This archive contains data files on pharmaceutical firms and inventors from ESRC Project ES/S015566/1 - **"Mergers and Inventors' Productivity in the Pharmaceutical Industry"**

The following STATA files have been archived:

DATA File name	Description
FIRM.DTA	This file contains the name, country and unique id (compcod) of pharmaceutical companies. Both names and countries have been standardised
FIRM_GROUP.DTA	This file contains the ID of the firms (compcod) and the ID of the group they belong to, in each year. The variable merger_status indicates when a company is the target of a M&As. This means that such company becomes part of a new group.
GROUP.DTA	This file contains the ID of the group and the corresponding name of the group. The ID and Name have been created by the authors.
PATENT.DTA	This file contains the correspondence between the firm ID (compcod) and the application ID (Appln_ID). The application ID "Appln_ID" is the patent ID used by the European Patent Office (EPO). As this is the same ID used by Patstat, this number can be used to link the dataset to all PatStat publications and also to the OECD data. <u>The version of PatStat used is 2017A</u>
PATENT_INVENTOR.DTA	This file allows to link the Appln_ID with the unique id of the inventors of that patent (codinv)
INVENTOR.DTA	This file contains the ID number (codinv) and name and country of inventors. Codinv2 indicates the original ID of the inventors in Patstat. Some codinv 2 have been aggregated to the same codinv using the disambiguation technique by Pezzoni et al (2014)[1]

[1] Pezzoni, M., Lissoni, F., & Tarasconi, G. (2014). How to kill inventors: testing the Massacrator© algorithm for inventor disambiguation. *Scientometrics*, *101*(1), 477-504.

Structure of the datasets:

The 3 datasets in the top part refer to the pharmaceutical firms and the groups they belong. Firms may change group following an acquisition.

The 3 datasets in the bottom part allows to link the pharmaceutical firms to applicant ID and Inventors ID. Note that Appln_ID is the same used in PatStat and the OECD Patent data.

FIRM
Compcod
Name
Country

FIRM-GROUP
Compcod
Year
Group_ID
Merger status

GROUP
Group_ID
Name

PATENT
Compcod
Appln_ID
Priority Year

PATENT-INVENTOR
Appln_ID
Codinv

INVENTOR
Codinv
Codinv2
Name
First Name
Country