

社会・人口統計体系のデータベースの 構築と利用の手段について

鷗野 公郎*

はじめに

経済が成熟化し所得水準が高まるにつれて、国民の関心は経済成長といったマクロ的なものから、生活環境・住環境といったものに移ってきた。それにともなって地域的な情報が求められる度合は、過去と比べものにならないほど増大している。しかし、そうした地域情報にたいするニーズにもかかわらず、これを体系的に提供し分析するシステムは未だに整備されるに至っていない。そのため地域をめぐる議論は、地方の時代を叫ぶ抽象的・情緒的なレベルにとどまるか、あるいは逆に利権をめぐる超現実的なものとなりやすい。本稿では既存の統計を基礎にして、それをいかにして体系的に利用するか、そのためにどのようなハードウェアとソフトウェアが必要とされるかを検討する。

1. 社会・人口統計体系とその利用

統計局では昭和51年度から社会・人口統計体系の整備をおこなってきた。その成果は逐次刊行されてきたので、統計体系としてのスタイルもかなり定着するとともに、利用者にとっても馴染み深くなってきたといえよう⁽¹⁾。この小論はこのような時をとらえて、社会・人口統計体系の今後の発展のためにひとつの試論を提示するものである。

社会・人口統計体系は、「自然環境、人口・世帯、財政、学校教育、健康、医療、労働、居住環境、社会保障など国民生活の諸側面にわたり、その実態を示す種々の地域別の統計基礎データを体系的に収集、編成し、地域特性の把握、分析等の利用に供することをねらい」とするものである⁽²⁾。そのスコープは、従来の経済面を中心とする統計体系からみれば周辺統計にあたるものである。しかし、国民生活の視点からみれば、生活の質にかんする中心的な関心項目であるといえよう。なぜならば、生産とか雇用、財政などは、国民生活から見ればいわば中間的な手段なのであり、究極的な目的ではないからである。経済の成熟化とともに手段よりも究極的な目的の達成状況に関心があつまるのは当然といえよう。その意味では社会・人口統計体系に期待されるところはきわめて大きい。

統計局の社会・人口統計体系は、国連の提唱したSSDS (A System of Social and Demographic Statistics) と関連をもっている。SSDS のスコープは1975年段階においては次の項目をふくんでいた⁽³⁾

人口の大きさと構成、出生、死亡、移住

* 社会工学系

家族形成，家族構成，家計
社会階級，階層，流動性
所得・消費・貯蓄・資産の分配
住宅と住環境
時間配分と余暇の利用
社会保障と福祉サービス
学習活動と教育サービス
稼得活動，雇用サービス，活動不能要因
健康と医療
治安と安全，犯罪者，被害者

SSDSの特徴としては，以上のスコープのほかに，次の点を指摘しておくべきであろう。

- ① 経済勘定としてSNAを想定している……SSDS自体は社会・人口にかんする統計体系であるが，経済的な側面を無視しているわけではない。サービス提供のためのインプット，費用や移転，資本支出といった経済情報は，SNA(A System of National Account)によって提供される。このためSNAの詳細な分類とSSDSの分類とが一致するように配慮されている。
- ② アプリオリな体系化……国連はSSDSの報告書のなかで，体系構築にあたって2つのアプローチを区別している。第1は，アプリオリな方法，第2は実証的な方法である。アプリオリな方法とは体系の理論的枠組みを先行させることをいい，実証的な方法とは既存の知識にもとづき漸進する方法である。SSDSの構想は両者を補完的に考えていると報告書はのべているが，どちらかという第1の方法にかたむき，実証統計は体系としてではなくケース・バイ・ケースの例示にとどまっている。
- ③ 社会指標の基礎統計……14分野におよぶ幅広い対象分野は社会指標分野の多くを含んでおり，社会指標作成のためのデータ・ベースとして利用することができる。実際に，SSDSにもとづく社会指標の体系案がしめされ，OECDの社会的関心項目(Social Concerns)との対応関係も明らかにされた。
- ④ 状態の移動を重視……SSDSの体系の大部分は，様々にぶんるいされた人間のストックおよびフロー情報に関係するが，既存の統計はストック情報に集中している。しかし，フロー情報は，サブシステム間の変化を分析できること，過去の状態が現在の状態にどう影響しているかを知り将来の状態の変化を計画できることなどのために重視されている。そのため，体系の各所に遷移行列をくみこんで状態の移動を統計的にとらえようとしている。
- ⑤ マイクロ・データ・セットの利用……体系の大部分は家計ないし世帯を単位としている。これまで経済統計が多くの場合マクロ的集計量を扱い，経済理論がアトミスティックな個人を分析するのにたいし，家計ないし世帯を分析単位とすることは経済社会における行動主体を重視することにつうじる。かんがえてみれば，マクロとマイクロが画然と区別されてきたのは奇妙なことである。生活の質の重視と，コンピュータの能力アップにともなって，最近はマイクロ・データ・

セットが利用される傾向にある。

- ⑥ 地域情報……生産とか物価、国際収支といった集計度の高い項目と異なり、生活の質の観点からは生活に関係の深い地域にかんする統計が必要になる。しかしSSDSにおいては地域統計としての体系化は正面にでていない。

わが国の社会・人口統計体系はこのようなSSDSの特徴に照らしたとき、どのような特徴をもっているであろうか。

- ① まず経済統計との関係については、社会・人口統計体系では経済基盤ならびに財政に関する主要項目を含めている。県民経済計算、事業所数、従業者数、産業別生産額、歳入、目的別歳出、性質別歳出などがそれである。ただし経済統計体系と社会・人口統計体系を体系として接合するという考えかたはとっていない。むしろ経済統計を社会・人口統計体系を利用するうえでの補助統計として扱っていると思われる⁽⁴⁾。
- ② 体系化にかんしては、実証統計の整理を第一にかんがえており、勘定体系の整備という視点はない。あるいは、勘定体系を改めて構築するよりは、既存の各種統計調査の体系を所与のものとかんがえている。
- ③ わが国の社会・人口統計体系は社会指標の基礎となる統計体系というよりはむしろ、指標そのものを提供している。報告書には、指標値と、指標算出の基礎となった基礎データが掲載されている。ただし、社会指標を体系として提示するものではなく、利用者が社会指標を作成するのに役立つように、指標の候補を提示するにとどまるというべきであろう。
- ④ クロス・セクションを主としているため、状態の変化は変化率などで示されている。
- ⑤ マイクロ・データ・セットの考えかたは特に意識されていない。家計ないしは世帯に関する調査統計を多く含んでおり、その意味では親近性をもつ。しかし目的が指標化にあるため個々の項目に重点がおかれており、主体としてのミクロ的経済単位にかんする統計体系とはいえない。なお、このことはつぎの地域区分と関連していると思われる。各種調査は、地域レベルではじゅうぶんな標本数をもっていないことが通例である
- ⑥ 地域区分については、社会・人口統計体系はすべて都道府県別の区分をおこなっており、もっとも重視されている。

もとより、わが国の社会・人口統計体系は国連のSSDSをそのまま日本に当てはめることを目的として開発されたものではない。むしろ生活の質重視という成熟社会における視点の転換に対応するため、SSDSのスピリットを受け継ぎつつ、わが国の事情にあわせてより実際的なアプローチをとったと言うべきであろう。しかし、社会・人口統計体系が一応の実用レベルに達した今日、今後の一層の発展のためにはいまいちどSSDSがもっていた特徴を想起することは有益であろう。社会・人口統計体系はこれまでのところ印刷物の形態で利用することが前提であった。しかしコンピュータの発達、とくにディスク・メモリーの大形化と価格低下、データベース理論の発達、データ・コミュニケーション技術の普及は、こうした前提をくつがえしつつある。そしてそれは、統計

の体系化、データの蓄積方法、指標などの加工統計の公表形態、関連統計の利用、など多くの面で、統計の作成者ならびに利用者の双方に新しい課題をなげかけている。

以下では、今日の政策分析上の要請、およびコンピュータ・サイエンスの発展状況から、社会・人口統計体系のデータベース構築にさいして考慮すべき点を検討することとしたい。

- (1) 既刊報告書としては、初期のものとして、「社会生活統計指標 体系と指標値」昭和52年12月、「社会生活統計指標 参考資料」昭和53年1月。最近のものとして、「社会生活統計指標 Statistical Indicators on Social Life」昭和59年3月、「統計でみる県のすがた」昭和53年3月、「社会・人口統計体系 基礎データ項目定義集」上下、昭和59年3月、などがある。
- (2) 総理府統計局『社会生活統計指標』昭和59年3月、まえがき。
- (3) United Nations Department of Economic and Social Affairs, *Toward a System of Social and Demographic Statistics*, ST/ESA/STAT/SER.F/18, New York, 1975.
- (4) 財政にかんする国連統計局の整理については、United Nations Department of International Economic and Social Affairs, *Classification of the Functions of Government*, ST/ESA/STAT/SER.M/70, New York, 1980, 参照。

2. 地域統計の特殊性とデータベース

社会・人口統計体系は地域別の表象を基本としている。統計における基本的な分類としては、産業、職業、年齢、などがあり、年次その他の時間軸をとまなう。これらについても社会・人口統計体系は随時利用しているが、比較的簡単にすませており、とくに体系をなしているということではできない。

また地域についても現状で公表されているのは都道府県別にとどまっている。この点は、全国集計量一本よりはるかに利用上の便宜を図ったものといえる。しかし、社会・人口統計体系の利用者の側からすると、十分なものとはいえない。しかし、これは直ちに一層の地域的細分化を要求するものではない。本節では、利用者が地域統計に何をもとめているかを検討することとしたい。

地域統計の利用者は多岐にわたることが予想されるが、

- ① 施策の立案、行政成果の評価
- ② 開発計画、環境影響評価、
- ③ 公共施設の立地
- ④ 交通計画
- ⑤ 宅地開発
- ⑥ 工場立地、大規模店立地

などが主なところであろう。

また、多くの場合、国民生活のたちばから見た地域は、行政区画とは一致しない。このことは、通勤圏、通学圏、買い物圏、レジャー圏、などなどを考えてみれば明らかであろう⁽¹⁾。

地域統計は、通常、市町村といった行政単位ごとに集計されているが、これらが行政上の理由や生活圏といった観点からいくつかにくくられて用いられることが多い。またわが国の場合、人口が密集しているため隣接市町村が外見的・機能的に区別できないほどとけあってしまっていることが多く、市町村という単位が実体的な意味をもたない場合がみられる。しかも、同じ県をとっても、地域区分は目的によって異なることが多い。

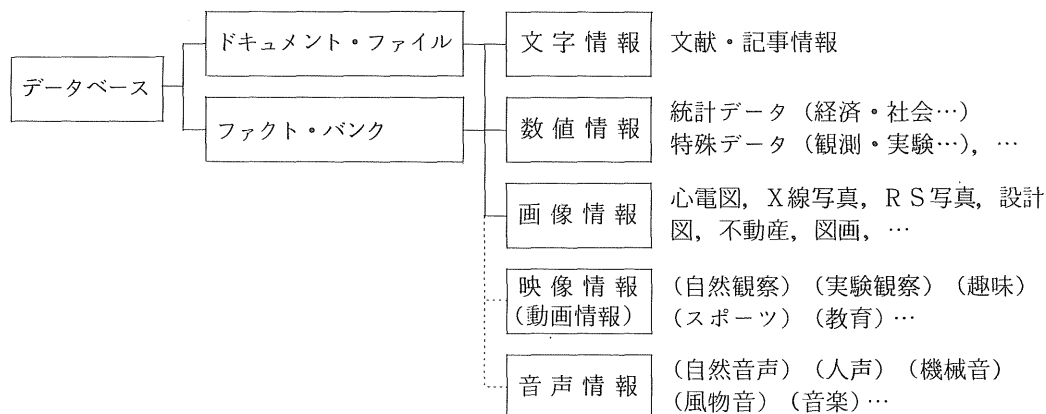
つぎに視点をかえて、地域統計として利用可能なものにどのようなものがあるかを検討してみよう。代表的なものを挙げれば、下記のようなものがある。

- ① 任意の面積を有するもの……行政単位、生活圏などに対応するもの。人口、生産、投資、財政、土地面積、等がその例である。
- ② 規則的な面積を有するもの……メッシュ・データ(グリッド・データ)、国勢調査、国土情報、住宅調査、事業所調査、等がその例である。
- ③ 線データ……道路、鉄道、河川、コンタライン、等がその例である。
- ④ 点データ……公共施設、事業所、文化財、等がその例である。地価のデータも特定地点におけるものである。環境汚染の測定値も観測点におけるデータである。
- ⑤ サーベイ・データ……世論調査、選好度調査などのサーベイ・データは、調査対象者が自らの住環境その他にかんする主観的評価を表明しているという意味で、調査対象者の所在地が意味をもつ。この点に関しては、上記の点データと類似性をもつ。
- ⑥ マトリックス・データ……特定の地点に関する情報ではなく、地点と地点の関係をしめすもので、O-D表などがその例である。上記②と異なり、面積を有さない。
- ⑦ 3次元データ……地形は本来立体であるが、通常は、コンタラインにより面として表現されている。大気の流れ、音の拡散なども本来は立体である。
- ⑧ 図面……地形図、土地利用図、地質図、等がその例である。人工衛星を利用したリモート・センシング・データもここにふくまれる。
- ⑨ 景観……山などの自然景観や、町並などの人工的景観がその例である。

これらは、デジタル・データとアナログ・データに大別できる。しかし、この区別は現在では画然たるものではない。土地利用現況、植生、道路、などの地図情報、リモート・センシングなどのカラーの図面はアナログ・データの典型であるが、コンピュータ・テクノロジーの進歩によって、容易にデジタル化することができるようになった。すなわち面積、座標、距離などの数値情報に置換して、分析に供することができるのである。また各種の情報を重層的にオーバーレイして利用することも可能である。従来もこれらのことは可能であったが、人手に頼らざるをえず、実用的ではなかった。

しかし、多数のデータを管理し、必要におうじて関連データを自由にとりだすことは、それほど容易ではない。従来のように、コンピュータ上のファイルを個別につくるやり方では限界がある。この機能を受け持つのがデータベース管理システムである。

データベースという言葉はいろいろな意味に用いられるが、『コンピュータ白書』は第1図のような分類を示している⁽²⁾。



第1図 データベースの分類

これによると、データベースはドキュメント・ファイルとファクト・バンクに大別される。前者は文献・記事情報などを内容とする文字情報の形をとる。これに対して、統計データは数値情報の一種として、ファクト・バンクのカテゴリーにふくまれている。しかし、データベースの分類としては、① 文献情報、② 事実情報、③ 統計情報、の3つのカテゴリーに分けることも可能である。あるいはこの分類のほうが優れているともいえよう。なぜならば、統計データの場合、データ同士の演算、加工、集計がおこなわれるのが普通であり、また、時系列的なアップデートがおこなわれることが多い、などの点で他の事実情報とはかなり異なるからである。

データベース・サービス業連絡懇談会によるデータベース関連用語の定義によれば、データベースとは「データを整理、統合し、コンピュータ処理が可能な形態でひとつの集合体にした情報ファイル」とされている。しかし、単にデータベースあるいはデータベース・サービスという場合、文献情報を指す場合が多い。これに対して、データバンクという言葉は「通常は、データベース・サービスを行なう組織、機関、部局といった意味」であるが、「データベースのうちファクト・データベース(数値など)を意味する場合もある」と注記されている⁽³⁾。このような用語の定義にもかかわらず、こと統計に関してはデータベースとデータバンクの相違は、注記がしめすように、それほど明確ではないのが実状である。その理由は、文献情報にかんしてはデータの管理にデータベース・マネージメント・システム(DBMS)が例外なく用いられており、また事実情報の管理(たとえば特許情報、法規、分子構造、など)についてもそうであるのにならして、統計データの管理については多数のファイルの集合といった形に止どまっているという実態がある。さらにいえば、これは現在のデータベース・マネージメントが個票を対象に行なうのには都合がいいが、統計のように一定の構造をもった集計量からできている場合には必ずしも適していないことがある。

統計のうちでもサーベイ・データのように個票を用いている場合には、データベース・マネジメント・システムの利用が適している。しかし、個票にアクセスすることができるのは、統計の作成にあたる部局、ないしは特別の許可をえた研究者に限られる。公表される統計データはある程度の集計データである。集計のフォーマットは統計毎にことなるが、地域、産業、職業など、基本的な分類基準やアップデートの頻度が利用者に共通に前提とされる知織となっており、その意味で一定の構造をもっているのである。したがって、その管理にあたっては、たとえば統計調査ごとのデータ・ファイルを作成しておくことがこれまでのところ便利であったわけである。

しかし、最近ではデータベース理論と統計データを架橋しようとする試みがでてきている。「…アイデアとしては、従来の統計データというのを一旦コンセプトチュアルなデータとして考える。そのもとでメタデータというのを全部記述する。その根元を押さえておいて、あとはいろんなアプリケーションで、こういう見方を見たいということを宣言した場合には、それに対しての変換プログラムというのを用意すればよろしいのではなからうか。」「メタデータというのは、データの内容とか、データを整理するときを使う記述内容とかの意味です。それがデータ本体と離れて、統計は存在するのがあたりまえである。多分、皆さん方はデータをつかうときに、データ自身は磁気テープに入っているんですが、その意味とか何とかは統計書の備考欄を見たり、あるいは頭の中にある知識を使ったり、ということでやられているのではないか。しかし、これは、意味も含めていろいろな解析をやっていこうというときに非常に妨げとなるわけですし、これが別にあるというのはちょっと困る。データベースでは、メタデータは実際はデータと一緒に扱うのが通常でありますので、そのテクニックが統計データにはないということが観察される。」「…実は、ビジネス・データでは、共有のメカニズムというのは、これは非常に簡単でありまして、エレメンタルなデータというのが、インディビジュアルなアトムのデータとして見えるわけです。したがって、共有もアトム単位でやればよろしいんですが、統計の場合にはアトムに当たるものがなくて、グループであるということで、共有の仕方は非常にむずかしいところがあるんですが、それにしても、何かがんばらないと統計の共有というのはできない。」⁽⁴⁾

議論を社会・人口統計体系にもどすと、データベース化に関連して検討すべきことは2つある。

- ① 都道府県を集計の単位とする現行体系をいかにしてデータベース化するか。
- ② 地域統計として利用可能なもののうち、都道府県のレベルのデータと補完的にもちいられるものはなにか。それらを①との関連のなかでいかにしてデータベース化するか。

このうち第2の点についてはメッシュ統計の活用が今後のポイントであろう。メッシュデータは単にそのメッシュ内の情報をあたえるだけではない。メッシュ情報を集計することによって、市町村その他の特定地域内の集計的情報をえることができる。メッシュ間の距離や時間距離といったデータを特定メッシュの属性としてとることも考えられよう。公共施設や道路などの点情報、線情報もメッシュにおとすことが可能であり、その他の情報と合わせて分析することができる。都道府県、市町村などの行政区分にもとづく統計データと、メッシュ統計とはこれまで別個のものとして扱われるきらいがあったが、それは単に分析のレベルの相異に帰すことはできない。むしろ、膨大

なデータ量に対応するハードウェア、ソフトウェアが無かったことが原因と思われる。データ量の点についていえば、特定地域内の全メッシュを対象とすることは必ずしも必要ではなく、いくつかのメッシュを抽出してもちいることも考えられてよい。

この場合、統計の作成者としてなすべきことと、利用者のがわで対応すべきこととの仕訳が必要であろう。

これらの検討事項は利用するハードウェア環境によって大きく左右される。そこでこの点について節を改めて論じたい。

(1) 行政目的に限っても、同じ県についても地域区分は、種々なものがもちいられている。茨城県の例では、商業統計、工業統計、農業基本調査のいずれも、県北地域、鹿行地域、県南地域、県西地域を区別しており、これらは県の出先機関である地方総合事務所の管轄区域と一致しており、この区分は県の福祉基本計画にももちいられている。神奈川県の場合では、工業統計について、横浜臨海地域、横浜内陸地域、川崎臨海地域、川崎内陸地域、三浦半島地域、湘南地域、県西地域、県央地域、県北地域、を区別しており、この地域区分は県の開発計画でも用いられているが、商業統計はこうした地域区分をおこなっておらず、農林業センサスについては異なったものもちいられている。

(2) 日本情報処理開発協会編、『コンピュータ白書』1982年版， p. 252.

(3) 前掲書， p. 253.

(4) 穂鷹良介，「データベース理論と統計データ」 *Statistical Data Bank*， 文部省科学研究費 特定研究 多目的総合統計データバンク News Letter Vol. 1， No. 4. また、実用化されている統計データベース・システムの一例については、佐藤英人，「共用クロスセクション・データベース・システムの概要」 *Statistical Data Bank*， Vol. 2， No. 1. 参照。

4. ハードウェア体系

地域統計の体系的利用にとっていくつかの制約がある。

- ① データ量が膨大になること……テープからの利用は繁雑であり、大形データ・ファイルをオンラインで持つことはコストがかかる。
- ② 既存統計が各種の出典に散在していること……データベースのデータベース（いわゆるメタ・データベース）が必要となる。しかし現実には包括的な情報は印刷物のかたちにおいてすら利用可能なものはない。
- ③ 分析の多様化……タイム・シリーズ分析とならんでクロス・セクション分析が必要となるケースが多く、またタイム・シリーズとクロス・セクションのプール化も可能でなければならない。
- ④ 既存データ以外に、業務資料など自己開発によるデータを併用することが多く、入力ソフトが必要となる。
- ⑤ 面・点・線にかかわるデータの処理……土地利用、道路・鉄道、公共施設などに関するデー

タを併用する必要がある場合がある。このためにはデジタイザーなどの設備が必要となる。

⑥ データの対象地域のレベルがまちまちである。……たとえば、都道府県レベルに関しても、個々の都道府県のみならず、全国8ないし9地域区分、表日本と裏日本、東海道メガロポリス、3大都市圏、農業県と工業県、などの分類がしばしば行なわれている。また原データから幾つかの地域を合計すること、あるいは必要な地域だけを抜き出すこともしばしばおこなわれる。このことはデータベース管理システムを必要とする。

⑦ データの表示……画面表示など、データ表示上の工夫がマクロ・データ以上に必要である。しかも、地域の事情がそれぞれ異なるため、そうしたソフトには汎用性が期待できない。

このうち、①⑤⑦はハード的な対応を必要とする。また①から⑦に至るいずれもソフト的な対応を必要とする。一般論はあまり意味がないと思われるので、これらに対する対応の例を紹介したい。いずれも筆者の研究室が関与しているものである。

筑波大学社会工学系ではこれまで数年にわたりミニコンピュータを制御機としてもちいる各種のシミュレータを開発導入してきた。この場合、大量のデータ処理・分析にはセンター・マシンを利用することが前提であった。しかし、昭和58年度に32ビットのスーパー・ミニコンピュータを導入した。現在のシステム構成を第2図に示す。センター・マシン以外に、専用のミニコンピュータを持つ必要性は、研究環境によって異なることはいうまでもない。場合によってはセンター・マシンを用いることができようし、またそのほうが望ましい。われわれの場合、導入のメリットとして次の点があげられる。

① 特殊機器の制御……グラフィック・レスポンス・アナライザー、プロッター、その他の特殊機器をローカルに制御する必要がある。その際、これまで制御機の能力が低く、機器毎にべつべつの制御機が必要となり、研究効率がわるかった点を改善する必要があった。導入後はそれまで散在していたシミュレータ類のネットワーク化が可能となった。

② ファイル転送が不要となる……シミュレータ類とセンターとの間のファイル転送時間がなくなり、処理時間を短縮できる。

③ 大容量計算への対応……アドレス空間がセンター・マシンの130倍まで拡張できるので多量の統計データを処理することが可能である。データ開発時には大きなファイルをメモリー上にもつ必要がある。シミュレーション・プログラムはかなり大形である例が多く、センターへ投入すると他のジョブへ支障をきたすことがあった。変数選択をおこはう統計処理用プログラム、大形のLP計算などもおおきなメモリー・サイズを必要とする。

④ 大容量ディスク・ファイルをもつ……最大オンライン・ストレージが18.5ギガ・バイトとなっており、大量の統計データをオンライン・ディスク上に記録し保存することができる。これまでは磁気テープで入手したデータは、必要なときにコンピュータのなかによみこみ、利用が終わると再びテープにおとしていた。テープによってデータ・フォーマットが異なるため、出典の異なるデータを組み合わせて利用することには、大きな困難がともなった。大容量ディスク・ファイルを持つことによって、データを標準フォーマット化しオンラインで常駐させることが可能

になった。

⑤ ソフト的は研究環境の安定……研究目的の場合、特殊な目的を迅速かつ効率的に達成するため、システム側が提供するサブルーチンを利用してアプリケーション・プログラムを開発する(たとえば、科学技術計算、線形数学、都市計画、グラフィック、データベース管理システム、など)。そのため機種の入れ換えはそのたびにアプリケーション・プログラムの書き直しを余儀なくさせるなど、望ましくない。この点で、汎用システムを提供するセンターと目的が異なる。

⑥ 稼働率の確保……センター・マシンは週末は停止するのが普通であり、また大学の場合は入試業務その他で一般の利用が制限されることがある。一部のシミュレータ制御などのためにセンターを特別に稼働させることは困難であるが、専用ミニコンピュータであれば土・日、休暇中なども時間的制約を受けずに処理が可能である。

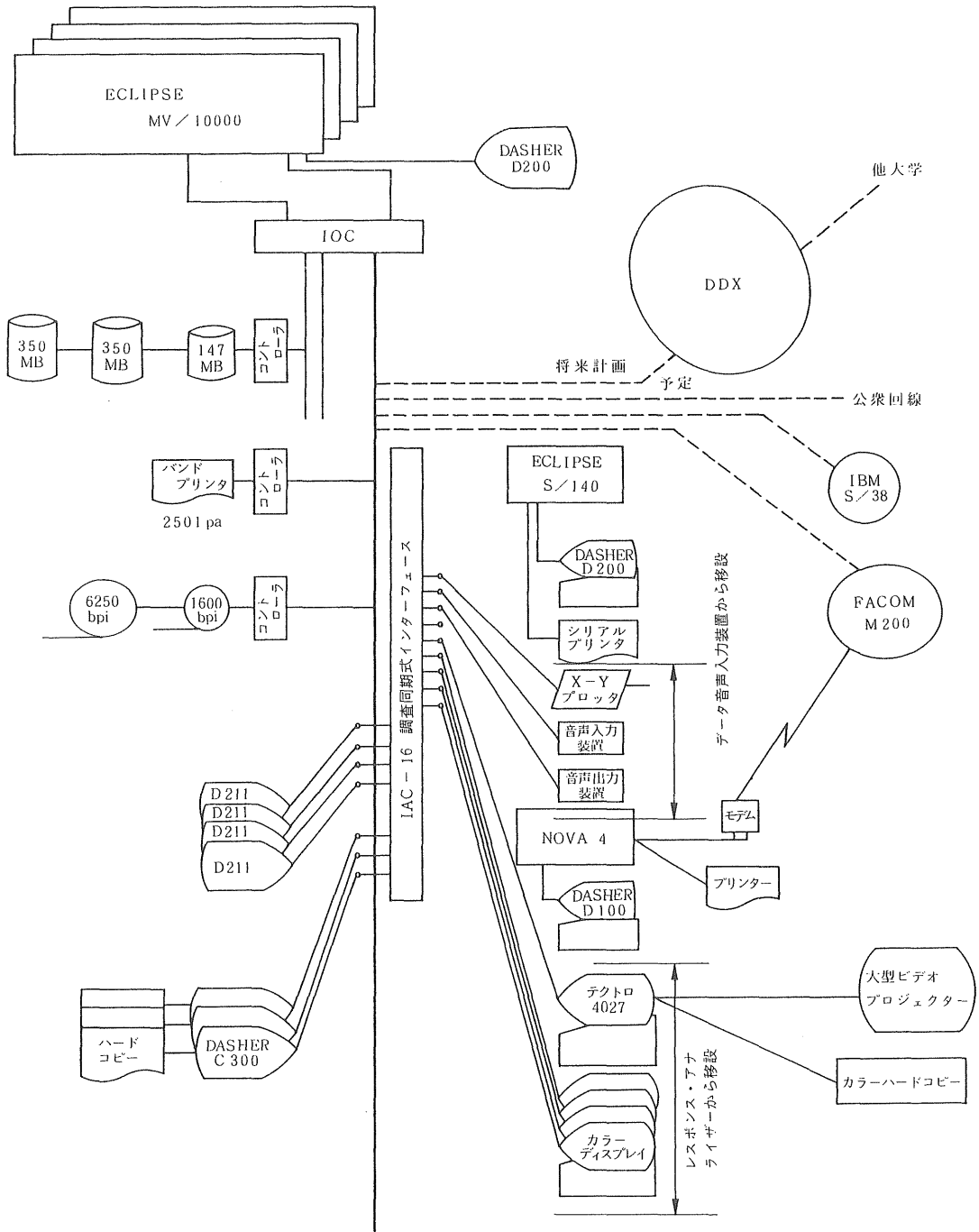
図に示したシステムでは、次の特殊装置を有している。

- ① グラフィック・レスポンス・アナライザー……カラー・ディスプレイ4台、高解像度カラー・ディスプレイ、それに接続した大形ビデオ・プロジェクターおよびカラー・ハードコピー・ユニット
- ② プロッター
- ③ データ音声入出力装置
- ④ ハードコピー装置付きグラフィック端末4台
- ⑤ 大容量ディスク・ユニット
- ⑥ コミュニケーション・システム……センター・マシン、データベース・マシン(IBM S/38), 公衆回線, 電電のDDX網, パソコンなどとのコミュニケーション機能

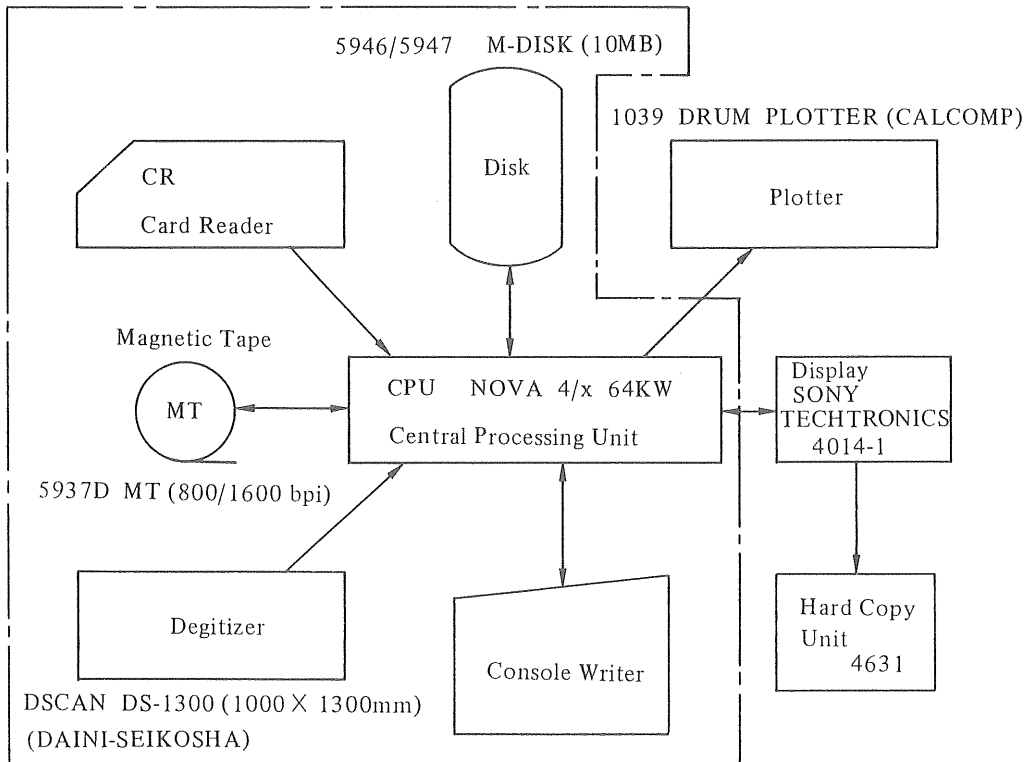
このシステムは地図情報の処理能力を持たないが、この点は環境シミュレータ、地域空間シミュレータが担当している。第2図は環境シミュレータのハードウェア構成をしめす。このシステムは次の特殊装置を有している。

- ① 図形情報変換器……地図などのアナログ情報のデジタル化装置
- ② ディスプレー端末……2次元・3次元情報の表示
- ③ ハードコピー・ユニット
- ④ 大形プロッター

この装置の制御機は小形なため、大規模な演算をおこなったり統計データベースを構築することはできない。将来的には上記のデータバンク用システムとネットワークを組むことが望ましい。



第2図 多目的統計データバンクのハードウェア



第3図 環境シミュレータの機器構成図

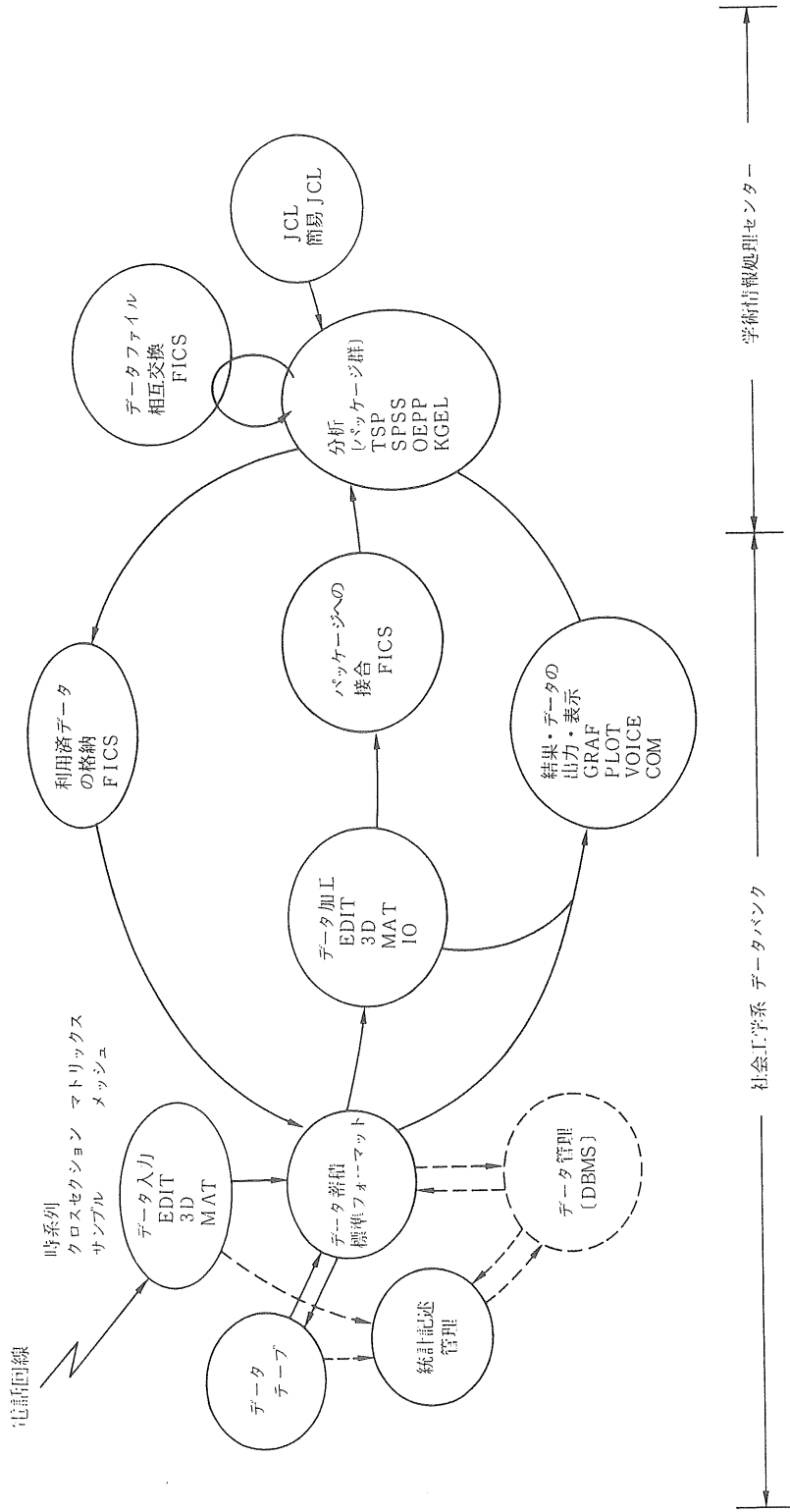
5. ソフトウェア体系

つぎにソフトウェア体系であるが、第4図は社会工学系において現在、実証的な運用テストの段階に入った多目的統計データバンク用ソフトウェアを示す。

ソフトウェア「体系」といっても1つのソフトウェア・システムがあるのではなく、機能ごとにモジュール化されている。設計思想としては1つの機能をもったソフトウェアの入口と出口についてだけ前もって標準化しておき、内容については必要に応じて機能アップを図っていくことをねらっている。それによってソフトウェアの変更がユーザーに及ぼす影響を最低限に保つことができ、ユーザーからすると安定したソフトウェア環境を保てることになる。また既存のソフトウェアではカバーしきれていない部分をユーザーの立場から埋めていくことになること、いくつかのソフトウェアを同時並行的に開発しなければならないこともモジュール方式とならざるをえない面もある。

さらに研究プロジェクト側が開発にあたるため、経費面で長期的開発プランをたてにくいこと、ソフトウェアの開発要因として学生をあてざるをえず長期にわたって熟練した人手を確保できないといった、実際的な問題もモジュール化を促した。

これらはいずれも標準フォーマットのデータ入力をおこなうだけでなく、データ加工およびかなりの程度の分析まで行うことができる。マトリックス・データのうち産業連関分析のように手順が



第4図 多目的統計データベース用ユティリティ・ソフトウェア

確立しているものについては、別個の分析プログラムが用意されている。

分析の手順にしたがって図を説明すると、まずデータテープが入手可能な場合が左端に示されている。テープのフォーマットはそれぞれ異なるので、統計データバンクとしてはこれを標準フォーマットに変換することによって、利用の効率化を図る必要がある。標準フォーマットは現在、時系列ないしクロスセクションデータとマトリックス形式のデータについて定められている。

データ開発の必要がある場合には、研究者がデータ入力をおこなう必要がある。データの種類に対応して、データ入力、編集、および簡単な分析をおこなうソフトウェアを開発した。大別すると、第1は時系列ないしクロスセクションのデータ系列に対応するためのもの、第2はマトリックス形式やメッシュ・データなどマトリックスとして取り扱えるデータに対応するためのもの、第3は表形式のデータを時系列に蓄積し、マトリックスとしてもあるいは系列としても取出すことができるものとなっている。

分析の段階では既存のパッケージ・プログラムに依存することが便利である。しかしパッケージはそれぞれ特定のファイル編成法をとっている。そこで標準フォーマットからそれぞれのパッケージが要求するフォーマットへ自動的に変換・接合するプログラムが用意されている。このプログラムはまた、パッケージ・プログラム内部のデータファイルを他のパッケージでの利用のために相互交換することができる。

センター・マシンの利用にあたっては、研究者はコンピュータの専門家ではないから、JCLの切り方にとまどうことが多い。そこでJCLのライブラリー・ファイルをつくり、利用者はジョブの種類によって必要なものを呼び出して使うようにした。

利用の終わったデータ・ファイルは、パッケージの中でさらに加工・開発され、保存に値するものも多い。そこで標準フォーマットに変換のうえ、データバンク側に格納することになる。

分析結果については、従来のようなプリンターへの出力だけでなく、各種のものが利用可能である。第1はグラフィック表示である。ハードウェアのところで触れたように、カラー・グラフィックへのデータやシミュレーション結果の表示をおこなうことができる。第2はプロッターへの出力である。第3は音声出力である。第4はコンピュータ出力を直接マイクロ・フィッシュ化するシステムである。

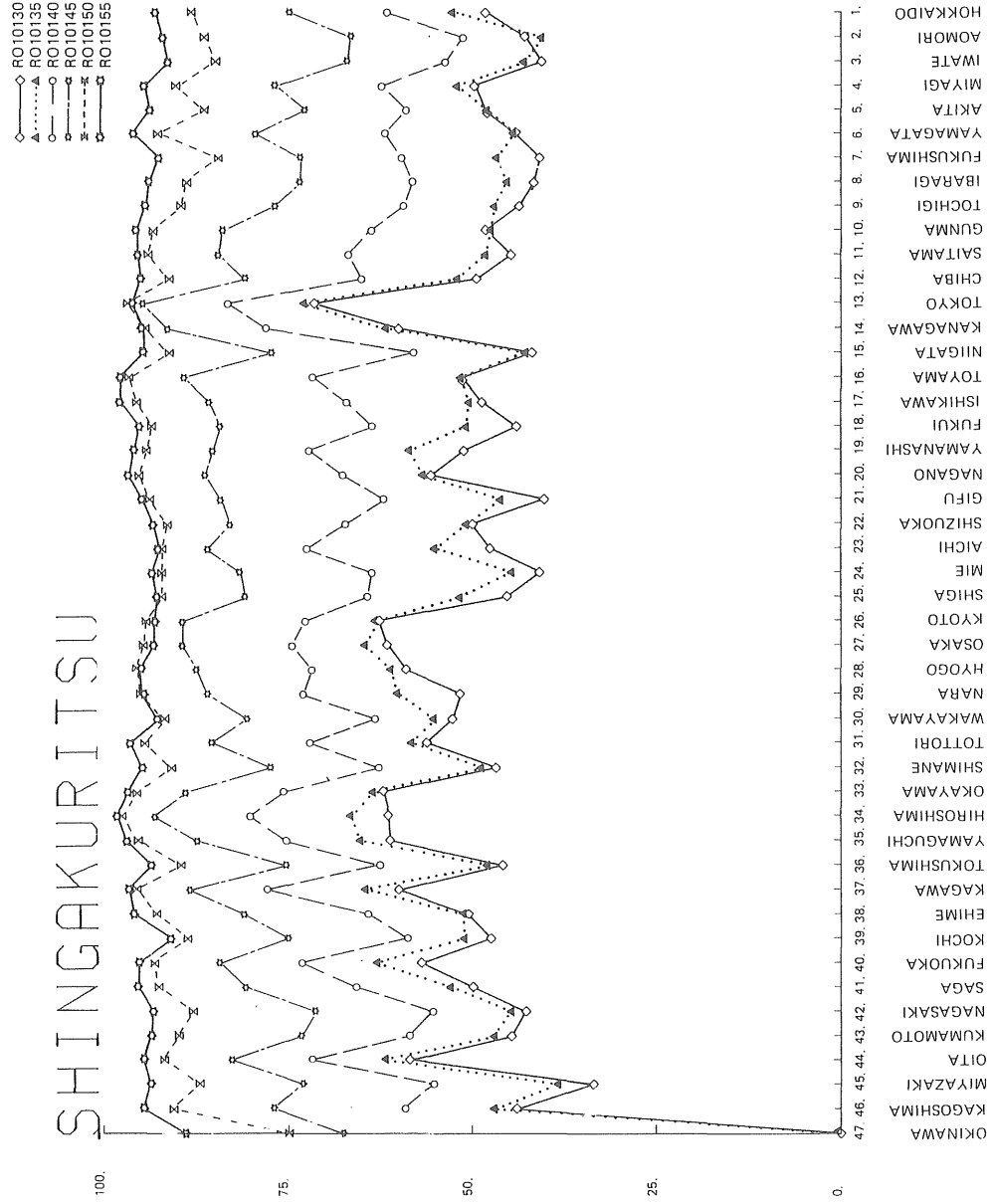
将来的には統計の定義や加工方法についても同一のシステムの上でコンピュータ・ファイル化することが必要である。現在のところの機能はパソコン上のワードプロセッサを利用している。また統計データのデータベース的管理については、IBM System 38を利用して別途研究が進行中である。

社会・人口統計体系をはじめとする地域統計の利用にあたって特に重視する必要があるアプリケーション・ソフトは、生データと既存のパッケージ・プログラムをつなぐ部分であろう。データバンク用ソフトの特徴は次の通りである。

① 統計データ編集用ソフトウェア (EDIT, Empirical Data Input and Transformation System)

- a. 各種のデータ加工をおこなうことができる。累計、ラグ、系列リンク、成長率、その他を含む。
 - b. データ音声による入出力を行なう。データ・チェックなどに利用。
 - c. ファイル内の統計データの属性リスト、検索、名称変更、削除、期間変更など、データ開発時に多用される機能をそなえている。
 - d. 月次、四半期、半期、暦年、年度など、各種の期種のデータを統一的にあつかえる。
 - e. クロスセクション・データのプール化など、複雑なファイル操作が簡易化されている。
 - f. 複数のファイルから任意の系列を選んで、新しいファイルを作ることができる。
 - g. 複数のファイルを容易にマージさせることができる。
 - h. 統計処理を行なうことができる。
 - i. 統計データをグラフィック端末にグラフィック表示できる。(開発中)
- ② プロッター・プログラム
- a. 統計データ編集用プログラムとリンクして利用できる。
 - b. 時系列、クロス・セクションの作図出力ができる。(第5図)
 - c. 構成比の作図出力ができる。
 - d. 都道府県別の作図出力ができる。ある統計データがある場合、基準を与えてぬりわけることができる。(第6図)
- ③ 3次元マトリックス編集用ソフトウェア(MAT, Matrix Automatic Transformer)
- a. 地域・産業・年次など3つのディメンジョンをもったデータは非常に多いので、これに対応することができる。マトリックスをどの次元でも取り出すことができるほか、時系列データあるいはクロス・セクション・データとしてベクターとしても取り出すことができる。
 - b. 各項目にラベルをつけることができる。また、マトリックスのます目に対応して注を付すことができる。
 - c. 取り出した系列は、系列名を付したうゑ、統計データ編集用ソフトウェアに送り込むことができる。
 - d. マトリックス演算機能を有している。
- ④ ファイル変換システム(FICS, File Conversion System)
- a. 現在流通しているアプリケーション・ソフトウェアはおのおの固有のデータ編成法を有しており、統一性が欠如している。本システムは標準フォーマットを介して、データ・フォーマットの相互変換を容易に行なうことができる。サポートしているアプリケーション・ソフトは第7図のとおりである。
 - b. アプリケーション・ソフトの内部データ・ファイルおよび外部ファイルのいずれもサポートしている。

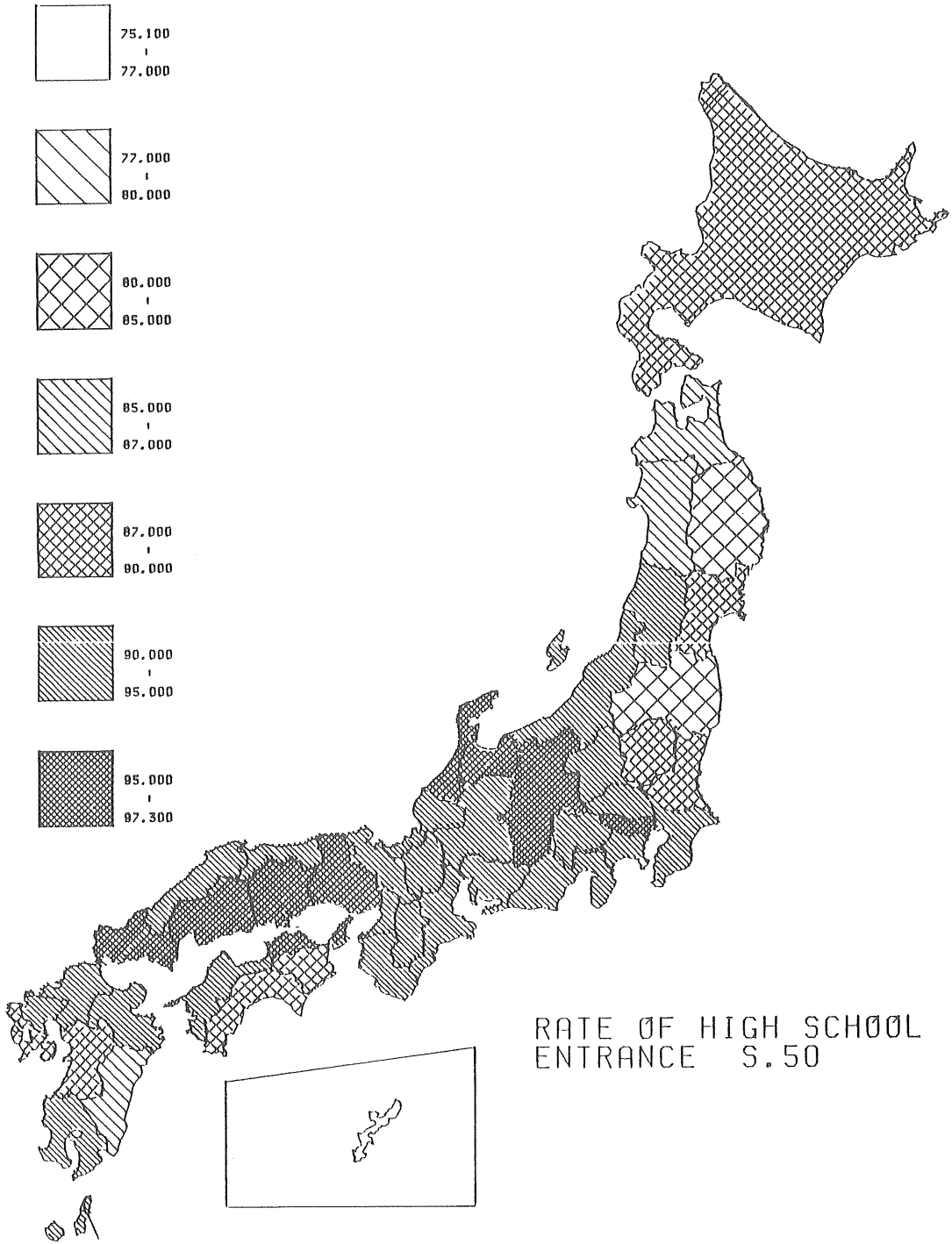
SHINGAKURITSU



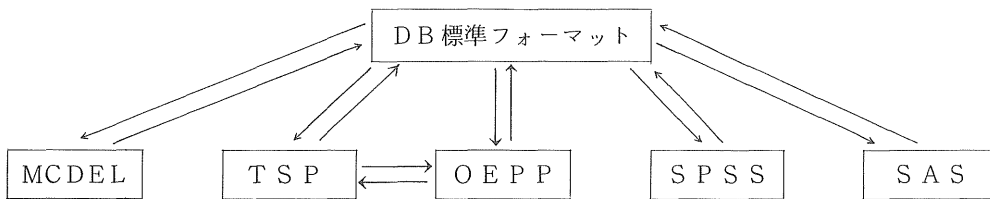
第5図 クロセクション・データのプロッター出力例
(高校進学率, 昭和30年—55年)

- Okinawa
- Kagoshima
- Miyazaki
- Oita
- Kumamoto
- Nagasaki
- Saga
- Fukuoka
- Kochi
- Ehime
- Kagawa
- Tokushima
- Yamaguchi
- Hiroshima
- Okayama
- Shimane
- Tottori
- Wakayama
- Nara
- Hyogo
- Osaka
- Kyoto
- Shiga
- Mie
- Aichi
- Shizuoka
- Gifu
- Nagano
- Yamanashi
- Fukui
- Ishikawa
- Toiyama
- Niigata
- Kanagawa
- Tokyo
- Chiba
- Saitama
- Gunma
- Tochigi
- Ibaragi
- Fukushima
- Yamagata
- Akita
- Miyagi
- Iwate
- Aomori
- Hokkaido

RATE OF HIGH SCHOOL



第6図 地域データの地図出力例（高校進学率，昭和50年）



第7図 データ・フォーマット相互変換プログラム

⑤ グラフィック・レスポンス・アナライザー（GRAF, Graphic Response Analyzer and Forecasting System）

- a. メッシュ・データのカラー・グラフィック表示，および大形ビデオ・プロジェクターへの出力ができる。さらにハードウェア体系の項でのべたように，グラフィック端末のRGB信号をうけて，画面のハード・コピーをとることができる。（第8図）
- b. 外部から基準をあたえて表示のカラーを変更できる。
- c. 統計データのメッシュ入力ができる。
- d. メッシュ・データのベクター化，およびベクター・データのメッシュ化ができる。これにより分析用ソフトウェアへデータを送り込むことができる。

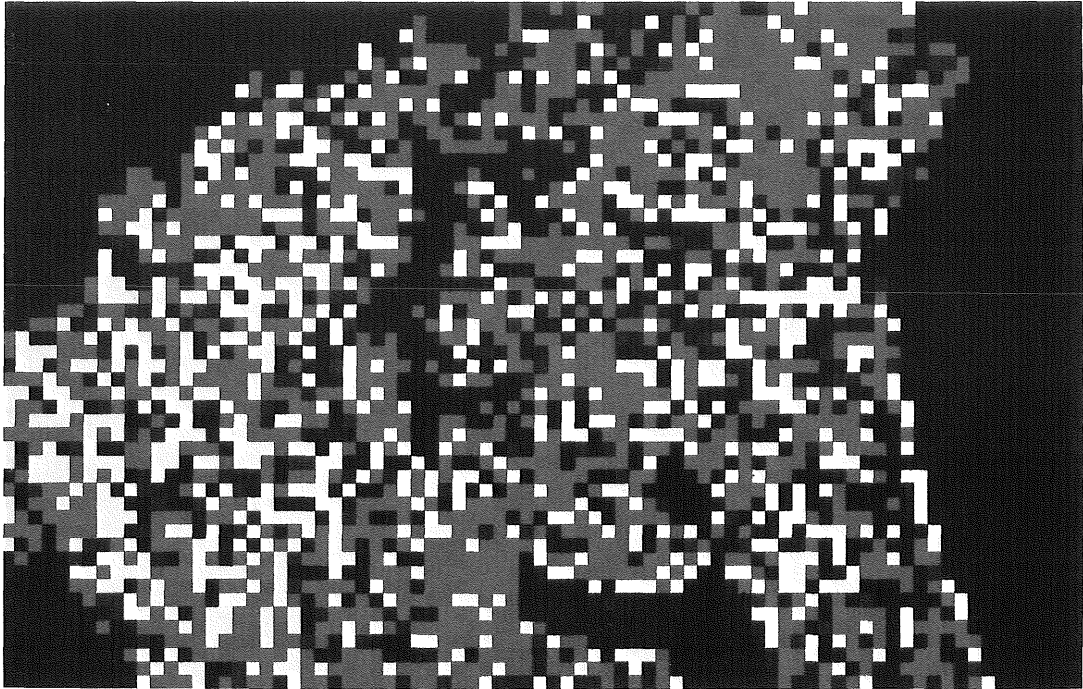
全体としての特徴は，ソフトが行き止まりにならないよう，統計データの相互変換に留意していることがあげられる。また数量分析にかんしては既に各種のアプリケーション・ソフトが流通しているのでこれを最大限利用しつつ，欠けている機能をデータバンクとして補う方針をとっている。

む す び

本稿では，社会・人口統計体系あるいはそれに類似した地域統計のデータの提供ならびに利用にあたって懸案となっている事項を検討した。それはつぎの点に要約できよう。

- (1) 地域統計の体系的提供と利用の必要性はきわめて高い。
- (2) 地域統計の多様さにかんがみてデータベース管理システムをもちいる必要がある。ただしその際，これまでのものは個々のサンプルや個票があることを前提に考えられており，集計量を扱う統計データに直ちに応用できるシステムは未開発である。
- (3) 地域統計の提供と利用にあたっては，ハードウェア的な対応を要する。恐らくそれは個々のユーザーにとっては手に余ることであり，それが地域統計の利用を妨げてきた。データ・コミュニケーション面での技術進歩は新しい提供と利用のシステムを可能にしつつある。
- (4) さらに，地域統計の分析，表示などのために，ソフトウェア的な対応も必要である。

以上の項目は研究テーマとしてはかなり膨大であるが，本稿で紹介した程度のプロトタイプはできており，一層組織的な努力がなされるならば，その実用化は決して夢物語ではないことが納得いただけたものと思う。



第8図 メッシュ・データの出力例(茨城県南部における第1次産業就業者数)

紺	$x = 0$
青	$0 < x < 50$
緑	$50 \leq x < 100$
黄	$100 \leq x < 200$
赤	$200 \leq x$

グリッドは標準メッシュをあらわす。

資料『国勢調査』