

人間・社会データ基盤整備
— 公的統計データのネットワーク型共同利用基盤 —
Development of Human and Social Data Basis
-Network Environment for Analyzing Official Statistics Micro Data-

Hiroe Tsubaki

The Institute of Statistical Mathematics

講演内容

- Needs: 人間・社会研究における公的統計活用
 - 国際競争力を低下させる日本の社会科学研究
 - リスクとエビデンスに基づく政策決定とその支援研究の在り方
 - 民間ビジネスとしてのリアルタイム統計
- Plan: 公的統計マイクロデータ活用のためのネットワーク基盤
 - 総務省「統計の2次利用研究会」
 - 川崎応用統計学会長(当時)「学術会議マスタープラン」提出
 - 内閣府統計委員会の「公的統計の整備に関する基本計画」
 - 基本計画2014年3月25日閣議決定
- Do 情報・システム研究機構の大学共同利用機関としての役割
 - (独)統計センターとの連携協定
 - 日本の公的統計以外の関連活動

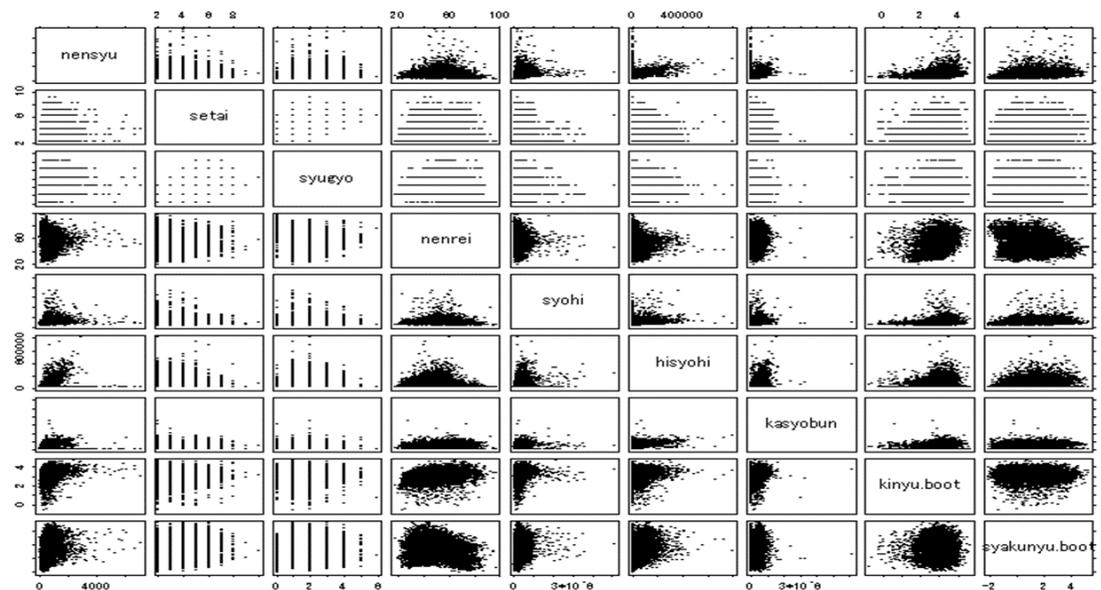
国際競争力を低下させる日本の社会科学研究
リスクとエビデンスに基づく政策決定とその支援研究の在り方
民間ビジネスとしてのリアルタイム統計

NEEDS: 人間・社会研究における公的統計活用

人間・社会研究における公的統計活用

- 国際競争力を低下させる日本の社会科学研究
 - 1980年代欧米で大きな変化
 - 自国公的統計マイクロ(個票)データの研究利用が可能に
 - 経済・社会研究の中心が集計データからマイクロデータ分析にシフト
 - 日本の経済学者の危機感: 科研費特定領域:1996-1999
 - 統計情報活用のフロンティアの拡大の総括的研究:松田芳郎(一橋大)
 - 13研究班の要請を集約:5省庁17調査について、総務庁長官に「目的外使用申請」
 - データ処理センターとして活動
→現行のオーダーメイド集計制度
 - リサンプリングデータの作成→現行の匿名データ提供制度
 - イミテーションデータの作成→現行の教育用擬似データ制度
 - データ処理結果および処理要求をデータの秘密保持を保ったうえで各研究班相互を計算機ネットワーク化して計算結果等の情報を流通させるシステムを開発
→今後のリモートアクセス制度
 - 日本評論社からのマイクロデータ分析シリーズにより啓発
 - 新統計法下のマイクロデータ研究利用の先駆け研究

一例たりとも実データが存在しない
家計調査+貯蓄動向調査のイミテーションデータ
重回帰分析位ならば再現する(椿:1998)

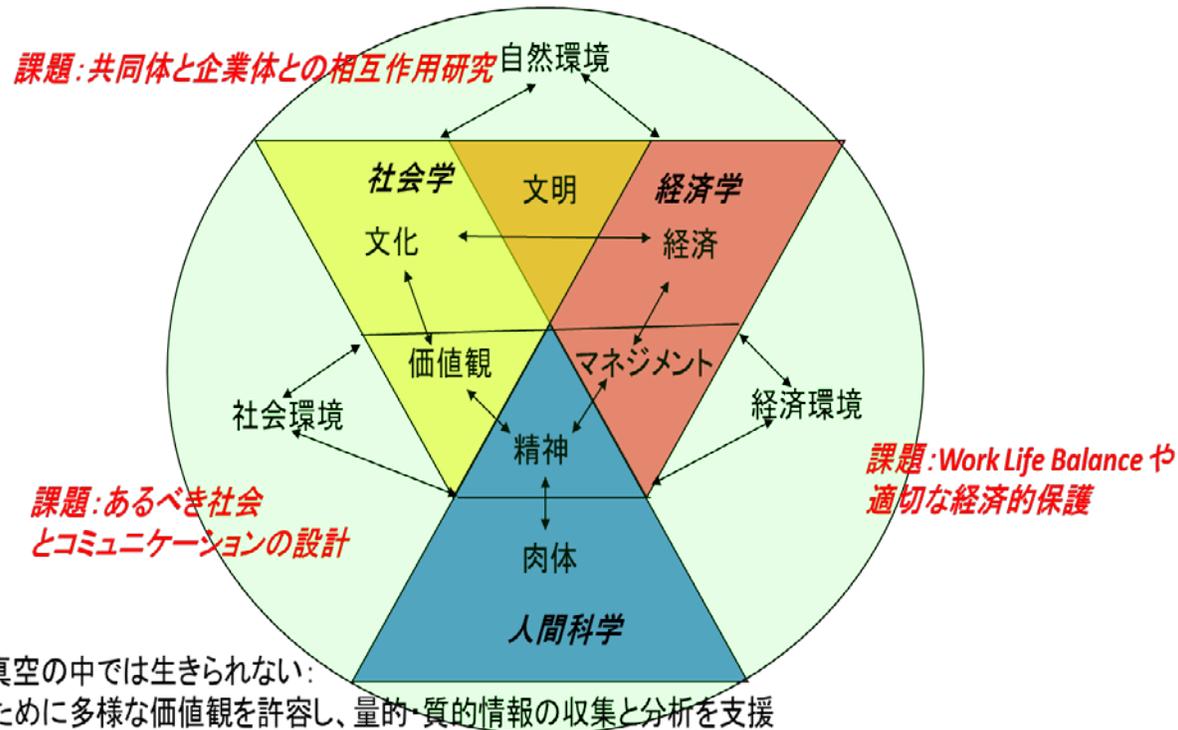


リスクとエビデンスに基づく政策決定とその支援研究の在り方

人(ミクロ)と共同体・経済(マクロ)との相互作用研究に必要なマイクロデータ

社会政策科学の難しさ: 人と社会と経済との相互作用をあるべき姿にする実学的活動
 対策の最前線・政策立案実施の現場とどうコミュニケーションできるのか?
 人文社会科学と健康科学とはどう融合し、政策の最適化を支援するのか?
 文理融合のみならず、文文融合(経済学+社会学+法学)をどう実現するのか?

- 2006/10/26: Scientific Advice,
- Risk and Evidence Based Policy Making: 英国下院科学技術委員会
 - 勧告14, 15: 政策立案総合職の要件: 証拠の利用と分析の専門的技量
 - 科学的方法, ピアレビュー, 証拠の多様な関連性の役割と重要性, それらを評価する方法に対する基本的理解
 - 科学的方法と多様な証拠の解釈と共に科学的インプットと分析の十分説明された要請が総合職である高級官僚の中で展開
 - 勧告16: 総合職と共に科学的力量を持つ専門職の価値
 - 経済的知識や法案作成能力同様, 科学的リテラシーが政策決定には必要.
 - 勧告50: 透明性
 - 政策決定に用いられる全ての証拠とそれをどのように使ったかの公表
 - 勧告60: 「予防原則」に代わる正しい予防アプローチの適用
 - リスクとベネフィット, 不確かさの一貫した説明. 予防原則を科学的知見が不確かなときに, 徹底したリスク分析に替わる政策決定原理として用いるべきでない
 - 勧告61: 科学的不確かさの下でのリスクマネジメント理論の実用化と意思決定過程の効果的コミュニケーション



統計家は、真空の中では生きられない:
 この統合のために多様な価値観を許容し、量的・質的情報の収集と分析を支援

民間ビジネスとしてのリアルタイム統計

- 公的統計の基本
 - 調査対象集団(標的母集団)を正確に定義
 - そこからランダムに対象を選ぶ
 - 正確性(偏りの無さ)
- **ビッグデータは母集団を代表するデータなのか否かは大問題!**
- 1936年米国大統領選挙
 - ギャロップ調査
 - Roosevelt 54%, n=3000 vs.
 - リテラシーダイジェスト: 電話調査
 - Landon 57%, n=200万
 - <http://www.stat.go.jp/teacher/c2epi4a.htm>
- 現在理論的にはPropensity Scoreが実用化
 - 補正技術
 - しかし, 補正の拠り所は必要
- 一藤・曾根原(NII)プロジェクト
 - **リアルタイム観光統計**
 - 情報はサイバースペースから収集
 - 公的統計データを用いて偏り補正
 - 年次統計→日次統計
 - リアルタイム統計
 - 産業利用可能性

総務省「統計の2次利用研究会」

川崎茂応用統計学会長(当時)「学会議マスタープラン」提出

内閣府統計委員会の「公的統計の整備に関する基本計画

基本計画2014年3月25日閣議決定

PLAN: 公的統計マイクロデータ活用のための ネットワーク基盤

2009年統計法全面改訂

- 旧統計法(1947年)法の目的
 - 第一条 この法律は、**統計の真実性**を確保し、統計調査の重複を除き、統計の体系を整備し、及び統計制度の改善発達を図ることを目的とする。
- 新統計法第一条
 - この法律は、**公的統計が国民にとって合理的な意思決定を行うための基盤となる重要な情報**であることにかんがみ、公的統計の作成及び提供に関し基本となる事項を定めることにより、公的統計の体系的かつ効率的な整備及びその有用性の確保を図り、もって**国民経済の健全な発展及び国民生活の向上に寄与**することを目的とする。
 - オーダーメイド集計や匿名データの作成・**研究目的などでの提供を可能とする統計データの二次的利用の制度**設立

統計データの二次利用促進に関する研究会

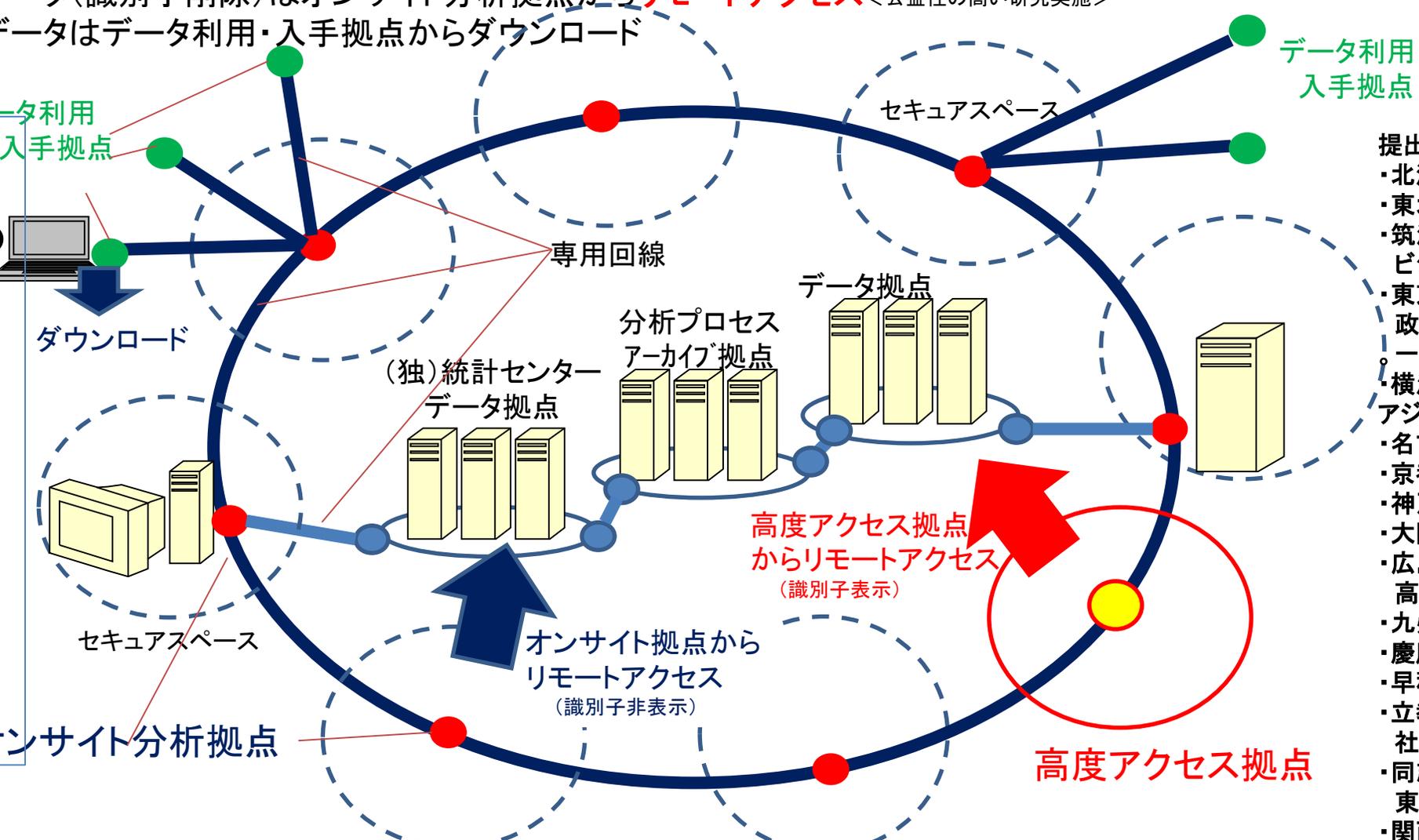
- 総務省政策統括官室: 2007年スタート
 - 廣松毅(東京大学, 現情報セキュリティ大学院大学)座長
 - 内閣府統計委員会へのInput
 - 統計法改正に向けた二次利用のスタイルと展開
 - オーダーメイド集計・匿名データの提供
- 利用目的としての公益性
- 統計法公布後の活動
 - 提供方法: 各国制度比較と日本の特異性
 - 目的外申請: 個人情報・法人情報が付随したデータをセキュアな環境を持たない研究者が管理する可能性
 - オンサイト拠点
 - 目的外申請で得た個票をセキュアな監視環境下で分析
 - 探索的なモデリングを可能とする
 - しかし, 各拠点ごとの人員・設備整備にかなりのコスト
 - リモートアクセスによるネットワーク形成: 応用統計学会からの学術会議マスタープラン提案
 - 各拠点にはデータは置かない: 監視は中央で一括
 - 統計データ・アーカイブの整備

- ◆ 秘密度超高のデータ(識別子付与)はごく一部の**高度アクセス拠点**からリモートアクセス<根幹を成す高度検証等に参画>
- ◆ 秘密度 高 のデータ(識別子削除)はオンサイト分析拠点から**リモートアクセス**<公益性の高い研究実施>
- ◆ 秘密度 低 のデータはデータ利用・入手拠点からダウンロード

公的統計マイクロ
データ等の
研究活用のための
全国ネットワーク
整備

川崎茂
応用統計学会長
2013年3月提出

2014年
学術会議
マスタープラン
登録



- 提出に協力いただいた組織:
- ・北海道大学
 - ・東北大学経済学研究科
 - ・筑波大学
 - ・ビジネスサイエンス系
 - ・東京大学経済学研究科
 - ・政策研究大学院大学
 - ・一橋大学
 - ・横浜国立大学
 - ・アジア経済社会研究センター
 - ・名古屋大学
 - ・京都大学経済研究所
 - ・神戸大学
 - ・大阪大学経済学研究科
 - ・広島大学
 - ・高等教育研究開発センター
 - ・九州大学基幹教育院
 - ・慶應義塾大学産業研究所
 - ・早稲田大学政治経済学術院
 - ・立教大学
 - ・社会情報教育研究センター
 - ・同志社大学
 - ・東アジア研究センター
 - ・関西大学

※ データ拠点: 2拠点 分析プロセスアーカイブ拠点: 1拠点 高度アクセス拠点: 4拠点
 オンサイト分析拠点: 18拠点 データ利用・入手拠点: 47拠点

公的統計の整備に関する基本計画への組み込み

- 内閣府統計委員会
 - 2013年10月30日総務大臣より諮問, 2014年1月31日答申
- 2014年3月25日閣議決定
- 調査票情報等の提供及び活用については、
 - セキュリティレベルや調査票情報等の匿名性の程度に応じた利用形態ごとの特性
 - 諸外国における取組状況等を総合的に勘案した上
- 法制度上の整理を含め、**以下の取組**を行う
 - オーダーメイド集計における利用条件の緩和に向けた検討
 - 調査票情報の提供における**リモートアクセスを含むオンサイト 利用やプログラム送付型集計・分析の実現に向けた整理・検討**
 - 匿名データの作成及び提供における提供対象統計調査の種類や年次 の追加等によるサービスの充実
- その際、効率性及び利便性の観点から、**政府一体として一元的な取組**を推進する。
 - 主担当総務省政策統括官, 府省横断の中核機関:(独)統計センター

DO 情報・システム研究機構の 大学共同利用機関としての役割

情報・システム研究機構の大学共同利用機関としての役割

- (独)統計センターとの連携協定: 2010/07/07
 - － 第2期中期目標・中期計画期間における新領域融合研究センターの研究プロジェクト
 - － 新たな研究テーマとして「人間・社会」
 - 新しい発想の先駆的な研究を強かに推進
 - その実施にあたって統計センターの持つ公的データ利用を行うため、本協定締結
 - － オンサイト拠点における探索的モデリングの可能性追求
 - － 連携協定機関として唯一非領域系機関
 - 共同利用データ基盤形成に軸足
 - － 中核公的統計機関(統計センター)と大学を繋ぐ中継拠点
 - 府省の壁を超えられるか?
 - － アーカイブ拠点: 人間・社会研究の知の再利用可能性
 - データリンケージ・エディットに大半の労力
 - 知財の壁を公益性は破れるか



その他の活動：活動の発展

- 公的統計情報から行政情報へ拡大
 - 厚労省レセプトNDBの研究利用
 - オンサイト拠点の厚生労働省からの認可
 - サンプルングデータセットに基づく医薬品安全研究を島根大学医学部と共に実施
 - 赤池ゲストハウス利用による滞在型研究
 - ヨーク大学との共同研究
- **公的統計マイクロデータ分析拠点をアジア10か国のマイクロデータに拡大**
 - 「マイクロデータラボラトリー」プロジェクトを公益財団法人統計情報研究開発センターと共に推進
 - アジア統計部局の統計家を毎年12月1週間3か国ずつ招へい
 - 実際に各国の匿名化マイクロデータを統計数理研究所で分析、プレゼンテーション



2014年12月に統計数理研究所で1週間合宿で開催されたマイクロデータラボラトリーWSに来日した、ベトナム、インドネシア、スリランカ統計局の統計家の樋口所長表敬訪問
後列は、研究所ならびに統計情報研究開発センターの方々

ROISのデータ基盤は、日本の人間・社会科学の進化と深化を強く支える苗代
「**マイクロデータネットワーク利用の全国研究機関コンソシアム**」を形成し、
実際に収穫を目指す
是非ご参画を

本プロジェクトを支えてきた、岡本基URAに謝意

結語