

DISCERN 2.0 Data Dictionary

By Ashish Arora, Sharon Belenzon, Larisa Cioaca, Lia Sheer, Hyun Moh (John) Shin, and Dror Shvadron
 Updated: August 2024

Researchers beginning to use DISCERN data should first read Section A of the *Data Appendix DISCERN DEC 2020* file available in Version 1.2.1. It provides detailed background information on how the firm panel was constructed. An updated *Data Appendix* file explaining the new subsidiary data and matches to patents and publications will be provided shortly.

The data repository contains two main folders. The folder *code_examples* contains six code files that demonstrate how to (1) preprocess the Compustat data, (2) calculate accurate stocks of patents, patent applications, and publications while considering the dynamic ownership of subsidiaries and firms over time, and (3) assemble the firm panel. It also contains code for standardizing the name lists.

The folder *output_files* contains Stata (.dta) comma-separated (.csv) versions of the eight DISCERN data files. Below are descriptions of the variables included in each data file.

1. Firm-Year Panel

DISCERN is a dataset of U.S.-headquartered, publicly traded, R&D-performing firms with at least three consecutive years of active records in Compustat. The main output file is *discern_firm_panel_1980_2021*. Each record pertains to an ultimate owner (UO) firm-year. Unlike in the original DISCERN dataset, the firm sample is no longer restricted to firms with at least a patent stock of 1 (including dynamic ownership).

Table 1. Contents of *discern_firm_panel_1980_2021*

Variable	Type	Variable label	Example	Notes
<i>permno_adj</i>	numeric	UO firm unique ID	10006	A firm is uniquely identified by its <i>permno_adj</i> , not by its <i>gvkey</i> . The latter establishes a connection to the Compustat financial data.
<i>gvkey</i>	string	UO GVKEY	001010	Was variable <i>gvkey_str</i> in the original DISCERN. Sourced from Compustat.
<i>fyear</i>	numeric	Data Year - Fiscal	1980	Following the Compustat convention, if the fiscal year-end month falls in January through May, use the calendar year minus 1 year. If the fiscal year-end month falls in June through December, use the calendar year. Was variable <i>year</i> in the original DISCERN. Sourced from Compustat.

Variable	Type	Variable label	Example	Notes
<i>name_permno_adj</i>	string	Standardized firm name (header)	ACF INDUSTRIES HOLDING CORP	The most recent representative firm name (standardized).
<i>pat_grant_fyear</i>	numeric	Flow of granted patents per permno_adj-year	50	Was variable <i>pat_yr</i> in the original DISCERN. Sourced from PatentsView. We count fractional ownership when calculating the flow of granted patents.
<i>pat_app_fyear</i>	numeric	Flow of patent applications per permno_adj-year	60	Sourced from PatentsView. We count fractional ownership when calculating the flow of patent applications.
<i>pub_fyear</i>	numeric	Flow of publications per permno_adj-year	2	Was variable <i>pub_yr</i> in the original DISCERN. Sourced from OpenAlex. We count fractional ownership when calculating the flow of publications.

2. Patents

The file *discern_pat_grant_1980_2021* contains the list of USPTO granted utility patents matched to DISCERN firms (including ultimate owners from 1980 to 2021 and subsidiaries from 1994 to 2021, for years when the firm was publicly traded). Each record pertains to a patent-firm match. Co-invented patents may be matched to multiple DISCERN firms.

Table 2. Contents of *discern_pat_grant_1980_2021*

Variable	Type	Variable label	Example	Notes
<i>patent_id</i>	string	Patent number	4206709	Was variable <i>publn_nr</i> in the original DISCERN. Sourced from PatentsView.
<i>patent_date</i>	string	Patent grant date	1980-06-10	Replaces the variable <i>publn_year</i> in the original DISCERN. Sourced from PatentsView.
<i>assignee_name</i>	string	Patent assignee name	ACF Industries, Incorporated	Sourced from PatentsView.
<i>fyear</i>	numeric	Data Year - Fiscal	1980	Fiscal year (following Compustat convention) of the grant date.
<i>name_std</i>	string	Standardized name matched	ACF IND INC	The standardization code is available in the <i>name_std.do</i> code file
<i>id_name</i>	numeric	ID of name_std matched	36	

Variable	Type	Variable label	Example	Notes
<i>sample</i>	string	U=UO match//S=subsidiary match	U	
<i>permno_adj</i>	numeric	UO firm unique ID at the time the patent was granted	10006	
<i>name_permno_adj</i>	string	Standardized firm name (header)	ACF IND INC	The most recent representative firm name (standardized).

The file *discern_pat_app_1980_2021* contains the list of USPTO patent applications (whether they were eventually granted or not) matched to DISCERN firms from 2001 (including both ultimate owners and subsidiaries, for years when the firm was publicly traded). Each record pertains to a patent application-firm match. Co-invented patent applications may be matched to multiple DISCERN firms.

Table 3. Contents of *discern_pat_app_1980_2021*

Variable	Type	Variable label	Example	Notes
<i>pgpub_id</i>	numeric	Patent document number	20030111736	Sourced from PatentsView.
<i>application_id</i>	numeric	Patent application number	10016939	Sourced from PatentsView. We use <i>current_pgpub_id_flag</i> (and <i>current_patent_id_flag</i> in cases when the application was granted) to retain the current application version (as indicated in the February 13, 2024 version of PatentsView). When multiple <i>pgpub_ids</i> are associated with an application, we choose the earliest <i>pgpub_id</i> by assignment date.
<i>filing_date</i>	string	Patent application filing date	2001-12-14	Sourced from PatentsView.
<i>assignee_name</i>	string	Patent application assignee name	DPAC TECHNOLOGIES CORP.	Sourced from PatentsView.
<i>patent_id</i>	string	Patent number	7081373	Sourced from PatentsView.
<i>fyear</i>	numeric	Data Year - Fiscal	2001	Fiscal year (following Compustat convention) of the filing date.
<i>published_date</i>	string	Patent application publication date		Sourced from PatentsView.

Variable	Type	Variable label	Example	Notes
<i>name_std</i>	string	Standardized name matched	DPAC TECH CORP	
<i>id_name</i>	numeric	ID of name_std matched	2691	
<i>sample</i>	string	U=UO match//S=subsidiary match	U	
<i>permno_adj</i>	numeric	UO firm unique ID at the time the patent application was filed	10012	
<i>name_permno_adj</i>	string	Standardized firm name (header)	DPAC TECH CORP	The most recent representative firm name (standardized).

A match between DISCERN firms and assignees in the Patent Assignment Dataset (PAD) is forthcoming.

3. Scientific Publications

The file *discern_pub_1980_2021* contains the list of publications (i.e., journal articles) matched to DISCERN firms (including ultimate owners from 1980 to 2021 and subsidiaries from 1994 to 2021, for years when the firm was publicly traded). Each record pertains to an author-publication-firm match. Co-authored publications may be matched to multiple DISCERN firms.

Table 4. Contents of *discern_pub_1980_2021*

Variable	Type	Variable label	Example	Notes
<i>openalex_id</i>	string	Publication number	https://openalex.org/W1487016295	Sourced from OpenAlex.
<i>earliest_pub_date</i>	string	Earliest publication date	01jan2002	Crossref_date if available and openalex_date otherwise.
<i>openalex_date</i>	string	Publication date from OpenAlex	2003-06-25	Sourced from OpenAlex.
<i>crossref_date</i>	string	Publication date from Crossref	2002-01-01	Sourced from Crossref.
<i>fyear</i>	numeric	Data Year - Fiscal	2001	Fiscal year (following Compustat convention) of the publication date.

Variable	Type	Variable label	Example	Notes
<i>name_std</i>	string	Standardized name matched	SUN MICROSYS INC	
<i>id_name</i>	numeric	ID of name_std matched	8876	
<i>sample</i>	string	U=UO match//S=subsidiary match	U	
<i>permno_adj</i>	numeric	UO firm unique ID at the time of publication	10078	
<i>name_permno_adj</i>	string	Standardized firm name (header)	SUN MICROSYS INC	The most recent representative firm name (standardized).
<i>author_id</i>	string	Author ID from OpenAlex	https://openalex.org/A5090599036	Sourced from OpenAlex.
<i>author_orcid</i>	string	Author ORCID	https://orcid.org/0000-0002-7761-1454	Sourced from OpenAlex.
<i>author_name</i>	string	Author name	Shashank Kumar Dubey	Sourced from OpenAlex.
<i>author_affil</i>	string	Author affiliation	Sun Microsystems Inc., Sunnyvale, CA, USA	Sourced from OpenAlex.

4. Dynamic Ownership

The file *discern_uo_names* contains the list of standardized names used by ultimate owners from 1980 to 2021. Each record pertains to a distinct combination of *name_std-permno_adj1*. A given firm (*permno_adj*) may use multiple names at the same time, and a *name_std* may appear under several *permno_adj*.

Table 5. Contents of *discern_uo_names*

Variable	Type	Variable label	Example	Notes
<i>id_name</i>	numeric	ID unique at sample-name_std-permno_adj1 level	5416	
<i>sample</i>	string	U=UO+Labs+traded-Subs	U	
<i>name_std</i>	string	Standardized name	KRAFT INC	
<i>fyear1</i>	numeric	First fiscal year assigned to owner 1	1986	

Variable	Type	Variable label	Example	Notes
<i>nyear1</i>	numeric	Last fiscal year assigned to owner 1	1988	
<i>permno_adj1</i>	numeric	Firm ID owner 1	61970	
<i>name_acq1</i>	string	Standardized name owner 1	KRAFT INC	
<i>fyear2</i>	numeric	First fiscal year assigned to owner 2	1989	
<i>nyear2</i>	numeric	Last fiscal year assigned to owner 2	2000	
<i>permno_adj2</i>	numeric	Firm ID owner 2	13901	
<i>name_acq2</i>	string	Standardized name owner 2	PHILIP MORRIS INC	
<i>fyear3</i>	numeric	First fiscal year assigned to owner 3	2001	
<i>nyear3</i>	numeric	Last fiscal year assigned to owner 3	2011	
<i>permno_adj3</i>	numeric	Firm ID owner 3	89006	
<i>name_acq3</i>	string	Standardized name owner 3	KRAFT FOODS INC	
<i>fyear4</i>	numeric	First fiscal year assigned to owner 4	2012	
<i>nyear4</i>	numeric	Last fiscal year assigned to owner 4	2014	
<i>permno_adj4</i>	numeric	Firm ID owner 4	13598	
<i>name_acq4</i>	string	Standardized name owner 4	KRAFT FOODS GRP INC	
<i>fyear5</i>	numeric	First fiscal year assigned to owner 5	2015	
<i>nyear5</i>	numeric	Last fiscal year assigned to owner 5	2021	
<i>permno_adj5</i>	numeric	Firm ID owner 5	23077	
<i>name_acq5</i>	string	Standardized name owner 5	HEINZ HJ CO	
<i>fyear0</i>	numeric	First fiscal year assigned to owner 0	1980	Owner "0" is usually the pre-IPO owner, if applicable.

Variable	Type	Variable label	Example	Notes
<i>nyear0</i>	numeric	Last fiscal year assigned to owner 0	1985	Owner “0” is usually the pre-IPO owner, if applicable.
<i>permno_adj0</i>	numeric	Firm ID owner 0	61970	Owner “0” is usually the pre-IPO owner, if applicable.
<i>name_acq0</i>	string	Standardized name owner 0	DART & KRAFT INC	Owner “0” is usually the pre-IPO owner, if applicable.
<i>matched_pat</i>	numeric	Indicator for <i>id_name</i> with patents matched	1	

The file *discern_sub_names* contains the list of standardized names used by subsidiaries from 1994 to 2021. Each record pertains to a distinct combination of *name_std-country_code-subdiv_code-permno_adj1*.

Table 6. Contents of *discern_sub_names*

Variable	Type	Variable label	Example	Notes
<i>id_name</i>	numeric	ID unique at sample-name_std-permno_adj1 level	111046	
<i>sample</i>	string	S=SEC Exhibit 21	S	
<i>name_std</i>	string	Standardized name	BIRD PROD CORP	
<i>country_code</i>	string	Country code	US	
<i>subdiv_code</i>	string	State code (US only)	CA	
<i>fyear1</i>	numeric	First fiscal year assigned to owner 1	1995	
<i>nyear1</i>	numeric	Last fiscal year assigned to owner 1	2000	
<i>permno_adj1</i>	numeric	Firm ID owner 1	62092	
<i>name_acq1</i>	string	Standardized name owner 1	THERMEDICS DETECTION INC	
<i>fyear2</i>	numeric	First fiscal year assigned to owner 2	2001	
<i>nyear2</i>	numeric	Last fiscal year assigned to owner 2	2006	
<i>permno_adj2</i>	numeric	Firm ID owner 2	89243	

Variable	Type	Variable label	Example	Notes
<i>name_acq2</i>	string	Standardized name owner 2	VIASYS HLTHCR INC	
<i>fyear3</i>	numeric	First fiscal year assigned to owner 3	2007	
<i>nyear3</i>	numeric	Last fiscal year assigned to owner 3	2009	
<i>permno_adj3</i>	numeric	Firm ID owner 3	21371	
<i>name_acq3</i>	string	Standardized name owner 3	CARDINAL DISTRIBUTION INC	
<i>fyear4</i>	numeric	First fiscal year assigned to owner 4	2010	
<i>nyear4</i>	numeric	Last fiscal year assigned to owner 4	2014	
<i>permno_adj4</i>	numeric	Firm ID owner 4	92988	
<i>name_acq4</i>	string	Standardized name owner 4	CAREFUSION CORP	
<i>fyear5</i>	numeric	First fiscal year assigned to owner 5	2015	
<i>nyear5</i>	numeric	Last fiscal year assigned to owner 5	2016	
<i>permno_adj5</i>	numeric	Firm ID owner 5	39642	
<i>name_acq5</i>	string	Standardized name owner 5	BECTON DICKINSON & CO	
<i>fyear0</i>	numeric	First fiscal year assigned to owner 0	1980	Owner "0" is usually the pre-IPO owner, if applicable.
<i>nyear0</i>	numeric	Last fiscal year assigned to owner 0	1987	Owner "0" is usually the pre-IPO owner, if applicable.
<i>permno_adj0</i>	numeric	Firm ID owner 0	36505	Owner "0" is usually the pre-IPO owner, if applicable.
<i>name_acq0</i>	string	Standardized name owner 0	SIGNAL CO INC	Owner "0" is usually the pre-IPO owner, if applicable.

The file *permno_gvkey* dynamically links firms to their Compustat financial data. A firm can have multiple GVKEYs and a GVKEY can be related to multiple firms over different periods. To merge with Compustat data, users should obtain a license for the Compustat North America Fundamentals Annual data file.

Table 7. Contents of *permno_gvkey*

Variable	Type	Variable label	Example	Notes
<i>gvkey</i>	string	UO GVKEY	014112	Was variable <i>gvkey_str</i> in the original DISCERN. Sourced from Compustat.
<i>permno_adj</i>	numeric	UO firm unique ID	10066	
<i>fyear1_adj</i>	numeric	GVKEY start fyear in Compustat (adjusted)	1987	
<i>fyearn_adj</i>	numeric	GVKEY end fyear in Compustat (adjusted)	2021	
<i>min_y_permno</i>	numeric	UO firm start fyear	1983	Sourced from <i>discern_uo_names.dta</i> .
<i>max_y_permno</i>	numeric	UO firm end fyear	2021	Sourced from <i>discern_uo_names.dta</i> .

Table 8. Contents of *name_permno_adj*

Variable	Type	Variable label	Example	Notes
<i>permno_adj</i>	numeric	UO firm unique ID	10066	
<i>name_permno_adj</i>	string	Standardized firm name (header)	ACF IND INC	The most recent representative firm name (standardized).

The DISCERN 2.0 dataset is freely available under the [O-UDA-1.0 License](#), permitting unrestricted use for research and commercial purposes. We request that users provide proper citations when utilizing the dataset. The license also allows for the creation of derivative datasets based on DISCERN, with the condition that creators ask their downstream users to cite the original authors appropriately.

If you use the data, please add these citations:

1. Arora, A., Belenzon, S., Cioaca, L., Sheer, L., Shin, H.M., & Shvadron, D. (2024). DISCERN 2.0: Duke Innovation & Scientific Enterprises Research Network [Dataset]. In Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.3594642>
2. Arora, A., Belenzon, S., Cioaca, L., Sheer, L., & Shvadron, D. (2024). Back to the Future: Are Big Firms Regaining their Scientific and Technological Dominance? Evidence from DISCERN 2.0 (available soon).