

デジタルアーカイブのための長期保存・長期利用保証ガイドライン(素案)

◎ 高度なレベル ● 標準レベル (○一般的なこと)

ガイドラインの要素を示すため、本委員会やヒアリングで出た意見、関連する資料等の記述を目次ごとに抽出し、ガイドラインの要素を示した。足りない要素、方向性として問題がある等あれば、ご意見いただきたい。

なお、典拠がないものは、事務局が次の資料を参考に作文した。

・「研究データリポジトリ整備・運用ガイドライン」(平成 31 年 3 月、国際的動向を踏まえたオープンサイエンスの推進に関する検討会)

<<https://www8.cao.go.jp/cstp/tyousakai/kokusaioopen/guideline.pdf>>

はじめに(長期保存・長期利用保証の意義)

- これまでに構築されたデジタルアーカイブの中には、保存されることなく消えていったものも多く、デジタルアーカイブ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。(「第二次中間取りまとめ」p.14)
- デジタルアーカイブの取組は、資料/作品を効率的に管理し、また、その魅力を最大限に活用するために重要なことであるが、そもそも、アーカイブ機関が所蔵する資料/作品の保存のためにも、それらのデジタル化とデジタルデータの長期保存・長期利用を保証するための取組は非常に重要である。(第 8 回実務者検討委員会)
- 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、原物(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。(第 8 回実務者検討委員会)
- 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・長期利用保証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、必須な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。
- なお、本ガイドラインでいうデジタルデータとは、原物のデジタル化データ及び収集したオープンデジタルのデータといったデジタルコンテンツのデータだけでなく、コンテンツに関するメタデータも含む。
- 本ガイドラインでいう長期利用保証とは、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。

1 デジタルデータ管理の基本的考え方

- 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。(ヒアリング)
- ◎ 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が望ましい。

- 人的・財政的リソースの確保については、原物の保存・利用とはまた異なる観点が必要となる。例えば、デジタルアーカイブシステムの開発・保守・運用及び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も、定期的に媒体を検査し媒体変換を行う必要があることに留意しなければならない。(ヒアリング)
- アーカイブ機関は、確保した(人的・財政的・技術的)リソースで実行可能な計画を作り、実施することが求められる。(ヒアリング)
- デジタルデータの管理方針は、アーカイブ機関の使命・目的に照らして検討する必要がある。原物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によって、管理・運用の在り方も異なる。例えば、小さなアーカイブ機関の場合、デジタルデータをバックアップするコストと原物を保存しておき再撮影するコストと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も考えられる。(ヒアリング)
- デジタルデータもメンテナンスする必要がある。デジタルデータに対し、定期的な品質チェックを行い、最新の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。(ヒアリング)

➤ アーカイブ機関が踏まえておくべき、デジタルデータ管理の基本的な考え方として足りているか。

2 デジタルデータの保存方針・計画策定

- アーカイブ機関は、デジタルデータの適切な保存・管理及び利用に向けて、組織としての取組方針及び計画を策定することが求められる。
- 方針の策定においては、長期に管理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、データを削除する判断基準等を示すことがよい。
- 保存す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、方針に盛り込むことが必要である。その際、バックアップの実行頻度、保持方式(フルか差分か増分か)、バックアップ及びリカバリの手順等について確認するのがよい。
- 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が必要である。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保存のためのスキーム及びデータ保存のための仕組み全体を含む。
- ◎ 災害等に備え、バックアップデータの複数の遠隔地での分散配置に対応することを方針に盛り込むことが望ましい。
- ◎ 長期アクセスを保証する観点から、デジタルアーカイブの終了といった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、方針で示しておくのが望ましい。
- 方針は、デジタル技術の進展を踏まえ、定期的に見直しを行う必要がある。
- 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等を認識

できる内容が盛り込まれることがよい。(ヒアリング)

- 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、管理システムにおけるリプレイス(4, 5 年単位)を盛り込むことが必要である。(ヒアリング)
- ◎ つなぎ役の場合は、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が求められる。(「デジタルアーカイブアセスメントツール」つなぎ役モデル1⑤)

- デジタルデータ保存の方針及び計画策定に当たって、アーカイブ機関が最低限決定しておくべきことは何か。
- アーカイブ機関が方針・計画を策定するに当たって参考となる媒体変換等の保存方法の選択基準はあるか。

3 デジタルデータ管理

- デジタルデータの管理に当たっては、データの更新(修正・削除等)作業や抽出作業等を簡便に行うことができるようにする必要がある。(「デジタルアーカイブアセスメントツール」標準・先進・つなぎ役モデル5⑤)
- データの更新(修正・削除等)作業に関しては、特定のアクセス権をもつものだけが作業できるようにすることがよい。(ヒアリング)
- 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が求められる。
- デジタル化作成時の状況が分かるよう、デジタル化の際のドキュメント等の記録を残しておくことが必要である。(「デジタルアーカイブの構築・共有・活用ガイドライン」p.13)
- また、収集したデジタルデータの受入れ時には、デジタルデータが作成された背景(内容、時期、目的、作成者、担当者、権利者、利用制限等)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うことが必要である。(ヒアリング)
- ◎ データマネジメントができるデータの管理部署又は管理者がいることが求められる。(ヒアリング)
- アーカイブ機関は、所蔵するデジタルコンテンツの保存に加え、メタデータやサムネイル/レビューの保存にも取り組むことが求められる。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、維持する必要がある。(「デジタルアーカイブの構築・共有・活用ガイドライン」p.12)
- 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必要がある。また、どのサーバ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必要がある。(ヒアリング)

- メタデータとデジタルコンテンツの保存の関係性について、記載しておくべき事項はあるか。

4 デジタルデータ保存

- 災害や大規模なシステム障害等への対応のため、バックアップによる公開用の提供データの複製、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が必要である。（「デジタルアーカイブアセスメントツール」標準モデル5⑦）
- ◎ 提供データに加え、保存用のマスターデータについても、バックアップや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が求められる。（「デジタルアーカイブアセスメントツール」先進・つなぎ役モデル5⑦）
- ◎ 保存用マスターデータ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が望ましい。（ヒアリング）
- デジタルコンテンツの保存用に作成する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、仕様書等が公開され、かつ広く普及している（国際標準等で定められた）データ形式を採用していることがよい。（「デジタルアーカイブアセスメントツール」標準・先進モデル5⑥）
- デジタルコンテンツの保存用に作成するマスターデータについて、前段に基づき広く普及している等の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がよい。（ヒアリング）
- ◎ 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として、欧米では OAIIS 参照モデル¹を基にした各種ガイドライン等が整備されており、我が国でも留意することが望ましい。（「デジタルアーカイブの構築・共有・活用ガイドライン」p.13）

- 保存用データがあったとしても、公開用の提供データは複数化・分散化によるバックアップが必要か。（保存用データと提供データの両方ともバックアップが必要か。）
- 複数のフォーマットの中から、保存用フォーマットを選ぶ際のポイントは何か。（「特定製品等に依存せず、仕様書等が公開され、かつ広く普及している（国際標準等で定められた）」という表現が適切か。）
- 複数のファイルフォーマットを保有せざるを得ないアーカイブ機関にとっては、どのような保存が望ましいか。
- 保存メディアとしては、特定製品への依存は避けられないものもある。HDD であれば依存は避けられるが、保存メディアとしては適切ではない。ガイドラインにはこういった点をどのように盛り込めばよいか。

5 メタデータ管理

- アーカイブ機関は、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がよい。
- 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が求められる。（「デジタルアーカイブの構築・共有・活用

¹ 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)

ガイドライン」p.8)

- 長期保存のために必要なメタデータの項目として、デジタルデータ作成者、作成日、…(必要な項目を議論のち追加)を管理することが望ましい。
- 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)を付与することが必要である。 (「デジタルアーカイブアセスメントツール」先進・標準モデル5①)
- (識別子の付与方法に関しては、(別紙 1)を参照)
- ◎ つなぎ役は、当該コミュニティ内に属する組織・団体に対し、機関識別子の付与・管理を行うことが考えられる。その際、コミュニティ内に対して付与・管理することも考えられるし、特定のコミュニティを超えた標準的な識別子の付与を推奨することも考えられる。また、当該コミュニティに対し、DOI(デジタルオブジェクト識別子)等のコンテンツへの永続的識別子の付与・普及に努めることも望まれる。(「デジタルアーカイブアセスメントツール」つなぎ役モデル5①、ヒアリングを受けて一部修正)

- 長期保存のために必要なメタデータ項目は何か。
- メタデータ管理のなかで、権利情報に関する取組も記載した方がよいか。その場合、どのような内容がよいか。

6 アクセス保証の仕組み

- 長期にアクセスを保証するため、アーカイブ機関は、公開メタデータについて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が求められる。(一部でも可。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。)(「デジタルアーカイブアセスメントツール」標準モデル5②)
- ◎ 用意した公開メタデータのための詳細表示ページに対し、DOI 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が望ましい。(「デジタルアーカイブアセスメントツール」つなぎ役・先進モデル5②)
- ◎ 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)で公開することが望ましい。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ていることがよい。(「デジタルアーカイブアセスメントツール」先進モデル2②)
- ◎ アーカイブ機関は URI(Uniform Resource Identifier)を識別子としてメタデータに付与し、提供することが求められる。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 5 で記載したメタデータ管理における識別子が URI に含まれるような形で設定すると、URI とメタデータの識別子の対応関係が容易に判別できる。(「デジタルアーカイブの構築・共有・活用ガイドライン」p.20)
- URI の付与をアーカイブ機関が自ら行えない場合は、活用者による URI 作成を可能とするため、メタデータの管理ファイルをウェブで公開することが求められる。その際、ファイル単位で管理番号を唯一にしておくことが必要である。ただし、メタデータのファイルをウェブ上の安定的な(URL が変更されない)ところに置くこととし、ファイルの形式は、(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が求められる。（「デジタルアーカイブの構築・共有・活用ガイドライン」p.21）

- 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は、永続的な固定 URL でアクセスできる（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。）ようにすることが望ましい。（「デジタルアーカイブアセスメントツール」標準・先進・つなぎ役モデル5③）

➤ ドメイン更新忘れでアクセスできなくなるケースがあるため、URL を数年にわたり維持するための計画・方法として、ガイドラインに盛り込むべきことは何か。

7 デジタルデータ移行性の担保

- プラットフォームやサイトについて、安定性を確保するため、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ことが重要である。（「デジタルアーカイブの構築・共有・活用ガイドライン」p.12）
- メタデータも含め、サムネイル/プレビューやデジタルコンテンツの管理において、特定の機器（システム、メディア）等に依存しないデータ形式で保持している、又はデータの抽出を可能とすることが求められる。（「デジタルアーカイブアセスメントツール」標準モデル5④）
- ◎ データ形式の確保に加えて、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が望ましい。（「デジタルアーカイブアセスメントツール」先進・つなぎ役モデル5④）

➤ 「特定の機器（システム、メディア）等に依存しないデータ形式で保持」とは、具体的にどのようなことに留意すればよいか。

➤ 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には、具体的にどのような内容を盛り込めばよいか。

8 安定稼働できるシステム

- アーカイブ機関は、自ら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、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・運用することが求められる。
- 管理システムでは、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も必要である。
- 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が求められる。（「デジタルアーカイブの構築・共有・活用ガイドライン」p.12）
- 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必要がある。（「デジタルアーカイブの構築・共

有・活用ガイドライン」Q&A p.41)

- ◎ つなぎ役は、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・運用することが望ましい。（「デジタルアーカイブアセスメントツール」つなぎ役モデル5⑧）

- 外部サービスの利用に当たって、アーカイブ機関が留意すべき点は何か。
- クラウドストレージを利用するに当たってどういった点に留意すればよい。選択する際の指針となるようなものはあるか。

9 その他（人材育成・セキュリティ等）

- 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が求められる。（ヒアリング）
- アーカイブ機関は、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が望ましい。（「デジタルアーカイブアセスメントツール」標準モデル1④）
- ◎ 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ができるとよい。（「デジタルアーカイブアセスメントツール」先進モデル1④）
- ◎ つなぎ役は、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が望ましい。（「デジタルアーカイブアセスメントツール」つなぎ役モデル1④）

- セキュリティに関してガイドラインに盛り込むべきことはあるか。

（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」などとするればよい。

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。つなぎ役が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）

(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)