

経情協 45-6

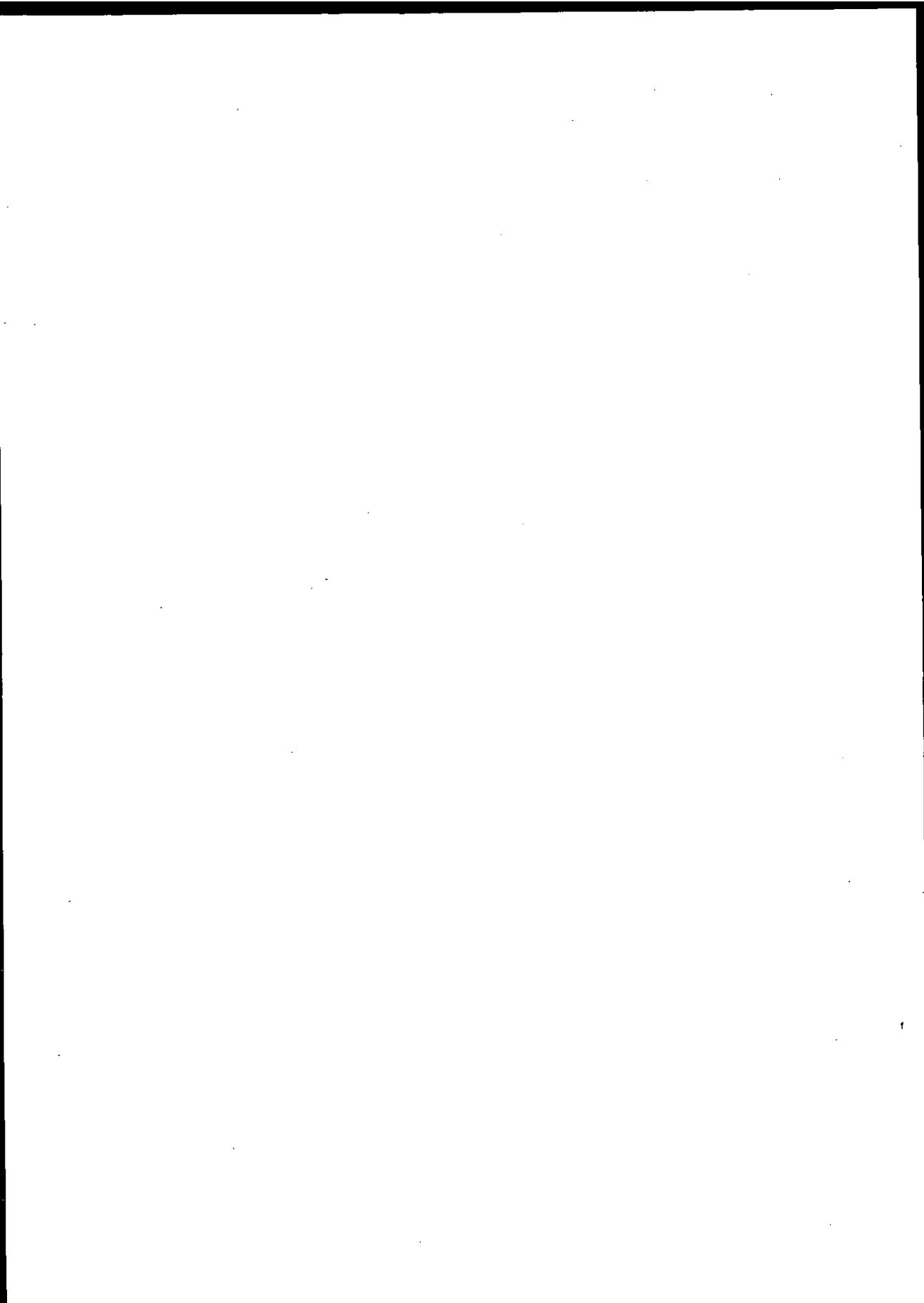
経情協45—6

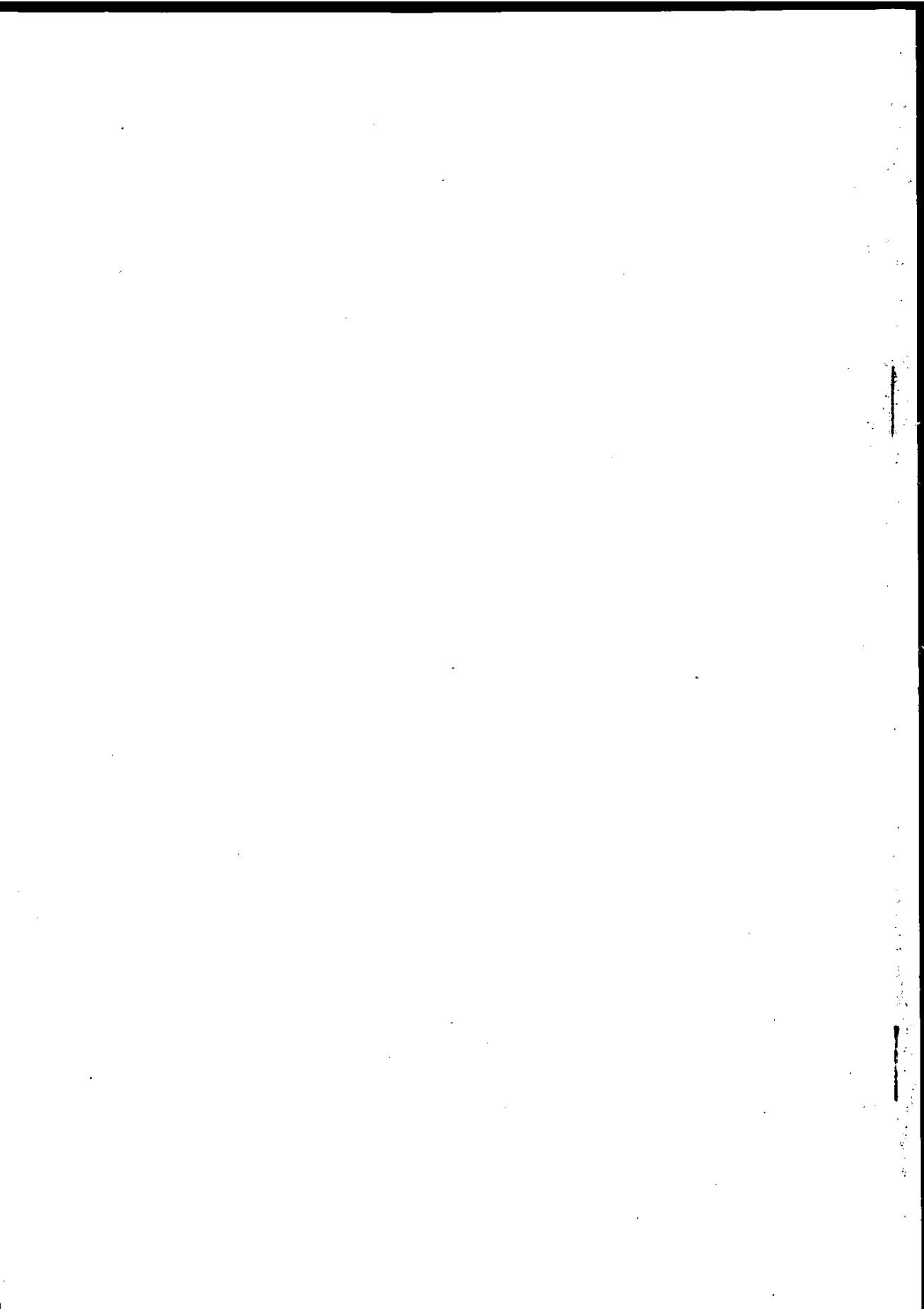
データベース研究に関する中間報告書

データベース研究資料 No. 3

昭和45年 6 月

財団法人 日本経営情報開発協会





は し が き

私たちに与えられた「データバンク」という研究課題は、極めて取り組みにくいテーマである。研究会を組織して正式な研究を始める前に、所謂ブレン・ストーミング的な自由討論を必要とすると考えた。

その意味において、まず、学識経験者に集って頂き、研究テーマを設定しない懇談会を数回開いた。この討論を通じて、データバンクとは何を意味するか、また、どのような研究アプローチが適切であるかを見出すことにつとめた。

つづいて、すでに活動を始めているデータバンクの現状と問題点を把握するために、アメリカと日本のデータバンクの現状報告を出発点とするシンポジウムを開催した。ここでは、広範囲に問題をとりあげるとともに、その本質を究明するための討論が展開された。

このようなかなり長期の予備討議を経て、データバンク基礎研究委員会が設置された。委員全員で「データ論」「データバンクの概念」などの基礎研究を行った後に、(1)情報ネットワーク (2)情報の価格形成 (3)システム開発 の3つのグループに分れて研究を進めてきた。

この間、昭和45年3月に本研究委員会の一部の者をアメリカに派遣し、データバンクの現状を調査する機会を持った。調査団員は1年近い研究により明確な問題意識を持っていたので立派な調査の成果をあげることができた。

上記のような討議、実態調査および研究会を通じて、データバンク研究は当初の混沌としたものから、次第に光明を見出したわけである。この研究はさらに昭和45年度も継続して実施されるが、その中間報告書

として、3つのグループの主査の論文を中心とした問題提起を行う。これらの問題の検討が今後の研究の中心的テーマになるものと思われる。

第1章の"システム開発とデータバンク"においては、NASAの開発した"FAME"のモデルがコンピュータ・システムを用いたデータバンクを活用している体験を、社会開発の分野で適用し得る可能性を追求しようとするものである。

第2章の"情報の価格形成"においては、わが国においてはクラブ財的な情報は少数の当事者間の協議により価格が決められているが、いまだ公開の情報市場というものは成立していないし、情報の価格形成の普遍的な原則は認められていない現状に基いて、今後の情報の流通に関する諸問題を追求しようとするものである。

第3章の"情報ネット・ワーク問題"においては、データバンクの基本的機能が需要指向型の情報流通機能であるとの理解から、まずデータの範囲を定め、次にデータバンクの運営主体や情報供給対象に関する概念を把握した上で、データの収集・分析・評価とタグファイルおよびデータファイル構成の技術的問題を追求しようとするものである。

なお、この中間報告書は、日本自転車振興会の機械工業振興資金による「経済統計データバンクに関する基礎調査」の昭和44年度の中間的研究報告に該当するものである。

昭和45年6月

データバンク基礎研究委員会

委員長 北川 敏 男

データバンク基礎研究委員会「50音順・敬称略」

- 委員長 北 川 敏 男 （九州大学・教授）
- 委 員 足 立 哲 朗 （日本興業銀行・調査部調査役）
- 今 井 賢 一 （一ツ橋大学・助教授）
- 印 東 太 郎 （慶応義塾大学・教授）
- 河 合 三 良 （行政管理庁・行政管理局長）
- 佐久間 孝 （電力中央研究所・電子計算室総括課長）
- 鈴 木 康 （日本開発銀行・設備投資研究所主任研究員）
- 高 瀬 保 （京都産業大学・教授）
- 中 井 浩 （科学技術情報センター・調査役）
- 野 口 照 雄 （興亜石油・専務取締役）
- 長谷川 寿 彦 （日本電信電話公社・データ通信本部・調査役）
- 堀 比呂志 （関西電力・企画部付）
- 若曾根 和 之 （通産省・電子政策課課長補佐）
- 渡 辺 竜 雄 （通産省・データセンター班長）
- 馬 越 善 通 （日本経営情報開発協会・理事）

▽ 情報ネットワーク・グループ

- 。中 井 浩
- 長谷川 寿 彦
- 渡 辺 竜 雄

▽ システム・グループ

。野 口 照 雄

高 瀬 保

堀 比呂志

渡 辺 竜 雄

▽ 情報価格形成・グループ

。印 東 太 郎

今 井 賢 一

佐久間 孝

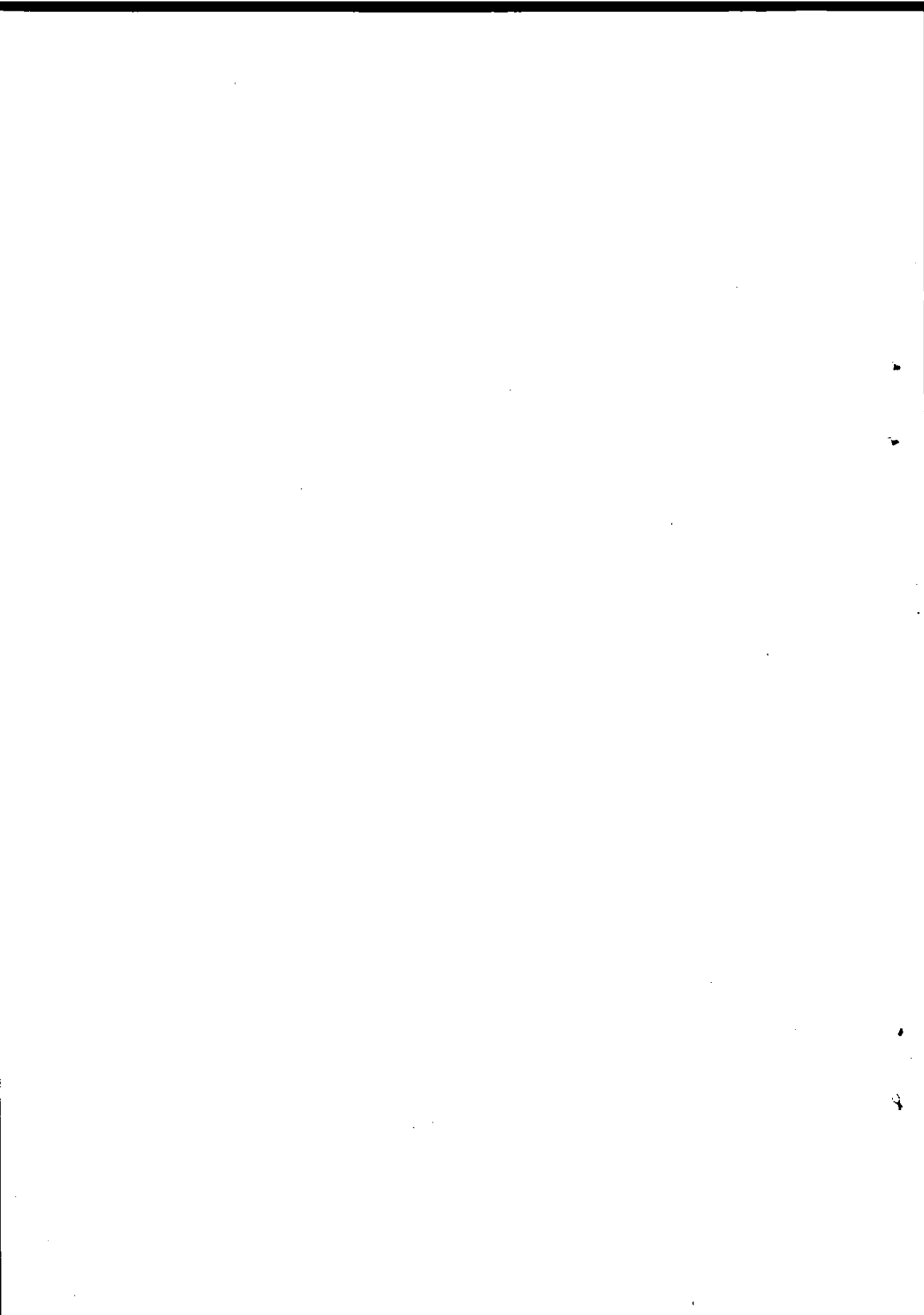
渡 辺 竜 雄

目 次

はしがき

第 1 章 システム開発とデータバンク	1
1 研究課題の選択	1
2 システムズ・アナリシスとデータバンク	2
3 FAMEの社会開発への応用	11
第 2 章 情報の価格形成	14
1 分科会開催経過	14
2 ケース・スタディの結果の概略	15
3 ケース・スタディにもとづく結論	17
4 情報の価値と価格形成をめぐる討論	18
第 3 章 情報ネット・ワーク問題	25
1 需要指向型情報流通とデータバンクの基本的機能	25
2 データバンクにおけるデータの範囲	27
3 データバンクの性格に関する概念的把握	29
4 データバンクを構成する成分・機能	31
5 要求のアクセスとマッチングの条件化	41
6 ファイル・サーチ	42
7 データ出力	48
8 回答の編成	50
9 データ(質問)の伝送	50
10 オペレーション・データとプランニング・データ	51
11 ネット・ワーク	53
12 システム記述と評価	59

(記録) 昭和44年度データバンク研究の経過報告



第1章 システム開発とデータバンク

1 研究課題の選択

「システム開発とデータバンク」に関する研究を進めるにあたり、我々はケーススタディをとりあげてみることにした。最初に沖縄の問題と関西近辺の開発問題（関西地方で万博会場のコンピュータを情報センターとして用いる計画がある）を考えた。しかし、両者ともすでに研究に着手されているテーマでもあるので、とりやめることにした。そして別の観点から、多摩ニュータウン、成田空港、四国の経済開発、筑波山の学園都市、東京湾の開発の問題等を候補として考えたが、その中では、成田空港と四国の開発の問題が、研究資料の入手などの点から、適当とみられる。

成田空港の方は、空港公団と航空政策研究会がそれぞれ研究を行っている。新東京国際空港公団が発足して研究を進めているが、空港内部の開発に限られている。外部の問題は、千葉県が空港周辺の住宅問題を研究したり、東京都の高速道路協会が、成田と東京の間の問題を扱うなど個別的には研究されているが、システムの的には考えられていない。航空政策研究会は、成田空港についてのシステム開発のデータをかなり持っているが、わが国の総合航空政策の一環として扱っている。

我々が成田空港のシステム開発をとりあげるのは、周囲の例えば交通問題からとか、住宅問題との関連はどうか等周辺の開発を含めてデータの体系的な収集と、その処理に視点を置いた研究を行ないたいためであり、そういう観点で研究を押し進めるつもりである。

四国の開発については、四国の経済発展は関西のそれと密接な関連があり、システムとしてのとりあげ方の例として、四国電力からデータがはいるのでとりあげることとした。その資料は四国経済連合会が持っている。

来年度は、これらのうちどちらか1つか、又は両方を研究してゆくわけであるが、システムそのものの開発を研究するのではなく、システムと関連してどのようなデータバンクが必要なのか焦点を合わせて研究をしてゆきたい。

2 システムズ・アナリシスとデータ・バンク

近年地域開発の総合的アプローチが特に関心を持たれ始めた。今までのように工場誘致計画、土地利用計画、交通体系計画、住宅計画等、それぞれの異なった組織によって、ばらばらに行なわれてきた。そのため、四日市市のような無茶な工場計画は、公害問題を惹起し、東京近郊の無計画な団地建設は、電車通勤地獄を現出せしめている。そのため地域開発計画のシステム化の必要性が認識され始めてきた。今までの社会開発計画は、流動的なものであるより、固定的であった。このことは、一つの計画がAからZまで一つの完成予定によって、完成するならば、それにて打ち切りといった傾向にあった。流動し発展する社会に対して、計画は、固定化され、社会の発展に取残される傾向にあった。その一つの例が、都心と羽田を結ぶ高速道路の混雑状況であり、現在の住宅建設のホープである団地のスラム化の可能性もある。このように計画の出来上った時点における陳腐化ということは、昔と較べて、これから益々そのテンポは、早まるであろう。

NASA の最近の研究によれば、現在までの計画は、O型であり、一つの計画は、完成までには、全然無修正であったため、完成した計画、進行中の計画、将来計画との間に関連性がなく、それぞれの計画の有機的な連関がなかった。しかしながらこれからの計画は、U型か又はL型か、図型通り閉鎖的なものでなく、常に修正が可能な状態にあるような計画、即ち一箇所の口がオープンになつてゐるような状態が望ましいと勧告している。例えば住宅建設計画の場合、向上する所得水準、建築技術の進歩、住宅要求の度合に関連して、建設しながら、片一方ではないものを破壊して行くという態勢がない限り、公共住宅建設は、応々にしてスラム化し易いものである。

このようなNASAの一つの教訓は、自己の宇宙船開発計画から生れた。それは十年間の長期計画の中で「FAME」というモデルを開発したのである。

「FAME」とは Forecast and Appraisals for management Evaluation の略称である。このモデルの開発目的は、1970年までに「人間を月へ」と、いくつかのロケットと宇宙船の開発計画の中で、限られた予算、限定された期限の枠の中で、1セントも安く、1オンスも軽いロケット、宇宙船を開発するため、常に予算と時間を自動的に記録し、少しでも余計な金や時間を超過するならば自動的に警戒を発し、修正を加えて行くモデルである。そしてロケット開発に伴い、常に新技術が計画の中に繰込まれて行くモデルであった。このようなモデルをつくるにあたって、計画全体のシステムの分析、即ちシステムズアナリシスなくしては、計画遂行はあり得ない。何故なれば、5万個以上の部品と20万人の量を必要としたアポロ計画において、

1つの部品に0.0001の誤差があってもそれが5万個に達すれば、全体で5割の誤差になってしまう。そのような意味で厳格な調整を必要とする。

この「FAME」モデルの中で一番重要なのは、データの収集とその分析である。刻々とうつり変わる状態に応じて、そのデータの更新性が基礎条件である。このモデルは、5つの重要な要素よりなっている。

1. 情報収集システム
2. データの予備プロセス技術
3. 歴史的傾向（過去の傾向）と総合予測を評価する統計技術
4. プログラム全体に対する臨界性という枠で、データと予測を評価する規準
5. 最終評価への報告システム。

データ収集と分析のシステム作りが、この「FAME」モデルの要件となっていることは、以上5つの要素より明解である。これらのシステムは、電子計算機を駆使して、自動的にデータを収集し、分析し、管理者、経営者に必要情報を継続的に提供し、デシジョンメイキングの判断をあたえるだけでなく、必要に応じて、システムに繰込んでおきさえすれば、オート・パイロットの様な自動修正さえ可能なのである。このことは、データ・バンクと長期計画の有機的な関係を呈示している。即ち、NASAでは、データ収集と分析に関して、各必要段階において、電子計算機システムを用いたデータ・ライブラリの重要性をのべているからである。データ・ライブラリとデータ・バンクは、同義語と考えてよいであろうからである。

次にNASAの長期計画のプロセスについて説明してみよう。これは

Probabilistic Long Range Programmingと言われる。これは、企業の長期戦略、社会開発の長期計画のどちらにも適用できるモデルであるが、こゝでは後者を中心にして論議をすすめてみたい。

6つの段階よりなっている。

1. 潜在能力の評価、計画実施機関の持つ能力
2. 環境要素の評価、社会経済指数、又は関連データの評価
3. 戦略設定、目標の明確化
4. 戦略のプログラミング、達成方法の研究
5. プログラミングの評価
6. その選択

以上6つの段階は、計画のシステムの思考の所産である。即ちシステムとは、目的を持った存在であり、その目的達成のために常に効率を求めている。システムズ・アナリシスは、その効率を求めるための手段である。その中で、①の段階では、M I Sという名のデータ・バンクとして、組織内に関するデータを必要に応じて、インスタントに把握させるものである。②の段階では、関連指数のデータ・バンクのことを意味する。

以上、システムズ・アナリシスとデータ・バンクとの関連を、NASAの長期計画モデルの中で明らかにした。

この手法が、社会開発の分野で、すでに有効な働きをし始めている例の典型は、カリフォルニア・プロジェクトという4つの計画、

- ① 地方行政における情報処理
- ② 公害防止モデル
- ③ 犯罪防止モデル

④ 交通コントロール・モデル

以上パイロット研究の中に見られる。これらの研究は、1964年に行なわれたものであるが、これを契機として、いくつかの適用例が、ロス・アンゼルス市、サン・ディエゴ市、ニューヨーク市で見られている。

それを交通計画の中に例を見てみよう。米国運輸省の Northeast Corridor Transportatin Project の調査計画によれば、次のような活動になる。

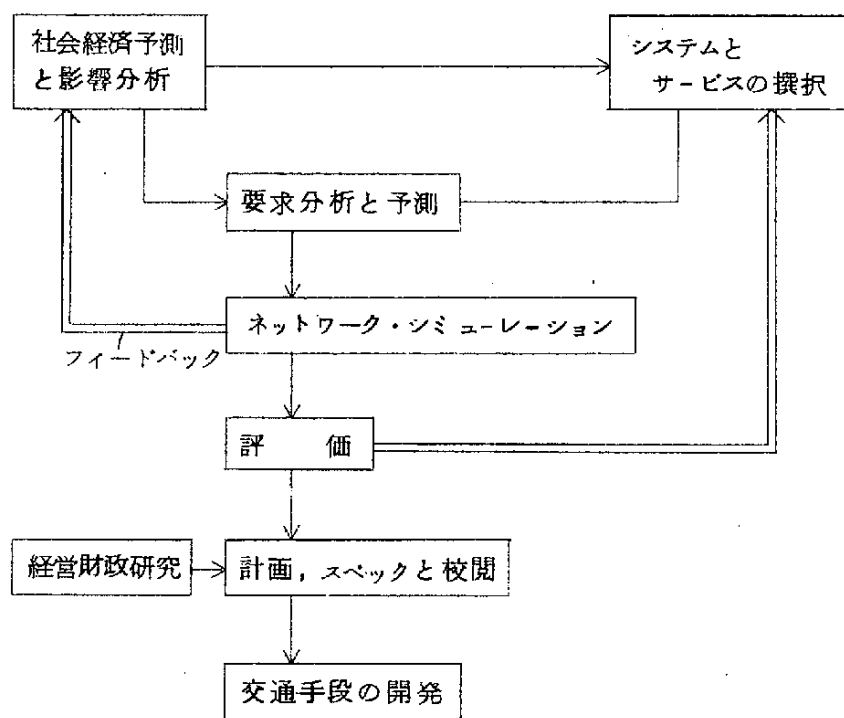
都市間交通の分析と計画は、ワシントンとボストン間の鉄道利用を中心に行なわれている。1980年と1999年の間における地域的都市間交通の要求を決定することが、本計画の目的である。予測される人口と経済活動のレベルにおいて、旅客、貨物の要求予測のモデルシステムが開発されている。鉄道設置地、活動の複合は、三つの手段によってきめられる。傾向の外挿法を基にした土地利用の変化、新鉄道設備とサービス（インプット）による土地利用の変化、他の公共政策に起因する土地利用の変化、の以上である。

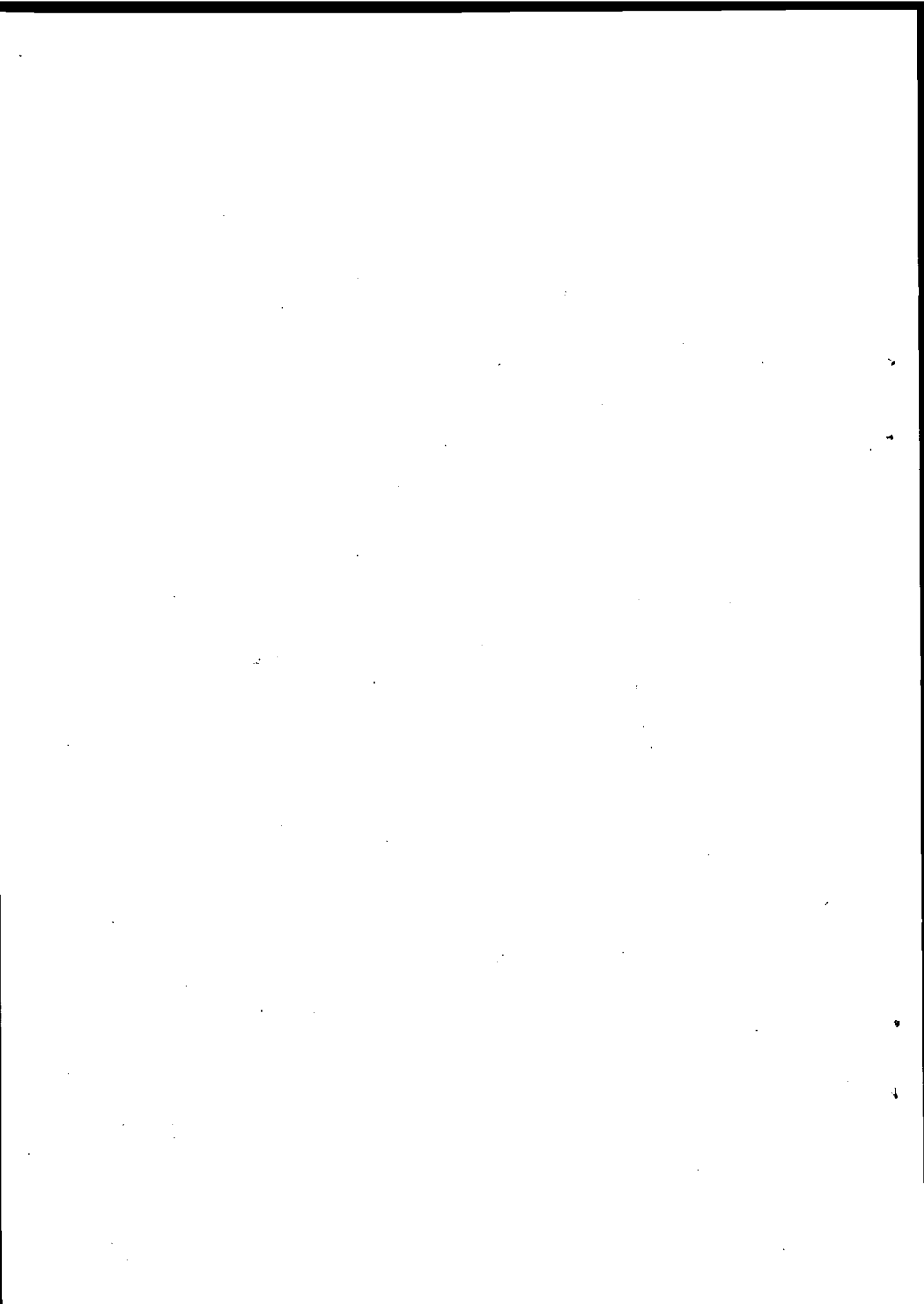
都市交通の場合には、人口、経済活動、土地利用、交通の流れの詳細なクロス・セクション的分析を必要としている。特に、都市とその周辺地域との関連においてである。人口予測、経済活動予測、土地利用形態予測のモデルが用いられる。これらの現在の交通網との関連において、交通要求の予測を都市とその周辺部との間に行わなければならない。

一般的に二段階に分れて研究しうる。一つは、自動車と大量輸送交通とその関係であり、二つは、もし二つ以上の大量輸送交通手段があ

るならば、その手段間の比較検討である。

以上の二つの研究を通じて、交通関係においては、非常に大きな、データ・収集システムを前提としている。即ち交通機関の異った形態の中での流れをつかむためにである。別の言い方をすれば、これは交通関係のMISとも言えるであろう。参考までにカリフォルニア研究の図表を示すことによって、そのデータ・バンクとシステムズ・アナリシスとの関係を明らかにする。





SYSTEMS ANALYSIS MODELS

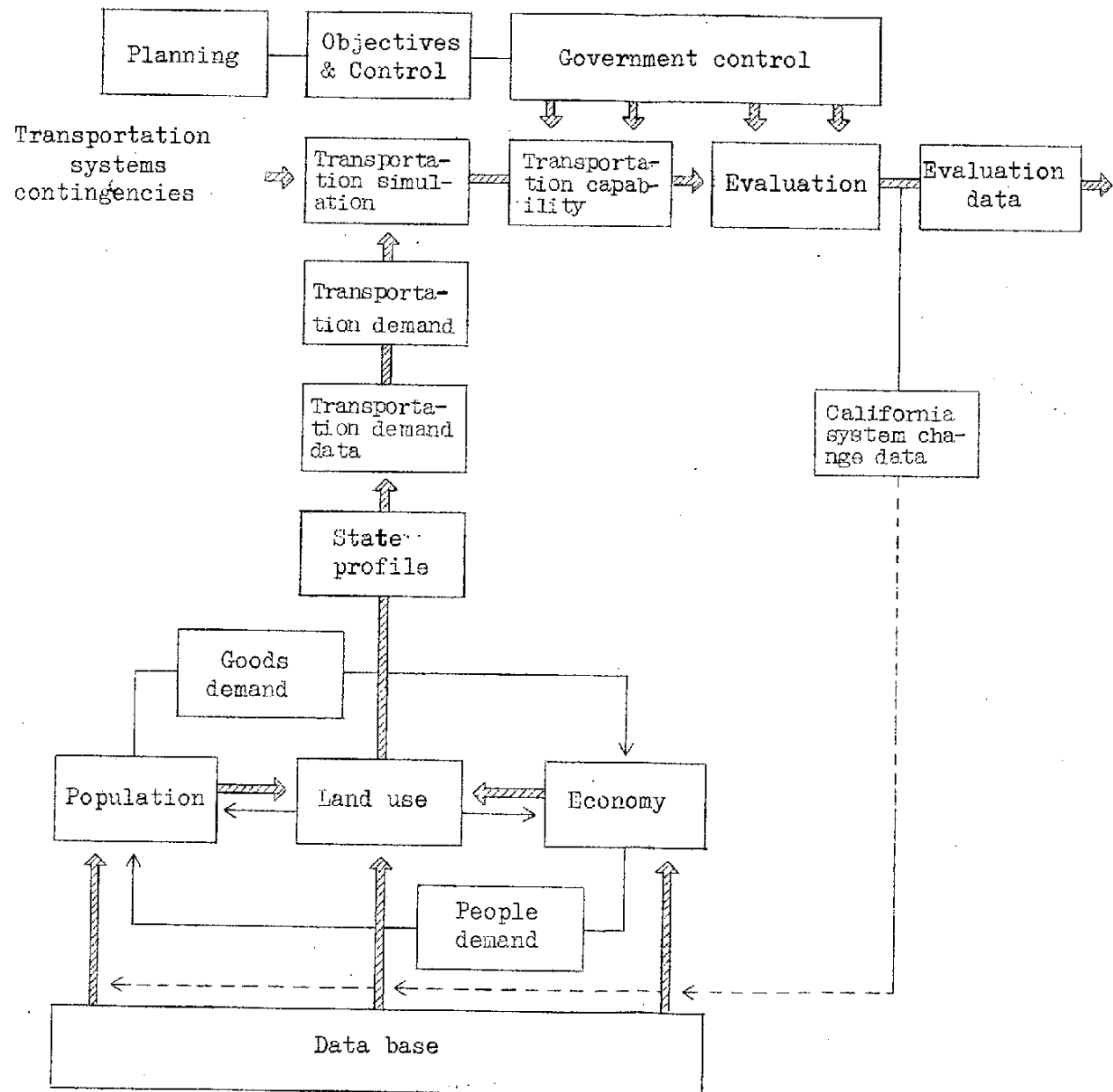


Figure 3. The California model.

3 FAMEの社会開発への応用

宇宙開発のような大規模なプロジェクトを計画どおり、効果的に管理していくためには、その過程で発生する膨大かつ多種類の情報を処理し、その情報のもつ意義を評価して、その結果を簡潔な形でマネジメントに伝達する情報システムが必要である。

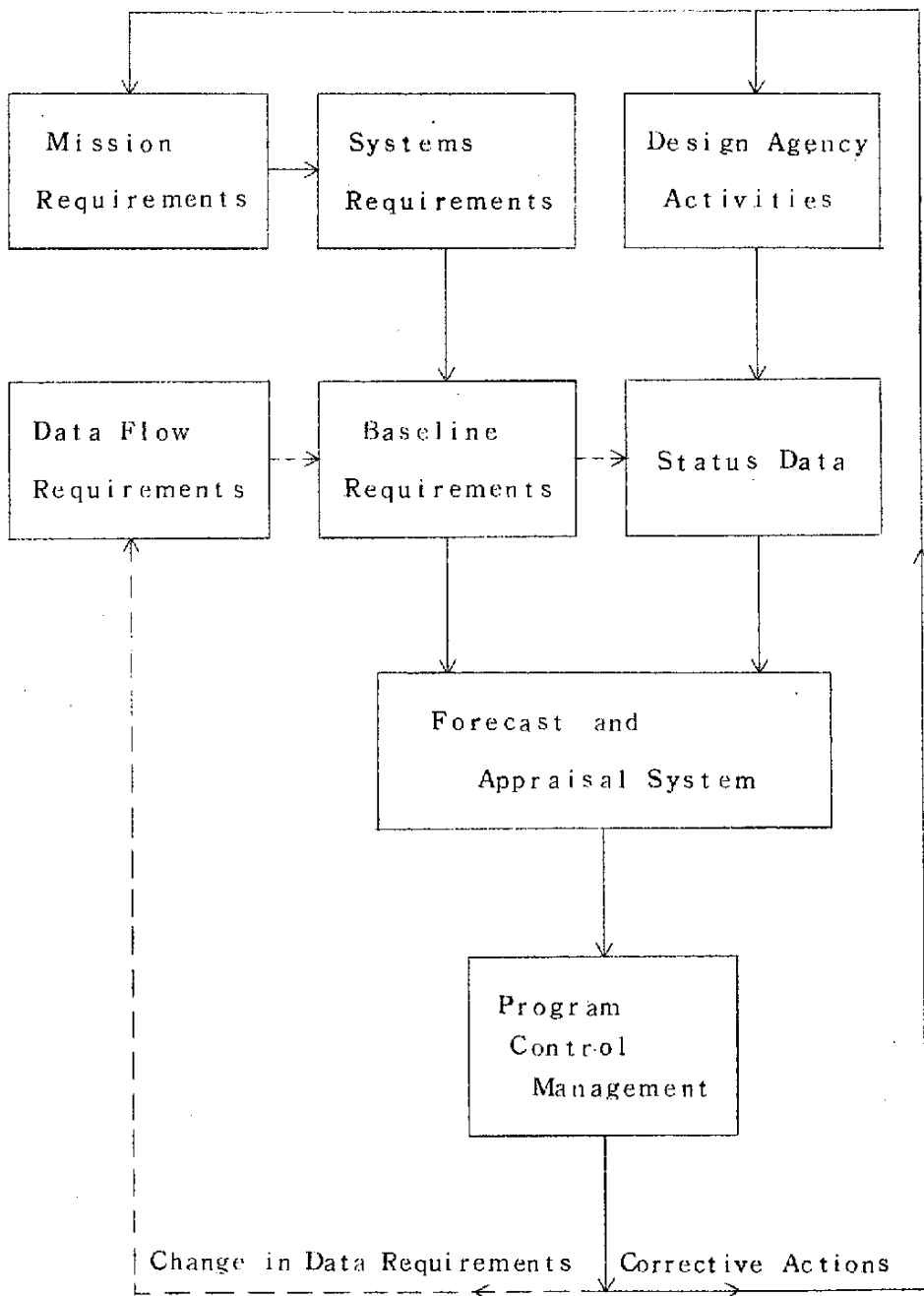
このような要求から生れたのがコンピューターを中心としたマネジメント・インフォメーション・システムであるが、NASAでは収集した情報をマネジメントの役に立つような形に加工するために数多くの技法を開発した。

FAMEは大規模なプロジェクトの管理に必要な計画の進行状態に関する予測を与えるもので、つまり、過去および現在の実績を評価して将来の傾向を予測するプロセスである。(次図参照) そのプロセスは、あらかじめ設定された要求条件と実績値とを比較分析し、その結果から今後の傾向を予測して、マネジメントに伝達する。FAMEではマネジメントに問題の所在を的確に認識させるために、分析と予測結果が正しく評価されるよう特に注意が払われている。

FAMEでは予測の手法として、次の4種類の科学モデルが用いられている。

- (1) 線形最大法
- (2) 非線形最大法
- (3) ロジスティク
- (4) フーリエ技巧法

数字モデルの適用によって求められた予測は、計画変更、計画の進行度(設計、開発、製造などの諸段階)に関連したファクタなどについて



Basic Organization and Operation of FAME

調整される。なお、マネジメントが効果的に評価予測できるようにFAMEではこの他の補足的な計算を行う。FAMEは最後に関係部門の意志決定者に対し、時間的データと予測データ等の出力を解り易い形で提供する。それには次のような内容のものが提供される。

- (1) プロジェクトの現在の状態
- (2) プロジェクトの将来の状態の予測
- (3) 問題領域
- (4) 問題領域の危険度
- (5) 問題領域の修正に必要なトレード・オフ

こういった出力機能を含むことによって、FAMEはマネジメントの情報と管理を目的とした一つの完全なシステムを構成する。このようにFAMEはダイナミックな変化に応じられるモデルであるので、宇宙船開発に伴う重量管理、コスト管理、日程管理、信頼性管理、性能評価などに適用できるとともに企業に於てはマーケットリサーチ、販売予測、コスト管理、品質管理、利益評価・分析・予測などに適用できる。さらに、最も適用が期待される分野として、公害、輸送、公衆衛生等公共的分野での適用が可能である。

第 2 章 情報の価格形成

調査の目的

調査の方法

1 分科会開催経過

1.1 昭和45年4月15日

本分科会の活動方針を、暫定的に、次のように決定した。

1. 情報の価格形成の純理論的考察の方は、今井委員に従来の研究¹⁾を更に発展、継続して頂く。

1) 日本経営情報開発協会、データバンク・シンポジウム
(データバンク研究資料 第1 P.195—215)

2. 本邦において実際に情報の価格決定が行われた事例についてケース・スタディを行う。

1.2 昭和45年5月20日

1.1の2の方針にもとづき、日本貿易振興会JETROから情報管理課長臼井孝氏においで頂き、事情を伺う。

1.3 昭和45年6月11日

同じく長期信用銀行事務管理部長伊沢勉氏、同調査役瀬尾道典氏においで頂き、事情を伺う。

- 1.4 その他、日本科学技術情報センターJTCSJからそのサービス料金に関する資料及び日本科学技術連盟JOSEからその物性定数推算プログラムAESOPPに関する資料を頂いている。

2 ケース・スタディの結果の概略

2.1 JETROの場合

1. 貿易に関する情報を、JETACと称する専用電子計算機によって検索し、販売している。その料金体系は次の通り。

基本料金

文字情報（世界6大洲，15文字を1件と数え），

1件 500円

数字情報（TSUSA 3桁を1件と数え）1件 500円

コピー料金

実 費

2. この価格の算定基礎は次の通りという。

JETACの導入費6000万円を，1ヶ月当り200時間，1時間当り検索5件，耐用10年として500円を算出。

$(6000 \text{万円} \div 120 \text{ヶ月} \div 200 \text{時間} \div 5 \text{件} = 500 \text{円})$

ただし，実際には1件当りの検索時間は平均40分という。

また，事前に業界30数社にアンケート調査を行い，検索費用は1件200～500円というのが一般に妥当と認められる線であったということも一つの根拠になっている。

3. JETACにインプットされるのは，世界各国の新聞33誌，雑誌550誌から年間23万件であるが，これら情報の蒐集費，インプット費用は上述の価格形成に算入されていない。
4. この業務による年間収入は，現在のところ，年間35万円に過ぎない。
5. 英国の商務省でも1件40円程度の費用で同様のサービスを行っているといわれ，JETROは基礎情報は公共財として社会に

提供し、それ以後のデータの加工は自由競争にまかせたという意見も出た。

2.2 J I C S T の場合

1. 情報検索サービスは昭和46年から開始の予定。
2. 現在は化学・化学工業に関する文献目録を年間36冊と、年間索引1冊を販売している。これは44年度は41,500円、45年度は48,000円という価格であるが、実際のコストからいえば、この3倍の価格にしなければひき合わない。

2.3 長期信用銀行の場合

1. 長銀では昭和40年以来、自行のために代表的企業および取引先をあわせ、計2200社につき、有価証券報告書、営業報告書などをソースとして、1社毎に260～270項目にわたる情報を、毎年計算機（IBM 360, model 40）に入れてきた。この間、項目の整備、インプットにおけるチェック、検索のプログラムも開発され、自行の用にあててきた。
2. 昭和44年になり、勧業銀行から依頼があり、このシステムを920万円で譲った。この場合は、価格の算定基礎は大略次の通りであった。現在このシステムを開発するとすれば、

開 発 費	3000万円
4年間のデータ・インプット費	3400万円
4年間のデータの価格	
（一年分を700万）	2800万円
	9200万円

を要するものとし、もともと販売を目的としたのではなく、本来、

自行使用のために開発されたものであるから、その $1/10$ を譲渡価格とした。

3. 以後は毎年データを700万の $1/10$ 、即ち、70万で譲り渡すことになっている。これは磁気テープの形で渡され、それをコピーして外部に出さないという条件がつけられており、また、データの分析用のプログラムはつけないことが原則になっている。
4. その後、東京銀行に対しても、全く同様の条件で契約譲渡している。
5. このデータは主として銀行にとってのみ有用なもので、その意味で、今井委員のいうクラブ財的性格をもっていることが指摘された。

2.4 AESOPPについては未だ直接に担当者から事情を伺う機会がなかった。

3 ケース・スタディにもとづく結論

- 3.1 いずれの機関においても、直接に採算をとる意図をもって情報が販売された訳ではない。JETROの場合も最終の検索に用する直接的経費だけしか考えられていないし、長銀の場合も、本来、その販売を意図してデータバンクを作ったのではなく、要した経費のある程度の部分が回収できればということで価格がきまっている。
- 3.2 現在のところ、未だ公開の情報市場というものは成立していないし、そういう形で価格が決定された例にはゆき当たっていない。
- 3.3 しかし、クラブ財として少数の当事者間の協議により価格が決められているにせよ、それには何らかの根拠が存在するはずである

う。ただ、個々の事例を通し、そこに普遍的な原則が認められるか否かは疑問である。現在までの事例においては3.1に述べたように、採算を度外視した価格形成が行われているという点だけが唯一の共通の原則である。

3.4 委員間の意見として、次の点が指摘された。

1. たとえ普遍的な原則が認められないにしても、現在の段階で、できるだけ多くの事例を記録しておくことは必要である。
2. 事例を前に各委員がそれぞれの立場から意見を出し合うところに、一つの生産的意義が認められる。未だそれによって一つの具体的な意見がまとまった訳ではないが、今後も1.1に述べた活動方針を継続し、徐々にまとめてゆく予定である。

4 情報の価値と価格形成をめぐる討論

1969年12月6日、7日に開催されたデータバンク・シンポジウムにおいてデータバンクに関する自由討論が行なわれたが、おのづから議論の集中したのは情報の価値と価格の問題であった。

- (1) データバンクの企業化が実現し、情報あるいはデータが販売提供されるようになった場合、むろん需給の経済原則によって価値は変動すると思うが、その価格をどのように決めたらよいかが、これら一つの大きな問題になる。

情報には迅速性が要求されるが、これを裏返せば時と共に情報の価値が下がるということである。しかしながら、その一方情報の価値は6年目位いまでは上昇カーブをたどる。そこで、上昇カーブにある間の情報の価格をどう決めるかなど、情報の販売価格をどのよ

うな基準で決めたらよいかという問題がある。

- (2) 民間機関が情報交換を行なえば必らず情報の価格をどうするかという問題が起こるであろうし、またそれ以前の問題として情報の価値ということが論ぜられねばならない。

情報というものは、一方では稀少価値が高いものが価値が高いという性質をもっており、個別的に接近すればするほど一般的でない情報が必要になる。ところがその一方では、それをオープンにして共同に利用する場合には共通の利用価値というものが必要になってくる。情報にはこのような相矛盾する2つの側面がある。

流通の側面ではあくまでオープンにして広く行き渡ることが望ましいのだが、もし生産者に生産者利潤みたいなものが全く認められないならば、創造的な情報を作ろうと努力する人がいなくなるおそれがある。しかし、作った情報を誰にも渡さないとなると、新しい情報のより広い利用には障害になる。こうした事は、情報の価値にまつわる一般的な問題だと思う。

財務情報にしても、現在数か所で個別的に開発しているが、作る側は他よりも独自性をもっていることを主張するし、利用者側はなるべくフリーに使いたがる。それに生のデータそのものという問題と、それに付加価値を与える分析手法の問題がからんでくる。価格の問題もこの辺の事情を考慮しないとすっきりした解決にはならないのではないか。

政府が集めて供給するデータにしても、必らずしも無償でなくてもよいのではないか。ユーザーに対し代価をとってもよいのではないだろうか。同様に企業の場合は開発費をどう扱うべきであろうか。

(3) 現段階で情報の価格形成を云々するのは時期尚早だという気がする。その前に先ず一般論を展開すべきである。データバンクの運営の仕方として、外国の例をも含めて2つに大別することができると思う。

1つはオープン・システムというもので、そこではプライシングの問題など考えずに、とにかく集めたものを誰にでもただで提供する。それに対し、クローズド・システムでは何か1つのチェックをおいて自由には出さないで条件をつける。そして、その条件の1形態がプライシングなのである。だから一般論としてはクローズド・システムの在り方というものが、どう成立するかといったところからつめて行く必要があるのではなからうか。

もちろん、データの種類を区分けして、このデータについては民間といえどもオープンにせよ、これから先は一定の条件をつけてもよいというようなことは出来ると思う。けれども元のデータは1つであり、企業内で作られたデータは別としても、外から入ったものをいくら加工しても、それに加工賃だけをつけても難しい問題が起きる。クローズとオープンの境界をどのように区分けするかというのが第1点であり、これを解決したら、クローズドの条件の側面としてはじめてプライシングの問題が出てくるのである。

またプライシングの問題にしても、保有と利用とが結びついている他の財とは違って、情報という財は、特定の人が利用した後もいくらでも他へ流動して行く。したがって、普通の経済学の原理にしたがってのプライシングでは少々難しいのではなからうか。

具体的な価格付けをしようとしても、最初は試行錯誤でやらざる

を得ないと思う。しかし、それは相当変動すると思うし、しかも一方にはオープンのものであるとなるとその動きは予想もつかない。プライシングの問題一つを取りあげても私はそう早く安定したデータバンク企業が実現するとは考えない。

- (4) 当然価値が議論されてから価格に入るべきであり、稀少価値があるときは価格は高くなるであろう。

しかし、その価値が他の財と違うから、買ってからの機密保護が問題になる。AからBへ売るときには当然そういう条件がつく訳であり、現にスタンダード・プア社は他へ転売しないという条件をつけている。

- (5) 戸籍謄本や何かの証明を貰う時と同様に、政府のデータバンクもコンピュータを動かせば経費がかかるのであるから、当然手数料のようなものをとってもよいと考える。

しかし、それはコンピュータ経費であって、情報の価値から生ずる価格ではない。そうしたプラス・アルファの方は考えやすいと思う。例えば政府統計にしても刊行物センターへ行けば売ってくれるが、あの値段も統計情報の価格ではなく、媒体の価格なのである。それと同様にデータバンクの場合も、加工賃そのものはプラスすることができるが、元の情報の価格となると理論的にはなかなか解決はむづかしい。

- (6) 価値の問題は需要者が決めるという感じがする。同じ公共財でも天気予報などは客観的な財なので、仮りに流用するとしても比較的問題なく、一定の価格なりあるいはただ同然のものとして利用できるのではないか。もっとも、ヒマラヤの山に登るときに利用するよ

うな、かなり地理的な形でサービスをしている会社もある。

そういう会社の場合には、山へ登るときに天気がどうかという問題であり、長期間の時系列をみて相関係数を計算するなど、相当の知識と加工費を要するから代価をとるのは当然である。一方テレビの天気予報などは、国へ税金を払ってみんなが平等に貰うのだが、その点ははっきりしているのではないか。

- (7) 判例のデータバンクにしても、外国の判例については大学の外国文献センターへ行って頼めばただで教えてくれる。そういうものが一方にあるとき、他方に企業化されたものができたとしたら、どうなるか問題である。

企業化された場合、地方にいてもサービスを受けられるとか、もっときめの細かいサービスが受けられるというので結構利用者はいると思う。

アメリカでも判例データバンクは企業としては難しいそうだ。セントラル・コンピュータのファイル能力を越えるテンポで判例が増えつつあり、経営的に非常に難しい。一方ではそういうものを非常に欲するムードがあるにも拘らず、経営的に成り立たないというのは、価格体系とは別に問題があるような気がする。コンピュータの容量、通信回線あるいは利用者の増加で将来解消する問題かも知れない。

- (8) 情報の値段について具体的に研究しているが、公共データバンクの場合でも、検索方法に特別な注文がついた場合は、政府刊行物が内容については金をとらぬが印刷について金をとると同様に、オペレーショナルなものについて、オプションの受益者負担の形で費

用をとるべきであろう。

公共財としてのデータを貰い、自分の創意工夫で加工編集して売り出そうとしている者もあるが、これは内容に顧客を惹きつけるような加工をほどこして価値を高めているよい例であり、これだけは顧客が値段を決めているような感じがする。

一番問題なのは、門外不出の貴重な情報がしまい込まれてしまうことである。これは国家経済全体からみてマイナスであり、そうしたものが取引きされるように価値付けを考えて行かねばならない。

Know-How のように取引きの材料になるという通念が確立されれば、その限りでしまい込まれることはなくなるであろう。これは国策的に考えるべきであろう。

- (9) 企業においては付帯サービスとして情報の蓄積を行っており、コストに見合うならオープンにしてもよいというのが実態である。しかしもう一步進んで、いまのままでやって行けるかどうかは問題であると思う。

一般業務をやっている中に蓄積されたのだから原材料がただであると考えるのは問題である。データ・ファイルに使用された人件費やコンピュータ時間を全部計算したとしても、それだけでは本当のコストとは言えない。前期までの効用や、再生産価値などを加えると随分高いものになると思う。

企業の中でオリジナルに作られた資料は特にそうだが、官庁統計にしても官庁毎に地方の人件費まで入れて原価計算をしてみると、みな単価が違って来る。それを全てただで提供しているのだが、その間の差をどうして埋めるか、そしてそれに再生産のことまで考え

合わせると非常に高くつくということになる。そうしたものに加工費はとるとしても、一方ではオープンにしておいて、他方では企業化されたものが出来るとその間のギャップをどうして埋めるかが問題になろう。

- ⑩ 商売を目的としてデータバンクをはじめたら恐らくコスト倒れになる。しかも集めたデータも古くなればどんどん捨てなければならぬから、1万件の中から検索するのと80万件の中から検索するのでは価値が違ってくる。1度も検索されずに捨てられるデータのコストも計算に入れていたらとても商売できるような価格ではデータを売ることはできなくなる。

それであるから、日常の個有業務があってその処理の副産物としてデータが蓄積されている。そしてそのデータをどのように評価するかというように機能の面から整理しないとデータの価格は出てこないように思われる。

第3章 情報ネット・ワーク問題

データバンク（以下DBと略称する）は、或るCommunityにおいて何らかのデータの発生があり、またそのデータの利用に対する需要があるとき、その間を効率よく結ぶ「働らき」として捉えられよう。この「働らき」が、如何なるデータに対して（例えば個人の信用データ、株価データ、各道路における交通量データ、各地の気象データ etc.）、如何なるCommunityにおいて（例えば米国、日本等の国を単位とするもの、証券業界のような業界を単位とするもの、特定の企業におけるMISのように一つの組織体を単位とするもの etc.）現在如何なる機関が、如何なる組織構成でそれを行なっているかを分析したり、近い将来に実現すべき理想形態は如何にあるべきかを設計したりすることは、他のワーキング・グループ（以下WGと略称する）においてなされるであろう。

WG第3は、この「働らき」をモデルの中に記述し、主としてIRとネット・ワーク構成の技術的問題を解明する方向で検討を行なった。昭和44年度は、先ずWGメンバーの間のコンセンサスを得る必要性和時間的制約も考慮し、この方向づけの下に、「先ず何を考えねばならぬか？」というレベルでの問題提起と整理を行なった。

1 需要指向型情報流通とデータバンクの基本的機能

データの発生からその利用の間を効率よく結ぶ機構は、情報の流通機構として捉えることができる。（商品の供給から消費の間を効率よく結ぶのが、商品の流通機構として捉えられるように）。商品の流通機構において、供給力が消費力に比して弱いとき、或いは消費からのフィード・バックによって供給が制御されないとき、供給指向型の流通と

なる。それに対し、消費者の要求の多様性を満たすだけの供給力があるとき、そして、この消費者の選択性向のフィード・バックによって供給のレギュレーションが可能になるとき、流通は消費者指向的となる。

情報の流通は、高速多量のデータ処理装置や通信技術の開発と普及によって漸次その様相を変えつつあるとはいえ、まだ「高度選択社会」というイメージには程遠い段階である。例えば報道・出版等のマス・コミュニケーションは、視聴率や購読部数等によって情報利用者のフィード・バックを受けているとはいえ、「企画のための判断材料」ではあっても、「需要によって供給が起る」という形の直接のレギュレーションはない。学術的情報に対しては、主として科学技術分野におけるドキュメンテーション・サービスを除いては、同業種（例えば学会を単位とする同じ学術分野の研究者間や、協会を単位とする同じ産業分野の企業間）の間でのコミュニケーションか、或いは自給自足型の情報収集に頼っているのが主流である。

時間的にも空間的にもまたそれを発生させる意図においてもランダムに発生している情報（すなわち、世界中のどこで、如何なる機関に属している人が何を目的にして、いつそれを発表したものであっても）の中で、それを必要とする人がいたならば、必要とした時に、その必要とする情報のみを、利用者にとって許容できる時間間隔の中で、必要とする詳しきで、利用者の望む媒体（印書、音声、画像、マシン・リーダブルな形態等）によって、利用者が適正と判断するコストによって入手できることを、非常に多くの利用者に対し保証する機構が、これからの社会では必要であり、その要求は漸次顕在化しつつある。

このような情報要求を満たすためには、発生した情報を集めてファ

イルを構成し、利用者の要求を満たす情報がファイルの中にあるとき、それを供給するシステムが必要である。このランダムに発生する情報をファイルの中に収め、需要が起ったときそれを供給する機能を、DBの基本的機能と考えるとき、DBは、需要指向型の情報流通機構と考えることができる。しかし、この流通機構は、情報発生機構そのもののレギュレーションを基本的課題とはしていない。情報需要に基づく、流通過程の効率化であり、発生している情報の中に需要を満たすものがないとき、その需要に対して流通機構は無力である。この需要による発生機構へのレギュレーションを直接行なうシステムの、1つのモデルが所謂シンク・タンクであろう。しかし、このシンク・タンク（以下TTと略称する）機能が効率よく働らくには、必要とする素材情報が必ず入手できる機構を必要とする。その意味でDBはTTの前提条件であるといえることができる。

2 データバンクにおけるデータの範囲

以上のようなDBの基本的機能の理解の下に、その機能モデルを技術的側面から記述するためにはデータの範囲を想定する必要がある。

データの属する知的生産の分野（物理学、化学、電気工学等々の所謂学術分野のみでなく、基礎研究、応用研究のような研究活動の諸段階、経営企画、製品企画、マーケティング等々の種々の知的活動の場面等、広く扱った場面における）の差違は、分類・索引等の技術の細目においてその特殊性が表われるが、これらはモデル記述においてはパラメータの差に表わされる程度であり、システム・モデル構成の際、代替手法を決定する大きい要因とはならない。

これに対し、データが数値データのみか、画像・音声等のアナログ・データを含むかは、用いるハード・ウェア、通信路の性格を変え、更に、開発すべき技術に対する Specification にも言及する必要があるので、技術的には大きい問題である。その意味で、データの範囲を想定する必要がある。

WG Ⅲ では、DB におけるデータを次のように極めて広く解釈することとした。

(1) 数値データ

項目名（或いはその数値が対象とするものの名前）、数値、その数値に固有の情報（発表者名、発表年月日、発表地等）によって構成されるデータ。この「数値」の中には四則演算のみでなく、何らかの定義をされた記号の「値」を含む。例えば加入者番号などは、「加入者」という集合の要素を表わす記号の値である。又人名、各機関名も大小比較（配列順序）一致の判定が可能な記号で表現されているかぎりと同様である。物理的・化学的常数、統計データ、諸種の観測データ、信用データ等。

(2) センテンシアル・データ

日常言語による文、およびそれらと一義的な変換アルゴリズムで結ばれた文、又は、特定の文法規則、意味規則を持った記号系における文によって表現された情報で、何らかの記録媒体の上に人間又は機械によって解読可能な形に記録されたもの。学術的論文、事実の報道等。

(3) アナログ・データ

再現および伝送可能な形に記録された平面像（写真，絵画，設計図，等の図面等），立体像（ホログラム等による），色彩，明暗の階調，音声等で表現できる情報。

3 データバンクの性格に関する概念的把握

DBは，その運営主体と，その情報供給対象によって，その各成分機能のあり方に対する制限条件が大きく異なる。

(1) 国家機関（行政・司法・立法の各部門における），又は地方公共団

体がその中で発生する情報を一般に供給するために設置する場合

この場合は，データは，その属する機関のオペレーショナル・ワークの中で多量に発生するし，又，特定のプロジェクト遂行のために集められる。従来は，そのプロジェクトやオペレーショナル・ワークのために用いられ，他の用途への考慮が殆どされなかった。これをすべての国民或いは企業体において利用できるように供給することがこのDBの目的である。このことから

(i) DBにデータを入力し，供給可能なワーク・ファイルを構成するに当って，データ内容の歪みを極力避けなければならない。

(ii) 情報供給の対象を限定してはならない。

(iii) データの発生とDBへの入力が（公開可能か否かのスラリーニングはあっても）直接的でなければならない。そのため情報需要からのフィード・バックによるレギュレーションが弱くなることは或程度はやむを得ないが，それを補うために，この運営主体は「公開可能なデータはすべての国民の共通の財産であり，国民

はこれを要求する権利があり、DBはそれを供給する責任がある」ことを、運営の理念として掲げるべきである。

- (iv) このDBの財政基盤は、その所属機関による資本投下と維持予算によって運営されるべきである。
- (2) 公共企業体又は公共性の強い民間企業体が、その中で発生するデータを供給するために設置したもの
これは(1)に準ずる。
- (3) 国又は地方公共団体が、知的生産活動の振興、効率化のために設立するDB

例えば科学技術情報における日本科学技術情報センター(JICS T)などがその代表的なものである。その性格からみて(i-i, ii)はそのまま適用される。また財政基盤も、社会資本としての性格が強いため、サービス・コストは、受益者が負担するとしても、資本投下と維持費の一部は、公共的支出がなされるべきである。従来この種の機関は、監督機関のレギュレーションが強く働らく傾向があるため、利用者側のレギュレーションを強化する制度が必要である。

- (4) 研究機関、民間団体、時には民間企業体などが、閉じたサークルの中だけでのデータ利用のために設けるDB

これは、例えば或る大学の研究室において学術的データを集め、同じテーマの研究者に配布したり、或る企業で系列会社に対してのみ情報を提供するような場合である。これをその閉じた利用者以外での利用を考え、より強化する必要があるときは、公共的な財政援助か、制度的バック・アップが必要である。このような公共化の際、

(1)の各項目が大体において適用できるが、公共的援助に際して、その本来の活動にネガティブ・レギュレーションを絶対に行なってはならない。

(5) 営利企業としてのDB

これは、全く自由な経営が許されるべきである。自由競争と利用者の選好の中で、みずからの方向を求める性格のものである。

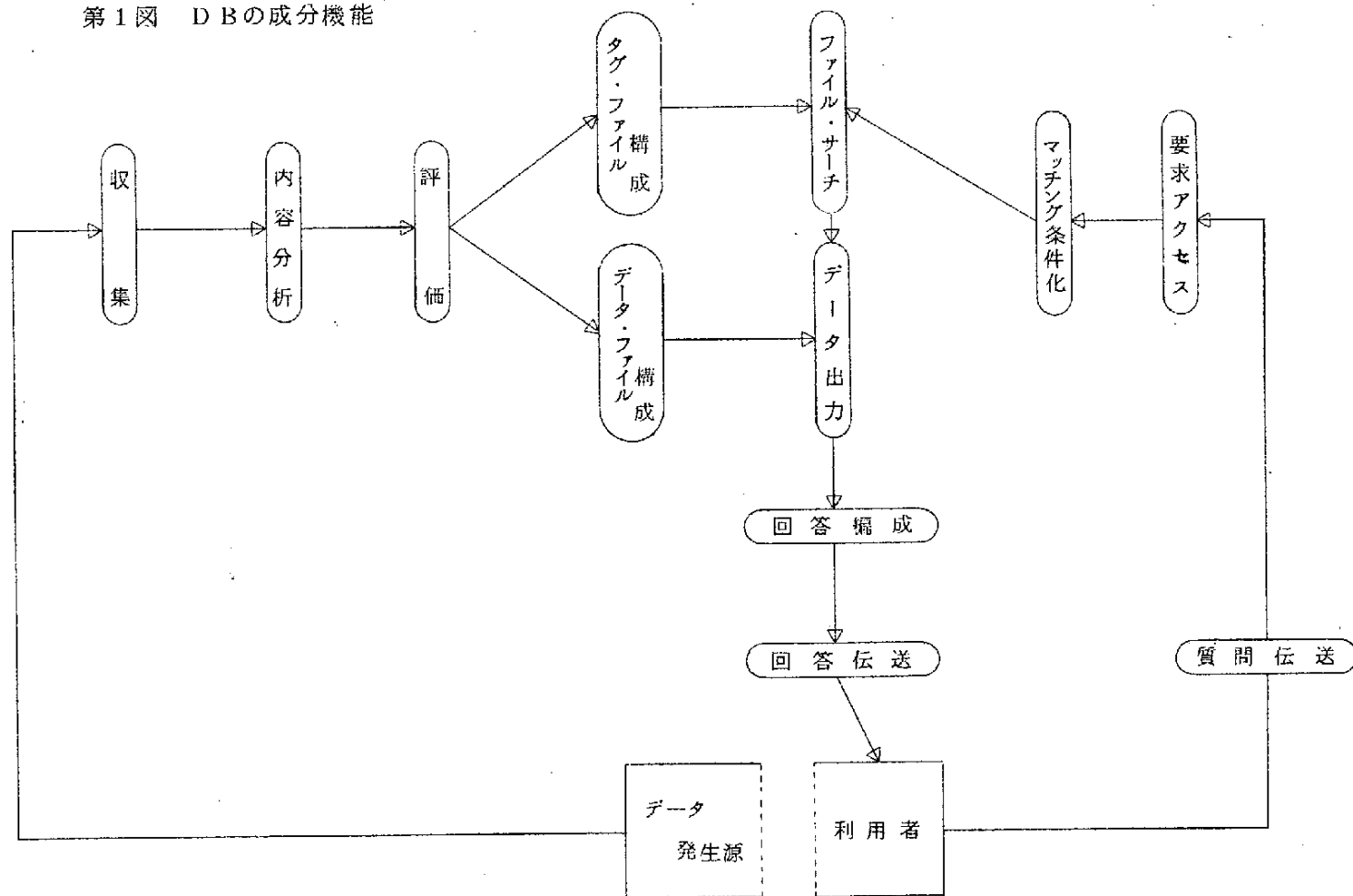
4 データバンクを構成する成分機能

DBを象徴するものはデータ・ファイルである。このデータ・ファイルは、データの分析技術とファイル構成技術に何を用いるかによって特性づけられる。その技術の適用は、データの種類と内容によって、その選び方の範囲が限定されるし、一つのDBがカバーするデータの種類、内容は、その対象とする利用者の要求によって、その範囲が限定される。

このため、あらゆるデータを一つのファイルに収めることは不可能であるし、又この世の中の情報需要を満たすためには多くのDBが存在しなければならない。すなわち、一つのDBの中にもいくつかのファイルが存在し得るし、利用者は、一つのDBの中のどのファイルも、また、世の中に存在するすべてのDBにも、容易にアクセスできることが必要である。

これはファイル間ネット・ワークを考慮すべき事を示しているが、その問題の前に一つのファイルを構成するための成分機能のつながりの観点をまず明確にしよう。

第1図 DBの成分機能



(1) データ収集

データの発生機構は、一応DBの外の機能であるとすれば、最初の成分は、データを集める機能である。

- (i) 技術的にはデータ発生即DBへの入力の方法が最も良い。特に、オペレーショナル・ワークに伴うデータの収集において有効である。しかし、この場合の問題点は、各データ発生源間の参加意志をまとめることと、データ・フォーマットの統一である。これについては、先ず行政機関においてその範を示すべきであろう。

又、データ発生機構の中でも、発生させたデータの流通に責任があること、流通過程にのせるために、データの発表者は何をせねばならぬかについて、広いコンセンサスを確立すべきである。

- (ii) 発生即入力とならない場合、DBが成立するか否かがこの収集機能にかかっている。収集ストラテジーへの利用者需要からのレギュレーションを有効に行なう仕組みの解明と、データ源の検索技術の開発がこの場面での課題である。

又、データ収集は労力コストを伴う仕事であることの認識が意外と低いことも、わが国においては問題である。収集コストがそれ以降のプロセスのコストとコンパラブルになるケースは多いし、数千人というマン・パワーを必要とするケースも多いことをDBの関係者はよく認識すべきである。

- (iii) DBのデータ源と利用者が全く異なる集合と考える場合と、同じ集合の元が或るときはデータ源となり、或るときは利用者となると考える場合では、DBモデルが大きく変わる。後者はクラブ

財的な性格を持つデータについて最も良くモデル化できるし、或意味では、この利用者でありデータ源である集合が構成するメディアム・コミュニケーションで結ばれる社会のモデル化である。今後の社会開発や、市民生活を考える上で、このようなメディ・コミ社会が大きい役割を果たす可能性が考えられる。そのメディ・コミを可能にする機能がDBの中にあることは充分注目に値する。

(2) データ内容の分析

データの内容分析技術は、まだ経験的技能の域を出ず、一般的な基準は定性的にも確立されていないのが現状である。この面で最も経験の豊富なドキュメンテーションにおいても、技術として解明する初歩的な段階である。

一般的に言えば、一つのデータの特徴を表現する項目数が限定されていて、各項目の名前で指定されている内容にあいまい性のないとき（例えばスキル・インベントリーのファイルにおいて、性別、学歴、住所等）内容分析は容易であり、ファイルの信頼性は高い。しかし、このようなデータは極めて限られたものである。数値データにおいても、統計データを例にとると、層別の方法、サンプリングの方法、検定の方法などが明確にされないと、そのデータの信頼性が得られない。又アナログ・データにおいても、例えば指紋をとっても現在は流れ方、円のかき方をいくつかに分類してデータ化しているにすぎない。（すべて人間の作業である）。センテシアル・データの場合、全く経験の中で何となく内容を分析しているにすぎない。

これには、データを構成する記号の記号論的 Semiotical な基礎研究を必要とし、少し込み入ったデータになると、人間の思考過程のモデル化をより伴うものである。

これに対して現在提言できることは、このような基礎研究を、より有効に加速するための研究開発のプロジェクト化と研究投資を行なうことと、各所で相互の連絡がなく実際に行なっている分析担当者の中で、経験的に行なっていることの整理を協力して行なうアクションを、何らかの形でとるべきであるということである。

以上の技術論的な難点と離れると、実際の問題として、サービス対象とする利用者の要求の範囲を限定すると（すなわち分析基準を限られた範囲で明確化すると）分析は容易である。この事は § 3 の DB の性格づけにおいて述べた(4), (5)の場合は比較的容易である。この場合も、その分析基準を、できるだけ明確に公表すべきである。

(3) データの評価

データの評価には2つの面がある。一つはデータ発生源および入手経路に対する評価と、データそのものが利用者にとって価値があるか否かの評価である。

前者については、過去のデータ発生・入手の頻度と利用者の利用頻度の統計的ランクづけによって、一応高い近似で基準を求めることができる。しかし後者の客観的評価基準を求めることはきわめて困難である。それは一般的に言えば、この評価基準は利用者によって異なり、その統計的性格を求める手がかりがまだつかめていないからである。

内容分析の基準が精密に定められ、その内容を表現する記号系の

記号特性が構造的に明らかになったとき、恐らくは統計モデルが画けるのであろう。

§ 3 で述べた DB の性格の中、(1)～(3)においては、原則としてデータ源・入手経路の評価に止め、内容の評価はむしろ行なうべきではない。それは、利用者の要求と離れた、勝手な取捨選択が行なわれ、データ・ファイルに歪みが生ずる危険性を伴うからである。(4)、特に(5)においては、その任意に定める基準によって評価することは自由である。それが適切でないときは利用者からのレギュレーションを強く受けるであろうからである。

(4) タグ・ファイルの構成

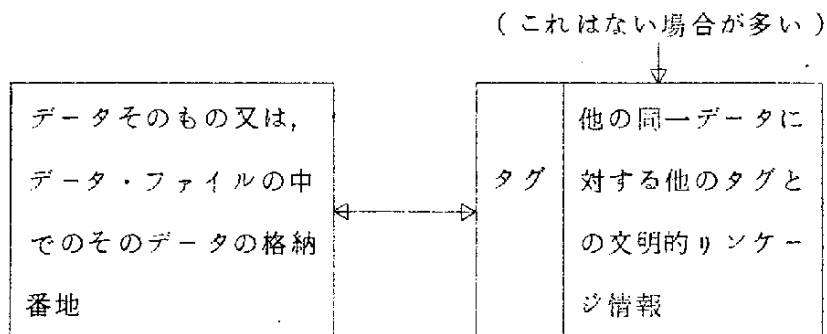
データの種類にもよるが、DB のファイルは、一般的にはデータそのもののファイルと、利用者の要求をどのデータが満たすかを判定するためのファイルとは別の性格を持つ。これを全く別のファイルとする場合と、同じレコードの中に（勿論定まったレコード・フォーマットによって）両者を入れる場合があるが、ファイル容量の大きい場合は、別のファイルに収める方が一般にはスペース効率および時間効率がよい。この後者の方に対する名称は、ドキュメンテーションにおいては2次情報ファイル、2次データファイル、検索用ファイル等と呼ぶが、今ここでは、マッチングのためのタグのファイルと呼ぶこととする。（この名称も研究会の進行とともに専門語の統一をはかる中で定着させる必要がある。）

タグ・ファイルの構成技術は2つの面をもつ。1つは内容分析、および評価の結果を、如何なる特性を持つ記号系によって表現するか？であり、もう1つは記号によって表現されたタグを、マッチン

グ効率が最もよい（主として時間効率）ように如何なるファイル構造に納めるか？である。

タグ記号については、シンボジウムで中井が報告したいいくつかのものがある。しかし、センテシアル・データに対しては、記号論的研究がまだ不充分であるため、信頼性のある記号系がまだ得られていない。アナログ・データについては、それを人間が分析してデジタル・データに変換するのが一般的である。これに対し、パターン認識の技術の実用化の試みをもっとなされるべきであろう。

ファイルのレコード構成は、



この対応関係をつける構造をもっている。1つのデータについて1つのレコードを構成し、その中にすべてのタグが記入されているとき、ファイルはシーケンシャル・ファイルとなる。逆に1つのタグの下にレコードを構成し、そのタグをもつデータのアドレスが記入されているとき、ファイルはランダム・ファイルの性格を持つ。各レコードのファイルの中での番地の割りつけは、待ち行列の計算を用いたりして、種々の工夫がなされている。

このファイル構成におけるソフト上での1つの問題は、ファイル

・プロテクションであり、もう1つはその維持更新である。又ハード上での問題は、安定したメモリーの問題である。

DBの性格によっては、特定の人しか利用できないデータを、他のデータと共にファイルするとき、他人からは引き出せないことを信頼度100%で保証しなければならない。これは、呼び出し者の判別コードを如何に構成するかの問題と、DB運営における機密保持体系の問題である。維持更新における最も大きい問題は、ファイル容量に制限のあるとき、ワーキング・ファイルを如何にコンパクトに納めるか？の問題である。一般にデータは、その利用頻度が、大体 $e^{-\lambda t}$ の指数函数的減衰をすることが知られている。ファイルの各レコードに対し、利用頻度をモニターしていて、一定基準を常に下回るとき、ワーキング・ファイルから外して Rare Use File とでも呼ぶべきものに移すオペレーションが必要である。更に、この Rare Use File のみを持つ公共的なDBを設けることも、充分考慮すべきである。

メモリーの問題は、実は深刻である。現在、Machine-Readableなレコードの記録媒体の主体は磁性体であり、安定性がない。あくまで、磁性体は中間的記録媒体と考えるべきで、長期間、多量のデータを、高速にとり出すための、高密度の安定したランダム・メモリーを実用化できるか否かが、DBの成立の1つの条件ですらあると言える。

(5) データ・ファイルの構成

データ・ファイルは、データの種類によってそれぞれ異なるものである。数値データの場合は、一般的に言って Machine-Readable

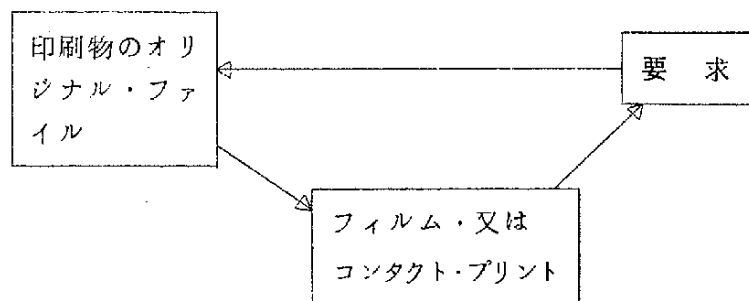
な形の方が有効であり、又可能である。この場合は、タグ・ファイルで述べたことが大体そのまま当てはめられる。

センテシアル・データの場合は多くの可能性がある。

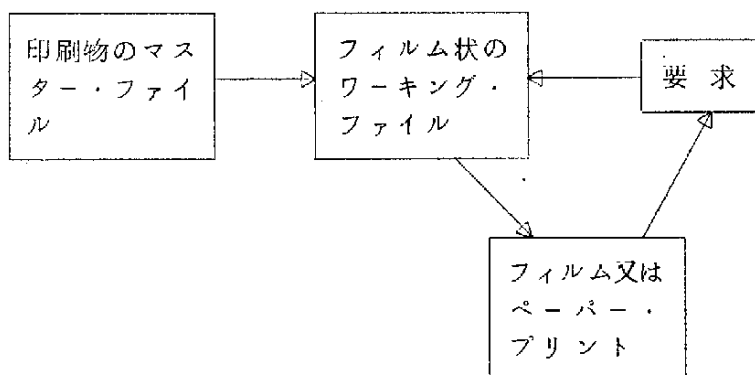
記 録 媒 体	格 納 形 式	対 象
印 刷 物	書 架 グァーティカル・ファイル	冊 子 体 小冊子、リーフ状のもの
フイルム (縮率 1/10 ~ 1/30)	検索コード付ロール・フィルム ストリップ・フィルム	図 書、カタログ、人事記録、 データ・シート 主として欧文、数枚のシート
フイルム (縮率 1/100 ~ 1/300)	フィッシュ・ジャケット (30P ~ 60P/枚) シート・フィルム (2000P ~ 4000P/枚)	同 上 図 書、多量のデータ表、多量 の図・表
Machin-Readable	MT, MD, MK...	○Fact-retrival 用データ ○入力コストがパフォーマンスからみて許されるもの ○OCRその他機械的入力の可能性なもの

しかし、一般には、

(i)

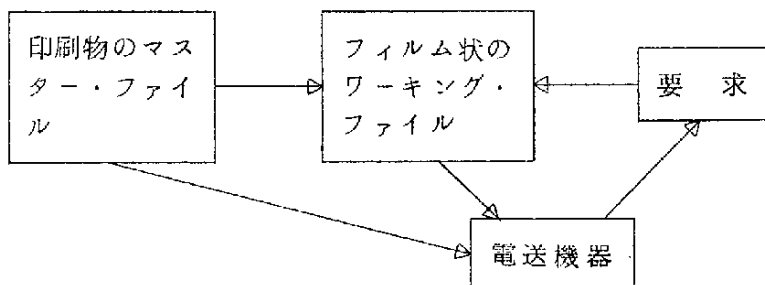


(ii)



が最も実用的であり、近い将来、極めて特殊なケースにおいて

(iii)



が実用される程度である。

5 要求のアクセスとマッチングの条件化

利用者の情報要求には、2つの型がある。ドキュメンテーションにおいては、それをカレント・アウェアネスとレトロスペクティブ・サーチと呼ぶ。カレント・アウェアネスは、刻々と発生するデータの中で、利用者の需要に relevant なものを入手しようというものであり、レトロスペクティブ・サーチは、利用者側で需要が起きたとき、過去をも含めて現存するデータの中から、relevant なものを入手しようというものである。

カレント・アウェアネス系の要求に対しては、利用者の自由なセレクションが許されるマルチ・チャネルの伝送系による発信指向的な流通システムが有効である。このセレクションを利用者側で行なう場合と、個々の利用者の需要をあらかじめシステムが受取っておいて（利用者毎の検索項目のプロファイルを造っておいて）システムがセレクションを行なう場合の、2つのケースがある。この利用者セレクションの例は、新聞やラジオ・TVのニュース、抄録雑誌、索引雑誌、等である。好みの新聞や抄録誌を購読して、その中から relevant なものを探したり、好みの放送局にダイヤルを合わせ、その中から relevant なニュースを聞く場合である。システム・セレクションの場合は、最近では S D I (selective dissemination of information の略) と呼ばれ、漸次普及しつつある。

レトロスペクティブ系の要求には（ワーク・ファイルとレア・コース・ファイルの区別は必要であるが）原則として刻々と追加され削除は許されないファイルが必要である。そのファイルからデータが流れるのは、利用者側の要求が起ったことによるのであり、本質

的には利用者指向的な流通システムである。この場合も、図書館の目録カードや、利用者が自由に図書を探すことのできる、Open-stech の書庫のような利用者セレクションのシステムもあり得るが、主流は利用者の要求をうけて、その内容を検索条件に翻訳し、システムが検索するシステム・セレクションである。

システム・セレクションの場合、要求の受けつけと、その内容を分析して検索条件に変換するプロセスが技術的問題となる。それを次表にまとめておく。

6 ファイル・サーチ

検索コードをガイドとして、それを意味的に含むが構造的に含むタグを持つデータを探すプロセスである。これは人間が行なう場合と機械が行なう場合では本質的に違うものがあるようである。

人間がこれを行なうためのタグ・ファイルは、最終的にはシーケンシャル・サーチを行なうものであることを前提とし、そのシーケンシャル・サーチを行なう範囲を限定するために、タグによるセグメンテーションを行なうという考え方が基本にある。その、最終的なシーケンシャル・サーチが必要な理由は、

1. 人間の情報要求を分析して追いつめて行っても、最後まである範囲の蓋然性を持つ。
2. データに目を通していている中に、分析し出せなかった要求に気づくことがある。
3. 当初は不必要だと思っていたデータによって、新たな情報要求を起す。

検索条件化		利 用 者 自 身		シ ス テ ム		
アクセス方法		1 W a y	2 W a y	分析スタッフ	自 動 化	
オ ー ラ ル	面 接			現状の技術では最も望ましいが、利用者にとってシステムとの距離に制限がある。		
	電 話	エラーの発生率が大きく相互のチェックが困難 (註1) 単純なデータ向き		利用者の要求内容が複雑なとき、分析者と利用者の会話の精度に問題がある。		
記 録 可 能	文 書	(1) 利用者の記述(コーディング)エラーの危険性がある。(註2) (2) 郵送時間、誤配の面で不利		(1) } 同左 (2) }	(1) } 同左 (2) }	
	加入 電信	(1) 同上 (2) オン・ライン化に問題がある。		(1) } 同左 (2) }	(1) } 同左 (2) }	
	データ通信	(1) 同上 (2) 回線自由化・使用料の面での問題がある (3) オン・ライン化には管理プログラムの開発が必要	(4) ディスプレー装置などのマン・マシン・インターフェースに問題がある。 (5) ソフト開発・システム維持のコストが膨大である。	(1) } 同左 (2) }	(1) } (2) } 同左 (3) } (4) } (5) }	
		これは利用者が自分の要求を自分で分析して、検索コードに翻訳してシステムに送って来ると、その検索コードでシステムがセレクションを行なうものである。このためには、 (1) 要求分析と検索コード化のマニュアルを完備すること。 (2) 利用者がその使用に習熟するか、その専門家を雇用する必要がある。	これは利用者とシステムのモニターとの間でのオン・ラインの会話によって、望むデータに接近するものである。IR技術としては、1つの理想であるが、現在の技術ではコストをあまり意識しないでより限られた場面で実施が可能である。	これは利用者の要求をシステムの専門スタッフが分析して、検索コードに変換してセレクションを行なうものである。この方式での一般的問題は、 (1) 分析スタッフの訓練と人員確保の問題点。 (2) 分析パターンの理論的解明にある。	利用者が日常言語で、その内容を記述したものをシステムで自動的に分析し、検索コードに変換するものであるが、これには言語と意味の問題の基礎的研究を必要とするので、ここ5年以内の実用は困難であろう。	(註1) J I C S T の複写サービスでこの方法を採用していた時期があるが、検索データの受け渡しエラーが大きく相互に確認することが困難なため廃止した。 このチェック機構は、厳密に行なうと面倒である。 (註2) J I C S T の複写申込において約5%は申込者の書誌データの記入エラーがある。

という人間の持つソフトでフレキシブルな性質によるものである。
このために「流し読み」Bransingや、会話形式の検索システムを必要とする。またこれが、検索効率を定量化するにあたって、

" relevancy " の規定が非常にむづかしい理由でもある。(「情報を入手することによって自分が変わり、その変わったことによって自分自身を1つの成分として含む全体に影響を与え、全体が変わることによってその成分である自分に再び影響を受ける」ということの起る系を記述することは、極めて困難である。それが人間を含む系のシステム概念を困難としているし、予測、評価の問題を困難にしているし、組織の概念や、タグ記号の構造の解明をも困難にしている。この問題は後で再び触れることにする。)

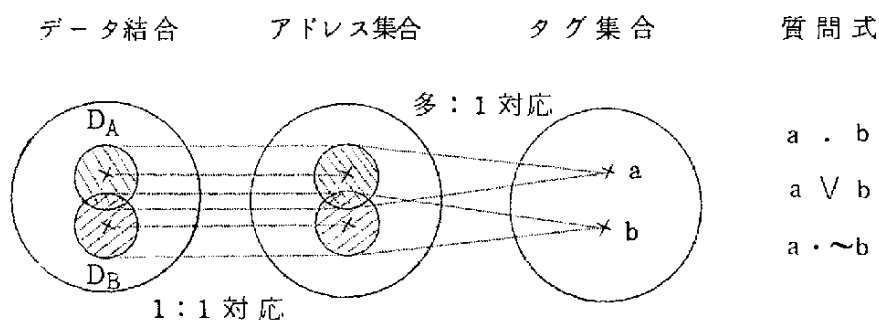
この事から単純な内容のデータは別として、少しこみ入った内容のデータを扱うあらゆる検索システムにおいては、その検索結果を人間がもう1度点検したり、再セレクションを行なうことを前提として、ノイズの混入を相当程度許容する設計をしているのである。

機械が行なうファイル・サーチは、人間の持つフレキシビリティを持ち得ないので、極めてハードなシステムとなる。タグ・ファイルには2つの部分から成っている。

1. データ・ファイルの中の個々のデータと1対1の対応を持つ
アドレス・ファイル(すなわち、データ・ファイルの中でそのデータの格納されている番地)
2. 個々のデータの内容を表現するタグ・ファイル。これは1つのタグと、いくつかのデータのアドレスとが1対多の対応を持

っている。

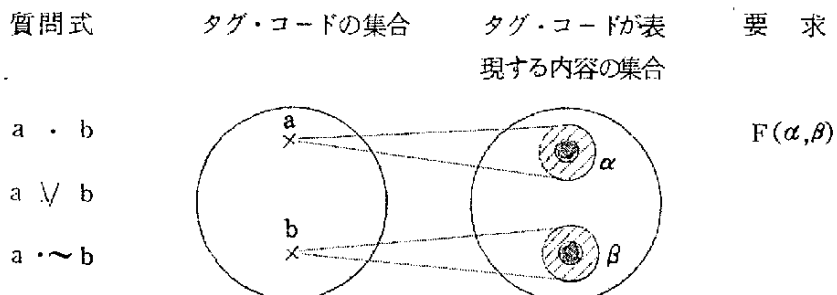
これに対し、検索コードは最も一般的には、タグ・コードの論理結合の構造をもつ、この関係を次表にあげておく。



質問式	値	条 件	出力内容
$a \cdot b = 1$	$A \cap B \neq \phi$		$A \cap B$ の各 エレメント
$= 0$	$A \cap B = \phi$		
$a \vee b = 1$	$A \neq \phi, B \neq \phi$	$A \cup B$ の 各エレメント	
	の少くともどちら かが成立するとき		
$= 0$	$A = B = \phi$		
$a \cdot \sim b = 1$	$A \neq \phi$ で $A \neq B$ のとき	$A \cap \sim B$ の 各エレメント	
$= 0$	$A = \phi$ 又は $A = B$ のとき		

a, b : タグ・コード
 A, B : タグコードと対応
 するアドレス集合
 の部分集合
 D_A, D_B : アドレス集合と
 1:1の対応を持
 つデータ集合
 $\cap$: 集合の積
 $\cup$: 集合の和
 ϕ : 空集合

この検索コードと要求との関係は次図のようになる。要求は、よく



分析するとタグ・コードが表現する内容の（まだよく解明されていない）関数となるであろう。（ $F(\alpha, \beta)$ と表現しておく。）その内容を、より広く把えたタグ・コードが、より狭く把えてタグ・コードの論理結合で近似して質問式の中に把えたものである。

単なる論理結合でなく統計的重みを考慮する試みもあるけれど、現在の所実用的ではない。又、グラフィックに表現できる（或いはすべき）データ（例えば化学構造式の検索など）アナログ・データの場合、それをデジタル表現した上で、論理結合の代りに結合構造パターンに表現し、パターンのマッチング、又は構造的包含関係を用い、アナログ・データを、アナログ形のままでマッチングさせる技術は、基礎研究の域を出ていない。

サーチのオペレーションには、ファイルを最初から終りまですべて一定順序で探すシーケンシャル・サーチと質問式を構成するタグ・コードをテーブル・ルックアップし、そのタグ・コードに対応するアドレス・コード集合の中で質問式の構造とマッチする（質問式の値を1にする）アドレス・コード集合エレメントを求めるものの2つがある。

シーケンシャル・サーチは

- (1) タグ・ファイルがコンパクトなとき。(例えばMTファイルであれば1本か高々システムの持つMT装置の台数以下のとき)
 - (2) 始どリーティングに近いような検索のとき(例えばすべてのデータを利用者の指定するカテゴリー別に仕分けるようなとき)
- に有効である。それ以外はルック・アップ方式が有利であるか、安定した高速のランダム・メモリーの開発が大きいファイルのときには問題となる。又ファイルを常置しておくコストも大きいものとなる。

7 データ出力

データ・ファイルからデータを取り出す手段は、多くの形式が可能になりつつある。それと、技術的に不可能なもの、経済的に成り立ち得ないものを除いて、現在および近い将来に可能になるであろうものを次表にあげておく。この中でOCR、音声の認識と発生、ホログラムなどは実用技術として早く完成されることが望まれる。

尚色彩および濃度の階調の入ったアナログ・データについては各手段で取扱いが可能であるが、忠実性と安定性にまだ問題がある。又、字画の多い漢字のマイクロ形式・超マイクロ形式による処理において、特にレンズの空間周波数特性に少し問題がある。

出力手段 データ ファイルの 記録手段	ペー パー プ リ ン ト	マイ ク ロ 形 式	超 マイ ク ロ 形 式	ホ ロ グ ラ ム	マ シ ン リ ー ダ ブ ル	C R T	立 体 像	音 声	V E T V R R
ペーパー・ プリント	コンタクト プリント	マイクロ カメラ		可干渉光源	OCR/パンチ	撮 像 管		OCR/パンチ 音声発生装 置	撮 像 管 記 録 装 置
マイクロ形式	引 伸 機 リーダー・ プリンター	コンタクト プリント			同 上	同 上		同 上	同 上
超マイクロ形式	リーダー・ プリンター	プリンター	コンタクト プリント		同 上	同 上		同 上	同 上
ホ ロ グ ラ ム	可 干 渉 光源 プリンター			コンタクト プリント	同 上	同 上	撮 影 機		
マシ ン ・ リ ー ダ ブ ル	LP, 特殊 プリンター	マイクロカ メラ/特殊 プリンター		特殊プリン ター	計 算 機	特殊ディス プレー装置	ホログラム を中間媒体 として	音声発生装 置	特殊記録装 置
音 声					音 声 認 識 装 置				
V E T V R R	CRTによ るコンタク トプリント	マイク ロ カ メ ラ			O C R				

8 回答の編成

出力したデータを回答に編成するプロセスは、大きく分けて次の4つの種類が考えられる。

- (1) データ出力そのものが回答となる場合
- (2) 出力したデータに取捨選択を加える場合
- (3) 出力したデータに定型的な加工を加える場合。(非常に簡単なものでは、出力した多くの数値に対し平均値、分散、偏差の計算を行なうとか、或いは、利用者側で指定したプログラムによる計算結果など)
- (4) 依託調査・研究の一環としてDB機能を持っている場合。

(1)～(3)は純然たるDBであるが、(4)はTTとの境界領域となる。公共的なDBの場合、(1)、(3)の機能だけで充分その存在意義がある。又学術的なDBの場合も同様である。しかし商業機関としてのDBが(1)、(3)だけで成立するか否かには問題がある。何らかのレベルで(4)の機能を必要とするであろう

9 データ(質問)の伝送

データの伝送技術としては常識的に郵送、メッセンジャー、電話、加入電信、データ通信があげられるが、TV電話(これからコンタクト・プリント可能)、電話機とキーボードの連結(組み込まれたもの以外に、差し込みによる入出力タイプライター)、小型ファクシミリ等、多くの種類の実用可能性がある。

これに対する利用側からの要求を仕様の形でまとめる必要がある。

10 オペレーション・データとプランニング・データ

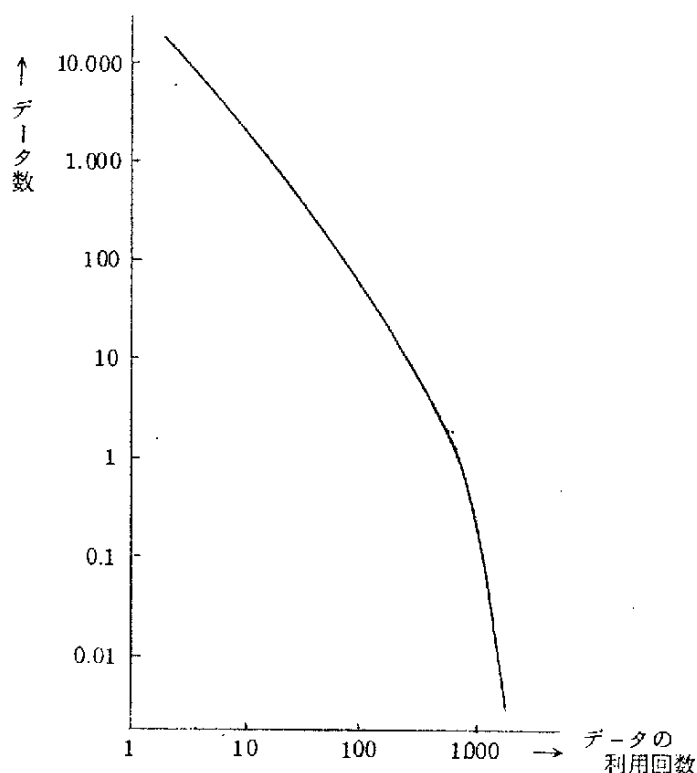
情報要求に2つの型カレント系とシトロスペクティブ系があることはすでに述べたが、要求するデータにも2つの型がある。それは定常業務の中で必要とするもの（オペレーション・データ）と、新しい計画や大きいプロジェクトを立てるときに必要なデータ（プランニング・データ）である。

オペレーション・データの方は、一般的に言えば、カレント系の要求に属する。シトロスペクティブ系の要素が入るとしても過去に逆上る年数が短い（すなわちデータの有効寿命が短い）ため、ファイルのサイズが比較的小さくてよい。又、検索条件も明確にしやすいため、SDIも可能である。

このために、オペレーション・データに関するDBを造ることは、比較的容易である。（カレント系の情報流通は発信指向的であるから、コスト・パフォーマンスを高くすることができる。）特に多量のデータ・ソースを身近かに持つ所ならば、独立したDBでなくとも、その役割を果し得る。

これに対し、プランニング・データの場合は、一般的にはシトロスペクティブ系の要求である。そして、相当古いデータも、又多くの種類のしかも信頼性の高いデータを必要とする。すなわち、高品質のデータの大きいファイルが、種々の分野、多種類のデータについて維持されねばならないし、そのいくつかのファイルをいつでも同時に利用できなければならない。

このファイルの中のデータの利用頻度については、次図に示すような傾向が知られている。



すなわち、利用回数の少ないものが圧倒的に多いのである。このことは、コスト・パフォーマンスを高めるためには利用頻度の少ないデータをカットした方が有利となる。これは商業的DBの場合には可能であるが、公共的性格を持つDBの場合は許されない。

プランニング・データを供給するためのDBは、公共的投資によらなければまず成立しない（少なくともDBの発生期においては、公共財と考えねばならない）とみるべきである。そこで問題となるのは、オペレーション・データに関して、需要があるものについてはDBが自然発生的に造られて行く可能性がある。一方、プランニング・データに関しては公共投資によって計画的に造らねばならない。この両者は、一方があれば他方はなくてもよいという性質のものではない。

情報流通の場面においては、国家統制的な考え方を極力警戒すべきである。公共投資の場合においても、民間のDBを圧迫する形であってはならない。

自然発生的にできるものと、計画的に造るものとが相互抑圧することなく共存させるためには、それらがネット・ワークで結ばれることと、それらの間を調整するモニター機構が必要となる。このためには、

1. 公共的なものは勿論、民間のDB機能を積極的に育成する、財政的・制度的バックアップ
 2. 民間では運営できない性質のDB設立のための投資
 3. ネット・ワーク構成のための投資
 4. DBに必要な技術、機器、ソフトウェア開発のための投資
- に対し、政府（或いは地方公共団体）は積極的姿勢を示すべきである。

11 ネット・ワーク

DBにおけるネット・ワークは、2つの場面がある。1つは、「DBの基本的機能」であげたいくつかの機能の間を如何に結ぶか？という場面と、もう1つはいくつかのフェイルを如何に結ぶか？の場面である。前者を主として1つのDB組織内の問題として考え、ここでは後者を中心としてネット・ワーク問題を取上げることにする。

Si : データ・ソース

CTj : 収集ターミナル

DBh : データ・バンク

STe : サービス・ターミナル

Um : 利用者

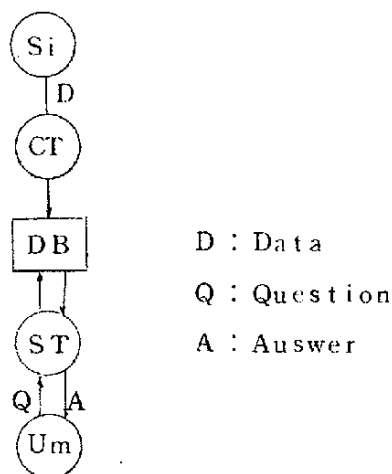
M : モニター

とすると、これらのネット・ワーク構成上の問題点を先ず摘出することから始めよう。

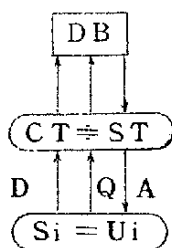
1. Si と Um の関係

ネット・ワークのモデルとして、ソースと利用者の位置づけに1つの問題がある。

モデル 1.



モデル 2.



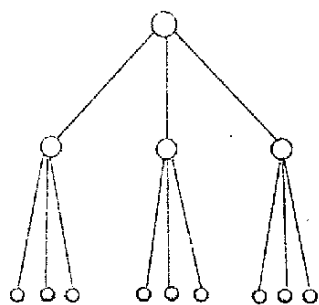
モデル 1 ではソースと利用者を全く異なる集合として捉える立場である。モデル 2 は、ソースと利用者を同じ集合の中で捉える立場である。

モデル 2 は、ソース、DB、利用者で閉じた系を造る場合で、メディア・コミ・社会のモデル化である。現在、情報社会の黎明期にあ

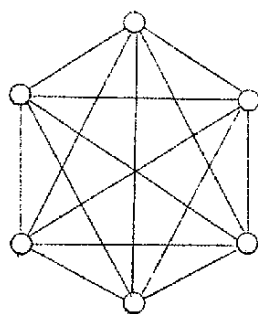
って、社会の中で情報現象に関する社会慣習が確立されていない。その慣習を確立して行くことが、情報に関係する者の1つの任務でもある。そのためには、モデル1によるよりは、モデル2で考える方が有利である。すなわち社会の各メンバーは、自分自身がソースであり、又利用者であることの自覚と、DBによって結ばれるコミュニケーション社会の一員として、ソースとしての権利と義務、利用者としての権利と義務を確立する必要性の認識を持たせやすいからである。そして、地域社会の開発において、モデル2の形でのDB機能を、開発計画の重要なポイントとして組み入れることが必要である。

2. トリー連結とメッシュ連結

ネット・ワークをノッドとノッドの連結としてみると、トリー連結とメッシュ連結が考えられる。



トリー連結



メッシュ連結

トリー連結の場合、コミュニケーション効率は、発信指向的には極めてよい。このため、現在のマス・コミュニケーションではこの

形をとる傾向が強い。(民放における系列化, 新聞社と配達会社の系列化など)そして, カレント系の情報サービスも, この形をとるものが多い。(出版業などは, 主としてこの形であるし, 天気予報, 交通情報の電話サービスなども, この形として捉えることができる。)利用者が自分で操作できる範囲のスイッチ機構でセレクションが可能であれば(例えばTVやラジオのダイヤルのような), この方式が効率的である。

いくつかの種類の情報が混合するような要求に対しては, トリー連結では問題が起る。すなわち, DB間のネット・ワークが必要になる要求に対しては, トリー構造は利用者指向的にみて効率が悪くなる。

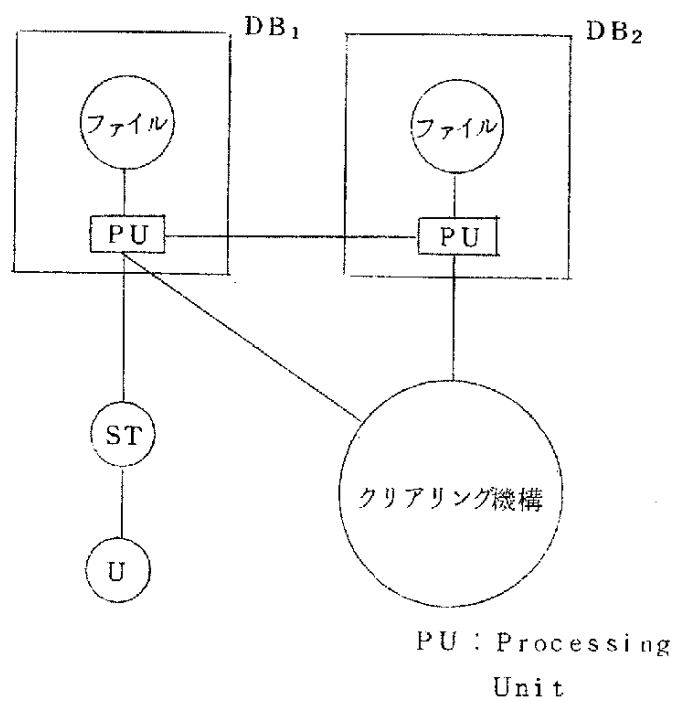
一応現在の段階で考えられるネット・ワークは,

S_i	—	CT_j	:	トリー連結
CT_j	—	DB_h	:	トリー連結
DB_h	—	DB_e	:	メッシュ連結
ST_m	—	DB_h	:	メッシュ連結
又は ST_m	—	ST_n		
ST_m	—	U_o	:	トリー連結

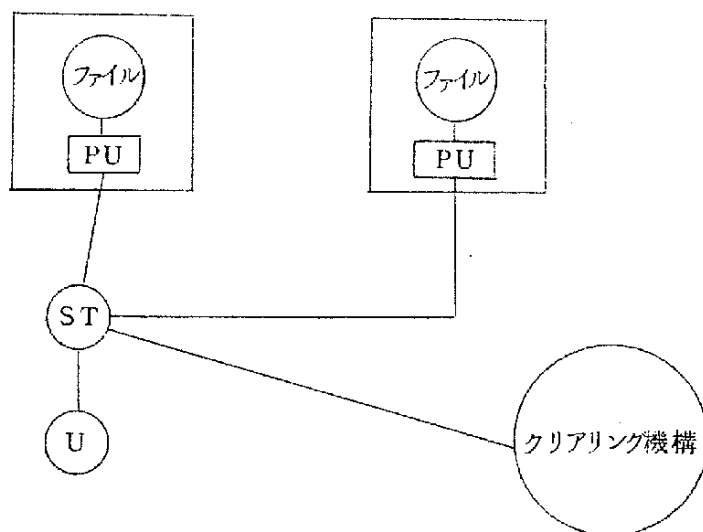
が合理的であろう。

3. DB間のネット・ワーク

DB間のネット・ワークについてはDB間が直接連結される場合と, DB間をSTが結ぶ場合が考えられる。



モデル 2.



モデル1では、ST-DBがトリ-連結になっていて、クリアリング機構がDBと結ばっていてDB間メッシュを維持するものであり、モデル2は、ST-DBがメッシュ連結になっていてクリアリング機構がSTと結ばっている場合である。この両者の優劣については、現在の段階では結論が出せないなので更に検討を要する。

しかし、原則として、独立したいくつかのファイルを、プロセッシング・ユニット間の情報交換でデータをやりとりする可能性がなければならない。このためのソフトウェアの開発も必要である。

4. モニター機構

DBが直接にせよ、STを通してにせよ相互に連結して全体のパフォーマンスを高めるためには、ソース、利用者を含めたコミュニケーション系としてのコントロールが必要である。このコントロールを可能にする、システム・モニターは、計算機システムの場合のように、システムの外から与えるのではない。システムの中で造り出すものである。(かつての行政府が、システムの外からコントロールするイメージは、これからの社会では成立しない。)

しかし、このモニター機構の成立と維持に関する理論は確立されていない。すなわち、自分で造り出したものによって、自分の属する集団の性格が変わり、その集団の変革が自分に影響し、それがリカーブに連結を持つ系の解明で、多くの問題が行きづまってしまっているのである。多くの社会問題、予測の問題、人間を含む系におけるシステム概念、そしてDBに関係する所ではタグ・コード管理の問題等々である。

このモニター機構を通してのシステム・メンバーの相互コントロ

ールに対する何らかの考え方を提示することも、DB基礎研委のテーマの1つとなるであろう。

5. ファイル・シェアと信頼性

DBネットワークがいくつかのファイルから構成されるとすれば、このファイル・シェアの問題が起る。ファイル内のデータに全くオーバーラップがないとき、通信距離のアンバランスが起り、チャネルが混むときの効率が悪くなる。又地震その他の災害などで1つのファイルがつぶれたときの危険が大きい。しかし、オーバーラップが大きすぎても効率が悪い。このオーバーラップに対する定量的基準が、DB網の計画において重要である。これもまた1つのテーマである。

12 システム記述と評価

以上にあげたような、技術的問題を更に細かく分析して、DB網を記述するシステム・モデルを造る必要がある。このとき、その評価を行なう上で問題となる要素をあげると次の項目にまとめられよう。

1. 網 ら 性 : 利用者が必要とするデータの中、どれだけが入手できるか
2. ノ イ ズ : 網ら性を高めようとしたとき、不必要なデータが混入して提供される率をどれ位に収めるか
3. 精 度 : データの正確さ、詳細さがどの程度であるべきか

4. 時 間 : カレント系の場合はデータ発生から利用者の入手まで、シトロスペクティブ系であれば、要求発生から利用者入手までの時間がどれ位であるべきか
5. アクロシビリティ : 利用者が必要なデータを求めるのに、システムアプローチするために要する労力をどの程度とすべきか
6. 信 頼 性 : どんな事態が起っても、DB組が維持できるか
7. コ ス ト : 利用者がデータを入手するに要するコストは如何にあるべきか

(この中に、投資コストも含めて考える)

以上のような要素によって、システム評価を行なうための

- (1) 各成分機能の分析
- (2) 成分機能間の関連の分析
- (3) それらを定量モデルで記述
- (4) パフォーマンス概念の決定

を通して、記述モデルを造ることも、次の段階でのテーマである。

昭和44年度データバンク研究の経過報告

1. データバンク懇談会

1969年8月7日，8月22日，9月9日，10月7日の4回懇談会が開かれた。第1回から第3回までの懇談会では，議題を設定せずにデータバンクの実情紹介を中心として自由討議を行った。

（資料） データバンクをめぐる討論

第4回懇談会では，高瀬教授の視察報告を中心として，アメリカのデータバンクについて検討した。

（資料） アメリカのデータバンク事情（CUDI 1969 10月号）

2. データバンク・シンポジウム

1969年12月6日～7日にわたり迎賓館でシンポジウムを開催し，アメリカの地方行政におけるデータバンクおよびわが国の銀行と電力業における情報処理システムの報告に始まり，データ論，情報の価格形成，さらにデータバンクのシステム構成など，広範囲の問題がとりあげられ討論が行われた。

（資料） データバンク・シンポジウム

3. アメリカのデータバンクの実情調査

1970年3月8日～28日の間，海外調査団をアメリカに派遣し，政府，企業，大学など，43ヶ所を訪問して，アメリカのデータバンクの現状と問題点を調査した。

(資料) アメリカのデータバンク

4. データバンク基礎研究委員会

1970年2月2日に研究委員会が発足し、5回の会合が持たれた。

この委員会の下部機構として、システム・グループ、ネット・ワーク・グループ、価値グループの3つの小委員会が設置された。

第1回 1970年2月2日

研究委員会の進め方

小委員会の設置と分担

第2回 1970年2月17日

データ論および官庁統計

第3回 1970年3月31日

海外調査団の出張報告

第4回 1970年6月2日

アメリカのデータバンクの検討

第5回 1970年6月24日

各グループのデータバンク研究に関する報告

(1) システム・グループ

1970年3月25日、4月1日、5月22日、6月2日、6月30日の5回会合を持ち、システム開発の事例を選び、データバンクのあり方を研究する方法論を検討した。

(2) 情報ネット・ワーク・グループ

1970年3月3日、4月13日、5月6日、6月15日の4回

会合を持ち、意見交換、問題の発掘、通信ネット・ワークとの比較、問題の整理と具体的アプローチの検討などを行った。

(3) 情報価格形成グループ

1970年4月15日、5月20日、6月11日の3回会合を持ち、情報価値の理論的研究とともに情報提供の諸事例につき、如何に価格形成が行われているかを調査した。

