

VuFind を利用した異種情報統合検索システムの構築

——昭和音楽大学オペラ研究所デジタル・アーカイブの事例——

岩崎 陽一¹ 吉原 潤² 根木 昭²

概要: デジタル・アーカイブを作成する目的は何か。資料の蓄積と保存という、広く認知されている目的の先に、個々のアーカイブはそれぞれ固有の目的をもっているはずである。昭和音楽大学オペラ研究所が文部科学省私立大学戦略的研究基盤形成支援事業の補助を受けて整備した、デジタル・アーカイブを根幹に据えたウェブサイト「オペラ情報センター」は、オペラ研究者・制作者の活動を助け、国内のオペラの振興に寄与することを目的としている。その目的のために、公演記録 DB、作品 DB、所蔵資料書誌 DB という三つのデータベースを統合したサービスを、オープンソースのディスカバリー・サービスである VuFind を利用して開発した。本稿では、本サービスの機能設計とデータ設計を解説し、そこから得られた知見を一般化して示す。これにより、学術データベースのひとつの新たな設計・開発方法を提示する。

Creating a Digital Archive and Information Service with VuFind: Opera Information Center of Showa University of Music

IWASAKI YOICHI¹ YOSHIHARA JUN² NEKI AKIRA²

1. はじめに

1.1 デジタル・アーカイブの目的

何のためにデジタル・アーカイブを作るのか。デジタル・アーカイブは知的資産の蓄積・保存を目的とするという趣旨が、総務省のガイドライン [1] をはじめ、各処でいわれている。では、知的資産の蓄積・保存は何を目的としているのか。蓄積・保存は、手段がデジタルであるか否かを問わず、MLA が担うべき役割である。その目的についてはさまざまな議論があるが、たとえば小島浩之氏は [2] において、図書館にとっての資料保存が「歴史を後世に伝えるという意思を伴った行為」であるとしている。また、図書館情報学の或る教科書には、「情報メディアの保存は、現存する情報を将来の利用に備えて残していくことを目的としている」ともある ([3] p. 106)。後者は、小島氏が言及する「利用のための保存」、すなわち「資料をいつでも利用可

能な状態に整えておくことが図書館における資料保存」であるとする考え方に類するものであろう。

たしかに、すべてのデジタル・アーカイブは、そのような目的のためにデジタル資料の蓄積・保存を行っているといえる。しかし、こういった普遍的な目的とは別に、個々のアーカイブ事業には、より直接的・具体的な目的があるだろう。何のために、誰に歴史を伝えようとするのか。誰が、何のために資料を利用することを願うのか。個々のデジタル・アーカイブの事業においては、このような事業目的を事業主体が明確に意識し、その実現を促すかたちで事業を進めなければ、事業主体が意識的または無意識的に期待するような結果を得られないだろう *1。本来の目的を見失い、資料のデジタル化ができたことで満足していれば、開店休業状態のデジタル・アーカイブが立ち並ぶことになる。ただ手元の資料をデジタル化してアップロードし、画面にインデックスを並べたり、汎用的なメタデータ項目で検索できるようにしたりするといった手法では、利用者は

¹ 日本学術振興会
Japan Society for the Promotion of Science

² 昭和音楽大学
Showa University of Music

*1 杉本重雄氏は [4] において「個々のデジタルアーカイブごとの利用者要求を満たすサービスの実現」の必要性を論じている。

訪れない。そして誰にも利用されなければ、コンテンツが伝えるべき「歴史」は忘れられてしまう*2。

1.2 オペラ研究所と「オペラ情報センター」

昭和音楽大学オペラ研究所（以下「オペラ研」）は、国内でのオペラ上演に関する調査・研究を通して、我が国におけるオペラをはじめとする舞台芸術の振興に貢献することを目的としている。そして、オペラ研が提供する、デジタル・アーカイブを中心とするウェブサイト「オペラ情報センター」*3 もまた、オペラ研究者およびオペラ上演関係者が求める情報を供給し、我が国でのオペラの振興に寄与することを目的としている。

オペラ研は、書籍『日本オペラ史』[6][7]や、文化庁の委託事業として制作する年報『日本のオペラ年鑑』を刊行しており、その編纂の一次資料となる国内オペラ公演のフライヤーと、いわゆる「プログラム」（会場で配布または販売される、主に演目や演奏者・出演者の解説を載せた冊子）を多数、保有している。また、古い公演の写真や録音テープ、ビデオ等も保存・管理している。こういった公演調査に関わる事業とは別に、国内で創作されたオペラ作品の作品詳細、出版状況等を組織的に調査・研究する事業も行った実績がある。こうして蓄積した情報資産により、これまで日本のオペラ研究とオペラ公演を支援してきた。しかし、資料のデジタル保存や共有については対応が遅れていた。そこで、所蔵資料のデジタル・アーカイブ構築を中心とし、その他の情報資産を統合してインターネット上で提供する事業を計画し、文科省の私立大学戦略的研究基盤形成支援事業による補助対象プロジェクト（プロジェクト名「オペラ資料のアーカイブ化を通じた情報センター機能の構築」）として採択された。2012年度から事業を運営しており、ウェブサイト「オペラ情報センター」は2014年にバージョン1を公開した（図1）。現在、2016年度上期にバージョン2を公開すべく、開発を継続している。

本稿では、「オペラ情報センター」がその目的である「オペラの振興」を実現するためにどのような設計を行ったか、そしてそれを限られたリソースで実装するためにどのような工夫を行ったかを詳説する。そして、そこから得られる知見を一般化して結論において提示する。図書館用ディスクバリー・システムを利用して検索サービスを作り上げる手法は、様々な学術データベースに応用可能であると考えている。



図1 オペラ情報センター バージョン1

2. サービス設計——三つのデータベースの統合

所蔵資料のデジタル・アーカイブを公開するには、ダブルリンク等々に規定されるメタデータを作成し、それを検索できる機能を提供するのが王道といえるだろう。しかし、それではたして、オペラ研が主な利用者として想定するオペラ研究者たち、またオペラ公演をつくっていく音楽関係者たちの需要を満たすことができるだろうか。筆者は、データ検索を、既知のデータの検索と未知のデータの検索とに分けて考えている。求める資料のタイトルや著者が分かっている利用者に対しては、それらのメタデータを正確にデータベースに入れておけば、資料に到達することができる。しかし、たとえば「〇〇さんの出演した公演の資料が欲しい」「〇〇年代のプッチーニのオペラの公演資料が欲しい」といった要望もあるだろう。むしろ、研究者や音楽関係者の需要を想定すると、このような未知の資料の検索の方が一般的であると推定される。そして、オペラの振興を目的とするオペラ研は、これらの要求に応えられるサービスを提供しなければならない。

では、どのようにしたら、上に例示したような要望に応え、所蔵資料のファイルを、それを必要とする利用者届けられるサービスを作成できるのか。それを考えるためには、資料の特性と、その用途を考慮しなければならない。オペラ研の保有する資料は、そのほとんどが、過去に行われたオペラの公演と結びついている。資料の検索は、実質的に、公演の検索に等しい。したがって、所蔵資料データベースは公演記録データベースと結びついていることが望ましい。さらに、公演記録データベースは、作品データベースと結びついていることが望ましい。というのも、オ

*2 富澤浩樹氏は、震災関連デジタルアーカイブの利用が進まないことに起因して、震災の記憶が風化するのではないかと、という懸念に言及している。そして、記憶を伝えるには、資料の利活用を継続的に促すための取り組みが必要であると論じている [5]。

*3 <http://opera.tosei-showa-music.ac.jp/>

ペラの公演は、すべて、何らかの作品の上演である。そして作品は、作曲者や台本作者といった多数の属性をもつ。作品データベースを設置しなければ、作品の属性を手がかりに公演を特定することはできない。つまり、「〇〇年代のプッチーニのオペラの公演資料が欲しい」という検索には結果を返せない。

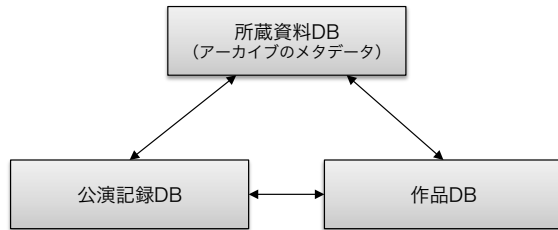


図2 オペラ情報センター サービス構成

以上のように、オペラ研の所蔵資料は、公演記録データベースと作品データベースを伴ってはじめて、現実的に利用可能なデジタル・アーカイブとして提供できるようになると考えられる。それゆえ、オペラ研では、所蔵資料データベース、公演記録データベース、作品データベースという三つのデータベースを組み合わせ、**「オペラ情報センター」**という単一のサービスを提供することにした。

これら3種の情報資産が、オペラ研が長年蓄積していた3種の情報——資料と、公演記録と、作品調査の成果——と一致するのは、偶然ではない。オペラの研究や公演を促進するには、デジタルであれ物理媒体であれ、これら3種の情報が必要であることを、上記の分析は示しているといえるだろう。

3. システム設計——VuFindの活用

以上のようなまったく質の異なる三つのデータベースをウェブで公開しようとするならば、それぞれのデータベースをRDBで作成し、それらを検索・表示するアプリケーションを何らかのフレームワークを用いて作成するという方法がはじめに思い浮かぶ。しかし、フレームワークが自動で作ってくれる検索インターフェイスはユーザビリティが低く、かといって全文検索やファセット機能を自力で追加していくならば工数が見るみる増加していく。フレームワークを元に自前で検索サービスを構築することは、ほぼ不可能と考えられた。

簡単にサービスを作る方法として、何かCMSの類を用いるという選択肢もあるだろう。デジタル・アーカイブについてはDSpace^{*4}やOmeka^{*5}といったCMSの採用実績が多いが、公演記録データベースまでカバーできるものではなく、別途作成するそれらの外部データベースとCMSを連携して運用するのは余計に煩雑となる。そもそも、DSpace

等はアーカイブやリポジトリのバックエンドとしては申し分のない機能を有しているのかもしれないが、検索機能は決して充実しているとはいえない。また、文献[8]等ではDrupal^{*6}やWordPress^{*7}といった汎用的なCMSを用いる方法も紹介されているが、やはり、満足のいく検索インターフェイスが自動でできるような方法ではない。

何とか楽に、低予算で、リッチな検索サービスを作れないものか。筆者らはここで、図書館用のOPACに注目した。現代のOPACは、書籍や電子ジャーナル、CD等、さまざまな質の異なるデータを包括的に扱うことができる。オペラ研の扱うデータの多様性も、OPACで処理することができるのではないかと。オペラ研の扱うデータには、所蔵資料、公演記録、作品という第1レベルの質の差異と、所蔵資料の下位区分として存在する紙資料(静止画ファイル)、録音資料(音声ファイル)、映像資料(動画ファイル)といった第2レベルの質の差がある。しかしどのような差異であれ、OPACが読み込める形式、たとえばMARCでデータを記述することさえできれば、OPACが吸収してくれるはずである。そして、MARCというデータ形式は、オペラ研の扱う各種データを記述するのに十分な論理構造をもっている。

また、OPACは多くの場合、権限種別を扱える認証機能を有している点も重要であった。オペラ研のデジタル・アーカイブに含まれるデータには、不特定多数には公開できないものも多い。したがって、未認証ユーザー、学内ユーザー、研究所内ユーザー等の権限区分を扱える認証機能が備わっているのは魅力的である。筆者らはKoha^{*8}、Next-L Enju^{*9}、Blacklight^{*10}、VuFind^{*11}などの主要なオープンソースのOPACおよびディスカバリー・サービスを試用したうえで、優れた検索インターフェイスをもち、セットアップが容易で、多言語に対応し、カスタマイズに関するドキュメントも充実しているVuFindを利用することにした。開発計画を具体化させた2013年当時、バージョン2.0の正式版がリリースされた直後であり、バグ修正版である2.0.1のリリースを待って、それをカスタマイズのベースとすることにした。

VuFindについて簡単に紹介しよう。これは米ヴィラノヴァ大学が開発するオープンソースのディスカバリー・サービスであり、検索機能はSolr^{*12}、その他の機能はZend Framework^{*13}を用いて実装されている。レコードはMARCまたはMARCXMLで登録する。SolrでMARCを扱う部分は、独立したオープンソースのSolrアドオンであ

*6 <https://www.drupal.org/>

*7 <https://wordpress.com/>

*8 <http://www.koha.org/>

*9 <http://www.next-l.jp/>

*10 <http://projectblacklight.org/>

*11 <http://vufind-org.github.io/vufind/>

*12 <http://lucene.apache.org/solr/>

*13 <http://framework.zend.com/>

*4 <http://www.dspace.org/>

*5 <https://omeka.org/>

る SolrMarc^{*14} を用いている。VuFind には外部の ILS (統合図書館システム) と相互運用するための「ドライバー」が各種用意されており、独自にドライバーを作ることもできる。オペラ情報センターにおいては、デジタル・アーカイブのファイル管理は別のサービスとして独立させ、それをひとつの ILS とみなし、オリジナルのドライバーを開発して相互運用を可能にした。図 3 は、オペラ情報センター全体のシステム構成を示したものである。

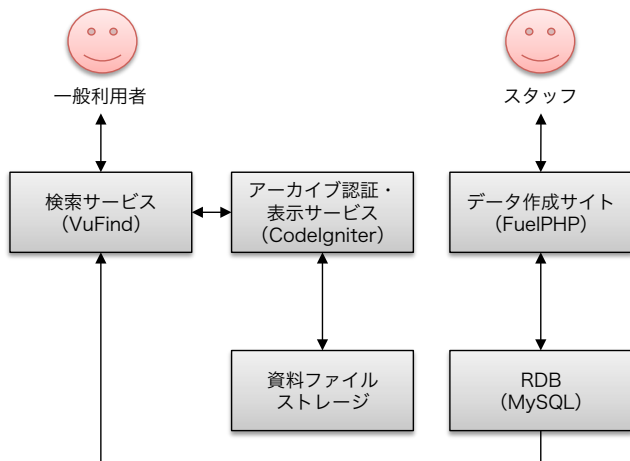


図 3 オペラ情報センター システム構成

4. データの MARC21 での記述

所蔵資料のメタデータを MARC21 で記述するのは造作ないが、他の 2 種のデータについては工夫を要した。そのうち、作品データの取扱いは比較的容易であった。タイトルや著者は、MARC21 の Field 245 (Title Statement) に、仕様に従って格納することができる。幕数、キャスト、出版状況、楽譜の所在といった、MARC21 の仕様でカバーされていない属性については、Field 500 (General Note) に格納して対処した。

公演データの取扱いもまた、大部分は MARC21 の仕様に従うことができた。MARC21 には、映画やオペラの DVD 資料を扱うために、出演者を格納するためのフィールド (511: Participant or Performer Note)、会場や日時を格納するフィールド (518: Date/Time and Place of an Event Note) 等が用意されている。これらで扱いきれない属性については、Field 500 や Field 59X (Local Notes) で記述した。

資料と関連づけられた、或いは公演に出演した人名・団体名は、上述の Field 245 や 511 に格納するだけでなく、Field 600/610 (Subject Added Entry) や Field 700/710 (Added Entry) にも登録した。そしてそれらの人名を、典拠データとして VuFind に登録した。こうすることにより、

*14 <https://github.com/solrmarc/>

特定の人物の出演作品・著作作品・関連資料等を高速に検索できる。

オペラ情報センターは日英の 2 言語に対応している。データの多言語化についても MARC21 の仕様に従い、Field 880 (Alternate Graphic Representation) に英文表記と日本語の読み仮名とを登録して対応した。

ただし、次節に提示するような、公演データの複雑な構造は、MARC21 の仕様の範囲内で表現することができなかった。そのため、XML で表現したデータを独自に定義したフィールドに格納した。VuFind は (或いは SolrMarc は) その XML を文字列として取得できるため、アプリケーションの側でパースし、画面表示に用いた。この XML は純粋に画面表示に関わるものであり、検索で用いられるデータはすべて通常の MARC21 フィールドに格納されている。表示用と検索用でデータを分けることにより、VuFind の機能を利用しつつ、分かりやすい画面表示が可能となった。

なお、所蔵資料書誌、公演記録、作品それぞれのデータは RDB に保持されている。それを VuFind にロードするに際し、独自ツールで MARCXML 形式のファイル出力し、VuFind に附属するデータ登録ツールで登録を行っている。

実のところ、オペラ情報センターのデータはそのまま他の OPAC に投入するつもりはないので、無理に MARC の仕様に合わせる必要もない。しかし、あえて仕様を無視する必要もない。なるべく仕様に従うことで、VuFind の標準機能を最大限利用できる。また、「なるべく MARC に準拠」という方針を拡大適用して、読み仮名の表記方法、補記の方法等についても、なるべく日本および世界の MARC に関する標準的手法を踏襲することにした。そのため、データの inputs は大部分を MARC 作成のエキスパートである企業に委託し、質の高いデータを作ることができたと考えている。

5. 公演データの概念モデル

5.1 バレエ研究所の公演概念モデル

オペラの公演記録をデータ化する概念モデルを作成するにあたっては、昭和音楽大学バレエ研究所 (以下「バレエ研」) のデータベースを参考にした。オペラ研の姉妹機関であるバレエ研は、オペラ研よりもひと足早く、バレエの公演記録のデータベース化を試みていた。そこで用いられている概念モデルは、文献 [9] で詳しく提示されている。ここでは公演記録が、FRBRoo (オブジェクト指向 FRBR) の考え方を利用しつつ、「興行—公演—演目」という 3 階層のエンティティと、「計画—遂行」という 2 階層のエンティティの組み合わせとして分析されている。

オペラの公演記録も、同様の概念モデルによって分析できる。おそらく、バレエ研の概念モデルは、音楽や演劇、さらには学術大会やシンポジウムの記録に至るまで、若干の拡張または変更を行えば広範に利用できるものであろう。

以下、オペラ研がどのような概念モデルを作成したか、バレエ研モデルからの改変点を中心に詳述する。

5.2 オペラ研究所の公演概念モデル

バレエ研が「興行（イベント）」というエンティティに担わせた、複数のステージをまとめる役割は、オペラの世界では「プロダクション」というエンティティが担っているといえる。プロダクションは公演団体と演目等により規定され、例えば「〇〇歌劇団の××年の《魔笛》のプロダクション」といったように認識される。フライヤーやプログラムといった物理資料は、多くの場合、プロダクションに從属する。それゆえ、プロダクションを基本単位となるエンティティとして設けるのが都合がよい。ひとつのプロダクションは、多くの場合、複数のステージにより構成される。個々のステージは、通常ひとつの演目で、ときには複数の演目で構成される。ここまでは、「興行」を「プロダクション」に置きかえただけで、バレエ研の概念モデルと相似形である。両者の大きな違いは、オペラ研においては、ひとつの興行において複数の異なる作品が日を変えて上演されるとき、それぞれの作品を互いに異なるプロダクションに割り当てているところにある。たとえばメトロポリタン歌劇場は2006年の来日興行において、《ワルキューレ》と《椿姫》と《ドン・ジョヴァンニ》の3作品を、それぞれ異なる日程で上演した。これらは三つの異なる制作物であり、データベースにおいては三つのプロダクションとして登録した。

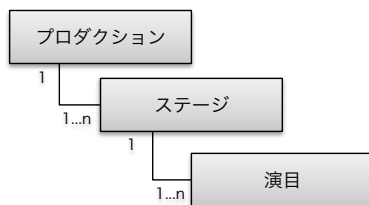


図4 概念モデル1

また、オペラ研の概念モデルには次のような違いもある。オペラ研のデータベースでは、個々のステージにおいて上演される演目は作品ではなく「作品表現」という別のエンティティであると考えられる。これはFRBRで「著作」の実現体とされる「表現形」に相当する概念である。オペラの公演を扱う際、この階層化を厳密に行うことが大きな意味をもつ。というのも、オペラの公演においては、同じ作品が、日本語訳で上演されたり、原語で上演されたりする。また、もともとはフルオーケストラ用に書かれた作品が室内楽編曲で上演されたり、長い作品が抜粋で上演されたりもする。これらを、元の作品の「表現形」とみなすことができる。個々の表現形では、原則として著作者が異なる。ひとつの表現形が複数のプロダクションにおいて上演されることも

あるため、表現形に關係する諸属性は演目エンティティから切り離し、別個の「作品表現」エンティティに從属させた。この概念モデルを用いることにより、たとえば、或る作品がこれまでどのようなかたち（言語、編成等）で上演されたことがあるのかを調べることができるようになる。

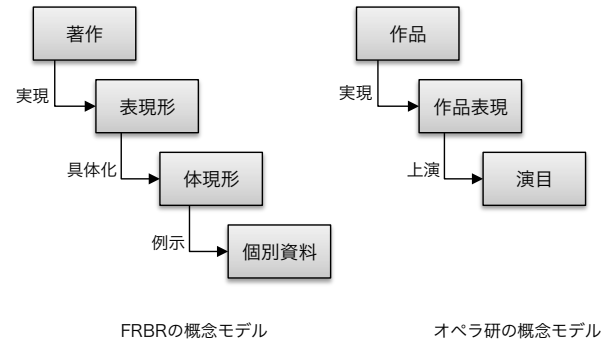


図5 概念モデル2

上演を計画と遂行のふたつのレベルで分析する点は、バレエ研に從った。オペラの公演においては、直前になって、ときには当日、出演者が変更されることがしばしば起きる。実際に出演しなかった人物については、出演の事実がないのだから記録する必要はない、という考えもあるだろう。しかし、その人物で上演が計画されたということは事実であり、その事実はオペラ研の所蔵資料であるフライヤーやプログラムに記されているので、それら所蔵資料にもとづいてデータをつくるという方針に照らせば、記録しておくべきと考えられる。また、たとえ出演がキャンセルになったとしても、出演が予定されていた歌手は、その役柄を演じることができたはずだと推定できる。これから新しいプロダクションを作ろうとするひとひとが「〇〇役をできる歌手にはどんなひとがいるかな」と検索するとき、それらの歌手も、キャンセルになったという情報を含めて、提示した方が検索者の期待に応えられるだろう。同じ理由で、もしキャンセルが発生した場合に代役で出られるようスタンバイしていた歌手たち（カバーキャスト、アンダースタディ）のデータも記録している。

5.3 公演データの提示

上述の概念モデルは、公演に関する事実を正確に記述することを目的に考案されたものである。そして、事実がどうであるかということ、それをひとがどう認識しているかということは、別の話である。ひとの認識に依拠せずに事実がどうであるかを記述することは不可能であろうが、上述のモデルの作成においては、ひとによる意味づけを極力排除し、客観的に確認可能な属性のみをデータ項目として設定した。しかしそれをウェブサイトにおいて提示する際は、利用者がそれをどう認識するかを想定し、事実の把握を容易にするかたちで提示するのが望ましい。

ひとがオペラの公演を認識するとき、プロダクションに附随するエンティティとして、「プログラム」というものが認識されていると考えられる。資料としてのプログラムではなく、演目のセットとしてのプログラムである。たとえば、或るオペラ公演団体が、東京と大阪では A プログラムを、名古屋では B プログラムを上演するということがしばしばある。事実としては 3 箇所それぞれ特定の演目が上演され、東京の演目と大阪の演目が一致しているだけなのだが、そのような出来事をひとは「プログラム」という枠組みで認識するだろう。そして、演目はステージに付随するものではなくプログラムに附随し、そのプログラムが複数のステージにおいて具体化されていると理解するだろう。オペラ情報センターでは、データをウェブサイトで提示する際、データ項目としては存在しない「プログラム」を仮想的に設けて提示している^{*15}。

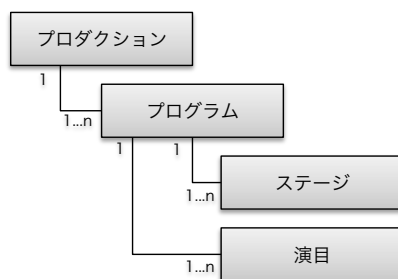


図 6 データ提示の際の概念モデル

6. 開発外注における工夫

以上のような設計をオペラ研内部で作成した上で、内部設計以降の工程を外部の開発会社に委託した。VuFind を利用したシステム開発のノウハウをもつ開発者は、筆者らが調べた限りでは首都圏には見当たらなかった。やろうとしていることが、VuFind のカスタマイズで実現できる保証もなく、システム開発の外注は困難をきわめた。いくつかの開発会社に受託を断られた末、数社から見積もりを取り、OPAC 関係の開発を得意とする開発会社になんとか開発を受託してもらうことができた。

本案件の外注が難航した原因は、VuFind のソースコードおよびドキュメントの解読に始まる調査・研究が必要な部分が多く、ウォーターフォール型の開発工程を想定して見積を立てることが難しいことにあると思われる。VuFind のような既製システムを使わず、一般的なフレームワークを使っていちから構築してほしいという依頼であれば、見

^{*15} 概念モデルをひとの認識に適合するように構築することもできる。その場合、プログラムを、RDB におけるひとつのテーブルとして定義することになるだろう。筆者も、そのようにして概念モデルを構築したこともあった [10]。今回のシステムにおいては、データが過度に複雑になることを恐れ、またデータへの意味づけを後から変更することができるよう、データの概念モデルは客観的事実に適合させることにした。

積を出せる開発会社はいくらでもあるだろう。しかしそれでは、本稿冒頭に示したような要件を満たすサービスは作れない。

本事業では、開発手順を 2 段階のスパイラル方式にすることにより、外注に際して生じる問題を軽減できたと思われる。第 1 段階では、内部設計は暫定的でも構わないので、とにかく VuFind をカスタマイズして、目標に向かって前進することにした。もちろん、公的予算を用いての事業であるので、バージョン 1 として外部公開できるサービスを作り上げることが前提となる。しかし、サービスの洗練や安定性には、第 1 段階では固執しなかった。

第 1 段階により VuFind の詳細が明らかになったところで、リファクタリングも含めて、サービスの完成に向けた第 2 段階に入る。この段階では、個々の機能の洗練と安定を図るほか、バージョン 1 に対して寄せられた、次節で示すような要望を満たすサービスの完成を目標とする。現在、この第 2 段階開発の最後の仕上げをしており、2016 年度の早い時期にバージョン 2 を公開することを目指している。

7. 問題点と改善計画

バージョン 1 のサービスを公開した後、実行パフォーマンスやアクセス数を測定し、改善点を調査した。また、オペラ研内外の研究者や学生に試用してもらいフィードバックを得たほか、有識者を招いて外部評価委員会を組織し、改善すべき点を指摘していただいた。以下、主要な問題点と改善計画を挙げていく。

(1) 検索結果の統合

所蔵資料、公演記録、作品データという、オペラ研のもつ情報資産をすべて包括的に検索できるシステムを作ろうという考え方にもとづいて、バージョン 1 のシステムでは、たとえば「ワーグナー」で検索すると、ワーグナーと関連する資料、公演記録、作品データのすべてがひとつのリストに統合されて表示されるようにした。もちろん、その後でデータ種別により絞り込みが可能である。(図 7 を参照。)しかし、どうやら、これら 3 種のデータをひとつのリストにまとめることは、むしろ利用者の混乱を招いたようであった。例えば、最近の図書館システムでは、雑誌記事と蔵書と、他の図書館にしかない資料とが一緒に表示されることがあるが、筆者自身もこれを不便と感じていた。冒頭で述べたように、利用者が何を求めているのかという問題に立ち返り、結果表示の方式を改める。バージョン 2 では、検索結果を統合せずに、三つのタブに分けて表示するように変更する。

(2) SEO

VuFind はもともと図書館での利用を想定しているためか、SEO をほとんど考えていないように思われる。たとえば、日本語ページと英語ページの切替は Cookie



図 7 統合された検索結果

を使って行われており、ひとつのデータに関しては全言語のページで同じ URL が用いられる。検索されるデータ、すなわち所蔵資料をリソースのノードと考えるならば、メタデータがどの言語で示されようとも、ノードに対して単一の URI が与えられる方が、Linked Data の考え方に照らすと望ましいのかもしれない。しかし、それは SEO の観点では好ましくない。検索エンジンのインデックスには、英語版のページも、日本語版のページも、どちらも登録しておいてもらいたい。Google で日本人が日本語で検索したならば、ページタイトル等も日本語化された日本語のページを提示し、そのページへの集客を図りたい。第 2 段階の開発では SEO の強化を予定しているが、その中で、Linked Data の思想との共生を図りたい。

(3) 実行速度

筆者はこれまでに Solr を用いていくつかのサービスを開発したことがあるが、本サービスは、異常とも言えるほど検索速度が遅い。メモリーリソースの消費も激しい。これからパフォーマンス・チューニングを実施する予定であるが、どうやら、MARC が想定していないデータを無理矢理 MARC で表現して読み込ませても、システムが想定しているようなパフォーマンスが出ないのではないと思われる。しかし、Solr のチューニングで充分改善の余地はあると見ている。

(4) データの可視化

図書館の蔵書検索システムで、検索結果の可視化が行われているのを見たことがない。VuFind もまた、リスト表示以外の可視化に関する機能は持ち合わせていない。しかし、データの件数も 10000 件を超えると、

データの「ディスカバリー」を支援するためにも、可視化が強く望まれるようになる。また、検索結果もグラフ等で可視化できれば、研究者やオペラ事業者の調査にも役立つだろう。バージョン 2 では、D3.js 等のモジュールを用いたデータの可視化を試みる計画である。年表や地図等、利用価値のあるモジュールは多数存在する。

(5) 受動的利用者の利用促進

データベースのサービスにありがちだが、キーワードを入れて検索しないと何もデータが表示されないということがよくある。調査・研究に使うサービスなのだからといって、何も調べたいものがないひとは対象外、とするわけにはいかない。まずはサービスの内容を見てもらい、それを使うと何が調べられるのか、どのような情報を得ることができるのかを知ってもらうことにより、利用者に調査・研究のアイデアを与えることもまた、データベースの果たすべき役割であるといえるだろう。本サービスでは、とにかくサービスを使ってもらうために、トップページに「よく参照されるページ」のリストを表示してみたが、それではまだ受動的利用者の行動を促すには不充分だったようである。上述の可視化も手段のひとつに含めて、トップページからの離脱率の低下を図る。

(6) 外部データベースとの連携

本サービスは現在のところ外部との接続を行っていないが、既に OAI-PMH 準拠のハーベスト用インターフェイスを開発済みであり、国立国会図書館サーチにデータを提供する準備を進めている。提供するデータは、所蔵資料の書誌データに限られる。将来的には Linked Open Data への対応を進め、より多くの国内外のデータベースと連携したいと考えている。公演記録のデータについても外部データベースと連携できるだろう。

8. 結論

以上、昭和音楽大学オペラ研究所における「オペラ情報センター」の開発事例を紹介・分析してきた。最後に、この分析から得られた主要な知見を一般化して提示する。

- デジタル・アーカイブを公開する際、資料の保存のさらにその先に目的があるならば、その目的を明確にしたうえで、その実現を促すようにサービスを設計・開発すべきである。本稿で示したサービスでは、オペラ公演に関する資料のアーカイブを研究および実務に利用してもらうため、オペラ公演記録およびオペラ作品のデータベースと連携させて提供するという手法を取った。アーカイブを公開したけれど誰もアクセスしてくれない、と悩んでいる事業者があれば、自分たちが何を目的としているのかをいまいちど問い直してみ

るとよいだろう。

- デジタル・アーカイブの検索サービスを開発するにあたり、オープンソースのディスカバリー・システムの利用を、選択肢のひとつとして検討できるだろう。本稿では VuFind 2.0.1 を用いた開発事例を紹介したが、eXtensible Catalog^{*16} 等のシステムでも同様のことができると思われる。これはデジタル・アーカイブの開発に限らない。複雑な構造をもつデータベースを公開しようとするとき、RDB を直接検索するアプリケーションを作るよりも、また Solr 等の全文検索サービスを立ち上げてそのインターフェイスをいちから開発するよりも、データを MARC で記述して VuFind に投入してしまう方が簡単に、より高度なアプリケーションを作れる場合は多々あるだろう。
- オペラ公演記録を分析するための概念モデルとして、文献 [9] で提案された、バレエ公演記録の概念モデルが有効であることを検証した。また、オペラ公演を扱うために必要とされた拡張について詳述した。この概念モデルは、他の音楽・舞台芸術の公演分析においても利用可能と思われる。

参考文献

- [1] 総務省『デジタルアーカイブの構築・連携のためのガイドライン』<http://www.soumu.go.jp/main_content/000153594.pdf>. 2012.
- [2] 小島浩之「資料保存の考え方——現状と課題——」『情報の科学と技術』60(2): 42-48. 2010.
- [3] 上田修一・倉田敬子編著『図書館情報学』勁草書房. 2013.
- [4] 杉本重雄「知のデジタルアーカイブへの期待」『デジタル図書館』44:44-57. 2013.
- [5] 富澤浩樹・阿部昭博「資料の利活用を前提とした震災関連デジタルアーカイブの検討」『情報処理学会研究報告』情報システムと社会環境研究報告 2014-IS-129(3): 1-8. 2014.
- [6] 増井敬二著／昭和音楽大学オペラ研究所編『日本オペラ史～1952』水曜社. 2003.
- [7] 関根礼子著／昭和音楽大学オペラ研究所編『日本オペラ史 1953～』水曜社. 2011.
- [8] H. Kawano. “Digital Archive Systems Using CMS and Gallery Tools—Implementation of Anthropological Museum—.” *International Journal of Heritage in the Digital Era* January 2012 1:125-130.
- [9] 海野敏・高橋あゆみ・小山 久美「舞台芸術のための情報組織化手法の開発～バレエ情報総合データベースの設計と試作～」『情報処理学会シンポジウム論文集』2011(8): 205-210. 2011.
- [10] 岩崎陽一「学術データベースとして、消費者サービスとして—図書館用コンテンツ配信サービスの二面性」『情報処理学会研究報告』2014-CH-102(7): 1-6. 2014.

*16 <http://www.extensiblecatalog.org/>