

## NDL デジタルアーカイブシステムの概要

国立国会図書館 関西館  
電子図書館課 研究企画係  
木目沢 司

### 表紙 (スライド 1)

韓国国立中央図書館の皆様、本日は国立国会図書館関西館によろこそおいでくださいました。私は国立国会図書館関西館電子図書館課の木目沢と申します。現在、国立国会図書館関西館では電子情報の収集・保存・提供を担う中核システムとして「NDL デジタルアーカイブシステム」を開発しています。本日はこの「NDL デジタルアーカイブシステム」の概要について、報告させていただきます。

### 目次 (スライド 2)

まず、本日の報告内容ですが、このスライドの 1 から 8 までで NDL デジタルアーカイブシステムについて説明します。それから 9 として、国立国会図書館における電子情報の長期利用保証の調査研究についても、報告させていただきます。

#### 1 . NDL デジタルアーカイブシステムの目的 (スライド 3)

NDL デジタルアーカイブシステムの目的は、3 点あります。

インターネットでは膨大な量の情報が提供されていますが、これらの情報は内容の更新、URL の変更、公開の中止などが生じる不安定な情報資源です。NDL デジタルアーカイブシステムは、こうした電子情報の消失や散逸等を防ぐため、広く電子情報の収集を実施します。

収集・保存した電子資料を後世に残し、将来にわたって利用を保証するために、技術の変化に迅速に対応できる拡張性や、最新の機器への移行を定期的に行える移行容易性を確保する長期保存の仕組みを開発します。

また、当館だけでは解決が困難な課題に対処するため、技術標準の整備等の推進にあたっては、他システム等との連携や協働を図ります。

## 2. システムの全体像 (スライド4)

NDL デジタルアーカイブシステムの全体イメージをお示しします。

すでに当館ではインターネット情報選択的蓄積事業 (WARP ; Web ARchiving Project) により、ウェブサイトサイトをサイトの形式のままで収集し、提供しています。NDL デジタルアーカイブシステムは WARP の機能を発展させ、サイト全体を一つの単位として収集・組織化するだけでなく、サイト内の白書や年報、電子雑誌のように、ある程度独立した著作物の形態で切り出せるものについても、個別にメタデータを付与し、組織化、提供する予定です。

次に、当館所蔵資料をデジタル化し、NDL デジタルアーカイブシステムに保存してインターネットからの利用を行えるようにします。当館ではすでに近代デジタルライブラリーや貴重書画像データベース等により、当館所蔵の紙資料をデジタル化してインターネットで提供するサービスを行っていますが、紙資料だけではなく、将来的には、レコードやビデオといったアナログ媒体に記録された録音・映像資料についても、デジタル化し保存・提供することを検討しています。

さらに、デジタルアーカイブポータル (PORTA) を入り口として、国立国会図書館以外の日本国内のデジタルアーカイブ機関で保存・提供しているコンテンツについても統合検索を可能とし、広く国民にサービスを提供することを目指しています。

## 3. システムの構成 (スライド5)

NDL デジタルアーカイブシステムのシステム構成についてご説明します。

NDL デジタルアーカイブシステムは、機能拡張やシステム移行を容易にしてシステムのライフサイクルコストの削減を図るため、全体を「アプリケーション層」、「保存システム層」、「ストレージ層」の3つの階層に分けています。

アプリケーション層は、ユーザーとのインターフェースとして、デジタルアーカイブポータル (PORTA) によって、利用者へのサービスを提供します。また、デジタルアーカイブポータルは、他機関が所有しているデジタルコンテンツのメタデータのハーベストや他機関との横断検索を実現します。

アプリケーション層では、インターネット上の著作物 (Works) や現在の WARP が収集しているウェブサイト (Websites) 、紙資料をデジタル化したコンテンツである近代デジタルライブラリー (Digital Library from the Meiji Era) の機能を統合していきます。アプリケーション層は、NDL デジタルアーカイブシステムが保存対象とするコンテンツ種別の増加や提供技術の変化に柔軟に対応できるように、次に説明する保存システムとは独立した層として設計しています。

保存システム層は、長期保存のための技術標準である OAIS (Open Archival Information System) に準拠し、電子情報の長期保存とアクセスを可能とするための層です。アプリケーション層から長期保存対象とするコンテンツを受け取り、メタデータとともに情報パッ

ケージ化して保存します。また、ストレージ層の機器を制御して、古くなったストレージから新しいストレージへコンテンツの移行を容易に行える仕組みを特定機器ベンダーに依存しない方法で実現します。

ストレージ層は、コンテンツを蓄積し、長期にわたって保存を行う大規模なストレージ装置です。データ量の増加に柔軟に対応するため、ストレージの増設を容易に行える機器構成としています。

またシステムの設計とあわせ、NDL デジタルアーカイブシステムで長期保存する電子情報の管理や利用のために必要なメタデータスキーマを設計しています。

#### 4 . NDL デジタルアーカイブシステムの業務フロー ( 1 )(スライド6)

NDL デジタルアーカイブシステム完成後の業務概要を代表的な業務フローにより説明します。

業務フローは現在当館が運用中の WARP 及び近代デジタルライブラリーの課題を踏まえ、業務ステップの簡素化や機能の共通化を図っています。特にシステムに登録するデータは、選定や許諾の業務で一度入力したデータを次の業務ステップで重複入力することがないようにし、合理化を図ります。またメタデータについては選定や許諾作業で入力した業務データから流用する仕組みになる予定です。

このスライドはインターネット情報を収集ロボットにより自動収集する際の業務フローを示しています。

まず「選定」で、収集対象の「タイトル」,「URL」,「公開者等」のウェブサイト特有のデータを登録します。

著作権などの問題に対処するため、「許諾」では、「選定」のデータを引継ぎ、許諾依頼書を作成し、サイトを公開している責任者に紙で送付します。返送された許諾依頼書に従い、収集除外対象やアクセス制限、すなわちインターネット公開か館内公開のみか等の許諾条件の設定を行います。この許諾条件は次の収集条件設定に引き継がれます。

「収集」では、初回収集時のみ収集頻度や収集速度等を設定し、以後はスケジューリングによって定期的に収集ロボットによる自動収集を行います。収集結果の確認は、リンク切れのチェック等可能な限り自動化する予定です。

「組織化」では、収集までの作業ステップで入力されたデータと、収集したコンテンツからの自動抽出により、メタデータを登録します。メタデータの手動入力が必要な場合のみ行う予定です。「組織化」が完了すると、コンテンツは「受入用情報パッケージ」(SIP)として「保存システム層」に渡され保存されます。保存システム層では、アプリケーション層から受け取った SIP に長期保存のための技術メタデータなどを自動的に追加して「保存用情報パッケージ」(AIP)に変換します。この AIP は電子書庫に保存され、「提供」が可能になります。

なお、インターネット情報の収集にあたっては、このフロー以外に、提供者からのコン

コンテンツの送信または媒体送付による収集や、OAI-PMH 等でハーベストしたメタデータから URL を抽出して収集ロボットによる収集方法もあります。

#### 4. NDL デジタルアーカイブシステムの業務フロー (2) (スライド7)

次のスライドは、近代デジタルライブラリーのように紙資料をデジタル化する場合の業務フローです。

「選定」では、電子化対象資料のタイトル、著者等を当館 NDL-OPAC の書誌データを元に電子化対象リストに登録します。

「電子化」では、電子化作業に必要な書誌データのリスト出力、電子化コンテンツ格納媒体の登録を行います。

「組織化」でのメタデータの登録は、「選定」で設定した書誌データを流用します。組織化完了後の流れは、インターネット情報収集と同様です。

なお、コンテンツのデジタル化と公開のための著作権処理も別途行います。

#### 5. 収集ロボット (スライド8)

NDL デジタルアーカイブシステムでは、インターネット情報の「収集ロボット」として、国際連携も考慮し、IIPC (International Internet Preservation Consortium) が開発を推進している「Heritrix」を採用しました。昨年度は、ソースコードの解析や現行 WARP の収集ロボットである wget との性能比較調査を実施しました。

調査の結果、Heritrix は、収集範囲の制御、ロボット排除指定の遵守、収集対象サイトへの負荷防止、リンクの抽出、収集データのディスクへの保存等の NDL デジタルアーカイブシステムの主要な機能を備えていることを確認しました。

一方、NDL デジタルアーカイブとして追加開発すべき部分として、国際化対応の強化、つまりマルチバイト文字への対応があります。

この昨年度の調査結果は、IIPC のメンバーであり、Heritrix 開発の中心であるインターネットアーカイブに報告する予定です。Heritrix の追加開発及び開発したモジュールのオープンソース化にあたり、今後さらに IIPC 及びインターネットアーカイブと連携を図っていく予定です。

#### 6. メタデータ (スライド9)

NDL デジタルアーカイブシステムでは、コンテンツとそのメタデータを OAIS 参照モデルに準拠した情報パッケージとして保存します。

NDL デジタルアーカイブシステムでは、情報パッケージ自身を記述するメタデータスキーマとして、デジタルコンテンツの保管と交換のためのメタデータ記述の標準である METS (Metadata Exchange and Transmission Standards) を使用します。

コンテンツの目録に相当する情報、すなわちタイトルや著者、出版者といった情報を記

述する記述メタデータスキーマには、記述性、拡張性、相互運用性といった観点から MODS ( Metadata Object Description Schema ) の要素および記述方法を使用することとしました。

CPU、ハードウェア、周辺機器、OS、アプリケーションプログラム等、保存対象のコンテンツを再生するために必要な技術要素を記述する技術メタデータスキーマ、保存対象のコンテンツの利用について規定する情報を記述する権利メタデータスキーマ、コンテンツが作られた理由やプロセス、マイグレーション履歴などを記述する保存メタデータスキーマは PREMIS に従うこととしました。

NDL デジタルアーカイブシステムでは、収集したインターネット情報やデジタル化コンテンツに対し、これら国際標準に準拠したメタデータを付与して、デジタル情報の永続的な保存を実現しようと考えています。

また将来的には、この NDL デジタルアーカイブシステムのメタデータを公開し、国内外のデジタルアーカイブ機関とのメタデータ交換等の相互運用を実現することも視野に入れています。

## 7. 保存システム(1) (スライド10)

NDL デジタルアーカイブシステムは、OAIS に準拠したシステムとして設計しています。OAIS に定められている、保存計画 ( Preservation Planning ) 受入 ( Ingest ) データ管理 ( Data Management ) アクセス ( Access )、データ管理・保存 ( Archival Storage ) システム管理 ( Administration ) の各機能を、主に保存システム層によって実現しようとしています。

## 7. 保存システム(2) (スライド11)

NDL デジタルアーカイブシステムにおける、保存システムの仕組みを少し詳しくご説明します。

まずコンテンツは、アプリケーション層によって収集、組織化が行われます。アプリケーション層は、コンテンツに記述メタデータを付与し、受入用情報パッケージ ( SIP ) として、保存システム層に投入されます。保存システム層では、受入 ( Ingest ) 機能により SIP を保存用情報パッケージ ( AIP ) に変換します。AIP には、長期保存に必要な再生環境に関する技術メタデータが自動付与されます。また Ingest は、メタデータをデータ管理機能 ( Data Management ) に渡します。Data Management はメタデータをデータベースに登録します。Ingest はさらに、AIP をデータ管理・保存機能 ( Archival Storage ) に渡します。Archival Storage は、ストレージ層の電子書庫に AIP を保存します。

さて、利用者からのコンテンツの検索や参照は、アプリケーション層を經由して、保存システム層のアクセス ( Access ) 機能で受け付けられます。利用者からの検索要求に対して Access は、Data Management に検索要求を行い、その結果をアプリケーション層に渡

します。アプリケーション層で検索結果がリスト形式あるいは、詳細な書誌データとして表示されます。利用者が、参照したいコンテンツをアプリケーション層に指示すると、保存システム層の Access 機能は、Archival Storage 機能を経由して参照したいコンテンツが含まれる AIP からコンテンツをアプリケーション層に返します。

なお、コンテンツは参照だけではなく、提供用情報パッケージ (DIP) としても提供可能とする予定です。

また、長期的な保存計画やシステム管理機能は、それぞれ、Preservation Planning と Administration 機能によって実現し、ストレージの状態を監視したり、AIP のマイグレーションを可能とする予定です。

## **8 . 開発スケジュール (スライド 1 2 )**

NDL デジタルアーカイブシステムの開発スケジュールについて、簡単にご説明します。2005 年度は概要設計を行いました。

昨年度はシステムの基本設計と一部詳細設計、及びメタデータスキーマの設計を行いました。

今年度は、詳細設計の残りとプログラム作成などの製造を行い、来年度、テストフェーズに入る予定です。

そして 2009 年度に NDL デジタルアーカイブシステムの運用を開始する予定です。

## **9 . 電子情報の長期利用保証調査 ( 1 ) (スライド 1 3 )**

次に、NDL デジタルアーカイブシステムの開発と平行して行っている国立国会図書館における「電子情報の長期利用保証調査」について説明します。

当館では、「電子情報の長期保存に係る調査研究」として、2002 年度には主に諸外国の動向や、電子情報の長期保存の参照モデルである OAIS の文献調査を行いました。

2003 年度から 2004 年度にかけては当館所蔵の「パッケージ系電子出版物」の再生可能性のサンプル調査を行いました。ここでパッケージ系電子出版物とは、フロッピーディスクや CD-ROM などの電子媒体を主とした出版物です。2003 年度は、PC で動作することが前提の 5.25 インチフロッピーディスク、3.5 インチフロッピーディスク、CD-ROM 資料 200 点について、調査当時の最新動作環境である WindowsXP 上での起動と簡単な動作確認を行いました。その結果、約 7 割の資料の利用に問題があることが判明しました。問題の理由のうち、OS の不適合によるものが最も多く、次に再生アプリケーションソフトウェアの入手不可及び再生アプリケーションソフトウェアとそのプラグインの不適合のため、残りは 5.25 インチ FD のドライブ入手ができなかったためでした。

2004 年度には、旧式 OS で稼働することが前提のプログラムを含む電子資料 100 点について、市販エミュレーションソフトにより、WindowsXP の上で MS-DOS などの旧式 OS 環境を再現して正常に再生するかの調査を行いました。しかし、インストールや起動不可

などが 16 点、音声や画像が正常に再生できないなどの一部不完全なものが 55 点あり、正常に再生したものは 29 点にすぎないことがわかりました。

このように、電子資料については再生技術の変化により、利用不可となる資料が増えてきており、現状のエミュレーション技術だけでは、長期的な再生可能性の保証が難しいことがわかりました。

そこで、国立国会図書館では、電子情報の長期利用保証について、2006 年度から本格的に取り組むを始めることになりました。

なお、2003 年度、2004 年度の調査内容の詳細は、“The Long-term Accessibility of Packaged Digital Publications. (NDL Research Report: No.6)”として刊行しておりますので、ご興味ございましたら、こちらをご覧ください。

## 9. 電子情報の長期利用保証調査(2)(スライド14)

当電子図書館課では、2006 年度に「電子情報の長期利用保証に関する調査研究実施計画」を策定しました。この実施計画では 4 つの目的を掲げました。

1 つ目は、マイグレーション及びエミュレーションに関する調査研究を行うことです。記録媒体、ファイル形式等の旧式化により再生が困難となりつつある電子情報の媒体移行、ファイル形式変換によるマイグレーション、及び再生環境の擬似的再現であるエミュレーションに関する調査研究を行います。

2 つ目は、レコードやビデオテープなどの媒体劣化や再生機器の入手が困難になりつつある旧式録音資料・映像資料のデジタル化に関する調査研究です。旧式録音資料・映像資料のデジタル化及びデジタル化後のデータの長期利用保証に関する調査研究を行い、今後実施するデジタル化実施計画に反映しようとするものです。

3 つ目は、国内外の動向を踏まえ、当館における電子情報の長期保存に関するガイドラインを策定・公表し、当館及び関係機関等における電子情報保存の推進に役立てようとするものです。

4 つ目は、調査研究の成果を国内外に広報し、関係機関等との情報の共有や連携・共同の推進を行おうとするものです。

2006 年度は、旧式電子資料のマイグレーション調査として、近年、再生技術及び再生環境の旧式化が懸念されているフロッピーディスク資料について、マイグレーションの調査を行いました。

当館所蔵の、PC-98 シリーズや MS-DOS などの 3.5 インチ及び 5.25 インチの旧式 OS 用フロッピーディスクからサンプル 100 点(216 枚)を選定し、市販のマイグレーションツールにより、フロッピーディスクから WindowsXP が稼働する PC へ、マイグレーションを実施しました。

その結果、3.5 インチフロッピーディスクについては、特殊なものを除き、マイグレーションが容易であることが分かりましたが、5.25 インチフロッピーディスクについては、2D という規格の資料が相当数存在していることが分かりました。2D 規格のフロッピーディスクには、これに対応したドライブが必要でマイグレーションができないことが分かりました。

また 5.25 インチフロッピーディスクについても、ディスクドライブ自体の入手が困難になってきており、早急に媒体移行によるマイグレーションを実施する必要があると思われる。

2007 年度以降のスケジュールですが、旧式電子資料については、昨年度に引き続きフロッピーディスクのマイグレーション調査を続けます。また、エミュレーションの技術調査を開始します。

さらにこの調査研究では、マイグレーション・エミュレーション機能を含む長期利用保証システムのプロトタイプシステムも作成しています。これは、将来 NDL デジタルアーカイブシステムに電子資料の保存と利用の機能を組み込む場合にどのような技術課題があるのかを検証するためのものです。

レコードやビデオテープなど旧式録音・映像資料についても、アナログからデジタルへ変換するための技術調査とデジタルファイル形式の規格の調査を行っています。

長期保存のガイドラインは、旧式電子資料の技術調査と、旧式録音・映像資料のデジタル化調査の結果を反映したものを検討し、2010 年度までに策定する予定です。

## まとめ（スライド 15）

以上、NDL デジタルアーカイブシステムの概要について説明させていただきました。これを機会に、今後は、韓国国立中央図書館様とより実務的な交流をしていきたいと思えます。

それにより NDL デジタルアーカイブシステムに関しては、メタデータの標準化やマルチバイト文字などの多言語対応、その他、収集・組織化・保存・提供に関する技術的あるいは制度的な課題について、お互いに情報交換ができるのではないかと思います。

また、電子情報の長期利用保証調査に関しては、マイグレーションやエミュレーションに関する技術、長期利用のための再生環境情報の管理方法などについて、情報交換ができるのではないかと思います。

ご静聴ありがとうございました。