uk/ena/data/view/Taxon:9606>

- The same data collection is often provided by many alternative physical locations

<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=9606>

<http://www.ebi.ac.uk/ena/data/view/Taxon:9606>



Challenges

- Multiple URLs for the same collection make object unification challenging
- Which resource should be used for annotation or citation
- A given location may be down, or change its URL, resulting in dead links



Compact Identifiers

- A unique prefix indicating the assigning authority
- A locally assigned database identifier sometimes called an accession
- An additional provider level prefix (provider_code) to identify individual hosts

prefix:accession

provider

code/prefix:accession



Which would you rather see?

A?

- <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=9606>
- <http://www.rcsb.org/pdb/explore/explore.do?structureId=2gc4>
- <http://www.wormbase.org/db/gene/gene?name=WBGene00000001;class=Gene>

B?

- taxon:**9606**
- pdb:**2gc4**
- wb:**WBGene00000001**



Support for compact in-line data citation

... the protein we used (Protein Data Bank 2gc4) in our tests ...

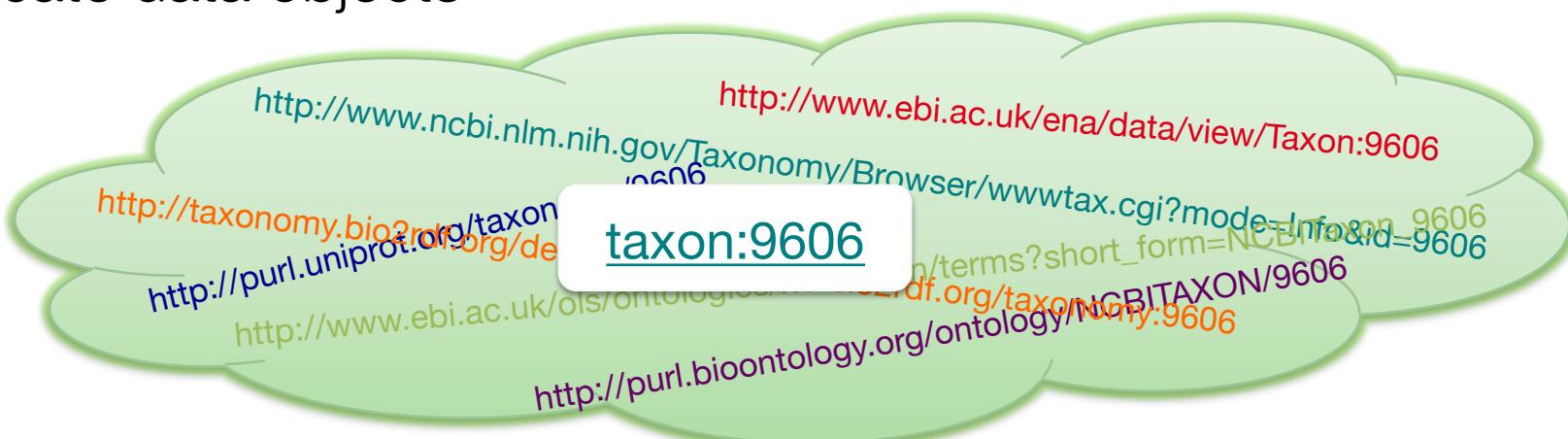
A persistent identifier
that provides access
and metadata.

... the protein we used ([pdb:2gc4](#)) in our tests ...



Compact Identifier Resolution Services

Actionable Compact Identifiers provide unique stable, resolvable and location-independent URIs to identify and locate data objects



<https://identifiers.org/taxon:9606>
<https://n2t.net/taxon:9606>

N2T RESOLVER
NAMES ➡ THINGS



EMBL-EBI

Meta-resolver harmoni(z|s)ation

SCIENTIFIC DATA

SCIENTIFIC DATA

Article | [OPEN](#) | Published: 08 May 2018

Uniform resolution of compact identifiers for biomedical data

Sarala M. Wimalaratne, Nick Juty, John Kunze, Greg Janée, Julie A. McMurry, Niall Beard, Rafael Jimenez, Jeffrey S. Grethe, Henning Hermjakob, Maryann E. Martone & Tim Clark 

Scientific Data 5, Article number: 180029 (2018) | [Download Citation](#) 



 **FORCE11**
The Future of Research Communications and e-Scholarship

Editorial | [OPEN](#) | Published: 08 May 2018

On the road to robust data citation

Scientific Data 5, Article number: 180095 (2018) | [Download Citation](#) 

Scientific Data is changing the way we incorporate links into our data citations. We will now be taking advantage of the resolver services offered by identifiers.org and N2T.net to provide more standardized and predictable links for biomedical datasets that have accession identifiers when they are cited in our publications.

MEMORANDUM OF UNDERSTANDING
BETWEEN THE CALIFORNIA DIGITAL LIBRARY AND
THE EUROPEAN MOLECULAR BIOLOGY LABORATORY :
COMPACT IDENTIFIER SUPPORT

N2T RESOLVER
NAMES → THINGS 



EMBL-EBI 

Identifiers.org

The screenshot shows the Identifiers.org homepage. At the top, there's a navigation bar with links to EMBL-EBI, Services, Research, Training, About us, and a dropdown for EMBL-EBI Hinxton. Below the header is a large banner with the Identifiers.org logo and a network graph background. The main content area includes a search bar, an Advanced search link, a Feedback link, and a "Connect with us" section featuring Twitter and GitHub links. A prominent teal box on the left contains the text: "Identifiers.org is an established resolving system that enables the referencing of data for the scientific community, with a current focus on the Life Sciences domain. It handles persistent identifiers in the form of URLs and CURIES. This allows the referencing of data in both a location-independent and resource-dependent manner. The provision of resolvable identifiers (URLs) fits well with the Semantic Web vision, and the Linked Data initiative." Below this text is a form field with placeholder text "Enter a prefix:identifier and press Enter". To the right of the teal box are four buttons: Registry, Request prefix, Web Services, and Download. Below these buttons is a "Data records" section with two rows of data: Collections (623) and Resources (768). Last updated is listed as 2018-01-17. To the right of the teal box is a "Meta-resolvers" section which states: "Identifiers.org share a common prefix registry with N2T.net, based at the California Digital Library, enabling users to resolve Compact Identifiers using the Identifiers.org or N2T.net meta-resolver." At the bottom of the page is a footer stating "Identifiers.org is part of the ELIXIR infrastructure" with a "Learn more" link.

Identifiers.org is an established resolving system that enables the referencing of data for the scientific community, with a current focus on the Life Sciences domain. It handles persistent identifiers in the form of URLs and CURIES. This allows the referencing of data in both a location-independent and resource-dependent manner. The provision of resolvable identifiers (URLs) fits well with the Semantic Web vision, and the Linked Data initiative.

Enter a prefix:identifier and press Enter

Registry Request prefix Web Services Download

Data records

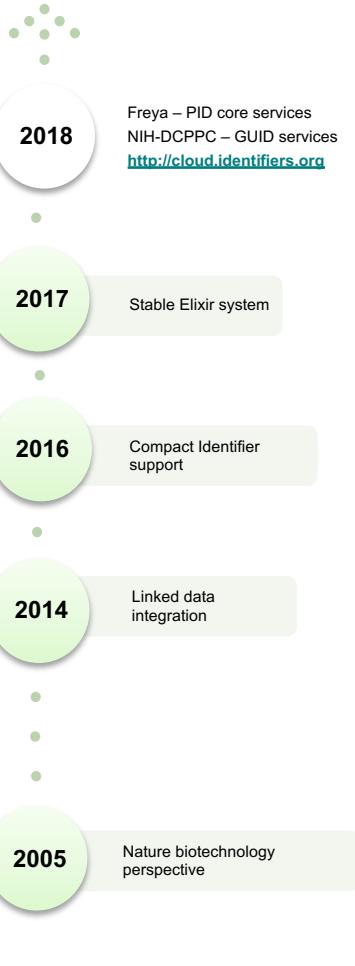
Collections	623
Resources	768

Last updated 2018-01-17

Meta-resolvers

Identifiers.org share a common prefix registry with N2T.net, based at the California Digital Library, enabling users to resolve Compact Identifiers using the Identifiers.org or N2T.net meta-resolver.

Identifiers.org is part of the ELIXIR infrastructure [Learn more](#)



<https://identifiers.org>

EMBL-EBI

Register Prefixes for Compact Identifiers

Identifiers.org Registry is a manually curated resource of over 650 life science data collections with metadata, alternative access URLs, and identifier patterns.

Request an identifier prefix

Please complete this form to register an identifier prefix that can be recognized by the meta-resolvers at identifiers.org and n2t.net. Completing all fields will enable swift processing of your request.

*Required

Resource details

The "resource" is the set of data being provided, such as a database, photo archive, document repository, or ontology.

Name:

Name of the resource. Example: Protein Data Bank.

Your answer

Description:

One or more sentences describing the resource. Example: The Protein Data Bank is the single worldwide archive of structural data of biological macromolecules.

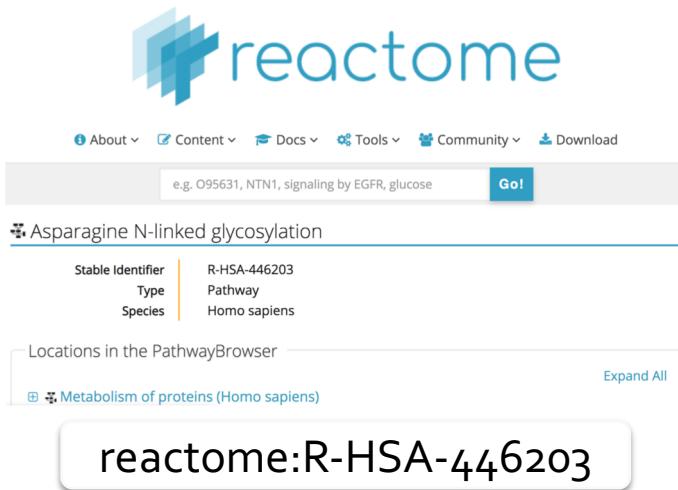
Your answer



<https://identifiers.org/request/prefix>

Compact Identifier Metadata

Identifiers.org metadata service provides access to the Compact Identifier Schema.org metadata encoded by the resource providers.



The screenshot shows the Reactome search interface. A search bar contains the query "e.g. 095631, NTN1, signaling by EGFR, glucose". Below the search bar, the result title "Asparagine N-linked glycosylation" is displayed. To the left of the title, there is a "Stable Identifier" section with "R-HSA-446203", "Type" as "Pathway", and "Species" as "Homo sapiens". Below the title, a "Locations in the PathwayBrowser" section lists "Metabolism of proteins (Homo sapiens)". At the bottom of the result card, the identifier "reactome:R-HSA-446203" is shown.

```
{  
    apiVersion: "1.0",  
    errorMessage: "No cached metadata for access URL 'https://reactome.org/content/detail/R-HSA-446203', score '40' No  
    Cached metadata found, running in-line metadata extraction for access URL 'https://reactome.org/content/detail/R-  
    HSA-446203', score '40' null",  
    - payload: {  
        - metadata: [  
            - {  
                @context: "http://schema.org",  
                @type: "Website",  
                url: "https://reactome.org/",  
                - potentialAction: [  
                    @type: "SearchAction",  
                    target: "https://reactome.org/content/query?q=(term)",  
                    query-input: "required name=term"  
                ],  
            },  
            - {  
                @context: "http://schema.org",  
                @type: "Organization",  
                url: "https://reactome.org",  
                logo: "https://reactome.org/templates/favourite/images/logo/logo.png",  
                email: "help@reactome.org"  
            },  
            - {  
                name: "Asparagine N-linked glycosylation",  
                description: "N-linked glycosylation is the most important form of post-translational modification for  
                proteins synthesized and folded in the Endoplasmic Reticulum (Stanley et al. 2009). An early study in  
                1999 revealed that about 50% of the proteins in the Swiss-Prot database at the time were N-glycosylated  
                (Apweiler et al. 1999). It is now established that the majority of the proteins in the secretory"  
            }  
        ]  
    }  
}
```

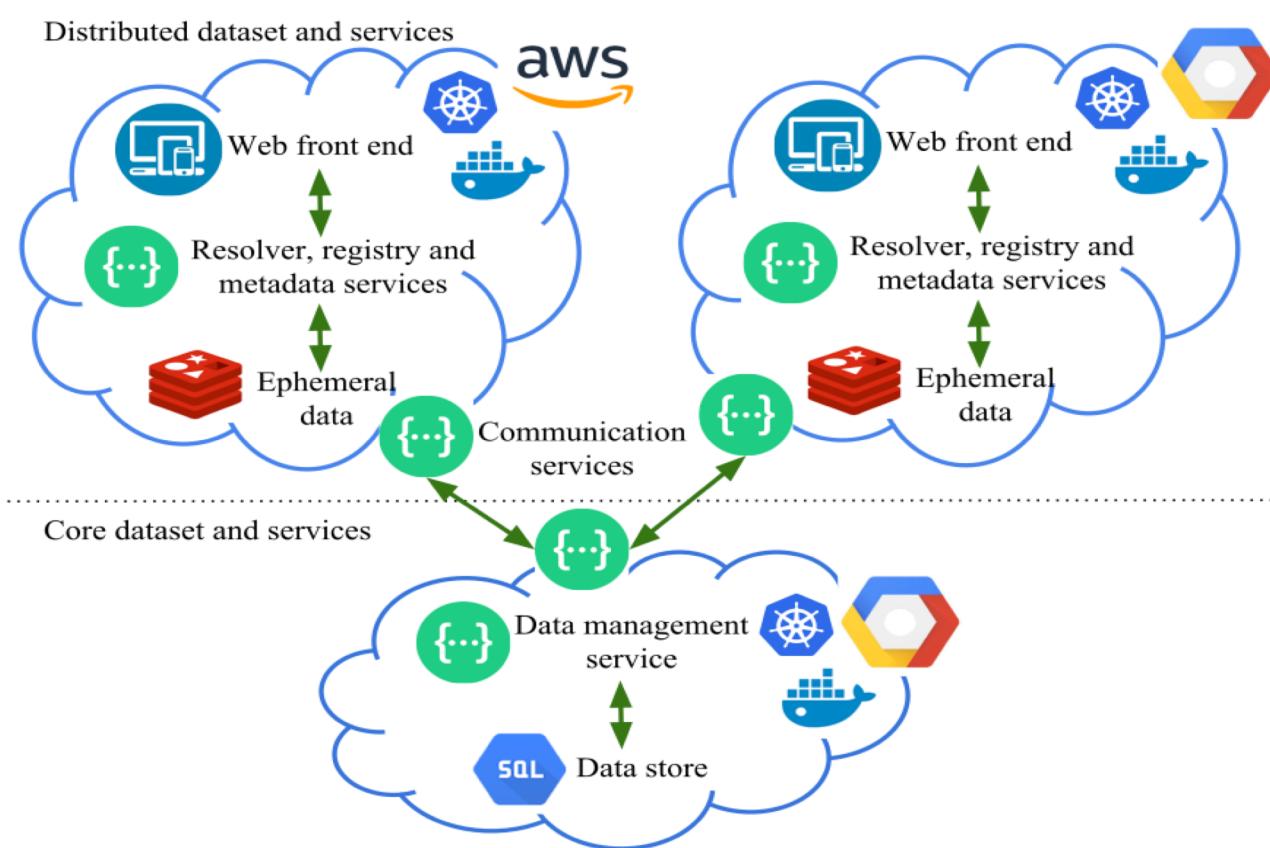
<http://metadata.api.identifiers.org/reactome:R-HSA-446203>



<http://metadata.api.identifiers.org/{Compact Identifier}>

EMBL-EBI

Identifiers.org in the cloud



<http://cloud.identifiers.org>

EMBL-EBI

Hackathon Activities

- Access Identifiers.org services via the [cloud-libapi](#)
- Test our cloud infrastructure at <http://cloud.identifiers.org>
- Deploy your own Identifiers.org services in the cloud
- Contribute to development of [identifiersJS](#) package for accessing Identifiers.org services



Acknowledgements

- EMBL-EBI
 - Henning Hermjakob
 - Sarala Wimalaratne
- California Digital Library
 - John Kunze
 - Greg Janee
- The University of Manchester
 - Nick Juty
- Harvard University
 - Timothy Clark
- Elixir Hub
 - Rafael Jimenez

Funding



Excelerate/676559, EU/312455,
THOR/654039, BB/J019305/1,
BB/E006248/1,
OpenPhacts/115191,
NIH/1U41HG006104
Contact

sarala@ebi.ac.uk





Annotate and Cite your data using **Compact Identifiers**

[taxon:9606](#)