

# 人文社会科学データベースの構築に関する考察

原正一郎<sup>†1</sup>

**概要:** 京都大学地域研究統合情報センターが開発した地域研究情報基盤は、人文社会科学分野における研究情報の保存・公開・共有・利活用の促進を目的としている。地域研究情報基盤は地域研究を支援する情報システムとして高く評価されている。しかし、地域研究情報基盤は10年以上前の技術に依っており、高い管理・運営コストや、研究者の多様なニーズに対応できないなどの問題も顕在化してきた。本稿では、これらの問題を考察するとともに、人文社会科学研究者が発信する研究情報の流通と利活用のさらなる促進を目指した新しい情報基盤の開発について述べる。

**キーワード:** 地域研究情報基盤, My データベース, 資源共有化システム, RDF, LOD

## Considerations about Humanities' Databases

SHOIHCIRO HARA<sup>†1</sup>

**Abstract:** The Center for Integrated Area Studies (CIAS), Kyoto University has developed the Information Infrastructure for Area Studies to facilitate preserving, distributing, sharing and utilizing various humanities' research information. This Information Infrastructure is appreciated as information system to support area study. However, as the Information Infrastructure was developed using technique more than ten years ago, some problems, such as the high cost of system maintenance and the difficulties to support various researchers' needs, become clear. This paper considers the background of these problems, and explains development of a new information infrastructure by CIAS.

**Keywords:** Information Infrastructure for Area Studies, My Database, Resource Sharing System, RDF, LOD

### 1. はじめに

多様な学術データベースがインターネット上に公開されているが、それらのメタデータは収集組織や資料の特性により異なる。アーカイブは個人や組織や団体などに関する記録資料の総体を対象とする。これに対して、図書館は一点ごとの書籍、博物館は一点ごとのモノを対象としている。大学では機関リポジトリが普及しつつある。これは大学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。機関リポジトリの機能はアーカイブに近いが、メタデータは図書館に近いものが多い。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は異なっている。その一方で、これらのデータベース（以下、機関DB）は、公開性が高いこと、資料収集や整理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、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。

大学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が、その様相は機関DBとはかなり異なる。研究者DBでは、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特徴がある。メタデー

タの標準化が進んでいない素材も多いため、標準メタデータが使われることは少ない。さらに、研究目的に応じてユーザインタフェースを設計しているため、汎用性や公開性は一般的に低い。そのため、他の研究者にはアクセスできなかったり、使い勝手の悪かったりするデータベースが多い。研究の進展によっては、メタデータやデータベースの設計方針を変更する場合もある。さらに、予算不足のために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、研究資料を安定的に蓄積・利活用でき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。

本稿では、研究資料の安定的な蓄積・利活用に加えて多様な研究応用にも耐えうる研究者DB、特に人文社会科学データベースについて、京都大学地域研究統合情報センター（以下、地域研）[1]の取り組みを中心に考察する。

### 2. 地域研究情報基盤

本章では、研究者DBの公開・利活用を支援する情報システムの事例として、京都大学地域研究統合情報センター（以下、地域研）が運用している地域研究情報基盤について述べる。地域研究情報基盤は、データベース構築支援機

<sup>†1</sup> 京都大学  
Kyoto University

能であるMyデータベースと、インターネット上に分散しているデータベースを統合検索するための資源共有化システムを中心として構築されている[2].

## 2.1 My データベース

データベースの公開には、メタデータ作成に加えて、データベースシステム、サーバ装置、ネットワーク機器の構築、維持、管理が必要である。特に後者については専門知識と継続的な管理コストが必要であり、研究者個人や小規模研究組織にとってデータベースを敷居の高いものとしている。そのため、資料のデジタル化とメタデータ作成まで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。さらにデータベースを公開しても、様々な理由で公開停止に至ったデータベースもかなりの数に上る。永続的なデータ公開を実現するには、システム構築が容易、システム管理が不要、かつコスト負担の少ない情報サービスの提供が必要である。

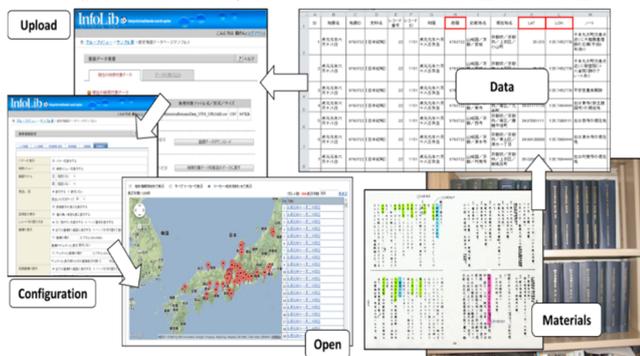


図 1 My データベースの構築手順

Figure 1 Construction Procedure of MyDatabase

### (1) My データベースの構築手順

Myデータベースは、データベースシステムの管理や運用法を見直して、メタデータの定義と修正、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工夫した情報処理システムである。幾つかの条件を満たした CSV ファイル、TSV ファイルあるいは XML ファイルと画像などのデータを用意したら、あとはMyデータベースの指示に従って操作するだけで、データベースの構築と公開を実現できる。Myデータベースの構築手順を以下に示す(図 1)。

- ① データ作成：スプレッドシートやエディタなどを利用してデータファイルを作成する
- ② データ登録：データファイルを登録画面からアップロードする
- ③ データ属性設定：コンテンツファイルへのパス、時空間情報の有無、標準メタデータへのマッピングなどを設定画面に従って指定する
- ④ 画面設定：各画面のレイアウト、背景色、画像表示法などを設定する

Myデータベースの構築から公開までに要する時間は

データ量と設定の詳細さに依存する。もしデータ量が少なければ、数分程度でデータベースを公開できる。Myデータベースの操作は簡単であるため、地域研における研究者DBの多くはMyデータベースにより構築されている。

### (2) My データベース API

データベースを構築しても、データベースの応用プログラムに対する要求は研究者ごとに異なり、変更や拡張などの頻度も高い。これらは研究者DBの特徴であるが、予算や人員が乏しいため、システム管理者は研究者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に対して研究者が抱く不満の一つとなっている。そこでMyデータベースでは、API (Application Programming Interface) の仕様を公開し、応用プログラムの作成は研究者に任せるとの方針を採用している。

MyデータベースのAPIでは、検索式として Context Query Language (CQL) [3]を採用している。地名データベースにおける検索式の例を示す。

```
http://Some_URI?operation=searchRetrieve&version=1.2&
query=(c2="宇太水分神社")&recordSchema=original
```

ここで **c2** はデータ項目名を指示する My データベース独自の手法である。通常のデータベースシステムでは、データ項目名に空白文字を使うことを許さない。しかし、スプレッドシートによりデータを作成する場合には、空白文字を含んでいることが多い。Myデータベースでは空白文字を許しているため、データベースシステムとの整合性を確保するために、上記の方法を採用している。

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ある。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 Web ブラウザに入力すると(ただし *Some\_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、返戻形式を Simple JASON[4]とすると)、以下のような返戻を得る。

```
{"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"}
 }
```

これは宇太水神社を検索した返戻で、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、研究応用に適したGUIの作成やマッシュアップの作成などに都合がよい。

### (3) My データベース API の応用例

My データベースとAPIの応用例を図2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、掲載できる数や大きさには限界がある。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ことが困難となる。

このような問題を解決する手段として、QRコードを介して誌面と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[5]。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、QRコード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、紙の書籍では困難であった地図などのズームイン・ズームアウトや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)の2冊の出版物について、Myデータベースとの連携を実現している。特に後者では、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2 APIを利用したMyデータベースの応用例  
Figure 2 Application Example of MyDatabase API

## 2.2 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の多くは個人の努力によるものである。機関DBのように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、類似のデータベースを連携させる必要がある。また研究の過程で複数のデータベ

スを併用する場合も多く、データベースを統合検索できれば便利である。そのためには、インターネット上に分散しているデータベースをシームレスに接続する必要がある。この機能を資源共有化機能と呼んでいる。

共有化を実現する簡単な方法はメタデータの統一である。例えばOPAC(Online Public Access Catalog)では、参加する全てのデータベースが同じメタデータを利用している。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可能性ある上に、多様な研究領域を包摂する人文社会科学分野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。

### (1) 資源共有化システムの構造

資源共有化システム[6]は、資源共有化機能の実現を目指した情報処理システムである。その構造を図3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対応関係を設定する。この対応関係をマッピング規則と呼んでいる。資源共有化サーバ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[7]またはSRW[8]を利用する。資源共有化サーバには、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。

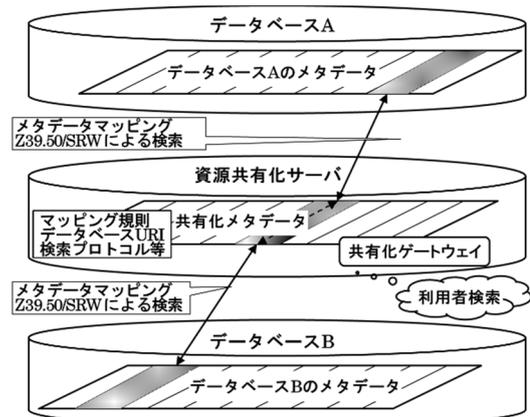


図3 資源共有化システムの構造

Figure 3 Structure of Resource Sharing System

資源共有化システムに与えられた検索命令は、データベースごと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。これら手順により、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51データベース(地域研および東南アジア研究所(22)、総合地球環境学研究所(5)、国立民族学博物館(19)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、東京外国語大学アジア・アフリカ言語文化研究所、カリフォルニア大学バークレイ

校)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。



図 4 資源共有化システムの検索例 (上から、検索画面、検索結果一覧、詳細表示、画像表示)

Figure 4 Retrieval Example of Resource Sharing System (From the Top: Search Screen, Search Result, Detail Information, Map)

### (2) 資源共有化システムの検索例

地域研が公開しているデータベースは、英語、タイ語、ロシア語、マレー語などの多言語で記述されている。その

ため、地域研共有の資源共有化システムでは、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が必須である。多言語間の語彙変換には言語グリッド[9]が提供している Web サービスを利用している。資源共有化システム[10]による検索例を図 4 に示す。

### 3. 地域研究情報基盤の問題点と対応

地域研究情報基盤が提供する My データベースおよび資源共有化システムにより、研究者 DB の構築、データベースの利活用およびデータベースの共有化を実現し、地域研究を支援する情報システムとして高く評価されている。しかし、地域研究情報基盤は 10 年以上前の技術に依っており、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き換えや共有化するデータベースの追加が容易ではない
- ② 重要な部分に独自の技術を採用しているため、サーバの導入・運用コストが高く、普及の障害となっている
- ③ 柔軟なデータベース共有が困難である。ネットワーク越しにデータベースを結合すること (join など) ができないので、API を使って同様の操作を手続き的に記述する必要がある。
- ④ 人文科学分野では一般的なデータ記述や操作を素直に行えない。例えば、家系図・人間関係・シソーラスなど、人文社会科学分野では一般的なグラフ型のデータを My データベースでは扱えない。さらに、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) 操作も、API を使って手続き的に記述しなければならない。
- ⑤ Web 上の非定型データ (HTML 文書など) を扱うことが困難である。

地域研究の場合、地域研究コンソーシアム[11]加盟機関だけでも 162 以上のデータベースが公開されている[12]。これら地域研究情報は、海外展開している企業などの直接の関係者以外にとっても、貴重な情報資源となっている。これらの地域研究情報を共有して横断的に利用できれば、地域研究を支援する強力な情報ツールとなる。しかし、データ構造、語彙、言語などがバラバラであるため相互に連結できず、共有は困難である。もちろん資源共有化システムを使えば共有化は可能であるが、小規模な人文社会科学機関には導入コストが高すぎて、コストパフォーマンスが悪すぎる。

また、一地域の変動が直ちに世界へ影響を及ぼす今日、地域研究が即時対応性の高い課題を扱うためには、データベースに蓄積された情報に加えて、ウェブ上で流通している膨大かつ多様な地域研究情報にも対応しなければならない

い。いまや地域情報はビッグデータとなりつつあり、人手で処理することは不可能である。しかし資源共有化システムでは、データベースの定型データと Web 上の非定型データを接続することが困難である。

これらの問題を解決するために、地域研では新しいデータモデルに基づいた地域研究情報基盤(以下、新情報基盤)の開発を試みている。その概要を以下に述べる。

### (1) Resource Description Framework(RDF)の導入

関係データモデルによるデータベースが普及した要因として、直感的で単純なデータ構造(関係データ、いわゆる表形式)と関係代数(あるいは関係論理)に準拠した問合せ言語(SQL)を上げることができる。書籍や画像などに関するメタデータは表形式で記述できるので、関係データベースの適用は妥当である。しかし、家系図やセンサーのようなグラフ構造のデータを関係データモデルで記述することは簡単ではない。さらに、インターネット上のデータをリンクするために、ネットワーク越しでテーブルを結合することも容易ではない。

そこで新情報基盤では、グラフ型のデータを記述でき、インターネットを介したデータリンクが容易であり、さらに機能の豊富な問合せ言語が考案されているデータモデルとして、Resource Description Framework(RDF) [13][14]を採用した。RDF は W3C で規定されたデータ表現モデルであり、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。

RDF では、情報を主語・述語・目的語の3つの要素で表現される文の集合として扱う。例えば A is B (A ならば B である) という文を RDF で表現すると、図 5 の上段に示す有向グラフとなる。ここで、A が主語、is が述語、C が目的語である。これを RDF の triple と呼ぶ。同様して B is C は図 5 の中段のようなになる。最初の文の目的語と次の文の主語が同じ B であれば、図 5 の下段のように2つの文を繋ぐことができる。このようにして、RDF は単純な文から複雑なグラフを組み立てる。また RDF では主語、述語、目的語を URI で表現するので、別のデータベース中にある情報資源どうしをリンクすることができる。

RDF によるデータ表現は図 5 に示すような有効グラフであり、データ記述用としては非実用的である。実際のデータ記述やデータ交換には XML や Turtle などを利用する。

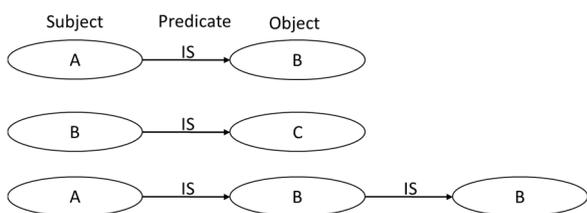


図 5 RDF によるデータ記述  
 Figure 5 Data Description by RDF

### (2) RDF データベースの構築

RDF を利用した新情報基盤の機能と性能を評価するために、RDF Triple Store (RDF データベース)の構築と、現在の地域研究情報基盤に蓄積されているデータの RDF 化を進めている。前章の地名データベースに登録されている「相国寺」を Turtle 形式の RDF に変換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 geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix gzt:<http://Some_URI/gzt/elements/1.0/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix skosxl: <http://www.w3.org/2008/05/skos-xl#>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix ns7: <http://supercluster.cias.kyoto-u.ac.jp/pn/pntype/>.
@prefix dcq: <http://purl.org/dc/terms/>.
```

```
<http://Some_URI/id/placename/10026682>
rdf:type geo:SpatialThing ;
rdfs:label "相国寺";
rdfs:comment "30048501 : 相国寺と重複".
gzt:country "山城";
gzt:locality "京都市上京区////";
gzt:localityOriginalNotation "京都市上京区////";
gzt:projection _vb2220681 ;
geo:Point _vb2220682 .
dc:format _vb2220679 ;
dc:source "大日本地名辞書";
gzt:countyAltLabel "カミキョウ";
dc:identifier "10026682";
dc:subject _vb2220680 ;
skosxl:altLabel "ソウコクジ";
gzt:county "上京区".

_vb2220679 rdfs:label "1".
_vb2220679 owl:sameAs ns7:Point .
_vb2220680 rdfs:label "寺院".
_vb2220680 dcq:subject
<http://Some_URI/id/placeattribute/13>.
_vb2220681 rdfs:label "1";
owl:sameAs ns7:Point .
_vb2220682 gzt:lat 35.03333333 ;
gzt:long 135.7627778 ;
geo:lat 35.0312057223273 ;
gzt:x 3876771.4929401 ;
gzt:zone "53 S";
geo:long 135.765576348059 ;
gzt:y 569834.290676553 .
```

ここで@prefix で始まる部分は名前空間の URI 定義である。<http://Some\_URI/id/placename/10026682>が、このデータベースにおける相国寺を表す URI である。これが主語となり、rdf:type を述語、geo:SpatialThing を目的語として、一つ文が構成される。同じ主語を共有する文が gzt:county"上京区"まで続く。以下同様である。

RDF データの問合せ言語としては SPARQL[15]がデファクトスタンダードある。SPARQL クエリを受け付ける RDF Triple Store の URI を SPARQL Endpoint という。SPARQL クエリを利用すると、グラフの階層を行き来するような検索命令を容易に記述できる。以下は「相国寺と同じ国に属する地名」を探すための SPARQL クエリの例である(前述

の RDF データを参照).

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

SPARQL クエリは Turtle の構文で triple パターンを記述するので、直感的で理解しやすい。例えば where 句の最初の {?s rdfs:label "相国寺"} は、述語が rdfs:label、目的語が相国寺の文を RDF データから探し、その文の主語を変数 s に代入することを示している。ここでは相国寺の URI、つまり http://Some\_URI/id/placename/10026682 が代入される。次の {?s gzt:country ?country} は、主語 s が相国寺の URI、述語が gzt:country である文を探し、その文の目的語を変数 country に代入することを示している。ここでは山城が代入される。これにより、相国寺の属する国名が検索される。次の {?place gzt:country ?country} は、述語が gzt:country、目的語 country が山城である文を探し、その文の主語を変数 place に代入することを示している。これにより、相国寺と同じ国にある別の地名 (URI) が検索される。select 句にしたがって、その地名 (URI) と名称を表示する。この操作を指示された文がなくなるまで繰り返す。結果は以下のようになる。このような形式のデータも加工が容易であり、研究応用に適した GUI の作成やマッシュアップの作成などに都合がよい。

```
http://Some_URI/id/placename/10023952 朱雀野
http://Some_URI/id/placename/10042178 真葛原
http://Some_URI/id/placename/10003404 池坊
.....
```

(3) RDF データベースの利用事例

RDF データベースの応用事例として、地名データベースを利用したアプリケーションを試作した。歴史史料などをデジタル翻刻する際に、地名に緯度と経度を紐付けできたら、いろいろと便利である。図 6 は簡単なテキストエディタと地名データベースを連携させた翻刻支援ツールの試作版である。

このツールでは、エディタ中 (図 6 上段) の文字列 (この例では相国寺) をドラッグ後にクリックすると、地名データベースを検索して候補の一覧を表示する (図 6 中段)。ここから正しいと考えられる地名を選択すると、地名情報を XML 形式あるいは RDFa 形式 [16] でテキストデータに埋め込む。RDFa 形式による出力例を以下に示す。

```
<gn:Feature typeof="schema:Place">
  <a typeof="nihugazetteer" href="http://Some_URI/id/placename/10026682"
    property="nihugazetteerLink"><span property="rdfs:label">相国寺</span></a>
  <div style="display:none">
    <rdf:RDF
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

```

```
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:gzt="http://supercluster.cias.kyoto-u.ac.jp/gzt/elements/1.0/"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dccq="http://purl.org/dc/terms/"
xmlns:skos="http://www.w3.org/2008/05/skos-xml#">
<rdf:Description rdf:about="http://Some_URI/id/placename/10026682">
  .....
</rdf:Description>
<rdf:Description rdf:nodeID="V62220682">
  <gzt:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#double">35.03333333</gzt:lat>
  <gzt:long rdf:datatype="http://www.w3.org/2001/XMLSchema#double">135.7627778</gzt:long>
  <geotlat rdf:datatype="http://www.w3.org/2001/XMLSchema#double">35.0312057223273</geotlat>
  <gzt:x rdf:datatype="http://www.w3.org/2001/XMLSchema#double">3876771.4929401</gzt:x>
  <gzt:zone="53 S"></gzt:zone>
  <geolong
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">135.765576348059</geolong>
  <gzt:y rdf:datatype="http://www.w3.org/2001/XMLSchema#double">569834.290676553</gzt:y>
</rdf:Description>
</rdf:RDF>
</div>
</gn:Feature">は、日本の禅寺。京都市上京区にある臨濟宗相国寺派大本山の寺である。山号を萬年山と称し、正式名称を.....
```

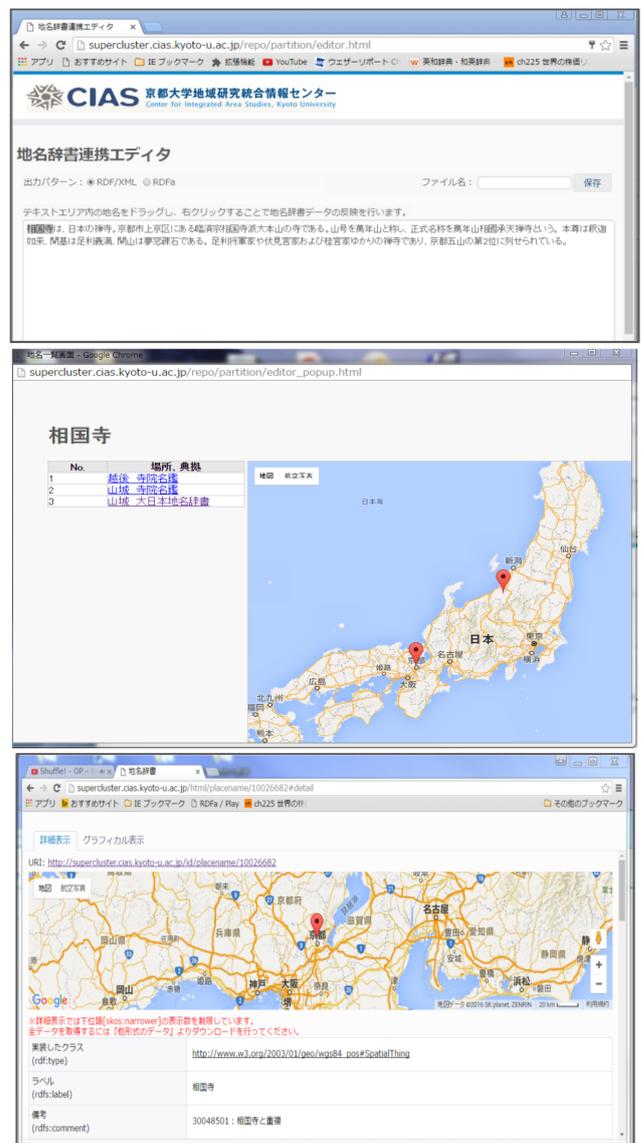


図 6 RDF 版地名データベースを利用した応用事例  
 Figure 6 Application Example using RDF Gazetteer Database

この文書を HTML ファイルとしてブラウザで表示すれば、2行目の<a>タグ内の href 属性を使って地名辞書の該当データへリンクできる。また、<rdf:Description>タグ内の情報を利用すれば、地名の位置を地図上で可視化することなどができる(図 6 下段)。さらに、この文書を LOD(Linked Open Data)[17]として公開すると、他の情報への中継知識として利用できる。

#### 4. 考察

新情報基盤のプロトタイプを今年度中に構築し、データ量と検索速度の関係、データ更新などの作業の容易さ、LOD としての利用可能性などに関する検証などを開始する予定である。以下では、現在の地域研究情報基盤が抱える問題点に対する、新情報基盤による問題解決の可能性を考察する。

- ① **利用者によるマッピング規則の書き換えや共有化するデータベースの追加が容易ではない**：RDF では地名データベースで示したようなファクト型データも、「データベース A の項目 a とデータベース B の項目 b は同じ内容である」といったマッピング規則のような知識データも、同じ RDF triple で表現され、RDF Triple Store に登録され、表現上は区別されない。したがって、現在の地域研究情報基盤に比べると新情報基盤の構成は単純になるので、データベース追加やマッピング規則変更などの機能を実装することは容易になると期待している。ただし、RDF データや SPARQL クエリの作成は人文社会学研究者にとって敷居の高い作業であるため、何らかの支援ツールが必要であると考えている。
- ② **サーバの構築・管理・運用コストが高く普及の障害となっている**：新情報基盤の構築では RDF Triple Store と SPARQL Endpoint の実装が中心となる。これらについては、信頼性の高い商業ソフトやフリーのソフトが利用可能である。新情報基盤は、これらのソフトを組み合わせで構築する。情報センターのように高速かつ安定的な運用を求められる場合は、商業ソフトを利用する選択肢があり、その場合、低コストを実現するのは困難であろう。一方、個人研究者や小規模研究室の場合、性能や安定性を多少犠牲にしても大きな問題となる可能性は低いため、フリーソフトを利用した低コストのシステム構築が適当である。速度や安定性の差はあるものの、基本的な機能はどちらも同じである。このような棲み分けは、現在の地域研究情報基盤では不可能である。つまり、新情報基盤では全体として低コストのサーバ構築・管理が可能になると期待している。
- ③ **柔軟なデータベース共有が困難である**：RDF は情報資源を URI で記述するので、ネットワーク越しのデータ

連携は容易である。しかし、ネットワーク上の多数の RDF Triple Store を繋いだ場合でも快適な検索速度を維持できるかは未知数である。概念の未整備という問題もある。例えば同じ研究者であってもデータベースごとに識別子、つまり URI が異なっているのが現状である。また同じタイトルという語彙が、あるデータベースでは書名の意味として、別のデータベースでは肩書きの意味として使われている。RDF によるデータリンクを実現するためには、基本的な実体や語彙に関する辞書の整備が必要である。

- ④ **人文科学分野において一般的なデータ記述や操作が困難である**：3章で示したように、RDF はグラフ構造を記述でき、SPARQL はグラフ構造を対象としたクエリを受け付けることができる。したがって、新情報基盤では家系図やシソーラスなどの処理が容易である。ただし、データ量が膨大になった場合やネットワーク上の多数の RDF Triple Store を繋いだ場合にも十分なパフォーマンスを維持できるかは未知数である。
- ⑤ **Web 上の非定型データを扱うことが困難である**：これも3章で示したように、RDFaなどを導入すればある程度は可能である。ただし、どの程度の処理が可能であるかは、今後の評価の重要なテーマになると考えている。

#### 5. まとめ

地域研究資料の保存・公開・共有・利活用の促進を目指して、My データベース、My データベース API、資源共有化システムを中核とする地域研究情報基盤の開発を進めた。My データベースによる研究者 DB の構築実績、My データベース API を利用した応用事例の実績の蓄積、資源共有化システムによる国内外地域研究関連 51 データベースの共有の実現など、当初の目的を達成しつつある。地域研究を支援する情報システムとして高く評価されている。

しかし、地域研究情報基盤は 10 年以上前の技術に依っており、高い管理・運営コストや、研究者の多様なニーズに対応できないなどの問題も明らかになってきた。これらの問題への対策として、RDF をデータモデルとした新しい情報基盤の開発に着手した。

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。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、アーカイブ、機関レポジトリあるいは研究者 DB といったデータベースの相違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。構築中の新情報基盤が、その議論の一助として貢献することを期待している。

## 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] 原 正一郎:「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–」, じんもんこん 2014 論文集, Vol. 2014, No.3, pp.185 – 192, 2014-12-06.
- [3] LOC : CQL Contextual Query Language SRU Version 1.2 Specifications, <http://www.loc.gov/standards/sru/specs/cql.html>
- [4] D. Crockford : RFC 4627: <http://www.rfc-editor.org/rfc/rfc4627.txt> (2006).
- [5] 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>
- [6] 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).
- [7] LOC: Z39.50, <http://www.loc.gov/z3950/agency/>
- [8] LOC: SRU- Search/Retrieve via URL, <http://www.loc.gov/standards/sru/>
- [9] 言語グリッド: <http://langrid.org/jp/>
- [10] 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-1g/cgi/Start.exe>
- [11] 地域研究コンソーシアム: <http://www.jcas.jp/>
- [12] 地域研究コンソーシアム: 「特集 地域研究コンソーシアム加盟機関所蔵の研究資源一覧」, 地域研究コンソーシアムニューズレター, No.15, 2013 年 12 月.
- [13] W3C: Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- [14] 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 年.
- [15] W3C : SPARQL 1.1 Query Language, <http://www.w3.org/TR/sparql11-query/>
- [16] W3C : RDFa Core 1.1 - Third Edition, <https://www.w3.org/TR/rdfa-core/>
- [17] 武田英明, 小出誠二: 特集 リンクト・オープン・データの活用, 情報処理, Vol.57, No.7, p.568-625, 2016.