



2 次的利用の推進方針 の在り方についての 検討資料(案)

～ 制度立ち上げ時における対応等 ～

平成20年6月10日

総務省政策統括官室

議論にあたって

【理想的目標】

基幹統計調査を含めた全ての統計調査について、個々の統計の特性()を踏まえた二次利用サービスが展開されることを理想的目標とする。

事業所・企業を対象とした調査は、大企業のデータを丸めたりすると、統計結果、分析結果に大きく影響することからこのような処理が行いにくい。大企業データは丸めないと個別の企業名が判明してしまうリスクが大きい。したがって海外においても事業所・企業のデータを匿名データ化する事例は少ない。

行政記録と調査票の結合によるデータの匿名データ化は、行政記録を提供する側の意向に大きく依存

【本資料提示の目的】

理想的目標の達成に当たっては、一定の制限、不確定要素があることから、これらの制限、不確定要素を整理しながら、一歩でも理想的目標に近づくための方策を議論・検討するために提示するもの。

**制度立ち上げ期における
制限、不確定要素の整理と
これらを加味した対応、基本
計画との関係について**

2 次的利用の立ち上げに当たっての留意事項(1)

信頼の確保

匿名データの提供、オーダーメイド集計については、相当程度の事務作業を行い、国民の共有財産である行政資源を費やすことになるものであり、国民、調査対象者の信頼を確保できる範囲でのサービス提供とすることが必要

特に、匿名データは秘匿措置が施されているものの、個体別の情報が提供されることになるものであり、国民・調査対象者等の統計調査に対する信頼を損なわないように留意することが必要

二次利用の試行的な取組の実績

平成16年度より統計局が一橋大学と連携して取り組んだ匿名データ提供の試行的取組の実績(平成18年度後半より提供対象調査()を3調査 4調査に増加、助教以上の研究者を対象)

H16年度

6件

H17年度

18件

H18年度

33件

H19年度

27件

全国消費実態調査、就業構造基本調査
社会生活基本調査、住宅・土地統計調査(追加)

専門職員リソースの現状

調査環境の悪化

行政記録の活用

統計の最終
工程である
集計への負荷

集計職員 の努力による
基本統計の精度維持

これらの職員が二次利用を実施

二次利用サービスの立ち上げに当たっての留意事項(2)

匿名データ作成に必要な能力(想定)

匿名化する調査の特性に関する知見

各種の秘匿処理技法の知見、各技法の個別調査への適用についての評価能力

(秘匿技法) ・ トップ(ボトム)コーディング ・ リコーディング(再符号化)
・ スワッピング ・ ノイズ付加 ・ リサンプリング 等

開示リスクの評価能力(数理評価を含む)

匿名化に伴う情報欠損の評価能力(数理評価を含む)

新たな秘匿処理技法に対する知見(国際的な研究動向の把握)

情報処理・情報システムへの対応能力

等

オーダーメイド集計に必要な能力(想定)

調査の特性に関する知見

統計表の秘匿技術(2次秘匿(統計表の周辺合計からの加算・除算によってデリケートなセルの値が導出されないようにする補足的秘匿))

発展途上の秘匿処理プログラム(Argus等)の特性に関する知見

情報処理・情報システム等への対応能力

高度な分析要求(多変量解析結果の要求等)に対応する統計数理の知見

統計パッケージソフト(SAS、SPSS、Stata等)の活用能力

(プログラム持ち込みによるオーダーメイド集計では、ソースコードの不正処理やバグ等がないかチェックが必要)

等

これらの能力を多く有する職員の養成が急務

留意事項から想定される制限、不確定要素

職員余力に
大きな期待
はできない

秘匿技術や情報技術を有する公務員の確保は大きな課題

職員は通常業務を確実に行う必要がある
(新たな仕事を行う余力はない)

見通せない
需要予測、
予算獲得

新たな制度であり、需要予測がないことに加え、実際の需要に過度な期待はできない

公表されている基本統計表が存在している
ので、国民生活や社会生活へのメリット等が
説明しにくい

立ち上げ期における利用可能なリソースと
現実的な初期需要を前提に計画を構築

制度立ち上げ期における限りあるリソースの有効活用

サービス開始までは、技術の検討等一定の蓄積が必要

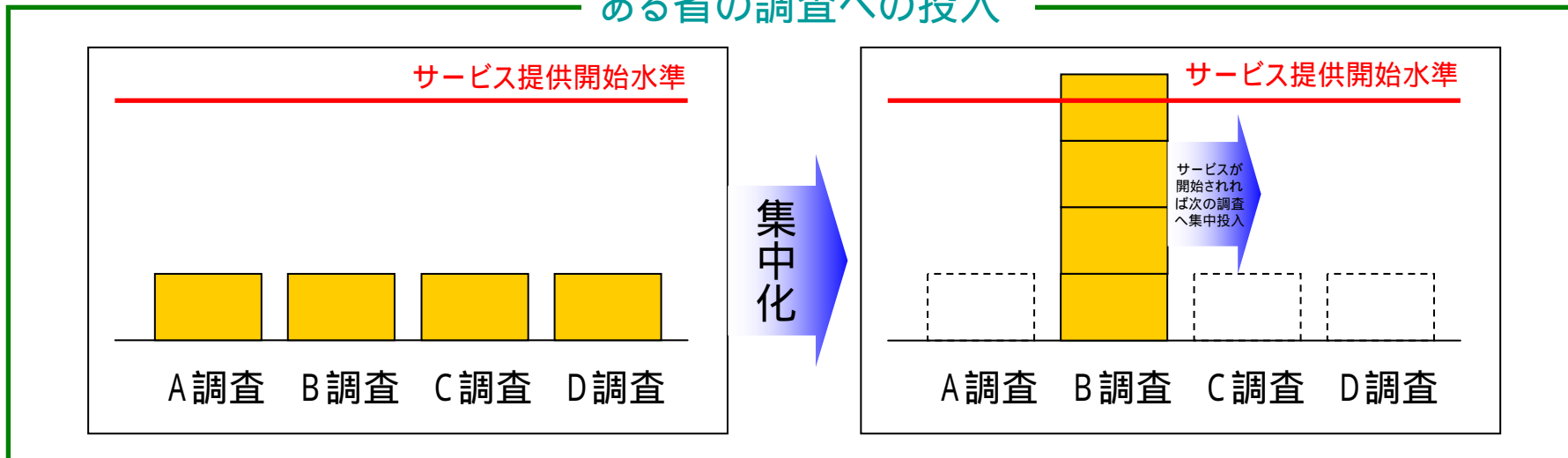
二次利用に対応できる能力を有する人材と、限りある予算をまんべんなく全ての統計調査、すべてのサービスに割り振れば、制度立ち上げ期におけるサービス提供開始に、さらに多くの時間がかかる懸念

利用可能な人材と予算を見極め、特定調査と特定サービスにこれらを集中的に投入することで、特定調査における技術の検討等の一定の蓄積を早く完了することが可能

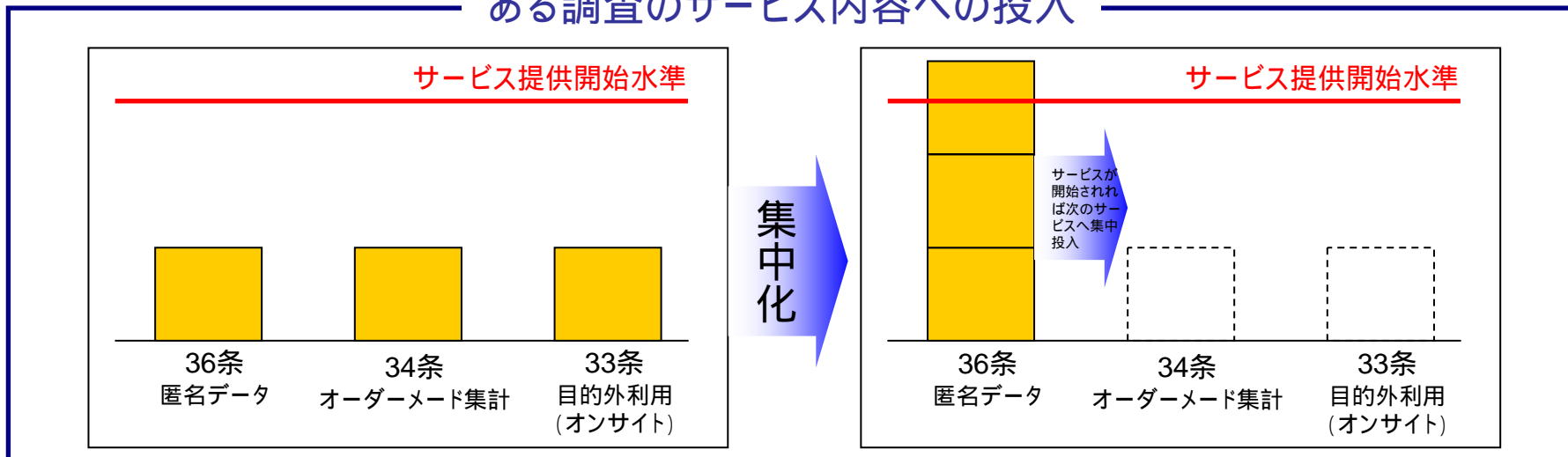
蓄積が完了しサービス提供が開始されれば、次に優先順位の高い調査、サービスに集中投入(このローリングによりサービスメニューを順次拡大)

イメージ

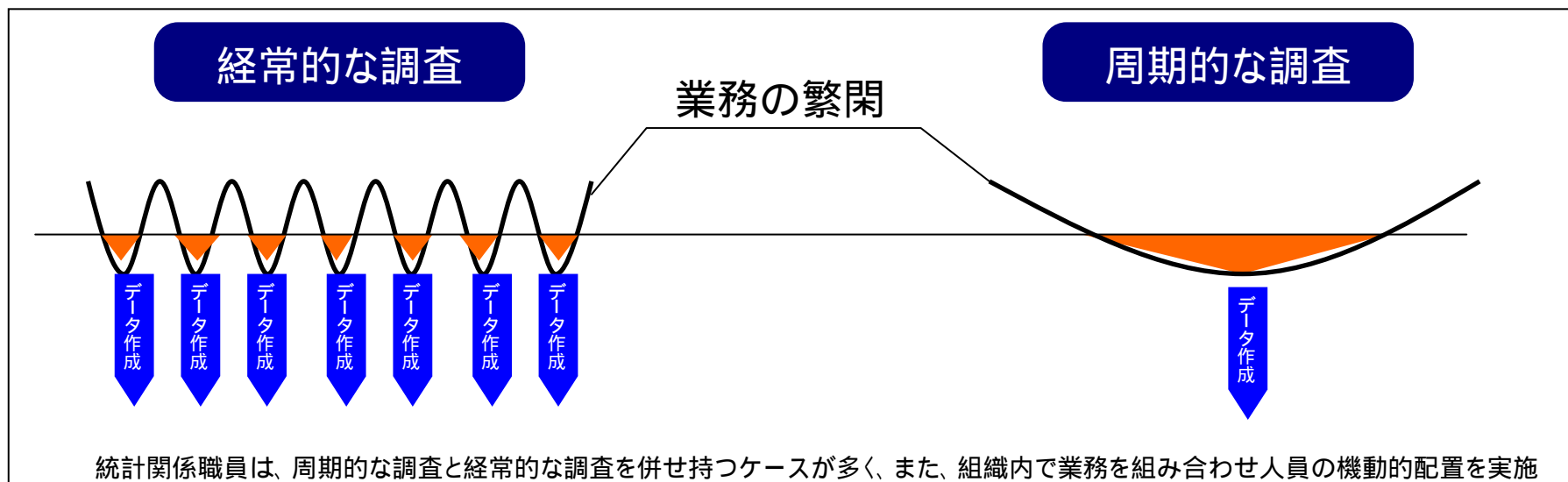
ある省の調査への投入



ある調査のサービス内容への投入



調査周期における特徴



余力のない中で、繁閑をぬって丹念なチェックなどにより事故リスクを小さくし、国民の信頼を確保できるのは、周期的な調査

仮に秘匿度に落ち度があり個人特定事故が発生した場合、行政機関の責任が問われ、ひいては統計調査全体に影響を与えかねないことから、事故リスクの低減は必須。

制度立ち上げ期における当面の方針

制度立ち上げ期の当面の取組方針として、各府省は、例えば、以下のような考え方を検討し得るのではないか

基幹統計をターゲットとした取組を推進

世帯対象調査は匿名データの提供、事業所・企業対象調査はオーダーメイド集計(オンサイト利用)をターゲットとした取組を推進

匿名データの取組は、周期調査から着手

調査に対する
集中投入

サービスメニューに
対する集中投入

匿名データの
リスク軽減

当面の取組のコンセプト

立ち上げ期に利用可能なリソースを集中配分し「小さく生んで大きく育てる」方針が現実的。(このような取組によりサービス開始を行うことで実績による需要把握と予測が可能。)

各府省が主導するサービス展開

制度立ち上げ期には、リソースの制限があることから、サービスメニューが限られる。したがって、受動的な対応ではなく、選択と集中等について各府省が自らの意思で主導するサービス展開が必要

例えば、基幹統計については、当該年度の初めに各府省が提供可能なサービスを判断し、サービス内容について事前に提示することなどが有効

期待される効果

行政機関が主導してサービス内容を決めることができることから、例えば受付期間を限定することで以下のような効果

公表時期など既存業務の繁閑にうまく新規サービスを組み合わせることができ生産性が向上する

民間委託する場合、調達手続きをまとめて行うことができるなど効率的

統計ユーザーがメニューを見て効率的に申請可能（行政機関に照会する度合いが減少）

統計ユーザーが新規サービスの要望を整理しやすい

基本計画との関係

各府省が横並びで主導的なサービス展開を図りやすいように、基本計画において、次のような取組を記載してはどうか

各府省が基幹統計に対し事前にサービスメニューを提示すること
利用実績の蓄積、申請を断ったケースの情報蓄積
サービスメニューと2次利用実績について取りまとめを行い、法
55条に基づき法の実施状況として統計委員会へ報告(委員会は意見陳述)
各府省が主導的にサービスメニューを決める際の参考となるサー
ビスメニュー(負担の少ないサービスメニュー)の例示

【 共 通 】

受付期間、サービス提供期間

【オーダーメイド集計】

再コーディングを伴わない表頭・表側形式の数次元クロス集計サービス、小地域集計
既存統計ソフトを利用した基本統計量計算(平均、各種%点、各種検定統計量、各種推定統計量)

【匿名データ】

統計局が試行的に行った取組をモデルとした匿名データ作成
各調査事項を大くりにまとめる大域的再符号化(個別データ全体にわたってカテゴリーをまとめて荒くする)

サービス提供の理想系に向けて (充実方策)

2 次的利用の特徴とリソースの確保

手数料収入

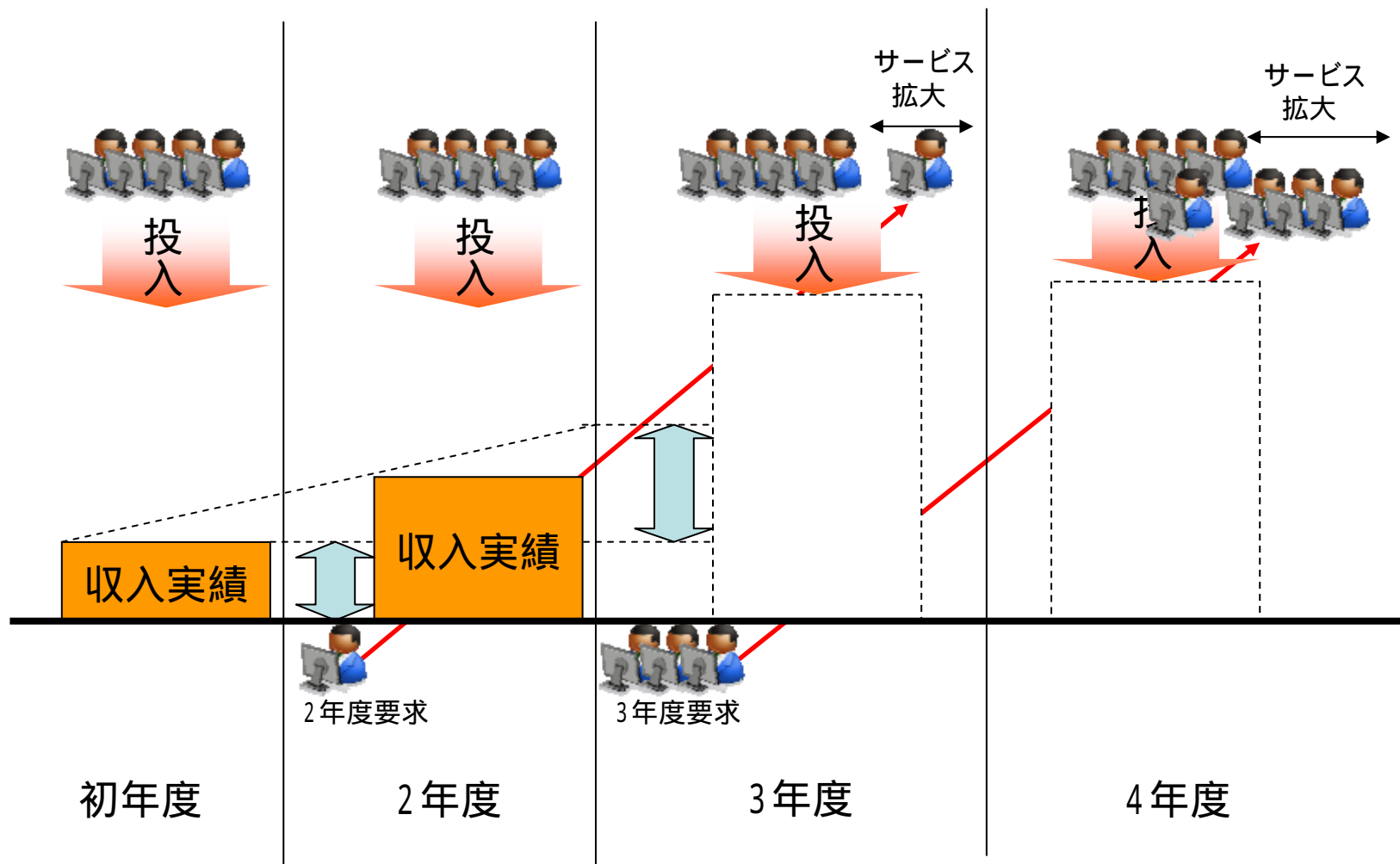
二次利用は手数料収入が伴う

集中化等により、小さくともなんとか事業を立ち上げ、数年間の利用実績を出来るだけ多く積み、収入見通しを立てることが将来のリソース確保に重要

利用実績

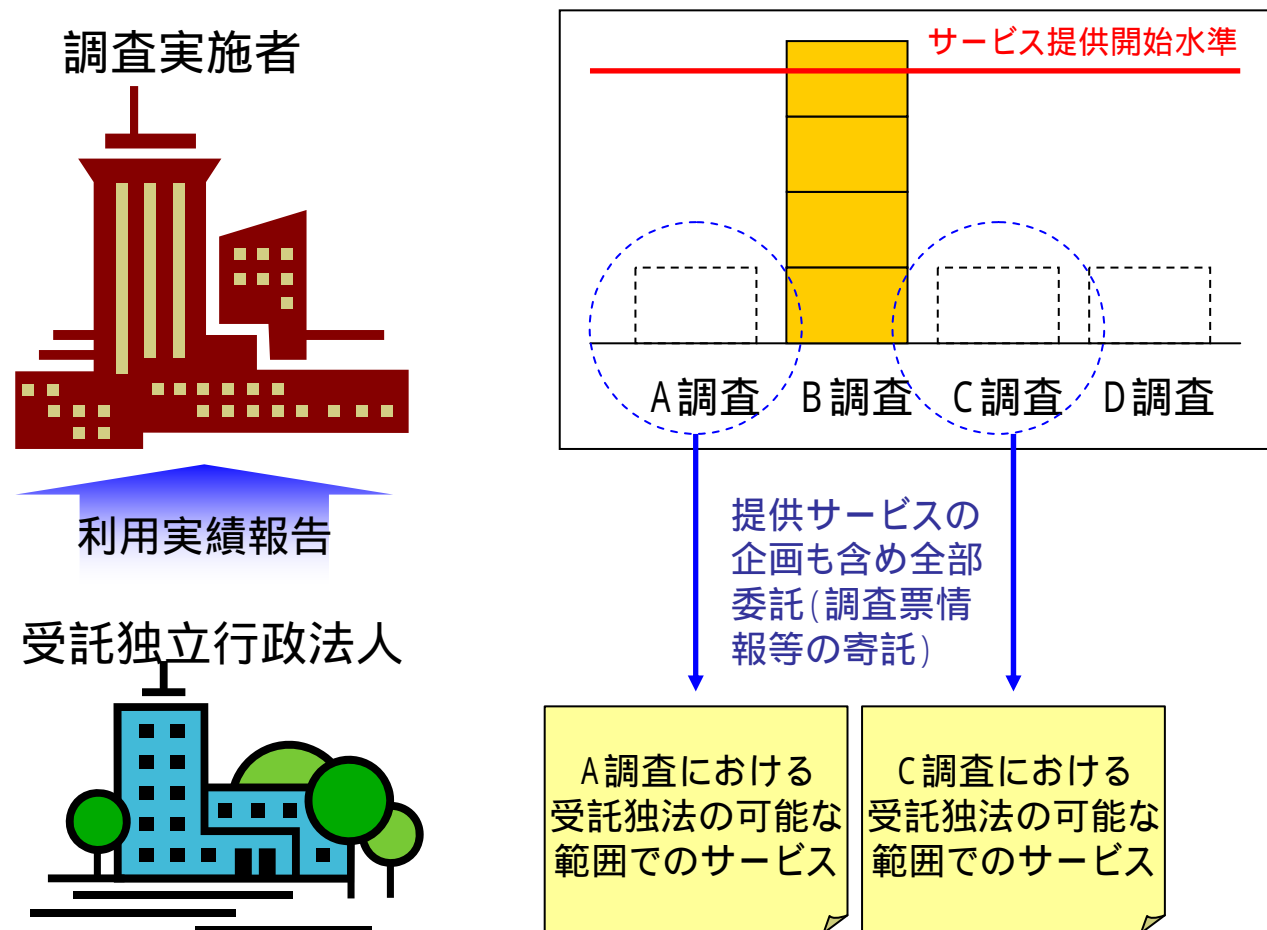
リソースを確保するためには、実績に基づく説得力ある将来性を示すことが必要

リソース確保のイメージ



利用実績の拡大のひとつの方策

受託独立行政法人制度を用いることで、サービス開始に至らない調査についても利用実績を蓄積可能

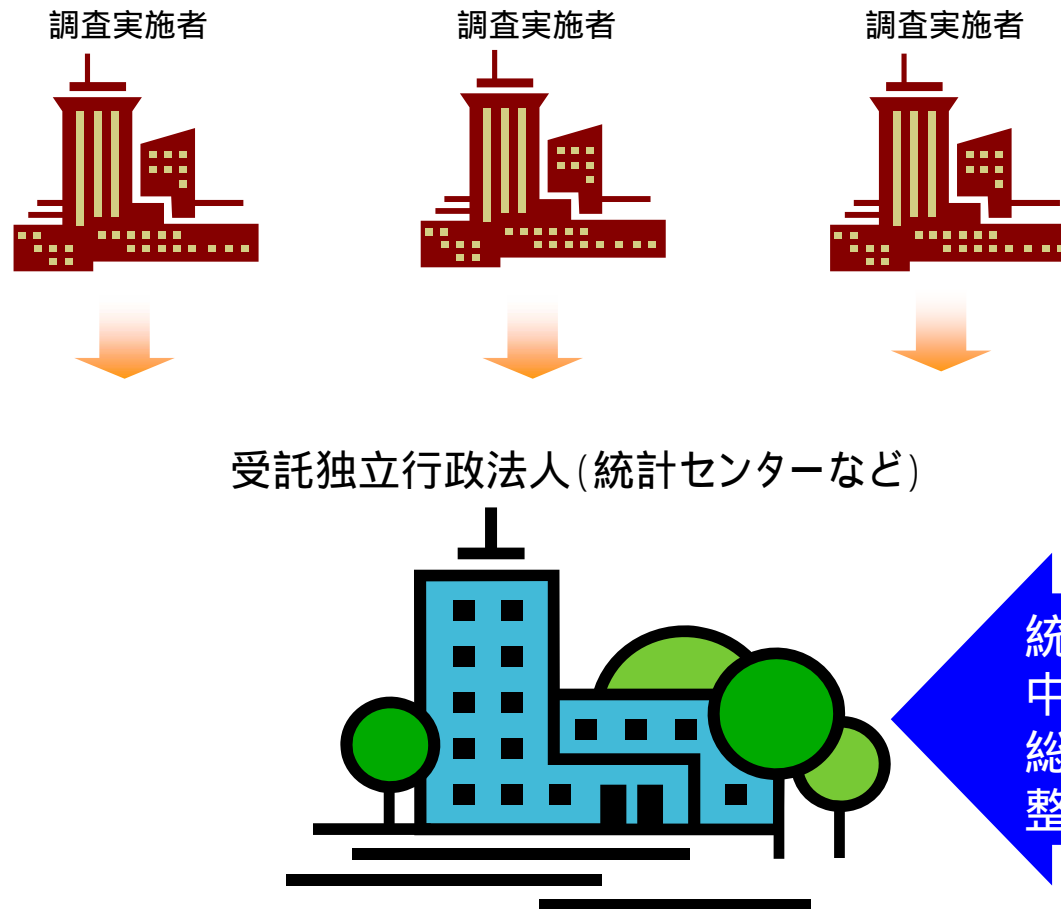


(事務の委託)

第三十七条 行政機関の長又は届出独立行政法人等は、その行った統計調査に係る調査票情報に 関し第三十四条又は前条の規定に基づき行う事務の全部を委託するときは、その業務の内容その他の事情を勘案して政令で定める独立行政法人等に委託しなければならない

実効性の確保

調査実施者の判断に任せるものの、その判断を行うためには、全部委託事務の受け皿となる法人が1つ以上存在することが必要



二次利用システムの課題 整理と当面の対応について

2次利用に関するシステムにおける制限の整理と対応

統計の集計は、各府省の調査のノウハウ及びそれを支える業務・システムと密接不可分であり、現行の分散型統計機構の元では集計システムの共有化は困難

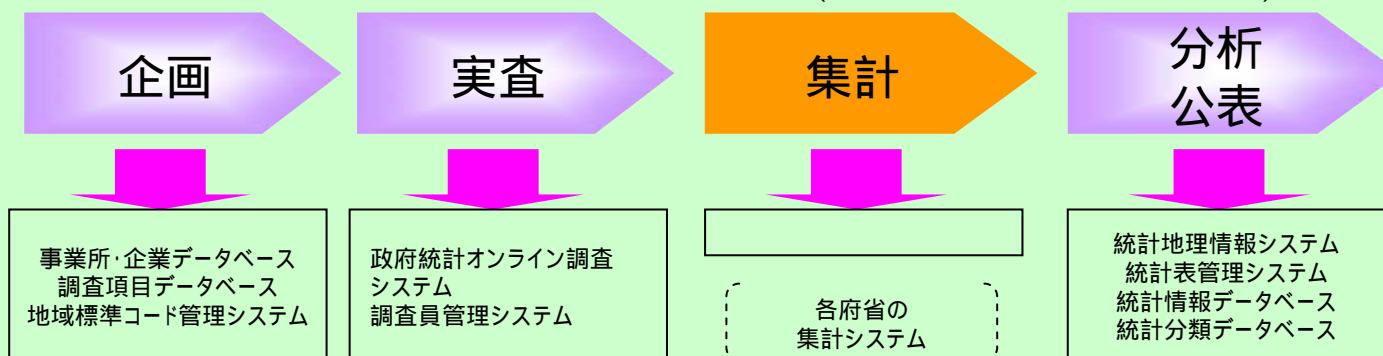
集計システムの共有化が困難な状況において、集計関連業務である二次利用のシステム共有化を進めることは困難であり、さらに各調査の特性、ニーズに応じた二次利用サービスの在り方や機動的な提供をかえって制限する恐れ

二次利用ユーザーに提供される匿名データ等のファイル形式が特定のパッケージソフトに依存しないようにする取組を、積極的に推進

二次利用に関する需要実績が一定程度明らかになる時期を目途にその時の各府省における二次利用の体制を踏まえ、オーダーメイド集計や匿名データの提供のシステムの共通化について改めて検討するものと整理することとしてはどうか

参考

統計調査の流れ図と政府統計共同利用システム(統計調査等業務最適化)



学界との連携

(学界からの支援)

学界から政府への協力

例えば、一橋大学においては研究費を取得し、システムの整備等を図り、自ら匿名データを作成するなどの取組を行った

制度立ち上げ期における行政側のリソースの限界を想定し、最大のユーザーと想定される学会も行政の補完（例えば、匿名データの作成方法の提案、統計から生成される学習用データの作成）などの協力を行ってもらえないか

作成された統計表に対し近い値の統計表を有するデータセットを一定のアルゴリズムで生成。生成データは作成された統計表に依存するが統計表を作成した調査票データ又は匿名データには依存しない