

# 「電子公文書等の長期保存フォーマットを含む 長期保存に関する調査検討」の状況について

令和5年7月

独立行政法人 国立公文書館



# 1. 令和4年度の調査検討内容の振り返り

## (1) 令和4年度前半の調査検討による方向性

- ・ すべての移管文書を一律に一定のフォーマットに変換する必要はない
- ・ 長期保存にあたってのリスクが低いフォーマットである「標準的フォーマット」(仮称)については、そのまま移管
- ・ 行政機関等には移管文書は標準的フォーマットで作成、保存するよう求める
- ・ 長期保存にあたってのリスクが高いフォーマットについてはフォーマットを変換して移管

## (2) 令和4年度・5年度に残されていた課題

- ① 長期保存にあたってリスクが低いと考えられる標準的フォーマットの特定、リスト化
- ② 長期保存にあたってリスクが高いと考えられるフォーマットの特定、リスト化
- ③ ①②でリスト化していないフォーマットの文書が移管された場合のリスク評価方法
- ④ 高リスクのフォーマットによる文書を変換する場合の「長期保存フォーマット」の特定

※過去に移管された文書、行政機関等への調査を行う。

今回調査検討を進めた課題

- ⑤ ④の「長期保存フォーマット」に変換する技術等の実装に向けての検討
- ⑥ 行政機関等と国立公文書館との役割分担の検討

※移管前の現用文書について、移管時まで見読可能な状態を維持(最新のバージョンにアップデートしておくなど)することは、公文書管理法第6条により、各行政機関に義務づけられていると考えられる。

令和5年度に調査検討を行う

## 2. リスク評価方法に関する調査検討

③のフォーマットのリスク評価方法について以下（１）～（３）の順にリスク評価の考え方の調査を行うと共に、専門家等（※）へのヒアリングを行った。今後さらに、国立公文書館がどのようにリスク評価をするか等を検討し「リスク評価方法」を策定する。

### （１）米豪の国立公文書館等におけるフォーマットのリスク評価の考え方を把握

| 米国国立公文書館（NARA）<br>及び米国議会図書館（LC）             | 豪州国立公文書館（NAA）                                 | リスク評価<br>の考え方 |
|---|---|---------------|
| 公開性（Disclosure）                             | 十分に文書化されている（Well documented）                  | A             |
| 普及性（Adoption）                               | 広く採用されている（Widely adopted）サポートされている（Supported） | B             |
| 透明性（Transparency）                           | 非圧縮（Uncompressed）                             | C             |
| 自己記述性（Self-documentation）                   | メタデータをサポート（Metadata friendly）                 | D             |
| 外部依存性（External dependencies）                | プラットフォームに依存しない（Platform independent）          | E             |
| 特許の影響（Impact of patents）                    | 制約がない（Unrestricted）                           | F             |
| 保護の仕組み<br>（Technical protection mechanisms） |   | G             |

（※）ヒアリング先

- 〈専門家〉 上原 哲太郎氏（立命館大学教授）、安岡 孝一氏（京都大学人文科学研究所 附属東アジア人文情報学研究センター教授）、杉本 重雄氏（筑波大学名誉教授）  
溝上 卓也氏（株式会社デジサイン DX支援事業本部フェロー）、
- 〈専門団体〉 公益社団法人 日本文書情報マネジメント協会（木村道弘氏、名護屋豊氏、相馬淳人氏）  
ARMA International 東京支部（宮崎一哉氏）
- 〈資料保存機関〉 国立国会図書館、国立映画アーカイブ

(2) 米豪の国立公文書館等におけるフォーマットのリスク評価の考え方を踏まえ、次のA~Gを低リスクと評価するための基本的な条件として整理した。

| リスク評価の考え方           | リスク評価に用いる理由   |
|---------------------|---|
| A 仕様が公開されていること      | 米豪では仕様が公開していることを評価項目としており、仕様が公開されたフォーマットは長期保存に必要な技術情報を確認・取得することができる |
| B 世界中で広く使用されていること   | 米豪いずれも「広く採用されている」、「普及性」を評価項目としている                                   |
| C 可逆圧縮等が可能なこと       | 圧縮の際に記録の内容が一部欠損するような圧縮技術は長期保存上のリスクとなり得るため                           |
| D メタデータを付与できること     | 米豪でもメタデータのサポートを評価項目としており、メタデータで管理できる必要がある                           |
| E 技術要素が独立していること     | 特定の環境（ハードウェア、OS、ソフトウェア）に依存していること、技術仕様が複雑であることは長期保存上のリスクとなり得るため      |
| F 特許やライセンス等の制約がないこと | 特許、ライセンスに過度の条件があると長期保存上のリスクとなり得るため                                  |
| G プロテクションの制約がないこと   | コンテンツの保護技術（DRM）、暗号化などがあると、将来利用できなくなる可能性があり、長期保存上のリスクとなり得るため         |

(3) その他、我が国の状況を踏まえ、次の点は必須と考えた。

- ・次期文書管理システム等においてフォーマットの変換や保存が低コストで行えること
- ・一般的な日本語が取り扱える（例：漢字、ひらがな）こと

# (参考)米豪の公文書館等の長期保存上のリスク評価の考え方

| 米国 NARA/LC共通 (グローバル要素)                      |   | 豪州 NAA 長期保存フォーマット                            |  |
|---|---|--|--|
| Disclosure<br>(公開性)                         | 完全でオープンなドキュメントと仕様書を通じて、そのフォーマットに関する技術情報を利用できるか。   | Well documented<br>(十分に<br>文書化されている)         | フォーマットは識別可能であり、仕様が公開されていること。仕様は十分に詳細であり、適切なスキルを持つ人はフォーマットを正しく表示するソフトウェアを作成できること。 |
| Adoption<br>(普及性)                           | そのフォーマットが、特に類縁機関で広く使用されているか。<br>複数のツールセットに統合されていたり、特定ベンダーの実装にロックされていないか。<br>コミュニティのユーザーグループがアドバイスやサポートを提供しているか。 | Widely adopted<br>(広く採用されている)                | 世界中で広く使用され、サポートされていること。  |
|   |   | Supported<br>(サポートされている)                     | ベンダー、コミュニティ、またはサードパーティが、テクニカルサポートを利用できること。                                       |
| Transparency<br>(透明性)                       | 基本的なツールで解析可能なフォーマットか。<br>標準的な文字エンコーディングがサポートされているか。<br>非可逆圧縮や暗号化がされているか。  | Uncompressed<br>(非圧縮)                        | フォーマットは圧縮されていない、可逆圧縮を使用していること。   |
| Self-documentation<br>(自己記述性)               | そのフォーマットのファイルは、埋め込みされるメタデータにより、自身のコンテンツや構造を説明できるか。  | Metadata friendly<br>(メタデータをサポート)            | メタデータへのサポートがあること。  |
| External dependencies<br>(外部依存性)            | 表示・再生 (レンダリング) や利用において、特定のハードウェア、OS、ソフトウェアに依存しないフォーマットであるか。   | Platform independent<br>(プラットフォームに<br>依存しない) | フォーマットは幅広いソフトウェアでサポートされているか、プラットフォームに依存しないこと。                                    |
| Impact of patents<br>(特許の影響)                | そのフォーマットに、長期的な使用を妨げるような条件の特許がないか。例えば、ライセンス条件に使用料が含まれている場合、コストが高くなり、予測できない可能性がある。                                | Unrestricted<br>(制約がない)                      | フォーマットは特許や法的制約を受けないこと。<br>オープンソースの原則を使用することが望ましい。                                |
| Technical protection mechanisms<br>(保護の仕組み) | そのフォーマットがDRM、暗号化あるいはその他の保護メカニズムの使用を必要とする場合、コレクションの所蔵者 (custodians) はコンテンツへの将来のアクセスを確実に維持することができるか。              |  |  |

### 3. 長期保存の観点からリスクが低いフォーマットリストの調査検討

米英豪の国立公文書館等において長期保存上のリスクが低いフォーマット等のリストが公開されており、上記機関で共通しているフォーマットを抽出し、専門家等に提示しフォーマットごとのバージョン等の確認をした上で、我が国におけるリスト案の敲き台を整理した。今後、リスト案を策定していく予定。

| カテゴリ      | 我が国におけるリスト案<br>(標準的フォーマット候補)            |
|-----------|---|
| 文書        | PDF/A-1、PDF/A-2                         |
|           | PDF ver. 1.7以降                          |
|           | Microsoft Word 2007以降 (OOXML)           |
|           | OpenDocument Text ver. 1.2              |
| 表計算       | Microsoft Excel 2007以降 (OOXML)          |
|           | OpenDocument Spreadsheet ver. 1.2       |
| プレゼンテーション | Microsoft PowerPoint 2007以降 (OOXML)     |
|           | OpenDocument Presentation ver. 1.2      |
| 画像        | JPEG 2000                               |
|           | TIFF 6 (Tagged Image File Format)       |
|           | PNG 1.2 (Portable Network Graphics)     |
|           | GIF (Graphics Interchange Format)       |
|           | JPEG (Joint Photographic Experts Group) |
| 音声        | WAVE (Waveform Audio File Format)       |
|           | MP3 (MPEG Layer III Audio Encoding)     |
| 動画        | MPEG-4                                  |
|           | MPEG-2 (MPEG2-PS, MPEG2-TS)             |

- ① 米国立公文書館 (NARA) が移管元機関に対するガイダンス上で、長期保存の観点からのリスク評価が"Preferred" (望ましい) 又は"Acceptable" (受入れ可能) としているフォーマット
- ② 英国立公文書館 (TNA) が移管元機関に対するガイダンス上で、「長期に維持できる」としているフォーマット
- ③ 豪国立公文書館 (NAA) が移管元機関に対するガイダンス上で、長期保存、アクセシビリティ、相互運用性の観点から"Standard" (標準) としているフォーマット

- 今後、リスク評価方法を策定し、標準的フォーマットのリスト案について、技術的動向を加味して、取捨選択を行うこととする（更に加える可能性もあり。）。
- 標準的フォーマット以外のフォーマットを移管しようとする場合、変換の際に、標準的フォーマットであればどれでもよいのか、一層リスクが低いものに変換することとするのか行政機関等と国立公文書館との役割分担も踏まえ検討したい。
- ※ なお、残された課題のうち②の標準的フォーマット以外のものを悉皆的に明らかにすることは難しいため、特にリスト化することはしない（低リスクのフォーマット以外を検出する機能を実装させれば、システム上でその発見もできるものとする。）。

## 4. 今後の進め方

### (1) 令和5年度

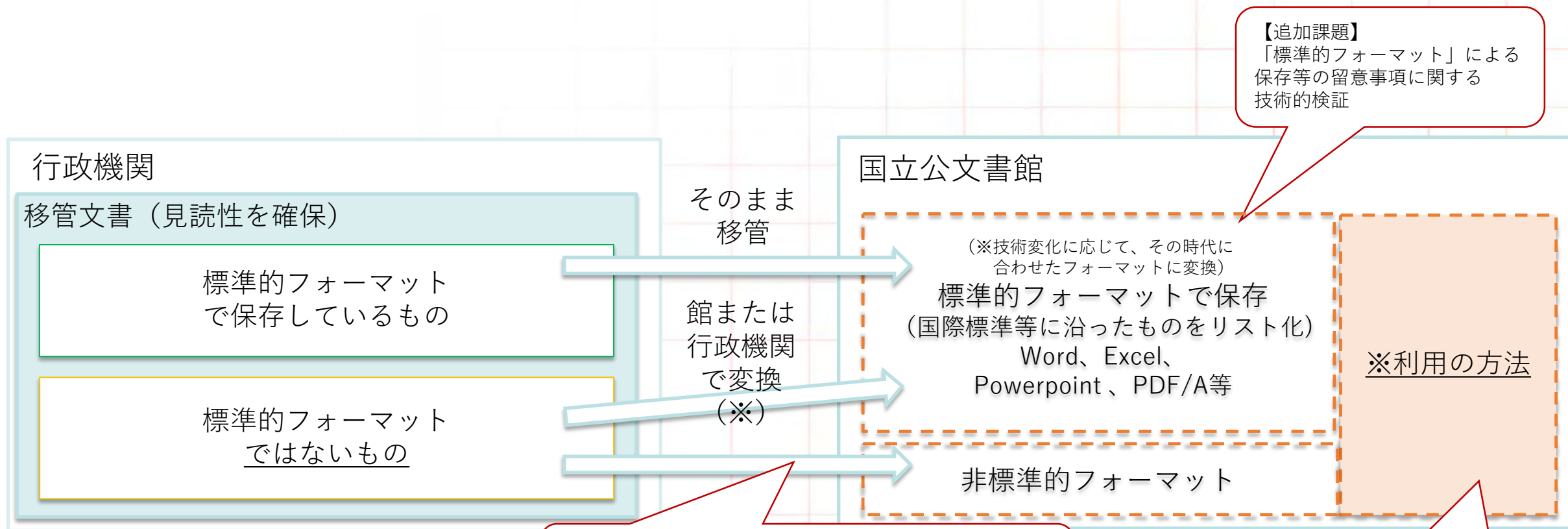
- リスク評価方法案の策定
- 「標準的フォーマット」リスト案の策定
- 「標準的フォーマット」に変換する技術等の実装に向けての検討
- 行政機関等と国立公文書館との役割分担の検討

### (2) 令和6年度

- (追加課題)「標準的フォーマット」で移管、保存したものを利用に供する方法  
→例: 利用制限情報のマスキング方法、電子的に紛れ込んでしまう情報の確認方法等
- (追加課題)「標準的フォーマット」による保存等の留意事項に関する技術的検証  
→特定の機能(マクロ、関数、外字、異体字等)を使用していた場合など、  
「標準的フォーマット」による長期保存の安定性について追加調査が必要



標準的フォーマットのリスト化、リスク評価方法の確立に加え、標準的フォーマットの保存等の運用、利用に供する際の方法について、更に検討を加えていく。



リスクが高いもので技術的に変換できない場合、特定歴史公文書等として移管が必要なものについてはそのまま移管

【追加課題】  
「標準的フォーマット」で移管、保存したものを利用に供する方法

※ 変換の際、標準的フォーマットであればどれでもよいのか、一層リスクが低いものに変換することとするのか、行政機関等と国立公文書館との役割分担の検討も踏まえ判断