

## 目撃情報の収集による歴史的状況記録パスファインダーの構築

市野美夏（情報・システム研究機構 データサイエンス共同利用基盤施設

人文学オープンデータ共同利用センター／国立情報学研究所）

橋本幸恵（フリーランスプログラマ）

平野淳平（帝京大学 文学部）

増田耕一（首都大学東京 都市環境学部）

北本朝展（情報・システム研究機構 データサイエンス共同利用基盤施設

人文学オープンデータ共同利用センター／国立情報学研究所）

歴史的状況記録（HSR）を用いた研究の資料探索や情報取得を支援するため、歴史資料に関するさまざまな人々からの目撃情報を収集し、それを利用したパスファインダーのようなサービスを提供するシステム構築を進めている。本システムのサービスに利用する情報は、多岐にわたる史料利用者の相互協力によるものであり、その収集および整理も重要である。一方、情報が増えることでより充実したパスファインダーとなっていくシステムである。このシステムを利用した歴史気候学の資料整理の事例と、西暦・和暦の変換、資料の記録期間検索機能などを紹介する。

### A pathfinder system for Historical Situation Records using eyewitness information

Mika Ichino (ROIS-DS-CODH / NII)

Yukie Hashimoto (freelance Programmer)

Junpei Hirano (Faculty of Liberal Arts, Teikyo University) Kooiti Masuda (Faculty of Urban Environmental Sciences, Tokyo Metropolitan University)

Asanobu Kitamoto (ROIS-DS-CODH / NII)

A system that features a pathfinder function by using collections of eyewitness information of Historical Situation Records (HSR) has been developed. This will help research with exploring and finding documents, including HSR. The database can contain information about something's bibliography and location, as well as its HSR and various descriptions such as the state of the materials, errors and reliability of HSR data. The system re-uses information from previous research in each field with the users' mutual corporation, which also creates new interdisciplinary collaboration, to avoid repeating the same processes. The functions that can handle various temporal information by using Htime web API also helped the research.

## 1. はじめに

### 1. 1 歴史ビッグデータ

現代のビッグデータ研究の根底にあるのは、データの大規模な収集と統合に基づき世界を復元して解析するという「データ駆動型アプローチ」である。このアプローチを過去に延長することで、過去の世界の復元と解析を実現することが「歴史ビッグデータ」研究の目的である。そして過去の天気や地震に関する自然の復元、あるいは江戸時代の人物情報などを用いた社会の復元など、歴史的記録に対してデータ駆動型アプローチを適用することにより、過去の歴史をよりよく知ることを目指す。

世界的に見ても日本は歴史的な記録が数多く残されている国である。一説によると20億点にも達する歴史的な文書が各地の図書館や旧家などに保存されており、潜在的には地域情報の宝庫として活用する余地がある。また江戸時代の出版業は世界有数の規模を誇り、高い識字率にも支えられて数多くの書籍が出版されてきた。こうした過去の日本の文化遺産の活用は重要な学術的課題となりつつある。例えば、国文学研究資料館を中心とした「歴史的典籍NW事業」[1]が、人文社会科学分野で初めての大型プロジェクトとして2014年から30万点の古典籍のデジタル化とオープン化を推進するなど、遅ればせながら研究データが充実し始めており、人文学分野もビッグデータ時代に入ったと言えるだろう。

そこで我々は、アクセス可能な史料から情報を抽出するというデータ駆動型のアプローチにより、過去世界の復元と統合解析に取り組む。ここで本研究が対象とする「データ」とは、過去に人間が書き記した記録であり、これを我々は「歴史的記録」と命名する。具体的には歴史的記録の分類として、以下の3種類を研究対象にすることを考えている。

【歴史的状況記録】自然の状況に関する記録であり、天気に関する記録や、地震の被害に関する記録など、古気候復元や古地震復元などで重要な記録である。

【歴史的行動記録】人間の行動に関する記録であり、人間の移動記録や観光地での購買記録など、生活や観光の復元などで重要な記録である。

【歴史的状態記録】人間やモノの状態に関する記録であり、家族の状態を表す戸籍や市場の取引を表す価格、モノの位置を表す輸送など、人口や経済の復元などで重要な記録である。

こうした記録を収集するだけでなく、統一されたスキーマ(データ形式)にしたがって構造化し、時空間や人物、イベントなどの固有名と紐づけることで、過去世界を情報空間に復元し統合的に解析するというのが最終的な目標である。

本研究が特に対象とするのは歴史的状況記録(Historical Situation Record, HSR)である。市野ほか[2]は、歴史学者以外の研究者にとって負担が大きい歴史資料の利用に向けて、HSR データを取りあつかう際の共通課題をまとめ、この課題を協働によって解決する HSR データベース構想を掲げた。

## 1. 2. HSR を含む資料の情報の現状

HSR のデータベースの構築には、翻刻からデータ構造化まで多くの課題があるが、そもそも「どの資料に目的の HSR があるのか?」を知ることが大変である。また、資料探査の後にも翻刻、データ構造化などの困難な作業を控えているため、手当たり次第に資料を探すより、HSR のほか、対象の時間(イベントの時期など)、空間(イベントの地域など)の情報が含まれている資料を効率的に探したい。

これまで HSR を利用する研究は分野ごとにバラバラに進められてきたが、研究で利用する HSR を含んだ古文書等は共通の場合が多い。例えば、石川日記[4]には、毎日の天気のほか、隕石や地震など、さまざまな記録が含まれており、各分野で独自にデータ化されている。HSR データベース構築は大変であるが、まずはこれまでの情報分野間で共有し、有効活用することで、いわゆる「二度手間」を減らすなど、資料探索や情報取得の負担軽減につながる。

しかし、古典籍や古記録などの研究者が利用する情報については、図書館情報学や人文情報学分野を中心に研究開発が進められている[3]。一方、

HSR に関する情報についての取り組みはほとんどなく、資料探索のためのデータベースもない。

## 1. 3. 目撃情報

研究者は自らが注目する記録以外は、目にしてきたとしても、その情報を発信したり共有したりすることはほとんどない。これまでの HSR の資料探索は、各研究者が自ら探索するほか、歴史学者やほかの HSR を利用する研究者などに相談し、情報を得るなどであった。これは、事件捜査などで、警察官が聞き込みをして情報を得る方法と似ている。

一方で目撃情報などを提供してもらう方法もある。同様に、生物多様性の情報サイトなどでも特定の生物についての目撃情報の提供を呼びかけたりしている[5]。加えて、これらの情報の提供者は、事件捜査や生物調査の専門家ではない上、その情報は不確かである。

HSR 利用研究においても、偽りの情報ではない限り、「天気の記録があった」といった漠然とした情報でも役立つことが多い。情報の不確かさへの対応は、事件捜査と同様に、研究者自ら情報を確認し、研究に利用すればよいわけであり、情報提供者はその分野の専門家に限る必要はない。

## 1. 4. HSR のパスファインダー

特定の主題に関する各種情報資源や探索方法を紹介・提供する初歩的なツールとして、パスファインダーがある[6]。各図書館などで、利用者に向けたパスファインダーとしてのサービスを行っている。また、国会図書館では特定の主題に関する問い合わせ一覧を WEB で後悔し検索可能としている[7]。本研究で目指すものはさまざまな HSR という特定の主題に関する各種情報資源や探索方法を提供するものであり、パスファインダーの一つとも言える。

そこで、資料に含まれる HSR に関する目撃情報を収集し、目的の HSR を含む資料へのアクセスを支援するパスファインダー的なシステムを構築することにした。現在、試験運用および改修を進め、公開を目指している。

## 2. 課題と要件

### 2. 1. 概要

本システムは HSR を利用した研究を支援するパスファインダーのような役割をするものである。

例えば、ある研究者が過去の出来事、大きな被害のあった台風の情報を得たとする。その台風について、その進路や被害などをより詳しく調べたいと思う。

そして、その情報が書かれている資料を探す。このような場合、本システムで、イベントの日時や場所のほか、「災害」、「天気」などのキーワードを利用し、この台風に関する情報が「得られそうな」資料一覧を得る。

その資料に関する情報を元に、実際の資料にアクセスする。本システムは資料に含まれる情報全てをデータとして提供しているわけではないため、史料を研究に利用するためには、翻刻やデータ構造化が必要な場合もある。一方、すでに誰かが整理していて、提供が得られる場合もある。このような情報は、「整理状況」として、原資料、翻刻済みといった項目で情報を得ることができる。

この研究者がさまざまな資料を利用し、研究成果を得たのちには、新たに得られた情報を、今度は情報提供者として HSR に関する情報を登録する。

以上が、本システムの利用イメージである。

## 2. 2. パスファインダーとしての要件

図書館のパスファインダーのようなサービスの実現のため、必要な機能を検討する。パスファインダーの作成の第一歩はトピックの選定であるが、まずは HSR またはその中の特定の主題 HSR とする。次に利用者の設定であるが、このシステムのユーザはさまざまな対象者が考えられる。本研究においては、著者らがテストユーザとなり、HSR を利用する研究者または場合により歴史気候学研究をする人を利用対象者とする。

### 元情報の取得

次にパスファインダーに利用する情報の収集である。パスファインダーは「通常、その図書館のコレクションやサービスを対象として作成される」のであるが、本研究では史料の利用者の目撃情報を利用しパスファインダーを作成しようというものである。そのため、パスファインダーを作成するための情報(これ以降、元情報とする)をどのように収集し、整理するかが大きな課題の一つである。

**パスファインダー**：主に図書館などの情報から作成される。情報提供する人は図書館員であり、データ作成は明確な規定のもとで行われる。個別の興味の偏りは少ない。あっても、職務なので平等に振る舞う。

**本システム**：目撃者による情報。目撃した人はさまざま。史料には詳しい人が多い。史料に対する自分の専門や興味にこだわりがあり、情報や視点、その精度に偏りがある。どんなものが情報として提供されるかがわからない。そもそも、どうやって、情報を集めかということが重要な課題である。一方で、利用者が増え情報が集まるにつれ、パスファインダーとして充実してくるという長所もある。

### 提供する支援方法

図書館のパスファインダーは、「個々の資料・情報資源が人為的に重み付けされた上で解題を付してリスト化され、調べ方に関する解説もなされているため、単なるリストやリンク集とは異なる」ものである。また、どのようなパスファイン

ダーにすべきかといった課題には、利用者のニーズを把握することが必要である。

本研究のユーザを著者らと想定し、レファレンス協同データベースを例に考える。データベースの検索画面のようなものは必要である。一方、結果として得られたリストも事例詳細は必要がない。結果としては事例詳細であるパスファインダーの編集に利用された資料一覧を得たい。リストにある「〇〇について書いた本はあるか」と同様に、「天保の飢饉を調べているが、1830年代の天気の詳細がある日記はあるか」といった質問を反映できる検索項目がほしい。

## 2. 3. 検索のニーズ

検索の際に重要なものはいくつもあるが、まず、目的の HSR の有無である。イベントや飢饉など、ある時間に資料があるかも知りたい。被害の広がりなどを研究したい場合は、面的に広がる地域の記録の有無も必要である。さらに、天気の記録であれば、どの程度毎日記録しているのかという記録の頻度、農家の史料であれば、農作物の種類や収量などの情報もあるとなお良い。そこで、天気や地震といった HSR の有無と時間情報の検索と、検索結果として、資料に関して収集された詳細な情報を取得できるようにする。

## 2. 4. HSR の目撃情報

1995年以降、都内の電車に乗ると「不審者や不審物を見かけたら・・・」という車内掲示を見る。どういうものが不審なのだろうか。多くは「怪しい」ペットボトルを爆発物や毒物と考えたり、様子が「普通に見えない」人が危ないのではないかと想像する。しかし、「怪しい」とか「普通に見えない」は非常に主観的であり、どんな状態をどう捉えるかは情報提供者それぞれである。

史料の情報という点から、ほとんどの記録者は故人であり、記録者や観測者以外の人からデータに関する情報を提供する。そのため、タイトルも故人である記録者ではなく、それを整理した人や所蔵者がつけることも多く、いろいろな名前を持つ可能性がある。翻刻が必要な場合、史料内容の詳細がわかるまでに多くの時間を要する。そのため、記録地や記録期間、さらにはさまざまな間違いなどの、研究データとして基本的な情報でさえも得ることは容易ではない。

本システムが利用しようとしている目撃情報は、対象が史料であることから、多面的な曖昧さを持っている。これらをどのように収集し、データベース化するか検討と工夫が必要である。

## 2. 5. 時空間情報

気候学や地震学など地球科学分野をはじめ、HSR を利用する研究では時空間情報は重要な要素の一つである。江戸時代以前の資料については、和暦(元号による表記)のものも多く、時間情報は和暦への対応が必要である。HSR を利用する研究者の多くは西暦を利用するが、図書館や資料

館のほか、市民の場合も含め、和暦、西暦のどちらを利用するかは、その目的や好みなどによりさまざまである。著者らはこれまでも文理融合プロジェクト等で、この暦のあつかいの違いによる分野間の障害なども経験した。われわれのニーズとして、和暦、西暦双方をあつかえる機能が必要である。

表 1: 主な登録項目

項目	備考: 説明
書誌タイトル*	資料の名称、シリーズ、1巻ずつなど、登録者の判断で入力
書誌タイトルほか	複数ある場合や、数巻をまとめて登録した際などに記録する
書誌情報ソース*	リンク先などや情報のソースを書く。史料およびそのメタデータへのリンクがある場合はそれを優先。所蔵および管理者、編者、作成者のみの場合は、それをこちらに記載。
登録者	情報を登録する人
史料の作成者	史料を公開できるようにした人または機関。書誌情報ソースと異なる場合。
史料の公開または管理者	所蔵および管理者が書誌情報ソースと異なる場合。
著作権	著作権に関する情報。システムのデータはあくまでデータカタログのため、登録された史料についての著作権の情報
観測地	歴史的状況記録が記録された場所
観測地の変遷	記録場所が異なる場合はその変遷を記述。例: 参勤交代で江戸にいた時期と藩にいた時があり、記録場所が複数ある
記録開始	和暦、西暦いずれも可
記録終了	和暦、西暦いずれも可
歴史的状況記録	どのような歴史的状況記録があるかにチェック。複数回答可。
チェックボックス	地震・天気・火山・災害・農業・その他
歴史的状況記録(記述)	上記、その他をチェックした場合は具体的な状況記録を記述。その他、詳しい状況記録の内容など。
記録の連続性	日記の場合、記録期間の日数における記録がある日数。毎日記録されているか、欠書があるかの情報
天気の詳細に関する情報	晴、雨などの簡単な記述か、時間など詳しい記述があるか
整理状況	資料の整理状況、翻刻済みかなどの情報を入力。複数の段階の資料へのアクセスが可能な場合は複数チェック可。
チェックボタン	原資料のみ画像・翻刻テキスト(紙媒体・電子媒体)・構造化テキスト(マークアップ・エクセルなどテーブル)・解析用データ・解析結果データ
その他	自由記述
資料の利用状況など	市町村史編纂や研究などで利用されている場合はそのCITATIONなど
その他	自由記述

史料によっては、未翻刻や全資料へのアクセスができていないことなどの理由で、記録開始年以外は明らかでないもの、HSRの研究では一部の情報のみの利用が多く、利用した部分の時間情報しかないものなどがある。しかし、ユーザである著者らは資料の記録された期間を調べたいと考えており、時間による検索機能が必要である。

### 3. システムと機能

#### 3. 1. 概要

本システムの主な機能は、目撃情報の登録とその検索という、非常にシンプルなものである。しかし、HSRを利用した研究に対するパスマインダーとして必要な検索機能、目撃情報を利用するためのデータベース構築にはいくつかの工夫が必要である。3章で、データベースの構築に向けた取組みを紹介し、4章で試験的な利用を通して見えてきた課題やそれに伴う改修について述べる。

#### 3. 2. 登録システム

パスマインダー作成のための情報として、目撃情報を利用するため、検索を意識した項目設定と目撃情報の曖昧さを考慮した入力方法が必要である。また、多くの目撃情報を得ることが、パスマインダーのデキにつながる。

まずは試験的な入力フォームの作成として、Google Drive[8]のGoogleフォームを利用し登録フォームを作成した。曖昧な目撃情報をフォームに落とし込むため、各入力項目名との対応は簡単ではない。そこで、具体例もつけて記載方法を添えた。また、フォームからデータを取り込んだスプレッドシートへ直接入力も可とした。ユーザによっては、表形式のフォームの方を好む場合もある。また、新たな項目が必要であった場合、Googleフォームの変更より、スプレッドシートの列を追加の方が容易い。

表1は主な項目一覧である。登録者が気づいたさまざまな情報を容易に提供できるよう、必須項目を極力減らし、自由記述欄を多くした。必須項目は登録者と書誌情報ソースだけである。書誌情報ソースは所蔵館やWEBサイトといったものから、登録者が所持しているといったものまで、資料に関する最低限の情報でも可としている。また、登録する単位(1レコードの単位)は、決めておらず、登録者の考えに任せている。

#### 頻出するHSRの項目

HSRの有無について、著者らがこれまでよく目にするHSR項目である天気、地震、災害、農業は、チェックボックスで登録する。一方、記述の信頼性、洪水、米の収量といった、さらに詳細な情報や、結氷や植物の変化といった登録者の興味や専門分野による事柄については、自由記述も可能である。自由記述欄で得られた頻出項目は、

今後チェックボタンなどの登録方法へと改修し、用語の統一を図っていく。

### 史料の整理状況の項目

史料が、原資料のみ、翻刻されているか、テキストはあるか、画像は公開されているか、といった史料の整理状況は、HSRの資料の情報として重要であるが、整理状況という項目名だけでは漠然としている。そこで、表1に示す項目のチェックボタンにした。さらに、自由記述欄も用意し、研究や市町村史などで利用され、整理されたといった詳しい情報も記述できるようにした。

### 記録期間および場所の項目

観測地とその変遷や記録の開始終了など時間に関する情報は、所蔵されている地点や、実際の記録地、西暦、和暦なども入力が可能であり、規則は緩い。時間情報のデータベースでのあつかいは、3.4節で述べる。

### 3.3. パスファインダーとしての検索と情報取得

登録されたデータはデータベースで管理し、以下の項目での検索、結果は資料の一覧表示、資料ごとに詳細表示が見られるようになっている。

#### 書誌タイトル

タイトルの一部「日記」などで検索可能である。

#### 時間（年月日）指定検索

記録時期についての以下の形式による検索が可能である。

- イベント型（1日指定）事象があった日
- 年代型（1880年代など）事象の発生した時代
- 元号型（和暦による指定）
- 年、年月のみでも検索可能である。
- 詳細は次節で詳しく述べる。

#### その他、位置情報など

現在は、記録地などの位置情報は文字列としてのあつかいである。また、自由記述として登録された情報は、資料ごとの詳細表示で取得できる。

### 3.4. 時間情報

時間情報については和暦・西暦双方への対応が必要である。そこで、入力データおよび検索データの西暦和暦変換と、データベースの時間情報のフォーマットの統一にHuTime[9]のWeb APIによる暦変換を利用した(図1)。それにより、登録時、検索時ともにさまざまな形式での時間記述、時間情報の曖昧さや不十分さに対応できた。

西暦は、気候モデルおよび地震・噴火史料データベースを参照し、「ユリウス/グレゴリオ暦・1582年改暦」を利用した。気候モデルはCFConvention[11]に従い、1582年10月15日からである。一方、地震・噴火史料データベース[12]は1583年初めからであり、厳密には同じではな

い。なお、本システムは日より細かい時間分解能については現時点ではあつかっていない。

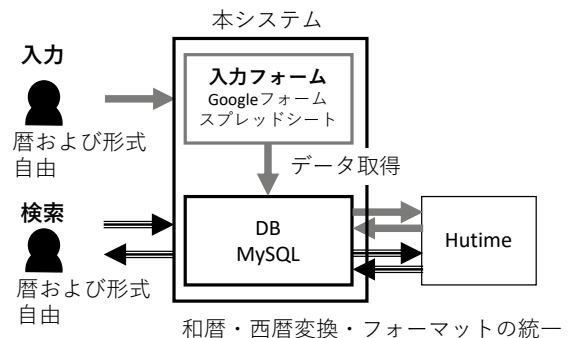


図1 本システムとHutime Web API

### 登録時の時間情報

登録項目において時間情報としてあつかうものは記録開始および記録終了である。それぞれ西暦と和暦の入力が可能である。また、異なるフォーマット、西暦、和暦の混在、記録開始または記録終了のどちらかのみも可としている。また、年のみ、年月のみ、元号のみなどはHutimeで年月日に変換している。年月日まで補完登録された入力値をtextとして保持し、期間検索用の西暦はdate型として保存する。

新規登録時と、修正や追加などの編集時における時間の入力値のあつかいについて、Hutimeを利用したデータベースでの和暦変換および期間検索用西暦を具体例で示す。

記録開始（西暦）：入力値[1852]  
→ 和暦[嘉永4年12月10日]  
期間検索性西暦[1852-01-01]

記録終了（西暦）：入力値[1864]  
→ 和暦[元治1年12月3日]  
期間検索性西暦[1864-12-31]

### 検索時の時間情報

検索は、特定年、年代などの時間指定も可能とした。検索の時間情報は、西暦、西暦（年代）、和暦のいずれかが利用可である。入力された時間情報はHutimeを利用し西暦に変換する。検索においても、年のみ、年月のみも可である。

## 4. 歴史気候学での利用事例

### 4.1. 開発メンバーによる試験利用

著者らは、本システムのテストを兼ねて、本システムを利用し、研究用の資料のデータカタログの作成および整理作業を行なっている。例えば、対象となる資料にはさまざまな形態があり、入力テストを繰り返して、項目などを検討している。

これまで利用した資料等の情報として、歴史天候データベースでの利用資料[11]、総合地球環境学研究所気候適応史プロジェクト(「中塚 PJ」とする)収集資料の一部[12]などが登録され、現在116件の資料情報がある。ここでは、この取り組みを通して得られた課題などを紹介し、今後の展開や改修方針などにも触れる。

## 4. 2. 本システムを利用した歴史気候学資料の整理

### 日記の存在期間一覧図の作成

水越允治元三重大学教授は1980年代から日記天候記録による気候復元推定の研究を進め、その関連資料を収集してきた。その資料および関連する資料は、著者にも参加している中塚 PJ に提供されたが、プロジェクトが終了間近となり、一著者の増田が首都大学東京に引き取り、整理を始めた。

資料のうちわけの主要な部分は、本や雑誌論文などの出版物(ここでは「刊本」と呼ぶ)、地方自治体などに所蔵された紙資料を水越教授が撮影したネガフィルムおよび印画紙(「フォトコピー」と呼ぶ)、水越教授が資料から毎日の天気を読み取った紙媒体の表(「ワークシート」と呼ぶ)である。

これまでの作業は、このうち刊本の日記の情報の整理である。出版物としての書誌情報とともに、記録されている日記が、だれが、どこで、いつからいつまで書いたものかを読み取り、本システムに入力した。入力した日付は和暦によった。本システム内でHuTimeの機能を使って西暦の日付が得られる。それを利用し、それぞれの日記が書かれた期間を線で表示し、一覧図をつくることのできた(図2)。ただし、日記の記録期間の連続性の確認はできていないので、図で線がつながっていても実際に記録があるとは限らないという問題が残っている。

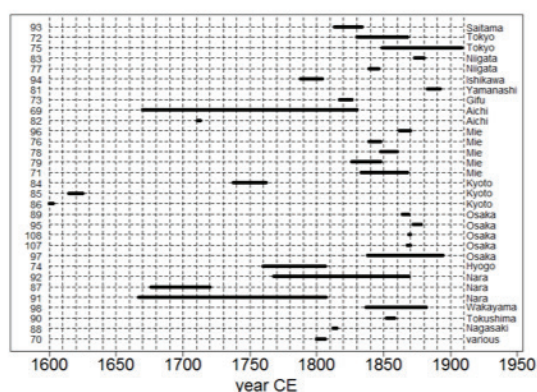


図2 資料の存在期間一覧図

左軸は資料のID, 右軸は記録地(都道府県)。

これまでの作業では現在の都道府県名, 市町村名を入れた。一覧図では、日記を都道府県別にまとめ、都道府県名を表示した。

### 日記以外の資料への対応

刊本のうちには日記以外もある。そのうち地方の年表や災害史は、時間の刻みは年単位であることが多いが、HSRを含む資料なので、同様に本システムへ入力を試みた。ただし、日記とは性格の異なる資料であり、図2には入っていない。資料の性格の判別は人(整理業者)による個別判断である。現在、本システムでは資料類別はなく、また、必要な種類も検討する必要がある、今後は、ある程度自動判断できる資料類別を改良したいと考えている。

フォトコピーの情報は、収納されている箱に書かれた文字列に対応するものが、水越[14, 15]の紀要論文の付録にあげた資料一覧にあるらしい、という認識できたところである。その確認のためには印画紙を1枚ずつ見る必要がある。今後その情報を本システムに登録する予定である。

フォトコピーの内容の多くは、未出版の手書きの資料である。中には、同じ人の日記が、一時期は刊本となっているが、残りの時期は未出版で、フォトコピーで補っている。HSRとして利用するには、翻刻が必要だが、本システムに登録した情報を提供し、「みんなで翻刻」などのシステムで史料が翻刻される流れを構築したいと考えている。

### 4. 3. 登録項目の再検討

登録項目の意味について判断のゆれが生じた。その対策として、項目を追加したところもある。本システムでは、項目等の再検討が起こることを前提に、Googleのフォームとスプレッドシートを利用している。この事例のように、必要に応じて登録者がその都度項目案をあげ、試行的に項目の追加修正しながら進めている。

### 史料の公開または管理者

「史料の公開または管理者」には、刊本の場合は、出版者を入れた。未出版資料の場合は、所蔵する文書館などを入れる予定である。刊本の所蔵者、刊本のもとになった原資料の所蔵者をどうあつかうか、また、今後フォトコピーの情報を入れる場合はネガや印画紙の所蔵者をどうあつかうかという問題が残っている。

日記の刊本の場合、日記を書いた人と、出版できるように編集した人は異なる。「史料の作成者」に日記を書いた過去の人、「編纂者」に編集した現代の人を入れることにした。しかし、刊本になった文献には、現代の人が書いた文章中に過去の記録が引用される形になったものもある。この場合、書誌情報上の著者となる現代の人を「史料の作成者」とすべきか「編纂者」とすべきかの判断が難しい。今回は「編纂者」とし、引用された過

去の記録が単一でないときは「史料の作成者」は空白とした。

### 書誌情報ソース

「書誌情報ソース」の項目には書誌情報そのもの、書誌情報へのリンク、書誌情報が含まれるリストなどがありうる。書誌情報が既存の図書目録データベースに収録されているものはそれにリンクすれば、目録データベースにある刊本の所蔵情報が利用できる。ここで作業した刊本の日記については、「書誌情報ソース」に CiNii Books[16] の書誌ページの URL を置くこととした。今後は国会図書館などと書誌情報の紐付けを検討する。

この作業における実際の書誌情報源は、水越資料に含まれていた刊本自体であった。しかし、このことは「書誌情報ソース」というよりもむしろ、資料がコレクションに含まれているという事実を記録しておくべきと考えた。そこで「資料シリーズ」という項目をたて、「水越資料」であることおよびそのうちの作業用整理番号を記した。

### 整理状況

「整理状況」の項目は、刊本の日記がいずれも定型表現の「翻刻テキスト（紙媒体）」とすればよい。しかし、さらに、翻刻されたものがどのような形の出版物になっているか（単行本か雑誌記事か、活字であるか手書き謄写版であるか、など）の情報も添えておきたいと考え、「整理状況の補足情報」という項目を追加した。

## 5. 今後の課題

共同研究の資料整理に利用しながら、項目の追加やその解釈も議論し変更するなどが行われ、それに伴うデータベースなどのシステム改修も続いている。そのため、著者以外のテストユーザへの公開が遅れている。本章では改修予定および今後の課題について述べる。

事例でも課題としてあげた位置情報については、まず歴史地名のあつかいの課題がある。そのほか、実際の記録は記録者の住居とは限らず、旅行などの移動時にも記録される可能性があり、曖昧な部分が多い。しかし、所蔵先や記録者の住居の位置などのいずれかを、地図表示する機能を構築する予定である。

目撃情報は主観的であり、多くの点で曖昧さを含んでおり、同じ対象物に異なる情報が提供されることもあるだろう。また、史料解読が進むにつれ情報は増えていく。研究に利用した場合は論文などの研究発表情報の提供が望まれる。そのため、図書館の書誌情報に比べて、多くの修正や追加が発生する。また、異なる情報を並列して保持する必要も出てくる。すでに登録されている資料に新たな情報が追加されることもある。史料解読は研究者、史料の専門家や資料館員、市民などさまざまな人により進められるため、一つの資料に対して複数の人から情報が追加される可能性も高い。

このようなことを想定し、修正機能のほかユーザ管理のための機能も必要となる。ユーザの展開にあたり、ログイン機能の追加、入力フォームも新たに作成する予定であるが、ユーザの権限や、修正方法、情報の履歴や管理などについては、多角的な検討が必要であり、HBD の課題として進めていく。

## 6. 利用ユーザの拡張と改修

### 6. 1. 他分野のユーザ

著者は、滋賀県立琵琶湖博物館[17]との共同研究に先立ち、館所属の研究者より、琵琶湖周辺の近世の資料をこのシステムで検索できないかという問い合わせを受けた。研究員の専門はフナやコイの消費量の変動である。幸いにも4の一覧図があり、コイ、フナといった HSR の記録は不明だが、1600 年以降の滋賀県の資料の有無は確認できた。やはり、地域を指定した検索が必要だと痛感し、とりあえず、利用事例と同様に所蔵先ないしは記録地の都道府県を入力することにした。

### 6. 2. 資料館の登録ユーザ協力

一部の資料を所蔵している機関、大学図書館員らは、資料をより多くの人に利用してもらうための試行錯誤を続けている。広島県の頼山陽史跡資料館[18]の石川良枝氏は、書誌情報カタログの作成業務を進めながら、従来の書誌情報だけではなく、HSR を利用する研究者の目に止まるような情報も含められたら、資料の利用者が増えるのではないかと考えていた。本システムを紹介したところ、広島県の文化財課の佐伯氏にも本システムのテストユーザとしてご協力いただけることとなった。現在、利用マニュアルの作成も始めており、改修が一段落した段階で、広島県との協働作業を進める予定である。今回のつながりを活かし、今後は提供者側のメリットを考慮した、協働で構築するシステムを目指していく。

### 6. 3. 研究ネットワークによるシステム連携

HBD はまだ始動したばかりであるが、プロジェクトへの発展のため、新たな研究グループの創設を目指し、第6回 CODH セミナー「歴史ビッグデータ～過去の記録の統合解析に向けた古文書データ化の挑戦～」を開催した[19]。8名の講演者と、96名の参加者があった。著者らの予想を超えて、「歴史ビッグデータ」という新しい研究への関心が高いことがわかった。それを機に人文科学オープンデータ共同利用センター (Center for Open Data in the Humanities, CDOH) は「歴史ビッグデータ研究会」[20]をスタートし、多種多様な分野からの参加者によるネットワークができつつある。

HBD に関連する研究集会も開催され、本システムの情報を HBD で進めているデータ構造化や、「みんなで翻刻」[21]などのシステムで利用することを検討している (図3)。例えば、翻刻の対

象の資料選択や本システムのデータを資料のメタデータとして利用するなどである。まずは、「整理状況」に、IIIF で公開されて場合はマニフェスト URL を登録できるように改修する予定である。そのほか、人口学などの歴史資料のデータベースとの連携も生まれつつある。

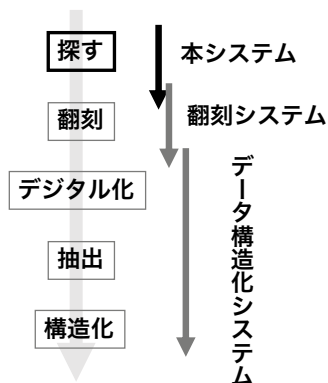


図3 史料の構造化における本システムと他のシステムの関係

## 7. おわりに

本研究では、史料に関わるさまざまな人々からの目撃情報を利用し、HSR を利用した研究のパスファインダー的なシステムの構築を紹介した。また、著者らによる利用事例を紹介し、本システムの特徴や課題を示した。本システムは多分野の研究者や資料館の職員の協力により、少しずつ価値を高めている。

本研究の背景には1章で紹介したHBDという大きな取り組みがある。HSR データベース構想をはじめ、HBD に必要な技術の多くは、機械学習によるくずし字OCRなど、情報学の発展が必要であり、実用化までには年月がかかる。

本システムは、既にあるデータベースの技術やAPIなどを利用し、研究における負担の軽減と研究の推進を目指すものである。じんもんこん2017やHBD研究会ほかさまざまな関わりの中で、本システムがより役立つ可能性が見えてきた。研究者に限らず、図書館や資料館などの職員をはじめ、市民の中にも史料についての知識が豊富な方が多い。このようなさまざまな人の参加により、情報や利用者が増えるとともに、充実したパスファインダーが提供できるようになると期待している。

## 参考文献

- [1] “日本語の歴史的典籍の国際共同研究ネットワーク構築計画”。  
<https://www.nijl.ac.jp/pages/cijproject/> (参照 2018-10-31)。  
 [2] 市野美夏, 増田耕一, 北本朝展, 平野淳平, 庄建治朗: 人文情報学の素材としての歴史気候学の経験, じんもんこん2017 論文集, 2017, 139-146  
 [3] 吉賀 夏子, 只木 進一, 伊藤 昭弘. 小城藩日

記データベースの構築. 研究報告人文科学とコンピュータ, 2018, vol.117, No.3, p.1-7.

[4] 八王子市郷土資料館編. 石川日記(七), 郷土資料館資料シリーズ第24号, 1985, p.122

[5] “広島県: ★生物多様性総合情報サイト★～広島県の生物多様性を知ろう!～”

<https://www.pref.hiroshima.lg.jp/site/tayousei/investigation-biodiversity-wanted.html/> (参照 2018-09-03)

[6] 日本図書館情報学会用語辞典編集委員会 (編): 図書館情報学用語辞典 第4版, 丸善出版, 2013.

[7] “レファレンス協同データベース”  
<http://crd.ndl.go.jp/reference/> (参照 2018-10-31) .

[8] “Google Drive” <https://www.google.com/drive/> (参照 2018-09-03) .

[9] “Hutime”  
<http://www.hutime.jp/index.html> (参照 2018-09-03) .

[10] “CF Conventions and Metadata”  
<http://cfconventions.org/index.html/> (参照 2018-10-31)

[11] “[古代・中世]地震・噴火史料データベース (β版)”

<https://historical.seismology.jp/eshiryodb/manual.html> (参照 2018-10-31)

[12] “総合地球環境学研究所気候適応史プロジェクト” <http://www.chikyu.ac.jp/nenrin/> (参照 2018-10-31)

[13] “歴史天候データベース”  
[http://tk2-202-10627.vs.sakura.ne.jp/index\\_hw.html/](http://tk2-202-10627.vs.sakura.ne.jp/index_hw.html/) (参照 2018-10-31)

[14] 水越 允治: 近畿・東海地方における17世紀以降の梅雨期の長期傾向, 皇学館社会福祉学部紀要, 2000, No. 3, p165-176.

[15] 水越 允治: 近畿地方における過去六百年間の梅雨の長期傾向, 皇学館社会福祉学部紀要, 2001No. 4, p15-31.

[16] “CiNii books” <https://ci.nii.ac.jp/books/> (参照 2018-10-31)

[17] “滋賀県立琵琶湖博物館”  
<http://www.biwahaku.jp> (参照 2018-10-31)

[18] “頼山陽史跡資料館”  
<https://www.pref.hiroshima.lg.jp/site/raisanyou/> (参照 2018-10-31)

[19] “人文学オープンデータ共同利用センター: 第6回CODHセミナー「歴史ビッグデータ～過去の記録の統合解析に向けた古文書データ化の挑戦～」”

<http://codh.rois.ac.jp/seminar/historical-big-data-2018-0312/> (参照 2018-09-03).

[20] “歴史ビッグデータ”  
<http://codh.rois.ac.jp/historical-big-data/> (参照 2018-09-03).

[21] “みんなで翻刻 - 歴史災害資料のオンライン翻刻プロジェクト” <https://honkoku.org/> (参照 2018-09-03) .