There are two folders in the export

1) Code: Exclusively contains the R code that has been created by the NEWPROD team as part of a research project aimed at analyzing the effects of product reallocation (product entry and exit) on firm dynamics. In particular, file Tools/fake\_data.R was used to create the dummy datasets in folder 2.

2) Data: Includes simulated and / or censored industry level data (no information for groups of less than four). Simulated data made in “2) code/2a\_make\_dummy\_versions.R” and proceeds in several steps:

1. Firm ids are randomly generated (not using any of the original siren ids) just to make the variables in the simulated data. align with those in the original data.
2. I identify groups of observations that are larger than the cutoff threshold for cell cize (IE greater than 4 firms)
3. I randomly assign observations in the simulated data to a group identified in step 1
4. the remainder of variables not used to identify groups, I generate group level summary statistics (eg mean and SD) and then use those summary statistics to generate simulated values

The data folder includes two xlsx files: one with agrupations of NACE codes into sectors and another with the classification of technology by nace sector according to EUROSTAT classification. All of these xlsx datafiles were created using external information and did not require CASD data to be created.

DECLARATION: THE SIMULATED DATA DOES NOT CONTAIN INFORMATION ON SINGLE FIRMS. THEY ARE SYNTHETICALLY CREATED WITH DISTRIBUTIONS AT THE NACE X YEAR OR PRODFRA X YEAR, EXCLUDING THOSE GROUPS THAT HAVE FEWER THAN 5 FIRMS.