

# 研究者が発信する学術情報の流通促進を目指した情報基盤 —京都大学地域研究統合情報センターの試み—

原 正一郎

京都大学地域研究統合情報センター

Web への常時接続が当たり前になり、あらゆる情報をネットワークから入手できる時代を迎えつつある。そのような情報化時代にあつて、研究者が公開しているデータベースはそれほど多くない。さらに、公開にこぎつけても利活用の進まないデータベースや、予算減あるいは管理者不在により公開停止を余儀なくされたデータベースは枚挙にいとまがない。本稿では、それら背景にある問題を整理するとともに、研究者が発信する学術情報の流通促進を目指した京都大学地域研究統合情報センター（以下では地域研）の取り組みについて述べる。

## Information Infrastructure for Facilitating the Distribution of Academic Information

Shoichiro HARA

Center for Integrated Area Studies, Kyoto University

Even in the information age we can get everything from the Web, contributions to information distribution on the Web by databases build by academic researchers are much limited. Furthermore, some of the databases have not been used so much, and some of other databases were compelled to stop their services due to budget decrease and/or manager absence. This paper considers the background of these problems, and explains the trials by CIAS (Center for Integrated Area Studies, Kyoto University) to facilitate the distribution of academic information.

### 1. まえがき

Web への常時接続が当たり前になり、あらゆる情報をネットワークから入手できる時代を迎えつつある。そのような情報化時代にあつて、研究者が公開しているデータベースはそれほど多くない。さらに、公開にこぎつけても利活用の進まないデータベースや、予算減あるいは管理者不在により公開停止を余儀なくされたデータベースは枚挙にいとまがない。本稿では、それら背景にある問題を整理するとともに、研究者が発信する学術情報の流通促進を目指した京都大学地域研究統合情報センター（以下、地域研）の取り組みについて述べる。

### 2. 研究者が発信する学術情報の特徴

多様な学術情報が、アーカイブ・図書館・博物館・機関レポジトリなどからインターネット上に公開されているが、その特性は相互に異なっている（表 1）。

アーカイブは記録資料の保存を意味し、日本では公記録保管所・公文書館・記録保存館など訳されることが多い。ここから推察されるように、アーカイブでは個人・組織・団体などに関する記録資料が主たる対象となる。記録資料の特徴は、文書に加えて地図・写真・動画などの多様なメディ

アが含まれていること、原資料をそのまま保存する必要があること、出所ごとに総体として整理すること、資料構造・来歴・関連資料などについての詳細な記述が重要などである。

図書館は一般的に書籍を対象とし、博物館はモノを対象としているので、両者のメタデータは異なっている。またアーカイブは資料を総体的に扱うのに対して、図書館・博物館は一点ごとが単位となっているので、アーカイブと図書館・博物館のメタデータも異なっている。

大学などでは機関リポジトリ（repository）と呼ばれる情報サービスが普及しつつある。機関リポジトリとは、大学などの研究機関の知的生産物をデジタル的に集積・保存・公開する仕掛けである。研究成果を自主的に保存・公開すること、灰色文献（出版されないものや失われやすい資料）の保存・公開を目指す点に特徴がある。素朴な意味ではアーカイブであるが、機関リポジトリのメタデータはアーカイブというよりも図書館に近いものとなっている。

このようにアーカイブ・図書館・博物館・機関リポジトリのメタデータ構造は相互に異なっているものの、一般（公衆）を対象としていること、公開性が高いこと、作業を組織的に行っていること、資料整理やデータベース構築の方針が一貫していること、標準メタデータに準拠していること、

	アーカイブ	図書館	博物館	リポジトリ	研究者DB
主要な利用者	一般(公衆)	一般(公衆)	一般(公衆)	学生・研究者	研究者・専門家グループ
公開の目的	一般的	一般的	一般的	教育・研究	研究/特定の目的
公開性	高い	高い	高い	高い	低い
構築主体	組織的	組織的	組織的	組織的	個人あるいは研究グループ
構築方針	一貫している	一貫している	一貫している	一貫している	変更されることが多い
システム等の更新頻度	低い	低い	低い	低い	高い
活用	検索(限定的)	検索(限定的)	検索(限定的)	検索(限定的)	多様
資料の多様性	大きい (記録・モノ)	必ずしも大きくはない (書籍)	大きい (モノ)	必ずしも大きくはない (教材, 論文)	大きい
コレクションの範囲	網羅的	網羅的	網羅的	網羅的	部分的
データ量	大きい	大きい	大きい	大きい	小さい
メタデータ	標準的	標準的	標準的?	標準的	特殊(非標準)
永続性	長期的	長期的	長期的	長期的	短期的

表1 図書館等と研究者によるデータベースの相違

データベースの永続性が高いこと、システムおよび応用プログラムの変更頻度が低いことなど、共通点も多い。一方で、検索方法が限定されており研究目的には使いにくい点も共通している。

アーカイブ・図書館・博物館・機関リポジトリに加え、大学・研究機関などの研究者が公開しているデータベース（以下、研究者DB）も学術情報発信の重要な一翼を担っている。しかし、その様相は図書館などとはかなり異なっている。研究者DBの特徴は、紙資料・画像・動画・音声・テキスト・数値データなど多様な研究素材を対象としている点にある。標準メタデータの整備が進んでいない素材も多く、また研究者の多くがメタデータへの関心が薄いため、メタデータはデータベースごとにバラバラである。研究目的に応じてデータベース・検索手順・ユーザインタフェースを設計しているので、汎用性や公開性は低い。そのため、他の研究者にはアクセスできなかつたり使い勝手の悪かつたりする事例も多い。研究の進展によっては、メタデータ構造やデータベース設計の方針を変更する場合もある。さらに、予算不足のために資料収集やシステム運用を中止するなど、永続性が低いという問題も抱えている。

学術データベースというと、これまでは図書館などが議論の中心であったように思われる。しかし研究者DBは貴重な一次資料を扱っているため、それらの離散・消滅を防ぐ意味でも、安定的に利用でき、研究にも資するデータベースのあり方を考慮すべきであると考えられる。3章では、研究者DBの公開・利活用を支援する情報ツールの研究開発事例として、地域研のMyデータベースについて述べる。

一方、アーカイブ・図書館・博物館・機関リポジトリ・研究者DBがインターネット上で公開されるようになると、史資料をデジタル化して保存・公開を行うものとして、十把一絡げにデジタルアーカイブと呼ばれるようになった。インター

ネットを新しいメディア空間と考えれば、図書館や研究者DBといった旧来の区別を議論するのではなく、研究資料をインターネットメディア空間でどのように統合していくかに研究の重点を移すべきである。その試みとして、4章では地域研の資源共有化システムについて述べる。最後に、次世代共有化システムを目指し、オントロジーに着目した情報システムの試みについて考察する。

### 3. My データベース

研究資料の公開にはデータベースシステムが必要であるが、その構築は情報学系以外の研究者にとって敷居が高い。またデジタル化・システム構築・管理にかかる経費も決して安くはない。そのため、資料のデジタル化とメタデータ作成までにはこぎつけたものの、データベース公開に至らない研究資料はかなりの数に上るものと推察される。またデータベース公開を実現しても、予算不足やシステム管理者の不在などの理由で公開停止に至った事例も枚挙にいとまがない。さらに研究目的や進捗状況に応じて、データ構造・検索手順・インタフェースなどを頻繁に変更・拡張する事例も多い。以上より、研究者DBは以下の要件を満たす必要がある。

- ① メタデータの定義や修正が容易であること
- ② データベースシステムに関する知識がなくともデータベースを構築できること
- ③ GUI やユーザアプリケーションの構築・修正・拡張を低コストで実現できること
- ④ システム管理が不要かつコスト負担が低いこと

このような条件を満たす研究者DBを目指し、地域研ではMyデータベースと称するデータベース構築支援ツールを開発している[1]。Myデータベースは、メタデータ登録と検索画面設定の機能だけを研究者に提供する。システム管理や運用などは地域研の責任において実施するので、研究

No	延喜式社名	比定社名	緯度	経度	式内社国名	式内社郡名	比定住所1
No	Shikinaisha	IdentifiedName	Lat	Lon	CountryName	CountyName	Address
20000493	阿爲神社	阿爲神社	34.854167	135.56278	播磨国	嶋下郡	大阪府茨木市大字安威
20002117	阿比多神社	阿比多神社	37.16	138.1664	越後国	頸城郡	新潟県上越市長浜
20000506	阿比太神社	阿比太神社	34.825278	135.45639	播磨国	豊島郡	大阪府箕面市桜ヶ丘
20002039	相見神社	相見神社	36.83028	136.7558	能登国	羽咋郡	石川県羽咋郡宝達志水町妻生
20000995	取国神社	取国神社	34.787222	136.16444	伊賀国	阿拝郡	三重県伊賀市一之宮
20002425	香澤神社	香澤神社(香波神社)	35.66694	139.5525	武蔵国	多磨郡	東京都調布市深大寺元町
20001857	香海神社	香海神社	35.4875	135.5103	若狭国	大飯郡	福井県大飯郡高浜町香
20002624	香念神社	熱田神宮境外摂社香念神社	35.12417	136.905	尾張国	愛智郡	愛知県名古屋市長久区白鳥
20002276	赤城神社	赤城神社	36.48444	139.1778	上野国	勢多郡	群馬県前橋市三夜沢町

表 2 データサンプル (式内社)

桶谷猪久夫氏 (大阪国際大学教授) 作成データを改変

者は各自のメタデータだけに關心を払えばよい。なお初期の My データベースには、GUI の不具合や使いにくい箇所があり、さらにセキュリティ上の問題もあったため、試行運用の状態が続いていた。これらの改良を進めた結果、今年度より一般サービスが可能となった。

### 3.1 My データベースの構築手順

My データベースで利用可能なメタデータ形式は、CSV・TSV・XML の 3 種類である。フィールド名あるいはエレメント名の付け方や、それらの出現順などについての制限はないので、独自のメタデータを自由に定義できる。My データベースの構築手順は以下のように簡単である。

- ① ユーザ登録: My データベースの登録を管理者に依頼すると、利用者 ID およびパスワードが発行される。同時に一定サイズの作業領域が My データベース内に確保される。利用者は与えられたサイズ内でメタデータおよびコンテンツデータの登録を行う。より多くの作業領域が必要な場合は、別途申請する。
- ② データファイル作成: スプレッドシートやエディタソフトを利用してメタデータファイルを作成する。CSV テーブルの例を表 2 に示す。これは地名辞書データの一部である。ここでテーブルの 1 行目はヘッダ行、2 行目は英語によるヘッダ行である。この例では、No フィールドがキーとなっている。

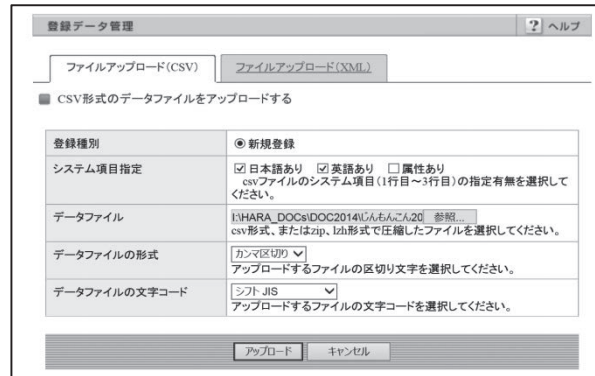


図 1 メタデータ登録画面例

- ③ メタデータ登録: My データベースの登録画面からメタデータファイルを登録する (図 1)。登録したメタデータの修正はレコード単位あるいはファイル単位で行う。データ追加についても、レコード単位あるいは差分ファイル単位で可能である。
- ④ データ属性の設定: My データベースでは、検索や検索結果表示の対象とするデータ項目を属性パラメータにより設定する。例えば図 2 において PRIME と設定された「延喜式社名」は、検索対象かつ検索結果詳細表示画面へのリンクとなる。また BASIC と設定されたデータ項目は、検索対象・検索結果一覧・検索結果詳細表示用となる。通常の検索

No	1	2	3	4	5	6	7
日本語見出し	No	延喜式社名	比定社名	緯度	経度	式内社国名	式内社郡名
英語見出し	No	Shikinaisha	IdentifiedName	Lat	Lon	CountryName	CountyName
属性	編集 ID フィールド追加	編集 PRIME フィールド追加	編集 BASIC フィールド追加	編集 SEARCH フィールド追加	編集 SEARCH フィールド追加	編集 BASIC input_type=listbox フィールド追加	編集 BASIC フィールド追加
属性(プレビュー)	ID	PRIME	BASIC	SEARCH	SEARCH	BASIC&input_type=listbox	BASIC
1	20000493	阿爲神社	阿爲神社	34.85416667	135.5627778	播磨国	嶋下郡
2	20002117	阿比多神社	阿比多神社	37.16	138.1663889	越後国	頸城郡
3	20000506	阿比太神社	阿比太神社	34.82527778	135.4563889	播磨国	豊島郡
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
2840	20002449	今城青八坂稲葉神社	天神社	36.24916667	139.1127778	武蔵国	賀美郡
2841	20002450	今城青坂稲葉荒御魂神社	長輪部神社に合祀	36.24305556	139.1108333	武蔵国	賀美郡
2842	20002451	今城青坂稲葉池上神社	今城青坂稲葉池上神社	36.26898889	139.1536111	武蔵国	賀美郡

図 2 データ属性の設定画面例



語彙の入力にはテキストボックスを利用する。しかし、国名・年号・少数の主題語などのように語彙数が限定されている場合は、チェックボックス・プルダウン・リストボックスなどの方が便利である。My データベースでは、このような入力方法の変更もデータ属性で行う。例えば「式内社国名」の属性は「input\_type=listbox」と設定されている。これにより、メタデータ登録時に「式内社国名」の内容から語彙リストを生成し、検索画面にリストボックスとして表示する (図 3)。

図 3 My データベースの検索画面例

- ⑤ 画面設定: My データベースを構成する各画面のレイアウトはほぼ固定されているが、ある程度の変更は可能である。検索画面を構成するヘッダ領域、データベース領域、フッタ領域などの配置・背景色・画像などは管理画面により変更可能である。同様に、カテゴリ検索・簡略表示・詳細表示画面などの変更も可能である。
- ⑥ 公開設定: データベースの公開基準を設定する。公開基準は、特権利用者・ユーザーグルー

図 4 My データベースの時空間検索画面例

プ・一般利用者に区別されている。IP アドレスによるアクセス制限も設定できる。

My データベースの構築に必要な時間はデータ量と設定の詳細さに依存する。メタデータファイルが完備しており、データ量が少なく、デフォルト設定でよければ、データベースの構築から公開まで数分程度である。

My データベースの検索画面例を図 3 に示す。この例では「式内社国名」をリストボックスに設定している。さらに、表 2 のデータベースのように時空間データが含まれている場合、My データベースでは地図あるいは時間スライダを利用した検索も可能である (図 4)。

### 3.2 My データベースの応用プログラム

研究者 DB の応用プログラムに対する要求は研究者ごとに異なり、変更・拡張などの頻度も高い。一方、サーバ管理者の側では人員や予算不足などにより、個別の要求に十分に答えることができない。これが学術情報システムの利活用を考えている研究者の不満となっている。この問題を解決するために、地域研では以下の方針を設定した。

- ① 地域研の情報システムはレポジトリ機能と共有機能に限定する。
- ② データベース構築および応用プログラム開発は各研究者に任せ、地域研はそのための支援環境を提供する。

データベース構築の支援環境は、これまで述べた My データベースである。My データベースの応用プログラムを構築する支援環境として、地域研ではデータベースの所在を示す URI (Uniform Resource Identifier) と My データベースの検索仕様である API (Application Programming Interface) を公開している。API の特徴は以下の通りである。

- ① 従来の SOAP[2] による Web サービスは実装がやや面倒であるため、My データベースでは HTTP の GET を利用した REST 的[3]な Web サービスを提供している。
- ② REST 的な Web サービスでは、サーバ・クライアント間の検索・返戻手順は CGI や Servlet と同じであり、応用プログラム作成における通信部分はかなり簡素化される。
- ③ 返戻形式として XML と JSON[4] を用意しているため、返戻データの変換などには様々なツールの利用が可能である。そのため、信頼性の高い応用プログラムを効率的に作成することができる。

My データベースの API では、検索式として Context Query Language (CQL) [5] を採用している。CQL は SRU (Search/Retrieve via URL) およ

び SRW (Search Retrieve Web services) [6]で利用される検索式である。SRU が検索要求に REST を用いるのに対して SRW では SOAP を用いる点が異なっている。SRU の検索式は

**Request URL:: Base URL?Request Part**  
**Request Part:: Parameter=value[&Parameter=value]\***

と定義されている。例えば検索式が

**http://Some\_URI?operation=searchRetrieve&version=1.2&query=(c2="宇太水分神社")&recordSchema=original**

の場合、**Base URL** は「http://Some\_URI」に相当する。My データベースでは、**Base URL** を利用して、各データベースの所在に加えて返信形式 (XML, JSON など) の設定も行う。「?」の後ろが **Request Part** である。さらに **Request Part** のうち「query=」の部分が CQL による検索式である。検索式中の「c2」はデータ項目名を指示する My データベース独自の方法であり、この例では表 2 の「延喜式社名」を指示している。一般的なデータベースシステムでは、データ項目名の中で空白を使うことを許さない。しかし、Microsoft Excel などデータテーブルを作成する研究者の多くが、データ項目名の中で空白を使っている。初期の My データベースではデータ項目名の修正を依頼していたが、別の誤りを招く事例が多かったため、最新の My データベースでは、このような対応を採用している。

CQL は簡潔かつ直感的な検索式であり、読みやすくプログラム処理も容易で、そのまま Web ブラウザに入力すれば動作確認も可能である。例えば、上記の検索式を入力で表示すると (ただし **Base URI** を正しく設定し、検索語の「宇太水分神社」を URL エンコードし、返信形式を Simple JASON とすると)、以下のような返信を得る。

```
{
  "numberOfRecords": "1",
  "recordData": {
    "original": {
      "c1": "20000207",
      "c2": "宇太水分神社",
      "c3": "宇太水分神社上宮",
      "c4": "34.35361111",
      "c5": "135.8730556",
      "c6": "大和",
      "c7": "宇陀郡",
      "c8": "奈良県宇陀市 (菟田野町) 古市場***"
    },
    "itemset": {
      "c1": "No",
      "c2": "延喜式社名",
      "c3": "比定社名",
      "c4": "緯度",
      "c5": "経度",
      "c6": "式内社国名",
      "c7": "式内社郡名",
      "c8": "比定住所 1"
    }
  }
}
```

これは「延喜式社名」が「宇太水分神社」のデータを検索した返戻で、「original」の部分は返戻データの本体、「itemset」の部分はデータ項目名のリストである。このような形式のデータは加工が容易であり、研究目的に適した GUI の作成や他のデータベースと連携したマッシュアップの構築などは、容易かつ低コストで実現できる。

### 3.3 My データベースの応用例

My データベースと API の応用例を図 5 に示す。学術出版物には多数の図表などが収録されているが、掲載できる数には限界がある。また地図の場合、対象地域の全体地図を載せると詳細部分を見ることができず、反対に詳細地図を載せれば全体を把握するのが困難となる。

このような問題を解決する手段として、QR コードを介して紙面と My データベースをリンクする新しい学術出版物の開発を、京都大学学術出版会と共同で試みている [7]。地図などに付与されている QR コードを、読者がスマートフォンなどの端末で読み込むと、そのデータは京都大学学術出版会のサーバに転送される。サーバでは API を利用して My データベースを検索し、必要な静止画などを取り出して、読者の端末に返送する。これにより、書籍では困難であった地図などのズームイン・ズームアウト、関連画像へのリンクなどが可能となった。

これまでに、布野修司 (著)「グリッド都市」(2013 年 2 月)と、小島敬裕 (著)「国境と仏教実践」(2014 年 2 月)を京都大学学術出版会と連携して実現している。後者では、踊りなどのカット写真から、その動画データへのリンクを実現している。



図 5 API を利用した My データベースの応用例

## 4. 資源共有化システム

My データベースで構築されたデータベースは研究者の個人的な努力の結晶であるため、図書館などにおいて業務として構築されているデータベースに比べると、規模は小さくコレクションと

しても完結していないことが多い(表 1)。コレクションとして完結させるためには、複数のデータベースを連携する必要がある。また研究の過程において複数のデータベースを利用する場合も多いので、データベース連携ができれば便利である。インターネット上のデータベースを連携させる機能を、本稿では共有化と呼ぶ。

共有化を実現するためには、各データベースのメタデータが単一であることが望ましい。例えば OPAC (Online Public Access Catalog) では、この方法を採用している。しかし、メタデータの強制は研究の独自性や発展を阻害する要因となる。そもそも、多様な研究領域を包摂する人文科学において単一のメタデータを定義すること自体が不可能である。そのため My データベースでは特定のメタデータを強制していない。

#### 4.1 資源共有化システムの構造

My データベースで構築されたメタデータの異なるデータベースを共有するために、地域研では資源共有化システム[8]を開発している。資源共有化システムの構造を図 6 に示す。

資源共有化システムでは、各メタデータから独立した共有化メタデータを定義し、各データベースのメタデータと共有化メタデータのデータ項目間には対応関係を設定する。これをマッピング規則と呼んでいる。資源共有化システムから各データベースへの検索には、標準検索プロトコルである Z39.50 または SRW を利用する。資源共有化システムのサーバには、データベースごとのマッピング規則、データベースの URI、検索プロトコルなどの情報が登録されている。これを共有化情報と呼んでいる。

資源共有化システムに与えられた検索命令は、

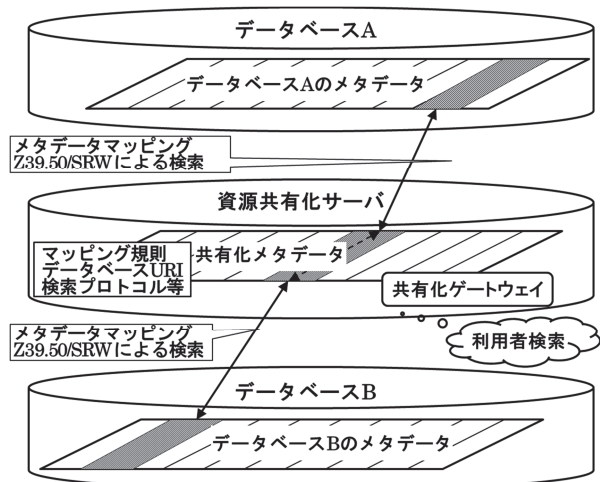


図 6 資源共有化システムの構造

各データベースのマッピング規則に基づいて変換され、指定された検索プロトコルに従って、指定された URI に送付される。この方法により、資源共有化システムでは異なるデータベースシステムの共有を実現している。本稿執筆時点において、地域研の資源共有化システムは 37 データベース(地域研および東南アジア研究所(21)、総合地球環境学研究所(5)、国立民族学博物館(8)および OPAC(地域研、京都大学東南アジア研究所、北海道大学スラブ・ユーラシア研究センター)を共有化し、さらに国内外地域研究機関との共有化の拡大を目指している。

#### 4.2 資源共有化システムの検索例

資源共有化システムには一般公開されている通常版[9]と、試行中(非公開)の多言語版の 2 種類がある。資源共有化システムに共有されている各データベースは、英語・タイ語・ロシア語・

マレー語などの言語で記述されている。そのためシステム的には共有されていても、実際に特定の言語で検索できるデータベースは限定されてしまう。この問題を解決するために、入力された検索語彙を各データベースの言語に変換する機能を追加した多言語版の資源共有化システムを開発した。多言語間の語彙変換には言語グリッド[10]が提供している

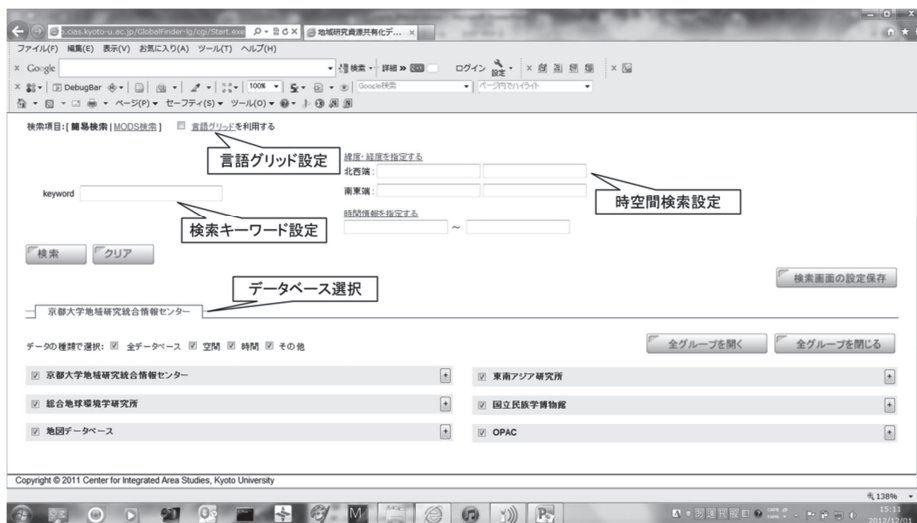


図 7 資源共有化システムの検索画面



検索結果	言語	機関
<input checked="" type="checkbox"/> 京都大学地域研究統合情報センター		
データベース名	英語	地域研
<input checked="" type="checkbox"/> 53件:英国議会資料地図データベース	英語	地域研
<input checked="" type="checkbox"/> 0件:タミル映画データベース	マレー語	地域研
<input checked="" type="checkbox"/> 0件:『カラム』雑誌記事データベース	インドネシア語	地域研
<input checked="" type="checkbox"/> 1件:Waktuデータベース	ロシア語	地域研
<input checked="" type="checkbox"/> 66件:トルキスタン集成データベース	日本語	地域研
<input checked="" type="checkbox"/> 0件:ポスト社会主義諸国選挙・政党データベース	マレー語	地域研
<input checked="" type="checkbox"/> 0件:マレーシア映画データベース	タイ語	地域研
<input checked="" type="checkbox"/> 0件:タイ映画データベース	日本語	地域研
<input checked="" type="checkbox"/> 3件:NEARDB:20世紀年表データベース(1918~1952年)	中国語	地域研
<input checked="" type="checkbox"/> 0件:NEARDB『北京特別市公署市政公報』目次検索データベース(1938~1944年)	英語	地域研
<input checked="" type="checkbox"/> 0件:NEARDB:上海租界工部局警務処文書件名索引データベース(1894年~1949年)	日本語	地域研
<input checked="" type="checkbox"/> 1件:NEARDB:スタンフォード大学フーワー研究所中国関係アーカイブ件名索引データベース	エストニア語	地域研
<input checked="" type="checkbox"/> 0件:NEARDB:モンゴル(人民共和)国科学アカデミー刊行人文社会科学学術定期刊行物記事索引データベース	日本語	地域研
<input checked="" type="checkbox"/> 0件:満洲国ポスターデータベース	日本語	地域研
<input checked="" type="checkbox"/> 1件:藍前期東アジア絵はがきデータベース	タイ語	地域研
<input checked="" type="checkbox"/> 24件:三印法典データベース		
検索結果 149件		

図 8 資源共有化システムの検索結果画面例

Web サービスを利用している。多言語版資源共有化システムの簡易検索画面を図 7 に示す。

ここで言語グリッドをセットし、キーワードを「運河」として検索すると、データベースごとの検索結果が表示される(図 8:ここでは地域研データベースの部分のみを示す)。日本語キーワードによる検索であっても、英語で記述されているデータベースからも検索結果が得られている。例えば英語で記述されている「英国議会資料地図データベース」を選択すると、検索結果一覧が表示される(図 9)。これより、「運河」を「canal」に翻訳してから検索していることが分かる。さらに最初のレコードを選択すると、検索結果の詳細検索画面を経て、地図画像を表示することができる(図 10)。

No.	title/info	genre	abstract	オリジナルDB	機関	データベース名
1	SECOND REPORT OF THE COMMISSIONERS for making and maintaining an Inland Navigation, commonly called The Caledonian Canal, from the Eastern to the Western Sea, by INVERNESS and FORT WILLIAM, in SCOTLAND./Ordered to be printed 30th May 1805.	Commons	GENERAL MAP OF THE INTENDED CALEDONIAN CANAL or INLAND NAVIGATION, From the EASTERN to the WESTERN SEA, BY INVERNESS AND FORT WILLIAM.	<a href="#">表示設定初期化</a>	地域研	英国議会資料地図データベース
2	THIRD REPORT OF THE COMMISSIONERS for making and maintaining an Inland Navigation, commonly called The Caledonian Canal, from the Eastern to the Western Sea, by INVERNESS and FORT WILLIAM, in Scotland./Ordered to be printed 2d June 1806.	Commons	GENERAL MAP OF THE INTENDED CALEDONIAN CANAL or INLAND NAVIGATION, From the EASTERN to the WESTERN SEA, BY INVERNESS AND FORT WILLIAM.	<a href="#">表示設定初期化</a>	地域研	英国議会資料地図データベース
	FOURTH REPORT OF the COMMISSIONERS for		PLAN of the			

図 9 資源共有化システムの検索結果画面例

### 5. より高度な情報システムを目指して

My データベース, API および資源共有化システムにより, 研究者 DB の構築, データベースの活用およびデータベースの共有化という当初の目的は達成しつつある。しかし資源共有化システムには以下の問題点が指摘されている。

- ① 共有化情報がサーバ側で管理されているため, 利用者によるマッピング規則の書換えや共有化するデータベースの追加ができない。

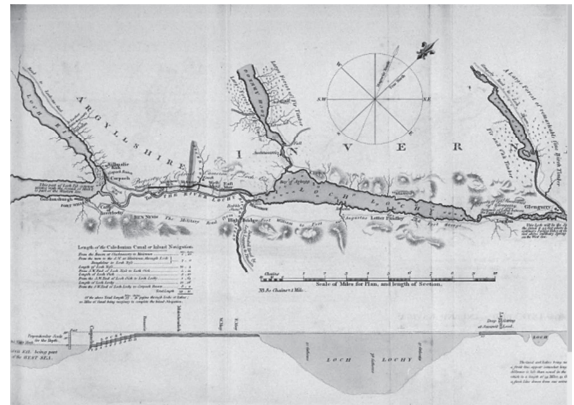


図 10 資源共有化システムの画像表示画面例

- ② 全てをサーバが実行しているため, サーバの構築・管理・運用コストが高い

さらに My データベースにも以下のような問題が指摘されている。

- ① 柔軟なデータベース共有が困難である: My データベースではデータベースを結合する join 操作ができないので, 同様の操作を API により手続き的に記述する必要がある。
- ② 人文科学分野では一般的なデータ記述や操作を素直に記述できない: My データベースでは, 家系図・人間関係・センサーなど, 人文科学分野では一般的なネットワーク型のデータを記述できない。また系譜を辿ったり(祖先や子孫を辿る)関係者を探し出したりする(仲間の仲間の...)操作を API で手続き的に記述することも容易ではない。XML を利用すれば, My データベースでもネットワーク構造を記述できる。しかしネットワークの接続などは困難である。

そこで, ネットワーク型データに対応した新しいデータベースシステムの研究を開始した。ネットワーク型のデータモデルは数多く提案されているが, ここでは RDF (Resource Description Framework) [11,12]を採用した。W3C の標準であること, Web 上にある情報資源の記述に適していること, ツールやアプリケーションが豊富であり応用実績が多いことを評価した結果である。

この研究では, 人間文化研究機構[13]で開発している地名辞書データベースを Web 環境に適したより使い勝手の良い情報資源とすることを目指して, データモデルの設計, 語彙の定義, RDF Triple Store (データベース) および SPARQL Endpoint (検索用アクセスポイント) [14]の構築を進めている。この地名辞書には, 地名を行政地名・建物などの属性で分類している, 地名を町村・郡市・県(歴史地名の場合は, 郷・郡・国など)に従って階層的に関連づけている, 出典情報

(地図)への参照を持っているなどの特徴がある。これらの特徴は構築した RDF Triple Store でも踏襲している。RDF Triple で「阿位郷」を記述した例を以下に示す。なお、語彙や URI は暫定的なものである。

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix dcq: <http://purl.org/dc/terms/>.
.....
<http://gazetteer.chikyu.ac.jp/id/placename/1000001>
  a geo:SpatialThing;
  rdfs:label "阿位郷";
  gzt:country "出雲";
  gzt:county "仁多郡";
  gzt:countyAltLabel "ニタ";
  gzt:locality "仁多郡阿位郷";
  gzt:localityOriginalNotation "仁多郡阿位郷";
  gzt:projection
    [ rdfs:label "1";
      owl:sameAs <http://gazetteer.chikyu.ac.jp/pn/pntype/Point>
    ];
  dc:format
    [ rdfs:label "1";
      owl:sameAs <http://gazetteer.chikyu.ac.jp/pn/pntype/Point>
    ];
  dc:identifier "1000001";
  dc:source "大日本地名辞書";
  dc:subject
    [ rdfs:label "郷・里";
      dcq:subject <http://gazetteer.chikyu.ac.jp/id/placeattribute/6>
    ];
  geo:Point
    [ gzt:lat "35.13";
      gzt:long "132.94";
      gzt:x "3889169.23606705";
      gzt:y "312534.868129257";
      gzt:zone "53 S";
      geo:lat "35.1279467847798";
      geo:long "132.942567549788"
    ];
  skos:altLabel "アイ".
```

RDF Triple Store の検索には SPARQL を利用する。SPARQL クエリを受け付ける API の URI を SPARQL Endpoint という。以下は、「相国寺と同じ国に属する地名」を表示する SPARQL クエリの例である。

```
select distinct ?place ?name
  from <http://gazetteer.chikyu.ac.jp/rdf/placename>
 where {
   {s rdfs:label "相国寺"}
   {s gzt:country ?country}
   {?place gzt:country ?country}
   {?place rdfs:label ?name}
 }
```

この結果は以下ようになる。

```
http://gazetteer.chikyu.ac.jp/id/placename/10023952 朱雀野
http://gazetteer.chikyu.ac.jp/id/placename/10042178 真葛原
http://gazetteer.chikyu.ac.jp/id/placename/10003404 池坊
.....
```

RDF Triple Store および SPARQL Endpoint の

プロトタイプができあがり、その有効性の検証を始めている。

## 6. まとめ

研究資料の保存・公開・共有・利活用を目指して、My データベース・API・資源共有化システムの開発を進め、当初の目的は達成しつつある。しかし研究者支援ツールとして、より柔軟なデータ記述・共有化・利活用を実現するために、RDF に基づいた新しい情報基盤の開発に着手した。プロトタイプができあがり、その有効性の検証を始めている。

## 参考文献

- [1] 原正一郎：資源共有化システムの機能拡張に関する試案, 人文科学とコンピュータシンポジウム論文集, Vol.2011, No.8, pp.147-154 (2011).
- [2] W3C : SOAP Version 1.2 Part 0: Primer (Second Edition) W3C Recommendation 27 April 2007, <http://www.w3.org/TR/soap12-part0/>
- [3] Bill Burke : RESTful Java with JAX-RS, O'Reilly Media (2009).
- [4] D. Crockford : RFC 4627: <http://www.rfc-editor.org/rfc/rfc4627.txt> (2006).
- [5] LOC : CQL Contextual Query Language SRU Version 1.2 Specifications, <http://www.loc.gov/standards/sru/specs/cql.html>
- [6] LOC : Search/Retrieval via URL, <http://www.loc.gov/standards/sru/>
- [7] 柳澤雅之准：出版と提携した京大フィールド・データベースの公開, <http://www.cias.kyoto-u.ac.jp/news/?p=26>
- [8] 原正一郎：地域研究のための資源共有化システムとメタデータに関する研究, 東南アジア 研究, 東南アジア研究所, Vol.46, No.4, pp.608-645(2009).
- [9] 京都大学地域研究統合情報センター地域研究資源共有化データベース : <http://app.cias.kyoto-u.ac.jp/GlobalFinder/cgi/Start.exe>
- [10] 言語グリッド : <http://langrid.org/jp/>
- [11] W3C: Resource Description Framework (RDF), <http://www.w3.org/RDF/> (2014).
- [12] トム・ヒース, クリスチャン・バイツァー (著), 武田英明 (監訳) : Linked Open Data, 近代科学社 (2013) .
- [13] 人間文化研究機構 : <http://int.nihu.jp/>
- [14] W3C : SPARQL 1.1 Query Language, <http://www.w3.org/TR/sparql11-query/> (2013).