

電子情報の長期的保存—国立図書館の新たな挑戦—

国立国会図書館
関西館事業部電子図書館課
武藤 寿行

はじめに

現在、情報通信技術の進展、情報通信基盤の整備、PC の普及に伴い、情報の生成、伝達の世界で劇的な変化が進行しています。社会のあらゆる分野で情報が、電子的に生成され、インターネットを通じて迅速に、容易に、そして大量に伝達されるようになりました。大変便利な社会になっています。

しかし、一方で、インターネットで流通する電子情報は、日々更新され、消滅していく危機にあります。このため、各国の国立図書館は、電子情報を収集し、蓄積し、将来にわたってアクセスを保証するという、従来の印刷資料の保存とは比較にならない困難な課題に直面することとなりました。

電子情報の長期的保存に関する困難性の要因は、電子情報の生成、流通を促進した情報通信技術の進展そのものに内在するものと言えるでしょう。電子情報は、その生成から伝達、蓄積、再生に至るすべての局面で、情報通信技術の技術要素—アプリケーションソフトウェア、基本ソフトウェア（OS）、ハードウェアなど—に依存しています。新しいアプリケーションソフトウェアの登場や OS のバージョンアップなどにより、旧来の記録形式の電子情報が再生不能となる事態が発生しています。

このような電子情報を巡る状況のなかで、国立国会図書館（NDL）は、電子情報の長期的保存のための取り組みとして、2002 年の関西館の設立を機に、ウェブサイトの選択的収集実験事業と併せて、調査研究プロジェクトを開始しました。

この報告では、2003 年度と 2004 年度に実施したパッケージ系電子出版物に関する実態調査結果の概要を御紹介したうえで、電子情報の長期的保存の課題と今後の取り組みについてお話しします。

1. パッケージ系電子出版物の長期的保存に関する調査

電子情報は、インターネットを通じて送受信される情報であれ、CD-ROM などで頒布される情報であれ、記録形式、再生方式など基本的な技術要素は同じです。電子情報の長期的保存を考える場合、むしろ以前に刊行され、現在ではその記録形式や再生方式が旧式化していると見られるパッケージ系電子出版物を用いる方が、その課題を認識し、対策を検討する上で有効であると考えられます。そこで、NDL が所蔵するパッケージ系電子出版物を用いて、再生可能性や、データ変換などの調査を 2003 年度、2004 年度の 2 か年にわたり行いました。

1.1 再生可能性調査

2003年度のパッケージ系電子出版物の再生可能性調査では、最新のPC環境で利用可能かどうかサンプル調査を行いました。調査対象として、NDLが所蔵するパッケージ系電子出版物のうち、ゲームなど再生環境が専用のプレーヤーによるものを除いた、録音資料(音楽CD)50点と、主にPCで再生されることを前提に作成された電子資料200点を抽出しました。

調査方法は、最新のPC環境(WindowsXPとMacOSX)を用いて起動と簡易な動作確認を行いました。その結果、電子資料については再生可能な資料が200点中62点で、全体の3割に留まり、7割弱の資料に利用上問題があることが判明しました(表1参照)。録音資料については、CDが比較的安定した規格であることもあり、すべて再生することができました。

表1：電子資料の再生可能性結果

受入れ年度	調査点数	再生可能点数	割合(%)
～1990年度	29	1	3
1991年度	20	2	10
1992年度	20	2	10
1993年度	20	3	15
1994年度	20	3	15
1995年度	20	8	40
1996年度	20	8	40
1997年度	20	16	80
1998年度	20	13	65
1999年度	11	6	55
計	200	62	31

電子資料の再生不可の原因は、図1のとおりです。

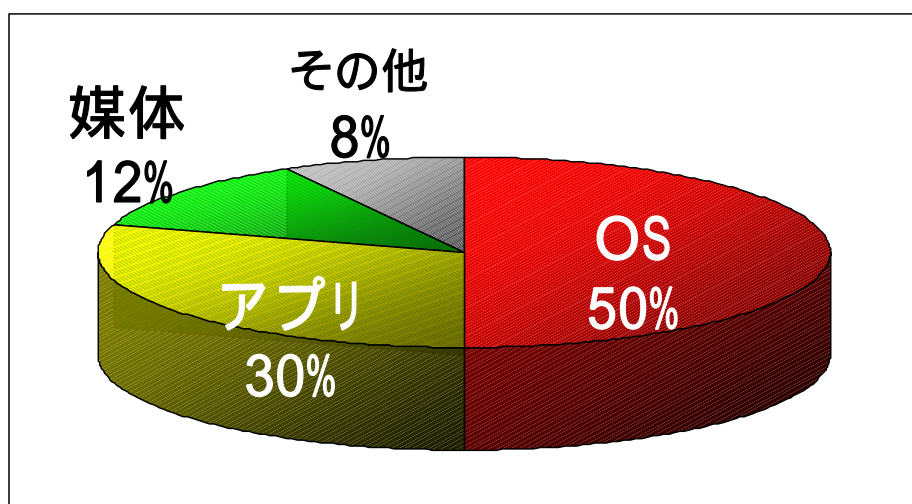
OSに原因があるものが全体の5割を占めています。電子情報は、通常特定のOSの特定バージョンでのみ動作しますので、調査で利用した最新のOSと電子資料が必要としている旧式OSのアーキテクチャの違いにより再生不可となったものです。

次に、アプリケーションソフトウェアに原因がある場合です。これは、電子資料には特定の再生用アプリケーションソフトウェアを必要とするものがあり、その再生用アプリケーションソフトウェアが特定できないか、あるいはもう存在しないため、再生できなかったもので、全体の3割を占めています。

また、記録媒体に原因がある場合で再生できなかったものが、全体の1割強に上っています。これは、5インチのフロッピーディスク(FD)のように記録媒体の旧式化により、最新のPCでは再生できないものや記録媒体自身に劣化が発生しているものなどです。

最後のその他の原因によるものとは、インストールに必要なFDが電子資料に附属しなかったことや必要な再生環境が特定できなかったことが原因で再生不可となったものです。

図 1：再生不可の原因



1.2 マイグレーション・エミュレーション試行調査

2003年度の再生可能性調査の結果を受け、2004年度は、電子情報の長期的保存対策として有効とされるマイグレーション¹とエミュレーション²を、パッケージ系電子出版物について試験的に実施し、これらの有効性と課題を探る調査を実施しました。

調査の対象は、NDL所蔵のパッケージ系電子出版物のうち、1999年度以前に受入れた国内刊行のCD-ROMであって、IBM-PC互換機用に作成されたものを対象（サンプル抽出）としました。

調査の方法は、次のとおりです。

マイグレーション（ハードディスクへのコピー）

市販のマイグレーションプログラムを使用して、サンプル354点のCD-ROMに記録された電子情報をハードディスクへ移行しました。その際、次の、の調査を行うため、電子資料に付随する取扱説明書、マニュアルなどから再生環境等のメタデータを作成しました。

エミュレーション

市販のエミュレータを用いて旧式ハードウェア環境を擬似的に再現し、この環境に旧式OSをインストールし、旧式のOS環境を構築しました。でマイグレーション（ハードディスクへのコピー）したサンプルから、プログラムを含むもの100点を選び、この旧式環境において再生確認を行いました。再生確認の内容は表2のとおりです。

¹ マイグレーション：新規記録媒体への移行やデータの変換などを指す。電子媒体の寿命は短く、データの記録形式も旧式化するため、この作業が必要となる。

² エミュレーション：再生環境の擬似的な再現。最新環境で旧式の環境を再現することで古い電子資料を再生可能とする。

表 2：エミュレーションの再生確認内容

	機能種類	機能の有無調査	再生確認内容
1	インストール	当該資料の再生にはインストールが必要か否か	インストール処理が正常に終了するか
2	起動画面	起動直後に表示される画面の有無	正常に起動して起動画面が表示されるか
3	指示に対する動作	各ボタン、エンターキーなどのオペレーションの有無	左記の各動作に対して正しく作動するか
4	静止画像	写真、テキスト、スライドなどの静止画像の有無	文字化け、画像の乱れなどの有無
5	動画	アニメーション、ムービーなどの動画の有無	文字化け、画像の乱れなどの有無
6	音声	音声の有無	雑音などの有無
7	計算、判定	計算、判定などのロジック有無	正常に計算されるかなど
8	検索抽出	検索、抽出機能の有無	検索、抽出結果がユーザの指示どおりか
9	データ記録	次回起動時までのデータ記録機能の有無	データが記録されるか
10	その他	上記以外の機能があれば、その機能概要を調査	左記の機能が正しく動作するか

マイグレーション（ファイル形式変換）

まず、 でマイグレーション（ハードディスクへのコピー）したサンプルから、データのみのも 100 点を選び、市販のファイル変換ソフトを用いてファイル形式の変換（マイグレーションの 1 種）を行いました。その上で、変換後のファイルを、そのファイル形式に対応したアプリケーションソフトウェアにより再生し、確認を行いました。また、同じサンプルによりマルチファイルビューワを用いた再生確認も実施しました。いずれの場合も、再生確認の内容は、再生されるかどうかをまず確認し、再生された場合は、文字化け、画像の乱れ、雑音の有無を確認するというものです。

調査結果は、次のとおりです。

マイグレーション（ハードディスクへのコピー）

354 点中 352 点のマイグレーションを行うことができました（表 3 参照）。マイグレーションできなかった 2 点はコピープロテクトが原因でした。なお、マイグレーションによりハードディスクにコピーされたファイル形式がマイグレーションプログラム特有の形式となり、長期的な利用に問題が生じることが明らかになりました。

表 3：マイグレーション結果一覧

刊行年度	1991-1994	1995	1996	1997	1998	1999	合計
対象点数	16	23	31	48	143	93	354
失敗点数				1		1	2
成功点数	16	23	31	47	143	92	352

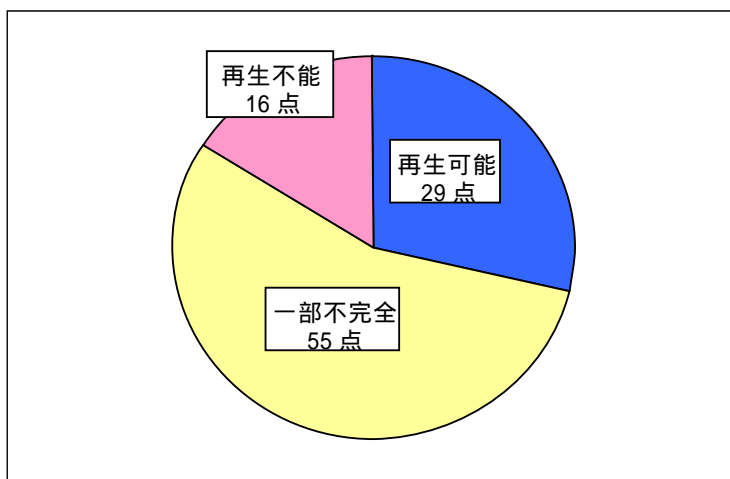
エミュレーション

全 100 点の調査対象資料のうち、問題なく再生できたものは 29 点、不完全ながらも再生できたものは 55 点、再生不能なものは 16 点でした（図 2 参照）。

インストール失敗や起動画面の不正表示、音声の乱れ、ユーザの操作通りに動作しない、文字化け、画像の乱れ、データが記録されない、などの問題が発生しました。特に音声の再生に係る問題が顕著でした。

この調査結果からは、特定のエミュレータの性能による部分があるとはいえ、現時点ではエミュレーションが必ずしも有効な手段ではないことが判明しました。

図 2：エミュレーションによる再生結果



マイグレーション（ファイル形式変換）

マイグレーション（ファイル形式変換）の対象として、まず、データのみ電子資料 100 点を選びました。このうち、HTML などの複数のファイルが関連性を持ちながらコンテンツとして再生される資料 61 点を除外しました。除外した理由は、ファイル変換ソフトはファイル単位で変換を行うため、変換によりファイル間の関連が失われてしまい、コンテンツとして再生することができなくなるためです。さらに、JPEG などファイル形式変換が不必要な規格化されたファイル形式の資料 24 点を除外しました。残りの電子資料 15 点について実施した結果、ファイル形式変換及びマルチファイルビューワによる再生確認ができた資料は 15 点中 2 点のみでした。

これは、使用したファイル変換プログラムやファイルビューワの適用可能な対象が非常に限定的であることに起因するものでした。また、調査対象から除外した HTML などの複数のファイルが関連性を持ちながらコンテンツとして再生される資料に対しては、ファイル単位での変換ではなく、資料単位で複数ファイルの関連性が維持される変換を行う必要があります。そのためには専用のプログラム開発を行う必要があります。

なお、データのみ資料 100 点について、補助的に実施した調査では、現時点で、再生に必要なアプリケーションソフトウェアを入手できるものが、指定バージョンで 73 点、最新バージョンで 96 点に上ることも判明しました。

2. 今後の取り組み

これまでパッケージ系電子出版物に関する調査結果について述べてきました。

電子情報を長期的に保存し、将来的なアクセスを保証することとは、データそのものの保存とともに、再生環境（擬似環境）もセットで長期的に維持管理することです。パッケージ系電子出版物の調査では、市販のアプリケーションソフトウェアを利用したエミュレーションやマイグレーション（ファイル形式変換）の有効性に疑問の結果が出ましたが、マイグレーション、エミュレーション自体は、電子情報保存の手法として不可欠なものです。また、再生環境の維持のためには、電子情報の属性情報、すなわち電子情報のファイル形式（データ移行した場合は、移行後のファイル形式を含む）、その再生環境（再生用アプリケーションソフトウェア、OS、ハードウェアなど）に関するメタデータの付与、維持管理が必須の条件ともなります。

一方で、電子情報の長期的保存の問題は、技術的課題を含め、保存機関だけで対応できない問題でもあります。今挙げたメタデータの維持管理・標準化やファイル形式の標準化・規格化などの課題は、国、学術機関等の電子情報の作成者、情報通信技術関連企業、アーカイブ機関等、関係機関が連携・協力して取り組まなければ解決できない課題です。

さらに、著作権の制限をはじめ電子情報の保存、将来的アクセス保証のための法制度の整備も不可欠です。この点に関する日本の取り組みについては、さきほど岡村から御報告したとおりです。

NDL は、2005 年度から、これまでの調査研究を踏まえ、NDL デジタルアーカイブシステムの開発に着手しております。その一環として、メタデータスキーマの設計を行うとともに、国際標準である OAIS に準拠した長期保存システムを構築し、これまで蓄積してきた「近代デジタルライブラリー」のデータをはじめ、制度化後に収集するインターネット情報の長期的保存に取り組んで参ります。

おわりに

貴館は、すでに電子情報のアーカイブに積極的に取り組んでおられます。電子情報の長期的保存という、我々に課せられたチャレンジングな分野においても、相互の知識技術の共有をはじめ貴館と NDL が連携・協力できる領域が多いのではないかと期待しております。