

CA2068

新しい「国立国会図書館サーチ」への
統合に至る道のりかわしまたかのり
川島隆徳*

2024年1月5日、国立国会図書館は統合的な検索サービスである「国立国会図書館サーチ」⁽¹⁾ (NDLサーチ) を公開した。本稿では、主にシステム面に着目し、開発の背景と実装について述べたい。

1. NDLサーチの概要

NDLサーチは、当館の蔵書目録「国立国会図書館オンライン」(NDLオンライン; CA1940参照)と全国の図書館の総合目録機能を持つ「国立国会図書館サーチ」(旧NDLサーチ; CA1762参照)、さらに調べ方を案内するリサーチ・ナビを統合した検索サービスである。これにより、当館資料の蔵書検索、全国の図書館等資料の横断検索、当館提供書誌のダウンロードやAPI利用、当館作成のレファレンス情報の提供が一つのサービスから行えるようになった。統合したことで完全に新しいサービスの提供を始めたということではなく、これまでは複数のウェブサービスを行き来する必要があったものが、単一のインターフェースで提供できるようになった、というのが実態である。例えば、検索窓の直下にある「国立国会図書館」のチェックボックスだけを選択すればNDLオンライン相当の挙動となるし、「全国の図書館」のチェックボックスも選択すれば、旧NDLサーチ相当の挙動となる。そういった訳で、本稿ではNDLサーチの個々の機能やサービスを詳説することはせず、統合に至る道のりを紹介することとしたい。

2. OPACの流れ

当館のOPACがWebに公開されたのは2000年のことで、当時のシステムはフルスクラッチで開発したものだ。その後、2012年には業務システムのパッケージ導入に伴って、一時OPACもその製品(ExLibris社のAleph)の機能で提供することになった(NDL-OPAC)。しかし、国立図書館としての業務・サービスとパッケージ製品の乖離が大きく、今後のサービスの展開の障壁ともなりうることから、改めて基幹業務システムのフルスクラッチ開発を行った⁽²⁾。この過程で2018年に公開されたのがNDLオンラインである。

3. 総合目録とデジタルアーカイブ関連の流れ

全国の図書館の目録を横断検索するという総合目録ネットワークシステムに対する当館の試みは1993年から開始していたが、Webサービスとして公開したのは2004年である。これとは別にデジタルアーカイブの横断検索サービス「PORTA」(E706、CA1677参照)を開発し、2007年に公開した。そして2012年、これらを統合する形で旧NDLサーチを公開した。

旧NDLサーチは、総合目録に加えて、様々なデジタルアーカイブや、当館の蔵書目録や国立国会図書館デジタルコレクション(以下、デジタルコレクション)も横断検索できるサービスとして開発された。開発当時は、検索は全て旧NDLサーチで行ってもらい、という想定もあったようだが、実際には主たる検索サービスはNDL-OPACで、NDLサーチはそれでも見つからないときに利用される、という使い方がヒアリングなどでは明らかになっていた。

その後、2019年にジャパンサーチ⁽³⁾(試験版)が公開されると、旧NDLサーチは書籍等分野の「つなぎ役」を担うこととなり、連携は書籍等分野に絞る、という方針が立てられ、博物館等の連携は徐々にジャパンサーチに移っていった。

旧NDLサーチはシステムとしては独自開発し、膨大なメタデータの同定処理⁽⁴⁾を行うために、2012年当時には最新の分散処理システムであったHadoopを採用したバッチ処理が実装されていた。Hadoopは複数台のサーバから成るインフラ構成が必要で、また処理のプログラムも複雑であった。このため、旧NDLサーチは全面的なシステム更改が難しく、大きく手を入れられないまま10年以上が経過し、技術的負債が増えてきていた。

4. NDLサーチへ

NDLオンラインは、内部的には蔵書目録とデジタルコレクションのメタデータの統合処理を行っており、旧NDLサーチほどではないが複雑なバッチ処理の仕組みを持っていた。それ以外にも、検索の仕組みやインターフェース等、NDLサーチとNDLオンラインでシステム上の要素は重複しており、アンケートなどでも両者の区別がつかないユーザがいることが分かっていた。このため、システムとしての統合自体は長らく検討されてきた。ジャパンサーチの登場により、旧NDLサーチは書籍等分野に特化できるようになり、メタデータの入れ物としても書籍等のことだけを考えれば良くなったため、システム統合の見通しが晴れた。これが契機となって詳細な検討が進み、2022年の冬頃から具体的なシステムの検討が始まった。

* 電子情報部電子情報サービス課

5. NDL サーチのシステム構成

NDL サーチの基本機能は旧 NDL サーチと大きく変わらず、メタデータの収集、組織化、検索、そして提供用の Web サービスという四つから成り立っている。

メタデータの収集については、基本的には旧 NDL サーチから収集プログラムを引き継いだ。ただし、メタデータの収集方式は、前回収集分からの差分を収集する差分更新方式よりも、毎回全件収集する方式を優先するように変更した。差分更新はデータ提供側も実装が難しい。一方で現代の計算機環境では、データ量にもよるが、全件データであっても大きな処理時間はかからない。例えば、デジタルコレクションとの内部的なデータ同期は、日々全件データ（約 600 万件）を用いて行っている。今後の NDL サーチと外部機関の連携は、OAI-PMH ではなく、定期的な全件出力ファイルの取込みという方法も選択肢となっていくだろう。

組織化は、これまで Hadoop を用いて分散処理をしていたところを、逆に 1 台の高スペックマシンで行う方法に変更した。毎日、組織化処理が開始する時点で収集されている数億のメタデータを、1Tb のメモリに全て読み込み、その上で書誌同定の処理等を行うようにしている。現在、同定処理にかかる時間は 3 時間ほどで、旧システムの半分程度まで高速化した。この仕組みを可能にしたのはインフラのクラウド化である。NDL サーチはクラウドに最適化された形で開発しており、当該サーバは組織化処理の時だけ利用料が課金される仕組みになっている。サーバを所有する必要がなくなったことで、高スペックのハードウェアを利用するハードルが大きく下がった。

検索については、従来の Apache Solr から Elasticsearch に切り替えた。実際には Elastic Cloud というクラウドサービスを利用しており、検索サーバの管理をクラウド側に任せることで、柔軟な構成が取れるようになった。例えば、外部 API 用の Elasticsearch と検索画面用の Elasticsearch はサーバが分離されており、外部 API に大きな負荷がかかっても、サービス画面には影響は及ぼさない。また、検索画面用については、アクセスが集中する時間帯にサーバを自動的に増強する仕組みなども実現した。

提供用の画面は、旧 NDL サーチと NDL オンラインを UI として分離する案や、CiNii のようにタブで分けるといった案などもあったが、最終的には統合した形に決定した。当館以外に所蔵がある場合、利用者によっては当館の所蔵を利用するよりも容易に利用ができる可能性もあるという点と、利用者の三分の二程度は Web からの利用（館外での利用）であり、「本を探す」サイトとして使っているのであって、必ずしも当館の

所蔵情報が最終目的ではないという点を考慮した。実際には、三分の二程度のユーザは特に「国立国会図書館」に絞込まないで検索を行っているようである。

画面のデザインも全体的に見直し、初見のユーザが使いやすいことを意識して、ユーザテストなども行いつつ開発した。アプリケーションは Kubernetes 上のコンテナとして実装しており、リリース作業も自動的に行えることから、運用しながらの継続的な改善が行いやすい構成が実現できた。

6. リサーチ・ナビとみなサーチ

リサーチ・ナビ⁽⁵⁾は、従来は独立したシステムとして運用していたが、元々持っていた機能が段階的に別システムに移っていき、最終的に CMS（コンテンツ・マネジメント・システム）のみが残ったような形となっていた。NDL サーチでも様々なページの作成に何らかの CMS は必要だったことから、NDL サーチの CMS を使って、リサーチ・ナビの機能を NDL サーチに統合することにした。これによって、検索結果や書誌詳細の下にリサーチ・ナビへのリンクを出すことが容易になり、またリサーチ・ナビから検索結果や書誌詳細への遷移も軽快に動くようになった。パスファインダーは、ある意味では検索システムの手引きでもあるため、統合されることで全体としてのユーザ体験は向上したのではないかと考えている。

みなサーチ⁽⁶⁾は、「点字図書・録音図書全国総合目録」に由来する旧 NDL サーチの「障害者向け資料」の検索メニューが別サービスとして分離されたものである。裏側のシステムは NDL サーチと完全に共有されているが、独自のドメインとデザインを持ち、NDL サーチとは別ブランドとして展開できるように開発した。旧 NDL サーチとの大きな違いとして、全文検索と全文テキストダウンロード機能の追加がある。これは、デジタルコレクションの全文検索機能を利用して実現されており、全文テキストについてはみなサーチからワンストップで提供できるようになっている。

7. 結びに

NDL サーチは、様々な流れを持って開発されてきた国立国会図書館の検索システムの現時点での到達点であると言える。当面は大きくシステム構成を変えることはせず、クラウド化によって得た柔軟性を有効活用しながら、新しい NDL サーチのサービスを育てていきたい。

- (1) 国立国会図書館サーチ。
<https://ndlsearch.ndl.go.jp/>. (参照 2024-06-04).
- (2) 今から振り返れば、この決断なしには、オンライン登録や個人送信サービスの実装などは非常に困難だったと考えられる。

- (3) ジャパンサーチ.
<https://jpsearch.go.jp/>, (参照 2024-06-04).
- (4) 複数館で所蔵する同一の資料の書誌をまとめる処理。例えば、
同じ MARC 番号や同じ ISBN を持っていれば同定される。
- (5) リサーチ・ナビ.
<https://ndlsearch.ndl.go.jp/rnavi>, (参照 2024-06-04).
- (6) みなサーチ.
<https://mina.ndl.go.jp/>, (参照 2024-06-04).

[受理:2024-08-06]

Kawashima Takanori

The Road to the New NDL Search