諸外国におけるウェブサイトの収集をめぐる状況

	日本	米国	英国		ドイツ	フランス	カナダ	オーストラリア	韓国	【参考】
	国立国会図書館 （National Diet Library : NDL）	議会図書館 （Library of Congress : LC）	英国図書館 （British Library : BL）	英国公文書館 （The National Archives : TNA）	ドイツ国立図書館 （Deutsche Nationalbibliothek : DNB）	フランス国立図書館 （Bibliothèque nationale de France : BnF）	カナダ国立図書館・文書館 （Library and Archives Canada : LAC）	オーストラリア国立図書館 （National Library of Australia : NLA）	韓国国立中央図書館 （National Library of Korea : NLK）	インターネットアーカイブ （Internet Archive : IA）
サービス名	インターネット資料収集保存事業 （WARP）	Library of Congress Web Archives	UK Web Archive （UKWA）	UK Government Web Archive （UKGWA）	Webarchiv der Deutschen Nationalbibliothek	Archives de l'internet	Government of Canada Web Archive （GCWA）	Australian Web Archive （AWA）	Online Archiving & Searching Internet Sources (OASIS)	Wayback Machine
URL	https://warp.ndl.go.jp/	https://www.loc.gov/web-archives/	https://www.webarchive.org.uk/	https://www.nationalarchives.gov.uk/webarchive/	https://webarchiv.dnb.de/	https://www.bnf.fr/fr/consulter-les-archives-de-linternet	https://webarchiveweb.bac-lac.canada.ca/	https://webarchive.nla.gov.au/collection	https://nl.go.kr/oasis/	https://web.archive.org/
開始時期	2002年（公的機関の網羅的収集は2010年-）	2000年	2004年	2003年	2012年	2002年	2005年	1996年（AWAの前身のPANDORAの開始時期、AWAは2019年-）	2004年	1996年
収集範囲・方法・頻度等	・公的機関は網羅的収集（年4-12回） ・民間は選択的収集（年1-4回）	・網羅的収集：立法府及び関連機関、連邦議会・委員会、国政選挙キャンペーンサイト、司法府 ・選択的収集：行政府、特定のテーマ、地域、イベント等（国外のサイトを含む）	・網羅的収集：.uk、.scot、.wales、.cymru、.londonドメイン、IPアドレスが英国内等英国ベースのウェブサイト（年1回） ・選択的収集：特定のテーマ、イベント等（年1回以上）	・網羅的収集：英国政府関連サイト（政府、GOV.UK、政府外公共機関、王立委員会、NHS） ・選択的収集：政府関連SNS・動画	・網羅的収集：選択的収集を補完するため、.deドメインを対象としたバルク収集を何年か1回の頻度で実施する方針を示しているが、詳細不明。これまでに少なくとも1回（2014年）実施。 ・選択的収集：連邦政府・州政府等（年2回（ただし、サイトによる））	・網羅的収集：.frドメイン、フランス人やフランス国内で作成されたウェブサイト（年1回）。ただし、全サイトの収集には至っていない（2022年時点で約60%）。 ・選択的収集：特定のテーマ・地域・イベント（選挙、オリンピック、COVID-19等）、ニュース・新聞、オーディオビジュアル（ポッドキャストやYoutube）、SNS等のコレクション（年1回以上）	・網羅的収集：全.caドメインサイト、連邦政府機関 ・選択的収集：記念行事・国家的に重要なイベント	・網羅的収集：.auドメインのウェブサイト（最低年1回）、連邦政府（年複数回） ・選択的収集：国家的重要性の高いテーマ・イベントに関するサイト、ニュースサイト等	・網羅的収集：韓国インターネット振興院（KISA）に登録されている韓国ドメインのサイト（年1回） ・選択的収集：国家的災難（新型コロナ等）、主要な国家イベント等の特定のトピックや問題に関するサイト、保存する価値がある個々のウェブサイト	公開されている世界中のウェブページを網羅的に収集（アクセスにパスワードが必要なウェブページ、ウェブサイトの所有者が登録を拒否したサイト等を除く）
収集根拠	・公的機関：国立国会図書館法第25条の3 ・民間：許諾	・政府関連、クリエイティブ・コモンズ・ライセンスが適用されたサイト：米国著作権法 ・上記以外：許諾（サイト所有者に対し最低1回通知）	・網羅的収集：2003年法定納本図書館法（2013年改正による）、2013年法定納本図書館（非印刷体）規則 ・選択的収集：許諾	・1958年公記録法第3条 ・1988年著作権・意匠・特許法第163条（国王の著作権（Crown copyright））	・ドイツ国立図書館法第3条、第16a条(1) ・著作権法第60e条(1)	・文化遺産法典L132-2ほか ・文化遺産法典R132-23ほか	・カナダ国立図書館・文書館法第8条(2)	・1968年著作権法第195CC条（2016年改正による） ・AGIMO Circular 2010/001（2010年通達）	・図書館法第22条	・米国著作権法第107条（フェアユース規定）
収集規模	・1万4,466タイトル ・合計3,645テラバイト （2025年3月）	・約1万4,600タイトル（※起点URL） ・合計3,700テラバイト （2023年8月）	・合計1,400テラバイト超（2023年）	・約6,400タイトル（2021年11月）	・9,945タイトル（2025年10月） ・選択的収集分73テラバイト（2025年4月）、2014年ドメインクロール分139テラバイト	・約588億ページ（URL）（2024年） ・2,300テラバイト超（2024年）	・130テラバイト超（国別ドメイン分440テラバイト超を除く）（2024年）	・約156億ファイル ・合計880テラバイト超 （2022年）	・合計1,122,106サイト（2024年） ・約1,000テラバイト（2022年）	・9,160億超のウェブページ（2025年10月）
選択的収集の対象	・私立大学、公益法人、学協会、第三セクター、業界団体、文化施設、政党、企業、イベント、震災、新型コロナ関連等	・イベントやテーマ（9.11、Fortune誌500掲載企業、新型コロナ等）	・テーマ（選挙、議会、企業、政治・社会的トピック、スポーツイベント、科学、文学、民族、ブログ、ニュースサイト、ウェブコミック等）	・政府関連SNS・動画	・機関単位（連邦政府・州政府、利益団体、文化・スポーツ団体、宗教団体、政党・政治家、科学・研究機関、福祉団体等） ・イベント単位（連邦議会選挙、2014FIFAワールドカップ、2017宗教改革500周年等）	・特定のテーマ・地域・イベント（選挙、オリンピック、COVID-19等）、ニュース・新聞、オーディオビジュアル（ポッドキャストやYoutube）、SNS等のコレクション ・選択的収集のタイトル等は、API et Jeux de données（BnF）やdata.gouv.frで公開。	・テーマ別コレクション ・記念行事・国家的に重要なイベント（2010バンクーバーオリピック、COVID-19等） ・首相、大臣等のSNS投稿 ・消滅が危惧されるウェブサイト	・選挙運動、森林火災、新型コロナ、記念行事等の関連サイト、ニュースサイト等	・国家的災難（新型コロナ等）や主要な国家イベントに関するサイト、主要機関（政府・公的機関、報道機関、企業、文化芸術関連機関等）のサイト等	—
提供方法	・全タイトルのうち8割を許諾を得てインターネット公開（公的機関も含む） ・上記以外は館内提供 ・URL、キーワード等で検索。全文検索可。	・許諾を得たものはインターネット公開（2万6,400件）、その他は館内提供（6,500件） ・URL、キーワードで検索。全文検索不可。 ・1年間のエンバーゴあり。	・原則法定納本図書館（※）内提供。許諾を得たもののみインターネット公開（約25%）。ほかにOpen Government Licenseが適用されるものもインターネット公開。 ・URL、キーワードで検索。全文検索可（一部のみ）。	・インターネット公開（国王の著作権の対象となるものは、Open Government Licenseに基づき二次利用可） ・URL、キーワード等で検索。全文検索可。	・原則館内提供。許諾を得たもののみインターネット公開（286タイトル） ・タイトル等は書誌カタログで検索可。全文検索は館内のみ可。	・研究パスを持つ者に限り、館内で専用アプリにより利用可。なお、法定納本連携センター（Pôles associés de Dépôt légal）のうち21館でも利用可。URLで検索。全文検索は一部のコレクションのみ可。	・原則インターネット公開 ・URL、キーワードで検索可。全文検索可。	・Trove（情報探索サービス）を通じて、インターネット公開 ・URL、キーワードで検索。全文検索可。	・原則として館内提供。許諾を得たものはインターネット公開。権利者の要請等による非公開も可。 ・URL、キーワード等で検索。全文検索可。	・インターネット公開 ・URLで検索。全文検索不可。
SNS・ニュースサイト収集	・アメーバブログ48タイトル（2019年-） ・ニュースサイト2タイトル（「共同通信社：東日本大震災（ニュース特集）」（2014-2015年）、「Slow News」（2022年-））	・Twitterの全テキストデータを同社からのデータ寄贈により収集（2006-2017年、2018年以降は選択的収集）。未公開。 ・ウェブアーカイブにFacebookページ等も含まれる。ただし、Facebook、Instagram、ポッドキャストやYoutube等のストリーミングメディアは現状では技術的に収集・公開が困難。 ・米国ニュースサイトコレクション（2014年-、Washington Post・New York Post・Breitbart等105タイトル。主に週1回） ・経済ブログコレクション（2018年-、経済学者・連邦準備制度理事会等82タイトル。月1回）	・政治家、ジャーナリスト、企業家、選挙等に関するSNS収集（Twitter、Facebook、Instagram。YoutubeやTikTok等のストリーミングメディアは対象外） ・英国オンラインニュースサイトコレクション（Guardian、BBC等の新聞社・放送局、地域新聞等。毎日又は週1回の頻度。Twitter、Facebook等も収集） ・選択的ブログ収集（2005年-。ニュース、政治批評、スポーツ、音楽、個人の日記等）	・政府関連のYouTube、Twitter、Flickr、Instagramを収集（2014年-。1,297アカウントを公開（2025年5月））。Instagramは未公開。Twitterは2023年のXへの変更後、技術的問題により収集を中断。	・SNSについて、体系的な収集は行っていないが、Twitter買収後の2023年2-4月に、580万アカウントの約2億2,000万件（2006年3月-2011年5月）のドイツ語ツイートを収集。640GB分。現在は、Digital Humanitiesの研究プロジェクトに提供可能としている。 ・高頻度ではないが、ニュースサイトや新聞のウェブサイトも収集されている。2013年にフィナンシャルタイムズ・ドイツが廃刊になったときには、ウェブサイト全体が収集された。	・SNSについて、Facebook（2007-2020年）とTwitter（2017-2023年）は、技術的課題（仕様変更）により収集できなくなるまでは毎日収集。 Instagramは2020年以降、TikTokは2022年以降、ごく一部を年数回収集。 ・ニュースサイトや新聞のウェブサイトもコレクションとして収集・保存（Le Monde、Libération等。毎日）。	・SNSについて、首相、大臣等のアカウントのほか、テーマ別コレクションではハッシュタグ付きの投稿の収集などを行っている。なお、一般個人のアカウントを対象とした収集は行っていない。 ・ニュースサイトや新聞のウェブサイトも収集されているが、CBC（ほぼ毎日）を除き、必ずしも高頻度ではない。	・SNS等のソーシャルメディアの収集は一部のみ ・Twitter accountsコレクション（政府機関や政治家、ニュースサイト等1,649アカウントを収集） ・ABC News、9News等のニュースサイトを定期的に収集	・中央省庁、地方自治体、公的機関のブログやTwitter等の主要なソーシャルメディアも選択的に収集 ・Twitterコレクションを構築（2018年-） ・報道機関（韓国の放送局や新聞社等）のウェブサイトを選択的に収集（月1回）	・SNS、ニュースサイトも収集
その他	—	・指定サイトの自動収集や関連調査を英MirrorWeb社に委託（2021年-） ※タイトル＝seed（起点URL）、件＝提供単位（コレクション等を含む）	・「Save a website」フォームで収集すべき英国サイトを一般公募 ・BLに対する2023年10月のサイバー攻撃により公開停止中（停止中も収集は継続し、Amazon Web Servicesのクラウドに保存） ※法定納本図書館は、英国図書館、ボドリアン図書館（オックスフォード大学）、ケンブリッジ大学図書館、トリニティ・カレッジ図書館（ダブリン）、スコットランド国立図書館、ウェールズ国立図書館の6つ。	・収集開始時にInternet Archiveが1996年から収集開始前までのアーカイブを寄贈 ・ウェブアーカイブシステムの開発・運用を英MirrorWeb社に委託（2017年-。2017年以前は蘭NPOのInternet Memory Foundation）。	・収集開始以降、システム開発を担当した独oia GmbH社に委託。ただし、2014年のドメインクロールはInternet Memory Foundationに委託。	・2006年以降、テレビ局・ラジオ局のウェブサイトは国立視聴覚研究所（INA）が収集 ・収集開始後5年間はInternet Archiveと協力。その後は全ての収集を内部で実施。 ・収集すべきフランスのウェブサイトの提案も受け付けている。	・Internet Archiveに委託。1996年から収集開始前までの国別ドメイン（.ca）のアーカイブをIAから購入・取得。	・PANDORA（1996年-、連邦政府の刊行物や選挙等テーマ・イベントの選択的収集）、Australian Government Web Archive（AGWA、2011年-、連邦政府のサイトの網羅的収集）、Australian Domain Harvest collection（2005年-、.auドメインの網羅的収集。AWAの80%以上を占める。）の3つのウェブアーカイブを統合して2019年にAWAを構築。 ・Internet Archiveに委託。IAが1996年からAWA開始前までの.auドメインのアーカイブを寄贈。	—	—
参考URL	—	https://www.loc.gov/programs/web-archiving/about-this-program/ https://www.loc.gov/preservation/outreach/preweek/preswekk2023/webarchivingslides.pdf?loclr=blogsig https://blogs.loc.gov/thesignal/2023/08/the-web-archiving-team-answers-questions-about-the-web-archives/ https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/ https://www.loc.gov/acq/devpol/webarchive.pdf https://www.loc.gov/acq/devpol/socialmedia.pdf https://www.loc.gov/collections/united-states-news-web-archive/about-this-collection/ https://www.loc.gov/collections/economics-blogs-web-archive/about-this-collection/ https://www.digitalpreservation.gov/meetings/DSA2023/loc_dsa2023_website_0124_stringer_MirrorWeb-WAStorageChallenges-DSA2023_KS-GB_Mar23.pdf	https://web.archive.org/web/20230911181851/https://www.webarchive.org.uk/en/ukwa/info/about/ https://web.archive.org/web/20241113194514/https://blogs.bl.uk/webarchive/2023/04/uk-web-archive-technical-update-spring-2023.html https://blogs.bl.uk/webarchive/2023/09/what-can-you-discover-and-access-in-the-uk-web-archive-collection-.html https://web.archive.org/web/20230911181901/https://www.webarchive.org.uk/en/ukwa/collection/138 https://blogs.bl.uk/webarchive/2024/10/archiving-social-media-with-browsertrix.html https://www.bl.uk/stories/blogs/posts/exploring-alternative-access-making-the-most-of-web-archives-during-uk-web-archive-downtime	https://www.nationalarchives.gov.uk/webarchive/about/ https://www.nationalarchives.gov.uk/webarchive/archive-a-website/ https://blog.nationalarchives.gov.uk/preserving-uk-government-history-for-25-years/ https://www.nationalarchives.gov.uk/webarchive/find-a-website/ https://www.nationalarchives.gov.uk/webarchive/find-archived-uk-government-social-media-channels-in-the-uk-government-web-archive/	https://www.dnb.de/webarchiv https://www.dnb.de/twitterarchiv https://www.dnb.de/EN/Service/Hilfe/Webarchiv/hilfeWebarchiv_node.html https://www.dnb.de/SharedDocs/Downloads/EN/Ueber-uns/zumSammelauftragEN.pdf?__blob=publicationFile&v=2 https://digital.library.unt.edu/ark:/67531/metadc2472445/m1/1/	https://www.bnf.fr/en/understanding-web-legal-deposit https://www.bnf.fr/en/discover-bnfs-web-archive-collections https://netpreserve.org/ga2024/web-archiving-in-france/	https://www.canada.ca/en/library-archives/services/government/web-social-media-preservation-program.html https://www.canada.ca/en/library-archives/corporate/about-us/policies/collection-development-published-heritage.html https://www.governmentinformationday.ca/uploads/1/1/2/0/112034557/gid_smyth_gcwa_dec2023.pdf	https://www.library.gov.au/discover/what-we-collect/australian-web-archive https://www.library.gov.au/sites/default/files/documents/2024-10/collection-development-policy-20200702.pdf https://web.archive.org.au/awa/20240612173623mp_/https://www.nla.gov.au/sites/default/files/2022-10/NLA%20AR%202021-22%20Web.pdf	https://nl.go.kr/oasis/main/pop/english_main_pop.do https://oak.go.kr/nl-ir/bitstream/2020.oak/739/1/국립중앙도서관 온라인 자료 납본 수집 지침.pdf https://www.nl.go.kr/afile/fileDownloadById/23033033587Df0xs	https://help.archive.org/help/wayback-machine-general-information/