

Title	データ・情報基盤整備に関する課題
Author(s)	富澤, 宏之; 岸本, 晃彦
Citation	年次学術大会講演要旨集, 27: 102-105
Issue Date	2012-10-27
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/10984
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

データ・情報基盤整備に関する課題

富澤宏之、○岸本晃彦（文科省・科学技術政策研）

1. 背景・目的

近年、科学技術の高度化・複雑化や科学技術と社会の関係深化を背景として、これまで以上に客観的根拠（エビデンス）と合理的な政策形成プロセスに基づいた、科学技術イノベーション政策の推進が求められている。こうした背景の下、文部科学省では、「エビデンスに基づく政策形成」の実現を目指した「科学技術イノベーション政策のための科学」（以下、「政策のための科学」）事業を2011年度から開始した。その一環として、科学技術政策研究所では、「政策のための科学」推進の基盤となるデータ・情報整備事業を開始した。本事業の進め方に関する2011年度の検討結果は、『科学技術イノベーション政策のための科学』におけるデータ・情報基盤構築の推進に関する検討（仮）と題して科学技術政策研究所から公開する予定である。本発表では、本事業において得られた結果に分析を加えて報告する。

2. 「政策のための科学」に資するデータ・情報基盤のあり方の検討

データ・情報基盤全般や関連事業の現状・課題、及び今後の進め方・あり方について幅広く議論を行うために、関係する有識者による委員会（委員長：元橋一之東京大学教授）を開催した。また、委員会での議論を深めるために、アンケート、及びインタビューにより多くの方々の意見を収集した。

(1) アンケート調査

電子メールによるアンケート調査を実施した。アンケート対象者は、主に科学技術総合リンクセンター（J-GLOBAL）から抽出した。まず、①「科学技術政策」、「政策科学」をキーワードとしてそれぞれ200人、50人を抽出した。次に、②「政策のための科学」との関連が想定される「研究・技術計画学会」（50人）、「科学基礎論学会」（20人）、「科学技術社会論学会」（20人）、「統計科学研究会」（20人）、「日本知財学会」（16人）から抽出した。さらに、③「政策のための科学」に関連するプロジェクトの代表研究者を加えた。これらの中から公開情報よりメールアドレスが取得できた120人を調査対象とした。

設問は、(A)既存データについて利用したいデータの内容・問題点、(B)新規（既存データには存在しない）データについての内容・問題点、(C)相互に接続・結合したいデータ、(D)事業推進に当たっての要望・意見、について回答を依頼し、17人から回答が得られた。

まず、アンケートから得られた、(A)既存統計・データの利用ニーズについて図1に示す。科学技術に関連する統計・データとして学校基本調査（文部科学省）、科学技術研究調査（総務省）が、イノベーションに関連する統計・データとして企業活動基本調査（経済産業省）、知的財産活動調査（特許庁）、海外事業活動基本調査（経済産業省）が複数の回答者から挙げられている。ただし、ニーズは必ずしもこれらの統計・データに集中してはおらず、多岐にわたっている。なお、「その他」として挙げられているものとしては、国立研究機関・独立行政法人・民間調査機関による各種アンケート結果や上場企業の財務データ、企業グループのリストなどであった。

次にこれら既存統計・データを利用するに当たって、現状の課題44件が挙げられ、これを図2に示す。利用上の課題としては、①マイクロデータ（機関・企業単位データ、個人単位データ）の利用ニーズが大きく、問題意識も高い。個票データ（マイクロデータ）利用の次に問題とされているのが、②時系列データの不整備や③国際比較の困難に関する課題である。調査対象の抽出方法の変更や、調査対象（主に企業）の名称変更・改組改編などにより、統計・データを時系列に接続することは技術的に難しい。また、統計調査の回答率や回答者の解釈などの問題から、④統計・データの精度・信頼性に関する疑問も挙げられている。

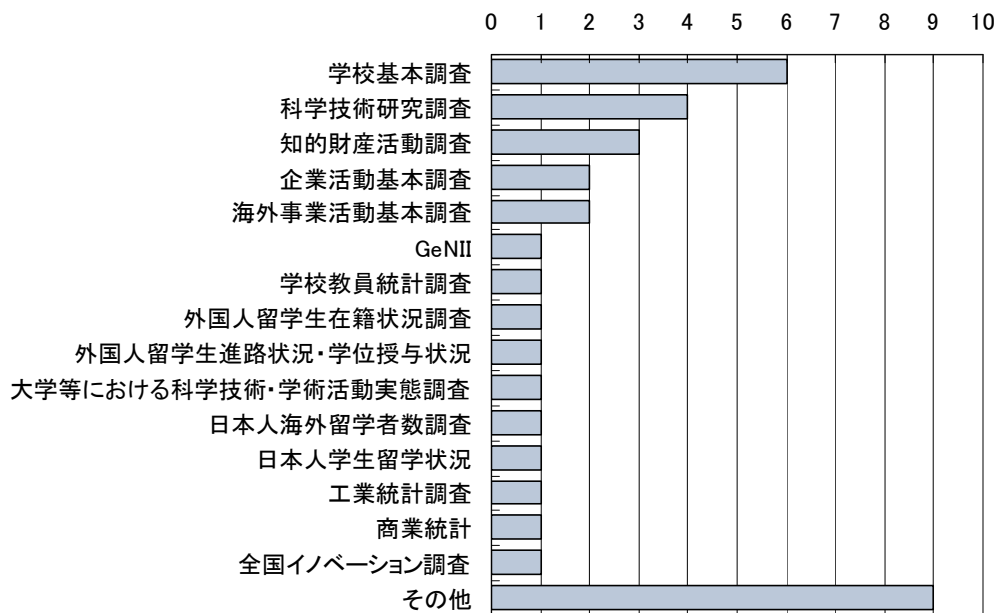


図1 利用したい既存統計・データ

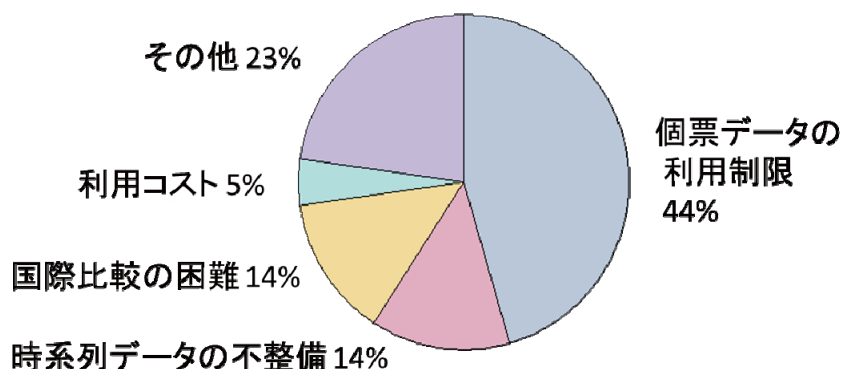


図2 既存統計・データ利用上の課題・問題点（全44件）

(2) インタビュー調査結果のクラスター分析

本事業ではアンケート調査と平行して、「政策のための科学」に関連する研究者・有識者 17 人に対してインタビュー調査を実施した。インタビュー調査の対象者は、例えば、『政策のための科学』に関連した研究開発プログラム」の採択課題関係者を始めとした科学技術政策関連の研究者、ファンディング機関におけるデータ整備・分析担当者、大学における学内データ整備・分析担当者などである。インタビュー調査における質問内容はアンケートの設問構成と同様である。

インタビュー調査で記録したテキストについて、統計解析用のフリーソフト「R」を用いてテキストマイニングの手法によりクラスター分析を試みた。以下、調査結果の中のデータ接続に関する 17 件の課題について分析した結果を述べる。まず、課題の文章(テキスト)を語句に分け、課題ごとに語句の出現頻度を求め、2 つ以上の課題に含まれる語句を抽出すると 94 種類、534 個あった。これから近い課題を塊(クラスター)として図示した結果を図 3 に示す。これを見ると、大きく 2 つのクラスターに分かれる。上部のクラスターには課題:10, 11, 6, 2, 13, 7, 14 が含まれ、324 個の語句があった。この上部クラスターでよく表れる単語に「人」があった。これは全体では 8 個あったが、そのうちこの上部クラスターに含まれていたものが 7 個であった。これを(7/8)と表す。「単位」については 13 個すべてがこのクラスターにあった(13/13)。「ID」(6/7)、「調査」(6/7)、「結合」(5/6)、「個人」(5/5)、「機関」(4/4) といった語句がこの上部クラスターに特徴的な語句として

使われていた。

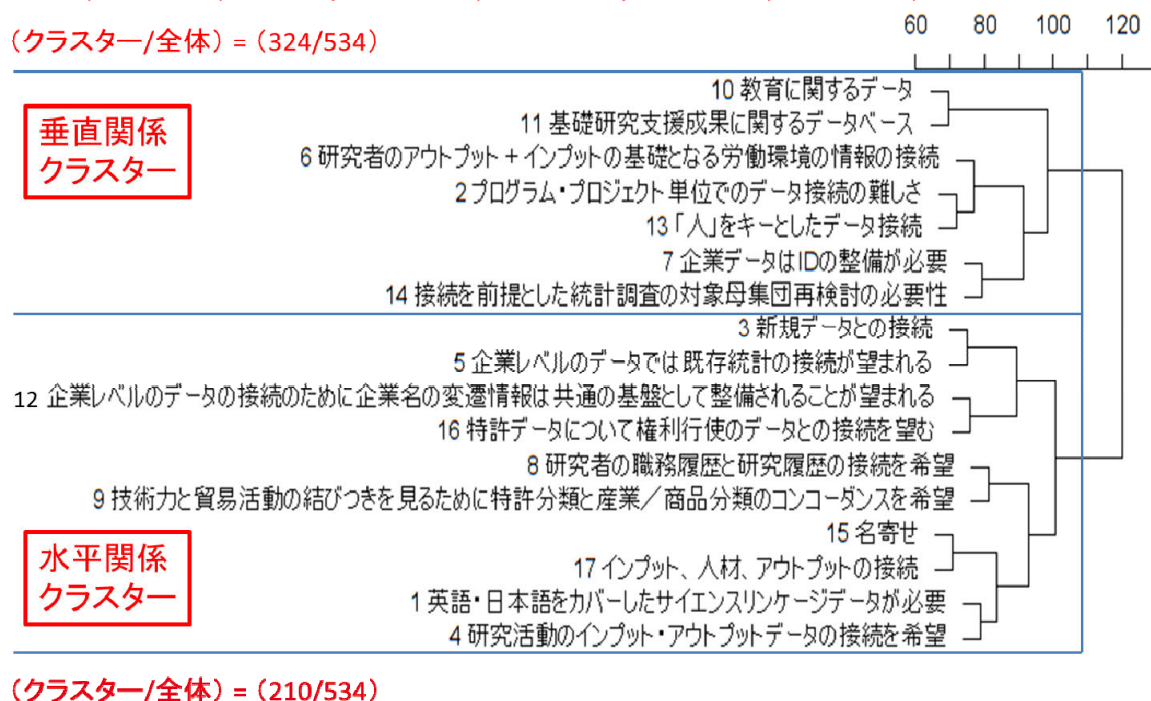
一方、もう一つの下部クラスター(課題:3, 5, 12, 16, 8, 9, 15, 17, 1, 4)では「特許」(13/16)、「論文」(5/7)、「活動」(5/6)、「履歴」(5/5)、「インプット」(4/5)、「アウトプット」(4/5)、「レベル」(4/5)といった語句が多く使われていた。

「政策のための科学」で扱うデータ・情報には、研究者、研究資金といった「インプット」、論文、特許といった「アウトプット」があり、「インプット」から「アウトプット」へと成果が「水平」に流れる関係にある。下部クラスターにある課題は、データ接続の課題の中でこれら「水平」の関係にある課題のまとめ、「水平関係クラスター」であると考えることができる。

一方、データ接続を実施するには、「人」、「個人」を「単位」としてデータを接続する場合がある。また個人の集合体である「機関」を「単位」としてデータを接続する場合もあり、これはデータ接続において「垂直」の関係にあるものである。上部クラスターにまとめられているものは、使われている語句の頻度からこの「垂直関係クラスター」であると考えることができる。

このように、課題を語句に分けて、近い課題をクラスターとしてみいていくと、データ接続に関する「水平」、「垂直」の構造を抽出してグループ分けできることが分かった。

単位(13/13)、人(7/8)、ID(6/7)、調査(6/7)、結合(5/6)、個人(5/5)、機関(4/4)



特許(13/16)、論文(5/7)、活動(5/6)、履歴(5/5)、インプット(4/5)、アウトプット(4/5)、レベル(4/5)

図3 インタビューにおける「データ接続」に関する課題のクラスター分析

3. 「政策のための科学」に資するデータ・情報基盤の課題に関するまとめ

データ・情報基盤の課題を収集し分析を試みたが、最後にこれら委員会での議論やアンケート、インタビュー調査を通して、「政策のための科学」に資する「データ・情報基盤構築」へのニーズや課題についてまとめる。

(1) ミクロデータの整備

委員会が出された意見、アンケート、及びインタビュー調査のいずれにおいても、「政策のための科学」に取り組む研究者にとって、政府統計の個票データをはじめとする各種ミクロデータ利用へのニーズは非常に大きい。政策の効果の測定や分析、大学などの機関についての構造的な分析には、従来の集計データのみでは不十分であり、ミクロデータに基づく要因分析が必要である。

(2) 複数データベースの接続

研究者や政策担当者が問題とする様々な課題は、個別のデータベースで対応できるものから、複数のデータベースを用いた横断的な分析が必要なものになりつつある。そのためのデータベースの接続・統合には、個別機関や研究者をキーとするなど、マイクロレベルでのデータ接続が必要である。データ接続を実現するためには全体的な構造や個別の技術的課題を詳細に検討する必要があるが、これが実現すれば、様々な分析が可能になるだけでなく、データ提供者の負担軽減につながる可能性もある。

(3) 政策評価に資することの重要性

委員会では具体的なデータ・情報基盤の議論とともに、「政策のための科学」における「データ・情報基盤構築」の目的について議論した。その結果、適切な政策評価に資することが重要であるとの指摘がなされた。政策評価を客観的なエビデンスに基づいて実施するためには、まず評価対象となる政策目標を明確化したうえで、評価可能な指標を設定し、それらのエビデンスに基づいた議論をする必要があるとの指摘がなされた。

(4) 既存データの利用ニーズ

「政策のための科学」に関連する研究の実施に当たっては、科学技術研究調査、学校基本調査、企業活動基本調査などを利用している、あるいは今後利用したいとの意見が比較的多く見られた。

4. 現状と今後の進め方

委員会での議論、アンケート及びインタビュー調査では、上記以外にも国家レベルで実施すべきことも含めて数多くの貴重な意見をいただいた。ここでは、特に科学技術政策研究所が実施、あるいは関与している「政策のための科学」における「データ・情報基盤構築」の現状と、今後の進め方についてまとめる。

(1) 科学技術政策研究所の実施する2つの個別事業

科学技術政策研究所が実施している2つの個別事業「公的研究開発システムにおける科学知識生産に関するデータ整備」と、「産業におけるイノベーションに関するデータ整備」は、ともに個別機関レベルでデータを整理し、論文、特許などのデータを研究成果、企業財務データなど、関連する他のデータに接続・拡張しようとするものである。これは調査の上明らかとなったマイクロデータの整備、複数データベースの接続という研究ニーズに応えるものといえる。

また、これらデータ・情報基盤の整備は、継続的に進めなければ意味がないとの指摘が多数あり、科学技術政策研究所でも長期的に取り組む課題であると考えている。また、現状では論文、特許といった、明確で定量可能なデータからの接続、拡張を計画しているが、さらに、他の定量化できる指標や、定性的指標なども政策提言のためには必要であると考えている。

(2) 次期 e-Rad 構築等におけるデータベース連携の取り組み

科学技術政策研究所を含む関係機関では、次期府省共通研究開発管理システム(e-Rad)の構築について、関連データベース(ReaD & Researchmap、科学研究費補助金データベースなど)との連携を図りながら検討を続けている。この取り組みは、複数データベースの接続とともにデータ提供者の負担軽減、データの精度向上を図る試みである。

(3) 研究者と政策担当者の相互交流

「政策のための科学」は「科学」であると同時に、「政策」決定のプロセスにも密接に関連している。従って、「科学」「政策」の両面を常に意識する必要があると、研究者と政策担当者のオープンな相互交流が不可欠である。科学技術政策研究所では、「科学技術イノベーション政策のためのデータ基盤の構築に関する国際会議」を2012年2月に開催し、国際的な研究者間の交流と、国内外の研究者と日本の政策担当者とのコミュニケーションを図った。今後も引き続き、「政策のための科学」における「データ・情報基盤の構築」を実質的に推進させるための研究者と政策担当者の活発な交流が必要であろう。

〔謝辞〕

本研究は、文部科学省の「科学技術イノベーション政策のための科学」事業の一環として、2011年度に科学技術政策研究所が三菱総合研究所に委託して実施した「データ・情報基盤の全体システム設計及びデータ提供事業の推進」事業の成果に基づき、これに分析を加えたものである。