ORCHESTRA

Pseudonymization Tool

– User Manual (for LibreOffice and Excel)

Change history

Version	Date	Changes	Authors
1.0	04.08.2021	First Draft	Fabian Prasser
			Hammam Abu Attieh
2.0	25.05.2023	Finalization	Hammam Abu Attieh

Table of contents

1. Background

The ORCHESTRA Pseudonymization Tool (OPT) is a crucial application designed to pseudonymize personal data for subjects and samples. However, due to the varying legal frameworks across countries participating in the ORCHESTRA project, it was challenging to find a standard software solution that fits all locations and technical infrastructures of hospitals and labs. Purchasing and implementing such software for each location would have been time-consuming and resource-intensive.

To address this issue, the decision was made to develop a custom tool tailored to the heterogeneous technical and legal frameworks of the ORCHESTRA project's sites. This tool was urgently required to ensure the implementation of the agreed data protection approach within the consortium. It provides a practical and pragmatic solution considering the complexity of ORCHESTRA's activities and offers a high level of personal data protection.

2. Introduction

This user manual provides comprehensive guidance for working with the OPT. The OPT provides a user-friendly solution for pseudonymizing sensitive information, thus enabling effective data analysis and processing.

The OPT is specifically designed for use in cross-site environments, facilitating seamless collaboration and data exchange among different research locations. It is available in both Microsoft Excel and LibreOffice versions to cater to the heterogeneous (technical) needs of users.

This manual is applicable to both the Microsoft Excel and LibreOffice versions of the OPT. It aims to provide all users with a comprehensive introduction to the various features and capabilities of the tool. It is crucial for all individuals working with the tool to familiarize themselves with this manual to ensure secure and efficient processing of health data.

The OPT tool offers a wide range of functionalities that empower users to carry out pseudonymization of health data in their research projects effectively. From pseudonym generation to data management and biosample-label generation, the tool provides a user-friendly interface and powerful features to ensure data privacy and security.

We hope this manual provides you with all the necessary information to successfully utilize the OPT tool. Please read it carefully!

We wish you success in using the OPT your scientific research projects!

3. Installation and Configuration

3.1 Installation

The OPT does not require any installation. Instead, it can simply be stored in a secure location. Ideally, a shared network that all users have access to is recommended for storing the tool.
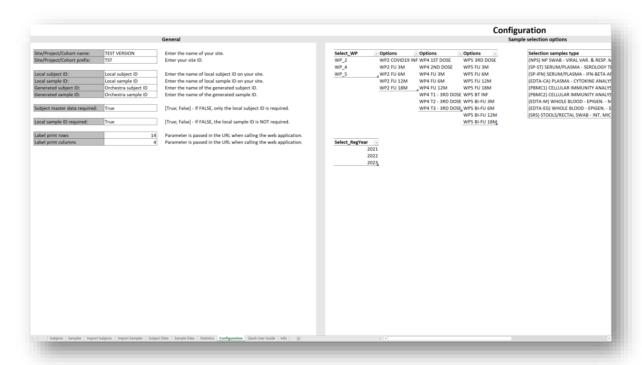
However, it is <u>strongly discouraged</u> to store the OPT tool in Excel Online, SharePoint or similar structures that can be accessed by multiple users simultaneously. Such environments often lead to problems and the risk of data loss. Therefore, it is strongly recommended to choose a secure location where only authorized users can access and work with the OPT tool sequentially.

Choosing a dedicated, shared network storage can ensure smoother operation and mitigate the risks associated with concurrent access. This approach helps maintain data integrity, minimizes potential conflicts, and protects against accidental data loss.

3.2 Configuration

The OPT allows comprehensive configuration options. The configuration must be carried out before the tool is used for the first time, as fundamental aspects are regulated there. The prefix in particular is crucial, as it is decisive for the generation of unique pseudonyms. If the OPT is to be used across locations, then the prefixes must differ from location to location, otherwise duplicates will occur! In addition to the prefixes, the field names and the selection options of the various dropdowns must also be adapted here. On the worksheet there are also the counters for the pseudonym generation. These should be set to 0 before the first use and never be changed by the users.

4. General information

Safety note: Despite the implemented security measures in OPT, it's important to note that a spreadsheet program does not provide high-level data security. Therefore, caution must be exercised when using OPT. Regular backups are strongly recommended to mitigate any potential data risks. Safeguarding your data through frequent backups ensures that you can restore it in the event of any unforeseen issues or data loss. Protecting the integrity and confidentiality of your data should also be a top priority.

Field input:

When using the LibreOffice version, please note that when you enter data into a field, the input is not recognized until you click out of the field after you have finished entering.

Date format:

When entering the date in a data field, the following formats can be entered depending on whether you use the Excel or LibreOffice version.

The allowed input types are shown in the table on the left.

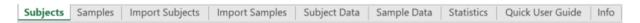| Date format	Excel	LibreOffice
dd.mm.yyyy	yes	yes
d.m.yyyy	yes	yes
d.m.yy	yes	yes

5. User interface and structure

In the following, first the structure of the tool and then the user interface will be explained.

The tool consists of ten individual worksheets. These are:

- Subjects
- Samples
- Import Subjects
- Import Samples
- Subject Data
- Sample Data
- Statistics
- Quick User Guide
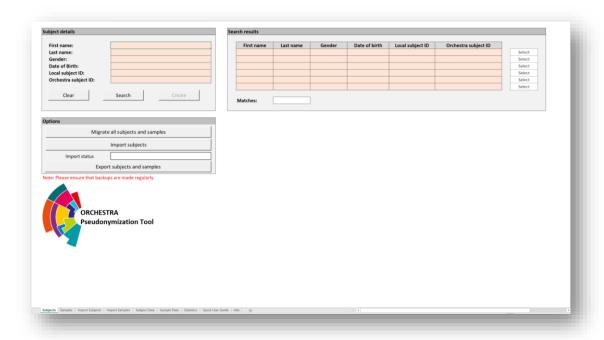- Info
- Configuration (*hidden*)

Use the Navigation bar at the bottom to see which worksheet you are on. Here it is also possible to navigate between the individual worksheets. When you open the tool, the last used worksheet is always displayed.

| Subjects | Samples | Import Subjects | Import Samples | Subject Data | Sample Data | Statistics | Quick User Guide | Info |

5.1 Subjects

The Subjects worksheet is structured into three windows. The first window "Subject details" contains a fillable form. This is where the search will take place later. The second window "Search results" shows a table, which is automatically filled by the tool during a search. In the third window "Options" there is a button to import subjects and another button to export all registrations for the subjects as well as for the samples in this tool in a new Excel/LibreOffice document.

5.2 Samples

The samples page is structured analogously to the subject page and has a very similar functionality. Only the buttons are slightly different. In the first window at the "Search details" there is another button, which serves for the Data-Matrix Code generation.

5.3 Import Subjects

The worksheet is an empty table. This will later be used to import existing data records into the tool. The column labels are similar to those on the "Subject Data" worksheet.
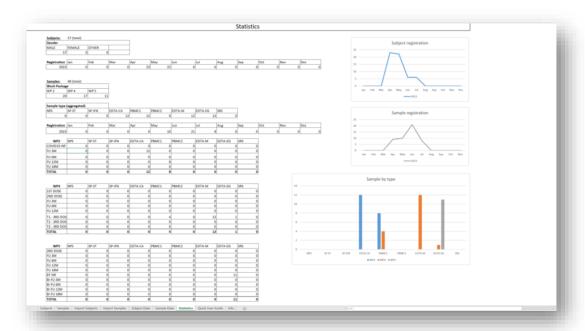
5.4 Subject Data and Sample Data

All registrations of subjects and samples are stored in these worksheets. In addition, the information that was entered in the Adding additional subjects or samples process is displayed here.

5.5 Statistics

On the Statistics Worksheet, the user gets a comprehensive overview of the data registered in the tool. Therefore, the data is broken down according to its time of registration, type and quantity.

5.6 Quick User guide

This document is a detailed description of the objectives, the structure and the functionalities of the tool. In addition, we have summarized the relevant core functionalities briefly and clearly in the "Quick guide" worksheet.

5.7 Info

If you need help or have any comments on how we can better customize the tool to your needs, please do not hesitate to contact us via the contacts on the "Info" worksheet. We appreciate any feedback!

You will also find the current version number of the tool you are working with as well as an update history.

6. Functions

In the following, the individual functionalities of the tool are explained in detail.
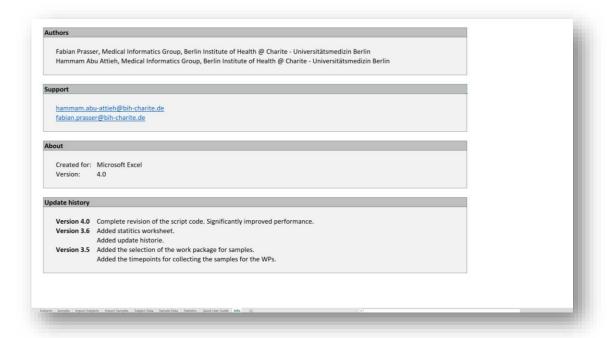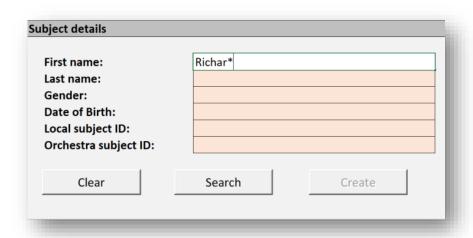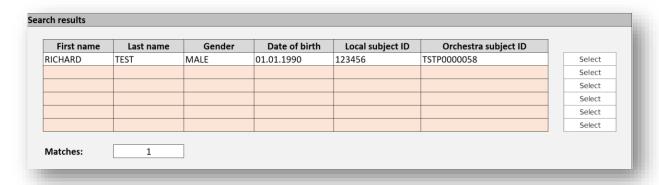
6.1 Search subject

Searching for subjects is one of the basic functionalities of the tool. This is a mandatory requirement for the registration of subjects. To perform a search, you must be on the "Subject" worksheet. In the first window "Search details" you will find a fillable form. Here you can insert the attributes of the subject you want to search for. You can search with one or more attributes. The use of wildcards (*) can be useful here. You can use them anywhere in the entered attribute.

Subject details

First name:	Richar*
Last name:	
Gender:	
Date of Birth:	
Local subject ID:	
Orchestra subject ID:	

[Clear] [Search] [Create]

For example, if you forget whether the person's name is *Richard* or *Richardt*: "Name: Richar*".

To search for a specific subject, it is useful to do this using the unique ORCHESTRA ID or local ID. With the other fields, such as the name or the gender, it cannot be guaranteed that the searched person can be found clearly.

After you have filled in the form, confirm the search by clicking on the "Search" button. Now the search results should be displayed in the "Search Results" window.

Search results

First name	Last name	Gender	Date of birth	Local subject ID	Orchestra subject ID	
RICHARD	TEST	MALE	01.01.1990	123456	TSTP0000058	Select
						Select
						Select
						Select
						Select
						Select

Matches:	1

6.2 Create subject

When a new subject is registered, a new unique ORCHESTRA pseudonymization ID will be created for them and their identification data will be stored. This creates a link between the subject and an ORCHESTRA-wide unique ID. Then the generated ID can be used in the context of other investigations

and evaluations in the ORCHESTRA project in a data protection compliant way, without anyone knowing the identity behind the ID.

Registering a new subject is done on the "Subjects" worksheet in three steps. Please do not wonder that the "Create" button is disabled. It will be enabled automatically after the *Step 2* of the following process.

Step 1: Fill out the form

In the first window "Search details" you will find a form where you must fill in all relevant information to identify a subject.

The individual fields are defined as follows:

- First name: The subject's first name.
- Last name: The subject's last name.
- Gender: The gender of the subject.
- Date of birth: The date of birth of the subject.
- Local subject ID (designation may differ): The subject's local ID, which is used in daily practice in the hospital or in the labs to identify the subject uniquely.
- ORCHESTRA ID: The field remains empty. The ID is generated automatically and completed after registration.

> *Important:* Please note that the OPT subject ID field must not be filled in. The ID is automatically generated and assigned to the subject when registering the subject as described above. Therefore, please leave this field blank.

Step 2: Search for the subject

Confirm the above input by clicking on the "Search" button. This will trigger the search. The search results are then displayed in the "Search Results" window. As this is a new registration of a subject, there must be no match, as each subject may only be registered once. Therefore, please make absolutely sure that this subject is not already registered.

By performing the search, the "Create" button will be enabled.

Step 3: Registration of the subject

After making sure that the subject has not been registered yet, you will be able to register the subject by clicking on "Create". The tool will prompt you in a pop-up window to check your input for correctness.

Please make sure that all information is entered correctly in the form. Also make sure that the spelling is correct.

Then confirm the message in the pop-up window. This completes the registration process. The automatically generated OPT subject ID should now be displayed in the form.

Now you can create another subject, clear the fields of the form by clicking "Clear" and search for another subject or just close the tool.

6.3 Import subjects

If you already have a data set with subjects for whom you need unique ORCHESTRA IDs, you can generate them automatically. The process of importing data consists of only two steps.

Step 1: Pasting the data

To import data into the tool, switch to the "Import Subjects" worksheet. There you will find a table with many columns. The columns with the colored headers are the default fields, which are also used in the normal registration of a subject. All other columns are additional and can be customized (add, delete, rename) by adjusting the "Subject Data" worksheet, since they are mirrored from there.

Copy the data you want to import into the corresponding column. Please ensure that the required fields are filled in for each subject.

The required fields are:

- First Name
- Last Name
- Date of Birth
- Local subject ID (e.g. Health insurance number, etc.)

Please note that it is not mandatory to specify the gender of a subject, but if an entry is made, it must correspond to the options "male", "female" or "other", otherwise it will not be recognized by the tool and import will not be possible.

Step 2: Perform the import

When you have pasted your data, switch back to the "Subjects" worksheet. Click on "Import Subjects" in the "Options" window. This will start the import.

After the import is complete, a pop-up window will appear showing how many subjects have been imported.

In the best case, all subjects are imported, and the process is completed.

If subjects were not imported, they will remain in the "Import Subjects" worksheet. If this is the case, check the following points:

1) A mandatory field has not been filled in
2) The Date of Birth does not correspond to a valid format.
3) A gender has been entered for a subject that does not match the expected value.
4) A subject with the same local ID is already registered in the tool.

6.4 Search samples

The search for samples is basically the same as the search for subjects. It is also a mandatory requirement for the registration of sample. To perform a search, you must be on the "Samples" worksheet. In the first window "Search details" you will find a fillable form (For the detailed description of the fields: See Create Sample). Here you can insert the attributes of the samples you want to search for. You can search with one or more attributes. The use of wildcards (*) can be useful here. You can use them anywhere in the entered attribute.

To search for a specific sample, it is useful to do this using the unique ORCHESTRA ID or local ID. With the other fields, such as the Sample-Type or the Timepoint, it cannot be guaranteed that the searched sample can be found clearly.

After you have filled in the form, confirm the search by clicking on the "Search" button. Now the search results should be displayed in the "Search Results" window.

Note: While searching for samples on a specific subject, you may be returned a larger number of search results, as a subject may have many samples.

6.5 Create samples

When a new sample is registered, a new unique OPT sample ID is generated for it. Since the OPT subject ID which was generated for the previously registered subject must be entered when registering the sample, a link between the subject and the sample is established and stored on a pseudonymized basis. Then the generated OPT sample ID can be used in the context of other studies and evaluations in the ORCHESTRA project in a data protection compliant way, without anyone knowing the subject to that the sample is assigned.

Registering a new sample is done on the "Samples" worksheet in three steps. Please do not wonder that the "Create" button is disabled. It will be enabled automatically after the *Step 2* of the following process.

Step 1: Fill out the form

In the first window "Search details" you will find a form where you must fill in all relevant information to identify the sample.

The individual fields are defined as follows:

- Local subject ID (if available; designation may differ): The subject's local ID, which is used in daily practice in the hospital or in the labs to identify the subject uniquely.
- OPT subject ID: Insert here the OPT subject ID of the subject belonging to the sample, which was generated during the registration of the subject. This is necessary to have a link to the subject from which the sample was taken.
- Local sample ID (if available): The sample's local ID, which is used in daily practice in the hospital or in the labs to identify the samples uniquely.
- Date: The date of taking the sample.
- Work package: By using the work packages field, a sample can be assigned e.g. to a work package in the research project.
- Sample type: Select one of the displayed options (options can be adjusted, see: configuration).
- Timepoint: Select one of the displayed options (options can be adjusted, see: configuration).
- OPT sample ID: The field remains empty. The ID is generated automatically and completed after registration.

Step 2: Search for the sample

Confirm the above input by clicking on the "Search" button. This will trigger the search. The search results are then displayed in the "Search Results" window. As this is a new registration of a sample, there must be no match, as each sample may only be registered once. Therefore, please make absolutely sure that this sample is not already registered.

By performing the search, the "Create" button will be enabled.

Step 3: Registration of the sample

After making sure that the sample has not been registered yet, you will be able to register it by clicking on "Create". The tool will prompt you in a pop-up window to check your input for correctness.

Please make sure that all information is entered correctly in the form. Also make sure that the spelling is correct.

Then confirm the message in the pop-up window. This completes the registration process. The automatically generated OPT sample ID should now be displayed in the form.

Now you can create another sample, clear the fields of the form by clicking "Clear" and search for another sample or just close the tool.

6.6 Import samples

The import of samples basically works in the same way as the import of subjects.

Step 1: Pasting the data

To import data into the tool, switch to the "Import Samples" worksheet. There you will find a table with many columns. The columns with the colored headers are the default fields, which are also used in the normal registration of a sample. All other columns are additional and can be customized (add, delete, rename) by adjusting the "Sample Data" worksheet, since they are mirrored from there.

Copy the data you want to import into the corresponding column. Please ensure that the required fields are filled in for each sample.

The required fields are:

- Local subject ID
- OPT subject ID
- Local sample ID
- Date
- Sample type*
- Timepoint

Please only enter the abbreviation here. You can find them in the Quick User Guide or in the Samples worksheet in the drop-down menu. E.g.: "NPS" for "NP swap.

Please note that the entered Locale Patient ID and OPT subject ID must already exist in this OPT. This will be checked by the OPT during import. If they do not exist, the import will fail.

Step 2: Perform the import

When you have pasted your data, switch back to the "Samples" worksheet. Click on "Import Samples" in the "Options" window. This will start the import.

After the import is complete, a pop-up window will appear showing how many samples have been imported.

In the best case, all samples are imported, and the process is completed.

If samples were not imported, they will remain in the "Import Samples" worksheet. If this is the case, check the following points:

5) A mandatory field has not been filled in
6) The Date does not correspond to a valid format.
7) A corresponding local subject ID or OPT subject ID does not exist.
8) A sample with the same local ID is already registered in the OPT.

6.7 Generate Labels

The OPT supports the generation of the following codes for the labels for the sample vials:

- QR-Code
- DataMatrix-Code
- Barcode

A label contains the following information:

- Code: The Code contains the OPT sample ID. This can be read by the Code scanners of the labs.
- Sample type: The sample type of the sample contained in the vial or the analysis that is performed on it.
- ORECHESTA subject: The OPT subject ID of the subject from whom the sample was taken.
- Timepoint: The time at which the sample was taken.

All this information is based on the entries made when the sample was registered.

There are two ways to create a label: **One-by-one** via the sample worksheet or **bulk printing** after import.

One-by-one:

The generation of a label with a Code is done on the "Samples" worksheet. According to the initial situation, there are two possibilities:

1) You have just registered a sample. The form in the "Search details" window is still filled completely and unchanged. The OPT sample ID has already been generated.

2) If the form in the "Search details" window is empty, you must run a search corresponding to Search sample by filling in the fields in the form and start the search by clicking on "Search". In the "Search Results" window you will see the search results. Select the sample for which you want to create a label by clicking on "Select" next to the corresponding entry. Now the form in the "Search details" window is automatically filled with all relevant information.

Then you only need to click the "Print" button. Your standard browser will open with the corresponding application, where our Code generator is running. Now you can print the generated labels directly.

Bulk printing:

After importing samples, the tool asks if you want to print the corresponding labels. If you confirm this, a separate Excel list with all the required information is generated and your default browser opens and displays the label generator application.

Now, just copy and paste the contents of the Excel into the application. Click on "Refresh Labels" and the labels will be generated. Now you can print the generated labels directly.

6.8 Adding additional subjects or samples information

Depending on the needs, additional information that is not covered by the fields in the forms on the Subject and Samples worksheets can be stored in the tool. For this purpose, the tool offers the option of adding individual columns on the "Subject Data" and "Sample Data" worksheets and adding the information on the corresponding subjects and samples.

The process is identical for adding information for subjects and samples.

Information is added as requested in the "Subject Data" or "Samples Data" worksheet. In the gray locked area are all registrations that have already been made. You can insert any number of attributes in the remaining open columns.

If you want to add information about a specific subject or a specific sample, you can at first perform a search based on the OPT subject ID or OPT sample ID by using the filter in the related column.

6.9 Export subject and samples (backups)

Exporting subject and samples data from the ORECHESTRA pseudonymization tool serves as a backup measure. All data entered during the registration of subjects and samples as well as the additional information in the individually created columns are exported to a separate file.

The export can be performed on both the "subjects" and "samples" worksheets by clicking on the "Export all subjects and samples" button.

A file will be generated and saved immediately next to the OPT in the Backup-file.

7. References

> _Important_: Please make sure that the export is stored separately from the tool in another place to have the data backed up in case of possible problems with the drive.

[1] ORCHESTRA [Online]. Available: https://orchestra-cohort.eu/