ディスカバリーサービスのさらなる「日本化」を目指して

飯野 勝則（佛教大学図書館専門員）

１．はじめに

皆さんこんにちは。佛教大学図書館の飯野と申します。本日はこのような場を頂戴いたしまして、ありがとうございます。

今、巷でディスカバリーサービスというのが、取り沙汰されています。ある程度の普及もしてきました。しかし、その中身については、分からないことが多いという学校も少なくないと聞いています。

ですので、今日は「ディスカバリーサービスのさらなる「日本化」を目指して」という題名で、ディスカバリーサービスとは一体どういうものか、日本で本当に使えるものにするにはどうしたら良いかを、日外アソシエーツのコンテンツとの関わりを含めてお話しいたします。

皆さんと意見を交換できればと思っています。よろしくお願いいたします。

２．佛教大学について

先ずですね、佛教大学は京都にございまして、関東や東北の大学から見たら遠い所にあるために、良くご存じない方もいらっしゃるかと思うので、一通り説明させて下さい。

佛教大学は、1912 年に開学しています。今年 101 年目です。浄土宗が母体となっております。京都市と南丹市にキャンパスを持っていまして、仏教学部をはじめ、保健医療技術学部など 7 学部体制でございます。学生数は、通学課程が 7,000 人位。それと共に、通信課程にも力を入れていまして、14,000 人位います。

図書館の蔵書数は、大体 97 万冊。年間の開館日数が 316 日と、比較的開いている図書館ではないかなと思っています。2011 年 4 月から、日本で初めてディスカバリーサービスの Summon を導入させて頂きまして、今日に至っている次第です。

３．ディスカバリーサービスとは？

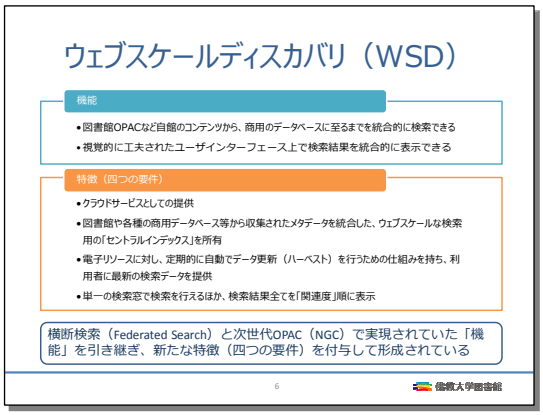
ディスカバリーサービスについて、ちょっとまとめてみます。名前は最近本当によく聞かれるようになってきました。ただですね、その実態は曖昧としている所がありますので、その辺に触れようと思います。また、すごく深い思想を秘めたシステムだということも述べたいと思っております。

さて、ディスカバリーサービスって一体何なの、と聞かれた時に見解がいろいろ存在しています。ここでは 3 つに分けて提示しています。先ず 1 番目。ディスカバリーサービスとは、次世代 OPAC のことである。次世代 OPAC（NGC）とは、Next Generation Catalog という形で、一時期随分取り上げられました。それじゃない、ディスカバリーサービスとは、ウェブスケールディスカバリと呼ばれるような、新しめの次世代 OPAC のことだ、という人もいます。更には、ディスカバリーサービスとは、ウェブ検索エンジンに決まっている、という見解もある訳です。

ディスカバリーサービスというと、皆さん 3 つ位のイメージを持たれるんですけど、最近の図書館業界としては、2 番目のウェブスケールディス

カバリを指すことが多いです。4 つほど製品名が並んでいますが、通称BIG4といいます。今日は、これを中心に話を進めたいと思います。ディスカバリーサービスといった場合には、ウェブスケールディスカバリを指している、というコンセンサスで聞いて頂ければと思います。

では、ウェブスケールディスカバリ、略称 WSD がどういうものかを話させて頂きたいと思います。



機能として、こういうようなものを持っています。先ず、図書館 OPAC など自館のコンテンツから、商用のデータベースに至るまでを、統合的に検索できる。それからもう一つ、視覚的に工夫されたユーザーインターフェース上で、検索結果を統合的に表示できる。視覚的に工夫されたとは、非常に見やすいといえば良いですね。今までの OPAC と違って、直感的で分かりやすい画面を持った検索システムということになります。

特徴として、4 つの要件があります。先ず、クラウドサービスとして提供されていること。大学や図書館の中にサーバーを設置することなく、Gmail のように何処か遠い場所にある、インターネット上のサーバーを利用して提供されている。次に、図書館や各種の商用データベース等から収集されたメタデータを統合した、ウェブスケールな検索用のセントラルインデックスを所有している。インデックスは索引のようなものですが、いろいろな商用データベースや OPAC などから集めたデータを一つの井の中に入れて、自分たちだけの索引、セントラルインデックスを作りそこに検索をかける、というようなシステムです。更に、電子リソースに対し、定期的に自動でデータ更新、ハーベストを行うための仕組みを持ち、利用者に最新の検索データを提供している。そして最後、単一の検索窓で検索を行える他、検索結果全てを関連度順に表示できる。こういった特徴があるといわれています。

これからですね、横断検索（Federated Search）という、串刺し式の検索システムや、次世代 OPAC で実現されていた機能を引き継いで、新たに 4 つの要件を付与して形成された、新しいシステムである、ということがいえます。画面で表現すると、右側にデータベース、真ん中にウェブスケールディスカバリのサーバーがあります。ウェブスケールディスカバリの場合、各データベース A、B、C、D から事前にデータを収集して、ウェブスケールディスカバリのサーバーに溜め込みます。「Hawaii」と検索する時には、ウェブスケールディスカバリのサーバーに対して、大学内或いは大学外から検索する、ということになります。つまり、予め「Hawaii」というキーワードを収集したデータを用いて作成したセントラルインデックスに対して検索を行い、その結果を表示するということです。従って、この検索を行っている段階では、A から D までのデータベースとは全く関係がありません。検索結果からデータベース部分にリンクなどして、データベースに飛ぶことはあります。しかし、検索の段階で A から D に対してのアクセスは発生しない、ということになるのです。

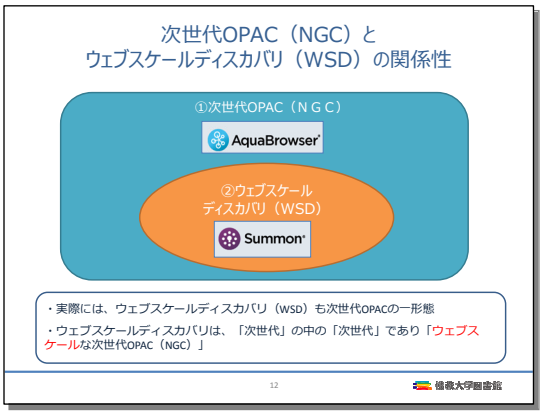
では、横断検索と何が違うのか。これは、かつて本学で使っていた、横断検索の画面になります。同じようにキーワードを一つ入れると、検索結果を一つの画面に表示するシステムでしたが、今見たのとは全く別の動きをしています。キーワードを入れると、各データベースに飛ばして、検索結

果をまとめて表示するだけのシステムです。増え続ける、特に英語圏のデータベースを、利用者に効率的に使わせたいという必要性から生まれたシステムでございます。画面的には今のとちょっと似ていますが、内容的には全く違います。仕組みは、大学内のインターネットから主に検索をするんですが、真ん中の横断検索サーバーへキーワードを投げます。そうすると、認証を通して各データベースに改めてキーワードを投げ、収集してきた結果を合成して表示するというものです。認証が早かったり遅かったりで、検索結果がばらばらになるといった弱点を持つシステムといえます。

次世代 OPAC というものも、ウェブスケールディスカバリには欠かせない要素の一つであります。ディスカバリーサービス以前の、次世代 OPAC の画面でございます。普通の OPAC と違って、書影や関連するキーワードを図で表示したりとか、ファセットといわれる絞り込みの技法を駆使する項目を持たせたり、視覚的には非常に優れたデザインではあった。そういうものが、ウェブスケールディスカバリを構成する要素の一つになってます。

横断検索や次世代 OPAC がそのまま使えなかったのは、弱点があったからですね。弱点を整理すると、横断検索は基本的に各データベースの反応順で検索結果が表示される。要するに、早いデータベースから返ってきた結果は上に表示されるけれども、遅いデータベースの場合は後ろに回ってしまう。更に、検索能力が各データベースの検索システム能力に依存していました。つまり、検索能力が高いデータベースであれば、いろいろな結果が返ってくるけれど、検索能力が低いデータベースだと、あまり良い結果が返ってこない。でも、それは利用者からは判断が出来ません。更にですね、キーワードを各データベースに投げて、そこで検索させた結果を引っ張ってきたために接続が非常に不安定だったんですね。例えば、あるデータベースがダウンすれば、そのデータベースからは結果が返ってきません。そうすると、検索結果が不足したものになる。気付けば良いんですが、気付かない場合も当然ある訳です。また、検索に対する反応もやっぱり遅いものがあります。

次世代 OPAC も、雑誌の他に論文項目の検索や、データベースの内容を表示させることも出来たのですが、そのためには横断検索システムを裏で動かさなければならなかった。ということで、横断検索の弱点は、次世代 OPAC の弱点でもあった訳ですね。これでは良くないだろうと、先程いった 4 つの要件、特徴を持ったシステムとして、ウェブスケールディスカバリが開発されました。



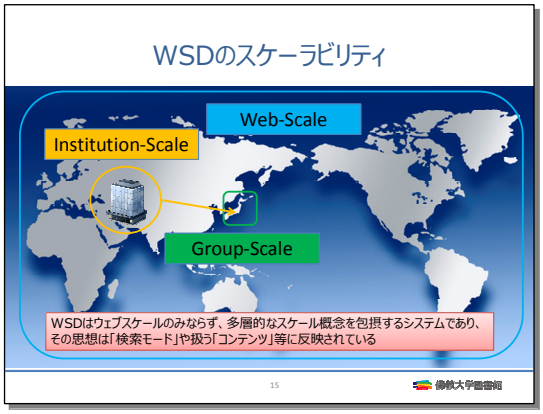
ウェブスケールディスカバリと次世代 OPAC の関係ですが、ウェブスケールディスカバリといっても、次世代 OPAC の一形態に過ぎません。つまり、ウェブスケールディスカバリとは、あくまでも次世代 OPAC の中の、ウェブスケールな次世代 OPAC であるとの位置付けで捉えることができる訳です。

４．スケーラビリティ

ウェブスケールとは何かという所と、スケーラビリティという考え方が、このシステムには盛り込まれているので、その話をさせて下さい。

そもそも、ウェブスケールとは何か。ウェブス

ケールディスカバリとはいいますが、図書館用語としてはウェブスケールディスカバリ以外にはあまり聞かないんですね。元々情報工学の言葉らしいんですけれども。図書館の場合には、図書館本来のスケールである、インスティチューションスケールの対義語として、ウェブスケールは存在する、という捉え方をします。要するに、インスティチューションスケールとウェブスケールは、相反する概念を包括するということです。



これを図として考えてみますね。ウェブスケールディスカバリがいう所のウェブスケールとは、全世界に周く広がる学術情報を繋ぐスケールです。それに対して、図書館一つ一つを、インスティチューションスケールとして捉えます。更にスケーラビリティという捉え方では、グループスケールという言葉があります。これはインスティチューションが幾つか集まった団体、例えばコンソーシアムとか地域とか、そういったものを繋げるスケールでございます。

インスティチューションスケールは、ローカルです。ローカル OPAC とかいいますが、このローカルにあたります。更にいうと、ウェブスケールとは、グローバルにあたります。

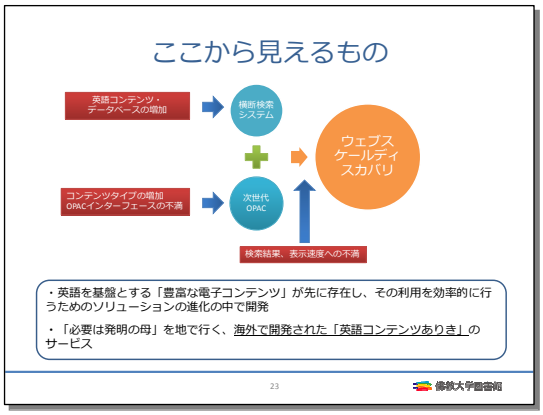
先程から、ウェブスケールディスカバリという話をしていますが、実際にはウェブスケールだけを扱うシステムではありません。ウェブスケールディスカバリは、インスティチューションスケールやグループスケールといった多層的なスケール概念を包摂するシステムであり、その思想は検索モードや、扱うコンテンツ等に反映されている、ということになります。これは、WorldCat Local の検索モードです。Libraries Worldwide と Summit Libraries、Portland State University と書いてありますが、検索の範囲をスケールに応じて決められるということです。Libraries Worldwide はウェブスケール、Summit Libraries は、彼らが属しているグループスケール。一番下に、自分たちの大学の、インスティチューションスケールがある。こういった、データの範囲をスケーラビリティで設定できる、という特徴を持っています。

更に、コンテンツを分析してみると、スケール毎にいろいろなデータベースの種類に分けられます。一番上のウェブスケールであれば、商用のデータベース。契約があれば、世界中で利用出来るものですね。更に、グループスケールであれば、地域コンソーシアムが作成したデータベースとか、グループとしての利用が中心となるもの。それから、インスティチューションスケールでは、図書館や大学が作成したデータベース、或いは OPAC といったものを利用出来る、ということになります。つまり、幾つものスケールが包摂されて、中でいろいろと切り替えが出来て、取り込んでいるようなシステム、これがウェブスケールディスカバリだ、ということになる訳です。

5．ウェブスケールディスカバリ

こういった特徴を持つウェブスケールディスカバリ、現状では有名なもの 4 つございます。これを BIG4 といいます。ウェブスケールディスカバリの先駆けとして、2007 年にリリースされたのが、OCLC の WorldCat Local（WCL）です。続きまして、2009 年に Serials Solutions 社からリリースされた、Summon。更に、2010 年にリリースされた、EBSCO Discovery Service（EDS）。

そして、4 番目が Primo Central で、2010 年に Ex Libris 社からリリースされました。何れも名前を見て分かる通り、日本のシステムではございません。



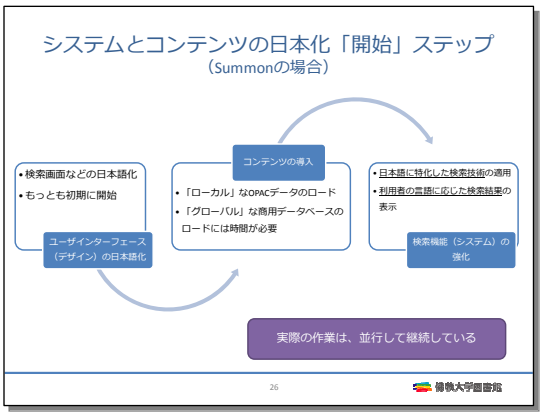
さて、BIG4 の存在を考えた時に、何が見えてくるのかを、少し考えてみたいと思います。この背後にあったのは、英語のコンテンツやデータベースの増加をどうにかしたい、という要求でした。そこから横断検索システムが出てきた。加えて、本、雑誌、論文、目次データといった、コンテンツタイプが増加してきた。それを今までの OPAC のインターフェースでは、十分に表現できなかった。それは困るということで、次世代 OPAC が出来てきた、ということになります。しかし一方で、検索結果や表示速度に対する不満が出てきたので、それを補う形で生まれたのが、ウェブスケールディスカバリということになる訳です。

つまり、ウェブスケールディスカバリとは、英語を基盤とする、豊富な電子コンテンツが先に存在し、その利用を効率的に行うためのソリューションの進化の中で開発されたシステムだ、といえると思います。必要は発明の母、といういい方が正しいか分かりませんけど、こういったものを地で行く、海外で開発された英語コンテンツありきのサービスであった、ということを前提で考えて頂ければ幸いです。

６．日本化

ウェブスケールディスカバリを日本化して行くには、どういったステップが必要なのかを考えてみたいと思います。

日本化の前提として、先ずは、現実の認識。BIG4 をはじめ、ディスカバリ製品は海外製品しかなかった。とにかく継続的な日本化への取り組みは避けられない。一方で、日本語というのは、ディスカバリベンダーにとっては、世界中のユーザーが使う言語の一つでしかない。英語のような、スペシャリティがある言語ではございません。なのでディスカバリベンダーの全面協力はなかなか得にくいところもある。次に、ディスカバリーサービスが生まれた背景も、明らかに日本とは違います。豊富な英語の電子コンテンツとか、そういったものは日本には全然ない。その上で、これまでの図書館システム、OPAC などとは異なる 4 つの要件を持っています。海外製品であるがゆえに、日本的感覚でみると若干違和感のある対応や仕組みも持っている訳です。海外製品であることの文化的な摩擦を甘受することも必要です。では、最終的に日本化をどのように進めるのか。システムとコンテンツの両面から、とにかく日本語の学術情報を十分に扱えるようにすることが必要である、と。その上で、日本的な細やかさが求められる部分をどう扱うか考えなければいけない。これが日本化の前提でございます。



そういった前提の上で、ウェブスケールディス

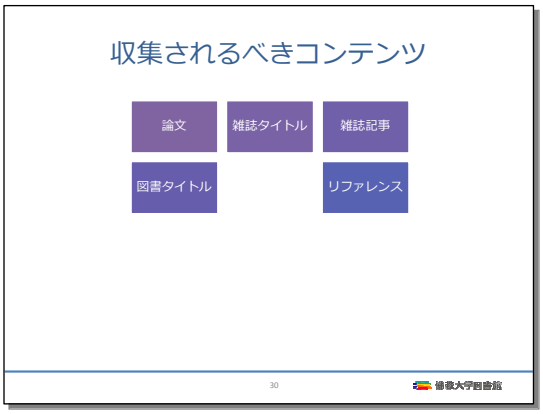
カバリの状況をもう一回考えてみたい。これは、うちで使っている Summon の、システムとコンテンツの日本化の開始、というステップです。最も初期には、ユーザーインターフェース、つまりデザインですね、サイトデザインの日本語化があった。サイトなのかサービスなのか、或いはディスカバリなのか分かりませんが、そういったものを訳すのかどうか、とかいった所から入ります。その次に、自分たちが持っているローカル、インスティチューションスケールの OPAC データのロードや、日本語コンテンツの導入。グローバルな商用データベースのロードには、時間が必要です。その先に、検索機能、システムの強化が出てきます。これは開始ステップということで、大体左側から始まった訳ですけれども、実際の作業は今も並行して継続しているのが、おそらく全てのウェブスケールディスカバリベンダーの状況だと認識しております。

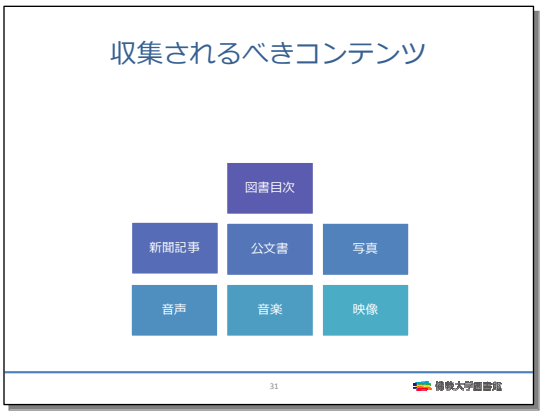
一番考えなければいけないのが、一番右側の日本語に特化した部分。日本語に特化した検索技術とは何かというと、これが例の一つですね。形態素解析を持たせる。形態素解析とは、「京都」という検索結果を示したい時に、「東京都」が入ってくるのはおかしい訳ですね。そういったものが入ってこないようにする技術です。ただですね、新語などが出てきた時には、対応が難しいこともあります。例えば、リニア新幹線の駅に、「東京都駅」という名前が付いていました。「東京都駅」と検索しようとすると、多分「東」、「京都駅」とかに分割されて、検索結果がおかしくなります。こういった、日本語的にこれとこれは似ているが全く違うだろう、というのをちゃんと判別して表示させる仕組みを持たす必要が最低限あります。

それから、ディスカバリーサービスは、世界中で同じシステム、セントラルインデックスを使っております。そうすると、中に中国語のコンテンツが大量に入っている。だからといって「夏目漱石」と検索した時に、検索結果で中国語コンテンツの「夏目漱石」がわらわら出てこられても困る訳です。最初はそういうこともあったのですが、今は、利用者がどんな言語のインターフェースを使っているのかをみて検索結果に反映させるような、細やかな気遣いをしています。これが現状の画面ですが、左側はインターフェースが繁體中文、中国語ですね。「夏目漱石」と検索すると、中国語の「夏目漱石」の検索結果がいっぱい出てくる。右側は、インターフェースが日本語。この場合「夏目漱石」と検索すると、日本語のコンテンツが上に行く。こういった、利用者の望む結果を返す仕組みも必要とされている訳です。

7．コンテンツの日本化

今のは検索の話だったんですが、今度はコンテンツの日本化を考えてみたいと思います。

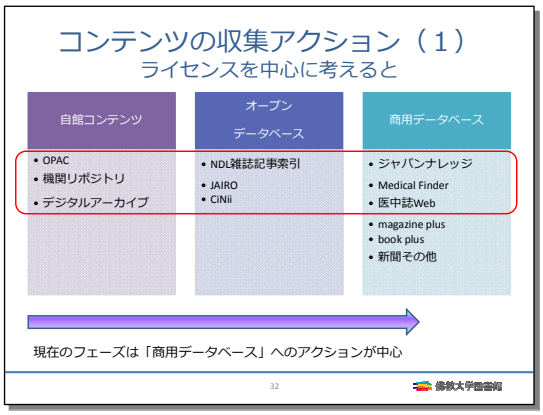




欧米とかのウェブスケールディスカバリを考えると、ここに出ているマスが、おおむね収集さ

れるべきコンテンツになりますね。一番手前の「論文」から一番下の「映像」まで、非常に幅広く、これらの検索結果が返ってくるのが望ましい。では、日本の現状はどうなのか。「論文」、「雑誌タイトル」、「雑誌記事」、「図書タイトル」、「リファレンス」、大体これだけです。さすがに今後どうするのかを考えなければいけない、ということになります。

コンテンツの収集アクションを、ライセンス中心に考えた場合、左側の自館コンテンツに関しては問題がありません。一番最初に手が着けられる。その後、オープンデータベース、NDL の雑誌記事索引、JAIRO、CiNii といったものは、今は取り込みやすくなっている。すると、現在のフェーズでは、一番右側にある商用データベースへのアクションが必要になります。赤色で囲っているコンテンツは、今現在いろいろなウェブスケールディスカバリで取り込まれて利用されているものです。Summon でいうと、ジャパンナレッジとか Medical Finder、医中誌 Web が対応している。日外アソシエーツの magazineplus はまもなく対応されますが、今のところは bookplus と共に囲みから外してあります。今後の課題ですね。



スケーラビリティで考えると、現在のフェーズは、グローバルとかウェブスケールというような、ディスカバリ対応の難易度が一番高い部分でのアクションを中心に行っているといえます。各ディスカバリベンダーともおそらく同じだと思います。

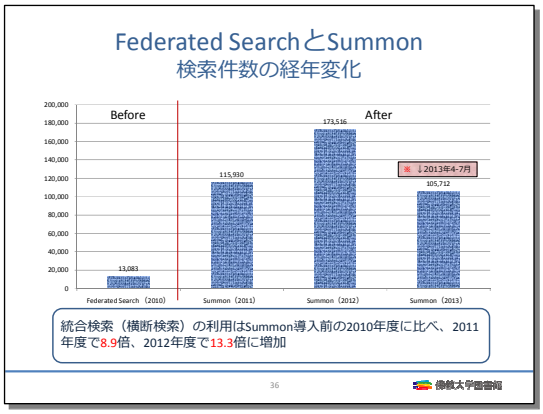
何故難易度が高いのかというと、いろいろな課題があるからです。その一つが、コンテンツベンダーからみた場合の課題です。コンテンツベンダーからみてディスカバリーサービスって、いろいろな不安があります。心理面でいったら、海外のベンダーなので、本当に大丈夫なのかという信頼の欠如や、メタデータの行方がどうなっちゃうのか不安だとか、そういったことがあります。ライセンス面では、そもそもポリシーの問題で、うちはそういうこと出来ないとか、データを作った契約の時には、そういったことはしないといっているんだから契約上無理、みたいなこともある訳です。金銭面では、提供した所で、どうやって収益を確保するの？とか、データの提供に掛けるお金はないよとか、そういう不安を持っている場合もあります。技術面では、提供したいけど、どうしたらいいのか分からないとか、そんな技術は難しくて自分たちでは対応が無理という場合もある。

ただ、こういった不安は当然あるんですけど、解決できる課題の場合、コンテンツベンダーに対して、図書館が共存共栄できるメリットを提示し、迷いを取る手助けをする必要があるんじゃないか、と最近思うようになりました。ですので、こういう話をしているんですけれども。

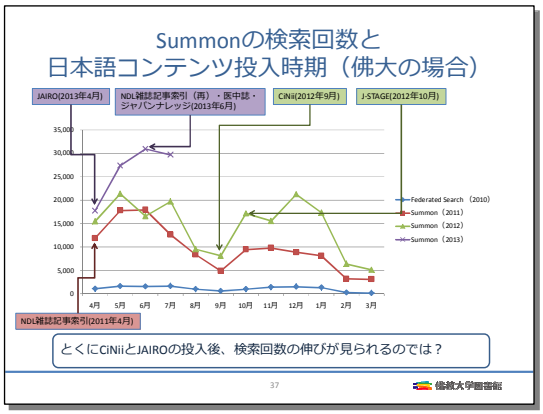
ディスカバリーサービスにコンテンツベンダーからデータを提供させると、すごく良いことが起こります。日本語コンテンツの場合はとても顕著で、勝手に Win-Win 関係だと思っています。その説明をさせて下さい。多分今日コンテンツベンダーの方が来ていると思うので、聞いて頂けると嬉しいです

その前に、本学のフェデレーテッドサーチ（横断検索）から Summon への流れで、いろいろな状況が起きたので、ちょっとその話させてもらいます。画面をみると、2010 年、2011 年の間で分かれています。ビフォーはディスカバリーサービス導入前です。ディスカバリーサービスの導入前

って、統合検索とか横断検索は年間 1 万 3 千件位しか使っていなかったんですね。ところがディスカバリーサービスを導入したら、11 万件になり、昨年は 17 万件と増えてきています。

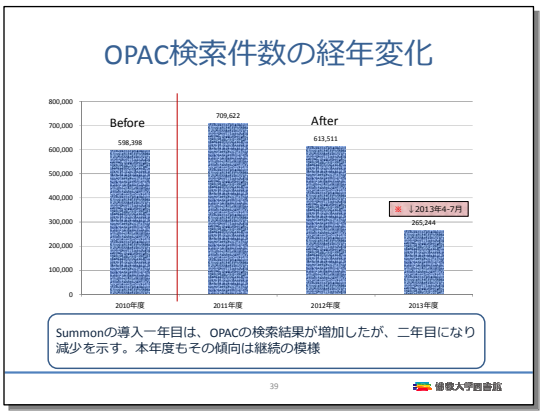


導入前に比べ 2011 年度で 8.9 倍、2012 年度で 13.3 倍というのが、現在の状況でございます。今年の状況を見るに、4 月から 7 月までの 3 ヶ月のデータで 10 万 5 千件でしたので、今年も去年の数値は超えるだろうと思っています。



Summon の検索回数がこんなに増えたのは、日本語コンテンツの投入時期が密接に絡んでいます。月別のグラフを示しています。一番下の青いものが、フェデレーテッドサーチ時代で、そんなに多くないです。2011 年に Summon が入って、この赤。その翌年 2012 年が緑。一番向こう、急速に上がっているのが、今年 2013 年のデータでございます。2011 年の時点では、OPAC は入っているんですが、外部のデータベースと結びつく主要な日本語コンテンツは、NDL の雑誌記事索引が唯一でした。2012 年 9 月、CiNii が搭載された瞬間に、アクセス数が急速に増えてきた。更にいうと、本学は他大学に比べるとちょっと遅かったんですが、2013 年 4 月に JAIRO を投入した所、一気にアクセス回数が増えた、と。JAIRO はご存じの通り、日本全国のリポジトリを集積している NII のデータベース。そのデータが流れ込んだことで、日本語の電子コンテンツが飛躍的に増えたということになります。6 月位になると、NDL の雑誌記事索引がもう一回入り、医中誌・ジャパンナレッジが入ってきて、アクセス回数が非常に増えた、ということになります。

当時の検索回数の増加は、CiNii と JAIRO が特に著しいということになります。ここ 4 年間の、CiNii のダウンロード推移でございます。CiNii が Summon に入った 2012 年 9 月から、CiNii のアクセスが一気に増えています。2013 年度も前年数よりも遙かに大きい数字で、グラフの方に出てきている。明らかに Summon と CiNii のダウンロード件数には関わりがある。つまり、ウェブスケールディスカバリである Summon に CiNii のデータが入ったことで、CiNii のアクセスが著しく増えたということが、ここからも分かる訳です。

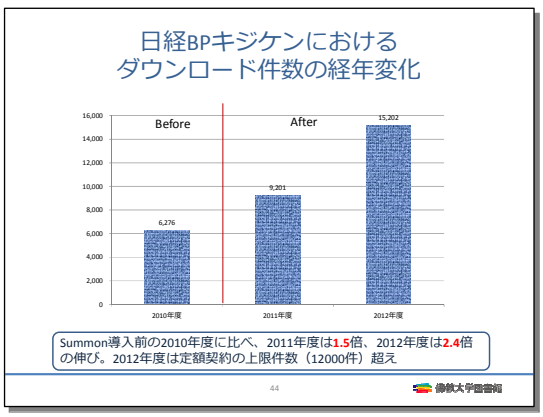


これは OPAC 検索件数の経年変化なんですが、おもしろい動きをしています。Summon 導入一年目の 2011 年度は、OPAC の検索結果は増加し

ています。しかし二年目は減少。今年も減少の傾向が出てきております。これは、一年目が特殊な環境であって、Summon の検索結果一覧で、本学の蔵書のデータが出てきた場合、これをクリックすると OPAC 上の詳細画面に飛んでいたんですね。これが検索としてカウントされていました。ところが昨年度から、Summon 自身に蔵書の詳細画面を持たせるようにしたんです。そうすると、Summon 上で、必要な情報が入手できてしまうので、利用者は OPAC へ行かないんですね。はっきりいえば OPAC の必要性が低くなって、OPAC への遷移が減少したということです。なので、ウェブスケールディスカバリが OPAC に代わる機能を持っているということがここからも分かります。

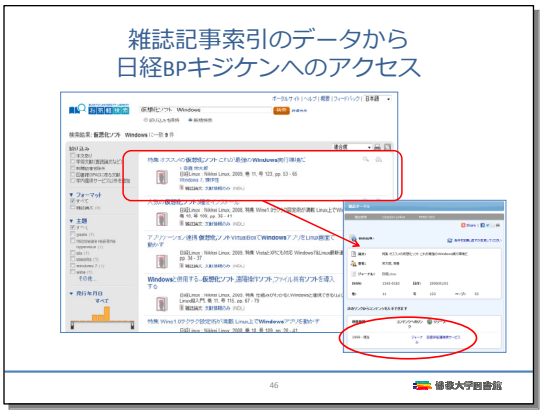
この現状を OPAC との利用比較ということで挙げてみたんですが、2010 年度はまだウェブスケールディスカバリが入っていない頃で、OPAC とフェデレーテッドサーチの割合で、横断検索は 2%しかなかったんですね。それが、初年度は 14%、昨年 22%、今年は現在の段階で 28%まで来ています。おそらく、今後もこういった傾向が続くんじゃないか、と。つまり、OPAC は、ウェブスケールディスカバリに少しずつ取って代わられる状況にある、ということが分かります。

8．日経 BP 記事検索サービス



さて、話がちょっとずれるんですが、一方的な Win-Win 関係というのも矛盾しているようで変ですが、棚からぼた餅的な話があります。これは日経 BP 記事検索サービス、キジケンと呼ばれるもののダウンロード件数を示したグラフです。キジケンは 2013 年 9 月の時点では、Summon にメタデータを提供していない、日本語データベースの一つです。キジケンのダウンロード件数の経年変化なんですが、Summon 導入前に比べて、2011 年度は 1.5 倍、2012 年度は 2.4 倍に増加しています。2012 年度は、定額契約の上限件数を超す 1 万 5 千件と、ものすごく増えました。

でも、おかしいですよね。メタデータを提供していないのに、何でここまでアクセスが増えるのか。一つ理由があります。Summon に昔から入っていた NDL の雑誌記事索引の採録対象には、多くの日経 BP 社の雑誌があります。このデータが Summon に流れ込んでいる。例えば「仮想化ソフト□Windows」で検索すると、日経の雑誌記事がバーッと出てくる。ここに NDL と書いてあります。これをクリックすると、本学の場合はリンクリゾルバというシステムを間に噛ませて、所蔵している日経 BP 記事検索サービスへのリンクが形成されるようになっています。つまり、利用者は日経 BP とは全く関係のない雑誌のメタデータをディスカバリーサービスで使って日経 BP の電子ジャーナルへのリンクを見つけ出し、それを通してアクセスすることが出来ている、ということになります。

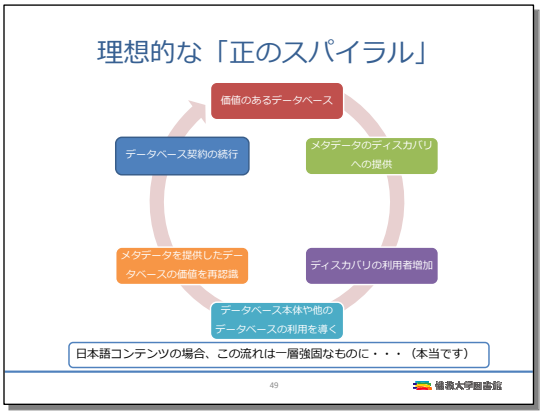


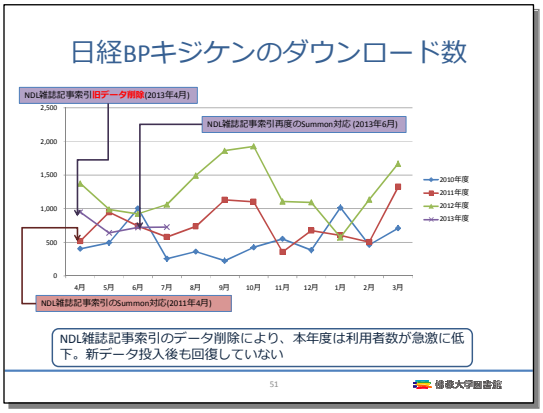
似たようなことは、冊子体でも出来ています。これは或るインド学の論文なんですが、これを検索します。すると、同じように雑誌記事索引のデータがリンクリゾルバへ出力されます。リンクリゾルバには、冊子の所蔵状況の「佛教大学図書館」と、電子ジャーナルの「Journal @rchive」、二つのリンクが成されます。「佛教大学図書館」を押すと、OPAC の詳細画面へと飛びます。つまり、OPAC で、所蔵している場所を見つけることができて、この雑誌を冊子で利用することもできる、ということになります。実際、Summon の NDL 雑誌記事索引のデータから、冊子の書誌へのアクセスが増え、冊子体雑誌の利用も増えるという、相乗効果が出てきたことも分かりました。

９．理想的な「正のスパイラル」

ここから分かることを、少し挙げてみます。Summon、ディスカバリーサービスですが、一度受容されると、利用者の資料探索行動に与える影響は非常に大きいです。間違いなくいえる。更に、日本語コンテンツの質と量というのが Summon、或いはディスカバリーサービスの評価に直結することも分かります。つまり、日本語コンテンツの質が良く、量も多ければ、利用者にとって非常に使いやすいディスカバリーサービスになるということです。もう一つ、ディスカバリーサービス、Summon にデータを提供した CiNii などの一次資料データベースの利用は、明らかに増えています。次は回りくどいのですが、重要なことなのでいわせて下さい。雑誌記事や論文タイトルなど、価値のあるメタデータを提供できるデータベースが Summon にコンテンツを提供することで、他のデータベースに多大な貢献をもたらします。ということで、それを手放すのは、非常に難しくなります。magazineplus も手放すことが難しいデータベースになることが予測可能です。つまり、これは図書館とそのステークホルダにとって、重要な共存共栄をもたらす、正のスパイラルになると考えています。

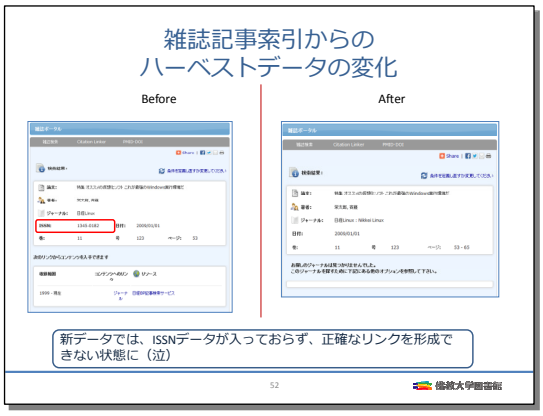
正のスパイラルとは何か。先ず、価値のあるデータベースが、ディスカバリへメタデータを提供します。そうすると、ディスカバリの利用者が増える。ディスカバリの利用者が増えると、そのデータベースや他のデータベースの利用も増える。メタデータを提供したデータベースの価値は高くなる。そうすると、そのデータベースは止められなくなり、契約は続けられる。それは価値のあるデータベースになる。こういうスパイラルが生まれてくることになります。日本語コンテンツの場合、間違いない。非常に顕著だし、強固だと私は考えています。というのは、それを実証するような例が実際にあったからです。





これは、先程挙げたキジケンのダウンロード数ですが、絶好調だったはずのダウンロード数が、若干減っております。この裏には、いろいろなことがありました。先ず、2011 年 4 月に対応した

NDL の雑誌記事索引が、2013 年 4 月に理由があって本学のデータが削除されたんですね。削除された瞬間、がくっと減りました。6 月に再度入ったんですけれども、まだ回復していないというのがこのグラフです。さっきいっていた話と矛盾するじゃないか、という訳ですけれども、秘密があるんです。それは、雑誌記事索引からのハーベストデータが変化していたことが理由です。



ビフォー側は、アクセスが絶好調に増えていた時です。アフターは、雑誌記事索引が入り直した 6 月位のデータなんですけど、ISSN のデータがなかったんですね。ということは、リンクリゾルバから正確なリンクが作れなくなった。利用者はディスカバリーサービスで該当の記事のデータを検索したにも拘わらず、その一データである日経 BP のサイトに行き着くことの出来ない状況が生じていたことになります。

もう一つ分かることがあって、これは NDL の雑誌記事索引のデータから、本学の OPAC に飛んだ件数を出力したデータ。2010 年、青色ですね。2011 年は赤。フェデレーテッドサーチ時代の 2010 年は、そんなに多くなかった訳です。それが、2011 年に Summon を導入して NDL の索引が直接使えるようになったことで、利用者が非常に増えました。2012 年、ある所までは 2011 年と同じ流れで来ていたんですが、この部分でがくっと減りました。それは何かというと、CiNii の Summon 対応です。CiNii のデータがディスカバリーサービスに流れ込んだ所で、電子で使えることに気付いた利用者が随分いるようです。今まではディスカバリーサービスを使っていて、CiNii の電子データはなかったので冊子を使っていた利用者が、いきなり使わなくなった、と。今年度、更に旧データを削除して、雑誌記事索引がなくなるといったことがあり、見てみると、2010 年の Summon 導入前と変わらないようなアクセスになって、冊子の利用が激減しました。これは非常におもしろい結果です。

ISSN がなくなったそもそもの原因は、API の仕様変更です。国立国会図書館は何も悪くないんですよ。PORTA から NDL サーチに変わった時に仕様が変更されて、その対応に Summon 側が手間取ったということです。これは NDL サーチの「日経 Linux」の検索結果ですが、ちゃんと ISSN 持っているんですね。NDL サーチの結果もね。更にいうと、API で取得したデータも持っているんです。URI 形式で持っており、はっきりと ISSN と書いていなかったので気付かれなかったのか、いろいろ理由があると思うんですけど、Summon への対応が遅れ、消えてしまったということです。とはいえ、データ上に ISSN が残っているわけで、何とかなりそうと思っていたんですが、実際どうにかなりました。先月、9 月 13 日から ISSN が戻ってきており、推移を注意深く見ている所です。なんと ISSN が復活した後、いきなり 1,041 件までダウンロード数が戻ってきております。復活して半月なので去年には及んでいませんが、本年度初めて 1 千件を超えまして、去年までは行かないにしろ、ここから先は、ある程度明るい見通しになるだろうと思っています。

ということで、我々は ISSN データを含む雑誌記事索引を、正のスパイラルの枠組みで非常に評価しているということになります。当然、同じようなことは、他のデータベースにもありうるでしょう。雑誌記事索引は商用データベースではありませんが、商用データベースであっても、スパイ

ラルとしての評価の過程に変化はないと思います。特に magazineplus ならば、雑誌記事索引と類似する特性があり、同じような評価も期待できるでしょう。メタデータを考えると、これ以上の波及効果を生み出せる可能性があるとも考えています。更に、商用データベースの契約者に、Summon 上でのみ、例えばメタデータの利用を許諾すれば、商用データベースの契約は間違いなく維持されると思います。一つ重要なことなんですが、日本語コンテンツが入っていても、ユーザビリティを確保できないと、Summon 或いはディスカバリーサービスは、十分な効力を発揮できないのが現実です。ISSN が消えてしまい、それに対応する手立てを持っていなかった。そしたら利用者が激減してしまった、というのはユーザビリティを確保できなかったからで、そういった所を考えていけば、今後明るい現実が見えてくると考えております。因みに、magazineplus は 2014 年の 4 月から、ディスカバリーサービスでメタデータ利用が可能になるということで、私どもも非常に楽しみにしています。

10.API によるユーザビリティ向上の試み

日本語対応から日本化へ、ということで、API による Summon のユーザビリティ向上の試みについてお話しさせて頂きたいと思います。

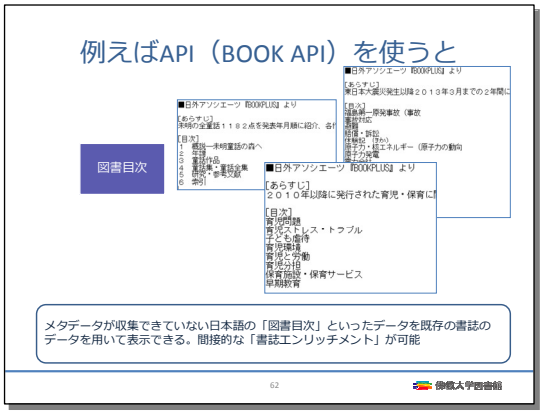
日本語対応における現実と課題を振り返ってみたいのですが、ハーベストを受け入れていない商用データベースが存在しているのは、事実でございます。更に、NDL-OPAC や OPAC のメタデータの内容は周知をされておりますけれど、これは従来からの目録規則に準拠した部分にとどまり、決して豊かなメタデータではございません。はっきりいって、まだまだ従来型のシンプルなメタデータである、ということになります。日本語によるディスカバリーサービスでの図書検索というものは、OPAC における図書検索と遜色ないものであってほしい。つまり、OPAC で可能なことは、ディスカバリーサービスでも可能なことが最低限必要だと思います。更に、ユーザビリティに関わることですが、ユーザーインターフェースに、日本的な細やかさを求めたいと思っています。こういった現実をみると、全てを集中することが出来ないのであれば、API を用いて外部情報資源を上手く利用することも必要だと考えられます。

API って最近よく聞きますね。API とは、あるソフトウェアが外部のソフトウェアに対して提供するインターフェース、接続規格でございます。データベースのデータを外部から呼び出す時とかに用いられます。これは CiNii Articles の API と呼ばれるものなのですが、「論文検索の OpenSearch クエリは以下の形式です」とあって、何か URL が書いてありますね。この呪文の通りブラウザ上に打ち、「search?q=ディスカバリ」と書きました。「q」というのは、フリーワードを指定するパラメータで、「q=ディスカバリ」と入れると、CiNii Articles で「ディスカバリ」というキーワードの検索結果を返してくる。これはあくまでも CiNii の API の一機能に過ぎません。実際には、いろいろな手法で外部からデータを呼び出せるわけです。

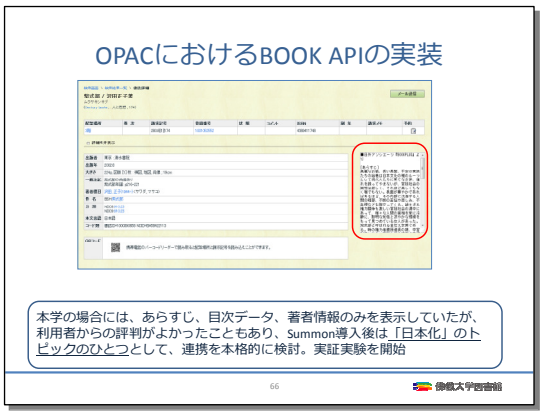
じゃあ、API を使うと、どんなことが出来るのか。例として、日外アソシエーツの BOOK API を出してみたい。これを使うとメタデータが収集できていない、日本語の図書目次というデータを、既存の書誌データを用いて表示することが出来るようになります。つまり、間接的ではありますが、書誌エンリッチメント、豊かなメタデータを作ることが可能になる。それを行うのが、BOOK API です。本学の方で、BOOK API とウェブスケールディスカバリの実証実験を行いました。その話を少しさせて下さい。

そもそも BOOK API とは、日外アソシエーツの BOOK データベースの内容、目次・要旨、著者紹介情報を表示するサービスでございます。表

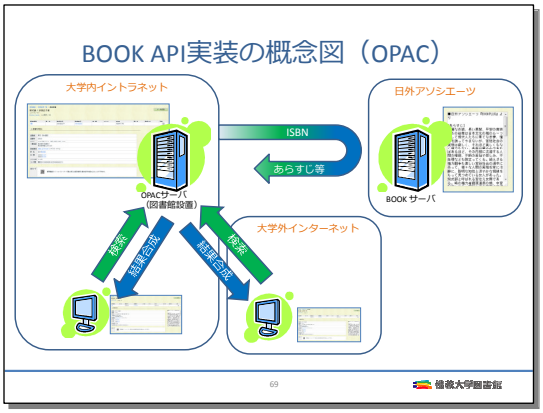
紙の書影も出せます。ISBN を検索キーとして、BOOK データベースからメタデータを XML 形式で、表紙書影を JPEG 画像で呼び出せます。



一番重要なのは、著作権が処理されていること。書影を出すのに、他の大きな書店とかのサイトのデータを利用したりしますが、それって本当に良いのかどうか。図書館員としては、ライセンス的な問題が気になります。日外アソシエーツの BOOK データベースには、ライセンス的な心配がありません。データは XML と呼ばれる形式で、title、youshi、mokuji、それから著者の情報が呼び出せる。更には、画像も取れます。



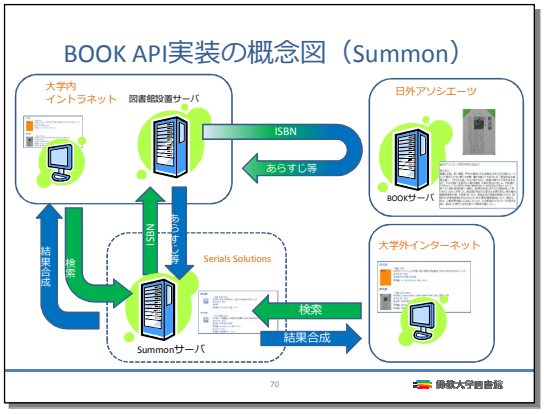
これは本学の古い OPAC の画面なんですけれども、あらすじ、目次データ、著者情報のみを表示していました。大学によっては、書影を出している所もございます。利用者からの評判が良くてですね、Summon を入れた後も、こういうのがあった方が良いんじゃないの、ということで、連携を本格的に検討しました。それで実証実験を開始した経緯がございます。ただ、OPAC とディスカバリーサービスは仕様が違うため、連携がなかなか難しいんですね。何が違うかというと、OPAC は学内にサーバーを設置する、オンプレミスと呼ばれる運用が一般的でございました。サーバー自体も所有や管理が可能なので、ベンダーにとっても非常に自由度が高いものだった。画面のカスタマイズとかも考えれば何とかなるよね、という所が結構あったんですよね。ところが、ディスカバリーサービス、本学の場合 Summon ですが、クラウドだった訳ですよ。クラウドということは、大学はサーバーを所有していないため、自由に触れられるような環境ではございません。ユーザーインターフェースもある程度共通化されているので、画面カスタマイズの自由度も高くない。更にいうと、自分たちのサーバーでもないし、セキュリティなどの面で、非常に考慮すべき要件が結構あります。Summon の方が、圧倒的に適応が難しい部分がある訳ですね。こういった前提での実証実験であったとお考え下さい。



BOOK API 実装の概念図でございます。OPAC の場合は、シンプルです。誰かが大学内の OPAC サーバーに対して検索を行う。検索を行うと、ISBN をキーにして、日外アソシエーツのサーバーへ API でデータを呼び出しにいきます。呼び出されたあらすじのデータを OPAC サーバーに返し、合成をして表示させるということになりま

す。

ところが、ディスカバリーサービスは経路が若干複雑です。大学内、大学外のインターネットで検索を行うと、本学の場合は先ず Serials Solutions 社の Summon サーバーを検索しに行きます。すると、Summon サーバーから、本学の図書館の設置サーバーへ ISBN を飛ばします。今度は図書館の設置サーバーが、ISBN を使って日外アソシエーツのサーバーへデータを引き出しに行き、あらすじを取得する。図書館の設置サーバーが取得したあらすじを Serials Solutions 社のサーバーへ飛ばし、そこで結果を合成して大学内外のインターネットで表示させる仕組みです。



検証結果は、決して悪いものではございません。図書に関する、目録規則準拠のメタデータを補う情報、例えば書影やあらすじが表示できるようになっております。本学の Summon の場合、書影が出ているんですけれども、日外アソシエーツの BOOK API を使って呼び出したデータでございます。

但し、技術的な課題もございます。先ず、ブラウザに由来する課題。Internet Explorer は、いろいろなセキュリティポリシーを持っております。多分そのせいなんですけれども、目次、あらすじの表示が難しい状況です。Firefox なんかも頻繁にアップデートしておりますので、その辺の懸念が出てきている。それから Summon 側の話なんですが、クラウドサービスでございますので、ベンダー側でのユーザーインターフェースのアップデートも非常に頻繁でございます。良いことでもあるんですが、こちらから組み込むという所で懸念がある。アップデートによって仕様や設計思想が変更された場合には、一からこちらが用意しているスクリプトを書き直して、設定をし直す必要が出てくる所が問題ではないかと。例えば、Internet Explorer と Firefox や Chrome のブラウザの比較すると、IE では、「あらすじ・目次」というのが表示できていません。回避策を模索中でございます。それから、Summon 側の仕様変更も入ってきて、今度新しい 2.0 がリリースされます。とっても楽しみなんですけれども、対応するスクリプトを急ピッチで頑張らなきゃいけない状況でございます。正式運用に向けて、Internet Explorer への完全対応、Summon2.0 への対応スクリプトをどうするのか、完全に解決しなければいけないと私どもは思っております。Serials Solutions 社にお願いをしておりますし、前向きな返答も頂いておりますので、多分何とかなるんじゃないかな、とは思っておりますが、対応に時間は必要かもしれません。

ということで、ディスカバリーサービスの日本化のステップは、うちの場合ですが、Summon に関してはある程度の見通しがついたといえます。ただですね、いろいろな注意点がございます。組み込みを認めるか否かは、ディスカバリーサービスベンダーの考え方によるので、事前の協議が必要でしょう。因みに、Summon とか EBSCO Discovery Service で対応できることは、既に確認済みであるということです。更には、ディスカバリーサービスベンダーとの技術的協議が必要で、クライアントとしての要望を伝えていかなければいけない。また、ディスカバリーサービスベンダーを介して各種設定を行うことで、持続的な保守体制みたいなものも確立出来るのではないかと思っています。こういった仕組みがないと、

正直持続可能なシステムとはいえない部分があります。

11. さらなる「日本化」に向けて

ここからは、近未来への展望ということで、更なる日本化を考えてみたいと思います。

先程、収集されるべきコンテンツの例を幾つか挙げました。それと共に、日本語においても、英語コンテンツと同じようなメタデータの厚みが必要だという結論に至っています。厚みの一要素は時間であり、過去に遡ってのデータを必要としています。将来的には、遡及の作業が必要じゃないかなと考えております。例えば、図書の目次、あらすじに絞ると、現状はハーベストデータはほとんどないんですよ。BOOK API の外部情報の取得で、エンリッチメントは可能ではありますが、ハーベストデータでは存在しない。また BOOK API で取得できる図書のあらすじや目次情報は、1986 年以降に限られてしまうという問題も存在しています。

時間的な厚みを考えると、我々の図書館の書庫には、1986 年以前の図書とか人文科学系の情報も結構あるんですよね。更に、NDL-OPAC や自館 OPAC の書誌によるメタデータは全て流れ込んでいますから、基本的なメタデータの遡及は十分に行われています。但し、現在の段階では、エンリッチメントが追いついていません。この辺と結び付けて考えると、遡及的処理は非常に必要性が高いといえるのではないかと思います。

では、日本語の図書目次メタデータの状況を考えてみたいと思います。必ずしも、ディスカバリーサービス対応しているとか、していないとか、そういう話ではないですよ。

1986 年から現在に至るまでは、BOOK データベースでいろいろなものが採られている。1968 年以前は、国会図書館のデジタル化資料である程度のデータが蓄積されている。ところが、1969 年から 1985 年に関しては、現時点では電子の存在があまり明らかじゃないんです。冊子目録みたいなものが唯一の情報源になるのかもしれません。我々としては、ここの部分がとっても重要だと思っていたんですが、日外アソシエーツがやってくれるそうです。BOOK データベースの拡張プロジェクトとして、ミッシングリンクである 1969 年から 1985 年にかけての入力事業が開始されると聞いております。これが出来ると、この辺のメタデータがリッチになって良いじゃないか、と思います。一定の課題が解決されれば、ディスカバリーサービスの一層の日本化が進むきっかけになります。

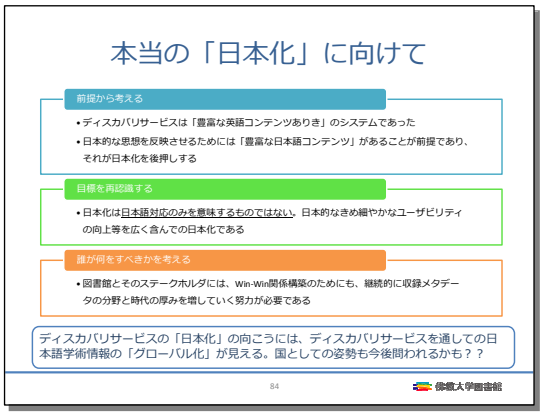
一定の課題とは、例えば、API を使った ASP サービスですが、遡及データにはほとんど ISBN がないんですね。そうすると、現状の仕様では連携できない。NCID なのか全国書誌番号なのか分かりませんが、代替のキーを考える必要も多分出てくる。じゃあ、ハーベストにしてもらうか。国会図書館のデジタル化資料と共に、ディスカバリーサービスにハーベストしてもらえれば、多分状況は非常に良い。その場合には、ディスカバリーサービス内での書誌マージ、OPAC、NDL の書誌との統合などを、ディスカバリベンダーと協調する必要があります。ただ、やっぱりハーベスト対応というのは、考えて頂かなくてはいけない今後の対応だと思います。

こういった課題を解決することで、多くの価値

ある日本語メタデータを、意識せずに提供する状況を作ることが多分可能。そうすれば、日本語コンテンツベンダーと大学との Win-Win 関係は、おそらく拡大するんじゃないかと思っております。

12. 本当の「日本化」に向けて

さて、本当の日本化に向けて、前提からもう一回考えてみましょう。ディスカバリーサービスは、豊富な英語コンテンツありきのシステムでした。日本的な思想を反映するためには、豊富な日本語コンテンツがあることが前提で、それが日本化を後押しすると思います。



目標をもう一回考えてみましょう。日本化は、日本語対応のみを意味するものではありません。日本的なきめ細やかなユーザビリティの向上とかを広く含んで日本化することで、ディスカバリーサービスはより受け入れられるようになるだろう、ということです。

もう一つ、誰が何をすべきかを考える。図書館とそのステークホルダは、Win-Win 関係構築のためにも、継続的に収録メタデータの分野と時代の厚みを増して行く努力が今後の展開として必要だと思っております。

ディスカバリーサービスの日本化の向こうには、ディスカバリーサービスを通した日本語学術情報のグローバル化も存在しています。ディスカバリーサービスは、クラウドで使われるようなシステムです。日本で入れたメタデータが海外で使われるのは当たり前のことになってくる訳ですね。そうすると、例えば日本で magazineplus というデータベースがあって、それを入れるといろいろなことができるらしい、という情報も海外で知られるようになる。海外でも重要性を認識するかもしれません。そうすると、コンテンツベンダーの販路の拡大にも繋がるし、日本語コンテンツの優位性も保たれるようになってくるでしょう。今後、国としての姿勢すらも問われることになるかもしれません。

これがですね、現状のディスカバリーサービスにおける日本化の課題というか、これからの成すべきことだという風に私は思っております。今日は長い間でしたけれども、ご静聴どうもありがとうございました。

〈フォーラム開催日　2013 年 10 月 30 日〉