

特集「デジタル・ヒストリーの諸実践」に寄せて

『クリオ』編集部

『クリオ』31号巻頭に、「歴史研究者のためのデジタル道具箱」という小論が掲載されたことがある。これは、小風尚樹氏が歴史研究に活用できるデジタルツールを紹介したものであるが、それからわずか3年の間に、歴史研究におけるデジタル技術の活用、いわゆるデジタル・ヒストリーの認知と実践は大きく進んだように思われる。『歴史情報学の教科書』のような書籍の出版、学術誌『西洋史学』におけるデジタル・ヒストリー特集、さらに若手研究者を中心とする Tokyo Digital History により2度にわたって開催されたシンポジウムなどは、そうしたデジタル・ヒストリーの広がりを示す例のごく一部である。いまやデジタル・ヒストリーが射程に収める対象は洋の東西、国内外を問わない。歴史研究者は地域や時代による区分を超えて、歴史研究にデジタル技術を活用し、あらたな成果を生み出す方法をともに考え、共有していくことになるだろう。このような「方法の共有」は、デジタル・ヒストリーもその一角を占めるデジタル・ヒューマニティーズにおける、「方法論的共有地 Methodological Commons」という概念によっても示されている。本特集は、この「方法の共有」の可能性を示すために、西洋・東洋・日本を幅広く対象に含めるとともに、あえて狭義の歴史研究に限ることなく、資料の管理や研究支援ツールの開発、研究者の協働といった多様な論点を扱っている。

冒頭では、歴史研究の土台ともいえる資料・情報管理の、デジタル時代における在り方について岡崎敦氏に論じていただいた。これはいわば全体の導入ともなっている。次に、早くからデジタル・ヒストリーの可能性を探求してきた藤川隆男氏には、これまでのデジタル・ヒストリーとの関わり、そして現在携わっている研究プロジェクトの内容についてご紹介いただいた。これは、歴史研究におけるビッグデータ活用の可能性、そして大規模な協働に基づくデジタル・ヒストリーの在り方を示唆する論考となっている。以上2つの論考に続いて、各分野における若手研究者が、各々の関心に基づいて進めている研究内容についての小論を配した。これらはいわば各論とも言うべきものであり、資史料の構造化、テキスト分析、地理情報分析など多様な事例におけるデジタル技術活用の方法を論じている。各分野における取り組みと方法論を知るとともに、若手研究者によるデジタル・ヒストリー実践の可能性を感じていただければ幸いである。

本特集が、デジタル・ヒストリーという分野のごく一部を示すに留まることは間違いない。しかしそれでも、ここに収めた示唆に富む諸論考を通じて、上で述べた、分野の壁を越えた「方法の共有」の一端を示すことはできたのではないかと思う。それゆえ読者諸兄に、ここで示された論点や方法が自らの研究分野においてどのように適用されるのかを考えていただけたならば、本特集は意義あるものとなるだろう。

最後に、本特集は2020年2月に開催された2020 Spring Tokyo Digital History Symposiumでの報告をもとに構成されたものである。このシンポジウムの内容はYouTubeで配信(<https://www.youtube.com/watch?v=G1zCMf0MGkU&t=11890s>)されているので、論考の内容に興味を抱かれた方はご視聴いただければ幸いである。

執筆者とタイトル一覧（ページ数）

岡崎 敦	「デジタル時代のアーカイブズ学と文書学」	119
藤川 隆男	「公共圏の歴史的構造」	126
佐治 奈通子 / 中村 覚	「歴史学と情報学のより良い協働を目指して」	133
田野崎アンドレア嵐	「証書研究における TEI マークアップの活用」	135
濱野 未来	「歴史学における計量テキスト分析の活用」	137
大邑 潤三	「歴史災害研究における GIS 活用の試み」	139

デジタル時代のアーカイブズ学と文書学

岡崎 敦

はじめに

アーカイブズ学と文書学は、ともに「資料・情報管理」を対象とし、特に後者は長い伝統を誇る学問分野であるが、デジタル時代において新しい展開を見せている。ここでは、その一端を紹介するとともに、この2つの学問がいま経験している課題が、同様に変容のさなかにある歴史学にとって、どのような意味があるのかについて考えてみたい。

1. アーカイブズの現代的変容

国際アーカイブズ評議会の語彙集によれば、「アーカイブズ」とは、「個人、家族、あるいは公的・私的の如何をとわずなんらかの組織が、その業務の過程で産出、受領し、その後、それらの資料が含む情報が持つ永続的な価値のゆえに、あるいは資料の作成者の機能や責任の証拠として、保存される資料の塊。特に、来歴、本来の秩序および共同コントロールの原則のもとに維持されるもの」とされる¹。ここで念頭に置かれる資料とは、権利証書あるいは組織の業務遂行の過程で生み出される行政内部資料などであり、近代国民国家においては、文書館という施設が最終的には管理するものと観念される。主観的な意見を恣意的に表明する、多くは大量のコピーの一つを収集する図書館資料と対立するものとして、これらの資料は、特定組織のルーティンな業務の過程で「有機的」に生成、蓄積されるオリジナル資料の塊で、一般には、業務的・証拠的価値とともに、歴史的価値をもつものとみなされている²。

人間が、個人であれ、集団であれ、社会生活の秩序を守りながらなんらかの行為を行うためには、それらを社会が相互承認する仕組みが必要であり、この意味で、法行為あるいは法的事実の証拠の確保は、人類の歴史のはじめから存在したであろう。事実、アーカイブズの歴史においては、権利の確保、適正な業務の証拠の保存や管理のあり方が研究されてきた。フランス革命における近代的な文書館制度の形成においても、少なくとも当初目指されたのは、国民主権にもとづく国民国家政府による市民への現用行政文書の公開であった。また、現在にいたるまで、アーカイブズ資料管理の根幹を形成する、いわゆる「フォンド維持原則」、つまりアーカイブズ資料は、図書館において採用される方法論的分類ではなく、資料を管理してきた（作成ではないことに注意）組織における資料「群」が内在的に有する秩序を維持するという原則も、本来は、19世紀半ばに、役

¹ 第1項と第4項参照。 <http://www.ciscra.org/mat/mat/term/64> (2020年3月25日最終確認)。

² アーカイブズ概念、および歴史については、DURANTI, L. and FRANKS, P. C., ed., *Encyclopedia of Archival Science*, Lanham, 2015の関係項目が役立つ。日本語では、以下の拙稿、およびそこで提示した参考文献リストも参照。岡崎敦「アーカイブズ、アーカイブズ学とは何か」『九州大学附属図書館研究開発室年報 2011/2012』, 2012, pp. 1-10。アーカイブズの歴史については、以下の文献も参照。ブノワ＝ミシェル・トック（岡崎敦訳）「西欧中近世におけるアーカイブズ（6-18世紀）」『古文書研究』65, 2008, pp. 71-88; 岡崎敦「フランス革命とアーカイブズ」『歴史学研究』984, 2019, pp. 57-66。

所の現局から文書館へ移管されたばかりの資料の取り扱いに関して出された省庁内の業務命令に由来するものであった。アーカイブズ資料の価値は、本来的に、コンテンツではなく、法的・業務的価値、つまり定まった規則、手続きにしたがって作成、管理されたものであることに存するとされ、その本質は、当該資料のアーカイブズの真正性（自称するものであることの証明）とよばれる。

他方、19世紀以後の近代歴史学は、科学としての歴史学の基盤として、年代記等の記述資料と異なる「客観的なナマの事実」を表す史料としてアーカイブズ資料を発見し、文書館が歴史家の巢窟になるにつれ、文書館＝歴史文書の収蔵庫という誤解が広がった。さらに、20世紀なかばには、特にアメリカにおいて、業務の現場で日常使用される「レコード」と、現用的価値を失って本来は廃棄されるはずの資料が「歴史学的な観点」から保存される「アーカイブズ」を対立させる理論が生まれ、特に北米と日本に強い影響を及ぼした。

しかしながら、20世紀末のICT環境の飛躍的發展と、政府や組織の社会的説明責任の強化の動きの中で、法行為および業務証跡としての情報資源の意義が高まるとともに、アーカイブズの認識自体が、理論、現場への実装双方において、根本的な変容を受けた。この動きを代表するのが、オーストラリア学派のレコードキーピング理論である³。ここでは、電子環境のもとでの、とりわけボーンデジタル記録について、現局で資料が生成されてから、最終的に廃棄あるいは永久保存されるすべての過程を、統一的に把握する必要性が強調される一方で、デジタル文書においては、伝統的にアーカイブズの基礎を構成していたオリジナルなどの概念が根本的に動揺すること、さらに、情報の価値自体が業務のコンテキストやプロセス、さらには関係者によって異なる意義や価値を与えられる構造などが論じられてきた。このような新たな認識を理論化したものとして著名なのが、いわゆるレコード・コンティニューム理論である⁴。ここでは、行為、主体、証拠、レコードキーピングの4つの座標軸のもとで、資料が個別の資料として発生し、組織内で共有された記録として補足され、さらに組織の共同記憶として組織化され、最後に組織を越える社会全体の中で多元化的な価値付けを受ける4つの次元が表象される。ここでは、資料とその前提となる主体や行為を取り巻くコンテキスト、およびその時間的な位置づけの変容プロセス自体が記述の対象として要請されているのである。

ところで、かつてはレコードマネジメントの、最近では統合されたアーカイブズ管理全体における必須要件として提示されてきた、資料の真正性、信頼性、完全性、利用可能性の4要件を⁵、デジタル環境においてどのように保証するかについて精力的に取り組んできたデュランチは、最近、デジタル記録（アーカイブズ）フォレンジックを提唱し

³ レコードキーピングについては、その代表的著者たちによるマニフェスト的著作の一部がつい最近邦訳された。McKEMMISH, S., PIGGOTT, M., REED, B. and UPWARD, F., ed., *Archives: recordkeeping in society*, Wagga Wagga, New South Wales, 2005（安藤正人他訳『アーカイブズ論』明石書店, 2019年）。

⁴ Cf. UPWARD, F., *Structuring the Records Continuum - Part One: Postcustodial principles and properties*, in *Archives and Manuscripts*, 24-2, 1996, pp. 268-285; Id., *Structuring the Records Continuum, Part Two: Structuration Theory and Recordkeeping*, in *Archives and Manuscripts*, 25-1, 1997, pp. 10-35.

⁵ Cf. ISO15489-1:2011 Information and documentation — Records management — Part 1: General ; 2016年に第2版が公表され、メタデータの重要性がさらに焦点化された。

ている⁶。そこでは、デジタル環境における資料管理、さらには業務管理を担うアーキビストは、資料あるいは業務の適正性それ自体を、法廷において強い立証能力を有するとみなされる証拠として管理することを使命とするとされるが、実はそこで議論の鍵を担ったのが文書学であった。

2. 文書学 *diplomatics*, *diplomatique* の拡張

文書学とは、17世紀末に、ベネディクト会修道士マビヨンによって著された著作 *De re diplomatica* のタイトルにその名を由来する学問である。国際文書学委員会編纂の「語彙集」では、「文書学」は「文書の伝来、形式および作成を研究するもので、その目的は、資料の批判、正しさの評価、テキストの質の吟味、書式部分と資料固有の内容の弁別、日付の確定、最終的には資料の刊行である」、「文書資料 *sources diplomatiques*」は、「あらゆる個人あるいは法人による法行為、および行政財務活動の結果生じる資料、および職権に基づいて発給され、なんらかの規則に基づいた形式を有する書簡。もっとも重要な部分は、アーカイブズ資料によって構成される文書資料は、一方で、文芸資料や歴史著作と、他方で、非文字資料と対立する」、さらに「文書 *acte écrit*」とは、「法行為の完遂、あるいは法的事実の存在、あるいはなんであれそこに法的有効性を与えるなんらかの形式のもとで作成されていれば、なんであれ事実が記載された書き物」、とそれぞれ定義されている⁷。つまり、文書学とアーカイブズ学の対象は事実上同じであるが、そこには、歴史的経緯も介在していた。

マビヨンの著作および文書に関する学問的体系の提示は、実は、ベネディクト会修道院が持つ権利証書、具体的にはメロヴィング王によって発給された王文書の偽作、つまり法的効力を否定する主張に対抗し、現用文書の法的価値を守るためになされた。ただ、マビヨンは、個々の文書の法的効力の立証という要求に対し、メロヴィング王を始めとする過去の大量の文書の体系的な検討と、そのなかでの個別資料の性格特定という方法論の開発によって応えたのであり、それはデカルトの原理、つまり網羅と列举、体系化を基盤としていた。そして、その具体的な検討対象は、当時の文献学と同様、内容のもっともらしさではなく、言語や文体、さらにはレイアウトや書体などからなる「形式」であったのである。19世紀中葉以降、歴史学が科学を標榜して、客観的な情報の批判的な吟味を志向した際、文書学の方法論は、丁度テキスト研究全般において文献学が果たした役割を担った。つまり、究極のオリジナル＝歴史的事実の確定であり、それは、コンテンツではなく形式、とりわけ「外層（資料の物理的特徴）」に関する専門的知識と個々の特徴を判別する特殊技能によって担われた。フランス文書学校が、「文書＝アーカイブズ資料」の研究として一体であった文書学の研究・教育を旗頭として、アーキビストと

⁶ DURANTI, L., From Digital Diplomatics to Digital Records Forensics, in *Archivaria*, 68, 2009, pp. 39–66; Id., The return of diplomatics as a forensic discipline, in A. AMBROSIO, S. BARRET and G. VOGELER, ed., *Digital diplomatics. The computer as a tool for the diplomatist?*, Köln/Weimar/Wien, 2014, pp. 89–98. Cf. ROGERS, C., art. Digital Records Forensics, in L. DURANTI, ed., *Encyclopedia of Archival Science*, *op.cit.*, pp. 166–170.

⁷ CARCEL ORTI, M. M., ed., *Vocabulaire international de la diplomatie*, Valencia, 1997, I-1 (Diplomatique), -2 (*sources diplomatiques*), -3 (*acte écrit*).

文書学者を養成してきたのも、この意味で当然と言える⁸。

しかしながら、このような状況は、文書の法的価値、およびオリジナルの確定と真偽の判定という狭い目標に、文書学を限定させる傾向を有していた。特にドイツにおいて顕著であったが、王や教皇という特殊な公権力が発給した多かれ少なかれ定式性が高い文書の真偽や作成過程に関心が集中した。研究の対象が、比較的古い時代に、つまり文書数が少ない時代に集中したのも、文書数が爆発的に増加する13世紀以降は、もはや偽文書の判定は意味がなく、また形式的特徴が研究（研究者の好奇心や知的能力）に値するほど多様で顕著ではなくなることも影響を与えたであろう。

これに対して、20世紀中葉以降、文書学の対象の拡張が叫ばれるようになった⁹。第一に、対象となる資料類型に関してである。マビヨンにおいてもまったく無視されていたわけではないが、初期の文書学のもっぱらの関心は、発給者が一人称で現れ、特権付与などを内容とする単葉の権利証書であった。文書の偽作がもっとも盛んに行われたのが、特に古い時代のこの種の権威の文書であったからであろう。しかしながら、文書＝アーカイブズ資料のなかには、行政内部の役人間の連絡や通知などの行政書簡に加えて、会計簿、議事録、徴税台帳、目録など、いわゆる帳簿系資料をはじめとする行政内部資料も大量に伝来している。ここで問題となるのが、文書学およびアーカイブズ学における真正性の概念の再構築である。単葉の権利証書における文書学的真正性は、書式はもちろん、レイアウトや書体、なにより署名や印章などの有効性保証の記号などによって保証されたが、行政内部資料については、多くの場合、このような示標が欠けているように見える。ここでの真正性は、資料が、行政部局のルーティンとして定められたやり方のもとで作成され、「管理」されているという事実自体にあると考えねばならないのである。第二に、文書学の対象の限定が取り払われ、「資料の形式」ではなく、業務の適正な運用が判断の基準となれば、時間的な限定が無用となる。つまり、従来好んで研究の対象となってきた、12世紀以前の古い時代だけではなく、ルーティンでの資料作成が普通となり、多様な資料類型の文書が大量に作成されるようになった中世末期¹⁰、近世、さらには近現代文書＝アーカイブズ資料もまた、文書学の方法論の適用が可能な領域とみなされるようになった¹¹。第三に、文書の発給手続きのみならず、伝来過程自体が特別な関心の対象となった。20世紀末の人社系諸学問を共通に襲ったのがいわゆる構築主義の嵐であり、歴史学においては「記憶」の問題系が一世を風靡したが、テキスト論、メディア論においては、これに先立って、資料の価値は、生成ではなくむしろ「受容環境」に依存し、多様なかたちで存在するコピーや「版」の意義を、オリジナルとは別個に評価

⁸ 文書学については、拙稿「文書形式学」、高山博・池上俊一編『西洋中世学入門』東京大学出版会、2005年、pp. 59-74。さらに、以下の文献は、本稿の観点からも重要である。DURANTI, L., *Diplomatics: New Uses for an Old Science*, Metuchen, 1998 (first published, in *Archivaria*, 28-33, 1989-1992).

⁹ Cf. GUYOTJEANNIN, O., The Expansion of Diplomatics as a Discipline, in *American Archivist*, 59, 1996, pp. 414-421.

¹⁰ Cf. BERTRAND, P., *Les écritures ordinaires: Sociologie d'un temps de révolution documentaire. (entre royaume de France et Empire, 1250-1350)*, Paris, 2015.

¹¹ Cf. GUYOTJEANNIN, O., Avant-propos pour Exportations de la Diplomatie, II, Documents contemporains, in *Bibliothèque de l'Ecole des chartes*, 161-2, 2003, pp. 495-500.

しようとする動きが強まっていた。文書＝アーカイブズ資料についても、究極のオリジナル復元や生成の手続き論よりも、むしろある文書が、生成とは異なるコンテキストのもとで、どのような取り扱いを受けるのか、新たな価値付けの対象となるのかなどの諸問題が議論の対象となったのである¹²。

しかしながら、最大の衝撃はむしろ情報学の発展から来た。20世紀中葉から盛んに行われていた、統計学的手法によるテキスト分析やデータベース解析、画像デジタル・アーカイブを活用した資料原本のヴァーチャル比較などの重要性ももちろんであるが¹³、ここでは、文書＝アーカイブズ学の関心や方法論自体に関わる問題について検討しよう。

3. デジタル時代の文書学とアーカイブズ学

デジタル環境のもとで一新された現代アーカイブズ学において、本来は中世初期の特殊な権利証書の真偽判定のために開発された文書学の方法論が本質的な重要性を有する、とする主張のチャンピオンはデュランチだが、同様の意見は、特に大陸系のアーカイブズ学研究者には珍しくない¹⁴。ここでは、いくつか論点を指摘したい。

第一に、デジタル環境のもとでは、文書＝アーカイブズ資料は、そのものとしては実存しない、つまり、なんらかの介入によって「その都度」システム内で生成され、可視化されて、場合によっては介入を受けるなにものかである。常に流動的で、物理的な実体が同定不可能な溶解性を持つ一方、紙媒体に比べてはるかに介入・操作が容易であり、さらに、その使用可能性が、基本的には、短期的に更新されるシステムに依存するからには、定期的にマイグレーション、すなわち別のなにものかへの変換が不可欠である。このような条件下においては、資料や業務の適正さの保証のためには、紙媒体がそうであったような、いわば放置しておいてもそこにあるような、物理的な証拠の実在を前提とすることはできず、資料の生成から廃棄、利活用までのすべての段階で受ける介入すべてを、証跡（痕跡）としてトレースする必要がある。つまり、情報管理の責任者は、資料を事後的に収集、整理するのではなく、業務のあらゆるプロセスに事前介入せねばならないのである。レコードマネジメントと非現用文書管理とは統合されるが、前者は事実上、文書学の方法論が適用されるべき領域であるからには、文書学とアーカイブズ学とは、かつてそうであったように再度一体化することになる。

第二に、文書学＝アーカイブズ学の根本的原理として、個々のコンテンツではなく、形式に代表されるような、なんらかの資料「群」に共通の諸特徴の重視がある。つまり、ここで重要なのは、資料個々ではなく、情報の「集合体」の性格の管理であり、アーカイブズ学における複数レベル記述がこれに対応する。ここでは、ある資料の真正性や

¹² Cf. GUYOTJEANNIN, O. and MORELLE, L., Tradition et réception de l'acte médiéval: Jalons pour un bilan des recherches, in *Archiv für Diplomatik*, 53, 2007, pp. 367–403.

¹³ Cf. VOGELER, G., ed., *Digitale Diplomatie. Neue Technologien in der historischen Arbeit mit Urkunden*, Köln/Weimar/Wien, 2009; AMBROSIO, ed., *Digital diplomacy. op.cit.*

¹⁴ Cf. CHABIN, M. A., Peut-on parler de diplomatique numérique?, in V. FREY et M. TRELEANI, ed., *Vers un nouvel archiviste numérique*, Paris, 2013, pp. 105–129; DELMAS, B., De la diplomatique contemporaine à la diplomatique numérique: retour sur l'expérience française, in L. GARNON-ARQUIN et M. LAJEUNESSE, ed., *Panorama de l'archivistique contemporaine : évolution de la discipline et de la profession*, 2015, pp. 261–287.

業務の適正さの保証は、「同じ特徴を有する」資料や行為のシリーズのなかでの位置づけの評価による一方で、個々の資料や行為を発生させるコンテキストが問題となってくる。他方、ある資料や業務過程における複数の介入の追跡は、同一と観念される業務を構成する一連の行為や資料の集合体として把握される必要もある。ここで重要なのは、このようなコンテキストとプロセスの複層的・複合的な記述＝メタデータによる管理には、デジタル技術が前提とするディレクトリ型の情報整理がよく適合することである。事実、最近提出されたアーカイブズ管理に関する新しいメタデータ標準は、資料それ自体というよりも、資料とそれに関わる主体や行為の総体を要素として、それらの間に紡がれる多様な関係を主語と述語で記述する、いわゆる「実体＝関係モデル」からなっている。そこでは、資料という実在ではなく、資料が世界との関係で持つ機能に関心があるともいえる¹⁵。

第三に、社会との関係である。電子環境の拡張と「民主化」は、当局や企業に対する情報公開の圧力を強めたり、ブロックチェーン技術を活用した貧困者への経済的手段の提供への寄与等を可能とする一方で、フェイクの拡散、個人・機密情報の漏洩や不法な利活用の拡大などの社会的問題を同時に招来しているが、本来両者は同じ盾の両面である。このような状況のもと、情報コンテンツ（行為）の善し悪しそれ自体ではなく、情報の性格の保証を旨とする文書学とアーカイブズ学は、情報の社会的共有のための基盤としてあらためて注目する価値がある。ことは、単に業務におけるアカウントビリティにとどまらない。コンティニューム理論が射程に入れているように、デジタル環境において拡散される情報資源は、当事者にとどまらず、世界中の多様な主体の介入によって、その都度新たに意味を与えられるが、いわゆる「文化遺産」としての価値付け問題は、このような理論的構造のもとで理解され、適正に運用されねばならない。

おわりに

本稿では、デジタル環境のもとでのアーカイブズ学、文書学の変容について論じてきたが、これらの議論は、古典的な意味での歴史学研究には無縁であるかのようにみえるかもしれない。最後に、この点について論じて、本稿を閉じたい。

第一に、本稿で論じたようなアーカイブズ学、文書学における問題関心と方法論の変容は、20世紀末以来の歴史学、ひいては人文諸科学全般において共通に認められてきた動向と共通点がある。事実の構築主義的理解、コミュニケーション・受容環境論、紛争解決研究、そして言うまでもなく史料論と呼ばれる一連の研究があげられよう。特に、歴史学固有の方法論の練磨にとって、史料や歴史情報の性格とその取扱いの方法の吟味は決定的に重要な意味を持つ。文書学とアーカイブズ学は、ながらく歴史学と密接な関

¹⁵ 国際アーカイブズ評議会が検討中の「コンテキストのなかの記録。アーカイブズ記述のための概念モデル」Records in Context (RiC). A Conceptual Model for Archival Description (2016)を念頭においている。この標準、およびこれと密接な関係にあると考えられる図書館情報学における資料記述標準との関係については、拙稿「情報管理専門職の人材養成問題：職務標準、メタ情報標準の動向からみるアーキビストのミッション」『九州大学附属図書館研究開発室年報 2017/2018』, 2018, pp. 1-7.

係を保ってきたが、むしろ今ほど本質的な重要性を有することはなかったといえる¹⁶。多様な個別の網羅的収集と分析によって、社会が共有する法や規則を「発見」して体系化し、それを新たに見いだされた個別の検討によって再評価するという方法論の原則は、情報やテキスト理解の基盤に他ならない。情報の真正性他の管理は、研究データやプロジェクト管理そのものでもあり、研究の妥当性の評価、さらには方法論自体の根本的吟味も包含する射程を持つ点も強調しておきたい。

第二に、デジタル時代の情報管理、特に真正性保証問題は、歴史学的素養と密接不可分な関係にあるように思われる。ある社会においてあらかじめ承認された規則や手続きへの準拠をもって真正性の要件と考える文書学、アーカイブズ学の方法論は、意味を発生させ受容させる諸条件が、恣意的でありながら強制力を持つ記号体系としての社会のあり方を研究するという意味で、相対主義的価値観の上にある。社会は多様であり、それは異文化として理解する努力を必要とするということ、そしてそれは中長期的な時間の変化の相で理解すべきという見方は、デジタル時代の歴史学研究にこそふさわしい。短期的視野による恣意的な価値の一般化に対抗する視座として、歴史学からの貢献可能性をあらためて確認しておきたい。

最後に、21世紀の情報管理の現場が、情報を意味づける環境の拡張、つまり新たな利活用と価値付けに満ちていることが挙げられる。世界遺産を始めとする「文化遺産」はその端的な表れであるが、情報の価値をめぐる問題は、個人や家族、民間企業や地域社会など、あらゆる場で共通に提起されるダイナミックな性格をもち、当事者とよそ者の区別も再定義される必要がある。欧米では、アーキビストの養成制度が整備されているが、重要なのは、近年そのカリキュラムのなかに「公共歴史学」的要素、つまり過去に関する社会の認識に関する教育が組み込まれつつあることである。歴史学の専門性を磨いた学生が、その能力を社会において認められるためにも、歴史学の「公共的」拡張は不可欠であり、それはデジタル時代にはかつてないほど求められているが、情報管理の現場こそ、その前衛に位置するのではないだろうか。

¹⁶ 以下の拙稿は、資料の生成と伝来状況から、その前提となる一連の行為と社会秩序のあり方の検討を試みたものである。現代の情報空間は、近代のそれに対立するかたちで、むしろ中世と類似性を持っていることに注意したい。岡崎敦「11世紀北フランスに文書史料の危機はあったか」『西洋史学論集』37, 1999, pp. 1-21; 同「西欧中世における記憶の管理とアーカイヴズ ―パリ司教座教会のあるカルチュレールをめぐる(『Liber Niger』)―」『史淵』146, 2009, pp. 57-89; 同「王のモノグラマンが付与された11世紀のシャルトル司教文書」『史淵』154, 2017, pp. 69-93.

公共圏の歴史的構造

——自然言語処理による新聞データの分析を通じた

19-20 世紀オーストラリアの公開集会と世論形成の構造の解明——

藤川 隆男

はじめに

本稿では、すでに行われた学術的研究の成果を明らかにするのではなく、歴史研究において、情報技術、あるいはデジタル・テクノロジーを用いることで、どのような刺激的な成果を新たにもたらしることができるのか、あるいはできないのか、その見通しを述べたい。適確な見通しを述べる能力が筆者にあるかどうかは少々疑わしいが、早い時期からこうした情報技術の応用に関心を抱き、現在も共同研究者とともにデジタル・ヒストリーに取り組む者として、1つの参考意見を述べていると考えていただきたい¹。

以下では、これまでの取り組みを簡単に紹介した後に、現在、デジタル・テクノロジーを利用して取り組んでいる研究課題について述べる。とは言っても、今のところ共同研究者が自然言語処理の手段を活発に開発し、学生が必要な正解データを作成するのを見守っているだけの役割しか果たしていないが、プロジェクトのリーダーとして、一応工程表を作成し、研究の方向を決めているのは確かである。最後に、私が期待するデジタル・ヒストリーの今後の展望を述べて本稿を締めくくりたいと思う。

1. これまでのプロジェクトの限界と教訓

1999 年に、インターネット講義を配信。2003 年度からは、オンラインで検索できるオーストラリア辞典とオーストラリア年表の運用を開始。2005 年には、オーストラリアの町を紹介するサイト Net Kelly を作成した。2007 年には、地図を利用した英語のデータベース the Ghostly Gazetteer of Australia を公表した。これは現在では用いられなくなったオーストラリアの歴史的地名の辞典である。また、2003 年度には、InDesign を用いた DTP による新しい雑誌『パブリック・ヒストリー』も創刊した²。

こうした試みは、西洋史研究室の学部学生と大学院生の協力とデータベースのプログラムを作成してくれる他学部の学生の存在に依存していた。学生は興味がなければ来て

¹ Arguing with Digital History working group, "Digital History and Argument," white paper, Roy Rosenzweig Center for History and New Media (2017): <https://rrchnm.org/argument-white-paper/> はデジタル・ヒストリーの様々な展開と射程を簡潔にまとめている。また、併せて Cohen, Daniel J., et al. "Interchange: The Promise of Digital History." *Journal of American History* 95, no. 2, (2008): pp. 452–491 を読めば、デジタル・ヒストリーにとっての課題がどのように変化してきたか、10 年近く未解決の問題が理解できて、参考になる。デジタル・ヒストリーの展開については、山中美潮「アメリカ史研究とデジタル・ヒストリー」、『立教アメリカン・スタディーズ』40、2018 年、7–31 頁の前半部がわかりやすく、まとまっている。

² 以下の URL を参照。インターネット授業: <http://www.let.osaka-ac.jp/seiyousi/brcast/index.html>; オーストラリア辞典・年表: <http://bun45.sakura.ne.jp/>; the Ghostly Gazetteer of Australia: <http://www.let.osaka-u.ac.jp/seiyousi/Ghost-Gazetteer/index.htm>; 『パブリック・ヒストリー』: <http://www.let.osaka-u.ac.jp/seiyousi/info-1.html>。

くれないし、辞典と年表のプログラムを書いてくれた学生、現在は奈良先端科学技術大学院大学の准教授となった石尾隆氏の協力を得られたことなど、10年近くこの種の活動は続いたが、その基盤は偶然性に依存し脆弱であった。研究に対する組織的なサポート（学部・大学レベル）が皆無であったばかりか、研究資金の目途もなく、この種の活動への評価も存在せず、一旦手仕舞いすることにした。

この時点では、アメリカのデジタル・ヒストリーに大きく遅れていたとは思わないが、撤退して、興味を持った別の活動に取り組むことにした。組織的サポートがなければ³、よほど情報技術に心酔している人でない限り⁴、個人として、継続的に行える領域ではなかったと思われる。人的・技術的・資金的リソースが不可欠の領域であることは明らかであった。

この時期のプロジェクトの一例だけ少し詳しく述べると、2005年から2007年に the *Ghostly Gazetteer of Australia* を制作した。これは、オーストラリアの失われた地名を地図上で検索できる英語のサイトである。最初に、私の先生であった故 D.W.A. Baker 氏の承諾を得て、氏が作った失われた地名の手書きのカード約 2600 枚の内容を、エクセルのスプレッドシートに項目別にすべて書き写して、デジタルデータ化した。20人以上の学部生と大学院生が参加するプロジェクトであった。次にこれらの地名と説明を *Geoscience Australia* からダウンロードした地図上に配置した。今は亡き *FireWorks* を利用して、オーストラリアの失われた地名に近似する位置にある都市を書き入れて、その都市をクリックすると失われた地名が現れるようにしたのである。現在では GIS が発達し、データの地図的表現が簡単にできるようになったが、当時はそういうシステムが利用できなかったのも、ずいぶん手間のかかる作業になった。*FireWorks* の利用については、当時大学院生で、現在は清風高校教員の米田誠氏が力になってくれた。地図を用いたデータベースの公開も、大規模な英語のデータの公開も、ネットで検索可能な本格的な歴史辞典や年表の公開も、歴史分野では最先端を進んでいたように思う。

デジタル・ヒストリーに回帰し始めたのは 2014 年である。オーストラリア辞典・年表を大学外のサーバーに移行し、石尾氏に新しい検索装置の開発を依頼。年表も紀元前を入力・表示できるタイプに改良してもらい、幅広く活用できる準備を整えた。2015 年度には久しぶりに辞書の比較的長い項目を追加した。さらに、代表になった雑誌『西洋史学』では、遅ればせながらデジタル化を進めた。2018 年度、授業もデジタル・ヒストリーに回帰した。回帰というよりも、デジタル・ヒストリーの様ざまな新しい道具について知識を持ちたいと思った。ゼミの学生とともに、*Exploring Big Historical Data: The Historian's Macroscopic* を読み始めた⁵。

この間、オーストラリアの地方の歴史博物館の研究では、25000 枚以上の写真を撮影しては、その処理と公開方法の問題に直面し、『妖獣バニヤップの歴史』（刀水書房、2016）では、歴史的新聞データベース、トローブ Trove（1955 年までのオーストラリアの主要

³ 研究センターやデジタル・ヒストリー学科のような、研究・教育組織。

⁴ 目立ちたい気質も必要。

⁵ See Graham, Shawn, Ian Milligan and Scott Weingart, *Exploring Big Historical Data: The Historian's Macroscopic*, London: Imperial College Press, 2015; see also URL: http://www.themacroscopic.org/?page_id=584.

な日刊紙と地方新聞を網羅したオーストラリア国立図書館の新聞データベース⁶⁾を用いて、約 14 万件、新聞記事だけでも 6 万件を超えるデータを活用する必要が生じた。デジタル・テクノロジーの有効的な活用なしに、研究活動を進められない状況になってきたと実感した。

2. 公共圏の歴史とパブリックなものへの関心

現在進行中のデジタル・ヒストリーのプロジェクトの説明に入る前に、このプロジェクトが解明しようとするテーマに対して私が抱いてきた関心について、手短かに説明したい。公共圏と言えば、ユルゲン・ハーバーマスを想起するほど、彼の影響力は大きい。他の分野と比較すると、歴史学の分野ではあまり注目されているようには思われない⁷⁾。その原因は、ナンシー・フレイザーが指摘しているように、彼のモデルがかなり強硬な「自由主義的なブルジョワ公共圏」の理念型であったからであろう。ジェンダー、階級、エスニシティを排除し⁸⁾、さらに宗教や娯楽の果たした役割を軽視してしまうと、近年の歴史的モデルとしてはポンコツである。また、ハーバーマスは、その後の研究でも、批判を受け入れつつ、初期の理念型を抜本的に修正しなかった。それも影響したと思われる。いずれにしても、ハーバーマスが熟議的な民主主義が成立する理想的な公共圏を歴史的過去に求めていたのは間違いない⁹⁾。

私自身は、最初の研究テーマ「白豪主義」を推進した反中国人運動に関心を持っており、その運動が新聞とパブリック・ミーティング（公開集会）によって推進された点に着目していた。当時の世界で最も民主的な社会において、公共圏における民主主義の根幹を支える二つのシステム、とりわけすべてのパブリック（市民や住民）¹⁰⁾に開かれた公開集会が差別を推進する原動力になった点に関心を抱いた。つまり公共圏の重大な欠陥に焦点を合わせたのである。ハーバーマスとは真逆と言っていいだろう¹¹⁾。

こうした観点から、約 35 年前に公開集会の研究を始めた。1871–1901 年の 2000 以上の公開集会の事例を *the Sydney Morning Herald* 紙から集めて、篩にかけ、この時期のニューサウスウェールズ植民地の首府シドニーにおける、公開集会による世論形成の構造

⁶⁾ Trove, digitized newspapers: <https://trove.nla.gov.au/newspape> を参照。

⁷⁾ ユルゲン・ハーバーマス著（河上倫逸、M・フーブリヒト、平井俊彦訳）『コミュニケーション的行為の理論』上中下、未来社、1985–1987；ユルゲン・ハーバーマス著（細谷貞雄訳）『公共性の構造転換』未来社、1973；ユルゲン・ハーバーマス著（細谷貞雄、山田正行訳）『公共性の構造転換：市民社会の一カテゴリーについての探究』第2版、未来社、1994；ユルゲン・ハーバーマス著（高野昌行訳）『他者の受容：多文化社会の政治理論に関する研究』（叢書・ユニベルシタス；803）、法政大学出版局、2004 など参照。

⁸⁾ ナンシー・フレイザー「公共圏の再考：既存の民主主義の批判のために」（山本啓・新田滋訳）『ハーバーマスと公共圏』未来社、1999 年、117–159 頁参照。

⁹⁾ 日本の「公」概念の問題については、佐々木毅、金泰昌編『日本における公と私』東京大学出版会、2002 及び佐々木毅、金泰昌編『欧米における公と私』東京大学出版会、2002 や拙稿「『パブリック・ヒストリー』とは何か。」「『パブリック・ヒストリー』17, 2020, 12–24 頁参照。

¹⁰⁾ パブリックの多重性については、Dean, David M. ed., *A Companion to Public History*, John Wiley & Sons, Inc.: Hoboken, NJ, 2018, pp. 3–4.

¹¹⁾ See Fujikawa, Takao, “Public Meetings in New South Wales: 1871–1901”, *Journal of the Royal Australian Historical Society* 76, Part 1, (1990): pp. 45–61.

を明らかにしようとした。しかし、研究の範囲を時間的・空間的に拡大するには、人手や資金がまったく不十分であることは明らかであった。

他方で、社会運動の歴史的誕生に強い関心を抱いていたチャールズ・ティリーは、多数の研究者の助力を得て、1758–1834年の間にロンドンと周辺部において、民衆の国民的な公的領域への参加が暴力的なものから平和的なものに変化し、その手段としての公開集会が劇的に増加したことを証明した。しかし、ティリーと研究協力者たちによる実証的研究も、より包括的ではあるが、地域的・時代的に限られており、その後、人手や資金を費やして19世紀中期以降の公開集会の全国的な実態を明らかにしようとする研究は現れなかった¹²。

ところが、デジタル・ヒストリー、というよりもデジタル技術の発達によって、大きな転機が訪れた。人手や資金をデジタル技術によって代替すれば、夢見ていた大規模な研究を実現するだけでなく、35年前には進めなかった世論形成の構造のはるかに複雑な分析も可能ではないかと考えるようになった。

3. プロジェクトの起動

この研究、科研のタイトルでは「オーストラリアの世論形成の歴史的解明：自然言語処理による公開集会データの分析」の一つの特徴は、既存のデジタル・ヒューマニティーズの技術を個々の歴史研究に応用する研究ではなく、必要に従って様々なデジタル技術を駆使し、場合によってはアルゴリズムを独自に作成するところにある。そうした構想を当然のごとく抱いた背景には、最初から自分の活動に必要なプログラムを書いてもらっていた経験があった。

研究は前述のトロープを利用し、1803–1955年の間の公開集会に関する全広告データ約39万件を集め、基礎的なコーパスを構築し、さらに約128万件の新聞記事から付加的な情報を集める2段階から構成される。

こうしたデータ収集と整理において、この研究のもう一つの特徴がある。それは、膨大な人手と資金により、既存の手稿資料やPDF・画像ファイルからテキストを人間が写し取り、機械的に読み取り可能な形でデジタル化するというプロセスを経るのではなく、単にスキャンされただけで、文字起こしなどによって記号化されていないデータを、自動的に電子化・記号化してコーパスを構築し、容易に利用できるようにするという特色である。もし成功すれば、一般的に機械的に処理できるデータ量は爆発的に増加し、多くの分野で、これまで扱えなかったレベルの大規模な歴史的現象を容易に分析できるようになる。本研究の手法は、他国の研究や他のテーマの研究にも応用可能であり、新聞を用いた近現代史の歴史研究の手法を一変できるだろう¹³。今回、このプロジェクトで

¹² See Tilly, Charles, "The Rise of the Public Meeting in Great Britain, 1758-1834", *Social Science History* 34, 2010, pp. 291–299; Tilly, Charles, "Social Movements as Historically Specific Clusters of Political Performances", *Berkeley Journal of Sociology*, 38, (1993-1994): pp. 1–30; Tilly, Charles, "Britain Creates the Social Movement", CRSO Working Paper No.232, University of Michigan, 1981; ティリーの構想の展開を簡便に見ることができる。

¹³ 次の論文に研究の詳しい内容は記述されている：藤川隆男, Chenhui Chu, 梶原智之, 長原一「歴史研究におけるビッグデータの活用：オーストラリアを中心に」『西洋史学』268, 2020年、50–61

お世話になったのは、大阪大学データビリティフロンティア機構(IDS)であった。デジタル・ヒストリーの本格的プロジェクトの実行には、情報技術を多面的に提供する IDS のような機関や情報人文学研究科のような組織的サポートが必要であろう。

4. 研究の目的と内容

研究は大きく分けて三つの部分からなる。

(1) オーストラリア全土における公開集会による世論形成の構造と変化を広告から明らかにする。

連邦運動や反囚人輸送運動、道路の建設運動のような一過性の世論形成のプロセスが全体として世論形成の構造を形成すると仮定し、その構造と変化を解明する。広告から集める主要なデータは、件数、開催曜日・時間、場所、開催要求者、招集者と招集される者、目的などである。時系列データマイニング技術などにより、データの単純な個別的・時間的变化だけでなく、データ間の関係を見ることで構造的連関が明らかになる。例えば、開催曜日・時間、場所、招集者の組み合わせからは、時間と空間に対する招集者の支配的な権力行使がわかり、世論統制の仕組みの一つを分析できる。

(2) 公開集会によって形成された社会的ネットワークの分析。

広告から取り出したデータからその関連記事を探し、記事からデータを収集することで、参加者数、主要な参加者氏名、重要なテーマ群、賛否の動向、女性やマイノリティの参加など広範な情報を集める。画期的な点は、最先端の SNS マイニング技術などを利用することで、世論形成に関わった人間の社会的ネットワークを検証する点である。インターネット上でのネットワークは広く研究対象になっているが、過去の世界における世論形成のネットワークを構造的に再現できることで、現代の研究手法を歴史世界に適用する可能性が開かれる。

(3) 市民的公共圏の実態の提示。

世論形成の構造を膨大な資料を用いて長期的に分析することで、19-20 世紀の市民的公共圏の実体と性質を明らかにし、歴史学の側から市民的公共圏にアプローチする。19 世紀後半から 20 世紀前半には、新聞と、参加自由な公開集会が世論形成装置の根幹であった。それは、既存の新聞やラジオやテレビと、SNS に代表される特定の組織に属さない人びとが形成するソーシャルメディアが併存する状況に似ている。この研究では、階級、ジェンダー、人種、宗教、娯楽などのすべての要素を含んだ可変的なモデルを提示することで、現代社会理解にも貢献できるようにするつもりである。

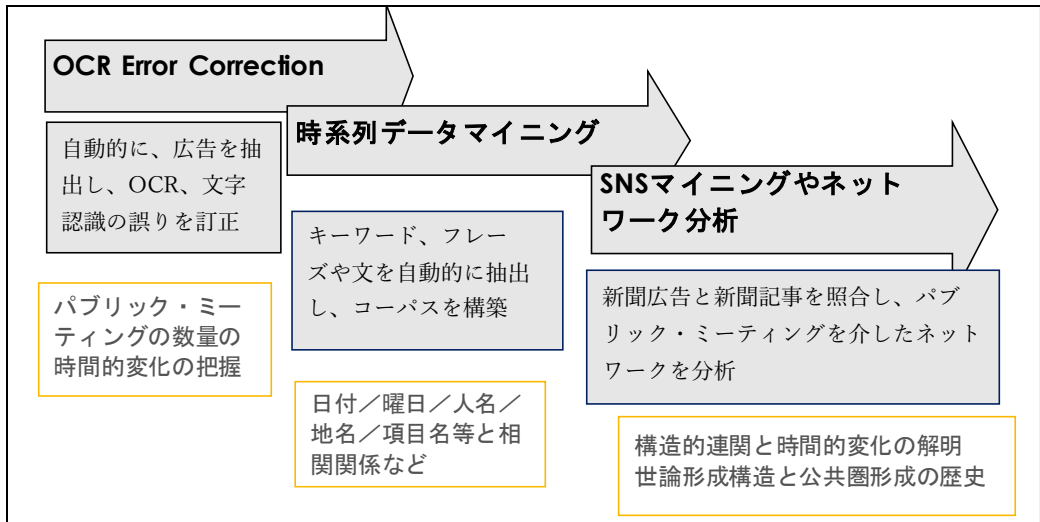
デジタル・ヒストリーが歴史学として評価されるためには、単なるテキスト分析や図像化・映像化などの手段の提供だけでなく、人文社会科学がこれまで問うてきた大きなテーマに対して、一般化が可能で斬新なモデルを提供できることが重要になるだろう。

5. 研究のモジュール化

プロジェクト型の研究においては、参加する研究者がどのような形で研究成果を共有するかが問題となる。その解決方法は、研究のモジュール化に行きつくように思われる。

頁。

図 1 公開集会研究プロジェクトのモジュール化



図一1を見てほしい。一番上の矢印は、研究に進展に伴って用いる主なデジタル技術の手法、中段はその内容の説明、下段はそれに基づく歴史的評価・考察などである。矢印は自然言語処理の専門家、下段は歴史研究者の担当であり、研究の進展にしたがって、それぞれが与えられたモジュールにおいて成果を発表する。ただし、勝手に活動しているわけではない。例えば、現在は左のOCR訂正の段階にあるわけだが、歴史研究者は、正解データの作成、デジタル化された地名辞典の提供、関連する固有名詞辞典の提示などによって、OCRの読み取り精度の向上や、自然言語処理の高度化に貢献している。また、第2段階への準備として、2000枚ほどの公開集会の手書きのカードをデジタル化して、次の段階の基礎資料として提供した。自然言語処理が進展しなければ、歴史研究者には活躍の場がない。しかし、膨大で詳細なデータを集めたとき、それを解釈し、構造的なモデルを完成させるには、広大な地域に関する詳細な歴史的知識が不可欠になる。研究対象のあらゆる側面に関心を持ち、それに生涯を喜んで捧げるような唯一無二の歴史家、そのうえ人文社会科学が直面する多くの課題に関心を持ち続ける研究者こそ、デジタル・ヒストリーの時代に生き残る歴史家だろう。

今のところ研究成果としては、OCR訂正を担当するChenhui Chu助教とRAの田中昂志が、言語処理学会やじんもんこんで報告しただけだが、COVID-19がなければ、今年5月に、マルセイユLRECや日本西洋史学会全国大会でのシンポジウムでも報告するはずであった¹⁴。歴史研究者としては、研究がもう少し進まないと手持ち無沙汰な感じがするが、業績リストは増えており、業績スクレーピングには成功しているのかもしれない。

¹⁴ 田中昂志, Chenhui Chu, 中島悠太, 武村紀子, 長原一, 藤川隆男「歴史新聞データからのコーパス構築」、言語処理学会第25回年次大会, 2019年; Chu, Chenhui, Koji Tanaka, Haolin Ren, Benjamin Renoust, Yuta Nakashima, Noriko Takemura, Hajime Nagahara, Takao Fujikawa, "Public Meeting Corpus Construction and Content Delivery", 人文科学とコンピュータシンポジウム, 2019.

おわりに

当面、研究としてのデジタル・ヒストリーは、トピック・モデリングのような他分野の技術を利用し、量が多いが均質で単一の資料に依存するだけでは、多種多様な史料を用いて、幅広いコンテキストのなかで考察する従来型の歴史に、大きな影響を及ぼすのは難しいだろう。歴史家が単独で利用するコーパスを構築する場合には、量的側面での優位性さえも怪しくなる。ビッグデータを活用するデジタル・ヒストリーは、圧倒的な史料の量という面でアナログ・ヒストリーを凌駕し、大きな成果も期待できるが、膨大な資金と労力が必要であり、今のところ広大な歴史世界に点在する小島のように、大きく歴史認識を変える力はないように思われる¹⁵。

しかし、快く思うかどうかは別にして、歴史研究のやり方、いや一般的に社会的な知の在り方そのものが急激にデジタル化しているのは事実である。従来型の歴史がこの動きと主体的に関わらず、デジタル・ヒストリーを他分野からの借用品として無視したり、あるいは観客になって珍獣として愛でたりするだけでは、歴史学自体の衰退を加速化するだけではないだろうか。日本の歴史学はデジタル技術を含めた社会との関係を再構築すべき時期に来ている。パブリック・ヒストリーが日本でもようやく認知されるようになってきたのは好ましい傾向であるが、大きな変化が起きているとは思われない。

既存の歴史学は、資料のデジタル化や単なるデータベースの拡張を超えて、歴史研究の手法、歴史的認識の在り方、つまり歴史哲学の問題も含めて、デジタル・ヒストリーをその一部として取り込まなければならない。デジタル・ヒストリーを歴史学に包含しなければならない。そのためには少なくとも既存の枠組みのなかで、

- (1) デジタル・ヒストリーを歴史教育の一部にすること、
- (2) 歴史研究（コンテキスト）を主体とするデジタル技術を開発すること、
- (3) 評価の枠組みを改変し、デジタル・ヒストリーを評価できるようにすること、
- (4) 論文や書物での成果発表、そのレビュー、国際的発信などの場自体をデジタル化すること、

が必要である。また、制度的には、すでに述べたように情報技術を多面的に提供する IDS のような機関や情報人文学研究科のような教育・研究機関が不可欠である。

私自身は微力ではあるし、退職まで5年と時間もないが、大学教育、雑誌『西洋史学』の改革、今日紹介したプロジェクトの推進、希望として言えば高大連携の斬新な企画や歴史博物館やスポーツ史のプロジェクトなどで、この方向への道を探りたい。

※本研究の一部は JSPS 科研費 JP19H01330 の助成を受けたものである。

¹⁵ See “Digital History and Argument,” pp. 18–20; 拙稿「歴史研究におけるビッグデータの活用」『西洋史学』268、61頁。

歴史学と情報学のより良い協働を目指して

——オープンな DH 支援ツールを用いたオスマン・トルコ語
文書群のデータ整理の一事例——

佐治 奈通子・中村 寛

本発表では、歴史学研究者と情報学研究者による共同研究について、背景、具体的な研究事例、協働の成果を紹介し、両分野による共同研究を実現・継続していく上で重要なことについて述べた。

まず本共同研究の背景として、互いの分野が抱える問題と関心、両分野の研究者をつなぐ場について述べた。概して、情報学研究者は、情報学の知識と自ら開発したシステムを持っており、それらを活用できる他分野での潜在的ニーズの掘り起こしや、開発システムに対するフィードバックを求めている。そのためには、データとフィードバックを提供する他分野の研究者との共同研究が必要であるが、その機会を得ることは簡単ではないという問題を抱えていた。一方、歴史学研究者は、歴史学の知識と研究対象である歴史史料を持っており、史料のデジタル化やそれに伴う研究プロセスの効率化、オンラインでの公開を見据えたデータ作成などに関心がある。しかし、新たにデジタル技術を学ぶための時間や費用を捻出できない、デジタル技術を自らの研究にどう活かせばよいか分からない、などの問題を抱えていた。このように、自分野の問題を乗り越えるために他分野に対する関心を持つ研究者が、互いに出会うことが協働の始まりであった。具体的には、東京大学人文情報学拠点が開講する人文情報学概論の授業と、大学院生を中心とする人文情報学の勉強と実践の場である Tokyo Digital History の活動が、本共同研究の出発点となった。

本研究で課題としたのは、歴史史料の画像データを、開発した DH 支援ツールを用いて整理・読解・分析することであった。対象は、クレシェヴォ修道院（在ボスニア・ヘルツェゴヴィナ）所蔵のオスマン・トルコ語文書である。本史料群は、1,000 点を超えるまとまった画像データであり、管見の限りこれまで体系的な整理・読解がなされていない新出史料であることから、本共同研究で一から検討する対象として適していた。

また開発した DH 支援ツールは、歴史研究プロセスにおける史料管理および分析を支援することを目的としている。具体的には、「1. Omeka を用いた画像とメタデータ管理」「2. TEI を用いた史料本文の記述」「3. 可視化による史料の整理・分析支援」の 3 機能を提供する。「1. Omeka を用いた画像とメタデータ管理」では、Omeka（オープンソースのコンテンツ管理システム）を用いることで、史料画像を IIIF（International Image Interoperability Framework: 画像共有のための国際規格）準拠の形で管理・利用可能とする。「2. TEI を用いた史料本文の記述」では、TEI（Text Encoding Initiative: 人文学資料の共有を意図した構造化ルール）と Oxygen（TEI/XML の編集ツール）を利用して、史料本文の記述を行う。TEI を利用したデータ作成の利点として、テキストの構造を残すことができる、多様なデータ活用が可能となる、といった点が挙げられる。作成した TEI ファイルは GitHub（データのホスティングサービス）を用いて管理すること

により、ファイルのバージョン管理に加え、ユーザ間におけるデータ共有を行う。「3. 可視化による史料の整理・分析支援」では、TEI で記述したテキストデータと、Omeka に格納した画像やメタデータを関連づけて可視化することにより、研究者への気づきの提案や、仮説の検証支援を行う。

以上の DH 支援ツールを用いて、①画像データの管理とメタデータの作成による史料整理、②TEI による史料本文の記述と情報抽出による史料読解、③それらに基づく可視化と史料分析という 3 つの作業を行った。個々の具体的な作業内容については紙幅の都合で割愛し、以下では協働の成果と、両分野による共同研究を実現・継続していく上で重要なことについて述べる。

まず、協働の成果のひとつとして、歴史研究プロセスの効率化が挙げられる。DH 支援ツールを利用することにより、ひとつのシステム上で史料の整理・読解・分析という 3 つの作業を行うことが可能となったため、個々の作業の進展と作業同士の連携が促された。従来、これらの作業を Word や Excel など別々のソフトを用いて段階的に進めてきた研究者も多いのではないだろうか。しかし、DH 支援ツールを用いて画像管理から分析までを一元化することにより、たとえば転写を作成しながら可視化機能を使って史料全体の性格や傾向を把握することが可能となった。また、TEI を用いて史料本文を作成することにより、データを、国際的な基準に沿った、様々なアプリケーションに適用可能な、長期保存に向く形式で作成することができた。

また、人文情報学に取り組むハードルが下がったことも、協働による大きな成果であった。本来、自分の研究に有用なアプリケーションを探索し、使い方を習得するまでには試行錯誤を要するが、共同研究により、情報学研究者の専門的なアドバイスを受けながら早い段階で適切なアプリケーションを知り、扱い方を学ぶことができた。その分、史料読解やデータ作成などに十分な時間を割くことができたことは、大きな利点であった。本共同研究により、必ずしも歴史研究者が情報学の知識を持ち合わせたオールラウンダーでなくても人文情報学に取り組むことができる、という可能性を見出せた。

一方、もちろん、歴史学研究者が情報学研究者に一方的に手伝ってもらっただけでは、積極的・長期的な共同研究は成り立たない。協働関係維持のためには、常に互いの研究にとっての Win-Win 関係とはどのようなものかを、共同研究者との間で考え共有していくことが大切なのではないだろうか。本共同研究においては、歴史学研究者の研究上の悩みや実現したいことをフィードバックとして提供することが、情報学研究者のシステム改良や今後の開発のヒントにつながった。たとえば、印影画像分析のための機械学習技術、文書の生成過程・人物関係分析のためのネットワーク分析などの研究開発を現在進めている。このように、共同研究で得られた知見が、それぞれの分野においても新たな研究成果につながることは、最も重要な Win-Win 関係のひとつであろう。

また、互いの分野へ関心を持ち、尊重し合うことは非常に大切である。初歩的ではあっても相手の分野の知識を学び理解を深めることで、共同研究への関わり方や提案できるアイデアの質が変わってくる。それぞれの専門分野を掛け合わせ、より研究を加速させアイデアを広げていくためには、やはり、分野を超えたひとつの研究プロジェクトとして、課題や目標を共有できる人間関係が何よりも大事なのではないだろうか。

証書研究における TEI マークアップの活用

——イングランド王妃アリエノール・ダキテーヌの宮廷随行員の分析を事例に——

田野崎 アンドレア嵐

本報告の目的は、人文学のテキストデータを XML でマークアップするための TEI (Text Encoding Initiative) という規格を用いて、西洋中世史での主要な史料形式の一つである証書の分析を行い、証書研究における TEI 活用の可能性の一端を示すものである。具体的には、中世盛期のイングランド王妃アリエノール・ダキテーヌ (1122/24-1204) の発給証書を題材として、TEI 化で証書内の各要素の共起関係を容易に抽出可能にする一例を提示する。また抽出したデータを、Palladio というビジュアルライズツールにより可視化する試みも行う¹。

アリエノールは 12 世紀のアキテーヌ公領の女子相続人で、イングランド王ヘンリー 2 世の妻であった。彼らの婚姻で生まれた広大な領域複合体 (いわゆる「アンジュー帝国」) では、ヘンリーと彼の息子のリチャード (1 世) とジョンという 3 人の王の統治に加え、王妃のアリエノールも専属の随行員を伴って独自に移動・統治を行っていた。このような、統治目的で移動する王族と随行員からなる人的集合体を、報告者は「宮廷」と定義する。この王妃の宮廷の随行員は、彼女が発給した証書の証人欄から導き出すことができる。

12 世紀のイングランド王・王族が発給した証書 (広義の charter) は、役人等への履行命令である令状 (writ ないし writ-mandate) と土地・特権を与える譲渡証書 (狭義の charter ないし writ-charter) という 2 種類に大別されるが、どちらも証人欄を有するため今回の分析には双方を利用する。証書の分析に TEI を用いる理由は、テキストの構成が極めて定型的で、XML による構造化に適するためである²。アリエノールの証書を利用した研究において、彼女の宮廷には文書発給の可能な機構が欠けている (=彼女の証書は受給者起草である) とかつて主張されていたが、王妃の証書の暫定的なリストを作成し現存証書の筆跡を分析したヴィンセントが、2006 年にこれを否定し発給者起草説を唱えた³。報告者は今回、証書を筆跡ではなく TEI を用いた証人分析で考察し、文書作成能力を持つ聖職者が彼女の宮廷随行員として活動していたか否か、すなわち発給者起草の可能性がどの程度あるかを調査するものである。

今回分析に用いるのは、アリエノールがイングランド王妃時代に発給し、テキストの残存する文書 120 通のうち、報告者未確認の 11 通、証書でない 8 通、発給年代の判別が極めて困難な 7 通を除く、94 通である。TEI 化に際し、各証書は冒頭定式・本文・末尾定式に対応する形で <front> <body> <back> と 3 分割し、文書学的な各要素を <seg type> で分類している (次頁図 1)。ここで、証人欄 (<seg type="testes">) に現れる証人に <persName> を、発給地欄 (<seg type="locatio">) に現れる地名に <placeName> を付し、それぞれに xml:id を与え、証書群全体の TEI Header に証人と地名のリストを作成した。証人リストでは、各証人を聖俗

¹ 'Palladio,' Humanities + Design (Stanford Univ.), accessed March 20, 2020: <https://hdlab.stanford.edu/palladio/>.

² TEI による証書分析の実例として、アングロ・サクソン時代の証書データベースである ASChart があげられる。'ASChart,' King's College London, accessed March 20, 2020: <http://www.aschart.kcl.ac.uk/>.

³ Nicholas Vincent, 'Patronage, Politics and Piety in the Charters of Eleanor of Aquitaine,' in Martin Aurell & Noël-Yves Tonnerre (eds.), *Plantagenêts et Capétiens*, Turnhout 2006, pp. 17-60 (especially pp. 34-38).

で分類し、かつ随行員とみなされる者(特定の役職名を持つ者又は複数の地名に登場する者)に特定のラベルを付している。地名リストでは、各地名がどの地域にあるかで分類した。

結果、各証書から証人と地名との関係データを容易に抽出可能となったので、このデータからネットワーク分析を試みることにした。前段階として、TEI 化された証書のテキストデータを、Python を用いて tsv ファイルに変換している⁴。このファイルを元に、オープンソースのビジュアライズツールである Palladio を用いて、先述の随行員のラベルを付された証人がどの地名に現れるかを、ヘンリー2世治世において可視化した(図2)⁵。

図1 TEI化した証書の例

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI>
  <front>
    <docId>1001</docId>
    <docTitle>[Title]</docTitle>
    <docText>[Text]</docText>
  </front>
  <body>
    <text>[Text]</text>
  </body>
</TEI>
```

図2 証人と地名との関係のPalladioによる可視化(ヘンリー2世治世)

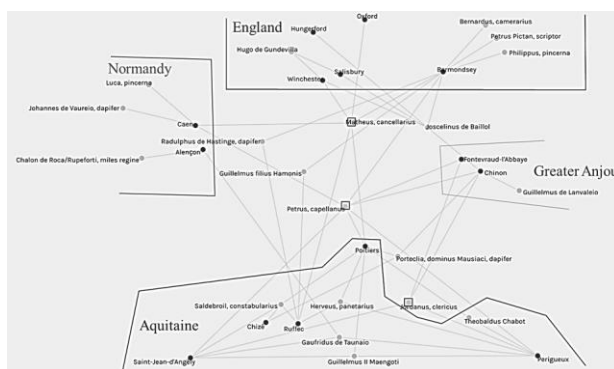


図2から、四角の枠で表される聖職者(特に中心部の2名)が、多数の地名・地域に登場することが分かる。すなわち、文書発給能力を持つ聖職系役人が、アリエノールの巡行に伴って各地域を越境移動しているのである。ここに提示していないリチャードとジョンの治世のデータにおいても、同様に聖職系役人が複数の地域を移動することが確認できた。この結果から、彼女がその宮廷に聖職者の随行をさせている、という事実は少なくとも断定できる。ただし当然ながら、この事実をもって彼女の宮廷の発給者起草を結論付けることはできない。Palladio では表示しきれなかったより細かな発給時による分析、聖職者が文書を作成している証拠(起草者明示)の検討、筆跡・フォーマットの共有の検討を同時に行った上で、ようやく発給者起草を論じることができよう。そのため、本報告では「証書発給人員を随行している可能性がある」という言及に留めておく。

今回の報告から、証書の TEI 化により各要素の比較検討やツールを用いての可視化が容易に可能となることが示された。報告後の進展として、証人に性別や出生地等の新たな属性を加えた他、受益者・発給時という他要素についてもタグ付けを行い、証書データの充実化に努めている。また目下、アリエノール以外のアンジュー帝国期の王族の証書について同様の TEI 化を行っており、長期的には、ASChart のような証書データベースや、その発展形としてプロソポグラフィや宮廷巡行路のデータベース構築に寄与できるよう望むものである。

今回の報告にあたり、小風尚樹氏に、報告の構想からデータの構築まで、幅広く援助・助言をしていただいた。また大向一輝先生、永崎研宣先生から技術面の専門的な助言を、内川勇太氏から歴史学の観点での助言を頂いた。ここに記して感謝の意を表する。

⁴ Python を用いての同変換は、小風尚樹氏に実行していただいた。

⁵ 黒点が地名を、灰色の点が証人を表す。各地名の地域分類と、聖職者を表す四角の枠は報告者が書き加えたもの。また、各地域を越境する証人が中心部に来るように配置を調整した。

歴史学における計量テキスト分析の活用

——歴史地震史料の分析を通して——

濱野 未来

本発表では、日本史学における計量テキスト分析の適用の試みとして、中近世の地震史料を対象に分析を行った結果と、実践から見えた利点・課題・展望を示した。

1. なぜ計量テキスト分析か

発表者の専攻は日本史学であり、中世～近世を対象に、災害（特に地震）の認識や災害対応を研究している。取り組んでいるテーマの1つに、「地震記録の表現の時代的変遷」がある。この研究テーマには2つの特徴がある。1つは、対象年代を限定していない、つまりタイムスケールが広い点、2つ目は地震記録の「表現」に着目する点である。この特徴は、①対象年代が広いために、扱う史料・記録数が膨大となる、②表現を評価するうえでは、客観的・数量的な分析も必要、といった課題も孕んでいる。こうした課題の克服のため、膨大な記録を統計的に分析できるツールとして挙げたものが計量テキスト分析であった。

2. 分析の実践

地震表現の時代的な変遷という問題関心から、全体の傾向を掴む第一歩として、今回は中世から近世への移行期に発生した慶長伏見地震の史料を対象とした。この地震は、文禄五（1596）年閏七月十三日に近畿地方を震源として発生した内陸地震である。近畿一帯で生じた大きな被害や長期間の余震により、それ以前の地震と比較して記録史料数も多い。

扱った史料は、『大日本地震史料』『新収日本地震史料』所収史料のうち、宣教師関連史料等を除いたもので、史料数は50、文字数は16,668語、総抽出語数は12,422語となった。解析は、計量テキスト分析用のフリーソフトウェア KH Coder を利用し、本ソフトに標準搭載されている IPA 辞書から、近世口語（洒落本）UniDic に変更して行った¹。古文用 UniDic のうち、対象史料の年代に合わせて、近世口語（洒落本）、中世口語（狂言）、中世文語（説話・随筆）、中古和文の4つでそれぞれ分析を行い、近世口語（洒落本）が最も適切な語の分解であると判断し、こちらに変更した。

3. 解析結果

まず、頻出語の計数結果（表1）をみると、「地震」「転倒」といった地震やその被害の表現が目につく。この点は地震史料の性質を反映している。「慶長」「閏」などの時間情報も頻出しているが、地震による改元後の「慶長」が多いのは特徴といえる。次に、地震の被害に関する語を抽出し、それらを計数したものが表2である。KH Coder の抽出語から原文に遡及できる機能を使い、最も多い「転倒」の主語をみてみると、その多くは伏見城や関連する建造物であった。また、「無為」という被害がなかったことを示す表現も一定数見られた。

表1 頻出語上位10語

順位	抽出語	出現回数
1	地震	141
2	事	60
3	云々	50
4	者	42
5	転倒	38
6	慶長	37
7	言う	37
8	閏	34
9	主計	33
10	所	33

今回使用したテキストデータには、暫定的に二次史料・同時代史料・後代史料という史料の性質情報を付与した。この史料性質ごとの特徴語を示したものが図3である。後代史料では「破壊」「仏像」（地震により破損）など、地震自体よりも被害情報の語が多い印象を受けた。二次史料には、人物を示す語（「主計」）や、天皇や秀吉を主体とした動詞（「聞こし召す」「仰る」）が多いという特徴がみられた。

こうした解析結果から、以下の2点の全体的な傾向や展望が得られた。1つは、被害の表現が多様になっているのではないかという点である。この点は、今後慶長伏見地震以前/以後と比較してみる必要がある。2点目としては、伏見城の被害に関する情報が非常に多くみられる点などから、地震による伏見城の甚大な被害は当時の人々にとって大きな衝撃であり、慶長伏見地震は地震認識の面において画期と位置づけられるのではないかという洞察を得られた。

4. 実践からみえたもの

実践から得られた発見として、「分析要素の補完」がある。史料を手作業で追う場合、発生した被害の表現のみに注目してしまうが、計量テキスト分析の場合、無被害を示す表現もあることに気づくことができた。その他、①共起関係にある語を視覚的に理解できる（図1）、②属性（史料性質等）ごとの特徴語がわかる、といった史料情報の視角化・可視化の機能もメリットに感じた。これらは、史料を語に分解しているからこそその利点であろう。その上で、歴史学で活用する際に重要なのは、解析結果から原文に遡及が可能な点であろう。当初は、史料に含まれる様々な文脈を無視することになるのではという懸念から、語に分解しての分析には抵抗もあったが、語に分解する利点を認識して抵抗感は薄まった。

一方、課題としてはデータの取捨選択等の解析前の処理の難しさが挙げられる。特に、データクレンジングの際に、史料に解釈が介入してしまうのではないかという点は、歴史学での活用ならではのジレンマであろう。日本史学で活用する上での課題としては、上記に加え、

前近代史料に完全に適した解析用辞書が未だないことや、テキストデータ化の進行状況などが挙げられよう。

今回の実践で、計量テキスト分析は、分析要素を補完できたり、新たな研究視角を得たり、史料の全体的傾向をみるためのツールとして、すなわち、研究の前段階で活用できるのではないかと感じられた。まずはそうした過程での実践事例が増えることで、活用の展望が広がれば幸いである。

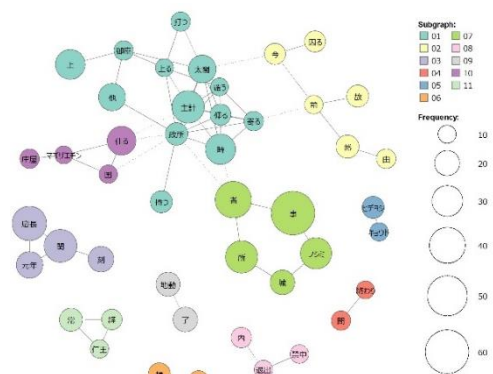


図1 共起ネットワーク図の例

表2 被害関係の特徴語

順	抽出語	出現回数
1	転倒	38
2	損	16
3	破壊	12
4	崩す	10
5	大破	9
6	無為	9
7	崩ずる	8
8	破る/破れ	6
9	裂く/裂け	6
10	崩える	6

表3 史料性質別の特徴語

二次	同時代	後代
主計	地震	元年
被せる	晴れる	慶長
聞こし召す	天晴	閑
寄る	云々	破壊
中	動く	者
仰る	地動	仏像

¹ 国立国語研究所コーパス開発センター:「UniDic」国語研短単位自動解析用辞書:<https://unidic.ninjal.ac.jp/>

歴史災害研究における GIS 活用の試み

大邑 潤三

1. 歴史災害研究における GIS 活用の利点

現在、様々な Web マップがインターネットを通じて提供され、だれもが閲覧・利用できるようになっている。またそれらの Web マップや、それをベースとして構築された Web GIS を利用して、個人が作成した地図データを公開することも容易になっている。報告者は過去に発生した災害について被害の復原や被害要因を分析しているが、これまでこれらの研究成果の発信手段として Web GIS を活用してきた。

災害は空間的広がりをもつ現象であるため、過去の災害現象を復原する場合も、災害記録をもとに被害分布図を作成するなどの地図化作業が必須である。従来このようにして作成された地図は、図として論文や報告書に掲載されてきたが、縮尺の異なる地図を何枚も用意する必要があるなど不便さも生じていた。その点 Web GIS を災害情報の整理に利用することには、1) 地理的範囲やスケールの制限を受けない、2) インターネット上で広く公開できる、3) 整理したデータの頒布や2次利用が可能である、4) 情報の加除修正が容易であるといった、多くの利点が存在する。

2. 歴史地震研究における震度分布図の課題と地震史料 GIS 化の試み

歴史地震研究では、史料の被害記述などを震度に変換し、震度分布図を作成して推定震央の位置を決めるといった手法がとられている。これらの震度分布図は図として先行研究に掲載されているが、図を見ただけでは示された場所が具体的にどの地点を示しているか分からないことが多く、さらに震度判定の根拠となった記録がどの史料であるか判明しにくいという課題がある。具体的な地点が不明な場合、その場所の地形や地盤の影響を考慮することができない。また根拠となった史料が不明であると再現性が確保されず、史料の信頼性や震度判定の妥当性について、第三者が検証することも難しくなる。

報告者は、Web GIS を利用して地震記録を整理することで、こうした問題も解決できると考え、1830 年に発生した文政京都地震（M6.5 程度）の地震史料を GIS データとして整理しデータベース化することを試みた。地震発生直後に作成された同時代史料 4 点を対象に、記録を地点・被害ごとに分解し、分解したフレーズごとに史料件名、記述の性質（実体験・伝聞ほか）、被害種別（建物倒壊・人的被害ほか）、地名とそれに対応する緯度経度などのメタデータを付し、表計算ソフトで整理した。

以上の作業により位置情報が判明する約 1100 件について Web GIS 上でマッピングすることに成功し、被害種別ごとの分布を容易に確認することが可能になった。また様々な記録を網羅的に整理したため、異なる視点で分析を行う場合に史料を改めて始めから読み直すといった労力が軽減される効果がみられた。一方で現時点では試作したデータベースをもとに震度分布図を作成するまでには至っていない。膨大な地点の震度判定を全てひとりで行うのは限界があり、他の研究者の協力やオープンサイエンスなどの手法を取り入れることも有効であると考えている。また本地震に関しては他にも信頼性の高い史料が多く存在しており、これらを追加することでさらに詳細な被害の全体像が明ら

かになり、新たな知見も得られると考えている。

3. 今後の展望

現在日本では、国土に関する様々な GIS データが、国土交通省等によって無償で頒布されている。これらは地理空間情報活用推進基本法に基づいたものだが、同法は 1995 年の兵庫県南部地震の際に、関係省庁が独自にシステムやデータを整備した結果、効率的な整備や相互利用を行えなかったことが教訓となって策定された。また防災科学技術研究所は災害が発生するとクライシスレスポンスサイトを立ち上げ、発生した災害に関する様々な情報を集約し Web GIS 上で提供している。これは迅速な災害対応を行う上で有用な情報となっている。

一方で近世以前に発生した災害に関しては、明治期から史料集が連綿と出版されてきているが、ようやくテキストのデジタル化が開始されたばかりで、GIS データとして整備されるまでには至っていない。また当該分野では多くの先行研究が存在するが、これらに掲載された震度分布図や火山灰の等厚線図などは GIS データになっていないため 2 次利用が難しく、他者による検証なども難しい状況である。

過去の災害履歴に関する情報は、地域の災害対策や被害予測、防災教育にとって重要な公共性の高いデータであり、GIS データとして整備して広く公開されるべき情報である。このような動機から、1) 歴史災害記録（先行研究の成果も含む）を GIS データとして整備公開し、2) 2 次利用可能な汎用性のある形式で頒布（オープンデータ化）することを進めたいと考えている。これが可能になれば、研究者は被害記録の内容解釈や震度判定、被害評価などにのみ注力でき、歴史災害を専門としない研究者も史料を扱いやすくなる。また一般社会においては、過去の被害を掲載したハザードマップの作成などで、より具体的な方法でリスクを伝えることが可能になる。さらに防災教育の面では教育現場の新たなニーズに対応することもできる。高等学校の学習指導要領の改訂により 2022 年から「地理総合」が開始される。「地理総合」では「防災」と「GIS」に関する教育が強化されることになっており、歴史災害に関する GIS データが整備されれば、有効に活用される可能性が高いと考えている。

一方で整備・公開には課題も多い。史料所有者の許可、記録内容の信頼性の確保、データの維持や追加・更新などである。これらの問題については、利用のハードルが低くオープンになっている史料の活用や、歴史学者や情報学の研究者との協力などで対応したいと考えている。



図1 データ提供イメージ