

第 23 回統計データの二次的利用促進に関する研究会 議事概要

- 1 日 時：平成 28 年 12 月 8 日（木）10:00～12:00
- 2 場 所：総務省第二庁舎 6階特別会議室
- 3 出席者：廣松座長、縣委員、安田委員
新井政策統括官、吉牟田統計企画管理官、中村調査官
《オブザーバー》
総務省（統計局、統計委員会担当室）、財務省、文部科学省、厚生労働省、農
林水産省、経済産業省、国土交通省、独立行政法人統計センター
《審議協力者》
（独立行政法人統計センター）椿理事長
（統計研修所）小林教授
（中央大学）伊藤准教授
《事務局》
総務省政策統括官（統計基準担当）付統計企画管理官室（佐藤専門官、越補佐）
- 4 議 題：（1）オンサイト利用の試行運用に当たっての基本的考え方
（2）第 22 回研究会における意見とその対応について
（3）分析結果等のチェック基準について
（4）オンサイト施設基準について
（5）その他
- 5 議事の概要及び意見等
（1）議題 1 オンサイト利用の試行運用に当たっての基本的考え方
事務局から、資料 1「オンサイト利用の試行運用に当たっての基本的な考え方（案）」
の説明が行われた。説明後、意見交換及び質疑応答が行われた。

（意見交換及び質疑応答の概要）
○ 利用者 PC 3 台というのはあくまで初期設定で、ニーズが上がれば増設を検討するの
か。また、ソフトウェアの持ち込みは、システムのセキュリティの観点で考えると、利
用者にデータ操作以外のこと（インストール作業等）を行わせて良いのか。（縣委員）
→ PC 3 台でスタートし、試行運用の中で支障があれば考える。仮想 PC の準備のため
にも、スタートとなる台数を決めたもの。ソフトウェアの持ち込みについては、利用者
から提供されたソフトウェアのインストール作業は統計センターが行う。（統計センタ
ー）
→ 利用者 PC 3 台は原則であり、増やしたい大学等は統計センターと協議することがで
きる。（事務局）
○ 資料 1 は、本日の研究会で合意が得られれば、案を取り日付が入るのか。なお、参加

拠点の開設予定時期には、平成 29 年と入れた方が良い。(廣松座長)

→ そのようにさせていただきたい。(中村調査官)

○ 監視カメラだけではなくキーログを残すことになっていたと思うが、それも書き入れた方が良い。また、保存期間 3 年は研究者のふるまいからすると短過ぎる。何か起こってからを想定すると、もう少し長期にするか、条件をつけて永久保存にするイメージを研究者に与えた方が良い。関連して、どういう時に開示するのかを研究者に徹底させるため、条件を書き加え、申請書にも記入した方が良い。Excel2013 の 2013 は外した方が良い。(安田委員)

→ キーログは検討させていただく。保存期間は、大学の設備の関係もあるので、試行運用の中で大学と協議しながら検討させていただく。Excel の件は修正する。(統計センター)

→ ログはバックアップ媒体でも良いので検討してほしい。(安田委員)

→ 一橋大学で従前のオンサイトの認証を受けた時は、キー入力記録をとることとし、そのためのソフトウェアを入れた。同じようにすれば良い。(小林教授)

○ ソフトウェアの持ち込みの関係で、ワープロソフトや Microsoft Office は大学側の持ち込みで仮想 PC に入れる想定か。(小林教授)

→ 資料 1 別紙で紹介しているのは統計解析ソフト。Microsoft Office や論文を書くソフト等は導入済み。(統計センター)

○ データの利用期間について、科研費の要件で利用する場合、科研費の期間に応じて利用期間を設定するのが通常だと思うが、今回の場合利用期間は原則 1 年以内となっている。弾力的に運用することを想定しているのか。(伊藤准教授)

→ 運用側からすると、システム容量の限界があるので、多数の利用者に長期間占有されると新たな利用が見込めなくなる。とりあえず 1 年単位で切りながら無駄のない運用ができるよう考えている。ただ、1 年で終わりではなく、1 年単位で延長申請してもらって運用を考えている。(統計センター)

→ 制度上は、「研究等の期間に照らして適切な期間（できるだけ短期間であることが望ましい）」となっており、システムの都合も含め、とりあえず 1 年と置いているもの。(事務局)

→ 今回の試行運用期間で走りながら考える中で、そうした条件も検討されることになるので、まずは手始めのルールと考えている。(統計局)

→ 科研費が申請条件になっているが、無条件に認めるということか。(安田委員)

→ 統計法第 33 条の要件について、オンサイトだから緩めるということは考えていない。(中村調査官)

→ 大学関係の利用施設では年度単位の申請なので、年度と書いた方が良い。(安田委員)

○ 長い間検討してきた課題なので、まずは動き出すことが重要。出された意見については、動きながら問題点を洗い出して本格運用に繋げてほしい。また、提供する調査票情報の種類について、準備が終わったものから順次提供するとのことであるが、統計局には是非準備を積極的に進めていただきたい。世帯系、事業所系といった種類やデータ容量もいろいろあるので、オンサイトの提供がうまく回るかの検証のためにも、なるべく

多くの統計を試行運用中に提供し、なるべく多くの方に利用されるようにしてほしい。

(廣松座長)

- 提供する統計調査について、検討の過程で個別の調査ごとに課題が出てきているが、試行しないと分からないところもあるので、できるだけ広く出せるように調整を進めているところ。(統計局)
- 試行の結果を各府省も注目していると思う。この研究会でも統計局以外のデータの提供に関する意見もあった。各府省が安心して協力できるような状況を作り上げてほしい。(廣松座長)
- 当面提供する側の状況が前提となっているが、オンデマンドで準備してもらえるのか。また、最終的にどこまでという原則はなく、状況をみてということか。(縣委員)
- 準備ができてサーバにアップした調査票情報について利用申請ができる形となる。オンデマンドでできるかということだが、準備していないものについて、調査票情報の提供を個々に対応することは困難。また、調査票情報の提供は、できるだけ広く提供したいと考えているが、具体的な限界がどこにあるかは何とも言えない。(統計局)
- どこまでオンサイトで提供するかについて、限界は最初に合意しておいた方が良いのではないか。(縣委員)
- どういった形で提供した場合にどういった不都合が生じるかを見ながらの判断になると考える。(統計局)
- 最終的にどこまで広げるかは中長期的な課題だと思うので、事務局でもそれを意識しつつ試行運用を評価してほしい。(廣松座長)

(2) 議題2 第22回研究会における意見とその対応について

事務局から、資料2「第22回統計データの二次的利用促進に関する研究会(5月31日開催)における主な意見と対応方針」の説明が行われた。説明後、意見交換及び質疑応答が行われた。

(意見交換及び質疑応答の概要)

- 「持ち出し」という言葉と「分析結果等のチェック」という言葉は主体が違う。これをパラレルに入れ替えて文意が通るのか。また、資料5はまだ「持ち出し申出書」となっている。作成したデータが別のルートで手元に来るのを「持ち出し」というのはおかしい。(縣委員)
- 「分析結果等のチェック」という言葉はタイトルの使おうと考えており、文脈の中で使う場合はうまく文意が通るような文章にしたいと考えている。資料5の「持ち出し申出書」は統計センターとも相談して考えさせてほしい。(事務局)
- 例えば主体を考えると、分析結果等の「提供」となる。(安田委員)
- 現時点で良い文言が見つかっていないので、もう少し検討させてほしい。(中村調査官)
- 氏名公表に関しては、廣松座長がアメリカで情報収集された際に、もっと厳しい対応をされている事例があったと記憶しているが、日本では難しいということか。(縣委員)

- アメリカの厚生省に属しているデータセンター（National Center of Health Statistics）では、オフィスにブラックリストが貼られていた。しかし、外部に出していたかまでは明確ではない。（廣松座長）
- それは一つのオプション。社会に公表しなくても、その場所だけで特定しておくことにも意味はあるし、利用者側の不利益も小さくなる。（縣委員）
- 例えば懲戒処分を受けた者を社内で公表して再発を防ぐような例もあるようで、閉じられたコミュニティ内で公表するやり方であり得る。今回は公表する場合の懸念点を挙げたもので、それよりも抑制的なペナルティがあり得るのではないかとということで説明したところ。ご意見も含めて検討する。（事務局）
- イギリス国家統計局でヒアリングした時に、担当者は法律上の罰則があることを強調していた。オンサイト施設やリモートアクセスを持っている大学では、罰則規定ではなく科研費のような研究費の使用停止という話をしていた。（伊藤准教授）
- 科研費を参考にするのは良いと思う。科研費の場合、研究不正が明らかになると研究が中止されることが公知される。また、不正を行った研究者だけでなく、研究班員全体が申請を数年間禁止される。そういう意味では、研究者はそういった情報が開示されることは当然と考えているのではないか。（樫理事長）
- アメリカやイギリスでは利用の入口が我が国と違っていて、研究ベースではなく、宣誓を行って身分を取得することで利用できるようになる。入口段階から厳しくなっているとところもあると思う。（統計局）
- 科研費は氏名が公表されていないわけではなく、何かの法律に触れているということで公になることもある。（安田委員）
- 刑事罰を受けた者は氏名公表されることになるのは当然の流れだと思っている。（事務局）
- ガイドライン上ではペナルティをかける時は各省に情報共有されるのではないか。（小林教授）
- 基本的には1～12か月の提供禁止措置（オーダーメイド集計、匿名データ含む）を定めている。各省に情報共有するというシステムは組まれていない。（事務局）
- ガイドラインの注書きでは全府省における一斉停止となっている。（安田委員）
- 2つの問題が残っている。一つは不適切利用に関してどういう処置をとるか。これは法制的な面もからむので、事務局や統計局・統計センターで十分検討していただきたい。また二つ目として、「持ち出し」と「分析結果等のチェック」の言葉遣いに関して、検討していただきたい。（廣松座長）

（3）議題3 分析結果等のチェック基準について

統計センターから、資料3「分析結果等のチェックの流れ（案）」、資料4「オンサイト利用における分析結果等に関するチェック基準（試行運用版）（案）」、資料5「分析結果等の持ち出し申出書兼審査報告書」及び資料6「課題」の説明が行われた。説明後、意見交換及び質疑応答が行われた。

(意見交換及び質疑応答の概要)

- 前回の研究会で紹介した「Five Safes Model」のうち、「safe output」に相当する部分の考え方が今回まとめられ、また、安全な統計という概念を EUROSTAT でガイドラインとして明記されているものが、資料3となっているものと認識。(伊藤准教授)
- 資料6の各課題について、
 - ・課題1は、調査データごとの特徴を考慮した基準ということで、具体的には閾値の問題になると思う。世帯・人口系と事業所・企業系で同じ閾値を適用する必要があるのかについて、今後具体的に分析をしていく中で課題として出るのではないかな。
 - ・課題2については、集中度ルールを適用するかはデータの特性としてスキューな分布になっているかが一つの論点。基本的には、事業所・企業系のデータに対して、集計表を作った時にスキューな分布になっていることで特定される可能性が高い度数に対して処理をすることが考えられるので、世帯・人口系のデータにも形式的に集中度ルールを適用するのが良いのかどうか、運用の段階で検討してほしい。
 - ・課題3の τ -ARGUS については、公表されている集計表の秘匿のためにオランダで使われているソフトウェアだが、オンラインで利用者が使うに当たり、例えば利用者が三次元よりも細かなクロス表を作った時に τ -ARGUS がうまく起動するか、また、利用者が τ -ARGUS をうまく使いこなせるかどうか論点となる。問題になるのは、利用者が個票データに準じたレベルの高次元の集計表を持ち出すこと。リモートアクセスの仕組みを持つデンマーク統計局で二次利用の担当者にヒアリングしたところ、彼らは持ち出すデータの容量を見ているとのこと。容量が大きい場合、個票データあるいはそれに準じた高次元の集計表である可能性が高い。また、全ての集計表をチェックしているわけではなく、ランダムに選定した集計表を、マニュアルでチェックしているとのこと。チェックの在り方として統計センター側でチェックする一方で、利用者側に高次元の集計表を作ることがないように、集計表の次元やセルの閾値を考慮しながら作成することを事前に訓練することが重要だと考える。
 - ・課題4については、ウェイトをかけた集計表についてセルの閾値を議論するのは簡単ではないと思う。ウェイトをかけた集計表から個人が特定される可能性は低いと思われる。ウェイトをかけた集計表のみほしいという場合には、それほど厳格な基準を適用する必要はないと考える。
 - ・課題5は、チェック基準に照らして提供できないと思われるものとして、例えばグラフのようなビジュアル表現のものがあるが、個人が特定されにくいようなものも考えられる。一律にグラフが出せないというのは、利用者にとって大きな制約になり得るので、その辺りで専門家パネルの判断を求めるといことが今後出てくるのではないかな。(伊藤准教授)
- 集中度ルールの適用については、スキューな分布（左右対称でない右や左にゆがんだ分布）かどうかで見るべき。世帯系でも年収などはスキューと言えるので、世帯系かどうかではなく項目の分布を見てのチェックではないかと思う。(小林教授)
- 一橋大学で行っていたオンラインによる提供の際に、 τ -ARGUS を使った例はあるか。(廣松座長)

- ない。全部マニュアルでチェックした。(安田委員)
- 閾値をかなり高くしておいて、少なくともその条件を通ればフリーで通して良いというやり方をしないと、数が増えた時に負荷がかかり過ぎる。後は原課との調整でここまで閾値を下げて良いという議論を個別調査ごとに今後行えば良い。(安田委員)
- この点については高度な統計分析の知識が必要であり、簡単にはまとめられない課題であるとする。試行運用の開始時には実績のある ESSnet に準じた基準でスタートするのはやむを得ない。試行運用中に、伊藤先生や安田委員から指摘があった課題を的確に把握した上で更に検討を進めていく必要がある。その検討内容をこの研究会に適宜報告いただきたい。(廣松座長)

(4) 議題4 オンサイト施設基準について

統計センターから資料7「調査票情報のオンサイト利用に係る施設基準(試行運用版)(案)」の説明が行われた。説明後、意見交換及び質疑応答が行われた。

(意見交換及び質疑応答の概要)

- 資料7の2①「パーソナルコンピュータ又はシンククライアント」という文言は、逆にするか、「端末(シンククライアント又はパーソナルコンピュータ)」にした方が良い。また、その後の「利用者PC」も「利用者端末」とした方が良い。PCよりもシンククライアントの方が推奨されると思われる。(安田委員)
- 資料7の基準を満たすと、資料1別紙にある200万円で収まるのか。(廣松座長)
- オンサイト施設内の設備部分だけを計算したもので、PC2台程度、監視カメラ、録画機器、入退室管理システム等で200万円程度になる。部屋の工事費や什器類は含んでいない。(統計センター)
- 試行運用時の拠点大学等ではそれらの設備は準備できているのか。(廣松座長)
- 従前から二次利用の関係で施設等を作っていたところが多いので、既存の施設を有効利用する形で進めてもらっている。(統計センター)

(5) 議題5 その他

事務局及び総務省統計局から参考資料1及び2の説明が行われた。

(以上)

《文責：統計企画管理官付高度利用担当》