

電子公文書等の適切な保存に係る調査検討報告書

令和2年7月

独立行政法人国立公文書館

目次

1. 調査の概要	1
1.1. 調査の背景及び目的.....	1
1.2. 調査の位置付け.....	2
1.2.1. 過年度における電子公文書等に係る取組経緯.....	2
1.2.2. 次期システムの更改に向けた本調査の位置付け.....	5
1.3. 調査の内容.....	6
1.3.1. 調査の方法.....	6
1.3.2. 調査の手順.....	8
2. 電子公文書等の保存に係る問題及び課題の把握	9
2.1. 電子公文書等の保存に係る政府の方針.....	9
2.1.1. 電子公文書等の管理に係る方針.....	9
2.1.2. 情報システムに係る方針.....	9
2.2. 館における電子公文書等の保存に係る問題及び課題.....	11
2.2.1. 館業務の概要.....	11
2.2.2. 電子公文書等システムの概要.....	16
2.2.3. 館における電子公文書等の保存に係る問題.....	17
2.2.4. 館における電子公文書等の保存に係る課題.....	21
2.3. 専門家ヒアリング.....	25
2.3.1. ヒアリングの概要.....	25
2.3.2. 専門家からの示唆.....	25
2.4. 海外における電子公文書等の保存に係る文献調査.....	27
2.4.1. 電子公文書等に係る館の保存対策に対する示唆.....	27
2.4.2. 受入れ等に係る業務の実施に対する示唆.....	29
2.4.3. 電子公文書等システムの運用に対する示唆.....	30
2.5. 館の課題と対応の方向性.....	31
2.5.1. 政府・館の方針に対する対応.....	31
2.5.2. 電子公文書等に係る館の保存対策に対する対応.....	31
2.5.3. 受入れ等に係る業務の実施に対する対応.....	31
2.5.4. 電子公文書等システムの運用に対する対応.....	32
2.5.5. 課題と対応の関係.....	32
3. 電子公文書等に係る館の保存対策に対する対応	34
3.1. 保存対策のPDCAサイクルの確立.....	34
3.2. フォーマット対応表の見直し.....	36
3.2.1. フォーマット対応表の見直しに係る運用ルール策定.....	36
3.2.2. 現行のフォーマット対応表の見直し.....	39
3.2.3. フォーマット対応表の管理項目の見直し.....	50

3.2.4. フォーマット対応表の改善案.....	50
3.3. オリジナルフォーマットの原則保存.....	54
4. 受入れ等に係る業務の実施に対する対応.....	55
4.1. 受入れ等に係る業務の効率化.....	55
4.1.1. 業務プロセスの見直し.....	55
4.1.2. 機械学習を利用したフォーマット判定機能.....	55
4.2. フォーマット変換処理結果の確認観点の明確化.....	56
4.3. 電子公文書等のフォーマットに係る技術情報蓄積.....	56
4.3.1. フォーマット対応表の履歴管理.....	57
4.3.2. フォーマット変換エラー情報の蓄積.....	57
4.3.3. 技術情報の蓄積に係る運用ルールの策定.....	59
4.3.4. 類縁機関との技術情報の共有.....	59
5. 電子公文書等システムの運用に対する対応.....	60
5.1. 次期システムの更改.....	60
5.2. 外部環境変化を踏まえた要件の具体化.....	63
5.2.1. システム方式の検討.....	63
5.2.2. 非機能要件.....	64
5.3. 電子公文書等に係る館の保存対策を踏まえた機能改善.....	69
5.3.1. 日本独自のファイルフォーマットの判定・登録機能.....	69
5.3.2. ファイルフォーマットの統計情報の出力機能.....	69
6. あるべき姿へ向けて.....	70
6.1. 移管元行政機関等での現用時の文書管理状況の把握.....	70
6.2. 移管元行政機関等に対する電子公文書等の管理及び保存にあたっての依頼.....	70
6.3. 電子公文書等の管理に係る方針の適宜見直し.....	70

目次

図 1-1	電子公文書等システム更新における本調査の位置付け	5
図 1-2	本調査の実施方法	7
図 1-3	調査の手順	8
図 2-1	受入れ等に係る業務の流れ	11
図 2-2	移管される電子公文書等のフォルダ構成	12
図 2-3	電子公文書等システム（現行システム）の全体概要図	16
図 2-4	行政文書ファイル数の実績と今後の予測	22
図 2-5	課題と対応の関係	33
図 3-1	業務の PDCA サイクルとシステムのライフサイクル	34
図 3-2	システム更改に向けたプロジェクト運営スケジュールのイメージ	35
図 3-3	フォーマット対応表の見直し等に関する PDCA サイクルのイメージ	38
図 3-4	SXF のファイル形式	43
図 3-5	図面類を電磁的記録で作成する場合のフォーマット	44
図 4-1	業務プロセスの見直し前後の業務フロー案	55
図 4-2	技術情報の蓄積のイメージ	57
図 4-3	システムと職員による技術情報の蓄積のイメージ	58
図 5-1	次期システムの構成イメージ①（クラウド型）	64
図 5-2	次期システムの構成イメージ②（ハイブリッド型）	64
図 5-3	長期保存用ストレージの試算手順	65

表目次

表 1-1	電子公文書等の保存に係る過年度の主要な取組.....	3
表 1-2	本調査においてヒアリングを実施する専門家.....	7
表 2-1	受入れ等に係る業務	11
表 2-2	現行のフォーマット対応表	13
表 2-3	電子公文書等システムで管理するメタデータ.....	14
表 2-4	行政文書ファイルの受入れ実績	15
表 2-5	電子公文書等システムで保存する電子ファイルのフォーマット概況.....	16
表 2-6	業務とサブシステムの対応	17
表 2-7	電子公文書等に係る館の保存対策に起因する課題.....	23
表 2-8	受入れ等に係る業務に起因する課題	23
表 2-9	電子公文書等システムに起因する課題	24
表 2-10	専門家に対するヒアリング実施概要	25
表 2-11	NARA が保存対象とするフォーマットタイプ	27
表 2-12	TNA が保存対象とするフォーマットタイプ	28
表 3-1	フォーマット対応表の見直しに係る作業項目案.....	36
表 3-2	非標準フォーマットの移管実績	39
表 3-3	標準的フォーマットへの追加に係る確認観点.....	39
表 3-4	テキスト形式のフォーマット毎の移管実績	40
表 3-5	テキストの標準的フォーマット追加の確認結果.....	40
表 3-6	Web ページのフォーマット毎の移管実績	41
表 3-7	Web ページの標準的フォーマット追加の確認結果.....	41
表 3-8	CAD のフォーマット毎の移管実績.....	42
表 3-9	CAD の標準的フォーマット追加の確認結果	44
表 3-10	標準的フォーマットの見直しに係る確認観点.....	46
表 3-11	長期保存フォーマットを見直す際の観点	46
表 3-12	PDF/A に対する確認結果.....	47
表 3-13	JPEG2000 に対する確認結果	47
表 3-14	MP3 に対する確認結果.....	48
表 3-15	MPEG-2 の確認結果.....	49
表 3-16	フォーマット対応表の管理項目見直し内容	50
表 3-17	フォーマット対応表の改善案	51
表 4-1	長期保存フォーマットへの変換における確認観点.....	56
表 4-2	電子公文書等のフォーマットに係る技術情報蓄積に対する対応.....	57
表 5-1	次期システムの対応事項	61
表 5-2	システム方式の概要比較	63
表 5-3	長期保存用ストレージ容量の試算	66
表 5-4	SBD マニュアルを用いたセキュリティ要件の検討結果.....	67
表 5-5	SBD マニュアルに基づくセキュリティ対策要件.....	67

用語の定義

No	用語	説明
1	電子公文書等の移管・保存・利用システム(略記「電子公文書等システム」)	国立公文書館が運用する、電子公文書等の受入れ、保存等を行う情報システムのこと。受入れた電子公文書等は、同システムにおいて長期保存可能なフォーマットに変換されたうえで保存し、整備された目録データ及び電子公文書等の複製データは、国立公文書館デジタルアーカイブ等を通じて一般の利用に供する。
2	国立公文書館デジタルアーカイブ	国立公文書館が運用する、同館が保存する特定歴史公文書等について、インターネットを通じ、目録情報の検索、デジタル画像等の閲覧等を可能とする情報システムのこと。
3	政府全体で利用可能な一元的な文書管理システム	総務省が運用する、文書管理業務の業務・システム最適化計画に基づき整備した政府全体で利用可能な情報システムのこと。
4	オリジナルフォーマット	国立公文書館に移管された電子公文書等のファイルフォーマットのうち、移管された時点における電子公文書等のファイルフォーマットのこと。
5	標準的フォーマット	国立公文書館に移管された電子公文書等のファイルフォーマットのうち、電子公文書等システムへの移管及び長期保存に適したフォーマットへの変換処理に対応するファイルフォーマットのこと。
6	非標準フォーマット	国立公文書館に移管された電子公文書等のファイルフォーマットのうち、標準的フォーマット以外のファイルフォーマットのこと。電子公文書等システムの長期保存に適したフォーマットへの変換において、一部の処理ができない等の理由で個別の運用対応が必要となる。
7	長期保存フォーマット	電子公文書等システムにおいて、電子公文書等を長期保存する際に用いる長期保存に適したファイルフォーマットのこと。
8	一般提供フォーマット	電子公文書等システムにおいて、電子公文書等の複製データとして国立公文書館デジタルアーカイブ等を通じて一般の利用に供する際に用いるファイルフォーマットのこと。
9	フォーマット対応表	国立公文書館が運用する標準的フォーマット、長期保存フォーマット及び一般提供フォーマットの関係を示した表のこと。
10	電子公文書等	移管元行政機関等から国立公文書館に移管される歴史公文書等のうち電子的方式で作られたもの。
11	電子ファイル	電子公文書等を構成する最小単位のもので、文書作成ソフトウェアや表計算ソフトウェアで作成された電子データのこと。
12	受入れ等に係る業務	国立公文書館による電子公文書等システムを通じて行う受入れ、保存、利用、運用管理からなる業務全体のこと。
13	メタデータ	電子データの意味などを説明する情報であり、電子データを適切に管理、保存、利用するために不可欠な情報のこと。
14	現用	文書管理のライフサイクルを示すものであり、国の行政機関等において公文書等が作成・整理・保存される段階のこと。
15	非現用	文書管理のライフサイクルを示すものであり、保存期間満了後の公文書等が特定歴史公文書等として、国立公文書館等において保存し、利用に供する段階のこと。

1. 調査の概要

1.1. 調査の背景及び目的

独立行政法人国立公文書館（以下「館」という。）では平成 23 年度より、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。）を用いて、受入れ等に係る業務を実施してきた。電子公文書等システムを用いた受入れ等に係る業務は、内閣府及び館が平成 21 年度までに実施した調査検討で得た知見をベースとしており、現在の ICT の技術水準や昨今の公文書を取り巻く社会情勢とは釣り合わないものとなりつつある。

近年は、館が受入れする電子公文書等の多様化及び量の増加のため、適切な保存及び利用提供が困難になっている。特に、「公文書管理の適正の確保のための取組について」（平成 30 年 7 月 20 日閣僚会議決定）及び同決定を踏まえて策定された「行政文書の電子的管理についての基本的な方針」（平成 31 年 3 月 25 日内閣総理大臣決定）」において、効率的・確実な文書管理の確立に向けて、今後作成する行政文書は電子的に管理することを基本とすること、また行政文書の作成から保存、廃棄及び館等への移管までを一貫して電子的に行うための仕組みの検討について取り組むことなどが必要とされた。これにより、館が受け入れる電子公文書等の多様化及び量の増加は、今後ますます加速することが見込まれる。

また、館は、平成 28 年度から運用している現行電子公文書等システム（以下「現行システム」という。）のシステム更改に係る検討を開始する時期に来ており、令和 2 年度から予定する次期電子公文書等システム（以下「次期システム」という。）の要件定義に向けた整理が必要な状況にある。

本調査は、政府における電子公文書等の管理や保存等に係る方針及び館の受入れ等に係る業務量などの将来的な見込みを踏まえ、適切な保存に向けた技術や方法、運用のあり方について調査検討することで、現行業務の見直し及び次期システムの更改のための論点整理に資することを目的とする。

1.2. 調査の位置付け

1.2.1. 過年度における電子公文書等に係る取組経緯

電子公文書等の受入れ等に係る業務及び現行システムは、平成 17 年度から実施された電子公文書等の管理、移管及び保存のあり方等に係る調査検討の結果を踏まえて、平成 21 年度に内閣府から示された「電子公文書等の移管・保存・利用の具体的方法に係る方針」に基づき、平成 23 年度より運用を開始している。

平成 23 年度には「公文書等の管理に関する法律」（平成 21 年法律第 66 号。以下「公文書管理法」という。）が施行され、同法の規定の趣旨や特定歴史公文書等における保存、利用及び廃棄を行う際の実務上の留意点について記した「特定歴史公文書等の保存、利用及び廃棄に関するガイドライン」（平成 23 年 4 月 1 日、内閣総理大臣決定。以下「ガイドライン」という。）を踏まえて、館は受入れを始めとする諸業務を実施している。

館は、公文書管理法第 15 条に基づき、特定歴史公文書等を永久保存する責務を負っている。具体的には、その内容、保存状態、時の経過、利用の状況等に応じ、適切な保存及び利用を確保するために、適宜必要な措置を講じた上で保存する責務である。これらの責務を全うするため、館においては、施設・設備や特定歴史公文書等の保存・利用に係る実態を踏まえた保存対策として、「特定歴史公文書等の保存対策方針」（平成 27 年 5 月 27 日館長決定）を定めている。本保存対策方針において、電磁的記録のうち電子公文書等については、原則として、電子公文書等システム内に保存することとしている。

現行システムは、平成 28 年度より運用を開始しており、平成 23 年度からのシステム運用から得られた課題や海外事例を参考に、要件定義がなされたものである。

過年度における電子公文書等に係る取組経緯を次頁に示す。

表 1-1 電子公文書等の保存に係る過年度の主要な取組

年度	主要な取組	補足(本調査に関連の深い事項)
平成 17 年度 ～18 年度	【内閣府】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(平成 18 年 4 月)	<ul style="list-style-type: none"> ● 公文書等の中間段階における集中管理の仕組み(中間書庫システム)の早期整備に向けた検討を実施。 ● 電子公文書を非電子媒体に変換せず、電子媒体のまま保存する。また、公文書の内容・コンテキスト・構造・機能と、記録媒体とを明確に区別し、電子公文書そのものを保存対象とする。 ● 電子公文書の記録としての価値を維持するのに必要不可欠な「エッセンス」を保存する。 ● 作成時から歴史公文書としての保存・利用段階までのライフサイクル全体の管理を行う必要があり、そのメタデータ、ファイルフォーマット等については、公文書の作成前から適切な対応をとる。
	【内閣府】公文書等の適切な管理、保存及び利用に関する懇談会(第二次報告書)(平成 18 年 6 月)	
平成 19 年度	【内閣府】平成 19 年度 電子公文書等の管理・移管・保存のあり方を検討する実証実験(調査研究)(平成 20 年 3 月)	<ul style="list-style-type: none"> ● メタデータのかなりの部分が、比較的簡便なツールによって電子ファイルのプロパティ情報から抽出できることを確認。 ● 特定の OS に依存する文字コード及び文字フォントについて、別の OS での漢字の字体変化及びプロパティ情報の喪失等を確認。 ● 国の各機関で電子記録作成に主に用いられているファイルフォーマットの多くが PDF/A に変換できることを確認。他方、音声及び動画等のフォーマット変換では、品質の劣化等の問題があることを確認。
平成 20 年度	【内閣府】平成 20 年度 電子公文書等の管理・移管・保存・利用システムに関する調査(平成 21 年 3 月)	(1)受入れ・保存管理機能、(2)行政利用機能及び(3)デジタルアーカイブ連携機能からなるプロトタイプシステムを構築・検証し、構築に向けた課題を抽出。
平成 21 年度	【館】前電子公文書等システムの要件定義書の作成及び構築ベンダーの調達	-
	公文書等の管理に関する法律(平成 21 年 7 月 1 日公布)	-
	【内閣府】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)	<ul style="list-style-type: none"> ● 館は、平成 23 年度から電子公文書等の各府省等からの受入れを開始する。 ● 各府省等は、可搬媒体又はネットワークにより電子公文書等を移管する。 ● 各府省等は、移管する電子公文書等について、フォーマット等に関する情報を可能な限り内閣府に提供する。 ● 館は、可搬媒体又はネットワークにより移管された電子公文書等について、媒体変換を行った上で検疫等を行う。 ● 館は、受け入れた電子公文書等を原則として長期保存フォーマットに変換した上で、電子公文書等システムで保存する。 ● 館は、社会情勢及び情報技術の変化等を注視し、適時適切な措置を講ずる。 ● 館は、電子公文書等システムで保存する電子公文書等の複製物を作成し、デジタルアーカイブ等により一般の利用に供する。 ● 館は、移管された電子公文書等について、移管元府省等職員が霞が関 WAN を通じて利用できるシステムを構築・運用する。
平成 22 年度	【館】前電子公文書等システムの構築	-

年度	主要な取組	補足(本調査に関連の深い事項)
平成 23 年度	公文書等の管理に関する法律(平成 23 年 4 月 1 日施行)	-
	特定歴史公文書等の保存、利用及び廃棄に関するガイドライン(平成 23 年 4 月 1 日、内閣総理大臣決定)	「公文書等の管理に関する法律」の規定の趣旨・意義や特定歴史公文書等の保存、利用及び廃棄を行う際の実務上の留意点が記されている。
	【館】前電子公文書等システムの運用開始	-
平成 24 年度	-	-
平成 25 年度	【館】電子公文書等の移管・保存・利用システムの次期システムの要件検討に向けて(平成 26 年 3 月)	<ul style="list-style-type: none"> ● イギリス国立公文書館の PRONOM 及び DROID を引き続き活用するとともに、国内製ソフトウェアに関しては、電子ファイルの情報を蓄積し、将来における技術情報の提供等に備えるものとする。 ● 相互運用性を確保するために、データ出入力に係る要件について、MoReq2010 を踏まえた要件を取り入れる。 ● フォーマット変換やメタデータ抽出に係る処理エラーの事例と個別の対応方法に係る情報を技術レファレンスとして蓄積する。 ● 遠隔地バックアップシステムを長期保存用光ディスクによる保存に変更する。メインデータのバックアップを光ディスクに保存し、その取り扱いに係るノウハウを蓄積しながら、引き続き長期保存用光ディスクに係る動向を注視していく。
平成 26 年度	【館】現行電子公文書等システムの要件定義書の作成及び構築ベンダーの調達	-
平成 27 年度	【館】現行電子公文書等システムの構築	-
平成 28 年度	【館】現行電子公文書等システムの運用開始	-
平成 29 年度	-	-
平成 30 年度	公文書管理の適正の確保のための取組について(平成 30 年 7 月 20 日、行政文書の管理の在り方等に関する閣僚会議)	<ul style="list-style-type: none"> ● 内閣府が主導して、文書管理者が一元的に文書管理できるよう、文書の所在情報を的確に把握しうる電子的な行政文書管理の仕組みを構築する。 ● 作成、保存、廃棄・移管を一貫して電子的に行う仕組みの構築について、内閣府が基本的な方針を策定する。
	行政文書の電子的管理についての基本的な方針(平成 31 年 3 月 25 日、内閣総理大臣決定)	<ul style="list-style-type: none"> ● 今後作成・取得する行政文書について、所在把握や探索等を容易にするとともに、電子媒体を正本・原本として体系的に管理することを基本とし、そのための枠組みを構築することとする。 ● 各行政機関における行政文書の作成、整理・保存、移管又は廃棄の各段階における利便性・効率性の確保と、機密保持・改ざん防止の確保の双方に留意しつつ、文書管理業務のプロセス全体を通じた電子的管理の枠組みを構築する。 ● 文書管理業務のプロセス全体を電子化する中において、従来手作業により行っていた作業工程の自動的な処理を可能とする。将来的には、確実かつ効率的に管理を行うことができる枠組みを構築することにより本格的な電子的管理を実現する。
令和元年度	【館】次期電子公文書等システムの調査検討	-

1.2.2. 次期システムの更改に向けた本調査の位置付け

電子公文書等システムの現行システムは平成 28 年度から運用を開始しており、館では、令和 4 年度から次期システムの運用開始を計画している。令和元年度は、令和 2 年度から始まる次期システムの要件定義の前年度であり、本調査は要件定義を行う上での検討材料及び前提条件を整理する工程と位置付けられる。



図 1-1 電子公文書等システム更新における本調査の位置付け

1.3. 調査の内容

1.3.1. 調査の方法

本調査は、過年度における電子公文書等に係る取組経緯、及び電子公文書等システムの更改に向けた本調査の位置付けを踏まえ、以下に示す2つの視点で調査及び検討を実施した。

1.3.1.1. 長期保存の技術・方法

調査の1つ目の視点として、多様なフォーマットかつ大量の電子公文書等の適切な保存に向けて、館における受入れ等に係る業務及び電子公文書等システムを構成する技術、長期保存の方法として考慮すべき事項を検討する。検討にあたっては、受入れ等に係る業務の要であるフォーマット変換に主眼を置いて、長期保存に適したフォーマット（以下「長期保存フォーマット」という。）の変換仕様や現行業務で顕在化している問題の解決に重点を置くこととする。

1.3.1.2. 長期保存に求められるシステム環境・運用管理

調査の2つ目の視点として、今後ますます増大することが見込まれる電子公文書等を保存するために、安定的かつ継続的にサービスを提供するためのシステム環境及びシステムを長期的にマネジメント、運用管理していくために考慮すべき事項を検討する。検討にあたっては、館に移管される電子公文書等のデータ容量に合わせた適切なデータ容量の試算や次期システムで求められる運用管理のあり方を踏まえ、システムの全体構成や非機能要件など、次期システムの要件検討に大きな影響を与える事項の整理に重点を置くこととする。

なお、本調査にあたっては、現行の受入れ等に係る業務やシステムにおいて、顕在化している事象を「問題」として抽出した。その問題から、あるべき姿とのギャップである「課題」としての整理を行い、各課題の解決に向けた「対応」について検討した。本調査の実施方法のイメージを以下に示す。

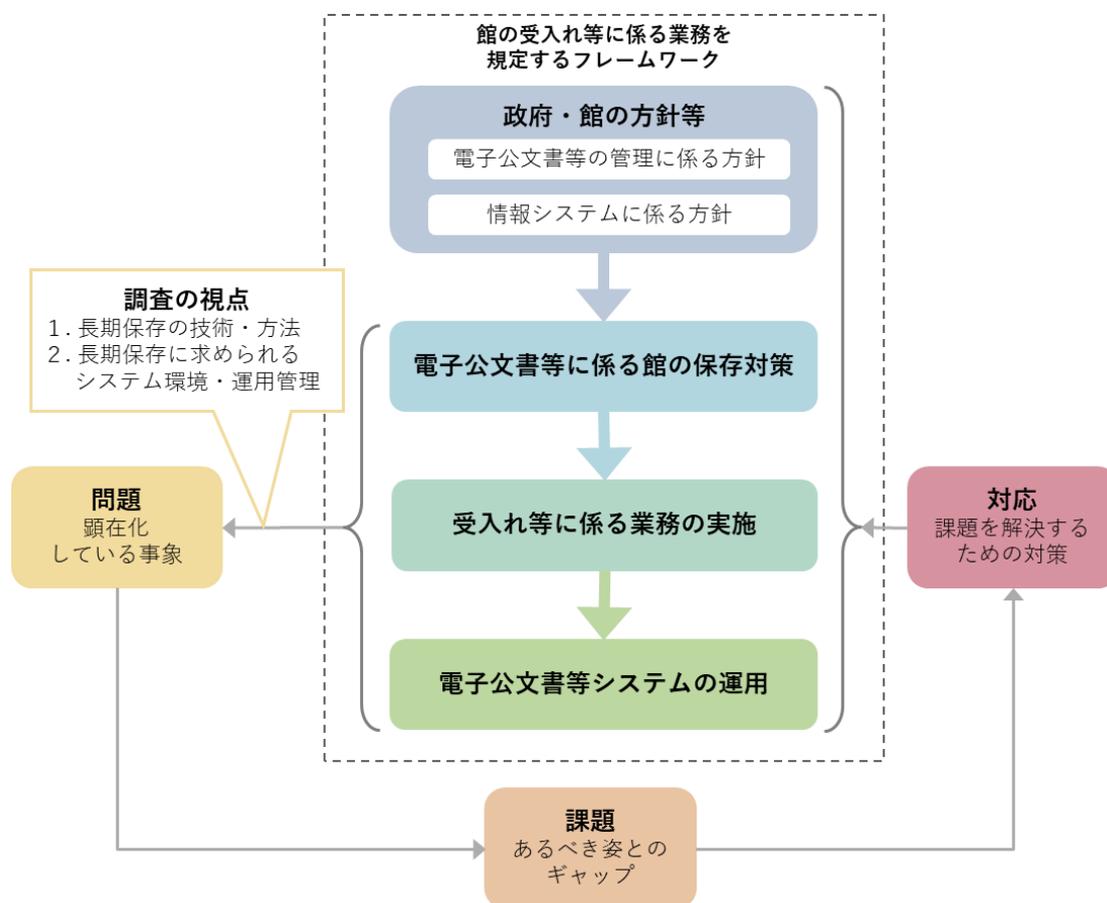


図 1-2 本調査の実施方法

なお、電子公文書等の適切な保存に向けた課題の整理及び対応の検討にあたっては、電子公文書等の保存について知見を有する専門家へのヒアリングを参考とする。本調査において、ヒアリングを実施する専門家を以下に示す。

表 1-2 本調査においてヒアリングを実施する専門家

氏名・所属等	専門分野	委員歴など
筑波大学 名誉教授 杉本重雄	図書館情報学 人文社会情報学	<ul style="list-style-type: none"> ● 内閣府公文書管理委員会委員(平成 22 年 7 月~平成 26 年 7 月) ● 内閣府知的財産戦略本部デジタルアーカイブの連携に関する実務者協議会構成員(平成 27 年 11 月~平成 29 年 3 月) ● 内閣府知的財産戦略本部デジタルアーカイブジャパン推進委員会及び実務者検討委員会 実務者検討委員会構成員(平成 29 年 11 月~平成 31 年 3 月) ● デジタルアーカイブ学会理事(平成 29 年 5 月~)

氏名・所属等	専門分野	委員歴など
東京大学文書館 准教授 森本祥子	アーカイブズ学	<ul style="list-style-type: none"> ● 国立公文書館アーキビストの職務基準に関する検討会議 構成員 (平成 29 年 5 月~平成 30 年 3 月) ● 国立公文書館アーキビストの職務基準に関する検討会議 構成員 (平成 30 年 9 月~平成 31 年 3 月) ● 内閣府公文書管理委員会専門委員(平成 30 年 9 月~令和元年 9 月)

また、電子公文書等の適切な保存に向けた対応について検討するにあたり、海外における電子公文書等の保存の現状を文献等調査により整理した。なお、本調査の調査対象国として米国と英国を選定した。選定理由として、電子記録の管理におけるクラウド技術の活用や業務における ICT 適用の事例等が抽出可能と考えたためである。

1.3.2. 調査の手順

本調査は、「1.3.1 調査の方法」で示した 2 つの視点に基づき、まず電子公文書等の保存に係る現状の調査を通じて問題と課題の把握を行う。そして、抽出された課題を踏まえた電子公文書等の適切な保存に向けた対応について検討を行う。

最後に、我が国における電子公文書等の保存のあるべき姿の実現に向けて、望まれる取組について述べることとする。

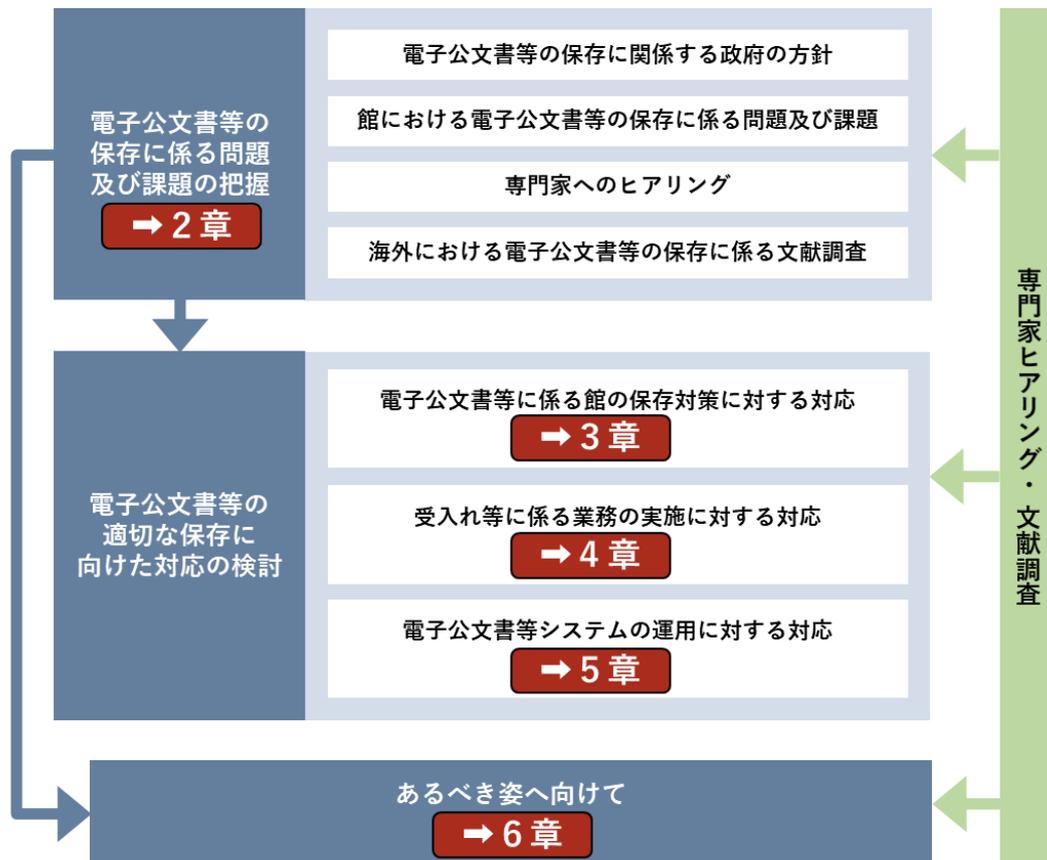


図 1-3 調査の手順

2. 電子公文書等の保存に係る問題及び課題の把握

2.1. 電子公文書等の保存に係る政府の方針

館における受入れ等に係る業務は、政府が示す電子公文書等の保存や基盤となる情報システムに係るガイドラインや方針、計画を根拠に実施されることになる。電子公文書等の保存に係る問題及び課題の把握の前提として、政府のガイドラインや方針等を十分に踏まえる必要がある。

2.1.1. 電子公文書等の管理に係る方針

今後の受入れ等に係る業務を検討する上で、公文書の電子化及び電子決裁への移行の加速が政府方針として示されていることを念頭に、受け入れる電子公文書等の量が増大することを考慮する必要がある。

2.1.1.1. 電子決裁移行の加速

「電子決裁移行加速化方針」（平成30年7月20日、デジタル・ガバメント閣僚会議）により、政府において電子決裁を困難としている要因への対応方針を始めとして、一連の業務プロセス見直しの中で電子決裁への移行を加速化するための方針が取りまとめられている。館へ移管される電子公文書等が増加することが想定されるため、次期システムの非機能要件の検討において考慮する必要がある。

2.1.1.2. 行政文書の電子的管理

「行政文書の電子的管理についての基本的な方針」（平成31年3月25日、内閣総理大臣決定）において、新たな館の開館時期を目途として、今後作成する行政文書は電子的に管理することを基本とすることが定められている。また、そのための仕組みとして、行政文書の作成から保存、廃棄及び館等への移管までを一貫して電子的に行うための仕組みの確立が必要であることが示されている。次期システムについては、移管元行政機関等の側で構築される仕組みとの連携を考慮する必要がある。

2.1.2. 情報システムに係る方針

政府では、内閣官房情報通信技術（IT）総合戦略室を中心に政府情報システムに係る各種方針が示されている。次期システムの検討にあたっては、これらの方針に整合した形でとりまとめを行う必要がある。

2.1.2.1. 業務改革（BPR）の徹底

「デジタル・ガバメント実行計画」（令和元年12月20日閣議決定）では、デジタル技術の徹底活用と官民協働を軸として、全体最適を妨げる行政機関の縦割りや、国と地方、官と民という枠を超えて行政サービスを見直すことにより、行政のあり方そのものを変革していくこと、すなわち、デジタル社会に対応したデジタル・ガバメントの実現を求めている。特

に、サービスデザイン・業務改革（BPR）の徹底による情報システムの整備及び管理の重要性が示されている。次期システムにおいても、業務改革（BPR）を意識したサービス視点での検討を進める必要がある。

2.1.2.2. クラウド・バイ・デフォルト原則

「デジタル・ガバメント実行計画」を基に策定された「政府情報システムにおけるクラウドサービスの利用に係る基本方針」（平成 30 年 6 月 7 日各府省情報化統括責任者（CIO）連絡会議決定）では、政府情報システムを整備する際に、クラウドサービス⁽¹⁾の利用を第一候補とすることを求めている。次期システムにおいても、各種クラウド（パブリッククラウド及びプライベートクラウド等）の特徴を踏まえた上で、クラウド利用を積極的に検討する必要がある。

⁽¹⁾ クラウドサービスとは、「事業者等によって定義されたインタフェースを用いた、拡張性、柔軟性を持つ共用可能な物理的又は仮想的なリソースにネットワーク経由でアクセスするモデルを通じて提供され、利用者によって自由にリソースの設定・管理が可能なサービスであって、情報セキュリティに関する十分な条件設定の余地があるものをいう」。

2.2. 館における電子公文書等の保存に係る問題及び課題

2.2.1. 館業務の概要

2.2.1.1. 業務概要

館における受入れ等に係る業務は、3つの業務（受入れ、保存及び利用）と運用管理から構成される。受入れ等に係る業務の一覧とその流れのイメージを以下に示す。

表 2-1 受入れ等に係る業務

業務分類	業務
受入れ	受入れ
	検疫
保存	フォーマット変換
	メタデータ付与
	長期保存
	遠隔地バックアップ
利用	行政利用
	審査・マスキング
	一般利用
運用管理	運用管理

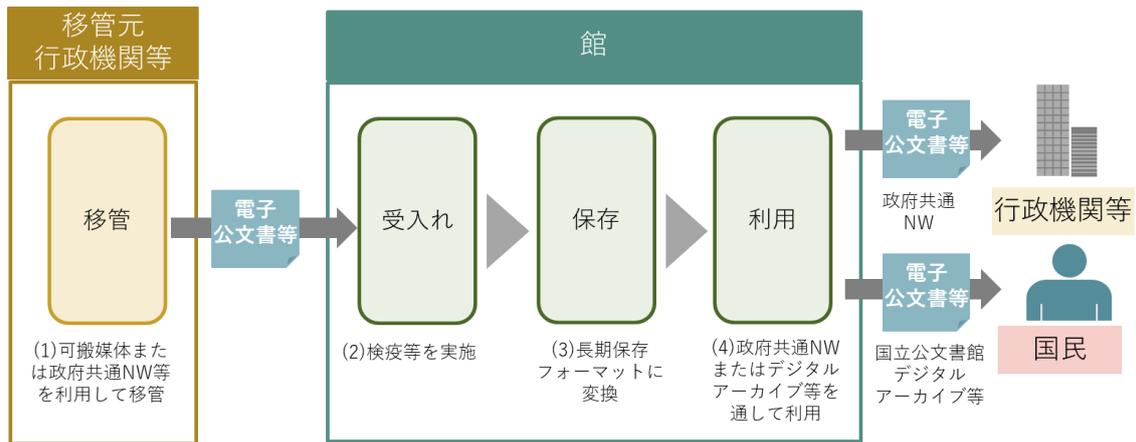


図 2-1 受入れ等に係る業務の流れ

2.2.1.2. 受入れ業務

行政機関及び独立行政法人等（以下「移管元行政機関等」という。）から移管の申出があった電子公文書等について、事前に移管数や搬出方法等の確認を行ったうえで、受入れ作業を行う。なお、移管元行政機関等からの電子公文書等の搬出方法としては、政府全体で利用

可能な一元的な文書管理システム（以下「一元的な文書管理システム」という。）から政府共通ネットワークを経由して館がダウンロードする方法と、可搬媒体で館が移管元行政機関等から引き取りを行う2種類が存在する。

移管される電子公文書等の構成は、移管元行政機関等、行政文書ファイル及び行政文書の順でフォルダが構成されており、行政文書のフォルダに電子ファイルが格納されている。

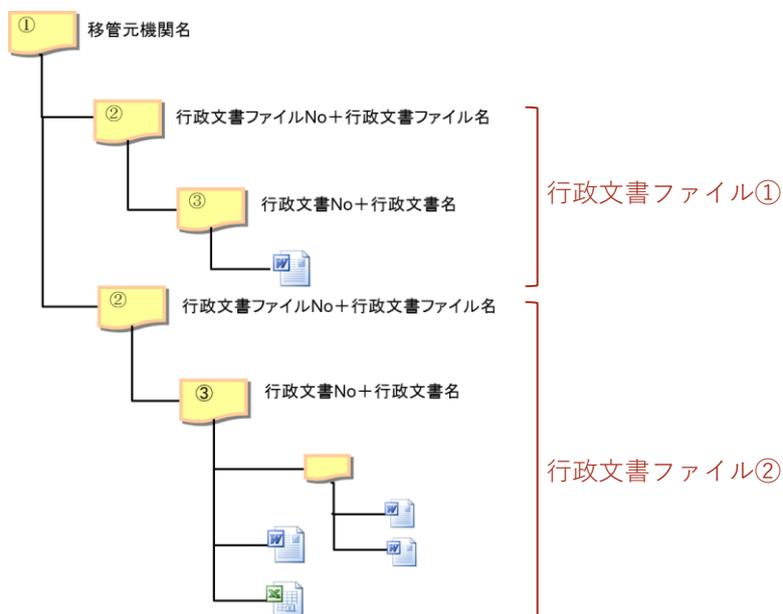


図 2-2 移管される電子公文書等のフォルダ構成

また、電子公文等書の受入れ後、電子ファイルが開けるか等のデータ確認を行い、電子ファイルが開けない等の場合は、移管元行政機関等に対して問い合わせを行う。

2.2.1.3. 検疫業務

受入れを行った電子公文書等について、コンピュータウイルスが混入していないことを確認するために検疫を行う。

2.2.1.4. フォーマット変換業務

館は、「電子公文書等の移管・保存・利用の具体的方法に係る方針」に基づき、受け入れた電子公文書等はその見読性を長期に確保することを図るため、原則として長期保存フォーマットに変換した上で、システムで保存することとされている。

そのため、前工程で検疫が行われた電子公文書等の電子ファイルに対して、電子公文書等システムにおいてフォーマットのチェックを行い、フォーマット対応表（表 2-2 現行のフォーマット対応表）に従って、長期保存フォーマットに変換している。

表 2-2 現行のフォーマット対応表

種類	標準的フォーマット	長期保存フォーマット
文書作成	OASYS 一太郎 8-2014 Word 97-2003 Word 2007-2013 PDF PDF/A Open Office Writer	PDF/A
表計算	Excel 97-2003 Excel 2007-2013 Open Office Calc	PDF/A
プレゼンテーション	PowerPoint 97-2003 PowerPoint 2007-2013 Open Office Impress	PDF/A
画像	JPEG JPEG2000 GIF TIFF BMP	JPEG2000(lossless)
音声	WAVE MP3 WMA	MP3(ビットレート 256kbps 以上)
映像	QuickTime ムービー WMV Real Video MPEG	MPEG-2

電子公文書等システムにおいて、システム内の登録、保存及び利用のためのフォーマット変換に対応したファイルフォーマットを「標準的フォーマット」、標準的フォーマット以外のフォーマットを「非標準フォーマット」と定義している。標準的フォーマットは長期保存フォーマットに変換するが、非標準フォーマットについては、運用している電子公文書等システムでは長期保存フォーマットへの変換に技術的に対応できていないため、受入れ時のオリジナルフォーマットのみを保存している。

2.2.1.5. メタデータ付与業務

電子公文書等システムに取り込んだ電子公文書等にメタデータを付与する。電子公文書等システムでは、一元的文書管理システムから移管された行政文書ファイルに関するメタデータを記録管理メタデータとして、また電子ファイルのプロパティや再生するソフトウェアなどの技術情報が抽出可能な場合には技術的メタデータとして機械的に付与する。その後、館として保存と利用の管理で必要な情報として、館職員によってアーカイバルメタデータを作成し、付与する。技術的メタデータとアーカイバルメタデータから、電子公文書等システムがコンテナメタデータを自動的に生成する。

また、一般利用者向けの国立公文書館デジタルアーカイブへの連携のため、DA 目録データを電子公文書等システムで生成し、提供する。

メタデータの種類を次頁に示す。

表 2-3 電子公文書等システムで管理するメタデータ

No	メタデータの種類	メタデータの情報
1	記録管理メタデータ	電子公文書等の作成・取得時の移管元行政機関等の名称、分類、文書名、作成者等及びその文書について説明し、管理するためのメタデータ。
2	技術的メタデータ	電子公文書等のファイル形式、作成したアプリケーションとそのバージョン、アプリケーションが動作する OS 等及びファイル作成時の技術的内容に関するメタデータ。
3	アーカイバルメタデータ	電子公文書等の移管後に、非現用の保存・管理するために必要な識別番号、受入情報、保存延長情報、公開情報及び破棄情報などのメタデータ。
4	コンテナメタデータ	記録管理メタデータ、技術的メタデータ及びアーカイバルメタデータの全てのメタデータを包含するメタデータ。電子公文書等一件に対して、1 個のコンテナメタデータが対応する。
5	DA 目録データ	館所蔵の歴史公文書等の資料群、簿冊及び件名・細目単位で作成した記録史料記述情報であり、国立公文書館デジタルアーカイブ上で管理・運用されているメタデータ。電子公文書等システムから抽出及び整理し、国立公文書館デジタルアーカイブへ受け渡される。

2.2.1.6. 長期保存業務

電子公文書等システムでは長期保存のため、前工程までのサブシステムとは独立した長期保存用ストレージに、電子公文書等（オリジナルフォーマット及び長期保存フォーマット）とコンテナメタデータを保存する。なお、長期保存された電子公文書等については、定期的にはサンプル再生及びハッシュ値比較を行い、改ざんのチェックを行う。

2.2.1.7. 遠隔地バックアップ業務

長期保存した電子公文書等のデータ及びコンテナメタデータは多重バックアップのため、バックアップしたデータを長期保存用光ディスクへ格納し、当該光ディスクを国立公文書館つくば分館（以下「つくば分館」という。）に搬送及び保存する。なお、当該光ディスクについては、定期的に媒体の品質チェックを行う。

2.2.1.8. 行政利用業務

移管元行政機関等の利用者に対し、政府共通ネットワークを通じて、電子公文書等の検索並びにオリジナルフォーマット及び長期保存フォーマットによる閲覧・ダウンロードのサービスを提供する。行政利用にあたっては、移管元行政機関の利用者に ID の発行を行う。

2.2.1.9. 審査・マスキング業務

電子公文書等の事前審査や利用請求がなされた場合、利用制限事由該当性の審査を行う。審査の結果、利用制限情報が含まれる場合、当該電子公文書等にマスキング処理を行った複製物を作成し、承認者が承認を行う。承認を受けたマスキング済の複製物は、一般提供用データとして変換等の処理を行い、国立公文書館デジタルアーカイブ等において公開する。

2.2.1.10. 一般利用業務

一般利用者からの電子公文書等の利用請求等に対して、国立公文書館デジタルアーカイブでの公開、又は来館利用により一般提供用データを提供する。来館利用の場合は、利用請求された電子公文書等の一般提供用データを来館利用用の端末に複製し、利用者はそれを参照する。

2.2.1.11. 運用管理業務

各業務における計画及び実績の進捗管理、並びに各業務完了の承認を行う。

2.2.1.12. 業務の現況

館で受入れた行政文書ファイルの件数は、平成 29 年度から平成 30 年度の 1 年間で約 33% 増加している。また、電子公文書等システムの運用開始翌年度である平成 24 年度から平成 30 年度までの間に、行政文書ファイル件数は年度平均で 58% 増加している。

平成 23 年度から平成 30 年度までに移管された電子公文書等の行政文書ファイル数を以下に示す。

表 2-4 行政文書ファイルの受入れ実績

移管元 行政機関等	平成 23 年度	平成 24 年度	平成 25 年度	平成 26 年度	平成 27 年度	平成 28 年度	平成 29 年度	平成 30 年度
行政機関	689	25	44	31	70	140	214	282
独立行政法人	0	0	0	1	1	1	1	1
司法府	11	0	0	1	0	1	0	1
寄贈・寄託	0	0	0	0	0	0	0	1
合計	700	25	44	33	71	142	215	285

なお、平成 23 年度の行政文書ファイルの受入れ件数には、広報資料 652 件を含む。

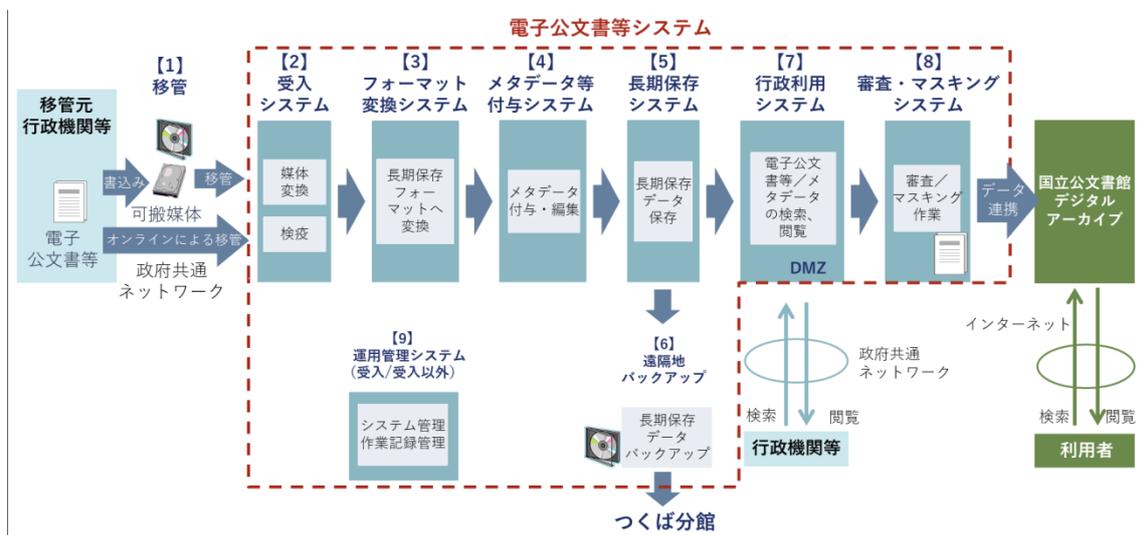
また、電子公文書等システムが運用を開始した平成23年度から平成30年度までに受入れた電子公文書等のうち、電子公文書等システムで保存している電子ファイルのフォーマット種類、ファイル概数、割合を以下に示す。

表 2-5 電子公文書等システムで保存する電子ファイルのフォーマット概況

フォーマット種類	概数	全体に占める割合
標準的フォーマット	5万件	約70%
文書作成(OASYS、一太郎、Word、PDF)	3万件	
表計算(Excel)	4千件	
プレゼンテーション(PowerPoint)	2百件	
画像(JPEG、GIF、TIFF、BMP)	8千件	
音声(WAVE)	20件	
映像(QuickTime)	数件	
非標準フォーマット	2万件	約30%
合計	7万件	100%

2.2.2. 電子公文書等システムの概要

館における受入れ等に係る業務を支える電子公文書等システムについて、現行システムの全体概要図を以下に示す。



電子公文書等システムは、受入れ等に係る業務毎にサブシステムで構成されており、各業務が終了し次工程へ進むのと合わせて、サブシステム間でデータの受け渡しが行われる。また、各業務を担当する館の担当係に応じてサブシステムの操作権限を管理するためのユーザー認証機能を設けている。

また、サブシステム間でのデータ受け渡し方法は、ネットワーク構成の関係上、人手を介した方法及びオンラインでのデータ転送の2種類がある。

なお、電子公文書等システムのネットワーク構成については、セキュリティを考慮しインターネットや館のLANと切り離された構成となっている。また、受入れデータを検疫する受入システムは、他のサブシステムのネットワークからも分離され、独立したシステムとして構築されている。

受入れ等に係る業務と電子公文書等システムを構成するサブシステム・機能の対応、サブシステムの概要を以下に示す。

表 2-6 業務とサブシステムの対応

業務分類	業務	サブシステム・機能	概要
受入れ	受入れ	【2】受入システム	移管時媒体から電子公文書等の受け付け、ディレクトリ構造等の確認、媒体変換、検疫(一次、二次及び最終)及びフォーマット変換システムへの受け渡し等、受入れ業務全般を行う。
	検疫		
保存	フォーマット変換	【3】フォーマット変換システム	電子公文書等をオリジナルフォーマットから長期保存フォーマットに変換するとともに、テキストデータ及びメタデータの抽出を行う。
	メタデータ等付与	【4】メタデータ等付与システム	電子公文書等に対するメタデータ付与及び編集を行うとともに、DA 目録データの生成及び出力を行う。
	長期保存	【5】長期保存システム	電子公文書等のオリジナルフォーマット、長期保存フォーマット、コンテナメタデータ及び媒体データを保存する。
	遠隔地バックアップ	【6】遠隔地バックアップ	つくば分館で保存する長期保存した電子公文書等及びコンテナメタデータのバックアップデータを出力する。
利用	行政利用	【7】行政利用システム	電子公文書等に関し、政府共通ネットワークにより、各府省庁職員による検索及び閲覧を可能とする。
	審査・マスキング	【8】審査・マスキングシステム	電子公文書等の一般利用に向けて、利用制限区分の設定、審査結果に基づくマスキング作業を行う。作業後、国立公文書館デジタルアーカイブのための一般提供用データを生成する。
	一般利用		
運用管理	運用管理	【9】運用管理システム(受入)	受入システムを利用する利用者及び権限を管理する。受入システムに関する作業記録を蓄積、管理及び提供する。蓄積した作業記録を運用管理システム(受入以外)に受け渡す。
		【9】運用管理システム(受入以外)	受入システム以外のサブシステムを利用する利用者及び権限を管理する。受入システム以外のサブシステムに関するにおける作業記録を蓄積、管理及び提供する。また、運用管理システム(受入)の作業記録についてと合わせて統合的に管理する。

2.2.3. 館における電子公文書等の保存に係る問題

2.2.3.1. 電子公文書等に係る館の保存対策に関する問題

受入れ等に係る業務及び電子公文書等システムは、「電子公文書等の移管・保存・利用の具体的方法に係る方針」及び「特定歴史公文書等の保存、利用及び廃棄に関するガイドライ

ン」を踏まえて業務の実施及び運用が行われている。そのため、受入れ等に係る業務及び電子公文書等システムの見直しを検討する上では、館の保存対策の見直しについても検討する必要がある。

館の保存対策について顕在化している問題を抽出し、整理した。

問題 1-1： 移管を想定するフォーマットを標準的フォーマットとしていたが、非標準フォーマットによる移管が一定数存在している

館では、「電子公文書等の移管・保存・利用の具体的方法に係る方針」を踏まえ、移管が見込まれるフォーマットとして標準的フォーマットを定めている。しかし、「表 2-5 電子公文書等システムで保存する電子ファイルのフォーマット概況」に示すとおり、電子公文書等システムの運用を通して、非標準フォーマットによる移管が全体の約 30%存在している。

非標準フォーマットについては、長期保存フォーマットへ変換しないため、長期的な見読性の確保の観点で懸念がある。

問題 1-2： 標準的フォーマット及び長期保存フォーマットについて、受入れ実績及び技術動向を反映した見直しをしていない

館では、前システムの運用を開始した平成 23 年度から「表 2-2 現行のフォーマット対応表」に基づき、現在まで運用を行ってきた。結果として、「問題 1-1」のとおり非標準フォーマットによる移管が一定程度存在することが明らかになり、また、受入れ後も技術的な変化が想定されることから、業務の現在の実態を踏まえた標準的フォーマット及び長期保存フォーマットの見直しが求められる。

問題 1-3： 標準的フォーマット及び長期保存フォーマットを見直しするための手順が定められていない

「問題 1-2」の背景として、標準的フォーマット及び長期保存フォーマットの見直しに係る手順を館として定めていないことがあげられる。

問題 1-4： 受入れ時のフォーマットをシステム内に保存することが原則とされていない

内閣府から示された「電子公文書等の移管・保存・利用の具体的方法に係る方針」において、「受入れ時のフォーマットによる保存が必要と認められる場合は、「長期保存フォーマット」及び受入れ時のフォーマットにより保存すること」とされていることを踏まえ、館では、受け入れた電子公文書等の長期保存フォーマットと、可能な場合には運用上の判断で受入れ時のオリジナルフォーマットをシステム内に保存している。「必要と認められる場合」の解釈による保存では、長期にわたるシステム運用において、受入れ時のフォーマットが適切に保存されず、見読性を確保できなくなる懸念がある。

2.2.3.2. 受入れ等に係る業務の実施に関する問題

「2.2.1 館業務の概要」で示した受入れ等に係る業務について、顕在化している問題を抽出し、整理した。

問題 2-1： 受入れ業務において、移管対象外となる電子ファイルを検出できない

移管元行政機関等による文書作成や館への移管において、行政文書ファイル内に人が意図しないコンピュータで自動生成されるファイルや空（ファイルサイズが0バイト）のファイル、対象外の電子ファイルが含まれる事例がある。移管対象となる行政文書ファイルに対する認識の問題も関わるが、現状ではこれらファイルの有無の検出や確認は困難であり、本来必要のない電子ファイルを保存している可能性がある。

問題 2-2： 受入れ業務において、一元的な文書管理システムと手動のダウンロードによる連携となっており、時間を要している

「2.2.1 館業務の概要」に示すとおり、受入れにあたっては、館の職員が一元的な文書管理システムから行政文書ファイルを手動でダウンロードの作業を行っている。今後、行政文書の電子的管理や移管が推進されることを踏まえると、当該作業を館職員の人手により実施している現状は、業務の円滑な遂行の阻害要因となることが懸念される。

問題 2-3： フォーマット変換業務において、長期保存フォーマットへの変換後の品質確認が不適合となる場合があり、再変換作業に時間を要している

現行システムでは、長期保存フォーマットへの変換作業のうち、自動変換が適切に実施できないケースが存在し、手動での再変換の実施に伴う追加作業負荷が発生するとともに、受入れ等に係る業務の円滑な遂行が阻害されている。

問題 2-4： フォーマット変換業務において、非標準フォーマットはフォーマットの種類を判定できていない

現行システムでは、DROID（ファイルフォーマットを自動的に特定するツール）を利用しているところ、全体の約7割のファイルフォーマット抽出は可能であるが、非標準フォーマットに係るフォーマット判定ができないこと、またCADや日本でのシェアが高かったとしても海外では販売が限られるフォーマットには対応していない。フォーマット変換業務において、手動での確認作業が必要であるため、受入れ等に係る業務の円滑な遂行が阻害されている。

問題 2-5： 運用管理業務において、作業記録管理機能に一覧性がなく、計画に基づく進捗確認等の機能が無いため、精度の高い運用管理ができない

現行システムでは、移管される各行政文書ファイルの受入れ等に係る業務において、サブシステムの運用管理機能は有しているが、館職員による各作業の詳細な工程管理や進捗状況の把握を行うことが難しい。

問題 2-6： 遠隔地バックアップ業務において、長期保存に係る新たな技術であった長期保存光ディスクを採用したが、その後の技術動向及びバックアップに求められる要件の変化を踏まえた検証が必要である

「1.2.1 過年度における電子公文書等に係る取組経緯」に示すとおり、館では「電子公文書等の移管・保存・利用システムの次期システムの要件検討に向けて」（平成 26 年 3 月）において、バックアップメディアとして長期保存用光ディスクを採用するとともに、その取扱いに係るノウハウを蓄積しながら、引き続き長期保存用光ディスクに係る動向を注視していくこととした。次期システムの検討に向けて、技術動向等を踏まえた検証が求められる。

問題 2-7： 電子ファイルの処理エラーの発生事例等に関する技術情報の蓄積に努め、業務上の技術レファレンスとして活用することとしていたが、そのための仕組みが体系化されていない

電子ファイルのフォーマット変換時に発生する処理エラーについて、館では、各処理エラーの発生事例等に関する技術情報の蓄積に努め、受入れ等に係る業務上の技術レファレンスとして活用することとしていたが、当該取組が館職員の個別・属人的な活動に留まっており、館としての組織的及び体系的な活動として定義されていない。

2.2.3.3. 電子公文書等システムの運用に関する問題

「2.2.2 電子公文書等システムの概要」にて示した現行システムについて、顕在化している問題を抽出し、整理した。

問題 3-1： 各業務の実施順序がシステムにより規定されているため、効率的な業務実施ができない

現在の受入れ等に係る業務は、現行システムの構成や機能を前提として業務手順が定められており、効率的な業務遂行に向けた業務改善にあたり制約となっている。

問題 3-2： 手動での操作による定型的な作業があり、ミスが発生しやすい

現行システムの定型的な作業のうち、システム間のデータ受け渡しや確認結果の様式への転記など、手動で操作する作業が残存しており、ミスが発生しやすい。

問題 3-3： オンプレミス型で構築・運用されており、クラウド・バイ・デフォルト原則の政府方針と整合していない

現行システムは、低コストで原本性を保証するため、外部ネットワークに接続していないオンプレミス型^②で構築・運用されているが、情報システムに係る政府の方針及び技術動向を踏まえて適切なシステム方式を選定する必要がある。

問題 3-4： 受入れデータが想定している容量以上の場合があり、システム内に保存できない

現行システムでは、ストレージの容量に鑑み、受け入れる電子ファイルの容量に上限を設けている。移管される電子公文書等で、1つの行政文書ファイルとして想定している容量以上となった場合、結果として、システム内に保存できない電子ファイルが発生している。

問題 3-5： システムの性能上の制約により、フォーマット変換処理に時間を要し、業務スケジュールの計画的な実施に支障となる懸念がある

現行システムの性能上の制約から、電子ファイルのフォーマット変換に時間を要することがあり、今後行う業務の円滑な遂行や利用に向けた速やかな目録公開に支障となる懸念がある。

2.2.4. 館における電子公文書等の保存に係る課題

「2.2.3 館における電子公文書等の保存に係る問題」から、あるべき姿とのギャップである課題の抽出を行う。はじめに、前提となる電子公文書等の保存を取り巻く外部環境について整理する。

2.2.4.1. 課題抽出の前提となる外部環境の変化

2.2.4.1.(1) 受け入れる電子公文書等の増加

「2.1.1 電子公文書等の管理に係る方針」で示した電子文書の管理に係る政府方針により、今後、館に移管される電子公文書等の件数は大幅に増大することが想定される。

館で受け入れる行政文書ファイルの件数は、平成 29 年度から平成 30 年度までの 1 年間では 33%増加している。また、運用開始の翌年である平成 24 年度から平成 30 年度までの間では、年平均で約 58%増加している。平成 24 年度から平成 30 年度までに移管された行政文書ファイル数の実績値、そして前述の増加率を高く予測した場合と後者の増加率を低く予測した場合の令和元年度以降における予測傾向を以下に示す。

^② オンプレミス型とは、情報システムに必要な設備を自社内やデータセンター等に設置し、情報システムの構築・運用を、自社で主体的に行う運用形態のこと。

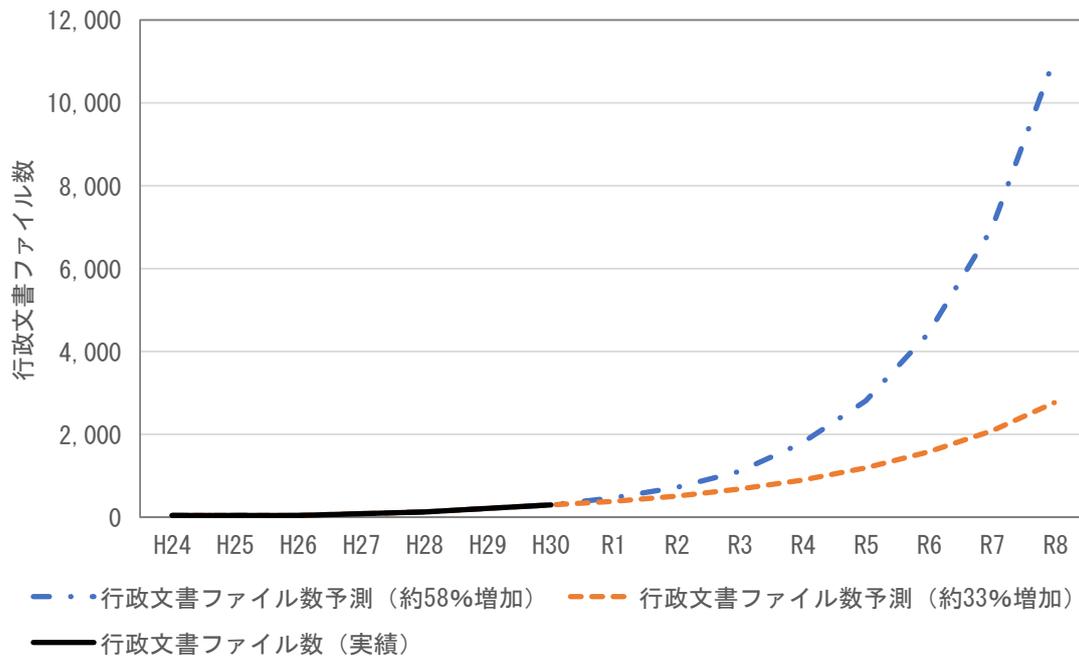


図 2-4 行政文書ファイル数の実績と今後の予測

2.2.4.1.(2) 受け入れる電子公文書等の多様化

デジタル化の推進やソーシャルメディアの定着に伴い、作成する行政文書には様々な種類・形式が含まれ、媒体のマルチメディア化の進展、また業務を行う環境も変化している。そのため、今後、館に移管される電子公文書等のフォーマットは更に多様化することが見込まれる。先述したとおり、「表 2-5 電子公文書等システムで保存する電子ファイルのフォーマット概況」で移管される電子ファイルのうち約 30%が非標準フォーマットであり、今後はこの割合が更に増えることが想定される。

2.2.4.2. 電子公文書等に係る館の保存対策に関する課題

館の保存対策に関する課題を以下に示す。

表 2-7 電子公文書等に係る館の保存対策に起因する課題

問題	課題
問題 1-1: 移管を想定するフォーマットを標準的フォーマットとしていたが、非標準フォーマットによる移管が一定数存在している。	課題 1: 館のフォーマット対応表を移管元機関に示し、標準的フォーマットによる移管を推進する必要がある。
問題 1-2: 標準的フォーマット及び長期保存フォーマットについて、受入れ実績及び技術動向を反映した見直しをしていない。	課題 2: 標準的フォーマット及び長期保存フォーマットの見直しに係る運用ルールを定め実施する必要がある。
問題 1-3: 標準的フォーマット及び長期保存フォーマットを見直しするための手順が定められていない。	
問題 1-4: 受入れ時のフォーマットをシステム内に保存することが原則とされていない。	課題 3: 受入れ時のフォーマットによる保存を、電子公文書等に係る館の保存対策として明文化する必要がある。

2.2.4.3. 受入れ等に係る業務の実施に関する課題

受入れ等に係る業務の実施に関する課題を以下に示す。

表 2-8 受入れ等に係る業務に起因する課題

問題	課題
問題 2-1: 受入れ業務において、移管対象外となる電子ファイルを検出できない。	課題 4: 業務上必要な機能を追加する必要がある。
問題 2-2: 受入れ業務において、一元的な文書管理システムと手動のダウンロードによる連携となっており、時間を要している。	
問題 2-3: フォーマット変換業務において、長期保存フォーマットへの変換後の品質確認が、不適合となる場合があり、再変換作業に時間を要している。	
問題 2-4: フォーマット変換業務において、非標準フォーマットはフォーマットの種類を判定できていない。	
問題 2-5: 運用管理業務において、作業記録管理機能に一覧性がなく、計画に基づく進捗確認等の機能が無いため、精度の高い運用管理ができない。	

2.2.4.4. 電子公文書等システムの運用に係る課題

電子公文書等システムに係る問題と課題を以下に示す。

表 2-9 電子公文書等システムに起因する課題

問題	課題
問題 3-1: 各業務の実施順序がシステムにより規定されているため、効率的な業務実施ができない。	課題 5: BPR による業務効率性向上を検討する必要がある。
問題 3-2: 手動での操作による定型的な作業があり、ミスが発生しやすい。	
問題 3-3: オンプレミス型で構築・運用されており、クラウド・バイ・デフォルト原則の政府方針と整合していない。	課題 6: 業務の特性やコスト、情報セキュリティなどクラウド利用の実現性を確認した上で、システム方式と要件を決定する必要がある。
問題 3-4: 受入れデータが想定している容量以上の場合があり、システム内に保存できない。	課題 7: 規模の拡張性を要件として定義する必要がある。
問題 3-5: システムの性能上の制約により、フォーマット変換等のシステム処理に時間を要し、業務スケジュールの計画的な実施に支障となる懸念がある。	課題 8: 性能の拡張性を要件として定義する必要がある。

2.2.4.5. 過年度の調査報告のフォローアップに係る課題

「表 1-1 電子公文書等の保存に係る過年度の主要な取組」に示したとおり、過年度に実施した館の調査報告において、受入れ等に係る業務の見直しに向けて、社会情勢及び技術動向を踏まえて適宜適切な措置を講じることとした。この趣旨を踏まえ、当該調査報告の内容に関して、以下の観点でフォローアップが必要である。

課題 9： 技術動向及びバックアップ要件の変化を踏まえ、システム外部に保存する媒体として適切な媒体を選択する必要がある。

課題 10： 電子ファイルの処理エラーの発生事例に関する技術情報の蓄積と技術レファレンスとしての活用に向けた手順を定める必要がある。

2.3. 専門家ヒアリング

2.3.1. ヒアリングの概要

館における受入れ等に係る業務の現状、問題及び課題を踏まえて、専門家へのヒアリングを実施した。

表 2-10 専門家に対するヒアリング実施概要

実施日時	ヒアリング先	目的
2019年11月19日	杉本重雄名誉教授	受入れ等に係る業務の現状を説明し、本調査の視点等に対する示唆を得る。
2019年11月25日	森本祥子准教授	
2020年2月19日	杉本重雄名誉教授	館にて検討・整理した課題及び対応案を説明し、対応の策定に係る示唆を得る。
2020年2月21日	森本祥子准教授	

2.3.2. 専門家からの示唆

専門家に対して実施した合計4回のヒアリングから、館における受入れ等に係る業務への対応に係る以下の示唆を得た。なお、示唆された各取組の実施にあたっては、電子公文書等に係る館としての保存対策を明確化することが重要である。

2.3.2.1. 現用から非現用までの連続した管理の実現

公文書管理においては連続性を意識することが重要である。移管元行政機関等においても、長期の電子公文書等の管理と保存が求められており、更に延長の必要があると認める場合には定める期間が経過するまで長期保存を行った上で、館へ移管されることになる。このことも踏まえ、公文書の現用段階から非現用まで連続した管理を意識した業務設計を行うべきである。特に、電子公文書等の場合、従前の紙による運用よりも、現用段階との連携を意識した業務設計が特に重要となる。

連続性を意識した業務設計にあたっては、館が移管元行政機関等における現用段階の電子公文書等の情報を常時把握可能とすることが重要である。これにより、館における受入れ等に係る業務の計画的な実施が可能となり、結果として、我が国における公文書管理全体の業務量の削減及びコスト削減に寄与するものと考えられる。

また、移管元行政機関等が、公文書の作成段階から館への移管を前提としてファイルフォーマットを選択することも、長期的な業務量削減及びコスト削減へ効果が期待できる。

2.3.2.2. 利用可能な状態を維持管理する環境の整備

電子公文書等の保存にあたっては、利用可能な状態を長期的に保つための維持管理が重要である。具体的には、館が有する電子公文書等の維持管理に係る知識及びノウハウを組織的・体系的に蓄積し、継承する仕組みを整えとともに人材育成に活かすべきである。

また、海外と比較すると、日本においてはフォーマットに関する技術情報や取扱いの経験などの各種情報を蓄積する体系的な仕組みが存在しないことが課題である。将来的な見読性の確保のために、フォーマットに係る各種情報を集積したデータベースを整備することが必要である。

2.3.2.3. 技術進歩等を踏まえた運用見直し

技術の進歩等を踏まえて、取り扱うフォーマットについては定期的に見直しを行うべきである。館として保存している電子ファイルの見読性を将来的にも維持し続けるためには、保存している各フォーマットについての技術情報、館の保存対策及び変換に係る実務等を継続的に更新していくことが必要である。

なお、取り扱うフォーマット等の見直しについて、電子公文書等システムの機能等が制約となることも想定されるため、原則として、システム更改のタイミングと歩調を合わせるものが現実的である。

2.3.2.4. 受入れ時のオリジナルフォーマットの保存の必要性

館が行う電子公文書等システムの保存において、受入れ時のオリジナルフォーマットでしか保持していない電子ファイルを含む電子公文書等があり、館の環境において再生及び内容を見読することが不可能になるリスクがある。そうした場合であっても、将来的な保存と利用を可能にするため、受入れ時の電子ファイルについてフォーマットを変換した長期保存に適したフォーマットに加え、受入れ時のオリジナルフォーマットを保存しておくことが必要である。

2.4. 海外における電子公文書等の保存に係る文献調査

電子公文書等の適切な保存に向けた対応について検討するにあたり、海外における電子公文書等の保存の現状を文献等調査により整理した。

海外の事例から得られた示唆を以下に示す。

2.4.1. 電子公文書等に係る館の保存対策に対する示唆

2.4.1.1. 受入れフォーマット及びその見直し

米国及び英国では、長期保存に適したフォーマットを提示し、移管されるファイルフォーマットの定義や整理を行っている。これらの事例からは、長期保存及び見読性の確保が容易なファイルフォーマットを推奨しておくことで、移管以前の連続した管理の重要性が示唆される。

2.4.1.1.(1) 米国の電子公文書管理におけるファイルフォーマット

アメリカ国立公文書記録管理院（National Archives and Records Administration、以下「NARA」という。）では、電子記録の作成・保存に用いられるフォーマットを長期保存上の持続可能性などの観点から評価し、「望ましい」又は「受入れ可能」等の類型化をした上で、受入れ可能なファイルフォーマットを示している⁽³⁾。「望ましい」又は「受入れ可能」としているものと異なるファイルフォーマットで記録を維持管理している機関があった場合には、永久保存対象の記録に対してフォーマット変換や変更等を実施する前に、NARAに連絡することを求めている。また、電子ファイルの保存において、文字フォントの埋め込みについて言及している。

表 2-11 NARA が保存対象とするフォーマットタイプ

No	フォーマットタイプ
1	CAD
2	デジタルオーディオ
3	デジタル動画: デジタルシネマ
4	デジタル動画: デジタルビデオ
5	デジタル静止画: デジタル写真
6	デジタル静止画: スキャンテキスト
7	デジタルポスター
8	地理空間 (Geospatial) フォーマット
9	プレゼンテーション・フォーマット
10	テキストデータ

⁽³⁾ The U.S. National Archives and Records Administration. “Format Guidance for the Transfer of Permanent Electronic Records. Bulletin 2014-04 Revised August 2018 per NARA Bulletin 2018-01”.
<https://www.archives.gov/records-mgmt/bulletins/2014/2014-04.html>, (accessed 2020-06-30).

No	フォーマットタイプ
11	構造化データ・フォーマット
12	電子メール
13	Web 記録
14	カレンダー

また、NARA では、保存対象とするファイルフォーマットに対して、リスク評価による受入れフォーマットの見直しを検討しており、今後も注視が必要である⁽⁴⁾。フォーマットのリスク評価は、具体的には長期的な見読性確保に係るリスクと保存の優先順位の観点で評価を行っている。評価にあたり加点要素としている観点の例を以下に示す。

- 広く採用されていること
- フォーマットに関する文書が利用できること
- ドキュメント化できること
- 特定のソフトウェアに依存しないこと
- 技術的な保護（例えば、暗号化など）の要件がないこと

同じく、減点要素としている観点の例を以下に示す。

- フォーマットの寿命
- ハードウェア又はソフトウェアへの依存性

2.4.1.1.(2) 英国の電子公文書管理におけるファイルフォーマット

イギリス国立公文書館（The National Archives、以下「TNA」という。）では、現時点で長期にわたって維持できる範囲を「移管用フォーマット」として定義しており、これらについては受入れ、保存及び公開を行うと明文化している⁽⁵⁾。

表 2-12 TNA が保存対象とするフォーマットタイプ

No	フォーマットタイプ
1	ワードプロセッサ及びテキスト・フォーマット
2	スプレッドシート・フォーマット
3	プレゼンテーション・フォーマット
4	グラフィック・フォーマット
5	オーディオ・フォーマット
6	ビデオ・フォーマット
7	電子メール・フォーマット

⁽⁴⁾ The U.S. National Archives and Records Administration . “U.S. National Archives Digital Preservation Framework”. <https://github.com/usnationalarchives/digital-preservation>, (accessed 2020-06-30).

⁽⁵⁾ The National Archives. “File formats for transfer”. <https://www.nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer/>, (accessed 2020-06-30).

No	フォーマットタイプ
8	その他フォーマット

なお、「移管用フォーマット」に指定されていないフォーマットの移管についても、TNA Aとして受入れ可能であるとしている。

2.4.1.2. 受入れ時のオリジナルフォーマットの保存

英国 TNA では、各機関により作成された電子記録のオリジナルフォーマットを保存した上で、提供時には当該ファイルフォーマットの利用・再生環境を維持する「エミュレーション」技術を用いることを原則とすることで長期的な見読性を確保している。フォーマット変換は、保存上のリスクを管理する方法が他に無い場合のみに用いることとしている⁽⁶⁾

一方、米国 NARA では、保存するデジタルコンテンツのファイル管理として、オリジナルフォーマットの重要な性質を保つフォーマットを選択し、標準的フォーマットに積極的に変換（ノーマライズ）することにより、管理するファイルフォーマットを最小限に絞り込むこととしている。なお、オリジナルフォーマットの電子ファイルも並行して保存しており、長期保存用のストレージとアクセス頻度が少ないストレージとを区別する形で保存している⁽⁷⁾。

今回調査対象の他国と比べて、館は「望ましい」や「受入れ可能」等のフォーマットの類型化が明確化されていない。また、館においては、海外でリスク評価において「危険」や「要評価」と判断されている種類のフォーマットを問わず受入れしており、システムの運用と業務で対応しているのが現状であり、長期保存のための変換ができない場合、電子公文書等を利用に供せない可能性がある。そのため、館における受入れ等に係る業務においても、長期保存及び見読性の確保が容易なファイルフォーマットを推奨するとともに、受入れ時のオリジナルフォーマットについても保存を行うことが有用と考えられる。

2.4.2. 受入れ等に係る業務の実施に対する示唆

英国 TNA では、移管される電子公文書等のフォーマット判定に機械学習を用いており、プレーンテキストファイル形式のフォーマット識別で高い精度を達成している⁽⁸⁾⁽⁹⁾。バイナリファイルのファイルフォーマット識別に関する方法はいくつか開発されているものの、テキストファイルのファイル形式を識別するための方法はこれまで開発されてこなかった。そこで、TNA では、教師あり学習モデルの機械学習を利用し、Python、Java、TXT、CSV、TSV の 5 つの形式を識別するためのプロトタイプを作成した。当該プロトタイプでは、最大 98.58% の識別精度を達成しており、今後の展開が期待される。

⁽⁶⁾ The National Archives, “Digital Strategy”. 2017.

<https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf>, (accessed 2020-06-30).

⁽⁷⁾ National Archives and Records Administration. “Digital Preservation Strategy”. 2017.

<https://www.archives.gov/preservation/electronic-records/digital-preservation-strategy>, (accessed 2020-06-30).

⁽⁸⁾ Santhilata Kuppli Venkata. “Motivation to Undertake File Format Identification Research for Plain Text Files”. 2019-09-02. <https://www.dpconline.org/blog/file-format-id-research>, (accessed 2020-06-30).

⁽⁹⁾ Santhilata Kuppli Venkata. “How to correctly identify the file type of a text file from its contents?”. 2019-09-13. <https://www.dpconline.org/blog/identify-file-type-of-a-text-file>, (accessed 2020-06-30).

次期システムにおいても、機械学習を始めとした先進的な ICT 技術を活用することによる業務効率性向上を検討する必要がある。

2.4.3. 電子公文書等システムの運用に対する示唆

米国 NARA では、現行の電子記録管理システムにおいて、変化する業務プロセスや多種多様な電子記録の保存に対し、システムの柔軟性を向上する必要があることから、次期電子記録管理システム（ERA）の開発にアジャイル手法やクラウド技術を採用している⁽¹⁰⁾。現在のワークフローや機能の提供におけるシステム的な制約から、記録の処理に必要な新しいソフトウェアツールが統合しやすいこと、また利用ニーズに応じてリソースの拡張が可能であることから、システムの拡張性を確保できるとしている。館における次期システムの要件定義においても、特に規模や性能の拡張性において、クラウドサービスの利活用は重要な対応となり得る。

⁽¹⁰⁾ Leslie Johnston. “ERA 2.0: The National Archives' new framework for electronic records preservation”. Proceedings of the Association for Information Science and Technology. 2017, vol.54, p.197-202. <https://doi.org/10.1002/pr2.2017.14505401022>, (accessed 2020-06-30).

2.5. 館の課題と対応の方向性

館における受入れ等に係る業務の課題、専門家及び海外事例からの示唆を踏まえ、電子公文書等の適切な保存に向けて取り組むべき対応の検討にむけて、「政府・館の方針」、「電子公文書等に係る館の保存対策」、「電子公文書等に係る館の保存対策」、「受入れ等に係る業務の実施」及び「電子公文書等システムの運用」の4つの方向性で整理した。

2.5.1. 政府・館の方針に対する対応

電子公文書等の保存に係る政府の方針について、今後、中長期的に期待される対応を以下に示す。

- 移管元行政機関等での現用時の文書管理状況の把握
- 移管元行政機関等に対する電子公文書等の管理及び保存にあたっての依頼
- 電子公文書等の管理に係る方針の適宜見直し

各対応の詳細については、「6. あるべき姿へ向けて」を参照されたい。

2.5.2. 電子公文書等に係る館の保存対策に対する対応

電子公文書等の保存に係る館の方針について、今後、実施すべき事項を以下に示す。

- 保存対策のPDCAサイクルの確立
- フォーマット対応表の見直し
- オリジナルフォーマットの原則保存

各対応の詳細については、「3. 電子公文書等に係る館の保存対策に対する対応」を参照されたい。

2.5.3. 受入れ等に係る業務の実施に対する対応

受入れ等に係る業務について、今後、実施すべき事項を以下に示す。

- 受入れ等に係る業務の効率化
- フォーマット変換処理結果の確認観点の明確化
- 電子公文書等のフォーマットに係る技術情報蓄積

各対応の詳細については、「4. 受入れ等に係る業務の実施に対する対応」を参照されたい。

2.5.4. 電子公文書等システムの運用に対する対応

電子公文書等システムについて、今後、実施すべき事項を以下に示す。

- 次期システムの更改
- 外部環境変化を踏まえた要件の具体化
- 電子公文書等に係る館の保存対策を踏まえた機能改善

各対応の詳細については、「5 電子公文書等システムの運用に対する対応」を参照されたい。

2.5.5. 課題と対応の関係

各対応と、これまで述べてきた課題、専門家からの示唆及び海外事例からの示唆の関係整理を次頁に示す。

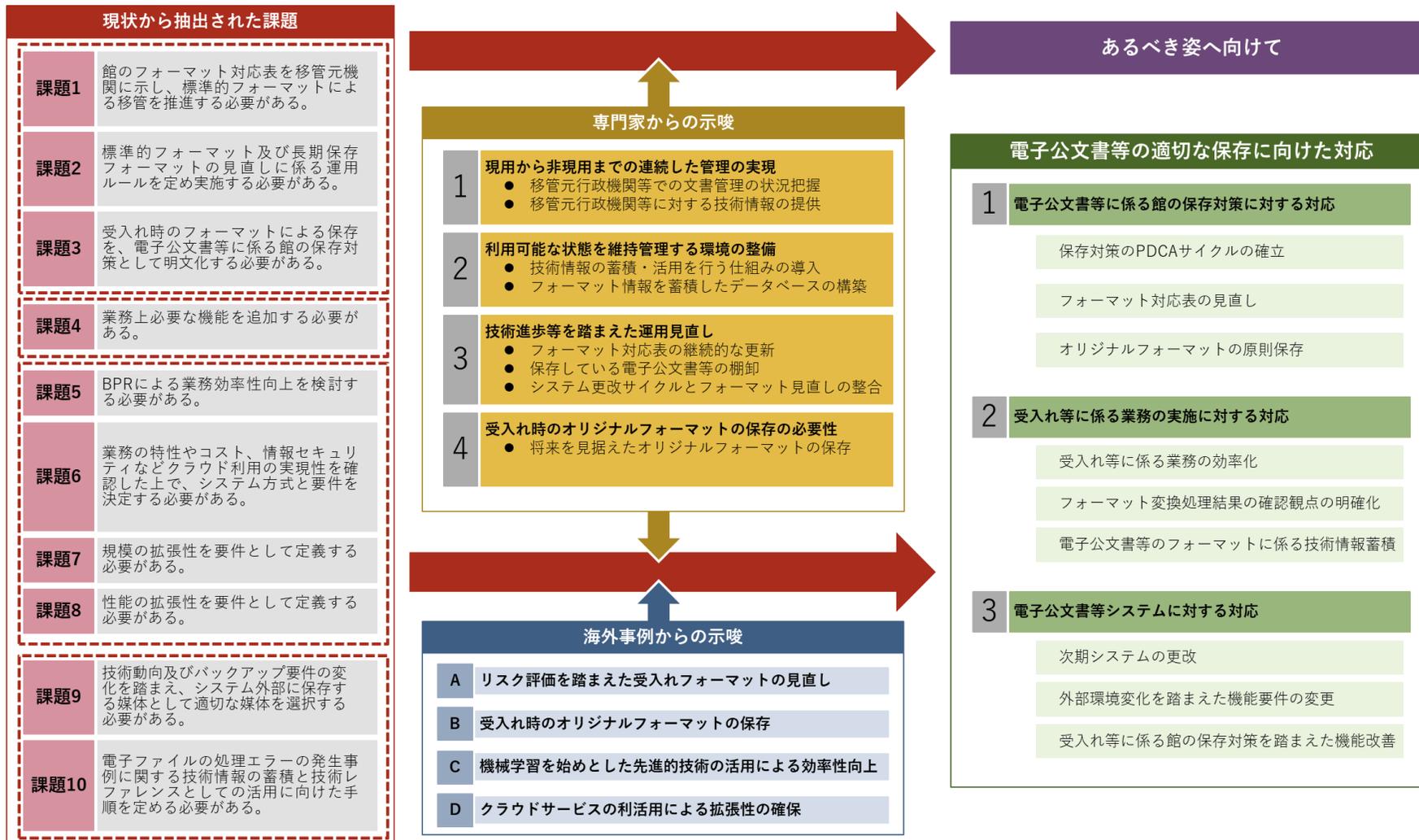


図 2-5 課題と対応の関係

3. 電子公文書等に係る館の保存対策に対する対応

3.1. 保存対策の PDCA サイクルの確立

政府では、「デジタル・ガバメント推進標準ガイドライン」（平成 31 年 2 月 25 日、各府省情報化統括責任者（CIO）連絡会議決定）において、個々のプロジェクトを全体的かつ適正に管理することで、情報システムに係る課題解決のみならず、各組織の目標を達成し、個々のプロジェクトをマネジメントするだけでは出し得ない価値（例：便益の実現、リスク及びリソースの適正化等）を生み出すことの重要性を示している。

電子公文書等に係る館の保存対策についても、システムの運用及び同システムを通じた受入れ等に係る業務、それらの実績を適切に評価し、必要な見直しを実施する PDCA サイクルの確立を目指す必要がある。

受入れ等に係る業務の PDCA サイクルと、システムのライフサイクルの関係を以下に示す。

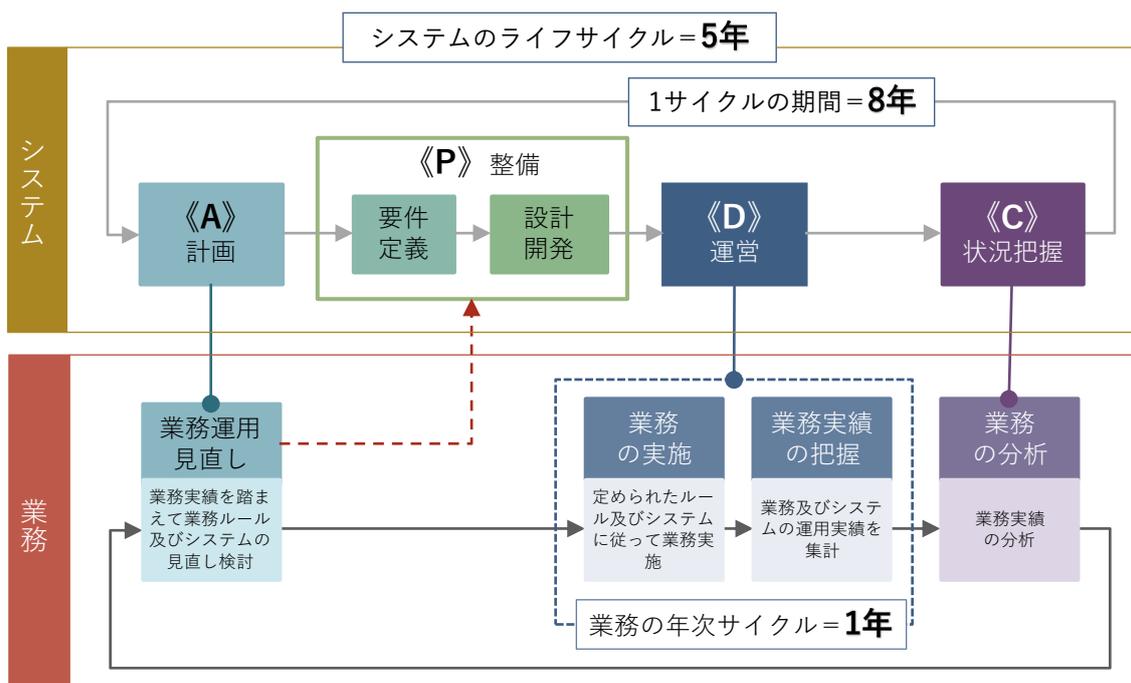


図 3-1 業務の PDCA サイクルとシステムのライフサイクル

なお、電子公文書等システムの更改について、更改前後で切れ目なくシステムを運用するためには、更改後のシステム運用開始の3年前程度を目安に計画を開始する必要がある。そのため、更改前のシステムに対する状況把握についても、プロジェクト活動の運営終了後ではなく、より早いタイミングで集計・整理を行う必要がある。

これらの実態を踏まえた、中長期的なシステム更改とプロジェクトの運営スケジュール（イメージ図）を以下に示す。

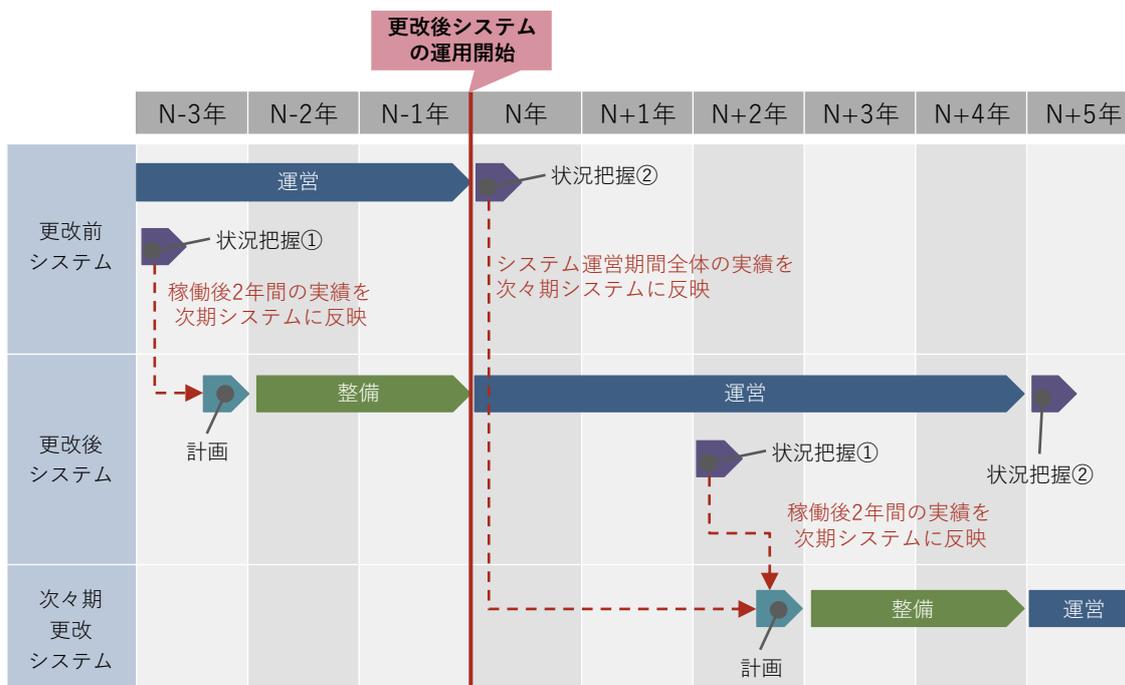


図 3-2 システム更改に向けたプロジェクト運営スケジュールのイメージ

なお、上図において、「状況把握 ①」は次期更改に向けてプロジェクトの運営途中の段階で実績に基づく状況把握を行うものであり、「状況把握 ②」は5年間のシステム運用後にプロジェクト全体の活動実績を包括的にとりまとめるものである。

3.2. フォーマット対応表の見直し

3.2.1. フォーマット対応表の見直しに係る運用ルール策定

館のフォーマット対応表（「表 2-2 現行のフォーマット対応表」参照）で定義されるフォーマット等については、一過性の更新ではなく、継続的に見直しが必要である。また、「3.1 保存対策のPDCA サイクルの確立」を踏まえて運用ルールを策定する必要がある。なお、見直しの実施にあたっては、見直しを効率的に実施するための機能を電子公文書等システムに実装することを始め、電子公文書等の受入れ件数が増加傾向であることを踏まえて対応する必要がある。

3.2.1.1. フォーマット対応表の見直し方針

3.2.1.1.(1) 見直しの頻度

フォーマット対応表の見直しは、最終的に電子公文書等システムでの機能要件に反映することを鑑み、システムのライフサイクルに合わせて実施することが妥当と判断する。ただし、各年度で移管される電子ファイルのファイルフォーマットに係る傾向の把握については、年間の業務サイクルに組み込み、年次で実施する必要がある。

3.2.1.1.(2) 見直しの基準

フォーマット対応表の見直しは、電子公文書等システムで保存している電子公文書等の移管実績の集計結果及び今後移管される予定の電子公文書等のファイルフォーマットに係る統計情報に基づくことが妥当である。後者については、電子公文書等の移管前に移管元行政機関等から当該情報の提供を受けた上で、フォーマット対応表見直しの実施可否を判断できることが望ましい。

3.2.1.1.(3) フォーマット対応表の見直しに係る作業項目

上記の検討を踏まえ、フォーマット対応表の見直しにあたって必要な作業を以下に示す。

表 3-1 フォーマット対応表の見直しに係る作業項目案

No	作業項目	頻度及び タイミング	目的	内容
作業 1	フォーマットに関する情報修正	フォーマット 変換時毎	フォーマット定義マスタの更新	フォーマット判定において、システム機能のフォーマット判定が出来ず、かつ事前定義マスタにも情報が無い場合に、事前定義マスタへの追加登録を行う。
作業 2	年度実績集計	1回/年 各年度末に 実施	電子公文書等システムに保存されている電子ファイルやフォーマット情報に関する統計データの取得	電子公文書等システムに保存されている電子ファイルやフォーマット情報について、抽出機能を利用して統計データを取得する。

No	作業項目	頻度及び タイミング	目的	内容
作業 3	技術動向の把握	1 回／年 各年度末に 実施	フォーマット管理情報の最新化	電子公文書等システムの運用業者から各フォーマットの技術動向、及びフォーマット変換ソフトウェアの技術動向の報告を受ける。
作業 4	フォーマット判定の再実行	1 回／5 年 システム更改の計画 の前段で実施	フォーマット管理情報の最新化	電子公文書等システムのフォーマット判定の再処理機能を利用し、電子公文書等システムに保存されている電子ファイルのフォーマット管理情報を更新する。なお、判定前のフォーマット管理情報は履歴として保持する。
作業 5	プロジェクトの 運営実績集計	1 回／5 年 システム更改の計画 の前段で実施	電子公文書等システムに 保存されている電子ファイル やフォーマット情報に関する 統計データの取得	電子公文書等システムに保存されている電子ファイルやフォーマット情報について、抽出機能を利用して統計データを取得する。
作業 6	非標準フォーマットの 見直し	1 回／5 年 システム更改の計画 時に実施	非標準フォーマットの標準 的フォーマットへの追加検 討	実績集計の結果を基にして、非標準フォーマットの登録件数や増加傾向を確認し、標準的フォーマットの追加検討を行う。
作業 7	標準的フォーマットの 見直し	1 回／5 年 システム更改の計画 時に実施	標準的フォーマットの技術 情報更新	標準的フォーマットについて、技術動向等を踏まえて、下記の確認を行う。 ● 標準的フォーマットのバージョンアップ情報 ● 標準的フォーマットのサポート状況 ● フォーマットリスク確認(NARA 等の情報)
作業 8	長期保存フォーマットの 見直し	1 回／5 年 システム更改の計画 時に実施	長期保存フォーマットの代 替フォーマットの有無の確 認	長期保存フォーマットについて、技術動向等を踏まえて、下記の確認を行う。 ● フォーマットのサポート状況・フォーマットリスク確認(例:NARA の情報等) ● 確認の結果、代替フォーマットやフォーマットリスクが高い場合は、代替フォーマットへの変換計画や措置を検討する。
作業 9	フォーマット対応表の 見直し	1 回／5 年 システム更改の計画 時に実施	各フォーマットの見直し結 果のフォーマット対応表へ の反映	非標準フォーマット、標準的フォーマット及び長期保存フォーマットの各フォーマットのリスク評価の実施、見直し結果をフォーマット対応表に反映する。

上記の作業項目について、PDCA サイクルとの関係のイメージを次頁に示す。

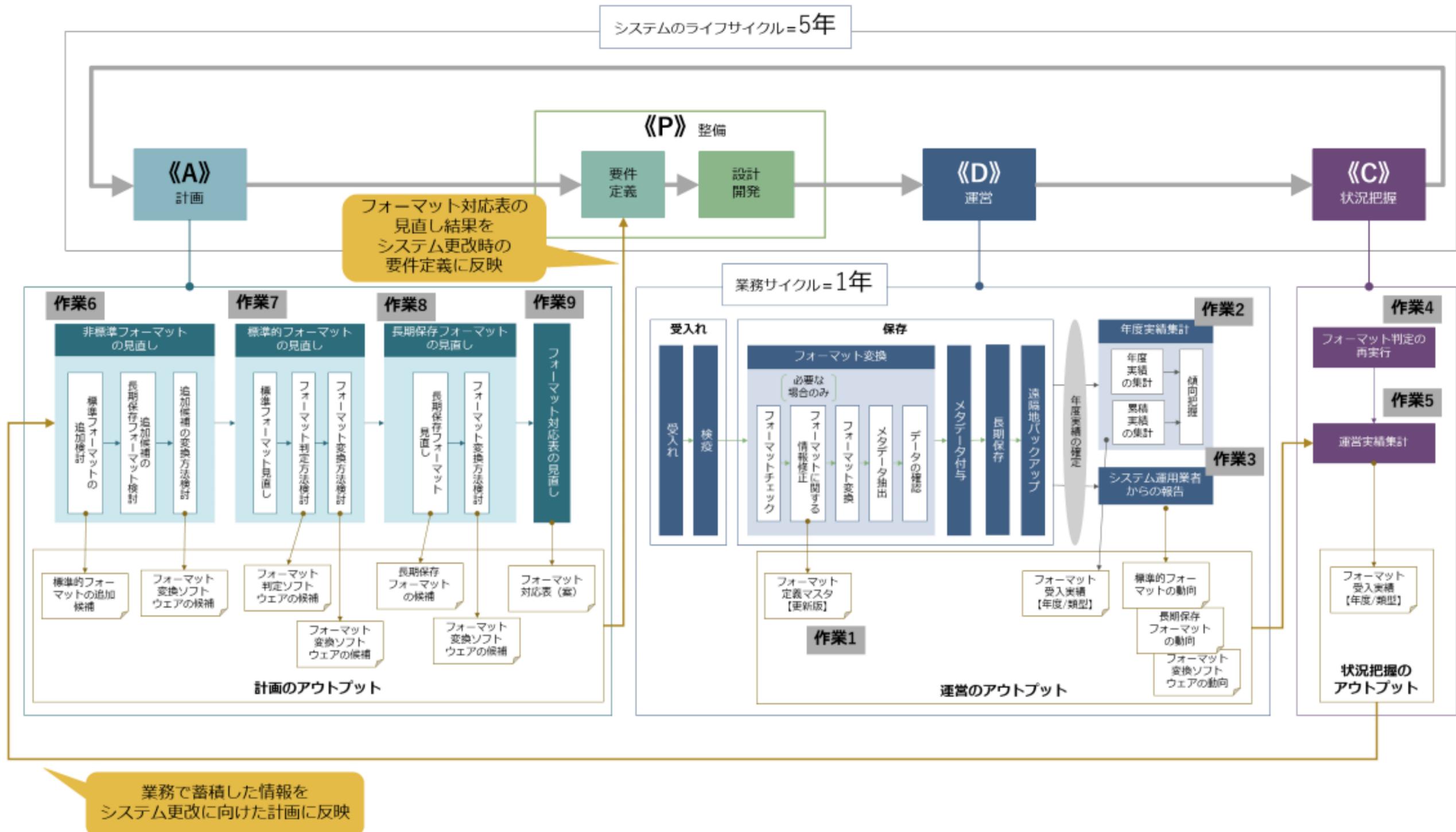


図 3-3 フォーマット対応表の見直し等に関するPDCAサイクルのイメージ

3.2.2. 現行のフォーマット対応表の見直し

3.2.2.1. 運用ルールに基づくフォーマット対応表の見直し

「3.2.1 フォーマット対応表の見直しに係る運用ルール策定」で示した運用ルールに従い、現行のフォーマット対応表の見直しを実施した。

3.2.2.1.(1) プロジェクトの運営実績集計

「表 2-5 電子公文書等システムで保存する電子ファイルのフォーマット概況」で概要を示すとおり、電子公文書等システムでこれまでに受入れた電子ファイルのファイルフォーマットについて集計した。

実績集計結果から整理した非標準フォーマットの移管実績を以下に示す。

表 3-2 非標準フォーマットの移管実績

フォーマット種類	移管電子ファイル数
テキスト形式(TXT、CSV 等)	約 15,000
Web ページ(HTML、スクリプト形式ファイル等)	約 3,000
CAD(HOC、JWW、SFC 等)	約 800
OS 関連ファイル(bin、chm 等)	約 300
その他ファイル(電子メール、画像編集ソフトウェア及びデータベース等)	約 300

3.2.2.2. 非標準フォーマットの見直し

「表 3-2 非標準フォーマットの移管実績」を踏まえて、非標準フォーマットのうち、標準的フォーマットへの追加が必要と考えられるファイルフォーマットの有無を確認した。非標準フォーマットを標準的フォーマットへ追加する際に確認した観点を以下に示す。

表 3-3 標準的フォーマットへの追加に係る確認観点

No	項目	補足(確認方法等)	
1	移管数	移管実績数は多いか。	移管実績を確認する。
2		移管実績が一過性のものではないか。	移管実績を確認する。
3	標準規格	国際標準、国内標準又はデファクトスタンダードであるか。	
4	環境の用意	いつまでに受け入れ環境を用意すればよいか。	運用の特性上、現用から移管されるまで時間があるため、受入れを行う実施時期を見てシステム更改等での対応計画を整理する。
5	長期保存の方法	適切な長期保存フォーマットが存在するか。	長期保存方法についての検討を行う。
6		長期保存フォーマットへの変換は可能か。	長期保存方法についての検討を行う。

なお、非標準フォーマットの見直しにあたっては、「表 3-3 標準的フォーマットへの追加に係る確認観点」への該当性の確認結果から、追加の検討対象となるフォーマットを抽出し、対応に係るコストや移管元行政機関等の状況なども踏まえて、判断を行う。

上記観点に基づく移管実績を有するフォーマット種類に対する評価結果を以下に示す。

3.2.2.2.(1) テキスト形式

テキスト形式の電子ファイルは、バイナリファイルと異なりシンプルな構成であるため、再生環境の影響が小さく見読性に優れている。また、ISO/IEC 8859-1:1998⁽¹¹⁾に準拠しており、汎用性に優れている。米国 NARA や英国 TNA でも、受入れを推奨するフォーマットとして、テキストフォーマットのうち、Plain Text 及び CSV (Comma Separated Values) 形式を定義している。

館におけるテキストのフォーマット毎の移管実績を以下に示す。

表 3-4 テキスト形式のフォーマット毎の移管実績

フォーマット	移管電子ファイル数
TXT	約 13,000
CSV	約 2,000
XML	約 180
XSL	約 30

テキスト形式に関する確認結果を以下に示す。

表 3-5 テキストの標準的フォーマット追加の確認結果

No	項目	確認結果
1	移管実績数	移管実績数は多いか。
		約 15,000
2		移管実績が一過性のものではないか。
		H23 年度～H30 年度までの 6 年間にわたり移管がされている。
3	標準規格	国際標準、国内標準又はデファクトスタンダードであるか。
		ISO 等で規定されている。
4	環境の用意	いつまでに受け入れ環境を用意すればよいか。
		既に移管実績があり対応が必要である。
5	長期保存の方法	適切な長期保存フォーマットが存在するか。
		長期的な見読性の観点からオリジナルフォーマットでの長期保存が妥当である。
6		長期保存フォーマットへの変換は可能か。
		オリジナルフォーマットのため、フォーマットの変換は必要ない。

これより、テキスト形式として移管実績を有するフォーマットを標準的フォーマットに追加するとともに、オリジナルフォーマットを長期保存フォーマットとすることが妥当である。

⁽¹¹⁾ ISO/IEC 8859-1:1988. Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1. <https://www.iso.org/standard/28245.html>, (accessed 2020-06-30).

3.2.2.2.(2) Web ページ

Web ページ形式のデータは、HTML (Web ページを記述するためのマークアップ言語) や JavaScript といったスクリプト言語の電子ファイルを含み、テキストファイルの拡張である。テキストファイルと同様にシンプルな構成であるため、再生環境の影響が小さく見読性に優れている。また、ISO/IEC 15445:2000⁽¹²⁾に準拠しており、汎用性に優れている。

館における Web ページに関するフォーマットの移管実績を以下に示す。

表 3-6 Web ページのフォーマット毎の移管実績

フォーマット	移管電子ファイル数
HTML	約 3,000
CSS	約 100
DTD	約 20
JavaScript	約 20

Web ページに関する確認結果を以下に示す。

表 3-7 Web ページの標準的フォーマット追加の確認結果

No	項目		確認結果
1	移管実績数	移管実績数は多いか。	約 3,000
2		移管実績が一過性のものではないか。	H23 年度～H30 年度まで全ての年度において移管されている。
3	標準規格	国際標準、国内標準又はデファクトスタンダードであるか。	ISO 等で規定されている。
4	環境の用意	いつまでに受け入れ環境を用意すればよいか。	既に移管実績があり対応が必要である。
5	長期保存の方法	適切な長期保存フォーマットが存在するか。	Web ページを構成する電子ファイルは様々であり、各電子ファイルのフォーマットの変換方針に従うこととする。
6		長期保存フォーマットへの変換は可能か。	Web ページを構成する電子ファイル毎に各フォーマットの状況に依存する。

これまでの館への移管実績及び米国 NARA において Web ページに係る電子ファイルを保存対象としていることを踏まえ、館においても HTML を始めとする Web ページに関連したファイルフォーマットに対する検討が必要である。なお、Web ページの見読性確保に関わる要素として、ページを構成する各電子ファイルを格納しているフォルダの構成、また CSS や DTD などのレイアウトや文書構造を指定する電子ファイルがある。これらの Web ページ全体の構成に関わるフォルダ構成の情報を保存しておくことが必要である。また、Web ページのコンテンツ内に含まれるテキスト、画像、音声などの多様な電子ファイルに対して一律的に長期保存フォーマットを定めることはせず、対象となる各ファイルのフォーマット

⁽¹²⁾ ISO/IEC 15445:2000. Information technology — Document description and processing languages — HyperText Markup Language (HTML). <https://www.iso.org/standard/27688.html>. (accessed 2020-06-30).

毎に変換方針を定めて対応を行うこととする。加えて、今後の移管にあたっては、ページのリンク先に外部ファイルが存在することを考慮し、移管対象となる電子公文書等についての範囲の確認や移管対象となるファイルそのものの見読性の確保がされるよう、移管元行政機関等に対して促す必要がある。

3.2.2.2.(3) CAD

CAD 形式の電子ファイルとして、これまで様々なフォーマットが館に移管されている。館における CAD の各フォーマットの移管実績を以下に示す。

表 3-8 CAD のフォーマット毎の移管実績

フォーマット	移管電子ファイル数	概要
HOC	約 500	フリーソフトウェアで作成されるフォーマット
JWW	約 90	フリーソフトウェアで作成されるフォーマット
SDF(SFC)	約 60	CAD データ交換標準開発コンソーシアム(SCADEC)が日本国内での CAD データ交換のために策定したフォーマット
JWC	約 40	フリーソフトウェアで作成されるフォーマット
SXF(P21)	約 20	CAD データ交換標準開発コンソーシアム(SCADEC)が ISO に準拠して策定したフォーマット
DXF	約 20	AutoCAD で作成されるフォーマット
DCZ	数件	DynaCAD で作成されるフォーマット

CAD ファイルのフォーマットについては、CAD データ交換標準開発コンソーシアム (SCADEC) が国際標準 ISO 10303-21:2016⁽¹³⁾に準拠した規格として SXF の P21 形式を定めており、国内専用の形式として SFC 形式を定めている。これらの規格については、国土交通省等が発注する工事完成図書等のファイルフォーマットとして定められており、国内においてはデファクトスタンダードとして利用されている⁽¹⁴⁾。

⁽¹³⁾ ISO 10303-21:2016. Industrial automation systems and integration — Product data representation and exchange — Part 21: Implementation methods: Clear text encoding of the exchange structure. <https://www.iso.org/standard/63141.html>, (accessed 2020-06-30).

⁽¹⁴⁾ 金澤文彦, 青山憲明, 今井龍一, 上坂克巳, 大野聡, 西木也寸志. “3.CAD データの交換標準 SXF”. 国土技術政策総合研究所研究資料. 2007, no.403, p.1-5, <http://www.nilim.go.jp/lab/bcg/siryou/tnn/tnn0403.htm>, (accessed 2020-06-30).

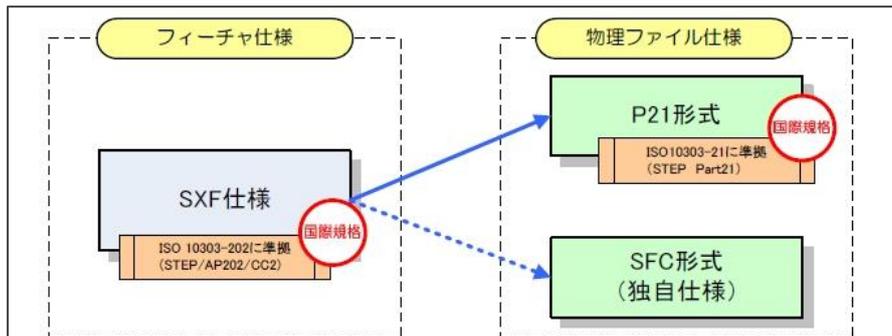


図 3-1 SXF のファイル形式

表 3-2 SXF のファイル形式の違い

項目	SXF(P21)形式	SXF(SFC)形式
フォーマット形式	国際標準 ISO 10303-21 (STEP/Part21)の形式	SCADEC で策定した国内専用の形式
ファイルの内容	国際標準 ISO 10303-202 (STEP/AP202)のサブセット	SCADEC で策定した仕様(フィーチャ仕様)に基づいたもの
ファイルの拡張子	P21	SFC
利用場面	CAD 製図基準(案)で納品用として定めている	打合せ段階
特徴	長期保存に利用できる	SXF(P21)形式に比べるとファイルサイズが小さいため、変換時間が短いなど取り扱いが容易【参考】

引用：金澤文彦他、前掲註(14) 3-2 ページ

図 3-4 SXF のファイル形式

SXF の平成 23 年度から平成 30 年度までの受入れ実績は、CAD ファイル全体の受入れ実績の 10%未満である。館における CAD ファイルとして受入れ実績が多いのは、全体の 76% を占める HOC 及び JWW であり、双方ともオープンソースではないフリーソフトウェアで出力されるフォーマットである。

また、公益社団法人日本文書情報マネジメント協会（以下「JIIMA」という。）の「建築設計業務における設計図書の電磁的記録による作成と長期保存のガイドライン⁽¹⁵⁾」によると、図面類を電磁的記録で作成する場合は、ISO 32000-2:2017⁽¹⁶⁾準拠の PDF を推奨しており、長期的な保存の観点では、PDF/A が望ましいとされている。

⁽¹⁵⁾ 公益社団法人日本文書情報マネジメント協会ガイドライン検討会. 建築設計業務における設計図書の電磁的記録による作成と長期保存のガイドライン. 公益社団法人日本文書情報マネジメント協会, 2017, https://www.jiima.or.jp/pdf/kenchiku_denshika_vol1_201712.pdf, (accessed 2020-06-30).

⁽¹⁶⁾ ISO 32000-2:2017. Document management — Portable document format — Part 2: PDF 2.0. <https://www.iso.org/standard/63534.html>, (accessed 2020-06-30).

8. 本ガイドラインでの推奨ファイル形式（フォーマット）

1) 図面類

① 図面類を電磁的記録で作成する場合

CAD ソフトウェアから直接作成する場合は、拡大表示しても端末の能力に合わせた表示ができ、拡大してもジャギーが発生しないことや、ファイルサイズが比較的小さいことなどから、ベクターデータ形式で構成された ISO 32000-2 準拠の PDF とする。また CAD データ（DWG、DXF、JWW 等）を直接読み取って PDF を作成するソフトウェアやクラウド変換サービスを利用する場合も同様の PDF 仕様を推奨する。

ソフトウェアによっては PDF/A を作成できるものもあるが、通常の PDF にフォントを埋め込むことは技術的には可能であり、あるいは、テキストフォントを使用せず、文字もベクターデータで表現するソフトウェアもあるため、PDF/A が望ましいものの採用については長期的な保存を考慮したうえで、作成者の判断となる。

引用：公益社団法人日本文書情報マネジメント協会、前掲註（15）30 ページ

図 3-5 図面類を電磁的記録で作成する場合のフォーマット

CAD ファイルについては、図面や CAD ソフトウェアに関する知見が必要とされるため、一般的な文書作成や表作成等とはフォーマットの特徴が異なる。そのため、受入れ実績数が多い CAD ファイルについては、標準的フォーマットとして追加する必要があると考える。

なお、CAD ファイルの PDF 又は PDF/A での長期保存を前提とする場合、CAD ファイルから PDF/A へ変換が可能な市販のソフトウェアも存在するため、一定の試行期間が必要であるものの、館として対応可能であると考え。

CAD に関する確認結果を以下に示す。

表 3-9 CAD の標準的フォーマット追加の確認結果

No	項目		確認結果
1	移管実績数	移管実績数は多いか。	約 800
2		移管実績が一過性のもではないか。	H23 年度～H30 年度まで 3 年移管がされている。
3	標準規格	国際標準、国内標準又はデファクトスタンダードであるか。	P21 は国際標準化機構(ISO)、SFC は国内標準として規定されている。
4	環境の用意	いつまでに受け入れ環境を用意すればよいか。	既に移管実績があり対応が必要である。
5	長期保存の方法	適切な長期保存フォーマットが存在するか。	JIIMMA のガイドラインでは PDF/A が推奨されている。
6		長期保存フォーマットへの変換は可能か。	受入れ実績から多様なフォーマットが移管されている。全てのフォーマットの変換を行うには、それに応じた変換環境の用意や、CAD の取扱いに関する知見が必要となる。

以上より、CAD ファイルの標準的フォーマットとして、標準化されている P21 及び SFC を追加するとともに、PDF/A を長期保存フォーマットとすることが妥当である。なお、移管実績が多いものの、HOC 及び JWW についてはオープンソースではないフリーソフトウェアで出力されるフォーマットであり、非標準フォーマットとして整理することが妥当であるとする。

3.2.2.2.(4) OS 関連ファイル

OS 関連ファイルは、例えば、パーソナルコンピュータのオペレーティングシステムや移管元行政機関等の業務システム等によって自動的に作られるファイル管理のためのデータファイルである。館の「問題 2-1」の事例にみられるとおり、OS 関連ファイルについては、移管される電子公文書等のフォルダに含まれるものの、移管する側が移管を意図していない電子ファイルである可能性が高い。

以上より、OS 関連ファイルについては、本調査検討の対象から外すこととした。

3.2.2.2.(5) その他ファイル

その他ファイル（電子メール、画像編集ソフトウェア及びデータベース等）については、移管実績が少なく、政府としての方針や管理のあり方についての検討状況等を踏まえて引き続き検討を行うこととする。

3.2.2.3. 標準的フォーマットの見直し

現行のフォーマット対応表で定められている、標準的フォーマットの見直しに関して確認した観点を以下に示す。

表 3-10 標準的フォーマットの見直しに係る確認観点

No	確認観点	内容	補足(確認方法等)
1	移管数	長期保存フォーマットへの変換ができるか。	実績を確認する。

上記観点に基づく標準フォーマットの評価結果、いずれも特段の問題なく長期保存フォーマットへの変換ができているため、今回の調査検討において見直しが必要な標準的フォーマットは存在しないと考える。

3.2.2.4. 長期保存フォーマットの見直し

現行のフォーマット対応表で定められている、長期保存に適した長期保存フォーマットの見直しを実施した。長期保存フォーマットの見直しにあたって確認した観点を以下に示す。

表 3-11 長期保存フォーマットを見直す際の観点

No	項目
1	標準化、規格化がなされている。
2	広く普及している。
3	技術内容は誰でも知ることができる。
4	オリジナルの情報に戻すことができる(可逆)圧縮か、そうではなく非可逆圧縮か。
5	移管データに対して長期保存フォーマットが品質の面で劣っていない。

上記 No.1～No.4 の観点は、総務省が公表した「震災関連デジタルアーカイブの構築・運用のためのガイドライン⁽¹⁷⁾」において示されたものである。

上記観点に基づく、長期保存フォーマットに対する評価結果を以下に示す。

3.2.2.4.(1) 文書作成に係る長期保存フォーマット

文書作成フォーマットで保存・移管される電子公文書等の長期保存フォーマットはPDF/Aである。PDF/A は、ISO 19005-1⁽¹⁸⁾として国際標準化されていることに加えて電子文書の配布形式として広く普及しており、長期保存における見読性確保に加えて移管元の行政機関等における利便性といった、館が長期保存フォーマットに求める要件からも、引き続き長期保存フォーマットとして利用することが妥当である。

⁽¹⁷⁾ 総務省. 震災関連デジタルアーカイブ構築・運用のためのガイドライン. 2013, https://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/02ryutsu02_03000114.html, (accessed 2020-06-30).

⁽¹⁸⁾ ISO 19005-1. Electronic document file format for long-term preservation -Part 1: Use of PDF 1.4(PDF/A-1). <https://www.iso.org/standard/38920.html>, (accessed 2020-06-30).

PDF/A に対する評価結果を以下に示す。

表 3-12 PDF/A に対する確認結果

No	項目	確認結果
1	標準化、規格化がなされているか。	ISO 19005-1 ⁽¹⁹⁾
2	広く普及しているか。	長期保存フォーマットとして広く利用されており、NARA 等も採用している。
3	技術内容は誰でも知ることができるか。	技術仕様は公開されている。
4	オリジナルの情報に戻すことができる(可逆)圧縮か、そうではなく非可逆圧縮か。	非可逆圧縮
5	移管データに対して長期保存フォーマットが品質の面で劣っていないか。	オリジナルフォーマットの論理構造等も保持可能である。

3.2.2.4.(2) 表計算に係る長期保存フォーマット

表計算フォーマットで保存・移管される電子公文書等の長期保存フォーマットは PDF/A である。「3.2.2.4.(1)文書作成に係る長期保存フォーマット」に示すとおり、引き続き長期保存フォーマットして利用することが妥当である。

3.2.2.4.(3) プレゼンテーションに係る長期保存フォーマット

プレゼンテーションフォーマットで保存・移管される電子公文書等の長期保存フォーマットは PDF/A である。「3.2.2.4.(1)文書作成に係る長期保存フォーマット」に示すとおり、引き続き長期保存フォーマットして利用することが妥当である。

3.2.2.4.(4) 画像に係る長期保存フォーマット

画像フォーマットで保存・移管される電子公文書等の長期保存フォーマットは JPEG2000 である。JPEG2000 は、ISO/IEC 15444-1:2004⁽²⁰⁾として国際標準化されており、現在も広く普及しているため、引き続き長期保存フォーマットして利用することが妥当である。

JPEG2000 に対する評価結果を以下に示す。

表 3-13 JPEG2000 に対する確認結果

No	項目	JPEG2000
1	標準化、規格化がなされているか。	ISO/IEC 15444-1:2004
2	広く普及しているか。	長期保存フォーマットとして広く利用されており、NARA 等も採用している。

⁽¹⁹⁾ ISO 19005-1:2005. Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1), <https://www.iso.org/standard/38920.html>, (accessed 2020-06-30).

⁽²⁰⁾ ISO/IEC 15444-1:2019. Information technology — JPEG 2000 image coding system: Core coding system — Part 1, <https://www.iso.org/standard/78321.html>, (accessed 2020-06-30).

No	項目	JPEG2000
3	技術内容は誰でも知ることができるか。	技術仕様は公開されている。
4	オリジナルの情報に戻ることができる(可逆)圧縮か、そうではなく非可逆圧縮か。	可逆圧縮、非可逆圧縮。
5	移管データに対して長期保存フォーマットが品質の面で劣っていないか。	要求仕様において、解像度やビット深度について高い品質を要求した上で作成されている。

なお、画像に係るファイルフォーマットとしては、TIFF が広く利用されているが、館としてこれまで採用してきた JPEG2000 について、「表 3-13 JPEG2000 に対する確認結果」への該当性の確認結果から、画像に係る長期保存フォーマットの見直しを行う合理性はないものと判断する。

3.2.2.4.(5) 音声に係る長期保存フォーマットの見直し

音声に係る現行の長期保存フォーマットは、MP3（ビットレート 256kbps 以上）である。MP3 は、ISO/IEC 11172-3:1993⁽²¹⁾として規格化されており、より高品質な後継フォーマットがリリースされていること、また各項目の確認結果を踏まえると、音声の長期保存フォーマットとして AAC の採用が適当であると判断できる。なお、現行のフォーマット対応表で定義されているビットレートについては、移管される電子ファイルのビットレートが想定できず、移管される音声のコーデックにも影響するため、長期保存フォーマットとして一律の定義は行わないことが妥当と考えられる。

MP3 に対する確認結果を以下に示す。

表 3-14 MP3 に対する確認結果

No	項目	MP3
1	標準化、規格化がなされているか。	ISO/IEC 11172-3:1993 として規格化。
2	広く普及しているか。	携帯音楽プレーヤや、ボイスレコーダーなどで採用されている。
3	技術内容は誰でも知ることができるか。	技術仕様は有償であるが公開されている。
4	オリジナルの情報に戻ることができる(可逆)圧縮か、そうではなく非可逆圧縮か。	非可逆圧縮

⁽²¹⁾ ISO/IEC 11172-3:1993. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio. <https://www.iso.org/standard/22412.html>, (accessed 2020-06-30).

No	項目	MP3
5	移管データに対して長期保存フォーマットが品質の面で劣っていないか。	AAC(Advanced Audio Coding)はMP3の後継フォーマットである。MPEG-2及びMPEG-4に採用されており、MP3とほぼ同じ音質ながら、圧縮効率を1.4倍ほど高めているのが特徴。開発から10年以上が経過したというMP3が2チャンネルであるのに対し、AACでは最大48チャンネルを利用できる。

3.2.2.4.(6) 映像に係る長期保存フォーマット

映像に係る現行の長期保存フォーマットは、MPEG-2である。MPEG-2は、ITU-T Rec. H.262⁽²²⁾及びISO/IEC 13818-2:2000⁽²³⁾として規格化されており、より高品質な後継フォーマットであるMPEG-4がリリースされていること、また各項目の確認結果を踏まえると、映像の長期保存フォーマットを、MPEG-4に変更することが適当であると判断できる。

MPEG-2に対する確認結果を以下に示す。

表 3-15 MPEG-2 の確認結果

No	項目	MPEG-2(映像)
1	標準化、規格化がなされているか。	ITU-T Rec. H.262 及び ISO/IEC 13818-2:2000 として規格化。
2	広く普及しているか。	地上デジタル放送や、DVD、映像の録画装置などで採用されている。
3	技術内容は誰でも知ることができるか。	No1の通り技術仕様は公開されている。
4	オリジナルの情報に戻すことができる(可逆)圧縮か、そうではなく非可逆圧縮か。	非可逆圧縮
5	移管データに対して長期保存フォーマットが品質の面で劣っていないか。	後継フォーマットとしてMPEG-4が存在する。Blu-ray等で採用されており、MPEG-2より画質は上位のものとなる。Blu-rayは、移管実績として存在し、今後も増加することが想定される。

⁽²²⁾ H.262 : Information technology - Generic coding of moving pictures and associated audio information: Video, <https://www.itu.int/rec/T-REC-H.262>, (accessed 2020-06-30).

⁽²³⁾ ISO/IEC 13818-2:2000. Information technology — Generic coding of moving pictures and associated audio information: Video — Part 2: <https://www.iso.org/standard/31539.html>, (accessed 2020-06-30).

3.2.3. フォーマット対応表の管理項目の見直し

現行のフォーマット対応表について、運用を適切に実施するために、管理対象とする項目の見直しを実施した。

表 3-16 フォーマット対応表の管理項目見直し内容

No	項目	見直し理由	対応
1	管理番号の項目を追加	フォーマット対応表の定期的な更新が実施されることを想定すると、管理番号が必要である。	標準的フォーマット単位での管理番号の項目を付与する。
2	各フォーマットの項目名及び記載内容の整理	フォーマットとして記載されている内容が、受入れの可否を表すものか(例:当該フォーマットにより作成される拡張子の全てを標準的フォーマットとして定義しているか 等)、長期保存を行うものか不明瞭である。また、一般利用を目的とする一般提供用データのフォーマットが直感的に分かりにくい。	各フォーマットで記載する項目及び内容を整理し、明確化する。 ● フォーマット ● バージョン ● 拡張子
3	目的別の表の作成	フォーマット対応表の目的に照らした整備が必要である。なお、フォーマット対応表の目的は以下のとおり。 ● 館の運用方針を対外に示す ● 電子公文書等システムで管理するフォーマット情報やフォーマット変換の詳細を示す	目的別に、簡易版と詳細版を2つの対応表を用意する。 簡易版:館が推奨するフォーマットでの移管を促す用途 詳細版:フォーマット管理や変換に係るシステム仕様を明確にする用途

3.2.4. フォーマット対応表の改善案

前項の検討を踏まえて、現行のフォーマット対応表の改善案を検討した。

3.2.4.1. フォーマット対応表（簡易版）

館が実施する研修や移管元行政機関等の外部への提示を目的とした、フォーマット対応表（簡易版）の改善案を次頁に示す。なお、簡易版については、その時点で館が受け入れる推奨のフォーマットを示すものとし、対応表の見直し後に技術的な理由などで対応困難となったフォーマットの記載は含めないものとする。

表 3-17 フォーマット対応表の改善案

管理番号	種類	標準的フォーマット			長期保存フォーマット			一般利用フォーマット		
		フォーマット	バージョン	拡張子	フォーマット	バージョン	拡張子	フォーマット	バージョン	拡張子
0001	文書作成	OASYS	V10	oa3	PDF/A	PDF/A-1b	pdf	PDF/A	PDF/A-1b	pdf
0002		一太郎	8-12	jtd						
0003		Word	97-2003	doc						
0004			2007-2016	docx, docm						
0005		PDF	1.3 ~ 1.7	pdf						
0006		PDF/A	PDF/A-1	pdf						
0007		PDF/A	PDF/A-2	pdf						
0008		OpenOfficeWriter	4.1.0-	odt						
0009	表計算	EXCEL	97-2003	xls	PDF/A	PDF/A-1b	pdf	PDF/A	PDF/A-1b	pdf
0010			2007-2016	xlsx						
0011		OpenOfficeCalc	4.1.0-	ods						
0012	プレゼンテーション	PowerPoint	97-2003	ppt	PDF/A	PDF/A-1b	pdf	PDF/A	PDF/A-1b	pdf
0013			2007-2016	pptx						
0014		OpenOfficeImpress	4.1.0-	odp						
0015	画像	JPEG	-	jpeg, jpg	JPEG2000 (lossless)	-	jp2	JPEG2000 (lossy)	-	jp2
0016		JPEG2000	-	jp2						
0017		GIF	-	gif						
0018		TIFF	-	tiff, tif						
0019		BMP	-	bmp						

管理番号	種類	標準的フォーマット			長期保存フォーマット			一般利用フォーマット		
0020	音声	WAVE	-	wav	MP4(音声コーデック: AAC)	-	mp4	MP4(音声コーデック: AAC)	-	mp4
0021		Windows Media Audio	-	wma						
0022		MP3	-	mp3						
0023		MP4	-	mp4, m4a						
0024	映像	QuickTime	-	mov, qtm	MP4(映像コーデック: H264、音声コーデック: AAC)	-	mp4	MP4(映像コーデック: H264、音声コーデック: AAC)	-	mp4
0025		WMV	-	Wmv						
0026		RealVideo	-	Rm, rmvb, ram						
0027		MPEG-2	-	mpg, mpeg, m1v						
0028		MP4	-	mp4, m4v						
0029	テキスト	Plain Text	-	txt	Plain Text	-	txt	Plain Text	-	txt
0030		Comma Separated Values	-	csv	Comma Separated Values	-	csv	Comma Separated Values	-	csv
0031		Hypertext Markup Language	HTML4.01, XHTML1.0, HTML5	htm, html	Hypertext Markup Language	HTML4.01, XHTML1.0, HTML5	htm, html	Hypertext Markup Language	HTML4.01, XHTML1.0, HTML5	htm, html
0032		eXtensible Mark-up language	1.0,1,1	xml	eXtensible Mark-up language	1.0,1,1	xml	eXtensible Mark-up language	1.0,1,1	xml
0033		XML Schema Definition	-	xsd	XML Schema Definition	-	xsd	XML Schema Definition	-	xsd

管理番号	種類	標準的フォーマット			長期保存フォーマット			一般利用フォーマット		
0034	WEB ページ	WEB ページを構成する各電子ファイルのファイルフォーマット	-	-	電子ファイルのフォーマット毎に対応する長期保存フォーマット			電子ファイルのフォーマット毎に対応する長期保存フォーマット		
0035	CAD	SXF(P21) (試験的)	-	P21	PDF/A	PDF/A-1b	pdf	PDF/A	PDF/A-1b	pdf
0036		SXF(SFC) (試験的)	-	sfc						

3.2.4.2. フォーマット対応表（詳細版）

フォーマット対応表（簡易版）に対して、電子公文書等システムのフォーマット変換の詳細仕様を明確にすることを目的とするフォーマット対応表（詳細版）では、以下の要素を追加する。また、各フォーマットについては、例えば米国 NARA の「Digital Preservation Framework⁽²⁴⁾」を始めとしたリスク評価のフレームワークを参照し、ファイルフォーマットのリスクが High とされるものについてはその管理方法や措置を適宜判断する必要がある。

- 各フォーマットの技術情報の詳細
- 各フォーマットについて英国 TNA が提供する PRONOM（ファイルフォーマット情報提供サービス）で付番されている識別子(PUID)
- 各フォーマットについてフォーマットリスクの評価結果
- 館が管理するフォーマットについての履歴情報

現行システムでは、標準的フォーマットに関する情報として技術的メタデータを管理しているが、長期保存フォーマットの変換作業に係る仕様や処理に使用したソフトウェア等の技術的な情報は管理していない。そのため、フォーマット対応表（詳細版）には、こうした技術情報を追加するほか、ファイルのアクセスに必要なドキュメントやマニュアル類を参照できるようにすることで、データ・ディクショナリーとしての活用も期待される。また、フォーマット変換の根拠となるフォーマット対応表自体の履歴管理を行い、長期保存フォーマットに変換された電子ファイルとメタデータと共に、一体的に管理することが望まれる。

3.3. オリジナルフォーマットの原則保存

電子公文書等の見読性は、時間経過に伴う再生環境の変化などによって変わり得るため、将来にわたって確保することは難しい。また、電子ファイルのフォーマットが多様化し、当該ファイルの記録としての価値を維持するのに必要不可欠なエッセンスの定義や判断が困難になっている現状に鑑みると、長期保存フォーマットの保存のみでは、電子公文書等の適切な保存の目的を達成できないとも考えられる。米国及び英国において、オリジナルフォーマットでの保存を採用していることも、その有用性の根拠となる。

以上より、電子公文書等システムによる長期保存フォーマットへの変換及び保存に加え、移管時のオリジナルフォーマットをシステム内に保存することを併用して、リスクに備えることが妥当である。

⁽²⁴⁾ The U.S. National Archives and Records Administration, “Digital Preservation Framework.” <https://github.com/usnationalarchives/digital-preservation>, (accessed 2020-06-30).

4. 受入れ等に係る業務の実施に対する対応

4.1. 受入れ等に係る業務の効率化

4.1.1. 業務プロセスの見直し

受入れ等に係る業務の更なる効率化のために、従来業務の関連性を整理した上で、同時並行的に業務を行うことを検討する。なお、業務の同時並行による実施に際しては、手戻りを考慮した上で対象を設定する必要がある。

業務の前後関係及び目的を加味した上で、同時に実施可能な業務、業務プロセス見直し前後の業務フローを以下に示す。

- 技術的メタデータ付与と目録整備からなるメタデータ付与業務の【統合】
- フォーマット変換業務とメタデータ付与業務の【並列化】
- 行政利用業務と一般利用業務の【並列化】

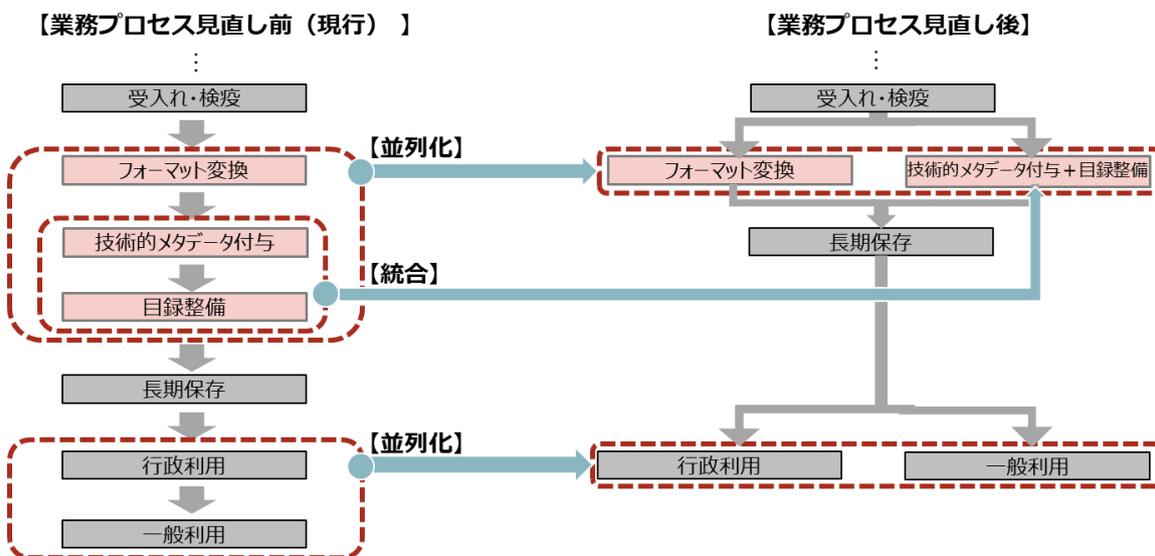


図 4-1 業務プロセスの見直し前後の業務フロー案

4.1.2. 機械学習を利用したフォーマット判定機能

移管される電子公文書等のフォーマット判定は、現行と同様に PRONOM と DROID を引き続き利用することとする。また、PRONOM 及び DROID では判定出来ない日本独自のフォーマットについて、英国の事例を参考に、機械学習等の先端的なデジタル技術を活用した自動化の検討が望まれる。具体的には、電子公文書等システムに保存されている、日本独自のフォーマットの電子ファイルを教師データとして学習させることが想定される。

4.2. フォーマット変換処理結果の確認観点の明確化

現行システムの運用では、長期保存フォーマットへの変換作業のうち、自動変換が適切に実施できないケースが存在し、再変換の実施や手動変換が発生して時間を要している。業務の円滑化に向けては、変換エラーとなる原因の分析、再処理又は手動作業の自動化の検討が望まれる。また、館が行うフォーマット変換後の品質確認については、処理工程の手戻りを防ぐという点のみならず、館が電子公文書等の適切な保存と利用を担うために重要な作業である。そのため、館としての長期保存フォーマットへの変換後の確認内容の基準を設定することが求められる。

標準的フォーマットから長期保存フォーマットへの変換における確認基準を明確化するため、長期保存フォーマット変換処理結果確認の観点を以下に示す。

表 4-1 長期保存フォーマットへの変換における確認観点

No	確認分類	確認項目	確認内容
1	電子公文書等システム上の確認	電子公文書等システムが出力する処理結果	電子公文書等システムが出力する処理結果として、正常であることを確認する。
2		ファイルサイズ	長期保存フォーマットのファイルサイズが、オリジナルフォーマットのファイルのサイズと比して所定の割合の範囲内であることを確認する。
3		処理時間	フォーマット変換の処理時間が、所定の時間内であることを確認する。
4		ファイルを開けるかどうか (確認分類上、館職員による確認でも可)	長期保存フォーマットの電子ファイルが開けるかを確認する。
5	館職員による確認	文字化けが起きていないか	館にて事前に定めた確認対象頁(例えば先頭ページ)等について、文字化けが起きていないかを確認する。
6		見読できる内容となっているか	館にて事前に定めた確認対象頁(例えば先頭ページ)等について、見読上の問題点(内容が白紙等)は無いか。

4.3. 電子公文書等のフォーマットに係る技術情報蓄積

専門家から得た示唆「2.3.2.2 利用可能な状態を維持管理する環境の整備」に示すとおり、電子公文書等を中長期的に利用可能な状態を維持管理していくための環境に向けて、電子ファイルのフォーマットに係る技術情報の蓄積を体系的・組織的に実施する必要がある。

受入れ等に係る業務を通して技術情報を蓄積するイメージを以下に示す。

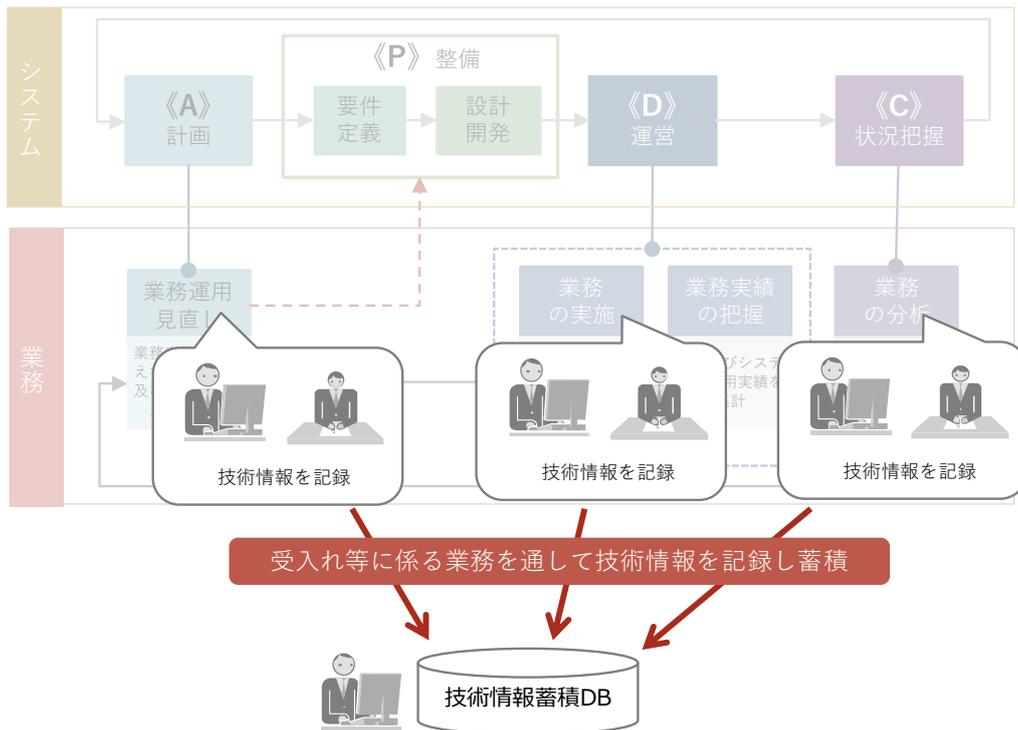


図 4-2 技術情報の蓄積のイメージ

技術情報の蓄積について、必要な対応を以下に示す。

表 4-2 電子公文書等のフォーマットに係る技術情報蓄積に対する対応

対応策の分類	対応
蓄積する技術情報の拡充	フォーマット対応表の履歴管理
	フォーマット変換エラー情報の蓄積
技術情報蓄積に係る運用改善	技術情報蓄積に係る運用ルールの方策
	技術情報の文書品質向上
類縁機関との連携	類縁機関との技術情報の共有

4.3.1. フォーマット対応表の履歴管理

「3.2.4.2 フォーマット対応表（詳細版）」に示したとおり、館におけるフォーマット変換に係る技術情報の履歴を蓄積するとともに、保存する電子公文書等と一体的に管理することで見読性を長期的に確保することが見込まれる。また、フォーマット対応表は、館が行うフォーマット変換業務の作業方針や根拠となるものであるため、見直す度にフォーマット対応表自体の履歴管理が行われる必要がある。

4.3.2. フォーマット変換エラー情報の蓄積

蓄積する技術情報として、標準的フォーマットから長期保存フォーマットへ変換した際のエラー情報も重要である。フォーマット変換に関するエラー情報の蓄積にあたっては、電子

公文書等システムによる自動的に取得するもの、館職員の記録によるものと2種類が想定される。それらは、必要な情報が効率的かつ適切に管理され、職員が利用可能な帳票や一覧形式で出力できるようにデータベースとして整備することが望ましい。

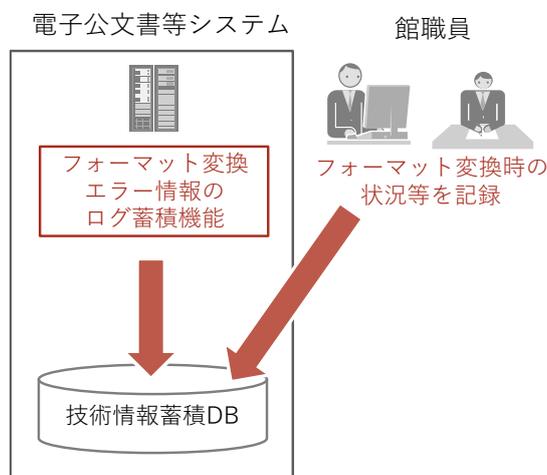


図 4-3 システムと職員による技術情報の蓄積のイメージ

4.3.2.1. システムによるフォーマット変換情報の取得

フォーマット変換に係る処理エラーに係る情報のうち、自動で取得可能な情報については電子公文書等システムにより自動的に取得されるべきである。電子公文書等システムにおいて、自動取得を想定する情報を以下に示す。

- 処理エラーに関するシステムログ情報
- 対象のファイル情報（例：ファイル種類、ファイルサイズ、タイムスタンプ 等）
- システム環境情報（例：動作環境、変換処理で使用するソフトウェアの情報 等）

なお、処理が正常時の情報も、エラーに至らないまでも類似状況の発生有無等の確認において有用と考えられる。そのため、取得ポイントを絞り、かつ、変換処理を行ったファイル名や処理に要した時間など、最低限の情報とすることが重要である。

4.3.2.2. 館職員によるフォーマット変換情報の取得

フォーマット変換に係る情報のうち、職員の手動による記録が想定される情報を以下に示す。

- 事象（例：状況、直前操作内容、対応方法、エラー有無及び画面表示内容 等）
- 原因・経緯
- 発生時の対処（例：発生時の確認、対処方法 等）
- 影響（例：作業的影響、品質的影響 等）
- 関連する事象

※ 下線：現行システムの課題管理表から有用と考えられる項目

4.3.3. 技術情報の蓄積に係る運用ルールの策定

技術情報の取得は、業務上で記録するタイミングや条件を明示化した上で、日常の業務の中で蓄積、体系化され、共有できることが重要である。情報の整理は定期的を実施し、フォーマット変換エラー等の原因の特定や改善の活動が、システムや業務の運営全体を通じて行われることが重要である。そのため、次期システムの運用保守要件として組み込み、エラー情報の整理及び分析を行う必要がある。

また、技術情報が中長期的な業務遂行に役立つレファレンスとなるように、PDCA サイクルを回すことが重要である。蓄積された情報による運用改善や機能改善、原因特定による業務の安定的な実施に向けた、運用ルールを作り上げる必要がある。

4.3.4. 類縁機関との技術情報の共有

将来的な方向性として、館と同じように電子公文書等の移管及び長期保存・システム化を検討する類縁機関が共有及び参照できる仕組みの構築を目指すことが望まれる。こうした保存や利用技術に係る蓄積情報は、都道府県の公文書館等の類縁機関において、システム化検討やシステム運用に際する参考情報として寄与することが期待される。

5. 電子公文書等システムの運用に対する対応

5.1. 次期システムの更改

電子公文書等の適切な保存を実現するためには、本調査で整理した各事項に対応しつつ、次期システムの継続的な運用につなげていくことが不可欠である。次期システムに求められる各要件の対応については、要件定義工程で検討を深める必要がある。

次期システムの更改に向けて、本調査で提示した次期システムの要件を次頁に示す。

表 5-1 次期システムの対応事項

要件定義事項		関係する対応	要件定義工程での検討事項
業務要件		フォーマット対応表の見直し	<ul style="list-style-type: none"> ● 標準的フォーマットとして追加したフォーマットを対象として定義する必要がある。 ● 長期保存フォーマットとして追加したフォーマットを対象として定義する必要がある。 ● 上記を含め、改善したフォーマット対応表に基づく業務を定義する必要がある。 ● フォーマット対応表の定期的な改善を業務として定義する必要がある。
		オリジナルフォーマットの原則保存	<ul style="list-style-type: none"> ● オリジナルフォーマットの原則保存を業務要件として定義する必要がある。
		電子公文書等のフォーマットに係る技術情報蓄積の運用改善	<ul style="list-style-type: none"> ● 技術情報の体系的な仕組み、運用改善に向けた整理・分析を業務要件として定義する必要がある。
機能要件		フォーマット対応表の見直し	<ul style="list-style-type: none"> ● 標準的フォーマットとして追加したフォーマットに関するフォーマット変換機能を要件化する必要がある。 ● 長期保存フォーマットとして追加したフォーマットへのフォーマット変換機能を要件化する必要がある。 ● 上記を含め、改善したフォーマット対応表に基づくフォーマット変換機能を要件化する必要がある。
		フォーマットへの変換結果の確認観点の明確化	<ul style="list-style-type: none"> ● 長期保存フォーマットの品質確認を容易に実施するための機能を要件化する必要がある。
		受入れ等に係る業務の効率化	<ul style="list-style-type: none"> ● 機械学習を利用したフォーマット判定機能の実現性について検討した上で、要件化する必要がある。
		電子公文書等の保存に係る館の方針を踏まえた機能改善	<ul style="list-style-type: none"> ● 以下の機能について、要件化する必要がある。 <ul style="list-style-type: none"> ・ 日本独自で扱う電子公文書等のフォーマット情報を登録する機能 ・ 保存している電子公文書等に係るフォーマットの統計情報を出力する機能
非機能要件	システム方式に関する事項	システム方式の見直し	<ul style="list-style-type: none"> ● 業務の特性やコスト、情報セキュリティなどクラウド利用の実現性を確認した上で、システム方式を決定する必要がある。
	規模に関する事項	外部環境を踏まえた要件の具体化	<ul style="list-style-type: none"> ● 電子公文書等の移管件数が指数関数的に増加することも見据え、規模の拡張性を要件として定義する必要がある。
	性能に関する事項	外部環境を踏まえた要件の具体化	<ul style="list-style-type: none"> ● 電子公文書等の移管件数が指数関数的に増加することも見据え、性能の拡張性を要件として定義する必要がある。

要件定義事項		関係する対応	要件定義工程での検討事項
	情報セキュリティに関する事項	外部環境を踏まえた要件の具体化	<ul style="list-style-type: none"> ● 政府のガイドラインやマニュアル等で求められる情報セキュリティ対策について、業務の特性やコスト、システム方式を踏まえて、要件として定義する必要がある。
	信頼性に関する事項 拡張性に関する事項 上位互換性に関する事項 中立性に関する事項 継続性に関する事項 情報システム稼働環境に関する事項 テストに関する事項 移行に関する事項 引継ぎに関する事項 教育に関する事項 運用に関する事項 保守に関する事項	特になし	<ul style="list-style-type: none"> ● 業務要件等を踏まえて、次期システムに求められる要件として定義する必要がある。

5.2. 外部環境変化を踏まえた要件の具体化

5.2.1. システム方式の検討

クラウド・バイ・デフォルト原則の政府方針、電子決裁への移行の加速化や行政文書の電子的管理に係る基本的な方針に伴い、次期システムとしては今後電子公文書等の増加が見込まれること、またクラウドサービスの活用等により拡張性や柔軟性の向上が求められる。現行システムと同様のオンプレミス型の構成に加え、クラウド型、長期保存用ストレージを分離したハイブリッド型の検討結果を以下に示す。

表 5-2 システム方式の概要比較

	オンプレミス型	クラウド型 (IaaS/PaaS)	ハイブリッド型 (IaaS/PaaS+オンプレミス)
構成	現行システムと同様、オンプレミスで構築を行う。	遠隔地バックアップを除き、全てのサブシステムをクラウドで構成する。	長期保存システムのセキュリティを考慮し、長期保存のみをオンプレミスで構成する。
セキュリティ	◎ 物理的にネットワークの分離をしているため、セキュリティレベルは他パターンと比較し高い。	△ インターネット上の環境を利用しているため、ヒューマンエラー(セキュリティ設定ミス等)による情報漏洩の可能性はある。	○ インターネット上の環境を利用しているため、ヒューマンエラー(セキュリティ設定ミス等)による情報漏洩の可能性はあるものの長期保存システムへの影響は避けられる。
技術革新対応	○ 特に制約は無い。	△ 移管元行政機関等で整備予定のシステムについての連携について、個別ネットワーク要件の検討が必要と想定されクラウド環境が制約になる可能性はある。	△ 移管元行政機関等で整備予定のシステムについての連携について、個別ネットワーク要件の検討が必要と想定されクラウド環境が制約になる可能性はある。
柔軟性	△ フォーマット変換システムのスケールアップや、長期保存用ストレージのディスク容量の拡張の柔軟な対応については懸念がある。	◎ クラウドサービスの提供サービスによって、フォーマット変換システムの柔軟なスケールアップや、長期保存用ストレージのディスク容量の拡張が可能。	○ 長期保存用ストレージのディスク容量の拡張については、柔軟な対応に懸念がある。
可用性	○ 現行運用でも稼働率は担保されているため、稼働率は担保が可能と想定される。災害対策としての冗長化は、クラウドと比較し高価となる。	◎ AWS 等著名なクラウドサービスにおける稼働率は、99.95-99.99%とされており、稼働率は担保可能。災害対策としての冗長化は、オンプレミスと比べて確実かつ安価にシステム構築運用が可能と考えられる。	○ 現行運用でも稼働率は担保されているため、稼働率は担保が可能と想定される。災害対策としての冗長化は、クラウドと比較し高価となる。

今後行う要件定義を通して、上記を総合的に判断し、適切なシステム方式を選定する必要がある。現時点で想定される、次期電子公文書等システムのクラウド型及びハイブリッド型の構成イメージを以下に示す。

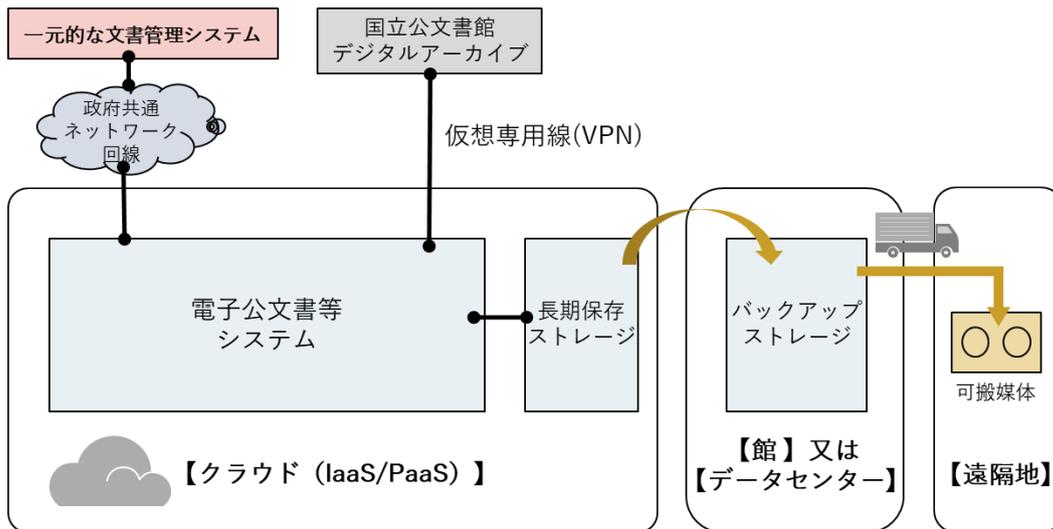


図 5-1 次期システムの構成イメージ①（クラウド型）

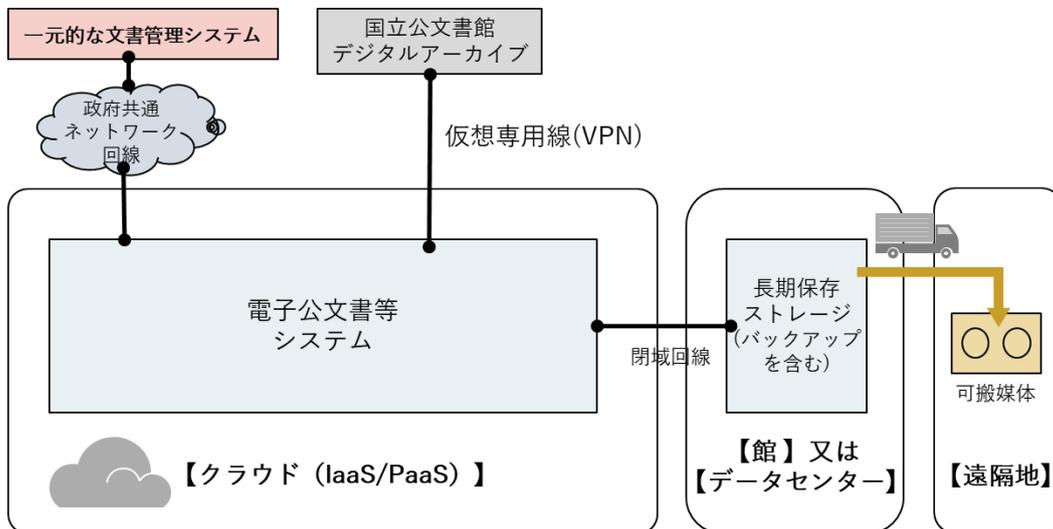


図 5-2 次期システムの構成イメージ②（ハイブリッド型）

5.2.2. 非機能要件

5.2.2.1. 性能要件の検討

電子公文書等の受入れ量の増加に伴い、フォーマット変換システムの性能拡張が必要となる。米国 NARA では、次期電子記録管理システム（ERA）をクラウドで構築することにより高い拡張性と柔軟性を確保しており、業務のニーズに応じてリソースを増減しやすい環境を採用している。

フォーマット変換システムについても、クラウドを採用することにより、性能の拡張性を担保することが望まれる。ただし、クラウドの採用にあたっては、「5.2.2.3 情報セキュリティ要件」に示すとおり、電子公文書等システムの業務の特性を踏まえた適切な情報セキュリティ対策を実施可能であることが前提となる。

5.2.2.2. 規模要件の検討

5.2.2.2.(1) 規模の拡張性の確保

現状、標準的フォーマットに依らない非標準フォーマットで作成されていたり、一つの行政文書ファイルとして想定を大きく超えるデータ容量であったりなど、例外的な事案の電子公文書等が移管された場合には、長期保存用ストレージの容量を柔軟に拡張が出来ないために、電子公文書等システムの外部に電子公文書等のデータを保存している。現用における電子公文書等の状況が把握できない状況は今後も継続するため、次期システムのストレージ容量を容易に拡張出来るようなシステム方式を採用することが望まれる。

規模の拡張性という観点では、米国 NARA にみられるようにソフトウェアやストレージまでをクラウド基盤上で構築するクラウド型が優位である。

5.2.2.2.(2) 長期保存フォーマットへの変換後のストレージ容量

次期システムの規模要件の検討にあたっては、長期保存フォーマットへ変換した後のデータ容量を試算する必要がある。本調査では、次期システムにおける長期保存用ストレージの容量を試算した。なお、「2.1.1 電子公文書等の管理に係る方針」にみられる今後は電子媒体を正本・原本とする電子文書管理の方向性を見据え、この増加率がどのようにどのように推移していくのか、注目していく必要がある。

「図 2-4 行政文書ファイル数の実績と今後の予測」に示す行政文書ファイルの移管数の予測結果を前提に、過年度における1行政文書ファイルあたりの長期保存用ストレージ容量を係数として適用することで、令和8年度における容量を試算している。長期保存用ストレージの試算手順を以下に示す。

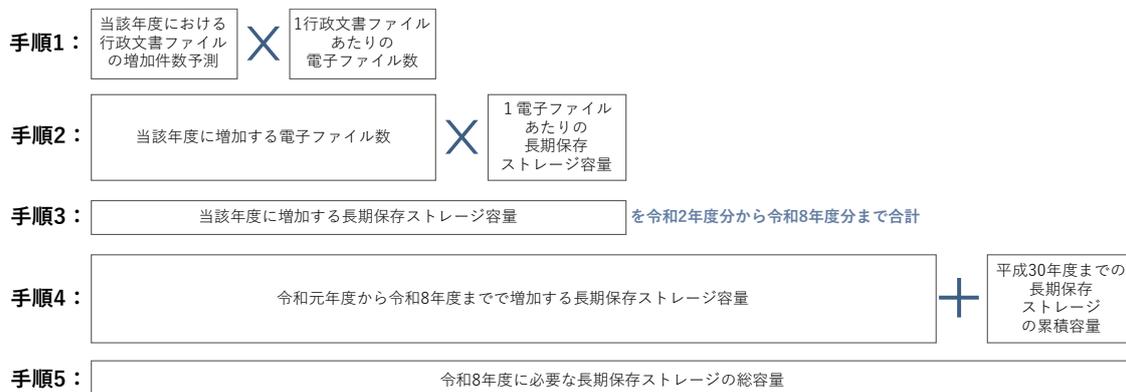


図 5-3 長期保存用ストレージの試算手順

試算にあたり採用した行政文書ファイル増加率を以下に示す。

- パターン①:平成29年度から平成30年度までの行政文書ファイル増加率=1年あたり33%増加
- パターン②:運用開始の翌年である平成24年度から平成30年度までの行政文書ファイル増加率の年平均=1年あたり58%増加

受入れ実績の増加予測に係る2つのパターンについて、長期保存用ストレージの容量を試算した結果を以下に示す。

表 5-3 長期保存用ストレージ容量の試算

年間の行政文書ファイル増加率	次期システムに求められる ストレージの想定容量 (令和8年度時点)
パターン①:33%増加	約 3,200GB (3.2TB)
パターン②:58%増加	約 8,900GB (8.9TB)

次期システムの長期保存用ストレージの容量は、これまで現行システムの制約によりシステム内に保存していなかったデータも可能な限り格納することも踏まえ、令和8年度には最大で約10TB程度の容量拡張に対応できるようにしておく必要があると試算される。加えて、これまで紙媒体で移管されていた文書が今後は紙媒体から電子媒体に変更となる加速化の度合い、また紙・電子媒体における割合の状況などを踏まえて、電子公文書等として移管されることを考慮する必要がある。このため、次期システムについては、規模の拡張性に対し、ストレージ容量の十分な確保と継続的な容量の精査が重要である。

5.2.2.3. 情報セキュリティ要件の検討

次期システムにおいては、クラウドサービスの採用可否に関わらず、電子公文書等を長期の保存と管理、安定的な利用提供を行うためのシステム基盤として、十分な情報セキュリティを担保することが求められる。「政府機関等の情報セキュリティ対策のための統一基準群⁽²⁵⁾」及び館の情報セキュリティポリシーの遵守事項に準拠して、情報システム全体としての情報セキュリティの要件化に向けた検討が必要である。

次期システムの要件定義の工程に先立ち、調達仕様の策定にあたっては、「情報システムに係る政府調達におけるセキュリティ要件策定マニュアル⁽²⁶⁾」（以下「SBD マニュアル」という。）を活用する。SBD マニュアルを用いたセキュリティ要件の検討結果は、以下のとおりである。

⁽²⁵⁾ 内閣官房内閣サイバーセキュリティセンター。「政府機関等の情報セキュリティ対策のための統一基準群（平成30年度版）」について、<https://www.nisc.go.jp/active/general/kijun30.html>, (accessed 2020-06-30).

⁽²⁶⁾ 内閣官房内閣サイバーセキュリティセンター。「情報システムに係る政府調達におけるセキュリティ要件策定マニュアル」の改訂について、https://www.nisc.go.jp/active/general/sbd_sakutei.html, (accessed 2020-06-30).

表 5-4 SBD マニュアルを用いたセキュリティ要件の検討結果

名称	観点分類	判断条件	判断結果	判断理由
A. 外部アクセスの有無	利用環境・手段	インターネット等の通信回線を介して(情報の管理ポリシーが異なる)外部から情報システムにアクセスしてサービスの利用、業務の遂行、情報システムの管理等を行うか。	○	次期システムは、政府共通ネットワークを経由して、各行政機関等が利用するものである。 利用する機関によって、情報の管理ポリシーは異なることが想定されるため、該当すると判断する。
B. 情報の重要度	情報	漏えいした場合、正常にアクセスできない場合或いは消失した場合に、深刻な損害を被る可能性がある重要性の高い情報を取り扱うか。	○	次期システムは、電子公文書等を長期的に保存することを目的とするものである。 当該電子ファイルの消失は社会的な問題になる可能性があるため、該当すると判断する。
C. 情報保存時の安全性	情報	入退室管理等の物理対策だけでなく、情報システムが保存する情報についてより一層の安全を期すために追加的対策をさらに行うべきと考えるか。	○	次期システムでは、秘密情報の中の極秘情報を扱うものではないが、情報の安全性は確実に担保する必要があることから、該当すると判断する。
D. 利用者の限定要否	主体	情報システムにアクセスする主体は、利用資格のある者、職員、グループのメンバー等の特定の者に限定されるか。	○	次期システムでは、運用管理サブシステムにより権限が付与された利用者のみが利用する前提であるため、該当すると判断する。
E. アカウントの多様性	主体	利用者によって利用可能なサービスや業務が異なる等、利用者の特徴にバリエーションがあるか。	○	次期システムは、館の担当係等によって利用する機能を制限する想定であることから、該当すると判断する。
F. 複数部局による利用	主体	情報の取り扱い方や利用目的等が異なる複数の部局等の中で共用されるか。	○	同上

次に、同マニュアルから導出された次期システムに係るセキュリティ要件として、考慮すべき事項を以下に示す。今後、次期システムの要件定義において、館としての業務要件や各対策要件から情報セキュリティ要件の具体化を図る必要がある。

表 5-5 SBD マニュアルに基づくセキュリティ対策要件

対策区分	対策方針	対策要件	
侵害対策 (AT: Attack)	通信回線対策(AT-1)	AT-1-1	通信経路の分離
		AT-1-2	不正通信の遮断
		AT-1-3	通信のなりすまし防止
		AT-1-4	サービス不能化の防止
	不正プログラム対策(AT-2)	AT-2-1	マルウェアの感染防止
		AT-2-2	マルウェア対策の管理

対策区分	対策方針	対策要件	
	セキュリティホール対策 (AT-3)	AT-3-1	構築時の脆弱性対策
		AT-3-2	運用時の脆弱性対策
不正監視・追跡 (AU: Audit)	証跡管理(AU-1)	AU-1-1	証跡の蓄積・管理
		AU-1-2	証跡の保護
		AU-1-3	時刻の正確性確保
	不正監視(AU-2)	AU-2-1	侵入検知
		AU-2-2	サービス不能化の検知
	主体認証(AC-1)	AC-1-1	主体認証
	不正監視(AU-2)	AU-2-1	侵入検知
		AC-2-2	アクセス権管理
AC-2-3		管理者権限の保護	
データ保護 (PR: Protect)	機密性・完全性の確保 (PR-1)	PR-1-1	通信経路上の盗聴防止
		PR-1-2	保存情報の機密性確保
		PR-1-3	保存情報の完全性確保
物理対策 (PH: Physical)	情報搾取・侵入対策 (PH-1)	PH-1-1	情報の物理的保護
		PH-1-2	侵入の物理的対策
障害対策(事業継続対応) (DA: Damage)	構成管理(DA-1)	DA-1-1	システムの構成管理
	可用性確保(DA-2)	DA-2-1	システムの可用性確保
サプライチェーン・リスク対策 (SC: Supply Chain)	情報システムの構築等の外部委託における対策 (SC-1)	SC-1-1	委託先において不正プログラム等が組み込まれることへの対策
	機器等の調達における対策 (SC-2)	SC-2-1	調達する機器等に不正プログラム等が組み込まれることへの対策
利用者保護 (UP: User Protect)	情報セキュリティ水準低下の防止(UP-1)	UP-1-1	情報セキュリティ水準低下の防止
	プライバシー保護(UP-2)	UP-2-1	プライバシー保護

5.3. 電子公文書等に係る館の保存対策を踏まえた機能改善

「3 電子公文書等に係る館の保存対策に対する対応」に示す各対応の実現に向けて、電子公文書等システムとして支援機能の実装が求められる。

5.3.1. 日本独自のファイルフォーマットの判定・登録機能

現行システムの運用では、電子公文書等システムで判定したフォーマット情報の妥当性を検証するために、DROID 及び PRONOM の情報を参考としている。しかし、この検証は標準的フォーマットに対してのみ実施されており、非標準フォーマットを含めて網羅的になされているものではない。また、日本独自で扱っているフォーマットに係る情報については、PRONOM に収録されておらず情報を抽出することが出来ない。例えば、CAD ファイルのフォーマットについては、国内のソフトウェアで利用しているフォーマットの移管件数が多く、PRONOM からは情報抽出が出来ないものが多い。そのため、これらのフォーマットについて、館でフォーマットの情報を管理する必要がある。

これまでの運用実績で、電子公文書等として移管され日本独自で扱うファイルフォーマットの種類は少数であることから、この登録・管理は可能と想定される。今後も DROID 及び PRONOM を引き続き利用すると共に、日本独自のフォーマットの判定、登録を可能にする機能が必要である。なお、電子公文書等システムで上記フォーマットの同定にあたっては、電子ファイルの拡張子に基づく判定となる点は留意しておく必要がある。

5.3.2. ファイルフォーマットの統計情報の出力機能

電子公文書等システムに登録されたフォーマット情報の統計情報を出力できるようにする必要がある。この機能を利用することにより、フォーマット対応表の更新にあたり、電子公文書等システムで保存しているフォーマットの情報収集が容易となることが期待される。

6. あるべき姿へ向けて

電子公文書等の長期保存のあるべき姿の実現に向けては、これまで記載した対応に加えて、館内の取組に留まらない、政府全体での取組が求められる。具体的には、「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）に沿うものであり、現用段階からの管理・保存・移管のライフサイクル強化に向けて、取組を進める必要がある。

6.1. 移管元行政機関等での現用時の文書管理状況の把握

「公文書管理の適正の確保のための取組について」（平成30年7月20日、行政文書の管理の在り方等に関する閣僚会議）では、“電子的な行政文書の所在情報管理の仕組みの構築”及び“作成から保存、廃棄・移管まで一貫して電子的に行う仕組みの検討”が示されている。

電子公文書等の適切な保存に向けては、館が現用段階から移管元行政機関等での電子公文書等の管理状況を把握することが必要不可欠であることから、上記仕組みの構築にあたっては、館に対して電子公文書等の作成や管理に関する技術的メタデータ等が共有されることが望ましい。

6.2. 移管元行政機関等に対する電子公文書等の管理及び保存にあたっての依頼

同じく「公文書管理の適正の確保のための取組について」では、今後作成する行政文書は電子的に管理することを基本とすることが定められている。移管元行政機関等からの電子公文書等の受入れを円滑に進めるためには、現用の行政文書ファイルの作成・取得段階から、標準的なファイルフォーマットや長期保存に適したファイルフォーマットの採用を考慮することが重要である。

移管元行政機関等における文書のフォーマット選定についての意識醸成に向けて、公文書管理に係る専門的・技術的な助言を行う立場の館による、移管元行政機関等への情報提供・依頼等が定常的に実施されることが必要である。

6.3. 電子公文書等の管理に係る方針の適宜見直し

電子公文書等の保存や管理、移管に係る館の方針は、内閣府等の方針に準拠している。例えば、現行のフォーマット対応表は、「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）を前提としている。しかし、本調査で整理した通り、ファイルフォーマットや活用する要素技術についてはその動向を踏まえた見直しが求められるものであり、より上位の方針が見直しの制約となることは望ましくない。電子公文書等の保存に係る実務及び実情に合わせて、同方針についても適宜見直しが必要とされるべきである。