

# メディアセンターにおける電子資源の課題と展望

さとう やすゆき  
佐藤 康之

(メディアセンター本部課長代理)

## 1. はじめに

大学図書館から発信される情報は、従来の紙を媒体とした情報に加えて、電子的な媒体による情報が急速に一般化している。電子媒体による情報は電子資源 (Electronic Resources) と総称されるが、その形態や提供、管理、利用の仕組みは一樣ではなく、大学図書館における管理方法の模索が続いている。電子資源は、既存媒体の電子化による資源から開発され、現在では Web 上の資源などのように電子的に生産・提供される電子資源が急速に増加している。メディアセンターでは現在「電子的に生産・提供される資源」について管理や利用方法の検討・整備を進めている。「既存媒体の電子化による資源」についてはデジタル変換技術及び提供方法の実証を試みている。

## 2. 「電子的に生産・提供される電子資源」

### 2-1. Born-Digital: その特性

1990年代後半から学術雑誌の出版社では出版工程の電子化が急速に進められてきた。論文や記事は、著者から投稿される段階から電子ファイルで扱われ、その後の査読や編集、印刷においても電子ファイルのまま処理され冊子体として刊行される。同時に最終的な電子ファイルは出版社のデータベースに格納され CD-ROM やネットワークを媒体として出版されている。また本文だけでなく、抄録や引用といった副次的な情報も容易に出版できるようになった (One Source Multi Use)。この生産工程の改革が生み出した電子ジャーナル (EJ, Electronic Journal) は、デジタルのまま流通し利用者までデジタルで届くことから生まれながらの電子資源 = Born-Digital の典型といえる。また国内外の研究機関では、研究者が研究成果の論文やワーキングペーパー、研究過程における各種データなどを所属研究機関の Web サイトに掲載するなど電子的な業績公開が一般化している。これらの資源も Born-Digital の事例の一つであろう。

「電子的に生産・提供される電子資源」は、生産の容易性、速報性が利点となる一方で、物理的な形態が存在せず改変が容易なことから公開・出版状態の不安定性も併せ持っている。電子資源を利用する側でも提供方法や利用実態の管理、収集管理などの課題を抱えている。

### 2-2. ライセンス (利用許諾) により提供される電子資源

「ライセンスにより提供される電子資源」とは外部の機関や出版社、情報提供会社などとの利用許諾契約にもとづいて提供される電子資源である。これらはネットワークを経由して外部の出版社サイトなどに直接アクセスすることによって提供される。「ライセンスにより提供される電子資源」へのアクセスを改善することは、メディアセンターの重要な課題になっている。以下に電子ジャーナルを例にその対応を紹介する。

メディアセンターでは全塾及び各地区で利用可能な電子ジャーナルの一覧を EJ-OPAC と呼ぶ Web ページで提供している。利用者は EJ-OPAC を参照することにより、電子ジャーナルタイトル毎のリンクから各出版社の提供する電子ジャーナルタイトルのトップページへと誘導される。また一方で EJ-OPAC とは別に既存の本学オンライン利用者目録 KOSMOSII-OPAC (以下、OPAC) に電子ジャーナルタイトルの目録レコードの追加を予定している。利用者は図書・雑誌と同様に電子ジャーナルの目録レコードを OPAC で検索できるようになるだけでなく、OPAC から直接電子ジャーナルタイトルのトップページへとリンクを辿る事ができるようになる予定である。さらに 2004 年 1 月より複数の記事索引データベースから電子ジャーナルに掲載された全文 (フルテキスト) へのリンクや OPAC への検索機能リンクを提供する Link Resolver<sup>1)</sup> の試験運用を開始した。前述の EJ-OPAC から出力した購読リストを Link Resolver へ登録することにより、EJ-OPAC と Link Resolver の更新作業を定期的な一連作業とし

て組み立てるべく準備を進めている。

また電子ジャーナルやデータベースの場合、購読を管理するために出版社が利用者の認証に使う IP アドレスやアカウント、ライセンス条件、システムの機能を制御する画面の URL などの情報が必要とされる。最近の海外図書館システムパッケージでは、このような電子資源の購読情報を管理する業務用のサブシステムとして ERM (Electronic Resource Management) システム<sup>2)</sup>が開発されている。ERM システムでは、出版社・代理店と図書館の間で取り決められた購読情報を交換するための標準メタデータ仕様<sup>3)</sup>がサポートされていて、提供者と連携した容易な購読情報の管理が実現されている。メディアセンターでは ERM システムのような本格的な電子資源の購読管理の仕組みは整備されていないが、購読情報の管理は出版社や代理店との交渉やコンソーシアム活動においても欠くことのできない情報となっているため、今後 ERM システムのような仕組みの整備が課題となるだろう。

### 2-3. 大学内で生産される電子資源

大学に所属する教員や研究者は研究業績として学術情報を生産している。近年これらの学術情報は電子的に生産される方向に向かっており、シラバスや講義資料についても同様の傾向が見られる。これらの「大学内で生産される電子資源」は本学の研究や教育活動を広く外部に発信し、他大学や企業などと連携して共同研究やインキュベーションの手がかりとすること、また本学の知的財産としての管理・運用が求められている。そのためには電子化された研究業績の組織的な収集と保存、公開が必須であり具体的な実践が開始されている。

#### (1) 研究者を視点とした研究業績の集約と公開： 「研究者ディレクトリ」

学内では研究支援センター本部を中心に本学の研究者ディレクトリとして「慶應義塾研究者情報データベース(K-RIS<sup>4)</sup>)」の開発が進められ、2004年6月に信濃町地区の教員の研究業績が統合されて全学的な運用基盤が完成した。さらに将来構想として研究業績の書誌的事項だけでなく、電子的に生産された論文などの電子資源を登録する構想も検討されている。K-RIS とメディアセンターの関わりは、統合プロジェクトに信濃町メディアセンターが関与するだけ

でなく、研究業績の登録に関する運用においても信濃町研究支援センターを支援する形で協力が行われている。また湘南藤沢メディアセンターでは、研究業績の書誌的事項の登録に湘南藤沢研究支援センターと共同して業務にあっている。この2地区の例が示すように、研究業績の書誌的事項を作成する作業には集約された情報を効率よく発信するために正確性や統一性が求められる。これらの作業にはメディアセンターが約350万の書誌レコードを管理・運用している経験や学術情報の取り扱いに関する経験を活用することができ、研究業績の電子資源に至っては外部の電子資源の取り扱いに関するメディアセンターの経験がさらに役立つことになるだろう。今後も研究業績の集約に積極的に関わっていく必要があると思われる。

#### (2) 電子化された研究業績の収集と発信：「機関リポジトリ」

他大学の事例では、機関リポジトリ(Institutional Repository)と呼ばれるシステム基盤を構築する試みも始まっている。機関リポジトリは、研究者が研究業績などの電子資源を直接投稿できるようにし、研究機関の業績として網羅的に公開することにより新たな学術情報コミュニケーションを生み出そうとする仕組み<sup>5)</sup>である。既にカリフォルニア大学やマサチューセッツ工科大学などで実質的な運用が開始<sup>6)</sup>されている。千葉大学附属図書館の試み<sup>7)</sup>によれば、機関リポジトリの運用によって大学の社会的な説明責任や機関に属する研究内容の視認性を向上させる効果もあるとされている。また国立情報学研究所では、機関リポジトリに関する英文技術資料の翻訳などを手がけ各大学における機関リポジトリ構築<sup>8)</sup>を支援している。さらに機関リポジトリ間のメタデータを収集する通信技術標準である OAI-PMH<sup>9)</sup>(Open Archival Initiative-Protocol for Metadata Harvesting)を利用して、各大学の機関リポジトリからメタデータの自動収集を進めている。国立情報学研究所は集約した各大学のメタデータを使って電子資源ポータル<sup>10)</sup>の試験運用を開始している。機関リポジトリは研究機関内において電子的に生産される電子資源を収集し発信する仕組みの標準として広く普及しており、メディアセンターにおいても研究支援センター等の学内関連部門と連携して、その構築を推

進する必要があるだろう。

### (3) 研究業績などの電子テキストへの展開

研究業績や学位論文など電子的に生産される電子資源には、前述のような収集・発信の仕組みだけでなく、電子媒体としての効率的な応用の検討も考えられる。以下にその一つである電子テキスト化について述べる。

電子テキストでは、情報や知識の表現を計算機が理解可能なコード化された情報として表現する。情報の生産の過程から電子テキストを活用することによって、全文(フルテキスト)検索のような広範囲な探索の手がかりを作り出せるだけでなく、メタデータの精度向上や情報の意味解析、効率的な再利用などが可能となる。例えば、著者はフルテキストを一定の基準に沿ってマークアップすることにより、主体的にメタデータ作成に関与することができるようになる。これを実現するためには単に電子的にテキストを生産するだけでなく、電子テキスト自体を「意味(コンテキスト)」を表現する部分と「見栄え(スタイル)」を表現する部分に分離し、かつ電子テキスト全体に標準的な文書定義(DTD: Document Type Definition)を適用することが必須となる。このような文書定義の例としてTEI<sup>11)</sup>(Text Encoding Initiative)-DTDやISO Article-DTD<sup>12)</sup>などがあり、TEIの場合はバージニア大学電子テキストセンターとピッツバーグ大学東アジア図書館の共同プロジェクト「Japanese Text Initiative」で日本語の古典資料の電子テキスト化が実践<sup>13)</sup>されている。電子テキストからスタイルを分離することでデータベース化の際にコンテキストを重視した検索システムなどの構築が可能となり、資源探索の精度を効率化できる。またコンテキストをHTMLやPDFなどスタイルに合わせて自動変換することにより、一つのコンテキストから複数の見栄えの電子テキストを生成することも可能となる。

学内で生産される電子テキストについても文書定義の標準を選択しデータベース化することにより、従来よりも効率的な探索や多様な表現が可能となり電子資源の利用に貢献するものと思われる。

## 3. 既存資源のデジタル化による電子資源

### 3-1. 高精細画像キャプチャ技術の学内標準化

既存の学術資源を電子資源として利用する場合、電子化による高精細画像がメディアセンターでも用いられている。一般的な既存資源の電子化プロジェクトでは電子化対象に合わせてアドホックな技術仕様に基づいて作業が行われるが、この技術仕様を固めるためには多くの試行が伴い大量の電子化処理の時間的、コスト的なボトルネックとなる。この課題に対する取り組みとしてカナダのトロント大学では電子化技術仕様の学内標準化を推進し、数多くの電子資源の開発<sup>14)</sup>に成功している。

本学の場合はHUMI(HUMANITIES MEDIA INTERFACE)プロジェクトによる先進的な高精細画像電子化技術が確立されており、グーテンベルク聖書をはじめとする三田メディアセンターの所蔵する貴重書が電子化され公開されている。HUMIプロジェクトにより開発された技術は、高精細画像の撮影方法やその機材構成、撮影環境、画像処理と電子化プロジェクトの全般にわたっており、これらの技術の移転を受けてメディアセンター内の標準仕様を確立することは電子化プロジェクトにおける時間とコストの削減に貢献し、通常業務の一環として既存資源の電子化プロジェクトを推進する基盤となるだろう。

一方で高精細画像を閲覧利用する場合の技術については標準となる技術がまだ確立されていない。岡山大学や東北大学などの先行事例<sup>15)</sup>を踏まえ、技術の検証と標準化が必要であろう。

### 3-2. 階層型メタデータの整備

既存の学術情報資源を電子化した電子資源を利用するときもメタデータは重要である。特に高精細画像として電子化された既存の学術情報資源の場合、書籍であれ図版や地図であれ一塊のコレクションとして収集された資料であり、このコレクションについての説明が電子資源の探索や発見に重要な役割を果たすことになる。例えば写真のような画像コレクションでは、写真一点一点に属するメタデータと、これらを束ねる形のコレクションに属するメタデータがあれば、利用者は各々の写真のメタデータを探索することもできるし、コレクションを探索することもできるようになる。また、将来にわたる電子資源の保存の観点で、どのような技術仕様に基づいて

作成された電子資源であるのかを説明するメタデータも考慮する必要があるだろう。メタデータ階層の最下層には、各々の画像情報を概説するコンテンツ・ディスクリプション・メタデータが配置される。これは現在でも一般的に作成されているメタデータで、その画像の書誌的事項を記述したメタデータである。一般に Dublin Core, MODS ( Metadata Object Description Schema ) などの標準メタデータ記述が使われることが多い。各々の画像の技術的な仕様を記述するメタデータとしては米国規格協会 NISO ( National Information Standards Organization ) が標準化した MIX ( Metadata for Images in XML ) などが候補となる。その上位層にコレクション・ディスクリプション・メタデータを配置する。これは複数の画像を一つのコレクションとして取り扱う時に、そのコレクションの書誌的事項を記述するメタデータである。このメタデータ標準としては EAD ( Encoded Archival Description ) や ISAD ( G ) ( General International Standard Archival Description ) などが使われる。また各々の画像やコンテンツ・ディスクリプションの関係性を集約するストラクチャ・ディスクリプション・メタデータの標準も今後必要となる。オランダの出版社であるエルゼビア社が開発した EFFECT ( Exchange Format For Electronic Components and Texts ) などがこの分野のメタデータ構造をもっており、米国議会図書館 Library of Congress が開発した METS ( Metadata Encoding & Transmission Standard ) などはコレクションの交換も視野に入れた構造を持っている。

このように画像として電子化された電子資源は、電子化の対象であった資源の特徴に配慮したメタデータの作成が必要であり、それが世界的な標準化の動向であることを意識して既存の学術情報の電子化に取り組む必要がある<sup>16)</sup>。

### 3-3. 画像からの電子テキスト変換の試み

既存の学術資源を電子化する場合においても、可能な限り電子テキスト化して効率的な探索や多様な表現が実現できることが望ましい。高精細画像から構造化された電子テキストを生成することができれば本格的な電子テキストデータベースの実現が可能となる。一方で以下にあげるような制限もあり慎重な検討が必要となる。

電子テキスト化の基本は計算機が理解できるコード化が必須となるが、印刷された文字を撮影した画像はビット形式の表現であり、これを文字コードとしてコード化するためには OCR ( 光学式文字読取 ) 処理が必要となる。OCR 技術は年々進歩を遂げ通常出版される書籍類であれば 90% 以上の識別率には達しているが、必ずしも 100% ではないために人手による確認処理が必要で多くの手間がかかることになる。

さらに計算機が表現できる文字種の問題もある。現在の標準で計算機が表現できる日本語の文字種は JIS に認定された文字に過ぎず、多量の旧字や旧仮名遣いが混在する資料においては OCR を使うまでもなくコード化自体が不可能となる。計算機のコード化方式として期待されている Unicode を使用したとしても、中国語などに文字種を広げるだけで時代的な文字種の違いを吸収することはできない。したがって画像から電子テキストを作成する場合は、元の資源の作成年代が電子テキスト化の可否を決める重要な要点となり慎重な検討が必要であろう。

また構造化された電子テキストへ変換するための課題もある。文書のページ全体をスキャンしただけでは文書の構成を構造化された電子テキストへ変換することができない。スキャンした画像のタイトル記述や抄録の記述などを計算機は判別できないからだ。そこで文書の構成要素に沿ってスキャンした画像を分別し、OCR 処理を施して生成された電子テキストに文書の構成を適用するゾーニング技術が必要となる。この技術により画像から文書の構成を維持しつつ電子テキストを作成することが可能となる。

## 4. 電子資源保存の本質的な課題

最後にいずれの電子資源にも共通する本質的な課題について言及したい。電子資源の技術的基盤がデジタル技術にあることは言うまでもない。デジタル技術が紙媒体と本質的に異なる点は、その媒体のみでは人間が情報に接することができないことにある。紙に書かれた文字や図表は、そのまま人間が認識し理解することができる。一方で電子的に表現された情報は人間が認識できる形に変換するハードウェアやソフトウェアがなければ単なる電子媒体上のビットの記録にしか過ぎない。電子資源保存の本質

的な課題はこの特徴に由来するといえる。

ハードウェアやソフトウェアは技術革新により次々と新しい機能が生み出され古くなった機能は淘汰されていく。今現在の技術に立脚した電子資源を未来にわたってアクセス可能な状態に維持するためには、基盤となった技術自体を保存していくか、基盤となった技術を再現する技術を開発するか(エミュレーション)新しく開発された技術に移行させるか(マイグレーション)、いずれかの方法<sup>17)</sup>を検討しなければならない。電子資源は生産されるときから保存に対する具体的な対策が必要であり、保存のための仕組みまでが揃って初めて電子資源といえるのかもしれない。

メディアセンターで利用されている電子資源、生産されている電子資源が保存のための仕組みを持ちえるのか、この点は今後検討すべき大きな宿題である。

#### 参考文献

- 1) Link Resolver については NISO が開催したワークショップの記録を参照  
[http://www.niso.org/news/events\\_workshops/MS-2003\\_ppts.html](http://www.niso.org/news/events_workshops/MS-2003_ppts.html) (2004.6.30 参照)
- 2) 図書館システム INNOPAC で著名な米国 Innovative Interfaces 社が ERM システムを開発している  
<http://www.iii.com/mill/digital.shtml#erm> (2004.6.30 参照)
- 3) <http://www.library.cornell.edu/cts/elicensestudy/dlfdeliverables/home.htm> (2004.6.30 参照)
- 4) <http://www.k-ris.keio.ac.jp/> (2004.6.30 参照)
- 5) 慶應義塾大学 DRM 機構とライブラリシステム研究会の合同研究会の記録を参照  
<http://libsys.lib.keio.ac.jp/DRM/040115.html> (2004.6.30 参照)
- 6) eScholarship :  
<http://repositories.cdlib.org/escholarship/> (2004.6.30 参照)  
 Dspace : <https://dspace.mit.edu/index.jsp> (2004.6.30 参照)
- 7) <http://libsys.lib.keio.ac.jp/DRM/doc/asoshina.ppt> (2004.6.30 参照)
- 8) <http://www.nii.ac.jp/metadata/oai-pmh/ref.html> (2004.6.30 参照)
- 9) <http://www.nii.ac.jp/metadata/oai-pmh2.0/> (2004.6.30 参照)
- 10) <http://ju.nii.ac.jp/> (2004.6.30 参照)
- 11) <http://www.tei-c.org/> (2004.6.30 参照)
- 12) <http://www.xmlxperts.com/article/dtd.htm> (2004.6.30 参照)
- 13) <http://etext.virginia.edu/japanese/index.euc.htm> (2004.6.30 参照)
- 14) <http://digital.library.utoronto.ca/> (2004.6.30 参照)
- 15) 岡山大学附属図書館「池田家文庫」  
<http://gill.lib.okayama-u.ac.jp/ikedah/hrs/hr.htm> (2004.6.30 参照), 東北大学附属図書館「狩野文庫画像データベース」  
<http://www2.library.tohoku.ac.jp/kano/mado/index.html> (2004.6.30 参照)
- 16) 階層的なメタデータや各種メタデータ標準については以下の資料を参照  
 Caplan, Priscilla. Metadata fundamentals for all librarians. Chicago, American Library Association, 2003. ISBN : 0838908470
- 17) 電子資源の保存については以下のページでチュートリアルとして総合的に紹介されている。  
<http://www.library.cornell.edu/iris/tutorial/dpm/index.html> (2004.6.30 参照)