

令和5年3月29日(水)
総務課デジタル推進室



「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)

はじめに

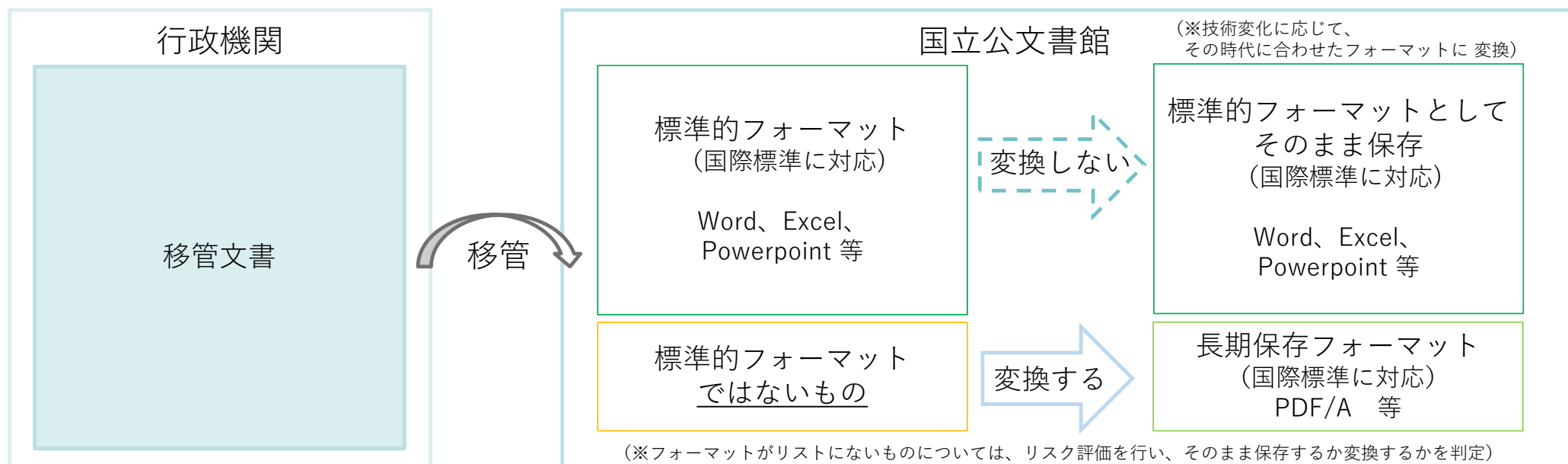
◎令和4年度の「電子公文書等の長期保存フォーマットを含む長期保存に関する調査検討」については、11月に状況を報告（令和4年11月9日 第99回公文書管理委員会）

(1) 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）（以下「現行方針」）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

(2) 検討の方向性

移管文書を一律一定に長期保存フォーマットへ変換せず、長期保存にあたってリスクが低い場合は「標準的フォーマット」（仮称）であれば、そのまま保存、リスクが高い場合は長期保存フォーマットへ変換する方向で検討



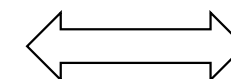
(3) 検討課題

- 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
- 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

(「第99回内閣府公文書管理委員会」より)

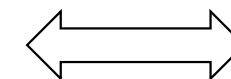
1. 調査検討の概要 (R4)

(1) 米英豪の公文書館等のファイル・フォーマットの考え方、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

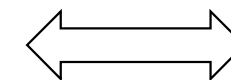


専門家等ヒアリング(表1)

(2) 日本ならではのローカル要素の確認(日本独自フォーマットや文字・フォント等現実的に対応可能な仕組み)



(3) 技術的動向を踏まえたフォーマット



(4) 検討課題①②③を踏まえた整理
フォーマットを特定・評価するための考え方、リスト案等を提示

※検討課題④⑤⑥については、引き続き調査検討を継続
(変換に関する技術的検証、役割分担や運用上における技術的論点の整理等)

表 1

ヒアリング先 ※記録管理やPDF等の専門家、長期保存に取り組まれている専門機関

(専門家、専門団体等)

ヒアリング先	日時（初回説明含む）	備考
上原 哲太郎 氏 (立命館大学教授)	令和4年7月、令和5年3月	内閣府公文書管理委員会委員（令和4年7月～）
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)	令和5年1月	文字情報等の専門家
杉本 重雄 氏 (筑波大学名誉教授)	令和5年3月	電子媒体による公文書等の管理・移管・保存のあり方に関する研究会委員（平成17年度～平成18年度）、内閣府公文書管理委員会委員（平成22年度～平成26年度）
溝上 卓也氏 (株式会社デザイン DX支援事業本部フェロー)	令和4年12月、令和5年1月	記録管理の専門家
公益社団法人 日本文書情報マネジメント協会(JIIMA) ・木村 道弘氏 (JIIMA 特別研究員) ・名護屋 豊氏 (ウイングアーク1st株式会社 Business Document 事業部執行役員兼副事業部長) ・相馬 淳人氏 (株式会社エリジオン 取締役副社長 CTO)	令和4年11月、12月	標準化、記録管理の専門団体 (電子記録保存、PDF、CAD等の専門家)
ARMA International 東京支部 ・宮崎 一哉氏 (ARMA International 東京支部理事)	令和4年12月、令和5年1月	記録管理の専門団体

(専門機関)

ヒアリング先	日時（初回説明含む）	備考
国立国会図書館 電子情報部電子情報企画課・関西館電子図書館課	令和4年12月、令和5年1月	デジタル資料の保存及び利用に関する機関
国立映画アーカイブ 学芸課映画室、学芸課展示・資料室	令和4年12月、令和5年1月	デジタル資料の保存及び利用に関する機関

本調査に関する全般的コメント

- 長期保存フォーマットの見直しに関する方向性（「標準的フォーマット」を移管時のまま保存することなど）は妥当。
- 検討にあたっては、「記録」と「記録」以外の取扱い、作成・保存・利用の一連の流れを踏まえた検討、例外も含む現実的なスキームとしておくべき。

専門家等のご意見例

- 長期保存の方針見直しの方向性について、国際動向からみて妥当である。
- リスクが低いとされるフォーマットでも、長期の保存で考慮すべき事項があるので注意。
- 「標準的フォーマット」について
 - 電子記録の作成、整理、保存、利用の一連の流れで整理が必要。
 - 一定の方向性を示しながら、それに当てはまらない場合も想定しておくことが必要。
 - 長期に利用可能性が見込まれるフォーマットの把握と選択が重要。
 - 移管が想定されるものを基本に検討してはどうか。
- 技術進歩が速い分野は、ツールやフォーマットを固定せず、機能・利用の場面に応じ使い分けている。
- フォーマットの標準化にあたっては、作成や保存の目的を踏まえて検討する必要。
（責任を持つ「記録（＝証拠）」と記録以外のファイルの区別など）

2.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

検討方法

- (1) 米豪の公文書館等の長期保存上の考え方（表2）、
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
- (2) 日本ならではのローカル要素の確認
(日本独自フォーマットや文字・フォント等現実的に対応可能な仕組み)

ファイル・フォーマットのリスク評価観点（案）を整理

共通要素

- ① 仕様が公開されていること
- ② 普及していること
- ③ 元に戻せること（可逆圧縮・復号化等）
- ④ メタデータに対応していること
- ⑤ 技術要素が独立していること
- ⑥ 制度面の制約がないこと
- ⑦ 技術面の制約がないこと

ローカル要素（日本ならではの要素）の考慮

- ① 実行可能性（現実的要素）
(処理できるか、保有する割合等の実績)
- ② 標準化活動への関与
(専門家等と密に連携取れることが大切)
- ③ 文字に関する取扱い
(文字コード、フォント等)

表 2

(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 (情報開示性)	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	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	世界中で広く使用され、サポートされていること。
		Supported	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	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。もし該当する場合、そのフォーマットは、クローズドキャプション (CC) のような ADA 法に準拠したアクセシビリティオプションがあるか。	Metadata friendly	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	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	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 (技術保護の仕組み)	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。		

米国 NARA (ローカル要素)	
Need (必要性)	そのフォーマット仕様が、最初に作成されたのはいつか。
	そのフォーマットの仕様が、最後に更新されたのはいつか。
	NARA ERA (電子公文書等の保存システム) で所蔵する 20 億ファイルにおける割合
Prevalence (保有率)	NARA が保有するファイル・フォーマット全体に占める割合で測定したフォーマットの採用レベル
Feasibility (実現可能性)	変換ツールがあり、NARA はフォーマット変換することができる

豪州 NAA リスクの高いフォーマット
<ul style="list-style-type: none"> • フォーマットの文書化が不十分であること • 特許やライセンスの制約があること • フォーマットを実装するソフトウェアの選択肢が限られていること • フォーマットがサポートされていないこと • 国際的なデジタル保存コミュニティが危険であると判断していること

長期保存上のリスクについての考え方

ファイル・フォーマットの評価観点を参考に、それらを満足できない条件が揃う場合に、長期保存上のリスク要因と考える（「豪州NAAリスクの高いフォーマット」の例）

- ・ 技術情報が公開されていない
- ・ 特許やライセンスの制約がある
- ・ フォーマットを実装するソフトウェアが限られている
- ・ フォーマットがサポートされていない
- ・ フォーマットの危険性に関する専門家らの合意

専門家等のご意見

- ・ 長期保存においては、技術が陳腐化する前に、次のフォーマットへ移せることがポイント。一般的には2世代前のバージョンまでは互換性保証していることが多い。
- ・ リスクの高いフォーマットの検出方法として、複数のツールで変換できないものは、リスクが高いと判断できる。
- ・ フォーマットにおけるリスクは、技術がベンダに囲い込みされていること。オフィスソフトは標準化されているが、企業動向をウォッチしておくこと。
- ・ 電子は陳腐化しやすいので、30年後の移管を考慮した技術的条件を確保しておく必要。
- ・ ファイル・フォーマットのリスク評価について、「誰が行うか」が重要。
- ・ 長期にわたる保存においては、結局どこかで変換等を伴う作業が必要となる可能性があり、上流工程で基本的な事項として対応してもらうことが望ましい。

【リスクとは？】

フォーマットの見読性が失われること、次のフォーマットへマイグレーションできないこと。

【いつ？】

技術が陳腐化する前。複数の利用可能なツールで変換できない時。

3.長期保存にあたってリスクが低いフォーマットのリスト化

- (1) 英米豪国立公文書館の受入れ可能なもの、リスク評価情報をもとに整理
3館における長期保存上のリスク評価等に関する情報をもとに、共通で取り扱われているものを基本とし、専門家等からの意見を踏まえ整理 (表3)

(参考) 英米豪国立公文書館の取扱状況・リスク評価等

(例)

No	A.カテゴリ (NAJ)	フォーマット名	拡張子	①NARA "Preffered"	①NARA "Acceptable"	②TNA "Sustain over the long term"	③NAA "Standard"	リスク点数 (NARA)	リスク評価 (NARA)	カテゴリ (NARA)	アクション (NARA)
5	文書	Microsoft Word for Windows 2007-onwards (OOXML)	docx	○	○	○	○	22.00	Low Risk	Textual and Word Processing	Retain

※Preffered 又は Acceptableのいずれか○

※ 米英豪のガイダンス
(最終参照日：令和4年12月)

※米NARAが公表しているフォーマットの
リスク点数、評価、アクションプラン
(令和4年8月時点)

表 3

長期保存にあたってリスクが低いフォーマット (= 「標準的フォーマット」 の候補)

カテゴリ	フォーマット	拡張子	コーデック	備考
文書	PDF/A-1、PDF/A-2	pdf		ISO 19005-1:2005 (※PDF/A-1) ISO 19005-2:2011 (※PDF/A-2)
	PDF ver. 1.7以降	pdf		ISO 32000-1:2008 (※PDF 1.7) ISO 32000-2:2020 (※PDF 2.0)
	Microsoft Word 2007以降 (OOXML)	docx		ISO/IEC 29500-1:2016
	OpenDocument Text ver. 1.2	odt, fodt		ISO/IEC 26300:2015
表計算	Microsoft Excel 2007以降 (OOXML)	xlsx		ISO/IEC 29500-1:2016
	OpenDocument Spreadsheet ver. 1.2	ods, fods		ISO/IEC 26300:2015
プレゼンテーション	Microsoft PowerPoint 2007以降 (OOXML)	pptx		ISO/IEC 29500-1:2016
	OpenDocument Presentation ver. 1.2	odp, fodp		ISO/IEC 26300:2015
画像	JPEG 2000	jp2		ISO/IEC 15444-1:2004
	TIFF 6 (Tagged Image File Format)	tif, tiff		
	PNG 1.2 (Portable Network Graphics)	png		ISO/IEC 15948:2004
	GIF (Graphics Interchange Format)	gif		
	JPEG (Joint Photographic Experts Group)	jpg		ISO/IEC 15444-1:2004
音声	WAVE (Waveform Audio File Format)	wav	リニアPCM (LPCM)	※BWF 推奨
	MP3 (MPEG Layer III Audio Encoding)	mp3, mpeg3等	MP3	ISO/IEC 11172-3:1993, ISO/IEC 13818-3:1998
	AAC (Advanced Audio Coding MPEG-4 Low Complexity Object)	aac, mp4等	AAC	ISO/IEC 13818-7, ISO/IEC 14496-3(※AAC)
動画	MPEG-4	mp4, mpa	(映像) H.264、H.265等 (音声) AAC 等	ISO 14496-10:2003 (※H.264/MPEG-4 AVC), ISO/IEC 23008-2 HEVC (H.265) 等
	MPEG-2 (MPEG2-PS, MPEG2-TS)	mpg, mpeg	用途に合わせ選択	ISO/IEC 13818-1 等
その他	※ CADやテキストファイルのほかに種類があるか、ファイル・フォーマットの利用ニーズを踏まえた調査検討を継続			

3-1. 文書/表計算/プレゼンテーション

長期保存にあたってリスクが低いフォーマット (= 「標準的フォーマット」 の候補)

カテゴリ	フォーマット名	拡張子	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
文書	PDF/A-1、PDF/A-2	pdf	○	○	○	○	文字情報はUnicodeが望ましい。PDF/A-1とPDF/A-2が適当。(PDF/A-3のマルチファイル機能は使わない方がよい。電子署名やタイムスタンプを用いるのであればPDF/A-4の必要。
文書	PDF ver. 1.7以降	pdf		○	○	○	バージョンは1.7以降を基本。ファイル内の情報の持ち方については標準化されておらずページの追加等が可能。長期保存の場合は、PDFやPDF/Aが適当。
文書	Microsoft Word 2007以降 (OOXML)	docx		○	○	○	word内のフォント、図に含まれる文字を、長期保存することが必須の場合は、PDFやPDF/Aが適当。
文書	OpenDocument ver. Text 1.2	odt, fodt	○		○	○	
表計算	Microsoft Excel 2007以降 (OOXML)	xlsx		○	○	○	関数。マクロ等、プログラム機能の部分については、将来的な動作が保証できず、保存できない。値を長期保存したい場合はPDFやPDF/Aが適当。
表計算	OpenDocument Spreadsheet ver. 1.2	ods, fods	○		○	○	
プレゼンテーション	Microsoft PowerPoint 2007以降 (OOXML)	pptx		○	○	○	アニメーション等の動的な要素はバージョンが変わると動作しないことが多い。長期保存する場合は、PDFやPDF/Aが適当。
プレゼンテーション	OpenDocument Presentation ver. 1.2	odp, fodp	○		○	○	

- オフィス系文書は、そのまま保存する傾向。Office Open XML (OOXML)推奨。
- ISO対応でも、中身に標準化していない要素 (Word、Excel等)、文字コードの定義がない (PDF、PDF/A等) など留意。
- マクロは、ウィルス混入やマクロを利用しないと記録が再生できないことなど、管理上のリスクが大きい。保存が必要な場合は、「プログラム」として、設計書 (理解に必要な情報) と合わせて保存するなどが望ましい。
- 複数のデータを1つにまとめて管理するマルチファイルの機能は使用せず、文書間の関係は、メタデータで管理することが現実的。機能的な要素のほか、文書内の外部・内部参照にも留意。

3-2. 画像

長期保存にあたってリスクが低いフォーマット (= 「標準的フォーマット」 の候補)

フォーマット名	拡張子	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
JPEG 2000	jp2	○	○	○	○	非可逆、可逆圧縮方式、両方に対応。 実装のサポート (タイリング機能が複雑なので使用しないなど) への留意が必要。
TIFF 6 (Tagged Image File Format)	tif, tiff	○		○	○	可逆圧縮方式が基本で、多様な圧縮形式に対応しているが、一般的な圧縮方法 (JPEG等) を使用するなど注意が必要。最新版で移管することが望ましい。
PNG 1.2 (Portable Network Graphics)	png	○	○	○	○	可逆圧縮方式。扱いやすく、圧縮アルゴリズムも公開。
GIF (Graphics Interchange Format)	gif		○	○	○	可逆圧縮方式。
JPEG (Joint Photographic Experts Group)	jpg		○	○	○	非可逆圧縮方式。圧縮アルゴリズムに留意。

- ISO化されていること、圧縮率などを考慮してJPEG2000を採用 (国立国会図書館)。
(補足：当館、国会図書館及び海外の国立公文書館等において、長期保存に適した画像のフォーマットとして採用)
- 標準的なフォーマットとしてはTIFF、PNG、GIF、JPEGなどが候補。ただし、JPEG、TIFFは、圧縮アルゴリズムに留意。
- 書類等のスキャンデータ、デジタル写真など、目的に応じてフォーマットは使い分ける。

(技術動向の観点から注視が必要なフォーマット)

フォーマット名	拡張子	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
HEIF (High Efficiency Image File)	heif					非可逆圧縮方式、Apple社デバイスの管理用フォーマットとして使われているが、現時点では、他のプラットフォームやブラウザで標準対応していないが、今後の動向に留意。

3-3. 音声

長期保存にあたってリスクが低いフォーマット (= 「標準的フォーマット」) の候補)

フォーマット名	拡張子	コーデック例	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
WAVE (Waveform Audio File Format)	wav	リニアPCM (LPCM)	○ ※BWF	○	○	○	非圧縮音声。アーカイブ機関で広く採用。(WAVの拡張形式として、映像業界で広く用いられるBWF (欧州放送連合策定、日本版BWF-J))。
MP3 (MPEG Layer III Audio Encoding)	mp3,mpeg3等	MP3		○	○	○	非可逆圧縮形式。 MP3、MP4は、フレーミング、圧縮方法がシンプル。
AAC (Advanced Audio Coding MPEG-4 Low Complexity Object)	aac, mp4等	AAC			○	○	非可逆圧縮形式。 MP3、MP4は、フレーミング、圧縮方法がシンプル。

- ファイル・フォーマットだけでなく、コーデックの整理が必要 (フォーマットは対応してもコーデックが対応せず開けないことがある)。フォーマットの変換にあたっては、コーデックを同じ、としておくと移行しやすい。
- フォーマットは1つに絞るより、候補を並べるのが現実的。可逆性の有無が重要。
- 長期保存フォーマットは、非圧縮のファイル形式が望ましい (国立国会図書館、国立映画アーカイブ)。WAVはCDとも密接に関係し普及しており、技術的に当面無くなることはない。

(技術動向の観点から注視が必要なフォーマット)

フォーマット名	拡張子	コーデック例	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
FLAC (Free Lossless Audio Codec)	flac	FLAC	○			○	可逆圧縮形式。オープンソースのフォーマットとして採用している機関がある (コンテナフォーマットに格納されている事例がある)

3-4. 動画

長期保存にあたってリスクが低いフォーマット (= 「標準的フォーマット」) の候補

フォーマット名	拡張子	コーデック例	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
MPEG-4	mp4, mpa	(映像) H.264、 H.265等 (音声) AAC 等		○	○	○	QuickTimeをベースに開発された国際規格の主に開発されたフォーマット。Zoomを始めとする動画配信にMP4が用いられることが多く、今後「記録」としての位置づけが高くなる見込み。変換処理にあたりライセンス留意。既に特許権が消滅しているので問題ない。
MPEG-2 (MPEG2-PS, MPEG2-TS)	mpg, mpeg	用途に合わせ選択		○	○	○	

- 音声同様、ファイル・フォーマットとコーデックの整理が必要。
- ISO化がされていても、フォーマット変換の処理等を行う際のライセンスに留意。
- 使用するソフトウェアが対応していること、機械処理ができるよう絞りこんでおくこと。
- 利用目的によるフォーマット選定、カラーマネジメント、字幕やクローズドキャプションが課題。

➡ **動画技術の特性**(技術が多様、変化の大きさと変わりやすさ、容量の大小も様々)、**利用目的や要求品質により条件が大きく変わるため、「標準的フォーマット」の明確化には困難さが残る**(保存・利用条件を整理し、さらに技術的な調査が必要)

(技術動向の観点から注視が必要なフォーマット)

フォーマット名	拡張子	コーデック例	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考 (ご意見等)
WMV (Windows Media Video 9 File Format)	wmv	VC-1		○	○	○	<u>近年触れる機会が少なく、陳腐化のリスクが高いような印象がある。</u>
MOV (QuickTime File Format)	mov	ProRes		○	○	○	<u>QuickTime依存。フォーマットとして良くとも、コーデックが将来的にひっかかる可能性もある。</u>
Matroska (Matroska Multimedia Container)	mkv, mka	(映像) H.264、 FFV1 等 (音声) MP3、AAC、FLAC 等				○	<u>コンテナフォーマット。動画・音声データの様々な規格のデータを格納できる。海外の文化機関を中心に採用が広がっている。Google社のオープンな動画規格であるWebMのコンテナのベース (Matroskaのサブセットを採用) で、今後注視が必要。</u>

3-5. その他フォーマット（CAD、テキストデータ）

カテゴリ	フォーマット名	拡張子	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考（ご意見等）
CAD	STEP (Standard for the Exchange of Product Model Data)	step, stp, p21	○			○	国際標準に準拠した規格のP21は、国土交通省の「CAD製図基準」等の電子納品基準に定めている。長期保存フォーマットとして挙げられる。
CAD	X3D (Extensible 3D)XML 3.0~3.2	x3d, x3dz	○			○	最新版が望ましい。
CAD	PDF/E-1(Portable Document Format/Engineering)	pdf		○		○	長期保存フォーマットとして挙げられる。

（技術動向で注意が必要なフォーマット）

CAD	E57 Lidar Point Cloud Format	e57	○				地方公共団体でよく使われている。
-----	------------------------------	-----	---	--	--	--	------------------

※上記のほか広く使われる3Dプリンタ等のデファクトスタンダードSTL(.stl)、点群のLAS (.las) やLeica PTS (.ptx) のご意見あり

- CAD文書では、保管から取り出し・参照までのプロセス全体で議論が進んでおり、変換する際の確認手順等も標準化、規格化が進められている。機械可読性も重視。
- 2次元データは、長期保存の観点でPDF (/A) 等にする。また、3次元データの扱いは、図面データ等での再利用性やデータの容量面など今後検討が必要。業界・分野ごとの整理が必要。
- 元の文書の再現にはタグの意味やスキーマの理解が重要、スキーマを一緒にきちんと残すことが望ましい。

カテゴリ	フォーマット名	拡張子	①米NARA 望ましい	①米NARA 受入れ可能	②英TNA 長期に維持	③豪NAA 標準	備考（ご意見等）
テキスト	UTF、PlainText	txt,utf, utf8. utf16	○		○	○	日本語は文字化けの可能性。特殊な文字を長期保存する場合は、PDFやPDF/Aが適当。
テキスト	CSV	csv	○		○		数字のみであれば可。数字以外の文字を長期保存したい場合は、PDFやPDF/Aが適当。
テキスト	XML	xml	○		○	○	階層構造を保存したい場合は適しているが、XMLスキーマの解釈情報が必要。テキストそのものも文字化けの可能性があり、長期保存する場合は、PDFやPDF/Aが適当。

(2) 日本ならではのローカル要素の確認

① 国立公文書館への移管実績を踏まえ、

日本ならではのファイル・フォーマットを確認し、専門家等へヒアリング

国立公文書館に移管実績があり、標準的フォーマット以外のもと思われるフォーマット（主なもの）

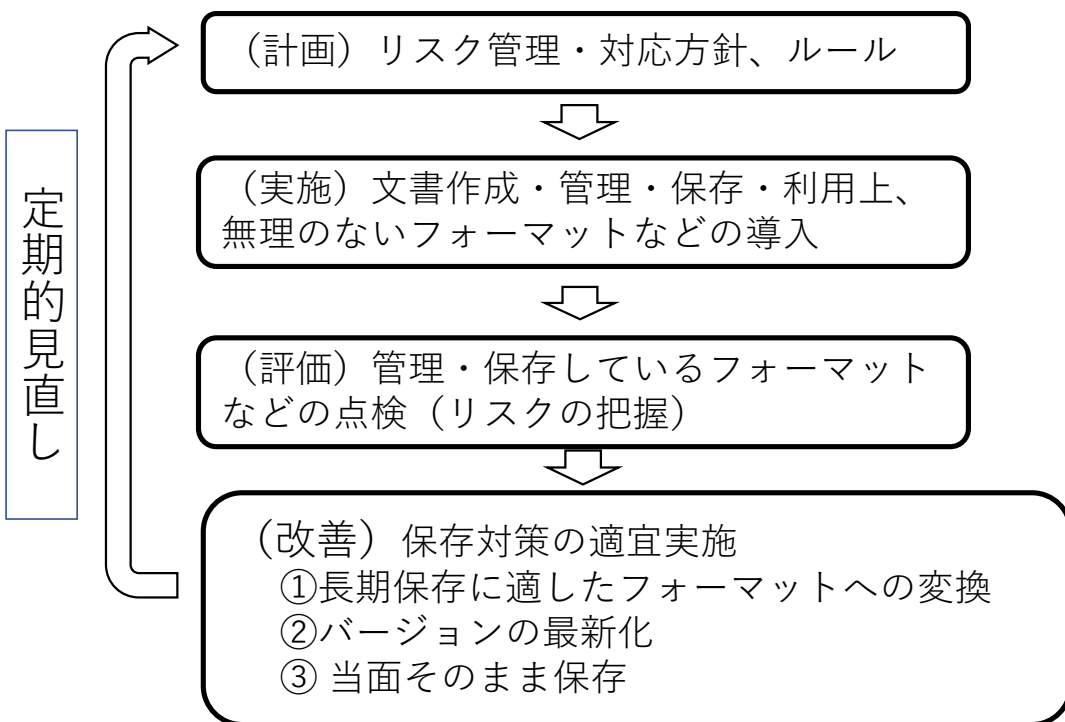
カテゴリ	フォーマット名	拡張子	備考（ご意見等）	館受入れ電子ファイル概数（R2年度まで）
文書	PDF 1.3～1.6	pdf	フォントが埋め込むことが望ましい。長期保存の場合は、PDF/Aが適当。	10万件の内数 ※PDF/A等含む
	Word97～2003	doc	文字コード、変換できない可能性がある。見読性確保のため、フォントは通常のもの、docx又はPDF/Aが適当。	3万件
	一太郎8～2021	jtd/ jtt	長期保存の場合は、PDFやPDF/Aに変換して移管することが望ましい。	8千件
	OpenOffice.org 3.1/3.2 Writer	odt	Solarisなど特別なOS環境を用意する必要がある。	数件
	DocWorks	xdw	長期保存の場合は、PDFやPDF/Aに変換して移管することが望ましい。	500件
	OASIS		現在サポートを受けることができない。	数十件
表計算	Excel97～2003	xls	見読性確保のため、xlsx又はPDF/Aが適当。マクロは複雑、機能部分は再現困難。	1万7千件
	OpenOffice.org 3.1/3.2 Calc	ods	Solarisなど特別なOS環境を用意する必要がある。	数件
プレゼンテーション	PowerPoint97～2003	ppt	見読性確保のため、pptx又はPDF/Aが適当。PowerPoint 97は現在の環境で開けない。	1千件
	OpenOffice.org 3.1/3.2 Impress	odp	Solarisなど特別なOS環境を用意する必要がある。	

- ・ 今から20年以上前の古いフォーマットが移管される場合があり、再生環境を確保する必要。
- ・ 日本では縦書き文書があるため、そのソフトウェア、フォーマットへの考慮が必要。

② 受入れ数の実績が多いソフトウェアについて動向のヒアリング

- ・ 「一太郎」：日本の文書作成ニーズに合わせて設計開発、古いバージョンのファイルもビューアが利用可能
- ・ 「DocWorks」：文書作成・管理全体の業務環境として設計開発、古いバージョンのファイルもビューアが利用可能

4.長期保存に向けたリスク評価方法



○リスク管理・対応のための技術的サポート

- フォーマットに関する技術情報の整理、提供
- フォーマットのチェックツール (フォーマット種別およびバージョン)
- フォーマット自動変換等ツール (長期保存フォーマットへの変換等ツール)

○長期保存に適したフォーマットへの変換等

- 長期保存上のリスクが高いと思われるフォーマットは長期保存フォーマットへ変換 (見読可能なうちに実施)
- 古いバージョンのファイルは最新化。

具体的な判断ポイントに関する専門家等意見

- コーデックやメタデータの不備がリスクとなるため、チェックを実施できる形が望ましい。
- 望ましいフォーマットの判断は、文書の体系や規則等のドキュメントの更新と合わせて、内容を定期的に見直す。 (例：ISOのレビュー)
- フォーマットの評価やサポートの情報など定期的に観測する必要がある。例えば、数年に1回評価をしてランクをさがったものをフォーマット変換することが考えられる。
- 変換した場合も、オリジナルのファイルは保存しておいた方がよい (後日、点検する必要がある場合がある)。

5. 今後の課題等

- 引き続き、以下の検討課題④⑤⑥に関して調査検討を実施

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

➡ フォーマット動向及び利用面からの検討を行い、長期保存フォーマットの特定を行う予定

- ・ 保存と利用しやすさの観点を踏まえたフォーマットのあり方について
- ・ 変換前の電子ファイルの保存と利用について（制度面、技術面から検討が必要）
例：統計データ（変換前Excel等）はPDF/Aに変換するとページが膨大、使い勝手の面で課題

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

➡ 変換に関する技術的な検証が必要であり、今後の調査で検討

- ・ 変換時の再現性（文書：見た目、画像：可逆性など）について
- ・ 変換時の方式やツール、運用性、性能などについて

⑥ 行政機関等と国立公文書館との役割分担の検討

➡ 行政機関等又は館における見読性の確保・リスク確認時の対応、運用上の注意点などの整理 (次頁参照)

検討課題⑥ 参考

文書作成～利用までの工程と検討すべきご意見

