

## 国立国会図書館デジタルアーカイブポータル概要 我が国のデジタルコレクションの窓口を目指して

国立国会図書館 総務部 電子情報企画室  
久古 聡美

当館におけるデジタルアーカイブに関する計画の概要と、その中でのデジタルアーカイブポータル(PORTA)の目的と位置づけを紹介する。その後構築した PORTA の機能概要を紹介し、最後に PORTA と外部システムの連携方法、及び今後の予定に触れる。

### 国立国会図書館と電子情報

近年、情報ネットワークの急速な発展によって、社会における情報流通のあり方が大きく変化しつつある。とりわけインターネットは、情報環境、社会環境に大きな影響を及ぼしている。こうした情勢を踏まえて、当館では、インターネットを積極的に活用し、情報発信の強化、図書館サービスの強化・拡充に取り組んできた。

国立国会図書館が 2004 年度に策定した「国立国会図書館ビジョン 2004」では、その使命と役割として、「国民の知的活動の成果を、印刷物から電子情報に至るまで広く収集し、国民共有の情報資源を構築すること」、「国会の立法活動を補佐すること」、「現在及び将来に渡り、情報資源へのアクセスを保証すること」を謳っている。また、当面、「立法補佐機能の強化」、「協力事業の推進」と並んで、「デジタルアーカイブの構築」、「情報資源へのアクセスの向上」に重点を置くことにした。

「電子図書館中期計画 2004」は 5 年間を目途とした計画だが、この計画の中で、電子図書館サービスの計画として、次のように定めている。

「デジタル・コンテンツを広範な利用者に提供するために、当館は国のデジタル・アーカイブの重要な拠点となる。また、国内外の多様な利用者層の需要に応じ、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。」

本日主にお話するのは、このうち、後者の「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」についてである。

### 「電子図書館サービス」の目標イメージ

当館の電子図書館サービスの目標イメージを示す。

デジタル情報を長期的に保存し、将来にわたり利用可能とすることを目的として、NDL デジタルアーカイブを構築することとしている。ウェブサイトをはじめとして、画像やテキスト等の形で電子化された図書資料、CD、DVD といったパッケージ系電子出版物など、広くデジタルコンテンツを蓄積し、永続的に保存、提供することを目指している。

このデジタルアーカイブの構築には、ウェブアーカイブ、デジタルデポジット、デジタル化の主に3つの要素がある。

ウェブアーカイブ(web archive)は、日本のウェブサイトキャプチャ(capture)して保存しておく。ウェブサイトは非常に消失しやすいもので、将来に渡って情報へのアクセスを確保することが必要となっている。

デジタルデポジット(digital deposit)は、訳すなら電子的納本だが、これまで当館が紙媒体やCD-ROM などのパッケージ物を納本対象としていたのと同様に電子雑誌・電子書籍等もデポジット(納本)してもらう枠組みである。

デジタル化(digitization)は、従来の紙媒体、マイクロの資料等の所蔵資料のデジタル化を行い、デジタルアーカイブとして保存するとともに、インターネットを通じて外部提供を行う。

一方、デジタルアーカイブのポータルは、これらの当館で構築したデジタルアーカイブと共に、外部の公的機関、民間機関、図書館のアーカイブ等の有用な情報をも対象に、ナショナルデジタルコレクションの窓口として、統合的にナビゲーションを行う役割を果たすことを目標としている。

## デジタルアーカイブポータルの目的

そもそも、「デジタルアーカイブ」は訳すとデジタル資料の保管所、「ポータル」はインターネットの中では、ウェブサイトで何かを探したいと考えた時に情報を探す起点となる場所だといえることができる。

当館の目指すデジタルアーカイブポータルは、当館が保有しているコンテンツに加えて、国の機関、民間に関わらず、広く有用なコンテンツおよびサービスへ案内するポータルサイトである。その機能の柱として、以下の2つがある。

### ワンストップナビゲーション(One Stop Navigation)

1つの検索窓から多数のアーカイブを統合的に検索可能とし、また、デジタル情報はその閲覧まで、紙資料はその入手先までナビゲーションを行う。

### カスタマイゼーション(Customization)

様々な利用者、利用形態でのニーズに対応する。

## ポータルの構築と公開

ポータル構築の前段として、2004年度からプロトタイプでの検証を行ってきたが、今年度に本格システムとして構築し公開する。本格システムの通称は PORTA である。ラテン語で「門」や「入り口」という意味があるが、デジタルアーカイブへの入り口をイメージした。

URLは<http://porta.ndl.go.jp/>である。

## 機能概要

現時点の PORTA の機能概要について紹介する。

検索対象は現時点で 25 種類、一部は外部公開に至っていないものもある。

メタデータの件数は、横断検索先のシステムでは随時変更され正確には把握できないが、約 1,000 万件以上と推定される。

PORTA の画面には日本語だけでなく、英語への切り替えも可能とした。

データの検索方法は、いずれもデータに付与されたメタデータの検索を行うものだが、特徴の違う複数の方法を検索提供する。キーワード検索、キーワードマッチングではなく概念からの検索を可能とする連想検索、分類検索として NDC(日本十進分類法)と NDLC(国立国会図書館分類表)の体系を用いたブラウザ検索がある。

その他、当館の典拠データを用いた辞書データベースの検索の仕組みと、キーワード検索時に辞書データも同時に検索してメタデータの検索を補助する検索支援機能がある。

## 様々な検索方法を提供

検索方法について、画面イメージを用いながら簡単に紹介する。

### 簡易検索

トップページに配置。キーワード検索であり、ユーザが入力したキーワードでメタデータ(情報資源に対する、タイトル、作成者、分類といった情報。本の書誌データもその一種。)の全文検索を行う。キーワード検索と連想検索の切り替えを可能とした。

### 詳細検索

簡易検索と同様、キーワードによる検索だが、メタデータの項目を指定した検索と細かなオプションが設定できる。

### 連想検索

キーワードと違って、自然文、フレーズからの連想的な検索を行う。ユーザが入力したキーワードから、関連性を数値的に測ることによって「連想」される言葉(関連語)を導き出して、結果を取り出す。オープンソースのソフトウェアを利用している。

### 分類による検索

日本十進分類法(NDC)のカテゴリ、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、データを検索可能。最大 3 階層目の分類を選択した時点で、そのカテゴリに該当するデータが一覧表示される。

各機能を使い分けることにより、それぞれ違った角度からデータを探ることができるようになっている。

## 検索・ナビゲーションの流れ

検索から一次情報へのナビゲーションのメインの流れは、検索 検索結果の一覧 データの詳細表示 データ提供元の書誌ページ 一次情報があるものは一次情報へ、となっている。また、データの詳細表示からは、関連情報リンク(リンクリゾルバ)を通じた関連資料へのナビゲーションも行っている。後述するが、ヒットしたデータ内容を用いて、書籍販売サイト、総合目録、インターネット上の関連ページを見に行くことを可能とする機能である。

また、辞書検索でヒットした語群を用いてデータの検索に遷移することも可能としている。

## 辞書検索機能と検索支援

### 辞書検索

当館の典拠データを用いた辞書データベースの検索である。統制語の数は約 84 万語。キーワードで検索するが、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グラフィックで表示可能とした。

### 辞書による検索支援

キーワード検索時には、データ検索と同時に辞書データも同じキーワードで検索して結果を画面右に表示する。統制語・読み・上位語・下位語、類義語、関連語が表示され、その語群の中から適切な語を選択してデータの再検索を行うことが可能。

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た。

## 関連情報へのナビゲーション

### 関連情報リンク

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である。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注文や書誌データや写真、資料のレビュー等を見たりすることもできる。また、インターネット上の関連サイトへの遷移も可能である。検索結果で一次情報がある場合はよいが、PORTA では目録や参考情報をも対象にしており、その場合は、検索結果のその先へのナビゲーションが重要になってくると考えている。

## カスタマイゼーション機能

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が可能である。

ユーザグループ、ブロックの配置、ダイレクトメニューの編集、デフォルトの検索対象の設定、デザイン等のカスタマイズ

### お気に入りの編集

PORTA では、登録ユーザがお気に入りのメタデータにタグを付けていき、その付加価値情報を皆で共有して参照利用できるようにしている。

このお気に入り機能については、フォークソノミー (folksonomy) の考えを参考にしている。フォークソノミーは、Web2.0 とも呼ばれている次世代 Web の中でよく言われている方法だが、フォークソノミーとは folks(民衆)と taxonomy(分類学)の合成語で、ユーザがタグや分類を付与していき、その情報を共有し、検索等の場面でその情報を活用可能とすることを指して言われている。

## PORTA の検索対象になる

検索対象になるためには

連携の共通仕様を各システムで実装していただく。

ここで言う連携の共通仕様は、PORTA でさまざまなシステム等と連携するために定めた仕様とガイドラインを意味する。各デジタルアーカイブが実装すべきメタデータの受け渡しや通信のルール、つまり、メタデータ記述要素、メタデータスキーマ、通信プロトコル等の事項をデータ連携にあたって定めている。

PORTA が多くのシステムを検索対象とする場合、個々のシステムのインタフェースに合わせた個別対応をしていくには作業量としても限界がでてきてしまう。メタデータ記述要素、メタデータスキーマ、通信プロトコルの仕様を、できる限り世の中の標準に沿った形で共通に定めて、連携元のシステムに実装していただくと、効率的に進めることができると考えている。また、共通仕様に沿った実装を行った場合、副次的には、PORTA との連携だけでなく、外部のシステム同士が連携可能となるメリットがあることも期待できる。

データプロバイダディレクトリに登録していただく。

データプロバイダディレクトリとは、検索対象候補のシステムが、データプロバイダの名称、連携方法、リクエスト先の URL、データの更新周期、その他システム連携のために必要な情報を登録するリポジトリである。ここに登録された情報をもとに、可能な限り PORTA とデータプロバイダ間の自動的なシステム連携を図ることを想定している。

## PORTA の検索機能を利用する

ユーザが画面を通して検索する方法の他、外部システムが PORTA の検索機能を体系的に利用可能とするため、外部提供インタフェースの API(Application Programming Interface) を提供する。外部システムは、PORTA の検索機能を利用することで、PORTA の検索結果を横断検索先の一つに加えて提供する等ができ、付加価値をつけたサービスを構築することが可能となる。また、間接的には、これを通じて PORTA が検索対象としているシステムの利活用にもつながると期待できる。

OAI-PMH ( Open Archives Initiative Protocol for Metadata Harvesting )

メタデータを収集 (ハーベスト) するための仕様。

SRW ( Search/Retrieve Web Service )

ウェブサービス (SOAP) を利用した横断検索のための仕様で、現在仕様は LC が管理している。

OpenSearch

A9 が管理する簡易で軽量の横断検索のための仕様。実装が比較的容易である。

OpenURL

目録等の元データから適切コピー (appropriate copy) へリンクする際に、その記法を定めた標準仕様。

以上の 4 種類の方法で PORTA の検索機能を利用可能となる。ただし、OAI-PMH を通じたデータ提供の可否は、データの 2 次的配布にあたるため、各検索対象の許諾条件に従うこととしており、今後調整を行っていく予定である。

## 今後の PORTA は？

共通仕様の普及を図り、コンテンツを他システムへ機械的に提供可能なシステム (= データプロバイダ) となるための呼びかけ、登録促進を行っていく。機械的な提供機能を有していないシステムも多いのが現状であり、積極的な普及を図る。

検索対象のコンテンツは、25 種類ではまだ十分とはいえない。今後は、検索対象を順次増やしていく予定である。具体的には、以下にあげるような優先順位が高いターゲットとしていきたい。

- ・政府機関の保有コンテンツ
- ・図書館の保有コンテンツ
- ・学術研究機関の保有コンテンツ
- ・地方公共団体の保有コンテンツ
- ・有用な民間のコンテンツ

扱う分野については、図書館が分け隔てなく多様な分野を扱ってきたのと同様に、基本的に全分野を対象としていく。

また、先に触れたように、当館では NDL デジタルアーカイブシステムの構築中である。将来的

には、PORTAとNDL デジタルアーカイブ・システムは連携・統合される予定である。

最後になるが、有用なコンテンツであっても、必要なときに探し出して利用されなければ意味が薄れてしまう。換言すれば、人々の目に触れて活用されてこそ意義を発揮するものだと言える。その考えの下で、コンテンツの利活用を促進するため、デジタルアーカイブポータル提供と検索対象の拡充を行っていききたい。