

人文科学研究資源の共有化 ～国文学研究資料館を例として～

原 正一郎

(文部科学省大学共同利用機関・国文学研究資料館)

1. 概要

1990年代の初頭、学術情報発信源の殆どは大学や研究所の大型計算機センターあるいは一部の大規模書誌情報ユーティリティであった。発信される情報も目録が中心で、全文データや画像データなどはごく一部に過ぎなかった。したがって、どの様な研究資源が、どこかのサイトから、どの様な形式で発信されているか、ということについて、それほど精力を注ぐ必要はなかった。しかしインターネットの拡大に伴い、人文科学の領域においても研究資源の種類や発信源の数が急速に増加してある。研究者にとって研究資源の多様化は望ましいことであるが、必要とする資源の探索はかつてほど単純ではなくなりつつある。

多様な研究資源へのアクセスを支援にする方法として Yahoo のようなサーチエンジンがある。しかし検索精度は低く、アクセスできる情報源も Web ページだけであり、データベースを検索することはできない。このような中で、情報資源の共有化を目指す様々な動きが見られるようになってきた。

2. 資源共有の形態

ネットワーク上に分散する情報資源を管理する方法としては、

- 1) センターサーバを構築して、ここに研究資源を集約する
- 2) データクリアリングハウスを構築して、所在・データアクセス法などのメタデータを集約する
- 3) 標準規約の導入により、分散データベースの共通検索を実現する

の3種類に分類できよう。センターサーバ型は国立情報学研究所などの情報センターや OCLC のような巨大書誌情報ユーティリティなどが該当する。また、カリフォルニア大学サンディエゴ校が中心となって組織されている PRDLA (Pacific Rim Digital Library Alliance) のような一種のコンソーシアムなども形成されつつある [http://www.prdl.org]。

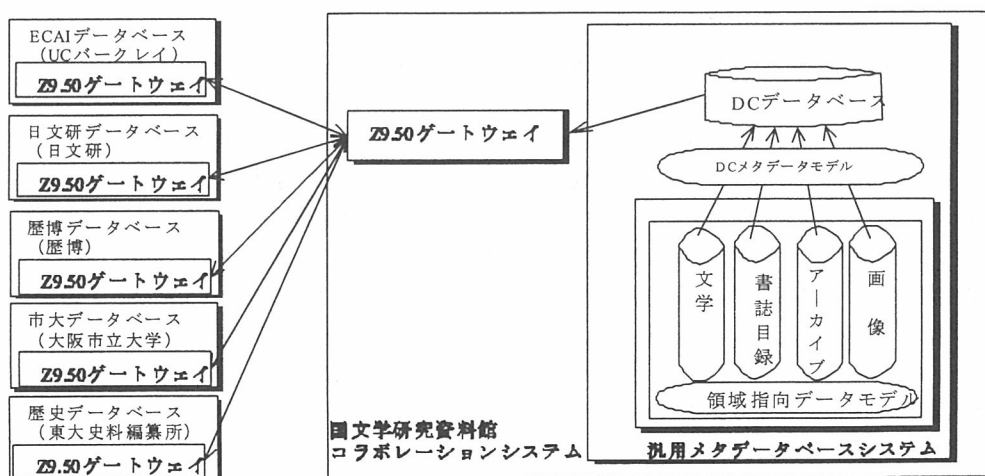
データクリアリングハウス (Data Clearinghouse) は、「手形交換所」などと訳され、情報処理の分野ではネットワークを活用した情報の流通機構、つまり情報の出所・入手方法などに関するデータを収集・検索できるシステムを指すことが多い。インターネット上に情報資源を提供している機関は、その資源に関するアクセス情報(つまりメタデータ)をデータクリアリングハウスに登録する。データ利用者はデータクリアリングハウスを検索することにより、どこに、どのような情報が、どのような形式で存在しているかを知ることができる。現在、このようなデータクリアリングハウスは増えつつある(例えば、人文科学系に関連するものとして Electronic Cultural Atlas Initiative [http://ecai.org] あるいは地理情報クリアリングハウス・ゲートウェイ [http://zgate.gsi.go.jp/] など)。

国文学研究資料館では「国文学研究資料館コラボレーションシステム」を開発中であり、これは第3の方法を採用している。このシステムでは、機関内の複数のデータベース間および複数の機関の情報システム間におけるデータアクセスの透過性を実現するために、検索用規約として Z39.50 を、データ構造の規約としてダブリンコア・メタデータを採用している点にある。

本システムにおけるダブリンコア・メタデータの役割は、データベースの種類を越えた相互利用性の実現である。具体的には、MARC に基づいた OPAC や国文学研究資料館独自のデータ構造を持つ画像データベースなどからダブリンコア・メタデータベースにマッピング可能な要素を抽出し、メタデータベースとして統合する。ユーザはダブリンコア・メタデータベースを検索のゲートウェイとして、全ての館蔵データを統合的に検索することが可能となる。

ところがダブリンコアはデータ要素の定義のみであり、システムの実装については言及していない。したがって、ダブリンコア・メタデータベースシステムといっても、ある機関では XML/SGML のタグを利用した文字列検索システムとして実装され、別の機関では関係データベースシステムとして実現することも可能である。このようにダブリンコア・メタデータだけでは、国文学研究資料館の全資料は検索可能であったとしても、機関を越えた検索を行うことはできない。これを解決する方法として、検索手順に関する標準規約である Z39.50 を導入した。データクリアリングハウスも機関を越えたデータ検索を可能とする。しかしデータクリアリングハウスには専門領域に特有なメタデータが蓄積されることが多い。またデータクリアリングハウスを構築するためには、関連する機関・領域団体との調整が必要であり、さらにデータセンターを構築・維持するためのコストも考慮しなければならない。Z39.50 は単なる規約であるため、内容に関する調整やデータセンターのための費用は不要である。このような理由から、本研究では Z39.50 を採用した。

国文学研究資料館コラボレーションシステムには、国文学研究資料館の目録・画像・論文目録・歴史史料所在・OPAC などのデータベースが統合されつつある。このシステムが目指すシナリオは、例えば、国文学研究資料館の史料所在データベースから「伊能家」を検索すると、やはり国文学研究資料館のマイクロ資料目録データベースから伊能忠敬の「日本経緯度実測」の所在情報さらに画像データベースからその画像情報など、関連するあらゆる情報を単一の GUI 操作でシームレスかつ高い精度で検索することである。ところで複数の図書館・博物館・文書館に分散している資料あるいは史料を検索する際にも、ユーザは上記と同じ問題に直面する。もしネットワーク上にデータを公開している機関が国文学研究資料館と同様なコラボレーションシステムを導入していれば、機関を越えたデータ検索も可能になると期待している。



コラボレーションシステムの構想