Opal Upload Vignette

A tutorial on how to upload clinical data onto the Opal server

www.eucanshare.eu

Dr. Andrej-Nikolai Spiess
Dr. Anna Lena Engels

v. 1.1 (03/12/2020)

Data Pre-processing

Step 01: Create the Raw Data File (best formatted in Excel)

1 2 5

ID	SEX	AGE	BMI	SMOKER	FSMOKER	SYSTC	DIASTC	HRATE	GLUC	BNP_0h	BNP_3h	BNP_6h
ID001	2	41	26.06	1	1	105	116	77	79.14	8.3	32	78.8
ID002	1	40	19.08	0	0	115	88	89	87.35	9.2	34.5	34.3
ID003	2	40	26.01	1	0	115	97	77	77.46	7.6	26.5	45.2
ID004	1	47	25.56	0	0	120	93	85	77.21	13.4	32.9	81.3
ID005	1	39	24.67	0	0	133	100	76	91.4	10	29.6	83.7
ID006	2	48	23.33	0	0	143	98	69	80.46	12.1	39.3	94.8
ID007	1	32	20.41	0	0	135	88	92	75.74	4.8	30.4	96.8
ID008	1	3	19.07	0	0	119	94	87	73.57	9.1	33.3	79.4
ID009	1	41	23.07	0	0	102	81	97	97.14	8.6	27.2	101.5
ID010	2	41	23.43	0	0	116	99	81	87.35	7.6	29.5	61.1
ID011	1	42	22.51	0	0	121	85	91	75.18	13.1	35.1	60.9
ID012	2	37	23.18	1	1	128	114	100	87.25	7.2	22	47.4
ID013	1	38	25.42	0	0	130	73	77	86.59	10.6	22	86.8
ID014	1	35	19.5	0	0	127	85	67	94.13	11.7	34.9	83.2
ID015	2	35	22.19	0	0		98	78	90.98	10.4	22.8	88.6
ID016	1	42	26.23	0	0	128	95	78	87.33	8.2	28.2	72.8
ID017	2	42	26.52	1	1	116	84	82		11.9	33.2	71.4
ID018	2	40	21.14	0	0	122	80	97	92.32	10.3	33.1	75.7
ID019	2	45	22.19	0	0	121	91	77	82.3	7.9	28.8	41.7
ID020	1	50	26.71	0	0	124	90	84	79.78	7.2	23.5	50.1
ID021	1	38	26.68	0	0	120	72	78	86.36	14.2	39.1	60.3
ID022	1	28	21.66	0	0	103	90	80	79.01	8.6	37.3	108.4
ID023	1	45	20.32	4 -99	0	127	92	83	94.46	6.3	30.7	68.8
ID024	3	36	25.34	0	0	124	92	93	82.94	11.1	21.4	87.2
ID025	1	37	19.93	1	1	117	79	81	73.96	10.6	37.6	117.9
ID026	1	45	19.3	0	0	121	95	87	97.75	7.3	27	73.2
ID027	1	39	27.99	0	0	121	104	88	77.86	6.1	28.7	84.4
ID028	1	34	19.31	0	0	122	95	89	82.95	9.8		47.9
ID029	2	41	22.05	0	0	136	79	74	88.89	12.3	31.8	65.4
ID030	2	39	27.24	1	1	118	86	96	88.97	11.3	30.9	82
ID031	1	40	24.56	1	1	122	93	76	80.86	9	34.1	90.1
ID032	1	42	21.58	0	0	132	84	79	75.75	8.3	31.9	66.7
ID033	1	4 -88	25.64	0	0	131	68	94	96.7	10.5	35.5	139.2
ID034	1	43	26.51	1	1	131	99	93	88.18	10.3	42.2	83
ID035	2	39	21.83	0	0	114	82	69	92.24	11.3	34.6	85.9
ID036	2	42	23.43	0	0	140	84	71	96.81	9.2	30.2	93.2
ID037	1	45	25.28	0	0	121	105	66	86.28	11.8	20.7	70.2

1 ID's in first column with name „ID", preferably without spaces and in capital letters.

2 All categorical data in numbers, so for instance SEX: 1 = male, 2 = female, 3 = transgender.

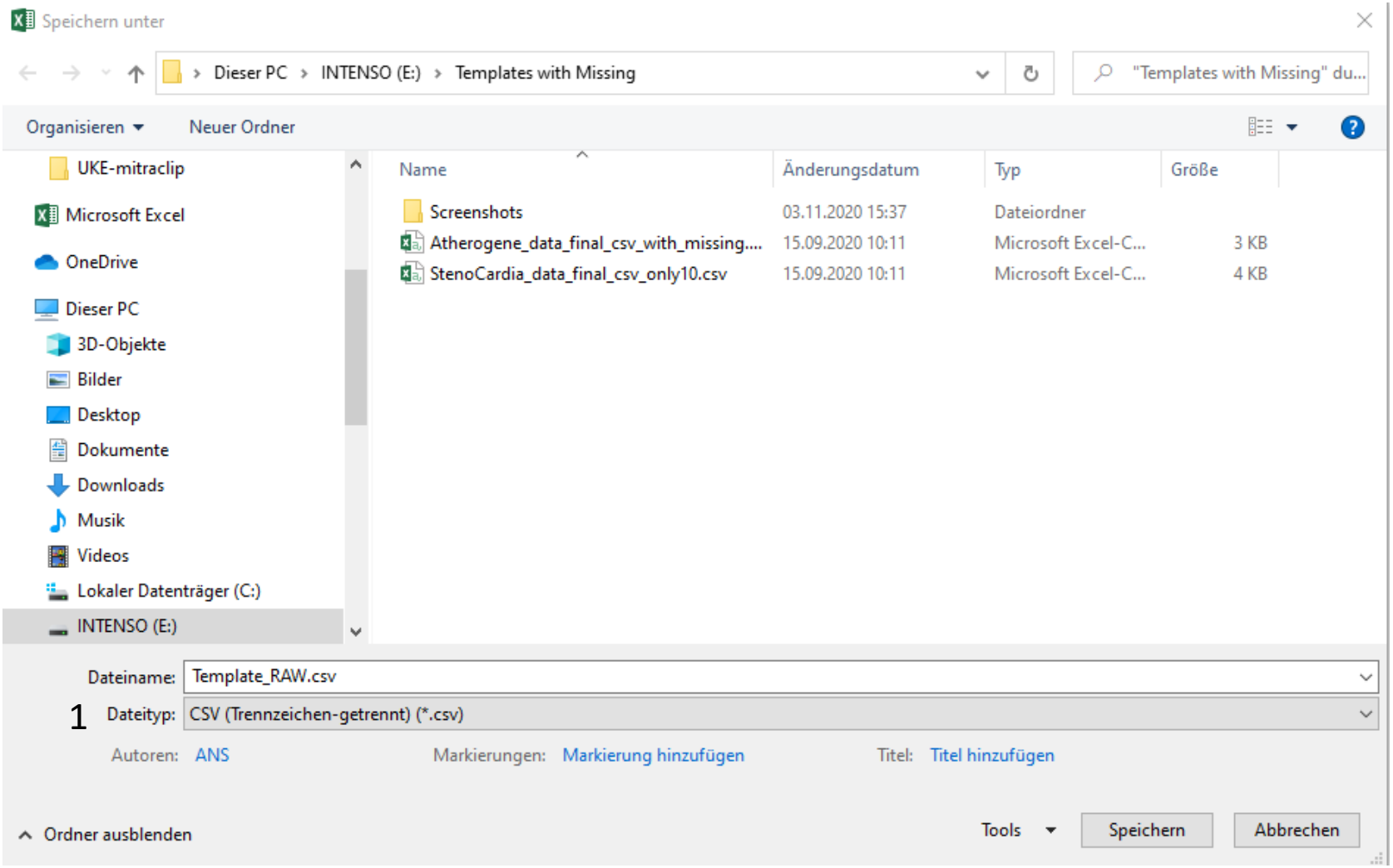
3 All technical empty data (e.g. non-measured) just as empty cells in Excel. Currently, Opal does not support definitions such as NA, NULL, #NV. This will translate to x1, x2, , x3, x4… when exported to csv-file.

4 For any other questionaire scenario, supply numbers that do not interfere with the data type. For example: -88 for „Don't know" in **AGE** or -99 for „Don't want to answer" in **SMOKER** do not interfere with the data ranges.

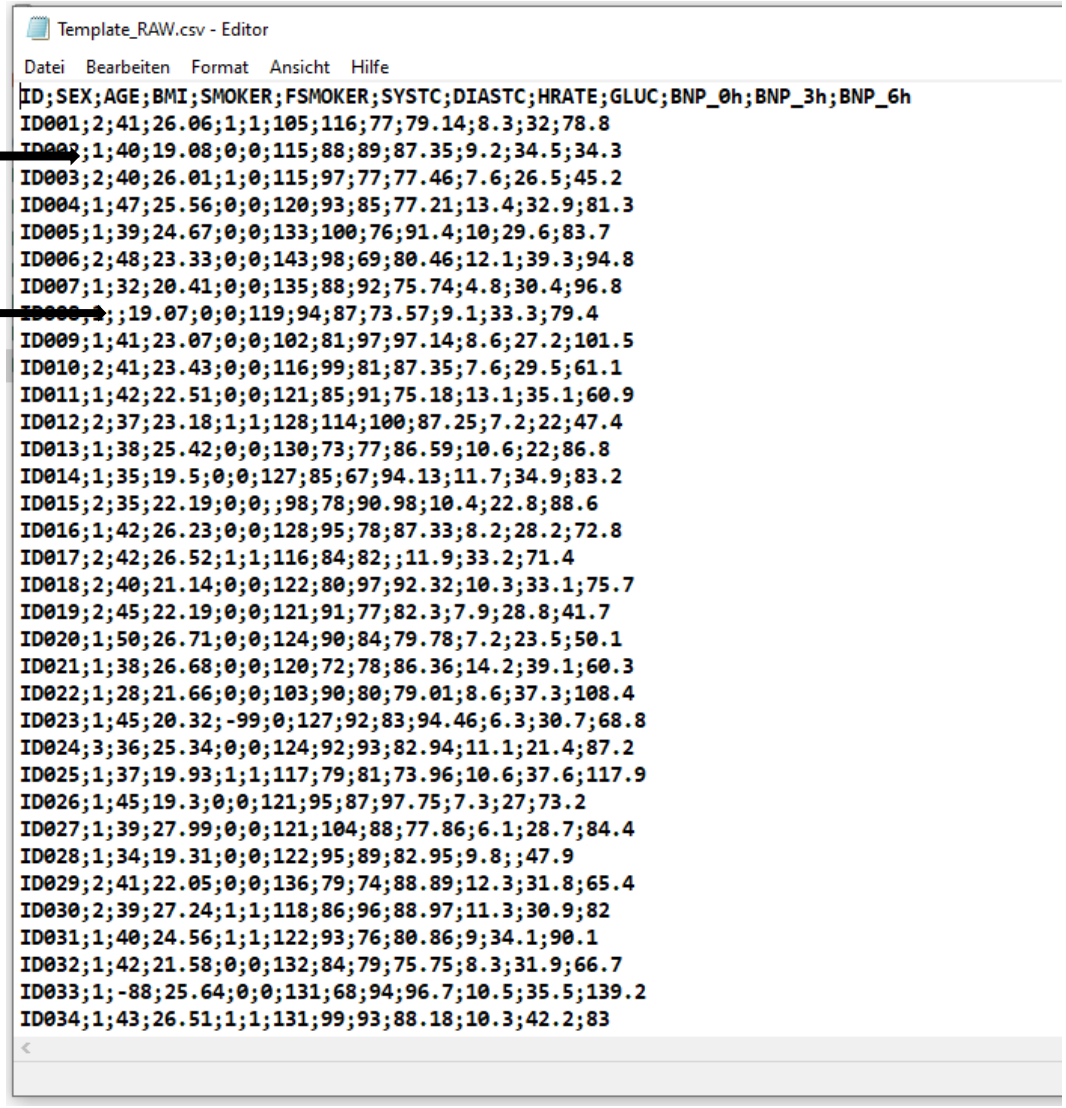
5 Sensible variable names without spaces, but underscores.

Step 02: Export to .csv file

1 Save your Excel file in .xlsx and ALSO in .csv (Comma delimited) format



Step 03: Inspect your exported .csv file



Depending on computer or locale,
Excel might export your data
comma-separated or semicolon-separated!
Open your .csv file with a right mouse click
=> Open with => Editor to inspect this.

1 In this case, the data was exported semicolon-separated.

2 We also see the empty cell in Age, ID008 => ;;

WE WILL NEED TO REMEMBER
THE SEPARATOR-SYMBOL (;) FOR LATER!

Step 04-1: Create your dictionary file

The dictionary file is an Excel file with two MANDATORY tabs, „Variables“ and „Categories“

The „Variables“ tab contains 7 columns (A-G) that have to be named EXACTLY like shown here. These contain:
table: a comprehensive name of the study, also later to be used for the uploaded project.
name: the column names of all the variables in the raw data file, EXCLUDING first column „ID“
valueType: the numeric type, if continuous data => „decimal“, if categorical => „integer“ (e.g. male = 1).
Also possible: „boolean“ (True/False).
unit: the measurement unit, for non-units (e.g. SEX) => „none“, else the clinical/chemical/physical unit.
label:en : a short label for the variable.
description:en : a longer and more detailed description of the study variable.
script: a special-formatted character string of the variables for database indexing. See next page on how to create this.

	A	B	C	D	E	F	G	H	I
1	**table**	**name**	**valueType**	**unit**	**label:en**	**description:en**	**script**	**makeScript**	
2	template	SEX	integer	none	Sex	Sex	$('SEX')	$('SEX')	
3	template	AGE	decimal	years	Age	Age at baseline	$('AGE')	$('AGE')	
4	template	BMI	decimal	kg/m^2	BMI	Body-mass-index	$('BMI')	$('BMI')	
5	template	SMOKER	integer	none	Current smoke	Current smoker	$('SMOKER')	$('SMOKER')	
6	template	FSMOKER	integer	none	Former smoker	Former smoker	$('FSMOKER')	$('FSMOKER')	
7	template	SYSTC	decimal	mmHg	Systolic	Systolic blood pressure	$('SYSTC')	$('SYSTC')	
8	template	DIASTC	decimal	mmHg	Diastolic	Diastolic blood pressure	$('DIASTC')	$('DIASTC')	
9	template	HRATE	decimal	1/s	Rate	Heart rate	$('HRATE')	$('HRATE')	
10	template	GLUC	decimal	mg/dL	Glucose	Blood glucose levels	$('GLUC')	$('GLUC')	
11	template	BNP_0h	decimal	ng/L	BNP 0h	B-type natriuretic peptide 0h after admission	$('BNP_0h')	$('BNP_0h')	
12	template	BNP_3h	decimal	ng/L	BNP 3h	B-type natriuretic peptide 3h after admission	$('BNP_3h')	$('BNP_3h')	
13	template	BNP_6h	decimal	ng/L	BNP 6h	B-type natriuretic peptide 6h after admission	$('BNP_6h')	$('BNP_6h')	
14									
42									
43									
44									
45									

Variables | Categories

Step 04-2: Create your dictionary file

The creation of the **script** column (G) is somewhat tricky. The created character strings in **script** need to be automatically generated from the variable **name** in column B. This can be achieved by

1 Creating a new column H called **makeScript** and typing in (EXACTLY) the string ="$('"&B2&"')" in H2.
Here it is important to not confuse single (') quotes (red) and double (") quotes (green)!
2 After pressing ENTER, an automatic string should appear with the variable name of B2 included
3 Click on the lower right rectangle grabber and pull it down until end of variable, which are then auto-filled.

1

H2 ="$('"&B2&"')"

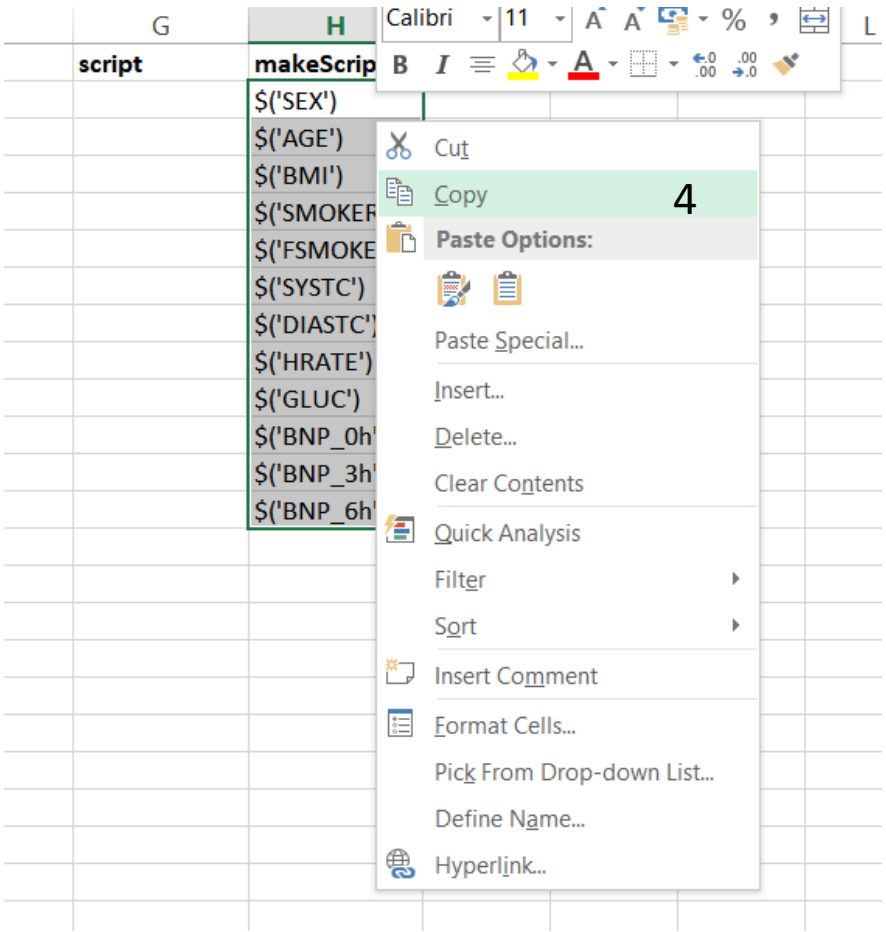
	A	B	C	D	E	F	G	H
1	**table**	**name**	**valueType**	**unit**	**label:en**	**description:en**	**script**	**makeScript**
2	template	SEX	integer	none	Sex	Sex		$('SEX')
3	template	AGE	decimal	years	Age	Age at baseline		$('AGE')
4	template	BMI	decimal	kg/m^2	BMI	Body-mass-index		$('BMI')
5	template	SMOKER	integer	none	Current smoker	Current smoker		$('SMOKER')
6	template	FSMOKER	integer	none	Former smoker	Former smoker		$('FSMOKER')
7	template	SYSTC	decimal	mmHg	Systolic	Systolic blood pressure		$('SYSTC')
8	template	DIASTC	decimal	mmHg	Diastolic	Diastolic blood pressure		$('DIASTC')
9	template	HRATE	decimal	1/s	Rate	Heart rate		$('HRATE')
10	template	GLUC	decimal	mg/dL	Glucose	Blood glucose levels		$('GLUC')
11	template	BNP_0h	decimal	ng/L	BNP 0h	B-type natriuretic peptide 0h after admission		$('BNP_0h')
12	template	BNP_3h	decimal	ng/L	BNP 3h	B-type natriuretic peptide 3h after admission		$('BNP_3h')
13	template	BNP_6h	decimal	ng/L	BNP 6h	B-type natriuretic peptide 6h after admission		$('BNP_6h')
14								

2

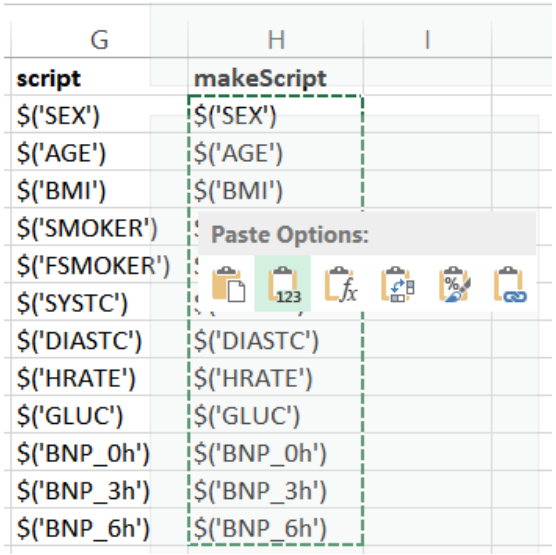
3

Step 04-3: Create your dictionary file

4 Unfortunately, this formula-based string can not be used for the server, so that we have to remove the formula-association. This is achieved by marking all variable cells in column H, and copying them to the clipboard.
5 We then go into the **script** column in G2 and right click => copy values („123“) and all cells in column G are automatically filled with the string in column **makeScript**, but with removed formula syntax.
Hooray! We are now finished with the „Variables“ tab...



5



Step 04-4: Create your dictionary file

The dictionary file is an Excel file with two MANDATORY tabs, „Variables“ and „Categories“

The „Categories“ tab contains 5 columns (A-E) that have to be named EXACTLY like shown here. These contain:
table: as in „Variables“, the same name of the study, also later to be used for the uploaded project.
variable: the name of all CATEGORICAL variables, multiplied by the number of levels of this variable. For instance, SEX has three levels (1 = male, 2 = female, 3 = transgender), then this need to appear three times in column B. Same with, for instance, SMOKER, replicated twice for the levels 0 = Not current smoker and 1 = Current smoker. Also, and this is very important, the variable names of ALL CONTINUOUS VARIABLES WITH AN INCLUDED CATEGORICAL LEVEL, i.e. since we had -88 for „Don‘t want to answer“ and -99 for „Don‘t know“ in variable AGE, the continuous variable AGE needs to be included here for these two levels.
name: the levels as they are found in the raw data file, e.g. 1/2/3 for SEX, -88/-99 for the two categories included in the age group, 0/1 for SMOKER etc.
missing: a binary definition, if the levels in name correspond to *bona fide* data => 0, or if they need to be seen as a missing value => 1. In this example -88/-99 correspond to answers that do not supply a value, hence „1“.
label:en : The level names corresponding to **name**, as they will appear on the server when displayed.

	A	B	C	D	E
1	**table**	**variable**	**name**	**missing**	**label:en**
2	template	SEX	1	0	male
3	template	SEX	2	0	female
4	template	SEX	3	0	transgender
5	template	AGE	-88	1	Don't know
6	template	AGE	-99	1	Don't want to answer
7	template	SMOKER	0	0	Not current smoker
8	template	SMOKER	1	0	Current smoker
9	template	SMOKER	-99	1	Don't want to answer
10	template	FSMOKER	0	0	Not former smoker
11	template	FSMOKER	1	0	Former smoker
12	template	FSMOKER	-99	1	Don't want to answer
13					

Variables | Categories

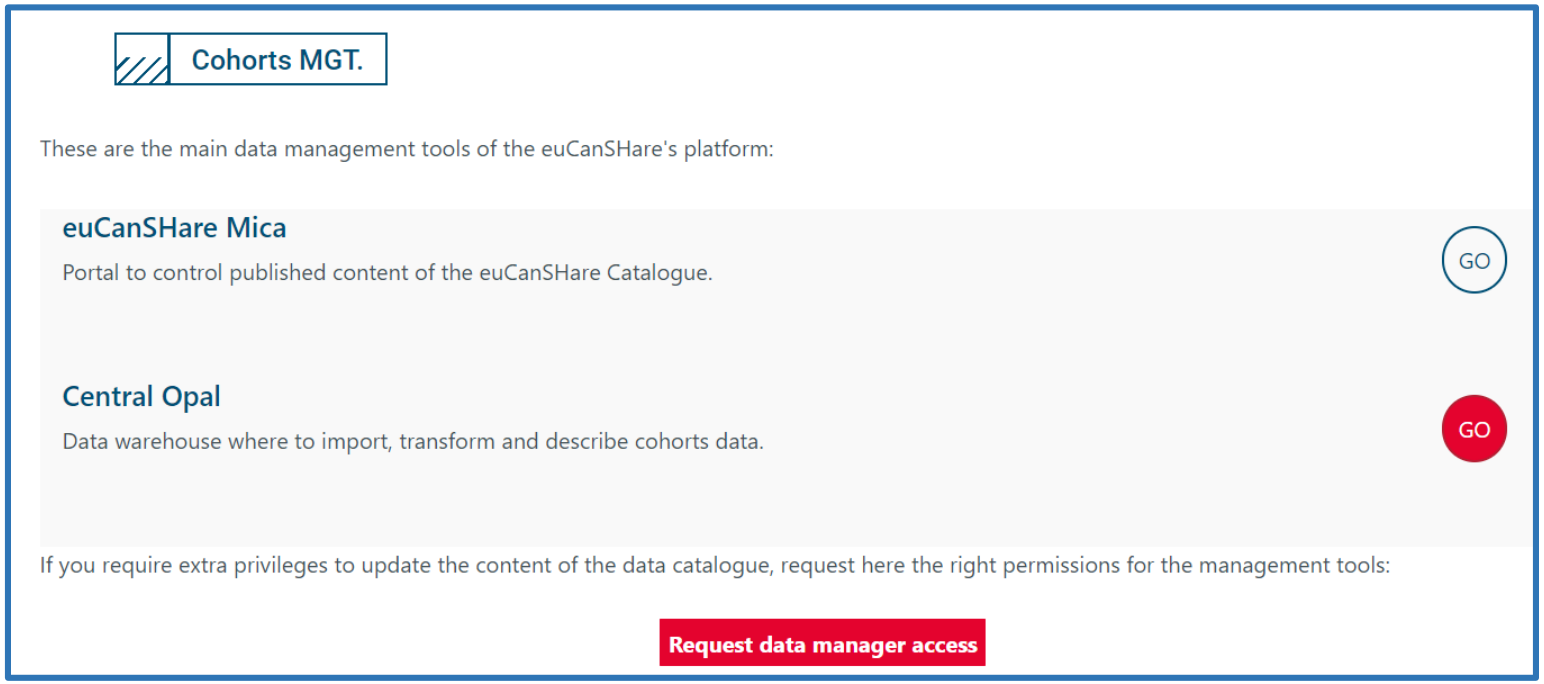
Good News:
We‘re (almost) done!
What remains is the...

Data Upload

Step 05-1: Log in

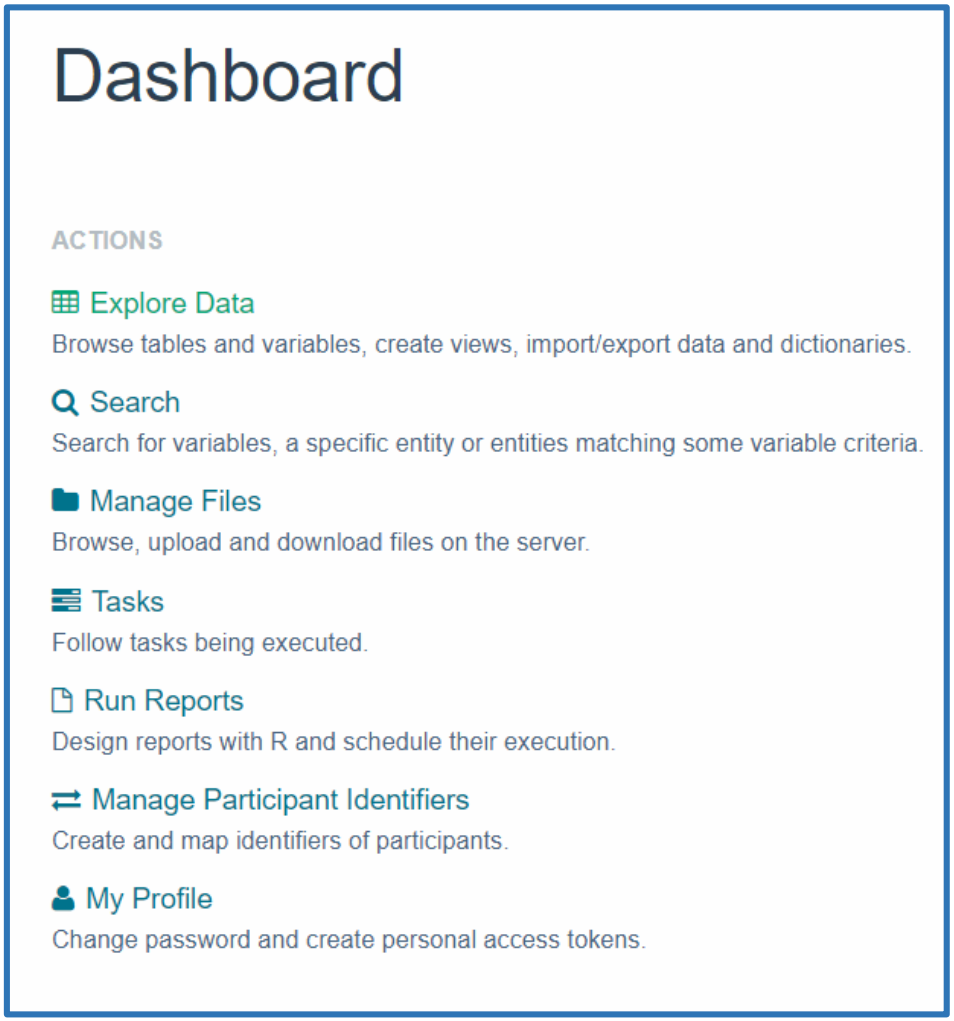
Log in at https://eucanshare.bsc.es/platform/ and click on „Cohorts Mgt.“ in the menu.
Then click on the „Central Opal GO“ button.
If you then have a „Sign in with EuCanSHare“ menu, just click on the
„Sign in with EuCanSHare“ button.

If you do not have access, apply for one with the „Request data manager access“ button.

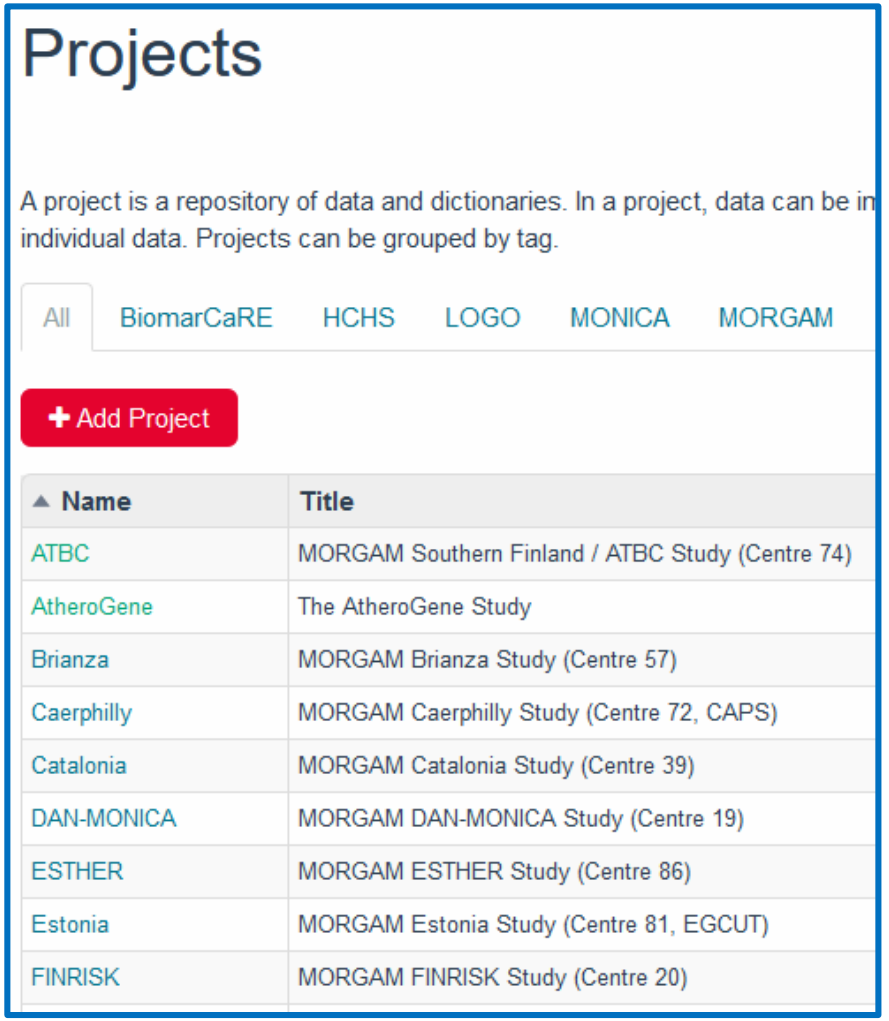


Step 05-2: Dashboard

Go to Dasboard and click on „Explore Data“

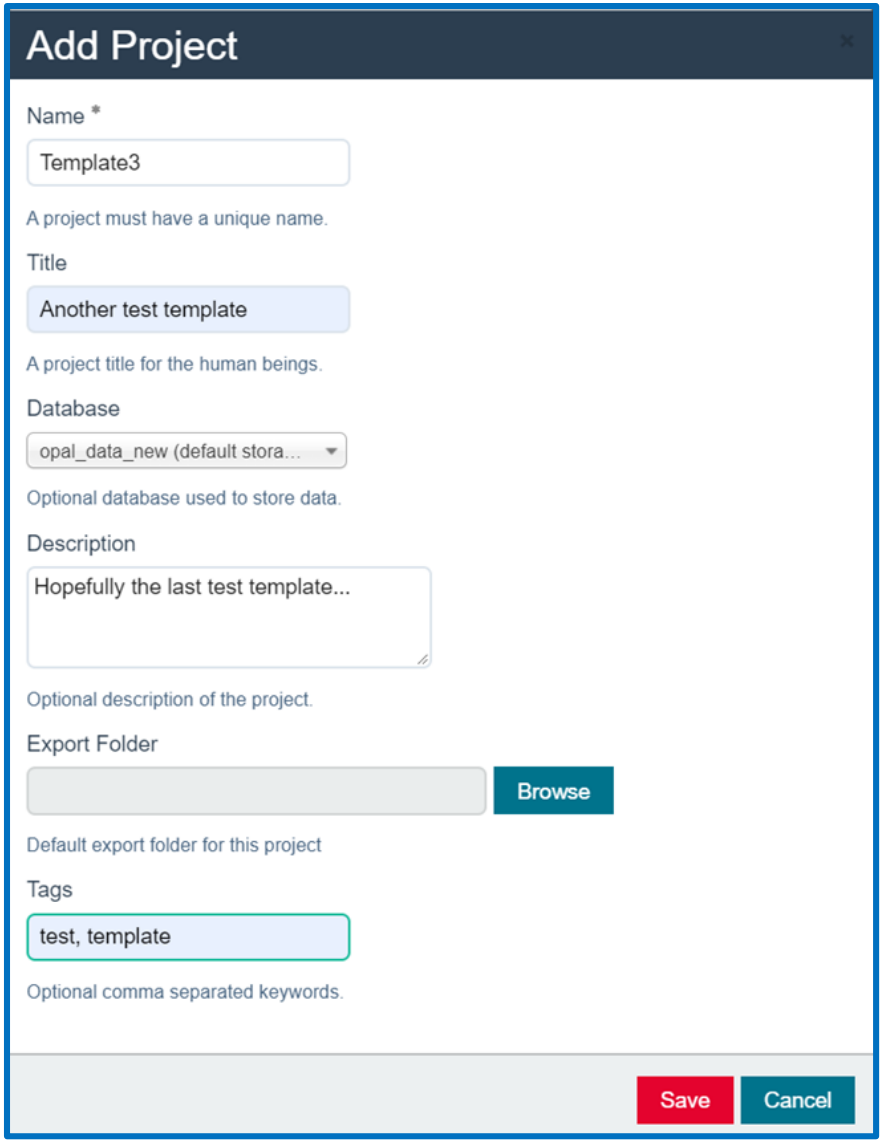


Click on „Add Project“



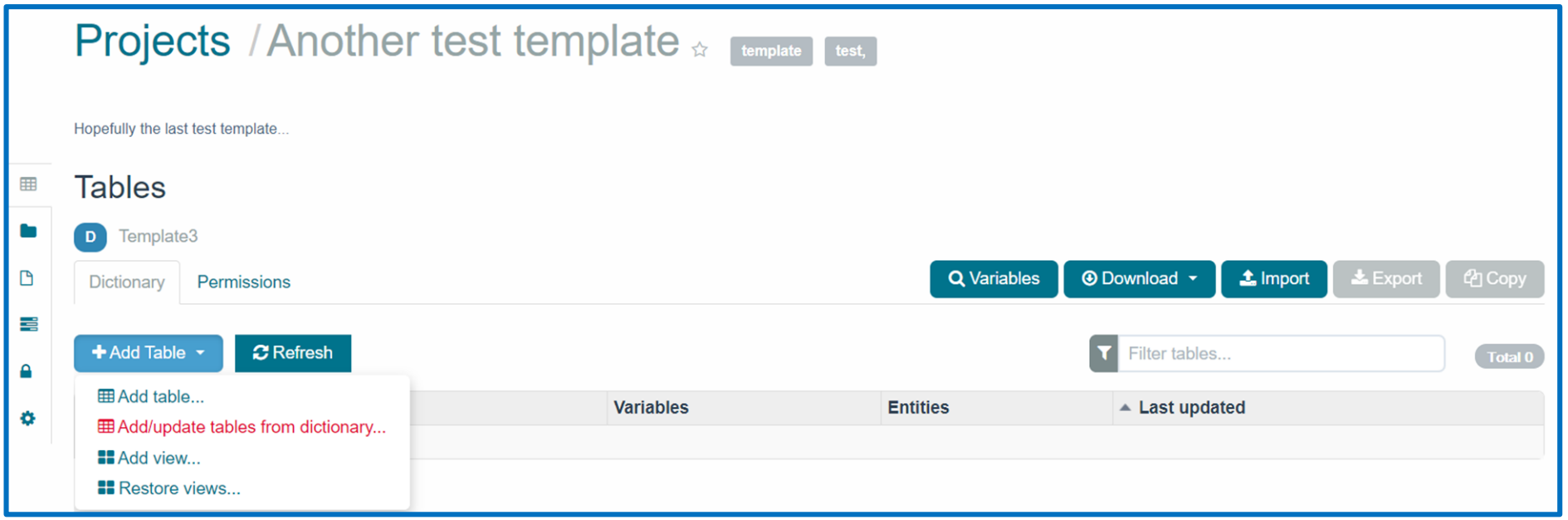
Step 05-3: Projects

Fill in project Name/Title/Description/Tags and click „Save“



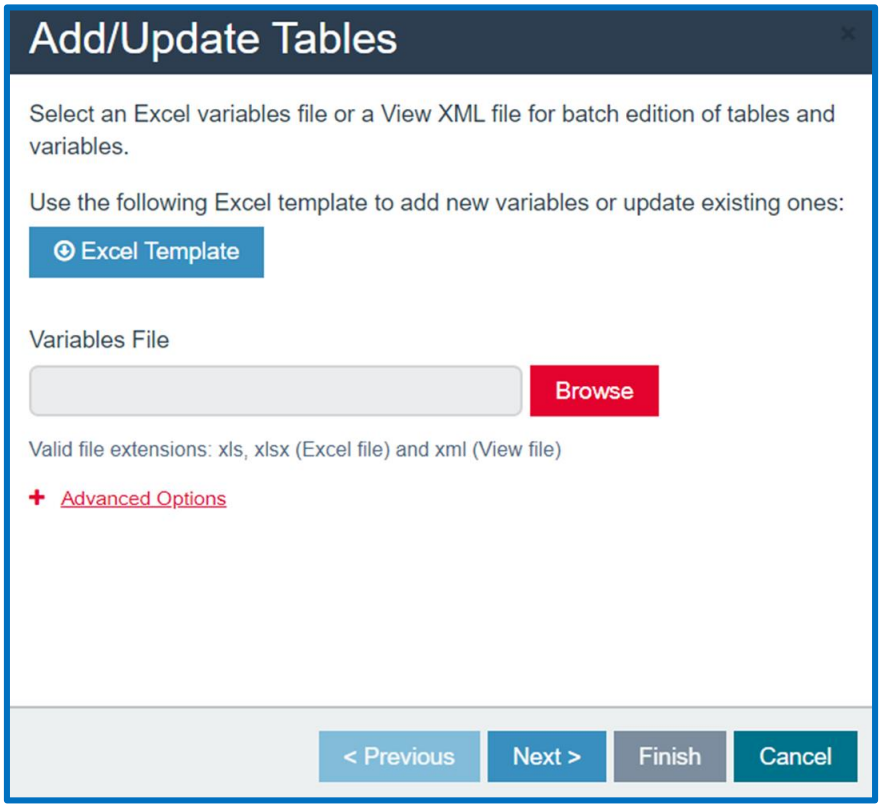
Step 05-4: Create Project Folder

We have now created a new Project with the **Title** name from the previous page. Click on „Add Table“ and then „Add/update tables from dictionary...“ to add a dictionary file.
IMPORTANT: The dictionary file always has to be uploaded first!

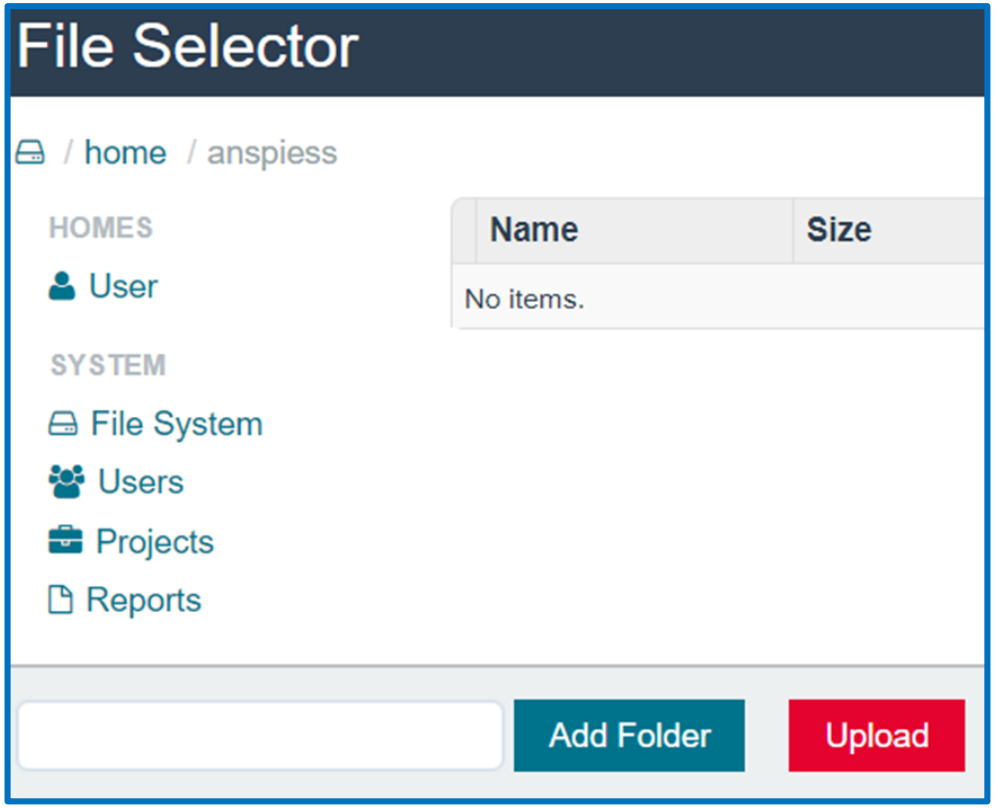


Step 05-5: Dictionary File

Click “Browse” to go to dictionary file

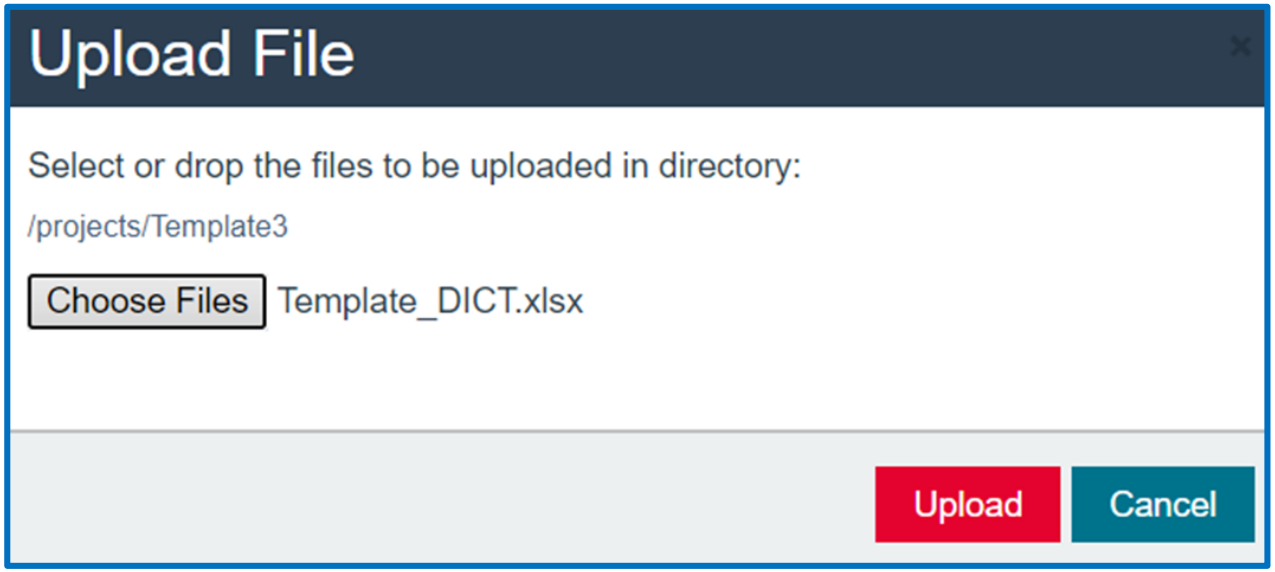


On the next page, click “Upload”



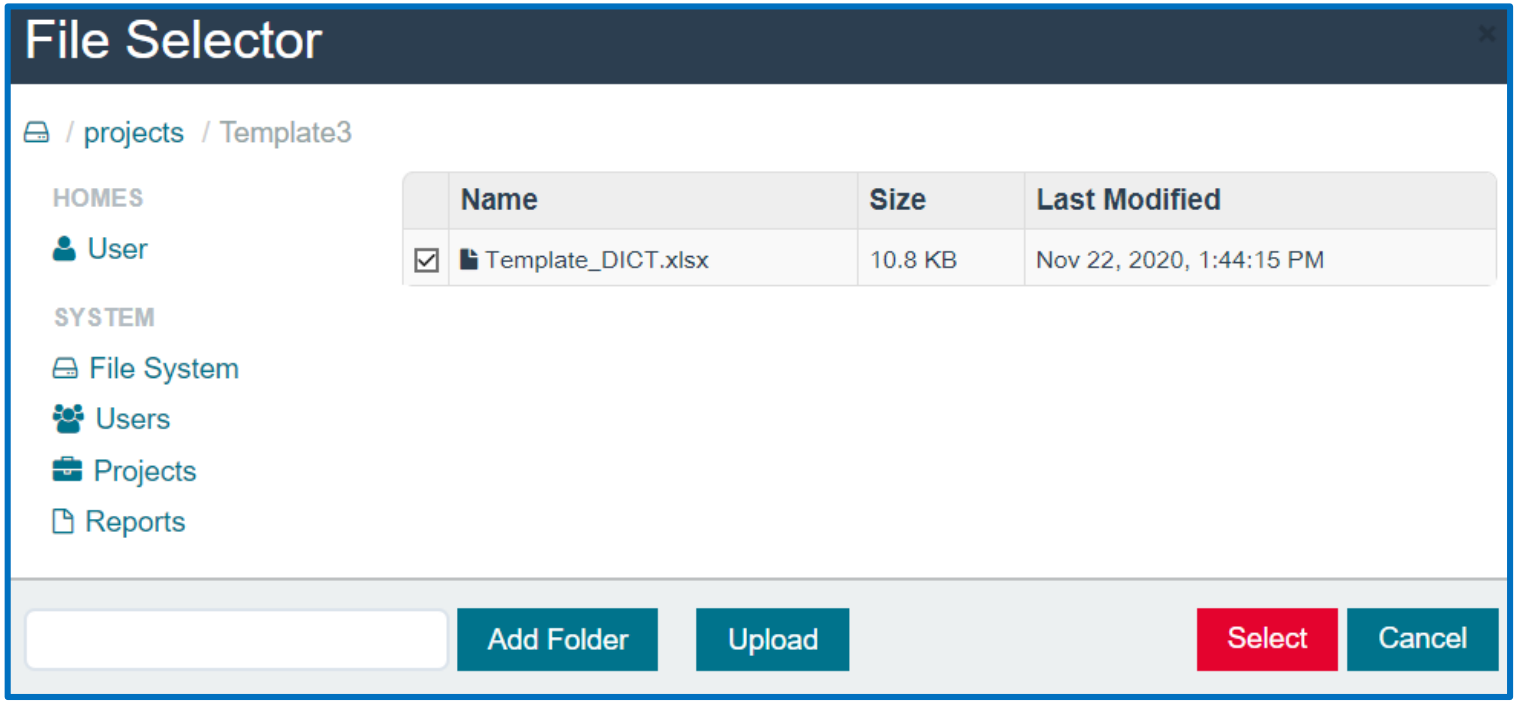
Step 05-6: Dictionary File

Select your dictionary file and click “Upload”



Step 05-7: Dictionary File

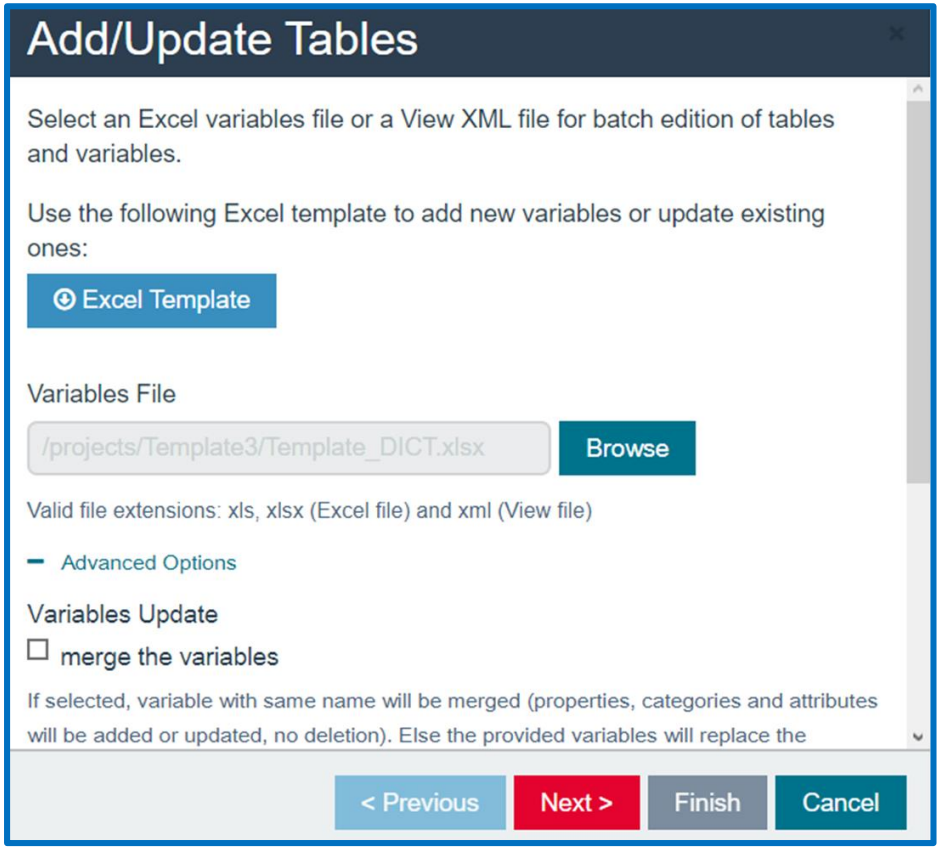
On the next page, tick-mark the dictionary file and the click on „Select“

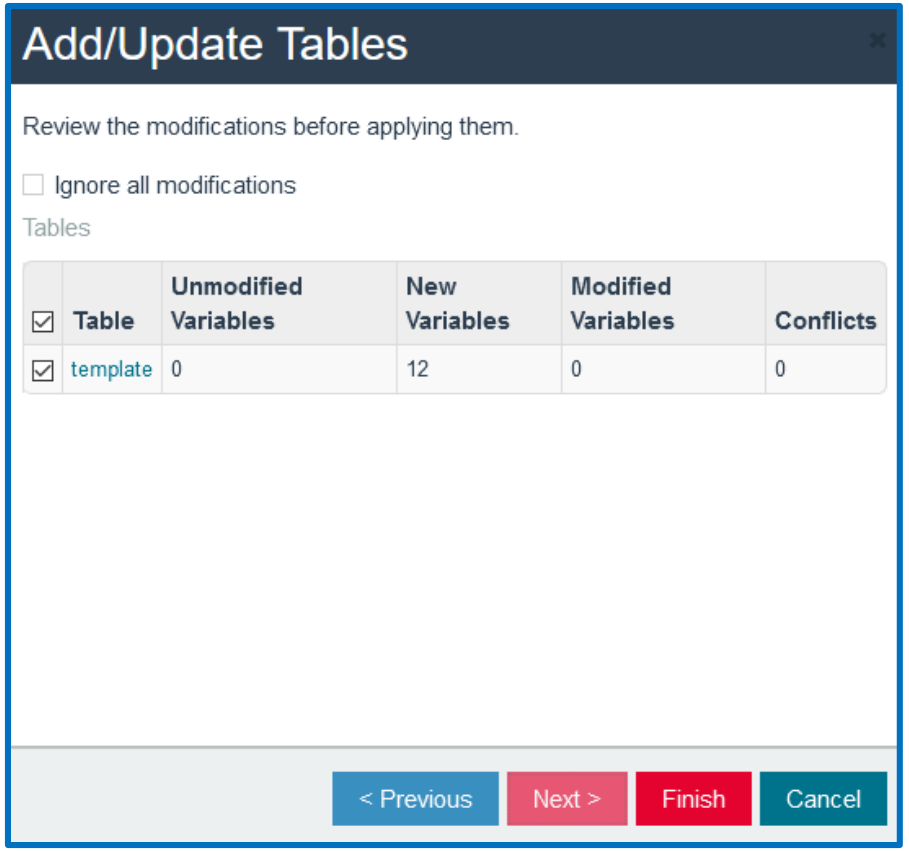


Step 05-8: Dictionary File

On the next page, click on „Next>“

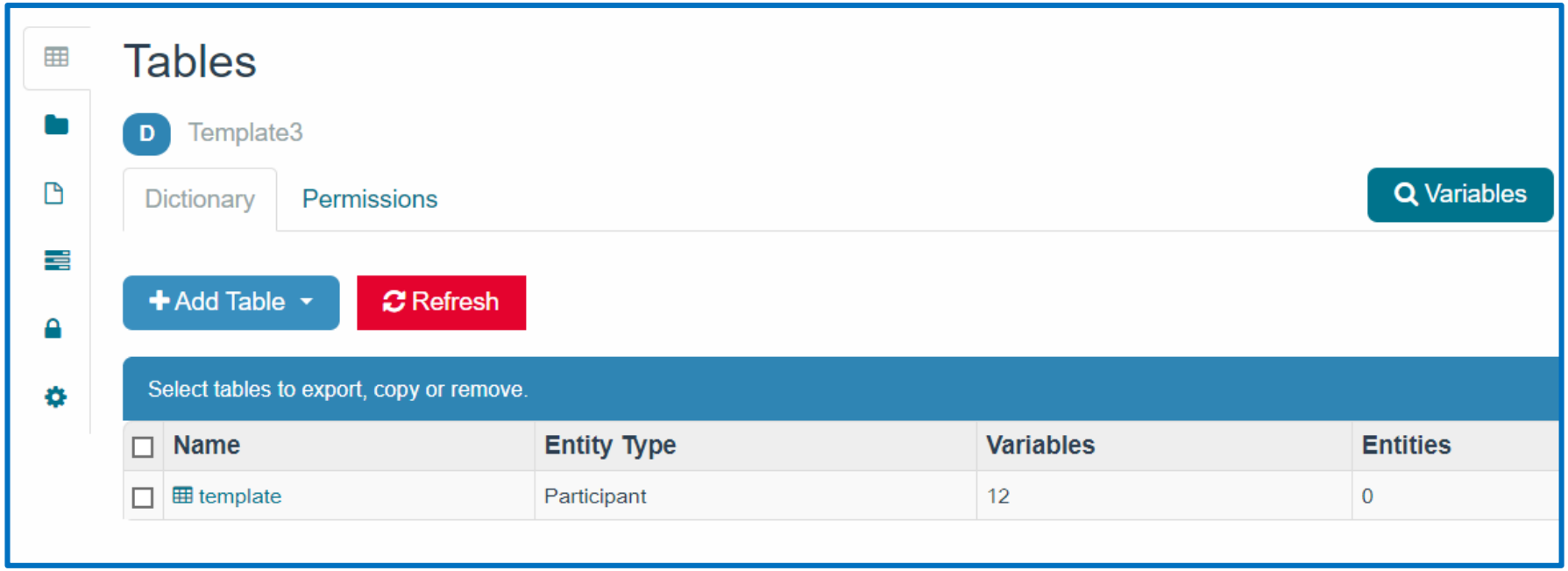
The next page shows the Table name (the **table** column in your dictionary file tab “Variables”), and the number of „New Variables“ should correspond to the number of entries in the **name** column of that tab. Then click „Finish“





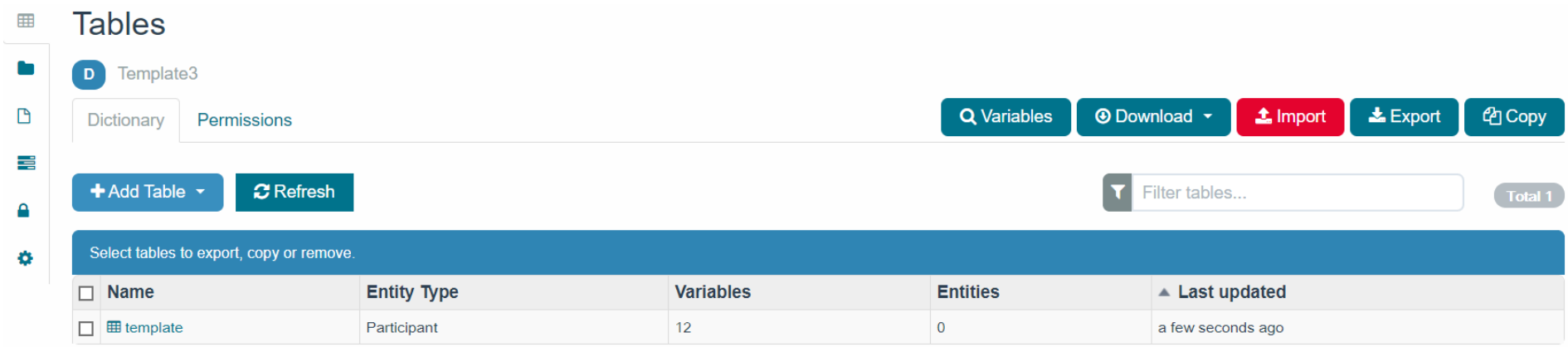
Step 05-9: Dictionary File

We are then on the Project page again, and the dictionary file has been added. If not, click on the „Refresh“ button. Note that „Entities“ are still zero because we have not added the raw data yet. We will do this now...



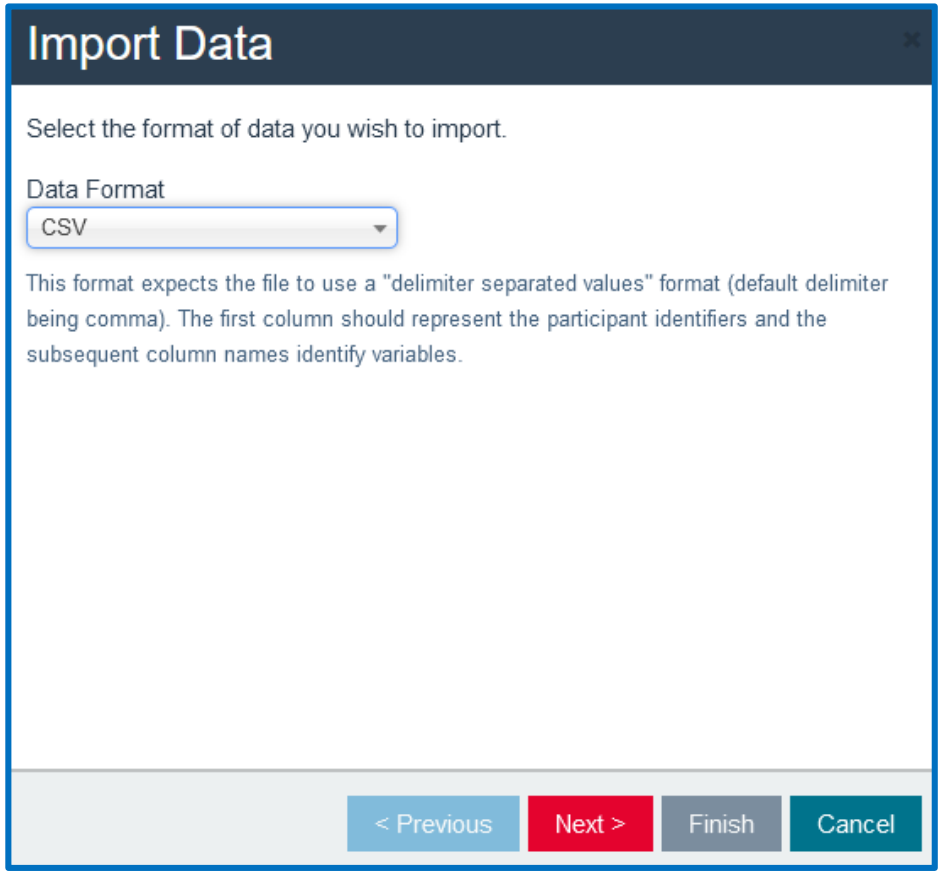
Step 06-1: Raw File

On the same Project page as before, click on the „Import“ button

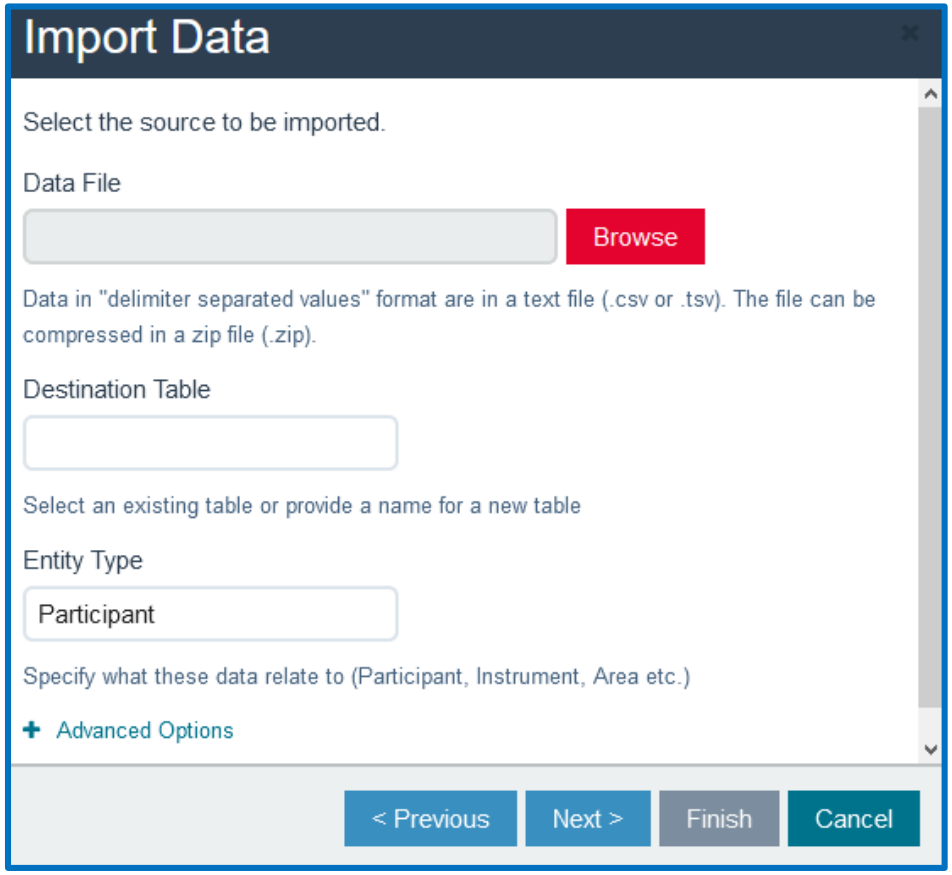


Step 06-2: Raw File

Select „CSV“ from dropdown menu in „Data Format“ and click „Next>“

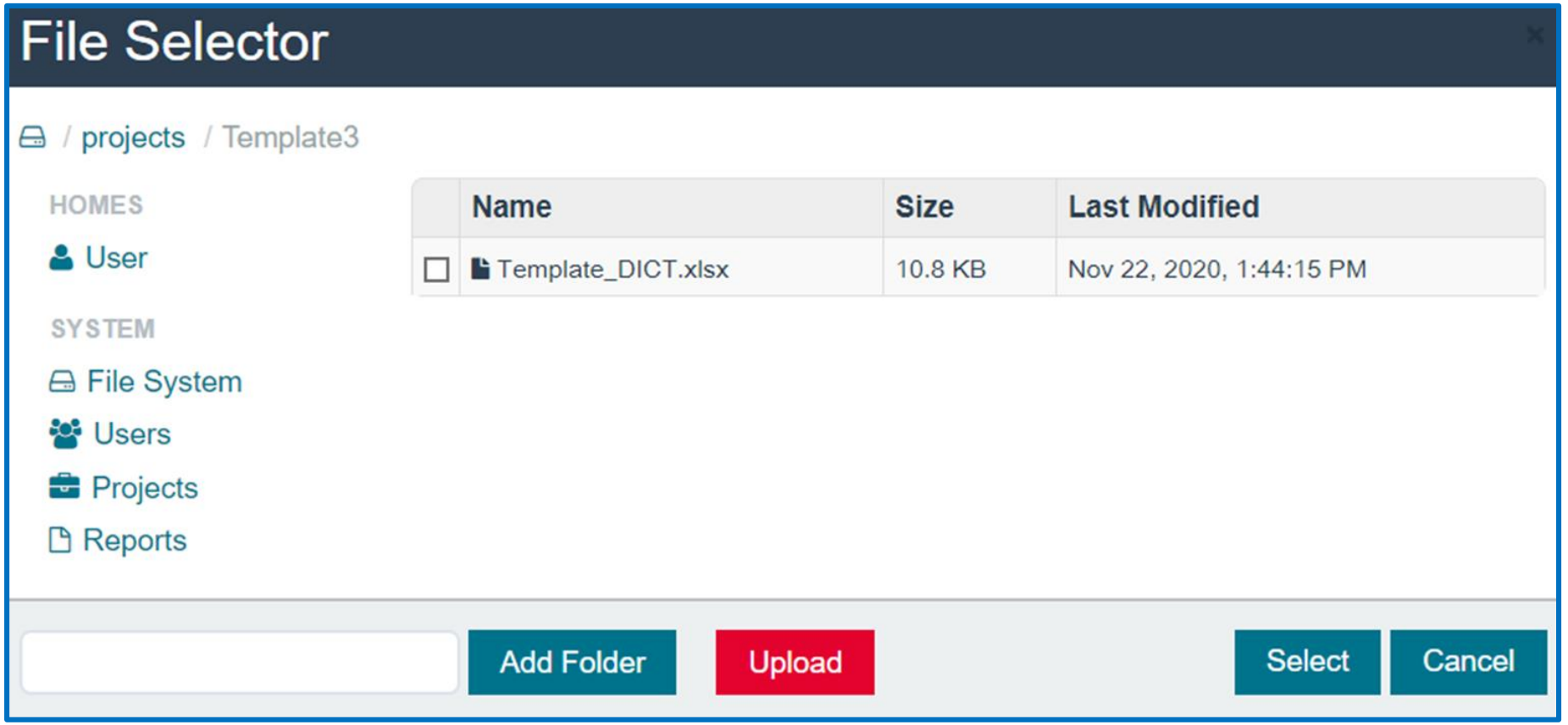


On the next page, click „Browse“ button and select your raw data file



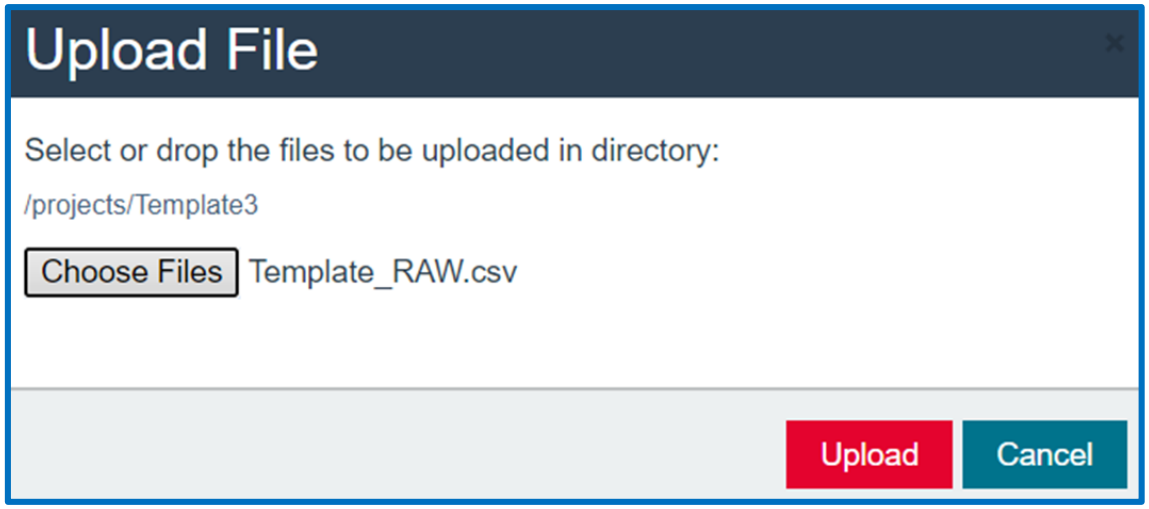
Step 06-3: Raw File

On the appearing page we see only the dictionary file. Click on „Upload“

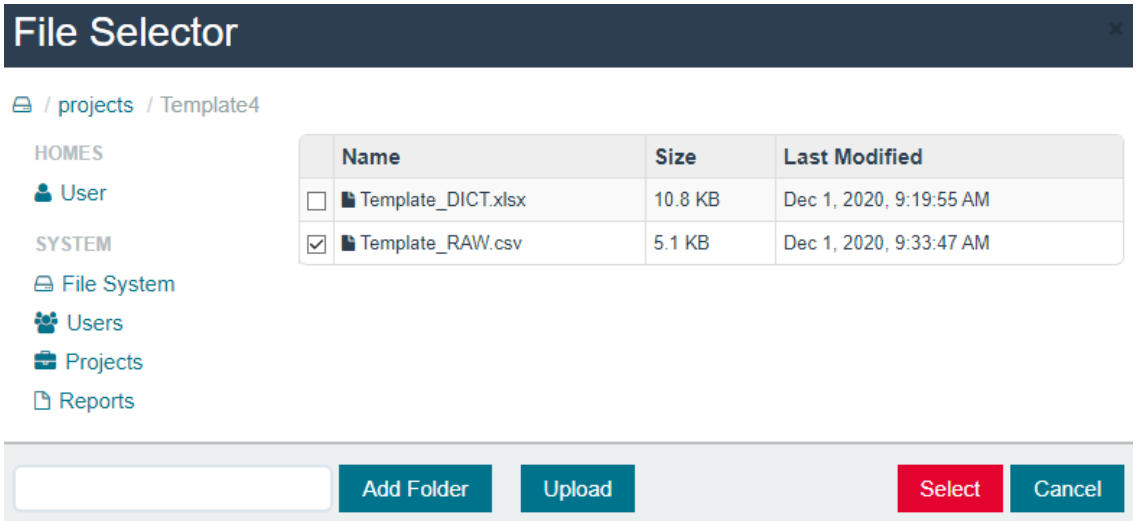


Step 06-4: Raw File

On the next page, choose your raw data file and click „Upload“

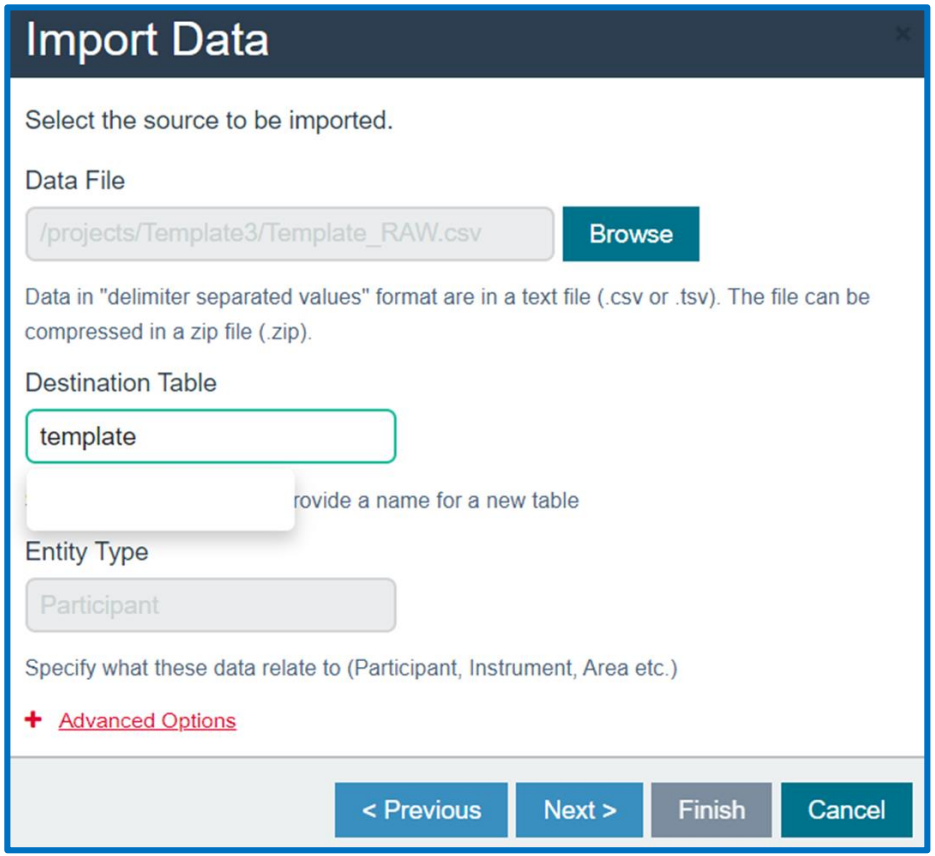


Then, the dictionary as well as the raw data should appear
Tick-mark the raw data file and click „Select“

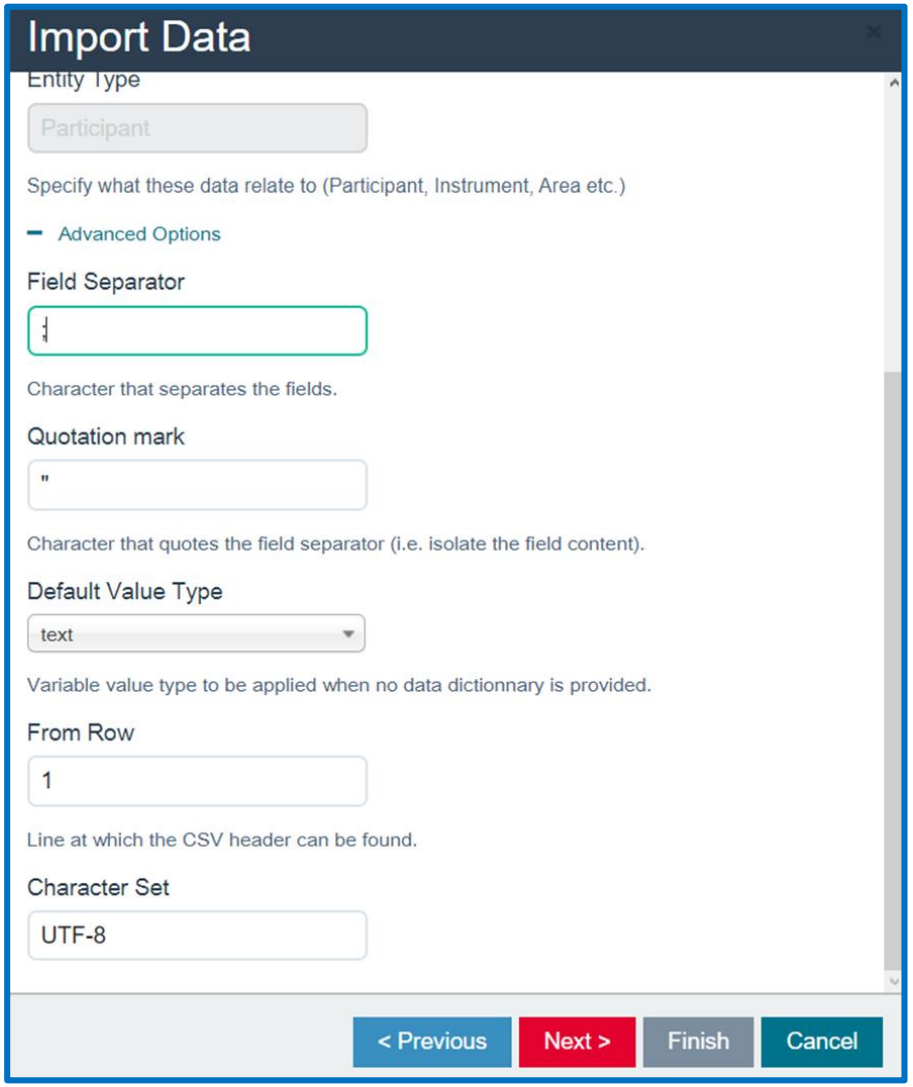


Step 06-5: Raw File

In the next page, the raw data has to be linked to the corresponding dictionary file („Destination Table“). The name that you define here has to be EXACTLY the same that the server has given for the dictionary file in Step 05-9 (in this case, „template“)!

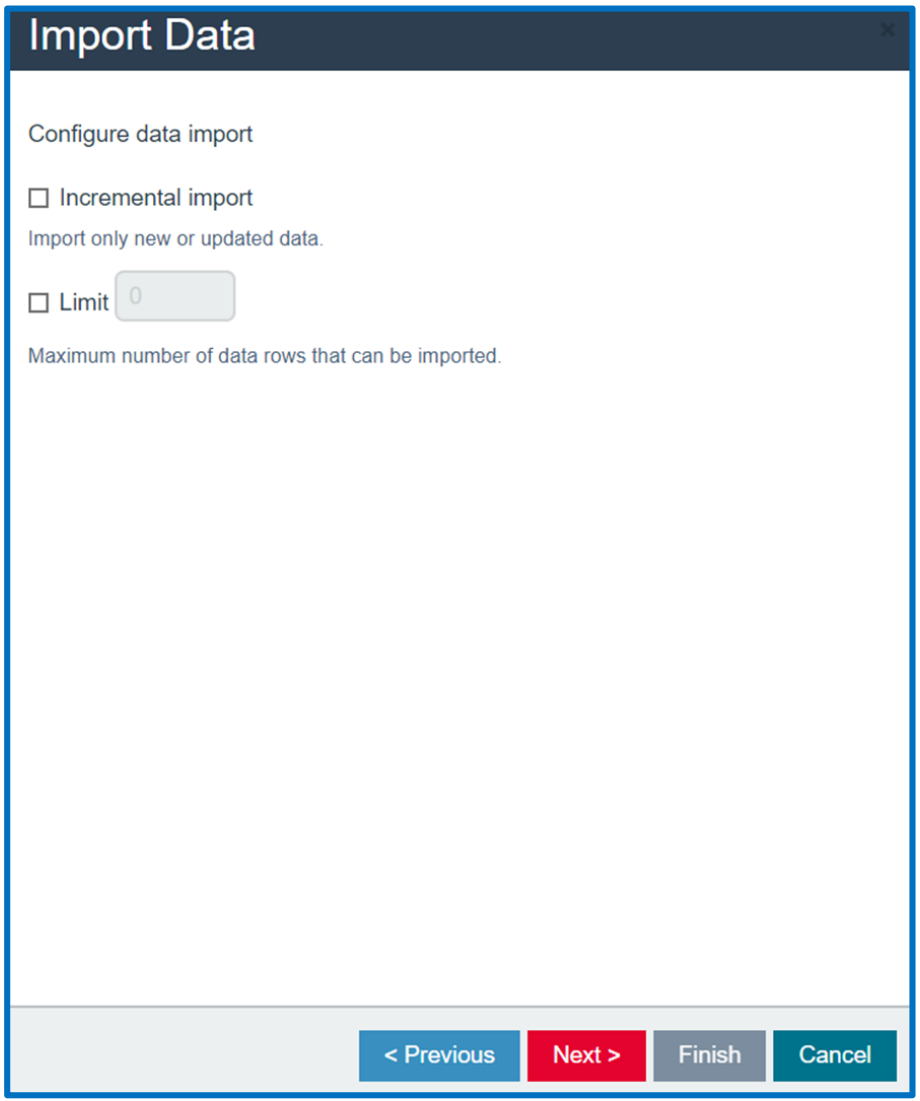


Click on „Advanced Options“ and fill in the Field Separator with your separator symbol from Step 03 (I told you to remember...). Click „Next>“

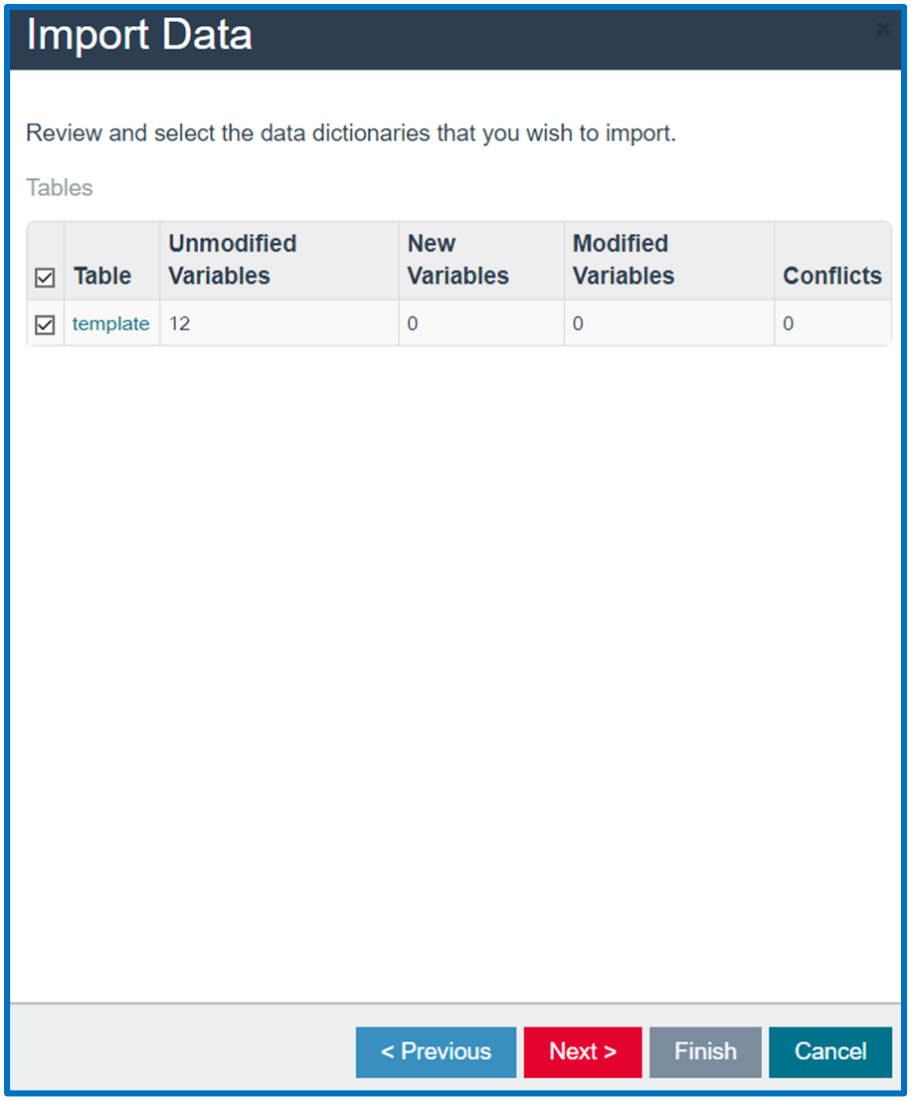


Step 06-6: Raw File

If this is a new Project and no incremental import, you can just Click on „Next>“

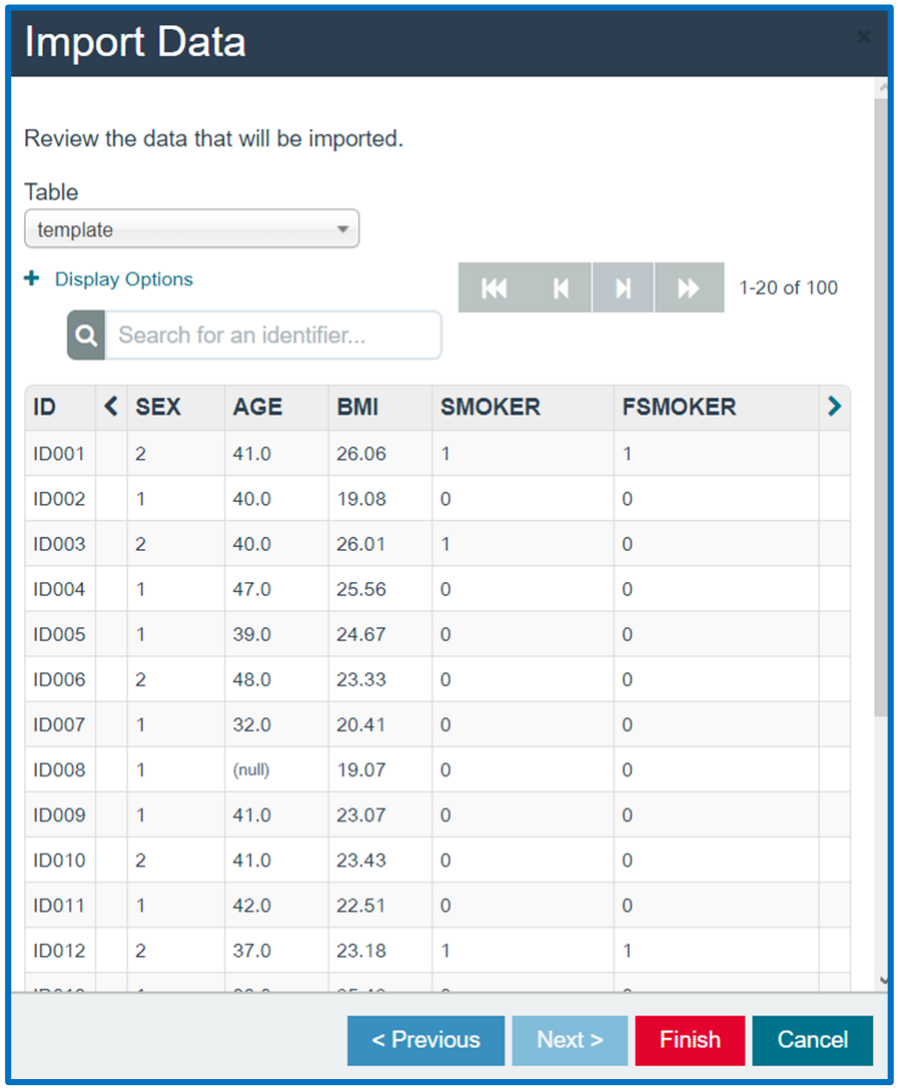


Your data dictionary should be activated, click on „Next>“



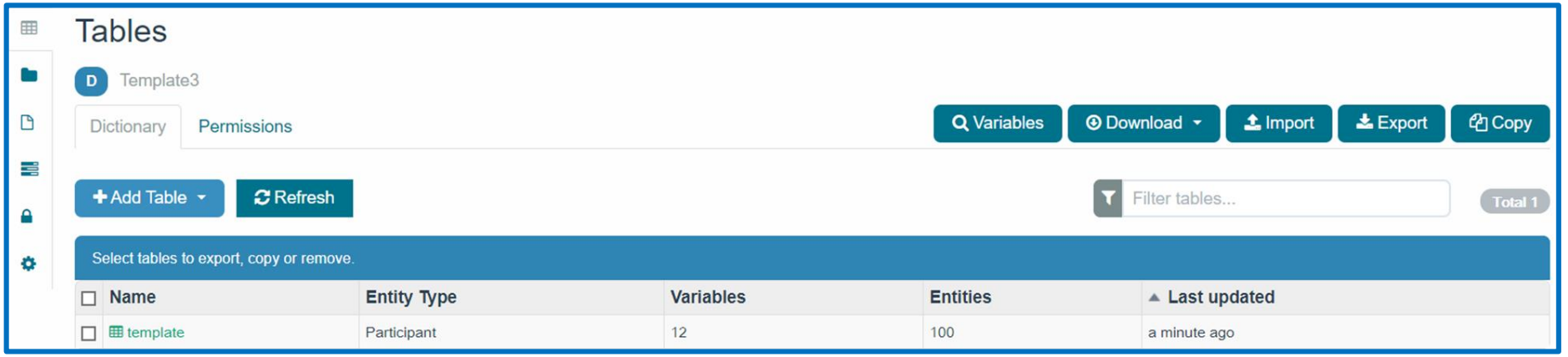
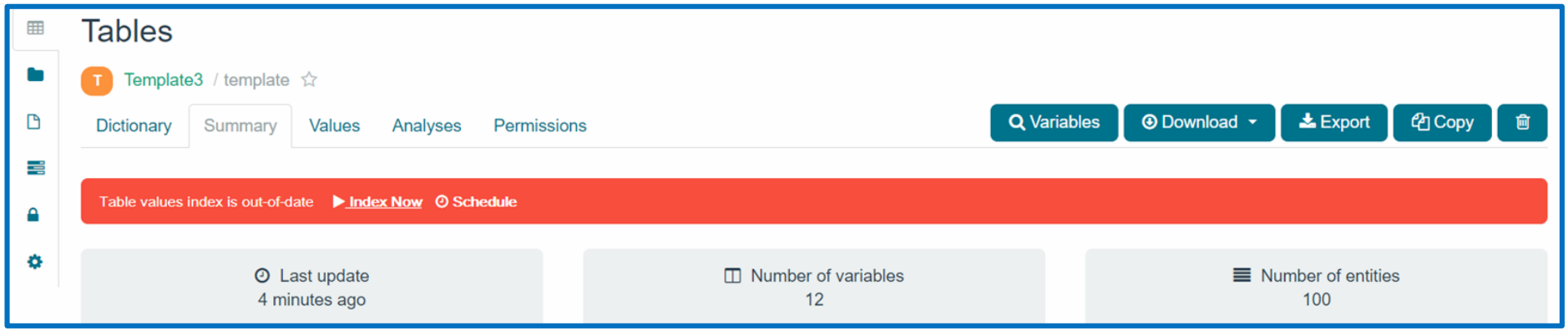
Step 06-7: Raw File

On the final Review page, you should see your raw data properly formatted and the correct dictionary table name. Note that formerly empty cells have been replaced by „(null)“. Click „Finish“



Step 06-8: Raw File

You should now see the correct number of „Variables“ (clinical parameters) and „Entities“ (participant ID‘s) in the Project view. If „Entities“ is zero, click on „Refresh“ a few times.



If that still doesn‘t work, the dictionary table needs to be re-indexed.
Click on your Project „Name“ (e.g. „template“) and on „Index now“, if this appears.