

**STANDARD OPERATING PROCEDURE (SOP) :**

**eCRF DATASETS**

# Collect repository organization:

The organized data files must look like this: (see appendix below for correspondence to former name)

**/rlink**

└──**rlink\_ecrf-upload**

└──**rawdata**

└──**participants\_version-<label>.tsv**

└──**dataset\_description\_version-<label>.pdf**

**└──dataset\_description\_version-<label>.xlsx**

└──dataset-clinical\_**mod-inclusion**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-preLi\_tab-csrih**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-preLi\_tab-med**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-preLi\_tab-csrimeet**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi\_tab-famhist**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi\_tab-antipdep**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi\_tab-antipsy**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi\_tab-benzos**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi**\_**tab-mood**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi**\_**tab-neurol**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi\_tab-dsrdr**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_form-postLi\_tab-hcnd**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_tab-med\_m-1**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_tab-med\_m-2**\_version-<label>.csv

└──dataset-clinical\_**mod-baseline\_tab-med\_m-...**\_version-<label>.csv

└──...(24 )

└──dataset-clinical\_**mod-visits\_form-visit**\_version-<label>.csv

└──dataset-clinical\_**mod-visits\_form-visit\_tab-med**\_version-<label>.csv

└──dataset-clinical\_**mod-visits\_form-visit\_tab-csrih**\_version-<label>.csv

└──dataset-clinical\_**mod-visits\_form-visit**\_**tab-csrimeet**\_version-<label>.csv

└──dataset-clinical\_**mod-visits\_form-visit\_tab-dsrdr**\_version-<label>.csv

└──dataset-clinical\_**mod-visits\_form-visit\_tab-hcnd**\_version-<label>.csv

└──dataset-clinical\_**mod-cm**\_version-<label>.csv

└──dataset-clinical\_**mod-cm\_form-cm\_tab-cmtrt**\_version-<label>.csv

└──dataset-clinical\_**mod-eos**\_version-<label>.csv

└──dataset-clinical\_**mod-dm**\_version-<label>.csv

└──dataset-clinical\_**mod-sae**\_version-<label>.csv

└──dataset-clinical\_**mod-sae\_tab-extrt**\_version-<label>.csv

└──dataset-clinical\_**mod-sae\_tab-cmtrt**\_version-<label>.csv

└──dataset-clinical\_**mod-sae\_tab-proc**\_version-<label>.csv

└──dataset-clinical\_**mod-pregnancy**\_version-<label>.csv

└──dataset-clinical\_**mod-pregnancy\_tab-newborn**\_version-<label>.csv

└──dataset-clinical\_**mod-pregnancy\_tab-cmtrt**\_version-<label>.csv

└──dataset-clinical\_**mod-pregnancy\_tab-extrt**\_version-<label>.csv

└──dataset-clinical\_**mod-pregnancy\_tab-proc**\_version-<label>.csv

**Filename structure :**

A file name consists of a chain of entities, or key-value pairs, a suffix and an extension. Two prominent examples of entities are **dataset** and **tab**.

The **version field** allows the provider to push an intermediate version. All of these versions will be kept and the final labeled version (**version-final**) will contain all the modifications of the previous versions. If there is only one version, the valid value for the version field is final:**version-final**.

**Modules :**

* Inclusion : identification and inclusion criteria.
* Baseline : participant consent and baseline.
* Visits : Medical appointments.
* CM : Concomitant medications.
* DM : Data Management.
* EOS : End of study.
* SAE : Serious Adverse event.
* Pregnancy : On-going pregnancy or planned pregnancy on the next 2 years.

**Forms :**

* preLi : Month 0 (M00), before taking the Lithium.
* postLi : Month 3 (M03), after taking the Lithium.
* m : Medical appointment (visit).
* cm : Concomitant medications.

**Tables :**

* med : medications.
* csrih : Clinical Service use Inventory for Hospital admission.
* csrimeet : Clinical Service use Inventory for any meeting relevant to mental or physical health.
* dsrdr : Psychological Disorders.
* hcnd : Health Condition.
* cmtrt : Concomitant medications, other than psychotic.
* extrt : Investigational drug(s) administered or not during pregnancy or exposure involving the father.
* Famhist: Family History
* proc : Procedures and care.
* newborn : Newborn information.

**M : Visit number.**

## 1. Content of the files :

## Be careful on the Regional Setting of the used text/table editors. All fields are expected in English (US) format. For instance for numbers: English - 123,456.89, French - 123 456,89 .

### Detailed content of the dataset-clinical csv files :

* All csv file contains one line per participant except for files with “visits” module.
* Each subject must be described in all csv files, except for files with ,”visits”, “pregnancy” and “sae” modules; also except for the files with a “tab” key.
* The columns correspond to descriptive information and to questions labels (question’s id of the questionnaire).
* The encoding of this file must be UTF8.
* Each csv header is checked per csv file. (according to the test files previously sent)

### Detailed content of the participants.tsv file :

Some information about the participant are needed. It is used to detect potential inconsistencies.

* The first column must be the **participant\_id (PSC1)**, followed by two columns describing the participants, namely **sex** and **birth\_date**. First level anonymous pseudocode (PSC1).A valid value for this field is 5 integers (in order to comply with the bids format, it is needed to remove the dash)
* The valid values for the sex field are **M** or **F** and the birth\_date field must be formatting with the following format 2009-06-01T00:00:00 (year, month, day, hour (24h), minute, second; this is equivalent to the RFC3339 "date-time" format, day must be set to 01, hour, minute, and second must be set to zero).
* Each participant needs to be described by one and only one row.

**Content of the dataset\_description files :**

* The dataset\_description.pdf corresponds to the RLINK\_CRF\_ANNOTE.pdf
* The dataset\_description.xlsx corresponds to the RLINK\_Data Handling Manual.xlsx

*Notes: the blood lithium level can be found on dataset-clinical\_mod-visits\_form-visit\_version-2.tsv in the columns “PLIMRI” as shown in the eCRF description PDF.*

*On the files from the visits module, the visits number starts 3 months after the lithium initiation (at M03) , visit 1 corresponds to the M03, visit 2 corresponds to the M04...*

# **2. Processings**

No processings has been done on the data.

# **3. Publication repository organization:**

Here are the publication repositories.

rlink

├── rlink\_actigraphy

├── rlink\_ecrf

├── rlink\_mri-anat

├── rlink\_mri-dwi

├── rlink\_mri-lithium

├── rlink\_mri-mrs

├── rlink\_omics-genotyping

├── rlink\_omics-metabolomic

├── rlink\_omics-methylomic

├── rlink\_omics-mirna

├── rlink\_omics-rna

└── rlink\_smartphone

When you ask access to the data you should specify each directory that you want access to.

## 1. ECRF repository content :

## Description of the files:

|  |  |  |
| --- | --- | --- |
| 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 antidep taken |
| dataset-clinical\_mod-baseline\_form-postLi\_tab-antipsy\_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 benzos taken |
| dataset-clinical\_mod-baseline\_form-postLi\_tab-dsrdr\_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-famhist\_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-hcnd\_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-neurol\_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-csrih\_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 new born 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 investigationnal 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 appointement 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-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”\* |
| 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 |

*\* On the files from the visits module, the visits number (FORM\_F\_VISIT) starts 3 months after the lithium initiation (M03) , visit 1 corresponds to the M03, visit 2 corresponds to the M04…:*

FORM\_F\_VISIT:

|  |  |
| --- | --- |
| **Code** | **Time point (Month)** |
| F\_VISIT\_1 | M3 |
| F\_VISIT\_2 | M4 |
| F\_VISIT\_3 | M5 |
| F\_VISIT\_4 | M6 |
| F\_VISIT\_5 | M7 |
| F\_VISIT\_6 | M8 |
| F\_VISIT\_7 | M9 |
| F\_VISIT\_8 | M10 |
| F\_VISIT\_9 | M11 |
| F\_VISIT\_10 | M12 |
| F\_VISIT\_11 | M13 |
| F\_VISIT\_12 | M14 |
| F\_VISIT\_13 | M15 |
| F\_VISIT\_14 | M16 |
| F\_VISIT\_15 | M17 |
| F\_VISIT\_16 | M18 |
| F\_VISIT\_17 | M19 |
| F\_VISIT\_18 | M20 |
| F\_VISIT\_19 | M21 |
| F\_VISIT\_20 | M22 |
| F\_VISIT\_21 | M23 |
| F\_VISIT\_22 | M24 |

## **Extraction codes examples :**

**# %% Use Case : Export age, sex and serum Liyhium level at M3**

file1 = os.path.join(folder, "dataset-clinical\_mod-inclusion\_version-2.tsv")

vars1 = ["participant\_id", "AGE", "SEX"]

df = pd.read\_csv(file1, sep="\t")

file2 = os.path.join(folder, "dataset-clinical\_mod-visits\_form-visit\_version-2.tsv")

vars2 = ["participant\_id", "FORM\_F\_VISIT", "PLIMRI"]

df2 = pd.read\_csv(file2, sep="\t")[vars2]

# Select Month 3 visit

df2 = df2[df2["FORM\_F\_VISIT"] == "F\_VISIT\_1"]

output\_filename = "clinical\_data.csv"

final\_df = pd.merge(df[vars1], df2[vars2], how="outer")

final\_df.to\_csv(output\_filename, sep="\t", index=False)

## **# Load a file and identify one line per subject, grouping by the key\_id**

## folder = "/path/to/the/folder"

## file = "dataset-clinical\_mod-visits\_form-visit\_version-2.tsv"

## df = pd.read\_csv(os.path.join(folder, file), sep="\t")

## key\_ids = ["participant\_id", "FORM\_F\_VISIT"]

## grouped\_df = df.groupby(key\_ids)

## for pid, data in grouped\_df:

## print(pid, data)

## **Appendix:**

|  |  |
| --- | --- |
| Old name | Bids compliant name |
| F\_PREGNANCY\_UTF8.csv | dataset-clinical\_mod-pregnancy\_version-<label>.csv |
| F\_VISIT\_UTF8.csv | dataset-clinical\_mod-visits\_form-visit\_version-<label>.csv |
| INCLUSION\_UTF8.csv | dataset-clinical\_mod-inclusion\_version-<label>.csv |
| M\_BL\_UTF8.csv | dataset-clinical\_mod-baseline\_version-<label>.csv |
| M\_CM\_UTF8.csv | dataset-clinical\_mod-cm\_version-<label>.csv |
| M\_DATA\_MANAGER\_UTF8.csv | dataset-clinical\_mod-dm\_version-<label>.csv |
| M\_EOS\_UTF8.csv | dataset-clinical\_mod-eos\_version-<label>.csv |
| M\_SAE\_UTF8.csv | dataset-clinical\_mod-sae\_version-<label>.csv |
| TAB\_ANTIDEP\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-antipdep\_version-<label>.csv |
| TAB\_ANTIPSY\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-antipsy\_version-<label>.csv |
| TAB\_BENZOS\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-benzos\_version-<label>.csv |
| TAB\_CMTRT\_UTF8.csv | dataset-clinical\_mod-cm\_form-cm\_tab-cmtrt\_version-<label>.csv |
| TAB\_CSRIH\_PRELI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-preLi\_tab-csrih\_version-<label>.csv |
| TAB\_CSRIH\_UTF8.csv | dataset-clinical\_mod-visits\_form-visit\_tab-csrih\_version-<label>.csv |
| TAB\_CSRIMEET\_PRELI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-preLi\_tab-csrimeet\_version-<label>.csv |
| TAB\_CSRIMEET\_UTF8.csv | dataset-clinical\_mod-visits\_form-visit\_tab-csrimeet\_version-<label>.csv |
| TAB\_DSRDR\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-dsrdr\_version-<label>.csv |
| TAB\_DSRDR\_UTF8.csv | dataset-clinical\_mod-visits\_form-visit\_tab-dsrdr\_version-<label>.csv |
| TAB\_FAMHIST\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-famhist\_version-<label>.csv |
| TAB\_HCND\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-hcnd\_version-<label>.csv |
| TAB\_HCND\_UTF8.csv | dataset-clinical\_mod-visits\_form-visit\_tab-hcnd\_version-<label>.csv |
| TAB\_MED\_M1\_UTF8.csv | dataset-clinical\_mod-baseline\_tab-med\_m-1\_version-<label>.csv |
| TAB\_MED\_M2\_UTF8.csv | dataset-clinical\_mod-baseline\_tab-med\_m-2\_version-<label>.csv |
| TAB\_MED\_PRELI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-preLi\_tab-med\_version-<label>.csv |
| TAB\_MED\_VIS\_UTF8.csv | dataset-clinical\_mod-visits\_form-visit\_tab-med\_version-<label>.csv |
| TAB\_MOOD\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-mood\_version-<label>.csv |
| TAB\_NEUROL\_PLI\_UTF8.csv | dataset-clinical\_mod-baseline\_form-postLi\_tab-neurol\_version-<label>.csv |
| TAB\_NEWBORN\_UTF8.csv | dataset-clinical\_mod-pregnancy\_tab-newborn\_version-<label>.csv |
| TAB\_NG\_CMTRT\_UTF8.csv | dataset-clinical\_mod-pregnancy\_tab-cmtrt\_version-<label>.csv |
| TAB\_NG\_EXTRT\_UTF8.csv | dataset-clinical\_mod-pregnancy\_tab-extrt\_version-<label>.csv |
| TAB\_NG\_PROC\_UTF8.csv | dataset-clinical\_mod-pregnancy\_tab-proc\_version-<label>.csv |
| TAB\_SAE\_CMTRT\_UTF8.csv | dataset-clinical\_mod-sae\_tab-cmtrt\_version-<label>.csv |
| TAB\_SAE\_EXTRT\_UTF8.csv | dataset-clinical\_mod-sae\_tab-extrt\_version-<label>.csv |
| TAB\_SAE\_PROC\_UTF8.csv | dataset-clinical\_mod-sae\_tab-proc\_version-<label>.csv |
| RLINK\_Data Handling Manual.xlsx | dataset\_description\_version-<label>.xlsx |
| RLINK\_CRF\_ANNOTE.pdf | dataset\_description\_version-<label>.pdf |