

Title	イノベーションデータベースの開発と研究者に対する提供
Author(s)	元橋, 一之; 鈴木, 潤
Citation	年次学術大会講演要旨集, 23: 522-525
Issue Date	2008-10-12
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/7616
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

2 A 2 2

イノベーションデータベースの開発と研究者に対する提供

○元橋一之（東京大学），鈴木潤（政策研究大学院大学）

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ードנס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。

イノベーションに関する分析を進めるためには、研究開発活動によって、それが新たな商品・サービスとなって市場に供給されるまで、あるいは新たなプロセスイノベーションが生まれて生産効率の向上が実現するすべてのプロセスについて理解を深めることが重要である。特許は、このプロセスの中間的なアウトプットと考えることができるが、その上流にあたる研究開発活動の状況やあるいは下流にあたる企業のパフォーマンスについて一体的に分析を行えるようなデータベース基盤の整備が重要である。従って、ここでは研究開発活動に関する「科学技術研究調査」（総務省）と企業活動に関する「企業活動基本調査」（経済産業省）も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ている。ただし、これらのデータは政府統計であるため統計法によって個票の利用は「目的外利用」制度によって個別に承認をとることが必要である。大きな問題としては、データそのものをデータベースとして管理し、公開することはできない。しかし、これらの個票データが相互に接続できるような企業コードのコンバータと整備・公表することはこれらのデータを相互接続した研究を進めるうえで意義が大きい。

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 パテントデータベース

日本を始めとする多くの国において、出願特許の情報は 18 ヶ月以内に公開される。この公開特許情報は、新たに特許出願を行う際の先行技術調査に関する情報として有用だが、技術分野別のパテントマップの作成や研究開発戦略の策定への応用など、より高度な使い方も存在する。その中でイノベーション活動を測定するための統計情報としての活用方法が脚光を浴びている。

公開特許データを用いたデータベースとしては、米国の研究者が中心になって作成した NBER パテントデータベースが有名であるが、最近では EPO（ヨーロッパ特許庁）が PATSTAT という世界の出願特許に関するデータベースを公開し、研究者による利用が活発化している。また、OECD においては、特許情報の統計利用を促進するために日米欧の 3 極特許庁からなる特許統計に関するタスクフォースを結成し、毎年会議を行ってきている。

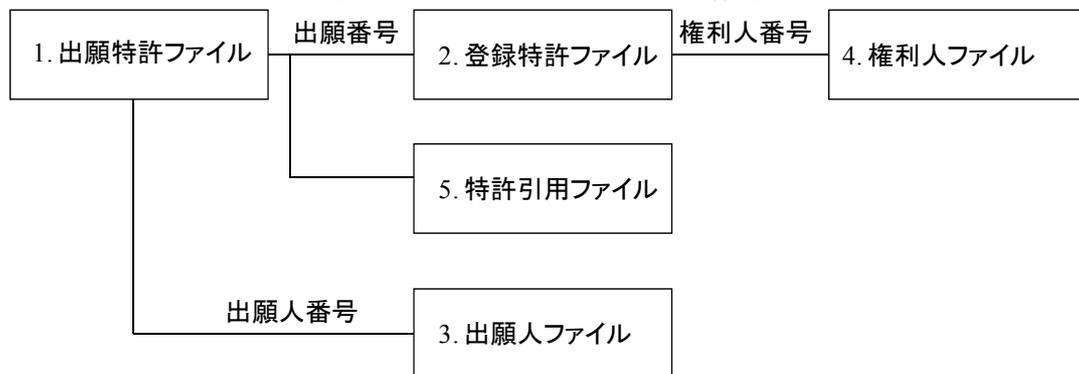
このような世界的な動きに対応して、日本においては「IIP 特許データベース」が開発された。このデータベースは、(財)知的財産研究所の HP 上 (<http://www.iip.or.jp/>) で公開されており、研究者向けに無料で提供されている。

IIP 特許データベースは、特許庁の整理標準化データから必要な項目を抽出し、実証分析用データベースとしてまとめたものである。このデータベースには、1964 年 1 月以降の出願データが含まれており、現時点では、2004 年 1 月までに公開された約 902 万件のデータが蓄積されている。さらに、各特許出願に対応した、審査・登録・引用データや出願人・権利人のデータなども含まれている。

データベース構築に当たって、延べ約 5,600 万件の整理標準化データから、イノベーション研究に有用な項目のみが選択・抽出されている。その上、更新時点ごとの出願人・権利人コードの違いが統一され、特許庁内部で使用される中間コードも変数としてまとめられている。さらに、データ形式は統計処理が行いやすい形式に変換されている。

こうした一連の作業により、実証分析に有益なデータセットが、(1) 特許出願ファイル、(2) 特許登録ファイル、(3) 出願人ファイル、(4) 権利者ファイル、(5) 引用情報ファイルという 5 つのカテゴリの下、コンパクトにまとめられ、容量も一般ユーザーにとって扱いやすい大きさに抑えられている。これらのデータの構成については、図 1 のとおりである。なお、データベースの作成方法や内容の詳細については、Goto and Motohashi (2007) を参照されたい。

図 1 : IIP 特許データの構成図



また、当データベースは、2008 年 5 月にアップデートが行われている。これにより、収録期間が 2007 年 5 月公開のもの（整理標準化データ平成 19 年度第 5 回提供分）へと延長されるとともに、発明者データが追加された。ただし、このアップデート・ファイルは、現時点では完全版ではなく、(株)人工生命研究所の HP 上 (<http://www.alife-lab.jp/>) で β 版が公開され、そこに設置された掲示板にてユーザーからのバグ出しや問題提起等の議論を重ねることで、質の改善が図られている。このようにオープンイノベーション形式によってデータの問題点を洗い出し、改善されたものを定期的に (財) 知的財産研究所の HP に公開することとしている。

3. イノベーションデータベース基盤の現状

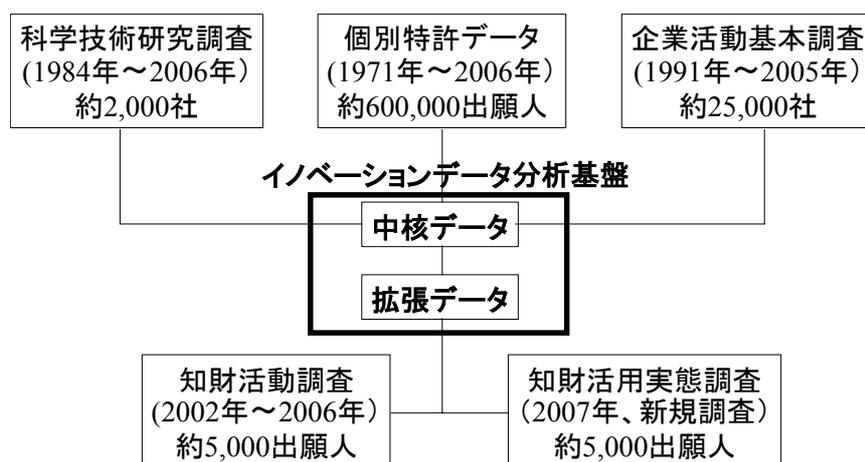
東京大学においては、経済産業省の委託研究（産業技術調査事業）などを受けながら、上記の IIP 特許データベースの更新・拡充とともに他の統計調査との接続などのデータベース基盤整備事業に取り組んでいる。前述したようにこのデータベース基盤の中核的なデータとして、IIP 特許データベースをはじめとして、科学技術研究調査（総務省）と企業活動基本調査（経済産業省）の企業レベル個票データの接続を行っている。

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スなど知財利用状況、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。

また、知的財産活動調査は出願数やライセンスに用いられる特許数など定量的な情報として有益なものが存在するが、特許データをイノベーション分析に用いるためには、特許性向（すべての発明が特許化されるわけではないので、発明案件のうちどの程度の割合のものを特許出願するか）や特許利用の実態に関する詳細な情報（例えば特許保有の目的として防御的な目的なのかライセンスアウトを目的としているのか）が必要となる。これらの情報は知的財産活動調査から得ることができないので、新規にアンケート調査（ライセンス活動実態調査）を行いデータベースの拡充を行った。

4. イノベーションデータ基盤を用いた分析可能性

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あろうか？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・技術コンコルダンステーブ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ので、これまでカナダ特許庁や OECD によってコンコルダンステーブルの作成が行われてきた。

このテーブルは上記のように特許データと企業データが接続されることによって作成することが可能となり、いくつかのコンコルダンステーブルを作成して、時系列的な変化や手法による結果の違いに関する検証作業が行われている。

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、コンコルダンス表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）。

また、図2のデータベース基盤を確立することによって、イノベーションに関する様々な実証研究を行うことが可能である。たとえば、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。具体的に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。同データベース基盤の利用が進みことによって、日本におけるイノベーション政策やイノベーションプロセスに関する実証研究が大いに進むことが期待される。

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