OECD REGPAT DATABASE

EPO & PCT Patent Applications at Regional Level

February 2016 edition

BACKGROUND INFORMATION

The OECD REGPAT database – February 2016 – presents patent data that have been linked to regions utilizing 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 published (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. For EU-28 countries, the REGPAT database integrates data previously regionalised by the Katholieke Universiteit Leuven (KUL, INCENTIM/ECOOM) on behalf of Eurostat.

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

DATASET COVERAGE

The OECD REGPAT database, February 2016, derives from two complementary sources of data: the *European Patent Office's (EPO) Worldwide Statistical Patent Database* (PATSTAT, Autumn 2015); and the OECD patent database that relies on *EPO's epoline® database* (EPO Bibliographic Database and Abstracts –EBD), covering all publications up to end November 2015. Two datasets are covered by REGPAT:

- **Patent applications filed to the EPO** from 1977 to 2012 – partial data from 2012 (*by priority date*)
- **Patent applications filed under the Patent Co-operation Treaty (PCT) at international phase**, from 1977 to 2013 - partial data from 2013 (*by priority date*).

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

REGIONAL BREAKDOWN

All regions are defined within national borders and, in most cases, correspond to administrative regions. The regional breakdowns provided in REGPAT correspond to the latest version of the Nomenclature of territorial units for statistics (**NUTS, 2010** *Eurostat*) for European countries (NUTS3), and are based on OECD's Territorial Level 3 (TL3) for other countries. Regional breakdowns for Brazil, China, India, the Russian Federation and South Africa refer to the highest administrative breakdowns (TL2). For Belgium, Greece, the Netherlands, NUTS2 level corresponds to TL3. For Australia, TL3 regions correspond to the new Australian Statistical Geography Standard (ASGS). The concordance to an intermediate micro-level at TL3 is provided for Germany (97 spatial planning regions) and the United States (179 BEA Economic Areas).

Addresses for the countries listed in the table were regionalised using REGPAT methodology. Thanks to several user contributions, data has been improved for some countries. Data for Turkey still require further improvement. Other 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	96.9
AT	Austria	35	NUTS3	9	NUTS2	98.6
BE	Belgium	44	NUTS3	11	NUTS2	98.3
BR	Brazil			27	TL2	97.5
BG	Bulgaria	28	NUTS3	6	NUTS2	92.0
CA	Canada	288	TL3	13	TL2	96.4
CL	Chile	53	TL3	15	TL2	95.3
CN	China			34	TL2	97.2
HR	Croatia	21	NUTS3	2	NUTS2	98.4
CZ	Czech Republic	14	NUTS3	8	NUTS2	98.6
DK	Denmark	11	NUTS3	5	NUTS2	98.5
EE	Estonia	5	NUTS3	1	NUTS2	99.1
FI	Finland	19	NUTS3	5	NUTS2	99.3
FR	France	100	NUTS3	26	NUTS2	99.3
DE	Germany	412	NUTS3	38	NUTS2	96.9
GR	Greece	51	NUTS3	13	NUTS2	98.3
HU	Hungary	20	NUTS3	7	NUTS2	98.2
IS	Iceland	8	NUTS3	2	NUTS2	87.7
IN	India			35	TL2	96.2
IE	Ireland	8	NUTS3	2	NUTS2	92.0
IL	Israel	15	TL3	6	TL2	94.5
IT	Italy	110	NUTS3	21	NUTS2	99.0
JP	Japan	47	TL3	10	TL2	98.8
KR	Korea	16	TL3	7	TL2	98.8
LV	Latvia	6	NUTS3	1	NUTS2	98.2
MT	Malta	2	NUTS3	1	NUTS2	82.8
MX	Mexico	209	TL3	32	TL2	93.0
NL	Netherlands	40	NUTS3	12	NUTS2	98.9
NZ	New Zealand	14	TL3	2	TL2	95.5
NO	Norway	19	NUTS3	7	NUTS2	99.1
PL	Poland	66	NUTS3	16	NUTS2	98.6
PT	Portugal	30	NUTS3	7	NUTS2	98.7
RO	Romania	42	NUTS3	8	NUTS2	90.4
RU	Russian Federation			83	TL2	97.4
SK	Slovak Republic	8	NUTS3	4	NUTS2	98.2
SI	Slovenia	12	NUTS3	2	NUTS2	96.6
ZA	South Africa			9	TL2	96.2
ES	Spain	59	NUTS3	19	NUTS2	99.3
SE	Sweden	21	NUTS3	8	NUTS2	99.6
CH	Switzerland	26	NUTS3	7	NUTS2	98.7
TR	Turkey	81	NUTS3	26	NUTS2	97.8
GB	United Kingdom	139	NUTS3	37	NUTS2	95.9
US	United States	3145	Counties	52	TL2	98.0

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 2016" 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.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

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 Y-codes from the Cooperative Patent Classification (CPC). It is also possible to identify PCT applications entering the EPO regional phase 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 NUTS3 and TL3 levels is provided for DE and 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, Autumn 2015

EPO_APP_REG — List of EPO applicants — ***3,123,893 rows***

App_nbr	EPO application number (13 character format "EPYYYYNNNNNNN")
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2015
Pub_nbr	EPO patent publication number
Person_id	Surrogate key - person identifier in PATSTAT, Autumn 2015
App_name	Applicant's name
Address	Address
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 — ***7,384,453 rows***

App_nbr	EPO application number (13 character format "EPYYYYNNNNNNN")
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2015
Pub_nbr	EPO patent publication number
Person_id	Surrogate key - person identifier in PATSTAT, Autumn 2015
Inv_name	Inventor's name
Address	Address
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 — ***11,646,893 rows***

Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2015
Prio_year	Priority year (first filing)
App_year	EPO filing year
IPC[4]	List of IPC classes - 8th edition

PCT-EPO Correspondance

EPO_PCT — PCT applications entering EPO regional phase — ***1,396,987 rows***

App_nbr	EPO application number
PCT_nbr	PCT Publication Number

REGION CODES

REGPAT_REGIONS *(inc. new AU codes)* **& REGPAT_REGIONS_NUTS3_TL3** *(DE / US only)* — 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

CPC Classes

CPC_CLASS_Y — Patents in the CPC class — ***1,625,658 rows***

Appln_id	Surrogate key - applications in PATSTAT, Autumn 2015
CPC_Class[5]	CPC classes Y only - e.g. specific technologies

PATENT APPLICATIONS FILED UNDER THE PCT

Source: OECD, Patent database, December 2015 update

PCT_APP_REG — List of PCT applicants — ***2,996,887 rows***

PCT_Nbr	PCT Publication Number ("WOYYYYNNNNNN")
PCT_App	PCT Application Number ("CCYYYYNNNNNN")
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2015
App_name	Applicant's name
Address	Address
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 — ***7,127,984 rows***

PCT_nbr	PCT Publication Number
PCT_App	PCT Application Number
Appln_id[1]	Surrogate key - applications in PATSTAT, Autumn 2015
Inv_name	Inventor's name
Address	Address
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 — ***9,476,322 rows***

PCT_Nbr	PCT Publication Number
Prio_year	Priority year (first filing)
App_year	EPO filing year
IPC[4]	List of IPC classes - 8th edition

1. Identifiers (surrogate keys) from PATSTAT, Autumn 2015.
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 2015. 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 2015. Further information on the CPC are available at http://worldwide.espacenet.com/classification?locale=en_EP