

イルへ変換する作業は不可避である。

長期保存のためのファイルフォーマットに関するこれまでの取り組みは以下のとおりである。

2. これまでの取り組み

2.1 平成15、16年度

「公文書等の適切な管理、保存及び利用に関する懇談会」報告書等で、電子記録への対応が課題として指摘された。

2.2 平成17、18年度

「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」報告書では「基本的視点」として、以下の点が電子記録の管理、移管及び保存のあり方として指摘された。

- 1) 電子記録を非電子媒体に変換せず、電子媒体のまま保存する原則
- 2) 電子記録の証拠価値のため不可欠な「エッセンス」の長期保存
- 3) 電子記録については、作成時から保存・利用段階までのライフサイクル全体の管理が必要

ここでいう「エッセンス」とは、長期保存上の必要から電子公文書等の持つ機能の一部停止、レベルダウンを行い、電子文書の記録としての価値を維持するのに必要な情報のことを指している。

2.3 平成19年度

内閣府において、電子記録の移管と長期保存を実現するという観点から、電子記録のメタデータ及びフォーマットのあり方等について、実証実験が実施された。その結果、長期保存に必要なメタデータとして付与すべきデータの多くが、ファイルのプロパティ情報から抽出可能であることが確認された。また、電子記録作成に用いられているファイルフォーマットの多くが、長期保存フォーマットとして各国政府で採用されている PDF/A (ISO19005 1) に変換可能であ

ること、および音声、画像、動画などのフォーマット変換では、品質の劣化などの問題があることも確認された。

2.4 平成20年度

内閣府において、平成19年度の実証実験結果を踏まえ、試験システムを構築し、総合的検証を実施した。データの移管、保存、利用の3つのフェーズごとに、業務やシステムの実態、運用上の課題等が把握された。

3. 電子公文書等移管・保存・利用システムの検討に当たって

3.1 フォーマットについて

電子公文書等の移管で取り扱うべきフォーマットについては、『平成20年度電子公文書等の管理・保存・利用システムに関する調査報告書』（平成20年3月）（以下、『平成20年度内閣府調査報告書』と略す。）を基に、標準フォーマットと長期保存フォーマット（表1）という観点から検討している。報告書でいう標準フォーマットとは、現場で利用されている一般的なフォー

表1 標準フォーマット、長期保存フォーマット

フォーマット	標準フォーマット	長期保存フォーマット
文書作成	OASYS、一太郎 8 12、Word 97 2003、Word 2007、PDF、PDF/A、OpenOffice Writer	PDF/A
表計算	Excel 97 2003、Excel 2007、OpenOffice Calc	PDF/A
プレゼンテーション	PowerPoint 97 2003、PowerPoint 2007、OpenOffice Impress	PDF/A
画像	JPEG、JPEG 2000、GIF、TIFF、BMP	JPEG 2000
音声・音楽	WAVE、MP3、WMA	MP3（ビットレート256 kbps以上）
映像	QuickTime、Windows Media、RealPlayer、MPEG	MPEG 2

出典：『平成20年度内閣府調査報告書』pp.43 44を元に野口が作成。

マットを指しているといえる。長期保存フォーマットでは国際標準に対応していることが要件になっている。これは開発した機関・会社が無くなっても、国際標準規格であれば現存する他の機関・組織によって再現する手段が提供されていると考えられるためである。なお、平成20年度の調査結果では結論を得られなかったものの、システムでの部分対応が可能と考えられるもの（電子メール、Web データ、データベースやCAD の汎用フォーマット等）も、当館では利用方法についての課題は残しつつ、保存対象には含めている（後述4.3参照）。

3.2 メタデータについて

メタデータモデルについては、平成20年度内閣府調査結果に基づき、OAIS 参照モデルを基本枠組みとしている。管理オブジェクトの基本単位は「行政文書ファイル」としている。行政文書ファイルには、複数の行政文書が収まっており、それら個々の行政文書の下にはドキュメント、さらにその下に電子ファイルというように枝分かれして行く。それら個々の電子公文書には、異なるファイル形式も含まれるだろう。そこで、ひと固まりの行政文書ファイルとして、個々の文書がどのように関連しているのかをメタデータとして付与しなくてはならない。これを紐づけという。実際には、表2のように1から3までの各メタデータを束ねたコンテナメタデータが必要とされる。

表2 メタデータの定義

項番	メタデータ名称	定義
1	記録管理メタデータ	電子公文書等の作成時の府省名、分類、文書名、作成者等行政文書ファイル管理簿の項目並びに、その他文書内容に係る情報を管理するためのメタデータを意味する。
2	技術的メタデータ	電子公文書等の電子ファイルのファイル形式、作成アプリケーション名称とそのバージョン、作成アプリケーションが動作するOS、電子署名やタイムスタンプ等の作成時の技術的内容のメタデータを意味する。また、

		移管・保存時に長期保存フォーマットに変換される場合は、変換時の技術内容も対象となる
3	アーカイバルメタデータ	電子公文書等の移管後に、非現用文書として管理する内容（受入情報、保存延長情報、公開情報、破棄情報）のメタデータとして定義する。
4	コンテナメタデータ	上記の記録管理メタデータ、技術的メタデータ、アーカイバルメタデータのすべてを包含したメタデータ。電子公文書1件に対して、1個のコンテナメタデータが対応する。

出典：『平成20年度内閣府調査報告書』p.60。

3.3 補足

電子公文書を扱う際には、できることとできないことを分けて、何が最も重要かを考える必要がある。

また、電子ファイルの保存においては、リスク、ミスは必ずあると考えなくてはならない。だからこそシステムに人為的リスクコントロールを盛り込む必要が生まれる。電子公文書自体の保存・変換作業が重要であるのみならず、電子ファイルやメタデータに関する処理や作業の記録を管理、保存することも必要である。ファイルの中身のデータとは別に、ファイルの生成・管理に関する情報（記録管理メタデータ、技術的メタデータおよびアーカイバルメタデータ）を管理しなくては、電子公文書は役に立たない。次の世代に文書を引き継ぐことを第一の目的として、機能・性能・業務負担を勘案しながらシステムを構築してほしい。

4. 電子公文書等の移管・保存・利用システムの概要

国立公文書館では、「重点計画2008」（IT戦略本部）に基づき、平成23年度にシステムを稼働させ、電子公文書等の移管・保存・利用を実現する予定である。行政機関と一般国民両方の利用を前提にしている。システム化に当たっての基本的な考え方は以下のようにまとめることができる。

4.1 システム化に当たっての基本的な考え方
 システム化に当たっては、以下のような考え方に配慮する必要がある。

- ・見読性の保証
- ・原本性の保証
- ・「いつでも」、「どこでも」、「だれもが」、「自由に」、「無料で」インターネットを通じ、いつでも利用可能
- ・システムの冗長化
- ・機能ごとの分離、検疫
- ・アクセス制御によるセキュリティ向上
- ・長期保存データの適切なバックアップ

4.2 電子公文書等の移管・保存・利用に係る4つの類型

基本的な考え方を、実際の業務フロー図にまとめると図2の類型に当てはまる。

4.3 電子公文書等の移管・保存・利用システムで対応する標準的フォーマット、標準的媒体

長期的な保存、将来の利用者が閲覧できるフォーマットへの変換を考えると、当館のシステムで対応する標準的フォーマット、標準的媒体は表

3のようにまとめることができる。電子公文書等の種類によっては、システムで部分対応するフォーマットや、個別対応を検討すべきもの、または保存のみで対応するものもある。

4.4 電子公文書等の移管・保存・利用システム概要

当館の電子システムの概要は図3のようになる。移管から一般利用までを6つの段階に分けて説明する。

1) 移管

移管の方法としては、電子媒体（標準媒体、必要に応じ媒体を貸与）、またはネットワークを介した移管を想定している。行政文書ファイル毎に、電子公文書等のフォルダ、ファイル等の構成、ファイルフォーマット、作成時のシステム等に関する情報を可能な限り書き込む。このメタデータが無ければ電子文書は意味がない。電子媒体やメタデータの取り扱いに係る技術的な支援も必要である。

2) 受入

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、媒体変換を行い、検疫を実施する。検疫は三回に分けて行う。1次検

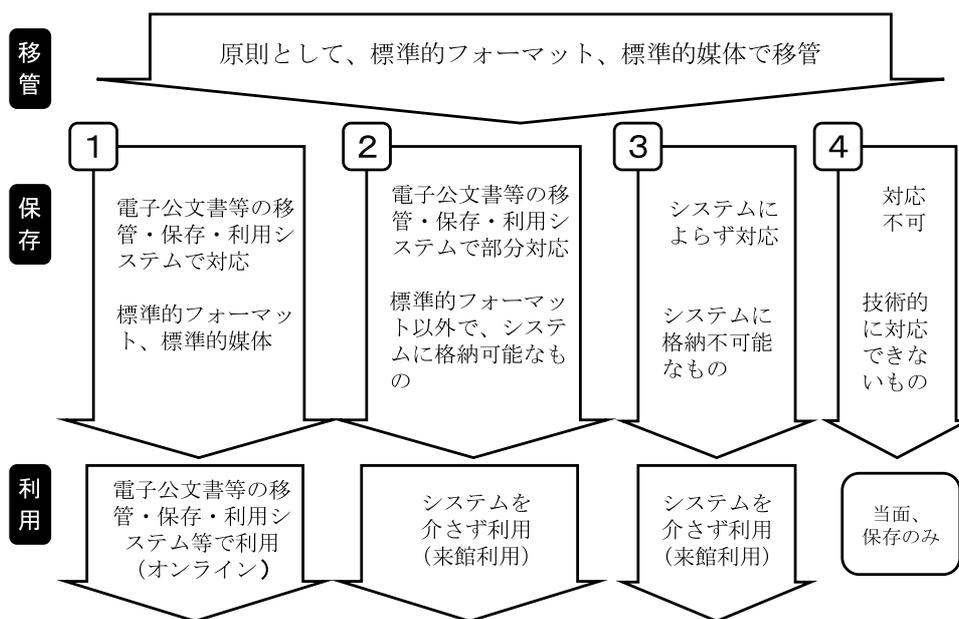


図2 電子公文書等の移管・保存・利用に係る4つの類型

表3 国立公文書館におけるフォーマット・媒体の対応

電子公文書等の移管・保存・利用システムで対応する標準的フォーマット、標準的媒体				
種類	標準的フォーマット		長期保存フォーマット	デジタルアーカイブ用フォーマット
文書作成	OASYS 一太郎 8-12 Word 97-2003 Word 2000	PDF PDF/A OpenOffice Writer	PDF/A	PDF/A
表計算	Excel 97-2003 Excel 2007	OpenOffice Calc	PDF/A	PDF/A
プレゼンテーション	PowerPoint 97-2003 PowerPoint 2007	OpenOffice Impress	PDF/A	PDF/A
画像	JPEG JPEG 2000 GIF	TIFF BMP	JPEG2000 (lossless)	JPEG2000 (lossy) PDF (JPEG2000)
音声	WAVE MP3	WMA	MP3 (ビットレート 256kpbs以上)	MP3 (ビットレート 256kpbs以上)
映像	QuickTime Windows Media	RealPlayer MPEG	MPEG-2	MPEG-2

標準的媒体: CD-R, DVD-R, HDD, FD

2	システムで部分対応するフォーマット、媒体		3	個別対応
				1及び2以外の フォーマット及び媒体
			4	保存のみ

種類	標準的フォーマット以外		その他媒体
Web、電子メール、CAD、データベース、等	htm, html	DWG, DXF, SXF Ai, vsd, vxd, wk	MO

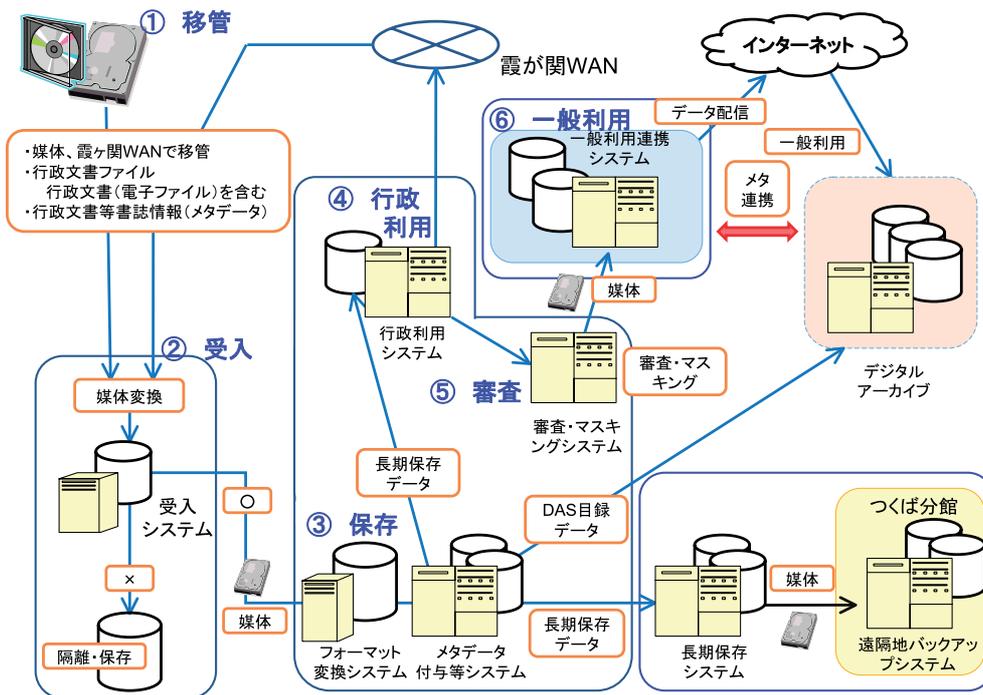


図3 移管・保存・利用システム概要図

疫はクライアントで行ってもらう。その後、1次と異なる検知システムを用いて当館が2次検疫をする。そして、2次検疫後、一定期間をあけて第3次検疫を行う。なお、ウィルスによる被害リスクを考慮し、検疫システムは当館のメ

インシステムとは物理的に隔離する。

3) 保存

電子文書の入った媒体を受け入れた後、そこからデータを取り出し、オリジナルファイルとフォーマット変換したファイルの二種類を保存

する。この際、ファイルからテキストデータとメタデータを抽出する。メタデータは元々含まれているものに加えて、当館が保存や利用のために新たに付与するものもある。付与されたデータは、例えば、一般利用者がデジタル・アーカイブで検索する時に利用することになる。

メタデータの作成が終わった後、つくば分館での遠隔地バックアップやミラーリングバックアップの措置を施しながら、長期保存されることになる。長期保存中もデータ内容の同期チェックなどを定期的に行う。

4) 行政利用

電子行政利用に当たっては、ユーザー認証を経て、霞ヶ関 WAN 経由でオンライン利用できるようにする。その際、オリジナルと長期保存フォーマットをそれぞれ利用できるようにする。もしも、オンライン利用が困難な場合は、来館利用で対応する。

5) 審査

個人情報や法人情報、あるいは国の安全など、要審査の扱いとなっている文書は現在も多く所蔵されている。公開扱いになっていない文書は、当館の専門官室で文書を審査した上で、一部非公開として該当ページを袋掛けするか、あるいは複製を作成して、非公開箇所を墨塗りしてから公開している。

電子文書の場合も審査作業が発生するため、長期保存されたファイルから作業用ファイルを作成し、PDFの「墨消し機能」で処理し、一部非公開として該当ページを袋掛けするか、あるいは複製を作成して、非公開箇所を墨塗りしてから公開している。

電子文書の場合も審査作業が発生するため、長期保存されたファイルから作業用ファイルを作成し、PDFの「墨消し機能」で処理することも可能である。JPEG2000の場合でも画像編集アプリでスクリーニング処理することが可能である。よっていずれの場合でも作業効率はあるだろう。ただし、このほかの音声・映像、

非標準フォーマット等については、今後の課題である。

6) 一般利用

閲覧には、長期保存フォーマット（審査・マスキング作業、変換処理済）を使用し、デジタル・アーカイブ上では従来の媒体（紙、マイクロ等）と電子文書の閲覧は、統合して利用するため、利用者の視点から見れば媒体の違いのない形で利用可能になる。インターネットを通じて、検索・閲覧できる点も同じである。

5. まとめ

電子公文書のシステム作りには、まず関係者が集まって業務的なルールや技術的なルール作りをしなくてはならない。その上で、人員・体制・予算の確保が必要である。システムは一度作って終わりではなく、将来のデータ移行への配慮も必要である。また、十分にシステム設計を行ったとしても、リスクは前提としておくべきことも忘れてはならない。

ともあれ、メタデータとファイルを保存するという原則が重要なのであって、それ以外の点については、できることから始めればよいと考え、システム作りに取り組んで頂きたい。

おわりに

講義を振り返ってみると、不断に新しいOS、アプリケーション、ファイル形式が生まれていくことが文書管理に大きな問題を投げかけていることが分かった。紙と違って、作った時点と利用する時点とでは、媒体やファイル形式が異なることも多いわけだから、変換作業は電子文書を保存する限り永遠についてまわる。誰もがいつでも簡単に使えるハードやファイル形式で電子公文書を利用するためには、公文書館側は絶えざる変換作業をしなくてはならない。