RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

BASIC TUTORIAL

RISIS Patent

Written by P. Laurens (UPEM) 08/2019

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

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

Outline

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

1 Basic Characteristics

The RISIS Patent database derives from the EPO PATSTAT (Version April 2017), a database provided by the European Patent Office and considered as a reference in the field of patent intelligence and statistics. It contains bibliographical patent data from leading industrialised and developing countries. Building on competencies accumulated over the years and mobilizing dedicated resources, IFRIS UPEM, as a research center specialized in quantitative STI, has incorporated in the RISIS Patent database a series of improvements, stemming from outside providers or from its own proprietary developments (filling of missing data, calculations of indicators, geocoding of addresses and their allocation to functional urban areas, classification of technologies, typology of applicants).

The RISIS Patent database is designed for the analysis of technological knowledge creation, using patent as a proxy. It focuses on 'priority patents', the very first patent application for any new invention, that represent the creation of new knowledge, while other non-priority patents that describe either technical ameliorations or market extensions are mobilised as indicators of the importance of the priority patent.

The RISIS Patent database is designed to be a user-friendly relational database using a Mysql querying environment. Including six tables, it provides essential patent information for analysing the geography and the actors of the knowledge production, the content and the value of the knowledge production.

This basic tutorial will provide condensed user-oriented information on the content of the RISIS Patent database for different kinds of research questions. It provides an overview on the main tables and data. For all technical details of the database refer to the RISIS Patent technical documentation (https://rcf.risis2.eu/dataset/4/metadata).

2 Data base content

The RISIS Patent database includes all the priority patents of invention applied by legal organisations from 2000 to 2015, i.e. 13,333,585 patent applications (patents applied in 2015 are not all included).

The data covers all priority patent applications worldwide, i.e. at all regional and national offices in the world.

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[1] and a text concatenating the definitions 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 the content available for semantic analyses.

1 Translation in English for titles and abstracts that remain in their native language will be included in a next release of RISIS Patent.

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

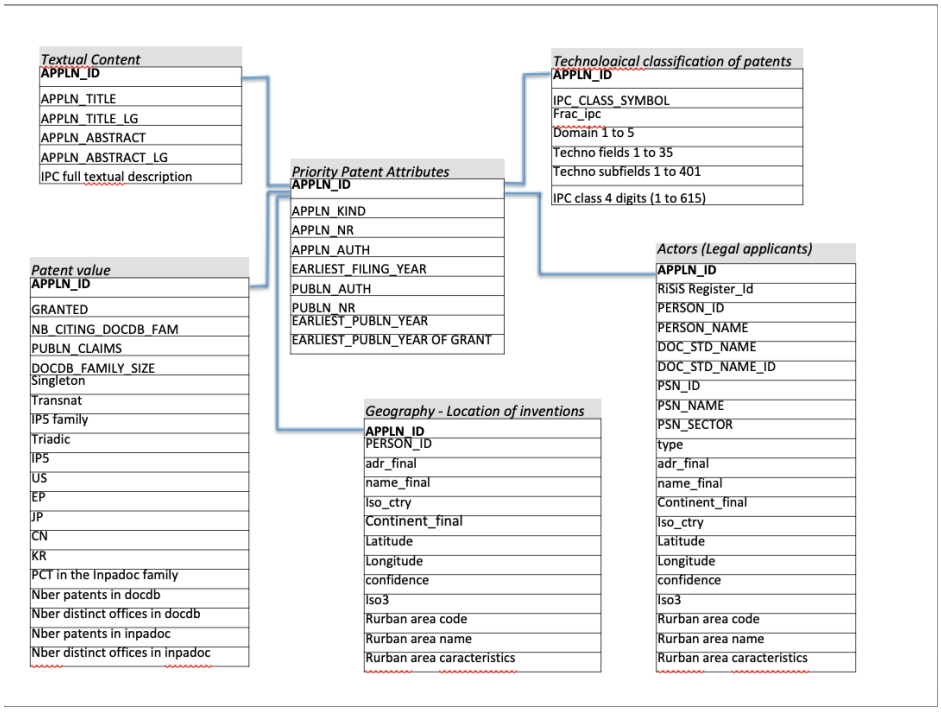
- **Their technological content using the standard technology classification:** (IPC subclasses, aggregation of IPC codes) by technological domains, fields or subfields. Patent allocation in the different classifications is realised on a fractional count basis according to their IPC classes.

- **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).

- **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 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 data model is shown below



For each priority patent application, the RISIS patent database gives:

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

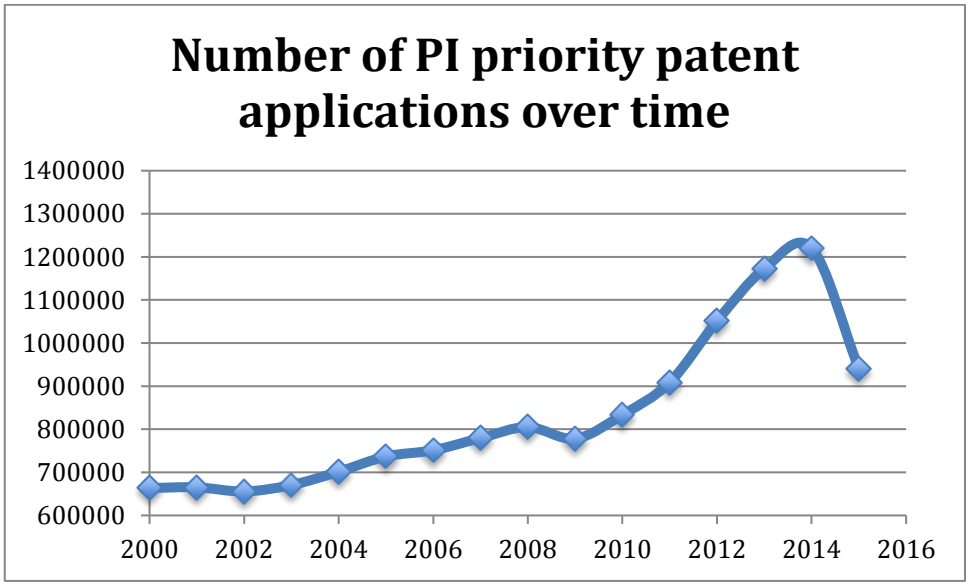
- Patent ID number
- Date of first filing
- Country of first filing
- Date of first publication
- Date of first granting
- Title
- Abstract
- IPC categories mentioned: their number and their language description: as many variables as IPC categories, only one linguistic description
- Whether the patent 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 he/she belongs to
- For each inventor: the inventor address, the inventor geo coordinates, the inventor urban or rural cluster he/she belongs to

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

3 Key figures

Number of patents over time

The RISIS Patent database includes all the priority patents of invention applied for by legal organisations from 2000 to 2015, i.e. 13,333,585 patent applications. The evolution of the number of patent applications over time is shown in figure 1 (patents applied in 2015 are not all included in the database).



Geography of the IP protection

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.
After a first priority patent application for a new invention, the IP protection can be extended in several geographical countries considered as future markets. Most of the patents, for a given invention are applied for 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).

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.

Geography of the inventions

In RISIS Patent, an address is identified for 75% of inventors (to be compared with 10% in the initial raw PATSTAT data) and 67.4% of the addresses are geocoded and associated to 'functional areas' (urban and rural) worldwide mobilizing the RISIS CORTEXT geocoding service. The share of geocoded addresses varies according to the countries. It exceeds 80% in most of the western countries.

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

Typology of 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[2]. 82,6% of the applicants are companies.

4 Querying the database

Currently, MySQL Workbench is used to deliver visual tools for creating, executing, and optimizing SQL queries. For the time being, querying the database, which requires an on-site visit at UPEM in Paris, is carried out with the researchers from LISIS in charge of the database.

5 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 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 University Paris-Est Marne-la-Vallee
- 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)

[2] 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).

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

References

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)

Schoen, A., Laurens, P., Yegros, A., Larèdo, P. (2017) *Evolving technological capabilities of firms; Complexity, divergence, and stagnation*, STI conference 2017, Paris.

Heimeriks, G., Schoen, A., Laurens, P., Yegros A., and Kogler, D. F. (2018). *Knowledge, networks and proximities - An analysis of knowledge dynamics in the Chemical and Pharmaceutical and Biotechnology sectors*, Eu-SPRI conference 2018, Paris.

Laurens, P., Schoen, A., Toma, P., and Daraio C. (2018). *Exploring the Innovative Efficiency of Big Multinational Firms through Conditional Efficiency Analysis,* STI conference 2018, Leiden.

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