

日本の大学ウェブサイトのアーカイブ状況 —— Internet Archive と WARP の比較

塩崎 亮

1. はじめに

日々、多数の人々に参照・利用されているウェブサイトは、紙媒体の書籍と異なり、安定した存在とはいえない。その内容は更新・削除されやすく、URLが変更されることもあれば、ウェブサイト自体が閉鎖される場合もある。このような消失リスクを回避するため、ウェブ上の情報の長期利用を保証する取り組みが多数生まれてきた。代表例は、米国の非営利組織Internet Archive（以下「IA」）が提供するWayback Machineだろう¹⁾。IAは「フェアユース」の考えにもとづき、著作権者から許諾を事前にとることなく、日本のものを含め、世界中の公開ウェブサイト等を収集してきた。著作権やプライバシー等の観点から削除の依頼があれば事後的に対応するオプトアウトの運営方針にもとづく。他方、国内では国立国会図書館が法制度に則り対応している。「インターネット資料収集保存事業（Web Archiving Project：WARP）」という²⁾。だが法制度で定められた収集範囲は公的機関に絞られる。民間部門のウェブサイトは権利者側より明示的な合意をえてから収集されているが、現状その範囲と量は極めて限定的である。

一利用者からすると、収集されたデータの保管場所は重要でないかもしれない。しかし、収集主体について懸念する立場もありうる。日本から発信され、日本語で記された、日本に関する情報については、IAと別に国内でも保管しておくべきではないか、という懸念である。だがその前提として、IAとWARPで収集されたコレクションがどの程度異なるのかは必ずしも明確でない。そこで本稿では、事例として「大学」のウェブサイトを取り上げ、IAとWARPのコレクション内容の差異について概観したい。大学を対象とするのは、IAとWARP双方で収集されてきた資料群のためである。WARPの収集範囲には、法制度の範疇にある国公立大以

外に民間部門の私立大等も含まれる。このような点から「大学」は、IAとWARP間、機関種別間で比較可能なカテゴリと位置づけられる。本稿の問いは次の2点とし、以下、現状の記述を行う。

- ・IAとWARPにおいて、どのくらい大学のウェブサイトはアーカイブされてきたか
- ・IAとWARP間、また機関種別間で、アーカイブされたコレクションの内容に差異はあるか

2. 関連文献

他の多くの組織と同様、大学でもウェブサイトは重要な広報・情報公開ツールの一つと位置づけられている³⁾。必然的に、このような情報発信・共有の側面に関心が向けられてきた⁴⁾。一方、各機関でそれらデータがどのように保管されているかは不確かである。あるいは、大学史の編纂、組織のガバナンスや説明責任の観点から、大学に関する「アーカイブズ資料」を管理する重要性は日本でも指摘されてきた⁵⁾。だが、大学ウェブサイトがアーカイブズ資料として管理されている国内事例までは、文献上、確認できない。

翻って、欧米の一部の大学ではウェブアーカイブを構築する事例が見られる⁶⁾。ただしその大半は独自に収集を行うのではなく、外部サービスを利用しており、なかでもIAのサービスArchive-It⁷⁾による構築例が多い。さらにそのほとんどは、著作権法で「フェアユース」または「フェアディール」の規定があり、オプトアウト方式での運営が可能な米国またはカナダの大学で占められている。ただし、米国におけるウェブアーカイブの実態調査によれば、あるテーマで第三者のウェブサイト／ページをアーカイブする事例より、自機関のものを収集する割合が増しているともいう⁸⁾。

複数の異なるウェブアーカイブ間の収録内容を比較した研究もある。Ahmed AlSumらによる調査では、IAや米国議会図書館、英国図書館のもの

を含む12の公開ウェブアーカイブを対象（WARPは対象外）とし、ウェブページや言語単位でそれらの収録状況が分析された⁹⁾。結果、IAの収録範囲が最大であることが確認されている。あるいは、複数の公開ウェブアーカイブを横断的に検索可能なサービスTime Travel¹⁰⁾を対象とし、インドの*Annals of Library and Information Studies*誌に掲載された論文の参照URLの収録状況を確認したD. Vinary Kumarらの論考も挙げられる¹¹⁾。ここでも、IAが最大の収録率だったという。別の観点として、アーカイブされた大学ウェブサイトを対象とした研究もある。Scott A. Haleらは、英国の大学ウェブサイト過去15年分を含むデータセットにもとづき、大学間のハイパーリンク構造の変遷等を分析している¹²⁾。しかし、そもそもIAを含むウェブアーカイブに収録されたデータは、ウェブサイト全体の一部に過ぎず、かつ、収集できているウェブページにも偏りが見られる、と指摘されてきた¹³⁾。

日本語でウェブアーカイブについて論じられた文献としては、WARPの紹介やIAを含む国際動向の整理¹⁴⁾、システム構築の事例報告¹⁵⁾が蓄積されてきた。しかし、IAとWARPのコレクションを比較した調査、ひいては大学ウェブサイトを対象とした調査は、これまでにない試みといえる。

3. 調査手順

3.1 対象

本稿の主対象は国公立の「大学」ウェブサイトである。ただし調査項目によっては、機関種別の現状を把握するため、より広い対象とした（なお、本章の作業はすべて2019年7月に実施）。具体的には、文部科学省のHP上のリンク集¹⁶⁾で「高等教育機関」と分類された1,168機関（リンク集には統廃合大学が1件含まれていたが除外）と、WARPで「大学」と分類された845機関の集合とした。前者（以下「文科省リスト」）は、国公立大のほか、短大・高専の現存機関から構成される（放送大学は対象外とした）。後者（以下「WARPリスト」）には、すでに統廃合等で現存しない機関（以下「統廃合大」）のほか、独立行政法人の研究機関や大学協会（以下「その他法人」）が含まれる。文科省リストとWARPリストには機関番号等のIDが含まれ

ないため、両リストは機関名で突合した。表記揺れがあった数機関についてはWARPリスト側を修正した。逆に、現存大学の名称変更が文科省リストで未反映の数機関については文科省リスト側を修正して対応した。結果、重複を除くと総計で1,243機関となった（以下「集合リスト」）。内訳は、国立86、公立93、私立610、短大322、高専57、統廃合大66、その他法人9、となる。過去に存在した大学一覧があればより包括的な調査が可能だったが、諸資源の制約上、実施していない。

3.2. 手順

まず文科省・集合リストを対象とし、IAでの収録率をWayback Availability JSON API¹⁷⁾で確認した。返戻ステータスコードが正常処理を示す「200」であれば収集済みと判断した。なお、IAの起点URLを特定できなかったため、引数は、文科省リストのURLのFQDN部分とした。そのため、四年制と短期の大学等、同一ドメイン名のもとに複数機関のコンテンツが配置されたウェブサイトの場合、同じFQDNで確認したことになる。また、WARPの起点URLとは異なる可能性もある。

WARPでの収録率は、文科省リストとWARPリストで機関名を照合して確認したが、例外が2点ある。一つ目は国立の高専51機関で、これらは「大学」でなく「法人・機構」として分類されており、同一起点URLの「国立高等専門学校機構」として一括収集されていた。二つ目の例外は短大である。それらは四年制大学のウェブサイトに含まれている可能性があった。そこで、FQDNが同じものはWARPで収集された最新のデータ断面を目視で確認し、「短大」まで含まれていれば収集済みと判断した。ただし館内限定公開の場合、収集済みか否かは館外から確認できないため、一律収集対象外とした。また、「学術情報基盤実態調査」で用いられている規模別の区分（A：8学部以上、B：5～7学部、C：2～4学部、D：単科大）でも集計を行った。この規模別大学一覧表は平成30年度版が最新のため、未掲載大学は各大学のウェブサイトで学部数を確認した。

補足的に、収録率に影響を及ぼす可能性を考慮し、クローラーのアクセス制御設定についても見

た。具体的には、各大学ウェブサイトのルートディレクトリ直下にrobots.txtファイルがあるか確認し、その有無とクローリング拒否の設定内容、さらにIA (archive.org_bot) とWARP (ndl-japan)、そしてWayback Machineにデータを提供してきたとされるAlexa社 (ia_archiver)¹⁸⁾のクローラーに関する指定有無を記録した。

あわせて、集合リストを対象とし、収集されたコンテンツの公開状況を確認した。IAについては収集状況の確認と同じく、FQDNを引数としたWayback Availability JSON APIの返戻ステータスコードで判断した。WARPについては、ウェブでの一般公開のほか館内限定公開という選択肢もあり、その場合は詳細画面上に記載されている。そこで、データ断面(収集時期)によって公開条件が異なるものもあるが、一律、調査時点の直近収集分を対象とし、「館内限定公開」の記載有無を機械的に確認した。なおIAとWARPともに、ウェブページ単位では非公開の場合もありうる。

また、文科省リストを対象とし、IAとWARPでの収録期間や収集頻度の比較ができるように、収録断面数と収録期間について詳細画面上の値をスクレイピングで収集した。ドメイン名が変更されている大学もありえたが、前述の通りIAについては文科省リスト記載のFQDNでの確認にとどまっている。収録期間については、収集開始日から調査時点の直近収集日までを月単位で集計した(直近収集日の当該月を含む)。なおIAの値については、その確からしさを確認するため、Wayback CDX Server API¹⁹⁾により、ステータスコードが「200」のもののみ集計する作業も補足的に行った。結果、すべての機関において、詳細画面上の値とAPIによりえた値とで若干の差があることを確認した。これがデータベースの更新タイミングに起因するのか、集計ロジックの違いによるのかまでは判然としない。ただし、数機関(東大等の大規模大学)についてはこの差が1,000以上となっていた。これら大学のウェブサイトは、少なくとも調査時、ルートディレクトリから下位ディレクトリへリダイレクトされる挙動で、何らかの関係があると推察される。とはいえ、いずれも収集はリダイレクト後に成功していたため、以降の集計では

スクレイピングで取得した画面上の値を一律採用した。

4. 調査結果

4.1. 収集状況

ここから調査結果について整理する。まず収集率について、IAでは文科省リストの全機関を収集していることが確認できた。集合リストであってもIAはすべて収集しており、収集率100%である。WARPの収集率は表1にまとめた。法制度の対象となる国公立大、国立高専は100%の収集率を示している。私立大も80%を記録したが、逆にいえば、残りの122大学については、IAで収集されているにも関わらず、いまだWARPに対しては収集を許諾していない状況となる。表2ではWARPでの私立大の収集状況を規模別に示した。若干だが、規模が小さいほど未収集の機関が多い傾向にある。また理由は定かでないが、WARP対象のうち1年以上収集されていない私立大が8機関あった。

加えて、短大の収集率は6割にとどかない結果となった。しかし、調査手順で示したとおり、館内限定公開の場合等、目視確認で把握できなかったものが存在する可能性を否定できない。そのため短大の収集率はもう少し高いのが実態だろう。

表1 WARPにおける大学等ウェブサイトの収集率

区分	機関数	WARP対象	
		件数	割合
国立	86	86	100%
公立	93	93	100%
私立	610	488	80.0%
短大	322	173	53.7%
高専	57	54	94.7%
全体	1,168	894	76.5%

表2 WARPでの私立大ウェブサイトの規模別収集状況

区分	収集済		未収集	
	大学数	割合	大学数	割合
A	38	88.4%	5	11.6%
B	70	88.6%	9	11.4%
C	213	78.9%	57	21.1%
D	167	76.6%	51	23.4%
全体	488	80.0%	122	20.0%

4.2. 収集拒否設定の対応状況

クローラーのアクセス制御設定の状況については結果を表3にまとめた。ここから、全体の半数程度である597機関（51.1%）が何らかの制御設定をしていることが見て取れる。制御設定がなされたもののうち、あるディレクトリ配下の収集を認めないことを表明する「拒否設定（Disallow）」が記されたウェブサイトも多く、全体の46.9%（548機関）を占める。ただしそのうち、IAのクローラーに対する拒否設定を記載した例はなく（Alexaについては4件）、WARPについても5件に過ぎなかった。

表3 クローラーの制御設定

区分	機関数	クローラー制御設定あり			
		収集拒否設定あり			
		WARP	Alexa		
国立	86	46	44	2	2
公立	93	51	47	1	0
私立	610	315	292	1	0
短大	322	153	135	1	1
高専	57	32	30	0	1
全体	1,168	597	548	5	4

4.3. 公開状況

次にIAでの公開状況だが、単純にWayback Availability JSON APIのレスポンス結果に従うと、少なくともウェブサイト単位ではすべて公開済みということが確認できた。WARPでの公開状況については表4にまとめた。表1のWARP対象894機関に加えて、統廃合大とその他法人も含めた969機関分となる。

結果、WARPで一般公開されている大学ウェブサイトは856機関（88.3%）、館内限定公開のものは113機関（11.7%）となった。内訳をみると、国公立とその他法人で館内限定公開の比率が相対的に高い。他の種別は一般公開が大半である。なお、統廃合大で限定公開が6機関ある。これらの担当窓口が不明確な場合、将来的に一般公開へ切り替わる可能性は低くなってしまふことが危惧される。

表4 WARPにおける大学等ウェブサイトの公開状況

区分	一般公開		館内限定		計
国立	59	68.6%	27	31.4%	86
公立	60	64.5%	33	35.5%	93
私立	457	93.6%	31	6.4%	488
短大	160	92.5%	13	7.5%	173
高専	53	98.1%	1	1.9%	54
統廃合	60	90.9%	6	9.1%	66
その他	7	77.8%	2	22.2%	9
全体	856	88.3%	113	11.7%	969

4.4. 収集開始時期と収録期間

IAとWARPにおける収集開始時期については、国公立大に絞り、図1にまとめた。

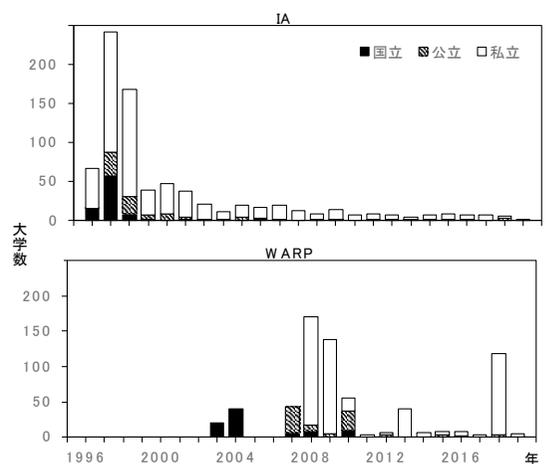


図1 IAとWARPにおける収集開始年別の大学数

IAは1996年から開始された取り組みだが、大学ウェブサイトについても、その開始年から収集が一斉に開始されたことが見て取れる。その後は、新設・機関名変更大学（あるいはドメイン変更）等に対応して新規収集を個別に開始している傾向もうかがえる。他方、WARPでは国立大の収集が2003年から始められ、徐々に対象を2007年から公立大、そして2008年から私立大へと拡大させていった状況が明瞭となった。2013年と2018年は収集開始対象数が突出しているが、おそらく、許諾依頼を再度実施した成果の反映だろう。なおこの作業を介して、IAよりもWARPの収集開始年が早い（収録期間が長い）機関が18機関あり、それらの一部はドメイン名の変更に起因することが実際に確認

できた。しかし、他にもドメイン名が変更された機関もありえたため、補正作業は行っていない。この点も本調査の制約となる。

収録期間は表5に示した。単純に、事業開始年の差が収録期間の差を生んだとってよく、総じてWARPよりIAの収録期間が長い。いいかえれば、IAは先行して大学ウェブサイトを収集し始め、かつ継続して収集している。もう一つ明瞭なのは、IAとWARP双方とも、国立大（あわせて国立の高専）の収録期間が長く、そのばらつきの度合いが小さい点である。これは、他の種別よりも新設・統廃合がなく、最初期から収集対象となっていることの裏返しといえる。またWARPにおいては、収集開始年が遅かった私立大、そして特に短大については（本調査によれば2018年以降に収集開始された短大が92機関 [53.2%]）、必然的に収録期間が短い。

表5 大学ウェブサイトの収録期間（単位：月）

区分	平均値	中央値	標準偏差	
IA	国立	261	266	20.8
	公立	220	251	68.5
	私立	220	247	63.8
	短大	214	229	54.3
	高専	244	263	47.5
WARP	国立	168	184	28.4
	公立	111	115	34.4
	私立	85	114	48.1
	短大	49	10	47.1
	高専	134	140	24.2

4.5. 収録断面数と収集頻度

収録断面数については表6にまとめた。断面数とそのばらつき度合い双方ともに、総じてIAの方がWARPよりも大きい。対してWARPの断面数は、IAと比較すると小規模にとどまっている。

表6 IAとWARPでの大学等ウェブサイトの収録断面数

区分	平均値	中央値	標準偏差	
IA	国立	827	609	628.8
	公立	366	357	191.9
	私立	425	336	373.1
	短大	262	210.5	201.6
	高専	260	244	93.4
WARP	国立	41	41	1.9
	公立	34	37	9.3
	私立	21	28	10.6
	短大	14	4	11.4
	高専	38	40	6.3

図2は、単純に収録断面数を収集期間で除算し、収集頻度（ひと月あたりの断面数）としてまとめたものである。IAでは国立大の収集頻度がもっとも高く、平均でも、ひと月に3回以上（3.14回）収集されていた。他の種別でも、平均でひと月に1回以上の収集頻度を示した。ただし国立・私立大では、ばらつきが大きく、外れ値が多い。一方、IAと比較するとWARPの収集頻度は低く、機関種別でもほとんど差がない。おおむね、3～4ヶ月に1回程度収集されてきたことがうかがえる。実際、WARPの「よくあるご質問」では、“国の機関は月に1回、それ以外の制度収集の対象機関は年4回収集し、“制度収集の対象機関以外のウェブサイトは、原則として年1回収集”していると記されている²⁰⁾。

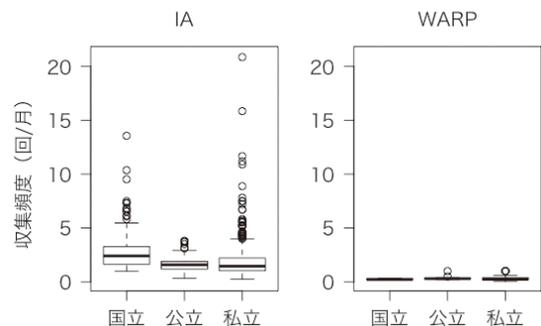


図2 IAとWARPの収集頻度の分布

5. 考察

まず一つ目の問い、IAとWARPにおける大学ウェブサイトの収集率について、前者は100%、後者も国公立大は100%を示したが、他の種別、特に私立大はその2割がWARPの収集対象外であった。とはいえ、収集を拒む理由までは明らかでない。著作権・肖像権といった権利関係や第三者による不正利用（例：偽サイトの構築、スパムアクセス先URLリストとして乱用）を考慮しているためなのか、ただ傍観的に回答が保留されているだけなのかを明らかにするには、インタビューや質問紙による追加の調査が欠かせない。だが、それらWARP対象外の機関は、すべてIAの対象に含まれており、かつ、Wayback Machineを介して公開されている。収集を拒否する強い理由があるのならば、IA側にデータの削除要求を行うべきだろうし、クローラーの制御設定やIP制限をかけていてもおかしくないが、その痕跡は見当たらなかった。そうだとすると、特段の理由なくWARPの収集対象外となってしまうことになるが、国のコレクションとして、これでは不完全といわざるをえない。

あわせて、WARPで限定公開とされている大学ウェブサイトが、少なくともウェブサイト単位ではIAで公開されていることも確認した。大学側はこの矛盾をどのように捉えているのだろうか。国内の国の機関が日本語のインタフェイスで全文検索も含めて提供しているサービスには制限をかけたい、ということなのだろうか。ただし、ウェブサイトを管理する立場からすると、これはある意味、WARPの方が自身らの意向を反映させやすい（積極的に活用できる）、ということでもある。自分らでデータを維持管理せずとも、IAとWARPの双方に収集されていれば、分散型バックアップとなり、消失リスクを低減できる。

次に二つ目の問い、IAとWARP間、機関種別間でのコレクション内容の差異について整理する。調査の結果、収集開始時期はIAの方が早く、必然的に収録期間もIAの方が長くなっており、かつ、収録データ断面数でもIAはWARPを凌駕していることが明瞭となった。特に、収集開始時期が遅い

私立大・短大ではその傾向が顕著である。収集頻度も同様の構図だったが、機関種別に見ると、IAでは国立大の頻度が高いのに対し、WARPではどの種別もさほど差がなかった。他方、IAでどのように収集頻度が決められているか（サイト構造等の技術的要因が関係するのか、更新頻度や人気度に応じて設定されているのか）は明らかでない。robots.txtの設定内容をIA（またはAlexa社）がどこまで遵守しているかも不確かである。加えて、IAの収集頻度は国立・私立大とも大学によって差が大きい。もっともデータのばらつきが大きい私立大について、大学規模別に収集頻度の中央値を見ると（A:2.9 > B:2.4 > C:1.4 > D:1.2）、少なくとも、大規模大学ほどIAではよく収集されている傾向にある。とはいえ、繰り返しになるが、その理由までは明らかでない。

WARPでは、大学の規模や種類に関係なく収集頻度が設定されている。自機関でウェブアーカイビングを行う例のほか、英米圏では、Archive-Itを含め、ウェブアーカイブの代行サービスがいくつか存在するが、これは、自機関で収集対象と収集頻度をコントロールしたいニーズが一定程度あることを示唆している。同様のニーズが日本にも存在するのであれば、さらにいえば、現行制度のもと民間部門の収集対象を広げる場合においては、収集・公開の是非に加えて、収集頻度についても収集される側の意思を反映可能なサービスが求められるのかもしれない。

6. おわりに

本調査では、大学ウェブサイトを対象とし、IAとWARPのコレクション内容を比較した。結果、収集率・収録期間・データ断面数ともにIAがWARPを凌駕していること、オプトインの方針で運営されているWARPは、大学側の意思をある程度反映したコレクションとなっていることを確認した。あわせて、IAで収集・公開されてきたにも関わらず、WARPでの収集を許諾していない大学が多数存在することも明瞭となった。国のコレクションとしては不完全といえ、収集を拒否する明確な理由がないのであれば、それらもWARPの収

集対象に含まれていくことが期待される。

ただし、本調査にはいくつかの制約がある。第一に、過去に存在した、あるいは機関名変更前の大学一覧にもとづいた調査でない点が挙げられる。特にWARPはIAよりも事業開始年が遅いため、IAにしかない大学ウェブサイトが存在するだろう。第二に、IAの収集状況について、文科省リストをもとにFQDNのみで確認した点が挙げられる。ドメイン名が変更された機関の場合、過去分の収集状況まで反映できなかった。WARPの過去の起点URLを利用するアプローチもありえる。

また、さらなる課題も残った。1点目は、ウェブページ単位での収集率の算出である。「2. 関連文献」で触れたとおり、IAもまた不完全なコレクションである。本稿では、ウェブページ単位の収集率まで調査できていない。対象範囲にもよるが、目視確認は量的に困難なため、データ欠損やレイアウト崩れをどこまで許容するかといった判定基準の策定と機械的な判別の仕組みが求められる。2点目は、ウェブサイト以外のコンテンツのアーカイブ状況である。代表例はSNSだが、将来的にこれらも貴重な大学の史料となりうる。IAを含め、欧米ではすでに関連事例の蓄積があり、国内でも議論を開始すべき時期ではないだろうか。

謝辞：この研究は聖学院大学の2019年度学長裁量経費(19-02)の助成を受けています。WARPの収集率算出等については国立国会図書館の福嶋聖淳氏から助言をいただきました。記して感謝いたします。

注・引用文献

- 1) <https://archive.org/web/>
- 2) <http://warp.ndl.go.jp/>
- 3) 「小特集 大学Webサイトリニューアルの取り組み」『大学時報』no.398, 2019, pp.76-95.
- 4) 「大学スマホ・サイトユーザビリティ調査」日経BPマーケティング, 2019. <https://web.archive.org/save/https://www.nikkeibpm.co.jp/item/1060/1060/index.html>
- 5) 菅真城『大学アーカイブズの世界』大阪大学出版会, 2013, 296p.
- 6) Brügger, N. *The Archived Web : Doing History in the Digital Age*. MIT Press, 2018, pp.96-97 ; Adler, P. S. *et al.*

Code of Best Practices in Fair Use for Academic and Research Libraries. ARL, 2012. <https://web.archive.org/web/20191030005538/https://www.arl.org/wp-content/uploads/2014/01/code-of-best-practices-fair-use.pdf>

- 7) <https://archive-it.org/>
- 8) Farrell, M. *et al.* *Web Archiving in the United States : A 2017 Survey*. National Digital Stewardship Alliance, 2017. <https://osf.io/ht6ay/>
- 9) AlSum, A. *et al.* "Profiling web archive coverage for top-level domain and content language," *Int J Digit Libr.* vol.14, no. 3-4, 2014, pp 149-166.
- 10) <https://timetravel.mementoweb.org/>
- 11) Kumar, D. V. *et al.* "Recovery of Missing URLs cited in Annals of Library and Information Studies : a study of Time Travel," *Annals of Library and Information Studies.* vol.66, no.1, 2019, pp.24-32.
- 12) Hale, S. A. *et al.* "Mapping the UK webspace : fifteen years of British universities on the web," *Proceedings of the 2014 ACM conference on Web science*. ACM, 2014, pp.62-70.
- 13) Hale, S., *et al.* "Live versus archive : Comparing a web archive to a population of web pages," *The Web as History*. Brügger, N. *et al.* (eds.) . UCL Press, 2017, pp.45-61.
- 14) 例として次を参照：前田直俊「ウェブアーカイブの活用に向けた動き：世界の潮流とWARPの取組」『カレントアウェアネス』no.331, 2017, pp. 9-13.
- 15) 例として次を参照：喜連川優ほか「Socio Sense：過去9年に及ぶWebアーカイブから社会の動きを読む」『情報処理』vol.49, no.11, 2008, pp.1290-1296；竹内勇希「ウェブアーカイブシステムの構築：新聞社サイトでのニーズと可能性」『新聞技術』vol.61, no.1, 2017, pp.71-74.
- 16) https://web.archive.org/web/20190722024633/http://www.next.go.jp/b_menu/link/1294885.htm
- 17) https://web.archive.org/web/20191113014329/https://archive.org/help/wayback_api.php
- 18) <https://web.archive.org/web/20191009160355/https://archive.org/details/alexacrawls>
- 19) <https://web.archive.org/web/20191111204556/https://github.com/internetarchive/wayback/tree/master/wayback-cdx-server>
- 20) https://web.archive.org/web/20190531013324/http://warp.dandl.go.jp/info/WARP_qanda.html#01_02

(URLの最終確認日はすべて2019-11-16)
(しおざき・りょう 聖学院大学基礎総合部准教授)