

# Phenotype Database V2

---

Short user manual for importing study (meta)data

Mariska Bierkens, Jolanda Strubel



## Contents

1 Introduction .....	3
2 Generic Study Capture Framework (GSCF) section: set up your study .....	4
2.1 Creating a new study .....	4
2.2 Creating subjects .....	4
2.3 Defining samples .....	5
2.4 Assays.....	6
3 Creating platforms and features .....	6
3.1 Creating a platform .....	7
3.2 Creating a feature .....	7
3.2.1 Manually creating a feature .....	7
3.2.2 Importing features .....	8
4 Importing data/measurements.....	8
5 Export .....	9
5.1 Study (meta) data .....	9
5.2 Measurements .....	9
Appendix I: Examples of files for subject, sample and measurements import.....	10
Appendix II: Example of importing features.....	11

# 1 Introduction

In this short manual a walkthrough will be given for creating a study in Phenotype Database.

It includes:

- Defining the study ([section 2](#)).
- Creating platforms and features ([section 3](#)).
- Uploading measurement data and linking them to samples ([section 4](#)).
- Exporting of study meta data and measurements ([section 5](#)).

In Figure 1 the general setup of the Phenotype Database and the relation between the different elements is shown.

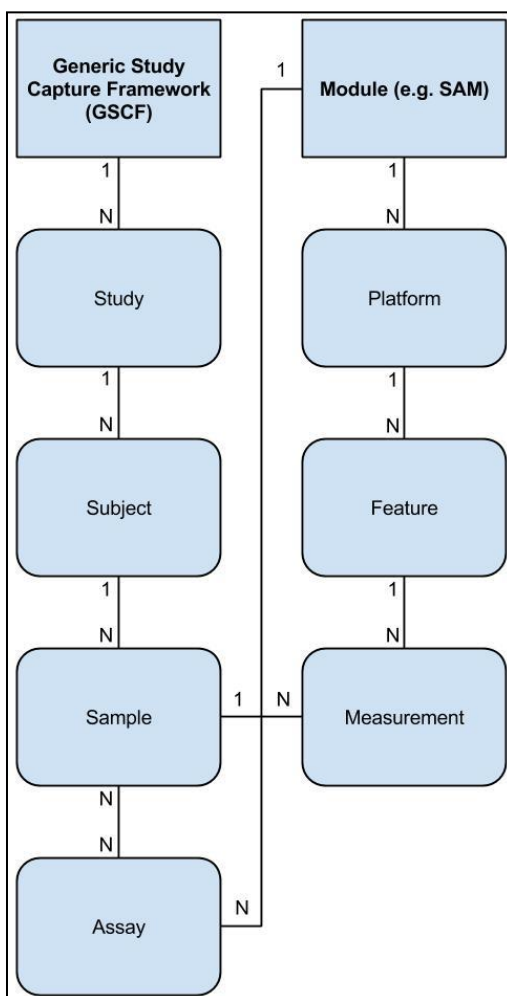


Figure 1: General overview of elements within the Phenotype Database.

One subject (indicated by '1') can have more than one (= N) sample(s) and one sample can have more than one measurement.

Please note before starting with uploading data in Phenotype Database:

- Templates need to be defined in the Phenotype database for all parts of the study. A template contains fields which describe the details for that study part (sample, subject etc). Creating templates beforehand will enable you to directly match data to particular template fields (see

the extended manual for a detailed description of templates).

You can create and edit templates, but only in the TraIT Sandbox environment. Only a 'Template administrator' can edit templates on the TraIT Production environment.

To move your templates from the Sandbox to the Production environment or to change/delete templates on the Production environment, send a request to Topdesk ([direct link](#)).

- Excel files containing the subject and sample details are needed to set up the study.
- Files that are uploaded should have a specific structure to allow matching of column headers to the desired template field (see Appendix I for some examples). When the column names do not match, you can manually select a field from the template.
- The size of the number of measurements (data points) that can be uploaded in one go is limited to 30.000 samples, 80 features and max 1.2 million measurements in total. If you have more data points divide your dataset in multiple datasets and upload these in separate assays.

For a full description of the Phenotype Database go to the extended user guide which will be available soon on the website.

## 2 Generic Study Capture Framework (GSCF) section: set up your study

In this section it is explained how to set up a study, create subjects and samples in the Generic Study Capture Framework (GSCF) section of the Phenotype Database. The subjects and samples are created by uploading files. If you need to setup subjects or samples manually, please refer to the extended user manual.

### 2.1 Creating a new study

1. Go to the menu 'Create' and select 'Create a new study'.
2. Select a predefined template or choose the add/modify option to create a new template.
3. Choose the 'General' Study template in case you are not sure of the fields you need. This template contains the fields 'title', 'description', 'code' and 'start date'. Help text is available under the question mark icons.
4. Complete the required fields (indicated by a green arrow).
5. Click on 'Next'.

### 2.2 Creating subjects

A subject is the source from which your sample is derived. In the 'subjects' step there are two methods to create new subjects: manually or by importing an Excel file containing the subject details. Here only the latter will be discussed.

The Excel file that is uploaded should contain unique subject names.

To import the subject data:

1. On the subjects screen click on 'Import'.
2. You are now on the 'Upload file: subjects' screen:
  - Click on the icon for 'choose a file'.
  - Select the Excel file with subject information.
  - Indicate sheet for 'Use data from sheet' (default value is '1').
  - Indicate row for 'Column header at line' (default is '1').
  - Leave the date format (dd/MM/yyyy) unless it is not appropriate.
  - Select type of the 'subject template' (e.g. human, mouse).
  - Click on 'next'.
3. The 'Match data' screen is shown:
  - Check the data in the overview.
  - Click on 'match' to try and automatically match the columns in the Excel file to similar fields in the subject template.
  - Check whether the matching is correct.
  - If the matching is not correct, you can manually choose another field by clicking on the dropdown menu to manually select another field or set to 'don't import'.
  - If no appropriate subject fields are present, go to 'Browse/edit', 'Templates', 'Subject templates' and click on 'add/modify' to see the contents of the current templates.
  - If the matching is correct, click on 'validate'.
  - If a list with problems appears, scroll through the list to see whether unexpected problems are indicated. 'Error number 7' only means that no value was present and can be ignored if correct.
  - If no changes in the file are needed, click on 'import' at the bottom of the screen.
4. Click on the 'edit subjects for study...' link to return to your study.

## 2.3 Defining samples

Samples can be created in two ways:

- Using the defined events in the design part (see the extended manual).
- Importing the sample details using an Excel file.

In this manual the design step will be skipped and the sample names will be imported directly using an Excel file. The sample names need to be unique in the file.

On the 'Samples' screen in the Wizard:

- Click on 'import'.
- On the 'Upload file: samples' screen select the file containing the sample data.
- The procedure is the same as described for importing subjects (see section 2.2).
- When ticking the box for 'attach to subjects' the data will be linked to the subjects already present in your study. The subject name should be present in the dataset to make this link.

## 2.4 Assays

Assays are used to group measured data for a particular test or experiment. You need to indicate for which samples which tests were performed. In a later phase, the measured data (measurements) are then imported into your assays.

First create your assay(s). On the 'assay screen' in the Wizard:

1. Click on 'Add'.
2. Indicate the number of assays you want to create.
3. Select an appropriate assay template.
4. Give the assay a meaningful name (e.g. the name of the measured biomarker and your study). If more than 1 assay is created the name will be followed by a number (\_1,\_2 etc). This name can be changed afterwards.
5. Complete the required and other appropriate fields.
6. Choose the appropriate module for this measurement (e.g. 'SAM').
7. Click on 'ok'.
8. Change the assay names (if needed) and save the changes.
9. Click on 'next'.

An overview screen appears, showing all sample names and assays in the study:

10. Check the boxes to indicate which samples belong to which assay.
11. To select all samples click on the tickbox top left in the table header. Then click on 'select all items on other pages' at the bottom of the table. Click on the first row to select that assay for all selected samples.
12. Click on 'next'.

## 3 Creating platforms and features

The measurement data are imported in the 'Modules' section of the Phenotype database. To import your data you need a platform (section 3.1) and features (section 3.2).

### 3.1 Creating a platform

Platforms are used to group features that belong together, for instance for a type of technique (e.g. tissue microarray, qPCR etc). Create a new platform if no appropriate platform is present yet or if a version update is required:

1. Select 'Browse' in the appropriate 'Modules' section. An overview of all available platforms is shown.
2. Click on 'Create new platform'.
3. Complete the fields : 'Name', 'Platform version ' (enter '1' when in doubt), 'Platform type', 'Comments' and 'Template'.
4. Click on 'create'.
5. Click 'list platforms' to return to the current module's platform list.

**Please note:** it is possible to change the 'platform type' (=module name) to another than the current module . If you have created a platform but can't find it under the expected module, check another module and change the 'platform type' to move it to the appropriate module.

### 3.2 Creating a feature

A feature describes the measured item (e.g. a biological marker, laboratory parameter etc) and the associated metadata for that item. These are defined in a feature template.

It is important to enter as much information as possible on a feature in order to reproduce or trace back the results (e.g. according to the 'Minimum Information Guidelines' such as MIQE).

Separate features should be present for all types of measurements in your study and should be uploaded to the appropriate platform. You can manually define features in Phenotype Database (section 3.2.1) or import them using an Excel file (section 3.2.2).

#### 3.2.1 Manually creating a feature

1. Go to the 'Module' menu, choose the appropriate module.
2. Select in the 'Browse' menu the option 'Features'.
3. All current features are shown. If no appropriate feature is available yet, select 'create new feature'.
4. Choose the platform for the new feature.
5. Enter a meaningful name (required) and, if applicable, a unit. It is also possible to upload the coding information for a feature in the unit field (e.g. 0=No, 1=Yes).
6. Select a feature template.
7. Complete the fields as much as possible.
8. Click on 'Create'.
9. The new feature is added to the list.

### 3.2.2 Importing features

An Excel file containing the contents of multiple features of the same type of measurement (using the same feature template) can be imported to create features quickly (see Appendix II for an example). Remove all unnecessary rows from the file before importing. Upon importing only columns can be omitted.

1. Choose 'SAM' from the 'module' menu.
2. Choose 'Import', 'Features'.
3. Select the Excel file containing information on the feature details.
4. Choose the appropriate feature template. The feature template is not required, but leaving this empty results will give problems in the next step, as the descriptive fields from the Excel file cannot be matched.
5. Select the appropriate platform.
6. Click on next.
7. Match the headers in the Excel file to the fields present in the selected feature template, either manually or using the 'match' button.
8. Set all columns that do not need to be imported to 'don't import'.
9. Click on 'validate'.
10. Check whether everything was matched correctly. If not, go back and change the file or matching accordingly.
11. Click 'import' when everything is correct.

You can now use these features to add measurements to your study.

## 4 Importing data/measurements

You can import your measured data in the 'Modules' section of the Phenotype database using Tab separated (TSV), Excel or comma separated (CSV) files either in sample or subject layout:

1. Go to the 'SAM' module.
2. Choose from the 'Import' menu 'Measurements (sample layout)' or 'Measurements (subject layout)' depending on the type of file you want to import (see Appendix I).
3. On the 'Upload file: Measurements...' page:
  - Select the appropriate file.
  - Indicate sheet for 'Use data from sheet' (default value is '1').
  - Indicate row for 'Column header at line' (default is '1').
  - Select the appropriate decimal separator used in the file (check in the preview whether the measurements look as expected).

If a message appears indicating that 'added data does not match known number of



columns', this means that not all columns in the file contain values. Click on 'ok' if this is correct.

- Leave the date format (dd/MM/yyyy) unless it is not appropriate.
  - Select the appropriate assay from your study.
  - Select the appropriate platform.
  - Click on 'next'.
4. If the message appears indicating that 'added data does not match known number of columns', click on ok again.
  5. The matching and importing procedure is the same as described for importing subjects (see [section 2.2](#)).
  6. Go to 'Browse', 'Assays' (in the 'Modules' part) to check the import.

## 5 Export

You can export data from Phenotype Database to perform calculations or visualisations or to upload your data into, for example, the data-integration platform tranSMART.

You can export just the meta data of the study or a combination of the meta data and the measurements.

### 5.1 Study (meta) data

1. Go to the GSCF.
2. From the 'Export' menu choose 'Studies'.
3. Select the study
4. "SimpleTox" will give a file with subject and sample-names. In this file you can check whether the samples were linked correctly to the subjects (if applicable).
5. "Combined Excel" gives an Excel file containing the study, subject and sample data including the measurements, on separate tabs.

### 5.2 Measurements

1. Go to the 'Export' menu and choose 'Assay data'.
2. Select your study and the assay(s).
3. Choose the appropriate decimal separator for the output file.
4. Click on 'assay data'. The downloaded file contains all data for all indicated assays.

## Appendix I: Examples of files for subject, sample and measurements import

For subject and sample imports the files should contain unique subject or sample names.

Example 1: a subject import file:

	A	B	C
1	<b>SubjectID</b>	<b>Species</b>	<b>Gender</b>
2	Subject_01	Homo sapiens	Female
3	Subject_02	Homo sapiens	Female
4	Subject_03	Homo sapiens	Male
5	Subject_04	Homo sapiens	Male

Example 2: a sample import file: (note that the subject id/name needs to be present if the samples need to be linked to a subject)

	A	B	C	D	E
1	<b>SubjectID</b>	<b>SampleName</b>	<b>SampleMaterial</b>	<b>sampleMaterial-particulars</b>	
2	Subject_01	Subject_01_TMA	Colon	Primary tumour	
3	Subject_02	Subject_02_TMA	Colon	Primary tumour	
4	Subject_03	Subject_03_TMA	Colon	Primary tumour	
5	Subject_04	Subject_04_TMA	Colon	Primary tumour	

Example 3: a measurements (sample layout) import file:

	A	B	C	D
1	<b>SampleName</b>	<b>SubjectID</b>	<b>EGFR primary CRC core membranous staining frequency score</b>	<b>EGFR primary CRC core membranous staining intensity score</b>
2	Subject_01_TMA	Subject_01	4	2
3	Subject_02_TMA	Subject_02	3	1
4	Subject_03_TMA	Subject_03	3	3
5	Subject_04_TMA	Subject_04	0	1

## Appendix II: Example of importing features

For each feature (measured item) details should be present in the file for the fields present in the feature template used.

For instance: for a “Tissue Micro array” feature additional fields are needed next to name, platform and unit:

- Scoring performed on (cellular localisation)
- TMA core scored
- Note on dichotomisation cut-off

Etc

To import the values for these ‘Tissue Micro array’ features the structure of the Excel file needs be like this:

	A	B	C	D	
1	<b>Marker name</b>	<b>Unit</b>	<b>Scoring performed on</b>	<b>TMA core scored</b>	<b>Note on</b>
2	EGFR CRCLM core cytoplasmic staining frequency	-4: ask pathologist, -3: not representative, -2: No cytoplasm		CRCLM	For each
3	EGFR CRCLM core cytoplasmic staining intensity	-4: ask pathologist, -3: not representative, -2: No cytoplasm		CRCLM	For each
4	EGFR CRCLM core membranous staining frequency	-4: ask pathologist, -3: not representative, -2: No membrane		CRCLM	For each
5	EGFR CRCLM core membranous staining intensity	-4: ask pathologist, -3: not representative, -2: No membrane		CRCLM	For each
6	EGFR primary CRC core membranous staining frequency	-4: ask pathologist, -3: not representative, -2: No membrane		primary CRC	For each
7	EGFR primary CRC core membranous staining intensity	-4: ask pathologist, -3: not representative, -2: No membrane		primary CRC	For each

In this example the Unit field is used to store the coding data.

The column ‘Marker name’ is matched to the feature field ‘name’ in the Phenotype database:

Home Browse Import User Guide Go to GSCF Hello jolanda | sign out

Match data Choose type / Upload file / Match data / Validation / Store data

Match the rows and columns from your file  
Below you can specify where and how to store the data you provided.

Show 10 entries Match Clear mapping Search:

Marker name	Unit	Scoring performed on cellular localisation
name	unit	Scoring performed on subcellular localisation