



**HAL**  
open science

# DOCUMENTATION OF RISIS DATASETS: RISIS Patent Database

Patricia Laurens

► **To cite this version:**

Patricia Laurens. DOCUMENTATION OF RISIS DATASETS: RISIS Patent Database. LISIS, Univ Gustave Eiffel, ESIEE Paris, CNRS, INRAE. 2022. <hal-03881028>

**HAL Id: hal-03881028**

**<https://hal.science/hal-03881028v1>**

Submitted on 1 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization





## Outline

1	Basic Characteristics .....	2
2	Database content .....	3
2.1	Definition and description of observations .....	3
2.2	Data acquisition and processing (e.g. data cleaning) .....	4
2.3	Information on all variables/indicators .....	6
2.4	Sectorial, temporal and geographical coverage .....	7
2.5	Quality and accuracy of data .....	16
3	Technical Specifications .....	21
3.1	Information on the data base system .....	21
3.2	Technical variable definition .....	21
3.3	Description of the Entity Relationship Model .....	25
3.4	Interfaces for access and to other infrastructures .....	27
4	Scientific use and main references .....	28
	Appendix .....	0
	Appendix 1: Artificial patents .....	1
	Appendix 2: Type of applicants .....	4
	Appendix 3: Classification of patent technologies .....	6
	Appendix 4: Building description for International Patent Classification .....	23
	Appendix 5: Data model of PATSTAT2021 .....	25



## 1 Basic Characteristics

The RISIS Patent database derives from the EPO PATSTAT. The database is designed for the analysis of technological knowledge creation, using patent as a proxy. It thus focuses on 'priority patents' that represent the creation of new knowledge, while other non-priority patents describe either technical ameliorations or market extensions and are mobilised as indicators of the importance of the priority patent. We shall explain in the documentation an important enrichment made to ensure a good coverage of the priority patents, that is the reintroduction in the database of so-called 'artificial priority patents' (11,7% of total priority patents).

Overall, 20,656,894 priority patents of inventions were applied for by legal entities from 2000 to 2020<sup>1</sup>. They include 59,150,156 inventors<sup>2</sup>.

As we are first interested in tracking knowledge creation, (more than the extension of the knowledge protection), the dates that matter, are the first filing date (when the organisation is applying for a priority patent, i.e., the first filing for a new invention) and the first publication date, as the time stamp of when the knowledge has been created is an important information for research.

Besides the date of the filing and the office where the patent was first filed, we specifically consider in RISIS Patent 5 core attributes on patents:

- **Their content, using textual pieces of information** such as the patent titles, the patent abstracts<sup>3</sup> and a text aggregating the definition of the patent IPC (International Patent Classification) codes. Based on the definition of some **75000 IPC classes** provided for by PATSTAT, the latter builds a rich vocabulary enriching widely the content available for semantic analyses.
- **Their technological content using the standard technology classification:** (IPC subclasses, aggregation of IPC codes) by technological domains, fields or subfields. Patent allocation in the different classifications is realised on a fractional count basis according to their IPC.
- **The geographical location of inventive activities.** As we are interested in the geography of knowledge creation, we focus on inventors' addresses (instead of using applicants' addresses which would more capture commercialisation). We shall explain the difficulties associated with this approach and the extensive developments made. These addresses are geocoded and associated to 'functional areas' (urban and rural) worldwide mobilizing the RISIS CORTEXT geocoding service.

---

<sup>1</sup> i.e. `ipr_type = PI` in PATSTAT

<sup>2</sup> Corresponding to **11,970,694** distinct `person_id`.

<sup>3</sup> Translation in English for titles and abstracts that remain in their native language will be included in a next release of RISIS Patent.

- **The legal organisations that apply for patents (the applicants).** We use both available (but tested and in part improved) types proposed by PATSTAT. But one central effort is to articulate it and mobilise the extensive work done by other RISIS databases to identify worldwide large firms (CIB), European fast growing mid-sized firms (Cheetah), European venture capital backed start-up firms (VICO) and European public research organisations (ORGREG covering universities, PROs and research hospitals). ORGREG entities as patent applicants are already recognized in this new updated RPD version. The link with CIB, VICO and CHEETAH databases will be implemented step by step in 2023.

The complete methodology used to build the RPD database is detailed in [https://gitlab.com/cortex/patents/patstat/-/tree/version-2021/version\\_2021](https://gitlab.com/cortex/patents/patstat/-/tree/version-2021/version_2021)

- **Characteristics linked to the ‘value’ of the patent:** In this context, the database provides information on whether the patent was granted or not, as well as the size of patent families, i.e. the number of different applications for a given invention (or group of linked inventions), which tells something about the interest of applicants for developing and protecting the knowledge; moreover, the presence in 5 world-level patent offices (EPO, USPTO, Japan Patent Office, Korean Patent Office, China Patent Office) tells something about the potential for future markets; and citations they have received indicate knowledge flows to other inventions and therefore its value for further technological development. .

The coverage of the database is 2000-2020 (using the PATSTAT 2021 Edition April as a source) but the PATSTAT coverage for patent applied in 2020 is very limited. The first release is in November 2022. Further releases including the tagging of Tirmreg entities (CIB, CHEETAH, VICO) as applicants will be displayed in 2023. The database will be only periodically updated using new PATSTAT version. The database is developed and maintained by UGE. The database is located on the servers of UGE at Marne La Vallee, France.

## 2 Database content

### 2.1 Definition and description of observations

As a proxy for knowledge creation, the unit of observation is a priority patent application, i.e. the very first patent application, anywhere in the world to protect an invention. The priority date is used to determine the novelty of the invention, which implies that it is an important concept in patent procedures. For statistical purposes, the priority date is the closest date to the date of invention.

In RPD, we consider only applications of priority patents of invention (i.e. `ipr_type = PI` in PATSTAT)<sup>4</sup>. Accordingly, the database covers priority patents applied since 2000, whether or not they turn later into granted patents. The 20,656,894 priority patents are distributed across three main types of applications:

---

<sup>4</sup> Other types of patent are: Utility Models (`ipr_type = 'UM'`) and Design patents (`ipr_type = 'DP'`).

- a large category of patents 'A', i.e. 'regular' patents following the Paris convention (appln\_kind = 'A'): 19,198,231 applications,
- provisional patent, i.e. preliminary and temporary patents proposed in some patent offices like USPTO (appln\_kind = 'P'): 1,153,226 applications,
- PCT patent in the international phase (appln\_kind = 'W'): 305,437 applications.

We also include priority patents tagged as 'artificial priority patents due to priorities' in PATSTAT (11,7% of the overall priorities in RPD but 50,6% in the priorities applied at the US patent office and 38,7% the priorities applied at the EPO). As these latter documents only contain the filing date and patent office in PATSTAT, we have devised a complex method to fill the information on their applicants, inventors, title, IPC classes. (See Appendix 1 for the description of technical developments and share of applications by patent offices and Laurens et al., 2018<sup>5</sup> for the importance of such an enrichment).

As we are interested in the development of organised knowledge, we focus only on patents taken by organisations, i.e., legal persons (firms, university, NGO...); therefore, the priority patents applied only by natural persons have been discarded.

Moreover, when a co-application of patent includes both legal and natural persons, the patent is maintained in the database with its legal applicant only.

In order to do so, we have first devised a method to recognise applicants that are 'natural persons', and we take them out from the database.

A fractional counting was employed to calculate the number of legal applicants in a patent applications: All applicants with a `fract_appl` below 1 have legal co-applicants.

## 2.2 Data acquisition and processing (e.g. data cleaning)

### a) Data sources

The RISIS Patent database uses four sources of information. It builds first and mainly on the PATSTAT public database from EPO (version April 2021). Secondary, sources are used for complementing this main data source, mainly for filling addresses that are void in PATSTAT:

- REGPAT data base (provided freely on request) from OECD that presents patent data that have been linked to regions according to the addresses of the applicants and inventors,
- Data from the French patent office (INPI) for filling addresses in patent applications from INPI that are missing in PATSTAT,

---

<sup>5</sup> Laurens et al., 2018, "The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?" *Scientometrics* 114,1, 91-112.

- Data from the Japanese patent office (JPO) for filling addresses in patent applications from JPO that are missing in PATSTAT.

## b) Data processing

The principal data processing includes:

### 1- Data enrichment steps of the initial raw patent data

- Filling missing addresses, missing information on IPC (international patent classification) and on inventors and applicants using the REGPAT, INPI and JPO data (see above),
- Re-allocation of addresses' information contained in other fields (person'' names)
- Propagating information (using fuzzy matching) within patent families for missing info on applicants and inventors,
- Propagating information for filling information in artificial applications using other applications within patent families,

### 2- Addition of complementary information

- Identifying priority patent applications, the artificial priority applications, singletons (i.e. lone patent applications in an Inpadoc family), transnational patents (applied in at least 2 patent offices), triadic patents (applied in at JP, EP, US), ip5 family patents (applied in at least 1 IP5 patent office), 2ip5 family patents (applied in at least 2 IP5 patent offices) and ip5 patents (applied in the 5 IP5 patent offices),
- Identifying the type of applicants (legal/person/unknown),
- Extraction, cleaning and harmonisation of country information for applicants and inventors by analysing respectively applicants' and inventors' addresses,
- Addition of geographical classification on continent for applicants and inventors,
- Geocoding of addresses (applicants and inventors),
- Attribution of the urban area for the address based on geocoded data,
- Addition of information on technology classification at three levels (5 domains, technology fields according to WIPO classification in 35 categories and technology subfields according to IFRIS classification in 400 categories),
- Addition of the IPC definition (full textual information),
- Extraction of year of application from the date of application,
- Addition of tags characterising the patent applications with no inventors, the patent applications with no IPC,
- Fractional counting at the patent application level according to different dimensions: technological classifications, number of applicants, number of inventors,
- Calculation of the priority patent geographical coverage (based on patent families): number of patent offices, applications in the IP5 patent offices (EPO, USPTO, JPO, SINO, KINO), number of patents in the docdb and inpadoc families of the priority applications.
- Assignment of an ORGREG-ID for applicants belonging the ORGREG repository (using ORGREG data available in winter 2021).

## 3- Filtering of patents

We select only priority patents of inventions applied from 2000 to 2020 with at least one legal applicant (including artificial priority applications).

## 4- Creation of the final tables

The final tables includes data on priority patents only and covering 6 dimensions: Basic patent data, Geographical data on inventors and applicants, Institutional data on applicants, Technological data based on classification, Textual data, Patent value data.

### 2.3 Information on all variables/indicators

For each priority patent application, the database gives:

- patent ID number
- date of first filing
- country of first filing
- date of first publication
- date of first granting
- title
- abstract
- IPC categories mentioned: their number and their language description: as many variables as IPC categories, only one linguistic description
- whether the patent is a singleton (only one single application) or not
- whether the patent is a transnational patent, i.e., was the priority application extended in at least a foreign country) or not
- size of the DOCDB family
- size of the INPADOC family
- presence (directly or through extension) in the 5 core offices (so called IP 5 families) (5 variables Y/N per office)
- For each applicant: the applicant 'natural name' and its ID; the applicant standardised name and its ID; the applicant standardised name by Leuven and its ID; the applicant RISIS standardised name and its ID when available
- For each inventor: same variables as for applicants
- For each applicant: presence or not in the CIB database (next release), name and ID of the group firm (GUO) (next release), presence or not in the CHEETAH

database(next release), ID of the Cheetah firm (next release), presence or not in the VICO database and ID of the VICO firm(next release); presence in firmreg database and Firmreg ID. (to be provided in 2023) (next release), presence or not in the ORGREG repository and Orgreg\_id

- For each applicant: The applicant address, the applicant geo coordinates, the applicant urban and rural cluster it belongs to, the NUTS3 regions for EU countries
- For each inventor: the inventor address, the inventor geo coordinates, the inventor urban or rural cluster he/she belongs to, the NUTS3 regions for EU countries

## 2.4 Sectorial, temporal and geographical coverage

### - Sectorial coverage

#### *Sectors of the legal applicants*

In the current RISIS Patent database, we rely on the sectorial information provided in the raw PATSTAT database<sup>6</sup>. Data are shown in the table below.

Table 1: Industrial sectors of priority patent applicants

psn_sector	Number of applicants	Share of applicants
NULL	33	0,0%
COMPANY	18,284,645	81,0%
COMPANY GOV NON-PROFIT	41,637	0,2%
COMPANY GOV NON-PROFIT UNIVERSITY	1,460	0,0%
COMPANY HOSPITAL	1,636	0,0%
COMPANY UNIVERSITY	4,164	0,0%
GOV NON-PROFIT	701,141	3,1%
GOV NON-PROFIT HOSPITAL	2	0,0%
GOV NON-PROFIT UNIVERSITY	201,021	0,9%
HOSPITAL	36,288	0,2%
INDIVIDUAL	511,261	2,3%
UNIVERSITY	2,361,015	10,5%
UNKNOWN	423,876	1,9%
Total	22,568,179	100,0%

For about 96% of the applicants, our identification of the applicant status (legal vs individual) follows the sector (legal/individual/unknown) assigned in Patstat. However, we

<sup>6</sup> Harmonizing names and allocation of assignee sectors in Patstat raw data was done by ECOOM (K.U. LEUVEN; <http://www.ecoom.be/en/EEE-PPAT>).

have discarded about 0.6% of the applicants assigned as legal in Patstat (and considered as individual according to our treatment and further checking steps). Conversely, we have included 2.3% of applicants considered as Individuals in Patstat. In this set, we have detected company names and we need further checking.

Moreover, our own matching of ORGREG entities with patent applicant names in RPD has shown that among the 22.5millions of applicants, 0,79% are ORGREG entities. They have applied for 156,132 patents (0,76% of RPD priority patents).

### *Technological coverage of patents*

There is a thematic coverage of the technology that the patents protect linked to the IPC classification of patents (one patent can belong to multiple classes). Patents are allocated to fields of technology on a fractional count basis according to their IPC. Building on the correspondence table developed by ISI FhG for WIPO - which defines domains (5) and fields (35) of technology, an additional level of sub fields (401) of technology was developed (See Appendix 3 on technological fields).

The distributions of patents with the domains and the technological fields are shown below.

Table 1: Domains of technology of priority patent applications

Domain code	Domain name	Number of patents (fractional counting)	Share of patents
TD01	Electrical engineering	6,605,360	32,6%
TD02	Instruments	2,996,713	14,8%
TD03	Chemistry	4,399,498	21,7%
TD04	Mechanical engineering	4,620,405	22,8%
TD05	Other fields	1,619,730	8,0%
Total		20,241,706	100,0%

Table 2: Technology fields of priority patent applications

Field code	Field_name	Number of patents (fractional counting)	Share of patents
TF01	Electrical machinery, apparatus, energy	1,460,025	7,2%
TF02	Audio-visual technology	745,623	3,7%
TF03	Telecommunications	550,958	2,7%
TF04	Digital communication	870,580	4,3%
TF05	Basic communication processes	154,328	0,8%
TF06	Computer technology	1,696,641	8,4%
TF07	IT methods for management	393,579	1,9%
TF08	Semiconductors	733,624	3,6%
TF09	Optics	697,791	3,4%
TF10	Measurement	1,136,741	5,6%

TF11	Analysis of biological materials	106,365	0,5%
TF12	Control	439,344	2,2%
TF13	Medical technology	616,473	3,0%
TF14	Organic fine chemistry	357,053	1,8%
TF15	Biotechnology	301,112	1,5%
TF16	Pharmaceuticals	582,232	2,9%
TF17	Macromolecular chemistry, polymers	387,679	1,9%
TF18	Food chemistry	390,867	1,9%
TF19	Basic materials chemistry	535,738	2,6%
TF20	Materials, metallurgy	546,088	2,7%
TF21	Surface technology, coating	352,656	1,7%
TF22	Micro-structural and nano-technology	37,708	0,2%
TF23	Chemical engineering	519,153	2,6%
TF24	Environmental technology	389,211	1,9%
TF25	Handling	603,590	3,0%
TF26	Machine tools	677,266	3,3%
TF27	Engines, pumps, turbines	463,621	2,3%
TF28	Textile and paper machines	405,759	2,0%
TF29	Other special machines	672,267	3,3%
TF30	Thermal processes and apparatus	388,143	1,9%
TF31	Mechanical elements	553,279	2,7%
TF32	Transport	856,482	4,2%
TF33	Furniture, games	528,471	2,6%
TF34	Other consumer goods	332,163	1,6%
TF35	Civil engineering	759,096	3,8%
Total		20,241,706	100,0%

## - Temporal coverage

The database of this first release covers patent applications from 2000 to 2020 (based upon PATSTAT2021 Version Spring). It is most likely that years 2019 and 2020 are still only partially filled. It includes 20,241,706 applications of priority patents

The number of PI priority patents applied every year is shown below. It has increased by 170% during the 18 years. Year 2020 is poorly completed.

Table 3: Number of priority patent applications over time

Filing year	Number of PI priority patent applications
2000	670,246
2001	678,265
2002	665,576
2003	680,682

2004	693,834
2005	730,992
2006	746,253
2007	765,074
2008	793,153
2009	775,383
2010	833,730
2011	906,741
2012	1,042,377
2013	1,160,214
2014	1,223,779
2015	1,339,724
2016	1,526,923
2017	1,716,362
2018	1,833,727
2019	(1,341,263) <sup>1</sup>
2020	(532,596) <sup>1</sup>
Total	20,656,894

## - Geographical coverage

### *Geography of the PI protection*

The data cover all priority patent applications worldwide, i.e. at all regional and national offices in the world (see table below). More than 50% of the priority applications are applied for at the Japanese or Chinese ones. The IP5 patent offices (US, EP, JP, CN, KR) cumulate together 86,2% of the applications.

Table 4: Number of priority patent applications and share of artificial priorities according to the patent offices

Patent office	Number of PI patent applications	Distribution of PI patent applications	Share of artificial priority patents
CN	8,034,493	38,9%	0,9%
JP	5,214,992	25,2%	11,7%
US	2,534,379	12,3%	50,6%
KR	1,696,261	8,2%	4,8%
DE	726,530	3,5%	10,2%
EP	328,008	1,6%	38,7%
GB	305,440	1,5%	0,5%
WO	305,437	1,5%	5,2%
RU	287,777	1,4%	0,5%
TW	270,273	1,3%	1,9%

FR	223,880	1,1%	2,0%
IT	114,231	0,6%	9,8%
AU	79,318	0,4%	44,1%
PL	45,934	0,2%	0,8%
BR	42,482	0,2%	1,3%
IN	36,576	0,2%	87,6%
SE	35,180	0,2%	20,4%
ES	29,644	0,1%	5,7%
NL	29,109	0,1%	10,3%
FI	24,594	0,1%	7,9%
UA	22,977	0,1%	2,9%
TR	22,840	0,1%	24,2%
Others	246,539	12,0%	
Total	20,656,894	100%	11,7%

After a first priority patent application for a new invention, the IP protection can be further extended in several geographical countries considered as future market. The following tables give information on the geography of the patent protection of a given priority patent using data available in its Inpadoc family. Most of the patents, for a given invention are applied in a single patent office and only 25% of the priority patents are transnational, i.e. further extended in another patent office (20% from 2 to 5 patent offices).

Table 6: Number of patent offices where patents are applied for a given invention

Number of distinct patent offices in the inpadoc family	Number of priority patents	Share of priority patents
1	15,500,510	75,0%
2	1,742,758	8,4%
3	1,000,989	4,8%
4	770,614	3,7%
5	583,166	2,8%
More than 5	1,642,023	5,1%
Total	20,656,894	100,0%

Almost 50% of the inventions are protected in China (either in the first (39%) or secondary (20%) subsequent filings), 28% in Japan, 24% in US, 12% at EPO, 12% in Korea. 16% of the families include a PCT patent<sup>7</sup>. Respectively 23,8% and 17,3% of the patents were applied (first filling or further geographical extension) in at least 1 or 2 IP5 patent offices. Only 1.3% of them have been applied in the 5 IP5 patent offices. Only 37% of the patents are granted (either because they were rejected or because the granting decision is still pending).

<sup>7</sup> The Patent Cooperation Treaty (PCT) provides a unified procedure for filing patent applications to protect inventions in each of its contracting states. A patent application filed under the PCT is called an international application, or PCT application.

Table 7: Number and share of patents (priority or secondary applications) applied in the five largest patent offices (IP5 patent offices) and other offices

Content of Inpadoc family <sup>8</sup>	Number of priority patents	Share of priority patents
Total	20,656,894	100%
US application	4,894,036	23,7%
EP application	2,526,412	12,2%
JP application	5,713,216	27,7%
KR application	2,472,223	12,0%
CN application	10,145,321	49,1%
W application	3,312,187	16,0%
Transnational patent	5,156,384	25,%
Triadic	719,293	3,5%
Ip5	273 ,801	1,3%
2ip5 family	3,567,686	17,3%
Ip5 family	4,912,441	23,8%
Granted	7,673,581	37,1%

### *Geography of the inventions*

In 80% of the patents, the information of the inventors' geographical location is given (country assignment); in 75% of the patents, the information of the applicants' geographical location is given (country assignment).

The shares of patents, inventors' addresses and applicants' addresses according to the inventors' and applicants' addresses are shown in the tables below.

Among the 59 million inventors, the geographical location (country assignment) was identified for 83% of them and for 61% of them, the addresses were geocoded and assigned to an URA (Urban and Rural Area). Among the 22 million applicants, the geographical location (country assignment) was identified for 76% of them and for 63% of them, the addresses were geocoded and assigned to an URA.

Almost 30% of RPD priority patents have at least one inventor located in China. Then come Japanese inventors (27% of the patents) and US inventors (10% of the patents). More than 65% of the patents have at least one Asian inventor and more than 75% of the overall set of inventors are located in Asia.

Table 8: Number of priority patent applications according to the country of inventors

Country of inventors	Number of patent applications with inventor from the country	Share of patent applications with inventor from the country
CN	6,049,743	29,3%
JP	5,545,334	26,8%

<sup>8</sup> Only ipr\_type : PI is considered in inpadoc family.

Unknown	4,135,671	20,0%
US	2,162,462	10,5%
KR	1,706,754	8,3%
DE	885,923	4,3%
TW	433,187	2,1%
FR	310,509	1,5%
RU	251,011	1,2%
GB	159,517	0,8%
IT	140,428	0,7%
CA	135,854	0,7%
NL	110,184	0,5%
CH	96,095	0,5%
IN	90,057	0,4%
SE	86,902	0,4%
AU	64,210	0,3%
IL	59,691	0,3%
ES	54,321	0,3%
PL	51,518	0,2%
FI	49,644	0,2%
BR	48,115	0,2%
AT	46,379	0,2%
BE	46,364	0,2%
HK	44,415	0,2%
DK	41,162	0,2%
SG	33,929	0,2%
TR	26,860	0,1%
UA	25,654	0,1%
NO	19,376	0,1%
Total number of patents	20,656,894	100%

Table 9: Number and share of inventors according to the country of inventors. Share of inventors allocated to URAs and NUTS3 regions

Country of inventors	Number of inventors	Share of inventors	Cumulated share	Share of addresses allocated to URA	Share of addresses allocated to NUTS3 (EU countries)
CN	18,512,361	31,3%	31,3%	60,0%	-
JP	12,647,940	21,4%	52,7%	67,8%	-
Unknown	9,831,894	16,6%	69,3%	0,0%	-
US	5,761,034	9,7%	79,0%	98,2%	-
KR	4,349,848	7,4%	86,4%	86,2%	-

DE	2,116,297	3,6%	90,0%	97,1%	96,9%
RU	901,027	1,5%	91,5%	94,7%	-
TW	898,469	1,5%	93,0%	92,4%	-
FR	658,218	1,1%	94,1%	83,0%	83,0%
GB	301,650	0,5%	94,6%	93,1%	92,9%
CA	295,339	0,5%	95,1%	94,3%	-
IT	262,756	0,4%	95,6%	84,6%	79,4%
IN	236,465	0,4%	96,0%	95,3%	-
NL	236,404	0,4%	96,4%	98,2%	98,2%
CH	190,423	0,3%	96,7%	96,3%	96,2%
SE	173,977	0,3%	97,0%	84,1%	83,8%
IL	146,473	0,2%	97,2%	85,6%	-
PL	143,842	0,2%	97,5%	98,4%	98,0%
ES	138,949	0,2%	97,7%	87,4%	87,3%
AU	120,053	0,2%	97,9%	80,1%	-
BR	117,344	0,2%	98,1%	64,1%	-
FI	107,527	0,2%	98,3%	97,5%	97,3%
UA	93,483	0,2%	98,5%	90,8%	-
BE	90,787	0,2%	98,6%	97,8%	97,6%
AT	89,993	0,2%	98,8%	96,7%	96,6%
DK	89,354	0,2%	98,9%	90,4%	89,6%
SG	70,330	0,1%	99,0%	95,6%	-
RoW	567,919	1,0%	100,0%		-
Total number of inventors	59,150,156	100,0%		61,7%	

Table 10: Number of priority patent applications according to the country of applicants

Country of applicants	Number of patent applications with applicant from the country	Share of patent applications with applicant from the country
JP	5,810,690	28,1%
Unknown	5,119,587	24,8%
CN	3,327,160	16,1%
US	2,041,026	9,9%
KR	1,581,758	7,7%
DE	851,613	4,1%
TW	423,951	2,1%
FR	275,960	1,3%
RU	240,333	1,2%
GB	226,183	1,1%
NL	117,804	0,6%

CH	115,186	0,6%
IT	101,487	0,5%
SE	91,825	0,4%
CA	88,490	0,4%
AU	61,093	0,3%
FI	53,925	0,3%
PL	46,200	0,2%
BR	42,494	0,2%
ES	42,315	0,2%
IL	41,897	0,2%
DK	37,504	0,2%
BE	37,401	0,2%
AT	36,398	0,2%
SG	30,406	0,1%
IN	29,685	0,1%
TR	25,421	0,1%
UA	23,055	0,1%
HK	21,719	0,1%
NO	18,057	0,1%
Total number of patents	20,656,894	100%

Table 11: Number and share of applicants according to the country of the applicants.  
Share of applicants allocated to URAs and NUTS3 regions

Country of applicants	Number of applicants	Share of applicants	Cumulated share	Share of addresses allocated to URA	Share of addresses allocated to NUTS3
JP	6,283,479	27,8%	27,8%	80,2%	-
Unknown	5,367,580	23,8%	51,6%		-
CN	3,642,558	16,1%	67,8%	89,1%	-
US	2,209,343	9,8%	77,6%	98,6%	-
KR	1,667,166	7,4%	84,9%	67,8%	-
DE	879,819	3,9%	88,8%	89,0%	88,9%
TW	438,804	1,9%	90,8%	71,8%	-
FR	299,716	1,3%	92,1%	78,3%	78,1%
RU	254,707	1,1%	93,2%	35,3%	-
GB	238,296	1,1%	94,3%	77,9%	76,0%
NL	122,247	0,5%	94,8%	91,6%	91,6%
CH	118,775	0,5%	95,4%	94,2%	94,1%
IT	105,883	0,5%	95,8%	90,5%	90,4%
SE	94,336	0,4%	96,3%	92,5%	92,5%
CA	94,151	0,4%	96,7%	87,5%	-

AU	64,764	0,3%	97,0%	73,6%	-
FI	56,903	0,3%	97,2%	97,9%	97,9%
PL	49,805	0,2%	97,4%	64,8%	64,1%
BR	48,829	0,2%	97,6%	52,7%	-
ES	47,533	0,2%	97,9%	78,5%	78,4%
IL	47,218	0,2%	98,1%	85,8%	-
BE	40,015	0,2%	98,2%	91,4%	91,2%
DK	38,867	0,2%	98,4%	92,0%	91,4%
AT	37,353	0,2%	98,6%	87,8%	87,8%
SG	32,457	0,1%	98,7%	96,7%	-
IN	31,466	0,1%	98,9%	90,5%	-
TR	26,739	0,1%	99,0%	51,3%	-
RoW	229,370	1,0%	100,0%		
Total number of applicants	22,568,179	100,0%		63,1%	

## 2.5 Quality and accuracy of data

Checking the data quality is not carried out on the initial PATSTAT raw data but on the further improvements done at LISIS to produce RISIS Patent. It includes tests of completeness and consistency.

### Tests of completeness

#### 1- Patent applications without any information on applicants (all applicants)

All patent applications of RPD have at least one applicant. If there were patents without any information on at least one applicant in the initial database and if no information was completed during the enrichment steps, such priority patents were not included in RPD

#### 2- Patent applications with legal applicants without any information on inventors nor IPC codes

Table 12: Share of priority patent applications with legal applicants without any information on inventors or IPC codes according to the patent office for the 25<sup>th</sup> largest patent offices

Patent office	Number of patent applications	Share of patent applications with at least one inventor	Share of patent applications with ipc information
Total	20,656,894	98,40%	98,20%
CN	8,034,493	99,97%	99,75%
JP	5,214,992	99,99%	99,83%
US	2,534,379	99,90%	99,53%
KR	1,696,261	99,99%	99,24%
DE	726,530	97,11%	99,95%

EP	328,008	97,68%	99,53%
GB	305,440	25,76%	35,68%
WO	305,437	99,81%	99,34%
RU	287,777	99,04%	99,69%
TW	270,273	97,01%	98,59%
FR	223,880	99,34%	99,73%
IT	114,231	98,67%	44,44%
AU	79,318	71,82%	78,23%
PL	45,934	99,74%	99,45%
BR	42,482	99,71%	98,89%
IN	36,576	99,55%	98,50%
SE	35,180	98,69%	98,07%
ES	29,644	99,75%	99,81%
NL	29,109	99,69%	99,80%
FI	24,594	98,18%	96,36%
UA	22,977	98,08%	92,21%
TR	22,840	98,50%	94,11%
AT	19,793	54,15%	96,72%
CA	19,274	97,32%	99,39%
CH	18,959	98,83%	99,80%
DK	18,044	99,29%	99,76%

\* appln\_id not in table rpd\_inventors

\*\* appln\_id not in table rpd\_technological classification

### 3- Patent applications with missing country of the inventors' or missing geocoded address of inventors

For 2/3 of the inventors, the country is identified either in the priority patent, a subsequent patent of the same Inpadoc family in Patstat or other patent resources (step 1). For 0,03% of the inventors, the country can be allocated after the geocoding and further geographical allocation steps using Cortext facilities (situation where the initial inventors' addresses include an address but where the country information cannot be isolated because it is missing or not recognized during the automated process). When still missing, the country of the applicant is allocated to the inventor' when the patent is also applied in the country of the applicant (16% of the inventors' country allocation).

The allocation of geographical information (country or URA) to patent inventors drastically varies according to the patent offices. The completed assignment of all the patent inventors is satisfying in most of the EU countries (except the United Kingdom) and the US but cannot be properly done for patents applied in China, Japan or Brazil.

Selecting priority patents with further subsequent extensions allows to significantly reduce the share of patents where inventors' allocation is missing. The share of unassigned

countries among the inventors drops from 16% to 1% when selecting transnat or 2ip5 priority patents; the share of unassigned URA among the inventors decreases from 38% to 12%.

Table 13: Assignment of the country of inventors: numbers and shares

Source of the inventors' country assignment	Number of inventors' addresses	Share of inventors' addresses
Initial Patstat data (priority or inpadoc family) or other data sources (step 1)	39,958,215	67,55%
Country allocation based on geocoded information (step 2)	18,339	0,03%
Country assigned based on applicant and patent office data (step 3)	9,341,708	15,79%
Unknown	9,831,894	16,62%

Table 14: Share of priority patent applications with information on the country of the inventor or with geocoded address of inventors according to the patent office for the 25<sup>th</sup> largest patent offices

Patent office	Number of patent applications	Share of patents without any information of the inventors*	Share of patents with unassigned countries among the inventors' addresses (after step1)**	Share of patents with unassigned countries among the inventors' addresses (after step3)**	Share of patents with unassigned URA among the inventors' addresses**	Total share of patents with unassigned countries among inventors' addresses**	Total share of patents with unassigned URA among inventors' addresses**
Total	20,656,894	1,6%	36,6%	15,8%	47,8%	17,4%	49,4%
CN	8,034,493	0,0%	78,4%	49,3%	78,9%	49,3%	78,9%
JP	5,214,992	0,0%	15,6%	0,4%	48,1%	0,4%	48,1%
US	2,534,379	0,1%	1,0%	0,4%	6,9%	0,5%	7,0%
KR	1,696,261	0,0%	13,0%	2,2%	23,8%	2,2%	23,8%
DE	726,53	2,9%	0,2%	0,1%	6,1%	3,0%	9,0%
EP	328,008	2,3%	0,3%	0,3%	9,7%	2,6%	12,0%
GB	305,44	74,2%	0,2%	0,1%	1,7%	74,3%	75,9%
WO	305,437	0,2%	0,9%	0,9%	6,1%	1,1%	6,3%
RU	287,777	1,0%	17,5%	15,1%	20,2%	16,1%	21,1%
TW	270,273	3,0%	0,4%	0,2%	14,6%	3,2%	17,6%
FR	223,88	0,7%	0,3%	0,0%	34,0%	0,7%	34,7%
IT	114,231	1,3%	38,3%	19,4%	41,4%	20,7%	42,7%
AU	79,318	28,2%	34,9%	15,0%	35,9%	43,2%	64,0%
PL	45,934	0,3%	0,7%	0,5%	3,6%	0,8%	3,9%
BR	42,482	0,3%	46,5%	3,6%	54,5%	3,9%	54,8%

IN	36,576	0,4%	4,1%	2,7%	13,3%	3,1%	13,7%
SE	35,18	1,3%	26,1%	3,7%	38,0%	5,0%	39,3%
ES	29,644	0,3%	11,5%	0,6%	18,5%	0,9%	18,8%
NL	29,109	0,3%	8,0%	3,8%	10,2%	4,1%	10,6%
FI	24,594	1,8%	0,3%	0,1%	5,4%	1,9%	7,2%
UA	22,977	1,9%	9,8%	1,9%	10,5%	3,8%	12,4%
TR	22,84	1,5%	0,6%	0,5%	2,9%	2,0%	4,4%
AT	19,793	45,9%	0,6%	0,2%	3,3%	46,1%	49,1%
CA	19,274	2,7%	0,9%	0,3%	13,2%	3,0%	15,8%
CH	18,959	1,2%	0,8%	0,3%	7,8%	1,5%	9,0%

\* appln\_id not in table rpd\_inventors

\*\* appln\_id present in rpd\_inventors but none or only partial information is available to assign a country/URA to each of the patent inventors

Table 15: Share of priority patent applications with information on the country of the inventor or with geocoded address of inventors according the patent value

Patent value	Number of patents	Share of appln_id without any inventor*	Share of patents with at least one inventor without assigned country**	Total share of patents with missing information on inventors' country	Total share of patents with missing information on inventors' URA
Transnat	5,156,384	2,5%	1,5%	4,0%	32,8%
2ip5	3,567,686	2,0%	1,3%	3,3%	35,3%
All	20,656,894	1,6%	20,0%	21,6%	38,4%

\* appln\_id not in table rpd\_inventors

\*\* appln\_id present in rpd\_inventors but none or only partial information is available to assign a country/URA to each of the patent inventors

Table 16: Share of inventors' addresses with information on the country of the inventor or with geocoded address of inventors according the patent value

Patent value	Number of inventors	Share of inventors without assigned country	Share of inventors without assigned URA
Transnat	14,325,880	1,2%	11,8%
2ip5	10,201,420	1,1%	12,4%
All	59,150,156	16,6%	38,3%

4- Patent applications with legal applicants with missing country of the applicants' or missing geocoded address of applicants

1/4 of the patents have unassigned country or unassigned URA for the applicants. These shares significantly vary according to the patent office.

Table 17: Share of priority patent applications with legal applicants with information on the country of the applicant or with geocoded address of applicant according to the patent office for the 25<sup>th</sup> largest patent offices

Patent office	Number of patent applications	Share of patents with missing country information among the applicants' addresses	Share of patents with missing URA information among the applicants' addresses
Total	20,656,894	24,8%	38,4%
CN	8,034,493	59,7%	64,8%
JP	5,214,992	0,5%	2,6%
US	2,534,379	2,8%	7,7%
KR	1,696,261	1,3%	12,1%
DE	726,530	0,1%	6,0%
EP	328,008	0,1%	280,2%
GB	305,440	29,8%	222,8%
WO	305,437	0,2%	11,2%
RU	287,777	16,3%	2,6%
TW	270,273	0,4%	1,6%
FR	223,880	0,0%	1,4%
IT	114,231	29,0%	3,6%
AU	79,318	19,5%	28,4%
PL	45,934	0,6%	19,6%
BR	42,482	0,1%	259,2%
IN	36,576	1,3%	6,1%
SE	35,180	0,1%	75,7%
ES	29,644	2,1%	6,0%
NL	29,109	4,3%	30,2%
FI	24,594	0,1%	366,2%
UA	22,977	8,2%	28,7%
TR	22,840	0,7%	13,4%
AT	19,793	0,2%	90,1%
CA	19,274	0,3%	166,8%
CH	18,959	0,4%	63,2%
DK	18,044	0,2%	1,5%

## Tests of consistency

### 1- Priority patents

We now rely on `earliest_filing_id` now provided in Patstat to define priority patent application. It is more reliable than our own methodology to avoid to miss some linkages with priorities.

## 2- Geocoding of the addresses of inventors

We compare the country of the inventors given in the RISIS patent database with the country attributed to the inventor after the geocoding step. When the 2 countries are known for a given address (in 62% of the total 59,150,156 addresses), the country is the same for in 97,40% of the addresses. After a careful manual checking of these country discrepancies, we have selected the correct country in `country_final_enriched`.

## 3 Technical Specifications

### 3.1 Information on the data base system

#### Current data base system used

The current data base system is My SQL 5.1.63 with MyISAM as the default storage engine. In term of maintainability and backup, the main advantage of this storage engine is to use three different files for each table of a database:

- the data file has a .MYD (MYData) extension;
- the index file has a .MYI (MYIndex) extension;
- the structure file has a .frm extension.

MySQL is optimized for an intensive usage: a high level of accessibility and efficiency, for a low amount of users

#### Planned future technical changes concerning data base system

None

### 3.2 Technical variable definition

Variables present in PATSTAT raw data are defined below.

Table 14: Name and definition of the variables already present in PATSTAT and included in RISIS Patent

<b>APPLN_ID</b>	<b>Application unique identifier</b>
<b>APPLN_KIND</b>	<b>Kind of Application</b> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> A patent <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> _W PCT application (in the international phase) <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> _T used by some offices (e. g. AT, DE, DK, ES, GR, HR, PL, PT, SI, SM, TR) for applications which

	<p>are "translations" of granted PCT or EP applications</p> <p><input type="checkbox"/> _P provisional application (US only)</p> <p><input type="checkbox"/> _F design patent</p> <p><input type="checkbox"/> _D2, D3 artificial</p> <p><input type="checkbox"/> _Other "exotic" kind codes</p>
<b>APPLN_NR</b>	<b>Application number</b>
	Number issued by the Patent Authority where the National, International or Regional application was filed
<b>APPLN_AUTH</b>	<b>Office where the patent was applied</b>
<b>APPLN_ABSTRACT</b>	<b>Abstract of the application</b>
<b>APPLN_TITLE</b>	<b>Title of application</b>
<b>APPLN_ABSTRACT_LG</b>	<b>Language of abstract of application</b>
<b>APPLN_TITLE_LG</b>	<b>Language of title of application</b>
<b>DOCDB_FAMILY_SIZE</b>	<b>Size of DOCDB simple family of a given application</b>
<b>DOC_STD_NAME</b>	<b>Standardised name as recorded in DOCDB</b>
	<p>Standard name attributed to applicant and inventor names for inclusion in DOCDB.</p> <p>It is not 100% certain that the DOCDB standardised names are always linked with the correct person name, in particular if the person information came from a source other than DOCDB. This is especially true for names in USPTO patents.</p> <p>In case DOCDB does not provide a DOCDB standardized name, this attribute will contain the same data as the attribute PERSON_NAME.</p>
<b>DOC_STD_NAME_ID</b>	<b>ID for the DOCDB standardized name</b>
<b>EARLIEST_FILING_DATE</b>	<b>Date of the earliest filing</b>
	The earliest date of the filing dates of the application itself, its international application, its Paris Convention priority applications, the applications with which it is related via technical relations and its application continuations. Only directly related applications are considered; this is unlike the INPADOC family, where applications might also be indirectly related.
<b>EARLIEST_FILING_YEAR</b>	<b>Year of the earliest filing date</b>
<b>EARLIEST_PUBLN_YEAR</b>	<b>Year of the earliest publication date of an application</b>
<b>GRANTED</b>	<b>Granted indicator</b>



	"1" if there exists a publication of the grant; "0" otherwise
<b>IPC_CLASS_SYMBOL</b>	<b>IPC classification symbol (IPC 8th edition)</b>  Classification symbol according to the International Patent Classification, eighth edition (entered into force January 1, 2006)
<b>NB_CITING_DOCDB_FAM</b>	<b>Number of forward citations on family level</b>  Number of distinct DOCDB simple families citing at least one of the publications or applications of the DOCDB simple family of the current application
<b>PERSON_ID</b>	<b>Person identification</b>
<b>PERSON_NAME</b>	<b>Person name</b>
<b>PSN_ID</b>	<b>ID for the PATSTAT standardized name</b>
<b>PSN_NAME</b>	<b>PATSTAT standardized name</b>  The attribute is populated for all persons. Names of persons which have not been harmonized are just copied from the attribute PERSON_NAME.
<b>PSN_SECTOR</b>	<b>Sector of the applicant</b>  INDIVIDUAL ; COMPANY ; UNKNOWN ; GOVERNMENT ; NON-PROFIT ; UNIVERSITY ; HOSPITAL
<b>PUBLN_AUTH</b>	<b>Publication Authority or Publishing office</b>  Patent Authority that issued the publication of the application
<b>PUBLN_CLAIMS</b>	Indicator of the <b>number of claims</b> in the given publication
<b>PUBLN_DATE</b>	<b>Publication date</b>  Date on which the publication was made available to the public
<b>PUBLN_KIND</b>	<b>Kind of Publication</b>  Publication kind attributed by the Patent Authority issuing the publication
<b>PUBLN_LG</b>	<b>Publication language</b>
<b>PUBLN_NR</b>	<b>Publication number</b>  Number given by the Patent Authority issuing the

	publication
--	-------------

Table 15: Name and definition of the variables added in RISIS Patent (non included in PATSTAT or modified)

<b>adr_final</b>	Addresses of inventors and applicants	Completed
<b>Iso_etry</b>	Addresses of inventors and applicants (country)	Completed
<b>Ctry_final_enriched</b>	Addresses of inventors (enriched country)	Completed
<b>Source_etry_final_enriched</b>	Method for enriched inventors' country assignment	Completed
<b>Latitude</b>	Addresses of inventors and applicants	New info provided
<b>Longitude</b>	Addresses of inventors and applicants	New info provided
<b>Rurban_area_code</b>	Allocation of addresses of inventors and applicants	New info provided
<b>Rurban_area_name</b>	Allocation of addresses of inventors and applicants	New info provided
<b>Rurban_area_characteristics</b>	Allocation of addresses of inventors and applicants	New info provided
<b>NUTS3 Id</b>	Allocation of addresses of inventors and applicants	New info provided
<b>RISIS register id</b>	ID Allocation to applicants included in RISIS registers (ORGREG, FIRMREG)	New info provided
<b>RISIS register</b>	ORGREG, FIRMREG (VICO, CHEETAH,CIB)	New info provided
<b>Singleton</b>	Priority patent without any extension	New info provided
<b>Transnat</b>	Priority patent further extended in a second country at least	New info provided
<b>IP5 family</b>	Priority patent further extended in a second country at least and applied in one IP5 patent office (info of Inpadoc family)	New info provided
<b>2_IP5 family</b>	Priority patent further extended in a second country at least and applied in two IP5 patent office (info of Inpadoc family)	New info provided
<b>Triadic</b>	Priority patent for which the invention is protected at JPO, USPTO and JPO (info of Inpadoc family)	New info provided
<b>IP5</b>	Priority patent for which the invention is protected at EPO, USPTO, SINO, KINO and JPO (info of Inpadoc family)	New info provided
<b>US</b>	Priority patent for which the invention is protected at USPTO (info of Inpadoc family)	New info provided
<b>EP</b>	Priority patent for which the invention is	New info provided

	protected at EPO, (info of Inpadoc family)		
<b>JP</b>	Priority patent for which the invention is protected at JPO, (info of Inpadoc family)	New provided	info
<b>CN</b>	Priority patent for which the invention is protected at SINO (info of Inpadoc family)	New provided	info
<b>KR</b>	Priority patent for which the invention is protected at KINO (info of Inpadoc family)	New provided	info
<b>PCT in the Inpadoc family</b>	Priority patent for which the invention is protected using a PCT patent (info of Inpadoc family)	New provided	info
<b>Number patents in docdb family</b>	Number patent applications in docdb family (restricted family)	New provided	info
<b>Number distinct offices in docdb family</b>	Number of distinct patent offices where patents where applied for in docdb family (restricted family)	New provided	info
<b>Number patents in inpadoc family</b>	Number patent applications in docdb family (extended family)	New provided	info
<b>Number distinct offices in inpadoc family</b>	Number of distinct patent offices where patents where applied for in docdb family (extended family)	New provided	info
<b>Domain</b>	Fractional counting of priority patent on the 5 domains	New provided	info
<b>Techno fields</b>	Fractional counting of the priority patent on the 35 technological fields	New provided	info
<b>Techno subfields</b>	Fractional counting of the priority patent on the 401 technological subfields	New provided	info
<b>IPC full textual description</b>	Aggregation of the hierarchical IPC description	New provided	info
<b>IPC class 4 digits</b>	Fractional counting of the priority patent on the 615 technological 4 digits classes	New provided	info
<b>Earliest publication of grant</b>	First notification of a grant	Reshaped from PATSTAT	info

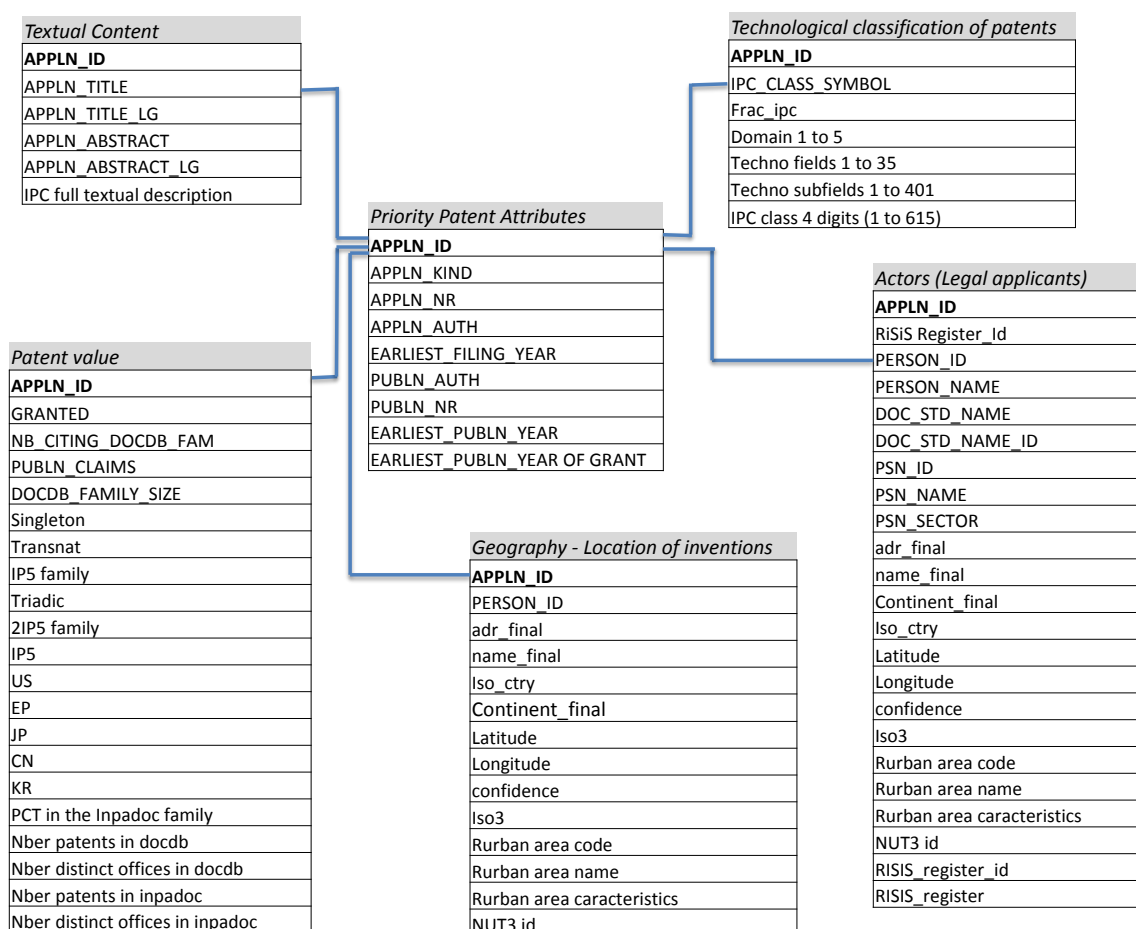
### 3.3 Description of the Entity Relationship Model

The data model of the RISIS Patent database is shown below. It includes 6 tables related using the identifier of the priority patent application (appln\_id):



- Table “Priority Patent Attributes” gives a set of information related to the patent application and publication places and dates,
- Table “Textual Content” includes the pieces of textual information useful for text mining. It aims at defining the core of the protected inventions beside the predefined technological classifications
- Table “ technological classification of patents” relates patent to technology on a fractional counting based using existing classification built on IPC codes
- Table “ Patent value” proposes several indicators to estimate the patent application value
- Table “Geography – Location of inventions” informs on the locations of the inventions based on the inventors’ addresses.
- Table “Actors” focuses on the institutions that patent and give information on their type, name and location.

Table 16: Data model of RISIS Patent



The data model of the PATSTAT2021 database is shown in Appendix 5.

### 3.4 Interfaces for access and to other infrastructures

RISIS Patent will be interlinked with the RISIS organisation registers, i.e. the firm register (FirmReg), and public sector (org register id of institutions). Furthermore, the database is designed to comply with RISIS integrative dimensions, actors, space as described in the following.

#### Geocoding and allocation to urban areas

Geocoding is done on inventors and applicants using the CORTEXT geocoding tools (see specific documentation on the CORTEXT Platform). It provides for each inventor and applicant its geo-coordinates. The % of addresses that can be geocoded is very high (around 98%), but the issue is to have addresses, and we face here a critical issue for the PATSTAT database.

We focus here on inventor addresses, i.e. pieces of detailed geographical information enabling address geocoding (city, county, street, ...). There are multiple strategies by firms where to take their patents. We thus do not use the address of applicants for analysing the geography of inventions. We focus on the addresses of inventors. Their presence and quality varies depending upon offices. For instance a study done by SPRU in the early 2000s demonstrated that the USPTO could not be trusted for addresses of inventors for patents taken elsewhere and extended to the US.

The raw coverage of addresses in PATSTAT for priority patents is limited and it has been a preoccupation by numerous teams to increase it.

We process data with the following steps to improve the address coverage:

- We first retrieve addresses that are not located in the adequate item in PATSTAT (addresses retrieved from natural or doc standardised names) (step 1-a)
- We incorporate external resources available: REGPAT from OECD, addresses provided for by INPI, Japanese addresses (thanks to colleagues in Japan) (step 1-b)
- For priority patents still without addresses, we then look into the Inpadoc family to look for inventors and addresses with specific rules linked to proximity non-US patents (step 1-c)
- We geocode and allocate URA addresses using the CORTEXT geographical modules. If a country is assigned after geocoding to a previously empty country address, we keep it (step 2).
- For priority patents still without countries, we allocate the country of the applicant to the inventor if the country of the applicant is the country where the patent was applied for (ctry\_appl=appln\_auth) (step 3)

Actors

We deal with actors that are legal applicants (individual applicants were discarded from the database in RISIS Patent. In the current RISIS Patent database, we only rely on the sectorial information provided in the raw PATSTAT database<sup>9</sup>.

Among the 59 million inventors, the geographical location (country assignment) was identified for 83% of them and for 61% of them, the addresses were geocoded and assigned to an URA (Urban and Rural Area). Among the 22 million applicants, the geographical location (country assignment) was identified for 76% of them and for 63% of them, the addresses were geocoded and assigned to an URA. Details are given above in the documentation in Tables 8 to 17.

### Integration with RCF

The current RISIS Patent is made available for access to researchers in RISIS is foreseen to be fully incorporated in RCF, under the condition of controlled access and that security of usage is given (i.e. access for selected users with a concrete research project to the parts of the dataset needed for the research). Linking to other datasets in the RCF will be realized via the RISIS registers (providing the respective identifiers to the registers in RISIS Patent). Technical issues for incorporation of RISIS Patent into RCF (e.g. database system, how can a user access which parts of the dataset, etc.) are to be defined in close cooperation with WP4 beginning with autumn.

## 4 Scientific use and main references

RISIS Patent is an accessible and rich data source via RISIS for research activities in the production of knowledge using patent data. It allows studying the dynamics of knowledge creation along different dimensions: space, actors and technologies.

Thanks to its links with other RISIS facilities, Risis Patent enables to access to these dimensions at a coarse level or at a fine grained level using either usual classification (for technologies, geography) or designing ad-hoc data subsets of patents in specific topics of inventions, for a particular type of institutions in given geographical spaces.

It had been recently used to:

.Observe **the distribution and location of the inventive** activities of a group of European public research centres in the field of marine biotechnology (the EMBRIC project)

.Analyse the **exploitation of new knowledge** in specific industries (pharmaceutical and chemical industries), done by researchers from Universite Gustave Eiffel

---

<sup>9</sup> Harmonizing names and allocation of assignee sectors in Patstat raw data was done by ECOOM (K.U. LEUVEN; <http://www.ecoom.be/en/EEE-PPAT>).

.Explore the **Inventive Productivity of Multinational Firms** using non parametric modelling (Conditional Efficiency Analysis)

.Analyse **the internationalisation of applied knowledge** production with a focus on special countries (Israel, central European countries)

.Analyse **the knowledge** production in worldwide Metropolitan Areas (Universite Gustave Eiffel team)

.Analyse **the knowledge** production specialization in EU NUTS regions Univeristy (Vilnius University)

## Recent References

Laurens, P., Toma, P., Schoen, A. *et al.* How does Internationalisation affect the productivity of R&D activities in large innovative firms? A conditional nonparametric investigation. Quality and Quantity (April 2022). <https://doi.org/10.1007/s11135-022-01391-z>

Patricia Laurens, Antoine Schoen, Philippe Laredo, International Patents: the role of large multinational firms in building competitive metropolitan areas, RISIS #6 Policymaker Session, 3th December 2020 (hal-03289712)

Patricia Laurens, Lionel Villard, Antoine Schoen, Philippe Laredo. The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?. Scientometrics, Springer Verlag, 2018, 114, pp.91-112. (10.1007/s11192-017-2578-5). (hal-01671905)

Laurens, P., Le Bas, C., Schoen, A. (2018), Worldwide IP coverage of patented inventions in large pharma firms: to what extent do the internationalisation of R&D and firm strategy matter?

Submitted to the International Journal of Technology Management

Laurens, P., Le Bas, C., Lhuillery S., Schoen, A. , (2018) Firm specialisation in clean energy technologies: the influence of path dependence and technological diversification. Revue d'économie Industrielle, n°164 (4eme trimestre 2018)

"Evolving technological capabilities of firms; Complexity, divergence, and stagnation"

Antoine Schoen; Patricia Laurens, Alfredo Yegros, Philippe Laredo  
STI conference Paris September, 2017

Gaston Heimeriks, Antoine Schoen, Patricia Laurens, Alfredo Yegros and Dieter Franz Kogler

Knowledge, networks and proximities - An analysis of knowledge dynamics in the Chemical and Pharmaceutical and Biotechnology sectors

Eu-spri 2018 conference Paris 2018

# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES



This project is funded by the European Union  
under Horizon2020 Research and Innovation  
Programme Grant Agreement n°824091

Laurens Patricia, Antoine Schoen, Pierluigi Toma and Cinzia Daraio  
Exploring the Innovative Efficiency of Big Multinational Firms through Conditional  
Efficiency Analysis  
STI 2018, Leiden (Netherlands) 12-14 September



## Appendix



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 824091

## Appendix 1: Artificial patents

The very first filing of a patent application that aims at protecting a novel invention, also known as a priority patent, is considered as a promising type of document for analysing the dynamics of knowledge creation in the field of technology (de Rassenfosse et al., 2013).

If most of the worldwide priority patent applications included in the PATSTAT database are the published patents, there is an additional group of patents, classified as « artificial patents » that are not part of the EPO database. The artificial priority patents are priority patents cited in further patent extensions included in the PATSTAT database but for these patents the application was not identified with 100% certainty by EPO.

An artificial priority patent in its original state contains very scarce information. The only information included in the Patent database provided by EPO for these documents relates to the patent office where the artificial priority patent has been applied for, the date and the type of the application and whether the patent has been cited as a priority patent or through another type of citation. All the other information, which is usually included in fully documented patent documents, is missing: the technology domains, the inventors' and applicants' addresses and names, the title and the abstract of the patent application.

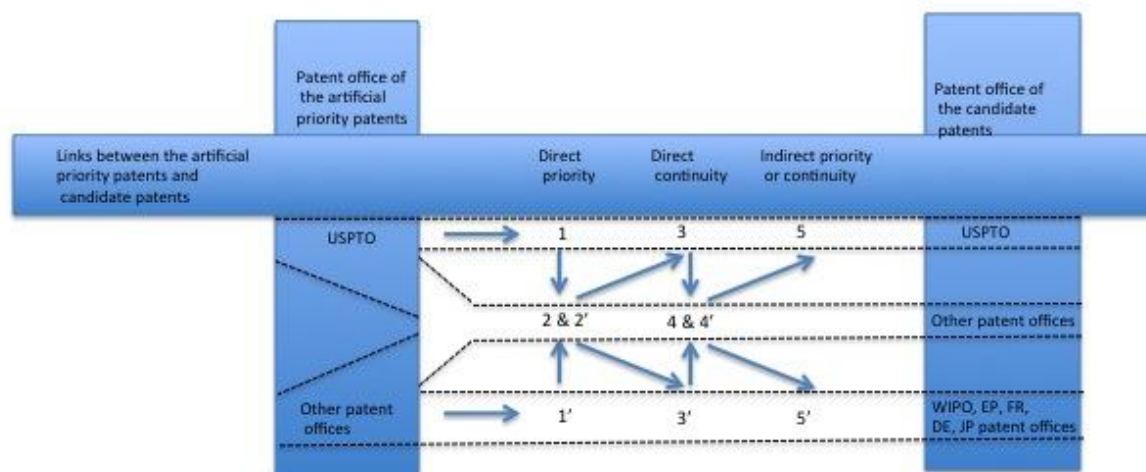
In RISIS Patent, we fill both the applicants' and the inventors' names and addresses but also of IPC codes in the artificial priority patents based on the identification and use of this piece of information retrieved from a non artificial patent belonging to the same INPADOC family (Laurens et al., 2018). To select the candidate from which information is retrieved, a set of sequential rules is defined in order to select a candidate patent as close as possible to the artificial one. The criteria relate to the type of information available in the candidate patent, the nature of the priority link between the artificial patent and the selected candidate patent, the time lag between the two applications and the patent offices. In order to be considered as a potential candidate, the patent must contain information concerning addresses, either the inventors' addresses or the applicants' addresses or both. A candidate with a direct priority link with the artificial patents is selected first (from table t1s204 of the PATSTAT database). In absence of such a link, a candidate with a direct continuity link is selected (continuation, division/continuation in parts/internal continuities in table t1s 216). At last, candidate patents with indirect links are selected. Besides the type of links, we also consider the time lag between the artificial and the candidate patents. For two candidates with the same link, we select the candidate with a date of filing closest to the filing date of the artificial patent (the maximum authorised time lag is fixed to five years). Finally, we give a higher priority to US candidates to fill US artificial patents. For non US artificial patent applications, candidates filed as PCT, or EP, JP, FR, DE received the highest priority. A schematic view of the rules of selection of candidates chosen for replenishing artificial patents is presented in figure A1.

As soon as a candidate patent is selected, its information related to the names and addresses and the IPC codes is used to fill the artificial patent. This procedure made possible to propagate information from a candidate patent in 99.3% of the artificial priority patents filed between 1995 and 2010

These high retrieval percentages show that the filling step of artificial patents was very effective. It thus provides a better coverage of the overall inventive activities when using the priority patents, in particular in patent offices where the share of artificial patents is

high (EPO, USPTO patent offices in India or Australia).

Figure A1: Schematic description of the different steps to fill inventors' and applicants' addresses in artificial priority patents.



Scripts are shown in [https://gitlab.com/cortext/patents/patstat/-/tree/version-2021/version\\_2021/artificial\\_patents](https://gitlab.com/cortext/patents/patstat/-/tree/version-2021/version_2021/artificial_patents)

## References :

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., van Pottelsberghe de la Potterie, B. (2013). The worldwide count of priority patents: a new indicator of inventive activity. *Research Policy* 42, 720-737.

Patricia Laurens, Lionel Villard, Antoine Schoen, Philippe Laredo. The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?. *Scientometrics*, Springer Verlag, 2018, 114, pp.91-112. (10.1007/s11192-017-2578-5). (hal-01671905)

## CorTexT Manager :

Breucker P., Cointet J., Hannud Abdo A., Orsal G., de Quatrebarbes C., Duong T., Martinez C., Ospina Delgado J.P., Medina Zuluaga L.D., Gómez Peña D.F., Sánchez Castaño T.A., Marques da Costa J., Laglil H., Villard L., Barbier M. (2016). CorTexT Manager (version v2). URL: <https://docs.cortext.net>

@software{cortext\_manager\_v2\_bibtex,  
keywords = {natural language processing, social network analysis, geospatial analysis, descriptive statistics, scientometrics, bibliometrics},

author = {Breucker, Philippe and Cointet, Jean-Philippe and Hannud Abdo, Alexandre and Orsal, Guillaume and de Quatrebarbes, Constance and Duong, Tam-Kien and Martinez, Cristian and Ospina Delgado, Juan Pablo and Medina Zuluaga, Luis Daniel and Gómez Peña, Diego Fernando and Sánchez Castaño, Tatiana Andrea and Marques da Costa, Joenio and Laglil, Hajar and Villard, Lionel and Barbier, Marc},

month = {10},

title = {CorText Manager},

url = {<https://docs.cortext.net>},

year = {2016}

}

Villard, Lionel & Ospina, Juan Pablo (2018). *CorText Geocoding service*, ESIEE Paris, Paris-Est University. <https://docs.cortext.net/geocoding-addresses/>

Villard, Lionel, Opsina, Juan Pablo & Medina, Luis Daniel (2019). *CorText Geospatial Exploration Tool*. ESIEE Paris, Paris Est University. <https://docs.cortext.net/cortext-geospatial-exploration-tool/>

## Appendix 2: Type of applicants

One of the main goals in our research activities is the study and analysis on the dynamics of R&D and innovations made by legal entities, such as corporations, universities, research centers and in general every class of entity that do not fit into the categorization of “natural person”. Considering the above, we have worked with the data contributed by PATSTAT to start with the data restructuring in order to begin on the classification process, in this way we have defined two entities classes: “legal” and “person”. A third one “unknown” contains the cases where it was not possible to allocate a type to the applicant.

In the entities that can be found within the PATSTAT applicants coexist a large proportion of homonym names, for instance, the word 'Ford' may be tagged as a company or as an individual, which means that, using a direct approach of detection where several functions that implement gazetteers and by combing regular expressions, could lead in a low quality results for the categorization of legal entities and individuals. Another ambiguity that we had to face was for companies names that do not have any corporate ending and consequently is not likely that we are able to label those entities with just a dictionary list of legal suffix.

Therefore, we turn the approach of the solution, instead of merely doing a direct match of terms we used a heuristic approach in the patent context in order to separate the different entities into three subsets named as 'probably legal', 'probably individual' and 'ambiguous'. The allocation of each entity is designated by a series of methods and rules that are relay on not only on the use of gazetteers but on the characterization of the applicants through the relation with the patent, besides some simple lexical identifiers.

We then compare our allocation of the applicant type with the one available in the Patstat raw data (psn\_sector). If the 2 allocations converge, we keep this type as a final allocation type of applicant. If not, we manually examine the set of applicants and allocate the most appropriate type (we were rather generous and allocate a legal type in case of doubt. We are aware that individual are still be present in particular when type UGE is “”‘legal’ and ‘psn\_sector is ‘Individual’.

At the end of the process, only applicants with a final\_UGE\_type set to legal were included in the table RPD\_actor.

Table A-2: Final allocation of the applicant type

Type_UGE	psn_sector	count(*)	final UGE_type
NULL	NULL	1842	legal
NULL	"	27	legal
	'COMPANY'	37	individual
NULL	'INDIVIDUAL'	7082	individual
NULL	'UNKNOWN'	203	individual
'legal'	"	154	individual
'legal'	'COMPANY'	7692509	legal
'legal'	'COMPANY GOV NON-PROFIT'	50602	legal
'legal'	'COMPANY GOV NON-PROFIT UNIVERSITY'	556	legal
'legal'	'COMPANY HOSPITAL'	2278	legal

'legal'	'COMPANY INDIVIDUAL'	4	legal
'legal'	'COMPANY UNIVERSITY'	1909	legal
'legal'	'GOV NON-PROFIT'	283713	legal
'legal'	'GOV NON-PROFIT HOSPITAL'	6	legal
'legal'	'GOV NON-PROFIT UNIVERSITY'	16876	legal
'legal'	'HOSPITAL'	19257	legal
'legal'	'INDIVIDUAL'	1432286	legal
'legal'	'UNIVERSITY'	178650	legal
'legal'	'UNKNOWN'	1243885	legal
'person'	"	4742	person
'person'	'COMPANY'	116319	legal
'person'	'COMPANY GOV NON-PROFIT'	388	legal
'person'	'COMPANY GOV NON-PROFIT UNIVERSITY'	1	legal
'person'	'COMPANY HOSPITAL'	17	legal
'person'	'COMPANY INDIVIDUAL'	2	legal
'person'	'COMPANY UNIVERSITY'	16	legal
'person'	'GOV NON-PROFIT'	4834	legal
'person'	'GOV NON-PROFIT UNIVERSITY'	121	legal
'person'	'HOSPITAL'	129	legal
'person'	'INDIVIDUAL'	6223141	person
'person'	'UNIVERSITY'	2437	legal
'person'	'UNKNOWN'	505611	person

Final results are shown in Table 1

We are currently matching RPD applicants' names with legal entities names included in RISIS register (ORGREG register for EU universities and PROs) and FIRMREG register (including 3 firms data set: CIB, VICO and CHEETAH). The results of the matching with ORGREG entities is already available: among the 22.5millions of applicants, 0,79% are ORGREG entities. They have applied for 156,132 patents (0,76% of RPD priority patents).

Details of the methodology are described in: [https://gitlab.com/cortex/patents/patstat/-/tree/version-2021/version\\_2021/applicants\\_classification](https://gitlab.com/cortex/patents/patstat/-/tree/version-2021/version_2021/applicants_classification)

## Appendix 3: Classification of patent technologies

Table A-3: Definition of the Domains, fields and subfields classification

Label domains	fields	Label fields	subfields	Label subfields	ipc	not_ipc	not_appl_n_id
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T01 F01	Lighting	F21		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T02 F01	Cables with Special Electric Properties	H01B		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T03 F01	Resistors	H01C		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T04 F01	Magnets	H01F		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T05 F01	Capacitors	H01G		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T06 F01	Electric Switches	H01H		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T07 F01	Discharge Lamps	H01J		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T08 F01	Electric Incandescent Lamps	H01K		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T09 F01	Batteries and related	H01M		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T10 F01	Electrically-Conductive Connections	H01R		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T11 F01	Spark Gaps	H01T		
Electrical engineering	TF01	Electrical machinery,	T12 F01	Boards For The Distribution Of Electricity	H02		

		apparatus, energy					
Electrical engineering	TF0 1	Electrical machinery, apparatus, energy	T13 F01	Electric Heating	H05B		
Electrical engineering	TF0 1	Electrical machinery, apparatus, energy	T14 F01	Electrical Device stopping human beings	H05C		
Electrical engineering	TF0 1	Electrical machinery, apparatus, energy	T15 F01	Static Electricity	H05F		
Electrical engineering	TF0 1	Electrical machinery, apparatus, energy	T16 F01	Other subjects electricity	H99Z		
Electrical engineering	TF0 2	Audio-visual technology	T01 F02	Displaying Advertising	G09F		
Electrical engineering	TF0 2	Audio-visual technology	T02 F02	Arrangements For Control	G09G		
Electrical engineering	TF0 2	Audio-visual technology	T03 F02	Information Storage Based Record Carrier	G11B		
Electrical engineering	TF0 2	Audio-visual technology	T04 F02	Scanning details of television	H04N 3		
Electrical engineering	TF0 2	Audio-visual technology	T05 F02	Details of television systems	H04N 5		
Electrical engineering	TF0 2	Telecommunica tions	T06 F02	Television systems	H04N 7		
Electrical engineering	TF0 2	Audio-visual technology	T07 F02	Details of colour television systems	H04N 9		
Electrical engineering	TF0 2	Telecommunica tions	T08 F02	Colour television systems	H04N 11		
Electrical engineering	TF0 2	Audio-visual technology	T09 F02	Stereoscopic television systems	H04N 13		
Electrical engineering	TF0 2	Audio-visual technology	T10 F02	Stereoscopic colour television systems	H04N 15		
Electrical engineering	TF0 2	Audio-visual technology	T11 F02	Diagnosis for television systems	H04N 17		
Electrical engineering	TF0 2	Audio-visual technology	T12 F02	Pictorial Communication	H04N 101		
Electrical engineering	TF0 2	Audio-visual technology	T13 F02	Loudspeakers Microphones	H04R		
Electrical engineering	TF0 2	Audio-visual technology	T14 F02	Stereophonic Systems	H04S		
Electrical engineering	TF0 2	Audio-visual technology	T15 F02	Printed Circuits	H05K		
Electrical engineering	TF0 3	Telecommunica tions	T01 F03	Transmission Systems	G08C		
Electrical engineering	TF0 3	Telecommunica tions	T02 F03	Waveguides	H01P		
Electrical engineering	TF0 3	Telecommunica tions	T03 F03	Aerials	H01Q		



Electrical engineering	TF03	Telecommunications	T04 F03	Transmission	H04B		
Electrical engineering	TF03	Telecommunications	T05 F03	Broadcas Communication	H04H		
Electrical engineering	TF03	Telecommunications	T06 F03	Multiplex Communication	H04J		
Electrical engineering	TF03	Telecommunications	T07 F03	Secret Communication	H04K		
Electrical engineering	TF03	Telecommunications	T08 F03	Telephonic Communication	H04M		
Electrical engineering	TF03	Telecommunications	T09 F03	Transmission or reproduction of documents	H04N1		
Electrical engineering	TF03	Telecommunications	T10 F03	Selecting (Switches/Relays)	H04Q		
Electrical engineering	TF04	Digital communication	T01 F04	Transmission Of Digital Information	H04L		
Electrical engineering	TF04	Digital communication	T02 F04	Selective content distribution	H04N21		
Electrical engineering	TF04	Digital communication	T03 F04	Wireless Communication	H04W		
Electrical engineering	TF05	Basic communication processes	T01 F05	Basic Electronic Circuitry	H03		
Electrical engineering	TF06	Computer technology	T01 F06	Mechanically computed Digital Computers	G06	G06Q	
Electrical engineering	TF06	Computer technology	T02 F06	Speech Anal. Or Synth.	G10L		
Electrical engineering	TF06	Computer technology	T03 F06	Static Stores	G11C		
Electrical engineering	TF07	IT methods for management	T01 F07	Data Processing Systems	G06Q		
Electrical engineering	TF08	Semiconductor s	T01 F08	Semiconductor Devices	H01L		
Instruments	TF09	Optics	T01 F09	Optical Elements Systems	G02		
Instruments	TF09	Optics	T02 F09	Photograph Apparatus	G03B		
Instruments	TF09	Optics	T03 F09	Photographic Processes	G03C		
Instruments	TF09	Optics	T04 F09	Apparatus For Processing Exposed Photographic Materials	G03D		
Instruments	TF09	Optics	T05 F09	Photomechanics of Surfaces	G03F		
Instruments	TF09	Optics	T06 F09	Photomechanical Production Of Surfaces	G03G		
Instruments	TF09	Optics	T07 F09	Holography	G03H		
Instruments	TF09	Optics	T08 F09	Devices Using Stimulated Emission	H01S		
Instruments	TF10	Measurement	T01 F10	Measuring Length Thickness	G01B		
Instruments	TF10	Measurement	T02 F10	Measuring Distances Levels	G01C		



Instruments	TF1 0	Measurement	T03 F10	Multivariable Measuring	G01D		
Instruments	TF1 0	Measurement	T04 F10	Measuring Volume Flow	G01F		
Instruments	TF1 0	Measurement	T05 F10	Weighing	G01G		
Instruments	TF1 0	Measurement	T06 F10	Measuring Vibrations	G01H		
Instruments	TF1 0	Measurement	T07 F10	Optical Measurements	G01J		
Instruments	TF1 0	Measurement	T08 F10	Measuring Temperature	G01K		
Instruments	TF1 0	Measurement	T09 F10	Measuring Force Stress Torque Work	G01L		
Instruments	TF1 0	Measurement	T10 F10	Testing Structures	G01M		
Instruments	TF1 0	Measurement	T11 F10	Chemical Physical Analyses	G01N	G01 N 33	
Instruments	TF1 0	Measurement	T12 F10	Measuring Speed Acceleration	G01P		
Instruments	TF1 0	Measurement	T13 F10	Scanning-Probe Techniques	G01Q		
Instruments	TF1 0	Measurement	T14 F10	Measuring Electric Variables	G01R		
Instruments	TF1 0	Measurement	T15 F10	Radio Direction-Finding	G01S		
Instruments	TF1 0	Measurement	T16 F10	Geophysics	G01V		
Instruments	TF1 0	Measurement	T17 F10	Meteorology	G01W		
Instruments	TF1 0	Measurement	T18 F10	Horology	G04		
Instruments	TF1 0	Measurement	T19 F10	Details Of Instruments	G12B		
Instruments	TF1 0	Measurement	T20 F10	Other subjects nucleonics	G99Z		
Instruments	TF1 1	Analysis of biological materials	T01 F11	Material Analysis by Chem/Phys Properties	G01N 33		
Instruments	TF1 2	Control	T01 F12	Control Systems in General	G05B		
Instruments	TF1 2	Control	T02 F12	Regulating Non Electric Variables	G05D		
Instruments	TF1 2	Control	T03 F12	Regulating Electric Variables	G05F		
Instruments	TF1 2	Control	T04 F12	Checking Devices	G07		
Instruments	TF1 2	Control	T05 F12	Signalling Systems	G08B		
Instruments	TF1 2	Control	T06 F12	Traffic Control Systems	G08G		
Instruments	TF1 2	Control	T07 F12	Educational Or Demonstration Appliances	G09B		

Instruments	TF1 2	Control	T08 F12	Ciphering Apparatus	G09C		
Instruments	TF1 2	Control	T09 F12	Railway Or Like Time Or Fare Tables	G09D		
Instruments	TF1 3	Medical technology	T01 F13	Surgery-Diagnosis	A61B		
Instruments	TF1 3	Medical technology	T02 F13	Dentistry	A61C		
Instruments	TF1 3	Medical technology	T03 F13	Veterinary	A61D		
Instruments	TF1 3	Medical technology	T04 F13	Blood Vessel Filters	A61F		
Instruments	TF1 3	Medical technology	T05 F13	Furniture for Patients	A61G		
Instruments	TF1 3	Medical technology	T06 F13	Therapy Apparatus	A61H		
Instruments	TF1 3	Medical technology	T07 F13	Medical Containers	A61J		
Instruments	TF1 3	Medical technology	T08 F13	Methods For Sterilising	A61L		
Instruments	TF1 3	Medical technology	T09 F13	Devices for Introducing Media in Body	A61M		
Instruments	TF1 3	Medical technology	T10 F13	Electrotherapy	A61N		
Instruments	TF1 3	Medical technology	T11 F13	X-Ray Technique	H05G		
Chemistry	TF1 4	Organic fine chemistry	T01 F14	Cosmetic Preparations	A61K 8		
Chemistry	TF1 4	Organic fine chemistry	T02 F14	Use of Cosmetics	A61Q		
Chemistry	TF1 4	Organic fine chemistry	T03 F14	Organic Chemistry	C07B		A61K
Chemistry	TF1 4	Organic fine chemistry	T04 F14	Acyclic or Carbocyclic Compounds	C07C		A61K
Chemistry	TF1 4	Organic fine chemistry	T05 F14	Heterocyclic Compounds	C07D		A61K
Chemistry	TF1 4	Organic fine chemistry	T06 F14	Compounds with Elts Other Than C H Halogen O N S Se or Te	C07F		A61K
Chemistry	TF1 4	Organic fine chemistry	T07 F14	Sugars Nucleotides	C07H		A61K
Chemistry	TF1 4	Organic fine chemistry	T08 F14	Steroids	C07J		A61K
Chemistry	TF1 4	Organic fine chemistry	T09 F14	Combinatorial Chemistry	C40B		A61K
Chemistry	TF1 5	Biotechnology	T01 F15	Compounds Of Unknown Constitution	C07G		A61K
Chemistry	TF1 5	Biotechnology	T02 F15	Peptides	C07K		A61K
Chemistry	TF1 5	Biotechnology	T03 F15	Apparatus For Enzymology Or Microbiology	C12M		A61K
Chemistry	TF1 5	Biotechnology	T04 F15	Micro-Organisms Enzymes Culture Media	C12N		A61K
Chemistry	TF1	Biotechnology	T05	Fermentation Or Enzyme-	C12P		A61K

	5		F15	Using Processes			
Chemistry	TF1 5	Biotechnology	T06 F15	Measuring Processes with Enzymes Or Micro- Organisms	C12Q		A61K
Chemistry		Biotechnology	T07 F15	Micro-Organisms-General	C12R		A61K
Chemistry	TF1 5	Biotechnology	T08 F15	Micro-Organisms To treat solid materials	C12S		A61K
Chemistry	TF1 6	Pharmaceuticals	T01 F16	Medical Preparations	A61K	A61K 8	
Chemistry	TF1 6	Pharmaceuticals	T02 F16	Therapeutic Activity Of Chemical Compounds	A61P		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T01 F17	Polysaccharides	C08B		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T02 F17	Treatment Of Rubbers	C08C		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T03 F17	Macromol with C-to-C Unsaturated Bonds	C08F		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T04 F17	Macromol without C-to-C Unsaturated Bonds	C08G		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T05 F17	Derivatives Of Natural Macromolecules	C08H		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T06 F17	Inorg Or Non-Macromolr Organ Subst	C08K		
Chemistry	TF1 7	Macromolecular chemistry, polymers	T07 F17	Compositions Of Macromolecular Compounds	C08L		
Chemistry	TF1 8	Food chemistry	T01 F18	New Plants	A01H		
Chemistry	TF1 8	Food chemistry	T02 F18	Treatment of Dough	A21D		
Chemistry	TF1 8	Food chemistry	T03 F18	Preseving food	A23B		
Chemistry	TF1 8	Food chemistry	T04 F18	Dairy Products	A23C		
Chemistry	TF1 8	Food chemistry	T05 F18	Edible Oils	A23D		
Chemistry	TF1 8	Food chemistry	T06 F18	Coffee and Tea	A23F		
Chemistry	TF1 8	Food chemistry	T07 F18	Chocolate	A23G		
Chemistry	TF1 8	Food chemistry	T08 F18	Proteins for Food	A23J		
Chemistry	TF1 8	Food chemistry	T09 F18	Fodder	A23K		
Chemistry	TF1 8	Food chemistry	T10 F18	Other Foods	A23L		

Chemistry	TF1 8	Food chemistry	T11 F18	Brewing Of Beer	C12C		
Chemistry	TF1 8	Food chemistry	T12 F18	Fermented Solutions	C12F		
Chemistry	TF1 8	Food chemistry	T13 F18	Alcoholic Beverages	C12G		
Chemistry	TF1 8	Food chemistry	T14 F18	Treatment of Alcoholic Beverages	C12H		
Chemistry	TF1 8	Food chemistry	T15 F18	Vinegar	C12J		
Chemistry	TF1 8	Food chemistry	T16 F18	Production of sugar juices	C13B 10		
Chemistry	TF1 8	Food chemistry	T17 F18	Purification of sugar juices	C13B 20		
Chemistry	TF1 8	Food chemistry	T18 F18	Crystallisation of sugar	C13B 30		
Chemistry	TF1 8	Food chemistry	T19 F18	Extraction of sucrose from molasses	C13B 35		
Chemistry	TF1 8	Food chemistry	T20 F18	Drying sugar	C13B 40		
Chemistry	TF1 8	Food chemistry	T21 F18	Working-up of sugar	C13B 50		
Chemistry	TF1 8	Food chemistry	T22 F18	Subject matter not provided for in other groups of this subclass	C13B 99		
Chemistry	TF1 8	Food chemistry	T23 F18	Sugar Juices	C13D		
Chemistry	TF1 8	Food chemistry	T24 F18	Sugar	C13F		
Chemistry	TF1 8	Food chemistry	T25 F18	Extraction Of Sugar	C13J		
Chemistry	TF1 8	Food chemistry	T26 F18	Synthesis Of Sugars	C13K		
Chemistry	TF1 9	Basic materials chemistry	T01 F19	Disinfectants	A01N		
Chemistry	TF1 9	Basic materials chemistry	T02 F19	Repellants	A01P		
Chemistry	TF1 9	Basic materials chemistry	T03 F19	Fertilisers	C05		
Chemistry	TF1 9	Basic materials chemistry	T04 F19	Explosives	C06		
Chemistry	TF1 9	Basic materials chemistry	T05 F19	Organic Dyes	C09B		
Chemistry	TF1 9	Basic materials chemistry	T06 F19	Dyeing Of Inorganics	C09C		
Chemistry	TF1 9	Basic materials chemistry	T07 F19	paints and inks	C09D		
Chemistry	TF1 9	Basic materials chemistry	T08 F19	Natural Resins	C09F		
Chemistry	TF1 9	Basic materials chemistry	T09 F19	Polishing Compositions	C09G		
Chemistry	TF1 9	Basic materials chemistry	T10 F19	Glue Or Gelatine	C09H		
Chemistry	TF1	Basic materials	T11	Adhesives	C09J		

	9	chemistry	F19			
Chemistry	TF1 9	Basic materials chemistry	T12 F19	Materials For Miscellaneous Applications	C09K	
Chemistry	TF1 9	Basic materials chemistry	T13 F19	Production Of Gas Coke Tar by distillation	C10B	
Chemistry	TF1 9	Basic materials chemistry	T14 F19	Working-Up Tar	C10C	
Chemistry	TF1 9	Basic materials chemistry	T15 F19	Working-Up Of Peat	C10F	
Chemistry	TF1 9	Basic materials chemistry	T16 F19	Cracking Hydrocarbon Oils	C10G	
Chemistry	TF1 9	Basic materials chemistry	T17 F19	Wet Production Of Acetylene	C10H	
Chemistry	TF1 9	Basic materials chemistry	T18 F19	Production Of Gas From Solid C-Material	C10J	
Chemistry	TF1 9	Basic materials chemistry	T19 F19	Modifying Combustible Gases Containing CO	C10K	
Chemistry	TF1 9	Basic materials chemistry	T20 F19	Fuels	C10L	
Chemistry	TF1 9	Basic materials chemistry	T21 F19	Lubrifiant	C10M	
Chemistry	TF1 9	Basic materials chemistry	T22 F19	Lubrifiant indexed	C10N	
Chemistry	TF1 9	Basic materials chemistry	T23 F19	Fats	C11B	
Chemistry	TF1 9	Basic materials chemistry	T24 F19	Fatty Acids	C11C	
Chemistry	TF1 9	Basic materials chemistry	T25 F19	Detergent Compositions	C11D	
Chemistry	TF1 9	Basic materials chemistry	T26 F19	Other subjects Chemistry	C99Z	
Chemistry	TF2 0	Materials, metallurgy	T01 F20	Casting and Powder Metallurgy	B22	
Chemistry	TF2 0	Materials, metallurgy	T02 F20	Inorganic Chemistry	C01	
Chemistry	TF2 0	Materials, metallurgy	T03 F20	Glasses	C03C	
Chemistry	TF2 0	Materials, metallurgy	T04 F20	Refractories	C04	
Chemistry	TF2 0	Materials, metallurgy	T05 F20	Metallurgy of Iron	C21	
Chemistry	TF2 0	Materials, metallurgy	T06 F20	Metallurgy	C22	
Chemistry	TF2 1	Surface technology, coating	T01 F21	Apparatus For Applying Liquids	B05C	
Chemistry	TF2 1	Surface technology, coating	T02 F21	Liquid Application Processes	B05D	
Chemistry	TF2 1	Surface technology, coating	T03 F21	Layered Products	B32	
Chemistry	TF2 1	Surface technology,	T04 F21	Coating Metallic Material	C23	

		coating					
Chemistry	TF2 1	Surface technology, coating	T05 F21	Electrolytic Electrophoretic Processes	Or C25		
Chemistry	TF2 1	Surface technology, coating	T06 F21	Crystal Growth	C30		
Chemistry	TF2 2	Micro-structural and nano- technology	T01 F22	Micro-Structural Technology	B81		
Chemistry	TF2 2	Micro-structural and nano- technology	T02 F22	Nano-Technology	B82		
Chemistry	TF2 3	Chemical engineering	T01 F23	Boiling	B01B		
Chemistry	TF2 3	Chemical engineering	T02 F23	Separation	B01D		
Chemistry	TF2 3	Chemical engineering	T03 F23	Evaporating and extraction and separation and degaseification	B01D 1		
Chemistry	TF2 3	Chemical engineering	T04 F23	Separation and Filters	B01D 2		
Chemistry	TF2 3	Chemical engineering	T05 F23	Distillation	B01D 3		
Chemistry	TF2 3	Chemical engineering	T06 F23	Regeneration of Filters	B01D 41		
Chemistry	TF2 3	Chemical engineering	T07 F23	Separating Liquids and Solids	B01D 43		
Chemistry	TF2 3	Chemical engineering	T08 F23	Other Separation	B01D 57		
Chemistry	TF2 3	Chemical engineering	T09 F23	Separation of isotopes	B01D 59		
Chemistry	TF2 3	Chemical engineering	T10 F23	Separation apparatus and processes	B01D 6		
Chemistry	TF2 3	Chemical engineering	T11 F23	Semi-permeable membranes for separation	B01D 7		
Chemistry	TF2 3	Chemical engineering	T12 F23	Mixing	B01F		
Chemistry	TF2 3	Chemical engineering	T13 F23	Chem and Phys Processes	B01J		
Chemistry	TF2 3	Chemical engineering	T14 F23	Chem Or Phys Lab Apparatus	B01L		
Chemistry	TF2 3	Chemical engineering	T15 F23	Crushing and Milling	B02C		
Chemistry	TF2 3	Chemical engineering	T16 F23	Separation of solid materials	B03		
Chemistry	TF2 3	Chemical engineering	T17 F23	Centrifuges	B04		
Chemistry	TF2 3	Chemical engineering	T18 F23	Spraying	B05B		
Chemistry	TF2 3	Chemical engineering	T19 F23	Generating Mechanical Vibrations	B06B		
Chemistry	TF2 3	Chemical engineering	T20 F23	Separating Solids and Sorting	B07		

Chemistry	TF2 3	Chemical engineering	T21 F23	Cleaning	B08		
Chemistry	TF2 3	Chemical engineering	T22 F23	Chemical Processing Of Skins Hides Or Leather	C14C		
Chemistry	TF2 3	Chemical engineering	T23 F23	Treating Textile	D06B		
Chemistry	TF2 3	Chemical engineering	T24 F23	Finishing Dressing Tentering Or Stretching Textile	D06C		
Chemistry	TF2 3	Chemical engineering	T25 F23	Bleaching	D06L		
Chemistry	TF2 3	Chemical engineering	T26 F23	Phase transformation Of Gases	F25J		
Chemistry	TF2 3	Chemical engineering	T27 F23	Drying	F26		
Chemistry	TF2 3	Chemical engineering	T28 F23	Plasma Technique	H05H		
Chemistry	TF2 4	Environmental technology	T01 F24	Fire-Fighting	A62C		
Chemistry	TF2 4	Environmental technology	T02 F24	Separating Gases	B01D 45		
Chemistry	TF2 4	Environmental technology	T03 F24	Filtering Gases	B01D 46		
Chemistry	TF2 4	Environmental technology	T04 F24	Separating by Liquids	B01D 47		
Chemistry	TF2 4	Environmental technology	T05 F24	Other Separating	B01D 49		
Chemistry	TF2 4	Environmental technology	T06 F24	Separating Combination	B01D 50		
Chemistry	TF2 4	Environmental technology	T07 F24	Pretreatment of gases	B01D 51		
Chemistry	TF2 4	Environmental technology	T08 F24	Treatment of gases	B01D 52		
Chemistry	TF2 4	Environmental technology	T09 F24	Separation of gases	B01D 53		
Chemistry	TF2 4	Environmental technology	T10 F24	Solid Waste and Contaminated Soils	B09		
Chemistry	TF2 4	Environmental technology	T11 F24	Gathering of Domestic Refuse	B65F		
Chemistry	TF2 4	Environmental technology	T12 F24	Water Treatment	C02		
Chemistry	TF2 4	Environmental technology	T13 F24	Absorbing noise from roads	E01F 8		
Chemistry	TF2 4	Environmental technology	T14 F24	Gas-Flow Silencers	F01N		
Chemistry	TF2 4	Environmental technology	T15 F24	Furnaces	F23G		
Chemistry	TF2 4	Environmental technology	T16 F24	Combustion Products	F23J		
Chemistry	TF2 4	Environmental technology	T17 F24	Measuring Nuclear Or X- Radiation	G01T		
Mechanical engineering	TF2 5	Handling	T01 F25	Manipulators	B25J		
Mechanical	TF2	Handling	T02	Packaging Machines	B65B		

engineering	5		F25				
Mechanical engineering	TF2		T03				
engineering	5	Handling	F25	Labelling Machines	B65C		
Mechanical engineering	TF2		T04	Containers for Storage of			
engineering	5	Handling	F25	Articles	B65D		
Mechanical engineering	TF2		T05	Transport of Storage			
engineering	5	Handling	F25	Devices	B65G		
Mechanical engineering	TF2		T06	Handling Thin Material	B65H		
engineering	5	Handling	F25	Lifting	B66		
Mechanical engineering	TF2		T08	Opening Bottles	B67		
engineering	5	Handling	F25				
Mechanical engineering	TF2		T01	Protection against Fire and			
engineering	6	Machine tools	F26	Chemicals	A62D		
Mechanical engineering	TF2		T02	Mechanical Metal-Working	B21		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T03	Machine Tools	B23		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T04	Grinding and Polishing	B24		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T05	Tools Or Benches	B25B		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T06	Nailing Or Stapling Tools	B25C		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T07	Percussive Tools	B25D		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T08	Multi-Purpose Tools	B25F		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T09	Handles For Hand			
engineering	6	Machine tools	F26	Implements	B25G		
Mechanical engineering	TF2		T10	Workshop Equipment	B25H		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T11	Other Hand-Held Cutting			
engineering	6	Machine tools	F26	Tools	B26B		
Mechanical engineering	TF2		T12	Cutting	B26D		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T13	Perforating	B26F		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T14	Working Wood	B27		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2		T15	Presses	B30		
engineering	6	Machine tools	F26				
Mechanical engineering	TF2	Engines, pumps, turbines	T01	Machines Or Engines for +			
engineering	7		F27	Displacement	F01B		
Mechanical engineering	TF2	Engines, pumps, turbines	T02				
engineering	7		F27	Piston Machines Or Engines	F01C		
Mechanical engineering	TF2	Engines, pumps, turbines	T03	Machines Or Engines for -			
engineering	7		F27	Displacement	F01D		
Mechanical engineering	TF2	Engines, pumps,	T04				
engineering	7	pumps,	F27	Steam Engine Plants	F01K		

		turbines					
Mechanical engineering	TF2 7	Engines, pumps, turbines	T05 F27	Cyclic Machines Or Engines	F01L		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T06 F27	Lubricating Of Machines Or Engines	F01M		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T07 F27	Cooling Of Machines Or Engines	F01P		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T08 F27	Combustion Engines In General	F02		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T09 F27	Propulsive Machines Or Engines For Liquids or Wind	F03		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T10 F27	Machines For Liquids Pumps Piston	F04		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T11 F27	Generating Combustion Products Of High P Or High V	F23R		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T12 F27	Other subjects Engines and Pumps	F99Z		
Mechanical engineering	TF2 7	Engines, pumps, turbines	T13 F27	Nuclear Eng.	G21		
Mechanical engineering	TF2 8	Textile and paper machines	T01 F28	Methofs for Clothes	A41H		
Mechanical engineering	TF2 8	Textile and paper machines	T02 F28	Repairing Footwear	A43D		
Mechanical engineering	TF2 8	Textile and paper machines	T03 F28	Production of Brushes	A46D		
Mechanical engineering	TF2 8	Textile and paper machines	T04 F28	Working Paper	B31		
Mechanical engineering	TF2 8	Textile and paper machines	T05 F28	Printing	B41		
Mechanical engineering	TF2 8	Textile and paper machines	T06 F28	Mechanical Processing Of Skins Hides Or Leather	C14B		
Mechanical engineering	TF2 8	Textile and paper machines	T07 F28	Threads or Fibers	D01		
Mechanical engineering	TF2 8	Textile and paper machines	T08 F28	Crimping of Threads or Fibers	D02		
Mechanical engineering	TF2 8	Textile and paper	T09 F28	Shedding Mechanisms	D03		

		machines					
Mechanical engineering	TF2 8	Textile and paper machines	T10 F28	Knitting	D04B		
Mechanical engineering	TF2 8	Textile and paper machines	T11 F28	Braiding	D04C		
Mechanical engineering	TF2 8	Textile and paper machines	T12 F28	Making Nets	D04G		
Mechanical engineering	TF2 8	Textile and paper machines	T13 F28	Textile Fabrics	D04H		
Mechanical engineering	TF2 8	Textile and paper machines	T14 F28	Sewing	D05		
Mechanical engineering	TF2 8	Textile and paper machines	T15 F28	Mechanical Or Pressure Cleaning of Textile	D06G		
Mechanical engineering	TF2 8	Textile and paper machines	T16 F28	Marking Inspecting Seaming Or Severing Textile	D06H		
Mechanical engineering	TF2 8	Textile and paper machines	T17 F28	Pleating Kilting Or Goffering Textile	D06J		
Mechanical engineering	TF2 8	Textile and paper machines	T18 F28	Other Treatments	D06M		
Mechanical engineering	TF2 8	Textile and paper machines	T19 F28	Dyeing Or Printing Textiles	D06P		
Mechanical engineering	TF2 8	Textile and paper machines	T20 F28	Decorating Textiles	D06Q		
Mechanical engineering	TF2 8	Textile and paper machines	T21 F28	Paper Making	D21		
Mechanical engineering	TF2 8	Textile and paper machines	T22 F28	Other Subjects textiles and papers	D99Z		
Mechanical engineering	TF2 9	Other special machines	T01 F29	Soil Working	A01B		
Mechanical engineering	TF2 9	Other special machines	T02 F29	Planting	A01C		
Mechanical engineering	TF2 9	Other special machines	T03 F29	Harvesting	A01D		
Mechanical engineering	TF2 9	Other special machines	T04 F29	Harvested Produce	A01F		
Mechanical engineering	TF2 9	Other special machines	T05 F29	Horticulture	A01G		
Mechanical engineering	TF2 9	Other special machines	T06 F29	Production of Dairy Products	A01J		
Mechanical engineering	TF2 9	Other special machines	T07 F29	Husbandry	A01K		

Mechanical engineering	TF2 9	Other special machines	T08 F29	Shoeing Of Animals	A01L		
Mechanical engineering	TF2 9	Other special machines	T09 F29	Traps for animals	A01M		
Mechanical engineering	TF2 9	Other special machines	T10 F29	Baking	A21B		
Mechanical engineering	TF2 9	Other special machines	T11 F29	Dough	A21C		
Mechanical engineering	TF2 9	Other special machines	T12 F29	Treatment of Meat	A22		
Mechanical engineering	TF2 9	Other special machines	T13 F29	Process of Harvests	A23N		
Mechanical engineering	TF2 9	Other special machines	T14 F29	Working Of Foodstuffs	A23P		
Mechanical engineering	TF2 9	Other special machines	T15 F29	Preparing Grain and Fruit	B02B		
Mechanical engineering	TF2 9	Other special machines	T16 F29	Working Cement or Stone	B28		
Mechanical engineering	TF2 9	Other special machines	T17 F29	Working of Plastics	B29		
Mechanical engineering	TF2 9	Other special machines	T18 F29	Other Micro and Nao Technology	B99Z		
Mechanical engineering	TF2 9	Other special machines	T19 F29	Manufacturind or Shaping	C03B		
Mechanical engineering	TF2 9	Other special machines	T20 F29	Processes Of Compounding	C08J		
Mechanical engineering	TF2 9	Other special machines	T21 F29	Pitching Machines	C12L		
Mechanical engineering	TF2 9	Other special machines	T22 F29	Reducing the size of material from which sugar is to be extracted	C13B 5		
Mechanical engineering	TF2 9	Other special machines	T23 F29	Expressing water from material from which sugar has been extracted	C13B 15		
Mechanical engineering	TF2 9	Other special machines	T24 F29	Evaporating or boiling sugar juices	C13B 25		
Mechanical engineering	TF2 9	Other special machines	T25 F29	Cutting machines specially adapted for sugar	C13B 45		
Mechanical engineering	TF2 9	Other special machines	T26 F29	Sugar Production	C13C		
Mechanical engineering	TF2 9	Other special machines	T27 F29	Evaporation Apparatus	C13G		
Mechanical engineering	TF2 9	Other special machines	T28 F29	Cutting of Sugar	C13H		
Mechanical engineering	TF2 9	Other special machines	T29 F29	Weapons	F41		
Mechanical engineering	TF2 9	Other special machines	T30 F29	Ammunitions Blasting	F42		
Mechanical engineering	TF3 0	Thermal processes and apparatus	T01 F30	Steam generation	F22		
Mechanical engineering	TF3 0	Thermal processes and apparatus	T02 F30	Combustion Apparatus Using Solid Fuel	F23B		

Mechanical engineering	TF30	Thermal processes and apparatus	T03 F30	Combustion Apparatus Using Fluent Fuel	F23C		
Mechanical engineering	TF30	Thermal processes and apparatus	T04 F30	Burners	F23D		
Mechanical engineering	TF30	Thermal processes and apparatus	T05 F30	Grates	F23H		
Mechanical engineering	TF30	Thermal processes and apparatus	T06 F30	Feeding Fuel To Combustion Apparatus	F23K		
Mechanical engineering	TF30	Thermal processes and apparatus	T07 F30	Air Supply	F23L		
Mechanical engineering	TF30	Thermal processes and apparatus	T08 F30	Combustion Chambers	F23M		
Mechanical engineering	TF30	Thermal processes and apparatus	T09 F30	Controlling Combustion	F23N		
Mechanical engineering	TF30	Thermal processes and apparatus	T10 F30	Extinguishing Devices	F23Q		
Mechanical engineering	TF30	Thermal processes and apparatus	T11 F30	Domestic Stoves For Solid Fuels	F24		
Mechanical engineering	TF30	Thermal processes and apparatus	T12 F30	Refrigeration Systems	F25B		
Mechanical engineering	TF30	Thermal processes and apparatus	T13 F30	Ice	F25C		
Mechanical engineering	TF30	Thermal processes and apparatus	T14 F30	Furnaces ovens	F27		
Mechanical engineering	TF30	Thermal processes and apparatus	T15 F30	Heat Exchange	F28		
Mechanical engineering	TF31	Mechanical elements	T01 F31	Fluid Pressure Actuator	F15		
Mechanical engineering	TF31	Mechanical elements	T02 F31	Engineering Elts or Units	F16		
Mechanical engineering	TF31	Mechanical elements	T03 F31	Storing Distributing Non Solids	F17		
Mechanical engineering	TF31	Mechanical elements	T04 F31	Mechanical Control Systems	G05G		
Mechanical engineering	TF32	Transport	T01 F32	Vehicles	B60		
Mechanical engineering	TF32	Transport	T02 F32	Railways	B61		
Mechanical engineering	TF32	Transport	T03 F32	Land Vehicles	B62		
Mechanical engineering	TF32	Transport	T04 F32	Ships	B63B		



engineering	2		F32				
Mechanical engineering	TF3 2	Transport	T05 F32	Launching of Vessels	B63C		
Mechanical engineering	TF3 2	Transport	T06 F32	Weapons on Vessels	B63G		
Mechanical engineering	TF3 2	Transport	T07 F32	Marine Propulsion	B63H		
Mechanical engineering	TF3 2	Transport	T08 F32	Auxiliaries on Vessel	B63J		
Mechanical engineering	TF3 2	Transport	T09 F32	Aircraft	B64		
Other fields	TF3 3	Furniture, games	T01 F33	Furniture and Domestic Equipment	A47		
Other fields	TF3 3	Furniture, games	T02 F33	Sports and Games	A63		
Other fields	TF3 4	Other consumer goods	T01 F34	Tobacco	A24		
Other fields	TF3 4	Other consumer goods	T02 F34	Shirts	A41B		
Other fields	TF3 4	Other consumer goods	T03 F34	Corsets	A41C		
Other fields	TF3 4	Other consumer goods	T04 F34	Outerwear	A41D		
Other fields	TF3 4	Other consumer goods	T05 F34	Suspenders	A41F		
Other fields	TF3 4	Other consumer goods	T06 F34	Wigs	A41G		
Other fields	TF3 4	Other consumer goods	T07 F34	Headwear	A42		
Other fields	TF3 4	Other consumer goods	T08 F34	Footwear	A43B		
Other fields	TF3 4	Other consumer goods	T09 F34	Laces	A43C		
Other fields	TF3 4	Other consumer goods	T10 F34	Fasteners and Braselets	A44		
Other fields	TF3 4	Other consumer goods	T11 F34	Hand or Travelling Articles	A45		
Other fields	TF3 4	Other consumer goods	T12 F34	Brushes	A46B		
Other fields	TF3 4	Other consumer goods	T13 F34	Methods For Life-Saving	A62B		

Other fields	TF3 4	Other consumer goods	T14 F34	Other Life saving and Amusement	A99Z		
Other fields	TF3 4	Other consumer goods	T15 F34	Bookbinding	B42		
Other fields	TF3 4	Other consumer goods	T16 F34	Writing Implements	B43		
Other fields	TF3 4	Other consumer goods	T17 F34	Decorative Arts	B44		
Other fields	TF3 4	Other consumer goods	T18 F34	Saddlery and Upholstery	B68		
Other fields	TF3 4	Other consumer goods	T19 F34	Trimmings	D04D		
Other fields	TF3 4	Other consumer goods	T20 F34	Laundering	D06F		
Other fields	TF3 4	Other consumer goods	T21 F34	Wall Floor Covering	D06N		
Other fields	TF3 4	Other consumer goods	T22 F34	Rope non electric cable	D07		
Other fields	TF3 4	Other consumer goods	T23 F34	Cooling Or Freezing Apparatus	F25D		
Other fields	TF3 4	Other consumer goods	T24 F34	Organs Harmoniums	G10B		
Other fields	TF3 4	Other consumer goods	T25 F34	Pianos	G10C		
Other fields	TF3 4	Other consumer goods	T26 F34	Musical Instruments	G10D		
Other fields	TF3 4	Other consumer goods	T27 F34	Automatic Musical Instruments	G10F		
Other fields	TF3 4	Other consumer goods	T28 F34	Aids For Music	G10G		
Other fields	TF3 4	Other consumer goods	T29 F34	Electroponic Musical Instruments	G10H		
Other fields	TF3 4	Other consumer goods	T30 F34	Sound-Producing Devices	G10K		
Other fields	TF3 5	Civil engineering	T01 F35	Permanent ways	E01B		
Other fields	TF3	Civil	T02	Surfaces for roads and sport	E01C		

	5	engineering	F35	grounds			
Other fields	TF3 5	Civil engineering	T03 F35	Bridges	E01D		
Other fields	TF3 5	Civil engineering	T04 F35	Platforms or refuge islands	E01F 1		
Other fields	TF3 5	Civil engineering	T05 F35	Landing for helicopters	E01F 3		
Other fields	TF3 5	Civil engineering	T06 F35	Draining of roads	E01F 5		
Other fields	TF3 5	Civil engineering	T07 F35	Protection again snow or sand drifts	E01F 7		
Other fields	TF3 5	Civil engineering	T08 F35	Road signs or traffic signals	E01F 9		
Other fields	TF3 5	Civil engineering	T09 F35	Sensitive and restricting and safety in roads	E01F 1		
Other fields	TF3 5	Civil engineering	T10 F35	Street Land Cleaning	E01H		
Other fields	TF3 5	Civil engineering	T11 F35	Hydraulic Eng. Foundations Soil-Shifting	E02		
Other fields	TF3 5	Civil engineering	T12 F35	Water supply sewerage	E03		
Other fields	TF3 5	Civil engineering	T13 F35	Building	E04		
Other fields	TF3 5	Civil engineering	T14 F35	Locking Safing	E05		
Other fields	TF3 5	Civil engineering	T15 F35	Openings in Building Ladder	E06		
Other fields	TF3 5	Civil engineering	T16 F35	Earth Or Rock Drilling	E21		
Other fields	TF3 5	Civil engineering	T17 F35	Other Subjects Building	E99Z		

## Appendix 4: Building description for International Patent Classification

Patents have fine grained information to describe the technologies that are combined by the applications. The International Patent Classification (IPC) is a hierarchical [patent classification](#) system used to classify the content of [patents](#) in a uniform manner.

with approximately 70 000 subdivisions <sup>10</sup>. Each subdivision is described with a short text that we have collected.

Each classification symbol is of the form « A01B 1/00 ». The first letter represents the "section" consisting of a letter from A ("Human Necessities") to H ("Electricity"). Combined with a two-digit number, it represents the "class" (class A01 represents "Agriculture; forestry; animal husbandry; trapping; fishing"). The final letter makes up the "subclass" (subclass A01B represents "Soil working in agriculture or forestry; parts, details, or

<sup>10</sup> <https://www.wipo.int/classifications/ipc/en/>

accessories of agricultural machines or implements, in general"). The subclass is followed by a one-to-three-digit "group" number, an oblique stroke and a number of at least two digits representing a "main group" or "subgroup"  
(reference [https://en.wikipedia.org/wiki/International\\_Patent\\_Classification](https://en.wikipedia.org/wiki/International_Patent_Classification))

**For each classification symbol we have** concatenated the description of all levels, rebuilding the IPC hierarchy. Each patent receives the concatenated description of different levels of its IPC codes.

A01B 1/00 will be described as : Soil working in agriculture or forestry; parts, details, or accessories of agricultural machines or implements, in general. Agriculture; forestry; animal husbandry; trapping; fishing. Human Necessities.

Details on the used methodology is given in: [https://gitlab.com/cortext/patents/patstat/-/tree/version-2021/nomenclatures/ipc\\_descriptions](https://gitlab.com/cortext/patents/patstat/-/tree/version-2021/nomenclatures/ipc_descriptions)

## Appendix 5: Data model of PATSTAT2021

The data model shown below is the model of the PATSTAT 2021 database.

The Patstat documentation is available in [https://gitlab.com/cortex/patents/patstat/-/blob/version-2021/version\\_2021/DataCatalog\\_Global\\_v5.17.pdf](https://gitlab.com/cortex/patents/patstat/-/blob/version-2021/version_2021/DataCatalog_Global_v5.17.pdf)

