PATSTAT: Complementary data sets

Table of content

1.   About this document

PATSTAT is worldwide data, designed to facilitate statistical analyses of patent data. PATSTAT is provided by the EPO since 2006. Over the years, a number of other data sets, which are based on PATSTAT's data or add value to PATSTAT data, have been made available by various organisations or persons. This document aims to collect basic information about these complementary data sets to promote them in the data analyst community.

Data sets included in this document are selected according to these criteria:
- They can be combined with PATSTAT with little effort, e. g. via the application identifier
- They are publicly available for free or at marginal costs
- They should be maintained, or at least the data should not be too outdated
- They are in the domain of industrial property

For details about the access, license conditions etc. please check with the respective data provider.

More information about PATSTAT on www.epo.org/patstat

Should you want to contact the EPO user support, please use the following links:
- PATSTAT access, subscription or prices
- other PATSTAT-related aspects

🚩 **NOTE**:
- **In the most recent version of this document, a number of data sets that are no longer available are removed, which does not mean that the remaining ones are all available and usable**
- **For the remaining data sets, a number of email addresses and/or websites may no longer be valid**

2. Data sets

2.1. Person data: Name harmonization and gender

Name	OECD HAN database
Content	Harmonised patent applicants' names
Linking data element to PATSTAT	PERSON_ID – person identifier APPLN_ID – application identifier
Provider	OECD
Access	https://oecd/ipstats
Comment	Registration required (fill in an on-line form); Contact: sti.microdatalab@oecd.org This data is already included in the PATSTAT table TLS906_PERSON.

Name	Gender of inventors
Content	Gender of inventors
Linking data element to PATSTAT	PERSON_ID – person identifier
Provider	UK Intellectual Property Office
License	Open Government License https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
Access	Latest "Gender inference table" with PERSON_ID and 2019 study paper in https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship-2019-edition "Name-gender dataset" (First names and their gender) in https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship
Comment	Study: Gender profiles in worldwide patenting: an analysis of female inventorship (2019 edition); see link above

2.2. Regionalization

Name	OECD REGPAT Database
Content	EPO and PCT patents by regions (based on addresses of inventors or applicants)
Linking data element to PATSTAT	PERSON_ID – person identifier APPLN_ID – application identifier
Provider	OECD
Access	http://oecd/ipstats
Comment	Registration required (fill in an on-line form); Contact: sti.microdatalab@oecd.org

Name	Geocoding of worldwide patent data
Content	Latitude, longitude, city, region, and country of the inventor and applicants for patent first filings.
Linking data element to PATSTAT	APPLN_ID – application identifier
Provider	ETH Zurich, EPFL Lausanne ; contact: seliger@kof.ethz.ch
Access	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OTTBDX
Comment	Research paper: Rassenfosse, Kozak, Seliger 2019: Geocoding of worldwide patent data, available at https://doi.org/10.1038/s41597-019-0264-6

2.3. Office-specific IP data

Name	Intellectual Property Government Open Data (IPGOD)
Content	100 years of Intellectual Property (IP) rights administered by IP Australia comprising patents, trademarks, designs and plant breeder's rights.
Linking data element to PATSTAT	PATSTAT_APPLN_ID (PATSTAT application identifier) from table IPGOD - 102
Provider	IP Australia
Access	https://data.gov.au/dataset/intellectual-property-government-open-data-2017
Comment	

2.4. Indicators

Name	OECD Patent Quality Indicators Database
Content	Series of indicators capturing the technological and economic value of EPO and USPTO patents
Linking data element to PATSTAT	APPLN_ID – application identifier
Provider	OECD
Access	http://oecd/ipstats
Comment	Registration required (fill in an on-line form); Contact: sti.microdatalab@oecd.org

2.5. Trademarks

Name	TM-link
Content	As of 2017, TM-link includes trade mark data from Australia, Canada, the European Union, New Zealand, and the United States. More countries will be added in 2018.
Linking data element to PATSTAT	PERSON_ID – person identifier
Provider	IP Australia
Access	Developers only in beta version; send a request to chiefeconomist@ipaustralia.gov.au
Comment	In November 2017 a beta-version was available, with a full version planned for 2018

2.6. Other data sets

Name	OECD Triadic Patent Families Database
Content	Set of patents jointly filed at EPO, JPO and USPTO
Linking data element to PATSTAT	APPLN_ID – application identifier
Provider	OECD
Access	http://oecd/ipstats
Comment	Registration required (fill in an on-line form); Contact: sti.microdatalab@oecd.org

Name	OECD Citations Database
Content	Consolidated references to patent and non-patent literature cited in EPO, USPTO or PCT patents
Linking data element to PATSTAT	APPLN_ID – application identifier NPL_PUBLN_ID – NPL publication identifier (check the version of PATSTAT)
Provider	OECD
Access	http://oecd/ipstats
Comment	Registration required (fill in an on-line form); Contact: sti.microdatalab@oecd.org

Name	JRC-OECD COR&DIP© Database, v.1, 2017
Content	IP bundle of top corporate R&D investors
Linking data element to PATSTAT	APPLN_ID – application identifier INPADOC_FAMILY_ID – Patent family identifier (PATSTAT Autumn 2016)
Provider	OECD
Access	http://oecd/ipstats
Comment	Registration required (fill in an on-line form); Contact: sti.microdatalab@oecd.org

2.7. Reference tables

Name	IPC8 -Technology Concordance
Content	Concordance table between IPC (version 8) and 35 technical fields
Linking data element to PATSTAT	IPC_CLASS_SYMBOL – IPC classification symbol
Provider	WIPO
Access	http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/xls/ipc_technology.xls
Comment	This data is already included in the PATSTAT table TLS901_TECHN_FIELD_IPC

Name	IPC8 - NACE REV.2 (Version 2.0) Concordance
Content	This concordance table maps IPC sub classes / IPC main groups to the first 2-4 digits of the hierarchical NACE code.
Linking data element to PATSTAT	IPC_CLASS_SYMBOL – IPC classification symbol
Provider	Eurostat in cooperation with INCENTIM/ECOOM - KU Leuven
Access	https://circabc.europa.eu/d/a/workspace/SpacesStore/1d3a4661-f37d-4cbe-9998-8bd5966f1c62/CONCORDANCE%20IPC%20V8_NACE%20Rev%202%2020150630.xlsx
Comment	This data is already included in the PATSTAT table TLS902_IPC_NACE2