Project Wizard Documentation, Version 1.1 - 12.02.18

1. Introduction 1

2. Step-By-Step Explanation 2

3. Special Cases 10
3.1 Experiments on Cell Lines and Single Cell Organisms 10
3.2 Multiple Levels of Exp. Factors 12

4. Batch Upload of Samples 18
4.1 General Format 18
4.2 Adding additional meta-information 19
4.3 Upload of the Experiment 20
4.4 Immunomics Format 21

5. Metadata Upload/Edit 21
5.1 Special Properties 24

1. Introduction

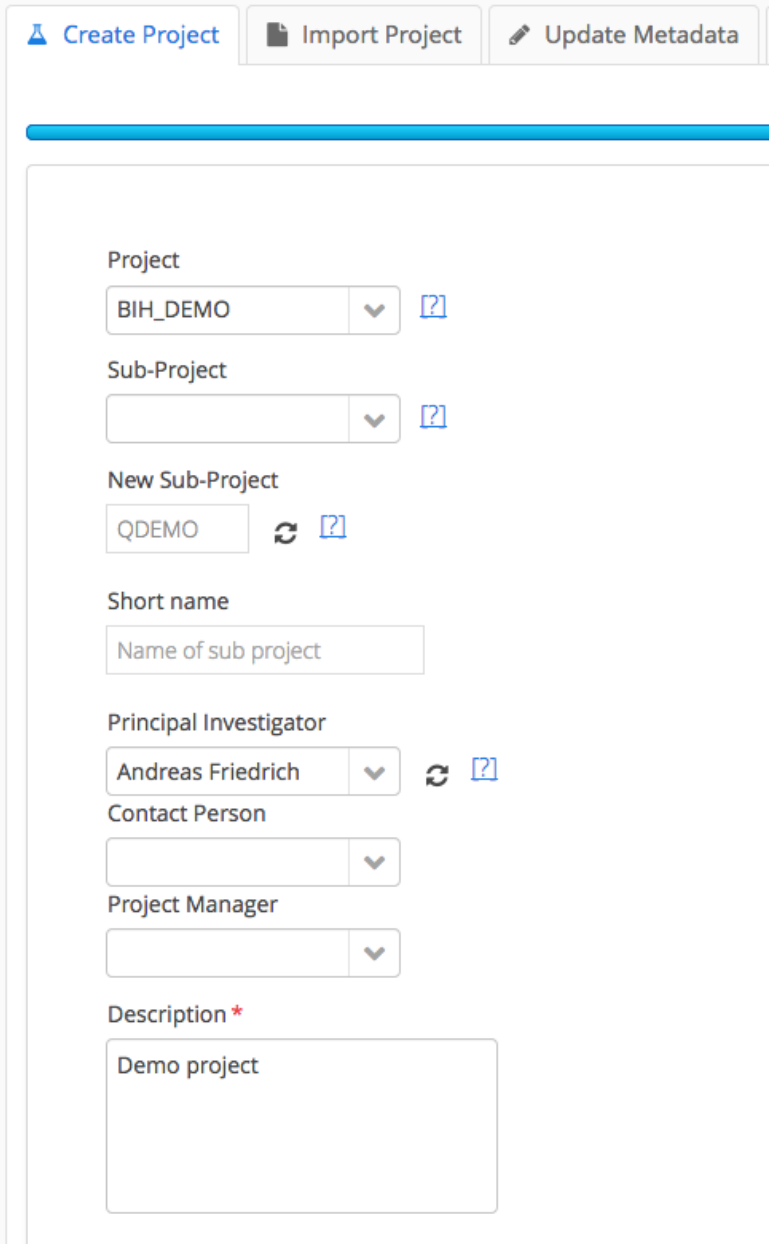
qWizard is a web-application embedded into the QBiC portal (https://portal.qbic.uni-tuebingen.de) allowing users to create a full-factorial experimental design. In a full-factorial design every experimental variable, e.g. genotype, different tissue types, treatments etc., as given by the user is multiplied by the number of replicates, leading to the number and type of samples. To give an example illustrating this principles: Selecting two different tissues from one patient will result in two samples, while selecting two different tissues from two patients will results in four samples. To this end, we see several key advantages in using qWizard:

(a) it speeds up experimental design
(b) it helps to perform statistically sound experiments

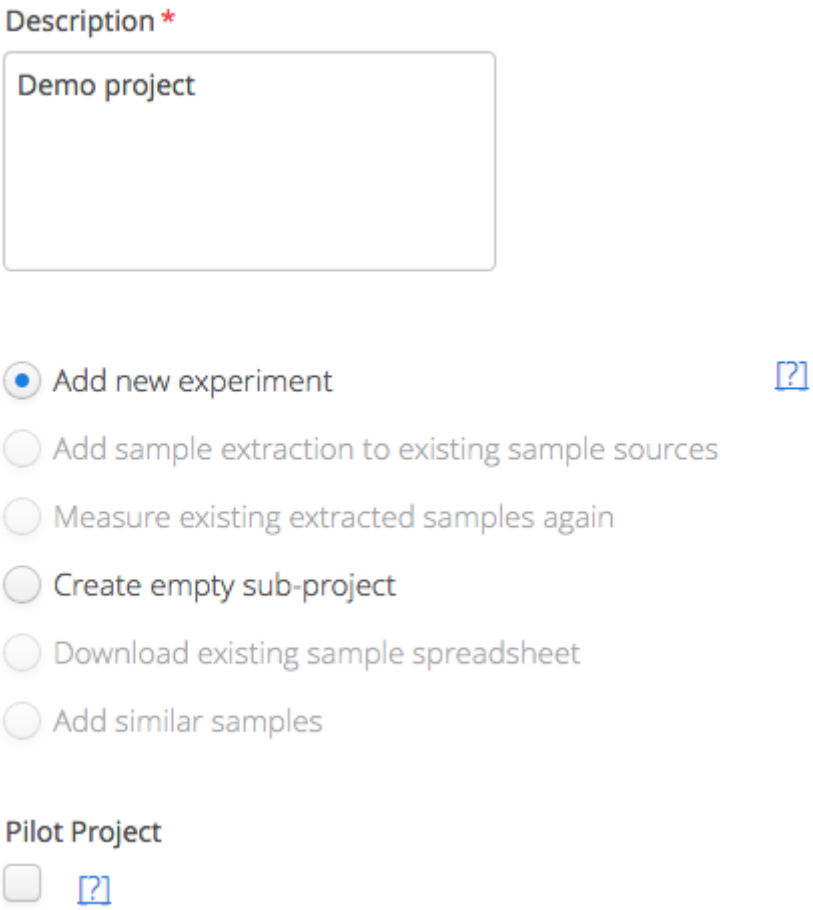
In reality, full-factorial designs are rare, so qWizard enables users to remove unnecessary samples as well as specify.
An additional consideration is that qWizard uses a multi-tier data model. This means that there are (currently) three different main levels to input information about experimental design and samples. These levels are the sample source (or patient) level, the tissue (or cell culture) level and the analyte (measured samples or ”test samples”) level. Further levels exist, for example for specific experimental steps like in MHC Ligand Extraction or the Mass Spectrometry runs themselves, but information about these is either collected at the end of the wizard or when the data is registered.

2. Step-By-Step Explanation

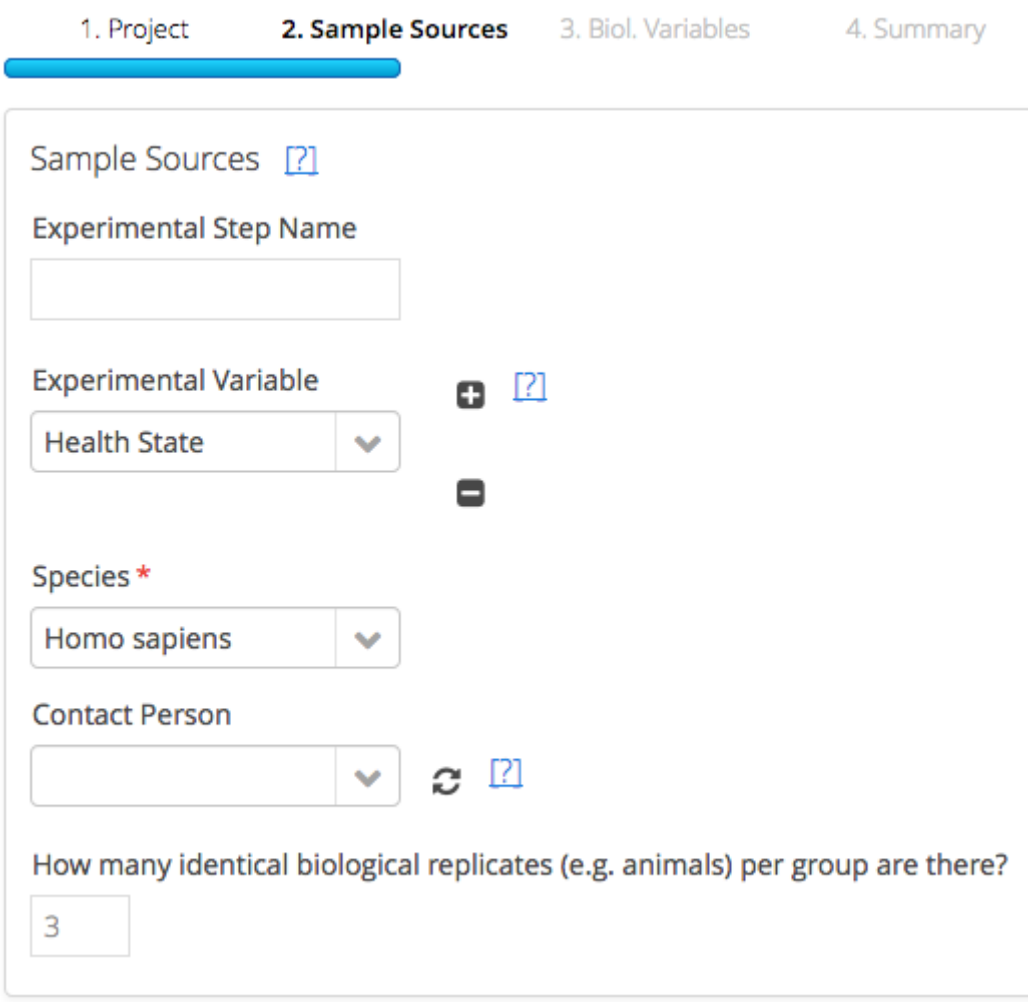
In the first step, a project space is selected. An existing sub-project can be chosen to add experimental steps with samples or a new sub-project can be created. If a new sub-project is created, a description is needed. Other information like a meaningful short name and people involved in the project can be added:



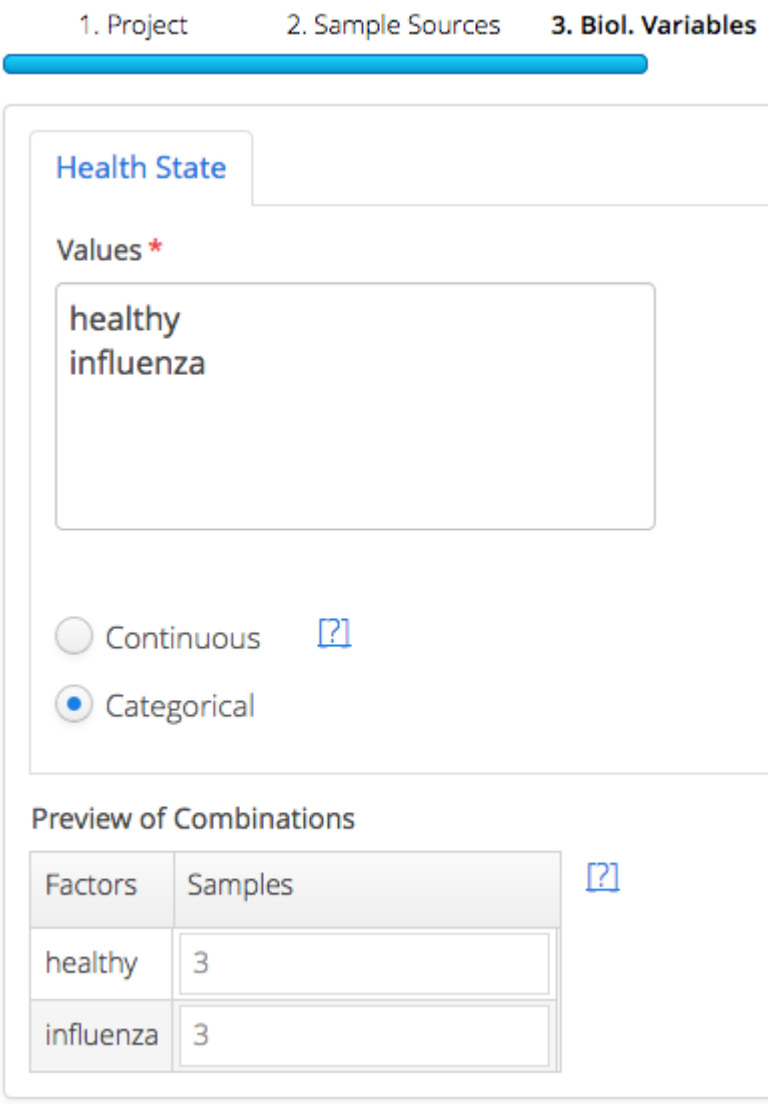
At the bottom of the first step, we can either add a completely new experiment or specify which kind of addition to an existing sub-project we would like to make. We can also create an empty sub-project and add samples later.
If our sub-project is a pilot study, the related box can be checked.



In the next step, the source of our samples has to be selected. In this example, we want to compare a group of healthy humans with a group of influenza patients. We select Homo Sapiens and the experimental variable “Health State”, as well as 3 replicates (each group will consist of 3 people):



Since we selected the “health state” variable, the next step asks us to specify which kind of health states are investigated in this project. After adding this information (one per line) and selecting the type of value (in this case “categorical” information), the amount (which can be changed) and type of patients are summarized:



In the next step, we can make some last changes to our patient information. We can change names, add a second identifier or delete superfluous samples:

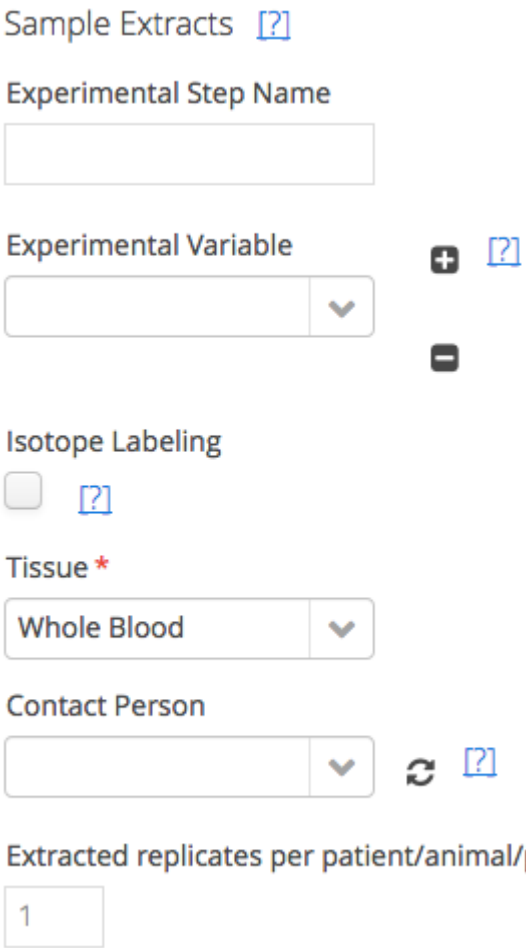
Sample Sources Tailoring [?]

6 Samples

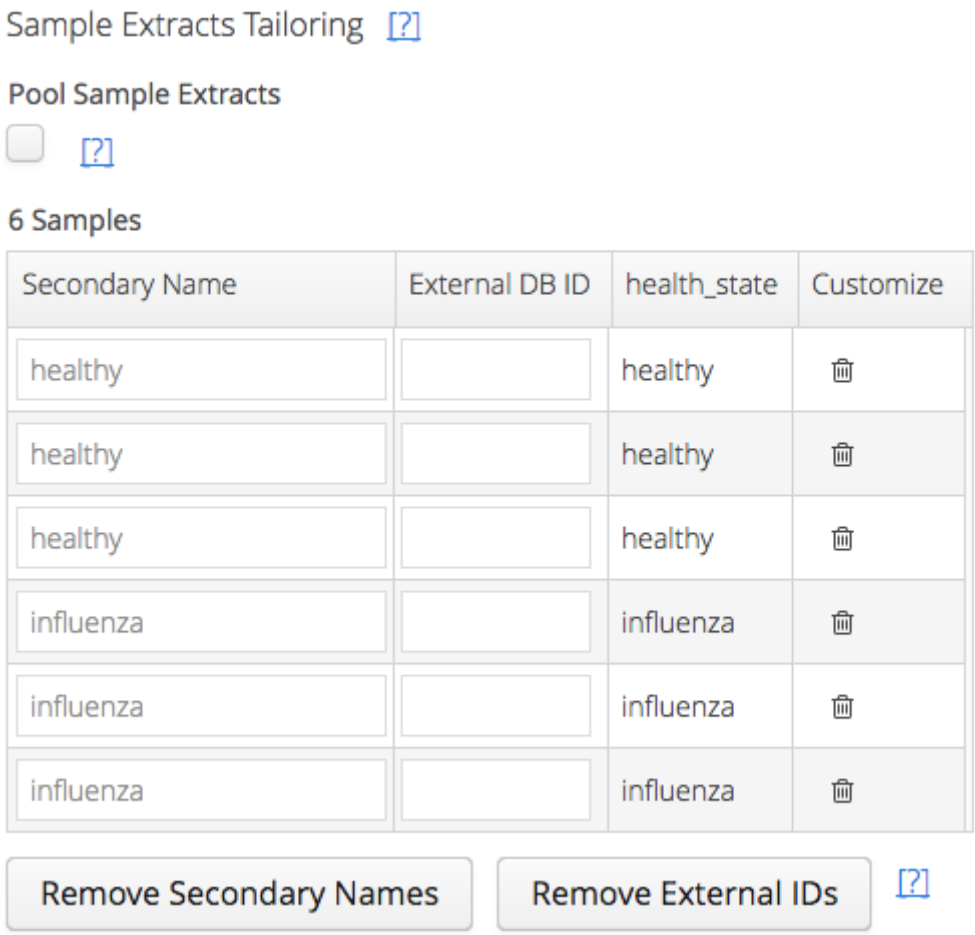
Secondary Name	External DB ID	health_state	Customize
healthy		healthy	
healthy		healthy	
healthy		healthy	
influenza		influenza	
influenza		influenza	
influenza		influenza	

Remove Secondary Names | Remove External IDs [?]

The following step handles sample extracts: the kind of tissue taken from patients for this experiment. For special cases (often cell-cultures), isotope labeling can be selected. In this case we only select that blood samples are taken. Since we draw only one blood sample per patient, we leave the number of replicates at 1:

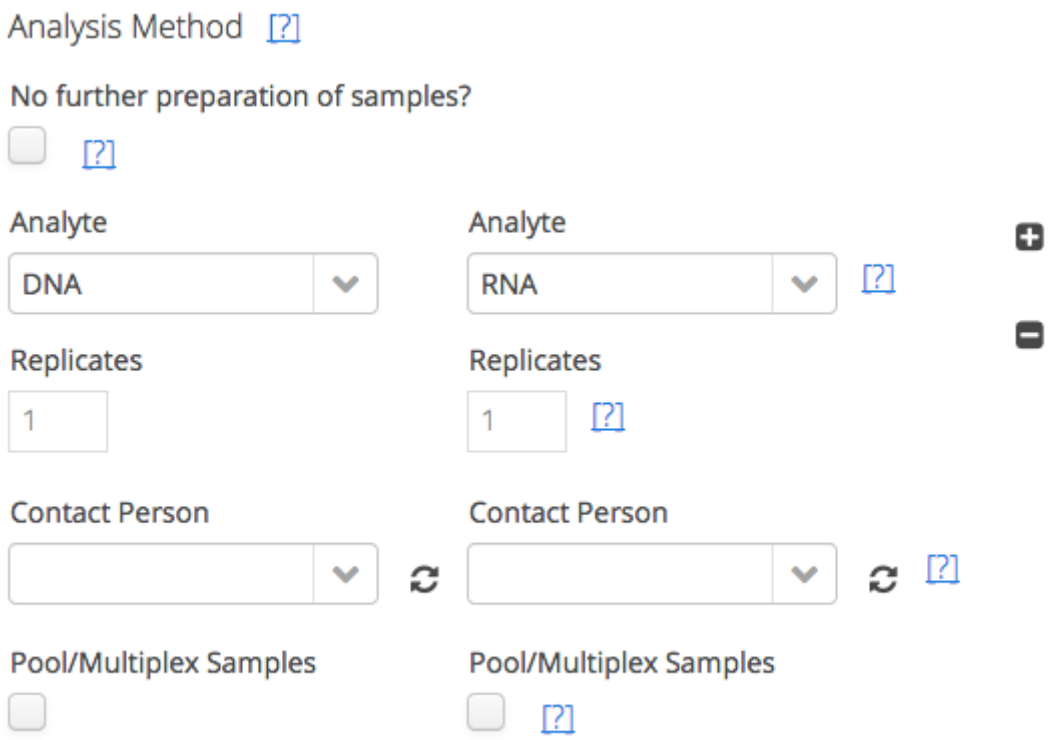


We could theoretically add other experimental variables at this step (more on that later), but since we did not, we are presented with the summary step of this (extract) level:

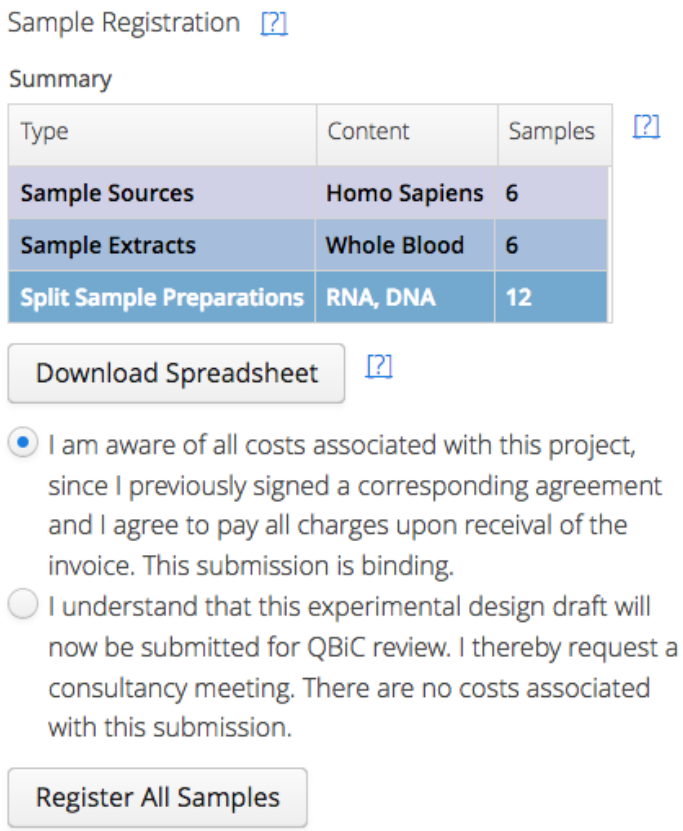


For cases using isotope labeling, samples can be pooled here. Again, names of the blood samples can be changed or single samples deleted.

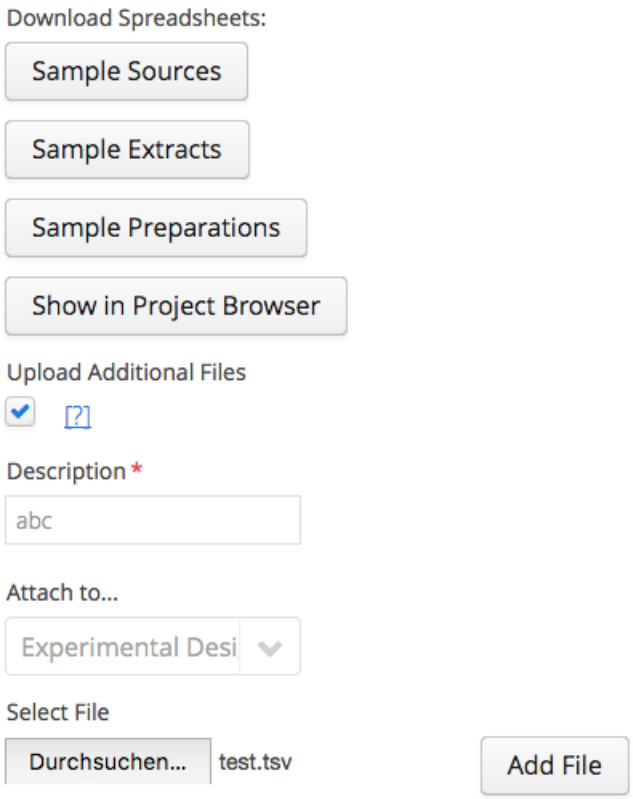
In the next step, the analysis method(s) are selected. Let's say we want to sequence DNA as well as measure RNA abundance of different active genes using a Microarray and extract one of each analyte samples from each blood sample (keep "1" as replicate):



In cases where no Analytes are extracted (e.g. medical imaging), we can stop the experiment process at this point by selecting “No further preparation”. In both cases we are presented with a summary in the next step:



We created 6 patient objects, 6 blood samples, and 6 (RNA) + 6 (DNA) samples, leading to 12 analyte samples in total. We can now select to register the experiment(s) and samples we created or send in the study design for review.



After registration of experiments and samples is completed, we can download spreadsheets of the different experiment levels. This is especially useful for easily updating metadata later (see Chapter 5).

We can now directly go to the project in the Project Browser or we can upload some additional information via small experiment-related files. To do this, we select the box, add a description and choose the file from hard disk. Multiple files can be added before pressing “Commit”, sending them to the database.

3. Special Cases

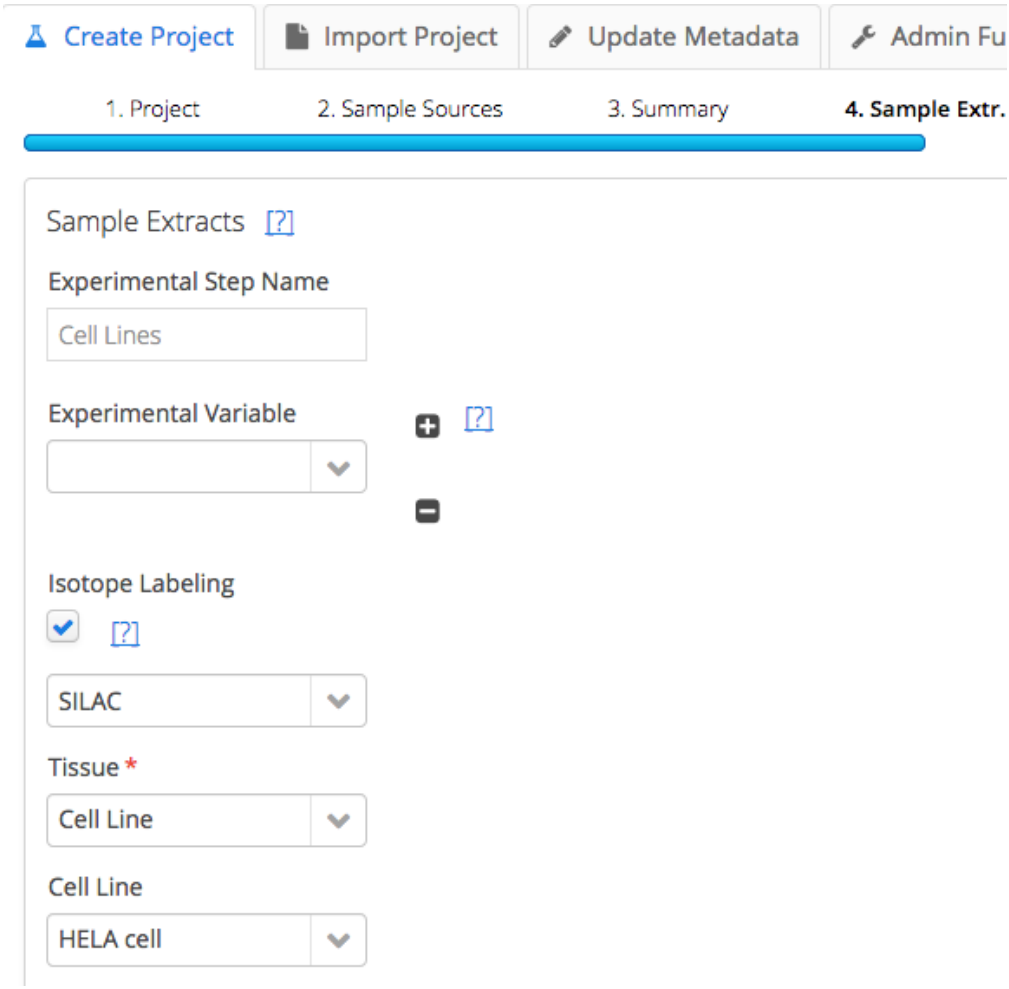
3.1 Experiments on Cell Lines and Single Cell Organisms

The three-tier-model of sample source, tissue and analyte preparation leads to some ambiguous solutions concerning the handling of single cell organisms and cell lines.

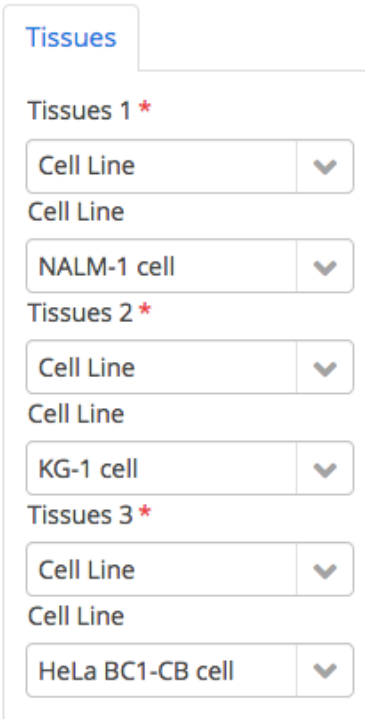
For single cell organisms like bacteria, that can be selected, a “tissue” has to be chosen at the second level. Here, selecting “**Full Organism**” is common. When selecting the option “**Other**”, something else can be specified.

Although it can be argued that the genetic makeup of some (cancer) cell lines makes them unique organisms themselves, qWizard models **cell lines on the level of Tissue Extracts**. This means that an organism has to be chosen in the first step. This can either be the origin of the cell line, e.g. **Homo sapiens** for HeLa cells, or “**Unidentified**” if no further information is available.

When selecting **Cell Line** from the drop-down menu in the **Sample Extraction** step, a cell line can be specified:



Multiple different cell lines can be used in the same experiment by selecting the **Experimental Variable “Tissue”:**

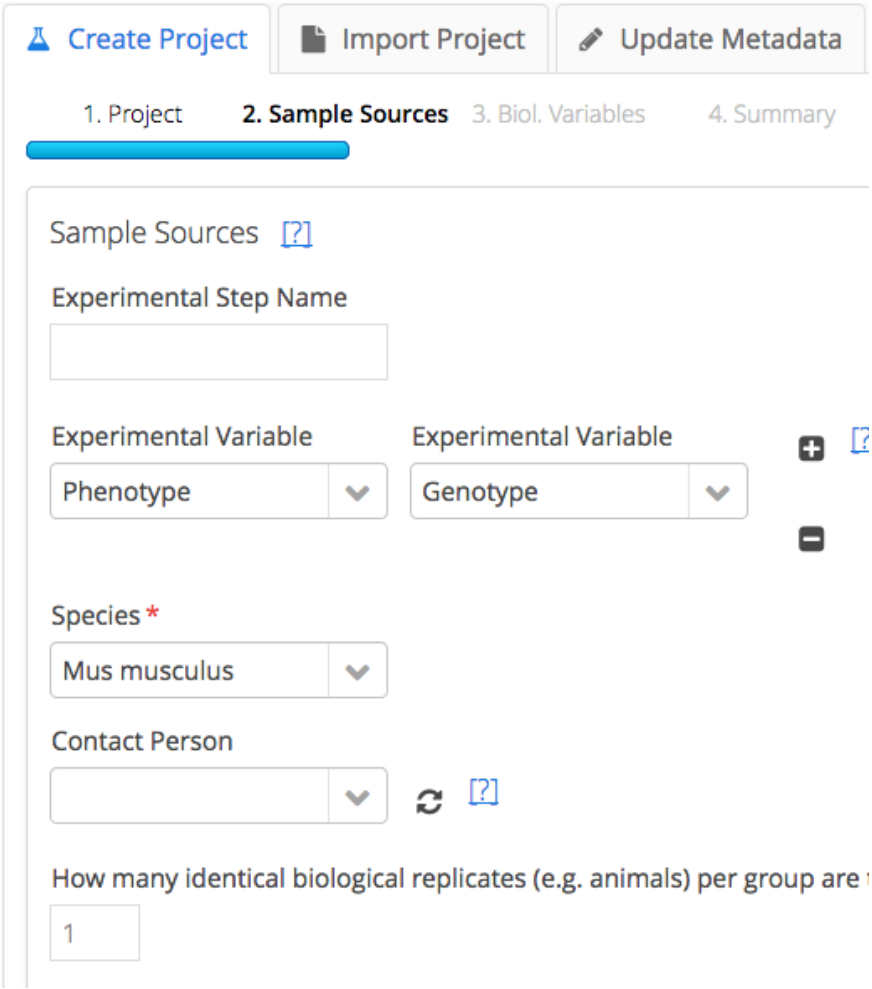


3.2 Multiple Levels of Exp. Factors

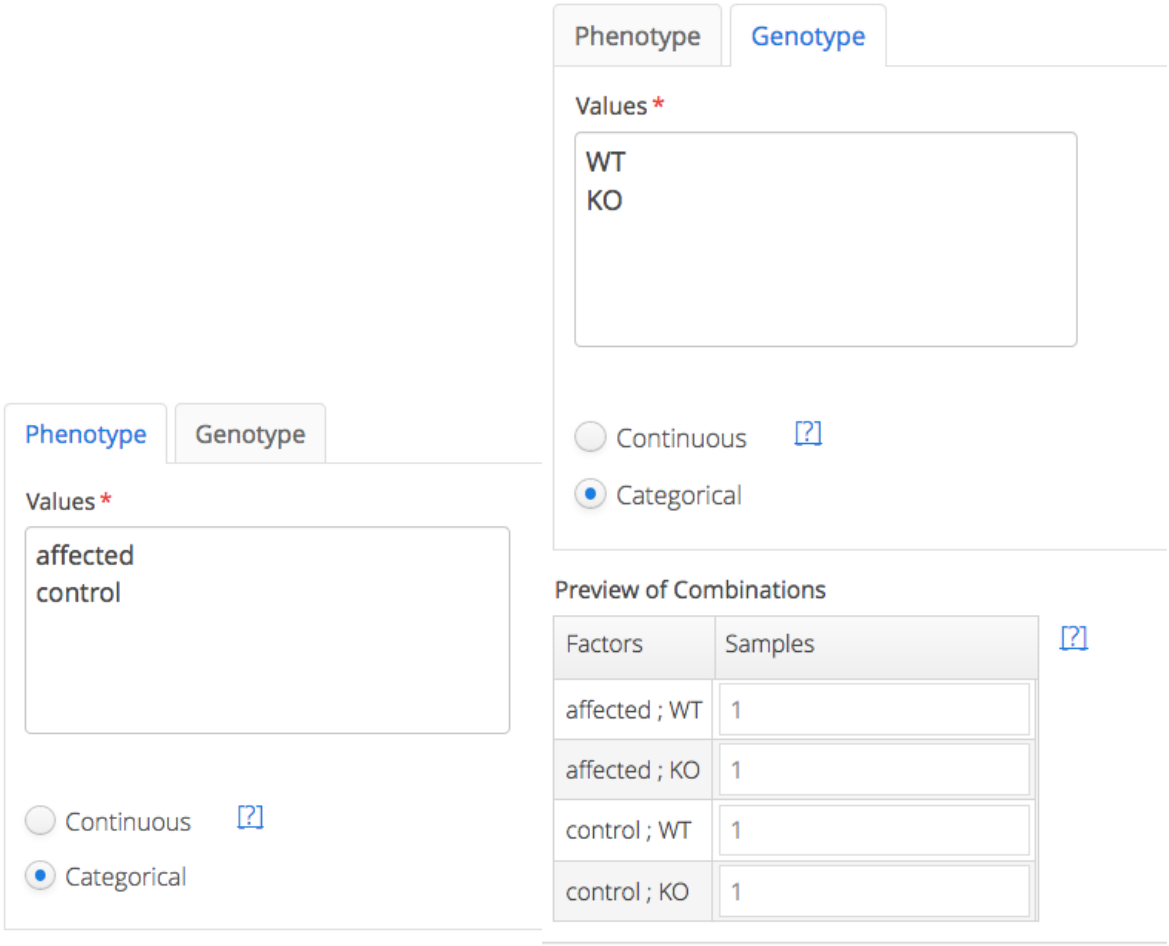
Complex experiments may not only examine a single experimental factor (with several test conditions) but the interplay of factors like genotype, treatment and different time points after the treatment. qWizard always multiplies factors within and between different levels of the experiment, creating all possible permutations of these attributes as a sample (or multiple samples, depending on the number of chosen replicates). Users can then fine-tune the numbers of the different sample types in their experiment.
For this to work, however, care has to be taken to input the conditions at their respective levels.

In this example, 6 mice are analyzed. They differ in genotypes (**3 WT and 3 KO**) and phenotypes (**2 affected and 4 control**) Furthermore, two different extracts (**blood and lymph - one for each mouse**) are sampled.

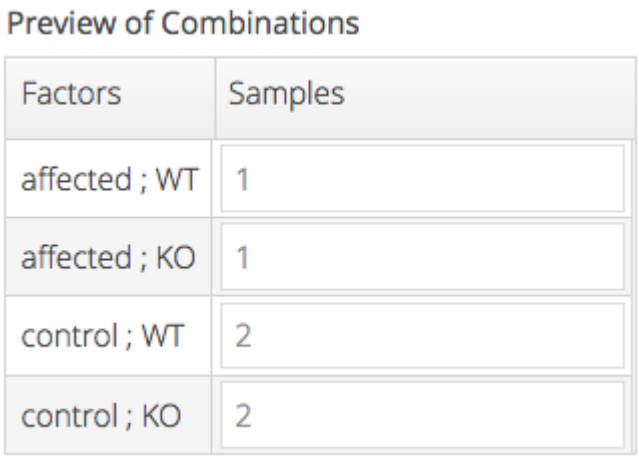
After adding the general information about the project (see Chapter 2), we select the two experimental variables concerning our mice in the **Sample Sources** step. We choose to leave the number of replicates at 1 and to fine-tune this information in the following step.



After selecting genotype and phenotype, the next step asks for the different **test conditions** of the two types of experimental variable. We add **affected** and **control**, as well as **WT** (wild type) and **KO** (knock out).



The resulting preview of all combinations represents four different types of mice: **affected+WT; affected+KO; control+WT and control+KO.**
Since we chose only one replicate in the previous step, qWizard assumes that the experiment contains one mouse of each type.
However, since we have two control mice of each genotype, we change the number accordingly:



In the following step we see the six mice in our experiment. We can rename the **Secondary Name**, which automatically contains an overview of all experimental factors or we can add

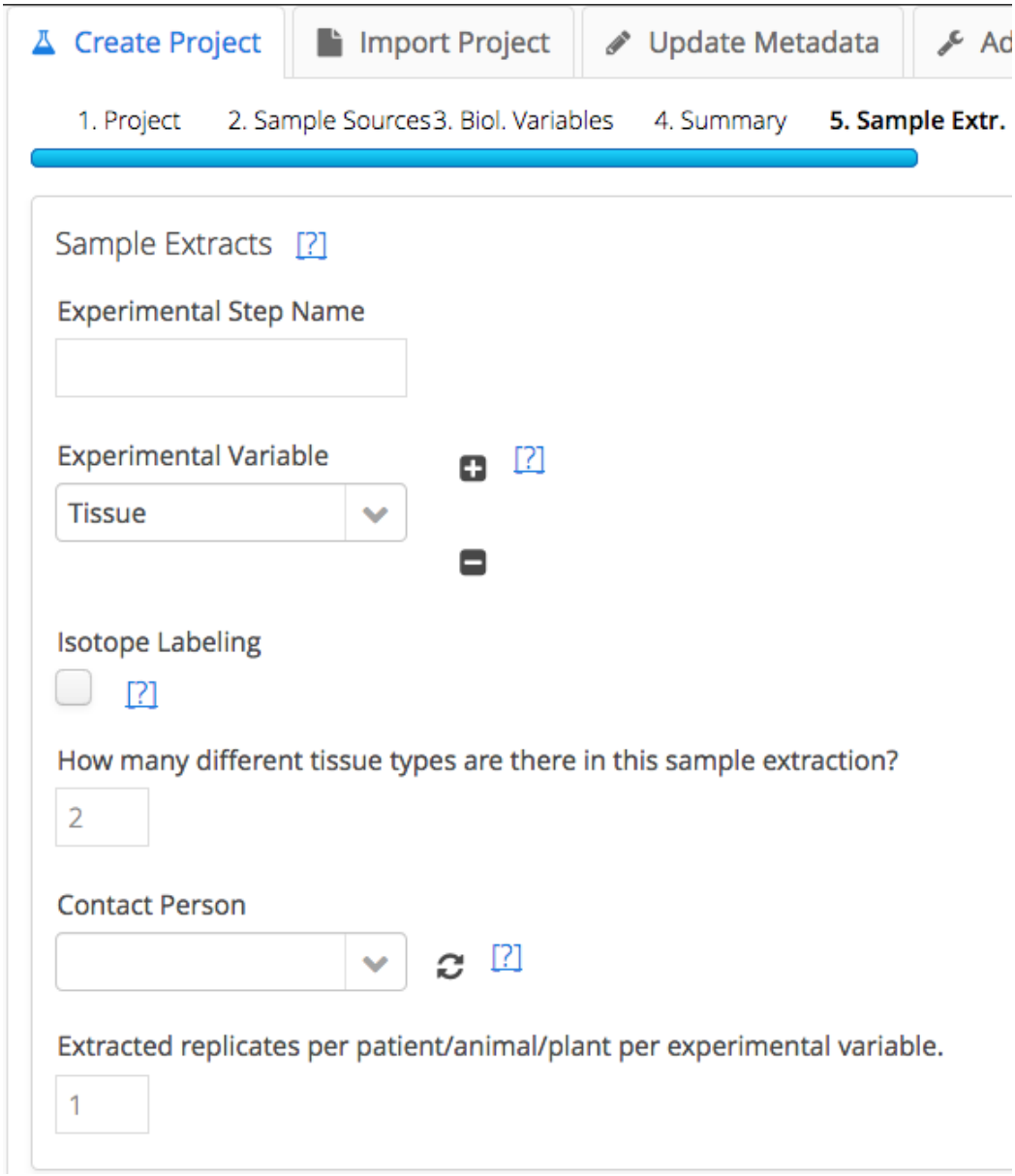
Lab IDs (in this case just “mouse 1” etc.). Both of these infos are used in many other parts of qPortal to make it easier to identify samples or sample sources.
Experimental Factors are saved after sample registration, even if we change the names of our sources/samples.

Sample Sources Tailoring [?]

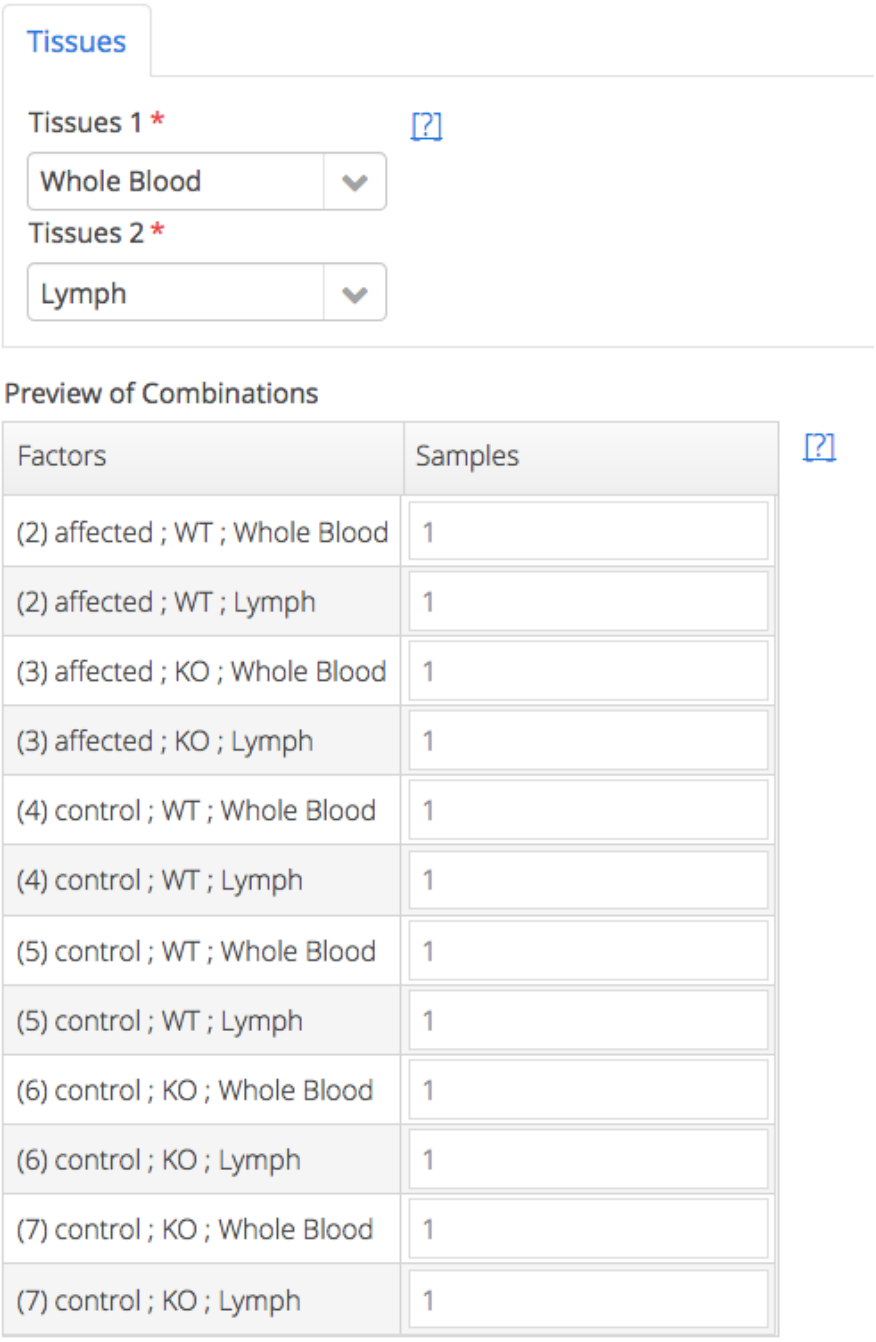
6 Samples

Secondary Name	External DB ID	phenotype	genotype	Customize
affected ; WT	mouse 1	affected	WT	
affected ; KO	mouse 2	affected	KO	
control ; WT	mouse 3	control	WT	
control ; WT	mouse 4	control	WT	
control ; KO	mouse 5	control	KO	
control ; KO	mouse 6	control	KO	

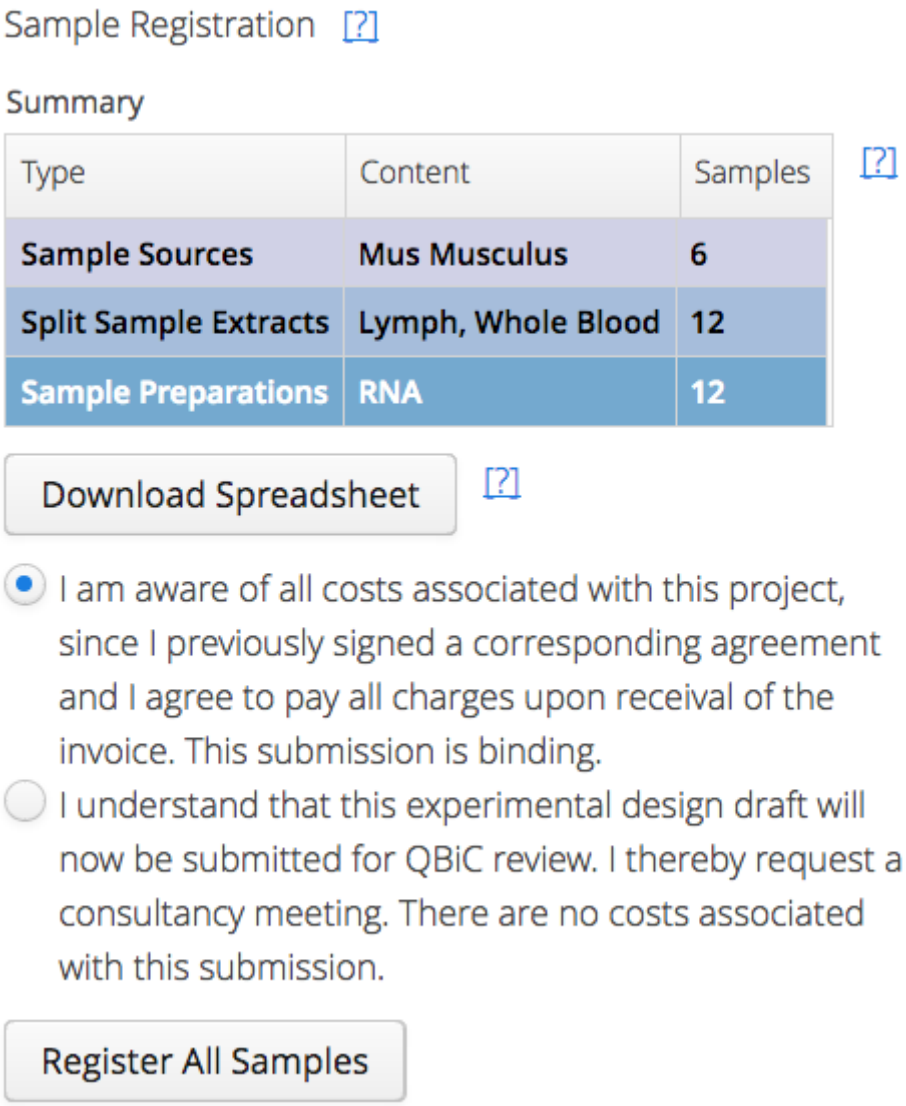
Since we only input information about our lab animals up to this point, we still need to specify which samples are taken from them. This happens in the next step: **Sample Extracts.**
Since we want to extract **blood** as well as **lymph**, we chose **Tissue** as an experimental variable and set the number of tissue types to 2. Again, we leave the number of extracted replicates at 1 for simplicity’s sake:



The next step allows us to choose the different “tissue types”, in this case **Whole Blood** and **Lymph** from a pre-defined vocabulary. Like in the case of different mice, we see the resulting samples below once we have made our choice. In this case, taking one blood and one lymph sample from each of our six mice leads to 12 extract samples:



After selection of Analyte (and no technical replicates), this leads to the following result which we can register in the system or send in for QBiC review:



The Batch Upload functionality (see Chapter 4) automatically recognizes to which tier of the experiment an experimental factor should belong. E.g. if two conditions of the same specified factor are different for the same mouse, then this condition can not be related to the mouse (like, for example, a genotype), but has to belong to different tissue extracts or analytes, like a time point of extraction.

4. Batch Upload of Samples

Creating large experiments with many samples and rich metadata by hand can sometimes be a hassle. To simplify and speed up this process, we have created several spreadsheet file formats that can be used to directly upload experiments or update the metadata of existing experiments afterwards (see Chapter 5).

4.1 General Format

- Columns are tab-separated
- No quotation marks or additional tabs should be used
- Order of columns can be random

All Batch Upload files must contain the basic information about Sample Sources, Tissues and Analytes (explained in Chapters 1 and 2) as well as label the different entities of each of these levels with unique IDs:

Organism	Organism ID	Tissue	Extract ID	Analyte	Analyte ID
Mus musculus	m1	Whole Blood	B1-1	RNA	R3001
Mus musculus	m2	Whole Blood	B2-1	RNA	R3002
Mus musculus	m3	Whole Blood	B3-1	RNA	R3003

In this example experiments, blood samples from 3 mice are taken and RNA is extracted. Since each sample and each mouse is unique, all the identifiers (on all three levels) are unique as well.
To illustrate how **replicates** are handled, we will add two additional RNA measurements: one further blood sample (B2-2) taken from the second mouse (m2), as well as one taken from the existing blood sample (B3-1):

Organism	Organism ID	Tissue	Extract ID	Analyte	Analyte ID
Mus musculus	m1	Whole Blood	B1-1	RNA	R3001
Mus musculus	m2	Whole Blood	B2-1	RNA	R3002
Mus musculus	m3	Whole Blood	B3-1	RNA	R3003
Mus musculus	**m2**	Whole Blood	**B2-2**	RNA	R3012
Mus musculus	**m3**	Whole Blood	**B3-1**	RNA	R3013

Keep in mind, that these identifiers are just examples. In reality, they can be based on lab-internal labeling, as long as they are unique for different entities.

4.2 Adding additional meta-information

Experimental factors (and other properties TODO) can be added in additional columns:

Organism ID	Extract ID	Analyte ID	Condition:	Condition:	Condition:

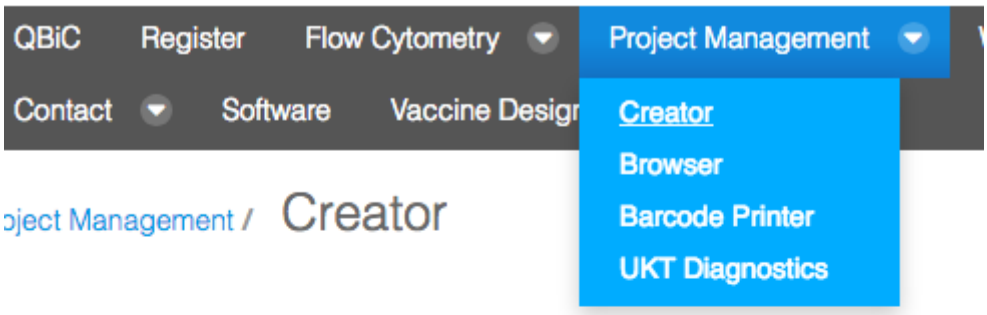
			genotype	extraction_time	protocol
m1	B1-1	R3001	gene+	t1	a
m2	B2-1	R3002	gene+	t1	a
m3	B3-1	R3003	gene-	t1	a
m2	B2-2	R3012	gene+	t2	b
m3	B3-1	R3013	gene-	t1	b

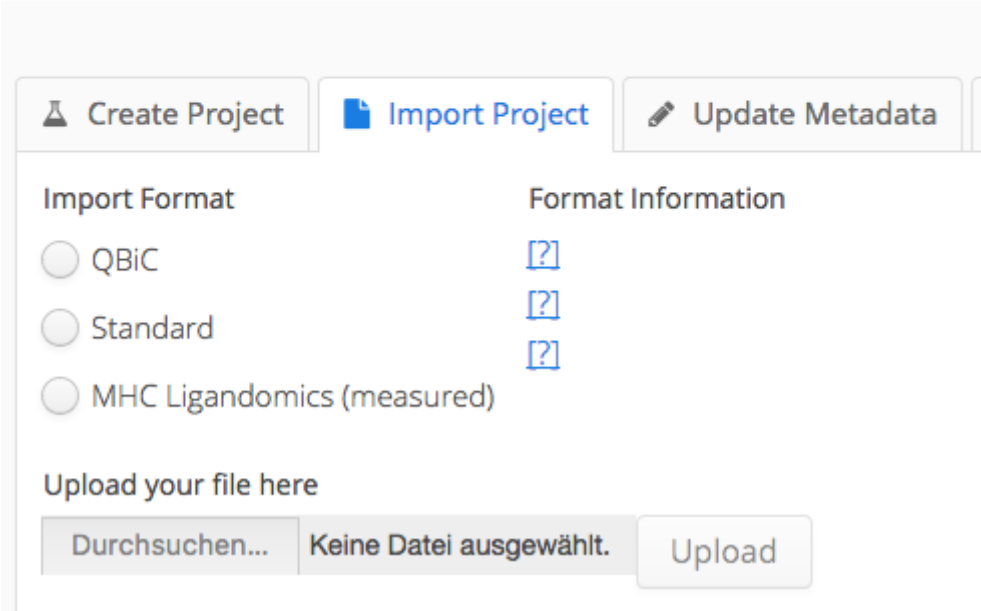
In this example, the conditions genotype (on Organism level), extraction time (of the blood, Extract level) and protocol of RNA preparation was added: Mice m1 and m2 have a certain allele, while m3 is a knockout; all blood samples but B2-2 were taken at the same time and the last two RNA measurements were performed using a new protocol (“b”).
Organism, Tissue and Analyte columns have been removed from this example for the sake of clarity.

To keep the basic upload format simple, additional properties can be added with the Metadata Upload functionality (see Chapter 5).

4.3 Upload of the Experiment

The upload functionality can be found as part of the Project Creator on qPortal’s Project Management Section:





To register your experiments follow these steps:

1. Create a TSV file using the information above or the example
2. Select “Import Project”
3. Select “Standard” and upload the file
4. Select Project and Sub-project and complete other missing information
5. Register experiments, a download link is created containing one barcode per row of the uploaded file
6. Download file use the first two columns to add the QBiC barcode to each file
7. Transfer the files (more information about this can be found…)

4.4 Immunomics Format

Coming soon

5. Metadata Upload/Edit

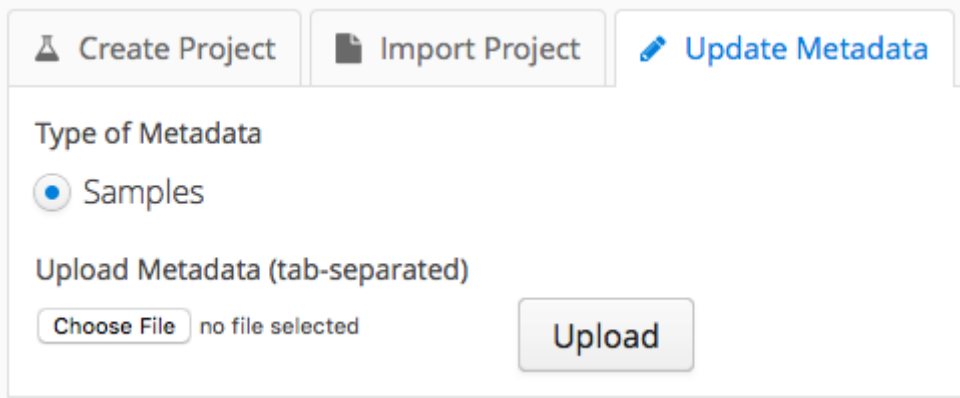
Since the multi-step registration as well as the batch registration formats are kept as simple as possible, qWizard provides an additional option to add or update metadata of existing

samples. For this, any type of spreadsheet (tab-separated values) can be uploaded, as long as it contains QBiC sample identifiers in one of the columns.

Sample Identifiers for specific groups of samples can be downloaded from the qNavigator portlet or in a more general form from qWizard itself (after registration or batch upload).

Every user that has access to a project can add new information in this way. However, to change (overwrite) existing information, additonal permissions are needed.

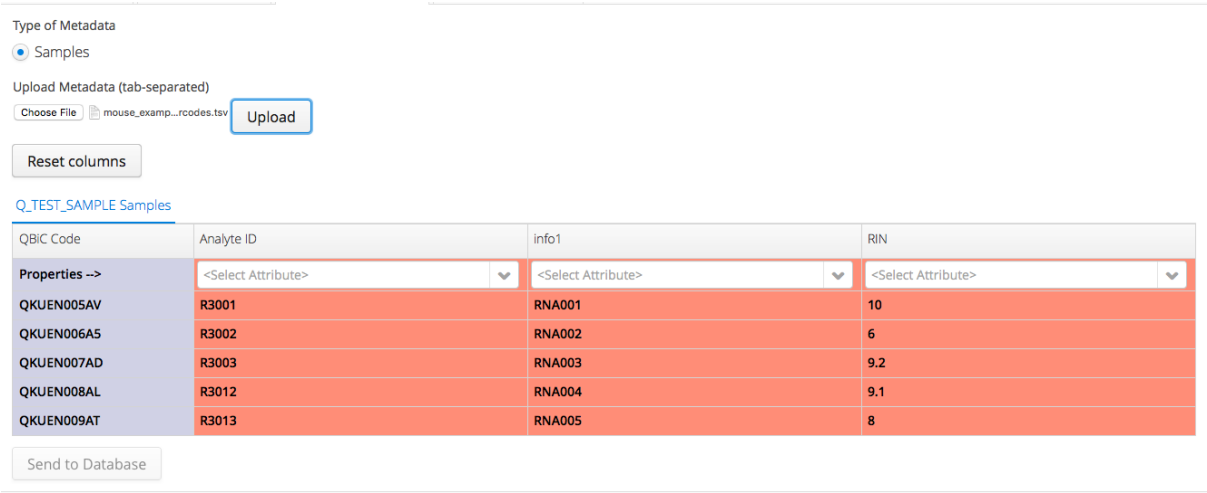
To upload information about your samples, go to the “Update Metadata” tab and select “Samples”:



As an example, we upload a file of our mouse project from Chapter 4:

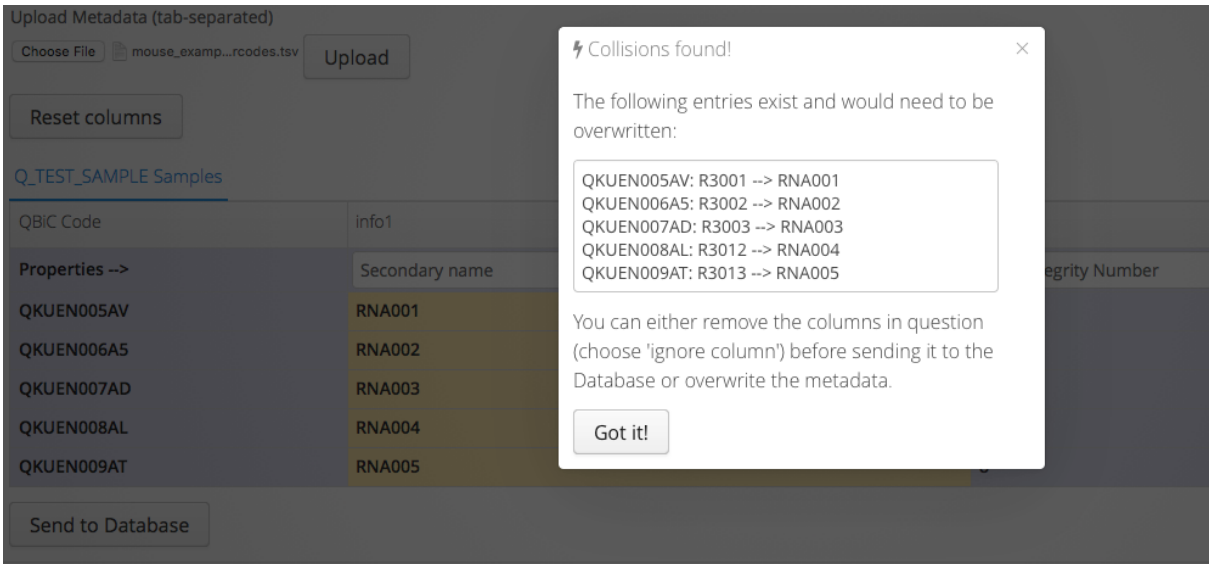
barcode	Analyte ID	info1	RIN
QKUEN005AV	R3001	RNA001	10
QKUEN006A5	R3002	RNA002	6
QKUEN007AD	R3003	RNA003	9.2
QKUEN008AL	R3012	RNA004	9.2
QKUEN009AT	R3013	RNA005	8

We are using the barcodes we received from the download of our newly created experiment. We would like to add new names for these RNA samples as well as the “RNA Integrity Number” (RIN), that has now been measured in the lab.
Notice how the column names in the header are unimportant: matching to a specific property will happen after the upload. (The old “Analyte ID” column is not important either, it was just added for clarity.)



After the upload, existing Sample IDs are automatically recognized (blue column). Other information is displayed, but has to be matched to the attribute of choice or ignored:

1. Since we don't care about "Analyte ID" (it's already saved), we select "IGNORE" from the dropdown menu. (We can reverse this choice clicking on "Reset columns", if it was a mistake.)
2. For the new column "info1" we select "Secondary Name".
3. For the column "RIN" we select "RNA Integrity Number" (we can only do this, since this attribute type is expected for this sample type.)

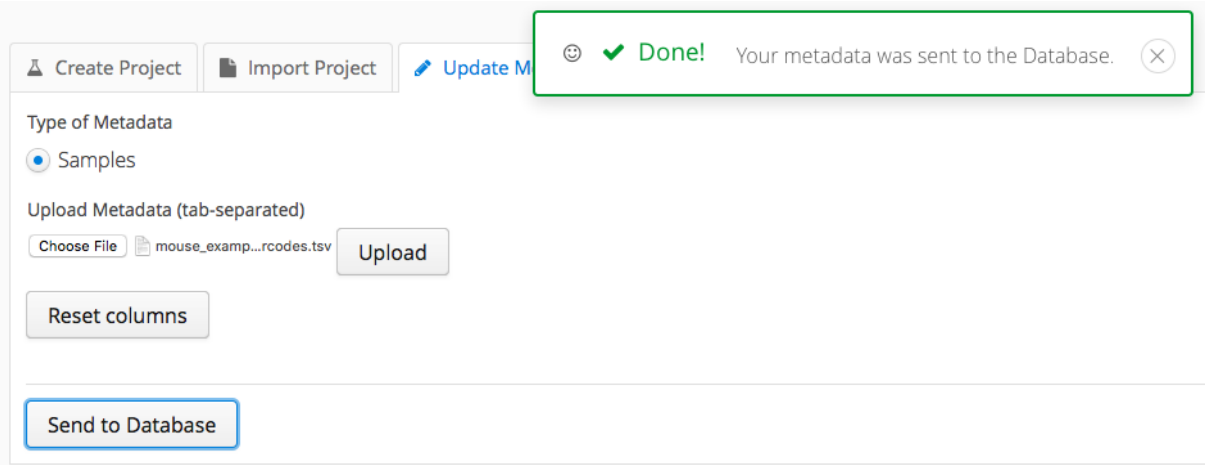


In the resulting table, a few things have happened:

1. the "Analyte ID" column was removed like we wanted
2. the "info 1" column is yellow - data would be overwritten, if we save it as "Secondary name"
3. the "RIN" column is blue - we can savely send this data to the database

Since there are not red columns left, we also get a popup summary of the status: there are collisions for all 5 samples.

We can either chose a different attribute to save this column (for example: “Additional Information”), overwrite it or change our mind and only upload the values for the RNA Integrity Number for now (IGNORE column).
In this case, we choose to save all the data and overwrite the old names by clicking “Send to Database”:



5.1 Special Properties

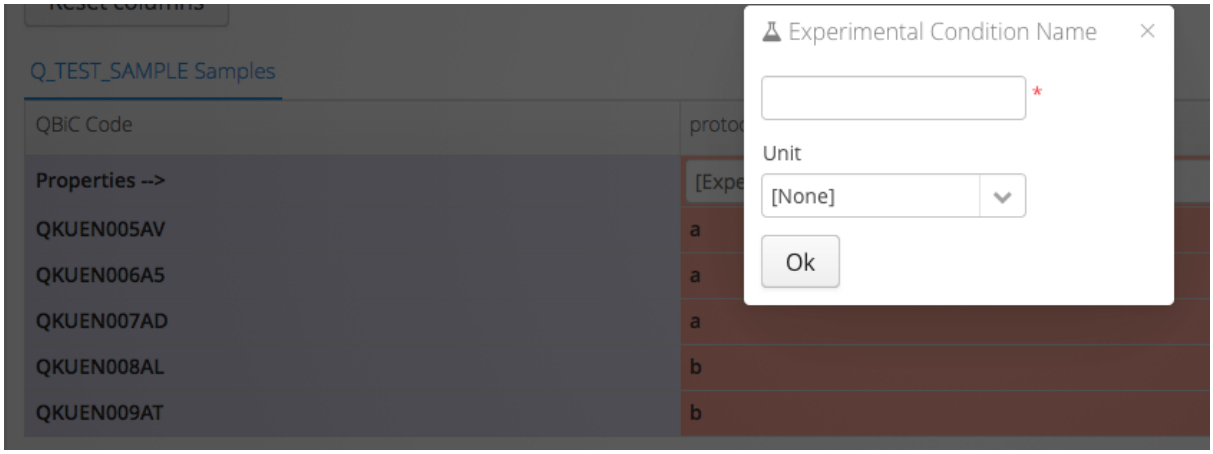
There are a number of different properties that need some more information to update.

Experimental Factors/Conditions

As described in previous chapters, one main idea of qWizard is to enable creation of factorial experimental designs. While it is best to add experimental factors directly via step-by-step registration or the batch upload, they can also be added or edited via Metadata upload:

barcode	protocol
QKUEN005AV	a
QKUEN006A5	a
QKUEN007AD	a
QKUEN008AL	b
QKUEN009AT	b

In this case, we want to add the factor “protocol” (see Chapter 4.2) to our RNA samples from before.
Similar to before, we upload our file and choose [Experimental Condition] from the dropdown menu:



We are then asked for the name of this factor, as well as a clarification what type of unit (if any) this condition has.
After input of the name “protocol” we accept. This time no collisions are found:

Q_TEST_SAMPLE Samples

QBiC Code	protocol
Properties -->	Condition: protocol
QKUEN005AV	**a**
QKUEN006A5	**a**
QKUEN007AD	**a**
QKUEN008AL	**b**
QKUEN009AT	**b**

Send to Database

We can now send this information to the database.

Experimental Properties

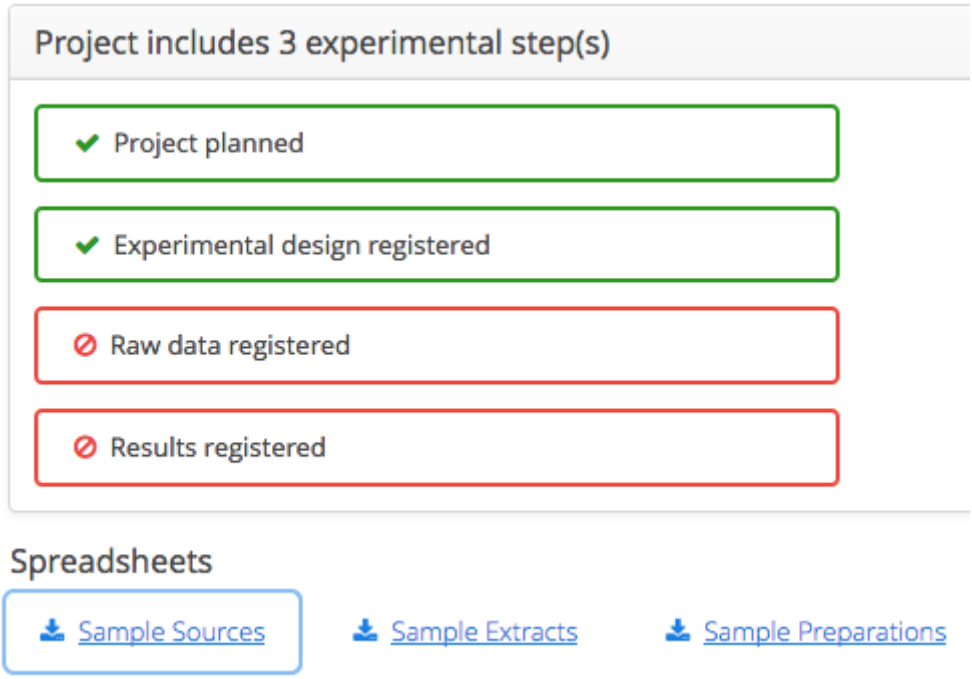
Experimental properties are handled in the exact same way as factors/conditions. The only difference is their importance to the experiment: properties can be used to store information that is not the main focus of the project, but should be saved somewhere.
Properties can be added by selecting “[Other Property]” from the dropdown menu.

Species, Tissues and Analytes

Since species, tissues and analytes are based on a “Controlled Vocabulary” of values, updating them with “free text” can be difficult. qWizard solves this problem by allowing users to choose which value from these vocabularies fits to the uploaded information.

As an example, we want to change the species of one of our aforementioned mice to rat (we always thought it looked kind of strange).

For now, we have only made changes to the analyte (RNA) level of our experiment. To change the species, we need the identifiers of the “Sample Source” level. We can for example find this information in the info page of our sub-project in the Project Browser:

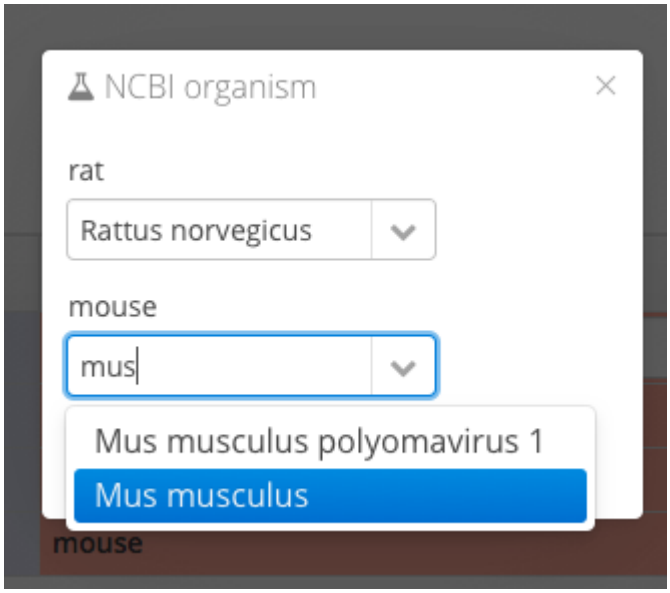


We upload the following table:

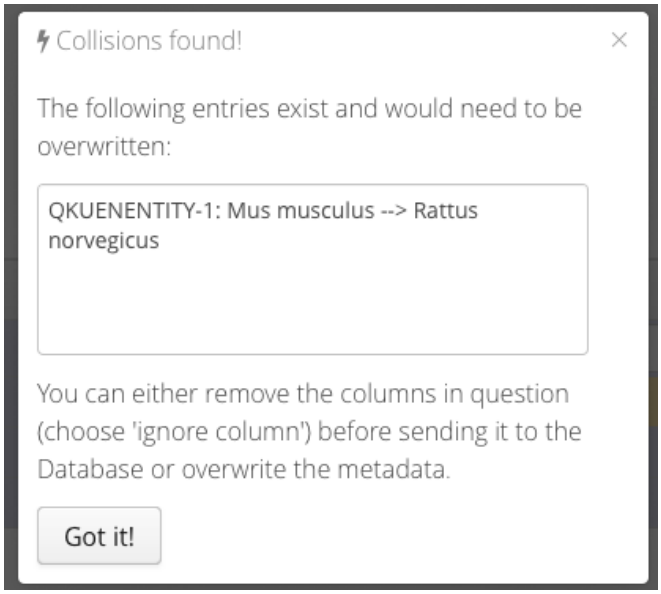
animals	species
QKUENENTITY-1	rat
QKUENENTITY-2	mouse
QKUENENTITY-3	mouse

qWizard recognizes the identifiers as belonging to sample sources of this project. We can now select “NCBI Organism” from the dropdown menu:

A popup asks us to assign the names from the vocabulary for both rat and mouse:



After confirming our selection, we get a summary:



Entity 2 and 3 are already mice, so only Entity 1 needs to be overwritten. We finish the update by clicking “Send to Database”.