

COVID-19のウェブアーカイブコレクション構築に関する 国際事例研究の予備的検討

常川真央*

A Preliminary Examination of International Case Studies on Building Web Archive Collections for COVID-19

TSUNEKAWA Mao

The novel coronavirus infection (COVID-19) spread worldwide from 2020 and had a significant impact on society and economy. In such a situation, the web became an important medium for information dissemination and transmission, and various contents were created. However, web contents is often updated or disappears, so web archives are needed to preserve it for the future research. This study investigates international cases of building web archive collections on COVID-19, and analyzes their purposes, methods, challenges and prospects. Specifically, we target major organizations and projects such as the Internet Archive (IA). The purpose of this study is to clarify the current situation and challenges of building web archive collections about social impacts of COVID-19, and to provide future directions and recommendations.

キーワード：COVID-19, ウェブアーカイブ, イベント中心コレクション, シビックテック

【目次】

1. はじめに
2. 関連研究
3. 方法論の検討
4. 対象とするイベントの検討
5. イベント中心コレクションの構築活動事例
6. おわりに

* 中央大学文学部助教

1. はじめに

ワールド・ワイド・ウェブ (World Wide Web; 以下, ウェブ) は, 情報社会に生きる市民にとって重要な情報資源として成長し, 発展している. 例えばマスメディアは, 紙媒体の出版物やテレビ放送といった従来のメディアから, インターネットによるニュースの発信へシフトしつつある. また, 2008 年以降のソーシャル・ネットワーク・サービス (Social Net Work Service; 以下, SNS) の登場と普及は, 個人が社会事象について反応したり, 意見を表明し, 広範囲に拡散したりすることを可能にした. 行政や政府など公的な領域においても, オープンガバメントの理念に基づき, 公的な情報がデジタル化され, ウェブ上のオープンデータ公開が積極的に推進されるようになった. このように, 現代社会において, ウェブは市民が社会情報を得るための重要な情報資源となっている.

しかし, ウェブ情報資源は容易に改変や逸失が起るため, 文献としては極めて脆弱であるという問題がある. この問題はウェブ創成期より指摘され続けてきた問題である. このような, 資料としてのウェブ情報資源の保存性の問題は, 情報社会に関する研究を行う上で支障を与えかねない.

そこで, ウェブ情報資源を適切に保存し, いつでも参照ができるようにする取り組みとして, ウェブアーカイブがある. ウェブアーカイブとは, ウェブサイトなどのウェブ上のデジタル情報資源を対象として, 特定の時点の内容 (スナップショット) を保存し, 任意のウェブサイトの時系列的な変更を捕捉できるようにする事業および情報サービスである. Internet Archive が提供する Wayback Machine をはじめとして, 様々な組織がアーカイブの構築事業を行っている. こうしたウェブアーカイブを活用することで, 特定の時点のウェブページを資料として引用・参照することが可能となる.

さらに, 近年では, ウェブアーカイブが収集した膨大なウェブページをもとに, 特定のトピックに関するコレクションをいかに形成するかが議論されている. 特に, 近年では特定の時期に発生した社会事象に関する内容を含むウェブページを組織化する「イベント中心コレクション」(Event-Centric Collection) を構築する手法についての研究が行われている. こうした研究を応用することで, 例えば COVID-19 の感染拡大など情報社会に重要な影響をおよぼした事象に関連したウェブアーカイブコレクションを構築し, 将来の研究に活用することが期待できる. 現在では, これまで世界中で蓄積されてきた膨大なウェブアーカイブから自動的にイベントに関するウェブページを検索し, 活用する手法が活発に研究されている.

一方で, ウェブアーカイブを市民が手動で構築し, 特定の社会事象に関するウェブアーカイブコレクションを構築する活動が存在する.

しかし、イベント中心コレクションの構築の研究と、現にウェブアーカイブコレクションを構築する市民活動の連携に関する研究は十分に行われていない。筆者は、これまで国内における COVID-19 と図書館に関する調査活動において、ウェブアーカイブがどのように活用されてきたのかをフィールドワークを通して調査してきた。さらに、既存のウェブアーカイブコレクションの評価も行ってきた。一方で特定のプロジェクトを超えて、国際的な視野で COVID-19 のウェブアーカイブコレクション構築にどのような課題があり、どのようなアプローチが有効かについての先行研究は管見の限り明らかでない。

そこで、本稿では COVID-19 を対象として、社会事象に関するウェブアーカイブコレクション構築プロジェクトの国際事例を調査する。さらに、事例ごとの特徴を比較した上で、今後のコレクション形成に関する課題や展望について議論する。

2. 関連研究

2.1. ウェブアーカイブの歴史と現在

ウェブアーカイブの歴史は、インターネットが一般に普及した初期の 1996 年より幾つかの非営利組織や図書館が始めた。2023 年現在では、世界中の行政機関や非営利組織によってウェブの広範囲なアーカイブ活動が行われている。

世界で初めてのウェブアーカイブは、1996 年にオーストラリア国立図書館、スウェーデン国立図書館、そして非営利組織 Internet Archive によって行われた¹⁾。Internet Archive は、1996 年に Brewster Kahle と Bruce Gilliat が創設した。2001 年には、“Internet Archive Wayback Machine” というウェブサービスを公開した。同サービスは、任意の URL を入力することで、誰もがウェブページに対する過去のスナップショットにアクセスできるサービスである。現在では、独自のクローラーを用いて自動的に全世界のウェブページを収集・保存しており、世界最大級のウェブアーカイブとして現在も運用されている。こうしたウェブアーカイブの活動は世界中に広がっていき、様々な国や非政府組織が独自にウェブアーカイブを構築していくことになった。

現在では政府主導のもとに構築されたウェブアーカイブとして、ナショナル・ウェブアーカイブが世界中で構築されている。日本国内でも国立国会図書館によるナショナル・ウェブアーカイブの取り組みとして、国立国会図書館インターネット資料収集保存事業 (Web Archiving Project; 以下、WARP) がある²⁾。2002 年以降の日本国内のウェブサイトを収集しており、主

1) Major, Daniela, and Daniel Gomes, “Web Archives Preserve Our Digital Collective Memory”, Daniel Gomes, Elena Demidova, Jane Winters, and Thomas Risse ed., *The Past Web* (Cham: Springer International Publishing, 2021), pp. 11–19. https://doi.org/10.1007/978-3-030-63291-5_2

2) <https://warp.da.ndl.go.jp/>

に公的機関や独立行政法人などのウェブサイトを集集対象としている³⁾。2020年時点では13,153件のタイトルを集集しており、ファイル数は約100億件にのぼる⁴⁾。さらに、ウェブアーカイブの国際的な相互連携のためのコンソーシアムとしてInternational Internet Preservation Consortium (以下、IIPC) が2003年に設立され、世界中のウェブアーカイブ運営組織が加盟するようになった。

2.2. ウェブアーカイブの技術的動向

ウェブアーカイブが組織的かつ国際的な取り組みとして発展してきた一方で、技術的な研究開発についても行われてきた。ウェブアーカイブは、ウェブの時間的な変化を考慮に入れた情報資源の保存を行う点で、デジタルアーカイブや一般的なウェブ検索システムと比べて特異である。本節ではウェブアーカイブの技術的特性に対応してどのような技術が研究開発されてきたかを概説する。

ウェブ情報資源の検索システムとしてのウェブアーカイブには、コンテンツを識別するURLの特性がある。ウェブ上の情報資源にはURLと呼ばれる識別子が付与されており、ウェブのユーザーはURLを使用してウェブコンテンツにアクセスする。一般的なウェブ検索システムは、収集した時点のウェブコンテンツを取り扱い、その変更履歴については保存することはない。したがって、URLとコンテンツは1対1の対応となる。しかし、ウェブアーカイブは、ウェブコンテンツの時間的な変化を考慮に入れるため、1つのURLに対してアクセス時間ごとに複数の異なるバージョンのコンテンツが紐づくことになる。また、ウェブ上では複数の異なるウェブアーカイブが同一のURLに対応するウェブコンテンツを集集・保存することがありうる。そのため、まったく同一の時刻にアクセスした同一のコンテンツであっても、複数のウェブアーカイブで重複して保存されていることが考えられる。以上の性質から、ウェブアーカイブでは、オリジナルのウェブコンテンツを表すURLとは別に、特定時点に特定のウェブアーカイブが保存したウェブコンテンツのスナップショットを表すURLが必要である。ソンペルらは、こうしたウェブコンテンツのスナップショットを「Memento」と呼び、オリジナルのウェブコンテンツのURLを「URI-R」、Mementoに対応するURLのことを「URI-M」と呼んでいる⁵⁾。

もう1つの特性として、現在のウェブアーカイブには複数のウェブアーカイブ同士での連携機能がある点である。ウェブアーカイブがユーザーに対してMementoを提供するには、URI-R

3) https://warp.da.ndl.go.jp/info/WARP_Intro.html

4) https://warp.da.ndl.go.jp/info/WARP_statistic.html

5) Van de Sompel, Herbert, Michael L. Nelson, Robert Sanderson, Lyudmila L. Balakireva, Scott Ainsworth, and Harihar Shankar, "Memento: Time Travel for the Web" (2009), pp. 3-3, <https://doi.org/10.48550/ARXIV.0911.1112>

とURI-Mの対応関係とMementoの保存日時を記録し、検索機能を提供する。ただし、こうした機能がシステム内に閉じていると、ユーザーはあるウェブページに対する履歴にアクセスしようとしても、そのシステム内で収集した範囲に限定されてしまい、他のウェブアーカイブシステムが収集・保存しているMementoに気が付かないおそれがある。こうした課題を克服するために、SompelらはMementoを検索・アクセスするプロトコルとしてHTTP Based Memento Framework(以下、Memento Framework)を策定し、ウェブアーカイブを横断して特定のURLに対応するスナップショットを検索することを可能とした⁶⁾。Memento FrameworkはInternet Archiveを含め、世界各地のウェブアーカイブサービスで実装されている。また、ウェブサービスTime Travelは、Memento Frameworkを活用して特定のURLに基づくアーカイブの横断検索を可能としている^{7),8)}。

URLの問題とは別に、ウェブアーカイブのもう1つの問題としてMementoのパッケージングの問題がある。ウェブページは、複数の画像やスクリプトファイルなどによって構成されており、こうした複数のファイルを1つの情報資源としてパッケージ化する必要がある。そこで、ウェブアーカイブでは単にウェブから収集したウェブコンテンツをデータベース内に保存するのではなく、ウェブページとしての内容を損なわないよう関連するファイル群を1つのアーカイブファイルとしてパッケージ化するアプローチがとられている。このようなパッケージファイルのフォーマットとして、WARC(Web ARChive)がある⁹⁾。こうした標準的なウェブアーカイブフォーマットが標準化されているため、ウェブアーカイブシステム間でMementoを交換することが可能となっている。

サービスの運用支援もまた、整備されている。Internet Archive Wayback Machine¹⁰⁾では、ユーザーが任意のURLの現時点でのスナップショットを保存する機能がある。この機能によって、誰もがいつでも、どこでも残しておきたいウェブページを保存することができるようになった。長期間に及ぶ大量のウェブページの保存についても、支援ツールが整備されている。Archive-It¹¹⁾は、Internet Archiveが運用するウェブアーカイブ構築ツールである。同ツールによって、技術基盤を持たない組織が自身のウェブサイトなどを保存し、バックアップとして利

6) Van de Sompel, Herbert, Michael L. Nelson, Robert Sanderson, Lyudmila L. Balakireva, Scott Ainsworth, and Harihar Shankar, "Memento: Time Travel for the Web" (2009), <https://doi.org/10.48550/ARXIV.0911.1112>

7) <http://timetravel.mementoweb.org/>

8) Van de Sompel, H., M. Nelson, and R. Sanderson, "HTTP Framework for Time-Based Access to Resource States - Memento" (RFC Editor, 2013).

9) International Organization for Standardization (ISO), "ISO 28500:2017 Information and documentation — WARC file format" (2017), <https://www.iso.org/standard/68004.html>

10) <http://web.archive.org/>

11) <https://archive-it.org/>

用できるようになっている¹²⁾。

2.3. イベント中心コレクション

このように、ウェブアーカイブは Internet Archive など少数の組織によってウェブの創成期より営まれてきたが、現在では政府を含む世界中の組織によって行われている国際的な取り組みとなっている。さらに、ウェブアーカイブの運営者や研究者は、Memento Protocol をはじめとしてシステム間の相互連携を可能とする技術も提案し、開発してきた。現在では莫大な規模のウェブページスナップショットを保存し、かつユーザーが相互参照可能な状況が整っている。

しかし、実際に人々がウェブアーカイブを活用する段階では、依然として様々な課題が存在する。そのうちの1つとして指摘されている問題が、ユーザーが情報欲求を満たすウェブページを特定できていない検索タスクの場合である。例えば、「東京オリンピックの延期に関する一連の経緯を説明する当時の報道や公式情報」について知りたいユーザーにとって、ウェブアーカイブは重要な情報源であるはずだが、利活用は容易ではない。なぜなら、ウェブアーカイブを利用するには URL を入力する必要があるが、ユーザーは知りたい情報が載っているウェブページを必ずしも把握しているとは限らないからである。また、全文検索のようなアプローチも、膨大な量のスナップショットに対して検索を行うのは現実的ではない。

こうした問題を解決する方法として、特定の時期の特定のトピックを表す「イベント」に関するスナップショット群を集約し、コレクションとして組織化する手法がある。このようなコレクションは、イベント中心コレクションと呼ばれており、2017年以降、膨大なウェブアーカイブからイベント中心コレクションを生成するための方法について研究が進んでいる¹³⁾。

こうしたイベント中心コレクションの研究は、既存のウェブアーカイブのスナップショットを自動的に検索し、イベントに関連するスナップショットを抽出するという遡及的なアプローチが中心である。このような自動的・遡及的なアプローチを実施することで、最終的にはインターネットの創成期から現在に至るまでの社会史や文化史をウェブアーカイブによって描き出すことが目標である。

しかし、自動的・遡及的なアプローチには次のような疑問がある。現在のウェブアーカイブは、イベント中心コレクションを構成する上で最適な収集方法をとっているのか、そして未来の社会的事象に関するイベント中心コレクションを構成する場合に、現在のウェブアーカイビング活動にはどのような問題点があるのか、という点である。ウェブアーカイビングは、ウエ

12) <https://archive-it.org/blog/learn-more/>

13) Gossen, Gerhard, Thomas Risse, and Elena Demidova, "Towards Extracting Event-Centric Collections from Web Archives", *International Journal on Digital Libraries*, Vol. 21, No. 1 (Springer Nature, 2020), pp. 31-45, <https://doi.org/10.1007/s00799-018-0258-6>

ブのすべてを完全に保存することは不可能である。常に膨大なウェブページが生まれ、そして更新されていく中でどのウェブページを保存していくのかは、ウェブアーカイビングを実施する組織の方針やクローラーなどの技術的制約によって変わってくる。

以上のようなイベント中心コレクションをめぐる課題を踏まえ、本研究では次のような問いを提起する。すなわち、特定の社会事象に対するイベント中心コレクションを形成するアプローチをウェブアーカイビングの組織体から見たとき、そこにはどのような類型が存在するのか。本研究では、以上の問いを明らかにするための方法論ならびに研究対象の検討を行う。

3. 方法論の検討

本章ではイベント中心コレクションの形成活動に対する研究の方法論について検討する。ウェブアーカイブ活動の研究を方法論の側面から検討するとき、そこには情報システム研究とイベントという概念を取り扱う研究という2つの特性を踏まえて考慮する必要がある。

情報システムは計算機による機械的な機構と、情報システムの構成要素としての人間や組織の機構が組み合わさって成立している。ウェブアーカイブもまた、2章で述べた通り、ウェブコンテンツを収集し、Mementoとして保存ならびに組織化していく機械的な側面と、ウェブアーカイブ活動を促進し維持していく組織の側面によって成り立っている。したがって、ウェブアーカイブの研究は純粋に計算機処理や検索アルゴリズムの改善といった計算機工学的なアプローチだけでなく、組織や人間がウェブ情報資源をどのように捉え、ウェブアーカイブという取り組みの中でどのような情報実践を行うのかを解明するという社会情報学のアプローチが必要となる。

本研究の方法論に関するもう1つの論点として「イベント」という概念がある。本研究における「イベント」は、ウェブアーカイブに保存されているMementoを組織化する際の主題として扱われる、特定の社会事象である。イベントは、ウェブアーカイブが収集するウェブページがデジタルデータとして実在するのに対して、社会の構成員が現実を解釈し、表現される現象である。そのため、イベント中心コレクションに対する実態の研究においては、常に人々が何を「イベント」と呼ぶのかという解釈上の問題がある。

以上の考察から、イベント中心コレクションの活動に対する研究は、統制された実験環境の中での何らかの一般法則を研究するアプローチではなく、現実のイベント中心コレクションの形成活動を観察し、比較分析するアプローチが適している。そこで、本研究では事例研究(Case Studies)が適している。事例研究は情報システムに関する研究領域において、多様なパラダイムから研究手法として採用されている方法論である¹⁴⁾。事例研究によって情報システムの個

14) Shanks, Graeme, and Nargiza Bekmamedova, "Case Study Research in Information Systems", *Research Methods: Information, Systems, and Contexts*, Kirsty Williamson and Graeme Johanson ed., (Cambridge: Chandos Publishing, 2018).

別具体的な事例を詳細に調査・分析し、記述することは、理論や設計上の情報システムのモデルと比べ、実際に開発・運用されている情報システムの社会的文脈や実践を捉えることができる。

他方で事例研究は、多様なパラダイムや分析手法を許容する方法論であるために、意義のある研究成果を生み出すために、拠って立つパラダイムや具体的な調査・分析手法を慎重に検討し、明確に記述する必要がある。情報システムに関する事例研究は、拠って立つ認識論によって、「実証的事例研究 (Positivist Case Studies)」、 「解釈学的事例研究 (Interpretivist Case Studies)」、 「批判的事例研究 (Critical Case Studies)」の3種類のアプローチがある¹⁵⁾。実証的事例研究は、情報システムならびに情報システムがおかれた社会的現実が客観的に実在することを前提として、研究者の提示した理論や仮説が現実の情報システムの開発・導入を適切に説明できるかを検証するために事例調査を行う。対照的に、解釈学的事例研究は、情報システムならびに情報システムがおかれた社会的現実が、人々の主観的、間主観的な認識によって構成されているという前提に立ち、人々がどのように情報システムを取りまく社会的現実を構成しているかを観察するために事例調査を行う^{16),17)}。また、批判的事例研究は、現実の情報システムの事例を調査することで、そこに存在する矛盾や課題を浮かび上がらせ、情報システムを取りまく社会の積極的な改善を目指すために事例調査を行う。情報システムの研究においては、その設計プロセスの認識や組織におけるシステム導入の文脈によって利害関係者の認識が情報システムの姿に強く影響を与えるという特性からも、解釈学や批判理論のような非実証主義的なアプローチの親和性がある。加えて、そのほかの認識論に依拠する情報システム研究として、批判的实在論 (Critical Realism) に基づくアプローチもある¹⁸⁾。批判的实在論は、客観的な社会的現実の实在を認めつつも、人々の現実へのアクセスは主観的認識のバイアスを受けており、結果的に、客観的に実在する現実と主観的認識によって経験的な社会的現実が構成されているとする認識論である。批判的实在論に立つ情報システム研究は、定量的な調査・分析によって客観的な実証研究を目指しつつも、情報システムの開発、導入、運用をめぐる一連のプロセス

15) Orlikowski, Wanda J., and Jack J. Baroudi, "Studying Information Technology in Organizations: Research Approaches and Assumptions", *Information Systems Research*, Vol. 2, No. 1, (Catonsville: INFORMS, 1991), pp. 3-6.

16) Shanks, Graeme, et al., op. cit., pp. 198-198.

17) Walsham, G, "Interpretive Case Studies in IS Research: Nature and Method", *European Journal of Information Systems*, Vol. 4, No. 2, (Birmingham: The Operational Research Society, 1995), pp. 74-75, <https://doi.org/10.1057/ejis.1995.9>

18) Mingers, John, Alistair Mutch, and Leslie Willcocks, "Critical Realism in Information Systems Research", *MIS Quarterly*, Vol. 37, No. 3, (Minneapolis: Management Information Systems Research Center, 2013), pp. 795-802.

において利害関係者の認識がその在り方に強い影響を与えることも配慮し、質的な研究プロセスも併用する。本稿では、本研究が目指すウェブアーカイブにおけるイベント中心コレクションの構築実践の事例研究において、どのようなパラダイムに基づくかについては確定していない。ウェブアーカイブを情報システム研究としてどのようなパラダイムに立つことが適切であるかを検討しつつ、位置づけを明確にする必要がある。

4. 対象とするイベントの検討

イベント中心コレクションの事例分析を行うにあたり、本研究では「COVID-19の感染拡大に対する図書館への影響」を対象とするイベントとして仮設する。このイベントについて、以下の通り概説する。

COVID-19は、2019年12月より中国の武漢にて初めて感染拡大が発見されて以降、急速に世界中で感染拡大が起こった。その結果として2020年3月11日に世界保健機関（WHO）は世界的な大流行である「パンデミック」であると宣言した。日本国内においても2020年1月に初めて感染が確認されて以降、急速に感染者が発生し、2020年4月7日には安倍首相（当時）が新型インフルエンザ等対策特別措置法に基づき、緊急事態宣言を発出した¹⁹⁾。これにより、人と人との接触を7～8割減らすよう呼びかけるなど、人々の日常生活に多くの制限をもたらすこととなった。このような、COVID-19の感染拡大による日常生活の制約は、教育文化施設の運営においても大きな影響を及ぼすこととなった。その1つに公共図書館の全国レベルでの休館が挙げられる。緊急事態宣言の発出により、2020年5月時点で全国の公共図書館の9割が休館となるなどの事態が発生した。また、国立図書館や大学図書館、専門図書館など様々な種類の図書館でも利用制限や休館の措置が行われた。

以上のイベントを本稿で取り上げる理由は、次の通りである。

1点目として、COVID-19の感染拡大は世界規模のパンデミックであり、短期かつ同時期に国際的な影響を及ぼしている点である。そのため、多種多様な関連するウェブページ群が存在することが予想される。さらに、同一のイベントに関するウェブアーカイブ活動の国際的な比較が容易である。

2点目として、図書館は世界中に同様の施設が存在し、そのサービスの在り方はある程度の類似性が見られるということである。そのため、COVID-19の図書館に与える影響に関するウェブコレクションで収集されるウェブページにはある程度の類似性があり、内容の分析が容易ではないかと推定するためである。

3点目として、図書館の休館や利用制限は、学術研究や生涯学習など、人々の文化的な生活

19) http://web.archive.org/web/20230306024220/https://corona.go.jp/news/pdf/kinkyujitai_sengen_0407.pdf

に影響を及ぼし、社会学的な研究上重要ではないかと考えられるからである。例えば、日本における人文社会学分野の研究者の研究活動に支障を及ぼすなど、文化的な負の影響を与えたことが調査から分かっている。

4 点目として、COVID-19 の感染拡大に対する影響は、各国の公衆衛生に関する政策や行政の意思決定に強い影響を及ぼしているという点である。そのため、行政機関などの公的なウェブページが多く参照されていることが推定される。これは、ナショナル・ウェブアーカイブがどの程度イベント中心コレクションの形成に役立つかを評価する上でも役立つ。

以上、4 点の理由から、本研究では「COVID-19 の感染拡大に対する図書館への影響」というイベントに関するウェブアーカイブコレクションの構築事例について調査・分析を行う。

5. イベント中心コレクションの構築活動事例

それでは、「COVID-19 の感染拡大に対する図書館への影響」というイベントに関するウェブアーカイブコレクションの構築活動にはどのような事例があるだろうか。本稿では、事例研究の予備的な段階として COVID-19 に関するウェブアーカイブ動向の概要と、対象となるイベント中心コレクションの例を概説する。

まず、COVID-19 に関する世界的なウェブアーカイブコレクション構築の取り組みとしては、Internet Archive による COVID-19 Web Archive がある²⁰⁾。同コレクションは、COVID-19 に関するウェブサイトのアーカイブを集約したポータルサイトである。Internet Archive が提供するウェブアーカイブサービスである、Archive-It 上で収集・保存されている²¹⁾。同団体が企画した COVID-19 Web Archiving Special Campaign の成果として構築している。現時点で、125 超の図書館、アーカイブや文化遺産に関する組織によって構築された 160 を超えるウェブアーカイブコレクションを収録しており、現在も継続的に更新を続けている。

一方で、国内のウェブアーカイブ活動にはどのような取り組みがあるだろうか。国内における COVID-19 と図書館に関するイベント中心コレクション構築活動には、saveMLAK COVID-19 図書館調査プロジェクトがある。同調査は、saveMLAK が 2020 年より実施している、COVID-19 に関する全国の図書館を対象とした調査プロジェクトである。saveMLAK とは、「博物館・美術館 (M)、図書館 (L)、文書館 (A)、公民館 (K) (M + L + A + K = MLAK) の被災・救援情報サイト」であり、「被災地域の各施設の被災情報を集め、必要とされている情報を発信」している²²⁾。同調査プロジェクトでは、全国の図書館ウェブサイトにアクセスし、COVID-19 の感染拡大を理由とした休館や利用制限が行われているか、代替となるサービスを提供してい

20) <https://covid19.archive-it.org/>

21) <https://blog.archive.org/2022/10/11/covid-19-web-archive/>

22) <https://savemlak.jp/wiki/saveMLAK>

るかを確認し、集計することで、国内における公共図書館の休館率や提供サービスの状況について明らかにする調査を実施している²³⁾。これまで、35回(執筆時点)に及ぶ調査を実施し、調査レポートとデータを公開した。同調査の特徴として、図書館ウェブサイトにアクセスする際、Internet Archive Wayback Machine を利用し、調査の根拠となる調査時点のウェブページを保存している。ウェブアーカイブを調査と同時に実施することで、調査レポートを参照した第三者がエビデンスを確認することができるようになっている。saveMLAKによるこのようなウェブアーカイブ活動は、COVID-19が公共図書館に与えた影響という社会事象に対応するイベント中心コレクションといえるだろう。saveMLAKが収集したMementoの多くは、WARPなどのナショナル・ウェブアーカイブに収録されていないことを筆者は予備的調査によって確認しており²⁴⁾、COVID-19の公共図書館に与えた影響という社会事象を明らかにするうえでの重要なコレクションとなることが期待できる。

6. おわりに

本稿では、ウェブアーカイブにおけるイベント中心コレクションの構築事例について国際的な事例研究を行うための予備的検討を行った。ウェブアーカイブの歴史と動向について触れ、情報社会の研究を支える基盤としての意義について説明した。そのうえで、ウェブアーカイブを利活用する際の課題を挙げ、課題を解決するための手法としてイベント中心コレクションのアプローチを紹介した。こうしたイベント中心コレクションを機械的な手法によって実現しようとする研究と、実際にイベント中心コレクションを構築している事例を紹介しながら、本稿ではイベント中心コレクションに対する研究アプローチとして事例研究を選択するということが、および事例研究のパラダイムの検討が必要であることを示した。そのうえで、本稿ではイベントの例としてCOVID-19の世界規模の感染拡大を試行的に取り上げ、COVID-19が社会に与えた影響に関するイベント中心コレクションの事例として、Internet Archive COVID-19 Web Archive Collection と、saveMLAK COVID-19 図書館調査を取り上げた。

今後の研究の展望としては、本稿で取り上げたCOVID-19に関するイベント中心コレクションの活動について、使用しているツールや活動形態を比較検討すると同時に、結果として構築したコレクションの内容を分析することで、イベント中心コレクションを構築する際の課題と

23) 常川真央「公共図書館における新型コロナウイルス感染症への対応」『月刊社会教育』編集委員会編『月刊社会教育』旬報社、65巻10号(2020年)、10-17ページ。

24) 常川真央「災害ならびに感染症の記録に向けたウェブアーカイブコレクションの評価：新型コロナウイルス感染症に関する図書館ウェブサイトのアーカイブを対象に」日本図書館情報学会研究大会事務局編『日本図書館情報学会研究大会発表論文集』Vol. 70、(仙台：日本図書館情報学会、2022)：pp. 5-8。

品質に影響を与える要因について調査・分析したい。並行して、ウェブアーカイブという情報システムに対する事例研究のアプローチとして、どのような認識論に立つことが望ましいかを議論し、ウェブアーカイブの研究領域全体に対する基礎づけへの貢献に努めたい。

付記：本研究は JSPS 科研費 JP19K20625 の助成を受けたものです。

参考文献

- Darke, Peta, Graeme Shanks, and Marianne Broadbent. 1998. "Successfully Completing Case Study Research: Combining Rigour, Relevance and Pragmatism", *Information Systems Journal*, Vol. 8, No. 4, (Wiley, 1998), pp. 273-89, <https://doi.org/10.1046/j.1365-2575.1998.00040.x>
- Gossen, Gerhard, Thomas Risse, and Elena Demidova, "Towards Extracting Event-Centric Collections from Web Archives", *International Journal on Digital Libraries*, Vol. 21, No. 1 (Springer Nature ,2020), pp. 31-45, <https://doi.org/10.1007/s00799-018-0258-6>
- International Organization for Standardization (ISO) , "ISO 28500:2017 Information and documentation — WARC file format" (2017), <https://www.iso.org/standard/68004.html>
- 常川真央「公共図書館における新型コロナウイルス感染症への対応」『月刊社会教育』編集委員会 編『月刊社会教育』旬報社, 65 卷 10 号 (2020 年), 10-17 ページ.
- 常川真央「災害ならびに感染症の記録に向けたウェブアーカイブコレクションの評価：新型コロナウイルス感染症に関する図書館ウェブサイトのアーカイブを対象に」日本図書館情報学会研究大会事務局編『日本図書館情報学会研究大会発表論文集』Vol. 70, (仙台: 日本図書館情報学会, 2022):5-8 ページ.
- Major, Daniela, and Daniel Gomes, "Web Archives Preserve Our Digital Collective Memory", Daniel Gomes, Elena Demidova, Jane Winters, and Thomas Risse ed., *The Past Web* (Cham: Springer International Publishing, 2021). https://doi.org/10.1007/978-3-030-63291-5_2
- Mingers, John, Alistair Mutch, and Leslie Willcocks, "Critical Realism in Information Systems Research", *MIS Quarterly*, Vol. 37, No. 3, (Minneapolis:Management Information Systems Research Center,2013), pp. 795-802.
- Orlikowski, Wanda J., and Jack J. Baroudi, "Studying Information Technology in Organizations: Research Approaches and Assumptions", *Information Systems Research*, Vol. 2, No. 1, (Catonsville: INFORMS, 1991), pp. 1-28.
- Shanks, Graeme, and Nargiza Bekmamedova, "Case Study Research in Information Systems", *Research Methods: Information, Systems, and Contexts*, Kirsty Williamson and Graeme Johanson ed., (Cambridge: Chandos Publishing, 2018).
- Van de Sompel, Herbert, Michael L. Nelson, Robert Sanderson, Lyudmila L. Balakireva, Scott Ainsworth, and Harihar Shankar, *Memento: Time Travel for the Web* (2009), <https://doi.org/10.48550/ARXIV.0911.1112>
- Walsham, G, "Interpretive Case Studies in IS Research: Nature and Method", *European Journal of Information Systems*, Vol. 4, No. 2, (Birmingham:The Operational Research Society,1995), pp. 74-81, <https://doi.org/10.1057/ejis.1995.9>