

Documentation Link Compustat – USPTO Patent Assignment Dataset

This page provides the data resulting from linking assignees and assignors in the USPTO Patent Assignment Dataset to Compustat gvkeys. We work with a version of the USPTO PAD that was gracefully shared with us by Stuart Graham. Such version precedes by one year the first release available at the USPTO website (<https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-assignment-dataset>). It covers 5,534,135 transactions recorded at the USPTO between January 1970 and January 2013 (inclusive). While the first transaction date is January 1970, the number of transactions recorded in the initial years is negligible. Data coverage seems sufficient for the years 1981-2012.

If you use the code or data, please cite the following two papers:

Arque-Castells, P., and Spulber, D. (2022). Measuring the Private and Social Returns to R&D: Unintended Spillovers versus Technology Markets. *Journal of Political Economy*.
<https://doi.org/10.1086/719908>

Arque-Castells, Pere and Spulber, Daniel F., Firm Matching in the Market for Technology: Business Stealing and Business Creation (September 17, 2021). Northwestern Law & Econ Research Paper No. 18-14, Available at SSRN: <https://ssrn.com/abstract=3041558> or
<http://dx.doi.org/10.2139/ssrn.3041558>

Data and code:

Data: <https://doi.org/10.5281/zenodo.6352358>

Paper and code: <https://doi.org/10.1086/719908> (code in the replication package of the JPE article)

Open access data files:

Original USPTO PAD data (reported in the Zenodo page)

We report the original USPTO PAD data files (see Graham et al., 2018): “assignee.dta”, “assignor.dta”, “assignment.dta”, “documentid.dta” and “assignment_conveyance.dta”.

Auxiliary input data files used in the match (not reported in the Zenodo page)

To generate corporate trees we draw on information from SDC Platinum, Osiris, Compustat and the NBER Patent Data Project (Bessen, 2009). To disambiguate assignees and assignors we use harmonized assignee codes from the USPTO Patent Technology Monitoring Team (PTMT) Custom Extract 2013. An exhaustive description of each one of the intermediate datasets used is provided in the readme file of the replication package of the JPE article.

Main outcome data files (reported in the Zenodo page)

These are the three main outcome files resulting from the matching effort:

The dta file “**dynass_nber_searle.dta**” is a disambiguated list of assignees and assignors in the “assignee.dta” and “assignor.dta” files of the USPTO PAD dynamically linked to Compustat gvkeys either through the parent company or subsidiaries collected from SDC Platinum, Osiris or the NBER

Patent Data Project link. A unique assignor/assignee identifier is stored in the variable iddr. The exact source of the company information used to create the corporate trees for assignees/ors linked to gvkey is reported in the variable source (Osiris, SDC Platinum, Compustat, NBER Patent Data Project).

The dta file “**assignee_standardized_dynass_aux.dta**” is the original “assignee.dta” file extended with gvkey information merged in from the “dynass_nber_searle.dta” file.

The dta file “**assignor_standardized_dynass_aux.dta**” is the original “assignor.dta” file extended with gvkey information merged in from the “dynass_nber_searle.dta” file.

Additional outcome data files (reported in the Zenodo page)

The following three data files are used as input data in the JPE article. They can be produced using either the original PAD files or the matched PAD files.

The file “**patents_conveyance.dta**” is a dataset of “transaction record-patent” observations. It supplements the USPTO PAD “patentid.dta” file with information mostly from other USPTO PAD files (e.g. execution date, assignment conveyance, names of assignor and assignee) and cleans the assignment type field. Serves as basis for creating the other datasets below (e.g. traded patents, original assignments, etc).

The file “**traded.dta**” is a dataset of “transaction record - patent” observations where the conveyance type of the transaction is “assignment of the assignor’s interest”. It includes a generated variable (traded) that indicates if the assignment is a patent trade (i.e. second or subsequent assignment or first assignments after the grant date).

The file “**issuedpat_assignee_gvkeys.dta**” is a dataset of “issued patent-gvkey” observations that includes patents linked to gvkeys through first assignment before issuance (presumably original assignments from inventors to companies). We supplement the data with external patent information mostly from the USPTO PTMT Custom Extract 2013 to merge in the main uspc technology class of each patent and forward citation counts.

References

Bessen, James. 2009. “NBER PDP Project user documentation: Matching patent data to Compustat firms.” Unpublished documentation. <http://users.nber.org/~jbessen/matchdoc.pdf>. Data available at: <https://sites.google.com/site/patentdatapoint/Home/downloads?authuser=0>

Graham, S. J., Marco, A. C., & Myers, A. F. (2018). Monetizing marks: Insights from the USPTO trademark assignment dataset. *Journal of Economics & Management Strategy*, 27(3), 403-432.