



Joining Firm Innovation data with Patent Data, a Practice of Combining NSI Data with Administrative Data

Jia GAO, Fengxin LI, Lei LIU, Peng ZHANG, Feng ZHEN

Paper prepared for the 16th Conference of IAOS
OECD Headquarters, Paris, France, 19-21 September 2018

Session 2.D.3., Day 1, 19/09, 14h30: From R&D statistics to innovation – China

Jia GAO

gaojia@sipo.gov.cn

State Intellectual Property Office of China

Fengxin LI

lifengxin@sipo.gov.cn

State Intellectual Property Office of China

Lei LIU

liulei_4@sipo.gov.cn

State Intellectual Property Office of China

Peng ZHANG

zhangpeng1979@stats.gov.cn

National Bureau of Statistics of China

Feng ZHEN *

zhen@ruc.edu.cn

Renmin University of China

**Joining Firm Innovation data with Patent Data, a Practice of Combining NSI Data with
Administrative Data**

DRAFT VERSION 10/09/2018

PLEASE DO NOT CITE

Prepared for the 16th Conference of the
International Association of Official Statisticians (IAOS)
OECD Headquarters, Paris, France, 19-21 September 2018

Note:

The views expressed in the papers are those of the authors and do not necessarily represent
the views of the affiliated institutions.

ABSTRACT

Joint data from statistical data and administrative data in different resources will improve efficiency of public administration. NBS China has firm level data of R&D input, new product output, and other criteria of production. SIPO China has administrative data of patent application and registration with detailed classification. Combining these two parts of data will improve understanding of systematic innovation system. The Context of “Big Data” in China promotes this data consolidation work across ministries. The process of data merging starts from clearing of patent data collected by registration. Then the firm registered patents transfer to NBS to merge in firm R&D and innovation data linked by firm id. Furthermore, the analysis and empirical results in aggregated sector level support understanding of innovation and relative policy making. The application of the merged data is limited due to micro data protection, but the working mechanism and analysis in sector level are still playing an active role in improving efficiency of official statistics.

Keywords: Innovation; Patent; Administrative Data

1. INTRODUCTION

Combining data from national statistical institutions (NSIs) with administrative data from other official institutions is an effective and realistic way of improving official statistics, especially in the background of "big data". Survey based data collection in NSI is facing more and more difficulties, and sometimes the result may even be different comparing with administrative institutions. The two kinds of institutions, i.e. NSI and administrative institutions, have natural barriers and it seems difficult to ask one side to follow the other side. The "big data" context forced strategy in China gave chances for them to consider about joining data from each other.

Patent is a kind of industrial intellectual property based on application and registration. Individual, firms or public organizations may apply for a patent in local intellectual property office, or for an international patent based on Patent Cooperation Treaty (PCT) mechanism. In China, the administrative office for patent is State Intellectual Property Office (SIPO China)¹. It once reported to the State Council directly like the National Bureau of Statistics of China (NBS China), but now reports to State Administration for Market Regulation after the institutional reform of the State Council in early 2018. Due to the view of enhancing data analysis and data based policy making, SIPO China started to search cooperation with NBS China no later than 2011. The mechanism has changed in these years and the points of analysis have increased since the cooperation with the third party, the academic institutions in 2017.

Innovation data collection in NBS China includes two main parts. The first part is R&D survey including mainly hard data of firms' R&D activity like input, personal, project, fixed capital formation and so on. The second part is innovation survey like Community Innovation Survey (CIS) in Europe with mainly soft survey data about innovation mechanism. On the side of NBS China, besides the final report of the research, joint data of firm innovation with patent details will bring a new understanding of firm innovation mechanism and innovation efficiency. Furthermore, it is also a double check of basic patent numbers from firm survey.

This paper will introduce the process of the cooperation, and some of interesting recent results based on the joint data. The second part will introduce patent data management in the troika system of intellectual property data management. The third part will introduce the process of data combination and its research work between SIPO China and NBS China. The fourth part will give some interesting results based on the 2017 round of joint work, and the final part will be a summary comment.

2. TROIKA MANAGEMENT SYSTEM OF INTELLECTUAL PROPERTY DATA

Patent, trademark and copyright are the three main fields of intellectual property rights. Figure 1 gives the structure of data management before and after the 2018 institutional reform. Before the reform, their registration and administration belonged to three independent institutions. Patent was managed by SIPO China, which is a dependent institution in the State Council. Trademark was managed by Trademark Office, a department in the State Administration for Industry & Commerce. In the reform plan in March 2018, SIPO China was merged into State Administration for Industry & Commerce, and the later institution

¹ The name changed to China National Intellectual Property Office (CNIPO) on 28 August 2018 due to the institutional reform of the State Council started early that year. The changing of administrative mechanism is still on going. We will keep using *SIPO China* in this discussion paper.

changed its name to State Administration for Market Regulation to enhance its regulatory function in the market economy development. The 2018 reform combined the administrative data management of patent and trademark into one ministry. However, it still needs time to facilitate data administration and data mining in one system.

Copyright was managed by General Administration of Press and Publication (GAPP), an independent ministerial administrative office before 2013. It was combined with the State Administration of Radio, Film, and Television to form the State Administration of Press, Publication, Radio, Film and Television. As a ministry of the State Council both before and after the merge, it keeps relatively independent and has relatively close relations to Publicity Department of CCCP.

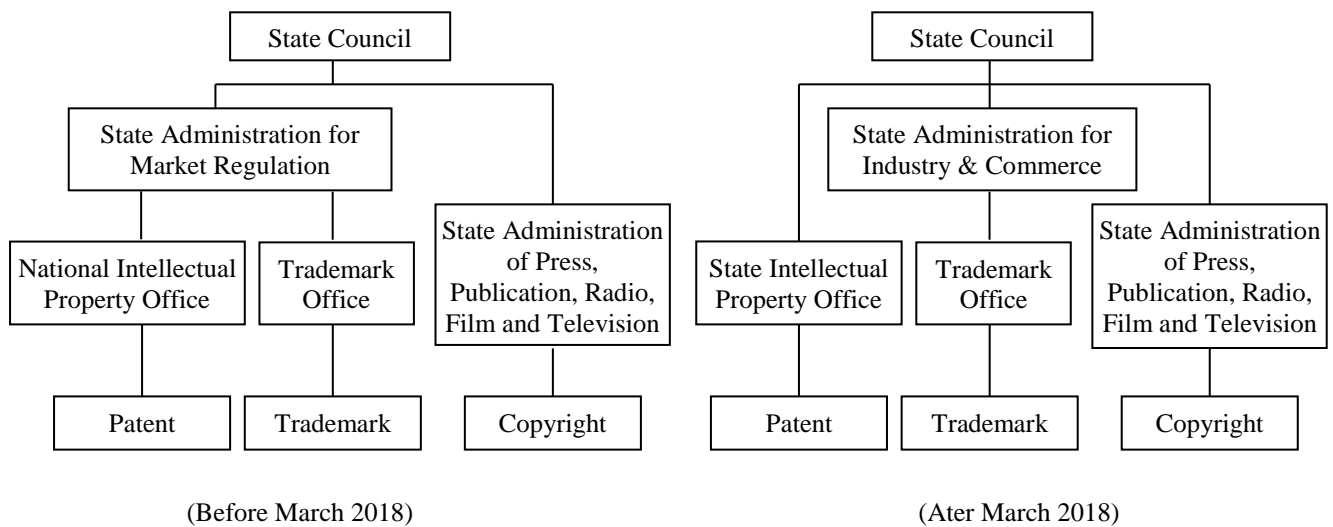


Figure 1. Intellectual Property Data Administration System

3. PATENT AND FIRM R&D DATA CONSOLIDATION

The thinking of joining firms' patent data and R&D data started from the statistics division, i.e. the Planning and Development Division of SIPO China. After the strategy of building a innovative country in 2006, each administrative institution thought about how to practice the strategy in its functional scope. Similar to the Lisbon Agenda in 2000 in European Union, the strategy emphasises R&D input and firm innovation. Based on a good relation and cooperation with NBS China, SIPO China raised the suggestion of joining data of patent from firms with R&D data in NBS, and preparing a report about patent output and firm's innovation efficiency in 2011. Starting from 2017, Renmin University of China (RUC) as an academic institution joined the plan and shared the responsibility of writing the draft report. Figure 2 gives the process of the data consolidation and application in 2017 and 2018 rounds of work.

SIPO China firstly cleans the patent data in each case, specify firm applicant from all the patents. One of the cleaning problem is comparing with legal person name, the applicant may use a different but similar name to apply the patent. Though not too much cased, it brings additional work and should be pay

attention in the application process in the daily work of SIPO China. Then the data lists in the date base by each patent, including patent application, granted, in force and PCT etc., indexed by legal person code. Then the data transfers to NBS China for joining with firm R&D data.

In NBS China, the patent data is joint with data from industrial firms above designed scale², for they are relatively large firms and the data will be collected every year by census in this scale³. However, due to relative strict data protection rules, the joint firm data is kept only in the NBS data system. The feed back data is aggregated firm data in different classification indicators like four digit industrial code, ownership, provincial region, size and some binaries like R&D or not, patent activity or not. Then the aggregated data is sent to research team and back to SIPO China. Besides number of patent in different level, R&D expenditure and personal, sales, sales of new products, export and profit are added in the database to support the multi-level analysis.

The research work includes two parts. The main part is opening exploratory data analysis to find relations among patent, R&D and firm efficiency. The additional part is to design more aggregated forms and open to public. Both of the results will be discussed among the three institutions and SIPO China has the final responsibility for the report and data opening.

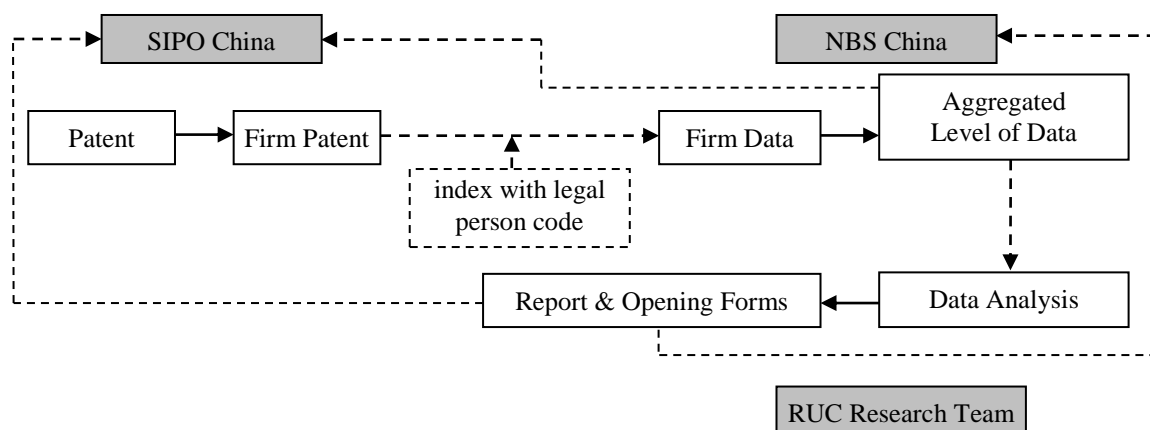


Figure 2. Process of Data Consolidation

4. RESULT

The results of the analysis are plenty. Hereby we would like to share some of the interesting results and the by-product, i.e. double check of the patent data by administration and survey.

² The designed scale is based on revenue from principal operations and partly on number of employees, but different standard in different sectors, and the standard changed upon years. Take industrial sectors and construction sectors as an example, firms above the designed scale means their revenues from principal operations are more than 20 million Renminbi, around 2.52 million Euro in September 2018.

³ "firms" in the contents below refers to industrial firms above designed scale.

4.1 Expected result from the data

As expected, innovation like patent and R&D expenditure has positive relation with firm performance. Take profit as an example, table 1 shows that firms with either R&D expenditure or patent activities has higher average profit. The profit is even higher if firm with both R&D expenditure and patent activities. Firms with PCT have the highest level of average profit, which is around ten times comparing with all firms though only 0.38% of total 378579 firms in 2016 has PCT application.

Table 1. Average Profit of Firms across R&D and Patent in 2016		RMB million	
Industrial firms above designed scale	Without R&D	With R&D	
	Profit	Profit	R&D expenditure
All firms	12.3	41.5	12.6
with patent application	20.1	55.6	17.9
patent granted	21.6	56.0	18.2
patent in force	18.5	48.8	15.6
PCT application	181.3	319.7	110.8

Sources: NBS and SIPO data based calculation. It is the same in the following tables.

4.2 Unexpected results of the report

Innovation like patent and R&D has its mechanism of improving firm performance. However, some of the results are not as expected, or should be pay more attention based on the numbers and facts.

Firstly, the mismatch of R&D and patent is large. For all firms above designed scale in 2016, 23.0% of them had R&D expenditure. However, 33.1% of those firms with R&D did not have any patent activities. On the contrary, 27% of total firms had at least one of the patent activities like application, granted, or patent in force, but 43.2% of those firms did not have any R&D expenditure. It is reasonable for a certain number of firms with either R&D or patent may not have the other activity. However the scale of the mismatch is relatively too large in China in 2016.

Table2. Mismatch of R&D and Patent in 2016	%	
Patent activity	With R&D but without patent activity	With patent activity but without R&D
Any patent activities	33.19	43.20
Application (domestic)	54.54	36.55
PCT application	98.58	15.16
Granted	56.25	34.87
In force	36.01	41.19

Secondly, the increase rate of R&D expenditure is much lower than the rate of patents, which will bring pressure for patent increase in the future. From 2012 to 2016, patent in force kept very high growth rate, and growth rate of invention patent⁴ in force kept even higher. (Table 3) However, the growth of R&D expenditure was lower in these years, especially in the recent two year. The background is the economy growth of China slowed down in recent years due to the pressure in both domestic reform and world political and economic environment. It is reasonable to see the reduce of R&D expenditure growth in a not so optimistic firm financial situation, but the patent strategy must adjust to fit for the new situation.

Table 3. Increase of Patent in Force and Invention Patent in Force from 2012 to 2016 %

Year	Patent in force	Invention patent in force	R&D expenditure
2016	13.01	26.31	8.52
2015	17.54	31.51	5.91
2014	14.89	21.52	13.17
2013	28.68	30.05	17.89
2012	49.42	51.96	-

Thirdly, patent activity is highly concentrated in three provinces, i.e. Guangdong, Zhejiang and Jiangsu. These three provinces are the most developed area in China. Nearly 35% of total firms are located in these area. However, for the patent activities, nearly half happened in these provinces. For PCT application, more than 75% of total 15452 pieces was from Guangdong in 2016. Moreover, more than half of the PCT cases was from Huawei and Zhongxing, the two largest manufacturer of telecommunication equipment.

Table 4. Patent Concentration in Selected Provinces in 2016

Region	% of firms	% of patents			
		Application	PCT application	Granted	In force
Jiangsu	12.65	17.93	6.81	15.92	14.95
Zhejiang	10.60	14.80	3.48	16.75	15.58
Guangdong	11.27	16.08	75.10	14.97	17.75
Sub-total	34.53	48.81	85.39	47.64	48.28

Fourthly, domestic funded firms has lower ratio of firms with patent activities comparing with those of foreign funded firms and Hong Kong, Macao, and Taiwan funded firms. Further more, private firms has even lower ratio. (Table 5) Considering about more than half of total firms are private firms, it is not

⁴ There are three kinds of patents in China, i.e. invention patent, patent for utility model, and patent for design.

optimistic for the industrial sectors to support innovation based sustainable fast development. The growth rate of economy may continue to slow down in the near future.

Table 5. Patent Activities by Firm's Ownership in 2016

Ownership	% of total firms	% of firms with patent activity			
		Application	PCT application	Granted	In force
Domestic funded	86.91	16.0	0.3	14.9	24.1
of which: Private	56.61	14.4	0.2	13.4	22.0
Funds from Hong Kong, Macau and Taiwan	6.19	20.3	0.8	19.6	31.4
Foreign funded	6.90	19.0	0.8	18.4	30.1

4.3 Double check of patent data

Survey based patent data is exist in NBS data system. The joint work gives chances to double check patent data between the administrative system and survey system. Take patent application as an example, the mismatch ratio is around 5%. (Table 6) On the one hand, frankly speaking, administrative data is more accurate than survey for the firms may not so clear about each patent situation when filling in the statistical forms in a short time. On the other hand, the difference can be partly explained by reasonable reasons that mismatch is always exist because firms may report non-PCT application aboard.

Table 6. Mismatch Ratio of Patent Application from 2011 to 2016 %

2016	2015	2014	2013	2012	2011
3.49	6.28	-1.99	10.29	8.03	3.94

* mismatch ratio = (SIPO number - NBS number)/SIPO number*100

5. SUMMARY

Though not complicated, this case of data consolidation between administration institution and NSI is a rare successful case in practice in China. It drives us the following thinking and suggestions of improving official statistics in China in the future.

Firstly, combining NSIs data with administrative data is a realistic application of "big data". Big data in official statistics may gives two main areas of application. The first one is combining data from different resources, especially from different official resources like administrative data. It has little difficulty in technique but it has buriers in practice. For technique, administrative data generally has good basis and it can cover population or sample well in a good format in the certain area. However, the non-technique reasons like the relations between the institutions or persons may deeply influence the practice. The second

is introducing non-traditional data like web data to official statistics. The problem is it may challenge the basis of sampling theory and central limit theorem based modern statistics.

Secondly, how to design a mechanism of mining administrative data is a request. The joint work should be based on not cooperation relations but mechanism of governance. It may start from institutions with good foundation of data collection system and the following areas may be considered in the first role. The first level is area with clear data standard, like patent and other intellectual property, financial and banking, customs. The second level is data that can be used as sampling frame of NSIs, like registration information of people and firms.

Thirdly, metadata is important for not only NSIs across regions and countries, but also meaningful for administrative data. Administrative data collection may be designed separately by different institutions and the need of each single institution will be considered with priority. If go further, all the administrative data in the government should be considered about consolidation and metadata should be emphasized to support the potential joint work.

In general, the simple case is just a start. It provides a more realistic direction for official statistics. Though not easy in practice, solving the problems will not only improve official statistics itself, but also give a better understanding of economy and society, and improve welfare of human life.

REFERENCE

[1] World Trade Organization (WTO), https://www.wto.org/english/tratop_e/trips_e/intel1_e.htm.