

【調 査】

北米における政府統計個票公開の現状に関する調査報告

——米国労働統計局、米国センサス局およびカナダ統計局のオンサイトリサーチを中心に¹⁾——

神 林 龍

欧米各国は政府統計個票の公開に際し、プライバシーの確保と利便性の追求とのトレード・オフに直面した結果、オンサイトリサーチと呼ばれる個票利用方法を編み出した。この方法は、特定の場所に設置された特定端末から直接各統計部局のデータ・サーバーにアクセスする。利用可能な場所が極度に制限され、物理的にデータの持ち運びが不可能なもの、固有名詞が落とされただけの、ほぼ調査個票原票に近い情報を利用することができる点有用とされる。本論文では、アメリカ合衆国労働統計局、同センサス局、カナダ統計局において設置・運用されているオンサイトリサーチの状況を報告する。特に法的な観点および利用に至るまでのプロセスに注目し、彼ら地でのオンサイトリサーチの運用実態を明らかにする。

1. はじめに

政府が所管する各種統計の個票データを公開する方法が議論されて久しい。議論の具体的な内容や背景については、2007年5月に成立した改正統計法などに関連した報告書がすでに様々に公開されているので、それらに譲る²⁾。本調査では、北米における政府統計の個票データ公開の現状を概観・報告することで、わが国における政府統計の個票データを利用する仕組みづくりの一助としたい。

具体的には、アメリカ合衆国労働統計局(Bureau of Labor Statistics, 以下BLSと略す)およびセンサス局(United States Bureau of the Census, 通称Census Bureau, 以下BOCと略す)、カナダ統計局(Statistics Canada, 以下StatsCANと略す)において設置・運営されているOnsite Researchの方法をまとめる。

2. 米国における政府統計個票の公開方法

よく知られているように、米国における政府統計は典型的な分散型システムに服しており、大小70以上の部局がそれぞれに政府統計を所管している。政府統計を統括的に扱う部局は連邦政府に存在せず、対をなして、日本の統計法にあたる政府統計の取り扱いを一律に定めた法律も存在しない。政府統計の取り扱いの法的根拠は、各統計の根拠法に定められる場合があるほかは、The Privacy ActやThe Trade Secret Actなど政府が扱う情報管理一般に関する法律に求められる。ただし、後に詳述するよ

うに、BOCだけは例外的に連邦法13条に管理規定が定められており、法的には他部局のおかれている状況と区別される。それゆえ、この調査報告ではBOC以外に、もう一部局、具体的にはBLSをとりあげ、なるべく米国の状況が一般的に整理できるように配慮した。

合衆国における行政府・立法府の資料公開には、建国以来の長い伝統がある。合衆国において重要政策が決定される際には、議会や行政府に調査会が設置され、調査検討された結果の報告を受け、知事・大統領が決定命令することが多かった。その際に調査会の議論や検討の内容は当初は立法府または行政府に、次いで公文書館(図書館)に保存され、遅くとも数十年後には公開される。もちろん、プライバシーを保護する必要があるため、すべての資料が完全に公開されるわけではないが、政策調査の際に用いられた統計資料も例外ではない。

読者の中には、政府が現行保管する公文書や情報に、一般市民がアクセスする権利を明示した情報公開法が連邦政府レベルで制定されるのは意外に遅く、1966年のFreedom of Information Actではないかと思う方がいるかもしれない。しかし州政府レベルでは、全米で最初の法律はおそらくウィスコンシン州のPublic-records Lawであり、実は正式州に昇格直後の1849年に制定されている。それ以前にも、イギリス起源の慣習法によって政府の保持する情報の公開が行われていたと考えられている(19世紀から20世紀前半にかけての合衆国慣習法における情報公開についてはCross(1953)を参照のこと)。統

計の個票データに関しては、大量のデータを取り扱うための計算機がない時代には公開要求はほとんどなく、議論の俎上にも上らなかった。かといって、公文書の性格をもつ個票データがすべて廃棄されたわけではない。たとえば国勢調査の個票は現在まで保管されており、これらの個票を遡及してデータベース化するプロジェクト(IPUM-USA)が進行している³⁾。

ただし、統計調査の個票データはそのまま公開されると被調査者のプライバシーを侵す可能性が非常に高い。それゆえに、公開に際しては細心の注意が必要とされ、その取り扱いについても法定されることが多い。合衆国は分散的な統計システムをとることから根拠法が統計によって異なり、したがって公開に関する規定もそれぞれ異なる。しかし、あえてまとめれば、政府統計個票の公開方法はおおむね次の3つのいずれかにあてはまる⁴⁾。

第一は、データ・アーカイブや民間会社を通じてリリースされる方法であり、Public-use Sampleと呼ばれている。利用資格制限はないが一般に有料で、利用可能な調査項目は限定されており、調査個票から何らかの基準で再抽出された標本が公開されることも多い。また、固有名が特定されやすい事業所データの個票は、原則としてこの方法では公開されていない。

第二は、使用条件や使用者個人を特定したライセンス契約を通じてリリースされる方法で、License ReleaseやOffsite Researchと呼ばれている。通常、利用資格や使用場所などに制限が伴うが、データ自体が物理的に持ち運ばれることとPublic-use Sampleよりもより秘匿必要性の高い調査項目を利用できることに特徴がある。

第三がOnsite Researchで、特定の場所に設置された特定端末から直接各統計部局のデータ・サーバーにアクセスする方法である。利用可能な場所が極度に制限され、物理的にデータの持ち運びが不可能なもの、固有名が落とされただけの、ほぼ調査個票原票に近い情報を利用することができる。この方法では、社会保障番号など個人情報を使用して複数の調査個票をリンクさせて利用することも可能である。ただし、リンク作業は多くの場合当該統計部局係員が行い、利用者が直接行うわけではない。以下、本稿では、上記3つの方法に従ってどのように調査個票が利用されているかを、具体的にはBLSおよびBOCの運用を取り上げて概観する。

2.1 BLSにおける個票公開

2.1.1 概況

BLSとは、米国の雇用・物価などの経済統計を司る連邦機関のひとつである。形式的には労働省(Department of Labor)の管轄下におかれているが、その責務の重要性に鑑み、たとえば政治任用はトップのコミッショナーのみに制限されるなど、人事や業務の独立性が図られている。日本の政府統計担当部局が組織上も人事上も各省庁に完全に従属しているのと比較すると興味深い。BLSの管轄する公表統計は一定の信頼を勝ち取っており、それは雇用・賃金指数や物価指数など、BLSの作成する月次経済指標が株式市場において景況判断の材料とされていることにもあらわれている。近年では、1998年に成立したWorkforce Investment Actによって積極的労働市場政策が展開されると、その進行状況などをモニターするために職業訓練や失業給付などを中心に統計が拡充された。

2008年1月現在、BLSでは、さきに紹介した3つの方法がすべて用いられて個票データが公開されている。どの統計がどの方法でリリースされるかは、Disclosure Review Boardにおいて調査項目ごとに決定される。したがって、たとえばNational Longitudinal Surveysのように、同一の調査の中で、Public-useに供される部分、Offsite Researchで使用が認められる部分、Onsite Researchでしか使用が認められない部分、まったく使用が認められない部分が混在することもしばしばある。調査の際には、どの方法でどの項目が公開されるかが事前に被調査者に説明される。事後的に公開方法を変更するのは実質的に難しいが、規則上は、Disclosure Review Boardの決定に基づいた変更が認められている。BLSにおいては、このDisclosure Review Boardは内部組織であり、外部識者は関与しない(Disclosure Review BoardについてはZarate, Hoy, Stamas, and Therriault(2000)を参照のこと)。ただし、外部識者が関与するTechnical Review Boardが別途設置されており、外部者が統計調査の公開方法に関して意見を述べる機会は確保されている。

2.1.2 Public-use

BLSの管轄するデータは事業所データ(Business Data)が中心であり、これらのデータは個票を用いて再集計すると容易に事業所を特定可能な集計表を作成することができる。このような集計表を公表する行為は、政府統計が一般的に服すべきPrivacy

Act, The Trade Secret Act)によって禁じられており、したがって、現在 Public-use の方法で個票が公開されている BLS 管轄の統計は世帯調査のみで、種類は多くない。しかし、世帯調査のなかでも重要なものが公開されており、わが国の家計調査にあたる Consumer Expenditure Surveys, 労働力調査にあたる Current Population Survey(BOC と共同), パネルデータである National Longitudinal Surveys(BOC と共同)がある。これらの統計調査の Public-use Sample は、直接 BLS より電子媒体で購入できるほか、ICPSR など各地のデータ・アーカイブにも寄託されており、合衆国国内のみならず外国においても教育や研究に頻繁に利用されている。

ただし、近年では、外部データとのリンクが無制限に可能であることが Public-use Sample の問題点として指摘されるようになった。計算機器の発達した現在では、他に手に入れることのできるデータと Public-use Sample としてリリースされたデータを、何らかの方法でリンクさせることで、個人を特定する可能性が無視できないほどになる場合がある。Public-use Sample を公開する統計部局には、潜在的にリンク可能な外部データがどれだけ存在するのか確かめる術はないので、当該データを Public-use Sample として公開するリスクを事前に推し量るのが困難な場合が生じてきている(この点に関する簡単なサーベイに Abowd and Lane(2004)がある)。

2.1.3 Offsite Research

先にも紹介したように、Offsite Research は、より厳格に秘匿する必要がある調査項目について、しかるべき手続きを経由した上で特定個人にデータをリリースする方法である⁵⁾。BLS が管轄する統計のなかでこの方法で利用可能なデータは、National Longitudinal Surveys Geocode Data および Census of Fatal Occupational Injuries に限られ、やはり事業所データは利用できない。このうち、National Longitudinal Surveys Geocode Data とは、パネルデータとして名高い NLSY79, NLSY79 Young Adult, NLSY97 のうち地理的項目を中心とした次の調査項目を指す。すなわち、

- 各調査年における回答者の居住地情報(州、郡、調査区)および当該地域の経済変数
- 誕生した時点の居住地情報(州または国)
- 14 歳時点の居住地
- 離婚後養育に関する合意が成立した州
- 誕生・結婚・離婚・入学などの月日(Public-

use Sample では年月まで)

- 入学大学の名称と所在地
- 取得博士号の専攻名
- 両親の居住地からの距離(Public-use Sample では階級値)

である。特に居住地情報については、Public-use Sample では利用できず、Offsite Research で初めて入手可能になるところに特徴がある。ただし、NLSY の Original Cohort についてのこれらの調査項目は、BOC での Onsite Research のみで対応されている。

BLS で Offsite Research を利用するためには米国民である必要はないが、米国在住で所属機関が米国にあることが必要である。利用者と BLS との間では BLS Agency Agreement が締結され⁶⁾、同時に利用者の所属機関と BLS との合意契約がなされる。この際、所属機関には、機関長(大学であれば副学長以上が必要で、学部長では不足する)の承認を得るほか、Project Coordinator と呼ばれる実質的責任者を指定した上、サーバーなど電子環境、ファイルや CD-ROM 保存のための物理的な準備を整えることが要請される。この所属機関と BLS との間の契約には、契約違反があった場合には所属機関が民事上の損害賠償義務を負うことも明記されている。したがって、守秘義務違反などがあった場合、利用者は連邦政府の Agent として刑事・民事訴追を受ける可能性が生じるのに加え、利用者の所属機関も民事訴追を受ける可能性が開かれていることになる。利用期間は学生 1 年、研究者 3 年(または所属機関に在籍中)で、利用場所は所属機関内の指定された場所と特定され、利用者の自宅等は認められない。利用者が所属機関を変更した場合には、利用契約はその時点で打ち切られる。利用終了時には、所属機関責任者によるデータ CD の返却とデータ消去の確認、BLS による成果物の検閲が求められる。

Offsite Research において BLS がもっとも留意しているのがデータ管理の実態である。したがって、利用期間中に BLS 職員による現場査察が行われる場合があり、過去に申請された NLSY の Offsite Research のうち半分以上は査察を受けている。査察時には、とくにデータ保存の物理的環境、サーバーやメインフレームのセキュリティ環境などがチェックされる。現在までのところ、Offsite Research における深刻な契約違反は、データを自宅に持ち帰ったという 1 件のみである。このときは利用契約が打ち切られ、当該利用者、当該所属機関における再

度の利用が禁止されたのみで、利用者に対する刑事・民事訴追、所属機関に対する損害賠償請求までには至っていない。査察で明らかになった違反のうち多くのものは、CD-ROMを保管しているキャビネットの鍵がうまくかからない、データを共有アカウントで扱うなどであったので、BLS職員がその場で改善を求め、所属機関が対応・報告することで取りまき、それ以上の重大な問題には発展していない。

Offsite Researchにおいては、データの利用可能範囲を、たとえば利用者の自宅などにも広げてほしいという要望が根強いが、厳格なデータ管理が求められることなどの理由からBLSは認めていない。

2.1.4 Onsite Research

2.1.4.1 概況

BLSにおけるOnsite Researchとは、BLSのワシントン事務所内の特定の端末を用いてデータサーバーにアクセスし、そこですべての作業を完結させる方法である。1997年9月より現在の方法で正式に運用が開始されて以来、総計120人の研究者が制度を利用している。BLSへの本ヒアリング調査が実施された2006年10月現在、Onsite Researchを利用中なのは36研究計画、50人であった。

Onsite Researchに用いられる端末は、BLSワシントン事務所内の3箇所、Office of Employment and Unemployment Statistics, Office of Prices and Living Conditions, Office of Compensation and Working Conditionsに6台ずつ設置されており、利用する統計の管轄によって使い分けられている⁷⁾。各端末は1室に集中的に配置されているが、パーティションにより6つのブースに仕切られており、部屋の外からすべてを見渡せるようには設計されていない。各ブースに備えられた端末には、SASがインストールされており、ローカルディスクを用いて計算することができる。またサーバーにはSASおよびSTATAがインストールされている。費用の関係上、Onsite Research用の専用サーバーは設置されておらず、各部署が通常業務で用いているサーバーのリソースが利用されているところにシステム上の特徴がある。

各ブースには端末のほか、棚や電源が設置されており、ラップトップコンピューターを含めた私物の持込が認められている。しかし、インターネットや端末への接続端子は設置されておらず、持ち込んだ電子機器を何らかの方法でインターネットに接続することも禁止されている。端末に自由に許される外部出力は画面のみに限られ、紙媒体・電子媒体での

出力にはその都度BLS担当者の検閲を通過する必要がある。この端末のおかれた部屋への出入りにはあらかじめ登録されたIDカードを用いた認証が必要で、これらを通じて、いつ誰がどこでのデータにアクセスし何をしたかを厳密に特定することができる。

2.1.4.2 Onsite Researchに至るまで

1997年にOnsite Researchの方法が確立する以前にも、外部研究者による政府統計個票データの利用は行われていた。ただし、この際には、公募制のフェローシップ・プログラムが用いられたほかは、BLS内のプロジェクトなどの特別な形をとった利用に限られていた。法的には、大学や州政府、あるいは適格と認められた組織の人員を最大2年間まで一時的に連邦職員として雇用することを認めた、Federal Intergovernmental Personnel Act(人事交流法)が根拠として利用され、外部の研究者に連邦職員と同様の法的義務を課し、刑事罰を適用可能にしたうえで、データにアクセスさせる方法がとられた。しかし、あくまで一時的に連邦職員として雇用されることが必要になるため、派遣元との雇用契約と矛盾がある場合(たとえば兼業禁止規定がある場合)や、人事交流法の適用外となる組織(たとえば外国の大学)に雇用されている場合などには利用できなかった。雇用主を持たない大学院生についても同様であるが、BLSでは大学院生を「ボランティア」として6ヶ月間受け入れるというプログラムを用意し、個票データへのアクセスの便を図っていた⁸⁾。このように、統計の個票利用に関するルールは存在したものの、職員任用(この場合はデータ利用許可と同義になる)に関する判断基準や権限などが不確かで、手続きも煩雑であるという限界ももっていた。

クリントン政権下の1993年にKatharine Abrahamがコミッショナーに就任すると、BLS Quality Councilを通じてMicrodata Access Review Teamを設立し、秘匿性を担保したうえでの個票データ利用の筋道を諮問した。同チームは“secured site”を利用した個票へのアクセス方法を確立することを答申し、具体的には専用のセンターを作ることを提案した(建議文書本文は非公開)。専用のセンターは予算の関係で実現をみなかったが、BLSワシントン事務所内に一角を設け、そこに“secured site”を設けるという現在のOnsite Researchの形が整えられた。

2.1.4.3 Onsite Research の仕組み

BLS の Onsite Research では様々な個票データが利用可能である。ただし、株式市場への影響を考慮し、発表までの秘匿が厳重に求められる統計が格納されているサーバーに存在する統計については、サーバーの設計上の理由で利用できない期間がある。

利用するには2つの条件を満たす必要があり、第一に「原則として米国民であること」、第二に「有資格機関に学生として在籍していること、または雇用されていること」が掲げられている。しかし、利用にあたっては Offsite Research と同様に利用者と BLS、利用者の所属機関と BLS との間に二重の契約を結ぶことから、第一の条件はそれほど厳格に解釈されておらず、第二の条件のほうが厳格に運用されている。有資格機関として、合衆国高等教育機関、NPO(内国歳入庁 (IRS) が定める基準 501c(1) または 501c(3) に適合する組織)、地方政府または地方政府の付属機関、地方政府を代表とする機関、連邦政府研究機関があげられている。したがって、外国籍の客員教授であっても、滞在研究機関が雇用主として Onsite Research に必要な義務を負う場合には、その限りで利用資格が認められる。問題の焦点は所属機関が十分な法的義務を果たすかにあるので、通常雇用契約を結ばない客員研究員などの資格で滞在する場合でも、滞在研究機関が同様の法的義務を負う限り、BLS は利用資格を認める方針である。

利用の際には、5~10 ページのプロポーザルの提出、審査、実際の計算、成果物の提出、というプロセスをたどる。BLS ではプロポーザルを提出する前に、ウェブサイト上で公開されているマニュアルをよく読み、目的の統計を管轄する部署に相談することを奨励している⁹⁾。プロポーザルには、もっぱら「統計的な利用」であることを明示する必要がある。「統計的な利用」とは、立法時には「他の行政目的で使用しない」ことが念頭に置かれていたが、現在では「個々の被調査者を特定しない」ことも含まれている。このほか、プロポーザルには研究内容について、統計技術的なメリットがあること、BLS の興味があること、BLS の政策を進めることを示す必要があるとされている。ただし、審査の際にもっとも重要視されるのは第一の「もっぱら統計的な利用であること」で、特に居住地の失業率など外部データを持ち込み、個票と接続して分析に用いる場合には個人を特定することを意図していると認識される可能性があり、注意が喚起されている。現在までのところ、BLS に興味がない、あるいは BLS の政策と矛盾するという理由で認められなかったプロポ

ーザルはない。また、研究内容に関しては、申請データが分析の目的に適合的であることを説得できるかどうかのポイントとなり、申請データでは分析の目的が明らかに達せられないという理由で却下されたプロポーザルが1件あったという。

そのほか、プロポーザルには成果物の公開方法、補助金、共同研究者の有無などを書き込むことが求められている。利用期間は最大で2年間で、さらに2年までなら延長が可能である。BLS 事務所への立ち入りが必要なことから、端末の使用は原則として月曜日から金曜日の8時から17時まで(特に申請すれば朝6時から可能)に限られている。

以上のようなプロポーザルの提出締め切りは年4回(1月、4月、7月、10月の各15日)で、ウェブサイト上の紹介では、審査に2ヶ月、審査後、法的契約などに2ヶ月要したあと、利用が開始されることが多いとしている。プロポーザルを受け付ける際には、利用開始まで3ヶ月から半年を見込むように注意を促しているという。

プロポーザルの審査は BLS-wide Microdata Access Review Board (MARB) によって担当される。この委員会には、BLS 内でデータの秘匿性について担当する職員のほか、被申請データを管轄するグループのシニア・エコノミストが出席し、審査する。純粋な外部研究者は参加しない。MARB は 2001 年 12 月から 2002 年 10 月までに 29 のプロポーザルを受けつけ(他の連邦行政機関からの依頼も含む)、うち 19 を承認、9 については改訂を要請し、却下したのは1つのみであった(Barkume and Spletzer(2003)p.7)。これ以降 2006 年 10 月までについても、却下されたプロポーザルは3つのみで、多いとはいえない。却下理由も、個票のデータを利用して追跡調査を行おうとしたものなど、匿名性保持に抵触すると考えられたものが大部分であった。ただし、新規に受け付けられるプロポーザルも減少傾向にあり、本ヒアリング調査直近の 2006 年 10 月 15 日締め切り分では3件のプロポーザルが届いたのみであった。

同一人物(グループ)によるプログラムの繰り返しについて明示的な制限はないが、最大4年間の利用期間が認められることから、現在まではほとんどみられなかった。ただし、2006 年 10 月 15 日締め切り分の3件のプロポーザルのうち、2件は過去の研究とほぼ同じ研究の延長に関わるものであった¹⁰⁾。

2.1.4.4 諸法規との関係

BLS がデータ公開時に服する諸法規の主なもの

は次の4つである。

- ◆ The Privacy Act
- ◆ The Workforce Investment Act
- ◆ The Trade Secret Act
- ◆ Confidential Information Protection and Statistical Efficiency Act(CIPSEA)

このうち、The Privacy Actは個人情報に関わる法律、Trade Secret Actは事業情報に関わる法律であり、統計目的以外の利用を禁じていること、連邦職員以外利用を禁じていることに大きな特徴がある。The Workforce Investment Actは、Katharine Abrahamが中心となり実行された雇用統計の拡充・刷新の根拠法で、この際に新設された統計について利用上の制限が明記されている。

合衆国の統計関連法規の特徴は、統計目的以外の利用を禁じると同時に、各統計について、調査客体本人も含めて個票が利用可能な範囲の主体を明示しているところにある。そして、上記3法には“officers and employees of the agency”などが統計情報に触れることができると明記されている。結局、合衆国における統計の個票利用の法的制限は、目的が統計目的に合致するかどうかと、利用者の身分が法的に是認されるかの2点が要だった。この2点を回避するために、人事交流法を用いて一時的に雇用関係を締結した上で個票が用いられてきたことはさきに触れたとおりである。

こうした中、2002年にConfidential Information Protection and Statistical Efficiency Act(通称CIPSEA)が制定され、違反時の罰則が重くなる一方(5年以下の懲役または25万ドル以下の罰金もしくはその両方)、雇用関係になくとも“agent”としての身分を保持できれば、個票データにアクセスすることが可能になった。こうして、大学院生や兼業禁止規定を有した研究者などが個票データにアクセスする道は広がったと考えられる。

ただし、BLSにおけるOnsite Researchの仕組みはすでにCIPSEA施行以前の1990年代後半に構築されており、利用者も着実に増加していた。現実には、大学院生を除けば人事交流法を適用できない潜在的な利用者はほとんどおらず、利用者には連邦職員の制約をかけるか否かという法改正は、人事交流法がある以上問題にはならなかったかもしれない。各研究者に対する私法上の規制は、研究者の所属する機関に明示的に責任を負わせることが要になっているからである。

2.1.4.5 Onsite Researchの資源

さきにも触れたように、当初Microdata Access Review Teamが提案した専用のセンターが設立できなかったのは、主として予算不足のためであった。現在でも、BLSはワシントン事務所以外の場所でOnsite Research施設を開設できるに至っておらず、後述のように全米9箇所にセンターをもつBOCと異なる。

また、BLSにおけるOnsite Researchは独自の予算が計上されておらず、人事局予算や各部局予算から費用を捻出している。2001年に実施されたBLS内の調査によると、1人の外部研究者に1週間対応するために、1300~1400ドルの機会費用が発生していたとされ、費用負担も少なくない。とくに、Onsite Researchへの外部研究者の応募は、1990年代はおおむね年間数件であったのが、1999年に12件、2000年に14件、2001年に20件、2002年に20件と急増したこと、研究施設がワシントンにしかないことから利用申請が休暇期間に集中すること、研究期間の長さがBLSの予想外に比較的長期にわたったことから、利用者の混雑現象が発生し、2003年にはOffice of Compensation and Working Conditionsは新規プロポーザルの受付を一時的に停止したことがある。

以上のように、BLS内でOnsite Researchに割くことのできる資源は限界がある。さらにワシントンに場所が限定されることは、旅費などの費用面で利用者に負担をかけるのみならず、利用可能期間に容易に混雑現象を起こす可能性がある。BLSではこの問題を解消するために利用料を徴収することのほか、BOCのResearch Data Centerに参加することも検討された。しかし、BOCは3.4.4に紹介する諸法規に加えて連邦法13条というより厳しい法律の規制下にあるため、同一の施設を用いた運用はできないとの結論に達した。

2.1.5 小括

合衆国における政府統計の個票利用は、秘匿性をどのように担保するか、法律上の利用制限をどのように解釈するかが大きな問題であり、第二の問題についてはCIPSEA制定によって一定の法的枠組みが整備されたと考えられている。しかし、実際には、BLSにおいてはCIPSEA以前においても人事交流法を用いるなど法的工夫が行われ、データ提供機関と利用者、データ提供期間と利用者の所属機関という二重契約をもとにしたOnsite Researchが利用されてきた。CIPSEA制定以降も、この二重契約をも

とにした利用者との法的関係は変わっておらず、この方法であれば秘匿性が担保できると考えられていることが示唆される。

一方、BLSのOnsite Researchには研究者や研究機関が運用に関わらない点にも特徴がある。データセットの利用可能性を検討するDisclosure Review Boardには外部研究者は参加しておらず、プロポーザルの事前審査・成果物の事後審査もすべてBLS職員によって担われている。そもそも、Onsite Research CenterはBLS事務所内部にしか設置されていない。BLSのOnsite Researchは、通常業務で使用する資源が流用されている側面が強い点は指摘しておきたい。

2.2 BOCにおける個票公開

2.2.1 概況

BOCとは、国勢調査などの主要統計を司るアメリカ合衆国連邦機関のひとつである。国勢調査は、建国以来、10年毎の調査のたびに臨時にセンサス局が設置され実施されてきた。20世紀に入ると国勢調査以外の統計調査の必要性も高まり、1902年にBOCとして恒久化され、当初内務省(Department of Interior)の管轄下におかれた。1903年に商務省(Department of Commerce and Labor)に移管された後、1913年に労働省(Department of Labor)が分離した際にも商務省にあり続け、現在に至っている(Anderson(1988), Ch. 5)。ただし、米国において国勢調査の実施は合衆国憲法によって規定されているため、それを司るBOCについても連邦法13条(Title 13 U. S. Code)により組織構成や職務内容、秘密保持義務などが明示的に規定されている。商務省の指揮命令を受けることはほとんどなく、憲法機関としての独立性を維持している。2008年1月現在、毎年100種類以上の調査を実施し、1万人を超える職員を雇用している。

現在のBOCは国勢調査以外にも多くの調査を実施しており、多種類の個票データを保存管理している。これらの統計はBLSなど他の連邦諸庁が行うものと異なり、連邦法13条に直接制約される法的特徴を持つ。したがって、その公開方法もBLSとは異なり、Public-useとOnsite Researchに限られ、利用者とライセンス契約を行い、BOC外へ個票データを持ち出すOffsite Researchは利用されていない。公開方法は調査項目ごとにThe Disclosure Review Board(通称、DRB)で決定される。

2.2.2 Public-use Sample

BOCは設立以来、個票の再利用あるいは他省庁の利用に関して保守的な態度をとってきた。有名な例では、太平洋戦争中の1942年に、軍需省より要請された日系人に関する個人情報の提供を拒否したことなどがあげられる(Bohme and Pemberton(1991))。しかし、戦後になって1964年に、1960年センサスの1000分の1の再抽出サンプル(標本サイズ約18万)から地域情報などを除去したものを公開したのを皮切りに、現在BOCでは、国勢調査や労働力調査(Current Population Survey, 通称CPS)について、誰でもアクセス可能なマイクロデータをPublic-use Sampleという形でリリースしている。

BOCを管轄する連邦法13条では、統計類の公表について、「特定事業所または個人の標章をしない」ことを定めるのみで、公表方法について特定の方法を指定しているわけではない(Title 13 U. S. C. Sec. 9(a)2, 条文は付録1を参照のこと)。現在では、この部分が「公表結果から事業所または個人を特定することができない」と解釈され、BOCによる国勢調査やCPSのPublic-use Sampleはこれに抵触しないと判断されている。

先述したように合衆国におけるプライバシーに関わる議論は、官庁統計のPublic-use Sampleに対しては概して好意的である。たとえば、BOCがはじめてPublic-use Sampleをリリースした当時、Carl Kaysenを座長とする予算局(Bureau of Budget)の委員会が、統計予算の効率化を理由として、National Data Centerの設立を提言したことがある。この提言の内容は、(a)主要な官庁統計を一元化してリンケージ可能にしたうえ重複を排除する、(b)マイクロデータへのアクセスを確保することを含んでいた。このうち(a)の点に関して、過度の情報集中を政府に許しプライバシーが侵されることを危惧する反対意見があいつぎ、結局日の目をみることにはならなかった(Hansen(1971), Flaherty(1979)など)。ところが、この間進行していたBOCによるPublic-use Sampleのリリースは、同じHansen(1971)のなかで、勧奨の対象となつてこそすれ、批判の対象とはなっていない。また、管見の限り、Public-use Sampleを巡るトラブルでBOCが訴えられた事件も確認できていない。後述するStatsCANがPublic-use Sampleを巡って世論の動向に敏感なのと、好対照をなしている。

2.2.3 Research Data Center, Center for Economic Studies

2.2.3.1 概況

BOCにおける Onsite Research は、センサス局内部の経済分析部門である Center for Economic Studies(以下、CES と略す)が担当する。CES では、Research Data Center(以下、usRDC と略す)と呼ばれる特定の敷地内に設置された特定の端末を用いて BOC のデータサーバーにアクセスし、そこですべての作業を完結させる方法を採用されている。ただし、BLS と異なり、BOC では usRDC 専用のデータサーバーが設置されており、usRDC から利用するデータは BOC 職員によって、あらかじめ固有名などが削除され専用サーバーに移し代えられている。usRDC からは、BOC が通常業務で使用しているデータサーバーにアクセスすることはできない。

また、BLS の Onsite Research Center がワシントンの BLS 事務局内に限られるのに対して、usRDC は 1994 年 1 月にボストンで最初に稼動して以来、現在では西海岸や中西部など全米 9 箇所で運用されており、その利用可能性については大きな差異がある。ただし、BLS の Onsite Research Center は利用料が必要ないのに対し、サイトでは、ワシントンの CES 直営サイト以外では、当該サイト関係機関以外に所属する利用者は、おおむね 1 年に 15,000 ドルと少なからずの利用料を支払わなければならない。各サイトの関係機関や利用料などについては付録 2 を参照されたい¹¹⁾。

usRDC は、ワシントンの CES 内部に設けられたもの以外は、現地の研究機関と BOC との合意契約に基づいて設置運営されている。この合意には、各サイトには BOC に直接雇用されている職員(RDC Administrator と呼ばれている)が少なくとも一人配置されることや、現地研究機関が BOC 職員の人件費も含む人員や敷地を提供することなどが記されている。また、各サイトと CES の専用サーバー間の回線の条件、各種セキュリティに関する設定要求などについてはハンドブックが編纂されている。それゆえ、usRDC を運営する現地研究機関の負担も少なくなく、たとえばニューヨークの例など、いくつかの研究機関が合同してコンソーシアムをつくり、一箇所のサイトを設置運用する場合もあり、ダーハムのように単独運営からコンソーシアムに運営形態を変更する例もみられる(各コンソーシアムの参加機関については付録 2 を参照のこと)。ただし、usRDC の設置・運営については National Science Foundation の経済学学会が特別の配慮を行ってお

り、設置後 3 年間について年間 20~30 万ドル相当の援助を行うのが通例となっている。usRDC は、利用料の設定や施設の増改築などについて、BOC の定める運用規定を外れない限り独自に運営することが認められており、たとえばブースが限られるためコンソーシアムメンバー以外の利用を制限している場合もある。また、当該サイトを利用する予定の研究プロポーザルの内容を審査する審査委員会は、各サイトに加盟する研究機関から審査員が選出される。

各サイトに配置されるセンサス局職員は後に見るように、簡単な開示審査やデータセットに関する相談に応じ、利用者の端末上での作業を大まかに理解する必要がある。そのため、統計・経済・計算機に関する専門知識と、センサス局の秘匿ルールに関する専門知識を同時に備えていなければならないとされてきた。現在のところ、9 つの usRDC では例外なく博士号を持ち、自らも分析に携わる研究者がこの仕事をこなしている(うち 8 つは経済学、1 つは社会学)。ただし、人件費を考慮すると、各サイトに常駐する Administrator が博士号所持レベルの研究知識を保持する必要があるかは、利用者からも疑問が呈されており、2008 年 1 月現在、学部卒業レベルの職員を配置する実験が行われている。センサス局職員以外に配置される職員は、いない場合から 3 名程度いる場合までサイトによって様々である。

2.2.3.2 RDC の開設から現在までの流れ

BOC において実施された統計の個票利用については、連邦法 13 条 9 項において厳格かつ明示的に規定が与えられている。しかし同時に a 節(3)にあるように、Special Sworn Status(通称、SSS)を有する被雇用者であれば、統計的な目的に限り個票データに触れることが可能であることも明示されている(条文は付録 1 を参照のこと。SSS は Title 13, U. S. C., Section 23(c)に規定されている)。CES では、usRDC が開設される 1994 年以前においてもこの条項を用いた統計の個票利用が行われていた。すなわち、研究者が必要書類を整え宣誓した後、形式的にセンサス局職員となることで、個票データを利用する方法である。

とくに 1986 年に Robert McGuckin が CES 部長に着任すると、事業所データの個票を利用した統計分析の重要性が強調された。そして、Longitudinal Research Database(通称 LRD)などの新たなデータベースが構築され、旧来収集されていたデータの利用可能性が拡充される一方、上記の SSS を用い

た研究者による個票利用もある程度行われた。そのほか、全米統計学会(the American Statistical Association)およびNSFC, CESが協力して個票データを用いるプロジェクトが実行されている(Dunne(2001))。この時期には、毎年2~3の研究で外部研究者が個票を利用しており、主な研究成果はDavis, Haltiwanger, and Schuh(1996)などに結実している。この際には、利用目的が統計的な目的に合致することを説明するためにプロポーザルが用意され審査されたものの、プロポーザルに何を書くのかが決められていないこと、誰がどのような権限で審査するのが曖昧であることなど、その手続きはきわめてインフォーマルであった。

当初LRDは、国勢調査やCPSと同様に匿名処理をほどこした後Public-use Sampleとしてリリースされる予定であったが、事後的に事業所の特定が不可能で、研究者の注目を集めるだけ豊かな情報量を保持したデータを作る方法は発見できなかった(McGuckin and Pascoe(1988))。McGuckinは、最終的にデータを公開する手段をPublic-useではなく、Onsite Researchに求め、NSFとNBERの協力をとりつけた後、1994年に最初のusRDCをケンブリッジのNBER敷地内に開設した。この時点での利用可能なデータセットは限られていたものの、年間20~30本のプロポーザルが提出された。

1996年にCES部長がMcGuckinからDaniel Weinbergに変わると、既存のデータセットを組み合わせることで事業所・世帯を通じたパネルデータを構築する計画がスタートした(Longitudinal Employer-Household Dynamics program, 通称LEHD)。この計画は、内国歳入庁(Internal Revenue Service, 通称IRS)の協力を仰ぎ、世帯データに付属する社会保障番号(Social Security Number, 通称SSN)および事業所データに付属する事業所納税番号を提供してもらい、同時に、各州の労働統計局に要請して、労働者の雇用保険番号を提供してもらい、これらの個人・事業所識別番号を鍵に、既存調査の個票を事業所単位あるいは労働者単位でつなぎ合わせてデータベースをつくりあげてを目的とした。このデータベースは現在も構築中であるが、1999年以降順次usRDCを通じて公開されている。

LEHDが開発されると、RDCを通じて研究者にアクセスを許すデータセットにIRSやBLSなど他省庁が関与することになった。このことから、RDCを通じたデータアクセスの際のルールがよりフォーマルに整備され、ほぼ現在の形に定められた。LEHDのように新規のデータセットだけではなく、

従来から存在していた調査についても利用可能範囲は随時拡大し、現在では60以上の調査を、usRDCを通じて利用できる。それに伴い、提出されるプロポーザルも増大し、時には年間100本を超えることもあった。調査時点現在では、60~70本の研究、200名程度の研究者が全米のusRDCを利用している。

2.2.3.3 プロポーザルから審査までの手順

こうしておおむね1999年に整えられた利用ルールを紹介しよう。

利用者はまずBOCのウェブサイト上で自分の利用したいデータセットを確かめた後、自らのアカウントを作り、最寄りのusRDCに滞在しているBOC職員とコンタクトを開始する。BLSのOnsite-Researchが原則米国民であることを要求し、例外的に国内研究機関に所属している外国人を許すのに対し、usRDCの利用要件に国籍や所属の制約はない。ただし、利用するにはSSSを取得し、無給ではあるが連邦職員としての雇用契約を締結する必要がある。したがって、これらが不可能な法的立場にある個人は実質的にusRDCを利用することができない。逆に言えば、BLSのOnsite-Research利用時に問題となった、大学院生や外国からの客員研究員であっても、SSSを得られる限りにおいてRDCを通じた政府統計の個票へのアクセスは保証されている。また、SSSを得られるか否かは、CESではなく人事局(Office of Personnel Management, 通称OPM)あるいは移民帰化局(Immigration and Naturalization Service, 通称INS)が決定しており、プロポーザルの内容の良否とは直接関係がない連邦政府の人事資格一般に関わる問題とされている¹²⁾。

プロポーザルの書き方、書類の準備の仕方については、BOCのウェブサイトには詳細なガイドラインが提示されている。それによると、15ページほどのプロポーザルには、研究内容などが記されなければならないほか、研究が次の5つの基準を満たすことを説得的に示さなくてはならない。

- a) 研究内容が連邦法13条に規定されたBOCの職務内容を満たすこと。
- b) 科学的に利点があること。
- c) Public-useのデータセットでは分析不可能であること。
- d) 分析の成功の見通しがあること。
- e) 秘匿情報・成果物の扱いに関する制限を認めること。

このうち、a)については、BOCが憲法機関であることから連邦法によって明確に職務が規定されていることに起因する。これらのプロポーザルに履歴書など必要書類を添付し、オンラインでサブミットする。プロポーザルは随時受け付けられており締め切りはない。

プロポーザルの審査プロセスは、まずサブミットされたサイトの審査委員会内で通常2名の担当が決められ6週間以内に各々2ページほどの審査報告が作成されることから始まる。この際に消費される時間は、1本のプロポーザルあたり大体4~5時間である。この審査報告がCESの担当者のもとに送られ、最終決定される。近年まで、審査委員会が物理的にワシントンに設けられ、一同に会した審査委員が議論して最終決定する方式がとられていたが、プロポーザルの増加から、現在はすべてメールなどでのやりとりで逐次的に審査が進行する方式に変更された。審査時間の多くは、利用予定のデータセットで申請された分析が可能かどうか、匿名性に抵触しないかどうかに費やされ、研究内容がBOCの職務内容に合致するかについては、実質的にはほとんど審査の対象とはならない。後でみるように、研究成果物は匿名性を保持することが必要であり、たとえば、クロス表の集計結果を成果物に含める場合などはそれに抵触する可能性がある。経済学に関するほとんどのプロポーザルは、計量経済学的な分析を目的としているためクロス表を最終成果物とすることはないが、他のディシプリンの研究者からのプロポーザルがまれに吟味の対象となることがある。

はじめて提出されたプロポーザルのうち、30~50%程度は記載事項の不備や、分析手法の不完全性が指摘され、再提出が要請される。しかし、再提出されて最終的に許可されないプロポーザルはほとんどない。例外的にプロポーザルをリジェクトした理由は、当該データセットで不可能な分析を提案した場合がほとんどである。また、各サイトの物理的なスペースの都合から、分析開始時期を相当程度遅らせることを要請することもある。複数データの接合のためにSSNなどのIRS所管のデータを用いる場合には、CESの審査プロセスを通った後、プロポーザルはIRSに回送され、審査を受ける。またLEHDについても、データ構造が複雑で複数の部局が関係することから、審査には余分な時間がかかる。

usRDCの利用を許可された後、SSSを得るための法的な手続きに移り、その後、連邦職員としての雇用契約が締結される。近年では、安全保障上の問題から、この手続きには多くの書類と指紋などの私

的情報、時間が必要とされるようになってきており、これらがすべて済んだ後、各サイトにおいてIDカードが作成され、実際の分析が可能になる。CESでは、こうしたプロセスを考えると、プロポーザルの提出からデータの利用可能までの期間を6~12ヶ月と見積もって研究計画を立てるよう推奨している。

2.2.3.4 実際の利用と成果物の公表

usRDCは独立した区画をもちIDカードによって出入りが記録されるため、基本的には24時間年中無休である。usRDC内では、利用者は端末を備えたブースを与えられ、そこで作業を完結させる。サイトによって異なるが、6~8台のブース(端末)が設置されている。各ブースに装備されている端末は、ローカルに何も持たない。したがって、利用者は、そこからCESのデータサーバーにアクセスし、そこですべての作業を完結させる。各端末に画面以外の出力デバイスは装備されておらず、途中経過などを紙あるいは電子媒体でサイトから持ち出す場合には、その都度各サイトに滞在しているBOC職員の承認を経なければならぬ。計算プログラムなどをアップロードする場合も同様である。BOC職員は24時間常駐しているわけではないので、この点では実質的に利用時間に制限がかかる。usRDCへの出入りの記録、データサーバーのログはすべて保存され、誰がいつどのようにデータにアクセスし、どのような作業を行ったかを事後的に確認することは容易である。

具体的な作業中にプロポーザルに示した計算内容を変更する必要が生じた場合には、原則として変更内容とその必要性を提出し、承認を受ける必要がある。ただし、利用するデータセットを変更する場合や、根本的に異なる研究をする場合以外は、事実上事後的な承認になっている。

またデータセットに関する質問については、まずは各usRDCに滞在するBOC職員に相談し、そこで解決しない場合はBOCの当該データの作成担当者に照会される。また、過去に同一のデータセットを使用した研究者がいる場合には、その研究者が斡旋される場合もある。

最終的な成果物は、匿名性保持に関する審査委員会で承認を得た後、CESのディスカッションペーパーとして公表され、学術雑誌への投稿も許可される。原則として、回帰分析および各変数の1次・2次の要約統計量が匿名性に触れることはない¹³⁾。しかし、多重クロス表や散布図を作成した場合は匿名性に抵触する恐れがでてくるため、審査が厳重になり、し

たがって成果物を公表するまでの時間も消費する。要約統計量であっても、中央値や最大値、最小値など特定サンプルを示す場合には吟味の対象となる。CESは、論文で掲載するクロス表はなるべく公表データを用いることを推奨している。

2.2.3.5 RDC利用者に対する法的拘束の構造

何度も指摘しているように、BOCが管轄する統計の利用条件は連邦法13条が詳細に規定しており、匿名性に対する違反などについては、すでに懲役5年または25万ドル以下の罰金(もしくはその両方)という刑事罰まで予定されている。利用者がコンソーシアムに参加している研究機関の所属者であれば、利用者が違反した場合の法的責任を、コンソーシアム契約を通じて所属機関に課すことは可能であるが、現在のところ利用者の所属機関に対するサンクションはなんら予定されていない。法的にSSSが確立している以上、usRDCを通じた個票利用者に対しては、利用者本人とBOCとの間の連邦職員としての契約関係があるだけで十分なサンクションを予定できるとCESでは考えられているからである。BLSにおけるOnsite-Researchでは、SIPSEA成立後も利用者本人とだけではなく利用者の所属機関とBLSとの間に契約関係を構成し、二重の方法で利用者に対する法的義務を担保しているのと対照的である。

BOCにおいては連邦法13条のSSSのみを用いる法的関係が中心で、usRDCができる以前から変わっていない。usRDCの設立・ルール確立は、これらの法的関係を変えるものではなく、SSSを承認するまでのプロセスを明確化したことが重要であるとまとめることができる。

2.2.4 小括

以上、BLSとBOCにおける政府統計の個票データの公開方法について、Onsite Researchを中心にまとめた。事前に許可された研究者のみに対して、特定の場所の特定の端末からのみデータへのアクセスを許し、すべての作業記録を保存可能にしようとして最終的な成果物(あるいはサイトの外に持ち出す中間生産物)について匿名性を審査するという、Onsite Researchの方法については、両者に顕著な違いはない。

また、統計の秘匿性に精通した研究者は大学にはそれほど多くなく、統計部局に比較的所属しているという事情もあり、両機関ともに利用可能な調査項目は統計の専門家が主導して決定し、成果物の審査

についても統計や法律の専門家が担当することが多い。ただし、この際の判断基準として「個人・事業所が特定されない」ことが明示されており、「回帰分析や一次・二次の要約統計量は抵触しない」など、一定の具体的な基準が事前に提示されていることには注意を促しておきたい。それに対して、事前のプロポーザルの審査には両者ともに経済学を専門とする研究者が中心的な役割を担っている¹⁴⁾。このとき主に審査される点は、申請の通りに分析が可能かどうかであって、法的には明示しなければならないBLSやBOCの設置目的との適合の可否はほとんど実質的な審査対象となっていない。BLSやusRDCでも他機関が関わるプロポーザルの場合には、追加的にデータ秘匿の観点から審査がなされるが、審査の中心はあくまで学術的な合理性が中心であり、ピア・レビューに基づく審査がなされているといえる。一方、両者で異なる特徴も散見される。

第一に、BLSの場合には、BLS本体のリソースを流用することでOnsite Researchに対応しており、利用可能性が地理的にワシントンに限定され、本体のサーバーにも追加的な負担を強いている。一方BOCの場合には、usRDCを通じて外部資金を獲得し、地理的・時間的な利用可能性を拡充し、専用サーバーの運用、専門の職員の常駐を実現させている。ただし、usRDC関係機関以外の所属者については、少なくとも利用料金を課しており、利用者の所属によって利用可能性に大きな格差があることは否定できない。

第二に、利用者は連邦職員もしくはagentとして各部局と法的関係を結び、それを通じて守秘義務が課されることは共通しているものの、BLSの場合には、利用者の所属機関とBLSの間にも法的関係を形成することで、利用者の違反を防ぐことを企図している。これに対してBOCは、潜在的に利用者の所属機関に対する罰則も発動可能な関係にありながら、利用者の規律付けについては、利用者との二者間関係のみで十分効果があると考えている。

3. カナダにおける政府統計個票公開の現状

3.1 概況

合衆国と異なり、カナダではほぼすべての政府統計を管理するカナダ統計局が設置されており、各種政府統計は統計法(Statistical Act)のもとに一元的に管理運営されている¹⁵⁾。したがって、政府統計の公開も、統計法の制限のもと、StatsCANによって統一的に管轄・実施されている。

StatsCANの特徴は、中央集権の構造を持つほか、

政府関連の情報に関する明確で強い権限を有している点にもある。統計法は、第3条において StatsCAN の設置目的を以下のように規定し、統計の種類、行政情報との関係、州政府との関係を明示的に特定している(原文は付録1を参照のこと)。

- a) 商業、産業、金融、社会、経済、国民の活動・状況に関する一般的統計情報の収集、編集、分析、要約、出版。
- b) 政府の活動により入手した行政情報を含む統計情報の収集、編集、出版における省庁との協力。
- c) 本法の定めるところによる、カナダの人口センサス及び農業センサスの実施。
- d) 省庁により収集される情報の重複防止。
- e) カナダおよび各州に関する、統合された社会経済統計の開発、統計の統合を目的とした計画の調整。

人口センサス、農業センサスは実施が特記されているほか、複数の統計の調整も役割のひとつと明記されている。これらに加えて、調査拒否者に対して刑事罰を規定することを認めるなど、カナダの統計法は、政府統計に関する強力な権限を統計局に与えているといわれている。経済情報の利用という観点からは、第3条において行政情報にアクセスする権限が与えられている点も注目すべきであろう。より具体的には、StatsCAN が租税情報や犯罪情報などにアクセス可能なことが統計法本体に明文規定されている¹⁶⁾。

公開情報の匿名性保持に関する強い要請もなされている。まず、17条(1)では、合衆国連邦法13条と同様に、個人情報と特定される情報の利用は、StatsCAN 職員または宣誓をもって認証された特別の身分(deemed employee)をもった者に限定する旨、明文化されている(原文は付録1を参照のこと)。

この一般原則の違反に対しては刑事罰が予定されており、たとえば権限のない情報へのアクセスを試みた場合には6ヶ月以下の懲役または1000ドルまでの罰金もしくはその両方が規定されている(30条)。さらに、資産価格に影響を及ぼしたり投機行為を行ったりするなど、個人を特定可能な情報を経済取引に利用しようとした場合には、5年以下の懲役または5000ドルまでの罰金もしくはその両方と、より重い刑罰が予定されている(34条)。

統計法本体では、一般原則以上の細かい規定はなされていないが、StatsCAN は独自に複数のルールを設け、匿名性の確保を担保しようとしている。ま

ず、Policy on the Security of Sensitive Statistical Information がある。この規定にはプライバシーに抵触する調査項目を扱う職員の資格要件や保管場所、廃棄の場合の方法などが具体的に定められている。EDP Security Policy は、StatsCAN が2重のネットワークを用意することを要求し、プライバシーに抵触する調査項目を含むデータを扱うネットワークと、より自由に扱うことのできるデータを扱うネットワークを物理的に切り離すことを定めている。その際に必要なセキュリティのレベルは Security Practices Manual によって定義されている。

近年の情報公開の発達に伴い、カナダにおいても政府情報の公開手順を示した Access to Information Act、政府が個人情報を扱う場合の規則を定めた Privacy Act が制定された。しかし、カナダにおけるこれらの法律は合衆国とは異なり、統計法管轄の情報を明確に例外扱いしており¹⁷⁾、StatsCAN は独立してデータを管理している。

3.2 Public-use Sample

1971年の統計法改正以前において、カナダでは政府統計の調査個票は公開されていなかった。統計法自体は調査個票の公開を明示的に禁止していなかったが、特に国勢調査の個票の公開が、たとえ秘匿処理が施されたとしても統計法に示された利用基準に合致するかどうかの法的判断が確定していなかったためである。それゆえ、政府統計の個票は個々の利用者が統計所管当局と個別特定の契約を結んだ上で利用され、利用場所も通常は統計所管当局内(StatsCANであればオタワ事務所)に限定されていた。

1966年に司法省から、統計法は秘匿処理をほどこした個票を公開することを何ら禁止していないとの回答を得た StatsCAN は、秘匿処理を施した個票の公開規定を含んだ統計法の改正に踏み切った。改正統計法は1971年に成立し、ついで個票データの公開に関する手順や基準をルール化した Microdata Release Policy が1973年に制定され、まず国勢調査の個票データが Public-use Sample としてリリースされた。以降、国勢調査のみならず家計調査や労働力調査などの世帯調査についても Public-use Sample が公開されている。

現在では、リリースされた Public-use Sample をストックし、データアーカイブ的な運用を行う会員制の組織である Data Liberation Initiatives(以下、DLIと略す)が設立されている。DLIは合衆国の ICPSR と同様の理念と形態をとっているが、ICPSR と異なり、政府統計部局である StatsCAN

が明確に運営に関わっていること、原則として国内の大学をすべて加盟させ、こと研究者や学生に対しての Public-use Sample の利用の便を平等に確保することには、特に意が割かれている。

各調査の Public-use Sample 作成に当たっては、Microdata Release Committee がリリースするサンプルや調査項目を吟味することが定められている。その際の検討基準も Microdata Release Policy に定められており、たとえば、個人を特定可能なすべての情報を削除することや、個人の特定につながる可能性がある調査項目(およびその組み合わせ)を削除するか集計することがあげられている。カナダの公開基準は合衆国と比較すると厳密に解釈・運用されている。その結果、特定されやすい事業所データのみならず、世帯データについても、健康・教育に関するデータやパネルデータは調査原票に近い形では公開されていない。また、同様の理由で、Off-site Research の方法も採用されていない。合衆国と比較して自由に利用できるデータセットが少ない点は注意されるべきであろう。

ただし、カナダ統計法と合衆国連邦法 13 条の明文規定を比較しても Public-use Sample の条件に関して大きな違いは認められず、カナダにおける Public-use Sample 作成時の制限的な運用の主な理由は法律的な枠組みの違いにあるわけではない。もちろん、事業所データの場合には、自然独占が成立しやすい産業が多いこともあり合衆国と比較して客体数が非常に少なく、簡単なクロス表で事業所が特定できてしまうという事情もある。しかし、インタビューのなかでは、Public-use Sample としての公開が被調査主体の協力姿勢に強い影響を与える可能性が指摘され、この点が合衆国と比較すると異なるという意見が聞かれた。StatsCAN が調査を行うときには、あらかじめ調査客体に調査結果の利用範囲を特定してもらうことが一般的である。他官庁であればその官庁での利用について、RDC での利用であれば“for use of research”での利用について同意するか否かが聞かれる。ここで No と答えた被調査主体の個票が他官庁に提示されることはないし、RDC で利用されることもない。カナダにおいては、この同意割合が時々の政治状況や官庁でのスキャンダルの有無などに大きく左右されるのが現実で、統計調査結果を公共財として意識し一般的に Public-use Sample として公開できる状況にはないという意見が複数のインタビューから聞かれた。Public-use Sample がはじめてリリースされたときの合衆国の世論の寛容さと好対照をなしているかもしれない。

3.3 Research Data Centre

3.3.1 概況

1980 年代に入ると、研究手法の一般的な発展からより詳細なデータの公開が要請されるようになった。しかし、事後的に個人が特定されないという統計法の規定に抵触する可能性があることから、Public-use Sample の調査項目を詳細にすることは控えられた。また、急速に整備されつつあったパネルデータも、その性格上 Public-use Sample として積極的に公開することはできなかった。

こうした中、政府内部に蓄積されているデータの利用促進について Social Sciences and Humanities Research Council (以下、SSHRC と略す)と StatsCAN との合同調査会がもたれ、1998 年 12 月に最終報告がまとめられた(Final Report of the Joint Working Group of the Social Sciences and Humanities Research Council and Statistic Canada on the Advancement of Research Using Social Statistics)。この報告では、1990 年代のカナダにおける統計利用を巡る問題点を吟味し、(1) 1980 年代にデータ開発が先行した反面、行政のダウンサイジングの影響で統計部局に(とくに継続データに関する)よく訓練されたスタッフが不足するようになったこと、(2) 統計法によるデータアクセスの制限によりデータが活用されにくく、データ利用者が合衆国に流れていること、(3) 研究者と統計作成者・政策担当者のコミュニケーションが不足していることの 3 点に集約した。そしてそれぞれの解決法として、(1) 人材育成に関しては Research and Training Group を設立し、Training Program や Fellowship Program を充実させること、(2) のデータアクセスに関しては Onsite Research である Research Data Centres (以下、caRDC と略す)を設立すること、(3) のコミュニケーションに関しては Social Statistics Communication Program を設立し、研究者、行政担当者らによる定期的な Research Forums を開催することを提言した。

この報告では、当時運用が開始されていた合衆国センサス局の usRDC を参考にして、StatsCAN および SSHRC が協力して caRDC を設立することを示した。そして、この提言に従い、“Anywhere, No pay, No two review,” の原則の下、2000 年 12 月に RDC Network が発足した¹⁸⁾。

報告のなかでも触れられているように、caRDC の物理的・運営上の構造は usRDC と酷似している。たとえば、データ利用者は宣誓の上、StatsCAN と無給の雇用契約を結び、統計法 17 条上の特別な雇

用者としての法的地位を得て、個票へのアクセス権限が与えられる。この契約を通じて、データ利用者は利用規定や罰則など、統計法上の義務を負うことになる。

また、実際のリサーチを行う場所は、利用者個人別に発行されたセキュリティカードで管理される独立した区画に設けられることも同様である。caRDCはStatsCANをはじめいくつかの大学・研究施設内に設置されており、2007年1月現在、カナダ国内に本所支所合計20箇所(関係機関は38)を数える¹⁹⁾。caRDCの開設マニュアルでは、全学の研究者にアンケート調査を行った後、最低限15~20名のサポートが得られることが望ましいとしており、2007~2008年にさらに4箇所が開設予定である。usRDCと同様に、これらの施設は法的にはStatsCANの建物として借用され、各施設にはStatsCAN職員が少なくとも1名常駐する。StatsCANの職員は、従来StatsCANでキャリアを歩んできた職員が転勤することもあれば、PhDを取得したばかりのPD段階の研究者を新たに雇用する場合もあり、各サイトに裁量が任されている。これらの職員は採用あるいは転勤時に数週間の初任者研修を受け、とくにデータセットの特徴や秘匿性に関するルールについて習熟することが求められている。

各施設にはワークステーションが備え付けられており、利用者はこの端末を用いて各caRDCに設置されたデータサーバーにアクセスし、データを利用する。各端末には画面出力のみ認められ、印刷物や電子ファイルで出力する場合には、StatsCAN職員の認証の下、原則としてStatsCAN職員が専用で使用する端末からのみ認められているのも同様である²⁰⁾。設立時の原則にのっとり、ネットワーク参加大学に所属する利用者には利用料は発生しない。セキュリティは、あらかじめ用意されているマニュアル“Security in the Research Data Centres”にしたがって構築され、開所時にHead Officeの監査を受ける必要がある。このマニュアルには、壁やドアの材質などからセキュリティチェックの仕組み、コンピュータ関係に至るまで規定がある。

usRDCとcaRDCのもっとも大きな違いは、運営形態と利用範囲に求められる。一方のusRDCは、基本的に個々の大学あるいは複数の大学とBOCとの契約に依拠し、センター間相互の資金のやりとり、連絡や協力はほとんどみられない。他方caRDCは、学界全体のバックアップのもとに成立したという事情もあり、RDC Networkのリーダーシップが強く、センター間の協力関係が強く意識されている。たと

えば、一度承認された研究計画はどこのセンターでも実行することができ、所属地の変更やサバティカルなどにも容易に対応できる。また、usRDCは研究目的の利用にほぼ限定されるが、caRDCは主に大学院レベルでの教育目的の利用も積極的に推進しており、実際いくつかのセンターではRDCを利用した講義もすでに実施されている。

もちろん、個々の物理的な環境についてもusRDCと異なる点が散見される。第一に、データを格納するサーバーは各サイトに設けられており、端末は各サーバーにアクセスするのみでStatsCANのデータサーバーにはアクセスしない²¹⁾。また、備え付けられる端末の構成は各サイトに任されており、SUNのようなServer-Terminal形式をとりOSもUNIXが用いられることもあれば、それぞれ端末にメモリーとCPU、ローカルディスクが装備されたPCが設置され、Microsoft Windows上からサーバーにアクセスする形式がとられることもある。また、ワークステーションの数も各サイトによりまちまちで、モントリオール大学に設置されているRDC(通称CIQSS)では30台のPCが、オタワ大学に設置されているRDC(通称COOL)では14台のPCが、クイーンズ大学に設置されているRDCでは12台のターミナルが装備されている。caRDCの場合にはセキュリティ・エリア内での利用者個人のPCの持込、PDAや携帯電話などの使用は原則として認められていない。ただし、CIQSSの場合には、セキュリティ・エリア内に特別な一室を設け、そのネットワークを本体のネットワークと隔離した上で、外部との接続を許すという形がとられている。

第二に、20箇所のcaRDCはFull-time Centre, Part-time Centre, Branchの3つの形態にわかれていることも指摘しておきたい。Part-time CentreにはStatsCAN職員が派遣され、通常の運用がなされるものの、端末の数が少なく、たとえば週3日などのように開設時間が限られる。この場合のStatsCAN職員は、たとえば当該大学のリサーチアソシエイトとの併任として雇用されることが多く、利用が少ないことが予想できる場合には財政的負担を少なくとどめることができる。実際、クイーンズ大学のセンターは週3日開館しているが、そのStatsCAN職員は経済学部非常勤講師もかねており、自らの講義でセンターを用いている。BranchにはStatsCANの職員は常駐しない。契約開始や出力などStatsCANの職員の認証が必要な場合には、直接出向か電子メールを介してFull-time Centerの職員とやり取りする必要がある。これらの設置形態

の選択には RDC Network の承認が必要で、主に利用人数や地理的条件などから判断される。また、caRDC は原則としてウィークデーの昼間という通常の開館時間を維持しており、usRDC のように年中無休ではない。

物理的な環境の特徴の第三として、セキュリティ・エリア内には責任者および StatsCAN 職員のオフィス、セミナールームを設けることが推奨されており、CIQSS のように端末を利用しながらの研究会も開催できるように配慮されていることがある。それゆえ、通常端末は 1 室に集中して配置され、CIQSS や Queens のように各端末間にパーティションは設けられないか、COOL のように設けられたとしても 30cm 程度の低いものが多い。

1 箇所 Full-time RDC を運営する費用は平均するとおおむね 279,020 カナダドル/年であるが、このうち設置大学によって賄われるのは 40% (およそ 80,000 カナダドル) で、残りの 60% (およそ 200,000 カナダドル) は、StatsCAN のほか Canadian Institutes of Health Research (以下、CIHR と略す) および SSHRC が RDC Network に資金を提供し、Network が加盟 caRDC に配分する形がとられている。この際、各 caRDC には固定的費用として 1 箇所 115,000 カナダドル/年が配分されるほか、利用者数やプロジェクト数などに応じて変動費も配分される。典型的な年間予算のうち、StatsCAN 職員の人件費に 100,000 カナダドルが見込まれるので、人件費分はネットワークより配分される固定的費用で賄われる計算となっている (Annual Local Costs)²²⁾。このネットワークは拡大の一途をたどっており、近年 Canadian Foundation for Innovation (以下、CFI と略す) が caRDC の設備更新のための資金提供をはじめた。また、the Department of Human Resources and Social Development (通称 HRSDC) も主に設備更新とネットワーク全体の活動のための資金提供をする予定である。

caRDC で利用可能データは調査時点で世帯データを中心にウェブの違いなどを含めると 47 種類を数える。The National Longitudinal Survey of Children and Youth (NLSCY) や Survey of Labour and Income Dynamics (SLID) などのパネルデータのほか、National Population Health Survey (NPHS) や Canadian Community Health Survey (CCHS) などの健康教育関連のデータセットが豊富である。また、Aboriginal Peoples Survey (APS) などの少数民族や移民に関するデータセットも利用可能である。活動報告書によれば、2006 年 10 月 31

日現在活動中の 776 本のプロジェクトで申請されているデータはのべ 1255 を数えるが、そのなかでもっとも多いのが NLSCY で 245 本、ついで CCHS が 223 本、そして SLID が 143 本の研究で申請されている。NPHS は 132 本の研究で利用が申請されているので、この 4 つのデータのうち少なくともいずれかを用いる研究が大部分をしめていることになる。カナダにおいては Public-use Sample が公開されていないパネルデータの利用と並んで、健康教育関連のデータセットが実際にも多く利用申請されているところに特徴があるといえよう。

3.3.2 プロポーザルから審査までの手順

usRDC と同様に、StatsCAN と雇用契約を結びデータを利用するにも、あらかじめ研究計画を提出したうえ、Selection and Review Committee による承認が必要である。ここでは、審査までの手順を要約しよう。

caRDC の利用資格は、原則として RDC ネットワーク参加機関に所属があるうち、以下のいずれかを満たすものに限定される。

- a) 研究者 (所属があれば国籍は問わない)
- b) 連邦政府機関の被雇用者 (州地方政府被雇用者は別途相談)
- c) 大学院生のうち修士論文または博士論文執筆の一部として利用する者
- d) 大学院生のうち修士論文または博士論文とは関連せず利用する者
- e) 学部学生のうち最終学年に属する者

usRDC と比較して、大学院生の利用が特に明確にルール化されている点、学部学生にも門戸を広げている点に特徴がある。所属・国籍ともにカナダと関連しない研究者もまったく利用が拒絶されているわけではなく、別途相談のことで決められており、かなり広範囲に利用者を想定しているといえる。

統計法上の擬似的雇用契約を結ぶときには BOC の方式を採用しているため、所属機関は関与しない。外国の研究者はネットワークに所属する機関に所属があれば、そのみで、その期間内においては caRDC を利用する可能性が当然に認められる。非ネットワーク参加機関に所属している場合には、大学院生と同様にあらかじめ SSHRC program officer の承認を得る必要があることを除くと、通常の研究者と同様のプロセスをなぞる。

大学院生については、c) の論文執筆に関連する場

合には、論文審査委員会に属する指導教員による研究計画の確認および推薦が必要で、同時に、指導教員を共同利用者として含めなければならない。ただし、指導教員を利用代表者とするか大学院生を利用代表者とするかは問わない。また、審査を集約する SSHRC program officer の承認をあらかじめ得た書状を揃える必要がある。d) の場合には、通常論文審査委員会が組織されていないため、一般的な指導教員による研究計画の確認および推薦に代わるほかは、c) と同様の手続きになる。caRDC では、カナダ国外の大学院生が利用することも予定している。この場合は、c) または d) のプロセスで必要な書類のほか、利用を予定している caRDC の責任者と事前に相談し、了承を得る必要がある。

学部学生については、利用代表者は必ず指導教員で、学部学生は共同利用者の位置づけとなる。また、プロポーザルには Public-use Sample で同様の分析が不可能であることを説得的に述べなければならない旨、特に注意がなされている。

用意するプロポーザルについてはウェブサイト上に執筆マニュアルが掲載されている。それによれば、原則として5ページ以内と定められており、usRDC と同様に、研究の目的や具体的内容、利用データやソフトウェアなどを含むべきことが決められている。そのほか、Public-use Sample では不可能な理由やプロジェクトの期限も記す必要がある。しかし、BLS や usRDC と異なり、統計部局の業務との関連を記す必要はない。通常、学術的な研究動機と統計部局との業務は無関係である。したがって、実質的に審査対象となっていないとはいえ、この関連をプロポーザル上合理的に説明しなければならないことは、BLS や usRDC での応募に少なくない心理的負担をかけていると考えられる。カナダではこの点を省略することで、よりスムーズな応募を実現しようとしている。また、各 caRDC には StatsCAN 職員が常駐しているので、データの利用可能性についてなど、プロポーザルは前もって当該職員と相談するとスムーズである旨、マニュアルには記述がある。利用の多い NLSCY や APS には、特に専用のプロポーザル執筆用のマニュアルが用意され、利用者の便宜が図られている。

このように、caRDC を利用するのに必要とされる申請書は、usRDC や BLS のオンサイトリサーチを利用するときほど厳密性が求められるわけではない。たとえば推定式を特定する必要はなく、説明変数や被説明変数、操作変数などをあらかじめ特定しておく必要もない。カナダでは申請書がしばし

ば“extended abstract”と称されることにも現れている。さらに、近年では、具体的な分析対象が明確に定まっていなくとも、分析の分野や方向性のみを示した“Research Program”に対して包括的なデータ利用を認める試みも行われつつある。

caRDC 利用希望者は、SSHRC のウェブサイト上にユーザアカウントを作成したうえ、研究計画書のほか利用者本人および共同利用者の履歴書などを添えて、ウェブサイトを通じて応募する。

ウェブサイトを通じて応募された研究計画書は、Selection and Review Committee に回送され、2名の研究者と StatsCAN 職員によって審査される。審査を担当する研究者は関連分野から選ばれ、研究目的とデータ・分析手法の合理性を主に審査する。統計局職員は個票へのアクセスの必要性や分析結果の匿名性にかかわる部分を主に吟味する。ただし、RDC 設立時の原則のひとつに“No Two Review”があり、利用資格のうち a) の研究者と c) の修士または博士論文執筆のために利用申請した応募者が、同じ研究で SSHRC または CIHR の補助金をすでに得ている場合には、研究者による審査は省略され、StatsCAN 職員による形式要件のみの審査となる。また、現在 c) の修士または博士論文執筆のための利用申請の場合には、指導教員による推薦状をこの審査プロセスに替えることが検討されている。2000年12月の開設より受け付けられた研究計画は1690本にのぼるが、このうち研究計画自体がリジェクトされたものは23本で2%にとどかない。その理由はほぼ「Public-use Sample で可能な分析」「被申請データでは明らかに分析目的が達せられない」という理由である。

マニュアルでは、審査期間は概ね8週間と記されている。実際には、2006年3月より9月までの6ヶ月間では、審査期間(休日など手続き上費消された時間も含む)の中位値は、学生の利用で35日、研究者の利用で44日となっている。75%点は学生の利用で51日、研究者の利用で58日なので、4分の3が8週間以内に審査されていると考えてよい。講義など教育目的の場合の利用などではおおむね2週間で手続きを終えられるように特別の便宜が図られている。

3.3.3 実際の利用と成果物の公表

研究計画が承認された後、StatsCAN との間に雇用契約が締結され、利用者は個票へのアクセス権限を得る。ただし、雇用契約を締結するためには、様々な個人情報や事前に StatsCAN に登録するこ

とが必要で、各サイトに出向いて職員の同席の下で書類を作成しなければならない。

これらの手続きが完了した後、各 caRDC では StatsCAN 職員によるオリエンテーションが行われ、この際に宣誓書が作成され、StatsCAN との雇用契約が結ばれる。また、承認された研究計画書は(雇用契約とは別個の)StatsCAN との契約内容の一部となる。具体的には、StatsCAN から提供されるデータセットの内容、実行される研究内容、研究の開始日と終了日、各サイトに定められたセキュリティポリシーの遵守、最終成果物の StatsCAN への提出の約定である。したがって、アクセス可能なデータはあらかじめ研究計画書の中で指定したデータセットに限られ、原則として研究計画書に記述された分析のみが実行できる。usRDC と同様に、SSHRC や StatsCAN、各サイトの StatsCAN 職員が、研究内容が変更されたとして研究計画書の変更・再審査を命じることもできるが、どこまでが同一の研究計画であるかについては、個々の職員などの判断に任されている。現在までのところ、変更が必要と認められた場合でも契約書を再締結した例はなく、すべてが既存契約書の変更によって対応されている。

研究遂行中、caRDC より持ち出す中間成果物については、各サイトの StatsCAN 職員の承認を得る必要はあるが、StatsCAN へ提出する必要はない。StatsCAN に対して提出義務があるのは最終成果物のみで、(1)RDC working paper, (2)Peer-reviewed journal article, (3)Book or Book chapter, (4)Thesis or Dissertation, (5)Commissioned Report のうちいずれかの形態で提出されることがあらかじめ契約される。ただし、何をもって最終成果物とするかは曖昧である。修士・博士論文の一部を書くために caRDC の利用を申請し学位を取得した場合や、成果物を含む論文が公刊された場合は分かりやすい。しかし、一般に所属大学のディスカッションペーパーとなった場合や、学術雑誌への投稿原稿を最終成果物とすることはない。最終成果物を提出した場合には caRDC の利用契約が打ち切られるので、レフリープロセスで再推定が要求された場合などに対応できなくなってしまうからである。したがって、学術雑誌への投稿後のレフリープロセスに乗ってしまったあと、相当の時間プロジェクトが休眠状態になることも多い。また学術雑誌に掲載が決まった後も、データ利用時から時間がたってしまったがために最終成果物としての caRDC への提出を失念する研究者も多い。caRDC では休眠中のプロジェクトについては一時利用を中断し、プログラ

ムの継続期間を実質的に延長する運用をしているが、プロジェクトが休眠中であるかどうかを個々に確認するのは容易ではなく、結果として多くの休眠プロジェクトが生まれているのが実情である。

3.3.4 近年の活動

caRDC の近年の活動内容を、直近の活動レポート(RDC Program Manager's Report November 2006)から概観したい。

caRDC が活動を開始したのは2000年12月で、本ヒアリング調査時点でほぼ6年間を経過している。この間、2006年10月31日までに提出されたプロポーザル総数は1690本にのぼる。そのうち306本の研究が完了し、最終成果物は313本提出された。このうち159本は学術雑誌に掲載されている。また、2006年10月31日現在、776本の研究が活動中で、516名の研究者、337名の学生、合計853名によって利用されている。usRDC では9箇所合計で60~70本の研究、200名程度の研究者が利用中であったことを考えると、カナダでは Public-use Sample が少ないことを考慮しても、usRDC よりも後発であるわりに、むしろ caRDC のほうが活発であるといえるかもしれない。

ただし、承認された後取り下げられた研究が53本、現在休眠中の研究が381本と少なくない。現在休眠中の研究は学術雑誌への投稿段階にあるとも考えられるが、利用が許可された研究のうち3~4%が途中で放棄されていることには注意すべきかもしれない。caRDC ではレビュープロセスの迅速化が進んでいることの裏返しとして、ある程度投機的な研究も進められている可能性がある。この点は新奇な分析を追及する学術研究としては望ましく、途中で放棄された研究の成果物は流出しないという On-site Research の仕組みをうまく利用した部分といえよう。また、最終成果物313本を主要な分野でわけみると、経済学が第一とされるのは107本で過半にみえない。人口学や社会学など、経済学以外の社会諸科学でよく利用されている実態が浮かびあがってくる。

3.4 小括

以上のように、caRDC の運用は、学術団体である SSHRC のサポートのもと形式的に整えられており、広範な利用を想定している点に特徴があるといえる。最後に、ネットワークの役割と州政府との関係について付言しておきたい。

RDC Network では集められた総資金の10%は

ネットワーク中央(The National Coordinating Committee of the RDC Program, 通称 RDCNCC)で利用される。ネットワーク中央は主に3つの役割、すなわち管理事務、ネットワーク相互の連携、コンファレンスの開催が与えられている。

とくにネットワーク相互の連携については、Knowledge Transfer Officer(通称 KTO)と呼ばれる専門の職員を雇用するなど、さまざまな努力がなされている。この職員の現在の主な役割は利用データセットに関わるノウハウの集積および共通化である。利用するデータセットを十分使いこなすには、個々のデータセットの細かな特徴をつかむ必要があり、研究者や学生が新たなデータセットを使い始めるときの困難の多くは、このプロセスにある。KTOは個々のデータセットに関する利用マニュアルやFAQの作成、利用プログラムの共通化をすすめており、NLSCYやAPSではその成果として専用のプロポーザル執筆用のマニュアルが作成された。KTOの他には、各サイトに所属するStatsCAN職員は採用時に2週間程度の共通の初任者研修を受けるほか、年に1度3日間程度の日程で1箇所に集まり、データセットや統計的手法についてなどの共通理解を深める場が組織されている。

カナダにおけるRDCは健康や教育分野のデータセットが豊富であることや、当初より社会科学全体の学界が深く関与していることなどから、学際的色彩が非常に強くネットワーク相互の連携に対する利用者からの需要も強い。usRDCの利用者がほぼ経済学の研究者に限られるのと好対照をなしている。年に一度開催されるコンファレンスも、学術分野別というよりは利用データセット別、問題関心別に組織されており、カナダの特徴をなしている。

また、2007年1月現在、中央では保守政党が政権を運営し、州政府ではリベラル政権が多いという事情もあり、州政府とcaRDCとの関係は微妙である。カナダでは健康保険や教育に関わる行政は基本的には州単位で行われているため、caRDCの強みである健康教育分野では、州政府に詳細な行政データが保持されている。また、州政府自身がこれらの分野の統計調査を実施することもたびたびある。本来、カナダの統計法は州政府を含めた他の行政官庁のデータを収集・利用することをStatsCANに認めているが、政治的な状況も影響し、現在のところ州政府所管のデータはcaRDCには格納されていない。むしろ、アルバータ、マニトバ、ケベックの州政府は、StatsCANの所管するcaRDCと形式的には独立であるが、形態や法的構造などをほぼ踏襲する形

でRDCタイプのデータセンターを設置しはじめている。たとえばケベック州統計局は、Centre d'accès aux données de recherche de l'Institut de la statistique du Québec(通称 CADRISQ)を設置運営して州政府のもつ統計の個票や行政データへのアクセスを許しているが、このセンターはネットワークもセキュリティも別に管理されているものの、QICSSと同一区画内に併設されている。CADRISQに常駐するケベック州政府職員もQICSSのStatsCAN職員と頻繁にやり取りしており、相互の協力関係を構築しつつある。

4. まとめ

わが国の政府統計の目的外利用は、米国におけるOffsite Researchに相当するといえるが、目的外利用を許可する際に用いられる「公益性」を判断する基準や利用者の所属機関の法律上の位置づけなど曖昧な点が多く、それゆえ目的外利用の審査も曖昧であることは否定できない。加えて、事前に詳細な推定方法や利用調査項目を提出しなければならず、プロポーザルの作成・審査に多くの資源が割かれる一方、利用中の状況や成果物に対する査察はほとんどなされないという欠点を有してきた(これらの諸点については山口(2008)を参照されたい)。

もちろん、これは、学術研究においては、事前に推定方法や利用調査項目を特定しコミットすることが一般に困難であることに起因している。事実、BLSや両RDCにおける典型的な利用期間は2年間を超えており、データを利用しながら事後的にプロポーザルの修正が適宜なされているのが現実である。元来、学術研究を目的とする目的外利用は、集計様式に事前にコミット可能な「オーダーメード集計」とは本質的に異なる性格をもつ。ところが、現在までのわが国における目的外利用の枠組みは、学術的な利用があまり念頭に置かれておらず、それゆえにこの二者を区別してこなかった²³⁾。確かに、2008年1月現在、改正統計法のもとに発足した統計委員会にて、目的外利用に関する手続き等も審議中ではある。しかし、改正統計法において目的外利用の位置づけそのものには変化が加えられなかったことをみても、この点については不透明感がぬぐえない。

学術研究を目的とした個票データの利用に際しては、研究手法に適した目的外利用方法の設計が望ましい。具体的には、可能な限り速やかに包括的な利用可能性を確保することである。他方、調査データの守秘義務からは公表される成果物に関して秘匿性を確保することが強く要請される。この二点を両立

させる仕組みとして、Onsite Researchの方法は最も適している。北米においては、物理的に隔離されたサイトを確保すること、利用可能なデータの整備・選定と秘匿性に関わる事後審査には統計部局が、プロポーザルの審査には研究者組織が主に関わることで比較的効率的な運用を可能にしており、日本への応用可能性も検討するべきかもしれない。

もちろん、北米のOnsite Researchの方法にも問題が垣間見られる。たとえば、プロポーザルの提出から利用可能になるまでに数ヶ月の長さが必要なこと、研究者の所属機関によって利用可能性に格差があることなどである。しかし、利用までの時間については、法的関係を結ぶことに消費される手続き上の時間のほかは、もっぱら研究者の事情が作用している。usRDCではプロポーザル1本あたりの審査時間を4~5時間と見込んでいるが、この審査結果を得るまでの標準的な期間を6週間としている。つまり、審査を担当する研究者にとって、usRDCのプロポーザル審査の優先順位は必ずしも高くなく、このことが審査期間の長期化につながっている。caRDCでは、この点を改善するために学術団体の全面的な支援と研究者相互の協力関係の構築を推進している。同時に、“No two review”の原則を明示することで各種補助金の審査との重複をさけるなど、比較的迅速なレビュープロセスを実現させており、参考となろう。

所属機関による利用可能性の格差は、Onsite Researchを設立・運用する費用の負担の問題と関わる。RDCのように、独立したサーバーの構築・運用、専門職員の常駐などに費用がかかる場合には、それを負担した機関に優先的利用権が与えられることはそれほど不合理ではない。しかし、現実的な研究者の費用負担の限度を考慮すると、可能な限りアクセスは平易なことが望ましいし、費用を学界全体に負担する方法も考えてしかるべきであろう。ただし、日本においては北米ほど地理上の利用可能性が限られることはなく、Onsite Researchを設立・運用するデメリットはそれほど大きくない。また、caRDCのPart-time方式やBranch方式をうまく利用することで、最小限の予算でOnsite Research Citeを作る方法もあり、この点も参考にすべきであろう。

この際、最も重要なことは、研究者組織内部の協力、研究者組織と統計部局との協力である。カナダの例で典型的にみられたように、Onsite Researchの方法が少なくとも失敗しないための条件として、プロポーザルの事前審査にかかる時間を可能な限り

短くすること、利用者の守秘義務違反に対する学界としての抑止力が効果的であること、などがあげられるが、これらの条件はいずれも研究者組織の助力が得られなければ成り立たない。逆に、個票データ自体の選定・整備のみならず、最終的な成果物に対する匿名性の審査についても研究者組織では不足し、統計部局の協力は不可欠である。BLSとBOC、StatsCANがそれぞれ異なる運営方法をとっていることからわかるように、サーバーの独立運営や、所属機関の法的位置づけ、統計部局雇用職員の常駐の有無など個別の仕組みについては、財政的な制約も考慮しなければならず様々な方法がありえる。しかし、研究者組織内部の協力、研究者組織と統計部局との協力がこの方法の要であることには変わりはない。

(一橋大学経済研究所)

注

1) 本調査は著者がYale University MacMillan Centerに滞在中に行われた。同センターに深く感謝したい。本調査報告の多くは、Chris Chapman(BLS)、Ron S. Jarmin(CES)、Raymond F. Currie(University of Manitoba)、Gustave Goldmann(Statistics Canada)、Saul Schwartz(University of Ottawa)の各氏に対する聞き取りおよび各氏から提供していただいた資料・情報に基づいている。さらに、Jim Davis(usRDC at NBER)、Margaret C. Levenstein(usRDC at University of Michigan)、Kirk White(usRDC at Duke University)、Becky Acosta(usRDC at UCLA)、Jean Poirier(QICSS)、Jean-Michel Billette(COOL)、Casey Warman(caRDC at Queens)の各氏にはメール・電話での聞き取りやRDCの現場の説明などにご協力いただいた。また、本調査報告を作成する過程において、川口大司(一橋大学)、松井博(総務省)の両氏、一橋大学経済研究所におけるセミナーの参加者からは適切なアドバイスをいただいた。長時間にわたる聞き取り調査にご協力いただいたChapman, Jarmin, Currie, Goldmann, Schwartz各氏、協力いただいた関係者の方々に深く感謝したい。

2) 広く公開されているものとしては、福重(2003)、山口(2008)などがある。

3) 2008年1月現在、部分的には1850年センサスまでOffsiteで利用可能である。<http://www.ipums.umn.edu/>を参照のこと。

4) そのほかに、ウェブサイトなどを経由した再集計プログラムも提供されており、Special Tabulationと呼ばれている。

5) 国立教育統計局(National Center for Education Statistics)のOffsite Researchを中心にまとめたものとしてSeastrom(2001)がある。

6) BLSの無給のAgentとなることの合意契約である。この契約の内容に関しては後段を参照のこと。

7) この3部署以外が管轄する統計についてはOnsite Researchでの利用が認められていない。

8) 後述のように、2002年にCIPSEAが施行され人事交流法の制限が実質的になくなったことから、大学院生を対象としたボランティア制度は廃止された。

9) <http://stats.bls.gov/bls/blsresda.htm>

10) 一般に論文出版に関わるレフリープロセスのなかで計算のやり直しなどが命じられることは多い。BLSでは利用期間が過ぎた後であれば、レフリー・コメントに対応するだけであっても再び利用申請をし直すことを建前としているが、Barkume and Spletzer (2003)によれば2003年までに関してはそのような要請はなかった。

11) Carnegie Mellon University (Pittsburg)に1997年に設置されたRDCは、利用数が伸び悩んだことと部局の方針変更により2004年に閉鎖された。

12) 2001年9月11日の同時多発テロの発生と、それを受けた愛国者法(Patriot Law)の施行に伴い、出入国管理のみならず連邦政府職員の任用についても人種的・思想的背景をより厳しく審査する傾向が強くなった。特に外国籍の研究者がSSSを得るには、サポートレターを含む多くの書類をOPM・INSに提出せねばならず、特に大学院生などのusRDCの利用費用が増大した。

13) 小さな標本サイズで説明変数にダミー変数を用いる場合には注意が喚起される。

14) BLSにおける審査では、明確に外部研究者が関わっているわけではない。しかし、BLSの各部署のシニア・エコノミストは博士号をもち、アメリカ経済学会でも活躍する研究者であることは指摘しておきたい。

15) 1867年の連邦国家成立(いわゆるConfederation)の後しばらくは、各省庁によって統計が管轄されていたもの、とくに人口・農業の両センサスが重要視されていたもの、各省庁に永続的な調査担当部局が置かれたのはようやく1905年になってからである。その後1918年には統計法・StatsCANが成立し、中央集権的制度に移行している。

16) 租税情報は24条および25条、犯罪情報は26条から29条に規定されている。

17) Access to Information Act 24条, Privacy Act 8条(2)

18) この際、Public-use Sampleの開発も引き続き行うことが特に明言されており、RDCネットワークとDLIは共通のボードメンバーをもつなど協力的な関係を構築している。

19) 大学に所属する利用者は基本的に各大学が所属するcaRDCを利用することが前提とされており、オタワにあるStatsCAN内のcaRDCは、政府系研究機関関係者の利用が想定されている。

20) CIQSSおよびCOOLでは、データ利用者がフリーで利用できるプリンターが用意されており、印刷される紙の色をグリーンにし、チェック後外で持ち出せる白い用紙と区別されている。

21) ただし、2007年1月の時点では、オタワのStatsCANにのみデータサーバーをおき、専用回線で各RDCを結ぶ形態が検討されていた。

22) これに対してStatsCAN職員が常駐しないBranchタイプのRDCでは年間79,800~126,000カナダドル程度の運営費(うち大学負担分22,800~36,000カナダドル程度)が見込まれる。

23) この点、1960年代には大量データの処理を可能にする計算機が一般には利用可能ではなかったという事情もあるものの、BOCがPublic-use Sampleのリリー

スと同時にSpecial Tabulationの受付を制度化したことは示唆的である。

参考文献

- (Reports and Manuscripts)
Annual Local Costs, manuscript.
 Currie, Raymond and Goldmann, Gustave (2005) "D'un noyau... From a Seed..." presentation material to the 2005 Census Research Data Centers Conference, Ithaca, N. Y.
 Currie, Raymond and Goldmann, Gustave (2006) "D'un noyau... From a Seed..." presentation material to the Canadian Foundation for Innovation External Review Committee.
Final Report of the Joint Working Group of the Social Science and Humanities Research Council and Statistics Canada on the Advancement of Research using Social Statistics, December 1998.
Guide for Researchers under Agreement with Statistics Canada, Oct. 2005.
 Hansen, Morris (1971) "The Role and Feasibility of A National Data Bank, Based on Matched Records, and Alternatives," Federal Statistics, *Report of the President's Commission*, Vol. 2, pp. 1-62.
Procedures for Opening a Research Data Centre, manuscript.
RDC Program Manager's Report, Nov. 2006.
 U. S. Bureau of the Census, *Center for Economic Studies Annual Report 1994-95*, October, 1995.
 U. S. Bureau of the Census, *Research at the Center for Economic Studies and the Research Data Centers : 2000-2004*, April, 2005.
 U. S. Bureau of the Census, *Research at the Center for Economic Studies and the Research Data Centers : 2005*, May, 2006.
 (Articles)
 福重元嗣(2003)「わが国における官庁統計の個票利用と経済分析——科研費プロジェクト以前の状況について——」『統計数理』第51巻第2号, pp. 373-388.
 山口幸三(2008)「政府統計の個票利用と統計法改正——試行的提供の経験を踏まえて——」『経済研究』第59巻第2号, pp. 139-152.
 Abowd, John and Lane, Julia (2004) "New Approaches to Confidentiality Protection: Synthetic Data, Remote Access and Research Data Centers," in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases*, Springer-Verlag, Berlin, pp. 282-289.
 Anderson, Margo (1988) *The American Census : A Social History*, Yale U. P.
 Barkume, Anthony J. and Spletzer, James R. (2003) "Microdata Access at the U. S. Bureau of Labor Statistics," mimeograph, August 2003.
 Cross, Harold (1953) *The People's Right to Know*, Columbia U. P.
 Davis, Steven; Haltiwanger, John and Schuh (1996) *Job Creation and Destruction*, MIT Press.
 De Wolf, Virginia A. (1995) "Procedures for Researcher Access to Confidential Microdata at the Bureau of Labor Statistics," mimeograph, 1995.

- Dunne, Timothy (2001) "Issues in the Establishment and Management of Secure Research Sites," in P. Doyle, J. I. Lane, J. J. M. Theeuwes, and L. M. Zayatz (eds.) *Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies* (Amsterdam: Elsevier), 297-314.
- Flaherty, David H. (1979) *Privacy and Government Data bank: An International Perspective*, London: Mansell Publishing.
- McGuckin, Robert and Pascoe, George (1988) "The Longitudinal Research Database: Status and Research Possibilities," *Survey of Current Business*, Nov. 1988; Vol. 68, No. 11, pp. 30-36.
- McGuckin, Robert and Nguyen, Sang (1990) "Public Use Microdata: Disclosure and Usefulness," *Journal of Economic and Social Measurement*, Vol. 16, No. 1, pp. 19-39.
- Reznek, Arnold and Nucci, Alfred (2000) "Protecting Confidential Data at Restricted Access Sites: Census Bureau Research Data Centers," *Of Significance: A Topical Journal of the Association of Public Data Users*, Vol. 2, No. 1, pp. 42-50.
- Ruggles, Steven (2000) "Foreword—A Data User's Perspective on Confidentiality," *Of Significance: A Topical Journal of the Association of Public Data Users*, Vol. 2, No. 1, pp. 1-5.
- Seastrom, Marilyn M. (2001) "Licensing," in P. Doyle, J. I. Lane, J. J. M. Theeuwes, and L. M. Zayatz (eds.) *Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies* (Amsterdam: Elsevier), pp. 279-296.
- Zarate, Alvan; Hoy, Easley; Stamas, George and Therriault, Gene D. (2000) "Disclosure Review Board of Federal Agencies: Characteristics, Defining Qualities and Generalizability," *Proceedings of the Section on Government Statistics*, American Statistical Association.

付録 1: 条文

付録 1-1: Title 13 U. S. C., Sec. 9:

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison may

(1) use the information furnished under the revisions of this title for any purpose other than the statistical purposes for which it is supplied; or

(2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or

(3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports. No department, bureau, agency, officer, or

employee of the Government, except the Secretary in carrying out the purposes of this title, shall require, for any reason, copies of census reports which have been retained by any such establishment or individual. Copies of census reports which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.

付録 1-2(a): Statistics Act (Canada)

3. There shall continue to be a statistics bureau under the Minister, to be known as Statistics Canada, the duties of which are

(a) to collect, compile, analyse, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people;

(b) to collaborate with departments of government in the collection, compilation and publication of statistical information, including statistics derived from the activities of those departments;

(c) to take the census of population of Canada and the census of agriculture of Canada as provided in this Act;

(d) to promote the avoidance of duplication in the information collected by departments of government; and

(e) generally, to promote and develop integrated social and economic statistics pertaining to the whole of Canada and to each of the provinces thereof and to coordinate plans for the integration of those statistics. 1970-71-72, c. 15, s. 3

付録 1-2(b): Statistics Act (Canada)

17. (1) Except for the purpose of communicating information in accordance with any conditions of an agreement made under section 11 or 12 and except for the purposes of a prosecution under this Act but subject to this section,

(a) no person, other than a person employed or deemed to be employed under this Act, and sworn under section 6, shall be permitted to examine any identifiable individual return made for the purposes of this Act; and

(b) no person who has been sworn under section 6 shall disclose or knowingly cause to be disclosed, by any means, any information obtained under this Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual person, business or organization.

付録2： onsite research site 一覧

付録2-1： センサス局 RDC の概要(2006年10月現在)

- ◆設置都市名(施設所属機関)
- 設立年
- 関係研究機関
- センサス局職員の数(バックグラウンド)
- センサス局職員以外のスタッフの数
- ブースの数
- 利用可能時間
- 利用料金
- 年間受諾プロポーザルの数
- 利用者にする大学院生の割合
- ◆Cambridge, MA(National Bureau of Economic Research)
 - 1994年設立
 - National Bureau of Economic Research
 - 1名(PhD in Economics)
 - 0
 - 8
 - 24時間/7日
 - 15000ドル/年
 - 12本程度
 - 約半数
- ◆Berkeley, CA(University of California, Berkeley)
- ◆Los Angeles, CA(University of California, Los Angeles)
 - 1999年設立
 - University of California
- (LA site)
 - 2名(1名はLEHD専門)(PhD in Economics)
 - 0
 - 大部屋ひとつでブースの設置はない。端末数は8
 - 24時間/7日
 - 3125ドル/ブース・月の利用料(週40時間(full-time)利用を想定)が必要。研究プロジェクトのうち典型的なもの(おおむね40%程度)は15000ドル/年程度かかっている。UCLAおよびUCBerkeleyの大学院生は無料。
 - 調査年2~3本。通例では7~8本/年の提出がある。
 - 3分の1
- ◆Upper Marlboro, MD (CES)
- ◆Chicago, IL (Federal Reserve Bank of Chicago)
 - 2002年設立
 - Federal Reserve Bank of Chicago, University of Chicago, the University of Illinois at Chicago,

Argonne National Laboratory, Northwestern University.

- ◆Ann Arbor, MI(University of Michigan)
 - 2002年設立
 - University of Michigan, ICPSR
 - 1人(PhD in Economics)
 - 2人(ミシガン大学の教員, ほかにパートタイマーあり)
 - 8
 - 24時間/7日
 - 15000ドル/年(ミシガン大学の教員・大学院生は無料)
 - 約15
 - 約半分
- ◆New York, NY(Baruch School of Public Affairs),
 - ◆Ithaca, NY(Cornell University)
 - 2006年設立
 - Baruch College, City University of New York, Columbia University, Cornell University, Federal Reserve Bank of New York, Fordham University, National Bureau of Economic Research, New York University, Pace University, Princeton University, Russell Sage Foundation, Rutgers University, State University of New York at Albany, State University of New York at Stony Brook, Yale University
 - 非参加機関の場合, 20000ドル/ブース・年の利用料が必要。
 - ◆Durham, NC(Duke University)
 - 2000年設立
 - Duke University, University of North Carolina at Chapel Hill, NC State University.
 - 1人(PhD in Economics)
 - 1人(Professor, Duke University)
 - 8
 - 24時間/7日(Administrator 滞在時間は40~50時間/週)
 - 15000ドル/年(関係大学の教員・大学院生は無料)
 - 5~6
 - 現在稼働中の14プログラムのうち, 4プログラムが大学院生によるもの。人数ベースでは31人中13人が大学院生。
- 付録2-2：カナダRDCの概要(2007年1月現在, 本所のみ)
 - ◆設置名(設置場所)
 - 設置場所以外の関係機関
 - 設置より2006年10月31日までの間の受付プロポーザル総数
 - 完了研究数
 - 2006年10月31日現在活動中の研究数, 2006年10月31日現在活動中の研究者数, 学生数
 - ◆The Atlantic Research Data Centre(Dalhousie University)
 - なし
 - 95本
 - 22本
 - 31本, 21人, 5人
 - ◆Research Data Centre at the Canadian Research

Institute for Social Policy (University of New Brunswick)

- >なし
- >101本
- >20本
- >36本, 13人, 5人

◆Quebec Inter-university Centre for Social Statistics (QICSS) (University of Montréal)

- >Laval University, McGill University, Sherbrooke University, Université du Québec à Montréal (UQAM)
- >276本
- >36本
- >154本, 138人, 110人

●Carleton, Ottawa, Outaouais Local Research Data Centre(COOL RDC) (University of Ottawa)

- >Carleton University
- >34本
- >N. A.
- >24本, 14人, 13人

●Queen's Research Data Centre(Queen's University)

- >なし
- >19本
- >N. A.
- >10本, 7人, 5人

●Toronto Regional Statistics Canada RDC(University of Toronto)

- >なし
- >326本
- >84本
- >133本, 109人, 60人

●McMaster University Research Data Centre (McMaster University)

- >なし
- >174本
- >34本
- >78本, 37人, 22人

●South Western Ontario Research Data Centre(University of Waterloo)

- >なし
- >94本
- >15本
- >44本, 22人, 25人

●University of Western Ontario Research Data Centre(University of Western Ontario)

- >なし
- >45本
- >3本
- >28本, 19人, 12人

●Manitoba Research Data Centre(University of Manitoba)

- >なし
- >39本
- >3本
- >29本, 38人, 13人

●University of Alberta Research Data Centre(University of Alberta)

- >なし
- >87本
- >23本
- >40本, 39人, 19人

●Prairie Regional Research Data Centre(University of Calgary)

- >なし
- >75本
- >23本
- >39本, 28人, 11人

●British Columbia Inter-university Research Data Centre(University of British Columbia)

- >University of Victoria
- >169本
- >26本
- >76本, 49人, 33人

●Federal Research Data Centre(FRDC)

- >なし
- >135本
- >11本
- >42本, 42人, 1人