

## 音声・画像とデジタル文献の永久保存

中国国家図書館 副館長

張 雅芳

### 一、はじめに

中国国家図書館（以下、「当館」）は、世界の国立図書館の中で第 5 位となる 2,456 万冊 / 件に上る各種文献を所蔵し、年間の増加数は 60～70 万冊 / 件である。

当館の蔵書構築は非常に早い時期から行われており、種類も多岐に亘り、1500 年以上前  
の手稿や 3000 年以上前の殷墟の甲骨から、技術発展により出現したマイクロ文献、音声・  
画像文献やデジタル文献（機械可読文献及びネットワーク文献）をも所蔵している。

各種文献とりわけ音声・画像文献やデジタル文献の保存と利用は図書館が直面している  
現実的な問題である。我々が注目すべきは、第一に各種媒体の文献、特に非紙媒体文献及  
びその他の肉眼で見ることのできない文献をどのようにして長期保存していくかというこ  
と、第二にどのような保存対策をとって、後世に対して内容、版、フォーマット、操作と  
使用環境の研究や再利用のために求められるものを保障していくかである。

音声・画像文献とデジタル文献の長期保存の問題は国際的にも各種図書館が最も重視し、  
解決のための研究に取り組んでいる難題のひとつである。ここでは主に、当館の音声・画  
像とデジタル文献の構築、蓄積や管理の現状、直面している問題・課題及び長期保存につ  
いての将来構想等を中心に紹介し、意見を交換したい。

### 二、文献保存媒体の変遷

文献媒体はその時代によって、古代文献媒体と現代文献媒体に分けられる。多様な媒体  
を長期保存することは、人類が長く直面してきた課題である。三千年以上に亘って保存さ  
れてきた古代の甲骨は、その記載内容、媒体の形態、筆記法や加工方法の特徴等が完全な  
状態で残されており、現代人が歴史を研究する上での重要な拠り所となっている。しかし、  
古代に文字を記した竹簡や木簡で千年以上に亘るものはごく少数である。媒体の違いによ  
って、文献の保存期間には差が生じることになる。

#### 1. 古代文献媒体

(1) 甲骨：亀甲と獣骨を含む。商代（訳注：殷のこと。BC1600～BC1100）に用いられていた。

金属：最も多いのは銅と錫の合金である青銅であり、鉄や金銀等も少量ではあるが使われていた。先秦時代（訳注：BC221 以前）に使用された。

- (2) 石：石を媒体とする文献には石碑、崖刻が含まれ、秦朝から現代まで用いられている。
- (3) 竹、木：竹あるいは木を細長く切断し、文字を記録したものを竹簡、木簡という。先秦から三国両晋時代まで（訳注：～AD420）広く用いられていた。幅が広い木板を媒体としたものが牘である。
- (4) 帛：絹織物のことである。最初に現れたのは商の時代で、秦・漢代（訳注：BC221～AD220）に流行した。
- (5) 紙：紙は前漢初年、西暦 150 年前後に発明され、後漢の宦官蔡倫が製紙法を改良し、東晋（訳注：265～420）に至ってほぼ竹、木に取って代わった。
- (6) 海外の主要な媒体：古代エジプトで使用されたパピルス、古代インドで使用された貝葉及び中世ヨーロッパで使用された羊皮紙が含まれる。

## 2. 現代文献媒体

- (1) 紙媒体文献：印刷を記録の手段としている文献形式で、現在の主要な文献である。その特長は利用のために特別な設備を必要とせず、持ち運びが可能で、いつでもどこでも読むことができるという点である。しかし、蓄積密度が低く場所をとるため、保存には不便である。
- (2) 感光性媒体文献：撮影を記録手段としている文献形式の一種で、マイクロフィルム、マイクロフィッシュ、マイクロカード等を含み、数的には多くない。体積が小さく収蔵・保存がしやすい、価格が安いという特長があるが、閲覧のために設備が必要となる。
- (3) 磁気・感光性媒体文献：磁気や感光性をもつ材料を媒体として音声・画像等のデータを記録する音声・画像文献フォーマットの一種である。その特長は迅速にアクセスできること、また、音を聞いたり形を見ることができるため理解しやすいことである。
- (4) コンピュータ、ネットワーク文献：コンピュータ技術やネットワーク技術の発展に伴って現れた、コンピュータの処理技術を用いて情報を記録するデジタル文献フォーマットの一種である。その特長は蓄積容量が大きく、迅速かつ多様な検索ができる、また使いやすいことである。デジタル型文献はさらに、磁気ディスク文献、光ディスク文献及びネットワーク文献に分けられる。

音声・画像文献やデジタル文献の保存及び利用の問題は、技術やコストに関する課題のほか、いかに文献の知識性・記録性・物質性という三つの属性を保存し、研究者の多様な利用目的に対応できるかという挑戦がある。

## 三、音声・画像、デジタル資源の構築・保存・管理の現状と課題

### 1. 構築の現状

- (1) 当館は主に納本、自館作成、購入、交換、寄贈等により音声・画像・デジタル文献を

入手している。1987年から書誌データベースの構築を開始し、文献の知識属性を提供することで利用者の便宜を図った。2000年には大規模な自館作成デジタル資源構築と音声・画像資源のデジタル変換を開始するとともに、納本、購入によって電子出版物、国内外のCD-ROMデータベース、ネットワーク型データベース及び図書・雑誌の付録CD-ROMを収集した。2003年にはネットワーク資源の収集を開始し、大量の音声・画像、デジタル文献を蓄積した。

(2) 2004年未までの自館作成のデジタル資源：

書誌データ	357万件
典拠データ	70万件
目次データ	275万件
全文画像データ	8,000万ページ
音声デジタル変換	40余万曲（中国・外国の音楽・歌曲）
画像デジタル変換	1.3万部（国内外の古典名作映画及び経済・芸術・科学技術等の分野のビデオ資料）

(3) 90年代中期から購入や納本によって導入を開始した電子資源の主なもの：

中国語データベース	28種
外国語データベース	102種
中国語電子出版物	2万件
外国語電子出版物	1千件
音声画像文献（録音テープ、録画テープ、レーザーディスク、VCD及びVHD等）	10万件（150,624件/冊）
中国語・外国語図書付属CD-ROM	2万余件

## 2. 保存の現状

(1) 所蔵する紙媒体文献原本の保存及び閲覧者の利便性のため、当館は一部の文献について大規模なデジタル加工作業を進め、CD-ROM34,778枚、磁気テープ227巻を作成した。その他のネットワーク資源やマルチメディア・ナレッジデータベース及び専用デジタル文献はハードディスクに保存している。

(2) 所蔵する音声・画像文献の原本の保存及び閲覧者の利便性のため、当館は音声・画像文献の移行・変換作業を順調に行ない、これらの資料のコンテンツはネットワーク・コンピュータディスクで利用可能である。411,803本の番組（歌曲、漫才、教育関係）をCD-ROM3,367枚に、13,361本の映画をCD-ROM23,906枚に、129本の児童映画をCD-ROM2,146枚に移行変換した。

(3) ベンダーによる種々の制約があり、当館では購入した大量のネットワーク型デジタル文献については、いまだローカルエリアでのバックアップを作成していない。

(4) 膨大なインターネット資源の長期保存の問題を解決するため、当館では2003年よりネ

ネットワーク資源のアクセス及び保存についての研究を開始するとともに、主題ごとの資源採集を進めている。

### 3. 管理の現状

(1) 管理モデルは現在のところ分散管理を採用している。当館では、図書の付属資料としてのデジタル文献の保管及び閲覧サービスは、典蔵閲覧部が主に担当している。デジタル文献、音声・画像文献と雑誌の付属資料の保管及び閲覧サービスは、報刊資料部が主に担当している。古典籍文献のデジタル資料については、善本部が保管を担当し、一般の古典籍や地方誌等当館所蔵資料のデジタル資料については分館が保管を担当している。

(2) 管理の単位は冊あるいは件である。当館では図書付属のデジタル文献、音声・画像文献や CD-ROM の目録作成に MARC を採用しており、冊・件ごとに管理している。しかし、インターネット資源やデータベースの目録規則については、いまなお検討・作成段階である。

### 4. 問題と挑戦

#### (1) 読み取り機器の更新・バージョン変更に伴う問題

以前の録音テープ、ビデオテープ、LD 等の音声・画像文献は、媒体自体の原因及び読み取り機器の更新により、既にそのままでは利用できなくなっている。80 年代後期のデジタル文献も、技術の更新あるいは機器のバージョン変更によってさまざまな問題が生じている。例えば、古い OS でないと動作しない CD-ROM は既に対応する環境がなく、一部の CD-ROM は複製版を作成せず直接利用したために CD-ROM の材質が劣化し利用不可能となっている。また、図書の付録であるデジタル文献のうち大型フロッピーについては既に読み取り機器がない。どのようにしてこうした音声・画像、デジタル文献の知識性・記録性・物質性それぞれの特徴を伝えてゆき、後世の人の研究利用を可能にするかは、これは一刻の猶予も許されない課題である。

#### (2) 保存条件と管理方法に伴う問題と挑戦

2004 年 12 月末までに当館は音声・画像及びデジタル文献 107TB 分をメディア変換によって作成している。しかし、温度、湿度、ph 値、配置方法等の保存条件が要求水準に達していないことに加えて、管理が十分に行き届いていないことなどが原因で、一部の文献は利用不可能な状態になっている。そのため、どのように最適な保存条件を作り最良の管理戦略を採用し、音声・画像・デジタル文献を長期保存するかについて、研究と規範化が必要である。

#### (3) コンテンツ管理に伴う問題と挑戦

当館では大量の物理的実体を有する文献とデジタル文献を所蔵しているが、現在はまだコンテンツによって関連文献と結びつけたり、コンテンツの特徴や属性を反映したりするにいたっていない。どうやって媒体としての管理からコンテンツ管理へと管理水準を高め

ていくのか、計画策定と解決方法についてさらに検討する必要がある。

## 四、永久保存という命題と将来構想

### 1. 保存の範囲

#### (1) 文献そのものの保存

図書館員は所蔵資料である一次文献及び実体のある派生文献を保存するとともに、今後ますます増加するインターネット資源などのヴァーチャルなオンライン資源をも保存しなければならない。

#### (2) 文献の価値属性の保存

「文献」の現代的な意味は「情報と知識を記録したすべての有形の媒体」である。具体的には、文献とは知識、情報を文字・記号・画像・音声等で一定の物理的媒体上に記録した結合体である。文献には知識性・記録性・物質性の三つの属性がある。図書館は文献そのものを保存すると同時にふさわしい保存・管理方策を策定し、文献の価値属性をも永久に保存するようにしなければならない。

### 2. 音声・画像文献とデジタル文献管理の新たな位置付け

(1) 音声・画像、デジタル文献が有する特殊性に基づき、物理単位、コンテンツ単位ごとに管理する方策を樹立することがぜひとも必要である。すなわち、各種の文献を有機的に統合することによって、文献の蓄積と管理、提供とサービスも統合することが可能となる。

(2) 従来型の文献の目録作業については、MARC フォーマットを適用することで問題はない。しかし、音声・画像文献やデジタル文献では相互間の階層と関係について MARC フォーマットを用いて詳しく記述するのは困難である。そのため、膨大で、階層関係が複雑な音声映像文献、電子デジタル文献の目録に関するフォーマットや基準を検討することが必要とされる。当館では既に担当チームを設け、この問題について精緻な検討を行っている。

(3) 音声・画像、電子デジタル文献が有する独特な属性を考慮し、叙述情報、技術情報、管理情報、保存情報等、これらの文献を永久保存するために記述すべき情報を確定することが必要である。これらの情報によって、後世の人が使用環境を再現あるいは擬似環境を構築することで、文献の利用が可能となる。

### 3. データの蓄積と保管

(1) 当館ではデジタル資源の保存について以下の方法を採用している。

利用率の高い発信・サービス型のデジタル資源はハードディスク上に保存し、バックアップは磁気テープを用いている。ハードディスクは読み取りが速いという利点があり、一度に大勢のユーザーが使用するのに適しているため、発信・サービス型デジタル資料は八

ードディスク上に保存している。

永久保存型の音声・画像、デジタル資源は磁気テープと CD-ROM に蓄積している。磁気テープはメモリ容量が大きく（200G / 巻あるいは 400G / 巻）コストが低いという利点があるが、読み取り速度が遅いという欠点がある。CD-ROM は保存しやすくコストが低いという利点があるが、容量が小さいという欠点がある。このため、磁気テープと CD-ROM はバックアップ用に適している。

(2) 磁気テープ及び CD-ROM のオフライン保存条件と媒体の寿命は密接に関連している。当館では第二期計画及びデジタル図書館プロジェクトにおいて、特にデジタル資源蓄積管理センターを設立し、デジタル文献の長期保存のために必要な条件を研究し、保存条件を段階的に改善する予定である。

(3) OAIS 標準に準拠したオープンアーカイブ情報システムを構築し、デジタル資源の保存と管理を行う予定である。現在は OAIS 標準についての調査・研究を行うとともに、デジタル資源長期保存構造モデルの担当チームを設け、デジタル資源の計画、加工、検収、保存対策、保存用情報パッケージ等の観点から当館の資源の実体に即した研究を行い、最終的には保存のための枠組みを確定する予定である。

(4) 電子帳簿、磁気テープ、CD-ROM の検査登録制度の構築や磁気テープ及び CD-ROM の複製制度の構築といった、デジタル資源の健全な管理制度を構築、整備していく。

#### 4. マイグレーション

(1) 当館では音声・画像文献についてマイグレーション作業が比較的順調に進んでいる。これらのコンテンツは、ネットワーク・コンピュータ及びコンピュータディスクを通じて利用できる。

(2) 購入デジタル文献中の低容量媒体の CD-ROM については、高容量の媒体へ移し替える。デジタル図書館システム完成後は、磁気ディスクに適したものは磁気ディスクへの移行を進め、利用も磁気ディスクを用い、バックアップには磁気テープを用いる。そのほかは、CD-ROM に複製し、利用に供する。

(3) 利用のために専用の電子的機器を必要とする音声・画像文献については、知識の利用・活用という観点から、既にコンピュータネットワークで利用可能なデジタル文献に変換している。同様に、大量のマイクロ文献も電子デジタル文献に変換していく。

(4) 当館ではすでに 2005～2007 年のデジタル資源構築計画を策定している。引き続き従来型の文献の大規模なデジタル化を進め、デジタル資源で利用者サービスをまかない、原本の汚れや破損を減らす。

デジタル資源の保存フォーマットは媒体によって異なる。従来の紙文献のデジタル化は TIFF 画像を保存フォーマットとしている。TIFF 画像は文献の真正性を保証することができるので、コンピュータの使用文字と文献に使用されている文字が異なるために文献が本来のものと違ってしまうことも避けられる。音声・画像文献は一定の圧縮率の保存フ

フォーマットを採用し、圧縮時に多くの情報が損失し文献を再現できなくなってしまうようにしている。

(5) 当館では状態が悪い従来型の文献に原本保護のためのマイクロ化、デジタル化を進めており、実際の破損程度を勘案して段階的に取り組みを進めている。例えば、清末民国初期の一部の文献は破損状態が深刻で、この段階の文献のデジタル化作業を強化する一方、画像と原本の保存と利用の組み合わせという観点から多くの作業をし、徐々にデジタル化資源で利用者サービスをまかなうことができるようになった。ポーンデジタルの音声・画像資料や電子文献はマイグレーションを行う必要があり、同時に大量に所蔵する伝統的文献についても同様の移行・変換作業が必要となっている。

(6) 当館ではベンダーが提供するネットワーク資源や、インターネット及びオープンアクセスの各種ネットワーク資源について、目録規則、資源のカプセル化フォーマット、収集方針の3点を中心に系統的な研究をすでに進めており、第二期計画及びデジタル図書館プロジェクトにおいて段階的に実行していく予定である。

## 五、公開と協力

(1) 保存媒体の選択については各方面の技術協力を求め、長期間にわたり保存していく過程で保存媒体の変化あるいは破損によって文献が利用できなくなる問題を解決する。同時に、媒体の長期安定性と大容量性を研究し、コストがさらに安く、保存期間がより長期にできる媒体を追求する。

(2) 国内外で協力し、OAIS 標準の概念モデルを実際のプロジェクトに速やかに用いることができるようにする。主に OAIS モデルのサブ構造において、実装の形式と方法、中国の各種図書館における汎用性などについて協力していく。

(3) 積極的に国際標準の研究と制定活動に参加し、国際標準を中国に適用する際に、中国語情報や中国図書館の特性により誤差が生じないようにする。

(4) 世界の主要な国立図書館と技術、標準及びプロジェクトに関する情報交換制度を構築し、各国図書館の研究事業に重複が生じないようにする。