

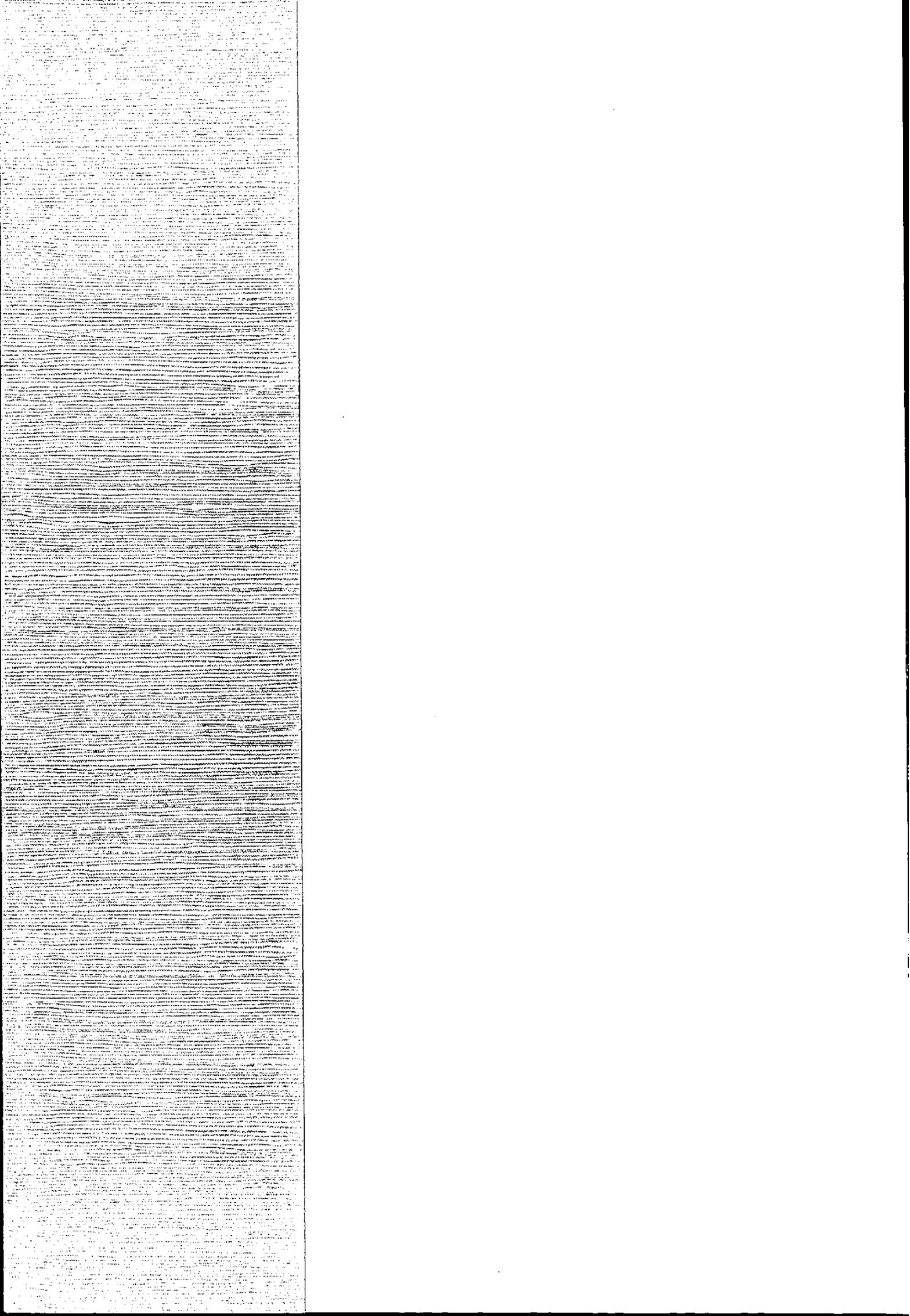
60-DPC-1

保存本

# データベース構築標準マニュアル

昭和61年3月

財団法人 データベース振興センター

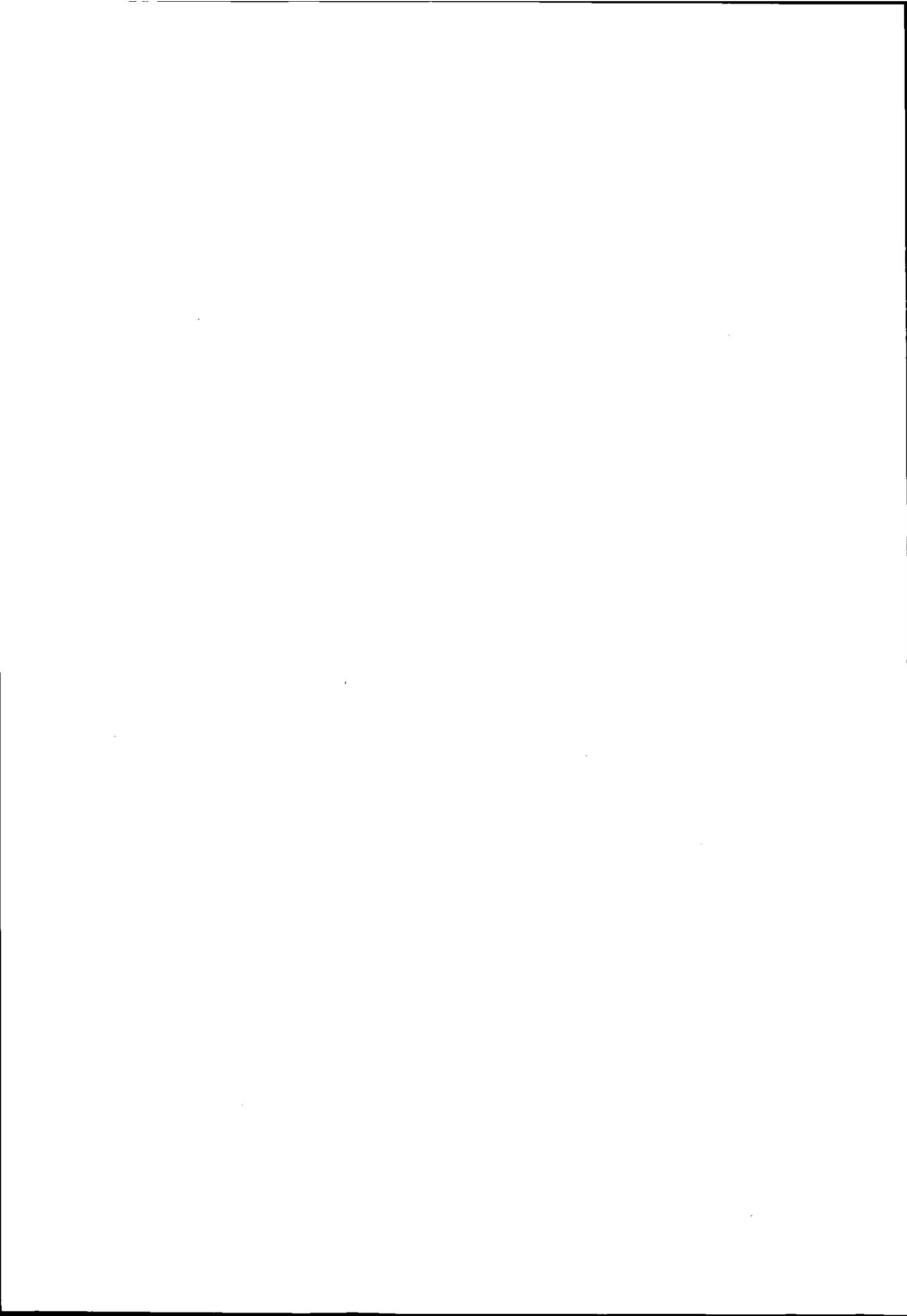


JKA 補助.

# データベース構築標準マニュアル

昭和 61 年 3 月

財団法人 データベース振興センター



## は　じ　め　に

データベース振興センターでは、わが国データベース振興のため発足以来各種事業を行っております。

最近になって、わが国においてもデータベースの重要性が認識され、大手企業の多くはデータベースの活用を始めています。しかし、わが国独自の開発によるデータベースは外国製に比べるとごく限られた数にしか過ぎないのが実情です。

一方、多くの企業ではすでに業務の遂行上自家用データの効果的活用には迫られ、自社内のデータの整備ならびにデータベース構築の必要性が急速に高まっています。こうしたデータには商品的価値を持つものも少なくなく、ここにビジネス、チャンスが生まれます。事実、現在のところ数は限られていますが自家用データをベースにサービスを提供し利益を上げ始めている企業もあります。

今回当財団が実施した調査は、自家用データの整備により、データベースを商品として展開できる可能性について実施したものです。したがって、本報告書は今後自社でデータベースを構築し、市場に参入しようとする企業にとって貴重な資料になるものと確信いたします。

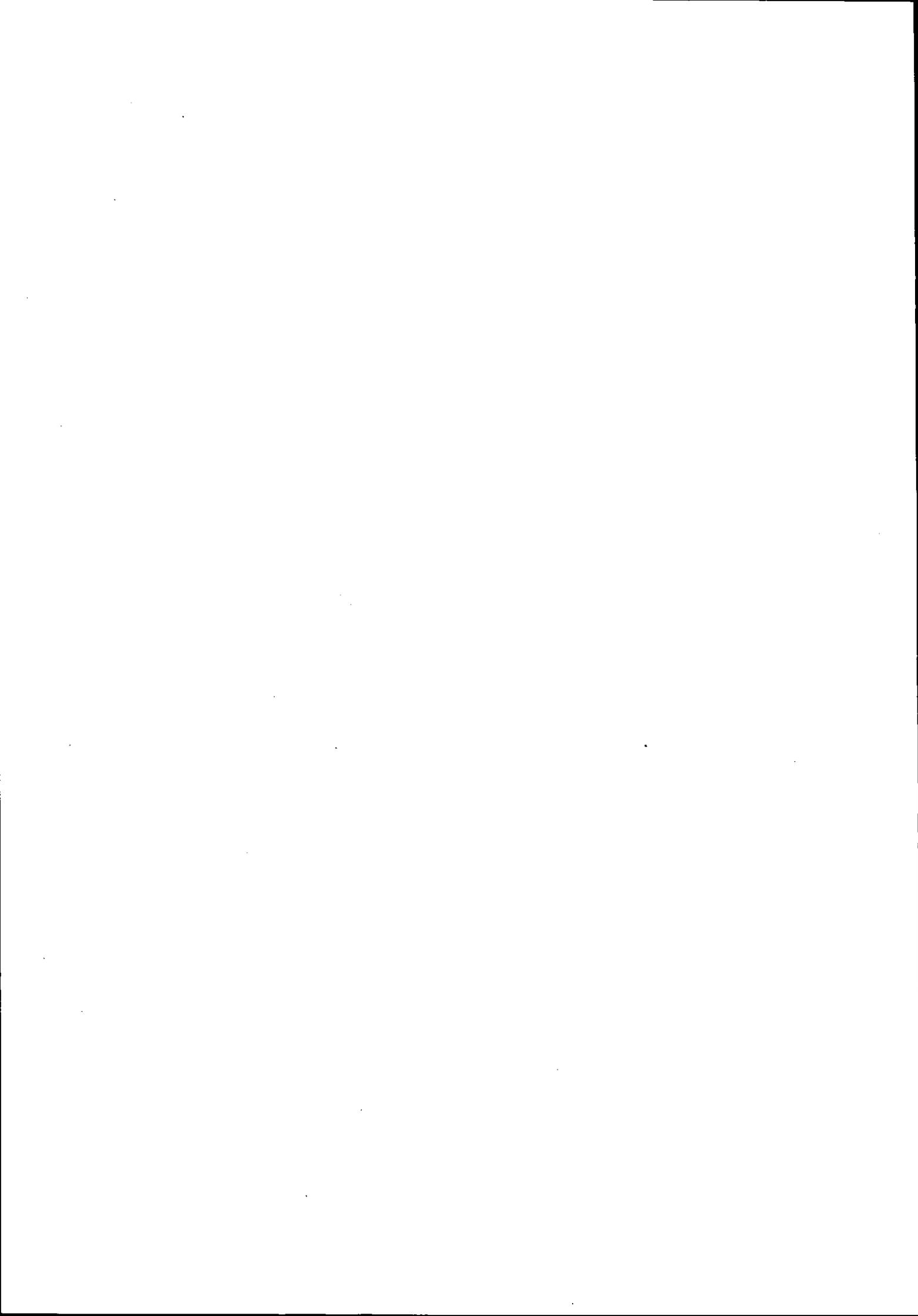
今回の調査は、㈱エポック・リサーチの協力を得て実施しました。

また、実例紹介において各企業のご担当の方々にさらに報告書の編集・作成にあたって専門家諸氏のご協力・ご指導を賜りました。ここに、厚くお礼申し上げます。

本報告が関係各位の参考になれば幸いに存じます。

昭和 61 年 3 月

(財) データベース振興センター  
理事長 円 城 寺次郎



# 目 次

## 第1部 序

1. 序 .....	1
1.1 本マニュアルの対象と目的 .....	1
1.2 本マニュアルの構成 .....	1
1.3 用語解説 .....	3
2. データベースとは .....	9
2.1 データベースの定義 .....	9
2.2 データと情報 .....	9
2.3 データベースの種類 .....	10
2.4 データベース利用の効果 .....	13

## 第2部 事前調査

3. ユーザーと市場の明確化 .....	15
3.1 ユーザーの明確化 .....	15
3.2 類似データベースの探索 .....	18
4. 情報源と情報収集方法 .....	20
4.1 データベース化対象情報の明確化 .....	20
4.2 情報源の明確化 .....	22
4.3 情報収集方法の検討 .....	25
5. データベースの機能の明確化 .....	26
5.1 情報システムの持つべき機能 .....	26
5.2 データベースの持つべき内容 .....	29

### 第3部 分析

6. データの構造化と入力情報の分析 .....	33
6.1 データの構造化 .....	33
6.2 入力情報分析の手法 .....	42

### 第4部 計画

7. DBMSの選定・開発 .....	55
7.1 DBMSとは .....	55
7.2 主な市販(汎用)DBMS .....	57
8. ファイルの設計 .....	60
8.1 データ項目の設定 .....	60
8.2 フォーマット .....	61
9. 入出力の設計 .....	66
9.1 入力的设计と入力手法 .....	66
9.2 出力的设计と出力手法 .....	67
10. 標準化ツールの開発と利用 .....	68
10.1 記述 .....	68
10.2 シソーラス .....	70
10.3 標準分類 .....	70
10.4 交換用フォーマット .....	71
10.5 規格 .....	71

## 第5部 文書化

11. 基本ドキュメントの作成 .....	75
11.1 基本計画書の作成 .....	75
11.2 基本計画書の作成要領 .....	75
11.3 基本計画書の内容 .....	75
11.4 その他の基本文書 .....	75
12. 各種マニュアルの作成 .....	78
12.1 前処理・入力マニュアル .....	78
12.2 運用マニュアル .....	80
12.3 利用マニュアル .....	82
12.4 マニュアル作成の方法 .....	86

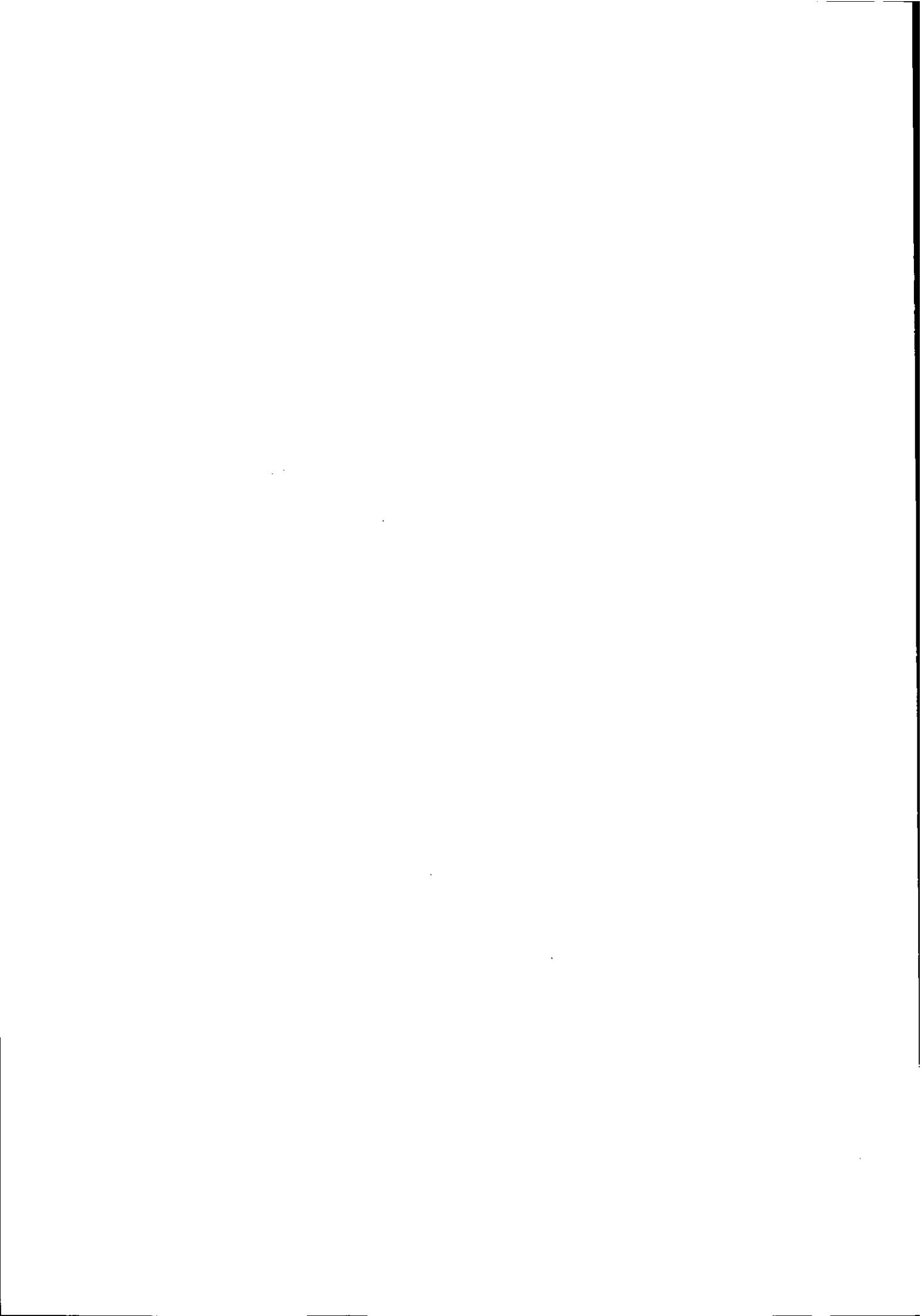
## 第6部 生産

13. 運用・開発面での体制の確立 .....	89
13.1 開発体制 .....	89
13.2 運用体制 .....	93
13.3 サービス体制 .....	94
14. 商品化のためのチェックリスト .....	99
14.1 商品 .....	100
14.2 市場 .....	101
14.3 流通システム .....	102
参考文献 .....	104

資料編 データベース構築例

A. (株)旭リサーチセンター (ID-IR) .....	107
B. (株)市況情報センター (QUICK) .....	115
C. (財)住宅産業研修財団 (HIC) .....	130
D. (株)ゼンリン (地図データベース) .....	139
E. (株)帝国データバンク (COSMOS) .....	147
F. (株)図書館流通センター (TRC MARC) .....	155
G. (財)日本情報処理開発協会 (データベースクリアリングシステム) .....	163
H. (財)日本造船振興財団 (MATRAX) .....	181
I. 米国国立医学図書館 (MEDLINE) .....	188

# 第 1 部 序



## 序

### 1.1 本マニュアルの対象と目的

本マニュアルは、データベースを初めて構築する方々のために、データベース構築の標準的プロセスを解説したものである。

ひとくちにデータベースといっても、その内容や形態は様々であり、各々のデータベースの構築手順も一様ではない。そこで本書では、文献データベースを中心に、基本的なデータベースの構成と構築手法を解説し、更に、必要に応じて、その他のデータベースにも言及するという方式を採用した。なお、本書で扱うデータベースは中型以上の電算機処理を必要とする汎用データベースであり、パソコンによるデータベースの構築は対象外である。

読者は、本書を通読することにより、文献データベース構築の基本的な手順を理解するとともに、データベース作成において一般に用いられている主な情報加工・分析技術と、各種の標準体系を知ることができる。また、データベースの構築とサービスの提供に必要な各種のマニュアルや計画書の概要も紹介している。

更に、各種データベースの構築プロセスの特性を把握するために、実際に構築されているデータベースの事例を資料編に収録した。これらの事例は、本文中の随処で、必要に応じて参照されている。

### 1.2 本マニュアルの構成

本書は、6部14章から成る本文および資料編から構成されている。各部・章の内容は以下のとおりで、その関連は図1.1に示すとおりである。

#### 第1部 序

##### 第1章 はじめに

##### 第2章 データベースとは

#### 第2部 事前調査

##### 第3章 ユーザーと市場の明確化

##### 第4章 情報源と情報収集方法

##### 第5章 データベースの機能の明確化

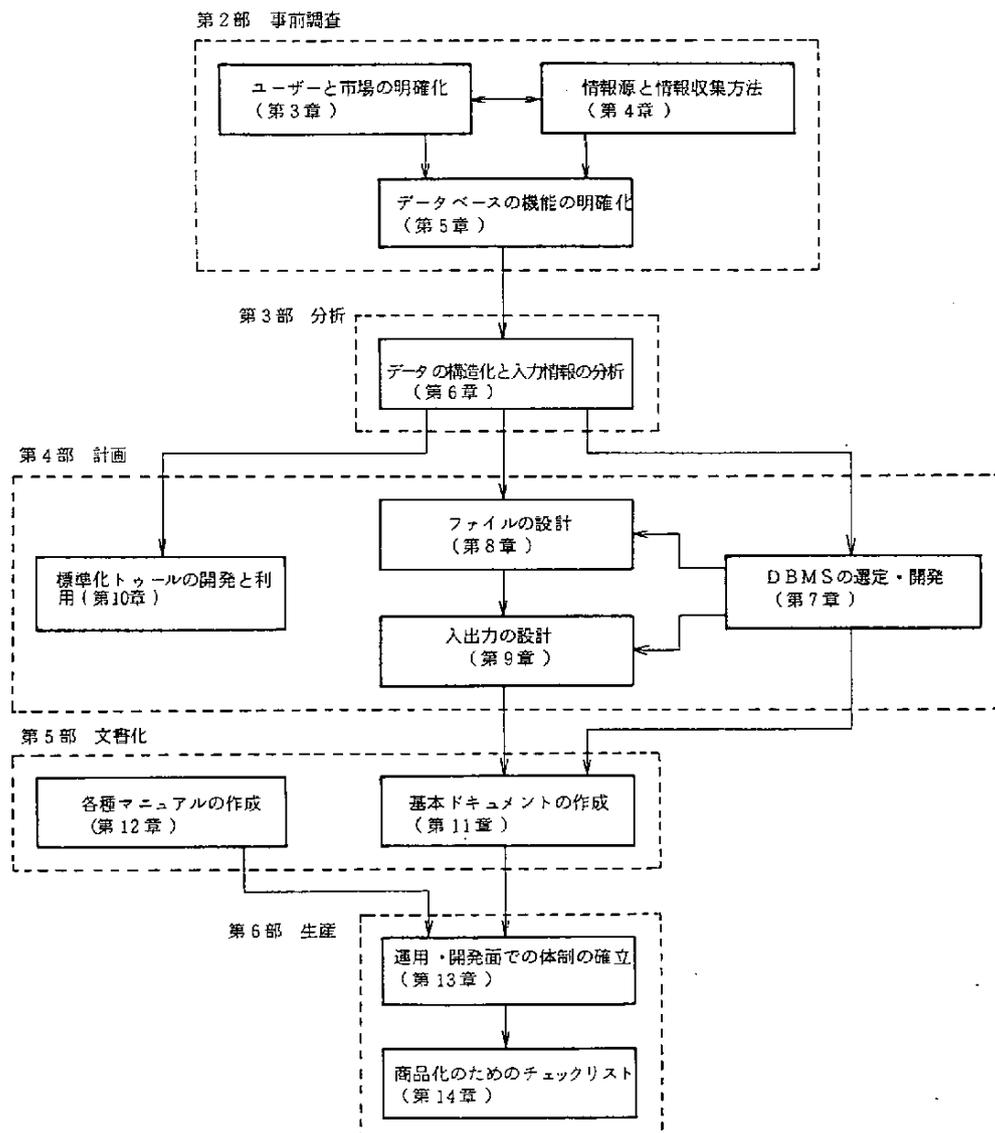
#### 第3部 分析

##### 第6章 データの構造化と入力情報の分析

#### 第4部 計画

- 第 7 章 DBMS の選定・開発
- 第 8 章 ファイルの設計
- 第 9 章 入出力の設計
- 第 10 章 標準化ツールの開発と利用
- 第 5 部 文書化
  - 第 11 章 基本ドキュメントの作成
  - 第 12 章 各種マニュアルの作成
- 第 6 部 生産
  - 第 13 章 運用・開発面での体制の確立
  - 第 14 章 商品化のためのチェックリスト

図 1.1 本マニュアルの各部・章の関連



### 1.3 用語解説

本書中で用いている用語のうち、定義、解釈が必要と思われるものを選定し、以下に語義の解説を示す。

#### (1) アクセスポイント

一般的には、利用者がファイル（これは機械可読のものに限らず、図書館のカード目録なども含む）中のレコード（同様にカード目録に書かれている記述なども含む）を検索し、識別、同定する時に用いる名称、用語、コード等のことである。機械可読の場合には、検索に用いることができるデータ項目をアクセスポイントと言い、原則的にはデータベースに入力された全項目がアクセスポイントとなりうる。

#### (2) 一次情報・二次情報

一次情報とは、それを利用するだけで利用者が他の情報源を参照することなく自らの要求を満たすことができる情報であり、二次情報とは、一次情報を見つけ出すための情報である。例えば、

一次情報：統計データ、物理・化学物性データ、文献の本文、等

二次情報：文献の書誌事項および抄録、ディレクトリー、等

#### (3) オンライン処理

オンラインというのは、中央処理装置と直結している状態であり、この状態でデータの処理を行うことをオンライン処理という。従って、処理すべきデータは発生すると直ちに処理される。データベースのオンライン検索では、端末機から通信回線で結ばれたホストコンピュータと会話をしながら検索する。検索結果を見ながら検索質問を修正し、求める情報に到達し、回答を得る。

#### (4) 索引語

索引作業において索引の対象となるものの概念を表わす語として採用された語。一般に名詞が用いられるが、名詞句でもよい。索引の対象となるものの中に出現したどのような用語（自然語）でも用いる場合と、あらかじめ定められている語彙、たとえばシソーラス中の用語（統制語）を用いる場合とがある。

#### (5) 索引作業

索引作業とは、対象となる文献の内容を分析して、その主題の中の主要な概念を抽出し、それを語もしくは分類コードで表すことである。ここで用いられる語を索引語というが、これには、自然語と呼ばれる統制されない語を用いる場合と、シソーラス等によってあらかじめ定められている統制語を用いる場合とがある。

## (6) 索引誌

索引は、一般的には、特定の記述資料(群)に含まれる各種の情報が探知できるように、それを項目として抽出し、所在指示機能を与え、一定の順序に配列した検索ツールと定義される。

索引誌は、一般的には、雑誌論文、図書などの文献に含まれている主題を探知できるように、それを項目として抽出し、そのもとに書誌事項(標題、著者、出版社、出版年、巻・号、ページ等)を与え、一定の順序に配列した逐次刊行物である。

## (7) サーチエイド

ユーザーが実際に検索を行う時に利用できる検索補助ツールをサーチエイドという。具体的には、シソーラス、分類表、利用マニュアル等を指す。このようなツールを用いることにより、効率的な検索が可能になり、ユーザーが必要としている情報を包括的に、少ないノイズで、少ない労力で検索できるようになる。

## (8) シソーラス

シソーラスとは元来、宝庫、百科事典、知識の宝庫、という意味である。情報検索システムにおいて、索引作業と情報検索で使用するキーワードとなる用語を統制し、用語間の相互関係を管理することにより検索効率を高めることを目的とした用語典拠リストを表す用語として用いられるようになった。

シソーラスは基本的には、ディスクリプタと非ディスクリプタから構成される。一般に、語順のリストと階層順のリストからなり、個々のディスクリプタにはスコープ・ノートと呼ばれる語の意味や使用範囲を規定したものと、他のディスクリプタや非ディスクリプタとの上下関係などが示されている。

## (9) 主題分析

文献を分類したり、あるいは索引作業を行う際に、対象となる文献の主題内容が何であるかを明らかにする作業を主題分析という。文献の内容を完全に理解するためには、本文すべてを読まなければならないが、時間等の制約によりこれが不可能な場合が多い。重要な情報を見落さないために、標題、抄録、目次、序論、挿図・表、強調部分、注、引用文献を詳読し、内容を把握する。

## (10) 抄録

抄録とは、文献の内容の概略を迅速に把握する目的で作られた文章で、主観的な解釈や批判を加えず、記事の重要な内容を簡潔かつ正確に記述したものをいう。

抄録は伝える内容の範囲によって、報知的抄録と指示的抄録に分けることができ

る。報知的抄録は、原文献を読まなくても内容の概要が把握できる程度の詳しい抄録である。抄録の記述内容は、目的、主題範囲、方法、結論などを含む。抄録の長さは、通常欧文で200語程度、和文では400字程度である。指示的抄録は、原文献で取り扱っている主題の範囲がわかり、利用者が原文献を読むべきか否かを判断しうる程度の短い抄録である。抄録の長さは、欧文で30語、和文では数10字程度である。

抄録作成者によって、著者抄録と第三者抄録に分けられる。著者抄録とは、原文献の著者自身によって作成された抄録であり、第三者抄録は原文献の著者以外が作成した抄録である。

#### (11) 抄録誌

文献の書誌事項（標題、著者、出版社、出版年、巻・号、ページ等）と抄録をひと組にして、これを主題によって分類配列してある逐次刊行物を抄録誌という。抄録誌は索引誌に抄録がついたものとも考えられるが、書誌事項と抄録がまとまると、情報量が多くスペースをとるため、索引誌のような自由な配列はできにくくなる。通常、抄録誌は、著者名と主題を表す語から検索できる。著名な大規模抄録誌である Chemical Abstracts では、これらに加えて特許番号索引、フォーミュラ索引など極めて詳細な各種の索引がある。

#### (12) 情報検索システム

蓄積された情報の中から、ある特定の属性を持つ情報を選択することを情報検索という。情報検索を行うためのシステムを情報検索システムと呼ぶ。情報検索システムは必ずしもコンピュータを用いているとは限らず、図書館なども情報検索システムの一つである。

コンピュータを利用した情報検索を機械検索といい、その中でもオンラインで検索できるシステムを特にオンライン情報検索システムと呼ぶ。

#### (13) 情報フロー

情報が生産されてから利用されるまでの一連の流れを情報フローという。情報は、さまざまなメディアによって伝達されるため、その流れは複雑で、同じ情報が異なるメディアにより異なる経路で伝えられる。情報フローを総合的に展望するためには、各情報メディアの役割と情報流通過程について把握している必要がある。

#### (14) ディスクリプタ・非ディスクリプタ

ディスクリプタ（優先語ともいう）とは、索引を作成する時に、ある概念を表わ

すために、一貫して用いられる統制語。つまり、検索する時に用いることができる語である。

非ディスクリプタ（非優先語ともいう）とはディスクリプタの同義語や準同義語で、索引作成の際に文献には付与されない。利用者に対して、用いることのできるディスクリプタへの指示（たとえば、「を見よ」「～を参照」「を使う」）があるので、検索の際には指示されたディスクリプタを利用する。

#### (15) データ項目

情報処理の目的で一単位として取扱われる一連の文字、または語の集合。これはユーザーにとって意味を持つ最少の単位であり、いくつかのデータ項目が統合されて1レコードを形成する。文献データベースの場合は、著者名、標題などのひとまとまりの書誌記述を構成する個々の要素、および検索のために付けられている索引語、分類記号などが、データ項目に相当する。

#### (16) データシート

データ入力の際に、入力すべきデータをあらかじめ記入しておく用紙のことをデータシートという。

データベース作成では、通常1枚のデータシートに1レコード分のデータを記入できるように作成するが、1レコードのデータ項目や記述内容が多く、1枚に入りきれない場合には、関連のある項目が同一用紙にまとまるように設計する。

#### (17) データ・ディクショナリー

D/DDとも書く。通常DBMS中にあり、データベース中の各データの型、名称、属性、アドレス等の記述を含む辞書である。これによって、データとその構造について集中的に管理することが可能になる。また入力データをコードで作成し、それを文字等の正式のデータに内部変換すること、および入力データのエラーチェック等が容易にできる。

#### (18) バッチ処理

データの処理を行う場合、処理すべきデータ（群）をまとめて、一括して処理するやり方をバッチ処理という。この方式では、データがある程度たまってから処理が行われるので、最初に処理すべきデータが発生してから、最終的に処理が終わるまでには、一定の時間の遅れがある。

#### (19) ファイル

ひとつの単位として取り扱われる関連したレコードの集合で同質のレコードが複

数個集ったものである。データベースは、通常ファイルの集合から構成される。

#### (20) 分類作業

分類作業とは、分類対象となる文献の主題分析を行い、明らかにされた主題概念を、分類表に基づいて記号で表すことである。手順としては、索引作業と類似している。索引作業と異なるのは、主題概念を記号で表す点と、与えることができる記号の数が制限されている点である。

#### (21) 分類表

分類表とは、分類法に基づいて知識を分類し何らかの理論的な順序に従って配列した表のことである。各分類項目は、数字や文字で表されている。分類表は、一般に以下の条件を備えていなければならない。

- ・体系の包括・受容性：あらゆる既存の主題を系統的に包摂でき、新主題に対しても体系表のどこかに吸収することができる。
- ・区分・名辞の明確性：各分類項目の意義が明確で、その概念規定も正確であり、外延も明快であること。
- ・柔軟な適用能力：綿密な表であるとともに、適用に際しては、精粗の使い分けができるように区分の各段階が伸縮可能な構造を備えていること。
- ・記号の代位性：分類の体系や仕組みを、そのまま忠実に代位できる記号法を備え、記号の持つ機能が分類効果を促進するような相乗作用が求められる。

#### (22) レコード

情報処理の目的で、1単位として取り扱われる、関連した項目、欄または語の集まり。文献データベースでは、データ項目にあたるものが、著者名、標題、抄録、索引語などであり、1レコードは、このような項目が集まったもので、書誌事項、抄録、索引語等から成る。これが、同質で個別化できる最低の単位であり、ファイルの基本要素である。

#### (23) DBMS (Data Base Management System)

DBMSとは、データベース管理システムの略称であり、データの入力、検索、出力などの管理を行うソフトウェアである。これにより、データの入力が容易にでき、また利用者が独自にデータベースの構造を決め、データの入力・更新ができる。

#### (24) ISO (国際標準化機構)

ISO (International Organization for Standardization) は、国際的な標準化の発展をはかることによって物とサービスの国際的交換を容易にし、知的、

科学的、技術的、経済的活動の領域における相互協力を発展させるという目的のために1946年に設立された。その活動範囲は、別に国際的な標準化機構を持っている領域を除くほぼ全分野にわたっている。

データベースに関連するものは、ISO Technical Committee 46 (ISO / TC 46) によって取扱われている。

⑤ S I S T (Standards for Information of Science & Technology)

S I S Tとは、1978年以降日本科学技術情報センターに設置されている、「科学技術情報流通基準作成委員会」を中心に作成されている、ドキュメンテーションおよび情報処理技術についての基準である。日本語では、「科学技術情報流通技術基準」と呼ばれる。書誌情報の作成および流通に関連する各種の基準が含まれる。

## 2. データベースとは

### 2.1 データベースの定義

各種のデータを収集し、必要なものを引き出しうるようにしたものをデータベースと呼ぶ。この条件を満たすような印刷物を含む各種メディアを総称してデータベースと呼ぶ場合もあるが、本書では、コンピュータで処理できる形で蓄積された上記の条件を満たすものをデータベースと定義する。

現在、わが国で用いられているデータベースの標準的な定義のひとつは、データベース協会のものである。そこでは、データベースを「データを整理統合し、コンピュータ処理が可能な形態にした情報ファイルもしくはその集合体」と定義している。

また、ひとつのデータベースは、ひとつのデータベース・ファイルのみからなる場合と、複数のデータベース・ファイルで構成される場合がある。

以上を総合すると、データベースとは、「多くの利用者のためにデータを供給するもので、ひとつあるいは複数のファイルから構成され、一般的に機械可読形式であり、利用希望者が情報検索システムを通じて利用できるもの」ということになる。

### 2.2 データと情報

データと情報の二つの用語は、しばしば混同して用いられているが、これらは基本的には全く異なる概念であり、その用法の区別をしておく必要がある。

情報という概念は、心理学、哲学、言語学、行動科学、通信工学、情報工学といった様々な分野から定義づけられているが、これらの定義は必ずしも同一のものを示すものではなく、相互に矛盾し合う部分が含まれている。

図書館・情報学分野における情報の定義は、シェラ (Shara)、ケント (Kent)、ペリー (Perry) らによって、以下のように述べられている。

- (1) 情報とは、ある特定の事実、対象、事象に関する知識であり、伝達可能な形式をとる。ドキュメンテーションの意味では、情報は、存在、利用、意味内容という三要素を持っている。
- (2) 実際の効果の点では、情報は既知のものに何かを追加したり、変化させるものである。
- (3) 情報理論における情報は、メッセージの意外性の尺度である。

一方、ヘイズとベッカー (Hayes and Becker) は、「情報とはデータ処理の結果として生じるデータである」と定義し、データと情報の概念を区別している。またマクドノウ (Mac Dough) も、データを未だ評価されていないメッセージ、情報を特定の状況における評価されたデータとし、両者を区別して用いている。

データは情報と区別される概念であるが、両者はきわめて関係の深い概念である。情報はデータから作られる。たくさんある雑多なデータの中の受け手に役立つデータだけが、情報の候補となりうるデータである。データが、受け手によって判断や意志決定、問題解決に使われたときに情報になるのである。

データは、「直接観察した事柄や考えられる事実」であり、それは、言語、数字、その他の記号を使ってあらわされた事実の代替物である。人間がデータを利用するのは、データという代替物によってあらわされる事実を知るためである。従って、事実を正確に認識するためには、客観的事実を正確に報告しているデータを入手することが重要となる。

### 2.3 データベースの種類

データベースは、主題内容、言語、発行機関、利用手法など様々な観点から分類することができる。

図 2.1 は、ヨーロッパの情報サービス機関の団体である EUSIDIC (European Association of Scientific Information Dissemination Centers) が 1977 年第 1 回オンライン国際会議において提示したデータベースの分類枠組である。ここでは、データベースを大きく書誌データベース (索引・抄録誌のデータベース) とそれ以外のものに分け、後者をデータバンクとしている。ヨーロッパでは、書誌データベース以外のデータベースを、一般にデータバンクと呼んでいる。

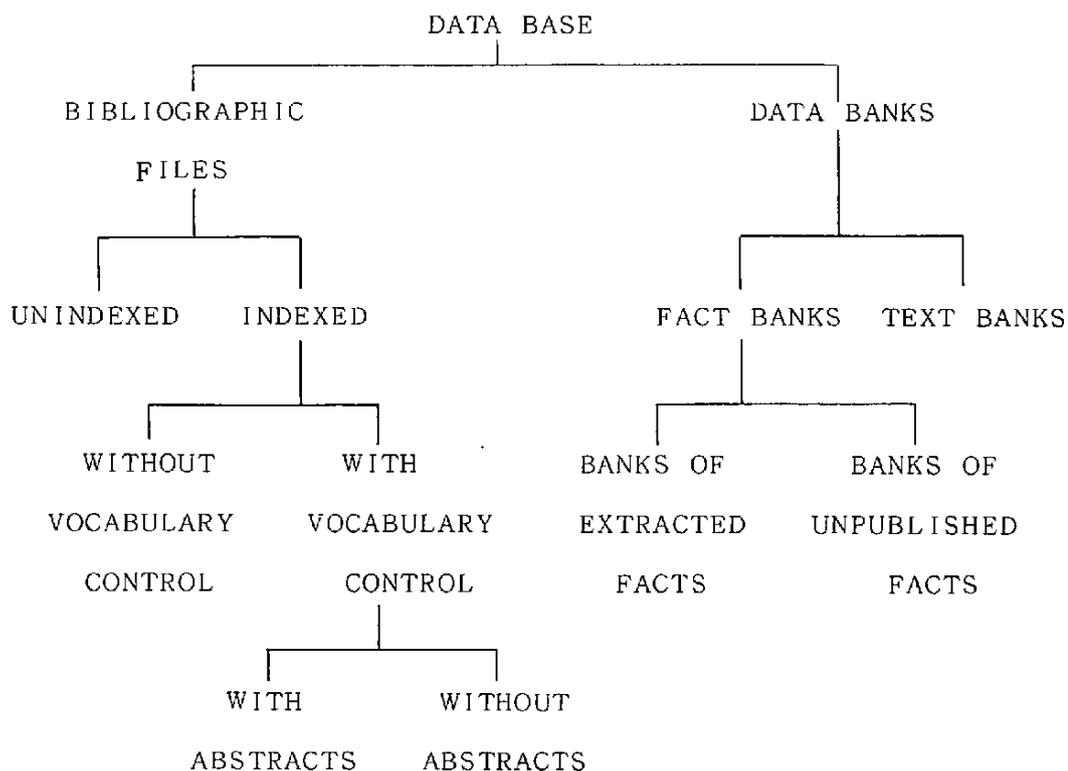


図 2.1 EUSIDIC の分類

図 2.2 は、米国の Cuadra Associates によるデータベースの分類枠組である。ここでは、データベースを二次情報としての Reference データベースと、一次情報としての Source データベースに大別している。

REFERENCE Databases

- Bibliographic ..... 書誌事項および抄録
- Referral ..... その他の案内情報

SOURCE Databases

- Numeric ..... 数値
- Textual - Numeric ..... 数値と文章の組合せ
- Properties ..... 物理的・化学的物性
- Full Text ..... 全文（文献・論文）

図 2.2 LINK社, C.A. Cuadra らの分類

図 2.3 は、日本データ通信協会によるデータベースの分類枠組である。ここでも、データ

ベースを案内と事実の提示という二つの機能に分けている。案内の中に、文献データベースと他の案内データベースが含まれている点は、図 2.2 の場合と同様であるが、事実の提示をファクト・データベースと命名し、そのなかに、数値、文字のほか画像情報を加えている点は、最近のニューメディアを利用した画像情報データベースの発展を予見したものといえる。

情報の目的	データベースの種類	情報の主要素	例
案内	文献データベース	文字情報	抄録誌 索引誌
	その他の案内データベース	文字情報	蔵書目録 機関案内
事実の提示	ファクト・データベース	文字情報	人やモノに関するリスト、 ディレクトリー、辞書・ 事典 全文情報（法令、議事録）
		数値情報	社会活動の諸指標および その統計  実験・観測で得られる データ 物質の構造式
		画像情報	地図・設計図

図 2.3 日本データ通信協会の分類

以上 3 種の分類枠組に共通してみられるように、データベースはその機能面から、二次情報を含むものと、一次情報を含むものに分けられる。前者は別名、参考データベース（Reference Databases）と呼ばれ、後者は、ファクト・データベース（Fact Databases）

あるいはデータバンク (Data Bank) と呼ばれる。ここでは、前者を二次情報データベース、後者を一次情報データベースと呼ぶことにする。

二次情報データベースは、抄録誌、索引誌などのデータベースである書誌データベース (Bibliographic Database) と、名鑑やディレクトリー等のデータベースであるリフェラル・データベース (Referral Database) を含む。

一方、一次情報データベースには、社会統計や経済指標・財務データ等の数値よりなる数値データベース (Numeric Database), 化合物の構造式や物性を含む物性データベース (Properties Database), 雑誌論文や法令等の全文よりなる全文データベース (Full Text Database), および画像や音声等のデータベースが含まれる。

## 2.4 データベース利用の効果

データベースの利用によって得られる効果は、利用者の立場や業務内容、用いるデータベースの種類により大きく異なるが、一般的に以下のような効果が得られる。

### (1) 必要情報の即時的な入手

データベース中に必要とされるデータが網羅的に収録されており、しかもそれらのデータを、あらゆる角度から検索できるようにアクセス・ポイントが付されていたならば、利用者は情報が必要になったときにデータベースを検索すれば、即時的に情報を入手することができる。

### (2) 大量情報の一元管理

データベース中に大量のデータが、一定の整理規準に順じて分類・整理して入力されていれば、それらのデータは一元的に管理される。このため、一人一人の人間が別個に自分の情報を収集・蓄積・管理することにより、同じデータが多くの人の手元や、多くの場所に分散して保管される必要がなくなる。

### (3) 情報探索時間の短縮

様々な場所に分散している情報源の中から、必要な情報を見つけ出すためには、莫大な時間と手間を要する。また、必要な情報の情報源がどこに存在するのかわからない場合には、情報の所在を探索するために、更に莫大な時間と手間を要することになる。このような人手による情報探索に要する時間と手間は、データベースの利用によって大幅に節減できる。

### (4) レベルの高い意志決定の促進

意思決定の根拠になる各種のデータが、データベースとして管理されていれば、

判断が必要なときに即時的に情報を入手できるので、意思決定のプロセスが短縮される。データベース利用により、根拠にすべきデータを網羅的に入手することができるので、幅広く、客観的な根拠に基づいた信頼性の高い意思決定を下すことが可能になる。

(5) 情報コストの節減

データベースの構築や利用には、一定のコストを必要とする。しかるに、データベースを利用せず、個々人の努力によって情報の収集・探索・利用が行なわれると、情報探索に莫大な時間を要し、更に個々人が別々に情報を管理することになるので、資料保管のために莫大なスペースを必要とする。

データベースを利用することにより、情報探索時間が節減されれば、そのために得られる時間を、他の本来の業務に振り当てることができるので、人件費、コストの削減が得られる。更に、個々人が独自に情報資料を保管することが不要となればスペース・コストの削減にもなる。

以上のようなデータベース利用の効果は、必要とされる情報がいつでも利用できるような形で様々なデータがデータベースとして構築されていることが条件である。データがデータベース化されていない状態では、情報検索を行なっても、必要情報の入手はできないのである。

## 第2部 事前調査



### 3. ユーザーと市場の明確化

#### 3.1 ユーザーの明確化

データベースを作成するにあたっては、第一に「ユーザーは誰か」ということを明確にすることが必要である。データベース利用の効果(2.4)で示したとおり、データベースを利用することによって、情報収集が効率化でき、これは業務の効率化や意思決定のサポートにつながる。ユーザーを明確にするということは、そのデータベースを構築することによって、誰の情報収集が迅速化され、誰の業務が効率化でき、そして誰の意思決定をサポートできるのかを明らかにすることである。データベース作成の事前調査の第一段階において、ユーザーの範囲を明確にしておかなければならない理由は、以下に示すように、ユーザーのニーズや情報行動や業務内容によって、求められるデータベースの詳細内容が大きく変化し、そして、ユーザーの数やユーザーの地理的位置によって、データベースのサービス形態が異なってくるからである。

##### 3.1.1 ユーザーの専門性のレベル

データベースには限られたユーザーを対象としたものと、一般ユーザー向けのものがある。

本書の資料編にとりあげた例の中で、旭リサーチセンターの「I D - I R」や、市況情報センターの「Q U I C K」等は、特定分野の専門家の業務をサポートするためのものである。「I D - I R」は、化学分野の専門家向のものであり、化学物質を同定する業務をサポートしている。又、「Q U I C K」は、主として証券アナリストの業務をサポートするためのものである。このような特定の専門職を対象としたデータベースを構築する場合には、その分野の専門家の情報ニーズや情報行動および業務内容を分析し、そのニーズに合った内容や形式を備えたデータベースを構築することになる。文献データベースの場合も同様である。日本船舶振興財団の「M A T R A X」データベースの場合には、船舶・海洋工学の研究者・技術者がユーザーであり、この人達の情報ニーズに合わせた専門的情報を収集してデータベース化している。抄録やキーワード等のデータベースの記述内容も、専門家間で通用する専門用語で書かれている。

限られたユーザー層を対象とした専門家向けのデータベースの場合は、データベース作成者側も同分野の専門家であることが必要である。この種のデータベースの場合は、ユーザー側の情報ニーズを作成者側がある程度把握しており、又情報ニーズ分析にあたっては、

限られた対象（母集団が限られているのでサンプル数も少ない）を調査することになり、かなり詳細にわたる分析が可能である。

一方、一般ユーザー向けのデータベースには、汎用性が求められ、記述内容や検索方式も使い勝手の良さが求められる。一般ユーザー向けデータベースのニーズ分析には、市場調査の手法を応用することになるが、情報ニーズは個々人のライフ・スタイルや価値観等により大幅に異なるので、的確に把握するのは容易ではない。この種のデータベースのニーズ分析は、一般的には社会動向やライフ・スタイルの変化等のファクターから、情報ニーズを予測するという方法をとる。

一般ユーザー向けのデータベースの開発の一方法としては、専門家向けに作ったものを一般向けに加工して提供するというやり方もある。これは、例えば新聞記事のデータベースを新聞記者やマスコミ関係者等の専門家用に作成し、その中の一般向けに公開できる部分、もしくは一般ユーザーの関心や情報ニーズと一致する部分のみを、商用データベースとして提供するというものである。本報告書の資料編でとりあげた、ゼンリンの「地図データベース」は、基本となる地図データのみをデータベースとして提供し、その上にユーザーが自由にデータを付加できるように考慮しており、一般ユーザーの個別のニーズに柔軟に対応しようとの姿勢がうかがえる。

このように、専門家向けのデータベースを汎用化する場合には、初期のオリジナルなデータベースから、二次データベースを作成することになる。この場合には、オリジナル・データベースを作成する際に、専門家の情報ニーズを十分に把握し、専門家にとって有用なデータベースを開発することが第一に必要なが、後に加工できるように、汎用性についても考慮しておくことが必要である。

### 3.1.2 ユーザーの組織上の位置づけ

企業や団体等の組織内のユーザーのみを対象とする、いわゆる「社内データベース」の場合と一般に公開する「汎用データベース」の場合とでは、データベースに求められる機能が異なるので、構築の際の取り組み体制や、サービス体制にも違いがみられる。

前者の場合には、社内の特定業務のサポートに用いるものと、社内で汎用データベースとして共用するものとはユーザーの範囲が異なる。特定業務のサポートに用いるデータベースでは、その業務に係る人間のみが調査対象となるが、社内の汎用データベースとして用いる場合には、社内の関連部門全体が調査対象となる。

社内の汎用データベースの開発においては、担当部門によって、情報ニーズや情報利用

の方法に違いがみられるので、各部門別業務別のニーズ調査が必要となる。又ユーザーの職位により、入手できる情報にレベル差を設けることが必要な場合もあるので、機密保持機能を持たせることも考慮する必要がある。例えば、人事や経理のデータベース等では上級管理職が利用できる範囲と、中間管理職が利用できる範囲と、更に一般社員が利用できる範囲には差をつけることが必要となる。

企業内の業務をサポートするために開発した社内データベースを、商用データベースとして外部に提供する場合には、上述の機密保持の問題のほかに、提供方法、料金徴収方法、エラー処理方法、サービスのサポート方法等の広範にわたる考慮事項が生じてくる。個々の事項の詳細については、後述する。

### 3.1.3 ユーザーの地理的位置

データベースのユーザー対象が社内であれ社外であれ、ユーザーの地理的位置（ユーザーの居る地点）はデータベースのサービス形態を決定する上で大きな要因である。ユーザーが地理上の1地点や1つの建物の中に集中していれば、データベースをホスト・マシン内に構築して、これに直結された端末から、ユーザーがデータベースを利用することは、現在の技術をもってすれば全く問題はない。又、データベースを利用する際の利用手順や手続上の問題もデータベース作成者とユーザーとの間の直接的なコミュニケーションによって処理できるので、ユーザー教育や、利用マニュアルの整備・変更もデータベース作成者の配転や退職という事態が起らない限り、特に問題とはならない。

一方、ユーザーが地理上のいくつかの地点に分散している場合には、データベース作成者とユーザーとの間の何らかのコミュニケーション手段を構ることが必要となってくる。この場合の処理方法としては、以下に示すような幾つかの代替案がある。

- (1) 磁気テープ等の二次記憶媒体にデータベースを複写し、各地点（建物）に同一内容のデータベースを整備する。
- (2) 他地域（他地点）のユーザーに対しては、バッチ処理サービスを行ない、郵送やファクシミリ等により出力結果を届ける。
- (3) ホスト・マシンの中にデータベースを構築し、他の地点にオンライン・リアルタイムのサービスを提供する。
- (4) 地域ネットワークをもつオンライン・サービス業者（ベンダー）に依託してオンライン・サービスを行なう。

いずれの場合にも、予め処理手順や処理日程を決めておき、サービス窓口を設置し、ユーザー教育や利用マニュアルの整備をすることが必要となる。

### 3.1.4 ユーザーの時間コスト

データベースを構築し、これを運用することには、様々なコストが伴う。また、データベースは常時内容を更新して、ユーザーが常に最新の情報を入手できるようにしておかなければならないので、作成したデータベースをメンテナンスし続ける仕事は永続的に行なう必要がある。データベース構築後にメンテナンスを中止すると、それまでにかけた労力とコストとは水泡に帰することになる。

データベース構築に伴う費用については、社内データベースの場合は、データベース構築によって得られた業務効率の向上によって評価するのが一般的である。業務効率の向上の尺度の一つに、ユーザーがどれだけの時間を節約できたかというものがある。この評価は、得られた時間とユーザーの時間単価の積（管理者分も含まれる）として算出される。ここでは誰の時間がどの程度節約できたのか、という点が大きな問題となる。それは、時間単価の高い人間の業務効率の向上の方が、低い人間のそれより高く評価されるからである。そこで、データベース構築の基本計画の段階で具体的な数値や金額を積算するためにも、「誰がユーザーか」ということを先ず明らかにしておくことが必要である。

一方、データベースを商品化して販売する場合には、ユーザーから何らかの利用料金を徴収するわけであるが、ユーザーがここでも、データベースの利用（もしくは情報入手）にどの程度支出できるかという点が焦点となる。一般に、時間単価の高い人間ほど情報収集の効率化による時間の節約を高く評価する傾向があるので、ユーザー対象としては望ましい。商用データベースの利用料金を決める際には、目標とするユーザー（群）の年齢、職種等から、そのユーザーの時間単価を算出し、これを料金設定の一つの目安とするからである。

## 3.2 類似データベースの探索

作成しようとするデータベースと類似した情報を収録しているデータベースの有無については、計画の初期段階において充分探索し、内容を把握しておく必要がある。その理由は、

- (1) 既に同内容のデータベースが存在するならば、それを購入して利用することができる。
- (2) 全く同一の内容でなくとも、多少の変更や改訂で済むものであれば、既に存在するものを購入し、これに必要なデータを追加・修正し、二次データベースを作成することができる。
- (3) 収録対象が異なっても、類似分野のデータベースであれば、そこで用いられ

ているフォーマット，索引方式，標準化手法等を取り入れることができる。この良い例は，類似内容の外国製データベースである。データベースの構築過程においては，様々な試行錯誤がくり返され，ユーザーのニーズを反映した様々な工夫がなされている。既存のデータベースにおいて採用されている様々な手法を学び，これを継承した上で，更により一層の工夫を試みることにより，優れたデータベースの構築が可能となるからである。

- (4) 類似データベースが全く存在しない場合は，その市場性がないためであるのか，あるいは，未だに誰もそのニーズに気付いていないかのどちらかである。これを見極めてから，データベースの構築にとり組むことが必要である。

## 4. 情報源と情報収集方法

### 4.1 データベース化対象情報の明確化

構築しようとするデータベースの対象とするユーザーが特定されると、データベースに収録すべき情報の範囲が限定されてくる。

次に、データベースに収録する情報の範囲や内容について、各側面からの検討を加えることにより、収録情報を個別要素に分解していく作業が必要となる。検討すべき要件としては、以下のようなものがある。

#### 4.1.1 一次情報と二次情報

第一に、収録すべき内容は一次情報であるか二次情報であるかを、明確にしておく必要がある。

一次情報とは、ファクト、即ち情報そのもののことであり、記事や文献の全文、人口統計や実験データ等の数値、地図や図面等の画像、音声など、各種のものが含まれる。

二次情報とは、人やモノや情報がどこに存在するのか、どうすれば入手できるのかを知らせるための情報であり、記事や論文の索引や抄録、電話帳、住所録、商品カタログなどが含まれる。

一般に、一次情報を収録したデータベースのことを「ファクト・データベース」と呼び、二次情報を収録したデータベースのことを「参考データベース」と呼ぶ。本書の資料編にとりあげた例では、以下のように分けられる。

##### (1) ファクト・データベース

- ・旭リサーチセンターの「I D - I R」
- ・市況情報センターの「Q U I C K」
- ・帝国データバンクの「C O S M O S」
- ・日本経済新聞社の「N E E D S - T S」
- ・ゼンリンの「地図データベース」

##### (2) 参考データベース

- ・日本造船振興財団の「M A T R A X」
- ・日本情報処理開発協会の「データベースクリアリングシステム」
- ・図書館流通センターの「T R C M A R C」
- ・住宅産業振興財団の「H I C」

#### ・米国医学図書館の「MEDLINE」

これらの例からもわかるように、ファクト・データベースを構築する場合には、収集した情報にほとんど手を加えずにデータベース化する。一方、参考データベースの場合には、索引、抄録、主題分析、データシートの作成など、収集した情報を整理・分析し、パッケージ化するための、前処理作業が必要になる。

参考データベース作成のための前処理作業は、現在の技術水準では、ほとんどが人手によって行なわれるため、データベースの規模が大きくなると、莫大な費用を要する。

#### 4.1.2 収録情報の属性

データベースに収録するデータの属性も、この段階で明確にしておく必要がある。データの属性とは、文字、数値、画像、音声といった、データの表現方法をいう。

統計情報や財務情報等の数量化されたデータによって構成されるデータベースのことを、数値データベースと呼ぶ。例としては、市況情報センターの「QUICK」や日本経済新聞社の「NEED-TS」がある。数値データベースの多くは、データベースから検索されたデータに、後に統計処理を加えてモデル化やシュミレーション等を行なうので、データは演算可能な数値（Value）として入力されている。データが数字で表現されていても、それを表示するだけで演算処理を必要としないのであれば、これは次に示す文字データとして入力されている。

記事や文献の全文もしくは索引や抄録等の文字列のみから構成されるデータベースもある。この種のデータベースのほとんどは文献データベースである。データベース中では、文字はコード化されて表現されている。

地図や図面や写真などの画像データによって構成されるデータベースもある。画像データベースの表現の仕方には様々な手法がある。

音楽や人の声等の音声データによって構成されるデータベースを音声データベースと呼ぶ。

データの属性によって、個々のレコードのフォーマットやデータベース全体の容量が異なり、要求されるデータベースの蓄積媒体や検索方法も異なる。

#### 4.1.3 収録情報の範囲と量

次に、データベースに収録する情報の量を測定する。これを可能にするため、まずデータベースの収録対象の範囲を決める。

ファクト・データベースの場合は、本書の資料編でとりあげた例にも明らかなように、収録対象が決まれば、収録範囲も決まるのが一般である。旭リサーチ・センターの「I D - I R」では、収録対象の化合物の種類が決まれば、収録件数も決まる。市況情報センターの「Q U I C K」では、対象とする証券取引所が決まれば、銘柄の件数が決まる。他の例も同様である。

参考データベースの場合は、必ずしも明解ではない。例えば、日本船舶振興財団の「M A T R A X」の場合には、収録対象誌を決めることが必要である。日本情報処理開発協会の「データベースクリアリングシステム」の場合は、“日本で利用できる”データベースと収録範囲を限定している。図書館流通センターの「T R C M A R C」の場合も、東販を経由して流通する図書に焦点が絞られている。

これは、参考データベースのほとんどが、外部の情報を収集して作られているため、収録対象が、比較的集めやすいものに限られるためである。また、二次データベースの場合は、データベース作成時に人手による前処理を必要とし、これに莫大な費用がかかるので、収録対象を制限しないと、データベース化が不可能になるためでもある。

特に、抄録・索引等の文献の二次情報のデータベースでは、ユーザーにとってメリットのある質の高い情報がユーザーから求められるため、情報源を質の面から限定することも必要である。即ち特定の専門分野の索引・抄録のデータベースを構築する場合には、その分野の専門家の間で高く評価されているコア・ジャーナルを情報源の中心とした、入力可能な数の情報源に制限すべきである。

#### 4.2 情報源の明確化

収録対象情報が明確にされたならば、次にその情報をどこから入手するのかを決定する。データベース構築前の段階において、ユーザーがその情報をどこから、どのようにして入手しているかが明らかにされていれば、その情報の入手先がデータベースの情報源である。

データベースの情報源とは、データベースに収録されるデータが発生する場所もしくは発生源となる人や機関のことである。例えば、資料編第5章の帝国データバンクを例にとると、「企業信用調査ファイル」の情報源は外部の企業であり、「COSMOS 3 個人情報ファイル」の情報源は、外部の個人と新聞社である。ただし、文献データベースの場合には、文献の本来の情報源は著者もしくは著者の所属機関であるが、実際の文献の入手先は雑誌や図書や報告書などの媒体であるため、これらの資料を便宜的に情報源と呼ぶ場合もある。

情報源がデータベース構築主体の組織内にある場合と組織外にある場合とでは、情報源を探索するための手法や手間が大きく異なる。以下にデータベース構築主体から見たデータベースの情報源とユーザーの位置づけの幾つかのパターンを示す。

情報源	利用者	例
外部	内部	顧客ファイル、仕入れ先ファイル、外部文献ファイル等の組織内の業務サポート用データベース
内部	内部	在庫ファイル、社員ファイル、社内文書ファイル等
内部	外部	オーダーエントリーファイル、蔵書目録等の外部顧客に公開されている社内データベース
外部	外部	一般の文献データベース等の商用データベース

#### 4.2.1 内部情報源

情報源が組織体内部にある場合には、組織体内での情報フローを探索すれば、情報源を明らかにすることができる。例えば、研究開発の成果として得られた社内の技術情報をデータベース化する場合には、その研究開発を担当している研究者・技術者が情報源である。また、社内の財務データをデータベース化する場合には、経理・財務部門もしくは経営管理部門が情報源となる。社内で行われているプロジェクトのデータベースを作る場合には、社内で行なわれている全プロジェクトに関係している全ての人間が情報源である。

#### 4.2.2 外部情報源

情報源が組織外に広く分散している場合には、組織内にある場合ほど単純ではない。これは、ひとつには外部の情報源が、組織体、人間、資料、モノ、という様々な形態をとっているからである。

##### (1) 人間が情報源の場合

外部情報源が人間である場合には、その人々の間における情報フローを探索する。即ち、情報源の間に張りめぐらされた、目に見えない情報のネットワークを見つけ出し、情報源をたぐり寄せていく。例えば、特定分野の先端技術動向のデータベースを構築するといったような場合は、その分野の研究者や技術

者によって、自然発生的に形づくられている情報網を見つけることから始める。

この種の情報については、学会、研究会というような公的な情報流通の場があり、ここに情報源である人間がメンバーとして参画している。科学分野での主要なデータベースのほとんどは、このような学会によって構築されている。これは、学会が最も情報源に近く、自然に情報が集まる仕組みが既に存在するからである。

科学分野以外では、情報流通があまり公式化されていないので、情報源の探索には困難が伴う。この場合には、業界団体や新聞・雑誌等の出版物を作成している機関など、おのおの情報が最も集まりやすい場所、もしくは、既に外部情報の収集を業務として行なっている場所に、情報が集まっている。これらの機関や企業は情報源を掌握しているので、これらをデータベース化することができる立場にある。

## (2) 情報源の代りに資料を用いる場合

情報源が図書、雑誌、報告書、写真、カタログ等の資料の形をとっている場合には、これらの情報資料を見つけ出すことになる。ほとんどの資料は、組織体もしくは個人から発行されているので、本来の情報源である発行機関を情報源として選定する方法もあるが、文献の流通構造の中でこれをとらえることもできる。国内の図書目録のデータベースは、出版された図書が必ず1冊は納本される国立国会図書館で、情報を入手することができる。日本の包括的な図書目録のデータベースである JAPAN MARC は、このようにして集められた図書に基づき国立国会図書館で作成されている。図書目録の場合は、日本には東販と日販という二つの図書流通の巨大ネットワークがあり、おのおのが、扱っている図書の目録データベースを作成している。

特定分野の文献データベースを構築する場合には、計量文献学の手法を用いて情報源の質の評価を行なう。特定分野のコア・ジャーナル(主要雑誌)を見つけ出すための計量文献学の一手法として、引用文献分析がある。これは、

- ・重要な論文、質の高い論文は、よく利用される雑誌に掲載される
- ・よく利用される雑誌に載った論文は、よく引用される

という二つの原則から、特定分野のレビュー誌や雑誌論文の引用文献を累積し、どの雑誌からの引用が最も多いかを調べる。引用されている頻度が高い雑誌であればあるほど、重要度も高いということから、よく引用される雑誌(群)を

コア・ジャーナルとして選定する。

#### 4.3 情報収集方法の検討

構築すべきデータベースの情報源の所在が明確になったならば、それらの情報源からどのようにして情報を収集し、これをデータベース化していくかについての方法を検討することが必要である。情報収集の方法には、以下の3種の形態がある。

##### (1) 集中処理方式

情報源（資料、データ、人材）を一カ所に集中し、そこで情報分析、入力前処理、データ入力を行なうやり方。この場合、情報源が、1カ所に集中していなければ特に問題はないが、外部に分散している場合には、情報源をどのようにして集めるかの検討が必要である。

##### (2) 分散方式

情報源の所在地において、情報分析、入力前処理、データ入力を行なうやり方。この場合は、情報源が分散していることが前提となる。同一社内であっても、地域が異なれば、この方法が望ましい場合もある。リアルタイムの情報が必要な場合はオンラインによる分散入力であるが、緊急を要さない場合は、フロッピー・ディスク等のオフライン・メディアを利用することもできる。

##### (3) 分散前処理集中入力方式

情報源の所在地において、情報分析、入力前処理を行ない、データ入力は1カ所（通常外注）で行なう方式。これも情報源が分散していることが前提となる。ただし、入力は1カ所なので、オンラインによる分散入力はさけることができる。

以上の中からどの方法を採用するかは、許されるタイム・ラグと、支出できる費用の範囲によって異なる。一般に、

- (1) 情報源が集中している場合は、集中方式を採用する。
- (2) 情報源が広く分散しており、しかもリアルタイムでの情報が必要な場合（例えば、複数の株式市場の市況情報など）には、分散方式が採用される。
- (3) 情報源が分散しており、これを分析・前処理できる人材が情報源の側で確保でき、しかも、タイム・ラグがあまり重要でない場合は、分散前処理集中入力方式を採用する。

## 5. データベースの機能の明確化

### 5.1 情報システムの持つべき機能

データベースは、DBMSと一体化して、情報システムを構成する。データベースの基本設計の第一段階においては、情報システムが全体として持つべき機能を具体的に定義することが必要である。

一般に、情報システムには、以下のような機能が要求される。

- (1) 大量情報の一元管理
- (2) 情報流通の円滑化
- (3) タイムリーな情報の提供
- (4) 高品質な情報の提供
- (5) 豊富なアクセス・ポイントの提供
- (6) 機密保持

情報システムが持つべき機能を具体的に定義するということは即ち、構築しようとする情報システムの環境、特に利用者の要求に基づいて、上記の1～6の項目について、求められている条件を明確にしていくことである。

#### 5.1.2 情報流通の円滑化

システム化されていない情報環境においては、情報は口頭、電話、メモ、文書、出版物等を通じて流通している。このような状況において、情報は人と人との間に目に見えぬ形で張りめぐらされているフォーマル、インフォーマルなネットワークを通じて流通している。

フォーマルなネットワークの例としては、企業内の組織体の上下関係、学会や専門家団体の公式会合の発表や会報、学校の授業、出版社から取り次ぎを経て書店や図書館等を通じて提供される図書・雑誌等の出版販売ルートのようなものが存在する。

一方、インフォーマルなネットワークとしては、同窓の友人、近隣の知りあい、同好会の仲間といった、プライベートな人間関係の間に張りめぐらされている、個人的な情報ルートがある。

フォーマルなネットワークにおいては、情報は構成されている組織体の中を順を追って流通するため、情報発生から情報入手までの間に一定期間のタイム・ラグを生ずる。インフォーマルなネットワークにおいては、情報の提供者と受け手が既に知り合いであること

が条件であり、又通常は対一もしくは少人数の集団の間で情報の授受が行なわれるため、伝達の範囲も限られている。この情報が第三者に伝えられるには、一定の期間を要し、又情報内容も途中で変形される可能性がある。

情報流通をシステム化することにより、以下のように情報流通の円滑化をはかることができる。

- (1) 情報発生から情報入手迄のタイム・ラグを短縮
- (2) 提供情報の流通過程における消失や変形を解消
- (3) 従来伝達することの出来なかった受け手への情報伝達が可能となり、情報流通の範囲が拡大される。

この側面における情報システムの持つべき機能として定義すべき条件は、以下のようなものである。

(1) 情報利用者の範囲

利用者は、組織体内部に限られるのか、外部の利用者にも提供するのか。利用者の地域分布等。

(2) 許されるタイム・ラグの範囲

発生した情報が利用できる状態になる迄にどの程度の期間（日数）の経過が許容されるのか。これは情報システムで流通する情報のタイプにより大幅に異なる。市況情報や為替相場や気象情報は即時であり、ニュースは即日である。一方、ディレクトリーや文献の場合は、数日から数週間の遅れが許される。

(3) 更新頻度

蓄積されている情報をどの位の頻度で更新することが必要か。これは上述のタイムラグの問題と関連する。株式市況等は秒単位の更新が必要とされる。又、ダイレクト・メール等に用いる住所ファイル等や新聞記事も、即日更新が必要である。一方、文献情報は、一般に一週間単位もしくは一カ月の更新・追加でも間に合う場合が多い。

### 5.1.3 タイムリーな情報の提供

情報流通をシステム化することにより、タイムリーに情報を入手することが可能となる。タイムリーな情報提供とは、以下の二つの要素を含む。

- (1) 必要時に必要情報を即時的に入手できる
- (2) 特定分野の最新情報を常時定期的に入手できる

以上の機能は、情報システムにおけるデータベースの提供状況によって実現できる。即ち、(1)については、データベースをオンライン対話型で提供することにより、(2)については関心テーマをプロフィールとして登録し、データベースの更新の度に適合レコードを自動的に出力するS D I機能により可能となる。

#### 5.1.4 高品質な情報の提供

情報システムから提供される情報には、誤りの少ない質の高いものであることが要求される。特に、一般利用者に有料で情報提供を行なうような情報システムの場合には、情報の質の高さが利用者獲得の一つの鍵である。

情報の品質の尺度は、提供される情報のタイプにより異なる。情報のタイプ毎の品質の尺度の重点を以下に示す。

##### (1) 市況情報

即時性、情報源の信頼性、数値の精密さと正確さ等。

##### (2) 統計情報

サンプルの包括性、統計手法の信頼性、数値の精密さ等。

##### (3) 文献情報

収録対象文献の品質（新聞・雑誌・出版社の信頼性）、収録対象の網羅性と包括性等。

##### (4) 物質データ

収録対象の網羅性、データの信頼性、測定手法の信頼性、データの精密さ、データの即時性等。

##### (5) ディレクトリー

包括性、更新頻度（住所や担当者名変更に対応しているか）等。

#### 5.1.5 豊富なアクセス・ポイントの提供

情報のアクセス・ポイントとは、検索可能な項目のことである。文献を例にとると、標題、雑誌名、著者名、会議名、所属機関名、出版日付、出版地、分類、主題を示すキーワード等がこれにあたる。情報システムの利用目的によって、検索可能な項目が異なるため、どのようなアクセス・ポイントが必要であるかを定義することが必要となる。

一般に、アクセス・ポイントは多ければ多いほど情報システムの利用方法が柔軟になるため、多数のアクセス・ポイントを持つシステムが歓迎されるが、一方アクセス・ポイン

トを増やすと、比例級数的にファイル容量が拡大するため、費用対効果の観点から、ある程度制限を加えることも必要である。

#### 5.1.6 機密保持

情報システム中に収録される情報の中には、利用者を制限しないものと制限の必要なものがある。例えば、企業内情報システム等では、企業機密に関する情報は、重役や一部の担当者にしか公開できない。また、情報システムの一部を外部に公開するような場合は、外部利用者に公開できる情報と公開できない情報、管理者のみが見ることの出来る情報、というように何段階かのアクセス・レベルを設定することが必要である。

この点に関する情報システムの機能の定義としては、何のために、どのレベルの機密保持が必要であるかを決定することである。

### 5.2 データベースの持つべき内容

5.1で定義された情報システムの持つべき機能の中で、データベースの内容に直接係わる部分については、更に詳細な分析が必要である。

データベースの内容について、明確にしておかなければならない事項としては、以下のようなものがある。

- (1) 収録情報の属性
- (2) 収録情報の範囲
- (3) 収録情報の量
- (4) 利用者の属性と数
- (5) 収録情報の内容
- (6) 品質管理の方法

#### 5.2.1 収録情報の属性

構築しようとするデータベースに収録する情報の種類とデータの属性を決定する。情報の種類と属性には、一般に以下のようなものがある。

- (1) 情報の種類
  - 二次情報
  - 一次情報
- (2) データの属性

文字

数値

画像

音声

以上のような選択肢の中から構築しようとするデータベースに該当するものを選定する。データの属性では、一つのデータベースの中で二つ以上のものがともに含まれることもある。

### 5.2.2 収録情報の範囲

データベースに収録すべき情報の範囲を明確にする。この限定のし方は、5.2.1で決定された収録情報のタイプによって異なる。以下に、書誌データベースの場合の例を示す。

書誌データベース（二次情報／文字データ）

- (1) 主題領域：対象とする主題領域の範囲を決定する。
- (2) 対象文献のタイプ：図書、雑誌、新聞、会議録、報告書、特許、学位論文、公文書等のどのタイプのものを収録するか決定する。
- (3) 対象文献の言語：対象とする文献の書かれている言語を指定する。
- (4) 地理的範囲：対象とする文献の出版されている地域の範囲を指定する。
- (5) 時間の範囲：対象とする文献の出版年の範囲を指定する。

### 5.2.3 収録情報の量

5.2.2で指定した対象情報の数を積算する。例えば、日本で出版された日本語の工学分野の雑誌論文と会議録と報告書をすべて収録すると決定したならば、一年間に出版される該当文献の数がどの位あるのかを積算する。この数が、データベースに収録される一年分のレコード件数となる。

### 5.2.4 利用者の属性と数

想定されたデータベースの潜在的利用者の属性と数を推定する。工学分野の日本語文献の書誌データベースであれば、想定される利用者の属性は、工学および周辺分野の研究者、技術者、学生等である。この属性をもつ人口が何人位存在するのかを推計することにより、潜在利用者の数を把握することができる。

### 5.2.5 収録情報の内容

データベースに盛り込むべき情報の内容を明確にする。前述の工学書誌データベースを例にとると、想定される情報の内容には、以下のようなものがありうる。

標題，雑誌名，著者名，著者の所属機関，出版年，ページ，主題（分野／キーワード），内容の要旨等。

これらのどの要素を収録するのか，又それ以外にどのような要素が必要であるのかを考慮し，決定する。

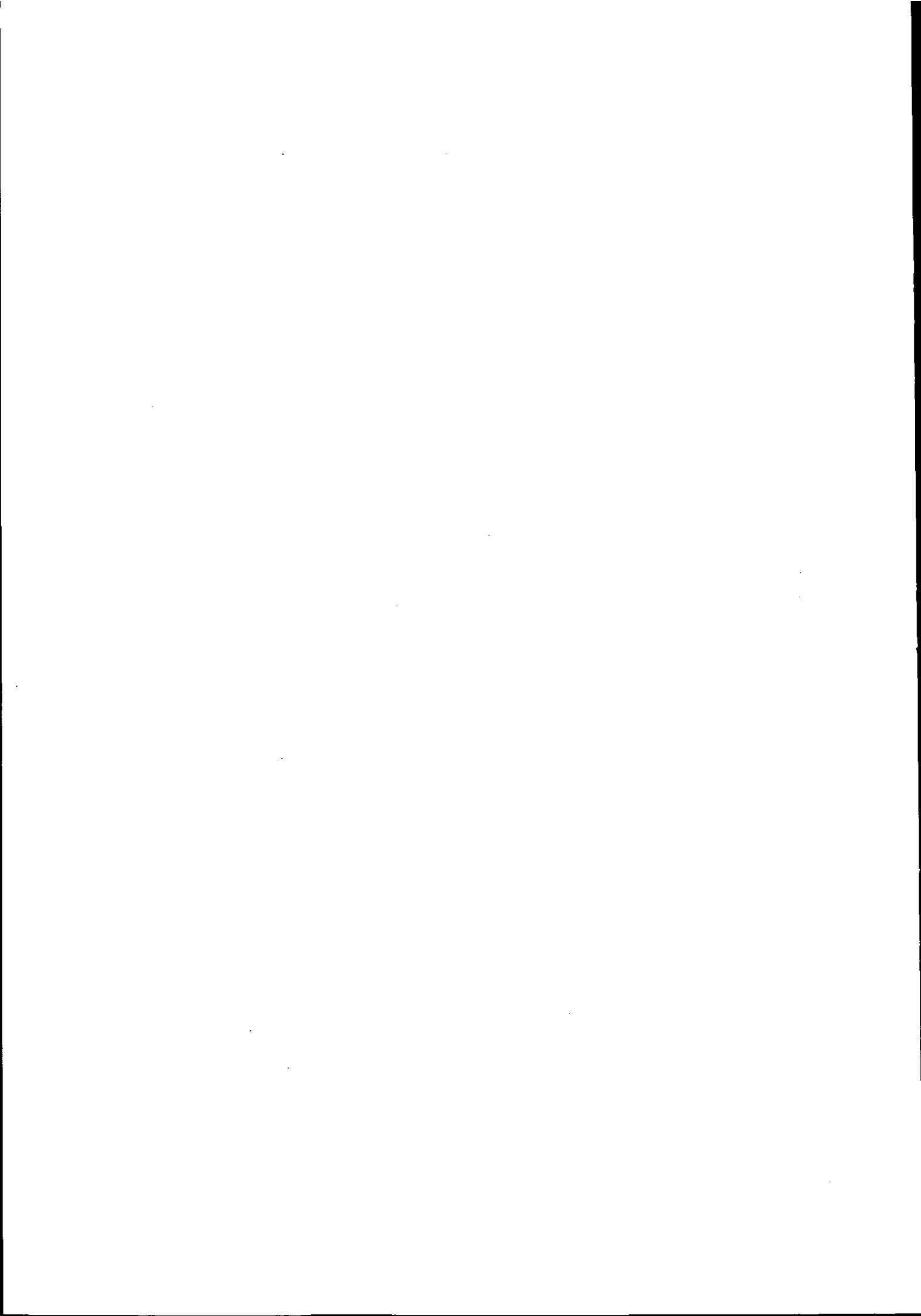
### 5.2.6 品質管理の方法

データベースの内容を質の高いものとして維持していくことは，データベースの利用効果を高めるためにも，又データベースの情報商品としての価値を高めるためにも必要なことである。データベースの品質を高める方法は，データベースの種類により異なるが，以下のような点に留意することが必要である。

- (1) 信頼性の高い情報源を選択して用いる。
- (2) 情報の発生からデータベース化するまでの時間を短縮する。
- (3) 標準化された手順で均質の情報を収録する。
- (4) 入力情報の加工，分析（前処理）に，均質でしかも質の高い人材を採用する。



### 第 3 部 分 析



## 6. データの構造化と情報分析

### 6.1 データの構造化

データベースの種別では、データベースそのものとする形態、データの形式、それに収録対象の情報やデータの性質とが混同されやすい。たとえば、「CD-ROMのデータベース」といった場合には、データベースの中でCD-ROMの形態をとるものと、CD-ROM化されたものの内容を収録したデータベースとがありうる。また、「数値データベース」は、収録されるデータの主要な表現形式が、数値であることを意味しており、これには文字形式のデータも含まれている場合が多い。データベースの対象とメディアの多様化により、こうした混乱が生じやすいが、ここでは、磁気テープなどの媒体により磁気ファイル化されたデータベースを対象とする。

#### <表現形式>

まず、表現形式すなわちデータベースに記録されるものとしては、通常は、

文字    画像    数値    音声

に分けることができる。実際のデータベースは、これらの組み合わせからなっている。

表現形式は、記録する媒体および入出力機器と密接に関連しており、画像や音声は、記録媒体、入出力機器が限られるため、現在はあまり多くはない。

また、文字や数値は、画像としても記録できる。たとえば、1冊の本を頁ごとにスキャナーやファクシミリなどで入力した時には、これは、画像と同じように扱われることになる。一方、その本に記載された文字を1文字ずつ入力した場合には、これは、文字として入力していることになる。

さらに、媒体上では、アナログによる表現とデジタルによる表現の相違がある。

#### <データベース化の対象>

データベース化の対象となりうるものは、様々であるが、例えば次のようなものがある。

印刷物（図書、雑誌記事、統計、データブックなど）

手稿（原稿など）

静止画像（地図、図表、絵画、写真など）

マイクロ形態資料

映画、ビデオ

録音物

機械可読資料

## 実物

コンピュータプログラムは機械可読資料であるが、これらを集めてデータベース化することができる。

### <一次、二次の区分>

4章で述べたように、対象をそのままデータベース化したものを「一次データベース」と呼んでいる。特にその中で、「数値」などを対象としたものについては、「ファクト・データベース」と呼ばれることがある。

一方、これらの対象に「関する」記録を収録したデータベースがあり、これを「二次データベース」と呼んでいる。現在、データベースの数では、圧倒的に「二次データベース」が多い。たとえば、文献データベースの多くは、図書や雑誌記事などについて著者やタイトルなどの書誌事項と検索手段である索引語や分類記号からなりたっており、原文書は収録していない。

一次データベースと二次データベースでは、利用目的をはじめ、データの収集、データやファイルの構造、データ項目について大きな相違がある。

一次データベースでは、その収録対象とするもの（印刷物、静止画像、録音物など）によって作成方法は、異なっている。

しかし、二次データベースにおいては、収録対象が異なっても、データ項目以外は、データやファイルの構造についてそれほど大きな差異があるわけではない。

そこで、以下では、二次データベースを中心にデータ構造を解説する。

### <レコード、データ項目、ファイル>

一般にデータベースは、次のような構造をとるものと理解されている。

第一に、データベース化の対象となるものは、同質で、個別化することができなければならない。例えば、「地球」に関するデータベースとして、普通の地図や地質図、気象データ、各国の国情の記述、「地球」を主題とする文献リストとを一つのデータベースとして作ることはできない。これらは、それぞれ別個のデータ要素からなり、データの表示方法も検索方法も異なるためである。探索する際にこれらに関連づけて探す方法は、開発されつつあるが、作成する際には、個々に同質のものとして分けられるもの、すなわち地図のデータベース、気象データのデータベース、各国の国情のデータベース、「地球」を主題とする文献のデータベースとして設計、構築しなければならない。

第二に、各レコードは、共通の複数のデータ項目からなっている。個々のレコードでは、かならずしも全てのデータ項目が出現しなくてもよい。しかし、設計時には、出現しうる全

でのデータ項目を考慮する必要がある。

同質で、個別化できる単位が、1レコードである。たとえば、人物に関するデータベースの場合、個々の人物が、1レコードとなる。また、企業に関するデータベースの場合、個々の企業が、1レコードとなり、企業に属する人物は、レコードを構成する要素、すなわちデータ項目として扱う。

通常、レコードは、データベースの場合はコンピュータ処理の上で、論理的、物理的な単位とみなされている。入力、出力、訂正、削除は、レコード単位で行われることが多い。

レコードは、データ項目から構成されている。データベースの対象の属性を示すものがデータ項目である。

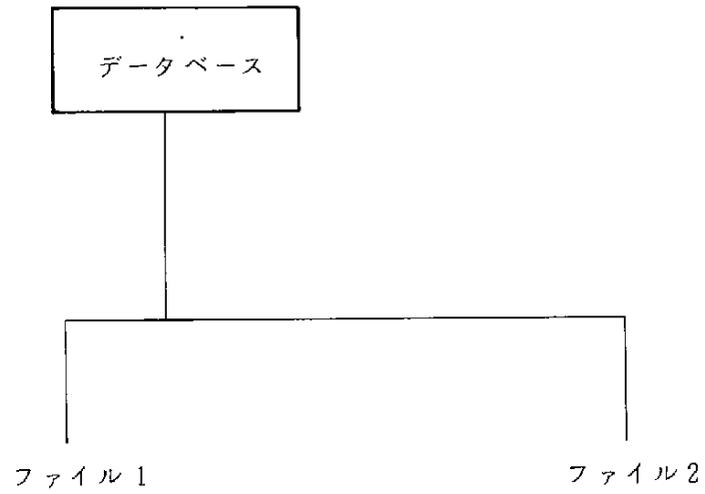
以下は、雑誌記事のデータベースのデータ項目の例である。

著者	刊行年
著者の所属機関	抄録
標題	索引語
掲載雑誌名	分類記号
巻、号、ページ	記事番号

レコードが複数個集まったものを、ファイルと呼んでいる。ファイルは、論理的には、なんらかの区分、たとえば、収録期間、収録地域によって分割されることが多い。また、物理的には磁気テープの1巻、あるいは磁気ディスク上のまとまった記憶エリアとなっていることが多い。

ファイルの集合がデータベースであるが、1ファイルのみでデータベースとなることもある。

これらを図示すると図6.1のようになる。



	レコード 1	レコード 2	レコード n
データ項目 1			
データ項目 2			
データ項目 3			
データ項目 4			
データ項目 n			

図 6.1 ファイル、レコード、データ項目の関係

### 6.1.1 データ項目

#### 1) データ項目の種類

データ項目は、次のようにグループ化される。

##### (1) 識別記号

これは個々のレコードを他のレコードと区別するもので、そのレコードのみを検索する時に必要となるものである。また、レコードの識別記号の他に、対象のもつ識別記号がある。たとえば、人物に関するデータベースの場合、データベース中で与えられた番号とその人物のID番号とがある。

データベースの識別記号の場合は、通常、数字やアルファベットが用いられる。

##### (2) コード化情報

コードを用いて表現されるデータ項目である。例えば、年月日や郵便番号はコード化情報である。コード化情報は、入力の際には、コード表によってチェックができ、また検索対象となる。

通常は、数字やアルファベット、カナが用いられる。

##### (3) 記述情報

内容の説明であり、文章形式のものである。主として、表示に用いるが、検索対象にもなりうる。漢字、カナ、アルファベットが用いられる。

##### (4) データ

フルテキスト・データベースでは、フルテキストそのもの、統計データベースであれば、統計そのものである。「二次情報のデータベース」では、この「データ」部分を持たない。

#### 2) データ項目の検討事項

個々のデータ項目について、以下のような点について検討しておかなければならない。

##### (1) 名称

まず内容を示す名称とその略称を定める必要がある。特に略称は、データを管理する上で必ず必要となるが、通常は、2～3桁のアルファベットを用いることが多い。

##### (2) データの表現形式

文字であるなら、漢字、カナ、アルファベット、数字などが表現形式である。数字は、バイナリー、デシマルなどの記録形式の区別が必要である。

##### (3) データの長さ

これは、「年月日」などのように長さが一定のもの、すなわち「固定長」の項目と、「住所」や「書名」などのように長さが不定の項目（「可変長」）とに分かれる。可変長の項目は、最大の長さを用いる。データベース内では、固定長の項目と可変長の項目とは扱い方が異なるのでこの区別をしておくことは重要である。なお長さの単位としては、「バイト」が用いられる。アルファベットやカナ、数字の1文字は、1バイトであるが、漢字1文字は、2バイトになる。

#### (4) 繰り返しの有無

データ項目の中には、同一のデータ項目が、繰り返して何回も現れる場合がある。例えば、複数の分類を行う場合にはデータ項目「分類」は、繰り返し項目である。繰り返し項目についても繰り返しの最大回数を定めておく必要がある。

#### (5) 出現

全レコードに全てのデータ項目が出現するわけではない。出現のタイプには以下のようなものがある。

A. 必須項目 : レコード識別記号のように全てのレコードに不可欠な項目

B. 準必須項目 : レコードによっては、欠如している場合があるもの。レコードにあれば必ず入力する。

C. 補助的項目 : 作成時に重要性を判断して入力するか否かが決められる項目。

#### (6) 入力方式

入力方式からデータ項目をみると、たとえば日付のように、ある1日に入力されるどのレコードも同じ値となるような、あらかじめ設定されている（デフォルト）値をとるものと、レコード毎に異なる値をとるデータ項目とがある。デフォルト値は、入力のプログラムによっては、自動的に入力できる。

また、データの形態によっては、入力機器の異なるものもある。

#### (7) チェック方法

データ項目によっては、数値項目のような場合のように、入力データの桁数や上限や下限を定めておいて自動的にチェックできるものがある。氏名などでもある程度のチェックは可能である。

また、たとえば国名のように、そのデータ項目がテーブルに登録されている一定の値以外とらない時にも、テーブルと照合することによってチェックしうる。

#### (8) D/D Dの参照の有無

DBMSには、データ・ディクショナリと呼ばれるデータベース中のデータを管

理する辞書を持つものがある。こうしたD/DDがあれば、記号などで入力しておき正式なデータに変換することや、入力データのチェックが容易に行なえる。

(9) アクセス・ポイント

通常、検索対象となるデータ項目をアクセス・ポイントと呼んでいる。データベースに入力されたどの項目も、原則としてはアクセス・ポイントとなりうる。しかし、一般に検索の際に使われる頻度の多いデータ項目については、入力時に検索しやすいように加工しておくことが行われている。

(10) 他のデータ項目との関連

前出の雑誌記事のデータベースの場合に著者が複数の時には著者の所属機関も複数になる。このような時には、著者というデータ項目と著者所属機関というデータ項目との間に関連付けを行なっておく必要がある。

これらの検討項目について表 6.1 のような一覧表を作成することが望ましい。

表 6.1 データ項目一覧表(例)

名 称	表現形式	長さ	繰り返し	出現	入力方式	チェック	D / DD	アクセスポイント	関連項目
記事番号	数字	8		必須	自動	あり		○	
著者(漢字)	漢字	可変	あり	必須		あり			所属
著者(ローマ字)	英字	可変	あり	必須		あり		○	所属
著者の所属機関	漢字	可変	あり	準					著者
論題(漢字)	漢字	可変		必須					
論題(英文)	英字	可変		必須				○	
掲載雑誌名	記号	4		必須		あり	参照		
巻, 号, ページ	記号	可変		必須					
刊行年	数字	4		必須		あり		○	
抄録	漢字	可変		補助					
索引語	英字	可変		必須		あり	参照	○	
分類記号	記号	6		必須		あり	参照	○	

### 6.1.2 レコード

レコードについては、以下の点を検討しておく必要がある。

- (1) 入力方法 (9.1 参照)
- (2) 出力方法 (9.2 参照)
- (3) 更新方法

既に作成されたレコードを追加・修正する際にはいくつかの方法がある。

A. 既存のレコードの内容に追加部分を含め、あるいは、修正してレコード全体を作り直す方法。

B. 追加、修正するデータ項目のみからなるレコードを作成する方法。

- (4) レコード間の関連

レコードとレコードに関連付けを必要とする場合があり、この際には、レコードになんらかの表示を必要とする。

A. 上記のように追加や修正のある時には、既存のレコードと追加・修正のレコードとの関連付けが必要になる。

B. 複数の部分から構成されるものと全体とを一つのデータベースに収録する場合に、相互の関連付けが必要となる。たとえば複数の報告からなるような会議録等では、個々の報告と会議録全体とを別個のレコードとする場合がある。

C. 本体と付録を別個のレコードとする場合に、相互の関連付けが必要である。

D. 逐次的に刊行され、途中で標題が変更になったものも両者の関連付けを要する。

### 6.1.3 ファイル

ファイルには含まれるレコードによって次の種類がある。

- (1) 初期ファイル：最初に作成されたレコードからなる
- (2) 追加ファイル：追加レコードから構成されているファイル
- (3) 更新ファイル：初期ファイルや追加ファイルの訂正レコードなどを含むファイル。

ファイルの作成にあたっては、以下の点に留意する必要がある。

ファイルの第一番目のレコードをヘッダー・レコードとして次のような内容を記録しておく。

- (1) データベースの名称
- (2) ファイルの名称
- (3) 著作権所有者名
- (4) 作成者の名称
- (5) ファイル作成年月日
- (6) ファイルの種類（初期，追加，更新の別）
- (7) ファイルの収録範囲

磁気テープの場合には，コピーされることを防ぐのは困難である。しかし，著作権所有者名等は，ファイルに明示しておいたほうがよい。

## 6.2 入力情報分析の手法

### 6.2.1 分析の目的と手法の種類

データベースの作成にあたって，情報をそのまま入力するだけでなく，その情報を分析することによって，その情報に付加価値を与えることができる。たとえば，その内容を要約した抄録を作成したり，また，内容を分類し，分類記号を付したり，あるいはキーワードを付したりすることはデータベースを分析し付加価値を与えることになる。これは知的な判断を要する作業であるため，多くの時間と労力を必要とする。しかしこうした付加価値を与える作業によって，データベースは独自のものとなり，その価値を高める一方，新たな検索手段や情報も与えることになり，その結果として利用範囲を広げることになる。

分析の手法は，システムの目的や対象とするデータによって様々である。例えば，経済統計に対してモデルによって分析した結果をデータベースとして提供することもできる。しかし，一方では，データベースは，元になるデータのみを提供することが望まれている場合も多く，こうした時にはデータの分析は，利用者側の作業となる。

従来からデータベース作成側でよく使われている代表的なデータの分析手法としては，索引，分類，抄録などがある。索引と分類は，文献の主題を分析し，それを言葉もしくは記号で表わしたものであり，抄録は文献の内容をまとめたものである。ここでは文献のデータベースを中心として取り上げるが，他の種類のデータベースにもこれらの分析手法は適用できよう。

### 6.2.2 索引作成のガイドライン

索引の作成方法についてはISOの指針があるので，これに従って，索引作成の一般的

な手順と基本的な留意事項をあげる。最初に索引作成作業の一般的な手順を示し、各段階において原則とされているものを説明する。そして、索引作成のツールであるシソーラスを作成するための基本的な手順を簡単に示す。

#### a. 索引作成作業の基本的な手順

図 6.2 は、索引作成作業の一般的手順と利用者の検索過程を合わせて示したものである。索引作成作業は、物理的な実体としての文献の記述（たとえば、その形態、出版社、出版年等の記述）とは別のものである。

索引作成作業は、実務上では重なる点があるが、基本的に次の三つの段階から成り立っている。

- (1) 内容を検討し、主題内容をはっきりさせる。
- (2) その主題の中の主要な概念を明確にする。
- (3) 選択された概念を索引語で表わす。

#### b. 内容の検討

##### <文献の場合>

印刷形態の文献を完全に理解するためには、できるなら本文を全て読まなければならない。しかし全文を完全に読むことは、実際的とは言えないことが多く、また常に必要であるわけでもない。ただ、どのような有益な情報も見落とさないようにしなければならない。

そこで、特に以下の点には注意を払う必要がある。

- (1) 標題
- (2) 抄録
- (3) 目次
- (4) 序論、章や節の最初の部分、結論
- (5) 挿図、図、表および図表の説明
- (6) アンダーラインが付されていたり、異なる字体で印刷されている語や語群
- (7) 注・引用文献

文献の内容を検討するために、これらの全ての要素を詳しく調べ、評価する。著者の作成した標題や抄録は、必ずしも内容を適確に表わしていないことが多いために、標題のみから索引を作成することは、望ましくない。また、抄録は本文を検討するための代わりものとしては用いない方がよい。

##### <文献以外の場合>

実物を含む視聴覚、映像、録音メディアのような印刷されていないものの場合、ひとつ

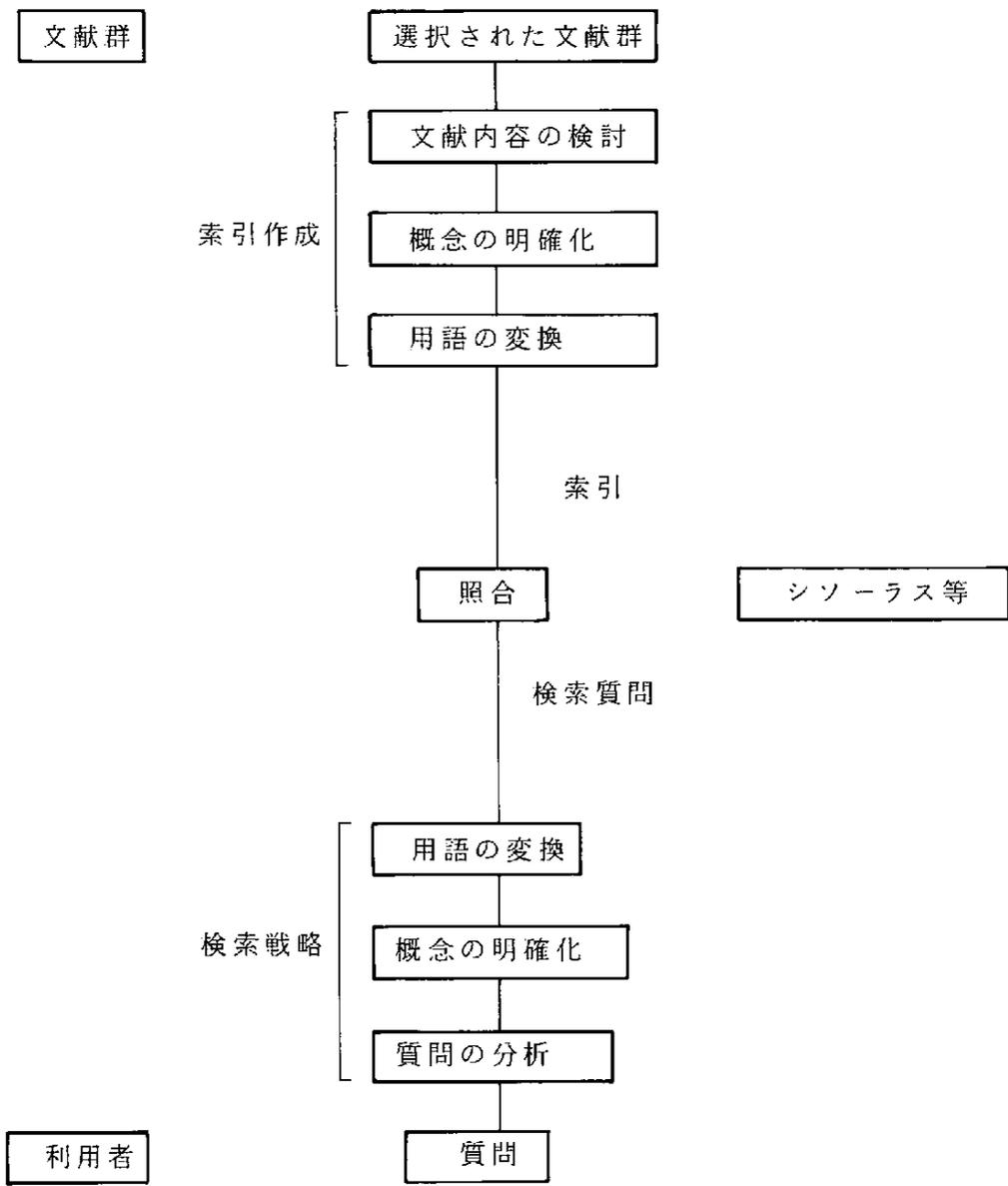


図 6.2 索引作成の一般的手順

出典 : Lancaster, F.W. Information Retrieval System.  
2nd.ed.を一部修正

の対象を全体にわたって（たとえば映画を上映するように）詳しく調べることは、実際上は難しい。

通常は、標題や内容を要約したものを対象として行なうことが多い。

#### c. 概念の明確化

内容を検討した後に、その主題を構成する主要な要素である「概念」を明確にするため、系統的なアプローチを行わなければならない。そのためには、対象としている分野で重要とみなされる構成要素のチェックリストを設定した方がよい。

こうしたチェックリストに通常含まれる一般的なチェック項目としては以下のようなものがある。

- (1) 特定の生産物、状態、現象を扱っているのか。
- (2) その主題は、能動的な概念（動作、操作、処理など）を含んでいるか。
- (3) その対象や被験者は、明確にされた動作による影響をうけているか。
- (4) この動作を行なう主体を扱っているのか。
- (5) その動作を成し遂げるための特別な手段（特別な機器や手法、方法）に言及しているか。
- (6) これらの因子は、特別な場所や環境に関連して検討されているのか。
- (7) どのような従属変数や独立変数が明らかになっているのか。
- (8) その主題はその研究分野とは通常は関連を持たない特別な観点（たとえば宗教の社会学的研究）から検討されているか。

これらは、どのような主題分野においても適用されそうな一般的な例として取り上げたものであるが、それぞれの分野では、他のチェック項目を設定する必要がある。

内容の検討を通して明らかにされた全ての「概念」を、索引を作成するために取り上げる必要は必ずしもない。「概念」の選択は、その索引の利用される目的によって異なるので、その目的を念頭において、その利用者集団にとって最も適切とみなされる「概念」を選べばよい。

検索質問からのフィードバックによって、索引作業のツール方法とを変更した方がよいこともある。

通常、索引作業においては網羅性と特定性が最も問題になる。

#### <網羅性>

網羅性とは、主題を構成する重要概念がどの程度、索引されているかを示すもので、一つのものに付与された索引語の数とみなすことができる。一つのものに与えられる索引語

の数が、多ければ、網羅性の高い（「索引の深度が深い」ともいう）索引をしていることになる。

この網羅性を重視するなら、そのデータベースの利用者にとって、潜在的に価値のある全ての概念を示しておいたほうがよい。また、データベースの利用層をあまり狭く考えないほうがよい。例えば、科学者とか技術者のために作成されたデータベースは、他の利用者グループ（例えば、経済学者）によっても利用されることもありうるので、科学・技術文献の場合は、その社会的・経済的側面のような、主題の他の側面を見落さないようにすることが望まれる。

また、ひとつの文献に付与することのできる索引語やディスクリプターの数をあらかじめ制限せずに、その文献中に含まれる情報量によって決めた方がよい。

#### <特定性>

特定性は、文献中に生じる特定の概念が、シソーラスなどの中で、どの程度正確に特定されるのかと云うことである。たとえば、「パーソナル・コンピュータ」を扱っているものを索引しようとした時、シソーラスには「コンピュータ」しかなければ、特定のでない索引語（つまり「コンピュータ」）を選ぶことになる。つまりシソーラスに特定性がないと、特定の概念がより一般的な意味をもつ語で表わされることになってしまう。

一般的に、概念はできるだけ特定の扱われた方がよい。

ただし以下のような場合には、例外的により一般的な概念を用いる。

- (1) あまりに特定のなものすぎて、逆に検索が困難になると考えられる場合。例えば、機器の非常に特殊なモデルは、その製造会社名や、モデルの属する製品名のようなより一層的な語で表現した方がよい。
- (2) 著者によってその概念が、たまたま引き合いに出されただけと考えられる場合。

#### d. 索引語の選択

自然語から索引語を選択する際には、その分野の辞書や百科事典、テキストブック、雑誌・抄録誌の索引などを参考にする。

シソーラスの中の語を用いて、概念を索引語に変換する時には、与えられた概念を表現することができる最も特定の用語を選択する。実際には、しばしば既存のシソーラスなどには表われていない概念に出会うこともあるが、こうした概念は、索引システムによって、様々な方法によって処理されている。たとえば、以下のような方法がある。

- (1) 直ちに、シソーラスなどに入れる
- (2) 一時的に、より一般的な用語で表わし、新しい概念は、後にシソーラスに付加

するための候補としてあげておく。

#### e. 質の管理

索引の作成の一貫性を保ち、その質を高めるためには、以下の点に留意する必要がある。

##### <索引作成者の資質>

- (1) 索引作成者は、不偏的、客観的に索引作業を行なわなければならない。
- (2) 索引作業の一貫性を保つことに配慮しなければならない。索引作業基準を遵守し、シソーラス中のスコープ・ノートなどの規制をよく知っていなければならない。

多人数での索引を作成する時には、一貫性を保つことがより困難となるが、探索する側からの索引作成者に対するフィードバックと集中化した点検作業が必要となる。

- (3) 索引作成者はできるだけ、索引を行なう文献がカバーしている分野の専門家であることが望ましい。シソーラスの様々な規則などとともに、文献中の用語を理解していなければならない。

##### <索引作成のためのツールなど>

- (4) シソーラス(索引言語)には、新しい専門用語や用語の変化を迅速にとり入れ、また利用者の新しい要求に応じなければならない。
- (5) 索引作成作業の質は索引したものを分析して評価する。たとえば再現率や精度を算出する。また、直接、利用者に接触することによって選択した用語あるいはディスクリプタが間違った連結を生じることになるかどうか、検索の段階でノイズを生じることが多いかどうかを確認する。

#### f. シソーラスの作成手順

シソーラスは、索引作業の中で「概念」を用語へ変換する時、および主題からの探索を行なう際に使うツールである。

現在、シソーラスを用いているデータベースはそれほど多くはない。また、一般的に利用者はデータベースを提供する側が期待するほどシソーラスを利用しない。

しかし、米国や英国の質の高いデータベースは、いずれもシソーラスを作成、維持している。

以下に、シソーラス作成の全般的な手順を概説する。

図 6.3 は、シソーラス作成の全般的な流れを示したものである。

最初に、システム全体について、利用者層であるとか、対象範囲、データベースの収録

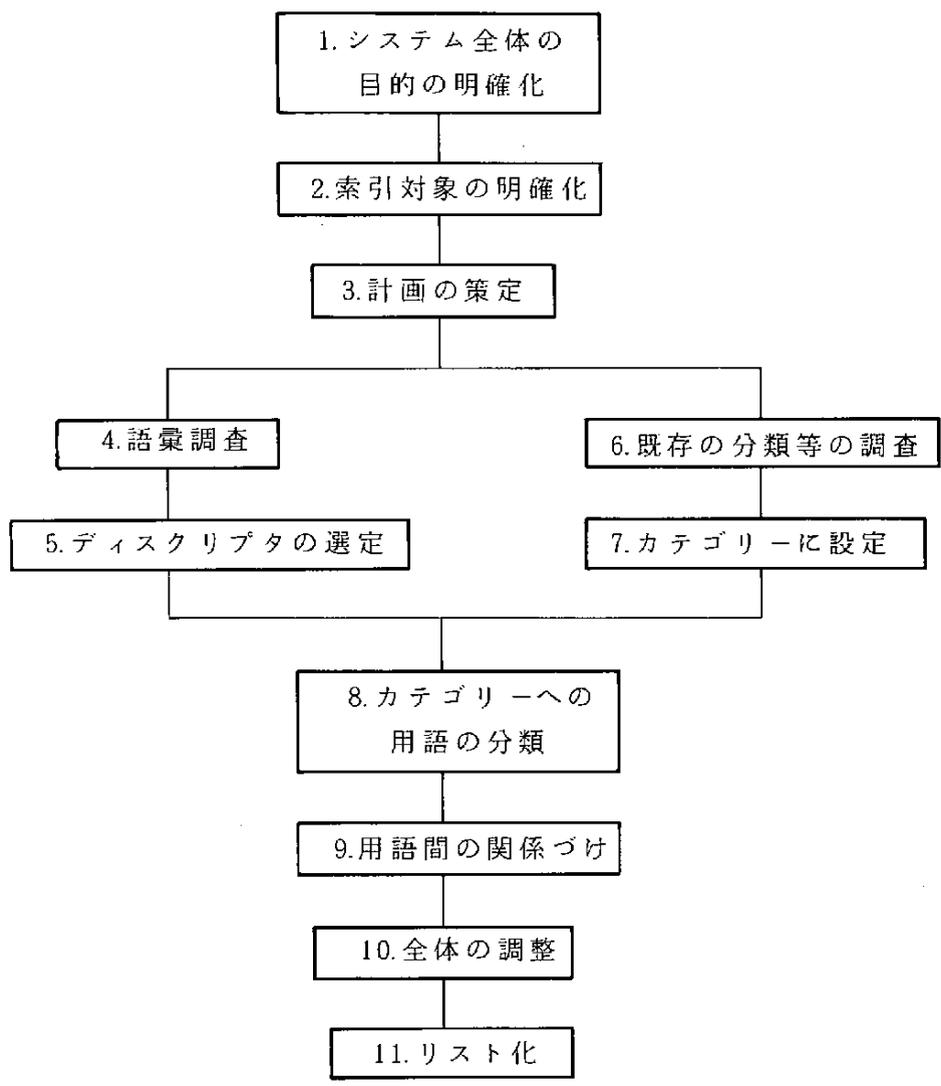


図 6.3 シソーラス作成の手順

項目などを明らかにしておく。

次に、索引方針を決定する。索引方針とは、

索引対象とする文献の種類やその言語

1 文献に付与する索引語の数

かけることのできる時間とコスト

などである。その上で、シソーラス作成の具体的計画がたてられる。

実際の作業としては、

1. 語の選択

2. (語を段階的に秩序づけるための) カテゴリーの設定

を行なわねばならない。語の選択にあたっては、網羅性のある語彙の収集と語の出現頻度の調査が必要となる。この語彙調査の結果から、シソーラスに採用する語(ディスクリプタ)を決定し、非ディスクリプタとの参照を作成する。

一方、カテゴリーを設定するためには、既存の分類などを調査する必要がある。その調査結果を参考として、仮のカテゴリーを設定する。その際、当該分野だけでなく他の分野をも包括したカテゴリーと地名などの汎用カテゴリーを別に作っておく。

次にディスクリプタとして選択した語をカテゴリーに分けてみて、ディスクリプタとカテゴリーの不備を調整するとともに、ディスクリプタを複数のカテゴリーに入れるか否かの方針をたてる。そして、個々のカテゴリー内のディスクリプタの上位、下位などの関係づけを行なう。全体の調整を図った上で、五十音順とカテゴリー別のリストを作成する。

### 6.2.3 分類の一般原則

#### a. 分類とは

分類は、知識を何らかの論理的な順序によって配列することと、そのようにしてできた分類表によって資料やその他の対象を配列する技術との二つのことを意味している。

#### b. 既存の分類表の基本的な考え方

現在までに多くの分類表が考案されているが、一般に次のような二つのタイプに分けられる。

#### <階層構造, 列挙型分類>

この分類では、まず知識の世界を全体としてとらえ、各段階の基盤となるような特性によって、その全体を類もしくは、下位区分の段階へと分けていく。概して、一般的なものから特定のなものへと進み、階層的な構造を形づくり、各段階は、上の段階よりは特定の

であり、その下の段階より一般的である。それぞれの段階の構成要素は、通常は相互排除的なカテゴリで、他のものに対しては同格となり、相互の関係の類似性によって配置されている。たとえば、文学という主題は言語、形式、時代という特性で表わされ、そして、文学—英文学—英国の詩—エリザベス朝時代の詩というように、一般的なものからより特定のものへと進んでいく。これが階層構造型分類である。また、従来の多くの分類表では、全ての主題とその下位区分を列挙し、記号を与えてきた。そこで、これらの分類表は列挙型分類とも呼ばれる。日本十進分類法（NDC）はこの種の代表例である。

#### <ファセット分類>

最近の分類理論では、ファセット分析とその統合とに重点がおかれている。これは、一つの主題を構成部分に分析あるいは分割しておき、分類対象に応じて、各構成部分を組み立て直すものである。

この考え方では、一つの階層構造中に全ての主題を列挙するかわりに、分類表は各学問分野や主要なクラスのもとに、ある特性やファセットに応じて分離された基本概念や要素を掲げることで、主題の基本的な構成要素を示している。さらに、形式、地理、時代のように類出する区分は、全てのクラスに適用するために別箇に掲げられる。こうした分類表を適用すると、分類作業は基本的に分類対象の文献の主題を構成している要素を見つけ出し、組合わせる作業となる。構成要素は、個々のクラスについてあらかじめ定められた順序に従って組み合わせられる。こうした原則にもとづく体系をファセット分類あるいは分析統合型分類と呼んでいる。国際十進分類法（UDC）はこの例である。

それぞれの分類表は、類と区分を表わす符号として記号の体系を持っている。分類記号法と呼ばれるこうした手段を用いるのは、たやすく主題を指示し、また主題間の関連を表わし、さらに論理的に配列するためである。記号法には、通常は数字か文字、あるいはその両方を用いる。分類の体系的構造や階層を反映した階層的記号法や、主題間の関連を表わした表示的記号法がある。また、この他に助記性を持たせた記号法がある。これは、分類の上で同じ内容のものが繰り返される時、同じ記号で表わすという方法である。たとえば、日本十進分類法では、日本は数字の1で表わす。

分類を利用する際には、既存の分類表を流用する場合と新たに分類表を作る場合とがある。

原則として、よく利用されている、あるいは知られている分類表がある時には、既存の分類表を使用したほうがよい。また、一部を拡張、変更して用いることも考えられる。既存の分類表は、一般に不備な点が多いが、利用者がよく知っているという利点は生かした

ほうが無難である。また、分類表の作成は、シソーラスの作成と同程度の時間と要員とを必要とする。

### c. 分類作業の一般的原則

分類作業は、分類対象となるものの主題内容を明らかにし、主題概念を見つけ出すという作業から始まる点では、既に述べた索引作業と同じである。索引作業では、概念は何らかの言葉に変換され、分類作業では、分類記号法による記号が使われると考えるのが普通であるが、分類記号を索引語の一種と考えれば、分類作業も索引作業のなかに含まれる。

分類と索引では、前者では記号を用い、後者ではことばを用いるという相違の他に分類作業では与える記号の数が制限されている点で、比較的多数の索引語を付与する索引作業と異なっている。

以下は、分類記号を一つだけ選定する際の一般的な指針である。

- (1) 文献が分類中の複数箇所に分類されうる時は利用しやすさを考慮して一つを選ぶ。
- (2) 形式よりも主題で分類する。
- (3) 最も特定の分類記号を用いる。
- (4) 分類表に付された索引のみを用いて分類しない。

また複数の内容を持つものを分類する場合の分類記号の選定については、明確な規則は存在しないが、一般的な留意点を示しておく。

- (1) 最も重点の置かれている主題のもとに分類する。

主題が個別に扱われている場合は、それが占めるスペースの大きさが手掛りとなる。

異なる主題を相互の関連性の点から扱ったもの場合には、その関連の仕方を検討して、重点を置いている面を明らかにする。

- (2) 冒頭に置かれた出題のもとに分類する。

重点のおかれた主題が決められない場合、たとえば二つ以上の主題を個別に扱った著作や、特に重点をおいたものを示すことなく比較している場合には、冒頭におかれた主題で分析する。

- (3) 広い主題のもとに分類する。

広い主題の下位区分であり、共に広い主題の主要部分を構成している二つないし三つの主題を扱った著作は、より広い主題で分類する。

## 6.2.4 抄録の作成

### a. 抄録とは

抄録とは、文献（記事）の内容の概略を迅速に把握する目的で作られた文章で、主観的な解釈や批判を加えず、文献（記事）の重要な内容を簡潔、かつ性格に記述したものである。

抄録には、一般に報知的抄録と指示的抄録の2種類があるとされている。

報知的抄録とは、もとの文献の内容について、結果や結論を含めて記述した抄録で、もとの文献を読まなくても、内容の要点が理解できるように作られたものである。

これに対して、指示的抄録とは、もとの文献の主題や扱っている範囲を説明した抄録で、もとの文献を読む必要が有るか無いかを、判断するのに役立つように作られたものである。

### b. 抄録の要件と作成における留意点

抄録の長さは、和文で200~400文字、欧文で100~200語が標準である。

抄録の作成にあたっての、一般的留意事項として、S I S T 0 1では14点を挙げているが、それを抄録の内容に関するものと表現に関するものに大別して、まとめると以下のようになる。

#### <抄録の内容に関する留意事項>

- (1) 著者が読者に伝えたい内容を重点的に取り上げる。
- (2) 常識的な内容は排除する。
- (3) 標題の内容に繰返しは避ける。
- (4) 主題の取扱い方を明示する（例：…を理論的に考察する，…の調査を行なった）。

#### <抄録の表現に関する留意事項>

- (5) 主観的解釈や批判を加えず、客観点に書く。
- (6) 字数の制限を鑑み、簡潔で明確な表現をする。
- (7) 一人称代名詞、および類似の主語（例：当研究所）は使わない。
- (8) 図・表は原則として使わない。
- (9) 図・表・数式番号などは引用しない。
- (10) 数式、化学式は使ってもよい。
- (11) 原則として、原記事で使われている専門用語を使う。
- (12) 略語、略称、略号は常識的なものを除いて、初出の箇所で説明を加える。
- (13) 単位記号、量記号は原記事に使用されているとおりに使う。
- (14) 商品名は、内容の理解に不可欠な場合に限り使ってよい。

c. 抄録の標準的な内容

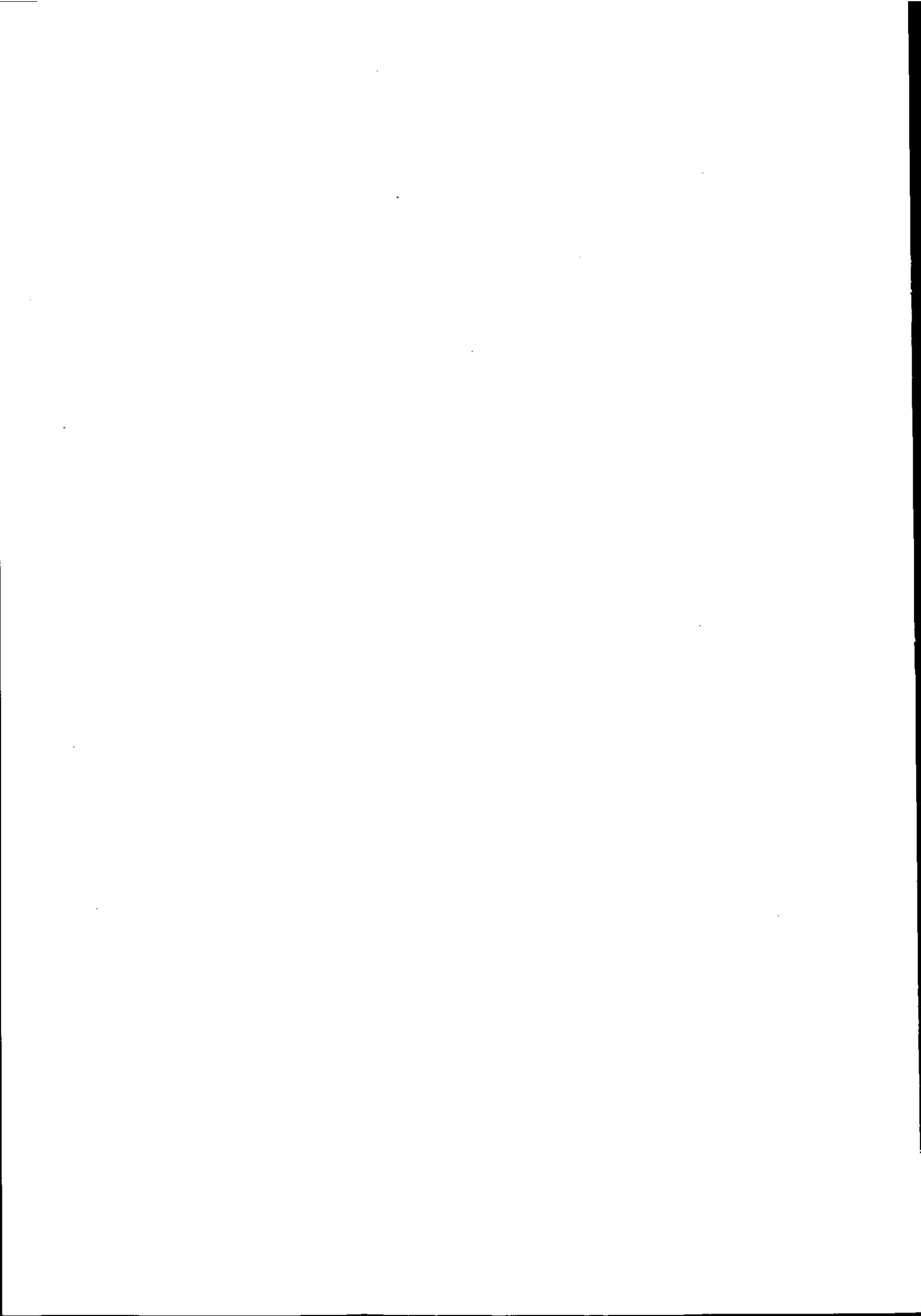
抄録は原記事の種類によって作成の仕方が異なる。

以下は、論文を例にした場合である。

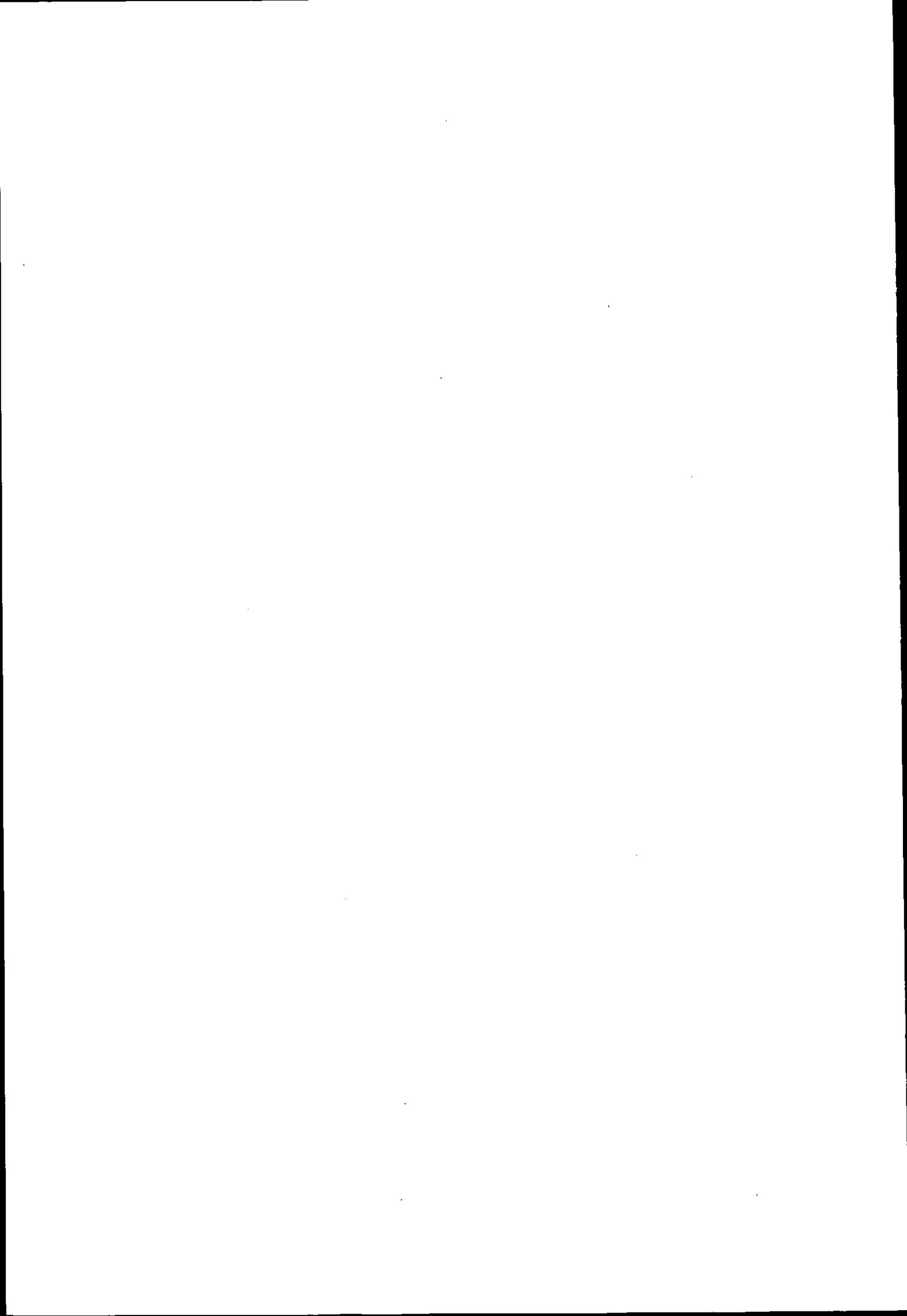
論文には報知的抄録をつけることとされており、特に以下の3点を詳しく記述する。

- (1) 研究、開発、調査の目的、取り扱っている主題範囲。
- (2) 用いた理論、方法、対象、条件等。
- (3) 実験的・理論的結果、認定された関係、得られた効果など。

その他に、研究・開発などの背景や経緯、結果の考察や今後の課題、主目的以外で重要な情報については、簡単に触れるか、場合によっては省略してもよい。



## 第4部 計 画



## 7. DBMSの選定・開発

### 7.1 DBMSとは

データベースの構築には、いくつかの方法があるが、直接にコンピュータのファイルに入力することもできる。DBMS「データベース管理システム(Data Base Management System)」は、データの入力、検索、出力などの管理を行うソフトウェアである。DBMSには、下記のように様々な種類があり、商品化されており、これらの機能も充実しており、新たにデータベース構築用のソフトウェアを開発する必要はなくなっている。

データを管理するためのソフトウェアとして当初はCOBOLに簡単なユーティリティ・プログラムを付与したソフトウェアが用いられていた。しかし、同じデータの入っているファイルを使っている場合、プログラムごとに定義を行なう必要があり、また、項目をふやしたりして、ファイルの中の構造を変えるたびに全ての関連するプログラムを変更する必要があった。不便であった。

そこで1960年代の中頃から後期にかけてデータの集まりをデータベースとして認識し、その集まりを「統合的に(重複や矛盾なく)管理されうる」専用のDBMSが盛んに開発されるようになった。

DBMSとJOIS, DIALOGなどの情報検索システムとは機能が異なっている。情報検索システムは、一度蓄積した後は変更されないファイルを高速で効率的に検索するような機能を持っている。一方、DBMSではデータの入力ができ、蓄積されたデータの追加、更新ができ、また利用者が独自にデータベースの構造を決め、データを入力できるようにしている。

一般にDBMSには、次のような特色がある。

- (1) データがプログラムから独立している。つまりプログラム作成時に、データがどのように記憶装置の中に配置されているのかを考慮しなくてもよく、データの配置を変えてもプログラムを変更しなくてもよい。
- (2) データを定義するデータ定義言語、データの検索や操作を行なうデータ操作言語、特に経験や知識の乏しい一般利用者向けの問合せ言語、それにデータ操作言語を用いて作成する定型処理を行なうための応用プログラムなどからなる。
- (3) 機密、データの一貫性を保持し、事故の際の回復処理、異なる利用者が同時に同じデータにアクセスするときのデータの保護を行なう機能を持つ。
- (4) データベースを作る時に対象となる「情報」を実体(entity)、属性(attribute)、

定義域 (domain), 関係 (relation) などの概念でとらえる。実体とは, 実在し, 識別可能なもの (例えば, 人間) であり, 実体は, 属性 (名前, 身長など) を持ち, 各属性は各実体について, 定義域の一つの値をとるとする。また実体間には, 何らかの関係がある。

DBMS というデータベースを論理的に表現するために, いくつかのデータ構造が考えられている。最近では,

- (1) 階層モデル (木構造)
- (2) ネットワーク・モデル (網構造)
- (3) リレーショナル・モデル (表形式)

の三種類に分けて論じられることが多い。

初期のDBMSには, 代表的なものとして, SDC社のTDMS (Time-shared Data Management System) (階層モデル), GE社のIDS (Integrated Data Store) (ネットワーク・モデル) などがあった。

しかし, 階層モデルやネットワーク・モデルにもとづくDBMSには以下のような大きな問題があった。

- (1) データの論理的な (利用者の) 視点と物理的な (記憶装置上の) 視点との区分が不明確であったため, 応用プログラムの作成者は, 目的とするシステムの論理的なデータ構造を考えるだけでなく, そのデータが記憶装置上にどのように記憶されているかということにまで常に配慮しなければならなかった。
- (2) 集合処理機能をもっておらず, 同一の属性を有する複数のレコードを同時に処理するコマンドがなかったため, 応用プログラムの作成者は, 常に頭の中に繰り返し処理を描きながらプログラムを作成せざるをえなかった。
- (3) 利用者が自由な形でデータベースへアクセスする必要性が十分に認識されていなかったため, 検索をするためには, 記憶装置上のデータがどのように関連づけられて配置されているかという経路をたどっていく手順を知っていなければならなかった。したがって, 一般利用者は, あらかじめ決められた形の検索しかできず, 自由に条件を与え対話形式で問合せることは難しかった。

以上の問題点のため, 階層モデルやネットワーク・モデルにもとづくDBMSは, 必要性は認識されてはいたが, 実際の業務への導入は, プログラム開発と保守作業面での負担が大きく困難であった。

そこで, これらの問題点を解消し, 応用プログラム作成の生産性を高めるためにIBMの

San Jose 研究所で新たなモデルの基礎を確立するための研究が行なわれ、その成果が1970年にCodd, E.F. によってリレーショナル・モデルとして発表された。このモデルをもとにして、1976年にINGRES, 1981年にSQL/DS (Structured Query Language / Data Base) などが開発された。

リレーショナル・データベース管理システム (以下RDBMSと略す) の特色としては次のような点をあげることができる。

- (1) データ構造が表形式で表現される。すなわちリレーショナル・モデルの集合的な構造では、 $n$ 項関係になっているが、実際のRDBMSでは分かりやすく表現するために2次元の表で表現されている。この表自体が「関係」であり、横方向の行をレコードあるいはタプル (tuple), 縦方向の列を項目あるいは属性と呼んでいる。この表はデータの論理構造を抽象的に表現したものであり、行の上から下への順序や列の左から右への順序には意味はないとされている。

このような表で示すことによって、応用プログラムの作成者も一般利用者もデータの論理構造が簡単に把握できることになり、また記憶装置上のデータ配置を最適化することをRDBMSに全面的に委ねることが可能となった。

- (2) 「関係」の操作機能については制約 (SelectあるいはRestrict), 射影 (Project), 結合 (Join) の三つの操作ができるようになっている。制約は、特定の表の中から1つ以上の制約条件をつけてレコードを抽出し、新たな表を作成する機能である。射影は特定の表からいくつかの項目を抜き出して新たな表を作成することであり、結合とは複数の表から共通の項目を元にして、新たな表を作成する機能である。

これらの機能では、表全体がまとめて集合演算の対象となり、プログラム上での繰り返し処理を不用としている。

DBMSには、最近では、「マルチメディア・データベース」とよばれるような多様な形態のデータを収録するDBMSの考え方がでてきている。

## 7.2 主な市販 (汎用) DBMS

現在、使用されている主なDBMSを表7.1に示した。

DBMSの選定にあたっては、以下の点に留意する必要がある。

- (1) 適用機種
- (2) 価格

表 7.1 主要DBMS

( 行政管理庁管理局「情報検索システムの設計・運用に関する調査研究報告書」( 1983 )などを参考に作成 )

名称	開 発	国内販売担当	運用機種	データ構造	利用 者 インターフェース	販 売 条 件 買取 リース/レンタル	ドキュメント	講習会	メンテナンス	テクニカルサポート
Model-204	Computer Corp. of America	三井情報開発 (1979-)	IBM IBM互換	関係	Host .L User.L	約 2,000 ~ 約 60- 4,000万 80万/月	有償	無償 (一部 有償)	0.751/ 月	有料
IMQ	NEC (1976)	NEC (1976-)	ACOS	関係	Host .L Natural	なし 約20万	一定範 無償	無償 (一部 有償)	無料	有料
ADABAS	Software A.G. (1971)	Software A.G. Fareast (1976-)	IBM IBM互換 Siemens	関係	Host .L Natural SCRIPT	約 4,000 万円	有償	有償		
QBE	IBM	日本IBM (1978-)	IBM	関係	QBE	約 7万 / 月	有償	有償		
AIM /RDB	富士通	富士通	FACOM Mシリーズ	関係	Host .L	約50万 /	有償	有償		
System 38	IBM	日本IBM (1979-)	System38	関係		約3,000万 約90万 /	有償	有償	無料	
RDBI	日立	日立	HITAC Mシリーズ	関係	Host .L					
ADM	日立	日立 (1976-)	HITAC Mシリーズ	階層	Host .L ACE		有償	有償		
MUMPS	DEC	日本DEC (1979-)	PDP /11	階層		無償	有償	有償		
ADBS	NEC	NEC (1976-)	ACOS	ネット ワーク	Host .L RIQS	ACOS-6は無償	有償	無償		
AIM /NDB	富士通	富士通 (1977-)	FACOM Mシリーズ	ネット ワーク	Host .L	22 ~ 50万 /	有償	有償		
DMS II	Burroughs Corp.	パロース (1973-)	パロース	ネット ワーク	DMS II	約 300 ~ 約 10 ~ 800万 23万/月	有償			
DMS 1100	Sperry Rand Corp.	日本ユニパック (1970-)	UNIVAC	ネット ワーク	QLP 1100 DSS 1180 AMDP 1100 PD1R	一部機種 のみ有償	一定 無償	無償		
DMS	日本ユニパック	日本ユニパック (1980-)	OURシリーズ 80	ネット ワーク	Host .L UNIQUE EASY	約 4.2万 / 月	一定 無償	無償		
TOTAL	Cincom System Inc.	シンコムシステム ジャパン (19 -)			Host .L T-ASK MARX-IV	約 1,000 ~ 約 15 ~ 2,000万 30万 /	一定 無償	無償		
PDM	日立	日立 (1976-)	HITAC Mシリーズ		Host .L ACE SCORE PSR	約 6万円 / 月	有償	無償		

(3) サポート体系

教育，マニュアル

(4) 実績

DBMS はかなり複雑なソフトウェアであるために開発されたばかりのものはたとえ機能が十分でも，初期にはその機能を発揮できないことが多い。ある程度の使用実績のあるものを選ぶほうが無難である。

(5) 機能

1. 処理量

2. パフォーマンス

検索，応答の速度など

3. 入力・修正機能

4. 検索機能

5. アプリケーション

6. データの表現形式

漢字，カナ，アルファベット，数字などが扱えることが最低限必要である。特にわが国では日本語処理は必須となっており，このため米国のDBMS は，使用できないことが多い。

7. データの長さ

「固定長」ばかりでなく，「可変長」のデータ処理が可能かどうか。

また，処理しうる最大長も問題になる。

8. 繰り返し項目の処理

9. チェック方法

10. D/D D機能

11. 他のデータ項目との関連づけ

## 8. ファイルの設計

### 8.1 データ項目の設定

データ項目の設計では、主として項目の選択と内容について検討しなければならない。

当然、収録対象によってデータ項目は異なるが、どのような対象でもデータベース化の際には、以下の点に留意する必要がある。

#### (1) データ項目の区分

基本的に次のデータ項目群から構成される。

##### A) 共通データ項目群

識別記号、タイトル、作成者、発行者、発行年などは、形態を問わず必要となるデータ項目である。

##### B) 対象特有のデータ項目群

収録対象に応じて不可欠なデータ項目がある。たとえば、地図であるなら、縮尺表示、経緯度は不可欠であり、録音物であるならば再生速度、録音・再生上の特性などを必要とする。新たにデータベースを設計、作成する際には、類似のデータベースなどを参考として、こうしたその対象に特有なデータ項目を網羅的に調べる必要がある。

##### C) 当該データベース特有のデータ項目群

以上に含まれないもので、そのデータベースの特色となりうるようなデータ項目である。

データベースの設計にあたり、各データ項目についてもこれらの中のどのグループに属するのかを検討しておく。

#### (2) データ項目の数

一般に、設計の際には、収録しようとするデータ項目は多くなりがちである。しかしながら、データ項目が多すぎるとデータ収集の労力や個々のデータ項目の質の維持の面で無理が生じやすい。

一方、データ項目が少ないと、データベースの利用価値は、小さくなる。そこで、6.1.1の2)で述べたようにある程度の数のデータ項目を収録する場合には、あらかじめ必須、準必須、補助などの区分を行っておきデータ収集の負担を軽減することが考えられる。

また、上述の共通データ項目群、対象特有のデータ項目群は、原則として含める必要がある。

### (3) データ項目の形式

データ項目の中で文字で表現されるものについては、特別な配慮が必要である。文字形式のデータ項目も検索と表示（目で読む）のために用いられる。しかし、漢字形式のデータは、現状の多くの検索システムでは、直接に検索できない。そのために、検索用に漢字で表示されているデータ項目のカナ表記、あるいはローマ字表記が必要になってくる。つまり実質的には一つであるデータ項目に対して、検索用と表示用の二つのデータ項目を用意しなければならない場合がある。

### (4) データ項目の区分

データ項目は、原則としては処理や表示方法の異なる最も細かい単位で設定する。たとえば、郵便番号、住所、住所コードは、個々のデータ項目としたほうがよい。郵便番号や住所コードは、数字か記号であり、特に入力時に郵便番号のチェックをしたり、編集の際に郵便番号と住所を分ける必要があったり、検索で郵便番号を用いたりする時には、別個にデータ項目となっているほうが処理しやすい。しかし、郵便番号を住所の一部としてしか扱わない場合もありうる。この時には、郵便番号と住所は、1つのデータ項目としても問題はない。

つまり、

A) 入力時の処理

B) 出力時の処理

C) 検索

において必要か否かを判断して、データ項目を分ける。

## 8.2 フォーマット

フォーマットは、レコード内のデータ項目の構成を示したもので次のような種類がある。

- (1) データシート
- (2) 入力ファイル・フォーマット
- (3) 蓄積ファイル・フォーマット
- (4) 交換用ファイル・フォーマット

これらの役割を図8.1に示す。

フォーマットには、もっぱら人手によって記入するデータシート等と、もっぱらコンピュ

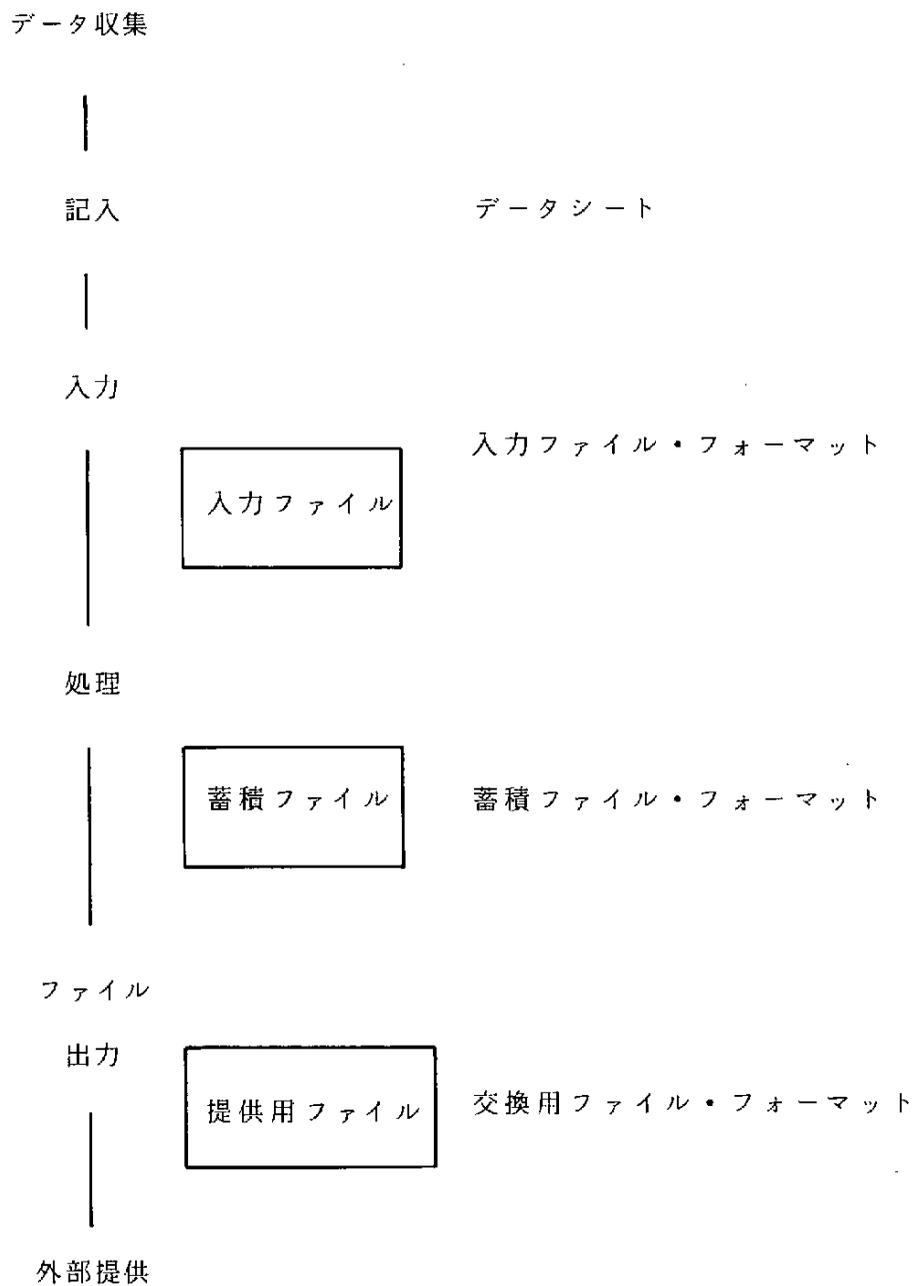


図 8.1 各ファイル・フォーマットの役割

ータ内部の記録様式である蓄積フォーマット等がある。

また、データベース作成機関の中で用いる「内部フォーマット」と、外部に提供するための「外部フォーマット」とがある。

「内部フォーマット」は、原則として外部には公開しないものであるので、独自に設計すればよい。「外部フォーマット」は、他の機関に提供するものであるために利用しやすくするために様々な配慮を必要とする。たとえば、標準的なフォーマットがあれば、できるだけそれに合わせたほうがよい。

### 8.2.1 データシート

データシートの役割は、

- (1) 入力するデータをレコード毎に一つにまとめる。
- (2) コンピュータ入力の際に用いる。

の2点である。

つまり複数の情報源からのデータをまとめる整理の機能と、入力用の機能とがある。

したがって、単一の情報源から入力する場合で、データ項目の数も少ない場合には、データシートを必要とはしないことがある。また、データ項目によっては、データシートに記入せず、あらかじめ入力過程を整備しておけば、入力対象となるものに記号などを記入するだけですむ場合がある。

また、入力にあたり、データが少量であれば、データ作成者自身が入力するシステムの場合やある程度訓練された入力要員が確保できる場合には、データシートは不用である。

一般にデータベースでは、データの量が膨大になるために外注して、バッチ入力することが多い。こうした場合には、データシートの作成は、不可決となる。また、第三者に理解しやすいように設計しておく必要がある。

原則としては、データシートを作ったほうがよいが、転記によるミスが生ずる可能性があるので、チェックの手段を設けるなどの配慮が必要である。

DBMSを使用する時には、ディスプレイ上にデータシートのフォーマットを作成して、ここに直接に入力する方法がある。通常、データシートは紙を用いるが、これもデータシート的一种である。

データシートの実例は、「資料編」(159頁)に掲載されている。

### 8.2.2 入力ファイル・フォーマット

これは、入力したデータを一時的に保存するための磁気媒体上のフォーマットである。このフォーマットは、主として入力と校正を目的としたものであるから、この目的にそった形で設計すればよい。

通常は、レコード別に、

一連番号    レコード番号    データ項目の識別記号（タグ）    データの本体

のようにデータ項目をならべておくだけで十分である。

### 8.2.3 蓄積ファイル・フォーマット

その機関の内部でデータの処理を行なうためのファイル・フォーマットである。

蓄積ファイル・フォーマットも内部フォーマットであるので、内部でのコンピュータ処理をしやすいように設計できる。

DBMSに収容する時には、DBMSの中のデータ構造がファイル・フォーマットとなる。

### 8.2.4 交換用ファイル・フォーマット

本来、データベースを提供、交換するためには、標準的なファイル・フォーマットが整備されていることが望ましい。

しかし、現在、文献用の二次データベースを対象としたISOの国際規格が存在しているのみである（10.4参照）。

こうした標準書誌データ交換用フォーマットのレコードの基本構造は、次のようになっている。

(1) レコードラベル（固定長）

各レコードの識別番号、長さ、書誌レベル、レコードの状態などを規定する部分。

(2) ディレクトリ（可変長）

各データフィールドのタグ（後述）、相対位置、長さを規定する目次にあたる部分。

(3) データフィールド（可変長）

実際のデータを収録する部分。データフィールド部は通常、幾つかの可変長の

サブフィールドからなる。

(4) サブフィールド（可変長）

個々のデータ項目。サブフィールドの先頭には、その内容を示すサブフィールド・コードが置かれている。

既存の標準的なフォーマットでは、各書誌データに対して3桁の数字か英数字のコード（「タグ（tag）」と呼ばれる）と1桁の英字（「識別子（identifier）」と呼ばれる）を与えることになっている。

## 9. 入出力の設計

### 9.1 入力的设计と入力手法

入力的设计においては、第3部で述べたような入力データの分析と決定に基づいて

- (1) データシート的设计 (8.2.1 参照)
- (2) 入力機器の決定
- (3) 入力方式の決定

などを行なう。

入力方式では、内部処理と外部処理(外注)、バッチ処理とオンライン処理、集中処理と分散処理について決定する。

#### <内部処理と外部処理>

入力は、内部処理あるいは外部処理(外注)によって行われる。外注の場合には業者の選定作業と外部用の入力仕様書の作成が必要となる。

#### <バッチ処理とオンライン処理>

データが少量の場合は、オンライン入力でもよいが、通常のデータベースではデータ量が多くなるためにバッチ処理で行なうことが多い。バッチ処理では、データシートの記入、各種の仕様の作成など多くの手間を要するが、データが大量である場合には効率的である。

#### <集中・分散>

入力を1カ所で集中的におこなう場合と数カ所で分散して行う場合とがある。特に学術的な分野では、国際協力によって各国の分担を決めて、分散入力することが多い。分散入力の場合は、詳細な入力規定を定め、マニュアルを整備しておくことが不可欠である。また、データベースの質を維持する上で、定期的に調整や評価を行なうことが必要となる。

データベースの作成の過程で、入力は以下の4段階に分けられる。

#### (1) データシートの記入・作成

各種の情報源からデータを収集し、確定し、データシートに記入する。

ただし、前述のように条件によっては、データシートを作ることは、常に必要ではない。特に、1レコードを作成するための信頼しうる情報源が一つである場合や単に転記するにとどまる場合には、データシートに記入するまでもない。

しかし、多数の情報源を用い個々のデータ項目を検討し、確定する必要がある時には、データシートを作成する過程は、不可欠である。

また、校正のためには、データシートを作成したほうがよい。

データシートの記入のためには、個々のデータ項目の記入規則と記入例とを詳細に記載した「データ記入マニュアル」を作成しておくことが必要である。

(2) データ入力

データシートをもとにコンピュータへの入力を行なう。

(3) 入力データの校正

オンライン入力でもバッチ処理であっても、データの確認、あるいは、校正作業は不可欠である。入力データのチェックは、以下のような点については、自動的にチェックを行ないうる。

1) 文字種 (数字, アルファベット, 漢字等)

2) 最大桁数, 最小桁数

3) 範囲 (最大値, 最小値)

チェック・ディジットによる正当性の確認, また, テーブル参照によって自動的にチェックを行ないうる。

しかし, 最終的には, データシートなどとの照合が必要である。

特に, 文字列に関しては, 出版の際と同様の校正作業を要する。

校正処理のために, 入力データを自動的にチェックし, 校正用のフォーマットで出力するプログラムと校正データによって修正するプログラムを作成する。

(4) 入力データの編集と蓄積フォーマットへの変換

入力データを編集し, 入力ファイル・フォーマットから蓄積ファイル・フォーマットに変換するプログラムを作成する。

## 9.2 出力の設計と出力手法

データベースの出力媒体は, 一般に磁気媒体である。従って, 出力形式は, 提供用ファイルのフォーマット, すなわち交換用ファイル・フォーマットとなる。

データベースは, 様々な用途に利用されると考えられるので, 提供用ファイルは, 利用側が扱いやすい形式になっていなければならない。具体的には, 各データ項目は, 利用側からみて, 識別しやすく, 個別に抽出できるようにしておく。さらに, コンピュータの機種によっては, 使用できなかったり, 変換に手間を要するような符号系やフォーマットを用いてはならない。

## 10. 標準化ツールの開発と利用

データベースの構成・構築に関する標準化は、全般的に未整備な状況にあるといえる。文献データベースを除いてはほとんど標準化はなされていない。

データベースに関連する国際規格および国内の規格、規準の制定状況を表 10. 1 に示す。また表 10. 2 は、表 10. 1 にあらわれる各規格、規準の標題を示したものである。これらの中で国際規格 ( I S ) と国内規格 ( J I S ) とは、日本規格協会 ( 東京, 赤坂 ) から入手できる。

### 10.1 記述

文献データベースの中で二次情報を中心としたデータベースは、通常、識別記号と、著者、標題、出版事項などのいわゆる書誌事項、および、ディスクリプタなどの検索のためのアクセス・ポイントと内容を示す抄録などからなっている。

従って、標準化の対象としては、アクセス・ポイントとして用いられる、識別のためのコード等、索引語や分類記号を与えるためのシソーラス、分類表、抄録の作成方法等が重要である。また、記述方法についても統一をはかる必要がある。

文献のデータ ( 書誌データ ) の記述では、以下の面で標準化が進んでいる。

(1) 書誌データ項目の識別方法

(2) 書誌データの記述方法

コード

書誌記述

略記法

表記法

翻字法およびローマ字化

特にデータベースの標準化においては、書誌データ項目の識別方法について努力がなされてきた。8.2.4 のような「タグ」とか「識別子」などは、一括して「内容識別子 ( content designator ) 」とよばれており、その与え方が標準化の対象となる。

元来は、米国の議会図書館で作られている図書目録ファイル US / MARC ( LC / MARC ) の内容識別子が標準とされており、英国でもほぼ同様なタグ等を用いていた。

しかし現在では、図書館関係だけでも国際的な目録情報交換フォーマットである UNIMARC、あるいはわが国の JAPAN / MARC などにおいては、異なった内容識別子を用いている。

さらに、二次情報データベースを作成する機関のためには、UNESCO の Reference

表 10.1 データベース関係の規格、基準

		国際規格、基準	国内規格、基準
データ項目の識別方法		UNIMARC II UNISIST-RM(1981) CCF	SIST 04-1983  JAPAN/MARC
データの 記述方法	コード 図書 逐次刊行 国名 日付 都道府県 市区町村 大学 書誌 文献 記述  特許 略記法 雑誌名 表記法 機関名 翻字法 中国語	IS 2108-1978 IS 3297-1975 IS 3166-1974  UNISIST-RM(1981)  (AACR II(1978)) IS 3388-1977  IS 7098	  JIS C6262-1977 JIS C6260-1973 JIS C6261-1973 JIS C6302-1983 SIST 10-1986 SIST 02-1984 日本目録規則  SIST 05-1985 SIST 06-1986
磁気テープ	物理的特性 記録様式 ラベル	IS 1864-1975 IS 1863 IS 1001-1977	JIS C6240-1980 JIS C6241-1980 JIS C6245-1978
符 号	漢字  エスケープ・ シークエンス 拡張ラテン文字 拡張キリル文字 文献用ギリシャ文字	IS 646-1973 IS 963-1973 IS 4873-1979 IS 2022-1975  IS 2375-1974 IS 5426-1983 IS 5427-1984 IS 5428-1983	JIS C6220-1976   JIS C6228-1984 JIS C6225-1979 JIS C6226-1983
情報交換用フォーマット		IS 2709-1981 CCF	SIST 03 JAPAN/MARC

Manualなどの規格が作られている。

そこで、既存の各種の規格や基準を検討し、これらの統合を図る必要が生じ、UNESCOのCCFが作成されている。

CCFにみられるように、内容識別子の標準化を図っていくと、最終的には一方ではファイル構造、他方では書誌データの記述の標準化にわたることになる。

一方、フルテキストのデータベースでは、二次情報のデータベースで必要とされるものの他にテキスト内の見出しや図表などの識別方法が標準化の対象となる。これは、現在は国際規格としては未検討ではあるが、米国の出版界とデータベース業との代表が規格の制定に着手している。

日本の図書館関係のデータベースでは、標準ツールとして従来から書誌記述のための「日本目録規則」が日本図書館協会の手によって作成、管理されている。前者は、国際標準に忠実に準拠しており、現在予備版が頒布されているが、近く本版が刊行される予定となっている。規格、基準の標題は、表10.2に一括して示している。

国立国会図書館が1981年から作成、頒布しているJAPAN/MARCは、国内で刊行される図書の網羅的なデータベースであるが、そのマニュアルは、実質的には、書誌記述とファイル・フォーマットの面で国内の標準化に貢献している。

## 10.2 シソーラス

全分野を対象とした標準的なシソーラスは存在しない。

国内では、科学技術分野においては日本科学技術情報センター(JICST)の「JICST科学技術用語シソーラス」が、産業経済分野では、日本経済新聞社の「日経シソーラス」が知られている。これらを用いている機関もあるが、最近では、独自のシソーラスを作成する傾向にある。

国外のものでは、工学関係を扱う英国の「INSPEC THESAURUS」、医学分野の米国医学図書館の開発した「MeSH」、そして、原子力関連分野を扱う「INIS THESAURUS」などが著名である。

## 10.3 標準分類

国内では、全主題に関する分類表として、日本図書館協会の「日本十進分類表」(ND C)が一般的に使用されている。ただし、特定の分野、例えば科学技術、産業経済等の分野で詳細な分類を必要とする場合には、不向きであり、また、新しい領域にも使えない。

国際的な分類とされている「国際十進分類」(UDC)も同様の弱点を持っている。  
既存のデータベースでは、独自の分類表を作成して用いることが多い。

#### 10.4 交換用フォーマット

一般に、文献データベースは、データ量が大量であるため、現在まで磁気テープによって交換されている。従って交換用の磁気テープの特性等の標準化が、一応問題となる。しかし、現在では全く支障はない。どの機種で作成された磁気テープでも他の機種で処理できる。

次のレベルとしては、符号系がある。符号に関しては、普及の点で問題は残るが、かなり整備されている。汎用の符号については、ISO/TC97とJISで規格化がなされている。しかし、現実には、IBMの符号系が広く用いられている。

また、文献データベースの面からみると漢字の国際規格の制定を必要とするが、これには、中国、台湾も関係があり、今後大きな問題となりうる。

書誌データ交換用フォーマットに関連する規格、基準としてよく知られているのは、LC/MARCフォーマットと呼ばれているもので、これは、1968年に、図書目録情報を機械可読形式で頒布するために米国議会図書館(LC:U.S. Library of Congress)によって開発された。

以後、これは、米国規格(ANSI Z39.2)となり、ついで国際規格(ISO 2709)となっている。

国内では、SIST03がこれに対応する基準となっている。

#### 10.5 規格

##### a. 国内規格

##### (1) 「国家規格(JIS)」

文献データベース関係では、ほとんど規格化がなされていない。書誌事項の一部として使われる可能性のある日付、地名などのコード、データベースによってはアクセス・ポイントとなりうる分類表がJIS化されている。

##### (2) 「科学技術情報流通技術基準」(SIST: Standards for Information of Science & Technology)

SISTは、その性格上、科学技術分野に限定されるが、この中には、書誌記述や抄録作成、あるいは機関名の表記、雑誌名の略記など文献データベースに関係の深いものを含んでいる。

b. 国際規格

(1) 「国際規格」( IS )

文献に関する情報の標準化は、ISO/TC46 (ドキュメンテーション) で扱われている。現在 100 をこえる事項の検討が行われている。国際規格となっているものの数は 40 である。なお、TC 46 の国際規格の翻訳作業が日本ドキュメンテーション協会標準化委員会によって行われている。

(2) 「UNESCO」

UNESCO は、書誌データの交換のための標準化には力を入れており、ISO や IFLA を支援しているほか、様々な標準化活動の調整を行っている。

表 10.2 主要規格の名称

〔国際規格〕

IS 646-1983	ISO 7-bit coded character set for information exchange
IS 963-1973	Guide for the definition of 4-bit character sets derived from the 7-bit coded character set for information processing interchange
IS 1001-1979	Magnetic tape labelling and file suructure for information interchange
IS 1863-1976	9-track, 12.7 mm (0.5 in) wide magnetic tape for information interchange recorded at 32 rpmm
IS 1864-1985	Unrecorded 12.7 mm (0.5 in) wide magnetic tape for information interchange -32 ftpmm NRZ1, 126 ftpmm (3200 ftpi) phase encoded and 356 ftpmm (9042 ftpi), NRZ1
IS 2022-1973	ISO 7-bit and 8-bit coded character sets - coded extension technique
IS 2047-1975	Graphic representations for the control characters of the 7-bit coded character set
IS 2108-1978	International Standard Book Numbering
IS 2375-1974	Procedure for registraction of escape sequence
IS 2709-1981	Format for bibliographic information interchange on magnetic tape
IS 3166-1981	Codes for the representation of names of countries
IS 3297-1975	International Standard Serial Number (ISSN)
IS 3388-1977	Patent documents : Bibliographic references : Essential and complementary elements
IS 4873-1979	8-bit coded character set for information interchange
IS 5427-1984	Extension of the Cyrillic coded character sets bibliographic information interchange
IS 5428-1984	Greek alphabet coded character sets bibliographic information interchange
IS 7098	Romanization of Chinese

〔UNESCO〕

UNISIST-RM(1981)	Reference Manual for machine readable bibliographic description
CCF	The Common Communication Format

〔IFLA〕

UNIMARC II (1980)	UNIMARC : Universal MARC Format , 2nd ed.
UDC (国際十進分類法)	

国内

〔国家規格〕

JIS C6220-1976	情報交換用符号
JIS C6225-1979	情報交換用漢字符号系のための制御文字符号
JIS C6226-1983	情報交換用漢字符号系
JIS C6228-1984	情報交換用符号の拡張法
JIS C6240-1980	情報交換用磁気テープ
JIS C6241-1980	NRZ-1方式による情報交換用磁気テープの情報記録様式
JIS C6245-1978	情報交換用MTのラベルとファイル構成
JIS C6260-1973	都道府県コード
JIS C6261-1973	市区町村コード
JIS C6262-1977	日付の表示(コード)
JIS C6265-1980	産業分類コード
JIS C6267-1977	商品分類コード
JIS C6302-1983	大学・高等専門学校コード

〔団体規格, 基準〕

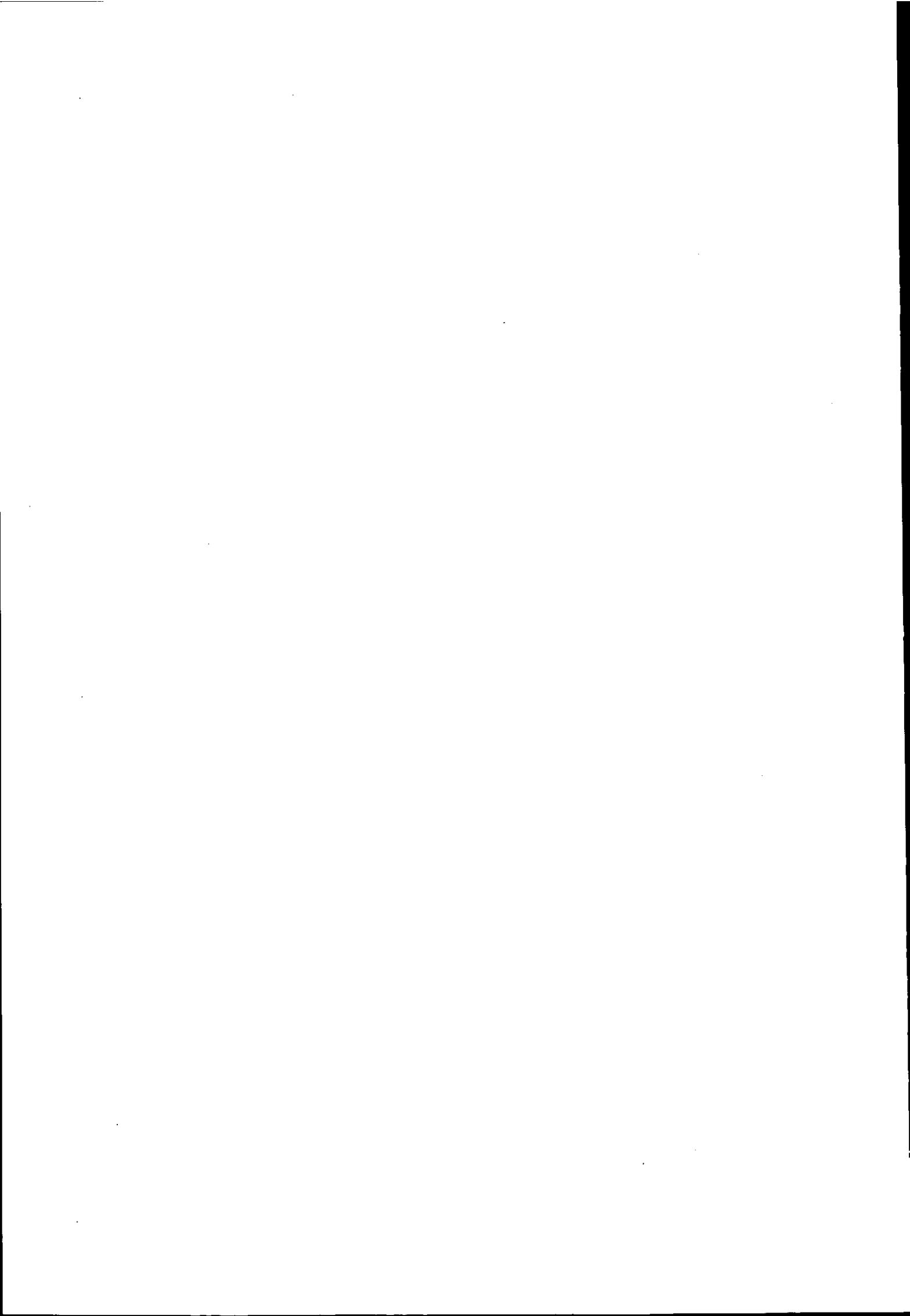
〔科学技術庁〕

SIST 01-1984	抄録作成
SIST 02-1984	参照文献の書き方
SIST 03-1984	書誌的情報交換用レコードフォーマット(外形式)
SIST 04-1983	書誌的情報交換用レコードフォーマット(内形式)
SIST 05-1985	雑誌名の略記
SIST 06-1985	機関名の表記
SIST 10-1985	書誌データの記述

〔日本図書館協会〕

日本目録規則(1977)	新版予備版
日本十進分類法	

## 第 5 部 文書化



## 11. 基本ドキュメントの作成

### 11.1 基本計画書の作成

基本計画書はデータベースの構築目的と構想を関係者及び意思決定管理者に伝達し、合意を得るための文書である。この内容によって、データベース構築の合意が得られて初めて、初期の目的に必要な人材、設備・機器、費用が得られ、データベースの構築に着手できるのである。

### 11.2 基本計画書の作成要領

本書の全体で述べられているような、データベース構築の全プロセスについての具体的な計画を、構築しようとするデータベースに関して作成し、構築の目的を明確にする。更に、データベース構築により得られる効果についても客観的な評価を行なう。

次に、構築しようとするデータベースの開発と運用の各段階において必要とされる人材、設備・機器、外注費等の全経費を積算し、構築の全工程と費用を明らかにする。

### 11.3 基本計画書の内容

基本計画書は概ね以下のような項目から構成される。

(1) データベース構築の目的

データベースを構築する目的を簡潔に述べる。

(2) データベース構築の方法

データベースを構築するために用いる方法を順序立てて示す。

(3) データベースの内容

データベースに収録する内容を項目別に示す。

(4) 想定される利用者

データベースの対象となる利用者層(属性)を明示する。

(5) 提供の方法

データベースをどのような手段や方法で利用者に提供するかを具体的に示す。

(6) データベース構築の効果

データベースを構築することによって発生する効果を具体的に示す。

(7) データベースのユニークさ

他の既存データベースと比較して当該データベースがどのような独自性を備え

ているのかを明確に示す。

(8) 緊急に必要とされる理由

そのデータベースを現時点で構築しなければならない理由を明確に示す。

(9) データベース開発の体制

データベース開発と運用に係わる、人員計画、設備・機器計画を提示する。これは、開発体制と運用体制に分け、各段階での所用期間、人材の量と質、機器を月別、年度別に表の形で示す。

(10) 費用見積り

データベースの開発と運用に係わる人材、機器、外注費等について、年度別の費用見積りを作成する。

(11) 収支計画

データベースの開発と運用に要する費用をどのような方法でどの程度の期間に回収するかについての計画を明確に示す。回収期間については年度別の回収額を表の形で示す。

#### 11.4 その他の基本文書

データベースの開発作業およびシステムの詳細についての逐次的な記録を、開発の進展に沿って作成し、保存しておくことが必要である。これらの記録文書は、システムのバージョン・アップやコンバージョンの際に必ず使用するので、第三者が読んでもわかるような一定の書式で作成し、大切に保管しなければならない。

この種の基本文書には以下のものが含まれる。

(1) システムの概略設計書

システムの概念設計・基本設計文書

(2) システムの詳細設計書

データベース、ソフトウェア、DBMS、ハードウェアの全体像と個々のサブシステムの相互関係の詳細を示した文書

(3) 各ファイルの仕様書

データベースの個々のファイルの論理的・物理的構造を詳細に示した設計文書

(4) ソフトウェアの仕様書

データベースの蓄積、検索、管理に関わるすべてのソフトウェアの詳細な設計文書。DBMSを使用する場合には、独自に開発した周辺ソフトウェアの仕様書。

(5) 打合せ議事録

データベースの計画段階から、開発、運用段階にかけての関係者間で行われた打合せにおいて決定された事項の記録文書

(6) 作業記録

データベースの計画段階から、開発、運用段階において実施された関連作業の記録

(7) メンテナンス・ノート

クレームとそれに対応して行った処理の記録

以上の文書の作成方法は、技術的には情報システムに関する一般的なドキュメンテーションと同じなので、本書では詳しくふれない。

## 12. 各種マニュアルの作成

### 12.1 前処理・入力マニュアル

データを入力するという作業は、まさしくデータベースを構築する作業であり、この作業の質がデータベースの質を大きく左右すると言える。この質を保つためにも、きちんとしたマニュアルを作成し、それによって作業上の曖昧さをなくし、個々の作業者の恣意的な判断の入る余地が全くない状態にしておく必要がある。以下に、前処理・入力マニュアルに必要な項目を挙げていく。

#### 12.1.1 データシート記入要領

データシートに何を記入すればよいのかということは、データシートを見れば大体の予想はつく。しかし、それでは厳密に正確という訳にはいかず、各項目（フィールド）に何を記入するのかということは、一般にあたりまえと思われることであっても、きちんと定義する必要がある。

記入の方法は、各項目ごとに詳細に説明する必要がある。論文標題などのように、データシート記入者が文字で記入しなければならない項目についていえば、表記方法だけに関しても、

- (1) 使用できる文字、特殊記号は何か
- (2) ローマ字の大文字と小文字はどのように使い分けるか
- (3) 日本語の促音便等に用いられる小さな文字は、それとわかるようにするためにどのように使うか
- (4) 濁点は一文字として扱うか

というように、何らかの決定を下さなければならない細かい事項が非常に数多く存在している。表現方法以外にも様々な留意点がある。マニュアルを作成する場合、これらすべてに注意を払い、考えられる限りすべてのケースに対応できるようにしなければならない。しかし、100%完全なものを初めから作るのは不可能であり、不都合な点が生じた場合には、マニュアルに追加あるいは変更を行なって柔軟に対応すべきである。だが、あまりにも頻繁な変更は、混乱を起こすので避けるべきである。

#### 12.1.2 コード表、分類表

入力作業の負担を軽くするためにも、各項目の中でコード化できるものはすべてコード

化すべきである。文献データベースの場合、コード化できるものとしては、

- ・使用言語
- ・情報源の種類
- ・掲載誌名
- ・論文の種類
- ・著者の所属機関、部署

が挙げられるが、データベースの規模によってはコード化が不可能なものもある。

分類表は、主題をコード化したものといえる。主題を表すことは、ディスクリプタ（キーワード）を用いることによっても可能であるが、あえて分類表を用いるのは、これによって、上位概念、下位概念、またその文献の学問体系の中での位置がすぐわかるという利点があるからである。また、MEDLINEに見られるクオリファイヤー（限定句）などもコード化したほうが便利であり、「医学中央雑誌」では、すでに限定句（医学中央雑誌では副標目という）は数字を用いたコード化がなされている。

コード表、分類表で問題となるのは、表にないものが出現した時にどう対処するかということである。定期的に表の見直しを行い、追加、更新を行う必要がある。

### 12.1.3 データシートチェックの要領

データの入力は、データシートに記入されている通りに行えばよい訳であるが、そのデータシートが必ずしも正確であるとは限らない。入力作業を開始する前にデータシートの以下のような項目のチェックを行う必要がある。

- (1) 表記方法に誤りがないか
- (2) 誤字、脱字はないか
- (3) 記入もれはないか

以上の3点については最低限チェックすべきであり、内容面のチェックを行うことが望ましい。チェックリストを作っておき、チェックもれを防ぐようにする。合格したものだけを入力作業に回し、不合格のものは、データシート記入者に差し戻す。

### 12.1.4 データ入力の要領

データの入力は、端末から直接行うか、あるいはパンチカードやワードプロセッサを用いて間接的に行う。前者の場合には、機器の使用法の説明も必要である。また、

- (1) データシート中の各項目をどの順序で入力すればよいのか

## (2) 項目と項目の境界となるデリミターは何か

など、入力作業に関連する細目にわたって、説明と指示を加えることも必要である。

パンチカードを用いて間接的にデータ入力を行う場合は、専門のパンチャーに依頼する方がスピードも早くまちがいの少ないので、コスト的にも有利である。パンチ業務を外注する場合には、テープのフォーマットについての詳細な仕様を作成し、誤りを防ぐ工夫が望まれる。

### 12.1.5 品質管理の方法

データシートの段階で誤りがなかったとしても、入力時のパンチミスは発生する。一定以上の質を維持するためには、データシートと照合してパンチミスの校正をすることが必要であるが、多量のデータの場合はなかなか困難である。そこで、

- (1) 同じデータシートに対して二人が入力作業を行う
- (2) 二人の入力結果をコンピューターで照合する
- (3) 二人の入力結果に違いがある場合のみデータシートとの比較を行い、誤りを修正する

といった方法を採用し、パンチミスのチェックを行う。この方法では二人が同時にパンチミスを行った場合は見逃されるが、そのようなケースはごく稀であり、一定の品質は保たれる。この方法は一例にすぎないが、いずれにせよ何らかの規準を設定し、これをマニュアルにも明記して、長期的な展望に立った品質管理を行う必要がある。

以上、前処理・入力マニュアルに必要な項目を列挙したが、これらは、ただ文章の形で記述するだけでなく、データシートの記入例を示しながら説明を加え、また2～5については、フローチャートなどで全体の流れを示しながら解説すると、よりわかりやすいものになる。

## 12.2 運用マニュアル

運用マニュアルは、データベース・サービス提供機関が、作成されたデータベースを利用者に提供するために必要な処理手順を指示するものである。これは、ホストコンピュータ中にデータベースを構築し、更にそのデータベースを運用し、定期的に更新していく作業の手順を指示するものである。

データベースの提供形態には、以下に示すような幾つかの方式がある。

- (1) データベース・プロデューサが磁気テープ等の形でデータベースをディストリビュータに提供する。ディストリビュータは、多くのプロデューサからデータベ

を購入もしくはリースして、ユーザーにオンラインもしくはバッチで検索サービスを提供する

- (2) データベース・プロデューサが、外部の情報処理業者のホストコンピュータを借りて、ユーザーに検索サービスを提供する
- (3) データベース・プロデューサが自前のホストコンピュータを使って、検索サービスを直接提供する
- (4) データベース・プロデューサがCD-ROM等の形で、データベースと検索システムをユーザーに販売する

利用マニュアルは、ホストコンピュータ中のデータベースを運用する業務の担当者に指示を与えるものであるから、上記の(1)~(3)の場合には必要であるが、これらの方式のどれを採用するかによって、記入すべき内容も異なる。

#### 12.2.1 ディストリビュータに提供する場合

データベース・プロデューサが構築したデータベースを磁気テープ等の形でディストリビュータに提供する場合には、データベース・プロデューサは磁気テープに収録したデータベースの入力書式と項目、内容等に関する詳細な仕様書を作成し、データベースとともにこれを提供する。

磁気テープに収録したデータベースの仕様書は、概ね以下のような事項から構成される。

- (1) 採用されているコード体系
- (2) データベースの論理的構造
- (3) データベースの物理的構造
- (4) 各項目の書式、記述方法、記号法
- (5) コード表・分類表

ディストリビュータ側では、このデータベースの仕様書に基づいて、磁気テープ中のデータベースをホストコンピュータ中のDBMSの中に構築していくために必要な変換ソフトウェアを作成する。

ディストリビュータは、プロデューサから定期的に納入される磁気テープを変換ソフトを用いて変換し、一定の手順でDBMS中にデータベースを構築する。この手順を詳細に指示したものが、運用マニュアルである。

この運用マニュアルは、概ね以下のような項目から構成される。

- (1) 磁気テープのフォーマット(上述のデータベース仕様書の内容と同様)

- (2) 変換ソフトウェアの仕様
- (3) 変換ソフトウェアの使用手順
- (4) DBMSの仕様
- (5) 作業時間と作業手順

### 12.2.2 外部のホストコンピュータを利用する場合

この場合は、運用マニュアルには、ホストコンピュータにデータベースを蓄積する作業手順、および構築されたデータベースの運用手順が示される。

### 12.2.3 自前のホストコンピュータを用いる場合

この場合は、データの入力段階から、ホストコンピュータへの蓄積を経て、データベースを運用する業務までの全プロセスが、運用マニュアルに含まれる。

## 12.3 利用マニュアル

データベースのユーザーが増え、多様になればなるほど、データベース提供者側がユーザーに対して、個々に十分な支援を行なうことは困難となる。従って、利用マニュアルの役割はますます重要になる。

利用マニュアルに求められる条件は、それを見ればユーザーが検索しようとするときに直面すると考えられる疑問全てを、解決できるということである。

### 12.3.1 実際に検索を行う前の段階で必要な項目

#### (1) システムの概要

データベースを利用するにあたり、利用者があらかじめ知っていなければならない一般的な事柄を記述する。細かい項目としては、

- ・利用可能な時間帯
- ・不明な点がある場合の問い合わせ先

が少なくとも含まれていなければならない。また、広く一般に有料でサービスを提供する場合は、

- ・料金体系
- ・利用可能な地域
- ・利用申し込み、解約方法

といった項目が必要である。

(2) 使用する機器について

- 必要な機器は何か（例えば、モデムなど）
- 利用できる機種は何か
- 機器の操作方法（特にタイプミスの訂正方法、送信方法）

といったキーボード上の操作については、使用する機種によって、若干の相違があるかもしれないので注意を要する。キーボード上の操作の説明は、キーボードを図示して行うとよい。

(3) システムへの接続、切断方法

接続方法の説明は、最初から手順通りに逐一丁寧に行なう必要がある。端末機の電源をONにするところから、検索式を入力する準備が完了するまで、実際に端末機上にホストコンピュータから送られてくるメッセージ、端末機から入力するメッセージ等、ひとつひとつに解説を加えなければならない。

切断方法についても同様で、切断（終了）のためのコマンドを入力するところから、端末機の電源をOFFにするまで、正常に切断された場合に現われるメッセージ等を、順を追って解説する必要がある。

### 12.3.2 検索の際必要な項目

(1) 論理式による検索について

文献データベースの場合、論理式を用いて検索する方法が一般的であり、

- 論理積（AND）
- 論理和（OR）
- 論理差（NOT）

について、おのおのがどのような意味を持ち、それらを用いたときにどのような結果が得られるのかを、ベン図等を用いて説明する必要がある。ただし、想定されるユーザーの専門性に応じて、この項目は簡潔にすることもできる。

(2) 検索コマンドの解説と使用方法

- どのようなコマンドがあるか
- 各コマンドの機能は何か
- 実際にはどのように使うか

という三点がコマンドに関して最低必要である。

どのようなコマンドがあるかということは、コマンドの一覧表を作ることによって簡潔に示すことができる。一覧表には、簡単な機能の説明と実例、さらに、コマンドの省略形がある場合には、それも含むとよい。

各コマンドの機能、使用方法は、検索例をあげながら説明すると、ユーザーは理解しやすい。また、ユーザーとしては、「Aというコマンドを使えば何ができるか」ということよりも、あることを実行するためには、どのコマンドを使えばよいか」ということを知りたいはずであるので、表現方法に工夫をしながらなるべくその点をわかりやすくすべきである。

### (3) ファイルの概要

データベースは複数のファイルで構成されているので、ユーザーが各自のニーズを満たすために、どのファイルを選択すればよいかを判断できるようにしておかなければならない。

- ・ファイルの特色についての簡単な解説
- ・主題の収録範囲はどこまでか
- ・情報源はどのようなものか
- ・収録期間／件数／追加頻度

を明示する必要がある。

また、一覧表を作り、どのようなファイルがあるのか、すぐに理解できるようにしておくとうい。

### (4) レコードを構成する項目とこれを用いた検索の方法

1レコードを構成している項目は、ファイルによって異なり、従って検索に利用できる項目も異なってくるので、原則として、ファイルごとに説明すべきであろう。

項目の説明にあたっては、完全な形のレコードのどの部分が、それぞれどの項目に該当するのかを示すとよい。

検索に利用できる項目については、それらの一覧表をつくり、

- ・各項目を示すための記号
- ・それらと共に使うことができるコマンド
- ・検索例

を示し、必要に応じて解説を加える。この部分では、=、/、;、-などの記号を多く使うことが考えられるので、混乱が起こらないよう注意を要する。こ

のほかにも、特殊記号を使用する場合には、これらをまとめた一覧表を作り、各記号の使用法の説明と用例を示しておくといよい。

(5) その他の検索機能

トランケーション機能があったり、近接演算子を使用できるなど、付加的な検索機能がある場合には、各機能の解説が必要である。必要な項目としては、

- ・各機能の役割の説明
- ・用いられる記号法
- ・実際にはどのように使うか

があげられる。このような機能の検索例は、ここだけではなく、コマンドの使用法を説明する際の検索例の中に取り入れるといよい。それによってユーザーの目に深くとまることになり、理解が進むことになる。

(6) サーチエイド

検索を助けるツールとして、分類表、シソーラスが考えられるが、まずこのようなツールの有無が示されなければならない。ある場合には、

- ・分類表・シソーラスとはなにか
- ・これらを使用する意義、利点は何か
- ・実際にはどのように使うか

という項目について解説する必要がある。分類表、シソーラスのいずれの場合も、

・どのようにして、分類番号（記号）あるいはディスクリプタに到達するかという点が最も重要であり、実例を示しながら順を追って解説しなければならない。また、ある分類番号、あるいはディスクリプタを用いて検索して、よい結果が得られなかった場合、再度どのように分類表、シソーラスを使用するかについても解説する必要がある。

(7) 検索例

コマンドの使用法など個々の項目を解説するために多くの検索例を用いているが、すべては断片的なものにすぎない。そこで、全体を通した検索例を示す必要がある。その場合には、できるだけ具体的に、ニーズの分析から最終的にニーズが満たされるまでの探索手順を示すべきである。その際、ユーザーが入力しなければならない部分がはっきりわかるように表現を工夫するといよい。

(8) 索引

マニュアルはその性質上、通読するものではない。何か不明な点がある時に、ユーザーが部分的に読むものであるから、どこに何が書かれているかを示す索引は充実させるべきである。同様の理由で目次もなるべく詳細なものにし、ユーザーにできるだけ多くの手がかりを与えるようにすべきである。

## 12.4 マニュアル作成の方法

マニュアルというものは、片手間で作れるものではない。わかりやすく、使いやすいマニュアルを作るためには、それなりの費用と人材と時間と情報が必要である。また、わかりやすい文章を書くために、正しい日本語のセンスを備えていることも求められる。更に、マニュアルは、常に使う人の立場に立って作成することが必要である。

### 12.4.1 マニュアル作成のツール

マニュアルを作るにあたって確保すべきものは、以下のとおりである。

- (1) 資料（対象に関する技術的な資料と、編集・制作に関する資料）
- (2) 人材（テクニカル・ライター、編集者、アシスタント）
- (3) 予算（開発費の一部に含まれる）
- (4) 日程（余裕があることが望ましい）
- (5) 道具（使いやすいワープロなど）

さらに、これらを管理し、開発担当者との連絡やスケジュール調整を行うディレクターが必要である。

### 12.4.2 考慮すべき事項

マニュアルを作成するにあたって、特に考慮すべき点を以下に示す。

#### (1) 利用者の水準

対象とする利用者がデータベース利用の初心者であるか熟練者であるか、また、特定分野の専門家であるか一般の利用者であるかによって、マニュアル作成の方針、記述内容、使用用語等が異なるので、利用者の水準を考慮することが必要である。

#### (2) テクニカル・ライターの起用

データベースや検索システムのマニュアルは、その出来ばえによって、利用者のデータベースに対する評価を左右するほど重要なものである。この種のマ

マニュアルの作成においては、システムやデータベースに関する十分な理解が必要とされるだけでなく、初心者にも理解できるような分かりやすい文章や表現方法を用いることも必要である。そこで、使いやすいマニュアルを作成するためには、第三者であるテクニカル・ライターという専門家に、マニュアルの作成を依頼することが望ましい。従来は、システム開発担当者がユーザー・マニュアルを作成することが多かったが、システム開発担当者が作成するマニュアルでは、利用者には理解出来ないカタカナ語や、専門家にしか理解できないような表現が用いられることが多く、データベースの価値を十分に表現しきれないためである。

### (3) 造本デザインの工夫

マニュアルの編集・制作の手順を熟知したうえで、判型やレイアウト、書体、文字の大きさ等を考慮して文章を書くことが必要である。さらに、重点を印象づけるために、イラストや色使いにも気を配る必要がある。

### (4) 目次・索引の充実

使いやすいマニュアルを使うためには、単にわかりやすい文章の組み立て方や全体構成や無駄のない説明方法を心がけるだけでは充分ではない。利用者が知りたいことを知りたい時に簡単にひきだせるような工夫も必要である。そのためには、マニュアルの読み方、使い方を最初に説明し、また重要な点は何度もくりかえし説明すること、そして、目次や索引を充実させることが必要である。

### (5) 社会性の具備

マニュアルは、分かりやすいだけでなく、ミスがなく正確であることも必要である。さらに、権利の表示や他者の著作権の保護、不適切な用語や表現の回避、著作物責任対策等の社会性を備えていなければならない。

## 12.4.3 文章表現上のポイント

マニュアルの文章は以下の点を考慮して書くことが望ましい。

- (1) 誰もが知っている平易な用語を用いる。
- (2) 短い文章で簡潔な表現で書く。
- (3) あいまいな表現を避ける。
- (4) 自然な日本語で書く。

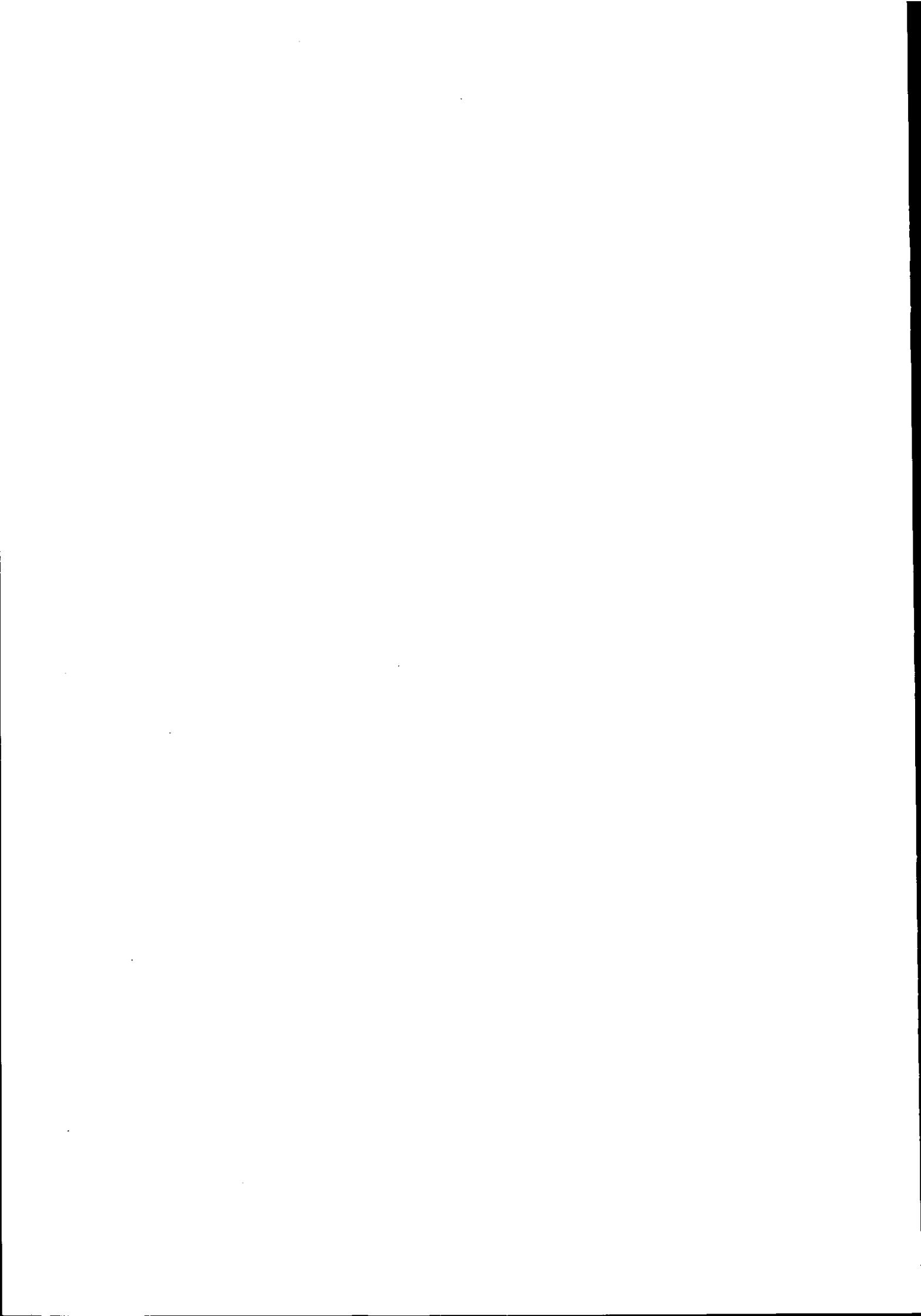
また、表記法や用字・用語の扱い方などについては、統一する必要がある。そのために、

独自の「文章表現集」や「用語集」を作ることも必要である。

自然な日本語で書くためには、以下の点に注意する。

- (1) 能動態で書く。
- (2) 現在時制を用いる。
- (3) なるべくカタカナを使い過ぎないように注意する。
- (4) 翻訳調にならないようにする。

## 第6部 生産



### 13. 運用・開発面での体制の確立

#### 13.1 開発体制

データベースの開発体制とは、データベースの構築の計画書が受理された後、データベースを実際に構築する際に必要な人材、機器、投資、期間の全ての資源と仕組みを総称するものである。

##### 13.1.1 開発スケジュール

データベースの開発に係る様々な業務を項目別に分類し、おのこの項目についての所要期間を設定し、更に各項目の実施順序を決めて、開発プロセスの全工程の所要期間を算出し、スケジュール表を作成する。書誌データベースのスケジュール表の例を(表 13.1)に示す。各項目の所要時間はデータベースの規模や種類により異なるが、作業手順は、概ね以下のようなものとなる。

作業項目	月											
	1	2	3	4	5	6	7	8	9	10	11	12
データ項目の設定	←→											
標準化ツールの開発	←→	←→										
ファイルの設計	←→											
入出力の設計			←→									
データシートの設計				←→								
機器の選定	←→	←→										
入力マニュアルの作成					←→							
データシートの記入						←→	←→	←→	←→	←→	←→	←→
入力作業							←→	←→	←→	←→	←→	←→
運用マニュアルの作成								←→	←→	←→	←→	←→
利用マニュアルの作成									←→	←→	←→	←→

表 13.1 書誌データベースの開発スケジュール表

### 13.1.2 人員計画

開発スケジュール表に基づいて、各作業項目において必要とされる人材のレベルと数を算出する。前出のスケジュールを例にとると、例えば、以下のような人材が算定される。

(1) データ項目の設定			
対象分野の専門家	3人	×	10日
情報専門家	2人	×	10日
(2) 標準化ツールの開発			
対象分野の専門家	3人	×	30日
情報専門家	2人	×	30日
(3) ファイルの設計			
情報専門家	2人	×	15日
(4) 入出力の設計			
対象分野の専門家	1人	×	15日
情報専門家	2人	×	15日
プログラマー	2人	×	30日
(5) データシートの設計			
対象分野の専門家	1人	×	15日
情報専門家	2人	×	15日
(6) 機器の選定			
情報専門家	2人	×	30日
(7) 入力マニュアルの作成			
情報専門家	2人	×	20日
(8) データシートの記入			
対象分野の専門家	5人	×	100日
(9) 入力作業			
パンチャー	2人	×	100日
校正者	2人	×	100日
(10) マニュアル類の作成			
テクニカル・ライター	2人	×	60日
編集者	1人	×	100日
(11) プロジェクト管理			

プロジェクト管理者 1人 × 200日

以上の算定された人材を開発スケジュールに合わせて振り分けると、開発の各段階において、どのような人材が何人必要であるかが明らかとなる。この人員計画に基づいて必要な人材を確保する。

### 13.1.3 使用機器

データベース開発に必要な機器を、開発スケジュールに合わせて導入する。

データベースを磁気テープの形で構築し、データのパンチ入力業務を外注すれば、全く機器を使用せずに、データベースを構築することも可能である。ただし、この場合にも、人員の作業スペース、机、書棚等の設備は必要である。

データベースを構築するために、コンピュータ等の機器を新たに導入する場合には、5.1で定義した情報システムの持つべき機能を満足しうる機器構成を選定する作業が必要となる。この場合には、各種の必要条件を明示した「要求仕様書」を作成し、これをコンピュータ・メーカーに提示する。メーカー側はこの要求仕様書を満たすことのできるハードウェア、ソフトウェア、DBMS等の構成を提案する。複数のメーカーから提案書と見積書を入手し、最も要求を満たしうる構成のものを選定する。

データベースの構築においては、データ量が増えるに従って、磁気ディスク等の二次記憶装置の規模が等比級数的に莫大なものとなるので、データ量の増加に合わせた装置の導入計画が必要である。

いずれの場合にも機器に仕事を合わせるのではなく、仕事に合わせて機器を揃えていくように心掛けなければならない。

### 13.1.4 予算とその積算方法

開発スケジュールと、人員計画、機器導入計画および設備計画に基づいて、開発経費を算定する。

データベースの開発に係る費用は、人件費、機器の購入（リース）料、使用スペースの費用、設備費のほか、パンチ業務等の外注費である。この中でも、費用の大部分を占めるのは人件費である。

#### (1) 人件費

データベースの開発においては、13.1.2の人員計画にも示したように、様々なタイプの専門技術者が必要であり、多数のしかも単価の高い人材が要求され

る。人件費の算定は、一般に以下のような方法で行なう。

- A) 各個人の技術レベルや職位に応じた平均年間所得（賞与・手当を含む）+会社が負担する社会保険料を、年間実務日数で割り、1日当りの単価(イ)を算出する。
- B) 各担当者の一日当りの単価(イ)を、各作業項目毎の必要人数および必要日数とかけ合わせ、各作業項目の必要人件費(ロ)を算出する。
- C) 各作業項目の必要人件費を集計し、開発に必要な全体の人件費を算出する。

## (2) 機器導入費

データベース構築のために機器を導入する場合は、一般にリース、またはレンタルの方式を採用する。

リースとは長期賃貸借のことで、購入資金の調達が困難であったり、買い入れるよりも賃借する方が費用面（節税、設備陳腐化の回避、財務比率悪化の防止）で有利な際に使用され、実際の契約はリース会社と行なう。一般にリースの月額使用料金は、次のように算定される。

3年リース : 買取り金額の約 3.3%

5年リース : 買取り金額の約 2.3%

レンタルは、本来、ユーザーがメーカーから直接賃借するものだが、実際には、レンタル専業会社から賃借する場合が多い。レンタルはリースと異なり、頭金を必要としない。価格は相手先や契約形態によって異なるが、短期の場合は一般にリースよりも高額である。

## (3) 外注費

パンチ入力業務は、専門のパンチャーに依頼した方が安価でしかも正確な結果が得られる。社内にパンチャーがない場合は、専門のパンチ業者に外注する。

パンチ業務は一般に単価制で行なわれ、概ね以下のような単価構成である。

数字 1文字	0.5 円～
アルファベット 1文字	0.6 円～
漢字 1文字	2 円～
仮名 1文字	1 円～

この中には、文字校正の費用も含まれている。

データベースの 1レコード当りの文字数が平均している場合は、使用文字の

組合せから、1レコード当りのパンチ入力費用を算定し、これにデータベース中に収録される全レコード数をかけ合せると、全体の費用が積算できる。

パンチ業務のほかにも、以下のような業務は外注する場合が多い。

- ・データシートの印刷
- ・利用マニュアルの印刷

#### (4) 一般管理費

開発費には、上述の人件費、機器のリース料、パンチ業務等の外注費のほかに、スペース・コスト、通信費、交通費等を含む一般管理費が必要である。また、開発プロジェクトの運営管理を行ない、業務を手順に従って進行させるプロジェクト・マネジメントの人件費や、開発担当の人材の給与支払や事務処理を行なう事務経費も、一般管理費の中に含まれる。

一般管理費の算出の仕方は、会社によって異なるが、一般的に、下記のA～Cの費用全体の10%、あるいは、人件費(A)の30%位が積算される。

開発に必要な経費の内訳（主要項目）

- A) 人件費
- B) 機器のリース料、設備費
- C) 外注費
- D) 一般管理費

### 13.2 運用体制

データベース構築の手順やツールが開発され、定常業務としてデータベースを構築・運用できる段階に入ると、開発体制において必要とされたものとは異なる人材、機器等の組合せが必要となる。

運用段階で必要な人材は、各種マニュアルに記述されている内容を忠実に実施し、定常的にデータベースを提供していく業務の担当者である。

運用体制において必要な人材は、データベースの提供形態によって全く異なる。文献データベースを磁気テープ等で構築し、外部のディストリビュータにこれを提供していく場合には、以下のような人材構成となる。

- ・業務管理者
- ・索引作成者
- ・抄録作成者

- ・編集・校正担当者
- ・パンチャー（外注）

データベース作成機関でホストコンピュータを導入して、利用者に対するサービスも行なう場合には、このほかに、以下のような人材が必要である。

- ・コンピュータ・オペレータ
- ・データベース管理者
- ・プログラマー

運用体制を維持するための経費についても、開発体制の場合と同様な方式で所要経費を算定し、月次、および年次予算計画を作成する。予算費目は、以下のような項目から構成される。

- (1) 人件費
- (2) 機器導入費
- (3) 外注費
- (4) 一般管理費

### 13.3 サービス体制

#### 13.3.1 サービス方式

完成したデータベースを商品として外部の利用者に提供していくためには、十分なサービス体制を整えることが必要である。

提供サービスの方式は、既に前章でも述べたとおり、以下のような三つの方式に分類できる。

- (1) データベース・プロデューサはデータベースを磁気テープ等の形で作成し、これをデータベース・ディストリビュータに貸与する。データベース・ディストリビュータは、プロデューサから入手したデータベースを、自前のホストコンピュータの中に導入し、DBMSの検索ソフトウェアとともに利用者に提供する。
- (2) データベース・プロデューサは、外部の情報処理業者の大型コンピュータのハードウェアとソフトウェアの環境の中に、データベースを構築し、情報処理業者のネットワークやDBMSを通じて、外部の利用者にサービスを提供する。
- (3) データベース・プロデューサが自前でホストコンピュータを導入し、DBMSを構築して、直接、利用者にサービスを提供する。

一般に、リアルタイムの情報を必要とするようなデータベースや、特殊な処理を必要とするものについては、(2)もしくは(3)の方式が採用されるが、書誌データベースのように標準化されたデータベースの場合には、(1)の方式を採用するのが一般的である。これは、プロデューサとディストリビュータの役割を明確に分離することにより、プロデューサの側はデータベースの開発に力を注ぎ、ディストリビュータは販売に専念できるからである。この場合には、プロデューサはディストリビュータの内容情報について責任を有し、ディストリビュータを通じて、料金を徴収する。

### 13.3.2 利用者サポート

直接、利用者に対して有料でデータベース・サービスを提供する場合には、十分な利用者サポート体制を整える必要がある。利用者サポートのために必要なサービスや業務は、以下のようなものである。

#### (1) 利用マニュアルの提供

初めての人がこれを見ればデータベースを利用できるような、詳細でしかも親切で読みやすい利用マニュアルを作成する。

(テクニカル・ライター、編集者等が必要)

#### (2) 講習会の開催

利用者にデータベースの利用法を指導するための、講習会を開催する。

(講習会講師が必要)

#### (3) サービス時間帯の厳守

データベース・サービスの提供時間帯を厳守し、その間、システムダウンや回線故障等に対応できるバックアップ体制を整える。

#### (4) 質疑応答サービス

サービス時間帯中は、利用者からの問い合わせに応じられるようにする。

(窓口業務の担当者が必要)

#### (5) 変更情報の伝達

データベースの内容や、システムの機能に変更が生じたときは、あらかじめこれを利用者に伝達しておく。

(ニューズレター等の発行)

#### (6) マーケティング活動

利用者開発のための、無料講習会の開催、見本市やフェアへの出品・参加

等の活動を行なう。

### 13.3.3 分析手法と分析ツールの提供

数値データベースや物性データベース等のファクト・データベースでは、検索結果を分析・加工するために、利用者用のアプリケーション・ソフトウェアを提供することも必要である。

一般に統計数値のデータベースでは、統計結果を分析・加工するために、様々な統計解析、モデル作成、シミュレーションのためのソフトウェアがデータベースとともにパッケージとして提供されている。利用者は統計データベースを検索して必要なデータをとり出した後、これらのデータを統計解析等のパッケージによって加工し、その結果を統計指標、グラフ、モデル等の形で入手することができる。

一方、物性データベースの場合にも、利用者が解析しようとする物性の測定データをデータベース中のデータと比較することにより、その物質の同定ができるようなアプリケーション・ソフトウェアを備えている。

### 13.3.4 サービスに伴う費用

サービス体制に伴う費用も、準備段階の初期経費と運用段階の定常経費に分けて考えることが必要である。また定常経費は、利用者数の増加に比例して増大するので、利用者開発計画と並行して算定する。

サービス準備段階の初期経費の費目は、概ね以下の項目から構成される。

#### (1) 人件費

サービス業務担当者の人件費で、これには、宣伝資料作成者の人件費、インストラクターおよび営業担当者の教育訓練期間の人件費、サービス部門の人員にデータベースとシステムに関する詳細を伝達するための、開発担当者の人件費、サービス部門の管理者の人件費等が含まれる。

#### (2) 広告宣伝費

データベース・サービスについての宣伝資料の作成と配布、および広告宣伝媒体を利用するためにかかる費用。この中には、紙代、編集・制作・印刷費、デザイン費、広告掲載費、展示会出品費、通信費等が含まれる。

#### (3) 機器導入費

データベースのサービスをサポートするために必要な機器の購入・リース・

レンタルに要する費用。この中には、講習会や展示会で、データベース検索の実演・試用を行うための端末装置、営業担当者に携帯させる端末装置等が含まれる。

#### (4) 一般管理費

データベースのサービスを行うためには、上述の人件費、広告宣伝費、機器導入費のほかに、サービス部門のスペース・コスト（机・椅子・電話等の設備を含む）、通信費、交通費等からなる一般管理費も必要である。

運用段階でのサービス体制の維持にかかわる定常経費も、概ね以上と同様の費目から構成されるが、利用者の増加に応じて、講習会の頻度の増加や講習内容の質の向上、営業活動の活発化、利用者との対応やクレーム処理の必要性に伴い、サービス要員の数と種類を増加させることが求められる。また、データベースの内容の更新や品質の向上に伴い、インストラクターや営業担当者の定期的な再教育を行い、サービスの品質を維持し、向上していかなければならない。そのため、利用者数の増加に伴って、サービス経費の特に人件費が増大する。また、データベースの内容が変化すると、マニュアル類の更新のほかに、ニューズレター等の定期的な配布も必要になるため、このための経費も考慮しておく必要がある。

### 13.3.5 価格の設定

データベースの価格の設定に際しては、慎重にこれを考慮することが必要である。一般的には、開発に要する初期コストと、運用に係る定常経費を10年かかって回収できれば良いという認識がある。

データベースは、出版物を作成したり、統計処理を行ったり、物性分析を行ったりする過程を効率化する目的でコンピュータ導入を計った結果、その副産物として生まれてきたものがほとんどである。

そのため、データベースのプロデューサの大半は、データベースの利用からではなく、出版物の発行や、統計処理結果の提供や、物性分析など、本来の業務から、主たる収入を得ており、機械化によって、この本来の業務が効率化され、従来方式より早く、少ない人員でより良い製品が開発されれば、それによって開発の初期経費を回収することになる。

このような形でデータベースを作成しているプロデューサの場合は、データベースの販売によって得られる収益は、副次的なものであるため、料金も低額となっている。

しかし、オンラインによるデータベースの利用が普及するに従って、データベース・プ

ロデューサの本来の商品であるべき、出版物や統計図表等の販売量が減少する傾向がみられ、この減少分をデータベースの販売によって回収しようとする動きが、主として文献データベースのプロデューサに見られる。

この結果、文献データベースの利用料金が高額になる傾向はみられるものの、それでも料金は1時間当たり6,000～20,000円位になっている。

一方、付加価値の高いデータベースを特定の利用者向けに提供している、専門プロデューサの場合は、データベースの売り上げによって、開発・運用経費を回収しなければならず、そのため、データベースの使用料も高額なものになっている。

データベースの販売料金を決定する場合には、上述のような様々な要因を考慮した上で判断を下すことになるが、プロデューサ側の市場調査や、販売戦略等によっても、その価格は異なってくる。

販売戦略は、以下の二つに分けられる。

(1) 多数の利用者に低価格のサービスを提供する。

(2) 少数の利用者に高価格のサービスを提供する。

プロデューサがディストリビュータを通じて、データベース・サービスを提供する場合には、プロデューサとディストリビュータの間で交わされる契約によって、収益が決定する。

プロデューサが利用者との間で直接、契約を結ぶ場合は、以下のような3通りの方式がある。

A) 完全従量制

利用時間(または利用量)に応じて、料金が徴収され、利用しない限り無料。

(DIALOG, SDC, 等)

B) 年間契約制

年間利用契約を一定金額で結ぶと、その間いくら利用しても料金は変わらない。

(ダウジョーンズ, QUICK, 等)

C) 最低料金制

従量制ではあるが、全く利用しない場合でも、一定の月額料金が請求される。

(NEEDS-IR, ニューズネット等)

一般に、個人利用者を対象として薄利多売をめざすならば、AまたはCの方式を、大口の機関利用者のみを対象とする場合にはBの方式を採用する。

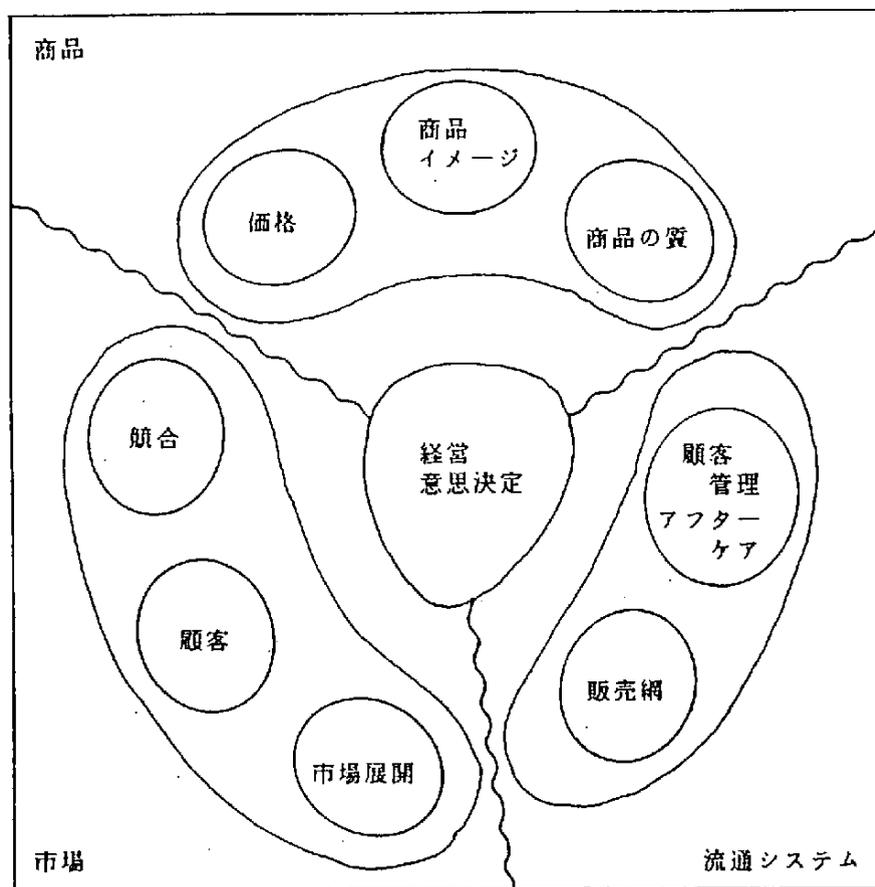
## 14 商品化のためのチェックリスト

データベースの構築が「営業」を目的としてなされる場合、あるいは既存の社内データベースの外販を始めようとする場合、そのデータベースの「商品」としての可能性の検討、および販売戦略の検討が必要である。ここでは、具体的な検討のためのチェック・リストを以下に示す。なお、このようなチェックに至る以前に、前章までに述べた各種の要件が十分に検討され、練りあげられているべきことは言うまでもない。

また、このリストの項目には、若干の配慮や表示の修正で済むレベルのものから、独立した調査・研究を必要とするようなものまで一緒に含まれている。後者のような項目については、その「調査法」まで紹介することができれば理想的であろうが、紙数の制約もあり、項目（課題）の提示にとどめる。

商品としてのデータベースをとりまく環境のイメージ・チャートを図14.1に示す。

図 14.1 イメージ・チャート



あるデータベースを「商品」として世に送り出し、さらにそれを事業として成立させてゆくためには、図に示すような

- 1) 商品自体の持つ可能性
- 2) 市場環境
- 3) 流通システム構築方策

という3つの角度からの検討が必要である。以下にこれらのチェックポイントについて述べる。

#### 14.1 商品

商品自体について検討する場合、商品化段階において戦略的に操作することの可能な「価格」などと、その段階では最早変えることのできない商品の「質」などとを区別して考えることが必要である。前者で求める情報は「どのように」という戦略であるのに対し、後者の場合は「評価」に基づく可能性になるのである。商品の「イメージ」の場合は、この両者の複合したものとなる。

##### 14.1.1 商品の質

- (1) 性格づけ
  - A) 構築目的は明らかか
  - B) 想定利用者は明らかか
  - C) ニーズは充分調査済みか
- (2) 品質水準
  - A) 情報源は確実か
  - B) 対象情報のカバー率は
  - C) 情報の新しさは
  - D) 「品質検査」の方法とシステムは

##### 14.1.2 商品イメージ

- (1) イメージ
  - A) 商品のイメージは明解か
  - B) 内容、ニーズに整合したイメージが確立されている(できる)か
  - C) “まぎらわしさ”はないか
- (2) 権威性

- A) 商品名は「ステイタス」をもっているか
- B) 企業名は「ステイタス」をもっているか
- C) その他の何らかのオーソリティーをもっているか  
(例：当該分野における著名人の参画など)

### 14.1.3 価格

#### (1) 価格設定

- A) 価格は類似商品と競合できるか
- B) 価格は想定利用者の負担能力と整合しているか

### 14.2 市場

例え商品自体がどんなに優秀なものであったとしても、市場のニーズに適合していなければ事業としては成立しない。したがって、事業化にあたっては、市場をどこにどのように設定するのか、対象とする市場の現状をいかに正確に把握しているか、といったことがきわめて重要になる。

#### 14.2.1 市場の展開

##### (1) 空間的展開

- A) どのエリアを対象とするか  
(例：世界・日本・関東・東京)
- B) サービスの展開能力は充分か
- C) 市場に関する情報の収集システムは

##### (2) 時間的展開

- A) 当該情報の「ニーズ・重要度」は高まっているか
- B) 当該情報の「量」は増加傾向にあるか
- C) 将来動向についての分析・予測はできているか

#### 14.2.2 顧客

##### (1) 顧客の確保

- A) 既に確保している顧客はあるか
- B) 既顧客は、新たなデータベース・サービスの顧客となりうるか

C) 通常の広告・宣伝・販売工作以外に、顧客を確保する手段・システムはあるか

(2) セグメンテーション

A) 顧客の種類をどのようにしぼっているか

B) 顧客の階層をどのようにしぼっているか

14.2.3 競合

(1) 競合商品

A) 競合商品はあるか

B) 相手先のシェアと市場余力は

C) 相手先商品の長所と短所は

(2) 競合企業

A) 相手先企業の動向、類似商品進出の可能性は

(3) 競合戦略

A) 商品の「質」の差別化の可能性は

B) イメージ・権威性による差別化の可能性は

C) 価格、販売力による差別化の可能性は

14.3 流通システム

データベースの事業化にあたっては、流通システムの構築も極めて重要な課題である。ポイントとしては、オンライン・リアルタイムサービスを実施するかどうか、外部ディストリビュータを利用するか自社で直販するか、といったことがまず挙げられる。これらは早い時期に選択決定しておかなければならない。また、「情報の流通」といっても、実際には「サービスの販売」としての性格が極めて強いのであり、その意味で顧客に対するきめの細かいサービスが重要である。

14.3.1 販売網の形成

(1) 外部ディストリビュータを利用する場合

A) 予定業者の組織は充分か

B) 予定業者の知名度は充分か

C) 予定業者の専門分野は商品と適合しているか

D) 予定業者のサービスシステムは商品と適合しているか

(2) 直接サービス

- A) ホストコンピュータは自社所有か
  - B) オンラインサービスのためのシステム開発はどうするか
  - C) 外部コンピュータの場合にはどこを利用するか(情報処理業者, VAN業者, NTTなど)
  - D) 回線網をどのように設けるか
  - E) サービス拠点をどのように設けるか
  - F) 課金および集金のシステムをどのように組織するか
- (3) 直接サービス(オフライン)の場合
- A) サービス拠点をどのように組織するか(支店, 営業所, 代理店など)
  - B) 販売単位をどのように設定するか(データ, レコード, ファイルなど)
  - C) 契約および課金の方式をどのように設定するか

#### 14.3.2 顧客管理およびアフターケア

##### (1) 顧客管理

- A) 顧客のデータを収集する方法は
- B) 顧客のデータを管理する方法は
- C) 顧客の要望・意見をモニターするシステムはあるか
- D) 顧客および顧客ニーズに関する情報の分析システムはあるか

##### (2) アフターケア

- A) マニュアル等の更新システムはできているか
- B) 顧客に対する講習・研修サービスは可能か
- C) 顧客とのコミュニケーションのためのメディアはあるか(ニューズレターなど)
- D) 顧客を組織化するしかけをどうするか(ユーザー会, 研究会など)

参考文献一覧

「情報検索システムの設計・運用に関する調査研究報告書」 行政管理庁行政管理局,  
1983, 231p.

「情報システム計画作成の手引き」 日本アイ・ビー・エム株式会社 1983, 159p.

高橋昭男 「わかりやすいマニュアルの作成法」 東京, 日経マグローヒル社, 1985,  
204p.

津田良成編 「情報システム論」 講座情報と図書館 東京, 雄山閣出版, 1983, 263p.

津田良成編 「図書館・情報学概論」 東京, 勁草書房, 1983, 239p.

原田 勝他 「データベース構築の理論と実際」 東京, コロナ社, 1985, 250p.

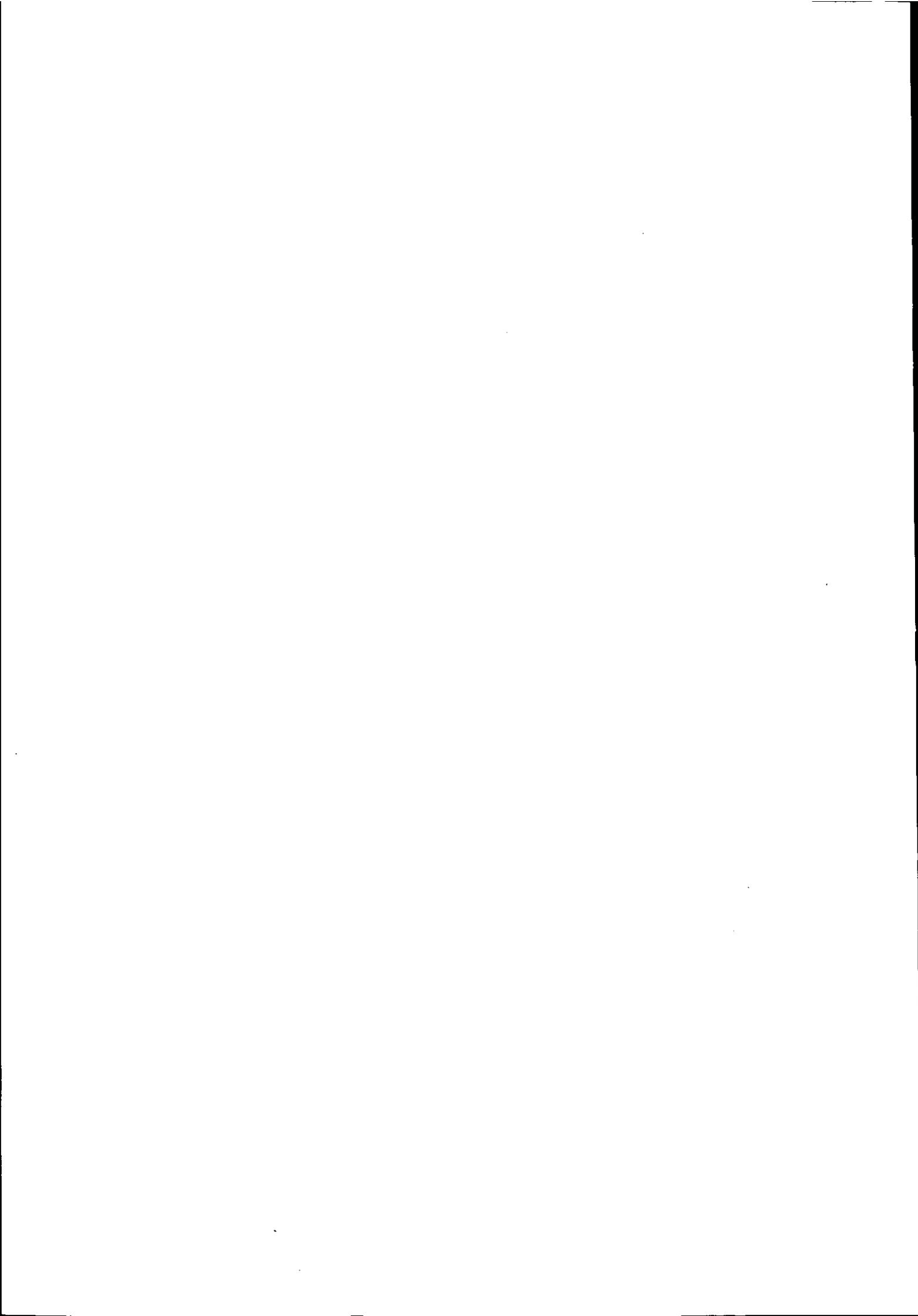
Atherton Pauline "Handbook for information systems and services", Paris,  
Unesco, 1977, 259p.

Chan, Louis Mai "Cataloging and classification : an introduction." New  
York, McGraw-Hill, 1981, 397p.

"ISO/DIS 5963 Documentation-Methods for examining documents,  
determining their subjects, and selecting indexing terms", International  
Organization for Standardization, 1984, 5p

"ISO/DIS 2788 Guidelines for the establishments and development of  
monolingual thesauri" International Organization for Standardization,  
1984, 65p.

## 資料編 データベース構築例



「データベース構築標準マニュアル」の作成にあたり、既にデータベースの開発・サービス提供を行なっている9機関の訪問調査を実施した。

調査期間は昭和61年1月10日より1月30日までとし、調査方法は、ヒヤリング方式を採用した。質問内容は、本文の章立てに沿って以下のような質問を設定した。

- (1) 設立にあたっての事前調査・モデルケース
- (2) データベースサービスとネットワーク
- (3) 情報源と情報収集方法
- (4) データベース機能の概要
- (5) データの構造化と情報分析
- (6) DBMSの選定・開発
- (7) ファイルの設計
- (8) 入出力の設計
- (9) 標準化ツールの開発と利用
- (10) 情報分析
- (11) 各種マニュアルの作成
- (12) 運用・開発面での体制の確立
- (13) 商品化のチェックリスト

なお、調査対象機関によって多少の項目の調整を行なった。

調査対象9機関のうち、一次情報データベースを開発しているものは6機関、二次情報データベースを開発しているものは3機関である。

一次情報データベースには、(株)旭リサーチセンターの物性データベースを初め、(株)市況情報センター、(株)帝国データバンク、(株)日本経済新聞社の数値データベース、(株)ゼンリンの地図データベース、(財)住宅産業研修財団の画像データベースがあり、それぞれに、ヒヤリング調査を行なった。

二次情報データベースでは、(株)図書館流通センターの書誌データベースを初め、(財)日本情報処理開発協会のクリアリングデータベース、(財)日本造船振興財団の文献データベースがある。

さらに、海外データベースの一例として、アメリカのNLM(National Library of Medicine)が開発したMEDLINEをとりあげた。

資料編の配列は、機関名のアイウエオ順になっている。

また、各機関が開発したデータベースの内容を表1に示す。

表1 収録データベース構築事例一覧

訪問先機関名	データベース名	内 容
A. ㈱旭リサーチセンター	ID-IR	◦ 化学物質の物性情報を入手する際に利用するもので、データの項目は、プリズムNo.、グレイディングNo.、構成元素、沸点等から成る。一次情報データベース。
B. ㈱市況情報センター	QUICK	◦ 証券、債券、株式等に関する情報を入手する際に利用するもので、データ項目は、株式・債券価格、各種指標、取引所の個別銘柄等から成る。一次情報データベース。
C. 財住宅産業研修財団	HIC	◦ 建築部材に関する情報を入手する際に利用するもので、データ項目は、商品名、会社名、用途、品番、基本材質等から成る。一次情報データベース。
D. ㈱ゼンリン	地図データベース	◦ 何処に、誰が住んでいるか、電線の配置はどうなっているか等の情報を入手する際に利用するもので、データ項目は、世帯主、住所、番地等から成る。一次情報データベース。
E. ㈱帝国データバンク	COSMOS	◦ 企業情報や企業財務、さらに個人情報等を入手する際に利用するもので、データ項目は、企業名、個人名等から成る。一次情報データベース。
F. ㈱図書館流通センター	TRC MARC	◦ 新刊書や既存図書等の目録カードを作成する際に利用するもので、データ項目は、書名、著者名、出版者名等から成る。二次情報データベース。
G. ㈱日本経済新聞社	NEEDS-TS	◦ 経済情報を入手し、分析・加工する際に利用するもので、データ項目は、民間最終消費支出、政府最終消費支出等から成る。一次情報データベース。
H. 財日本情報処理開発協会	データベース クリアリング システム	◦ データベースサービスに関する情報を入手する際に利用するもので、データ項目は、データベース名、分野、内容等から成る。二次情報データベース。
I. 財日本造船振興財団	MATRAX	◦ 船舶・海洋工学関係の技術文献に関する情報を入手する際に利用するもので、データ項目は、雑誌論文名、著者名、抄録等から成る。二次情報データベース。
J. 国立医学図書館*	MEDLINE	◦ 医学・生命科学に関する情報を入手する際に利用するもので、データ項目は、書誌事項、抄録等から成る。MeSH(統制語彙)あり。二次情報データベース。

\* 文献調査のみを実施した。

## A. (株)旭リサーチセンター (ID-IR)

### 0. データベースの概要

- ID-IRとは、化学物質の構造を決定する際に行なわれる赤外線分析で得られたスペクトルをもとに、パターン・マッチングを行ない、その物質が何で、どのような構造を持っているかを明らかにする解析システムである。

### 1. 会社概要

- 旭リサーチセンター (Asahi Research Center, 略称ARC) は、昭和49年に、旭化成グループの総合調査会社として、幅広い分野にわたる研究テーマについて、クライアントへ情報を提供するために、旭化成工業(株)の100%出資会社として発足した。

### 2. データベース構築にあたっての事前調査とモデルケース

- ID-IRの前に、社内サービスを目標としたより小さなデータベースであるX線分析システム(データ件数2,600件)を作成していた。

### 3. データベースサービスとネットワーク

#### ① システムサービス

- 現在提供されているデータベースは、ID-IRのみである。
- ID-IRの他に、NMRスペクトル解析(ID-NMR)、質量スペクトル解析、および、X線スペクトル解析のサービスを計画中である。

#### ② ネットワーク

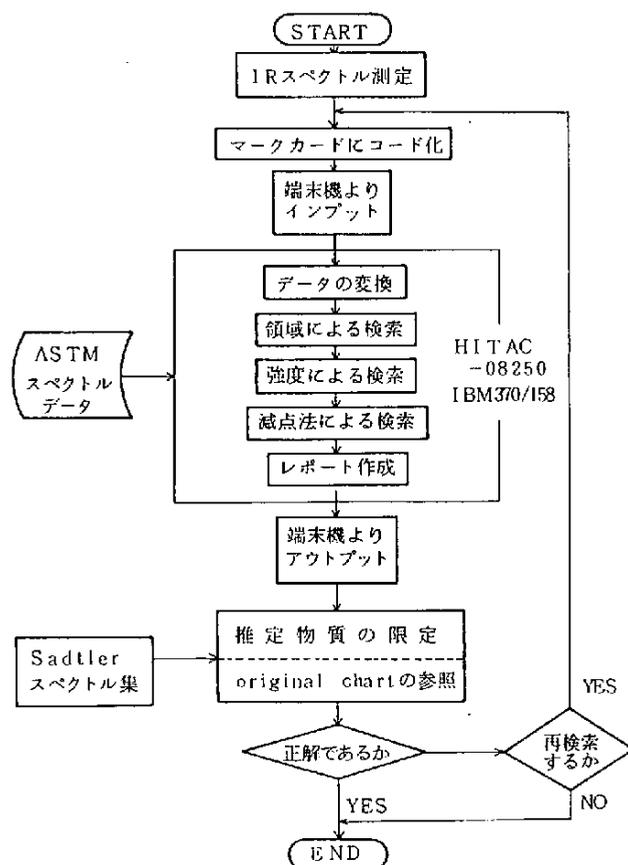
- オンライン・サービスは、昭和54年下期から、電通国際情報サービスのMARK-IIIを通じて行なわれている。
- データ源の契約上の制約から、サービスは日本国内に限られている。
- 端末には、公衆網用端末を使用。
- データベースを購入して利用する際はミニコン以上の機種が必要。
- 旭化成内では、いくつかの研究所にデータベースを置き、ローカルなオンラインサービスを行なっている。富士の研究所の場合、ACOSの中型機、本社の場合は、FACOMの大型機を利用している。端末は、各メインフレームの端末であるが、最近はパソコン端末も増えてきている。

#### 4. 情報源と情報収集方法

- ・情報源は既存のデータベース（約150万円で購入）および、印刷体のチャート集。
- ・ASTM（American Society for Testing and Materials）の赤外線スペクトルデータベース（データ数、約145,000件）があるが、これは数社のデータベースの複合体であり、同一物質の重複、基本的な誤りをもつデータも存在している。そこで、最も体系的かつ広範に収集された物質をもとにした、信頼性の高いサトラー社のデータを主として、選択、付加を行ない、約43,000件のデータを含むデータベースにした。
- ・すべてのデータの信頼性が高い訳ではなかった。従って、チェックをして誤りのあるものについては修正をした。このチェック・修正作業には大変手間がかかった。
- ・更新の頻度は少ない。新しい物質が出てきても、すぐにMTにつけ加えられる訳ではないからである。

#### 5. データベース機能の概要（レコード内容、検索機能等）

- ・システムフローは以下の通り。



（出典：長田洋ほか、赤外線スペクトルの自動解析システム、システム工学会誌、

VOL. 3, No. 1, P. 67 (1979)）

- レコードの内容は、プリズムNo (古いコード体系の物質番号)、グレイティングNo (新しいコード体系の物質番号)、常温での状態、スコア (スペクトルのピークのパターンマッチングでパターンが完全に一致した時を100点とした時の一致度を示す得点)、構成元素、沸点あるいは融点、NMR. No.、部分構造 (主要原子の原子数、基本的な官能基) である。
- 情報源としたサトラー社のデータベースにはなく、新たに付け加えられたのは、グレイティングNo.、NMR. No. である。
- 出力例を次頁に示す。
- スペクトル・データのみならず、官能基の存在の有無や、原子数からの検索もできる。
- 出力では、各物質は番号で示されており、その番号をもとにコードブックを用いて物質名を探すことになる。
- このデータベースに、特殊な自家データを加え、専用ファイルとして保管することができる。これにより、検索の精度を高めるだけでなく、分析実験室の情報管理にもつながる。
- 検索時間は、1件当たり平均 cpu time が 3.8 秒 (IBM 370/158)。

## 6. データの構造化と入力情報の分析

- 赤外線吸収スペクトルのデータは、本来アナログ情報であるが、入力の際、デジタル化している (コード化)。
- コード化は、スペクトル全体の特徴を把握するように行なわれる。
- コード化には、専門知識が必要であるため、当初は研究員クラスが行なっていた。
- 入力フォーマット、コード化のルールが決まってからは、ある程度の専門教育を受けているアルバイトなどが入力した。その場合は、研究員が後からチェックした。
- データベースの性質上、数値そのものの意味がきわめて重要であり、データのチェックを十分行なった。構成元素などについてもクロスチェックを行なった。

## 7. DBMS の選定・開発

- 専用のDBMSを独自に開発。

## 8. ファイルの設計

- ソースファイルと検索ファイルを持っているが、MARK-III を通じて利用できるのは検

SAMPLE I NAME SHPL-1

SEQ	P. NO	G. NO	ST	SCORE	ELEMENT 1234567R	POINT	NUM NO		TOWS				
							(1)	(2)	C	N	O	S	
1	16250	213R0	S	95	* *	M 240	4939		5	3	0	0	
2	1251	15314	L	90	*	H 164	7557		7	1	0	0	
3	28550	3R820	S	79	**	M 172	10953		14	2	3	0	
4	12976	47303	S	75	**	M 214			4	1	4	0	
5	27888	3614	S	74	** *	M 164			16	1	2	0	
6	10087	45558	S	74	** *	M 211	10697		19	2	1	0	
7	37544	16540	S	74	***	M 126			22	4	4	1	
8	28520	422R	S	70	** **	M 91	365		8	1	3	0	
9	15631	1R666	S	70	***	X 0			12	2	4	1	
10	12968	47295	S	70	** *	M 245			6	1	4	0	
11	23012	3R27R	S	70	*	M 197			9	2	0	0	
12	35703	1369R	S	70	* *	M 230	7395		6	1	0	0	
13	195	4R	S	70	*	M 109	10206		16	1	0	0	
14	28596	4302	S	70	***	M 210			12	3	3	1	
15	23550	767	L	69	* *	X 0	13670		8	1	0	0	
16	16811	25509	S	69	**	M 312			7	4	0	1	
17	16063	40775	S	69	* *	X 0			16	3	0	0	
18	257	89	L	69	*	B 230	525		11	0	1	0	
19	9049	45092	S	69	**	M 202			10	2	2	0	
20	22882	10589	L	69	**	X 0	2540		9	1	1	0	

FUNCTIONS

SEQ	IR (M)	FUNCTIONS			
		C-C 123456	C-O 123456789	C-N 1234567	C-S 12345678
1	213R0	*		*	*
2	15314			*	*
3	38820	**		*	*
4	47303		**	*	*
5	3614	*	**	*	*
6	45558	*	*	*	*
7	16540	**	**	*	*
8	422R		*	*	*
9	1R666	*		*	*
10	47295		**	*	*
11	3R27R		*	*	*
12	1369R	*		*	*
13	4R	*		*	*
14	4302	***		*	*
15	767	*		*	*
16	25509	*		*	*
17	40775	**		***	*
18	89	**	*	*	*
19	45092	**	*	*	*
20	10589	*	*	*	*

SUBST. C-H

SEQ	IR (M)	SUBST.	
		YX01234567R9	X01234567R0
1	213R0		
2	15314	*	
3	38820		
4	47303		
5	3614	*	*
6	45558	*	*
7	16540		*
8	422R	*	
9	1R666	*	
10	47295		
11	3R27R	*	
12	1369R	*	*
13	4R		*
14	4302	*	*
15	767	* *	*
16	25509	* * * *	*
17	40775	* * * *	*
18	89	* * * *	*
19	45092	* * * *	*
20	10589	* * * *	*

C-C	C-O	C-N	C-S	C-N-O
1 ISOBUTYL	1 -C(O)OH	1 -CN	1 -C(S)S-	1 -NCO
2 SEC-BUTYL	2 -C(O)O-	2 -NH2	2 -C(S)SH	2 -OCH
3 TER-BUTYL	3 -C(O)OH	3 >NH	3 >CS	3 -ONO2
4 PHENYL	4 >C(O)	4 >NH-	4 -SC(S)S-	4 -NO2
5 >C=C	5 -C(O)O-	5 >C=N-	5 -SH	5 -NH
6 -C=C-	6 -OH	6 >NH	6 -S-	6 HETEROCYCLIC
	7 -O-	7 HETEROCYCLIC	7 -SS-	
	8 -OR-		8 HETEROCYCLIC	
	9 HETEROCYCLIC			

C-O	SUBST.	C-H	ELEMENT
1 -S(O)O-	Y hexa	4 CONJUGATED	1 O
2 -S(O)O-	X PENTA	0 -CH=CH2	2 N
3 >S=O	0 MONO	1 >C=CH2	3 S
4 >S=O	1 1,2	2 =CHCH3	4 F
5 -OSO-	2 1,3	3 -CH2CH=CH2	5 Cl
6 HETEROCYCLIC	3 1,4	4 -CH=CHCH3	6 Br, I
	4 1,2,3	5 -C(CH3)=CH2	7 R, RI
	5 1,2,4	6 =C(CH3)2	8 ETAL
	6 1,3,5	7 -C=CH	
	7 1,2,3,4	8 -C=CCH3	
	8 1,2,4,5	9 -CH2=CH	
	9 1,2,3,5		

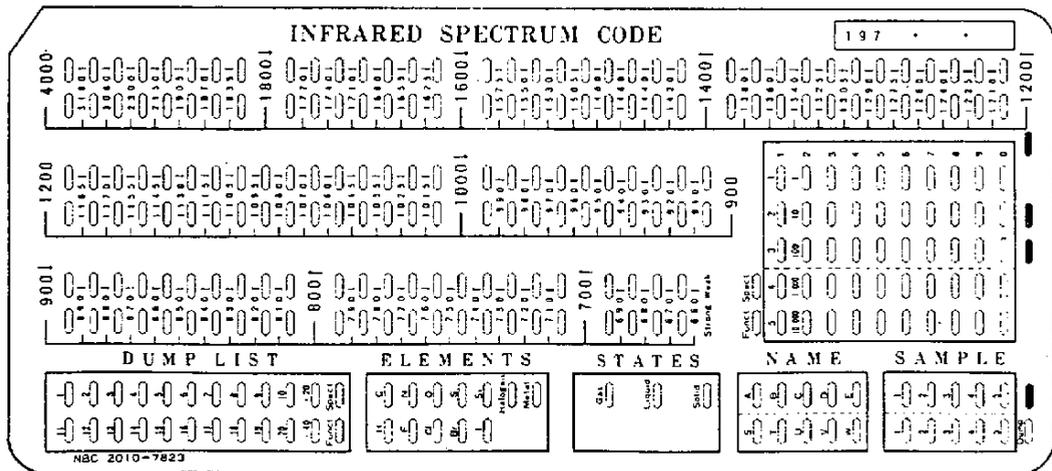
(出典: ID-IR 赤外線吸収マススペクトル解析パンフレット)

索ファイルのみ。

- 検索スピードを上げるために、検索ファイルはデータの長さを圧縮。分析者の名前等は除いてある。
- 社内データベースとして持つ時のみソースファイルを持つ。ソースファイルは、実験室の管理などに使えればよいと考える。

### 9. 入出力の設計

- 入力的设计をする際、どのようなファイルを持つのか、どのような出力にするのか、ということをおあらかじめ決定しておき、それをチェックする必要がある。
- 少ないデータを用いてパイロットスタディーを行ない、修正を加えていった。都合3回のパイロットスタディーを行なったがその時のデータ数は、1回目100程度、2回目300程度、3回目1000程度である。
- 入力は、マークカードを用いて簡易化している。端末からの入力も可能である。
- 入力用マークカードを以下に示す。



( 出典 : 長田洋, 赤外線スペクトルの自動解析システム, システム工学会誌,  
VOL. 3, No. 1, P. 69 (1979) )

## 10. 標準化ツールの開発と利用

- 入力する際の標準化として、アナログ情報で示される赤外線吸収スペクトルのどこをコード化するか、という問題がある。
- 古い分析法を行なっている場合、新しいものとは目盛のふり方が異なるので、古いものにも対応できるようにしている。
- 入力用マークカードを開発し、利用している。

## 11. 情報分析

- 利用者がデータを入れて、検索すること自体が情報分析である。検索して得られた結果を利用者がさらに処理、分析しやすいように、項目が設定されているが、処理、分析を行なうのは利用者であり、システム側が特別の手段を提供している訳ではない。

## 12. 各種マニュアルの作成

- 利用者向けに、
  - 1) 解析、検索のための入力方法
  - 2) 結果の見方、利用の仕方
  - 3) 解析、検索のプロセスの説明のマニュアルを作成している。
- データベース作成用に、
  - 1) データWP追加するための入力方法
  - 2) データをチェックする方法
  - 3) データベースに登録したあとのチェックのマニュアルがある。
- システムマニュアルは、プログラムレベルのものである。

## 13. 運用・開発面での体制の確立

- ファクト・データベースの場合、サービス提供者が内容をよく知っていることが必要である。
- しかし、主題専門家だけでデータベースを作れる訳ではない。システム専門家との共同作業が必要である。データベースシステムの構築をするのはシステム専門家である。また、主題専門家からは、自分が使うためにはどうするか、という意見はでても、サービ

スする側としてどうするかという意見はでてこないだろうし、サービスする時必要な技術を持っていないだろう。

- 主題専門家とシステム専門家との協力は、いろいろな面で可能である。例えば ID-IR の場合、専門家がデータをチェックする際のソフトウェアをシステム専門家が開発している。
- システム専門家の中に、主題及びデータベース、サービスに対する関心の強い人がいなければならない。

#### 14. 商品化のチェックリスト

- 商品化するのは相当むずかしいと考えなければならない。
- 開発と同程度の時間が必要である。( ID-IR の場合、企画段階も含め約 1 年かかっている )
- 社内用に作っていたシステムには、特殊性が伴うので、商品化する際には市場調査を行ない、利用者(この場合、チャート集を使っている製薬会社や化学会社などの研究者)の意見を聞き、システムにとり入れる。
- 価格をどう設定するかが問題である。
- ディストリビュータの評価としては、
  - 1) ハードウェア
  - 2) オペレーティング・システム
  - 3) ツール
  - 4) ネットワーク
  - 5) サービス・運用, PR 体制などを評価し、どのディストリビュータがよいかを選ぶ必要がある。
- プロデューサ自身がディストリビュータを兼ねる場合も上記は重要なチェックポイントになる。

#### 15. アドバイス, 意見

- データベースを作るには、体制づくりが大切である。
- データベースを作るのも投資であり、費用対効果ということを考えなければならない。
- 多量のデータが必要なので、長期的プロジェクトを作る必要がある。
- ソフトウェアをアセンブラで書いておくと、後から機種変更があった時に困ることもあ

る。

#### 16. 国・政府に対する要望

- データベースの作成を効率的なものとするために、国が持っているデータを公開し、情報収集の面での負担を軽くしてほしい。
- データベースに対しても投資減税を認めてほしい。
- データベース構築に対して、継続的にサポートしてほしい。
- データベースの国際流通のためのサポートをしてほしい。1社が1つか2つのデータベースしか持っていない現状では、販売体制、メンテナンス、ニューズレターの発行、バージョンアップ、などを考えると国際的なサービスは、難しい。故に、JETRO等を通じて海外に宣伝などをしてほしい。
- 発展途上国への援助の一部として、現地でシステムを作るためノウハウを提供するといった型の援助を考えてほしい。
- データベース産業がビジネスとして定着するように計ってほしい。民間がやるのは困難なことを公的機関が国の補助をうけてやるのはよいが、それらはいずれ民間に委譲してほしい。現状では、どこまでを官が負担し、どこを民が負担するのかという役割分担が明確ではない。

#### 収集資料

- 1) 長田洋ほか、赤外線スペクトルの自動解析システム、システム工学会誌、VOL. 3, No. 1, P. 63-73 (1979)
- 2) H. Abe et al. A Carbon-13 and Proton NMR Special Data Retrieval System : Spires, Data for Science & Technology, P. 287 ~ 290 (1980)
- 3) ID-IR 赤外線スペクトル解析 (パンフレット)
- 4) 株式会社旭リサーチセンター (パンフレット)

## B. (株)市況情報センター (QUICK)

### 0. データベースの概要

- QUICKの愛称で経済情報を中心に、オンラインでのサービスを目的としており、リアルタイム証券情報サービス「QUICK・ビデオI」をはじめ、株式のほか金融、為替を総合的にサービスする「QUICK-10」や、リアルタイムで自動更新表示をする証券、金融情報のボードサービス、債券、金融のリアルタイム情報サービス「ビデオBM」および店頭株式情報システム等をサービスしている。

### 1. 会社概要

- 市況情報センター (Quotation Information Center K. K. 略称QUICK) は、昭和46年に、刻々変わる経済市況情報をオンラインで提供することを目的として、日本経済新聞社を中心に、日立製作所、証券・銀行35社の共同出資により設立された。
- 事業所は東京本社をはじめ、大阪支社、名古屋支社、東京コンピュータセンター、大阪コンピュータセンター、海外でロンドン事業所がある。

### 2. データベース構築にあたっての事前調査・モデルケース

- 設立にあたってまず、日本経済新聞社と野村証券との共同プロジェクト構想があったが広く証券、銀行から出資を仰ぎ中立的な機関としての性格を持たせた。
- 米国 Bunker Ramo, 同 Standard & Poor's Corporation, Quotron その他先進米国のサービスシステムの調査研究を行なった。
- 東京証券取引所の取引き機械化の研究を進めていた日立製作所の技術協力を得た。

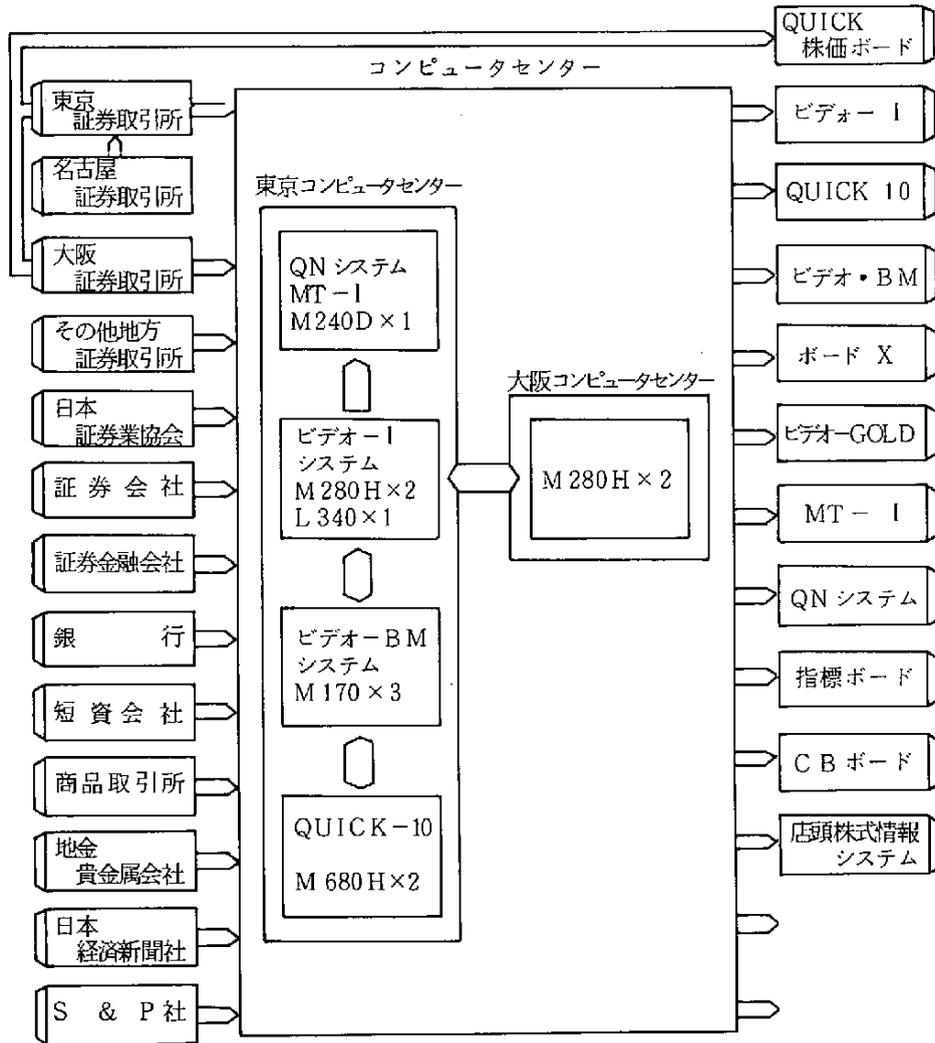
### 3. データベースサービスとネットワーク

#### ① システムサービスの沿革

開始年月日	システム名 (内容)
昭和49年 4月	QUICKビデオーI (株式)
10月	QUICKボードN (株式)
50年 3月	QUICKボードH (株式)
51年 3月	QNシステム (新聞社向け情報システム)
52年 4月	QUICKビデオーX300 (外国為替)
8月	QUICKボードM (株式)
53年 2月	QUICKホームプリンター (株式)
4月	QUICKビデオーII (株式)
54年 6月	QUICK MT-I (株式)
10月	NRS (ニュース検索)
55年 7月	QUICKビデオーBM (債券・金融)
56年 4月	QUICKボードM-II, N-III (株式)
57年 3月	QUICKビデオーGOLD (金)
58年 1月	QUICKボードX (外国為替)
59年 4月	QUICK指標ボード (株式, 外国為替, 債券, 金融)
	QUICK CBボード (株式, 転換社債)
7月	店頭株式情報システム (株式)

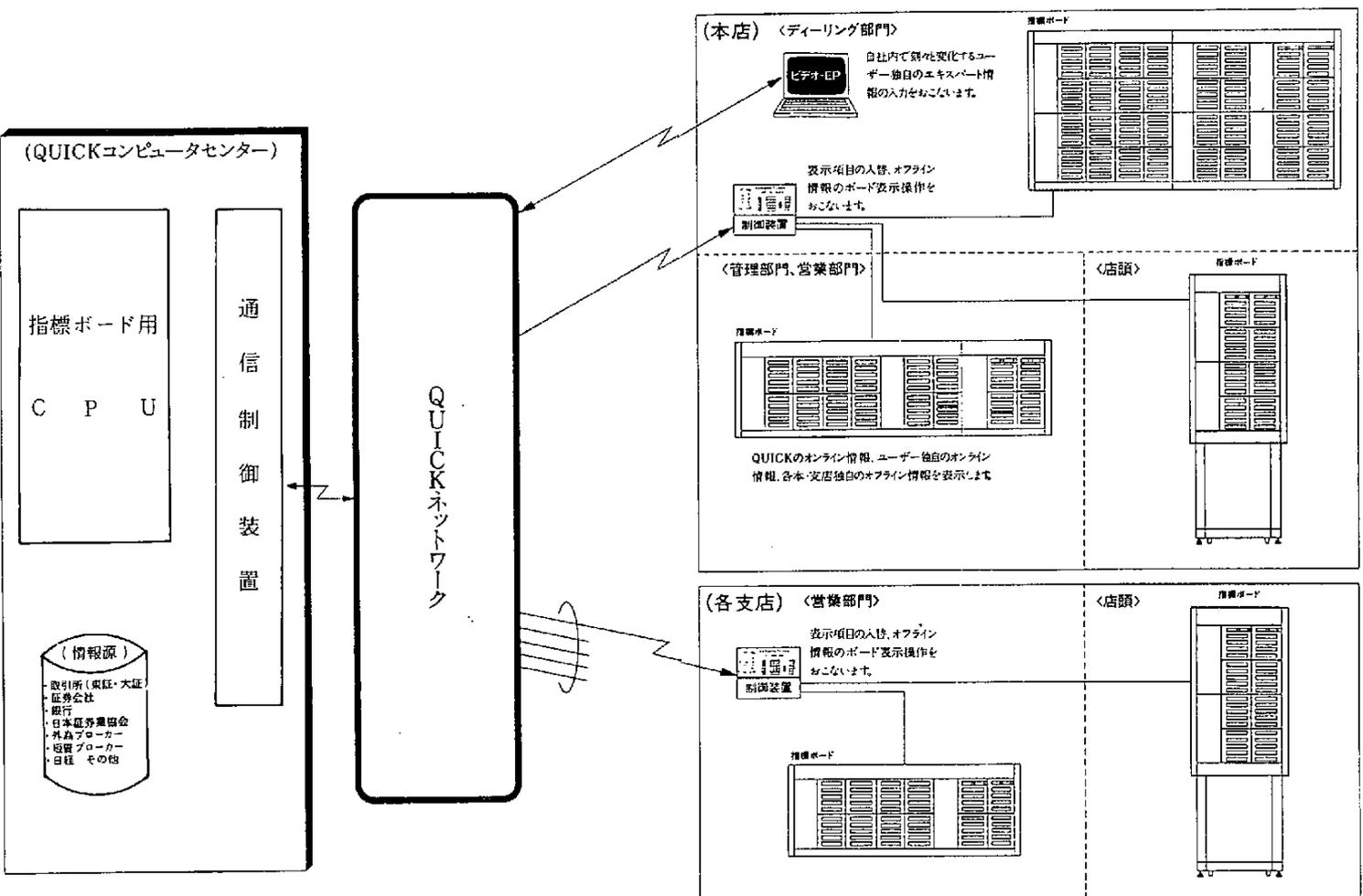
② QUICKの情報システム

- ・下にQUICKの情報システムの概略図を示すが、この他に、日電のMS140を3台、機械翻訳システム開発用にM240H、プログラム開発用にM240D、オンラインデータ伝送用にDATA100X3その他を設置している。



( 出典：市況情報センターパンフレット )

・ QUICK指標ボード（リアルタイム指標一覧表示システム）は次のとおり。

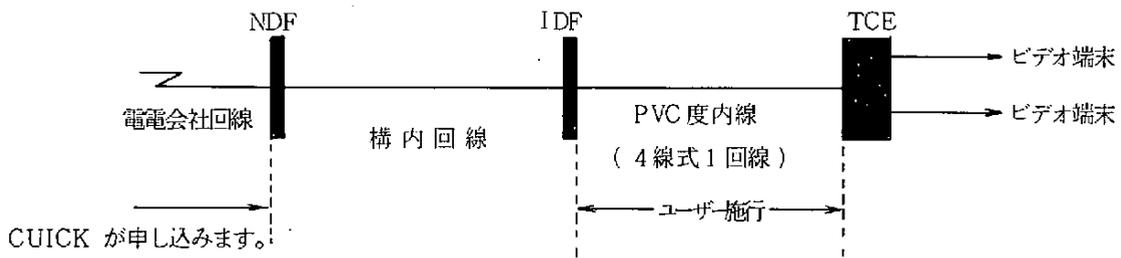


（出典：リアルタイム指標一覧表示システム QUICK指標ボードパンフレット）

- ・ビデオ I における通信回線等，設備工事参考仕様を次に示す。

1) 通信回線

TCE (コントローラ) に接続する回線数は TCE 1 台あたり，通信回線 (通常の電話回線) が 2 対必要で，この回線の申し込みは QUICK が電電会社に対して行なう。QUICK 「ビデオ I」は，D-規格 (0.3 ~ 3.4 KHz 帯域使用) 特定通信回線を使用する。



2) TCE と VD のコンセント諸元 ----- ユーザー 施行 -----

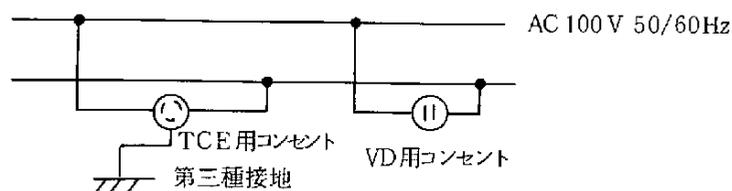
(1) TCE コンセント

- ・ 3 極引掛形コンセント (単相用)
- 10 A 250 V / 15 A 125 V (2 重規格)
- ・ 第 3 種アースが必要 (大地電流は，TCE 1 台あたり最大 3.5 mA)
- ・ 250 V 20 A 用は使用できない
- ・ 使用例：
- ナショナル (埋込型) WF 2315 B
- ナショナル (露出型) WK 2315

(2) VD 用コンセント

- ・ 通常の家庭用平型コンセント
- ・ 引掛型は必要ない

(3) 電源系統図



3) TCE～VD間ケーブル(5C-2V)収容規格

(1)プロテクター(モール)

プロテクター	収容本数(本)
3号	1
4号	3
5号	4
特5号	5

(2)金属管

①薄鋼電線管

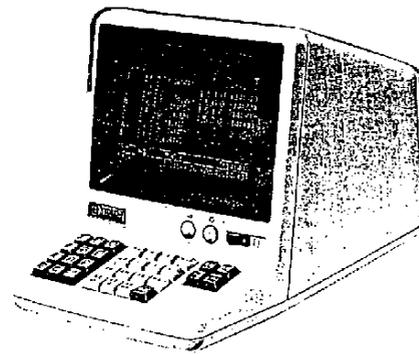
金属管の太さ(mm)	収容本数(本)
19	1～2
25	2～3
31	4～6
39	7～9
51	10～16
63	17～24

②厚鋼電線管

金属管の太さ(mm)	収容本数(本)
16	1～2
22	2～3
28	3～4
36	5～8
42	9～11
54	12～18
72	19～30

(出典：リアルタイム証券情報システムQUICK新ビデオ-I)

・現端末機の機能は以下のとおり。



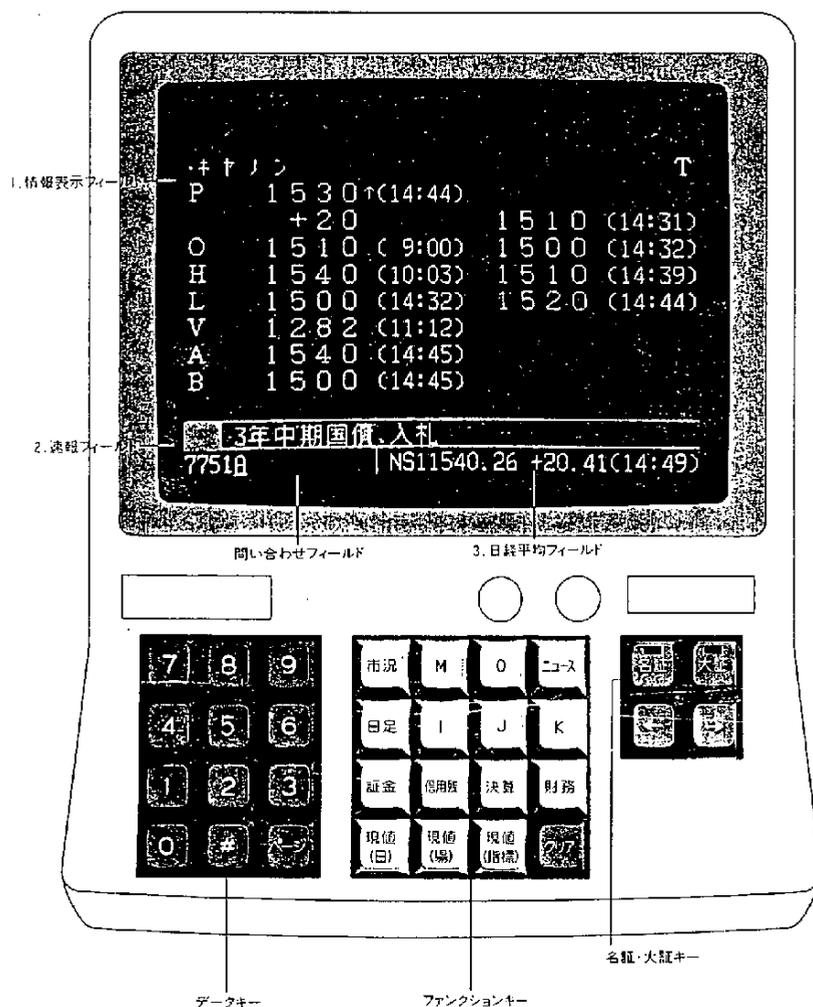
項 目		新 端 末	
ビ デ オ	表示部	ブラウン管	9インチ
		画面構成	半角モード520字/画面 全角モード260字/画面
		文字ドット	半角モード7(ヨコ)×13(タテ) 全角モード15(ヨコ)×16(タテ)
		文字サイズ(mm)	半角モード3.0(ヨコ)×5.3(タテ) 全角モード6.5(ヨコ)×6.5(タテ)
		表示色	モノクローム
		表示可能文字	漢字(約6000種=JIS 第1・第2水準)、ひらがな、カナ、英字、数字
		その他表示属性	リバース プリント
	諸元	高さ×奥行×幅(mm)	270×380×240
		重量(kg)	約10
		電圧(V)	AC 100
		周波数(Hz)	50/60
		電力(VA)	約50
		発熱量(Kcal/h)	約43
		コンセント	2極平型コンセント
電源コード(m)	5		
コ ン ト ロ ー ラ ー	諸元	高さ×奥行×幅(mm)	580×400×220(ただし、脚部は320)
		重量(kg)	約40
		電圧(V)	AC 100
		周波数(Hz)	50/60
		電力(VA)	約250
		発熱量(Kcal/h)	約215
		コンセント	3極引掛型コンセント(3P15A 125V)
		電源コード(m)	5
ケ ー ブ ル	信号ケーブル(コントローラ→ビデオ)	—	
	同軸ケーブル(カ)	100m(現行ケーブルを適用)	
回 転 台	諸元	高さ×奥行×幅(mm)	37×240×230
		重量(kg)	約3
		回転角度	270度(左右135度ずつ)
		電源コード(m)	5
ハ ー ド コ ピ ー	諸元	諸元	モノクロ 感熱記録方式 約20秒/画面
		信号ケーブル(ビデオ→ハードコピー)	3
		電源コード(m)	3
		コンセント	2極平型

(出典：リアルタイム証券情報システムQUICK新ビデオ-Iパンフレット)



5. データベース機能の概要（レコード内容、検索機能等）

- ・昭和61年9月に、株式を中心とするビデオ-Iに合わせて株式、債券、為替、商品を総合した情報サービスシステムQUICK-10が加わり、総合的なサービスとして実働する。
- ・検索機能等については、ハード、ソフト、ファイル構造、通信機能や端末機の機能にすべて依存している。ユーザーの任意性（例えば、提供者のデータベースを利用者が編集し、private fileとして保存すること）も、考慮している。
- ・ビデオ-Iの端末機のファンクションキー等は、以下のとおりで、極力、操作が簡単でスピーディに情報が検索できるようにしてある。ニュース・フラッシュ、平均株価は自動更新するので問い合わせる必要はない。



(出典：リアルタイム証券情報システムQUICK新ビデオ-Iパンフレット)

- ビデオ - I の表示フィールドは、次の 3 種類から成る。

1) 情報表示フィールド……株式、債券価格のほか、各種指標、信用取引、決算、財務、ニュース本文、海外データ等を表示する。

2) 速報フィールド……取引所の個別銘柄の特別な動き（売買停止、中断、板寄せ等）や最新ニュース見出しを 30 秒毎に自動更新する。

3) 日経平均フィールド……日経平均株価 225 種の最新データを常に表示し、1 分毎に計算し自動更新する。

- サービスを提供する際の一番の問題点は、回線の容量である。端末は 2400 ビット / 秒だが、基幹回線は国内では高速デジタルを使用、海外は例えばロンドン - 東京間は 56,000 ビット / 秒と 14,400 ビット / 秒 2 回線の 3 回線を使用している。
- ネットワーキングは重要なポイントとなる。

## 6. データの構造化と入力情報の分析

- リアルタイムの数値情報は、データの更新、加工、検索をスピーディーにすることが最重要であり、システム設計思想の基本となる。
- データの手入力は、QUICK 本社の情報管理部と銀行・証券会社その他で行なう。各証券取引所の数値データを中心に利用者へは加工しないで提供することを基本とするが、指標データとして加工するものもある。
- リアルタイムシステムとはいえ分析的なデータも要求されるが、市況情報センターとしては、分析・予測等についてはそれほど重厚なものをやる必要性は乏しいと思っている。それよりも、情報の信頼性をどこに置くかが大切である。
- 利用者からの要求で一般にも使用できるグラフは QUICK で作成するが、9 月からの総合サービスでは、ユーザーが独自でグラフ処理等の加工をできるようにする。
- 次に、ビデオ - I の情報の流れを示す。



## 7. DBMSの選定・開発

- 自社開発。  
新システム開発中。

## 8. ファイルの設計

- 自社開発。  
新システム開発中。

## 9. 入出力の設計

- 自社開発。  
新システム開発中。

## 10. 標準化ツールの開発と利用

- 自社開発。  
新システム開発中。

## 11. 情報分析

- アウトプットの表示の仕方にグラフ化などの工夫をしているが、これまでは計算機内でグラフの形でファイルを作っていた。出力結果を利用者が処理・分析するような仕組みは、QUICK-10で行なう。

## 12. 各種マニュアルの作成

- マシンセンターの運用はNBSに外注しているので、Operation マニュアルは外注用に作成している。
- ユーザーマニュアルは、情報管理部運用課で作成しているが、今後は画面に表示するようになりたい。マニュアルの中で最も大事なものと思う。
- 「QUICKビデオーボンドマネー(BM)プライベートページ入力操作マニュアル」と「QUICKビデオーEP操作マニュアル」を次に列挙する。

1) QUICK ビデオーボンドマネー (BM) プライベートページ入力操作マニュアル	
1. プライベートページの概要 .....	1
2. キーボード .....	2
3. 入力操作 .....	7
3-1 入力操作の概要 .....	7
3-2 入力操作にあたって .....	7
3-3 入力画面の説明 .....	12
3-4 入力操作の手順 .....	15
4. 操作メッセージ .....	20
5. システムメッセージ .....	21
6. 使用上の注意事項 .....	23
・障害時の操作手順を連絡先 (コール・センター) .....	24
2) QUICK ビデオーEP操作マニュアル	
1. 概 要 .....	1
2. 機器外観 .....	2
3. キーの種類と機能 .....	3
4. 画面構成 .....	9
5. 操作方法 .....	10
5.1 操作一覧 .....	10
5.2 電源の投入・初期設定及び電源の切断操作 .....	10
5.3 項目コード予約操作 .....	13
5.4 項目名登録操作 .....	14
5.5 項目名修正操作 .....	15
5.6 項目データ (当日値・前日値) 入力操作 .....	16
6. 日本語 (漢字) 入力の操作手順 .....	18
7. 表示メッセージとその対処 .....	30
8. 関連部署への連絡 .....	32
9. サービス時間と開始前・終了後の状態 .....	33
10. 機器の仕様 .....	34
11. QUICKからのお願い .....	36

### 13. 運用・開発での体制の確立

- ・システム運用面では、①開発用、②Back up、バッチ処理用、③Operation用のコンピュータ資源の配分や組み合わせのつめ方が最大の問題となる。
- ・ソフト要員が100人ほどいるが、開発に応じ時間的にも、労力的にも波がある。これは計算機運用の管理面での最大の問題となっている。
- ・大型コンピュータを設置すれば、場所の問題と人員の問題が生じ、かつ、新しい技術や手法に対し見通しのきく人間の必要性が出てくる。
- ・全体の整合性、バランスをとるのは非常に難しい。
- ・人材教育を要する。

### 14. サービス料金体系（ビデオ-Iの場合、オプションサービスは別途料金）

- ・情報提供料 10万円/月。
- ・情報提供料のなかには情報料、機器の利用料、特定通信の回線料、保守料、機器の取付工事費などが含まれている。

### 15. 商品化のチェックリスト

- ・商品テストを行ない、テストユーザーを使う。

### 16. アドバイス、意見

- ・供給が需要を創出すると考えること。
- ・ユーザーは気まぐれであることを念頭に置くこと。
- ・開発と商品化には多大な時間と労力を要する。
- ・商品として売り出す情報システムと、内部的に使うシステムでは考え方が全く違い、かかる手間も全く異なる。
- ・利用者の環境条件がすべて異なり、システム稼動の際、考えられないような事が起こりうる。
- ・株価ファイル等、単純なものほど難しい。
- ・競争者のいない産業は、発展の可能性が乏しい。商品の品質の維持・向上のためには、競争が必要である。
- ・人間は与えられる情報に流されやすい。ユーザーが自分で情報を選択することが必要でその時代が来るのは早い。

- ・プロデューサーが情報提供メディアを確保するためには、情報と媒体の方向性を見失わないようにすることが必要である。

#### 収集資料

- 1) 第14期事業報告書（自昭和59年8月1日～至昭和60年7月31日）
- 2) 会社経歴書
- 3) リアルタイム証券情報システム QUICK新ビデオ-I パンフレット
- 4) リアルタイム指標一覧表示システム QUICK指標ボード パンフレット
- 5) QUICK
- 6) QUICKビデオボンドマネー（BM）プライベートページ入力操作マニュアル
- 7) QUICKビデオEP操作マニュアル

## C. (株)住宅産業研修財団(HIC)

### 0. データベースの概要

- HIC(ハウジング・インフォメーション・センター)は、増え続ける建築部材関連の大量の情報を、データベース化し、オンラインで全国に提供する世界でも初めての情報システムである。

文字情報については、ホストコンピュータから各利用者の端末へ電話回線で直接送られ、画像情報は、パソコンとレーザーディスクを組み合わせたHIC端末装置で、文字情報とリンクさせて見ることができる。昭和60年11月からサービスを開始している。

### 1. 会社概要

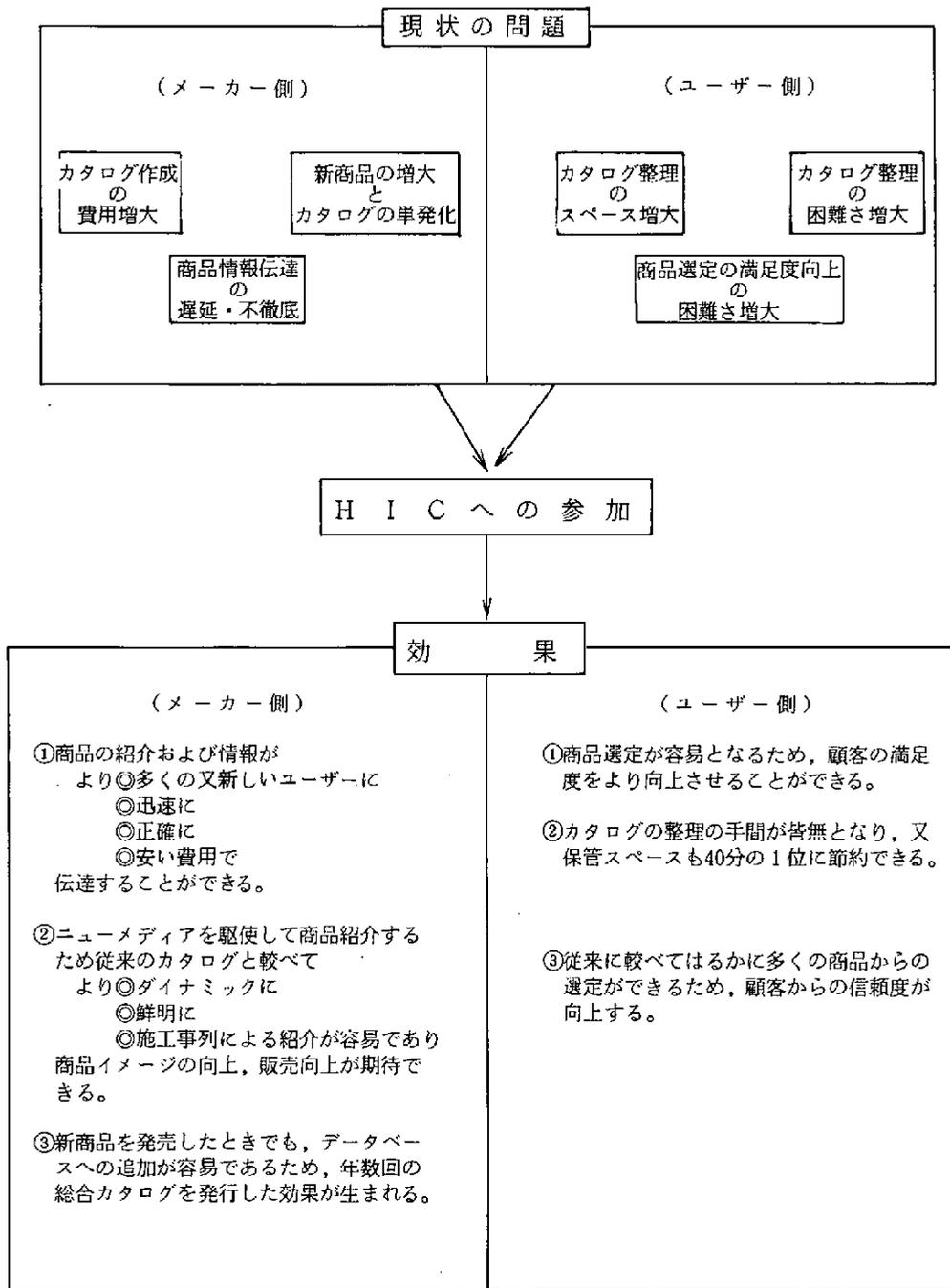
- (財)住宅産業研修財団は、新日本製鉄、トヨーサッシ、殖産住宅相互、三井不動産、竹中工務店、清水建設、松下電工、ミサワホーム、森ビル開発などが理事会社であり、基金3,000万円によって、建設省住宅局の許可法人として昭和52年12月24日設立された。情報サービスセンターである「HIC東京」を赤坂に設置している。

### 2. データベース構築にあたっての事前調査・モデルケース

- モデルケースといったものではなく、世界でも初めてのもの。
- ビデオディスクの方式、オンラインプロトコル、データベース、パソコンOSの選定に当っては米国で調査し決定した。
- 全国オンラインネットワークと分散型オンラインの比較検討。
- カラー画像電送手段(デューダー)の検討。
- 映像撮影方法の検討。
- 分類体系の確立に多大の時間を要した。
- 他にソフトウェア(D/B内容)は建築部材の利用者の立場から望ましい姿を設定した。
- 操作方法の容易さは特に留意した。

### 3. データベースサービスとネットワーク

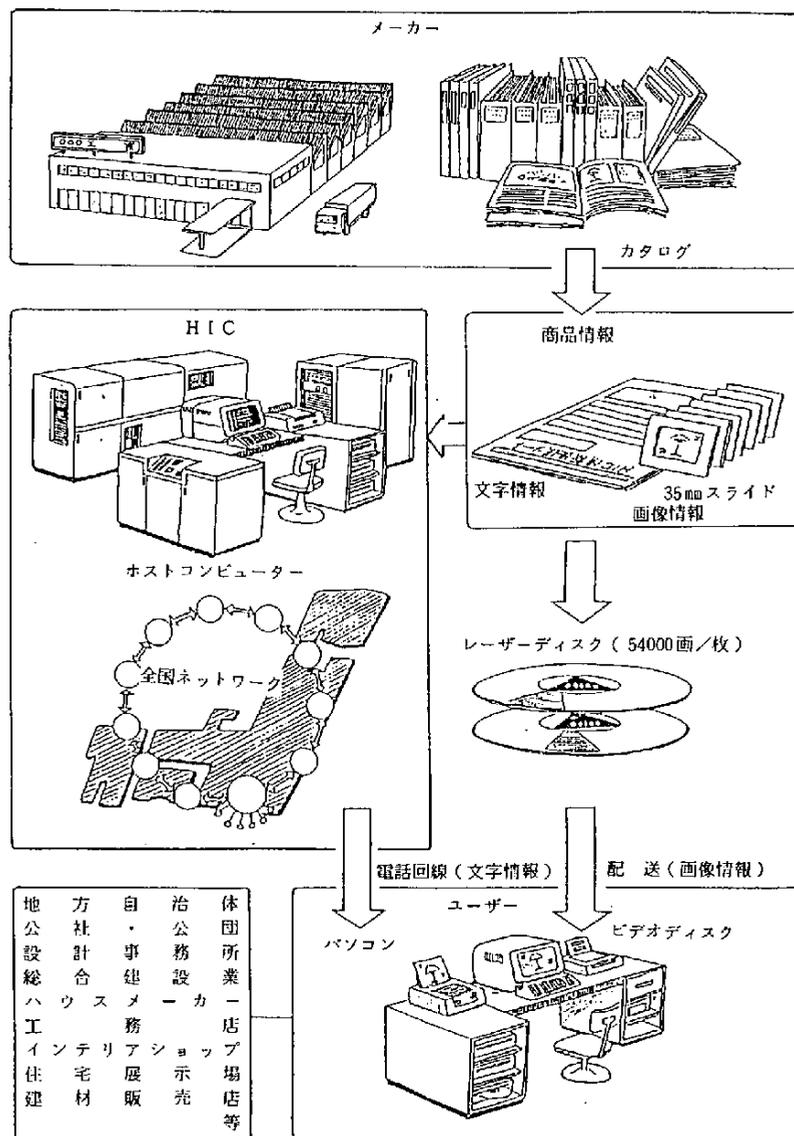
#### ① データベース化の目的とその効果



② ネットワーク

- ホストコンピュータにはプライム社のスーパーミニコンを用いている。
- ホストコンピュータは、現在東京にしかないが、昭和61年5月に大阪にも設置する予定である。将来的には一県一センターとし、分散型のシステムを構築することを考えている。
- 文字情報は、電話回線によって送られる（1200 BPS 全二重）。
- 画像情報は、レーザーディスクに入れて各利用者のもとに配送される。
- 地方用として、文字情報をハードディスクに収録し、ホストコンピュータとオンラインしないで情報を検索、使用できるシステムを今夏をめどに開発中。
- さらに、CD ROM が正式に実用化された場合への対応も検討中。
- 情報の流れを次の図で示す。

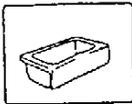
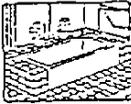
H I C シ ス テ ム



- 端末装置（標準タイプ）を構成するのは、パソコン本体、（レーザーディスク）プレーヤー、ビデオモニター、文字プリンター、スーパーインポーター、オンライン用機器（モデム）である。
- 現在画像のハードコピーをとるためのカラープリンターは7月にオプションとして発表する。

#### 4. 情報源と情報収集方法

- 情報源は、住宅、ビル建築、土木関連商品を製造、販売している企業（昭和60年11月現在448社）。
- 各企業は、情報掲載費と制作費をHICに対して支払う。費用は、冊子体の総合カタログに掲載する際の費用に比べ安く、小企業でも情報を掲載できる。費用は以下の通り。

	画像情報		文字情報	
	（例 1品目 2画像の場合）		（表）	（裏）
			商品データ(必須) №1シート _____ _____ _____	商品説明(オプション) №2-1 ----- №2-2
情報掲載料 （月額）	100円/画	100円/画	0円	0円
情報制作費 （初回のみ）	1,500円/画	1,500円/画	(№1) 400円/シート	(№2-1) 400円/シート (№2-2) 400円/シート

- 各企業は、文字情報については、「HIC情報提供シート」に記入し、また画像は、35mmスライドにして提供する。

#### 5. データベース機能の概要（レコード内容、検索機能等）

- 提供される文字情報の内容

商品名称、会社名、名称・用途、品番、問合せ先、住所、電話番号、形状・寸法、基本材質、認定マーク、価格、単位、フレームNo、商品紹介、（漢字600文字以内）。

- 画像情報は、部品、部材の単位写真、施工例、性能値、特徴などを図表にした静止画像である。1件の文字情報に対し、画像は何枚でもよい。
- 画像情報を収めるレーザーディスク1枚につき54,000件の画像を入れることができる。

- ・検索方法はオンライン方式として4種類用意されている。

① 小区分番号検索とは、分類表の番号6桁を指定する方法。

② フレーム番号検索とは、あらかじめレーザーディスクの画像をスタンダアロン方式でみておきその画像のフレーム番号(画面左上の5桁の番号)をメモしておき、オンラインにしてフレーム番号をキーインすることにより、画像文字情報を得る方法。

③ 会社記号検索とは、3桁のアルファベット記号を指定し(会社別に決まっている)。(1社1記号)会社別に商品を見る方法。

④ 大区分番号検索とは、大区分→中区分→小区分と分類表をみないで画面上の指示による操作のみで、指示してゆく方法。

である。

- ・このあと、会社名、あるいは、最高価格、最低価格でさらに絞り込むこともできる。
- ・スタンダアロン方式とは、センターのホストコンピューターと電話回線で結ばないで(オンラインしないという事)レーザーディスク・プレーヤーの操作(リモコンユニットを使用)のみで、画像情報を見る方法をいう。
- ・レーザーディスクは、年2回入れ替えられる。文字情報はリアルタイムで更新される。
- ・文字情報のレスポンスタイムは必ずしも速くない。今後利用者が増え、回線が混んでくるとさらに遅くなる可能性もあるので検討を要す。データの更新が頻繁に行なわれる訳ではないので、文字情報についても記憶容量の大きい媒体に入れて、画像情報のように提供するほうがよいかもしれない。
- ・レスポンスタイムを上げるために、パソコンのソフトをうまく活用している。
- ・データ数は現在40,000件程度だが、10万~20万に増やすことを考えている。

## 6. データの構造化と入力情報の分析

- ・各企業から持ち込まれる情報は、シート記入方法の誤り等は訂正するが、内容は手を加えずに入力する。内容についての責任は、一切各企業側にある。
- ・各製品に対し、分類表に従って6桁の小区分番号を付る。この作業も各情報提供者が行なう。
- ・複合製品の場合(例えば、卓上灯と時計がいっしょになったもの)には、基本的にはその主目的で分類番号を付るが、必要に応じて3つの区分にまで登録することが出来る。

## 7. DBMSの選定・開発

- ・既存のソフトでそのまま使えるものはなかった。プライム社の「Information」というソフトの日本語版を採用したが、それ以外はすべて独自に開発した。ソフトウェアとしては、ホストコンピュータ、及びレーザーディスクとパソコンのインターフェース、パソコンの各々のためのものが作られた。

## 8. ファイルの設計

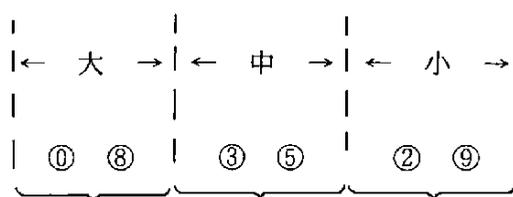
- ・検索時の応答時間を速めるために、小区分番号検索、大区分番号検索を行なった時あらわれる画面（収載件数、会社数）は、あらかじめ別ファイルにしてある。

## 9. 入出力の設計

- ・文字情報については、入力用フォーマット（HIC情報提供シート）を作り情報提供者（企業）に記入させている。入力のためのパンチ作業は外注している。
- ・35mmスライドになっている画像をレーザーディスクに入れる作業は外注している。
- ・どのような出力が必要か、ということを考えることは、すなわち利用者は何を求めているかということを考えることである。このデータベースでは、検索されて出てくる建築部材等が、自分にとって使えるものであるかどうか判断するために必要なものは何かということ考えた。

## 10. 標準化ツールの開発と利用

- ・HIC・INDEX一分類表と情報提供会社一覧から成り、コード体系は、次のようになる。  
(1) HIC分類表
  - ・アメリカ、カナダの土木、建設業界の仕様書のコード表である「マスター・フォーマット」をもとに修正を加えたもの。4000の小区分がある。
  - ・HIC分類表は、6桁の数字より構成されている。



大区分      中区分      小区分

- ・大区分は、「公共住宅建設工事共通仕様書」をはじめとして、建築業界で一般に用いられている工事別・部位別区分を中心として構成されている。

(例) 大区分：040000 外構工事

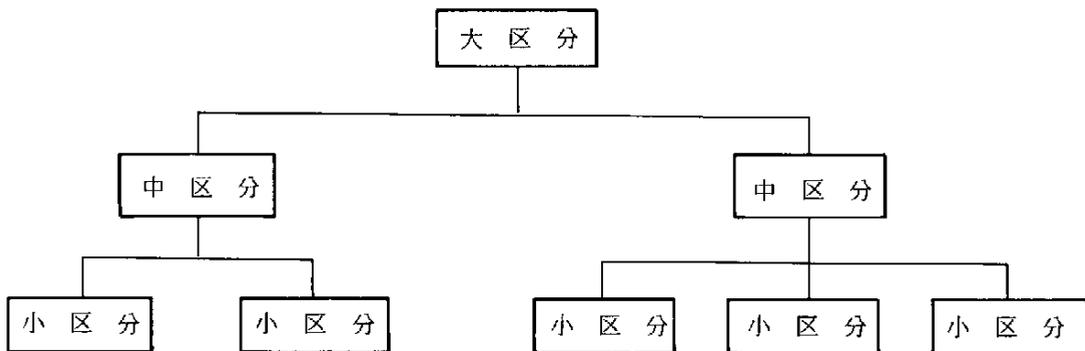
- ・中区分は、システム、材質別が主体となった区分となっている。

(例) 中区分：040100 門・塀  
041500 公園・遊園地施設

- ・小区分、形状、色、柄等が主体となり中区分をさらに細分化したものである。

(例) 小区分：040101 コンクリート、現場打ち  
041500 コンクリートブロック門

### 分類体系



- ・具体例

24：照明器具

240100 ランプ(光源)

240101 白熱灯

240111 蛍光灯

240121 高圧放電灯・ロゲンランプ  
・レフランプ

240300 安定器および付属品

240301 安定器および付属品

## (2) 情報提供会社一覧

- ・情報提供会社名・会社記号・フレーム番号が、情報提供会社名の五十音順に配列されている。会社記号は英文字3桁で、情報を受付た順にHICが付与する。
- ・具体例

### 情報提供会社名

---

あ

(株)アイエム換気工業	A C C
〃	A C C
アイエルビー(株)	A K F
アイカ工業(株)	A A M
愛知県陶器瓦工業組合	A K U
明石窯業(株)	A E Y
アキレス(株)	A B M

## 11. 情報分析

- ・文字情報と画像情報がドッキングした出力結果をもとに、利用者の主観的な判断が出来るような一次情報の提供とし、出力結果をさらに処理、分析するような仕組みはない。

## 12. 各種マニュアルの作成

- ・利用者マニュアルと情報提供者のデータ記入用マニュアルの簡略化したものがある。

## 13. 運用・開発面での体制の確立

- ・開発には約2年を費やしている。
- ・開発は、当財団が中心となり、コンサルタント、ソフトウェア会社、ハードウェア会社(NEC、パイオニア、エプソン、プライムなど)の協力を得て行なった。
- ・社内では、10～15人が係わっており、データ記入ミスのチェック、パンチミスのチェック、画像のチェックなどを行なっている。

#### 14. サービス料金体系

- 会員として 40,000 円／月。これには、レーザーディスクの提供料、INDEX の提供料、文字情報の提供料が含まれる。
- HIC 端末は利用者が購入する場合約 200 万円である。リースの場合は、5 年リースで 40,000 円／月である。
- これらの料金の中には回線料金は含まれない。

#### 15. アドバイス、意見

- システム開発には大変な費用と人数がかかっており、とても一企業でできるものではない。ビジネスとして成り立つには難しいだろう。HIC の場合は、理事会社の協力、建設省、通商産業省、各種団体の後援があったことが大きな力となっている。
- ハードウェアの進歩が、大きな影響力を持つ。このデータベースの発想自体は 10 年以上前からあったが、レーザーディスクによって、静止画像の提供ができるようになるまで、システムを実現できなかった。
- 利用者志向でデータベースを作るべきである。また、ターゲットを明確に決めておく必要がある。
- 利用者（このデータベースでは、総合建設業者、ハウスメーカー、設計事務所、工務店など）がはっきりしていれば、システム、ソフトウェアの開発は、比較的やさしいのではないかと。
- 情報提供者と利用者を啓蒙するのが大変であった。
- データベースを作ろう、として作ったものではだめで、必要があって作ったものなら使える。
- 利用者は、ほしい情報が少ししかない、データベースの利用はコスト的に割高だと感じる。注意を要する。
- 利用者がデータの二次加工できるようにする必要があるのではないかと。そのためにフォーマットやレコードのレイアウトなどを公開するようにした方がよい。

## D. (株)ゼンリン(地図データベース)

### 0. データベースの概要

- ・(株)ゼンリンの地図データベースは、誰が、どこに住んでいるか等を、地形図情報と戸別名称などの文字情報、および住所コードなどの基本属性から検索できるシステムをいう。また、地図情報利用システムを介して、ユーザー保有の属性情報と結びつけ、顧客管理、不動産管理、施設管理等への展開も可能である。

### 1. 会社概要

- ・株式会社ゼンリンは、昭和23年に創立した善隣出版社を母体とし、住宅地図の出版・販売をベースに、多様化するニーズに対応するよう、情報企業としての飛躍を期し昭和58年にグループ3社を合併して発足した。
- ・事業所は、北九州市小倉北区の本社をはじめ、全国に支社及び営業所を持ち、1,323名の従業員を擁している。

### 2. データベース構築にあたっての事前調査・モデルケース

- ・昭和56年4月、各種のショウでテストをしてもらった。このときはプロッタの性能が悪くて、成功しなかった。
- ・同年10月に、日立がさきのテストをふまえて、1) トータルな印刷システム、2) データベースサービスのサポート、を考えると提案してきた。

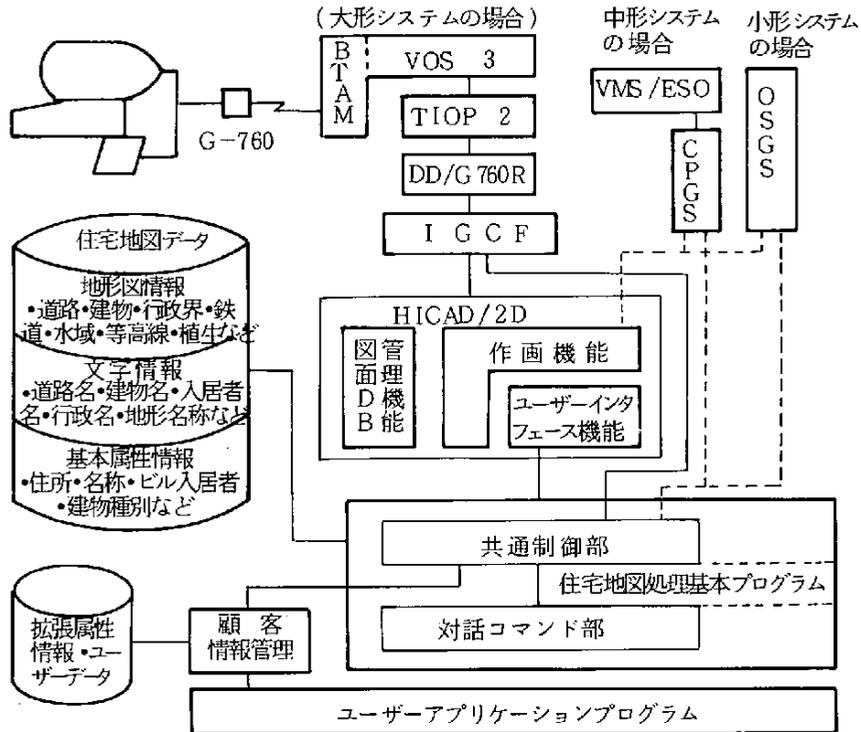
### 3. データベースサービスとネットワーク

#### ① システムサービスの沿革

- ・昭和60年5月、日立製作所との共同開発により住宅地図製作システムが完成、東京23区のデータが整備され、販売を開始した。

② システム構成

・以下に、住宅地図情報利用システム構成を示す。



注：略語説明

- VOS3 (Virtual-storage Operating System 3)
- VMS/ESO (Virtual Machine System/Extended Option)
- CPGS (Control Program for GRADAS)
- OSGS (Operating System for GRADAS)
- BTAM (Basic Telecommunication Access Method)
- TIOP2 (Terminal Input Output Program 2)
- DD/G760-R (Device Driver/G760-Raster Type)
- IGCF (Interactive Graphic Control Facility)
- HICAD/2D (Hitachi Computer Aided Design System for 2D Geometry)

(出典：住宅地図情報利用システムの開発，日立評論 Vol.66 No.12 1984-12p.43-46)

#### 4. 情報源と情報収集方法

- 2,500分の1の都市計画図をベースにし、誰が、どこに住んでいるかという情報を収集する。
- 情報の内容は、世帯主名、住所、番地等である。
- 情報収集方法は、原図については2,500分の1地図を1,000分の1に拡大したものをもとにし、自治体の都市計画課・土木課・住居表示担当課より資料を入手し、また地下街については担当部所から、一方通行については警察からデータを得ることにより、それぞれをつきあわせて、より正確な図面とする。住宅部分については、調査員が実際に現地を歩いて回って調査しており、表札・看板などのオープン情報からデータを得る。明確なデータが得られない場合は、電話帳あるいは町内名簿等で再調査をしたり、ヒアリングを行ったりする。地下や（立入り可能な）マンション・アパートの入居者についても調べる。

#### 5. データベース機能の概要

- 基本となる地図データベースからは、約100種類の階層化されたデータを自由に組み合わせ、様々な地図を描くことが可能である。さらに、ユーザーデータを組み合わせることにより、適用業務に対応した地図データベース利用システムとなる。このアプリケーションには、下表のようなものがある。

	機 能	適 用 業 種
1. 不動産管理	<ul style="list-style-type: none"> <li>• 物件の所在位置の高速検索</li> <li>• 周辺環境（公共施設、交通機関等）の表示</li> <li>• 間取り図・外観図の表示</li> <li>• 物件の最寄り駅までの経路の詳細表示</li> </ul>	不 動 産 建 設 デベロッパー
2. 顧客管理	<ul style="list-style-type: none"> <li>• 各種顧客情報の経年管理</li> <li>• 商品種類別の顧客の表示・リスト</li> <li>• 得意先分布地図</li> <li>• 地域別顧客リスト</li> </ul>	銀 行 保 険 証 券 デ ィ ー ラー 各 種 販 売 店
3. 施設管理	<ul style="list-style-type: none"> <li>• 地下・地上施設の設置状況を地図上に表示</li> <li>• 施設の属性（配管口径、種類、配線の系統敷設年月日、番号など）を高速に検索表示</li> <li>• 施設の稼働・保守情報の表示</li> <li>• 大規模工場敷地内施設の総合管理</li> </ul>	電 力 ガ ス 上 下 水 道 電 信 電 話
4. 集配送管理	<ul style="list-style-type: none"> <li>• 集配送先の高速検索・表示</li> <li>• 交通規制情報の表示</li> <li>• 集配送地図の出力</li> <li>• 交率的な集配送計画の立案</li> </ul>	運 送 メ ッ セ ン ジ ャ ー タ ク シ ー
5. 都市情報管理	<ul style="list-style-type: none"> <li>• 総合司令用地図、警備用地図、防災作戦図の作成</li> <li>• 事故・災害現場の高速検索</li> <li>• 連絡所、防災関連施設、防火用水や消火栓などの位置検索</li> <li>• 避難場所、避難路の色分け表示</li> </ul>	警 察 消 防 警 備 保 障 自 治 体

## 6. データの構造化と入力情報の分析

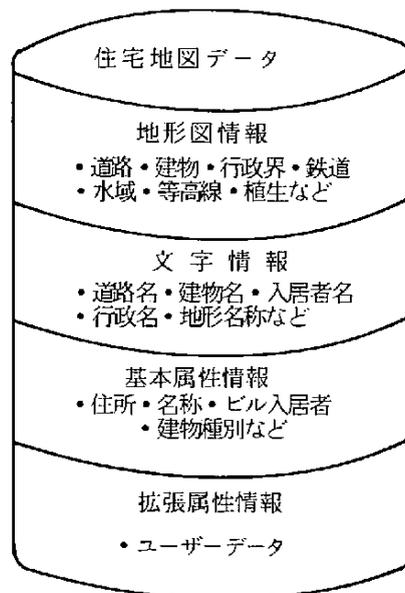
- 地図のデータは、道路、建物、行政区画等、約100種類の様々なデータで構成され、これらを自由に組み合わせて抽出できるように、階層化して管理されている。
- 地名の階層化については、データ入力時に行っている。

## 7. DBMSの選定・開発

- 日立との共同開発による。
- 利用者は、リレーショナル・データベースを使用する。
- それぞれのアプリケーションにもよるが、例えば福岡市ではパソコン上のR:BASE 4000を使用して管理している。
- 現在、ゼンリンにて、利用者向けの新しい基本システムを開発中であり、61年10月には発売の予定である。

## 8. ファイルの設計（ファイル構造）

- データファイルのおおよその構造を下図に示す。



（出典：住宅地図利用システムパンフレット（株ゼンリン））

・ IGES 準拠の標準フォーマット

地形図データは、IGES (Initial Graphics Exchange Specification) 形式のファイルに保存されている。

IGES では全てのデータをカードイメージすなわち 1 レコード 80 バイトである。

地形図ファイルは、5 つのセクションで構成される。以下の表にその概要を示す。

セクション項目名		内 容
スタートセクション項目		作成者氏名, 作成年月日, 地図名, 図名
グローバルセクション項目		デリミタ定義, ファイル名, システム名, ANSI 版番号, 定数語長指定, モデル空間指定値, 単位・線幅日付, 精度他
ディレクトリ	要素第 1 レコード	要素番号, ポインタ, IGES 版番号, 線フォント, ステータス他
セクション項目	要素第 2 レコード	要素番号, 線幅, ペン番号, データ数, 文字要素番号, 基本項目番号他
パラメータデータセクション項目		要素番号, X・Y・Z 座標, その他パラメータ ディレクトリセクション項目で指定した要素により, 次の 5 種類のデータが記述できる。 1) 線分要素 2) 折線要素 3) 円弧要素 4) 自由曲線要素 5) 注記要素
ターミネートデータセクション項目		前 4 セクションの最大シーケンス番号, 当該レコードのシーケンス番号

・ 建設省白地図 DB 標準フォーマットへの対応

フォーマット案が固定されたのち, 対応すべく検討中。

## 9. 入出力の設計（機器）

- 図形の入力には、スキャナ式のオートデジタイザ3台、同じくハンドデジタイザ12台を使用している。オートデジタイザは、1台は日立（イメージ入力と同時に、ベクトル化）、2台はキャディクス（リニア式、イメージ修正後、自動処理にてベクトル化）のものである。日立のスキャナは、もとはプリント基盤用に開発されたものを共同開発しており、現在は建物の枠をスキャンするのに使っている。これを使用することにより、例えば江東区の密集地等では20分の1の工程短縮ができる。
- 図面内の文字の入力にはペンピュータを80台、また文字の合成には日立製のGMMを使用中だが、スキャナーによる自動方式を目下研究中である。
- 図面の出力は、グラフィックディスプレイの他に、カラーハードコピー、16本/mmの静電プロッタ、住宅地図の印刷原版とするために1mm当り32本というレーザープロッタから印刷原版にダイレクトに出力する。

## 10. 標準化ツールの開発と利用

- 基本属性情報のうちの住所、地番は、下記のようなデータ構造となっている。下9桁はゼンリンで独自に作成したものである。

( )内は桁数

住所コード (11)			街区 番号 (2)	地番 (7)
県市町村(5) (自治省コード)	大字コード (3)	小字コード (3)		

- ユーザーの持つデータベースに、この形の住所、地番コードを与えることにより、マッチング処理が施せる。それによりオリジナルなデータベースの合成が可能となる。

## 11. 情報分析

- 地図の解析的な利用については、アプリケーションがないため、現在のところほとんど行っていない。

## 12. 各種マニュアルの作成

- ・社内用マニュアルは備えているが一般向けのものではなく、ユーザーごとにオリジナルのマニュアルを作成する。
- ・昭和61年春には、データ構造のみのマニュアル及びアプリケーション用マニュアルが完成する予定である。

## 13. 運用・開発面での体制の確立

- ・システム開発（センター・オペレーター兼務）は37名、またデータ入力・校正・前処理・マネジメントは70名で3交代制、その他、文字入力（製作部兼務）のために数十名擁している。
- ・コンピュータ関連部分の人材は、3名の出向者を除き、全て自社内で賅っている。

## 14. 費用対効果

- ・システム開発に投入した初期投資は、業務の効率化による人件費のコストダウンと地図の生産性向上で充分まかなわれている。更に、本データベースの作成により、（株）ゼンリンの知名度が上るとともに、冊子体の住宅地図の需要が増えたことも考えると、既に充分の効果をあげている。当初、昭和61年まではデータベースの売上げを予算に計上していなかったが、既に販売見込みが出ている。このような事情から商店主等の一般の人々が入手可能な価格で、住宅地図データベースを販売していく方針をもっている。

## 15. 商品化のチェックリスト

- ・データの提供単位は1市町村、1エリア、1画面と3種類あり、料金は、例えば1画面の月額使用料は約2,000円、東京23区全域では同じく100万円程度である。
- ・データは基本的には磁気テープに記憶させているが、小さいエリアであればフロッピーディスクに収まる。本年中頃には、CD-ROMによる提供を考えているが、CDのドライブとパソコンのOS（MS-DOSを基本に考えている）とのインターフェースに若干の問題がある。
- ・オンラインサービスについては、将来的には行うつもりである。現在行っていない理由は、(1)自社のセンターの管理が未整備、(2)データ圧縮が不十分、(3)ネットワークの未整備である。このうち、(3)のネットワークについては、NTTとの検討を行うとともに、CSKの共同VANに出資して、整備を待っている状況である。

- ・ FAX サービスについては、現在別会社（MRDゼンリン）にて検討中であるが、回線数や料金面で問題がある。需要については、不動産業等でかなり見込めそうである。
- ・ 現在データベース化済みのエリアは、東京23区、武蔵野市、三鷹市、横浜市、川崎市（一部）、福岡市である。昭和61年中に、政令指定都市の6～7割をカバーする予定である。
- ・ 一般向けの商品化については、社内データベースの入力が完備する61年度末～62年度頃になると思われる。価格については、どの程度であれば使ってもらえるかを念頭に現在検討中である。

#### 収集資料

- 1) 住宅地図利用システム（株）ゼンリン パンフレット
- 2) （株）ゼンリン パンフレット
- 3) 「住宅地図製作自動化システムの開発について」
- 4) 住宅地図データ交換フォーマット

## E. (株)帝国データバンク(COSMOS)

### 0. データベースの概要

- COSMOSとは、企業財務、財務分析、企業情報、個人情報等をオンラインや磁気テープ、フロッピー・ディスクで提供している(株)帝国データバンクのデータベース・サービス・システムのことをいう。

### 1. 会社概要

- (株)帝国データバンク(旧称:帝国興信所)は、信用調査機関として明治33年に創業。現在は、企業の信用調査を中心に活動している。
- 東京本社を中心に、大阪に支社、63支店、14出張所を国内に持つほか、海外にも30カ所の提携先を持っている。

### 2. データベース構築にあたっての事前調査とモデルケース

- 昭和43年からデータベース作成の発想と着手があった。
- 1963年より米国のDun & Bradstreet社がオンライン・サービスを開始しており、これを参考とした。

### 3. データベースサービスとネットワーク

#### ① 提供データベース・サービス

- 提供データベース
  - 1) 企業信用調査報告書
  - 2) COSMOS1(企業財務、財務分析)〔財務分析システムは長銀経営研究所と共同開発〕
  - 3) COSMOS2(企業情報など6ファイル)
  - 4) COSMOS3(個人情報)

#### • オンライン・データベース

上記4種のデータベースのうち、オンラインでサービスされるのは以下に示すものである。

- 1) については、東京都内および神奈川県のみ
- 2) については、企業財務ファイル

- 3) については、企業情報ファイル
- 4) については、個人情報ファイル
- その他サービス
  - ハードコピー（ファクシミリでのサービスもある）、磁気テープ、フロッピー・ディスク、アドレスタグなどの製品で提供される。
- NEEDS, TelecomにCOSMOS1, 2を提供予定。
- COSMOS1, COSMOS2については、英文化も行なっている。（一部完了）

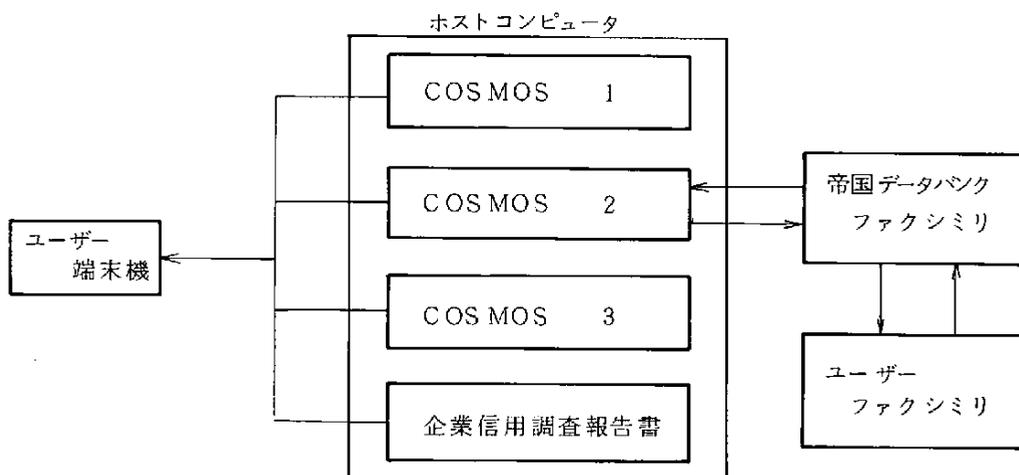
② データベース作成の目的

- 顧客への資料提供の迅速化。
- 省力化。これまでは、書類作成のためにタイピストが必要であり、そのために人件費が増し、これをコンピュータ化することにより、余力を調査にまわすことができる。
- 各地に保管されている多量の印刷資料をどこにいても利用できるようにすること。
- 企業信用調査報告を電算化することが第一目標であった。

③ データベース化の利点

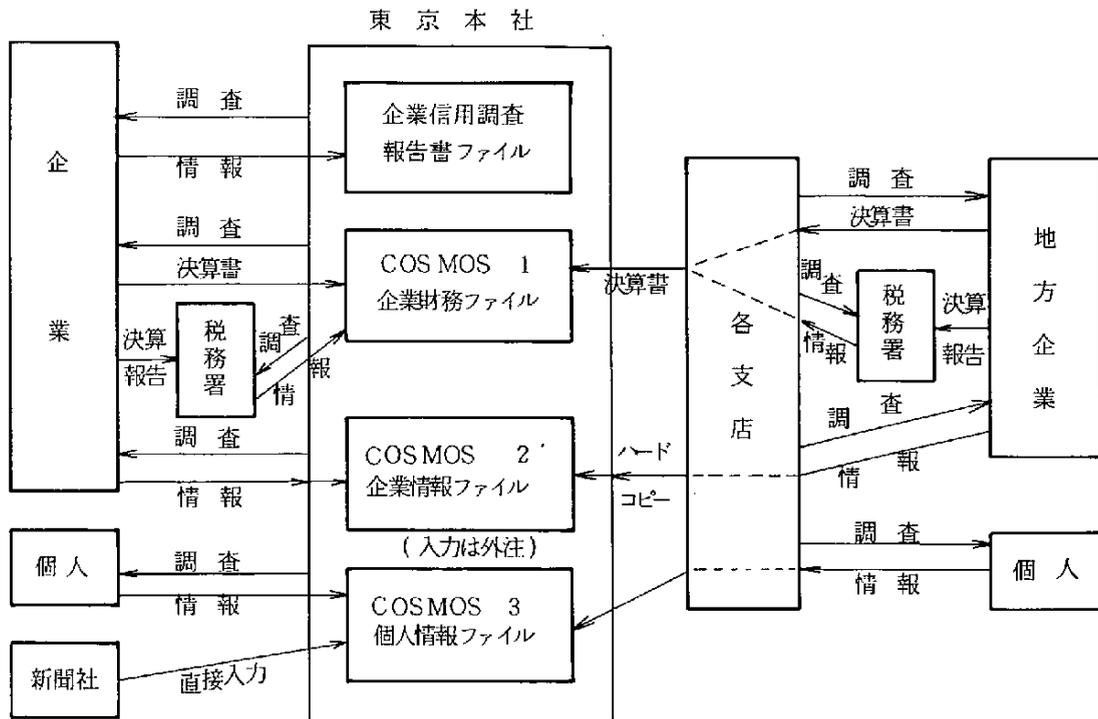
- データに基づいた経営分析が容易になった。
- これまで調査員の勘に頼っていたのが、評点（100点満点）という形で客観的に即時判断できるようになった。
- 評点に対する利用者の信頼が高くなり、それが経営上大きな影響を持つようになったため、企業側が調査に対して協力的になった。

④ ネットワーク



- オンラインサービスは、現在では限られた利用者に行っていない。
- 回線は、都内は専用回線（4800 BPS）、遠隔地はDDX（パケット）を使用している。
- 使用コンピュータ
  - 1) 日立 240H — 企業信用調査報告書、および各種オンラインサービス用
  - 2) 富士通 170F — ハードコピー、磁気テープ提供用
- 利用できる端末機
  - 1) 日立 T-560/20Ⅲ ビデオデータターミナル
  - 2) 日立 PT-1、PT-1EX パーソナルターミナル
  - 3) 日立 B16、B16-EX パーソナルコンピュータ
  - 4) 日立 2020、2050
  - 5) リコー SP150、250、350

#### 4. 情報源と情報収集方法



- まず、現在までに集めた多量の印刷資料があった。
- 情報収集は、全国79事業所の約2,000人の調査員が行なう。調査員は、各担当地域の事情に精通しており、信頼は高い。優秀な調査員を確保しておくために、待遇はよくしてある。
- COSMOS2は、メンテナンスは月1回、行なう。顧客から調査依頼があった時、調査し直して新データを入手するが、決算以後8カ月経ても調査依頼がない場合は、独自に調査を行なう。
- 企業信用調査報告書は、社内に入力セクションで処理している。入力については2交替制を採用している。
- 情報収集の新しい手法として、情報収集依頼者側から、収集先について既に得ている情報の提供を受け、それを基に情報を収集する、ということがある。Dun & Bradstreet社は、債券回収情報についてこの手法で成功している。

#### 5. データベース機能の概要（レコード内容、検索機能等）

- レコード内容、検索機能はファイルごとに異なる。以下にオンラインサービスを行な

っているファイルについて、内容と検索キーを示す。

1) 企業信用調査報告書ファイル

内容：摘要企業の概要，登録事項，大株主，代表取締役，設立の経緯，現状と将来性，仕入先，得意先，金融機関からの借入，推定資産，不動産登記写，決算書。（従来の企業調査報告書の内容すべて）

検索キー：企業名（カナ15文字以内），企業番号。

2) COSMOS1 企業財務ファイル

内容：貸借対照表，損益計算書，利益処分書など，財務諸表。

検索キー：企業名（カナ15文字以内），企業番号。

3) COSMOS2 企業情報ファイル

内容：商号，所在地などの企業概要，代表者氏名，住所など。

検索キー：企業名（カナ15文字以内），企業番号，地域，業種，資本金，年商，利益，申告所得，申請年月，取引銀行，従業員，評点，設立年月，上場区分など。これらについて16項目まで組み合わせ可能。

検索結果を売上高順，社名順など自由にソートできる。

4) COSMOS3 個人情報ファイル

内容：氏名，通称，住所，生年月日，家族，学位など約20項目。

検索キー：カナ氏名，出身地，出身校・卒業年，生年月日，趣味，電話番号，役職，住所，企業番号。

## 6. データの構造化と入力情報の分析

- COSMOS1の場合，各社の発表する決算報告は必ずしも信頼できる訳ではないので，税務当局から公示されたものと比較し，チェックを行なう。
- 入力されたデータをもとに利益率などの各種財務分析比率を出している。
- 収益性，効率性，安定性，流動性，成長性，規模等の要素により評価を行なって評点をつけ，総合評価を100点満点で示す。また，評価要素ごとに解説が加えられる。
- 決算書のあらゆる勘定科目をコード化して，各社の決算書の標準化を計っている。
- 業種別，地域別に同業種社数中の売上高順位でランキングを出している。

## 7. DBMSの選定・開発

- 1985年からMODEL 204を導入。
- それ以前は自社開発のものをいっパッチ検索をしていたが、かなり時間がかかった。
- MODEL 204とADABASを比較したが、MODEL 204のほうが処理スピードが速かったので採用した。

## 8. ファイルの設計

- COSMOS 1, 2, 3は、各々全く別のファイルとなっている。現在、MODEL 204を用いて、一体化することを考えている。

## 9. 入出力の設計

- COSMOS 1では、決算書を入力するが、勘定科目が企業によって異なる。それに対応するためにすべての勘定科目を4万に細分化してある。勘定科目はコードで標準化しているので、勘定科目様式が異ったものでも一元的に処理し、また比較できる。必要があればもとの勘定科目の原本に復元できる。決算書の入力は、端末から行なっている。
- COSMOS 1では、わかりやすくするために、出力にレーダーチャート、図型、グラフを多くとり入れている。
- COSMOS 2では、目的別の出力が可能である。
  - 1) サマリーリスト（地域別、規模別など用途別フォーマットで出力できる）
  - 2) 顧客管理カード
  - 3) ダイレクト・メール用ラベル
- 入力データのデータエラーチェックシステムを持っている。

## 10. 標準化ツールの開発と利用

- 帝国産業分類表：日本標準産業分類（総務庁告示）に準拠し、産業を673業種に分類し、コード化してある。
- 帝国データバンク独自の住所コード（10桁コードで、町、丁目レベルまで識別できる）。

## 11. 情報分析

- ・出力結果には、すでに相当量の分析結果が含まれている。それをさらに利用者が処理、分析するようなツールは現在のところ提供していない。

## 12. 各種マニュアルの作成

- ・社内用マニュアルと外注用パンチマニュアル、磁気テープ利用者用マニュアルを備えている。
- ・マニュアルの標準化はまだ不十分なので、整備について研究中である。

## 13. 運用・開発面での体制の確立

- ・システム部門には約30人の技術スタッフがおり、ソフト作成、開発、運用に携わっている。
- ・アプリケーション・ソフトは、細かい仕様になると社内のほうがよくわかるので、社内で作ることのほうが多い。細かくないものについては外注している。
- ・利用者の要求に対応するために、システム部門の6～7人がサービスを担当している。この利用者用は主に磁気テープのユーザーで、ニーズに合わせて磁気テープを加工することが多い。
- ・オンライン利用者に対しては、出張指導等（半日）を行なう。
- ・COSMOS1, 2, 3, 企業信用調査報告書の各ファイルは、オンラインシステム用としては、さらに向上のため開発を続けている段階である。

## 14. サービス料金体系

- ・企業信用調査報告書ファイル

- 1) 一般ファイル：調査問合票 1枚 / 1社  
調査問合票 0.15～0.3枚 / 1画面
- 2) 専用ファイル：調査問合票 50～100円 / 1画面

\*調査問合票は、帝国データバンクの会員のチケットのようなもの。オンラインサービスは、現在はごく限られた会員にしか行われていない。

- ・COSMOS1：企業財務ファイル 4,500円 / 1社
- ・COSMOS2：企業情報ファイル 企業索引 100円 / 1回  
企業概要 1,400円 / 1社

## 15. アドバイス、意見

- 通信網の整備により、利用が広がりつつあるので、より多くの人々のニーズに応えるデータベースの構築を検討すべきである。
- 料金的にも、システムのにも利用しやすいものを作るべきである。
- 質が高く、up to dateなものを作るべきである。そのためには、データ提供者の協力を得なければならない。データ提供者の協力を得るには、give and takeの関係になれるようなものでなければならない。
- 利用者サイドに立ったきめの細かいサービスが必要。
- 今後データベースが発展するためには、日本全体のデータベースという考えに立ち、データベースの共有と、データの収集、入力、加工、提供の分業化が必要ではないか。

## 16. 国・政府に対する要望

- データベース化したいデータを国は数多く所有している。費用は民間が負担するから速かに利用可能の途を開いてほしい。(例えば、法務省の不動産登記の変更書)
- 税務署単位で公示されている法人申告所得や高額納税者については、国側でまとめて、磁気テープの形で提供してほしい。
- データセンターを作る際は民間も参加できるように要望する。
- データベースの構築費用は税法上繰延資産扱いされ、単年度では、その5分の1しか費用化されないが、構築当初ほど費用に対応する利益が少ないので全額費用化できるようこれを改めてほしい。
- しかしながらデータベースを作る際開銀から資金を借入れるためには、担保としての資産性がなくてはならない。この事情も考慮してデータベースを資産にするかしないかは構築者が決める自由を認めてほしい。
- データベースの共有化とインターオペラビリティについて、国が標準化など、統一するための場を作ってほしい。

## F. (株)図書館流通センター (TRC MARC)

### 0. データベースの概要

- TRC MARCは、新刊書ならびに所蔵図書の書誌事項を機械可読目録の形で提供することのできる図書整理システムをいう。

### 1. 会社概要

- 株式会社図書館流通センター(以下TRCと称す)は、出版界18社と社団法人図書館協会(以下JLAと示す)の共同出資によって、昭和54年12月に設立された。
- JLAの整備事業部が行っていた図書館に対する印刷カードの提供、図書の装備等を引き継ぐ。JLAの技術を基に、目録情報と本の物流とを結びつける整理システムを確立。また、コンピュータ時代に即応した形態、つまり、TRC MARCとしての全国の図書館に向けてサービスを提供している。

### 2. データベース構築にあたっての事前調査、モデルケース

- モデルは特にない。MARC(Machine Readable Catalogue: 機械可読目録)でのサービスが利用され得るという明確な見通しが当初からあった訳ではないが、時代の要請が強かったように思われる。

### 3. データベースサービスとネットワーク

#### ① システムサービスの沿革

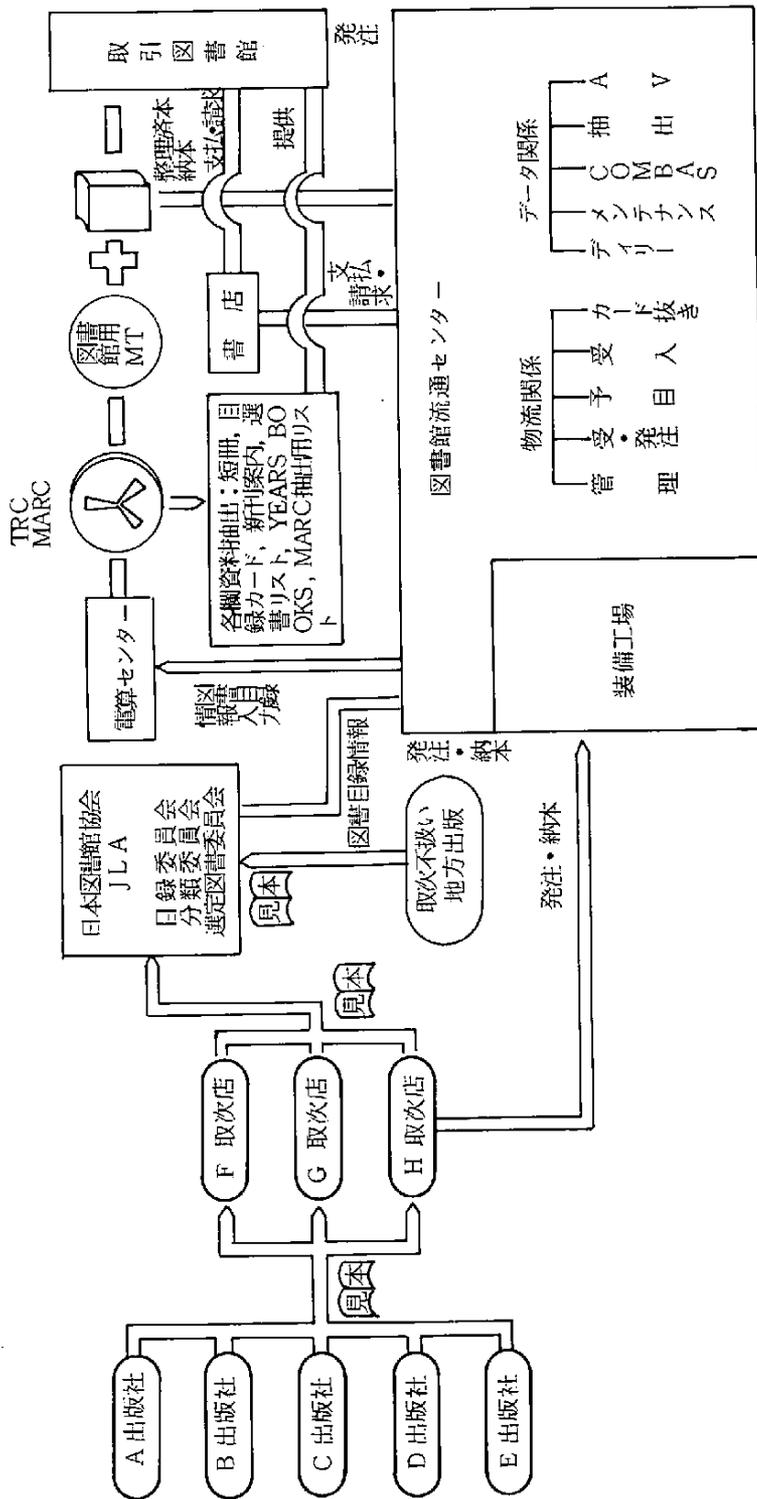
- TRC MARCは、昭和57年2月に発表され、当初の利用図書館は、所沢市立図書館、和歌山市民図書館など数館からスタートした。
- その後、わずか数年の間に、市町村立図書館はもとより、県立図書館、専門図書館、高校・大学図書館へとサービス提供が拡がり、現在200館以上の利用館がある。
- 200館の内訳は、公共図書館が約175館で全体の87.5%と圧倒的に多いが、大学図書館が約20館、高校図書館が4館、専門図書館が2~3館と年々、増加している。

#### ② ネットワーク

- このデータベースのオンラインでのサービスは、東販の図書受発注システムのTONETがあり、これを使ってTRCのデータをオンラインで検索することは可能である。
- TRC独自のオンラインでのネットワークは、現在、行っていない。

#### 4. 情報源と情報収集

- 最近までは、見本として取次店に入った新刊書を、JLAの目録情報センターに送り、そこで書誌事項等を指示してもらい、TRCで電算入力のための記号を付加し、外注先でパンチ入力していた。地方出版物、自費出版物は、JLAに直接送付されるので、届いた時点で書誌事項を指示し、その後、TRCで処理していた。
- 昭和61年からは、見本を取次店等からTRCに収集し、データを作成する流れに変えた。データ作成後、JLAの目録情報センターでデータチェックを行ない、データを完成させる。
- 新刊書は、楽譜、学習参考書も含めて、年間約4万件を処理している。
- 新刊書の他には、各図書館の要望に応じ図書データを作成しており、年間約2～3万件のデータを入力している。
- TRC MARCは、現在約36万件の図書情報を蓄積している。
- TRCシステムの従来フローは次のとおり。



5. データベース機能の概要（レコード内容、検索機能等）

- データベースの内容は、国立国会図書館でサービスを行っている JAPAN MARCの項目を全て網羅しており、さらに、副書名のヨミ、出版社のヨミ、JAL NO. 等、独自の項目を付与している。

## 6. データの構造化と入力情報の分析

- シリーズものは一括記入方式ではなく、個別記入方式をとっている。図書館の運営のことを考えると図書対カードは1対1が望ましい。
- データ作成、データチェック、情報分析（例えば分類付与等）は、司書の資格を有し、図書館業務等の経験を積んだ人達が行なっており、殆どが専任である。
- 書誌事項の記入や判断をするには、センスと閃きと幅広い知識が要求され、現場でも実務の中でこれを鍛えている。
- 図書の主題分野による分担制はとっていない。

## 7. DBMSの選定・開発

- TRCでは、TRC MARCを作成するまでのデータ入力作業を行なっており、DBMSの選定・開発については、今後の課題となる。

## 8. ファイルの設計

- データベースの提供は、図書データを各利用館のフォーマットに変換し、MT（磁気テープ）やFD（フロッピーディスク）の形で利用館に届けている。
- このフォーマットは、TRCで設計した。
- データフォーマットは、固定長（Ⅲ型、Ⅴ型、Ⅵ型）と可変長（JAPAN MARC仕様）とAV用（Ⅶ型）と様々な種類が揃っている。
- 一例として、固定長Ⅴ型のフォーマットを以下に示す。

FILE LAYOUT

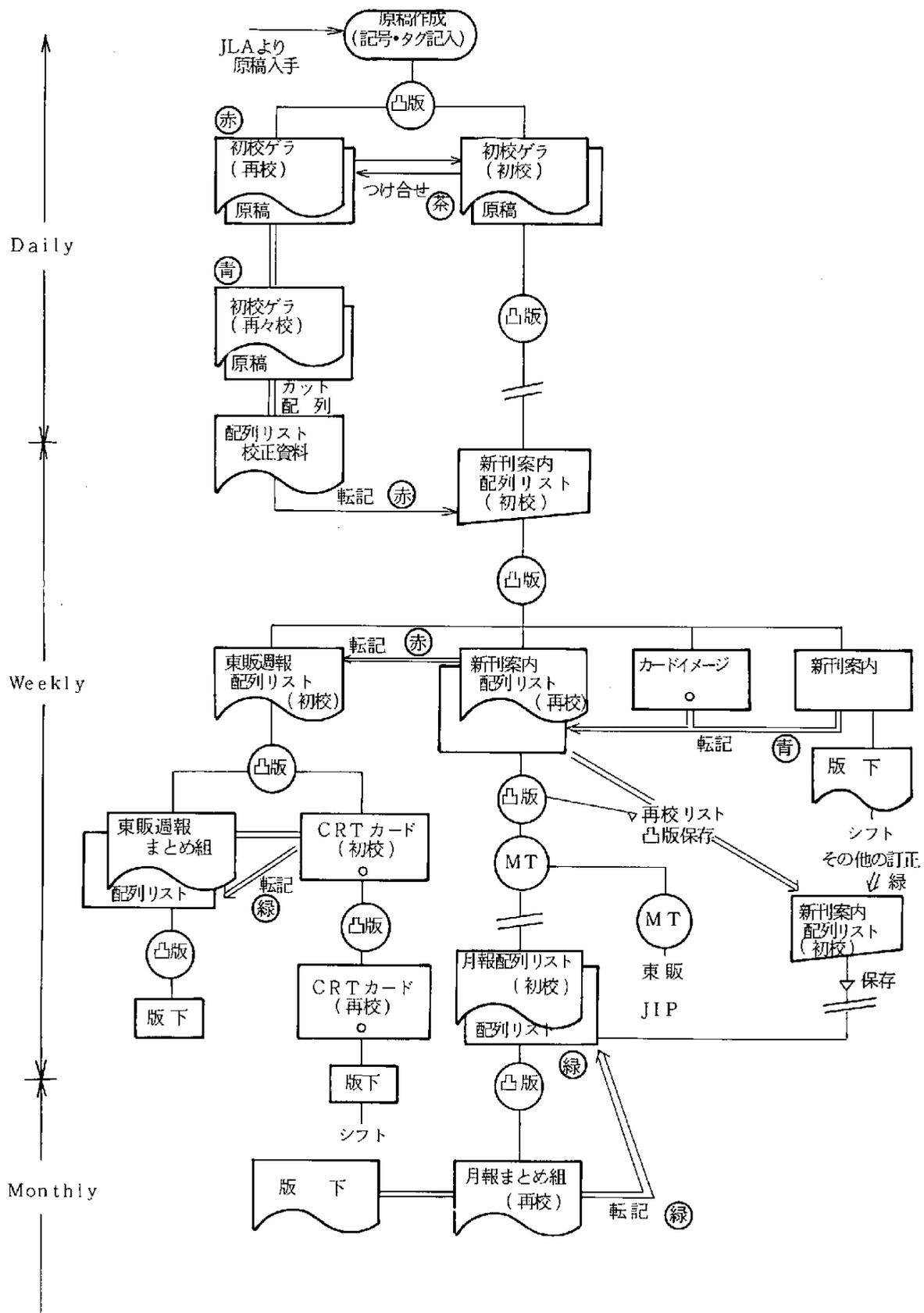
顧客名	図書館流通センター	File-ID	
業務名	TRCMARC	RECORD LENGTH	256 B
File-Form	固定長タイプⅤ	RECORD USAGE	1:DIS 3:MIX
		BLOCKING	12 REC/HT 1 REC/EP LABEL 1:ST 2:

項目名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
項目名	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100



・データ作成からMARCまでの従来フローチャートは次のとおり。



## 10. 標準化ツールの開発と利用

- TRC で扱っている図書は和書が殆どであるが、書誌事項記入の際には、日本目録規則（NCR）新版予備版や日本十進分類法（NDC）8版、7版に準拠している。
- また、TRC独自に、NCRの解釈の仕方の規定や、表紙・標題紙・奥付・裏表紙の4情報源の掛け合わせ方（例えば、タイトルの優先順位の決定の仕方）等の規定を設けている。
- 典拠ファイルは、著者名典拠ファイルをはじめ、叢書名典拠ファイル、出版者典拠ファイル等、TRCが自ら開発し、蓄積しているファイルを所有している。

## 11. 各種マニュアル作成

- コーディング用のマニュアルとしてかなり詳細なものを作成している。

## 12. 運用・開発での体制の確立

- 人材の確保、育成には時間がかかる。
- データの作成は人間が行なうものであるからミスが生じる事もある。いかにミスを少なくするか、ミスが出た場合には、どうフォローするかシステムとして考えておく事が必要である。
- 図書館の電算機導入に伴い、TRC MARCは、速いスピードで利用されはじめた。

## 13. サービス料金体系

- 1件、80円（Localデータ入）が基本だが、図書館側の必要情報や、提供フォーマットによって価格は違う。

## 14. 商品化のチェックリスト

- 図書館（利用者）が何を欲しているかを理解することが大切のように、提供者と利用者の理解の食い違いをなくすことが大切である。
- パンフレット（アピールする媒体）の内容重視と営業活動への展開。
- TRCの場合、スピードを一つの視点としている。本の注文から納品までをできるだけ速く処理し、活用できるようにする。
- ハード、ソフト会社との協力、工場との連携、内部にいる人間のコミュニケーションの円滑化を念頭に置き、業務展開を行なっている。

15. アドバイス, 意見

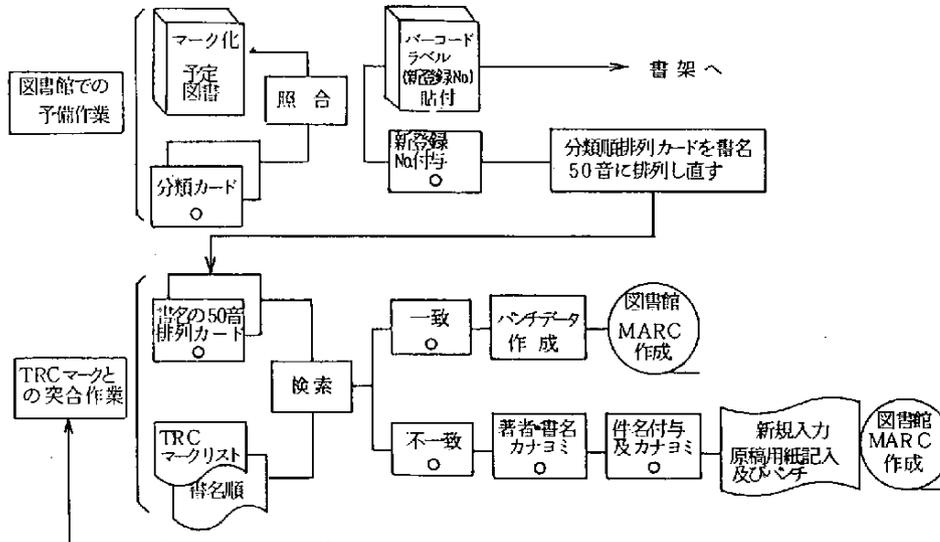
- ・データ作成業務で一人前になるには時間がかかる。

16. TRC MARCを使っている図書館の作業フロー

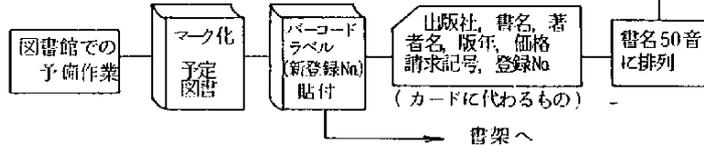
- ・既存図書館の場合

マーク作成フロー概略

<基本カード有りの場合>



<基本カード無しの場合>



蔵書の状況により、いろいろなパターンがあります。準備の段階で詳細な検討をいたします。

収集資料

- 1) TRC 営業の御案内
- 2) いろいろな方向からTRC MARCをとらえてみました。TRC コンパスシリーズ
- 3) コーディングシート
- 4) 目録作成からマークまで
- 5) FILE LAYOUT TRC MARC 固定長タイプV
- 6) (TRC MARCのJ/MARCフォーマット変換)データ要素一覧

## G. (財) 日本情報処理開発協会(データベースクリアリングシステム)

### 0. データベースの概要

- データベースクリアリングシステムとは、データベースサービスに関するクリアリング情報(案内情報)をデータベース化したもので、データベースユーザーが必要とするデータベースの内容、所在、利用方法を含むデータベースに関する情報と、データベースの流通に関わる業者に関する情報を、文字データとイメージデータとで提供しようとしているシステムをいう。

### 1. 機関概要

- 昭和42年12月20日(財)日本情報処理開発センターとして発足。昭和51年3月名称を(財)日本情報処理開発協会と変更、同年4月(財)日本情報開発協会、(財)情報処理研修センターと統合し現在に至る。附属機関として、産業情報化推進センター、情報処理研修センター、情報処理技術者試験センターがある。
- 現在、(財)日本情報処理開発協会はコンピュータシステムを用いた各種の情報処理方式の調査、研究開発、および情報処理技術者研修・試験の実施等を行なっている。

### 2. データベース構築にあたっての事前調査・モデルケース

- 事前作業として、国内外で出版されている既存のディレクトリーの調査と、国内外のデータベースプロデューサー、データベースディストリビュータ、データベースブローカーよりデータベースやデータベースサービスに関する資料の入手を行なう。
- 国内外で既に稼働しているクリアリングシステムの調査、解析を行なう。
- (財)日本情報処理開発協会で、商用データベース利用についてアンケート調査を実施し、利用者からクリアリングサービスの必要性に関する回答を得ている。
- データベースの所在について通商産業省や関連機関への問い合わせが増えてきている。それに応えるとともに、多くの利用者にも使用してもらうために Machine Readable の形で利用可能にするため、クリアリングシステムの開発を開始した。

### 3. データベースシステムとネットワーク

- ・クリアリングシステムの開発は昭和59年度より3カ年計画で始まっており、現在、進行中である。昭和59年度から昭和61年度までの開発計画は下記のようなになる。

			59年度	60年度	61年度
ク リ ア リ ン グ シ ス テ ム 開 発	ク リ ア リ ン グ の 作 成	文 字	58年度台帳中心 データベース 約600件 業 者 約300件	—————	データベース 1200件 業 者 600件
		イメー	58年度台帳 約100ページ	—————	( 約 1800 ページ )
	会 話 形 式	コマンド方式	コマンド方式 メニュー方式 } 併用	—————	
	システムの実現形態	文字データ系とイメージ データ系の独立形	文字データ系とイメージ データ系の統合形	—————	
	アクセス形態	チャンネル直結型	回線経由型	—————	
同義語辞書の作成		—————	一部構築	—————	

(出典：データベースクリアリングシステムの開発について)

### 4. 情報源と情報収集方法

- ・情報源は、通商産業省から出版されている「データベース台帳総覧」と、国外の各種のディレクトリ、国内外データベースのプロデューサー、ディストリビュータ等からのパンフレット等。
- ・情報収集方法は、国内外データベースのプロデューサー、ディストリビュータ、ベンダーへは直接連絡してパンフレット等を送ってもらう。

## 5. データベース機能の概要（レコード内容、検索機能等）

- 文字系データベースの検索例
- データベースクリアリングシステムには、文字系データベースとイメージ系データベースがあり、共にデータベースに関するデータベース（以下「データベースDB」と称す）と、業者に関するデータベース（以下「業者DB」と称す）を構築している。文字系データベースのサブシステム（コマンド方式）を用いた検索例を以下に示す。

（検索テーマ）

「データベースについての情報を含むデータベースについて知りたい。できれば、コンピュータ分野のものがよい。該当するデータベースがあれば、それをサービスする業者についても知りたい。」

\* 下線部は利用者の入力を表す。

```
RS> SEL DBDB .....データベースDBを選択
```

データベース「DBDB」の選択（SEL）

```
RS> SEA KWD EB データベース .....キーワードを指定
```

キーワード「データベース」の指定（SEA）

◆ ◆ 集 合 名 ◆ ◆	◆ ◆ 件 数 ◆ ◆	◆ ◆ 履 歴 ◆ ◆	◆ ◆ 検 索 式 ◆ ◆
Y1	3	SEA KWD EQ	F*~277*-2

図3-9で入力した検索式  
 ヒット件数  
 右の検索式による検索の結果として生成された  
 集合の名称

3 件 (データベース) 見つかりました。 ( Y1 )

RS> SEA KWD EQ コンピュータ

### サーチ結果とキーワードの指定 (SEA)

◆ ◆ 集 合 名 ◆ ◆	◆ ◆ 件 数 ◆ ◆	◆ ◆ 履 歴 ◆ ◆	◆ ◆ 検 索 式 ◆ ◆
Y1	3	SEA KWD EQ	F*~277*-2
Y2	1	SEA KWD EQ	コンピユータ

1 件 (データベース) 見つかりました。 ( Y2 )

RS> AND Y1 .....集合Y1とY2の論理積指示

### サーチ結果と論理積指示 (AND)

◆ ◆ 集 合 名 ◆ ◆	◆ ◆ 件 数 ◆ ◆	◆ ◆ 履 歴 ◆ ◆	◆ ◆ 検 索 式 ◆ ◆
Y1	3	SEA KWD EQ	F*~277*-2
Y2	1	SEA KWD EQ	コンピユータ

条件を満たすレコードがありません。

RS> OUT SEI(Y1) .....集合Y1に属するデータベースの内容表示指示

### 論理積の結果と内容表示指示 (OUT)

3 件 (データベース)

#1 データベース登録番号  
DF000121

データベース名 正式名称  
DIALOG Database Index

概要 DIALOGデータベースのデータベース索引。与えられた課題に適合するデータベースを選択することができる。適合したデータベースは、検索時の出現数が多い順に出力される。

DB項目名 データベース名  
検索の出発点

分野 社会科学 (索引ガイド/補助ファイル)

キーワード データベース

データのタイプ 文書 (その他)

収録情報の地域 世界

データベースの言語 ENG

カレント収録 カレント

OUT>

内容表示 (第1ページ)

3 件 (データベース)

更新周期 年  
更新回数 7  
更新回数 4

公開/非公開 公開

プロデューサ登録番号  
VF000200

ベンダ登録番号  
VF000200

S・システム登録番号  
SF0018

代理店登録番号  
VJ000051

OUT>

内容表示 (第2ページ)

3 件 (データベース)

#2 データベース登録番号  
DF000124

データベース名 正式名称  
DIALOG DIALOG提供データベースの索引。与えられた課題に適合するデータベースの選択ができる。

DB項目名 データベース名  
キーワード

分野 社会科学・人文科学 (図書学・情報科学)

キーワード データベース

データのタイプ 文書 (その他)

収録情報の地域 世界

データベースの言語 ENG

カレント収録 カレント

公開/非公開

OUT>

内容表示 (第3ページ)

3 件 (データベース)

```

公開
プロデューサ登録番号
VF0000081
ベンダ登録番号
VF0000081
S・システム登録番号
SF00005
データベース名正式名称
DIALOG INDEX
代理店登録番号
VJ0000015
VJ0000074

```

OUT>

内容表示 (第4ページ)

3 件 (データベース)

```

#3 データベース登録番号
0F000125
データベース名正式名称
DIALOG Publications
概要
DIALOG社の出版物に関する書誌情報を収録。
DB項目名
標題
ファイルナンバー
出版年
価格
分野
社会科学・人文科学 (図書館学・情報科学)
分野
0302
キーワード
データベース
ニューズマニュアル
データのタイプ
文書 (書誌)
文書 (抄録)
データベースの言語
ENG
カレント収録
カレント

```

OUT>

内容表示 (第5ページ)

3 件 (データベース)

```

収録件数 150
更新周期 年
更新回数 7
更新回数 4
プロデューサ登録番号
VF0000081
ベンダ登録番号
VF0000081
S・システム登録番号
SF00005
データベース名正式名称
DIALOG Publications
代理店登録番号
VJ0000015
VJ0000074

```

OUT> HISTORY .....会話履歴表示指示

内容表示 (第6ページ)

会話履歴は次の通りです。

```
*SEL DDDD
*SEA KWD EQ 1-99-99
3件 (データベース) 見つかりました。( *1 )
*SEA KWD EQ 001-99
1件 (データベース) 見つかりました。( *2 )
*AND *1
条件を調べるレコードがありません。
*OUT SET(*1)
```

RS> SEL VENDRDD .....業者DBを選択

会話履歴表示と業者DBの選択

RS> SEA VNO EQ VF000081 .....ベンダ登録番号として、VF000081を指示

ベンダ登録番号によるサーチ

```
◆ ◆ 検索履歴 ◆ ◆
舞台名 件数 検索式
*1      1  SEA VNO EQ 0F000081
```

1件 (実行) 見つかりました。( \*1 )  
RS> OUT .....内容表示指示

サーチ結果と内容表示指示

```

                                                    ( 件 (業者) )
#1 業者登録番号          VF000081
業者正式名称          D i a l o g I n f o r m a t i o n S e r v i c e , I n c .
連絡先                3 4 6 0 H i l l v l o w A v e . P a l o A l t o , C A
                      9 4 3 0 4 ; T E L : 4 1 5 / 8 5 8 - 3 7 8 5 ; 8 0 0 / 2 2 7 - 1
                      9 2 7 ; T e l e x : 3 3 4 4 9 9 D I A L O G
所属国コード          HUS
業者種別              ベンダ (ディストリビュータ)
業者種別              2
サービスDB登録番号   D F 0 0 0 1 2 4
                      D F 0 0 0 1 2 5

OUT> END .....検索サブシステム終了指示

```

内容表示と検索サブシステム終了指示

```

DOCS> END .....D B C S 終了指示
DBCS 終了.
READY
LOGOFF .....T S S セッション終了指示

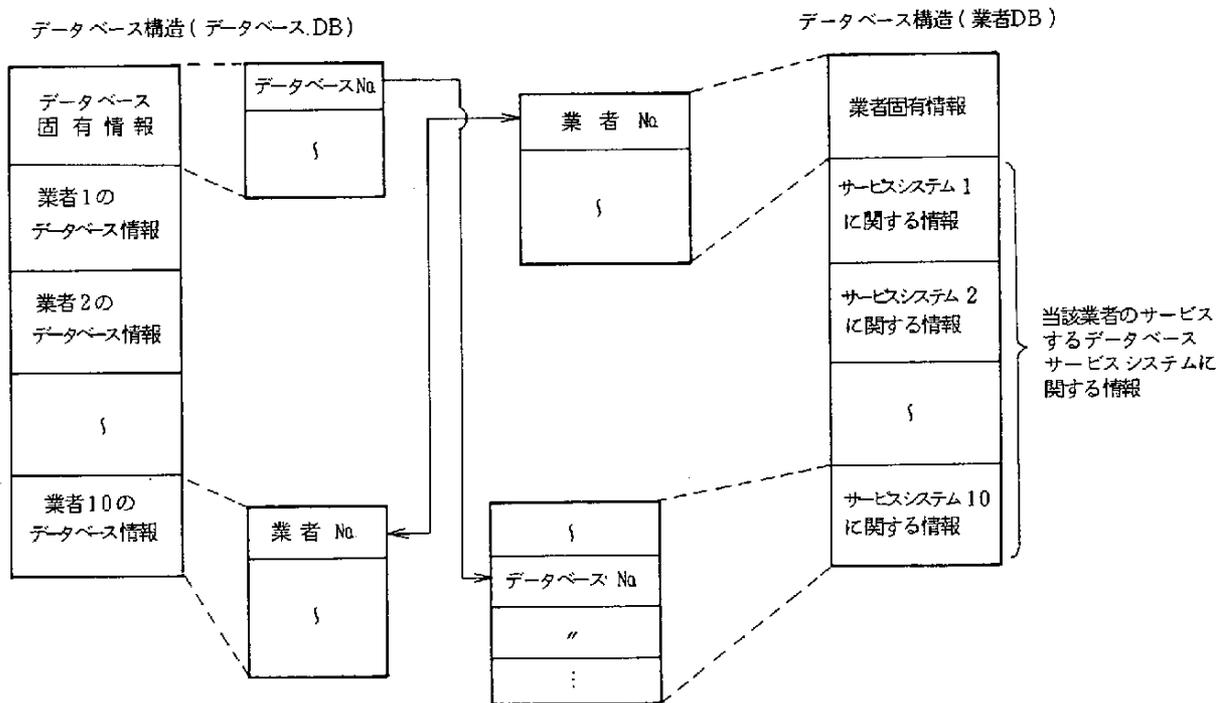
```

D B C S 終了と T S S セッション終了

( 出典 : 高度データベースシステムの開発に関するフィジビリティ報告書 ( 3 )  
 — オンラインデータベースクリアリングシステム — P. 204 - 209 )

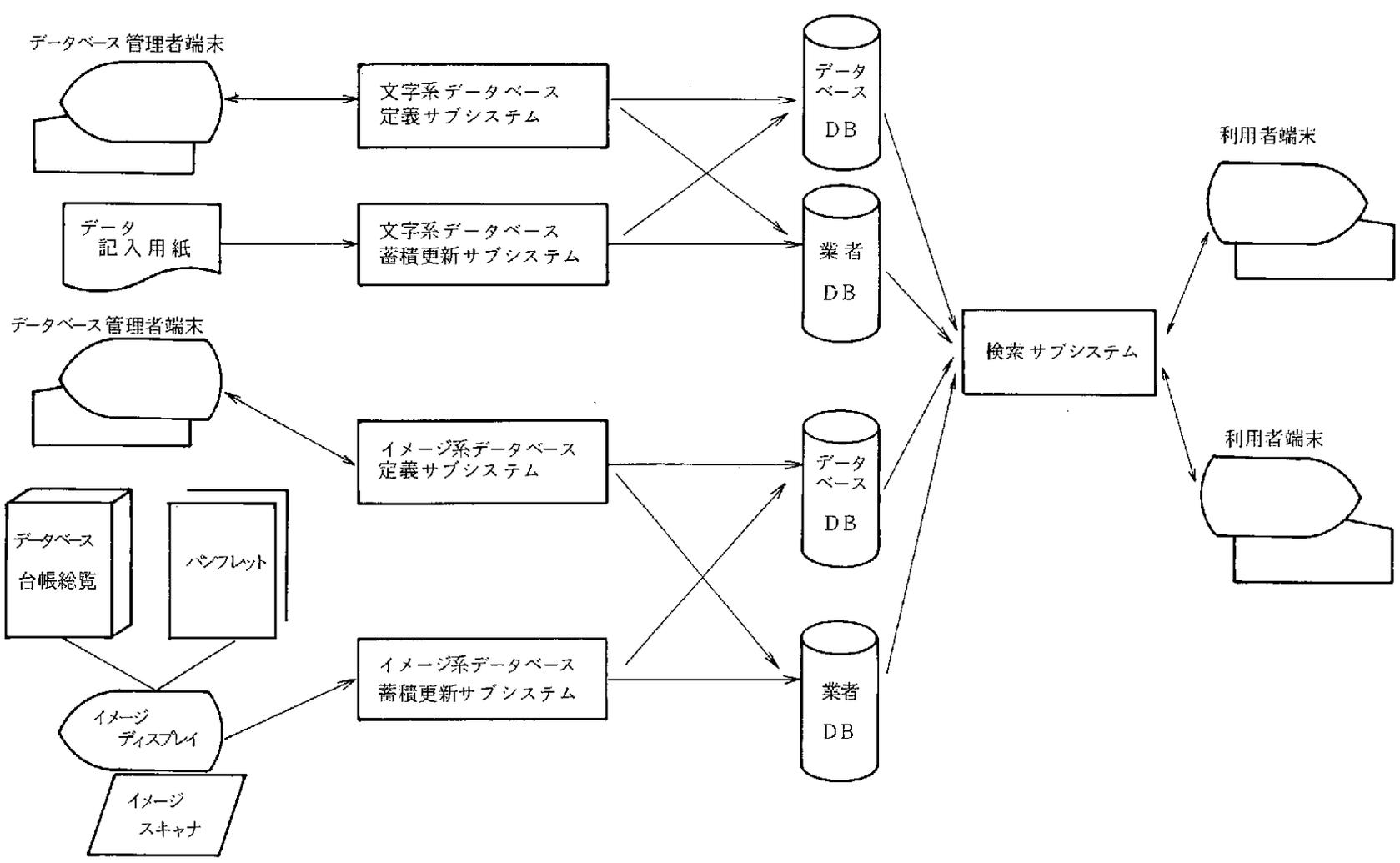
## 6. データの構造化と入力情報の分析

- データベースクリアリングシステムにおける文字系データベースは、データベース DB と業者 DB の2つのデータベースを持っており、それぞれのデータベース構造は以下のとおり。



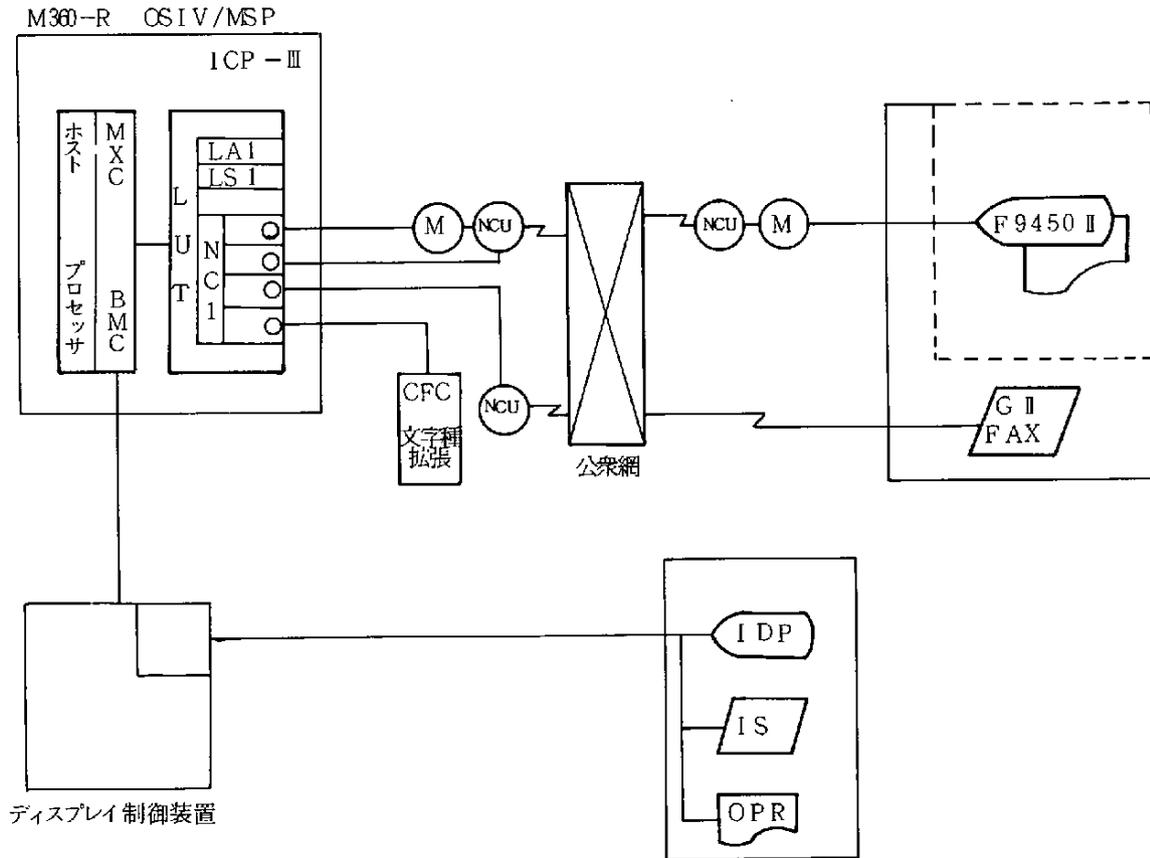
( 出典：クリアリングDBの構造 )

- データベース DB の利用目的は、データベースに関する案内情報の蓄積と検索である。そのため、データベース DB は、データベースの名称、分野、内容等データベースに関する種々の情報を収録している。
- 業者 DB の利用目的は、データベースサービス業者に関する案内情報の蓄積と検索である。そのため業者 DB は、業者名、連絡先、サービスシステム名等データベースサービス業者に関する種々の情報を格納している。
- データベース DB と業者 DB 作成にあたり、「データベース台帳総覧」や各種ディレクトリを情報源として用いている。情報の種類別様式に分けられたデータ記入用紙（10種）に必要事項を記入する。初期入力は大量となるため、パンチ入力で作成。
- 「データベース台帳総覧」の原稿（申告書）記入は、データベースサービス業者に依頼している。
- 文字系データベースの蓄積・更新サブシステムで用いるデータ記入用紙は、内部の専門業者に委託した。
- 蓄積されたデータは、蓄積・更新サブシステムにより、データのチェックを行った後、データベース DB と業者 DB として構築される。
- オンラインでのデータの蓄積・更新も可能である。
- 本システムは、データベース運用に関する3つの機能を有しており、それらは(1)データベース定義機能、(2)データベース蓄積・更新機能、(3)データ検索機能である。
  - (1)のデータベース定義機能は、データベースを構成するファイルの物理的領域の確保、およびデータベースの論理構造の定義を行うものである。
  - (2)のデータ蓄積・更新機能は、データベースの案内情報の蓄積、および既に蓄積されている情報の更新を行うものである。
  - (3)のデータ検索機能は、案内情報の検索と検索結果の表示、検索時の利用者サポートを行うものである。
- 本システムは、データベースの案内情報である文字系データベースのほかに、データベース台帳総覧やデータベースサービス機関のパンフレット類をデジタル画像として蓄積したイメージ系データベースを有する。本システムの提供する3つの機能を有したサブシステムの関連図は次のようになる。



## 7. DBMSの選定・開発

- DBMSの選定に際しては、文字系データとイメージ系データの連動が可能で、財団設置の汎用コンピュータ（2機種）で動作可能なものを対象とした。最終的には、富士通のFAIRS-Iを核として、周辺の蓄積・更新および検索等のユーザープログラムを開発している。
- データベースクリアリングシステムの開発は、日本情報処理開発協会の内部で行っている。
- ハードウェア構成は以下のとおり。



ICP-III : 通信制御装置 (関連機器 LUT, LA1, LS1, NC1)

- M : モデム
- NCU : 網制御装置
- CFC : コンピュータ・ファクシミリ制御装置
- IDP : イメージディスプレイ装置
- IS : イメージスキャナ装置
- OPR : オフィスプリンタ装置

( 出典 : データベースクリアリングシステムの開発について )

## 8. ファイルの設計

- ・クリアリングシステムは、文字系データベースとイメージ系データベースから構成されている。文字系データベースにはデータベース DB と業者 DB とがあるが、蓄積・更新サブシステムで使用するファイルには、DB 展開ファイル、DB 編集済ファイル、業者ファイル、業者展開ファイル、業者編集済ファイルがある。
- ・また、イメージ系データベースには文書イメージファイル、文書コードファイルと、文書定義ファイルがあり、総称して文書ファイルと呼んでいる。さらに、インデックス検索を効率良くさせるために、インデックスファイルを用意している。
- ・これらのファイル設計では、文字系データベースとイメージ系データベースのファイルの連動が可能となっている。

## 9. 入出力の設計

- ・データベースクリアリングシステムで使用されている 2 つのデータベースのうち、文字系データベースは、「データベース台帳総覧」や国外のディレクトリ数種を参考にして記入された、蓄積・更新サブシステム用のデータ記入用紙（10 種）の項目を基に構成されている。
- ・文字系データベースのデータベース DB のうち、「キーワード」については、あらかじめ決めた統制語を用いてデータ作成を行っていたため、検索時においても、原則として統制語で検索する必要がある。
- ・文字系データベースでは、データが大量の場合、入力はバッチ処理方式をとっているが、オンラインでの蓄積・更新も可能である。
- ・一方、イメージ系データベースは、「データベース台帳総覧」とデータベースのカタログを基に、イメージ・スキャナで各ページ（データベース台帳総覧の場合は 1 ページ単位、パンフレットの場合は複数ページになる場合もある）を読み取り、必要部分の切り出し作業を行ない、イメージファイルに蓄積される。
- ・コード情報は、既に蓄積されたイメージ情報と文字情報を結びつけ、文書（本システムでのファイリングの単位をいう）を構成するもので、対となるイメージ情報をユニークに定める役割を有しており、文書を検索するためのアクセスポイント（索引）として、次にあげる 5 つの項目が指定されている。(1) データベース名、(2) サービス企業名、(3) データの分野、(4) データの種類、(5) データベース台帳総覧ページ、である。これらは全てインバーテッド・ファイルとして構築されている。

## 10. 標準化ツールの開発と利用

- クリアリングデータベースのデータベース DB 中の「概要」項目から抽出した約 6,000 語をもとに関連語で 73 グループに分けた同義語ファイルがある。
- 蓄積・更新サブシステムで使用するデータ記入用紙は、様式 1～10 まであり、データベース固有の情報、ディストリビュータ毎に異なる情報、業者固有の情報、サービスシステム毎に異なる情報という様に、情報の種類別の様式が決まっている。
- 「データベース台帳総覧」の記入用紙（申告書）は、通商産業省で定められたものである。

## 11. 各種マニュアルの作成

- 現在あるマニュアルは、検索マニュアル、蓄積マニュアル、検索操作説明書（メニュー方式、タッチ方式）があり、目次は以下のとおり（昭和 60 年度作成予定を含む）。

### 1) 検索マニュアル・目次

1. 概要	1
2. 検索処理の流れ	3
2.1. 単一データベースの検索	4
2.2. データベース DB と業者 DB の関連検索	9
3. コマンド文法	13
3.1. コマンドの構成	13
3.2. 記述規則	14
3.3. コマンド一覧	16
3.4. 用語定義	18
3.5. コマンド各論	22
3.5.1. DBCS の起動と終了	22
3.5.1.1. DBCS 起動コマンド	22
3.5.1.2. DBCS 終了コマンド	23
3.5.2. 検索サブシステムの起動終了	24
3.5.2.1. 検索サブシステム起動コマンド	24
3.5.2.2. 検索サブシステム終了コマンド	32
3.5.3. 検索基本コマンド	33

3.5.3.1.	データベース選択コマンド	33
3.5.3.2.	検索コマンド	35
3.5.3.3.	検索結果出力コマンド	56
3.5.4.	検索補助コマンド	62
3.5.4.1.	管理情報表示コマンド	62
3.5.4.2.	索引語表示コマンド	69
3.5.4.3.	保存コマンド	75
3.5.4.4.	消去コマンド	77
3.5.4.5.	分類コマンド	79
3.5.4.6.	質問終了コマンド	81
3.5.4.7.	会話経過表示コマンド	82
3.5.5.	会話補助コマンド	84
3.5.5.1.	利用説明コマンド	84
3.5.5.2.	会話進行補助コマンド	86
3.5.5.3.	アテンション	87
3.5.6.	ディスプレイ端末補助コマンド	88
4.	ディスプレイ端末からの検索	89
4.1.	画面構成	90
4.2.	PFキー	91
4.3.	検索の流れ	92
4.3.1.	フルスクリーンモードの開始	93
4.3.2.	検索	94
4.3.3.	結果表示	95
4.4.	カナ(ローマ字)漢字変換入力	99
4.4.1.	入力規約	100
4.4.2.	変換規則	102
5.	検索例	104
6.	メッセージ	115
付録	1. コマンドの入力形式一覧	( 1 )
	2. 項目名一覧	( 5 )
	3. コード表, キーワードリスト	(14)

## 2) 蓄積マニュアル・目次

1. データ記入要領 .....	1
1.1. データ記入用紙の構成 .....	2
1.1.1. 各様式の概要 .....	2
1.1.2. 各様式の情報源 .....	3
1.2. データ記入用紙共通項目の説明 .....	3
1.2.1. 用語・用字法 .....	3
1.2.2. 処理区分 .....	4
1.2.3. 各様式共通項目 .....	5
1.3. データ記入用紙各様式の説明 .....	7
1.3.1. 様式 1 .....	8
1.3.2. 様式 2 .....	8
1.3.3. 様式 3 .....	9
1.3.4. 様式 4 .....	11
1.3.5. 様式 5 .....	13
1.3.6. 様式 6 .....	16
1.3.7. 様式 7 .....	17
1.3.8. 様式 8 .....	20
1.3.9. 様式 9 .....	21
1.3.10. 様式 10 .....	22
1.4. データ記入例 .....	23
2. ジョブ一覧 .....	62
3. オペレーション指示書 .....	64
4. メッセージ .....	67
4.1. データベース DB 更新時の出力メッセージ .....	68
4.2. 業者 DB 更新時の出力メッセージ .....	75

## 3) 検索操作説明書（メニュー方式）・目次

1. 概要 .....	1
-------------	---

2. 基本的機能及び操作 .....	2
2.1. 基本的検索機能 .....	2
2.2. 検索の流れ .....	2
2.3. 検索例 .....	4
3. 画面の種類 .....	25
4. PFキー機能役割り当てと画面還移 .....	28
4.1. PFキー機能割り当て .....	28
4.2. 画面還移 .....	29
5. 画面説明 .....	33
5.1. 画面構成 .....	33
5.2. 画面各論 .....	35

#### 4) 検索操作説明書(イメージ系データベース)・目次

1. イメージ系データベース .....	74
1.1. 文書ファイルの登録・定義 .....	75
1.2. イメージ情報の蓄積 .....	78
2. 文字系データベースとイメージ系データベースの結合 .....	83
2.1. 文字系データベースとイメージ系データベースとの結合 .....	84
2.2. 文字系データベースとイメージ系データベースとの統合検索 .....	86
2.2.1. 内容表示画面からイメージ情報表示モードの入口までの操作 .....	86
2.2.2. イメージ情報参照モードでの操作 .....	90
2.2.3. 使用例 .....	91

#### 5) 検索操作説明書(タッチ方式)・目次

1. 概要 .....	97
1.1. 構成 .....	98
1.2. 操作 .....	99
1.3. 索引の構造 .....	101
1.4. 索引の種類 .....	103

1.5	基本操作キーの機能	104
2.	検索例	106
2.1.	検索手順	107
2.2.	画面例	109
付録	1. データベース台帳総覧索引(一部)	115
	2. タッチ方式表示画面	119

## 12. 運用・開発面での体制の確立

- ・クリアリングデータベース作成のためのデータ記入とパンチは外部に委託している。
- ・システムは内部職員だけで開発している。
- ・(財)データベース振興センターのクリアリング委員会等でシステム運用のための検討が行なわれている。

## 収集資料

- 1) クリアリング委員会資料より(資3-1)
- 2) クリアリングDBの構造
- 3) 高度データベースシステムの開発に関するフィジビリティスタディ報告書(3)-オンラインデータベースクリアリングシステム-
- 4) コード化項目一覧表
- 5) データベースDBの項目一覧
- 6) 業者DBの項目一覧
- 7) システム構成図
- 8) 申告書提出について(通商産業省機械情報産業局情報処理システム開発課)
- 9) データベースクリアリングシステムの開発について
- 10) オンラインデータベースクリアリングシステムの開発及び運用上の諸問題

## H. (財)日本造船振興財団

### 0. データベースの概要

- ・MATRAXとは、(財)日本造船振興財団造船資料センターが蓄積している「船舶・海洋工学技術文献速報」のデータベース名で、現在、磁気テープで保管している。

### 1. 会社概要

#### ・目的及び事業

造船業及び造船関連工業を営む企業の近代化・合理化を促進するとともに、造船関係事業に関連する公益法人等の業務の推進に協力し、もってわが国造船関係事業の振興に寄与することを目的とする。

- (1) 造船関係事業に関連する図書その他の資料を調査し、収集し、及び提供すること。
- (2) 造船関係事業に関連する図書その他の資料を刊行し、又は製作すること。
- (3) 造船関係事業に関連する海洋油濁防止及び海洋環境技術に関する試験・研究を行なうこと。
- (4) 造船関係事業を営む企業に対し経営上及び技術上の指導を行なうこと。
- (5) 造船関係事業を営む企業の近代化・合理化のための設備のリース事業を行なう者に対し資金の融通を行なうこと。
- (6) 造船関係事業に関連する公益法人等に対し、その業務の実施に必要な資金の融通を行なうこと。
- (7) 造船技術及びこれに関連する技術の研究開発を行なう者に対し、資金の融通又は補助を行なうこと。
- (8) 造船関係事業に関連する公益法人等の業務の推進に協力するため、会議室、事務室その他の施設を提供すること。
- (9) その他この財団の目的を達成するために必要な事業を行なうこと。

### 2. データベース構築にあたっての事前調査・モデルケース

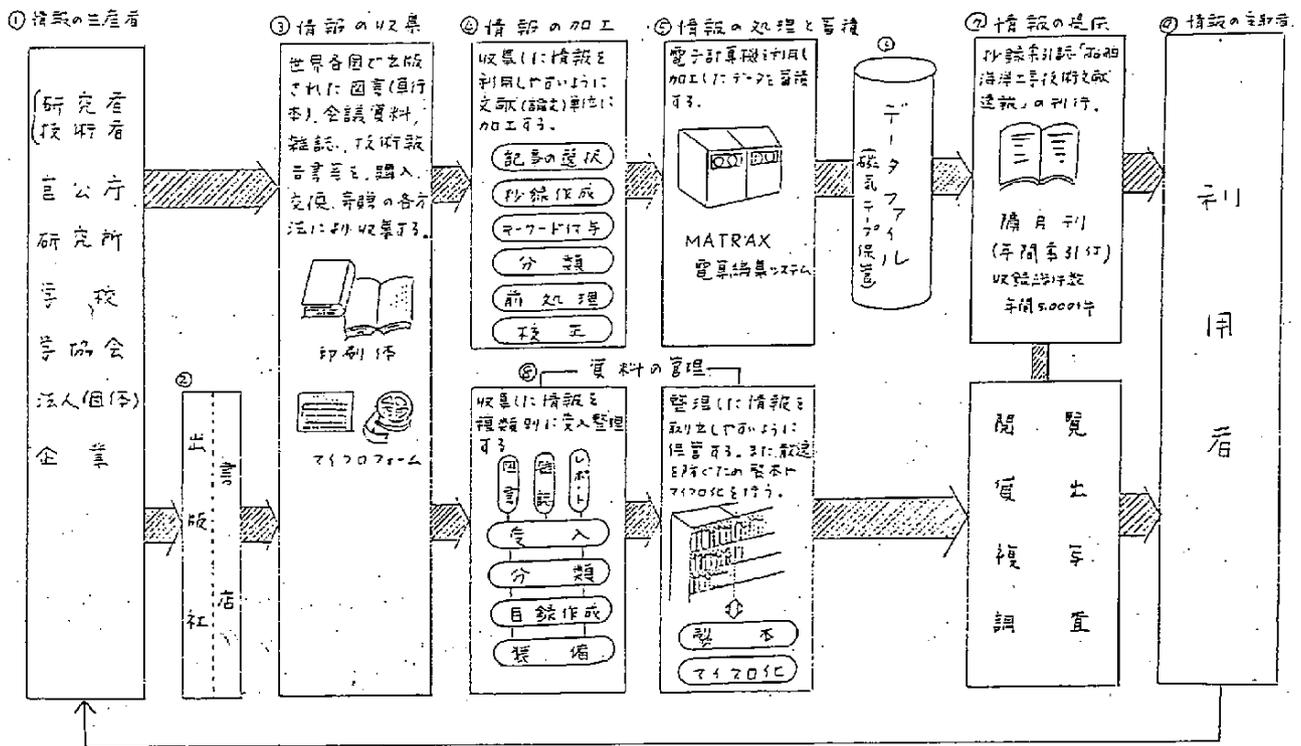
- ・日本科学技術情報センターのJICST科学技術文献ファイルや、米国BRAネットワーク、欧州BSRAネットワークの内容については、事前調査を行なっている。

### 3. データベースサービスとネットワーク

- コンピュータ化の過程
- 昭和39年より昭和56年度まで、抄録索引誌「船舶・海洋工学技術文献速報」(MATRAX)を手作業で作成、刊行していた。
- 昭和56年8月、データベース化計画のための調査を開始、同年10月、造船資料センター運営委員会において57年度MATRAXのデータベース化について承認を得る。
- 昭和57年2月、データベース化実施計画を策定、同年4月、造船資料センター運営委員会において57年度MATRAXデータベース化実施計画および予算について了承を得る。実施計画に基づき、システム開発を開始。5月には、データ入力開始、シソーラス編集のため「船舶・海洋工学に関する文献検索用語集編集委員会」を設置し、6月にMATRAX第17巻1号が完成(以後2カ月毎に納本される)。電算処理に関しては船舶振興会より80%の補助金を受ける。
- 昭和58年3月、MATRAX第17巻6号が完成。著者名索引、収録会議資料一覧、NTISレポート索引を付す。同年12月、「船舶・海洋工学技術文献シソーラス(英語版)」の編集終了。この年より、自主事業となる。
- 昭和59年3月、MATRAX第18巻の年間索引(著者名、キーワード、収録会議資料一覧)を刊行。「船舶・海洋工学技術文献シソーラス(英語版)」発行。59年には日本語版も入っているがprint outしていない。対訳も調整済。

### 4. 情報源と情報収集

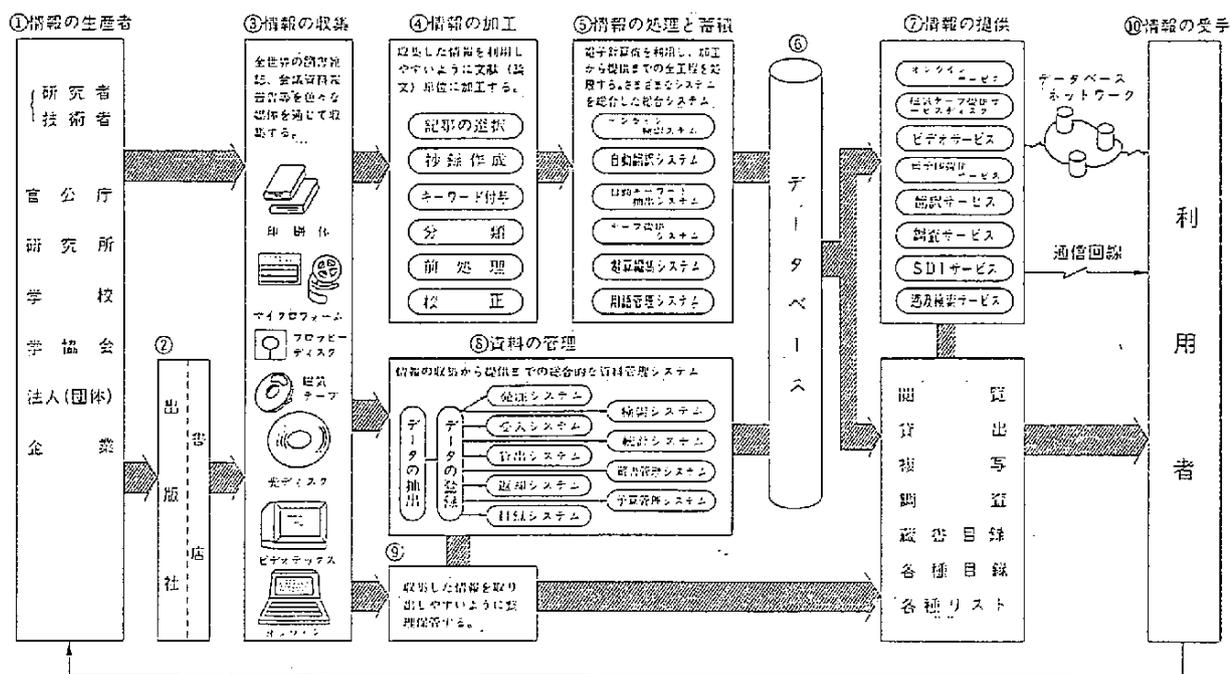
- 情報源は、大学、官庁、民間の各研究所等からの学術書や研究レポート、会議資料が多く、分野は国内外を問わず造船一般、船舶設計、船体構造、船舶流体力学、船舶用機械、船体艤装、造船工作および造船所施設、航海、海洋構造物の設計・構造・流体力学、海洋・港湾工学、海洋環境、海洋作業を対象としている。
- 収集資料は日本をはじめ、中国、韓国等アジア地域に重点を置いている。
- 一次資料は、購入、交換、寄贈の方法で入手している。
- 情報の流れは現在次のようになっている。



( 出典 : 造船資料センターの情報の流れ )

## 5. データベース機能の概要（データ項目、検索機能等）

- データの中身は、雑誌、会議資料、レポート、図書等の書誌事項（翻訳論文名、原論文名、著者名、出典、文献番号、記事区分、資料種類）と抄録から構成されている。
- 現在「船舶・海洋工学技術文献速報」をデータベース化しており、磁気テープに蓄積しているところである。
- コンピュータ利用による将来像は下の図のように考えている。

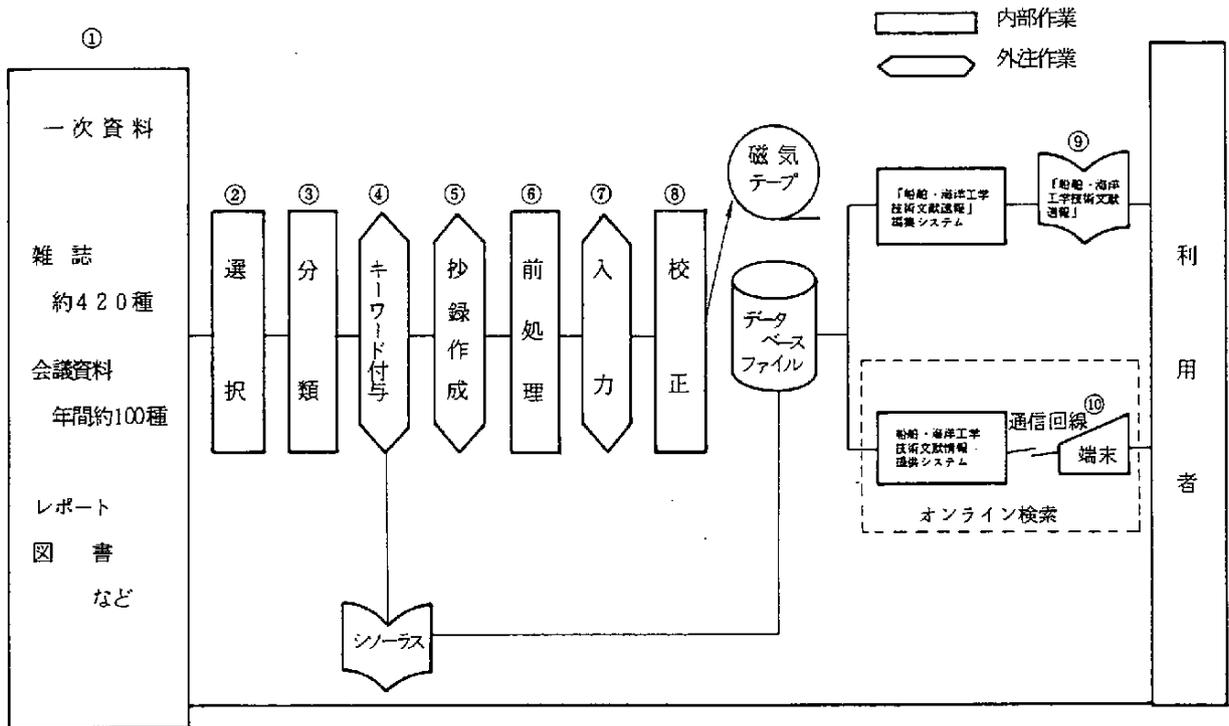


（出典：造船資料センターの情報の流れ（電算機利用による将来像））

## 6. データの構造化と入力情報の分析

- 資料の選択から「船舶・海洋工学技術文献速報」が出来上がるまでの作業過程は次のようになっている。
- 一次資料の選択と分類は専門知識を有する者が行っており、Indexingや校正等の前処理は職員が3～4人で作業している。抄録は、企業の研究所、大学の教官、また退職された方で、船舶・海洋工学の分野に通じている主題専門家10～12人に外注している。
- 一次資料の選択の際には、日本科学技術情報センターで収録している科学・技術関係の文献となるべく重複しないようにしている。

・収録件数は現在 51,000 件で、年間 5,000 件追加している。



( 出典：海洋工学技術文献速報 (MATRAX) 業務の流れ )

## 7. DBMSの選定・開発

- ・ホストコンピュータを独自に設置するか、企業のコンピュータセンターとオンラインで接続するかは検討中である。
- ・加工したデータのパンチ入力と磁気テープへの蓄積は外注している。

## 8. ファイルの設計

- ・外注先で設計している。

## 9. 入出力の設計

- ・「5. データベース機能の概要」のコンピュータ利用による将来図参照。

## 10. 標準化ツールの開発と利用

- ・分類、言語、記事区分、キーワード(統制)、キーワード(フリー)、分類・キーワード備考、前処理備考の項目の入ったデータシートがあり、前処理の際に使用している。

## 11. 各種マニュアルの作成

- ・内部用に「船舶・海洋工学技術文献速報」前処理マニュアル改訂版と、船舶・海洋工学技術文献情報システム実施計画マニュアルがあり、前者は、前処理作業の内容やデータ項目の一覧、各項目の解説と記入要領等から成る。後者は、SIST(科学技術情報流通技術基準)を参考に、データベースシステムの概要や構成についてまとめられたものである。

## 12. 運用・開発面での体制の確立

- ・出発点は図書館にあると考えている。
- ・図書館所蔵の文献情報を中心にデータベース化する場合には、図書館の知識、ドキュメンテーションの知識を有する人がいた方が望ましい。
- ・システム開発、運用面に対する体制確立は考慮中。

## 13. 商品化のチェックリスト

- ・一次資料蓄積の必要性。
- ・船舶・海洋工学関係の資料の特徴をふまえて、データの蓄積、サービスを行なうようにする。

- ・利用者の確保。
- ・料金の設定方法。
- ・既存データベース（日本科学技術情報センターの JICST 科学技術文献ファイル）となるべく重複しないように心掛けながら当該データベースの特質をだす。

#### 14. 国・政府への要望

- ・特定主題にわたる専門分野の文献情報のデータベース化の整備（開発）に助成金をだして欲しい。

#### 収集資料

- 1) 船舶・海洋工学技術文献速報（MATRAX）業務の流れ図。
- 2) 造船資料センターの情報の流れ。
- 3) 「船舶・海洋工学技術文献情報データベース-MATRAX-」, 情報管理 vol. 27, No. 9, Dec. 1984, p. 817-p. 830.
- 4) 船舶・海洋工学技術文献速報, vol. 20, No.1, May 1985.
- 5) 造船と情報, No. 18, Jan. 1986.
- 6) 造船技術情報利用実態調査報告書, 1873.
- 7) 船舶・海洋工学技術文献情報データベース抄録索引誌「船舶・海洋工学技術文献速報」の電算編集一, 1984,9.
- 8) 船舶・海洋工学技術文献情報システム 実施計画マニュアル, 1982,2.
- 9) 「船舶・海洋工学技術文献速報」前処理マニュアル 改訂版, 1983,4.
- 10) 昭和 59 年度 海難防止等データ利用システム開発調査研究報告書, 1985,3.
- 11) 船舶・海洋工学技術文献シソーラス, 1984,3.

# 1. 米国国立医学図書館 (MEDLINE)

## 0. データベースの概要

- MEDLINE (Medical Literature Analysis and Retrieval System (MEDLARS) on-Line) は、米国国立医学図書館 (NLM: National Library of Medicine) が作成、提供する書誌データベースである。対象分野は、基礎および臨床医学、歯科学、獣医学、生体工学、環境衛生医学、医療工学、看護学、法医学、生物学、生化学、薬学、医学史など生物医学 (Biomedicine) 全般に及ぶ。収録記事は、原著論文、総説、臨床病理検討記事、会議報告などである。
- サービスシステムによっても異なるが、1966年の文献まで検索することができ、現在約450万件の文献情報を蓄積している。

## 1. 機関概要

- NLM (National Library of Medicine, 国立医学図書館) は、1836年 Library of Surgeon General's Office として発足した。NLMへと改称したのは1956年でそれ以後、NIH (National Institute of Health, 国立保健研究所) に所属することとなった。
- NAL (国立農学図書館), Library of Congress (国立議会図書館) とならび、アメリカの3大図書館の1つと言われる。
- 1983会計年度のサービス状況と1984会計年度の予算を以下に示す。

	1983	1984
蔵書数 (単行書, 非単行書)	3,149,000	3,242,000
決算額, 予算額	\$46,406,000	\$49,613,000
フルタイム職員数	521	546
受入雑誌数	23,470	22,291
MEDLARS 索引件数	310,445	306,000
うち Index Medicus 索引件数	284,856	279,000
Index Medicus 索引雑誌数	2,709	2,695
目録作成件数—完全	11,322	11,243
—簡略	7,126	6,267
—計	18,448	17,510
収書費	\$ 1,861,000	\$ 2,390,000
貸出件数 (謝絶を除く)	340,000	308,000
うち相互貸借	177,000	116,000
うち来館者	163,000	192,000
コンピュータ検索 (米国内, 全データベース)	2,412,000	2,802,000
オンライン	2,020,000	2,406,000
オフライン	392,000	396,000

(National Library of Medicine News, vol. 39, no. 11, Nov. 1984)

## 2. データベースサービスとネットワーク

### 1) データベースサービスの沿革

- 1964. 1 Index Medicus 誌を電算化するために、バッチコンピュータシステムである MEDLARS が作られた。
- 1967. オンライン文献検索サービスの実験始まる。
- 1970. 6 Western Union 社の Teletype Writer Exchange を使った AIM-TWX が実験システムとして稼動。
- 1971.10 MEDLINE 稼動。
- 1975. 1 MEDLARS II へ移行。
- 1979 MEDLARS III 計画開始。

### 2) ネットワーク

- 米国内のネットワークは下図の通り (□はノードである)。

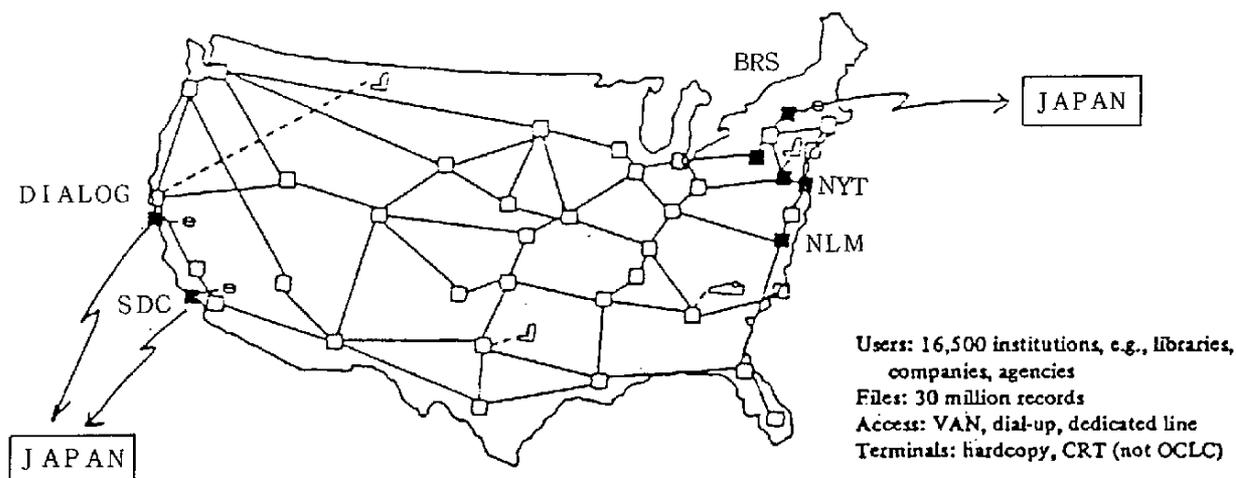


FIGURE 2. Search services (Lockheed, SDC, BRS) and value-added networks (VANs) (TYMNET, TELENET).

- この他に、海外14カ国にMEDLARS センターを設置し情報提供を行なっている

#### 1) NLMと直接オンラインで接続している国

- カナダ、コロンビア、フランス、イタリア、メキシコ、南アフリカ、英国

#### 2) NLMからソフトウェアと磁気テープの提供を受け、自国のコンピュータを用いてサービスをしている国 (機関名)

- オーストラリア、スウェーデン、ブラジル (Pan American Health Organization)

#### 3) NLMから磁気テープのみの提供を受けている国 (機関名)

- ・西ドイツ (DIMDI), スイス, 日本 (JICST)
- ・その他, 商用オンラインシステムであるDIALOG, BRS, SDC, DATA-STARを通じても利用できる。日本国内では, この4種とJOISによってMEDLINEを利用することができる。
- ・NLM自身もELHILLと呼ばれるオンライン検索プログラムに制御されるオンラインシステムを持っており, MEDLINEのみならず, NLMが作成している他のデータベースもオンラインでサービスされている。

#### NLMオンラインシステムの発展過程

期 間	1970-1971	1972-1974	1975-1979
システムの名称	MEDLARS Iの AIM-TWX phase	MEDLARS Iの MEDLINE phase	MEDLARS II
オンライン検索 プログラム	ELHILL I	ELHILL II	ELHILL III
計 算 機	IBM 360 / 67	IBM 370 / 155	IBM 370 / 158 (1980年末 IBM 370 / 168)
データベースの種類	1	1 - 6	6 - 17
契約機関数	90	190 - 430	430 - 1020
質問数/年 (オンラインとオフ ラインの合計)	-	212,000 } 503,000	503,000 } 1,404,000

(出典: 科学技術情報ハンドブック改訂版)

#### ・ハードウェア構成

ハードウェア構成 (1978年1月時点) は, 下記のとおりである。

中央処理装置: IBM 370 / 158 2台 (合計 4 MB)

磁気ドラム装置: 11 MB 1台

磁気ディスク装置: 70 MB × 8 ; 200 MB × 16 ; 300 MB × 24

(出典: 科学技術情報ハンドブック 改訂版)

・NLMから直接利用している利用者の利用状況は以下の通り（1981会計年度）。

利用者グループ	機関数	利用時間	%（総利用時間に対する）
医科系大学	195	33,076	20
非医科系大学，図書館等	77	5,879	4
病院	684	54,908	34
政府，研究所，学会等	204	16,712	10
営利企業	368	21,195	13
その他	22	469	1（以下）
米国外機関	342	11,667	7
NLM	—	19,397	12
合計	1,892	163,303	100%

### 3. 情報源と情報収集方法

- ・情報源は1983年度から雑誌のみに限られている。
- ・収録雑誌タイトル数5,499（うちIndex Medicus収録数は，2,695誌）（1984年）。
- ・年間収録文献数301,418件（1983年度）＜1カ月に約25,000件ずつつけ加えられている＞。
- ・1984年には，これまで対象とされてきた原著，原著類似文献，総説に加え，次の文献群が収録対象に加えられた。
  - (1) 米国公衆衛生局（U. S. Public Health Service）または，米国立保健・医療研究所（U. S. National Institutes of Health）職員のMEDLINE収録誌掲載原著論文のすべて（医学，非医学論文を問わず）
  - (2) 上記2機関の援助（資金）によっておこなわれたと記述のある収録誌論文のすべて
- ・雑誌タイトル数等の年次変化は以下の通り。

年度 *1	1980	1981	1982	1983
雑誌タイトル数 *2	3,283	3,558	3,643	5,346
IM雑誌タイトル数	2,661	2,664	2,697	2,709
収録文献数	266,730	279,105	282,950	310,445
IM収録文献数	243,873	256,112	259,874	284,856
著者英文抄録数	111,629	126,742	124,511	149,851

\*1 雑誌タイトル数以外は、すべて米国会計年度の統計。たとえば、1983会計年度は、暦年の1982年10月1日から1983年9月30日までである。

\*2 参考文献2) - 6) から算出。1984年度雑誌タイトル数は、5,499。

2) List of serials and monographs indexed for online users - 1980. U.S. National Library of Medicine, January 1980.

3) List of serials and monographs indexed for online users - 1981. U.S. National Library of Medicine January 1981.

4) List of serials and monographs indexed for online users - 1982. U.S. National Library of Medicine, March 1982.

5) List of serials indexed for online users - 1983. U.S. National Library of Medicine, March 1983.

6) List of serials indexed for online users - 1984. U.S. National Library of Medicine, February 1984.

(出典：後藤敬治，1985年度MEDLINEファイル。医学図書館，32(1)，p. 37-46 (1985))

#### 4. データベース機能の概要 (データ項目、検索機能等)

レコードの内容は、MEDLINE中の各書誌データユニット・レコードに含まれるデータ・エレメントは37種類あり、以下の表にそれらのデータ・エレメントと、そのエレメントが、直接、検索可能 (Directly Searchable) か、STRING SEARCHかどうか示され、そして出力について3種 ("PRINT", "PRINT FULL", "PRINT DETAILED") の方法のそれぞれに、コンピュータから印字されるエレメントの種類が、(X)で指示されている。

検索、出力についてのコマンドは、ELHILLのものである。

#### MEDLINE データ・エレメント

Category Qualifier	Data Element Name	Directly Searchable	STRING-SEARCHABLE	"PRINT" "PRINT"	"PRINT FULL"	"PRINT DETAILED"
AU	Author	X	X	X	X	X
TI	Title		X	X	X	X
CN	Citation Number	X	X			X
ED	Entry Date	X	X			X
MC	MeSH Class Number	X				
IS	ISSN	X	X			X
LA	Language	X	X		X	X

PL	Place of Publication	X	X				X
PN	Place of Publication Class						
	Number	X	X				X
MH	Main Headings	X	X		X		X
SH	Subheadings	X					
YP	Year of Publication	X					
PD	Publication Date	X	X				X
LI	Special List Indicator	X	X				X
SO	Source			X	X		X
CI	Citation Identifier				X		X
AB	Abstract		X				X
AS	Abstract Source		X				X
TA	Journal Title Abbreviation	X	X				X
SU	Subject as a Name	X	X				X
JC	Journal Title Code	X	X				X
EA	English Abstract Indicator		X				X
JI	Journal Article Identifier— Supplementary Data Indicator					X	X
CC	Catalog Citation Number		X				X
VD	Revision Date		X				X
SB	Journal Subset	X	X				X
LV	Last Major Revision		X				X
IP	Issue /Part /Supplement		X				X
ME	Monograph Editor/Compiler		X				X
MT	Monograph Title		X				X
NA	No-Author Indicator		X				X
NR	Number of References		X				X
PG	Pagination		X				X
RP	Reprint Source		X				X
RI	Revision Indicator		X				X
TT	Vernacular /Transliterated Title					X	X
VI	Volume /Issue		X				X

(出典：野添. 医学図書館, 22(2), P. 67~86 (1975))

・ ELHILL の主要なコマンドを以下に示す。

コ マ ン ド	機 能
COMMENT	NLMに対して質問または意見等を伝える。
DIAGRAM DISPLAY	特定のステートメントの質問式を表示したり、登録された質問式を表示したりする。
ELEMENTS	入力タームのタグ種別の省略時解釈を変更する。
ERASEALL	それまでのステートメントをすべて削除する。
ERASEBACK BACKUP	特定のステートメントを削除する。
EXPLAIN	システムメッセージの意味やコマンドの意味をくわしく知りたいとき使う。
FILE	データベースを指定する。
FILES	ユーザが利用できるデータベース全部を表示する。
FIND	直前のシステム側からの質問を無視してステートメントを入力する。
HELP	このコマンドを入力するとシステムからユーザが直面している問題のタイプが6とおり表示されるので、そのどれかを指定すれば、システムから指示がなされる。
MESHNO	特定の MeSH term の MeSH Tree Number を表示。
NEIGHBOR	インデックス・ファイルに登録された特定の用語の前後にある用語を表示する。
NEWS	NLMからのお知らせを表示する。
OFFSEARCH	オフラインサーチ用の質問を入力する。
PRINT	オンライン出力（標準出力様式のほかに自分の指定する項目だけの出力も可能）。
PRINT OFFLINE	オフライン出力（300件まで） 送付先の指定が可能。
PURGE SEARCH	登録質問を削除する。
RENAME	コマンド名または論理演算子を変更する（ANDを+に変更するなど）。
RESTACK KEEP	特定のステートメントだけを残し、他を削除する。
RESTART	それまでの会話過程をすべてクリアして、会話の初期状態にもどる。
SAVE（後宣言） SAVESEARCH（前宣言）	オンラインをひんぱんに利用する質問を登録する。
STOP	システムとの接続を切断する。
STORESEARCH	オフラインで使用する複雑な質問やSDI用の質問を登録する。
SUBS APPLY	副見出語の一括入力。
TIME	それまでに要した接続時間を表示する。
TREE	特定の MeSH 用語の階層関係を表示する。
USERS	その時点で会話しているオンラインユーザ数を表示する。



実験動物は何か（11種）

妊娠しているか

年齢（新生児から老年まで8区分）

in vitro か

症例報告，臨床研究，比較研究

総説

英文抄録付か，である。

- 抄録がついているものがすべてという訳ではない。抄録はすべて著者自身が書いた英文抄録で，入力される文献情報の48.2%に抄録がある。

#### 5. データの構造化と入力情報の分析

- 英語以外で書かれている論文標題は英訳される。
- キリル文字etc. は翻字される。
- 各論文に対し，indexer が（MEDICAL Subject Headings に基づいて），キーワード（dicriptor）を付与する。
- 一論文に10程度のキーワードを与える。
- 収録対象となる雑誌は，毎年，複数の専門家からなるコンサルタントによって検討される。1983会計年度の場合，207誌が，新たに加えられる候補として選ばれ，最終的に55誌が加えられた。また，胃腸病学，スポーツ医学，法医学，家庭医学，環境衛生，各分野の現行収録誌の見直しも行なわれ，71誌が，収録対象から削除されている。
- 検索作業および入力作業については，ファイル中の重要（最多利用）雑誌の90%以上は，1983年3月以降，雑誌到着語，26日以内にファイル化されている。これは，全入力雑誌文献の35%，または，全入力雑誌タイトルの13%に相当する。残りの文献は，3カ月以内にファイルに入力される。この入力時間の短縮は，バーコード方式の雑誌管理，およびオンライン対話方式の入力システムの稼働の結果である。この入力システムは，索引づけ担当者が check lag, PREGNANCY を入力すると，システムは自動的に FEMALE を入力し，さらに HUMAN か ANIMAL のいずれを入力するかを求めるメッセージが端末機に表示され，作業の正確さと迅速性を付加する機能等を有する。
- Index Medicus 用文献の索引づけ作業は，米国国立医学図書館（NLM）職員と NLM と契約した請負索引づけ担当者がそれぞれ27%ずつ，米国外 MEDLARS センター6機関または，それら機関の米国内委託業者が46%の入力作業を行なっている。

- また、従来、索引の質的レベルを一定水準に保つため、歯科学、保健管理 (health administration) 等の特定分野に関する索引作成作業はその分野の専任者が作業を行う方法がとられてきた。
- 1983年度入力作業から、この入力の質的レベル保持システムに加えて、作業の迅速化を目的とすると思われる書誌事項入力作業 (descriptive indexing) と主題分析入力作業の分離システムが実行されている。この結果、書誌事項入力作業の主体は editor と呼ばれる専任者の作業となり、索引作成者の作業はかなり従来より負担が軽くなっている。
- 索引作業に用いられる Tool 類。
  1. Medical Subject Headings (MeSH)
    - (1) Alphabetic List
    - (2) Categorized Lists
    - (3) Tree Structures
    - (4) Permuted Medical Subject Headings
  2. MeSHに関連したTools
    - (1) New Main Headings. 1969 -
    - (2) List of Provisional Headings. 1969 -
  3. MEDLARS Indexing Integrated Authority File. 1968.
  4. Indexing Manual. 1964 -
  5. Technical Notes. 1967 -
  6. Library Network / MEDLARS technical bulletin. 1969 -
  7. MEDLARS Indexing Instructions
    - (1) Genetics. 1968.
    - (2) Respiration Physiology. 1968.
    - (3) MEDLARS Classification of Steroids. 1968.
    - (4) Cell Lines. 1969.
    - (5) Dentistry. 1969.
    - (6) Parasitology. 1969.
    - (7) Pharmacy and Pharmacology. 1969.
    - (8) Veterinary Medicine. 1969.
    - (9) Eponymous Syndromes. 1970.
    - (10) Main Heading / Subheading Combinations. 1970.

- (1) Tumor Manual. 1970.
- (2) Provisional Cross-References. 1972.
- (3) Immunology. 1972.

(出典：後藤，医学図書館 19(1) p. 35-54 (1972))

## 6. DBMS の開発・選定

- DBMSは，どのオンラインサービスシステムを通じて，サービスされているかによって異なる（DIALOG, BRS等は各々DBMSと見ることできる）。
- NLMのオンライン検索プログラムは，ELHILLであるが，これは，NLMから開発を委託されたSDC社が，自社のORBITを基に作成したものである。ELHILLは現在ELHILL3の段階まできているが，NLMとSDC社の共同開発によるものである。

## 7. ファイルの設計

- MEDLINEはサブファイルである，
  - Index Medicus
  - Index to Dental Literature
  - International Nursing Index
 から構成される。
- また，データの新鮮さによってもファイルを分割している。
 

例えばBRSの場合，

( 職別名 )	( 提供年度 )	
M E S H	1979 ~ 現在	カレントファイル
M S 7 8	1975 ~ 1978	バックファイル
M S 7 4	1971 ~ 1974	”
M S 7 0	1966 ~ 1970	”
M E S Z	1966 ~ 現在	統合ファイル
M E S X	最新 1 カ月	

といった分割を行なっている。

## 8. 入出力の設計

- 索引，入力用のフォーマットが決まっており，qualifier については，チェック（）すればよいようになっている。
- 1984年1月からオンライン端末機による入力が始まった。  
システム名Automated Indexing and Management System（AIMS）で，入力作業全般の効率化を意図して設計されている。実験結果によれば入力所要時間は50%短縮されている。
- 出力は，どのサービスシステムを用いるか，どのコマンドを用いるかによって異なる。MEDLARS（ELHILL）を用いた場合，出力されるエレメントについては，「4. データベース機能の概要」の表を参照のこと。

## 9. 標準化ツールの開発と利用

### 1) MeSH (Medical Subject Headings)

MEDLINEでの索引作業の際，キーワード（descriptor）となる語のリスト（シソーラス）。

- キーワードは毎年変化する。

1985年版MeSHでは，総語数14,123語で，新たに加えられたのは122語で，それぞれに，定義，参照語，Tree番号，注記がある。

また，13,193語には，利用上の注記が付されている。

- MeSHは，

アルファベット順リスト

カテゴリーリスト

カテゴリー階段リスト（tree structures）

によって構成されている。

- MeSHは，

MeSH Vocabulary Fileとして，データベース化されている。

### 2) Name Authority File

著者名の典拠ファイルで，データベース化されている。

## 10. 各種マニュアルの作成

### 1) MEDLARS Indexing Manual (PART 1, および2)

索引作業のためのマニュアル

### 2) NLM On Line Service Reference Manual (1982)およびSupplement (1984)

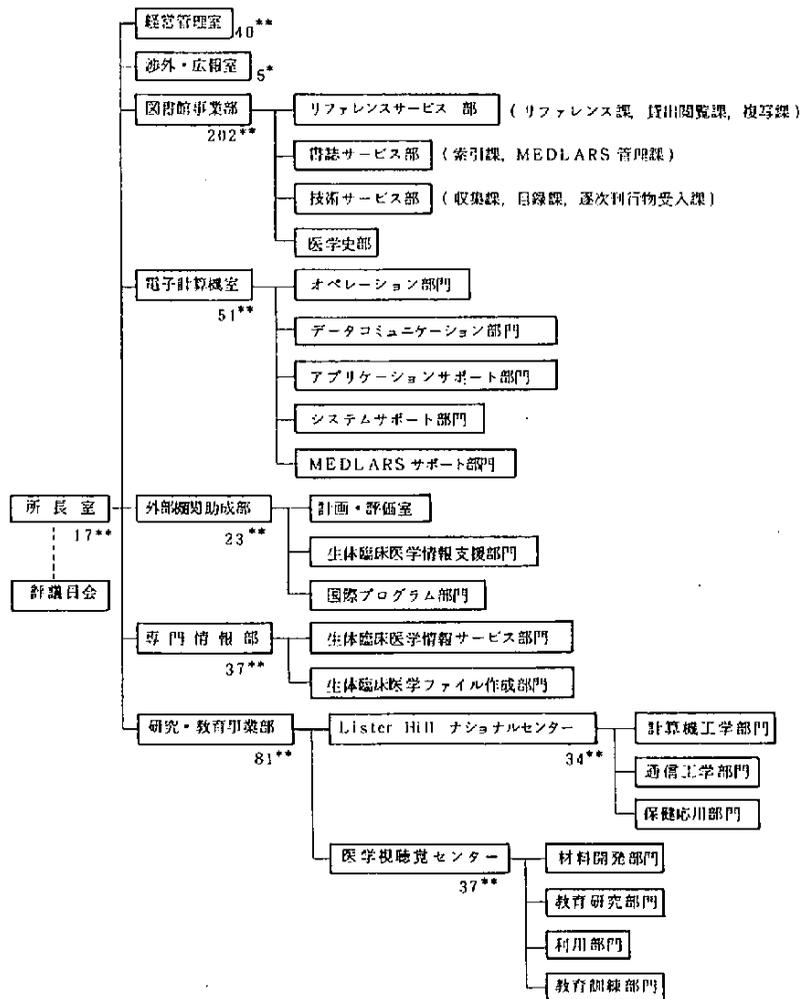
NLMから直接オンラインでデータベースを利用する人々のための検索ガイド

### 3) The Basics of Searching MEDLINE: A guide for the health professionals

一般利用者のための検索ガイド

## 11. 運用・開発面での体制の確立

・NLMの組織図を示す。



注) \* 副所長, 企画担当, 国際関係担当等のスタッフを含む。

\*\* 部室における実人員 (1980 会計年度)

(出典: 菅宮, 情報管理, 25(4)(1982))





データベース構築標準マニュアル

昭和61年3月

財団法人 データベース振興センター

東京都港区浜松町二丁目4番1号

世界貿易センタービル7階

TEL 03-459-8581

禁・無断転載

