

オケージョナル・ペーパー No.36

場所特性変数の付加による個体レコードの拡張について

2012年12月

法政大学

日本統計研究所

場所特性変数の付加による個体レコードの拡張について

森 博美(法政大学・経済学部)

要旨

統計調査の対象となる個人、世帯、事業所、企業といった統計単位はバーチャルな空間の中に浮かんでいるのではなく、時空間さらには個体間の関係といった軸によって張られたリアルな社会の中に存在している。このため、統計単位に帰属する各種情報を写し取ったとされる調査票情報は、個体それ自体に関する情報だけでなく、具体的な時空間さらには個体間の関係といった場によって規定された要素も同時に反映している。言い換えれば調査票情報は、時間(把握時点)、場所(個体存在の場)、そして個体間の関係というフィルターを通した個体情報の実現値に他ならない。その意味で調査票情報は、いわば四重の情報性格を有している。このことは、個体特性そのものを統計的に析出するためには、調査票情報から他の諸情報特性の作用を取り除く必要がある。

近年、統計ならびにその周辺で注目すべき二つの展開が見られる。ビジネス・レジスター等のデータベースの整備と数値位置情報(GPS)の観測精度の飛躍的向上がそれである。前者は、個体ベースでのデータ統合(data integration)のデータ基盤を、また後者は、数値位置情報を介した場所特性情報の個体レコードへの変数としての付加の技術的基盤を整備するものである。

本稿では、社会的存在としての個々の統計単位の反映とされる調査票情報がどのような情報特性を持つかをまず明らかにし、数値位置情報をリンクキーとすることによって、これまで統計においてその存在が事実上無視されてきた場所特性情報を個体レコードに付加することによって調査票情報の情報価値の拡張が可能であることを論じる。

はじめに

統計調査では個人や世帯、企業あるいは事業所といった個体が統計単位(statistical unit)として、また行政行為の遂行過程でも同様な個体が法的単位(legal unit)として把握される。こうした個体は、何よりもまず社会的存在である。それらの存在がまどう社会的性格は、それらの個体が相互に取り結ぶ関係、さらにはその関係を含め、個体の存在を規定する外的環境条件として、時間的、空間的要素に依存する。実査過程を経て調査票情報が得られるが、この調査票情報は、第一義的に当該個体の原単位情報という側面を持つものであるが、それは次式に示すように、同時に個体間の関係、さらには個体存在の場としての自空間による特性要素の作用もまた反映している。

調査票情報 = 個体情報 + 個体間の関係情報 + 時間情報 + 場所情報

これまで統計においては、把握対象となる個体がまどう空間的、時間的規定性については、行政区画や町丁目といった行政区分あるいは調査区、地域メッシュといった地域表章、さらには調査実施時点といった時間情報によって集計結果表に反映されてきた。このことに関して、ある特定の時点において把握された特定の地域に該当する集計結果表が果たして統計単位がまどう空間的、時間的規定性を十全に反映し切れているのであろうかという素朴な疑問が、今回の考察のそ

もその動機である。なお、以下では、調査票情報の予想のうち、特に空間(場所)的規定性に焦点を絞って考察を進めることにする。

1. 場所特性顕在化の背景

(1) マクロ統計からマイクロ統計へ

これまで統計、特に政府統計の分野では、「統計＝集計量(マクロ統計)」とする統計観が支配的であった。そこでは、統計論理ならびに社会的制度的論理に関係した次のような二つの事情が、このような統計観の妥当性の根拠となっていた。

第 1 は、調査によって得られる個々の調査票情報は本来的に誤差を伴うものであり、値それ自体が揺らぎの要素を持つとの認識をその根拠としている。統計に不可避的につきまとうこうした揺らぎの要素は個体情報を統合することで相殺され、安定的な結果が得られるというのがそれである。

もう一方は、統計に係る秘密の保護という社会的制度的側面に関するものである。統計は、報告者から提供される個々の統計単位に関する原単位情報をその情報源としている。従って、報告者による調査協力を継続的に得るためには、提供者は調査への協力に起因するいかなる不利益も排除されなければならない。そのためには、統計調査に係る秘密の保護が調査の円滑な遂行、高品質の原単位情報の収集の前提となる。集計量が、統計に係る秘密保護要件の充足にとって適切なデータ形式であるというのがそれである。

1970、80 年代以降、国内国外を問わず、調査実施環境の悪化や行政事務の効率化要請の一環としての統計予算・機構のスリム化など、統計作成をめぐる条件は厳しさを増している。そのような中で、各国の政府統計機関では、業務記録情報の統計への積極的活用を含め、既存情報資源の積極的活用より、日々拡大、多様化している統計ニーズへの対応をはかるという方向を強めている。

このような中、行政記録情報も含めた多様な情報源情報を有機的に連結してその有効利用をはかるためには、個体ベースでのデータリンケージによる利用可能な変数の拡張、データフュージョンによる疑似的なデータ拡張等を含むデータ統合(data integration)が、新たな統計の在り方として次第に形をとりつつある。

このようなデータ統合という視角から既存の集計結果表を改めて捉え直してみると、属性別あるいは地域別集計といったマクロ統計の一連の集計結果表も、見方によっては集計量ベースによる疑似的なデータ統合(macro-based data integration)とも言うことができる。例えば、『社会生活統計指標』やメッシュ統計は、都道府県という行政区分あるいはメッシュという集計単位を疑似的な個体とみなせば、一定の空間的広がりを持つものとして切り取られた集計単位について、種々の統計指標レイヤー群からそれぞれの集計単位に該当する集計値や平均等の導出統計量をデータ本体(data body)として統合したマクロレコードとみなすことができる。さらに敷衍すれば、性、年齢、学歴、資本金、従業員といった属性別集計表も同様に、そのような属性集団を疑似的な個体とみなせば、各調査から得られる集計量をデータ統合したものといえる。

なお、このようなマクロデータ統合に特徴的なデータ特性として、集計単位に属する個体の変数値を集計することに伴う平均化作用によりデータ統合の感度が鈍くなることがあげられる。このことは行政区分あるいはメッシュ等の地域表章においても同様であり、小地域レベルでは確認でき

ていた特異性が、広域集計としての地域平均や都道府県や市区町村レベルでの結果表では平均化されその中に埋没することがしばしば起こりうる。

1970、80年代以降、多くの国において、企業や事業所(local unit)等を統計単位としたビジネス・レジスターが整備され、また世帯レジスターについても北欧諸国やオランダ等ではすでに整備され、海外の主要国においても人口センサスの精度改善の一環としてその整備に向けての取り組みが進められている。

この種のレジスターは、基本的に調査基盤情報としてのフレーム整備を主たる目的として整備されているものである。しかしこれらのレジスターは同時に統計体系論の視点から捉えた場合、各種の調査から得られる統計票情報、さらには行政機関から提供される行政記録情報を文字通り個体ベースで統合(micro-based integration)したリレーショナル型データベースの基幹情報(backbone)としての役割も担わされている。レジスターの整備に伴い、各国の政府統計については、個体ベースでのデータ統合が次第に現実味を帯びつつある。

(2)位置情報観測精度の向上

冒頭にも指摘したように、調査票情報には、統計単位としての個体が持つ社会的性格の故に、三つの外的要因も同時に反映している。近年、これらのうち空間的規定性(場所特性)をこれまでとは全く異なる形で統計に反映できる情報的基盤が整備されてきた。全地球測位システム(Global Positioning System:GPS)が与える経緯度情報の統計への活用がそれである。もともとGPSは1970年代に軍事技術の一つとして開発された。そのため、民生利用向けには意図的に実際の観測精度よりは精度を劣化させた経緯度情報が提供されてきた。またデータ取得に必要な端末も高価であった。これらが隘路となって位置情報の民生利用には多くの制約があった。

しかし、一方で2000年5月の提供されるデータ精度の改善、観測可能な天頂衛星数の増加、携帯基地局等の情報も活用したマルチ測位による観測精度の飛躍的改善がみられ、他方で携帯電話にGPSソフトが標準装備されるなど比較的低廉なGPS情報取得端末の広範な普及が契機となって、民生用途での位置情報の利用可能性が一気に広がった。

従来から、統計単位的位置情報については、多くの統計調査で住所(所在地)として調べられてきた。その意味では統計単位それ自体の存在の場に関する情報は調査票情報の中に存在していた。しかし、住所(所在地)情報が記述的な文字情報であることから、それらが集計目的に用いられることはなかった。調査結果の表章に際しては専ら行政区分、調査区、メッシュといった空間区分が集計単位として用いられ、住所情報が潜在的に保有していた統計目的での価値にはこれまで何の関心も向けられてこなかった。

上述した集計量から個体レコードへの統計観の転換とGPSの観測精度の飛躍的改善ならびに位置情報取得の容易化は、統計におけるそれまでの場所的規定の取り扱いを一変させる契機を内在させているように思われる。なぜなら、統計単位存在に係る空間的位置情報を容易にデータ処理可能なデータ形式で個々の調査票情報と1:1で関係づけることが可能となったからである。住所というアナログな文字情報を経緯度情報という数値処理可能なデータ形式に変換することで、経緯度情報を媒介した個体ベースでのデータの相互のリンケージが可能になるだけでなく、各統計個体がどのような場所特性によってどの程度規定されているかを解明できる諸変数を既存の個体レコードに取り込むことが可能になる。

以下本稿では、①統計単位としての個体を規定しうる場所特性、②経済学と地理学が個体の

場所からの規定性あるいは場所特性についてどう取り扱ってきたか、③統計情報(調査票、行政記録情報)の多面的性格、④統計における場所の取り扱いとその帰結、⑤個体レコードの構造と拡張、⑥拡張された個体データが拓く新たな分析価値、といった点に関する考察を試みる。

2. 統計単位とその場所による被規定性

統計調査ではしばしば個人や世帯、それに事業所や企業といった個体が統計単位として調査対象とされてきた。また、政府(中央・地方)も、行政事務遂行の過程で、個別行政法規を根拠として、行政による把握単位としての個体(法的単位)から各種の報告を申告あるいは届出等の形で徴収し、あるいはこれらの活動を独自に記録することによって様々な行政記録情報を作成してきた。

統計単位あるいは行政単位として把握される個人、世帯、事業所、企業といった個体は、一般に常住地(所在地)を拠点に様々な社会・経済活動を行っている。これらの中で特に世帯や事業所については、その所在地を移転させる場合もないわけではないが、その存在の場は、通常、常住地(所在地)にある。例えば家計消費統計に反映される世帯の購買行動は、世帯規模、世帯のライフステージといった世帯属性だけでなく、交通利便性や商業施設等への時間距離といったその世帯の常住地の置かれた場所の場所特性にも当然影響を受ける。また事業所が展開する事業活動は、従業員規模や保有する技術水準といった自らの経営資源に関わる属性要素だけでなく、同業他社との競争性やその地域の事業所の集積度、さらには物流に関わる交通利便性といった立地環境特性もまた事業活動の成否あるいは効率に関わっている。

世帯や事業所といった統計単位としての個体と異なり、個人の場合には基本的に動的な存在である。しかしながら、個人についても、どのような立地環境特性を持つ場所に常住しているかは、その者の意識や行動に少なからず影響を及ぼしているものと考えられる。同じ無業者であっても、交通至便で雇用環境に恵まれた地域に居住しながら無業状態にある者と諸事情により就業機会の少ない地域に居住している無業者とでは、その層には自ずと差異があるであろう。逆に就業行動の面で同様の個人属性を持つものであっても、居住地域における労働市場の条件の違いによって就業確率は当然異なるはずである。なお企業については、単一事業所企業である場合、その所在地における場所特性が事業所としての活動を制約し、その結果が企業にも反映されるであろう。これに対して複数事業所企業の場合には、事業活動の展開は各事業所で行われ、本社(本所)では専ら投資等の経営判断といった司令塔的機能を遂行する。本社の所在地も会社のブランドイメージや情報収集環境といった点で企業としての事業活動に関係するが、事業所に比較すれば、企業の場合、所在地が持つ場所特性に制約される側面は相対的に小さいと考えられる。

3. 経済学・地理学における場所特性の取り扱い

(1) 経済モデルにおける場所特性の取り扱い

基本的に集計量として捉えられてきたこれまでの統計では、個々の個体の場所に関する情報は集団の中に埋没し、都道府県、市区町村といった行政区分、町丁目や調査区あるいはメッシュ統計といった小地域による地域表章として取り上げられてきた。等しい地域特性変数値をもつ空間的広がり(ポリゴン)がこれらの境域と一致するケースはむしろ稀で、両者の境域が非整合的である

のがむしろ一般的であると考えられる。その意味で、既存の地域表章は個体に関わる地域特性情報の統計への反映という点で有効に機能してきたとは言い難く、従って、集計量に基づく分析では個体特性の中に地域特性によって制約された要素が混在しており、後者をコントロールした分析とはなっていない。

近年わが国でも、学術研究・教育目的での利用のために匿名リサンプリングデータとしてマイクロデータが作成、提供されるようになった。マイクロデータでは、個人、世帯レコードについても、個体が識別されるリスクを考慮して、調査票に記載されている住所情報は削除され、また行政区分についても大幅な統合が行われている。このため、統計法第 33 条(行政機関等による使用の場合は第 32 条)による調査票情報の利用を除けば、分析資料としての個体レコードは、個体の存在に係る場所情報としては限定的な情報しか持っていない。このような事情から、個体レコードの変数値に匿名化措置を施して作成されたマイクロデータを分析資料として用いるマイクロ経済分析では、分析対象としての個人、家計、あるいは事業所といった個体は、個々の個体が存在する場所とは完全に切り離されている。他方、統計法第 33 条による識別個体レコードを分析資料とした従来の分析では、疾病に対する疫学的研究を除けば個体が現実に存在する場所特性に対して関心が払われることはなかった。少なくとも経済学の分野では、場所特性を独立変数として加えたモデルアプローチはなされてこなかったように思われる。

現実には独立変数も従属変数もその個々の変数値はそれらの存在する場所特性要素の作用というフィルターを通した実現値に他ならない。しかしながら、一方では利用可能なデータの欠如という資料制約、他方では分析者の関心の欠如といった事情もあり、これまで経済学の分野で構築されてきたマイクロ計量モデルは、事実上、個々の存在の場から切り離され、いわば無重力空間に浮かんだかのような個体を分析対象として取り扱ってきた。その結果、モデルで採用された諸変数には個体によって異なる場所特性が混在しているにもかかわらず、調査票情報の各変数値が全面的に個体特性を反映した情報であるかのような分析が行われてきた。

(2) 地理学における場所性への関心

人文・社会系の地理学では、分析の対象となる個体あるいは集団の空間的分布、空間的関連性、種々の指標を用いた場所特性による地域の類型化といったような分野での研究に中心的な関心が注がれてきた。特に地域の類型化については研究者の関心の向きは多様であり、類型化を行うにあたって採用される指標や類型方法も異なる。その結果、様々な問題関心や分析視角に応じてそれぞれ独自の諸類型が提案されてきた。しかし、個々の問題関心や分析視角が異なることから、析出された地域特性は相互に関連性を持たないそれぞれ独自の類型として併存し、類型相互間の関連性や体系性を取り上げる研究は、この学問分野においては主たる関心事項とはなっていないように思われる。

ところで、個々の地域特性指標による地域の類型化を個人や世帯、事業所といった統計単位としての個体の側から見た場合、個々の地域特性指標はそれぞれ境域を異にするポリゴンから構成されるレイヤーの集合として捉えることができる。地理学は、自然的、社会的、政治的、制度的、経済的、文化的等々の分野について様々な境域区分を持つポリゴンを地域構成要素とするレイヤーの存在を提案してきた。

その一方で地理学においては、こういったレイヤー相互間の関連性やレイヤー群の構造、さらには個体の存在の場所と関係づけられた各レイヤー内の場所特性値間の重層的関係、さらにそ

れぞれの場所特性が社会的存在としての個体の活動、行動、それに意識等をどのように規定しているかという分析視点は、これまでどちらかといえば希薄であったように思われる。

4. 統計調査と場所情報－調査票情報の多面的性格

これまで統計は、実査過程で調査票が個体に関する記録として写し取った調査票情報を、地域、個体属性、さらには種々の関心事項とクロスさせ、その結果を集計表の形で提供してきた。

調査票情報はその対象となる個体の存在場所に関する情報として、都道府県、市区町村、調査区といった各種のコードを持っており、それらを用いていろいろな地域階層レベルでの集計結果表が作成されてきた。また、調査区による集計結果をベースにして追加的な同定作業を施すことによって地域メッシュ統計なども作られている。その意味では、既存の表章形態では、個々の調査票情報が集計量という形で面としての場所と関連づけられている。

ところで、統計の調査対象としての個体の在りかたを制約する要因は、こういった行政区画レベルあるいは調査区といった境域によってすべて代表されるわけではない。こういった境域内の全個体に対して特定の要因が同じ強度で影響力を及ぼすことはむしろ稀である。各個体は、様々な場所的広がりと強度を持って各個体に対して影響力を及ぼす諸要因の中に身をおき、他の個体や集団などと多様な関係を取り結びながら存在し、活動している。

ここで、事業所間の競合関係という具体的事項を参考に、個体の存在を規定している場所特性の統計的把握について考えてみよう。

これまでの調査では、ある事業所が他の事業所と競合関係にあるかどうかというそれぞれの個体の存在を規定する場所特性について、あくまでも調査対象としての個体の側からその把握が行われてきた。例えば、平成9年、14年「全国物価統計調査」では、調査対象の店舗に対して、近隣地域における競合店舗の有無に関して、「価格競争を意識する競合店がある」と回答した店舗に対して、競合店の業態とその店舗までの距離を100m未満、100m～500m未満、500m～1km未満、1km以上といった距離帯別に調査している。

ところで、このような競合店舗の有無については、見方を変えて当該事業所(店舗)が立地する地点の場所特性という側面からもそれを捉えることができる。すなわち、個体に係る住所情報を手掛かりにして、既存の店舗から所定の規模のバッファを発生させ、その域内にその店舗の所在地点が含まれるか否かによって、事後的なデータ処理によってそれを判断するというのがそれである。この方式の場合、前者の調査による把握と異なり、報告者には何らの追加的報告負担を求めることなくこの種の情報の取得ができるだけでなく、店舗間の時間距離についても、単に報告者の主観的認識としての直線距離ではなく、徒歩あるいは車による厳密な移動時間距離としてその計測を行うことができる。

ここでは店舗間の競合関係の計測を一つの例示として住所情報を介しての場所特性情報の事後的把握の可能性を見てきた。その場合、住所情報というアナログな情報を経緯度情報というベクター情報に変換することによって、調査票情報には含まれていなかった個体に関する新たな情報の作成が可能となる。このような試みは他にも見られ、例えば経済産業省では「商業統計調査」における店舗の立地環境特性に関して、平成11年までの調査では調査区単位で行っていた立地環境特性の格付けを平成14年調査からは事業所(店舗)単位で行うことで、本稿末尾に【付表1】として掲げたような店舗が立地する立地環境特性の区分を行っている。

個体そのものの属性はともかくとして、その意識や行動、活動の結果には、好むと好まざるとにかかわらず、場所特性による作用が投影されている。言い換えれば、調査票情報のそれぞれの変数値は、実はこういった場所特性というフィルターをくぐってはじめて形をとった実現値であるといえる。調査票情報には個体そのものに係る個体特性情報とそれに対して規定要因として一方向的に影響を及ぼす場所特性情報要素とが実は混在しており、これまでの統計(学)は、これらのうち全面的に前者を反映した情報が調査票情報であるとして取り扱ってきたのである。

店舗の競合性に関する事例が暗示しているように、場所特性に関しては、個体の所在に関する経緯度情報さえ与えられていれば、事後的データ処理によって、いわば後づけ的に既存の調査票情報に付加することができる。この点に関しては、後に経緯度情報による個体レコードの拡張として改めて論じることとする。

5. 統計調査での場所把握と地域表章

(1) 統計調査での場所把握

(i) 実査の制度的装置としての調査区

国勢調査の『調査区関係資料』(総務省統計局)は、調査区設置の目的を、「国勢調査員の担当区域を明確にし、調査の重複・脱漏を防ぎ、もって調査の正確性を期するとともに、調査結果の集計及び各種統計調査実施のための基礎資料を得ること」と規定している。また、平成 21 年経済センサス基礎調査も調査区設定の目的を、「調査員の担当区域を明確にし、調査の重複、脱漏を防ぎ、調査の正確な実施を期する」としている。

国勢調査の調査区は、おおむね 50 の世帯から構成され、各市町の区域において町丁・字が原則として 1 調査区となるように設定することになっている。しかし、実際には地域によっては町丁・字の区域の境界が複雑に入り組んでいたり飛び地が存在する場合がある。このような地区では、実査に際して調査員が自らの担当地域を認識することが困難になり、また調査員の行動範囲が広がることも予想される。従ってこのような場合には、町丁・字とは別に調査区が設定される。一方、経済センサスでは、町丁・字内の企業数が 50 を超える場合には、企業数がおおむね 30~40 となるように町丁・字を分割してそれぞれ単独の調査区とされている。このように、調査区は、調査員の業務量を勘案し、基本的に調査における調査対象把握の完璧を期するために調査対象地域を網羅的にカバーした実査の基盤的枠組みとして設置されている。

しかし、上述の国勢調査の調査区の説明文でも指摘されているように、それは国勢調査以外の調査も含めた調査実施の制度的基盤であるだけでなく、調査結果の集計の基盤的枠組みとしても捉えられている。国勢調査の調査員の選任が市区町村の業務であるという事情に加え、市区町村あるいは都道府県といった行政区画による表章と整合性をもたせた形で調査区の境域設定が行われている。経済センサスについても調査区は原則として町丁・字レベルで設定されているが、それらが上位の地域行政区分との整合性が持たせてあるのは国勢調査の調査区と同様である。

(ii) 地域集計の基礎単位としての調査区

調査票には、調査事項の他にも、報告者(記入者)情報記載欄、調査員記入欄、調査対象の識別番号などとともに、調査区コード、町丁・字(町丁・目)コード、市区町村コード、都道府県コードが、また多くの調査票には、調査の対象となった統計単位の住所(所在地)記載欄が調査事項として設けられている。調査結果の地域集計に関しては、現在、一般には住所情報ではなく、このよ

うな各種の地域コードがキー変数として用いられている。

調査票情報は、それぞれの調査の対象となる統計単位を住所(所在地)という地点(ポイント)において捉えている。にもかかわらず、調査票情報回収後の統計作成過程では、調査票が持っていた具体的な地点情報要素は事実上削ぎ落とされ、調査区あるいは行政区分といった面的広がりを持ったポリゴン情報へと縮約され、それらに基づいて地域別表章は行われている。このような統計情報の取り扱いの結果、統計単位である個体の存在する地点にもともと関連づけられていた調査票情報は、具体的な地点としての場所情報から切り離され、ポリゴン構成要素の一つとして位置づけ直され、ポリゴンコードによって代表される。このような場所情報のポリゴンコードへの縮約処理の過程で、調査票情報が本来保有していた個体の場所情報に関する本質的な情報喪失が発生している。

しかし、「統計＝集計量」という統計観に基づいてこれまで行われてきたような結果の集計量による地域表章という限りでは、この種の情報喪失が内在的に持つ問題が表面化することはない。行政区分、町丁・目、あるいは調査区といったポリゴンに属する統計単位としての個体に関する情報を集計することによる個体データが持つ情報のゆらぎの相殺効果ならびに統計に係る秘密の露見リスクの回避という二重の意味で、場所情報の面(ポリゴン)的把握は、むしろ積極的な意義を持ってきた。

(2) 場所情報のポリゴンによる縮約の問題点

ここでは、調査票情報が地点として本来有していた位置情報のポリゴンによる縮約の過程で発生する情報喪失が統計利用においてもたらす種々の問題について検討する。

(i) 多様な地域区分の存在

現在、集計結果の地域表章は、都道府県、市区町村といった行政区分、町丁・字や調査区といった小地域統計での地域表章が一般に行われている。しかしこういった階層化された地域表章では対応できない多様な統計利用者の利用ニーズが存在している。

地域経済圏、商圏、ニュータウンのような居住圏は行政境域を跨がった形で広がっており、上下水道、電力といったライフラインや鉄道・バスといった公共交通機関が形成する通勤圏、河川の流域圏なども行政区分による境界を越えて、線あるいは面状に広がっている。他方で、広域避難所や地域コミュニティ施設が提供するサービス圏域、学区、町内会、郵便番号、商業施設や公共交通機関への *accessibility* といった既存の行政区分による境域の一部を切り取った形で設定される区域による結果数値についても種々の統計ニーズが存在する。これらの中には町丁・目、あるいは調査区レベルまで遡及することで対応できるものもあるが、調査区と必ずしも整合的でない境域設定が必要なケースも少なくない。

(ii) 境界変更による比較可能性の喪失

昭和や平成の大合併のように、わが国では政府主導でしばしば自治体の境域の再編が行われてきた。この他にも自治体では、市への昇格や政令指定都市の指定要件を充足するためにしばしば周辺の市町村を吸収合併するケース、あるいは人口増加に伴う分区といった形で境域の変更がしばしば行われてきた。こういった行政区画だけでなく、新たな道路の開通や住宅団地、集合住宅等の建設によって調査区そのものもしばしば見直される。調査実施に先立って調査区地図の整備が行われるが、いくつかの調査区では境域が再設定される結果、調査区レベルにおいても必ずしも調査結果の時系列比較可能性は担保されていない。

調査区の境域の変更による統計の時系列的な比較可能性の喪失に対処するために、わが国では平成2(1990)年国勢調査の調査区設定に際して、「小地域統計の時系列比較分析ばかりでなく、多様な地域区分による小地域統計の作成をも可能とする」とのふれこみで、調査区の下位区分で「調査結果の集計上の恒久的かつ最小の地域単位」〔総務省統計局 2005 5頁〕として基本単位区が導入された。

基本単位区は、時系列比較可能性の要件を前面に出した調査結果の利用から出発した区域概念である点で、主として調査実施基盤として設定されてきた境域概念である調査区とはやや質を異にしている。それは「固定的な境域として恒久化されて意味を持つ」〔(同上) 5頁〕ことから、そのポリゴンは、住居表示に関する法律(昭和37年法律第119号)に基づく街区又は街区に準じた境域として設定され、市町村境界の変更、街区方式による住居表示の実施、町・字界の変更等による市町村内の住居区画の変更、それに区画整理、地域再開発、水面埋立て、道路・河川等の新設・改修、災害等による地形・地物の著しい変化がない限り将来においても変更しないもの文書上では規定されている〔(同上) 5頁〕。

しかし、基本単位区についても、その境域は一定時点における地形や地物の状態を反映して設定されることから、それが恒久的な地域区分として維持されることはありえない。事実、それが変更されたケースも少なくない。さらに調査票情報に影響を及ぼしていると考えられる個々の場所特性との関係で言えば、基本単位区が調査区の下位区分として相対的に微小な境域をもつポリゴンであるとはいえ、依然として面的情報には変わりがない。そのため個々の基本単位区は一般に複数の統計単位を構成要素として持ち、それらの調査票情報が集計量として一つの基本単位区コードに紐づけられている。また仮にそれらを個体レコードという非集計量として捉えた場合にも、各レコードはそれがもともと保有していた場所情報ではなく、基本単位区コードを共有している。さらに、単一の統計単位しか持たない基本単位区であっても、それが持つ場所情報は基本単位区コードによって代替されている。その意味で、ポリゴン内に存在する統計単位の個数こそ異なるものの、調査区と同様に基本単位区でも、一つのポリゴンコードが一般に複数の個体情報を背負うという関係が維持されている。そこでは、1対1にせよ1対複数にせよ、特定の経緯度情報と調査票情報とが直接的に紐づくという関係は解消されている。

(iii) 地域メッシュ統計

地域メッシュ統計とは、地域をメッシュといわれる方形に区分し、個々のメッシュ(グリッド)内に所在する統計単位に関する調査票情報を、総計や比率といった指標(集計量)として表示する表章方式である。地域メッシュを地域表章単位とすることのメリットについては、①面積がほぼ一定でありメッシュ間の相互比較が容易であること、②行政区域の境域変更や地形・地物などの変化の影響を受けることなく時系列比較が容易であること、③地域内のメッシュデータを集計することで必要な地域についての統計を得ることができること、④任意の地域間距離に関連した比較・分析が用意に行えること、⑤位置情報が明確であることから地図化が容易であること、などが指摘されている〔総務省 2006 13頁〕。

政府は1960年代に地域メッシュ統計の基盤としての地域メッシュ画定のための研究を省庁間共同プロジェクトとしてスタートさせ、1970年には「標準地域メッシュ体系」が告示された。その後地域メッシュ統計については、1973年に「統計に用いる標準地域メッシュおよび標準地域メッシュ・コード」が告示され、また1976年にはその作成に係る統一基準が「JIS C 6304」として策定された。このような動きと併行して、1960年代後半には、センサスあるいは大規模標本調査に基づく地域メッ

シュ統計が整備された。メッシュによる統計表章の基盤となるのが、表1に示したような一連の区画体系を有する地域メッシュである。

表1 地域メッシュの区画体系と基準地域メッシュ

区画の呼称	一辺の長さ	区画数	区画コード	備考
第一次地域区画	約80km	176	4桁	
第二次地域区画	約10km	4,885	4+2桁	第一次地域区画の一辺を8等分
第三次地域区画 (基準地域メッシュ)	約1km	386,877	4+2+2桁	第二次地域区画の一辺を10等分

表1中の第三次地域区画(基準地域メッシュ)からは、その一辺を2等分、4等分、そして8等分した一辺約500mからなる1/2地域メッシュ、同じく250mの1/4地域メッシュが、さらに人口集中地区については、1/8地域メッシュ(同約125m)も作成されている。なお、これら各階層レベルの地域メッシュにはそれぞれ1桁のコードが追加的に付与されており、例えば現時点でメッシュの最小表章単位である1/8地域メッシュは、全体で11桁からなるメッシュ・コードを持つ。

地域メッシュ統計は、電子地図、メッシュ枠情報、国勢調査区、基本単位区の境界情報、それに事業所等の所在情報、さらには基本単位区ベースでの集計量情報を用いて作成される。メッシュ統計の作成には、かつての事業所・企業統計調査では調査区が、一方、国勢調査の地域メッシュについては基本単位区による集計値が用いられている。メッシュの境界線が境域を2つ以上に区分している調査区あるいは基本単位区については、同定によって集計量の各グリッドへの按分が行われる。なおその同定には、例えば、国勢調査メッシュの場合、面積割合同定、住宅建物同定、人口分布点同定、事業所建物同定、図心同定、といった方法が用いられている〔同 23 頁〕。なお、事業所に関する地域メッシュの作成の場合、調査された事業所の所在地情報(住所)をアドレスマッチングによって経緯度情報へと変換し、経緯度情報を用いて事業所の該当メッシュへの同定が行われている〔総務省 2006 1 頁〕。

メッシュという統計の表章形式は、メッシュという共通のプラットフォーム上で統計データや行政情報を接合、加工することで防災地図や危険度マップ等の作成にも活用されている。そこでは、メッシュといういわば物理的に設定された区画が、市町村合併のようにその時々での行政区画の改変に左右されることなく、時系列比較あるいは空間的比較にも対応可能である。その一方で、地域メッシュ統計については、逆に境界区分が行政区画に対応していないことから、具体的な行政業務の遂行のための情報基盤としてはいま一つ適合的でないという利用上の難点を持っている。このため地域メッシュ統計は、調査区等の面的に統合集計された統計から作成する場合には、同定に関わる膨大な事務的作業量を必要とする割には、統計の表章形式の一つとしては、その後必ずしも広範に普及したわけではない。

(iv) 災害の被害評価と位置情報の必要性

大雨、高潮、津波などによる浸水・水没被災地域、火災による延焼地域の広がり、海拔高度や堤防の状態、降雨量や津波の高さ、風向きや風力等によって異なる。このような被災地域の広がり、調査区あるいは行政区の境域とは自ずと独立であり、迅速な被害規模の確定や事前の被

害予測のためには、これまでのようなポリゴンをベースとする面的な集計量情報では対応できない。こういった統計ニーズに対応できる情報の形態としては、個々の事業所や住戸等と関連づけられた全住民位置情報のような非集計量によって与えられる情報が必要である。

2005年8月に合衆国南部を直撃した巨大ハリケーン「カトリーナ」は、ニューオリンズ市を中心に各地に甚大な人的、経済的被害をもたらした。ニューオリンズ市における水没による被害について、米国商務省センサス局は、一部推計も含め個々の事業所の経緯度情報を取得し、事業所個人情報とリンクさせることで、水没によって失われた雇用数、取引額等の経済的被害額の推計を行った【Jarmin S.Ron and Miranda J. 2009】。この研究は、その推計の迅速性ととも、経緯度情報と調査票情報とを結合した被害評価分析の嚆矢として注目されている。

(v) 弾力的なゾーニングへの対応の困難

上記の(i)以外にも統計のユーザーはそれぞれ独自の利用目的に応じて境域を設定し、分析を行うことがある。自治体では地域コミュニティ施設や公共避難場所等の最適な設置場所を決定するのに、既存の施設を中心とする徒歩あるいは乗用車等でのアクセス可能境域を設定し、サービス享受住民の規模の比較あるいは境域外人口の分布等に関する計測、さらには既存施設の統廃合の住民サービスへの影響評価を行うといったような政策評価ニーズを持っている。人口の高齢化の進展に伴い、いわゆる買い物難民や医療施設難民の地域別分布の検出などにも既存の施設からの時間距離等の算定といったニーズの適正な把握も重要な行政課題となっている。

一方、民間においても、流通事業者は公共交通機関による交通移動時間等も考慮した潜在的集客能力や既存の他の店舗等との競合度あるいは集中立地のメリットといった様々な要素を考慮した上で出店を決定し、製造事業者も工場の進出、移転場所の選定にあたって、企業側では実に多様な要因を考慮してその決定を行っている【付表2】。

こういった分析ニーズは、それぞれの分析目的によって異なる固有の設定境域を持つ。そこでは、都道府県や市区町村といった行政区分による集計結果はもとより、町丁・字や基本単位区、あるいは地域メッシュ統計といった小地域統計で採用されている境域も、こういった利用者がそれぞれの目的に応じて設定する境域とは一般には対応していない。既存の地域区分を分析目的に対応した境域に調整するためには、地域メッシュ統計の作成の際の同定作業のような膨大な手作業が必要となる。

1 対複数の照応も含め、経緯度情報と個々の調査票情報とが関連づけておりさえすれば、強力な空間集計、空間解析機能を用いることによって、専ら技術的なデータ処理としてそれを遂行することができる。

6. 調査票情報と個体レコード

(1) 調査票に記載された情報

調査票には、本来の調査項目の他にも、調査実施者側で事前に記入する調査区情報、調査対象となる統計単位の識別情報、さらには調査員が目視による観察結果を記載する欄等が設けられている。以下の図1、図2は、世帯調査の調査票、事業所調査票の調査票情報をレコード形式としてそのイメージを要約的に示したものである。なお、図中のデータ本体情報とは、フェース項目以外の調査項目、すなわち data body 部分を意味する。

7. 位置情報と個体レコード

(1) わが国における住所表記

調査票情報には、本来の調査項目の他にも、名前や事業所等の名称、世帯や事業所の住所（所在地）、報告者の電話番号や氏名といった情報も記載されている。こういった項目は調査票に記入された調査票情報の内容審査過程で、記入漏れや矛盾した回答内容の照会のための連絡情報として利用されるもので、それら自体は集計の対象となる統計項目ではない。このためこれまでの統計学では、こういった集計対象外項目に関してはこれまでほとんど関心が払われてこなかった。

ところで、これらの項目のうち住所情報は、その活用方法によっては、調査票情報に新たな利用可能性を拓き、また調査票情報が内在的に含み持っている問題の解決方向の手掛かりとなる重要な契機を含んでいるように思われる。そこで以下では、調査の対象である統計単位の住所情報の活用面に焦点を当てて考察してみたい。

住所情報については、「住居表示に関する法律」（昭和37年法律第119号、最終改正：平成11年12月22日法律第160号）の第2条に、（住居表示の原則）として、「市街地にある住所若しくは居所又は事務所、事業所その他これらに類する施設の所在する場所を表示するには、都道府県、郡、市、区及び町村の名称を冠するほか、表2に示したようないずれかの方法による」として、次の二種類の住所表示方式が規定されている。

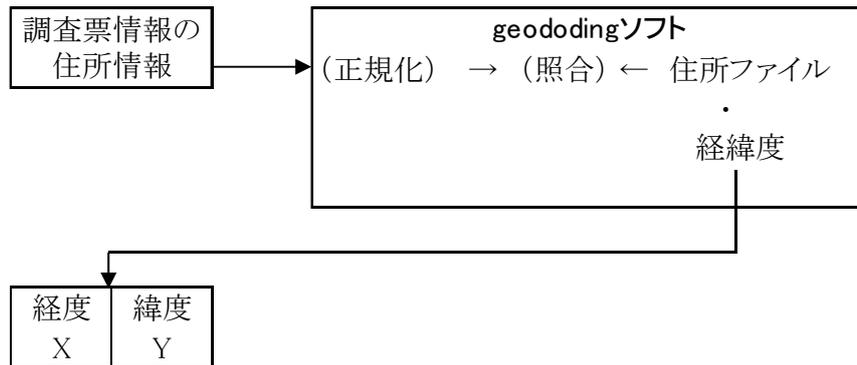
表2 わが国における二種類の住所表示方式

街区方式	市町村内の町又は字の名称並びに当該町又は字の区域を道路、鉄道若しくは軌道の線路その他の恒久的な施設又は河川、水路等によって区画した場合におけるその区画された地域（以下「街区」という。）につけられる符号（以下「街区符号」という。）及び当該街区にある建物その他の工作物につけられる住居表示のための番号（以下「住居番号」という。）を用いて表示する方法
道路方式	市町村内の道路の名称及び当該道路に接し、又は当該道路に通ずる通路を有する建物その他の工作物につけられる住居番号を用いて表示する方法

(2) アドレスマッチングによる経緯度情報の取得

経緯度情報は、各種GPS端末による直接取得の他にも、アドレスマッチング（geocoding）によって取得することができる。なお、図5にも示したように、アドレスマッチング用のソフト（アドレスマッチングサービス）は、位置（経緯度）情報と紐づけられた独自の住所データベースを持っており、依頼者の住所情報とデータベース上の住所情報との照合を行い、照合度情報も付与して経緯度情報を依頼者に提供する仕組みになっている。なお、わが国の場合、住所が漢字と数字が混在している他、同一の住所についても様々な表記が用いられている。このためデータベースが持つ住所情報との照合に先立って、その前処理として依頼者が提供する照合元の住所表記のゆらぎに対処するための正規化が行われる。

図5 アドレスマッチングの仕組み



現在、国土交通省、東京大学空間情報科学研究センター（CSIS）、民間のデータプロバイダー等がアドレスマッチングサービスを提供しているが、これらの機関が提供するサービスの中には、街区レベルの住所に位置情報を付与するものと、より詳細な住居レベルの住所に対するデータサービスを行っているものがある。また、付与される位置情報についても、敷地（筆）や建造物の中心点を与えるものと敷地への出入口を指示するものがある。

ところで、現行のアドレスマッチングにはいくつかの問題点が存在する。

まず、街区表示による住所については、一般に町丁・目（大字・字に相当）、街区（地番に相当）、住居番号の三段階で表示されるが、集合住宅等ではこれらに加えて棟や号といった枝番号が存在する。また、単一の住居番号を持っていた敷地が再開発され複数の住戸等が新築されたような場合、一般にこれらは同一の住居番号によって表示される。

一方、道路方式による住所表記の場合、単一の住居番号が複数の建物に照応する場合も少なくない。長谷川・卯田は、現行のアドレスマッチングが照合元の住所とマッチングソフトが持つ住所との照合度に関する情報を与えている一方、提供される経緯度情報が照合元の住所と1対1対応しているかどうかについての情報を与えていないという問題点を指摘し、今後の改善の方向を提案している〔長谷川・卯田(2012)〕。

わが国の場合、包括的な建物レジスターはまだ未整備であり、住戸あるいは事業所の建物と位置情報との間には必ずしも1対1の相応関係は成立していない。しかし、このことは、経緯度情報の統計目的での利用可能性を否定するものではない。なぜなら、照合の主体を建物に所在する世帯や事業所といった統計単位の側に置くのではなく、見方を変えて位置（経緯度）情報の側から個々の地点（ポイント）が単一あるいは複数の統計単位を担っていると考えることによっても、本稿で提起しているような利用目的には十分に妥当性を持ちうると考えられるからである。

（3）経緯度情報の情報特性

住所がアナログな文字情報であるのに対して、経緯度情報という数値化された情報は、その情報特性面で多くの優越性を持っている。

住居表示の指針を規定した法律が存在するにもかかわらず、わが国では地域によっては独特な表記による住所が日常的に使用されている。さらに、漢字、かな、数字の使用について住所表記は様々なゆらぎを持っている。このため、アドレスマッチングソフトに内蔵された住所表記の正規化プログラムも、あらゆる表記形態に対応できていない。

道路方式の街区方式への住所表記の変更によって住所が改められた場合、町丁・目レベルで

統計を時系列的に接続するためには、追加的な照合作業が必要となる。もし住所情報が経緯度情報に変換されてさえいれば、変更前の地域区分にもまた変更後のそれにも容易に対応でき、統計の時系列比較に際しても、特に追加的な作業を必要とすることなく、任意の時点での地域区分に対応したデータの編成を行うことができる。

また、経緯度というベクター型変数に変換された住所情報は、データ解析的にも多くの追加的価値を持っている。個体相互間のクランプ分析、あるいは公共交通機関や広幅員道路等から各個体までの距離についても、様々な距離測度による計測に基づく多様な分析を行うことができる。

さらに、経緯度情報については、統計単位が所在する任意の経緯度(X,Y)を指定することで、多様な形のポリゴンから構成されるレイヤー群から該当地点のポリゴン値を場所特性情報として正確に取り出し、個体レコードに追加変数として付加することができる。なお、この点については後に詳述する。

(4) 場所特性情報とそのレイヤー構造

(i) 場所特性情報

(a) 個人・世帯を規定しうる場所的特性

個人の意識や行動、世帯(家計)の行動や活動に作用を及ぼしうる要因は様々ある。このうち、住居の所有形態や居住スペースの広さといった調査項目は、国勢調査や住宅・土地統計調査の調査票に設けられている。なお、住宅・土地統計調査では、世帯側で作成、報告する調査票とは別に、調査員や市区町村の調査担当者が、調査区(基本単位区)の住環境を独自に把握している。例えば、調査員が概観の目視により作成する建物調査票で敷地に接している道路の幅員が調べられており、また同じく調査員が調査実施に先立って作成する単位区設定図に、郵便局、銀行、バス停、鉄道駅から当該単位区までの距離が、さらに調査実施者である市区町村では、医療機関、公園、公民館・集会所、避難場所、老人施設、保育所、小学校、中学校までの距離が距離帯別に把握されている。さらに、広い意味での住環境としては、常住地の人口密度、住・生活環境の快適性(緑被率、騒音、大気汚染度、地域の犯罪率)、学校区、用途地域区分、地価といった社会経済的要因だけでなく、日照、標高、気象といった自然条件も含まれ、それらもそのような環境の中に存在する個体に対して、直接的、間接的に影響を及ぼしうると考えられる。

このように、場所特性が統計調査によって把握されている場合にも、現状ではあくまでも調査区といった面的広がりを持った地域単位として捉えられた距離概念であり、個々の統計単位から計測した距離とはなっていない。さらに、場所特性に関する情報は一般に独立したファイル中の個々の集計された変数値として存在しており、これらの変数を調査事項あるいは調査員等による独自把握事項として持っていない多くの調査では、これらの変数値をそれぞれの調査票情報と有機的に連結できる状態には今のところなっていない。

(b) 企業・事業所を規定しうる場所的特性

企業が工場や店舗といった事業所の立地場所を選択し決定する場合、用途地域区分、容積率、建蔽率、環境基準といった立地の制度的制約要因以外にも、それぞれの場所が持つ様々な場所特性が考慮される。どのような要因が主要あるいは副次的であるかは業種によって自ずと異なるであろう。例えば、一般消費者を対象とする流通業や対個人サービス事業所の場合、立地事業所周辺の人口密度、通過人口規模、公共交通機関からのアクセスの利便性、地価やテナント賃借料率、他の店舗等との競合性の有無などが恐らく事業所の設置場所を判断する際の主要な規定要

因となるであろう。一方、製造事業所の場合、敷地の広さ、標高差、冠水の可能性、高速道路のインターチェンジ・幹線道路・空港・港湾等からの距離、労働年齢人口規模、関係業種の集積度、競合度、地価の他に、減免税や投資優遇措置、環境規制といった行政による制度的要因も当然考慮の対象となりうる。ちなみに山本は、ドイツの車メーカーBMW 社が工場の新設に際して誘致をはかる各国の自治体当局に対して提示した立地条件点検項目一覧【付表2】を紹介している〔山本 2003〕。この資料は事業所立地に関する製造企業側の関心事項を示すものであると同時に、われわれはこれから製造業に関する場所特性変数の候補となるものを窺い知ることができる。

(ii) 場所特性情報のレイヤー構造

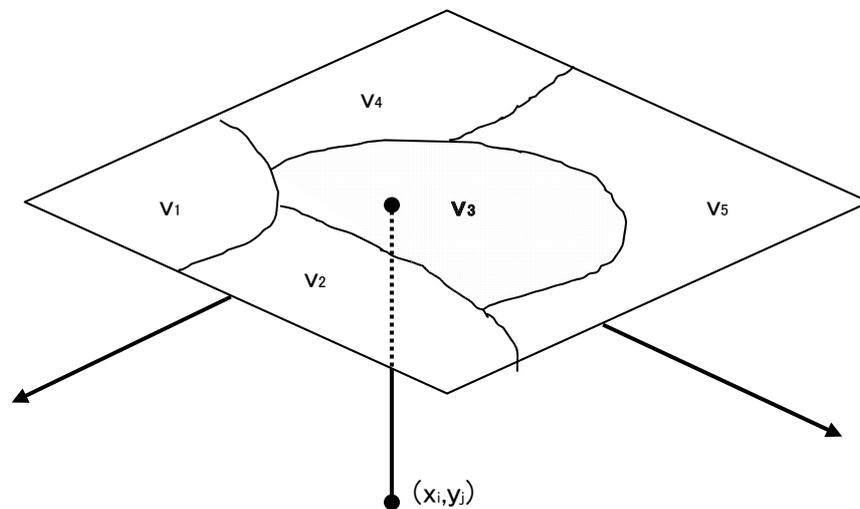
ここで統計単位としての個体情報を構成する各変数を経緯度情報と関連づけて見てみよう。

いま、変数値 v に経緯度座標 (x, y) をリンクさせた情報の組 (x, y, v) を考える。任意の経緯度情報に対して何らかの変数値が対応している場合、対象地域全体は特性値を異にするいくつかの排反なポリゴンの集合から構成されている。このように図式的に捉えた場合、個々の変数は一つの平面(レイヤー)とみなすことができ、そこでは各地点は必ずいずれか一つだけのポリゴンと関係づけられる。ここで X, Y をそれぞれ経度、緯度とすると、経緯度と場所特性値 V の間には

$$v = f(x, y)$$

が成立している。例えば図6に示したように、ある地点 (x_i, y_j) を与えた場合、その地点が属するポリゴンの変数値(ここでは v_3)が返される。

図6 レイヤー中のポリゴン値と経緯度情報の対応関係



すなわち、 $v_3 = f(x_i, y_j)$ として、この変数 V の経緯度座標 (x_i, y_j) に対応したレイヤー上の実現値として v_3 が与えられる。ここでは変数値 V は統計単位としての個体が存在する地点上で定義されていればよい。

次に、個体の存在を規定する外的環境要素としての場所特性変数レイヤーについて考えてみよう。レイヤー全体が相互に排反なポリゴンによって網羅的にカバーされていれば、任意の経緯度情報に対して、該当する場所特性変数の変数値を付与することができる。

ところで、場所特性情報の中には、特定の地点(あるいは線分)の上でしか変数値が定義(採取)されていない場合も少なくない。そこでは、統計単位としての個体が存在する地点と場所特性値が得られる地点とが一致しないのがむしろ一般的である。このような場合、個体情報と場所特性

情報を個体ベースでリンクできるようにするためには、地点(あるいは線分)上で定義された特性値を面的に拡張、補完することにより、各ポリゴンに特性値を付与する必要がある。

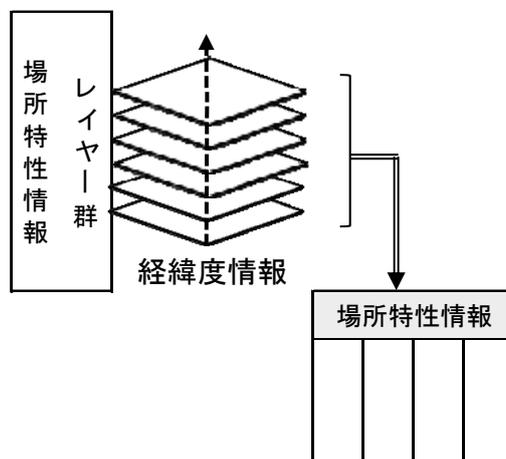
一例として、路線価等の地価情報に基づく地価レイヤーの編成方法について考えてみよう。周知のように、路線価は線弧に対して与えられている。線弧を用いたボロノイ補完によって構成されるポリゴンに対して当該線弧の路線価を与えることで、価格が採取されていないポリゴン内の任意の経緯度に対してもその路線価を地価の第一次近似情報として与えることが考えられる。

路線価については、実勢地価はもとより公示地価に対しても、その評価額は下方バイアスを持ち、またその乖離度は自治体によっても異なるといわれている。路線価による第一次近似地価を実勢地価に近づける第二次近似の方法としては、例えば、公示地価採取地点に最も近い路線価情報を空間検索することによって抽出し、それぞれの平均乖離率を市区町村別に求め、それを用いて地価の第二次近似値をもつポリゴン群を持つ地価レイヤーを編成するといった方法が考えられるであろう。

場所特性変数はそれぞれ一つのレイヤーを構成している。このため、統計単位としての個体に対して外部環境条件として作用を及ぼしうる場所特性変数の総体は、重層的に積み重ねられたレイヤーの集合体、すなわち場所特性情報レイヤー群として捉えることができる。

これらのレイヤー群をそれぞれの統計単位としての個体の存在の場(地点)が与える特定の経緯度情報を用いていわば串刺し的にリンクすることによって、個々のレイヤー上でその地点情報が属するポリゴンが与える場所特性値を要素とする場所特性値ベクトルが構成できる。図7は、経緯度情報を用いた場所特性値ベクトルの編成の様子を図示したものである。

図7 経緯度情報によるレイヤーのリンケージ



(5) 個体レコードの場所特性情報への拡張

(i) 個体属性情報とデータ本体情報のレイヤー構造

調査票情報から転写された個体レコードを構成するフェース項目にあたる個体属性情報とデータ本体 (data body) 情報は、これまで一般には統計単位に帰属する原単位情報を当該個体に関して写し取った一組のレコードとして捉えられてきた。

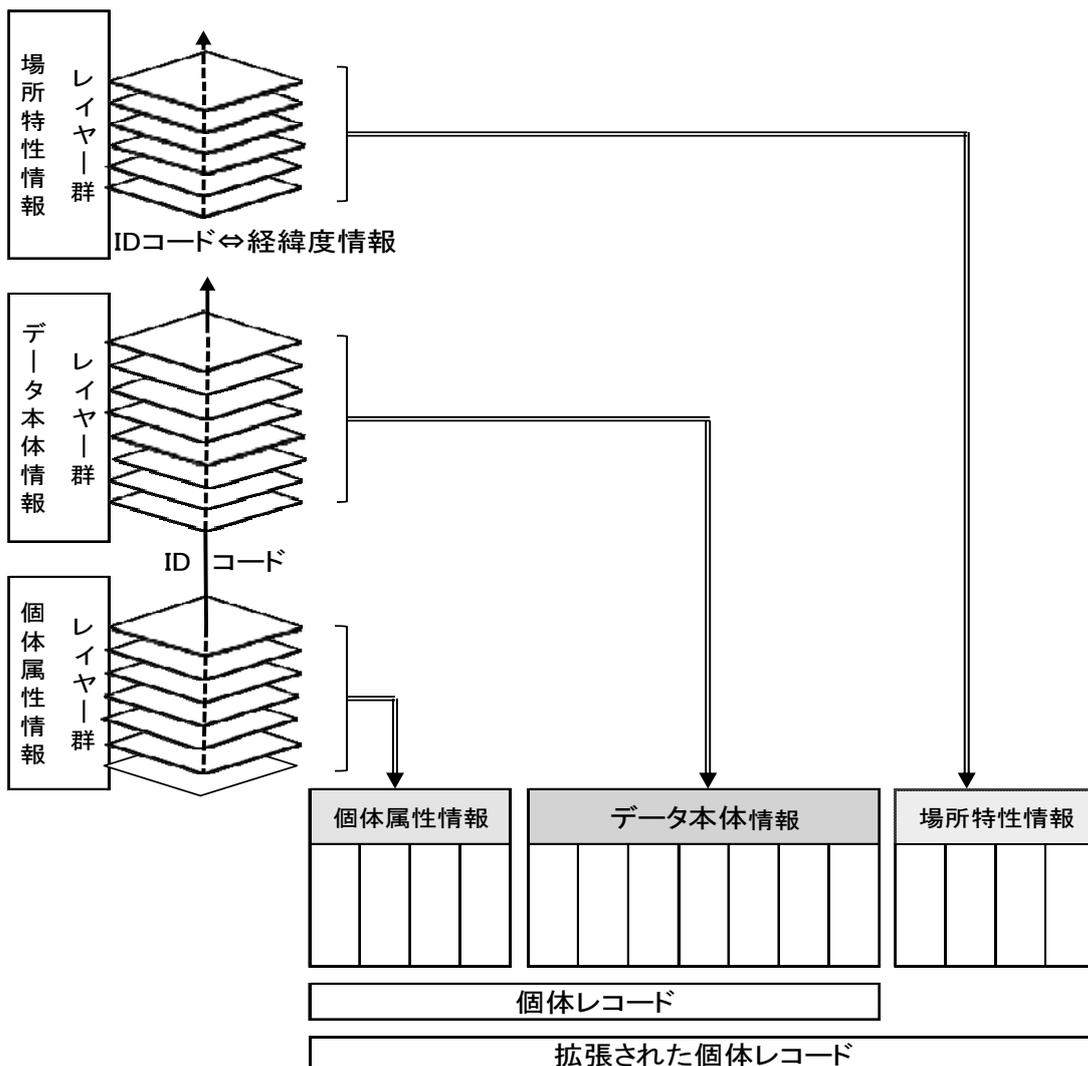
調査票情報から得られた各個体に関する一組のレコードは、個体識別 (ID) コードによって integrate されたレイヤー上のポリゴン値として解釈することもできる。すなわち、各レイヤーは個体識別 (ID) コード情報を持たせた変数値を持ち、それらの変数値はレイヤー上でバーチャルなポリ

ゴンを構成する。そこでは各レイヤーはそれぞれ排反なポリゴンによって網羅的におおわれており、個々の統計単位が持つ個体識別(ID)コードはいずれかのポリゴンの個体属性値あるいはデータ本体値と紐づけされている。このため、個体識別(ID)コードと個体属性・データ本体レイヤー群との間には、個体識別コードを特定すればそれぞれのレイヤーから、いわば串刺し的に、一組の該当する個体属性値、データ本体情報の値を引き出すことができる関係にある。

(ii) 場所特性変数の付加による個体レコードの拡張

すでに上述したように、図7は、経緯度をキー情報とした場所特性レイヤー群からの場所特性変数の data integration の過程を示している。一方、統計単位の個体特性情報、データ本体情報についても同様に、個体識別(ID)コードをリンクキーとしたそれに他ならない。リンクキーとなる変数が経緯度と個体識別(ID)コードと異なるものの、両者はいずれもこれらをキー変数として持つリレーショナルなデータ構造にある。このことは、統計単位としての個体と経緯度情報とを関連づけることができれば、これらを介して個体属性情報とデータ本体情報に場所特性変数を付加することによって既存の個体レコードの拡張ができることを意味する。図8は、既存の個体レコードへの場所特性変数の付加によるレコードの拡張の仕組みを模式的に示したものである。

図8 場所特性情報を考慮した個体レコードの模式図



ここで、統計単位である個体の存在の場に注目してみよう。個体が統計単位(あるいは行政単位)として存在している場所を経緯度情報として捉えた場合、すでにアドレスマッチングとの関連で上述したように、経緯度情報と個々の個体識別(ID)コードとの間には1対1(あるいは場所によっては1対複数)といった照合関係が成立している。言い換えれば、経緯度情報と個々の個体識別(ID)コードとの照合リスト(de-identification table)では、同一の経緯度に1つないし複数の個体識別(ID)コードが対応している。

両者がこのような照合関係にあることから、経緯度情報を共有する個体の場合、複数の個体レコードに対して同じ一組の場所特性値からなる場所特性情報が付加されることになる。このことは、複数の個体が同一の外部環境の影響下にあること、またそれをデータ論的に言えば、これらの統計単位が与える調査票情報に対して、同じ変数値の組からなる場所特性が作用を及ぼしていることを意味する。

8. データの秘密保護を考慮したデータベースの構造

集計量と異なり、個体レコードの場合には個体が識別されるリスクは飛躍的に増大する。個人情報保護法の施行以来、特に識別情報の取り扱いについては慎重を要することが情報管理者、利用者に求められている。

人口減少社会において行政には限られた資源制約のもとでのよりきめ細かな施策の立案、遂行が求められており、行政の効率化との両立を図る情報政策として、行政内部での情報の共有化が進められている。このような行政事務の遂行過程においても、秘匿性の担保は充足すべき要件であり、それと整合的なデータベースのシステム設計が必要となる。

一方、統計の分野では、統計単位である個体の側から提供される原単位情報に基づいて統計は作成される。その意味では、調査対象者側の調査への協力が統計作成の不可欠の前提である。このような事情から統計作成機関側では、統計作成に係る秘密の保護を最重要事項の一つとして様々な秘密保護措置が取られてきた。

個体の識別という点では、経緯度情報は直接的な識別可能性を与える。従って、たとえ個体レコードから氏名や名称といった個体の特定につながる識別子を除去したとしても、それが経緯度を変数として保持していれば容易に個体が特定できることになる。

以下では、情報の利用可能性を維持しつつ経緯度情報による識別リスク回避の方法について、海外での事例を参考に検討してみることにする。

(i) 秘密保護要件を考慮したデータベースの構築事例

各国では、ビジネス・レジスターや世帯レジスターをコアシステムとした政府統計のアーカイビングシステムを設計するにあたって、それに格納された統計個体情報と秘密保護との両立を図ることのできる構造設計とするように特別な政策的配慮を行ってきた。そこでの中心となっているのが、個体識別情報の取り扱いである。

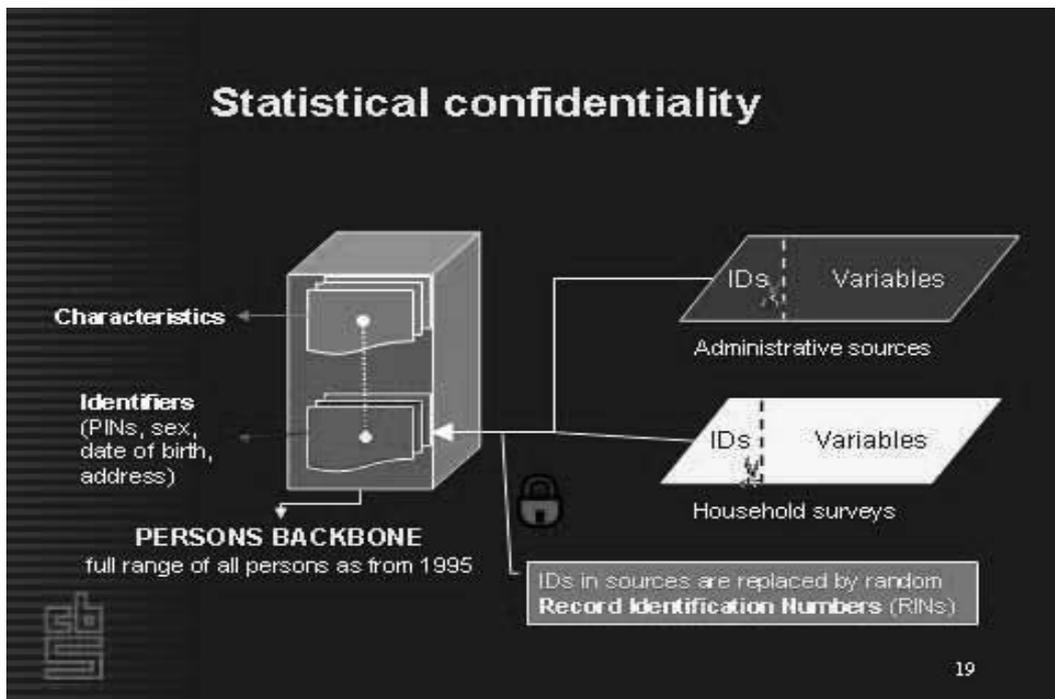
レジスターベースの統計システムを採用している北欧諸国やすでに統計調査と行政記録情報に共通に適用される統一個体識別コードに従って情報の収集を行っているオランダ、さらには強力な制度的強制によって企業の登録を管理しているフランス等においても、統計目的のために使用する識別コードは行政統一コードとは別系統のものとして設定されている。そこでは、収集された個体レコード中の行政統一識別コードは統計目的での代理識別子(surrogate identifier)に

変換された上でデータベースに格納される。様々な情報源から提供される同一個体に関する個体レコードは共通の統計用識別コードを持つことから、レコード間のマッチング等には変換後のコードによって連結可能である。なお、政府統計機関が他の行政機関から行政統一識別コードを持つ行政記録情報を新規に提供を受けた場合にも、それを受理した時点でそのコードが代理識別子に変換され、共通識別情報を持つリレーショナルなデータセットとして保管され、必要に応じて既存の個体レコードと連結される。なお、行政統一識別コードと統計目的での代理識別子との照合リスト(de-identification table)については、厳格な秘密保護を担保するために、限られた管理者権限を有する少数の職員だけがアクセスすることができる特別なファイルとして、アーカイブの特別のファイアーウォールの中で管理される。

図9、図10は、オランダ中央統計局(CBS)が維持管理している個人・世帯レジスター(Social Statistical Database: SSD)で採用されている個体レコードの格納方式を示したものである。なおオランダでは行政機関が収集、作成するすべての行政記録情報に行政統一個体識別コードの付与が義務づけられており、また2003年の改正統計法によって、すべての行政機関は中央統計局に対して行政記録情報の提供義務を負うことになった。

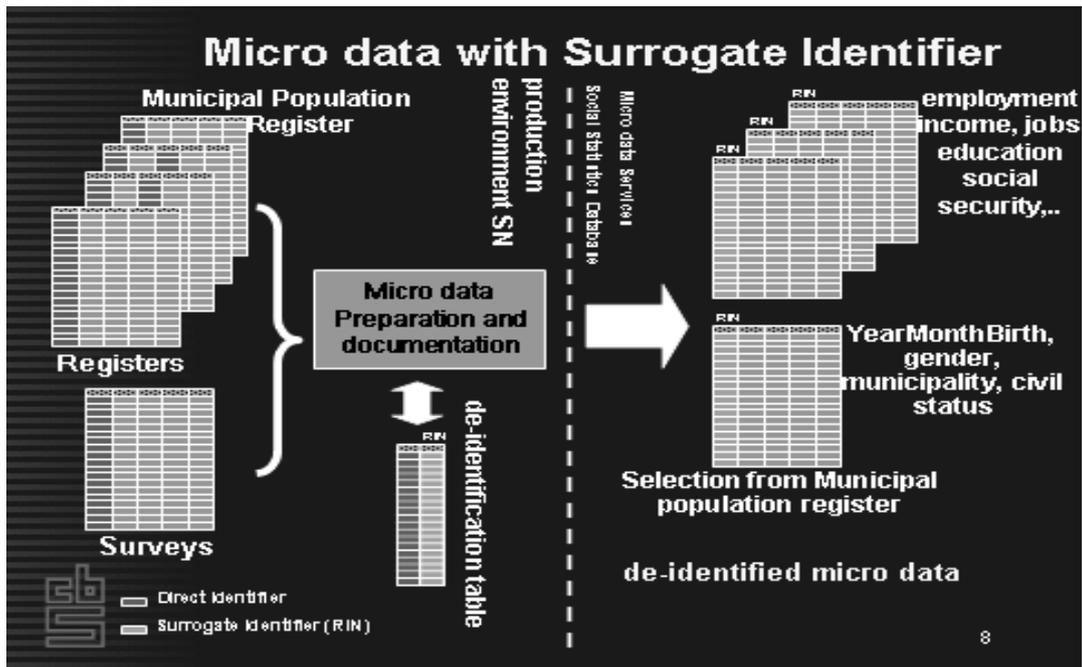
このため、中央統計局に提供される行政記録情報には、すべて行政統一個体識別コード(図9中のIDs)がつけられている。この識別コードは実際の行政事務遂行に用いられているものである。このため中央統計局では、秘密保護と行政事務への支障をきたさないようにそれを乱数を用いて作成した代理識別子である統計用識別コード(図10中のRINs: Record Identification Numbers)に置き換え、各個体レコードに個体識別情報としてRINsを持たせた個体レコードから構成されるリレーショナルなデータベースとして政府統計データのアーカイビングを行っている。

図9 オランダ中央統計局におけるSSDのアーカイビングシステム



〔出所〕 Nordholt (2007) p.19

図10 代理識別子による個体レコードの格納



〔出所〕 Nordholt (2007) p.8

図9からもわかるように、統計用識別コードと行政で用いられている行政統一個体識別コードとの照合表(de-identification table)は、個別レコードデータがアーカイブされている一般のファイルとは別に特別な鍵のかけられた秘密ファイルとして厳格に管理することで、SSD では統計的リンク可能性と秘密保護との両立を図っている。

一方、データアーカイブの本体部分ではデータベースは階層構造を持たせた形で構築されており、データ更新等のメンテナンス情報をその履歴情報も含めて包括的に記録した基幹ファイル、サンプリングフレーム用ファイル、レジスター統計作成用ファイル、研究目的での利用に供するファイル、Statline と呼ばれる外部の利用者がクエリーによるオンデマンド処理を行うための data cube からなるファイルといった位階的構造を持っている。なお、このような政府統計データのアーカイブを行うデータベースの構造については、森(2008)による展望的研究がある。

(ii) 経緯度情報の代理変数への変更による匿名化

経緯度情報は、個体の識別に直結する情報である。個体識別のリスクを回避しつつ経緯度情報を統計に有効に活用するための方策を講じる必要がある。上記のオランダ中央統計局が SSD において採用している方法が、経緯度情報の匿名化にも活用できるように思われる。

図9、10の説明でもすでに触れたように、オランダ統計局では、政府統計データのアーカイブに際して、個体識別リスクを考慮して、それが本来有していた個体識別コードをレコード識別番号(RINs)という代理変数に変換してデータベースに格納しており、またその照合用ファイル(de-identification table)は厳格な管理下に置いている。このようなデータの取り扱い方式は、経緯度情報の匿名化にも援用できるように思われる。

〔第1段階〕経緯度情報の代理変数への変換

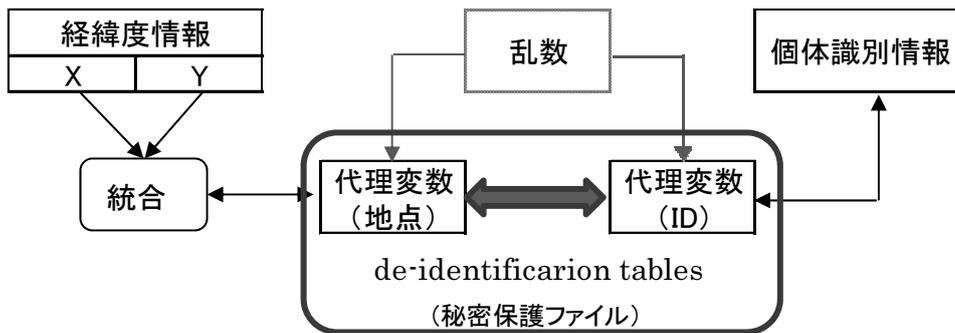
まず、経緯度情報を特別な意味を持たない乱数によるコードに置換し、それを経緯度の代理変数とする。具体的な方法としては、例えば GPS の精度を、住戸や事業所の特定に十分な精度と

思われる小数点以下 5 桁レベルで表示した場合、国内の地点の緯度は 7 桁、経度は 8 桁で表わされる。それらに 10^5 を乗じて整数化し、両者をマージして 15 桁からなる単一の変数を作成する。この数値には経緯度情報の痕跡が残されているので、これを適当な桁数の乱数にチェックデジットを追加した数値を当該地点の経緯度情報の代理変数とする。このようにして作成した代理変数からは、地点の特定はできない。なお、この代理変数と経緯度情報との照合ファイルは、当然、**de-identification table** として厳格な管理下に置かれる。

[第 2 段階] 経緯度情報の代理変数と個体識別コードの照合ファイルの作成

経緯度情報を統計の個体レコードと個体ベースで関係づけて使用できるようにするためには、この他に経緯度情報と個体識別コードとの照合ファイルが別途必要となる。このため、すでに代理変数化されている個体識別コードと経緯度情報との代理変数相互の照合ファイルも整備しておかなければならない。この照合ファイルについても、**de-identification table** として厳格な管理下に置かれる。図11は、それぞれの識別変数の代理変数化と照合ファイルの関係を概念的に示したものである。

図11 識別変数の代理変数化と照合ファイル



以上のような照合システムを構築することにより、図8中の「ID コード⇔経緯度情報」のマッチングの部分、より詳しくは、「ID コード⇔(ID 乱数コード)⇔(経緯度乱数コード)⇔経緯度情報」と記載されることになる。

アーカイブに格納する個体レコードに経緯度と個体識別情報の代理変数をそれぞれ持たせることによってデータ管理者は、リアルな地点と調査単位としての個体への遡及可能性を排除しつつ、適宜必要に応じて照合テーブル情報を介して個体レコードへの新たな場所特性変数の追加を行うことができる。

このような代理変数を用いた照合システムを構築することによって、秘密保護とデータインテグレーションの要件を充足することができよう。なお、経緯度情報の代理変数が実際に使用されている事例としては、カナダの **civic number** がある。これは、各住戸や施設の経緯度情報とリンクさせた乱数コードを各住戸や事業所等に付与したものである。これによって、消防、救急、警察といった各種の事案が発生した際にその番号を行政に通知するだけで、通報を受けた行政側では、この番号を用いて行政データベースに格納されている当該の世帯や事業所等に関する各種情報(世帯員の病歴、事業所での危険物貯蔵など)やそこへのアクセス情報等を取り出し、迅速かつ適切な行政サービスを提供することができる。

9. 利用に際してのデータの構成

(1) 個体レコードの編成

単一の統計調査の調査票情報からなるレコードあるいは行政記録情報を個体レコードとして利用する場合、転写レコードそのものが利用資料となる。一方、異なる種々の情報源からなる情報であってもそれが共通の個体識別(ID)コードが付与されておりさえすれば、それらはバーチャルにデータ統合されていることになる。なぜなら、単にそれらをリンクキーとして複数の個体レコード同士をマッチングすることによって拡張した個体レコードを編成することができるからである。

(2) 個体識別コードと位置情報の照合による個体レコードへの場所特性情報の付加

個人や世帯あるいは事業所の個体レコードへの場所特性情報の付加については、次のような方法が考えられる。

まず、統計識別コードと経緯度情報との照合リスト(de-identification table)を用いて統計単位である個体の常住地(世帯あるいは事業所の所在地)の経緯度情報の確定を行う。次いで、これらの経緯度をキー情報として各レイヤー上のポリゴン値を確定する。これによって、場所特性情報レイヤー群の個数(レイヤーの枚数)に対応するポリゴン値が場所特性変数ベクトルの形で与えられる。このような方法によって、既存の個々の個体レコードにそれぞれ一組の場所特性値が付加され、拡張個体レコードが構成できる。図12、図13は、こういった拡張個体レコードのイメージを例示したものである。

図12 場所特性情報の付加による拡張個体レコード(経緯度情報なし)

ID	個体属性情報				データ本体情報						場所特性情報			

図13 場所特性情報の付加による拡張個体レコード(経緯度情報つき)

ID	経緯度		個体属性情報				データ本体情報						場所特性情報			
	経度	緯度														

こういった個体識別(ID)コードあるいは経緯度情報を持つ拡張個体レコードは個体識別リスクが大きい。このためその利用には当然、慎重とならざるを得ない。秘密保護との両立を図るためには、海外の先行事例がそうであるように、このような識別レコードはデータアーカイブ内に厳格に管理され、基本的に行政内部での行政目的での使用に限定する必要がある。一方、学術目的での利用はあくまでも宣誓職員によるオンサイト利用あるいはオンデマンド型の利用に限定し、分析結果の公表に関しては秘匿性の確保という観点から適切な秘密保護委員会による事前審査といった制度的仕組みを構築する必要がある。

10. レコードの拡張による新たな分析可能性

統計識別コードと経緯度情報との照合リストが持つ個体識別コードと経緯度情報を介して既存の個体レコードに場所特性変数を追加変数として付与することができる。このようにして拡張された個体レコードは、後述するように、それまでのデータセットでは行うことができなかった新たなタイプの分析を可能にするように思われる。

拡張個体レコードを構成するのに用いた経緯度情報は、個体存在の場所をピンポイントで示すことから個体を識別する直接的な手掛りを与えることになる。個体識別という統計上の秘密保護の脅威を取り除くための方策としては、図14に示したように、拡張個体レコードから統計単位に関する個体識別(ID)コードおよび経緯度情報を削除することが考えられる。こうすることによって、拡張個体レコードからなるデータセットを個体識別のリスクを排除しつつ場所特性情報も含めた包括的な分析に対応できるデータセットへと作り変えることができる。なお、既存の個体レコードに付加された場所特性情報それ自体も、レイヤー上のポリゴン値の分布によっては識別リスクをもちうる。これは、位置情報を持つ変数の匿名化という、今後の重要な検討課題である。

図14 識別情報除去済み拡張個体レコード

個体属性情報	データ本体情報	場所特性情報

場所特性変数を既存の個体属性変数ならびにデータ本体変数とともにモデルに取り込むことで、既存の調査票情報の諸変数値に内在していた場所による被規定性の部分を本来の調査票情報から分離した分析ができるであろう。場所特性を変数としてモデルに組み込むことによって、それまで既存の変数に混在してきた個体特性要素と場所特性要素とを峻別し、それぞれに係る寄与部分を定量的に評価できるであろう。このような拡張個体レコードから構成されるデータセットを用いることで、場所特性を混在させた諸変数に基づく分析結果は、場合によっては本質的な見直しを求められる可能性がある。

経緯度情報並びに個体識別(ID)コード除去済みの拡張個体レコードを用いて様々な空間分析を行うためには、その過程で各個体レコードは何らかの形で経緯度情報と関係づけられなければならない。この関係づけと秘密露見リスクの排除との両立を図る方法としては、データ解析プロセスを二段階に区分する方法が考えられる。

[ステップ1] 経緯度情報による経緯度乱数コード集合要素の抽出

経緯度情報・経緯度乱数コード照合ファイル上で経緯度情報を用いて空間集計によって分析目的に応じた空間集合(バッファ)に該当する統計単位の経緯度乱数コードの集合要素(サンプル)を抽出する。

[ステップ2] 抽出された拡張個体レコードによる分析

このようにして抽出した個体サンプルを分析対象として、経緯度情報ならびに個体識別(ID)コー

ド除去済みの拡張個体レコードを用いて分析を行う。

このような空間分析ニーズと統計の秘密保護との両立を図るためには、分析結果について秘密保護の視点からの審査制度である秘密保護パネルと併用した特別な宣誓職員制度等によるオンラインでのデータ使用あるいはオンデマンド型の利用といった利用システムの構築が必要となるであろう。

むすび

GIS の最大の特徴は、経緯度情報をマッチングキーとして、異なる種々のデータソースに基づいて個体ベースのリレーショナルデータベースの構築ができる点にある。本稿では、個体識別(ID)コードとリンクした経緯度情報をマッチングキーとすることで、場所特性変数の実現値として捉えた場所特性に関する各レイヤーにおけるポリゴン値を場所特性情報として個体レコードに付加することで既存の個体レコード情報を拡張できる可能性を検討してきた。以下では、今後取り組むべきいくつかの課題を指摘することで本稿のむすびとしたい。

まず、場所特性に関するレイヤーの整備面については、次のような課題がある。

その1は、既存の利用可能な場所特性に関するレイヤーに関するメタ情報の整備である。現在、すでに場所特性に関するレイヤーとして有料あるいは公開情報が存在している。それぞれのレイヤーがどのような利用上の特徴を持っているかを含めて、その案内情報についての一覧リストの作成を行う必要がある。

レイヤー整備に関する第2の課題は、独自のレイヤーの整備である。今後、GISは研究面だけでなく行政面においても広範に使用されるようになると予想される。このように利用ニーズが広がりを見せる中で、経緯度情報を照合キー情報として持つデジタル情報が質・量ともに飛躍的に拡大すると考えられる。こういったデジタル情報の中には、当初からポリゴンによって全対象空間を網羅的にカバーしているデータセットだけでなく、交通事故や火災の発生に関する記録、災害の際に避難援助を必要とする住人の所在情報等、行政が保有する情報の中には、経緯度という地点情報を持つ変数値からなるデータセットも少なからずある。後者のようなタイプのデータセットについては、特別な同定作業を経ることなく、空間検索、空間集計、さらには距離(時間距離を含む)等の分析手法を直接適用することができる。

これらの地点と関係づけられた情報の中には、その数値を面的に補完、拡張することで、数値が定義されていない地点についても網羅的にポリゴン値を推計によって与えることができる場合がある。本稿で考察したような個体レコードの場所特性情報をキー変数とした拡張を行うためには、いずれのレイヤーも相互に排反でしかも全地点漏れることなくいずれかのポリゴンによってカバーされている必要がある。このようなレイヤー編成作業には、地域補完やボロノイ分割による推計といった手法の活用が考えられる。なお、今後整備すべき場所特性レイヤーには、当該地点周辺の状況(外部環境)も取り込んだレイヤー、例えば、近隣外部効果を考慮した事業所の集積度に関する指標なども考えられる。

さらに、同一種のレイヤーについて、様々な統合度によるポリゴンからなる位階的なレイヤーの整備という課題もある。空間補完の手法の違いあるいは補完パラメータの設定いかんで同一の場所特性について複数のレイヤーを構成することができる。分析者に対して多様な分析上の変数オプションを提供できるレイヤーの整備も課題の一つとなりうる。

第2の課題として、場所特性レイヤーそれ自体の構造の解明がある。

場所特性レイヤーには、当該特性に対する人間の関与の程度に応じて、自然的レイヤー(標高、気候等)、行政的レイヤー(用途地域区分、建蔽率区分、容積率区分、税制措置等)、社会・人口レイヤー(人口密度、犯罪率等)、社会・経済レイヤー(事業所集積度、交通利便性)といったサブレイヤー群から構成される。

種々のレイヤーが整備されるにつれて、レイヤーの中には、地域分布のパターンの点で類似したものもありうると考えられる。このようなサブレイヤー群を構成する各レイヤーは、相互にどのような関連性を持っているのであろうか。この検証のためには、例えばレイヤーが対象地域とする全体の中から、例えば無作為に観測点を標本として選定し、それらに対応する場所特性値を各レイヤーから求め、それらのデータに基づいて、クラスターあるいは主成分等を算出することで、各レイヤー相互間の関連性についての情報を得ることができるようになる。

第3の課題としては、既存の実証研究の再評価がある。場所特性情報の利用面に関しては、既存研究で使用されたモデルに場所特性情報を説明変数として新たに追加したモデルを用いることによって、既存の研究の見直しを行うとともに、場所特性情報が既存のパラメータに及ぼす影響の程度を検証することが考えられる。

識別個体レコードからなるデータセットにおいて個人や家計の経済行動、あるいは生産額や売上高といった事業所の経済パフォーマンスに関する情報が個体の存在場所情報とともに利用可能であれば、場所特性変数値付きのモデルとそれを持たない従来型のモデルによる解析結果を比較することで、場所特性変数の被説明変数に対する説明力、あるいはそれを加えることによる既存の説明変数のパラメータをどの程度変更させるかを実証的に検証することができよう。またこれによって、説明力を持つ場所特性変数とそうでないものとを区別する中で、有効な場所特性変数の絞り込みもできるようになる。

さいごに、統計制度面での課題としては、場所特性情報が付加された個体レコードからなるデータセットの利活用システムの設計という問題がある。本文でもオランダ中央統計局のSSDを事例として政府統計データのアーカイビングシステムにおける秘密保護措置について紹介したが、経緯度情報に関する匿名化あるいは秘密保護の担保に関しては、これまでの個体レコードの秘密保護とはまた別の取り組みが必要となるように思われる。この点についても統計学の新たな課題として引き続き関心を持っていきたい。

【参考文献】

山本健児(2003)「BMWによる新規向上立地選択プロセス」『経済地理学年報』第49巻第4号

総務省統計局(2005)「平成17年国勢調査 調査区関係資料 利用の手引き」

総務省統計局(2006)「平成18年事業所・企業統計調査に関する地域メッシュ統計地図」

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』経済統計学会第94号

森博美(2010)「ジオコードによる統計の情報価値の拡張可能性について」『経済統計学会第54回研究大会報告予稿集』22-23頁

森博美(2011)「統計調査における地点情報の把握による統計の情報価値の新たな展開可能性について」『経済志林』法政大学経済学部学会 78-3 249-281頁

森博美(2011)「位置情報を用いた調査票情報の情報価値の拡張とその分析的意義について」『オ
 ケージョナルペーパー』法政大学日本統計研究所 No.25 1-25 頁

長谷川普一・卯田強 (2012)「災害時における全住民位置情報の重要性(その2)アドレスマッチン
 グに係わる諸問題と解決方策について」地理情報システム学会講演論文集 Vol.21

森博美・坂本憲昭・小西純・長谷川普一(2012)「近隣外部効果を考慮した事業所の立地集積度の
 計測ー八王子市を事例として」地理情報システム学会講演論文集 Vol.21

Eric S. Nordholt (2007) "Record matching for census purposes in the Netherlands" Joint
 UNECE/Eurostat Meeting on Population and Housing Censuses in Astana, 4-6 June.

Jarmin S.Ron and Miranda J., (2009) "The Impact of Hurricanes Katrina, Rita and
 Wilma on Business Establishments: A GIS Approach" *Journal of Business Valuation
 and Economic Loss Analysis*. Vol.4

【付表1】 立地環境特性の区分及び定義

特性番号及び区分		定 義
	商業集積地区細分	
10	商業集積地区	主に都市計画法第8条に定める「用途地域」のうち、商業地域及び近隣商業地域であって、商店街を形成している地区をいう。 概ね一つの商店街を一つの商業集積地区とする。一つの商店街とは、小売店、飲食店及びサービス業を営む事業所が近接して30店舗以上あるものをいう。また、「一つの商店街」の定義に該当するショッピングセンターや多事業所ビル(駅ビル、寄合百貨店等)は、原則として一つの商業集積地区とする。
	11 駅周辺型商業集積地区	JRや私鉄などの駅周辺に立地する商業集積地区をいう。ただし、原則として地下鉄や路面電車の駅周辺に立地する地域は除く。
	うち、駅改札内事業所	
	12 市街地型商業集積地区	都市の中心部(駅周辺を除く)にある繁華街やオフィス街に立地する商業集積地区をいう。
	13 住宅地背景型商業集積地区	住宅地又は住宅団地を後背地として、主にそれらに居住する人々が消費者である商業集積地区をいう。
	14 ロードサイド型商業集積地区	国道あるいはこれに準ずる主要道路の沿線を中心に立地している商業集積地区をいう(都市の中心地にあるものを除く)。
	15 その他の商業集積地区	上記「駅周辺型商業集積地区」～「ロードサイド型商業集積地区」までの区分に特性付けされない商業集積地区をいい、観光地や神社・仏閣周辺などにある商店街なども含まれる。
20	オフィス街地区	主に都市計画法第8条に定める「用途地域」のうち、商業地域及び近隣商業地域であって、上記「10商業集積地区」の対象にならない地区をいう。
30	住宅地区	主に都市計画法第8条に定める「用途地域」のうち、第一種・第二種低層住宅専用地域、第一種・第二種中高層住宅専用地域、第一種・第二種住宅地域及び準住居地域をいう。
40	工業地区	主に都市計画法第8条に定める「用途地域」のうち、工業専用地域、準工業専用地域及び工業地域をいう。
50	その他地区	都市計画法第7条に定める市街化調整区域及び上記「10商業集積地区」～「40工業地区」までの区分に特性付けされない地域をいう。
	うち、有料道路内事業所	

(出所) 経済産業省『平成19年商業統計表 立地環境特性格別統計編(小売業)』13頁

【付表2】 BMW 社が提示した新工場のための立地条件

大項目	具体的条件
1 用地の位置と規模	200～250 haのまとまりを持つ用地
2 地形条件	最高標高、最低標高、道路の標高、最寄りの水面(川、湖)の標高
3 技術的な調達、廃棄処理条件	最寄りの電力(110 kV/40 MW)調達可能地点までの距離 最寄りのガス(6600 m ³ /h)調達地点までの距離 最寄りの工業用水(450 m ³ /h)調達地点までの距離 最寄りの通信(2×各 60AL のPMA、12×光回線、各 34MB/sec)接続地点までの距離 最寄りの工業排水溝(250 m ³ /h)までの距離、その標高 雨水処理施設(下水溝、自然水面)までの距離、標高 廃棄物処理(固形物 2000t/a、汚泥、油脂 1500t/a、薄め液 95t/a)
4 交通	最寄りの鉄道駅、最寄りの鉄道線路までの距離、標高 工場用地に接する道路 最寄りのアウトバーンの特徴 アウトバーンまでの道路距離(物理的距離、時間(分))
5 近隣の建築状況	最寄りの住宅地まで最低限 800m 以上離れていること 外部からの煤煙、粉塵その他の汚染物質が工場用地に来る可能性 半径5km以内で災害が発生する可能性
6 最寄り空港	国内空港、国際空港までの距離(1 時間以内が条件)
7 用地の地質、建築阻害条件	地盤の固さ 建築物を支える地盤の深さ 地下水面の高さ 地下に空洞部分や充填部分があるか否か 考古学的に価値あるものがあるか否か 汚染物質が蓄積されているか否か 地震の危険性 上空や地下に構築物(電線、管など)があるか否か 建築物や記念建築物があるか否か
8 建築法(権利)	土地利用計画による指定状況 建設計画 建築基準の規制(高さ 30m以上まで建物を建設できること)、例外条件(煙突の高さなど) 建築許可交付までの期間(2002 年初めには工場建設を開始できること) エコロジー、自然保護、樹木などの条件
9 雇用条件(半径 50 km以内の地域)	地域名 人口の年齢構成(0-5 歳、6-15 歳、16-25 歳、26-45 歳、46-65 歳、66 歳以上) 人口動態(出生数、1980 年、1990～1999 年、死亡数、自然人口動態 1980 年、1990～1999 年) 住民の学歴(卒業資格なし、基幹学校卒、実業学校卒、大学入学資格保持者、学業継続中) 住民の職業教育受講状況(なし、職業教育、マイスター・テクニカー、大学卒、職業教育受講中) 学校種類別の生徒数 学校卒業生数(1980 年、1990～1999 年)(総数、基幹学校卒業資格なし、基幹学校、実業学校、ギムナジウム、職業学校) 大学生(総合大学、専門大学) 大学生の専攻分野(自然科学および数学、工学) 就業データ 失業者、失業率 金属、電機分野での規定労働時間数(週、有給休暇日数、祝日日数) 平均病欠日数(年間、従業員数当り、労働日当り、工業分野) 労働争議による操業不可能日数(年間、従業員千人当り、工業分野) 産業分野別就業者数(農業、工業、商業、運輸、サービス) 工業分野における平均時間給 工業分野における平均月給 金属・電機部門における付加的な人件費(法律に基づくものと協約に基づくもの。実労働時間当り)

10 生活環境(半径 50 km以内の地域)	基礎学校、基幹学校、上級学校のある都市までの距離 大学のある都市までの距離 最寄りの病院までの距離 最寄りの中規模都市、大規模都市までの距離 医療インフラ(1999 年) 人口 10 万人当り医師数 人口 10 万人当り歯科医師数 人口 10 万人当り薬局数 病院病床数 犯罪発生率(最新の統計で人口 10 万人当り犯罪件数) ドイツ語学校、国際学校の有無(外国の場合)
11 用地取得の可能性(権利を持つ第三者)	用地を一人の土地所有者から取得できるか否か 基礎的サービスを請け負うところがあるか否か 土地返還要求権利者がありうるか否か 用地内の耕区数 用地の土地所有者数(個人、農民、企業、公的機関)
12 年間販売額1千万マルク以上の供給企業	半径 10 km以内における供給企業名、所在地 半径 10～50 km以内における供給企業名、所在地 半径 50 km以内で供給企業が拡大できる見通し 半径 50 km以内に供給企業が新規立地する見通し

(出所)山本健兒(2003)「BMWによる新規工場立地選択プロセス」『経済地理学年報』第49巻第4号65頁第1表(一部修正)

〔謝辞〕

本論文は、平成24年度日本学術振興会科学研究費補助金基盤研究B「政府統計データのアーカイビングシステムの構造と機能に関する国際比較研究」(課題番号 22330070 研究代表者 森博美)および挑戦的萌芽研究「GPS情報の活用による公的統計の新たな展開可能性に関する多角的研究」(課題番号 23653060 研究代表者 森博美)による研究成果の一部である。

Exploring the New Dimensions of Statistical Records with GPS

Hiromi MORI (Hosei University)

ABSTRACT

Statistical units such as establishments, local units, households, individual persons are not entities which exist in virtual universe, but exist in real time/space world. Respective data written on the returned questionnaires which are regarded to be the reflection of information relevant to that unit, thus, consist also of information constraints by time/space dimensions. Put differently, returned questionnaires give a set of data relevant to the respective units filtered with time (reference date) and space (location of units' presence). In this sense the questionnaire information usually carries treble information characteristics. In order to identify unit's inherent attributes one should, therefore, exempt meaningful affects influenced by other two information factors.

In recent years one can witness two remarkable developments in statistics and related arena. Equipment of registers such as business and household registers in one hand, and a notable improvement in accuracy of GPS data on the other. The former provides the framework for the individual-based data integration and the latter information technological bases for adding a set of location specific variables to individual records through GPS. They pave the road for the successful expansion of existing records.

This paper examines first the particular attributes of questionnaires information inherent to the reflection of individual statistical units as social entities and then proposes the possible expansion of statistical records by adding location specific variables through GPS, which used to have disregarded their existence. This paper also intends to explore a new dimension of statistical data.

オケージョナル・ペーパー(既刊一覧)

号	タイトル	刊行年月
7	わが国における外国人の国籍別出生率について	2001.09
8	東京の消費構造－東京都生計分析調査	2002.10
9	Wide Variations in Statistics Data Sets on the Same Subjects－Reconsidering the Report of the Indian National Statistical Commission	2003.12
10	日中 1995 年産業別購買力平価の推計	2004.04
11	日本における「統計法」の成立	2005.06
12	「統計法」と法の目的	2005.07
13	諸外国におけるマイクロデータ関連法規の整備状況とデータ提供の現状	2005.09
14	統計に係る個人情報の秘密保護について	2006.08
15	若年層における雇用状況と就業形態の動向－『就業構造基本調査』のマイクロデータによる実証分析	2006.12
16	社会生活行動から見た若年層の不安定就業化・無業化の分析	2008.03
17	国勢調査による従業地把握の展開と従業地別就業データの意義	2009.06
18	無償労働の評価と世帯生産サテライト勘定	2009.10
19	エンゲルとザクセン王国統計	2009.12
20	第一次統計基本計画と政府統計の直面する課題	2010.01
21	エンゲルとプロイセン統計改革	2010.02
22	エンゲルと 1875 年ドイツ帝国営業調査	2010.03
23	調査形態論再論	2011.03
24	統計を規定する諸要因との関連から見た時空間個体データベースの可能性について	2011.04
25	位置情報を用いた調査票情報の情報価値の拡張とその分析的意義について	2011.06
26	ジオコード情報の活用による統計の把握精度改善の試み	2011.09
27	統計的マッチングによる疑似パネルデータの作成と精度検証	2011.11
28	駿河国人別調沼津・原政表再論	2012.01
29	ザクセン王国統計協会(1831-50 年)	2012.01
30	ザクセン王国における初期人口・営業統計	2012.02
31	フィンランドのビジネス・レジスター	2012.03
32	エンゲルのザクセン王国統計局退陣をめぐって	2012.04
33	フランスのビジネス・レジスター	2012.05
34	タウンページ情報を用いた事業所の自然・社会動態の把握	2012.07
35	疑似景況パネルによる予想パフォーマンスの計測	2012.11

オケージョナル・ペーパー No.36

2012 年 12 月 10 日

発行所 法政大学日本統計研究所

〒194-0298 東京都町田市相原 4342

Tel 042-783-2325、2326

Fax 042-783-2332

jsri@adm.hosei.ac.jp

発行人 森 博美