
submissionUI Documentation

Release 1.0

Erica, Deepak, Daofeng, Ting

Oct 25, 2018

Contents

1	Introduction	3
2	Submitting data and metadata to the TaRGET II DCC	5
2.1	Data submission	6
2.2	Metadata organization	13
2.3	Metadata submission via the Accession Registry	14
2.4	Metadata bulk submission via the web UI	19
3	Report an issue	23
4	Indices and tables	25

Contents:

CHAPTER 1

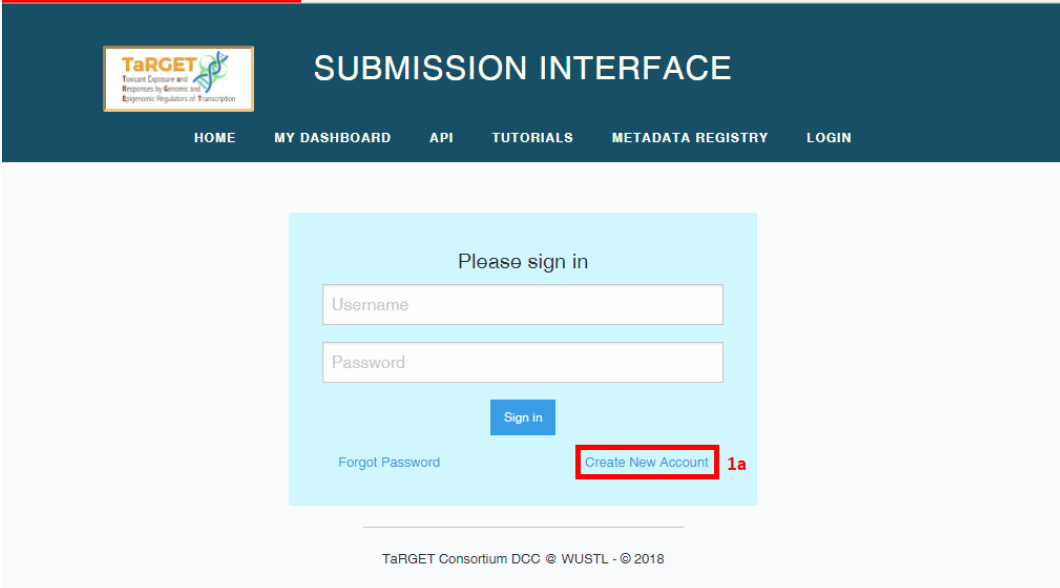
Introduction

Welcome to the TaRGET DCC Submission Pipeline! Here, you can upload data files to the consortium database as well as record information about the experiment that generated the data.

Submitting data and metadata to the TaRGET II DCC

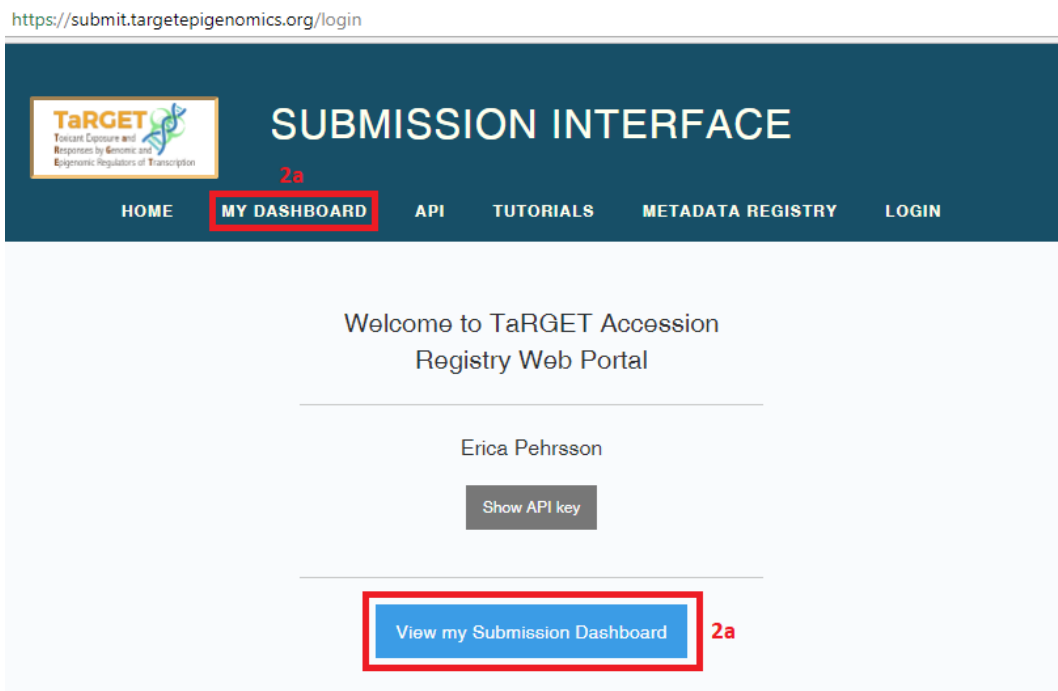
1. Log in to the Submission Interface (<https://submit.targetepigenomics.org>)

<https://submit.targetepigenomics.org/login> 1.



TaRGET Consortium DCC © WUSTL - © 2018

- (a) On your first login, you will need to register as a new user (“Create New Account”). After completing all fields, select “Create”, then “Login”. If you have a consortium wiki account, you should use the same username and email for this account.
2. Go to your Submission Dashboard (<https://submit.targetepigenomics.org/dashboard>)
 - (a) You can get to your Submission Dashboard by selecting the “View my Submission Dashboard” button on your homepage or selecting “My Dashboard” from the banner.

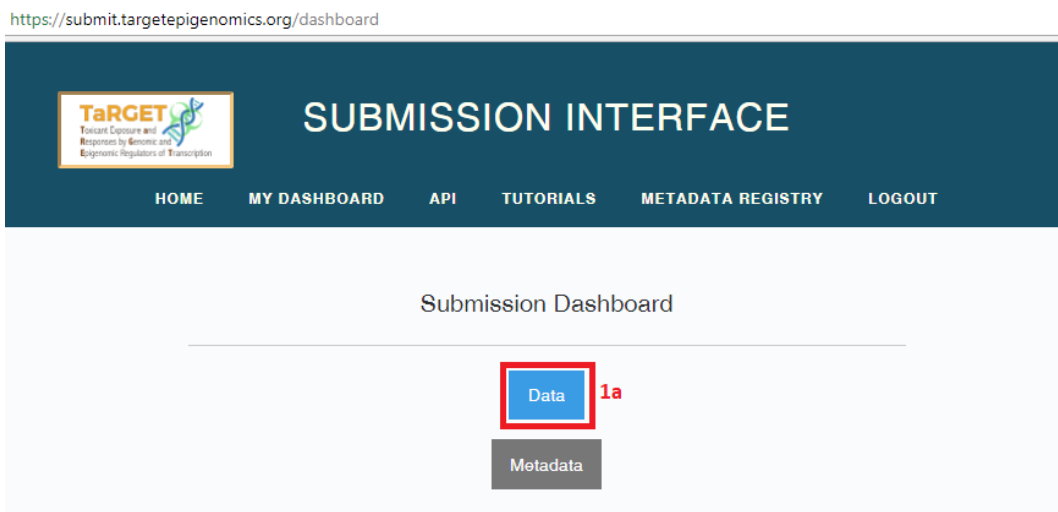


- (b) From here, you may choose to register metadata or data by selecting the appropriate button.

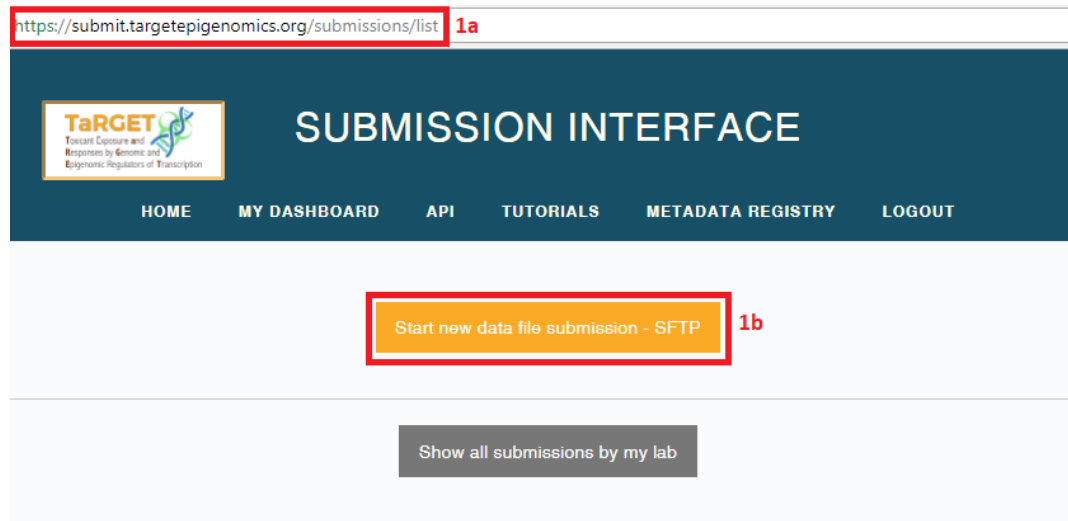
2.1 Data submission

1. Create a data submission


- (a) Select the “Data” button on your dashboard. You will be taken to a page (<https://submit.targetepigenomics.org/submissions/list>) that lists your previous data submissions (if any) and allows you to start a new submission.



- (b) To create a new submission, select “Start new data file submission - SFTP”. The Aspera upload option has been deprecated.



- (c) Specify your lab and username and other basic information for your submission. You may only include files from one assay (e.g., ATAC-seq, RNA-seq) in a single submission, and they must all be single-end or paired-end. Select "Submit".



New Submission

Lab:

Wang Lab ▼

Submitter Username:

coolsubmitter

Assay:

ATAC-seq ▼

Read type:

☒ Single-end
☐ Paired-end

Data phase:

☒ Pilot
☐ Production

Number of files

4

Comments/Notes for DCC

Any information

Submit

- (d) Review your entries and either confirm that they are correct (“Yes”) or return to the form to make changes (“No”).

Confirm metadata

No Yes ✓

Looks good ?

```
{
  "lab": "Wang Lab"
  "username": "ecpehrsson"
  "assay": "ATAC-seq"
  "data-phase": "pilot"
  "read-type": "paired-end"
  "quantity": "2"
  "comments": "TEST"
}
```

- (e) Next, follow the instructions to generate an md5sum list for your files, which will be used to ensure that the files were completely uploaded. **Please also note the new file naming requirements.** Submitted files must be in fastq or fastq.gz format. After pasting the md5sum list in the window, select “Submit” at the bottom of the page. An error will be returned if the naming convention has not been followed or the number of rows in the md5sum list does not match the number of files specified on the previous page.

New submission registration

Basic Information Declare Files Experiment Design Confirm

Steps:

- On your local computer/server using terminal/command line, cd into the directory where the data files to be uploaded are available
- Please ensure the files have extensions like this

Single-end reads

some-filename.SE.fastq.gz

Paired-end reads

other-filename.PE.R1.fastq.gz

other-filename.PE.R2.fastq.gz

Note:

other-filename string should be IDENTICAL for paired-end files

Filenames not in this format will be ignored
- Run md5sum calculation command for these files

```
md5sum * | awk -F' ' '{print $1","$2}' > md5sum.csv
```

Please make sure same files are uploaded as part of this submission

Paste md5sum, files (comma separated)

Hint: Step 3, left panel

```
635505428ae71ad62f84421c5851291a,WangT_29-4F5T_SIC0812_CCGCCACGA_S3.PE.R1.fastq.gz
585bc95945147b9528157fd1dcc7bed1,WangT_29-4F5T_SIC0812_CCGCCACGA_S3.PE.R2.fastq.gz
```

Submit

- (f) For pilot data, you will fill out a limited set of metadata fields during data submission. To fill out metadata, select files in the left pane, then select the relevant metadata for those files in the right pane and hit “Save”. On the review panel, select “Yes” if the metadata is correct. Repeat for all files. Files will be removed from the list once metadata has been registered for them, but

you will have the chance to correct all metadata in the next step.

New submission registration

✓ Basic Information
✓ Declare Files
 Experiment Design
 Confirm

Details Filer

Click on files on the left pane, fill out information on the right pane and hit Save button

Click Review on the right when done

<input type="checkbox"/>	md5sum	filename
<input checked="" type="checkbox"/>	635505428ae71ad62f84421c585129...	WangT_29-4F5T_SIC0812_CCGCCCA...
<input checked="" type="checkbox"/>	585bc95945147b9528157fd1dcc7be...	WangT_29-4F5T_SIC0812_CCGCCCA...

Previous
Page 1 of 1
Next

10 rows

Add Details for Selected Files

Biosample:

Tissue or cell population collected from mouse (press Cmd + click to select multiple)

Kidney epithelium
Liver
 Liver - accessory lobe
 Liver - caudate lobe

Treatment:

Specific exposure

☒ CONTROL

OR

☐ BPA (Bisphenol A)
 ☐ TBT (Tributyltin)
 ☐ Genestein

Confirm metadata

No
Yes ✓

Looks good ?

Selected Files :

635505428ae71ad62f84421c5851291a

585bc95945147b9528157fd1dcc7bed1

```

{ 10 items
  "biosample" : "Liver"
  "protocol" : "NA"
  "strand-specific" : "false"
  "starting_cells" : ""
  "starting_cell_count" : ""
  "starting_nucleic_acid" : ""
  "previously_frozen_tissue" : "false"
  "detergent-added" : "false"
  "filtering_method" : "Celltrics column"
  "comments" : ""
}
          
```


Submission UI Screenshot with annotations:

- 2f**: Status column with green checkmarks.
- 2b**: ID column with submission IDs.
- 2b**: Submitter column with names like yemitan.
- 2b**: Assay column with assay types like RNA-seq, ATAC-seq.
- 2b**: Lab column with lab names like Bartolomei Lab, Wang Lab.
- 2b**: Description column with details like "Bartolomei 20171108_164616 RNA seq data".
- 2b**: Data Wrangler column with names like Ivy Chen, Yiran Hou.
- 2c**: Registered on column with dates like 2017-11-08.
- 2c**: md5sum status column with green checkmarks.
- 2c**: Data Phase column with phases like production, pilot.
- 2d**: Metadata Status column with buttons like "Metadata Status".
- 2c**: View Files column with buttons like "View Files".
- 2e**: Update experiment design column with buttons like "Update experiment design".

- (c) The md5sum status column shows the results of initial validation performed on your uploaded files. If the number, name, or md5sum of the uploaded files does not match those registered, the status will be a red X and the submission must be corrected before proceeding. To view files that were uploaded as part of a submission, select the “View Files” button. Once files have been uploaded to the DCC server, they will be listed here along with their own UUID and upload date.
- (d) Submissions are not considered complete until metadata has been registered for each file. You can view a report showing the level of completeness of each metadata object attached to each file in the submission by selecting the “Metadata Status” button. Selecting the “Upload Metadata” button will take you to instructions for bulk upload of production metadata (see below).

Submission UI Screenshot showing the Submission Interface and a table of file details.

Submission Interface Header:

- HOME MY DASHBOARD API TUTORIALS METADATA REGISTRY LOGOUT

Table of file details:

file_uuid	AR accession	Submitter	Lab	Bioproject	Diet	Treatment	Litter	Mouse	Biosample	Library	Assay	Reagent	File	MergedFile
Saa7b3e03b7077bca2c20e	TRG1TF000358	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b3e03b707712a2c20f	TRG1TF000359	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b3e04f3b7073d33a2c210	TRG1TF000360	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b41d5bae83a370c32667	TRG1TF000361	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b45a0bae83b0cc32668	TRG1TF000363	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b45af3b7071572a2c211	TRG1TF000362	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b496f3b7079c0a2c212	TRG1TF000364	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b4d15bae836273c3266a	TRG1TF000365	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b4d15bae8369dd3266b	TRG1TF000367	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Saa7b4d25bae835e12c3266d	TRG1TF000366	yemitan	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%

- (e) Pilot metadata can be updated using the “Update experiment design” button on the submission dashboard - NOT through the Accession Registry.
- (f) Once the submission metadata is complete, your data wrangler will sign off on your submission, and the “Status” column will have a green check. Then, the QC pipeline will be run on the submission, and a QC report will be linked to each file under “View Files”.

bb8b7c39-1412-4cab-b8ff-16bdec26e128

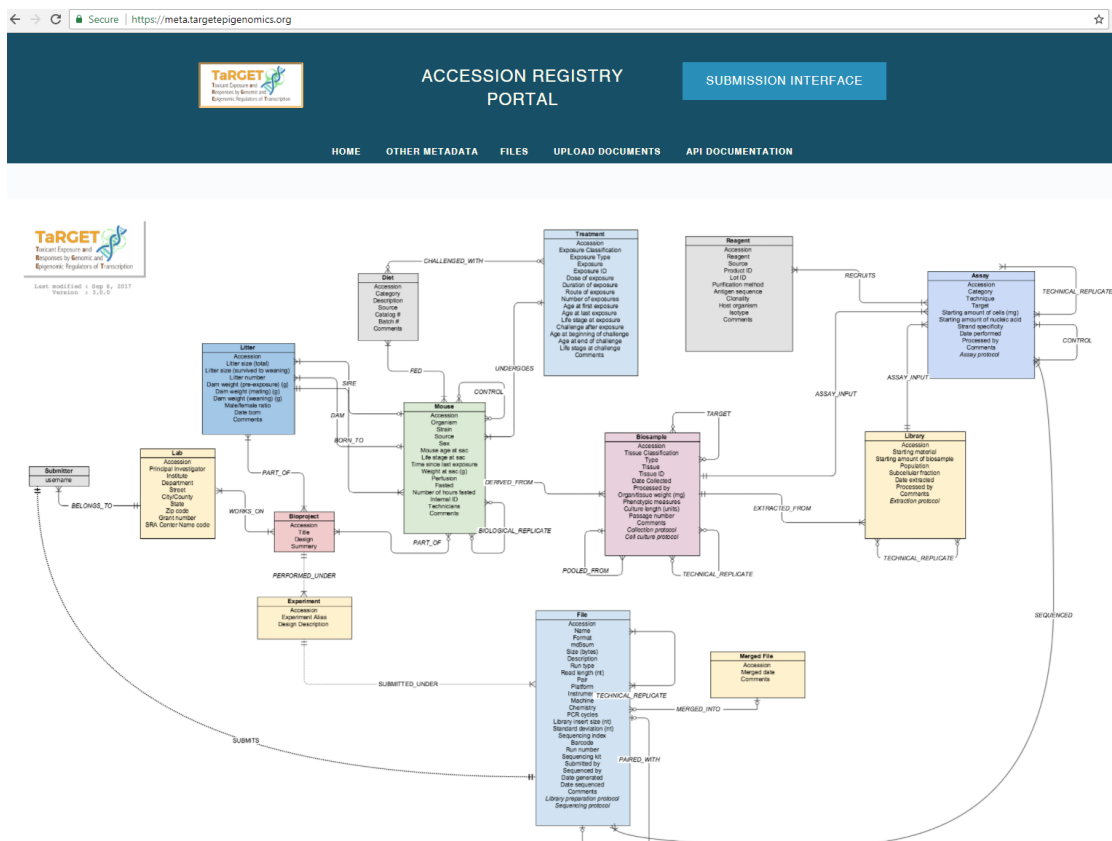
Note: For paired-end files, only .R1.fastq.gz will have a QC report

Files			
Id	Filename	Uploaded On	QC report
5b501b4a84510a6b283593f6	26-1F5T-ATAC_SIC0808_CTCCAGGGTA_S2.PE.R1.fastq.gz	2018-07-19T00:01:18.000Z	Link 2f
5b501b4cca0071a2326d3f8b	26-1F5T-ATAC_SIC0808_CTCCAGGGTA_S2.PE.R2.fastq.gz	2018-07-19T00:01:29.000Z	Link
5b501b85ca00711a0c6d3f8c	26-1F5T-redo-ATAC_SIC0801_CGTGGGTTCa_S3.PE.R2.fastq.gz	2018-07-19T00:02:41.000Z	Link
5b501b8684510a00583593f7	30-5F5T-ATAC_SIC0813_CCGAAGTTTA_S4.PE.R1.fastq.gz	2018-07-19T00:02:52.000Z	Link
5b501b8784510a85663593f8	26-1F5T-redo-ATAC_SIC0801_CGTGGGTTCa_S3.PE.R1.fastq.gz	2018-07-19T00:02:05.000Z	Link
5b501bc1ca0071fd226d3f8d	31-6F5T-ATAC_SIC0814_GATTAACCCA_S5.PE.R1.fastq.gz	2018-07-19T00:03:15.000Z	Link

2.2 Metadata organization

For a submission to be marked complete and to be available on the Data Portal, you must fill out several pieces of information about how the files were generated (e.g., how the mouse was treated, how the assay was performed). For production data, complete metadata constitutes all required fields and relationships.

The metadata is organized into discrete categories (such as Mouse, Assay, Reagent) that are linked together. The entity relationship diagram on the Accession Registry Portal home page (<https://meta.targetepigenomics.org/>) displays the relationships between the metadata categories. Some categories will have only one or a few unique instances per lab (e.g., Bioproject), while others (e.g., “Mouse” for individual mice, “Assay” for experimental assays performed on tissue obtained from a mouse) will have many. By storing metadata as unique objects, we can avoid entering redundant data (e.g., multiple mice may link to the same Diet and Treatment).

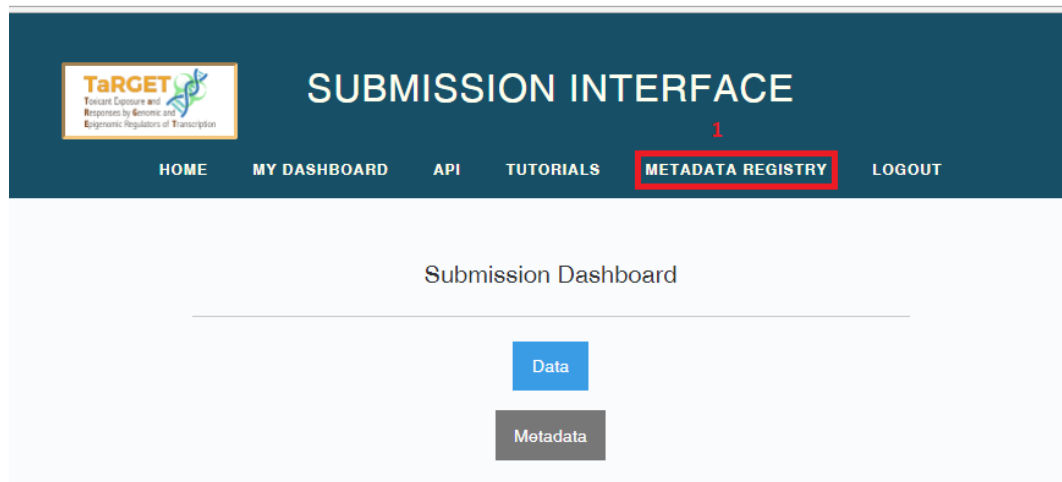


The following instructions can be used to register production metadata in the TaRGET II DCC metadata database. You can use them to: 1. Upload new metadata to the database; 2. Update existing records in the database; 3. Establish relationships between metadata records. You can register metadata one-by-one via the Accession Registry (for Lab, Bioproject, Diet, Treatment, and Reagent) or in bulk via the web UI (for Litter, Mouse, Biosample, Assay, and File). Bulk upload via the command line can be performed by request. Please note that pilot metadata should be updated only through the “Update experiment design” button on the submission dashboard.

2.3 Metadata submission via the Accession Registry

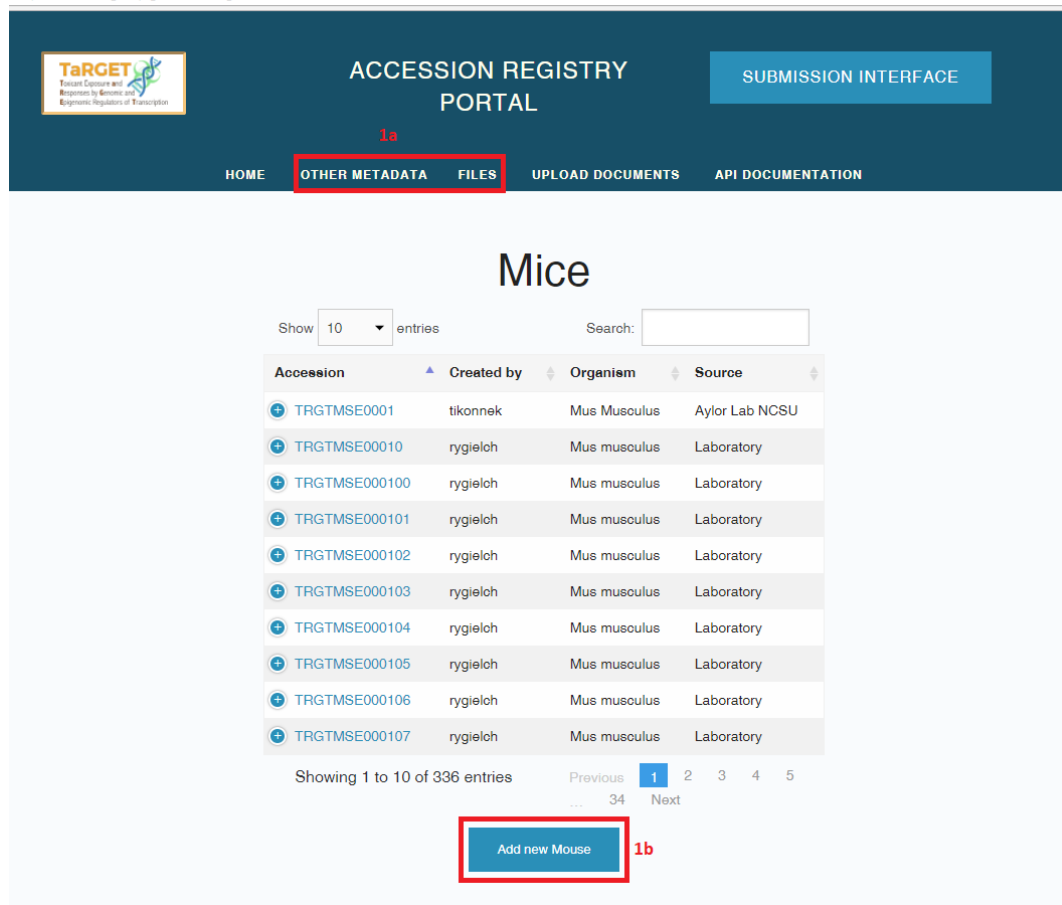
1. To register metadata one-by-one, go to the Accession Registry Portal (<https://meta.targetepigenomics.org/>).

<https://submit.targetepigenomics.org/dashboard>



- (a) Fill out metadata for your samples by clicking on the metadata objects listed under “Other Metadata”.

<https://meta.targetepigenomics.org/mice>



- (b) To create a new metadata object, fill out all of the required fields under the “Add new ___” button.
- (c) Some fields will include a description or a drop-down menu of available terms. After you submit the object (“Create”), a notification will appear that the object was successfully created, and its randomly generated, permanent accession number will become available in the list of current

objects.

* Weight at sac (g)

0.00

* Perfusion

Yes

* Fasted

Yes

Number of hours fasted

(if applicable)

* Technicians

Individuals handling mice


Comments

Create

1c

- (d) To view the details of a metadata object, select the accession number for that object. If an object has already been registered, you should not need to register it again when you submit new metadata; however, you should check to make sure that all of the fields match your requirements.

https://meta.targetepigenomics.org/mice



ACCESSION REGISTRY
PORTAL

SUBMISSION INTERFACE

HOMEOTHER METADATADATAFILESUPLOAD DOCUMENTSAPI DOCUMENTATION

Mice

Show 10 entriesSearch:

Accession	Created by	Organism	Source
+ TRGTMSE0001	tikonnek	Mus Musculus	Aylor Lab NCSU
1d + TRGTMSE00010	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000100	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000101	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000102	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000103	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000104	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000105	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000106	rygieloh	Mus musculus	Laboratory
+ TRGTMSE000107	rygieloh	Mus musculus	Laboratory

Showing 1 to 10 of 336 entries

Previous12345
...34Next

Add new Mouse

(e) To edit a metadata object, alter the relevant fields, and select “Save changes”.

Number of hours fasted

4

* Technicians

ULAM, Christine Rygiel, Kari Neier

Comments

Weight at weaning (g): 9.58

Save changes 1e

- (f) To link a metadata object to another metadata object (e.g., associate a Bioproject with its Lab), select the object from the drop down menu and select “Add”. Links can also be deleted without deleting the object (“X”).

undergoes

This mouse -- undergoes -- treatment :

TRGTTMT0004

x

part_of

This mouse -- part_of -- bioproject :

TRGTBPR0004

1f

x

biological_replicate

This mouse is not associated with any mouse yet.

1g Add mouse associated with this mouse :

TRGTMSE0001

TRGTMSE0002

TRGTMSE0003

TRGTMSE0004

Add

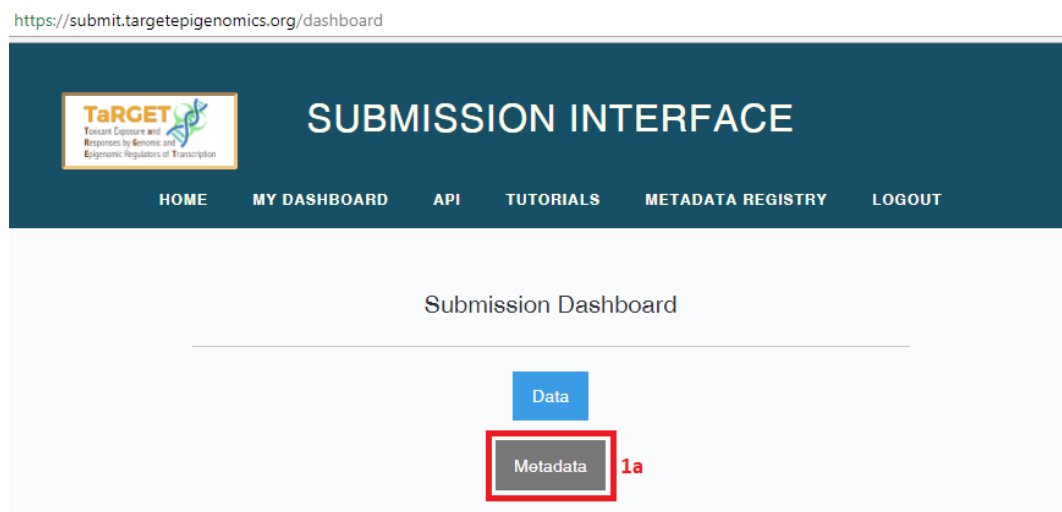
(g) To delete a metadata object, select “Delete _”. All links between the object and other objects must be deleted before the object can be deleted.

Delete Mouse

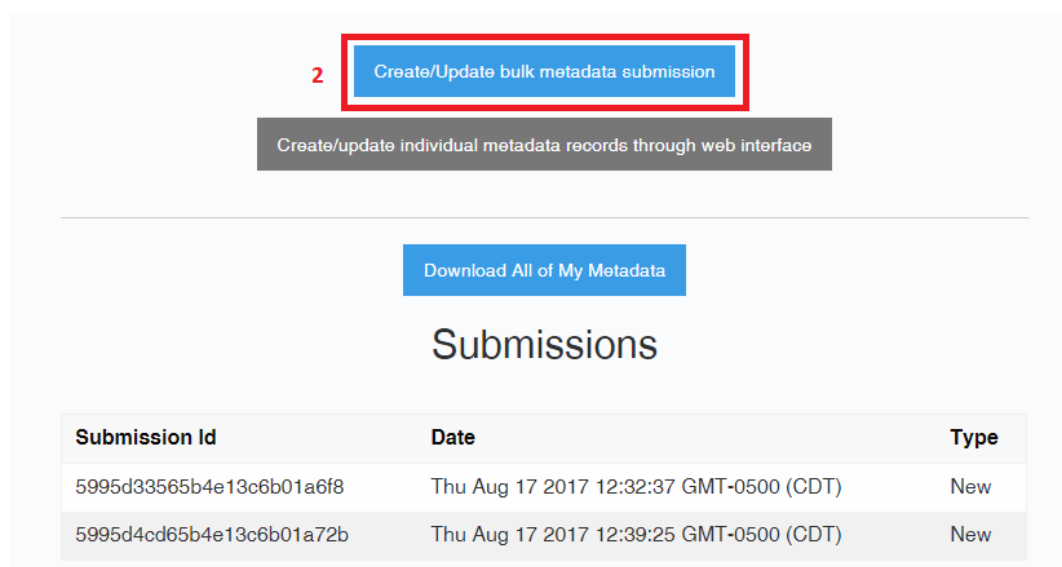
5g

2.4 Metadata bulk submission via the web UI

1. To register metadata in bulk, on your Submission Dashboard (<https://submit.targetepigenomics.org/dashboard>), select “Metadata” to go to your metadata submission dashboard.



2. Select “Create/Update bulk metadata submission” to access the web UI for bulk upload of metadata.



3. Login and follow the tutorial available on this site to download a new blank template or re-download previously submitted metadata into the most recent metadata version.
4. Instructions for filling out the Excel template:
 - (a) All required fields (red headers) and relationships (blue headers) must be populated.
 - (b) Many fields are restricted to a list of pre-determined values available in a drop-down menu. To add additional values to the list, contact the Metadata WG.
 - (c) Link metadata entries together by entering the relevant System Accession in the blue/teal relationship columns. Metadata can be linked to other records already in the metadata database using their System Accession.
 - (d) Enter dates as a string with format “YYYY-MM-DD”.
 - (e) All mouse ages should be specified in weeks.
 - (f) Please use the following missing data fields: Text, “NA”; Number, “999999999”; Date, “1970-01-01”

- (g) All protocol fields should be linked to finalized protocols that are available on the wiki or the metadata registry, including lab-specific protocols.

Thank you for using the TaRGET DCC submission pipeline! Please contact us with any questions.

CHAPTER 3

Report an issue

Please go to our github repo issue page <https://github.com/target2/SubmissionSystem> to report an issue, we will get back to you quickly.

CHAPTER 4

Indices and tables

- `genindex`
- `modindex`
- `search`