

第6回東京大学学術資産アーカイブ化推進室主催セミナー
「デジタル資料の保存管理の実際」
令和5年2月6日



国立国会図書館における取組

— デジタル資料の長期保存とは？ —

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長 徳原 直子



目次

1. デジタル資料とは？—国立国会図書館のデジタル資料
2. 国内機関のデジタル資料の収集・保存状況と課題
3. デジタル資料の長期保存のための取組とは？
4. 国立国会図書館における最近の取組と課題

1. デジタル資料とは？

国立国会図書館のデジタル資料

- **パッケージ系電子出版物**

フロッピーディスク、CD、DVD、Blu-ray Disc 等

- **インターネット資料・オンライン資料**

ウェブサイト、電子書籍・電子雑誌、電子版博士論文 等

- **デジタル化資料等**

当館が保存のために作成したデジタル化成果物（画像、音声、動画等）等

【長期保存の対象・保存期間】

国内で刊行又は発信された資料は**原則、永続的に原資料保存**

保存のために作成された原資料を代替する成果物（デジタル化資料）も、原資料の劣化に備えて長期に保存

国立国会図書館のデジタル資料

・パッケージ系電子出版物

フロッピーディスク、CD、DVD、Blu-ray Disc 等

磁気ディスク1.2万点

光ディスク87万点

・インターネット資料・オンライン資料

ウェブサイト、電子書籍・電子雑誌、電子版博士論文 等

23万件 (2.6PB)

165万点

・デジタル化資料等

保存目的で作成したデジタル化成果物（画像、音声、映像等） 等

画像317万点（3億画像枚）、音声5万点、映像700点

【長期保存の対象・保存期間】

国内で刊行又は発信された資料は**原則、永続的に原資料保存**

保存のために作成された原資料を代替する成果物（デジタル化資料）も、原資料の劣化に備えて長期に保存

注) 数字は概数。パッケージ系電子出版物の点数は2020年7月時点、他は2022年12月時点

誰が保存・管理？

- 電子情報部が計画の調整及び進捗管理を行うが、**部局横断的・全館的に取り組む**課題としている

「デジタル資料の長期保存は、当館として部局横断的・全館的に取り組むべき課題であり、電子情報部門、収集・書誌・資料保存部門及び利用提供部門が連携して施策を実施する。本計画に係る調整及び進捗管理は、電子情報部が行う。」

- 調査研究は、電子情報部（次世代システム開発研究室）と関西館（電子図書館課）が担当

※ 「国立国会図書館デジタル資料長期保存基本計画 2021-2025」より

https://www.ndl.go.jp/jp/preservation/dlib/pdf/NDLdigitalpreservation_basicplan2021-2025.pdf

デジタル資料の長期保存に関する 国立国会図書館の調査研究

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

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/research.html> 参照

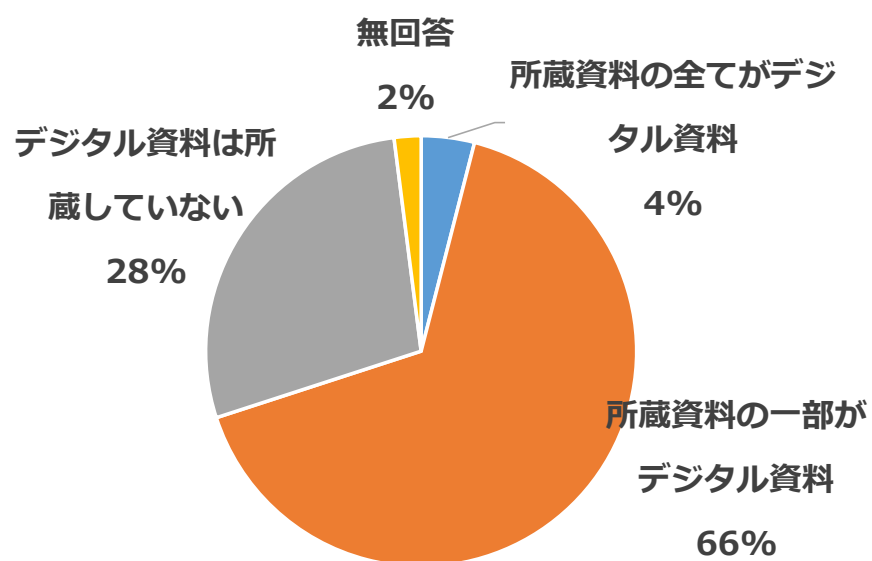
2. 国内機関のデジタル資料の 収集・保存状況と課題

国内アーカイブ機関の実態調査

国立国会図書館による、2021年度「デジタル資料の長期保存に関する国内機関実態調査」(図書館、博物館・美術館、文書館・資料館、大学・研究機関、地方自治体等 **2,921機関**が回答)

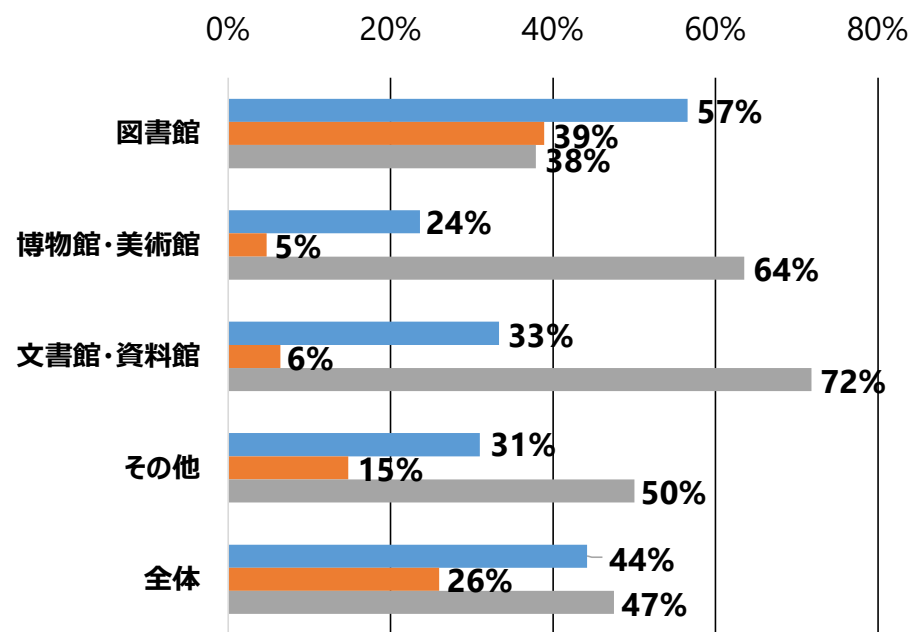
<https://www.ndl.go.jp/jp/preservation/dlib/research.html>

① デジタル資料の所蔵状況



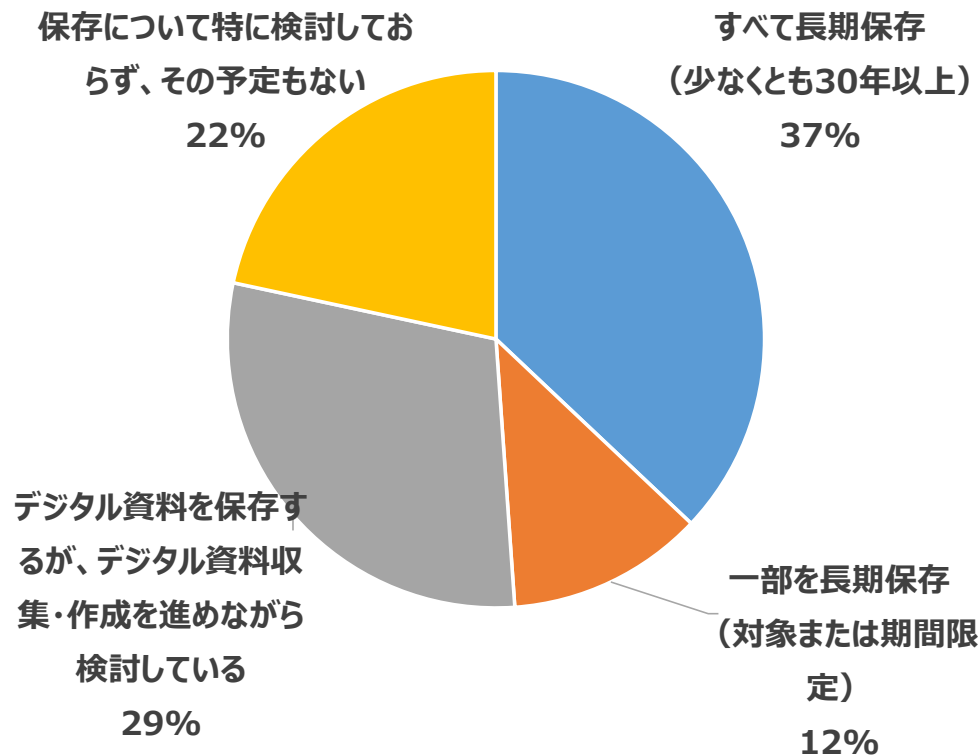
n=2921 (有効回答数)

② デジタル資料の収集方法

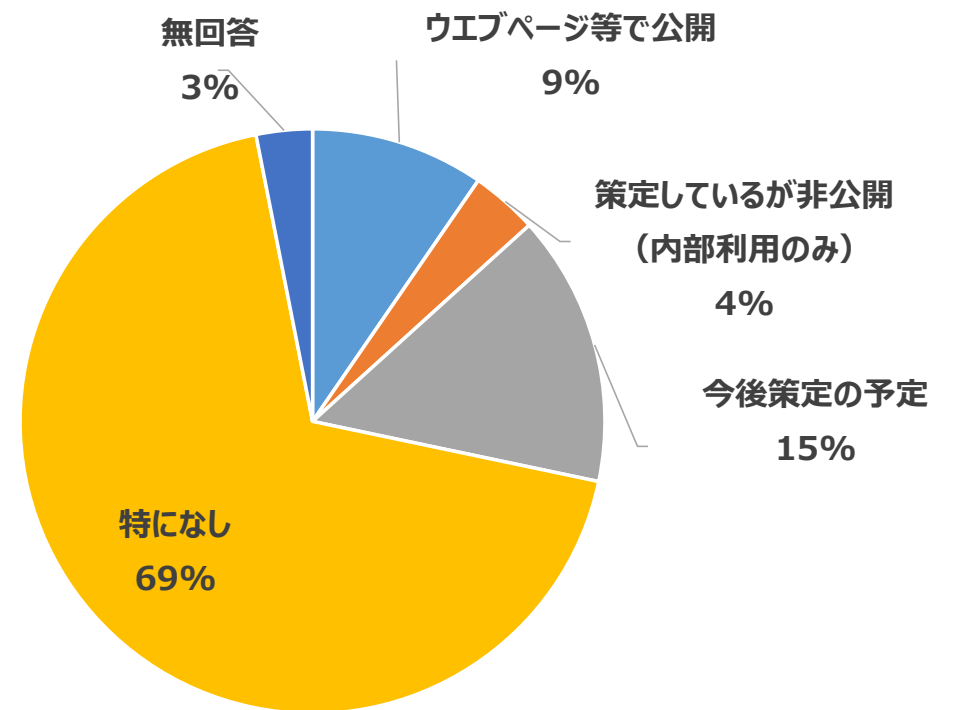


- パッケージ系電子出版物として受入
- ネットワーク系電子資料として受入
- 自機関での資料のデジタル化/ポーンデジタル資料として作成

③ デジタル資料の保存想定

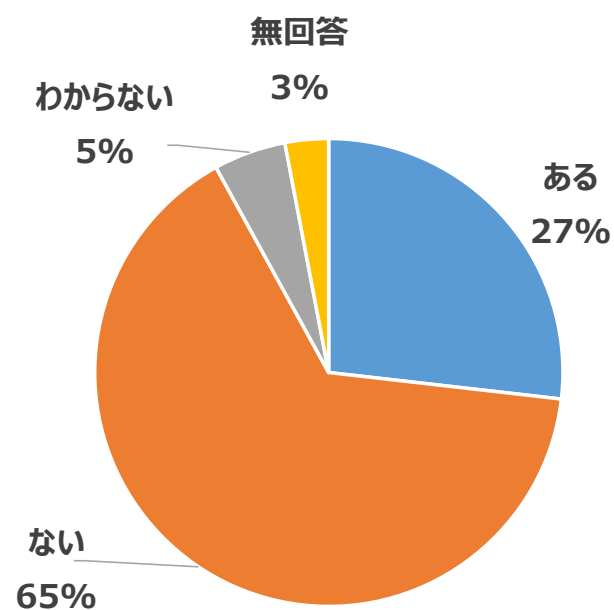


④ デジタル資料の保存方針・計画

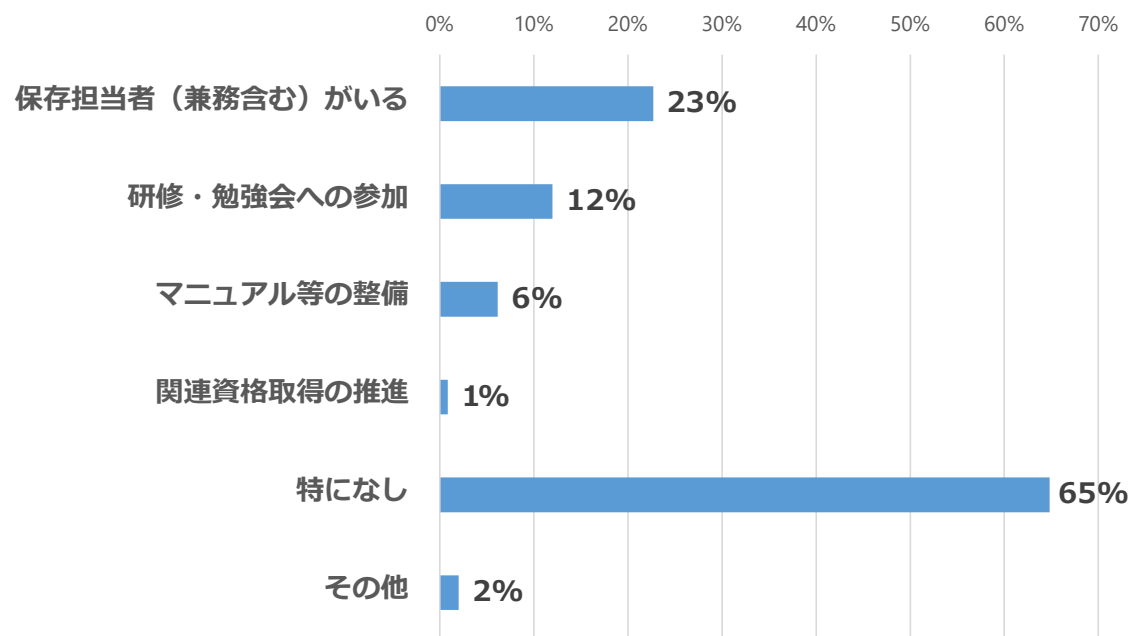


n=2041 (デジタル資料を所蔵すると回答した機関)

⑤ デジタル資料保存の予算措置

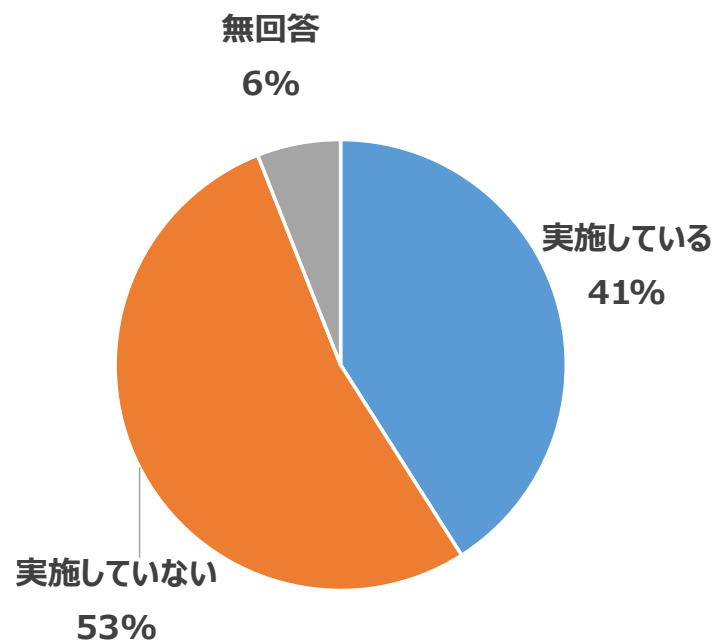


⑥ デジタル資料保存の実施体制・人材育成

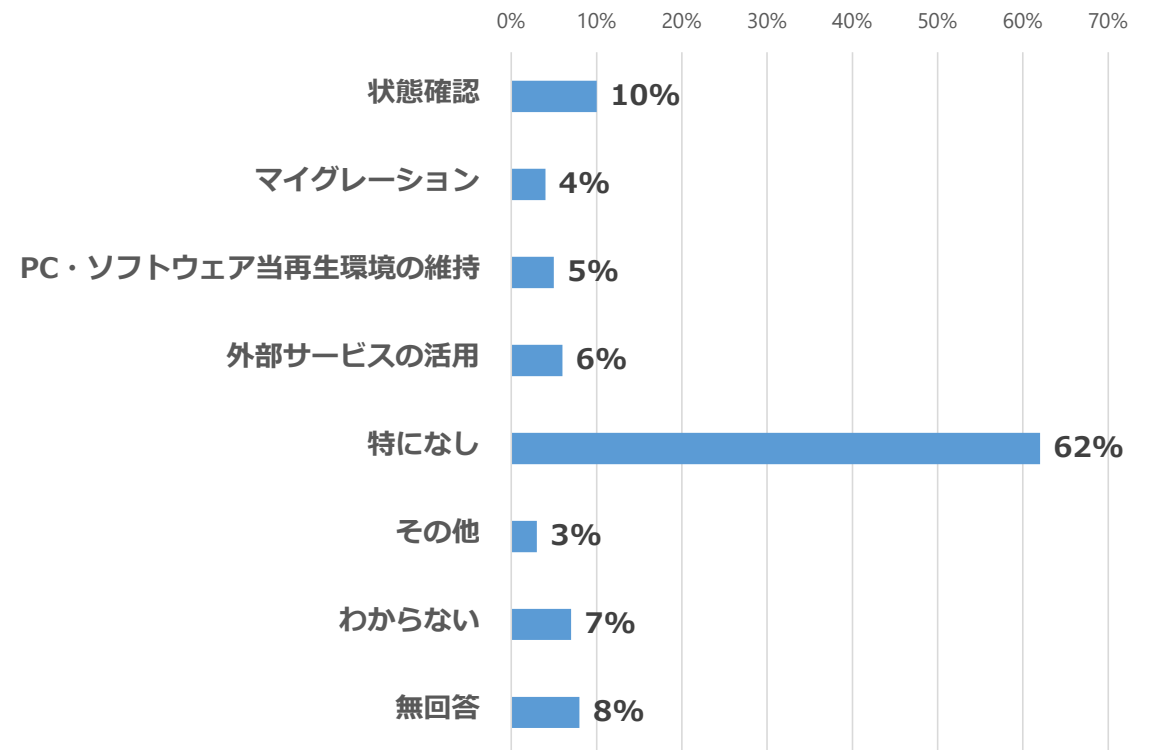


n=2041（デジタル資料を所蔵すると回答した機関）

⑦ バックアップの実施状況

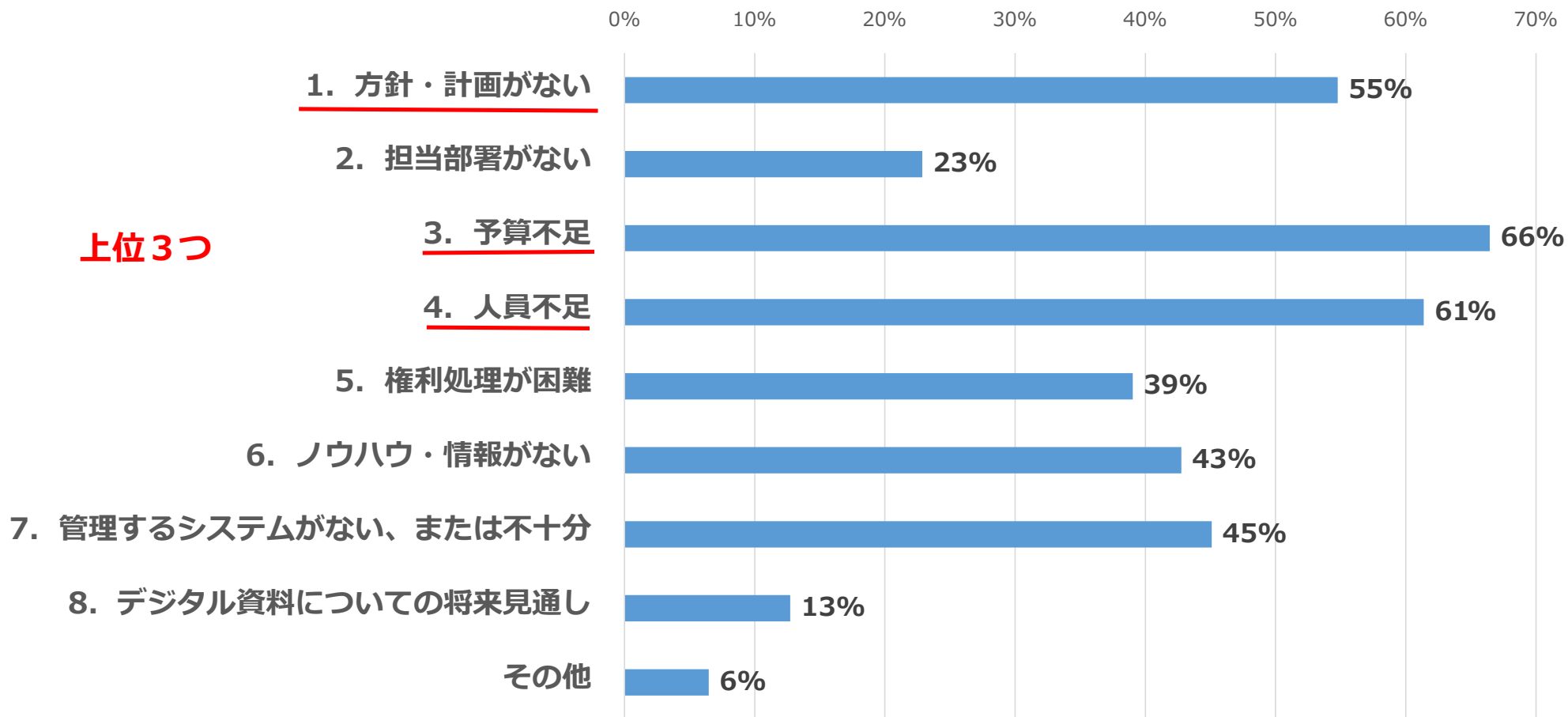


⑧ 保存対策の取組



n=2041 (デジタル資料を所蔵すると回答した機関)

⑨ デジタル資料の収集・保存・提供の課題



n=2921 (有効回答数)

3. デジタル資料の長期保存のための取組とは？

デジタル資料の長期保存とは？

「利用目的に応じた方法で、長期にわたって利用可能性を保つこと」

メタデータの整備も含め、デジタルデータが適切に管理されており、必要なときにデジタルコンテンツにアクセスでき、そのデジタルコンテンツが利用可能な状態をいう。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。



デジタル資料の長期保存とは？



デジタル媒体の特性

- 保存メディアによっては数年で利用不能、データそのものが消失することもある
- 紙等の物理的資料と比べてデータの劣化や消失に気付きにくい
- 再生するための機器（PC、OS等）の変化が激しい
- 再生にはそれに対応したアプリケーション等が必要になることから利用保証の維持が非常に難しい

定期的かつ適切なメンテナンスが必要

「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

デジタルデータの長期保存に必要な
検討事項を提示

(前2つのスライドはこれに基づき作成)

◆基本的考え方◆

保存と利用の両面に対応できるよう、
安定的に管理・運用できる体制が必要

ジャパンサーチを運営する「デジタルアーカイブ
ジャパン推進委員会実務者検討委員会」作成

※同委員会で「[デジタルアーカイブの構築・共有・活用ガイドライン](#)」と一緒に改定作業中

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline.pdf

- 1 デジタルデータ管理の基本的考え方
- 2 デジタルデータの保存方針・計画策定
- 3 デジタルデータ管理
- 4 デジタルデータ保存
- 5 安定稼働できるシステム
- 6 メタデータ管理
- 7 アクセス保証の仕組み
- 8 デジタルデータ移行性の担保
- 9 人材育成

「デジタルアーカイブアセスメントツール」

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/files/assessment_tool_kaitei.xlsx

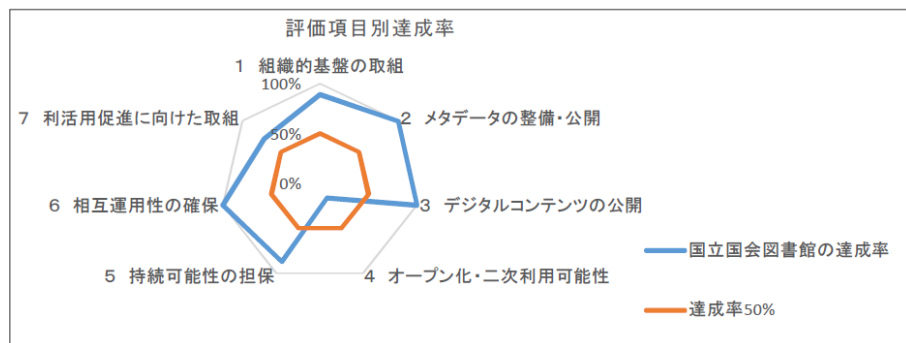
デジタルアーカイブの質や取組の達成状況を把握するための自己点検ツール

デジタルデータの保存・管理の状況もチェックできる

デジタルアーカイブジャパン推進委員会実務者検討委員会作成

※同委員会で改定作業中

- 1 **組織的基盤の取組** (計画、予算、人材等)
- 2 メタデータの整備・公開
- 3 デジタルコンテンツの作成・公開
- 4 オープン化・二次利用可能性
- 5 **持続可能性の担保** (アクセス保証、システム運用)
- 6 **相互運用性の確保** (データ流通を促すための取組)
- 7 利活用促進に向けた取組



← (例) 国立国会図書館「つなぎ役モデル」自己点検例 (2020年8月)
(「3か年総括報告書(令和2年8月)」補足資料より)

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/r0208_3kanen_houkoku_honbun.pdf

4. 国立国会図書館における 最近の取組と課題

デジタル資料長期保存基本計画2021-2025

https://www.ndl.go.jp/jp/preservation/dlib/pdf/NDLdigitalpreservation_basicplan2021-2025.pdf

◆保存対策の優先順位◆

- ① **再生環境の陳腐化、媒体の脆弱性・劣化状況に応じた保存の緊急性**
再生装置・ソフトウェアが陳腐化し入手困難となっている資料、媒体の特性上脆弱なもの、劣化が進行している資料
- ② **唯一性・希少性**
他媒体との内容の重複がない資料、他機関における所蔵が確認できない資料
- ③ **長期保存への社会的ニーズ**
長期保存への社会的ニーズがあり、当館の使命、他機関との役割分担等に照らして当館が対策を行うことが妥当な資料
- ④ **保存のための対策手段の確立状況及び対策に要するコスト**
保存対策の技術的手段及び利用提供方法が確立している資料並びに保存対策の実施に要するコストが過大とならない資料

デジタル資料長期保存基本計画2021-2025

https://www.ndl.go.jp/jp/preservation/dlib/pdf/NDLdigitalpreservation_basicplan2021-2025.pdf

◆保存対策◆

- ① 資料の状態検査：特に光ディスクの検査手法の検討 ⇒後述
- ② **メタデータの整備**：長期保存に必要なメタデータ項目の検討
[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30年8月）「4. 保存メタデータに関する調査結果」（p.7）参照](#) <http://dl.ndl.go.jp/info.ndljp/pid/11133855>
- ③ **適切なファイルフォーマットの選択**：
[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> 参照
- ④ **マイグレーション等の実施**：⇒後述
- ⑤ **保存環境の整備**：⇒後述
- ⑥ 再生環境の維持
- ⑦ **対策後の利用環境等の整備**：⇒後述
- ⑧ 原資料等の保存環境の維持
- ⑨ **長期保存の必要性及び各種情報の周知**：
[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/research.html> 参照、遠隔研修の公開⇒後述

最近の主な取組

1. パッケージ系電子出版物の内製マイグレーション（主に媒体変換）

- 2021年度より本格実施

- フロッピーディスク（FD）は内製で作業中



E2294 – フロッピーディスク保存の新デバイス「ポリーナ」について | カレントアウェアネス・ポータル 参照
<https://current.ndl.go.jp/e2294>

- 光ディスク（CD、DVD、BD等）の状態検査手法の調査

劣化状況：エラーレート (※)

種類の特定制も：再生専用型（-ROM）、追記型（-R）、書換型（-RW）

[光ディスクの状態検査の手法に係る基礎的調査報告書（令和4年6月）](#)

<https://dl.ndl.go.jp/info:ndljp/pid/12300246> 参照



- USBメモリ、書換型光ディスクは受入れ時に判明したら速やかに内製で作業中



(※) JIS X 6257:2022 「長期データ保存用光ディスクの品質判別方法及び長期保存システムの運用方法」 参照

最近の主な取組

2. デジタル化資料等の保存媒体の見直し（保存環境の整備）

◆2021年度まで◆

- アナログ媒体をデジタル化した際、提供用データ（非可逆圧縮）はサーバに保存し、保存用データ（可逆圧縮）は光ディスク（約19万枚※1）で保存
- パッケージ系電子出版物のマイグレーション後データも、光ディスクで保存

◆2022年度から◆

- **LTO（※2）又はクラウドへの移行へ**
- デジタル化資料の保存用データは、光ディスク⇒LTOへの移行を試行中※2021年度～
- マイグレーション後データは、提供用も兼ねて、電子書庫（クラウドとHDDを併用）へ移行準備中

（※1） JIS Z6017 :2013「電子化文書の長期保存方法」に基づくサンプル調査を実施し、エラー値の高い群から実施

（※2） Linear Tape-Open：容量が大きい磁気テープの媒体。定期的な規格のバージョンアップ（新LTOへの移し替え）が必要

最近の主な取組

3. マイグレーション後データの利用環境の整備に向けた調査

- マイグレーションしてデータを取り出しても、現在のパソコン環境で閲覧できない場合もある
- 古いOS、ソフトウェア等を現在の環境上に再現する（エミュレーション・仮想化）技術について、海外の動向調査を実施

[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に関する調査報告書（令和3年7月）](https://dl.ndl.go.jp/info:ndljp/pid/11703206) <https://dl.ndl.go.jp/info:ndljp/pid/11703206> 参照



今後の取組と課題

1. 図書付録等の大量の光ディスクのマイグレーション

- 先ずは2000年以前の図書付録から（2022年度補正（第2号））
- 大量に効率的に高精度な作業の実施方法？
- 正しく複製できたかどうか、複製データの検査・確認方法？
媒体種別も様々、中身のファイルフォーマットも様々
- コピーガード措置等への対応？
コピーガードを解除しなければ、長期保存に適した複製ができない
⇒**マイグレーションシステムの要件調査**へ

2. エミュレーション技術による利用環境の整備

- 国立国会図書館デジタルコレクションで古いOS等の環境を再現する方法？
⇒**技術検証の実施**へ
- クラウドで仮想的にOS等を利用可能？
OS等の利用契約ライセンスの権利処理が課題（古いものは利用保証できないことを理由に新たな契約もできない）

今後の取組と課題

3. デジタル資料の長期保存に係る知見・情報の共有

- 国内の他機関の具体的な対策・取組が知りたい
- 情報共有できる場の構築のため、コミュニティ形成へ？
みんなが知りたいことは優良事例だけでなく、失敗事例も必要か（こういった状況で保存していたらデータが消えてしまった等）

遠隔研修「デジタル資料の長期保存に関する基礎知識」

基礎知識 1	基礎知識 2
<p>1. デジタル資料とは何か</p> <p>2. デジタル資料の長期保存の課題</p> <ul style="list-style-type: none">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 <p>3. デジタル資料の長期保存のための対策</p> <ul style="list-style-type: none">・ビット列の保存のための対策・コンテンツの保存のための対策・その他の対策	<p>1. デジタル資料の長期保存に関する動向</p> <p>2. 「デジタル資料の長期保存に関する国内機関実態調査」の概要—各機関の課題とは—</p> <p>3. 長期保存の課題への対応</p> <ul style="list-style-type: none">①ノウハウ・情報②予算、人員、方針③権利処理④保存システム



(終了後補足) 2/14に公開しました！「デジタル資料の長期保存に関する基礎知識」

<https://ndl.go.jp/jp/library/training/remote/digitalpreservation.html>

遠隔研修のページ | 国立国会図書館 <https://ndl.go.jp/jp/library/training/remote/> 参照

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

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