

# ソーシャルメディアのアーカイブ化に対する“国民の意識”

代表研究者 塩崎亮 聖学院大学 基礎総合教育部 教授

## 1 概要

ソーシャルメディア上のコンテンツは、現代の日常生活を反映したこれまでにない歴史史料と将来なりうる。しかし、そこには私的な内容のコンテンツも含まれるため、たとえ公開されているものであっても、それらを第三者が保存する行為は法的・倫理的な課題を伴う。本研究では、日本の国立国会図書館がソーシャルメディア上のコンテンツをアーカイブするという架空のシナリオを一般個人に対して提示し、その反応を確認するため、全国規模の質問票調査を実施した。結果、ブログサイトと Twitter の公開ツイートを保存するシナリオを許容しないと回答した件数は全体 (n = 1,126) の 35% を占めた。他方、すでに国内で法制度にもとづき収集されている公的機関のウェブサイトをアーカイブするという条件付きシナリオを許容するとした回答者は 44% にとどまった。また順序ロジスティック回帰分析の結果によれば、プライバシー感度の高いものほど非許容傾向を示すことも明らかとなった。さらに、非許容理由に関する内容分析の結果、非許容理由の多くが情報プライバシーに関するものであることを確認した。これらの調査結果は、懸念を緩和するには事前の同意取得やデータの匿名化が特に有効な手段となりうることを示唆している。

## 2 背景と目的

ソーシャルメディア上のコンテンツには個人／私的な情報が多々含まれており、それらは現代社会の諸相を反映した資料群ともとらえられる[1]。実際にこれまで多数の文献が、ソーシャルメディア上のコンテンツを分析対象として、たとえば COVID-19 が世界中で蔓延する中で個人の「感情」や「心情」などを記録してきた[2, 3]。しかし、そのようなコンテンツを第三者がアーカイブ（収集保存）する行為は著作権やプライバシーなどの個人の諸権利を侵害してしまう可能性がある[4]。仮に図書館などの文化遺産機関によるソーシャルメディア・アーカイブを構築する試みが制度化されたとしても、あらゆる法的・倫理的問題を事前に解消できるとはいいがたい。パーソナルなコンテンツを保存することと個人の諸権利を保護することとの間で社会的に最適なバランスを模索する必要がある[5]。

ソーシャルメディア上のコンテンツには誤情報[6]や偽情報[7]、悪意のある情報[8]も含まれているが、それらも（ある意味では）現代社会を表す歴史史料と将来なりうる。たとえばスコットランド国立図書館は、コロナ禍における誤情報やフェイクニュースの蔓延自体を後世に伝えることを目的として、プロジェクト *The Archive of Tomorrow: Health Information and Misinformation in the UK Web Archive* を 2021 年に立ち上げている[9]。とはいえ公的な文化遺産機関にとっては、そのようなコンテンツを公的支出にもとづき集め保存することを正当化する必要がある、継続的に説明責任を負わざるをえない。

これらの側面は、ソーシャルメディアを持続的にアーカイブする取り組みが、市民／国民の態度や意識にある程度依存せざるをえないことを示唆している。たとえば日本では、国立国会図書館（*National Diet Library: NDL*）が法制度にもとづき収集可能なウェブサイトは公的機関のものに限定され、私人のブログサイトなどは制度収集の対象外となっているが、このような限定的な収集方針の骨格は 1999 年に示された納本制度調査会の答申内容にまでさかのぼる[10, 11]。そこでは、ネットワーク上で公開された著作物を公権力が収集してしまうことは公開者が通常予期する範囲を超えているため、言論活動に対する「萎縮効果」を生じさせてしまう可能性もあり、「国民の意識」が変化しない限りは法制度にもとづき一律的に収集することは難しい、と議論が整理されてきた。

しかし、実際に一般個人がソーシャルメディアのコンテンツを第三者に保存されることに対して懸念を示すかどうかは、実証研究がこれまでほとんどなされておらず、必ずしも明確でない[12]。他方、ウェブアーカイブやパーソナルデジタルアーカイブに関する各種取り組みでは、個人がもつかもしれない不安をできるだけ緩和させるために具体的な措置がいくつか提唱されてきた。たとえば、事前の交渉や契約[13]、収集範囲の限定化[14]、利用の制限[15]、申し出に応じたコンテンツの非公開化[16]、データ提供者との信頼関係の構築[17]、データの匿名化[18]などである。だが、それらの緩和措置に関する実証研究もこれまでほとんど

なされておらず、それらの実質的な効果がどれくらいのものなのかは明らかでない。先行研究ではむしろ、研究倫理などの文脈で、より短期的な視座から、研究目的でソーシャルメディアのコンテンツを利用することに対する人々の反応に焦点が当てられてきた[19–23]（なお、欧米におけるソーシャルメディアのアーカイブ事例[24–29]、ソーシャルメディア・アーカイブの法的・倫理的課題[30–33]を含む先行研究のレビューについては冗長と判断し、本報告書では割愛した）。

本調査では、日本の国立国会図書館がソーシャルメディアをアーカイブすることの倫理的課題に焦点を当てる。前述のとおり、NDL ではそれらコンテンツを集められていないが、そもそも、国民の否定的な反応が示されたエビデンスまでは提示されていない。国立の図書館を対象とする理由は、それら機関が国際インターネット保存コンソーシアム（IIPC）などの国際コミュニティを先導して形成してきたことに見てとれるように[34]、この領域をもっとも代表するプレイヤーとみなせるためである[35]。さらにここでは対象コンテンツの範囲をブログと Twitter の公開ツイートに限定する。基本的には両者ともにパスワードやアクセス制限なしに広く利用できるものであるため、また、前者は分散的なもので、後者は日本で広く用いられているプラットフォームのひとつであり、それぞれ異なる特性もつメディアとみなせるためである。なおログインが必要なウェブコンテンツも技術的には収集可能だが[36]、より私的な空間に国が介入することになり、議論がより複雑化するためここでは対象外とした。

本研究の目的は、国立図書館がソーシャルメディアをアーカイブするという架空のシナリオに対する人々の反応について、比較可能なエビデンスを提示することにある。各国のアーカイブ方針は一般個人の意見に厳格に従うべきだと主張したいわけではない。むしろ、政策立案時に配慮すべき倫理的な課題を提示することを狙いとしている。この作業により、エビデンスにもとづく政策立案の道筋を描くことが可能になると考えた。日本に限定した調査ではあるが、著作権や個人データ保護に関する法制度、また国の納入制度（legal deposit）は国際的な潮流に沿って整備されてきたものといえるため、本研究の成果は他国での検討にもある程度資することが期待される。あるいは、ソーシャルメディアのデータを収集・保管する研究者や研究機関にとっても今後の方針を立てる上で役立つと思われる。

まとめると、本研究のリサーチクエスチョン（RQ）は次となる。

- ・ RQ1：国立の図書館がソーシャルメディア上のコンテンツをアーカイブするという架空のシナリオに対して人々はどのような反応を示すのか。
- ・ RQ2：人々の反応に差異があるとしたら、その差異をもたらす要因は何か。
- ・ RQ3：架空のアーカイブシナリオに拒否反応を示すものがあるとしたら、その理由は何か。

### 3 調査手法

#### 3-1 概要

先行研究[19–23]にもとづき、ソーシャルメディア上のコンテンツを日本の国立図書館がアーカイブすることに対する人々の認識を明らかにするために、代表性の高いサンプルを対象とした質問票調査を実施した。（特によく知られていない事項に関する）国民の意識あるいは民意や世論といった不可視のものを調査者が設計した質問項目にもとづき完全に再現することなど期待できないものの[37]、ここでは一定数の回答を短期間でえる目的から質問票調査の手法を採用することとした。実査は調査会社（日本リサーチセンター）に委託し、乗合形式にもとづく全国規模の個別訪問留置法による調査を行った（2021年5～6月）。対象者は日本の人口構成比と同様の性・年代構成で割り付けられた日本全国における15～79歳の男女個人1,200名である。サンプリングは住宅地図データベースにもとづき世帯抽出後に個人が割り当てられている（1地点6人×200地点で質問票を回収）。なお本研究は、著者が所属する機関における倫理委員会の承認をえて実施された。

#### 3-2 従属変数

質問票では、ソーシャルメディア上のコンテンツをアーカイブするという架空のシナリオを許容できるか否かを尋ねた。メディアの種類で回答内容に違いがあるかどうかを確認するために、ウェブサイト（シナリオA）、ブログサイト（シナリオB）、Twitterの公開ツイート（シナリオC）という3種類のシナリオを設けている。許容度の選択肢は順序尺度の従属変数として扱えるように、「分からない」といった選択肢は設けず、無条件で許容、条件付けで許容、非許容の3段階とした。

条件付き許容の場合、その懸念の理由を探るために、どのような条件であれば許容できるかを多肢選択式

で尋ねた。具体的には「事前の同意取得」「データの匿名化」「申し出に応じた非公開化」「収集範囲の限定化」「利用の制限」である。シナリオAの「収集範囲の限定化」では、公共部門のウェブサイトに収集範囲を限定する条件であることを具体的に記した。認知度は低いかもかもしれないが、NDLは2010年から公共部門のウェブサイトの制度収集を実施している。つまり、これらのウェブサイトを収集することは、潜在的な批判はあるかもしれないが、社会的には実質的に受け入れられているといえ、いわば「実質的な許容率」とみなすことも可能だろう。

### 3-3 独立変数

独立変数は先行研究に従って設定した。質問票では、ソーシャルメディアなどの閲覧経験や自身での公開／投稿経験、図書館資料の利用経験、プライバシー感度に関する項目（人口統計学的属性の項目を含む）を設けた。プライバシー感度についての質問では、ウェブ上のセキュリティ全般、履歴データ等の開示、個人データの二次利用に関する懸念の有無を尋ねた。便宜的にひとつ以上以上の項目に懸念を示した回答者はプライバシー感度が高いと判断した。本研究では、各アーカイブシナリオに対する許容／非許容の反応を分かつ要因を探るため、順序ロジスティック回帰分析を行う。なお相関分析と分散拡大係数の算出を行い、多重共線性が確認されたため、いくつかの独立変数はモデルから除外することとした。

### 3-4 非許容理由のテキストデータ

ひとつ以上のシナリオで非許容を選択した回答者には、自由記述形式でその理由を尋ねた。ここでは日本語のテキストマイニングツール（RMeCab）で非許容理由の内容分析を行った。表記揺れを整形して不要語を除外のうえ、名詞（非自立語などを除く）と副詞に限定して単語単位で頻度上位の語を確認後、頻出語のバイグラム単位での共起パターンをネットワーク図で描画した。このネットワーク図にもとづき、目視で非許容理由の観点をコーディングする作業を調査者1名により実施した。複数の観点が含まれている場合は複数のコードを付与している。また、拒否理由を記した回答者の特性を探るため、拒否理由の観点と他の変数間の対応分析も行った。その際、非許容の度合いに応じて、3件のシナリオすべてを非許容とした「強い拒否者」と、1件または2件のシナリオを非許容とした「弱い拒否者」というカテゴリカル変数を追加した。なお対応分析で用いる変数は、非許容理由の観点のほか、非許容レベル、プライバシー感度、閲覧経験、年代とした。

## 4 調査結果

### 4-1 概要

本質問票調査でえられた回答は1,200件だったが、ここでは単純に欠損データを一律除外することとし、結果、有効回答数は1,126件となった。表1は独立変数を示す。詳細については省略するが、インターネット利用者の割合などの全体的な傾向は総務省による『情報通信メディアの利用時間と情報行動に関する調査』の結果とおおむね合致しており、サンプルの妥当性はおおむね担保されている。

表1. 独立変数の概要 (n = 1,126)

区分	n	%	区分	n	%
ウェブコンテンツの閲覧経験あり <sup>a</sup>	720	63.9	インターネット	171	15.2
- 組織のウェブサイト	596	52.9	利用していない		
- 個人のブログサイト	408	36.2	利用時間/日 <sup>b</sup>		
- Twitterの公開ツイート	374	33.2	1時間以内	301	26.7
コンテンツの公開／投稿経験あり <sup>a</sup>	264	23.4	1～5時間	504	44.8
- ウェブサイト	101	9.0	5～10時間	119	10.6
- ブログ	67	6.0	10時間以上	31	2.8
- Twitterの公開ツイート	194	17.2	性別		
図書館資料の利用経験あり	94	8.3	女性	561	49.8
プライバシー感度高 <sup>a</sup>	855	75.9	男性	565	50.2
- ウェブ上のセキュリティ全般に不安	663	58.9	就業者	773	68.7
- 私的な内容がウェブ上で公開されていることに違和感	347	30.8	21大都市在住者	333	29.6
- 各種履歴データを民間企業へ提供することに違和感	309	27.4	年代 <sup>b</sup>		
- 個人データを公的機関へ提供することに違和感	293	26.0	10代	71	6.3
- 各種履歴データの二次利用に違和感	412	36.6	20代	146	13.0
			30代	181	16.1
			40代	213	18.9
			50代	179	15.9
			60代	198	17.6
			70代	138	12.3

<sup>a</sup> 下位項目のいずれかひとつ以上に当てはまる場合に該当するものと判断した。

<sup>b</sup> 順序ロジスティック回帰分析では各カテゴリの中央値を採用した。

#### 4-2 許容／非許容率 [RQ1]

表 2 では、シナリオ A・B・C に対する許容／非許容の回答件数と割合を示した。おおむね、シナリオ A の非許容率は 25%、シナリオ B と C の非許容率は 35% である。「強い拒否者」は 261 件 (23.2%)、「弱い拒否者」は 178 件 (15.8%) を示した。逆に、条件付きまたは無条件で許容のもの合計はシナリオ A で 75.5%、シナリオ B で 64.9%、シナリオ C で 64.5% となった。ただし文字どおり、条件付き許容とは「条件」に依存する (後述)。

収集対象となるメディアの種類で差があるかを検証するためにカイ二乗検定を行った。結果、メディアの種類と許容／非許容レベルとの間には連関があることが示唆された ( $\chi^2(4) = 41.479$ , Cramer's  $V = .078$ ,  $p < .001$ )。比較すると、残差分析の結果、回答者はシナリオ A (ウェブサイト) では許容する一方、シナリオ B と C では許容しない傾向が見られた。ただし、シナリオ B (ブログ) と C (公開ツイート) の間に差異は確認できなかった。

表 2. 従属変数の概要 (n = 1,126)

シナリオ		許容／非許容レベル			
		許容	条件付許容	非許容	
A	[ウェブ] 日本の国立図書館 (以下、「国立国会図書館」) は、法制度にもとづき、国内で刊行された出版物を網羅的に収集・保存するとともに、それら出版物を一般の利用者へ提供してきました。仮に、この国立国会図書館が日本国内に住所をもつ「組織・団体」から発信されたウェブサイトを集集・保存・提供することになったとしたら、あなたは許容できますか。	n [28.3]	319 [47.2]	531 [24.5]	276 [24.5]
	ASR <sup>a</sup>	(3.1)	(3.2)	(-6.4)	
B	[ブログ] 仮に、国立国会図書館が日本国内に住所をもつ「個人」から発信されたウェブサイトやブログサイトを収集・保存・提供することになったとしたら、あなたは許容できますか。	n [22.8]	257 [42.1]	474 [35.1]	395 [35.1]
	ASR <sup>a</sup>	(-2.1)	(-1.0)	(3.0)	
C	[ツイート] 仮に、国立国会図書館が日本語で記された Twitter の公開ツイートを収集・保存・提供することになったとしたら、あなたは許容できますか。	N [24.0]	270 [40.5]	456 [35.5]	400 [35.5]
	ASR <sup>a</sup>	(-1.0)	(-2.3)	(3.4)	

<sup>a</sup> ASR = adjusted standardised residuals

表 3 は、条件付き許容時に選択された条件の内訳を示している。すべてのシナリオにおいて、「事前の同意取得」と「データの匿名化」がもっとも好まれる選択肢となったが、この 2 条件だけですべての懸念が解消されるわけではないことも明確になった。「発信主体で収集範囲の限定化」の条件はシナリオ A では上位だったが、シナリオ B と C では順位が下がっている。これはおそらく、シナリオ A では公的機関が公開したものに収集範囲を制限、シナリオ B と C では特定個人、あるいは特定個人／組織が公開したものに収集範囲を制限、と異なる条件にしていたことも影響するだろう。興味深いことに、アーカイブズの文脈でよく採用されてきた方策だが、「50 年間非開示」という選択肢がもっとも低い値を示した。なお回答「その他」のほとんどは「分からない」という内容のものであった。

表 3. 条件付き許容の内訳

条件	[A] ウェブ (n = 531)		[B] ブログ (n = 474)		[C] ツイート (n = 456)	
	n	%	n	%	n	%
事前の同意取得	274	51.6	282	59.5	251	55.0
データの匿名化	236	44.4	213	44.9	213	46.7
発信主体で収集範囲の限定化 <sup>a</sup>	176	33.1	82	17.3	91	20.0
申し出に応じた非公開化	154	29.0	160	33.8	142	31.1
図書館内でのみに利用制限	127	23.9	93	19.6	87	19.1
研究者のみに利用制限	113	21.3	95	20.0	86	18.9
テーマにもとづき収集範囲の限定化	104	19.6	84	17.7	92	20.2
50 年間非開示	29	5.5	25	5.3	24	5.3
その他	8	1.5	7	1.5	4	0.9

<sup>a</sup> 「発信主体」は次のとおりシナリオに応じて異なる設定とした：[A] 公的機関、[B] 特定個人、[C] 特定個人／組織

表 3 には前述した「実質的な許容率」が含まれる。シナリオ A では、176 名の回答者が公的機関のウェブサイト収集範囲を限定する条件を選択していた。繰り返しになるが、これは NDL がすでに収集している資料群である。この値と無条件で許容すると回答した 319 名を足し合わせると有効回答総数の 44.0% (1,126

名中 495 名) を占めた。単純ではあるものの、法制度化済みで社会的にはすでに許容されているという点から、この値を「実質的な許容率」とみなすことができる。いいかえると、「実質的な許容率」を超える条件等の組み合わせであれば、正当化される可能性があるとも主張できるだろう。たとえば、「事前の同意取得」という条件が加われば、全体の許容率はシナリオ B (47.9%) と C (46.3%) においても 44% を超える。「事前の同意取得」という条件でなくても、「データの匿名化」とあわせて、「50 年間非開示」以外の条件がもうひとつ追加されればこの基準を超える。「事前の同意取得」と「データの匿名化」は抜きにして 3 つの条件ではどうかとなると、「申し出に応じた非公開化」に、「50 年間非開示」以外の条件があとふたつ追加されれば、同じく「実質的な許容率」を超える。

#### 4-3 懸念要因 [RQ2]

ソーシャルメディアのアーカイブ化に関する人々の認識の差異を分かつ要因を探るため、順序ロジスティック回帰分析を行った (表 4)。すべてのシナリオで、許容/非許容の傾向に対して閲覧経験とプライバシー感度の 2 要因が統計的に有意となった。シナリオ A については、ウェブ上のコンテンツをよく閲覧するものの方がウェブサイトのアーカイブを許容する傾向にある一方 (オッズは 0.487 で減少 [95% CI: 0.373-0.636])、プライバシー感度の高いものの方が非許容となる傾向にある (オッズは 1.399 で増加 [95% CI: 1.057-1.850])。シナリオ B についても、閲覧経験ありのオッズは 0.593 で減少 (95% CI: 0.455-0.773) となる一方、プライバシー感度の高いもののオッズは 1.530 で増加 (95% CI: 1.161-2.016) となった。同様に、シナリオ C についても、閲覧経験ありのオッズは 0.610 で減少 (95% CI: 0.469-0.794) となる一方、プライバシー感度の高いもののオッズは 1.502 で増加 (95% CI: 1.140-1.989) となった。

表 4. 順序ロジスティック回帰分析の結果

	[A] ウェブ		[B] ブログ		[C] ツイート	
	B	Exp(B)	B	Exp(B)	B	Exp(B)
インターネット利用時間/日	-0.021	0.979	-0.002	0.998	-0.016	0.984
閲覧経験あり	-0.720 ***	0.487	-0.522 ***	0.593	-0.494 ***	0.610
公開/投稿経験あり	-0.051	0.950	-0.010	0.990	0.178	1.194
図書館利用経験あり	0.080	1.083	0.226	1.253	0.286	1.331
プライバシー感度高	0.336 *	1.399	0.426 **	1.530	0.407 **	1.502
性別	-0.128	0.880	-0.130	0.878	-0.057	0.945
就業者	0.029	1.030	0.118	1.103	0.078	1.045
21 大都市在住者	-0.004	0.996	0.098	1.125	0.044	1.081
Model Fit						
-2 log likelihood	2335.609		2382.225		2401.303	
Model chi-square	43.453		24.919		22.177	
df	8		8		8	
sig.	0.000		0.002		0.005	
Nagelkerke pseudo R <sup>2</sup>	0.043		0.025		0.022	

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05

#### 4-4 非許容理由 [RQ3]

非許容理由に関する有効回答数は 282 件となった。ここでは非許容理由の観点として、「情報プライバシーの保護」「感覚的な拒否」「残す価値に対する疑念」「悪用リスクの回避」「権力の介入に対する嫌悪感」の 5 点を抽出した。表 5 は各観点の頻度を示すが、「情報プライバシーの保護」と「感覚的な拒否」が主な拒否理由であることを確認できた。そのなかでも「情報プライバシーの保護」がもっとも多く、順序ロジスティック回帰分析の結果とも整合する。なお、ひとつ以上のシナリオを非許容としたもののうち、拒否理由を回答しなかったものは 157 名であった。

表 5. 非許容理由の観点 (n = 282)

観点	n	%
情報プライバシーの保護	91	32.3
感覚的な拒否	86	30.5
残す価値に対する疑念	68	24.1
悪用リスクの回避	56	19.9
権力の介入に対する嫌悪感	28	9.9

観点「情報プライバシーの保護」は、プライバシーや個人の権利という文脈でなされたコメントであった。たとえば、「プライバシーの侵害があると思う」、「個人情報、そもそも収集すべきではない」、「個人が特定できる情報ばかり」といったものである。プライバシーの懸念だけでなく著作権を侵害する可能性について触れていた回答者は 3 名であった。本調査では、明確な記述がなく、感覚的に拒否反応を示していると考えられたものを「感覚的な拒否」とカテゴリ化した。具体的には、「理由はない」、「なんか嫌」、「興味がない」、

“今のままでいい”といったコメントがあげられる。観点「残す価値に対する疑念」は、ソーシャルメディアをアーカイブすることの価値に批判的なコメントの場合とした。たとえば、“個人の主観に基づいており”、“内容に偏向がある”、“信用できない情報が多い”、“信頼がおけるかどうか分からないものを、集める意味が分からない”、といった表現があげられる。わずかではあるが、“コストがかかる”、“税金を投入することを認められない”、“公的機関が収集保存する必要性がない”など、費用対効果についてまで言及した回答者もいた。想定していない悪用／誤用に対する不安が表明されたものは、観点「悪用リスクの回避」のコードを付与した。たとえば、“情報の漏洩がありうる”、“犯罪が多発する恐れ”、“悪用されるのではないかと不安がある”といった表現があげられる。ある回答者は“トラブルに巻き込まれることが嫌”ともコメントしていた。「権力の介入に対する嫌悪感」の観点は、国の機関が介入することに対する忌避感が示されたものである。たとえば、“思想の管理に使われたくない”、“自分の Twitter 等を国に管理されたくない!”、“個人の自由が侵されることが出てくるのでは？と危惧”といった表現が確認できた。表現の自由の萎縮効果の可能性について明示的に言及した回答者は1名にとどまり、“Twitter は自由に発言できる場所なので、国がからんでくるとなんとなく自由に発言できない感じがする”と表現されていた。

拒否理由の特性を探るために対応分析も実施した。累積寄与率は低いものの、ここでは解釈可能性の観点から、二次元上にプロットした結果にもとづき考察した。次元1の寄与率は15.6%、次元2の寄与率は10.2%である。次元1で高い寄与率を示した変数は「70代」(30.7%)、「感覚的な拒否」(23.4%)、「強い拒否者」(17.0%)と「弱い拒否者」(11.0%)となった。この次元は非許容の度合いを反映しているように思われる。70代の回答者は特別な理由もなく強い拒否者である傾向にある。ひるがえって、より具体的な理由にもとづき非許容の反応を示したものは、ウェブのコンテンツに親しんでいる一方でプライバシー感度が高い層であることがうかがえる。次元2で高い寄与率を示した変数は、「30代」(27.5%)、「悪用リスクの回避」(27.3%)、「残す価値に対する疑念」(16.2%)となった。この次元は、収集と提供の2側面を反映しているように思われる。20代と50～60代は残す価値に疑問をもち、権力の介入に嫌悪感を示すものが多い傾向にあるが、これは集められることに対する懸念を表しているのかもしれない。一方、30代はリスク回避傾向にあることが示唆されたが、これはむしろ、事後に提供されることに対する不安の方が強いことを表しているのかもしれない。

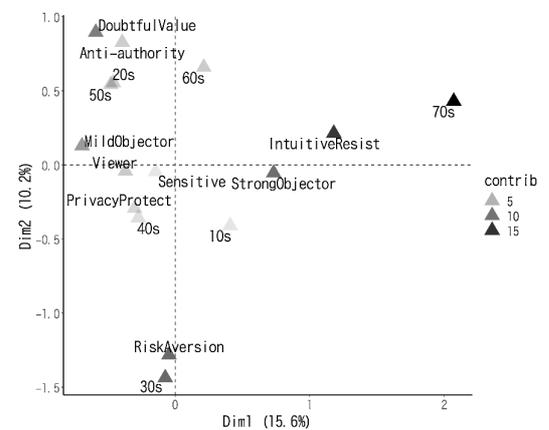


図1. 対応分析の結果

## 5 考察

### 5-1 実質的な許容率の基準

本研究の目的は、国立の図書館がソーシャルメディアのコンテンツを保存すべきか否かという二択の問いに答えることではない。実際に本調査の結果は、無条件で許容するものもいれば、どのような条件でも許容しないものもいる中間に、何らかの条件付きであれば許容できると回答したものが一定程度いることを示すものだった。おおまかに、許容および条件付き許容の割合を足し合わせると60～70%を示したが、この値をもってして、NDLによるソーシャルメディアのアーカイブ構築は正当化されると主張するものはいらぬかもしれない。逆に、20～30%の非許容層を無視することはできないと主張するものもいるかもしれない。とはいえ、いずれにせよ、本調査では「実質的な許容率」が44%であることを確認した。ナイーブな基準であることは否定できないが、それでも今後議論していくうえでの参照点とはなるだろう。この参照点に従えば、どのシナリオの非許容率も改善可能なレベルにあるように見える。

### 5-2 プライバシーに対する懸念

プライバシーがこのアーカイブ問題の中心に位置づけられることになるだろう。人々は、ソーシャルメディア上のコンテンツよりも、組織や団体のウェブサイトをアーカイブするシナリオであれば受け入れやすい傾向にあった。シナリオB（ブログ）とC（ツイート）の間では差異が確認できなかったとはいえ、全体と

しては、プライバシーの懸念が生じやすいシナリオの方が許容されにくい傾向にあることが示唆されていた。また本調査では、人々が受け入れられやすくなる条件についても確認したが、「事前の同意取得」「データの匿名化」「申し出に応じて非公開化」といった方策がより選択されていた。このような傾向は、人々がいわば自身のアイデンティティをコントロールする権利の重要性を感じていることの表れかもしれない。順序ロジスティック回帰分析の結果からすると非許容の態度をもたらす決定的な要因までは特定できなかったが、相対的には、プライバシー感度と閲覧経験が回答者の意識を分かち要因となっていることが確認できた。繰り返しになるが、プライバシー感度の高い回答者は許容しにくく、閲覧経験があるものは逆に許容しやすい関係にある。利用経験のあるユーザはウェブ上に存在する、つまりはアーカイブ対象となるコンテンツの内容をよく理解しているが、非ユーザはよく分からないままに回答せざるをえなかったことの反映かもしれない。利用経験のあるユーザはまた、ウェブ上で公開されたコンテンツが、基本的には、許諾なしで誰もが技術的には収集できてしまうものであることを理解していたともいえるだろう。

とはいえ、非許容理由の観点でもっとも多かったプライバシー保護に対する懸念は、他の非許容理由の遠因となっている可能性も否定できない。感覚的な拒否は、プライバシー概念の言及が明示的にはないにせよ、私的な情報空間への介入を拒む姿勢の表れともとらえられる。リスク回避者は、仮に個人に関する情報が漏えいした場合、予期せぬ形で自身のアイデンティティが脅かされてしまうと感じたことの表れかもしれない。権力の介入に拒否反応を示したものは、個人の情報空間に侵入する「主体」こそを問題視したともいえる。

### 5-3 実務面の示唆

本研究からは実務上の示唆もえられる。調査結果にもとづけば、「実質的な許容基準」を参照しながら、プライバシーの懸念を緩和するためのアーカイブ戦略を策定するのが望ましい、とまとめられる。ただし、この基準はあくまで暫定的な参照点に過ぎない。

プライバシーに対する懸念を緩和する方策として第一にあげられるのは、事前に同意を得ることだろう。とはいえ、たとえば Twitter 上の公開ツイートとブログでは同意を得るやり方は異ならざるをえないけれども、いずれにせよ、すべてのユーザから（あるいは場合によっては非ユーザからも）同意を得るというのは現実的でない。「すべてにインフォームドコンセントを」というよりも「ほとんどの事柄について何らかの同意を」という方針をとらざるをえないのではないだろうか [38]。その際、個人の諸権利を規制する法制度が導入されれば、各権利者から同意を得ることにかかる取引コストを大幅に減らすことが可能である。しかしながら、そのような法制度が仮に社会実装されたとしても、倫理的な観点からすると、アドボカシー活動などの別の方策も併用して立てていくのが望ましいだろう。

事前に同意を得るアプローチとは異なり、データを匿名化する、あるいは申し出に応じてコンテンツを非公開化する方策は事後的なアプローチとなる。後者は（削除対象となる分量にも依存するが）比較的に実現しやすいだろうが、未検出の違法コンテンツなどが含まれているかもしれない。前者の匿名化もまた骨の折れる作業である。大量のアーカイブコンテンツを対象とした機微データの自動検知技術の開発が欠かせない [39]。もちろん完全にリスクを解消することは難しいだろうが、しかしそれでも、本調査結果からすると、データの匿名化は有効な方策となりうる。

「実質的な許容基準」にまで不確かなリスクを下げるうえで、利用制限も有効な選択肢のひとつとなりうる。施設内での利用に限定する、利用者を研究者に限定する、一定期間は非公開とするといった方策は、これまでも図書館やアーカイブズ領域でとられてきた。しかし本調査結果にもとづけば、これらの方策は「収集範囲の限定化」よりも効果が見られなかった。興味深いことに、「50年間非開示」がもっとも人気のない方策だったのは先に触れたとおりである。このような傾向はおそらく、ほとんどの回答者が二次利用者とならないことに起因するだろう。ソーシャルメディアのアーカイブは、現代世代の二次利用者でないものには直接的な便益をもたらさないためと考えられる。

最後に、対応分析の結果にもとづく混合戦略についてまとめる。図1では、あくまで相対的にはあるが、70代の回答者が特に理由もなく許容できないと回答する傾向にあったことを示していた。この傾向は、年代が上がるにつれてソーシャルメディアの利用経験が少なくなる傾向と重なる。このような層に対しては、ソーシャルメディアを保存することの意義を訴え、その価値を提示する活動が特に有効となるだろう。対して、プライバシーの感度が高く、「弱い拒否者」であるのはウェブに慣れ親しんだ世代が多い。このようなタイプはおおきく二つに分けられる。ひとつは誤用リスクを懸念するもの、もうひとつは費用対効果や国の介入に疑念を抱くものである。前者に対しては、利用を制限することも実質的に有効となるだろうが、情報セキュ

リティ体制の向上を推進する方針がより説得力を増すと思われる。後者に対しては、アドボカシー活動や信頼関係の構築が欠かせない。

## 6 制約と今後の課題

本研究にはおおきく2点の制約がある。ひとつは内的妥当性で、質問票調査の結果が、人々がもつ認識をどこまで反映できているかという問題である。より深く分析するには定性的な調査が欠かせない。また、回答者はソーシャルメディアをアーカイブすることに関してよく分からないままに回答していただけかもしれない。たとえば仮に、他国で同様の取り組みがすでになされていることを知らされたら、その判断は変わったかもしれない。今後の課題として、第三者である国立図書館等の文化機関がソーシャルメディア上のコンテンツを収集保存する行為に対する拒否感が、一般個人においてどのくらい強固なものなのかを探る必要がある。具体的には、ソーシャルメディア上の公開コンテンツをアーカイブする価値に対する認識が高まればその拒否感は緩和されるのかを実験的手法にもとづき検証する予定である。

もう一つの制約は外的妥当性で、調査対象があくまで日本に限定されたことに起因するものだ。本研究の結果が他国にも当てはまるとはいえない可能性がある。「実質的な許容率」の妥当性、非許容の理由、想定されるアーカイブ戦略の妥当性について検証しなければならない。この点については、すでにソーシャルメディアのコレクションを構築しだしている国立図書館がいくつかあるため、それら先行国/地域との比較調査を実施する予定である。

### 【参考文献】

- [1] Milligan I. History in the age of abundance? How the web is transforming historical research. Montreal: McGill-Queen's University Press, 2019.
- [2] Medford RJ, Saleh SN, Sumarsono A, et al. An 'infodemic': leveraging high-volume Twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infect Dis* 2020; 7(7): ofaa258. DOI: 10.1093/ofid/ofaa258.
- [3] Tsao SF, Chen H, Tisseverasinghe T, et al. What social media told us in the time of COVID-19: a scoping review. *Lancet Digit Health* 2021; 3(3): E175–E194. DOI: 10.1016/S2589-7500(20)30315-0.
- [4] Thomson SD. Preserving social media. Digital Preservation Coalition, 2016. DOI: 10.7207/twr16-01.
- [5] Murray A. Part VI Data privacy. In: Information technology law: the law & society. 4th ed. Oxford: Oxford University Press, 2019, pp. 563–675.
- [6] Brennen JS, Simon FM and Nielsen RK. Beyond (mis)representation: visuals in COVID-19 misinformation. *Int J Press/Polit* 2021; 26(1): 277–299. DOI: 10.1177/19401612200964780.
- [7] Allcott H and Gentzkow M. Social media and fake news in the 2016 election. *J Econ Perspect* 2017; 31(2): 211–236. DOI: 10.1257/jep.31.2.211.
- [8] Adewole KS, Anuar NB, Kamsin A, et al. Malicious accounts: dark of the social networks. *J Netw Comput Appl* 2017; 79: 41–67. DOI: 10.1016/j.jnca.2016.11.030.
- [9] National Library of Scotland. Funding secured to explore misinformation around the pandemic, [www.nls.uk/news/media/health-websites-project/](http://www.nls.uk/news/media/health-websites-project/) (2021, accessed 6 November 2021)
- [10] 納本制度調査会『(答申)21世紀を展望した我が国の納本制度の在り方:電子出版物を中心に』国立国会図書館, 1999, [dl.ndl.go.jp/info:ndljp/pid/1001007](http://dl.ndl.go.jp/info:ndljp/pid/1001007) (accessed 6 November 2021)
- [11] 納本制度審議会『(答申)ネットワーク系電子出版物の収集に関する制度の在り方について』国立国会図書館, 2004, [dl.ndl.go.jp/info:ndljp/pid/999243](http://dl.ndl.go.jp/info:ndljp/pid/999243) (accessed 6 November 2021)
- [12] Velte A. Ethical challenges and current practices in activist social media archives. *Am. Arch.* 2018; 81(1): 112–134. DOI: 10.17723/0360-9081-81.1.112.
- [13] Bailey S and Thompson D. UKWAC: building the UK's first public web archive. *D-Lib Mag* 2006; 12(1). DOI: 10.1045/january2006-thompson.

- [14] Cunnea P. Selective web archiving in the UK: a perspective of the National Library of Scotland within UK Web Archiving Consortium (UKWAC). *SCONUL Focus* 2005; 34: 44-49.
- [15] Aubry S. Introducing web archives as a new library service: the experience of the National Library of France. *LIBER Q* 2010; 20(2): 179-199. DOI: 10.18352/lq.7987.
- [16] Black S. The implications of digital collection takedown requests on archival appraisal. *Arch Sci* 2020; 20(1): 91-101. DOI: 10.1007/s10502-019-09322-y.
- [17] Redwine G, Barnard M, Donovan K, et al. Born digital: guidance for donors, dealers, and archival repositories. Council on Library and Information Resources 2013, [www.clir.org/pubs/reports/pub159/](http://www.clir.org/pubs/reports/pub159/) (accessed 6 November 2021)
- [18] Henttonen P. Privacy as an archival problem and a solution. *Arch Sci* 2017; 17(3): 285-303. DOI: 10.1007/s10502-017-9277-0.
- [19] Fiesler C and Proferes N. 'Participant' perceptions of Twitter research ethics. *Soc Media Soc* 2018; 4(1): 1-14. DOI: 10.1177/2056305118763366.
- [20] Williams ML, Burnap P and Sloan L. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. *Sociology* 2017; 51(6): 1149-1168. DOI: 10.1177/0038038517708140.
- [21] Williams ML, Burnap P, Sloan L, et al. Users' views of ethics in social media research: informed consent, anonymity, and harm. In: Woodfield K (ed) *The ethics of online research*. Bingley: Emerald Publishing, 2017, pp. 27-52.
- [22] Marshall CC and Shipman F. On the institutional archiving of social media. *Proc 12th ACM/IEEE Jt Conf Digit Libr* 2012: 1-10. DOI: 10.1145/2232817.2232819.
- [23] Marshall CC and Shipman F. Who owns the social web? *Commun ACM* 2017; 60(5), 52-61. DOI: 10.1145/2996181.
- [24] Burnap P, Rana O, Williams M, et al. COSMOS: towards an integrated and scalable service for analysing social media on demand. *Int J Parallel Emergent Distrib Syst* 2015; 30(2): 80-100. DOI: 10.1080/17445760.2014.902057.
- [25] Galarza A. Documenting the Now. *J Am Hist* 2018; 105(3): 792-793. DOI: 10.1093/jahist/jay444.
- [26] Winters J. Giving with one click, taking with the other: electronic legal deposit, web archives and researcher access. In: Gooding P and Terras M (eds) *Electronic legal deposit: shaping the library collections of the future*. London: Facet Publishing, 2020, pp.159-178.
- [27] Laursen D and Møldrup-Dalum P. Looking back, looking forward: 10 years of development to collect, preserve, and access the Danish web. In: Brügger N. (ed) *Web 25: histories from the first 25 years of the World Wide Web*. NY: Peter Lang, 2017, pp.207-227.
- [28] Zimmer M. The Twitter archive at the Library of Congress: challenges for information practice and information policy. *First Monday* 2015; 20(7). DOI: 10.5210/fm.v20i7.5619.
- [29] Pehlivan Z, Thièvre J and Drugeon T. Archiving social media: the case of Twitter. In: Gomes D, Demidova E, Winters J, et al. (eds), *The past web*. Cham, Switzerland: Springer, 2021, pp.43-56. DOI: 10.1007/978-3-030-63291-5\_5.
- [30] Chambers S, Birkholz J, Geeraert F, et al. BESOCIAL final report Work Package 1: an international review of social media archiving initiatives. 2021. [orfeo.kbr.be/handle/internal/7741](http://orfeo.kbr.be/handle/internal/7741) (accessed 6 November 2021)
- [31] Vlassenroot E, Chambers S, Di Pretoro E, et al. Web archives as a data resource for digital scholars. *Int J Digit Humanities* 2019; 1: 85-111. DOI: 10.1007/s42803-019-00007-7.
- [32] Vlassenroot E, Chambers S, Lieber S, et al. Web-archiving and social media: an exploratory analysis. *Int J Digit Humanities* 2021. DOI: 10.1007/s42803-021-00036-1.
- [33] Bingham NJ and Byrne H. Archival strategies for contemporary collecting in a world of big data: challenges and opportunities with curating the UK web archive. *Big Data Soc* 2021; 8(1). DOI: 10.1177/2053951721990409.

- [34] International Internet Preservation Consortium. About IIPC, [netpreserve.org/about-us/](http://netpreserve.org/about-us/) (accessed 6 November 2021)
- [35] Masanès J. Web archiving. Berlin: Springer, 2006. DOI: 10.1007/978-3-540-46332-0.
- [36] Ximm A. Active personal archiving and the Internet Archive. In: Hawkins DT (ed) Personal archiving: preserving our digital heritage. Medford, NJ: Information Today, 2013, pp.187–213.
- [37] ブルデュー P. 「世論なんてない」『社会学の社会学』田原音和監訳, 藤原書店, 1991, pp.287-302.
- [38] サルガニック M. 『ビット・バイ・ビット: デジタル社会調査入門』瀧川裕貴ほか訳, 有斐閣, 2019, p.315.
- [39] Mcdonald G, Macdonald C and Ounis I. How the accuracy and confidence of sensitivity classification affects digital sensitivity review. ACM Trans Inf Syst 2021, 39(1), 1–34. DOI: 10.1145/3417334.

### 〈発表資料〉

題名	掲載誌・学会名等	発表年月
People's perceptions on social media archiving by the national library of Japan	Journal of Information Science	2022年 (in press)
関連する業績は次のとおり		
A note on law and economics of legal deposit systems	Libri	2022年 (in press)
Future uncertainties for preserving tweets: Peoples' perceptions in Japan	Journal of Librarianship and Information Science	2021年
公開ツイートを第三者がアーカイブすることに対する個人の意識：質問票調査の集計結果	図書館界	2021年