

JRC-OECD, COR&DIP© database v.1, 2017

IP bundle of top corporate R&D investors

BACKGROUND INFORMATION

The IP bundle of top corporate R&D investors database (*JRC-OECD COR&DIP©*) results out of the collaboration between the European Commission, JRC Directorate B – Growth and Innovation, and the OECD Directorate for Science, Technology and Innovation (STI).

The *COR&DIP©* database contains information about the R&D activity and IP assets (i.e. patents, designs and trademarks) of the top 2000 corporate R&D performers worldwide. Information about the R&D investors is taken from the [2015 EU Industrial R&D Investment Scoreboard](#). Industrial property (IP)-related information is taken from EPO's Worldwide Patent Statistical Database (PATSTAT, Autumn 2016) for patents and design patents filed at the USPTO (United States Patent and Trademark Office), from USPTO for US trademarks, and registered designs and trademarks from the European Union Intellectual Property Office (EUIPO) complement the data. These cover:

- 1) Families of patent applications filed at the five top IP offices (IP5) in the world, namely: EPO (European Patent Office), JPO (Japan Patent Office), KIPO (Korean Intellectual Property Office), SIPO (State Intellectual Property Office of the People's Republic of China), and USPTO. IP5 patent families are composed of members filed at least in one of the IP5, excluding single filings (see Daiko et al., 2017);
- 2) Trademark applications filed at EUIPO and the USPTO;
- 3) Design applications filed at the EUIPO and USPTO design patents.

IP data were linked to enterprise data using the names of the top corporate R&D investors and of their subsidiaries (as of 2014) and matching them to the applicants' names provided in IP documents. The linking was carried out on a by-country basis using a series of algorithms contained in the Imalinker system (Idener Multi Algorithm Linker) developed for the OECD by IDENER, Seville, 2013.

DATABASE STRUCTURE & COVERAGE

The *JRC-OECD COR&DIP© v.1* database is articulated over 8 flat files using a pipe character field separator ("|"). The text is enclosed in quotes "". Data are organised as a set of relational tables that can be linked using three identifiers (see schema below):

- 1) The *Company_id* identifier should be used to match the 'Top Corporate R&D Investors' table with the 'Patent portfolio - 2012-14', the 'Trademark portfolio - 2012-14', the 'Design portfolio - 2012-14' and the 'Top Corporate R&D Investors – Financial' tables.
- 2) The patent, trademark and design identifiers (*Patent_Appln_id*, *TM_Appln_id*, *Design_Appln_id*) contained in the first two tables should be used to link the 'Patent classes', 'Trademark classes' or 'Design classes' tables.

RESTRICTION SOURCE & CONTACTS

Please note that the " JRC-OECD COR&DIP© data " data are made available to the public for research and analytical purposes.

When referring to the database, please use the full text detailed below in references and quote the database as "*JRC-OECD COR&DIP© database, v.1, 2017*" in the text.

Comments and questions should be sent to :

STI.Microdatalab@oecd.org or JRC-B3-IRITEC@ec.europa.eu

Further information about the methodology developed for building the database can be found in the references mentioned below.

For further information on EPO's PATSTAT, please contact patstat@epo.org.

For internal purposes, and to help improving future versions of the dataset, we would be grateful if you would provide feedback about possible issues encountered when using the data, the purpose/analysis for which the data were used, as well as possible desiderata.

REFERENCES

Daiko T., Dernis H., Dosso M., Gkotsis P., Squicciarini M., Vezzani A. (2017). *World Corporate Top R&D Investors: Industrial Property Strategies in the Digital Economy*. A JRC and OECD common report. Luxembourg: Publications Office of the European Union.

The report can be downloaded [here](#).

COR&DIP© v.1: SDatabase structure

Top Corporate R&D Investors - Financial	
Company_id*	Unique company identifier (from 1 to 2000)*
Year	2011-2014
RD	Research and Development investment (million €)
NS	Net sales (million €)
CAPEX	Capital expenditure (million €)
OP	Operating profits (million €)
EMP	Number of employees
Top Corporate R&D Investors	
Company_id*	Unique company identifier (from 1 to 2000)*
Company_name	Company name, as listed in the 2015 Scoreboard
Ctry_code	ISO2 country code
Worldrank	From 1 to 2000, as ranked in the 2015 Scoreboard
ICB-3D	Industry sector, as listed in the 2015 Scoreboard
NACE2	NACE, rev.2 sectors
ISIC4	ISIC, rev. 4 sectors
Patent portfolio - 2012-14	
Company_id*	Unique company identifier*
Patent_appln_id	Patent application identifier (<i>Appln_id</i> from PATSTAT, Autumn 2016)
Publn_auth	IP5 Offices (EP, JP, KR, US, CN)
Patent_publn_nr	Patent publication number (from PATSTAT, Autumn 2016)
Patent_filing_date	Application date
Inpadoc_family_id	Patent family identifier (from PATSTAT, Autumn 2016)
Family_filing_date	Earliest filing date of the family (also based on non-IP5 members)
IP5_2_offices	1 if family has been filed at two or more IP5 offices
Patent classes	
Patent_appln_id	Patent application identifier (<i>Appln_id</i> from PATSTAT, Autumn 2016)
IPC_class	IPC Class (<i>Class_symbol</i> from PATSTAT, Autumn 2016)
WIPO_tech	WIPO Technology class
Trademark portfolio - 2012-14	
Company_id*	Unique company identifier*
TM_appln_id	Trademark identifier
Publn_auth	IP Offices (EU, US)
TM_number	Trademark number
TM_filing_date	Application date
Trademark classes	
TM_appln_id	Trademark identifier
NICE_class	NICE Class
Design portfolio - 2012-14	
Company_id*	Unique company identifier*
Design_id	Design identifier
Publn_auth	IP Offices (EU, US)
Design_number	Design number
Design_filing_date	Application date
Design classes	
Design_id	Design identifier
Locarno_class	Locarno Class

* Company identifiers have been randomly allocated to companies and are different to those provided in COR&DIP© v.0, 2015.