and SPPM VA Technical Manual

DG EMPL’s Joint Assessment Framework (JAF), version 2, migrated to [R](https://www.r-project.org/about.html), and Social Protection Performance Monitor Vertical Assessment (SPPM VA)

Alek

2024-05-15

# Inventory of files

All the “input” files (scripts and data files) are located in one folder, no subfolders are used for storing these “input” files.

### Scripts

For JAF: the scripts, from JAF\_KEC.R onwards, correspond to the sub-folders of the output folder.

- JAF\_MAIN.R  
- JAF\_functions.R  
- JAF\_indicators\_\_definitions.R  
- JAF\_output.R  
- JAF\_KEC.R  
- JAF\_Compendium.R  
- JAF\_Country\_Compendium.R  
- JAF\_Main\_Indicators.R  
- JAF\_PAs.R  
- JAF\_Country\_Profiles.R  
- JAF\_pdf.R

For SPPM VA:

- SPPM\_VA\_\_MAIN.R  
- JAF\_functions.R —— the same file as for JAF  
- SPPM\_VA\_indicators\_\_definitions.R  
- JAF\_output.R —— the same file as for JAF  
- SPPM\_VA\_KEC.R

### Data files

The following LSF data files are received from Eurostat annually at the end of April.

- IESS\_10\_PA5\_C3\_mod\_Y.csv  
- IESS\_11\_PA7\_2\_S1\_Y.csv  
- IESS\_15\_PA7\_1\_C4\_2064\_FTE\_AA.csv  
- IESS\_16\_PA7\_1\_C5\_AA.csv  
- IESS\_17\_PA7\_1\_C6\_N1\_N2\_AA.csv  
- IESS\_PA2\_C3\_AA.csv  
- IESS\_PA2\_S5\_v2\_Y.csv

For JAF, these template files are also needed, but don’t have to be updated:

- bar\_chart\_template.xml  
- line\_chart\_template.xml  
- Blank\_16x9.pptx

### Additional tools for converting the indicator definitions R script <–> Excel file

The definitions of the indicators are stored in the respective R scripts: JAF\_indicators\_\_definitions.R and SPPM\_VA\_indicators\_\_definitions.R.

It is possible to convert JAF\_indicators\_\_definitions.R into a tabular form i.e. Excel file JAF\_indicators\_\_definitions.xlsx with the script with a long name Convert\_\_JAF\_indicators\_\_definitions.R\_\_to\_\_JAF\_indicators\_\_definitions.xlsx\_\_.R

In turn, the (manually modified) tabular definitions from JAF\_indicators\_\_definitions.xlsx can be converted back to the R script JAF\_indicators\_\_definitions.R with the script with a long name Convert\_\_JAF\_indicators\_\_definitions.xlsx\_\_to\_\_JAF\_indicators\_\_definitions.R\_\_.R

When run (sourced), each of these two scripts will ask interactively the user for the path to the folder where the respective input file (R script or Excel file) is located.

# Running the scripts

It is recommended to use RStudio IDE to run the files.

### JAF

In the code below replace XXXX with your user name and YYYY with your password used in the web browser for web proxy authentication and replace the path in the 6th code line:

Sys.setenv(http\_proxy =  
 "http://XXXX:YYYY@proxy-t2-bx.welcome.ec.europa.eu:8012")  
Sys.setenv(https\_proxy =  
 "http://XXXX:YYYY@proxy-t2-bx.welcome.ec.europa.eu:8012")  
# Replace the path below with the actual path, use / instead of \  
setwd("/path/to/the/folder/with/JAF/scripts")  
source("JAF\_MAIN.R") # This script will run all the needed sub-scripts

### SPPM VA

In the code below replace XXXX with your user name and YYYY with your password used in the web browser for web proxy authentication and replace the path in the 6th code line:

Sys.setenv(http\_proxy =  
 "http://XXXX:YYYY@proxy-t2-bx.welcome.ec.europa.eu:8012")  
Sys.setenv(https\_proxy =  
 "http://XXXX:YYYY@proxy-t2-bx.welcome.ec.europa.eu:8012")  
# Replace the path below with the actual path, use / instead of \  
setwd("/path/to/the/folder/with/SPPM\_VA/scripts")  
source("SPPM\_VA\_\_MAIN.R") # This script will run all the needed sub-scripts

# Output files

A new output folder named on the basis of the date and time when the scripts were run, e.g. JAF output 2024-04-10 11.24.27 will be automatically created in the folder where the scripts are located. This output folder will contain the appropriate sub-folders with specific types of outputs (Excel files, PNG charts, PDFs).

# The key elements of JAF2R / SPPM VA

All scripts rely heavily on the R package data.table, which allows for high-perfromance transformations of data.tables (alternative to the native R data.frame).

They also rely heavily on the R package magrittr and its pipe operator (%>%) which is even more expressive than the native R pipe (|>) and allows writing a cleaner R code by avoiding deeply nested function calls or creating too many intermediate objects; basically:

a %>%  
 f(x=1) %>%  
 g(y=2, .)

is an equvalent of

g(y=2, f(a, x=1))

or

second\_object <- f(a, x=1)  
third\_object <- g(y=2, second\_object)

Excel fles are created with the R package openxlsx2.

### In JAF\_functions.R

This script defines the functions for different data sources used in JAF\_indicators\_\_definitions.R and SPPM\_VA\_indicators\_\_definitions.R. Each function whose name starts with from defines how data is imported from the respective source, e.g. the function fromEurostatDataset defines how data is imported from Eurostat’s SDMX 2.1. web API and fromLMPdataset defines how data is imported from LMP REDISSTAT.

Each from… function must unambigously return a non-empty table (data.frame/data.table) with columns *geo*, *time*, and *value\_*. The columns *geo* and *time* must together uniquely identify each row/observation.

There is an “escape path” for some very idiosyncratic sources: fromSpecialCalculation. It allows calling an arbitrary function and for the moment it is used to calculate only a vacancy rate with the vacancy\_rate function.

The following sources and functions are currently defined:

- fromEurostatDataset  
- fromOECDdataset  
- fromLMPdataset  
- fromBenefitsAndWages  
- fromLFSspecialFile  
- fromDESI  
- fromAMECO  
- fromSpecialCalculation

### In JAF\_output.R

Each of the objects described below (as well as other objects) can be inspected via the RStudio’s *Environmanr* pane or, in the console, with R’s functions print or str.

JAF\_INDICATORS is an R list of hundreds of 10-element lists, all produced by running JAF\_indicators\_\_definitions.R or SPPM\_VA\_indicators\_\_definitions.R. Each of 10-element list contains metadata (*name*, *indicator\_groups*, *unit\_of\_level*, *unit\_of\_change*, *source*, *high\_is\_good*, *calculate\_score\_change*, *calculate\_score\_change\_with\_break\_in\_series*, *reference\_in\_scores*) as its first 9 elements and a raw data.table as the final 10th element. The data.table always contains columns *JAF\_KEY*, *high\_is\_good*, *geo*, *time*, *value\_* and it may also contains columns like *a*, *b*, *c*, etc. and *flags\_a*, *flags\_b*, *flags\_c*, etc., if the indicator was calculated via a JAF formula. E.g.:

List of 10  
 $ name : chr "Overall employment growth"  
 $ indicator\_groups : chr "OUTPUT CONTEXT COMPENDIUM 1 COUNTRY"  
 $ unit\_of\_level : chr "% growth"  
 $ unit\_of\_change : chr "pp"  
 $ source : chr "Eurostat, EU Labour Force Survey"  
 $ high\_is\_good : logi TRUE  
 $ calculate\_score\_change : logi TRUE  
 $ calculate\_score\_change\_with\_break\_in\_series: logi TRUE  
 $ reference\_in\_scores : chr "EU27\_2020"  
 $ value :Classes ‘data.table’ and 'data.frame': 609 obs. of 9 variables:  
 ..$ JAF\_KEY : chr [1:609] "PA1.C1." "PA1.C1." "PA1.C1." "PA1.C1." ...  
 ..$ high\_is\_good: logi [1:609] TRUE TRUE TRUE TRUE TRUE TRUE ...  
 ..$ a : num [1:609] NA NA NA NA NA ...  
 ..$ b : num [1:609] NA NA NA NA NA ...  
 ..$ geo : chr [1:609] "AT" "AT" "AT" "AT" ...  
 ..$ time : chr [1:609] "2003" "2004" "2005" "2006" ...  
 ..$ flags\_a : chr [1:609] ":" ":" ":" ":" ...  
 ..$ flags\_b : chr [1:609] ":" ":" ":" ":" ...  
 ..$ value\_ : num [1:609] NA NA NA NA NA ...

JAF\_GRAND\_TABLE is the “flattened” (one big data.table) version of the stacked all 10th-element data.tables described above.

POP\_WEIGHTS is a data.table with the following columns: *JAF\_KEY*, *geo*, *time*, *popweight*. It is used for a few indicators where the population-weighted variants of scores need to be also calculated. Those few indicators are enumerated in the IndicatorsWithPopulationWeigths data.table.

JAF\_SCORES is an data.table which contains the scores calculated according to the methodology agreed in the Committees (EMCO IG and SPC ISG). Scores make it possible to compare different indicators which are originally expressed in different units.