

要 望

電子媒体学術情報の恒久的な蓄積・保存・
利用体制の整備・確立

平成 17 年 9 月 15 日

日 本 学 術 会 議

要 望 の 内 容

1. 要望提案の背景

近時図書館の変容は目覚しく、これまでの印刷媒体の延長線上でのマイクロフィルムによる保存から、電子情報媒体による保存に依存するように変化するだけではなく、出版形態そのものが電子情報媒体による比重が急速に増えている。

他方、電子情報媒体による情報の蓄積は、その情報を取り出すための機器(ハードウェア)とそれを読み取るためのプログラム等(ソフトウェア)が適切に保存されるか、あるいは媒体の技術変化に即応して、新機種用に書き換えて保存するという処置をとらなければ、これまでの印刷媒体・マイクロフィルムと異なって再現利用はほとんど不可能であり、蓄積自体が無に帰着する可能性を持っている。このような点を考慮して、いかに既存の情報を恒久的に保存するかをの体制を検討することが必要である。

2. 現状と問題点

学術基盤情報のデジタル化

近年の情報技術の進展は、学術情報のデジタル化を進めた。学術研究情報のデジタル化は、研究成果である論文から研究のための基盤である各種データまで、研究のあらゆる方面におよび、従来の媒体である紙と印刷による情報の量を凌駕するほどになってきた。

デジタル情報の保存に対する問題点

しかしながら、これらのデジタル情報にはその利便性の反面、保存性に関して大きな問題がある。媒体の耐久性の問題に加え、ハードウェア、オペレーティングシステム、アプリケーションソフトウェアといった稼働環境も保存されないと、利用可能な形での保存にならない点がそれである。

保存問題の解決に関する技術の現状

この問題を解決するための方法は、これまで、紙媒体での保存、新しい環境に対応してコピーをくり返すマイグレーション、恒久的に保存する書式を標準化して限定する方法、新しい環境の上で古い環境を稼働可能にするエミュレーションなどが提唱されているが、いずれも完全な解決にはならず、決定的な解決策と考えられるものは見つかっていない状況である。

保存に関する制度的問題

紙・印刷媒体の情報は、出版物の納本制度と、図書館による保存、あるいは公文書館・アーカイブ等の保存および利用体制が一定程度整備されており、少なくとも重要な情報が雲散霧消することのないような体制が作られている。これに対し、デジタル情報に関する保存体制は、パッケージ系電子出版物と呼ばれるごく一部を除き整備されていない。

わが国の取り組みの現状

国立国会図書館の納本に関する審議、電子図書館計画、および国立情報学研究所の電子ジャーナルアーカイブなど、問題点の認識はされており、いくつかの関連する取り組みはあるものの、実験的、限定的な範囲にとどまっている。将来の研究のための基盤的資産として情報を蓄積・保存・利用するという観点から見れば、とても十分なものとは言い難い。

3. 改善策、提言等の内容

デジタル情報の本格的蓄積・保存・利用体制をすみやかに構築する方策を立てるために、学術基盤情報常置委員会は以下のことを提言する。

- 1) 学術の世界においてデジタル情報の保存技術・制度の研究の推進。
- 2) 学術機関におけるデジタル情報の蓄積・保存・利用のための組織の確立、およびそれに対する環境整備と支援。
- 3) 民間はもとより国家全体におけるデジタル情報の蓄積・保存・利用体制の確立。

目 次

1	はじめに	1
2	学術基盤情報のデジタル化	1
3	デジタル情報の保存に対する問題点	2
4	保存問題の解決に関する技術の現状	3
5	保存に関する制度的問題	4
6	我が国の取り組みの現状	5
7	提 言	7

電子媒体学術情報の恒久的な蓄積・保存・利用体制の整備・確立

1. はじめに

近年の情報技術の進展は、学術情報のデジタル化を進めた。これによって学術研究の方法や社会と科学のかかわりかたに大きく変化をもたらした。また、研究に必要な情報へのアクセスも、格段に早く幅広くなった。

しかし、その一方でデジタル化され、電子媒体に記録された情報の保存については、技術的にも制度的にも大いに問題をかかえている。電子媒体の情報は、従来の紙媒体の情報に比べ保存性という点でははるかに脆弱である上に、出版物での納本制度のような保存制度も、全くといってよいほど整備されていないからである。その状況は、未来の社会と科学の発展を阻害する危険をはらんでいる。

この報告は、デジタル化された情報の蓄積・保存・利用に関する問題点を検討し、とるべき方策について提言する。

2. 学術基盤情報のデジタル化

デジタル化は学術研究のあらゆる方面におよび、デジタル化された学術情報の利用は、従来の紙・印刷媒体を凌駕するほどになっている。

学術研究情報のデジタル化は、研究成果である論文から研究のための基盤である各種データまで、研究のあらゆる方面におよび、ほとんどの局面で利用されるようになってきている。

第 1 に、現在論文を手書きの原稿で作成、提出することはほとんどなくなった。また、電子ジャーナルがこの 5 年ほどの間に急速に普及し、国際的に流通する大手学術出版社の学術雑誌は、そのほとんどすべてが電子ジャーナルの形で流通するようになった。非商業的な学会誌などでも、電子ジャーナル化が進む傾向にある。読者である学生や若手研究者は、デジタル媒体と紙媒体とがあった場合、圧倒的にデジタル媒体を選択するとされている。

第 2 に、古典籍、古版本や古写本などの初期出版物などアクセスの難しかった図書、雑誌が、多く電子図書館としてインターネット上でアクセス可能になり、あるいは CD-ROM 化されて商業的に流通している。古典籍資料は、著作権上の問題が少なく、また、デジタル化が利用による資料の劣化を防ぐ手段として有効であることから、所有している図書館がデジタル化して電子図書館のコレクションとすることが多い。明治期や戦前の出版物で研究資料としての価値は大きくても、一般的に広く売れるほどの利用の見込まれないものは、従前マイクロフィルムで流通していたが、最近は CD-ROM として出版されることが多い。

デジタル化によって索引の作成が容易になり、読者の利便性が増すという利点もある。

第 3 に、学術研究の基盤となる実験データ、統計データ等は、ほとんどすべてデジタル情報となってきた。現在では、アンケート調査を行なっても、その他の統計調査を行なっても、その結果の整理にコンピュータを利用しないことは、ほとんど考えられない。実験結果も同様で、かつてはノートに書きとられてから始まったデータ処理が、現在では直接コンピュータに入力されていることが多い。これらの基盤的情報は、商業的流通や公的情報としての流通に乗ることはあまり多くない。

これらに限らず、学術的な議論の行われるメーリングリストや、プレプリント、また、研究上必要な参考情報など研究上重要な多くの情報が、デジタル媒体で作成され、インターネット上に流通し利用されている。また、その形態も、文字によるテキスト中心であった従来の紙媒体の情報に比較して多様になってきた。写真にあたる静止画像はもとより、音や動画像にいたるまで、同一媒体の中に収録して流通・利用することが可能となってきている。これらのデジタル化された学術情報は、利用される度合いからすれば、従来の媒体である紙と印刷による情報の量を凌駕するほどになってきた。

3. デジタル情報の保存に対する問題点

デジタル化された学術情報には、その利便性の反面、保存性に関して大きな弱点がある。

これらのデジタル情報の保存性の問題点は、その媒体の問題から、利用のための環境の保存という点におよぶ。これまで指摘されている問題点の中から、特に学術情報に対して重要と思われる問題点を拾うと次のような点があげられる。

第 1 に、デジタル情報はコンピュータのハードディスク等の磁気ディスク、CD-ROM、DVD-ROM 等の光ディスクなどに蓄積・保存されるが、これらの媒体の寿命は数年から数十年と考えられ、一般的に紙よりも寿命が短い。光学的媒体のある種のものは、百年以上の寿命があると言われているが、これも実績のあるものではなく、千年の実績のある和紙と墨には遠く及ばない。

第 2 に、媒体の寿命が保ったとしても、それを読み書きするハードウェアが稼働する状態で保存できるとは限らない。たとえば、初期の 8 インチフロッピーディスクに記録されたファイルを現在利用しようとしても、それを読むことのできるドライブを入手することは不可能に近いだろう。

第 3 に、媒体が正常で、入力装置が稼働したとしても、これらの情報は何らかのアプリケーションソフトウェアに依存して作成されている。アプリケーションソフトウェアが保存されていなければ、データだけがあっても実際には利用できないということも多い。

特に文字情報にとってコード体系の変更は、重要な影響を及ぼすことは、漢字コードを例

にとっても明らかである。

第4に、データとアプリケーションソフトウェアが読み込み可能であったとしても、そのアプリケーションソフトウェアの稼働するハードウェアおよびオペレーティングシステム環境が保存されていなければ、実際には利用できない場合が多い。たとえば、20年前のオペレーティングシステム DOS 用のアプリケーションソフトウェアは、現在のオペレーティングシステムのもとで正常に動作することは通常期待できない。

デジタル情報は、広範囲な機能性やコピーの容易さといった利便性の裏腹として、アクセスのための環境も共に保存されなければならないという点で、伝統的媒体に比べ保存性に弱点がある。

4. 保存問題の解決に関する技術の現状

保存問題の解決方策としては、紙媒体での保存、マイグレーション、標準化、エミュレーションなどの方策が考案されているが、限定的な解決にしかならない、あるいは有効性の検証が十分でないなど、完全な解決策は見つかっていない。

上記のような、問題点を克服するための技術的方策は、いくつか提案されている。

第1の方法は、すで実績のある紙媒体にして保存する方法である。この方法は単純で、ある程度までは有効な方法であるが、いくつかの限界がある。ひとつはプリンタ出力の寿命で、もし本格的にこの方式を採用するのであれば、保存に耐える印字方法を採用または開発しなければならない。さらに大きな問題は、音や動画など紙媒体で記録できない情報の存在である。これらの情報については、別の方式をとらなければならないし、文字情報であってもハイパーテキストのようなデジタル情報特有の機能は、紙の上には表現できない。

第2の方法として、定期的に、あるいはシステム環境が変わるごとに、すべての情報を新しい環境にコピーし直して、アクセスを保証するという方法がある。マイグレーションと呼ばれるこの方法は、少なくとも媒体の劣化による脆弱性を解決することができる。デジタル情報の場合にはコピーを繰り返しても劣化はしないためである。しかし、これだけではアプリケーションやオペレーティングシステム環境の変化に対する対策にはならない。これを解決するためには、データを新しいアプリケーション、新しいオペレーティングシステム環境に合わせて変換して行くことも必要となる。これは、データの量や種類が多い場合、大変な人手とコストを要する仕事となる。

第3の方法は標準化によるものである。保存する情報をひとつ或いはいくつかの標準形式に限定し、この標準形式のデータは、恒久的に将来のシステムでもアクセスを保証して行く方法である。例えば、国際標準化機構 ISO ですすめられている PDF-A は、現在広く使

用されている PDF フォーマットのひとつの版を公的な標準として、アーカイブ用に使用するという計画である。保存すべき資料がこの標準の範囲に納まって、保存用の資料をすべてこのフォーマットにできるという環境のもとでは有効な方法である。しかし、そのような前提の成り立つ範囲は限られるので、これも全面的解決にはならない。

第 4 の方法は、エミュレーションである。システム環境が変わったときに、それまでのシステムを新しいシステムで動かせるよう、古いハードウェアおよびソフトウェアの代わりをする仕組みを用意する方法である。第 2 の方法がデータをシステムに合わせて生かし続けるのに対し、古いシステムごと生かし続けようという方法といえよう。この方法は、未だ技術的に十分検証されていない。また、アプリケーションソフトウェアやオペレーティングシステムを含めて保存するための制度的保証も必要で、構想の域をでていない。

上記のように、デジタル情報の保存に関する技術的基盤はまだ弱く、完全な解決策と考えられるものは見つかっていない状況である。

5. 保存に関する制度的問題

伝統的な紙・印刷媒体に対しては、納本制度、公文書館などの保存・利用体制が一定程度整備されているが、新しいデジタル情報については、このような制度が確立されていない。

一方、蓄積・保存・利用に関する制度、体制の面でも問題が存在する。紙・印刷媒体の情報は、出版物の納本制度と図書館による保存、あるいは公文書館・アーカイブ等の保存および利用体制が一定程度整備されており、少なくとも重要な情報が雲散霧消することのないような体制が作られている。

これに対してデジタル情報の保存・利用体制は、学術情報に限らず社会一般の情報に対して未だ制度が整っておらず、極めて遅れた状態にある。現在、日本でデジタル情報の保存制度のあるのは、パッケージ系電子出版物とよばれる CD や DVD などの媒体で、出版という形で流通しているものだけである。

商業的に流通している電子ジャーナルは、多くの場合ある年度を区切った利用契約によってアクセスできるのみであり、契約年度が終了すれば利用できなくなるという契約形態が通常である。これは、雑誌を購入すれば、その雑誌を保存しておく限り半永久的に利用が可能であった従来の媒体の場合と大きく異なる点であり、学術情報の保存という点からみればきわめて脆弱な体制である。

国内で出版している電子ジャーナルは数多くないが、その保存に関しては、サーバを運用している組織にまかされている。安定しない組織の場合、組織の消滅とともに情報も消滅するという事は十分考えられる。

電子図書館などの形でインターネット上に流通しているデジタル情報の場合も、事情は同

じでサーバを運用している組織次第である。図書館のような組織が消滅するということは多くないが、そのようなことがおこった場合、冊子と異なり移管等の処置がとられる保証がない。

統計データや実験データの場合、大きな組織が運用しているものは前述の電子ジャーナルや電子図書館と同じような事情になるが、より小さな組織、たとえば研究室といった単位の場合、保存に関する対策はほとんど考えられていないというのが実情であろう。

また、行政文書の電子化に関しては、e-Japan 戦略に沿って様々な検討がなされている。しかしながら、文書としての真正性の確保、機密の保持などに重点が置かれ、利用を保証する見読性の確保もほとんど文書の現用期間内の考慮にとどまっている。現用として保管される限りの利用の保証は考えられているが、公文書館に移管後の長期の保存にわたってまで十分に検討されてはいない。学術研究には、現用期間を過ぎた後の長期保存が必要であり、そのための体制が取られているとは言い難い。

インターネット上に公表されたその他の情報では、事情はさらに悪い。管理保存体制のあるものは限られていて、多くは、一時的に公表されるものとして扱われている。多くの人々がアクセスし利用されている情報であっても、なんらかの特別な手段を講じない限りは消滅するであろう。

このような脆弱性の問題を解決するためには、紙・印刷媒体における納本制度にあたる何らかの社会的制度が必要である。その制度は必ずしも納本制度とナショナルライブラリのような集中的な制度でなくてもかまわない。

いずれにしても現在の状況をこのままに放置すれば、将来 21 世紀初頭を情報の暗黒時代と位置づけることになりかねない。また、蓄積された確かな情報をもとに築いていかれるべき科学の発展と社会の成長をも阻害しかねない。この状況を改善するには、技術的、制度的両面からの大規模な取り組みが必要である。

6, 我が国の取り組みの現状

我が国における取り組みは実験的・限定的であり、学術研究の基盤的資産として情報を蓄積・保存・利用するという体制にはなっていない。

これらの問題点に対して多くの先進国は、国家のプロジェクトとして対応策を検討している。たとえば米国では、議会図書館が中心となってデジタル情報長期保存のための立案活動 NDIIP (National Digital Information Infrastructure and Preservation Program)が行われている。

我が国では、国立国会図書館がいくつかの構想を発表し、事業を行なっている。また、国立情報学研究所でも、限定的ながら関連事業を行っている。

インターネット資源選択的蓄積実験事業 (WARP)は、限定的にインターネット上の情報資源をアーカイブしている。目的は、納本制度審議会での審議に供するためであり、2002年から電子雑誌、政府や地方自治体、政府系団体、大学などのウェブサイトで管理責任者の許諾が得られたものについての収集保存を行っている。実験事業でもあるため、2005年1月現在で電子雑誌は約1500タイトルである。また、官公庁等のウェブサイトも完全に収集されているわけではないなど、限定的なものにとどまっている。

一方、制度的保存体制としては、納本制度審議会ネットワーク系電子出版物の収集に関する検討が行われてきた。2004年12月にでた答申では、納本制度とは別の制度により収集すること、国会議員の職務遂行等に必要な公表されたネットワーク系電子出版物を収集すること、発行者による収集拒否の申し出を認めること、著作権関係で法律による制限が必要になることなどを述べている。範囲を国会議員の職務遂行に必要なものに限定している点は、たとえその範囲に「学術的な情報」が含まれていたとしても、学术界の学術研究上の必要性和合致するとは限らない。

ただし、国会図書館内部の構想である電子図書館中期計画2004においては、デジタル・アーカイブの構築を目標として掲げている。また、内部的な審議会の報告においては、デジタル情報を利用可能な形で維持していくための保存技術の問題について、正面に据えて取り組むべき問題としている。現段階では、問題の認識および今後の目標については、デジタル情報の保存問題を解決しなければならないことを十分認識しつつ、現実においては限定的取り組みを行っている段階といえよう。

国立情報学研究所のNII-REOは、出版社との契約のもとに電子ジャーナルを図書館に統一した枠組で提供する仕組みであるが、この中でアーカイブする事業も行っている。出版社との契約が許せば、最新号購読の契約終了後もアクセスを可能にするために、電子ジャーナル本体を保存する。この方式によって、冊子体の雑誌で可能であった恒久的な利用を電子媒体でも可能にする。しかしこれは電子ジャーナルに限られた体制であり、また、個々の出版社との契約に基づくものであるため、現段階では保存されているタイトル数は多くない。

以上のように、現段階での国家的取り組みは、実験的、限定的なものにとどまっている。将来の社会と科学のための基盤的資産として情報を蓄積・保存・利用するという観点からみれば、とても十分なものとは言い難い。特に電子ジャーナルの問題は、国際的競争の観点から早急に方策を講じるべきである。

その解決の際、留意すべきことは、デジタル情報の脆弱性に対応するための技術は、まだ未成熟であり、実用的レベルにするためには、なお一層の研究・開発が必要である。我が国の学术界もこの問題の重要性を認識し、その解決のための研究を進める必要がある。さらに制度的問題の解決には、法的観点や社会制度の観点からの研究が必要であり、分野を

超えた幅広い問題解決型の研究を推進することが重要である。これらの実現のためには、国家的な大型の研究プロジェクトを立てて推進することが望ましい。

また、これらの研究成果を待ってから保存を始めたのでは、現在生み出され流通しているデジタル情報が消滅するおそれ大きい。これらを救うためには、学術情報を生み出している機関である大学や研究所が、自らの生み出している冊子体でないデジタル情報を蓄積・保存・利用可能にする組織を既存の組織の再編成または新設によって作り、まずは収集・保存し始めることが必要である。機関ごとに保存することによって、一組織あたりの量が減少するので、長期間にわたらなければ、現在知られている方法で利用可能性を維持しつつ保存することが可能であろう。少なくとも、学術機関が必要とする情報、創り出した情報について、自らの責任において保存・公開する体制を創ることが重要である。また国としても、学術機関での保存・公開体制の確率を支援し、そのための環境整備を行うことが望ましい。

しかし、学術機関の生み出した情報だけが学術研究に使われるわけではない。政府情報もあれば、広く社会で生み出され、流通している情報も学術研究の対象となりうる。これらを含め、日本で生み出されたデジタル情報を包括的に対象とするナショナルアーカイブ・ライブラリのような組織を国家として設立することが望まれる。可能な限りの情報を永く後世に残すことは人類の文化の発展の基本であり、このためにも国としての情報の蓄積・保存・利用体制を確立し、維持していくことは国家としての人類に対する責務として行うべきである。

7. 提言

デジタル情報の本格的蓄積・保存・利用体制をすみやかに構築する方策を立てるために、当委員会は以下のことを提言する。

- 1) 学術の世界におけるデジタル情報の保存技術・制度の研究の推進
- 2) 学術機関におけるデジタル情報の蓄積・保存・利用のための組織の確立、およびそれに対する環境整備と支援
- 3) 民間はもとより国家全体におけるデジタル情報の蓄積・保存・利用体制の確立

文献

Rothenberg, Jeff, "Avoiding Technological Quicksand: Finding a viable technical foundation for digital preservation", Council on Library and Information Resources, Washington DC, 1999.

武邑光裕,「デジタル・アーカイブにおける課題と展望」, 情報科学技術協会編,「特集=デジタル情報資源のアーカイビング」, 情報の科学と技術, Vol.54, No.9, 2004, pp.440-446

今野篤,「デジタル情報の長期的保存の政策：アメリカと日本」, 情報科学技術協会編,「特集=デジタル情報資源のアーカイビング」, 情報の科学と技術, Vol.54, No.9, 2004, pp.447-452

中山正樹,「国のデジタル・アーカイブ・ポータル構築：国立国会図書館『電子図書館中期計画 2004』の実施に向けて」, 情報科学技術協会編,「特集=デジタル情報資源のアーカイビング」, 情報の科学と技術, Vol.54, No.9, 2004, pp.453-460

栗山正光,「OAIS 参照モデルと保存メタデータ」, 情報科学技術協会編,「特集=デジタル情報資源のアーカイビング」, 情報の科学と技術, Vol.54, No.9, 2004, pp.461-466

肥田康,「電子画像情報の利用と保存」, 情報科学技術協会編,「特集=デジタル情報資源のアーカイビング」, 情報の科学と技術, Vol.54, No.9, 2004, pp.467-474

尾城孝一, 杉田茂樹, 阿蘇品治夫, 加藤晃一「日本における学術機関リポジトリ構築の試み：千葉大学と国立情報学研究所の事例を中心として」, 情報科学技術協会編,「特集=デジタル情報資源のアーカイビング」, 情報の科学と技術, Vol.54, No.9, 2004, pp.475-482

National Digital Information Infrastructure and Preservation Program, "Digital Preservation", < <http://www.digitalpreservation.gov/>>, [cited 2005-03].

納本制度審議会答申,「ネットワーク系電子出版物の収集に関する制度の在り方について」, 国立国会図書館, 2004

国立国会図書館,「国立国会図書館電子図書館中期計画 2004」, 国立国会図書館, 2004, http://www.ndl.go.jp/jp/aboutus/elib_plan2004.html

国立情報学研究所, "NII-REO", < <http://reo.nii.ac.jp/journal/HtmlIndicate/html/index.html>>, [cited 2005-03].

文書の電磁的保存等に関する検討委員会,「文書の電磁的保存等に関する検討委員会中間報告書：文書の電磁的保存等の要件について」, 2005, <http://www.meti.go.jp/press/20050111002/050111bunsyo1.pdf>