

## 国立国会図書館のデジタルアーカイブシステムとポータル

国立国会図書館関西館電子図書館課長  
佐藤 毅彦

### 1. はじめに

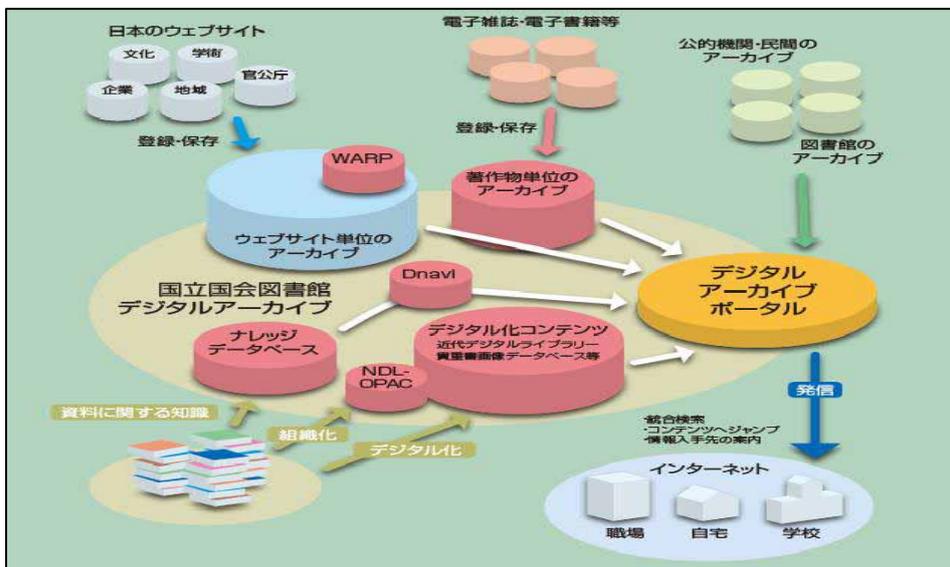
当館は「電子図書館中期計画 2004」で、電子図書館事業の目標として、(1)国のデジタルアーカイブの重要な拠点となること、(2)日本のデジタル情報全体へのナビゲーションの総合サイトを構築すること、を掲げています。

「デジタルアーカイブの重要な拠点」となるために、デジタル情報を収集・組織化・保存するシステムとして、当館は国立国会図書館デジタルアーカイブシステム(以下「DAシステム」という。)を構築しています。2005年度から開発を開始し2009年度中に運用を開始する予定です。

また、「日本のデジタル情報全体へのナビゲーション」を実現するために、当館は「国立国会図書館デジタルアーカイブポータル(PORTA)」を構築し、10月から公開しています。

本日は、この2つのシステムの内容を中心に報告します。

最初に、昨年もお示したのですが、電子図書館サービスの全体像を示します。



DAシステムでは、インターネット情報(ウェブサイト単位及び著作物単位)デジタル化コンテンツ、パッケージ系電子出版物といった各種のコンテンツを収集・保存します。DAシステムで保存するコンテンツのほか、国内の他の図書館などで保存・提供されている電子情報は、PORTAを入り口としてワンストップで統合的に検索、利用できるような仕組み

みを、当館は目指しています。

## 2. DA システム

### 2.1 DA システムの目標

DA システムの開発に当たっては、次の 3 点の実現を目指しています。

1 点目は、電子情報を広く収集・組織化・保存・提供することです。

DA システムでは、インターネット情報、紙資料などからデジタル化した情報及びパッケージ系電子出版物を広く収集し、組織化、保存、提供します。

2 点目は、電子情報を長期保存し、かつ将来にわたって利用保証することです。DA システムでは、将来にわたっての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柔軟性を確保した長期保存の仕組みを開発します。

3 点目は、当館だけでは解決が困難な課題への対処、技術標準の整備等の推進にあたって、当館以外の組織・システムとの連携や協働を実現するための柔軟性及び相互運用性を確保することです。

### 2.2 DA システムの構築

#### (1) 既存システムと DA システム

現在提供している電子図書館サービスは、システム面で二つの問題を抱えています。

一つは、類似のシステムが複数存在していることです。業務・システムの統合・連携を図り、利用者本位の簡素で効率的なサービスを実現する必要があります。

もう一つは、既存のシステムでは、電子情報の長期保存や長期利用保証が考慮されていないことです。

これらの問題点を解決するため、DA システムでは、関連するシステムを順次統合するとともに、電子情報の長期保存・長期利用保証を実現する仕組みを確立します。

#### (2) DA システムの特徴

DA システムの構造は、コンテンツの収集、提供等を行う「アプリケーション層」、アプリケーション層から受け取ったコンテンツの保存、管理を行う「保存システム層」、コンテンツを蓄積する「ストレージ層(システムインフラ)」の 3 層構造とします。各層間のインターフェースを標準化するとともに、可能な限り各層の独立性を高めることにより、各層ごとの機能拡張やシステム統合を容易にするためです。

##### アプリケーション層

アプリケーション層には、ウェブサイト(Websites)を収集している WARP、紙資料のデジタル化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)、深層情報(Deep Web)の入り口まで利用者を案内する Dnavi を統合していきます。

インターネット情報の「収集ロボット」として、DA システムは、International Internet Preservation Consortium (IIPC)が開発を推進している「Heritrix」を採用しました。

また、組織化機能として、膨大な量のメタデータ付与を自動的に行う機能を実装するための検討を行っています。

アプリケーション層から利用者へのサービス提供は、PORTA によって行います。

#### 保存システム層

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。保存システム層では、アプリケーション層から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の形態で保存します。

#### ストレージ層

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仕組みを備える想定です。

### 3. 国立国会図書館デジタルアーカイブポータル (PORTA)

#### 3.1 PORTA の目的

PORTA は、我が国の電子情報資源や情報提供サービスに案内する総合的なポータルサイトとして、当館だけでなく、日本国内のあらゆるデジタルアーカイブ機関が保存・提供しているコンテンツ及びサービスにワンストップで案内することを目指しています。

#### 3.2 PORTA の利用機能

PORTA では、連携する複数のデジタルアーカイブや紙資料の目録、各種参考情報を統合的に検索することができます。

##### (1) 検索機能

PORTA では、様々な特徴を持つ複数の検索方法を提供しています。

##### 簡易検索

利用者が入力したキーワードでメタデータの全文検索を実施します。

##### 詳細検索

簡易検索と同様、キーワードによる検索ですが、詳細なオプションが設定できます。

##### 連想検索

キーワードと異なり、自然文、フレーズからの連想的検索を行います。

##### 分類による検索

日本十進分類法 (NDC) または国立国会図書館分類表 (NDLC) のカテゴリをたど

りながら、データを検索します。

#### 辞書検索

当館の典拠データを用いた辞書データベースによる検索です。統制語の数は約 84 万語です。結果を文字列で表示するだけでなく、統制語を中心に、上位語、下位語、関連語、類義語がどのような相関をもっているかを視覚的に表示することができます。

その他、Wikipedia で該当する語がある場合はそのリンクも表示します。

#### (2) 関連情報リンク

PORTA には、検索結果のデータを表示するだけでなく、さらにそのデータから関連する情報源へナビゲートする機能が備わっています。検索でヒットしたメタデータの内容情報を用いて、総合目録を検索して他の図書館の所蔵状況を調べたり、書籍販売サイトから本を注文したりすることができます。また、インターネット上の関連サイトへの遷移も可能です。

#### (3) カスタマイゼーション機能

PORTA では、ユーザ登録を行うことにより、個別ユーザに合わせた設定やデザイン等のカスタマイズを可能にしています。

### 3.3 システム連携のための機能

#### (1) データの収集等のための機能

PORTA の検索対象となるデータは、OAI-PMH や RSS によりメタデータを収集（ハーベスト）し、あるいは、SRW(SRU/SOAP)や Z39.50 により横断検索されます。

#### (2) データ提供のための連携機能

PORTA はまた、外部のシステムが PORTA の検索機能を系統的に利用することを可能とするために、外部提供インターフェースの API(Application Programming Interface)を提供します。これにより、外部システムが PORTA の機能を利用し、さらに付加価値をつけたサービスを提供することが可能になります。

### 3.4 PORTA の課題：検索対象の拡大

PORTA の検索対象となっているデジタルアーカイブは、現在のところ、当館のアーカイブを除くと 8 種類にとどまっていますので、今後の協力機関の拡充が大きな課題です。

#### (1) システム連携のための共通仕様の実装

メタデータ記述要素、メタデータスキーマ、通信規約を可能な限り共通化するとともに、連携先システムに実装してもらうための働きかけを行っていきます。

また、PORTA との連携のためのガイドライン、メタデータ基準のガイドライン、資料

デジタル化ガイドライン、外部提供インターフェース仕様を整備し、普及に努めます。

## (2) データプロバイダディレクトリへの登録

PORTA は、共通仕様を実装したデータプロバイダとの間で、自動的にシステム連携できる機能を装備しました。この機能の実効的なものにするため、データプロバイダに、システム連携のために必要な情報の登録を呼びかける必要があります。

## 4 . メタデータ

### 4 . 1 DA システムのメタデータ (保存用メタデータ)

2006 年度は、DA システムで保存するコンテンツの管理、利用のために必要なメタデータスキーマを設計しました。

DA システムのメタデータは、長期的な可読性、相互運用性を確保するため、XML ベースのメタデータスキーマとし、国際的に広く普及している標準的なメタデータスキーマを利用して設計することとしました。また、将来的なコンテンツの種類に対応するため、拡張性と柔軟性を備えたメタデータスキーマとなるよう配慮しました。

メタデータの構成は次のとおりです。

情報パッケージのメタデータ

メタデータスキーマには、METS( Metadata Exchange and Transmission Standards ) を使用しています。

記述メタデータ

書誌情報に相当します。MODS ( Metadata Object Description Schema ) を使用します。

技術メタデータ

保存対象コンテンツを再生するために必要な技術要素を記述します。

権利メタデータ

保存対象コンテンツの利用関係情報を記述します。

保存メタデータ

コンテンツの作成やマイグレーションの履歴などを記述します。

管理メタデータ

インターネット情報の収集関係情報やメタデータの更新・登録情報を記述します。

技術、権利及び保存メタデータについては、PREMIS(PREservation Metadata: Implementation Strategies)に独自要素を追加し、管理メタデータについては、独自設計しました。

貴館にもお知らせしているとおり、当館は、これらのメタデータスキーマ (NDL-DA メタデータスキーマ) を、7月に館ホームページで公開し、意見を募っています。いただいた

意見を参考に、今年度中に DA システムのメタデータスキーマを確定し、システムに実装する予定です。

#### 4.2 交換用メタデータ

記述メタデータについては、DA システムのメタデータのほかに、ダブリンコアに基づく「国立国会図書館ダブリンコアメタデータ記述要素 (DC-NDL)」を策定し、メタデータの交換を円滑に行うことを想定しています。

DA システムでは、ダブリンコアと MODS とのマッピングが自動的に行える仕組みも実装する予定です。

#### 4.3 国際連携

当館は、将来的には、メタデータスキーマを可能な限り共通化し、メタデータの交換や、さらには、情報パッケージによる交換を実現したいと考えています。今回の業務交流では、将来の連携に向けた情報交換及び意見交換を行いたいと考えています。

### 5. 電子情報の長期利用保証のための取組み

一昨年業務交流では、当館が 2002 年度から 2004 年度に実施した電子情報の長期利用保証調査について報告しました。

その調査結果をふまえ、当館は 2006 年度に「電子情報の長期利用保証に係る調査研究に関する実施計画」を策定しました。

この計画の目的は次のとおりです。

フロッピーディスクを中心とした旧式電子情報について、マイグレーション (媒体移行・ファイル形式変換) 及びエミュレーション (再生環境の擬似的再現) の調査を行い、DA システム開発に反映する。

録音・映像資料について調査し、今後のデジタル化実施計画に反映する。

ガイドラインを策定、公表し、電子情報長期保存の推進に資する。

国内外にも調査成果を広報し、関係機関との情報共有や連携・協働の推進に資する。

DA システムへの長期利用保証システムの実装は 2010 年度を想定しています。

旧式電子資料については、昨年度から今年度にかけて、マイグレーションに係る技術的課題の調査を行っています。

旧式録音・映像資料については、昨年度から、媒体の規格、再生機器の技術要件とその入手可能性などの基礎的な調査を行っており、来年度、録音・映像資料デジタル化の手引きを作成する計画です。

電子情報の長期利用保証について、貴国は国際的な連携体制を構築していると伺ってい

ます。今回の業務交流では、貴館における取組みの状況を伺うとともに、国際的な連携のあり方についても意見交換を行いたいと考えています。

## **6 . おわりに**

電子図書館のシステム構築、業務運営の様々な局面で、国際的な連携が必要であり、実際の連携も着実に進んで来ています。

当館の DA システムは今まさに開発中であり、ポータルシステムはサービスが開始されたばかりの状況です。今回の業務交流では、貴図書館の経験についてお話を伺い、情報や意見を交換し、当館の DA システム開発やポータルサービスの参考にさせていただきたいと思っています。