

諸外国の国立図書館等におけるネットワーク系出版物収集への対応状況

平成19年7月27日
国立国会図書館

国	主導的機関	名称	収集方法	規模	開始年	IIPCへの参加	法制化	特記事項
アメリカ	インターネット・アーカイブ(NPO団体)	—	バルク	2PB	1996年	○	—	<ul style="list-style-type: none"> 1996年以降の全世界のWeb情報550億ページを包括的に収集し(=バルク収集)、2001年からネットで公開。 2003年に11の国立図書館とコンソーシアム(IIPC)を締結し、活動継続中。(2007年5月29日現在のIIPCの参加機関数は24) ウェブアーカイブ構築ツールキット開発の中心メンバーとして、高性能収集ロボット(Heritrix)と提供システム(WERA、Open Source Wayback)等の開発を実施、継続中。この開発成果は適宜オープンソースソフトウェアとして公開され、IIPCの世界的なウェブアーカイブの推進活動、特に技術的側面において貢献している。 2007年6月から収集ロボットHeritrixによる世界のウェブサイト200億ページを収集するプロジェクト(「around the World in 2Billion Pages」)を開始。2007年12月には収集結果がインターネット上で公開される予定。
アメリカ	米国議会図書館	MINERVA	選択的	約37,000サイト	2000年夏	○	—	<ul style="list-style-type: none"> 2000年、議会図書館は関連機関と連携し、全米デジタル情報基盤整備・保存プログラム(NDIIPP)を開始(議会の承認を経て、当初総額約1億ドルの予算を得ていたが、2007年4,700万ドルが減額された)。NDIIPPは、議会図書館の主導で、産官学共同で重要なボーンデジタルの情報収集、保存、長期アクセスの保証を目的とする。その予算執行の権限は議会図書館が持っており、大学や民間等の電子情報保存等に係る各種プロジェクトにつき、助成を行っている。議会図書館ではその枠組みの中で、「MINERVA」と称するウェブアーカイブプロジェクトにおいて、2000年大統領選挙、2001年9月11日同時多発テロに関するウェブサイトコレクションを初めとする、ウェブサイトの選択的収集を実施し、インターネットで提供している。 IIPCのコミュニケーションを担当している(2007年現在)。
アメリカ	米国政府印刷局	—	選択的	—	2005年	○	—	<ul style="list-style-type: none"> 2005年、収集ロボットにより、米国政府の各省庁ウェブサイトを収集することで、ウェブサイト内にある個々の電子著作物を、コンテンツ内の文字等から自動的に抽出し、アーカイブする実験を実施。成果報告書が2007年2月に発表。2008年精度をあげるため、さらに継続調査予定。
イギリス	英国図書館等	①Domain.uk ②UK Web Archiving Consortium	選択的	①100サイト(30MB) ②1000サイト	①2001年5月 ②2004年6月～ 2006年6月	○	○	<ul style="list-style-type: none"> 2003年10月、Webサイトも対象とした「2003年法定納本図書館法」が成立。これによって、英国図書館がインターネット情報を包括的に収集することが可能となり、近年バルク収集に着手予定。 英国図書館は、2001年に「Domain.uk」と称するプロジェクトにおいて、英国のウェブサイトの1部を実験的に収集し、ウェブアーカイブに係る課題を検討後、2004年～2006年「UK Web Archiving Consortium」と称する、国内の他機関(英国公文書館等5機関)と協力連携した、選択的収集プロジェクトを展開。成果はインターネットで公開。 ニュージーランド国立図書館とともに、選択的収集ツールオープンソースソフトウェア「Web Curator tool」を開発、公開(IIPCの開発スケジュールの枠組みでの成果)
オーストラリア	オーストラリア国立図書館	PANDORA	選択的	1.2TB	1996年6月	○	△	<ul style="list-style-type: none"> 国内の社会や文化に関するオンライン出版物やサイトを中心に選択的に収集し、ネットで公開。 1999年2月、著作権法審議会(CLRC)は報告書を提出し、①「図書館資料」に電子的資料を含むよう定義拡大、②納本義務の継続、③納本資料の限定的アクセス下での利用、を勧告。現在、著作権法の納本規定改正を視野に入れている。 2005、2006年に、.auドメインの包括的収集を、インターネット・アーカイブに委託して、IIPCが開発を推進する収集ロボットHeritrixで実施。
オランダ	オランダ国立図書館	e-Depot	選択的	6TB	1995年	○	—	<ul style="list-style-type: none"> ヨーロッパ納本図書館ネットワーク(NEDLIB)において電子納本モデルを策定。 エルゼビア社、シュプリングァー社等と協定を結び、電子ジャーナルを収集。 オンライン出版物の納入義務の法的根拠は存在しないが、1999年の出版者協会との合意に基づき、オンライン出版物を収集。 長期保存を視野に入れて、論文等を著作単位で大規模収集。(デジタル・デポジットの例) 2005年から選択的ウェブアーカイビングのプロジェクトを開始。
オランダ	ユーロピアンアーカイブ	European Archive	選択的	—	2006年9月	—	—	<ul style="list-style-type: none"> ユーロピアンアーカイブは、IIPC設立に尽力したJulien Masanes氏が統括する、ヨーロッパのウェブアーカイブを中心としたNPO団体。インターネットアーカイブと協力関係。2010年を目途にドメインレベルでの包括的なウェブアーカイブのためのシステム基盤の構築を目指す。 2007年5月現在、ヨーロッパ憲法関係コレクション、週次、6か月単位の英国政府ウェブサイトコレクション、2006年のイタリアのドメインレベルのスナップショットを公開。
カナダ	カナダ国立図書館・公文書館	—	選択的	1.8TB	1994年	○	○	<ul style="list-style-type: none"> 2004年4月、「カナダ国立図書館・公文書館法」が成立し、第8条(2)において、.caドメインのインターネット情報の包括的収集ができるようになった。2005年12月～2006年1月に国政サイト、2006年10月に国政及び地方公共団体のサイトを集中的に収集した。また、法制化の延長線上において、オンライン電子出版物(電子雑誌等)を納本対象とする規定が作られた。収集コンテンツをインターネットで公開。

諸外国の国立図書館等におけるネットワーク系出版物収集への対応状況

平成19年7月27日
国立国会図書館

国	主導的機関	名称	収集方法	規模	開始年	IIPCへの参加	法制化	特記事項
スウェーデン	スウェーデン国立図書館	Kulturarw3 projekt	バルク	9TB (347,642サイト)	1996年	○	○	・1996年からKulturarw3 projektで.seドメインサイトの収集を行ってきたが、2002年「国立図書館のデジタル文化遺産計画における個人情報の処理に関する政令」が制定され、包括的収集の法的根拠を整備した。 ・国内サイトの包括的収集は、既に10回以上なされている。
デンマーク	デンマーク王立図書館	netarchive.dk	バルク	5.3TB	1997年6月	○	○	・2004年12月、「公表資料の義務的納入に関する法律」を改正し、2005年7月に施行された。このことによって、包括的な自動収集が可能になった。 ・「netarchive.dk」と称するプロジェクトで、IIPCが開発を推進する収集ロボットHeritrixにて、2005年から国内サイトを年4回の頻度でバルク収集を行っている。収集成果は現在研究者に限定公開。 ・2007年秋には、ウェブアーカイブ提供に係る、検索エンジン、提供協機能の開発開始する計画。開発成果はオープンソースソフトウェアにする予定。
ドイツ	ドイツ国立図書館	—	バルク	—	2006年	○	○	・2006年に「ドイツ国立図書館法」が公布された。16条により、ドイツ国内で発行されたすべての出版物、インターネット情報を一元的に収集できることになった。 ・Webサイトの包括的収集は、未だ行われていないが、ドイツを中心にした190の出版社から、電子ジャーナル等の納本を受けている。
フランス	フランス国立図書館	—	バルク 選択的	約3000サイト	1999年	○	△	・従来は、独自基準に基づきつつ、フランス国内のサイトを選択的に収集していた。特に、2002年大統領選挙に関するサイト並びに2004年の地方及びEU選挙サイトが、集中的に採取された。また、2004年から実験的に、フランスドメインの包括的収集が行われている。 ・2006年8月1日「情報社会における著作権及び著作権隣接権に関する法」のとりわけ39条から47条において、国立図書館が、インターネット情報の包括的収集を行うことが規定された。 ・IIPCの調整、技術担当(2007年現在)。
ニュージーランド	ニュージーランド国立図書館	Archives New Zealand	バルク 選択的	—	2002年	○	○	・2003年4月、「2003年ニュージーランド国立図書館法」が成立し、インターネット情報及びオンライン出版物が納本対象となった。31条(3)で、国立図書館長が、いつでも、何度でも、その裁量で、インターネット情報を複製可能であることが明記され、34条(4)で、その情報もともとインターネット上でアクセス制限なく閲覧されていた場合、国立図書館は、それをインターネット上に自由に公開できることが規定された。2006年5月に、上記法を受けた収集範囲の拡大が明確化され、2006年8月から効力を持つに至った。 ・英国図書館とともに、選択的収集ツールオープンソースソフトウェア「Web Curator tool」を開発、公開(IIPCの開発スケジュールの枠組みでの成果)。
ノルウェー	ノルウェー国立図書館	—	バルク	350,000ページ	2000年9月	○	○	・Webサイトも対象としたアーカイビング・トータルシステムを構築しバルク収集を行っている。 ・1989年、「国会関連法の中の公開資料の提出義務に関する法律」、1990年、「王令」、「施行規則」、「指令」により、ネットワーク系電子情報を納本制度の枠組みで収集。2001年8月から2004年12月まで、Paradigmaプロジェクトが行われた。 ・現行「納本法」に、「オンライン通信手段で利用可能な電子文書は二部納本」と規定されている。 ・2005年から.noドメイン全体の収集を開始した。
韓国	韓国国立中央図書館	OASIS	選択的	約3万タイトル	2001年12月	—	—	・2008年の国立電子図書館開館に先駆け、OASIS(Online Archiving & Searching Internet Sources)プロジェクトとして、2004年3月から、公的機関を中心にインターネット情報(ウェブサイト、電子著作物)を選択的に収集している。 ・インターネット情報の収集に当たっては、ウェブサイトの所有者(owner)に対し、基本的に電子メールにて許諾事務を行う。 ・再収集については、収集対象のウェブサイトをクローラが継続的に監視しており、収集されているサイトと現在のサイトの変更情報を数値を比較し、管理者がその変更の規模を判断し、収集・保存を行う。 ・メタデータは現在ダブリンコアを採用。(「国立電子図書館運用基本戦略」2006年8月によると、将来別途「標準メタデータ」を検討中) ・「図書館及び読書振興法」の改訂中であり、オンラインデジタル資源を納本対象に含ませるように関連条項を改正する予定。
中国	中国国家図書館	WICP、ODBN	選択的	120TB(デジタル化コンテンツ等も含む)	2003年1月	—	—	・WICP(Web information Collection and Preservation)と称する、ウェブアーカイブプロジェクトを展開中。 ・収集コンテンツは2万件以上の政府ウェブサイトの収集、105タイトルの電子雑誌(2007年) ・現在デジタル図書館を建設中(2003年～2007年)。 ・2001年10月から、中国政府は中国デジタル・ライブラリー・プロジェクトを発足。ウェブアーカイブについては、その中核であり、中国国家図書館がプロジェクトの中心である。 ・WICPは表層ウェブを扱う。収集にはロボットを使用し、以下の2種類を行っている。 a)サイト単位のミラーアーカイブ…政府情報、無料の電子新聞・電子ジャーナル、中国学 b)ウェブページ単位で主題アーカイブ…北京オリンピック、SARS、中国の有人宇宙飛行等 ・ODBN(オンライン・データベース・ナビゲーション)は深層ウェブを扱う。館内提供のみ。 ・インターネット情報等の「ネットワーク系電子出版物」の納入対象化については、2003年5月に中国国家図書館長から中国図書館法起草委員会に対し提案がなされ、現在審議中である。