

## 中国国家図書館における情報資源の長期保存事業 —戦略計画と政策措置

中国国家図書館 情報ネットワーク部・デジタル資源部

### 一、概説

コンピュータ、ネットワーク等の技術の急速な発展に従い、伝統的な媒体による文献の図書館での比率は次々と下がり、伝統的な媒体から変換されたデジタル資源とポーンデジタル資源がますます多くなって、長期保存が急務のデジタル化情報資源は私たちの事業における新たな挑戦となっています。デジタル資源は特殊性があり、ひとたび壊れたり更新されたりすると、元の姿に戻すことができず、情報の伝達も継続することができなくなってしまいます。デジタル資源の数量・種類が日々増加するのに従い、デジタル資源の長期保存の重要性と緊急性はますます際立っています。

国家の総書庫・全国書誌センター・国家古籍保存センターとして、中国国家図書館は文明を伝承し、社会的な目的に寄与し、国内外の文献の収集と保存の職責及び対外文化交流の機能を担うという使命を果たしています。近年、中国国家図書館はデジタル資源の構築と保存などの面に大量の人的・物的資源を投入しており、所蔵するデジタル資源は絶えず豊かになり、デジタル資源の保存の面にも新たなニーズが生まれています。

中国国家図書館のデジタル資源を保存するという職責をよりよく履行し、科学的にデジタル資源の保存と再利用を管理するために、当館は、デジタル資源の長期保存事業において統一的に計画し、合理的に分業し、「保存を主とし、利用に役立て、特色を強調し、知的財産権を重視する」という原則を定め、先進的な技術・設備を利用して万全なフローを定めます。それにより価値あるデジタル資源を保存し、最大限に所蔵デジタル資源の安全を確保し、的確かつ簡便にその能力を引き出し、今後のデジタル資源情報に係る利用提供を保障します。

### 二、中国国家図書館所蔵資源保存の現状

中国国家図書館の主管による国家デジタル図書館プロジェクトは2005年に構築を開始しました。2010年末までに、中国国家図書館のデジタル資源の総量はすでに480テラバイト

(TB) に達しています。主な資源としては、電子新聞の納本 (3 TB)、購入データベース (71 TB)、特色あるコレクションのデジタル化 (388TB)、ネットワークナビゲーション及びネットワーク資源収集 (18TB) の 5 つの部分になります。ネットワーク資源は、ネットワークアーカイブ (WARC フォーマット)、電子書籍、論文、電子新聞等で、主な構築方法はインターネットのクローリングを通じてネットワーク情報資源を形成するというものです。現在の国家デジタル図書館のデジタル資源は、媒体の形態別に、光ディスク・磁気ディスク・磁気テープなどのリムーバブルメディアの形で出版された有形デジタル資源と、ネットワーク上で出版されたネットワーク情報資源とに分けられます。資源の使用方式としては、国家デジタル図書館でインターネット方式・LAN 方式及び光ディスク検索方式によるアクセスと検索を提供しており、知的財産権を保護すると同時に利用者の通常の使用を保証しています。

現在、国家デジタル図書館のデジタル資源の領域は以下のように分けられています。現実の所蔵部分、すなわち当館コレクションは、中国国家図書館自体が収集・所管・整理したデジタル資源です。中国国家図書館が資源の所有権を持っており、かつ、利用・統制・更新・修正等の権利も持っています。一方、バーチャルな所蔵部分は、コンピュータネットワークと通信設備を通じてインターネット上で共同利用している館外の情報資源です。すなわち中国国家図書館は使用权だけを所有しているデジタル資源であり、この部分の資源は使用权のみの所有ですから、中国国家図書館デジタル資源の保存範囲とはなっておりません。

現在、国家デジタル図書館はすでに先進的なストレージ技術を採用し、ネットワーク化されたストレージフレームワークシステムを設計して、ストレージネットワークに基づいた大量デジタル資源のライフサイクル管理を中心とする、デジタル資源のストレージに係るバックアップ/アーカイビングシステムを成功裏に構築しました。これと併せて、統一的な管理対策を立案し、レベル別のストレージ機能を採用することで、中国国家図書館デジタル資源長期保存システムの堅実な基礎を築くことができました。中国国家図書館のデジタル資源に係るストレージのバックアップ/アーカイビングシステムは、1340TB の保存容量を備え、ストレージネットワークシステムを通じてデジタルデータの自動バックアップと集中管理を行い、安全で信頼できる効率の良いバックアップとリカバリーの管理メカニズムを構築しました。2012 年 10 月までに、中国語図書、博士論文、デジタル家譜、ウェブページ収集などを含む 10 種類の資源に係る長期保存事業を完成し、その数量は 472TB、146 万 3,582 冊/種、2 億 6687 万 8344 ページとなりました。

データの安全を確保し、データ記録媒体が古びてしまうことによるリスクをできる限り避けるために、中国国家図書館は光ディスクをデジタル資源長期保存の媒体としています。2012 年 10 月までに、音楽映像出版物は 21 万 4856 枚 (このデータは実際よりもかなり少ないですが、複本を数えていないためです)、電子出版物は 7 万 7936 枚、自館によるデジタル化資源は 38 万枚を数えています。これらの資源の保存に際しては、除湿・防湿、定期的なほこりの除去、防磁などの複数の対策を実施し、人員・運用を規格化することで、保

存上必要な環境を満たしております。

### 三、長期保存事業の政策措置

中国国家図書館は、デジタル資源を含めた各種情報資源に関する国の法定保存機関であり、各種情報資源の保存管理を行うという、一般の情報サービス機関と比べると特異な性質を持っています。長期保存事業をよりよく完成させるため、中国国家図書館は関係する政策・基準・法律などについて一貫して研究と実践に努力し、保存事業の中で発生した問題に取り組み、万全な長期保存システムを構築するために経験を重ねてきています。伝統的な文献の保存に比べると、デジタル資源の保存はその開始が遅かったのですが、大量のデジタル資源の出現により、保存政策・基準・法律等の面で厳しい挑戦を受けております。また、前例がないことですから、多くのことについて中国国家図書館により初めて切り開き実践される必要がありました。中国国家図書館はデジタル図書館の最前線として、長期保存に係る政策措置についても積極的に尽力し、研究と実践を積んできております。次に、中国国家図書館がデジタル資源保存に対処するために採用したいくつかの政策と方策についてご紹介します。

#### 1. レベル別保存の方策

国家図書館はデジタル資源の保存と利用のそれぞれ異なる特質に的を絞り、デジタル資源保存を長期保存レベルとアーカイビングレベルに分け、デジタル資源の現在及び将来の利用要求にそれぞれ応えるようにしています。

(1) 長期保存レベル：国家図書館としての職責を果たし、国家デジタル資源保存センターとしての任務を担うに際して永久保存に値するデジタル資源を長期保存レベルとしています。長期保存レベルのデジタル資源は、所蔵する全てのメタデータ（書誌データ、標準データ、分類主題データ、所蔵資料データ）、デジタル化した特色あるコレクションのデータ、主要な中国語ネットワーク資源、海外から返還された貴重古典籍資源、永久保存権を得た中国語資源データベース及び国外の重要なレファレンスツールデータベースです。長期保存レベルのデータは同時に3部以上の保存が必要です。

(2) アーカイビングレベル：現在の利用要求を満たすために、永久保存すべきかどうか当面は確定できないデジタル資源及び長期保存条件に満たない資源は、アーカイビングレベルとしています。アーカイビングレベルで保存しているデジタル資源は、中国国家図書館が発信しサービスしているデジタル資源、データプロバイダーが提供するミラーリングされたデジタル資源、所蔵する特色あるコレクションのデジタルデータ（メタデータは全てまたは部分的に欠落）、所蔵する音楽映像電子出版物資源及び対象データとリンクするメタデータ資源（つまり、当館が購入または使用権を得ているものは商業データベースのメタデータですが、リンクによってその対象となるデータも使用できるということです）ならびに資源のメタデータ（記述メタデータ、管理メタデータ、構造メタデータを含む）と対象データのコンテンツとフォーマットの完全性を保証することができない、あるいは

付帯する使用説明の完全性を保証することができないデジタル資源です。アーカイビングレベルで保存するデータは、1部または2部を保存し、関連する基準に照らして長期保存レベルとなった時には、長期保存レベルで求められる条件で保存します。

このほか、中国国家図書館は大量の購入資源及びその他不確定なデジタル資源を持っていますが、永久保存庫での保存管理はせずに、オフライン方式として光ディスクを作成し、5つの倉庫に保存しています。

## 2. 標準規格

デジタル情報資源の長期保存に関する標準規格の研究分野においては、すでにいくつかの国際的な保存規格があります。例えば OAIS (Reference Model for an Open Archival Information System, 開放型アーカイブ情報システムのための参照モデル)、METS (Metadata Encoding & Transmission Standard, メタデータ記号化・伝送標準)、PREMIS (Preservation Metadata Implementation Strategies, 保存メタデータ：実装戦略) などです。現在、国家図書館はすでに『国家図書館長期保存規格』によって、国家図書館の長期保存メタデータ規格、情報パッケージのカプセル化標準規格及びアプリケーションガイドを策定しました。同時に、中国語デジタル情報資源の特性と、当館が独自に作成したりソースフローに基づいて『図書館デジタル資源長期保存メタデータ規格』を策定しました。

現在構築中の国家デジタル図書館保存システムは、国際標準規格 OAIS のシステムアーキテクチャに準拠し、また、国家図書館の現有デジタル資源を保存するために、実際の業務フロー及び国家図書館デジタル資源の特性などの要素をふまえ、総合的に考慮したものです。次の段階として、国家図書館は、図書館としての業種の特性と国家図書館としての特色をふまえた保存システムを構築する中で、上述した標準規格の実践と応用を完成させていきます。

## 3. 規準の整備

デジタル資源の保存に関する規準については、国家図書館はすでに『国家図書館デジタル資源長期保存管理規則』を策定することによって、デジタル資源の保存原則、保存レベル、保存フロー等を明確にしています。同時にデジタル資源保存の実際の業務をふまえて、『長期保存デジタル資源交換フロー』『デジタル資源長期保存アーカイビング検査規格』『長期保存デジタル資源保存業務フロー』『長期保存デジタル資源定期検査業務フロー』などの一連の現実的に実行できる運用規格とフローを作成しました。特に、デジタル資源の保存に係るフローと措置については細分化を図り、データ受理、品質検査、電子書庫への登録、電子台帳登録、定期サンプリング、アクセスなどに関する詳細な運用規則を定め、専門の担当者を置くことにしました。

## 4. 法律・政策

デジタル時代に入り、出版者は認証システムを取り入れています。つまり今のところ当

館は、購入した多くの電子資料について、いずれも所有権を持っておらず、単に「使用权」という方式でデジタル資源を購入・利用しているにすぎないのです。

現在国家図書館は、『出版物サンプル納本管理条例』〔国の法規に相当〕を成立させるために活動しており、この条例が成立して納本制度が完全なものとなることを望んでいます。出版者にデジタル資源の納本に係る責任と義務があることが法律の形式によって明確となり、国による管理の下でデジタル資源の長期保存が保証されることを期待しています。

ここまでで述べた通り、国家図書館のデジタル資源の長期保存事業は、政策措置の面でもすでに少なからず成果を上げておりますが、デジタル資源が膨大な量であるのに対し、依然として保存事業に投入できるリソースが限られているという矛盾があり、バランスのとれた対応が必要となっております。

#### 四、長期保存事業戦略計画

国家図書館は、デジタル資源を含めた各種情報資源に関する国の法定保存機関であるため、戦略計画は、構築と使用の両面を重視しなければなりません。そのために一方では爆発的に増え続けるデジタル資源に対応し、他方では最先端技術をフォローし、技術的なサポート体制を整備し続けることが重要です。

ここ10年の間に、国家図書館のデジタル化情報資源は爆発的に増え、「第12次5カ年計画」(注：2011～15年)の末には、国家図書館のデジタル資源の総量は1万テラバイト(世界の主要な先進国の国立図書館が所蔵する文献総量の合計を超え、約26億冊の図書、あるいは926万時間の動画に相当します)、そのうち電子書籍は200万タイトル、電子ジャーナルは1万2千タイトル、電子新聞は2千タイトル、音楽データは20万時間・100万曲、動画データは30万時間・150万編、インターネット情報資源は1千テラバイトとなります。これにつれて生じるストレージの需要も跳ね上がっていますが、特に全文テキスト情報とマルチメディア情報が大幅に増加したことから、膨大な情報の長期保存に係る効率性と安全性につきより高いレベルでの対応が必要となっております。

デジタル資源の長期保存システムは、膨大な資源のストレージを実践するための技術的なサポートを提供するものとなります。このシステムはデジタル図書館プロジェクトの重要項目の一つであり、主に全館の長期保存デジタル資源の資源管理とストレージ業務を完成させるものです。これによって文献デジタル化加工、デジタル資源化加工、電子資料納本、ネットワーク資源収集などが生み出す各種のデジタル資源をアーカイビングして保存し、完備された電子アカウントを構築し、長期保存デジタル資源のオーダリング管理を実現します。同時に資源の状態とアプリケーションの要件に応じて、それらをオンライン・ニアライン・オフラインの三層ストレージシステムに配分し、デジタル資源の長期保存と利用を実現します。そのシステムには二つの面があり、一つはデジタル資源の変更と破壊の防止、もう一つはデジタル情報資源の長期的な信憑性とアクセス可能性の保証です。デジタル資源保存のマイグレーション方針、データ検査対策、不測の事態によってデータが破損した際の対策、データの安全なストレージの実現などといった技術的な保護メカニズ

ムは、長期保存システム構築の中で段階的に計画を策定し、一步一步実現していきます。デジタル保存技術システムの複雑性と変動性という特色により、完全に規格化された技術システムを作り上げることは難しく、これが長期保存システム構築の課題となっています。国家図書館デジタル資源長期保存システムは、国際規格 OAIS (Open Archival Information System) のシステムアーキテクチャと国家図書館の現有デジタル資源保存業務フローを元にして構築を進めることとなります。現有技術に対して適切な整理、総括をすることによって、当館の現実にふさわしい長期保存管理モデルとサービスモデルを模索し、国家図書館デジタル資源を効果的に保存し、歴史・文明を継承するという責任を果たしていきます。

## 五、展望及び構想

2011年5月、文化部と財政部は『「デジタル図書館推進プロジェクト」の実施に関する通知』を発し、デジタル図書館推進プロジェクトの全国的な実施につき手配を行いました。推進プロジェクトの全体目標は、国家デジタル図書館プロジェクトや各地方のデジタル図書館建設における既存の成果に基づき、各図書館と各地で構築済のデジタル図書館システムとを連結することで、高度な共同利用のための統一プラットフォームを構築し、全国のデジタル図書館をカバーするサービスシステムを順次形成しようとすることにあります。このサービスシステムは、各地域のデジタル図書館システムと省・市・県のデジタル図書館とを組み合わせる構築されます。「デジタル図書館推進プロジェクト」は、国家デジタル図書館プロジェクトの理念・技術・基準を推奨し、「一庫一网三平台 [“一つのデータベース、一つのネットワーク、三つのプラットフォーム”]」の構築を通じて、新たなメディアに基づく図書館サービスの新たな業態を打ち立てようとするものです。「一庫一网三平台」の構築とは、すなわち各レベル分散型のデジタルデータベース群の構築 [「一庫」]、全国の公共図書館におけるデジタル図書館をカバーする仮想ネットワークの形成 [「一网」]、「優秀な中華文化展示に係るプラットフォーム」・「開放型情報サービスプラットフォーム」・「国際文化交流プラットフォーム」の構築 [「三平台」] を指します。また、携帯機器・デジタルテレビ・移動式テレビ等の新興メディアの力を借り、インターネット・移動通信網・ブロードキャストネットワークを介して、政府の立法・政策決定、教育・科学研究、公民の生涯学習等に向けて多層的・多層的・専門的・個性的なデジタル図書館サービスを提供しようとするものでもあります。更に、国家デジタル図書館が各方面で打ち立てた成果の使用を推奨することで、全国を範囲とする各レベル分散型のシステムを構築していきます。ここで言うシステムとは、①「工業化された文献デジタル化加工システム」、②「連携型の保存システム」、③「便利にして簡易な文献検索システム」、④「統一されたサービス調達システム」のことです。資源構築、加工、ストレージ、検索及び調達の面から統一された計画と管理を進め、全国のデジタル図書館をカバーする各レベル分散型の通信プロトコルやサービスシステムを形成していきます。

「工業化された文献デジタル化加工システム」の構築とは、デジタル図書館という大きな枠組みの下で、全面的な計画と統一的な基準に係るシステム構築を実行し、紙媒体文献、

音楽・映像製品、マイクロ化加工製品、有形物等のデジタル化生産ラインを構築して、工業化された生産を実現するものです。

図書館における書誌情報やデジタル資源の取得については、公民の生活習慣に寄り添うようなワンストップ型の検索システムを構築し、ユーザーの検索プロセスにおいて応答する情報の数量や品質を向上させます。

国家デジタル図書館を中心とするデジタル資源サービス調達プラットフォームを構築し、情報資源に係る登録の統一、表示の整合、調達の統一、サービスの整合を実現します。

「連携型の保存システム」の構築とは、全国を範囲とする資源の分散型ストレージを実現し、完備された電子アカウントを構築し、長期保存デジタル資源のオーダリング管理を実現することを指します。同時に、資源の状態とアプリケーションの要件に応じて、オンライン・ニアライン・オフラインの三層ストレージシステムへ資源を配分し、デジタル資源の長期保存と利用を実現します。

国内から条件が満たされている地域を選んでデジタル資源保存センターを建設することにより、全国を範囲とするデジタル資源の長期的な保存という任務を各レベル分散の形で完成させます。これにより、まず国内におけるデジタル資源保存の各種活動を調整し、資金の浪費や保存の漏れを防ぐことができます。その一方で、保存に際してのリスクを効果的に防止し、保存費用を分担することもできます。全国的なデジタル資源保存システムの構築を通じて、各地における図書館の所蔵の中から、中華の悠久な歴史と豊富な文化の内包を十分に反映した希少な資源や、地域の特色ある文化を十分に反映した資料を、広範に探し集めることができます。また、大量のボーンデジタルや非ボーンデジタルの中華優秀文化デジタル資源を収集・ストレージ・長期保存・利用に供し、文化伝承の目的を達成します。

デジタル資源の長期保存事業については、関係する研究分野が広範囲に及んでおり、技術・方法・設備等の問題だけでなく、異なる国家の経済・法律・文化・社会・環境等とも個別に関係を有しています。このことは、デジタル資源の長期保存事業が巨大さと複雑さを備えることを決定づけています。デジタル資源の長期保存を将来的に進めていく中では、以下に掲げる分野の強化に着手する必要があります。①国際か国内かを問わず先進技術を積極的に学習し、実践の中で信頼できる保存方策を継続して追い求めること、②業務担当者の積極性を調達し、関連業務担当者に適切な訓練を施すこと、③フォーマット・規則・規格やフォーマット登録に係るシステムを整備し、資源構築のフォーマットに係る統一な管理を進めること、④法律・法規の研究を重視し、著作権の保護を強化すること。

将来的な発展において国家図書館は、全国の主要な科学研究・教育機関と協調して責任を分かち合い、系統的に分業し、信頼できる管理を実施し、公共サービスを展開することで、デジタル資源の長期保存事業につき持続可能にして効果的な発展をなし得るようにします。対外交流においては、各国の技術・能力と効果的に連結し、長期保存事業の経験をシェアし、情報資源の保存・共同利用を促進することを期しています。