**Xptdef 2 Tutorial**

**(by Guillaume Bonnefois)**

1. **Package installation from Github**

* devtools::install\_github("leonpheng/xptdef2")
* devtools::install\_github('davidgohel/ReporteRsjars')
* devtools::install\_github('davidgohel/ReporteRs')
* devtools::install\_github(“[benjaminrich](https://github.com/benjaminrich)/[PCSmisc](https://github.com/benjaminrich/PCSmisc)”)

Install additional packages from CRAN such as plyr, dplyr, SASxport, Hmisc and certara logo:

* xptdef2::install.packages()

1. **Library to be loaded in R Studio**

* library(xptdef2)

1. **Setup the working folder to save all define documents:**

Example:

* working.folder="C:/Users/lpheng/Desktop/Test"

1. **Use the helps() function for a quick step-by-step procedure**

* helps()

stepBYstep Procedure

1 require(xptdef2)

2 working.folder= enter full path

3 step1()

4 EDIT list of files.csv then save

5 define.library= path/libraryfile.csv

6 step2()

7 EDIT studydefinelist.csv and save.

8 step3(title)

1. **Run step1**

* step1()

A file entitled “list of files.csv” will automatically be generated

* 1. **Edit** list of files.csv **file according to the following:**
* COLUMN A (filename) and COLUMN N (sourcepath):
  + enter file name including extension in column A and copy/paste the source path in column N. Note that dataset files accepted are csv and NONMEM tables, and program files in txt or csv format. For other file formats, it is necessary to convert them to appropriate format prior to running step2.
  + **TIPS:** The list should be ordered from top to bottom (ex: pk dataset.csv, Residuals.csv, Eta.csv, model.mod, Rscript.R, etc.)
* COLUMN B (t*ype*) is for identifying the appropriate file type, **dataset** or **program**
* COLUMN C (rename) is for renaming the files in short name (~8-10 characters) (ex: pkdataset, residualfin, etafin, finalmod) without extension.
* COLUMNs D (*keyvar*) and E (*Structure*) are for datasets only (not applicable for program cells). The *keyvar* and *Structure* are to specify of key variable (or combination of variables) that is unique. For example, if ID and RTIME are variables to identify unique row of PK dataset, then the *keyvar* = ID, RTIME and *Structure* = per subject per time point.
* COLUMN F (P*rogram*): if applicable, enter the name of the program used to generate the dataset. Use the new name in column C to fill out corresponding cells (ex: if Residuals.csv and Eta.csv were derived using finalmod then enter finalmod in Residuals and Eta cells under P*rogram* column).
* COLUMN G (*description*): enter detail to describe each dataset and program (ex: PK dataset for metabolite M, NONMEM control file of final PK model, etc.).
  1. **Setting Program Table for PMDA submission Using list of files.csv**

Note 1: For FDA and EMA submission, these columns H to M should remain blank (see Figure 1).

Note 2: For PMDA submission, all items (datasets, programs, scripts, dependency, etc.) related to TFL in the final report must be included. Ex: pkdataset (for modeling), eta (for boxplots and scatter matrix), residuals (for GOF), vpcoutputs (for VPC), programs, model, R script used, etc. Columns H to M should be completed as below (see Figure 2 for example).

COLUMN H (*progNo*): label all programs with unique tag (example, the BAD labels are 1, 10, 11, 111, since 1 is nested in other tags and 11 is nested in tag 111. The GOOD labels are unique like *#1a*, *#10a, #11a*).

COLUMN I (*Software*): enter platform version (ex: “*NONMEM version 7, R version 3.0.0”*)

COLUMN J (*Purpose*): example, “*the script used to generate Figures 5, 6 and 7 in the report and tables 2.1.1, 3.1.1 and 4.1.1. in Appendix 12*”

COLUMN K (*proNo.input*): ex: If *residuals* dataset was used as input for program *#1a* and *#2a* then enter *#1a#2a* in the *residuals* row.

COLUMN L (*proNo.output*): ex: If residuals is the output generated by program *#1b* then type in *#1b* in the residuals row.

COLUMN M (*progNo.dependent*): ex: If programs *#1b* and *#11c* are dependencies for programs *#1a* and *#2c* then type in *#1b#11c* in rows *#1a* and *#2c*.

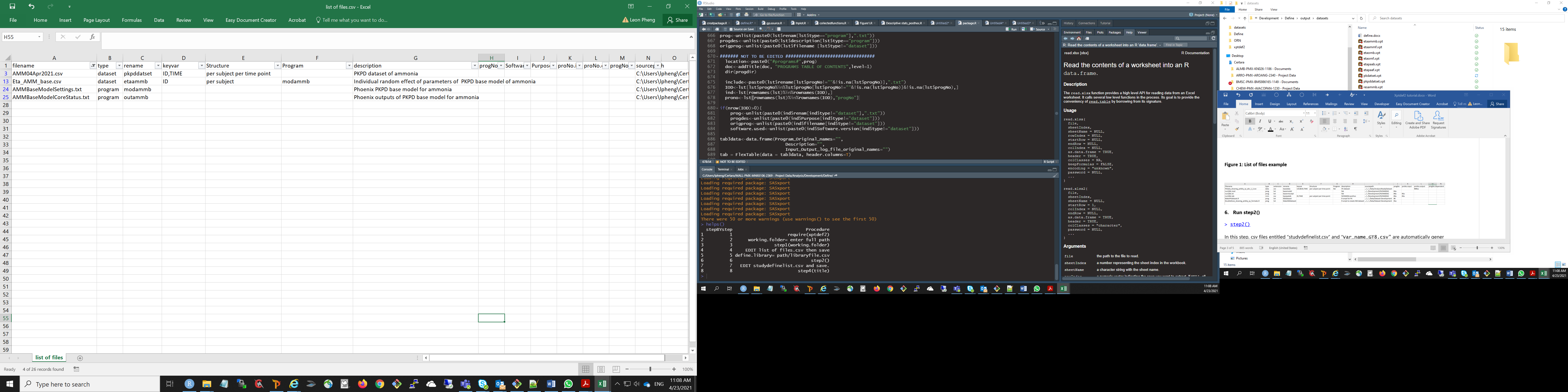
Save "list of files.csv" and run line below:

* + - define.library="no"

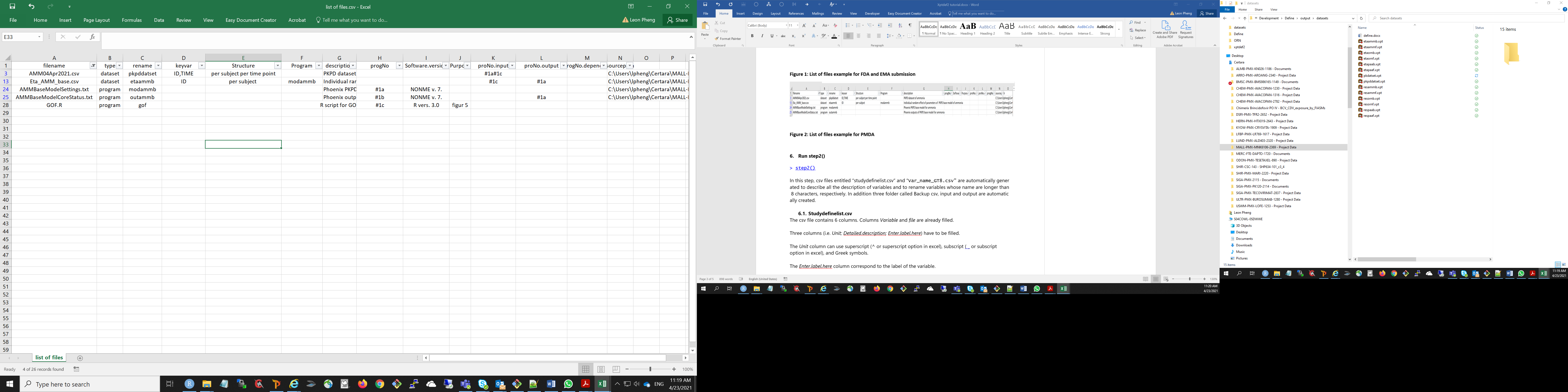
or if library is available,

* + - define.library=“path+library name”

**Figure 1: List of files example for FDA and EMA submission**



**Figure 2: List of files example for PMDA**



1. **Run step2()** 
   * + step2()

In this step, csv files entitled “studydefinelist.csv” is automatically generated.

Note that if library is available, the list will be merged with the library information.

* 1. **Edit Studydefinelist.csv**

Columns A (*Variable)* and F (*file) contain information extracted from datasets*.

Columns B, C and D (i.e. *Unit; Detailed.description; Enter.label.here*) have to be filled out.

Note: *Unit* column accepts superscript (^ or superscript option in excel), subscript (or subscript option in excel), and Greek symbols.

*Enter.label.here* column (D) correspond to the label of the variable. The number of characters should not be more than 40.

.

NOTE: Information could be found in the SAP or previous define document. Pay attention to common labels such as TAD, DV, RTIME, etc.. It should be defined as Time after dose, not “Time after last dose” or “Time after previous dose”.

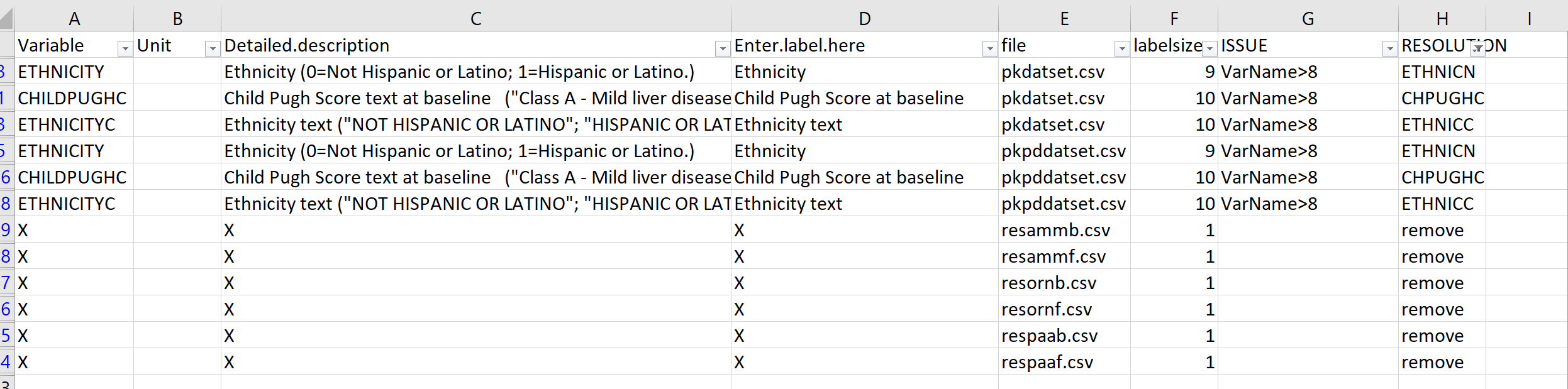
The *Detailed.description* column (C) enable a precise description of the label. Example, for categorical variables, the numerical code must be provided, e.g. Sex of patients (0 = Female; 1= Male)

NOTE: Information could be found in the SAP or previous define document. Pay attention to common labels such as TAD, DV, RTIME, etc.. It should be defined as Time after dose, not “Time after last dose” or “Time after previous dose”.

The *RESOLUTION* column (H) could be used to remove unwanted variable (H= remove) or to rename variable with “>8 characters” just by entering the new name in column H (see Figure 3).

**TIPS:** Create a library with the same format as **Studydefinelist.csv** for common variables. Information from dataset specification or SAP document can be used to merge with the **Studydefinelist.csv** using R or Excel (VLOOKUP function)**.**

**Figure 3: Studydefinelist example**



1. **Run step3()**

The step3 will correct the variable name (remove or rename), convert csv to xpt file and create define document.

Add title and run step3 as example below.

* + - title="Population Pharmacokinetic model for XX project"
    - step3(title=title)

Overall, xpt files and define.docx will be generated in output\datasets subfolder in step3 and txt files generated in output\programs subfolder in step2.

**TIPS:** to change the define location, used option define\_location in step3. As default define\_location= "./output/datasets/".

1. **Refine manually the Word document generated**
2. The word document entitled “define.docx” is located in the output\datasets sub-folder created in step 3. If applicable, replace “XXXX” with “\_” and “#” with “\” using find/replace function in word.

TIPS

1. Edit and format all sections in the word document except for the Label column in section VARIABLE DEFINITION TABLES. The issues found in the Label column require corrections in **Studydefinelist.csv** and re-run of step3.
2. Certara logo could be added and convert the word document in pdf. Archive the define.docx copy and remove it from datasets folder.
3. Make document ready to be sent to client by putting “datasets” and “programs” folders in one zip file.