

資料1

現地ヒアリング調査(ワシントンDC、オタワ)報告

2013年3月26日

株式会社三菱総合研究所

報告の流れ

本日のご報告の流れは以下の通りです。

- ◆ 調査スケジュール
- ◆ 調査先機関の特徴
- ◆ 各機関におけるデータの取り扱い
- ◆ 3機関に共通する考え方

ヒアリングスケジュール(日程は現地時間)

2/18 アメリカ保健統計センター(NCHS)

- ◆ RDC、リモート集計、パブリックユースファイルを提供
- ◆ ヒアリングはRDCが中心
- ◆ システム設置環境の見学も実施

2/19~20 アメリカセンサス局(US Census)

- ◆ RDC、パブリックユースファイルを提供
- ◆ 2日間のヒアリングを実施したが、話題の中心は秘匿処理手法
- ◆ その他として、初日は情報開示の考え方、や構想検討中のリモート集計サービスの現状、2日目はRDCの運営形態や新たな取り組みについて調査(2日間で10セッション)
- ◆ システムの現物は見学できなかったが、近年の取り組みとしてタブレットやスマートフォン向けのパブリックユースファイル公開システムを開発している部隊なども調査

2/22 カナダ統計局(StatCan)

- ◆ RDC、リモート集計、パブリックユースファイル、オーダーメイド集計を提供
- ◆ さらに新サービスであるビジネスデータの分析支援についてもヒアリング
- ◆ リモート集計サービスについては昨年本格稼働を開始した新システムについても調査

各機関の特性等

NCHS

- ◆機関としてのセキュリティチェックが非常に厳しい。これは組織としてCDCに属してはいるものの保健福祉省の業務に近く、また統計データを持つことからランクはCDCより高いことにも起因すると思われる。(実際に面談者の制服は保健福祉省のものであった)
- ◆ただ、陣容は豊富ではなく、リサーチ部門全体でスタッフは16人(うち3人はIT担当)という限られたリソースの中でサービスを提供している。
- ◆結果、インフラはシンプルでセンターサーバーも可動式ハーフラックに収まる規模(その他、フルラックの2/3程度のストレージ機器)であり、そこに収められたデータについても業務処理上の負荷低減のために所有権を持たず、バックアップをメディアに保管するといったことを実施していない。

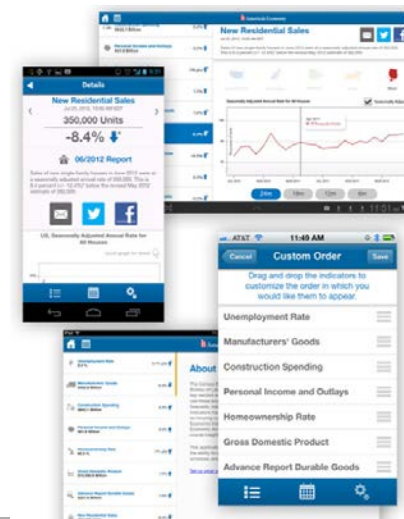


各機関の特性等

USセンサス局

- ◆ 研究リソースが豊富。今回訪問した本局で5000～6000人が就業しており、その多くが統計技術の研究に携わっているとのこと。
- ◆ 各分野に精通したプロ集団というイメージがあり、今回ヒアリングしたメンバーも「マイクロデータの秘匿性技術」「地理情報の秘匿基準」「インターネット上のデータからのRe-identification」「パブリックユースファイルの取り扱い」といった個々のテーマの専門家であった。
- ◆ 基本的には人口動態関係のデータ(10年センサス、ACSとそれに準ずる調査)を持ちRDCを運営する主体であるので、それらについてはセンサス局の範疇で検討・改善等がなされている。
- ◆ しかし、健康・保険関係データ(NCHS所管)と納税・所得関係データ(IRS所管)は所管外で、それらとのリンケージなどはRDCで行われるものの、国のデータ全体に関する意識は低め(カナダ統計局対比)。
- ◆ リモート集計機能の提供が遅れているのも、こういった組織特性が影響している可能性あり(ユーザー側ニーズは影響力が弱い印象)。

- ◆ 独自の新たな研究も行われており、PUFの公開用スマートフォン・タブレット用アプリケーションの独自開発なども進んでいる。



各機関の特性等

カナダ統計局

- ◆全てのデータの所管元という意識がある(実際には健康関係データは各州に帰属するが、その処理をRDCで行う際には主体的に参画する、など)ように思われ、自己の組織収支に対する責任も強いと思われる。
- ◆逆にRDCの運営そのものは大学に依るところが大きいなど、予算と執行責任が近接な関係にある。
- ◆結果、利用料の支払い元であるユーザーやRDCの予算を握る大学からのニーズ・意見には積極的に対応する流れがあるようで、RDAからRTRAという新システムへの切り替えなども進んでいると思われる。
- ◆ただ、やはり基本はUSセンサス局と同じく研究者チームであり、先般センサスのロングフォームを首相が廃止したことに対して統計局長が調査意義の見地から抗議辞任するといった事象も発生している。
- ◆オーダーメイド集計に当たる「Census Custom tabulations」においては、地域オフィスに営業担当者を置くといった民間リサーチ会社のような活動体系もある。



- ◆また、研究目的に限定するものの、ビジネスに関するマイクロデータの分析支援サービスを昨年立ち上げるなど、新たな分野へも進出している。

各機関におけるデータ公開手法の取り扱い比較

		保健統計センター	USセンサス局	カナダ統計局
RDC	センターの運営	センサス局が主体となって運営し、RDCの職員も局から派遣する (NCHSはデータセットを供給)		主体は大学であり、設置大学でコストも手配する (RDC職員給与も大学負担)
	システム構成	センターサーバーにすべてのデータが保存されており、ユーザーは各RDC設置のThin-Client端末でデータを参照・操作する		RDCごとにサーバーがあり、必要なデータセットのみをコピーして配備
	宣誓要否	不要 (個別契約にて対応)	要 (罰則適用のため)	要 (みなし職員とすることで責任を明確化するための措置)
	利用料	運営費の5~10%程度 ハードウェア購入費にしている	運営費を賄うに遠く及ばない 対ユーザー牽制機能的意義	大学にて収支管理 必要に応じて補助金も申請
リモート	方式	プログラム送付型 (レアケースとしてスタッフ支援型もあり)	(Microdata Analysis Systemを構想中)	RTRA
	データリリース チェック	自動アルゴリズムで実施しつつ、 10%のみスタッフチェックに回る	(極力自動化の構想)	全自動処理で、早ければ依頼から1 時間程度で結果返答
PUF	秘匿化手法	ノイズ、カテゴリ統合、トップ(ボトム)コーディングなど	ノイズ、合成 スワッピングは限られたデータ のみに適用(大量の表を公開する ための措置)	スワッピングは行わない (データ改ざん行為に準じるという考 え方)(※利用に制限があり、日本 の匿名データに相当)
その他		RDCデータの正式なバックアップは実施 しない(データの所有権の問題)	近年ではiOSアプリを用いたPUD公開も 開始	RDCのThin-Client化も検討されているが、 これまでの大学主導体制からの変革が課題

3機関で共通の考え方

RDC関連

- ◆ 利用者はまずはパブリックユースファイルでの対応を十分に検討し、どうしても不足する場合により詳細なデータ参照のためにRDCの使用申請をする、という考え方に基づいている
- ◆ RDCの利用申請に際しては
 - 「目的」
 - 「必要なデータセット」
 - 「パブリックユースファイルでは足りない理由」を明記したプロポーザルを提出しなければならない(適宜、RDCからの支援もあり)
- ◆ 数週間以上の審査があり、承認されたのちに必要な教育訓練を経て利用者にRDCが解放される
- ◆ RDCで利用者が作成した集計作業結果については必要最低限なもののみ、各機関のDRB(Disclosure Review Board)の承認を経て提供される
- ◆ RDC利用の場合、情報漏えいを起こした利用者には罰金・禁固刑も含めた罰則があり(カナダの場合は契約に基づく訴追)、いずれの場合にも二度とRDCは利用できなくなる

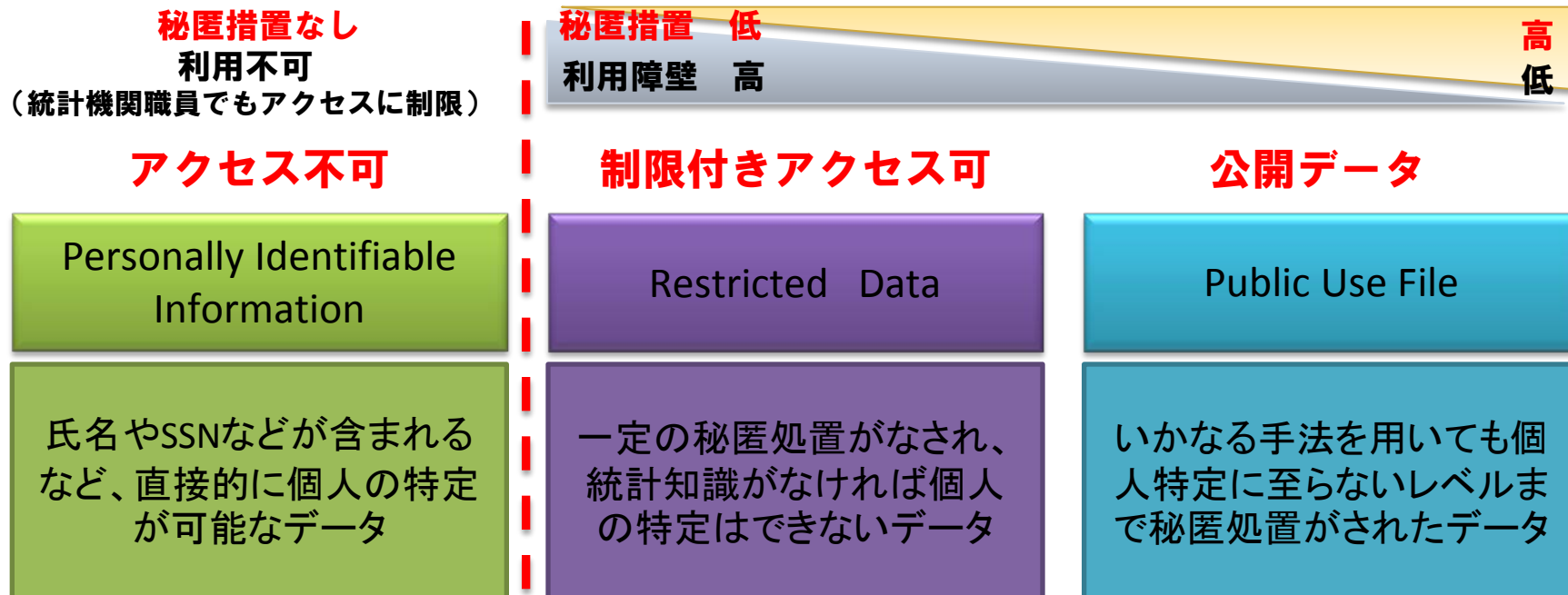
リモート集計関連

- ◆ ユーザーから見た場合、パブリックユースデータで足りない場合には数か月単位の事前準備期間やコストを必要とするRDC利用を申請しなければならず障壁が高い。この間を埋めるものとしてリモート集計が位置づけられる
 - NCHS リモート集計
 - USセンサス局 Microdata Analysis System(構想中)
 - カナダ統計局 RTRA
- ◆ リモート集計は極力データセットの準備やリリースチェックを自動化し、運営負荷を下げることで時間的・金銭的なユーザー負担を低減することを目的としている

パブリックユースファイル関連

- ◆ パブリックユースファイルはその特性上、データ品質に影響が出るとしても秘匿性を第一に考えた処理が行われる
- ◆ 重要視されるポイントは「地理情報」あまり小さな母集団ができてしまうと個人特定が容易になってしまうため、各機関とも一定の基準以下の母集団となった地域は自動検出して対応処理を施す設定をしている(州レベルの地域設定でも個人特定できるケースはあるという前提)

データの秘匿レベルと公開手法について



RDCの利用申請

リモート集計サービス

インターネット上での公開も含めて利用可能

RDCとPUFの間を埋める位置付け

プロポーザルの作成・提出

統計機関での審査

プロポ承認
研修受講

集計作業

開示申請・
審査データ
提供