

中国国家図書館のデジタル資源長期保存の課題と対策

中国国家図書館 文化教育部
主任 曹寧

この十年来、社会文化の記録機関の積極的な参与により、情報資源コンテンツのデジタルでの作成（ポーンドジタル情報）及び伝統文献に対するデジタル化作業（リポーンドジタル情報）は重要な成果を収め、現在の社会デジタル資源の重要な一部分を構成することとなった。これらのデジタル資源を広く利用することで、人類の社会生活に対し全く新たなデジタル時代を切り開き、情報伝達・知識共有・娯楽及び教育に対して新しい方法を提供する。この歴史の過程の中で、デジタル資源の長期保存は、次第に社会の各領域が共通して直面する課題となりつつあり、あらゆるデジタル資源生産者と消費者とが直面する共通の課題かつ一致した要求である。このため、国家デジタル資源体系を構築する上での重要かつ不可欠な仕事の一環にもなった。

1. 中国国家図書館デジタル資源の構築と保存の現状

中国国家図書館のデジタル資源は以下の三つの部分から成る。第一に当館自ら構築したデジタル資源であり、第二に当館が自発的に収集したネットワーク情報資源であり、第三は外部からの購入等のルートを経て入手した商用デジタル資源である。

1.1 当館自ら構築した資源の分野では、2008 年末までに中国国家図書館自ら構築したデジタル資源の総容量は 180TB に達し、その中で現代の文献は以下のとおりである。電子図書 34 万タイトル、1 億 1,868 万ページ、博士論文 13 万タイトル、1,677 万ページ、中国語電子新聞 291 紙、音楽・歌曲 51.5 万曲、映像資料 5.12 万時間、マルチメディア出版物のデジタル化 1,300 枚、オンラインビデオ講座 386 回。

民国時代の文献は以下のとおりである。図書 11,259 タイトル、286 万ページ、逐次刊行物 4,515 タイトル、8.6 万期、504 万ページ、法律 11,118 件、1,620 万字、年画 607 種、734 幅。

中国語古典籍は以下のとおりである。甲骨実物の画像 6,575 種、11,732 コマ、甲骨拓本の画像 5,773 種、6,990 コマ、敦煌写本 820 巻、2.3 万コマ、石刻拓本 2.4 万種、2.9 万コマ、地方志 6,868 件、334 万葉、西夏文献全文画像 1.4 万ページ、善本の古い写真 7,310 種、

7328 コマ、中国古代の典籍 1,084 種、1 億字。

中国国家図書館の上述デジタル資源に対する保存は、基本的にはそのメディアを保管及び維持するという段階に留まっており、これらのデジタル化資源は、一般に光ディスクやハードディスク、磁気テープ等の物理メディアに格納・保存している。技術的な制限やメディアの劣化等の原因により、統計によると、これら当館が構築したデジタル資源オブジェクトの再生不可能率は 2.75%となっており、ソフトウェア環境の陳腐化率は 3%となる。

1.2 ネットワーク情報資源の分野については、中国国家図書館デジタル資源部李春明副主任が詳細に紹介するので、ここでは贅言を弄さない。

1.3 商用デジタル資源の分野では、2008 年末までに、中国国家図書館は計 136 件のデータベースを外部から購入し、中国語データベース 59 件、外国語データベース 77 件から成る。その内、中国語逐次刊行物全文データは 3 万余種、外国語逐次刊行物全文データは 1.5 万種、中国語図書全文データは 27.1 万タイトル、57.3 万冊、外国語善本図書全文データは 33.8 万タイトル、中国語新聞全文データは 1,639 紙、外国語新聞全文データは 681 紙、中国語学位論文は 194.5 万本、外国語学位論文全文データは 3.2 万本、中国語会議論文全文データは 212.5 万件、外国語会議論文全文データは 6,279 件である。

上述の商用デジタル資源のうち、同時に物理メディア（例えば光ディスク）を伴って出版される出版物や、物理的にデータのバックアップを提供できるデータベースについては、中国国家図書館は納本及び購入等の方法でバックアップの保存を進めてきた。しかし保存された有形メディアのデジタル資源は、往々にしてデジタル資源の生データであり、サービスプラットフォームを備えておらず、その内容は断片的かあるいは不連続である。このほか、中国国家図書館はビジネス交渉を経て、契約によって主要な中国語商用デジタル資源のミラーサイトを構築している。これはある意味ではバックアップ保存の役割を果たすことになるが、決して完全な意味での商用デジタル資源のローカル保存とはならない。統計によると、上述の商用デジタル資源オブジェクトは再生不可能率が 3～5%となり、ソフトウェア環境の陳腐化率は 11%になる。

2. 中国国家図書館デジタル資源長期保存の事業進展

中国国家図書館はデジタル資源の長期保存の領域において、既に長年の理論的研究と具体的実践を展開させてきた。戦略的任務、全体目標、法律や制度の整備、業務要件、業務計画及び実験的プロジェクト等、多くの分野で一連の成果を得た。

2.1 戦略的任務

「国家図書館十一五計画綱要」の中に明文で指摘しているように、「文献情報資源の収集と保存を強化し、高品質な国家文献情報資源保存拠点を構築し、内容豊富であり、メディア

が多様であり、あらゆる文献が揃い、明確な特徴をもつ文献情報資源体系を形成する」必要がある。また「所蔵文献情報資源の保存及び保護に力を入れ、文献の保存条件を改善する。電子資源及び音楽映像資料の保管庫を建設し、中心的、貴重、あるいは特徴的な所蔵資料のマイクロ化やデジタル化といった、再生性を維持する事業を強化し、各種のデジタル情報資源の分析研究と戦略的保存を強化する」必要がある。

2.2 全体目標

中国国家デジタル図書館はオンライン、ニアライン、オフラインを互いに組み合わせる方法で、データ保存を進めるつもりである。オンラインストレージには磁気ディスク、磁気テープ、光ディスクといった三種のメディアを採用し、ニアラインストレージには SATA ハードディスクアレイをメディアに採用し、オフラインストレージには光ディスクと磁気テープの二種のメディアを採用している。総保存能力はオンラインで 150TB 以上、ニアラインで 150TB 以上、オフラインで 360TB 以上に達するであろう。

2.3 規則や制度の整備

中国国家図書館は 2007 年 5 月に「国家図書館デジタル資源保存管理暫定条例(草案)」を公布した。この条例は以下のように明確な規定を行っている。

2.3.1 中国国家図書館デジタル資源の保存原則

デジタル資源保存の全体的な原則は次のとおりである。中国語デジタル資源を重点的に保存し、また外国語デジタル資源を選択的に保存する。あわせて、様々なデジタル資源の特徴とライフサイクルに基づき、技術の発展に従って、保存と管理の方法を絶えず補完する。デジタル資源保存の具体的な原則は、独立性の原則、重点保存の原則、再現性の原則、相補性の原則、分業協調の原則等を包含する。

2.3.2 中国国家図書館デジタル資源のクラス別保存制度

長期保存クラス：国家図書館の職責を履行するために、国家デジタル資源保存センターの任務を引き受けて長期保存するデジタル資源は、長期保存クラスとなる。長期保存クラスのデータはローカル及びリモートに分散して保管しており、光ディスクと磁気テープを保存メディアとし、そして同時に 3 部を保存してバックアップとする。

不定期保存クラス：永久保存が必要かどうか、しばらくは確認できないデジタル資源、及び当初は保存する価値があったものの、一定期間後しだいに保存の価値を失ってしまう可能性があるデジタル資源は不定期保存クラスとする。不定期保存クラスのデータは 1 部ないし 2 部をバックアップとする必要がある。一定期間後、関連する基準に基づいて長期保存クラスとしたものは、長期保存クラスの要求基準に応じたバックアップが必要となる。

臨時保存クラス：オンラインサービスのデジタル資源に異常が発生して消失あるいは損壊

した時に、それがただちに復旧してサービス提供が可能になることを確実にするデジタル資源が臨時保存クラスである。このグループの資源は中国国家図書館が公開してサービスするデジタル資源、ベンダーがミラーを提供するデジタル資源、及びオブジェクトデータのリンクを伴うメタデータ資源(すなわち当館が購入及び権利の委託を受けた単なる商用データベースのメタデータであるが、リンクを通じてその対象データを使用できるもの)からなる。

2.4 業務要件分析

中国国家デジタル図書館建設事業の中で、デジタル資源組織化システム及び保存システムの業務要件分析は中核作業の一つであった。現在、すでに完全版の業務要件説明書ができている。その文書の叙述によると、中国国家図書館の資源保存機能は資源ファイリング機能、資源管理機能及び資源永久保存の3つの内容からなる。

2.4.1 資源ファイリング機能：データ受取り機能、クラス別に保存する機能、エラーチェック機能、AIP 取得機能からなる。

2.4.2 資源管理機能：この機能は主に各グループの資源を対象としたユーザーポリシー、関係する権限、データフォーム、データの唯一性、データインデックス情報等の管理を達成する。データベース管理機能、クエリ実行機能等からなる。

2.4.3 資源永久保存：この機能は主に資源のバックアップ、ファイリング、ディザスタリカバリ用リモートバックアップ、データ資源のマイグレーション等の機能を達成し、同時にシステム及び関連環境に対するモニタリング、永久保存規則とデータマイグレーションの戦略の制定等の作業を達成する。

2.5 業務計画

中国国家図書館は、デジタル資源の長期保存分野において、図書館ブランド発展戦略と結びつけた「中国アーカイブ」計画を提出した。「中国アーカイブ」の目的は、デジタル情報資源コレクションを通じて、「保存可能な価値あるデジタル資源を全て保存する」目標を実現することである。これにより、各種の公共デジタル情報資源の統一収集、保存と管理の集中が実現する。異なる起源、異なる形式、異なるフォーマットのデジタル情報資源の集中的なアーカイビングと有機的な統合が進行する。デジタルアーカイブを通じて、信頼に足る長期的持続的なアクセスが実現する。このプロジェクトは、デジタル資源の種類とフォーマットに従い、「中国の画像」「中国の動画」「中国の音楽」「中国のゲーム」「中国の web」「中国の blog」「保存サービス」「中国のメールサービス」などのサブブランドに分析することができるだろう。

2.5.1 インターネット情報資源アーカイブ

中国国家図書館は、インターネット資源保存の面では、インターネット資源の歴史的情報の整理に力を入れ、米国の Internet Archive 社と北京大興天網（スカイネット）実験室から過去のインターネット資源を取得するべく力を尽くす。さらにインターネット情報資源アーカイブの長期保存の実践を拡大し、早々に立法化を促進することによって、ネットワーク情報資源の収集、保存、提供のための免責権を獲得し、国内の主要ネットワーク情報資源の刊行者や仲介サービス業者との間に友好的な協力関係を確立し、共同設立方式により、中国語のインターネット情報の体系的保存と分類サービスを実現したいと考えている。

2.5.2 政府公開情報アーカイブ

中国国家図書館は「政府公開情報統合サービスプラットフォーム」構築を計画し、政府公開情報の収集・保存方針及び制度の策定指針を制定して、系統的に中央政府ならびに各省政府公開の情報を収集・保存し、また、専用のメタデータスキームを制定し、全国の省級公共図書館と共同で中国政府公開情報アーカイブ連盟を設立し、我が国の政府公開情報に関する付加価値情報サービスと検索サービスを可能にした。

2.5.3 オープンアクセス資源アーカイブ

中国国家図書館は、我が国のオープンアクセス資源の保存計画と管理方針のドラフトを制定して、国家レベルの分散式のオープンアクセス資源の保存基盤となる環境を確立し、関連するアーカイブの権利と責任の分担を実現し、オープンアクセスの雑誌とオープンアクセスのコレクション中の学術資源を重点的に保存する。国家図書館は、早期に、オープンアクセス資源のメタデータ及びオブジェクトデータの転送とカプセル化のプロトコルを制定するであろう。OAI のサポートに基づき、第三機関による信頼性の高い保存サービスの選択的ハーベスティングと逆ハーベスティングのメカニズムを確定するであろう。オープンアクセス資源のデジタルオブジェクトについて、様々なレベルで表現するモデル（ビットトレント等）によって、カプセル化を実現するツールを開発し、関連する技術プロトコルとワークフローの確立を通じて、オープンアクセス資源の長期保存を保障するであろう。

2.5.4 ブログ資源アーカイブ

中国国家図書館は中国国家デジタル図書館のインフラを利用し、早期にブログ収集のパイロットプロジェクトを展開するであろう。実験を通じて、ブログ収集における法律的、技術的、社会的経済的問題を明らかにする。その上で、ブログ資源の取り込み、格納、組織化及び公開のための管理システムを構築し、完成する。国家図書館は関連立法と法定免責権の整備を進めることにより、ブログアーカイブ構築の権利を獲得する。このインフラ整備によって、ブログ資源の内容選択基準を明確化し、専門的な収集、組織化及び保存技術の基準を制定する。また、共同構築の仕組みを取り入れ、ブログ資源の長期保存の戦略的提携を確立して、学術的社会的価値のあるブログの内容を、適切に保全保護できるようにするであろう。

2.5.5 電子メールアーカイブ

中国国家図書館は、電子メールのアーカイビング業務を展開するであろう。まず、電子メールの選択保存政策と保存システムを確立する必要がある。公共性のある政府や機関の電子メールを優先的に保存し、私人のメールについては、有名人による学術的な価値、保存する価値のある電子メールの保存を優先的に考慮する必要がある。電子メールアーカイブのインフラを確立し、電子メールの収集、保存、公開に到る全工程の標準化された管理を実現し、著作権管理のフレームワークを定め、政府、科学研究機関、教育機構、電子メールサービスプロバイダ及び個人と提携して、電子メール保存の実験研究を展開するであろう。

2.5.6 デジタル出版物アーカイブ

中国国家図書館は国家の総書庫として、デジタル図書、デジタル雑誌、年鑑、レファレンスツール、デジタル新聞と全文データベースなどの正規出版物デジタルアーカイブを重点的に行い、その上に、学位論文、会議文献、政府出版物等の灰色文献のデジタルアーカイブ計画を徐々に実施していく計画である。これを行うために、近代的なデジタル出版物の納本・保存の制度体系の構築を急ぎ、デジタル出版物、特にネットワーク出版物納本の立法化を促進するだろう。納本義務の主体を明らかにし、電子出版物納本の類型と範囲、数量と方法を確定し、全国各レベル各種出版機関のデジタル出版物納本の管理を強化し、効率的なデジタル出版物の納本・アーカイブを行うための技術な方法と補償の仕組み及び明確なデジタル出版物アーカイブサービスの標準的な規範を定める。デジタル納本の法的地位と、図書館及び利用者の権利を明らかにし、デジタル出版物アーカイブの協力体制を形成するであろう。

2.5.7 デジタルアーカイブサービス

中国国家図書館は中国国内に知的財産権があるソフトウェアとハードウェアの登録、図書館及び文書館システム中の保存システム及び保存メタデータ登録、知的財産権のファイルフォーマットでの登録を進め、デジタル資源長期保存システムが使用する統制語の登録と解析のサービスを提供する。中国の国情に基づいたデジタル情報保存システムの認定と評価指標体系を制定する。

3. 展望

3.1 中国国家図書館は政府の承認により、ネットワーク資源の収集、保存とサービスを積極的に進めるだろう。法令の制定または行政の許可により、国家図書館とその他の指定図書館の権限を明確にする。また、長期保存中国語デジタル資源の権利と責任を有し、権利侵害についての免責を得る。すなわち著作権者の承認を得ていない状況においても中国語デジタル資源、特に Web サイトや Web ページの保存が可能になり、フェア・ユースが可能となるであろう。

3.2 中国国家図書館は、国家図書館をはじめ、他の国家レベルの専門図書館や大型の省レベルの公共図書館、大学図書館の参画する中国語デジタル資源の保存体制を構築するであろう。インターネット上の価値ある中国語ネット情報資源を重点的に保存し、商業機関が作成した高品質の中国語デジタル資源を保存し、我が国に世界中国語デジタル資源の保存拠点とサービスセンターを建設するであろう。国家レベルの専門図書館や重点大学図書館において導入された学術デジタル資源の長期保存体制に、国家図書館は積極的に参加・協力する。