

記 録

文書番号	S C J 第 2 0 期 200904-20530800-011
委員会等名	日本学術会議 情報学委員会 ウェブ・メディア社会基盤分科会
標 題	我が国におけるウェブアーカイブ構築の 意義および諸課題
作成日	平成 2 0 年 (2008 年) 9 月 4 日

※ 本資料は、日本学術会議会則第二条に定める意思の表出ではない。掲載されたデータ等には、確認を要するものが含まれる可能性がある。

この記録は、日本学術会議 情報学委員会 ウェブ・メディア社会基盤分科会
における審議の結果をとりまとめ、記録として公表するものである。

日本学術会議情報学委員会ウェブ・メディア社会基盤分科会

委員長	喜連川 優（連携会員）	東京大学 教授
幹事	辻 ゆかり（連携会員）	日本電信電話株式会社 主幹研究員
委員	石塚 満（連携会員）	東京大学 教授
	酒井 善則（連携会員）	東京工業大学 教授
	坂内 正夫（会員）	国立情報学研究所 所長
	高野 明彦（連携会員）	国立情報学研究所 教授
	田中 克己（連携会員）	京都大学 教授
	所 眞理雄（連携会員）	（株）ソニーコンピュータサイエンス研 究所 代表取締役社長
	長尾 真（連携会員）	国立国会図書館長
	古井 貞熙（連携会員）	東京工業大学 教授
	松田 晃一（連携会員）	情報処理推進機構 IT人材育成本部長

報告書及び参考資料の作成にあたり、以下の方々に御協力いただきました。

岡本 真	ヤフー株式会社検索事業部企画部	リーダー
角谷 和俊	兵庫県立大学環境人間学部	教授
瀬高 隆裕	内閣府 参事官補佐（前 総務省情報通信政策局	課長補佐）
田中 久徳	国立国会図書館総務部企画課	電子情報企画室長
豊田 正史	東京大学生産技術研究所	准教授

目 次

1. はじめに	1
2. ウェブアーカイブの動向	2
2. 1 各国国立図書館による取り組み	2
2. 2 学術的取り組み	3
2. 3 我が国における企業等における取り組み	3
3. ウェブアーカイブの意義	4
3. 1 国家としての意義	4
3. 2 文化・社会・生活面からの意義	5
3. 3 学術としての意義	7
3. 4 経済的意義	8
4. ウェブアーカイブ構築に関する諸側面	9
4. 1 国によるアーカイブ	9
4. 2 法制度面について	9
4. 3 ウェブ情報の収集蓄積の技術的側面	9
4. 4 文化・社会・生活面の側面	10
4. 5 学術的利用に関して	10
4. 6 経済・ビジネス利活用に関して	10
5. おわりに	11

日本学術会議情報学委員会
ウェブ・メディア社会基盤分科会「記録」

我が国におけるウェブアーカイブ構築の意義および諸課題

1. はじめに

1997年に生まれたウェブ技術は、大きな発展を遂げ、現在、人類にとって、日常生活においても、また、ビジネスにおいても不可欠なメディアと進化しつつある。ウェブはいわゆるサーチエンジンに関して目覚ましい技術革新がなされ、極めて容易にアクセス可能な情報源としての地位を確立し、今日、ウェブを利用しない日は無い程の社会情報基盤となりつつある。また、CGM (Consumer Generated Media)、UGC (User Generated Contents)とも呼ばれるブログ情報は多様な目線を記す容易なメディアとして広く受け入れられ、社会コミュニティの形成に大きな役割を果たしつつある。このようなウェブ情報は時代時代を民意の表出としての文化ともみなすことが出来、極めて重要なメディアであるにも関わらず、とりわけポーンデジタルなメディアの特性としてその脆弱性から極めて失われやすく、ユネスコは、2003年、Charter on the preservation of the digital heritage (デジタル遺産の保存に関する憲章) を採択し、その保存の重要性を訴えている。ウェブは学術研究の基盤としても極めて重要な意義を有する。欧米においても、ウェブアーカイブの取り組みが開始されており、我が国においても本格的な取り組みが強く望まれる。

このような状況を背景とし、「日本学術会議情報学委員会ウェブ・メディア社会基盤分科会」第20期においては、ウェブ上で発信されるポーンデジタルな情報のアーカイブ（以下、ウェブアーカイブと呼ぶ）に関して世界での諸活動を俯瞰するとともに、その推進の方策について検討し、提言をしていくことを目指している。即ち、ウェブアーカイブを取り上げ、本分科会委員、あるいはオブザーバーからの話題提供、調査および現状報告などをもとに、その構築に関する諸課題についての議論を行ってきた。

今後の学術の新たな情報基盤となりうるウェブアーカイブの構築に関し、その重要性を明らかにするとともに、その実現に向けた具体的な課題を提示することは日本学術会議が果たすべき非常に重要な事項であり、日本の学術のレベルを高め、長期的な発展に大きく資するものと考え次第である。

2. ウェブアーカイブの動向

2. 1 各国国立図書館による取り組み

欧米諸国では、国立図書館が中心になってウェブアーカイブへの取り組みを進めている。

欧州では、1996年にスウェーデン王立図書館が Kuitura3 プロジェクトを開始したのを嚆矢として、翌年には欧州8ヶ国による NEDLIB (Networked European Deposit Library) プロジェクトが行われた。その後、デンマーク王立図書館 (1997年)、フランス国立図書館 (1999年)、ノルウェー国立図書館 (2000年)、英国図書館 (2001年) 等で、実験的アーカイブの取り組みが続き、方法についても、選択的収集から、クローラーによる自国内ドメインの包括的自動収集へと展開した。

また、欧州各国では、国立図書館の法定納本の枠組みによって、ウェブサイトの包括的収集を実施することを規定する法律の整備も進んでいる。(1988年ノルウェー、2002年スウェーデン、2003年英国、2004年デンマーク、2006年ドイツ、フランス)。

米国では、1996年にNPO法人(現在はカリフォルニア州認可図書館)インターネット・アーカイブ(IA)が設立され、全世界のインターネット情報を対象とした包括的アーカイブを開始し、アーカイブしたサイトを公開している。一方、電子情報の消滅を懸念する連邦議会が主導して、ポーンデジタルのコンテンツを保存するための国家戦略、「全米デジタル情報基盤・保存プログラム(NDIIPP: National Digital Information Infrastructure and Preservation Program)」が2000年に立法化され、米国議会図書館が、その中心機関として活動している。議会図書館は、IAが収集したウェブ情報をもとに大統領選挙や各種イベント、9・11同時多発テロ関係等のテーマによって編集する、選択的アーカイブのプロジェクト、MINERVAを2000年に開始した。個別に著作権者の許諾を得ることで、現在、37,000以上のウェブサイトが公開されている。

このほか、2003年には、IAとフランス国立図書館が主導して、主要国の国立図書館によるインターネット保存のための国際組織であるIIPC(International Internet Preservation Consortium)が設立され、共通技術課題への対応、標準化の推進、ガイドラインの策定等の連携活動を行っている。

我が国では、国立国会図書館が、2002年から「インターネット情報資源選択的蓄積実験事業(WARP)」を開始している。この事業は、同館が選択したウェブサイト及び電子雑誌について、個別に著作権許諾契約を結び、収集、保存、提供を実施するもので、2006年に実験事業から本格事業に移行した。現在、国、地方自治体、大学、法人等の公的機関、イベント事業等のサイト約2,000、電子雑誌約1,500が継続的に収集されている。

2. 2 学術的取り組み

ウェブアーカイブの構築には、膨大なウェブ情報の収集・蓄積技術が必要なことから、計算機科学・工学の分野において先進的な研究開発が行われてきた。膨大な情報を効率的に収集するには、多数の計算機による協調作業が必要であり、大規模なデータの蓄積には高度なストレージ技術を要する。また、蓄積した大規模ウェブ情報の利活用には、高速かつスケーラブルなデータマイニング技術が必須である。

米国スタンフォード大学においては WebBase なるプロジェクトにおいて、大規模ウェブ情報の収集技術、索引技術、および蓄積技術の研究が行われており、米国検索エンジンのグーグルは本プロジェクトからのスピンアウトである。比較的大規模なウェブ情報を蓄積しており、他大学の研究プロジェクトへのデータ提供等も行われている。

我が国においても、早稲田大学、東京大学等において、大規模なウェブ情報の収集・利活用に関する研究開発が実施されている。早稲田大学においては、ウェブ情報の瞬時収集技術が開発されており、これを用いて 144.5 億ページを収集し、言語分布等の分析が行われている。東京大学においては 1999 年から時系列的な日本語ウェブ情報収集に取り組み、可変周期収集技術により 100 億ページ規模のウェブアーカイブを構築した。また、アーカイブのデータを社会学、マーケティング等へ利用するための実証実験も行われている。

また、文部科学省特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究（平成 17～平成 22 年度）」においては、ウェブ情報をはじめとした爆発的に増大する情報の利活用研究が行われている。大量かつ多様な情報から真に必要な情報を効率良くかつ偏りなく安心して取り出すことを可能とする技術、大量の情報を管理する大規模な情報システムを安定・安全に運用するための新しいサステナブルな技術等を目指しており、ウェブ情報技術に関しては京都大学、国立情報学研究所、東京大学を中心として研究開発が推進されている。

2. 2 我が国における企業等における取り組み

経済・ビジネス利活用におけるウェブアーカイブの目的としては、内部統制強化、及び企業活動への活用の 2 つがあげられる。

平成 18 年（2006 年）6 月 7 日、金融商品取引法¹が成立し、新たな内部統制のルールとして「J-SOX（日本版 SOX 法）」が実施されることになって以来、管理責任の観点から、各企業内で内部統制の仕組みを構築する必要が生じ、様々な企業情報の蓄積が必須となった。蓄積する情報としては、会議文書などドキュメント類や電子メール、インターネットを通じた社外への公開情報などが含まれる。これは、特に企業に対して何か問題や疑念が生じたときに、その疑念を晴らすために必要となる証拠を保持しておかねば

¹ 金融商品取引法：<http://law.e-gov.go.jp/htmldata/S23/S23HO025.html>

ならないためである。例えば、電子メール・アーカイブを民事訴訟の際に証拠提示するのは代表的な適用事例である。

企業活動にウェブアーカイブを使う流れとしては、一つには、蓄積情報の分析をマーケティングや社内情報共有に活用する事例がある。例えば、きざしカンパニー²は、毎日15~20万のブログエントリをランダムに収集し、この情報をもとに、旬な話題ランキングや、話題検索・ブログ検索機能を提供している。さらに、ブログの書き込み内容を解析し、書き手の感情や話題の属性を判定し、感情の変化をグラフで表示する仕組みなども有している。また、NTTのBLOGRANGER（ブログレンジャー）³は、ブログの記事を分析し、仮想の地形で話題やトピックスの分布を可視化する。これらの分析情報は、次の商品企画に用いられ、顧客対応の注意点として顧客窓口のオペレータが共有するなど、社内情報共有基盤として利用されたりする。もう一つ、インターネット上の情報を蓄積し、集団の力を業務委託の領域で活用しようとする動きとして、クラウドソーシングがある。クラウドソーシングとは、企業や団体が不特定多数のネットワーク上の人々に対して業務を委託することをいう。従来であれば、自社の従業員や外部にアウトソースして実施していた業務のうち、例えば、製品設計用のソフトウェア開発や映像編集などにおいて、不特定多数の参加者からの膨大なアイデアや知識が蓄積され、整理され、淘汰されることにより、巨大な成果が生み出される。ウェブ上での作業過程で発生した中間生成物や参加者間のメールのやりとりがアーカイブされることにより、参加者はそれらを検索して、作業過程の振り返りをおこなったり、情報を再利用することが可能となるため、より効率的な作業が期待できる。

3. ウェブアーカイブの意義

3.1 国家としての意義

国（政府）として、ウェブアーカイブの意義を考えると、大きく3つの観点から、その必要性を指摘することができる。具体的には、①政府の活動記録の保存と国民への情報公開、②事実確認のための公的証拠の担保機能、③民意の補足と国政活動での活用といった方策である。

政府・自治体の活動記録としては、これまでは冊子の出版物として刊行されてきた行政府省、国会、裁判所等の公的記録の、電子政府の推進がはかられ、電子形態で広く公開されるようになり、速報性の観点も含めて国民のアクセスの利便性が向上している。

一方、情報の永続性の観点から見れば多くの課題が残っており、原則として、情報の発信主体とは異なる機関により、公開された記録の包括的保存と長期利用の保証が図られる必要がある。国、独立行政法人、地方自治体、国立大学等の機関の活動や政策審議

² きざしカンパニー：<http://www.niche-ad.com/kuchikomi/kizasi.php>

³ ブログレンジャー：<http://ranger.labs.goo.ne.jp/>

過程の公的記録、各種統計、予算、法案審議、調査研究報告、選挙記録、政党・議員等の政治活動記録といった情報のアーカイブが重要である。これらの記録は、国民に対する情報開示の観点に加えて、新たな政策立案、国会における国政調査、法案審議、裁判実務等の政府活動に不可欠な情報基盤を提供することにもなる。

次に、事実確認のための「証拠」としての側面では、政府機関のみならず、企業や民間団体、報道機関等の広い意味で公的性格を持つ組織・機関がウェブに発信した記録の第三者機関によるアーカイブが必要である。現実には、ウェブの世界では、事件や問題の発覚後に発信されていた情報の消去、改ざんが行われている現実があり⁴、公的情報の隠蔽を防止し、事実証拠の提供を保証するための社会的機能として、公的機関によるアーカイブの検討が望まれる。

最後に、ウェブ上では、CGM (Consumer Generated Media) やブログといった新規のメディアが機能することで、これまで社会に対し意見表明の機会を持たなかった広範な国民が自由に意見表明する状況が生じている。こうした多数の多様な意見をアーカイブし、世論（社会・市民）の動向を分析し、新しい意見を掘り起こすことができれば、これまでは把握困難な深いレベルでの民意を補足し、国や地方自治体における政策立案に役立てる可能性が期待される。

3. 2 文化・社会・生活面からの意義

日本の歴史と文化に関わる文化遺産は、現在、博物館・美術館、図書館・公文書館、大学・研究機関、自治体・教育委員会などにおいてデジタルアーカイブ化され、ウェブ情報として収集・蓄積され発信されている。文化遺産のデジタルアーカイブは、古来より継承・蓄積されてきた文化遺産を劣化させることなく後世に正確に伝えることができるとともに、いつでも再生・復元することができるという意味で、大きな意義がある。また、デジタルアーカイブは、我が国や各地域のアイデンティティのための情報基盤であり、経済のグローバル化は進行したが、個人レベルのグローバル化のためには国際的な相互理解・情報交流が重要であり、その基盤となる我が国の歴史・文化に関わる情報の収集・蓄積・発信は重要である。我が国における文化遺産等に関するデジタルアーカイブおよびウェブによる情報発信については、文化庁による「文化遺産オンライン」⁵や、デジタルアーカイブ推進協議会 (JDAA) によるデジタルアーカイブWebサイト事例集⁶、

⁴ 国内の例ではないが、第168回臨時国会のテロ特措法延長審議の過程で、江田衆議院議員が政府に提出した質問趣意書では、「米海軍中央司令部&第五艦隊 (United States Naval Forces Central Command and 5th Fleet)」の公式ウェブサイトが、2007年9月3日から完全に削除された事例が問題となっている。

⁵ 文化遺産オンライン (Cultural Heritage Online) (<http://bunka.nii.ac.jp/Index.do>) は、文化庁が運営する我が国の文化遺産に関する電子情報広場 (ポータルサイト) である。

⁶ デジタルアーカイブ推進協議会 (JDAA) (Webサイトは2005年7月1日閉鎖) がまとめた、我が国の博物館・美術館、図書館・公文書館、大学・研究機関、自治体・教育委員会等のデジタルアーカイブのWebサイト事例集であり、同協議会のデジタルアーカイブ白書2005から引用されたもの。

<http://www.dcaj.org/jdaa/web/index.html>

(財)京都国際文化交流財団⁷による「京都デジタルアーカイブ」などがある。

社会面からのウェブ情報のアーカイビングは、社会の情勢を敏感に反映しうるブログやSNS（ソーシャル・ネットワーク・サービス）のコンテンツを蓄積・分析することで、時系列的な社会分析などに活用可能である。テロや犯罪行為の兆候をリアルタイムで監視するという公安活動の基礎資料としても利用可能であると考えられる。このような場合、我が国の生活・文化に関して我が国から発信されるもの以外に、海外ニュースサイト、特にアジア圏での日本に関わるニュースやブログ等の報道コンテンツ等のアーカイビングや分析も重要である。文化・芸能分野では、国内よりも報道が早い⁸ことも少なくなく、海外から見た我が国に関するウェブ情報の収集・蓄積は、海外からの日本の見え方を分析し、外交を展開していく上でも、アーカイビングによる動態把握は重要不可欠であると考えられる。

国民生活の面からは、ウェブアーカイブの「利活用」の可能性に関して総務省において、「ウェブ情報アーカイブの普及促進に向けた利活用シーン等に関する調査研究」⁹が実施されている。この調査は、ウェブ情報アーカイブの社会的評価の把握を目的として、インターネットアンケート調査の実施と分析、および、企業・団体等ヒアリングの実施と分析を行ったもので、ウェブ情報アーカイブ化の社会的効果（利活用シナリオ別などで、ユーザ数等可能な限り定量的に整理）について言及している。インターネット利用上の課題として、ウェブ情報アーカイブが解決に資する「リンク切れ」「ページがなくなる」「少し前の情報がない（新しい情報ばかり）」との声があり、それに応える形で、以下のような調査結果が得られている：

- ウェブ情報アーカイブの利活用シナリオとして作成した9つのサービスタイプについて、1つでも「関心がある」とする回答（該当なし以外）は、職場では約5割、家庭では8割を超え、高いニーズが確認できている。
- 特にニーズの高いものは、職場では「過去ニュース提供」、「過去製品情報提供」、「企業活動」の順である。一方、家庭では「料理レシピ提供」、「過去ニュース提供」、「過去製品情報提供」であり、これらの関心度は2割を超えている。
- さらに利活用シナリオを閲覧後に利用意向を尋ねたところ、概ね7～8割が「利用してみたい」以上の高い利用意向を示している

総じて、家庭における属性別の利活用への関心は男女・年齢の別なく高い。また、ウェブへの接触時間が増えつつあると想定される60代以上で高いことが特徴である。60歳代

⁷ (財)京都国際文化交流財団が、京都デジタルアーカイブセンター作成した京都の文化遺産に関するデジタルアーカイブコンテンツ等をウェブ発信している (<http://www.kyo-bunka.or.jp/>)。

⁸ 政治・経済はもとより、文化・芸能の分野においても日本への注目度が高まっている（クールジャパン等）。特にアジア圏においては、すでに市場が形成され、国内より報道が早いことも少なくない。例えば、「浜崎あゆみの難聴告白」報道は、国内よりも海外の報道の方が早く、香港 - 蘋果日報（壹蘋果網絡）<http://www.atnext.com/>や、中国 - 新華社（新華網）<http://www.xinhuanet.com/>から報道されたが、現在、その記事そのものはすでに削除されている。

⁹ (株)三菱総合研究所：ウェブ情報アーカイブの普及促進に向けた利活用シーン等に関する調査研究報告書、平成17年度総務省委託調査、平成18年3月

では「トピックス」「過去ニュース提供」などの情報を深掘するサービスへのニーズが高いことが、また30代の子育て世代で「キッズサイト」へのニーズが高くこの世代を通じて、子供世代でのアーカイブの利活用が進む可能性を示している。

3. 3 学術としての意義

ウェブ上には様々な事象に対する人々の経験、感想、意見、および評価が現出し、変化を続けている。ウェブアーカイブに蓄積されるこれらの情報は、人々の意識の歴史的変遷ともみなせることから、言語学、社会学、政治学等、多様な学術分野においてこれまでにない新たな研究調査の対象および情報源として幅広く利用可能である。

言語学の観点からは、継続的なウェブアーカイブは貴重な言語資源とみなせる。例えば、ブログ、掲示板等のメディアに既存の辞書にはない新しい造語が頻繁に用いられることから、ウェブアーカイブを用いて新造語の時間的な普及の度合いを調査し、一過性の流行で終わるか、定着していくかなどを観測することができる。また、既存の語彙に関しても、時間の経過とともにその使われ方が変化していくことから、重要な言語学的調査の対象となる。

社会学、社会心理学の観点からは、ウェブアーカイブは人々の社会活動、および社会的概念に関する意識の履歴とみなせ、またウェブ上での人々の活動そのものが新たな社会現象とも捉えられ、調査対象として注目を集めている。米国コーネル大においては、2006年より米国インターネット・アーカイブより一部データの提供を受けて、Web Labなる社会科学への応用を目的としたウェブ分析基盤が構築されており、社会学者との共同研究によりイノベーションの普及過程等の調査が行われている。

政治学の観点からは、ウェブアーカイブはその時々世論を反映したものとみなせ、選挙、政府の施策、政治家の発言等に対する人々の反響や意見を調査することが可能となる。近年、政治家および政党はホームページや動画サイト等で情報発信を行っており、ウェブ上の情報も政治活動を観測するうえで重要な情報源となっている。米国議会図書館のMINERVAプロジェクトにおいては、大統領選挙に関して、候補者、政党、評論家等のホームページやブログ、および関連する情報を選別して収集、蓄積しており、同様に、米国政府、上院、議会に關係するホームページや、先の教皇選挙に關係するホームページ等、重要な政治的イベントに關するアーカイブを作成して研究者に提供を行っている。ウェブ情報を用いた政治学的研究としては、ブログネットワークの構造を用いた支持政党による意見対立分析などが行われている。

こうした様々な学術調査のみならず、論文、学会発表など学術成果の影響力をいち早く知るためにもウェブアーカイブを利用することが可能である。通常、論文の影響力の評価は他の論文からどの程度参照されたかを基準とするインパクトファクターを用いて行われるが、論文出版のスピードを考慮すると評価に必要な参照数を得るまでには数年の時間を必要とする。しかし、ウェブ上では論文が発表された直後からその内容につ

いて議論が行われるため、論文の参照よりも早く様々な反応を知ることが可能となる。

以上述べてきたように、ウェブアーカイブは多様な学術研究調査のための貴重な情報源として今後重要な役割を果たすものであり、これまでにない新たな調査手法を創出するものである。また学術成果そのものの影響力を調査するための情報源としても期待される。

3. 4 経済的意義

経済・ビジネスにおけるウェブアーカイブの効用については、3つの側面、すなわち、商品・サービス売上の側面、企業情報の側面、市場を取り巻く環境の側面、が考えられる。

企業は商品やサービスの売上により、少しでも高い利益を得るため、様々なマーケティング活動を行っている。近年、CGM やブログの普及により、消費者が自由に情報を発信し、企業や他の消費者とのコミュニケーションを深めており、その結果、賢くなった多数の消費者の意見が「集合知」として、少数のマーケティング専門家の意見より優れたものが出てくるかもしれないという状況になってきた。ジェームズ・スロウィッキー氏が” The Wisdom of Crowds¹⁰⁾” の中で述べているように、集団を構成する個人から意見が発せられていて（独立性）、様々な分野からの意見が集まり（多様性）、各自の意見に基づいた意思決定がなされており（分散性）、分散された情報が集約される（集約性）という状況下で、集合知は発揮される。インターネットを通じて発信されたブログの蓄積情報を活用したトレンド分析は、まさに集合知の代表事例といえる。ウェブ上から収集し蓄積したブログ情報から、ユーザーニーズのトレンド分析を行うことにより、効率的なマーケティング活動ができ、経済活動をより活性化することにつながる。

企業情報の側面についていえば、現在、各企業は自社内に散在する様々な情報の集約・蓄積・分析をおこない、戦略を決定している。内部統制強化の観点からも、ウェブ上での公開情報を含む様々な企業情報の蓄積が必須となっている。蓄積された情報は、社内情報基盤として使われるのみならず、企業内で問題が生じた際の証拠としても保持される。また、その時代を表す経済活動の記録としての意味ももつ。

企業・市場を取り巻く環境の側面からみた場合には、グローバルな企業活動が活発化している現在、リスクマネジメントの観点から、各国における知的財産の扱いや法律の変遷を常時把握しておくことが不可欠である。万が一、訴訟にいたった場合、迅速に対応しなければならず、そのためには、時系列で変化情報を蓄積しておき、迅速にその変化分を含めて検索・分析できるような仕組みの構築が望まれる。また、我が国の国際競争力向上の観点からは、各国の経済活動や企業分析も非常に重要と考えられ、そのためには各国の経済情報アーカイブとの連携を整えていく必要があると思われる。

¹⁰ The Wisdom of Crowds : <http://www.randomhouse.com/features/wisdomofcrowds/>

4. ウェブアーカイブ構築に関する諸側面

4. 1 国によるアーカイビング

国（政府）によるウェブアーカイブにおいては、インターネットの世界への国家権力の介入と受け止められることのないよう配慮が必要である。紙の出版物と比べて、書き換えや消去が容易なメディアの特性から、ウェブでの情報発信は、これを一時的、暫定的なものとして行っていることも考えられる。国の機関がウェブ情報を「永続的」なものとして収集、蓄積することが、情報の国家管理と受け止められることのないよう、ウェブアーカイブを社会的に利活用することについて、発信者を含めた社会全体の理解を啓発することが肝要である。

この他の課題としては、ウェブ上では、不正な著作物の使用、児童ポルノ等の違法コンテンツの発信、公序良俗の観点で問題となる有害情報の氾濫といった状況が生じており、違法・有害情報をアーカイブすることの問題、また、アーカイブ情報を裁判係争（名誉毀損、権利侵害等）の証拠とすることへの対応、個人情報保護の観点からの収集及び提供等の法的検討も必要である。

4. 2 法制度面について

法制度面では、アーカイブに伴う知的財産権との抵触が重要な課題である。著作権については、収集蓄積の段階で著作権者の複製権許諾が必要であり、さらにネット上で発信を行うためには公衆送信の対応も必要となる。公益的観点からの権利バランスを踏まえた制度の実現が必要となる。

また、諸外国の国立図書館では、出版物の納本法制の枠組みの中で、ウェブアーカイブを整備している事例が多いが、我が国においても、著作権法の整備に加えて、国立国会図書館法の改正により、納本法制の枠組みの中で、ウェブアーカイブの制度的実現をはかることが、喫緊の課題である。

4. 3 ウェブ情報の収集蓄積の技術的側面

ウェブ上の情報は現在も爆発的な増加を続けており、音声、画像、動画、自動生成コンテンツ、対話的コンテンツなど多様なメディアへの進化も日々進んでいることから、完全な収集蓄積は原理的に困難となってきている。特に、ウェブで発信される音声、画像、動画等のマルチメディア情報のアーカイビングには、経済的にも技術的にも課題が多い。これまでに述べてきた多様な利用を可能とするためにはできる限り網羅的な収集または分散収集が必要となるが、必要かつ十分な収集範囲を規定することが重要な課題となる。

さらに、収集/蓄積した情報を保存するだけでなく、効率的効果的に利活用できるようにするための検索技術やユーザインタフェース技術の開発も重要である。また、多く

のウェブサイトは背後のデータベースから動的にウェブ情報を検索・生成する、いわゆる深層ウェブ(deep Web)という構成を取っており、このようなdeep Webのアーカイビングや、既存の検索エンジンが生成した索引情報のアーカイビングも重要な課題である。また、博物館・美術館などの場合にはアーカイビングが分散して行われることが普通であるが、利用者の立場からは、これらを横断的に検索・利活用することができるような検索技術も重要である。

総じて、ウェブ上の情報メディアは進化し続け、これに伴う技術開発が今後も重要であることから、中核的・永続的保存機関と最先端の収集・検索・分析技術を有する大学・研究機関との連携¹¹が重要である。

4. 4 文化・社会・生活の側面

文化遺産のデジタルアーカイブ化については、まずは、経済的な問題(コスト)が大きな課題である。さらに、現状では、デジタル化されていないコンテンツは地方に多々あり、デジタルアーカイブの推進には全国的な視野が必要である。また、各機関それぞれがアーカイブ化したコンテンツ相互を連携した利活用が可能となることが重要である。

デジタルアーカイブ化、および、ウェブアーカイブ化は、保存という観点から国家主導で推進するという視点に加えて、アーカイブされるコンテンツの国民による利活用や国民に対する説明責任という視点や、脱工業社会における生産手段としての知識や知識の循環に基づく新産業の創出という視点から推進することもきわめて重要である。

4. 5 学術的利用に関して

様々な学術利用に対応するためには、調査対象に関する情報をウェブアーカイブから適切に抽出可能とすることが必要となる。コンピュータの専門家でなくとも必要なデータ抽出を可能とするデータアクセス機構を構築することが課題となる。また、学術成果の影響力を観測する際には、既存の学術情報データベースとの比較対照が行えるよう柔軟な連携機構を構築することが課題となる。

4. 6 経済・ビジネス利活用に関して

経済・ビジネスにおけるウェブアーカイブを考えると、その主目的が経済活動、すなわち、企業や個人が利益を得ることであることから、アーカイブをビジネスに活用する場合には、企業側がある程度の負担をする必要があると考えられる。

想定されるすみ分けとしては、企業が各社独自の事業をおこなう上で発生する情報に関しては、各社が自社責任で情報蓄積し、企業を超えた取り組みにより日本の国際競争

¹¹ 検索エンジン Google は、米国の国家プロジェクトである DLI (Digital Library Initiative)から生まれたことは著名である。

力向上に貢献できると考えられる事柄や、企業が証拠として第三者による保存を希望する情報に関しては、国やその他の第三者機関としての対応が望まれる。

例えば、経済分野のウェブアーカイブについて、企業独自の努力にのみ頼った場合、次にあげるような問題が発生すると思われる。

- 企業や国家組織が存続している間は、内部統制からもほぼ確実に自社組織内で必要な情報を蓄積し管理し続けられるが、組織の統廃合にともない、蓄積主体が消えてしまうと、企業情報の継承ができない。統廃合の変遷も包含できるような統合管理をおこなうことはできないか。
- 意義の項で述べたように、日本企業の国際競争力を向上するため、各国地域の法律や知的財産権の扱いなどの時系列変化を蓄積し、各社共通に検索や分析ができるような、グローバル経済活動を下支えするインフラ的なシステム構築をできないものか。

経済分野において、国やその他の第三者機関と企業の役割分担を明確に引くことは非常に難しいことではある。いずれの事例においても、極めて限られた範囲であれば企業努力でカバーし得ると考えられる。しかしながら、経済活動全体がグローバル化している現在、まずは、日本企業の国際競争力を強化するという観点から、国のインフラとして持つべきウェブアーカイブの範囲や、産官学や中央・地方など、各種組織間でどう連携するかについて議論することは、非常に意味のあることと思われる。

5. おわりに

本記録では、学術の新たな情報基盤となりうるウェブアーカイブ構築の動向、ウェブアーカイブの意義を明らかにするとともに、その実現に向けた具体的な課題についても取りまとめた。

ウェブは新しいメディアであるが、急速に、かつ、広範に広がりつつあり、今や、重要な社会情報基盤の位置を占めつつある。その範囲は、学術活動の情報基盤のみならず、国家、文化、社会、国民生活、経済活動など広範囲に及び、過去に発信されたウェブ情報を、その同一性を保証しつつ収集・保存するウェブアーカイブの構築の重要性および緊急性は論を待たない。ポーンデジタルな情報を発信するウェブは、ポーンデジタルであるが故に利活用が容易である反面、過去のウェブ情報は消滅しやすいという特性を併せ持っているからである。ウェブアーカイブの構築を行うことの意義は大きいが、一方、以下に述べるように解決すべき課題も山積している。

ウェブアーカイブに関する学術的な観点からの課題は、学術分野においてはその収集保存範囲が我が国に閉じていないこと、言語の壁、国立研究機関や学会などの収集保存主体の連携をいかにして行うかなどである。また、ウェブというメディア自体が依然発展途上にあるため、ウェブアーカイブの構築およびその検索・利活用のための技術研究開発が重要であるが、これをどのように推進するかも最重要課題の一つであろう。

アーカイブの主体の選択に関しては多様な考えがあり、結論を得るには至っていない。一つの組織が全ウェブ空間のアーカイブを実施することも考えられるが、コンテンツに応じて複数のステイクホルダーが参画し、互いに連携することも考えられる。国の機関がアーカイビングを行うことに対しては、プライバシーに関する懸念や、血税を使って構築することに対する説明責任などに課題がある。反面、国家や地域のアイデンティティのための情報基盤となること、情報公開の促進につながることなど国家が責任を持ってアーカイブすべきところも大きい。また、プライバシーに関する懸念等に対しては、「永久保存は行うがアクセス制限を行う」などの技術と社会制度によって解決が期待できるものも多いと考えられる。

ウェブからの情報発信の影響力が著しく増大しつつある今日、ウェブアーカイブは膨大な個人の情報発信に基づくデジタル民主主義の促進につながる大きな力を有しているが、それ故に発信情報の同一性保証等による安全安心な社会活動の基盤を構築することは、重要な課題となる。経済的な側面だけではない利点への配慮が必要であるが、技術的な課題や妥当な経済的負担モデルについても多様かつ詳細な方策検討が不可欠である。即ち、一般に時間の経過とともに多くの情報は利用者にとっての価値が減少する傾向にあり、アーカイブは新鮮な情報だけを対象とする現行のサーチエンジンとは異なる収益モデルを確立する必要があり、持続可能な形態については議論を深める必要がある。

ウェブアーカイブの収集対象も結論を得るに至っていない。多くの有害なページ、スパムページも存在するウェブ空間の中で、何をアーカイブの対象とすべきかは、利用者によって大きく異なるところである。ウェブ空間の健全性を維持するためには、スパム構造の解析は不可欠であり、一見無駄にも見えるスパムの保存は、スパムの時代的変遷を見る興味深い研究素材とも捉えられる。多様なステイクホルダーが独自の空間を保持することが考えられるが、収集対象を効率的に分別することは技術的に容易でないという課題もある。

知財財産推進計画 2008 にも著作権の制限について触れられているように、今後の著作権法の制度改定などの議論はウェブアーカイブに強く関連しており重要である。ブログ等の自由意思での発信とともに、ウェブ上には多くの有料コンテンツも存在する。ウェブアーカイブの構築にあたっては、後者に関してはオリジナルサイトへの流動を担保するなど、コンテンツホルダーへの特段の配慮が必要であろう。納本制度の改定に関する議論も重要である。

文化、社会、国民生活に資するウェブアーカイブの構築という視点からは、単なる保存収集のためのウェブアーカイブではなく、これをどのように利活用するかという視点が重要である。アーカイブされるコンテンツの国民による利活用や国民に対する説明責任という立場からの構築・運営が重要であるとともに、脱工業社会における生産手段としての知識や知識の循環に基づく新産業の創出という経済的視点から推進することが

重要である。

経済的な観点からは、ウェブアーカイブ構築に関しての国と民間の役割分担とともに、産官学連携や中央/地方連携のもとにいかに関築を行うかが重要な論点である。この課題については、米国の議会図書館と大学・NPO 等との連携、米国グーグル社と New York Times 社や Internet Archive (NPO) との連携を参考にしていくべきであろう。また、我が国の国際競争力向上という視点でウェブアーカイブを論じることもきわめて重要である。

この様にウェブなるメディアは極めて大きく社会を変えつつある新しいメディアであり、そのアーカイブは意義深いものの、その実現のためには多様な課題が残されており、次期学術会議ウェブ・メディア社会分科会において議論を重ねて行きたい。