

文化財デジタルデータ長期保管の実務

高田祐一（奈良文化財研究所）

Long-term Storage of Digitized Heritage Data

Takata Yuichi (Nara National Research Institute for Cultural Properties)

・データ長期保管／Long-term data storage・ファイルフォーマット／File format
・保存媒体／Storage medium

1. 調査研究成果は情報資産

考古学・文化財の調査研究成果は、貴重な情報資産であり、恒久的な保管が求められる。なぜなら、考古学において、発掘調査は不可逆的行為であり、その成果となる発掘調査報告書は遺跡の身代わりとも位置付けられるからである。また、考古学は蓄積型学問であり、過去の成果に基づいて学問分野が深化し広がっていく。そのため、過去の調査研究成果に容易にアクセスし、利活用することが可能となれば、調査機関あるいは個人の効率的・網羅的・高次の研究を支援する基盤となる。

よって、将来にわたり調査研究成果を情報資産として適切に管理していくにあたっては、まずデータの長期保管の実現が不可欠である。

2. データが消失する要因

データが消失する要因は様々である。DataONEによれば、Natural disasters（自然災害）・Facilities infrastructure failures（インフラの障害）・Storage failure（ストレージの障害）・Server hardware or software failure（サーバの物理故障あるいはソフトウェア障害）・Application software failure（アプリケーションソフト障害）・Human errors（人為ミス）・Malicious attack（悪意ある攻撃）・Format obsolescence（電子フォーマットの廃絶）・Loss of competencies（実施能力の消失）・Loss of funding

（予算の消失）・Loss of institutional commitment（制度根拠の消失）が挙げられている（DataONE）。

3. データ長期保管のリスクと対応

DataONEは「Costs of not doing data management can be very high!（データ管理を実施しないコストは高くなる!）」と警告している。データ消失が起きた後で、データ復旧をするためには、膨大な費用をかけることになるか、もしくはデータが永遠に失われる事態に直面する。普段から適切にコスト（費用・手間など含む）を投下することで、結果的に全体のコストを低減させることができるだろう。起こりうるリスクを整理したうえで、業務やサービスを維持するための予防コスト（Preventive Cost）と損失が起きた際の失敗コスト（Failure Cost）の観点から検討する必要がある。

デジタルデータ長期保管の際のリスクは、文化庁によれば以下のとおりである（図1）（文化庁2017）。

- ①物的リスク：機器故障、媒体不良等の物的要因により安定保存が妨げられる状態
- ②人的リスク：作業者の作業ミス等の人的要因により安定保存が妨げられる状態
- ③災害リスク：地震、火災等の外的要因により安定保存が妨げられる状態
- ④情報セキュリティリスク：セキュリティの脆弱要因により安定保存が妨げられる状態

それぞれのリスクへの対策は以下の通りである。

①物的リスク

- ・機器を冗長化（サーバやネットワークなどの機器故障等による負荷の急増に備えて、必要とされる設備よりも多めに予備設備を準備しておくこと）する。
- ・データをコピー分散して保存する。機器が故障していないか毎日点検する（チェック体制の構築）。

②人的リスク

- ・データを操作できる人間を特定する（不必要な人には触らせない）。
- ・データをコピー分散して保存する。

③災害リスク

- ・データをコピーし別地保存（遠隔地保存）する。
- ・分庁舎にもデータ保存。
- ・提携自治体とのデータ相互持合い。

④情報セキュリティリスク

- ・データを保存した機器のセキュリティを万全にする。
- ・パッチ（ソフトウェアの欠陥修正等を目的に、ソフトウェアの書き換えを行うプログラム）適用と

ウイルス対策を万全にしたパソコンでデータを操作する。

- ・データは必要に応じてアクセス制限を設定する。
- ・バックアップデータの暗号化。機器の施錠（持ち去り防止）。

各組織は、有限の予算・人員配置・機材でデータ長期保管に取り組む必要がある。それぞれの各組織の文化・リソースなどは異なるため、組織ごとにベストミックスの最適解を模索する必要がある。結局はどのリスクをどの程度受忍するかというリスク評価の問題に決着する。優先度高で守るべきデータは何か、最悪の事態にあきらめざる得ないデータは何か、代替できるのかという整理と組織的合意が必要である。

4. ファイルフォーマットの選択

一般的にデジタルデータ自体は保管媒体よりも寿命は長い。しかしデジタルデータ自体にも寿命は存在する。それは電子ファイルのフォーマットが死んだときである。フォーマットに関してリスク

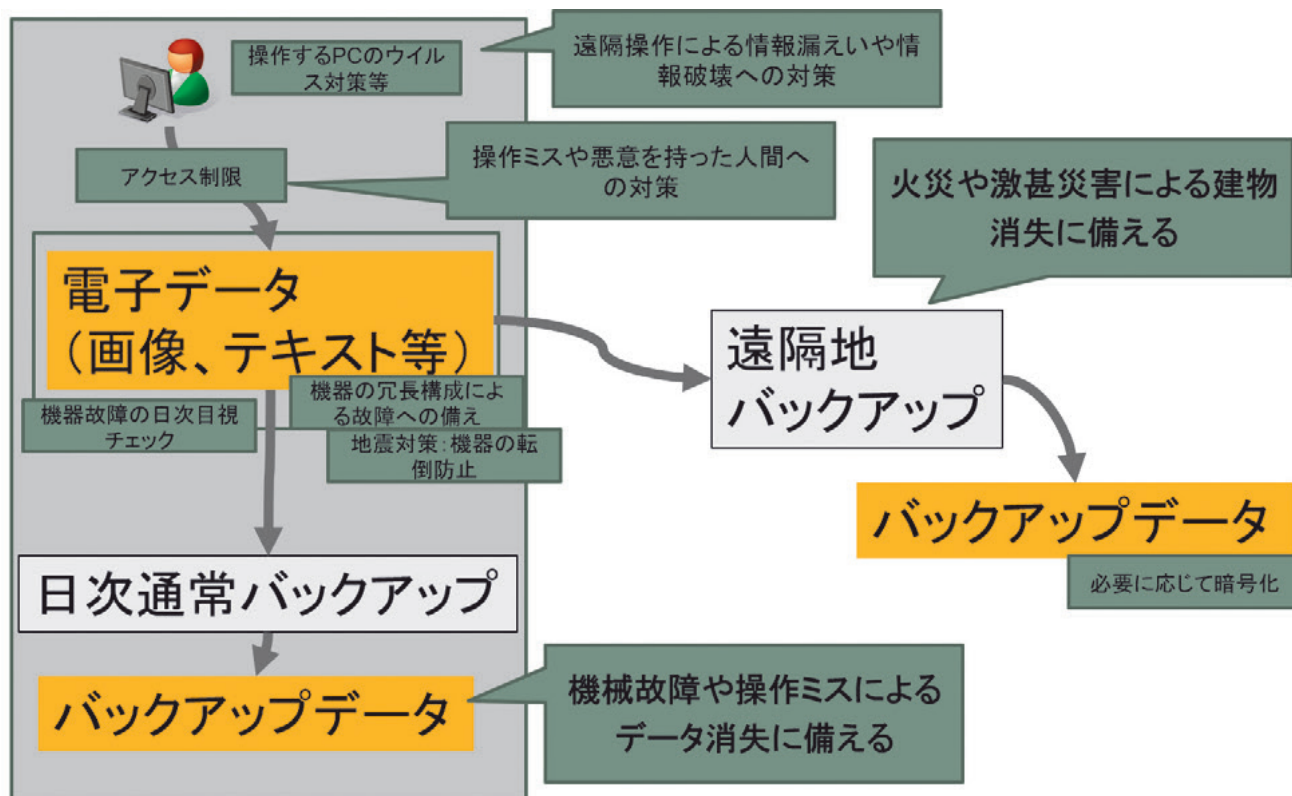


図1 リスクと対策案（文化庁『埋蔵文化財保護行政におけるデジタル技術の導入について1』（報告）より転載

が高いのは、特定の会社、特定の製品に依存しているフォーマットである。会社倒産・事業終了・ソフトウェア製品の開発終了などに影響されるほか、ソフトウェア製品の値上げによってソフトウェアを維持できなくなるケースもありうる。データ保管におけるフォーマットへの考え方は、イギリスの Archeology Data Service とアメリカの Digital Antiquity が協力して作成した「Guides to Good Practice」にて整理されている（Archeology Data Service / Digital Antiquity）（図 2）。各フォーマット形式の評価については執筆者による日本語化のうえ、別表にて要約を示す（本誌別稿「文化財デジタルデータ長期保存の為のファイル形式」）。

データ長期保管に適したフォーマットは、特定の会社・製品に依存しない形式である。つまり、フォーマット仕様がオープンになっていること、ISO など国際標準化されていること、技術的に単純な構造であることなどである。しかし単純な構造であるゆえに、複雑なことを実現できず、利便性の低下につながることもある。使用に不便な形式は、将来的に利活用のメリットも減少するため、その他の形式と合わせて複数の形式で保管する方が良い。また、その際は、デファクトスタンダードとなっている形式がのぞましい。もしその製品が廃絶となった場合でも、社会的影響の大きさから互換性のある形式が登場する確率が高いためである。陳腐化した形式であってもコンバージョンを繰り返すことによって、

データ自体を将来にわたって継承していくことが可能である。

5. 保存媒体の選択

2019年現在、様々な保存媒体がある。大別すると、磁気テープ、光ディスク、HDD、クラウドサービスなどである。それぞれ長短があるため、特性を理解したうえで媒体を選択する必要がある。それぞれの特徴は表 1 の通りである。媒体の長短を考慮しながら組織の予算・人員体制やデータ量などによって検討する必要がある。

参考例として、組織に IT 専門家がない場合のモデルケースを示す。まず通常の運用においてハードディスクを活用する。ハードディスクは安価で大容量を容易な操作で保存できる。ただしハードディスクは故障しやすいため、定期的なバックアップが必須である。可用性（安定稼働）が求められるなら RAID 構成によるディスクの冗長性が有効である。PC がウイルス感染した際、ハードディスクがオンラインもしくは USB 接続していた場合、ハードディスク内のデータまで消失する可能性がある。建物が火災にあった際、データがハードディスクか所だけの保存であれば、消失するリスクがある。これらを解決するために、物理的な媒体に定期的にバックアップし、その媒体を別地や別建物などに保管することが有効である。オフラインであれば、ウイルス

表 1 デジタル化された文化資源の保存環境（2012 現在）
『博物館学Ⅲ - 博物館情報・メディア論 * 博物館経営論』
2012より引用

項 目	内 容
ハードディスク	一度に大容量のデータを扱うことが可能であるが、定期的な動作チェックを必要とする。また書き込みが容易であるため、ウイルスなどの侵入などに注意が必要。
Blu-ray	容量は多いが、メディア、読み取り、書き込み機器とも高価。永続性の点でもまだ実績が十分とはいえない。
DVD	メディア、読み取り、書き込み機器とも安価だが、大容量のデータを扱う場合は、HDD と比較して複数枚に分割する必要があるなど手間がかかる。またメディアの永続性の点で注意が必要。書き込み方式については DVD-R、DVD-RW など複数の規格が存在する。
CD	メディア、読み取り、書き込み機器とも安価だが、DVD と比較して容量が少なく、メディアの永続性の点で注意が必要。書き込み方式については CD-R、CD-RW など複数の規格が存在する。
テープ（DAT など）	保存メディアとして、長い実績がある。速度がディスクと比較して低速であり、テープデバイスが高価。
その他	USB メモリ、フラッシュメモリなどのリムーバブルメディアは、長期保存に安定的なメディアとはいえない。



図 2 Guides to Good Practice

感染のリスクもない。媒体を施錠環境にて保管することで悪意ある犯行からも保護できる。その媒体は、データ量が少なければ長期保存用 DVD、データ量が大きければ ODA（オブティカルディスク・アーカイブ）が候補となる。磁気テープは、速度が遅い点と非専門家にはハードルが少し高い点がネックとなりうる。

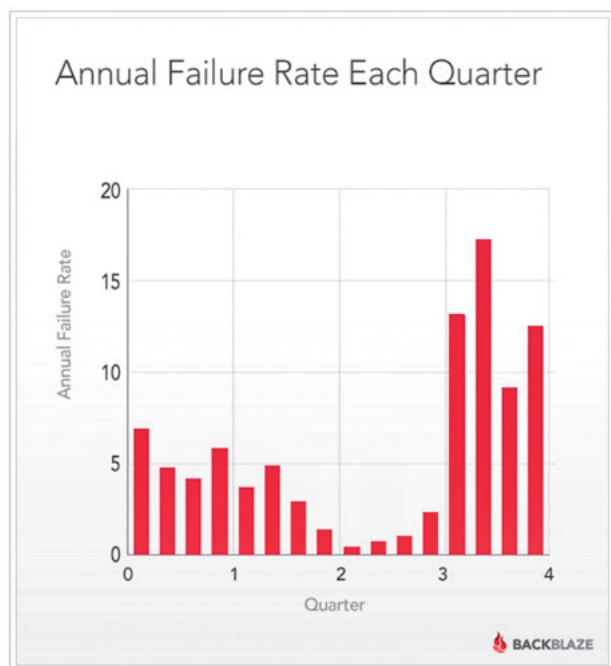


図3 ハードディスクの四半期毎故障率

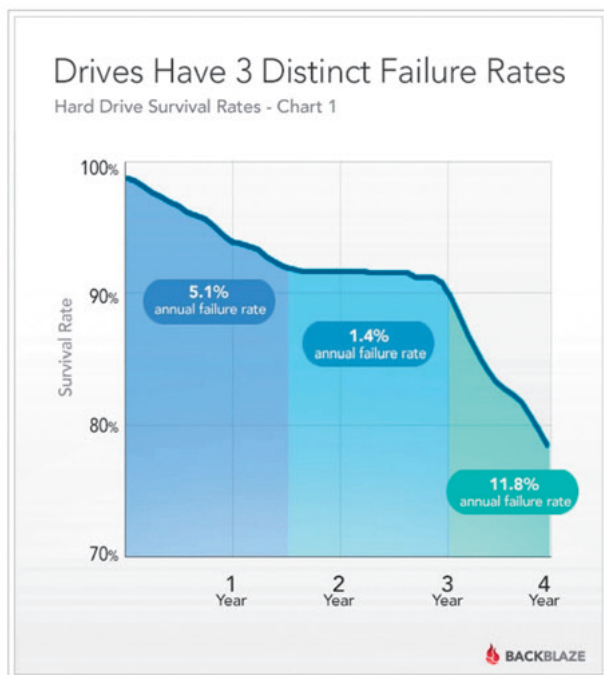


図4 ハードディスクの年間故障率

媒体の選択に絶対解はない。技術や製品は消長がある。超長期の未来を見通すことは困難である。その前提で、ある程度見通せる10年ごとに最適なものを見直すのが良いと考える。

6. ハードディスクの寿命

ハードディスクはコスト面や利便性に優れていることから多数使われているが、障害リスクは高い。ハードディスクの障害は、論理障害と物理障害がある。論理障害とはデジタルデータに不整合が起こることである。物理障害とは物理的に機器が破損することである。これらの障害が起きる原因として、株式会社バッファロー社は次の点を挙げている（BUFFALO）。経年劣化、衝撃、動作中の急な電源オフ、水漏れ、高温・強い磁場など過酷な環境での使用、分解・改造である。

では、ハードディスクの寿命はどのくらいか。ストレージサービスを展開している Backblaze 社は、ハードディスクの故障率を公開している（Brian

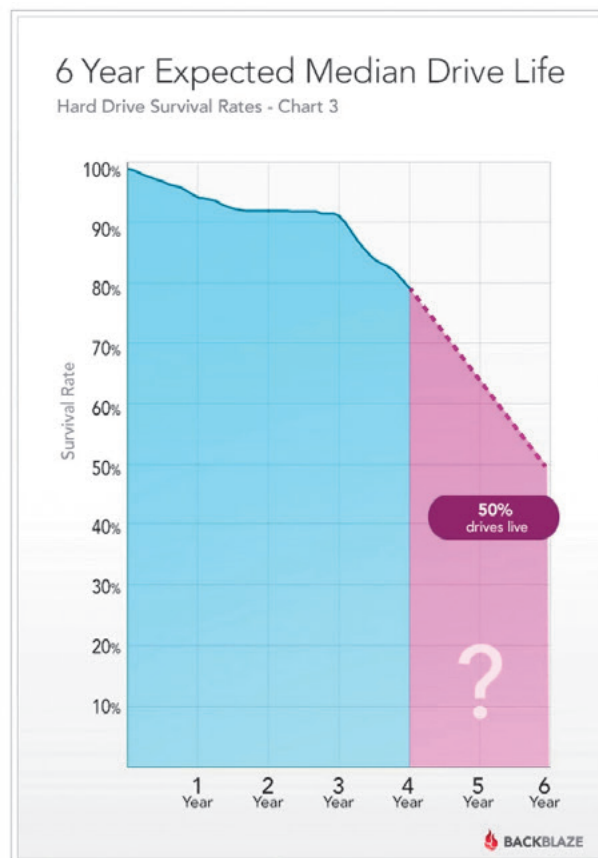


図5 ハードディスクの年間生存率

Beach 2013)。2013年11月に公開した記事を紹介する。故障率は最初の1年は5.1%である。2年目は安定し1.4%となるが、3年目は11.8%となり、バスタブ曲線の傾向をたどる（図3・4）。満4年で生き残っているハードディスクは約80%である。4年目以降は急速に故障率が高まり、6年目には生き残っているものが50%と予測される（図5）。運が悪ければ1年以内に故障するが、運が良ければ10年以上稼働するかもしれない。ただし恒久的に残すべき調査研究の成果を賭けによって保存するのは適切とは言えないだろう。ハードディスクは壊れるという前提で、2台のハードディスクに同じ内容を書き込むなどして冗長化を実現する必要がある。4年後故障の中央値が20%であっても冗長化していれば、2台とも故障する確率はかなり下がる。さらに2台がまったく同じ瞬間に壊れる可能性は極めて低い。ハードディスクを冗長化あるいはバックアップすることで、貴重なデータを保護することができる。

7. 運用体制はいかにあるべきか

データを確実に保全し将来に継承するためには、日々のモニタリング作業が不可欠である。エラーが発生していないかストレージの監視、バックアップ処理の結果確認、ストレージや保管媒体を格納している部屋の温湿度確認等である。これらのモニタリングは相応の手間が発生するため、自動監視を導入し、異常時だけアラートメールを発出するなどの仕組みによって効率化した方が良い。ただし自動監視自体が異常の場合、機能しないため必ず人間でも定期的にチェックすることが必要である。諸作業を日次・週次・月次などに整理し、ルーチン化することで、属人性を抑制することができる。

8. HotデータとColdデータ

奈良文化財研究所では膨大なデータを長期保管するために、データ自体をHotデータとColdデータ

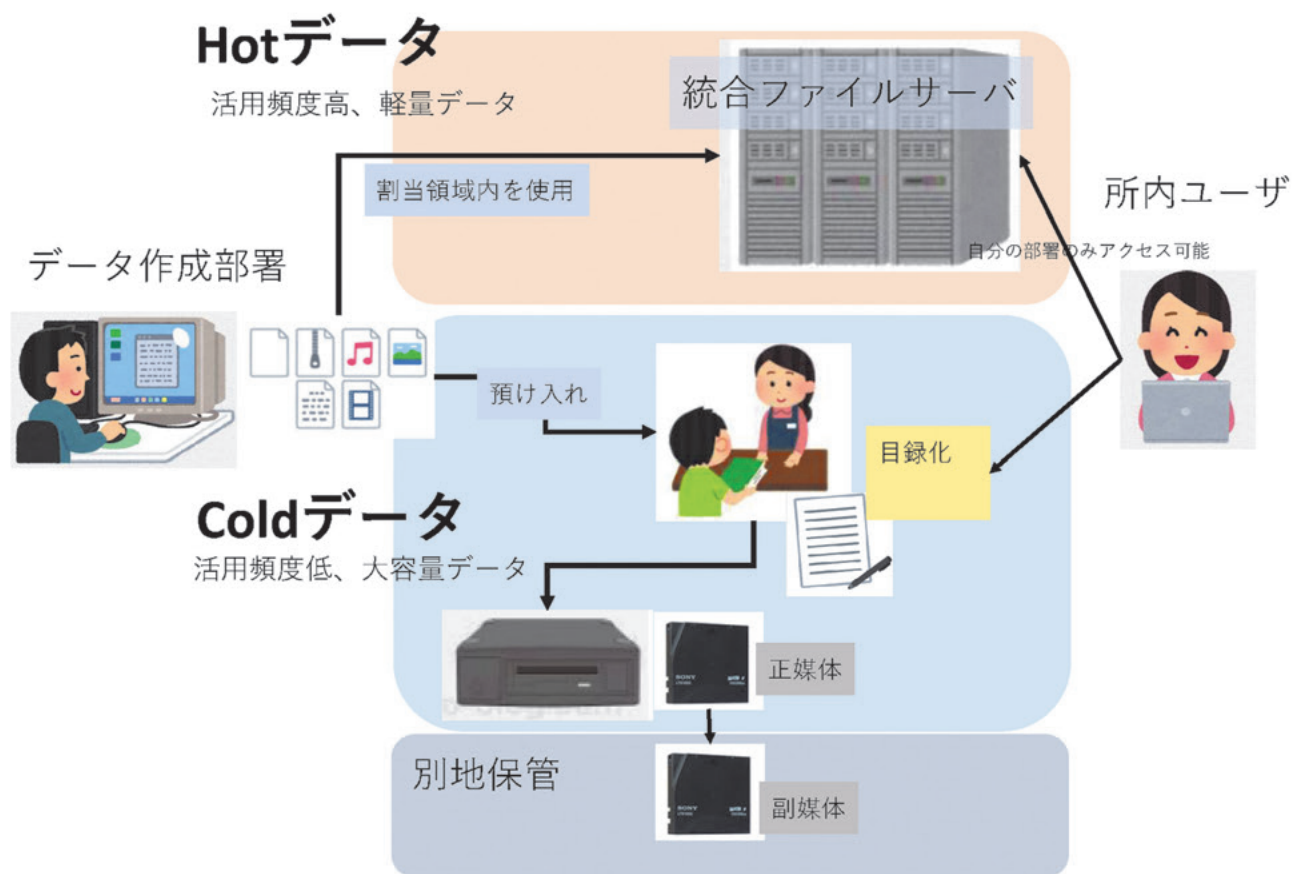


図6 奈良文化財研究所での運用例

に分類した（図6）。Hotデータは活用頻度が高いかつ軽量データである。活用頻度が高いため、オンラインのファイルサーバに格納することで、利便性が向上する。しかしファイルサーバ容量は限りがあるため、3次元データのような大容量のファイルを格納するのは困難である。そのため、ある程度活用頻度が低下している大容量ファイルは、Coldデータとして保管媒体に格納することで、データ保管を図った。媒体は内容が同一のものを正副2つ作成することで、どちらかが何らかの理由で消失した場合でも、一方は残るため、確実に保管できる。

9. おわりに

ここで述べた課題は、将来的に技術進化や新製品の登場で解決できるかもしれない。技術的課題は技術で解決できるだろう。しかし、実際にデータ長期保管を運用するのは人間である。運用体制を整備し、各組織にあった方式・技術・製品をローカライズするのは所属職員の仕事である。本稿によって各機関がデータ長期保管を実現させ、調査研究の基盤

となることへの一助となれば幸いである。

【註】

DataONE Lesson 1: Data Management (https://www.dataone.org/sites/all/documents/education-modules/handouts/L01_DataManagement_Handout.pdf) 2019年12月5日確認

文化庁 2017『埋蔵文化財保護行政におけるデジタル技術の導入について 1』（報告）埋蔵文化財発掘調査体制等の整備充実に関する調査研究委員会

Archaeology Data Service / Digital Antiquity「Guides to Good Practice」(<https://guides.archaeologydata.service.ac.uk/g2gpwiki/>) 2019年12月2日参照

BUFFALO「ご存知ですか？ハードディスクは壊れます」(<https://www.buffalo.jp/contents/topics/knowledge/safety/hdd.html>) 2019年12月2日確認

Brian Beach 2013「How long do disk drives last?」(<https://www.backblaze.com/blog/how-long-do-disk-drives-last/>) 2019年12月2日確認