

Dataset creation

Create dataset

The screenshot shows a form for creating a dataset. It has three sections: 'NAME', 'KIND', and 'TYPE'. The 'NAME' section has a text input field with a pink circle containing the number '1' next to it. The 'KIND' section has two radio buttons: 'RAW' (selected) with a pink circle containing '2', and 'PROCESSED' with a pink circle containing '3'. The 'TYPE' section has two radio buttons: 'TABULAR' (selected) with a pink circle containing '4', and 'NON-TABULAR' with a pink circle containing '5'. At the bottom right, there are two buttons: a pink 'SUBMIT' button and a white 'CANCEL' button with a pink border.

1. Enter the name of the dataset
2. **Raw data** is the primary data collected (e.g. instruments readings) which has not been subject to any manipulation;
3. **Processed data** refers to the data obtained after their analysis (e.g. calculation of averages, standard deviations, etc.), removal of outliers or errors, normalisations, etc.
4. **Tabular** format refers to the data arranged or structured in rows and columns according to the data schema concept of INTERVALS (e.g. in Excel or csv files);
5. **Non-tabular** format refers to data in an arbitrary format (e.g. images, text, omics CEL files or pdf files as well as tables not formatted according to the data schema concept).

Dataset creation

Dataset information

DATASET INFORMATION

Edit

Edit the dataset details. The description should be a few paragraphs long and can contain images, videos, links... All relevant references should be added directly at the end of the section.

NAME
Test dataset

DESCRIPTION ? 1
None


EXPERIMENTAL SYSTEM ? 2
None

1. Enter the description of the dataset. The description should not exceed a few sentences.
2. Add or edit the study experimental system(s). If your experimental design is not listed below, contact the admin to get it added. It is important to thoroughly select the experimental design(s) as they will be used as filters when searching for datasets on the platform.

Dataset creation



Protocols

PROTOCOLS

Edit 

Add or remove protocols associated to this study. You can associate protocols that have been created separately (through your dashboard).

PROTOCOLS

 *at one or more* 

CONFIRM

1. Add or remove protocols associated to this study. You can associate protocols that have been created separately (through your dashboard).

Metadata

METADATA FILES

In this section the metadata is collected, followed by a validation step. The metadata is represented by two separate files, the metadata schema and the metadata file.

This step is required in order to support the documentation of the experiment and also the interoperability goal by using controlled vocabularies.

- The **metadata schema** is represented by a tabular form (e.g. Excel) with eight predefined columns: StudyStage, ColumnName, Role, Type, Ontology, Unit, UnitOntology and Description.
- The **metadata file** template is generated following the entries under ColumnName in the metadata schema, and will be used to store the metadata of your experiment.

Detailed tutorials on metadata schema and metadata file creation, as well as on the validation step are available [here](#).

Examples of a metadata schema and metadata files are available [here](#).

METADATA SCHEMA Choose file +

1 e

METADATA Choose file +

2 e

VALIDATION Validate ✓

3 ation hasn't started yet

1. The **metadata schema** is represented by a tabular form (e.g. Excel) with eight predefined columns: StudyStage, ColumnName, Role, Type, Ontology, Unit, UnitOntology and Description.
2. The **metadata file** template is generated following the entries under ColumnName in the metadata schema, and will be used to store the metadata of your experiment.
3. Click "Validate" button to validate the uploaded files.

Dataset creation

Tabular data

TABULAR DATA FILES

In this section the tabular data is collected, followed by a validation step. The data is represented by two separate files, the data schema and the data file.

The **data schema** has the same structure as the metadata schema created and uploaded in the previous step and is represented by a tabular form with the following columns headings: StudyStage, ColumnName, Role, Type, Ontology, Unit, UnitOntology and Description.

The **data file** is generated based on the data schema as in the previous step for metadata file, and it will be used to store the data generated in your experiment.

Detailed tutorials on data schema and data file creation, as well as on the validation step are available here.

Examples of data schema and data files are available here.

DATA SCHEMA Choose file +

1.

DATA Choose file +

2.

VALIDATION Validate ✓


3. Validation hasn't started yet

1. The **data schema** has the same structure as the metadata schema created and uploaded in the previous step and is represented by a tabular form with the following columns headings: StudyStage, ColumnName, Role, Type, Ontology, Unit, UnitOntology and Description.
2. The **data file** is generated based on the data schema as in the previous step for metadata file, and it will be used to store the data generated in your experiment.
3. Click "Validate" button to validate the uploaded files.

Dataset creation

Category

DATASET CATEGORY

Test Items Edit 	Endpoints	Organisms	Tested Sample Types	Diseases & pathways
--	-----------	-----------	---------------------	---------------------

Add or edit the study test item(s). A test item is the item on which the dataset was generated. If your test item of interest is not listed below, [contact the admin to get it added](#).

TEST ITEMS
None

1. Add or edit the study test item(s). A test item is the item on which the dataset was generated. If your test item of interest is not listed below, contact the admin to get it added.
2. Add or edit the organism(s) on which the dataset was generated. If the study's organisms is not listed below, contact the admin to get it added.
3. Add or edit the organ tissue(s) on which the dataset was generated. If the study's organ tissue is not listed below, contact the admin to get it added.
4. Add or edit the dataset disease and pathways. If the dataset's disease or pathway is not listed below, contact the admin to get it added. It is important to thoroughly select the disease(s) and pathway(s) as they will be used as filters when searching for datasets on the platform.