令和６年３月28日(木)
総務課デジタル推進室

独立行政法人 国立公文書館
NATIONAL ARCHIVES OF JAPAN

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報告

（令和４年度・令和５年度検討取りまとめ資料）

はじめに

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

目的

「電子公文書等の移管・保存・利用の具体的方法に係る方針」（平成22年3月26日内閣府大臣官房公文書管理課 ）の見直しに向けて、最新の技術情報、変換のための業務負担、利用しやすさ等の課題があることを踏まえ、調査検討を実施

検討の方向性

- 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
- 長期保存に当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

実施内容

令和４年度は、上記の方向性を示した上で、米英豪の公文書館等のリスク評価の考え方やフォーマットへの評価を整理し、「標準的フォーマット候補」を整理した。
令和５年度は、それらの整理結果にもとづき、リスク評価方法案の策定、フォーマット判定や変換の技術的な検証を実施し、「標準的フォーマット案」を策定し、行政機関等と国立公文書館との役割分担について検討を実施した。

本資料について

本資料は、令和４年度の中間報告及び令和５年度に公益社団法人日本文書マネジメント協会に委託した調査及び技術的な検証等の結果報告に基づき取りまとめた資料である。本資料における参照資料は、最終ページの「参考資料」に資料名およびURLを記載している。

1. 調査検討の全体概要

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

令和4年度

【令和４年度の実施内容】

（1）リスク評価の考え方

米豪の国立公文書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整理

リスク評価の考え方
A 仕様が公開されていること
B 世界中で広く使用されていること
C 可逆圧縮等が可能なこと
D メタデータを付与できること
E 技術要素が独立していること
F 特許やライセンス等の制約がないこと
G プロテクションの制約がないこと

（2）標準的フォーマットの候補

米英豪の国立公文書館において長期保存上のリスクが低いとされているフォーマットを抽出し、専門家等に確認をした上で「標準的フォーマット候補」を整理

カテゴリ	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)

令和5年度

【令和５年度の実施内容】

① リスク評価方法案の策定
② 「標準的フォーマット」リスト案の策定
③ 「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
④ 行政機関等と国立公文書館との役割分担の検討

委託により実施

2.令和４年度の調査結果の振り返り　※p.3-6【令和４年度調査】

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

（1）フォーマットのリスク評価の考え方と標準的フォーマットの候補

長期保存にあたってリスクが低いフォーマットの考え方

- 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実施

（※）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

- 米豪の公文書館等の長期保存上の考え方、具体的なファイル・フォーマットごとの技術的な評価情報を把握
- リスク評価の考え方において共通要素を整理

米国国立公文書館（NARA）及び米国議会図書館（LC）	豪州国立公文書館（NAA）	リスク評価の考え方
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（Technical protection mechanisms）		G

※【令和４年度調査】の詳細は「「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」について　中間報告（令和４年度検討取りまとめ資料）」を参照。

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

参考１　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 （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 （十分に 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 （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。 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。 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 （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 （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 （透明性）	基本的なツールで解析可能なフォーマットか。 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。 非可逆圧縮や暗号化がされているか。	Uncompressed （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 （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 （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 （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 （プラットフォームに 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 （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 （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。 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 （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。		

● **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、低リスクと評価するための基本的な条件を整理**

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

● **日本での状況を踏まえた条件の考慮**

- 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）文書作成～利用までの工程と検討すべき専門家等からのご意見

作成 → 整理・保存 → 利用（各府省） → 移管 → 受入れ → 保存 → 利用（館）

● 運用ルールや手順、フォーマットチェック等の仕組みの整備　● 国立公文書館からの実務上の技術情報やノウハウの関係機関等への提供

主なご意見

- 作成：
  - 文字コード、フォント、外字、日本の独自のフォーマット、関数・マクロ、アニメーション等の扱い
  - コーデック取り決め
- 整理・保存：
  - 変化が許容できない文字、重要な文書や情報は、記録として管理・保存
  - 真正性等の確保（電子署名、タイムスタンプ、ハッシュ値等）
- 利用（各府省）：
  - 記録として利用（例：情報公開等）
  - データとして利用（例：機械可読性、アクセシビリティ）
- 移管：
  - フォーマットを標準化する以前（90年代）の受入れへの対応
  - 確認業務の自動化の検討
- 受入れ：
  - 技術情報やノウハウの蓄積
  - リスク評価の運用とフォーマットチェックの仕組み、変換等の技術検討
- 保存：
  - 長期保存フォーマットの運用見直し
  - 変換前のフォーマットの保存ルール
  - マイグレーション
- 利用（館）：
  - 特歴としての一般利用、行政利用(例：提供時のファイル・フォーマット)
  - 利用しやすさの観点からの検討

長期保存にあたってリスクが低いフォーマットの候補

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

- 米英豪の国立公文書館において長期保存上のリスクが低いとされているフォーマット(以下①～③) を抽出し、専門家等に確認をした上で「標準的フォーマット候補」を整理

カテゴリ	我が国におけるリスト案 (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”Preffered”（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

3. 標準的フォーマットリスト案の検討 ※p.7-17【令和５年度委託調査】

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

（1）フォーマットのリスク評価

長期保存のリスク評価方法の検討

- 「リスク評価の考え方」ごとの評価基準、評価点を設定
- 様々なフォーマットのリスク評価を行う「リスク評価基準表」を作成

リスク評価の考え方		評価基準	評価点		備考
A	仕様が公開されていること	仕様が公開されており、仕様を維持するための広く認められた国際的なコミュニティ活動の実態がある（例：ISO、IEC、ITU、W3C）	国際標準	2	＜３段階評価＞ 評価基準に基づき、相対的なレベル設定が可能な場合は、３段階（２点、１点、0点）で評価
		仕様が公開されている	仕様公開	1	
		仕様が公開されていない、又は公開されていても広く認知されていない	非公開	0	
B	世界中で広く使用されていること	ツールが相当数あり、世界中で広く使用されている	世界中で使用	2	
		ツールが少ないが、国内もしくは一部コミュニティで一定程度使用されている	限られた使用	1	
		ツールが少なく、あまり使用されていない。又はツールのサポートが終了しているか、終了が予告されている	使われていない	0	
C	可逆圧縮等が可能なこと	無圧縮を含め圧縮品質を選択できる	圧縮可能（圧縮品質指定可）	2	
		圧縮（可逆圧縮又は非可逆圧縮）ができる	圧縮可能	1	
		圧縮ができない、又は不明	圧縮不可、又は不明	0	
D	メタデータを付与できること	任意のメタデータを付与でき、メタデータ項目が標準に準拠している	標準準拠	2	
		任意のメタデータを付与できる	付与できる	1	
		任意のメタデータを付与できない、又は不明	付与できない	0	
E	技術要素が独立していること	特定の機器、OS等に依存しない	独立	2	＜２段階評価＞ 評価基準に基づき、条件の有無により評価するものは２段階（２点、0点）で評価
		特定の機器、OS等に依存する、又は不明	依存	0	
F	特許やライセンス等の制約がないこと	特許等が存在しないか、使用許諾を得ることなく無償で使用することができる	制約なし	2	
		特許等の主張があり使用許諾が必要	制約あり	0	
G	プロテクションの制約がないこと	【評価対象外】運用リスクであるので、個別のフォーマットでは評価しない。（長期保存上、必要に応じ、暗号化等のプロテクションを解除して保存・利用可能なことが必要）		-	

※【令和５年度委託調査】の詳細は「電子公文書等の適切な保存に係る調査検討　調査検討報告書（令和６年３月　公益社団法人日本文書マネジメント協会）」（以下、「外部委託調査検討報告書」という。）を参照。

フォーマットの技術的特性の整理

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

リスク評価の考え方を踏まえ、「標準的フォーマットの候補」の技術情報を再整理

- **令和４年度調査結果：「標準的フォーマット候補」の技術情報の整理**
  - ✓ **フォーマット毎のバリエーション（バージョンや仕様への準拠の違い）ごとの技術的特性の確認**

（技術情報は「外部委託調査検討報告書」を参照）

①　最新の技術動向を考慮
- **PDF、PDF/Aの最新のバージョン追加：PDF2.0、PDF/A-3、PDF/A-4**

②　フォーマットのさらなる情報収集
- **Microsoft Office（OOXML）には「完全準拠型（strict)」と「移行型（transitional）」が存在**
  →通常設定であれば、旧バージョンのOfficeファイルと互換性のある「移行型」で保存されるため、「移行型」を基本として調査
- **Open Document（ODF）には「フラットXML形式（非圧縮）」と「ZIP形式」が存在**
  →XML構造に２つ表現方法があり、XML言語による一般的な公文書を作成する可能性が低いことから、「ZIP形式」を基本として調査

- **「標準的フォーマット候補」以外で保存している主な移管予定のフォーマットの技術情報の整理**
  - ✓ **当館での受入れ実績から「標準的フォーマット候補」以外の主なフォーマットをピックアップ**
  - ✓ **各フォーマットの技術的特性を確認**

＜「標準的フォーマット候補」以外の主なフォーマット＞
- **一太郎（4、5、6、7、8～12、2004～2023）**
- **旧バージョンのOfficeファイル（Microsoft Office 97-2003）**
- **OASYSファイル**

フォーマットのリスク評価結果

● フォーマットごとに「リスク評価基準表」に基づき評価を実施

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

リスクが低いフォーマット	拡張子	カテゴリ	評価点合計	評価コメント
PDF/A-1	pdf	文書・プレゼンテーション	10	ISO19005-1。PDF1.4をベースに見読性の維持に適さない機能を制限。フォント埋込必須
PDF/A-2			11	ISO19005-2。PDF1.7をベースに見読性の維持に適さない機能を制限。フォント埋込必須。長期保存にあたり安定しているフォーマット
PDF/A-3			11	ISO19005-3。PDF/A-2に加えて、多様なファイル（PDF/Aへの変換前のファイルなど）の埋込みが可能となったが、埋込みファイルの見読性やセキュリティなどに考慮が必要
PDF/A-4			9	ISO19005-4。PDF2.0をベースに見読性の維持に適さない機能を制限している。普及は限定的
PDF1.7			11	ISO32000-1。広く普及している。一般的なソフトで作成可能。フォント埋込が必須ではない
PDF2.0			10	ISO32000-2。最新のPDFフォーマットであるが、普及は限定的
Word 2007以降 (OOXML)	docx	文書	11	ISO-IEC29500。国際標準化されており、一般的なオフィスソフトで作成可能（※１）
OpenDocument Text	odt		10	ISO-IEC26300。無償で提供されているソフトで作成可能だが普及していない（※２）
Excel 2007以降 (OOXML)	xlsx	表計算	11	（※１）と同様
OpenDocument Spreadsheet	ods		10	（※２）と同様
PowerPoint 2007以降 (OOXML)	pptx	プレゼンテーション	11	（※１）と同様
OpenDocument Presentation	odp		10	（※２）と同様
JPEG2000	jp2	画像	12	ISO-IEC15444-1。可逆圧縮で圧縮品質が選択可能。長期保存の目的で広く利用
PNG 1.2	png		10	ISO-IEC15948。可逆圧縮に対応。広く普及
JPEG	jpg		12	ISO-IEC10918。非可逆圧縮だが圧縮品質が選択可能で、広く普及
TIFF 6	tif / tiff		11	仕様は公開されているが、国際標準ではない。可逆圧縮で圧縮品質が選択可能。
GIF	gif		9	仕様は公開されているが、国際標準ではない。可逆圧縮に対応。解像度は低いが、データサイズが小さくなるため普及している
WAVE	wav	音声動画	10	非圧縮方式であり、幅広く利用されているが、国際標準ではない。
MP3	mp3		12	ISO-IEC11172-3。非可逆圧縮だが圧縮品質が選択可能で、高音質。様々なデバイスで再生可能のため広く普及
MPEG-4	mp4		12	ISO-IEC14496-14。非可逆圧縮だが、圧縮品質が選択可能で、高音質。広く普及
MPEG-2	mpg /mpeg		12	ISO-IEC13818-1。非可逆圧縮だが、圧縮品質が選択可能で、高音質。広く普及

候補以外のフォーマット					
	一太郎（4～12／2004～2008）	jtd/jfw/jbw/jaw/jsw	文書	3	公用文のルールに対応しており日本国内で一定程度利用されているが、仕様が非公開であり、将来的なファイルの見読性を確保に課題があり
	OASYS	-		0	ツール及びサポートが終了しているほか、仕様が非公開であり、バージョンによって見読性に問題あり
	Word 97-2003 Excel 97-2003 PowerPoint 97-2003	doc xls ppt	文書 表計算 プレゼン	6	Microsoftの後継製品で「互換モード」で開くことができる。Microsoftから仕様が公開されているが、国際標準でない

フォーマットごとの評価点詳細

（詳細は「外部委託調査検討報告書」を参照）

リスク評価の考え方	A 仕様が 公開されていること	B 世界中で広く 使用されていること	C 可逆圧縮等が 可能なこと	D メタデータを 付与できること	E 技術要素が 独立していること	F 特許やライセンス等 の制約がないこと	評価点
評価項目	2:国際標準	2:世界中で使用	2:圧縮品質指定可	2:形式が標準準拠	2:独立	2:制約なし	
	1:仕様公開	1:限られた使用	1:圧縮可能	1:任意要素可			
	0:非公開	0:使われていない	0:圧縮不可・不明	0:固定・不明	0:依存・不明	0:制約あり・不明	
PDF/A-1	2	2	1	1	2	2	10
PDF/A-2	2	2	1	2	2	2	11
PDF/A-3	2	2	1	2	2	2	11
PDF/A-4	2	0	1	2	2	2	9
PDF1.7	2	2	1	2	2	2	11
PDF2.0	2	1	1	2	2	2	10
Word 2007以降（OOXML）	2	2	1	2	2	2	11
OpenDocument Text	2	1	1	2	2	2	10
Excel 2007以降（OOXML）	2	2	1	2	2	2	11
OpenDocument Spreadsheet	2	1	1	2	2	2	10
PowerPoint 2007以降（OOXML）	2	2	1	2	2	2	11
OpenDocument Presentation	2	1	1	2	2	2	10
JPEG2000	2	2	2	2	2	2	12
PNG 1.2	2	2	1	1	2	2	10
JPEG	2	2	2	2	2	2	12
TIFF 6	1	2	2	2	2	2	11
GIF	1	2	1	1	2	2	9
WAVE	1	2	2	1	2	2	10
MP3	2	2	2	2	2	2	12
MPEG-4	2	2	2	2	2	2	12
MPEG-2	2	2	2	2	2	2	12
一太郎 4～12/2004～2008	0	1	0	0	0	2	3
OASYS	0	0	0	0	0	0	0
Microsoft Word 97-2003	1	0	1	0	2	2	6
Microsoft Excel 97-2003	1	0	1	0	2	2	6
Microsoft PowerPoint 97-2003	1	0	1	0	2	2	6

（2）「標準的フォーマット」に変換する技術等の実装に向けた検証

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

文書作成

①文書作成時の確認

○ **一般的なソフトウェアによる「標準的フォーマット候補」での文書作成等の点検、留意事項を把握**
Word（OOXML）等は、一般的な業務環境で作成、保存が可能。

②フォーマット判定機能

（試験環境等は「外部委託調査検討報告書」を参照）

保存・整理

○ **「標準的フォーマット候補」は、英国国立公文書館が公開しているフォーマット判定ツール（DROID）をはじめ複数のツールで判定できることを確認（「標準的フォーマット候補」以外を自動検出することも可能）**
○ **幅広いフォーマットを判定可能で、実装コストも不要な「DROID」が、フォーマット自動判定ツールとして適当**

③フォーマット変換・検証機能

（試験環境及び変換方法等は「外部委託調査検討報告書」を参照）

○ **各省庁で「標準的フォーマット」以外を保有する場合、見読性を確保するため「標準的フォーマット」へ変換しておくことが望ましい。そのため、「標準的フォーマット候補」以外を、「標準的フォーマット候補」への自動変換・自動検証する試験を実施**
※試験環境における変換・検証結果は次項の【表１】を参照

- **試験環境では「標準的フォーマット候補」への個別変換の自動化は可能（※メーカーで推奨・保証するものではない。）**
- **OASYSは変換不可。複数シートの一太郎は手動変換のみ可**
- **変換結果の検証機能については、**
  変換前後の文字（テキストの比較）、表示内容（レイアウト）の比較ツールにより一定程度は可能
  ✓ 変換・検証の精度や信頼性については更なる研究が必要
  ✓ システム化に向けて、必要となる処理能力や実装コストの確認が必要

【表１】フォーマット変換及び変換結果の一覧

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

凡例　○：変換可能、△：条件あり、×：変換不可

変換元フォーマット	変換先の標準的フォーマット候補	①変換の可否			②変換結果の検証		③推奨する変換先
		手動変換	自動化	留意事項	自動検証	変換後の検証結果	
一太郎（4～12，2004～2008） 拡張子：jtd, jfw, jbw, jaw, jsw	PDF/A-1～4	○	△	複数シートの場合、自動化ができない	一部可	注釈、変更履歴、隠し文字は設定によって出力されない	PDF/A-2
	PDF1.7、2.0	○	△	複数シートの場合、自動化ができない			
	Word 2007以降（OOXML）	×	×				
Word 97-2003 拡張子：doc	PDF/A-1～4	○	○		一部可	注釈、変更履歴、隠し文字は設定によって出力されない	固定化：PDF/A-2 再利用：OOXML
	PDF1.7、2.0	○	○				
	Word 2007以降（OOXML）	○	○			微細だがレイアウトがずれる	
Excel 97-2003 拡張子：xls	PDF/A-1～4	○	○	印刷範囲を正しく設定する必要あり		注釈、変更履歴、隠し文字は設定によって出力されない	固定化：PDF/A-2 再利用：OOXML
	PDF1.7、2.0	○	○	印刷範囲を正しく設定する必要あり			
	Excel 2007以降（OOXML）	○	○			差異なし	
Excel 2007（マクロ有効ブック） 拡張子：xlsm	PDF/A-1～4	△	△	マクロは移行しない		・注釈、変更履歴、隠し文字は設定によって出力されない ・マクロは移行されない	固定化：PDF/A-2 再利用：OOXML
	PDF1.7、2.0	△	△	マクロは移行しない			
	Excel 2007以降（OOXML）	△	△	マクロは移行しない、或いは移行されてもマクロの動作保証をしない		マクロ機能は移行されるが動作保証されない	
PowerPoint 97-2003 拡張子：ppt	PDF/A-1～4	○	○			・注釈は設定によって出力されない ・アニメーションは最終状態となる	固定化：PDF/A-2 再利用：OOXML
	PDF1.7、2.0	○	○				
	PowerPoint 2007以降（OOXML）	○	○			差異なし	
OASYS	—	×	×	ソフトウェアが入手できないため（現在販売終了、サポートも終了予定）、変換は不可。			

（3）標準的フォーマット案の検討

内閣府における「標準的フォーマット」の指定に向けて、リスクが低いフォーマットである「標準的フォーマットの候補」の中から以下の項目を条件に「標準的フォーマット案」として整理

① 国際標準（ISO）化されていること

→国際標準は、国際的な組織で検討・承認されているものであり、定期的（5年ごと）に技術や仕様の点検・見直しが実施され、継続的にフォローされている。国際標準化されたものはフォーマットの技術的な情報が利用可能な形で公開されており、見読性を確保するために必要なソフトウェアなどの環境を構築することが可能。

② 行政機関からの移管実績がある程度あること（行政文書の保存フォーマットとして広く利用されていること）

→行政機関からの移管実績が一定程度あり、行政文書の保存フォーマットとして広く利用されていると判断されるもの。

※国立公文書館で受け入れた電子公文書等のうち、「電子公文書等の移管・保存・利用システム」（以下、「電子公文書等システム」という。）で受入れ、保存を実施した子ファイルのフォーマット概況は次頁の【参考２】のとおり

標準的フォーマット案一覧

ファイル・フォーマット類型	「標準的フォーマット案」	国際標準
文書作成・ 表計算・ プレゼンテーション	PDF/A-1、A-2 PDF1.7 Word2007以降 Excel2007以降 PowerPoint2007以降	ISO 19005 ISO 32000 ISO-IEC 29500 ISO-IEC 29500 ISO-IEC 29500
画像	JPEG2000 PNG JPEG	ISO-IEC 15444 ISO-IEC 15948 ISO-IEC 10918
音声・動画	MP3 MPEG2 MPEG4	ISO-IEC 11172 ISO-IEC 13818 ISO-IEC 14496

※標準的フォーマット案の整理の詳細は【表２】を参照

参考２　国立公文書館で受け入れた電子ファイルのフォーマット概況（電子公文書等システムで保存しているものに限る）

フォーマット種類	概数	全体に占める割合	令和５年度受入れ分（概数）	令和５年度受入れ分の割合
PDF 1.7	2,600	約14％	1,500	約19％
Word 2007以降	23,000		14,000	
Excel 2007以降	8,600		3,500	
PowerPoint 2007以降	2,000		1,400	
JPEG	16,600		400	
JPEG2000	30		0	
PNG	130		30	
MP3	240		20	
MPEG	20		10	
GIF	7,000	約86％	40	約81％
TIFF	440		35	
WAVE	15		0	
一太郎	12,900		4,100	
OASYS	360		90	
PDF 1.0-1.6　※	146,000		45,000	
Word 95	160		130	
Word 97-2003	49,800		19,700	
Excel 95	2,200		200	
Excel 97-2003	21,000		5,700	
PowerPoint 97-2003	3,000		2,000	
その他	76,500		12,000	
合計	**37万件**	**100%**	**11万件**	**100%**

「標準的フォーマット」案

「標準的フォーマット」案以外

※PDF1.4にはPDF/A-1が含まれている可能性がある。

【表２】 文書/表計算/プレゼンテーション

フォーマット名	拡張子	標準的フォーマット案	国際標準	標準的フォーマット案の補足説明
PDF/A-1	pdf	●	ISO19005-1	従来の課長通知２－６において、「長期保存フォーマット」として定められていたフォーマットである。 国際標準化されており、フォントの埋め込みをはじめ、文書の長期保存に必要な文字及び色の再現性が担保されているため、文書が環境や時間によって変化することを防ぐことができる。
PDF/A-2	pdf	●	ISO19005-2	平成23年（2011年）に国際標準化されたPDF/Aの新たなバージョンであり、PDF/A-1にはなかった機能（透明効果やレイヤー、PDF/A準拠のファイル埋め込み等）が追加され、より一般的な文書を対象とする長期保存が可能となっている。 PDF/A-1がベースとしているPDF1.4は国際標準ではなかったが、PDF/A-2がベースとしているPDF1.7はISO-32000-1として国際標準化されている現在の主流であり、より一般的で、長期保存にあたり安定したフォーマットとなっている。
PDF/A-3	pdf		ISO19005-3	平成24年（2012年）に国際標準化されたPDF/Aの新たなバージョンであり、PDF/A-2との違いは、多種多様なファイルを埋め込むことが可能になったことである。ただし、長期保存に適さないファイルも埋込可能となっており、埋込みファイルの見読性やセキュリティに課題がある。
PDF/A-4	pdf		ISO19005-4	令和2年（2020年）に国際標準化されたPDF/Aの新たなバージョンであり、最新のPDF技術（PDF2.0）に基づくが、ツールの提供が進んでおらず、利用普及はこれから。
PDF1.7	pdf	●	ISO32000-1	平成20年（2008年）に国際標準化され、現在使用されているPDFの主流として広く普及している。フォントの埋め込みが必須とされていないため、文字の再現性については課題があるものの、PDF/Aに比べて著しく保存上のリスクが高いわけではない。
PDF2.0	pdf		ISO32000-2	国際標準化された最新のPDFフォーマットであるが、現時点ではフォーマット及びそれを作成するツール自体が普及していない。PDF2.0をサポートしないPDFソフトでは開くことができない場合がある課題がある。
Word 2007以降 (OOXML)	docx	●	ISO-IEC29500	平成20年（2008年）に国際標準化されており、一般的なオフィスソフトで作成可能で広く普及している。定期的に仕様が更新されているが、旧バージョンとの互換性も高く、長期保存にあたってのリスクも高くないと評価した。（※１）
OpenDocument Text	odf		ISO-IEC26300	平成18年（2006年）に国際標準化されており、無償で提供されているソフトで作成可能だが、行政文書の保存フォーマットとして普及していない。（ヨーロッパでは普及している）（※２）
Excel 2007以降 (OOXML)	xlsx	●	ISO-IEC29500	（※１）と同様
OpenDocument Spreadsheet	ods		ISO-IEC26300	（※２）と同様
PowerPoint 2007以降 (OOXML)	pptx	●	SO-IEC29500	（※１）と同様
OpenDocument Presentation	odp		ISO-IEC26300	（※２）と同様

画像/音声/動画

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

フォーマット名	拡張子	標準的フォーマット案	国際標準	標準的フォーマット案の補足説明
JPEG2000	jp2	●	ISO-IEC15444-1	従来の課長通知２－６において、「長期保存フォーマット」として定められていたフォーマットである。 国際標準化されており、可逆圧縮で用途に応じて圧縮効率と品質を選択することが可能であり、長期保存の目的でも広く利用されている。
PNG 1.2	png	●	ISO-IEC15948	国際標準化されている。圧縮品質の指定ができず、メタデータの形式が標準化されていないため評価点は高くないが、可逆圧縮に対応しているので保存を繰り返しても画質が悪くならずに高品質の画像が作成できる。ライセンスを取得せずに多くのツールで表示、編集することが可能であり、WindowsOS標準の保存形式であるなど、広く利用されている。
JPEG	jpg	●	ISO-IEC10918	国際標準化され、非可逆圧縮であるが圧縮品質を選択することができ、ほとんどのブラウザー、ソフトウェア、アプリケーションで利用できる、一般的な画像ファイル形式である。
TIFF	tif|tiff			可逆圧縮で品質が選択可能であり、仕様は公開されているため、広く普及している。ただし、仕様を維持するための広く認められた国際的なコミュニティ活動における国際標準ではなく、自由度が高い反面、長期保存上の観点から将来のファイルの再現性やバージョンアップ等があった際の互換性を考慮する必要がある。
GIF	gif			広く利用されているが、日本では2004年まで特許が存在した。現在では特許の有効期限が切れており、仕様は公開されているが、国際的なコミュニティの活動という面では他のフォーマットと比べ活発ではない。国際標準ではなく、長期保存上の観点から将来のファイルの再現性やバージョンアップ等があった際の互換性を考慮する必要がある。
WAVE	wav			非圧縮方式であり、国立国会図書館においても保存用途に利用されているが、仕様を維持するための広く認められた国際的なコミュニティ活動における国際標準ではなく、長期保存上の観点から将来のファイルの再現性やバージョンアップ等があった際の互換性を考慮する必要がある。
MP3	mp3	●	ISO-IEC11172-3	国際標準化されており、非可逆圧縮だが圧縮品質が選択可能で、高音質で保存できることから広く利用されている。多くのプレイヤーやデバイスが対応している。
MPEG-4	mp4	●	ISO-IEC14496	国際標準化されており、非可逆圧縮だが、圧縮品質が選択可能で、高画質・高音質で保存することができる。また、国立国会図書館等においても、保存用途に利用されている。ストリーミングやダウンロードなどのインターネット上のビデオ配信に適した形式として広く普及しており、多くのプラットフォームやデバイスが対応している。
MPEG-2	mog|mpeg	●	ISO-IEC13818	国際標準化されており、非可逆圧縮だが、圧縮品質が選択可能で、高画質・高音質で保存することができる。デジタルテレビ放送やDVDなどに適したファイル形式として広く普及しており、多くのプラットフォームやデバイスが対応している。

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

その他（「標準的フォーマット案」以外のフォーマット）

国立公文書館への一定数の移管実績があるフォーマットのうち、「標準的フォーマット」の候補に該当しないフォーマットについても改めてリスク評価を実施。

フォーマット名	拡張子	標準的フォーマット案	国際標準	補足説明
Word 97-2003	doc			Microsoftから仕様が公開されており、Microsoftの後継製品（Office2007以降のバージョン）で「互換モード」として開くことができるが、国際標準化されていない。
一太郎	jtd/jfw/jaw/jsw			公用文のルールに対応しており日本国内で一定程度利用されているが、仕様が非公開であり、将来的なファイルの見読性を確保に課題がある。
OASYS	oas等			日本国内で利用されていたが、ツールについては個人向けは2020年9月30日、法人向けは2021年5月31日に販売が終了し、サポートについても、個人向けは2023年9月30日、法人向けは2024年5月31日に終了している。仕様が非公開であり、見読性に課題が生じている。
Excel 97-2003	xls			Microsoftから仕様が公開されており、Microsoftの後継製品（Office2007以降のバージョン）で「互換モード」として開くことができるが、国際標準化されていない。
PowerPoint 97-2003	ppt			Microsoftから仕様が公開されており、Microsoftの後継製品（Office2007以降のバージョン）で「互換モード」として開くことができるが、国際標準化されていない。

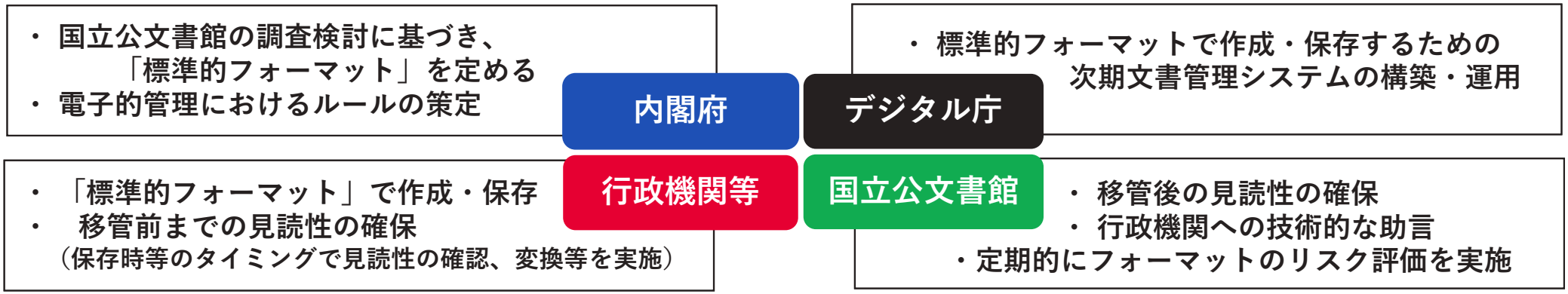
4．標準的フォーマット案による運用について

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

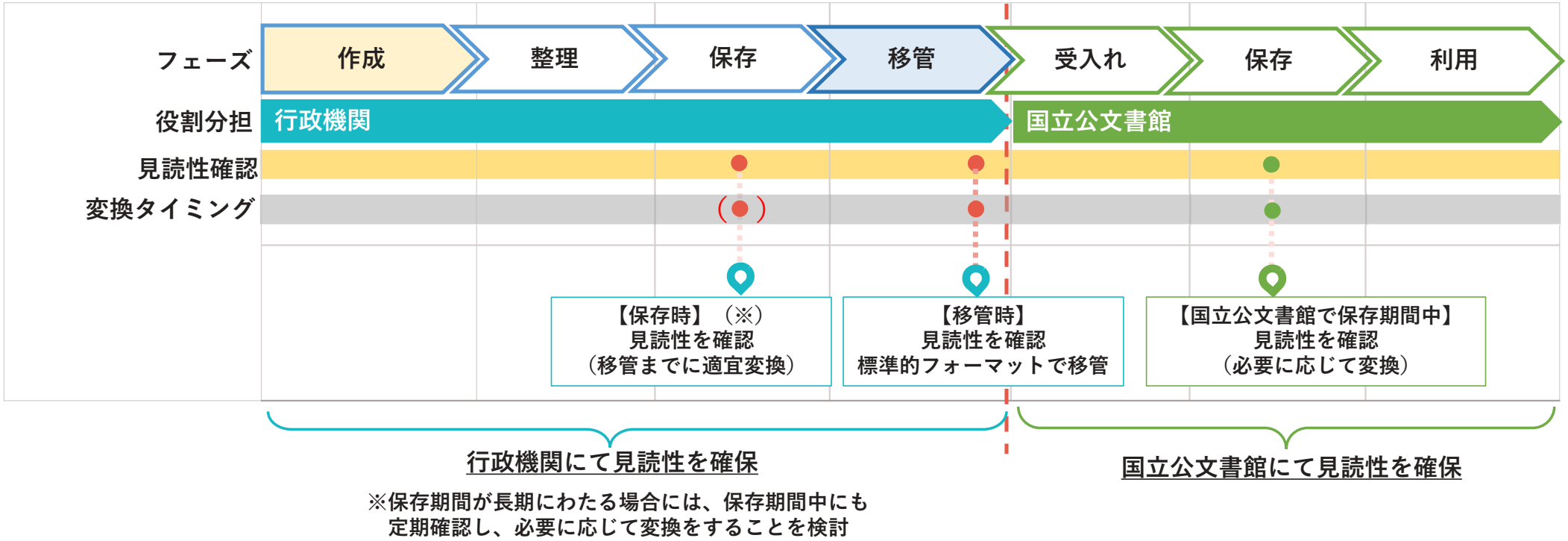
（1）行政機関等と国立公文書館との役割分担の検討

○基本的な考え方

- 行政機関等は、長期に保存するものは、原則「標準的フォーマット」で作成、保存する
- 行政機関等は、「標準的フォーマット」については、そのまま移管
- 行政機関等は、「標準的フォーマット」以外については、「標準的フォーマット」に変換して移管
- 国立公文書館は、「標準的フォーマット」への変換について、行政機関等へ技術的な助言を実施
- 国立公文書館は、定期的にフォーマットのリスク評価を実施



○見読性の確認と「標準的フォーマット」への変換の運用



※保存期間が長期にわたる場合には、保存期間中にも
定期確認し、必要に応じて変換をすることを検討

○ 行政機関では、移管前までの見読性を確保

保存時、移管時に見読性確認を行い、「標準的フォーマット」以外やパスワード保護等がないか確認する

移管までに、「標準的フォーマット」以外は標準的フォーマットに変換し、パスワード等は解除する

○ 国立公文書館では、移管後の見読性を確保

国立公文書館で永久保存する期間中、見読性確認を行い、必要に応じて変換する

（２）技術等の確認結果からの論点

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

フォーマット判定について

論点１　新規作成する電子公文書については、
記録領域に保存時に「標準的フォーマット」かどうかをチェックする機能を設けてはどうか。

論点２　行政機関等での見読性確保の観点から、既に保存している「標準的フォーマット」以外の電子公文書について把握する機能を設けてはどうか。
（※不要なファイル：一時保存ファイルやジャンクファイルの点検含む）

論点３　見読性の確認機能（再生確認）を設けてはどうか。

【不要なファイルの例】
～$公文書管理.docx
○○○.sys
._公文書管理.xlsx
公文書管理.$$$

フォーマット変換・検証について

論点４　自動変換・検証機能についてシステム化及び技術の確立にむけた調査研究が必要。

論点５　今後自動変換を実現した場合も変換前後において文書の本質的な内容が失われていないかの確認する必要があると考えられる。

「標準的フォーマット」以外のフォーマットへの対応

論点６　変換できないフォーマットの保存・利用のため、引き続き継続的に検討が必要
（仮想的な再生環境等）。

5. 今後について

独立行政法人 国立公文書館 NATIONAL ARCHIVES OF JAPAN

○「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の検討
→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。

○「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
→「標準的フォーマット」であっても、
特定の機能（環境依存文字、特殊フォント、電子署名等）を使用していた場合の長期保存の安定性について調査。

○「標準的フォーマット」以外のフォーマットの保存及び利用に向けた調査研究の継続
→「標準的フォーマット」以外のフォーマットの自動変換及び自動検証、仮想的な再生環境に関する調査研究。

○ 「標準的フォーマット」の普及・技術的支援
→「標準的フォーマット」への変換方法の紹介、電子公文書の作成・保存・利用に資する情報をまとめたガイドブックの作成。

参考資料

- 「「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４年11月　独立行政法人国立公文書館）
https://www.archives.go.jp/about/report/pdf/chousa_2022_11.pdf
- 「「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」について　中間報告（令和４年度検討取りまとめ資料）」
（令和5年3月29日　国立公文書館総務課デジタル推進室）
https://www.archives.go.jp/about/report/pdf/chousa_2023_03.pdf
- 「電子公文書等の長期保存に係る調査検討報告書」（令和６年３月　公益社団法人日本文書マネジメント協会）
https://www.archives.go.jp/about/report/pdf/chousa_2024_03_01.pdf

本調査検討に当たって、多くの方々に御協力いただきました。深く感謝申し上げます。