

5. 2 学術情報のデジタル化への対応

国立情報学研究所情報学資源研究センター長

安達 淳

1 はじめに

電子図書館ないしデジタルライブラリーは、インターネットの隆盛とともに、いろいろな側面から脚光を浴びている。本稿は、学術情報をインターネットの上で活用していくための情報の電子化とそれをサービスするための活動について、特に学術に関係した部分に焦点を当てて、動向を探ろうとするものである。また、特に筆者の所属していた旧学術情報センターおよび現在の国立情報学研究所における活動の経験を踏まえた事例の紹介を行う。

2 学会活動、学術情報と電子化

2.1 学術情報の特質

学会や研究活動に関係して生み出される学術情報は、商業的な出版とは異なる性格を持っている。近年、紙による情報生産・流通に代わり、CD-ROM データベースの利用等も脚光を浴びてきた。学術的活動の性格に即して、情報電子化の歴史を振り返る。

流通している学術情報の典型は学術雑誌である。学術雑誌を出版する主体は、大学等の研究機関、学協会、そして商業出版社である。

その特徴は、

- ・ 紀要など商業流通ルートにのっていないものも多い
- ・ 営利を第一目的とせず、著者の気持ちとしては複写の禁止よりも、積極的に研究者コミュニティで流通する方をよしとする

などである。

出版に関係する電子化活動の一般的な分類を試みると、

- 1 原稿の電子化
- 2 出版物の電子化
- 3 ネットワークなどの利用

の三つを挙げられる。

2.2 学協会における出版

わが国の学協会の特徴は、

- ・組織的規模は米国と比べはるかに小さい
- ・日本語出版物は、流通が国内に限定される。一方、国際流通を図ろうとしても困難が多い。
- ・学協会内での出版の電子化等の技術革新は米国に比べて遅れがちである。

などである。

2.3 電子原稿の利用

組織的なデータベース化の問題点は次のようである。

- ・ワープロの多様なファイル形式のための統一された形式が望まれる。
- ・全文データの標準形式に関する国際標準には SGML や XML がある
- ・標準化が定着すれば、学会は容易に電子化原稿を集めることができ、その副産物としての全文データベース化の推進を図ることができる。
- ・しかし、現実には統一は大変難しい
- ・図や写真の統一的な扱いが難しい。今後のマルチメディアの課題である。

3 電子図書館の機能とコンテンツの設計

3.1 情報利用形態による分類

「電子図書館」を、何らかの「物」の形態で流通している情報を電子的な形態でしかも組織的に蓄積し、提供サービスするシステムととらえる。例えば、冊子、CD 等の物理的媒体の中に格納されている情報を指し、特に冊子の形態の情報は出版という形ですでに安定した社会システムが確立している。これが電子出版やネットワーク化により「電子化」の渦中にあるわけである。

まず、電子化情報の利用形態に着目すると、(1)スタンド・アロン、(2)ネットワークによるもの、の二つに大別される。今後はネットワークベースの情報サービスが中心となっていくと考えられる。

また、利用形態はともかく、目下情報の「値付け」が最大の関心事であり、学術出版社を始めこれを模索するためのプロジェクトが多く走っている。

3.2 蓄積する情報による分類

「電子図書館」の蓄積する電子化されたコンテンツについて、ドキュメント情報を大別すると、

- ・従来の紙の形態の情報形式に依存した電子化
- ・新しいネットワーク環境に適合できるように電子化情報を構成

になる。前者は、ページをスキャナによりデジタル画像にして蓄積する手法を採ることになる。また、ポストスクリプトや PDF などのページ記述言語による方法も考えられるが、あくまでも紙の上にレイアウトされた画像情報を対象とするものである。

一方、後者はコンピュータ上での表示と利用の容易さを狙って、例えばハイパーテキストのように、情報の構成そのものを再検討して、提供するものである。

現在各所で行われている電子図書館プロジェクトをみると、スキャニングしたデジタル画像を対象とするシステムとコード化された全文(full text)情報を扱うものがあり、次第に後者に移行していくような傾向にある。前者の利点・欠点を列挙すると、

- ・印刷して読み易いレイアウトになっている
 - ・膨大な紙の情報の遡及的電子化に適用し易い
 - ・言語やフォント、外字等に依存せず適用できる
 - ・慣れ親しんだ表現形式なので、紙のシステムから移行し易い
- 最近の情報はすでに発生時から機械可読であるが、これを有効活用していない

— 検索機能を補うデータベースが必要である

のようになる。これを逆に考えれば、おおむね全文情報の利害得失になる。

従来、学術情報センターでは、SGML によるフルテキストデータベースの実施を試みてきたが、最近では SGML をより簡素化した XML が標準として広まる気配を見せている。

4 NACSIS/NII の電子図書館システム

4.1 事業の位置付け

文部省の大学共同利用機関である学術情報センター(NACSIS、2000年4月からは国立情報学研究所)では、1986年の発足以来、大学図書館のネットワーク化や学術情報のデータベース形成を行ってきた。

この目録事業と並行して、わが国の学会と協力したデータベース形成も行ってきた。中でも学会発表データベースには65の学会が参加し、学会の行う大会、研究会、シンポジウム等での研究発表に関する概要情報、すなわち発表表題、著者、アブストラクト等の情報が蓄積されてきた。

学術情報センターはこれ以外にも学会への支援事業を行っており、その最近の例として Academic Society Home Village というサービスがある。これは学会に無料で WWW サービスのためのコンピュータ資源を提供しようというもので多くの学会で活用されている。

このような事業展開のなかで、次の情報サービス電子図書館システムの開発を行い、1995年2月からの試行サービスを経て1997年4月から公開サービスを行っている。

学術情報センターの活動には、文部省管轄下の機関としての役割とも関連し、学会活動やその発行する学術雑誌を対象としたものが多い。このような背景から、NACSIS-ELS の開発では、学会活動に関連した情報形成・提供支援に寄与することを強く意識して設計してきた。

4.2 NACSIS-ELS のねらい

電子図書館サービス NACSIS-ELS は、雑誌のすべてのページを画像としてデータベースに蓄積し、利用者の手元に高速ネットワークを通してセンターサーバから直接供給する機能を実現したものである。

すなわち、NACSIS-ELS のデータベースサーバは、

- ・二次情報データベースの検索機能
- ・文献のページのブラウズ機能

の二つの機能を統合したものである。

4.3 開発の経緯

NACSIS-ELS の開発は 1990 年頃から始めた。システムの稼働する環境としては、当時勃興しつつあったインターネット、すなわち TCP/IP をベースにしたネットワーク環境を前提として、client/server 型のネットワークアプリケーションとして実現した。システムの設計と実現には、それほど時間を要しはしなかったが、最大の課題はデジタルコンテンツの入手であった。

まず最初に 1993 年に情報処理学会に実験目的での学会誌や論文誌のデジタル化の許可を申し出た。さいわい快諾を得ることができ、1994 年には国際会議で実現システムのデモンストレーションなども実施できた。

並行してコンテンツデジタル化の規模を拡大すべく、文部省の科学研究費補助金を申請し、1995 年から 2 年間の試行実験を行った。この際には、情報処理学会に加え、電子情報通信学会、電気学会から雑誌のデジタル化の許可を得ることができた。この試行では、あくまでもサービスの実用化実験として、著作権使用料を支払わずに、限定した利用者にサービスを提供してその実現上の問題を検討することが目的であった。

2 年間の試行期間中には、いろいろな学会との間で著作権使用料に関する考え方の調整を行い、このような情報サービスにおける新しい課金方式の検討を進めた。これと同時に雑誌のデジタル化についての許可を他の学会にも求め、コンテンツを拡大する努力を行った。1996 年末には 28 の学会がこのプロジェクトに参加し、62 の雑誌がデジタル化の対象となり、約 56 千論文、45 万ページのコンテンツが蓄積できた。

一方、課金方式については、他に類をみないため、なかなか方針が定まらなかった。試行実験は 1997 年 3 月に一応終了し、4 月からは公開サービスとして広い範囲の利用者が利用できるような制度でサービスを公開した。この際には参加学会には当分の間無料で情報を提供する旨の特別の許可を得てサービスを続けた。そのため、試行に参加した学会の一部は改めて意思決定するのに時間を要するなどの都合もあったため、公開当初は 14 学会、20 誌程度に下がった。しかし、並行して、多くの学会にサービスの説明を行い、雑誌のデジタル化を進め、雑誌の提供を許可する学会も 9 月には 23 学会、45 誌へと徐々に増加した。

1997年度公開当初は無料にしたものの、課金単価を定める手続きに手間取り、最終的に著作権料の徴収を始めたのは1999年1月からである。それ以前には利用者の登録は7000人規模であったが、料金徴収開始のための利用者再登録の結果、当然のことながら1600人くらいまでに利用者数は低下した。今後徐々に増加していくと期待している。

2002年5月では、183学会が参加し、418タイトルの雑誌についての122万論文あまり、400万ページ近くのコンテンツが公開されている。

収録雑誌の数は今後も増えていくと予想している。登録利用者は約5000人である。

5 著作権処理

5.1 共通するサービス上の問題

ネットワーク上の文献提供に関する学会側での懸念を要約すると、

- ・雑誌発行部数の減少
- ・会員の減少

の二点になる。雑誌の情報が容易に入手できるようになれば、定期購読の予約数が減り、また会員になる利便性が薄くなることにより、会員が減り、学会活動に悪影響が出るというものである。

これらが果たして真実かどうかは不明であるが、NACSIS-ELS 開発に当たっては、学会側のこれら懸念に十分配慮し、結果として多くの料金設定パラメータを用意した。学会側がそれらを自由に設定できるようにすることで、問題を回避するような方針をとってきた。

5.2 料金方式

1999年1月から実施した課金方式の骨子は次のようである。

- ・個々の著者は学会へ著作権を委譲する
- ・利用した量に応じて課金する従量制
- ・個人利用を対象
- ・新しく発行される雑誌を対象
- ・遡及的なデジタル化も並行して実施

というものである。

5.3 学会の決定項目

著作権使用料金設定に関して学会は、

・画面表示の場合のページ当たり単価 ・印刷の場合のページ当たり単価 ・会員、非会員による料金の差別化 ・ページ属性による料金の変更 ・ネットワーク公開までの時間遅れの設定 ・無料にするページの設定可能

が可能である。要は、雑誌ごとにかなりきめ細かく料金を変えることができるようにした。会員かどうかにより料金を変更でき、学会の中には会員に関しては印刷しても料金をとらないという方針をとることもある。なお、学術情報センターの機関としての性格から、会員だけへの閉じたサービスは許していない。料金さえ支払えば有資格者すべてが利用できるような体系を設定している。

以上のような複雑な料金制度は、学会との協議の上で定まってきたもので、従来の活動に与える影響を極小に押さえながら徐々に新しいネットワーク志向のサービスに転換していく際に必要な考え方で、NACSIS-ELS に特徴的なものであると考えている。

6 NACSIS-ELS の課題

6.1 技術的な課題

1998 年度から学術情報センターは「オンラインジャーナル」プロジェクトを開始して、学会が XML を用いて電子的に雑誌を編集していくためのソフトウェアシステムを開発し、学会に提供していくことになった。このプロジェクトの成果として、今後は学会からは XML ベースのデジタル文書として学術情報が生産されてくることになる。このように新しく生産される情報については、XML、PDF など新しい文書形式が採用されるようになっている。

学術情報の電子化が進んできたことから、既存の種々の情報の統合により、さらに付加価値を生み出していくことが進んでいる。すなわち、

・ページ画像の NACSIS-ELS ・抄録情報の NACSIS-IR 学会発表データベース ・XML による新しいオンラインジャーナル ・引用索引データベース

を統合するような利用環境の実現を目指している。現在「学術情報ポータルサイト」から逐次公開していくことにしている。

6.2 制度的な課題

今後出版社が学会が独自に電子ジャーナルの提供サービスを始めていくと思われるが、決して情報の入手がすんなりと容易になっていくとは思えない。課金処理が介在するためである。

米国ではすでに電子ジャーナルの利用に当たり、appropriate copy という問題が指摘されており、技術的解決と共に、今後のコンテンツ利用の仕組みづくりの課題になっている。

一方で、並行して研究者や学生の文献アクセスの態様が変化していくことも分析していかなければならない。電子化されていくことにより、雑誌単位のアクセスから記事単位へのアクセスへと、アクセス粒度が小さくなっていく。これは、情報を生産する学会側からは大きな問題であるといえ、一方、研究者側からも従来のような単純な右肩上がりの文献発表では済まなくなる環境になるとも思われる。研究者コミュニティの中での文献の利用の仕方の変化の方向にも注目する必要があり、簡単に予測しがたい。

7 むすび

学術コンテンツの電子化を紹介するなかで、最後に、システム開発とともにコンテンツの作成があいまって進んでいくことの必要性を強調したい。

新たに発足した国立情報学研究所では、学術情報センター時代から引き続いて、画像ベースの電子図書館システムの開発と並行して、全文情報を扱うデータベースの研究開発も行い、画像、全文がコヒーレントにつながるシステムを目指している。一方、電子文書では PDF が広まっている。今後は多くの情報が直接電子的な形態で入手できるようになると期待されるため、電子図書館システムも、これらに対応したものに変わっていく必要がある。

一方、改めて「図書館」的機能やサービスに関心が集まっている。現在、「電子図書館」といえば、ネットワークを通じて行われる情報サービスの典型として注目されており、今後、教育、学術研究、生涯教育などすべての面での情報化と密接な関係を持っているといえる。

図書館に対しては、ネットワーク上で流通する学術情報を積極的に収集していく営みがますます強く期待されるようになっていくと思われる。