

第4章 デジタルデータの長期保存・利用について

デジタルデータには寿命があります。第4章は、アーカイブしたデジタルデータの長期保存に関して、なぜ長期保存の対策が必要なのか、どのような対策が必要なのかを説明しています。「3. 震災関連デジタルアーカイブでの事例」では、運用実証調査の事例を紹介します。

1. デジタルデータの長期保存について

デジタルデータは記憶媒体と再生機器のセットで考えましょう。アーカイブしたデジタルデータの長期保存に際しては、記憶媒体、再生機器それぞれについて次の課題が残されており、これらに留意しないと、せっかく構築したデジタルアーカイブを失うことにもなりかねません。後述する適切な対策の実施が必要となります。

- 記憶媒体に関する課題
 - ✓ 記録媒体には寿命がある
 - ✓ 保存環境が悪いと、寿命が極端に短くなる場合がある
- 再生機器に関する課題
 - ✓ 記録媒体を適切に保管しても、内容を読み出す機器がなくなる

表 4-1 デジタルデータの保存可能期間

	メディア（記憶媒体）	再生機器（ハードウェア）	再生機器（ソフトウェア）
保存期間	メディアがデータを保存していること(媒体寿命)	メディアからデータを読み出し、システムのメモリに格納できること	ハードウェアで読み出したデータを他のプログラムやユーザーが取り扱えること
3年	通常使用可能	データの保存で使用したハードウェアが使用可能	対応ソフトウェアで使用可能
10年	テープ・光ディスクなら読み出し可能	新製品は無くても、動作するものが残っている可能性が高い	上位互換システムで使用できる場合もある
30年	テープ・光ディスクなら読み出し可能という資料もある	通常、残っていない	通常、残っていない
100年	電子媒体では難しい	残っていない	残っていない

出所：(社) 電子情報技術産業協会 テープストレージ専門委員会
「コンピュータ用テープによるデジタルデータの長期保存」より MRI において作成

http://home.jeita.or.jp/upload_file/20111221165535_MvQbSIymHi.pdf

デジタルデータにはこのような課題が残されているため、バックアップが重要です。記録媒体による保存の他に、別途バックアップシステムを構築することも重要です。バックアップシステムは、なるべく複数の方法を用意することが望まれます。さらに災害に備えるには、最低一つは、同時に被災することのないように遠隔地に保管しておくことも検討したほうが良いでしょう。

2. 長期保存のために考慮すべきポイント

(1) 記憶媒体の選択

アーカイブしたデジタルデータは、その保存期間中に媒体の劣化等の理由によって使用できなくなることがないように、「保存性」に留意しなければなりません。保存性を確保するためには、保存対象となるデータを記録するための記録媒体の寿命が重要です。記録媒体には、長期的な保存に耐えうる、高品質で信頼性の高い媒体を使う必要があります。ただし、記録媒体の寿命は、一般に、紙（中性紙）ならば適切に保管すれば 100 年以上は持つと考えて良いですが、電子媒体の場合は、数十年が限界とされています。

(2) 記憶媒体の適切な保存

デジタルデータの記憶媒体は光や高温多湿といった環境要因によって、極端に寿命が短くなってしまうケースもあります。擦れや反りを防ぐための媒体の格納方法への留意、劣化の大敵である光と高温多湿を防ぐといった保存環境への配慮を行ったうえで、データが正しく記録され、保存した内容が正しく読み出されるかを定期的を確認する必要があります。

(3) 再生機器（ハードウェア、ソフトウェア）の選択

紙や写真等の場合は、情報の内容を必要に応じて肉眼で容易に見ることができます²⁷が、電子媒体の場合は、適切な再生用のハードウェアとソフトウェアがなければ、そのままでは内容を見る事ができないという問題があります。そのため、再生用のハードウェアとソフトウェアの寿命や将来的な入手可能性を考慮する必要があります。

JEITA の委員会²⁸ の資料²⁹によれば、「長期保存分科会による想定であり、実際のケースにおいて必ずしもあてはまるものではない」としつつも、保存期間が 30 年になると、記録媒体はまだしも、再生用のハードウェアとソフトウェアについては、「通常、残っていない」としています。データが再生できない事態に陥った場合のための救済手段として、JIS Z 6017 という規格³⁰があります。

²⁷ このような性質を「見読性」という

²⁸ 一般社団法人 電子情報技術産業協会 テープストレージ専門委員会

²⁹ 「コンピュータ用テープによるデジタルデータの長期保存」

http://home.jeita.or.jp/upload_file/20111221165535_MvQbSIymHi.pdf

³⁰ JIS Z 6017 では、記録媒体に管理番号を印字し、それとリンクさせた形で管理台帳を作成・維持するという方法を定めている。また、管理台帳に記載する項目をダブリンコア（メタデータを記述する標準的な項目のうち広く普及しているものの一つ）の規定に準拠させており、ダブリンコアのメタ情報の基本項目 15 個との対応関係を一覧として示している。この方法により、仮に当該の媒体に記録されている情報が容易には利用できない事態になったとしても、それがどのような情報であったのかを把握することで、復元のために必要

(4) ファイルのフォーマット

再生用のハードウェアとソフトウェアについては、記録時点と同一の機能を持つ機器が入手できなかったとしても、その後継機種が利用できる可能性があります。

ただし、後継機種や代替機種で利用可能であるためには、記録時点で用いたフォーマットが、後継機種や代替機種でも利用可能でなければなりません。そのためには、記録時点において（完全に将来を予測する事はできないまでも）できるだけ長期的・安定的に利用できる標準的なフォーマットを選択し、そのフォーマットにしたがって記録しておくことが望まれます。

現時点では「長期保存用デジタルフォーマット」と呼べるような、長期保存が保証されたフォーマットは存在しません。実際にフォーマットを選択する際には、次のような要素を総合的に判断して決めることが重要と思われれます。

- 標準化、規格化がなされているか
- 一般に広く普及しているか
- 技術内容は誰でも知ることができるオープンなものか
- オリジナルの情報に戻ることができる（可逆）圧縮か、そうではなく非可逆圧縮か

採用するフォーマットは、こうした要素と、ストレージやマイグレーションにかかる費用等を勘案して決めることとなります。

デジタルアーカイブの実績がある各種機関が長期保存を意識して採用しているフォーマットを採用するのもよい方法でしょう。参考として国立国会図書館が、平成 21 年度に「国内外の機関に対してアナログ形式の録音・映像資料のデジタル化実施状況」等について実施したアンケート調査結果³¹をご紹介します。

表 4-2 国内外の機関における利用フォーマット

区分	フォーマット
録音保存用	回答機関のほとんどが WAV 又は BWF を採用
録音提供用	回答機関の半数以上が MP3 を採用
映像保存用	Motion JPEG2000、JPEG2000、MPEG2 を採用する機関が複数あり その他、MPEG4 等
映像提供用	MPEG2 を採用する機関が複数あり

出所：国立国会図書館「国内外の機関に対してアナログ形式の録音・映像資料のデジ

な機材、費用等がある程度見積もることができる。

³¹平成 21 年度 国立国会図書館

「録音・映像資料のデジタル化と電子情報の長期保存に関するアンケート調査」

http://www.ndl.go.jp/jp/aboutus/d-library/pdf/summary_2009_jp.pdf

タル化実施状況」を元に（株）三菱総合研究所が作成しました。

また、国立公文書館が採用しているフォーマットについては、本ガイドライン「第3章 資料・記録のデジタルデータ化について」をご覧ください。

3. 長期保存のための対策

(1) 定期的なデータ移行（マイグレーション）

上記の配慮をしたとしても、デジタルデータの長期保存のためには、定期的なデータ移行が必要です。予算の確保を含め、アーカイブを構築する段階から移行計画を立てておくことをお勧めします。

古いフォーマットや媒体を、それらを再生する機器がまだ存在する間に、新しいフォーマットや媒体に変換しておくことをマイグレーションと言い、マイグレーションのためのツールが市販されています。例えば、市販のマイグレーションツールにより、CD-ROMの内容をHDDに移行するなどの方法があります。

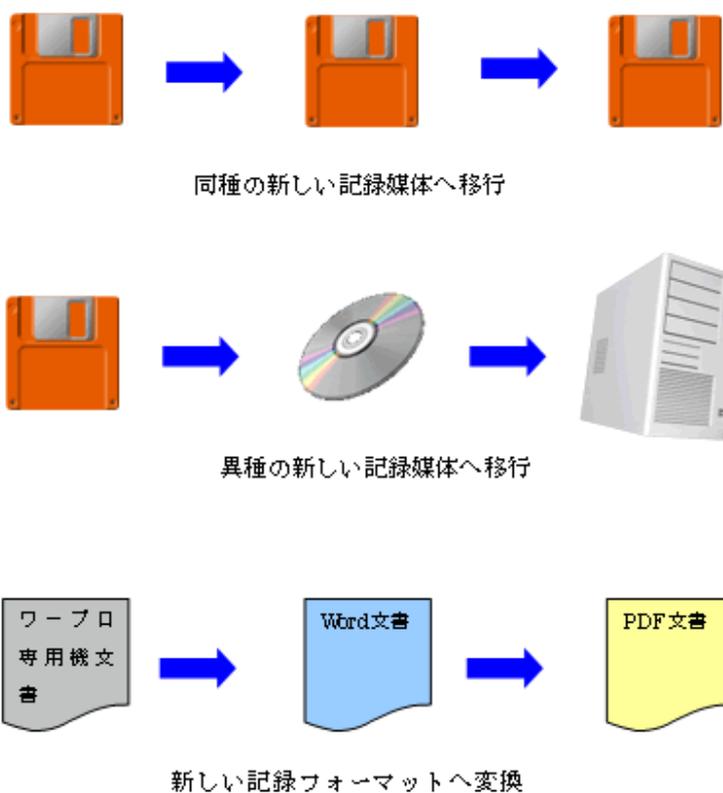


図 4-1 マイグレーションの例

出所：国立国会図書館：電子情報の長期的な保存と利用 > FAQ（よくある質問）より

http://www.ndl.go.jp/jp/aboutus/preservation_03.html#q09

(2) エミュレーション

エミュレーションとは、新しいハードウェアや OS 等の上に、エミュレータと呼ばれるソフトウェア等を搭載し、古いフォーマットや媒体のままのオリジナルのデータを疑似的に再生できるようにすることです。

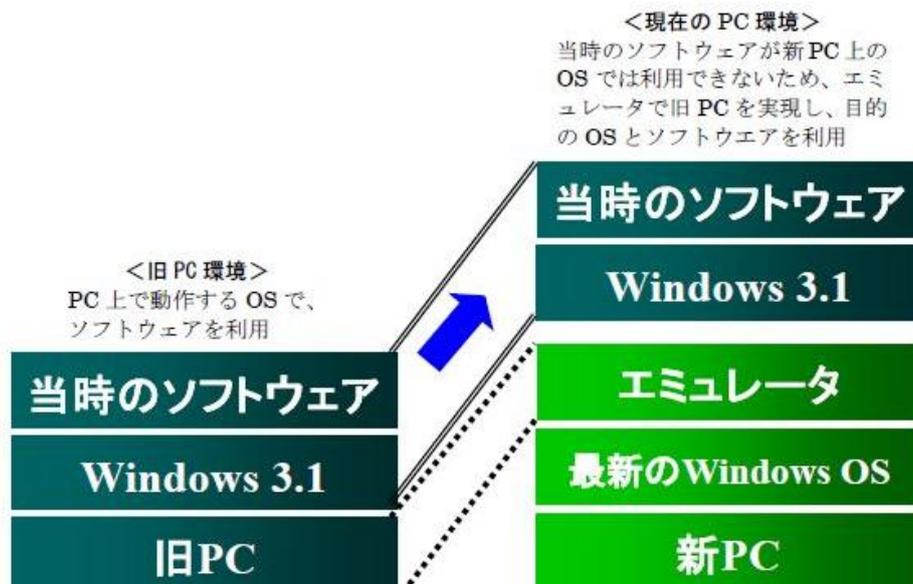


図 4-2 エミュレーションの構造

出所：国立国会図書館：電子情報の長期的な保存と利用 > FAQ（よくある質問）より

http://www.ndl.go.jp/jp/aboutus/preservation_03.html#q09

マイグレーションするに当たりオリジナルのデータを再生できなくなってしまうような場合には、エミュレーションが必要となります。

4. 運用体制・運用費用の整備

上述したような技術面での課題の他に、デジタルアーカイブを維持するために必要となる適切な人材を育成したり、そうした人材を長期的に安定して維持したりするための運用体制の整備もとても重要な課題です。加えて、年を追って増えていくことが想定されるデジタルデータを維持するために必要な予算の確保も避けて通れない重要な課題となります。さらに、人材、体制、予算等を管理するマネジメント能力も重要な要素の一つです。

(1) 運用体制と予算の確保

デジタルデータを蓄積すればするほど、マイグレーションに要する費用は増大します。そのため、UNESCO の「デジタル遺産保存のためのガイドライン」では、「デジタル情報

の長期保存に際して、前提として知っておくべきこと」として、「保存対象の選択が必要」であることに言及しています³²。

本来であれば、データの保存性を確保するために、定期的に、かつなるべく早めに（あまり長期にならないうちに）広くマイグレーションを実施することが望まれますが、マイグレーションの対象が増え、費用が増大する中で、マイグレーションのみに限りのある予算を投入するわけにもいきません。そのため、投入できる予算と必要な費用を勘案しつつマイグレーションのタイミングと範囲を調整していく必要があります。

マイグレーションの間隔を長くすれば、費用は下げられますが、データを喪失するリスクは増大します。逆に、マイグレーションの間隔を短くすれば、データを喪失するリスクは下げられますが、費用が増えてしまいます。

現実問題としては、マイグレーションのタイミングを決める大きな要因として、費用の要因は避けられません。そのため、あらかじめマイグレーションに要する費用も含めて、デジタルアーカイブ構築・運用の計画を立案したり、フォーマットや媒体を決定したりする必要があります。

ただし、仮に一定のマイグレーション間隔（例えば「〇年ごとに」等）を定めたとしても、その期限が到来するより前に、何らかの要因でデータが利用不可能となるおそれもあります。そのため、真に長期保存が必要なデータであれば、マイグレーションの間隔よりもさらに短い間隔で定期的に点検を行い、その結果としてデータ喪失のリスクが懸念されるようであれば、速やかにマイグレーション等を行うべきであると考えられます。

デジタルアーカイブは持続可能性（sustainability）が重要であり、長期にわたる継続的な維持や保守のためには、運用を目的とする組織や部署を新たに設置するか、既存の組織や部署が運用を担当する場合は、そこでの正式な業務として位置付けを明確にしておく（明文化しておく）ことをお勧めします。

(2) クラウドの活用

長期保存のために必要なシステム環境を構築したり、体制を確保したりすることの代替手段の一つとして、クラウドによるデータストレージサービスやアーカイブサービスを利用できる可能性があります。

すなわち、バックアップシステムとしてクラウドを利用する事で、サービス利用側としては記録媒体の事を意識する必要はなくなりますし、またオンラインで自動的に複製することができれば、容易にデータのバックアップシステムを構築できる可能性もあります。

³² 国立国会図書館：「電子情報の長期的保存とアクセス手段の確保のための調査報告書」、平成 16 年 3 月、http://www.ndl.go.jp/aboutus/report_2004.pdf

SaaS（Software as a Service）型³³のアーカイブサービスを利用すれば、再生に必要なソフトウェアの更新もサービス提供者で実施するため、サービス利用者側が気にする必要はありません。

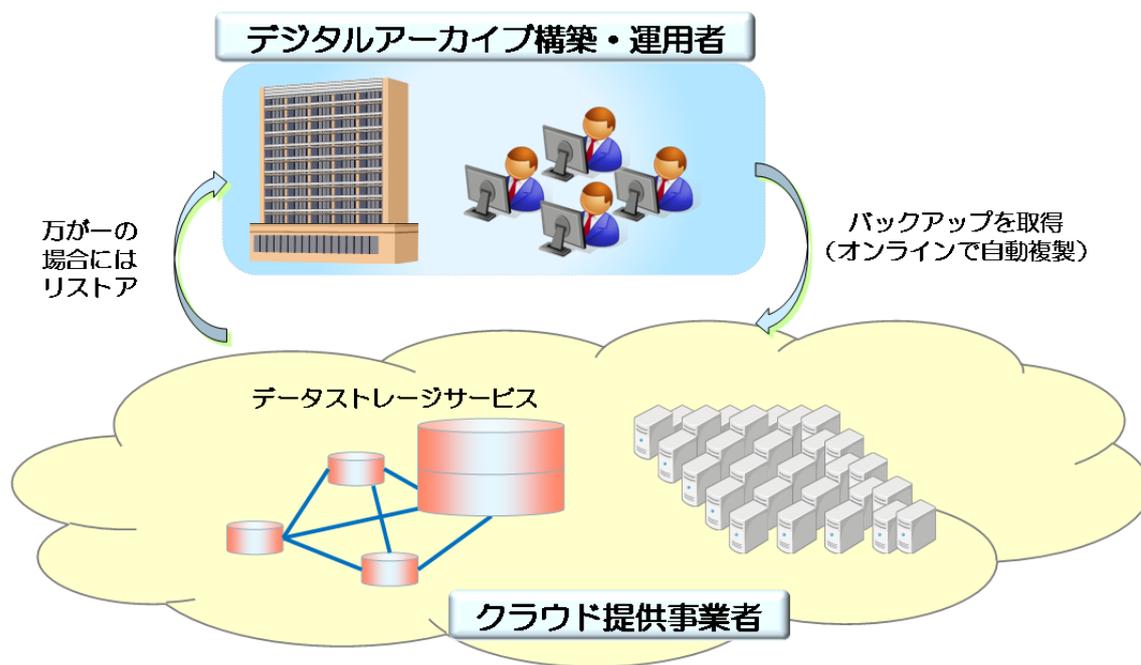


図 4-3 クラウドによるデータのバックアップシステムの実現イメージ

また、バックアップシステムの運用体制の検討・構築・運用を省力化でき、デジタルアーカイブ運用側の作業（特にバックアップ作業）負荷の軽減につながることも可能です。

加えて、無料のデータストレージを利用することができれば、デジタルアーカイブの構築・運用に要する費用の低減を図ることも可能となります³⁴。

ただし、クラウド側のシステム事故や、クラウド提供事業者の事業環境の急変（例えば倒産や企業買収など）によって、データの保全がおぼつかなくなったり、最悪の場合にはデータが消失したりしてしまう可能性も考えられます。そのため、仮にクラウドを利用するとしても、万が一に備えて、ローカルにもバックアップを取得しておくことをお勧めします。また、海外にサーバを置いているクラウド事業者もあります（日本の企業でも海外にデータセンターを設置・運用しているケースもあります）が、その場合、サーバが設置されている国の法律に従うこともありますので、選定の際にはコンテンツを預けるサーバの設置場所も確認しましょう。

³³ 必要な機能をサービスとして提供するタイプのクラウドサービス。利用者はハードウェア、ソフトウェアとも不要だが、インターネットでサービスを利用するため、インターネットが利用できる環境を持っていることが条件となる。

³⁴ 例えば、無料のブログ作成サイト、画像（写真）共有サイト、動画共有サイト等を用いたアーカイブの構築事例もある。

なお、クラウドサービスは、ハードウェアを購入したりソフトウェアを開発したりすることなく利用できるため、初期費用は発生しませんが、利用し続ける限りサービス料金を支払う必要があります。長期的な利用を考えている場合には、アーカイブ構築者自身で運用する場合の費用と比較する必要があります。

(3) 非公開のアーカイブ（ダークアーカイブ）の活用

学術情報資源の保存を目的としたものですが、デジタルコンテンツの長期保存を目的とした CLOCKSS³⁵（Controlled Lots of Copies Keep Stuff Safe）というアーカイブのプロジェクトがあります。図書館と大手出版社が共同で設置した非公開のアーカイブ（ダークアーカイブ）で、論文等のデジタルコンテンツを世界各国の大学図書館等に分散して保管しておき、例えば出版社の倒産等により、万が一、デジタルコンテンツへのアクセスが失われる危機に瀕した場合は、このアーカイブを活用してアクセスできるようにすることを目的としています。

このプロジェクトは目的限定ですが、公開しないことを前提とした保存のためだけのアーカイブをクラウドサービスなどを利用してアーカイブ構築者自身で運用することも考えられます。

5. NDL 東日本大震災アーカイブについて

NDL 東日本大震災アーカイブでは、他機関の震災関連デジタルアーカイブのメタデータ及びコンテンツを受け入れており、バックアップ機能として活用することも考えられます（ただし、国立国会図書館の運用ポリシー等に従う必要があります。）。また、アーカイブ活動が維持困難となった場合、他に適切な引き受け機関が見つからないときには、国立国会図書館がその記録等を受け入れる、記録保管のラストリゾート機能としての役割を担うことも想定しています。

6. 震災関連デジタルアーカイブでの事例

(1) 運用実証調査における長期保存・利用のための検討

(a) 長期保存のためのファイル形式について

デジタルコンテンツを長期保存するため、運用実証調査では、以下の基準で保存のためのファイル形式を決定しました。

- 写真は、広く使われている、JPEG 形式での保存とする。

³⁵ <http://www.nii.ac.jp/content/justice/project/clockss.html>

- 音声は、オリジナルデータ又は内容をテキスト化し、テキスト形式での保存とする。
- 映像は、様々なブラウザで視聴可能な形式での保存とする。
- 文章は、広く使われている PDF、Word、Excel 形式での保存とする。

提供されたデジタルデータのファイル形式が上記の基準と異なっていた場合には、そのままの形式とするか、ファイル形式を変換するかを各プロジェクトで判断し、保存することとしました。ただし、ファイル形式を変換した場合、変換前のデジタルデータも保存対象としました。

以下に各プロジェクトが、採用したコンテンツ種別ごとのファイル形式を示します。

表 4-3 運用実証調査におけるファイル形式

地域	写真	音声	映像	文章
青森	JPEG TIFF	—	MP4 ASF MOV	PDF Word Excel
	上記形式で受領	—	ASF、MOV 形式で受領したものは MP4 に変換	JTD (一太郎) 形式で受領したものは PDF に変換
岩手	JPEG	MP3	MP4 FLV AVI MTS	PDF Word Excel
	上記形式で受領	上記形式で受領	MTS は、プレビュー生成可能な MP4 形式に変換	上記形式で受領
宮城 (東北大学)	JPEG	テキスト	対象なし (MP4/webm を想定)	PDF
	解像度オリジナル、圧縮率オリジナル、JPEG が広く流通しているため	音声は、そのままアーカイブしても内容がわからないため	HTML5 形式でブラウザ毎に対応する形式が違うため	広く使われている形式であるため

地域	写真	音声	映像	文章
宮城 (河北新報社)	JPEG	—	対象なし (MP4/webm を想定)	—
	解像度オリジナル、圧縮率オリジナル、JPEG が広く流通しているため	—	HTML5 形式でブラウザ毎に対応する形式が違うため	—
福島	JPEG BMP	MP3 WAV	MP4	PDF Word Excel TXT
	上記形式で受領	WAV 形式で受領したものはファイルサイズから MP3 に変換したが、オリジナルデータも保存対象	受領した MOV 形式では再生ソフトが限定されるため	上記形式で受領

(b) 長期保存のためのファイルサイズについて

運用実証調査のうち、宮城県、福島県の 3 つのプロジェクトでは、長期保存のため、保存形式はオリジナルデータのままとし、公開用にリサイズを実施しました。

公開用にリサイズを実施し、オリジナルデータを保存することは以下の観点から、二次利用が行いやすくなると考えられます。

- ▶ システム負荷の観点から、閲覧用にはプレビュー画像を用いることでネットワーク負荷を軽減することが可能です。
- ▶ プレビューのサイズについては、詳細な画像がわかることを目標にサイズを決定すべきであるものの、実証の結果 72dpi 程度で問題ない場合が多く、1 つの目安として考えられます。