

## 統計システムの高度利活用に関する三つの提言

WG 4 用資料 出口弘、榎俊吾

### 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。課題2に関しては実際に内閣府経済社会総合研究所と経済産業省などを中心に加工統計の処理のための研究開発のコンソーシアムを形成し、SNAやI/Oのための加工統計の構築プロセスなどを事例に研究開発を可能な範囲で行っていく。

### 目次

- （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  
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ムワークに関する提案
- （2）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とそのコンソーシアムによる具体化提案
- （3）調査客体に負担をかけない電子的なデータの統計用途での抽出と、その集計結果を素早く調査客体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

## 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 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

#### ( 1 · 1 ) 基本的な考え方

基本計画の5年間で様々な統計のデータ間のリンクを用意できるようなシステムをデザインすることで、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。府省庁内研究者、政策担当部局がオンラインでオーダーメイド加工を実施し得る情報システム環境を整備する。

#### ( 1 · 2 ) 取組の方向性

過去のデータにとどまらず、リアルタイムで府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。

##### ア) エビデンスベースの政策立案機能の強化

本提案の目的は、様々なステークホルダーのための広義の政策策定の高度化のために加工統計を利活用するための統計システムの整備にある。また官学連携により府省庁が責任を負う形で研究者のオンラインによる分析実施の可能性は開かれ得る。

##### イ) 府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、三次加工のための府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ソース推計、CO<sub>2</sub>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、そのための加工統計を用意に構築できる統計システムの最適設計を行う必要がある。

##### ウ) 統計の多元性（根拠を示した多様な加工統計の提示）と高次加工統計としての将来推計の構築環境整備の必要性

統計の多元性（根拠を示した多様な加工統計の提示）は、高次加工統計では重要な考え方となる。様々な帰属計算やサテライト勘定は異なる視点からの加工統計の多元性を明示的に示したものである。更にエビデンスベースの政策のための根拠となる将来推計を高度加工統計（三次加工統計）として基幹統計の一部として提供するにあたっても、そこでは推計根拠を明確にした多元的なモデルの提示と選択、政策結果からモデルへのフィードバックが必

要となる。このような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。また政策上の必要性から二次加工や、一次統計にフィードバックをかけることができる。

### **エ) 統計データのリンクエージの必要性**

様々な統計データがリンクエ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政策的必要性等に応じて、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、統計の今後の利活用で重要な柱の一つになると考えられる。

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。ただしここでは統計の多元性を前提としている。高次加工統計は唯一のモデルを特定することを目指した物ではなく、根拠を明示して、多元的な視点での二次加工統計や将来推計のモデルを提供できることを目指す。二次加工では、既にサイト勘定という形で一般的である。また推計方法そのものについても多元的なパスの検討がなされている。

### **(1・3) 具体的な措置、方策等**

#### **ア) データとプログラムの分離**

各府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）。これにより現況の府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、府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。

府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。

#### **イ) 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。これについては、内閣府経済社会総合研究所、経済産業省等がコンソーシアムを作つて対処する。

2) システムのリンクエージのためには各府省庁のデータからの抽出・利用するにあたり、何らかの形で（仮想或は実態の）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テスト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オンデマンドでの串刺し集計や、加工が可能な統計データの抽出・加工集計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。

4) 行政情報で秘匿性が高く、それぞれの府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）は府省庁内で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することもできる。

## （2）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：コンソーシアムによる研究開発の具体化の提案

### （2・1）基本的な考え方

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整備を基本計画の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押し進める。

統計システムに関しては、科学的研究上の課題から技術的に解決すべきものまで多くのイノベーション上の課題がある。これは次の三つに大別される。

#### 1) エビデンスベースポリシーを担保する加工統計に関するR&D

エビデンスベースポリシー（根拠に裏打ちされた政策立案）に基づいた政策立案や産業構造、社会構造の迅速な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についての継続的なR&Dを行うことで時代に対応した統計の高次加工を可能とする必要性。

#### 2) ツール開発に関するイノベーション

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抽出と集計加工のためのツールに関する継続的なイノベーションを行うことで、様々な高次加工統計を支援する必要性。

### 3 ) 統計の多元性に関するシステムデザイン

統計の多元性（根拠を示した多様な加工統計の提示）を可能とする迅速な統計・推計データの提供のためのシステムはどのように構築されるどのように利用可能になるかについてのシステムデザインの必要性。

#### （2・2）取組の方向性

##### ア) 現状

1 ) 現状府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ていない。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が、集計する機能は各府省庁が個別の言語とノウハウを持っているので現状では府省庁間のデータを串刺しにした集計利用は難しい（付録図1）。

2 ) エビデンスベースの政策立案のための統計利用は、統計作成目的であれば現在でも目的外利用制度が整備されており、制度上の問題はない。しかし、政策決定に影響を与える高度な活用事例は経済モデルなどごく一部に限られており、今後さらに広範囲な分野での活用の余地がある。

##### イ) 取組の方向性

###### 1 ) エビデンスベースポリシーを担保する加工統計に関するR&D

官学の連携で幾つかの具体的なテーマを定め、R&Dのためのコンソーシアム的枠組みでこれを研究開発することで、次の5年間の間で統計の高度利用に関する問題解決のためのR&Dとプロトタイピングを行う。

###### 2 ) ツール開発に関するイノベーション

オンデマンドでの串刺し集計や、加工が可能な統計データの抽出・加工集計のための情報システム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。

###### 3 ) 統計の多元性に関するシステムデザイン

統計の多元性を可能とする迅速な統計・推計データの提供のためのシステムをデザインする。

#### （2・3）具体的な措置、方策等

##### （1）統計の高次利用に関する官学連携によるR&Dのためのコンソーシアム

具体的なイノベーションスキームとして、内閣府経済社会総合研究所と経済産業省などを中心に加工統計の処理のための研究開発のコンソーシアムを形成し、大学、学会等とも連携し、SNAやI/Oのための加工統計の構築プロセスなど、高度IT利活用による様々な加工統計作成や統計の高次利活用のための研究開発を推進する。

コンソーシアムでは具体的に下記のテーマの検討が行われる。

### 1 ) SNAやI/Oのための加工統計の構築プロセスに関する研究

様々な一次統計や関連データを用いて、内閣府はSNAを、経済産業省は産業連関表を策定する。これらの作業は相互に密接に関連し、これらの統計やデータをベースに遅滞なく二次加工統計を作成するためには、府省庁の壁を越えた個票から加工統計まで新しいデータリンクエージと統計システムの高度利用のスキームの開発が必要となる。プロジェクトでは、このようなシステムの基盤設計と実際の推計作業に可能な限り使えるシステムのデザインを行う。

### 2 ) 省庁内の多様なデータソースの統合的利用に関するシステム分析

各種の統計、データ等を用いて作成されているIIP(鉱工業生産指数(Indices of Industrial Production)、第三次産業活動指数等についても、必要な二次加工をタイムリーに行えるようなシステムの設計とプロトタイプを試みる。さらに、これを発展させ、様々な課題に対して省庁内外での高度利用を可能とすることも検討する。

### 3 ) 企業に負担を与えない統計データ収集法に関する研究

日銀が銀行の考查に用いている、XBRL(eXtensible Business Reporting Language)など、企業の情報システムからの情報の直接的抽出は、企業に負担やリスクを与えずに、企業から早いレスポンスで産業や雇用に関するデータを収集することは今日の政策運営で極めて意義が大きい。また産業界にとっても自社が全体の中でどのような位置にあるかの情報の素早いフィードバックが受けられるのであれば、情報を出すことへのインセンティブは高い。

また金融等の分野では特に今後社会の安心安全を担保するためにも、リアルタイムのセンサリングが市場や機関に対して必要となる。このようなシステムの利活用には、国民や企業の広範な合意も必要となる。本プロジェクトでは、このような業務プロセスからの統計データの抽出による新たな統計データの収集法を関連部局の協力の下でデザインする。

#### 4 ) 統計の高度加工、統計システムのソフトのオープンソース化等のビジョンの構築。

上記のような具体的課題の研究開発を通じて、統計の高度加工（二次処理としての加工統計の作成や、さらにそれを利用した将来推計などの三次加工）のあり方や、統計システムのオープンソース化等について、技術に裏打ちされた枠組みとビジョンを構築する。

#### （3）民間電子データ集計の利活用と集計データを民間の意思決定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

##### （3・1）基本的な考え方

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、調査客体に負担をかけない電子的なデータの統計用途での抽出と、その集計結果を素早く調査客体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する。またそのような統計情報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。

##### （3・2）取組の方向性

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 ) 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なども課題となる。

4 ) 社会経済の現状の迅速かつ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。

### （3・3）具体的な措置、方策等

#### ア) 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（enterprise resource planning）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する、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、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。これは当該組織の意思決定情報の抽出のためにも有益となる。

#### イ) 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、附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。

#### ウ) 民間電子業務データ利用に関する課題整備とインセンティブの提供

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することで、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かを示すことが必要。その上で、データの提供に関する課題を洗い出す。

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院内感染対策サーバ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業界のコンセンサスを得られる可能性を模索する。実際、自らの位置を知るというフィードバックが（回答負担軽減だけでなく）が情報提供の積極的なインセンティブとなり得る。このほかにも自社の属する業界および取引先業界における平均的な成長性、安全性等の財務指標、倒産、製造物責任等のリスク、商品、研究開発力等の指標を早いフィードバックで業界に投げ返すことは情報提供の重要なインセンティブとなる。

#### 参考図

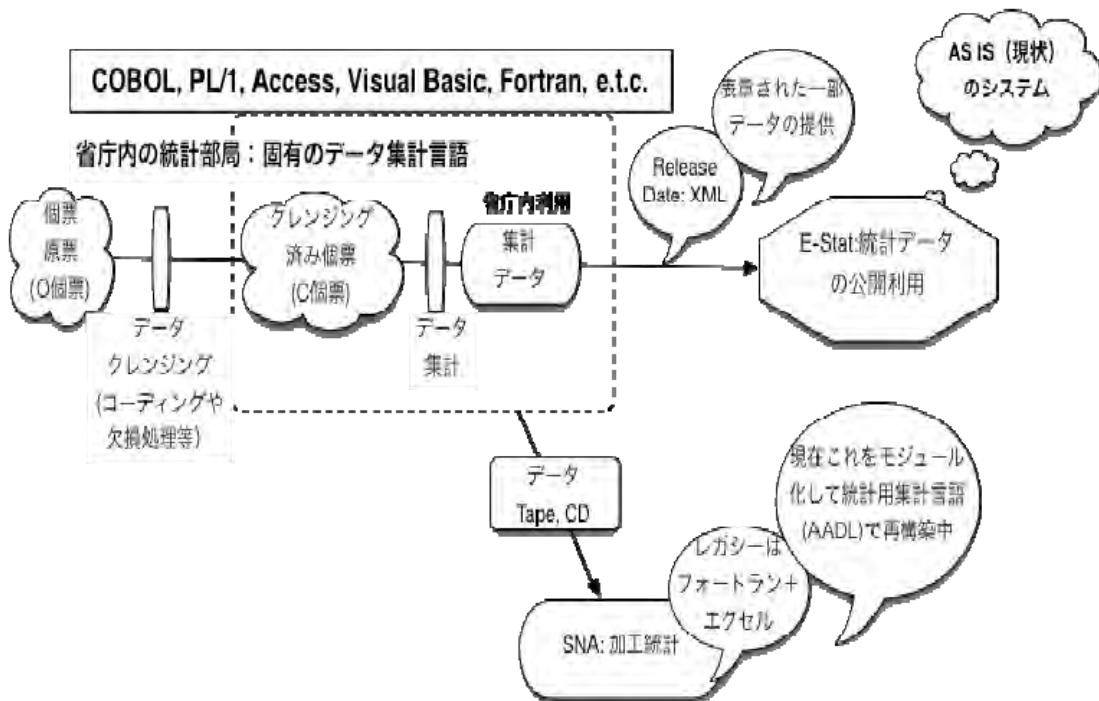


図 1 現状の統計システム

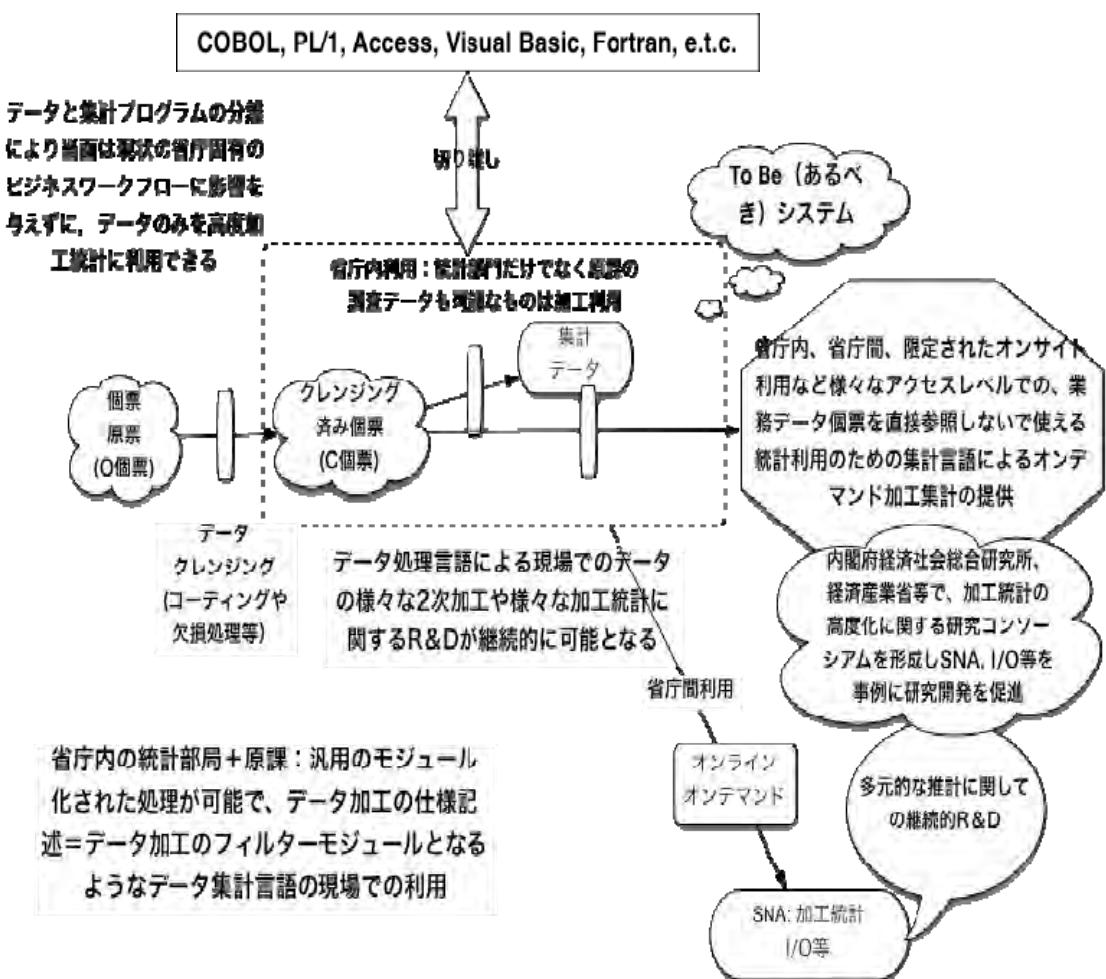


図2 エビデンスベースポリシーのためのあるべき統計システム