Sample information form specification

This specification describes sample information to be used with metabolomics samples.

This information should be saved in **.xlsx-file** without any formatting.

Required column

Column name	Description
SAMPLE_ID	Identification for the sample.

Common columns

Column name	Description
SUBJECT_ID	Identification for the study subject (e.g. patient number).
TIME	If samples are collected from multiple timepoints (e.g. intervention study), identifier for the time point. Examples: "1,2,3,4", "BASELINE, END", "A,B", "0MIN, 30MIN, 60MIN"
GROUP	If samples are from multiple experimental groups, group identifier.

Other columns

Besides these columns, there can be any number of additional columns with more phenotype information (e.g. columns for age, gender, date).

The column names should follow these rules:

- Only alphanumeric characters (A-z, 0-9) and underscore "_" to denote space
- No umlauts/other special characters (e.g. äöå)
- Start column names with a letter instead of a number

Examples: AGE, sex, DATE_OF_BIRTH, fP_GLUK_OVER_5_5

Description of the additional variables may be added as a separate sheet in the .xlsx file. This sheet should have two columns:

Column name	Description
VARIABLE	The name of the variable e.g. CHOLESTEROL
DESCRIPTION	Description of the variable e.g. Cholesterol level measured 2 hours after last meal.

Example files

SAMPLE_ID	ORDER	GROUP	SUBJECT_ID	TIME	AGE	SEX	SMOKING
S011	1	Control	1	A	45	M	5
S012	2	Treatment	2	A	52	M	10
S045	3	QC					4
S013	4	Control	3	A	40	F	0
S034	5	Control	3	B	41	F	0
S042	6	Treatment	4	A	38	M	7

VARIABLE	DESCRIPTION
AGE	Age in the beginning of the study.
SMOKING	Number of cigarettes smoked per day in the beginning of the study.