Patents filed in India

Nishant Chadha*, Satyaki Chakravarty† and Piyasha Majumdar‡

Objective We describe the construction of a novel data set on patents filed in India between 2000 and 2023 and using the data set, we analyse the spatial pattern of innovation and its evolution over time.

Big picture value of the data set The Indian Patent Office's patent applications, also called Intellectual Property India (IPI), remain under-explored in the innovation literature and the main constraint is its poor coverage in the existing databases (PATSTAT, Derwent Innovation, etc.) along with an inadequate level of detail available per patent. India fairly represents the developing world and is uniquely placed to connect a developed country to the developing through invention linkages. There are a range of questions that could be answered if we had the appropriate data; a few of them being — how are inventions from and to developing countries adopted in the country of origin and elsewhere? Why do some U.S. inventors file for patents in India and others do not? How did the international patent law harmonization exercise – the Trade Related Intellectual Property Rights (TRIPS) agreement influence the proliferation of collaboration among inventors within and across countries, and did it in turn correspond to an increase in follow-on and original inventions? Complete patent data would allow researchers to study such and a range of other questions, and importantly — adequately control for applicants' and patent office's strategies. An example of this is the *request for an examination date* that accounts for the time taken between an application's filing and the actual start of the examination of that patent application, which is governed by an applicant's strategy, either to delay, or expedite the patent application's prosecution. Such an event does not occur at the USPTO but does occur elsewhere. This, if not taken into consideration would introduce bias to the conclusions of any study involving this or related measures.

Until now measures like this were an impediment to studies focusing on patenting strategy, patent examiners' efficacy, and related areas. Fortunately, a recent updation of the IPI website made available a series of new information and this provides us with an opportunity to answer and revisit the aforementioned questions. The only challenge that remains is — the patents are hosted on the IPI website, but are not downloadable in bulk.[1] Thus, in this project we propose construction of a patent data set by scraping the Intellectual Property of India website. We cover new patent applications that are unavailable in popular databases and we provide a lot additional details about patent applications that are already available.

*Indian School of Business, Hyderabad, India; nishant_chadha@isb.edu
†Corresponding Author, Università Cattolica del Sacro Cuore, Milano, Italy; satyaki.chakravarty@unicatt.it
‡India Development Foundation, New Delhi, India; pmajumdar@idfresearch.org
[1]https://ipindiaservices.gov.in/publicsearch

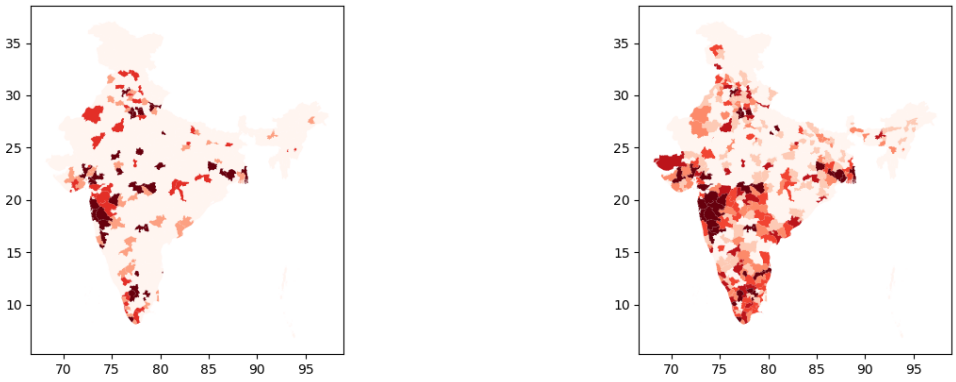
Demonstrations to highlight the new components The most popular worldwide patent database among researchers is the European Patent Office's (EPO) PATSTAT, and we use this as a benchmark for existing databases covering patent applications applied in India. As an example, in Table 1 we provide a few statistical facts about PATSTAT's coverage of patents filed in India and compare them with the information we have collected from IPI's public search website for the patents filed in calendar year 2000 to illustrate the gaps in PATSTAT that could be filled through this project. We can see that firstly, the number of patent applications covered by PATSTAT is lower than IPI — for PATSTAT it is 417 while for IPI it is 829. Secondly, for the year 2000, abstracts and addresses for Indian applicants are unavailable in PATSTAT.

Table 1: PATSTAT vs. Intellectual Property India (only year 2000)

Data	PATSTAT	IPI
Applications	417	829
Abstract	0	84
Applicant with address	0	100
Indian applicants' pin codes	0	88
Inventor with address	0	100
Indian inventors' pin codes	0	90
Attorney/Agent	0	xx

This table compares PATSTAT and Intellectual Property India for the calendar year 2000 (1 Jan 2000 to 1 Jan 2001). *Applications* report the number of applications covered by both sources and the rest of the rows report the **percent of patents** that contains the respective data.

Next, we pinpoint the pin codes (postal codes in India) of inventors and illustrate the emergence of inventors from districts over time in India in Figure 1.



(a) 2005 (b) 2013

Figure 1: District-wise proportion of patents in India

The two Figures show district-wise proportion of inventors for the years 2005 and 2013 from left to right.

The next table, Table 2 provides a snapshot of what is already available in public. PATSTAT's coverage for abstract, assignee, and inventor address is low for the years 2000-2008, and improves thereafter, as given in the Table 2. Its coverage does improve over the years, and especially after 2008, but it could be improved further to include a larger sample of patent applications, where each of them will have detailed information on addresses, legal events, examiner, and attorney/agent names. Also, there are no information on the legal events that happen for each patent, and some examples of those are the request for a patent examination date, the first examination report date, and the description of the report as added by the examiner, patent renewal or abandonment date, etc.

Table 2: PATSTAT coverage

Data	Period	Coverage (%)
Abstract	2000-2008	15
	2009-2011	80
Applicant with country	2000-2011	85
Applicant with address	2000-2008	10
	2009-2011	80
Inventor with address	2000-2008	15
	2009-2011	80
Legal event	-	No coverage

The percentages are obtained from the Autumn 2022 PATSTAT, source: https://forums.epo.org/mapping-data-completeness-of-patstat-global-and-inpadoc-data-tls231-7984

Public dissemination In summary, we plan to scrape the entire IPI website that includes the patents filed at the Indian office and to also scrape the patent E-Register. Our primary plan is to provide a systematic method of scraping the IPI public search to researchers and store a scraped full sample on https://iiindex.org/. But, our larger goal is to set up a system where the data sets could be bulk downloaded in real time by researchers, eliminating the need to replicate downloading and parsing millions of files each time.