

学術情報システムの機能拡張について —京都大学地域研究統合情報センターの試み—

原正一郎^{†1}

地域研究統合情報センターでは、地域研究を支援する学術情報システムとして地域研究情報基盤の研究開発を継続している。地域研究情報基盤を構成するデータベース群は、サーバ機能として設計されているため、検索手順やユーザインタフェースを利用者のニーズに応じて変更することは困難である。そのため研究用情報ツールとしては柔軟性が低く、地域研究情報基盤の利活用を考えている研究者が抱く不満の一つとなっている。

最近の WEB および XML 技術を利用することにより、この状況の改善をめざした取り組みを行っている。それらのうち、本稿では、研究者自身がデータベースを容易に構築できることを目指した「Myデータベース」と、Myデータベースにおいて柔軟な検索機能を実現するための REST 形式の API について述べる。

Expansion of Academic Information Systems - Case Study of Center for Integrated Area Studies, Kyoto University -

SHOICHIRO HARA^{†1}

The Center for Integrated Area Studies Kyoto University has developed the Information Infrastructure for Area Studies (IIAS). As databases of IIAS have designed as server-side functions, researchers are difficult to modify retrieval logics and user interfaces by their purposes. This low flexibility of the system is a cause of dissatisfaction of researchers who want to use IIAS. We are trying to improve this situation by introducing latest WEB and XML technologies. Among our trials, this paper will describe two issues; My-Database that is designed for researchers easily to develop and modify databases by themselves, and REST-like API to give My-Database flexible retrieval functions.

1. はじめに

京都大学地域研究統合情報センター（以下、地域研）[1]では、地域研究に関する資料収集・蓄積・公開・共有・知識処理を支援する学術情報システムとして、「地域研究情報基盤」の研究開発を継続している。しかし、地域研究情報基盤を構成するデータベースはサーバ機能であり、データ蓄積・検索・ユーザインタフェース等の制御は全てサーバ側で行っている。そのため、データ構造・検索手順・画面構成等に対する制限が強いうえに、サービスの受容者である研究者のニーズに迅速に応じることが困難であり、これらが地域研究情報基盤の利活用を考えている研究者の不満の一つとなっている。上記の問題を改善する手法として、地域研では、最新の Web と XML(eXtensible Markup Language) 技術を利用した Web サービスの利用を試みている。

本稿では、地域研の地域研究情報基盤を事例として取り上げ、研究者指向の情報システムの備えるべき機能について考える。まず第2章において、地域研究と地域研究情報基盤の関連およびシステムの課題についてまとめる。次に第3章では、研究者自身がデータベースを容易に構築できることを目指した「Myデータベース」について述べる。第4章では、Myデータベースにおいて柔軟な検索機能を実現するための REST (Representational State Transfer)形式の API(Application Program Interface)と、その事例について

述べる。最後に第5章において、地域研究情報基盤の残された課題と今後の展開等について考察する。

2. 地域研究と地域研究情報基盤

地域研究は、地域の間人活動を様々な視点から総合的に理解するための学問であり、歴史・地理・政治・経済学・文化人類・環境等の研究領域を包摂している。地域を理解するためには、多くの専門家がこれまで努力して蓄積してきた膨大な研究成果を共有する必要がある。しかし、概念・語彙・方法論等の異なる史資料や知識を共有化するのは容易ではなく、これが地域の総合的な理解を阻む大きな要因の一つであったと考えられる。

地域情報学は、地域研究に関する情報の共有、地域の分析と理解、さらに地域の知の創生をめざす情報学的パラダイムである[2]。地域研は地域情報学の構築をミッションとしてかかげ、研究を継続している。地域研究情報基盤は、地域研における地域情報学の研究成果であり、地域研究を情報学の立場から支援する情報システムである。

地域研究、情報学および地域研究情報基盤の関連を俯瞰するために、地域研究のプロセスをデータ・情報・知識のフローという観点から整理した地域情報学モデルを図1に示す[3]。

(1) 資料の収集・整理

地域研究はフィールドから始まる。フィールドではインタビュー・観察・資料調査などの研究活動が展開される。

^{†1} 京都大学地域研究統合情報センター
Center for Integrated Area Studies, Kyoto University

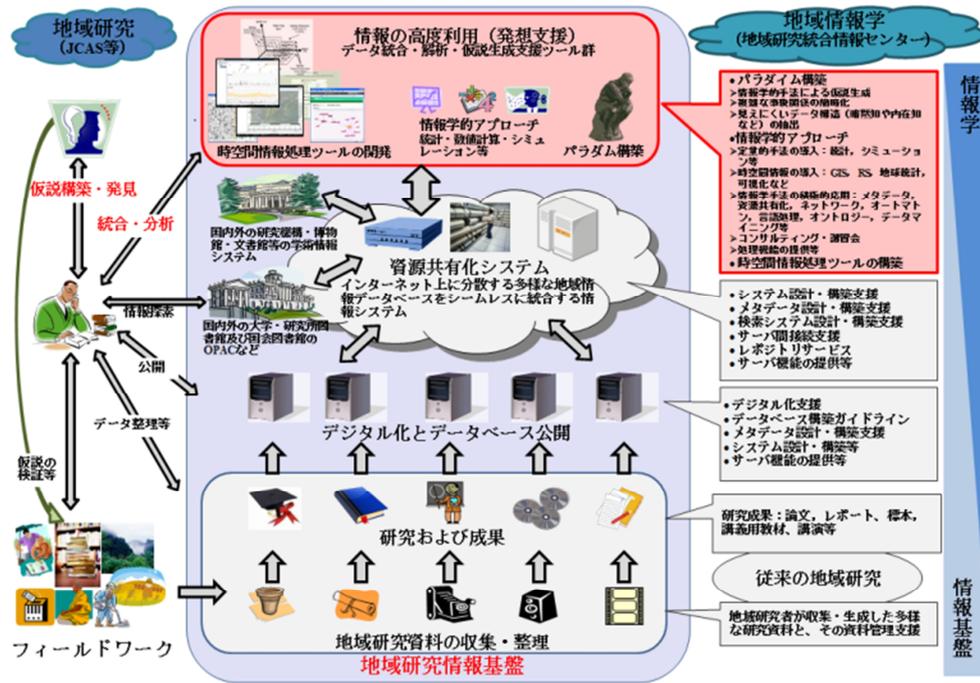


図 1 地域情報学モデル (地域研究と地域研究情報基盤の関係)

Figure 1 Area Informatics Model (Structure of Information Infrastructure for Area Studies).

その成果として、静止画・動画・音声・メモ・標本・文献・数値データなどの多種多量の資料が収集・生成される。資料は研究者あるいは研究領域の手法に基づき整理・分析される。研究成果は論文や書籍等の形態で公開され、フィールドにもフィードバックされる。これが従来の地域研究のフローであると考えられる。

(2) 資料のデジタル化とデータベース公開

近年のデジタル機器の普及により、デジタル撮影やデジタル録音等によるポーンデジタル資料が増えつつある。また資料の可視化や統計分析を実行するためには、資料をデジタル化しなければならない。これらの理由により、地域研究においても資料のデジタル化とデータベース公開は進みつつある。

資料は、適切なメディアモデルとデータモデルに基づいて、データベースとして組織化される。メディアモデルは、資料のメディア特性に応じたデジタル化の手法を決定する。データモデルは、資料管理と検索に適したメタデータ的设计、つまりデータ項目・粒度・言語・符号化法と記述規則を定義する。これらのモデルに従って資料のデジタル化とメタデータを作成する。デジタルデータとメタデータからデータベースを構築する。データベースは言葉通りに地域研究のベースであり、その設計と実装の良し悪しが以降の情報処理の速度・精度・利用法などに影響を与える。本稿執筆時点において、地域研では35データベースが公開中あるいは公開に向けて開発中である[4]。

(3) 資源共有化システム

これらの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、図書館等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。コレクションとして完結させるためには、関連データベースを結合する必要がある。さらに研究の過程では、複数のデータベースを利用する場合も多いので、データベースの連携機能があれば便利である。インターネット上のデータベースを結合させたり連携させたりする機能を、以下では共有化と呼ぶ。

共有化を利用者の側から眺めると、一回の検索で複数のデータベースが同時に検索できるということである。

これ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。大学図書館 OPAC(Online Public Access Catalog)等は、この手法を採用している。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。また、多様な研究領域を包摂する地域研究においては、単一メタデータを定義すること自体が困難である。そのため地域研の共有化では、データベース構築に際して特定のメタデータを強制しない。そのかわり、共有化メタデータと称する、各データベースのメタデータから独立した単一メタデータを定義する。個別メタデータと共有化メタデータの要素間には予め関連づけを設定する。共有化検索時は、個別メタデータではなく、共有化メタデータを検索する。これにより、異なるデータベースをシームレス(seamless)に接合することを実現している。このような仕組で共有化を実現している情報システムを、資源共有化システムと呼んでいる[5]。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, 東南アジア研究所, 北海道大学スラブ研究センター))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[6]。

(4) 情報の高度利用

地域情報学の目標はデータベースの構築と共有化にとどまるものではなく、共有化されたデータに情報学の手法を駆使して、データを検索し、それらを統合・解析し、新しい仮説形成や知識発見を試みることである。具体的には可視化・語彙分析・統計・時空間などの基本的なデータ分

析、データ間の関連付け、分類・因果関係の抽出・予測などを行う。

とりわけ語彙は研究分野ごとに異なっているため、資源共有化システムを駆使したデータ検索や統合を効果的に実行するためには、語彙の違いを乗り越えた知識の枠組みが必要である。その枠組みとしてオントロジーに注目している。地域情報学では、多様な地域研究情報を関連づける語彙として場所と時間が重要と考えており、地名語彙に関してはデジタル地名辞書[7]、暦日については暦日テーブル[8]を構築している（歴史地名の性質上、デジタル地名辞書へのアクセスは制限されている）。また基本学術語彙に関する試みとして日本図書協会基本件名標目標[9]と国立国会図書館件名標目標[10]、さらに AIMS(Agricultural Information Management Standards)の農業関連語彙集 AGRVOC[11]を Topic Mas の Web アプリケーションとして公開している。

地域研究では、対象の空間的・時間的関連が特に重要である。そのため、地域研では時空間に注目した情報処理手法の研究を積極的に進めている。また時空間情報処理ツールとして、HuMap という空間情報処理ツールと HuTime という時間情報処理ツールの開発も進めている[12]。

研究成果は再びフィールドへフィードバックされ、新しい研究プロセスが始まる。地域研究情報基盤は、上記(1)から(4)の各研究プロセスを支援するための情報システムおよびツールの総体である。構想していた地域研究情報基盤はほぼできあがりつつあるが、研究現場への普及を図る上で解決しなければならない問題点も明らかになってきた。以下では、特にデータベースについての課題と取り組み整理する。

(5) データベースシステムの課題と取り組み

地域研究情報基盤のデータベースは機関レポジトリとして設計されている。そのため、管理が複雑なうえに利用

者ごとの要求への柔軟な対応が困難で、研究用情報ツールとしては使いにくいという意見が多かった。そこで地域研では、地域研究情報基盤を研究用情報ツールとすることをめざして、以下の三つの課題に取り組んでいる。

一つ目は My データベースと呼ぶ個人データベース機能である。地域研究の現場では多様な資料が収集され、それらの多くはスプレッドシートなどを使って整理されており、データベース化の一手前までの準備ができています。しかし地域研究は基本的に人文系研究セクターであり、データベースシステムを構築・管理・維持できる人材や資金に乏しいため、データベース公開は進んでいない。一方、地域研究情報基盤は、①管理・運用が複雑である、②共有化を念頭においているために標準メタデータをベースとしたメタデータ設計を要求している、③メタデータ定義の頻繁な修正に対応しにくいなど、研究用としては柔軟性に欠けている。そこで My データベースでは、データベースシステムの管理・運用法を見直して、研究者個人による操作が簡単に行えるようにした。またデータベース構築用の設定情報や関連ファイルをコンパクトにして、メタデータの定義・修正、検索機能の設定、検索画面の作成などを簡単に行えるようにした。

二つ目は、My データベースにおいて柔軟な検索機能を実現するために開発した REST 形式の API である。資源共有化システムは横断検索システムと呼ばれることがある。これは、資源共有化システムが共有化メタデータを介したメタデータ語彙の名寄せという検索手法をとっている、つまり各データベースの意味的に同じデータ要素を繋いでいるようなイメージがあるからと想像している。しかし研究者の検索ニーズは語彙の名寄せ程度では実現できないものが殆どである。例えば、あるデータベースから地名を検索し、それを地名辞書で緯度・経度に変換し、その地点を中心とした範囲を設定し、その範囲内の情報を別のデータベ

ースから抽出するような場合である。横断検索機能に対して、複数のデータベースを逐次検索する機能を縦型検索機能と呼ぶことにする（これは Web2.0 で普及している mashup とほぼ同じ機能であるが、本稿では縦型検索機能とする）。REST 形式の API は縦型検索を実現するために導入した手法である。

三つ目は多語化である。資源共有化システムが統合しているデータベースは英語、タイ語、ロシア語、マレー語など多様な言語で記述されている。そのためシステム的には共有化されていても、同時に検索できるデータベースは同一言語のものに限

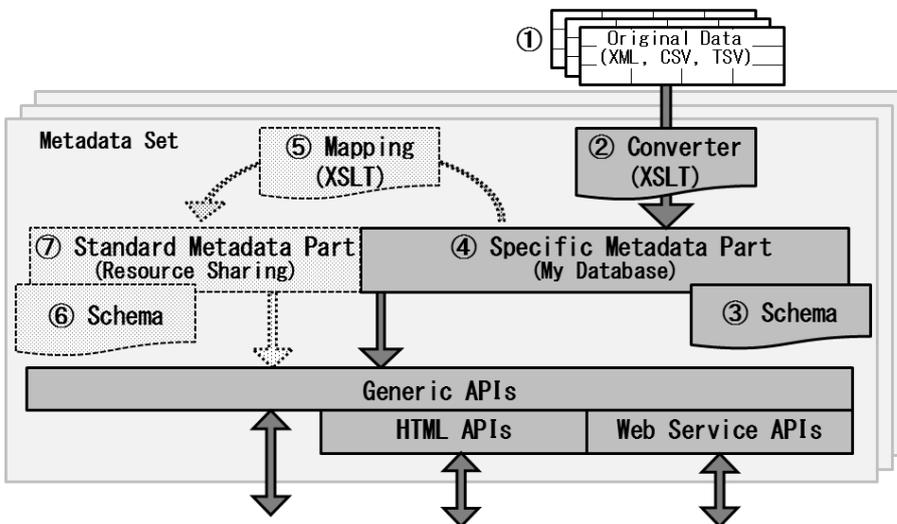


図 2 My データベースの構造
Figure 2 The Structure of My-Database.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	地震名	地震ID	史料名	レコード番号	レコードID	和暦	西暦	記載地名	現在地名	LAT	LOIN	ノート
2	1	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/宮城	京都府/京都市/上京区/小山西町	35.015	135.7452778	千本丸太町交差点近くに大極殿基壇跡の石碑(平成6年あり)
3	2	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/宮城/八省院	京都府/京都市/上京区/小山西町	35.015	135.7452778	千本丸太町交差点近くに朝堂院(=八省院)跡のプレートあり
4	3	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/豊楽院	京都府/京都市/中京区/聚楽廻西町	35.015	135.7452778	平安宮豊楽殿跡
5	4	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/東寺	京都府/京都市/南区/九条町	34.97777778	135.7494444	左は東寺(教王護国寺)の現在地
6	5	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/西寺	京都府/京都市/南区/唐橋平垣町	34.97861111	135.7386111	左は西寺の現在地
7	6	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/清水寺	京都府/京都市/東山区/清水一丁目	34.99138889	135.7869444	左は清水寺の現在地
8	7	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/円覚寺	京都府/京都市/上京区/柳風呂町	35.02638889	135.7694444	左は円覚寺の現在地

図 3 データサンプル例

Figure 3 Example Dataset.

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多言語システムの開発を進めている。

これら三つの取り組みのうち、以下ではMyデータベース機能と REST 形式の API について説明する。

3. My データベース

Myデータベースの概要を図2に示す。登録可能なデータ形式は XML, CSV(Comma Separated Values), TSV(Tab Separated Values)の3種類である。CSV および TSV データの場合は関係データモデルの第一正規形, XML データの場合は整形形式という条件を満たせば、フィールドあるいはエレメントの名称や出現順などについての制限はない。なおシステム内部は XML データ形式であり、検索には XML 検索エンジンを利用しているため、CSV および TSV データは登録初期段階で中間形式の XML データに変換される。

登録された XML データから検索用と表示用の XML データを生成する。資源共有化用の XML データを生成するこ

ともできる。

Myデータベースでは3種類の検索問合せ方法を用意している。一つ目はMyデータベース固有の問い合わせ言語である。MyデータベースのXML検索エンジンにはOpenTextという商用ツール採用しているため、検索プログラムはOpenText専用の問合せ言語で記述されている。ただしセキュリティ上の理由から、一般利用者レベルからのアクセスはできないようになっている。

二つ目の問い合わせ方法は、HTML(Hyper Text Markup Language)

のFORMを利用するもので、Myデータベースでは通常の使い方である。この方法では、クライアントからの問い合わせをアプリケーションサーバにおいてOpenText用の問合せ手順に変換してから検索を行う。OpenTextからの返戻データはXMLであるため、HTMLに変換してからクライアントに返す。XMLからHTMLへの変換にはXSLTを利用している。

三つ目の方法は、HTTPのGETあるいはPOSTメソッドを使い、XMLデータを返戻として利用するものである。これについては次章で説明する。

以下では、CSVデータからMyデータベースを構築する手順の概要を述べる。

(1) データの準備

EXCEL等のスプレッドシートやエディタを利用してデータファイルを作成する。その例を図3に示す。これは歴史地震データのメタデータをEXCELにより作成したものの一部である。文字コードはShift-JISでも構わないが、多言語対応やシステム間のデータ共有を考慮するならUTF-8とすることが望ましい。

図のテーブルの1行目はヘッダ行でフィールド名のリストである。例えば、1列目はIDでレコードの主キーとなっている。また11行目は緯度、12行目は経度である。2行目以降はデータレコードである。EXCELで作成しているが、データ構造としては関係データモデルの要件を満たす必要がある。

(2) 登録用データの作成

作成したデータファイルに制御情報を指定する行を加えて、Myデータベース用の登録データファイルを作成する。具体的

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	地震名	地震ID	史料名	レコード番号	レコードID	和暦	西暦	記載地名	現在地名	LAT	LOIN	ノート
2	SID	ENAME	EID	RNAME	RID	RECID	JDATE	NDATE	ONAME	PNAME	LAT	LOIN	NOTE
3	ID	BASIC	PRIME	BASIC	BASIC	BASIC	BASIC	BASIC	BASIC	BASIC	BASIC	BASIC	BASIC
4	1	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/宮城	京都府/京都市/上京区/小山西町	35.015	135.7452778	千本丸太町交差点近くに大極殿基壇跡の石碑(平成6年あり)
5	2	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/宮城/八省院	京都府/京都市/上京区/小山西町	35.015	135.7452778	千本丸太町交差点近くに朝堂院(=八省院)跡のプレートあり
6	3	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/豊楽院	京都府/京都市/中京区/聚楽廻西町	35.015	135.7452778	平安宮豊楽殿跡
7	4	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/東寺	京都府/京都市/南区/九条町	34.97777778	135.7494444	左は東寺(教王護国寺)の現在地
8	5	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/西寺	京都府/京都市/南区/唐橋平垣町	34.97861111	135.7386111	左は西寺の現在地
9	6	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/清水寺	京都府/京都市/東山区/清水一丁目	34.99138889	135.7869444	左は清水寺の現在地
10	7	貞元元年六月十八日	9760722	〔日本紀略〕	22	1151	貞元元年六月十八日癸丑	9760722	山城国/京都/円覚寺	京都府/京都市/上京区/柳風呂町	35.02638889	135.7694444	左は円覚寺の現在地

図 4 Myデータベース用の入力データ例

Figure 4 Example Import Dataset for My-Database.

	詳細検索画面	一覧表示画面	並べ替え項目	詳細表示画面
PRIME	○	○(リンク表示)	○	○
BASIC	○	○	○	○
SEARCH	○		○	○
LINK			○	○(リンク表示)
DATA				○
THUM		○(サムネイル表示)		
IMAGE				○(画像表示)
ID			○	○
PCONTENTS				
SEARCH_ONLY	○			
COMMENT				
PRIVATE				
IMAGE_XML		○(画像表示)		○(画像表示)
LINK_XML				○
RACLGROUP				

図 5 My データベースの属性パラメータと機能の概要

Figure 5 The List of Attribute Parameters and Functions of My-Database.

にはヘッダ行の下に 2 行を加えて図 4 のようにする。ここで 2 行目はヘッダの別言語表記である。多くの場合、1 行目は原語表記、2 行目は英語表記となる。2 行目の値は空でもよい。3 行目は各フィールドの属性パラメータを設定する。属性パラメータにより、検索や検索結果表示の際の各フィールドの機能を指定する。属性パラメータと機能の概要を図 5 に示す。

(3) 利用者情報の登録

管理者に依頼して、My データベースの利用者情報を登録する。それ以降は、該当する My データベースの管理者の資格でシステムにログインして作業を行う。

(4) データ登録

利用者登録されたデータベースの URI(Uniform Resource Identifier), 利用者 ID およびパスワードを使って My データベースへログインする。それ以降は管理画面の指示に従って操作を行う。最初にデータベース名を定義し、引き続きデータの登録を行う。登録したデータは、レコード単位で修正することができる。差分ファイルによる修正やファイルの入れ替えも可能である。

(5) 検索機能の設定

検索項目、簡易表示項目、詳細表示項目、画像等コンテンツファイルへのパス設定、時空間情報利用の有無、サムネイル表示法など、検索や表示機能に関する設定を行う。

(6) 画面の設定

My データベースの各画面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。管理画面より、検索画面を構成するヘッダ領域、データベース領域、フッタ領域などのレイアウトや背景色・画像などを変更できる。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。

(7) 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、利用者グループ、一般利用者となっている。IP アドレスによるアクセス制限を設けることもできる。

My データベースの構築作業にかかる時間は、登録するデータ量と、どの程度の設定を行うかにより異なる。ただしデータファイルが完備していて、デフォルトの設定をそのまま使うのであれば 1 時間もかからない。図 6 に図 4 のデータファイルから構築した My データベースの検索画面の例を示す。

My データベースには幾つかの便利な検索機能が用意されている。例えば検索語の入力には

テキストボックスを利用するが、主題語のように検索語彙数が限定されている場合、チェックボックスやプルダウン等を使って選択できたら便利である。このように検索条件の入力を簡略化する機能は、登録用データファイルの 3 行目の属性値に属性パラメータを加えることで実現できる(図 4)。例えば 2 列目の地震名フィールドに

`input_type=dropdown&input_order=asc`

属性パラメータを加えると、データ登録時に該当フィールドを走査して語彙リストを生成し、プルダウンリストとして表示する。図 6 の検索画面は、このプルダウン機能を利用した例でもある。

また多用される検索フィールドについては、予め検索を実行して、結果をカテゴリに分類しておくとも便利な場合がある。これを My データベースではカテゴリ検索機能と呼

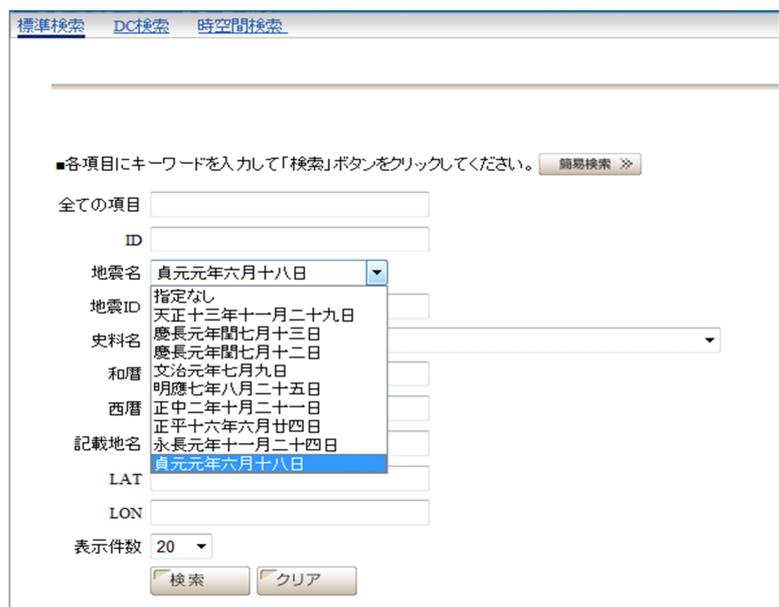


図 6 My データベースの検索画面例

Figure 6 Example Display of Data Retrieval by My-Database.

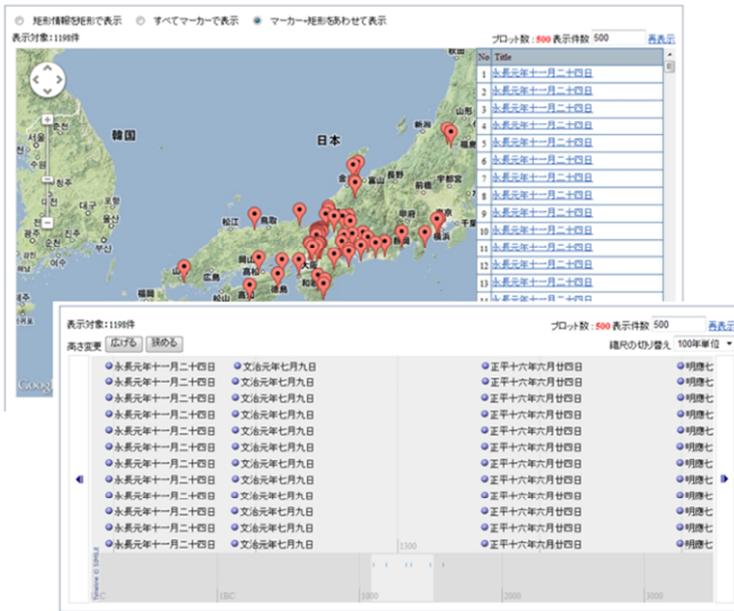


図 7 My データベースの時空間表示例

Figure 7 Example Display of Spatiotemporal Data by My-Database.

んでいる。カテゴリは階層化できる。この機能もカテゴリ化したフィールドに属性パラメータを加えることで実現できる。

さらに緯度・経度や時間に関するフィールドがあれば、地図やタイムスライダを利用した検索・表示機能を利用することもできる。図 7 に検索結果を地図とタイムスライダで可視化した例を示す。

4. 縦型検索機能と REST 形式の API

地域研究情報基盤を構成するデータベースは、サーバ機能として設計された(図 8 左)。利用者が発行した要求や表示の制御(図の Controller)、データベース検索ロジック(図の Model)、表示画面の生成(図の View)といった全ての処理はサーバが実行する。利用者のブラウザは単なる表示装置にすぎない。

縦型検索では、前段のデータベース検索結果を分析し、その結果に応じて後段の処理を適切に選択するロジックが必要である。ロジックは利用者・利用目的ごとに異なるので、この機能は利用者コンピュータに実装する必要がある。また縦型検索では、異なるプラットフォーム上のデータベースサービスを組み合わせる必要がある。つまり縦型検索機能は、利用者コンピュータに Model, View, Controller のプログラムを実装した、分散コンピューティング環境ということが出来る(図 8 右)。

Web 環境下において、プラットフォームに依存せず、しかも利用者に大きな負担をかけずに RPC (Remote Procedure Call) 機能を実装する適切な手段を見極めることができなかったため、縦型検索機能の実装が遅れることとなった。

地域研究情報基盤の開発を開始した 2006 年ころ、HTTP と XML を組み合わせた SOAP[13]と呼ばれる分散コンピューティング技術(Web サービス)が注目されており、縦型検索の手法として利用することを検討した。しかし仕様が複雑かつ巨大であり、利用者側のプログラミングの負担が大きくなると懸念された。その後、RPC スタイルに合わせた XML+HTTP による REST と呼ばれるアーキテクチャスタイルが我が国でも注目されるようになった[14]。REST には、①URI と HTTP でリソースを容易に操作できる、②初期の学習コストが低い、③利用者側のロジックは HTTP と XML 解析用の DOM(Document Object Model) API を使えば比較的簡単に開発できるなどのメリットがある。そこで地域研究情報基盤では、縦型検索機能を実現するために REST の利用を試みている。

縦型検索機能の実現には二つの設計が必要である。一つ目は利用者側とサーバ側の機能設計である。地域研究情報基盤において、サーバ側は汎用的なデータベース機能(汎用的な Controller, Model および View)を分担し、外部に対してはサーバ自身の所在を示す URI とアクセス手段である API を公開する。利用者側では、目的に応じた Controller, Model および View を実装する。サーバへのアクセスには、サーバ側が公開している URI と API を利用する。

二つ目はサーバ側の API と検索結果の返戻データ構造の設計である。縦型検索機能では REST 形式の Web サービスを利用している。その特徴を要約すると以下ようになる。
 ① データ交換に HTTP の GET あるいは POST を利用する。

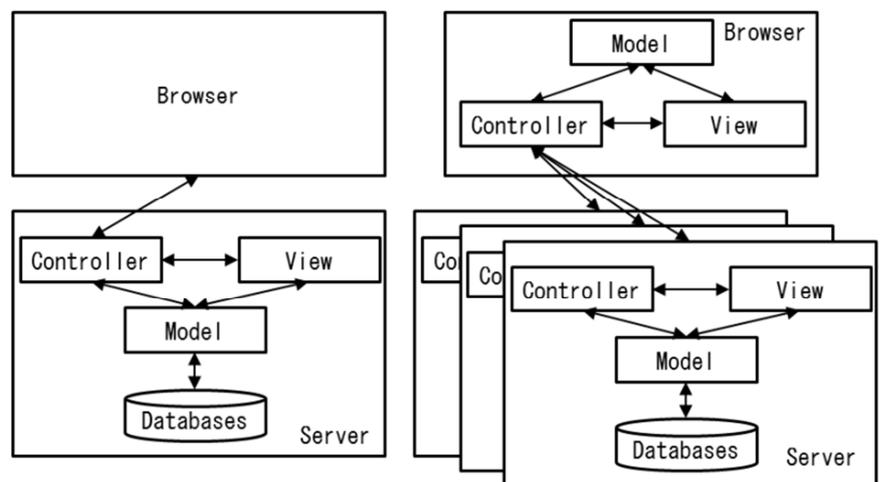


図 8 従来のデータ処理(左)と REST 型データ処理(右)

Figure 8 Conventional Data Processing (left) and REST-like Data Processing (right)

- ② ロジックを記述するプログラムは、CGI あるいは Servlet のクライアントプログラムと似ているので、ユーザ側のプログラミングが容易になると期待される。
- ③ 返戻が XML であるためプログラム実装の自由度が高く、またデータ変換などに際しても既存のツールを利用できる。

地域研究情報基盤の機能拡張では、サーバ側の既存機能には手を加えず、REST 型 API を Web サービス API として新たに加えた (図 2)。利用者側の機能は JavaScript や XSLT などの利用を試みている。

(1) 地域研究情報基盤における REST 形式の API

地域研究情報基盤の REST 型 API では、検索式として CQL(Context Query Language)[15]を採用している。CQL は Web 環境における情報検索用プロトコルである SRU(Search/Retrieve via URL)および SRW(Search Retrieve Web services) [16]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://app.cias.kyoto-u.ac.jp/?operation=searchRetrieve&version=1.2&recordSchema=originalxml&query=(c1="京大")**

の場合、Base URL は **http://app.cias.kyoto-u.ac.jp/****であり、Request Part は?より後ろの部分である。また Request Part のうち **query=**以下の部分が CQL による検索式である。SRU 上の CQL は簡潔かつ直感的な検索式であり、人にも読みやすく、プログラム処理も容易で、そのまま Web ブラウザに入力するだけで動作確認を行うこともできる。

地域研究情報基盤の REST 型 API の返戻形式は XML のみであり、フォーマットとしては DC(Dublin Core Metadata) [17], MODS(Metadata Object Description Schema) [18]および ORIGINAL (インポートされた状態) の 3 種類がある。

(2) 縦型検索の例

縦型検索機能の例として、前述のデジタル地名辞書を取りあげる[7]。地名辞書は、人間文化研究機構[19]と H-GIS 研究会[20]が中心となって開発を進めている、日本の歴史地名に関するデジタル辞書である[21]。地名辞書には歴史地名、歴史地名の包含関係 (郷、郡、国等)、現在地名、現在地名の包含関係、地名属性 (寺院、記念碑、河川等)・緯度・経度等が登録されている。本稿執筆時点で、大日本地名辞書 (49,557 地点)、式内社調査報告書 (2,842 地点)、日本寺院総鑑 (78,588 地点)、旧 5 万分の 1 地形図 (192,122 地点) の 323,109 地点が登録されている。

以下では XSLT を利用した簡単な例を示す。デジタル地名辞書で「相国寺」を検索した結果の返戻を XML で表示すると、以下ようになる。

```
<?xml version="1.0" encoding="UTF-8"?>
<soapenv:Envelope
  xmlns:srw="http://www.loc.gov/zing/srw/" . . . . . >
<soapenv:Body>
<searchRetrieveResponse xmlns="http://www.loc.gov/zing/srw/"
  . . . . . >
  <records>
  <record xmlns="">
  . . . . .
  <recordData>
  <original>
  <item pid="10026682">
  <country>山城</country>
  <county reading="カミキヨウ">上京</county>
  <placename reading="ソウコクジ">相国寺</placename>
  <proma1>so^kokuji</proma1>
  . . . . .
  <pname1>京都市上京区</pname1>
  <shp>1</shp><loc>1</loc>
  <lat>35.2.0</lat><long>135.45.46</long>
  . . . . .
  </item>
  </original></recordData>
  . . . . .
```

同様の検索を XSLT で実行するには、XSLT を起動するための簡単な XML ファイルを用意する。ここから起動される XSLT ファイルにおいて、

```
<xsl:variable name="Place" select="document(Query)/Xpath"/>
```

のように document 関数を利用して XML データを読み込む。ここで、**Query** は SRU 検索式、**Xpath** は読み込んだ XML データの文書要素からの Xpath である。例えば **Xpath** が

```
/Envelope/Body/searchRetrieveResponse/records/record/recordData/original/item
```

であれば (名前空間を省略)、1 つ以上の **item** が **Place** に代入される。この **Place** から地名 ID、地名、国名、郡名を抽出し HTML の TABLE に変換する XSLT を作成すれば、地名一覧表となる。さらに以下のような XSLT を利用すると、

```
<xsl:for-each select="$Place">
  <xsl:variable name="PName">
  <xsl:text>Pos</xsl:text><xsl:value-of select="position()"/>
  </xsl:variable>
  <xsl:variable name="MName">
  <xsl:text>markerOptions</xsl:text><xsl:value-of select="position()"/>
  </xsl:variable>
  <xsl:if test="number(lat2) >= 0">
  var <xsl:value-of select="$PName"/> =
  new google.maps.LatLng(
  <xsl:apply-templates select="lat2"/>,
  <xsl:apply-templates select="long2"/>);
  var <xsl:value-of select="$MName"/> = {
  position : <xsl:value-of select="$PName"/>,
  title : <xsl:text></xsl:text>
  <xsl:apply-templates select="placename"/> (
  <xsl:apply-templates select="lat2"/>,
  <xsl:apply-templates select="long2"/>)<xsl:text></xsl:text>,
  map : map};
  var marker =
  new google.maps.Marker(<xsl:value-of select="$MName"/>);
  </xsl:if>
</xsl:for-each>
```

Place から緯度と経度のペアを抽出して Google Maps に点として表示する JavaScript の一部

```
var Pos2 = new google.maps.LatLng(35.02958418, 135.7566909);
var markerOptions2 = { position : Pos2,
  title : '相国寺(35.02958418,135.7566909)',
  map : map };
var marker = new google.maps.Marker(markerOptions2);
```

が生成される。

以上を表示すると図 9 のようになる。図の上段は地名辞書の検索結果の一覧の TABLE 表現、下段は地名の位置を Google Maps 上に表示したものである。この程度の処理であれば、起動用 XML ファイルと XSL ファイル (デジタル



図 9 REST 型データ処理による地図表示例

Figure 9 Example Map Display by REST-like Data Processing

地名辞書への問い合わせ、返戻 XML の加工およびテーブル表示と地図表示を併せたものをそれぞれ作成して利用者側のパソコンに載せるだけで実現できる。

5. まとめ

地域研究情報基盤を研究用情報ツールとするために、研究者自身が容易に構築できる My データベースと、My データベースにおいて柔軟な検索機能を実現するための REST 形式の API の開発を継続している。

My データベースについては、センター内の研究者を対象としたデータベースの試作を通じ、不具合や改良すべき箇所の検証を進めている。不具合等の解消を進め、My データベース機能を学外の地域研究者に公開する予定である。

REST 形式の API についても並行して検証を進めており、幾つかの利用事例もできている。縦型検索では JavaScript が多用されており、その XMLHttpRequest には特定ドメインのサーバとしか通信できないという制約がある。これをクロスドメイン制約と呼ぶが、そのためにネットワーク上の複数のデータベースを利用するという縦型検索機能の機能が制限されてしまう。これを回避する方策として、JSONP (JSON with padding) を利用する、中継プロキシを設置する (現時点における暫定的解決法)、XMLHttpRequest の代わりに DOM を利用する等が提案されており、検討を進めているところである。

地域研は共同利用・共同研究拠点として、地域研究に関する情報の組織化をミッションの一つとしている。My データベース機能は、情報組織化のツールとして利用可能である。ところで一つの研究センターが、データセンター機能と検索サービス機能を提供しつづけることは困難である。地域研はデータの維持・管理を行うデータセンターを基本機能とし、高度な検索機能は研究者に任せる方針である。

REST 形式の API は、研究者のニーズに応じた検索機能を実現するための地域研究情報基盤の機能と捉えることができる。これらにより地域研究情報基盤の利活用が進むことを期待している。しかしデータベース、XML、プログラミング等に関する情報リテラシは不可欠であり、地域研究情報基盤の普及に向けた教育活動が重要になると考えている。そこで大学院生や研究者を対象としたデータベース講習会を昨年度に開催したが[22]、今後も継続する予定である。

参考文献

- 1) 京都大学地域研究統合情報センター：
<http://www.cias.kyoto-u.ac.jp>
- 2) 柴山守，原正一郎：総論 地域情報学の目指すところー地域研究における GIS の応用，アジア遊学，No.113，pp.28-35(2008).
- 3) Shoichiro Hara: Area Informatics – Concept and Status-, Lecture Note in Computer Science 6259, Springer, pp.214-228(2010).
- 4) 地域研究統合情報センターデータベース：
<http://www.cias.kyoto-u.ac.jp/database/>
- 5) 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).
- 6) 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>
- 7) デジタル地名辞書：
http://area.net.cias.kyoto-u.ac.jp/infolib/meta/CsvDefault.exe?DEF_XSL=default&GRP_ID=G0000016&DB_ID=G0000016pdic&IS_TYPE=csv&IS_STYLE=default
- 8) 暦日テーブル：
http://area.net.cias.kyoto-u.ac.jp/infolib/meta_pub/G0000017wareki_conv
- 9) 日本図書協会基本件名票目標トピックマップ：
<http://infos.net.cias.kyoto-u.ac.jp:8083/bsh1/>
- 10) 国立国会図書館件名票目標トピックマップ：
<http://infos.net.cias.kyoto-u.ac.jp:8083/ndlsh1/>
- 11) AGRVOC トピックマップ：
<http://infos.net.cias.kyoto-u.ac.jp:8083/agrovoc/>
- 12) 原正一郎，関野樹：時空間情報処理ツール HuTime・HuMap の開発と利用，歴史 GIS の地平 景観・環境・地域構造の復原にむけて，HGIS 研究会編，勉誠出版，pp.13-24(2012).
- 13) SOAP Version 1.2 Part 0: Primer (Second Edition) W3C Recommendation 27 April 2007,
<http://www.w3.org/TR/soap12-part0/>
- 14) Roy Thomas Fielding : Architectural Styles and the Design of Network-based Software Architectures,
<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>
- 15) LOC: CQL: Contextual Query Language (SRU Version 1.2 Specifications), <http://www.loc.gov/standards/sru/specs/cql.html>
- 16) LOC: Search/Retrieval via URL, <http://www.loc.gov/standards/sru/>
- 17) LOC: Metadata Object Description Schema,
<http://www.loc.gov/standards/mods/>
- 18) Dublin Core Metadata Initiative: <http://dublincore.org/>
- 19) 人間文化研究機構：<http://www.nihu.jp/>
- 20) H-GIS 研究会：<http://www.h-gis.org>
- 21) 桶谷猪久夫：地図・地名データベースの構築，歴史 GIS の地平 景観・環境・地域構造の復原にむけて，HGIS 研究会編，勉誠出版，pp.79-88(2012).
- 22) 地域研究コンソーシアム：「データベース構築ガイド」開催，http://www.jcas.jp/event/2011/08/post_166.html