PIA Research Data Export

This ZIP file contains data exported from PIA for one study. Based on the chosen export configuration it may contain one or more of the following files:

- `answers/answers_*.csv` - answers given by participants (recommended format) (read more)
- `answers.csv` - answers given by participants (old format) (read more)
- `blood_samples.csv` - blood sample meta-data (read more)
- `codebook_*.csv` - meta-data of one questionnaire (read more)
- `files/*` - files uploaded by participants or study assistants as part of answering a questionnaire (read more)
- `lab_results.csv` - laboratory results available in PIA (read more)
- `questionnaire_settings_*.csv` - settings of the selected questionnaires (read more)
- `samples.csv` - sample meta-data of nasal swab samples (read more)
- `settings.csv` - participants' settings, e.g. test participant incl. external consents (read more)
- `README.md` - this file

Of note: The * above point to the fact, that PIA fills in a concrete content-related name automatically (see below).

Data records are exported in CSV format. CSV is a text-based format where each record is written on a separate row. Every row contains **semicolon-separated** values. The first row of each file contains the column names.

CSV files can be opened with any text editor or spreadsheet application which supports **UTF-8 encoding**, like LibreOffice Calc, Microsoft Excel or Apple Numbers.

However, **Excel on Windows might have difficulties if you open a CSV file**. Special characters might not be displayed correctly. If you encounter such problems, please refer to the "PIA Handbuch". Briefly, open an empty excel file -> data -> "Externe Daten abrufen" -> "Aus Text": The drop-down menu for "Dateiursprung" needs to display "65001 : Unicode (UTF-8)".

Answers (recommended format)

Answers are written into separate files for each questionnaire and each questionnaire version. Therefore, a questionnaire that has been changed 10 time, will produce 10 export files. The file name has the structure (square brackets enclose a variable value)

`answers_[questionnaire name]v[questionnaire version]_[questionnaire id]_[date of export]`

where the date of export is the current UTC date formatted as an ISO 8601 date with the format `YYYY-MM-DDThhmm` ("T" is a delimiter). For example, CEST being 7 am is exported as 0500.

The following generally applies to the content of one answers file:

- Each **file** contains the **answers of one questionnaire in one specific version**.
- Each **row** contains **all answers to one specific questionnaire instance**.
- A **questionnaire instance** is a single questionnaire which is, was or will be shown to the participant or study assistant. Its answers will always be released all at once or not at all. If the underlying questionnaire is a questionnaire that is displayed continuously (e.g. spontaneous questionnaire) or cyclic, each appearance of the questionnaire is represented by a new instance.
- As a result there might be **multiple rows for the same participant** in cyclic questionnaires.
- The file contains rows for every questionnaire instance. A file does not contain answers the participant or study assistant has entered but not yet released ("abschicken"). **Answers will only be exported if the questionnaire instance is already released otherwise it will be marked as a missing** (read more).
- The file contains **fixed columns** which appear in every answers export file and **dynamic answer columns** which appear based on the questionnaire structure.
- Generally **all columns** by definition of a CSV are **strings** (text). Those strings have to be converted to a more specific format (e.g. number, date, etc.) outside of PIA. However, the following conventions apply to the content of all columns:
  - **Boolean** values are represented as `T` (= true) and `F` (= false)
  - **Dates** are represented as ISO 8601 date with the format `YYYY-MM-DD`
  - **Dates with time** are represented as ISO 8601 date with the format `YYYY-MM-DDThh:mm:ss±hh:mm` **with ± giving the deviation from universal time coordinated**

(**UTC**) and "T" being a delimiter

Fixed Columns

- `participant` - pseudonym of the participant
- `is_test_participant` - `T` if the participant is a test participant, `F` otherwise
- `questionnaire_name` - name of the questionnaire the answers belong to
- `questionnaire_id` - id of the questionnaire the answers belong to
- `questionnaire_version` - version of the questionnaire the answers belong to (starting with 1)
- `questionnaire_cycle` - the number of the iteration of a cyclic questionnaire (starting with 1)
- `questionnaire_date_of_issue` - the date the questionnaire instance was issued to the participant (ISO 8601 date with time zone)
- `answer_date` - the date the questionnaire instance was last answered by the participant (ISO 8601 date with time zone)
- `answer_status` - the status of the questionnaire instance (read more)

Answer Status The answer status provides information about the progress of the processing of a questionnaire instance by participants. The following values are possible:

- `pending_participant_answer` - answers from the participant or study assistant are pending, i.e. there are no answers yet
    - export of questionnaire instance will only contain missings (read more)
- `in_progress_participant_answer` - answers were given by the participant or study assistant but not yet released ("abgeschickt")
    - export of questionnaire instance will only contain missings (read more)
- `modifiable_participant_answer` - answers were released by the participant but can still be modified once (PIA allows answers by study participants to be modified once.)
    - export contains released answers
- `final_participant_answer` - answers were released twice and cannot be modified anymore
    - export contains latest released answers
- `latest_study_assistant_answer` - last answer of a study assistant (study assistants can modify, i.e. release answers unlimited times)
    - export contains latest released answers
- `expired_answer` - the possibility to answer has expired before an answer was released
    - export of questionnaire instance will only contain missings (read more)

Dynamic Answer Columns

The fixed columns are followed by the dynamic answer columns. The dynamic answer columns are **structured based on the questions/items within the underlying questionnaire**. Each column contains the answer to one question. In case of multiple choice questions the file contains one column for every answer option. In case of answer option "sample" the file contains two columns - one for each sample id. In all other cases there is one column per question.

A **detailed description of all dynamic answer columns** including its question texts, answer code to answer text mappings, validations and conditions can be found in the questionnaire's **codebook** (read more).

Column Names The format of the **column names** of dynamic answer columns is as follows:

`[questionnaire id]_[first 3 letters of the questionnaire name]_[variable name]`,
e.g., `13_hea_today`

In case of **multiple choice questions** the variable name is followed by the answer option text:

`[questionnaire id]_[first 3 letters of the questionnaire name]_[variable name]`
`_[answer option text]`, e.g., `13_hea_var10_influenza`

In case of **sample questions** the variable name is followed by two sample IDs:

`[questionnaire id]_[first 3 letters of the questionnaire name]_[variable name]_ProbenID[1|2]`,

```
e.g., 24_nas_swab_ProbenID2
```

The **variable name** uniquely identifies a question or (if any) its answer options within a questionnaire, even across different questionnaire versions. It can be configured by researchers in the questionnaire editor. If no variable name is configured, a random label starting with auto___ (_ sequence of 8 random digits) will be generated when creating or updating the questionnaire.

If a questionnaire was created or last updated **before the introduction of automatically generated variable names** (PIA version 1.32.0, Dec 2022), the format of the column names is as follows (square brackets enclose a variable value):

```
[questionnaire id]_v[questionnaire version]_[answer position in questionnaire]
```

The same suffixes for multiple choice and sample questions apply as described above.

Answer Values   Whenever **predefined answer options** were provided for a question (e.g., yes, no, I don't know), the answers given to that questions will be exported as **coded values**(e.g., 1, 0, 2). The mapping between coded values and the actual answer values is configured in the questionnaire editor and displayed in the codebook (read more).

For **non-predefined answer options** the answer values will be exported **as string** with the following conventions:

- **Photo** and **file** question columns will contain the filename (read more)
- **Sample** will result in two columns, which contain the respective sample ID (read more)
- **Date** question columns will contain the date in the format `YYYY-MM-DD` (see dates)
- **Timestamp** question columns will contain the date in the format `YYYY-MM-DDThh:mm:ss±hh:mm` **with ± giving the deviation from universal time coordinated (UTC)** (see dates with time)

Missings   Whenever a question was answered and its questionnaire instance released, the answer will be exported. However, there are several cases where no answer will or can be exported. In those cases the answer will be marked as missing by special **missing codes**:

- `-9999` (`unobtainable`) - The questionnaire instance was released by the participant or study assistant but the answer to this item was not given (= real missing)
- `-8888` (`notapplicable`) - The question could not be answered because it was hidden due to conditions in the underlying question or questionnaire (= legitimate/qualified missing)
- `-7777` (`no_or_unobtainable`) - Only applicable to multiple-choice answers: The questionnaire instance was released by the participant or study assistant but the answer to this item not given OR answered with `false`
- `-6666` (`notreleased`) - The question might have been answered or might be empty but the questionnaire instance was never released ("abgeschickt")

Missings will always be exported as code, even if the corresponding question's answers would be exported as text otherwise.

Additionally, there are **two cases where questionnaire instances might have existed, but they do not appear** in the export:

1. The questionnaire was deactivated by the research team before the participant answered it
2. The participant revoked its consent to save and process his/her answers fully or partially

There will be no entry for such cases in the export; even no missing code.

Answers (old format)

The `answers.csv` file contains the answers given to the questions of **one or more questionnaires**, based on the chosen export configuration. The **usage of this format is not recommended anymore** as it misses information like the correct specification of missings or a corresponding codebook. This answers export format may be removed in a future version of PIA.

Blood Samples

The `blood_samples.csv` file contains the blood samples that were collected during the study. Each row represents one blood sample. The following columns are available:

- `Blutproben_ID` - the sample id
- `Proband` - the participant's PIA pseudonym
- `IDS` - the participant's IDS if available
- `Status` - takes the value `genommen` if the sample was carried out, `nicht genommen` otherwise
- `Bemerkung` - the remark given related to taking the sample

Codebook

The codebook is a **detailed description of all questions, their answer options, code mappings, validations and conditions within** a questionnaire (conditions on other questionnaires are given in the file `questionnaire_settings_*.csv`). It is a valuable source of information to understand and evaluate the contents of the answers export.

Codebooks are written into separate files for each questionnaire and questionnaire version. The file name has the structure:

`codebook_[study name]_[questionnaire name]_v[questionnaire version].csv`

The following generally applies to the content of one codebook file:

- Each **file** contains the description of one questionnaire in one specific version.
- Each **row** contains one of the following entry types:
  - a **question**
  - a question's **answer option** in case of multiple or single choice questions
  - a question's **introductory text**
- Generally **all columns** by definition of a CSV are **strings** (text). However, **boolean** values are always represented as T (= true) and F (= false)

Columns

The following columns are available in the codebook:

- `questionnaire_id` - the questionnaire's id
- `questionnaire_version` - the questionnaire's version
- `questionnaire_name` - the questionnaire's name
- `variable_name` - unique identifier for a question or (if any) its answer options (then attached automatically) (read more)
- `column_name` - column name under which a question's answers can be found in the answers export (read more)
- `answer_position` - display position of the question or answer option in the questionnaire
- `text_level_1` - a text preceding one or more questions
- `help_text_level_1` - a help text preceding one or more questions adding any information probably needed to answer the question(s) correctly
- `text_level_2` - a text displayed next to the questions input column
- `answer_option_text` - a text displayed next to the corresponding checkbox in case of multiple choice questions, empty otherwise
- `answer_type` - the type of the question's answer (read more)
- `answer_category` - the category of a possible answer (read more)
- `answer_category_code` - the code mapping of an answer category (read more)
- `valid_min` - the minimum value of a numeric answer
- `valid_max` - the maximum value of a numeric answer
- `answer_required` - T if the question is mandatory, F otherwise
- `condition_question` - T if a condition is configured for the question, i.e., if the question is only displayed given a certain condition applies, F otherwise
- `condition_question_type` - whether the condition originates from the same or another questionnaire
- `condition_question_questionnaire_id` - the condition target questionnaire's id

- `condition_question_questionnaire_version` - the condition target questionnaire's version
- `condition_question_column_name` - the condition target question's column name
- `condition_question_operand` - the comparison operand (`</>/<=/>=/==/\=`)
- `condition_question_answer_value` - the condition target answer value as text
- `condition_question_link` - the condition link (`AND/OR/XOR`)

Answer Types   Answer types are defined for each question and describe the type of the answer that can be given to the question.

The following answer types are available:

- **single choice** - multiple answer values are given, only one can be selected
- **multiple choice** - multiple answer values are given, zero to all can be selected
- **numeric** - only numbers are allowed as answer
- **text** - free text is allowed as answer
- **date** - a date can be selected from a date picker
- **sample** - allows sample id 1 and sample id 2 to be entered
- **pzn** - allows a Pharmazentralnummer (PZN) to be entered
- **image** - PNG and JPEG files can be uploaded
- **timestamp** - allows to set the current timestamp as answer
- **file** - PDF or CSV files can be uploaded (only available for study assistants)

Answer Categories   Answer categories describe the **possible answer values related to a question**. Answer categories are exported as human-readable text. However, the answers export contains only the answer category codes. These are code mappings of the answer categories to allow the data to be more machine-readable.

The **answer category** may contain

- the **text of one of the predefined answers** in case of **single choice** questions,
- `yes` or `no` in case of **multiple choice** questions,
- may be **empty if answers are not predefined**
- or may contain one of the four **missing values** (read more)

The **answer category code** may contain

- the **code of one of the predefined answers** in case of **single choice** questions,
- `1` or `0` in case of **multiple choice** questions,
- may be **empty if answers are not predefined**
- or may contain one of the four **missing codes** (read more)

Files

Files can be uploaded by participants and study assistants to answer questions of type **photo** or **file**. The files are exported into the `files` folder and are referenced in the answers export.

Please keep in mind, that those files are uploaded by participants and study assistants and are not checked for viruses by PIA. It is the **responsibility of the researcher to check the files for viruses before using them**.

However, PIA will check the file size and file format of uploaded files and will not allow files that are too large or are not of a supported format. The supported formats are:

- **JPEG images** (allowed for participants and study assistants)
- **PNG images** (allowed for participants and study assistants)
- **PDF documents** (allowed for study assistants only)
- **CSV documents** (allowed for study assistants only)

Lab Results

The `lab_results.csv` file contains the laboratory result observations that were collected during the study. Each row represents one laboratory result observation. The following columns are available:

- `Bericht_ID` - sample id
- `Proband` - participant's PIA pseudonym
- `IDS` - participant's IDS
- `Datum_Abnahme` - date of sampling
- `Datum_Eingang` - date of delivery
- `Datum_Analyse` - date of analysis
- `PCR` - name of the pathogen analysed with PCR
- `PCR_ID` - organizational ID of analysis
- `Ergebnis` - result of the laboratory analysis as text
- `CT-Wert` - result of the laboratory analysis as number, here: cycle-threshold value
- `Auftragsnr` - order id
- `Arzt` - performing doctor
- `Kommentar` - comment given to the observation

Samples

The `samples.csv` file contains meta-data of the nasal swamp samples that were collected during the study. Each row represents one sample. The following columns are available:

- `Proben_ID` - sample id 1
- `Bakt_Proben_ID` - sample id 2 (RNAlater)
- `Proband` - participant's PIA pseudonym
- `IDS` - participant's IDS
- `Status` - information about the status of the sample (read more)
- `Bemerkung` - comment given to the sample

Sample Status

The sample status can take the following values:

- `neu` - the sample ID has been read into PIA but not yet taken
- `genommen` - the sampleing has been carried out
- `analysiert` - the sample has already been analyzed in the laboratory
- `gelöscht` - the sample was deleted

Questionnaire Settings

The `questionnaire_settings_*.csv` file contains the general settings of all questionnaires which were selected for the export. Each row represents one questionnaire. The following columns are available:

- `questionnaire_name` - the questionnaire's name
- `questionnaire_id` - the questionnaire's id
- `questionnaire_version` -the questionnaire's version
- `questionnaire_version_start` - the date the questionnaire has been created
- `questionnaire_version_end` - the date the questionnaire has been replaced by a newer version if a newer one exists
- `questionnaire_type` - `for_probands` if questionnaire is to be filled out by participants, `for_research_team` if questionnaire is to be filled out by study assistants
- `cycle_unit` - the unit of the cycle the questionnaire is being released in
- `cycle_amount` - how often the questionnaire is being released in the cycle
- `cycle_per_day` - how often the questionnaire is being released per day
- `cycle_first_at` - the time the questionnaire is being released for the first time
- `activate_at_date` - the date questionnaire instances are being displayed to participants or study assistants
- `activate_after_days` - days after which a questionnaire instance is being displayed to the participant or study assistant. Days from the participant's first registration or the creation of the questionnaire or the fulfilled condition (the later date is valid)
- `deactivate_after_days` - days after which a questionnaire instance cannot be filled out anymore. Days from the participant's first registration or the creation of the questionnaire (the later date applies, only for perpetual questionnaires).

- `expires_after_days` - days from activation of the questionnaire instance to expiration of the FB instance (only if the questionnaire instance has not been released, not relevant for spontaneous questionnaires).
- `non_modifiable_after_days` - days from the first release of the questionnaire instance to the automatic final release after which answer cannot be modified anymore
- `notification_tries` - how often questionnaire instance reminders will be sent out to participants
- `notification_title` - the title of the questionnaire instance reminder
- `notification_body_new` - the body of the questionnaire instance reminder for new questionnaire instances
- `notification_body_in_progress` - the body of the questionnaire instance reminder for questionnaire instances that have not been filled out yet
- `compliance_samples_needed` - `T` if the questionnaire should only be displayed to participants who complied to take samples, `F` otherwise
- `visibility` - `hidden` if the questionnaire should not be displayed to participants, `testprobands` if the questionnaire should only be displayed to test participants, `allaudiences` if the questionnaire should be displayed to all participants of the study
- `despite_end_signal` - `T` if answers to this questionnaire should be preserved after the final signal of SORMAS (only relevant if PIA is configured to connect to a SORMAS system), `F` otherwise
- `deactivated` - `T` if the questionnaire is deactivated and cannot be answered or edited, `F` otherwise
- `deactivated_at` - the date the questionnaire was deactivated, if applicable
- `condition_questionnaire` - `T` if the questionnaire is only displayed based on a condition to another questionnaire, `F` if no condition is set
- `condition_questionnaire_name` - the name of the questionnaire to which the condition relates
- `condition_questionnaire_id` - the id of the questionnaire to which the condition relates
- `condition_questionnaire_version` - the version of the questionnaire to which the condition relates
- `condition_questionnaire_question_id` - the id of the question to which the condition relates (format: `q_[question id][answer option id]`)
- `condition_questionnaire_question_column_name` - the column name of the question to which the condition relates and to which the question can be found in the codebook (read more) and answers export (read more)
- `condition_questionnaire_question_operand` - the comparison operand (`</>/<=/>=/==/\=`)
- `condition_questionnaire_question_answer_value` - the condition target answer value as text
- `condition_questionnaire_question_link` - the condition link (`AND/OR/XOR`)

Questionnaire Cycles

Questionnaires can be configured to appear based on specific rules. These rules are called cycles. The following cycles are available:

- `once` - the questionnaire appears only once based on the `activate_after_days` setting
- `day` - the questionnaire appears every x days, where x is equal to the value of `cycle_amount`
- `week` - the questionnaire appears every x weeks, where x is equal to the value of `cycle_amount`
- `month` - the questionnaire appears every x months, where x is equal to the value of `cycle_amount`
- `hour` - the questionnaire appears every x hours starting at time t, where x is equal to the value of `cycle_per_day` and t is equal to `cycle_first_at`
- `spontan` - the questionnaire is always visible. After one questionnaire instance was released, a new one appears immediately
- `date` - the questionnaire appears only once at a specific date based on the `activate_at_date` setting

Settings

The `settings.csv` file contains the participant's settings and external consents. Each row represents the data of one participant. The following columns are available:

- `Proband` - the participant's pseudonym
- `IDS` - the participant's IDS
- `Testproband` - `Ja` if the participant is a test participant, `Nein` otherwise
- `Einwilligung Ergebnismitteilung` - `Ja` if the participant externally agreed to receive the results of lab observations, `Nein` otherwise (read more)
- `Einwilligung Probenentnahme` - `Ja` if the participant externally agreed to take nasal swabs, `Nein` otherwise (read more)

- `Einwilligung Blutprobenentnahme` - `Ja` if the participant externally agreed to give blood samples, `Nein` otherwise ([read more](#))

External Consents

PIA supports two types of consents: **internal** and **external** consents. Internal consents are fully managed and obtained inside of PIA. External consents are consents that are obtained externally, but its results can be imported into PIA. They control the visibility of certain features and questionnaires for the participants.

Currently only the **export of external consents is supported**.