

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)

はじめに

(本ガイドラインの目的)

デジタルアーカイブには、文化的・学術的・社会的資源が含まれており、社会的課題等を解決していくための基盤となるものである。そのため、これらの資源を長く保存して利活用されるよう長期保存に係る取組は非常に重要である。しかし、これまでに構築されたデジタルアーカイブの中には、保存・継承されることなく消えていったものも見受けられる。貴重なデジタル資源が多数失われることが起きないように、デジタルアーカイブとその資源への長期的なアクセスを保証することは大きな課題の一つである。

ここでいう長期保存とは、利用目的に応じた方法¹で、長期にわたって利用可能性を保つことをいう。つまり、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。

本ガイドラインは、アーカイブ機関がデジタルデータの長期保存のためにどういった観点での検討が必要かを示すためのものである。それぞれの機関の使命に照らして取組を決められるよう、必須の取組と実施が望ましい取組とが分かるようレベルを示すことに留意した。

(長期保存の意義)

デジタルアーカイブは、資料/作品を効率的に管理し、その魅力を最大限に活用するために重要な取組である。特に、アーカイブ機関が所蔵する資料/作品のデジタル化とデジタルデータの長期保存を保証するための取組は、資料/作品の保存と活用の両立にとっても非常に重要である。

近年、災害により貴重な文化財が消失する事態が相次いでいる。文化財に係るデジタルデータは、消失した文化財を復元・復旧するための貴重な情報であり、原資料(オリジナル)の保存と合わせて、デジタルデータの保存・管理・利用提供の在り方の検討が求められている。

1 デジタルデータ管理の基本的考え方

(用語の定義)

本ガイドラインでいうデジタルデータとは、デジタルアーカイブに含まれるデジタルで表現されるもの全てを対象とし、アナログ媒体の原資料をデジタル化したデータ

¹ 利用目的によって、オリジナルのビット列を残さねばならない場合もあれば、中身のテキストが読めればよいという場合も考えられる。

及び収集したボーンデジタルのデータといったデジタルコンテンツのほか、コンテンツに関するメタデータも含む。

（デジタル媒体の特性）

デジタルデータを含む電磁記録資料は、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く、保存メディアによっては数年で利用不能になったり、データそのものが消失したりすることもある。また、保存されたデジタルデータを再生するための機器（PC、OS等）の変化が激しいことに加え、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ことにも留意しなければならない。定期的かつ適切なメンテナンスが必要である点は、紙等の物理的な資料と同じであり、デジタルデータの長期保存のための取組は、所蔵資料全体の保存対策の一部として検討する必要がある。検討に当たっては、デジタルデータの全体像を俯瞰できるようにしておき、そのメンテナンスのためには、定期的な品質チェックを行い、再生機器・アプリケーション等の旧式化や最新の技術動向を踏まえつつ適宜に必要な技術を用いながらデータの品質をリフレッシュしていくことが求められる。

（取組の基本的考え方）

デジタルデータが利用可能な状態で長く維持されるためには、保存と公開（限定された範囲での利用提供を含む。）の両面に対応できるよう、安定的に管理・運用できる体制の構築と人的・財政的リソースの確保が必要である。その際、デジタルアーカイブシステムの場合は開発に加え保守・運用及びそのデジタルデータの保存に係るコストが定常的にかかることや、デジタルデータを光ディスク等で保存する場合は定期的に媒体を検査し媒体変換を行う必要があること等、利用可能な状態の維持に係るデジタルデータ特有の事情に留意して（人的・財政的・技術的）リソースを確保する必要がある。

アーカイブ機関は、確保したリソースで実行可能なデジタルデータの長期保存の方針・計画を作り、実施することが求められる。方針・計画の作成に当たっては、アーカイブ機関の使命・目的に照らして実施すべき事項を検討する必要がある。原資料（オリジナル）を長期保存する必要があるのか、複製物であるデジタルデータだけ保存すればよいのか、何を保存すべきマスターと考えるのかによって、管理・運用の在り方も異なる。例えば、デジタルデータをバックアップするコストと原資料を保存しておき再撮影するコストのどちらが大きいかの比較検討を行い、保存方針を決めることが考えられる。また、文献資料等を扱う場合は、データのビット列保存よりもテキスト等の中身のコンテンツを取り出して保存する方針を立てることも考えられる。

以下、アーカイブ機関が取り組むべき事項を示す。●はアーカイブ機関が必須として取り組むべき事項であり、○は実施することが望ましい事項である。

2 デジタルデータの保存方針・計画策定

(方針・計画策定の考え方)

- 所蔵する資料全体の保存方針・計画の中で、デジタルデータの適切な保存・管理及び利用に向けて、組織としての取組の方針及び計画を策定すること。
- 方針・計画は、デジタル技術の最新動向を踏まえ、定期的に見直しを行うこと。

(方針に盛り込む基本事項)

- 方針の策定においては、長期に管理しマスターとするデジタルデータの範囲を決定し、デジタルコンテンツに関してはその品質、ファイルフォーマット、メタデータ（形式・スキーマ）等の要件、公開・非公開、利用条件、保存期間、データを削除する際の判断基準等を示すこと。
- デジタルコンテンツの長期保存を適切に行えるよう、メタデータの整備を行うことを方針に盛り込むこと。メタデータの整備に当たっては、公開したデータの長期アクセス保証の観点にも留意した取組を含めること。
- 方針には、デジタルデータ管理におけるリスク、例えば使用する保存媒体の寿命、時間経過による権利管理への影響等を把握できる内容を盛り込むこと。
- デジタルアーカイブを終了する場合のデジタルデータの移管やデータの返却といった対応についても、方針で示しておくこと。

(計画に盛り込む基本事項)

- 計画の策定においては、方針に基づき、保存媒体に関すること、保存媒体の寿命に応じたマイグレーション方法等の具体的な保存対策、メタデータの整備方法、デジタルデータを管理するシステムのリプレース（4，5年単位）時の対応等を計画に盛り込むこと。

(方針・計画のための検討事項)

- 保存対象とするデジタルデータの特徴を踏まえた、利用可能な状態を確保できるデータ保全の方法を検討し、方針・計画に盛り込むこと。例えば、物理的な保存メディアの定期検査の方法、デジタルデータの利用可能性の確認方法（メタデータにファイルサイズやハッシュ値等を記録しデータ整合性チェックに利用する等）、マイグレーション又はエミュレーションといったデジタルコンテンツの利用可能性を確保するための保存方法等を検討することが考えられる。
- デジタルデータの破損や紛失に備えたバックアップ方法も検討し、方針・計画に盛り込むこと。その際、バックアップの実行頻度、保持方式（フルか差分か増分）、バックアップ及びリカバリの手順等について確認すること。また、バックアップデータの保管及びデータのリストアを実施するためのデータバックアップシステムの確保についても検討すること。バックアップシステムとは、いわゆるサーバストレージや保存媒体を用いたバックアップの仕組みのほか、組織内又は組織間でのデジタルデータ保存のための仕組み全体をいう。

- 災害等に備え、保存対象とするデジタルデータの複製物を複数の遠隔地で分散配置すること、又は信頼できるクラウドストレージを用いる等の対応を方針・計画に盛り込むこと。

(つなぎ役の役割)

- つなぎ役に相当するアーカイブ機関の場合は、コミュニティ全体のためのデジタルデータ管理や長期保存に関する方針・文書等を用意すること、また情報共有の仕組みを用意すること。

3 デジタルデータ管理

(管理者の配置)

- データマネージメントができるデータの管理部署又は管理者を配置すること。業務の継続性に配慮し、可能な場合はデータの管理を行う担当者を複数人配置すること。
- どうしても組織としての管理者の配置が難しい場合は、デジタルデータの状態を把握できる担当者を配置し、定期的に状態を確認できる体制を構築すること（例えば、1年に1度サンプル抽出したデータを確認するなど）。

(関連情報の整理)

- デジタルデータが作成された来歴情報・権利情報等（内容、時期、目的、作成者、担当者、権利者、利用条件等）について、デジタルデータの受入れ又は作成時点で十分な記録を行い、その記録をデジタルデータとともに残しておくこと。特にデジタル化作成時は仕様書等のドキュメントも保管しておくこと。
- 所蔵するデジタルコンテンツの保存に加え、更新作業が伴うメタデータやサムネイル/プレビューの保存にも取り組むこと。メタデータ項目については、それぞれの項目の意味が将来的にも把握できるようドキュメントを整備し、維持すること。
- 公開用の提供データと保存用データが異なる場合、その対応関係が把握できるように管理すること。また、どの保存メディアにどのデータがあるかといったことにも留意してデータを管理すること。

(データ管理のための取組)

- 計画に基づき、適切なマイグレーションとバックアップを行うこと。
- デジタルコンテンツを管理者の必要に応じて簡便に抽出できるようにメタデータを整備しておくこと²。

2 デジタルコンテンツを発見するためのメタデータ整備においては、コンテンツの内容等のメタデータと一緒にサムネイル画像の URL やデジタルコンテンツの URL の抽出も可能としておくことが望ましい。

- デジタルコンテンツの改変作業に関しては、特定の権限をもつものだけが作業できるようにし、改変内容を把握できるようメタデータの来歴情報を更新すること。

4 デジタルデータ保存

(マスターデータの保存対策)

- マスターとして長期に保存すべきデジタルデータ（マスターデータ）に対しては、災害や大規模なシステム障害等への対応のため、データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えること。
- マスターデータの複製物の保存場所の分散化においては、災害発生リスクの異なる複数の遠隔地での配置に考慮すること。

(マスターデータのデータ形式)

- マスターデータについて、長期的な利用可能性を考慮し、特定製品等に依存せず、仕様等が公開され、かつ広く普及している（国際標準等で定められた）データ形式（フォーマット）を採用すること。
- マスターデータについて、前項に基づくデータ形式を採用するに当たって、同じ組織内では種類を制限してできる限り同じデータ形式を用いるようにすること。ただし、画像、動画、音声などコンテンツの種類によってそれぞれに対応したデータ形式を採用すること。
- マスターデータのデータ形式は、オリジナルのデジタルコンテンツをより正確に再現しやすくするため、可逆圧縮方式の採用や、カラープロファイルの埋め込みが可能な形式の採用を検討すること。

(メディアに応じた保存対策)

- 光ディスク、LTO テープ、HDD などの保存するメディアの特性に応じて、メディアの寿命や記録・再生装置の互換性等も踏まえ、定期的に新しいメディアへの移行を適切に行う必要がある。特に HDD の場合は、複数台の複製物を用意し、4～5年置きに移行すること。

(アーカイブシステムにおける保存対策)

- アーカイブシステムの運用・管理と、デジタルデータの長期保存とは区別して対策を講じ、システムにあるデジタルデータはいつでも抽出できるようにしシステムから切り離しても再現できる状態を担保できるようにしておくこと。
- 長期的な保存を考えたアーカイブの仕組みに関する国際標準であり、欧米で各種ガイドライン等が整備されている OAIIS 参照モデル³に留意した取組を行うこと。

3 デジタル情報の長期保存アーカイブシステムに関する枠組みを規定した国際標準規格 (ISO 14721:2012)

5 安定稼働できるシステム

(サービスレベルに応じた運用)

- アーカイブ機関毎に自らのデジタルアーカイブのサービスレベルを検討し、サービスレベルに応じた機密性、可用性及び安全性の確保に留意し、デジタルデータの公開及び管理システムを確保・運用すること。
- 管理システムでは、長期間のシステム運用性の視点から、デジタルデータの保存領域の拡張性及び移行可能性を確保すること。また、組織のセキュリティポリシーを遵守すること。

(持続可能とする運用コストの確保)

- ストレージ装置や各種デバイス、システム自体は数年ごとのリプレースが必須であり、そのための経費、さらにメンテナンスに従事する人員の確保も含め、運用面のコストをデジタルアーカイブ構築時に見込んでおくこと。

(ベンダーロックインの防止)

- いわゆる「ベンダーロックイン」(ある特定業者のシステムに依存する状態。業者を変えられずコストが上がり、業者がシステム提供を終えたことでデジタルアーカイブが維持できなくなるなどの弊害がある。)にならないよう留意すること。

(つなぎ役によるポータル構築)

- つなぎ役に相当するアーカイブ機関の場合は、コミュニティに属するアーカイブ機関が自らデータの整備・公開ができるよう、安定的な統合プラットフォームを構築・運用すること。

(外部サービスの利用時の留意点)

- 長期保存に係る外部サービスを利用する場合は、利用規約を精査し、組織のセキュリティポリシー及びデジタルデータの権利状態に抵触しないことを確認すること。
- 外部のクラウドストレージを利用する場合は、利用規約において、組織のデータの位置付けやセキュリティポリシーに照らして、デジタルデータの保持に関する保証内容に問題がないことを確認する必要がある。また、ISO等の認証⁴を得ているかどうかを確認すること。

4 ISO/IEC 27001:2013 (JIS Q 27001:2014「情報技術—セキュリティ技術—情報セキュリティマネジメントシステム—要求事項」と同等)、ISO/IEC 27002:2013 (JIS Q 27002:2014「情報技術—セキュリティ技術—情報セキュリティ管理策の実践のための規範」と同等)、ISO/IEC 27017:2015 (JIS Q 27017:2016「情報技術—セキュリティ技術—JIS Q 27002に基づくクラウドサービスのための情報セキュリティ管理策の実践の規範」と同等)といった情報セキュリティ管理に関する国際規格等。

6 メタデータ管理

(メタデータの更新・抽出)

- メタデータの更新（修正・削除等）作業や抽出作業等は、必要に応じてアクセス権の制限を設けた上で、簡便に行うことができるようにすること。

(メタデータの形式・項目)

- 相互運用性の観点から、分野で標準としている、又は分野内で広く用いられているメタデータ形式によるメタデータの管理を行うこと。その上で、可能な限り、時間の経過によって変化しないものをメタデータの必須項目として選びつつ、継続的にメンテナンスしていくこと。
- 長期保存のため、資料ごとに、デジタルデータ作成者、作成日、更新日、原資料の権利状態、デジタルデータの権利状態（著作権保護期間中かどうか、保護期間満了の時期、権利者、利用条件等のライセンスなど）の情報をメタデータとして管理すること。
- どうしても個別の資料ごとにメタデータを整備することが難しい場合は、同様の資料をまとめていくつかのデータセットとし、そのセットごとにデジタルデータの来歴がわかる情報として、作成内容、目的、時期、作成者、作成日/更新日、デジタルコンテンツのマスターデータを改変・差替え等を行った場合はその履歴情報等を管理すること。

(管理用識別子の付与)

- 長期にわたり安定的にデータを保存・管理するため、個別のコンテンツを判別し認識できる識別子（重複しない一意の管理番号）を付与すること。（識別子の付与方法に関しては、（別紙1）を参照）。

7 アクセス保証の仕組み

(固定 URL の提供)

- ウェブで公開しているデジタルデータへの長期アクセスを保証する取組として、公開しているデジタルコンテンツやメタデータを紹介する詳細表示ページを用意し、永続的な固定 URL でアクセスできるようにすること。（つなぎ役等のポータルを通じて提供されるようにすることもよい。）
- 詳細表示ページに加え、公開しているサムネイル/プレビュー又はデジタルコンテンツのそれぞれについても、永続的な固定 URL でアクセスできること（つなぎ役等のポータルを通じて提供されている場合を含む。）
- ドメインドロップキャッチ⁵の被害を避けるため、組織のサブドメインを使うこと。独自ドメインを使わざるを得ない場合は、定期的にリンク切れの確認を行い、維持に努めること。

⁵ デジタルアーカイブ用に取得したドメイン名の有効期限が切れて、第三者に使われてしまうこと。

(アクセスを保証するための永続的識別子の付与)

- 公開データの詳細表示ページ又はデジタルコンテンツに対し、DOI (Digital Object Identifier)⁶等の永続的識別子を付与するなどして長期アクセス保証を意識して公開すること⁷。
- つなぎ役に相当するアーカイブ機関の場合は、属するコミュニティに対し、DOI等のコンテンツへの永続的識別子の付与・普及に努めること。

(メタデータの提供方法)

- 公開された詳細表示ページを探せるよう、メタデータを利便性の高い方法(検索機能、地図表示等)で公開すること。また、公開するメタデータを最新の内容に保つ仕組みを備えること。
- アクセスを保証しデータの相互運用性を担保するため、URI (Uniform Resource Identifier)を公開用識別子としてメタデータに付与し、提供すること。なお、URIの付与においては、既存の管理番号などを用いて一括生成するのが合理的である。特に6で記載したメタデータの管理用識別子がURIに含まれるような形で設定すると、URIと管理用識別子の対応関係が容易に判別できる。
- メタデータをファイルで提供する場合のファイル形式は、JSON、XML、エクセル等機械可読性の高いフォーマットを採用すること。また、エクセルや表形式テキストの場合は、機械可読性を担保するため、(別紙2)「活用できる表形式のデータとは？」の要件を満たすこと。

8 デジタルデータ移行性の担保

(データ形式の移行性確保)

- システム更新や組織の統廃合によってデータが失われないよう、データ移行性を確保するためには、メタデータも含め、全てのデジタルデータの管理において、データ形式は特定製品等に依存せず、仕様等が公開され、かつ広く普及している(国際標準等で定められた)形式とすること。また、全てのデジタルデータの抽出を可能とすること。

(包括的な権利処理)

- デジタルアーカイブで提供しているデータについて、組織統廃合時のデータ移行や他機関へのデータ譲渡に対応できるよう、包括的な権利処理を行うこと。例え

6 コンテンツの電子データに付与される国際的な識別子。ISO 26324:2012として国際標準化されている。DOIの登録には国際DOI財団が認めるDOI登録機関への参加が必要であり、我が国のDOI登録機関としてはジャパンリンクセンター(JaLC)がある。なお、DOIはデータの管理主体が変わっても識別子自体は変わらないため、この観点からも持続可能なデータの提供を行うことに資するものである。

<https://japanlinkcenter.org>

7 永続的アクセスを保証する仕組みの一つに、国立国会図書館のウェブサイト収集保存事業(WARP)<URL:<http://warp.ndl.go.jp/>>がある。WARPにおいて、ウェブサイトの収集対象となっている場合(データベースを除く)、WARP上で永続的アクセスが保証される。

ば、デジタルデータの受入れ時に、著作権の譲渡契約を結ぶこと、オープンなライセンスを付与しておくこと（これにより権利処理を要せず移行可能性を担保できる。）、デジタルアーカイブを運営する組織が、アーカイブの運営が困難になった場合に組織が認めた第三者に対してその組織と同じ条件で利用できることを契約に盛り込むこと等が考えられる。

9 人材育成

（体制整備）

- アーカイブを運用面、技術面で管理できる能力をもった人材を育成し、継承していける体制づくりを行うこと。
- 担当者が変わっても、体制が維持できることに配慮した仕組みづくり（組織・文書等）を行うこと。
- 体制の構築に当たっては、デジタルデータを管理・運用するスタッフが継続的な研修と専門的な能力開発を受けられるよう環境を整備すること。
- デジタルアーカイブに関する研修会や資格認定制度等を活用し、自館のデジタルアーカイブ担当者の教育に努めること。
- 自館でデジタルアーカイブに関する勉強会等を開催し、担当者以外にも知識の共有を図るよう努めること。
- デジタルアーカイブを運営する他のアーカイブ機関との間で人的交流を進め、新しい知識・情報を共有する場をつくること。

（つなぎ役による人材育成支援）

- つなぎ役に相当するアーカイブ機関の場合は、デジタルアーカイブに関する研修会・講習会の実施等によって、当該コミュニティのアーカイブ機関に関する人材育成を支援する活動を行うこと。

(別紙1) 識別子の付与について

(識別子の付与方法)

識別子の付与には、2種類の方法がある。組織内で決めたルールに基づき、一つは手作業で一意となる(重複しない)管理番号を付与する方法、もう一つは機械的に管理番号を付与する方法である。手作業で管理番号を付与する場合は、管理番号が重複しないように運用することが求められる。識別子(管理番号)は数字でもそれ以外でもよいが、国際的な流通を意識する場合、漢字等は使わない方がよりよい。

表形式により手作業でメタデータを整備している場合は、第一列目に、同一ファイル内で重複しない管理番号をつけるなどすればよい。ファイル内で一意であれば、「ファイル名」+「管理番号」を機関における識別子として利用できる。複数のファイルを複数のフォルダごとに管理している場合には、それらを「/(スラッシュ)」でつないで「フォルダ名/ファイル名/管理番号」などとすればよい。

機械的に管理番号を付与する場合は、一意に特定するための管理番号を付与するルールとなるよう、プログラムなどを調節する必要がある。データベースシステムを用いてメタデータ等を整備している場合は、一意の管理番号が自動で付与されるように設定することがよい。

手動にせよ機械的な付与にせよ、一度付与した識別子(管理番号)は、原則として変更しないようにしなければならない。同じ番号を別のデータに付与するような使い回しをしてはならないし、機械的な付与においては、削除した場合に自動で新たな番号が振り直されることのないよう気をつける必要がある。

なお、メタデータを新規に整備する場合、自館の所蔵品目録や展示会図録等から全作品の一括付与を仮に行う方法なども考えられる。その識別子を用いて自ら保有する作品情報や解説等のデータを整備・公開するなどの活用が可能になる。

(識別子の種類)

識別子の付与は、データを共有する際の相互運用性の担保においても重要な取組である。URI (Uniform Resource Identifier) のほか、デジタルコンテンツの識別子として、国際標準規格である DOI (Digital Object Identifier) がある。アーカイブ機関が個別のコンテンツごとに URL を指定してアクセス可能なウェブページを提供している場合は、国際的な流通促進を考慮して、DOI を付与することも考えられる。DOI は永続的な識別子であるため、DOI のメタデータのメンテナンスが必要であるが、URL が変更されても長期アクセスが保証される点にメリットがある。DOI の付与には、我が国で唯一の DOI 付与機関であるジャパンリンクセンター (JaLC) の会員になる必要がある。

機関を特定するための識別子としては、国際標準規格である ISIL (International standard identifier for libraries and related organizations) があり、我が国では、国立国会図書館が付与・管理を行っている。つなぎ役は、機関コードの管理を行う場合に ISIL の活用が考えられる。

([「デジタルアーカイブの構築・共有・活用ガイドライン」](#) p.11 より)

(別紙2)

●活用できる表形式のデータとは？

エクセル等のアプリケーションソフトを用いた表形式のデータを作成する際は、機械可読性を担保するため、以下の点に留意して作業を行わなければならない。

- 表データは、列を項目、行をレコードとして構成する。
- 1つのデータセルには、1つの要素のみ記入する。
- データセルに、整形や位取りのための文字（スペース、改行、数値でのカンマ等）を含めない。
- 数値等のデータの値やタイトル、単位以外の情報を、セルに含めない。
- 値が存在しない場合を除き、データセルを空白にしない。（データ値を省略しない）
- 英数字は半角とし、ひらがな・カタカナは全角とする。
- データ列の内容を示す項目名は、1行で構成する。
- レイアウトのための空行・列は使わない。
- データセルにふりかな、コメント・注釈などの加工は行わない。
- 全てのセルは、他のセルと結合しない。
- ファイルのヘッダやコメント等の内容が残っていないか、プロパティ情報（ファイルの属性情報）に問題がないか十分に確認してから公開する。

(参考) 「数値(表)、文章、地理空間情報のデータ作成に当たっての留意事項」
(二次利用の促進のための府省のデータ公開に関する基本的考え方(ガイドライン) 別添2 (2013年6月25日 各府省情報化統括責任者(CIO)連絡会議決定 2015年12月24日 改定))
http://www.kantei.go.jp/jp/singi/it2/densi/kettei/data/gl26_betten2.pdf

(「[デジタルアーカイブの構築・共有・活用ガイドライン](#)」 p.39)