

中国国家図書館の文献資源の統合と表示

中国国家図書館 デジタル資源部主任
曹 寧

現在、IT（情報技術）は、かつてない質的变化を遂げつつあり、社会経済の発展とライフスタイルの転換にとって極めて重要な役割を果たしています。過去30年の図書館の発展を振り返りますと、従来型の図書館からデジタル図書館、さらに複合型図書館へと変わってきましたが、それぞれの変化は技術のバージョンアップ、社会の進歩と密接に関わっています。人類の文化遺産を保存し、知識・情報を伝播する中心基地としての図書館は、現在、非常に重要な発展の時期に来ており、デジタル情報環境の変化に適応し、ユーザの閲覧行為や情報に対するニーズの高まりに対応し、サービスを刷新し、空間を拡張しなければ、図書館の新しい生命力や活力を呼び起こして新しい図書館の業態を作り上げることはできません。

1 前期の取り組み

中国国家図書館は、国の総書庫であり、国の書誌センターであり、国の古籍保存センターであり、国の典籍博物館であり、人類の文化遺産を保存し、知識・情報を伝播する中心基地として、常に文献資源、とりわけ、デジタル資源の構築・管理・利用に注力してきました。中国国家図書館のデジタル資源の構築は1990年代に始まり、情報技術（IT）が進歩しデジタル化が進展するにつれて、ゼロの状態から急速に発展し、デジタル化・ネットワーク化の新しい段階に入りました。デジタル資源をめぐる作成・組織化・保存・公開及びサービスのライフサイクルの全体構造が概ね出来上がり、当館は世界最大の中国語文献保存センターとなりました。2016年1月現在、中国国家図書館のデジタル資源の総量は1181.57TBに達し、類別としては電子テキスト、画像、視聴覚資料、ネットワーク資源等をカバーしています。これには、電子図書74.7万タイトル、電子雑誌5.58万タイトル、電子新聞1.5万タイトル、学位論文約640万件、会議論文約507万件、音声資料約109.9万曲、映像資料約14万時間が含まれます。また、特色あるデジタル化資源は2.6億ページを超えます。海外の中華民国時代の文献のデジタル化、国の貴重な古籍のデジタル化、所蔵音楽映像資料のデジタル化を展開し、インターネット情報の保存と保護を実施しており、保存されたインターネット情報は、既に130TBを超えています。購入データベースは259件ののぼり、中国語データベースが126件、外国語データベースが133件あります。開発した文津検索システムには、3億件以上のメタデータが蓄積され、統一ユーザ管理システムの実名ユーザは755万人を超えており、一意識別子システムによって175万件以上の一意識別子の登録・解析サービスを実現しています。デジタル化資源は現在、インターネット・移動通信ネットワーク・放送ネットワーク等のチャネル、タッチスクリーン・ス

スマートフォン・デジタルテレビ等の端末を通じて伝送され、大量の知識・情報が利用者に提供されています。

資源の種類が増え、チャンネルが多様化したことで、中国国家図書館デジタル資源のメタデータの構成パターンも多様化しました。デジタルインクルージョンはもはや阻むことのできない流れであり、私たちは大きな試練に直面しています。長期にわたるデジタル資源の構築において、私たちは、デジタル資源の作成・加工・記述・組織化・サービス・管理・保存等、各段階の一連の基準を相継いで制定しましたが、メタデータの分散構築と分散管理がリソースの有効利用とサービスの最適化を難しくしています。

2 具体的な実践

「文献情報資源の統合と表示機能の強化」の業務配置に基づき、当館は、2014年から本格的に国家図書館文献資源統合事業を始動しました。

中国国家図書館の文献資源統合事業は、長期計画、持続的な刷新に立脚し、従来型の文献提供サービスとデジタル図書館サービスの高度な融合を推し進め、全国の公共デジタル文化資源の最適配置と情報共有を積極的に推進しています。古典籍、中華民国時代の文献、1949年以降の中国語/外国語文献といった資源の各種メタデータの収集と整理を重点的に行い、メタデータの統一検索機能を強化し、資源を複数のプラットフォームで統一的に表示・公開するサービスを実現しました。多次元的な整理による情報の関連性に基づいた統合的利用を推進し、現代の情報環境におけるサービス機能を発掘・拡張し、それによって包括的で、調和のとれた、持続可能な事業の発展を実現しています。

(一) 超大型メタデータウェアハウスを構築し、文献資源の利用率を向上させる。

メタデータの役割は、資源のキーとなる情報を記述することであり、超大型のメタデータウェアハウスを構築することは、ユーザが資料を発見し、利用する上で重要です。自館構築、購入、募集、収集した各種資源のメタデータを集め、既に2.3億件を超える超大型メタデータウェアハウスが形成され、メタデータの効率良く遅延のない包括的管理が実現されました。購入資源については、受入段階でのメタデータの収集水準と品質コントロールを強化し、メタデータの構築を通常の収集計画に組み入れています。また、メタデータ更新システムを確立し、メタデータ収集の持続可能性を保証しています。段階的に中華民国時代の逐次刊行物を全面調査の範囲に組み入れ、豊富な中華民国時期文献資源のメタデータを超大型メタデータウェアハウスに提供していきます。そして、文津検索システム、ALEPHシステム、公開閲覧サービスシステムのそれぞれにおいて、一意識別子に基づくメタデータの科学的管理を実現しています。

デジタル図書館推進プロジェクトにより、主要商用メタデータウェアハウスプラットフォームとの提携を進め、段階的に全国各レベルの公共図書館のデジタル資源メタデータの共同構築と集中保存管理を行っています。中国国家図書館の主導するデジタル図書館推進プロジェクトは、文化部の三大国民福利事業のひとつで、全国各レベルの公共図書館のデジタル資源のメタデータを対象とした共同構築と集中保存管理を進めているところです。111の図書館が69.37

万件のメタデータの構築を計画し、79 の図書館が 31.42 万件の一意識別子データの構築を計画し、229 の図書館が 293.648 万件の政府情報データの構築を計画し、196 の図書館が 463.165 万ページ分の文献のデジタル化を計画しています。現在、資源の統合と公開プラットフォームの構築を大々的に進めています。プラットフォームにはメタデータを集中させ、対象データは会員図書館に分散して保存する方式を採用しています。2016 年には、各会員図書館の 90 万件以上のメタデータの統合作業を完了させ、推進プロジェクトの成果を段階的に実現し、事業の社会的認知度をさらに高めていきます。

(二) メタデータの典拠の構築を強化し、メタデータの表示・検索機能を高める。

文献資源の統合においては、メタデータの典拠の構築を重点的に強化し、メタデータ生成の規則性・統一性・汎用性・公開性・共有性を確保し、メタデータの一括統合・マルチ利用を実現させる必要があります。個人責任者名称典拠データを構築・整備し、機関典拠メタデータを構築・整備し、タイトル典拠について関連書誌へのリンクを整備し、書誌的関連の表示機能を強化し、逐次刊行物の関連リンクを完了させ、基礎的なデータの指向性と適応性を向上させます。

自館構築・募集・共同構築によるデジタル資源の整備プロジェクトにおいてメタデータの改修と整備を推進し、所蔵デジタル資源の表示・サービスの機能を高めるための基礎固めをします。2016 年末には、200 万件近いメタデータの改修を完了する予定です。この改修では、デジタル資源の特徴および所蔵情報等を記述した管理用項目を大幅に増やし、物理的な所蔵形態の特徴を記述した情報を削除し、また、資源のファイル形式と媒体種別を反映した項目を増やし、学際的特徴を示す情報を増やし、所蔵デジタル資源の公開用アドレスや一意識別子情報を増やすといった業務も含まれます。所蔵デジタル資源メタデータの改修によって、時間軸・地域軸等多次元的な可視化された分類による表示を実現できます。図書館公開講座は、中国国家図書館が主導し全国公共図書館と共同で行っている歴史・文化関係の社会講義です。私たちは現在、質の高い映像資源を利用しながら、積極的にメタデータ統合の実験的事業を推進し、図書館公開講座の Linked Data の加工と公開を行っています。RDF を活用して知識を集積し、静的ウェブページの提供、URI アクセス、sparkle クエリ等の方式を通じたデータの公開を実現し、科学的で合理的なナレッジ組織化システムの構築およびスマート情報サービスの提供に向けた基礎固めを行っています。

(三) 資源を統一的に公開・表示するプラットフォームを構築し、ユーザの利用満足度を高める。

資源統合の目的は、ユーザにより良いサービスを提供することです。中国国家図書館は、科学的な情報組織化方式と先進技術を活用した表示手段を通じて、ユーザの利用満足度を高めていきます。

1. マルチメディアポータルサイトの情報サービス機能を高める。

国家図書館のウェブサービスをさらに充実させるべく、現在、サイト項目の統合とレイアウト

トの変更を行っているところです。モバイル端末向けデジタル資源の編成と表示を最適化し、表示機能を向上させます。情報発信の効率性を高め、オンライン、タブレット PC、モバイル端末でのデータの同一表示を実現させ、文津検索システムのモバイル端末版を開発し、様々な主要モバイル端末に自動で適合するようにします。ウェブサイトと館内のその他の業務システム、特に、資源公開閲覧システムとの融合を加速させ、データベース資源が最大限に活用されるようにしていきます。

2. 所蔵デジタル資源の統一表示機能とサービスの質を高める。

当館の自館構築・募集・購入によるデジタル資源について、統一した所蔵デジタル資源公開プラットフォーム上で、システムごとのメタデータの統合と表示を行います。データベースの公開レベルの統合によって、構築済みのデータベースへの統合を実現します。

3. 文津検索システムプラットフォームの統合機能を高める。

文津検索システムの機能を最適化・発展させ、ユーザ向けにカスタマイズされたサービス機能を拡充し、利用満足度を高めていきます。関連関係データベースを構築し、メタデータ間の関連関係に基づく知識集積型の提示を実現させます。また、検索機能の多様化を実現させます。

4. 国家図書館のユーザポータルサイトの統合機能を高める。

「私と国家図書館の物語」、利用者年度閲覧明細、各ユーザに対応した資源のリコメンド機能等を実現させます。ユーザ管理システムの提供する WeChat、微博、QQ 等へログイン機能を実現・統一します。

(四) データ分析技術を利用し、資源統合の持続可能な発展能力を高める。

ユーザ行動解析を強化し、ユーザのニーズを研究し、資源利用者と資源構築者の間の動的双方向関係を確立することで、所蔵デジタル資源構築の計画と意思決定のために、データの根拠を提供します。

現在構築中の中国国家図書館統計分析プラットフォームは、ビッグデータ技術を応用することにより、サイトアクセス統計分析・資源アクセス統計分析・ユーザ行動統計分析を実施することができ、当館の運営状況・資源利用状況・ユーザの習慣と関心を表示することができ、当館の運営および意思決定に有力な参考データと裏付けを提供します。ここでいくつかのデータをご紹介します。2015年現在、国家図書館利用者総数は363万人、このうち、身分証による認証を経た実名利用者は293万人、利用者カードによる利用者が198.8万人で、利用者総数は毎年60万人のペースで増えてきています。利用者は広大な地域に分布していますが、相対的に多いのは華北華中地域の利用者です。また、利用率では、言語・文字分野・文学分野の資源が最高です。教育レベルが高くなるほど、閲覧する資源の種類の種類が少なくなります。なお、前記のビッグデータの管理および分析により、一部の基本的な研究や試行は既の実現しています。

(五) 主題データベースの構築を強化し、資源の社会的認知度と影響力を高める。

社会発展に係る重大かつ注目の問題、重要な学術研究分野および特定ユーザ層の情報ニーズ

について、テーマを際立たせた質の高い主題データベースを構築しています。「一帯一路（シルクロード経済ベルトと21世紀海上シルクロードからなる経済・外交圏構想）」、「大衆起業、万衆創新（大衆による起業、万民によるイノベーション。中国の経済政策のスローガン）」等、特定テーマの資源の統合や、図書鑑賞動画データベースの構築を進め、既に構築した中華民国時期新聞史料データベースでは、当時の歴史や社会状況を客観的に反映でき、かつ、ある程度代表的である中華民国時期の新聞タイトルを厳選しており、タイトル数、ページ数、記事数ともに現在国内で最大の規模を誇っています。

3 次なる事業構想

中国国家図書館の文献資源統合事業は、所蔵文献資源の現状を全面的かつ詳細に把握したうえで、文献資源の特徴と技術的条件を結びつけ、段階的・計画的に実施することにより、統合事業の有効性・科学性・完全性を保証します。構築の規範化と標準化を強化し、サービスのモデルや手段を絶えず最適化し、統合事業の精度を高めていきます。また、文献資源の統合では、特色をさらに際立たせ、特色ある資源を秩序立てて整理し、知識ユニットを統一した規則にのっとり有機的に組み立て上げ、ひとつの完全な知識ネットワークを作り上げます。文献資源の統合では広く長所を取り入れる必要があり、国内外の先進的な統合技術と統合経験から学び、これらを活用し、将来を見据える意識を高め、所蔵資源統合事業の持続可能性と拡張性を保証していきます。

（一）データ面：基礎データの質と機能を全面的に高める。

各種文献資源の書誌作成と統合を強化し、科学的な整理を行い、各種メタデータをより規範化・系統化されたものにし、よりレベルが深く、より粒度が細かい資源の表示のための基礎固めをし、ひいては膨大な所蔵資源の合理的な開発を実現します。

（二）サービス面：コンテンツに基づく多面的な知識ネットワークの構築を推進する。

科学技術との融合を強化し、性質の異なる資源のクラスタリングと再編を通じて、文献レベルでの表示・展開から、コンテンツレベル・知識レベルまで含めた高度なサービスへと転換させ、所蔵資源のワンストップ検索、関連情報の表示、有効利用を促進します。

（三）展開面：資源の発見・表示方式を絶えず充実させる。

可視化技術の応用を強化し、図書館の各端末のユーザの利用満足度を全面的に高め、図書館の所有する資源の発見および伝達プロセスを改善します。それによって、国内の業界のリーダーとして中国国家図書館の公共文化サービスの実力と影響力をさらに高めます。

（四）管理面：ビッグデータ技術を存分に活用して科学的な意思決定を実現する。

資源データ、ユーザデータ、業務データ、管理データの処理と分析を強化し、各種ビッグデータに含まれる情報価値を十分に掘り起こし、中国国家図書館のビッグデータの全貌を系統的

に提示します。それによって、科学的な意思決定の水準を高め、事業の持続的発展を実現します。

デジタル図書館の構築と発展は、漸進的なプロセスです。デジタル図書館にとって、サービスはユーザと情報資源をつなぐ恒久的な絆です。文献資源の統合を研究・模索・実践することで、中国国家図書館は、情報保障の整備・社会教育の発展・知識資源の提供等の面でより大きな役割を果たしていきます。

ご清聴ありがとうございました。