

オーストラリアのオンデマンド集計に関する調査研究

NSTAC

Working Paper No.33

平成 28 年 8 月

独立行政法人 統計センター

製表技術参考資料は、独立行政法人 統計センターの職員がその業務に関連して行った製表技術に関する研究の結果を紹介するためのものである。

ただし、本資料に示された見解は、執筆者の個人的見解である。

目 次

要旨	1
1 本研究の背景と目的	3
2 オーストラリアにおけるオンデマンド方式による統計作成の実態	4
2-1 オーストラリアにおけるオンデマンド集計の状況	4
2-2 オーストラリアのオンデマンド方式による統計作成	6
3 オーストラリアの TableBuilder における法的根拠について	11
4 オーストラリアの TableBuilder における秘匿処理について	14
5 オーストラリアの TableBuilder における運用面の特徴	21
6 むすびにかえて - オーストラリアの TableBuilder における今後の方向	25
参考文献	26
付録1 “Census and Statistics Act 1905” <抜粋>	27
付録2 pTable の作成方法	30

オーストラリアのオンデマンド集計に関する調査研究

谷道正太郎*、伊藤伸介**、小島健一***

要旨

諸外国の政府統計(公的統計)のマイクロデータにおいては、①匿名化マイクロデータ(個票データに匿名化処理が施されたデータ)の提供、②個票データの提供、③オーダーメイド集計、④オンデマンド型の提供サービス(リモート集計)といった様々な形態による提供が進められてきた。一方、我が国においては、利用者が調査項目(表頭・表側)を選択するだけで集計結果を自動的に出力する、オンデマンド型の集計機能サービスに関する研究が現在進められている。平成26年3月25日に閣議決定された第Ⅱ期「公的統計の整備に関する基本的な計画」では、統計データの有効活用の推進を図る観点から、オンデマンドによる統計作成に関する技術的検証等について実用化に向けた検討を行うことが言及されており、オンデマンド型の集計機能サービスについても、その実用性に関する議論が展開されている。

諸外国においては、政府統計データの提供形態の1つとして、オンデマンドによる集計システムが展開されている国が存在する。このような諸外国におけるオンデマンド集計システムの安全性(調査客体の匿名性)の確保とニーズへの対応を検討するために、独立行政法人統計センターは、「政府統計オンデマンド集計等に関する調査研究」を実施することとした。本稿では、「政府統計オンデマンド集計等に関する調査研究」の成果に基づき、オーストラリアを対象としてオンデマンド集計システムの特徴を明らかにしている。

オーストラリアでは、**Table Builder** というオンデマンド集計システムが開発されている。**Table Builder** では、インターネットを介してシステムにアクセスし、調査項目を指定することによって、自動集計された集計結果表をリアルタイムに受け取ることができる。本研究では、オーストラリア統計局が開発したオンデマンド集計システムである **TableBuilder** の政府統計の二次的利用における位置付けを捉えるために、**TableBuilder** の法的根拠と秘匿処理の方法、さらに運用面での特徴を考察した。また、オンデマンド集計では、利用者のリクエストに応じて集計表を提供することになるため、集計表に含まれるセルの秘匿性をどのようにして確保するかが求められる。そのために、秘匿処理の方法として、出力された集計結果表のすべてのセルにノイズが導入されていることは、**Table Builder** における重要な特徴として指摘することができよう。

* (独) 統計センター経営審議室

** (独) 統計センター統計情報・技術部統計技術研究課非常勤研究員(中央大学経済学部准教授)

*** 株式会社 野村総合研究所 ICT・メディア産業コンサルティング部

オーストラリアのオンデマンド集計に関する調査研究

谷道正太郎、伊藤伸介、小島健一

1. 本研究の背景と目的

諸外国の政府統計(公的統計)のマイクロデータにおいては、①匿名化マイクロデータ(個票データに匿名化処理が施されたデータ)の提供、②個票データの提供、③オーダーメイド集計、④オンデマンド型の提供サービス(リモート集計)といった様々な形態による提供が進められてきた(伊藤(2016a))。一方、我が国の統計法は、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るため、公的統計の整備に関する基本的な計画を策定することや、統計データの二次的利用を促進すること等を踏まえ、2007年に改正され、2009年に全面施行された。現行の統計法では、学術研究等を主要な対象として、調査票情報の提供(統計法33条)、オーダーメイド集計(統計法34条)、匿名データの提供(統計法36条)等の二次的利用が可能となっている。

その中で、わが国においては、利用者が調査項目(表頭・表側)を選択するだけで集計結果を自動的に出力する、オンデマンド型の集計機能サービスに関する研究が現在進められている。平成26年3月25日に閣議決定された第Ⅱ期「公的統計の整備に関する基本的な計画」では、統計データの有効活用の推進を図る観点から、オンデマンドによる統計作成に関する技術的検証等について実用化に向けた検討を行うことが言及されており、オンデマンド型の集計機能サービスについても、その実用性に関する議論が展開されている。さらに、平成27年7月に開催された「第20回統計データの二次的利用促進に関する研究会」では、「統計データの二次的利用の検討状況(全体像・イメージ)」が提案され、オーダーメイド集計に関しては、「利用制限の緩和、『オンデマンド集計』の実用化」が今後の課題として指摘されている。

諸外国においては、政府統計データの提供形態の1つとして、オンデマンドによる集計システムが展開されている国が存在する。その中に、オーストラリアで開発された Table Builder がある。Table Builder では、インターネットを介してシステムにアクセスし、調査項目を指定することによって、自動集計された集計結果表をリアルタイムに受け取ることができる。このような諸外国におけるオンデマンド集計システムの安全性(調査客体の匿名性)の確保とニーズへの対応を検討するために、独立行政法人統計センターは、「政府統計オンデマンド集計等に関する調査研究」を実施することとした¹。

¹ 本調査研究においては、法制度、技術、運用の3つの観点を幅広く調査し、わが国でオンデマンド集計を実施する際の課題を多面的に明らかにすることを目的としていることから、事前に Web や文献、論文等の公開情報を調査した上で、諸外国に実際に聴取またはアンケートすべき項目が何かを検討した。その上で、オーストラリアについては、2016年2月22日～24日の3日間で、オーストラリア統計局を訪問

本稿では、「政府統計オンデマンド集計等に関する調査研究」の成果に基づき、オーストラリアを対象としてオンデマンド集計システムの特徴を明らかにすることによって、オンデマンド集計システムにおけるその可能性を追究することにしたい。

2. オーストラリアにおけるオンデマンド方式による統計作成の実態

2-1 オーストラリアにおけるオンデマンド集計の状況

「オンデマンド集計」は、利用者の要求に応じた集計表の作成及び提供として捉えられる。表1は、諸外国における政府統計の二次的利用について類型化を行ったものである。表1を見ると、統計データの二次的利用の種類に関しては、個票データへの直接的なアクセスの可能性および分析結果の提供形態によって、類別化が可能である。政府統計の個票データの二次的利用については、直接利用型、プログラム送付集計型とオーダーメイド型に大別される。そして、オーダーメイド型でリアルタイム提供型に整理される手法が、本研究で検討しているオンデマンド集計の方向性に近いものと考えられる。

政府統計においては、集計結果表が公表される場合、調査客体が特定されないような秘匿措置が求められるが、一般には、結果表が詳細になるにつれ、また、関連する結果表が増えるにしたがって、秘匿の度合いや困難さは増大する。ゆえに、オンデマンド集計システムにおいて、利用者の様々な要求に応じて集計表を作成・提供するためには、作成された集計表の秘匿に関する十分な検討が求められる。

一方、諸外国では、結果表の有用性と秘匿性のトレードオフ関係を何らかの方法で克服し、オンデマンド集計の提供を実現させている国が存在する。例えば、オランダ、オーストラリアが知られているが、特にオーストラリアは先進的な取組みを行っていることが確認されている。また、Andersson *et al.* (2015) によると、スウェーデンがオーストラリアで運用されている方法を試みていることが紹介されている。

そこで、本研究では、オンデマンド方式による統計作成・提供を実現しているオーストラリアを対象とした調査研究を行うことによって、オンデマンド方式による統計作成について、その技術的な側面を明らかにするだけでなく、実現に寄与していると思われる外的要因についても考察を行う。

して聴取を行った。オーストラリア統計局については、以下の担当者（役職は、2016年2月当時）に対してヒアリング調査を行った。記して謝意を表したい。

- John D Moore(Director, Customised and Microdata Delivery)
- Michelle Gifford(Acting Director, Customised and Microdata Delivery)
- Kristine Creek(Customised and Microdata Delivery)
- Todd Sweeting(Acting Director, Microdata Access Strategies)
- Phillips Wise(Director, Statistical Risk and Quality Assurance)
- Victoria Leaver(Assistant Director, Business Statistics Methodology)

表1 諸外国の政府統計の二次的利用の種類

二次的利用の種類		特徴	状況
直接利用型	オンサイト型	● 調査実施者の指定する施設内において、調査実施者が提供するデータを利用することができる	わが国において、既に実施済み
	直接提供型	● 利用者の研究室等、特定の施設以外の場所で、一定の条件のもとに調査実施者が提供するデータを利用することができる	
	リモート提供型 (リモートアクセス)	● リモートではあるが、直接統計局等のサーバにアクセスし、データ分析等を行い、結果表を受け取ることができる	わが国では実現していない
プログラム送付集計型	参照可能型	● ネットワークを経由して調査実施者が提供するデータを閲覧した上で、作成したプログラムを送付し、集計された結果を受け取ることができる	
	参照不可能型	● ネットワークを経由して調査実施者が提供する参考情報を利用した上で、作成したプログラムを送信し、集計された結果を受け取ることができる	
オーダーメイド型	従来型	● 調査実施者に対して集計の委託を行い、調査実施者によって集計された集計結果票を受け取ることができる	わが国において、既に実施済み
	<u>リアルタイム提供型</u>	● インターネットを介してシステムにアクセスし、集計項目を指定することによって、自動集計された集計結果表をリアルタイムに受け取ることができる	わが国では実現していない

出所) 「諸外国における二次的利用の現状について」(総務省)を一部追記・加工

2-2 オーストラリアのオンデマンド方式による統計作成

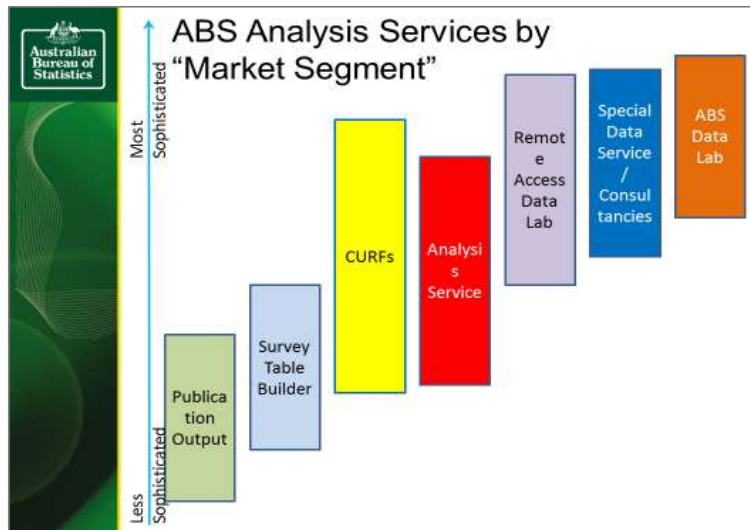
オーストラリアにおける統計の作成及び提供のサービスの全体像は、図1で示されている。オーストラリア統計局は、集計結果表の作成・公表を行うだけでなく、匿名化マイクロデータ (Confidentialised Unit Record Files=CURFs) の作成・提供、さらには、リモートアクセス (Remote Access Data Lab=RADL) による政府統計の個票データの利用サービスやオンサイト施設 (ABS Data Lab) による個票データの利用サービスを行っている。オーストラリアのオンデマンド方式による統計作成については、図1の中で集計結果表 (Publication Output) の隣に位置する「TableBuilder」という集計システムが該当する。このシステムは、マイクロデータから表、グラフを作成するためのオンラインツールである。オーストラリア統計局は、このシステムがテーブルの表頭と表側を選ぶことによって、統計表やマップを「簡単に」作成できるツールであることを強調していた。なお、TableBuilderについては、2009年より、サービスが本格的に開始されている。

TableBuilderで利用可能なサービスは、センサスデータ (Census) とそれ以外のデータ (Others) に関する集計サービスに大別されるが、それは表2のように体系づけられる。基本的な構成は、TableBuilder Pro と呼ばれる有料のサービスと TableBuilder Basic と呼ばれる無料のサービスの2種類であり、アクセス可能な政府統計データもセンサスのデータとそれ以外のデータに分けられる。また、政府統計データの利用可能な年次 (あるいはセンサスの実施時期) によって、異なるサービスが行われる。

このように、サービスの体系は複雑であるが、それがセンサスのデータとそれ以外のデータに分けられる理由は、オーストラリア統計局の統計データを管轄している部局が異なるためであって、秘匿処理の違いといった技術的要素によって分かれているからではない。ただし、ツールの機能から見ると、センサスのデータとそれ以外のデータでは、機能がやや異なっている。センサス向けの TableBuilder では、マッピング機能を使用することが可能であるが、センサス以外の TableBuilder においては、標本調査のデータセットから算出された基本統計量 (ウェイト、RSE、連続データ項目、平均、カスタム範囲、分位点を含む) のみが、算出され、提供されている。

つぎに、表3は、センサスデータを対象とした TableBuilder Pro と TableBuilder Basic に関する特徴をまとめたものである。TableBuilder Pro と TableBuilder Basic の場合、課金をするか否かという大きな違いがあるが、両者のいずれも、センサスの全てのデータを分析対象とすることができる。ただし、利用できる機能と利用可能な項目に違いがあることから、課金の有無が生じていることがわかる。とくに、TableBuilder Basic では、集計表に用いる変数は、予め設定された分類区分のみが利用可能であって、それらの変数を用いて作成された統計表がオンデマンドで提供可能であるのに対して、TableBuilder Pro においては、利用者がデータベースの中から希望する変数や分類区分を探索的に選定した上で、統計表を作成することが可能である。

図1 オーストラリア統計局が提供する統計作成サービスの体系



出所 Elazar (2013)

表2 TableBuilderの種類²³

	センサス	その他のデータ
有料	<ul style="list-style-type: none"> ● TableBuilder Pro data series for Census 2011 data ● TableBuilder Pro data series for Census 2006 data 	<ul style="list-style-type: none"> ● TableBuilder data series for all other TableBuilder data
無料	<ul style="list-style-type: none"> ● Census TableBuilder Basic 2011 ● Census TableBuilder Basic 2006 	—

※TableBuilder Basic 2006 とは、以前、CDATA Online 2006 として提供されていたサービスの後継である。

なお、TableBuilder Basic の場合、国外から利用することも可能であるが、TableBuilder Pro については、国内に居住する者に限定されている。また、TableBuilder Pro の利用者は通常、自身が所属する組織・団体を介して利用の申請を行うが、オーストラリア統計局は、その利用申請の段階で、組織・団体の確認を行っている⁴。したがって、国内外を問わず、組織の構造等の観点から、オーストラリア統計局が TableBuilder Pro の利用組織を判断しているものと推察される。さらに、表3では、TableBuilder のHPで紹介されている Census

2 <http://abs.gov.au/websitedbs/censushome.nsf/home/tablebuilderdetails?opendocument&navpos=240>

3 <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Microdata+prices>

4 <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/registration+centre+-+Answers+to+FAQs#hasorg>

表 3 TableBuilder Pro と TableBuilder Basic の違い⁵

	Basic	Pro
対象者	● センサスの利用経験があり、基本的な統計表を作るスキルを持つ者	● Basic に比べてさらに進んだスキルを持つ者
価格	● 無料	● 有料
対象 DB	● Census Topics の大半	● ほぼ全ての Census Dictionary
操作	● 変数の分類区分が予め設定された統計表の利用が可能	● 利用者のDBやシステムのインポートをし、構築したい統計表の抽出が可能
得られる表の例	<ul style="list-style-type: none"> ● センサスの表の構築・カスタマイズ ● 地図連動をした図表の作成と保存 ● 各種フォーマットでのグラフや図の閲覧、抽出等 	<ul style="list-style-type: none"> ● 異なる地域間のデータ比較 ● 特定のニーズにマッチした、カスタマイズされた地域データの作成と共有 ● 目的に合ったグラフや図表の作成、抽出等

Census Topic	Census Dictionary
<ul style="list-style-type: none"> ● 文化・言語の多様性 ● 教育および学歴 ● 雇用、収入と無償労働 ● 障害者、要支援者 ● 子供、保育 ● 通常の住居、国内の人口移動 ● 人民関連 ● 持ち家の状況 ● 持ち家以外の状況 ● 家族構成・特徴 ● 住居の特徴 ● 世帯収入及び支出 	<ul style="list-style-type: none"> ● 通常の居住に関する継続的な情報 ● 人民に関する継続的な一連の情報 ● 雇用者および勤務地に関する継続的な情報 ● 家族及び通常の居住に関する継続的な情報 ● 居住に関連する継続的な一連の情報

Topic、Census Dictionary の詳細を示している。それによれば、文化、教育、雇用、所得、福祉等、様々な調査項目が集計項目として利用可能であることがわかる。

TableBuilder は、現在、表 4 に示されるような 44 個の統計データを利用することができる。この 44 個の統計データを見ると、その多くが世帯調査に係る政府統計であること、さらには、行政記録に基づく統計も含まれていることがわかる。

利用可能な統計の選定は、計画的になされたものではなく、各統計担当部局の裁量によっ

⁵ <http://abs.gov.au/websitedbs/censushome.nsf/home/tablebuilderdetails?opendocument&navpos=240>

表 4 TableBuilder で利用できる統計 (ABC 順に記載) ⁶

TableBuilder が対象とする統計の種類
Australian Aboriginal and Torres Strait Islander Health Survey, Core Content - Risk Factors and Selected Health Conditions, 2012-13
Australian Aboriginal and Torres Strait Islander Health Survey, Detailed Conditions and Other Health Data, 2012-13
Australian Aboriginal and Torres Strait Islander Health Survey, Nutrition and Physical Activity, 2012-13
Australian Census and Migrants Integrated Dataset, 2011
Australian Census Longitudinal Dataset, 2006-2011
Australian Health Survey, National Health Survey, 2011-12
Australian Health Survey, Core Content - Risk Factors and Selected Health Conditions, 2011-12
Australian Health Survey, Nutrition and Physical Activity, 2011-12
Australian Health Survey, Core Content - Risk Factors and Selected Health Conditions, 2011-12
Barriers & Incentives to Labour Force Participation & Retirement & Retirement Intentions, 2010-11
Census of Population and Housing, 2011
Census of Population and Housing, 2006
Childhood Education and Care, Jun 2014
Childhood Education and Care, Jun 2011
Community Engagement with Nature Conservation, 2011-12
Crime Victimization, 2012-13
Crime Victimization, 2011-12
Crime Victimization, 2010-11
Attendance at Selected Cultural Venues and Events, 2013-14
Participation in Selected Cultural Activities, 2013-14
Participation in Selected Cultural Activities, 2010-11
Disability Ageing and Carers, 2012
Disability Ageing and Carers, 2009
Disability Ageing and Carers, 2003
Education and Work, May 2014
Education and Work, May 2013
Education and Work, May 2012
Education and Work, May 2011
Family Characteristics, 2012-13
General Social Survey, 2014
Household Use of Information Technology, 2010-11
Learning and Work, 2010-11
Characteristics of Recent Migrants, Nov 2013
Characteristics of Recent Migrants, Nov 2010
Census of Motor Vehicles, 2015
Census of Motor Vehicles, 2014
Census of Motor Vehicles, 2013
Patient Experiences in Australia, 2011-12
Preschool Education, 2014
Preschool Education, 2013
Participation in Sport and Physical Recreation, 2013-14
Participation in Sport and Physical Recreation, 2011-12
Work Related Training and Adult Learning, April 2013
Work Related Injuries, 2013-14

⁶ <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Expected+and+available+Microdata>

ている。TableBuilder の担当者は、各統計担当部局に、TableBuilder で利用できる統計データの提供を要請しているが、その判断は、各統計担当部局に委ねられている。最終的には、各部局の判断で、提供される統計データが決定される。その意味では、44 個の統計は、利用者のニーズに対応したものでもなければ、行政側の業務上の利用者に対する還元というものでもないと言える。

TableBuilder によるオンデマンドの集計表の提供については、その大半が、人口・世帯系のデータに基づいている。なお、オーストラリア統計局の担当者によれば、将来的には、度数表 (frequency table) ではなく、量的な (Magnitude type) データを含む経済系のデータの提供を考えているとのことで、現在、そのための調査研究がオーストラリア統計局で行われている。

TableBuilder の利用者数と利用目的は、表 5 で示される⁷。TableBuilder Pro と TableBuilder Basic に関する調査時点 (2016 年 2 月) における利用者数はそれぞれ、約 10,000 人と 85,000 人である。本調査時点においても、利用者は増え続けており、Basic の利用者数は 1 ヶ月に 220 名、Pro は 110 名程度増加しているとのことである。TableBuilder Pro の場合、利用者は、主として研究機関や政府・地方公共団体であり、その利用目的は、地方公共団体の場合、人口や年齢の概略把握であるが、大学・研究機関では教育目的のために利用している。

TableBuilder の利用者が 1 万人規模で存在する理由として、オーストラリア統計局は、TableBuilder が表、グラフやマップを作成するシンプルなツールであるだけでなく、センサスについて言えば、TableBuilder において集計の対象となっている標本の規模が、他の二次的利用に係るサービスと比較して大きいことを指摘している。先述した図 1 の通り、オーストラリア統計局は、二次的利用に係る様々なチャンネルを提供しているが、それによって提供されている政府統計マイクロデータは、標本データであることが少なくない。例えば、人口センサスについては、CURF 等の匿名化マイクロデータには、母集団の 1% 程度のレコード群が含まれているが、Remote Access Data Lab および ABS Data Lab においては、標本抽出率が 5%~10% の人口センサスの個票データにアクセスすることが可能である。その一方で、TableBuilder においては、人口センサスの全数データを対象にした集計を行うことができる⁸。

⁷ 聴取対象によって、利用者数に対する認識が異なっていたが、表 5 では、一例としての回答を記載している。

⁸ オーストラリア統計局は、人口センサスに関する匿名化マイクロデータ、さらには Remote Access Data Lab や ABS Data Lab で利用可能な人口センサスの個票データの標本抽出率を公表していない。今回実施したオーストラリア統計局におけるヒアリング調査によって、人口センサスのマイクロデータの標本抽出率に関する情報が明らかになったものである。

表 5 TableBuilder の利用者数と利用目的

	利用者	利用者のタイプ	利用目的
Census			
TableBuilder Pro	10,000 人 (290 機関)	<ul style="list-style-type: none"> ● 研究機関 (スタッフ・学生) ● 政府・地方公共団体 ※政府機関は、連邦政府、州政府 	<ul style="list-style-type: none"> ● 地方公共団体では人口や年齢の概略把握に利用している ● 研究機関では教育の一環としても利用している ● 大学が教育目的で利用する際、費用面での優位性が効く
TableBuilder Basic	85,000 人	<ul style="list-style-type: none"> ● 多種多様 (民間企業含む) 	<ul style="list-style-type: none"> ● 多種多様 (民間ではマーケティング等に用いるケースもある)
Census 以外	8,000 人 ～ 10,000 人	<ul style="list-style-type: none"> ● 多種多様 	<ul style="list-style-type: none"> ● Census データ以外は、地理的に粗すぎるため、地方政府にはあまり利用されない

3. オーストラリアの TableBuilder における法的根拠について

TableBuilder は、十分に秘匿された集計表を提供するサービスであると考えられており、TableBuilder の直接的な根拠となる法律や内規は存在しないことが知られている。このことは、オーストラリア統計局が提供する他の政府統計の二次的利用サービスと、決定的に異なる点だと指摘することができる。なお、表 6 は、オーストラリア統計局によって行われている TableBuilder を含む各種の政府統計の二次的利用のサービスに関する特徴を示したものである。

その一方で、法制度上では直接的に TableBuilder を規制するものはないが、後述する利用規約 (Terms & Condition) において、TableBuilder の利用上の規定や罰則が定められている。

TableBuilder で扱うデータセットを保護する義務は、オーストラリア統計局全体に課されており、“Census and Statistics Act 1905” で規定されている。付録 1 は、“Census and Statistics Act 1905” の抜粋であるが、法第 12 条の第 1 項では、「統計局長は、本法律に基づいて収集された統計情報を集約 (集計) して分析しなければならず、そしてその集約・分析、又はそれら集計結果や要約 (統計量) を、公表・提供しなくてはならない」ことが明記さ

れている。また、法第 12 条の第 2 項では、「細節(1)に係る集計結果、又は要約は、特定の個人または組織の識別が可能になるようなやり方で公表・提供してはならない」ことが述べられている。これは、TableBuilder においてオンデマンドで作成された集計表における秘匿処理の必要性を表す条文であるだけでなく、オーストラリア統計局で公表されている集計結果表についても、攪乱的手法の適用を含む秘匿処理を必要とする法的根拠となっている。

オーストラリアは、第二次世界大戦時に、欧州からの移民を多く受け入れてきた経緯もあり、政府統計の守秘義務に対して強い規制を強いてきた。とりわけセンサスにおいては、“Census and Statistics Act 1905” で規定された法第 7 条、第 12 条と第 19 条の遵守が、強く求められている。こうした法的根拠に基づいて、TableBuilder による二次的利用のサービスが展開されたと言えなくはないだろうと推察される。

表 6 オーストラリア統計局が提供する政府統計の二次的利用サービスの比較⁹

	TableBuilder	CURF	RADL	ABS Data Lab
概要	ABS ミクロデータを用い、表やグラフを作成するツール	秘匿処理がされた ABS の詳細データの集合ファイル	CURF の詳細なマイクロデータをオンラインでプログラム送付集計できるサービス	ハイレベルなデータを ABS オンラインサイトで分析するサービス
レベル (イメージ)	公表値より分析の範囲は広い	Table Builder より分析の範囲が広い (可能性が大きい)	CURF より精度が分析の範囲が広い	もっとも分析の範囲が広い
利点	結果の秘匿を自動で担保可能 テーブルを csv 等で出力可能	自身の PC で動作させることが可能	詳細な分析が可能 ほとんどの表や分析結果をリモートで取得可能	最も詳細な分類区分を用いて複雑な分析が可能
制約	一部有料 法制限なし、ただし利用規約への同意が必要 利用履歴は ABS に残る	有料 法律の制約あり	有料 法律の制約あり 利用履歴は ABS に残る	有料 法律の制約あり 場所の制約あり 利用履歴は都度消去される

※) この表ではオーストラリア統計局を ABS と略記している。

9
<http://www.abs.gov.au/websitedbs/D3310114.nsf/89a5f3d8684682b6ca256de4002c809b/c00ee824af1f033bca257208007c3bd5!OpenDocument>

なお、センサスに関しては、オーストラリア統計局の役職員に対して厳しい法令遵守が求められている。その内容は、TableBuilder User Manual で紹介されている。その内容の一例を以下に抜粋する。

- 本法制度の対象者は、常勤、非常勤に関わらない。
- 間接的直接的に関わらず、この法のもとで集められた、どのような秘匿情報を漏洩する、現在の職員または過去の職員に対し、違反と見なす。
- 違反のペナルティは、最大 5,000 ドル（マニュアルでは 5,000 ドルと記載されていたが、ウェブサイトでは、13,200 ドルと記載されている）の罰金、二年以内の禁固のどちらか一方、または両方である。
- 違反者は、オーストラリア統計局から懲戒される。
- オーストラリア統計局は Privacy Act 1988 に基づく原則を遵守する義務を負う。

また、センサスに関しては、情報の収集から保存に関するプロセスがマニュアルで詳細に定められることによって、その秘匿性が確保されている。その内容の一例を以下に抜粋する。

(情報の収集に係るプロセスの規定)

- ① センサス収集者は、データフォームを守るために、常時、収集ルートが安全であることに責任を負う。
- ② ABS は、センサスデータが収集者によって見られることを懸念する世帯に対し、①プライバシー封筒の提供、または、②eCensus での回答の選択肢を与える。
- ③ 疑心の強い世帯は、センサス収集者に、データが ABS に到着したことを照明するメールバックを求めることができる。

(収集した情報の処理・保存に係るプロセスの規定)

- ① Collection Center に集められたデータは、Data Processing Center に、安全な(セキュアな)状態で移される。
- ② 常時、警備が、権限の無い人の Processing Center への侵入を防ぐ。
- ③ 処理された後の書類は、オーストラリア統計局職員の立会のもと、溶解処理される。
- ④ 処理後、データとしての保管は、ABS computer system で実施される。日常的にパスワードは変更され、アクセスコントロール、認証の形跡もチェックされる。
- ⑤ データは、暗号化処理をかけられており、eCensus で回答する場合も、オーストラリア統計局と回答者以外の人が、データを見られないように処理を行うことによって、不正アクセスへの対策が取られている。

なお、オーストラリアでは、連邦プライバシーコミッショナー制度が存在するが、連邦プライバシーコミッショナーが、法制度の側面から、直接、政府統計データの秘匿処理の方法に関与してくることは無いとのことである。

4. オーストラリアの TableBuilder における秘匿処理について

先述の通り、“Census and Statistics Act 1905”では、いかなる秘匿情報も公開をさせないように、処理をした上で、結果を公表することが求められている(12 Publication etc. of statistics (2)が直接的に関係するものと推察される)。そのため、TableBuilderは、個票データ(非識別データ、deidentified data)を集計した「生の集計表」に対して、意図的に「ノイズ(攪乱値)を付加する」ことで、データの秘匿性を担保している。したがって、TableBuilderにおける秘匿処理の重要な特徴は、TableBuilderの集計システムによって得られる全ての集計表に対して攪乱的措置が施されていることにある。

オーストラリア統計局は、秘匿性の担保の方法として、「わずか」なノイズ(攪乱値)を入れる手法が、利用者にとって、もっとも満足いく技術であると主張している。この「わずか」な攪乱を行うことによって、真の値から乖離することから、オーストラリア統計局は、TableBuilderの利用に際して、「小さい値やRSE(標本標準誤差)の大きい結果に対しては、そのデータに信頼を置かない(“No reliance should be placed on small number cells or cells with large percentage Relative Standard Error (RSE) values”)」よう、注意喚起している。また、「生の集計表」を作った後でセルにノイズを付与していることから、作成された集計表内のセルの値を積み上げると(「加算性」)、合計値と合わなくなる場合もあることを注意喚起している(加算性については、後述で詳細に説明する)。

以下では、この攪乱をどのように行っているのかについて、説明を行うことにしたい。

わずかとは言え、ノイズを入れることで、分析の対象となる統計表の中の数値が大幅に変わった場合、利用者にとっての情報損失は小さくないと思われる。特に、表の平均値がずれたり、表のデータの分散・共分散構造が崩れたり、また、「0」という結果が「0でない」結果となってしまったりすることは、いずれも、分析対象となる集計表の価値を著しく減じることが考えられる。したがって、オーストラリア統計局は、以下を前提とした攪乱的手法を適用している。

- ① ノイズ付与に用いる数値の平均値は0である。
- ② 真の結果が0である場合、その値にはノイズを入れない。
- ③ 標本調査についても、ウェイトやウェイトをかける前の値にノイズを入れるのではなく、あくまで集計表に対してノイズを入れる(悉皆調査と標本調査で、ノイズ(攪乱値)の入れ方を極力変えない)。
- ④ 同じ分析をし、同じ表を生成する場合は、常に同じノイズ(攪乱値)を加える。

特に第 4 の特徴は、” Differencing ” と呼ばれる秘匿情報へのアクセスに対抗する手段である。攪乱する値が、集計表の生成ごとに異なる場合、何度か集計表を生成し、中身を比較していくことで、集計表の中の小さい数値を判別し、その内容を推測することが可能となる場合もある。この行為に対抗するため、TableBuilder では、同じ集計表を作成した場合には、必ず同じ値のノイズを入れる工夫がなされている。

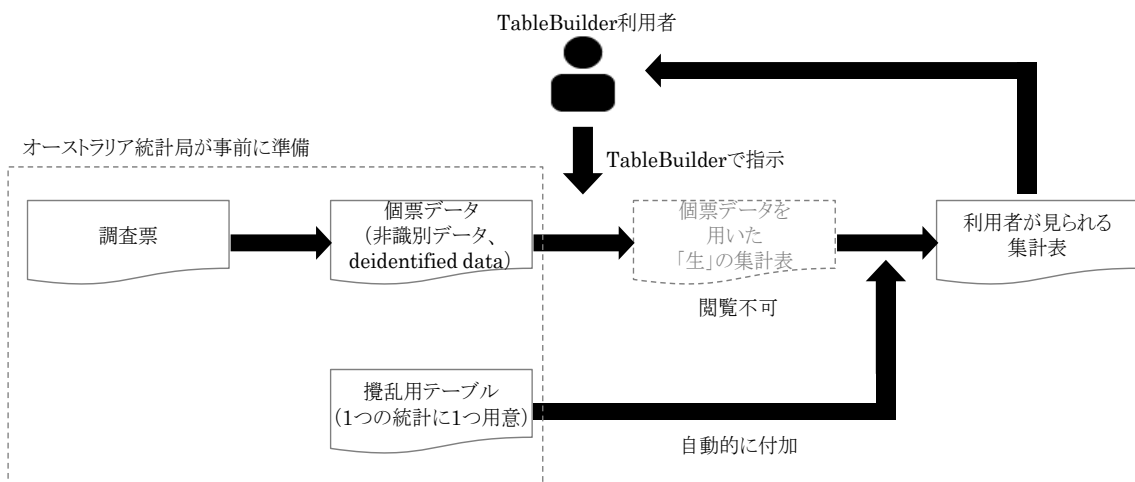
図 2 は、TableBuilder におけるノイズ付与のプロセスとノイズ付与がなされた集計表の提供のプロセスをまとめたものである。以下に、それぞれの特徴を述べる。

最初に、TableBuilder に用いられるデータであるが、このデータについては、TableBuilder Pro、TableBuilder Basic のいずれも、調査票情報から名前や住所等の直接的な識別子を削除した非識別化 (deidentified) されたマイクロデータを用いている点は、1 つの特徴だと言える。

個票データ (非識別データ、deidentified data) 自体には、ノイズやスワッピング等の処理はされておらず、また、データキューブのような中間集計表も用いられていない。従って、幾つかのデータセットは非常に詳細なデータ項目を有する場合もある。ただし、TableBuilder では、最終的な生成物を得る前に、攪乱を入れているために、安全性に問題は無いとしている。

つぎに、攪乱に用いる表 (オーストラリア統計局はこの表を、pTable (Perturbation Table) と呼んでいる) は、TableBuilder によって作成される集計表が利用者に提供される前に、事前に用意されており、1 つの統計に 1 つの pTable が設定される。pTable は、255 行×最大 30 列の行列から構成される。この pTable の作成には、熟達した技能と経験が必要となっており、「秘匿性を担保しつつも」、「政府統計の収集からなるべく早期に TableBuilder でのサービス提供ができるよう、過去の経験を活用」した表となっている。実際、秘匿性の十

図 2 TableBuilder におけるノイズ付与のプロセス



分な確保を考慮しないのであれば、オーストラリア統計局は、半日程度で自動的に（プログラム等によって）pTable を作成することが可能としている。もっとも、様々な角度からpTable が秘匿性に問題が無いかを検証し、もし問題があれば、pTable を再度作成するという作業を人手で行っていることから、実際には数週間を経て、作成が完了した集計表が利用者に提供される場合もある。

このようなpTable の作成方法については、今回のヒアリング調査に基づき、統計実務の観点から以下の5点を指摘することができる。

- ① pTable の波形は、0 を基軸に左右対称なものである（ただし正規分布とは異なる模様）。
- ② 左右対称であるため、pTable には0 も含まれること（攪乱と言っても、必ずなんらかしらのノイズを入れるのではなく、0 が付与される（全く攪乱がされない）ケースもありうる）。
- ③ ノイズの最大値、最小値は、経験則を基に決定している。
- ④ 人力で修正する作業を、50～60 回程度の試行錯誤によって達成している。
- ⑤ 経験則に則って pTable を作成しているため、この攪乱的措置が、どの程度、もとの政府統計の情報に損失を与えるのか、どの程度、秘匿に係るリスクが内在したままであるのかに関する評価は実施されていない。

上記の内容は、2006 年に、Fraser and Wooten(2006)で論じられた TableBuilder に関する内容とは、大きく異なるものとなっている。論文の中では、pTable は、攪乱の効果を最大限にするために、数理的かつコンピューターを用いた数値計算によって決定されていた¹⁰。オーストラリア統計局は、この記述と実態の違いを基礎研究と実務との違いと説明しており、論文に記載した内容は、基礎研究として、あくまで学識者等との「コミュニケーション」のためのツールだとオーストラリア統計局の担当者は認識している。しかしながら、統計実務の面から見ると、pTable の作成方法に関するこのような基礎研究と実務での適用可能性を勘案した上で、pTable を作成していることが考えられる。

最後に、オンデマンドで作成された表に、pTable に含まれるノイズを付与するために、参照用の Record Key を使っている点が攪乱的措置の特徴として指摘される。先述のように、pTable は、255 行×最大 30 列の行列で構成されるが、その中からどの数字を、どのように参照するかに関しては、Record Key が用いられている。Record Key とは、個票データ（非識別データ）の 1 レコードごとに割り振られた値であり、²³²の数字の中からランダムに付与された値である。この値は、1つのマイクロデータに対して一意に決められることから、途中で変わることは無い。利用者が TableBuilder に集計表作成のための指示を行うと、TableBuilder 上で、集計表の中のセルの数値を見ながら、Record Key との対応を図り、こ

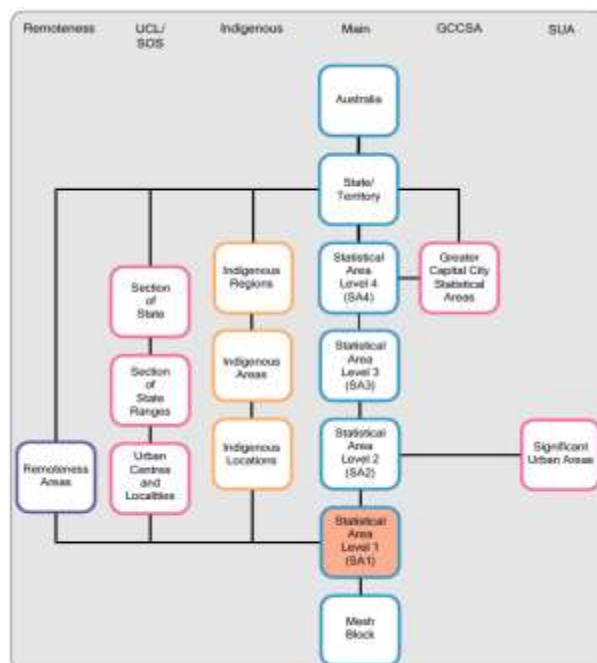
¹⁰ ptable の作成方法に関する数理的な説明については、付録 2 を参照されたい。

の Record Key に対して数学的処理を行うことによって、pTable のどのセルを参照すればよいのかが決定される¹¹。これによって、同じ統計表を生成する場合は、常に同一のノイズ(攪乱値、pTable のセル値)を加えることが担保される。

このような手法を用いて作られた、ノイズが付与された集計表は、秘匿上の安全性が高いとみなされ、実運用上では、オーストラリア統計局職員等の人手のチェックを介することなく、利用者に提供されている。利用者は、表頭、表側を自由に選ぶことができる。それによって、TableBuilder においては、任意の粒度で分析することも可能である。多次元クロス集計を、探索的に実施できる点は、TableBuilder の重要な特徴と言えるだろう。ただし、実運用から見ると、6~7次元のクロス集計表は、0 が非常に多く入ることから、そうした集計表の提供は困難なものと言える。

一方、地図へのマッピングについても、粒度の細かいレベルでの分析も可能であるが、オーストラリア統計局は安全性が担保されているとしている(もっとも、ノイズ(攪乱値)の付与だけでは秘匿が担保できない、または、0しか入らない場合には、そもそものサービスとして成立しないため、利用者に提供をしない(表頭・表側を選択させる際に、そもそも選択肢に載せない))ようにしている)。なお、地図の粒度は、オーストラリア統計地理標準(Australian Statistical Geography Standard=ASGS)に準拠しており、最も細かい粒度(Statistical Area Level 1=SA1)であれば、オーストラリアの54,805地区すべてに関する詳細な地域区分について地図上での分析が可能である(図3参照)。

図3 オーストラリアにおける地図の粒度の区分 (ASGS)¹²



¹¹ この数学的処理には、高度な数学的な方法が用いられていることから、詳細については、オーストラリア統計局が UNECE 等で発表した、Thompson *et al.*(2013)等の論文を参照されたい。

¹² <http://www.abs.gov.au/websitedbs/censushome.nsf/home/factsheetsas>

秘匿処理に関する技術的な論点として集計表内の数字の整合性（特に加算性）に関しては、以下のように述べることができる(表 7)。

先述の通り、TableBuilder は、各セルに対してノイズを入れる手法を採用しているため、基本的にはセルの合計と、予め提示された合計値とは一致しない。しかし、センサスに対しては、この加算性に係る整合性を担保するような処理がなされている。それによって、TableBuilder Pro および TableBuilder Basic で作られる全ての表は、加算性が担保される。もっとも、加算性は特定の 1 つの集計表の中でのみ担保されることから、例えば上述した SA1 のデータは、SA2 (SA1 よりも粗い粒度) のデータと整合しない。

なお、集計表に対して攪乱を行っている時点で、数字の加算に係る整合性を取る必要はないのではないか、という疑問も生じる。この点については、オーストラリアの国民性とも関係があるものと推察される。オーストラリアの感覚では、最低限、総計のみでもあっていればよいと考えられている可能性も否定できない。

このように、TableBuilder の秘匿処理技術の特徴を明らかにしてきた。一方で、こうした攪乱的処理を用いた統計データの「正確性」については、我が国とは基本的な考え方が異なる可能性もある。オーストラリア統計局に対するヒアリング調査に基づけば、政府統計に対する攪乱的措置の考え方については、つぎの 4 点にまとめることができる。

- ① 利用者の一部は、攪乱の影響を懸念している。ただし、この処理によって、ランダムな誤差が含まれるものの、集計表における有用性を大きく損ねるものではない。
- ② TableBuilder は、全てのセルにノイズを入れている。ノイズを付加されたデータではあるが、ノイズはあまり大きくない。したがって、統計量としての分布特性は、ノイズをいれたとしても変わらない。
- ③ データセットに対して適用される攪乱の値は一定であるため、この数字をもって、「真の値」として報告している。
- ④ もっと言えば、オーストラリアの政府統計は、なんらかしらの形で“Anonymize” (匿名化) されており、公表される政府統計においても適用されている。Spontaneous Recognition (偶発的な個体特定) を避ける処理もしている¹³。

特に、オーストラリアの政府統計が、法 12 条第 2 項などの影響を受けてか、公表されるものも含めて、秘匿化されている点は、TableBuilder をサービスとして成立させるための重要な要件と言えるだろう。オーストラリアでは世帯・人口系の政府統計に対しては、TableBuilder と同様のノイズ(攪乱値)を入れているだけでなく、事業所・企業系の政府統計に対しては、個体情報の秘匿性を確保するために閾値を用いた秘匿処理を行う等、別の手法も導入している。

13 オーストラリア統計局においては、匿名化マイクロデータの作成においても、偶発的な個体特定の考え方が用いられている。具体的には、オーストラリア統計局では、偶発的な個体特定の回避という観点から、CURFs を作成するために、SUDA(=Special Uniques Detection Algorithm)を用いた秘匿処理を行っている。

表7 加算性に関する整理¹⁴

加算の整合性	あり	なし
対象となる統計	<ul style="list-style-type: none"> ● Census のみ 	<ul style="list-style-type: none"> ● Census 以外全て
秘匿処理の仕方	<ul style="list-style-type: none"> ● テーブルの各セルに、ノイズ(攪乱値)を挿入 ● セルの行列の総和が保たれるように、さらに調整を実施 	<ul style="list-style-type: none"> ● テーブルの各セルに、擾乱を挿入
留意点	<ul style="list-style-type: none"> ● 1つの表の中で、データの整合は取れている ● ただし、整合はあくまで1つの表の中に閉じており、調整は、それぞれの表でなされるため、異なる地理レベル等のデータ間での整合は取れていない 	<ul style="list-style-type: none"> ● 表の中で、データの整合が取れていない

つぎに、オンデマンド集計システムの安全性を確保するために、TableBuilder の情報システムの構成について述べることにしたい。TableBuilder の情報システムの構成は図4に示されている。その特徴としては、第1に、TableBuilder が参照しているデータは、表4における”ABS Database”に記載した通り、個票データ(非識別データ、deidentified data)であることが指摘される。ただし、前述の攪乱的な方法を用いることによって、TableBuilder においてはレコードの属性情報は露見されないことから、個人情報(人、世帯、車、ビジネス等)の秘匿性は全て確保される。それゆえ、個人等のプライバシーが漏洩することがなく、Database 層で、TableBuilder に係る特別なセキュリティ対策は実施されていない。同様に、秘匿性が破られることに対する危機管理マニュアル等も作成されていない。

従来、セキュリティ対策は、「侵入者モデル(Intruder Model)」が念頭に置かれていたが、近年では「人間モデル(Human Model)」が重視されるようになり、ヒューマンエラーのリスクを低減する必要性が認識されている¹⁵。

この考え方にに基づき、近年はFive Safes Model(Safe Data, Safe Settings, Safe Project, Safe People, Safe Output)という考え方が提唱されている¹⁶。TableBuilder については、

¹⁴ <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1406.0.55.005~Jun%202013~Main%20Features~Confidentiality~810>

¹⁵ 「侵入者モデル(Intruder Model)」と「人間モデル(Human Model)」については、伊藤(2016c)を参照。

¹⁶ Five Safes Model の概要については、伊藤(2016b, 2016c)を参照されたい。

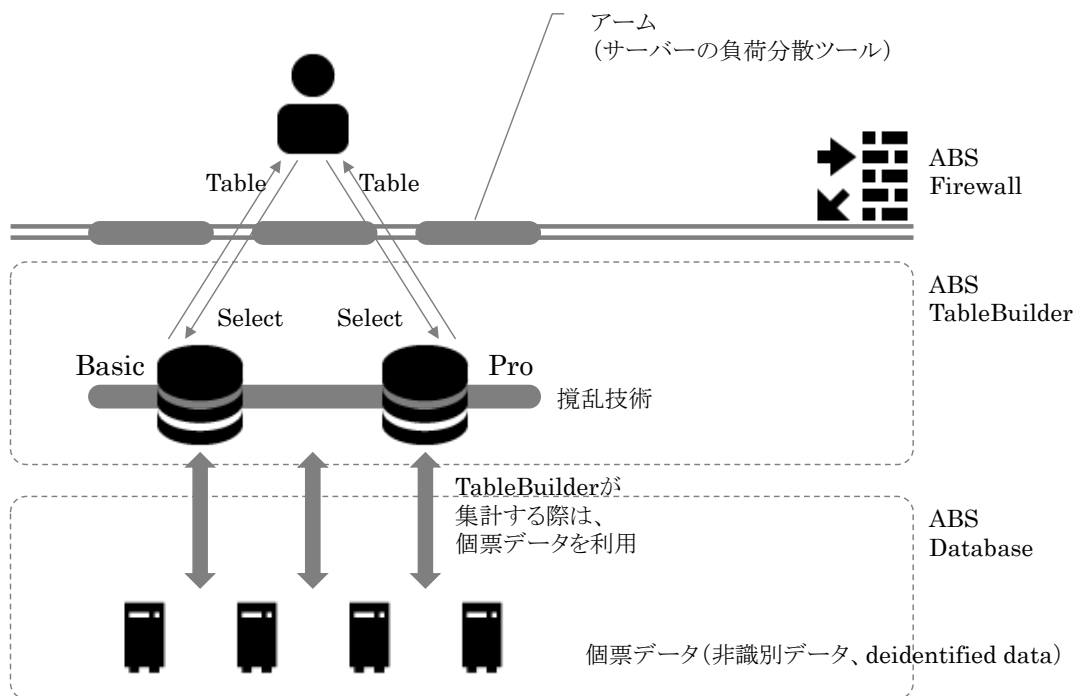
Safe Environment を実現しているとオーストラリア統計局は自認している。

つぎに、システム構成に関する特徴としては、データベースの構成と、TableBuilder のシステム開発において、民間企業のリソースを長らく活用していることが指摘される。この民間企業とは、Space Time Research Pty Ltd. (<https://spacetimeresearch.com/>) であつて、彼らはオーストラリア統計局と 1986 年以來関係を構築してきた。

TableBuilder の攪乱に係るアルゴリズム自体は、オーストラリア統計局によって開発されているが、システムの開発、コーディングおよび、データベース構造の設計・開発は、Space Time Research 社によって行われている。

特にデータベースの設計に関しては、Space Time Research 社の長年のノウハウが活かされている。それは、具体的には、オラクルや SQL サーバのような汎用のデータベース製品ではなく、Space Time Research 社によって独自に開発されているが、その詳細については、オーストラリア統計局自体も把握していない。ただし、政府統計データの高速処理は可能となっており、データ処理の多くは、即時に対応することが可能となっている（マッピング等で区分を細かくした場合には、その限りではない。オーストラリア統計局は、例えば地図上に 500 万のセルを表現しようとした場合には、システム負荷が高く、処理に 24 時間以上かかるものと考えている。よつて、実運用上は、500 万セルを上限にし、それ以上に負荷のかかる処理については、原則として受け付けないことにしている）。

図 4 TableBuilder の情報システムの構成



注 ABS=オーストラリア統計局 (Australia Bureau of Statistics)

このようなシステムにかかる負荷が利用者側にとっての負担とならないように、システム構成には工夫がなされている。具体的には、利用者が重たいジョブをかけることで、サーバを独占してしまうことが無いよう、図 4 に示したアームを 3 つ設け、それによって負荷分散を図っている。

現在、データベースを含め、すべての情報システムをオーストラリア統計局内部のサーバで運用がなされている。しかしながら、局外のサーバで運用することは禁止されていないとのものであり、将来的に安全性が担保されれば、局外で運用することも十分に考えられるということであった。連邦政府で運用するクラウドについては、そのような可能性も高いものと考えられている。

5. オーストラリアの TableBuilder における運用面の特徴

TableBuilder を利用するには、利用者がオーストラリア統計局に登録をすることが求められる。これは、TableBuilder Pro と TableBuilder Basic のいずれにおいても必要である。

利用登録の手続きは、以下のとおりである。最初に、利用者は、オーストラリア統計局のウェブサイトにある Registration Center に、自身で登録を行う。登録の際は、氏名、所属組織、電話番号、住所、メールアドレス、パスワード等の入力が必要である。その上で、オーストラリア統計局より、ユーザーID とパスワードを獲得する。一度、自身のアカウントを作れば、利用者は、全ての無償サービスにアクセスすることができる。なお、利用者が TableBuilder Pro を利用したい場合、利用者が所属する組織を介しての登録が、別途求められる。

つぎに、TableBuilder の利用者は、表 8 に示されるような利用規約を遵守することが求められる。具体的には、ID およびパスワードに関しては、利用者に厳重な管理を求めており、違反者には罰則も設けている。具体的には、TableBuilder の利用者については、本人の ID 以外からの利用は認められておらず、また、自身の ID を他人に貸し出すことも禁じられている。

本人の ID を利用しているか否かの確認は、システムで自動的に行うのではなく、オーストラリア統計局職員が、利用者に直接電話して行っている。なお、電話で確認した結果、利用者以外の人物が TableBuilder を利用していることが判明した場合には、その者に罰則を適用することとしている。

なお、オーストラリア統計局のウェブサイトでは、「当人及び当人の所属する組織が解約されるケースもある」という記載のみがされているが、オーストラリア統計局の担当者に聴取を行ったところ、違反者（初回）に対しては、3 ヶ月の TableBuilder の利用停止、2 回目

表 8 TableBuilder 利用者が遵守すべき規約^{17,18}

名称	内容
Terms & conditions (利用規約)	① 自分の ID、パスワードを他人、または他の組織に提供しないこと。 ② サービス利用においてセーブした図表等は、自身の責任で管理し、オーストラリア統計局は一切関知しないこと。 ③ 組織を介してアカウントを保有する場合、退職等をした際に、それを申し出ること。
ABS Conditions of Sale※	特に著作権について以下の対応をすること ・ TableBuilder の再生産、再販、流通、機能の一部取り出しや他機能との重ね合わせの禁止 ・ Copyright Act 1968 の 40～43 条に当てはまるレベルでの引用、過度な抜書きの禁止 ・ 出所における ABS の明記 ・ ABS の定義した用語の書き換えの禁止 ・ データ分析後のデータに対する「ABS による」との記載の禁止（自身で加工したことの明記）

※ABS Conditions of Sale は、TableBuilder 以外の政府統計の二次的利用にも適用される。

の違反者には 12 ヶ月の利用停止措置を行っているとのことである。なお、3 回以上の違反者については、これまでの運用上、経験はないとのことであった。

TableBuilder の申請有効期間および、価格体系は、表 9 で示されている。TableBuilder のアカウント有効期限は、最大 1 年である。毎年 6 月 30 日に、TableBuilder 内のデータが更新されることから、その時点で有効期限は切れることになっている。したがって、継続の意向を持つ利用者は、再申請を行う必要がある。

そのために、利用者は、最大で 1 年、短い場合には数ヶ月しか TableBuilder を利用することができない。これらの不公正を是正するために、TableBuilder では、期限までの期間に応じた傾斜料金を採用しており、その課金体系は、4 半期によって表 9 のように分類されている。

なお、価格には商品サービス税 (GST) が含まれている。また、組織の規模は、従業員規模で区分されており、従業員 20 名未満が small、200 名未満が medium、それ以上は Large とされている。利用者数が多くなりうる Large には高い課金をしているが、これについては、オーストラリア統計局は、厳密にユーザー ID 数等で管理をしていない。

オーストラリア統計局の課金体系の考え方は、利益を出すことではなく、コストを回収することのみにある (ABS Pricing Policy による)。オーストラリア統計局は、利用者が

17 <http://www.abs.gov.au/websitedbs/d3310114.nsf/home/ABS+Conditions+of+Sale>

18 <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/registration+centre++Answers+to+FAQs#difference>

表9 TableBuilder の価格体系¹⁹²⁰

TableBuilder の種類	価格	組織の規模
TableBuilder Pro data series for Census 2011 and 2006 data	\$1,950	大企業
	\$1,100	中堅～中小企業
TableBuilder data series for all other TableBuilder data	\$975	大企業
	\$550	中堅～中小企業
Census TableBuilder Basic 2011 and 2006 data	無料	-

	価格 (GST を含む)				組織規模
申込時期	7-9月	10-12月	1-3月	4-6月	
利用期間	2015年7月 ～ 2016年6月	2015年10月 ～ 2016年6月	2016年1月 ～ 2016年6月	2016年4月 ～ 2016年6月	
TableBuilder					
TableBuilder pro data series for Census	\$3,900	\$2,925	\$1,950	\$4,875	大企業
	\$2,200	\$1,650	\$1,100	\$2,750	中堅～ 中小
TableBuilder data series for all other TableBuilder data	\$1,950	\$1,465	\$975	\$2,440	大企業
	\$1,100	\$825	\$550	\$1,375	中堅～ 中小

TableBuilder の情報システムへアクセスすることによって発生するシステムの運用及びデータセットのローディングにかかる費用を回収するが、データ収集やファイルの提供にかかる費用の回収は行っていない。

ところで、TableBuilder では攪乱を用いた表を利用者に提供している。この攪乱的措置の結果、どの程度データに信頼をおけるのかに関して、周知することが必要と考えられる。この点について、オーストラリア統計局はつぎのような対応を行っている。

19 <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Microdata+prices>

20 <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/registration+centre++Answers+to+FAQs#sharelogin>

- ① 利用者にはまず、TableBuilder 利用者ガイド (オンラインで提供され、TableBuilder における利用上の制約を記載している) を読むことを要請している。
- ② 政府統計に関する情報は個々の TableBuilder のデータセットに対して提供される。すなわち、どのようにデータが収集されたかや、サンプルやデータアイテムのリスト等の情報が提供される。
- ③ TableBuilder のウェブサイトの至るところに、「小さい値や RSE (標本標準誤差) の大きい結果に対しては、そのデータに信頼を置かない (“No reliance should be placed on small number cells or cells with large percentage Relative Standard Error (RSE) values”)」と記載し、利用者に注意喚起している

ただし、具体的に「どの数字以下を小さい」とするか、「RSE が大きいとはどの程度か」等を示したものは、TableBuilder の中では提供されない。オンデマンドで作成された集計表によっては、なお、RSE に関する記載がなされていない場合もあることから²¹、利用者は利用目的と照らし合わせながら、自身の判断で、データに対する信頼の範囲を決めることが求められている。

TableBuilder に係るオーストラリア統計局の人員体制については、まず、TableBuilder の専門組織というものは存在しないため、各部門 (秘匿処理の部門、運用の部門等) の中に TableBuilder に関する担当者が含まれている。また、システムエンジニアに関する体制については、基本的に TableBuilder のコーディング等は、Space Time Research が担当していることから、統計局内で担当職員を確保する必然性は低い。ただし、オーストラリア統計局は、TableBuilder の利用のサポートのために、技術スタッフを採用している。このスタッフの業務としては、新しいデータセットをリリースする (自身でデータセットを作り、テストする)、システムを停止する、または、システムを更新することが含まれる。

オーストラリア統計局は現在、Census TableBuilder の更新と、TableBuilder ver9 のデータセットの準備を行っている。また、TableBuilder を周知するための広報・啓発活動に関しては、オーストラリア統計局は、TableBuilder のみに焦点を絞っているのではなく、政府統計の二次的利用に係るサービスを各種提供しており、全体としてサービスを捉えている。そして、利用者は、それぞれのサービスの特徴を踏まえながら、ニーズに合わせた形で、政府統計の二次的利用を行っている。なお、TableBuilder の主要利用者である、大学や政府、地方公共団体に対しては、オーストラリア統計局の職員は、それらの各部門を訪問し、TableBuilder や他の ABS のプロダクトに関する情報提供を行っている。

オーストラリア統計局は、TableBuilder において用いられる攪乱的な秘匿処理技術を、プロダクトとして販売したり、ライセンス供与したりしていない。オーストラリア統計局は、プログラムに関するアルゴリズムを全て公開にしておき、TableBuilder の技術やアイ

21
<http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1406.0.55.005~Jun%202013~Main%20Features~View%20Relative%20Standard%20Errors~270>

デアを知的財産として保護していない。実際、カナダ統計局も、TableBuilder のアルゴリズムを活用している。また、ニュージーランド統計局も、秘匿処理の技術を導入しようとしている²²。特に、カナダとの取り組みは、CSPA (Common Statistical Production Architecture) のサービスカタログにも掲載されている。なお、オーストラリア統計局が開発した攪乱的措置の方法論は、論文等で外部に公表されており、オーストラリア統計局は、この方法論が自由に活用されることについて期待を持っている²³。

6. むすびにかえて - オーストラリアの TableBuilder における今後の方向

本稿は、オーストラリア統計局が開発したオンデマンド集計システムである TableBuilder の基本的な特徴を明らかにした。オーストラリアで展開されてきた政府統計の二次的利用における TableBuilder の位置付けを明らかにするために、本稿では、TableBuilder の法的根拠と秘匿処理の方法、さらに運用面での特徴を明らかにした。オンデマンド集計では、利用者のリクエストに応じて集計表を作成・提供することから、集計表の秘匿性をどのようにして確保するかが重要な点だと言える。Table Builder の特徴としては、集計表における秘匿処理の方法として、出力された集計結果表のすべてのセルにノイズが導入されていることが指摘できる。

ところで、オーストラリア統計局によれば、TableBuilder に関して、現在以下のような計画が進められている。

第1は、TableBuilder の統合である。具体的には、センサスとそれ以外の標本データでは、それぞれ別のタイプの TableBuilder が開発されてきたが、将来的には複数のタイプの TableBuilder の統合が計画されている。

第2は、利用可能な政府統計の範囲の拡大である。マーケティング等、ビジネスに関するデータは、これまで TableBuilder で提供されたことがなかったため、現在、個人や世帯とは異なる秘匿処理の手法を用いて提供することが検討されている。

第3は、複数の統計調査のリンケージによるサービスの提供である。例えば、移民は、オーストラリアの人口の32%を占めているが、移民に関するデータとセンサスデータとのリンケージデータも TableBuilder で分析可能になることが計画されている。

こうした、オーストラリア統計局における TableBuilder の展開方向については、我が国におけるオンデマンド型の提供サービスを議論する上でも参考になる点が少ないように思われる。このような諸外国における政府統計の二次的利用の動向を踏まえつつ、我が国におけるオンデマンド集計の今後のあり方を検討することが求められよう。

²² スウェーデン統計局が TableBuilder の攪乱手法に関心を持っていることについては、オーストラリア統計局は認識していない。

²³ 統計実務上における秘匿処理の詳細（どのような攪乱の手法を具体的に利用しているか等）については、オーストラリアにおける政府統計の秘匿性を守る観点から、公表されていない。

参考文献

伊藤伸介(2016a)「わが国における政府統計のデータシェアリングの現状と課題」『情報管理』, Vol.58, No.11, 836～843 頁

伊藤伸介(2016b)「政府統計におけるリモートアクセスと秘密保護について—イギリスを例に—」, 『経済学論纂(中央大学)』 第 56 巻第 5・6 合併号, 1～19 頁

伊藤伸介(2016c)「諸外国における政府統計データの提供の動向について」中央大学経済研究所 Discussion Paper No.267, 1～13 頁

Andersson, K., Jansson, I., Kraft, K. (2015) “Protection of frequency tables – current work at Statistics Sweden” Paper Presented at UNECE/Eurostat Work Session on Statistical Data Confidentiality, Helsinki, Finland, pp1-20.

Elazar, D. (2013) “Innovative Microdata Access - Confidentialising on the fly”, UNECE Work Session on Statistical Data Confidentiality, Ottawa, Canada, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_7_ABS.pptx

Fraser, B., Wooten, J. (2006) “A Proposed Method for Confidentialising Tabular Output to Protect against Differncing”, Paper Presented at UNECE/Eurostat work session on statistical data confidentiality, pp.1-6.

Thompson, G., Broadfoot, S., Elazar, D. (2013) “Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics”, Paper Presented at UNECE Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-37.

付録1 “Census and Statistics Act 1905” <抜粋>

筆者注：各条文の下に記載した日本語は、あくまで訳例であり、法律家からの確認を踏まえたものではない点に注意されたい。

7 Undertaking of fidelity and secrecy

Every officer executing any power or duty conferred or imposed on any officer under this Act or the regulations, shall, before entering upon his or her duties or exercising any power under this Act, sign, in the presence of a witness, an undertaking of fidelity and secrecy in accordance with the prescribed form.

7 忠誠と機密保護の履行

本法またはその規則に基づいて与えられた権限と義務を遂行する全ての職員は、本法に基づいて職務に入ったり、権限を行使したりする前に、規定の様式に、忠誠と機密保護の履行を証人の前で署名をしなければならない。

12 Publication etc. of statistics

(1) The Statistician shall compile and analyse the statistical information collected under this Act and shall publish and disseminate the results of any such compilation and analysis, or abstracts of those results.

(2) The results or abstracts referred to in subsection (1) shall not be published or disseminated in a manner that is likely to enable the identification of a particular person or organization.

(3) The Statistician may make charges for results and abstracts published and disseminated under this section.

12 統計の公表等

(1) 統計局長は、本法律に基づいて収集された統計情報を集約(集計)して分析しなければならず、そしてその集約・分析、又はそれら集計結果や要約(統計量)を、公表・提供しなくてはならない。

(2) 細節(1)に係る集計結果、又は要約は、(1)特定の個人または組織の識別が可能になるようなやり方で公表・提供してはならない。

(3) 統計局長は本節の下で公表・提供する集計結果や要約に係る費用を徴収してもよい。

19A Non-disclosure of Census information

Non-disclosure of Census information to Agencies

(1) A person who is or has been the Statistician or an officer must not, at any time during the period of 99 years beginning on the Census day for a Census:

- (a) be required to divulge or communicate to an Agency any information that is contained in a form that is given to the Statistician or an authorised officer under section 10 in relation to that Census; or
 - (b) voluntarily give such information to an Agency;
- other than in accordance with this Act.

Non-disclosure of Census information to a court or tribunal

- (2) A person who is or has been the Statistician or an officer must not, at any time during the period of 99 years beginning on the Census day for a Census:
 - (a) be required to divulge or communicate to a court or tribunal any information contained in a form that is given to the Statistician or an authorised officer under section 10 in relation to that Census; or
 - (b) voluntarily give such information in evidence in proceedings before a court or tribunal.

Definitions

- (3) In this Act:
 - “Agency” has the meaning given by section 7 of the Public Service Act 1999.
 - “Census day” , for a Census, means the day appointed, under subsection 8(2), as the Census day for that Census.

19A センサス情報の非開示

政府機関へのセンサス情報の非開示

- (1) 現在又は過去に、統計局長又は職員であった者は、当該統計の公表日より 99 年間いかなる時も、本法に基づく範囲以外で、
 - (a) 当該センサスに関して、第 10 条に基づき、統計局長又は権限を有する職員に与えられる様式に含まれるいかなる情報も、政府機関へ開示する、又は口外することを要求されてはならない。
 - (b) 自発的にそのような情報を政府機関へ提供してはならない。

裁判所や裁決機関へのセンサス情報の非開示

- (2) 現在又は過去に、統計局長又は職員であった者は、当該統計の公表日より 99 年間いかなる時も、
 - (a) 当該センサスに関して、第 10 条に基づき、統計局長又は権限を有する職員に与えられる様式に含まれるいかなる情報も、裁判所や裁決機関へ開示する、又は口外することを要求されてはならない。
 - (b) 自発的にそのような情報を裁判所や裁決機関へ提供してはならない。

定義

- (3) この法律では、

“Agency” は、” the Public Service Act 1999” のセクション7の定義に基づく。
“Census day” は、本法の8(2)の記述に基づく。

付録2 pTableの作成方法

pTableの作成方法は、各種論文で説明されているが、本調査研究では、TableBuilderの研究を分かりやすく、詳細に説明した“Protection of frequency tables - current work at Statistics Sweden”（スウェーデン統計局作成、2015年UNECE）の記述を一部抜する形で説明をする。

pTableは、 $m \times n$ の行列で定義され、行数（ m に相当）は256、列数（ n に相当）は30以下と設定される。

真の値を u 、攪乱する値を p_{ij} とすると、TableBuilderから得られる値は $u+p_{ij}$ 与えられる。この i 及び j は、pTableの中で、どのノイズ(攪乱値)を参照するかを示すものであるが、参照する値の決め方は、別途説明をする。

この時、攪乱する値全体（これを Π_i で表記する）は、情報エントロピーを最大化するように設定される。

$$-\sum_{p \in \Pi_i} P_i(p|u) \log[P_i(p|u)]$$

ただし、前述の通り、攪乱には平均が0であること等を含め、幾つかの前提条件が存在するため、下記要件を満たしながら、情報エントロピーを最大化することが求められる。

$$\begin{aligned} P_i(p|u) &\geq 0 \\ -\sum_{p \in \Pi_i} P_i(p|u) &= 1 \\ E(P_i(p|u)) &= 0 \\ \text{Var}(P_i(p|u)) &\leq v_i \end{aligned}$$

この計算は、ラグランジュの未定乗数法を用いて計算されるが、線形な式ではないことから、数値計算手法（プログラム）によって算出することとなる。

得られた結果を元に、R-U図（Rはリスク、Uは効用を指す）を用いて評価を行う。評価方法は各種存在するが、引用した論文では、以下の手法が利用されている。

(リスク測定手法①)

分散の逆数をベクトルで表し、そのユークリッドノルムでリスクを表現。攪乱の度合いが低い場合は、算出結果が大きくなる。

$$\begin{aligned} V(t) &= \left(\frac{1}{v_1(t)}, \frac{1}{v_2(t)}, \dots, \frac{1}{v_M(t)} \right) \\ R_1(t) &= \|V(t)\|_2 \end{aligned}$$

(リスク測定手法②)

攪乱されなかったセルの数の確率で定義する。されていない確率を算出するため、リスクが大きいくほど、算出結果が大きくなる。

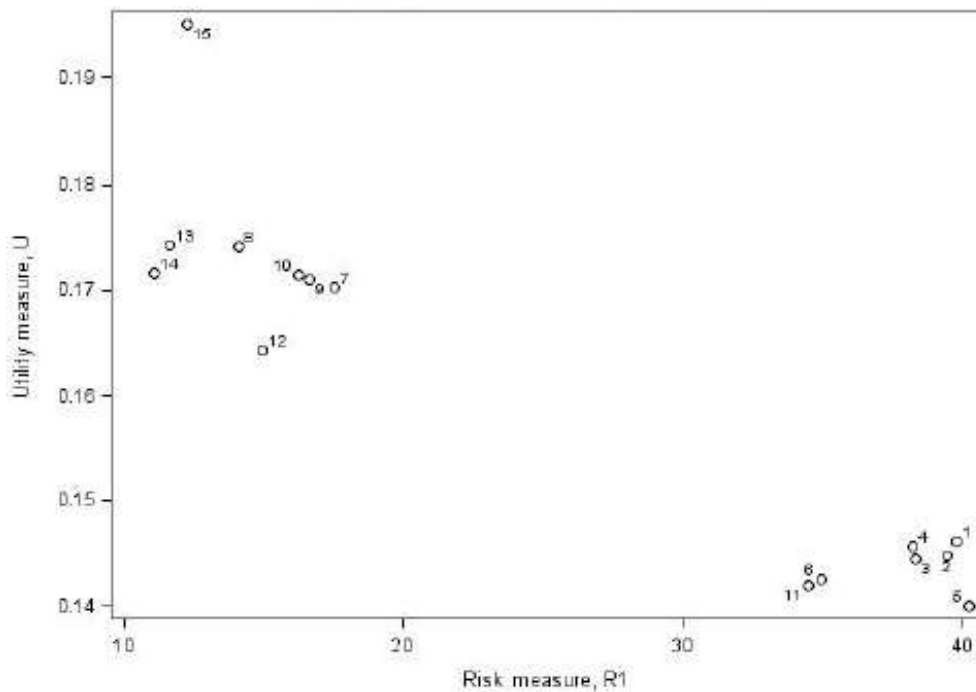
$$R_2(t) = \frac{|P_{zero}|}{K}$$

(効用測定手法)

ここで紹介する論文では、Hellinger distance を用いた効用測定手法が用いられていた。もともと、効用といっても、原データと攪乱されたデータとの差が大きいほど、値が大きくなるように設計されているため、効用損失を測定していると言った方が、より正確な表記であると考えられる。

$$u(t,c) = HD(t,c) = \frac{1}{\sqrt{2K}} \sqrt{\sum_{i=1}^M (\sqrt{|I_i(t)|} - \sqrt{|I_i(c)|})^2}$$

これらの数値を元に、R-U 図を作成する。紹介した論文では下図のような評価図が作成されていたが、具体的に、ここからどのように評価をするかについては記載が無かった。おそらく、R と U のそれぞれに何らかの閾値を設け、その閾値を満たすような pTable が存在するかを確認することになるものと想定される。



製 表 技 術 参 考 資 料 33

平成 28 年 8 月発行

編集・発行 独立行政法人 統計センター

〒162-8668

東京都新宿区若松町 19-1

電 話 代 表 03 (5273) 1200

掲載論文を引用する場合は、事前に下記まで連絡してください

統計情報・技術部統計技術研究課 TEL : 03-5273-1368

E-mail : research@nstac.go.jp