



**HAL**  
open science

## DOCUMENTATION OF RISIS DATASETS - RISIS Patent Database

Patricia Laurens, Philippe Laredo, Lionel Villard, Juan Pablo Ospina  
Delgado, Luis Daniel Medina, Antoine Schoen

► **To cite this version:**

Patricia Laurens, Philippe Laredo, Lionel Villard, Juan Pablo Ospina Delgado, Luis Daniel Medina, et al.. DOCUMENTATION OF RISIS DATASETS - RISIS Patent Database. [Technical Report] LISIS UPEM; RISIS. 2019. hal-02361266

**HAL Id: hal-02361266**

**<https://hal.science/hal-02361266>**

Submitted on 13 Nov 2019

**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.

# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES

## DOCUMENTATION OF RISIS DATASETS *RISIS Patent*

*Written by P. Laurens (UPEM), P. Laredo (UPEM) 07/2019*

Risis Patent was constructed by L. Villard and J.P. Ospina Delgado (UPEM).  
P. Laredo, P. Laurens, L.D. Medina and A. Schoen have also contributed



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



## 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 .....	5
2.4	Sectorial, temporal and geographical coverage.....	6
2.5	Quality and accuracy of data.....	12
3	Technical Specifications .....	16
3.1	Information on the data base system.....	16
3.2	Technical variable definition.....	17
3.3	Description of the Entity Relationship Model .....	20
3.4	Interfaces for access and to other infrastructures .....	21
4	Scientific use and main references.....	24
	Appendix.....	0
	Appendix 1: Artificial patents.....	1
	Appendix 2: Type of applicants.....	3
	Appendix 3: Classification of patent technologies .....	5
	Appendix 4: Building description for International Patent Classification.....	19
	Appendix 5: Data model of PATSTAT2017 .....	21

## 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 insure a good coverage of the priority patents, that is the reintroduction in the database of so-called artificial priority patents (13,3% of total priority patents).

Overall 16,406,977 priority patents of inventions were applied for from 2000 to 2015<sup>1</sup>. They include 39,761,538 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.
- **The legal organisations that apply for patents (the applicants).** We use available information proposed by PATSTAT for the harmonisation of applicants' names and allocation of assignee sectors. 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

---

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

<sup>2</sup> Corresponding to 9,096,610 distinct `person_id` and 6,057,257 distinct `doc_std_name_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.

public research organisations (ORGREG covering universities, PROs and research hospitals). This will be implemented step by step in the next RISIS Patent release.

- **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 2000-2015 (using the PATSTAT 2017 Edition April as a source). The first release is in September 2019. A second one will occur in spring 2020. The database will be only periodically updated using new PATSTAT version (the next update should be in 2021). The database is developed and maintained by UPEM. The database is located on the servers of UPEM at Marne La Vallée, 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.

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 16,406,977 priority patents are distributed across three main types of applications:

- a large category of patents A, i.e. regular patents following the Paris convention (`appln_kind = A`): 15,030,308 applications,
- provisional patent, i.e. preliminary and temporary patents proposed in some patent offices like USPTO (`appln_kind = P`) : 1,146,128 applications,
- PCT patent in the international phase (`appln_kind = W`): 224 188 applications.

We also include priority patents tagged as artificial priority patents due to priorities in PATSTAT (12,7% of the overall priorities but 41,1% in the priorities applied at the US patent office and 33,8% the priorities applied at the EPO). As these latter documents only contain the filling 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).

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

<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.

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 2,373,199 patents (15 % of the patent applications) applied only by natural persons have been discarded. The resulting perimeter of the current version of the database is 13,333,585 patent applications.

When a co-application of patent includes both legal and natural persons, the patent is maintained in the database with its legal applicant only. This is the case for 1,252,657 patent applications, i.e. 8 % of the overall applications.

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. In 0.8% of the case, the type of applicant (legal or person) could not be determined. These applicants are tagged as Unknown (see Appendix 2 for the technical developments).

A fractional counting was employed to calculate the number of legal applicants in a patent applications: All applicants with a `fract_applt` 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 April2017). 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,
- 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 (persons 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),
- 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.

### 3- Filtering of patents

We select only priority patents of inventions from 2000 to 2015 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, name and ID of the group firm (GUO), presence or not in the CHEETAH database, ID of the Cheetah firm, presence or not in the VICO database and ID of the VICO firm; presence in firmreg database and Firmreg ID.
- For each applicant: The applicant address, The applicant geo coordinates, the applicant urban and rural cluster it belongs to
- For each inventor: the inventor address, the inventor geo coordinates, the inventor urban or rural cluster he/she belongs to

## 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

Sector of legal applicants	Number of applicants	Share of applicants
COMPANY	615,724	82,6%
COMPANY GOV NON-PROFIT	4,764	0,6%
COMPANY GOV NON-PROFIT UNIVERSITY	8	0,0%
COMPANY HOSPITAL	87	0,0%
COMPANY UNIVERSITY	112	0,0%

<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>).



GOV NON-PROFIT	21,368	2,9%
GOV NON-PROFIT HOSPITAL	3	0,0%
GOV NON-PROFIT UNIVERSITY	1,406	0,2%
HOSPITAL	1,233	0,2%
INDIVIDUAL	13,571	1,8%
UNIVERSITY	17,041	2,3%
UNIVERSITY HOSPITAL	2	0,0%
UNKNOWN	70,548	9,5%
TOTAL	745,879	100,0%

### *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	4,524,707	34,6%
TD02	Instruments	1,949,793	14,9%
TD03	Chemistry	2,692,039	20,6%
TD04	Mechanical engineering	2,922,951	22,3%
TD05	Other fields	994,353	7,6%
Total		13,083,843	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	955,139	7,3%
TF02	Audio-visual technology	583,762	4,5%
TF03	Telecommunications	434,316	3,3%
TF04	Digital communication	562,319	4,3%
TF05	Basic communication processes	122,523	0,9%
TF06	Computer technology	1,057,050	8,1%
TF07	IT methods for management	219,395	1,7%
TF08	Semiconductors	590,203	4,5%

TF09	Optics	553,166	4,2%
TF10	Measurement	670,222	5,1%
TF11	Analysis of biological materials	69,652	0,5%
TF12	Control	248,645	1,9%
TF13	Medical technology	408,108	3,1%
TF14	Organic fine chemistry	238,838	1,8%
TF15	Biotechnology	198,223	1,5%
TF16	Pharmaceuticals	405,962	3,1%
TF17	Macromolecular chemistry, polymers	242,901	1,9%
TF18	Food chemistry	200,591	1,5%
TF19	Basic materials chemistry	323,096	2,5%
TF20	Materials, metallurgy	329,110	2,5%
TF21	Surface technology, coating	237,130	1,8%
TF22	Micro-structural and nano-technology	24,341	0,2%
TF23	Chemical engineering	278,038	2,1%
TF24	Environmental technology	213,809	1,6%
TF25	Handling	359,404	2,7%
TF26	Machine tools	384,030	2,9%
TF27	Engines, pumps, turbines	332,934	2,5%
TF28	Textile and paper machines	289,628	2,2%
TF29	Other special machines	381,389	2,9%
TF30	Thermal processes and apparatus	240,448	1,8%
TF31	Mechanical elements	374,598	2,9%
TF32	Transport	560,520	4,3%
TF33	Furniture, games	332,844	2,5%
TF34	Other consumer goods	209,155	1,6%
TF35	Civil engineering	452,353	3,5%
Total		13,083,843	100,0%

## - Temporal coverage

The database of this first release (summer 2019) covers patent applications from 2000 to 2015 (based upon PATSTAT2017 Version April). It is most likely that years 2016 and 2017 are still only partially filled. It includes 13,333,585 applications of priority patents

The number of PI priority patents applied every year is shown below. It has doubled during this 15 years. Year 2015 is probably not fully completed

Table 3: Number of priority patent applications over time

Filing year	Number of PI priority patent applications
2000	663,553
2001	665,235
2002	655,632

2003	670,504
2004	701,440
2005	736,782
2006	751,220
2007	779,526
2008	805,129
2009	778,235
2010	833,469
2011	908,657
2012	1,051,103
2013	1,172,448
2014	1,220,297
2015	(940,355) <sup>7</sup>
<b>Total</b>	<b>13,333,585</b>

## - 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 87,8% 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
Total	13,333,585	100,0%	12,7%
JP	4,421,954	33,2%	9,5%
CN	3,205,972	24,0%	1,0%
US	2,206,472	16,5%	41,1%
KR	1,265,278	9,5%	3,9%
DE	610,583	4,6%	9,9%
EP	278,071	2,1%	33,8%
TW	215,233	1,6%	2,3%
GB	193,746	1,5%	0,5%
FR	175,357	1,3%	3,0%
RU	174,021	1,3%	0,6%

<sup>7</sup> This number of patents in 2015 should be considered with caution. It should not be interpreted as a decreasing trend of patent application but rather as a lack of completeness of the database for this very year.

IT	92,378	0,7%	7,3%
AU	79,895	0,6%	31,2%
SE	34,249	0,3%	15,6%
PL	31,203	0,2%	1,2%
BR	27,072	0,2%	1,7%
ES	26,643	0,2%	5,4%
IN	24,830	0,2%	76,9%
NL	23,623	0,2%	12,5%
FI	22,686	0,2%	4,9%
UA	19,976	0,1%	2,4%
Other	204,343	1,5%	

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 27% 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	9,775,432	73,3%
2	1,146,279	8,6%
3	689,257	5,2%
4	544,891	4,1%
5	389,437	2,9%
More than 5	788,289	5,9%
Total	13,333,585	100,0%

40% of the inventions are protected in Japan (either in the first or secondary subsequent filings), 36% in China, 31 in US, 16% at EPO, 15% in Korea. 19% of the families include a PCT patent<sup>8</sup>.

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

Content of Inpadoc family <sup>9</sup>	Number of priority patents	Share of priority patents
Total	13,333,585	100%
US application	4,080,093	31%
EP application	2,170,029	16%

<sup>8</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.

<sup>9</sup> Only ipr\_type : PI is considered in inpadoc family.

JP application	5,362,865	40%
KR application	1,963,807	15%
CN application	4,733,883	36%
W application	2,541,004	19%
Transnational patent	3,558,153	27%

## *Geography of the inventions*

In 80% of the patents, the location of the inventions (inventors geographical location) is known and in 84.7% of them, the location of the applicant is indicated. The shares of patents according to the inventors and applicants addresses are shown in the tables below.

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
Total number of patents	13,333,585	
JP	4,443,833	33,3%
Unknown	2,661,403	20,0%
CN	2,435,340	18,3%
US	1,876,575	14,1%
KR	1,253,756	9,4%
DE	731,139	5,5%
TW	365,109	2,7%
FR	241,406	1,8%
RU	150,593	1,1%
GB	128,201	1,0%
CA	113,051	0,8%
IT	108,064	0,8%
NL	86,513	0,6%
CH	76,174	0,6%
SE	64,687	0,5%
IN	60,865	0,5%
IL	46,406	0,3%
AU	42,298	0,3%
FI	41,556	0,3%
ES	40,691	0,3%
AT	35,786	0,3%

Table 9: Number of priority patent applications according to the country of applicants

Country of applicant	Number of patents with applicant from the country	Share of patents with applicant from the country
Total number of patent	13,333,585	
JP	4,461,707	33,5%
Unknown	2,168,294	16,3%
US	1,910,643	14,3%
CN	1,419,625	10,6%
KR	1,260,410	9,5%
DE	697,682	5,2%
TW	317,614	2,4%
FR	234,668	1,8%
GB	185,910	1,4%
RU	144,748	1,1%
NL	103,813	0,8%
CH	100,988	0,8%
CA	94,427	0,7%
SE	77,818	0,6%
IT	70,122	0,5%
IL	55,721	0,4%
FI	47,115	0,4%
AU	41,071	0,3%
IN	38,090	0,3%
ES	36,146	0,3%
PL	32,018	0,2%

## 2.5 Quality and accuracy of data

Checking the data quality is not be 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)

Table 10: Share of priority patent applications without any information on the applicants according to the patent office for the 25<sup>th</sup> largest patent offices

Patent office	Number of PI patent applications	Share of PI patent applications without any applicant*
JP	4635006	1,3%
CN	4047877	2,2%
US	3004923	7,1%
KR	1700903	5,1%
DE	699263	0,1%
RU	380052	28,5%
GB	302037	0,0%
EP	291545	0,4%
TW	256584	1,9%
FR	207753	0,1%
IT	132644	0,2%
AU	107909	0,6%
BR	59205	3,1%
PL	43726	0,5%
SE	42100	0,9%
CA	40058	0,0%
ES	37619	0,1%
UA	34780	13,6%
IN	31792	3,1%
NL	30793	0,1%
FI	28474	0,1%
AT	21555	0,1%
CH	20595	0,4%
DK	16739	0,5%
IL	15631	3,2%
<b>Total</b>	<b>16406977</b>	<b>3,5%</b>

\* appln\_id not in table applt\_address\_ifris

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

Table 11: 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 PI patent applications	Share of PI patent applications without any inventor*	Share of PI patent applications without any IPC codes**
<b>Total</b>	<b>13333585</b>	<b>1,5%</b>	<b>1,8%</b>

JP	4421954	0,0%	0,1%
CN	3205972	0,1%	0,3%
US	2206472	0,1%	0,3%
KR	1265278	0,0%	0,8%
DE	610583	2,8%	0,1%
EP	278071	2,2%	0,2%
TW	215233	3,7%	1,8%
GB	193746	67,8%	53,6%
FR	175357	0,7%	0,5%
RU	174021	0,7%	0,0%
IT	92378	1,1%	64,6%
AU	79895	22,6%	17,7%
SE	34249	1,1%	1,9%
PL	31203	0,2%	0,5%
BR	27072	0,4%	1,4%
ES	26643	0,2%	0,2%
IN	24830	0,4%	1,6%
NL	23623	0,3%	0,1%
FI	22686	1,9%	3,8%
UA	19976	2,0%	0,1%
CA	19870	1,1%	0,2%
CH	16208	1,1%	0,2%
AT	15121	42,4%	4,0%
DK	14990	0,9%	0,5%
IL	10431	73,2%	25,0%

\* appln\_id not in table invt\_address\_ifris

\*\* appln\_id not in table ipc\_technology\_frac\_ifris

3- Patent applications with legal applicants with missing country of the inventors or missing geocoded address of inventors

Table 12: Share of priority patent applications with legal applicants 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	Nber of PI patent applications	Share of PI patent applications with at least one inventor without assigned country*	Share of PI patent applications with at least one inventor without geocoded address*
Total	13333585	21,5%	41,9%
JP	4421954	4,6%	14,9%
CN	3205972	69,8%	82,1%
US	2206472	0,4%	8,2%



KR	1265278	4,6%	75,2%
DE	610583	3,1%	41,1%
EP	278071	2,7%	12,6%
TW	215233	4,2%	73,3%
GB	193746	68,1%	86,2%
FR	175357	1,6%	10,4%
RU	174021	17,2%	97,5%
IT	92378	36,2%	65,4%
AU	79895	65,2%	67,3%
SE	34249	26,7%	32,0%
PL	31203	0,3%	82,8%
BR	27072	70,6%	78,7%
ES	26643	11,6%	76,8%
IN	24830	5,7%	31,4%
NL	23623	11,3%	45,1%
FI	22686	2,4%	7,5%
UA	19976	2,1%	99,0%
CA	19870	2,0%	63,7%
CH	16208	2,7%	40,0%
AT	15121	42,8%	69,7%
DK	14990	2,9%	29,5%
IL	10431	75,6%	85,8%

\* either no inventor or the inventor is known but missing country/missing geocoded information

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

Table 13: 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	Nber of PI patent applications	Share of PI patent applications with at least one applicant without assigned country*	Share of PI patent applications with at least one applicant without geocoded address*
Total	13333585	16,3%	30,9%
JP	4421954	1,3%	7,8%
CN	3205972	57,8%	66,3%
US	2206472	2,1%	9,0%
KR	1265278	2,9%	48,9%
DE	610583	0,1%	19,0%
EP	278071	0,1%	6,6%

TW	215233	0,6%	72,9%
GB	193746	22,3%	34,2%
FR	175357	0,1%	15,1%
RU	174021	16,7%	97,3%
IT	92378	47,3%	60,7%
AU	79895	45,5%	50,6%
SE	34249	0,1%	22,4%
PL	31203	0,1%	58,2%
BR	27072	0,1%	56,4%
ES	26643	3,5%	44,7%
IN	24830	2,6%	23,7%
NL	23623	9,7%	44,7%
FI	22686	0,1%	4,0%
UA	19976	0,1%	98,5%
CA	19870	0,5%	51,4%
CH	16208	0,8%	39,2%
AT	15121	0,4%	35,3%
DK	14990	0,4%	17,4%
IL	10431	40,5%	61,8%

\* either no applicant or the applicant is known but missing country/missing geocoded information

## Tests of consistency

### 1- Priority patents

We compare the `appln_id` of the patent with the `earliest_filing_id`. For priority patent applications (characterised by `appln_first_priority_year = 0`, a field added by UPEM), the `appln_id` and the `earliest_filing_id` should be identical.

This is not the case for 456 387 `appln_id` (3,42% of the 1 313 333 585 priority patent applications).

### 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 99,77% of the case, the country is the same.

## 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> . _A patent _W PCT application (in the international phase) _T used by some offices (e. g. AT, DE, DK, ES, GR, HR, PL, PT, SI, SM, TR) for applications which are "translations" of granted PCT or EP applications _P provisional application (US only) _F design patent _D2, D3 artificial _Other "exotic" kind codes
<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> Standard name attributed to applicant and inventor names for inclusion in DOCDB. 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

	<p>USPTO patents. 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>	<p><b>Date of the earliest filing</b></p> <p>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.</p>
<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>	<p><b>Granted indicator</b></p> <p>"1" if there exists a publication of the grant; "0" otherwise</p>
<b>IPC_CLASS_SYMBOL</b>	<p><b>IPC classification symbol (IPC 8th edition)</b></p> <p>Classification symbol according to the International Patent Classification, eighth edition (entered into force January 1, 2006)</p>
<b>NB_CITING_DOCDB_FAM</b>	<p><b>Number of forward citations on family level</b></p> <p>Number of distinct DOCDB simple families citing at least one of the publications or applications of the DOCDB simple family of the current application</p>
<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>	<p><b>PATSTAT standardized name</b></p> <p>The attribute is populated for all persons. Names of persons which have not been harmonized are just copied from the attribute PERSON_NAME.</p>
<b>PSN_SECTOR</b>	<p><b>Sector of the applicant</b></p> <p>INDIVIDUAL ; COMPANY ; UNKNOWN ; GOVERNMENT ; NON-PROFIT ; UNIVERSITY ; HOSPITAL</p>
<b>PUBLN_AUTH</b>	<p><b>Publication Authority or Publishing office</b></p> <p>Patent Authority that issued the publication of the application</p>
<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_ctry</b>	Addresses of inventors and applicants	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>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>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 protected at EPO, (info of Inpadoc family)	New info provided
<b>JP</b>	Priority patent for which the invention is protected at JPO, (info of Inpadoc family)	New info provided
<b>CN</b>	Priority patent for which the invention is protected	New info provided

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

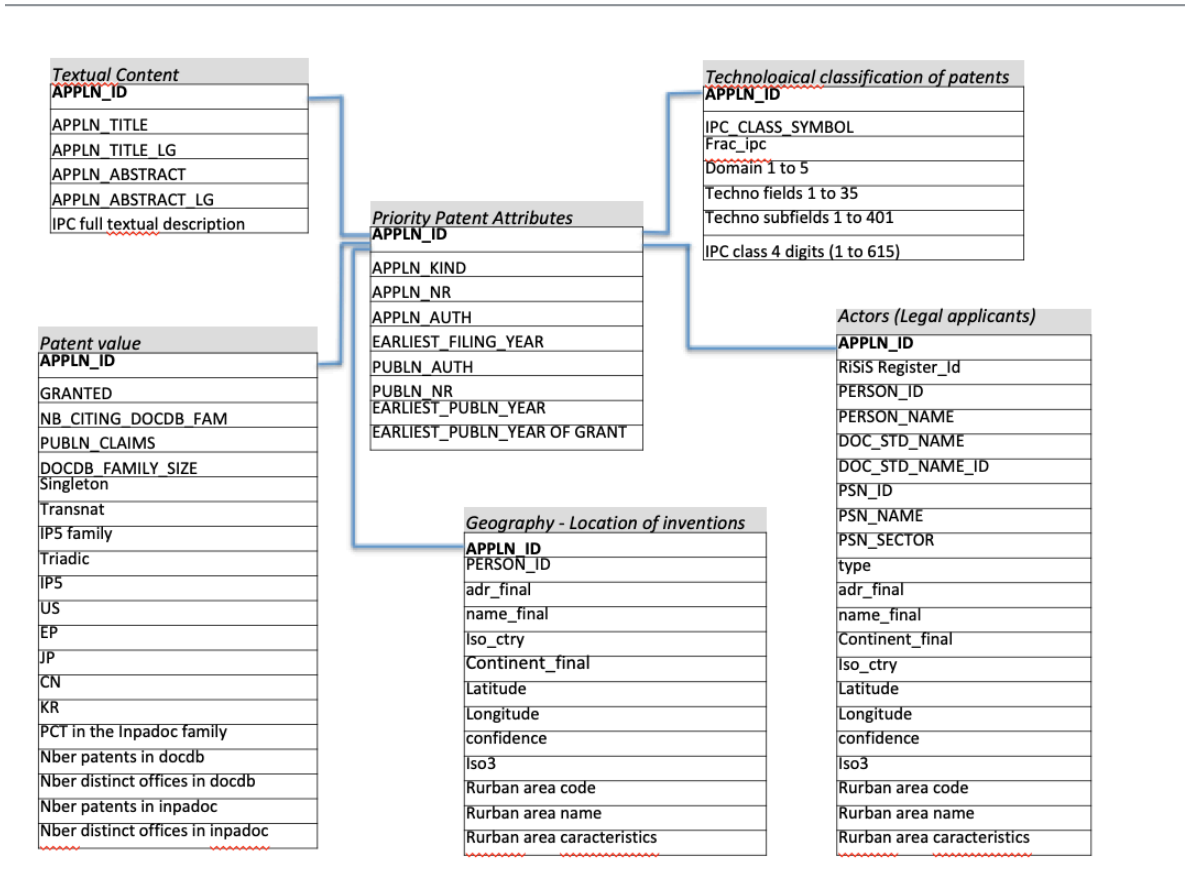
### 3.3 Description of the Entity Relationship Model

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 PATSTAT2017 database is shown in Appendix 5.

### 3.4 Interfaces for access and to other infrastructures

RISIS Patent will be inter-linked 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 rurban 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)
- We incorporate external resources available: REGPAT from OECD, addresses provided for by INPI, Japanese addresses (thanks to colleagues in Japan)
- 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.

The process allows to end in RISIS Patent with 75% of inventors with an address (to be compared with 10% in the initial raw PATSTAT data) and 67.4% of the addresses are geocoded. The share of geocoded addresses varies according to the countries (see table below). It exceeds 80% in most of the western countries.

Table 17: Steps for completing the addresses of inventors

Steps in filling inventor's addresses	Number of inventor's addresses	Share of inventor's addresses
PATSTAT2017 (raw data)	3,669,681	10,3%
REGPAT info	349,016	1,0%
INPI info	271,431	0,8%
JPO info	3,940,587	11,1%
Propagation of addresses within Inpadoc families	14,144,527	39,8%
Filling Artificial patent addresses	4,345,018	12,2%
Missing addresses in RISIS Patent	8,785,687	24,7%
Total	35,505,947	100,0%

Table 18: Completeness of the addresses of inventors and geocoding

Country of inventor	Raw PATSTAT data with geo info	Risis Patent with geo info	Share of addresses present in Risis Patent (more than country info)	Number of geocoded addresses in RISIS Patent	Share of geocoded addresses in RISIS Patent
All	3,669,681	35,505,949	75,3%	23,942,110	67,4%



JP	61,109	9,970,566	99,7%	9,310,142	93,4%
Unknown	718	6,150,087	5,9%	1,754	0,0%
CN	37,657	5,252,217	86,4%	4,237,224	80,7%
US	2,578,308	5,015,934	98,4%	4,840,760	96,5%
KR	14,155	3,093,326	77,1%	1,423,080	46,0%
DE	165,991	1,738,042	83,6%	1,387,630	79,8%
TW	122,250	728,935	72,4%	405,879	55,7%
RU	12,442	569,052	9,1%	41,581	7,3%
FR	52,261	500,759	94,9%	466,554	93,2%
CA	92,644	248,754	92,3%	209,858	84,4%
GB	50,807	235,926	82,3%	167,786	71,1%
IT	37,243	186,966	92,7%	127,946	68,4%
NL	30,187	182,245	92,9%	160,646	88,1%
IN	64,402	161,550	96,3%	137,160	84,9%
CH	41,613	146,093	90,4%	126,250	86,4%
SE	26,015	124,340	96,8%	115,794	93,1%
IL	38,094	112,700	97,4%	88,386	78,4%
ES	16,060	99,414	65,2%	54,141	54,5%
PL	15,601	97,771	29,0%	26,359	27,0%
FI	59,292	89,157	96,7%	84,030	94,2%
AU	12,781	82,815	97,8%	78,186	94,4%
UA	658	76,519	5,6%	2538	3,3%
DK	11,230	69,667	96,0%	57413	82,4%
BE	20,331	68,511	93,9%	60060	87,7%
AT	12,921	67,387	82,3%	53035	78,7%

## 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>10</sup>.

## 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

<sup>10</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>).

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 researcher from Université Paris-Est Marne-la-Vallée

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

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

### Recent References

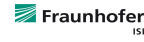
Laurens, P., Le Bas, C., Schoen, A. (2018), Worldwide IP coverage of patented inventions in largepharmafirms: to what extent do the internationalisation of R&D and firm strategy matter? Submitted to the International Journal of Technology Management

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<sup>2</sup>Patricia Laurens<sup>1</sup>, Alfredo Yegros<sup>3</sup>, Philippe Larjdo<sup>4,5</sup>  
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

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.

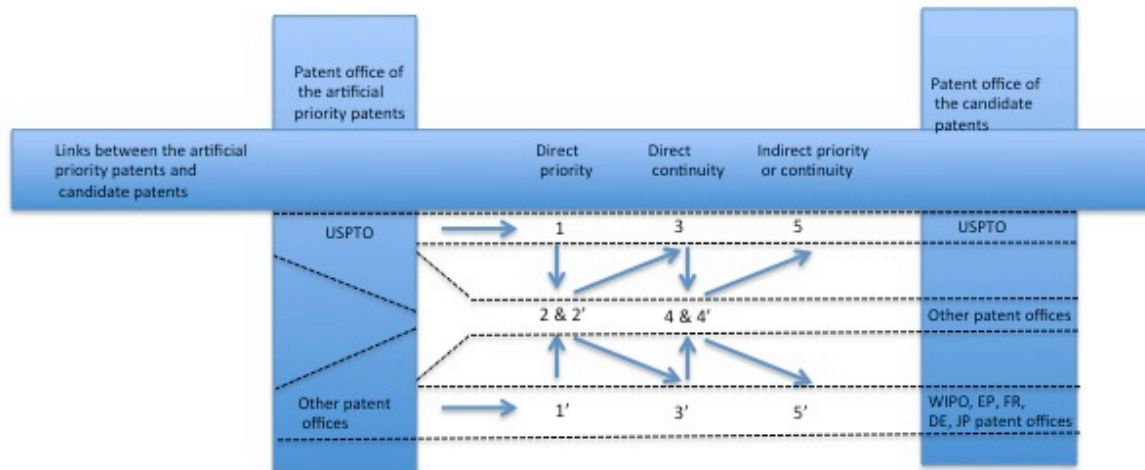
An artificial priority patent in its original state contains very scarce information. The only information included in the Patent database provided by EPO for these documents relates to the patent office where the artificial priority patent has been applied for, the date and the type of the application 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. 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://github.com/cortex/PATSTAT/blob/master/artificial/README.md>

## Reference:

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.

## 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 to 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 an 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 used of gazttees but on the characterization of the applicants through the relation with the patent, besides some simple lexical identifiers. Details of the methodology are described in:

<https://github.com/cortext/PATSTAT/tree/master/Actors>

Results are shown below. Almost 1/3 applicants are persons. They are present in 23% of the patents applications either alone (15% of the applications) or as co-applicant of legal entities (8%).

Table A2-1: Distribution of applicants according to the type of applicants

Type of applicants	Number of applicants	Share of applicants
Undefined	166,834	0,8%
Legal	14,355,395	67,8%
Person	6,646,629	31,4%
Total	21,168,858	100,0%

Table A2-2: Distribution of type of applicants in priority patent applications

Type of applicant	Number of patents including the type of applicants	Share of patents including the type of applicants
Undefined	154,161	1,0%
Legal	13,333,585	84,3%

# 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

<b>Person</b>	<b>3,625,856</b>	<b>22,9%</b>
<b>Total patents</b>	<b>15,824,851</b>	<b>100%</b>

## 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_appln_id
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T01F01	Lighting	F21		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T02F01	Cables with Special Electric Properties	H01B		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T03F01	Resistors	H01C		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T04F01	Magnets	H01F		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T05F01	Capacitors	H01G		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T06F01	Electric Switches	H01H		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T07F01	Discharge Lamps	H01J		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T08F01	Electric Incandescent Lamps	H01K		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T09F01	Batteries and related	H01M		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T10F01	Electrically-Conductive Connections	H01R		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T11F01	Spark Gaps	H01T		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T12F01	Boards For The Distribution Of Electricity	H02		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T13F01	Electric Heating	H05B		
Electrical	TF01	Electrical	T14F	Electrical Device stopping	H05C		



engineering		machinery, apparatus, energy	01	human beings			
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T15F 01	Static Electricity	H05F		
Electrical engineering	TF01	Electrical machinery, apparatus, energy	T16F 01	Other subjects electricity	H99Z		
Electrical engineering	TF02	Audio-visual technology	T01F 02	Displaying Advertising	G09F		
Electrical engineering	TF02	Audio-visual technology	T02F 02	Arrangements For Control	G09G		
Electrical engineering	TF02	Audio-visual technology	T03F 02	Information Storage Based Record Carrier	G11B		
Electrical engineering	TF02	Audio-visual technology	T04F 02	Scanning details of television	H04N 3		
Electrical engineering	TF02	Audio-visual technology	T05F 02	Details of television systems	H04N 5		
Electrical engineering	TF02	Telecommunicati ons	T06F 02	Television systems	H04N 7		
Electrical engineering	TF02	Audio-visual technology	T07F 02	Details of colour television systems	H04N 9		
Electrical engineering	TF02	Telecommunicati ons	T08F 02	Colour television systems	H04N 11		
Electrical engineering	TF02	Audio-visual technology	T09F 02	Stereoscopic television systems	H04N 13		
Electrical engineering	TF02	Audio-visual technology	T10F 02	Stereoscopic colour television systems	H04N 15		
Electrical engineering	TF02	Audio-visual technology	T11F 02	Diagnosis for television systems	H04N 17		
Electrical engineering	TF02	Audio-visual technology	T12F 02	Pictorial Communication	H04N 101		
Electrical engineering	TF02	Audio-visual technology	T13F 02	Loudspeakers Microphones	H04R		
Electrical engineering	TF02	Audio-visual technology	T14F 02	Stereophonic Systems	H04S		
Electrical engineering	TF02	Audio-visual technology	T15F 02	Printed Circuits	H05K		
Electrical engineering	TF03	Telecommunicati ons	T01F 03	Transmission Systems	G08C		
Electrical engineering	TF03	Telecommunicati ons	T02F 03	Waveguides	H01P		
Electrical engineering	TF03	Telecommunicati ons	T03F 03	Aerials	H01Q		
Electrical engineering	TF03	Telecommunicati ons	T04F 03	Transmission	H04B		
Electrical engineering	TF03	Telecommunicati ons	T05F 03	Broadcas Communication	H04H		
Electrical engineering	TF03	Telecommunicati ons	T06F 03	Multiplex Communication	H04J		
Electrical engineering	TF03	Telecommunicati ons	T07F 03	Secret Communication	H04K		
Electrical engineering	TF03	Telecommunicati ons	T08F 03	Telephonic Communication	H04M		
Electrical engineering	TF03	Telecommunicati ons	T09F 03	Transmission or reproduction of documents	H04N 1		
Electrical	TF03	Telecommunicati	T10F	Selecting (Switches/Relays)	H04Q		



engineering		ons	03				
Electrical engineering	TF04	Digital communication	T01F04	Transmission Of Digital Information	H04L		
Electrical engineering	TF04	Digital communication	T02F04	Selective content distribution	H04N21		
Electrical engineering	TF04	Digital communication	T03F04	Wireless Communication	H04W		
Electrical engineering	TF05	Basic communication processes	T01F05	Basic Electronic Circuitry	H03		
Electrical engineering	TF06	Computer technology	T01F06	Mechanically computed Digital Computers	G06	G06Q	
Electrical engineering	TF06	Computer technology	T02F06	Speech Anal. Or Synth.	G10L		
Electrical engineering	TF06	Computer technology	T03F06	Static Stores	G11C		
Electrical engineering	TF07	IT methods for management	T01F07	Data Processing Systems	G06Q		
Electrical engineering	TF08	Semiconductors	T01F08	Semiconductor Devices	H01L		
Instruments	TF09	Optics	T01F09	Optical Elements Systems	G02		
Instruments	TF09	Optics	T02F09	Photograph Apparatus	G03B		
Instruments	TF09	Optics	T03F09	Photographic Processes	G03C		
Instruments	TF09	Optics	T04F09	Apparatus For Processing Exposed Photographic Materials	G03D		
Instruments	TF09	Optics	T05F09	Photomechanics of Surfaces	G03F		
Instruments	TF09	Optics	T06F09	Photomechanical Production Of Surfaces	G03G		
Instruments	TF09	Optics	T07F09	Holography	G03H		
Instruments	TF09	Optics	T08F09	Devices Using Stimulated Emission	H01S		
Instruments	TF10	Measurement	T01F10	Measuring Length Thickness	G01B		
Instruments	TF10	Measurement	T02F10	Measuring Distances Levels	G01C		
Instruments	TF10	Measurement	T03F10	Multivariable Measuring	G01D		
Instruments	TF10	Measurement	T04F10	Measuring Volume Flow	G01F		
Instruments	TF10	Measurement	T05F10	Weighing	G01G		
Instruments	TF10	Measurement	T06F10	Measuring Vibrations	G01H		
Instruments	TF10	Measurement	T07F10	Optical Measurements	G01J		
Instruments	TF10	Measurement	T08F10	Measuring Temperature	G01K		
Instruments	TF10	Measurement	T09F10	Measuring Force Stress Torque Work	G01L		
Instruments	TF10	Measurement	T10F10	Testing Structures	G01M		
Instruments	TF10	Measurement	T11F10	Chemical Physical Analyses	G01N	G01N33	



Instruments	TF10	Measurement	T12F 10	Measuring Speed Acceleration	G01P		
Instruments	TF10	Measurement	T13F 10	Scanning-Probe Techniques	G01Q		
Instruments	TF10	Measurement	T14F 10	Measuring Electric Variables	G01R		
Instruments	TF10	Measurement	T15F 10	Radio Direction-Finding	G01S		
Instruments	TF10	Measurement	T16F 10	Geophysics	G01V		
Instruments	TF10	Measurement	T17F 10	Meteorology	G01W		
Instruments	TF10	Measurement	T18F 10	Horology	G04		
Instruments	TF10	Measurement	T19F 10	Details Of Instruments	G12B		
Instruments	TF10	Measurement	T20F 10	Other subjects nucleonics	G99Z		
Instruments	TF11	Analysis of biological materials	T01F 11	Material Analysis by Chem/Phys Properties	G01N 33		
Instruments	TF12	Control	T01F 12	Control Systems in General	G05B		
Instruments	TF12	Control	T02F 12	Regulating Non Electric Variables	G05D		
Instruments	TF12	Control	T03F 12	Regulating Electric Variables	G05F		
Instruments	TF12	Control	T04F 12	Checking Devices	G07		
Instruments	TF12	Control	T05F 12	Signalling Systems	G08B		
Instruments	TF12	Control	T06F 12	Traffic Control Systems	G08G		
Instruments	TF12	Control	T07F 12	Educational Or Demonstration Appliances	G09B		
Instruments	TF12	Control	T08F 12	Ciphering Apparatus	G09C		
Instruments	TF12	Control	T09F 12	Railway Or Like Time Or Fare Tables	G09D		
Instruments	TF13	Medical technology	T01F 13	Surgery-Diagnosis	A61B		
Instruments	TF13	Medical technology	T02F 13	Dentistry	A61C		
Instruments	TF13	Medical technology	T03F 13	Veterinary	A61D		
Instruments	TF13	Medical technology	T04F 13	Blood Vessel Filters	A61F		
Instruments	TF13	Medical technology	T05F 13	Furniture for Patients	A61G		
Instruments	TF13	Medical technology	T06F 13	Therapy Apparatus	A61H		
Instruments	TF13	Medical technology	T07F 13	Medical Containers	A61J		
Instruments	TF13	Medical technology	T08F 13	Methods For Sterilising	A61L		
Instruments	TF13	Medical technology	T09F 13	Devices for Introducing Media in Body	A61M		
Instruments	TF13	Medical technology	T10F 13	Electrotherapy	A61N		
Instruments	TF13	Medical	T11F1	X-Ray Technique	H05G		

		technology	3				
Chemistry	TFI4	Organic chemistry	fine T01F 14	Cosmetic Preparations	A61K 8		
Chemistry	TFI4	Organic chemistry	fine T02F 14	Use of Cosmetics	A61Q		
Chemistry	TFI4	Organic chemistry	fine T03F 14	Organic Chemistry	C07B		A61K
Chemistry	TFI4	Organic chemistry	fine T04F 14	Acyclic or Carbocyclic Compounds	C07C		A61K
Chemistry	TFI4	Organic chemistry	fine T05F 14	Heterocyclic Compounds	C07D		A61K
Chemistry	TFI4	Organic chemistry	fine T06F 14	Compounds with Elts Other Than C H Halogen O N S Se or Te	C07F		A61K
Chemistry	TFI4	Organic chemistry	fine T07F 14	Sugars Nucleotides	C07H		A61K
Chemistry	TFI4	Organic chemistry	fine T08F 14	Steroids	C07J		A61K
Chemistry	TFI4	Organic chemistry	fine T09F 14	Combinatorial Chemistry	C40B		A61K
Chemistry	TFI5	Biotechnology	T01F 15	Compounds Of Unknown Constitution	C07G		A61K
Chemistry	TFI5	Biotechnology	T02F 15	Peptides	C07K		A61K
Chemistry	TFI5	Biotechnology	T03F 15	Apparatus For Enzymology Or Microbiology	C12M		A61K
Chemistry	TFI5	Biotechnology	T04F 15	Micro-Organisms Enzymes Culture Media	C12N		A61K
Chemistry	TFI5	Biotechnology	T05F 15	Fermentation Or Enzyme-Using Processes	C12P		A61K
Chemistry	TFI5	Biotechnology	T06F 15	Measuring Processes with Enzymes Or Micro-Organisms	C12Q		A61K
Chemistry		Biotechnology	T07F 15	Micro-Organisms-General	C12R		A61K
Chemistry	TFI5	Biotechnology	T08F 15	Micro-Organisms To treat solid materials	C12S		A61K
Chemistry	TFI6	Pharmaceuticals	T01F 16	Medical Preparations	A61K 8		
Chemistry	TFI6	Pharmaceuticals	T02F 16	Therapeutic Activity Of Chemical Compounds	A61P		
Chemistry	TFI7	Macromolecular chemistry, polymers	T01F 17	Polysaccharides	C08B		
Chemistry	TFI7	Macromolecular chemistry, polymers	T02F 17	Treatment Of Rubbers	C08C		
Chemistry	TFI7	Macromolecular chemistry, polymers	T03F 17	Macromol with C-to-C Unsaturated Bonds	C08F		
Chemistry	TFI7	Macromolecular chemistry, polymers	T04F 17	Macromol without C-to-C Unsaturated Bonds	C08G		
Chemistry	TFI7	Macromolecular chemistry, polymers	T05F 17	Derivatives Of Natural Macromolecules	C08H		
Chemistry	TFI7	Macromolecular chemistry, polymers	T06F 17	Inorg Or Non-Macromolr Organ Subst	C08K		
Chemistry	TFI7	Macromolecular	T07F	Compositions Of	C08L		

		chemistry, polymers	17	Macromolecular Compounds			
Chemistry	TF18	Food chemistry	T01F 18	New Plants	A01H		
Chemistry	TF18	Food chemistry	T02F 18	Treatment of Dough	A21D		
Chemistry	TF18	Food chemistry	T03F 18	Preseving food	A23B		
Chemistry	TF18	Food chemistry	T04F 18	Dairy Products	A23C		
Chemistry	TF18	Food chemistry	T05F 18	Edible Oils	A23D		
Chemistry	TF18	Food chemistry	T06F 18	Coffee and Tea	A23F		
Chemistry	TF18	Food chemistry	T07F 18	Chocolate	A23G		
Chemistry	TF18	Food chemistry	T08F 18	Proteins for Food	A23J		
Chemistry	TF18	Food chemistry	T09F 18	Fodder	A23K		
Chemistry	TF18	Food chemistry	T10F 18	Other Foods	A23L		
Chemistry	TF18	Food chemistry	T11F 8	Brewing Of Beer	C12C		
Chemistry	TF18	Food chemistry	T12F 18	Fermented Solutions	C12F		
Chemistry	TF18	Food chemistry	T13F 18	Alcoholic Beverages	C12G		
Chemistry	TF18	Food chemistry	T14F 18	Treatment of Alcoholic Beverages	C12H		
Chemistry	TF18	Food chemistry	T15F 18	Vinegar	C12J		
Chemistry	TF18	Food chemistry	T16F 18	Production of sugar juices	C13B 10		
Chemistry	TF18	Food chemistry	T17F 18	Purification of sugar juices	C13B 20		
Chemistry	TF18	Food chemistry	T18F 18	Crystallisation of sugar	C13B 30		
Chemistry	TF18	Food chemistry	T19F 18	Extraction of sucrose from molasses	C13B 35		
Chemistry	TF18	Food chemistry	T20F 18	Drying sugar	C13B 40		
Chemistry	TF18	Food chemistry	T21F 18	Working-up of sugar	C13B 50		
Chemistry	TF18	Food chemistry	T22F 18	Subject matter not provided for in other groups of this subclass	C13B 99		
Chemistry	TF18	Food chemistry	T23F 18	Sugar Juices	C13D		
Chemistry	TF18	Food chemistry	T24F 18	Sugar	C13F		
Chemistry	TF18	Food chemistry	T25F 18	Extraction Of Sugar	C13J		
Chemistry	TF18	Food chemistry	T26F 18	Synthesis Of Sugars	C13K		
Chemistry	TF19	Basic materials chemistry	T01F 19	Disinfectants	A01N		
Chemistry	TF19	Basic materials chemistry	T02F 19	Repellants	A01P		
Chemistry	TF19	Basic materials	T03F	Fertilisers	C05		

		chemistry	19				
Chemistry	TF19	Basic materials chemistry	T04F19	Explosives	C06		
Chemistry	TF19	Basic materials chemistry	T05F19	Organic Dyes	C09B		
Chemistry	TF19	Basic materials chemistry	T06F19	Dyeing Of Inorganics	C09C		
Chemistry	TF19	Basic materials chemistry	T07F19	paints and inks	C09D		
Chemistry	TF19	Basic materials chemistry	T08F19	Natural Resins	C09F		
Chemistry	TF19	Basic materials chemistry	T09F19	Polishing Compositions	C09G		
Chemistry	TF19	Basic materials chemistry	T10F19	Glue Or Gelatine	C09H		
Chemistry	TF19	Basic materials chemistry	T11F19	Adhesives	C09J		
Chemistry	TF19	Basic materials chemistry	T12F19	Materials For Miscellaneous Applications	C09K		
Chemistry	TF19	Basic materials chemistry	T13F19	Production Of Gas Coke Tar by distillation	C10B		
Chemistry	TF19	Basic materials chemistry	T14F19	Working-Up Tar	C10C		
Chemistry	TF19	Basic materials chemistry	T15F19	Working-Up Of Peat	C10F		
Chemistry	TF19	Basic materials chemistry	T16F19	Cracking Hydrocarbon Oils	C10G		
Chemistry	TF19	Basic materials chemistry	T17F19	Wet Production Of Acetylene	C10H		
Chemistry	TF19	Basic materials chemistry	T18F19	Production Of Gas From Solid C-Material	C10J		
Chemistry	TF19	Basic materials chemistry	T19F19	Modifying Combustible Gases Containing CO	C10K		
Chemistry	TF19	Basic materials chemistry	T20F19	Fuels	C10L		
Chemistry	TF19	Basic materials chemistry	T21F19	Lubrifiant	C10M		
Chemistry	TF19	Basic materials chemistry	T22F19	Lubrifiant indexed	C10N		
Chemistry	TF19	Basic materials chemistry	T23F19	Fats	C11B		
Chemistry	TF19	Basic materials chemistry	T24F19	Fatty Acids	C11C		
Chemistry	TF19	Basic materials chemistry	T25F19	Detergent Compositions	C11D		
Chemistry	TF19	Basic materials chemistry	T26F19	Other subjects Chemistry	C99Z		
Chemistry	TF20	Materials, metallurgy	T01F20	Casting and Powder Metallurgy	B22		
Chemistry	TF20	Materials, metallurgy	T02F20	Inorganic Chemistry	C01		
Chemistry	TF20	Materials, metallurgy	T03F20	Glasses	C03C		
Chemistry	TF20	Materials, metallurgy	T04F20	Refractories	C04		
Chemistry	TF20	Materials, metallurgy	T05F20	Metallurgy of Iron	C21		
Chemistry	TF20	Materials, metallurgy	T06F20	Metallurgy	C22		
Chemistry	TF21	Surface	T01F	Apparatus For Applying Liquids	B05C		

		technology, coating	21				
Chemistry	TF21	Surface technology, coating	T02F 21	Liquid Application Processes	B05D		
Chemistry	TF21	Surface technology, coating	T03F 21	Layered Products	B32		
Chemistry	TF21	Surface technology, coating	T04F 21	Coating Metallic Material	C23		
Chemistry	TF21	Surface technology, coating	T05F 21	Electrolytic Or Electrophoretic Processes	C25		
Chemistry	TF21	Surface technology, coating	T06F 21	Crystal Growth	C30		
Chemistry	TF22	Micro-structural and nano- technology	T01F 22	Micro-Structural Technology	B81		
Chemistry	TF22	Micro-structural and nano- technology	T02F 22	Nano-Technology	B82		
Chemistry	TF23	Chemical engineering	T01F 23	Boiling	B01B		
Chemistry	TF23	Chemical engineering	T02F 23	Separation	B01D		
Chemistry	TF23	Chemical engineering	T03F 23	Evaporating and extraction and separation and degaseification	B01D 1		
Chemistry	TF23	Chemical engineering	T04F 23	Separation and Filters	B01D 2		
Chemistry	TF23	Chemical engineering	T05F 23	Distillation	B01D 3		
Chemistry	TF23	Chemical engineering	T06F 23	Regeneration of Filters	B01D 41		
Chemistry	TF23	Chemical engineering	T07F 23	Separating Liquids and Solids	B01D 43		
Chemistry	TF23	Chemical engineering	T08F 23	Other Separation	B01D 57		
Chemistry	TF23	Chemical engineering	T09F 23	Separation of isotopes	B01D 59		
Chemistry	TF23	Chemical engineering	T10F 23	Separation apparatus and processes	B01D 6		
Chemistry	TF23	Chemical engineering	T11F 23	Semi-permeable membranes for separation	B01D 7		
Chemistry	TF23	Chemical engineering	T12F 23	Mixing	B01F		
Chemistry	TF23	Chemical engineering	T13F 23	Chem and Phys Processes	B01J		
Chemistry	TF23	Chemical engineering	T14F 23	Chem Or Phys Lab Apparatus	B01L		
Chemistry	TF23	Chemical engineering	T15F 23	Crushing and Milling	B02C		
Chemistry	TF23	Chemical engineering	T16F 23	Separation of solid materials	B03		
Chemistry	TF23	Chemical engineering	T17F 23	Centrifuges	B04		
Chemistry	TF23	Chemical engineering	T18F 23	Spraying	B05B		

Chemistry	TF23	Chemical engineering	T19F 23	Generating Vibrations	Mechanical	B06B		
Chemistry	TF23	Chemical engineering	T20F 23	Separating Solids and Sorting		B07		
Chemistry	TF23	Chemical engineering	T21F 23	Cleaning		B08		
Chemistry	TF23	Chemical engineering	T22F 23	Chemical Processing Of Skins Hides Or Leather		C14C		
Chemistry	TF23	Chemical engineering	T23F 23	Treating Textile		D06B		
Chemistry	TF23	Chemical engineering	T24F 23	Finishing Dressing Tentering Or Stretching Textile		D06C		
Chemistry	TF23	Chemical engineering	T25F 23	Bleaching		D06L		
Chemistry	TF23	Chemical engineering	T26F 23	Phase transformation Of Gases		F25J		
Chemistry	TF23	Chemical engineering	T27F 23	Drying		F26		
Chemistry	TF23	Chemical engineering	T28F 23	Plasma Technique		H05H		
Chemistry	TF24	Environmental technology	T01F 24	Fire-Fighting		A62C		
Chemistry	TF24	Environmental technology	T02F 24	Separating Gases		B01D 45		
Chemistry	TF24	Environmental technology	T03F 24	Filtering Gases		B01D 46		
Chemistry	TF24	Environmental technology	T04F 24	Separating by Liquids		B01D 47		
Chemistry	TF24	Environmental technology	T05F 24	Other Separating		B01D 49		
Chemistry	TF24	Environmental technology	T06F 24	Separating Combination		B01D 50		
Chemistry	TF24	Environmental technology	T07F 24	Pretreatment of gases		B01D 51		
Chemistry	TF24	Environmental technology	T08F 24	Treatment of gases		B01D 52		
Chemistry	TF24	Environmental technology	T09F 24	Separation of gases		B01D 53		
Chemistry	TF24	Environmental technology	T10F 24	Solid Waste and Contaminated Soils		B09		
Chemistry	TF24	Environmental technology	T11F 24	Gathering of Domestic Refuse		B65F		
Chemistry	TF24	Environmental technology	T12F 24	Water Treatment		C02		
Chemistry	TF24	Environmental technology	T13F 24	Absorbing noise from roads		E01F 8		
Chemistry	TF24	Environmental technology	T14F 24	Gas-Flow Silencers		F01N		
Chemistry	TF24	Environmental technology	T15F 24	Furnaces		F23G		
Chemistry	TF24	Environmental technology	T16F 24	Combustion Products		F23J		
Chemistry	TF24	Environmental technology	T17F 24	Measuring Nuclear Or X-Radiation		G01T		
Mechanical engineering	TF25	Handling	T01F 25	Manipulators		B25J		
Mechanical engineering	TF25	Handling	T02F 25	Packaging Machines		B65B		
Mechanical engineering	TF25	Handling	T03F 25	Labelling Machines		B65C		





Mechanical engineering	TF25	Handling	T04F25	Containers for Storage of Articles	B65D		
Mechanical engineering	TF25	Handling	T05F25	Transport of Storage Devices	B65G		
Mechanical engineering	TF25	Handling	T06F25	Handling Thin Material	B65H		
Mechanical engineering	TF25	Handling	T07F25	Lifting	B66		
Mechanical engineering	TF25	Handling	T08F25	Opening Bottles	B67		
Mechanical engineering	TF26	Machine tools	T01F26	Protection against Fire and Chemicals	A62D		
Mechanical engineering	TF26	Machine tools	T02F26	Mechanical Metal-Working	B21		
Mechanical engineering	TF26	Machine tools	T03F26	Machine Tools	B23		
Mechanical engineering	TF26	Machine tools	T04F26	Grinding and Polishing	B24		
Mechanical engineering	TF26	Machine tools	T05F26	Tools Or Benches	B25B		
Mechanical engineering	TF26	Machine tools	T06F26	Nailing Or Stapling Tools	B25C		
Mechanical engineering	TF26	Machine tools	T07F26	Percussive Tools	B25D		
Mechanical engineering	TF26	Machine tools	T08F26	Multi-Purpose Tools	B25F		
Mechanical engineering	TF26	Machine tools	T09F26	Handles For Hand Implements	B25G		
Mechanical engineering	TF26	Machine tools	T10F26	Workshop Equipment	B25H		
Mechanical engineering	TF26	Machine tools	T11F26	Other Hand-Held Cutting Tools	B26B		
Mechanical engineering	TF26	Machine tools	T12F26	Cutting	B26D		
Mechanical engineering	TF26	Machine tools	T13F26	Perforating	B26F		
Mechanical engineering	TF26	Machine tools	T14F26	Working Wood	B27		
Mechanical engineering	TF26	Machine tools	T15F26	Presses	B30		
Mechanical engineering	TF27	Engines, pumps, turbines	T01F27	Machines Or Engines for + Displacement	F01B		
Mechanical engineering	TF27	Engines, pumps, turbines	T02F27	Piston Machines Or Engines	F01C		
Mechanical engineering	TF27	Engines, pumps, turbines	T03F27	Machines Or Engines for - Displacement	F01D		
Mechanical engineering	TF27	Engines, pumps, turbines	T04F27	Steam Engine Plants	F01K		
Mechanical engineering	TF27	Engines, pumps, turbines	T05F27	Cyclic Machines Or Engines	F01L		
Mechanical engineering	TF27	Engines, pumps, turbines	T06F27	Lubricating Of Machines Or Engines	F01M		
Mechanical engineering	TF27	Engines, pumps, turbines	T07F27	Cooling Of Machines Or Engines	F01P		
Mechanical engineering	TF27	Engines, pumps, turbines	T08F27	Combustion Engines In General	F02		
Mechanical engineering	TF27	Engines, pumps, turbines	T09F27	Propulsive Machines Or Engines For Liquids or Wind	F03		
Mechanical engineering	TF27	Engines, pumps, turbines	T10F27	Machines For Liquids Pumps Piston	F04		

Mechanical engineering	TF27	Engines, pumps, turbines	T11F 27	Generating Combustion Products Of High P Or High V	F23R		
Mechanical engineering	TF27	Engines, pumps, turbines	T12F 27	Other subjects Engines and Pumps	F99Z		
Mechanical engineering	TF27	Engines, pumps, turbines	T13F 27	Nuclear Eng.	G21		
Mechanical engineering	TF28	Textile and paper machines	T01F 28	Methofs for Clothes	A41H		
Mechanical engineering	TF28	Textile and paper machines	T02F 28	Repairing Footwear	A43D		
Mechanical engineering	TF28	Textile and paper machines	T03F 28	Production of Brushes	A46D		
Mechanical engineering	TF28	Textile and paper machines	T04F 28	Working Paper	B31		
Mechanical engineering	TF28	Textile and paper machines	T05F 28	Printing	B41		
Mechanical engineering	TF28	Textile and paper machines	T06F 28	Mechanical Processing Of Skins Hides Or Leather	C14B		
Mechanical engineering	TF28	Textile and paper machines	T07F 28	Threads or Fibers	D01		
Mechanical engineering	TF28	Textile and paper machines	T08F 28	Crimping of Threads or Fibers	D02		
Mechanical engineering	TF28	Textile and paper machines	T09F 28	Shedding Mechanisms	D03		
Mechanical engineering	TF28	Textile and paper machines	T10F 28	Knitting	D04B		
Mechanical engineering	TF28	Textile and paper machines	T11F 28	Braiding	D04C		
Mechanical engineering	TF28	Textile and paper machines	T12F 28	Making Nets	D04G		
Mechanical engineering	TF28	Textile and paper machines	T13F 28	Textile Fabrics	D04H		
Mechanical engineering	TF28	Textile and paper machines	T14F 28	Sewing	D05		
Mechanical engineering	TF28	Textile and paper machines	T15F 28	Mechanical Or Pressure Cleaning of Textile	D06G		
Mechanical engineering	TF28	Textile and paper machines	T16F 28	Marking Inspecting Seaming Or Severing Textile	D06H		
Mechanical engineering	TF28	Textile and paper machines	T17F 28	Pleating Kiltling Or Goffering Textile	D06J		
Mechanical engineering	TF28	Textile and paper machines	T18F 28	Other Treatments	D06M		
Mechanical engineering	TF28	Textile and paper machines	T19F 28	Dyeing Or Printing Textiles	D06P		
Mechanical engineering	TF28	Textile and paper machines	T20F 28	Decorating Textiles	D06Q		
Mechanical engineering	TF28	Textile and paper machines	T21F 28	Paper Making	D21		
Mechanical engineering	TF28	Textile and paper machines	T22F 28	Other Subjects textiles and papers	D99Z		
Mechanical engineering	TF29	Other special machines	T01F 29	Soil Working	A01B		
Mechanical engineering	TF29	Other special machines	T02F 29	Planting	A01C		
Mechanical engineering	TF29	Other special machines	T03F 29	Harvesting	A01D		
Mechanical engineering	TF29	Other special machines	T04F 29	Harvested Produce	A01F		
Mechanical engineering	TF29	Other special machines	T05F 29	Horticulture	A01G		

Mechanical engineering	TF29	Other machines	special	T06F29	Production of Dairy Products	A01J		
Mechanical engineering	TF29	Other machines	special	T07F29	Husbandry	A01K		
Mechanical engineering	TF29	Other machines	special	T08F29	Shoeing Of Animals	A01L		
Mechanical engineering	TF29	Other machines	special	T09F29	Traps for animals	A01M		
Mechanical engineering	TF29	Other machines	special	T10F29	Baking	A21B		
Mechanical engineering	TF29	Other machines	special	T11F29	Dough	A21C		
Mechanical engineering	TF29	Other machines	special	T12F29	Treatment of Meat	A22		
Mechanical engineering	TF29	Other machines	special	T13F29	Process of Harvests	A23N		
Mechanical engineering	TF29	Other machines	special	T14F29	Working Of Foodstuffs	A23P		
Mechanical engineering	TF29	Other machines	special	T15F29	Preparing Grain and Fruit	B02B		
Mechanical engineering	TF29	Other machines	special	T16F29	Working Cement or Stone	B28		
Mechanical engineering	TF29	Other machines	special	T17F29	Working of Plastics	B29		
Mechanical engineering	TF29	Other machines	special	T18F29	Other Micro and Nao Technology	B99Z		
Mechanical engineering	TF29	Other machines	special	T19F29	Manufacturind or Shaping	C03B		
Mechanical engineering	TF29	Other machines	special	T20F29	Processes Of Compounding	C08J		
Mechanical engineering	TF29	Other machines	special	T21F29	Pitching Machines	C12L		
Mechanical engineering	TF29	Other machines	special	T22F29	Reducing the size of material from which sugar is to be extracted	C13B5		
Mechanical engineering	TF29	Other machines	special	T23F29	Expressing water from material from which sugar has been extracted	C13B15		
Mechanical engineering	TF29	Other machines	special	T24F29	Evaporating or boiling sugar juices	C13B25		
Mechanical engineering	TF29	Other machines	special	T25F29	Cutting machines specially adapted for sugar	C13B45		
Mechanical engineering	TF29	Other machines	special	T26F29	Sugar Production	C13C		
Mechanical engineering	TF29	Other machines	special	T27F29	Evaporation Apparatus	C13G		
Mechanical engineering	TF29	Other machines	special	T28F29	Cutting of Sugar	C13H		
Mechanical engineering	TF29	Other machines	special	T29F29	Weapons	F41		
Mechanical engineering	TF29	Other machines	special	T30F29	Ammunitions Blasting	F42		
Mechanical engineering	TF30	Thermal processes apparatus	and	T01F30	Steam generation	F22		
Mechanical engineering	TF30	Thermal processes apparatus	and	T02F30	Combustion Apparatus Using Solid Fuel	F23B		
Mechanical engineering	TF30	Thermal processes apparatus	and	T03F30	Combustion Apparatus Using Fluent Fuel	F23C		

		apparatus					
Mechanical engineering	TF30	Thermal processes and apparatus	T04F30	Burners		F23D	
Mechanical engineering	TF30	Thermal processes and apparatus	T05F30	Grates		F23H	
Mechanical engineering	TF30	Thermal processes and apparatus	T06F30	Feeding Fuel To Combustion Apparatus		F23K	
Mechanical engineering	TF30	Thermal processes and apparatus	T07F30	Air Supply		F23L	
Mechanical engineering	TF30	Thermal processes and apparatus	T08F30	Combustion Chambers		F23M	
Mechanical engineering	TF30	Thermal processes and apparatus	T09F30	Controlling Combustion		F23N	
Mechanical engineering	TF30	Thermal processes and apparatus	T10F30	Extinguishing Devices		F23Q	
Mechanical engineering	TF30	Thermal processes and apparatus	T11F30	Domestic Stoves For Solid Fuels		F24	
Mechanical engineering	TF30	Thermal processes and apparatus	T12F30	Refrigeration Systems		F25B	
Mechanical engineering	TF30	Thermal processes and apparatus	T13F30	Ice		F25C	
Mechanical engineering	TF30	Thermal processes and apparatus	T14F30	Furnaces ovens		F27	
Mechanical engineering	TF30	Thermal processes and apparatus	T15F30	Heat Exchange		F28	
Mechanical engineering	TF31	Mechanical elements	T01F31	Fluid Pressure Actuator		F15	
Mechanical engineering	TF31	Mechanical elements	T02F31	Engineering Elts or Units		F16	
Mechanical engineering	TF31	Mechanical elements	T03F31	Storing Distributing Non Solids		F17	
Mechanical engineering	TF31	Mechanical elements	T04F31	Mechanical Control Systems		G05G	
Mechanical engineering	TF32	Transport	T01F32	Vehicles		B60	
Mechanical engineering	TF32	Transport	T02F32	Railways		B61	
Mechanical engineering	TF32	Transport	T03F32	Land Vehicles		B62	
Mechanical engineering	TF32	Transport	T04F32	Ships		B63B	
Mechanical engineering	TF32	Transport	T05F32	Launching of Vessels		B63C	
Mechanical engineering	TF32	Transport	T06F32	Weapons on Vessels		B63G	
Mechanical engineering	TF32	Transport	T07F32	Marine Propulsion		B63H	
Mechanical engineering	TF32	Transport	T08F32	Auxiliaries on Vessel		B63J	

engineering			32				
Mechanical engineering	TF32	Transport	T09F 32	Aircraft	B64		
Other fields	TF33	Furniture, games	T01F 33	Furniture and Domestic Equipment	A47		
Other fields	TF33	Furniture, games	T02F 33	Sports and Games	A63		
Other fields	TF34	Other consumer goods	T01F 34	Tobacco	A24		
Other fields	TF34	Other consumer goods	T02F 34	Shirts	A41B		
Other fields	TF34	Other consumer goods	T03F 34	Corsets	A41C		
Other fields	TF34	Other consumer goods	T04F 34	Outerwear	A41D		
Other fields	TF34	Other consumer goods	T05F 34	Suspenders	A41F		
Other fields	TF34	Other consumer goods	T06F 34	Wigs	A41G		
Other fields	TF34	Other consumer goods	T07F 34	Headwear	A42		
Other fields	TF34	Other consumer goods	T08F 34	Footwear	A43B		
Other fields	TF34	Other consumer goods	T09F 34	Laces	A43C		
Other fields	TF34	Other consumer goods	T10F 34	Fasteners and Braselets	A44		
Other fields	TF34	Other consumer goods	T11F 34	Hand or Travelling Articles	A45		
Other fields	TF34	Other consumer goods	T12F 34	Brushes	A46B		
Other fields	TF34	Other consumer goods	T13F 34	Methods For Life-Saving	A62B		
Other fields	TF34	Other consumer goods	T14F 34	Other Life saving and Amusement	A99Z		
Other fields	TF34	Other consumer goods	T15F 34	Bookbinding	B42		
Other fields	TF34	Other consumer goods	T16F 34	Writing Implements	B43		
Other fields	TF34	Other consumer goods	T17F 34	Decorative Arts	B44		
Other fields	TF34	Other consumer goods	T18F 34	Saddlery and Upholstery	B68		
Other fields	TF34	Other consumer goods	T19F 34	Trimmings	D04D		
Other fields	TF34	Other consumer goods	T20F 34	Laundering	D06F		
Other fields	TF34	Other consumer goods	T21F 34	Wall Floor Covering	D06N		
Other fields	TF34	Other consumer goods	T22F 34	Rope non electric cable	D07		
Other fields	TF34	Other consumer goods	T23F 34	Cooling Or Freezing Apparatus	F25D		
Other fields	TF34	Other consumer goods	T24F 34	Organs Harmoniums	G10B		
Other fields	TF34	Other consumer goods	T25F 34	Pianos	G10C		
Other fields	TF34	Other consumer goods	T26F 34	Musical Instruments	G10D		
Other fields	TF34	Other consumer goods	T27F	Automatic Musical Instruments	G10F		

		goods	34				
Other fields	TF34	Other consumer goods	T28F 34	Aids For Music	G10G		
Other fields	TF34	Other consumer goods	T29F 34	Electroponic Musical Instruments	G10H		
Other fields	TF34	Other consumer goods	T30F 34	Sound-Producing Devices	G10K		
Other fields	TF35	Civil engineering	T01F 35	Permanent ways	E01B		
Other fields	TF35	Civil engineering	T02F 35	Surfaces for roads and sport grounds	E01C		
Other fields	TF35	Civil engineering	T03F 35	Bridges	E01D		
Other fields	TF35	Civil engineering	T04F 35	Platforms or refuge islands	E01F 1		
Other fields	TF35	Civil engineering	T05F 35	Landing for helicopters	E01F 3		
Other fields	TF35	Civil engineering	T06F 35	Draining of roads	E01F 5		
Other fields	TF35	Civil engineering	T07F 35	Protection again snow or sand drifts	E01F 7		
Other fields	TF35	Civil engineering	T08F 35	Road signs or traffic signals	E01F 9		
Other fields	TF35	Civil engineering	T09F 35	Sensitive and restricitng and safety in roads	E01F 1		
Other fields	TF35	Civil engineering	T10F 35	Street Land Cleaning	E01H		
Other fields	TF35	Civil engineering	T11F 35	Hydraulic Eng. Foundations Soil-Shifting	E02		
Other fields	TF35	Civil engineering	T12F 35	Water supply sewerage	E03		
Other fields	TF35	Civil engineering	T13F 35	Building	E04		
Other fields	TF35	Civil engineering	T14F 35	Locking Safing	E05		
Other fields	TF35	Civil engineering	T15F 35	Openings in Building Ladder	E06		
Other fields	TF35	Civil engineering	T16F 35	Earth Or Rock Drilling	E21		
Other fields	TF35	Civil engineering	T17F 35	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>11</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

<sup>11</sup> <https://www.wipo.int/classifications/ipc/en/>

husbandry; trapping; fishing"). The final letter makes up the "subclass" (subclass A01B represents "Soil working in agriculture or forestry; parts, details, or 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://github.com/cortext/PATSTAT/tree/master/nomenclatures/ipc\\_descriptions](https://github.com/cortext/PATSTAT/tree/master/nomenclatures/ipc_descriptions)

## Appendix 5: Data model of PATSTAT2017

The data model shown below is the model of the PATSTAT 2017 database.

