



Guidelines to Manipulate Clinical Data

Contact:

rlink@cea.fr

Authors:

celine.bourdon@inserm.fr

dimitri.papadopoulos@cea.fr

edouard.duches@cea.fr

Obtain data from central database

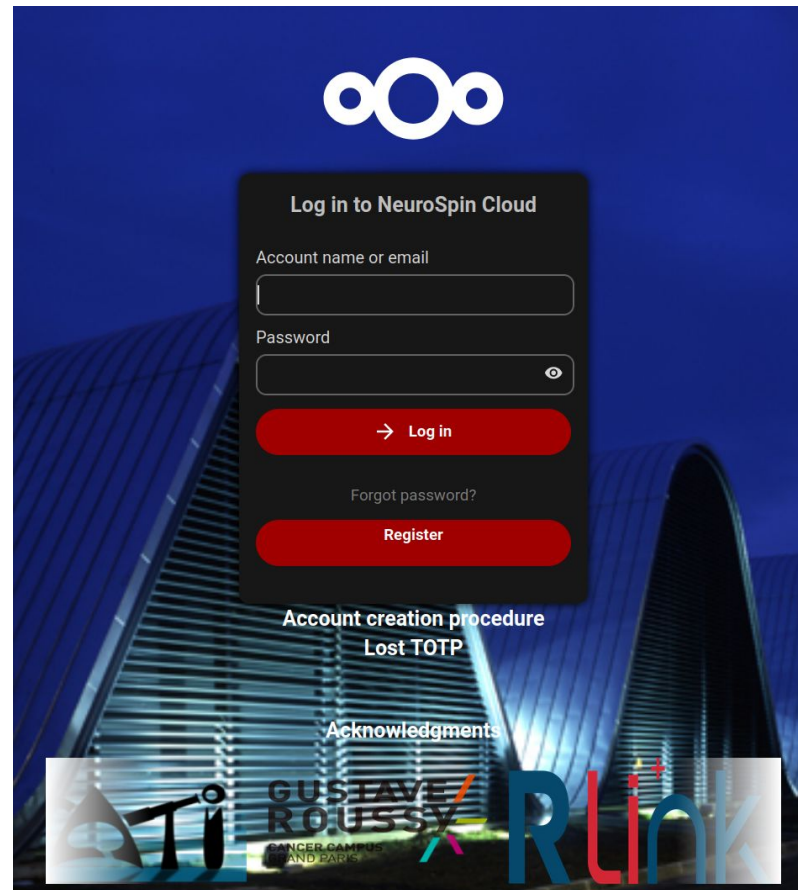
Contacts

1. mlink@cea.fr
2. dimitri.papadopoulos@cea.fr



Procedure

1. Go to <https://neurospin-cloud.cea.fr>
2. Request account creation
3. Download data



Clinical data = set of “tsv”* text files

*tsv = Tabular Separated Values. Similar to csv. Open with Excel, R or Python
See Appendix for details

File	Key ID	Content, of the file, Results of the aggregate
dataset-clinical_mod-baseline_version-2.tsv	participant_id	participant consent and baseline.; 1 line per subject
dataset-clinical_mod-baseline_form-postLi_tab-antidepress_version-2.tsv	participant_id	Antidepressants taken during the last 2 years; A varying number of lines per subject corresponding to antidep taken
dataset-clinical_mod-visits_form-visit_b-csrimeet_version-2.tsv	participant_id + FORM_F_VISIT	Clinical Service use Inventory for any meeting relevant to mental or physical health; A varying number of lines per subject corresponding to meeting relevant to mental or physical health on the visit number “FORM_F_VISIT”*

File name

Keys for this file:
participant_id + FORM_F_VISIT:
One line per participant and per visit

Keys for this file: participant_id
One line per participant

Description

Use case 1: Extract demographic, smoking status, Lithemia, Treatment Adhesion and response

Clinical variables

- Age
- Gender
- Smoking status
- BMI
- Lithemia at M03
- Treatment adhesion (MARS score) before M00 Lithium (M00) and after (M03)
- Depression and Mania at M00 (QIDS and BRMS)

Outcome

- Response.Status.at.end.of.follow.up

Procedure

1. Identify and list the desired variables in:
 - **Dataset-clinical_version-2_description.pdf**
 - **Dataset-clinical_version-2_description_file-information.xlsx**, sheet “**Data Handling Manual RLINK**”
2. Identify the file containing the variables
 - **Dataset-clinical_version-2_description_file-information.xlsx**, sheet “**FileInformation**”
3. Gather all variables and corresponding file into a table
4. Load files into tables and select variables
5. Merge tables into one table using “**participant_id**” and Save to excel or csv

CRF annoté

Protocole : [RLINK](#) Code centre : _____ Patient : _____

CRF annoté

Patient : _____

Protocole : [RLINK](#)
Version eCRF : V.266

1. Identify and list the desired variables in file:
- **dataset-clinical_version-2_description.pdf**

Plasma and erythrocyte Li on the day of Li-MRI 14769S4FC1

ID14769S4V1 / DATBIOMRI

Date (biology at MRI) : ____/____/____

ID14769S4V2 / TIMEBSIMRI

Time of blood sample intake : ____:____ (hh:mm)

ID14769S4V3 / BIOMRIYN

Performed : ☐ [1] Yes ☐ [0] No

ID14769S4V4 / **PLIMRI**

Plasma or Serum **Lithium** : |_|_|,|_|_|_| mEq/L = mmol/L

ID14769S4V5 / **ERYLMRI**

Erythrocyte **Lithium** : |_|_|,|_|_|_| mEq/L = mmol/L

Lithemia at M03

- PLIMRI
- ERYLMIR

2. Identify the file containing the variables using:

- file “dataset-clinical_version-2_description_file-information.xlsx”
- sheet “FileInformation”

1262	participant_id	dataset-clinical_mod-visits_form-visit_version-2.tsv	2449	participant_id + FORM_F_VISIT	Medical appointment information: A varying number c
1263	MODULE_M_VISITS	dataset-clinical_mod-visits_form-visit_version-2.tsv	2449		
1264	FORM_F_VISIT	dataset-clinical_mod-visits_form-visit_version-2.tsv	2449		
1265	DATBIOMRI	dataset-clinical_mod-visits_form-visit_version-2.tsv	2449		
1266	BIOMRIYN	dataset-clinical_mod-visits_form-visit_version-2.tsv	2449		
1267	PLIMRI	dataset-clinical_mod-visits_form-visit_version-2.tsv	2449		

1) Variable
PLIMRI

Optional
Keys for this file:
participant_id + FORM_F_VISIT
One line per participant and per visit

2) is in file
dataset-clinical_mod-visits_form-visit_version-2.tsv

Optional
Information about the file:
Medical appointment information: A varying
number of lines per subject corresponding to
hospital admission on the visit number
“FORM_F_VISIT”*

3. Gather all variables and corresponding file into a table

Variable	eCRF	File	
Age	AGE	dataset-clinical_mod-inclusion_version-2.tsv	inclusion
Gender	SEX	dataset-clinical_mod-inclusion_version-2.tsv	
Smoking status	WHO1A_PLI	dataset-clinical_mod-baseline_version-2.tsv	Baseline clinique
	WHO1A2A	dataset-clinical_mod-visits_form-visit_version-2.tsv	
BMI	HEIGHT_PRELI	dataset-clinical_mod-baseline_version-2.tsv	
	WEIGHT_PRELI	dataset-clinical_mod-baseline_version-2.tsv	Clinic at visits
	Formula		
Lithemia at M03 (FORM_F_VISIT = F_VISIT_1)	PLIMRI	dataset-clinical_mod-visits_form-visit_version-2.tsv	
	ERYLIMRI	dataset-clinical_mod-visits_form-visit_version-2.tsv	
	PLI	dataset-clinical_mod-visits_form-visit_version-2.tsv	
	ERYLI	dataset-clinical_mod-visits_form-visit_version-2.tsv	
	PLI2	dataset-clinical_mod-visits_form-visit_version-2.tsv	
MARS Score at M03 (FORM_F_VISIT = F_VISIT_1))	MARS1V	dataset-clinical_mod-visits_form-visit_version-2.tsv	Outcome
	
	MARS10V	dataset-clinical_mod-visits_form-visit_version-2.tsv	
	MARS1_PRELI	dataset-clinical_mod-baseline_version-2.tsv	
	...		
M00	MARS10_PRELI	dataset-clinical_mod-baseline_version-2.tsv	Outcome
	QIDSTSC_PRELI	dataset-clinical_mod-baseline_version-2.tsv	
QIDS and BRMS at M00	BRMSTSC_PRELI	dataset-clinical_mod-baseline_version-2.tsv	
Response status	Response.Status.at.end.of.follow.up	dataset-clinical_version-2_outcome.tsv	Outcome

Variables are found in **five** different files

4. Load Excel file and select variables

Read Inclusion data file “**dataset-clinical_mod-inclusion_version-2.tsv**”

- one line per participant
- Select AGE and SEX

R code

```
vars <- c("participant_id", "AGE", "SEX")
df1 <- read.csv(paste0(INPUT, " dataset-clinical_mod-inclusion_version-2.tsv"),
               sep = "\t", header = TRUE, na.strings = "ND")
df1 <- df1[, vars]
```

Python code

```
vars = ["participant_id", "AGE", "SEX"]
df = pd.read_csv(INPUT + " dataset-clinical_mod-inclusion_version-2.tsv",
                 sep="\t", na_values=['ND'])
df1 = df[vars]
```

4. Load Excel file and select variables

Read clinic at visits: “**dataset-clinical_mod-visits_form-visit_version-2.tsv**”

- one line per participant per time point
- Select (Smoking, Lithemia & MARS baseline) at month 3 (**F_VISIT_1**)

R code

```
vars <- c("participant_id", "FORM_F_VISIT", "WHOA2A", "PLIMRI", "ERYLIMRI", "PLI",  
         "ERYLI", "PLI2", "MARS1V", "MARS2V", "MARS3V", "MARS4V", "MARS5V",  
         "MARS6V", "MARS7V", "MARS8V", "MARS9V", "MARS10V")  
df3 <- read.csv(paste0(INPUT, "dataset-clinical_mod-visits_form-visit_version-2.tsv"),  
               sep = "\t", header = TRUE, na.strings = "ND")  
df3 <- df3[, vars]  
df3 <- df3[df3$FORM_F_VISIT == "F_VISIT_1", ]
```

Python code

```
vars = ["participant_id", "FORM_F_VISIT", "WHOA2A", "PLIMRI", "ERYLIMRI", "PLI",  
       "ERYLI", "PLI2", "MARS1V", "MARS2V", "MARS3V", "MARS4V", "MARS5V",  
       "MARS6V", "MARS7V", "MARS8V", "MARS9V", "MARS10V"]  
df = pd.read_csv(INPUT + "dataset-clinical_mod-visits_form-visit_version-2.tsv",  
                 sep="\t", na_values=["ND"])  
df3 = df[vars]  
df3 = df3[df3.FORM_F_VISIT == "F_VISIT_1"]
```

Merge all tables in a single large table

Use `participant_id` as key to align (merge) tables into a single **table**
Save to excel file

R code

```
table <- merge(merge(merge(df1, df2), df3), df4)  
write.xlsx(table, file.path(OUTPUT, "data_demoSmokLiMarsResponse_r.xlsx"))
```

Note

`Participant_id`
can be omitted:
Python & R
automatically
detect that it is the
only column
present in all
tables, and uses it
to align tables

Python code

```
table = pd.merge(pd.merge(pd.merge(df1, df2), df3), df4)  
table.to_excel(OUTPUT + "data_demoSmokLiMarsResponse_python.xlsx", index=False)
```

Output

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	participant_id	AGE	SEX	WHOAI1_PLI	HEIGHT_PRELI	WEIGHT_PRELI	MARS1_PRELI	MARS2_PRELI	MARS3_PRELI	MARS4_PRELI	MARS5_PRELI	MARS6_PRELI	MARS7_PRELI	MARS8_PRELI	MARS9_PRELI	MARS10_PRELI	OIDSTC_PRELI	BRMSTSC_PRELI	BMI
2	sub-10289	59	1	0	175	88.5	1	1	0	0	1	1	0	1	0	0	4	10	28,8979591836735
3	sub-11327	18	2	0	165	50	1	0	0	1	0	1	1	1	1	1	9	12	18,3654729109275
4	sub-11841	41	1	1	170	70	0	0	0	0	0	0	0	0	1	1	0	15	24,2214532871972
5	sub-14116	28	1	1	178	85	1	0	0	0	0	0	1	1	0	0	6	0	26,8274207802045
6	sub-14393	64	2	1	164	55	1	0	0	0	0	0	1	0	0	1	20	0	20,4491374182035
7	sub-16007	47	2	1	177	85	1	1	1	1	1	1	0	0	0	0	9	13	27,1314117909924
8	sub-16572	51	1	0	187	108	1	1	0	0	0	0	1	0	0	1	16	5	30,8844976979611
9	sub-17468	59	1	0	173	67	0	0	0	0	0	0	0	0	1	0	6	0	22,3863142771225
10	sub-17909	40	1	1	185	86.6	0	0	0	0	0	0	0	0	1	0	18	1	25,3031409788167
11	sub-19676	40	1	1	73.6	0	0	0	0	0	0	0	0	0	0	1	10	0	
12	sub-19987	26	1	1	180	73	1	1	1	1	0	0	1	1	1	1	3	10	22,5308641975309
13	sub-20807	27	2	1	173	70	0	0	0	1	0	1	0	1	1	1	1	15	23,3886865581877
14	sub-22378	52	1	1	62.1	0	0	0	0	0	0	0	0	0	0	0	5	5	
15	sub-22549	62	2	0	163	97	1	0	0	0	0	1	0	1	0	1	9	1	36,5087131619557
16	sub-23305	57	2	0	170	62.5	1	1	0	0	0	1	0	1	0	1	10	0	21,6262975778547
17	sub-23989	26	2	1	155	62	1	1	0	0	0	1	1	1	1	1	3	5	25,8064516129032
18	sub-25322	56	1	1	190	109	1	0	0	0	0	1	1	1	0	0	3	16	30,1939058171745
19	sub-25880	47	2	1	164	79.1	0	0	0	0	0	0	0	1	0	1	1	0	29,4095776323617
20	sub-27002	43	2	1	170	56	0	0	0	0	0	1	0	0	1	1	8	5	19,3771626297578
21	sub-27367	40	2	1	164	62	1	0	1	1	1	1	0	1	1	1	2	9	23,051754907793
22	sub-27813	54	1	1	183	111.5	1	1	0	0	0	1	0	1	1	1	17	5	33,2945146167398
23	sub-28561	46	2	1	160	53.8	0	0	0	0	1	1	1	0	0	0	0	19	21,015625
24	sub-28825	25	1	1	187	68	1	1	0	1	0	1	1	1	1	0	0	17	19,4457948468644
25	sub-29041	48	1	0	180	103	0	0	0	0	0	0	0	1	0	0	1	0	31,7901234567901
26	sub-29119	49	1	1	173	81	0	0	0	0	0	0	1	1	0	0	9	0	27,0640515887601
27	sub-29658	31	2	1	175	57	0	0	0	0	0	1	0	1	0	0	2	2	18,6122448979592
28	sub-30670	26	1	1	180	69	1	1	0	0	0	0	0	0	1	1	4	1	21,2962962962963
29	sub-30916	51	1	1	176	56	0	0	0	0	0	0	1	1	0	1	3	27	18,0785123966942
30	sub-32220	48	2	0	165	60	0	1	1	0	1	1	0	0	0	0	5	23	22,038567493113
31	sub-32363	54	2	1	171	81	0	0	0	0	0	0	1	1	0	0	12	2	27,7008310249308
32	sub-33001	69	2	1	163	63.9	0	0	0	0	0	0	1	1	0	0	8	1	24,0505852685461
33	sub-33139	36	1	1	182	91	0	0	0	0	0	0	1	0	1	1	7	2	27,4725274725275
34	sub-35218	32	1	1	0	0	0	0	0	0	0	0	1	1	0	1	1	7	
35	sub-36581	52	2	0	182	117.2	1	0	0	0	0	0	1	1	0	0	23	8	35,3822002173651
36	sub-36632	54	2	1	167	88.2	0	0	0	0	0	0	1	1	0	0	0	2	31,6253720104701
37	sub-36977	33	1	1	176	70	0	1	0	0	1	1	1	0	1	1	2	25	22,5981404958678
38	sub-37822	41	2	0	170	83.2	1	1	1	0	0	0	1	1	0	0	14	15	28,7889273356401
39	sub-37985	52	2	1	160	58	0	1	0	1	0	0	1	1	0	0	12	3	22,656625
40	sub-38348	51	1	0	76	1	1	0	0	0	0	0	0	1	0	0	10	0	
41	sub-38687	26	1	1	175	72	0	0	0	0	0	0	1	1	1	1	3	13	23,5102040816327
42	sub-39023	43	2	0	163	65	0	1	1	1	0	1	1	1	1	1	8	21	24,4646016033724
43	sub-40078	53	1	1	185	98.5	1	0	0	0	0	1	1	1	0	0	15	3	28,7801314828342
44	sub-40567	37	1	1	183	99.6	1	1	1	0	0	0	1	1	1	1	14	0	29,74110920208725
45	sub-40899	55	1	1	192	90											5	5	24,4140625
46	sub-40905	37	2	0	161	56	1	0	0	0	0	1	0	0	1	1	13	7	21,6041047799082
47	sub-41252	57	1	1	190	104.4	1	1	1	1	1	0	0	0	1	0	18	1	28,9196675900277

Appendix: file description

File	Key ID	Content, of the file, Results of the agregat
dataset-clinical_mod-baseline_version-2.tsv	participant_id	participant consent and baseline.; 1 line per subject
dataset-clinical_mod-baseline_form-postLi_tab-antipdep_version-2.tsv	participant_id	Antidepressants taken during the last 2 years; A varying number of lines per subject corresponding to antidepressants taken
dataset-clinical_mod-baseline_form-postLi_tab-antipsych_version-2.tsv	participant_id	Atypical Antipsychotics taken during the last 2 years; A varying number of lines per subject corresponding to antipsychotic taken
dataset-clinical_mod-baseline_form-postLi_tab-benzos_version-2.tsv	participant_id	Benzodiazepines and Others taken during the last 2 years; A varying number of lines per subject corresponding to benzodiazepines taken
dataset-clinical_mod-baseline_form-postLi_tab-disorders_version-2.tsv	participant_id	Psychological disorders; A varying number of lines per subject corresponding to psychological disorder
dataset-clinical_mod-baseline_form-postLi_tab-family_history_version-2.tsv	participant_id	Family history of mood disorders of biological member [1st degree only (adopted excluded)]; A varying number of lines per subject corresponding to the family history
dataset-clinical_mod-baseline_form-postLi_tab-health_condition_version-2.tsv	participant_id	Health condition (physical comorbidity); A varying number of lines per subject corresponding to health condition
dataset-clinical_mod-baseline_form-postLi_tab-mood_version-2.tsv	participant_id	Mood stabilizers taken during the last 2 years; A varying number of lines per subject corresponding to mood stabilizers taken
dataset-clinical_mod-baseline_form-postLi_tab-neuroleptics_version-2.tsv	participant_id	Conventional neuroleptics taken during the last 2 years; A varying number of lines per subject corresponding to neuroleptics taken
dataset-clinical_mod-baseline_form-preLi_tab-clinical_service_use_version-2.tsv	participant_id	Clinical Service use Inventory for Hospital admission in the 6 months prior to inclusion; A varying number of lines per subject corresponding to hospital admission

dataset-clinical_mod-baseline_form-preLi_tab-csrimeet_version-2.tsv	participant_id	Clinical Service use Inventory for any meeting relevant to mental or physical health in the 6 months prior to inclusion; A varying number of lines per subject corresponding to psychological or physical meeting
dataset-clinical_mod-baseline_form-preLi_tab-med_version-2.tsv	participant_id	Medications taken over the 6 months prior to inclusion at least once a week; A varying number of lines per subject corresponding to medications taken
dataset-clinical_mod-baseline_tab-med_m-1_version-2.tsv	participant_id	Medications taken over the last month at least once a week; A varying number of lines per subject corresponding to medications taken
dataset-clinical_mod-baseline_tab-med_m-2_version-2.tsv	participant_id	Medications taken over the last month at least once a week; A varying number of lines per subject corresponding to medications taken
dataset-clinical_mod-cm_form-cm_tab-cmtrt_version-2.tsv	participant_id	Concomitant medications, other than psychotropic; A varying number of lines per subject corresponding to the concomitant medications taken
dataset-clinical_mod-cm_version-2.tsv	participant_id	Concomitant medications, other than psychotropic; 1 line per subject
dataset-clinical_mod-eos_version-2.tsv	participant_id	Study discontinuation; 1 line per subject
dataset-clinical_mod-inclusion_version-2.tsv	participant_id	identification and inclusion criteria; 1 line per subject
dataset-clinical_mod-pregnancy_tab-cmtrt_version-2.tsv	participant_id	Concomitant medications, other than psychotropic; A varying number of lines per subject corresponding to the concomitant medications other than psychotic
dataset-clinical_mod-pregnancy_tab-extrt_version-2.tsv	participant_id	Concomitant medications, other than psychotropic during pregnancy; A varying number of lines per subject corresponding to investigational drugs or exposures
dataset-clinical_mod-pregnancy_tab-newborn_version-2.tsv	participant_id	Newborn information; A varying number of lines per subject corresponding to the newborn information
dataset-clinical_mod-pregnancy_tab-proc_version-2.tsv	participant_id	Procedures and cares during pregnancy; A varying number of lines per subject corresponding to procedures

dataset-clinical_mod-pregnancy_version-2.tsv	participant_id	On-going pregnancy or planned pregnancy on the next 2 years; A varying number of lines per subject corresponding to pregnancy information
dataset-clinical_mod-sae_tab-cmtrt_version-2.tsv	participant_id + MODULE_M_SAE	Concomitant Medication(s) at the time of the SAE, excluding those used to treat the SAE; A varying number of lines per subject corresponding to concomitant treatment taken for the “MODULE_M_SAE”
dataset-clinical_mod-sae_tab-proc_version-2.tsv	participant_id + MODULE_M_SAE	Additional procedures or medical cares performed during the clinical trial; A varying number of lines per subject corresponding to procedures and cares for the “MODULE_M_SAE”
dataset-clinical_mod-sae_tab-extrt_version-2.tsv	participant_id + MODULE_M_SAE	Investigational Medicinal Product(s) (IMP) or related product(s) administered prior the occurring of the SAE; A varying number of lines per subject corresponding to investigational drugs or exposure for the “MODULE_M_SAE”
dataset-clinical_mod-sae_version-2.tsv	participant_id + MODULE_M_SAE	Serious Adverse Event (SAE) information; A varying number of lines per subject corresponding to the SAE information for the “MODULE_M_SAE”
dataset-clinical_mod-visits_form-visit_version-2.tsv	participant_id + FORM_F_VISIT	Medical appointment information: A varying number of lines per subject corresponding to hospital admission on the visit number “FORM_F_VISIT”*
dataset-clinical_mod-visits_form-visit_tab-csrih_version-2.tsv	participant_id + FORM_F_VISIT	Clinical Service use Inventory for Hospital admission; A varying number of lines per subject corresponding to hospital admission on the visit number “FORM_F_VISIT”*
dataset-clinical_mod-visits_form-visit_tab-csrmeeet_version-2.tsv	participant_id + FORM_F_VISIT	Clinical Service use Inventory for any meeting relevant to mental or physical health; A varying number of lines per subject corresponding to meeting relevant to mental or physical health on the visit number “FORM_F_VISIT”*
dataset-clinical_mod-visits_form-visit_tab-dsrdr_version-2.tsv	participant_id + FORM_F_VISIT	Psychological Disorders; A varying number of lines per subject corresponding to psychological disorder on the visit number “FORM_F_VISIT”*

dataset-clinical_mod-visits_form-visit_tab-hcnd_version-2.tsv	participant_id + FORM_F_VISIT	Health Condition; A varying number of lines per subject corresponding to health condition on the visit number “FORM_F_VISIT”*
dataset-clinical_mod-visits_form-visit_tab-med_version-2.tsv	participant_id + FORM_F_VISIT	medications taken over the last month at least once a week; A varying number of lines per subject corresponding to medications on the visit number “FORM_F_VISIT”*
dataset-clinical_version-2_description.pdf		
dataset-clinical_version-2_description.xlsx		
participants_version-2.tsv	participant_id	1 line per subject