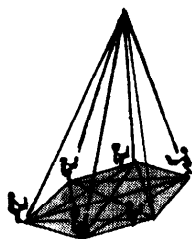


解説

通信網の変革と情報処理



データベース・ネットワーク†

浅野 正一郎††

1. はじめに

情報処理資源の1つであるデータベースの共用を目的としたデータベース・ネットワーク構築への試みは、米国では1960年代半ばから事例が散見できる。我が国では数年来、官庁、民間、学術研究機関等でネットワークが構築され、又計画が進行中である。

データベースの共用形態は、大きく2形態に大別できると思われる。その第一は、集中化され設置されているデータベースの共用により、従来個別に行われていた業務の統一化を企てることにより、全体としての効率化を目指すものである。個別に作成・維持・管理されていた同一種類のデータを集中化することにより、欠落が無く完備されたデータベースを構成することによる利便の向上を指向したものや、種類の異なるデータを集中させることで、システム維持・管理の効率化を図ると共に、新たなデータ利用を開発することを指向したもの等、属性は異なるが第一の類形に入るものは多数存在する。これらの背景には、大群化効果による利便の向上があるといえよう。

データの作成・維持・管理に高度な専門知識を必要とする場合には、専門家の近傍に分野別データベースを置くことが必然と見做される場合もある。そこにおいても、関連する分野のデータを横断的に利用することにより、総合的な知見を導出することへの期待が高まっている。即ち、第2の形態は、分散設置されたデータベース相互間を、最新のネットワーク・アーキテクチャ技術を用いて接続するものである。

我が国で現在開発中の代表的データベース・ネットワークとしては、文部省における第一の形態を骨子としたシステムと、科学技術庁・科学振興調整費による第二の形態のシステムとを挙げることができる。

† Database Network by Shoichiro ASANO (Institute of Interdisciplinary Research, Faculty of Engineering, University of Tokyo).

†† 東京大学工学部境界領域研究施設

本稿は、これらシステム開発の概要を解説したものであるが、それに先立ち、これらシステム開発の参考とされている海外事例を次章に概観することにする。

2. 海外事例

2.1 図書館業務の自動化

米国において、1960年代半ばから計画・開発された図書館ネットワークが、今日大規模に運用されている。これは、中央システムに設置されているデータベースにより、学術研究の一次資料である学術図書・雑誌の目録作成を統一化し、同時に所在情報を伴った目録・所在情報データベースを形成することにより、一次資料の相互流通を図り、日常的図書館業務のオンライン化を実現しようとするものである。

この構想は、Columbia-Harvard-Yale 各医学図書館相互の書誌情報共有計画 (Shared Cataloging Project) に始まり、今日 OCLC (Ohio College Library Center) システムに総合的に発展されている^{1),2)}。また図書業務の情報処理化にあたり、米国議会図書館 (Library of Congress) は、標準化された図書の書誌情報 MARC (Machine Readable Cataloging) を作成する計画をたて、MARC II と称されるフォーマットのテープサービスを1968年から開始した³⁾。いわゆる LC-MARC テープサービスである。

LC-MARC を用いて OCLC は1971年からオンライン業務を開始し、その後論理的構成を大きく変えることなく発展し、約2000の図書館にサービスする全米一のシステムとなり、又1980年から全米図書相互貸借サブシステム (ILL; Interlibrary Loan) を運用開始している。同様なシステムとして UTLAS⁴⁾ (University of Toronto Library Automation System; 1973年運用開始) や、ネットワーク技術の活用を志向した RLIN⁵⁾ (Research Libraries Information Network; 1981年運用開始) があり、これらの中には我が国から利用可能なものがある。これらは共に、目録

の作成・所在の登録, ILL がオンラインで可能であることであり, データベース技術が随所で活用されている。又, 使用上の利便を向上させるために, 検索機能が整備されており, 主題・主題分類検索を初めとする機能や, 一般商用データベース・サービスでの知見が内包されている。

これらの事例は次章の学術情報システムの参考とされている。

2.2 商用データベース・サービス

米国における Lockheed DIALOG, SDC (Systems Development Corporation) ORBIT は商用データベース・サービスの代表となっている。これらが実現しているデータベース検索機能は, 後に述べるシステムに大いに参考となっている。即ち, これらサービスは当初は検索機能の設定に差異が認められたが, 互いに行われた機能強化の結果, 今日ではほぼ同様の能力が設定されている。

具体的には, キーワード検索における前方一致やマスクング, 並びに複数のキーワードの連結; 特定語彙の前後に登録されているものを求める Browsing 機能; 質問式の保存・実行; 検索結果の質問化; 等は今後のシステムでは基本機能として設定されるべきものとなっている。

さらに, 多種・多様なデータベースを取り扱う際の運用知見, 例えばデータフォーマットの変換・統一化等は, データベース開発を伴う後述のシステムで共に参考とされている。

2.3 検索コマンドの標準化

EURONET では, そこに加入している種々の情報検索システムへアクセス可能とするために, 検索コマンドの標準化が検討された⁶⁾。1977年の活動の段階では, データベースへの接続・開放; 検索; 検索結果の表示; 案内等に関する基本機能の設定に留まっているが, このような方式による標準化の効用は, データベース・センタ間の相互接続を前提とする「化合物情報の共用ネットワーク」(4章)の方式設計への参考とされている。

3. 学術情報システム

3.1 概要⁷⁾

人文・社会・自然科学の全分野にわたる学術研究に従事する研究者数は, 大学及び関連する研究機関に既に14万人に達しているといわれている。ここにおいて, 国内外の各種学術情報の蓄積が進展し, 学術図

書, 雑誌を初めとする情報資源の大きな部分を保有するに至っている。

一方, 学術研究の進展に伴い, 国内外で流通する学術情報は急激に多量化, 多様化し, 研究者が必要とする情報を迅速かつ適切に入手可能とする状況を創り出すことは, 今後の独創的・先駆的かつ学際的研究の展開に必須であるとする認識が高まっている。

全国7大学に設置されている大型計算機センターを相互接続する大学間コンピュータ・ネットワーク¹⁰⁾(N-1 ネットワーク)の運用は, 現在も規模の拡大を続けており主要コンピュータ・システムの参画が行われ, 又計画されているが, これも前記の状況を背景とした文部省の施策として位置付けられている。

しかしながら現状は, 個別的に運用されている二次情報の提供をネットワークにより活用範囲を広げているに留まっており, 学術全分野にわたる二次情報提供の総合的实现や, 一次情報(原文)自体の提供にいたる一貫した情報提供サービスの実現に対応しているとはいえない。

このような状況を抜本的に改善し, 二次情報の検索から一次情報の提供にいたる一貫した情報提供サービスを行い, 又研究者集団によるデータベース形成を促進し, さらに研究開発機能を内包するような, 総合的学術情報システム実現を画ることが, 昭和55年1月の学術審議会答申「今後における学術情報システムの在り方について」の指摘するところとなっている。

文部省は答申に即し, 学術情報流通体制の確立を促進するために, その中枢機関となる「学術情報センター(仮称)」の開発に必要な調査を昭和55年から実施している。ここでは, 同センターの活動を大きく,

- (1) 学術情報にかかわるデータベース・サービス
 - 二次情報データベースによる情報提供サービス
 - 一次情報の目録・所在情報形成, 並びに提供サービス
 - 学術情報ネットワークの活動に関するクリアリング(案内)サービス
- (2) サービス実施者・利用者等に対する教育・訓練
- (3) 実施業務に関する調査・計画・調整
- (4) 同システムにかかわる研究開発

に分類している。この内第一項は, データベース・ネットワークの形態により実現される機能である。

3.2 データベース・サービスの特徴⁹⁾

前節に述べたように、データベース・サービスは基本的に三つの機能を有する。この内クリアリング・サービスは他の二者の延長として考えられるものとなっており、以下に主要二者の機能の特徴を述べる。

3.2.1 二次情報データベース・サービス

二次情報データベースを利用(検索)する端末は、全国の研究者の手元に設置されているものを使用する。利用されるデータベースは、数並びに量が膨大なものを想定(当面約50種、約250Gバイト)している。これらから、本サービスのために具備すべき機能要件は、

- (1) 多大のデータ量に対して効率的なDBMSの存在が必要。
- (2) 多様な利用技術(例えば学際的研究分野への対応)及び利用水準への対応が必要。
- (3) 既存のN-1ネットワークを経由するアクセスへの対応と、多様な端末からのアクセスへの対応が必要。
- (4) 管理すべきデータベース並びに利用者管理情報が複雑であり、これへの対応が必要。
- (5) 利用技術の指導についてシステムの対応が必要。
- (6) 効率的なデータベース維持(更新)が必要。等が挙げられている。

3.2.2 目録・所在情報データベース・サービス

一次情報の目録・所在情報の形成について、当面緊急に整備すべきものとして、学術図書・学術雑誌に関するものが挙げられる。学術図書については、各国の出版目録データベース(MARC)があり、又学術雑誌については全国大学図書館等総合目録(学総目)がある。この両者の活用を図ることは必須であるが、前者については、MARCに収録されていない図書については独自に書誌情報の入力を行いつつ、かつ全国の大学図書館での受入図書を即時に入力し、矛盾のない総合目録を形成する必要がある。一方雑誌についても同様な書誌データの追加、所蔵データの入力を必要とする。

これらの処理は、図書館の日常的業務の一環として、一定の方式・様式に基づき均一に行われることにより、全国図書業務の重複を省き、効果を発揮することになる。これらの観点から本サービスの機能要件として、

- (1) 管理運用すべきデータベースは、ネットワー

クを介して端末からオンライン更新されることが必要。

(2) 多数の端末(最終的に1000端末以上)からの多量の目録処理(1日当たり2万冊以上)を、効率的に実現するDBMSが必要。

(3) 統一的機能を実現するための図書館端末が必要。

(4) 設置されるすべての端末がほぼ定期的に稼動する状況でのシステム設計が必要。

が挙げられる。又、本業務に近い位置付けにILLがあり、図書館端末を介したILL処理が必要となる。

3.2.3 サービスの相互関連

前述の目録処理により作成される全国総合目録データベースは、二次情報の検索結果に対してその資料の所在を与えるものとして一般利用者の検索対象となる。この意味で、存在するデータベース群は総合的に管理運用されることになる。したがって本システムのデータベース・システムは、

(1) 複数のデータベースを有機的に組み合わせ検索を行うこと(データベースの横断的利用)を可能とする必要がある。

(2) 前記形態や目録処理に適する効率的なDBMS(リレーショナル形のものに適する)が必要。

(3) 各種データベース管理やシステム運用の自動化に資する効率的DD/D(Data Directory and Dictionary)の開発・運用が必要。

(4) 日本語を初めとする各種文字を取り扱うこと、及びこれら文字からのアクセスを可能とすることが必要。

等が主要要件として挙げられている。

3.3 ネットワーク形態

本システムは、データベースがセンタ・システムに数多く設置されるが、図書業務の地域センタの整備も別途進められており、又、既存大型計算機センターに設置されている二次情報をも活用可能とするために、これらはN-1ネットワーク方式により相互接続されることを前提としている。同時に、図書館端末にもネットワーク機能を具備することが検討されている。以上により、既存N-1ネットワークを拡充した、データベース・ネットワークが構成されることになる。

3.4 今後の動向

昭和55年度から調査が開始されている本計画は、文部省学術国際局情報図書館課を中心として検討が進められており、今年度は目録・所在情報サービスの技

表-1 「ネットワーク共有による化合物情報等の利用高度化に関する研究」への参画システム

| システム名称 | 所有するデータベース |
|------------------------|---|
| 通産省工業技術院情報計算センター・システム | スペクトル・データベース |
| 農林水産省農林水産研究計算センター・システム | バイオケミカル・データベース |
| 環境庁国立公害研究所システム | 環境データベース |
| 国立国会図書館システム | 文献・MARC データベース |
| 日本科学技術情報センター・システム | 化合物辞書データベース 農業・バイオリジカル・安全性・ 熱物性・医薬品・急性中毒・化学 文献データベース |
| 日本特許情報センター・システム | 文献データベース 特許情報データベース |

術開発と試行のための「文献情報センター」が東京大学に設置され活動が開始されている。同時に最終的システムの実現に向けた調査・検討も継続されており、次年度以降に「文献情報センター」を中心としたシステム構成が、試行の形態で提示される予定となっている。

4. 化合物情報の共用ネットワーク

4.1 概要

科学技術庁では、科学技術振興調整費による「ネットワーク共有による化合物情報等の利用高度化に関する研究」を昭和56年度から開始している。

化合物は、現在約600万種以上存在し、年々30万種程度増加しているといわれるが、これらに関する科学的な情報を収集・整理し、利用を促進することは、化合物に関する広範な研究の推進に不可欠と見做される。一方、現状の化合物情報は米国のデータベース・サービスに強く依存しており、情報の海外依存に対する危惧が発せられていることも、本計画推進の原動力となっている。

以上の背景に則して、本計画は次の特長を持つ我が国独自の化合物データベース・システムを開発するものとなっている。

- (1) 日本語による化合物辞書を開発する。
 - (2) 我が国の特色(必要性、地域性、法令、日本語)を有する化合物データベースを開発する。
 - (3) 化合物辞書を中心に、各化合物データベースに整合性を持たせ、これらデータベースを共用ネットワークで結合することにより、データの高度利用を達成するネットワーク・システムを開発する。
- これらの内、第3項は特に本稿に深い関わりを持っており、以降これに焦点を当てて解説する。なお表-1

には本計画に参画するセンタ・システムの一覧を示しており、これらシステム関係者、学識経験者、メーカ代表並びに方式設計の委託を受けている日本電信電話公社が、科学技術庁研究調整局総合研究課並びに同振興局管理課情報室のもとに研究組織を構成している。

4.2 共用ネットワーク方式⁹⁾

異なるメーカ・機種種の装置により、又異なるDBMS並びに検索方式により運用されている参画センタ・システム相互間をネットワークにより結合するために、本計画では最新のコンピュータ・ネットワーク技術を活用して実現を図っている。本データベース・ネットワークは中央集中形の資源配置とは異なり、各センタ・システムに設置されているデータベース資源を、センタ間に標準的に設置されるプロトコルに従って相互利用する特色を有する。現状では、同一種のデータベース自体が分散設置されるいわゆる分散形データベースの形態は採用していないが、利用者が通常加入しているセンタ・システム(ユーザ・ホスト)を介して、目的のデータベースを所有するセンタ・システム(サーバ・ホスト)に対してアクセスを可能とするために統一的に適用される、センタ間のデータベース・アクセス・プロトコルを設定している。これにより、利用者が所属するユーザ・ホストの検索(利用)手順に従って、他サーバ・ホストのデータベースを利用可能としている。

4.2.1 ネットワーク形態

前掲表-1に示したセンタ・システム群に電電公社DEMOS-Eを加えた共用ネットワーク形態を図-1に示している。図中の共用ネットワークは、通信網を示しており、これには当面のトラフィック属性を勘案して、DDXパケット交換網の利用が想定されている。

一方、センタ間通信に適用されるネットワーク・アーキテクチャとしては、電電公社を中心として異機種間接続を目的として研究開発されたDCNA¹¹⁾(データ通信網アーキテクチャ)の適用が計画されている。

図-2は、本ネットワークに適用されるネットワーク・アーキテクチャの機能階層構造を示している。この内、データリンク並びにトランスポート制御機能は、DDXパケット交換網とのインターフェースに適合するように機能設定が図られている。又、高位プロトコル共通機能は、DCNAの機能制御レベルの中でのデータユニット制御層(DUC層)、基本属性処理層(FAP層)がこれに対応しており、プロセス間論理パスの設定/解放機能、順序制御機能、応答制御機能、

割込み制御機能、障害処理機能、コード変換機能、データ圧縮機能、プレゼンテーション制御機能等が位置付けられている。

一方、高位プロトコルの内、メッセージ転送固有機能は、一般の情報転送に活用されるべく設定されており、メッセージの分解/組立て機能、障害回復機能が位置付けられている。データベース・アクセス固有機能には、検索・管理系機能と更新系機能が用意されており、これらについては次項で述べている。さらにネットワーク管理固有機能には、運用開始/終了通知機能、利用者の識別/資格チェック機能、利用者の通信許可機能等が位置付けられている。

4.2.2 データベース・アクセス固有機能

本ネットワークで想定されるデータベース利用の一例を 図-3 に示している。このように、利用者は一連のデータベース検索を行う過程で、目的とするデータベースに順次アクセスを行う。この際、目的データベースが設置されているサーバ・ホストに対して、ユーザ・ホストは利用者が入力した検索レベルのコマンドを標準プロトコルに変換しつつ転送し、結果を受け取る処理を実施する。即ち、ここで実施されるセンタ・システム間のプロトコル処理が、データベース・アクセス・プロトコル(DBAP) に従うことになる。

現在参画しているセンタ・システムには、文献検索時の効率を重視した DBMS から、リレーショナル形の DBMS に至るまでの多様な DBMS が運用されている。本 DBAP は、これら DBMS の差異を包含しつつ、文献検索に特徴とされる検索機能を機能的に実

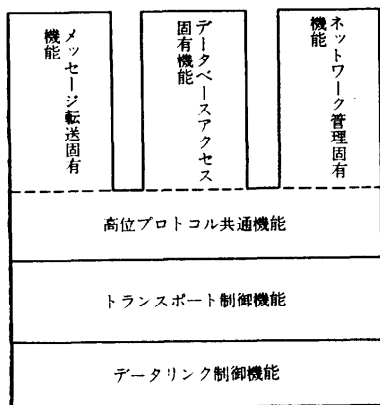


図-2 機能階層構造

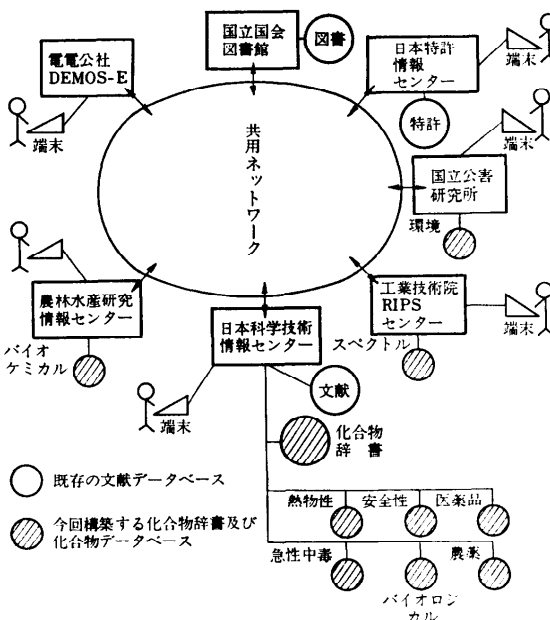


図-1 ネットワーク形態の概念図

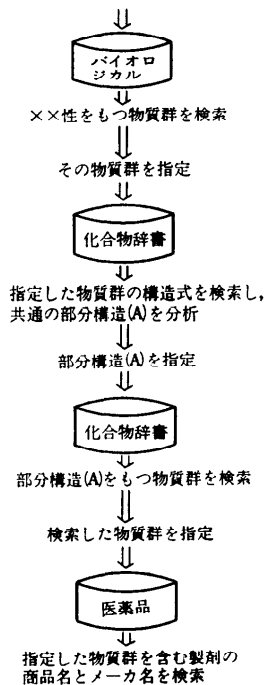


図-3 データベース・アクセス例
例] ある生物学的作用(××性)をもつ物質群に共通する部分構造を調べ、その部分構造をもつ物質群が既存の医薬品製剤中に含まれていれば、その商品名とメーカー名を調べる。

表-2 DBAP 検索機能の概要

| 機能 | 説明 | |
|------------------|--|--|
| 転送種別指定機能 | 転送するレコードオカレンスを指定する機能。 | |
| 一時レコード作成指定機能 | 検索結果のレコードオカレンスをサーバ側に一時的に保存する機能。 | |
| 検索指定機能 | 算術式 | 加減乗除及びべき乗が指定可能。 |
| | レコード選択条件式 | (1)論理条件式、(2)単位比較条件式、(3)集合比較条件式、(4)連結条件式、等が設定されている。 |
| | 単位検索指定 | 検索条件指定の基本単位であり(1)射影条件指定、(2)選択制限上限指定、(3)レコード・タイプ指定、等が設定されている。 |
| | 複合検索指定 | 選択されたレコードオカレンス群の間の集合演算を行うために指定する。 |
| 利用者固有仕様連結指定 | レコード選択条件式に利用者固有の条件式を指定可能とする。これにより複数キーの連結、トランケーション、マスキング、等を可能とする。 | |
| ソート指定機能 | 検索指定で得られたレコードオカレンス群に対して指定のデータ項目の値の順にソート可能とする。 | |
| 分割転送/継続指定機能 | 結果の分割転送/継続を指定する。 | |
| 検索結果のプレゼンテーション制御 | 印字制御、サブスクリプト等の制御文字の挿入を可能とする。 | |

現し得るものとして、DCNA のデータベース・アクセス・プロトコルの機能強化を図りつつ設定された。

本 DBAP のデータベース・モデルは、基本的に繰り返しグループを許すリレーショナルモデルとなっている。各センタ・システムのデータベースは、この仮想モデル（仮想データベース；VDB）にモデル化される。ユーザ・ホストから発せられるアクセス・コマンドは VDB に対する操作として位置付けられる。

本 DBAP の検索は、VDB を利用可能（初期化）とするフェーズ、検索フェーズ、VDB の利用を終了するフェーズから成る。この内検索フェーズの機能を表-2 に要約している。

一方 DBAP には、データベースの管理機能として、データベース属性参照機能（データベース構成情報、データベース状態管理情報の参照）、課金情報取得機能が設定されている。さらに、データベース更新機能として、更新データの転送機能が用意されているが、オンライン更新は現状では仕様を含められていない。

4.3 現状と将来計画

本計画は、昭和 56 年度から同 60 年度にかけた 5 カ年間の研究計画に従って進行中である。過去 2 カ年の間に、データベースの形成並びに共用ネットワーク構

成に関する基本設計が完了しており、今年度のソフトウェア製造を待って、次年度から総合試験、評価が行われることが予定されている。

我が国では、複数の機関にわたる初めてのデータベース・ネットワーク実現の試みであり、今後の情報ネットワーク構成への一つのあり方を示すものとして注目を集めている。

5. おわりに

本稿では、我が国のデータベース・ネットワーク開発事例を中心に解説した。これらは共に現在開発中の事例であり、技術的詳細については未定の部分も多く深く立ち入ることを敢えて避けている。

我が国のデータベース整備の現状は、データベースを海外に求める傾向が強いことを大きな理由として、脆弱であるとする見方が多い。これら事例の開発・運用を通して、データベース並びにシステムの両面で独自の整備が図られてゆくことを期待する。

参考文献

- 1) Evans, G. T.: On-line Library Networking. Bulletin of American Society for Information Science, Vol. 5, No. 5, p. 11 (1979).
 - 2) Epstein, H.: The Networking of Network, 同上, p. 14 (1979).
 - 3) Maruyama, L. S. et al.: Networking Activities at the Library of Congress, 同上, p. 16 (1979).
 - 4) Duchesne, R. M.: Computerized Library Networking in Canada, 同上, p. 18 (1979).
 - 5) Farmer, S. C.: RLIN as a Reference Tool, ONLINE, Vol. 6, No. 5, p. 14 (1982).
 - 6) EURONET Guideline; Standard Command for Retrieval Systems, (Source; INSPEC) (1977).
 - 7) 文部省学術国際局情報図書館課; 学術情報システム—これからの学術情報流通のあり方—(1980).
 - 8) 同上; 学術情報センターシステム開発調査—概要—(1983).
 - 9) 科学技術庁振興局管理課情報室; ネットワーク共用による化合物情報等の利用高度化に関する研究—化合物共通プロトコル仕様書—(1983).
 - 10) 猪瀬 博監修; コンピュータネットワーク技術, 情報処理学会 (1980).
 - 11) 日本電信電話公社; DCNA 基本概念, DCNA PS 010-1980, データ通信協会(1981).
- (昭和 58 年 8 月 25 日受付)