資料３

統計システムの高度利活用に関する三つの提言

出口弘、榊俊吾

提言要旨

統計のシステムは、社会の情報を集約し，それを意思決定に利活用する手段として，情報技術の発展や社会の複雑化の中でその姿を変化させつつある。特に求められているのが、社会経済の姿を的確かつ時間遅れなく把握し政策や経営に活用するためのフレームワークとしての統計システムの高度利活用に関する社会的アーキテクチャの確立である。この提言ではそれを三つの具体的な項目にブレークダウンし、そのアーキテクチャに関する提案を行っている。

その上でここで提案された統計のシステムの提言事項を、中期，長期的課題として、基本計画の5年間で部分的には実現し，或は検討すべき事項として、統計委員会の報告事項に盛り込むべきと考えこれを提案する。これらの内中期的課題は，次の5年で道筋をつけ、さらに長期的課題は２０１３年にとりまとめる次期の基本計画で具体的な整備計画について閣議決定事項とすることを目標とする。なお以下で提案する内容は、技術的には現在でも十分実現可能なものであり、実現に関する道筋をつけることと制度的実施可能性に関する合意形成が委員会での大きな課題となる。

目次

項目の詳細

（１）エビデンスベースの政策実現のための統計データの高度利活用案ー政策策定機能の向上を目的とした2次、３次の加工統計の構築を可能とする統計システムの最適化のためのフレークワークに関する提案ー

提案要旨：基本計画の5年間で様々な統計のデータ間のリンケージを利用できるようなシステムをデザインすることで、１）自在に必要な加工統計や将来推計モデルを構築できエビデンスベースのポリシーを提供できる基盤となるシステムを整備する。省庁内研究者、政策担当部局がオンサイトで加工を実施し得る情報システム環境を整備する。２）一般の統計目的には、オーダーメード集計を安全に提供する枠組みを提供する。そのために個人情報につながるような集計はフィルタリングされるような、専用の集計言語でオンラインで受け付けし、自動的に得られた集計結果をオンライン公開できる体制を整備する。

（１－１）目的

過去のデータにとどまらず、リアルタイムで省庁などの政策策定の高度化のためにデータを利用できる環境を提供する。

（１）政策立案機能の強化と学術目的のための利用

本提案の目的は、様々なステークホルダーのための広義政策策定の高度化のために統計を利活用するためのシステムの最適化にある。むろん学術レベルでの先端的なデータ解析は、データの新たな政策利用の可能性を切り開く重要なルートであり、官学連携により省庁が責任を負う形で研究者のオンサイトによる分析実施の可能性は開かれ得る。

なお個票情報の特定できないオーダーメード集計データ等の提供については、学術目的か商用利用かを問わず、広く一般に公開される必要があり、そのためのシステムの構築も課題となる。

（２）省庁でのエビデンスベースポリシーのための利用のための最適設計

統計の高次利活用のためには、その二次加工、三次加工のための省庁間でのデータ共有と利用（SNA, 産業連関表、人口統計による将来の様々なリソース推計、CO2排出量と税制の関係に関する総合的な推計、国土計画等）が不可欠であり、そのためのシステムの最適設計を行う必要がある。

（３）統計の高次加工としての将来推計の必要性

人口はじめ確度の高い将来推計を新たな統計（三次加工統計）として提供できる環境を整備することで、エビデンスベースの政策立案が容易になる。またこれらの必要性から二次加工や、一次統計の必要性についてのフィードバックをかけることが重要となる。

（4）統計データのリンケージの必要性

様々な統計データがリンケージできる体制を整備することにより、現在はない新しい加工統計を容易にオンデマンドで構築できる環境を整備する。

注釈：1）エビデンスベースポリシーは、黒田委員他により提起されている。
2）ここでいう二次加工はSNAのような加工統計。三次加工はそのうえの将来推計等のことである。例えば人口動態に関する将来推計等を示している。人口推計等の将来推計も海外で統計の範囲に含まれていることは、貝沼統括官により報告されている。

（1－2）方法と効果

（1）データとプログラムの分離

各省庁のデータとプログラムを切り離してデータを抽出する。（システム特有のデータ形式から一般的に可読なデータ形式への最小限の変換）これにより現況の省庁の固有の利用方式については大きく変える必 要はない。他方で新たなETLツール等が利用可能になる事で，省庁横断的な串刺し集計や、新たな加工統計の構築が可能となる。

更に省庁が抱えている統計人材の縮小等による統計の質の確保等 の、業務最適化と質の維持の確保と言う課題に貢献できる。

（2）システム実現に向けてのプロトタイプ開発

1）システムの実現には、研究開発が必要なため、産官学が協力し てR＆Dの共同プロジェクトで具体的な期限を区切って，プロトタイプを作成し、機能や運用コスト等の評価を行う。
2）システムのリンケージのためには各省庁のデータからの抽出し利用するために、何らかの形でデータウェアハウスを設計することが必要となる。これはプロトタイプレベルでは、データとプログラムを分離することで二次加工のためのプラットフォームとして比較的容易に構築できる。
3）データウェアハウスからデータを抽出変換する ツールを提供することで、オンデマンドで加工データを作成しそれを保存できるようにする。(ETLツール)
4）行政情報で秘匿性が高く、それぞれの省庁に置かざるを得ない物（例えば現状では税務データがそれに相当すると主張されている）は、ETLツールを使ってデータを統計的に抽出し、それをデー タウェアハウスのデータとあわせて利用できるように、ETLツールを整備する。
5）データウェアハウスの利用と高次統計の構築によるエビデンス ベースの政策の

ために、二次加工、三次加工側のユーザからの視点 を取り入れることが不可欠となる。政策に関するPDCAサイク ルを考えるとき、チェックとアクションは、政策立案や利用に関す る広範なステークホルダーを含む必要がある。特に統計では二次加工統計側からの参加が不可欠となる。例えばSNAのような政策 に直結する二次加工統計では、第二WGで日銀から改革案が具体的に出 ているように、ユーザからのフィードバックが効果的に機能している。

注釈：１）ETLツール
　ETLツールとはデータベースから抽出(Extraction)、変換(Transformation)、データベースへの新たな書き込み(Loading)を行うためのツールで、データウェアハウスの高度利用のために近年開発が進んでいるもの。統計データの情報処理のコンテクストでは、変換にあたるものが様々な加工集計に対応する。経済社会総合研究所と東工大の連携で、2007年度に開発されたAADLという言語は、国民経済計算のためのデータ集計と加工のための一種のETL言語である。

（3）データ・アクセス情報の管理の徹底について
　データの管理については、データが個人の情報に関してトレーサブルとなることで、個人情報の保護が問題とされる。この管理に関しては、特定の個人や組織に関するトレーサブルなデータへアクセスする行為そのもの記録として管理しそれをトレーサブルにすることで管理する方式（ダブルトレーサビリティ）により管理する方式がある。
　データそのものを管理することには限界があるが、データへアクセスした記録を管理する（ダブルトレーサビリティ基準の実施）により、より実効的にデータ秘匿を管理することができる。個票データへのアクセス者のリストアップにより、万一データ漏洩が生じた場合の漏洩元の特定が可能になり、厳罰を課すことによりデータ漏洩の実効的な防止を図る。

（2）民間電子データを公共財としての利活用し、その集計データを民間へフィードバックするためのフレームワークに関する提案

提案要旨：官庁の業務データのみならず、民間でも多くの経済社会のデータが電子データ化されつつある。あらゆる社会の業務プロセスが電子化される中で、民間の電子データから電子的に抽出されたデータの統計目的での利用に関しての明確な指針を基本計画の5年間で策定し、併せてその技術的基盤と実証例をパイロットモデルで具体的に検証し、公共財としての民間電子データのあり方について法的位置づけを行う。またそのような公共材としての統計がどのように民間にフィードバック

され、様々な組織の意思決定に有効に用いられるかのビジョンも明確にする。

（２－１）目的

１）官庁の業務データのみならず、民間でも多くの経済社会のデータが電子データ化されつつある。それらの情報そのものは様々な形で統計的に利用されているが、現状では多くは改めて「紙」の調査票へ記載する形で用いられる。これらを業務システムから直接電子的に抽出することで、調査客体の負担の大幅軽減と調査の質の確保が両立できる。
２）現状の調査票設計で、調査客体の負担が調査項目を増やす事に対する限界となっている。更に例えば、労働時間等の業務プロセスのデータから、本来当該組織の業務に無関係の見なし雇用人数の計算等、統計目的の計算を調査客体の側に強いることがある。電子的な抽出が可能となれば、調査項目を増やす事や、統計的なデータ収集に関する計算を調査客体に負担をかけず行う事が可能となる。
３）公共財としての統計データを様々な業務データから抽出することに関する広範な社会的合意の形成が必要になる。これらには、地方自治体の業務システムからの決算統計のフィルタリングや、行政情報からの抽出（提案１のデータウェアハウスの活用）など多くのテーマが含まれる。これらを整理することで、公共財としての電子的抽出データの統計利用に関して、国民や産業界等多くのステークホルダにとってのコストと便益が明らかになる。
４）社会経済の現状の十分な把握なくしてエビデンスベースの政策立案や諸提言は機能しない。

（２－２）方法と効果

（１）抽出フィルターの開発と提供

ビジネスプロセスから必要な統計データを抽出するフィルターを開発提供する。ビジネスプロセスから必要な統計情報を電子的抽出するには、業務プロセスのデータから統計に関係する部分を抽出するフィルターを実装することになる。その技術的課題は大きくは困難はないが，具体的な事柄についてはR＆Dが必要となる。

会計ソフトがXBRLを出力するモジュールを持つのはその一つの事例。公会計のシステムや、ERPなど様々なシステム向けにこのような抽出モジュールを提供することで、複雑な労務データに関して、調査客体に記入負荷をかけずにより詳細な統計データの抽出が可能となる。

このようなフィルターを政府の開発で提供することで、結果として民間に負担をかけることなく、結果としては安価で高度なデータを様々な電子的な業務プロセスから業務プロセス固有の秘匿すべきデータに影響を与える事なく抽出することが可能となる。

（2）既にあるXBRL、JANISなどの成功事例を参考とする

本来抽出したいデータは、業務プロセスの中で電子的に存在しており、付加的な抽出機能や統計利用に際してのコストを負担することで、客体の負担を最小として個別の業務プロセスの中から統計に必要なデータだけを抽出することが可能となる。このようなモデルの成功例に、日銀による銀行のデータのXBRLによる抽出がある。さらにEDIネットでの財務諸表の公開もXBRLで行われている。

同様に、JANIS院内感染対策サーベイランス(http://www.nih-janis.jp/)は、病院の検査システムからの感染データの抽出によるデータを集計し，医療機関で実施されている院内感染対策を支援している厚生労働省の承認統計であり、ここでも病院のオーダリングシステムを通じてデータを抽出している。

（3）公共財としての民間電子業務データ利用に関するコンセンサスと法的な課題整備

現状、POSデータ等は購入することが多いが、社会にとって基盤となる統計データ収集の一環として、統計的な抽出に対して、民間側の負担を細小にすることで、公共財としての電子データの利用に関する，国民、企業などの幅広い合意を形成する。そのためには、そのデータが二次加工、三次加工（将来推計）などを通じて、どのように役に立つのかを示すことが必要。その上で、データの提供に関する法的な整備を行う。

更に業務データからの抽出が、企業や組織にとって利点となるには、早いスピードでの二次加工とそのフィードバックにより各々の組織が自らの位置を知ることができることが肝要となる。前出のJANIS院内感染対策サーベイランスが受け入れられているのも、自組織の位置がわかるからである。企業も投資や様々なレベルで早期の経済指標や産業構造の変化等を知り、自らの位置を知るニーズはあり、それが提供できることでコンセンサスを得られるようにすべきである。

（4）先導的な電子データの抽出可能領域

1）病院の電子カルテ、オーダリング、レセプトシステムなどからの抽出

1、レセプトデータの活用：レセプトデータは電子化が進みつつある。レセプト病名という概念に見られるように補正は必要だが、動的な病態変化等、患者実態調査では得られない動的な変化に関するデータが取得可能となる。（2008年5月から、400床以上の病院はオンライン請求が義務化されたが、2013年にはすべての医
療機関のオンライン請求が義務化される予定。）

2、総合的なオーダリングシステムからの抽出：薬の使用状況など

３、電子カルテからの抽出は、感染症等の危機管理に有効。これについては国立感染症研究所の大日先生の研究等で，効果が示されている。
２）企業のERPなどの業務システムからの抽出：取引実態を正確に反映した財務データが得られるばかりでなく、働き方に関する広範な統計の設計が，調査客体に負担をかけずに可能となる。また企業の主業，副業による事業所ベースでない仮想按分等もR＆Dにより可能となる。
３）POSデータの利用：POSデータを利用することで、従来の調査員調査では得られない、特売データの価額を含むデータなど多くのデータが得られる。これらの利用に関しては、様々なR＆Dは必要であるが、現状のパソコンに関するヘドニック法を利用した現状の数品目だけの利用はあまりに消極的。
４）パスモ、スイカなどの利用：現在SUICAはサーバ上で26週間、テープで１０年保管されている。これを利用することによりパーソントリップ調査では得られない、様々な外的事象による交通利用動態の変化など質の高く、災害時等に有用な多くの知見が得られる。

（３）統計システムの高度利活用のために産官学連携でのR＆Dを可能とするためのフレームワークに関する提案

提案要旨：基本計画の5年間で統計システムの継続的イノベーションプロセスを可能とする産官学連携の体制を確立するために、パイロットイノベーションプロジェクトでこれを具体的に検証しつつ、制度設計を行う。具体的には、１）タスクフォース方式あるいはミッションクリティカルな検討の委員会を立ち上げ、２０１１年までに（3年間で）統計システムの継続的イノベーションに関する体制構想を具体的に取りまとめる。２）具体的な課題を選定し、イノベーションの実証例を民間の協力を得てパイロットモデルで具体的に検証する。

（３－１）目的

　統計システムは、社会の変化に対応する必要がある。米国の2008年度大統領経済報告でも１章を割いて統計システムが時代に対応することの重要性、統計のイノベーションの必要性を強調している。そのために産官学で開かれた統計システムの重要課題に対して継続的なイノベーションが可能となる体制に関する構想を基本計画の5年間で策定し、併せてパイロットイノベーションプロジェクトでこれを具体的に検証する。
　統計システムに関しては多くのイノベーションすべき課題があり、それには科学研究上の課題もあるが、より現実的で目標を持って技術的に解決すべき課題が多い。そこで

１）政策立案や産業構造、社会構造の把握のための統計の2次加工、将来推計のための加工などについてどのような加工統計を作成可能か、それはどのように利用できるかについての継続的なR＆Dを行うことで課題や時代に対応した統計の高次加工を可能とする。
２）様々なデータソースから統計データを抽出し加工集計するための情報システム(ETLツール等）に関する継続的なイノベーションを行うことで、１）で述べたような様々な高次加工統計を構築することが可能となる。

（３－２）方法と効果

（１）加工集計の方法や技術に関するR＆D

１）現状でも原課（政策担当部門）の政策立案に使うヒアリングのデータと、統計データの間に乖離が生じつつあり、統計データが蓄積されるが、個別の原課レベルの政策立案や新規課題などに有効に用いられなくなる可能性がある。これに対して継続的にデータウェアハウスからの高次加工統計についてのR＆Dを行う事で現実問題に即した社会の統計データの構築が可能となる。

（２）統計データの抽出集計ツールに関するR＆D

１）現状の省庁別の統計システムでは集計のシステムは一体化できない。集計した結果を公開する機能は現在共同化が進められている、しかし集計する機能は各省庁が個別のノウハウを持っているので共同利用化は現状では難しい。他方でこの状況が続くと省庁間のデータを串刺しにした集計利用は難しくなる。オンデマンドでの串刺し集計や、加工は、提案（１）で述べたデータウェアハウスを構築し、その中で、統計データの抽出・加工集計のための情報システム(ETLツール等）に関する継続的なイノベーションを行い、膨大なデータに対する高度集計加工を可能とするツール群を開発する。（これらは集計加工が目的で、現状のRやSASなどの統計処理ソフトとは目的も機能も異なる。）
２）統計目的の大規模データウェアハウスに関する技術的R＆D
　提案１で述べたデータウェアハウスなどの基盤技術を確立するためのR&Dを行う。

（３）統計調査に関する技術や手法のR＆Dの事例

１）調査員調査の調査客体とのハンドシェイクや、リサンプリング、調査の確認などの一連のプロトコルの標準化のためのR＆D。調査のときの客体とのハンドシェイクのプロセスなどの標準化がなされることで民間からの参入も基準が明確になる。
２）電子的な抽出データに関するサブ母集団の偏りからの母集団推計に関するR＆D。これによりPOSデータなど電子的な大規模データだが偏りのある部分母集団か

ら得られたデータからもとの母集団統計を推測可能となる。
３）将来推計のモデルの共有のためのR＆D：例えば市町村単位での人口動態推計の標準化とそれによる医療、教育、年金などの諸リソースの推計。これらがエビデンスベースの政策として提起され、そのモデルにフィードバックがかかることで、継続的に改善可能という意味で有効な政策評価が可能となる。
4)データメンテナンスのフェール・セーフとデータクリアランス
　データの溯及や欠損値処理などに対するメンテナンスの研究は実用的に非常に重要。

（４）R&Dとイノベーションの遂行主体

　現在省庁内には十分なR＆Dの機能がなく、他方で産業界や学会だけでもニーズを明確にした迅速なR＆Dは難しい。そこで、産官学が連携できるプロジェクト式のスキームを用意する必要がある。例えば2007年度に内閣府経済社会総合研究所と東京工業大学のエージェントベース社会システム科学研究センターとの間で作られた，社会会計コンソーシアムは産官学連携のバーチャルな研究体として、1年でSNAの構築に関する新しいシステムに関しての多くのプロトタイピングや手法の開発を行うことができた。このようなミッションクリティカルなプロジェクトを必要に応じて実施できるスキームを構築することが重要である。

以上