

Title	エビデンスベースの科学技術政策の実現に向けたイノベーションデータベースの構築
Author(s)	元橋, 一之; 鈴木, 潤
Citation	年次学術大会講演要旨集, 25: 225-228
Issue Date	2010-10-09
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9283
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

エビデンスベースの科学技術政策の実現に向けた イノベーションデータベースの構築

○元橋一之（東大），鈴木 潤（政策研究大学院大）

1. はじめに

少子高齢化時代において一定の経済成長を実現するためには生産性の上昇が重要であり、そのためのイノベーションの促進は政策面での最優先項目となっている。イノベーションを加速するための政策は、民間企業の研究開発の促進の他、大学や公的研究機関における研究成果をベースとした産学連携によるイノベーション、ハイテクベンチャーの育成など多岐にわたる。これらの政策を企画・立案していくためには客観的なデータに基づく分析を行い、公的な財政状況が厳しくなっている中で効果的な政策資源の配分を行っていくことが必要である。

幸いにして、日本は科学技術統計、知的財産活動統計、企業の経済活動に関する統計など、イノベーションの政策分析を行っていく上で必要になる政府統計は、欧米諸国と比較しても豊富に存在する。しかし、それらのデータを用いた政策分析については遅れているといわざるを得ない。その背景として最も大きな要因の1つは、統計間の相互接続が困難であることである。具体的には、産業分類の不一致や技術分類と産業分類のコンコダンス表が整備されていないこと、企業レベルの分析を行うために統計調査ごとで管理されている企業コードが異なり、相互接続ができないことなどの問題がある。

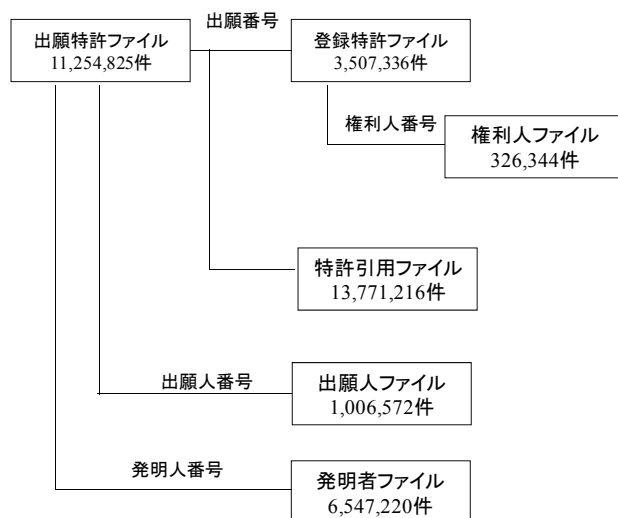
このようなイノベーションに関する統計間の相互接続を進め、より効果的な政策分析を進めるために東京大学においては「イノベーション・データベース・プラットフォーム」の構築と研究者に対する提供を行っている。その中核をなすのが、特許庁の整理標準化データを加工・修正して研究者に使い勝手がいい形式で公開している IIP パテントデータベースである。特許に関するデータは特許公報等ですでに公開されているものであるが、数百ギガバイトにおよぶ膨大なテキストデータであり、イノベーションに関する分析に用いるのは困難である。従って、我々は特許庁の了解も得て、研究用途に限って無償で利用可能なデータベースの開発・公開を行うこととした。これにイノベーションプロセスの上流に位置する研究開発費に関する「科学技術研究調査」（総務省）と下流の企業活動に関する「企業活動基本調査」（経済産業省）を加えたものをイノベーションデータベース基盤の中核データと位置付けた。ここでは、これらの中核データを中心としたイノベーションデータベース基盤の整備状況を紹介するとともに分析事例を交えたデータの活用可能性について述べる。

2. IIP データベース

IIP パテントデータベースは、概ね月 2 回のペース公表されている「整理標準化データ」をベースに作成されている。「整理標準化データ」は SMGL や XML などのタグ付きテキストファイルとして特許情報が収録されたものである。ここでは、これらのテキストファイルをデータの統計的処理を容易にするために SQL データベースに変換し、更に研究者においてもっともニーズの高いと思われるものを CSV 形式のテキストファイルとして公開している。現時点では、1964 年 1 月以降の出願から 2009 年 10 月時点で公開されたもの（整理標準化データの 2009 年度第 15 回公表分）までを取り込んだものとなっている。

IIP パテントファイルとして CSV 形式で公開しているデータには、特許出願データ（出願番号、出願日、審査請求日、技術分野、請求項数等）、特許登録データ（登録番号、権利消滅日等）、出願人データ（出願人名、個法官コード、国・県コード等）、権利者データ（権利者名等）、引用情報データ（引用・被引用特許番号等）、発明者データ（発明者名称、住所）が含まれている。データベースの構成とテーブル毎のデータ数については、図 1 のとおりである。例えば出願特許数でいうと 11,254,825 件の特許データが収録されており、そのうち 3,507,336 件の特許が登録されている。それぞれに出願人、権利人に関するテーブルが接続しており、また引用データは審査官引用（審査請求があった特許に対して、審査官が拒絶理由を付す際に引用された過去文献）に関するデータが収録されている（Goto and Motohashi, 2007）。

図 1：IIP パテントデータベースの構成



IIP パテントデータベースは「整理標準化データベース」における情報を忠実に取り出して、データベース化したものであるが、このデータを用いて分析を行う上ためには、オリジナルのデータにおいていくつかの問題がある。そのうち最も重要なのが、出願人、権利人、発明者などの情報の表記の揺れの問題である。例えば 1960 年代などの古い時代のデータはこれらの名称がカタカナ表記されているのに対して、最近では漢字表記になっているのでオリジナルのテキスト情報のみからは名寄せはできない。また、企業の名称変更や表記方法の変更によっても、本来であれば同じ企業であってもデータベース上では違うものとして認識されてしまう。そこで我々は主に出願人情報について名寄せ作業に取り組んでいる。

ここでの作業は、まず特許庁の出願人コードを活用することから始める。ただし、この出願人コードは現在の 9 ケタコードに至るまで、コードの変換が何回も行われていることから、これをまず補正する。なお、特許庁の出願人コードは、False Negative（本来同一の出願人に対して違うコードが振られる）という問題があるが、False Positive（違い出願人に対して同じコードが振られる）という問題はない。

この状態から出願人の名称情報を用いて出願人のタイプとして、(1)個人、(2)企業、(3)非営利機関（官庁、公的研究機関など）(4)大学に分類を行う。このうち、(3)と(4)については件数が少ないためマニュアルで名寄せ作業を行っている。

更に企業出願人を取り出して企業名称の標準化を行い、そのうえで住所情報から同一市町村の存在に存在する同一名称の企業を同一企業として新たな ID 番号を付与している。なお、この方法によると企業名称の標準化が不十分な場合や名称変更が行われた場合などにおいて False Negative の可能性がある。また、同一名称で違う企業が同一所在地に存在する場合は False Positive の可能性もある。これらの問題を解決するためには、所在地情報を含んだ正確な企業名称に関する情報が必要である。イノベーションデータベース整備にあたって企業活動基本調査との接続を行っているが、現時点ではこの情報を特許データの企業名寄せには用いていない。また、日本におけるすべての事業所・企業をカバーする事業所企業統計の名簿情報を用いればより広範囲の出願人名寄せに関する False Positive の問題を解決することができる。更に、ここでの作業は主に日本に所在する出願人に対して行われたものであることに留意することが必要である。欧米の企業などの外国における出願人についても今後の作業として残っているところである。なお、欧米の企業の名寄せについては OECD や NBER グループなどによって作業が進んでいるところで、これらのグループとの連携することによってデータの改善を行っている。

3. イノベーションデータベース基盤の現状

東京大学においては、上記の IIP パテントデータベースの更新・拡充とともに他の統計調査との接続などのデータベース基盤整備事業に取り組んでいる。前述したようにこのデータベース基盤の中核的なデータとして、IIP パテントデータベースをはじめとして、科学技術研究調査（総務省）と企業活動基本調査（経済産業省）の企業レベル個票データの接続を行っている。

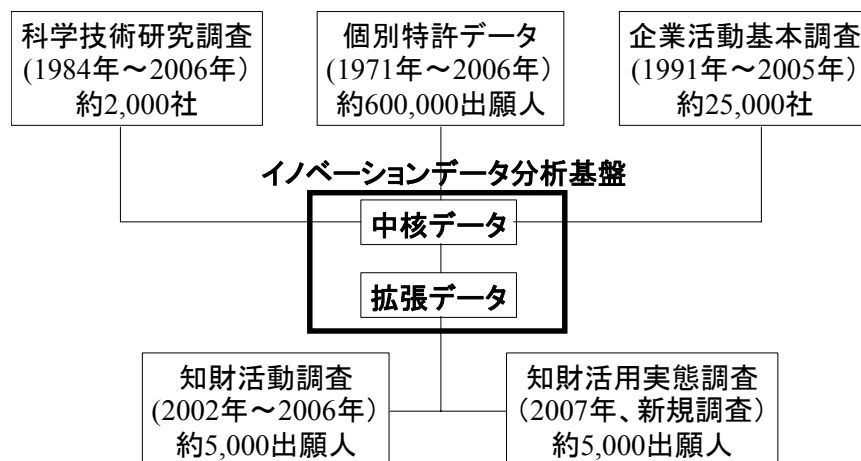
まず、科学技術研究調査の企業その他、大学や公的研究機関における研究開発活動を総合的に調査しているものであるが、ここでは資本金 1 億円以上の企業に対して行われている企業等 A の調査項目について、1984 年からのパネルデータを作成している。科学技術研究調査においては、科学コードという番号で企業データの整理が行われているが、コードの付け替えが行われている年があり、パネルデータの作成にあたっては新旧コードの対応関係を把握することが適当である。この作業によって、5414 企業のアンバランスパネルデータ（最多年の 2002 年の企業数が 3650）を作成した。

一方、企業活動基本調査は 1991 年に開始された比較的新しい統計調査であるが、こちらは永久企業番号という期間を通じて統一的なコード体系が整備されており、パネルデータの作成は容易である。資

本金 3000 万円以上でかつ従業員数 50 人以上の製造業または卸小売業（2001 年から一部のサービス産業に対して業種が拡大）に属するすべての企業に対する調査であり、毎年約 2.5 万社のサンプル数となっている。

これらの統計調査の企業パネルデータに IIP パテントデータの出願人（日本に所在する企業のみ）約 60 万社を接続させたものがイノベーションデータベース基盤の中核的な構成要素となっている。

図 2：イノベーションデータ分析基盤のフレームワーク



また、これらまでのデータ接続作業としては、これらの中核データの整備に加えて、「知的財産活動調査」（特許庁）や新規に行ったライセンスに関するアンケート調査（知財活用実態調査）を接続し、データ拡充を行っている。知財活動調査は特許庁において行われている承認統計であり、毎年一定数の特許、実用新案、商標及び意匠の出願を行っている出願人（企業の他、大学や個人発明家含む）に対して毎年行われている。内容は、各種知財の出願状況やライセンシングなど知財利用状況、知的財産活動費や知財侵害の実態に関するものである。このデータは特許庁における出願人番号をベースとして管理されているため、中核データの個別特許データ（同じ出願人番号の情報が存在する）との接続が可能である。

4. イノベーションデータベースの活用事例

（1）技術・産業コンコordanステープル

それではこのようなデータを活用することによって、どのような分析が可能になるのでしょうか？まず、イノベーション分析に関する基礎的な集計データとして、産業・技術コンコordanステープルをあげることができる。特許データは IPC 分類などの技術分類に従って集計することは可能であるが、それがその発明がどの産業によって行われ、どの産業によって使われているのかという点は不明である。この技術分類と産業分類の対応関係は、イノベーションの分析を行っていく上で重要な情報となる。

ここでは、企業レベルの特許保有と産業別売上高のデータを用いて IPC コードと産業分類のコンコordanステープルの作成を試みた。本来は、カナダ統計局のデータのように特許毎の IOM や SOU が分かれば理想的であるが、日本にはそのようなデータが存在しない。従って、企業レベルの集計データを用いて IPC と産業分類の関係について明らかにしたものである。

具体的には IPC 分類のサブクラスレベルと企業活動基本調査の 3 桁産業分類のテーブルを 1995 年と 2005 年のデータで作成した結果、おもに電気機械工業と化学工業の分野でイノベーションが盛んにおこなわれるようになってきていることが分かった。また、コンコordanステープルの作成にあたっては、技術分野に対応する産業が当該技術のサプライヤー（IOM：Industry of Manufacturer）なのか、技術のユーザー（SOU：Sector of User）なのかといった情報が重要である。これは図 2 のデータベース構造の中で知財活動調査が加わったことによって、企業毎に所有特許の活用状況（自社実施、インライセンスやアウトライセンスの状況）が分かるようになっており、この情報を活用することによって、より詳細なテーブルを作成することができる。日本企業の特許利用率（保有特許に占める自社実施かライセンスによって利用している特許の割合）は平均 6 割といわれているので、特に SOU の分析については当該デー

タを用いて補正をおこなうことが重要である（東京大学、2008）。

（2）技術政策（R&D 補助金や税制）に関する評価分析

図2の研究開発費と企業のパフォーマンスに関するデータを用いることによって、研究開発税制や補助金の効果分析といった政策分析を行うことも可能である。例えば、研究開発税制が企業の研究開発促進に影響を及ぼしているか、及ぼしているとしたらどの程度の大きさになるのかについては分析を行った。東京大学のプロジェクトにおいては、科学技術研究調査データと公開財務データ（財務諸表）を用いて、これまで行われてきた研究開発投資の決定要因に関するモデル推計を行い、研究開発税制の変化によって研究開発投資に関するコストの変化を推計し、どの程度の投資誘発効果をもたらしたかについての分析を行った（元橋・袁、2008）。日本の研究開発税制については、2003 年度にこれまでの増加試験研究費方式から研究開発費総額に対して税額控除が行われる方式に大きな変更が行われた。この改正によって、(1)実効控除税率は2002 年度には0.1%にまで下がったが、2003 年度には一気に10.6%に上昇したこと、(2)研究開発費の税引き後価格を10%程度下げる効果を持つこと、(3)その結果、研究開発投資を8%程度押し上げる効果を持つことなどが分かった。

さらに、企業に対する研究開発補助金の効果については、企業における公的資金受け入れはすべて基礎研究に用いられると仮定し、基礎研究の公的受け入れ分と、自己負担分の関係について分析を行っている（鈴木、2008）。その結果、公的資金は基礎研究費シェアに対してはプラス、自己負担基礎研究費に対してはマイナスの効果を持つことが分かった。この結果からは、公的資金が自己負担の基礎研究費に対して代替的な効果が存在すると示唆される。

この他、研究開発費に関する外部支出（大学や公的研究機関に対する支出）や特許の共同出願、共同発明の状況を用いた産学連携とイノベーションの研究、特許発明者や企業の住所の情報を用いた地域的な技術スピルオーバーの分析（クラスター政策に関する分析）、研究開発費や特許にみる技術分野の多角化活動（あるいは選択と集中）と企業パフォーマンスの関係など、さまざまな分析を行うことが可能である。

5. まとめ

東京大学を中心とするグループにおいて、特許庁における整理標準化データや全文特許テキストデータを用いた特許データベースを構築するとともに、研究開発に関する科学技術研究調査（特許庁）や企業パフォーマンスに関する企業活動基本調査（経済産業省）の企業データを接続した中核的データを中心としたデータベースの構築を行ってきた。また、このデータベース基盤を活用することによって、科学技術政策に関する定量的な分析事例を紹介した。

今後の課題としては、特許データベースについては、アップデートと名寄せ作業によるデータ精度の向上を国際的な研究者の協力の下行うことがあげられる。また、発明者情報と大学などの研究者リストを接続して、アカデミックパテンティングの研究を行うことで産学連携政策のあり方に対する貴重な示唆が得られると考えている。更に、企業統計との接続については日本におけるすべての企業をカバーする事業所・企業統計との接続を行っているところである。この作業によって、特許情報における出願人名寄せの精度向上につながるとともに、日本におけるイノベーション活動の全体像を押えることができる。また、企業の開廃業と特許活動の関係についての分析を行うことで日本のベンチャー政策に対する提言を行う予定である。

このように今後ともイノベーションデータに関する専門家やデータを保有する政府関係部局との協力関係を構築しながら、イノベーションデータベースの拡充を行い、さらにこれらのデータベースを公開することによって、エビデンスベースの科学技術政策の立案に今後とも貢献していきたいと考えている。

【参考文献】

- 鈴木潤（2008）、「企業が受け入れた公的資金と基礎研究の関係」、東京大学（2008）4－2章
東京大学（2008）、『平成19年度産業技術調査事業：イノベーションデータ分析基盤に関する調査事業報告書』、2008年3月
元橋一之・袁媛（2008）、「研究開発税制に関する定量的分析」、東京大学（2008）4－1章
Goto and Motohashi (2007), Construction of a Japanese Patent Database and a first look at Japanese patenting activities, Research Policy, 36(9), 1431-1442