

研究資源共有化システム ニュースレター

第10号
2015年
3月27日

大学共同利用機関法人 人間文化研究機構 研究資源共有化事業委員会
Inter-University Research Institute Corporation National Institutes for the Humanities The Committee of Resource Sharing Project
<http://www.nihu.jp/sougou/kyoyuka/system/index.html>

■ 人間文化情報の社会知化への実践

人間文化研究機構 理事 榎原 雅治

昨年4月より人間文化研究機構で非常勤の理事として情報資源共有化事業を担当している。情報学に関する知識は乏しく、このような場に一文を書くことは面はゆいが、この場を借りて、本務先である東京大学史料編纂所において私がかかわっている、歴史学・考古学と地震学の情報統合の試みについて紹介させていただきたい。

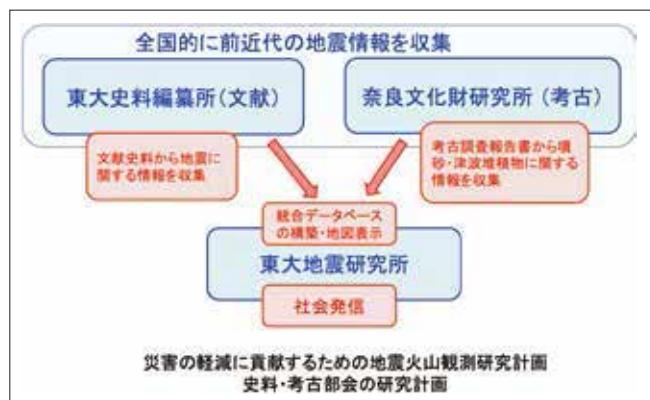
東日本大震災以後、前近代の地震、津波に対する社会的な関心はにわかに高まり、本年度より、科学技術・学術審議会測地学分科会の建議「災害の軽減に貢献するための地震火山観測研究計画の推進について」を受け、東京大学地震研究所に設置されている地震・火山噴火予知研究協議会のもとに史料・考古部会が設けられた。この部会では、文献史料に記録された前近代日本で起きた地震に関する記述、考古発掘報告書から判明する地震痕跡についてそれぞれデータベースをつくり、地理情報を与えることによって時間と空間をキーとして統合検索を可能とし、さらに結果を地図上に表示するシステムを構築することをめざしている。

地震学と歴史学の連携は長い経験をもっている。濃尾地震を機に文部省震災予防評議会（地震研究所の前身）によって編纂された『大日本地震史料』が1904年に刊行されて以来、『増補大日本地震史料』『日本地震史料』『新収地震史料』『日本の地震史料拾遺』と、現在に至るまで全37冊、27000頁に及ぶ前近代の地震史料が、地震研究所のほか武者金吉・宇佐美龍夫氏らの努力によって刊行されている。これらは貴重な史料集であるが、1世紀にわたって増補につぐ増補を重ねてきたものであるため、史料を年代順に通覧することができない。また研究者が史料調査中に地震関係の史料を一つ見つけても、それが地震史料として既知のものであるか否か確認するには30冊以上の本をめくらなければならないという状態となっている。無論、こうした状態の打開の必要性はかなり以前から指摘されており、既刊地震史料集のうち古代中世部分については、石橋克彦氏を代表とする古代中世地震史料研究会によって全文フルテキスト化が行われ、2009年よりネット公開されている。それに数十倍する近世部分についても、中越地震のうちに地震研究所を拠点として実施されたひずみ集中帶プロジェクトにおいて、

新潟地域を中心とした史料のフルテキストが行われ、震源・震度の推定も表示されるシステムが構築されている（次頁）が、大半の近世史料については手付かずの状態となっている。

一方、考古発掘報告書から過去の噴砂遺構を検出する試みも行われており、文献史料に残る地震記録と関連付けられたり、文献史料に記されない地震が発見されたりという重要な研究が進められているが（寒川旭『地震の日本史』中公新書）、現在のところ、卓越した研究者の努力に頼っている状態と言ってよいだろう。ある研究者が文献史料によって知りえた地震について、大地に残る痕跡を求めようとしても、探索する術がないというのが現状である。

皮肉な話であるが、東日本大震災後の歴史地震に対する社会的関心の高まりは、歴史地震研究を進めるうえでの積年の課題を解決に向けて一挙に進めることとなった。本年度開始された地震・火山噴火予知研究協議会史料・考古部会で進めている研究計画は下図のようなものである。文献史料については、東大史料編纂所が中心となって、既刊地震史料集のフルテキストデータベースを構築し、地名にはタグ付けして緯度経度による位置情報を与える。さらに国文学研究資料館から公開されている日本古典籍総合目録データベースとリンクさせ、出典史料の性格を把握しやすくすることも予定している。考古資料については、奈良文化財研究所が担当して、既刊の発掘調査報告書から噴砂、津波堆積物などの地震痕跡を抽出し、年代情報、位置情報を与えたデータベースを構築する。この二つのデータベースを統合検索することによって、年代や場所から前



近代の日本で起きた地震についての基礎情報に誰でもアクセスできるシステムを作ろうというのが、今回始まった計画である。

既存の史料集や考古発掘報告書をベースにするという点では、既存の知識の電子化といえる。しかしこれまでそれらの知識は別個に存在し、高度に習熟した専門家でなければ使いこなせない状態にあった。今回の計画は、情報学を媒介にして、地震学、歴史学、考古学の知識を関連付け、社会に発信しようというものである。それは文理連携による人間文化情報の社会知化への実践といつていだらう。

無論、課題は多い。一つの県だけでも1万冊を超える数になる発掘調査報告書からの地震情報の抽出や、27000頁の既刊史料のテキスト校正の問題もあるし、月日、場合によっては時刻単位で記される文献史料の時間情報と世紀単位で付与される考古の時間情報、逆に正確な緯度経度で示される考古の位置情報と伸縮の大きい文献の位置情報をどう統合させるか、

それをどう表示させるか、具体的なイメージはまだ検討途上である。情報学研究者の教示を願うところである。



より効果的な研究資源共有のための プロトタイプ構築事業について

人間文化研究機構 後藤 真

1. nihuINTが可能にしたことと課題

人間文化研究機構では、現在のnihuINTをより効果的に検索する手段の検討を行っている。nihuINTは、平成27年1月現在、150のデータベースを同時に検索でき、そのデータ件数は520万を超えるものとなっている。この大量のデータは人間文化の研究を支えるための重要な基盤であり、多くの情報へとアクセスするための重要な手段として位置づけることができるまでになりつつある。

しかし、nihuINTも構築しておよそ5年が経過し、さまざまな問題が指摘されている。特に以下のような問題点については、現状のシステムをそのまま維持しているだけでは困難な部分がある。

①柔軟な統合検索の問題: 現在のシステムにおいて、共有メタデータはダブリンコアのみであり、各データベースと共有メタデータとマッピング規則も固定されている。つまり、共有化のパターンが固定されてしまっている状況がある。書誌的データベース同士の共有検索であれば、それなりに有効な部分はある。しかしながら、たとえば地名辞書の階層情報を共有することは困難である。

②データベース連携が困難である: データベース連携においてはデータベース間でのテーブル結合や副クエリなどの検索操作が必要であるが、現在の共有化システムでは実現されていない。外部データベースとの連携はより困難である。

③データマイニングが困難である: 現在のシステムでは、各データベースをプログラムにより検索するAPI(Application Program Interface)がないため、データベースごとに検索画面をいちいち開いてキーワードを入力しなければならない。そのため、あるデータから関連するデータを検索することが困難である。

そこで、後述するプロトタイプを作成し、実験を行うこととした。

このプロトタイプ構築は、これらの課題を解決するための情報学的な手法を具体的に検討することにあり、今後の資源共有化システムをより発展させることを目指している。

2. 実験の内容

実験対象とするデータベース(人間文化研究機構関係)は以下の通りである。いずれもnihuINTに収録されているものである。

1. 日本莊園データベース(国立歴史民俗博物館)
2. 日本古典籍総合目録DB(国文学研究資料館)
3. デジタル地名歴史辞書(総合地球環境学研究所)

これらのデータベースをRDFストアとして再構築し、新しい共有化システムのプロトタイプとして構築を開始した。また、SPARQLエンドポイントを構築し、検索エンジンとして、作成を行った。これにより、上記の問題点を解決することができるか検証を行っている。

具体的には、上記データベースの中から、莊園データベースの情報のうちから備中に関係する100件程度の情報を抽出、それに関係する古典籍総合目録データベース・デジタル歴史地名辞書の情報を抽出している。そのうえで、後述のようなデータベースのリンクによる連結とインターフェース作成を試みている。なお、日本莊園データベースは、以下のようないくつかの項目を持つている。

莊園コード／国名／郡名／莊園名／重複コード／参考市町村／市町村コード／明治村字名／史料村郷名／領家・本家／初見年和暦／初見年西暦／出典／遺文番号／記録類／地名辞典／備考／関係文献

これらの各項目から以下のような連携を行うこととした。1. 郡名、明治村字名、史料村郷名から地名の情報を抽出し、デジタル地名歴史辞書とリンクする。2. 出典、記録類から、莊園が出てくる史料名の情報を抽出し、日本古典籍総合目録データベースと

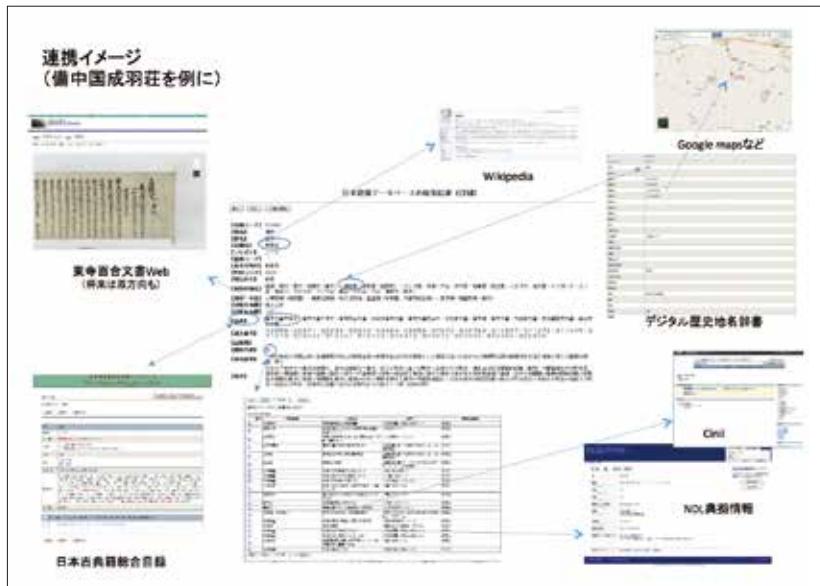


図1 リンク関係の全体図

リンクする。

また、機構外のデータベースについても、プロトタイプからのリンクを作成し、連携が可能かどうかの実験を試みている。具体的には、以下のようなデータベースにリンクを行っている。

1.出典から、京都府立総合資料館が所蔵する東寺百合文書Webの個別画像へのリンク。2.関係文献の論文名から、国立情報学研究所のCiNiiおよび国立国会図書館のNDLサーチへのリンク。3.莊園名からDBpediaへのリンク。全体像は図1のようになる。

まずは、内部のデータベースから外部の資源へのリンクを試験的に作成している。なお、プロトタイプでは、1レコードにつき、1URIを付すことも実験しているので、このプロトタイプが公開に至った段階では、外部のデータベースから、これらのデータベースへのリンクを行うことも可能となる。

3. 研究シナリオ

さらに具体的には、以下のような研究上の要求シナリオを想定した。それらの要求を満たすリンクの方法と例を示す。

3-1.中世莊園を調べた際にデータベースで提示された史料の特徴・性質をちょっと確かめたい場合は、日本古典籍総合目録データベースと莊園データベースの「出典」「記録類」に関係を作ることを試みることで、成果に結びつけられると考えられる。たとえば、備中国の「井原莊」の情報が出てくる「九条家文書」には、ほかにどのような莊園が出てくるのか、この史料にアクセスするには、どうすればよいかなどが、データベースを行ったり来たりすることなくシンプルに井原莊の情報にアクセスすればよく、あとは、リンクをたどることで情報を得ることができる。(図2)

3-2.中世莊園に関係すると推定される地名がどのような広がりを持っているか、視覚的に地図で確認したい場合には、莊園データベースの「郡名」「明治村字名」

「史料村郷名」と地名辞書に関係を作ればよい。たとえば、成羽莊にある地名情報(福地)から、地名辞書を通じて、緯度経度を確認し地図上で時期的な限界はあるというものの、当該データを見ることができます。

3-3.ある史料について(莊園関係で)言及している論文や著者の情報を詳しく知りたい場合には、古典籍総合目録と、莊園データベースを結びつけたうえで、莊園データベースの「関係文献」の論文名から、CiNiiの論文へと関係を作ることとなる。たとえば、「『東鑑』に出てくる莊園のことを書いている論文の詳細な書誌や(ある場合には)そのPDFデータ」を探すことが可能になる。

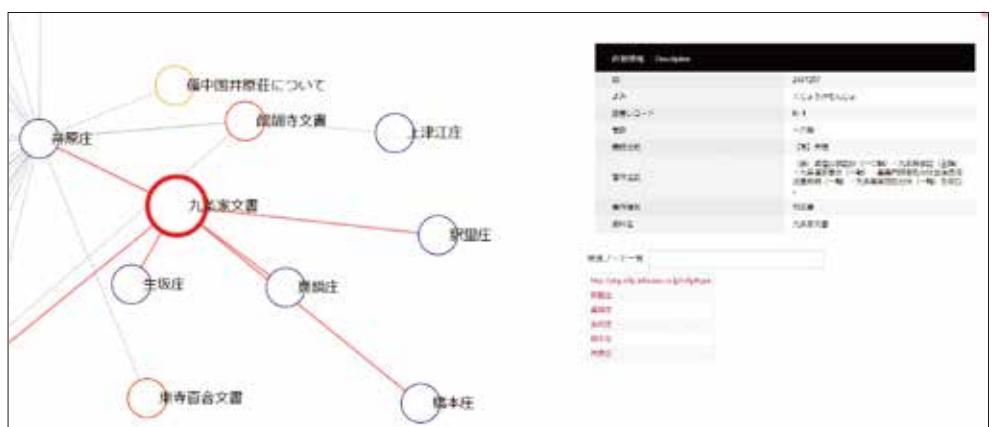
3-4.莊園データベースの情報にかかる外部の情報資源を見たり、また、外部の情報資源には、莊園のデータベースの情報資源とかかわるものはないのか確かめたい場合には以下のようなリンクとなる。東寺百合文書Webと莊園データベースの史料名間で対応を取ればよい。たとえば、新見莊に関する東寺百合文書の画像がリンク等で閲覧できるようになる。

また、1レコード1URIを実装することによって、一般の検索エンジンからも発見が可能となる土台を作ることができ、個別データへの容易なアクセスを実現することになると考えられる。

このように、現在実験中のRDFベースのプロトタイプは、URL同士の結合によるデータベースの内部・外部との連携を可能とするとができると考えられる。また、リンクをたどるしくみやSPARQLの応用により、より柔軟な検索手段も提供できるようになるであろう。現在、これらのリンク確立や使いやすいインターフェースなどについて、課題の洗い出しを行っているところである。

これらのRDFによるデータの提供は、より複雑なものやより広範な検索システム・サービスへのデータ提供を可能にするものともいえる。大学共同利用機関法人として、各大学での人文系データの発見やそれを支援するための技術開発のための重要な資源提供になるととも考えられる。

このプロトタイプの成功が、あらたな情報資源の共有と資源基盤の確立になるとと考えられる。

図2 井原莊のリンク先である「九条家文書」のリンク情報
九条家文書に記述されている莊園名が表示されている。右側には所蔵など古典籍総合目録の情報が記される。

2014年度の統合検索システム活動概要

2014年7月に、人文科学研究資源への統合検索システムnihuINTによるアクセスを支援する「人名一覧表示システム」を公開した。このシステムは、『古事類苑』より抽出した近代以前の人名に芳賀矢一編『日本人名辞典』掲載の解説文を付与した一覧を表示する。一覧の人名をキーワードとしてnihuINTでの検索を実行する機能を有している。様々な人名を一覧で示すことにより、利用者に新たな知識発見の機会を提供することを目的としている。2014年度後半からは、人名情報を増強とともに、他のデータベースと柔軟に情報連携するため、HTMLおよびRDFの人名単位ページデータの作成を進めている。

2014年度は、nihuINTの検索対象に、民博の「ジョージ・ブランドン・コレクション」(英語版を含む)、nDPの「現代南アジア・イスラーム復興思想: マウドゥーディ著作目録と解題」、「中東現代文学邦訳・研究文献」、「インドの統計地図集」を追加した。さらに、

2014年度の時空間解析システム活動概要

地名情報にかかる合同の研究会開催

地名は、時空間情報の解析や研究資源の探索を行う上で重要な要素である。このことから、地名にかかる情報技術に関する課題に取り組んでいる2つの研究組織（京都大学地域研究統合情報センター・地域情報学プロジェクト「地域の知」の情報学－時間・空間・語彙に注目した地域情報学の展開－（H-GIS研究会）、および、国立情報学研究所公募型共同研究「GeoNLPプロジェクト：自然言語文を対象とした地名情報処理とコミュニティの展開」と）との合同の研究会「地名にかかる情報技術に関する研究会」を2014年7月19日に国立情報学研究所において開催した。以下の5件の報告があり、緯度経度などの座標との対応付けや地名同士の時空間的な関係の表現、地名にかかる情報の公開などのさまざまな技術的な課題について検討を行った。

榎原雅治(東京大学史料編纂所／人間文化研究機構)

「歴史資料としての地名の活用」

桶谷猪久夫(大阪国際大学)

「歴史地名辞書をいかに作るか?」

関野 樹(総合地球環境学研究所)

「資源共有化における地名の役割」

北本朝展(国立情報学研究所) 「地名情報処理環境G-NLPの紹介と歴史的・地名

「地名情報処理環境GeoNLTの紹介と歴史的な地名に関する課題」

CONTENTS

人間文化情報の社会知化への実践 榎原 雅治	1
より効果的な研究資源共有のためのプロトタイプ構築事業について 後藤 真	2
2014年度の統合検索システム活動概要 古瀬 藏	4
2014年度の時空間解析システム活動概要 関野 樹	4

国語研の「米国議会図書館蔵『源氏物語』翻字本文」などを加えるため、nihuINTへのメタデータマッピング作業を行っている。また国立国会図書館に続く、新たな機関間連携を進めており、京都大学地域研究統合情報センターのデータベースもnihuINTの検索対象に加える予定である。

総合地球環境学研究所 関野 樹

原 正一郎(京都大学地域研究統合情報センター)

「地名データのセマンティックWebへの応用」

研究会の内容は報告書としてまとめられ、Webページなどからダウンロード可能である。

時空間解析ツールのBDF対応

別途進められている「より効果的な研究資源共有のためのプロトタイプ構築事業」に合わせて、時空間解析ツールでセマンティックWeb技術を応用するための改良を進めている。今年度は、上記「プロトタイプ」で中心となるRDF(Resource Description Framework)データを時空間解析ツールに取り込むための準備作業を進めた。現段階では、RDFで記述された内容を表現することはできないものの、RDFデータを時空間解析ツールに読み込んだ場合の動作の基本設計を行うとともに、現状の機能に障害が発生しないよう、対策を講じた。今後、研究資源の共有化において、プロトタイプで検証が進められている技術を活用しながら、時空間解析ツールがより効果的に活用されることが期待される。

その他

研究資源共有化ニュースレター9号において、時空間解析ツールの使用方法の解説を掲載した。また、旧5万分の1地形図の地名データについて、当時の測量精度に起因する誤差を補正する作業を進めている。

研究資源共有化システム ニューズレター 第10号

発行日 2015(平成27)年3月27日

発行・編集 大学共同利用機関法人 人間文化研究機構
研究資源共有化事業委員会

〒105-0001 東京都港区虎ノ門4-3-13 ヒューリック神谷町ビル2階
TEL.03-6402-9200(代表)