オランダの社会統計データベースSSDについて

MORI, Hiromi / 森, 博美

(出版者 / Publisher)
法政大学経済学部学会
(雑誌名 / Journal or Publication Title)
経済志林 / The Hosei University Economic Review
(巻 / Volume)
76
(号 / Number)
4
(開始ページ / Start Page)
5
(終了ページ / End Page)
28
(発行年 / Year)
2009-03-09
(URL)
https://doi.org/10.15002/00003969

オランダの社会統計データベース SSDについて

森　　博美

1．SSD構築の背景

オランダにおいて社会統計データベース（Social Statistical Database: SSD）が構築された背景には，同国の政府統計の実情並びにその将来をめぐる次のような事情があった。すなわち，①調査環境の悪化により伝統的な実査（enumeration）に基づく人口センサスは1971年を最後に実施できず，オランダ統計局（SN）ではそれに代替する方式としてvirtual census（VC）を追求してきたこと[(1)]，②小さな政府実現のために行政の効率化が求められ，将来的に統計予算の大幅な削減[(2)]が見込まれたこと，③作成される統計相互間に無視できない非整合性が存在していたこと[(3)]，④グローバリゼーションが進展する中，経済統計を中心に，統計作成の情報源として行政情報への依存度が高まってきたこと，などがそれである。

60年代終盤以降のプライバシー意識の高まりを受けて調査拒否が増加する中，政府統計には，一方で統計作成にかかる報告負担の軽減と経費の削減が，他方では詳細でかつ品質の高い相互に整合的な統計情報の提供が要請されてきた。このように統計の作成をめぐる条件が大きく変貌を遂げる中で，それまでのような調査を中心とした統計作成だけではもはや限界があるとして，オランダ統計局は最新のICT技術を前提とした統計作成方

式の抜本的見直しを行ってきた。オランダ統計局では他の行政当局との長期に亘る交渉の結果，部分的に行政情報の統計への活用が図られてきた。その後オランダでは，2003年末に統計法の抜本的改正[(4)]を行うことで，統計作成以外の目的で収集された情報の統計への活用（large-scale statistical recycling of information）を前提とした新たな統計作成システムが構築されることになる。

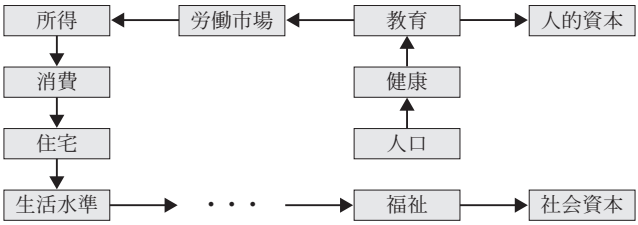
オランダ統計局には，行政情報を含めた既存の情報資源を最大限に活用し同一事項について並存していた相互に非整合的な集計結果をいわゆる one number として一本化し政府の公式数字として公表することが要請された。これを受けて同局は，1990年に integrated system of social statistics の構築による社会統計の体系化に着手した。そこでは，構想の初期段階から，登録情報を基盤とした広範なデータをデータベース化し集中管理することが想定されていた。しかし，初期のデータベースは単にいくつかのファイルを相互にリンクしただけのもので，データリンケージの結果生ずるデータ間の非整合問題などの研究が中心であった。

その後，各種の行政情報や統計データファイルの中に保存されていた人口，地理，所得，労働，教育，社会保障，健康といった諸分野の情報の統合調整がはかられ，1996年に初めてミクロ（個体）ベースのデータベースが構築された。そして1998年に同局は，SSDのプロトタイプの構築に成功した。なお，SSDは当初，virtual census への利用を直接の目的としていたが，その後，センサス目的を超えた独自のデータベースとして展開を遂げることになる。

SSDの基本的コンセプトを構成するのは次の６つの要素，すなわち，①登録情報の集約的活用，②異種の登録レコードの相互リンケージ，③世帯関連調査データ間の整合性の確保，④登録情報と調査データのリンケージ，⑤異なる情報源情報の個体レベルでのリンケージ，⑥ウエイト付けによる推計結果の整合性の確保，がそれである〔(11) p.3〕。

SSDは個人のライフサイクルの全局面を包含し，種々の人口集団につい

図１　SSD の背景となっているライフサイクルの模式図



〔出所（5）p.4〕

て，その社会・人口面，社会・経済面，さらに社会・文化面を反映するデータベースとして構築されている。そこでは，図１のようなライフサイクルの模式図がその背景に想定されている。なお，それらの各項目は，同時に政府の政策的関与事項でもある〔(5) p.4〕。

２．SSDの情報源とその構造

(1) SSDの情報源

データベースの中核（コア）部分を構成するのは，各地方自治体（municipalities）が維持管理している人口登録（Population Register: PR）[(5)]である。それがデータベースとして保有する変数は，社会保障番号（Social security and Fiscal number: SoFi）[(6)]，出生年月日，性別，郵便コード，住戸番号，RIN-person，RIN-address，変数の期間情報[(7)]である〔(8) p.6〕。また，PRは，世帯規模，世帯構成，世帯のタイプ，世帯員の続柄といった一連の世帯に関する情報を持っており，それは毎年更新される。

SSDでは，PR以外にも各種の行政情報や調査データがデータベース化されている。

まず，調査データとしては，EWS（Employment and Wages Survey），LFS（Labor Force Survey），ISSS（Integrated System on Social Surveys

（Living Conditions Surveyに相当）），SEE（Survey on Employment and Earnings）が現在SSDには収録されている。

他方，行政情報ファイルとしては，EIS-E（Employee Insuranse Scheme Registration System-Employment Insurance）EIS-DI（Employee Insurance Scheme Registration System-Disablement Insurance），EIS-UI（Employee Insurance Scheme Registration System-Unemployment Insurance），SABA（Social Assistance Benefits Administration），FiTap（Register of Final income Tax assessments on profits of self-employed persons），FiBase（Fiscal Database）がある。

表１は，SSDに収録されている主要ファイルのレコード数と変数名を一覧したものである。

表１　SSDの主な情報源泉

	ファイル名	レコード数	主要変数
基幹データ	PR	4644万(*)（1633万人）	社会保障番号（SoFi），出生年月日，性，郵便コード，住戸番号，RIN-person，RIN-address，変数の期間情報，世帯規模，世帯構成，世帯のタイプ，世帯員の続柄
調査データ	SEE	300万	労働時間，就業の場所
	LFS	23万(**)	学歴，職業，求職活動
	EWS		賃金，労働時間
	ISSS		教育，健康
行政情報	PR	1600万	性別，年齢，世帯規模，世帯の種類，同居者の有無，世帯員の続柄
	EIS-E		
	EIS-UI	44万	
	EIS-DI	100万	
	SABA	58万	
	FiBase	560万	労働，年金，生命保険
	FiTap		

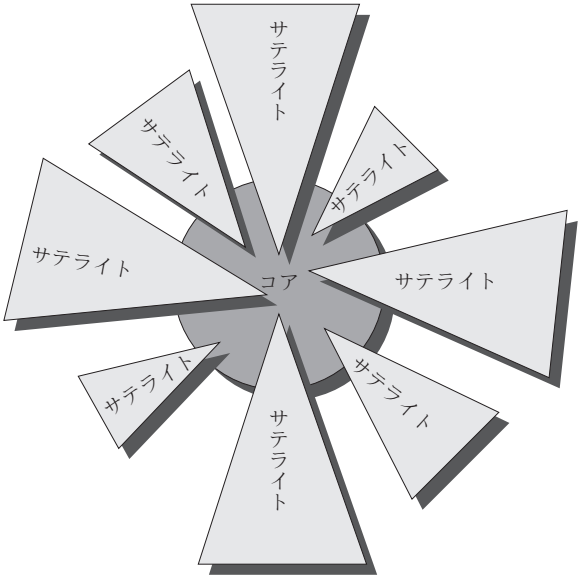
(*) 2007年４月現在，他は2000年末現在
(**) 2000年と2001年の合計標本数

（2）SSDの構造

SSDの所蔵ファイルは，コアファイルとサテライトファイルという２つのカテゴリーのファイル群からなる。PRファイルがSSDのコア部分を構成し，それ以外のファイルがサテライトファイルとしてコア部分にリンクされている。

一方，変数レベルでSSDを捉えた場合，各レコードは，コア部分を構成する変数とサテライト部分に属する変数要素からなる。このうちコア部分を構成するのは，PRから得られる情報，SSDの運営上特に重要とされる人口，社会・経済情報，それに二つ以上のサテライトファイルで使用されている諸変数である。他方，コア変数から複写作成された変数や複数の変数から導出される派生変数，それに各種の登録や調査データのうち特定のテ

図２　SSD のコアとサテライト



〔出所（8）p.14〕

ーマに関する変数がサテライト部分を構成する。図2は，SSDのコアとサテライトの関係を模式的に示したものである。

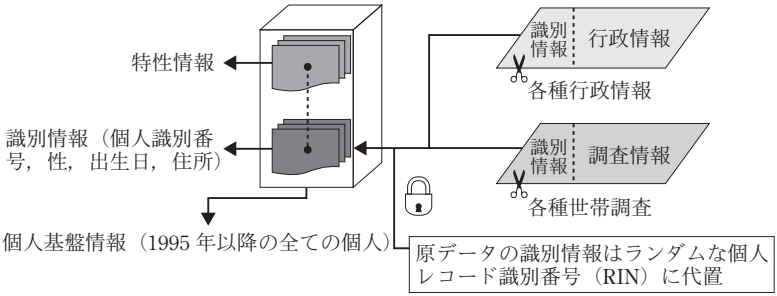
SSDがこのようなコアと複数のサテライトからなるファイル構成を持つのは，単一の大規模ファイルとしてデータベースを維持するのが更新やデータの秘密保護の観点から適切ではないとの判断によるものと考えられる。

分散的に維持されたファイルの各レコードは，共通のリンクキーによってコアとサテライトさらにはサテライト同士が相互にリンクされる。これは，個人に関する行政記録に識別情報としてSoFiがつけられており，またSSDの整備に伴いSSDが標本調査における標本抽出のフレームとして用いられることになったことから，調査データについても各レコードに統一的な識別情報を持たせることで可能となったものである。

1996年の改正統計法は，政府が保有する行政情報に対する広範なアクセス権をオランダ統計局に対して保証することになった。SSDの構築との関連で特に重要なのが，同法が統計局に対して統計目的でのSoFi番号の使用権限を制度的に保障した点である〔(1) p.7〕。これによって，各種の行政情報並びに調査データの個体ベースでマッチングの効率と精度が飛躍的に改善された。

ところで，SoFiは社会保障当局と税務当局が共有使用している社会保障や納税に関係する特に取扱い上慎重を要する極めてsensitiveな個人情報である。このため，オランダデータ保護庁（Dutch Data Protection Authority）では，その使用を厳しく規制している。SoFiの利用上の制約からオランダ統計局では，SSDでのSoFiのリンクキーとしての使用について，次のような方法で個人情報の保護とデータのマッチング機能との両立をはかっている。すなわち，SSDに新たに追加される情報についてはSoFiを介してSSDの既存レコードに個体ベースでリンクされるが，このうちマッチングに成功したケースについては，追加される情報ファイルの各レコードにSoFiに代って統計局が独自にランダムな数値として作成したその代替識別情報（Surrogate Identifier）であるRecord Identification Numbers（RIN-person）

図３　SSD における統計の秘密確保



〔出所（8）p.19〕

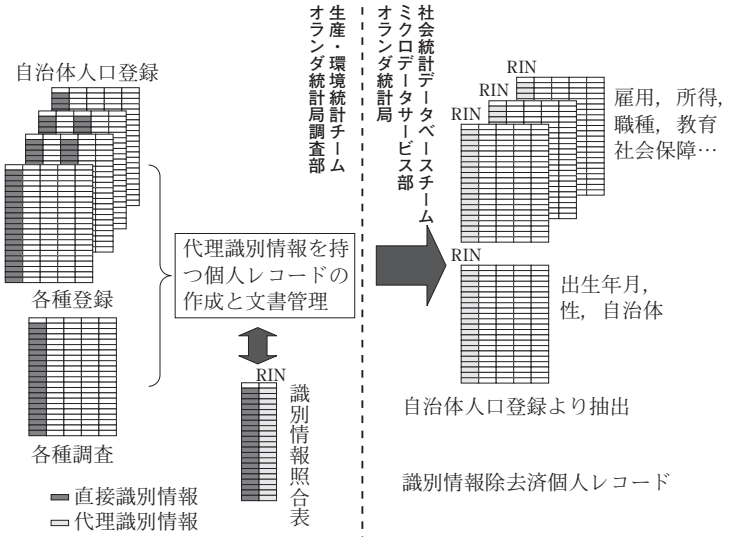
を付加することでそのリンク可能性が保証される。なお，この他にもRINでは，秘密保護の観点から，SoFiと連動した情報である出生年月日は基準日現在の年齢に，また住所はRIN-addressに置き換えられる〔(7) p.249〕。

SoFi番号とRIN-personの照合ファイルは個人の特定に直結する。このため照合ファイルについては，厳格な保護管理が求められる。ちなみにオランダ統計局では，SSDの維持管理業務部門であるSocial Economic State Central Section（SESCS）から完全に独立させた特別な部署Link Person Record Identity（LPRI）に所属する限定された数の職員[(8)]が，専従的にその管理に当たっている。

図３は，SSDにおける秘密保護の仕組みを図示したものである。

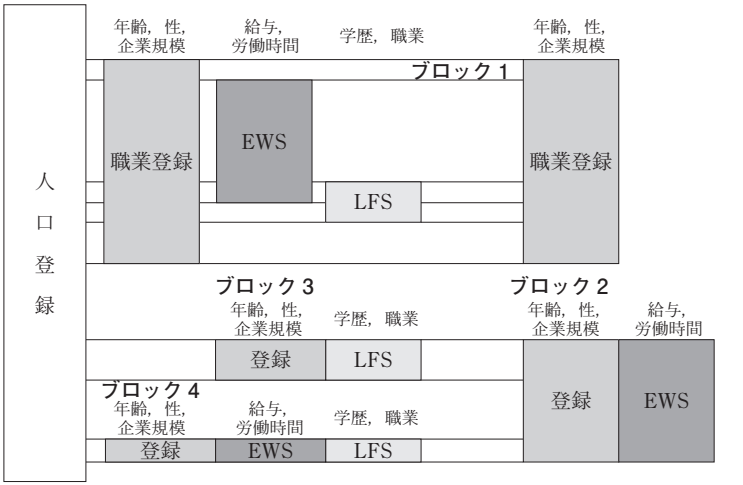
このようにSSDでは，コア並びにサテライトを構成するファイルの各レコードに共通のリンクキーとしてSoFiに代わるRIN-personを持たせることで，必要に応じてリンク済データをデータベースから引き出すことができるようになっている。図４は，RIN-personが付与された各ファイルからどのようにSSDが構成されるかを模式的に図示したものである。

図４　代理識別情報を持つ個人レコードの作成システム



〔出所（8）p.8〕

図５　SSD における行政情報と調査データのリンク



〔出所（2）p.58〕一部加筆修正

図５は，現在，SSDにファイルとして維持管理されている主要な登録並びに調査データ[(9)]のリンク状況を模式的に示したものである。

３．SSDの構築過程

(1) マッチングによるレコードのリンケージ

オランダでPRに登録を行っている者及びオランダから所得を得て納税義務を有する国外居住者は全員SoFi番号を持つ。不法滞在者の中にSoFi番号を不正に取得し使用している者〔(7) p.249〕がいないわけではないが，不適切にSoFi番号が付与されるケースは極めて稀である。ちなみに，1995年から2007年におけるSoFiのカバレッジは99.97%を超える。SSDで各レコードを相互にリンクする際に第一義的なリンクキー情報としてSoFi番号並びに関連するSoFi情報が用いられるのはこのためである。

PRを基盤情報としてそれに他の行政情報さらには各種調査によって収集された個体情報をSoFi番号さらにはSoFi情報によってミクロベースでマッチングさせることでSSDの個体レコード（原レコード）は作成される。その具体的なマッチングの手順は，主に次の３つのステップからなる。

〔第１ステップ〕リンクされるレコードがSoFi番号並びにSoFi情報を持つ場合，PRの登録情報とリンクされるレコードとでSoFi番号が一致し，しかも出生年，出生月，出生日，性の全てが一致するかあるいは不一致が高々１個の場合，両レコードは「照合」と認定される。

〔第２ステップ〕上記の照合作業でマッチングできなかったケースについて，郵便コード，住宅番号，出生年，出生月，出生日，性の全てが一致する場合，両レコードは「照合」と認定される。

〔第３ステップ〕第２ステップでマッチングできなかったケースについて，他の情報を無視してSoFi番号だけを使って照合される〔(8) p.7〕。

なお，第３ステップによってもマッチングできないレコードについて，

さらにマッチングの可能性を追求するために，誤記入の可能性にも配慮して，住居番号情報の無視あるいは郵便コードに一定の許容範囲を設けるなどの方法も用いられている。ちなみに，LFSの個人レコードはSoFi番号を持っていなかった。そのため，LFSの個人レコードとのマッチングは，性，出生年，出生月，出生日，住所（4桁数字と2桁文字）および住居番号を用いて行われてきた。なお，同性の双生児，出生年月日の記載ミス，移転の未届出などに起因するミスマッチのために，LFSの個人レコードのうちの2～3％はPRとマッチングできない。移動率が特に高い10代後半から20代前半世代の若者，都市（アムステルダム，ロッテルダム，ハーグ，ユトレヒトの4大都市）居住者の間で，マッチングに成功した割合が低かった。その後，SSDをLFSのサンプリング・フレームとして使用するようになったことから，LFSの非照合問題は解消した〔(9) p.7〕。なお，オランダでは，個人情報を持つ登録にはすべてSoFi番号を付与することが義務づけられている。このため，LPRI チームが管理しているSoFiとRIN-personの照合情報を介してSSDの既存の他のファイルとこれらの登録ファイルとは相

表2　EWSについてのマッチング情報

マッチングできた数		3,747,976	98.6%
	1．SoFi，誕生年，月，日，性	3,577,090	94.1%
	2．郵便コード，　〃	164,267	4.3%
	3．SoFiのみ	6,619	0.2%
マッチングできなかったケース		53,270	1.4%
	SoFiが有効な者	21,194	0.6%
	郵便コードが有効な者	5,799	0.2%
	郵便コード無効な者	10,294	0.3%
	非居住	5,101	0.1%
	SoFi不明者（あるいは無効者）	32,076	0.8%
	郵便コードが有効な者	8,718	0.2%
	郵便コード無効な者	20,052	0.5%
	非居住	3,306	0.1%
合　　計		3,801,246	100.0%

〔出所（8）p.9〕

互にリンク可能である。

表２は，2003年のEWSの調査サンプルのマッチング結果を示したものである。

(2) データの調整（micro-integration）

調査統計の場合，統計の比較可能性を担保するために産業分類や職業分類など一定の統計基準に基づいて統計が作成されている。他方，行政活動の遂行過程で収集ないしは作成される行政情報の場合，分類や統計項目の区分は個々の行政目的に従って行われるのがむしろ一般的である。このため，SSDに収録されているファイルの中には，同一の統計項目について，しばしばカバレッジを異にする複数の類似した定義が並存している。あるいは，情報源が異なるため，定義が等しいにもかかわらず変数値が異なるケースも発生しうる。また，統計の定義そのものは同じであっても，その回答（選択肢）区分が調査間あるいは調査と登録とで異なる場合もしばしば見られる。このように，同一の事項に対していくつかの異なる数字が並存する場合，そのいずれを正式の数字としてSSDに収録するかというデータ選択の問題が発生する。

このような定義調整に関する問題に加えてSSDでは，提供されるデータそのものの補正の必要性も発生する。なぜなら，SSDに原情報を提供する各種の行政情報や統計調査の結果データが，全て完備データであるという保証はないからである。各種の行政情報や統計調査の結果データの中には，一部の変数（項目）についてデータそのものが欠損しているものがある。また，マッチングの結果，幼児のレコードに有業者としての属性を持つデータが接続されるなど，論理的整合性を欠くケースが発生する可能性も排除できない。以上の諸問題は，いずれもSSDが母集団を反映した有効なデータベースとして可能な限りバイアスを最小化したデータを提供するために対処しなければならない問題である。

従来からも個々の調査結果について，データの論理チェックや記入漏れ

の補定（imputation）といったデータの編集（editing）が集計作業に先立って行われてきたが，SSDが性質や品質の異なる様々な情報源情報から構成されることから，より多角的な調整が必要となる。リンク済レコードのデータ補正に関わる一連の作業工程が，データの個体ベースでの調整，いわゆるmicro-integrationとしてSSDには組み込まれている。それは，SSDに収録される変数を相互に一貫した，整合的でしかも完備なものとするための調整過程である。なお，integrationはSSDに含まれる全変数に対して行われるのではなく，公表が予定されている変数についてだけ実施される。

調整過程での具体的な作業内容としては，①統計単位の調整，②対象期間の調整，③母集団の完全性（カバレッジ），④異なる定義間の調整，⑤分類の調整，⑥定義調整後の変数値の違いについての調整，⑦回答欠損項目の補定，⑧異なるデータソースからの新変数の創出，⑨全体の整合性のチェック，などがある。なお，オランダ統計局は独自にソフトウエアを開発，実用化しており，これらの作業工程は完全に自動化されている〔(7) pp.251-2〕。

SSDのmicro-integration工程では，このようなリンク済みレコード上のデータの調整といういわば狭義のmicro-integrationに加え，SSD上の既存の変数を用いた新たな変数の創出（derivation）も行われる。SSDにおけるmicro-integrationとは，それらを総称したものである。

(3) レコードの重み付け(Weighting)

SSDにストックされるLFS等の調査データは，いずれも計画サンプルのうち調査客体から実際に回答が得られたデータに他ならない。なお，後に見るように，SSDは，その機能の一つとして，集計結果表を提供するためのデータベースであるStatLineに対して原データを供給する役割も担っている。

一般に標本調査の結果データは，センサスが把握した母集団情報を元に，集計結果が母集団に一致するように抽出率の逆数を乗率として修正され

る。SSDがセンサスデータではなくPRを基盤情報としていることから，各種の標本調査データは，例えばPRが与える20歳以上の登録人口のように，PRが与える人口が母集団となるように復元される。各標本調査の間でそれぞれ抽出率が異なることから，initial weightsと呼ばれる各レコードに与えられたウエイト，すなわち復元のための乗率も調査の間で当然異なる。

さらに，SSDの母集団反映性という点で最も重要な問題の一つが，調査データにおける計画サンプルと回答サンプルとの乖離の問題である。調査環境悪化の進展に伴い，政府統計調査においても有効回答率はおしなべて低落傾向にある。計画サンプルとして調査は試みられたものの，種々の事情で回答が得られなかったいわゆる欠損サンプルの問題は，SSDが母集団分布を適切に反映したものとなるためには無視することができない。このためオランダ統計局では，推計結果のバイアスを可能な限り小さくするように，non-response biasの削減にも取り組んでいる。

集計表の作表は，個体レベルで調整（integrate）された重み付きレコードに基づいて行われる。その場合，理論的には，ある集計表の各セルの計数は，追加的な変数を加えた多重クロス表の周辺分布に一致しなければならない。言い換えれば，あるクロス表の各セルの数値は，追加的な変数を用いて作成される多重クロス表において追加変数を統合することで元のクロス表が再現される必要がある。しかし，実際の作表過程では，本来一致すべき計数が食い違うケースも起こりうる。このため，作表にあたっては，多数の変数を用いた多重クロス表の場合，相互の計数の間で整合性が維持されるような形での計数調整が行われる。

また，変数の中には，例えば年齢のように，集計表において各歳，５歳階級，10歳階級，生産年齢など統合の程度を異にするいくつかの区分が存在するものがある。このような変数については，相互に位階的（hierarchical）整合性の要件を充足することが求められる。このため，各変数並びに変数間に位階的な序列を設定し，集計結果に反復的に重みづけ（Repeated Weighting : RW）を行うことで，整合的な結果表の体系を得るための事後

調整が行われている。なお，これらの一連の操作はマクロ統計（集計値）に加えられることから，macro-integrationと呼ばれている。なお，オランダ統計局はそのための方法論を1997年に考案し，試行運用を経て2000年からこの自動処理システムは本格的に稼働している。また，この自動処理システムの採用により，作表作業が以前に比べてより迅速に行われるようになった[(10)]。ちなみに，整合的な結果表の体系を得るために，オランダ統計局が現在採用している作業原則とは下記のようなものである〔(2) p.3〕。

1．データベースの中で当該表の作表に必要な全変数が利用可能なファイルの中で最大のデータブロックを使用すること

2．該当するデータブロックのブロックウエイトと最も整合的なクロス表を最初に推計すること

3．低次元のクロス表から順次高次元のクロス表へと作表を進めること

4．最も適当なデータブロックのブロックウエイトと整合的なクロス表が得られない場合，反復加重（RW）法による補正を行うこと

(4) SSDの作成過程

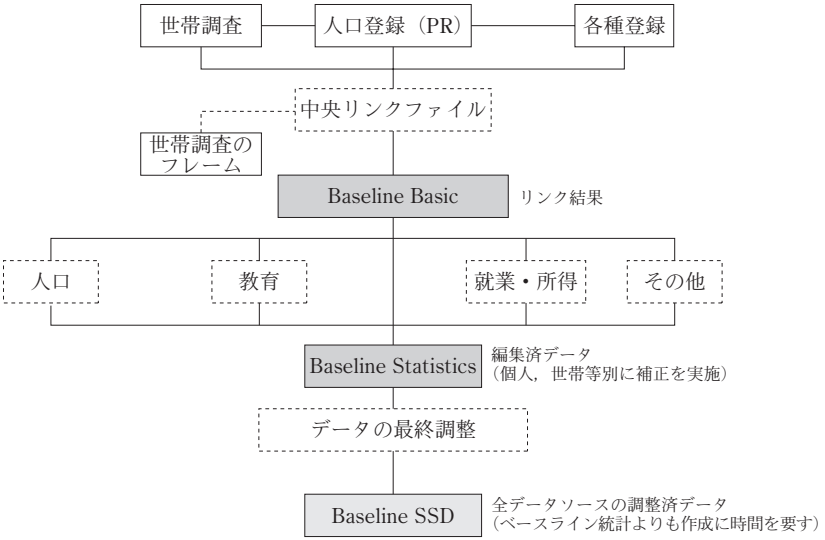
PRを基盤情報として各種の世帯調査と行政情報とを個体ベースでリンクすることでBaseline Basicファイルが構築される。これは，人口，教育その他の情報を原情報として持つレコードをPRの登録レコードにリンクさせただけのもので，SSDに収録される原レコードからなるデータベースである。

このBaseline Basicが有する原レコードに個人あるいは世帯ベースで登録情報や調査データを用いて様々な補正を施すことで，Baseline Statisticsと呼ばれる次の段階のデータベースが作成される。

その後，Baseline Statisticsの各レコードにウエイトが付加され，データベースからの作表結果が母集団を適切に反映するようなデータの補正が施されて，最終的に社会統計データベースとしてBaseline SSDが構築される。

図6は，SSDの作成過程の概要を図式的に示したものである。

図６　SSD 作成の概念図



〔(5) p.6〕一部加筆等修正

４．SSDの利用

政府が作成，提供する統計に対しては様々な利用ニーズが存在する。オランダ統計局では，SSDを単に既存の統計情報の維持保管のためだけでなく，政府が作成する統計の原データとして位置づけている。このため，最終的に維持更新されるBaseline SSDだけでなく，その作成過程で構築される作業データベースであるBaseline Statisticsは，政府統計に対する様々なニーズへの対応を求められることになる。

(1) 速報値の作成・提供

Baseline SSDのデータ更新作業の多くがすでに自動化されているとはいえ，収録データの調整（micro/macro-integration）には一定の作業量を必

要とする。政府が作成する統計の中には，速報性が求められるものもあり，この種の利用ニーズに対応するためにオランダ統計局では，Baseline Statisticsに基づいて暫定的な結果数字としての速報統計を作成，公表している。なお，この速報暫定値にかわる確定値については，Baseline SSDのデータ更新作業が終了次第提供される。

(2) 統計表提供用データベースとしてのStatLineの作成

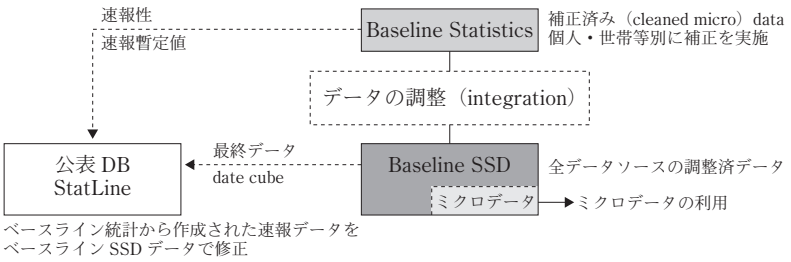
オランダ統計局は，政府統計の中心的な提供チャンネルとして，利用者の様々なニーズに応えるために構築した統計表提供用データベースStatLineを持っている。StatLineでは，Baseline SSDのレコードから作成したdata cubeと呼ばれる超多次元の集計表 [11] をそのデータ構造とすることで，統計上の秘密保護と多様な集計ニーズの充足との両立を図っている。StatLineに対しては，利用者が直接インターネットを介してアクセスし，個々の利用目的に応じて必要なデータを検索し，必要な集計表を作表，入手できるようにシステムが設計されている。

なお，StatLineの使用にあたっては，秘密保護の観点から，情報が100%安全な形で提供されねばならない。このために，集計結果表のセルの出力表示について，計数が 0 ～ 4 の場合には 0，また 5 ～ 9 については10がそれぞれ自動表示される仕組みになっている。

(3) SSDのミクロデータ提供機能

SSDの中には，RIN-person付きファイルだけでなく，匿名個体データ（ミクロデータ）も保存されている。このような個体データについては，秘密保護への配慮からStatLineには収録されておらず，Baseline SSD上のファイルとして別建てでその運用が図られている。なおミクロデータについては，on siteでの利用となっており，処理結果について利用者自身は出力することができず，オランダ統計局では Micro Data Statistics Section が統計上の秘密保護に関する点検を行った上で出力結果を提供している。

図７　SSD と StatLine



〔(5) p.6〕一部加筆等修正

図６は，SSDの作成に至る流れとそのそれぞれのステップにおけるデータベースの使用を図式化したものである。

(4) SSDの世帯フレーム機能

オランダでは，1971年を最後に，全数調査として実施された人口センサスは存在しない。このため，PRを基盤情報として構築され，維持更新されているSSDが，センサスに代って最も包括的な個人・世帯情報を提供している。図６にも示したように，基盤ファイルであるPRに各種の登録情報と調査データのレコードをリンクして作成される中央リンクファイルは，世帯を対象に実施される標本調査のサンプリング・フレームとしても使用されている。

なお，SSDで特筆されるのは，それが税務当局から提供される所得情報を保有している点である。このためオランダでは，Living Condition Surveyの調査票には所得に関する調査事項は設けられていない。またBudget Surveyでは，調査票に税務当局が把握した所得額をpre-codingし，調査客体にその内容確認を求めるという形で調査が実施されている。

このように，SSDが所得情報を持つフレームとして機能していることから，オランダ統計局はこの所得分布データを標本調査結果データの補正に活用している〔(12) p.51〕。特に，所得分布データは，欠損レコードの検出などnon-response biasの補正さらには各レコードへの適切なウエイトを

付与する上で有効であると考えられる。

むすび

1990年代以降，オランダ統計局は，①外部組織[13]からの直接的な電子的データの確保，②行政登録からのデータの確保，それに③統計調査と利用可能な登録情報との連携による弾力的な統計作成という３つのアプローチを内容とする”from assembly line to electronic highway junction”[12]を掲げ，厳しい予算制約の中で政府統計の積極的な改善に取り組んできた〔(10) p.5〕。

実査によって収集した統計原情報を集計し公表するという伝統的ないわゆるassembly line型の政府統計のあり方から新たな方式への脱皮を目指すオランダ統計局にとって，SSDの構築並びにその管理・更新は，その中核部分を構成する最も重要なシステム要素の一つといえる。それは，統計作成面では同国のvirtual censusの存立根拠となる情報基盤をなすだけでなく，様々なモジュールの下に企画・実施されているLFSを初めとする各種標本調査に対するサンプリング・フレームとして，その有効性を保証する役割も担っている。さらにSSDは，統計の提供面では，統計表提供用データベースであるStatLineに対する情報の供給源にもなっている。

SSDはまた，政府統計の保管のあり方という観点から見ても極めて興味深い。第１に，それはマッチングキー情報を保有した形でコアとサテライトの分散処理型システムとして設計されている。第２に，SSDではRINpersonとSoFiとの照合情報ファイルを厳格な管理下に維持することで，SoFiの付与された新たな個体情報レコードについて，極めて高い精度で既存のレコードとのマッチングの可能性を保証している。特に後者は，新たな個体レコードに時点情報を付与することで，そのための特別な調査を実施することなく，いわば電子的な処理としてパネル型のデータベースが半ば自動的に構築できることを含意している。

このように，政府統計の新たなあり方の一つとして注目されるSSDではあるが，現時点でなお多くの問題を抱えているのも事実である。例えば，SSDに行政機関から提供される行政情報の中には，はなはだしい場合には提供に 2 年以上を要するものもある。このため，利用可能な行政情報が存在することが確認されていても，速報ニーズの要請に対応するために，オランダ統計局では敢えて統計調査を実施しその実態把握を行っている場合もある〔(5) p.4〕。

今後のSSDの拡充方向としては，①SSDに蓄積される情報をより包括的なものとするために，今後，さらに多くの調査データにリンクキーを付与すること，②横断面でのいわゆるsnapshotとしてのリンク済みデータの蓄積から，将来的には縦断面分析への対応が可能なevent-based data warehouseとしての展開方向 (14) などが考えられる。

行政情報と調査データとの有機的連携を基調とするオランダ政府統計の今後を見る上で不可欠なシステム要素としてのSSDの今後の展開に引き続き注視したい。

注

（1）オランダ統計局で2008年８月に筆者が実施した訪問調査の際のEric Schulte Nordholt氏の説明によれば，1981年，1991年のvirtual censusの推計結果の精度については，満足できる水準にはないとして，研究者だけでなく行政の側からもその改善が求められていたとのことであった。

（2）1994-2002年の間に統計作成に係る業務の60%削減，また2002-2006年にはさらに25%の削減が求められた。

（3）オランダ統計局で2008年８月に筆者が実施した訪問調査でのもう一人のインフォーマントBart Bakker氏は，オランダ統計局がSSD構築を決断する背景の一つとして，集計結果表の計数が相互に整合性に欠けていたことを指摘している。同氏の説明によれば，economic accounts，labour accounts，education accounts等の間に計数の非整合性が存在し，それらをミクロ（個体）レベルでデータの調整（integrate）を図ることでそれを最小化するシステム並びにそのための方法論を構築ことが当時求められていたとのことである。

（4）オランダ統計法（2003年11月）の条文構成・内容については，拙訳『海外統計制度研究資料』（法政大学日本統計研究所）No.2　2008年10月を参照。

（5）PRは，2007年4月現在で16,334,210の個人に関する46,436,060レコードを有する。

（6）SoFiは９桁の数字から構成される。なお，現在はcivil service numberと呼ばれている。

（7）継続就業年数のような期間情報

（8）専従職員数４名

（9）現在，登録情報としては毎年47のファイルが，また調査データについては，年によって本数は多少異なるものの，５～10の調査データが維持更新されている。

(10) 2001年人口センサスでオランダ統計局は，結果表の作表作業を1年で完了することができた。その結果オランダは，EU加盟諸国の中でEurostatに対して2000年ラウンドセンサスの結果の最初の提供国となった〔(6)〕。

(11) かつてわが国では，集計表の系統別に結果原表という当該系統に関係する変数（項目）を網羅した超多元集計表が作成され，それを逐次積み上げることで公表結果表の作成が行われていた。集計結果表の公表後も，個体レコードではなく結果原表が，廃棄される調査票の代わりに永年保管されてきた。

（12）オランダ統計局The Strategic Development Plan‘CBS 2000’〔（10）p.5〕。
（13）例えば，SEEは企業を対象とする標本調査として実施されているが，その情報源である企業情報は，源泉徴収機構（payroll administrations）から電子データ流通（electronic data interchange: EDI）を通じて得られる。
（14）SSDが少なくとも定期的に提供される登録情報についてはすでに現時点でパネル的データ構造を持つことが，今回実施した関係者からのヒアリングでも確認された。

参考文献

（1） Statistics Netherlands (1999) Informed Consent: Buzzword or Panacea, an invited paper submitted to the Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, Greece, 8-10 March 1999.

（2）Marrianne Houbiers (2002) Towards a social statistical database and unified estimates at Statistics Netherlands, *Journal of Official Statistics*, Vol.20. No.1

（3）Laihonen, Aarno, European Union and Population and Housing Censuses around the year 2001

（4）Bart F.M. Bakker (2002) Statistics Netherlands' Approach to Social Statistics: The Social Statistical Dataset, *The Statistics Newsletter for the extended OECD Statistical Network,* No.11

（5）Pieter C.J. Everaers and Paul van der Laan(2003) The Dutch System of Social Statistics: Micro-Integration of Different Sources. ESA/STAT/AC. 88/06

（6）Eric Schulte Nordholt (2004) The Dutch Virtual Census of 2001, *ISI Newsletter,* Vol.28, No.3 (84)

（7）Eric Schulte Nordholt, Marijke Hartgers and Rita Gircour eds.,(2004) *The Dutch Virtual Census of 2001-Analysis and Methodology.* Statistics Netherlands, Voorburg/Heerlen.

（8）Eric Schulte Nordholt (2007) Record matching for census purposes in the Netherlands, http:// www.unece.org/ stats/ documents/ ece/ ces/ ge. 41/ 2007/ 6. e. ppt# 256.1. Record matching for census purposes in the Netherlands.

（9）Statistics Netherlands (2007) Record matching for census purposes in the Netherlands, paper submitted for the Conference of European Statisticians 4-6 June 2007

(10) Gosse van der Veen (2007) Changing Statistics Netherlands Driving Forces for Changing Dutch Statistics, paper presented at the Seminar on the Evolution of National Statistical Systems, Commemorative Event for the 60th Anniversary of the United Nations Statistical Commission

(11) Pieter Everaers and Paul van der Laan, The Dutch Virtual Census. http:// unstats. un. org/ unsd/ dnss/ docViewer.aspx?docID=249#start

(12) 神田玲子・棚川真宏（2007.10）「ヨーロッパにおける「家計調査」の課題とそれへの対応－その１　オランダ－」『統計』（日本統計協会）

〔付記〕本稿は，平成20年度科学研究費補助金（基盤研究Ｃ）「センサス機能の変質，新展開およびその統計制度，統計体系への影響に関する総合的研究」（課題番号19530188）の研究成果の一部として公刊するものである。なお，本年８月に実施した海外調査において，統計局のLeidschenveen新Officeへの移転後間もない業務多忙な中，SSDの実際に関して貴重な情報提供をいただいたオランダ統計局の社会・経済部門の部長Bart F. M. Bakker氏並びに社会・空間統計部門上席研究員・プロジェクトリーダーのEric Schulte Nordholt氏のお二人にこの場を借りて謝意を表したい。

The Dutch Social Statistical Database

Hiromi MORI

《Abstract》

Statistics Netherlands (SN) is now regarded as one of the first runner group in world statistical entities in terms of their data policy and setting by furnishing the system based on the latest developments in electronic technologies. As the catch words "from assembly line to electronic highway junction" symbolize, SN has established a new compilation system of official statistics, which are based on micro based matching and integration of administrative records with survey data.

This paper has elucidated the characteristics of Social Statistical Database (SSD) that works as a core system of Dutch contemporary data practices. SSD has several dimensions in its function. It works not only as the data source for the Dutch virtual census but also works as the official statistical data archive. Remote access system to official statistics called "StatLine" is also derived from SSD as an annex database for dissemination purposes.

Official statistical data are now regarded as a sort of durable social assets with historical importance. The structure and function of SSD seem to offer a lot of suggestions for future designing statistical data archive in Japan.