

ネットワーク系電子情報の収集、組織化と利用

(国立国会図書館関西館事業部長 内海 啓也)

はじめに

昨年は、国立国会図書館にとって、関西館開館という 20 数年来の大プロジェクトが成し遂げられ、また国立国際子ども図書館の全面開館など、インターネット時代における図書館サービスの新たな展開のための出発点となる年でした。

「関西館の開館を期して」という合言葉の下、国立国会図書館では、電子情報システムの基盤整備や所蔵資料の個体管理が進められ、同時に関西館の主要な機能の一つを象徴する電子図書館課という組織も新たに設置されたのでした。この電子図書館課では、電子情報の利用・提供や保存に関する研究開発、所蔵資料の電子化、ネットワーク系電子情報の収集と提供、電子情報の発信に関する業務を行っています。

今回は、私は国立国会図書館関西館で行われているネットワーク系電子情報に関する最先端の業務の実際についてのお話を中心にしながら、ネットワーク系電子情報に係る収集、組織化、そして利用について報告したいと思います。

1. 収集

国立図書館としては、言うまでもなく、いわゆる書籍と同様にネットワーク系電子情報の収集のための法的な整備を進める必要がありますが、紙資料の納本と異なって、ネットワーク系電子情報の収集に関する基本的な難しさは、公刊 (publication) という行為が意識されないままに、情報が発信されているものかなりの数を占めるだろうと思われることです。政府や大学や出版社によって発信される情報は publication (公刊) という行為の自覚の下に発信されるわけですから、これらの収集にはそれほど問題があるとは思えませんが、民間の私人が作成し、発信する情報を悉皆的に収集するには様々な問題が生じるだろうとも言われております。

国立国会図書館では、現在、館外の有識者で構成する納本制度審議会で、ネットワーク系出版物の収集のための法制度をどのように整えるべきかについて審議中です。審議会の中間報告は来年の春頃にはなされる予定ですが、答申の方向性としては、国によるものは別として、基本的に発信者の通知による「納本」という仕組みが有力な考えになっているように思われます。

このような状況の中で、2002年4月1日に関西館が開庁すると同時に、同館に新たに設置された電子図書館課では、これからお話しする、二つの事業を開始しました。

これらは、いずれもインターネット上に存在する情報をどのように収集し、利用者に提供していくのかというものです。この事業により、法制度が整備された段階で現実には始まるネットワーク系電子情報の収集と提供に関する業務について、試行的な経験を重ねつつあり、事業を通じて収集・整理された情報は当館のホームページを通じて公開されています。

これらの収集・提供にあたってはさきにも申し上げたように、まだ法制化以前の段階ですので、すべて相手方機関の許諾を得て行っています。これは収集に係る複製権や提供に係る公衆送信権などいわゆる著作権の問題があるためです。

それでは具体的にこの二つの業務について紹介したいと思います。

(1) WARP (Web Archiving Project) について <http://warp.ndl.go.jp/>

この事業の正式の名称は『インターネット資源選択的蓄積実験事業』とありますが、頻繁に更新または削除されていくインターネット上の情報資源を収集し、保存していくというものです。現在収集されているものは、

- ・電子雑誌コレクション（大学を含む官民の研究機関が発信する812件）
- ・政府ウェブコレクション（当館及び国際子ども図書館も含めて7機関）
- ・協力機関コレクション（地方自治体、大学、国際イベントのホームページ等125件）

となっています。

この3つに分類された電子情報の中で特徴的なことを二、三、申し上げておきます。

現在、日本ではいわゆる『構造改革』の政策によって、「平成の大合併」と言われている市町村の合併が推進されています。これらの合併が進みますと、旧市町村が維持公開していたインターネット情報がほとんど消滅してしまうこととなります。日本の歴史的資料としてこれらを保存していく必要があるとの判断の下で、重点的に収集を行っています。国立大学についても同様で、運営及び財政に自由性を持たせた「国立大学法人」への移行に伴った国立大学同士の統合が、今年は数多く行われました。これらについても統合以前のインターネット情報を保存していく必要があります。

また、国際イベントのホームページというのは、例えば去年は日中国交回復30周年でしたが、これを記念したホームページが公開されていましたが、そのイベントが終了すると発行主体がなくなるため、消滅してしまうということになります。これらのホームページも重点的な収集の対象としています。

ところで、これらの業務の一連の流れは次のようになっています。

推薦または選定 受付・確認 管理責任者に収集許諾依頼文書を発送 回答の返送 自動収集ソフトウェアによる収集 WARP に収載

収集に際しての望ましい条件としては、権利関係が複雑に錯綜していないことが挙げら

れます。これは著作権許諾契約に係る手続きの煩雑さを避けるためです。著作権許諾契約に際しての当館側の条件は、複製をすること、利用提供すること、無償とすることで、これに同意してもらうことが必要です。その際、相手方は複製の範囲、利用提供の範囲（インターネットでの公開は可か、館内での利用のみか）、利用提供の時期（即座に公開するか、ある一定の期日に提供を開始するか、ある一定期間経過後に提供を開始するか）などの条件を提示することができます。

また、当然のことながら、現行の WARP で技術的に収集できるコンテンツである必要があります（フラッシュ、ストリーミングデータなどは不可）。

次に収集の頻度についてですが、電子雑誌コレクションと政府のウェブコレクションは月 1 回の更新頻度、協力機関のウェブコレクションについては当該機関との協議に基づく頻度としています。

昨年の 4 月以来、この実験的業務を行う中で浮かびあがってきた主な課題は、いかに実験とは言え、インターネット上の情報を常に監視し、収集すべき情報資源を選択する手間が非常に大きいということや、収集範囲を指定することの難しさや、それから著作権上の手続きと言いますか、許諾手続きに係る事務に多大の人手と時間を要するなどのほか、自動収集ソフトウェアの性能改善、品質管理、URL に代わる安定した持続的な識別子などの検討課題が明らかになっています。

また、「ネットワーク系電子出版物」の収集について、網羅的な収集（バルク収集）ではなく、現在納本制度審議会で検討されているような通知による『納本』の場合の実験的対応として、さきに業務の流れのところで述べました推薦による実験的収集も 10 月末から実施しております。

(2) Dnavi (Database Navigation Service) について <http://dnavi.ndl.go.jp/>

この事業は、今紹介しました WARP と対をなす事業であります。

御承知のようにワールド・ワイド・ウェブ上には数十億のウェブページがあると言われていますが、これはウェブ全体の情報の表層部分にしか過ぎず、有用で貴重な情報資源の多くはデータベース等の『深層』に格納されており、これらの情報資源についてナビゲーションするものです。

対象とするデータベースは主として国内に置かれるウェブサイト上で公開されているものです。その条件は次のようになっています。

ウェブでその情報が開示されており、一定のフォーマットに基づいて標準化された内容の複数のレコードが、検索やブラウジングすることを目的に組織化されており、データベース内では、全レコードが共通の要素で関係付けられているもの

で、次のいずれかの機能を備えているものとしています。

利用者による何らかの行為（検索語投入、ボタン押下等）によって、背後のデータベースの検索が実行され、検索結果が動的に HTML として生成・表示される

静的な情報源であっても、階層ディレクトリ（ディレクトリ=フォルダによる階層型リンク構造によるデータ集合）や簡易検索などのナビゲーション機能が提供されている FTP（ファイル転送プロトコル）等によるダウンロードまたは媒体による、データの一括送付を提供している

このプロジェクトによってナビゲーションが可能なデータベースは、現在約 8 千件で、この中には相当数（約 15%）の画像データベースも含まれていて、日本国内における文化芸術関係のデータベースに案内する一種のポータルサイトの機能を果たし得ているのではないかと思います。

この業務は次のような一連の手続きによってなされています。

国立国会図書館がデータベースを発見または推薦通知 データベースの調査（内容、運営主体、管理責任者、重複の有無等） 書誌情報の確認、分類付与 登録 公開

これらの業務に関係する日本国内の動きをここで紹介しておきたいと思います。日本では今、いわゆる IT 関連プロジェクトとして行政府による e - JAPAN 計画が積極的に推進されています。これは電子政府の構築などを始めとして「5 年以内に世界最先端の IT 国家になる」という 21 世紀における日本の重要な国家戦略なのですが、この動きの中で、ウェブアーカイビングに関する施策の推進も強力な後押しを受けていて、文化遺産アーカイブ、政府が保有する各種データベースなどの利便的な利用を可能にする総合的なポータルサイトを立ち上げるという構想が提案されたりしています。

このようなウェブ情報の収集・利用あるいは保存に関する行政府の動きの中で、国立国会図書館も積極的な関与が期待されており、当館の今後の電子図書館サービス政策もこのような動きと積極的に連携を図りながら進めていくことになるのだろうと思われます。

2 . 組織化

さて、このように収集されたネットワーク系情報資源はどのように組織化が図られているのかについて紹介したいと思います。

国立国会図書館では 2001 年に、『国立国会図書館メタデータ記述要素』をダブリンコアに準拠して策定しましたが、要素の概要は次のようなものです。

タイトル、著者、主題、内容記述、公開者、寄与者(Contributor)、日付、資源タイプ(Type)、フォーマット、資源識別子 (Identifier)、情報源、言語、関係 (Relation)、時間的・空間的範囲 (Coverage)、権利関係

WARP や Dnavi ではこれに基づいて、次のようなメタデータを付与しています。

- ・電子雑誌コレクション タイトル、編者、出版者、主題（NDC）、巻号、資源識別子（URL）、ISSN、別メディア版 ISSN、NDL 資源タイプ（データベースの種類）
- ・政府及び協力機関ウェブ・コレクション タイトル、公開者、主題（NDC）、公開日、資源識別子（URL）、NDL 資源タイプ

ネットワーク系電子出版物の組織化に係る規則等の制定については様々の論点について現在検討がなされているところです。

論点とされている主なところを挙げますと、

収集の方法とメタデータの作成（例えば発信者が付与することがありうるのか、など納本制度のありようと密接に関係してくる）、全国書誌に掲載する範囲をどこまでとするか、メタデータをどの単位で付与するか（粒度の問題）、題情報をどうするのか、など決定すべき事項が多々あり、ネットワーク系電子情報出版物の納本制度の答申と並行して解答を出していかなければならない問題となっています。

3．利用（と保存）

次にネットワーク系電子情報資源の利用に関してですが、国立国会図書館のホームページから誰でも利用することができます。WARP は、公開した 2002 年 11 月から 2003 年 6 月までの平均利用数は 1 日 2,000 ページとなっています。Dnavi も同様の期間の平均利用数は 1 日約 800 ページであり、安定的な利用状況となっています。

このほか、当館が作成した代表的なネットワーク系電子情報に近代デジタルライブラリーという一次情報のデータベース（約 47,000 件）がありますが、これは昨年 10 月の公開と同時に国内に大きな反響を呼び、一日の平均利用数は 5,000 ページ強となっています。これは明治時代に刊行された国立国会図書館の蔵書の本文をイメージで提供しているものであり、今まで利用の便が良くなかった古い時代の蔵書が、利用者自身のパソコンから閲覧できるということで人気を呼んでいるのではないかと考えられます。また、電子展示会をホームページ上で開催していますが、その中の昨年 5 月の憲法記念日に公開された『日本国憲法の誕生』と題した電子展示会は、公開日に約 28 万件の利用を記録しました。

このように、インターネットの発達とデジタル技術の進歩によって、私たちはその利用について今までに経験したことのない便利さを享受しています。国内的にも国際的にも連携協力的手段としてインターネットとデジタル化はなくてはならない道具となっています。

しかし、この便利さは同時に脆さというか、はかなさというか、保存しつつ後世の利用のために伝えていくという点では実に厄介な課題を抱えています。関西館の電子図書館課では昨年 4 月から 3 か年計画で、電子情報保存に関する調査研究を続けています。昨年度の調査報告はホームページ上に公開されておりますので、参考にいただければと思います（<http://www.ndl.go.jp/jp/aboutus/preservation.html>）。今年は継続的な研究調査に加え

て、当館が所蔵するパッケージ系の電子出版物の保存状態について調査することを計画しており、その結果は2004年の3月までには明らかになるだろうと思います。

4. オンライン電子ジャーナルについて

これは、有料で提供される電子雑誌の発行機関やアグリゲーター（aggregator：複数の発行者から電子雑誌を集積して提供する業者）と国立国会図書館が契約を結び（つまり国の予算で利用に係る契約をし）、来館利用者に無料で閲覧提供しているものです。

関西館の建設は図書館サービスの新たな展開を期すものですが、東京と京都と、その距離は約500kmになります。情報通信回線の発達はその距離を克服するというという考えの下、関西館の建設は進められたのですが、このような距離を具体的に克服するものの一つがオンラインジャーナルでした。国立国会図書館は今後ともこの種の資料の充実を図っていく方針です。

現在利用が可能な電子ジャーナルは、『OCLC ECO』（自然科学、人文科学、社会科学）『ScienceDirect』（主に自然科学、医学）『ProQuest 5000（PQ 5000）』（全分野）『ProQuest Newsstand（PQ News）』（世界各国の新聞）で、これらは、いずれも館内での利用に限られ、プリントアウトが許可されている場合、有料で提供しています。

終わりに（電子化と著作権）

今までネットワーク系電子情報の収集、組織化、利用について述べてきましたが、これらの業務を展開していくに当たって、大きな課題は著作権の問題です。これまで紹介しましたネットワーク系電子情報の収集や提供に当たってはすべて著作権処理を行っています。

デジタル情報の生産は中断なく続けられ、紙の出版物に発生した権利とは比較にならないほどその権利関係が複雑になっています。個人の財産権でもあるこの権利に安易に制限を加えることはそう簡単ではないにしても、デジタル化時代において公共的利益と個人の権利について新たな関係を構築していかなければならない時代に私たちは今立っているのだと思います。この関係について早急に社会的な合意を得る必要性が生じています。

現在、納本制度審議会で審議中である「ネットワーク系電子出版物」の収集制度にしても、著作権の問題を抜きにしては、円滑な収拾が非常に難しくなるのではないだろうかということは想像できます。国による悉皆的収集は個人の表現の自由を萎縮させるのではないかという危惧や、それとは反対に、すべての創造的生産物は過去の生産物の恩恵の上に立っているという、いわゆるパブリックドメイン（public domain=公有財産）の考え方の下に著作権は制限されるべきであるとする考え方が存在しますが、この問題にどう対処するのかは21世紀初頭の大きな課題だと私は考えています。