

オンラインデジタル資源の収集及び保存管理

国立中央図書館 電算事務官
チェ・ギョンホ（崔景浩）

情報技術の発展は、社会の変化及び人間の日常生活に多くの影響を及ぼしてきた。その影響力は、これからも更に大きくなることは明らかである。特に、人類の文化遺産である情報を蓄積、維持している図書館の役割と機能は、デジタル・パラダイム(Paradigm)が強調される環境の中で、インターネット上の価値ある情報資源の収集と体系的な管理が求められている。

国家文献情報の網羅的な収集・保存の責任機関として適切で求心的な役割の遂行、国策事業として推進している国立デジタル図書館の建設にあわせた多様なデジタル情報資源の確保、将来のデジタル環境におけるデジタル資源の納本体系についての研究、オンラインデジタル資源の収集分野に対する国家的な標準モデルの提示などのためには、インターネット上の価値あるウェブ資源の収集・管理はこれ以上引伸ばすことのできない国立中央図書館の業務課題である。

ゆえに国立中央図書館は、2003 年 12 月からインターネット上の価値あるウェブ情報資源、オンラインデジタル資源の収集・保存プロジェクトを推進するに至った。日本の国会図書館で 2002 年から推進中である WARP プロジェクトと比較すると短い期間であるが、本日この席で、国立中央図書館が今まで推進してきたオンラインデジタル資源の収集及び保存管理業務の現況について説明する。

1. オンラインデジタル資源収集推進の現況及び業務手順

国立中央図書館は、オンラインデジタル資源の収集及び保存のため、2001 年 12 月に関連する試験システムを開発したが、推進に必要な諸般の条件が整わず、2003 年 12 月になって実質的な推進計画の策定と専門担当収集チームの組織が行われ、本格的なオンラインデジタル資源収集のための基本的枠組みを整えた。初期には主に 2001 年開発システムについての点検及び補完、オンラインデジタル資源収集関連の国内外における事例調査に重点を置いて事業を進め、2004 年 2 月に人文、社会、自然科学分野の学界及び研究界の専門家によるオンラインデジタル資源審議運営委員会を発足させた。オンラインデジタル資源の本格的な収集は、2004 年 3 月から開始し、3 名（チーム長 1 名、電算職 1 名、司書職 1 名）

の人員で出発した専門担当収集チームは、現在、計 22 名の多様な学問分野の専攻者で構成され運営中である。

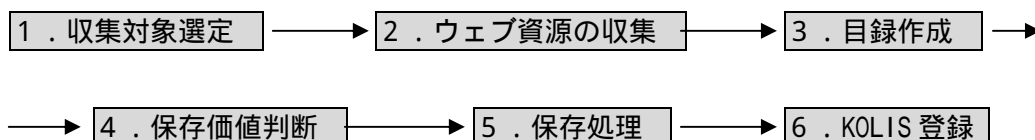
現在、国立中央図書館で収集しているオンラインデジタル資源の類型は、政府機関が生産したオンラインデジタル資源である研究報告書、刊行物、政策資料、統計資料等、主に Web サイト内の公開資料室で提供されている資料である。このような資料は、印刷形態の資料をスキャナー等を利用してデジタル形態に加工した資料もあるが、始めからデジタル形態で刊行しインターネットでサービスをしている資料も多く、いつ消えるかもしれない現代の貴重な文化遺産だといえる。国立中央図書館は、2004 年 8 月時点で、20,000 件余りのオンラインデジタル資源を収集し、特に地域の祝祭、国会議員選挙、新行政首都移転等、韓国国内の社会的に話題となっているもののデジタル資源も収集している。代表的な例として、去る 4 月 15 日には、大韓民国第 17 代国会議員候補者の個人ホームページ 900 件余りを収集した。7 月には、人文、社会、自然科学の 28 主題分野別の価値あるオンラインデジタル資源に関する所在情報 1,400 個 (URL) を関連主題分野の専門家から推薦を受け、収集対象範囲と収集量を漸次拡大している。〈表 1 参照〉

〈表 1〉 . 主題分野別オンラインデジタル資源の推薦現況

区分	主題名	推薦数量
人文科学 (10 分野)	文献情報、語学、文学、歴史地図、コンピュータ、音楽、美術、哲学、宗教、観光	500
社会科学 (9 分野)	経済、経営、企業情報、行政、政治、法学、教育学、風俗 / 民俗、社会学	450
自然科学 (9 分野)	医 / 薬学、生命科学、建築 / 機械、電気 / 電子、天文、地球科学、数学、物理、化学	450
	総計 28 主題分野	1,400 個

当館で推進しているインターネット上の価値あるオンラインデジタル資源の収集業務は、大きく分けて 6 段階に区分することができる〈図 1 参照〉。第 1 段階は、収集対象とするデジタル資源の選定段階であり、現在、オンラインデジタル資源は、選定指針に基づき収集チームがインターネット上の各種資源を収集している。今後、各主題分野別専門家で構成される予定のデジタル資源審議分科委員及び主題専門司書が、定期的に推薦するデジタル資源も収集対象に含む計画である。第 2 段階は、第 1 段階で選定したデジタル資源を、オンラインデジタル資源収集システムを利用して収集する。第 3 段階では収集したデジタル資源の目録作業段階として、題目(タイトル)、URL、発行者、要約文等の基本 DC(Dublin Core) 項目の入力と主題分類を行なう。第 4 段階では、主題専門司書による目録点検と修正、収集したウェブ資源の保存価値の最終判断、保存対象デジタル資源の著作権処理を行なう。第 5 段階は保存処理の段階として、収集したオンラインデジタル資源の保存ファイ

ルフォーマットの変換、CD-ROM 及び DVD に保存するなどの作業をする。最後の第 6 段階は、第 5 段階で保存を決定したオンラインデジタル資源を国立中央図書館統合情報システム (KOLIS: Korea Library Information System) に登録し、保存及び利用者サービスのために準備する段階である。現在、収集したオンラインデジタル資源については、第 3 段階の目録作成まで完了しており、当館が推進予定の主題専門司書制度が導入されれば第 4 段階以降も本格的に進行する予定である。



< 図 1 > オンラインデジタル資源の収集業務手順図

2. オンラインデジタル資源の収集対象資料の選定

収集対象資料の範囲を決定する業務は、資料の保存及び後世における利用価値を予測して決定する非常に重要な作業である。オンラインデジタル資源に対する体系的で効率的な収集のため、国立中央図書館はオーストラリアの PANDORA プロジェクトの収集指針を参考に、2004 年 2 月、オンラインデジタル資源選定指針を策定した。

収集対象とするオンラインデジタル資源の範囲は「韓国語で作成された韓国 (.kr) ドメイン」に限定されている。これはオフライン上で作成された資料に対して納本を国内出版物に限定している事実をオンライン上に拡大した概念で、収集初期ではまず、韓国語で作成された資料に限定して収集し、漸次収集対象範囲を拡大していくものである。

国立中央図書館は、オンラインデジタル資源の収集対象資料を選定する際は、ウェブサイトやホームページにある各種デジタル情報で、国立中央図書館に納本される印刷バージョンが存在しない場合を基本収集対象とする。さらに、将来の情報要求に応える有用性、情報提供者の評判、提供された情報の独自性、学術的内容、情報の最新性、更新の頻度なども参考にしている、主な内容は以下のとおりである。

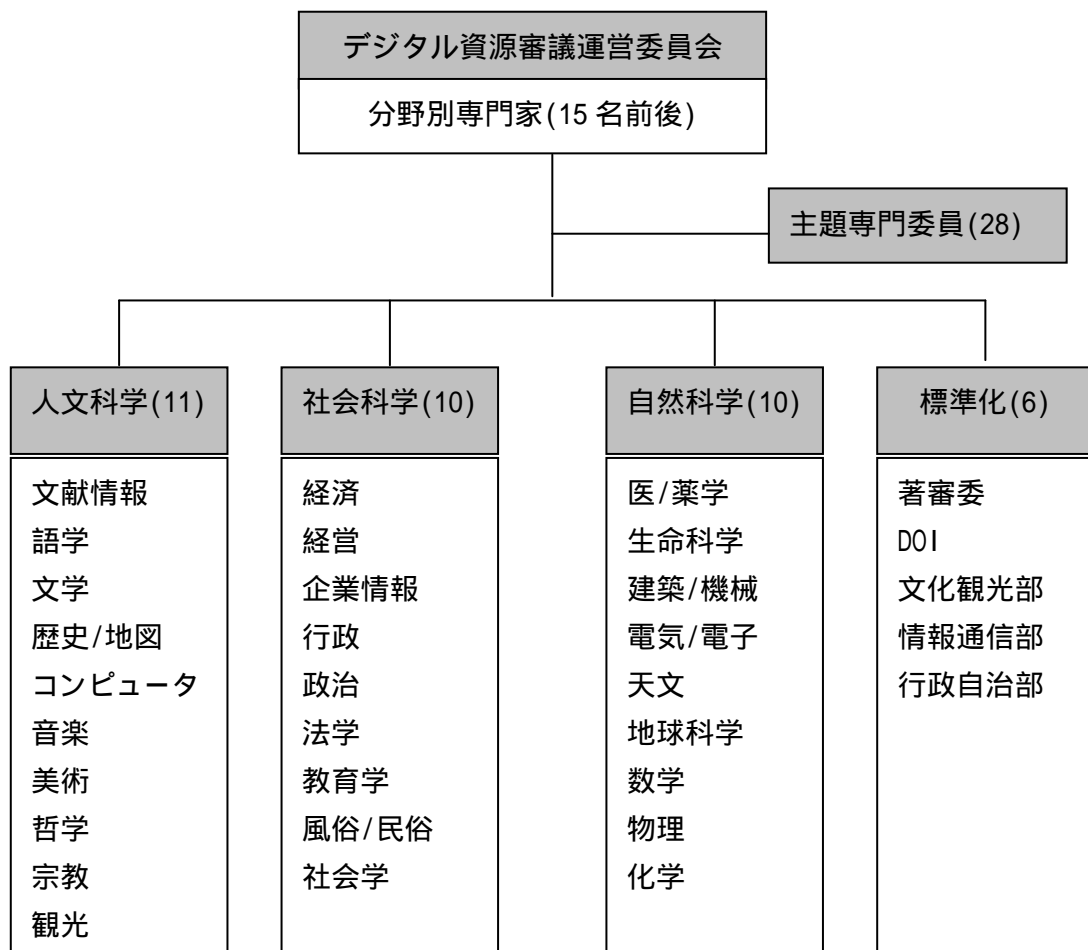
- 韓国のオンラインデジタル資源を最優先に選択する。ここでは、国内で刊行または制作された資料、外国で出された韓国を主題 (対象) にしたデジタル資源、韓国人が執筆 (生産) した資料が含まれる。この時の判別基準はオンラインデジタル資源のメタデータに明示された地理的位置、第一著作者の住所と国籍、ネットワークに搭載された出版物の一次的位置などである。
- オンラインデジタル資源の内容及び形態的側面から、国家の公的刊行物 (報告書、政策報告書、統計資料など)、地方自治団体の刊行物、主題別の代表的な参考資料 (書誌、辞

- 典、名鑑、地図資料など)、人文・社会・自然科学分野の基本資料などの順に収集する。
- あるオンラインデジタル資源のフォーマットが複数ある場合は、標準フォーマットとして作成された資料を優先的に選択する。
 - 国立中央図書館オンラインデジタル資源審議委員会規定により分科委員会の分科委員に委嘱された該当主題分野の専門家が推薦するオンラインデジタル資源を収集する。
 - 国会議員選挙、新行政首都など、最近、韓国国内で話題となっているデジタル資源を収集する。
 - 次のようなオンラインデジタル資源は収集対象から除外する。
 - ウェブに自分の考えを記事のように書いて提示しているウェブサイト
 - 討論リスト、チャットサイト、掲示板及びニュースグループ
 - 印刷物が存在しているオンライン日刊新聞など

3 . オンラインデジタル資源審議委員会の運営

国立中央図書館は、インターネットで流通しているオンラインデジタル資源の網羅的収集・保存の、専門家への諮問などのために、オンラインデジタル資源審議委員会の組織を2004年2月に開始した。オンラインデジタル資源審議委員会は、すでに組織・運営中である運営委員会と分科委員会、専門委員で構成される予定である。<図2参照>

運営委員会はオンラインデジタル資源の収集業務全般にわたって推進状況の点検と評価、広報、関連法制度改善などに対する諮問と共に、審議・議決権を持っており、人文、社会、自然科学各主題分野に対する3つの分科委員会では、現世代の重要かつ価値あるオンラインデジタル資源の所在情報(URL)推薦任務が付与されている。さらに、標準化分科会ではオンラインデジタル資源の収集業務と関連して、著作権、DOI、業務の標準化などに対する事項が審議される予定である。専門委員は分科委員が推薦したオンラインデジタル資源の保存価値などを審議する。



<図 2> オンラインデジタル資源審議委員会 構成図

4. オンラインデジタル資源収集システムの現況

デジタル情報化時代に直面する図書館の役割の変化に伴って、国家図書館として国立中央図書館が、良質のデジタル情報資源の拡充と未来の世代のため、現代のデジタル文化遺産を永久に保存する目的で、2001年12月、オンラインデジタル資源収集システムを開発した。

しかし、当時のオンラインデジタル資源収集システムは、テスト環境の構築段階にある試験的なシステムであり、関連する業務フローについての体系的な分析モデルが不十分な状態で開発された。2004年2月、同システムはアクセス可能なオンラインデジタル資源について選択的に収集し、該当する著作者に、永久保存及びサービスに対する同意を受けることができるという形に拡大・補完された。システムの主な機能は以下の通りである。

▶ オンラインデジタル資源の収集管理

資料の収集には収集対象となるデジタル資源の簡略情報、フォーマット、著作者情報、URL 情報などのメタ情報を直接入力し、WEB Robot に収集させるという方法を採用している。また、収集された文書についての自動分類と要約文の抽出機能があるが、オンラインデジタル資源の非定型的な特性により、正確な結果を期待するのが難しい現実がある。

▶ オンラインデジタル資源の目録作成

収集したオンラインデジタル資源について、長期保存及び利用者サービスを提供するために、主題識別及び整理などの作業を行う機能である。目録作成の第 1 段階は収集したオンラインデジタル資源に対する固有番号を割り当てることであり、割り当てられた固有番号単位で、分類記号参照、分類支援システム、キーワード生成などの機能支援をもとに目録作業を行なうものである。目録作成のためのメタデータは、全世界的に最も多く利用されているダブリンコア (DC) メタデータ形式によって作成する。

▶ オンラインデジタル資源の受付管理

目録が作成されたオンラインデジタル資源について、保存価値を判別し、価値がある資料に対して著作権許諾を要請し、著作権者が著作権の許諾をすることができる機能を提供する。著者に対して著作権を得る方法としては郵便、ファックス、電子メール、ホームページなど、多様なルートが可能であるが、開発されたシステムでは電子メール、著作権許諾のホームページを通じて著作権許諾を得ることができる機能を提供する。しかし、現在、インターネット上のオンラインデジタル資源の匿名性、多様性などにより著作者が見つからなかった場合、無数にあるデジタル資源について著作権の許諾を得るということは相当な困難が伴うと判断される。そのため現在、収集したオンラインデジタル資源についての著作権の解決法について模索中である。

▶ オンラインデジタル資源の保存管理

目録作成が完了したオンラインデジタル資源、書誌情報及び運営環境の情報を、長期間 (Long-Term) 保存するために必要な作業を行う機能である。保存管理で重要なのは「将来のアクセス可能性」である。オンラインデジタル資源を将来においても完璧に判読するために、収集されたデジタル資源の現在のハードウェア、ネットワーク、運営体制、アプリケーションソフトなど、必要となる全ての情報を保存している。また、長期保存のための保存媒体として現在は CD-ROM もしくは DVD を活用している。

5 . オンラインデジタル資源収集の今後の課題

インターネット上の価値あるウェブ資源の効率的な収集のために、これから国立中央図

書館が解決すべき課題を挙げれば以下のとおりである。

- (1) 価値あるデジタル資源の網羅的収集のためには、開かれた推薦体系の導入が必要である。各主題分野の専門家による価値あるオンラインデジタル資源の推薦が受けられる体系の樹立とシステムの開発が必要である。
- (2) インターネット情報サービスの急激な発達とともに、最近のウェブ資源は ASP、JSP、Flash のような動的な形態に発展してきている。こうしたウェブ資源は現在の技術では収集することにおいて多く限界点がある。すなわち特殊な形式の文書である場合、これらの文書に含まれるリンクを分析できる技術研究が進まなければならない。
- (3) 無数にあるオンラインデジタル資源を全て収集することは、事実上不可能である。より体系的に、多様なオンラインデジタル資源を収集するためには、当該資料について自発的な納本がなされなければならない。現在、フランスなど一部の国ではウェブサイト、電子ジャーナル、電子ブックなどのオンライン資源に対する納本が実施されている。こうしたオンラインデジタル資源に対する納本が実施されるためには、関連する法の改定が求められる。
- (4) デジタル資源のアーカイビング部門では、ハードウェアの装備と、デジタルファイルを永久的に保存するため、これを支援するインフラを構築する過程において、かなりの労力と費用を必要とするだろう。定期的なハードウェアの交換、ソフトウェアの更新、保存するデジタル資源のメディア変換作業などに必要な予算確保が必須である。
- (5) また、多様なデジタル情報資源に付与するメタデータの表現技術についての研究が必要である。現在、国立中央図書館において収集しているオンラインデジタル資源は主にウェブ文書資料であり、こうした資源についてのメタデータの表現はダブリンコアを使用しているが、今後収集予定である動画、写真など多様な形態のデジタル資源は、DC で目録を作成するには足りない点が多い。

以上はオンラインデジタル資源の収集及び保存管理に関する国立中央図書館の活動及び成果などについての紹介である。韓国国立中央図書館はこの分野に足を踏み入れてから間もないため、及ばない所もまだ多いが、今日、この場で皆様の御意見を通じて多くの支援を受けたいと思う。最後に、オンラインデジタル資料の収集及び保存管理において韓・日両図書館のさらに深化した業務交流が継続して行われることを心から希望する。