EC-JRC/OECD COR&DIP© database, 2015
IP bundle of top corporate R&D investors

BACKGROUND INFORMATION

The IP bundle of top corporate R&D investors database (***EC-JRC/OECD COR&DIP©***) results out of the collaboration between the EC-JRC Institute for Prospective Technological Studies (IPTS) and the OECD Directorate for Science, Technology and Innovation (STI).

The ***COR&DIP©*** database contains information about the R&D activity and inventive output (i.e. patents and trademarks) of the top 2000 corporate R&D performers worldwide. Information about the R&D investors is taken from the 2013 EU Industrial R&D Investment Scoreboard. Intellectual property (IP)-related information is taken from EPO's Worldwide Patent Statistical Database (PATSTAT, Autumn 2014) database for patents and from selected IP offices in the case of trademarks:

1) 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 (United States Patent and Trademark Office).
2) Trademark applications filed at the USPTO, OHIM (Office for Harmonization in the Internal Market) and IP AUS (IP Australia).

IP data were linked to enterprise data using the names of the top corporate R&D investors and of their subsidiaries (as of 2012) and matching them to the applicants' names provided in patent and trademark 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 (see also Squicciarini, M. & Dernis, H, 2013).

Patent data relate to families of patent applications with members filed at least in one of the IP5, excluding single filings. This is done to identify patents of relatively higher and comparable value. Hence, applications filed only in one of the IP5 offices, i.e. EPO, JPO, KIPO, SIPO and USPTO, are considered only in so far as another family member has been filed in any other office worldwide (anywhere in the world, not necessarily at another IP5 office). Data built following a stricter definition of IP5 patent families, which restricts the patent sample to patent families having members filed in at least two of the IP5 offices considered, are also provided (i.e. IP5_def3).

DATABASE STRUCTURE & COVERAGE

The ***EC-JRC/OECD COR&DIP©*** database is articulated over 6 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 - 2010-12', the 'Trademark portfolio - 2010-12' and the 'Top Corporate R&D Investors – Financial' tables.
2) The patent and trademark identifiers (*Patent_Appln_id* and *TM_Appln_id*) contained in the first two tables should be used to link the 'Patent classes' and 'Trademark classes' tables.
3) The *Patent_Appln_id* helps retrieving additional patent-related information directly from PATSTAT (Autumn 2014 version).

RESTRICTION SOURCE & CONTACTS

Please note that the "Top corporate R&D investors worldwide & IP bundles" data are made available to the public for research and analytical purposes.

When referring to the database in a text, please quote the database as ***"EC-JRC/OECD COR&DIP© database, v.0. 2015"***, and use the full text detailed below in references.

Comments and questions should be sent to :

STI.Microdatalab@oecd.org or jrc-ipts-iri@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

Dernis H., Dosso M., Hervás F., Millot V., Squicciarini M. and Vezzani A. (2015). World Corporate Top R&D Investors: Innovation and IP bundles. A JRC and OECD common report. Luxembourg: Publications Office of the European Union.

The report can be downloaded here.

Database structure

Top Corporate R&D Investors - Financial	
Company_id	Unique company identifier
Year	2009-2012
Worldrank	From 1 to 2000
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
Company_name	Company name as listed in 2013 Scoreboard
Ctry_code	ISO2 country code
ICB-3D	Industry sector, as listed in 2013 Scoreboard
NACE2	NACE, rev.2 sectors
ISIC4	ISIC, rev. 4 sectors

Patent portfolio - 2010-12	
Company_id	Unique company identifier
Patent_appln_id	Patent application identifier (PATSTAT, Autumn 2014)
Publn_auth	IP5 Offices (EP, JP, KR, US, CN)
Patent_publn_nr	Patent publication number (as provided in PATSTAT)
Patent_filing_date	Application date
Inpadoc_family_id	Patent family identifier (PATSTAT, Autumn 2014)
Family_filing_date	Earliest filing date of the family (also based on non-IP5 members)
IP5_Def3	1 if family under definition 3

Patent classes	
Patent_appln_id	Patent application identifier (PATSTAT, Autumn 2014)
IPC_class	IPC Class (as provided in PATSTAT)
WIPO_tech	WIPO Technology class

Trademark portfolio - 2010-12	
Company_id	Unique company identifier
TM_appln_id	Trademark identifier
Publn_auth	IP Offices (OHIM, US, AU)
TM_number	Trademark number
TM_filing_date	Application date

Trademark classes	
TM_appln_id	Trademark identifier
NICE_class	NICE Class