

国立国会図書館パッケージ系電子出版物を対象とした
長期保存対策
調査報告書【概要】

平成 29 年度

国立国会図書館
(電子情報部電子情報企画課次世代システム開発研究室)

(はじめに)

本報告書は、国立国会図書館が収集・保存している「パッケージ系電子出版物」について、長期にわたってアクセス可能な形態で保存していく方法に関し、平成 29 年度に株式会社データリソースに委託して行った調査の概要を取りまとめたものである。

目次

1. 調査の目的	2
2. 調査手法	2
3. マイグレーションに関する調査結果	3
(1) USB メモリに記録された資料	4
(2) CD/DVD-RW に記録された資料	4
(3) MO ディスクに記録された資料	5
(4) 録音資料 (CD-ROM、MD、DVD-Audio)	5
(5) 映像資料 (DVD-ROM)	5
(6) 再生にアプリケーションが必要な資料 (CD/DVD-ROM)	6
4. 保存メタデータに関する調査結果	7
5. 今後に向けて	8

1. 調査の目的

国立国会図書館（以下「NDL」という。）は、日本における納本図書館・保存図書館として出版物等の文化遺産を保存し、後世に残していく役割を担っている。本調査は、この役割を認識した上で、「パッケージ系電子出版物」¹の長期保存対策として、「国立国会図書館デジタルコレクション」（以下「NDL デジタルコレクション」という。）への登載を視野に入れたマイグレーションの手順確立を目指すとともに課題抽出を行うものである。

パッケージ系電子出版物には、媒体の寿命²が尽きることによるデータ消失の危険性や、再生環境の陳腐化により再生できなくなるといった問題があり、長期的な保存及び利用保証のための対策を講じることが必要である。中でも内容の変更の可能性がある、かつ媒体の寿命に関しても懸念がある CD-RW や USB メモリといった媒体は早急な保存対策が求められている。

現在パッケージ系電子出版物の利用提供は、NDL 内の閲覧室に設置された館内サービス用のネットワークに接続した端末（Windows 8.1）で行っているが、その環境では再生できない資料のためにネットワークと隔離した専用の端末（Windows XP）も用意して利用者に提供している。この専用の端末が今後用意できなくなった場合に新しい OS で再生できない資料をどうするのか、そもそも現時点でもこれらの端末で再生できない資料は多く、大きな課題となっている。さらに、アプリケーションのセットアップが必要である資料の場合は、利用のたびにインストールする手間が生じるといった運用面での課題もある。

また、本調査においては、長期保存に必要とされる保存メタデータ項目（以下「保存メタデータ」という。）の検討と付与の試行も併せて行った。

2. 調査手法

本調査の対象として 33 点のパッケージ系電子出版物を選定した（表 1 参照）。媒体としては USB メモリ、CD、DVD、MO 及び MD を対象とした。作業環境は Windows 3.1、Windows 98 (Second Edition)、Windows XP (SP2)、Windows 7 (32bit/64bit)、Windows 10 の OS を使用し、また、CD・DVD・MO・MD 用の外付けドライブを別途準備した。

作業の始めに各資料を再生できる現在の環境について調査した。その後、複製やディスクイメージへの変換を含むマイグレーション作業を試行し、NDL デジタルコレクシ

¹ フロッピーディスク、MO、MD、CD、DVD、USB メモリ、Blu-ray Disc 等の有形の媒体に情報を固定した電子出版物。平成 12 年に納本制度による収集対象となった。録音・映像資料だけでなく、企業情報や新聞のデータベース、地図アプリといったアプリケーション・ソフトウェア（以下「アプリケーション」という。）を含む CD-ROM や、学会の予稿集を収録した USB メモリなどもある。

² 媒体の物理的な経年劣化の影響を受けずに、データの読み出しに問題のない期間。

ョンへの登載に向けたフォーマット変換等を検討の上、実施した。

表1 調査対象について

媒体の種類	刊行年	点数
USB メモリ	2009 年、2012 年	2 点
CD-RW	2011 年、2012 年	2 点
DVD-RW	2010 年、2015 年	2 点
MO	2001 年	1 点
CD-ROM	1994～2007 年	9 点
DVD-ROM	2009～2015 年	3 点
(録音資料) CD-ROM	1990 年代、2014 年	3 点
(録音資料) MD	1992 年	1 点
(録音資料) DVD-Audio	2000 年	1 点
(映像資料) DVD-ROM	1990 年代～2009 年	9 点

注) 刊行年は推定を含む。

3. マイグレーションに関する調査結果

本調査では、33 点全ての資料についてマイグレーション手順を確認できた。起動時又はインストール時に認証を必要とするアプリケーションを含む 2 点の資料については、画面上の表示内容の確認はできなかったが、それ以外の資料については、マイグレーション後の画面上の表示内容も確認できた。

NDL デジタルコレクションへの登載に関しては、再生にアプリケーションが必要な資料 11 点以外の 22 点は、単純なフォーマット変換で提供可能となることが分かった。ただし、提供時に必要と考えられるメタデータの付与方法、目次情報の作成手順及び媒体特有の機能の再現に向けては更なる調査が必要である。再生にアプリケーションが必要な資料については、仮想マシン³ (VMware 等) を用いたエミュレーション⁴により、動作環境を再現することができる資料が多いことが確認されたものの、長期的な利用保証という観点では、エミュレーション技術そのものが陳腐化することが考えられること、また、こうした機能を NDL デジタルコレクションに実装すること自体が難しいといったことが課題として挙げられた。そのため、引き続き NDL デジタルコレクションへの登載を可能とするマイグレーションの実現を目指すとともに、海外で進められているエミュレーションの研究動向を注視することが必要である。

³ ここでは専用のアプリケーションを用いて 1 台のコンピュータ上に構築される、仮想的なサーバや OS を意味する。

⁴ 電子情報の利用に必要な技術的環境を擬似的に再現すること。詳しくは次のページを参照。
「よくあるご質問：電子情報の保存」<http://www.ndl.go.jp/jp/help/preservation2.html#anchor07>

以下、媒体ごとの長期保存の観点から見た特徴と今回のマイグレーションに関する調査結果について紹介する。

(1) USB メモリに記録された資料

USB (Universal Serial Bus) メモリは標準化されたシリアルバスを用いてコンピュータに接続する半導体メモリを用いた補助記憶装置であり、簡便な記録メディアとして広く使用されている。

USB メモリが読み書きできなくなる代表的な要因としては次のことが考えられる。

- (a) USB の電力供給不足
- (b) USB ドライバの不具合 (PC 側に原因がある場合)
- (c) USB 端子の接触不良 (主に端子部分の汚れ)
- (d) メモリの寿命 (電子回路の寿命)
- (e) USB メモリの異常 (フォーマットエラー、熱暴走、書き込み中の接続解除)

このうち、(a)及び(b)については、PC の交換や PC 側のドライバを入れ替えるなどで対応することができるが、(c)～(e)については USB メモリそのもののエラーであるため、対応が困難となる。

また、USB メモリはデータを記録する構造の観点から、長期的な保存には適さない媒体であるといえることができる。そのため、USB メモリで納本された資料は、できるだけ早い時期に USB メモリ以外の媒体にマイグレーションすることが望ましいと考えられる。

本調査では、エラーになったものではなく、USB メモリに記録された資料は PDF ファイルであったため、NDL デジタルコレクションで提供可能なことが確認できた。調査対象の資料は、目次情報を HTML ファイルとして持っており、この場合、HTML ファイルを基にして PDF ファイルを作成することによって、目次データを NDL デジタルコレクションに取り込む方法が考えられる。

(2) CD/DVD-RW に記録された資料

CD-RW 又は DVD-RW は、ディスクエラーが存在する可能性が高い媒体とされている。本調査においても、対象の CD-RW 及び DVD-RW 資料の全てに、事前の検査時にディスクエラーが存在していた。だが、調査環境では問題なく読み取ることができた。エラーの要因としては、読み取りに使用したドライブが原因の場合、ディスクが原因である場合、さらに両者の相性などが考えられる。ドライブが原因であるケースとしては、ピックアップレンズ⁵の汚れ、ピックアップの出力不足や寿命⁶、ピックアップレンズの

⁵ ディスクにレーザー光を照射し、データの読み書きを行う装置のレンズ部分。

⁶ データの読み書きに必要なレーザー光の出力不足や、レーザー光を発生させる装置の劣化。

脱落、サーボ回路⁷の調整のズレ（アナログサーボの登載機）、ディスク（CD）回転用のモーターの故障などが考えられる。また、ディスクが原因の場合の理由も複数考えられるが、その中で頻度が高いものとしては、ディスク表面に傷がついた場合と、記録面の劣化の2つが代表的であると考えられる。

本調査では、CD-RW 及び DVD-RW に記録された資料は全て PDF ファイルであったため、NDL デジタルコレクションで提供可能なことが確認できた。

(3) MO ディスクに記録された資料

MO ディスクはサイズが多様で、一般的な 3.5 インチや 5.25 インチのディスク以外に、3 インチ、2 インチ、8 インチ、12 インチなどのディスクも存在している。MO ディスクに関しては再生機器による読み込みの問題より、ドライブの製造が中止され入手できなくなるという問題が大きい。さらに、それらの機器を OS で認識させるためのドライバソフトウェアが新しい OS では作成されなくなるおそれもあり、できる限り早い時期に他のメディアに対してマイグレーションを行うことが望ましい。

本調査では、対象資料に対応した外付けの MO ドライブを用いて再生を行うことでデータの読み取りは可能であった。ディスク中には、マイクロソフト社の Word や Excel で作成された DOC ファイルや XLS ファイル、ジャストシステム社の一太郎で作成された JTD ファイルなどがあったが、今回の調査で用意した環境で再生可能であり、NDL デジタルコレクションで提供するフォーマットとして PDF への変換を行うことで提供が可能であることが確認できた。

(4) 録音資料（CD-ROM、MD、DVD-Audio）

デジタル形式で録音された資料としては、代表的なものとして音楽・音声などを記録した CD-ROM、MD、DVD-Audio などが存在している。近年では、これらのデジタル録音資料として販売、流通しているメディアよりも、ネットワークなどを経由して音楽・音声を内容とする MP3 や WAV などの形式のデジタルファイルを流通させることの方が多くなっているが、パッケージ系電子出版物に含まれる録音資料も、マイグレーション後は同様の形式のデジタルファイルに変換して保存することが想定される。

本調査では、対象とした録音資料は全て MP3 や WAV 形式など NDL デジタルコレクションで提供可能なフォーマットへの変換が可能であることを確認できた。しかし、MD については、前段の (3) MO ディスクと同様、再生環境の確保が今後問題となりうる。

(5) 映像資料（DVD-ROM）

映像資料で用いられるファイルの構造としては、一般的な DVD ビデオで使用されている DVD-Video 形式以外に DVD-VR など複数の形式がある。

本調査で対象とした映像資料は全て標準的な DVD-Video 形式であり、NDL デジタ

⁷ ディスクの回転数やレーザー光の照射を調整するための回路。

ルコレクションで提供可能なフォーマットに変換できることが確認できた。ただし、今回対象とした映像資料については、フォーマット変換に際して検討すべき項目として、次の点が挙げられた。

- (a) メニュー
- (b) 字幕
- (c) 多言語・副音声
- (d) その他（マルチアングル等）

(a)については、ファイルからメニューを抽出することが難しいため、スクリーンショットを撮影するといった作業が必要である。(b)及び(c)については、各言語又は音声と動画をそれぞれ組み合わせた複数種類のビデオファイルを作成する必要がある。また、(d)のマルチアングルについても同様の問題がある。大量の資料を対象として、このような編集作業を行うことは時間もコストもかかるため、NDL デジタルコレクションへの登載に向けて引き続き検討が必要である。

(6) 再生にアプリケーションが必要な資料（CD/DVD-ROM）

再生にアプリケーションが必要な資料群の定義は極めて曖昧であり、単純には分類できない。例えば、一般にはアプリケーションが必要な資料には分類されないことが多い PDF ファイルであっても、実際には Adobe Acrobat 又は Adobe Acrobat Reader というアプリケーションが必要であるという意味では「再生にアプリケーションが必要な資料」ということができる。

本調査では、NDL デジタルコレクションへの登載を目的に、「一般的に使用されている Web ブラウザの多くで広く使われているプラグインを使用しただけでは表示できない資料」を「再生にアプリケーションが必要な資料」と定義した。本調査における録音・映像資料以外の CD-ROM 及び DVD-ROM の資料は全てここに分類される。

これらの資料のうち、起動時又はインストール時に認証を必要とするアプリケーションを含むもの 2 点については、画面上の表示内容を再生することができなかったが、それ以外は、Windows10 環境上で VMware 等を用いて仮想マシンを構築し、動作する OS 環境を再現すれば実行可能であった。しかし、マイグレーションを行った資料について、仮想マシン上で動作したように見える場合でも特定の機能が動作しない資料や、Windows10 環境上の仮想マシンでは動作しない資料も存在した。また、OS や各種アプリケーションのバージョンアップなどによって、仮想マシン上で当該環境が動作しなくなる可能性もある。エミュレーション技術の NDL デジタルコレクションシステムへの実装という観点では、NDL デジタルコレクションに大規模な改修が必要となるため、今後の検討課題である。現時点で再現性を担保する長期保存の有効な方法が無いため、次善の策として、代表的なページを PDF ファイルや画像ファイルに変換することで資料の一部を保存する対策も考えられる。根本的な対策については、エミュレーション技術の調査研究を実用可能なレベルまで進め、NDL デジタルコレクションの大規模な改修が可能な機会に実装し、OS の旧式化等の再生環境の変化に対応できるように維持し

ていくこと、またそれを可能とするための体制の構築と維持が必要である。

4. 保存メタデータに関する調査結果

長期保存に必要とされる保存メタデータとは、保存対策作業やコンテンツの移管といったコンテンツの来歴情報や権利情報等を保持するために必要とされるメタデータである。本調査では、NDL デジタルアーカイブシステムを開発する際に検討した内容と、PREMIS Data Dictionary for Preservation Metadata, Version 3.0⁸（以下「PREMIS 3.0」という。）を基に検討を進め、NDL の環境と実際に行ったマイグレーション作業を踏まえた項目案を作成した。

表 2 に示したのが今回作成した項目案である。マイグレーション作業に関する項目について、PREMIS 3.0 では event と agent として個別に扱われている内容を統合して設定することや、ファイルフォーマットの判定方法を記述する項目を独自に設定することが本調査に基づいた提案である。

メタデータについて検討を進めるに当たっては、PREMIS 3.0 等のメタデータ標準で要求されるデータ項目を全て同一のメタデータファイルに含める必要は必ずしもない。しかし今後、外部機関にメタデータを提供したり、あるいはシステムリプレースなどで他のリポジトリへメタデータを移行したりする場合には標準として広まっているフォーマットに変換できるよう考慮してメタデータの設計を行うことが重要であると考えられる。

表 2 保存メタデータ項目案

項目	説明
納品アイテム番号	当該資料の識別子としての役割を担う
書誌 ID	
請求記号	
フォーマット名	マイグレーション作業前のフォーマット形式
フォーマットの版	フォーマットの版の情報
フォーマット判定方法	「フォーマット名」及び「フォーマットの版」の判定に使用した機器又はアプリケーションの情報
再生環境	マイグレーション作業前の再生環境。本作業においては基本的にアプリケーションに関する事項を記載する
再生環境の版	「再生環境」の版の情報
再生環境の製造元	「再生環境」の製造元の情報
再生環境注記	「再生環境」に関する注記事項

⁸ 長期保存に特化したメタデータフォーマットのデファクトスタンダードとされており、長期保存の中心的な機能を提供するために必要な意味単位を規定することを目標としている。現在は米国議会図書館が主に維持管理を担っている。

コピーガード	コピーガードに関する事項
マイグレーション日時	マイグレーション作業の実施日
マイグレーション詳細	マイグレーション作業に関する注記事項
マイグレーション環境	マイグレーション作業に用いるソフトウェア・ハードウェアの名称
マイグレーション環境の版	「マイグレーション環境」の版の情報
マイグレーション環境注記	「マイグレーション環境」に関する注記事項
マイグレーションファイルのファイルサイズ	マイグレーション作業により作成されたファイルのサイズ
マイグレーションファイルのフォーマット名	マイグレーション作業により作成されたファイルのフォーマット形式
マイグレーションファイルのフォーマットの版	マイグレーションファイルのフォーマットの版の情報
マイグレーションファイル製作者	マイグレーションファイルの製作者
マイグレーションファイルの再生環境	マイグレーション作業により作成されたファイルの再生環境
マイグレーションファイルの再生環境の版	「マイグレーションファイルの再生環境」の版の情報
マイグレーションファイルの再生環境注記	「マイグレーションファイルの再生環境」に関する注記事項

5. 今後に向けて

本調査の結果を受け、マイグレーション作業が比較的容易であると考えられる録音・映像資料については、作業前後のバリデーション（原資料のデータとマイグレーション後のデータの不変性の確認作業）の課題を今後検討する必要がある。また、MD等の再生環境の陳腐化がすでに懸念されている資料については、別媒体での収集状況等を考慮しつつ、再生環境の確保を第一に考えた対策を行う必要がある。

アプリケーションを含む資料については、海外でのエミュレーションの研究動向を注視しつつ、仮想マシン（VMware等）を利用し再生環境を確保する以外に、内容をどのような方法で再現して提供することが可能か、今後も検討を続ける必要がある。

保存メタデータについては、本調査ではPREMIS 3.0で必須事項とされる項目を採用した一般的かつ最低限の項目案が提案されたが、NDLのシステム環境・業務プロセスの変化を考慮してさらなる検討が必要であろう。また、今回採用しなかった権利情報

に関する項目については、資料の記載項目とメタデータ付与作業のコストを比較しながら今後検討を進める必要がある。