

土木研究所資料

第 1239 号

データ・バンクに関する研究

昭和52年 3 月

建設省土木研究所
企画部システム課

土木研究所資料

第 1239 号

データ・バンクに関する研究

昭和52年 3 月

建設省土木研究所
企画部システム課

ま え が き

情報化社会といわれるようになってから久しく、多くの分野に大型電子計算機が導入利用され、膨大な量の情報の処理加工が行われている。また、これに伴い多くの情報収集が行われている。しかし、これらの収集、処理、加工された情報が、有効に利用されているとはいいがたい。その理由の一つに、情報システムの整備、とりわけ、情報を蓄積し、提供するためのデータ・バンクの整備が、不十分であることがあげられる。一方、建設事業をとりまく環境の変化に伴い、建設行政を円滑に実施するためには、従来にもまして多種多様かつ、大量の情報が必要となっている。

このような認識のもとで、システム課においては、建設行政のあり方およびこれを実現するために必要となる方法論を検討するとともに、建設省における情報システムのあり方に関する調査研究を実施してきた。

本研究は従来から収集、蓄積されてきた大量の情報や、今後収集される情報を有効に利用するために、情報システムの一環としてのデータ・バンクのあり方を検討したものである。

本報告書では、従来の研究成果をもとにして、建設行政にふさわしいデータ・バンクの構成を提示するとともに、ケース・スタディとして流水管理業務を対象とした、流域データ・バンク・システムをとり上げ、電子計算機を活用したデータ・バンク・システムの、設計上の諸問題について検討を行った。

本報告書が、今後の情報システムの整備のための、一助となることができれば幸いである。

なお、資料の収集整理等は、奈良大橋設計センターが担当した。

昭和52年3月

土木研究所企画部システム課長

岩 松 幸 雄

(研究担当者)

三 好 逸 二

関 正 和

吉 川 勝 秀

目 次

第1章 序	1
第2章 データ・バンク・システムの現状	3
第3章 建設行政のためのデータ・バンク	5
3-1 情報システムの全体構成とデータ・バンク・システム	5
3-2 データ・バンクのシステム構成	5
3-3 普遍系による情報の流通	8
第4章 流域データ・バンク・システムの設計概念	10
4-1 流域データ・バンク・システム	10
4-2 情報インデックス	12
4-3 情報の国土空間への対応	14
4-4 情報の構成要素	19
第5章 データ・ベース	21
5-1 流域データ・バンク・システムのデータ・ベース	21
5-2 ソフトウェア体系と機能	22
5-3 ファイル処理の概要	24
5-4 データの構成と内部表現	29
5-5 ファイル設計	31
5-6 データ・ベースとの対話	38
第6章 結 語	42
付 録 1. 既存のデータ・ベース・マネージメント・システム	45
2. データ・バンク・システムの事例	64
3. 地理情報の表現	83
参考文献および参考資料	87

第 1 章 序

公共事業の規模の増大とそれを取りまく環境の変化に伴い、建設行政に対する要請が多様化し、多方面から評価がなされるようになってきた。したがって、事業の実施や管理に際しては、多種多様な要請に対する合理性を明らかにするとともに、それらが地域社会におよぼす派生的な影響を事前に把握することが求められている。これらの要請に応えるためには、建設行政を取りまく環境を認識し、事業の発揮する効果を派生的な影響も合わせてもれなく予測し、その妥当性を検討することが重要である。

各種事業（河川・道路・住宅・下水道等）を有機的に結合させて、建設行政に対する多角的な要請に応えるためには、計画立案に関する方法論（あるいは計画論¹⁾）の確立とともに、これに必要な情報を提供する情報システムの整備が必要である。

本報告では後者の情報システムの整備の一環として、そのシステムを構成する重要なサブ・システムであるデータ・バンク・システム^{2) 3) 4) 5)}を対象とし、収集、蓄積されてから利用に供される情報のシステムティックな取り扱い方および建設行政にふさわしいデータ・バンク・システムのあり方について検討する。

従来から河川・道路等の事業の実施のために、それぞれの事業主体は目的に応じて必要とする情報を収集し、将来の利用に備えて蓄積してきている。しかしながら、現状ではそれらが必ずしも有効かつ便利に活用されているとはいいがたく、死蔵されているものも少なくない。また、建設行政と地域社会の相互関係がますます重視されるようになると、各事業主体は従来から収集・蓄積してきた情報以外に、経済・社会・環境等の地域社会の情報をも必要とするようになってきた。このような新たな種類の情報は、河川・道路・住宅・下水道等の部門で共通に使用されることが多いので、相互利用しうる形で蓄積しておくことが望ましい。このような情勢に対して、近年、各方面で情報の有効な利用を行うために情報システムの整備、特にデータ・バンクの必要性に対する認識がたかまっている。体系的なデータ・バンクに関する研究は電子計算機関係の分野を中心として進められており、後に述べるようにすでにいくつかのデータ・バンク・システムが実用化されている。

データ・バンク（情報の蓄積）について考える場合には、次のようなことを念頭ににおいておくことが重要である。

a) すでに収集・蓄積されている情報の活用

効率的な情報の検索・提供および死蔵されている情報の活用がこれに含まれる。

b) 新たに収集する情報の効果的な蓄積・管理

収集した情報のうち将来利用されうるものを効率的に蓄積し、必要に応じて提供できるように管理する。この中には、従来から集取してきた情報とは異なる種類のものも含まれる。

c) 情報の相互利用

必要とする情報がすでにどこかで収集・蓄積されており利用可能ならば、それを新たに収集する労力が省かれる。したがって、共通に利用する情報は相互利用が可能なように蓄積・管理すべきである。

データ・バンクの検討においては上記 a)~c) を可能にするために、情報自体が具備すべき要件および情報を収集・蓄積・検索・提供するシステムの構成について調査・研究する必要がある。我々が対象とする情報は属地性の情報（位置を指定してはじめて意味をもつ情報）が多く、建設行政のためのデータ・バンクはこれを前提として作りあげる（あるいは再構成する）ことが重要である。

本報告書では、第2章でデータ・バンクに関する従来の研究とすでに実用化されている情報機器を利用したデータ・バンク・システムの事例を紹介し、第3章で建設行政にふさわしいデータ・バンクのシステム構成を明らかにし、第4章でケース・スタディとして電子計算機の活用を前提にした流水管理のための流域データ・バンクの考え方を示し、第5章でそれを構成するために必要なデータ・ベースの細部構成を述べ、第6章で本研究の結論を記す。なお、ここでデータ・ベースという言葉を用いたが、この用語は電子計算機を活用したデータ・バンクを考える人々の間ではデータ・バンクとほぼ同義語として解釈されている。最近の電子計算機を活用したデータ・バンクは汎用的な利用を前提としていることが多く、データ・ベースはそのためのソフトウェアとして用いられているものであり、今後はそのような意味でこの言葉を使用することにする。

第2章 データ・バンク・システムの現状⁶⁾

国内でのデータ・バンク・システムの実施例をみると、閉じた利用目的を持ったものが大半を占めている。例えば、国鉄のみどりの窓口で代表されるような座席予約システムであるとか、建設省の特殊車輛通行許可業務を行う特認システムなどがあり、これらは、単一の利用目的を持ったシステムである。一方、種々の業務に共通に利用される、いわゆるデータ・ベース機能を持ったシステムの実施例は数少ない。なお、データ・バンク・システムの実施例はその取り扱い情報量の大量なことから電子計算機の利用を前提としたものがほとんどである。

データ・バンクに関する研究、開発は各方面で活発に行われて来ており、その一例として、昭和48年に建設省大臣官房政策課の委託により地理情報システム (Geographic Information System; 略称GIS)⁸⁾ 研究委員会が設立された。ここでの研究は、国土情報の地理的表現に重点を置き、情報は地理への属性として定義される。地理は、不定形多角形としてとらえられ、点、線、地片に分解される。情報はこれらの属性として定義づけられる。また、地理はグラフ構造としてとらえられている。この方式は、地理と情報が一体となっていることから、地理構造にそった情報の検索が可能であり、また、地理とそれに含まれる情報を現実の詳細を損なわない形で蓄積される特徴を持っているが、しかし、広地域に対するデータの収集、作成には膨大な時間と経費を要し、現実的に具体化するにはかなりの困難が予想される。

具体的な開発例としては、東京都のデータ・バンク・システムがある。^{9) 10) 11)} 東京都は、昭和48年にデータ・バンクの開発を目的とした組織を設立させ、昭和50年後半において、地域計画支援パイロット・システムを完成させている。このシステムは、東京都の行政区画単位で収集された情報を時系列的に蓄積したもので、会話型による検索を可能としているが、点、線および地理情報は現段階では含んでいない。

一方、一般にデータ・ベース理論¹²⁾と言われるデータ・バンクの理論的研究も活発に行われている。¹³⁾ データ・ベース理論とは、情報をいかにモデル化(形式化)するかということに主眼をおいたものであり、データ・ベースは、情報のデータ値のみを扱うものでなく、情報相互間でもつ論理関係をも一緒にバンクするという事に基づいている。モデルは、従来から様々のものが提案されており、最近では、E・F Codd の提唱した、リレーショナル・モデル¹⁴⁾(関係形式モデル)が、データ・ベースを研究する技術者間でもてはやされている。しかし具体的な実用化には、まだまだ困難な問題を含んでいる。

データ・バンク・システムの構築のためのソフトウェアは、汎用データ・ベース・マネージメント・システムとして電子計算機メーカなどで商用化されているが、これらはユーザから見た場

合、単にデータ・ベースを構築する場合のデータ構造の枠組にしかすぎず、現実の情報をいかにしてこれに当てはめていくかという問題は残ってしまう。実はこれが、データ・ベースを構築する場合の一番困難な問題なのである。

データ・バンクは色々と困難な問題を含み、我国のデータ・バンクの実用化は、まだその緒についたばかりといってよい。

なお、データ・ベース・マネージメント・システムおよび、データ・バンク・システムの事例については、付録 1、2 を参照されたい。

第3章 建設行政のためのデータ・バンク

ここでは、建設行政で取り扱う情報の性質および利用方法に基づいたデータ・バンク・システムについて考察する。

3-1 情報システムの全体構成とデータ・バンク・システム

建設省のデータ・バンク・システムのあり方について考察する前に、情報システムの全体構成と、その中におけるデータ・バンク・システムの位置づけを明確にしておくことにする。というのは、我々が行ってきたデータ・バンクに関する調査・研究は、情報システム全体に関するものではなく、一度収集・蓄積されてから利用される情報の具備すべき要件およびその取り扱い方に関するものだからである。情報の中には、収集されてから蓄積されることなく利用されるものもある。一般に、情報がどのような経路を経て利用に供されるかを図3・1に示す。

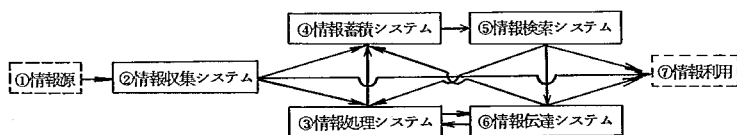


図3・1 情報システムの全体構成

情報は、情報システムに①から入力されて⑦へ出力される。データ・バンクの研究で対象とする情報は、①から⑦へ伝達する間に、一度は、④、⑤の情報蓄積・検索システムを経るものである。したがって、データ・バンク・システムは図3・1に示される全てのサブ・システムを包含しているけれども、情報システム全体ではない。

我々がデータ・バンク・システムの検討において主として考察しているのは④および⑤の情報蓄積・検索システムの機能と、それに必要な情報の具備すべき要件についてである。

3-2 データ・バンクのシステム構成

一般に、データ・バンク・システムは、情報の蓄積の仕方によって集中蓄積型（資料館型）システムと分散蓄積型システムに大別され、各々次のような特徴がある。集中蓄積型システムは、情報の収集・管理・運営に膨大な経費と労力を要する上、大規模な初期投資が必要である。もし小規模なものであれば利用者全体にとってはメリットが少ないし、逆

に情報の収集量（範囲）が多ければ、システム全体の利用効率は低下するというディレンマが生じる。一方分散蓄積型システムでは、情報を最も利用する所に分散して蓄積するため、情報の利用効率は高くなる。しかし、情報が各所に散在しているために、情報検索・伝達機能がきわめて重要な役割を果たし、これらを充実させる必要がある。

現在の建設省関係のデータ・バンク・システムは分散蓄積型であり、各地建・事務所が各々独自の方法で膨大な情報を蓄積している。しかし、これら個々のデータ・バンク間の情報のやりとりはあまり行われていない。将来はこれらのデータ・バンクを結合した建設省としてのデータ・バンク・システムを整備して行く必要があると考えられる。また、我々が対象とする情報は事務関係の情報とは異なり属地性のものが多くその土地で使用されることが多いこと、さらには経済的な問題、現存するシステムの有効利用、将来の段階的な発展等を考慮すると、建設省にふさわしいデータ・バンク・システムとしては分散蓄積型が適しているといえよう。

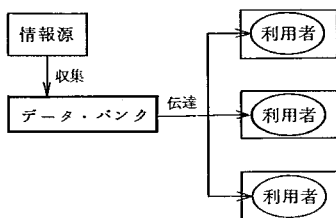


図3・2 集中蓄積型データ・バンク・システム

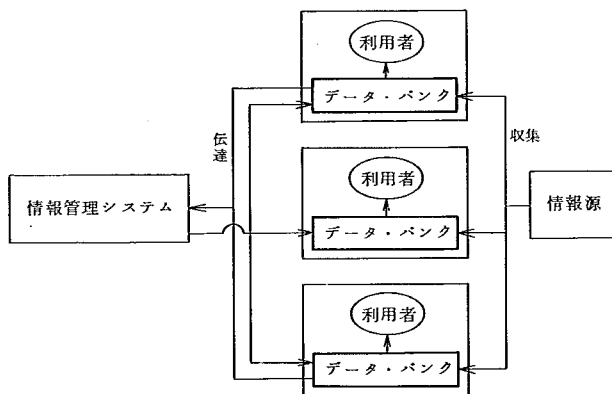


図3・3 分散蓄積型データ・バンク・システム

前述のように分散蓄積型システムでは、情報検索・伝達機能を整備する必要がある。この検索方法としては電話番号簿型システム（さらに発展させた電話番号案内型システム）が考えられる。すなわち、各所に散存する個々のデータ・バンクでは情報が蓄積されると、それを電話番号簿と同様の形式で登録する。これにより、情報利用者は自分の必要とする情報がすでに収集されているか否か、収集されているとすればどこにあるか（データの有無・所在）を知ることができる（この方法による検索簿の作り方等については昭和51¹⁵⁾年度の建設省技術研究会河川部門指定課題報告書を参照）。図3・4に建設省にふさわしいデータ・バンク・システムの全体構成を示す。

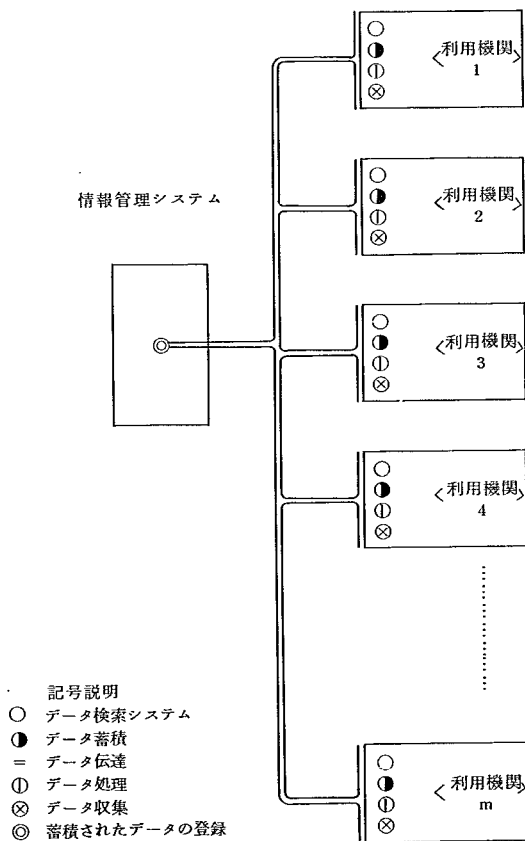


図3・4 建設省のためのデータ・バンク・システム

また、各サブ・システムのもつ機能はつぎのとおりである。

- a) 情報収集システム……必要情報を観測・測定調査等によって人手・自動計測機等により収集する機能。
- b) 情報処理システム……すでに収集または蓄積されている情報から解析手法の適用により必要情報を加工・処理する機能。
- c) 情報蓄積システム……将来の必要に備えて、すでに収集・加工された情報を蓄積しておく機能。
- d) 情報検索システム……必要情報をすでに蓄積されている情報群の中から検索する機能。
- e) 情報伝達システム……収集された情報または検索された情報を蓄積場所・利用場所に伝達する機能。

3-3 普遍系による情報の流通^{3) 4)}

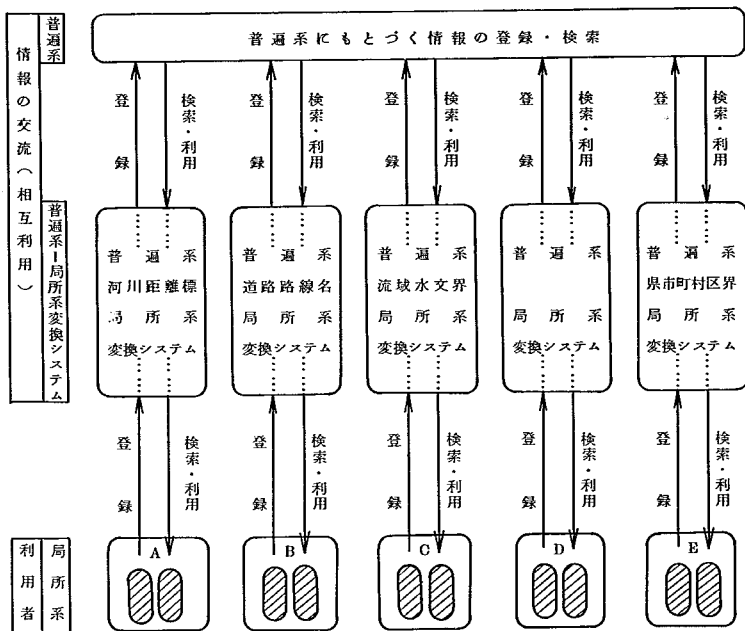
前述のように分散蓄積型システムでは、情報はそれを最も必要とするところ（機関）に最も利用しやすい形態で蓄積される。したがって、各機関で蓄積されている情報を相互利用するためにはある種の工夫が必要となる。ここでは、相互利用を行う方法として、図3・5に示す普遍系の導入による情報の流通を行う方法を提案する。

図3・5に示される個々の機関が用いている情報の整理方式を「局所系」とよび、それぞれの局所系の相互間で情報を流通させるために用いる情報の整理方式を「普遍系」とよぶ。

今、局所系Aの情報を局所系Bで利用するためには、まず局所系Aの情報を、普遍系との変換システムにより普遍系の形式で表示する。その後、この情報を普遍系と局所系Bとの変換システムにより、局所系Bの形式で表示しなせばよい。

このように、普遍系による情報の整理方式を設定するとともに、各局所系と普遍系との変換システムを作成しておけば、局所系相互間の情報の流通が可能となり、分散蓄積型システムにおいてもあたかも集中蓄積型システムと同様な情報の利用が可能となる。

この普遍系による表示は、共通的なデータの利用を可能にするための表示形式であり、分散蓄積型システムに、いわゆるデータベースとしての機能を発揮させるためのものである。



各種局所系にもとづく情報の登録・検索・利用

(河川距離標による局所系) (道路路線名による局所系) (流域・水文界による局所系) (県市町村区界による局所系)

図 3・5 普遍系の導入による情報の流通

第4章 流域データ・バンク・システムの設計概念^{16) 17) 18)}

ここでは、建設行政のためのデータ・バンクの考え方を明らかにするために流水管理に必要な情報を対象とした流域データ・バンク・システムをケース・スタディとしてとり上げ、電子計算機の活用を前提としたデータ・バンク・システムの検討を行った。

4-1 流域データ・バンク・システム

データ・ベースは、データの独立性を重要な設計要因としているが、流域の情報に限定して考えても、その情報は膨大であり、また情報形態も種々雑多である。このような情報に対してデータ・ベースの設計概念を汎用性に置いて議論を始めると、議論は発散し、結果的にはデータ構造が定まらない事になる。そこで、データ・ベースを設計する場合は、システムを性格づけるための概念の導入が必要になってくる。

流域データ・バンク・システムは、あらゆる情報を対象とした汎用的なデータ・バンク・システムをめざすものではなく、汎用的なデータ・バンクのサブ・システムとして、すでに収集されている流域情報を対象とし、それらの情報をシステムティックに蓄積しておくものである。本システムを性格づける概念は、次節以降で示して行くが、本システムのアウトラインは、次のとおりである。

a) システムの適用

本システムは、主に流域の水問題に関する計画業務および流水管理業務の支援システムとして利用される。

b) システム構成

流域情報の収集、利用の現状からみて、本システムは、分散蓄積型システムとすることが適当であるが、この場合、多数の現存するコンピュータ・システム（ハードウェア、オペレーティング・システム）下でも稼働できる事が必要である。

また、システムは段階的な開発により構成されるものであり、初期のシステムにおいてはごくシンプルなシステムが望ましい。よって、ハードウェアの構成は図4・1のよう
に想定し、処理形態はバッチ処理とする。システムの概要を図4・2に示す。

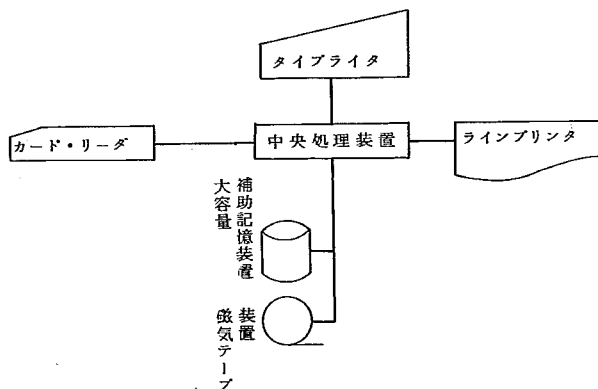


図 4・1 最小機器構成

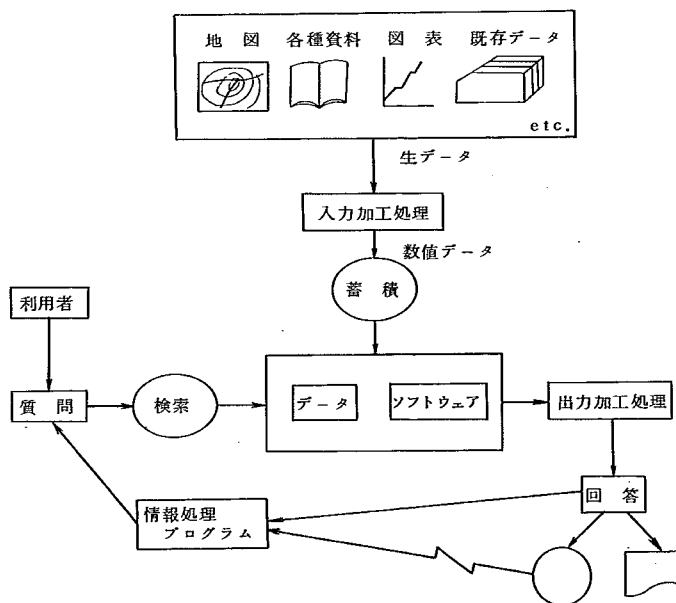


図 4・2 システム概要

4-2 情報インデックス

情報は、いつ(時刻)、どこ(位置)、何に関する(項目)情報なのかが規定されて、はじめて意味ある情報となる。本システムではこの時刻、位置および項目の三つの指標を情報インデックスと呼ぶ。例えば、「A地点の1976年の年平均流量は20.5 m³/secであった。」という情報は、表4・1のように表わされる。

情報インデックス			内 容
時 刻	位 置	項 目	
1976年	A	年平均流量	20.5 m ³ /sec

表4・1

(1) 時 刻

時刻は、データが調査、観測された時点または期間を表わし、ほとんどの情報は、規則的な時刻サイクルで調査、観測されているので、データは時系列として蓄積される。

時刻は、西暦または年号による年月日、日本標準時の時分秒で表示される。

(2) 位 置

位置は、各位置につけられた固有のコード、または名称で表示される。ここで、位置は、情報が調査、観測された有義領域の違いにより、ゾーン(行政区画単位)、メッシュ、および不定形の三つに大別できる。これらの表示方法は次節でのべる。

(3) 項 目

項目は、項目名または、項目コードによって表示され、項目に付けられるコードは、情報の分類に基づいて決められる。

情報の分類は、従来から図書や文献の整理のために種々の分類法が提案されて来たが、情報項目は、科学技術の発展などにもない常に更新が余儀なくされるものであり、固定的なものとしてとらえる事はできない。また、既に提案されている種々の分類法が、あらゆる目的にとって納得のいく十分なものであるか否かは、はなはだ疑問であり、これをそのままとりあげて用いる事は問題が多い。

一般的な分類手法としては、図書館などで使われている日本十進分類法(NDC)や、国際十進分類法(UDC)がある。これらは、体系的分類の代表的なものであり、分類は木構造で、逐次上位概念から下位概念に分かれていく。

日本十進分類法は、すべての同一概念は十分類され、それぞれ0から9までのコードが与えられる(図4・3参照)。

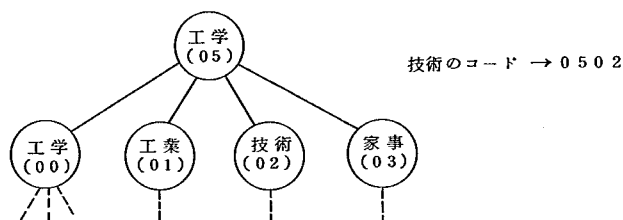


図 4・3 日本十進分類法

簡単な分類手法として、項目を核化して音順などにより線形的に分類する方法があるが、これは更新に不便をきたし適当ではない。

本システムにおいては、項目の更新が容易であり、かつ情報の蓄積、検索が分類されたコード体系にそってなされる事を前提とすると、階層分類で、コード体系は、階層レベルを表現できる方式が望ましい。

以上によって、本システムでは、日本十進分類法に準じた方式で図 4・4 のような分類およびコード体系を提案する。

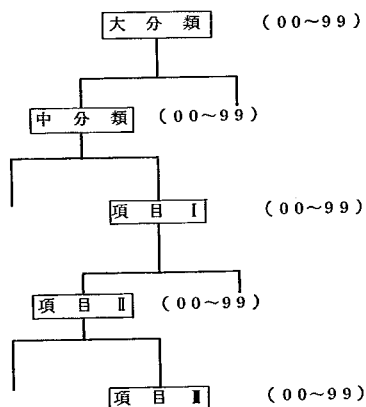


図 4・4 分類およびコード体系

図 4・4 は、階層レベルを 5 段階で制限したものであり、情報のコードは最高 10 桁

である。

(4) 内 容

時刻、位置、および項目で示されるデータの内容は、一つのデータ値を持つものと、多次元の配列データを持つものがある。例えば、2次元で、 (3×2) の配列の大きさが与えられた場合は、データの内容は6個のデータ値で構成される。

4-3 情報の国土空間への対応

(1) ゾーン・データ

ゾーン・データは、単一行政区画単位、または、複数行政単位で収集された時系列的データであり、情報源の行政区画単位により次の3種に大別することができる。

(ⅰ) 都道府県別ゾーン・データ

都道府県単位で収集されたデータ。

(ⅱ) 市区町村別ゾーン・データ

市区町村単位で収集されたデータ。

(ⅲ) ブロック別ゾーン・データ

複数の行政区画単位で収集されたデータ、例えば、東海、京浜地区とか県南、県北地区などである。

この他にも、町丁目単位などのゾーン区画が考えられるが、初期の流域情報システムでは考える必要がないと思われる。

この大別したゾーンは、それぞれ階層構造でゾーン間の関係づけがなされる。例えば東京都は、下位の行政単位の市区町村を含み、中央区と千代田区は同レベルのゾーンであるといえる。

本システムでは、各ゾーン・データの情報インデックスにおける位置インデックスは、都道府県コード、市区町村コードおよびブロック・コードで表現する。都道府県コードについてはJIS(C6260¹⁹⁾)を、市区町村コードについてはJIS(C6261¹⁹⁾)をそれぞれ使用し、ブロック・コードについては一般的に規定されたものがない。

各コードは表4・2、表4・3でもわかるように、次のような値をとる。

(ⅰ) 都道府県コード…………… 01~47

(ⅱ) 市区町村コード…………… 01100~47382 (頭2桁は都道府県コード)

(ⅲ) ブロック・コード……………

データは、ゾーンの属性として、図4・5のように対応づけられる。

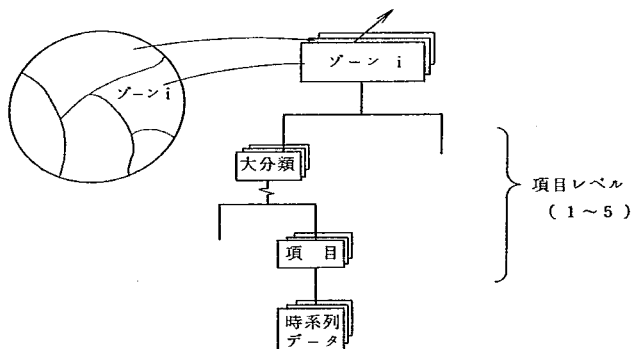


図 4・5 ゾーン・データ構造

01 北海道	13 東京都	25 滋賀県	37 香川県
02 青森県	14 神奈川県	26 京都府	38 愛媛県
03 岩手県	15 新潟県	27 大阪府	39 高知県
04 宮城県	16 富山県	28 兵庫県	40 福岡県
05 秋田県	17 石川県	29 奈良県	41 佐賀県
06 山形県	18 福井県	30 和歌山県	42 長崎県
07 福島県	19 山梨県	31 鳥取県	43 熊本県
08 茨城県	20 長野県	32 島根県	44 大分県
09 栃木県	21 岐阜県	33 岡山県	45 宮崎県
10 群馬県	22 静岡県	34 広島県	46 鹿児島県
11 埼玉県	23 愛知県	35 山口県	47 沖縄県
12 千葉県	24 三重県	36 徳島県	

表 4・2 都道府県コード

北 海 道

100	札 幌 市	さっぽろし
101	中 央 区	ちゅうおうく
102	北 区	きたく
103	東 区	ひがしく
104	白 石 区	しろいしく
105	豊 平 区	とよひらく
106	南 区	みなみく
107	西 区	にしく
201	削 除	
202	函 館 市	はこだてし
203	小 樽 市	おたるし
204	旭 川 市	あさひかわし
205	室 蘭 市	むろらんし
206	釧 路 市	くしろし
207	広 市	おびひろし
208	北 見 市	きたみし
209	夕 張 市	ゆうばりし
210	岩 見 沢 市	いわみざわし
211	網 走 市	あばしりし
212	留 萌 市	るもいし
213	苫 小 牧 市	とまこまいし
214	稚 内 市	わっかないし
215	美 唄 市	びばいし

表 4・3 市区町村コード（一部）

(2) メッシュ・データ²⁰⁾

メッシュ・データは、メッシュ区画で収集されたデータが時系列的に蓄積されたものである。メッシュへの情報の対応は、ゾーンと同様である。

メッシュは表4・4の方法で区画割がなされ、原理的にはいくらでも細分化できるが、現実的なデータ収集においては、第4次区画である100mメッシュまで十分ではないかと思われる。

名 称	作 成 方 法	辺 長	備 考
第1次地域区画	全国の地域を1度ごとの経線と偶数緯度及びその間隔を3等分した緯度における経線とに1つて分割した区画	約80Km	20万分の1地勢図と同区画
第2次地域区画	第1次地域区画を経緯線方向に8等分した区画	約10Km	2万5千分の1地形図
第3次地域区画	第2次地域区画を経緯線方向に10等分した区画	約1Km	基準地域メッシュ
第4次地域区画	第3次地域区画を経緯線方向に10等分した区画	約100m	
第5次地域区画	第4次地域区画を経緯線方向に10等分した区画	約10m	
第6次地域区画	第5次地域区画を経緯線方向に10等分した区画	約1m	
第7次地域区画	第6次地域区画を経緯線方向に10等分した区画	約0.1m	

表4・4 メッシュ区画

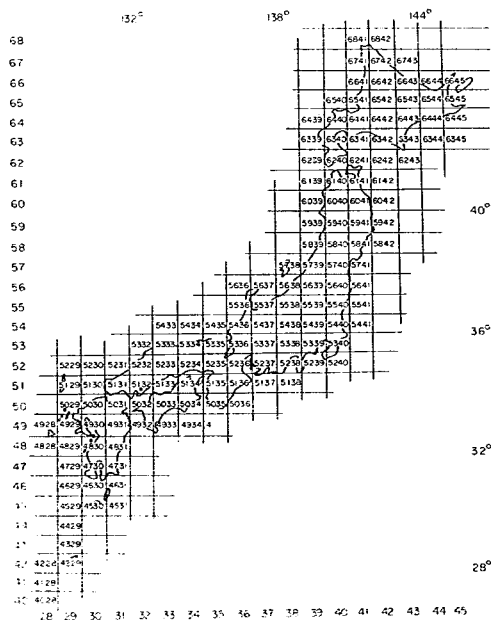


図4・6 第1次地域区画(80kmメッシュ)

メッシュの第2次区画は第1次区画を縦横に8等分(図4・7参照)し、第3次区画より下位のメッシュはそれぞれ上位のメッシュを10等分(図4・8参照)し番号づけがなされる。

7 0	7 1	7 2	7 3	7 4	7 5	7 6	7 7
6 0	6 1	6 2	6 3	6 4	6 5	6 6	6 7
5 0	5 1	5 2	5 3	5 4	5 5	5 6	5 7
4 0	4 1	4 2	4 3	4 4	4 5	4 6	4 7
3 0	3 1	3 2	3 3	3 4	3 5	3 6	3 7
2 0	2 1	2 2	2 3	2 4	2 5	2 6	2 7
1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7
0 0	0 1	0 2	0 3	0 4	0 5	0 6	0 7

図4・7 第2次区画

9 0	9 1	9 2	9 3	9 4	9 5	9 6	9 7	9 8	9 9
8 0	8 1	8 2	8 3	8 4	8 5	8 6	8 7	8 8	8 9
7 0	7 1	7 2	7 3	7 4	7 5	7 6	7 7	7 8	7 9
6 0	6 1	6 2	6 3	6 4	6 5	6 6	6 7	6 8	6 9
5 0	5 1	5 2	5 3	5 4	5 5	5 6	5 7	5 8	5 9
4 0	4 1	4 2	4 3	4 4	4 5	4 6	4 7	4 8	4 9
3 0	3 1	3 2	3 3	3 4	3 5	3 6	3 7	3 8	3 9
2 0	2 1	2 2	2 3	2 4	2 5	2 6	2 7	2 8	2 9
1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7	1 8	1 9
0 0	0 1	0 2	0 3	0 4	0 5	0 6	0 7	0 8	0 9

図4・8 第3次区画以上

メッシュ・コードは第4次区画まで考慮するとすると10桁のコードとなる。

第1次区画…… 4028 ～6842

第2次区画…… 402800 ～684277

第3次区画…… 40280000 ～68427799

第4次区画…… 4028000000～6842779999

(3) 不定形データ

不定形データはゾーンやメッシュ単位以外で収集されたデータであり、例えば、ある流域の流域面積、流量観測点での流量値などである。

不定形データの位置には、ゾーン、メッシュと同様にコードが付けられる。

以上のように情報はゾーン、メッシュ、不定形への属性としてまとめる事ができる

(図4・9参照)。

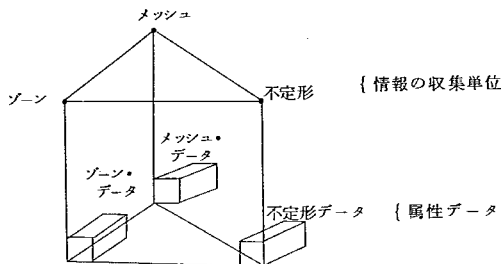


図4・9 国土空間に対する情報の対応

4-4 情報の構成要素

情報の定義尺度は先にのべた三つの情報インデックスである。しかし、これだけでは、情報を完全なものとしてとらえる事はできない。情報は表4・5のような情報が附加されて完全なものとしてとえられる。

情報 イン デックス	位 置	データが定義されている位置名および位置コード
	項 目	データ項目名および項目コード
	時 刻	データが定義された時点および期間
データ内容		データ属性値
データ 操作 情報	時刻間隔	データの時系列間隔
	単 位	ディメンション
	精度、タイプ	データの有効桁数およびデータ・タイプ
	配列データ	データの配列次元および配列の大きさ
補足 情報	観 測 者	データの調査、観測者名または機関名、出典等
	観 測 方 法	観測機器等
	加 工 方 法	データの加工処理方法
	備 考	データ利用上の注意等

表 4・5 情報の構成要素

表 4・5 において、データ操作情報は、データの検索時に参照されるものであり、情報インデックスと同様にシステム内部に蓄積しておく必要がある。補足情報は、情報の注釈的なものであり、文章でないと理解しにくいものである。したがって、この情報はデータベースとは分離し、データ索引誌に掲載する。

第 5 章 データ・ベース

5-1 流域データ・バンク・システムのデータ・ベース

コンピュータの基本ソフトウェアは、オペレーティング・システムと汎用言語である。オペレーティング・システムは、我々ユーザには、直接みえてはこないが、コンピュータを稼動する上でなくてはならないものである。

一方、直接的にユーザとかかわるもので、問題解決手順のアルゴリズムを記述する言語である。最近、この二つの基本ソフトウェアに加えて、第3の基本ソフトウェアといわれるようになってきた、汎用データ・ベース・マネージメント・システムがある。これは、前者の二つのソフトウェアと比べると、開発の歴史は浅く、十分な使いこなしが行われていない。また、既存のシステムの仕様はまちまちで、十分な理論的根拠にも乏しい。このような現状で、安易に既存の汎用データ・ベース・マネージメント・システムの導入は危険である。

流域データ・バンク・システムへの、汎用データ・ベース・マネージメント・システムの導入は、前述の問題の他にも次のような問題がある。

- a) 一旦ある特定のシステムを導入してしまうと、すべての設計基準が、そのシステムの仕様にしぼられ、簡単な仕様変更も困難となる。
- b) 分散蓄積型システムの指向のためには、分散システム間の互換性が必要である。しかし、既存のシステムは、異なるオペレーティング・システム間での互換性がない。
- c) 汎用性に主眼を置いている事から、簡単なものでも、複雑としているきらいがある。実際、情報全域をカバーするような汎用システムが存在するかどうかは疑問である。
- d) データ・ベースの運用、管理は複雑であり、これにたずさわる人々の特別な知識を要する。

以上のような問題をみると、流域情報システムとしては、既存の汎用システムの導入は、適当でないと思われる。しかし、既存のシステムの技術は、本システムにも利用される。

本システムのデータ・ベースは、汎用的な情報を扱うものでなく、すでに収集されてある流域情報を対象としたものであり、図5・1のように整理された情報を対象とする。

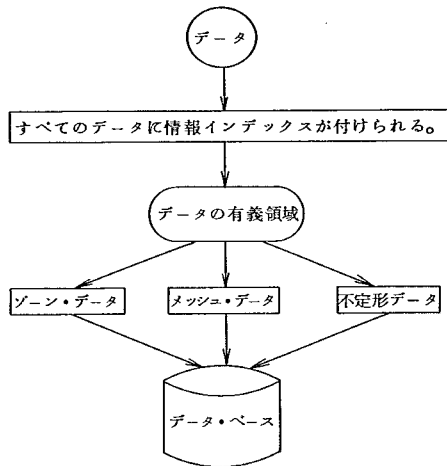


図 5 ・ 1 対象とする情報

5-2 ソフトウェア体系と機能

流域情報システムのソフトウェア体系は図 5 ・ 2 のようになる。

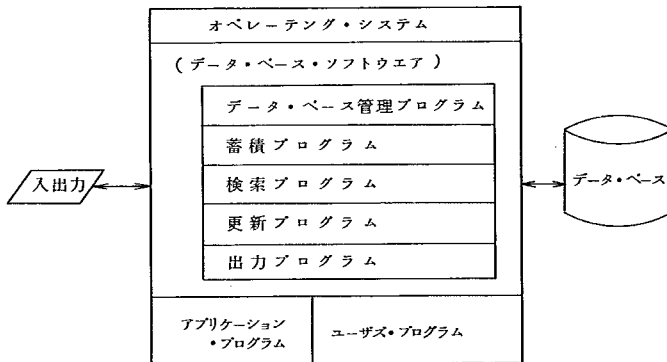


図 5 ・ 2 ソフトウェア体系

a) オペレーティング・システム

コンピュータを効率的、効果的に利用するため、一つまたは、複数のジョブの流れを総合的に制御するシステムである。通称OSと呼ばれている。オペレーティング・システムはメーカよりコンピュータに付属されて提供されるものであり、ハードウェアと一体なものとして理解されてよい。

b) データ・ベース管理プログラム

データ・ベースを総合的に管理するプログラムである。機能の主なものは、次のとおりである。

- (i) 記憶スペースの管理（記憶スペースの利用状況出力等）
- (ii) データ・ベースの退避、復元（磁気テープへ）
- (iii) 蓄積情報の出力（蓄積されている項目のリスト・アップ等）

c) 蓄積プログラム

初期のデータ・ベースを作成するプログラムである。主な機能は次のとおりである。

- (i) 記憶場所の割りつけ
- (ii) 構造定義
- (iii) 索引テーブルの作成
- (iv) 蓄積（決められた記憶場所にデータをストアする。）
- (v) 入力データの検査

d) 検索プログラム

蓄積されてあるデータを要求により取り出すプログラムである。主な機能は次のとおりである。

- (i) データの記憶場所の検索
 - ・構造検索
 - ・索引テーブルの参照
 - ・直接検索
- (ii) データの転送（2次記憶（ディスク）からバッファへデータを転送する）
- (iii) 条件検索
- (iv) コマンド・インタプリタ

e) 更新プログラム

データ・ベースを更新するプログラムである。主な機能は次のとおりである。

- (i) データの追加

(ii) データの削除

(iii) データの変更

(iv) データの再配置 (更新がたびたびかさなるとむだなスペース、レコード・ポインタなどが増えファイルの効率が低下するためファイルの中味を整理する。)

f) 出力プログラム

検索されたデータをレポートするプログラムである。出力機器は、ラインプリンタ、磁気テープであり、図化機とのインターフェースも可能である。

g) アプリケーション・プログラム

コンピュータが対象とする問題の解決に対する支援を目的にして作られたプログラム群である。例えば、流出解析、降雨予測、内水解析などがあり、固定化された問題をルーチン化してシステムに登録しておく。

h) ユーザズ・プログラム

ユーザが問題解決のために作成したプログラムである。ユーザズ・プログラムの中で検索ルーチンをコールする事によりデータ・ベースとインターフェースがとられる。

以上が、本システムのソフトウェアの体系と概略の機能である。

次節以降において、データ・ベース・ソフトウェアの中心である蓄積・検索のためのファイル処理について具体化してのべる。

5-3 ファイル処理の概要

情報には、情報の種類や性格があり、その特徴は異なっている。このような情報をいかに蓄えておくかがファイル設計である。したがって、ファイル設計は、情報の性質を調べ、それに適したファイル処理の方法、および、蓄積媒体を選定する必要がある。

(i) 蓄積媒体

蓄積媒体としては、磁気テープ、磁気ディスク、磁気ドラム等がある。

磁気テープは、データを順に書き読みするシーケンシャル・ファイル (一定の順番で読み書きされるファイル) として利用され、特に即時性の要求されない大量の情報を廉価に蓄積できる。すなわち、データの利用は、特定の順番でデータが要求される処理には向いているが、乱順処理には検索に多大の時間を要する。一方、並び順が不規則な情報を蓄積処理する場合は、磁気ディスク、磁気ドラムが使用される。磁気ドラムは、特にアクセス (ファイルに格納されているデータを主記憶装置にとりだす) 時間の速さを要求される情報の蓄積に適している。磁気ディスクは、アクセス時間は、長くなるが、

大量の情報を蓄積できる。最近では、ギガ (10^9) 単位の容量のディスクも開発されるようになって来ている。

データ・ベースのファイル処理は、乱順処理が必要不可欠であるとともに、大容量のファイルを必要とする事から蓄積媒体は、磁気ディスクが選定されよう。また、補助的な蓄積媒体としてデータ・ベースの退避や、部分的なデータ・ベースのコピーとして磁気テープが使われる。

(2) ファイル設計の条件

ファイルの設計は、次の様な条件を考えながら、データの記憶方法やアクセス方式が決定されなければならない。

- a) アクセスに要する時間の短縮
- b) 記憶効率の向上
- c) ファイルの更新の容易さ
- d) 将来の拡張性

(3) 情報の物理表現

データの記憶構造上での物理的な単位を、ワード、レコード、ファイルおよびエリアの四つに分割する事ができる。ワードは、記憶の最小単位であり、例えば一つの流量値がこれに該当する。レコードは、複数のワードの集合で構成され、書き込み、読み込みの最小単位である。

ファイルは、同一の概念でまとめられたレコードの集合である。例えば、気象データ、水位・流量データなどがこれに該当する。また、エリアは、論理的にも物理的にもデータ・ベースを分割するものであり、一つまたは、複数のファイルで構成される。エリアは、ゾーン・データ、メッシュ・データなどがこれに該当する (図 5・3 参照)。

情報の物理表現を考える場合、記憶の長さは、一定 (固定長) として考えるほうが、ファイル操作は簡単である。しかし、情報は、時系列データや配列データであり、記憶の長さは、固定的でない。但し、固定的に時系列や、配列の長さを決定すれば、この限りでないが、あきらかにこれは、記憶面のむだを生ずることになる。したがって、ここでは、上記したレコードの固定長化を図り、データは、複数組のレコードで収容される (図 5・4 参照)。

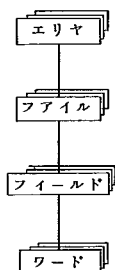


図5・3 データ・ベースの物理的分割

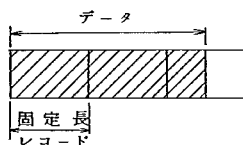


図5・4 固定長レコード

(4) ファイルの編成方法

データの蓄積・検索は、ファイルの編成からみると一体なものである。このファイルの編成の特徴は、ファイルのアクセス方法によって異なる事ができる。アクセス方式の基本方式は、次の三つに大別できる。

a) 順次アクセス(図5・5)

連続的に蓄積されてあるデータを順次読み書きする方法である。そのため、ファイルの途中でデータを追加したり、削除したりするには、ファイル全体を書きかえる操作が生ずる事であり、ファイルのメンテナン스에不便をきたす。反面、記憶面の効率の利用、およびファイル操作のアルゴリズムが、簡単である。

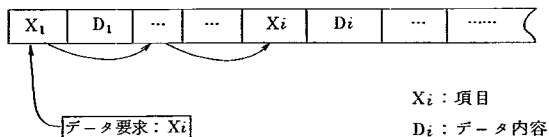


図5・5 順次アクセス

b) 直接アクセス(図5・6)

記憶番地を何らかの計算値で与え、データを直接読み書きする方法であり、必要なデータを迅速にアクセスする事ができる。また、ファイルの更新は容易であるが、番地計算の処理および記憶場所のかけはなれなどがあり、パフォーマンスは低下し、ま

た、アドレス関数 F とキー値との 1 : 1 の対応が簡単に何時もみいだせるとは限らず、乱数化や、それにともなう修正が必要になってくる。

$$\text{Address}(X) = F(\text{id}(x))$$

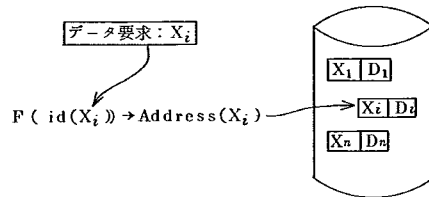


図 5・6 直接アクセス

c) 索引付直接アクセス

索引テーブルを設けて、そのテーブルを参照する事によって、要求されたデータの格納位置にたどりつく方法である。索引テーブルは、索引キーとそのキーのデータが格納されてある記憶番地へのポインタからなる。この方法は、直接アクセス方式において、一定のアドレス関数が得られないような場合に利用される。なお、索引テーブルが大きくなると、索引テーブルの中から要求されたキー値を見つけだすのに時間がかかるが、時間短縮の方法として 2 分検索法、分散記憶法などがある。

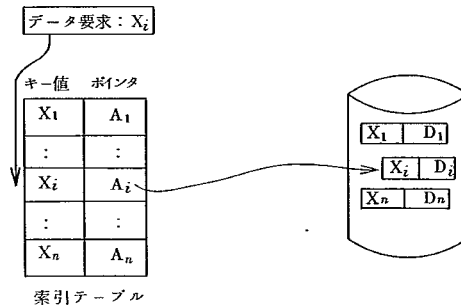


図 5・7 索引付直接アクセス

以上のような基本的なアクセス方法の組合せて種々のアクセス方法の設計が可能である（詳細は付録 1. 参照の事）。

380

(5) 接続関係の表現

分類されたデータ構造による構造検索を可能とし、迅速なデータ検索をするためには、構造の接続関係を何らかの手法で表現してやらなければならない。ここで、分類構造、つまりデータ構造は、階層木構造であり、関係の基本は、親子、兄弟の関係である(図5・7参照)。

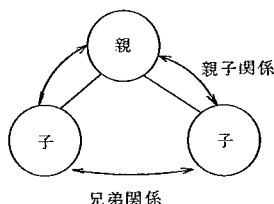


図5・7 親子、兄弟関係

図5・7の関係をポインタによって表わす方法がある。ポインタは、ある項目に接続関係を持つ項目に対して接続情報を属性として持たせるものである(図5・8参照)。

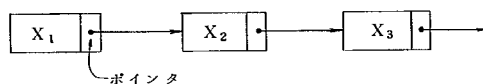


図5・8 一方向接続

図5・8でわかるとうり、ポインタは、方向性を持っている。つまり、 X_1 から X_2 を得る事ができても、その逆である X_2 から X_1 は直接の関係としては求められない。これを可能としたのが図5・9のような両方向接続である。

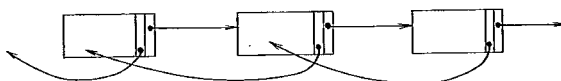


図5・9 両方向接続

この二つの接続手法で、図5・7の親子関係を表現する場合、親によって、子の数は、任意である。したがって、親の持つポインタの数は可変個となってしまいます。この問題に対処する方法として、リング構造による表現がある。これは、図5・10のように親は、最初の子へ、子は兄弟へ、最後の子は親への一方向ポインタを持つ方法であり接続関係

関係は輪状になる。また階層構造では、ある項目は親であり、子である事があり、一つの項目について二つのポインタが必要となる。

図5・10でAを与えてB、Eを検索する場合は、A-B-D-Eの経路をたどり、Oを与えてBを検索する場合は、O-A-Bの経路をたどってデータが検索される。

リング構造において、データの更新は、関係するポインタのみを変更すればよいわけであり、また各項目の記憶場所も任意であることから、データ・ベースのメンテナンスは簡単である。本システムでは、接続関係の表現は、リング構造を採用している。

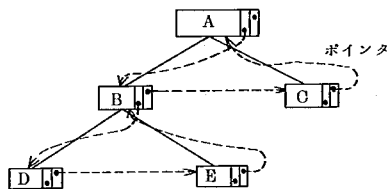


図5・10 リング構造

5-4 データの構成と内部表現

情報の最小単位は、三つの情報インデックスとデータ操作に必要な情報およびデータ値で構成される(図5・11参照)。

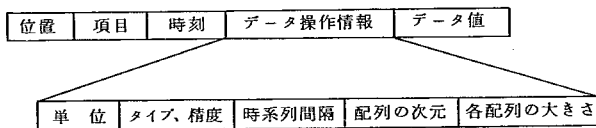


図5・11 データの構成

図5・11の各要素はそれぞれ次のような形式で、内部表現される。

a) 位置

位置には、位置の名称およびコードが付けられているが、位置の名称は、位置コード ↔ 名称のテーブルを設ける事によって得られる。

b) 項目

項目コードで表現される。

c) 時刻

382

時刻は、時系列の最初の時刻を記録しておく。時刻は、西暦年月日、時分秒で表わす。
 ここで、年については、現実的に対象とするデータは、1900年以降のものとしてさ
 しつかえないと思われるので、実際の西暦年より1900を引いた値で記録する。

Y Y Y M M D D h h m m s s
 年 月 日 時 分 秒

d) 単 位

単位の普遍化は、困難である。ここでは、JISで規定されてある「情報交換用単位
 記号表」を基にして作成した、表5・1のようなテーブルのコードを持つ。

コード	記 号	変換係数
長さ	1	M
	2	MICRN
	3	ANG
	4	NAM
	5	

表5・1 単位コード・テーブル

e) タイプ、精度

タイプは、データ値が、整数であるか、実数であるかを指示する。また、精度は、デ
 ータ値の有効桁数を表わす。

F 0 7 実数で有効桁数は7桁である。

I 1 0 整数で最大10桁である。

f) 時系列間隔

時系列のサイクルを次のように記録する。

時刻 単位	サイクル
----------	------

g) 時刻の単位は、1文字の記号で表現する。

年	月	日	時	分	秒
Y	M	D	H	M	S

h) 配列の次元および配列の大きさ

データ値は、時系列の他に配列で与える事ができる。例えば、表5・2のような風向

・風速頻度表を蓄積する場合は、配列の次元数は3、配列の大きさは1、6、4と与えられる。ただし、ここでは時系列も配列に含める。したがって、一値のデータの時は、配列の次元数は1、配列の大きさは時系列の数となる。

	0	2	4	6	8	10	m/sec
N	1	1	2	1	0	0	
E	0	3	0	4	2	1	
S	2	0	1	5	1	0	
W	1	1	1	1	0	2	

表5・2 風向・風速頻度表

i) データ値

データの個数は、配列の大きさで計算され、指定された精度およびタイプによってデータが格納される。

5-5 ファイル設計

(i) 三つのデータ・ベース

データ・ベースは、蓄積されるデータの特徴により、ゾーン、メッシュ、不定形のデータ・ベースとしてとらえる事ができる。この三つのデータ・ベースは、論理的にも、物理的にも分離して考える事ができる。

データ・ベースの構築は、この三つのデータ・ベースの任意の組合せが自由にできる。

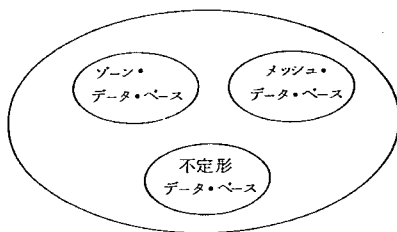


図5・12 三つのデータ・ベース

以下においては、三つのそれぞれのデータ・ベースについてファイル設計を行う。

(2) ゾーン・データ・ベース

ゾーン・データのデータ構造は、図5・13のとおりである。

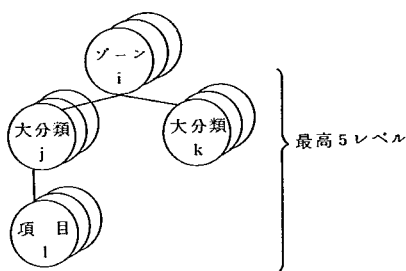


図 5・13 ゾーン・データ構造

ゾーン・データへのデータ要求パターンは、大別すると次の三つのパターンとして、とらえることができる。

(i) 単一データの要求

例えば、東京都の人口は？

(ii) 複数ゾーンの同一項目データの要求

例えば、東京都の 23 区のそれぞれの人口は？

(iii) 同一ゾーンの複数項目データの要求

例えば、東京都の人口、面積、工業出荷額は？

このような要求に対して、迅速なデータ検索を行うため、次のようなデータ検索の方法を準備する。

a) 索引付直接アクセス方式

索引付直接アクセス方式は、前述したように、索引テーブルを設けて、そのテーブルを参照する事によって、要求されたデータの格納位置に直接アクセスする方法である。索引テーブルに、図 5・14 のように、縦方向には、項目コードを、横方向には、ゾーン・コードを示し、要求された項目コード、ゾーン・コードにより、条件に一致した、行と列で示される部分のポインタが検索の開始アドレスである。なお、索引テーブルは、最高 3 レベルまで可能とする。例えば、3 レベルとした場合の各レベルの概念は、大分類、中分類、項目の索引テーブルとする事ができる。また、ゾーンに關しても都道府県、市区町村、およびブロックの三つに分割する事ができる。しかし、データの集積が少ない場合は、一つにまとめてもかまわない。以上のような、索引テーブルの種類は、システムごとに任意に決定されるものである。

b) 構造検索

構造検索は、分類構造にそった検索を行うものである。一つの木構造は、先にのべた、親子、兄弟の関係づけを輪状で示すリング構造で可能である。しかし、ここで、各データ項目は、位置に関して付帯している。このような場合、複数ゾーンの同一項目のデータが要求されると、かなりの道をたどる事になり、検索のパフォーマンスは低下するが、このためには、他ゾーンの同一項目へのポインタを各項目で持つ事にすればよい。このポインタで結ばれる関係を義兄弟と名付けると、各データは、親子、兄弟、義兄弟の関係を結ぶ三つのポインタを持つ事になる（図5・15参照）。

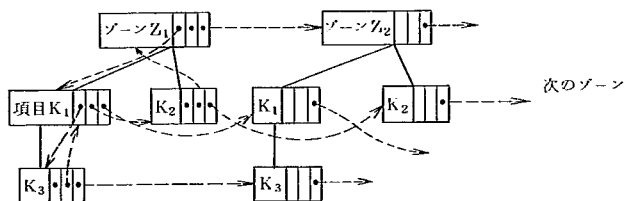


図5・15 三つのポインタ

c) 索引付直接アクセス方式と構造検索の組合せ

索引付直接アクセス方式と構造検索を組合せて、効率的な検索が可能である。例えば、ある項目は、ダイレクトに検索し、以降のデータは、構造検索を行うような組合せ利用が考えられる（図5・16参照）。

この方式は、一般にマルチ・リスト・ファイルと呼ばれている。

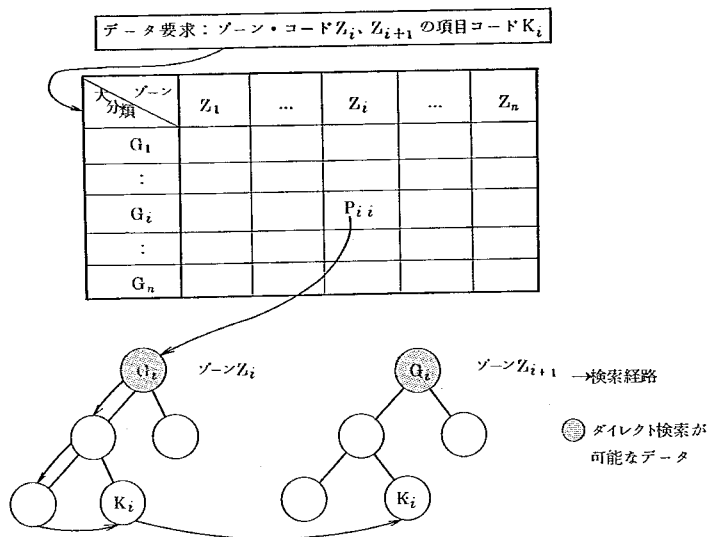


図5・16 マルチ・リスト・ファイル

d) ゾーン・データのデータ構成

ゾーン・データは、三つのポイントを持つ事を提案したが、このポイントは、図5・17のようにセットされる。

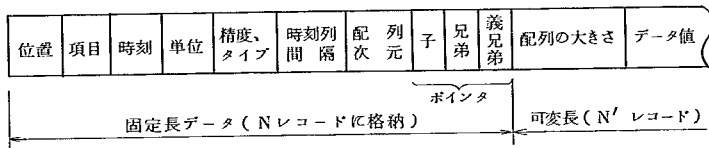


図5・17 ゾーン・データの構成

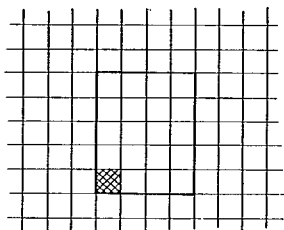
図5・17において、レコードの長さの検討は十分になされなければならないが、ここでは、詳細は、はぶく。また、レコードは、連続した記憶場所になくともよいとすると、各レコードには、レコード間の連結ポイントが必要となる。

(3) メッシュ・データ・ベース

メッシュの特性は、新しいメッシュの発生や、削除がない事であり、メッシュはその

コードとともに普遍的なものとしてとらえる事ができる。また、各メッシュはほぼ一様の情報量を持ち、データ項目数も十数種、多くて数十種と想定できる。

メッシュ・データは、国土をあるまとまった矩形地域で収集され、この矩形地域は、南西端のメッシュ・コードと縦横のメッシュ数で定義できる。



太枠のメッシュは  のメッシュ・コードと横のメッシュ数4、縦のメッシュ数5を与える事によって定義される。

図5・18 メッシュのブロック分解

図5・18のようなブロック内のメッシュ内のメッシュ・データは、順編成ファイルで管理されるのが妥当と思われる。メッシュ、項目に関する蓄積順序は図5・19、図5・20の通りである。

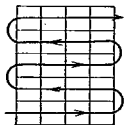


図5・19 メッシュの順序付け

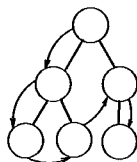


図5・20 項目の順序付け

ここで、メッシュ・コードと項目コードを与えて、データの格納番地を簡単に得るためには、図5・21のようにデータを固定長とし、固定長よりあふれたデータをあふれ領域の中に格納すればよい。また、データ操作情報は、同一項目について、全メッシュで共通であるとしてさしつかえないだろう。したがって、データ操作情報は、テーブルとして管理する。

以上がメッシュ・データファイルの基本部である。

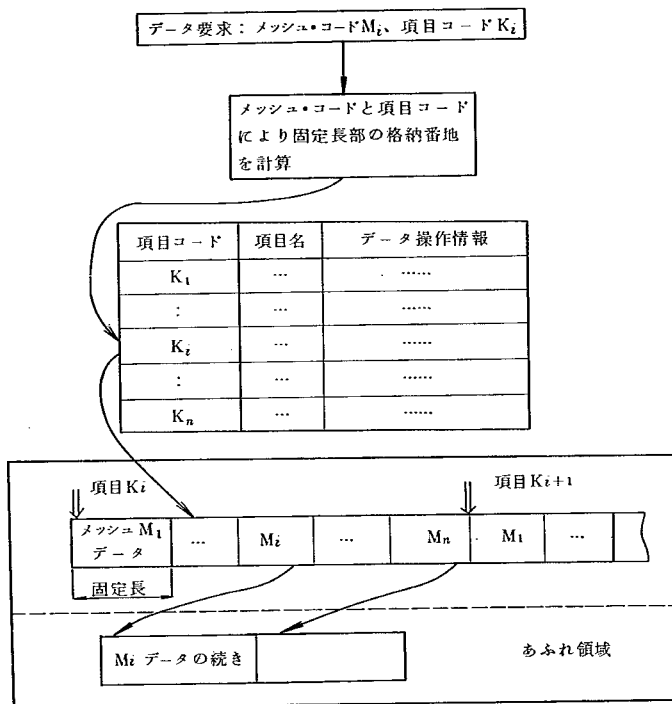


図5・21 メッシュ・データの検索の流れ

(4) 不定形データ・ベース

不定形データは、ある特定の点、線、面の属性データである。

不定形データの特徴は、位置と項目の関係が、固定的な事である。例えば、流量、水位は流量観測点への属性であり、流量観測点は流量、水位の属性のみを持っている。このように位置と項目は、1：1の対応関係があり、位置は、一般に事象で特徴づけられる。したがって不定形データは図5・22のようなデータ構造として、とらえることができ、ファイルの構造はゾーン・データと同様である。

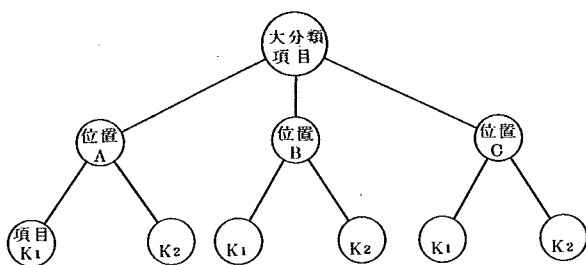


図 5・22 不定形データ構造

5-6 データ・ベースとの対話

データ・ベースの利用者は、プログラマと一般利用者と分けて考える事ができる。プログラマは、プログラムの記述の中で、データ・ベースのデータを要求し、回答されたデータをそのまま処理に用いる。また、一般利用者は、プログラムとは関係なくデータを要求し、回答は、レポートまたは 磁気テープなどで要求する。このように、データ・ベースとの対話の形態は 2 種類のものとしてとらえる事ができる (図 5・23 参照)。

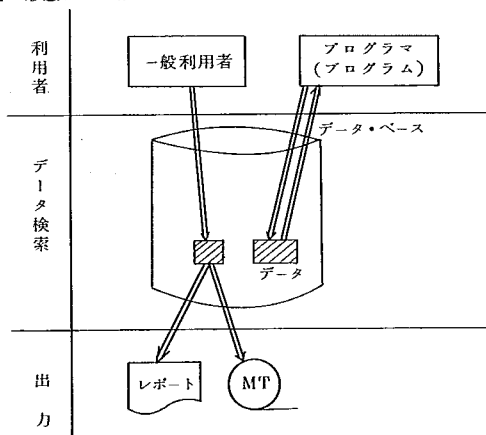


図 5・23 利用形態

ここで、2 種類の利用者集団に対してそれぞれに適した対話の方法を与えてやらなければならない。

本システムでは、プログラマに対しては、親言語ルーチン方式、一般利用者に対しては

コマンド・インタプリタ方式を準備する。

(1) 親言語ルーチン方式

親言語ルーチン方式は、例えば、処理プログラムがフォートラン言語で記述されていたとすると、データ検索機能は、サブルーチンとして与えられる。プログラマは、サブルーチンの検索パラメータを指定し、そのサブルーチンをコールする事によって、データ・ベースに蓄積されてあるデータを利用できる。

例えば、東京都(ゾーン・コード→13)の1970年の年間降雨量(項目コード→1001)の値をDATAという変数にほしい場合は既略次のようになるだろう。

GALL GET(1970, 13, 1001, ... , DATA)

(2) コマンド・インタプリタ方式

コマンド・インタプリタ(C/I)は、一般に問題向言語と呼ばれる言語を逐次解釈し、実行するためのプログラムである。

問題向言語とは、技術者が問題を解決する用語に近いことばで、問題の処理手続きを記述するための言語である。すなわち、コンピュータの知識がとほしい人でも、何のデータがどのような形式でほしいかさえわかっていたら、データ・ベースとの対話が可能である。

例として示すと、

```
OPEN DB(MESH); .....メッシュ・データ・ベースの開始命令
SET INDEX; .....情報インデックスのセット命令
TIME=1960-1970; .....1960~1970年の時系列データ
MESH=5439; .....メッシュ・コード5439
ITEM=1001; .....項目コード1001
GET(MT(1)); .....データを検索したのち機番1の微気テープ
                        にデータを出力する
END; .....終了
```

コマンド・インタプリタの処理の流れは図5・24のようになる。

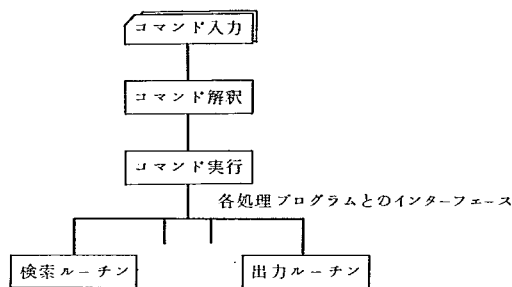


図 5・24 C/I の流れ

(3) 検索の方式

前述した二つの方式でデータ・ベースと利用者のインターフェースを図る事ができ、それらによって次のような検索要求が可能である。

a) 事項検索

事項検索は、三つの情報インデックスを与えてデータを検索する方法で、次のような検索要求が可能である。

(i) 単一データ検索

位置、項目、時刻を与えてデータを得る。

例えば、東京都の1970年の人口は？

(ii) 複数データ検索

複数位置、または複数項目、時刻を与えて複数データを得る。

例えば、東京都、大阪府の1970年の人口は？

東京都の1970年の人口、工業出荷額は？

(iii) 時系列データ検索

位置、項目、時系列期間を与えて時系列データを得る。

例えば、東京都の1970～1975年の人口は？

b) 条件検索

条件検索は、データ値の条件を与え、それを満たしたデータ値と情報インデックスを回答とする検索であり、回答は複数個となる場合がある。

条件は、次の4種の関係演算子と3種の論理演算子の組合せで与えられる。

関係演算子

論理演算子

=

∧

≠

∨

<

┐

>

例えば、

(i) 1970年で人口200万人以上の都道府県は？

$\{1970, *, 人口\} > 200 \text{ 万人}$

回答は、条件を満たした人口200万人以上の都道府県およびそれぞれの人口。

(ii) 1970年で人口3万人～5万人でかつ工業出荷額100億円以上の市区町村は？

$3 \text{ 万人} < \{1970, *, 人口\} < 5 \text{ 万人} \wedge \{1970, *, 工業出荷額\} > 100 \text{ 億円}$

回答は、条件を満たした市区町村とそれぞれの人口、工業出荷額。

第 6 章 結 語

本研究の成果と今後の課題は次のとおりである。

- (i) データ・バンクに関する従来の研究およびデータ・バンク・システムの事例について調査し、データ・バンクの現状を明らかにした。その結果、何らかの目的に対してデータ・バンクを整備しようとするれば、経済的な問題さえ解決できればそれが可能であること、データ・バンクは何らかの目的意識があってはじめて作りあげられるものであり、将来の発展を考慮してシステムを構成すべきであることなどが明確になった。
- (ii) データ・バンクを作りあげる（再構成する）ためには、情報自体が具備すべき要件（情報インデックス）およびシステムの構成について検討する必要がある。前者については、位置・時刻・項目のインデックスの普遍的な表示法を示し、これをもとにした検索方法を提示した。我々の対象とする情報は経理・事務等の情報と異り属地性のものが多いために、位置インデックスの表示方法について特に検討を加えた。
- (iii) 建設省にふさわしいデータ・バンクとして、情報の属地性、経済的な問題、現状のシステムの有効利用、将来の発展・整備等を考慮して分散蓄積型データ・バンク・システムを提案した。また、このような個別に作られるデータ・バンクの間で情報を相互利用するために必要となる情報のつき合せ（同定）方法として、普遍系と局所系の変換システムを用いる方法を提示した。
- (iv) 流水の管理を前提とした流域データ・バンク・システムの設計概念について検討し、電子計算機を活用したデータ・バンクを作りあげる場合の設計概念を提示した。すなわち、データ・ベースのあり方（蓄積・検索のための論理）について検討し、検索を容易にするためのデータの蓄積方法およびポイントを利用した検索方法等を提示した。
- (v) 本研究では取り扱うデータの範囲や量について具体的な検討は行っていないが、流域データ・²⁷⁾バンクの検討においては流水管理のための水配分計画で利用するデータを想定して考察を行った。電子計算機を活用したデータ・バンクには利用価値の高い情報から順次くみ込んでいけばよいと考えられるが、その範囲や量については今後の課題として残されている。
- (vi) 電子計算機等の情報機器を活用したデータ・バンクを作る場合の費用や労力は、現在すでにある程度実用化されている事例から推定することができるが、それに対する便益や効果等の面からの評価は今後検討を要する問題である。

付 録

目 次

1. 既存のデータ・ベース・マネージメント・システム	45
1-1 概 要	45
1-2 CODASYL 方式のデータ・ベース	46
1-3 RAPIDデータ・ベース・マネージメント・システム	50
1-4 IMS データ・ベース・マネージメント・システム	58
2. データ・バンク・システムの事例.....	64
2-1 東京都における地域環境データ・バンク実験システム	64
2-2 大阪府におけるメッシュ・データ・バンク・システム	69
2-3 北上川水系統合管理システム	74
2-4 広域水系統合管理システム	76
2-5 省庁におけるデータ・バンクの事例.....	80
3. 地理情報の表現	83

1. 既存のデータ・ベース・マネージメント・システム

1-1 概 要

データ・ベース・ソフトウェアは、データの蓄積、検索を効率的に行うためのアルゴリズムを記述したものである。このソフトウェアは、基本的にはデータをどのようにモデル化するかによって決定される。データ・モデルは、現在発表されているもので20数種類があり、それらの一般的なものは、メーカ、ソフトウェア・ハウス等で汎用データ・ベースとして開発され、実用化されている。

データ・モデルは、情報間で存在する論理関係をいかに表現するかであり、データ構造といわれるものである。データ構造は、一般に線形構造、木構造およびネットワーク構造の3種の構造としてとらえられる(図1.1参照)。

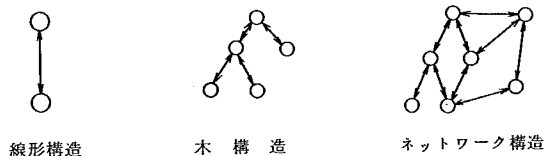


図 1. 1 データ構造

情報間の関係は線形構造においては1:1の対応であり、木構造においては1:nその逆は1:1の対応を示し、ネットワーク構造においては1:n、その逆は1:mの複数の関係を持ち、より複雑なデータ構造の表現が可能である。またこれらの3種の構造には次の関係がある。

線形構造 $\subset$ 木構造 $\subset$ ネットワーク構造

ソフトウェアは、大別すると、上記のようなデータ構造によってデータを蓄積、検索する機能と、データ・ベースと利用者のインタフェースとしての言語機能があり、この言語は、親言語(COBOL, PL/I)の中で記述できるようになっている。また、汎用データ・ベースは、一般的にOSをも含んだ形で設計されてあるので、OSの異なるシステム間の互換性は持ち合せていない。

21)

1-2 CODASYL 方式のデータ・ベース

(1) 概 要

COBOL言語などの開発を行なったCODASYL (The Conference on Data System Language) の作業グループであるDBTG (Data Base Task Group) によって提案されたのが、このCODASYL 方式のデータ・ベースの概念である。この提案は、データ・ベース用言語を通じ、データ・ベースの共通のイメージと基本的にデータ・ベースが具備すべき機能を具体的に示してくれた。この方式は、データ構造をネットワーク構造としてとられている代表的なもので、国内でもこの提案を取り入れた汎用データ・ベース・マネージメント・システムが多数実用化されている。RAPID (富士通)、DBMS (日電)、IDS-II (東芝) などがその例である。

CODASYL 方式では、データ・ベースを含めてデータ・ベース全体の管理を司どる、データ管理者 (DA: Data Administrator) と実際にデータ・ベースを利用する利用者集団の二つをその運用体系の環境と考えている。よって、言語体系も二つに分離して考える事ができる。前者に当る言語としてはデータ・ベースを作成するために使用される固有データ記述言語 (DDL: Data Description Language for the Schema) があり、後者は、利用者プログラムとデータ・ベースとのデータの転送を図るためのデータ操作言語 (DML: Data Manipulation Language) がある。DDL, DML はデータ管理者、利用者が親言語 (処理プログラムの記述言語: COBOL, FORTRAN) の中で記述することができる。DDL は共通なデータ・ベースを作成するものであるが、各利用者が、個別にデータ構造を記述する事ができる個別データ構造記述言語 (Sub Schema) が準備されている。これによって、利用者はあたかも自分のみのファイルであるという概念のもとにデータが利用できる。

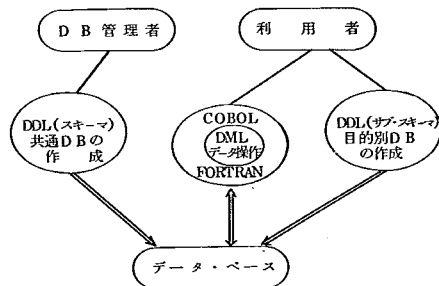
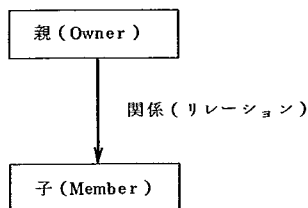


図 1.2 CODASYL 方式の言語体系

(2) データ構造

データ構造は、階層構造を成し、データ構造の基本要素は次の親子集合の概念である。



この基本構造を基に親子関係は木構造が可能となり図 1・3 のような 1 対 n の親子関係が成り立つ。また複数の木構造の組み合わせとしてネットワークに近い構造を表現する事ができる。しかし、子は、複数の親集合には参加する事は禁止されているので親と子との間で図 1・4 のような n 対 m の対応関係を持つ事はできないので、厳密なるネット・ワーク構造とはいえない。

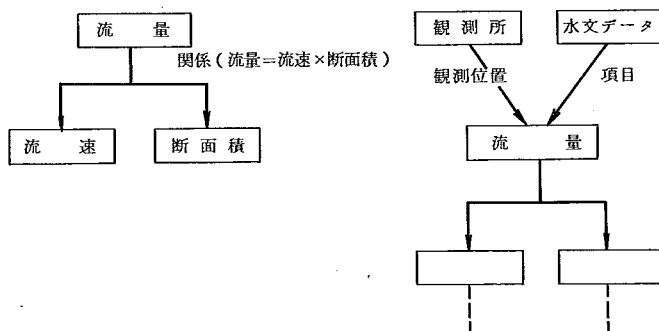


図 1・3 許されるデータ構造

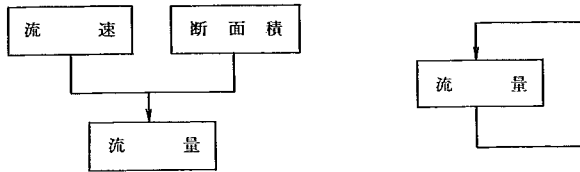


図 1 . 4 許されないデータ構造

(3) 言語機能

DDL, DML のデータ・ベース言語は、言語そのものとしての FORTRAN, ALGOL 等の汎用言語や GPSS, APT など問題向き言語とは同列の言語として考える事はできない。あくまでも、データ・ベース言語は文法的な諸取り決めの問題からのみ考察するのは、順序として正しくない。データ・ベース言語の問題は、ユーザとデータ・ベースとがどの段階でいかにインターフェースを持つかという問題の一部として存在するものである。

CODASYL 方式の言語構造と機能の概要は次のとおりである。

a) DDL (スキーマ) ; データ記述言語

DDL はデータ・ベース全体を記述するために使用する言語であり、データを操作する命令などはいっさい含まない。

データ・ベースの構造は、次のような階層構造を持っている。

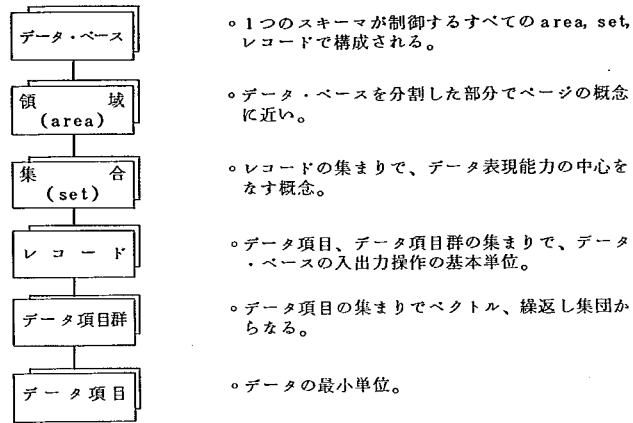


図 1 . 5 データ構造

b) DML ; データ操作言語

DMLは作られたデータ・ベースに対してユーザがアクセスするために使用するもので、親言語 (FORTRANについては開発中) の中で使用出来る。

これは制御用、検索用、更新用の3種の言語体系から成り次の表のような命令句が準備されている。

	1971年DBTG提案	COBOL データ・ベース機能
制 御 用	OPEN CLOSE IF	READY FINISH IF
検 索 用	FIND GET KEEP FREE MOVE	FIND GET KEEP REMONITOR FREE ACCEPT
更 新 用	STORE MODIFY DELETE ORDER INSERT REMOVE	STORE MODIFY ERASE ORDER CONNECT DISCONNECT
特 殊		USE

表 1 . 1 データ操作命令 (DMLの命令)

c) DDL (サブスキーマ)

サブスキーマは、スキーマの中から特定のプログラムが使用する部分だけを定義するものであたかもデータ・ベースが利用者独自のものであるかのように思わせる論理構造定義言語である。

言語の内容は、ユーザが利用する領域、レコード、集合等の定義が出来るようになっている。

1-3 RAPIDデータ・ベース・マネージメント・システム²²⁾

(1) 概 要

RAPID(Retrieval And Production for Integrated Data)は、富士通によって開発された汎用データ・ファイル・マネージメント・システムであり、データ・ベースを次の3種類の概念に分けて、データ・ベースの問題を議論している。

a) フィジカル・データ・ベース

データ・ベースに含まれるすべての情報が、外部記憶装置上に、どのように格納されているかといった、物理的構造の面からとらえたデータ・ベースである。

b) ロジカル・データ・ベース

データ・ベースに含まれる情報を、論理的な構造からとらえたデータ・ベースである。フィジカル・データ・ベースが主にデータ・ベース管理者から見たデータ・ベースのとらえ方であるのに対して、ロジカル・データ・ベースは、実際にデータ・ベースを利用する人からとらえたデータ・ベースである。

c) サブ・データ・ベース

データ・ベースを利用する側から見れば、一つのプログラムで全部のデータ・ベースを必要とする事はなく、ある部分のデータ・ベースを必要とする。このような、目的別部分データ・ベースがサブ・データ・ベースである。

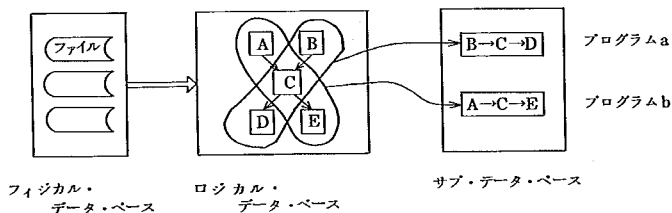


図1.6 フィジカル、ロジカル、サブ・データ・ベースの関連

(2) ロジカル・データ・ベース

ロジカル・データ・ベースは図1.7のような論理的体系となっている。

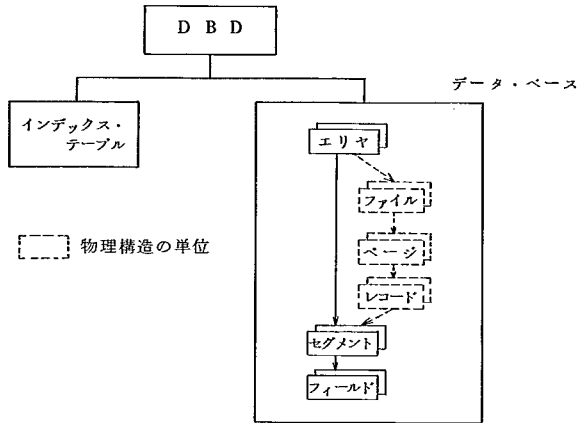


図1.7 ロジカル・データ・ベース

a) データ・ベース

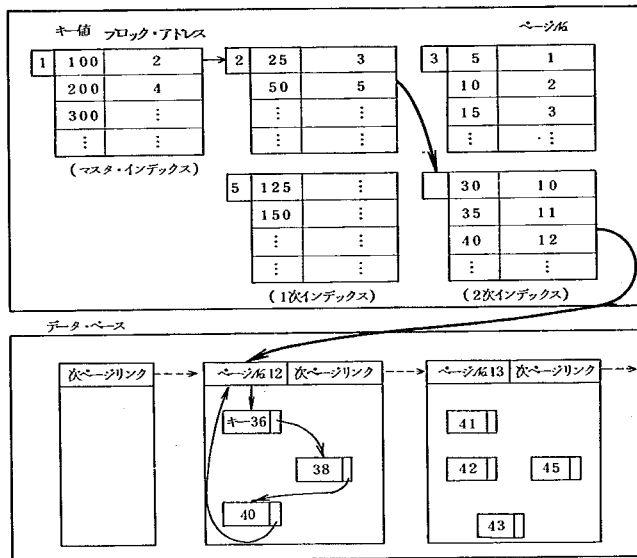
互に関連性を持つデータの集合体であり、一つまたは、複数のエリアで構成される。

b) D B D (Data Base Description)

データ・ベースの論理的構成、物理的構成などに関する定義情報を持つテーブルで、データ・ベースを総括的に管理する。

c) インデックス・テーブル

インデックス・テーブルは、要求されたセグメントをダイレクトに検索するために用いられるディレクトリ・テーブルであり、図1.8のように階層的に配置されたテーブルからキー値を与える事によりそのキー値を持つセグメントが検索される。



キー値 40 のセグメント検索の場合

図 1. 8 インデックス・テーブル

d) エリヤ

エリヤは、ロジカル・データ・ベースを理論的に、またフィジカル・データ・ベースを物理的に分割する単位であり、互に関連を持ったセグメントの集合で構成される。

e) セグメント

セグメントは、データ間の関連づけを行なう単位であり、ユーザがデータを処理する場合の基本単位である。セグメントは、一つ、または複数のフィールドで構成される。

f) フィールド

フィールドは、セグメントを構成する論理的に意味を持つ最小データ単位である。

(3) フィジカル・データ・ベース

フィジカル・データ・ベースは、データの物理的な格納形態から見たデータ・ベースであり、物理的な記憶単位を図 1. 9 のように規定している。フィジカル・データ・ベ

スは、ロジカル・データ・ベースに対応してまったく同様の構成になっている。

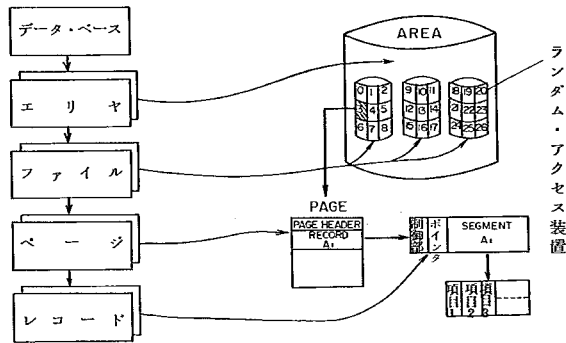


図 1.9 フィジカル・データ・ベース

(4) サブ・データ・ベース

サブ・データ・ベースは、ロジカル・データ・ベースとの独立性を確立する役割を持つ。すなわち、業務目的別にデータ構造が定義できるため、ユーザズ・プログラムは、複雑なロジカル・データ・ベースを意識する事なくデータ操作が可能である。

(5) データ構造

データ構造は、データに上下関係を持つ階層構造である。

データの関連は、セグメント単位でつけられ、階層関係の基本形態は、図 1.10 のように定義されている。

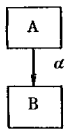
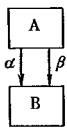
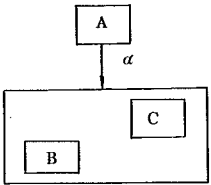
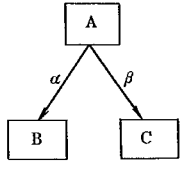
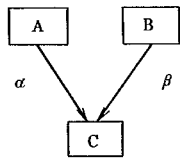
基本形態Ⅰ 	最も単純な階層関係であり、セグメント間に一つの階層関係が定義されるもの。
基本形態Ⅱ 	セグメント間に性質の異なる複数の階層関係が定義されるもの。
基本形態Ⅲ 	ディテール・セグメントとして異なる複数のセグメントがあり、それらが、一つの階層関係で定義されるもの。
基本形態Ⅳ 	ディテール・セグメントとして異なる複数のセグメントがあり、それらが、それぞれ別の階層関係で定義されるもの。
基本形態Ⅴ 	マスタ・セグメントとして異なる複数のセグメントがあり、それらが、それぞれ別の階層関係で定義されるもの。

図 1 . 1 0 階層構造の基本形態

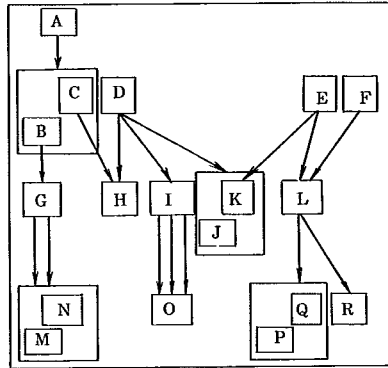


図 1 . 1 1 階層関係例

階層構造は、図 1 . 1 0 の基本形態の組み合わせで、複雑なネット・ワーク構造も表現できる。また、この階層は、ポインタによるリング構造（図 1 . 1 2 参照）で表現される。

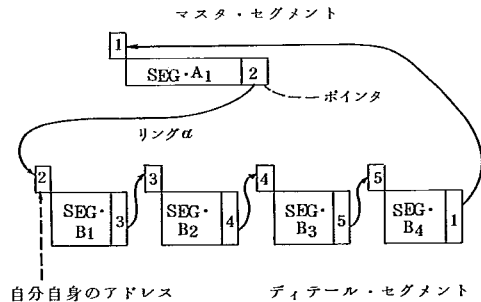


図 1 . 1 2 リング構造

(6) データの検索方式

検索方式は、表 1 . 2 のような検索モードが準備されている。

検索の単位	検索の形態	説 明	備 考
セグメント 単位の検索	データ・ベ ス・エントリ	データ・ベース中のセグメ ントをランダムイズ・アル ゴリズム、インデクス・テ ーブルなどによって直接 検索する。	これらの形態 を自由に組み 合せて、デー タ・ベースから の検索を行う ことができる。
	ロジカル構 造による検 索	リングによるデータ構造 にそった検索、データベ ースの関連データの検索 はこれによって行われる。	
	フィジカル 構造による 検索	フィジカル構造にそつ た検索、格納されたセグ メントの物理的シーケン スにしたがって検索する。	
ロジカル・レ コード単位 の検索	条件検索	RAPIDがデータ・ベー ス・エントリとロジカル構造 による検索を組合せて、 ユーザの指定した条件に あうセグメント群を検索 する。	

表 1 . 2 検索モード

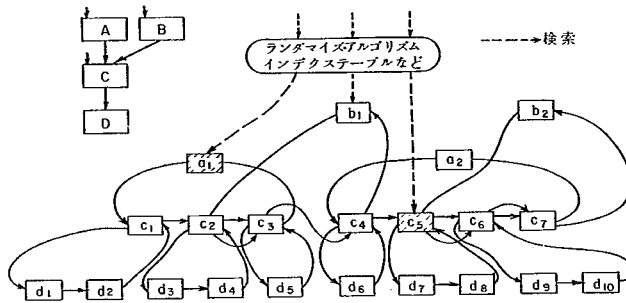


図 1.1.3 データ・ベース・エントリによる検索

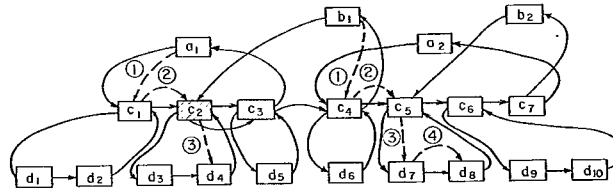


図 1.1.4 論理構造による検索

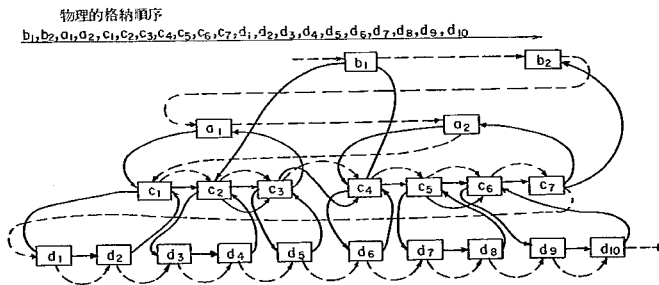


図 1.1.5 物理構造による検索

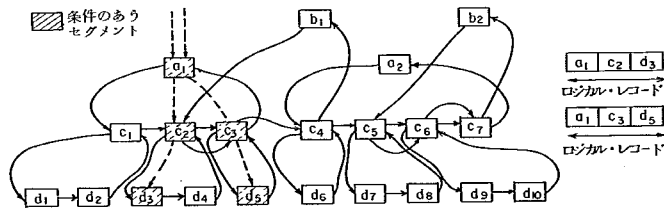


図 1.1.6 条件検索例

1-4 IMSデータ・ベース・マネージメント・システム^{23) 24) 25)}

(1) 概 要

データ・ベースのデータ構造として木構造を前提として、階層を成しているデータ・ベースを一般に、階層構造データ・ベースという。この代表的なものとして、IBM社の汎用データ・ベース・マネージメント・システムIMS(Information Management System)を例にとって説明を加える。

IMSは1960年代のはじめに、米国の航空会社が多年の経験を生かし、アポロ計画の宇宙船の製造部品の生産管理のために開発されたもので、200万点におよぶ部品の管理がこれによって行われた。以後、IBM社がこのシステムを引き受け、改良が積みかさなれ、今日のIMS/Vが商用化された。

(2) データ構造

先にのべたようにデータ構造は木構造による階層構造となっている。

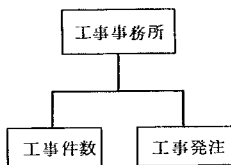


図1.17 階層構造

図1.17において、長方形はセグメント・タイプを表わす。セグメントは、データ間の関係がなされる単位で、論理的に関係のあるデータ・フィールドを単数または複数含んでいる。例えば図1.17において工事発注のセグメントは、請負者名、工事名、発注額などを含むようにセグメントを設計しても良いわけである。

図1.18には四つのセグメントがある。

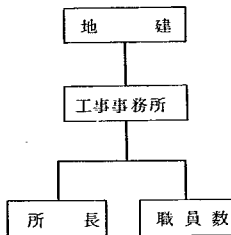


図1.18 論理データ構造

ここで、工事事務所セグメントは、根セグメント地建に依存している。根セグメントはもっとも高いレベルのセグメントであり、従属セグメントは、その内容が上のレベルのセグメントに依存しているセグメントである。

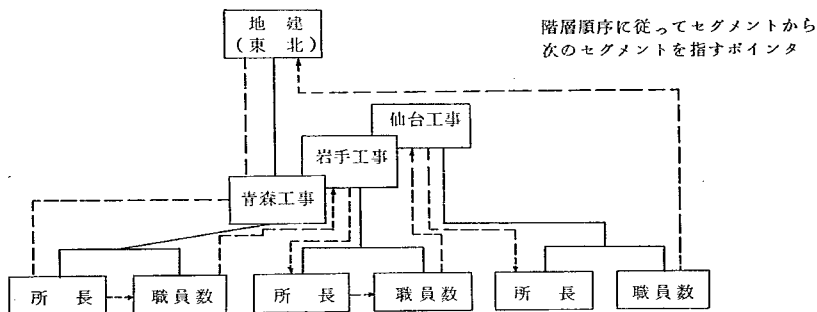


図 1.19 論理データ構造 (詳細)

図 1.18 を詳細に書くと図 1.19 のようになり、この例だと東北地建は、青森、岩手および仙台の三つの工事事務所で構成されている。また、セグメントのタイプは地建、工事事務所、所長および職員数の 4 種であるが、工事事務所、所長、職員数は、複数回出てくるので 10 個のセグメントからなっている。これにおいて、着目したセグメントのすぐ上のセグメントを親セグメント、すぐ下のセグメントを子セグメントと呼んでいる。また、同レベルの階層のセグメント（ここでは青森、岩手、仙台）を兄弟セグメントと呼んでいる。

基本的なデータ構造は、前述したとおりであるが、図 1.17 と図 1.18 において、工事事務所セグメントが重複して存在しているような場合は一つのセグメントを複数の論理データ構造に結びつけられる。この方法は、ロジカル・リレーションと呼ばれ図 1.20 のようにセグメントを指すポインタでおきかえる事ができ、IMS のデータ構造は、木構造 + ロジカル・リレーションという事ができる。

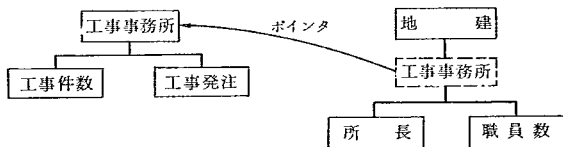


図 1.20 ロジカル・リレーション

(3) データ・ベースの多重入口（副次索引）

図 1.21 において、工事事務所をキーとして、工事発注（例えば工事名）のデータを得る事は容易であるが、工事名を与えて工事事務所を得る事は容易でない。このような場合、工事名に対して、副次的な索引をつける事によって、索引の対象となるセグメントからアクセスが開始されるので、迅速にデータを取り出す事ができる。この方法を副次索引と呼び、主な内容は、索引目的のセグメントを指すポイントと、索引項目データである。この例の場合は、索引目的のセグメントは工事事務所であり、索引項目は工事名である。

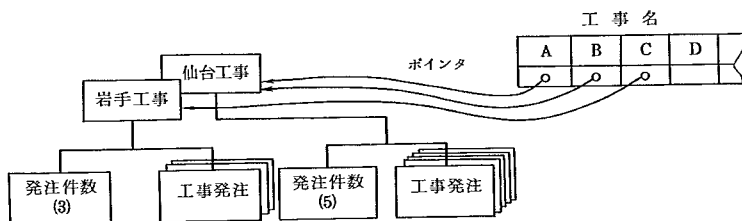


図 1.21 副次索引

比較項目	論 理 関 係	副 次 索 引
用 途	・複雑データ・ベースを関連づける。 ・データの重複保有を排除する。	・特定の検索目的を主眼にする。
データ構造と範囲	・1個以上のデータ・ベースを対象とする。 ・ほとんどすべてのセグメントをとり入れる。 ・論理的データ・ベースの根セグメントは物理データ・ベースの根セグメントでなければならない。 ・1組の論理関係で論理親から論理子へ、またその逆へと2方向のパスにより、大別して2種類の論理データ構造が作れる。	・1個のデータ・ベース内に限定。 ・索引目的セグメントの従属セグメントと根セグメントに至る階層パス上のセグメントのみとり入れる。 ・根/従属いずれのセグメントでも、副次データ構造の根セグメントにできる。 ・1方向のみであり、1種類の副次データ構造を作る。
ア ク セ ス	・探索には、すべてのセグメントのデータ項目を利用できる。 ・論理親/子の間でのみユニークなデータを保有・利用できる。(部品所要数など)	・副次データ構造の根セグメントの探索は予め選んだ5種類のデータ項目に限定される。 ・副次索引データ・ベースのみを対象とした利用ができる。
機能利用の指定	・物理データ・ベースに論理子セグメントを設ける必要がある。 ・物理D B Dに加えて論理D B Dが必要である。	・副次索引データ・ベースを作成する必要がある。

表1・3 論理関係と副次索引の比較

(4) データ・ベースの物理構造

データ・ベースの記憶装置への格納方法として、階層順次、階層索引順次、階層直接、および階層索引直接の4種の編成モードが準備されている。この物理構造は、そのまま検索のモードとなるものである。

a) 階層順次編成 (図1・2・2)

階層順次編成は、各セグメントを階層順序通りに配列し、ポインタは使用しない。ランダム、順次アクセスの両方に使用できる。

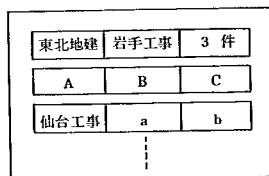


図1・2・2 階層順次編成

b) 階層索引順次編成 (図1・2・3)

階層索引順次編成は、階層順次編成の一種であるが、根セグメントに索引キーを用いて、索引による乱順処理を可能としている。

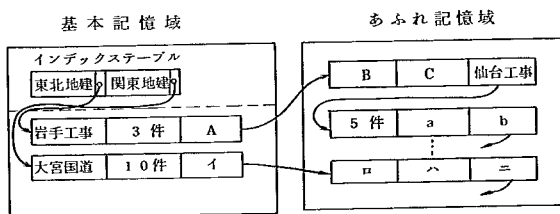


図1・2・3 階層索引順次編成

c) 階層直接編成 (図1・2・4)

階層直接編成は、乱順処理の効果を主眼とする編成法であり、セグメントは、階層順に根アドレス可能域に入り、残りはあふれ領域に入る。これらは、ポインタでつながれ、また、親子、兄弟の関係もポインタでつながれる。根セグメントのアドレス(記憶番地)は、根セグメントのキーをランダム処理する事によって決定さ

れる。

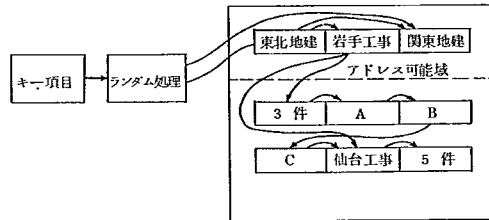


図 1.24 階層直接編成

d) 階層索引直接編成 (図 1.25)

階層索引直接編成は、階層直接編成と較べ多数の共通点を持つが、データ域は一本化し、索引テーブルを持っているのが主な相異点である。

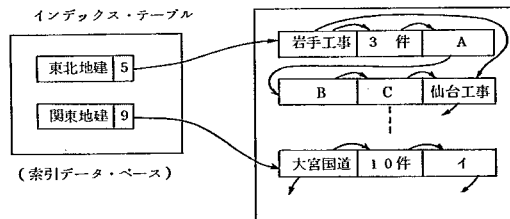


図 1.25 階層索引直接編成

2. データ・バンク・システムの事例^{9) 10) 11)}

2-1 東京都における地域環境データ・バンク実験システム

(1) 概 要

東京都は、昭和48年から都のデータ・バンク・システムの開発をめざし、組織が設置された。以来、研究開発が進められ、地環境データ・バンクのパイロット・システムが昭和50年に一応の完成をみている。

このシステムは、都行政における計画管理面へのコンピュータの適用の一環として開発されたものであり、その役割は、行政計画立案、業務遂行に必要な判断の為の資料の作成、提供にある。

本パイロット・モデルは、開発に当り既に策定された都市計画をもとに（図2・1参照）、その中の情報の流れを分析し、それを、コンピュータ・システムと対比させる事により基本的な考えの設定を行っている。

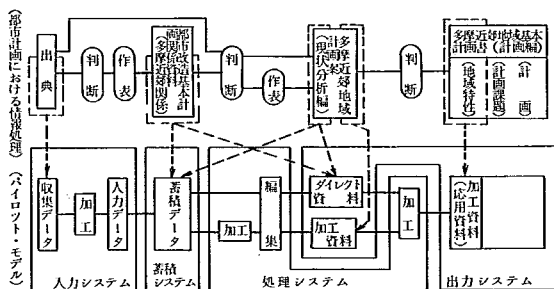


図2・1 多摩近郊地域基本計画とパイロット・モデルの対応

(2) データ構造

本システムで取り扱うデータは「行政データ」であり、あらゆる部門、分野に渡り膨大な量となり、かつ利用体系は多種多様である。このような問題に対処するため、次のような概念でデータ構造を決定している。一般的な行政データは、法律等によって、名称、内容が定まっているものや、多くの施策を理解しやすく整理、分類されているものは、共通データとしファイルを構成させている。一方、例えば、住宅局専用データとか、各計画単位のもの、目的別データとして前者と分離している。ここで、共通データは、目的別に任意に組み立てができるように、蓄積のためのデータ

構造はできる限りシンプルになっている。

データ構造は階層木構造であり、図2・2のように5段階に整理して蓄積されている。

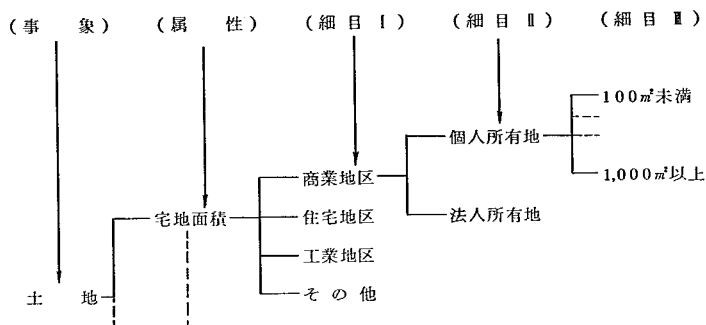


図2・2 データ構造

(3) 蓄積データ

現時点での蓄積データは、共通データとして、約5,000項目、約56万件が用意され、この他、個別指標約9,000件、都政への要望に関するデータ約1,100件、宅地利用に関するメッシュ・データ(500mメッシュ)約25,600件、および目的別データ約103,900件を含めて全部で約70万件のデータが蓄積されている。

またデータの地域集約単位は、原則として区市町村レベルである。

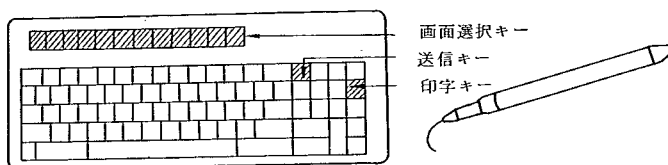
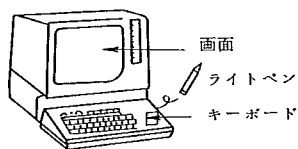
事象別共通データ一覧

		データ項目数	データ件数
人	口	常住人口(性別・年齢別等)	約 5,040件
		就業者数(職業別・産業別・地位別等)	約 559,800件
		流出入人口	約 1,020件
		増加人口	約 107,700件
		世帯(世帯数・人員)	
		将来(常住・就業・世帯)	
土地	地	行政区域面積、都市計画面積	約 280件
		課税地積(地目別・地区別)	約 122,000件
		非課税地積(区一施設別、市町村一丁目別・地区別)	
		宅地(所有形態別・規模別面積・所有者数)	
建物	物	評価家屋(構造別・種類別)	約 280件
		特別区家屋(用途別)	約 110,000件
		着工建築物(建築主別・用途別・構造別)	
		滅失建築物(面積・戸数)	
		高層・地階建築物(戸数)	
所得	得	区市町村民税納税義務者	約 750件
		区市町村民税課税状況(所得種別)	約 55,400件
事業所	所	事業所数(業態別・規模別)	約 1,120件
		事業所数(形態別・規模別)	約 67,500件
福祉	社	生活保護(人員・施設数・在籍者数・定員数等)	約 460件
		老人福祉(人口 " " ")	約 28,100件
		児童福祉(人口 " " ")	
		身体障害者福祉(施設数・在籍者数・定員数等)	
		精神薄弱者援護(" " ")	
医療	療	医療(医療種別・施設数・設備・診療科目等)	約 210件
		病院(病床数・患者数・従事者数等)	約 15,000件
		死亡数(死因別)	

		データ項目数	データ件数
ゴミ	収集量	約 20件	約 200件
	処理量		
住宅	住宅数（所有関係別・構造別・日照時間別・敷地規模別等）	約 170件	約 10,500件
	着工住宅戸数（利用関係別）		
	住いの広さ（世帯別・種類別・所有関係別）		
道路	面積（管理主体別・所有者別等）	約 80件	約 5,700件
	延長（ " " ）		
	安全施設設置数		
公園・緑地	公園（種類別・設置別一施設数・面積）	約 80件	約 4,100件
	遊び場（施設数・面積等）		
	森林（面積・材積）		
教育・文化	学令前教育（幼稚園数・在園児数等）	約 280件	約 19,000件
	義務教育（小学校・中学校一学校数・学級数・生徒数等）		
	高等教育（高等学校一学校数・学級数・生徒数等）		
	文化施設（体育施設数・図書館数等）		
交通	施設数（停留所・駐車場等）	約 60件	約 3,700件
	自動車（保有台数・交通量）		
	道路現況（道路面積・延長等）		
	交通事故（傷害別・事故形態別・発生件数等）		
上・下水道	上水（給水・使用水量・使用件数等）	約 40件	約 1,400件
	下水（普及面積・人口等）		
公害	固定発生源（届出数・事業所数）	約 20件	約 1,100件
	苦情陳情受付件数（発生源別・現象別）		
消防	火災・震災（発生件数・災害者数・火災棟・面積・損害額等）	約 170件	約 8,400件
	火災状況（建物階別・原因別）		
	救急活動（処置別・事故別件数）		
	危険物等取扱施設数		

(4) 利用方法

データ・バンクの操作は、会話型で行う即時処理と、バッチで行う非即時処理が準備されている。東京都の端末機の操作を図2・3に示す。



キー操作

画面選択キー	操作の開始、終了をはじめとして、このシステムで作成される画面を切替えるチャンネルのようなものである。 開始キーを押した後は、どのキーを押してもかまわない。
内容指定キー	一つの画面にある選択肢を選ぶキーで通常は数字の1, 2, …… 9, 0を用いる。 ライトペンを用いる場合は選択肢の頭にある「?」マークをペンで指し、ペン先のスイッチを押す。
送信キー	一つの画面における選択が終了したことをシステムに知らせるキーである。 ライトペンを用いる場合は、画面右下に出る送信マークをライトペンで指す。
印字キー	画面にでている内容をそっくりプリントしたいときに用いるキーである。

ライトペン操作

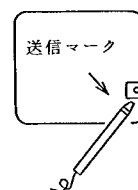
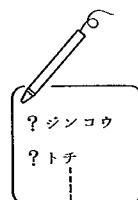


図2・3 端末機の操作

2-2 大阪府におけるメッシュ・データ・バンク・システム²⁶⁾

(1) 概 要

府政の各分野における事業計画の適確な推進をめざすために、その基礎となる統計、その他のデータをメッシュ・データ・ベースとして蓄積し、広く利用できるようにしている。

現在は、研究段階であるが、昭和50年にパイロット・モデルを作成し、それを基礎としてシステム開発をすすめている。

本システムの主な特徴は次のとおりである。

a) ゾーン・データ・ベース指向

ゾーンは、原則としてメッシュで構成されているが、すべてのゾーンに対するメッシュ化が期待できないため、主となるのはゾーンであるとする概念。

b) 回答における図形処理指向

(2) 位置の表現

位置の表現は、メッシュとゾーンである。メッシュについては、次の2方式が準備さ

a) 経緯度メッシュ

b) 14座標系メッシュ

ゾーンは、次の3種に対して規定されている。

a) 大地域区分（北大阪、大阪市域、南大阪など）

b) 市区町村

c) 学 校 区

これらの異なる座標系の変換は、次のようにして行われる。

a) 単一の変換式による方法

14座標系メッシュ → 経緯度メッシュ

b) 変換テーブルによる方法

経緯度メッシュ → 市区町村ゾーン

町字による地点 → 経緯度メッシュ

(3) データ・ベース構成

データ・ベース構成は図2.5のようにFCT, ICT, IAT, CTMD, およびDataの5種のファイルで構成される。

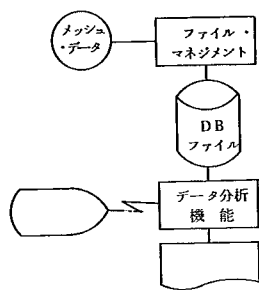


図 2.4 システム構成

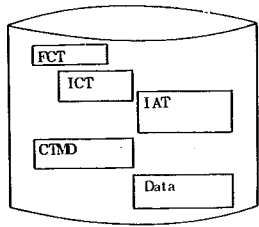


図 2.5 ファイル構成

a) FCT (File Control Table)

データ・ベースの総合的な管理に必要な情報が図 2.6 図のように格納されている。

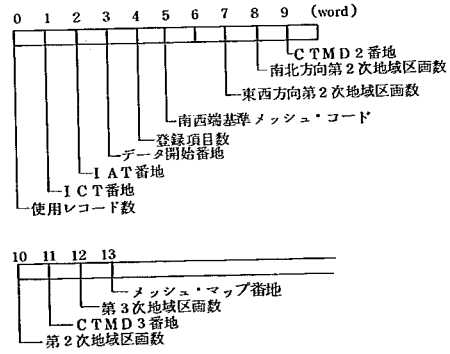


図 2.6 FCT の構造

4ZZ

b) ICT (Item Control Table)

データを検索するための情報が各項目ごとに図 2.7 のように格納されている。



図 2.7 ICT の構造

c) IAT (Item Attribute Table)

各項目の属性が図 2.8 のように格納されている。

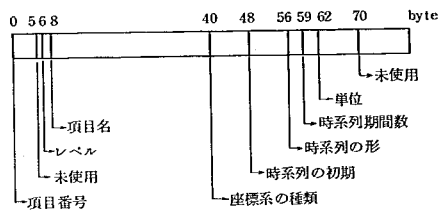


図 2.8 IAT の構造

d) CTMD (Conversion Table of Mesh Code into Data address)

メッシュ・コードから実際のデータの記憶されている番地を算出するために必要な情報が図 2.9 のように格納されている。

C T M D 2

[illegible]

CTMD 3

A blank 10x10 grid for graphing, consisting of 10 columns and 10 rows of squares.

図 2.9 CTMDの構造

e) Data

メッシュの属性値が図2・10のような記憶構造で格納されてある。

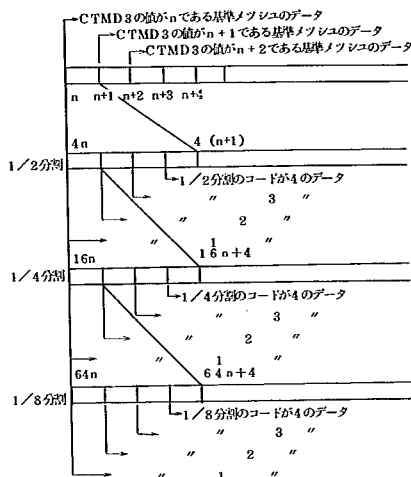


図2・10 データの記憶構造

(4) 図形処理

利用者はデータ・ベースからの回答として次のような図形を得る事ができる。

a) 多項目表示、時系列表示等の図表化

b) 地図の図化(表章のバック・グラウンドとして地図を用いる。)

大阪府では、地域データの一部として地理データがファイル化されており、これを本システムのサブ・システム(ベース・マップ・システム)として用いている。

ベース・マップ・システムの主な機能は次のとおりである。

a) 縮尺機能(地域の拡大、縮小)

b) 項目選択機能

(i) 市町村区画界

(ii) 市区町村名(漢字)

(iii) 海岸線、河川、および湖沼

(iv) 道路、鉄道

c) 地域選択機能

利用者は選択のキーとして、市区町村コードまたはこれらの集合、あるいは、メッシュ・コードによる特定地域の選択が可能である。

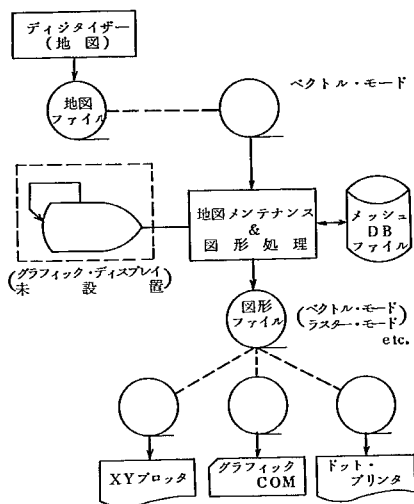


図2.11 ベース・マップ・システムの構成

2-3 北上川水系統合管理システム

(1) 概 要

北上川水系の流水管理を行うためのシステムである。

本システムでは、流水管理を、高水、低水、水質の各方面からとらえ降雨予測、流出予測、洪水追跡、流水配分、ダム操作、取水調節等を円滑、正確かつ、迅速に行う。

a) 高水管理機能

高水管理機能は、高水時に、洪水予報指定河川について、建設省と気象庁とが共同で行う洪水予報、ならびに、水防警報指定河川について洪水、または、災害が起る恐れのあるとき、水防を行う必要がある事を通報する水防警報、ダム等の放流量を制御する流量調節機能を持っている。

b) 低水管理、水質管理

低水時に、水の有効利用をはかるため、ダム放流量の調節、導水施設などによる水の移動など、諸河川の流況をふまえて広域的な流水配分を行う必要がある。また、水量的余裕のない場合、異常水質が発生した場合の即時的な水配分への応答機能が必要であるが、現在のところ、このシステムは立遅れている。

(2) 入力データ

水	位	河川水位
流	量	河川流量
雨	量	
水	質	水位、水温、導電率、溶存酸素、PH、濁度、シアニオン、アンモニア、その他合計15種
ダ ム	諸 量	貯水位、貯留量、放流量、流入量、その他合計7種
高 潮	情 報	潮位、波高(平均、周期)、風向、風速、気圧
レ ー ダー	雨量	雨量、反射電力
気 象	情 報	

表 1 . 4 入力データ

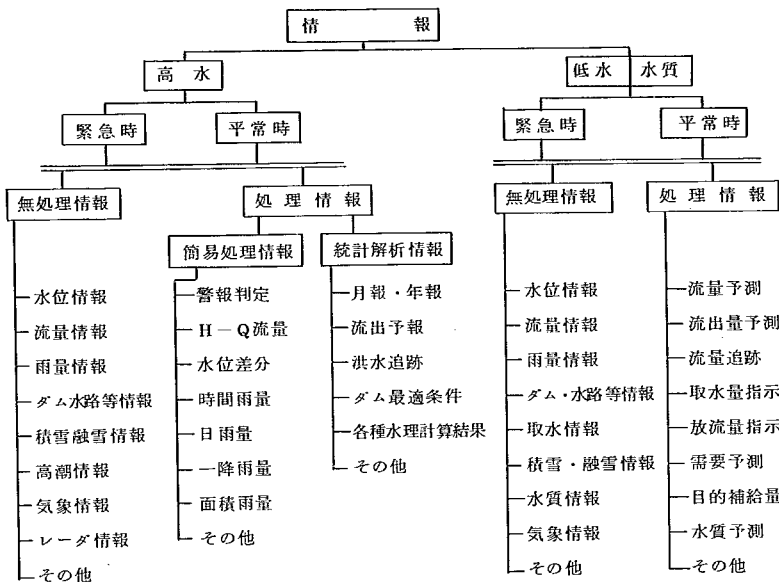


図 2. 1 2 情 報 体 系

2-4 広域水系統合管理システム²⁷⁾

(1) 概 要

本システムは、関東地方建設局管内の利根川水系を中心とする複数河川を対象とした、河川に関する各種情報を、建設省専用無線回線を伝送路として利用し、収集処理するオンライン・リアルタイムおよび、バッチ処理システムを形成する事により、高水時の即応対策、低水時の効果的調整、水資源の有効利用、水質の監視等の治水、利水、環境保全管理強化を目的とするシステムである。

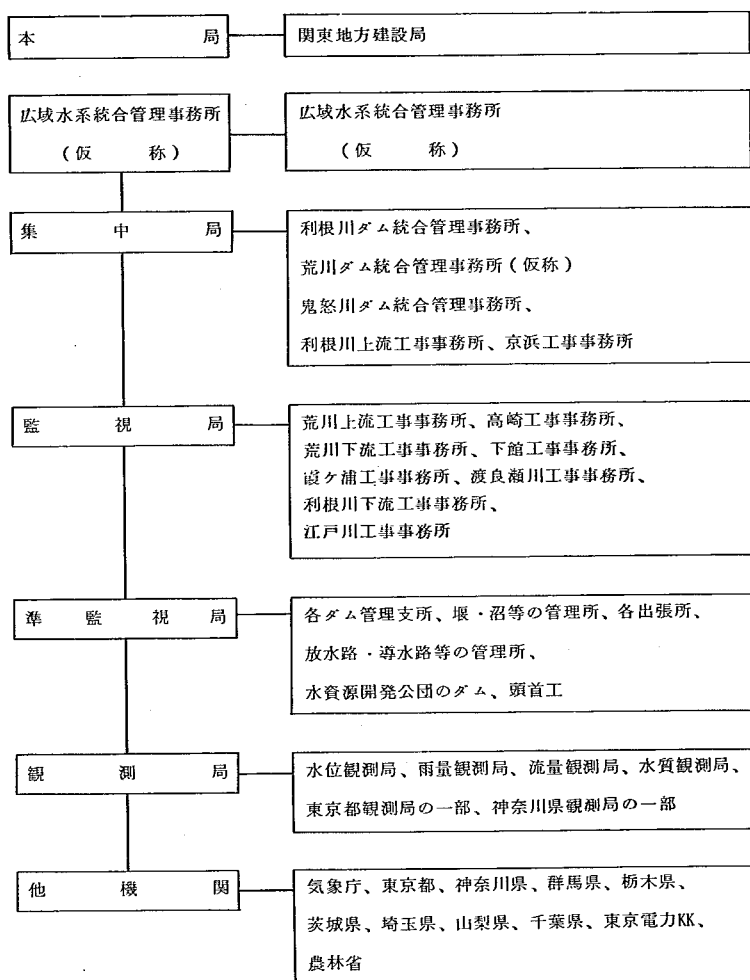


図 2 . 1 3 組 織 図

(2) システム機能

システム機能、および機能分担、情報の流れを図2・14、図2・15に示す。

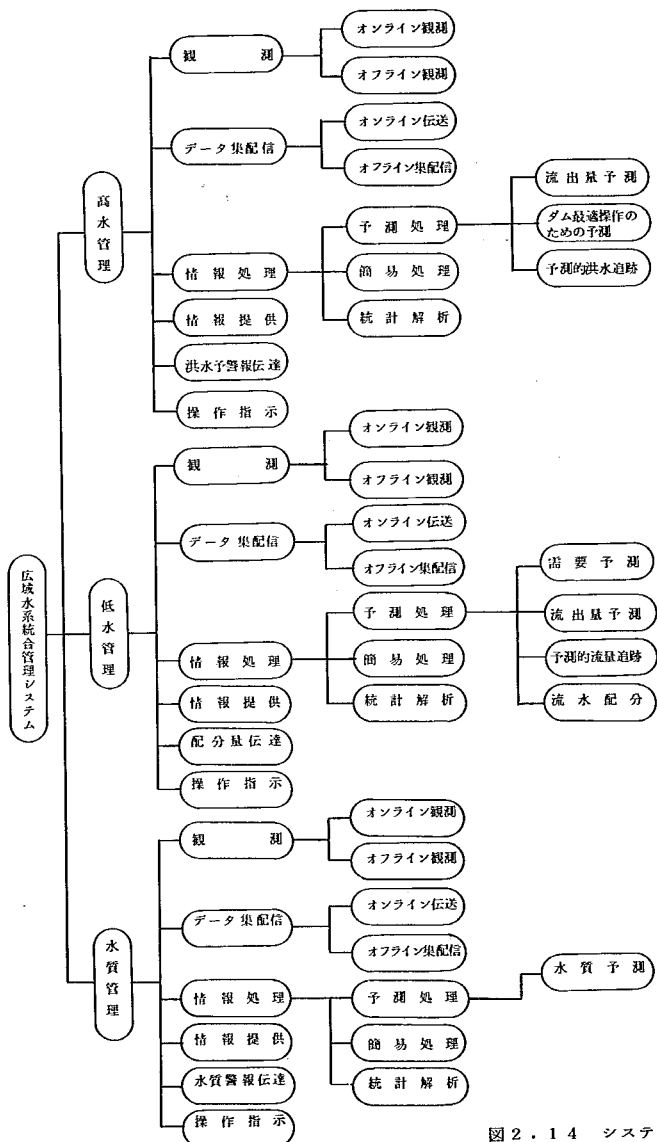


図2・14 システム機能

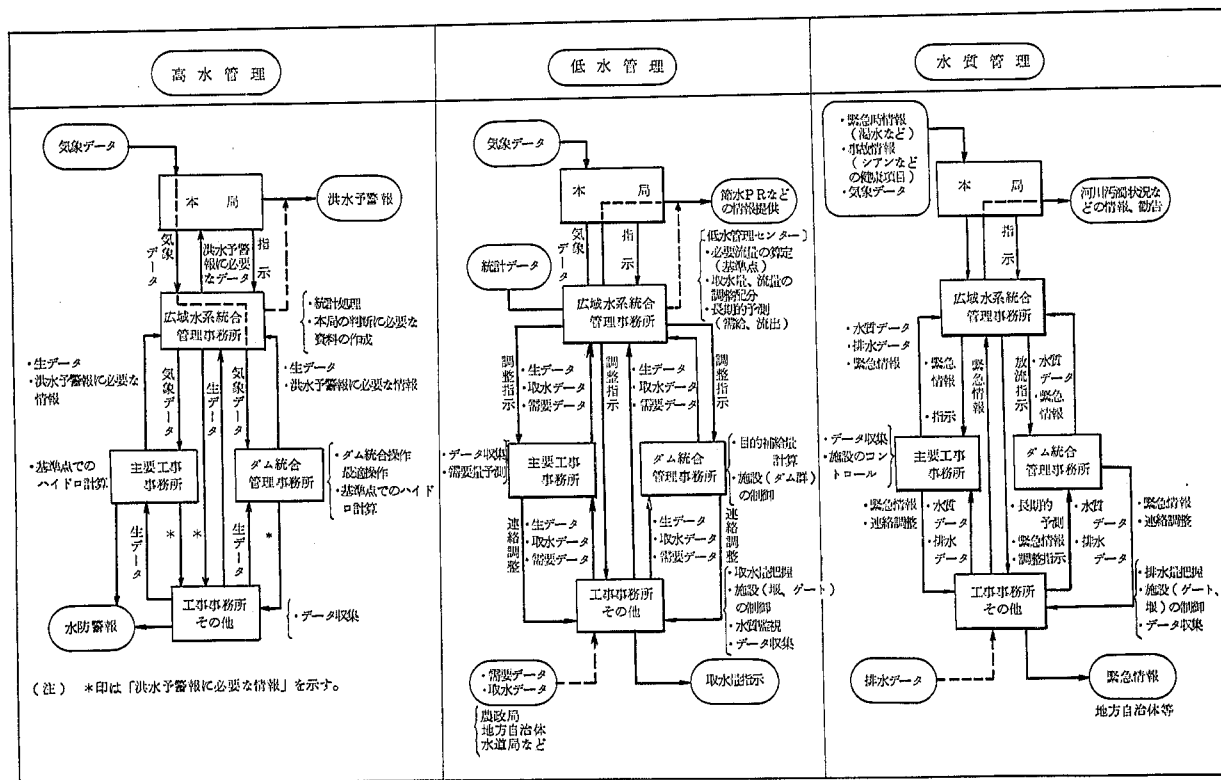


図2.15 情報の流れ

2-5 省庁におけるデータ・バンクの事例

省 庁 名	システム名	概 要	備 考
国 土 庁	国土情報管理システム	国土情報の総合的な管理を目的とし、それに対するデータ・ベースの研究開発を昭和50年度より行っている。	研 究 中
大 蔵 省 (主計局)		予算の作成および国会審議を通じて必要な情報を電子計算システムによって提供するもので当面は国会議事録および類似関連予算の蓄積・検索システムの開発を進めている。今後は更にその対象の拡大を予定している。	
経済企画庁		国民所得、物価、貿易、機械受注、労働賃金等のデータが約12,000係列(1系列最大300期)の時系列データが収録されている。DBのFile編成は順インデックス編成であり、処理プログラムDBへのアクセス・ルーチンはすべてFORTRANで記述されている。 なお、クロス・セクション・データの収納検索システムは現在検討中である。	
通 産 省	統計データ・バンク	各種の統計、調査データの総合管理、処理のほか、情報検索システムとして文字情報(英文、漢字)を扱う事ができる。	
	総合貿易情報システム	貿易関連システムは、通産省のシステムのほか大蔵省が電々公社のシステム・サービスで52年度から運用稼働する予定の通関システム、運輸省の国際運輸貨物情報システムなどが検討されている。本システムは、他省庁とのこれらのシステムとのコンセプトをとろうとしているが	

省庁名	システム名	概 要	備 考
		まだまとまっていない。 システムは次の4種のシステムで構成されている。 ・商取引システム ・商取引と物流システム ・通関システム ・輸出入統計システム	計画段階
行政管理庁	法令検索システム	対象とする法令は法律、政令、勅令など法務省司法法制調査部による「現行日本法規」である。 検索は法令用語および書誌的事項で行い次の4種のモードがある。 ・該当条文検索 ・関係条文検索 ・改廃経過検索 ・判例検索	53年度稼働をめざす
防衛庁 (航空自衛隊)	部品補給システム	航空自衛隊は補給処から基地、工場へ通信機、航空機、車両などの部品を補給しているが年々取扱い品目が増大してきた。それに対処するため、RAPIDを用いたオンライン・ネットワーク・データベースを作成して用いる。	インクワイアリ
農林省	生鮮食料品流通情報システム	野菜・果物・肉類などの生産出荷の数量および各地の市況を収集し、これを生産者や流通業者に提供する。	オフライン
労働省 (労働市場センター)	労働市場情報	全国466の公共職業安定所、47都道府県庁、労働省とを通信回線で結び広域職業紹介業務、失業保険関係業務などを行っている。	

省庁名	システム名	概 要	備 考
運 輸 省	自動車検査登録システム	昭和45年3月から開始し、翌年4月から全国65の渾運事務所を通信回線で結んだ全国規模のネットワークである。	
警 察 庁		運転車管理、犯罪手口照会、犯罪統計の業務をコンピュータにより行っている。運転者管理センターは運転者約2,500万人(1970)の免許台帳と違反、事故歴、処罰記録などを収録している。	

3. 地理情報の表現

情報は、属性データとしてそれぞれ、ゾーン、メッシュ、不定形に対応づけられる。しかし、例えば、東京都というゾーンは、単に地理に東京都というアイデンティフィケーションをつけたものにすぎない。これだけでは東京都がどのような位置関係にありどのような地理的形状しているのかは知り得ない。この問題を知るためには、ここでのべる地理情報の蓄積が必要である。

地図上に、表現されている地理図形を分解すると、点、線、面の3種の次元の異なる要素となる。この3種の情報を表現する方法としてメッシュ方式と座標方式が考えられる。

(1) メッシュ方式

国土を長方形区画に階層的に分割したもので、メッシュはそれぞれメッシュ番号を有し、その番号により地理的な普遍位置は算出される。

例えば、80Kmのメッシュでメッシュ番号5339が与えられると、その位置は、次のように簡単に求まる。

$$\begin{array}{c}
 \begin{array}{|c|c|c|}
 \hline
 5 & 3 & 9 \\
 \hline
 \end{array} \\
 \begin{array}{cc}
 I_1 & I_2 \\
 \hline
 \end{array}
 \end{array}$$

$$\begin{array}{l}
 G. \\
 (x_0, y_0) \\
 \hline
 (x_1, y_1)
 \end{array}$$

$$\begin{aligned}
 x_1 &= 100 + I_2 && = \text{東経 } 139^\circ \\
 y_1 &= 36 + (I_1 - 54) \times \frac{4}{6} = \text{北緯 } 35.67^\circ \\
 x_0 &= x_1 + 0.5 \\
 y_0 &= y_1 + 0.5 \times \frac{4}{6}
 \end{aligned}$$

メッシュ方式における困難さは、任意形状の形に関する情報の取り扱いである。例えば、ある河川をメッシュ・データに埋込む事を考えると、河川は貫いているメッシュの集合として表現する事になる。よって、河川の形状は、ぼけてしまう恐れがある。もちろん、メッシュの分割を細くすれば高精度で近似はできるがデータ量は膨大となる。

メッシュによる線、面の表現は図示すると図3.1～図3.3のようになる。

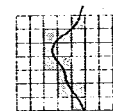


図3.1 線の表現

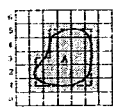


図3.2 面の表現-1



図3.3 面の表現-2

地域A: $\{(11)u(12)u(13)u(14)u(15)u(21)u(22)u(23)u(24)u(25)u(31)u(32)u(33)u(34)u(35)u(42)u(43)u(44)u(45)u(52)u(53)u(54)u(55)\}$

メッシュ方式による線、面の近似は図3.1～図3.3のごとくであり、このデータから直線の長さ、面積などを求める問題が与えられた場合は、何らかの概念によって近似計算をすることができ。その一つの方法としては、線、面が所属するメッシュ番号とメッシュとクロスする点の比率データを与える方法があるが、この方法はデータの収集が困難であり現実的ではない。他の方法としては、図3.2の点線のように部分的に含まれるメッシュはその $\frac{1}{2}$ を着目地域に含むという考えである。したがって図3.2で1メッシュの区画を1とすると、

面積 $A = \text{全部 } A \text{ に含まれるメッシュの数} + \text{隣接地域と共有するメッシュ数} \div 2$

となる。一般には、後者の方式が妥当と思われる。

(2) 座標方式

座標方式は、地理をX、Y座標値で与え、点、線、面の関係をグラフ構造として捉える方法である。この方式は、GIS⁸⁾(地理情報システム)として研究されていて、かなり正確な地理表現が可能であり、地理表現としてもっと期待されている。

情報は、定義された点、線、面に対してそれらの属性データとして結合される。点、線、面の関係は次のようなグラフ構造で表現できる。

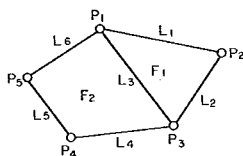


図 3.4 地理の多角形分割

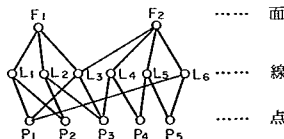


図 3.5 地理のグラフ表現

このグラフ構造を関係行列で表現すると次のようになる。

a) 面と線の関係行列

線 \ 面	面の属性データ					
	L ₁	L ₂	L ₃	L ₄	L ₅	L ₆
F ₁	1	1	1	・	・	・
F ₂	・	・	1	1	1	1

表 3.1 面と線の関係行列

1 : その線がその面の構成要素である。

・ : その線とその面は無関係である。

b) 線と面の関係行列

点 線	P ₁	P ₂	P ₃	P ₄	P ₅
L ₁	1	1	·	·	·
L ₂	·	1	1	·	·
L ₃	1	·	1	·	·
L ₄	·	·	1	1	·
L ₅	·	·	·	1	1
L ₆	1	·	·	·	1

線の属性データ	
項目 A	項目 B
...	...
...	...
...	...
...	...
...	...
...	...

点 座標 項目 E	...	...	...	...	...
...	...	...	...	...	...

表 3.2 線と面の関係行列

この例では、線に対する方向を持たせない無向グラフ表現であるが河川、下水道のネットワークのように方向を示す場合には、有効グラフで表わす事ができる。

この関係行列の演算により次のような事が可能である。

(i) 複数面の集合による外郭線の抽出

$F_1 \cup F_2$ の外郭線

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, A \cdot X = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{matrix} \cdots L_1 \\ \cdots L_2 \\ \cdots L_3 \\ \cdots L_4 \\ \cdots L_5 \\ \cdots L_6 \end{matrix}$$

(ii) 隣接面の抽出

(iii) 接続線の抽出

(iv) 面の分割線の抽出

メッシュ方式と座標方式のそれぞれの変換性は、次のとおりである。

- 座標方式 → メッシュ方式 (可能)
- メッシュ方式 → 座標方式 (可能であるが現実的でない)

参 考 文 献 お よ び 参 考 資 料

- 1) 建設省土木研究所企画部システム課「土木計画論」土木研究所資料第971号
1975年11月
- 2) 中 原 啓 一 「情報検索」昭和49年8月 社団法人 電子通信学会
- 3) 三重野博司編 「データバンクその理論と実際」昭和48年8月 大河出版
- 4) 橋 本 昌 幸「情報検索のABC」 昭和50年2月 日本出版協会
- 5) 渡辺龍雄編 「データバンク入門」 昭和46年2月 日刊工業出版社
- 6) 東京都企画調整局「国内及び諸外国におけるデータバンクの現状と問題点」
昭和49年3月
- 7) 建設省大臣官房情報管理室 「国土情報全体システムの研究」 昭和51年9月
- 8) " 「地理情報システムの方法」 昭和50年7月
- 9) 東京都企画調整局「東京都におけるデータバンクのあり方」 昭和50年5月
- 10) " 「データ索引誌」 昭和51年6月
- 11) 茶谷達雄 他 「東京都における情報管理システムのパイロット・モデルの設計思想」
1976年 情報処理 VOL17 情報処理学会
- 12) 石田 喬 也 「データベース理論」 1976年12月 社団法人ソフトウェア産業振興協会
(研修資料)
- 13) 西 野 博 二 「データベース・システムの研究開発動向」1976年 情報処理
VOL17 情報処理学会
- 14) E. F Codd 「A relational model of data for large shared
data bank」 1970. CACM Vol13, No. 6
- 15) 建設省技術研究会「河川の流水管理システムに関する研究」
建設省 技術研究会河川部門指定課題論文集 1976年11月
- 16) 建設省土木研究所企画部システム課 「電子計算機を活用した地域情報システムの構想」
土木技術資料 1976年8月
- 17) " 「電子計算機による地域情報システムに関する基礎調査」
土木技術資料 1976年3月
- 18) " 「山積する情報の整理と交流のために」
昭和50年3月

- 19) 日本規格協会編 「JISハンドブック 情報処理」 1976年
- 20) 奥平耕造 「メッシュ・データの基礎」 財団法人 日本科学技術連盟
- 21) 植村俊亮 「CODASYL方式のデータベース・システム」 1976年 情報処理
VOL17 情報処理学会
- 22) 富士通 「INIS/RICS RAPID解説書I」 87SR-0020-2
- 23) IBM 「情報管理システム(IMS/VS) 概説書」 N:GH20-1260-1
- 24) 渡辺純一 「階層構造のデータベース」 1976年 情報処理 VOL17 情報処理学会
- 25) 加藤文夫・斉藤正男 「階層型データベース」

1976年3月 情報処理学会

(MI研究会資料9)

- 26) 富士通 「FACOM ジャーナル」 1976年5月
- 27) 建設省土木研究所企画部システム課「水質源の利用配分利用に関するシステム分析」

土木研究所資料 第1125号 昭和50年3月