

第1章

アーカイブズ共有化計画の現状

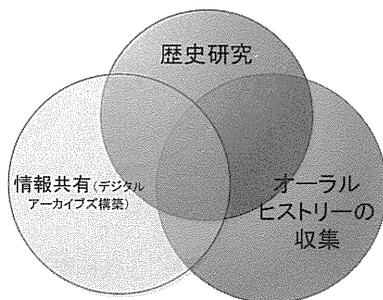
安倍 尚紀 総研大葉山高等研究センター 上級研究員

1. プロジェクトの概略

日本の巨大科学は、大学共同利用機関という世界的に見てもユニークな組織形態において研究が行われてきました。今回のワークショップのテーマは、日本における共同利用機関の歴史研究です。この背景には、科学史や科学社会学という学問的見地からの関心がある一方、社会的要請として、巨大科学についての説明責任や、社会との科学コミュニケーションが求められているという現代的課題があります。

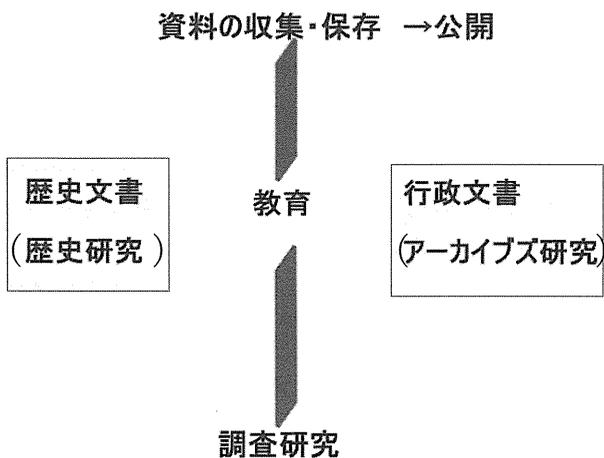
このプロジェクトでは、大学共同利用機関の歴史に着目し研究を進めるわけですが、その主要な方法として私たちは、アーカイブズを収集し、プロジェクト自らがそれを利用するとともに、社会への公開もめざしていきたくと考えています。このために現在、物理学、生物学、人類学、社会学をはじめ多様で脱領域的な研究チームをつくり研究を進めている状況です。こうしたなかでも、物理的な資料収集と並行してオーラルヒストリーも収集するのは特徴的だといえるでしょう。また、これは私が主に担当しますが、デジタル・アーカイブズの構築も含めた共同利用機関の情報共有もはかかっていく予定です。歴史研究、オーラルヒストリーの収集、デジタル・アーカイブズの構築というこれらの手法は、以下のように重なりあうものとして図示できるでしょう(【図1】参照)。

【図1】プロジェクトの手法



従来のアーカイブズ活動は、歴史文書、古文書などの収集・調査研究が主軸になっていました。今回のプロジェクトにおいても、そうしたノウハウをしっかりと踏まえながら、共同利用機関の諸アーカイブズと連携していきたいと考えています（【図2】参照）。

【図2】プロジェクトの概要



2. 「情報共有化」についての現状報告

2.1. ウェブページ(Webpage)等の情報発信について

次に、私が総研大に着任してから1年間の「情報共有化」の成果についてお話ししたいと思います。情報共有化とは、会社であれば、社内のグループウェアや共有サーバのように何らかの場所を設置し、情報をシ

エアするシステムを念頭においています。なにより、個々の情報発信が基本になっています。

以下、いくつかの情報発信の方法について、実施した経験にもとづいてまとめてみましょう。どれも、おなじみのインターネットブラウザから閲覧できる形式の情報発信です。たとえば、われわれが普通に閲覧しているウェブページもアーカイブズの1つと言えますが、ウェブ上の情報発信方法には、次のようにいくつかのバリエーションがあります。

- ・ **html**

ホームページによる発信に代表されるように、一元型の管理には効率的な方法です。

- ・ **html+cgi**

閲覧者による検索、書き込みなどダイナミックな相互作用が可能ですが、サーバの負担やセキュリティ上の問題などが発生します。以下にこの派生形態として、3つのカテゴリーを挙げておきます。

- ・ **ブログ**

ウェブログの略です。ここでは単なる **html** を用いた日記等のことではなく、Nucleus や Movable Type などのツールを用いたものを指します。簡単にアップロードでき、タグなどを自動編集してくれるので、日記・日誌によく使われており、有名なサイトとしては、「関心空間」「はてなダイアリー」「mixi」等があります。

- ・ **wiki**

ウィキウィキ(WikiWiki)とも呼ばれます。「複数人で編集できるコンテンツ作成ツール」です。ウェブブラウザを利用して WWW サーバ上の **html** 文書を編集するという感じで、何人もの人たちがネットワーク上のどこからでも、同一の文書を書き換えることができます。ウィキペディア(wikipedia)のように、世界中の人が知識を書き足しアップデートすることにより、共同作業で知識を構築することも可能です。このように共同作業で文書を作成するのに向いているので、私も dokuwiki を運用してみました。

・ XOOPS

ウィキや掲示板の“いいとこどり”をしたシステムで、多くの wiki が誰でも自由に書き込めるのに対して、こちらはメンバーシップを重視しており、認証を行った上でログインし、与えられた権限の範囲内で個々のページや共通コンテンツを構築していくものです。昨年(2005年)、eラーニング用の Netcommons を試験運用してみました。

しかし、いろいろ試した結果、先にも述べたように、やはり個々の情報発信とモチベーションが情報共有の基本であることを実感しました。またアーカイブズの情報だけでなく、データのフォーマットや手法も「共有化」の一つであると認識するようになってきました。

2.2. 実際のアーカイブズの検索手段 (finding aid)

あるアーカイブズで資料を探するとき、昔ならば図書館式のカードなどを用いて検索を行っていました。現在は電子的な検索手段が主流になっています。

もちろん、このプロジェクトについても独自の検索手段を確立していく必要があります。現在、KEK や NIFS でも電子的な検索手段を構築しようとしています。そこで手始めに Javascript 等いろいろ試しましたが、現在のところ CSV ファイルと Perl によるデータベース構築に決定して、簡素なものをいくつか試験運用しています。CSV ファイルはコマで区切られたテキストファイルで、エクセルで扱えるため変換も容易であるという利点があります。これによって膨大な資料の収集、整理が可能になりました。それぞれの公開資料は下記のとおりです。

・ KEK の公開資料

<http://hayama.soken.ac.jp/~n-abe/kek/no34/db.cgi>

・ NIFS の公開資料

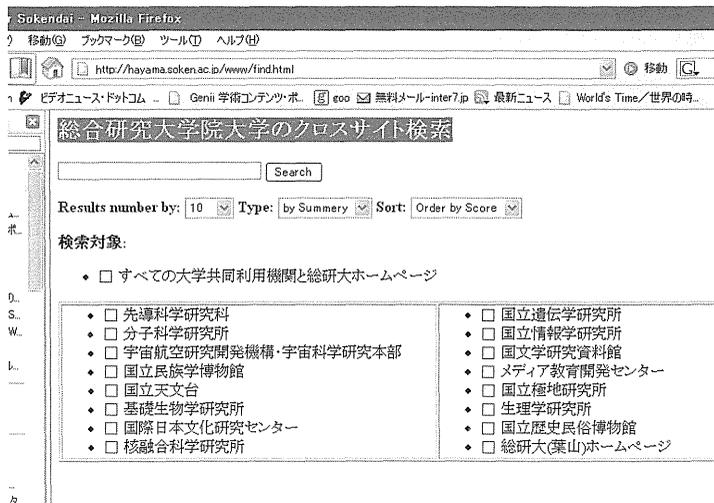
<http://hayama.soken.ac.jp/nifs/data/db.cgi>

確定したデータの流れですが、KEK、NIFS はファイルメーカーPRO というソフトを利用してデータベースをまとめています。その他、共同利用機関でもアーカイブズのデータベース構築の計画がいろいろありますが、まずエクセルで調整できる CSV ファイル化に落とし込むのが効率的なのではないかと思います。最終的には、後述のように XML の特定フォーマットに沿って MySQL というデータベースに格納し、もう少し高速での検索を実現したいと考えています。

2.3. クロスサイト検索による共有化

次のレベルとして、クロスサイト検索による共有化も構想しています。

【図3】総研大のクロスサイト検索 <http://hayama.soken.ac.jp/www/find.html>



【図3】は総研大のクロスサイト検索ですが、構成する研究所、研究機関のいくつかを自由に選んで、その中でのキーワード検索ができます。これはサーチエンジンの技術の適用ですが、通常のサーチエンジンがい

ろいろなウェブサイトを回って巡回するのに対して、これは、一定の指定されたウェブサイトだけを巡回してデータ収集し、インデックス化するものです。検索システムとしては、当面 **namazu** を用いています。

2.4. メタデータの統一による共有化

クロスサイト検索は、いろいろな既存サイトのデータを集めてその上で共有化をはかっていく手法です。これに対して、メタデータの統一による共有化は、そもそもデータを作成する段階から共有化を志向したものです。

われわれのプロジェクトでは共通フォーマットとして、**EAD(Encoded Archival Description)**を選択しました。EAD は、アーカイブズをインターネット上で公開する際の有力な国際標準で、DTD のセットが公開されています。また EAD の表現は、**SGML (Standard Generalized Markup Language)**と **XML (Extensible Markup Language)**によっています。

EAD 化に向けての課題としては、アーカイブズの編集項目を統一する必要がありますが、それぞれアーカイブズの編集方針があるために機械的にはできません。個々の特色を活かしつつ統一をめざしていきたいと考えています。これについては、日本における EAD 研究の第一人者である五島敏芳先生（国文学研究資料館）に協力いただいております。また私自身も、米アーカイブズ学会の認定資格研修等を通して理解を深めています。

2.5. 公開用のインターフェイスについて

EAD はインターネット上のデータ公開を念頭においてつくられていますが、オーラルヒストリーの場合、すべてのデータをインターネット上で公開していかどうかという問題があります。そこでアクセス制限を設定したり、セキュリティやプライバシーの問題を考慮しなければならないでしょう。必ずしもオンラインに限定せず、DVD 等によるオーラルヒストリーのデジタル・アーカイブズ化も検討の必要があるかもしれません。たとえば、カリフォルニア工科大学などでは、画像、音声など

をDVD化して見るようにするFlashのような研究データ公開用アプリケーションも使われています。非公開ですが、これまでの映像データやトランスクリプトを使って試験的に作ったアプリケーションがあります。

いずれにしても、アーカイブズと言っても、「共同利用機関の研究における(資料としての)価値」を重視する方向と「共同利用機関の歴史に関心をもってもらい実際にユーザーに使ってもらう」方向はまったくベクトルが違いますので、どこに重点を置くかについて、十分検討していく必要があると思っています。

〈質疑応答〉

—— メタデータによって検索システムを共有化する方向と、EADを使う方向は競合的なのですか？

安倍 そうです。EADは効率よく情報にたどりつくために、タグをたくさん埋め込んでいます。たとえば1冊の本の場合、題名、著者情報、本の色、紙質などの情報が詳細に埋め込まれているイメージです。

—— EADをベースにした総研大の基準で、関係研究所の横のつながりができてきていると考えていいのですか。アーカイブズでは、インターディシプリナリーなコラボレーションが形成されて、それらが有効に活用できることが重要ですから。われわれのところにも写真データなどはたくさんあるので、そのような方向になることが望ましいわけですが。

安倍 このプロジェクトでは歴史研究もちろんですが、どうすればアーカイブズをうまく情報共有できるかという方法論の研究も重視しています。できるだけ関係機関と密接に連絡をとらせていただき、方法論の共有化をはかっていきたいと考えています。