



PROJECT DELIVERABLE REPORT

DELIVERABLE NUMBER AND TITLE	D3.6.1
TITLE	REPORT ON THE TESTING ACTIVITIES
AUTHOR(S)	G. FELICI, F. TAGLINO, A. INGROSSO, R. LOMBARDO, F. VITALI
WORK PACKAGE	WP 3
TASK	TASK 3.6
WP LEADER	R. LOMBARDO
BENEFICIARIES CONTRIBUTING TO THE DELIVERABLE	JPI MEMBERS, CNR-IASI, COSBI, CNR-IBIMET
STATUS – VERSION	FINAL - VERSION 1.0
DELIVERY DATE (MONTH)	M18
SUBMISSION DATE	M26
DISSEMINATION LEVEL – SECURITY*	CO
DELIVERABLE TYPE**	R

* Security: PU – *Public*; PP – *Restricted to other programme participants (including JPI Services)*; RE – *Restricted to a group specified by the consortium (including JPI Services)*; CO – *Confidential, only for members of the consortium (including JPI Services)*

** Type: R – *Report*; P – *Prototype*; D – *Demonstrator*; - O - *Other*



CONTENTS

INTRODUCTION	4
LARGE DATASET GENERATOR.....	4
SCENARIO 1. IMPORTING DATA IN dbNP	5
Test Case 1.1 – dbNP dataset loading: 15000 records without wrong values.....	6
Test Case 1.2 – dbNP dataset loading: 150000 records without wrong values.....	6
Test Case 1.3 – dbNP dataset loading: 200000 records without wrong values.....	7
Test Case 1.4 – dbNP dataset loading: 250000 records without wrong values.....	7
Test Case 1.5 – dbNP dataset loading: 15000 records with Not Available values	8
Test Case 1.6 – dbNP dataset loading: 15000 records with invalid values types	8
Test Case 1.7 – dbNP dataset loading: 15000 records with wrong structure.....	9
SCENARIO 2. EXPORTING DATA FROM dbNP Test.....	10
RECAP FOR DBNP.....	10
SCENARIO 3. IMPORTING DATA IN OPAL.....	11
Test Case 3.1 – OPAL dataset loading: upload and insert 25K records without wrong values.....	12
Test Case 3.2 – OPAL dataset loading: upload and insert 250K records without wrong values.....	13
Test Case 3.3 – OPAL dataset loading: upload and insert 300K records without wrong values.....	13
Test Case 3.4 – OPAL dataset loading: upload and insert 350K records without wrong values.....	14
Test Case 3.5 – OPAL dataset loading: upload and insert 25K records with incorrect values types	14
Test Case 3.6 – OPAL dataset loading: upload and insert 25K records with not assign values	14
Test Case 3.7 – OPAL dataset loading: upload and insert 25K records with a wrong data structure.....	15
SCENARIO 4. ExportING DATA FROM OPAL.....	15
Test Case 4.1 – OPAL DATASET EXPORT: export 25K RECORDS and download file.....	15
Test Case 4.2 – OPAL DATASET EXPORT: export 250K RECORDS and download file.....	16
Test Case 4.3 – OPAL DATASET EXPORT: export 300K RECORDS and download file.....	16
Test Case 4.4 – OPAL DATASET EXPORT: export 350K RECORDS and download file.....	17



JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

Conclusions to Scenario 4 and 5	17
SCENARIO 5. MICA TESTING ACTIVITIES	17
SCENARIO 6. DASH-IN FEDERATED ANALISYS TESTING ACTIVITIES	19
SCENARIO 7. MONITORING SERVERS RELIABILITY	21
ANNEX 1. DBNP GUIDE	25
ANNEX 2. OPAL GUIDE	52
ANNEX 3. MICA GUIDE	67



INTRODUCTION

The objective of this document is to report about the activities, within the Task 3.6, aimed at testing the software infrastructure released by the project. These activities were focused on testing the stability of the different systems used within the DASH-IN infrastructure.

The activities have been organized in accordance with seven scenarios, each of which grouping a set of test cases. The list of the scenarios is reported below:

- **Scenarios 1 and 2** are focused on the Phenotype database (www.dbnp.org) through the execution of workflows to perform stress tests on import and export tasks, respectively.
- **Scenarios 3 and 4** are focused on OPAL database through the execution of workflows to perform stress tests on import and export tasks, respectively.
- **Scenario 5** is focused on testing the MICA infrastructure
- **Scenario 6** is focused on testing DASH-IN Federated Analyses through the execution of new instance connection and exploration of main functions
- **Scenario 7** is about testing quality of service through the execution of remote monitoring

For performing test cases pertaining to scenarios 1-4 a number of datasets have been automatically created. The description of the implemented procedure “Large Dataset Generator” is described as first.

All the testing activities were executed on server instances available at UNIBA and Cosbi.

LARGE DATASET GENERATOR

The aim of this task is to stress the infrastructure and in particular to prove that it is strong enough to manage huge datasets with hundred thousand or million records. To make this possible in a short time, we decided to develop a R application that takes a realistic dataset as input, and generates a new dataset based on columns values.

A generated dataset dimension, which is a parameter defined by user, specifies the total number of records created in the new dataset. The application recognizes the column type and chooses different actions for each one; recognized types are: *character*, *integer* and *double*.

- If the type is *character*, the application looks for unique values, and use them as a source to randomly fill the new column in the new dataset.
- If the type is *integer*, it looks for minimum and maximum values and generate a sequence of integer with an increment equal to one. As said previously, also in this case, this sequence is used as a source to randomly generate values in the new dataset.
- At last, if the type is *double*, the application performs the same action done in the case of *integer* type, but with a little difference, in this case we consider the decimal digits in the increment. As seen in the other case, the sequence is used as a source to randomly generate values in the new dataset.

The approach used by the application leaves out existing relations between columns in a realistic scenario. It is possible that two or more columns' values have some kind of relation, for example in the case of age and height, it is quite unfeasible that a baby is taller than an adult. For the sake of our objectives this is not a relevant issue, since the tests on data loading intend to assess if the infrastructure is reliable enough to



manage huge amount of data. As such, the tests themselves do not require any constraint on the meaning of data.

So far, we discussed about the creation of a dataset compliant with a given data schema, but we need also to test the behavior in the case of malformed datasets, e.g., with not assigned values, having values whose type is different than the expected one, with an unknown data structure.

The application is able to manage all these cases through three different functions enabled by parameters. There is a parameter to generate a dataset with randomly not assigned value, and another one that generates a dataset with different types of value per column. In both cases, it is also possible to specify the percentage of wrong values, with respect to the overall dimension of the generated dataset. An additional parameter, if true, makes column label anonymous, making the data difficult to be imported.

SCENARIO 1. IMPORTING DATA IN DBNP

The Phenotype Database Infrastructure, ([www.dbNP¹.org](http://www.dbNP1.org)), is an open source application suite to store biological studies.

The first scenario addresses the import of data in the Phenotype Database instance. Seven test cases have been grouped under this scenario. They differ for the imported dataset. The datasets generated for the first four test cases (namely, TC1.1-TC1.4) are *valid* datasets that differ only for the number of records, which increase from 15,000 in the case of TC1.1, to 250,000 in the case of TC1.4. The datasets for the next two test cases (namely, TC1.5-TC1.7) are *not valid* datasets, and present different kinds of errors.

All the seven test cases under this scenario repeat the same workflow, which performs first the *upload* of the dataset, then the *validation* of the dataset, and finally, at least for the first four test cases that work on valid datasets, the *import* of the data. Details on how to perform all the steps are reported in the Phenotype Database guide attached to this document (see annex 1).

The test cases concerning the importing data scenario have two objectives:

- To measure the time performance of the three intermediate steps, i.e., upload, validate, and import.
- To assess the capability of Phenotype Database to properly load data, if data are valid, and to properly communicate the failure, if data are not valid.

The upload, validation and import steps have been triggered through the graphical user interface of the Phenotype Database client, since Phenotype Database does not provide API (application program interface) to perform the three steps programmatically. To properly perform the time measurement, we used a Chrome plugin called "Performance-Analyser". It gave us the exact time spent to upload, or validate, or import a dataset. All times reported below were captured with this plugin.

¹ <http://www.dbnp.org/>



TEST CASE 1.1 – DBNP DATASET LOADING: 15000 RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in a Phenotype Database instance, of a dataset with 15,000 records and no wrong values with respect to the types in data schema (template). The expected result for the test case is the import of the whole dataset with no errors. Table 1 recaps the execution of the Test Case 1.1.

Table 1: Information about the elaboration of Test Case 1.1

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.1
Description	15K records without wrong values
Code	TC1.1
Start Date	10/01/2017
Study Type	Human Intervention
Objectives	Import dataset performance test
Dataset name	Subjects_Data_15000.csv
Dimension	1060 KB
Upload time	3044 ms
Upload rate	~348.23 bytes/ms
Validation result message	"Your data looks OK. Your data was validated and no errors were found. You can safely continue importing this data."
Validation time	96476 ms
Records/s	155.44
Import result message	"Data imported successfully. 15000 line(s) from your file have been successfully imported and no errors"
Import time	~1200 sec (~20min)

TEST CASE 1.2 – DBNP DATASET LOADING: 150000 RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the Phenotype Database instance, of a dataset with 150,000 records and no wrong values with respect to the types in data schema (template). The expected result for the test case is the import of the whole dataset with no errors. Table 2 recaps the execution of the Test Case 1.2.

Table 2: Information about the elaboration of Test Case 1.2

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.2
Description	150K records without wrong values
Code	TC1.2
Start Date	10/01/2017
Study Type	Human Intervention
Objectives	Import dataset performance test
Dataset name	Subjects_Data_150000.csv
Dimension	10728 KB
Upload time	30364 ms
Upload time rate	~353.31 bytes/ms
Validation result message	"Your data looks OK. Your data was validated and no errors were found. You can safely continue importing this data."



Validation time	964760 ms
Records/s	155.6
Import result message	“Data imported successfully. 150000 line(s) from your file have been successfully imported and no errors”
Import time	~12000 sec (~3h20min)

TEST CASE 1.3 – DBNP DATASET LOADING: 200000 RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the Phenotype Database instance, of a dataset with 200,000 records and no wrong values with respect to the types in data schema (template). The expected result for the test case is the import of the whole dataset with no errors. Table 3 recaps the execution of the Test Case 1.3.

Table 3: Information about the elaboration of Test Case 1.3

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.3
Description	200K records without wrong values
Code	TC1.3
Start Date	10/01/2017
Study Type	Human Intervention
Objectives	Import dataset performance test
Dataset name	Subjects_Data_200000.csv
Dimension	14342 KB
Upload time	41656 ms
Upload time rate	~344.30 bytes/ms
Validation result message	“Your data looks OK. Your data was validated and no errors were found. You can safely continue importing this data.”
Validation time	~1197936 ms
Records/s	~167.08
Import result message	“Data imported successfully. 200000 line(s) from your file have been successfully imported and no errors”
Import time	~16000 sec (~4h 26min)

TEST CASE 1.4 – DBNP DATASET LOADING: 250000 RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the Phenotype Database instance, of a dataset with 250,000 records and no wrong values with respect to the types in data schema (template). The expected result for the test case is the import of the whole dataset with no errors. Table 4 recaps the execution of the Test Case 1.4.

Table 4: Information about the elaboration of Test Case 1.4

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.4
Description	250K records without wrong values
Code	TC1.4
Start Date	10/01/2017
Study Type	Human Intervention

Objectives	Import dataset performance test
Dataset name	Subjects_Data_250000.csv
Dimension	17952 KB
Upload time	51319 ms
Upload time rate	~349.81 bytes/ms
Validation result message	“Your data looks OK. Your data was validated and no errors were found. You can safely continue importing this data.”
Validation time	~1565144 ms
Records/s	~159.73
Import result message	“Data imported successfully. 250000 line(s) from your file have been successfully imported and no errors”
Import time	~20000 sec (~5h 33min)

TEST CASE 1.5 – DBNP DATASET LOADING: 15000 RECORDS WITH NOT AVAILABLE VALUES

This test case concerns the loading, in the Phenotype Database instance, of a dataset with 15,000 records some of which containing NA values. The expected result for the test case is a proper failure message during the validation step. Table 5 recaps the execution of the Test Case 1.5.

Table 5: Information about the elaboration of Test Case 1.5

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.5
Description	15K records with NA values
Code	TC1.5
Start Date	10/01/2017
Study Type	Human Intervention
Objectives	Import dataset performance test
Dataset name	Subjects_Data_15000_withNA.csv
Dimension	1024 KB
Upload time	3015 ms
Upload time rate	~339.63 bytes/ms
Validation result message	“The following problems were found when validating your data: [3] Error casting Hip circumference of type Double with value NA (class java.lang.String) :: For input string: "NA" for column 8 ...”*
Validation time	96676 ms
Records/s	~155.16

* As expected the validation process reports error in each cell that contains a NA value.

TEST CASE 1.6 – DBNP DATASET LOADING: 15000 RECORDS WITH INVALID VALUES TYPES

This test case concerns the loading, in the Phenotype Database instance, of a dataset with 15,000 records some of which containing invalid values with respect to the types of the data schema (template). The expected result for the test case is a proper failure message during the validation step. Table 6 recaps the execution of the Test Case 1.6.

Table 6: Information about the elaboration of Test Case 1.6

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.6
Description	15K records with invalid values types
Code	TC1.6
Start Date	10/01/2017
Study Type	Human Intervention
Objectives	Import dataset performance test
Dataset name	Subjects_Data_15000_withIVT.csv
Dimension	1034 KB
Upload time	2977 ms
Upload time rate	~347.3 bytes/ms
Validation result message	<p>“The following problems were found when validating your data:</p> <p>[3] Error casting Body height of type Double with value m (class java.lang.String) :: For input string: "m" for column 7</p> <p>[3] Error casting Waist hip ratio of type Double with value f (class java.lang.String) :: For input string: "f" for column 10</p> <p>[3] Error casting Waist hip ratio of type Double with value j (class java.lang.String) :: For input string: "j" for column 10 ...”*</p>
Validation time	97073 ms
Records/s	~154.52

*As expected the validation process recognizes types mismatch where values types don't match with column type.

TEST CASE 1.7 – DBNP DATASET LOADING: 15000 RECORDS WITH WRONG STRUCTURE

This test case concerns the loading, in the Phenotype Database instance, of a dataset with 15,000 records with a wrong structure. The expected result for the test case is a proper failure message during the validation step. Table 7 recaps the execution of the Test Case 1.7.

Table 7: Information about the elaboration of Test Case 1.7

Type of info	Value
Template	Intervention/Observation study
Title	Test Case 1.7
Description	15K records with wrong structure
Code	TC1.6
Start Date	10/01/2017
Study Type	Human Intervention
Objectives	Import dataset performance test
Dataset name	Subjects_Data_15000_withRNM.csv
Dimension	1059 KB
Upload time	3035 ms
Upload time rate	~348.93 bytes/ms
Validation result message	Match data process doesn't recognize the structure and it is necessary a manual action to assign to each column a consistent label.



SCENARIO 2. EXPORTING DATA FROM DBNP TEST

In this testing activities, we have used the API interface available on Phenotype Database². The client application was built in R using GSCFClient³ library and the code contained few statements. Everything starts with a connection to the remote Phenotype Database instance, after that the application tries to download for each study its subject, every download is timed and each data stream sized.

Unfortunately, this test has failed to obtain Studies, the `getStudies()` method ended with the following error:

```
Error in .getUrlErr(url) : <html>
<head>
<title>Grails Runtime Exception</title>
<style type="text/css">
  .message {
    border: 1px solid black;
    padding: 5px;
    background-color: #E9E9E9;
  }
  .stack {
    border: 1px solid black;
    padding: 5px;
    overflow:auto;
    height: 300px;
  }
  .snippet {
    padding: 5px;
    background-color:white;
    border:1px solid black;
    margin:3px;
    font-family:courier;
  }
</style>
</head>

<body>
<h1>Grails Runtime Exception</h1>
<h2>Error Details</h2>

<div class="message">
<strong>Error 500:</strong> Executing action [getStudies] of controller
[api.ApiController] caused exception: Runtime error executing action<br/>
<strong>Servlet:</strong> grails<br/>
<strong>URI:</strong> /gscf/grails/api/getStudies.dispatch<br/>

<strong>Exception Message:</strong> Executing action [getStudies] of controller
[api.ApiController] caused exception: Runtime error executing action
```

RECAP FOR DBNP

Table 8: Recap of the results

Dataset	Upload (B/ms)	Import (R/m)	Export (R/m)	Download (B/ms)
Subjects_Data_15000	384.23	750	FAILED	FAILED
Subjects_Data_150000	353.31	750	FAILED	FAILED
Subjects_Data_200000	344.30	750	FAILED	FAILED
Subjects_Data_250000	349.81	750	FAILED	FAILED

² <http://dbnp.cloud.ba.infn.it/api>

³ <https://github.com/PhenotypeFoundation/GSCF-RClient>



Below are summarized the results achieved:

- Uploading performances are on average equal even though datasets have important difference in number of records;
- Upload streams never went below 339 Bytes/ms;
- Validating process recognizes both failure cases, with not assign values and with wrong types values per column;
- Import process recognize a dataset with a wrong structure and ask to the user to manually match the correct column label;
- Datasets with correct structure, without not assigned values and without types mismatch values, passed validation and import process;

SCENARIO 3. IMPORTING DATA IN OPAL

Obiba⁴ software consists of a suite of stand-alone applications that support various study's data management activities. Opal is the Obiba's core database application for epidemiological studies. Using Opal, studies can import, validate, derive, query, report, analyze and export data.

This scenario addresses the import of data in the OPAL instance by grouping seven test cases. They differ for the imported dataset. The datasets generated for the first four test cases (namely, TC3.1-TC3.4) are *valid* datasets that differ only for the number of records, which increase from 25,000 in the case of TC3.1, to 350,000 in the case of TC3.4. The datasets for the next three test cases (namely, TC3.5-TC3.7) are *not valid* datasets, and present different kinds of errors.

All the seven test cases under this scenario repeat the same workflow, which is based on the Opal guide attached to this document (see annex 2). We were built six Test Case, three of them end with success, and the others end with expected failures.

We started by creating a new project in the OPAL instance, namely "Test case" with code TC1. We used this project to test all cases, because each case is made of a table with the same dictionary (data schema) but different contents.

Project form was filled with the following information:

Project name: TC1

Title: Test Case

Description: Import dataset performance test

Once that project was created, we prepare seven different tables with the same data dictionary and we have called them as follows:

- Test_Data_25K, this table contains 25K records well formed;
- Test_Data_250K, this table contains 250K records well formed;
- Test_Data_300K, this table contains 300K records well formed;
- Test_Data_350K, this table contains 350K records well formed;

⁴ <http://www.obiba.org>

- Test_Data_25K_withIVT, this table contains 25K records with a certain number of incorrect values types;
- Test_Data_25K_withNA, this table contains 25K records with a certain number of not assign values;
- Test_Data_25K_withRNM, this table contains 25K records in a structure with misleading labels.

These activities were done following the add table dictionary procedure (see Figure 1).

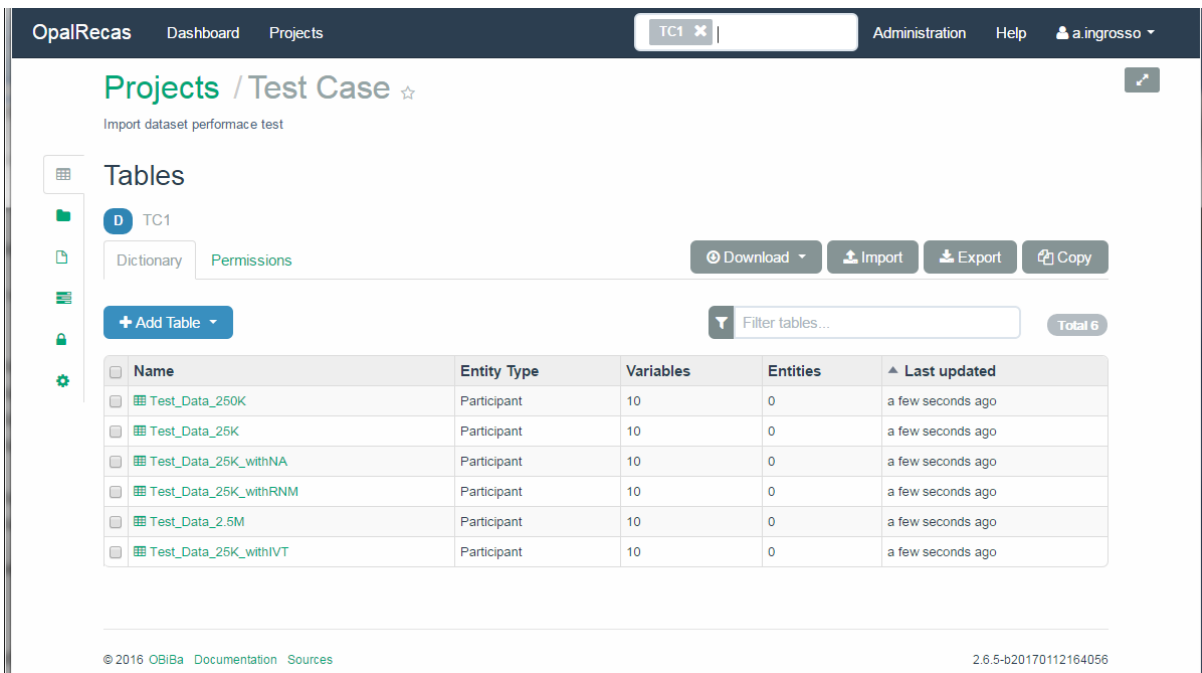


Figure 1: Screenshot of the OPAL web client to manage tables in a project

After that, we used import and export API made available by Opal to load datasets on it. We developed an R application that invokes the API commands and sends them to Opal. We measured the time necessary to upload files.

TEST CASE 3.1 – OPAL DATASET LOADING: UPLOAD AND INSERT 25K RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the OPAL instance, of a dataset with 25,000 records and no wrong values with respect to the types in data schema. The expected result for the test case is the import of the whole dataset with no errors. Table 8 recaps the execution of the Test Case 3.1.

Table 9: Information about the elaboration of Test Case 3.1

Type of info	Value
Description	25K records without wrong values
Code	TC3.1
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_25000.csv
Dimension	1073.407 KB



Upload time	3512.527 ms
Upload time rate	~305.59 bytes/ms
Import time	25 min
Records/m	1000

TEST CASE 3.2 – OPAL DATASET LOADING: UPLOAD AND INSERT 250K RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the OPAL instance, of a dataset with 250,000 records and no wrong values with respect to the types in data schema. The expected result for the test case is the import of the whole dataset with no errors. Table 9 recaps the execution of the Test Case 3.2.

Table 10: Information about the elaboration of Test Case 3.2

Type of info	Value
Description	250K records without wrong values
Code	TC3.2
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_250000.csv
Dimension	10986.473 KB
Upload time	31145.9 ms
Upload time rate	~352.74 bytes/ms
Import time	221 min
Records/m	1131.22

TEST CASE 3.3 – OPAL DATASET LOADING: UPLOAD AND INSERT 300K RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the OPAL instance, of a dataset with 300,000 records and no wrong values with respect to the types in data schema. The expected result for the test case is the import of the whole dataset with no errors. Table 10 recaps the execution of the Test Case 3.3.

Table 11: Information about the elaboration of Test Case 3.3

Type of info	Value
Description	300K records without wrong values
Code	TC3.3
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_300000.csv
Dimension	25287123KB
Upload time	70126.30 ms
Upload time rate	~360.59 bytes/ms
Import time	~190 min
Records/m	1578.95

TEST CASE 3.4 – OPAL DATASET LOADING: UPLOAD AND INSERT 350K RECORDS WITHOUT WRONG VALUES

This test case concerns the loading, in the OPAL instance, of a dataset with 350,000 records and no wrong values with respect to the types in data schema. The expected result for the test case is the import of the whole dataset with no errors. Table 11 recaps the execution of the Test Case 3.4.

Table 12: Information about the elaboration of Test Case 3.4

Type of info	Value
Description	350K records without wrong values
Code	TC3.4
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_350000.csv
Size	29521114 KB
Upload time	81975.91 ms
Upload time rate	~360.12 bytes/ms
Import time	~125 min
Records/m	~2800

TEST CASE 3.5 – OPAL DATASET LOADING: UPLOAD AND INSERT 25K RECORDS WITH INCORRECT VALUES TYPES

This test case concerns the loading, in the OPAL instance, of a dataset with 25,000 records containing invalid values with respect to the types of the data schema. The expected result for the test case is a proper failure message during the import task. Table 12 recaps the execution of the Test Case 3.5.

Table 13: Information about the elaboration of Test Case 3.5

Type of info	Value
Description	25K records with invalid values
Code	TC3.5
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_25000_withIVT.csv
Size	1490.396 KB
Upload time	4929.135 ms
Upload time rate	~302.36 bytes/ms
Import message	"Unable to get value for entity ID1 and variable PITUUS: not a decimal value: n"

TEST CASE 3.6 – OPAL DATASET LOADING: UPLOAD AND INSERT 25K RECORDS WITH NOT ASSIGN VALUES

This test case concerns the loading, in the OPAL instance, of a dataset with 25,000 records containing not assigned (NA) values. The expected result for the test case is a proper failure message during the import task. Table 13 recaps the execution of the Test Case 3.6.

Table 14: Information about the elaboration of Test Case 3.6

Type of info	Value
Description	25K records with not assigned values
Code	TC3.6
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_25000_withNA.csv
Size	1062.347 KB
Upload time	3556.266 ms
Upload time rate	~298.72 bytes/ms
Import message	"Unable to get value for entity ID6 and variable DIAS2: not a integer value: NA"

TEST CASE 3.7 – OPAL DATASET LOADING: UPLOAD AND INSERT 25K RECORDS WITH A WRONG DATA STRUCTURE

This test case concerns the loading, in the OPAL instance, of a dataset with 25,000 records with a wrong data structure. The expected result for the test case is a proper message during the import task. Table 14 recaps the execution of the Test Case 3.7.

Table 15: Information about the elaboration of Test Case 3.7

Type of info	Value
Description	25K records with wrong data structure
Code	TC3.7
Start Date	13/01/2017
Objectives	Import dataset performance test
Dataset name	Opal_Test_Data_25000_withRN.csv
Size	1073.494 KB
Upload time	3909.196 ms
Upload time rate	~274.61
Import time	17 min
Records/m	~1470.59
Comments	"Import process ends with success but there was no matching with table dictionary, all columns were imported as new with value type text."

SCENARIO 4. EXPORTING DATA FROM OPAL

This scenario addresses the extraction of data from the OPAL instance through four test cases that differ for the size of the exported datasets.

TEST CASE 4.1 – OPAL DATASET EXPORT: EXPORT 25K RECORDS AND DOWNLOAD FILE

This test case extracts and download 25,000 records. Table 15 recaps the execution of the Test Case 4.1.



Table 16: Information about the elaboration of Test Case 4.1

Type of info	Value
Description	25K records export and download
Code	TC4.1
Start Date	13/01/2017
Objectives	Export and download dataset performance test
Exported file name	Test_Data_25000.csv
Size	1574469 Byte
Export time	1 min
Records/min	25,000
Download time	1645.27 ms
Export time rate	~956.96

TEST CASE 4.2 – OPAL DATASET EXPORT: EXPORT 250K RECORDS AND DOWNLOAD FILE

This test case extracts and download 250,000 records. Table 16 recaps the execution of the Test Case 4.2.

Table 17: Information about the elaboration of Test Case 4.2

Type of info	Value
Description	250K records export and download
Code	TC4.2
Start Date	13/01/2017
Objectives	Export and download dataset performance test
Exported file name	Test_Data_250000.csv
Size	15996331 Byte
Export time	7 min
Records/min	35,714.28
Download time	53551.58 ms
Export time rate	~498.71

TEST CASE 4.3 – OPAL DATASET EXPORT: EXPORT 300K RECORDS AND DOWNLOAD FILE

This test case extracts and download 300,000 records. Table 17 recaps the execution of the Test Case 4.3.

Table 18: Information about the elaboration of Test Case 4.3

Type of info	Value
Description	300K records export and download
Code	TC4.3
Start Date	13/01/2017
Objectives	Export and download dataset performance test
Exported file name	Test_Data_300000.csv
Size	9753370 Byte
Export time	12 min
Records/min	25,000
Download time	22650.84 ms
Export time rate	~430.59

TEST CASE 4.4 – OPAL DATASET EXPORT: EXPORT 350K RECORDS AND DOWNLOAD FILE

This test case extracts and download 350,000 records. Table 18 recaps the execution of the Test Case 4.4.

Table 19: Information about the elaboration of Test Case 4.4

Type of info	Value
Description	350K records export and download
Code	TC4.4
Start Date	13/01/2017
Objectives	Export and download dataset performance test
Exported file name	Test_Data_350000.csv
Size	11397457 Byte
Export time	13 min
Records/min	26,923
Download time	26229.35 ms
Export time rate	~434.53

CONCLUSIONS TO SCENARIO 4 AND 5

In the following table, we compared performance concerning datasets that were involved in upload and download tasks.

Table 20: Recap of the results

Dataset	Upload (B/ms)	Import (R/m)	Export (R/m)	Download (B/ms)
Test_Data_25K	305.59	1000	25000	956.96
Test_Data_250K	352.74	1131.22	35714.28	498.71
Test_Data_300K	360.59	1578.95	25000	430.59
Test_Data_350K	360.12	2800	26923	434.53

Below, the achieved results are summarized:

- Uploading performances are on average equal even though datasets have important difference in the number of records;
- Upload streams never went below 305 Bytes/ms;
- Importing process recognizes both failure cases, with not assign values and with wrong types values per column;
- Importing process recognize a dataset with a wrong structure as a new dataset structure, new attributes were added;
- Datasets with correct structure, without not assigned values and without types mismatch values, passed import process;
- Exporting performance never went below 25000 records per minute;
- Downloading performance never went below 498 Bytes/ms.

SCENARIO 5. MICA TESTING ACTIVITIES



Mica⁵ is a web application used to create web data portals for epidemiological studies or consortia. Mica includes modules to create study catalogs, searchable variable dictionaries, online data access request forms and to query remotely datasets stored in Opal databases.

For Mica testing activities we used the workflow reported in the Mica guide attached to this document (see annex 3). The main objectives are to verify if add a study and add population procedures are successfully completed. The second objective underlay an existing connection with data stored in Opal server that must be visible and potentially combinable in a study.

First step, add a study, we used the following information:

- Name: LifeLines Cohort Study & Biobank (TEST)
- Acronym: llcsb_test
- Objectives: Overall aim of the study: unravel the interaction between genetic and environmental factors in the development of multifactorial diseases...
- Start year: 2004
- End year: 2036
- Study Design: Cohort
- Supplementary information: Funding Agencies: Government, 3 northern provinces of The Netherlands...
- Source of recruitment: Individuals, Families
- Website: <https://www.lifelines.nl/>
- Target number of participants: 165000
- Supplementary information: the project will include 165000 participants...
- Access: data, biosamples
- Marker Paper: ...three generation population-based study. Eur J. Epidemiol 2008...

This first step ended with success, we created a new study called “llcsb_test”. We also added a contact to the study.

Second step, add population, we used the following information:

- ID: pop_test_1
- Name: LifeLines population TEST
- Description: ...(18-65 years) or the elderly cohort (>65 years). The participants are followed for at least 30 years...
- Minimum age: 25
- Target number of participants: 165000
- Supplementary information: current number of participants recruited: 93000 (on June, 2012)...
- Recruitment supplementary information: general population, volunteer enrolment

This second step ended with success, we created a population section in our study.

Third and last step, add data collection events, we used the following information:

⁵ <http://www.obiba.org>



- ID: data_coll_test_1
- Name: Baseline recruitment / initial data collection TEST
- Start year: 2007
- End year: 2013
- Data sources: questionnaires, physical measures, biosamples

This last step ended with success, we created a data collection section in our study.

With the study created, we tested the association of a dataset to study data collection to combine data following the creation process.

We added a new dataset with the following information:

- Name: Test_Data_250K
- Acronym: TD250K
- Description: this table is stored in opal database and we want to create a relation with llcsb_test study
- Entity type: participant

Once the dataset on Mica was created we added study table by filling the form with this information:

- Study: llcsb_test
- Population: pop_test_1
- Data Collection Event: data_coll_test_1
- Data source project: TC1
- Data source table: Test_Data_250K
- Name: TD250K
- Description: test table with 250K records

The association of dataset Test_Data_250K within llcsb_test study was created with success.

Below, the achieved results are summarized:

- Study creation process was successfully completed
- Dataset creation process was successfully completed

SCENARIO 6. DASH-IN FEDERATED ANALISYS TESTING ACTIVITIES

DataSHIELD⁶ enables advanced statistical analysis across a network of Opal databases without the need of pooling and accessing individual-level data. This is a powerful tool for studies that can't share data for common ethical-legal reasons but still want to co-analyse in-depth their data in an extensive way.

In addition to the work done through the usability survey⁷, we decided to test the setting up of data sources. In this case the objective is to verify if the connection to an external Opal server works fine, where works fine

⁶ <http://www.obiba.org>

⁷ <http://survey-xact.dk/LinkCollector?key=YCSEVTT43P11>



means that all Opal data are correctly retrieved and queried. The objective was verified by adding a new Opal server into Dash-in.

Add New Opal Server

Server name: Opal Server BA

Server URL: <http://opal.cloud.ba.infn.it:8080/>

Username and password associated to a.ingrosso account

Result: server added with success.

After this first step we used Analisis tool to verify that was possible to load database and table from Opal remote server. We have selected database project called TC1 and we tested setup remote database procedure with four different tables.

Setup remote database with table called Test_Data_25K

After selecting TC1 data server and table Test_Data_25K we received this message “Data sources were successfully setup on remote servers! Visit the Federated Analysis section in left side-pane”, means that setup task was concluded correctly. Federated Analysis section displayed all data correctly.

Setup remote database with table called Test_Data_250K

After selecting TC1 data server and table Test_Data_250K we received this message “Data sources were successfully setup on remote servers! Visit the Federated Analysis section in left side-pane”, means that setup task was concluded correctly. Federated Analysis section displayed all data correctly.

Setup remote database with table called Test_Data_300K

After selecting TC1 data server and table Test_Data_300K we received this message “Data sources were successfully setup on remote servers! Visit the Federated Analysis section in left side-pane”, means that setup task was concluded correctly. Federated Analysis section displayed all data correctly.

Setup remote database with table called Test_Data_350K

After selecting TC1 data server and table Test_Data_350K we received this message “Data sources were successfully setup on remote servers! Visit the Federated Analysis section in left side-pane”, means that setup task was concluded correctly. Federated Analysis section displayed all data correctly.

Below are summarized the results achieved:

- Server Opal is added correctly into Dash-in application
- Data from Opal server are retrieved correctly into Federated Analysis tool



SCENARIO 7. MONITORING SERVERS RELIABILITY

Server reliability is one of the most important qualities that a server must have to deliver value to users across services. For this reason, we decided to monitor the reliability of services available into the ENPADASI federated infrastructure. In particular, we monitored the following three servers:

- An instance of Opal with URL `opal.cloud.ba.infn.it` on port 8080
- An instance of Mica with URL `micaclient.cloud.ba.infn.it` on port 443
- An instance of DBNP with URL `dbnp.cloud.ba.infn.it` on port 80

To monitor server’s reliability, we used an online tool called Uptime Robot (<https://uptimerobots.com>). This tool monitors services every 5 minutes and records an event if a site is down.

Below, the step-by-step monitoring activity of Uptime Robot is reported:

- It asks for your website headers and gets status code like “200-ok”, “404-not found”, etc. every 5 minutes;
- If the status code doesn’t indicate a problem, we are good;
- If the status code is 400+ and 500+, then the site is not loading;
- In order to make sure the site is down, Uptime Robot makes several more checks in the next 30 seconds;
- If the site is still down, it sends an alert.

Monitoring activity started on December 2nd 2016, and is currently active, since this date Uptime Robot has registered this percentage of reliability:

- Opal: 99.94%
- MICA: 100%
- DBNP: 100%

Below the events registered until January 28th 2017.

Table 21: Events Report

Event	Monitor	Date-time	Reason	Duration	Duration in mins
Up	OPAL BA	19/01/2017 09:44	OK	223 hrs, 26 mins	13407
Down	OPAL BA	19/01/2017 09:34	Timeout	0 hrs, 10 mins	10
Up	OPAL BA	17/01/2017 12:08	OK	45 hrs, 25 mins	2726
Down	OPAL BA	17/01/2017 12:01	Timeout	0 hrs, 6 mins	6
Up	OPAL BA	08/01/2017 06:52	OK	221 hrs, 9 mins	13269
Down	OPAL BA	08/01/2017 06:44	Timeout	0 hrs, 8 mins	8
Up	OPAL BA	05/12/2016 07:01	OK	815 hrs, 43 mins	48943
Down	OPAL BA	05/12/2016 06:53	Timeout	0 hrs, 7 mins	7
Up	OPAL BA	02/12/2016 02:59	OK	75 hrs, 54 mins	4555
Down	OPAL BA	02/12/2016 02:58	Timeout	0 hrs, 0 mins	1
Up	MICA BA	01/12/2016 17:37	OK	1391 hrs, 33 mins	83494
Down	MICA BA	01/12/2016 17:36	Timeout	0 hrs, 0 mins	1
Up	MICA BA	01/12/2016 15:12	OK	2 hrs, 24 mins	144
Down	MICA BA	01/12/2016 15:12	Timeout	0 hrs, 0 mins	1
Up	MICA BA	25/11/2016 10:22	OK	148 hrs, 49 mins	8930

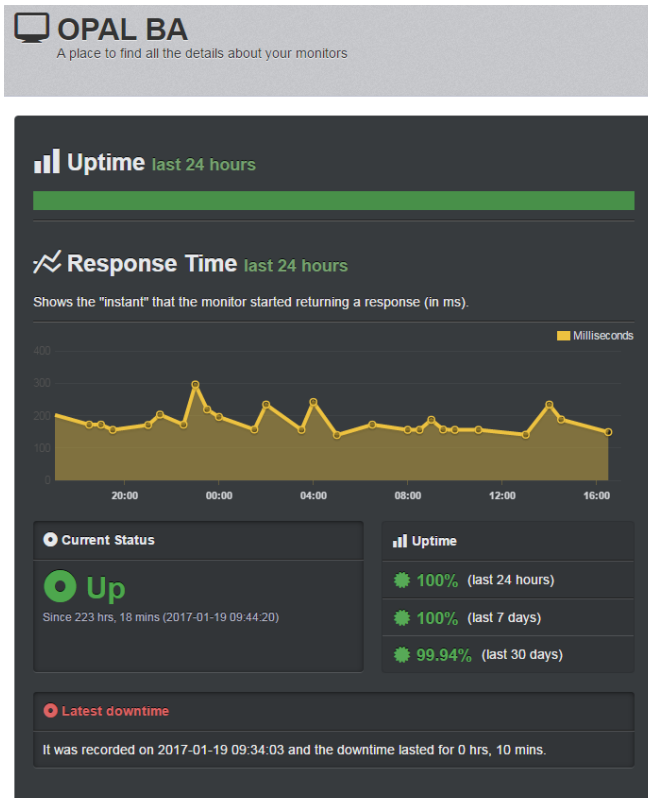


JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

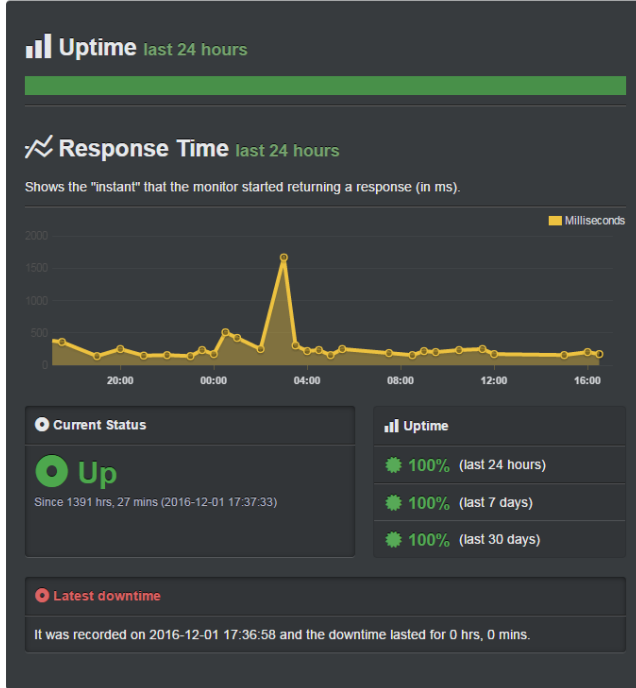
Started	MICA BA	25/11/2016 10:22	Timeout	0 hrs, 0 mins	0
Up	DBNP BA	25/11/2016 10:21	OK	1542 hrs, 49 mins	92569
Started	DBNP BA	25/11/2016 10:21	Timeout	0 hrs, 0 mins	0
Up	OPAL BA	25/11/2016 10:21	OK	160 hrs, 37 mins	9637
Started	OPAL BA	25/11/2016 10:21	Timeout	0 hrs, 0 mins	0

Opal site is the only server for which downtime events have been observed.

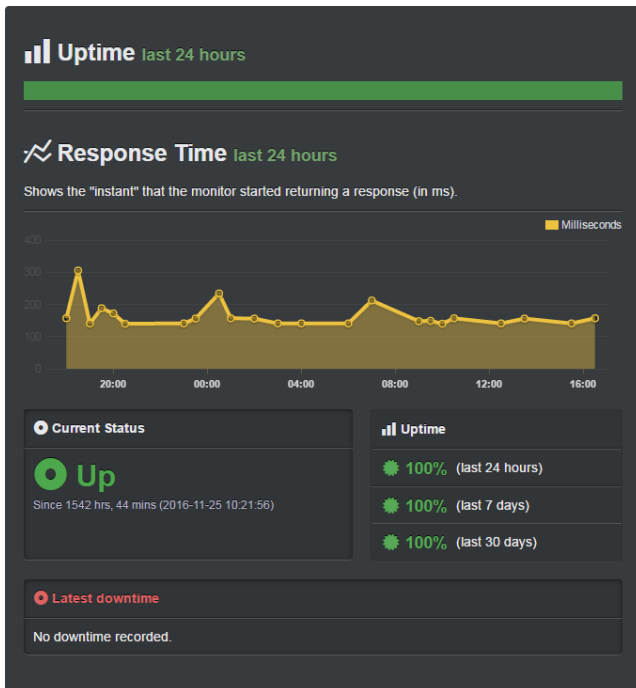
Below are reported single site performance card that are referred to the last 24 hours and are made on January 28th 2017:



MICA BA
A place to find all the details about your monitors



DBNP BA
A place to find all the details about your monitors



Below are summarized the results achieved:



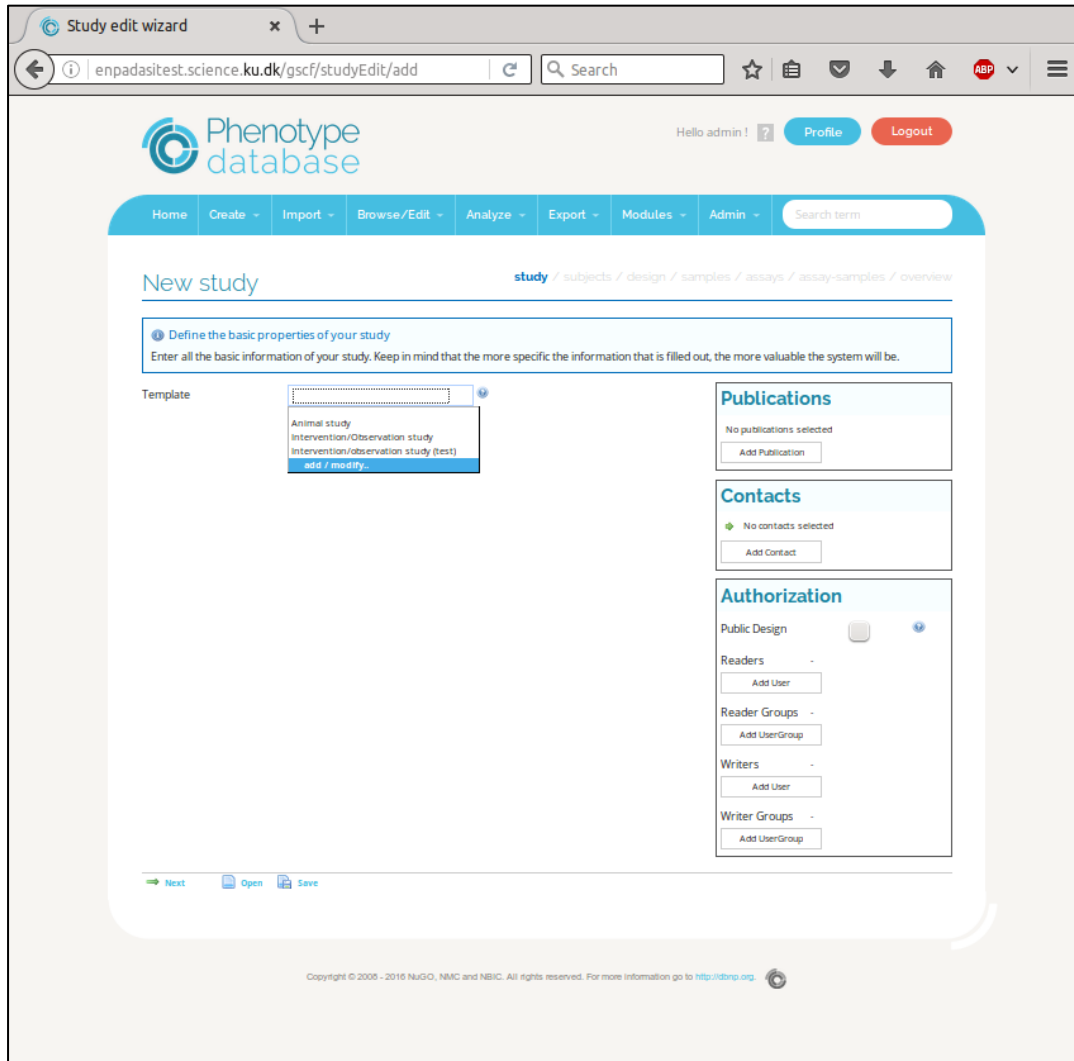
JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

- Reliability percentage never went below 99.9%;
- Every timeout events were closed at most in 10 minutes;
- Response time in the last 24 hours never went over 350ms.



ANNEX 1. DBNP GUIDE

Phenotype Database (dbNP) - create a new study



Study edit wizard

enpadasitest.science.ku.dk/gscf/studyEdit/add

Phenotype database

Hello admin! Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

New study [study](#) / [subjects](#) / [design](#) / [samples](#) / [assays](#) / [assay-samples](#) / [overview](#)

Define the basic properties of your study

Enter all the basic information of your study. Keep in mind that the more specific the information that is filled out, the more valuable the system will be.

Template

- Animal study
- Intervention/Observation study
- Intervention/observation study (test)

add / modify

Publications

No publications selected

Add Publication

Contacts

No contacts selected

Add Contact

Authorization

Public Design

Readers -

Add User

Reader Groups -

Add UserGroup

Writers -

Add User

Writer Groups -

Add UserGroup

Next Open Save

Copyright © 2005 - 2016 NUGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>



The screenshot shows a web browser window titled "Study edit wizard - Mozilla Firefox" with the URL "enpadasitest.science.ku.dk/gscf/studyEdit/propertie". The page header includes the "Phenotype database" logo, a user greeting "Hello admin!", and buttons for "Profile" and "Logout". A navigation menu contains "Home", "Create", "Import", "Browse/Edit", "Analyze", "Export", "Modules", and "Admin", along with a search bar.

The main content area is titled "New study" and includes a breadcrumb trail: "study / subjects / design / samples / assays / assay-samples / overview". A blue box contains the instruction: "Define the basic properties of your study. Enter all the basic information of your study. Keep in mind that the more specific the information that is filled out, the more valuable the system will be."

The form is divided into several sections:

- Template:** A dropdown menu set to "Intervention/observation study (be)".
- Title:** A text input field with the placeholder "This is the Title".
- Description:** A text area with the placeholder "This is the description".
- Code:** A text input field with the placeholder "abcde-code".
- Start Date:** A date picker set to "2010-09-01".
- Study Type:** A dropdown menu set to "Human Intervention".
- Principle Investigator:** A text input field.
- Primary Endpoints:** A text input field.
- Objectives:** A text area with the placeholder "This the objectives".
- Central Conclusion:** A text input field.
- Main Health-Related Outcome:** A text input field.
- Exclusion Criteria:** A text input field.
- Inclusion Criteria:** A text input field.
- Institute:** A text input field.
- Collaborators Involved:** A text input field.
- Country (For Multicentre Study Overall PI):** A text input field.

On the right side, there are three panels:

- Publications:** Shows "No publications selected" and an "Add Publication" button.
- Contacts:** Shows "No contacts selected" and fields for "Person:" and "Role:" with "Add" and "Close" buttons.
- Authorization:** Includes a "Public Design" toggle switch, and sections for "Readers", "Reader Groups", "Writers", and "Writer Groups", each with "Add User" and "Add UserGroup" buttons.



Study edit wizard - Mozilla Firefox

Study edit wizard

enpadasitest.science.ku.dk/gscf/studyEdit/propertie

Phenotype database

Hello admin! Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

New study

study / subjects / design / samples / assays / assay-samples / overview

Define the basic properties of your study

Enter all the basic information of your study. Keep in mind that the more specific the information that is filled out, the more valuable the system will be.

Template: Intervention/observation study (be)

Title: This is the Title

Description: This is the description

Code: abode-code

Start Date: 2010-09-01

Study Type: Human Intervention

Principle Investigator:

Primary Endpoints:

Objectives: This the objectives

Central Conclusion:

Main Health-Related Outcome:

Exclusion Criteria:

Inclusion Criteria:

Institute:

Collaborators Involved:

Country (For Multicentre Study Overall PI):

Publications

No publications selected

Add Publication

Contacts

No contacts selected

Person:

Role: Sando, Finn
Tullin, Mikkel

add / modify persons...

Authorization

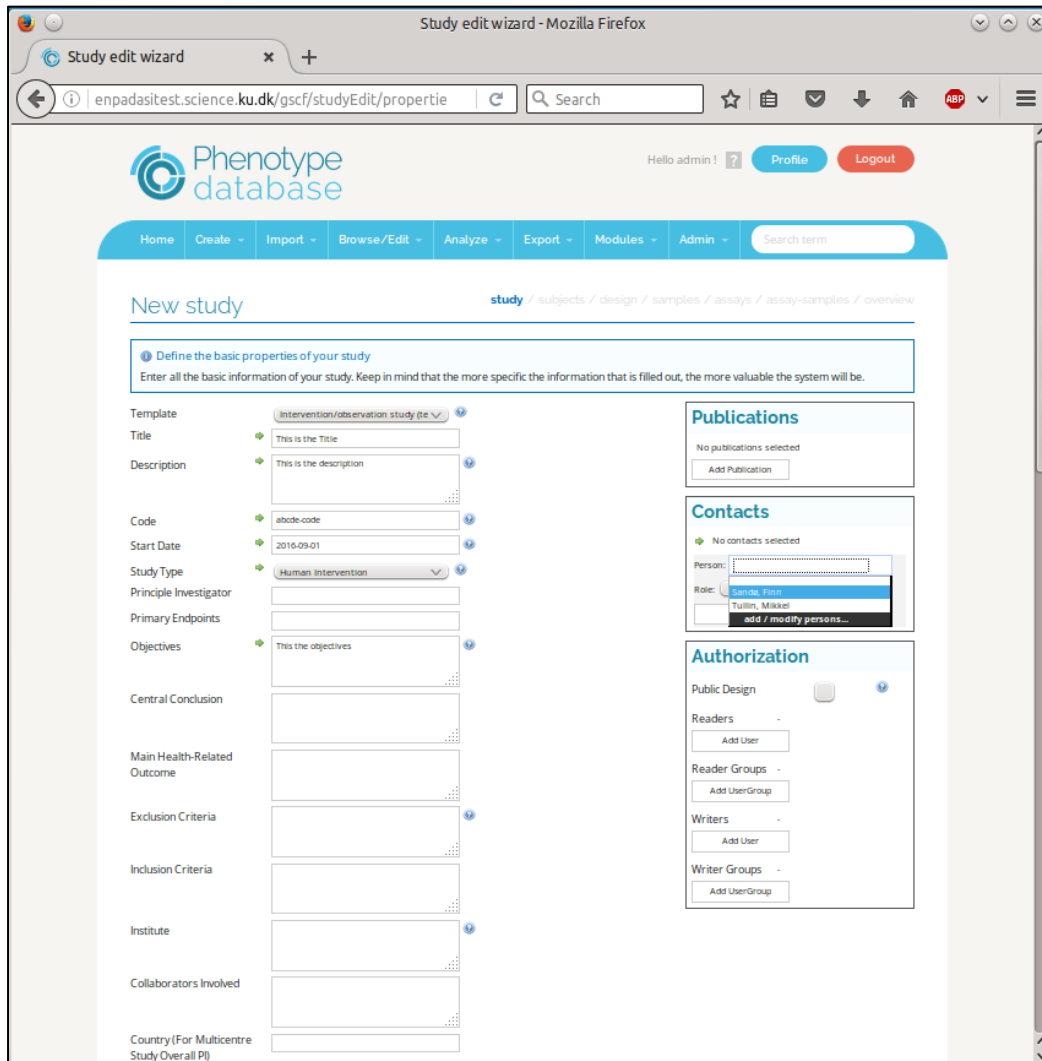
Public Design:

Readers: Add User

Reader Groups: Add UserGroup

Writers: Add User

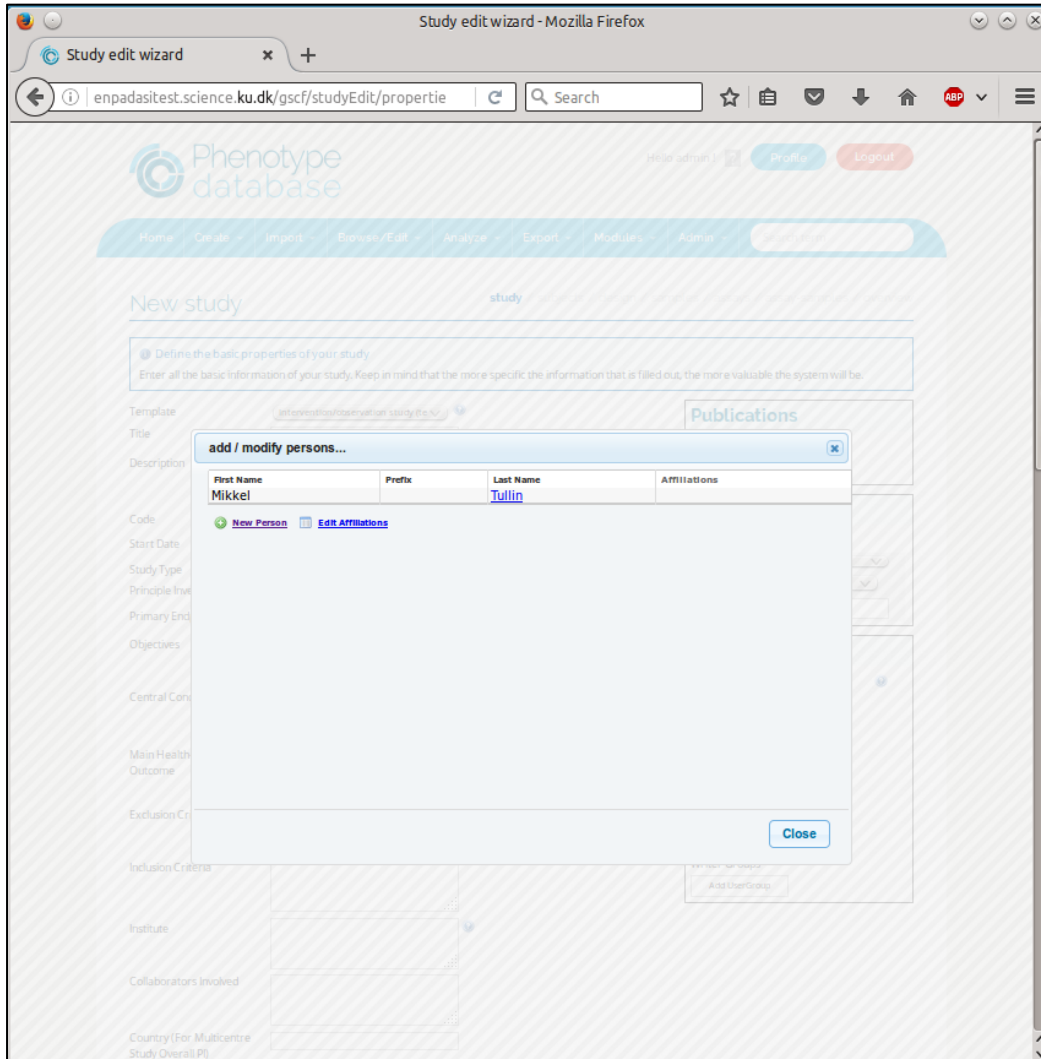
Writer Groups: Add UserGroup

These images show part of the process to create a new study in the Phenotype Database (dbNP).



Phenotype Database (dbNP) - add a person to the person list

Study edit wizard - Mozilla Firefox

Study edit wizard

enpadasitest.science.ku.dk/gscf/studyEdit/propertie

Phenotype database

Home Create Import Browse/Edit Analyse Expert Modules Admin

New study

Define the basic properties of your study

Enter all the basic information of your study. Keep in mind that the more specific the information that is filled out, the more valuable the system will be.

Template

Title

Description

Code

Start Date

Study Type

Principle level

Primary End

Objectives

Central Con

Main Health Outcome

Exclusion Cr

Inclusion Criteria

Institute

Collaborators Involved

Country (For Multicentre Study Overall PI)

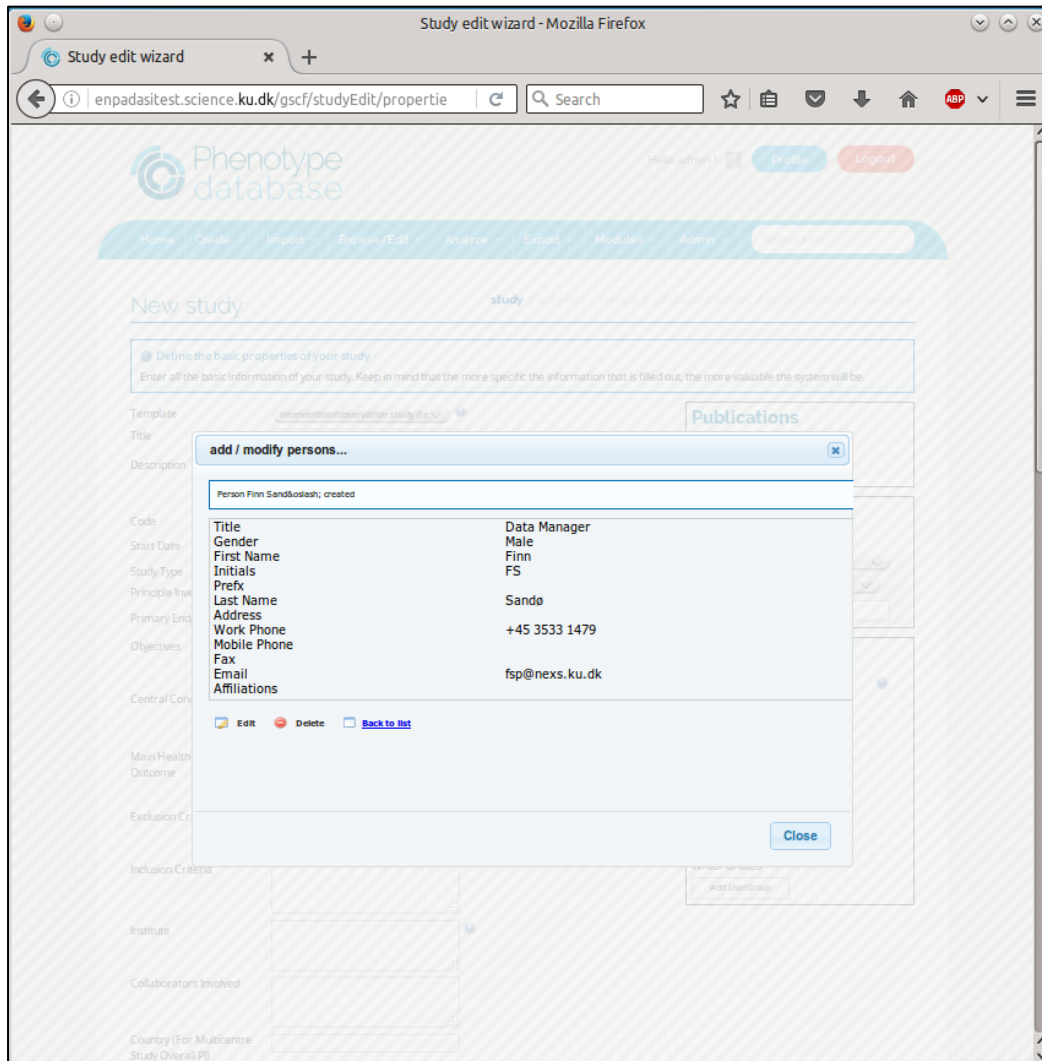
Publications

add / modify persons...

First Name	Prefix	Last Name	Affiliations
Mikkel		Tulin	

New Person Edit Affiliations

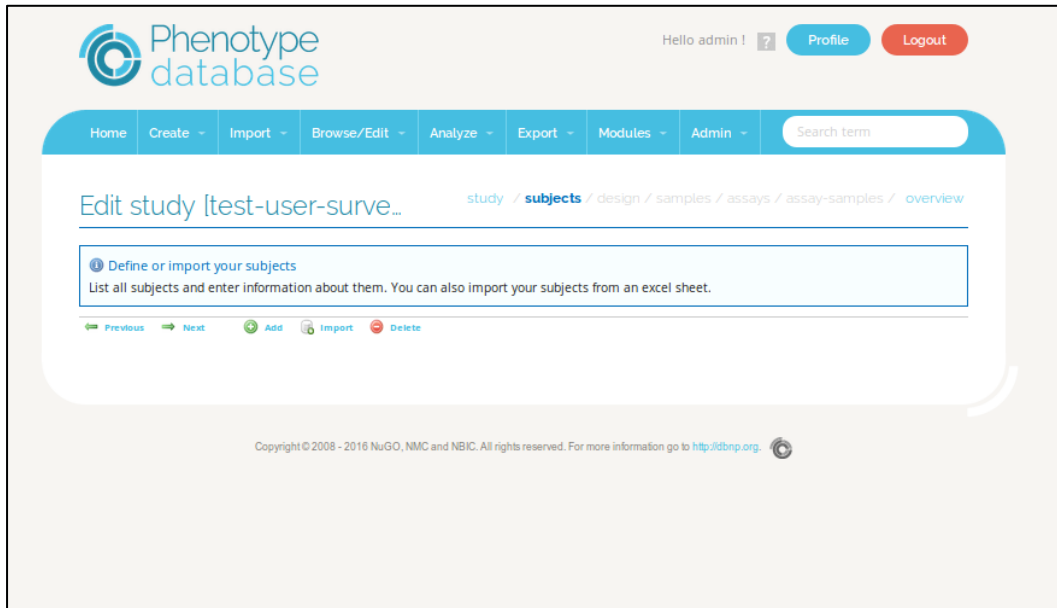
Close



These images show the process to add a person to the person list in the Phenotype Database (dbNP)



Phenotype Database (dbNP) - upload subjects



Phenotype database

Hello admin! ? Profile Logout

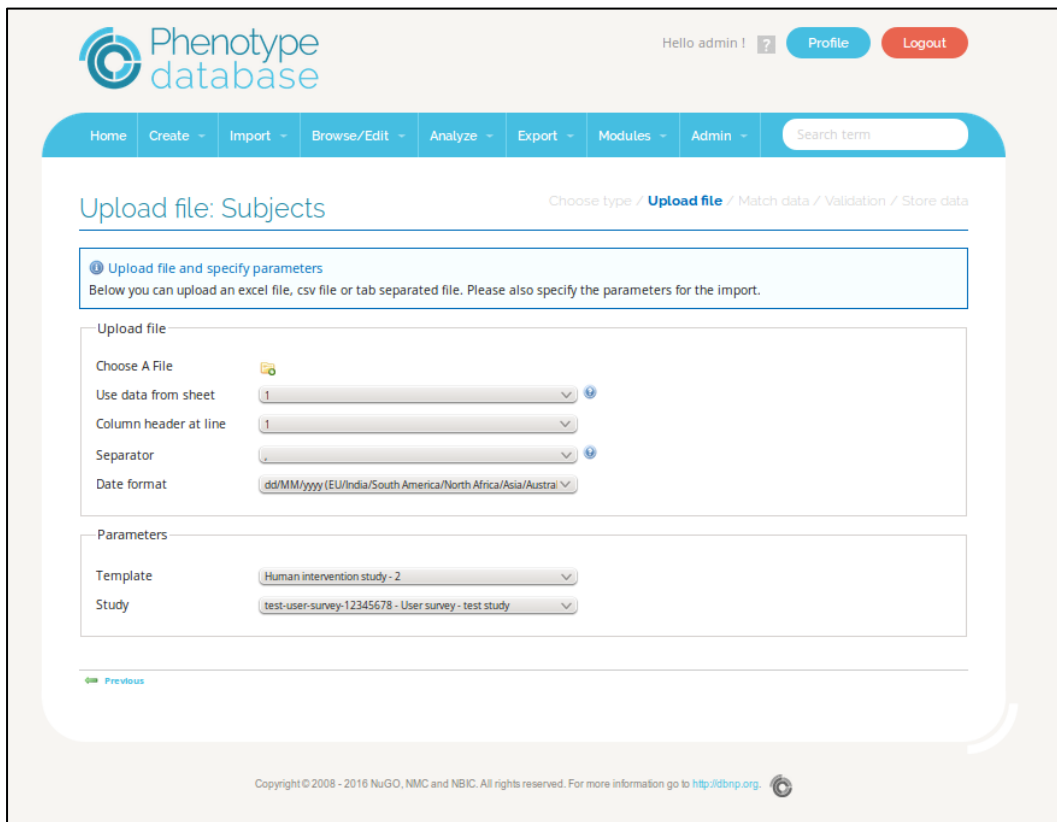
Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Edit study [test-user-surve... study / subjects / design / samples / assays / assay-samples / overview

Define or import your subjects
List all subjects and enter information about them. You can also import your subjects from an excel sheet.

Previous Next Add Import Delete

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.



Phenotype database


Hello admin! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Upload file: Subjects 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/Austral)

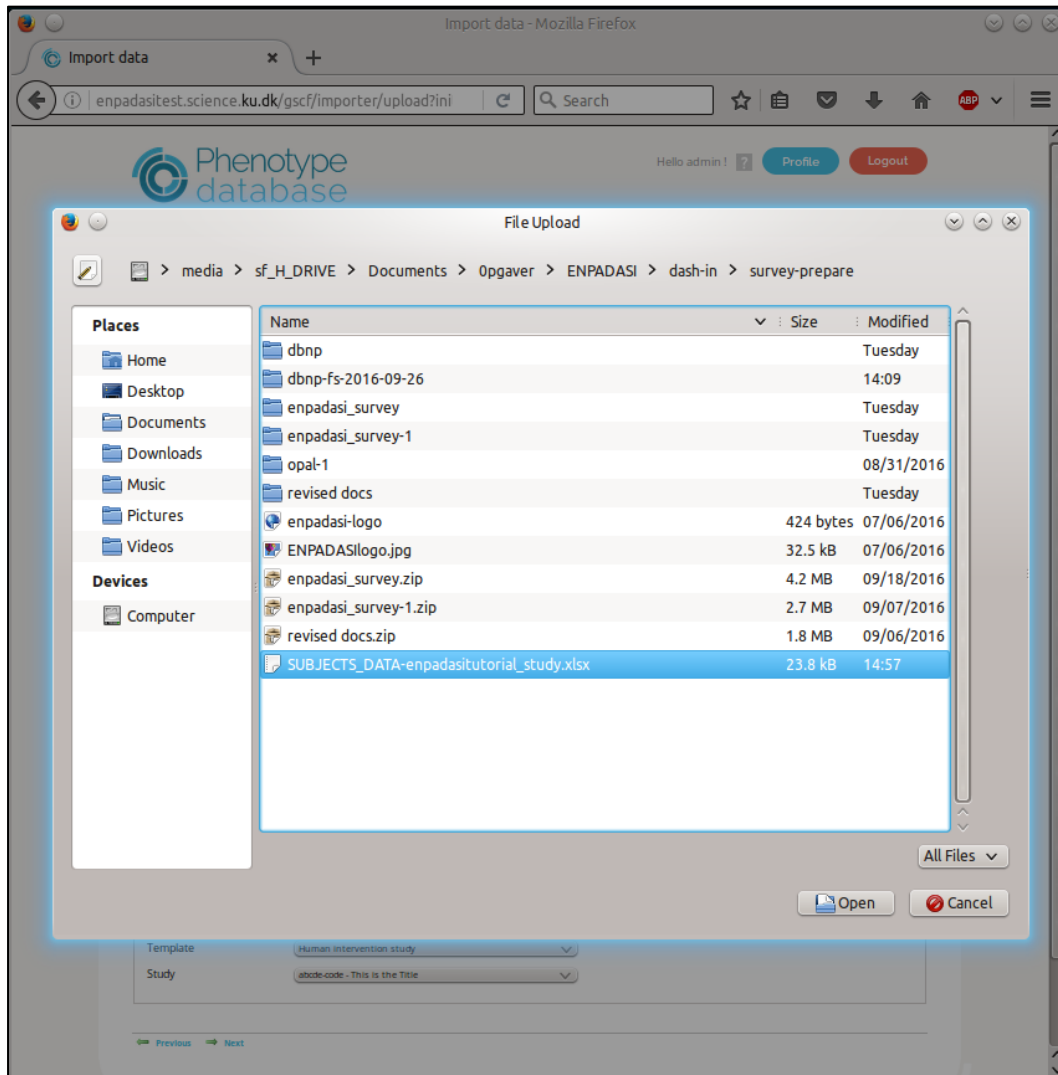
Parameters

Template Human intervention study - 2

Study test-user-survey-12345678 - User survey - test study

Previous

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.

The screenshot shows a web browser window titled 'Import data - Mozilla Firefox' with the URL 'enpadasitest.science.ku.dk/gscf/importer/upload?ini'. The page displays the 'Phenotype database' interface with a user profile for 'admin' and a 'Logout' button. A 'File Upload' dialog box is open, showing the file system path 'media > sf_H_DRIVE > Documents > Opgaver > ENPADASI > dash-in > survey-prepare'. The dialog contains a table of files:

Name	Size	Modified
dbnp		Tuesday
dbnp-fs-2016-09-26		14:09
enpadasi_survey		Tuesday
enpadasi_survey-1		Tuesday
opal-1		08/31/2016
revised docs		Tuesday
enpadasi-logo	424 bytes	07/06/2016
ENPADASIllogo.jpg	32.5 kB	07/06/2016
enpadasi_survey.zip	4.2 MB	09/18/2016
enpadasi_survey-1.zip	2.7 MB	09/07/2016
revised docs.zip	1.8 MB	09/06/2016
SUBJECTS_DATA-enpadasitutorial_study.xlsx	23.8 kB	14:57

Below the file list, there are 'All Files' and 'Open' buttons, and a 'Cancel' button. The background page shows a 'Template' dropdown set to 'Human intervention study' and a 'Study' dropdown set to 'abcde-code - This is the Title'.



Hello admin! [Profile](#) [Logout](#)

Home
Create
Import
Browse/Edit
Analyze
Export
Modules
Admin

Upload file: Subjects

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 Uploaded SUBJECTS_DATA-Sea bu...

Use data from sheet

Column header at line

Separator

Date format

Data preview

Name	Start group	Species	Country (humans)	Ethnicity (or strain)	Gender	Age	Initial bodyweight	Initial Body Height	BMI	Initial waist circumference	Commer
5493	1	Homo sapiens	Denmark	Caucasian	Male	26	118.2	1.855	34	110.5	Test c seque 1Sea
5502	1	Homo sapiens	Denmark	Caucasian	Male	44	89.9	1.76	29	111	Test c seque 1Sea
5506	1	Homo sapiens	Denmark	Caucasian	Male	23	111.3	1.83	33	102.5	Test c seque 1Sea
5491	2	Homo	Denmark	Caucasian	Male	20	106.1	1.79	33	104.5	Test c seque 1Sea

Showing rows 1 to 5 out of a total of 9 example rows

Parameters

Template

Study

← Previous
Next →



Phenotype database

Hello admin! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

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.

Match Clear mapping

Name	Start group	Species	Country (humans)
[Don't import]	[Don't import]	[Don't import]	[Don't import]
5493	1	Homo sapiens	Denmark
5502	1	Homo sapiens	Denmark
5506	1	Homo sapiens	Denmark
5491	2	Homo sapiens	Denmark
5492	2	Homo sapiens	Denmark

Showing rows 1 to 5 out of a total of 9 example rows

Previous Validate Import

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://fdbnp.org>.

Phenotype database

Hello admin! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

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.

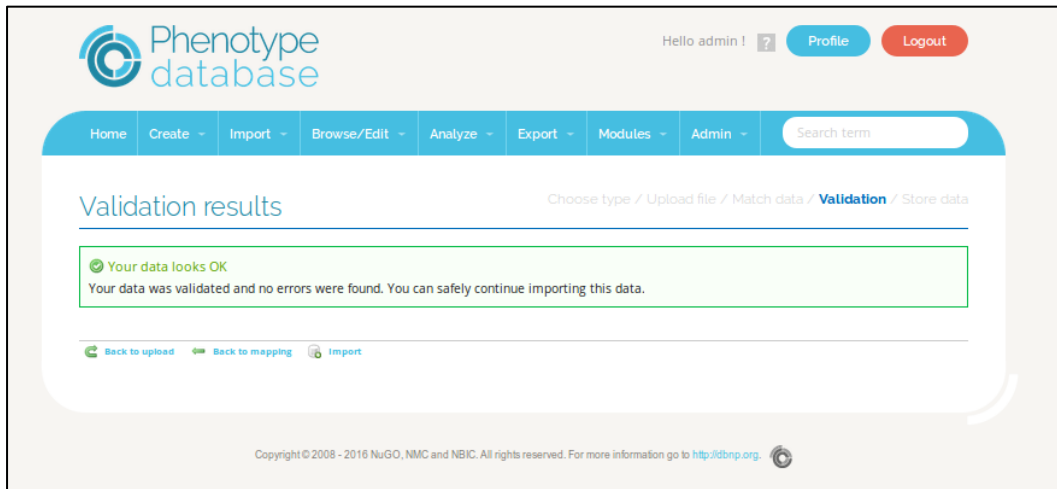
Match Clear mapping

Name	Start group	Species	Country (humans)
name	Start group	species	Country
5493	1	Homo sapiens	Denmark
5502	1	Homo sapiens	Denmark
5506	1	Homo sapiens	Denmark
5491	2	Homo sapiens	Denmark
5492	2	Homo sapiens	Denmark

Showing rows 1 to 5 out of a total of 9 example rows

Previous Validate Import

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://fdbnp.org>.



Phenotype database

Hello admin ! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

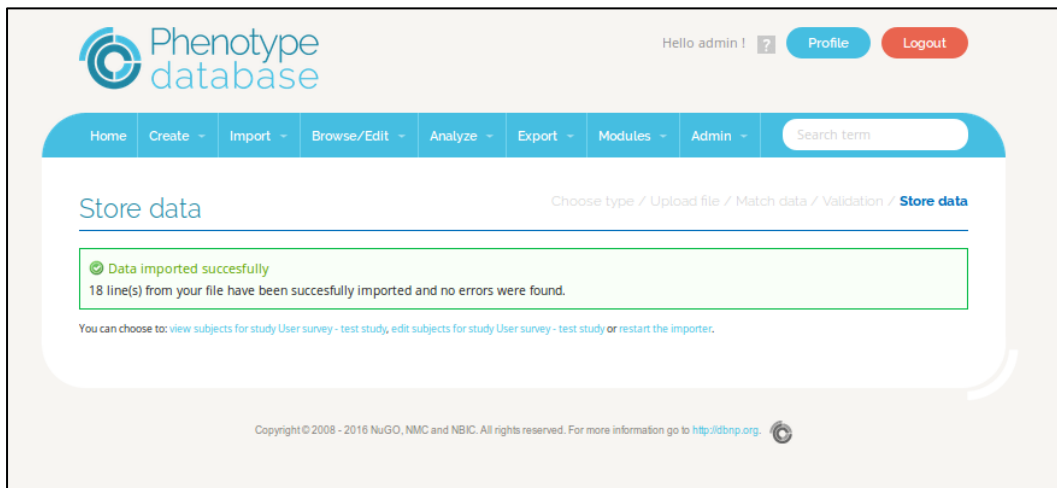
Validation results

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

✔ Your data looks OK
Your data was validated and no errors were found. You can safely continue importing this data.

Back to upload Back to mapping Import

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.



Phenotype database

Hello admin ! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Store data

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

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

You can choose to: [view subjects for study User survey - test study](#), [edit subjects for study User survey - test study](#) or [restart the importer](#).

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.



The screenshot shows the Phenotype Database interface. At the top, there is a navigation bar with 'Home', 'Create', 'Import', 'Browse/Edit', 'Analyze', 'Export', 'Modules', and 'Admin'. A search bar is also present. The main content area displays 'Study [test-user-survey-1...' with a breadcrumb trail: 'study / subjects / design / samples / assays'. A message box states: 'This view shows your subjects. For every template, a list of subjects is shown.' Below this, the 'Template: Human intervention study' is selected. A table lists 18 subjects, with the first 10 shown. The table has columns for name, species, start group, intervention protocol, gender, age (years), ethnicity, and waist circumference.

name	species	Start group	Intervention protocol	Gender	Age (years)	Ethnicity	Waist circumference
5491	Homo sapiens	2	Test diet sequence: 1Sea buckthorn; 3Control; 2Strawberry	Male		Caucasian	104.5
5492	Homo sapiens	2	Test diet sequence: 1Sea buckthorn; 3Control; 2Strawberry	Male		Caucasian	89.5
5493	Homo sapiens	1	Test diet sequence: 1Sea buckthorn; 3Control; 2Strawberry	Male		Caucasian	110.5
5494	Homo sapiens	2	Test diet sequence: 1Sea buckthorn; 3Control; 2Strawberry	Male		Caucasian	100.0
5495	Homo sapiens	6	Test diet sequence: 3Control; 2Strawberry; 1Sea buckthorn	Male		Caucasian	95.0
5496	Homo sapiens	4	Test diet sequence: 2Strawberry; 3Control; 1Sea buckthorn	Male		Caucasian	101.0
5497	Homo sapiens	2	Test diet sequence: 1Sea buckthorn; 3Control; 2Strawberry	Male		Caucasian	109.0
5500	Homo sapiens	2	Test diet sequence: 1Sea buckthorn; 3Control; 2Strawberry	Male		Caucasian	103.5
5501	Homo sapiens	6	Test diet sequence: 3Control; 2Strawberry; 1Sea buckthorn	Male		Caucasian	102.5
5502	Homo sapiens	1	Test diet sequence: 1Sea buckthorn; 2Strawberry; 3Control	Male		Caucasian	111.0

Showing 1 to 10 of 18 entries. Navigation: First, Previous, 1, 2, Next, Last.

These images show the process to upload subjects to the Phenotype Database (dbNP)



Phenotype Database (dbNP) - study design, add subject groups

Phenotype database Hello admin ! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Study Itest-user-survey-1... study / subjects / **design** / samples / assays

This page shows your study design
 The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Double click on a subjectgroup or sample & treatment group to see details.

12h	0	-12h	0	-12h	0	12h	0	12h	0	12h	0	12h	0
-3d	-2d		-1d		0		1d		2d		3d		4d

[edit](#) [back to list](#)

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.

Phenotype database Hello admin ! ? Profile Logout

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Edit study Itest-user-surve... study / subjects / **design** / samples / assays / assay-samples / overview

Define or import your study design
 The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.
N.B. You can edit your subject groups by double clicking on its name.

The first step is to add one or more subject groups.
 After that, you can create sample & treatment groups and assign them to the subject groups.

[Add subjectgroup](#)

Available sample & treatment groups

[Add new](#)

[Previous](#) [Next](#)

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.



Add subjectgroup ✕

ⓘ Edit the details of the subjectgroup
Check the subjects that are in this group

Name:

Show 100 entries Search:

<input type="checkbox"/>	Name	Template	Species
<input type="checkbox"/>	5491	Human intervention study	Homo sapiens
<input type="checkbox"/>	5492	Human intervention study	Homo sapiens
<input checked="" type="checkbox"/>	5493	Human intervention study	Homo sapiens
<input type="checkbox"/>	5494	Human intervention study	Homo sapiens
<input type="checkbox"/>	5495	Human intervention study	Homo sapiens
<input type="checkbox"/>	5496	Human intervention study	Homo sapiens
<input type="checkbox"/>	5497	Human intervention study	Homo sapiens
<input type="checkbox"/>	5500	Human intervention study	Homo sapiens
<input type="checkbox"/>	5501	Human intervention study	Homo sapiens
<input checked="" type="checkbox"/>	5502	Human intervention study	Homo sapiens
<input type="checkbox"/>	5503	Human intervention study	Homo sapiens
<input type="checkbox"/>	5504	Human intervention study	Homo sapiens
<input type="checkbox"/>	5505	Human intervention study	Homo sapiens
<input type="checkbox"/>	5506	Human intervention study	Homo sapiens
<input checked="" type="checkbox"/>	5508	Human intervention study	Homo sapiens
<input type="checkbox"/>	5509	Human intervention study	Homo sapiens
<input type="checkbox"/>	5510	Human intervention study	Homo sapiens
<input type="checkbox"/>	5511	Human intervention study	Homo sapiens

Showing 1 to 18 of 18 entries (3 selected) [clear selection](#) First Previous 1 Next Last



Home Create Import Browse/Edit Analyze Export Modules Admin

Edit study [test-user-surve...

study / subjects / **design** / samples / assays / assay-samples / overview

Define or import your study design

The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.

N.B. You can edit your subject groups by double clicking on its name.

Group6															
Group5															
Group4															
Group2															
Group1															
	0	12h	0	12h	0	12h	0	12h	0	12h	0	12h	0	12h	
	0		1d		2d		3d		4d		5d		6d		

+ Add subjectgroup

Available sample & treatment groups

+ Add new

← Previous
Next →

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://idbnp.org>.

Add eventgroup ✕

Edit the details of the sample & treatment group

Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.

Please note: changes to this sample & treatment group will affect all instances of the group.

Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

Save name
Cancel

Available sample & treatment groups

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Edit study [test-user-surve... study / subjects / **design** / samples / assays / assay-samples / overview

1 Define or import your study design
 The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.
N.B.You can edit your subject groups by double clicking on its name.

Group6													
Group5													
Group4													
Group2													
Group1													
	0	12h	0	12h	0	12h	0	12h	0	12h	0	12h	0
	0		1d		2d		3d		4d		5d		6d

[Add subjectgroup](#)

Available sample & treatment groups

Control	edit del
Sea Buckthorn	edit del
Strawberry	edit del

[Add new](#)

[Previous](#) [Next](#)

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>

These images show the process to study design, add subject groups to a study design in the Phenotype Database (dbNP)



Phenotype Database (dbNP) - study design: add treatment types and sample types

Edit sample & treatment group

① Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
Please note: changes to this sample & treatment group will affect all instances of the group.
Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

0	12h	0	12h	0	12h	0	12h	0	12h	0	12h	0	12h
0		1d		2d		3d		4d		5d		6d	

Available treatments types **Available sample types**

Add treatment type

New treatment type

Template

- Blood sampling
- Body fluid sampling
- Bone scan prep
- Compound challenge
- Compound intervention
- Diet challenge**
- Diet intervention
- Diet/Physical challenge
- Humanization
- Observation
- Physical activity challenge
- SHAM-UQ procedure
- Screening visit
- Surgery
- Urine Sample
- add / modify...

Add treatment type

New treatment type

Template:

Template for fasting or feeding intervention or challenge.

Name:

Route:

Event Name (STRING):

Intervention/Challenge:

Event-Type:

Description:

Diet Description:

Diet Carbohydrate-Level:

Diet Fat-Level:

Diet Protein-Level:

Migration:



Edit sample & treatment group

Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
Please note: changes to this sample & treatment group will affect all instances of the group.
Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

0	12h	0	12h	0	12h	0	12h	0	12h	0	12h	0	12h
0		1d		2d		3d		4d		5d		6d	

Available treatments types

Ad Libitum edit del

+ Add new

Available sample types

+ Add new

Save name
Cancel

Add sample type

New sample type

Template

- Body composition
- Body fluid
- Dietary intake
- Excretion
- Genetic variation (SNPs)
- Imaging
- Organ/tissue
- Physiology
- Questionnaire
- Tissue
- add / modify...

Save name
Cancel

Add sample type

New sample type

Template

Name + Anthropometry

Sample Template + Anthropometrics

Save name
Cancel



Edit sample type

Edit sample type [Blood plasma]

Template: Body fluid
 Template for Body fluids (e.g: blood, lymph).

Name: Blood plasma

Sample Template: Blood sample

Sampling Name Short: []

Sampling-Type: []

Body Fluid: []

Related Event/Chall.: []

Relative Time In Related Event: 0s

Sampling Method: []

Sample Fraction: []

Sampling Description: []

Fasting Period: 0s

External Treatment: []

Migration: []

Buttons: Save name, Cancel

Edit sample & treatment group

ⓘ Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
 Please note: changes to this sample & treatment group will affect all instances of the group.
 Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name: Control

-12h	0	-12h	Control	12h	0	12h	0	12h	0	12h	0	12h	0	12h	0
-2d	-1d	0		1d	2d	3d	4d	5d							

Available treatments types

- Ad Libitum [edit del]
- Control [edit del]
- Sea Buckthorn [edit del]
- Strawberry [edit del]
- Add new

Available sample types

- Anthropometry [edit del]
- Blood plasma [edit del]
- Urine [edit del]
- Vas [edit del]
- Add new

Buttons: Save name, Cancel

These images show the process to add treatment types and sample types to a study design in the Phenotype Database (dbNP)



Phenotype Database (dbNP) - create study design for a single subject group

The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be edited.

Edit sample & treatment group

Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
Please note: changes to this sample & treatment group will affect all instances of the group.
Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

-12h	0	-12h	Control	12h	0	12h	0	12h	0	12h	0	12h	0
-2d	-1d	0	1d	2d	3d	4d	5d						

Available treatments types

- Ad Libitum edit del
- Control edit del
- Sea Buckthorn edit del
- Strawberry edit del
- Add new

Available sample types

- Anthropometry edit del
- Blood plasma edit del
- Urine edit del
- Vas edit del
- Add new

Please specify the duration of this treatment type.
 Use 's', 'm', 'h', 'd', 'w' or 'y' to denote seconds, minutes, hours, days, weeks or years.
 Specify 0 for no duration.

Examples: '1d', '2d 12h', '3h 45m'



The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be...

Edit sample & treatment group

Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
Please note: changes to this sample & treatment group will affect all instances of the group.
Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

	Control														
m	-15m	0	15m	30m	45m	0	15m	30m	45m	0	15m	30m	45m	0	15m
	-1h	0				1h				2h				3h	

Available treatments types

- Ad Libitum edit del
- Control edit del
- Sea Buckthorn edit del
- Strawberry edit del
- Add new

Available sample types

- Anthropometry edit del
- Blood plasma edit del
- Urine edit del
- Vas edit del
- Add new

Save name Cancel

The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be...

Edit sample & treatment group

Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
Please note: changes to this sample & treatment group will affect all instances of the group.
Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

Available treatments types

- Ad Libitum edit del
- Control edit del
- Sea Buckthorn edit del
- Strawberry edit del
- Add new

Available sample types

- Anthropometry edit del
- Blood plasma edit del
- Urine edit del
- Vas edit del
- Add new

Save name Cancel



The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be...

Edit sample & treatment group

Edit the details of the sample & treatment group
 Drag treatment types and sample types into the group. Changes will be saved immediately. However, changes in the name require a click on the 'save name' button.
Please note: changes to this sample & treatment group will affect all instances of the group.
Please note: the timing of treatment and sample types also depends on the timing of the group within the study.

Name:

Available treatments types

Ad Libitum	edit del
Control	edit del
Sea Buckthorn	edit del
Strawberry	edit del
Add new	

Available sample types

Anthropometry	edit del
Blood plasma	edit del
Urine	edit del
Vas	edit del
Add new	

[Save name](#) [Cancel](#)

These images show the process to create a study design for a single subject group in the Phenotype Database (dbNP)



Phenotype Database (dbNP) - create over-all study design

Home
Create
Import
Browse/Edit
Analyze
Export
Modules
Admin

Search term

Edit study [ttest-user-surve...](#) study / subjects / **design** / samples / assays / assay-samples / overview

1 Define or import your study design

The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.

N.B. You can edit your subject groups by double clicking on its name.

Group6														
Group5														
Group4														
Group2														
Group1														
	-12h	0	-12h	0	12h	0	12h	0	12h	0	12h	0	12h	0
-2d		-1d		0		1d		2d		3d		4d		

3 Add subjectgroup

Available sample & treatment groups

Control 8x Vas, 6x Blood plasma, 4x Urine...
edit del

Sea Buckthorn 8x Vas, 6x Blood plasma, 4x Urine...
edit del

Strawberry 7x Vas, 6x Blood plasma, 4x Urine...
edit del

4 Add new

[Previous](#) [Next](#)

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>



Search term
Home Create Import Browse/Edit Analyze Export Modules Admin

Edit study *ltest-user-surve...* study / subjects / **design** / samples / assays / assay-samples / overview

1 Define or import your study design

The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.

N.B. You can edit your subject groups by double clicking on its name.

Group6									
	Control								
Group5									
Group4									
Group2									
Group1									
	0	4h	8h	12h	16h	20h	0		
	-1d	0					0		1d

2 Add subjectgroup

Available sample & treatment groups

Control	8x Vas, 6x Blood plasma, 4x Urine...	edit del
Sea Buckthorn	8x Vas, 6x Blood plasma, 4x Urine...	edit del
Strawberry	7x Vas, 6x Blood plasma, 4x Urine...	edit del

3 Add new

[Previous](#) [Next](#)

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>.

Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Edit study [test-user-surve... study / subjects / **design** / samples / assays / assay-samples / overview

1 Define or import your study design
The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.
N.B. You can edit your subject groups by double clicking on its name.

Group6	Contrc							Strawb										Sea Bu
Group5	Contrc							Sea Bu										Strawb
Group4	Strawb							Contro										Sea Bu
Group2	Sea Bu							Contro										Sea Bu
Group1	Sea Bu							Strawb										Contro
	0	1d	2d	3d	4d	5d	6d	0	1d	2d	3d	4d	5d	6d	0	1d	2d	
	0							1w							2w			

Add subjectgroup

Available sample & treatment groups

- Control 8x Vas, 6x Blood plasma, 4x Urine... edit del
- Sea Buckthorn 8x Vas, 6x Blood plasma, 4x Urine... edit del
- Strawberry 7x Vas, 6x Blood plasma, 4x Urine... edit del

Add new

Previous Next

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>

These images show the process to create the over-all study design in the Phenotype Database (dbNP)

Phenotype Database (dbNP) - generate samples

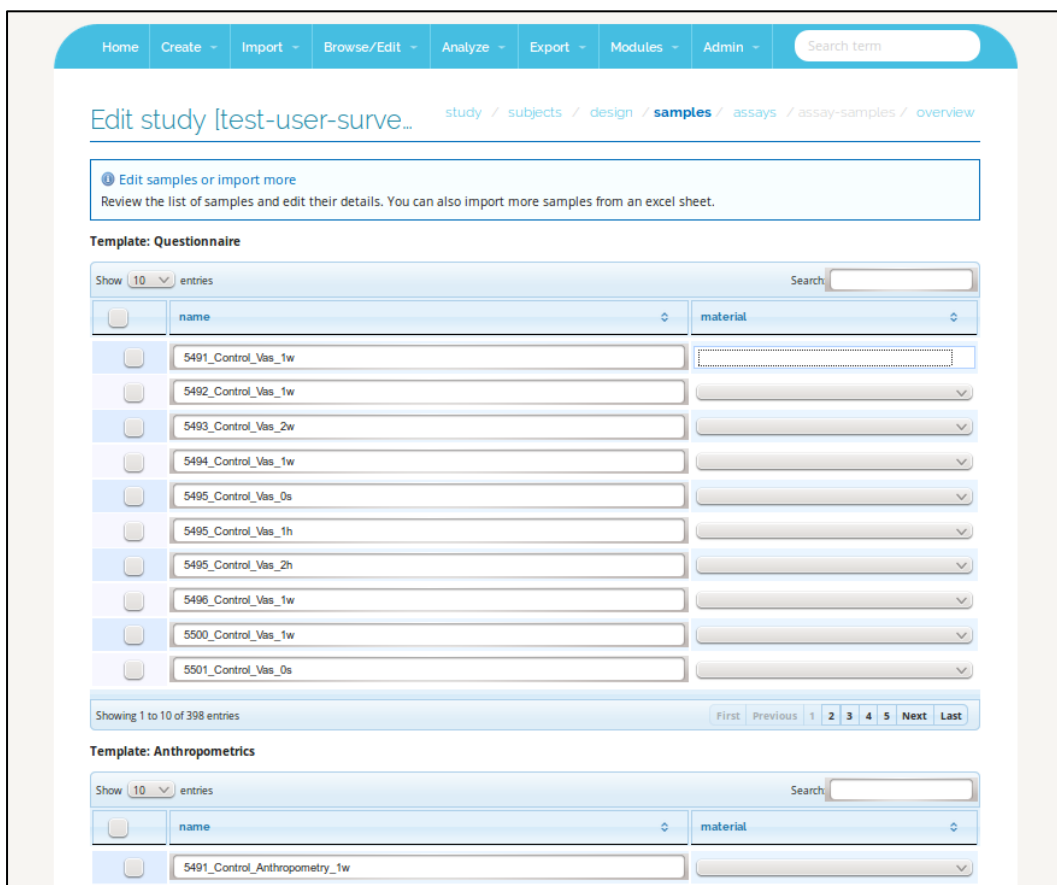
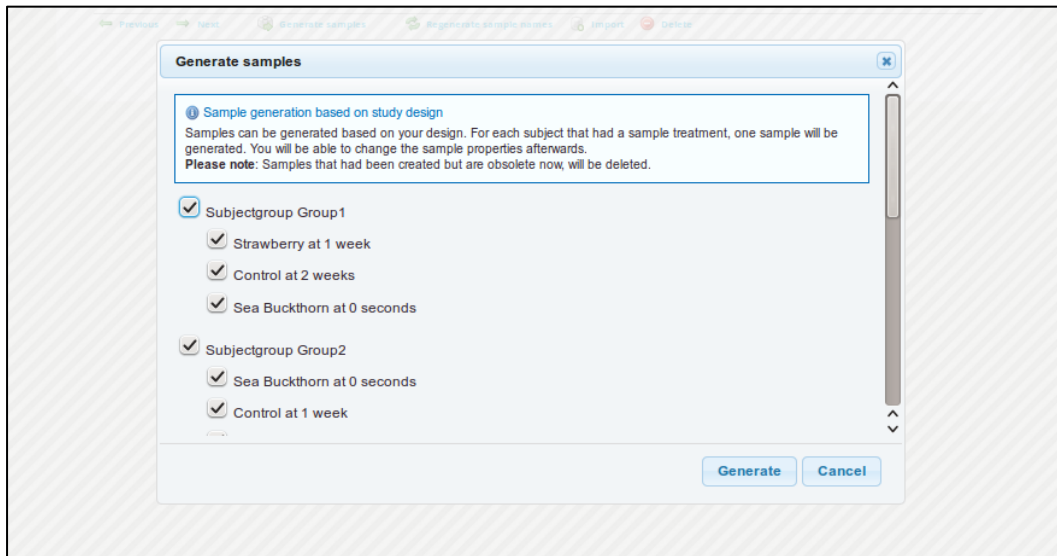
Home Create Import Browse/Edit Analyze Export Modules Admin Search term

Edit study [test-user-surve... study / subjects / design / **samples** / assays / assay-samples / overview

1 Generate samples
Click 'Generate samples' to initiate the sample generation based on your study design.

Previous Next Generate samples Regenerate sample names Import Delete

Copyright © 2008 - 2016 NuGO, NMC and NBIC. All rights reserved. For more information go to <http://dbnp.org>

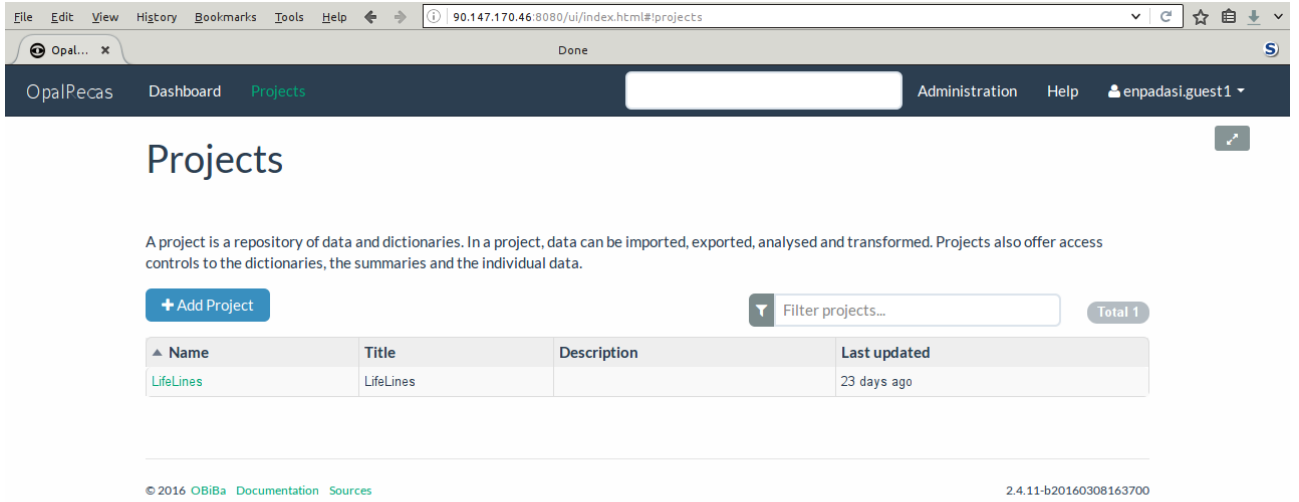


These images show the process to generate samples in the Phenotype Database (dbNP)



ANNEX 2. OPAL GUIDE

Add a new project to OPAL



OpalPecas Dashboard **Projects** Administration Help enpadasi.guest1

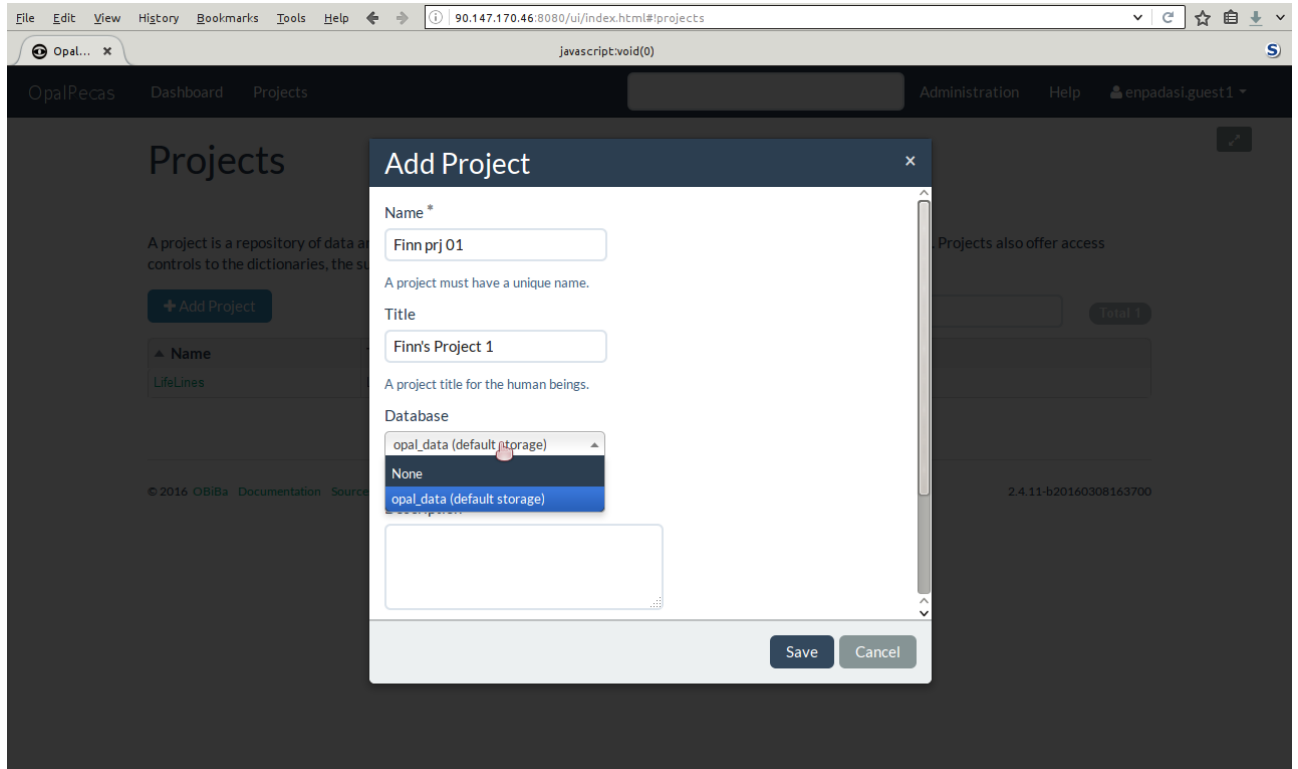
Projects

A project is a repository of data and dictionaries. In a project, data can be imported, exported, analysed and transformed. Projects also offer access controls to the dictionaries, the summaries and the individual data.

[+ Add Project](#) Total 1

Name	Title	Description	Last updated
LifeLines	LifeLines		23 days ago

© 2016 OBiBa [Documentation](#) [Sources](#) 2.4.11-b20160308163700



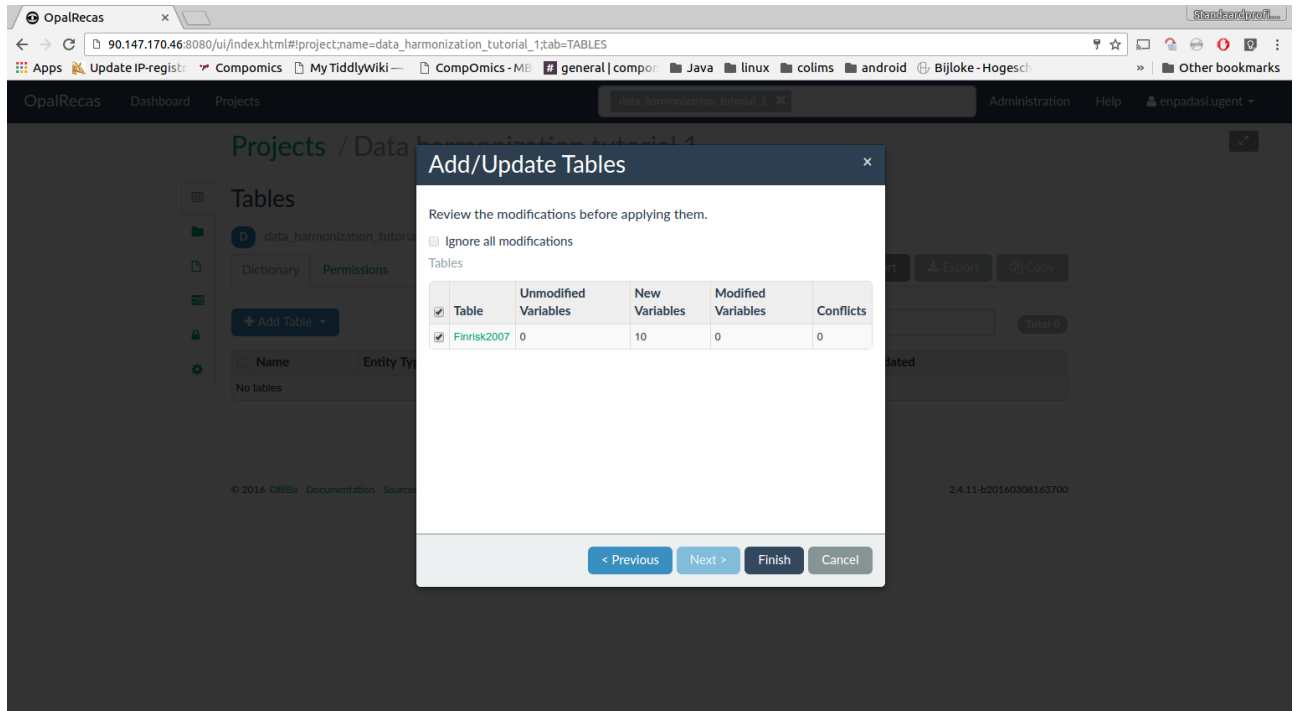
These images show part of the process to add a new project to OPAL

OPAL – Add table dictionary

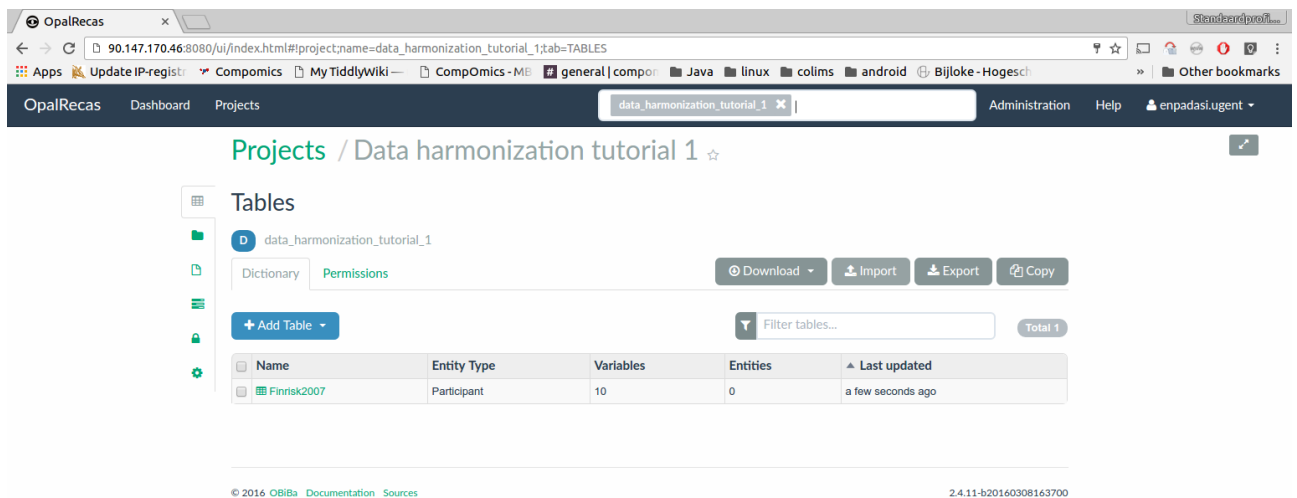
javascript;

Before uploading table data in .csv format, a data dictionary needs to be added. In case of data in .sav format, both the data dictionary and the data are included in the same file.

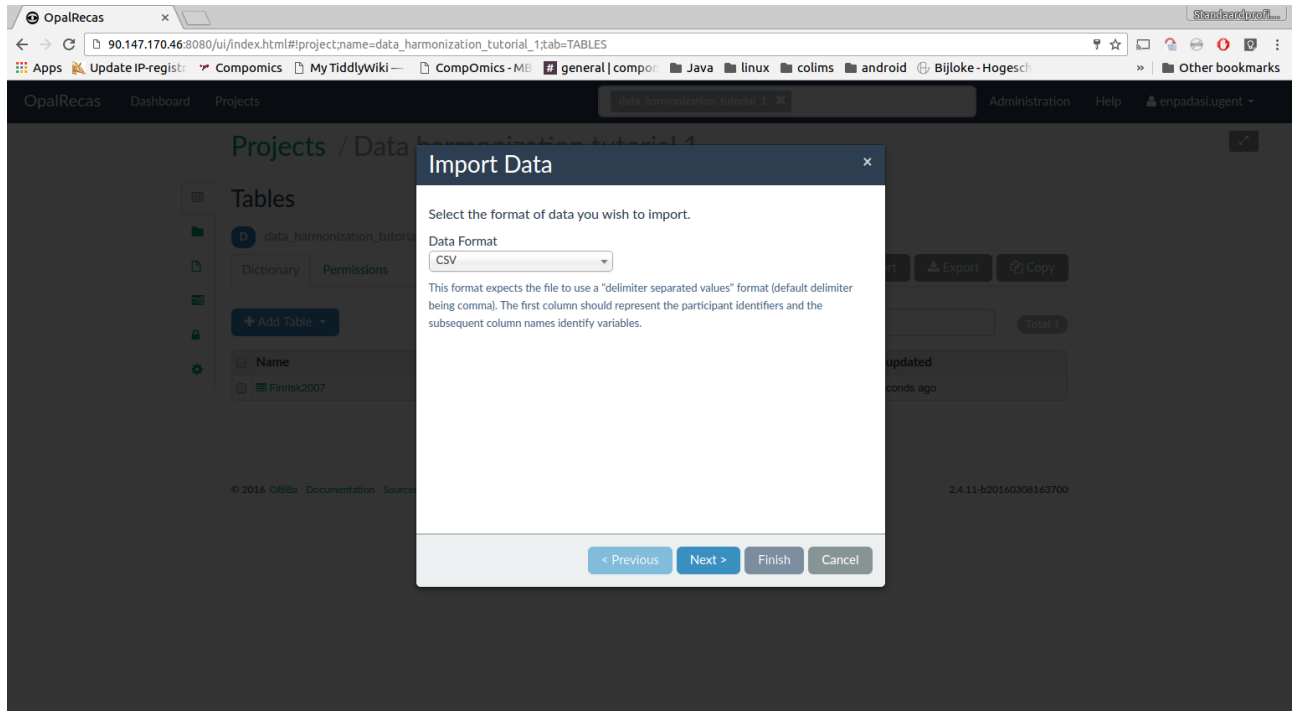
After uploading the data dictionary to the Opal server, the user can select it (.xls, .xlsx and .sav are supported).



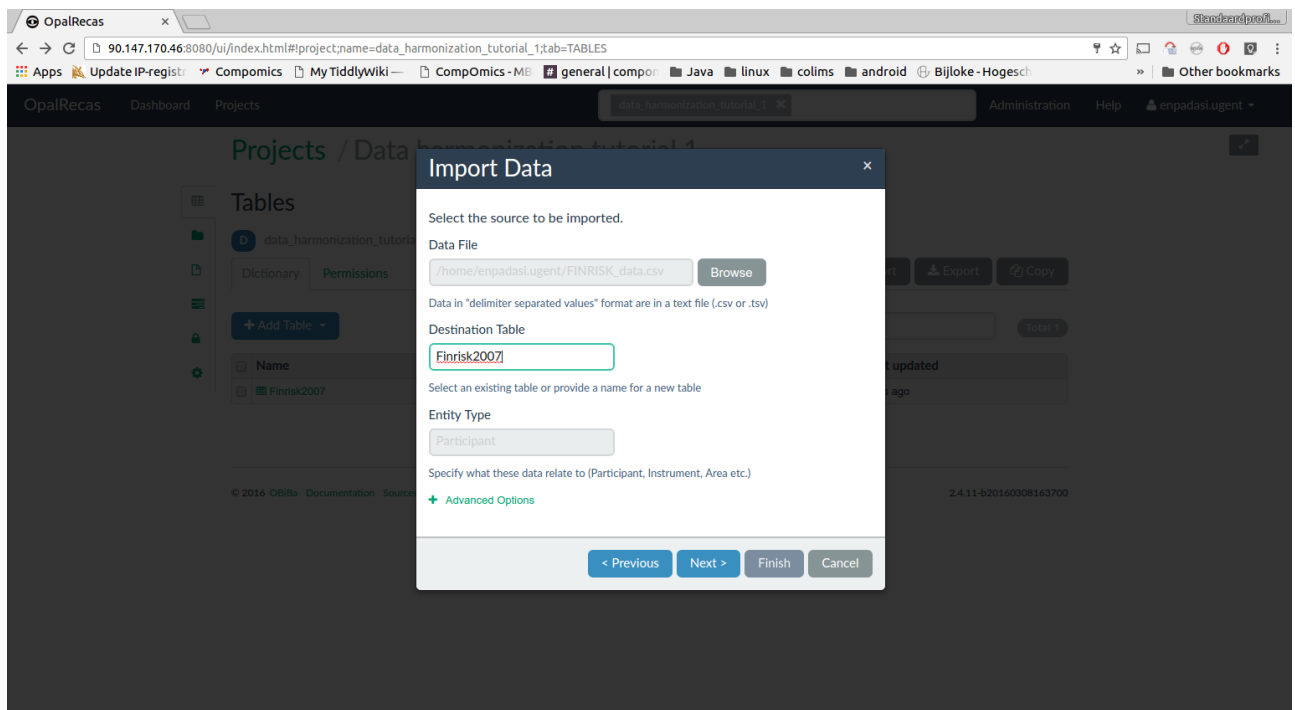
Select the table of interest and click on the "Finish" button.



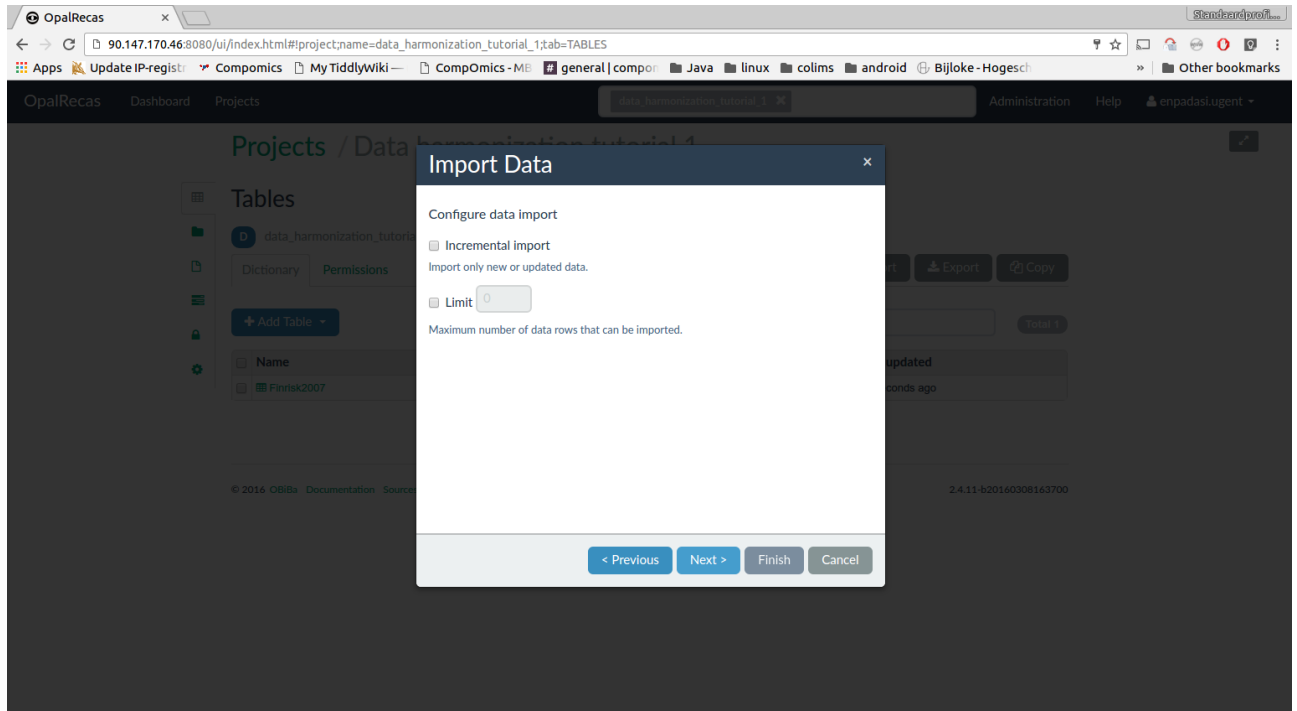
Once the data dictionary has been added, click on the "Import" button to import the table data.



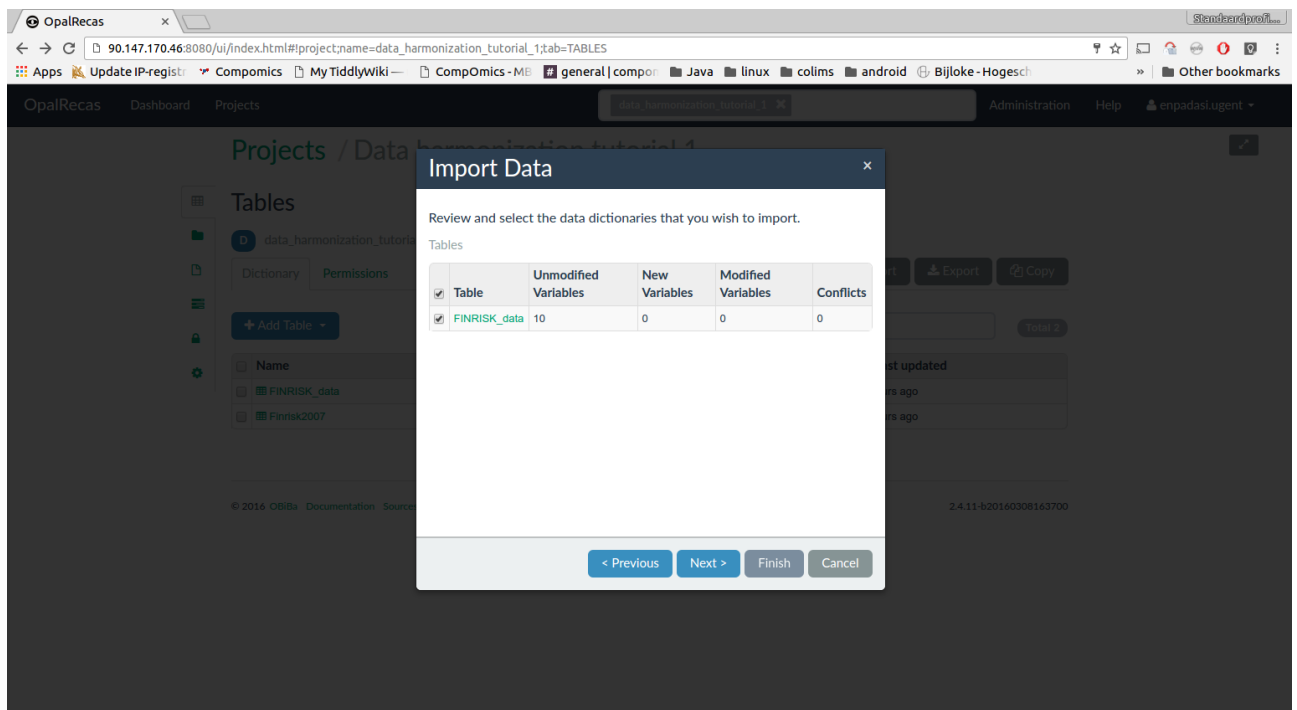
Leave the default "CSV" setting and click on the "Next" button.



Select the data file and the destination table and click on the "Next" button.



Keep the default settings and click on the "Next" button.



Select the appropriate table and click on the "Next" button.



JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

Review the data that will be imported.

Table: FINRISK_data

+ Display Options

ID	SUKUP	PAINO	PITUUS	FR
1000948357610215153	2	89.38351149082827	168.2201360071805	1
1002400312040470929	2	94.28620400965093	172.98381515846853	1
1004168722520747805	1	58.05548910267279	168.59543649669538	1
1007800521831721882	1	81.44858356305257	174.25781292787622	1
1010975981543513267	2	79.9910303261611	181.97687099030263	1
1013334134400726098	1	47.11228817996795	145.27003920443656	1
1013735151969205356	2	79.80130160335341	169.52511808291368	1
1015429109149314392	1	52.32337587441517	166.3947009852287	1

< Previous Next > Finish Cancel

Click on the "Next" button.

Specify whether the data file is to be archived.

Once a file has been successfully imported:

- leave it where it is
- move it to another folder:

Browse

< Previous Next > Finish Cancel

Click on the "Finish" button.

Opal - Add harmonized schema



OpalRecas Dashboard Projects data_harmonization_tutorial_1 Administration Help enpadasi.ugent

Projects / Data harmonization tutorial 1

Tables

data_harmonization_tutorial_1

Dictionary Permissions Download Import Export Copy

+ Add Table Filter tables... Total 1

- Add table...
- Add/update tables from dictionary...
- Add view...

Name	Type	Variables	Entities	Last updated
Finrisk2007	Table	10	0	7 hours ago

© 2016 OBIBA Documentation Sources 2.4.11-b20160308163700

Click on the "Add table" button and select "Add view..."

OpalRecas Dashboard Projects data_harmonization_tutorial_1 Administration Help enpadasi.ugent

Projects / Data harmonization tutorial 1

Tables

data_harmonization_tutorial_1

Dictionary Permissions Download Import Export Copy

+ Add Table Filter tables... Total 1

- Add table...
- Add/update tables from dictionary...
- Add view...

Name	Type	Variables	Entities	Last updated
Finrisk2007	Table	10	0	7 hours ago

Add View

Name *
harmonized_schema

A view must have a unique name.

Table References *
Finrisk2007

The table(s) from which the entities are extracted.

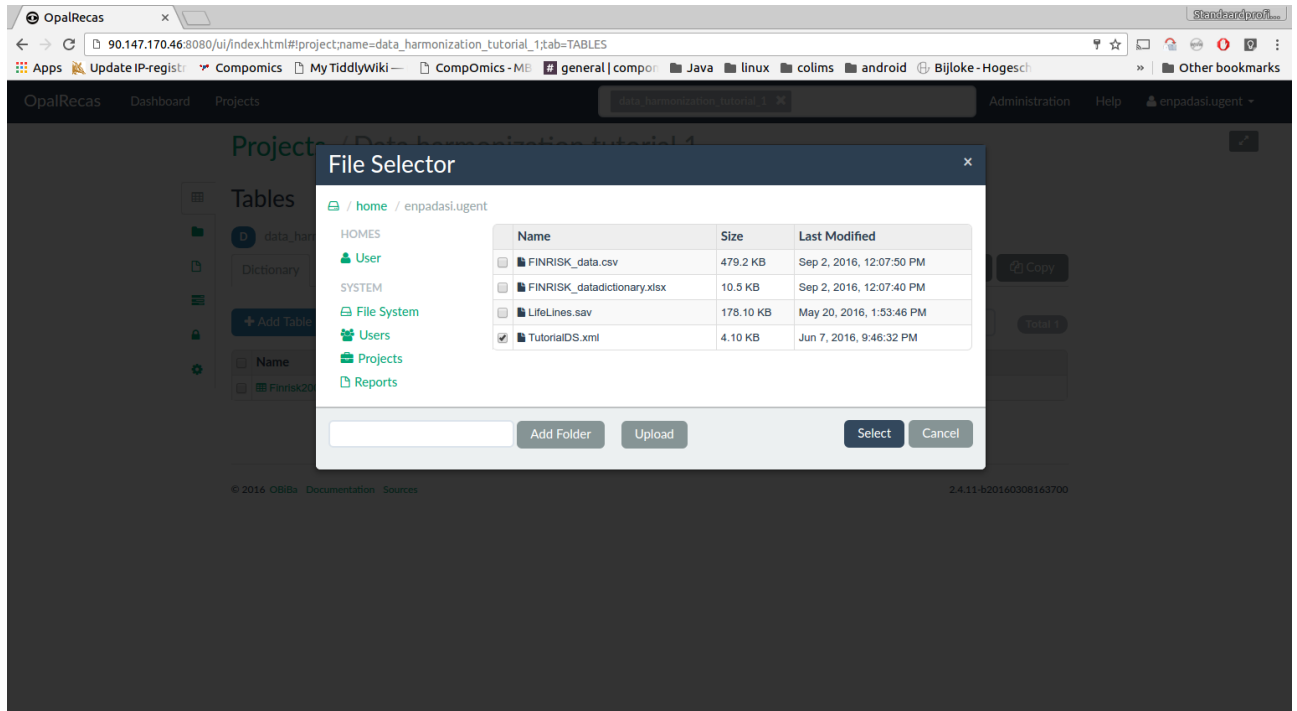
Variables
/home/enpadasi.ugent/TutorialDS.xml Browse

Provide derived variables from a view XML or a table dictionary in Excel, optional.

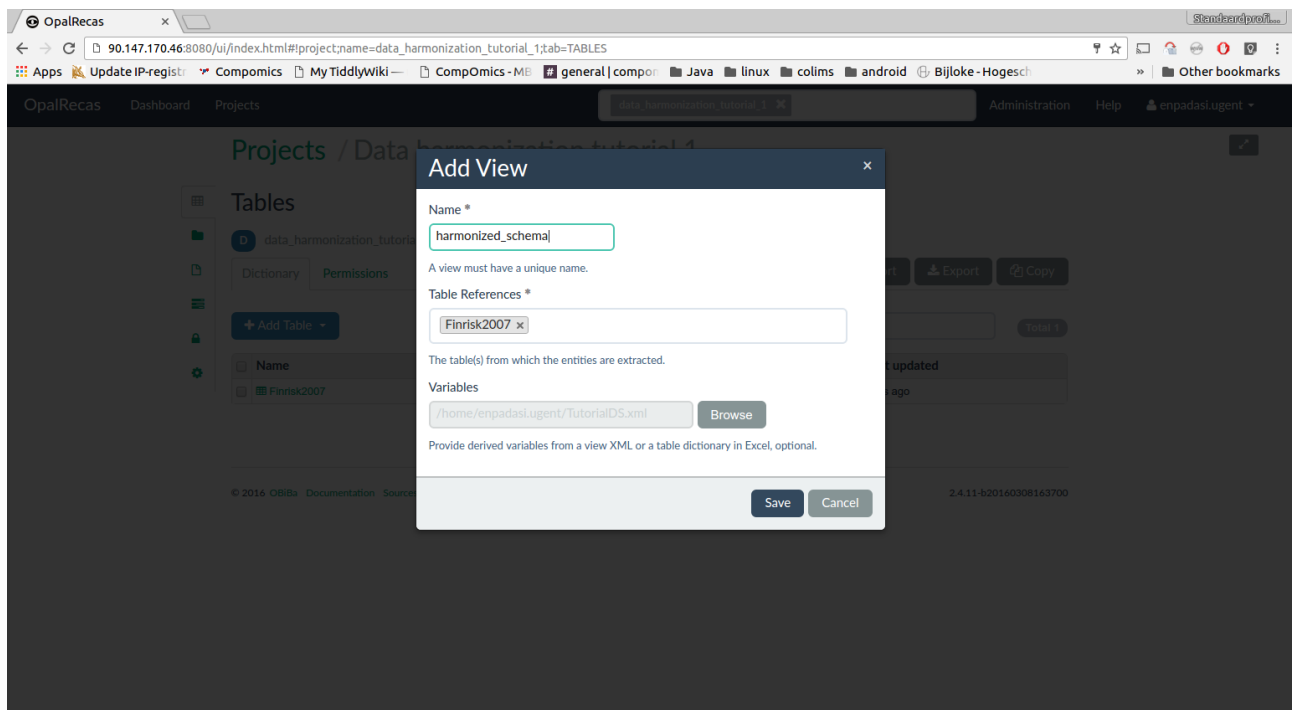
Save Cancel

© 2016 OBIBA Documentation Sources 2.4.11-b20160308163700

Select the previously added data table in the "Table References" field. Click on the "Browse" button next to the "Variables" field for selecting the harmonized variables.



Upload the harmonized variables in view XML format or table dictionary Excel format and the select it.



Click on the "Save" button to add the harmonized view.

Opal - Variable harmonization

The screenshot shows the OpalRecas web interface. The browser address bar indicates the URL: `90.147.170.46:8080/ui/index.html#project=name=data_harmonization_tutorial_1;tab=TABLES`. The page title is "Projects / Data harmonization tutorial 1".

The main content area is titled "Tables" and shows a list of tables for the project "data_harmonization_tutorial_1". The table list is as follows:

Name	Entity Type	Variables	Entities	Last updated
harmonized_schema	Participant	5	0	a few seconds ago
Finrisk2007	Participant	10	0	7 hours ago

At the bottom of the page, there is a copyright notice: © 2016 OBiBa Documentation Sources and a version number: 2.4.11-b20160308163700.

Click on the "Save" button to add the harmonized view.

The screenshot shows the OpalRecas web interface with the "Properties" view selected for the "harmonized_schema" table. The browser address bar indicates the URL: `90.147.170.46:8080/ui/index.html#project=name=data_harmonization_tutorial_1;tab=TABLES;path=data_harmonization_tutorial_1.harmonized_schema`.

The "Properties" section shows the following details:

Name	harmonized_schema
Entity Type	Participant
Table References	data_harmonization_tutorial_1.Finrisk2007

The "Entity Filter" section shows: `// no filter`.

The "Variables" section shows a list of variables for the table:

Name	Label	Value Type	Categories
Gender		integer	0, 1, 9
PM_BMI_CAT		integer	1, 2, 3, 9
DIS_DIAB		integer	0, 1, 9
PM_DIASTOLIC		integer	999
SMK_CIG_CURRENT		integer	0, 1, 9

At the bottom of the page, there is a copyright notice: © 2016 OBiBa Documentation Sources and a version number: 2.4.11-b20160308163700.

Click on the "Gender" variable. It will be used as an example of an "easy" variable to harmonize.



JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

The screenshot shows the OpalRecas web interface. The main content area displays the 'Properties' section for a variable named 'Gender'. A dropdown menu is open over the 'Derive' button, showing the following options:

- Add variable to view...
- Categorize this variable to another...
- Categorize another variable to this...** (highlighted in green)
- Custom derivation...

The 'Properties' section includes the following table:

Name	Gender
Entity Type	Participant
Value Type	integer
Repeatable	No

Below the properties is a 'Categories' table:

Name	Label	Missing
0	Male en Male	
1	Female en Female	
9	Missing	✓

Click on the "Derive" button and select "Categorize another variable to this..."

The screenshot shows the 'Derive Variable' dialog box open over the web interface. The dialog contains the following information:

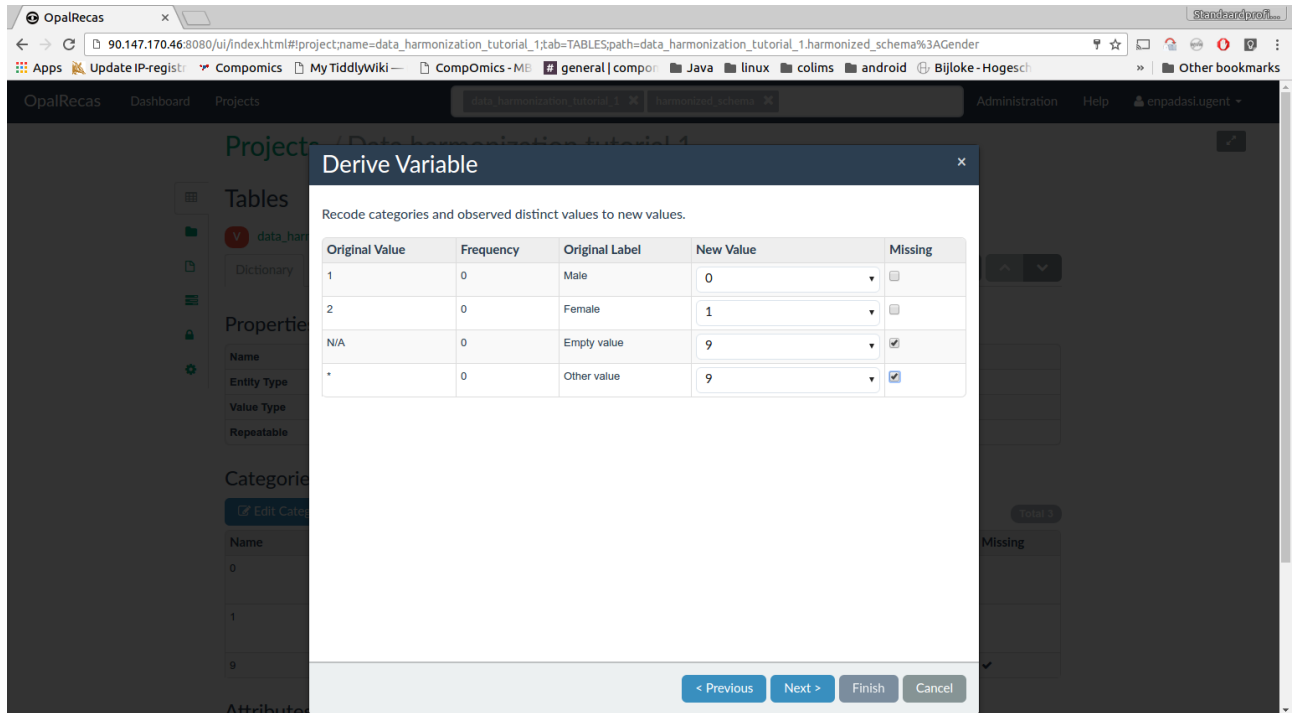
Select from which variable derivation will be performed. The resulting derived variable will be the current variable.

Table:

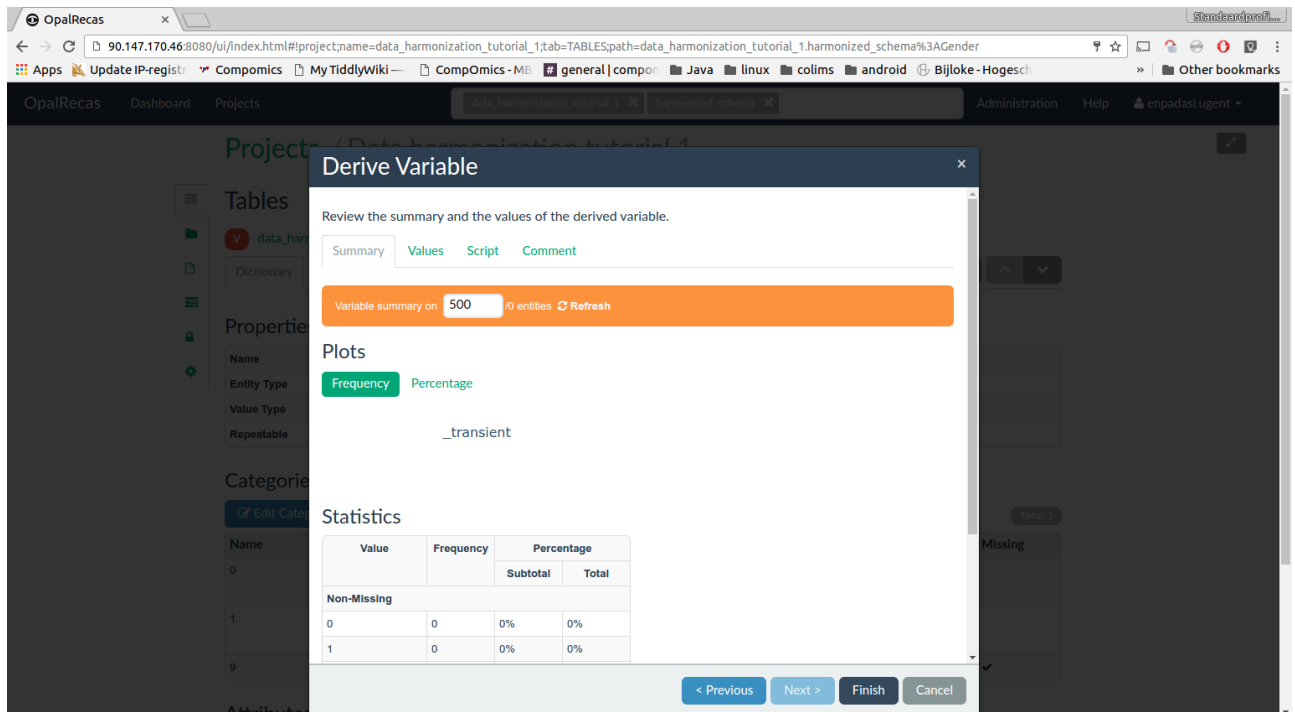
Variable:

At the bottom of the dialog are the buttons: < Previous, Next >, Finish, Cancel.

Select the "Finrisk2007" table and the "SUKUP" variable.



Give each original value a new value in the "New Value" column using the drop-down boxes. Click on the "Next" button to review the variable derivation.



Click on the "Finish" button to finalize the harmonization of the "Gender" variable.



OpalRecas Dashboard Projects data_harmonization_tutorial_1 harmonized_schema Administration Help enpadasi.ugent

Projects / Data harmonization tutorial 1

Tables

data_harmonization_tutorial_1 / harmonized_schema / Gender

Dictionary Script Summary Values Permissions

Derive Edit History

```

$( '*SUKUP*' ).map( (
  *1*: *0*,
  *2*: *1*
),
),
*9*,
*9*);

```

© 2016 OBIba Documentation Sources 2.4.11-b20160308163700

Click the "Script" tab of the "Gender" variable to see the generated harmonization script.

OpalRecas Dashboard Projects data_harmonization_tutorial_1 harmonized_schema Administration Help enpadasi.ugent

data_harmonization_tutorial_1 / harmonized_schema

Dictionary Summary Values Permissions Download Export Copy

Properties

Name	harmonized_schema
Entity Type	Participant
Table References	data_harmonization_tutorial_1.Finrisk2007

Entity Filter

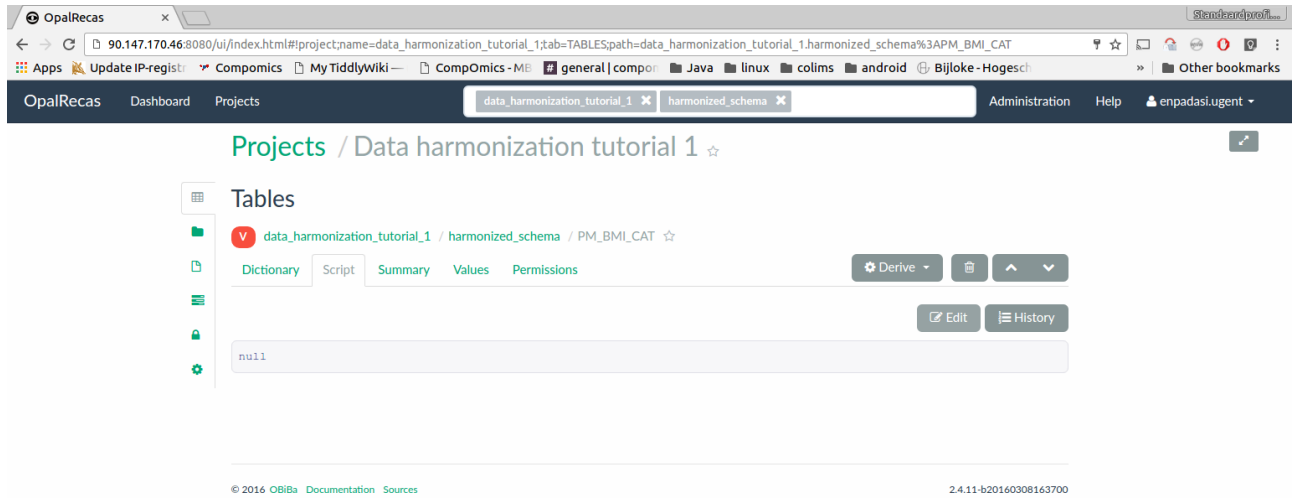
// no filter

Variables

+ Add Variables Filter variables... Total 5

Name	Label	Value Type	Categories
Gender		integer	0, 1, 9
PM_BMI_CAT		integer	1, 2, 3, 9
DIS_DIAB		integer	0, 1, 9
PM_DIASTOLIC		integer	999
SMK_CIG_CURRENT		integer	0, 1, 9

Click on the "PM_BMI_CAT" variable (Body Mass Index). It will be used as an example of a more complex transformation.



OpalRecas x Standardprofil...

90.147.170.46:8080/ui/index.html#project=name=data_harmonization_tutorial_1;tab=TABLES;path=data_harmonization_tutorial_1.harmonized_schema%3APM_BMI_CAT

Apps Update IP-registr Compomics My TiddlyWiki CompOmics - ME general | compo Java linux colims android Bijloke - Hogesch Other bookmarks

OpalRecas Dashboard Projects data_harmonization_tutorial_1 harmonized_schema Administration Help enpadasi.ugent

Projects / Data harmonization tutorial 1

Tables

data_harmonization_tutorial_1 / harmonized_schema / PM_BMI_CAT

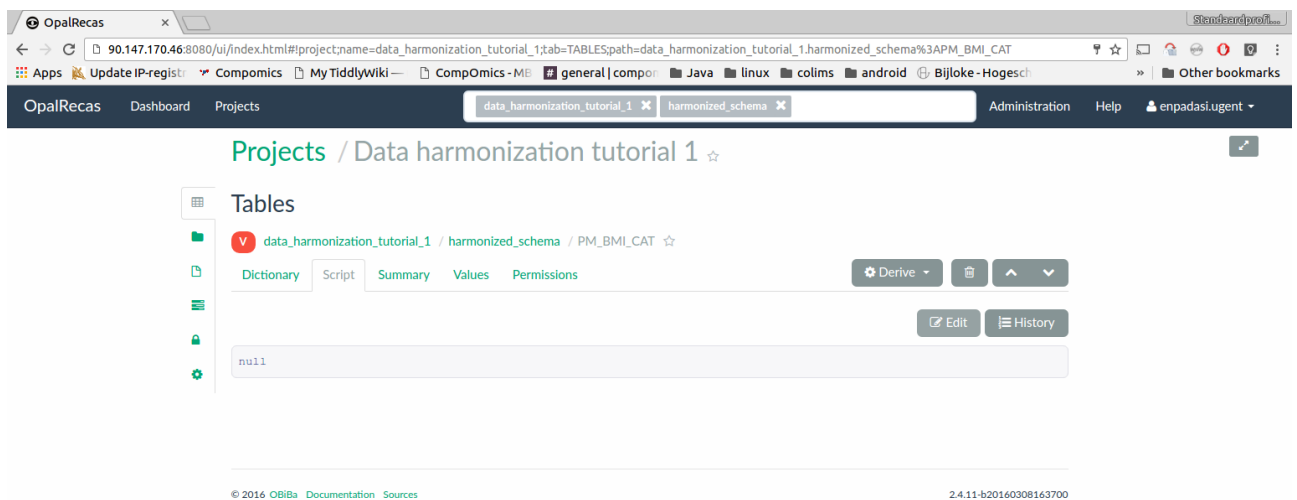
Dictionary Script Summary Values Permissions

Derive Edit History

null

© 2016 OBIba Documentation Sources 2.4.11-b20160308163700

Click on the "PM_BMI_CAT" variable (Body Mass Index). It will be used as an example of a more complex transformation.



OpalRecas x Standardprofil...

90.147.170.46:8080/ui/index.html#project=name=data_harmonization_tutorial_1;tab=TABLES;path=data_harmonization_tutorial_1.harmonized_schema%3APM_BMI_CAT

Apps Update IP-registr Compomics My TiddlyWiki CompOmics - ME general | compo Java linux colims android Bijloke - Hogesch Other bookmarks

OpalRecas Dashboard Projects data_harmonization_tutorial_1 harmonized_schema Administration Help enpadasi.ugent

Projects / Data harmonization tutorial 1

Tables

data_harmonization_tutorial_1 / harmonized_schema / PM_BMI_CAT

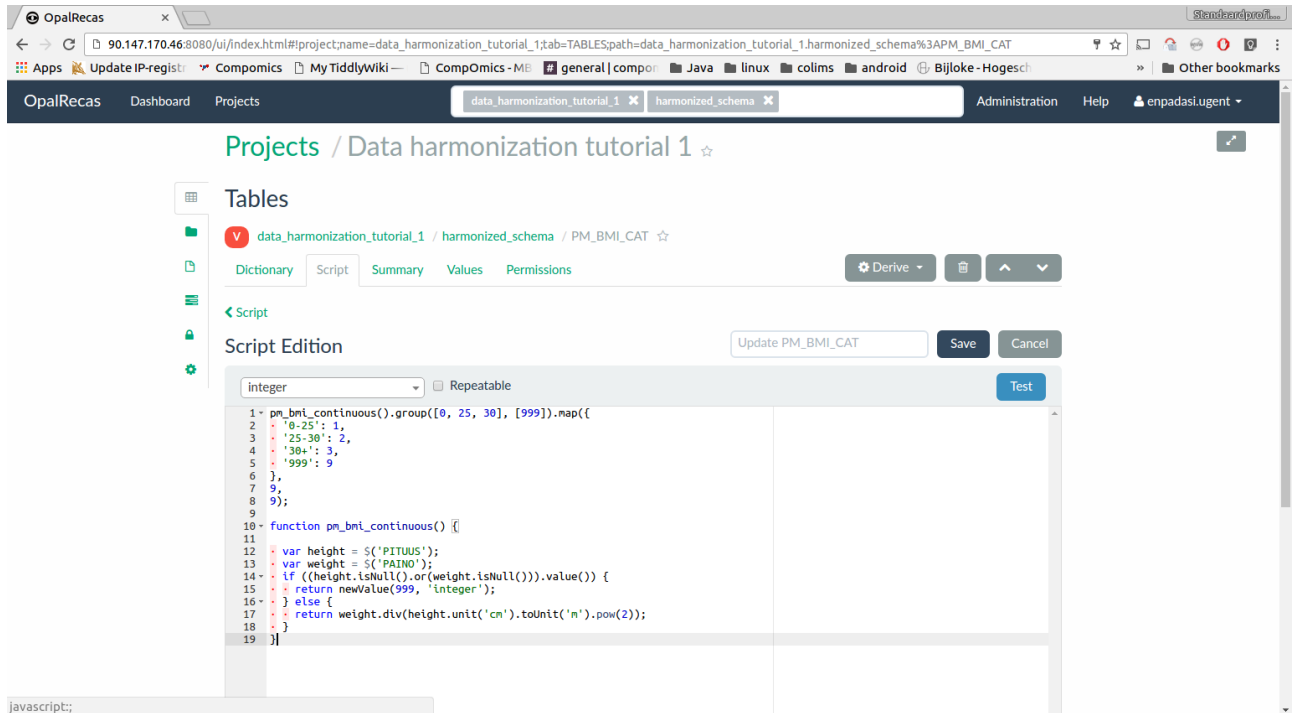
Dictionary Script Summary Values Permissions

Derive Edit History

null

© 2016 OBIba Documentation Sources 2.4.11-b20160308163700

After selecting the "Script" tab, click on the "Edit" button.



The screenshot shows the 'Script Edition' interface for the 'PM_BMI_CAT' variable. The script is as follows:

```

1 pm_bmi_continuous().group([0, 25, 30], [999]).map({
2   '0-25': 1,
3   '25-30': 2,
4   '30+': 3,
5   '999': 9
6 },
7 9,
8 9);
9
10 function pm_bmi_continuous() {
11
12   var height = S('P1TIUUS');
13   var weight = S('P2IINO');
14   if ((height.isNull().or(weight.isNull()).value()) {
15     return newValue(999, 'integer');
16   } else {
17     return weight.div(height.unit('cm').toUnit('m').pow(2));
18   }
19 }

```

Copy the transformation script for the "PM_BMI_CAT" variable from the "Tutorial_Algorithms.doc" file, paste it into the script editor and click on the "Save" button. Do the same thing for the remaining variables.



ANNEX 3. MICA GUIDE

MICA - add study

ID	Acronym	Name	Published	Status	Actions
atlantic-path	en Atlantic PATH	en Atlantic Partnership for Tomorrow's Health (The)	★	Draft	✎
bcgp	en BCGP	en BC Generations Project	★	Draft	✎
cag	en CaG	en CARTaGENE	★	Draft	✎
cls	en CLS	en Canberra Longitudinal Study	★	Draft	✎
clsa	en CLSA	en Canadian Longitudinal Study on Aging	★	Draft	✎
finrisk-2007	en FINRISK 2007	en National FINRISK Study 2007	★	Draft	✎
frele	en FRéLE	en FRéLE: Fragilité, une étude longitudinale de ses expressions [Frailty, a longitudinal study of its expressions]	★	Draft	✎
lasa	en LASA	en Longitudinal Aging Study Amsterdam	★	Draft	✎
lbs	en LBSL	en Long Beach Longitudinal Study	★	Draft	✎
lifelines	en LifeLines	en LifeLines Cohort Study & Biobank	★	Draft	✎
nuage	en NuAge	en Quebec Longitudinal Study on Nutrition and Successful Aging	★	Draft	✎
ohs	en OHS	en Ontario Health Study	★	Draft	✎
shin	en SHIP	en Study of Health in Pomerania	★	Draft	✎

After navigating to the studies tab in the main menu, click on the "Add Study" button.

General Information

Name *

Official study name

Acronym *

Acronym of the study

Objectives *

What are the principal objectives of the study?

Start Year

End Year

Study Design

Cohort
 Case-control
 Case only
 Cross-sectional
 Clinical trial
 Other

Other

Follow Up

Supplementary Information

Sources of recruitment
 Individuals
 Families



Studies / New Study

English

General Information

Name *
LifeLines Cohort Study & Biobank

Official study name

Acronym *
LifeLines_demo

Acronym of the study

Objectives *
Overall aim of the study: Unravel the interaction between genetic and environmental factors in the development of multifactorial diseases, their concurrent development in individuals and their complications as a complex trait. Primary objectives: Which are the disease

What are the principal objectives of the study?

Start Year
2004

Study Design

- Cohort
- Case-control
- Case only
- Cross-sectional
- Clinical trial
- Other

Other
[Empty text box]

Follow Up
[Empty text box]

Supplementary Information
Funding Agencies: Government, 3 northern provinces of The Netherlands (80%), Uni

Sources of recruitment
[Empty text box]

lifelines.jpeg

SHOW ALL

End Year
2036

Website
https://www.lifelines.nl/

Logo Choose file
lifelines.jpeg 4.72 KB

+ Add Attribute

Access
Access to external researchers or third parties provided or foreseen for

- Data
- Biosamples

Other sources
[Empty text box]

Number of participants

Target number of participants
165000

No Limit

Target number of participants with biological samples
[Empty text box]

No Limit

Supplementary Information
The project will include 165000 participants: 45000 probands, 30000 partners, 55000 |

Marker Paper
threegeneration population-based study, Eur J Epidemiol 2008; 23(1): 67-74

lifelines.jpeg

SHOW ALL

Fill in all the required fields and click on the "Save" button on the bottom of the page



On the study main page, click on the "Add contact" or "Add investigator" button. Fill in the required fields and click on the "Save" button.

MICA - add population



On the study main page, click on the "Add Population" button.

Studies / lifelines_demo

Add Population

English

General Information

ID *
pop_1

Name *
Lifelines population

Description
(18-65 years) or the elderly cohort (>65 years). The participants are followed for at least 30 years and are invited every 5 years for a visit to the LifeLines research facility. In the years in between the participants receive follow-up questionnaires.

Selection criteria

Gender
N/A

Minimum age
25

Maximum age

Health Status
[+ Add health status](#)

Ethnic Origin
[+ Add ethnic origin](#)

Criteria

SHOW ALL

Number of participants

Target number of participants
165000

No Limit

Target number of participants with biological samples
[Empty field]

No Limit

Supplementary Information
Current number of participants recruited: 93 000 (On June, 2012). Current number of [Empty field]

Recruitment

Supplementary information
[Empty field]

General population
 Specific population
 Participants from existing studies
 Other

General Population | Specific Population | Participants from existing studies

Other sources

Volunteer enrolment
 Selected sample
 Random digit dialing

Cancel Save

© 2016 OBIa Documentation Sources 1.3.1

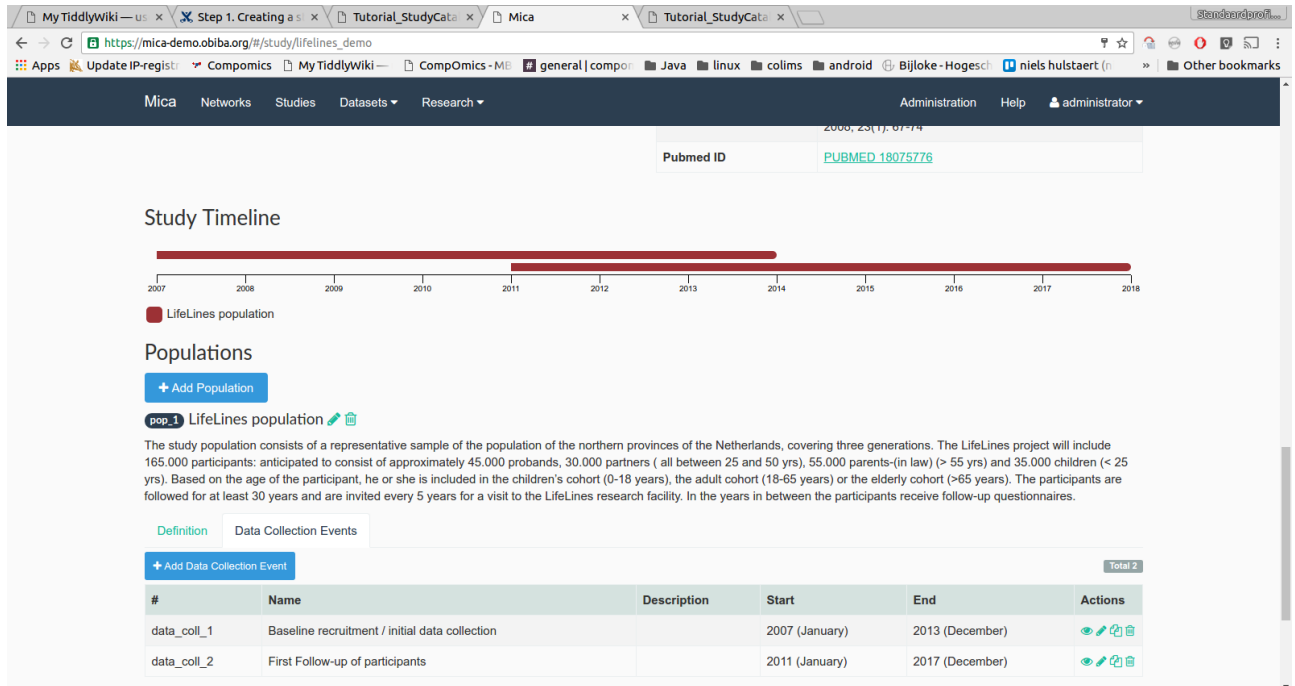
Fill in the required fields and click on the "Save" button.



The screenshot shows the Mica web application interface. At the top, there are navigation tabs for 'Mica', 'Networks', 'Studies', 'Datasets', and 'Research'. Below this, the 'Study Timeline' section is displayed, with a sub-section for 'Populations'. A button labeled '+ Add Population' is present. Underneath, the 'LifeLines population' is listed with a small icon. A detailed description of the population is provided: 'The study population consists of a representative sample of the population of the northern provinces of the Netherlands, covering three generations. The LifeLines project will include 165.000 participants: anticipated to consist of approximately 45.000 probands, 30.000 partners (all between 25 and 50 yrs), 55.000 parents-(in law) (> 55 yrs) and 35.000 children (< 25 yrs). Based on the age of the participant, he or she is included in the children's cohort (0-18 years), the adult cohort (18-65 years) or the elderly cohort (>65 years). The participants are followed for at least 30 years and are invited every 5 years for a visit to the LifeLines research facility. In the years in between the participants receive follow-up questionnaires.' Below the description, there are tabs for 'Definition' and 'Data Collection Events'. The 'Data Collection Events' tab is selected, and a button labeled '+ Add Data Collection Event' is visible. Below this, the 'Authorization' section shows a table with two rows, both indicating 'Not authorized' for different authorization sources. At the bottom of the page, there is a footer with copyright information and version numbers.

On the study main page in the "Population" section, click on the "Data Collection Events" tab. Then click on the "Add Data Collection Event" button.

The screenshot shows the 'Add Data Collection Event' form in the Mica web application. The form is titled 'General Information' and contains several input fields and checkboxes. The 'ID' field contains 'data_coll_1'. The 'Name' field contains 'Baseline recruitment / initial data collection'. The 'Description' field is empty. The 'Start Year' field contains '2007'. The 'Start Month' field is empty. The 'End Year' field contains '2013'. The 'End Month' field is empty. Below the form, there is a section for 'Data sources' with checkboxes for 'Questionnaires', 'Physical measures', 'Biosamples', 'Administrative databases', and 'Others'. The 'Questionnaires' checkbox is checked. At the bottom of the form, there are 'Cancel' and 'Save' buttons.



2006, 28(11), 67-74

Pubmed ID [PUBMED 18075776](#)

Study Timeline

2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

LifeLines population

Populations

+ Add Population

pop_1 LifeLines population

The study population consists of a representative sample of the population of the northern provinces of the Netherlands, covering three generations. The LifeLines project will include 165.000 participants: anticipated to consist of approximately 45.000 probands, 30.000 partners (all between 25 and 50 yrs), 55.000 parents-(in law) (> 55 yrs) and 35.000 children (< 25 yrs). Based on the age of the participant, he or she is included in the children's cohort (0-18 years), the adult cohort (18-65 years) or the elderly cohort (>65 years). The participants are followed for at least 30 years and are invited every 5 years for a visit to the LifeLines research facility. In the years in between the participants receive follow-up questionnaires.

Definition Data Collection Events

+ Add Data Collection Event Total 2

#	Name	Description	Start	End	Actions
data_coll_1	Baseline recruitment / initial data collection		2007 (January)	2013 (December)	View Edit Delete
data_coll_2	First Follow-up of participants		2011 (January)	2017 (December)	View Edit Delete

Fill in the required fields and click on the "Save" button on the bottom of the page. The added date collection events are shown in a time-line on the study main page.