【『統計学』創刊60周年記念特集論文】
特集B：政府統計ミクロデータの作成・提供における方法的展望
（『統計学』第120号 2021年3月）

公的統計における統計メタデータ・アーカイブの展開可能性

小林良行*

要旨

本稿では，公的統計全体を対象とする統合統計メタデータ・アーカイブ及びオントロジーの構築を提案し，その展開可能性について指摘している。統計業務過程を通じて生み出される統計情報は統計データとそれに関係する統計メタデータが一体となっている情報である。本稿では，二次利用を前提とする統計業務過程のモデルと統計情報の性質について考察している。また，個別統計調査の統計情報アーカイブがデータウェアハウスで実現できることを指摘している。統計データ・アーカイブと統計メタデータ・アーカイブは統計情報アーカイブを構成する。公的統計の統計メタデータは，大局的，半大局的，局所的の3層に分類できる。この3層構造に基づき統合統計メタデータ・アーカイブを構築すると，公的統計の用語の標準化や統計情報のトレーサビリティなどが実現可能である。公的統計のオントロジーを確立することにより，統計メタデータの概念の標準化が図られる。

キーワード

統計メタデータ，統計業務過程，データウェアハウス，トレーサビリティ，オントロジー

1．はじめに

法制度や技術の変化に伴い，統計業務の形態や範囲は変化していくものであり，統計業務を構成する一連の業務過程の集まりである統計業務過程[1)]も変化していくものである。新統計法（平成19（2007）年法）の二次的利用制度の下では，統計業務過程は二次利用[2)]を前提とする業務を最初から含むようなものに変化することになる。調査票や統計表を二次利用しようとするためには，それらを適切に保存し，いつでも利用できるようにしておく必要がある[3)]。森（2008）は調査票を二次利用のために保存したものを情報資産と呼んでいる。また，森（2008），山口（2019）は，公的統計における調査票情報の保管・管理体制を集中型とする利点について，データの管理と利用の面から指摘している。

公的統計の統計業務過程では，調査票から統計表を作成する業務過程は情報システムで行うのが一般的である。公的統計調査では近年，オンライン調査の導入が進んできているが紙媒体の調査票は依然として多くの調査で利用されている。また，統計調査結果はイン

* 正会員，総務省統計研究研修所
e-mail：ykobayashi@nstac.go.jp

ターネットを通じて提供されるとともに紙媒体の調査報告書に収録され利用されている。情報は実世界に具体的に存在するものではなく，情報を直接見るには紙媒体に記録（記述・印刷）したもの（たとえば，調査票）が必要になる[4]。本稿では，調査票情報とは記述済み調査票の記録内容を情報システムで扱えるようにしたものあるいは画面に表示された調査項目に直接回答を入力して情報システムで扱えるようにしたものとする。

友安（1957）は統計実務家向けに統計表の表構造について詳しく解説している。小林（2019）では，1つの調査票情報を複数の項目の直積とみて，調査票情報の集合から集計値の集合を作成する業務過程をモデル化し，統計表の表体は各集計値を表の各セルに対応付けて配置したものであると述べている。また，対応のさせ方を変えると統計表の表現形（表の見え方または表の形式のこと）は変わるが，表体に配置する元となる集計値の集合は変わらないとも述べている。本稿では，統計表情報とは調査票情報の集合から作成した複数の集計値を表形式で表現した情報とし，統計表とは統計表情報を編集して紙媒体上に記録したものとする[5]。なお，本稿中で，「調査票」の利用や「統計表」の利用と表現しているのは，「調査票上の記述内容」の利用や「統計表上の統計その他の印刷された内容」の利用を意味している。

調査票，調査票情報，統計表情報及び統計表は，そこに記録されている内容（以下，「情報本体」）と項目の配置のしかたのように情報本体の形式的構造を表現する仕様（以下，「様式」）から構成されている[6]。通常，情報本体と様式は一体として扱われる。情報本体は複数の項目（たとえば調査票なら各調査項目）の集合と考えてよい。なお，様式は，上記でいう情報本体とは別種の情報である。調査票，調査票情報，統計表情報及び統計表の情報本体は，様式と分離独立させることにより，共通のデータベース構造を用いて表現可能である[7]。このことは，公的統計調査全体を対象とした一元的なアーカイブが構築できることを示唆する。

一般に「アーカイブ」には，①組織的に収集し保存された記録や資料，②記録を保管する組織や機関，③記録を保管する施設という3つの意味がある（たとえば，国立国語研究所（2003），Pearce-Moses（2005）など）。本稿では，記録媒体の種類（紙媒体，電磁的媒体，光学的媒体）にかかわらず，収集された個々の記録や資料の集積全体を①の意味のアーカイブと考え，単に「アーカイブ」としたときには①の意味で用いている。一方，②及び③の意味で用いるときは，それぞれアーカイブ組織及びアーカイブ施設と表すことにする。公的統計の統計調査の調査票，調査票情報，統計表情報及び統計表を，将来の利用に供するため，収集，保存したものは，①の意味のアーカイブと言ってよい。

統計数値は，それがどのように作られ，何を表しているのかを理解していなければ適切に利用することはできない。統計数値のようなデータは，その意味を理解するのに必要となる情報と結びつかなければ単なる抽象的な数に過ぎない。一般に，データが表す意味を理解するのに必要なデータのことを「メタデータ」という。メタデータの最も簡明な定義は「データに関するデータ」であり，統計分野の情報システムにおいてメタデータに相当する語を最初に使用したのはSundgren（1973）である。UNSC and UNECE（2000）によれば，統計メタデータとは統計データのメタデータのことである。

美添（2005），森（2008）では，調査票情報を保存する統計データ・アーカイブの必要性とともにメタデータの整備，蓄積の必要性についても指摘している。小林（2012）は，統計データの作成過程と並行して統計メタデータを収集すること及び統計データ・アーカイブ

と統計メタデータ・アーカイブを一体的に整備することの必要性を指摘している。統計改革推進会議（2017）では，統計，統計ミクロデータ及び行政記録情報にメタデータを含めた統計等データの利活用を社会全体において促進することとされており，そのための基盤の整備に言及している。諸外国と比べ，日本の公的統計分野では統計メタデータあるいは統計メタデータ・アーカイブの整備の必要性を指摘する研究等が散見されるに留まっている。統計メタデータは，統計データを解釈する上で本質的な重要性を持っているにもかかわらず，公的統計においては従来，統計データの補助的，附属的な情報として取り扱われており，あまり重要視されてこなかった。

統計情報と統計業務過程は互いに密接に関係する存在である。本稿では，統計情報とは統計業務過程の生起，進行，終了に従って発生，変化，消滅，存続していく情報及びそれらの情報相互の関係を表す情報[8]からなるものであるとする。したがって，調査票（に記述された内容），調査票情報，統計表情報及び統計表（に印刷された内容）に限定するものではない。Radermacher et al.（2009）は，統計情報とは統計データとそれに関係する統計メタデータが一体となっている情報であるとしている。本稿における統計情報アーカイブとは，統計情報が統計データ及び統計メタデータから構成されるという視点で，統計情報を保存したものである。

各府省は，総務省政策統括官（統計基準担当）（2019a）に沿って，調査票情報及び匿名データを符号表やレイアウトフォームを含むドキュメント類と共に保存し，適正に管理することが求められている。しかし，利用者が調査票情報を利用しようとする際には符号表とデータレイアウトは電子的な情報として提供されるものの，それ以外の統計メタデータ，たとえば項目の定義，調査方法，回収数（率）や有効回答数（率），エディティング規則，エディティングによる項目値の変遷情報，標本調査における推定方法や推定精度などについては提供されず，それらは調査報告書を参照せざるを得ない[9]。また，統計表情報についても，分類項目と調査項目の関係，集計項目の算出方法など，やはり調査報告書を参照せざるを得ないのが現状である。

調査報告書は，統計調査に関する情報が網羅的，一元的に集積されているものである。今日では，調査票や統計表は電子的な情報として提供されるようになったものの，それらの利用上必要な統計メタデータは，調査報告書のように網羅的，一元的な形で電子的な情報として整備されているわけではない。すなわち，調査票情報や統計表情報の利用時に必要となる統計メタデータを，わざわざ調査報告書のような別の媒体上に存在する情報源から探し出し，参照しなくてはならないという不便さが解消できていない[10]。また，電子化されていても統計メタデータが再利用可能な形式で保存されていない場合や統計データと統計メタデータあるいは統計メタデータ相互の関係付けがされていないような場合には，やはり利便性が低いと言える。統計メタデータ・アーカイブ構築の必要性の一つは，このような問題を解決することにある。

以下，第2節では統計データ及び統計メタデータの種類並びに統計メタデータ標準の動向に触れる。第3節では二次利用を前提とする統計業務過程のモデルと統計情報が具えるべき性質について述べる。第4節では公的統計の統計情報アーカイブの構造とアーカイブ実現のための技術について述べ，次いで，第5節では統計メタデータ・アーカイブの展開可能性について示す。最後に第6節で今後の展望について述べる。

2．統計データと統計メタデータ

2.1　統計データの種類

日本では，統計データという語の意味が集

計値から個別データや匿名化データを指すものへと変化してきた[11]。一方，UNSC and UN-ECE（2000）によると，統計データとは統計ミクロデータと統計マクロデータの両者[12]を包含する概念とされており，日本と比べ明確にされていると言える。

従来，統計データは個別性の有無で統計ミクロデータと統計マクロデータに分けられていた。調査票の一次利用により統計表を作成・提供する統計業務過程では，統計作成の中間段階でサマリーデータあるいは中間集計表と呼ばれるものがある。このデータは，ミクロレベルとマクロレベルの中間（メソレベル）に位置するので，メソデータ（mesodata）と呼ばれる（Radermacher et al.（2009））。メソデータは，ミクロデータを集計したものであり，マクロデータより粒度（データの集約レベル又は詳細レベルのこと）が相対的に細かいデータである。諸外国では2000年代に統計データの提供形態が多様化し，メソデータをデータキューブの形態で提供する例がみられるようになった（小林（2012））。しかし，管見の限りでは，日本の公的統計分野でデータキューブの提供は実現していない。表1は統計データをその特徴と粒度の違いで整理したものである。

表1　統計データの種類

名称（略称）	特徴	データの粒度	該当する例
統計ミクロデータ（ミクロデータ）	個別的，非集約的	小 ↑ 細	調査票情報，匿名化データ
統計メソデータ（メソデータ）	集約的，非個別的		データキューブ
統計マクロデータ（マクロデータ）		大 ↓ 粗	統計表情報

メソデータはマクロデータに対して相対的に決まるものである。統計調査ではさまざまな統計が作成される。本稿では，調査計画で当初公表することにしていた統計をマクロデータ，それを作成する中間段階で得られる統計をメソデータとする。しかし，ある統計調査で公表される複数の統計表間の関係を考えると，ある粒度の統計がマクロデータでありメソデータでもあるという二重性を持つことがある。たとえば，都道府県レベルの統計表と全国レベルの統計表が公表される場合，都道府県別表にある統計は二重性を有するデータである。ある統計調査において絶対的なマクロレベルのデータを決めるのは現実には困難である。また，メソデータの粒度を細かくしていくとミクロデータと変わらない粒度のデータが出現するようになる。メソデータを提供する場合，このような疑似ミクロ性が生じないような粒度を探索的に見出すことで，この問題は実務的には解消可能である。

2.2　統計メタデータの種類と統計メタデータ標準を巡る動向

欧米における統計メタデータの整備や統計メタデータを用いた統計情報システムの構築に関する研究は，1990年代から2000年代にかけて盛んになった。90年代半ば以降になると，統計メタデータに関する用語集やガイドラインが作られている。近年では，統計業務過程を一般化した汎化統計業務過程モデル（GSBPM: Generic Statistical Business Process Model）及び統計メタデータの概念モデルである汎化統計情報モデル（GSIM: Generic Statistical Information Model）が提案されている。

これらの動きの中で，統計メタデータに対する認識は，当初の簡明で抽象的なものからより広範で具体性のあるものに変化している。たとえば，Dippo and Sundgren（2000）は，統計メタデータを①統計データを解釈，理解，分析するのを助けるもの，②統計データを識別し，その所在を見つけ，検索するのを助けるもの，③統計調査の設計・企画プロセス及び実施プロセスに関して記述及びフィードバックに使われるものの3つに分類している。③はパラデータ（paradata）とも呼ばれて

いる（Radermacher et al.（2009）。ただし，この語自体はCouper（1998）が作ったとされる）。パラデータには，たとえば調査票の回収数（率），項目補完率などがある。また，Radermacher et al.（2009）は，①構造的メタデータ（structural metadata）と②参照メタデータ（reference metadata）の2つに分類している。前者は統計データを識別，形式的に記述，検索するのに利用するもの，後者は統計データの意味的視点による内容と品質を記述するものである。現在では上記の例のように，統計メタデータには統計データと直接的に結びつくものと必ずしも直接的に結びつくとは限らないものがあると考えられている。

統計メタデータの実装レベルの標準として代表的なものにミクロデータ向けのDDI（Data Documentation Initiative）と集計データ向けのSDMX（Statistical Data and Metadata eXchange）の2つがある[13)]。DDIでは，従来の標準（DDI Codebook）とは別に，統計業務過程を対象にして統計データのライフサイクルを記述するDDI Lifecycleが作成されている。DDI codebookは，CESSDA（Consortium of European Social Science Data Archives），CRDCN（Canadian Research Data Center Network）などで利用されている。また，DDI Lifecycleは，フランス，オランダなどの統計局で利用されている。

一方，SDMXでは，EurostatがSDMXを拡張してより多くのデータ品質に関する情報を含むようにしたESMS（Euro SDMX Metadata Structure）を作成し，メタデータ標準として採用している。近年では，GSIMとGSBPMを概念モデルとして，DDIとSDMXといった既存の標準間の相互運用可能性を高めたり，統合化を進めたりしていこうという動きが出てきている[13)]が，管見の限りでは統計業務過程全体を通して包括的，体系的に記述可能なメタデータ標準の実現には至っていないようである。

3．統計業務過程と統計情報

旧統計法（昭和22（1947）年法）下の統計業務過程は，調査票と統計表の一次利用を前提としており，調査票の二次利用は例外的，附随的な業務であった。法制度上，調査で使用する調査票及び作成される統計表の様式と種類は，調査設計者が調査設計時に決定しておかねばならないため固定的であり，事後的な追加や変更は認められなかった。したがって，統計業務過程の処理形態は定型的である。一方，新統計法では二次的利用制度が創設されたことから，統計業務過程は初めから二次利用を前提とする業務過程，すなわち統計情報の保存と提供の2つの業務過程，を組み入れたものに変化することになる。統計表情報の二次利用のニーズはもともと不定期に発生するものである。また，調査票情報の二次利用のニーズは不定期に発生するものであり，作成される統計表の様式と種類は統計利用者が自由に決めることも可能である。さらにこれらのニーズは内容も求められる提供形態も多様であるため，二次利用のための提供業務は一般に非定型的である。

本節では，旧統計法下の統計業務過程と新統計法下の統計業務過程を区別するため，前者を「一次利用型統計業務過程[15)]」，後者を「二次利用型統計業務過程」と呼ぶことにする（図1参照）。一次利用型統計業務過程は，法制度の変化に関わらず基幹的な役割を担っている統計業務過程である。二次利用型統計業務過程の下で保存と提供の対象となる統計情報は，一次利用型統計業務過程全般に由来するものである。以下，単に「統計情報」と言った場合は一次利用型統計業務過程に由来する統計情報のことを指すものとし，調査票，調査票情報，統計表情報及び統計表を「狭義の統計情報」と呼ぶことにする。統計情報の保存と提供は，一次利用型統計業務過程が終了した後も継続して機能していくことになる[16)]。

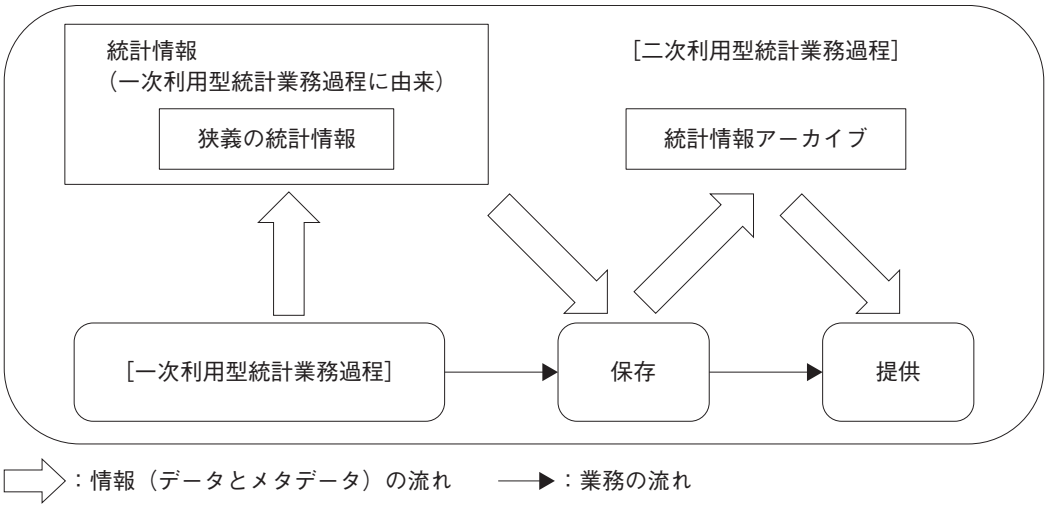

図1　二次利用型統計業務過程の概念図

実世界では調査対象が時間とともに変化するため，一度実施した統計調査とまったく同一内容の統計調査を再現することはできない。特に，調査票上の記述内容は，一部でも失われてしまうと復元ができないものなので，統計情報アーカイブの中では最も重要な情報である。統計情報アーカイブの中で狭義の統計情報は再現性を具えている必要がある。ここでいう再現性とは，保存している調査票情報又は調査票を用いて，一次利用型統計業務過程と同じ統計的方法，同じ処理手順を施すと，保存している統計表情報又は統計表と同じ成果物が得られるという性質とする。再現性は，保存している調査票情報，処理手順などの正当性を保証するものである。また，統計情報は完全性，すなわち保存している統計情報のみを用いて再現性が実現できるという性質を具えている必要がある。完全性は，再現性を実現するのに必要な情報が，保存している統計情報の範囲で充足していることを保証するものである。二次利用に当たり再現性と完全性は重要な性質と言える。保存している統計情報が再現性と完全性を備えていることにより，再集計に必要なデータがないため過去の集計結果を再現できないという事態は避けられることになる。過去の集計結果をいつでも再現できることで，統計作成者だけでなく統計利用者も，統計作成の業務過程と生成物の正当性を評価することが可能となる。

統計情報アーカイブの主要な関係者として統計利用者，統計作成者及びアーカイブ管理者（以下,「アーカイブ関係者」）が考えられる。以下では再現性及び完全性について，これらのアーカイブ関係者との関係の観点から見てみることとする。現行の二次的利用制度の下では，脚注15で述べているように調査票情報の二次利用で提供対象となるのは，エディティング済みのクリーンな調査票情報であるから，統計情報アーカイブから調査票情報の二次利用者に提供するのもクリーンな調査票情報となる。したがって，統計利用者にとっての再現性は，この調査票情報から統計表情報又は統計表までの範囲での実現ということになる。一方，統計作成者とアーカイブ管理者はエディティング前の調査票情報や調査票を利用できるので，再現性はこれらの情報から統計表情報又は統計表までの範囲での実現ということになる。特にアーカイブ管理者にとっての再現性は，保存している統計情報に誤りや欠落，記述のあいまいさといった

ことがないかを検証する際に不可欠な性質であると言えよう。一方，完全性はアーカイブ関係者が第1節後半で述べた統計情報利用時のメタデータ参照における不便さを解消するために必要な性質である。

4．統計情報アーカイブとデータウェアハウス

統計表は，調査報告書の形で図書館に保存され，利用されるという形態が，長い間一般的であった。図書館は，現在も統計表のアーカイブ施設として，あるいはアーカイブ組織としての役割を担っていると考えてもよいだろう。統計審議会（1985）の提言後，情報通信技術の進展に伴い，現在では統計表（調査報告書掲載表だけでなく非掲載表も含む）は，紙媒体の印刷上の制約を受けない統計表情報として保存され，インターネットを通じて提供されるようになっている。

一方，調査票は，伝統的な視点では「統計材料[17]」として位置づけられており，成果物である統計表ができてしまえば不要と認識されていた。諸外国におけるデータ・アーカイブ組織の成立・発展は，社会調査や統計調査のミクロデータの二次利用に対する社会的要請と密接に関係して進んできた[18]。日本では，新統計法施行後，公的統計の基本的な計画を定めた統計委員会（2009）及び同（2014）の中で，各種統計調査の調査票情報の蓄積を政府として推進していくものとされている。分散型の統計機構をとる日本では，統計調査を所管する府省が第一義的に調査票情報の保存を行っている。総務省政策統括官（統計基準担当）（2019a）は，各府省が統計調査によって収集した調査票情報を国民の共有財産として将来にわたり利活用可能とするための調査票情報等の管理に関する指針を示している。その中では，符号表とレイアウトフォームに加えて調査票情報を利用するのに必要となるドキュメント類を調査票情報と併せて保存することとされている。

調査票の様式は統計調査によって多様であるし，統計調査によっては複数種類の調査票が用いられることがある。また，統計表の様式は一つの統計調査の中にあっても複雑・多様である。さらに，統計編成過程の情報システム設計者が設計する調査票情報と統計表情報も多様な様式で表現される。既存の統計表を様式と共に標準化して，データベース（以下，「DB」）を構築しようとする試み[19]はあったものの，公的統計全体の範囲では実現していない。調査項目の標準化に関する研究会（2006）の検討結果に基づき，調査票情報や匿名データは政府標準レイアウト記述や符号表によって記述することが決められている。しかし，政府標準レイアウト記述では個別統計調査に由来する様式部分まで標準化されているわけではない。

二次利用型統計業務過程で統計情報をアーカイブに保存する場合，狭義の統計情報の多様性，複雑性を解決しておく必要があるが，それには発想の転換が必要であろう。第1節で述べたように狭義の統計情報は，情報本体と様式が一体の形で表現されている。様式は情報本体の見せ方を与える上で必要なものである（小林（2019）は統計表の見せ方を「表現形」と呼んでいる。この概念は狭義の統計情報に拡張できる）。様式が既定されていると，調査票の記入，統計表の集計などといった狭義の統計情報を作成する作業の操作性は向上する。しかし，調査票の情報本体は収集する調査項目が既定されていれば様式に依存せずに収集し得るし，統計表の情報本体は目的とする統計の作成方法が既定されていれば様式が既定されていなくとも作成可能である。狭義の統計情報が見せる多様性，複雑性は様式に由来するものであって，情報本体に由来するものではない。狭義の統計情報から様式という一種のメタデータを分離すること，すなわち様式独立にすることにより，情報本体を

後述する多次元DBの形で保存でき，保存時のDB構造の統一性と提供時の表現形の柔軟性を両立させることが可能となる。中間集計表の様式は通常，統計表情報の様式を想定して定めるので，上述の考え方は統計メソデータにも適用できる。DB構造の統一性はアーカイブされた様式独立な狭義の統計情報の管理容易性を高める。さらに，どの統計調査も保存時のDB構造を同一のモデルで表現できることから，公的統計全体の統計情報アーカイブ構築を進めることが期待できる。また，表現形の柔軟性は利用者自身が希望する様式で統計情報を利用することができるという利点がある。

図2は，個別統計調査の統計情報アーカイブの構造と利用に関して図示したものである。図中の両矢印は，両端に示すものの間に参照，生成などの関係があることを表す。統計情報アーカイブは，相互に関係する統計データ・アーカイブと統計メタデータ・アーカイブから構成される。統計データは，情報本体を構成する項目のうち量的項目に該当するものである。たとえば，様式独立な統計ミクロデータは，調査票情報の量的項目に該当するものである。統計メタデータには，統計データに直接的に結びつくものと間接的に結びつくものがある。図2では，統計データと直接的に結びつく統計メタデータのうち秘匿方法と様

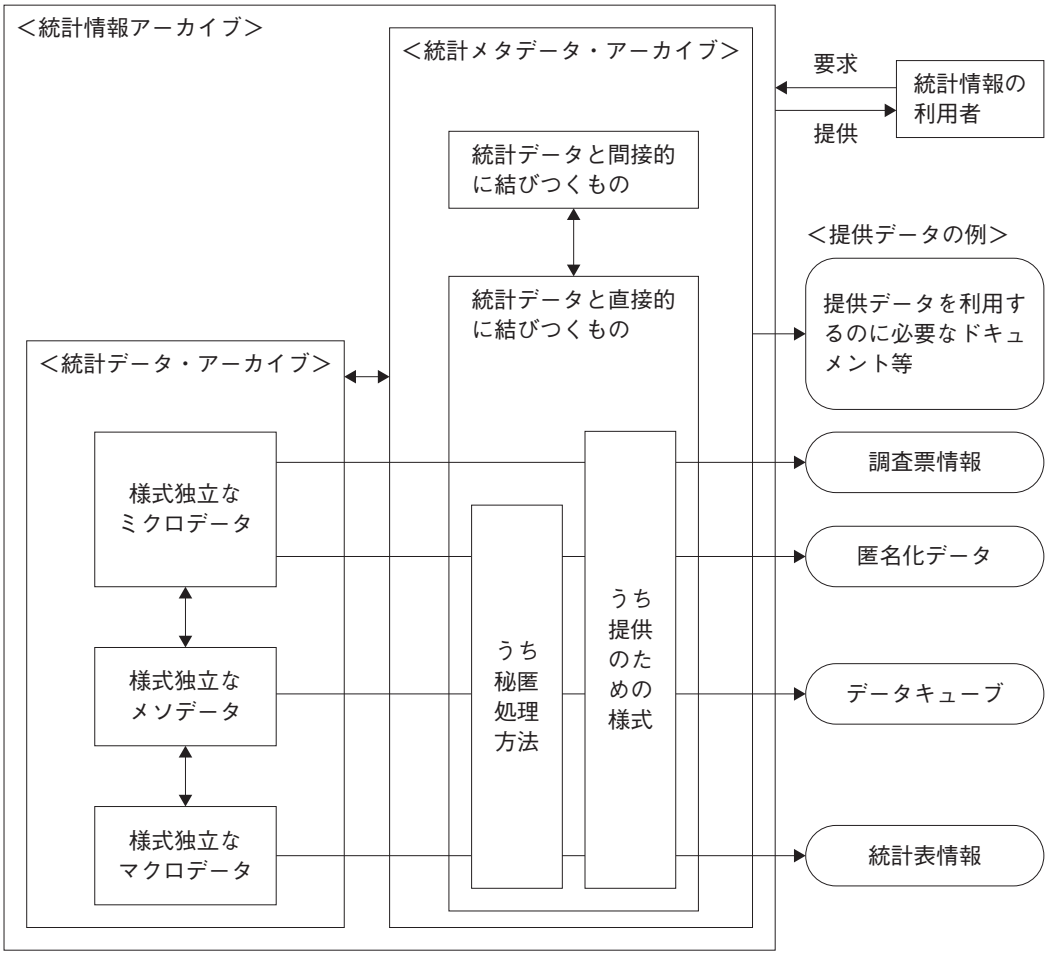

図2　統計情報アーカイブの構造と利用の概念図

式については特に明示してある。二次利用者の要求があると，統計情報提供者は必要に応じて統計データの秘匿処理を行った上で，調査設計時の様式又は二次利用者が指示する様式に従って，統計データ及び関係する統計メタデータを選択して配置した形態で統計情報を提供することになる。また，併せて提供データを利用する上で必要となるメタデータを提供する。

統計情報は一次利用型統計業務過程の進行に沿ってアーカイブに格納していくのが望ましい。特に調査票情報から統計表情報までを順次アーカイブに格納する情報システムが実現すれば，図2の統計情報の利用者には統計作成者も含むことになる。ただし統計的開示管理の観点から，利用者の立場によってアクセス可能な統計メタデータ及び/又は統計データの範囲には制約を設ける必要がある。

統計情報アーカイブを実現するには，DB技術が重要な要素技術となる。1980年代にDB技術が発展し，DBの設計，構築のシステム開発において，メタデータの有用性が認識されるようになった[20]。1990年代に入ると，企業活動の日常で発生する個別データとそれを集約したデータを多次元DBとして蓄積するデータウェアハウス（Data WareHouse。以下，「DWH」）とオンライン分析処理（OnLine Analytical Processing。以下，「OLAP」）ツールが，企業の経営意思決定支援システムとして登場した[21]。多次元DBは，検索や分析の対象となる数値をセルに，検索や分析に用いる複数の属性を配列の軸にした多次元配列で実現される。統計データは多次元DBのセルに対応し，統計メタデータは多次元DBの軸になる複数の属性に対応するものと考えてよい。OLAPとは，経営意思決定のように不定期に発生しかつ非定型な処理を要するようなエンドユーザーの要求を実現するため，蓄積されているデータを多面的な視点で検索・分析し，結果を迅速に提供する処理をオンラインで行うものである。UNSC and UNECE（1999）は，政府統計機関のクリアリング機能を実現する統計情報システムの一つとしてDWHを取り上げている。OLAPの基礎となる多次元DBは，データキューブとも呼ばれ，統計分野では統計情報提供の一形態として利用されている。

図2の右側には，統計情報アーカイブの利用と提供データの例を示している。一般的な流れとして，利用者は，利用したい統計データ又は統計メタデータを要求し，アーカイブから提供を受ける。本稿で提案する統計情報アーカイブは，第1節で述べたような狭義の統計情報の利用上の不便さを解消するものであり，提供されるのは基本的に電子的な情報になる。統計データの利用要求と提供は，技術的にはオンラインで処理可能なものの，データの特徴（表1参照）によって法制度上の扱いが異なるため，OLAPのような処理が一律に利用可能となるわけではない。一方，統計メタデータのみの利用要求と提供は，技術的にはオンラインで処理可能である。

DWHは，通常，時系列でデータを保存し，一度保存したデータは更新しないので，アーカイブとしての性質を満たすものと言える。DWHは統計データと統計メタデータの一体的な保存を実現する技術と言えよう。DWH構築の際には，統計情報の完全性を満たすために，狭義の統計情報を包含する統計情報全体をDWHとすることが必要である。

5．統計メタデータ・アーカイブの展開可能性

5.1　統合化による展開可能性

調査項目の標準化に関する研究会（2006）では，基幹統計調査（検討当時は指定統計調査）の調査項目について定義の標準化の研究が行われ，その成果は「政府統計の総合窓口（e-Stat）」で見ることができる。しかしながら，同研究会で問題提起された統計表の表章項目の定義の標準化は，今日に至っても進ん

でいない。

個別統計調査レベルの統計メタデータ・アーカイブを整備する際に考慮すべきことの一つは，調査票情報と統計表情報の各々が持つ項目間の相互参照性を確立することである。統計表の分類項目と集計項目は，もともと調査項目等に由来するものである。項目間の相互参照性の確立は，後述するトレーサビリティの確立に寄与し得るものである。項目間の相互参照性が確立していない状態であると，調査報告書を参照しなくてはならなくなる可能性があり，統計情報の完全性を満足しないことになってしまう。また，DDIとSDMXのように調査票情報と統計表情報で別々に統計メタデータを整備するのは，個別の統計調査の統計メタデータ・アーカイブ内で統計メタデータの重複や脱漏を引き起こす可能性が高く，望ましくない。調査票情報と統計表情報の項目間の相互参照性の確立は，公的統計全体を通じた統合的な統計メタデータ・アーカイブ（以下,「統合統計メタデータ・アーカイブ」）構築の際の共通語彙基盤の整備につながるものでもある。

統計メタデータは，公的統計全体に共通する大域的なメタデータ，個別の統計調査に固有の局所的メタデータ及び両者の中間的な半大域的なメタデータに分類できる（図3）。この分類に従って統計メタデータ・アーカイブも，個別統計調査レベルから公的統計全体レベルまで，階層的に構築していくことが考えられる。

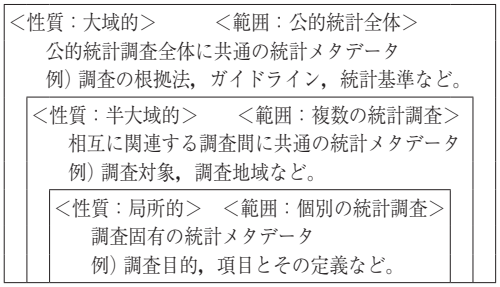

図3　統計メタデータの階層

統計調査ごとの統計メタデータが集積され，統合統計メタデータ・アーカイブが構築されることで，さらに可能となり得ることとして，下記の点が考えられる。

1点目は，公的統計における用語の標準化の実現である。個別統計調査の統計メタデータ・アーカイブの統合化を進めることにより，統計メタデータの調査横断的な相互参照性や共通化のための整合性の確立を進めることが可能となる。共通語彙基盤の構築における統計メタデータやその記述で使われる用語の標準化は，調査設計者が統計メタデータを二次利用して新規統計調査を設計する際や統計利用者が複数の統計調査の調査結果を比較する際に有用である。

2点目は，トレーサビリティの実現である。統計データのトレーサビリティとは，ある統計データに着目したときに当該データの作成元に該当する統計データは何か（トレースバック），また当該データの利用先に該当する統計データは何か（トレースフォワード），及びその統計データはどのような統計的方法により作成されたかといったことを明らかにし得ることである。

二次統計では，トレーサビリティは統計の品質保証，品質評価において必要なものであり，統計作成者の業務過程の透明性，成果物の適正性を示すものと言える。また，統計利用者にとっても統計の理解の向上につながるとともに，利用者自身が新たな加工統計を作成する際の参考になり得る。これは，二次統計に関する統計メタデータを包含する統合統計メタデータ・アーカイブの構築により初めて可能となる。一次統計においても，トレーサビリティは重要であって，頻度表や数量表の集計値が，どの個別データに由来するのかを知ること[22]により，集計方法の誤りなどのチェックに利用できるなど，統計編成過程の品質評価に役立つ。もちろん，個別データにまで遡るトレーサビリティ情報へのアクセス

は，統計的開示管理の観点から統計作成者の範囲に限るのは言うまでもない。

また，統計メタデータのトレーサビリティも考えることができる。これはたとえば，ある導出項目の導出元になる調査項目は何か，その調査項目を利用した導出項目は何か，その導出項目はどのような導出方法で作成されたのかといったことを明らかにし得ることである。統計メタデータのトレーサビリティは，統計メタデータの概念間の関係を含めて考えるのがよいので，後述するオントロジーの確立が必要となる。

3 点目は，統計情報アーカイブの情報セキュリティとオープンデータ化の両立の実現である。メタデータだけをドキュメントとして調査票情報と別に保存する方法では，情報の更新があった際に相互の同期を確保することが困難であるし，誤りも生じやすい。その点，DWH はデータとメタデータ相互のリンクは常に同期し，整合するようにできるという利点がある。個別統計調査の統計情報アーカイブにおいて，ミクロレベルの統計データとその統計メタデータとのリンケージ情報は秘匿して，承認された利用者のみがこのリンケージ情報を含めた統計ミクロデータへのアクセスを許可される仕組みを作ることは技術的に可能である。また，統合統計メタデータ・アーカイブにある調査項目等とその定義記述を，調査横断的，網羅的に検索し，比較できるような情報システムを構築することは技術的に可能である。これにより，統計ミクロデータの秘匿は保証しつつ，アーカイブされた統計メタデータを広く利用者にオープンデータとして公開することが可能となるであろう[23]。

4 点目は，データリンケージのキー項目候補を調査横断的に探索することの実現である。データリンケージにおけるキー項目の候補が統計データにアクセスする前に探索可能となると，調査票情報の二次的利用の申出前にリンケージ方法を検討する上で有用である。特に，リンケージデータによる実証研究では，分析のためのデータ調製作業の時間を短縮することにつながるだろう。複数の調査票情報を組み合わせたリンケージデータは，単独の統計調査の調査票情報に比べ，はるかに高い開示リスクを持っていると予想できる。統計データのリンケージキーファイルの管理と利用は厳格に行う必要がある。それにもかかわらず，リンケージデータの作成は新たに調査を行わないで有用なデータを得る可能性を持つ有効な方法と言える。データリンケージに関する先行研究において作成されたリンケージキーファイルとその作成方法を，統合統計メタデータ・アーカイブに保存[24]し，一種の共有財とすることできるならば，リンケージ項目の探索の際に役立つだろう。そのためには，制度面の検討，リンケージキーファイル作成者の積極的貢献などが不可欠である。

5.2　オントロジー構築による展開可能性

統計情報は，実世界をどのように観測し，その観測によりとらえた実世界がどのような姿に見えるかを表すものと考えてもよいだろう。統計メタデータが表すものとは対象物を理解するのに必要な概念である。統計情報アーカイブを構築する際に筆者が最も重要と考えるのは，統計メタデータが表す概念や概念間の関係の構造体系（オントロジー）を考え，確立することである[25]。

用語の標準化は，用語の背後にある概念の標準化と密接に関係していると言ってよいであろう。現実には，本来同じである概念が調査ごとに異なる用語で呼ばれる場合や同じ用語が異なる概念を指している場合が生じ得る。既存の統計調査にはすでに長い期間用いられている用語があり，それらをすべて標準化して共通語彙基盤を実現するのは必ずしも実際的とは言い難い。この問題を解決するには，公的統計分野のオントロジーの構築が有効で

あろう。統計メタデータやその記述に使われる用語のオントロジーを整備することによって，調査内や調査間で概念の統一性，概念相互の整合性を持った統合統計メタデータ・アーカイブの整備が期待できる。公的統計全体を対象として整備されるオントロジーは，概念の標準化を実現し得るものと言えよう。

オントロジーの構築によって可能となることの一つとして，統計メタデータが表す概念に基づく統計データ検索の実現が考えられる。統計データの二次利用ではまず，どのようなデータがどこにあるかを知ることが必要である。データの内容，種類，所在などを表すものはメタデータであるから，統計データの検索は本質的には統計データに結びつく統計メタデータの検索に帰着すると考えられる。通常のキーワード検索では，人が統計メタデータや用語の意味を解釈し，適切なキーワードを考えて統計メタデータ・アーカイブを検索し，統計データを得ることになる。オントロジーが構築されると，検索に用いる統計メタデータや用語に結びつく概念間の関係を用いて，統合統計メタデータ・アーカイブから個別統計調査の統計メタデータ・アーカイブに至り，統計メタデータと結び付いている統計データを得るといった検索の実現が期待できる。

6．今後の展望

統計データは，統計メタデータと関係付けられていなければ単なる抽象的な数に過ぎない。それゆえ，統計メタデータは，統計作成者と統計利用者の双方にとって，統計データの意味を理解して利用する上で重要な役割を担う不可欠な存在であると言えよう。さらに言えば，統合統計メタデータ・アーカイブは，公的統計の体系化，標準化だけでなく，統計調査の品質管理の視点からも重要なものとなる。

統計メタデータの収集，保存は統計データをアーカイブに収録する時点で行うのでは遅く，情報の散逸なく統計メタデータの収集を可能とするためには，統計調査の最初の段階から統計業務過程の進行に沿って，統計データと一体的に収集できるような仕組みを各業務過程に組み込んでおくことが必要である。

いったん作成された統計メタデータをアーカイブに収録するために組み替えたり，不足する情報を補ったりする作業をアーカイブ組織だけに委ねるべきではない。このような作業は労働集約的で時間がかかるものである。分散型の統計機構をとる日本では，統計調査を所管する府省が分担する部分は大きいと考えられるので，各府省が作業を進める上で，収集すべき統計メタデータ，統計メタデータが表す概念などに関する記述規約やガイドラインといった統一的な基準の整備が必要となる。

統合統計メタデータ・アーカイブの整備を行うには，個別統計調査の統計メタデータ・アーカイブを先に構築し，統合統計メタデータ・アーカイブの構築に向かうボトムアップ型アプローチとその逆のトップダウン型アプローチを双方向で進めていくのが実際的だろう。公的統計における統計メタデータ・アーカイブの構築は政府全体として取り組むべき事業と言える。それは，社会の情報基盤として，統計情報がより有効に利用される上で必要な事業だからである。

本稿で行ってきた議論は，行政記録情報を基に作成する業務統計のアーカイブを考える際にも適用できる。ただし，行政記録をコンピューター処理可能な行政記録情報とする取り組みの推進が必要である。また，行政記録情報自体を二次利用のためにアーカイブ化することが可能なのかについては法制度的な検討が必要である。

筆者が考える統計情報アーカイブは，DWHで保存される範囲より広いものである。DWHは，提供を考慮して，文字記号列として表現

される電子化された情報の範囲のみをカバーするものである[26)]。しかし，少なくとも紙媒体の調査票と調査報告書の掲載統計表は実世界の物理的存在であり，そこに記録されている情報は電子化された情報以上の情報量を含む可能性がある。たとえば，調査票の自由記入欄や回答欄外に書き込まれた回答に関係する記述，調査報告書の構成や統計表の配列順など，電子化された情報の範囲では捨象されてしまっている情報である。調査票や統計表をマイクロフィルムなどの光学的記録又はスキャナー入力したイメージデータとして保存することも統計情報アーカイブの範囲として考えてよい。たとえば，符号化される前の産業や職業に関する自由記入は，分類基準の見直しや再構成後の基準による再集計結果を用いた実証研究といった新たな研究分野につながるだろう。調査票は調査票情報と異なり一定年数保存後に滅却されるし，統計表を収録した調査報告書も物理的なスペースが不足すればごく少部（少なくとも 1 セットづつは国会図書館と調査主体の元で）保存されるだけである[27)]。このことは，誰もが，いつでも，どこでも，望みの統計情報にアクセス可能とする上で，将来にわたり考えていくべき課題といえよう。

謝辞

査読者からは多くの有意義なコメントをいただき，本稿の改善に資するところ大であった。ここに記して深く感謝の意を表したい。

注

1）統計業務過程のモデルには必ずしも統一的な認識の合意があるわけではない。実務では，大きく分けて調査設計，実地調査，集計，公表の 4 業務過程からなるとの認識が比較的普及していると言える。各業務過程は，さらに下位の業務過程から構成されている。

2）一般に，当初定められた利用目的（公的統計では調査の目的の中で記述）の範囲での利用は「一次利用」，一次利用以外の利用（当初定められた利用目的以外の目的での利用）は「二次利用」と呼ばれる。

3）本稿では公的統計の中でも中核をなす調査統計の場合を取り上げて議論を展開している。業務統計や加工統計の場合も，統計材料が行政記録や一次統計であるという違いがあるものの，統計を作成する業務過程は調査統計の場合とほぼ同様である。

4）アーカイブの保存対象となるものには，存在のありようによって，実世界の具体的存在物と情報がある（溝口（2012）は，実世界の実在物は具体物，抽象物及び準抽象物に分類でき，情報は準抽象物に含まれるとしている）。報告者が記述した紙媒体の調査票などは具体的存在物である。一方，情報は，通常，電子化された形態で存在する。本稿ではオントロジー的観点を踏まえて調査票と調査票情報，統計表と統計表情報のそれぞれの前者と後者を分けて考えている。調査票情報（に記録されている内容）は，紙媒体の記述済み調査票（に記録されている内容）を情報システムで扱えるように変換（一般的には入力機器を用いて電子的な媒体に入力）したもの（実際には集計で用いる付加項目，たとえば都道府県番号，標本調査のウェイトなども含まれる）である。オンライン調査では記述済み調査票から調査票情報への変換過程が省略され最初から記述済みの調査票情報が作成されていると考えられる。PDF 形式の調査票もオンライン調査票も調査票と呼んでいるが，電子的な形態で存在しているので本稿の定義では調査票情報に当たる。情報を見るには何らかの表現媒体（紙媒体や電子的媒体など）が必要であるが，紙媒体に記録されている情報は，直接視認できるものである。ディスプレイに表示されている情報は，コンピューター内にある情報をディスプレイという機器の画面を通してはじめて視認できるものである。しかし，回答時には調査項目や回答記述は回答者自身が視

認できている点は紙媒体の調査票の場合と変わらないと言えよう。

5）統計表情報は，統計原表に相当するものと考えてもよいであろう。統計表情報を紙媒体の物理的な大きさに収めるためには表頭分割や表側分割をして印刷する必要がある。また，2つの統計表情報を結合して見かけ上1つの統計表とすることがある。それゆえ，統計原表と統計表では表番号が異なる場合が生ずる。紙媒体という物理的制約を受けない統計表情報の方が，表としての様式の自由度ははるかに大きい。

6）調査票と統計表は，情報本体と様式に加えて，これらを実世界で視認可能とする紙媒体から構成される。

7）小林（2019）によれば，様式独立にすると，調査票や調査票情報は質的項目と量的項目の組で表現できる。また，統計表や統計表情報はセルを量的項目，表頭などの分類項目を質的項目と考えることができ，やはり質的項目と量的項目の組で表現できることがわかる。

8）森（2008）はアーカイブに保存すべきメタデータの例としてエディティングの処理プログラムを挙げているが，エディティングの処理プログラムを含む一連の集計プログラムは調査票情報と統計表情報の関係をプログラム言語で一連の手続きとして記述した仕様であり，統計業務過程の中で情報相互の関係を表す情報の例と言える。

9）管見の限りでは，どの統計調査もエディティングに関する情報など一部の情報は調査報告書にも掲載されていない。また，調査員事務などの実務に関するマニュアルは，一般にはアクセスが難しい。

10）たとえば，符号表からは調査票情報に「労働力状態」という項目とその項目値に「完全失業者」があることはわかっても，「完全失業者」の定義を知ろうとすれば調査報告書を参照する必要がある。また，分類項目に「社会経済分類」という項目が使われている統計表情報を，各府省のホームページや総務省のe-Statで提供されている統計表情報の中に見つけたとき，その分類項目がどの調査項目に該当するのかあるいはどのような調査項目どうしを組合せて導出したのかを知ろうとすると，表とは別に調査報告書にある分類項目一覧や用語の解説を参照する必要がある。用語の解説が電子的な情報としてウェブ上に存在していても，統計表情報の分類項目などと関係付いていないなら，表中の数値を見ているときに別のファイルを探して開き，参照しなければならない手間がかかるという不便さがある。

11）統計審議会（1985）では統計調査結果，すなわち集計値の意味で使用されている。統計審議会（1995），各府省統計主管部局長等会議（2003）では，集計結果データの意味と個票データの意味の両義が混在している。統計委員会（2009b）では個票データや匿名化データの意味で使用されている。

12）統計ミクロデータ（statistical microdata）とは「ある個別オブジェクト，すなわち統計単位，に関して収集された観測データ（an observation data）」と定義されており，一方，統計マクロデータ（statistical macrodata）とは「統計的方法論に従って統計ミクロデータをある目的/意図をもって集約すること（a purposeful aggregation）により得られた観測データ」と定義されている。

13）DDIについてはhttp://www.ddialliance.org/what，SDMXについてはhttp://sdmx.orgを参照。

14）DDIとSDMXの比較を行っているものとして，Gregory and Heus（2007），Gregory（2008），Pellegrino and Grofils（2013）などがある。

15）本稿では，一次利用型統計業務過程のモデルを調査設計，統計材料収集，統計編成，分析・公表及び評価の5過程からなるものと考えている。記述済み調査票は統計編成の過程の中でクリーンな調査票情報となる。この調査票情報が二次利用の提供対象となる。なお，本稿で考えている「評価」過程は，統計業務過程全般を対象に包括的，体系的に一連の業務過程（プロセス）とその生成物（プロダクト）の品質評価を行うものである。

16）一次利用型統計業務過程の実行とは独立して統計情報アーカイブに情報を追加することがある。例としては公的統計全体に影響する標準分類の改定が該当する。このような処理は，二次利用者に影響を与える変更だが，不定期に発生することから非定型的な業務と言える。

17）「統計材料」という用語は，統計を作成するための材料という意味で少なくとも明治期には使用され出しており，調査票に限らず行政記録も含む概念と認識されていた（呉（1892）；高橋（1906a, 1906b））。

18）諸外国のミクロデータ提供の歴史的経緯，提供形態，法制度などに関するまとまった調査研究は，

松田ほか（2000）が嚆矢であろう。また，総務省が行ったデータ・アーカイブ組織に関する調査委託研究としては，総務省政策統括官（統計基準担当）（2007）及び三菱総合研究所（2011）がある。前記2資料の中では，社会調査を中心に収集・提供しているデータ・アーカイブ組織の事例として米国のICPSR（Inter-university Consortium for Political and Social Research），イギリスのUKDA（UK Data Archive），ドイツのZA（Central Archive for Empirical Social Research in Cologne）のほか，日本のSSJDA（Social Science Japan Data Archive），JEDI（Japan Educational Data-archive Initiative）などが挙げられている。一方，政府統計では政府統計機関がデータ・アーカイブ組織の役割を担っているのが一般的である。なお，統計委員会（2009）以前と以後では政府文書で使われる表記が「データ・アーカイブ」から「統計データ・アーカイブ」に変わっている。

19）佐藤（1988），佐藤（1995）は，メタデータを用いて統計表（佐藤（1995）は「要約データ」と呼んでいる）のDB設計・構築の実証研究を行っている。これらは，既存の統計表に関して様式を含めた形態でDBを構築しようとしたもので，現実の統計表様式の多様性，複雑性を解消する工夫がなされているものの，管見の限りでは政府統計分野の中で同種の研究が広がることはなかった。

20）DBにおけるメタデータの集りはデータディクショナリ（以下，「DD/D」）と呼ばれる。DD/Dについては，Leong-Hong and Plagman（1982），椿（1982）を参照。

21）DWH及び多次元DBについてはKimball（1996），Inmon（2002）が，またOLAPについてはCodd et al.（1993）が論じている。日本では主に情報処理分野で研究が行われた（たとえば，豊島（1996），田中ほか（1996），田中（1997），石井ほか（1998）など）。池田（2006）はDWH，OLAP及びOLTP（On-Line Transaction Processing）を用いた民間企業のデータ蓄積及び利用について論じている。一方，本稿ではDWHと多次元DBを用いた公的統計の統計情報アーカイブについて論じている。

22）このような機能はOLAPの機能の一つとして確立されており，ドリルスルーと呼ばれている。たとえば，Excel2016のピボットテーブルには，この機能が実装されている。

23）無制限に公開するということではない。たとえば，エディティング規則やそれを適用したときのデータの変遷情報は，統計作成者と統計利用者に対して，公開の可否や公開の程度と範囲に制約があるのが現実である。また，提供時に適用する秘匿方法は公開の対象に含めないのが諸外国でも一般的である。

24）現行の二次的利用制度にあっても，作成したリンケージキーファイルを一定期間保存しておくことは可能である（総務省政策統括官（統計基準担当）（2019b））。しかし，データリンケージの方法論的研究やリンケージデータによる実証研究のための共有財として長期にわたり保存しておくことはできない。

25）オントロジーの基礎的知識，理論及び適用事例などは，溝口（2005），溝口（2012）を参照のこと。來村（2012）は医療など各種分野への応用事例が豊富である。

26）本稿で「電子化された情報」，「電子的な情報」と呼んでいるのは，電子的記録媒体上に存在する情報のことである。調査票情報や統計表情報はこれに該当する。文字には記号としての文字と図形としての文字の2つの側面があり，電子化された情報には文字記号列としての情報と文字図形としての情報の2種類がある。集計プログラムやDWHが扱う調査票情報などは電子化された文字記号列としての情報である。一方，報告者の記述済み調査票などをマイクロフィルムやイメージデータとしたものは文字図形としての情報である。本稿では，利用者に提供する電子化された情報の範囲を文字記号列としての情報に限定して議論を展開しているが，それは既存の一次利用型統計業務過程の情報システムが文字記号列としての情報を扱ってきたこと，及び統計データ・アーカイブと統計メタデータ・アーカイブの構築上の情報システム技術的な理由による。

27）電子的記録媒体の保存期間は技術的に半永久的とは言えないこと，及び記録された情報を読み取る機器がないと内容を知ることができないため機器の技術的寿命に制約されることが，紙媒体が残っている理由と考えられる。それゆえ，アーカイブに保存しておくものは，電子化された情報（文字記号列として表現されたものと文字図形として表現されたもの）だけでなく紙媒体による記録も必要と言えよう。

参考文献

池田伸（2006）「民間企業におけるデータの蓄積と利用 ― マーケティングリサーチ，データマイニング，統計 ― 」『社会科学としての統計学　第4集』第4章，pp.45-57，産業統計研究社
石井義興・豊島一政・木村哲（1998）「データウェアハウス/データマート」『電子情報通信学会誌』第81巻第10号，pp.1034-1041，電子情報通信学会
各府省統計主管部局長等会議（2003）『統計行政の新たな展開方向』
來村徳信（2012）『オントロジーの普及と応用』人工知能学会編，オーム社
呉文聰（1892）「統計ノ本体及問題」『統計集誌』第129号，pp.168-170
国立国語研究所（2003）「第2回　「外来語」言い換え提案」国立国語研究所「外来語」委員会資料，https://www2.ninjal.ac.jp/gairaigo/Teian2/iikae_teian2.pdf（最終アクセス：2020年9月11日）
小林良行（2012）「公的統計ミクロデータ提供の現状と展望 ― 一橋大学での取り組みをもとに」『日本統計学会誌』第41巻第2号，pp.401-420，日本統計学会
小林良行（2019）「統計表の構造とIPF法による教育用擬似個別データの作成方法」『公的統計情報 ― その利活用と展望』坂田幸繁編著　中央大学経済研究所研究叢書 75，pp.23-38，中央大学出版部
佐藤英人（1988）『統計データベースの設計と開発 ― データモデルと知識ベースの応用 ― 』穂鷹良介監修，第1版，オーム社
佐藤英人（1995）「要約データの基礎概念とデータベース内での推論 ― 世界貿易統計データベースを例として ― 」『世界貿易データベースシステムの整備と利用』アジア経済研究所統計資料シリーズ 67，pp.95-106，日本貿易振興機構アジア経済研究所
総務省政策統括官（統計基準担当）（2007）「諸外国の統計データの二次的利用の状況」，第1回統計データの二次的利用促進に関する研究会（平成19年10月22日開催）資料7
総務省政策統括官（統計基準担当）（2019a）『調査票情報等の管理及び情報漏えい等の対策に関するガイドライン』（平成21年2月6日総務省政策統括官（統計基準担当）決定. 平成31年4月19日改正）
総務省政策統括官（統計基準担当）（2019b）『調査票情報の提供に関するガイドライン』（平成20年12月24日総務省政策統括官（統計基準担当）決定.　令和元年6月27日改正）
高橋二郎（1906a）「技術統計論」『統計集誌』第308号，pp.487-496
高橋二郎（1906b）「技術統計論（二）」『統計集誌』第309号，pp.541-552
田中聡（1997）「データウェアハウスと多次元データベース」『情報処理』Vol. 38 No. 9，pp.745-750，情報処理学会
田中聡・木村哲・豊島一政・石井義興（1996）「多次元データベース入門～オンライン分析処理を中心として～」情報処理学会データベースシステム研究会報告103（1996-DBS-110），pp.1-7
調査項目の標準化に関する研究会（2006）『調査項目の標準化に向けて ― 統計の標準化への実践的な第一歩 ― 』（平成18年11月15日「統計調査等業務の業務・システム最適化計画」（各府省情報化統括責任者（CIO）連絡会議決定，2006年から継続中）に基づく決定）https://www.stat.go.jp/info/guide/public/hyoujun/pdf/houkoku.pdf（最終アクセス：2020年9月11日）
椿正明（1982）「データディクショナリ/ディレクトリ」『情報処理』Vol. 23 No. 10，pp.925-930，情報処理学会
統計委員会（2009）『公的統計の整備に関する基本的な計画』（平成21年3月13日閣議決定）
統計委員会（2014）『公的統計の整備に関する基本的な計画』（平成26年3月25日閣議決定）
統計審議会（1985）『統計行政の中・長期構想』（昭和60年10月25日諮問第207号の答申）
統計審議会（1995）『統計行政の新中・長期構想』（平成7年3月10日諮問第242号の答申）
統計改革推進会議（2017）『統計改革推進会議最終取りまとめ』
友安亮一（1957）『統計表の表わし方』，一粒社
豊島一政（1996）「多次元データベースとRDB（OLAPの紹介）」『情報処理学会全国大会講演論文集』，第52回平成8年前期(4)，pp.157-158
松田芳郎・濱砂敬郎・森博美（編著）（2000）『講座　ミクロ統計分析1　統計調査制度とミクロ統計

の開示』，日本評論社
溝口理一郎（2005）『オントロジー工学』人工知能学会編，オーム社
溝口理一郎（2012）『オントロジー工学の理論と実践』人工知能学会編，オーム社
三菱総合研究所（2011）.『統計データ・アーカイブの整備に関する調査研究報告書』総務省委託研究
森博美（2008）「情報資産としての統計と政府統計データ・アーカイブ」『統計学』第94号，pp.15-25，経済統計学会
山口幸三（2019）「改正された統計法と二次的利用の現状と課題」『公的統計情報 ― その利活用と展望』坂田幸繁編著　中央大学経済研究所研究叢書 75，pp.3-21，中央大学出版部
美添泰人（2005）「統計データの保存と再利用の体制」『統計』第56巻第 6 号，pp.32-37，日本統計協会
Codd, E.F., Codd, S.B. and Salley, C.T.（1993）, *Providing OLAP（On-line Analytical Processing）to User-Analysts: An IT Mandate*, Codd & Date Inc., なおCodd & Associatesからも出版されている
Couper, M.（1998）, “Measuring Survey Quality in a CASIC Environment”, *Proceedings of the Survey Research Methods Section of the ASA at JSM1998*, pp.41-49
Dippo, C.S. and Sundgren, B.（2000）, “The Role of Metadata in Statistics”, *Proceedings of the Second International Conference on Establishment Surveys*, pp.909-918
Gregory, A.（2008）, “Status on the Mapping of Metadata Standards: ISO/IEC 11179, SDMX, and Others”, WP.9, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata（METIS）
Gregory, A. and Heus, P.（2007）, “DDI and SDMX: Complementary, Not Competing, Standards”, Open Data Foundation
Inmon, W.H.（2002）, *Building the Data Warehouse（Third Edition）*, New York: John Wiley & Sons,
Kimball, R.（1996）, *Data Warehouse Toolkit*, John Wiley & Sons Inc.，藤本康秀監修（1998）『データウェアハウス・ツールキット』，日経BP社
Leong-Hong, B.W. and Plagman, B.K.（1982）, *Data Dictionary/Directory Systems: Administration, Implementation and Usage*, John Wiley & Sons, Inc., 成田光彰訳，穂鷹良介監訳（1986）.『データディクショナリ/ディレクトリシステム』第 1 版，オーム社
Pearce-Moses, R.（2005）, *A Glossary of Archival and Records Terminology*, the Society of American Archivists, http://files.archivists.org/pubs/free/SAA-Glossary-2005.pdf（最終アクセス：2020年 9 月11日）
Pellegrino, M. and Grofils, D.（2013）, “DDI-SDMX Integration and Implementation”, WP.5, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata（METIS）
Radermacher, W., Baigorri, A., Delcambre, D., Kloek, W. and Linden, H.（2009）, “ANNEX: TYPES OF STATISTICAL INFORMATION”, *Terminology relating to the Implementation of the Vision on the Production Method of EU Statistics*, pp.22-26, Eurostat
Sundgren, B.（1973）, *An Infological Approach to Data Base*, Doctoral Thesis, University of Stockholm
United Nations Statistical Commission（UNSC）and United Nations Economic Commission for Europe（UNECE）（1999）, *Information Systems Architecture for National and International Statistical Offices Guidelines and Recommendations*, Conference of European Statisticians Statistical Standards and Studies-No. 51, United Nations
UNSC and UNECE（2000）, *Terminology on Statistical Metadata*, Conference of European Statisticians Statistical Standards and Studies-No. 53, United Nations

【Special Section: The 60th Anniversary of the Journal】
Special Topic B: Methodological Perspectives in the Creation and Release of Official Microdata

Potentiality of Statistical Metadata Archives in the Official Statistics of Japan

Yoshiyuki KOBAYASHI*

Abstract

This study proposes the construction of integrated statistical metadata archives referred to herein as Integrated SMA and the establishment of domain ontology in official statistics and indicates the potentiality of both.

The statistical information generated through the statistical business process (SBP) is defined as the information comprising both statistical metadata and data with the associated statistical metadata. This study examines an SBP model presupposing the secondary use of statistical information and the properties of the statistical information to be satisfied. Furthermore, it is pointed out herein that the data warehouse technology plays an essential role in constructing the statistical information archives of respective surveys. The statistical data and metadata archives constitute statistical information archives.

The statistical metadata in official statistics can be classified into three layers: global, semi-global, and local metadata layers. The standardization of the terminology on the official statistics and the traceability of statistical information are feasible when the Integrated SMA based on the three-layer structure is constructed. The establishment of ontology on the official statistics can standardize statistical metadata concepts.

Key Words

Statistical metadata, Statistical business process, Data warehouse, Traceability, Ontology

* Statistical Research and Training Institute, Ministry of Internal Affairs and Communications

機関誌『統計学』の編集・発行について

『統計学』編集委員会

みなさまからの投稿を募集しています。ぜひ研究成果の本誌上での発表をご検討ください。

1．原稿は編集委員長宛に送付して下さい（下記メールアドレス）。
2．投稿は常時受け付けています。
　なお，書評，資料および海外統計事情等の分類の記事については調整が必要になることもありますので念のため事前に編集委員長に照会して下さるようお願いします。
3．次号以降の発行予定日は次のとおりです。
　第121号：2021年9月30日
　第122号：2022年3月31日
4．原則として，すべての投稿が審査の対象となります。投稿に際しては，「投稿規程」，「執筆要綱」，および「査読要領」の確認をお願いします。最新版は，本学会の公式ウェブサイト（http://www.jsest.jp/）を参照して下さい。
5．編集委員会は2021年4月から次の体制となります。引続きよろしくお願いします。
　2021年度編集委員会委員長　村上雅俊（関西）
　同副委員長　佐藤智秋（東北・関東）
　同委員　水野谷武志（北海道），山口幸三（東北・関東），西村善博（九州）

投稿，編集委員会についての問い合わせや執筆の推薦その他とも，下記編集委員長のメールアドレス宛に送付して下さい。

editorial@jsest.jp

編集後記

2020年度の日常は新型コロナ発生前とは大きく変わりました。そのような中でも『統計学』の投稿者のみなさま，そしてお忙しい中快く論文の審査をお引き受けいただきました査読者のみなさまに改めてお礼申し上げます。副編集委員長の村上先生をはじめ編集委員の水野谷先生，山田先生，松川先生には，大変お世話になりました。また，『統計学』創刊60周年記念事業委員会は特集の編集ありがとうございました。　（小林良行　記）

執 筆 者 紹 介

小林良行 （総務省統計研究研修所）　武内真美子 （愛知学院大学経済学部）

支部名	事務局		
北　海　道 ……	062-8605	札幌市豊平区旭町 4-1-40 北海学園大学経済学部 （011-841-1161） mizunoya@econ.hokkai-s-u.ac.jp	水野谷武志
東北・関東 ……	192-0393	八王子市東中野 742-1 中央大学経済学部 （042-674-3421） ysakata@tamacc.chuo-u.ac.jp	坂田幸繁（代行）
関　　西 ……	580-8502	松原市天美東 5-4-33 阪南大学経済学部 （072-332-1224） m-murakami@hannan-u.ac.jp	村上雅俊
九　　州 ……	890-0065	鹿児島市郡元 1-21-30 鹿児島大学法文学部 （099-285-7601） matsukawa@leh.kagoshima-u.ac.jp	松川太一郎

統　計　学　No.120　　定価 1,760円（本体1,600円）

2021年3月31日　発行

発 行 所　経　済　統　計　学　会
〒112-0013　東京都文京区音羽1-6-9
音羽リスマチック株式会社
TEL/FAX　03（3945）3227
E-mail：office@jsest.jp
http://www.jsest.jp/

発 行 人　代表者　金　子　治　平

発 売 所　音羽リスマチック株式会社
〒112-0013　東京都文京区音羽1-6-9
TEL/FAX　03（3945）3227
E-mail：otorisu@jupiter.ocn.ne.jp
代表者　遠　藤　　誠

昭和情報プロセス㈱印刷

Statistics

No. 120 2021 March

Special Section: The 60th Anniversary of the Journal

Special Topic B: Methodological Perspectives in the Creation and Release of Official Micro-data

Articles

JSES Activities

Japan Society of Economic Statistics

4-7 会員以外の者，機関等によるウェブ転載申請については，前号を準用するものとする。

4-8 転載を希望する記事の発行時に，その執筆者が非会員の場合には，4-4，4-5項を準用する。

1997年7月27日制定（2001年9月18日，2004年9月12日，2006年9月16日，2007年9月15日，2009年9月5日，2012年9月13日，2016年9月12日一部改正）

『統計学』創刊60周年記念特集掲載号発行規程

『統計学』創刊60周年記念特集論文（以下，記念特集論文）の掲載号の編集・発行作業は，経済統計学会2014年度会員総会の決議にもとづき『統計学』創刊60周年記念事業委員会（以下，事業委員会）が行なう。記念特集論文の掲載号（以下，記念特集掲載号）の発行は，本規程にしたがって処理される。

1．総則

1-1　テーマの確定及び原稿執筆者の選定と資格

特定テーマに関わる論文構成の確定及び執筆者の選定は，企画案と執筆計画にもとづき，事業委員会が行なう。

1-2　未発表

原稿は未発表ないし他に公表予定のない原稿に限る。

1-3　原稿の採否およびレフェリー制の導入について

提出された原稿の採否は，レフェリーによる厳格な審査の結果にもとづき，事業委員会が決定する。レフェリーの選任は事業委員会が行なう。事業委員会は原稿の書換え，訂正を求めることができる。

1-4　執筆要綱

原稿作成は別に定める『統計学』創刊60周年記念特集掲載号執筆要綱にしたがう。

2．原稿の提出

2-1　原稿の締切り

本誌発行の円滑のため，締切り日を設ける。締切り日以降に原稿が到着した場合や，訂正を求められた原稿が期日までに訂正されない場合，掲載されないことがある。

2-2　原稿の送付

原稿は原則として，PDFファイル（『統計学』の印刷レイアウト）を電子メールに添付して事業委員会委員長へ送付する。

2-3　原稿の返却

提出された原稿は，採否にかかわらず原則として返却しない。

2-4　校正

掲載が決定した原稿の著者校正は初校のみとし，内容の変更を伴う原稿の変更は原則的に認めない。内容の変更を伴う変更の場合は，事業委員会およびレフェリーの許可を必要とする。初校は速やかに校正し期限までに返送するものとする。

2-5　執筆などにかかわる費用

投稿料は原則として徴収しない。別刷は，執筆者の希望により，作成するが，実費を徴収する。校正段階で原稿に大幅な変更が加えられた場合，実費の徴収などを行うことがあ

る。

3．著作権

記念特集論文の著作権は経済統計学会に帰属する。詳細は，『統計学』の投稿規程に準ずる。

『統計学』創刊60周年記念特集掲載号投稿原稿査読要領

1．経済統計学会（以下，本会）の機関誌『統計学』創刊60周年記念特集掲載号に掲載する「論文」の査読制度について，この要領を定める。

2．『統計学』創刊60周年記念事業委員会（以下「事業委員会」）委員長に送付された原稿については，事業委員会による第一次審査を行い，事業委員会が別に定める「執筆要綱」に準拠しているかどうかを判定する。

3．「論文」の掲載にあたっては，第二次審査を必要とする。

4．第一次審査を経た「論文」の原稿は，速やかに第二次審査へ付されるものとする。

5．事業委員会は，次の事項を審議決定する。

⑴ 第一次審査結果の確認

⑵ 第二次審査を担当する2名のレフェリーの選任

6．第二次審査にあたるレフェリーは会員から選任する。

7．第二次審査にあたって，レフェリーについては匿名性を確保する。

8．第二次審査における判定は，⑴論文として掲載可，⑵論文として条件付掲載可，⑶掲載不可とし，レフェリーはその理由を明示するものとする。

9．第二次審査でレフェリー間での審査結果が異なる場合には，事業委員会はレフェリーと協議し，掲載の可否について最終的な判断を下すものとする。