OECD REGPAT DATABASE

EPO & PCT Patent Applications at Regional Level

February 2022 edition

BACKGROUND INFORMATION

The OECD REGPAT database – February 2022 – presents patent data that have been linked to regions utilising the addresses of the applicants and inventors. The data were 'regionalised' at a very detailed level covering more than 5 500 regions across OECD countries, EU 28 countries, Brazil, China, India, the Russian Federation and South Africa. REGPAT allows patent data to be used in connection with other regional data such as GDP or labour force statistics, and other patent-based information such as citations, technical fields and patent holder's characteristics (industry, university, etc.), thus providing researchers with the means to develop a rich set of new indicators and undertake a broad range of analyses to address issues relating to the regional dimension of innovation.

By making regionalised patent data available to all researchers interested in the field, the OECD aims to stimulate research and contribute to a better understanding of the regional dimension of innovation. In addition, the methodology used for the construction of REGPAT is publicly available (see *Maraut et al., 2008*), to give users the opportunity to suggest modifications and thus contribute to improvements in the quality of REGPAT.

In general, a region is allocated according to the postal code identified in the address field of the patent record. Where postcodes are not available, regions were allocated according to town names - this required correct identification of a town in the address field and then reference to a detailed town/region lookup table. However, since towns can be difficult to identify in address fields and since the town/region lookup table may not be optimal for some countries, there are inevitably some erroneous allocations.

Any feedback on incorrectly allocated addresses (false positives) would be highly appreciated.

DATASET COVERAGE

The OECD REGPAT database, February 2022, fully derives the *European Patent Office's (EPO) Worldwide Statistical Patent Database* (PATSTAT Global, Autumn 2021), namely **PATSTAT Biblio** and **EP Register**. REGPAT covers:

- **Patent applications filed to the EPO** from 1977 (priority date)
- **Patent applications filed under the Patent Co-operation Treaty (PCT) at international phase**, from 1977 (priority date)

REGIONAL BREAKDOWN

All regions are defined within national borders and, in most cases, correspond to administrative regions. The regional breakdowns provided in REGPAT use the 2013 version of the Nomenclature of territorial units for statistics (**NUTS, 2013** *Eurostat*) for European countries (NUTS3), and are based on OECD's Territorial Level 3 (TL3) for other countries. For the United Kingdom, the NUTS3 levels refer to the 2010 classification for the London area. Data for the United States are presented at the level of the county, which can in turn be aggregated into 179 BEA Economic Areas (see table *REGPAT_REGIONS_US*).

Addresses for the countries listed in the table were regionalised using the REGPAT methodology. The REGPAT database also integrates data regionalised by the Katholieke Universiteit Leuven (KUL, INCENTIM/ECOOM) on behalf of Eurostat, as well as regional data provided in the PATSTAT database. Countries for which the addresses have not been regionalised are considered as one region.

Country		Micro region		Macro region		% success
AU	Australia	49	TL3	8	TL2	97%
AT	Austria	35	NUTS3	9	NUTS2	99%
BE	Belgium	44	NUTS3	3	NUTS2	98%
CA	Canada	294	TL3	13	TL2	98%
CL	Chile	54	TL3	15	TL2	97%
CZ	Czech Republic	14	NUTS3	8	NUTS2	99%
DK	Denmark	11	NUTS3	5	NUTS2	99%
EE	Estonia	5	NUTS3	1	NUTS2	99%
FI	Finland	19	NUTS3	5	NUTS2	99%
FR	France	101	NUTS3	27	NUTS2	100%
DE	Germany	402	NUTS3	38	NUTS2	97%
GR	Greece	52	NUTS3	13	NUTS2	98%
HU	Hungary	20	NUTS3	7	NUTS2	96%
IS	Iceland	8	NUTS3	2	NUTS2	88%
IE	Ireland	8	NUTS3	2	NUTS2	92%
IL	Israel	6	TL2	6	TL2	94%
IT	Italy	110	NUTS3	21	NUTS2	100%
JP	Japan	47	TL3	10	TL2	99%
KR	Korea	17	TL3	7	TL2	99%
LV	Latvia	6	NUTS3	1	NUTS2	98%
LU	Luxembourg	1	NUTS3	1	NUTS2	100%
MX	Mexico	209	TL3	32	TL2	95%
NL	Netherlands	40	NUTS3	12	NUTS2	99%
NZ	New Zealand	14	TL3	14	TL2	97%
NO	Norway	19	NUTS3	7	NUTS2	99%
PL	Poland	72	NUTS3	16	NUTS2	99%
PT	Portugal	25	NUTS3	7	NUTS2	99%
SK	Slovak Republic	8	NUTS3	4	NUTS2	99%
SI	Slovenia	12	NUTS3	2	NUTS2	97%
ES	Spain	59	NUTS3	19	NUTS2	99%
SE	Sweden	21	NUTS3	8	NUTS2	100%
CH	Switzerland	26	NUTS3	7	NUTS2	99%
TR	Turkey	81	NUTS3	26	NUTS2	98%
GB	United Kingdom	139	NUTS3	12	NUTS2	98%
US	United States	3144	County	51	TL2	99%
BR	Brazil			27	TL2	97%
BG	Bulgaria	28	NUTS3	6	NUTS2	96%
CN	China	35	TL3	34	TL3	98%
HR	Croatia	21	NUTS3	2	NUTS2	98%
IN	India	36	TL3	36	TL3	97%
LT	Lithuania	10	NUTS3	1	NUTS2	87%
MT	Malta	2	NUTS3	1	NUTS2	81%
RO	Romania	42	NUTS3	8	NUTS2	96%
RU	Russian Federation	83	TL3	83	TL3	98%
ZA	South Africa	9	TL3	9	TL3	97%

RESTRICTIONS, SOURCE & CONTACT

Please note that the REGPAT database is provided for research and analytical work. Make sure it is quoted as:

"OECD, REGPAT database, February 2022" when publishing the results of your analysis.

For further information about OECD patent related work, the methodology beyond REGPAT database and access to patent indicators, please read more at: oe.cd/ipstats.

Comments and questions about this dataset should be sent to STI.Microdatalab@oecd.org.

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

PRE-DEFINED INDICATORS

Statistics on patent at regional level are provided on the OECD statistics portal -OECD.Stat. Patent counts are broken down by regions (using OECD's territorial grids - TL3) for selected technology domains such as ICT, biotech, nanotech and some environment-related technologies.

REFERENCES

Eurostat (2011), Patent statistics at Eurostat: Methods for regionalisation, sector allocation and name harmonisation, *Eurostat Methodologies and Working Papers*.

Maraut S., H. Dernis, C. Webb, V. Spiezia and D. Guellec (2008), "The OECD REGPAT database: a presentation", *STI Working Paper 2008/2*, OECD, Paris

OECD (2009), *OECD Patent Statistics Manual,* OECD, Paris

DATABASE STRUCTURE

OECD REGPAT database is proposed as 2 sets of independent files that can be linked using the patent number or the application identifier, providing : the list of applicants and inventors for each patent, along with their addresses and regional codes; the patents filing and priority dates with the list of International Patent Classification (IPC) codes, and the list of codes from the Cooperative Patent Classification (CPC). It is also possible to identify the links between PCT applications and EPO patents using EPO_PCT table. Another correspondance table provides the regional code names at the lowest level (NUTS3/TL3) along with the upper regional level when available (the correspondance between county and TL3 levels is provided for the US). Data are presented as flat files: extension .txt using UTF 8 codepage and the pipe "|" character as a field delimiter.

PATENT APPLICATIONS TO THE EPO

Source: PATSTAT Biblio, Autumn 2021

EPO_APP_REG — List of EPO applicants — ***4,151,395 rows***

App_nbr	EPO application number (13 character format "EPYYYYNNNNNNN")
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2021
Pub_nbr	EPO patent publication number
Person_id	Surrogate key - person identifier in PATSTAT, Autumn 2021
App_name	Applicant's name
Address	Address
City	City name (as identified in the REGPAT process)
Postal_code	Postal code (as identified in the REGPAT process)
Reg_code	NUTS3/TL3 region code
Ctry_code	ISO 2 country code
Reg_share[2]	Multiple allocation to a region - Share ≤ 1
App_share[3]	Applicant's share ≤ 1

EPO_INV_REG — List of EPO inventors — ***10,230,273 rows***

App_nbr	EPO application number (13 character format "EPYYYYNNNNNNN")
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2021
Pub_nbr	EPO patent publication number
Person_id	Surrogate key - person identifier in PATSTAT, Autumn 2021
Inv_name	Inventor's name
Address	Address
City	City name (as identified in the REGPAT process)
Postal_code	Postal code (as identified in the REGPAT process)
Reg_code	NUTS3/TL3 region code
Ctry_code	ISO 2 country code
Reg_share[2]	Multiple allocation to a region - Share ≤ 1
Inv_share[3]	Inventor's share ≤ 1

EPO_IPC — List of IPC classes and selected dates — ***15,771,821 rows***

Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2021
Prio_year	Priority year (first filing)
App_year	EPO filing year
IPC[4]	List of IPC codes - latest edition

PCT-EPO Correspondance

EPO_PCT — PCT applications entering EPO regional phase — ***4,067,661 rows***

App_nbr	EPO application number
PCT_nbr	PCT Publication Number

REGION CODES

REGPAT_REGIONS — Description of regional codes

Ctry_Code	Country Code (ISO 2 characters)
Reg_Code	NUTS3 level code;TL3; or equivalent
Reg_Label	Micro-level region's name
Up_Level_Code	NUTS2 level code; TL2
Up_Level_Label	Macro-level region's name

REGPAT_REGIONS_US — County level concordance

Ctry_Code	Country Code (ISO 2 characters)
Reg_Code	County level
Reg_tl3	TL3 level code
Reg_tl3_label	TL3 level's region name

CPC Classes

CPC_CLASS — List of CPC classes related to the invention — ***53,979,472 rows***

Appln_id	Surrogate key - applications in PATSTAT, Autumn 2021
CPC_Class[5]	CPC codes

PATENT APPLICATIONS FILED UNDER THE PCT

Source: PATSTAT EP Register, Autumn 2021

PCT_APP_REG — List of PCT applicants — ***4,517,989 rows***

PCT_Nbr	PCT Publication Number ("WOYYYYNNNNNN")
internat_appln_nr	PCT Application Number - from PATSTAT Register
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2021
App_name	Applicant's name
Address	Address
City	City name (as identified in the REGPAT process)
Postal_code	Postal code (as identified in the REGPAT process)
Reg_code	NUTS3/TL3 region code
Ctry_code	ISO 2 country code
Reg_share[2]	Multiple allocation to a region - Share ≤ 1
App_share[3]	Applicant's share ≤ 1

PCT_INV_REG — List of PCT inventors — ***11,184,511 rows***

PCT_nbr	PCT Publication Number
internat_appln_nr	PCT Application Number - from PATSTAT Register
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2021
Inv_name	Inventor's name
Address	Address
City	City name (as identified in the REGPAT process)
Postal_code	Postal code (as identified in the REGPAT process)
Reg_code	NUTS3/TL3 region code
Ctry_code	ISO 2 country code
Reg_share[2]	Multiple allocation to a region - Share ≤ 1
Inv_share[3]	Inventor's share ≤ 1

PCT_IPC — List of IPC classes and selected dates — ***14,557,940 rows***

PCT_Nbr	PCT Publication Number
Prio_year	Priority year (first filing)
App_year	EPO filing year
IPC[4]	List of IPC codes - latest edition

1. Identifiers (surrogate keys) from PATSTAT, Autumn 2021.
2. Region share, when an address was allocated to more than one region – see methodological documentation (*Maraut et al., 2008*).
3. For fractional counts, when more than one applicant/inventor per patent, the applicant/inventor share (e.g. contribution) is provided.
4. IPC classes – data extracted from PATSTAT, Autumn 2021. Further information on the IPC are available at http://www.wipo.int/classifications/ipc/ipc8/?lang=en .
5. CPC classes - data extracted from PATSTAT, Autumn 2021. Further information on the CPC are available at http://worldwide.espacenet.com/classification?locale=en_EP . The variable CPC value is no longer available in REGPAT.