



# (GIGA)<sup>n</sup> DB

## Submission wizard to enable authors to curate their own metadata

**Xiao Si Zhe,**  
Robert L Davidson,  
Peter Li,  
Scott C Edmunds,  
Laurie Goodman,  
Christopher I Hunter.

GigaDB (<http://www.gigadb.org>) is an integrated database of 'big-data' studies from the life sciences. The initial goals of GigaDB are to assign Digital Object Identifiers (DOIs) to datasets to allow them to be tracked and cited, and to provide a user-friendly web interface for providing easy access to GigaDB datasets and files. The new GigaDB submission wizard can help authors curate their own metadata (information about the dataset), which allows searching and indexing.

The submission wizard is based on the GigaDB database schema. It is divided to three main parts: Study, Authors and Sample details, with the ability to include links to Projects, genome browsers, related DOI(s) and even other external URL links. Authors simply follow the steps and fill in the information on the web-forms. After curators review the submission, authors will be given credentials to allow upload of data files to the GigaDB server. The submission wizard can then be used to add and modify any metadata about the files (description, format, links to samples etc...). GigaDB creates DataCite Metadata standard compliant XML and after our curators review your submission, this is used to assign a unique DOI to the dataset, which can be used for citation purposes.

GigaDB is an open-access data repository closely linked to GigaScience journal, we aim to promote reproducibility of published research by allowing access to all the data, and maximize the reuse of published data by giving a means to credit data publication by citation. The new GigaDB submission wizard plays an important role in the GigaDB structure. It provides a new simple to use web-interface for authors to curate their own data with the most relevant and accurate information. We wish solicit open discussions from others to help our future developments. ([database@gigasciencejournal.com](mailto:database@gigasciencejournal.com))

1.

**Create Dataset**

Study Author Project Link External Link Related DOI Sample

Fields with \* are required.

Submitter \*  No image

Types

- Epigenomic
- Genomic
- Metagenomic
- Transcriptomic
- Workflow
- Software
- Imaging
- Metabonomic
- Proteomic
- Neuroscience
- Network-Analysis
- Genome-Mapping

Image Upload

Image Source \*

Image Tag

Image License \*

Image Photographer \*

Title \*

Estimated Dataset Size  Bytes

Description

3.

**Part of a project?**

Study Author Project Link External Link Related DOI Sample

Project Name

If your dataset is part of a larger collaborative international project, please select it from the list below.

Project

### Cross reference data

Study Author Project Link External Link Related DOI Sample

Link Type	Link
ext_acc_mirror	BioProject: ERP012345

Please add the database and accession numbers of any data in this dataset that is also stored in other repositories, e.g. sequences in the Sequence Read Archive (SRA).

Database

Accession number

### Add link to a genome browser or website?

Study Author Project Link External Link Related DOI Sample

Url	External Link Type
No results found.	

If your data is a genomic assembly that is represented in a public genome browser, please add the direct URL here.

External Link Type

Url

4.

**Add Relationship with another DOI?**

Study Author Project Link External Link Related DOI Sample

Related DOI	Relationship
No results found.	

If your dataset is directly related to any other GigaDB published DOI, please add the DOI number here and the type of the relationship.

Related DOI

Relationship

### Add sample details

Study Author Project Link External Link Related DOI Sample PX Info

Please provide the details of all samples represented by the data being submitted, this should include their taxonomic identification, and as much information about the sample and its collection as possible.

Sample ID	Species name	Attribute Name	Attribute Value	Attribute Unit
belly	soil metagenome	weight loss in last three months	5	kilogram
		PX_sample_processing_protocol	this is where i write about	
		Attribute name	Attribute Value	
		alternative accession-SRA_sample	SRA012345	absorbance unit
		PX_sample_processing_protocol	this is where i write about	
		health	Attribute Value	
		amniotic fluid/foetal health status		
		amniotic fluid/maternal health status		
		health or disease status of specific host at time of collection		
		Age sample		

2.

### Add Authors

Study Author Project Link External Link Related DOI Sample

First name	Middle name	Last name	ORCID	Order
Christopher		Hunter		1

First Name  Middle Name (Optional)  Last Name  ORCID (Optional)

### Thanks to:

Huayan Gao (*GigaScience*), Yong Zhang (BGI; China National Genebank) Shaoguang Liang (BGI-SZ), Alex Wong, Dennis Chan (BGI-HK), Tin-Lap Lee (CUHK), Qiong Luo, Senghong Wang, Yan Zhou (HKUST), and Mark Viant (Birmingham Uni)



Submit your next manuscript containing large-scale data and workflows to *GigaScience* and take full advantage of:

- No space constraints, and unlimited data and workflow hosting in *GigaDB* and *GigaGalaxy*
- Open access, open data, and highly visible work freely available for distribution
- Collation of Data Citation Numbers by Thompson-Reuters
- Indexing of papers in PubMed Central, Google Scholar, etc.



With financial support from:

