

デジタル資料の長期保存に関する 基礎知識 1

—記録媒体の特性の違いや課題を理解する—

電子情報部電子情報企画課
次世代システム開発研究室

本研修の目的

- デジタル資料を長期的に保存するにあたって考慮すべき事項や課題などを理解すること
 - 特に、紙媒体の資料との特性の違いについて説明できるようになること

目次

1. デジタル資料とは何か
2. デジタル資料の長期保存の課題
 - ① 記録媒体の物理的な破損
 - ② データの部分的な破損
 - ③ 記録媒体の寿命
 - ④ データの改ざんや上書き
 - ⑤ 読取機器の維持
 - ⑥ ソフトウェア等の利用環境の維持
 - ⑦ その他

3. デジタル資料の長期保存のための対策

ビット列の保存のための対策

適切な保存媒体の選択/保存媒体の運用方法/バックアップ/ハッシュ
(チェックサム) /マイグレーション

コンテンツの保存のための対策

マイグレーション/エミュレーション・仮想化

その他の対策

データ作成時の工夫/管理のための工夫 (メタデータ)

有用な情報源の紹介

1. デジタル資料とは何か

デジタル資料の定義

- 本研修では、デジタルデータや、それを記録した媒体を「デジタル資料」と呼ぶこととする

デジタル資料の例	デジタル資料でないものの例
<ul style="list-style-type: none">• 光ディスク（CD、DVD等）• 画像ファイル、文書ファイル• ウェブページ• 図書、レコード等アナログ媒体をデジタル化して作成したデータ	<ul style="list-style-type: none">• 図書、雑誌、地図• マイクロフィルム• 彫刻、標本資料• レコード、カセットテープ• VHS、LD

デジタルデータとは

□ 情報を 0 (ゼロ) と 1 (イチ) の並び (ビット列) で表現したもの

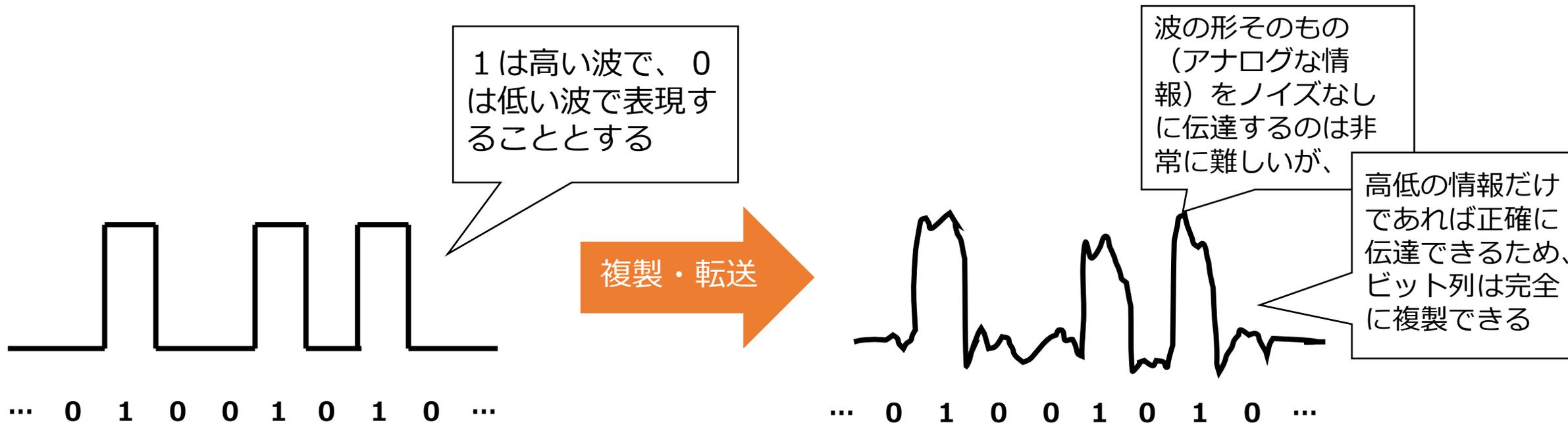
```
01010000 00011000 01001001 11011101
10100010 01000001 10000110 01000100
01010001 00110100 01000100 11000010
10001010 01010110 10010001 11001001
01010101 00100110 01000000 01000110
.....
```

- 中身が何であるか (テキスト/画像/音声/映像/...) に関わらず、すべて一律にビット列で表現する
- デジタルデータの「容量」とはこの長さのことを指す

※容量の単位として有名な「ビット (bit)」とはこの桁数のこと
※以後、特に断らない限り「データ」という言葉はデジタルデータに限定して用いる

デジタルデータの特徴

- デジタルデータは、アナログデータと異なり、もとの情報と全く同一の情報を簡単に複製、伝達できる



デジタルデータの特徴（つづき）

- ビット列がどのような規則（フォーマット）で並んでいるかを知らなければ、人間が理解できるテキスト、音声、画像等に変換することができない

```
01010000 00011000 01001001
10100010 01000001 10000110
01010001 00110100 01000100
10001010 01010110 10010001
01010101 00100110 01000000
.....
```

例えば画像ファイルの場合……

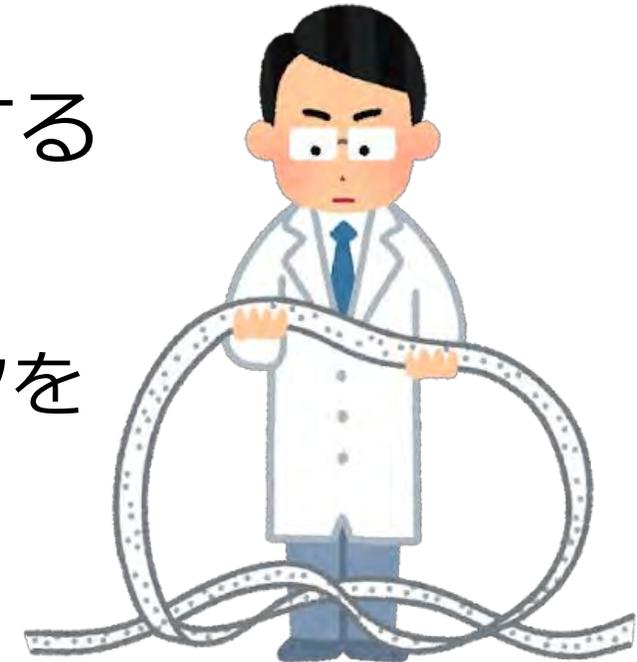
ファイルサイズ、画像の縦横のピクセル数、解像度、(0, 0)ピクセルの色、(0, 1)ピクセルの色、…

といった情報を決まった規則（フォーマット）で並べている。

この規則に沿ってビット列を読み解き、画面に色を表示することで、元の画像が再現できる

デジタルデータの記録

- 0-1の2状態を表現できれば、原理的にはあらゆる媒体にデータの記録が可能
 - 過去には紙テープ上に穴の有無で表現したこともあった
- 現在はより高密度で効率的な媒体（CD,DVD,HDD,SSD等）にデータを記録するのが普通
 - 人間の目では読み取れない
 - → 専用の機器（ドライブ）等がなければデータを読み出せない



デジタル資料の分類

□デジタル資料の長期保存を考える際、さらに次のように分類するとわかりやすい※

- デジタル化資料（自機関作成デジタル資料）
- パッケージ系電子出版物
- ネットワーク系電子出版物

※国立国会図書館の資料所蔵状況等を参考に今回作成したものであり、一般的な分類ではない。主に、図書館、博物館・美術館、文書館等で取り扱う資料を想定している

デジタル化資料

□ 図書・雑誌等の紙資料やレコード等のアナログ媒体をデジタル化して作成したデータ

□ 各機関で作成する（外部委託含む）ので、データの仕様をある程度指定可能

- ・ファイルフォーマット（解像度等も含む）、容量、保存媒体…



パッケージ系電子出版物

- 記録媒体にデータを記録した状態で出版されたもの
 - 音楽、映像、ソフトウェア、図書の付録、ゲームソフト等
 - デジタルデータの記録媒体としては、CD, DVD, Blu-ray, MOディスク, MD, SD, フロッピーディスク (FD), USBメモリ, ROMカセット等
- コピープロテクトがかかっている等、媒体から正しくデータを複製することが難しい場合がある



ネットワーク系電子出版物

□ インターネット等、ネットワークを通じて提供されるデジタル資料のこと

- 電子書籍
- 電子ジャーナル
- インターネット上の情報
 - ウェブサイト、ブログ、SNS…



国立国会図書館では、国立国会図書館インターネット資料収集保存事業（WARP）により、2002年より、公的機関を中心にウェブサイトを収集・保存している

デジタル資料の長期保存とは何か

□ これらデジタル資料の消失・改変を防ぎ、長期にわたって利用可能に保つこと

□ 長期とは

- 30年というのが1つの目安

- 「長期とは、デジタルデータが必要とされる限りできるだけ長い期間を意味する。その具体的な長さは組織の使命によって異なるが、最低でも30年以上を想定する。」 [1]

[1] デジタルアーカイブ推進委員会及び実務者検討委員会 「デジタルアーカイブのための長期保存ガイドライン (2020年版)

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/guideline2020.pdf

2. デジタル資料の長期保存の 課題

デジタル資料の長期保存における主要な課題

- ① 記録媒体の物理的な破損
- ② データの部分的な破損
- ③ 記録媒体の寿命
- ④ データの改ざんや上書き
- ⑤ 読取機器の維持
- ⑥ ソフトウェア等の利用環境の維持
- ⑦ その他



①記録媒体の物理的な破損

□ デジタル資料の記録媒体は紙資料と比べ非常に高密度・高集積

- 例えばCDの場合…
 - 直径12cmの円盤に700MB程度（半角で7億文字程度）の情報を記録
 - 1倍速で1分間に数百回転。（現在は数十倍速で読み取るのが普通）
- 小さな領域の物理的な破損が大きなデータの破損につながる
- 軽微な歪み等でも正常に動作しなくなる
- その他、磁気や静電気の影響など、紙資料では問題にならない破損要因について、記録媒体ごとに考慮が必要

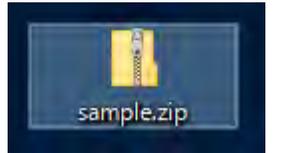
②データの部分的な破損

- データの部分的な破損が全体に影響を与えうる
 - データ（ファイル）は、厳密な規則に沿って解釈されるビット列
 - 部分的な破損でもその規則に従わなくなってしまう、通常の方法では正常に中身を再生できなくなることがある

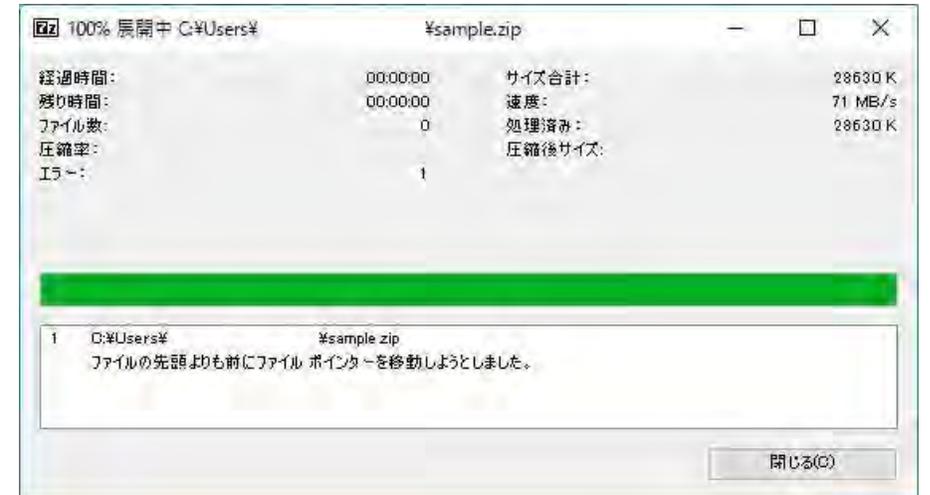


Zipファイルから1文字分データを削除すると…

- 適当なzipファイルを作成し、メモ帳などで1文字分データを消す。（数バイトのデータの欠損）
- 展開しようとするとき……
 - エラーになることがある



※展開時の挙動は元のzipファイル、データを削除した位置、展開に使用したソフトウェアによって異なり、ここで示したのはあくまで一例である



③記録媒体の寿命

- 記録媒体にも寿命が存在する
- デジタルの記録媒体が誕生してからあまり時間がたつておらず、現時点では確たることは言えない
 - CD : 1982年～ DVD : 1996年～ Blu-ray : 2003年～
- ただし、紙媒体と比べ短いということは確実
- 推定寿命と呼ばれているものはある

③記録媒体の寿命（つづき）

- 記録媒体の種類（CD/DVD/BD/USBフラッシュメモリ/HDD/FD/…）によっても、寿命が異なる
- 推定寿命は保存対策への優先順位設定のための指標となりうる。そのため、国立国会図書館では推定寿命等の知見をもとにある程度暫定的な評価をしている
- あくまで2022年時点の内容であり、今後関連分野での研究や知見の蓄積によって変化しうる

※次スライドからの記述は、以下の内容に基づく

国立国会図書館 「電子情報の恒久保存メディア及びそれを用いたシステムに関する調査報告書」

2010年 <https://doi.org/10.11501/8977840>



光ディスクの故障要因と寿命

□ 光ディスクの仕組み

- CD、DVD等。盤面に光の反射率が変わるような微小部分（ピット）を形成することで情報を記録するメディア

□ ピットの形成方法

- 読取専用（ROM）：物理的な凹凸で形成
- 追記型（R）：強いレーザー光によって変化する素材で形成
- 書換型（RW）：レーザー光の強さにより違った変化をする素材で形成・書換

□ 劣化

- 物理的な凹凸で形成するピットの方が安定とされている
- RとRWではRWのピットの素材の方がより劣化が早いとされている

光ディスクの故障要因と寿命（つづき）

□ 寿命

- 10年程度という説から100年以上という説まで様々

□ 国立国会図書館の方針

- 2022年現在では「ROMは比較的長く保存可能。Rは30年が最大、RWは可能な限り早期に対策が必要」という認識
 - RWを「早期の対策が必要」としているのは、データの書換を防ぎたいというのも一つの要因（詳細は「④データの改ざんや上書き」を参照）

ハードディスク(HDD)の故障要因と寿命



□ 仕組み

- アルミニウムやガラス等の硬い（hard）ディスク表面に磁気記録層をコートし、ディスクから浮いたヘッドでデータを読み書きする

□ 故障要因・寿命

- 記録媒体と読取機構が一体となっており、読取機構は精密（ディスクとヘッドの隙間はnm単位）。温度、湿度、気圧、振動、突発的な衝撃等によって破損する
- 5年以上同じものを使うのは避けたほうがよいとされている

その他の媒体について

- フラッシュメモリ（USBメモリ等）
 - 寿命は10年程度とされており、長期保存には向かない
 - 書き換えが容易（後述）

- フロッピーディスク（FD）
 - カビや磁気の消失が劣化要因
 - 寿命は5～10年という説も



媒体の取扱い方法

□ 光ディスクの取扱い

- 理想的な保管環境は15℃～25℃、湿度40%～60%の直射日光を避けたホコリの少ない環境
- ディスクの記録面には傷、指紋、汚れ等がつかないように取り扱う
- ディスクは曲げたり落としたりしないようにする
- レーベル面に文字を記入する場合は先の柔らかい油性フェルトペンなどを用いる
- ディスクに保護シートやラベルなどを貼らない

※詳細は以下を参照

一般社団法人日本記録メディア工業会 「光ディスク取扱い上の注意」 平成23年

https://www.jiima.or.jp/pdf/disk_20110810.pdf

媒体の取扱い方法

□ 外付けHDDの取扱い

- 倒す、落とすなどの物理的な衝撃を与えない
- 給排気口の周りに広い空間を取り、冷却を妨げないようにする
- USBケーブル、ACアダプターなどは純正品（製品に付属していたもの）を用いる
- 正しい手順で取り外す
 - データの読み書き、削除中は特に注意
- なるべくUSBハブ経由で接続しない
 - ハブ経由の方が通信エラーが多い

参考

- BUFFALO | 故障リスクを減らし、データを守る、正しい外付けハードディスクの扱い方
https://www.buffalo.jp/topics/trouble/detail/recovery_0016.html
- 高田祐一「文化財デジタルデータ長期保管の実務」 奈良文化財研究所研究報告 第24冊 2020年 pp.49-54 <http://hdl.handle.net/11177/7253>
- 文化財写真保存ガイドライン検討グループ 「文化財写真の保存に関するガイドライン～デジタル画像保存の実情と課題～」 日本写真学会誌 75巻4号 2012年 pp. 302-322
<https://doi.org/10.11454/photogrst.75.302>

④データの改ざんや上書き

□ ハードディスク(HDD)やフラッシュメモリ、書換型光ディスク(RW)などは、容易にデータを改変できる

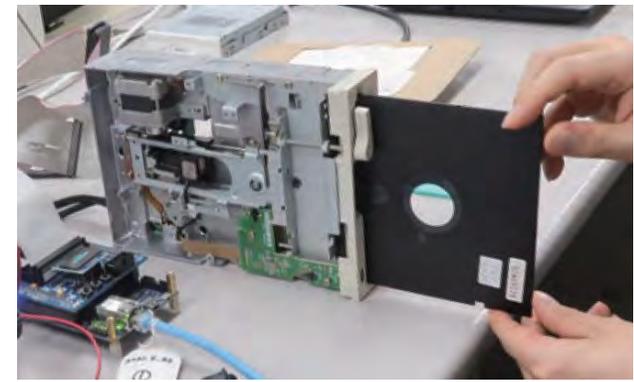
- ウイルス対策ソフトが誤動作したり、OSが管理情報を書き込んだりと、意図せず上書きしてしまうケースもある
- 改ざんや上書きの検知自体も大きな課題。受入時の状態を記録できていなければ改ざんに気づきようがない
- 書込みを防止する専用の装置なども市販されている



⑤ 読取機器の維持

- 記録媒体は読取機器が使えなければただの物と変わらない
- 読取機器も機械なので寿命があり、生産も終了する
 - 記録媒体の評価にあたっては、その媒体の寿命だけでなく、読取機器の継続的な入手可能性も考慮に入れる必要がある
 - 特定のメーカーが性能（推定寿命、容量）の高いメディアを開発したとしても、その読取機器が普及していなければリスク要因となる

⑤ 読取機器の維持（つづき）



□ 例えば…（国立国会図書館所蔵資料から）

- 5.25インチFD
 - 5.25インチFDドライブは、現在は中古でのみ入手可能。入手できても、USB等の規格が一般的でない時代の機器なので、特殊な機器を用いないと現在のPCと接続できない
- ゲームのROMカセット等
 - そもそも中身をデータとして読み取る機器が市場に流通しない

※現在は広く普及しているCD等のメディアも、ドライブを組み込んだPCの生産が減っていたり、音楽CDの流通量が減っていたりと、注視が必要

⑥ソフトウェア等の利用環境の維持

データ（ビット列）を保存し、正確に読み出すことに成功したとしても、利用できなければ意味がない

□ 利用環境の維持

- 読取機器だけでなく、対応するソフトウェアも必要
 - 対応ソフトウェアがなければデータはただのビット列
 - しかし、ソフトウェアもOSが新しくなるにつれ動作不良等が起こるようになり、ソフトウェアが依存するソフトウェアの維持が必要となる……

⑥ソフトウェア等の利用環境の維持 (つづき)

利用環境の維持を別側面から見ると…

□ ファイルフォーマットの旧式化・陳腐化

- そのファイルのフォーマットを正常にサポートできるソフトウェアがなくなっていくこと
 - ファイルフォーマットの例：txt, pdf, docx, xlsx等
- メジャーなファイルフォーマットであっても、マイナーな仕様についてはサポートや使い勝手が変わったりする
- 3Dデータなどは特に利用環境が失われやすい

国立国会図書館所蔵資料の例

- CD-ROM×3枚 + FD×5枚のMS-DOS用学会論文集
(1990年代)
 - FDのうち1枚が検索プログラム、残り4枚が書誌データや検索インデックス
 - CD-ROM内には独自の画像形式で保存された論文データ
 - 画像だけを抽出することは困難であり、論文データの利用には同梱の専用アプリケーションの維持が必要

比較的最近の資料でも……

□ USBメモリ資料。中にindex.htmlがありWebブラウザでの利用を想定（2010年代後半の資料）

- FlashとHTML5が選択できるが、Flashは2022年時点ではサポートが切れ、通常のWebブラウザからはアンインストールされている
- HTML5の方も、index.htmlをダブルクリックするだけでは現在の国立国会図書館の提供環境では利用不可
- 幸い中に電子ブックリーダーを起動できるEXEが入っているのでそこから利用可能

⑦ その他

□ 増大するデータ量への対応

- 保存すべきデータの量が、紙では考えられない規模で増える

□ 法的課題

- コピーガードへの対応に関する法的な懸念
- エミュレーションなど保存対策のための措置と、アプリケーションの利用許諾やライセンス条件との衝突

□ 組織体制

- デジタル資料の長期保存は組織全体に関わる問題
- 長期保存に必要なスキルセットの網羅的な涵養が難しい一方で、専門知識を持った人員の確保が必要

3. デジタル資料の長期保存の ための対策

課題の分類

課題再掲

- ①記録媒体の物理的な破損
- ②データの部分的な破損
- ③記録媒体の寿命
- ④データの改ざんや上書き
- ⑤読取機器の維持
- ⑥ソフトウェア等の利用環境の維持
- ⑦その他

データのビット
列での保存に関
係

コンテンツとし
ての利用保証

(1) ビット列の保存 (bit level preservation等)

- データを0-1の並びとして欠損なく完全に保存すること
- データの中身が何であるかは問わない（文書、画像、音声、アプリもすべて同じ）ためシステムティックかつ一律に実行可能

(2) コンテンツの保存 (logical preservation等)

- データを人間が利用可能な状態のままとする
- データの中身や想定するユーザ、組織のポリシー（何を残したいのか）に依存した対策が必要

※ビット列の保存のみではデジタル資料の長期保存として十分ではないということが重要なポイント

適切な保存媒体の選択

□ ハードディスク（HDD）

- 最も広く用いられている保存媒体
- HDD単体では耐障害性に難があるので、通常はRAID等の手法で複数組み合わせ、壊れたものを順次交換するという運用をする

□ LTO等の磁気テープ

- ランダムアクセスはHDDより低速だが、運用面含めた長期的なデータ保存のコストはHDDより低いとされているため、データのバックアップ用媒体として広く用いられている

□ 光ディスク

- 1枚で300GB保存できるような規格も出てきている。適切に運用すれば、長期的なデータ保存のコストはHDDより低いとされている

保存媒体の運用方法

- HDD、磁気テープ及び光ディスクを単体で用いるのではなく、適宜組み合わせることもよく行われる
 - アクセス頻度が高いものは読み書きの性能が高い媒体に保存し（ホットストレージ）、頻度が低いものはテープや光ディスクに入れる（コールドストレージ）など

- 機関が保存媒体を自分で持たず、クラウドでデータを保存するケースも
 - その場合は中でどのようにデータを保存しているかは不明だが、（各機関が自前で用意するよりは）容易に複数遠隔地点でのバックアップ等ができる

バックアップ

- オリジナルのデータが破損しても回復できるように、データを別の媒体に複製すること
- 災害等が起こった場合に備え、複製した媒体は地理的に離れた箇所に保管するとなおよい
- オリジナルと複製で別の技術を採用するとよい
 - ・ 同じ媒体を採用して旧式化すると両方使えなくなる
- 3-2-1ルール
 - ・ バックアップの理想的な方法
 - ・ 「データは3重に持ち、記録媒体は2種類に分け、1つは遠隔地で保存する」

ハッシュ (チェックサム)

膨大な数の0-1が間違いなく保存されていることをどのように確認するか？

- 任意長のビット列を短い固定長の値 (ハッシュ値) に変換する
 - 一番単純な例：32バイトごとに切り (不足は0で埋める)、繰り上げは無視して全部足す
- 資料受入時点で計算した値をメタデータ等に記録しておき、再度計算して違う値が出れば、受入時点からビット列が変化していることがわかる

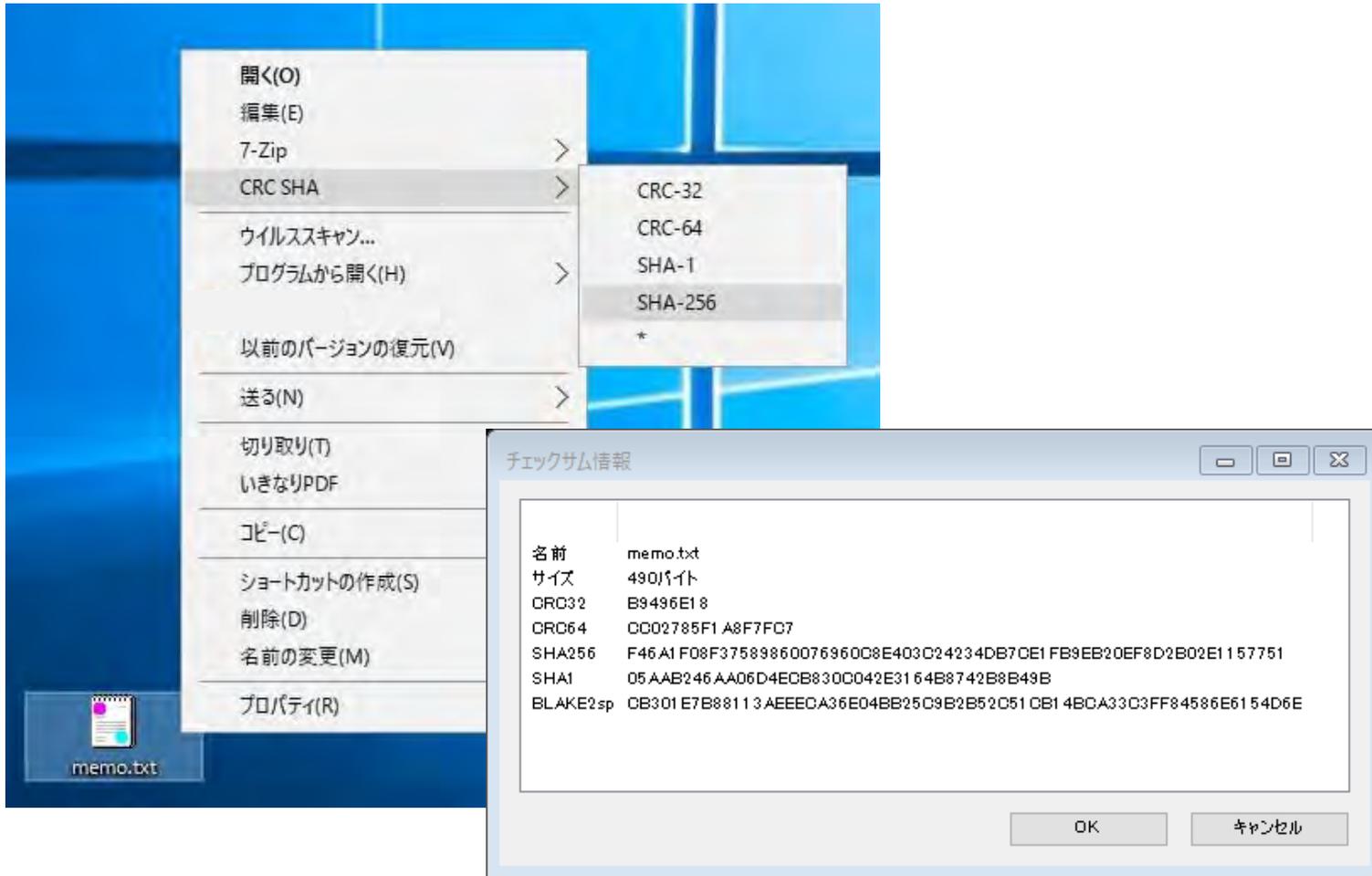
ハッシュ (チェックサム)

- ハッシュ値の計算方法 (アルゴリズム) は様々
- よりランダムで、計算コストの低い計算方法がよいハッシュアルゴリズム
- CRC, MD5, SHAなどのアルゴリズムがよく用いられる

ハッシュ値取得方法の例

7-Zip等の圧縮ソフトにハッシュ値の計算がついていることもあり、右クリックメニュー等から簡単に計算可能

(大容量の圧縮ファイルをインターネット経由で配布する際、転送や正当性の確認のためにハッシュ値とあわせて公開することがよく行われる)



マイグレーション

□ データを旧式化した形式から新しい形式へと移し替えること

□ 次の2つの意味で使われる

① 物理的な媒体の移行（ビット列はそのままに記録媒体を変える）

② ファイルフォーマットの変換

• マイグレーションのイメージ図



同種の新しい記録媒体へ移行



異種の新しい記録媒体へ移行



新しい記録フォーマットへ変換

マイグレーション（つづき）

□国立国会図書館での実施例

- USBメモリ、書換型光ディスク(RW)、MD等の記録媒体のデータを長期保存用光ディスクなど、より長期に保存が可能な媒体への移行を実施
 - USBメモリ及び書換型光ディスク(RW)は媒体の寿命等リスク、MDについては再生環境の旧式化、が主たる懸念

著作権法第31条第1項の図書館等の場合は、権利処理することなく、資料を保存のため又は再生可能とするためにマイグレーションすることは可能（※）

エミュレーション・仮想化

□ エミュレーション・仮想化技術

- 旧式化した環境を現在の環境で再現する技術のこと
 - PCの中に仮想の古いPCを作る、といったイメージ
- 精度には限界がある
- 過去の環境は容量が小さかったり性能が低かったりするので、近年ではWebブラウザ上で動作するエミュレーション環境も存在する
- エミュレーションを行うソフトウェアをエミュレータとよぶ

エミュレーション・仮想化の例

□ 一般的なPCで利用可能なもの

- Hyper-V
 - Windows 10/11 Proに標準で搭載されている仮想化ソフトウェア。パッケージ版のOSを所有していれば古いOSを現在の機器にインストールできる。
- DOSBox
 - MS-DOS環境を現在のOSのPC上で再現するソフトウェア
- その他多数



※エミュレーション・仮想化等の実施にあたっては、該当するソフトウェアのライセンス条項の確認や、サポートが切れたOSを動作させるリスクなど、事前の検討と注意が必要

□ Internet Archive

- Webサイトだけではなく、ソフトウェアやゲーム等を収集
- 一部をWebブラウザ上で動作するエミュレータ上で起動して提供

□ EaaSI (Emulation as a Service Infrastructure)

- イェール大学図書館のデジタル保存サービスチームが運用
- 電子資料に対し、それに対応する環境のエミュレータをオンデマンドで準備する仕組み

データ作成時の工夫

- ファイルフォーマットの旧式化の問題を防ぐためにも、データ作成時にフォーマットの長期的な利用可能性に留意するとよい
 - パッケージ系電子出版物などでは不可能だが、デジタル化資料等では可能

- フォーマットを選択する際の基準は？
 - 基本的には広く利用されていて、かつ、オープンなフォーマットほど長期的な利用が可能なのがよいとされる
(=使いやすいライセンスであり、仕様書等が公開されており、原理的にはゼロから再生ソフトを作れるもの)

データ作成時の工夫

□ フォーマットを選択する際の基準は？（続き）

- 機関が推奨できるフォーマットの情報を公開していることも
- 米国議会図書館：Recommended Formats Statement
<https://www.loc.gov/preservation/resources/rfs/TOC.html>
- 国立国会図書館：国立国会図書館が採用するファイルフォーマット
<https://www.ndl.go.jp/jp/preservation/dlib/formats.html>

国立国会図書館が採用するファイルフォーマット

- 資料デジタル化において保存用データ作製の際に採用しているファイルフォーマット

□ 紙資料・マイクロフィルム

- JPEG 2000

- 可逆・非可逆の圧縮が選択可能だが、可逆の方を採用

□ 録音資料

- WAVE

□ 映像資料

- MP4

- 映像コーデックはMPEG-4 AVC/H.264

ファイルフォーマットの特定方法

- 保存対象となるデータに対し、ファイルフォーマットを特定してメタデータとして保存しておくことが望ましい
 - ファイルフォーマットの特定自体も正確には難しく、大きな技術的課題となっている

□ 拡張子による特定

- ファイル名の最後の「.」以降の箇所での特定
 - .pdf, .docx, .csv, .txt, .json, .jpg, .mp3...
- 単純で適用範囲が広い方法だが、人為的に書き換え可能であり、違うフォーマットに同じ拡張子が使われる場合もある

□ ビット列のパターンによる特定

- あるファイルフォーマットに特有のビット列のパターン（Signature, magic number）を利用して特定
- 英国国立公文書館が開発しているDROIDやUnix/Linuxのfileコマンド、Apache Tikaなどのツールが有名

ファイルフォーマットの特定方法（つづき）

□ フォーマットの表現方法

- 拡張子
- フォーマットレジストリ
 - ファイルフォーマットの情報を収集し、IDを振ったデータベース
 - PRONOM (The National Archives (英))
 - <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>
 - Wikidataをそのように利用しようという研究も
- MIME Type
 - image/jpeg等
 - Webブラウザが取得したファイルに対する動作を決定する際などに使用

PRONOMの表示例

PUIDはPRONOMが
ファイルフォーマットに対し
発行しているID

<https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailedReport&id=667>

The National Archives

Search our website

About us | Education | Records | Information management | Archives sector

You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > [PRONOM](#) > [Search by format](#) > [Details: Summary](#)

The technical registry
PRONOM

Welcome : About Add an entry
Search ? Help Information resources

? Help : detailed report on file format

Details: File format summary

Simple search | File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles | Migration Pathways

Details for: **JPEG File Interchange Format 1.00** Save as... XML | CSV Print

Go to: [Summary](#) | [Documentation](#) > | [Signatures](#) > | [Compression](#) > | [Character encoding](#) > | [Rights](#) > | [Reference files](#) > | [Properties](#) >

Summary

Name	JPEG File Interchange Format
Version	1.00
Other names	JFIF (1.00)
Identifiers	PUID: fmt/42 MIME: image/jpeg Apple Uniform Type Identifier: public.jpeg
Family	
Classification	Image (Raster)
Disclosure	Full
Description	The JPEG File Interchange Format (JFIF) is a file format for storing JPEG-compressed raster images. It was developed by the Independent JPEG Group and C-Cube Microsystems, in the absence of any such format being defined in the JPEG standard, and rapidly became a de facto standard; this is what is commonly referred to as the JPEG file format. A JFIF file comprises a JPEG data stream together with a JFIF marker. It begins with a Start of Image (SOI) marker, immediately followed by a JFIF Application (APP0). This is followed by the JPEG image data, which is terminated by an End of Image (EOI) marker. JFIF supports up to 24-bit colour and uses lossy compression (based on the Discrete Cosine Transform algorithm). Other types of compression are available through JPEG extensions, including progressive image buildup, arithmetic encoding, variable quantization, selective refinement, image tiling, and lossless compression, but these may not be supported by all JFIF readers and

管理のための工夫（メタデータ）

□ 長期保存用のメタデータ

- ハッシュ値、利用環境、ファイルフォーマットの情報、権利情報、保存対策の履歴などを管理しやすく、機械的に判読しやすい形で記述するのに適したスキーマが必要

□ メタデータ標準（いずれも現在は米国議会図書館が管理）

- PREMIS (Preservation Metadata Implementation Strategies)
 - 長期保存活動に関する基本的な語彙等を規定
- METS (Metadata Encoding and Transmission Standard)
 - 主にXMLへのメタデータのエンコードとリポジトリ間での交換に焦点を当てている

有用な情報源の紹介（国立国会図書館）

□調査報告書

- 2016年度からほぼ年次で調査を実施し、報告書としてまとめてHPで公開している
- 国立国会図書館 | 電子情報の長期利用保証に関する調査研究
 - <https://www.ndl.go.jp/jp/preservation/dlib/index.html>

国立国会図書館の調査報告書

2021年度	<ul style="list-style-type: none">資料デジタル化及びデジタルデータ長期保存に係る国内実態調査作業フロッピーディスク資料の長期保存対策に係る調査光ディスクの状態検査の手法に係る調査
2020年度	<ul style="list-style-type: none">電子情報の長期保存におけるエミュレーション技術の利用
2019年度	<ul style="list-style-type: none">電子情報の長期保存対策に係る調査（文献調査＋海外機関の取り組みについてアンケート・インタビュー調査）
2019年度～ 2018年度	<ul style="list-style-type: none">光ディスクを対象とした媒体劣化状況分析
2017年度	<ul style="list-style-type: none">パッケージ系電子出版物を対象とした長期保存対策（USB, MO, CD/DVD等のマイグレーション手法の調査と保存メタデータの検討）
2016年度	<ul style="list-style-type: none">恒久的保存基盤の構築に向けた技術

※記載は調査の実施時期であり、報告書の公開時期とは異なる

有用な情報源の紹介（日本国内）

- デジタルリソースの長期保存に関する概観
 - http://www-archive.cseas.kyoto-u.ac.jp/ipres2017.jp/wp-content/uploads/jtutorial_sugimoto.pdf
 - 杉本重雄氏（現筑波大学名誉教授）のiPRES2017での発表
- デジタルアーカイブ推進委員会及び実務者検討委員会 「デジタルアーカイブのための長期保存ガイドライン（2020年版）」
 - https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/guideline2020.pdf
- 電子情報保存 | カレントアウェアネス
 - <https://current.ndl.go.jp/taxonomy/term/131>

有用な情報源の紹介（海外）

□ Digital Preservation Handbook

- デジタル資料の長期保存について包括的に記載されたハンドブック。英国のデジタル保存連合(Digital Preservation Coalition)が管理
 - <https://www.dpconline.org/handbook>

□ Open Preservation Foundation開発のツール

- ファイルフォーマットの判定、バリデーションのためのツールを提供
 - <https://openpreservation.org/>

有用な情報源の紹介（最新の動向）

□ DPC Technology Watch Publications

- 英国のデジタル保存連合が不定期に刊行しているレポート。録音映像, 3Dデータ, Eメール等テーマを絞って紹介している
 - <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>

□ iPRES (International Conference on Digital Preservation)

- デジタル資料の長期保存をテーマとする国際会議。最新の研究等が把握できる
 - <https://ipres-conference.org/>

本研修のまとめ

□ デジタル資料はアナログ資料とは異なった特徴を持つ

□ デジタル資料の長期保存には特有の課題がある

- 記録媒体・データの破損の影響
- 記録媒体の寿命
- データの改ざんや上書きの影響
- 読取機器・ソフトウェア等利用環境の維持の必要性

□ ビット列の保存とコンテンツの保存

ビット列の保存

- 適切な媒体の選択、バックアップ、ハッシュ、マイグレーション等

コンテンツの保存

- マイグレーション、エミュレーション・仮想化

その他

- 適切なメタデータの付与、他機関との連携等