Phenotype Database Extended User Manual

Jolanda Strubel, Mariska Bierkens, Annemieke Hiemstra

Contents

Definitions

Assay	analysis or test containing a number of assessments (features)
Feature	assessment done for a certain type of test or analysis (e.g. cholesterol level, staining frequency score, gene expression level)
Measurement	result for an assessment (feature) for a certain sample
Modules	grouping of measurements (results) for the same type of assays (analysis)
Platform	grouping of features (assessments) for certain tests and analysis
Sample	material taken from a subject (sometimes at more than one time point) used for analysis or testing
Sample type	type of sample taken from a subject (e.g. blood, DNA)
Subject	source of the samples on which assessments are done (e.g. mice, patients, plants)
Template	predefined set of fields that describe a study element in the Phenotype database
Treatment	the use of an agent, procedure, or regimen, such as a drug or diet, over a period of time

Abbreviations

CSV	Comma-separated value file
GSCF	General Study Capture Framework
PHdb	Phenotype database
SAM	Simple Assay Module
TSV	Tab-separated value file

1 Introduction

The Phenotype database (PHdb) enables you to store the metadata of your study design in the same tool as the actual measurements.
Data can be viewed in the PHdb after importing and can also be exported. Sharing data and comparing the set-up and data of other studies is now possible.
You can grant persons view or write access to your study or even go completely public.

The PHdb consist of 2 sections (Figure 1.1):
1. "Generic Study Capture Framework" (GSCF): in this part the description of a study design is stored. This includes the subject data, sample data and other study details.
2. "Modules" part: in this part the actual measurements are imported and stored in so called features.
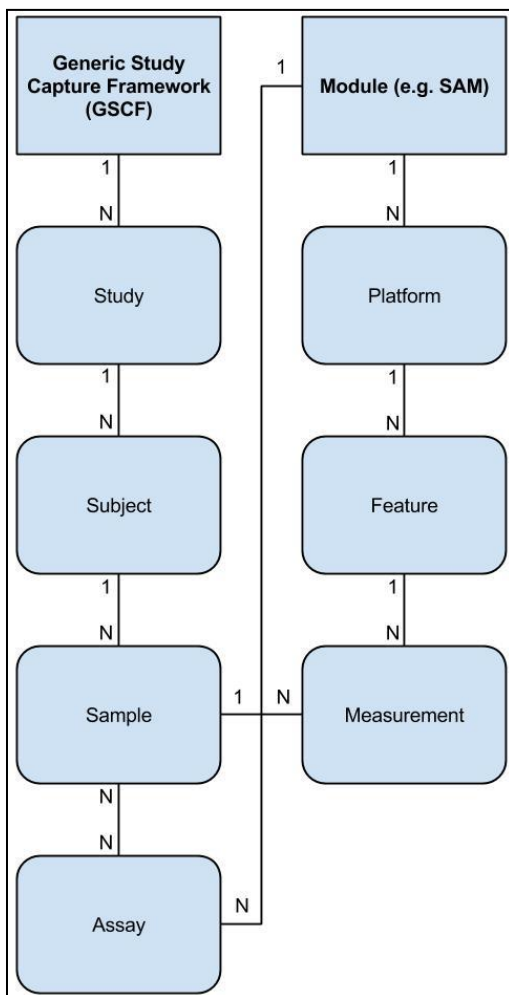
Figure 1.1: General overview of elements within the Phenotype Database.

As indicated in Figure 1.1, one subject can have more than one (N) sample. Each sample can be assigned to one or more assays (different experiments). The measurements (results of the tests) are indirectly linked to the assays in the Module part via platforms. In the current release of PHdb only the Simple Assays Module (SAM) is present.

The PHdb uses so called "Templates" to store all details on the different study elements (study details, subject data fields etc). See Chapter 2 for more details on templates.
The set-up of a new study in the GSCF, using the "Study Wizard", is described in detail in Chapter 3.
The measurements are imported in the Small Assay Module (SAM). They are grouped in so called platforms which contain features that are used to store the actual data. The measurements are linked to the study data in GSCF by the assays (see Chapter 4).

2 Templates

The PHdb uses customizable templates and ontologies to store the metadata of a study. This allows for flexibility to capture all information on your study. Templates are used in all elements of the PHdb.

Create your templates before setting up your study. This will enable you to directly enter data in the template fields.
First create your templates in the Sandbox 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 (trait.topdesk.net) by using the respective forms:
- Go to "diensten en aanvraag formulieren"
- Go to "Experimental data", "Phenotype database"
- Choose the "PheDB Template New, Import or Modification in Production" from the forms indicated on the right

Only a "Template administrator" can add or modify templates in the Production environment.
In section 2.1 the possible fields for each type of template are discussed; section 2.2 describes how to create and modify templates.

2.1 Template contents

Templates are present both in the GSCF and Modules part of the PHdb. A template contains default standard fields and can be extended with additional fields that can be created to fit your study needs. The following information is captured in the different types of templates:
- **Study (GSCF)**: fields in this template describe the study characteristics. Standard fields within the template are: title, description, code, startdate and published. Other fields could include: study aim, study protocol, participating institute(s) etc.
- **Subject (GSCF)**: fields in this template describe subject characteristics. Standard fields are name (for the subject identifier) and species. Other fields that can be added are for instance gender, age and weight. It is not allowed to store person identifying data in the PHdb, meaning a person's name, address and in most cases date of birth. Exceptions for date of birth are possible, for instance in studies on babies.
- **Sample (GSCF)**: fields in this template describe sample characteristics. Standard fields within the template are: name (as in sample name/ID) and material. Additional fields could include time-point etc.
- **Event/Treatment type (GSCF)**: fields in this template describe treatment type characteristics. The only standard field is the template name. Other types of field that could be added are duration or type of diet etc.
- **Sampling event/Sample type (GSCF)**: this template describes the sample type details. The standard fields are (template) name and sample type. Other types of field could include sample volume, storage conditions etc.
- **Assay (GSCF)**: fields in this template describe assay characteristics. Standard fields within the template are: name and module (here only "SAM"). For most assays these fields are sufficient.
- **Platform (Modules)**: fields in this template describe platform characteristics. Standard fields within the template are name, platform version, platform type (=module) and comments. For most assays these fields are sufficient.
- **Feature (Modules)**: fields in this template describe the characteristics of the assessments that were done. The standard fields are template name and unit (the latter is not required but should be entered wherever applicable). For each type of assessment in your study, you need to define a separate feature template and specific feature fields. For example for the feature "Weight" you could create the template fields "Instrument used", "Time of day", "Shoes on or off". These details will make your method reproducible. For molecular profiling experiments the Minimal Information Guidelines should be adhered to.

2.2 Create/Modify Templates

Templates can only be modified in the Sandbox environment or by a template admin.
There are two ways to create or modify templates:

1. Go to "Browse/Edit" (in GSCF), choose "Templates" and choose the appropriate template.
2. Choose the option "Add/modify" from the dropdown list for a template while creating a study (Figure 2.1).

Figure 2.1:Create or change a template via the "Add/modify" functionality for templates in the study setup wizard.

Both methods will take you the same screen (Figure 2.2).

Figure 2.2: Change an existing template or create a new one.

To create a new template
- Click on "Create new template" at the bottom of the list (Figure 2.2)
- Enter a name and description for the new template and click on "Save".

The new template is now added to the list.

To modify the (new) template:
- Click on the "Add/remove template fields" icon for that template.

A number of standard fields are present in a template which can't be modified (indicated in yellow):
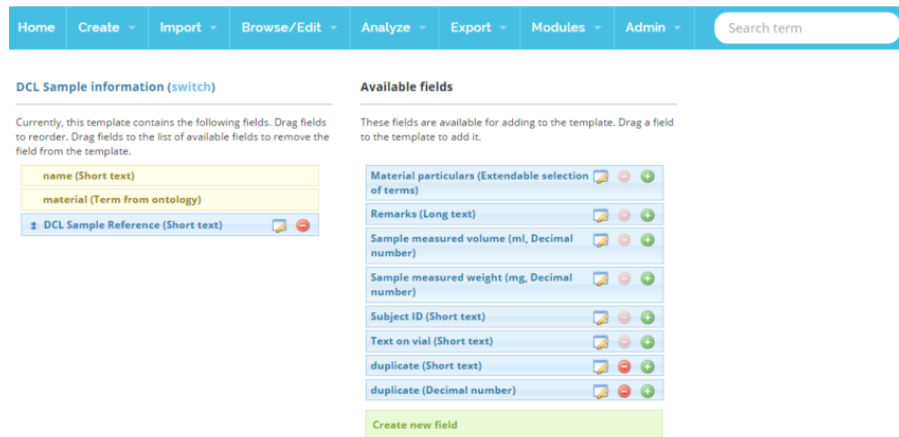

Figure 2.3: Add/remove template fields for a template.

On the left side you see the fields present in this template, on the right side the fields available for this type of template.

To add an existing field:
- Click on the green plus sign behind the appropriate field. The field will now be added to the template and shown on the left.

To remove a field from the current template:
- Click on the red minus sign for the field in the current template (left hand side of the screen). You can only remove a field from a template if no data has been entered in this field for any study using this template. The minus sign will be grayed out as soon as data is present.

To remove a general field from a template type:
- Click on the red minus sign for the field on the right hand side of the screen. This option is only available for fields that are not used in any template (Figure 2.3.).

To change the settings of a field:
- Click on "Show template field properties" (folder icon) for that field.
- For fields already in use in other studies, only the comments can be changed (Figure 2.4, all other options are greyed out). The change you make is valid for all studies using this template field.
- The option "Required" can only be changed if the field is not in use in any study as yet. If you need a field to be required but it is in use as not required in other studies, you need to create a new template field with another name (for instance "Template-name_req").
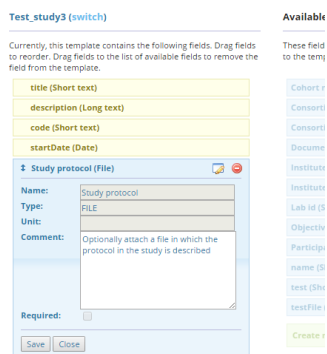- Click on Save.

Figure 2.4: Change template field settings.

To create a new template field:
- Click on "Create new field" at the bottom of the "Available fields" list (see figure 2.3). The screen to create a new field appears (figure 2.5).
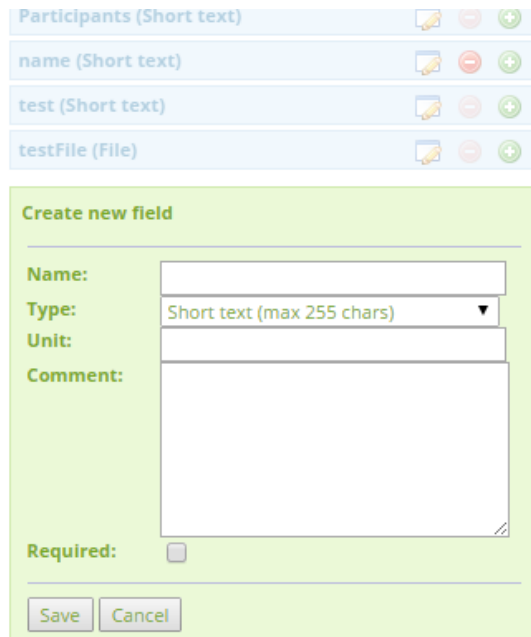
Figure 2.5: Create a new template field.

- Enter a unique name for the new field (add a descriptive suffix if a name is already in use).
- Click on the dropdown list button to show the possible field types.
- The unit field is optional (and in most cases not applicable).
- Text entered in the "Comments" field will be visible when clicking on the question mark icon next to that field.

The following types of template fields are available:

Text:
- *Short text (max 255 characters)*: if a longer text entered is entered, only the first 255 characters will be stored
- Long text (unlimited number of characters): for text longer than 255 characters.
- Dropdown selection of terms: indicate the needed response options (for instance yes and no). Only these values will be possible for this field. If a dataset that is imported contains another value, the value will not be imported and left empty.
- Extendable selection of terms: choose this option when you do want response options, but are not sure of the exact responses present for this field. Every response not present in the response options will be added automatically to the dropdown list.

Numerical:
- Decimal number (1.31): in the PHdb the decimal sign is a dot. Upon importing or exporting data a decimal sign can be chosen.
- Natural number (100): integer numbers only (no decimals).

Other:
- Term from ontology (A term that comes from one or more selected ontologies): **Please note:** this option is not functioning properly yet.
  Choose one of the ontologies already present in the list or click on "Add new". After clicking on "Add new" a search screen for ontologies present in the PHdb appears (see Figure 2.6).
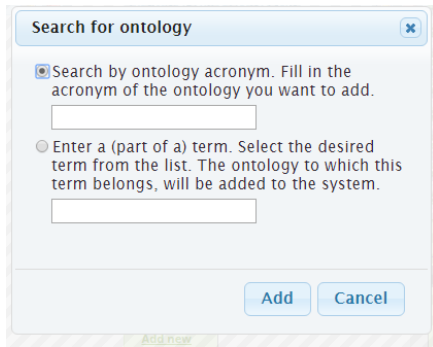
Figure 2.6: Search ontologies.

- If you know the name of the ontology you need use the top search field. To search the complete list enter a space.
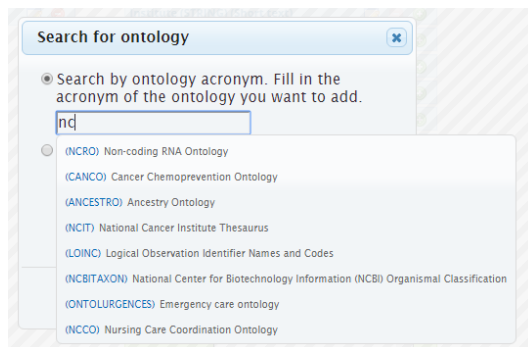
Figure 2.7: Search on ontology name.

If you want to search by concept, use the bottom search field.
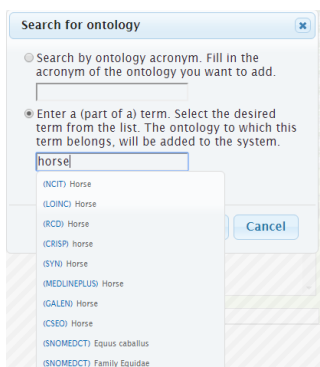
Figure 2.8: search by concept.

Select the appropriate ontology and click on "Add". The ontology is added to the field. You can add several ontologies to a field.

To remove the ontology from a field, select the ontology in the field and click on "Remove".

Figure 2.9: Remove ontology.

- File: option to upload a file and attach it to the template.
- True/false: this will result in a checkbox on screen.
- Template: dropdown list of the available templates for that study part. Not in use.
- Omics module: dropdown list of available modules in the PHdb.
- Date:
    - Date (2010-01-01): the date can be entered using a calendar or as yyyy-mm-dd. An example of the entered date will be shown to check whether the correct date was entered.
   - Relative time: not in use.

After choosing the file type and completing the unit and comments field (if appropriate), click on "Save" to save the field settings.
The new field is now added to the "Available fields" list for this type of template and can be added to your template.

It is possible to add fields to templates even after data have been entered in that particular study part using the template. Data need to be entered manually for the new fields in the applicable study part.

Click on "Switch" to return to the overview of the available templates.

Figure 2.10: Click on "Switch" to return from a template details screen to the templates overview.

To get an overview of which fields are currently present in which template, click on compare:

Figure 2.11: Compare template contents for a certain template type.

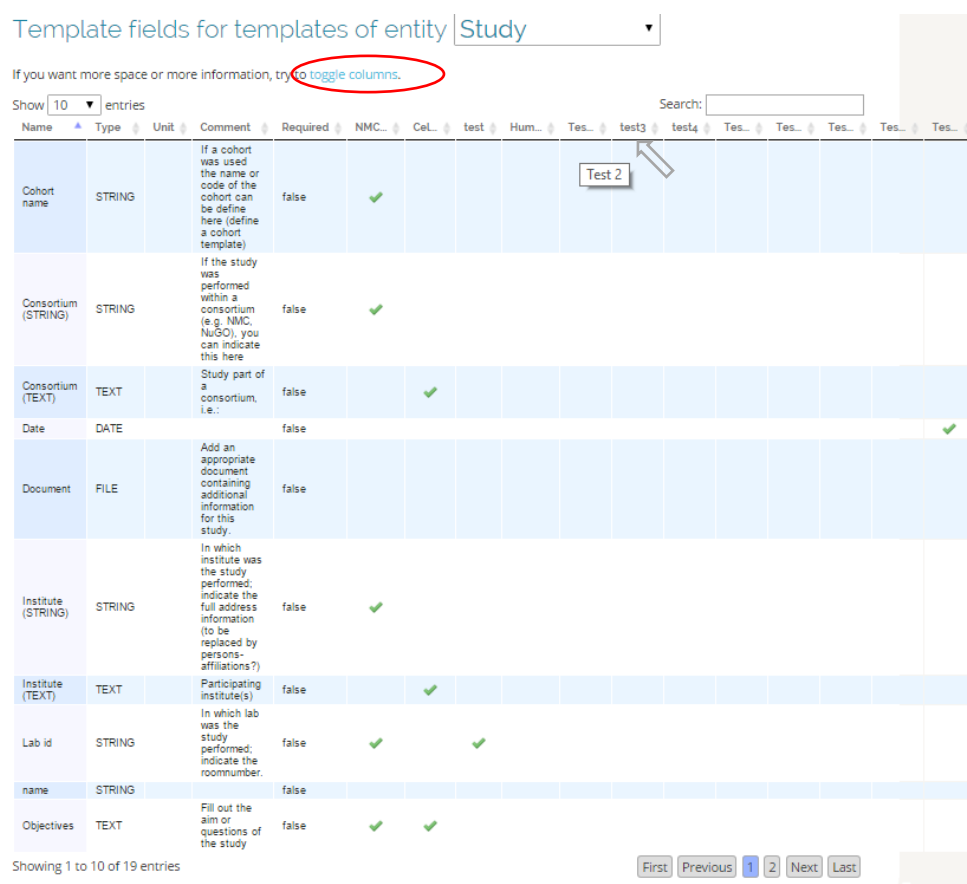By hovering over a template name, the complete name will be shown:

Figure 2.12: Toggle columns to hide part of the columns

Click on "Toggle columns" to hide the general field columns.
Choose a different type of template at the top to go to the overview of those templates.

3 Creating/editing a study

Creating a new study and editing an existing study is done in the same "Wizard" environment.
You can save your study settings and leave the wizard at any point and return to finish the setup (see also section 3.6 on editing an existing study). You can also jump to a next step in the wizard if it is indicated in blue. Click on that step at the top of the screen (Figure 3.1).
Before you create a new study you need to create the appropriate templates or know exactly what should be in them (see Chapter 2).

3.1 Step 1: Study details

To create a new study,

- Go to the "Create" menu and select "Create a new study".
- The wizard is opened on the first step: define your study details. Only the "template" field is shown.
- Choose the appropriate template or click on "Add/modify" to create/change a template.
- Extra fields will appear depending on the template you choose. Here, "Cell line use case" will be used as an example (Figure 3.1).

Figure 3.1: Step 1 of creating a study: entering the study details.

- At the top of the screen the current step of the study set-up is shown.
- Enter the required fields (indicated with a green arrow). Click on the question mark icon for more information on the expected content for a field.
- The title and code need to be unique within PHdb.
- Add at least one contact person for this study. You can add a new person and role if needed.

- User access: you can authorize persons with a PHdb account to view your study (readers and reader groups) or to view and edit your study (writers and writer groups).
  If a person is not present in the list, request an account:
  -go to Topdesk, "Diensten en aanvraag formulieren" "Experimental Data", "PhenotypeDB".
  -for the sandbox environment use the form "PheDB request new user Sandbox environment"
  -for a production account use "PheDB request new user production"

  A user can also request access to your study on the production environment using the form "PheDB Study view and/or edit access request". Requests will always be verified with the study owner (listed in the "Contacts part of the study screen").
- Ticking the "Public" box will make your study available on the Phenotype home page for everyone (no PHdb account is needed).
- Click on next to save your settings and go to the next step. You can always return to the "Study" screen in the wizard and make changes.

Clicking on "Save" will result in leaving the wizard and redirection to the "Study list".

3.2 Step 2: Subjects

The next step is to add subjects and subject details to your study. A subject is the source from which your study material(s) are taken. This could be human subjects, but also animal subjects.
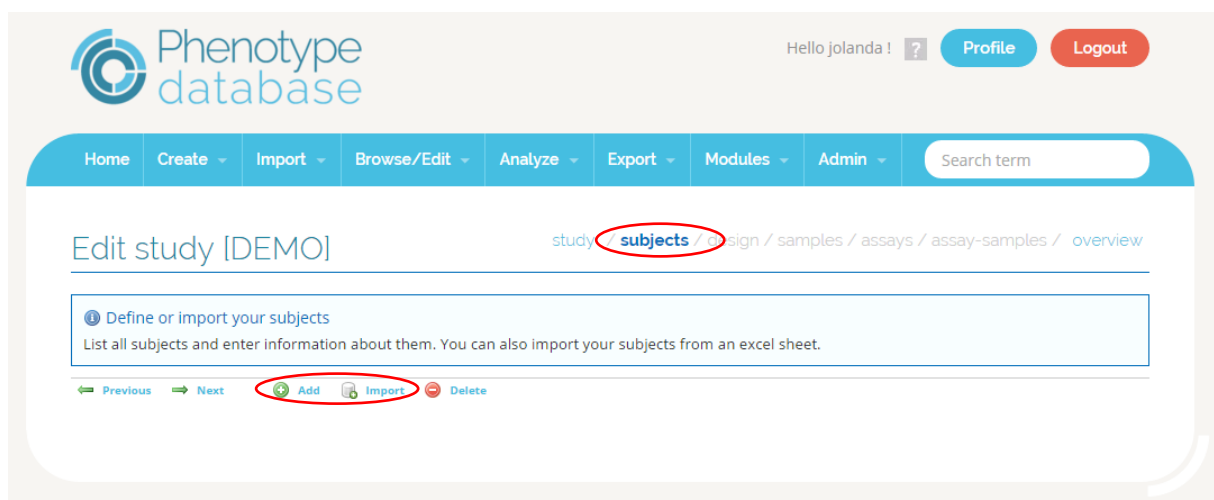
Figure 3.2: Creating subjects

You can add subjects in two ways: manually or by importing a file containing the subject details.

3.2.1 Adding subjects manually

To add subjects manually:
- Click on the "Add" icon at the bottom of the screen.
- Enter the number of subjects you want to create (default is 1).
- Choose the appropriate template (or "Add/modify" a template).
- Enter the name of the subject(s).
  If you are creating more than one subject, this name will be used for all subjects and followed by a number (e.g. if you enter the name "Subject" this will result in subject_1, subject_2 etc).
- Complete all required fields (depends on the template used). The entered values will be used for all subjects. The values can be changed later on.
- If a field uses an ontology (for instance "Species", which is a standard field for the subject template), but the value you need is not present, add it by choosing "Add more" at the bottom of the dropdown list (Figure 3.3).
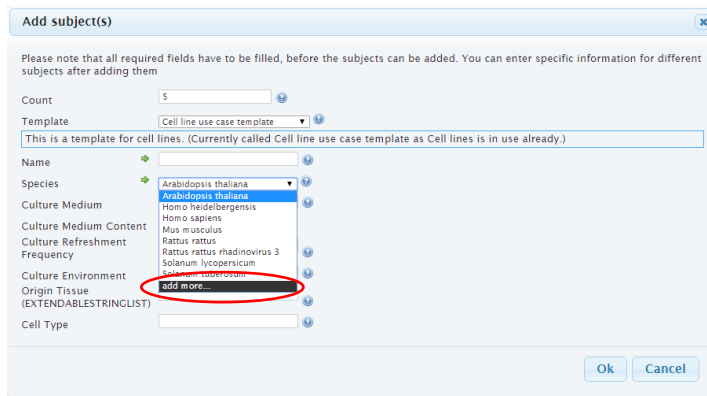
Figure 3.3: Add a value from an ontology.

A new screen opens to enter a search term. All concepts for the ontology available for this field are shown.
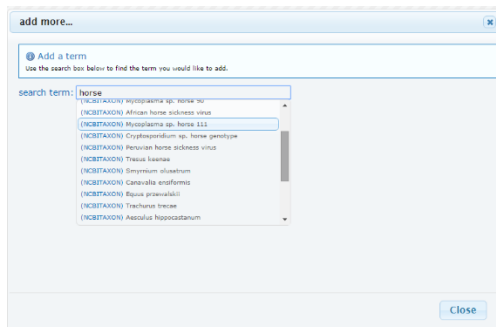
Figure 3.4: Search for a term in the ontology.

Select the appropriate term and click on "Add term". You can add multiple terms this way. Click on close when you have finished.
The new item has now been added to the template field. These new entries can be seen by all studies using this template field.

- Click on "Ok" to add the subjects.

The subjects are now added to your study and present on the "Subjects" screen.
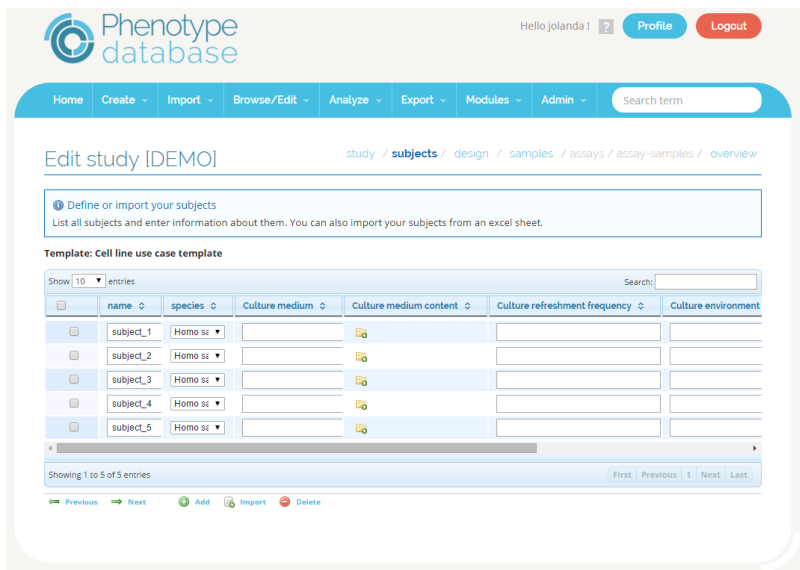
Figure 3.5: Current subjects present in study.

You can change or add data in the table. Changed fields will be indicated green.
Click on "Save" at the bottom of the table (Figure 3.6). If you click on the next button without saving, the changes will be discarded without warning.
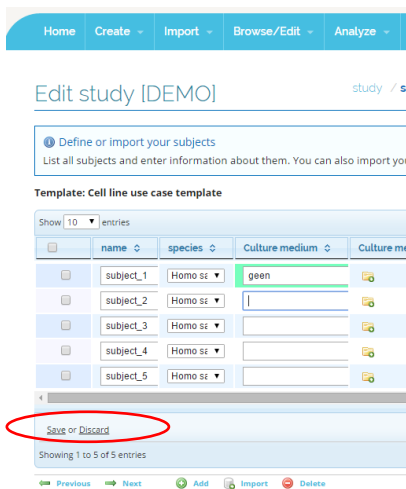
Figure 3.6: Save changed data in the subject table.

3.2.2 Importing subject data

You can also add subjects by importing an Excel, CSV or TSV file containing the subject data. Only one row per subject should be present and the subject names should be unique within the file and your study.
Don't use Excel files for large datasets (>5000 subjects), but convert it to a CSV or TSV first.

To import your subject data:

- Click on "Import" at the bottom of the "Subjects" page. The "Upload file: subjects" screen appears (Figure 3.7).
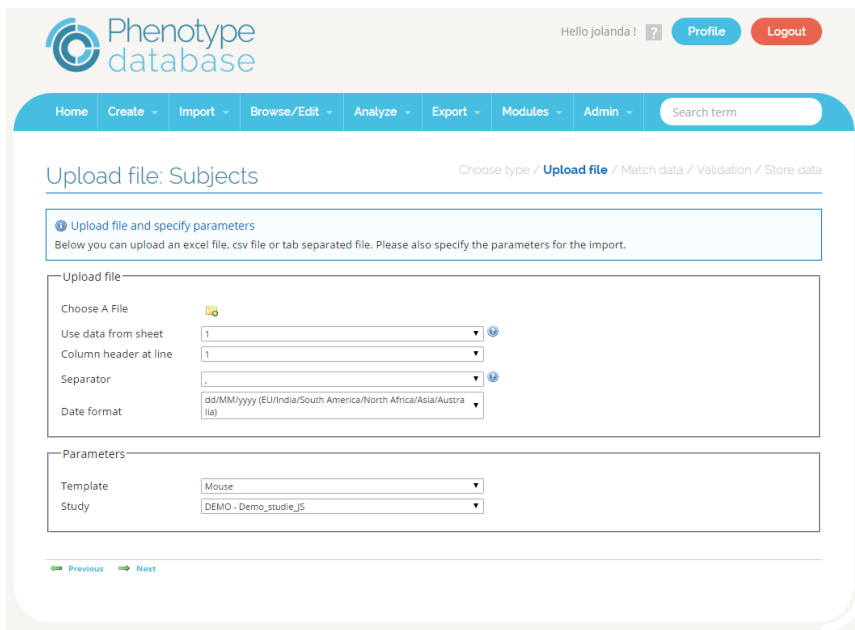
Figure 3.7: the "Upload file: subjects" screen.

- Click on the icon for "Choose a file".
- Select the file with subject details.
- A preview of the data in the file is shown on the screen.
- Indicate sheet for "Use data from sheet" (default value is "1").
- Indicate row for "Column header at line" (default is "1").
- If you are uploading a CSV or TSV file, select the appropriate separator: a comma or semicolon for a CSV (check your file) and tab for a TSV.
- Leave the date format (dd/MM/yyyy) unless it is not appropriate.
- Select type of the "Subject template" (e.g. "Human subject template"). Click on "Next". The "Match data" screen is shown (Figure 3.8).
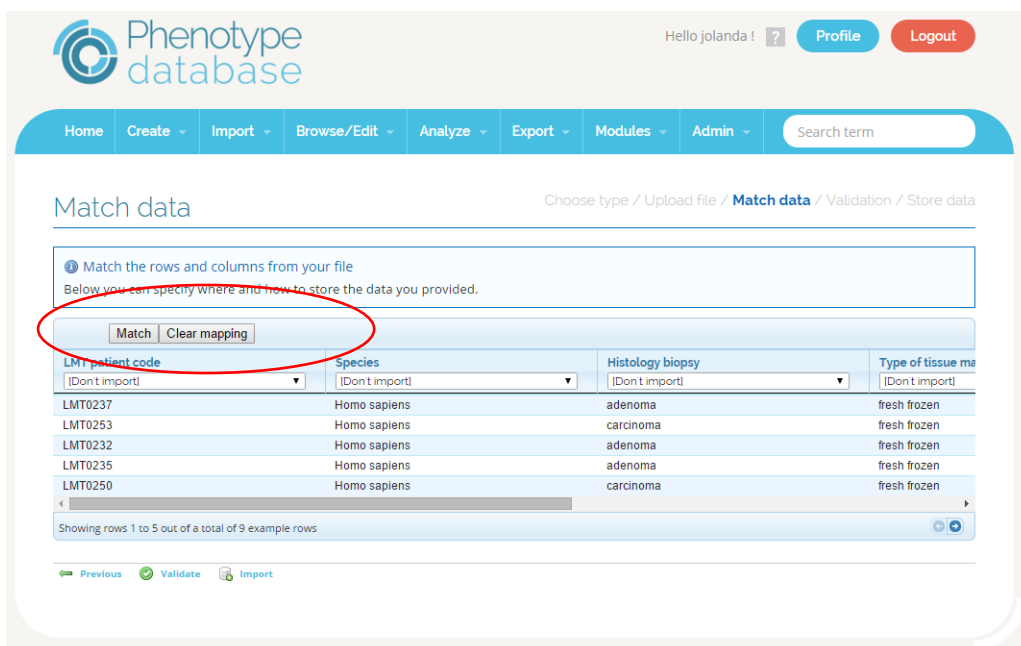
Figure 3.8: "Match data" screen.

- Check the data in the overview.
- Click on "Match" to automatically match the columns in the file to the fields in the template. The names of the columns in the file are shown at the top of the column.
- Check whether the matching is correct.
- If the matching is not correct, manually choose another field in the PHdb from the dropdown menu or set to "Don't import". The column will then be skipped when importing the data.
- 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.
- A message indicating how many rows have been imported will appear when the import has finished. The import of large datasets (>5000 subjects) can take a while. Don't use Excel files for such large datasets.
- Click on the "Edit subjects for study" link to return to the wizard (Figure 3.9).

Store data

Choose type / Upl

✓ Data imported succesfully
19 line(s) from your file have been succesfully imported and no errors were found.

You can choose to: view subjects for study Demo_studie_JS, edit subjects for study Demo_studie_JS or restart the importer.
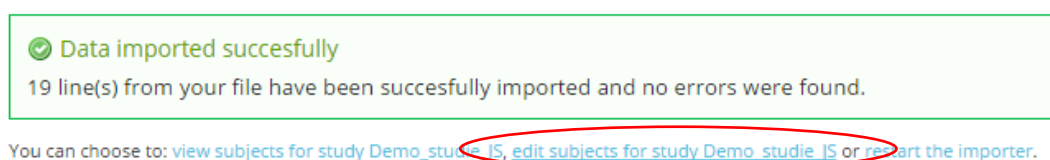
Figure 3.9: Subjects successfully imported.

The subjects are added to the study and shown in the subject screen.

3.3 Step 3: Study design

This step is not required. If you only have one time point on which one sample is taken or if the time aspect of the measurement is not important for your experiment, skip this part and continue with "Step 4: Samples" to import your sample names.
In the design step you can specify the details of your study design if applicable, e.g. treatments and their duration, samples taken at different time points for various study groups.
You need to define at least one "Subjectgroup" (section 3.2.1) and one "Sample&Treatment group" (section 3.2.2). You then indicate for each Subjectgroup which "Sample&Treatment group(s)" were used.

3.3.1 Creating a subject group

To create a Subjectgroup:
- Click on "Add a Subjectgroup"
- Enter a name for the group and select the subjects belonging to this group.
  Click on the checkbox in the top left corner of the table to select all subjects on this page. Click on "Select all items on other pages" at the bottom of the table to select all subjects in your study.
- Click on "Ok" to save.

The new group will be shown in the "Timeline" at the top of the design screen.
Repeat this process to create all groups you need.
To open a Subjectgroup and change the settings, double click on the name of the group in the timeline.

3.3.2 Creating a Sample&Treatment group

A "Sample&Treatment group" contains sample and treatment definitions that belong together. If your study is designed in blocks of repeated or interchanged treatments, you can recreate those details here.
To create a "Sample&Treatment group":
- Click on "Add new"
- Enter a descriptive name (e.g. placebo)
- Click on "Save name"

The "Sample&Treatment group" is added to the design screen.
Click on "Edit" for this "Sample&Treatment group" to add samples and treatment definitions. An overview is shown of all the available treatment and sample types in the current study.
Any changes in the sample or treatment types will affect all instances within the study (not just the current "Sample&Treatment group".

To add a new treatment type:
- Click on "Add new" under "Available treatment types".
- Choose the appropriate template.
- Enter the required fields and click on "Save name".

The new treatment type is now added to the overview.

To add this treatment to the current "Sample&Treatment group":
- Click on the name of the group and drag it to the timeline on this screen. This timeline represents the relative time point within the current "Sample&Treatment group".
- Drop the treatment on the desired (relative) timepoint.
- Indicate the treatment duration and click on "Ok".

The treatment is now added to the timeline of the current "Sample&Treatment group".
Adding a sample type follows the same process, except no duration needs to be indicated.
In section 3.2.3 the functionality of the timelines are explained.

Click on "Save name" to save the current "Sample&Treatment group" settings and to return to the design screen.

3.3.3 Using the timelines

The timelines in both the design and the "Sample&Treatment group" screen behave the same. To manipulate the timeline or any elements (samples, treatments or Sample&Treatment groups):
- Add an element: Click on the name of the element you want to add and drag it to the timeline. Drop it at the desired timepoint.
- Move an element: Double click to select the element on the timeline. Drag this element to another time point and drop.
- Delete an element: Double click to select the element on the timeline. Click on the red cross to remove.
- Zoom: To change the time window shown on the timeline put your cursor on the timeline and use your scroll wheel.
- Scroll left or right in timeline: Click on the timeline itself and drag to scroll left or right.

3.3.4 Finishing your design

Add the "Sample&Treatment groups" to the appropriate subject group(s) in the timeline. Check your design before going to the next step.
Please note that the timepoints indicated inside the "Sample&Treatment group" will be relative to the timepoint indicated in the design screen for the "Sample&Treatment group". So in case a sample is indicated at "1 day" in the "Sample&Treatment group", but the group is placed at "2 days" in the design screen, the actual time point of the sample will be "3 days".

3.4 Step 4: Samples

There are two ways to create samples for your study:
- Automatically, using the design set-up created in step 3.
- Importing sample names from a file.

These methods are explained in detail in the following sections.

3.4.1 Automatically generated sample names

PHdb can create sample names based on the the study design set-up in step 3 of the wizard. This way you are certain that all possible samples are present in your study.

To create sample names:
- Click on "Generate samples" at the bottom of the table.
- The sample name is always structured as <subject name>_<Sample&Treatment group name>_<sample type name>_timepoint.
- The timepoint is the combined timepoint, adding up the timepoint indicated in the "Sample&Treatment group" and the timepoint of the complete group on the design screen.

If you have to make changes to the design and need to change the sample names accordingly, choose "Regenerate sample names" to incorporate the changes.

3.4.2 Importing sample names

You can import the sample names using an Excel, CSV or TSV file containing all names. You could also use the file containing your measurements for those samples. The sample names however need to be unique in the file.

To import samples:
- Click on "Import" on the "Samples" screen in the wizard.
- 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 3.2.2).
- When ticking the box "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 and matched to the subject name for this link.

3.5 Step 5 and 6: Assays and Assay/Samples

Assays are used to group results for a particular test or analysis. Create a separate assay for each sample type you defined in de Design step (if applicable).

Next you need to indicate for which samples which assays were performed. In a later phase, the results (measurements) are imported into your assays.

To create an assay :
- Go to "Assays" in the wizard.
- Click on "Add".
- Indicate the number of assays you want to create.
- Select an appropriate assay template.
- Give the assay a descriptive name (e.g. the name of the measured biomarker but also part of your study name. This will make it easier to identify when uploading measurements).
  If more than one assay is created the name will be followed by a number (_1,_2 etc). This name can be changed afterwards.
- Complete the required and other fields (depends on the template used).
- Choose the appropriate module for this measurement (e.g. "SAM").
- Click on "Ok".
- Continue until you have defined all required assays.

To link your assays to the samples:
- Click on "Next".
- An overview screen appears, showing all sample names and assays in the study.
- Check the boxes to indicate which samples belong to which assay.
- 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.
- Click on "Save".

Saving the assay-sample settings can take a long while, depending on the number of samples. If you have more than 5000 samples, select one assay at a time for all samples and save.

3.6 Step 7: Overview

In the final step of the Wizard an overview is given of the study parameters for your study. This includes the number of subjects, subject groups, treatment and sample types, samples and assays. If you see any strange numbers, click on the corresponding step in the wizard to return to that step.

Your study set-up is now ready and you can continue with importing your measurements.

3.7 Editing/viewing existing studies

To change or view the settings of existing studies, go to either "My studies" (listing the studies created by you) or "All studies" (listing all studies you created, plus studies for which you have reader or writer access). Depending on your access rights for a certain study, next to the "View" icon an "Edit" icon will be available. Click on the appropriate icon to open the study. The wizard opens on the "Study" step. Click on the step to directly navigate to those settings.

4 Modules

The experimental data (measurements) are captured in the so called "Module" section. In this case the "Simple assay Module" (SAM) is present. On the start page of a Module a short introduction is given of the different Module functionalities (Figure 4.1).

Figure 4.1: Module "SAM" start page.

A module contains a number of platforms. A platform groups together features belonging to a certain type of experiment (e.g. Tissue Micro Array (TMA) or Multiplex ligation-dependent probe amplification (MLPA) etc). Features are assessments for a certain analysis performed on your samples.

This chapter describes the creation of platforms (section 4.1) and features (4.2) and importing measurements (4.3).
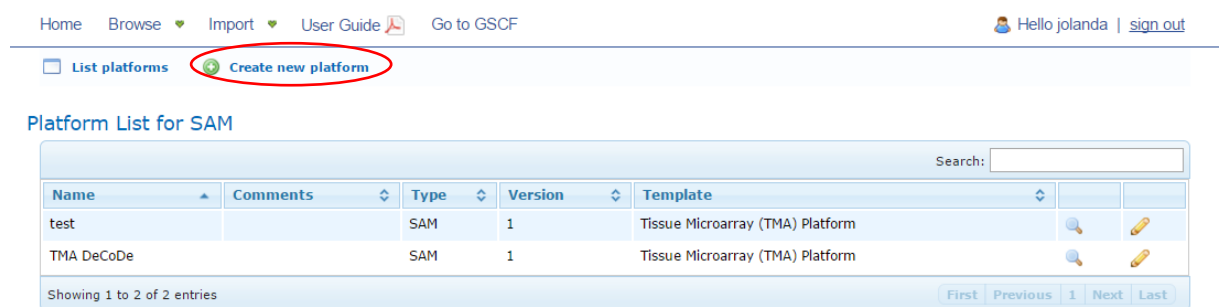
As in most cases the platforms will be created in the system, the functionality for importing them will not be explained here.

4.1 Platforms

Platforms are used to group features that belong to the same type of experiment (e.g. tissue microarray, qPCR etc). Type part of a platform name into the search box or go through the table using the buttons at the bottom of the table.

Create a new platform if no appropriate platform is present or if a version update is needed:
- Go to "Modules".
- Select the appropriate module ("SAM").
- Select "Browse".
- An overview of all available platforms for this module is shown.

Home Browse ♥ Import ♥ User Guide 🔺 Go to GSCF					👤 Hello jolanda	sign out

☐ List platforms   ⊕ Create new platform

Platform List for SAM

Search: [   ]

Name ▲	Comments ⇕	Type ⇕	Version ⇕	Template ⇕		
test		SAM	1	Tissue Microarray (TMA) Platform	🔍	✏️
TMA DeCoDe		SAM	1	Tissue Microarray (TMA) Platform	🔍	✏️

Showing 1 to 2 of 2 entries

First | Previous | 1 | Next | Last

Figure 4.2: Overview of available platforms in the SAM Module.

- Click on "Create new platform".
- Complete the fields : "Name", "Platform version " (enter "1" when in doubt), "Platform type", "Comments" and "Template". The combination of the platform name and version number needs to be unique in the PHdb.
- Click on "Create".
- 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 (sandbox only). 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.

4.2 Creating features

A feature describes an assessment or measurement(e.g. expression level of a biological marker, laboratory parameters such as pH etc.) and any associated metadata. The metadata is predefined by the feature template used.

Enter as much information as possible on a feature in order to reproduce or check the details (follow the "Minimum Information Guidelines", such as e.g. MIQE, as much as possible when applicable).

Separate features should be present for all types of assessments in your study and should be created in the appropriate platform.

Browse the features currently available in the module using the search box or the buttons at the bottom of the table. If no appropriate feature is available, create a new feature either manually (section 4.2.1) or import multiple features using an Excel file (section 4.2.2).

4.2.1 Manually creating features

To create features one by one:
- Go to the "Module" menu.

- Choose the module "SAM".
- Select in the "Browse" menu the option "Features".
- All current features are shown. If no appropriate feature is available yet, select "Create new feature".

Figure 4.3: overview of available features in the SAM Module.

- Choose the platform for the new feature.
- Enter a descriptive name (required) and, if applicable, a unit.
  It is also possible to enter coding information for a feature in the unit field (e.g. -4=Ask pathologist, -1 Not evaluable, etc.).
- Select the appropriate feature template.
- Complete the fields as much as possible.
- Click on "Create".

The new feature is added to the list.

4.2.2. Importing feature details

An Excel file containing the details of multiple features with the same feature template can be imported to create multiple features (see Appendix I for an example of such an Excel file).
Remove all unnecessary rows from the file before importing. Upon importing only columns can be omitted.
To import the features:
- Choose "SAM" from the "modules" menu.
- Choose "Import", "Features". Or go to "Browse", "Features" and click on "Import features" (Figure 4.3)
- Select the Excel file containing information on the feature details (Figure 4.4).

Upload file: Features   Choose type / **Upload file** / Match data / Validation / Store data

ⓘ Upload file and specify parameters
Below you can upload an excel file, csv file or tab separated file. Please also specify the parameters for the import.

```
┌─Upload file──────────────────────────────────────────────────────────────┐
│                                                                            │
│  Choose A File          📁                                                  │
│  Use data from sheet    [ 1                                      ▼ ] ❓     │
│  Column header at line  [ 1                                      ▼ ]        │
│  Separator              [ ,                                      ▼ ] ❓     │
│                         [ dd/MM/yyyy (EU/India/South America/North Africa/Asia │
│  Date format              /Australia)                           ▼ ]        │
│                                                                            │
└────────────────────────────────────────────────────────────────────────────┘

┌─Parameters───────────────────────────────────────────────────────────────┐
│                                                                            │
│  Template    [ Tissue Microarray (TMA) Feature              ▼ ]            │
│  Platform    [ TMA DeCoDe                                    ▼ ]            │
│                                                                            │
└────────────────────────────────────────────────────────────────────────────┘
```

⬅ **Previous**   ➡ **Next**

Figure 4.4: Upload screen for importing features.

- Select the appropriate feature template.
- Select the appropriate platform.
- Click on next.
- Match the headers in the Excel file to the fields present in the selected feature template, either manually or using the "Match" button (see for more details on importing section 3.2.2).
- Choose "Don't import" for all columns that do not need to be imported.
- Click on "Validate".
- If errors are indicated change the file (if needed).
- Click "Import" if everything is correct.

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

4.3 Importing measurements

After a platform has been defined and the appropriate features have been defined within that platform you can start importing your measurements.

Assays are used to link your measurements to the samples created in the GSCF (see also Figure 1.1).

Important note: there is a maximum number of measurements that can be imported in an assay. The maximum is 30.000 samples, 80 features or 1.2 million data points/measurements.

To import your measurements:
- Go to the SAM module.
- Choose from the "Import" menu "Measurements (sample layout)" (see Appendix II for file set-up). "Measurements (subject layout)" is not in use at the moment.

Upload file: Measurements (sample layout)   Choose type / **Upload file** / Match data / Validation / Store data

> ⓘ Upload file and specify parameters
> Below you can upload an excel file, csv file or tab separated file. Please also specify the parameters for the import.

┌─ Upload file ───┐

Choose A File   📁

Use data from sheet | 1   ▼ | ⓘ

Column header at line | 1   ▼ |

Separator | ,   ▼ | ⓘ

Date format | dd/MM/yyyy (EU/India/South America/North Africa/Asia /Australia) ▼ |

└───┘

┌─ Parameters ──┐

Assay | AURKA   ▼ |

Platform | TMA DeCoDe   ▼ |

└───┘

⇐ Previous   ⇒ Next

Figure 4.5: Upload page for importing measurement.

On the "Upload file: Measurements" page:
- Select the appropriate file. You can import Excel, CSV and TSV files.
- 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 in case you import a CSV or TSV file.
- A preview is shown on screen. Check whether the measurements look as expected.
  If a message appears indicating that "added data does not match known number of columns", this means that columns in the file are empty. Click on "ok" if this is correct. Otherwise check your file.
- Leave the date format (dd/MM/yyyy) unless it is not appropriate.
- Select the appropriate assay from the list (please note that all assays from all studies you have access to will be shown).
- Select the appropriate platform.
- Click on "Next".

The message "Added data does not match known number of columns" will appear, ignore it and click on "Ok".
The matching and importing procedure is the same as described for importing subjects (see section 3.2.2).

Go to "Browse", "Assays" (in the "Modules" part) to check the import.
When importing data for the same assay at a later stage, the current data will be overwritten. It is however possible to add more samples and measurements to an already existing assay.

5 Analyze

In this menu two options are present for a quick overview of the data present in the PHdb: "Search" and "Visualization".

5.1 Search

This functionality is not working correctly at the moment.
You can search for fields with certain values both in the GCSF and the Module section. The results can be shown as Studies, Assays or Samples. Please note that this functionality is not working properly at the moment, especially for text values.
- Go to "Analyze", "Search".
- Click on the "Field" entry box. Start typing the field name or enter a space to show the list.

- A list with possible fields appears. In Italic the type of template is indicated in which the field is present.
- Choose a field.
- Click on the "Operator" list and select an operator (e.g. <, = etc).
- Enter a value. For text fields containing a dropdown list, you need to know the values in that list. The values shown upon starting entering a text in this field are not the options in the actual dropdown list.

Figure 5.1: Search database screen.

- Add another criterion (if needed) for your search by clicking on the "Plus" sign. You can also view the result first and then decide to search in that subset.
- Choose the "output type" of your search (Studies, Samples or Assays). When choosing "Samples" the loading of the result will take long, due to the number of samples present in the PHdb.
- Click on "Search".
- The result of your search is given. You can export the result as a simple ToX file (for studies) or a CSV file. If the output was "Study", the downloading time depends on the number of subjects/samples in the selected study.
  When choosing "Simple ToX file" <u>all</u> studies indicated in the overview are exported whether they are ticked or not.

5.2 Visualization

To get a quick overview of the data in a study, you can plot values from subject fields and measurements.
The plot types available depend on the type of values selected (see Appendix III)
To view your data:
- Go to "Analyze", "Visualization".

Figure 5.2: Visualization screen.

- Select your study.
- Click on the X-axis field. A list of available fields appears.
- Start typing the name of the field you want to use or scroll through the list.
- Click to select the field.
- Do the same for the Y-axis.
- Upon selecting the Y-axis a visualization is shown.
- You can group by another field (optional).
- Choose the aggregation type ("Sum", "Average", "Count", "Median" or "No aggregation"). Upon choosing another type, the visualization on the right will change.
- Choose the desired visualization type ("Barchart", "Linechart", "Table", "Horizontal Barchart", "Scatterplot" and "Boxplot"). The types of visualizations that are available depend on the field types and aggregation chosen (see Appendix III).
- To change the field for the X-axis, Y-axis, "Group by" or to choose another study, click on "Clear" for that field.

To change the settings for showing the data, click on "Advanced settings" in the bottom left corner of the table.

6 Export

You can export measurements, study metadata and template settings 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 metadata of the study or a combination of the metadata and measurements.

6.1 Measurements

To export the measurement data from the PHdb:
- Go to the GSCF part.
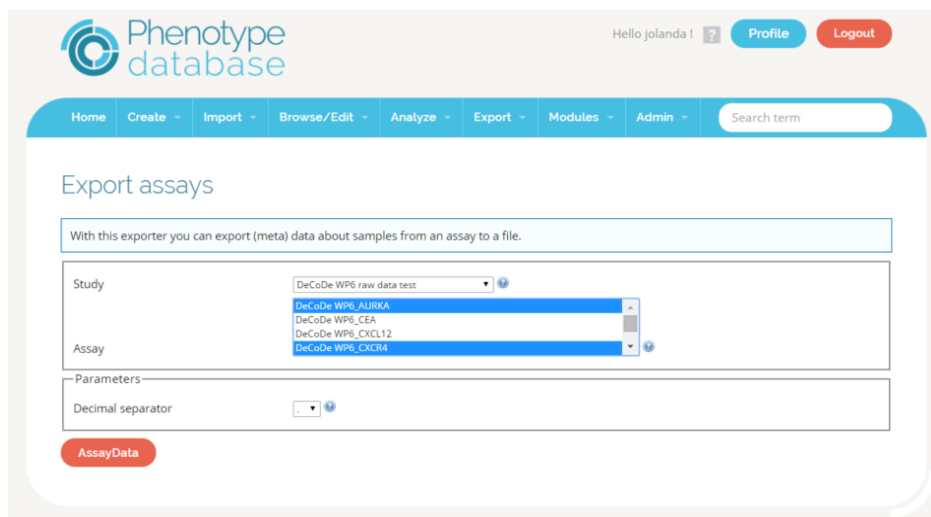- Go to the "Export" menu and choose "Assay data".

Figure 6.1: Export assay data for a study.

- Select your study and the assay(s). Use the "Ctrl" or "Shift" button to select multiple assays.
- Choose the appropriate decimal separator for the output file.
- Click on "Assay data".

The downloaded file contains all subject data, the details on the features and measurements for all indicated assays. In this file you can check whether the samples were linked correctly to the subjects (if applicable; see section 3.3.2). Downloading time depends on the number of measurements.

6.2 Study (meta)data

To export your study metadata (including subjects and sample names):
- Go to the GSCF.
- From the "Export" menu choose "Studies".
- Select the study/studies you want to export.
- The "SimpleTox" downloaded file contains information on subjects, sampling-events (if applicable) and sample-names. In this file you can check whether the samples were linked correctly to the subjects (if applicable. See section 3.3.2).
- Clicking on "combined Excel" results in one Excel file containing the study, subject and sample data including the measurements, on separate tabs. Additional sorting may be needed for the sample data in case more than one study was selected.

6.3 Templates

The contents of all types of templates used in the PHdb can be exported as a XML file. This file can be imported to create the templates in another environment (for instance the Production environment).
Creating/importing of templates in the Production environment can only be performed by a template administrator. Contact Topdesk (https://trait.topdesk.net/) to move templates from the Sandbox to Production.
In Topdesk:
- Go to "diensten en aanvraag formulieren"
- Go to "Experimental data", "Phenotype database"
- Choose the "PheDB Template New, Import or Modification in Production" from the forms indicated on the right

To export templates:
- Go to the GSCF.
- From the "Export" menu choose "Templates".
- Scroll through the list to search the desired template.
- Click to select a template. Use "Ctrl+click" to select more than one template.
- Click on export.
- A XML file will be created containing the selected template(s).

7 Additional Import options

In the Import menu two options are available: "A part of the study design" and "Templates". Both options are explained in detail in the next sections.

Importing of features and measurements is situated in the "Modules" part of the PHdb (see section 4).

7.1 Importing a part of a study design

It is possible to import subject, sample, event (=treatment type), sampling event (=sample type) and assay details (template fields, not results for measurements) into the PHdb.

For treatment and sample types and assay details, it will be more convenient to create them in the GSCF part. Choose the type of data you want to import and click on "Next". See for more details on the general functionality on importing of subject data and sample names sections 3.2.2 and 3.3.2 respectively.

7.2 Importing templates

Only use XML files created by PHdb to import template data or an XML file with the same set-up.

Importing of templates on the production environment can only be performed by a "template administrator". Contact the Topdesk helpdesk to move templates from the sandbox to production by completing the form "PhenotypeDB New, Import or Modification of a Template in the production environment" (directory "Webformulieren Experimental data").

To import a template:
- Go to "Import".
- Choose "Templates".
- Click on the "No file selected" text.
- Select the XML file containing the template(s) details.
- Click on "Import".
- An overview will be given of the templates present in the file (Figure 7.1). If a similar template is already present in the PHdb, this will be indicated on the right for each template.
- If a template with the same name is already present, this will be indicated. You can change the name of the template you are importing.
- Tick the boxes for the templates you want to import.
- Click on "Verzenden" to import the templates.

Figure 7.1: Importing templates.

An overview will be given of the templates which were or were not imported (including the reason why it was not imported).

8 Admin

This part of the PHdb is only available for persons with the "ROLE_ADMIN" role.
The following options are available:

List users:	Overview of all users in this environment (see section 8.1 for more details)
Create users:	Used to create new users (see section 8.2)
List groups:	Overview all user groups currently defined (see section 8.3)
Create groups:	Used to define user groups (see section 8.4)
Manage modules	Not in use
Setup wizard	Not in use
Application information	Detailed, technical information on the current application version

8.1 List users

An overview is shown of all users in the current environment.
You can search for users based on user name and/or account status. Enter the search criteria and click on "search". The appropriate users will be shown in the table.

To open the details for a certain user, click on the user name.
You can:
- Set a new password.
- Change the email address.
- Change the status of the user account.

For all active accounts "User confirmed" and "Admin confirmed" need to be ticked (see also section 8.2).
Don't use the options "Account expired" and "Password expired". To lock an account, tick the "Account locked" box.
Click on "Save" to save the changes.
To return to the "List of users", choose "List users" from the Admin menu.

8.2 Create users
To create a new user:
- Go to "Admin".
- Choose "Create users".
- Enter the username (this will be the login name).
- Enter an initial password (this password needs to be send to the user).
- Enter the email address.
- Tick the boxes for "User confirmed" and "Admin confirmed" to activate the account.
- If a person should have "Template administrator" and/or "Administrator" access, go to the "Roles" tab. Choose "ROLE_TEMPLATEADMIN" and/or "ROLE_ADMIN" respectively.
- Click on the "User groups" tab to select the a user group (if applicable). Click on the name of the group in the "Not in group" list to select the group for this user or vice versa.

8.3 List groups
Overview of all current groups are shown. You can search for a group name at the top of the screen.
- Click on the group name to open the group details.
- Click on the "Users" tab to view and change the persons in that group.
- Click on a user in the "Users not in group" list to add it to the group.
- Vice versa: click on a user in the "User in group" list to remove from the current group.
- Click on save to save the changes.

8.4 Create group
Use user groups to easily add a number of users as readers or writers to a study.
To create an users group:
- Go to "Admin".
- Choose "Create group".
- Enter a (unique) group name.
- Enter a group description.
- Click on "Save" to create the group.
- An "Users" tab is now added. Click on this tab to add users as described in section 8.3.

Appendix I: Example for 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	Marker name	Unit	Scoring performed on	TMA core scored	Note o
2	EGFR CRCLM core cytoplasmic staining frequency	-4: ask pathologist, -3: not representative, -2: No	cytoplasm	CRCLM	For ea
3	EGFR CRCLM core cytoplasmic staining intensity s	-4: ask pathologist, -3: not representative, -2: No	cytoplasm	CRCLM	For ea
4	EGFR CRCLM core membranous staining frequenc	-4: ask pathologist, -3: not representative, -2: No	membrane	CRCLM	For ea
5	EGFR CRCLM core membranous staining intensity	-4: ask pathologist, -3: not representative, -2: No	membrane	CRCLM	For ea
6	EGFR primary CRC core membranous staining freq	-4: ask pathologist, -3: not representative, -2: No	membrane	primary CRC	For ea
7	EGFR primary CRC core membranous staining inte	-4: ask pathologist, -3: not representative, -2: No	membrane	primary CRC	For ea

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:

Appendix II: 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	SubjectID	Species	Gender
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. See also section 3.4.2)

	A	B	C	D	E
1	SubjectID	SampleName	SampleMaterial	sampleMaterial-particulars	
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	SampleName	SubjectID	EGFR primary CRC core membranous staining frequency score	EGFR primary CRC core membranous staining intensity score
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 III: Available visualizations depending on the field types selected for the X- and Y-axis

The available visualizations depend on the type of fields selected and on the aggregation. In the table below the available options are indicated:

X-as[1]	Y-as[1]	Group by	Aggregation[2]	No aggr.	Barchart	Linechart	Table	Hor. Barchart	Scatterplot	Boxplot
Category	Numeric	-	x		Yes	Yes	-	-	-	-
Category	Numeric	-		x	Yes	Yes	-	-	-	Yes
Numeric	Category	-	x		-	-	-	Yes	-	-
Numeric	Category	-		x	-	-	-	Yes	-	-
Category	Category	-	(Count)	NA	-	-	Yes	-	-	-
Numeric	Numeric	-	x		-	-	-	-	Yes	-
Numeric	Numeric	-		x	-	-	-	-	Yes	-
Category	Numeric	Category	x		Yes	Yes	-	-	-	-
Category	Numeric	Category		x	-	-	-	-	-	Yes
Numeric	Category	Category	x		-	-	-	Yes	-	-
Numeric	Category	Category		x	-	-	-	Yes	-	Yes

1 Examples of a "Category" field are gender and ethniticity.

2: Available aggregation types are: Sum, Average, Count, Median.