

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  
報 告 書

平成 21 年 3 月

内閣府



## 【 目 次 】

<b>1. 全体概要</b> .....	<b>- 1 -</b>
1.1. 調査の目的.....	- 1 -
1.2. 平成 20 年度の調査の位置づけと調査内容.....	- 3 -
1.3. 各府省向けデモンストレーションでの意見.....	- 9 -
1.4. 委員会等.....	- 16 -
1.5. 調査実施体制.....	- 17 -
1.6. 調査実施スケジュール.....	- 18 -
<b>2. ファイルフォーマット</b> .....	<b>- 19 -</b>
2.1. 調査概要.....	- 19 -
2.2. ファイルフォーマット調査報告.....	- 20 -
2.3. 結論.....	- 42 -
<b>3. メタデータスキーマ</b> .....	<b>- 47 -</b>
3.1. 調査概要.....	- 47 -
3.2. 各種メタデータ標準等の調査.....	- 47 -
3.3. 電子公文書におけるメタデータの定義.....	- 60 -
<b>4. 受入・保存管理機能</b> .....	<b>- 83 -</b>
4.1. 調査概要.....	- 83 -
4.2. 検疫・媒体変換機能実証結果報告.....	- 84 -
4.3. フォーマット変換機能実証結果報告.....	- 87 -
4.4. アーカイバルメタデータ作成機能実証結果報告.....	- 113 -
4.5. 保存機能実証結果報告.....	- 130 -
4.6. 結論.....	- 135 -
<b>5. 行政利用機能</b> .....	<b>- 139 -</b>
5.1. 調査概要.....	- 139 -
5.2. 行政利用向けデータ管理機能実証結果報告.....	- 140 -
5.3. 行政利用向け検索機能実証結果報告.....	- 143 -
5.4. 行政利用向け閲覧機能実証結果報告.....	- 149 -
5.5. 業務管理方法.....	- 157 -
5.6. 結論.....	- 158 -
<b>6. デジタルアーカイブ連携機能</b> .....	<b>- 161 -</b>
6.1. 調査概要.....	- 161 -
6.2. デジタルアーカイブ用フォーマット変換機能実証結果報告.....	- 162 -
6.3. デジタルアーカイブ用目録データ作成機能実証結果報告.....	- 168 -

6.4.	デジタルアーカイブ・システム変更調査報告 .....	- 179 -
6.5.	結論 .....	- 181 -
<b>7.</b>	<b>業務実施要件報告 .....</b>	<b>- 183 -</b>
7.1.	業務フロー .....	- 183 -
7.2.	業務実施にあたっての問題・課題 .....	- 186 -
<b>8.</b>	<b>全体総括 .....</b>	<b>- 189 -</b>
8.1.	今年度の調査結果のまとめ .....	- 189 -
8.2.	次年度以降の検討課題 .....	- 195 -

# 1. 全体概要

## 1.1. 調査の目的

平成 17 年度より、内閣府では、保存期間の満了した電子公文書等について、電子媒体による国立公文書館への移管及び保存を平成 23 年度から開始するという計画に基づいて、調査研究等の取組みを行っている。本調査研究は、その取組みの一環として内閣府の委託を受けて実施したものである。

電子公文書等の移管はその移管に係る取り決めがないため、当該移管を開始するに当たっては、移管後の長期保存フォーマットへの変換等を含む現行の紙媒体公文書等の移管ルールにはない新たな電子公文書等専用の取り決めを、公文書移管制度を所管する内閣府として各府省と調えなければならず、その前提として平成 19 年度までの電子公文書等の移管関係の調査研究結果も踏まえた電子公文書等の移管制度設計に必要な調査を実証実験により総合的に行う必要がある。本調査は、電子公文書の移管等を平成 23 年度より着実に実施するため、受入・保存から行政利用、一般国民の利用までの、移管後の電子媒体の公文書の具体的な流れに即して、各機能間の連携のとれた一連の環境により、システムを構築するために必要な諸要件および課題等の抽出・整理を行うことを目的とする。

平成 19 年度までの調査研究では、将来の移管・移送を確保するための長期保存フォーマットの策定、「原本性」確保のルール化等について専門的実証的な調査研究を行ったが、平成 20 年度の調査においては、これまでの調査結果を踏まえつつ、プロトタイプシステムの構築と実証試験による技術的側面の調査を中心に行った。プロトタイプシステムでは、電子媒体の公文書を、電子のまま受入れてから電子媒体として一般の利用に供するまでの「受入・保存管理機能」、「行政利用機能」及び「デジタルアーカイブ連携機能」の 3 つの機能及び各機能間の連携等について調査・検証を行った。

また、各府省に対しては、実際の環境により近い形で、電子媒体の公文書等の移管後の具体的な提供方法等を提示することにより、今後の電子媒体の移管についてのルール化の協議等に資するとともに、平成 23 年度稼働予定のシステムにその意見を反映させるものとする。

### 1.1.1. 本調査研究が対象とする電子公文書等の範囲について

保存期間の満了した公文書等のうち国立公文書館に移管されるのは、「歴史資料として重要な公文書等」と認められるものである。現在、国立公文書館への移管は、

- (1)「歴史資料として重要な公文書等の適切な保存のために必要な措置について」(平成 13 年 3 月 30 日閣議決定)」
- (2)「歴史資料として重要な公文書等の適切な保存のために必要な措置について(平成 13 年 3 月 30 日閣議決定)の実施について」(平成 13 年 3 月 30 日各府省庁官房長等申合せ(改正 平成 17 年 6 月 30 日))
- (3)「歴史資料として重要な公文書等の適切な保存のために必要な措置について(平成 13 年 3 月 30 日閣議決定)等の運用について」(平成 13 年 3 月 30 日各府省庁文書課長等申合せ(改正 平成 17 年 6 月 30 日))

のいわゆる「移管基準」に基づいて、行われている。これらの「移管基準」では、

(1) 国政上の重要な事項等に関する意思決定に係る決裁文書、当該意思決定に至るまでの過程(審議、検討、協議等)の記録

(2) 当該意思決定に基づく施策の遂行過程等の記録

などの文書が移管対象とされている。つまり、公文書等作成のコンテキスト(背景・状況・環境)としての各府省庁の活動、当該活動が結果として社会に及ぼした影響、文書自体の内容等の重要度によって、移管の適否が判断されているのである。従って、公文書等が「歴史資料として重要な公文書等」とであると判断されれば、その記録様式、記録媒体等がいかなるものであっても、移管対象となる。

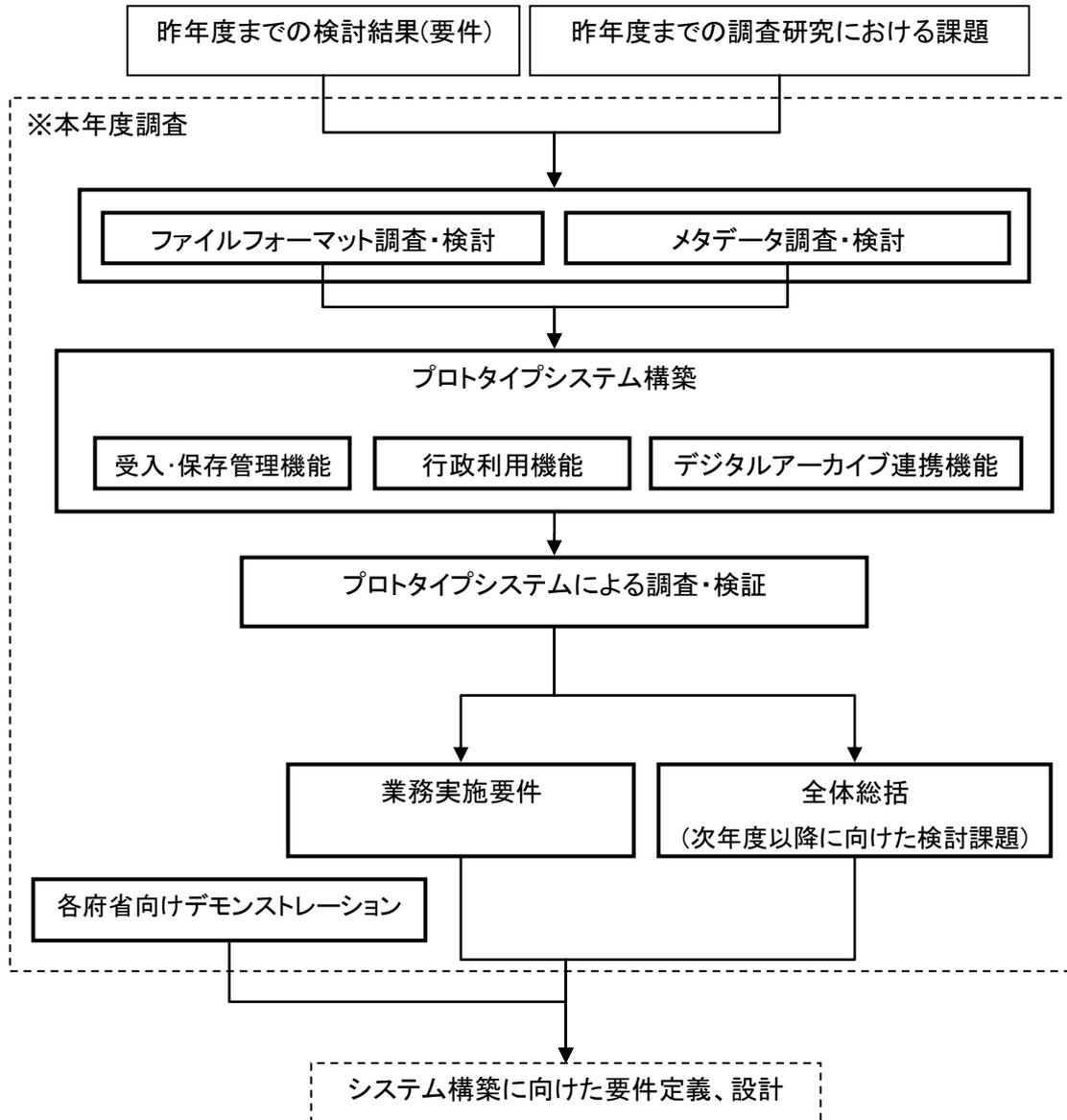
このような移管基準の存在に鑑みて、基本的には、記録様式、記録媒体等のいかににかかわらず、あらゆる電子公文書等を本調査研究の対象とするのが適切であると考えられる。ただし、過去数年の調査等の結果から、各府省庁等の活動の過程で作成される公文書等の多くは、ワープロ、表計算及びプレゼンテーション用のソフトウェアで作成されるテキスト形式の文書であると考えられるので、この種の記録様式の電子公文書等を本調査研究の対象として優先的に取り扱うこととした。また、行政機関がその施策等を一般に周知させることを目的として作成した広報資料が移管対象となっている。これらの広報資料には、画像、音声及び映像などの記録様式による電子公文書等が一定の比率で存在していると考えられることから、この種の記録様式の電子公文書等も本調査研究の対象として取り扱うこととした。一方、電子公文書等のなかには、データベース、CAD等動的システムに依存し、これと一体化してはじめて存在し得るものもあるが、これら動的システムについては、現時点では関連業界、研究機関等でも長期保存に関する検討が正に「検討段階」であって一定のソリューションが得られていないことに鑑み、今回の調査研究の対象からは除外することとした。

なお、現在、「公文書等の管理に関する法律」案が検討中であるが、同法律案が成立・施行しても、現在の移管基準のような考え方に基づいて、記録様式、記録媒体等のいかににかかわらず、移管の適否が判断されるものと思われる。

## 1.2. 平成 20 年度の調査の位置づけと調査内容

### 1.2.1. 平成 20 年度の調査の位置づけと流れ

平成 20 年度の調査の位置づけと流れは以下の通りである。



## 1.2.2. 前年度までの報告書で提案された方針と本調査での調査研究内容

### (1) 前年度までの報告書の要件等

本調査の実施にあたり、平成 19 年度までの調査結果における論点、検討結果について表 1-1 にまとめた。

表 1-1 平成 19 年度までの調査結果における論点及び検討結果

<p>1.移管対象となる電子公文書等の範囲と内容</p> <p>(1)移管対象となる電子公文書について</p> <ul style="list-style-type: none"><li>・ 今後電子公文書の移管が増加することが予想されるが、各府省の現用電子文書は国際規格かつ普及率の高いファイルフォーマットで作成されている。</li><li>・ 移管対象の中心となる行政文書は主にテキスト文書、表グラフ、プレゼンテーション文書であるが、動画・静止画像・音声ファイル等のテキスト以外の公文書も移管されている。</li></ul> <p>(2)テキスト文書以外の電子公文書等について</p> <ul style="list-style-type: none"><li>・ データベースや CAD 等の電子ファイルについて、電子公文書としての移管方針や移管方法、保存方法について今後検討する必要がある。</li><li>・ 電子メール、Web コンテンツについて、電子公文書としての位置づけ、移管方針の検討が望まれる。</li></ul> <p>2.電子公文書等のメタデータ、フォーマット</p> <p>(1)電子公文書等の長期保存におけるメタデータの要件と今後のあり方</p> <ul style="list-style-type: none"><li>・ 「アーカイバルメタデータ」について、国際的な標準や相互運用性を考慮した項目を維持しつつ、技術情報(技術的メタデータ)等についても保持する必要がある。格納(交換)形式は XML 形式とし、技術的メタデータについては可能な限り自動的に付与することが望ましい。</li><li>・ メタデータの長期利用のために、作成時の文字コードを附帯情報として保持しておく必要がある。</li><li>・ 現状では移管時のメタデータとして利用できる最低限の情報は行政文書ファイル管理簿から取得することができるが、移管後の電子公文書等を長期に利用可能な状態で保存するためには記録管理メタデータの内容を充実させていくことが不可欠となる。そのためには文書管理システムの適切な運用、行政職員の作業負荷軽減に配慮することが重要となる。</li><li>・ メタデータに含まれる行政組織に関する情報は流動性を持っているため、長期の利用にあたっては組織変更に対応でき、また組織間での相互運用性も考慮したスキーマが必要となる。また、メタデータを長期にわたって維持・管理するための体制、仕組みも必要となる。</li></ul> <p>(2)電子公文書等の標準フォーマット、長期保存フォーマットの考え方</p> <ul style="list-style-type: none"><li>・ フォーマット変換の実験結果より、電子公文書等の作成時の標準フォーマットとして以下を推奨する。</li></ul> <p>文書:PDF(PDF/A を含む) 音声:MP3(MPEG-1 Audio Layer-3) 動画:MPEG-2</p>
--

静止画: JPEG(JPEG 2000 については今後検討が必要)

- ・ ISO 32000-1(PDF)<sup>1</sup>規格は長期保存の目的で制定されており、原本を添付ファイルとして埋め込むことも可能であり、長期保存フォーマットとして有望である。
- ・ 長期保存フォーマットへの変換に際して、保存対象となる電子公文書等の種類、保存目的、エッセンス要件に応じたフォーマットを選択する必要がある、そのためのガイドラインが必要である。

#### (3)長期保存における文字コード、文字フォントに関する要件

- ・ 長期保存の観点からは、OS やアプリケーションに依存しない文字コード、文字フォントを文書の作成段階で使用することが望まれるが、現時点では十分な解決策がない状況である。文字コードテーブルや文字フォントセットを長期にわたり維持・管理していく仕組みが必要となる。

#### (4)電子公文書等のエッセンス

- ・ 「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っている機能の一部の停止、または機能の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である。
- ・ 電子公文書等の保存目的と照らし合わせると、「見読性」「完全性」「可用性」の 3 つが重要となる。
- ・ ファイルフォーマットの変換により情報や機能の欠落等が発生するため、「完全性」を担保するには原本データによる保存が望まれるが、原本データのフォーマットが将来にわたり再生可能であるかは保証されないため、エッセンスの抽出が必要となる。
- ・ 見読性の確保を優先する場合はエッセンス保存を行うが、定期的なフォーマットのマイグレーション、見読性の検証を行っていくことが望ましい。

#### (5)ライフサイクル管理に対応した電子公文書等の移管、保存のあり方

- ・ 現状の枠組みにおいては、各行政文書は保存期限に達し非現用となった時点で、国立公文書館へ移管されるか若しくは廃棄される。移管に際して、現状の紙文書と同様に、記録管理メタデータが移管される各文書ファイルに紐付けされて移管される。
- ・ 国立公文書館では、移管された電子公文書ファイルから、技術的メタデータの自動抽出、作成を行い、長期保存のための管理情報を付与して「アーカイバルメタデータ」として一元化し、保存・参照履歴、複製ファイルの作成等の記録を保存する。

## (2) 本調査における調査研究内容

本調査では、平成 19 年度までの調査結果を踏まえながら、次の事項について調査・検証を行った。

### ① 対象とする電子媒体の公文書のファイルフォーマット

---

<sup>1</sup> 平成 19 年度までの調査では「ISO/DIS 32000」であったが、2008 年 7 月に ISO により「ISO 32000-1」として標準化されたため、以降「ISO 32000-1」と表記する。

② 長期保存のためのメタデータスキーマ調査

③ 受入・保存管理機能調査

受入・保存管理に必要な、検疫・媒体変換、フォーマット変換、アーカイバルメタデータ作成、保存の各機能について、各機能の評価、連携方法、実運用における課題と対策を調査する。

④ 行政利用機能調査

行政利用に必要な、行政利用向けデータ管理、行政利用向け検索、行政利用向け閲覧の各機能について、実稼動環境等での検証等を踏まえて、各機能の評価、連携方法、実運用における課題と対策を調査する。

⑤ デジタルアーカイブ連携機能調査

「行政利用機能」に保管されている電子媒体の公文書のうち、保存年限を満了した電子媒体の公文書を、国立公文書館デジタルアーカイブで一般の利用に供するために必要な、デジタルアーカイブ用フォーマット変換、デジタルアーカイブ用目録データ作成の各機能について、実稼動環境等での検証等を踏まえて、各機能の評価、連携方法、実運用における課題と対策を調査する。

### 1.2.3. 前年度までの残課題の解決状況

平成 19 年度までの調査結果で提起された次年度以降の検討課題を表 1-2 にまとめるとともに、本調査においてそれぞれの課題に対してどのような解決を図ったか(または未解決のままとなっているか)を記載した。

表 1-2 平成 19 年度までの調査結果で挙げられた検討課題

検討課題	本調査での解決	
(1) 全般的事項	移管対象とする電子公文書等の定義、範囲、条件等の確認	「業務実施要件」において、ガイドライン案として一定の方向性を提示した。
	移管文書の原本に関する考え方の確認	「業務実施要件」に記載したが、明確な結論は出ていない。今後も引き続き議論が求められる。
	移管公文書等の保管目的に応じたメタデータ、保存フォーマットの関係整理	保管目的や公文書のライフサイクルという観点からの関係整理は十分に行えていない。今後も検討が必要である。
	一元的な文書管理システム稼動後における移管公文書等に関する想定要件の確認	「一元的な文書管理システム」との連携可能性は見えたと、具体的な要件までは踏み込めていない。
(2) メタデータ関連	「アーカイバルメタデータ」(試案)の検証、評価	試案を検証・評価し、メタデータスキーマを作成した。
	移管対象文書の府省における作成段階での記録管理メタデータの内容充実に向けた運用ガイドラインの検討	「業務実施要件」において、ガイドライン案として一定の方向性を提示した。ただし各府省との調整や検討が必要である。

検討課題		本調査での解決
	技術的メタデータに関する運用実験	メタデータスキーマにおいて定義した技術的メタデータに関して、プロトタイプシステム上で検証を実施した。
	アーカイバルメタデータの付与に関する業務フローの検討と運用システムの仕様検討	プロトタイプシステム上でメタデータ付与の仕組みを検証。業務フローへ反映した。
(3) 文書フォーマット関連	文書作成用の標準フォーマット(推奨)から、長期保存のための移管フローと文書フォーマット変換のためのワークフロー作成	プロトタイプシステム上でのワークフローは作成したが、本システムの運用を想定したワークフローの検討が必要。
	テキスト文書の長期保存フォーマットとしての ISO 32000-1 仕様の継続調査・検証	今回の調査では PDF/A を中心として調査を実施したため、ISO 32000-1 については十分なフォローはできていない。2008 年 7 月に標準化されたばかりであり、今後も継続して調査・検証する必要がある。
(4) 動画、音声、HTML、電子メール等に関する移管、保存に関する指針の検討	移管、保存に関する指針の作成	動画、音声、HTML、電子メール等については具体的な指針までは作成できなかった。引き続き検討が必要である。
	作成用のファイルフォーマット、アプリケーション(推奨)の検討	今回の調査では検討できなかった。フォーマット変換における技術的な課題もあるため、引き続き検討が望まれる。
	エッセンスの作成方針や作成方法に関する検討	今回の調査では検討できなかった。引き続き検討が望まれる。
データの原本性を担保するためのセキュリティ技術の検討		電子ファイルの同一性を担保する手段としてのチェックサムを追加した。署名情報をメタデータとして追加したが、電子ファイル自体の原本性、真正性については引き続き検討が必要である。
その他	長期保存フォーマット変換等における文字コード、文字フォントに関する課題への対応	文字フォントについては明確な対応策は見出せていない。全府省の横断的な仕組みや体制が必要となるため、今後の検討課題とする。
	パスワード付きの文書やファイル圧縮された文書の取り扱い方針とその運用方法の検討	「業務実施要件」において、ガイドライン案としてセキュリティ設定の解除という方針を提示した。圧縮された文書についてはシステム上で展開した上で処理することは可能であるが、原本性やセキュリティ面での課題があるため、引き続き検討が必要である。

検討課題		本調査での解決
	メタデータに登録された機関名・組織名、各種コード名等の長期にわたる情報更新(情報メンテナンス)等の長期保存上の対応策の検討	今回の調査では明確な回答は出せなかったため、行政利用における利用者管理方法も含め、今後の検討課題とする。

### 1.3. 各府省向けデモンストレーションでの意見

本調査で構築したプロトタイプにより、各府省の関係者にデモンストレーションを実施した。以下にデモンストレーションの実施概要と、参加者からの意見として挙げられた内容をまとめる。(プロトタイプシステムの概要については「8. 全体総括」の「8.1.1 今年度構築したプロトタイプシステムについて」を参照)

#### 1.3.1. デモンストレーション実施概要

##### 【実施日時】

2009年3月10日(火)

以下の3回に分けて実施した。

第1回目 10:00～11:30

第2回目 14:00～15:30

第3回目 16:00～17:30

##### 【実施場所】

内閣府会議室

##### 【参加者数】

第1回目 10名

第2回目 11名

第3回目 8名

合計 29名

##### 【目的及び内容】

デモンストレーションでは、電子公文書等の移管から保存、利用、国立公文書館デジタルアーカイブ・システムへの連携までの具体的な流れを提示するとともに、平成23年度からのシステム運用に向けた問題、課題の情報共有を図ることを目的とした。

また、ヒアリングシートを準備し、デモンストレーション終了後に参加者に記入を依頼した。

デモンストレーションの実施内容は以下の通りである。

1. システム全体概要説明
2. 受入・保存機能説明
3. 行政利用機能説明(プロトタイプシステムの画面を使ったデモンストレーション)
4. デジタルアーカイブ連携機能説明
5. 質疑・応答
6. ヒアリングシート記入、回収

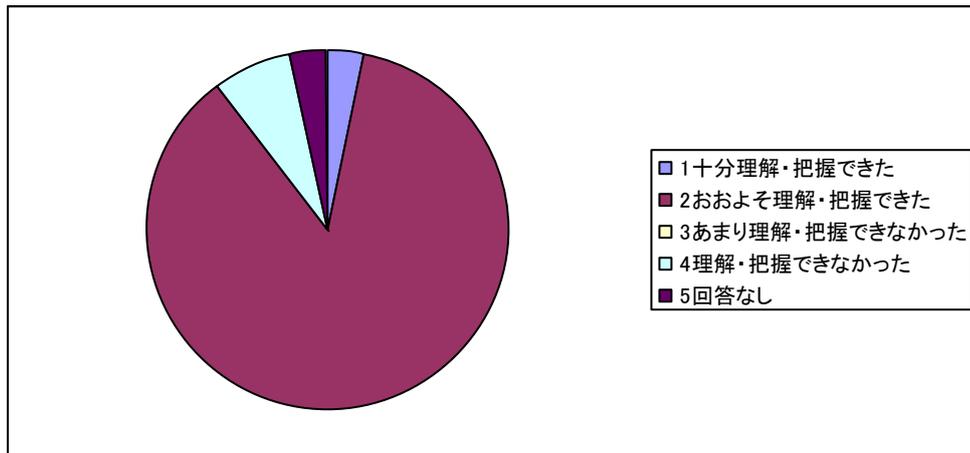
なお、デモンストレーションで使用した資料、ヒアリングシートについては資料編を参照のこと。

### 1.3.2. 参加者からの回答集計結果、意見等

参加者にヒアリングシートに記入してもらった回答内容を以下にまとめる。

【設問 1-1】 電子公文書等の移管からデジタルアーカイブ連携までの流れについて、理解できたか。

1 十分理解・把握できた	1
2 おおよそ理解・把握できた	25
3 あまり理解・把握できなかった	0
4 理解・把握できなかった	2
5 回答なし	1



今回の説明した電子公文書等の移管からデジタルアーカイブ連携までの流れについては、おおよその参加者から「理解・把握できた」との回答があった。

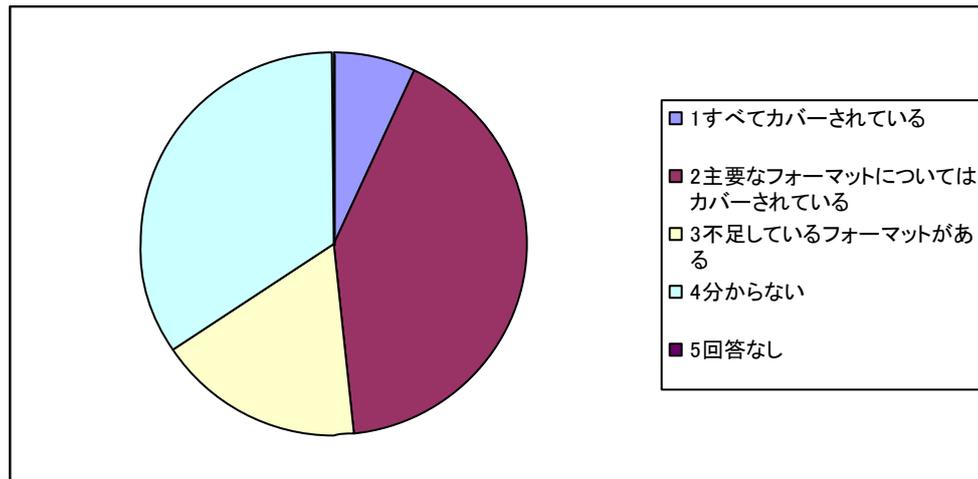
しかし、個別の意見として

- やろうとしていることは理解できたが、一元的文書管理システム等とのシステム連携など見えてないため全体像を把握できない。
- 本日初めてこのようなシステムが作られていることを知ったので、まったく理解できなかった。
- 電子公文書の範囲・イメージが定まっていない状況にあるため移管業務につなげにくい。
- 何について準備などをしていかなければならないのかが、良く分からない。また、認証システムについてよく理解できなかった。

等の意見が挙げられており、特に各府省側でどのような対応が必要となるのか、といった点に質問や意見が多かった。

【設問 2-1】 標準フォーマットとして定義されているフォーマットについて、府省での文書フォーマットと照らし合わせてすべてカバーされているか。

1 すべてカバーされている	2
2 主要なフォーマットについてはカバーされている	12
3 不足しているフォーマットがある	5
4 分からない	10
5 回答なし	0

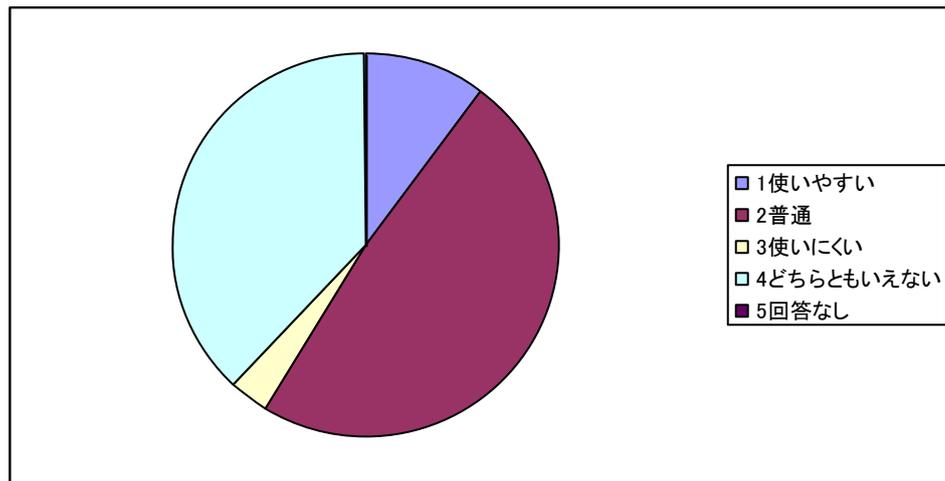


標準フォーマットの定義については、すべてまたは主要なフォーマットがカバーされているとの回答が半数であったが、「分からない」という回答も多かった。

不足しているフォーマットとして挙げられたフォーマットの中では「DocuWorks」が4件と最も多く、少数ではあるがCAD 図面、プレーンテキスト、HTML、XML 等の意見があった。

【設問 3-1】 行政利用向け検索機能、閲覧機能の各インターフェースについて、使いやすいと感じたか。

1 使いやすい	3
2 普通	14
3 使いにくい	1
4 どちらともいえない	11
5 回答なし	0

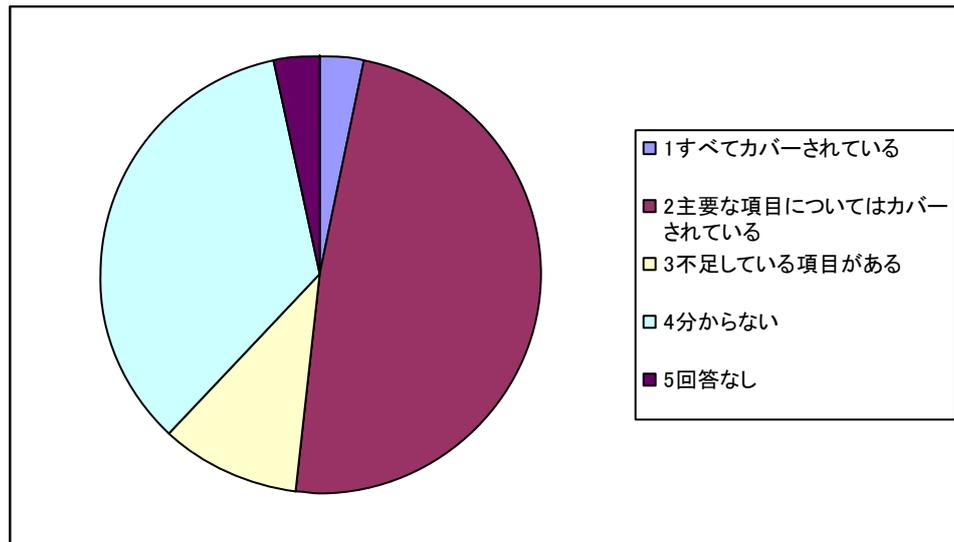


行政利用向け検索機能、閲覧機能の各インターフェースについての評価は、「普通」という意見が半数で、「どちらともいえない」という意見も多かった。

今回のデモンストレーションでは説明者が操作を行なう形だったため、「実際に使ってみないと分からない」という意見が実際のところであろう。

【設問 3-3】行政利用向け検索機能の詳細検索で、検索項目は行政利用の観点から十分な項目となっているか。

1 すべてカバーされている	1
2 主要な項目についてはカバーされている	14
3 不足している項目がある	3
4 分からない	10
5 回答なし	1

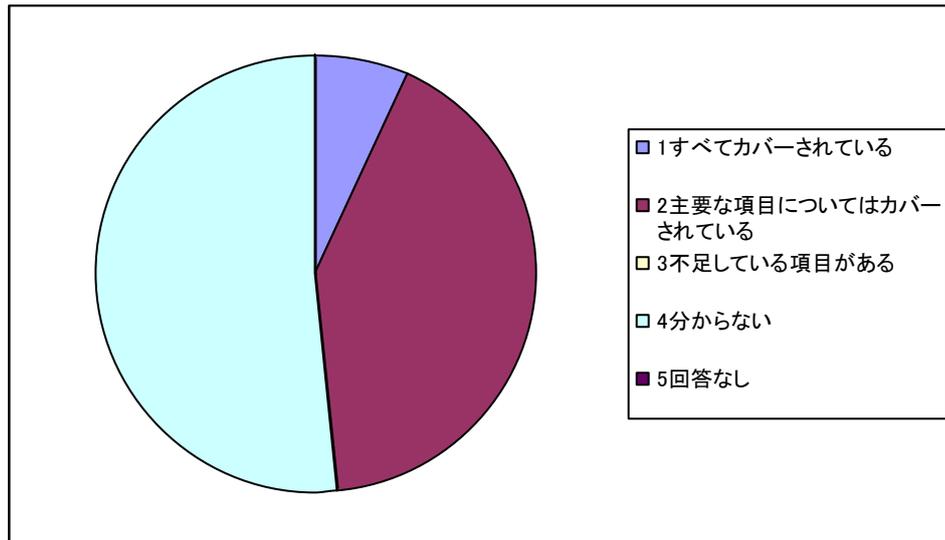


行政利用向け検索機能での詳細検索項目としては、すべてまたは主要な項目がカバーされている、との回答が半数以上であったが、「分からない」という回答も多かった。これは設問 3-1 と同様に、実際に使ってみないと判断できない、ということが大きいと考えられる。

不足している項目として、年度や作成年月日、移管年月日といった日付項目が多く挙げられた。また、文書種別や管理担当課・係を挙げる意見もあった。

【設問 4-1】メタデータ項目について、行政利用の観点から十分な項目となっているか。

1 すべてカバーされている	2
2 主要な項目についてはカバーされている	12
3 不足している項目がある	0
4 分からない	15
5 回答なし	0



今回のメタデータ項目については、すべてまたは主要な項目がカバーされている、との回答も多くみられたが、「分からない」という回答が半数以上であった。各担当者に直接関係のある記録管理メタデータ以外に、技術的メタデータやアーカイバルメタデータといったなじみの薄い項目が多いことが影響しているのではないかと考えられる。また、「一元的な文書管理システム」の項目との比較を望む意見も挙げられた。

【設問 5】平成 23 年度から電子公文書等の管理・移管・保存・利用システムの本格稼働が予定されているが、移管にあたってご関心のある点、重要と考える点

本設問に対しては、主に

- 改ざんなどのセキュリティ対策、大規模災害に対する対策はどうなるのか。
- 各府省でのシステム構築・改修の必要があるのか。
- 「一元的な文書管理システム」と連携する仕組みとなるのか。
- 機密性や原本性をどのように担保するのか。
- 移管された電子ファイルのマスキングにかかわる箇所の確定方法、確認方法等はどのようになるのか。

といった意見・質問が挙げられた。

多かったのは本システムが稼働した場合に各府省にどのような作業が発生する(どれぐらいの負担がある)のか、という点と、機密性や原本性の保証についての意見であった。

【設問 6】 その他意見、要望、指摘等

本設問に対しては、主に

- マスキングの徹底
- 標準フォーマットから長期保存フォーマットに変換する場合の真正性の定義の提示を求める。
- 各府省の担当者が具体的に何をしなければならないのかを教えてほしい。
- 行政文書ファイルとして管理されていないものは移管できないのか。

といった意見・質問が挙げられた。

ただし、実際に電子公文書の移管が開始されてみないと何ともいえない、という意見もあり、システムの稼働に向けては各府省に対しての十分な説明と理解を得ていくことが重要であると考えられる。

## 1.4. 委員会等

### 1.4.1. 委員会の目的

本調査の実施にあたり、電子公文書等の管理・移管・保存・利用システムに関する有識者を構成員とする委員会を設置し、以下の内容について検討していただいた。

- 調査方法及び計画
- 調査の進捗状況等
- 調査結果の妥当性
- 調査報告書の内容

### 1.4.2. 委員会の委員構成

委員等区分	氏名（敬称略）	所属等
委員長	杉本 重雄	筑波大学図書館情報メディア研究科知的コミュニティ基盤研究センター教授
委員	山田 洋	一橋大学大学院教授
	長谷川 英重	ISO/TC171 国内委員会委員長
	臼井 信昭	株式会社 PFU ニュービジネス推進統括部担当部長
	本田 実	内閣府 CIO 補佐官 城西国際大学 IT 教育センター教授
	平林 元明	内閣府 CIO 補佐官
内閣府関係者	福井 仁史	内閣官房公文書管理検討室参事官 内閣府大臣官房公文書等保存・利用推進室次長
	梅本 昌丈	内閣府大臣官房管理室 公文書館担当主査
独立行政法人国立 公文書館関係者	大賀 妙子	独立行政法人国立公文書館 業務課業務第二担当補佐
	中島 康比古	独立行政法人国立公文書館 業務課利用係長

### 1.4.3. 委員会開催状況

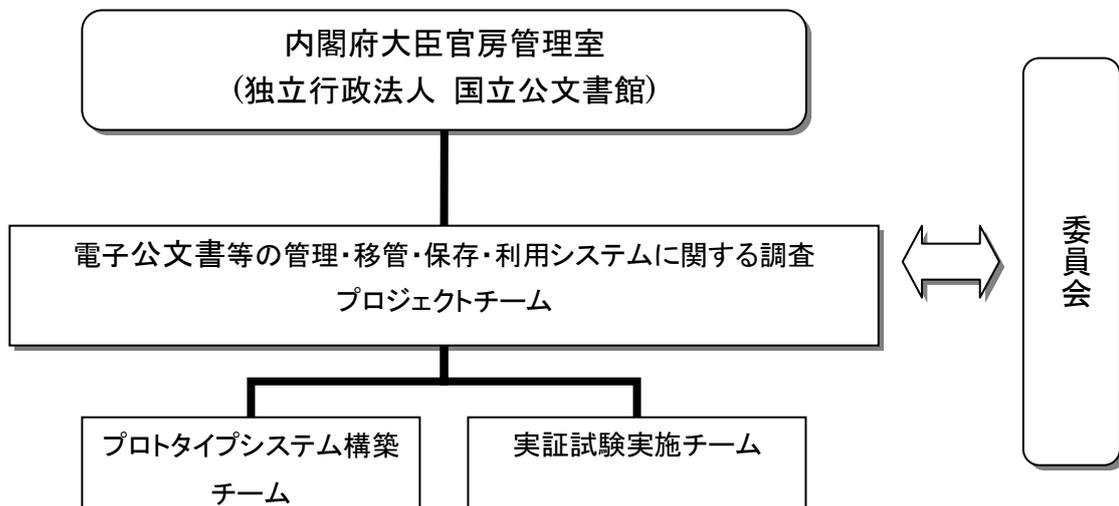
委員会は以下の期日で3回に分けて実施した。各委員会での主な議題は以下の通りである。

委員会	開催期日	主な議題
第1回	平成20年 12月8日	(1) 本調査業務の進め方について(調査概要・スケジュール) (2) 検証システム設計方針について (3) 検証システム構築方針について (4) 今後の委員会開催計画について

委員会	開催期日	主な議題
第2回	平成21年 2月23日	(1) 検証システムの構築結果の報告 (2) 調査結果の報告 (3) 各府省へのデモンストレーション計画について (4) 調査報告書概要について
第3回	平成21年 3月18日	(1) 各府省へのデモンストレーション結果の報告 (2) 調査報告書(案)の説明

## 1.5. 調査実施体制

本調査の実施体制は以下の通りである。



## 1.6. 調査実施スケジュール

項目		2008年	1月	2009年	
		12月 12月8日		2月 2月23日	3月 3月18日
委員会の実施					
検証システム開発	システム設計	←→			
	受入・保存管理機能		←→		
	行政利用機能		←→		
	デジタルアーカイブ連携機能		←→		
調査の実施	調査方針の検討・確定	←→			
	対象とする電子媒体の公文書のファイルフォーマット	←→			
	受入・保存管理機能		←→		
	行政利用機能		←→		
	デジタルアーカイブ連携機能			←→	
実験のための稼動環境作成		←→			
各府省へのデモンストレーション実施、意見聴取					←→
検証結果報告書の作成				←→	

## 2. ファイルフォーマット

### 2.1. 調査概要

今年度の調査における対象ファイルフォーマットについては、表 2-1 の種類とした。平成 19 年度までの調査研究の結果を踏まえつつ、各ファイルフォーマットの特徴、今後の動向および対応するアプリケーションを中心に調査を実施した。

PDF、PDF/A についてはどの分類に含めるか難しい部分もあるが、本調査では文書作成フォーマットの中に含めることとした。

表 2-1 調査対象ファイルフォーマット

分類	ファイルフォーマットの種類
文書作成	OASYS、一太郎 8-12、Word97-2003、Word2007、PDF、PDF/A、OpenOffice Writer
表計算	Excel97-2003、Excel2007、Lotus1-2-3、PDF、PDF/A、OpenOffice Calc
プレゼンテーション	PowerPoint97-2003、PowerPoint2007、PDF、PDF/A、OpenOffice Impress
画像	JPEG、JPEG 2000、GIF、TIFF、BMP
音声・音楽	AU、WAVE、MP3(MPEG Audio Layer-3)、WMA
映像	QuickTime、Windows Media、RealPlayer、MPEG

以降で各ファイルフォーマットについて調査した結果を記載する。

## 2.2. ファイルフォーマット調査報告

### 2.2.1. 文書作成フォーマット

本項では、文書作成フォーマットについて調査した結果をまとめる。なお、文書作成フォーマットに対する長期保存フォーマットとしては、PDF/Aとして出力して保存することが平成19年度までの調査結果において推奨されている。

#### (1) ファイルフォーマットの概要

文書作成フォーマットのファイルフォーマット概要として、「仕様策定関連情報」「特徴」「今後の動向」の3点について表2-2にまとめた。ここでの仕様策定年については、現在のフォーマットを扱うアプリケーションおよび機器が発売された年としている。

表 2-2 文書作成フォーマット:仕様策定関連情報

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
OASYS	1980年	OASYS 一号機である OASYS100 は、ワードプロセッサ販売初期の日本語ワードプロセッサ機器として発売された。 2009年3月時点でのアプリケーションの最新バージョンは「OASYS V10」であり、本調査(検証)も「OASYS V10」上で実施した。	富士通
一太郎	1983年	PC-100 用日本語ワードプロセッサとして「JS-WORD」という商品名で発売された。 2009年3月時点でのアプリケーションの最新バージョンは「一太郎 2009」であるが、本調査(検証)は「一太郎 2008」上で実施した。	ジャストシステム
Word97-2003 Word2007	1983年	Microsoft Office Word は、Multi Tool Word という名称で発売された。Microsoft 社の製品で、初めて GUI を採用した製品となっている。 2009年3月時点でのアプリケーションの最新バージョンは「Word 2007」であり、本調査(検証)は「Word 2003」及び「Word 2007」上で実施した。	Microsoft
PDF	1993年 (ISO 32000-1 としては 2008 年)	Portable Document Format の略称で、Adobe Systems が開発・提唱を始めた電子文書ファイルフォーマットである。 2009年3月時点での Adobe による最新バージョンは「1.7 Adobe Extension Level 3」である。 PDF1.7 をベースとして 2008年7月に ISO 32000-1	Adobe Systems ISO

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
		として標準化されている。	
PDF/A	2005 年	ISO によって、電子文書を長期保存するために制定された。PDF 1.4 をベースとして 2005 年 9 月に ISO 19005-1 として標準化された。	ISO
OpenOffice Writer	2002 年	Sun Microsystems が 1999 年に StarDivision 社(ドイツ)を買収し、StarOffice5.2 を無償公開するとともに、2000 年 10 月、StarOffice のソースコードを公開した。2005 年以降は GNU LGPL となり、オープンソース・コミュニティによる維持・管理が行われている。 バージョン 2.0 より、国際標準である OpenDocument Format (ODF) (ISO/IEC 26300)に標準対応している。 2009 年 3 月時点でのアプリケーションの最新バージョンは「OpenOffice Writer 3.01」であり、本調査(検証)も「OpenOffice Writer 3.01」上で実施した。	Sun Microsystems ISO

各フォーマットの特徴を、国際標準規格への対応、業界シェア、機能面などを中心に表 2-3 にまとめた。

表 2-3 文書作成フォーマット:特徴

フォーマット	フォーマットの特徴
OASYS	富士通が販売するワードプロセッサアプリケーションである。発売当初は、ワードプロセッサ専用機として販売されていた。しかし、パーソナルコンピュータ分野で、Microsoft Windows が圧倒的シェアを獲得している現在は、Windows をプラットフォームとするワードプロセッサアプリケーションとして販売されている。主な特徴としては、「独自開発キーボード(親指シフトキーボード)の採用」、「プロフェッショナル向け」という点が挙げられる。
一太郎	ジャストシステムが販売するワードプロセッサアプリケーションである。同じくジャストシステムの日本語入力フロントエンドプロセッサ(ATOK)とともに日本語のワードプロセッサとして、Microsoft Word が現在のシェアを獲得するまではデファクトスタンダード(事実上の業界標準)の地位を占めていた。一部の企業・官公庁においては現在も利用されている。 主な特徴としては、原稿用紙、縦書き、罫線の自由度、日本語文書の編集とレイアウトに特化している点が挙げられる。「一太郎 2006」からは、「一太郎 2006 OpenDocument 対応モジュール」により ODF フォーマットへの対応が可能となっている。
Word97-2003	Microsoft が販売するワードプロセッサアプリケーションである。Word 95 が Windows 95 とともに発売され、Windows の普及とともにシェアを拡大し、現在ではワードプロセッサ分野において圧倒的シェアを誇る。
Word2007	Microsoft が販売するワードプロセッサアプリケーションである。Windows Vista とともに発売され、Word シリーズの最新版となっている。 Office 2007 より、ファイルの内部保存構造が OOXML(Office Open XML)に変更された。

フォーマット	フォーマットの特徴
	<p>OOXML 形式は Word2003 以前の形式とは互換性がなく、Word2003 以前で作成された文書の処理は「互換モード」と呼ばれる機能により実現されている。</p> <p>OOXML は 2006 年 12 月に Ecma International により ECMA-376 として標準化され、また 2008 年 4 月には ISO/IEC 29500 として ISO/IEC により標準化が承認された。</p>
PDF	<p>PDF の特徴としては、「異なる OS 間でレイアウトを保持できる」、「セキュリティ設定」、「マルチメディア対応」などがある。Adobe Systems が仕様を公開しており、また閲覧用ソフトウェアである Acrobat Reader が無償で配布されていることから、デファクトスタンダードとなっていたが、2008 年 7 月に PDF 1.7 をベースとして ISO 規格 (ISO 32000-1) として標準化された。</p>
PDF/A	<p>ISO が制定している標準規格 (ISO 19005-1) であり、電子文書の長期保存を目的として策定された。主な特徴としては、「異なるアプリケーション間でのレイアウト保持」、「メタデータ保持フレームワーク」、「構造と意味を記録するフレームワーク」などが上げられる。PDF/A は、PDF 1.4 をベースとしながら、その中で定義される各種機能、オブジェクトについて必須機能、制限付き機能、使用禁止機能に分類して規定している。長期保存という観点から、フォントの埋め込みや色情報の再現性の保証が要求され、一方でファイル添付や外部参照など外部に依存する機能の使用が禁止されている。</p> <p>また、準拠レベルに「PDF/A-1a」、「PDF/A-1b」があり、それぞれ「完全準拠」、「一部準拠」としている。1a と 1b の大きな違いは、1a ではドキュメントの論理構造を表すタグを付与することが必須となっている点である。</p>
OpenOffice Writer	<p>オープンソースのフリーソフトウェアであるが、機能が豊富であり、様々なプラットフォーム上で動作する。バージョン 2.0 以降では国際標準である ODF (OpenDocument Format) に対応し、バージョン 3.0 以降では OOXML にも対応しているため、Microsoft Office 製品との高い互換性を持っている。</p> <p>世界的にシェアを伸ばしており、ベンダーに依存しないことから官公庁や自治体などでの採用も増えつつある。</p>

平成 19 年度までの調査結果における各ファイルフォーマットの PDF (PDF/A を含む) 形式での出力可否及び出力時の精度、さらに業界でのシェアや国際標準規格への対応状況を加味し、今後の動向を表 2-4 にまとめた。

表 2-4 文書作成フォーマット: 今後の動向

フォーマット	今後の動向
OASYS	<p>今後利用が進むと想定される標準文書フォーマットへの対応がされておらず、ワードプロセッサ機能への特化により、Microsoft Office や OpenOffice のような、いわゆる「オフィススイート」の持つ各種文書の連携の容易さ、操作性の統一といったメリットを上回る特徴を出すことが難しいと考えられる。将来的にみても普及状況は現在よりも厳しいものになると予想される。</p>
一太郎	<p>国際標準である ODF フォーマットに対応しているが、上記 OASYS と同様にワードプロセッサ以外の文書との連携が難しく、また ODF フォーマットの扱いは OpenOffice と競合するため、</p>

フォーマット	今後の動向
	将来的な普及拡大は厳しいものになると予想される。ただし、国内では一時期デファクト・スタンダードの位置を占めた経緯もあり、一太郎で作成された過去の文書が存在すること、また日本語文章作成に特化した機能を求めるユーザのニーズがある程度見込まれることから、日本国内における需要は今後もある程度残ると予想される。
Word97-2003 Word2007	現時点のワードプロセッサアプリケーション市場において大きなシェアを占めており、Microsoft Windows が今後もパーソナルコンピュータ用 OS における高いシェアを維持することが予想されることから、それとともにオフィススイートの一部として Word シリーズも当面は高いシェアを維持すると予想される。 Word2007 からは国際標準規格 (OOXML) に対応しているが、以前のバージョンとのファイル形式の互換性がなくなったことと、アプリケーションのバージョンアップが比較的短いサイクルで行われることから、今後も動向については注視していく必要がある。 また、Windows XP SP2 上の Word2003 で作成されたファイルを Windows Vista 上の Word2007 で表示する際に文字フォントの問題が発生する可能性が指摘されており、文書作成及び閲覧時の環境について注意が必要である。
PDF	プラットフォームを問わない可搬性、再現性を重視した設計により、情報交換の手段として広く普及している。2008 年 7 月より ISO で標準化されたこともあり、なお一層利用が促進されていくと考えられる。
PDF/A	元々電子文書の長期保存を目的として仕様が制定された経緯があるため、長期保存の観点からは望ましいフォーマットであると言える。 しかし、現在 ISO で PDF/A-2 の策定が進められており、ISO 32000-1 をベースとした仕様となる可能性もあり、今後の動向に注視していく必要があるといえる。
OpenOffice Writer	特定のベンダーに依存しないオープン性と国際標準規格 (ODF) への対応、また Microsoft Office 製品との互換性も高く、導入コストが低く抑えられることから今後普及が拡大していくことが予想される。 しかし、官公庁や企業においては、サポート体制が弱いこと、ユーザインタフェースが変わることによるユーザの再教育が必要となるため、普及に時間がかかる可能性もある。

## (2) 対応アプリケーション

それぞれのファイルフォーマットに対応する代表的なアプリケーションを閲覧、作成の観点から表 2-5 にリストアップした。なお、対応アプリケーションに含まれている場合でも、元フォーマットの作成アプリケーション以外で開く場合はレイアウトが崩れたり、機能が欠落したりする可能性があるため注意が必要である。

表 2-5 文書作成: 対応アプリケーション一覧

フォーマット	閲覧: 対応アプリケーション、バージョン	作成: 対応アプリケーション、バージョン
OASYS	OASYS V10	OASYS V10
一太郎	一太郎 2008	一太郎 2008
Word97-2003	Microsoft Word2003 OpenOffice Writer 3.01	Microsoft Word 2003 OpenOffice Writer 3.01

フォーマット	閲覧:対応アプリケーション、バージョン	作成:対応アプリケーション、バージョン
Word2007	Microsoft Word2007 Microsoft Word2003 *1 OpenOffice Writer 3.01	Microsoft Word 2007 Microsoft Word2003 *1 OpenOffice Writer 3.01
PDF	Acrobat Reader 9.0	Adobe Acrobat Professional Antenna House PDF Driver V4.0 Professional
PDF/A	Acrobat Reader 9.0	Adobe Acrobat Professional PDFlib 7 Antenna House PDF Driver V4.0 Professional ※ただし、Antenna House PDF Driver については、PDF/A-1b のみ対応
OpenOffice Writer	OpenOffice Writer 3.01	OpenOffice Writer 3.01

- ・ PDF 作成アプリケーションについては、ISO 32000-1 に準拠した PDF を閲覧・作成できるアプリケーションのみ対象とする。
- ・ 対応 OS は WindowsXP とする。
- ・ \*1 「Word/Excel/PowerPoint 2007 ファイル形式用 Microsoft Office 互換機能パック」をインストールすることで閲覧、編集が可能となる。

## 2.2.2. 表計算フォーマット

本項では、表計算フォーマットについて調査した結果をまとめる。なお、表計算フォーマットに対する長期保存フォーマットとしては、PDF/Aとして出力して保存することが平成19年度までの調査結果において推奨されている。

### (1) ファイルフォーマットの概要

表計算フォーマットのファイルフォーマット概要として、「仕様策定関連情報」「特徴」「今後の動向」の3点から表2-6にまとめた。ここでの仕様策定年については、フォーマットを扱うアプリケーションおよび機器が発売された年としている。

表 2-6 表計算フォーマット:仕様策定関連情報

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
Excel97-2003 Excel2007	1982年	表計算ソフト「Microsoft Multiplan」として発売され、グラフ作成ソフト「Microsoft Chart」と統合されて後のExcelとなっている。 2009年3月時点でのアプリケーションの最新バージョンは「Excel 2007」であり、本調査(検証)は「Excel 2003」及び「Excel 2007」上で実施した。	Microsoft
Lotus1-2-3	1983年	MS-DOS上で動作する表計算アプリケーションとして、発売された。MS-DOS全盛時には、表計算アプリケーション市場で大きなシェアを占めていた。 2009年3月時点でのアプリケーションの最新バージョンは「2001」となっている。	Lotus
OpenOffice Calc	2002年	OpenOffice Writerの内容を参照。 2009年3月時点でのアプリケーションの最新バージョンは「OpenOffice Calc 3.01」であり、本調査(検証)も「OpenOffice Calc 3.01」上で実施した。	Sun Microsystems ISO

各フォーマットの特徴を、国際標準規格への対応、業界シェア、機能面などを中心に表2-7にまとめた。

表 2-7 表計算フォーマット:特徴

フォーマット	フォーマットの特徴
Excel97-2003	Microsoftが販売する表計算アプリケーションである。表を作成し、合計の計算やグラフ作成等に利用される。数百種類ある関数を用いることで高度な計算も可能になっている。また、ワークシート上にオブジェクト(図形・画像)を貼り付け、設計書や進捗管理表に利用されることもある。マクロ言語を利用した簡易的なアプリケーションの作成等、純粋な表計算以外にも用途は多岐にわたる。現在、Microsoft Officeスイート製品には必ず含まれてい

フォーマット	フォーマットの特徴
	ることもあり、広く利用されている表計算ソフトと言える。
Excel2007	Microsoft が販売する表計算アプリケーションである。Windows Vista とともに発売され、Excel シリーズの最新版となっている。 また、Office 2007 より、ファイルの内部保存構造が OOXML(Office Open XML)に変更された。Word2007と同様に、OOXML形式は Excel2003 以前の形式とは互換性がない。
Lotus1-2-3	MS DOS時代の Lotus 社が販売していた代表的な表計算アプリケーションである。しかし、Microsoft Windows の普及とともにシェアを失い、2008 年現在、日本における販売は終了している。また、製造元の IBM から Windows Vista への対応を行わないことが発表されている。
OpenOffice Calc	OpenOffice Writer の内容を参照。

平成 19 年度までの調査結果における各ファイルフォーマットの PDF(PDF/A を含む)形式での出力可否、さらに業界でのシェアや国際標準規格への対応状況を加味し、今後の動向を表 2-8 にまとめた。

表 2-8 表計算フォーマット:今後の動向

フォーマット	今後の動向
Excel97-2003 Excel2007	表計算アプリケーションにおけるデファクトスタンダードであり、現状では OpenOffice を除けば有力な競合ソフトウェアが存在しないため、今後も高いシェアを維持すると予想される。また、Excel2007 ではサービスパック(SP)2を適用することにより PDF/A や ODF での出力に対応していることから、国際標準への対応という点でも評価できる。 ただし、Word2007と同様に、バージョン間の互換性の問題もあるため、今後の動向を注視する必要がある。
Lotus1-2-3	2008 年現在、日本での販売が終了しており、また 2001 年以降バージョンアップが行われていないため、今後の普及拡大は非常に厳しい状況にある。
OpenOffice Calc	OpenOffice Writer の内容を参照。

## (2) 対応アプリケーション

それぞれのファイルフォーマットに対応する代表的なアプリケーションを閲覧、作成の観点から表 2-9 にリストアップした。なお、対応アプリケーションに含まれている場合でも、元フォーマットの作成アプリケーション以外で開く場合はレイアウトが崩れたり、機能が欠落したりする必要があるため注意が必要である。

表 2-9 表計算:対応アプリケーション一覧

フォーマット	閲覧:対応アプリケーション、バージョン	作成:対応アプリケーション、バージョン
Excel97-2003	Microsoft Excel 2003	Microsoft Excel 2003
	Microsoft Excel 2007	Microsoft Excel 2007
	OpenOffice Calc 3.01*2	OpenOffice Calc 3.01*2
Excel2007	Microsoft Excel 2003*1	Microsoft Excel 2003 *1
	Microsoft Excel 2007	Microsoft Excel 2007

フォーマット	閲覧:対応アプリケーション、バージョン	作成:対応アプリケーション、バージョン
	OpenOffice Calc 3.01*2	OpenOffice Calc 3.01*2
Lotus1-2-3	Lotus1-2-3 2001 Microsoft Excel 2003	Lotus1-2-3 2001 Microsoft Excel 2003
OpenOffice Calc	OpenOffice Calc 3.01	OpenOffice Calc 3.01

対応 OS は Windows XP とする。

- ・ \*1 「Word/Excel/PowerPoint 2007 ファイル形式用 Microsoft Office 互換機能パック」をインストールすることで閲覧、編集が可能となる。

## 2.2.3. プレゼンテーションフォーマット

本項では、プレゼンテーションフォーマットについて調査した結果をまとめる。なお、プレゼンテーションフォーマットに対する長期保存フォーマットとしては、PDF/Aとして出力して保存することが平成19年度までの調査結果において推奨されている。

### (1) ファイルフォーマットの概要

プレゼンテーションフォーマットのファイルフォーマット概要として、「仕様策定関連情報」「特徴」「今後の動向」の3点から表2-10にまとめた。ここでの仕様策定年については、フォーマットを扱うアプリケーションおよび機器が発売された年としている。

表 2-10 プレゼンテーションフォーマット:仕様策定関連情報

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
PowerPoint97-2003 PowerPoint2007	1994年	もともとは、アウトラインプロセッサとして開発されていたが、表示機能が豊富になるにつれ、現在のPowerPointとして登場することになった。 2009年3月時点でのアプリケーションの最新バージョンは「PowerPoint 2007」であり、本調査(検証)は「PowerPoint 2003」及び「PowerPoint 2007」上で実施した。	Microsoft
OpenOffice Impress	2002年	OpenOffice Writerの内容を参照。 2009年3月時点でのアプリケーションの最新バージョンは「OpenOffice Impress 3.01」であり、本調査(検証)も「OpenOffice Impress 3.01」上で実施した。	Sun Microsystems ISO

各フォーマットの特徴を、国際標準規格への対応、業界シェア、機能面などを中心に表2-11にまとめた。

表 2-11 プレゼンテーションフォーマット:特徴

フォーマット	フォーマットの特徴
PowerPoint97-2003	Microsoftが販売するプレゼンテーションアプリケーションである。現在は、パーソナルコンピュータ上でプレゼンテーション資料を作成するアプリケーションとして広く普及しており、デファクトスタンダードとなっている。主な特徴としては、「スクリーン映写用資料作成が容易」、「文字・図・絵・写真・動画・音声等の貼り付け・微調整が容易」等が挙げられる。

フォーマット	フォーマットの特徴
PowerPoint2007	Microsoft が販売するプレゼンテーションアプリケーションである。Windows Vista とともに発売され、PowerPoint シリーズの最新版となっている。ただし、Microsoft Office Personal Edition には含まれておらず、ビジネス用途のアプリケーションとして位置付けられている。 Office 2007 より、ファイルの内部保存構造が OOXML (Office Open XML) に変更された。Word2007 と同様に、OOXML 形式は PowerPoint2003 以前の形式とは互換性がない。
OpenOffice Impress	OpenOffice Writer の内容を参照。

業界でのシェアや国際標準規格への対応状況から、プレゼンテーションフォーマットとしては今後も PowerPoint が利用され、OpenOffice Impress の利用も増加していくと予想される。

表 2-12 プレゼンテーションフォーマット: 今後の動向

フォーマット	今後の動向
PowerPoint97-2003 PowerPoint2007	プレゼンテーション資料作成で利用するアプリケーションにおけるデファクトスタンダードであり、現状では OpenOffice を除けば有力な競合ソフトウェアが存在しないため、今後も高いシェアを維持すると予想される。また、PowerPoint2007 ではサービスパック (SP)2 を適用することにより PDF/A や ODF での出力に対応していることから、国際標準への対応という点でも評価できる。 ただし、Word2007 と同様に、バージョン間の互換性の問題もあるため、今後の動向を注視する必要がある。
OpenOffice Impress	OpenOffice Writer の内容を参照。

## (2) 対応アプリケーション

それぞれのファイルフォーマットに対応する代表的なアプリケーションを閲覧、作成の観点から表 2-13 にリストアップした。なお、対応アプリケーションに含まれている場合でも、元フォーマットの作成アプリケーション以外で開く場合はレイアウトが崩れたり、機能が欠落したりする必要があるため注意が必要である。

表 2-13 プレゼンテーション:対応アプリケーション一覧

フォーマット	閲覧:対応アプリケーション、バージョン	作成:対応アプリケーション、バージョン
PowerPoint97-2003	Microsoft PowerPoint 2003	Microsoft PowerPoint 2003
	Microsoft PowerPoint 2007	Microsoft PowerPoint 2007
	OpenOffice Impress 3.01	OpenOffice Impress 3.01
PowerPoint2007	Microsoft PowerPoint 2003*1	Microsoft PowerPoint 2003*1
	Microsoft PowerPoint 2007	Microsoft PowerPoint 2007
	OpenOffice Impress 3.01	OpenOffice Impress 3.01
OpenOffice Impress	OpenOffice Impress 3.01	OpenOffice Impress 3.01

対応 OS は Windows XP とする。

- ・ \*1 「Word/Excel/PowerPoint 2007 ファイル形式用 Microsoft Office 互換機能パック」をインストールすることで閲覧、編集が可能となる。

## 2.2.4. 画像フォーマット

本項では、画像フォーマットについて調査した結果をまとめる。なお、画像フォーマットについては、平成19年度までの調査結果より JPEG 若しくは JPEG 2000 を長期保存フォーマットとすることが推奨されている。

### (1) ファイルフォーマットの概要

画像フォーマットのファイルフォーマット概要として、「仕様策定関連情報」「特徴」「今後の動向」の3点から表 2-14 にまとめた。

表 2-14 画像フォーマット:仕様策定関連情報

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
JPEG	1994 年	カラー静止画像を扱う汎用的なアプリケーションへの適用を目的として開発された。JPEG 自体はファイルフォーマットではなく、画像圧縮方式を指すものである。 「JPEG」とは、仕様を策定した ISO/IEC JTC 1/SC 29/WG 1(Joint Photographic Experts Group)の略称である。	ISO/IEC JTC 1
JPEG 2000	2000 年	ISO/IEC によって、JPEG より画質と圧縮率を向上させることを目的として策定された。 画像圧縮方式と画像フォーマットの両方を指す。	ISO/IEC JTC 1
GIF	1987 年	CompuServe 社が、ネットワーク上での画像データ交換用フォーマットとして開発した。電話回線を使用したネットワークを想定し、圧縮アルゴリズムとして LZW を採用している。 Graphics Interchange Format が正式名称である。 GIF 規格には GIF87a(1987 年 6 月 15 日公開)と GIF89a(1990 年 7 月 30 日公開)の 2 種類あるが、2009 年 3 月時点では GIF89a が主に使われている。	CompuServe
TIFF	1986 年	Microsoft 社及び Aldus 社(現在は Adobe Systems に合併)によって開発された。 Tagged Image File Format が正式名称である。 2009 年 3 月時点で「Revision 6.0」と呼ばれるバージョンが主流となっている。	Aldus(Adobe Systems) Microsoft
BMP	1991 年	Microsoft と IBM が共同で OS 開発を行っていた時期に策定された画像ファイル形式である。Microsoft Windows 3.0 から OS の標準ファイルフォーマットのの一つとして採用されている。	Microsoft

各フォーマットの特徴を、国際標準規格への対応、業界シェア、機能面などを中心に表 2-15 にまとめた。

表 2-15 画像フォーマット:特徴

フォーマット	フォーマットの特徴
JPEG	<p>離散コサイン変換(DCT:Discrete Cosine Transform)を利用した非可逆画像圧縮である。符号化段階では、情報の一部が失われ、画像品質の劣化が発生する。1999 年に ISO/IEC 14495-1 で、劣化無しのロスレス JPEG が標準化されている。課題として、超低ビットレートでの急激な画像劣化、ベースライン以外の符号化の互換性が困難、CG 画像や文字・線画等の 2 値画像において圧縮性能が低い等がある。</p> <p>拡張子は、jpg、jpeg が使用される。</p>
JPEG 2000	<p>ISO で標準化された JPEG の後継であり、圧縮方式として離散ウェーブレット変換(DWT: Discrete Wavelet Transform)を利用している。JPEG と比較して、圧縮性能の向上、特定領域圧縮機能等の豊富な符号化機能、各フレーム独立に JPEG 2000 符号化する動画像符号化方式を標準とする等 JPEG での課題を含めて仕様に盛り込んでいる。一方で圧縮にかかる時間は JPEG に比較して長くなる。その他の機能としては、特定の領域の画像を向上させる技術や、著作権保護技術、端末やネットワーク状況によって適切なデータ量を転送する技術が採用されている。JPEG 2000 の規格は複数の Part に分かれている。</p> <p>拡張子は、jp2、j2c が使用される。</p>
GIF	<p>1ピクセル当たり色情報が8ビットのため、最大256色という制限があるが、データ量は少なく済む。多くの Web ブラウザで標準サポートされており、Web 上での普及度が高い。</p> <p>LZW 圧縮アルゴリズムの特許を取得していた UNISYS 社が特許使用についての有償ライセンス契約を要求するようになったため、画像編集用のフリーソフトでは敬遠される時期があったが、2003年6月20日(日本では2004年6月20日)を以ってLZW圧縮アルゴリズムに関する特許が失効したため、それ以降は再び Web 上で自由に使える画像フォーマットとして認識されるようになっている。</p> <p>拡張子は、gif が使用される。</p>
TIFF	<p>OS 機器間で画像ファイルを交換できるようにするために、画像ファイルの先頭にタグと呼ばれる識別子を付けて、画像データの属性を分かるようにした画像フォーマットである。画像データの符号化方式、解像度、圧縮方式等に関して自由度が高いことが特徴であるが、逆に言えば厳密なルールが存在しないため、同じ TIFF フォーマットでも完全互換にはならないことがある。また、ひとつのファイルの中に複数の画像を格納したマルチページファイルを構成できるという特徴がある。</p> <p>拡張子は、tiff、tif が使用される。</p>
BMP	<p>Windows での標準画像形式であり、無圧縮のため、ファイルサイズは画像サイズに比例して大きくなる。</p> <p>拡張子は、bmp が使用される。</p>

業界シェア及び標準規格の観点から今後の画像フォーマットに関する動向を表 2-16 にまとめた。

注目すべき点としては、JPEG 2000 の今後の動向である。JPEG の後継として標準規格化されているが、対応しているアプリケーションの少なさや JPEG との互換性の問題など、普及に向けた課題が多い。

表 2-16 画像フォーマット:今後の動向

フォーマット	今後の動向
JPEG	JPEG の利用される範囲(デジタルカメラ、インターネット等)は多岐にわたる。「ISO 規格」、「高圧縮率」、「高画質」の 3 つの特徴に加え、多くのアプリケーションでサポートされていることもあり、今後も画像フォーマットの主流として利用されることが予想される。
JPEG 2000	2008 年現在、JPEG 2000 は国際標準となっているものの、一般利用者レベルでの普及はあまり進んでいない状況にある。OS 及び Web ブラウザでトップシェアを持つ Microsoft が、JPEG 2000 を標準的にサポートしていないこと、近年のストレージの急激な価格低下により圧縮保存の重要性が低下していること、計算コストの割にはサイズがあまり小さくならないこと等が、普及を拒む要因として考えられる。 Microsoft が対抗規格である HD Photo を開発しており、その動向如何によっては JPEG 2000 の実用レベルでの普及は厳しいものになる恐れがある。
GIF	LZW 圧縮の特許に関する問題がクリアされたため、今後もインターネット上で利用されていくものと予想される。さらに、アニメーション GIF は、他の画像フォーマットにはない機能であり、有用性があると判断できる。 ただし、近年のネットワークの高速化に伴い、ファイルサイズの圧縮自体がそれほど問題ではなくなってきたこと、フォーマットの仕様としては今後拡張される可能性が低いことから、Web 上でのボタンやバナーといった用途に限定されていくと考えられる。
TIFF	現在でも汎用的な画像データ交換用ファイルフォーマットとして使用されており、圧縮アルゴリズムを選ばない仕様であるため、今後も利用されることが予想される。 TIFF(version 6.0)で導入された JPEG 圧縮の仕様に不具合があり、最新版では解決されているものの、JPEG 圧縮については独自のエンコード/デコードを行うアプリケーションもあるため(アプリケーション依存がある)、JPEG 圧縮を利用する TIFF 形式を扱う際は、アプリケーションの選択に十分留意する必要がある。
BMP	圧縮はされないものの、BMP を扱える変換アプリケーションは多いため、PC 上の画像を劣化させずに保存したい場合は有用である。パーソナルコンピュータを利用する上でのスクリーンショット採取時に利用されることが多く、Microsoft Windows がシェアを持つ限り、今後も利用されていくと予想される。

(2) 対応アプリケーション

それぞれのファイルフォーマットに対応するアプリケーションを閲覧、作成の観点から表 2-17 にリストアップした。

表 2-17 画像:対応アプリケーション一覧

フォーマット	閲覧:対応アプリケーション、バージョン	作成:対応アプリケーション、バージョン
JPEG	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32 Microsoft Internet Explorer 6/7	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32
JPEG 2000	Adobe Photoshop CS4 (プラグインが必要) *IrfanView32 *LizardTech Express View Plug-in	Adobe Photoshop CS4 (プラグインが必要) *IrfanView32
GIF	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32 Microsoft Internet Explorer 6/7	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32
TIFF	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32
BMP	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32 Microsoft Internet Explorer 6/7	Adobe Photoshop CS4 / Elements 6 Microsoft ペイント *IrfanView32

\*はフリーウェアまたはフリー版を指す。

Microsoft Internet Explorer についてはプラグイン等を利用せずに閲覧できるものとした。

対応 OS は Windows XP とする。

Internet Explorer 6 を利用して、画像を表示した例を以下に示す。プラグインなどをインストールしない状態では、JPEG 2000、TIFF フォーマットの画像を表示することができない。JPEG 2000 については、閲覧のためのプラグインを別途インストールする必要がある。BMP は Web サーバから配信されている場合はプラグインなしで表示されるが、ローカルディスク上のファイルを直接 Internet Explorer で表示することはできない。



図 2-1 Internet Explorer で各画像フォーマットを表示した例

## 2.2.5. 音声・音楽フォーマット

本項では、音声・音楽フォーマットについて調査した結果をまとめる。なお、音声・音楽フォーマットについては、平成 19 年度までの調査結果よりビットレート 256kbps 以上の MP3 で音声ファイルをアーカイブすることが推奨されている。

### (1) ファイルフォーマットの概要

音声・音楽フォーマットのファイルフォーマット概要として、「仕様策定関連情報」「特徴」「今後の動向」の 3 点から表 2-18 にまとめた。

表 2-18 音声・音楽フォーマット:仕様策定関連情報

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
AU	不明	Sun Microsystems が策定。UNIX では標準的な音声フォーマットとされていた。	Sun Microsystems
WAVE	1991 年	WAV とも呼ばれる。 RIFF(Resource Interchange File Format)と呼ばれる、タグ付きのデータを格納するための汎用メタフォーマットの一つであり、データ形式は任意の形式を使用できるようになっている。	Microsoft IBM
MP3	1992 年	MPEG Audio Layer-3 の略称。動画の標準規格である MPEG-1 の音声規格(圧縮方式)の 1 つとして策定された(ISO 11172-3)。	ISO/IEC JTC 1
WMA	1999 年	Windows Media Audio の略称。 Windows Media の音声規格として Microsoft 社が開発した。 2009 年 3 月時点での最新バージョンは WMA9.2、WMA9.2 Lossless、WMA10 Pro となっている。	Microsoft

各フォーマットの特徴を、国際標準規格への対応、業界シェア、機能面などを中心に表 2-19 にまとめた。

表 2-19 音声・音楽フォーマット:特徴

フォーマット	フォーマットの特徴
AU	Sun Microsystems 社が策定した非圧縮音声ファイルフォーマットであり、現在は、UNIX 系 OS での標準音声ファイルフォーマットとなっている。 拡張子は au が使用される。
WAVE	Microsoft 社と IBM 社が共同で開発したフォーマット。基本的には非圧縮音声フォーマットであ

フォーマット	フォーマットの特徴
	り、任意のサンプリング周波数とビットレートのデータを格納できる柔軟性が特徴である。非圧縮で録音するため、録音対象が複雑な音楽でも全くの静寂であっても、単位時間当たりと同じ量のビットを記録するが、ADPCM方式による圧縮にも対応している。Windowsでは標準の音声フォーマットとされており、「サウンドレコーダー」アプリケーションで再生・録音が可能である(ただし、Windows Vista以降ではWMA形式がサウンドレコーダーの標準形式とされている)。拡張子はwav、waveが使用される。
MP3	高圧縮率にも関わらず音質の劣化が少ないため、音楽CDからPCへの取り込みでよく使用される。対応アプリケーションも豊富であるため、携帯用音楽プレイヤーの普及とともに、現在最も普及している圧縮音声ファイルフォーマットと言える。拡張子はmp3が使用される。
WMA	WMA9.1以降では、可変ビットレート圧縮や可逆圧縮に対応している。ストリーミング配信、音楽配信分野では、ACCに並ぶファイルフォーマットとして利用されている。WMAをサポートするデバイスは多いが、ファイル形式はMicrosoft独自形式である。拡張子は、wmaが使用される。

平成19年度までの調査結果、業界シェア及び標準規格の観点から今後の音声・音楽フォーマットに関する動向を表2-20にまとめた。

表 2-20 音声・音楽フォーマット:今後の動向

フォーマット	今後の動向
AU	非圧縮音声ファイルフォーマットかつ対応しているアプリケーションも少ないことから、今後の普及の可能性は低いと予想される。
WAVE	非圧縮音声ファイルフォーマットなので、原音に近い保存を用途とする場合は有用である。さらに、Windowsでの標準音声フォーマットであり、他の音楽ファイルフォーマットに変換するアプリケーションも豊富なため、今後も引き続き利用されていくものと予想されるが、他の圧縮可能な音声フォーマットのシェアを崩すまでには至らないと考えられる。
MP3	圧縮効率、圧縮・再生アプリケーションの対応状況、国際標準であることから、今後も広く使用されていくことが予想される。
WMA	ストリーミング配信分野・音楽配信分野で広く利用されており、WindowsであればWindows Media Playerでの再生が可能であるため、今後も使用されていくと予想される。しかし、Microsoft独自の形式であるため、今後の普及動向を注視する必要がある。

## (2) 対応アプリケーション

それぞれのファイルフォーマットに対応する代表的なアプリケーションを、再生、作成・変換の観点から表2-21にリストアップした。作成・変換アプリケーションについては、長期保存フォーマットのMP3(ビットレート256kbps以上)に変換できることを前提としている。

表 2-21 音声・音楽ファイル:対応アプリケーション一覧

フォーマット(拡張子)	再生:対応アプリケーション	作成・変換:対応アプリケーション
AU	Windows Media Player 11 *Real Player 11 *QuickTime 7 Player	*BatchWOO! QuickTime 7 Pro
WAVE	Windows Media Player *Real Player 11 *QuickTime 7 Player iTunes 8	*BatchWOO! *Free MP3 WMA WAV Converter 2.0 QuickTime 7 Pro
MP3	Windows Media Player 11 *Real Player 11 *QuickTime 7 Player iTunes 8	*BatchWOO! *Free MP3 WMA WAV Converter 2.0
WMA	*Windows Media Player 11	*BatchWOO! *Free MP3 WMA WAV Converter 2.0

- ・ \*はフリーウェアまたはフリー版を指す。
- ・ 対応 OS は WindowsXP とする。

## 2.2.6. 映像フォーマット

本項では、映像フォーマットについて調査した結果をまとめる。なお、平成 19 年度までの調査結果では、映像フォーマットでの利用フォーマット、長期保存フォーマットともに MPEG-2 が推奨されている。

### (1) ファイルフォーマットの概要

映像フォーマットのファイルフォーマット概要として、「仕様策定関連情報」「特徴」「今後の動向」の 3 点から表 2-22 にまとめた。下記のフォーマットについては、映像のみでなく音声等も含めたマルチメディアコンテナフォーマットが中心となっている。

表 2-22 映像フォーマット:仕様策定関連情報

フォーマット	仕様策定年	成立までの経緯	仕様策定機関
QuickTime	1991 年	QuickTime は、Macintosh の System 7 に初めて搭載された。厳密には動画、音声、画像等のマルチメディアを扱うための技術の総称である。 2009 年 3 月時点での最新バージョンは 7.6 となっている。	Apple
Windows Media	1991 年	Windows 3.0 マルチメディア拡張版に、デジタル オーディオ、ビデオ機能を追加した。 音声・動画の複合形式として AVI(Audio Video Interleaving)と呼ばれるコンテナフォーマットを使用する。 2009 年 3 月時点での最新バージョンは WMV(Windows Media Video) 9 となっている。	Microsoft
Real Player	1995 年	ストリーミングを再生できるメディアプレイヤーとして登場した。 音声フォーマットである RealAudio と、動画フォーマットである RealVideo の 2 つを合わせた RealMedia と呼ばれる形式が中心となっている。 2009 年 3 月時点での最新バージョンは RealVideo、RealAudio とも 10 である。	Real Networks
MPEG-2	1995 年	ビデオ、オーディオの他、システム等についても規格が策定されている。様々なメディアでの利用を想定し、ISO/IEC 13818 として標準化された。	ISO/IEC JTC 1

各フォーマットの特徴を、標準形式、業界シェア、機能面などを中心に表 2-23 にまとめた。

表 2-23 映像フォーマット:特徴

フォーマット	フォーマットの特徴
QuickTime	Apple 社が開発するマルチメディア技術で、音楽・動画・画像・テキストデータ等を取り扱うことができる。QuickTime フォーマット、国際標準 MPEG-4 規格のファイルフォーマット、Adobe Flash 等多くのフォーマットをサポートしている。さらに、第三世代携帯電話等でも利用されている、3GPP および 3GPP2(両フォーマットともに MPEG-4 規格ベース)もサポートしている。 拡張子は、mov、qt が使用される。
Windows Media	Microsoft は AVI の後継として ASF(Advanced Systems Format)を開発し、Windows Media の標準ファイルフォーマットとしている。Windows Media Video はその動画コーデックとなっている。 Microsoft Windows において標準で再生環境がサポートされるため、PC 向けに広く普及している。低ビットレートで映像の破綻が少なく、ストリーミングに対応しているため、主にインターネットのネット配信時のストリーミングフォーマットとして利用されている。ただし、ASF コンテナ構造は Microsoft の特許技術である点に注意が必要である。 拡張子は、wmv、avi が使用される。
Real Player	Real Networks 社が開発したメディアプレイヤーであり、狭義の映像フォーマットとしては RealVideo となる。コンテナとしての RealMedia は独自規格であるが、RealVideo のほか 50 種類以上のメディアファイルに対応している。PC 向けだけでなく携帯電話でも利用されている。 拡張子は、rm、rmvb、ram が使用される。
MPEG-2	ビデオ、オーディオの他、システムとしての規格が決められている。特徴としては、復号化方式のみが規定されており、符号化方式は規定されていない点が挙げられる。圧縮効率は MPEG-1 と同程度だが、今後、様々な環境に適用できるよう「インターレース」「多重化」「色情報フォーマット」「圧縮効率の最適化」等について拡張されている。DVD-Video ファイルフォーマットでは映像として MPEG-2 規格が採用されている。 近年、MPEG-2 の後継規格として、H.264 と H.264/AVC が登場してきている。圧縮効率では、従来方式である MPEG-2 の 2 倍以上の圧縮効率を実現するといわれている。 拡張子は、mpg、mpeg が使用される。

業界シェア及び標準規格の観点から、今後の映像フォーマットに関する動向を表 2-24 にまとめた。

表 2-24 映像フォーマット:今後の動向

フォーマット	今後の動向
QuickTime	QuickTimeについては、幅広いフォーマットをサポートしていること、また ISO ベースメディアファイルフォーマット(ISO/IEC 14496-12)のベースとなるなど、オープン化が進んでいる状況にあることから、今後も引き続き利用されていくと予想される。

フォーマット	今後の動向
	しかし、PC(Windows)環境での再生にはアプリケーションのインストールが必要となるため、他の再生アプリケーションによる対応にも今後注意していく必要がある。
Windows Media	Windows で標準的に再生環境が用意されているため、PC 向けに広く普及している。 Windows Media Video 9 を SMPTE(米国映画テレビ技術者協会)が規格化し、VC-1 コーデックとなっているが、符号化方式は MPEG-4(Part 2)をベースとした仕様であり、競合する MPEG-4 AVC などの規格も含め、今後の動向に注意する必要がある。
Real Player	Real Player 自体は多くのフォーマットに対応しているものの、独自規格である RealMedia(RealAudio、RealVideo)形式が Windows で標準的にサポートされていないため、今後シェアを伸ばすことは難しいと予想される。 ただし、ストリーミング配信用のコンテンツとして今後もある程度の利用が見込まれる。
MPEG-2	国際標準規格であること、DVD-Video や Blu-Ray Disc における動画圧縮方式として採用されていること、高精細度(HD)にも対応した仕様となっていることから、今後も広く利用されていくと予想されるが、前述の H.264 や VC-1 といった圧縮率で優る規格と競合するため、今後の動向には注意が必要である。 DVD-Video、Blu-Ray Disc のパッケージからの変換では、チャプター、メニュー情報、字幕データ等が欠落してしまうため、長期保存の観点では注意が必要である。

## (2) 対応アプリケーション

それぞれのファイルフォーマットに対応する代表的なアプリケーションを再生、作成・変換の観点から表 2-25 にリストアップした。作成・変換については、長期保存フォーマットの MPEG-2 に変換できることを前提としている。

表 2-25 映像:対応アプリケーション一覧

フォーマット(拡張子)	再生:対応アプリケーション	作成・変換:対応アプリケーション
QuickTime	*QuickTime 7 Player	TMPGEnc 4.0 Xpress QuickTime 7 Pro
Windows Media	Windows Media Player 11	TMPGEnc 4.0 Xpress MediaCoder 0.6.2
Real Player	*Real Player 11	RealProducer Plus
MPEG-2	Windows Media Player 11 *Real Player 11	MediaCoder 0.6.2

- ・ \*はフリーウェアまたはフリー版を指す。
- ・ 対応 OS は WindowsXP とする。
- ・ 各アプリケーションともにプラグイン等を利用しない。

## 2.3. 結論

### 2.3.1. 各ファイルフォーマットのまとめ

ここでは、ファイルフォーマット調査における結論を述べる。

今回の調査においては、使用する「フォーマット」関連の用語の定義は以下の通りとする。

オリジナルフォーマット	電子媒体の公文書が移管された時点でのファイルフォーマット。 オリジナルフォーマットには標準フォーマットが含まれる。
標準フォーマット	電子媒体の公文書の管理・移管・保存・利用システムにおいて、 長期保存フォーマットへの変換の対象とする電子媒体の公文書の フォーマット。
長期保存フォーマット	電子媒体の公文書の管理・移管・保存・利用システムにおいて、 標準フォーマットから変換される長期保存のためのフォーマット。
抽出テキストデータ	標準フォーマットのうち、文書作成、表計算、プレゼンテーションのい ずれかに該当する電子媒体の公文書の電子ファイルから、テキスト データのみを抽出したデータ。

#### (1) 標準フォーマットの考え方

文書作成、表計算、プレゼンテーションの各フォーマットについては、調査結果により、現時点では今後も Microsoft Office シリーズが広く利用されていくと考えられる。ただし、近年 Microsoft Office と同等の機能を利用できる Open Office も普及しつつあり、両者を現時点での標準フォーマットとすることとした。また、OASYS、一太郎については、平成 19 年度までの調査結果において限定的ではあるものの依然として利用されていることから、標準フォーマットに含めている。PDF は移管時点でのフォーマットとして既に使われているケースも想定されるため、同様に標準フォーマットとして扱うこととする。PDF/A で移管されるケースは現時点ではほとんど無いと予想されるが、今後 PDF/A で移管される可能性もあるため、標準フォーマットに含めることとした。

画像、音声、映像フォーマットについては、普及度、国際規格、利用範囲の観点から、今後も利用される可能性が高いものを標準フォーマットとした。

#### (2) 長期保存フォーマットの考え方

文書作成、表計算、プレゼンテーションの各フォーマットについては、レイアウトやページ構成の保持及び見読性の要件を満たし、改ざんされにくいといったメリットを持ち、多様な文書作成アプリケーションからの変換が可能であること、閲覧アプリケーションが容易に入手できること等から、電子文書の長期保存を目的とした国際標準規格である PDF/A を採用する。

画像フォーマットについては、国際標準規格であり、圧縮率が高いことから、JPEG 2000 を採用する。ただし、現時点では普及率が低く、将来性に対する懸念があるため、JPEG XR などの競合する仕様も含めて、今後の動向に注意する必要がある。

音声フォーマットについては、圧縮効率、圧縮・再生アプリケーションが豊富であること、国際標準規格で

あることから、MP3を採用する。

映像フォーマットについては、普及度、再生アプリケーションが豊富であること、国際標準規格であることから、MPEG-2を採用する。ただしMPEG-2の後継規格であるH.264/AVC、VC-1等が今後普及していく可能性もあるため、今後の動向に注意する必要がある。

(3) 今回の調査における標準フォーマット及び長期保存フォーマット

上記(1)、(2)より、今回の調査(プロトタイプシステム)において、文書作成、表計算、プレゼンテーション、画像、音声・音楽、映像の各フォーマットに対応する標準フォーマット、長期保存フォーマットを表 2-26 に示す通りとした。

また、国立公文書館デジタルアーカイブ・システムで一般の利用に供するときに採用するフォーマットを「デジタルアーカイブ用フォーマット」として定義し、原則として長期保存フォーマットと同じフォーマットを採用することとした。ただし、現在のデジタルアーカイブ・システムでは音声・音楽、動画は公開されていないため、「デジタルアーカイブ用フォーマット」からは除外した。

表 2-26 標準フォーマット、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一覧

フォーマット	標準フォーマット	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 2007 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	JPEG 2000	JPEG 2000
音声・音楽	WAVE MP3 WMA	MP3 (ビットレート 256kpbs 以上)	-
映像	QuickTime	MPEG-2	-

フォーマット	標準フォーマット	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
	Windows Media RealPlayer MPEG		

### 2.3.2. PDF/A の今後の動向

現在、ISO 19005-2(PDF/A-2)という PDF/A-1 の後継規格の策定が進行中である。PDF/A-2 は、PDF1.7 仕様に基づくとされているが、最終的には ISO 32000 ベースになる可能性もある。PDF (ISO 32000) のマルチメディア化が促進されることによって、将来的に文書作成、表計算、プレゼンテーション、画像、音声・音楽、動画すべてにおいて長期保存フォーマットを PDF/A に統一することも可能性としては有り得る。

長期保存フォーマットを全て同一規格かつ国際標準規格に統一することで、長期保存の見通し期間が現状よりも一層長く見込めることになるという利点もあるが、PDF/A-1 と PDF/A-2 の互換性によっては、マイグレーションにおける問題が発生する可能性も考えられるため、今後も PDF/A 関連の仕様の動向に注意が必要である。



## 3. メタデータスキーマ

### 3.1. 調査概要

本項では、電子公文書等の受入・保存管理に必要なメタデータについて、まず国際標準であるスキーマやモデルについての調査を行い、その結果を踏まえてプロトタイプシステムで使用するメタデータスキーマの定義を行った。

### 3.2. 各種メタデータ標準等の調査

本項では、電子情報の長期保存に有効なフレームワークである OAIS 参照モデルのほか、現時点で国際的なメタデータ標準である EAD、METS、MODS、PREMIS のそれぞれについて調査を行い、電子公文書の長期保存のためのメタデータスキーマを検討する材料とした。以下に調査結果を示す。

### 3.2.1. OAIS 参照モデル

OAIS 参照モデルの基本情報については、下表の通りとなる。OAIS は参照モデルとして定義されているが、文書の長期保存のための有効なフレームワークであるため調査対象とした。

表 3-1 OAIS 参照モデルの基本情報

正式名称	Reference Model for an Open Archival Information System (OAIS 参照モデル)
仕様策定機関	宇宙データシステム諮問委員会 (CCSDS)
概要	宇宙観測データのみならず、情報という概念を視野に入れ、アーカイブ責任、機能詳細、保存方策およびアーカイブ連携を扱う総合的な内容として定義された。1999 年 5 月にドラフト版 (Red Book, Issue 1) が、2001 年 7 月に改訂版 (Red Book, Issue 2) が出され、2002 年 1 月に正式勧告版 (Blue Book, Issue 1) が採択され、国際標準規格 (ISO 14721:2002) として認定された経緯を持つ。
参照 URL	<a href="http://nssdc.gsfc.nasa.gov/nost/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf">http://nssdc.gsfc.nasa.gov/nost/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf</a>

OAIS 参照モデルは、情報パッケージ (Packaging Information) という「保存記述情報 (Preservation Description Information)」、「内容情報 (Content Information)」をまとめた形式をとる (図 3-1 参照)。

OAIS 参照モデルでいう内容情報とは、文字通り内容そのものを表現する情報を指す。また、内容情報は、内容データオブジェクトとも呼ばれ、保存の観点からすると、内容情報こそが保存対象となるオリジナルデータとされる。情報パッケージ、保存記述情報は、内容情報に対するメタデータという扱いとなる。

保存記述情報とは、内容情報の管理・理解を手助けするサポート的位置づけとなる。ここでは、言及しないが、保存記述情報には、「参照情報」、「コンテキスト情報」、「不変性情報」、「来歴情報」の 4 種類に分類される。

保存データの探索情報 (パッケージの説明、アーカイブ内での位置や内容特定情報) となる記述情報 (Descriptive Information) については、この情報パッケージに含まれていない (図 3-1 参照)。

また、OAIS 参照モデルの情報パッケージは、処理段階に応じ「提出 (受入) 用情報パッケージ Submission Information Package (SIP)」「保管用情報パッケージ Archival Information Package (AIP)」「配布用情報パッケージ Dissemination Information Package (DIP)」の 3 種類が規定されている (図 3-2 参照)。

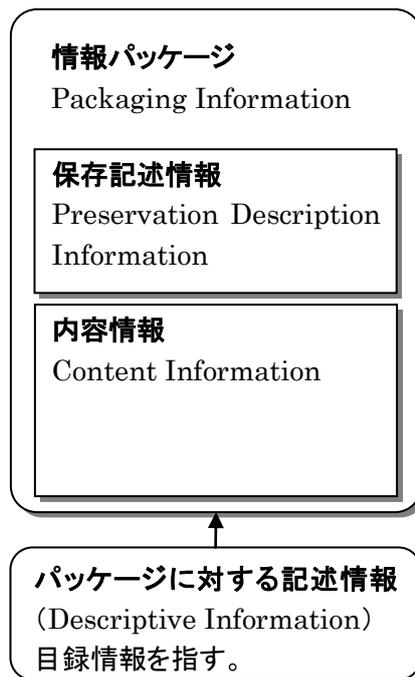
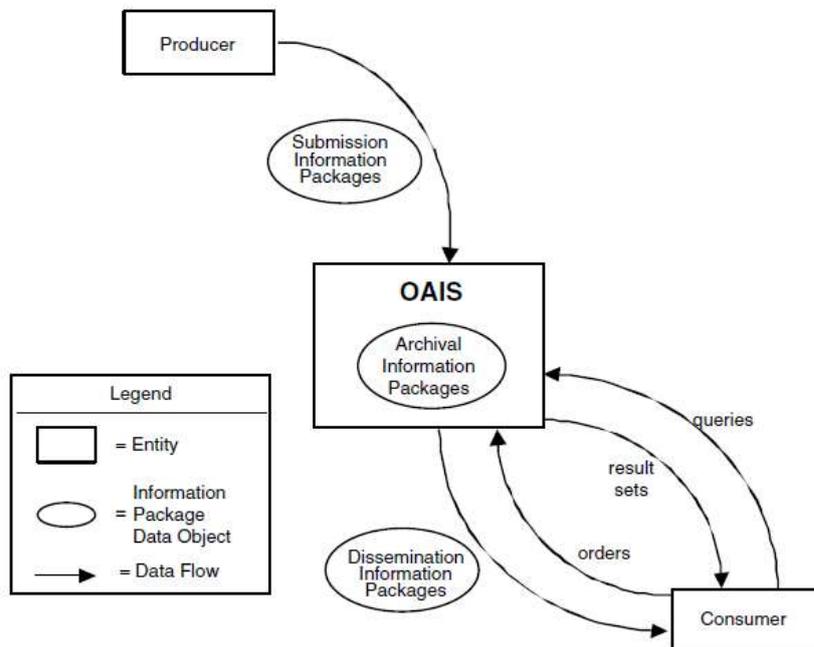


図 3-1 OAIS 参照モデルの情報パッケージ



(Reference Model for an Open Archival Information System (OAIS) Blue Book, Issue 1, Page 2-8 より)

図 3-2 OAIS における外部データフロー

こうした情報モデルに基づき、いくつかのデジタル情報保存プロジェクトにおいて、具体的なメタデータの項目や表記法が検討されてきている。代表的な例としては、METS (Metadata Encoding and Transmission Standard) が該当する。METS スキーマについては、本報告書内でも言及しているため、詳細は「3.2.3 METS」を参照のこと。

※OAIS 参照モデルについての参考文献、Web サイトについては、ページ下部の注釈を参照のこと。

2)3)4)5)

- 
- 2) 杉本重雄, “電子図書館概説” (図書館職員長期研修資料),  
<http://avalon.slis.tsukuba.ac.jp/~sugimoto/Articles/Chouken2003.pdf>
  - 3) 杉本重雄, “デジタルコンテンツのアーカイブとメタデータ”, 人工知能学会誌, Vol.18, No.3, pp.217-223, 2003
  - 4) 栗山 正光, “長期保存型電子図書館と OAIS 参照モデル”,  
[http://www.tokiwa.ac.jp/~mtkuri/presentations/preserv\\_dl\\_oaisrm.ppt](http://www.tokiwa.ac.jp/~mtkuri/presentations/preserv_dl_oaisrm.ppt)
  - 5) 国立国会図書館, “電子情報の長期保存とアクセス手段の確保のための調査報告書”,  
[http://www.ndl.go.jp/jp/aboutus/report\\_2004.pdf](http://www.ndl.go.jp/jp/aboutus/report_2004.pdf)

### 3.2.2. EAD

EAD の基本情報については、下表の通りとなる。

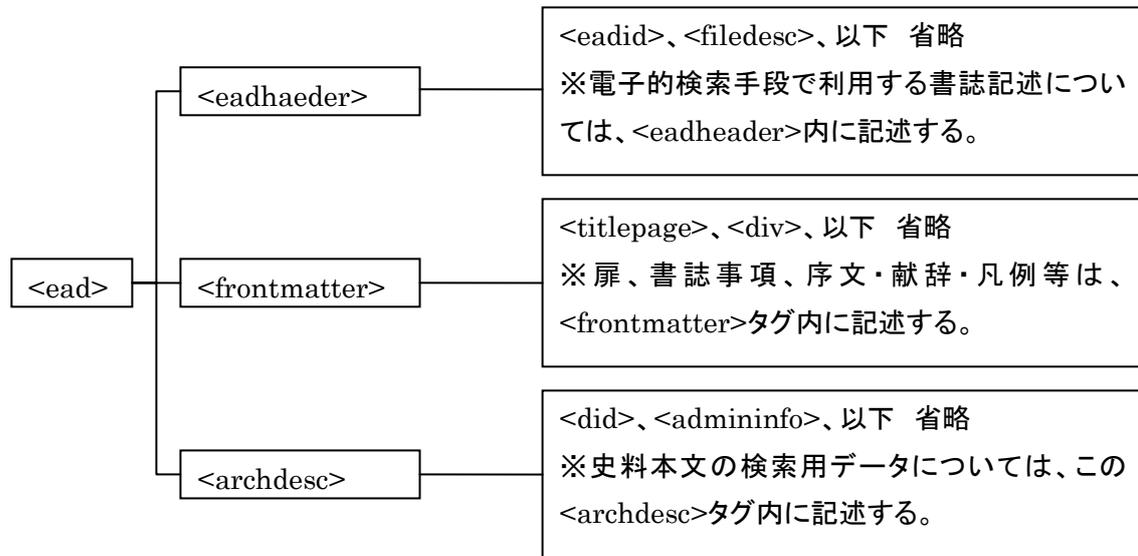
表 3-2 EAD の基本情報

正式名称	Encoded Archival Description (EAD)
仕様策定機関	カリフォルニア大学バークレー校 (University of California, Berkeley) アメリカアーキビスト協会 (SAA)
概要	EAD は、カリフォルニア大学バークレー校の図書館で発足したプロジェクトが発端となっている。アーカイブズ検索手段を電子的に符号化するための国際標準である。 1998 年に EAD DTD Version 1.0(1998)がリリースされたが、項目の見直しと XML に特化する形で EAD DTD Version 2002 がリリースされており、2009 年 3 月時点での Version 2002 が最新となっている。Version 2002 は XML Schema 形式でも公開されている(下記参照 URL)。
参照 URL	<a href="http://www.loc.gov/ead/">http://www.loc.gov/ead/</a> <a href="http://www.loc.gov/ead/ead.xsd">http://www.loc.gov/ead/ead.xsd</a> (EAD 2002 の XML Schema)

EAD (Encoded Archival Description) は、記録史料を記述するために必要な要素は何かという意味内容を具体的に構成するための標準規格である。EAD とは、アーカイブズを検索し、かつ記述・作成するための DTD である。元来独自の規格だったが、ISAD(G)との互換性を検討したうえで整備されていった。EAD の DTD (バージョン 2002) は、ISAD(G)の記述を含む形となっているため、階層構造を維持した記述とコンピュータでの検索機能を実現している。ISAD(G)との互換性があることから記録史料等に最も適したメタデータとして適用されるようになってきたが、適用例については、多いとは言い難い現状がある。その主たる原因としては、各収蔵機関の複雑かつ伝統的な目録管理手法が挙げられる。一方日本の現状に対して、欧米では目録の規格化(国単位での統一)が進んでおり、欧米の多くの記録資料保存機関で EAD が、広く採用されていることから、事実上の標準となっている。EAD の利点・特徴としては、以下のものが挙げられる。

- 電子的検索手段として、マルチレベル記述の表現が可能
- 電子記録・電子書類として扱うための枠組みが用意されている。

記録史料メタデータを記述するうえで、EAD 基本構成タグは、大きく分けて 3 つに分類される。



EAD スキーマに関する詳細は、「EAD Official Site」を参照のこと。また、EAD の代表的なタグの解説及び ISAD(G)との対応表については、脚注を参照のこと。

表 3-3 EAD の適用事例

<p>国立公文書館 デジタルアーカイブ・システム</p>	<p>国立公文書館 デジタルアーカイブ・システムでは、移管された公文書や古典籍史料の目録情報を EAD で管理している。一般に公開されている検索システムでは、EAD の階層情報を資料群に対応させることで、階層的に目録情報の検索を行えるようになっている。</p> <p>目録情報と EAD のマッピング定義が Web サイト(<a href="http://www.digital.archives.go.jp/howto/pdf/naj_ead107.pdf">http://www.digital.archives.go.jp/howto/pdf/naj_ead107.pdf</a>)上で公開されている。</p>
<p>アジア歴史資料センター</p>	<p>アジア歴史資料センターは、国の機関が保管するアジア歴史資料をインターネット上で提供する電子資料センターとして、国立公文書館により運営されている。アジア歴史資料センターにおいても国立公文書館と同様に EAD による目録情報の管理が行われており、検索システムにおいても階層的な検索が実現されている。</p>
<p>国立国語研究所</p>	<p>国立国語研究所では、「資料情報検索システム」という EAD XML を利用した検索システムを公開している。EAD 要素 (element) の中から、ISAD(G)の記述要素と対応するものを記述項目としている。</p> <p>「入力システム」「EAD ファイル自動生成プログラム」「検索システム」から構成される。管理者が入力したデータから、EAD ファイルを自動生成し、検索システムで生成されたファイルを利用する仕組みになっている。</p>

※EAD についての参考文献・サイトについては、ページ下部の注釈を参照のこと。<sup>6)7)8)9)10</sup>

6) EAD Official Site,<http://www.loc.gov/ead/>

7) 五島敏芳,“EAD の概要と日本における動向:国文学研究資料館の事例紹介を中心に”,  
[http://archives.nijl.ac.jp/DAS/projects/eadfa/20051020\\_gotoh.htm](http://archives.nijl.ac.jp/DAS/projects/eadfa/20051020_gotoh.htm)

8) 五島敏芳,“検索手段電子国際規格 EAD/EAC と記録史料記述国際標準”,  
[http://archives.nijl.ac.jp/DAS/projects/eadfa/20051020\\_gotoh-ref.pdf](http://archives.nijl.ac.jp/DAS/projects/eadfa/20051020_gotoh-ref.pdf)

9) 国立国語研究所,“資料情報検索”, <http://ead.kokken.go.jp/ead/>

10) ISAD(G)については、<http://www.ica.org/en/node/30000> を参照のこと。

### 3.2.3. METS

METS の基本情報については、下表の通りとなる。

表 3-4 METS の基本情報

正式名称	Metadata Encoding and Transmission Standard (METS)
仕様策定機関	電子図書館連合 (Digital Library Federation: DLF)
概要	METS は Encoded Archival Description(EAD)などでの経験をもとに、Digital Library Federation を中心に作られたデジタルコンテンツのためのメタデータ記述の標準であり、デジタル化資料のアーカイブの方向性を示したMETSは、XML による記述形式を定義している。 最新版のスキーマとしては、Version 1.7(2008 年 12 月現在)が発表されている。
参照 URL	<a href="http://www.loc.gov/standards/mets/">http://www.loc.gov/standards/mets/</a>

METS は、OAIS 参照モデルに準拠しており、OAIS 参照モデルにおける情報パッケージ (SIP,AIP,DIP) の形態として適用可能である。METS の記述単位 (セクション) として以下の内容を含んでおり、XML データとして記述する形式が定義されている。

METS は、以下の 7 つ記述単位から構成される。各特徴は、下表の通りである。

表 3-5 METS における記述単位の特徴

METS ヘッダ <metsHdr>	METS ファイル (XML テキストデータ) 自体を記述し、内容の属性 (作者等) を含むセクションとなる。
記述メタデータセクション <dmdSec>	METS は、「記述メタデータ」「管理系メタデータ」に大きく分かれる。記述メタデータセクションには、検索などに利用される記述的メタデータとなり、外部の記述メタデータへの参照または、記述の埋め込みをする。
管理系メタデータセクション <amdSec>	ファイルの管理・記録・知的財産情報等を記述する。 OAIS 参照モデルでは、「保存記述情報 (Preservation Description Information)」に該当する。
ファイルセクション <fileSec>	デジタルオブジェクトを構成するすべてのファイルをリストアップする。または、同一内容を、異なる方式によって作成したファイルのリストとなる。
構造マップ <structMap>	デジタルオブジェクトの階層構造を記述する。ファイルセクションのファイル要素 <div> が、コンテンツの構成部分を表現する。<div> の入れ子には、制限がないため複雑な階層を表現できる。
構造リンク <structLink>	構造マップで表現されたコンテンツ同士をハイパーリンクで表現する。
動作セクション <behaviorSec>	METS の XML 文書と実行可能な動作との関連付けを記述する。

METS の要素の ID 属性と参照属性を利用し、メタデータとコンテンツが関連付けされる。METS の特徴

としては、構造マップを利用することでデジタルオブジェクト間の階層情報を記述することが可能となっている点が挙げられる。

表 3-6 METS の適用事例

国立国会図書館	NDL-DA システムは、OAIS 参照モデルに準拠しており、OAIS 参照モデルの「情報パッケージ」を METS で定義・構築している。
---------	---

なお、海外での事例は METS の Web サイト(<http://www.loc.gov/standards/mets/mets-registry.html>)を参照されたい。

※METS についての参考文献・サイトについては、ページ下部の注釈を参照のこと。<sup>11)12)</sup>

---

11) METS Official Web Site , <http://www.loc.gov/standards/mets/>

12) 国立国会図書館, “NDL デジタルアーカイブシステム メタデータスキーマガイドライン”, <http://www.ndl.go.jp/jp/standards/da/da.pdf>

### 3.2.4. MODS

MODS の基本情報については、下表の通りとなる。

表 3-7 MODS の基本情報

正式名称	The Metadata Object Description Schema (MODS)
仕様策定機関	米国議会図書館ネットワーク開発 MARC 標準オフィス (Library of Congress Network Development and MARC Standard Office 略称、LC-NDMSO)
概要	米国議会図書館 (Library of Congress、以下 LC) が、MARC21 から抽出されたサブセットの記述エレメントを有する XML スキーマを目指したのが始まりである。 MODS は、Dublin Core のような非常に簡易な形式と MARC21 のような多数の記述エレメントと複雑な構造を持つ詳細な形式まで対応できる XML スキーマである。MARC21 に対応した記述エレメントから構成されており、同じ意味構成を継承しているため、MARC レコードとの互換性が非常に高い。 最新版のスキーマとしては、Version 3.3 (2008 年 12 月現在) が発表されている。
参照 URL	<a href="http://www.loc.gov/standards/mods/">http://www.loc.gov/standards/mods/</a>

MODS の最大の特徴としては、MARC21 の書誌フォーマットに対応しており、MODS スキーマに MARC21 のどのタグに対応しているかが記述してあることから、MARC レコードとの親和性が高いと言える。MODS の利点・特徴としては、以下のものが挙げられる。

- Dublin Core、出版 MARC として開発された ONIX スキーマよりも豊かな表現が可能
- MARC 形式より人間にとってわかりやすい記述が可能
- MARC の意味を継承
- 最小限のフィールドかつ階層構造も持つ

今後 MODS は、さまざまな利用展望がある。Z39.50 の後継規格である SRW (Search/Retrieve Web Service) は、メタデータスキーマを特定した検索サービスであり、図書館のデータと互換性・親和性が高い MODS のようなスキーマの利用が望まれている。さらに、LC では、MODS の導入を始めており、資料 (図書、地図資料、写真、初期のころの映画や楽譜、様々な印刷物がある) に対して、MODS を利用することで、豊かな表現を実現している。

MODS は、MARC21 のサブセットであるため、MARC21 からの移行の際には、情報の欠落が発生する可能性がある。しかし、LC はその情報欠落を防止するため、MARC21 から MARCXML スキーマを利用して完全レコードで XML 表現することを実現させた。現在では、MARC21 から MODS への変換を、[MARC21] → [MARCXML] → [MODS] の順に変換する MODS 変換ツールが LC より提供されている。

MODS は、公式サイトによると、以下の用途で将来的利用できる可能性があると言及している。(すでに可能になっているものも含む)。

- SRU/W がサポートしている XML スキーマ

- METS への拡張スキーマ
- ハーベスティング用のメタデータを表現する XML スキーマ
- MARCXML の MODS 表現
- 電子ファイルなどに、MODS で表現されたメタデータを埋め込む

MODS が適用されている例については、下表の通りである。

表 3-8 MODS の適用事例

視聴覚プロトタイププロジェクト	LC では「視聴覚プロトタイププロジェクト」という、音響・映像資料のデジタル保存に関するプロジェクトを進めていた(1999-2004年)。このプロジェクトでは、大半のメタデータがリレーショナルデータベースに蓄積されていく。その後のアーカイビング時に XML へと変換される。本プロジェクトでは、METS を利用したメタデータパッケージ化(記述メタデータとして MODS を利用)の実験も同時に実施している。LC にある既存データに対しては、項目の消失を最小限に抑えるために MARC から MODS への変換で対応した。
Minerva プロジェクト	このプロジェクトは、LC が他大学と協業し展開しているプロジェクトであり、ウェブアーカイブを目的としている。ウェブサイトのメタデータに MODS を利用している実績がある。
国立国会図書館	METS の適用事例で挙げたとおり、NDL-DA メタデータは OAIS 参照モデルに準拠している。「保存記述情報」の記述メタデータに該当する部分で MODS (version 3.2) を利用している。

※ MODS についての参考文献・サイトについては、ページ下部の注釈を参照のこと。<sup>13)14)15)</sup>

<sup>13)</sup> MODS Official Web Site , <http://www.loc.gov/standards/mods/>

<sup>14)</sup> 鹿島みづき,酒井由紀子,“メタデータ オブジェクト ディスクリプション スキーマ”,  
[http://www2.aasa.ac.jp/org/lib/j/netresource\\_j/guenther0306/3.1guenther\\_j.pdf](http://www2.aasa.ac.jp/org/lib/j/netresource_j/guenther0306/3.1guenther_j.pdf)

<sup>15)</sup> Allene Hayes,Rebecca Guenther , Leslie Myrick , “The 3M’s MINERVA ,MODS,and METS”,  
<http://www.diglib.org/forums/Spring2004/presentations/hayes-2004-04.pdf>

### 3.2.5. PREMIS

PREMIS の基本情報については、下表の通りとなる。

表 3-9 PREMIS の基本情報

正式名称	PREservation Metadata: Implementation Strategies (PREMIS)
仕様策定機関	Online Computer Library Center Research Library Group
概要	PREMIS は、2000 年に OCLG と RLG の合同ワーキンググループが組織され、OAIS 参照モデルやその他メタデータスキーマを参考にメタデータの枠組みとして構築された。 最新版のスキーマとしては、Version 2.0(2008 年 12 月現在)が発表されている。
参照 URL	<a href="http://www.loc.gov/standards/premis/schemas.html">http://www.loc.gov/standards/premis/schemas.html</a>

PREMIS には、実装可能な保存メタデータのコア要素を定義・記述している「PREMIS Data Dictionary version 2.0」(以下 データディクショナリー)が公開されている。データディクショナリーによると、PREMIS の適用範囲は、「生存の維持」「分かりやすさの維持」「信頼性の維持」「アイデンティティの維持」等のデジタルコンテンツアーカイブ活動に限定されている。

PREMIS は、OAIS 参照モデルから構築され、5 つの構成要素(エンティティ:Entity)を定義している。5 つの構成要素については、表 3-10 の通りである。

表 3-10 PREMIS の各 Entity の特徴

知的エンティティ (Intellectual Entity)	デジタルコンテンツ 1 つのユニットを構成する単位を指す。知的エンティティは、他の知的エンティティを含めることが可能である(つまり 1 つまたは複数のデジタル表現を保有できる。) 例) 画像(同じ内容の画像 JPEG TIFF) ※知的エンティティはデータモデルとして定義されているため、保存メタデータの対象外となる。
オブジェクト(Object)	デジタル形式の個々の情報単位である。 オブジェクト(Object)の下位階層では、以下のように分類される。 ファイル(file): コンピュータにファイルとして保存されたビット列を指す。 表現(representation): 表示に必要なファイル群を指す。 ビットストリーム(bitstream): 定義された開始位置と長さをもったファイル内のビット配列を指す。
イベント(Event)	オブジェクト(Object)、エージェント(Agent)に影響を与える動作または行事である。 例) 行政文書 PDF を、デジタルアーカイブ検索システムに登録する
権利(Rights)	オブジェクト(Object)、エージェント(Agent)に関する権利または許諾

	を指す。 例) 行政文書を長期保存するためのフォーマット変換許諾
エージェント (Agent)	イベント (Event) にかかわる個人・組織・またはソフトウェアを指す。 例) PDF ファイルより抽出したテキストデータに対して PREMIS で表現する場合は、エージェント (Agent) は、PDF からテキストデータを抽出するためのソフトウェアとなる。

また、PREMIS のメタデータ要素の実装は、事実上の標準として利用される METS でエンコードすることを承認されているため、今後は PREMIS を利用する機関などが増加していくことが十分に考えられる。

表 3-11 PREMIS の適用事例

国立国会図書館	NDL-DA メタデータでは、OAIS 参照モデルの「保存記述情報」内の「技術的メタデータ」「権利メタデータ」「保存メタデータ」に該当する部分で PREMIS を利用している (ただし独自拡張の要素も含まれている点に注意が必要)。 メタデータの詳細な項目や解説については、「NDL デジタルアーカイブシステム メタデータスキーマガイドライン」を参照のこと。 <sup>16)</sup>
---------	--

※PREMIS についての参考文献・サイトについては、ページ下部の注釈を参照のこと。<sup>17)18)</sup>

<sup>16)</sup> 国立国会図書館, “NDL デジタルアーカイブシステム メタデータスキーマガイドライン”, <http://www.ndl.go.jp/jp/standards/da/da.pdf>

<sup>17)</sup> PREMIS Official Web Site , <http://www.loc.gov/standards/premis/>

<sup>18)</sup> 後藤敏行, “デジタル情報保存のためのメタデータ: 現状と課題”, [http://www.jstage.jst.go.jp/article/johokanri/50/2/74/\\_pdf/-char/ja/](http://www.jstage.jst.go.jp/article/johokanri/50/2/74/_pdf/-char/ja/)

### 3.3. 電子公文書におけるメタデータの定義

今回のプロトタイプシステム構築にあたって、電子公文書におけるメタデータ定義を表 3-12 の通り整理した。このうち、「コンテナメタデータ」は平成 19 年度までの調査結果等において「アーカイバルメタデータ」として定義されていた内容と同一であるが、混乱を避けるために「コンテナメタデータ」として定義し直している点に注意されたい。

表 3-12 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	上記の技術的メタデータ、記録管理メタデータ、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。

### 3.3.1. メタデータ項目の定義

平成19年度までの調査結果等も参考にして、今回のプロトタイプシステム構築にあたり策定したメタデータ項目の一覧を表3-13に示す。表内の「DAS」の記述は「デジタルアーカイブ・システム」の略とした。

表 3-13 メタデータ項目一覧

※は行政文書ファイル管理簿から取得					メタデータ区分		
	No.	大項目	中項目	小項目	記録管理	技術的	アーカイバル
※	1.1	文書分類	行政文書ファイル管理簿大分類		○		○
※	1.2	文書分類	行政文書ファイル管理簿中分類		○		○
※	1.3	文書分類	行政文書ファイル管理簿小分類		○		○
	1.4	文書分類	文書分類		○		○
※	2.1	タイトル	簿冊名		○		○
	2.2	タイトル	件名		○		○
※	3	作成者			○		○
※	4.1	日付	作成年月日		○		○
	4.2	日付	移管年月日				○
※	5.1	保存条件	保存期間		○		
※	5.2	保存条件	保存期間満了時期		○		
※	5.3	保存条件	保存場所		○		
※	5.4	保存条件	保存時期満了時の措置結果		○		
※	6	媒体種別					○
※	7	管理担当課・係 <sup>*1</sup>			○		
※	8	備考			○		
	9	各府省個別情報			○		
	10	資料内容					○
	11.1	受入情報	受入方法				○
	11.2	受入情報	受入年月日				○
※	11.3	受入情報	移管元府省名		○		○
※	11.4	受入情報	移管時管理担当当局 <sup>*1</sup>				○
	11.5	受入情報	保存場所				○
	11.6	受入情報	移管年度				○
	11.7	受入情報	移管年月日				○
※	12.1	文書番号	行政文書ファイル管理簿番号		○		○
	12.2	文書番号	受入文書番号		○		○
	13	言語			○		○
	14	関連資料、参考文献			○		○

※は行政文書ファイル管理簿から取得					メタデータ区分		
No.	大項目	中項目	小項目	記録管理	技術的	アーカイバル	
15	権利関係			○		○	
16.1	公開条件	公開状況				○	
16.2	公開条件	公開年月日				○	
16.3	公開条件	審査状況				○	
16.4	公開条件	審査年月日				○	
16.5	公開条件	公開に関する移管元の可否		○			
16.6	公開条件	公開に関する移管元の可否(否の理由)		○			
17	媒体情報				○		
17.1	媒体情報	合計サイズ			○		
17.2	媒体情報	ボリューム名			○		
18	電子ファイル情報				○		
18.1	電子ファイル情報	電子ファイルID			○		
18.2	電子ファイル情報	作成時技術情報			○		
18.2.1	電子ファイル情報	作成時技術情報	作成環境(OS名)		○	○	
18.2.2	電子ファイル情報	作成時技術情報	作成環境(OSのバージョン)		○	○	
18.2.3	電子ファイル情報	作成時技術情報	作成アプリケーション名		○	○	
18.2.4	電子ファイル情報	作成時技術情報	作成アプリケーションバージョン		○	○	
18.2.5	電子ファイル情報	作成時技術情報	フォーマット名		○	○	
18.2.6	電子ファイル情報	作成時技術情報	フォーマットバージョン		○	○	
18.2.7	電子ファイル情報	作成時技術情報	メッセージダイジェスト		○		
18.2.8	電子ファイル情報	作成時技術情報	ファイルサイズ		○		
18.2.9	電子ファイル情報	作成時技術情報	元ファイル名		○		
18.2.10	電子ファイル情報	作成時技術情報	文字コード		○		
18.2.11	電子ファイル情報	作成時技術情報	資料区分		○		
18.2.12	電子ファイル情報	作成時技術情報	媒体上のファイルパス		○		
18.2.13	電子ファイル情報	作成時技術情報	ストレージ上のファイルパス		○		
18.2.14	電子ファイル情報	作成時技術情報	フォント情報		○		
18.3	電子ファイル情報	作成時附帯情報			○		

※は行政文書ファイル管理簿から取得					メタデータ区分		
No.	大項目	中項目	小項目	記録管理	技術的	アーカイバル	
18.3.1	電子ファイル情報	作成時附帯情報	ウイルスチェック情報 (ソフトウェア名)		○		
18.3.2	電子ファイル情報	作成時附帯情報	ウイルスチェック情報 (バージョン)		○		
18.3.3	電子ファイル情報	作成時附帯情報	作成時タイムスタンプ		○		
18.3.4	電子ファイル情報	作成時附帯情報	デジタル署名情報 (署名者)		○		
18.4	電子ファイル情報	移管時技術情報			○		
18.4.1	電子ファイル情報	移管時技術情報	作成環境(OS名)		○	○*2	
18.4.2	電子ファイル情報	移管時技術情報	作成環境(OSのバージョン)		○	○*2	
18.4.3	電子ファイル情報	移管時技術情報	作成アプリケーション名		○	○*2	
18.4.4	電子ファイル情報	移管時技術情報	作成アプリケーションバージョン		○	○*2	
18.4.5	電子ファイル情報	移管時技術情報	フォーマット名		○	○*2	
18.4.6	電子ファイル情報	移管時技術情報	フォーマットバージョン		○	○*2	
18.4.7	電子ファイル情報	移管時技術情報	メッセージダイジェスト		○		
18.4.8	電子ファイル情報	移管時技術情報	ファイルサイズ		○		
18.4.9	電子ファイル情報	移管時技術情報	ファイル名		○		
18.4.10	電子ファイル情報	移管時技術情報	文字コード		○		
18.4.11	電子ファイル情報	移管時技術情報	資料区分		○		
18.4.12	電子ファイル情報	移管時技術情報	ストレージ上のファイルパス		○		
18.4.13	電子ファイル情報	移管時技術情報	抽出テキストデータファイル名		○		
18.4.14	電子ファイル情報	移管時技術情報	抽出テキストデータファイルサイズ		○		
18.4.15	電子ファイル情報	移管時技術情報	フォント情報		○		
18.5	電子ファイル情報	移管時附帯情報					
18.5.1	電子ファイル情報	移管時附帯情報	ウイルスチェック情報 (ソフトウェア名)		○		
18.5.2	電子ファイル情報	移管時附帯情報	ウイルスチェック情報 (バージョン)		○		

※は行政文書ファイル管理簿から取得					メタデータ区分		
No.	大項目	中項目	小項目	記録管理	技術的	アーカイバル	
18.5.3	電子ファイル情報	移管時附帯情報	作成時タイムスタンプ		○		
18.5.4	電子ファイル情報	移管時附帯情報	その他附帯情報		○		
19	保管履歴情報					○	
20	利用履歴情報					○	
21	文書ID			○	○	○	
22.1	DAS目録情報	レファレンス・コード				○	
22.2	DAS目録情報	簿冊番号1				○	
22.3	DAS目録情報	簿冊番号2				○	
22.4	DAS目録情報	簿冊番号枝番				○	
22.5	DAS目録情報	件名番号				○	
22.6	DAS目録情報	件名事項SEQ				○	
22.7	DAS目録情報	目録種別				○	

\*1 行政文書ファイル管理簿のデータを基本とし、同一の値が格納される。

\*2 デジタルアーカイブ用フォーマットに変換した場合の情報を格納する。

今回の調査においては、表 3-12 に挙げた各メタデータを、OAIS 参照モデルに基づいてパッケージ化を行うこととした。各メタデータと OAIS 参照モデルにおける情報パッケージとの対応関係を図 3-3 に示す。

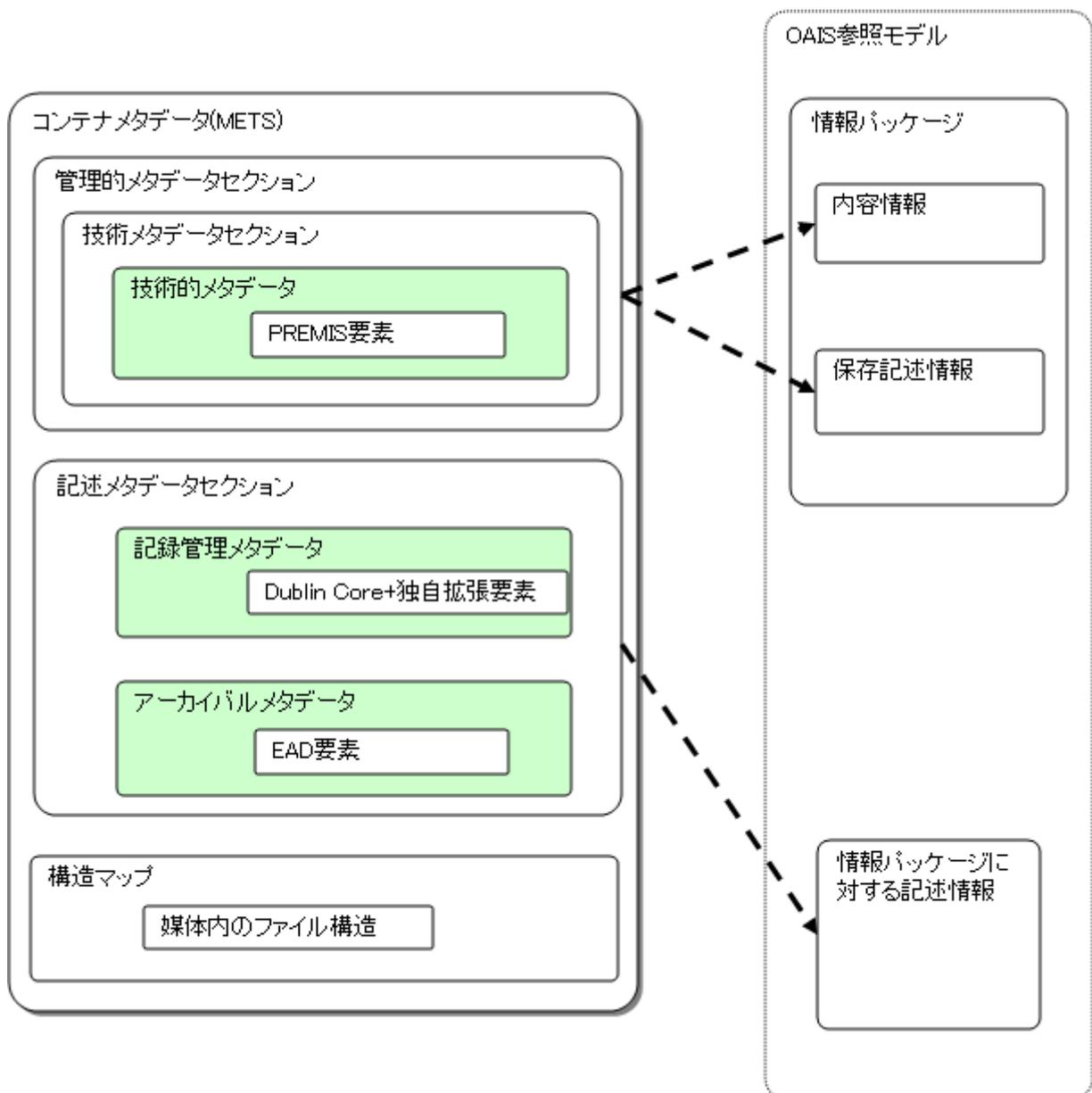


図 3-3 各メタデータと OAIS 参照モデルの対応関係

### 3.3.2. 管理対象オブジェクトとしての電子公文書等の構造について

#### (1) 行政文書ファイル

現在、各府省庁等が現用段階で管理する公文書等は、基本的に「行政文書ファイル」という単位で管理されている。また、歴史資料として重要な公文書等の移管についても、「行政文書ファイル」を基本の単位として行われている。

現用段階における公文書等の管理については、「行政文書の管理方策に関するガイドラインについて」(平成12年2月25日 各省庁事務連絡会議申合せ)(以下「ガイドライン」という。)に基づいて、管理されている。ガイドラインでは、「行政文書ファイル」を次のように説明している。

「『能率的な事務又は事業の処理及び行政文書の適切な保存の目的を達成するためにまとめられた、相互に密接な関連を有する行政文書(保存期間が1年以上のものであって、当該保存期間を同じくすることが適当であるものに限る。)の集合物』であり、小分類の下で保存及び廃棄について同じ取扱いをすることが適当であるものである。」

また、「いわゆる『簿冊』と同義ではなく、複数の簿冊が1ファイルである場合、一つの簿冊の中に複数のファイルが存在する場合等種々の態様が想定される」としている。

#### (2) 行政文書ファイルの構造

これらから見て、「行政文書ファイル」は、1件又は2件以上の「行政文書」から構成される集合体であると考えられる。

では、2件以上の「行政文書」から構成されている「行政文書ファイル」の場合、1件の独立した単位としての「行政文書」を他の「行政文書」と識別することができるか。また、複数の「行政文書」相互の関係はどのようになっていると考えられるか。ガイドラインは、「行政文書ファイル」の設定方法について、次のような方法(または、複数の方法の組み合わせ)を例示している。

##### 1) 内容(主題)別

行政文書に書かれている内容(主題)をとらえて、その内容(主題)ごとにまとめる方法(例:〇〇制度各国調査結果ファイル、〇〇審議会議事録ファイル等)

##### 2) 形式別

行政文書の内容や相手方とは関係なく、その形式をとらえてまとめる方法(例:〇〇課例規ファイル、〇〇関係照会・回答ファイル等)

##### 3) 様式・標題別

帳票類や伝票類のように、行政文書の様式・標題が定められている場合に、その標題をそのままファイル名称とし、まとめる方法(例:〇〇申請書ファイル、閲覧申出書ファイル、〇〇届書ファイル等)

##### 4) 案件(一件)別

許認可の申請から処分まで、工事の計画から完了までの行政文書など、一つの案件に係る行政文書を順序立ててまとめる方法(例:〇〇許可(認可)一件ファイル、〇〇訴訟一件ファイル等)

##### 5) 相手方別

行政文書に係る提供者、提出先等の相手方をとらえて、その相手方ごとにまとめる方法(例:法人台帳(特殊法人、事業者等)、国会提出・説明資料ファイル等)

## 6) 時期別

同種の内容の行政文書を一定の期間ごとにまとめる方法(例:○年○月相談案件ファイル、○年○月支払書ファイル等)

以上のような多様な設定方法が可能であることを考慮に入れると、時間又は業務プロセスの前後関係、関係する案件、関係者等の違い等により、1件の「行政文書」を他の1件又は複数件の「行政文書」と識別することは可能であるが、識別のために必要なメタデータがない場合は、自動的に識別することは容易ではないと考えられる。また、「雑件ファイル」というような相互に関係が全く無い又はほとんどない複数件の「行政文書」が1つの「行政文書ファイル」を構成している場合も、1件の「行政文書」と他の1件又は複数件の「行政文書」を識別することは容易ではないと考えられる。

### (3) 行政文書の構造

では、1件の「行政文書」は、どのような構成になっているか。

典型的な決裁文書の場合、決裁面(鑑)及び別紙案から構成されている。また、決裁面(鑑)及び別紙案とは別に、参考資料等が添付されていることもある。電子決裁システムにより作成される電子公文書等の場合、決裁面(鑑)は、電子決裁システム上で作成・決裁され、それ以外の別紙案等は、一般的なオフィス・アプリケーション等により作成されていることが想定される。

次に、許認可案件等の申請に係る決裁文書の場合は、決裁面(鑑)及び別紙案等のほかに、申請書が添付されている。その申請書も、申請書本体のほかに、図面等の参考資料等が添付されていることがある。

明示的な決裁行為がなされていない文書の場合、例えば、会議の配布資料や部内の検討資料等では、「資料1」～「資料〇」と付番された1つ以上の資料の集合体が1件の「行政文書」を構成している。

以上のように、1件の「行政文書」は、決裁面(鑑)、別紙案、参考資料、申請書等1つまたは複数のドキュメントから構成されていると考えられる。複数のドキュメントから構成されている場合、ドキュメント相互の関係は、並列的又は階層的であることが多いと考えられるが、並列的とも階層的とも分類できない関係になっていることも想定される。また、並列的又は階層的な関係になっている場合を含めて、ドキュメント相互の関係を外形的又は自動的に判別することは、それらに関する情報を記録したメタデータがない場合は、きわめて困難になると考えられる。

### (4) ドキュメントの構造

では、「行政文書」を構成する個々のドキュメントは、どのような構成になっているのだろうか。1つのドキュメントが1つの電子ファイルから構成されている場合もあり得るが、多くの場合は、1つのドキュメントは複数の電子ファイルから構成されることが多いと考えられる。

1つのドキュメントを構成する複数の電子ファイルは、同一のアプリケーションで作成された1つのフォーマットの電子ファイルである場合もあれば、複数の異なるアプリケーションで作成された複数の異なるフォーマットの電子ファイルである場合もあると考えられる。例えば、決裁文書の「別紙案」がテキスト文書である場合は、「別紙案」ドキュメントはワープロ・ソフトで作成されたワープロ・フォーマット(例:マイクロソフト doc ファイル)のファイルが1つ又は複数集まった集合体であることが多いと考えられる。一方、「別紙案」に、テキスト文書のほかに図表等が含まれている場合は、ワープロ・フォーマットのファイルのほかに、表は表計算フォーマットのファイル、図はプレゼンテーション・フォーマットのファイル、画像ファイ

ル等であることも考えられる。

ドキュメントを構成する複数の電子ファイルが1つのフォーマットで統一されていようと、複数の異なるフォーマットであったとしても、複数の電子ファイル相互の関係は、識別のために必要なメタデータがない場合、自動的又は外形的に識別するのは容易ではないと考えられる。

なお、今回の調査研究では、ワープロ、表計算及びプレゼンテーション用のフォーマットのファイルについては、個別に「長期保存フォーマット」として PDF/A に変換した(変換可能な場合に限る。)が、1つのドキュメントを構成しているファイルを PDF/A に変換した上で、複数の PDF/A を結合して1つの PDF/A ファイルとして管理すれば、ファイル相互の関係を固定化できるというメリットがある。この手法は、電子公文書等の移管・保存・利用システムだけでなく、現用段階での文書管理でも活用可能であると考えられる。

#### (5) オリジナルフォーマット、長期保存フォーマット及びテキスト抽出データ

本調査研究においては、1つの電子ファイルに対して、オリジナルフォーマットのほか、オリジナルフォーマットから変換した長期保存フォーマット(変換が可能な場合に限る。)及びテキスト抽出データを生成して管理することを試みている。つまり、ドキュメントを構成する電子ファイルは、さらに最大3つのフォーマットのファイルから構成されるということになる。

#### (6) まとめ

電子公文書等の移管・保存・利用に係るシステムにおいて管理すべきオブジェクトの基本単位となるのは、(4)で述べた電子ファイルである。その場合、電子ファイルのフォーマット及びメタデータを適切に設定する必要があるのはもちろんであるが、行政文書ファイルー行政文書ドキュメントーファイルという階層構造及び同一階層の構成要素相互の関係を適切に把握し維持管理(そのためのメタデータの付与及び維持管理を含む。)する必要がある。

一方、電子公文書等を1つの情報パッケージ(OAIS 参照モデル)として管理する場合、4で述べた電子ファイルを情報パッケージの基本単位とすることが考えられるが、この場合、基本単位の数が膨大となるため、マクロなシステム全体の管理が難しくなる可能性もある。一方、行政文書ファイル又は行政文書を情報パッケージの基本単位とする場合、マクロなシステム全体の管理にメリットがあるほか、行政文書ファイルの構造を安定的に保存する点でも望ましいと考えられる反面、一つ一つの情報パッケージが収納する情報量が大きくなることが考えられる。

上記の概念を図に表したものが図 3-4 である。

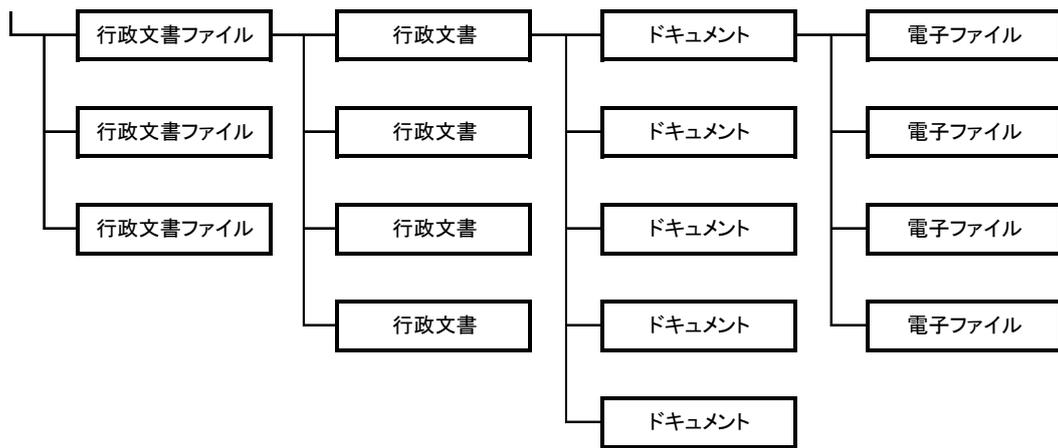


図 3-4 電子公文書等の構造

### 3.3.3. メタデータスキーマの定義にあたっての検討事項

#### (1) 電子公文書の構造

前述の電子公文書等の構造を踏まえて、電子公文書の移管・受入時に各府省より提供される基礎情報として、行政文書ファイル管理簿の情報が添付されることを前提とした。しかし、行政文書ファイル管理簿の情報だけでは行政文書の単位を機械的に判断することは難しいため、プロトタイプシステムでは 1 件の電子公文書の単位を行政文書ファイル管理簿の 1 レコード=1 行政文書ファイルとした。

#### (2) 公文書の持つ階層性

行政文書ファイル管理簿の項目には分類情報として大分類、中分類、小分類の 3 階層に該当する項目が存在している。

今回のプロトタイプシステムでは公文書の階層性を表現するための情報として、これらの分類情報及び階層構造を用いる仕組みとした。ただし、この分類情報は各府省における組織階層を表す場合もあれば、純粋な文書の分類を表す場合もあり、分類の基準、その数においてばらつきがある点に注意が必要である。

(3) 電子公文書とメタデータ、ファイルの対応関係

プロトタイプシステムにおいて、行政文書ファイル(行政文書ファイル管理簿のレコード)とメタデータ、電子公文書の媒体(に含まれるフォルダ、電子ファイル)の対応関係を図 3-5 に示す。

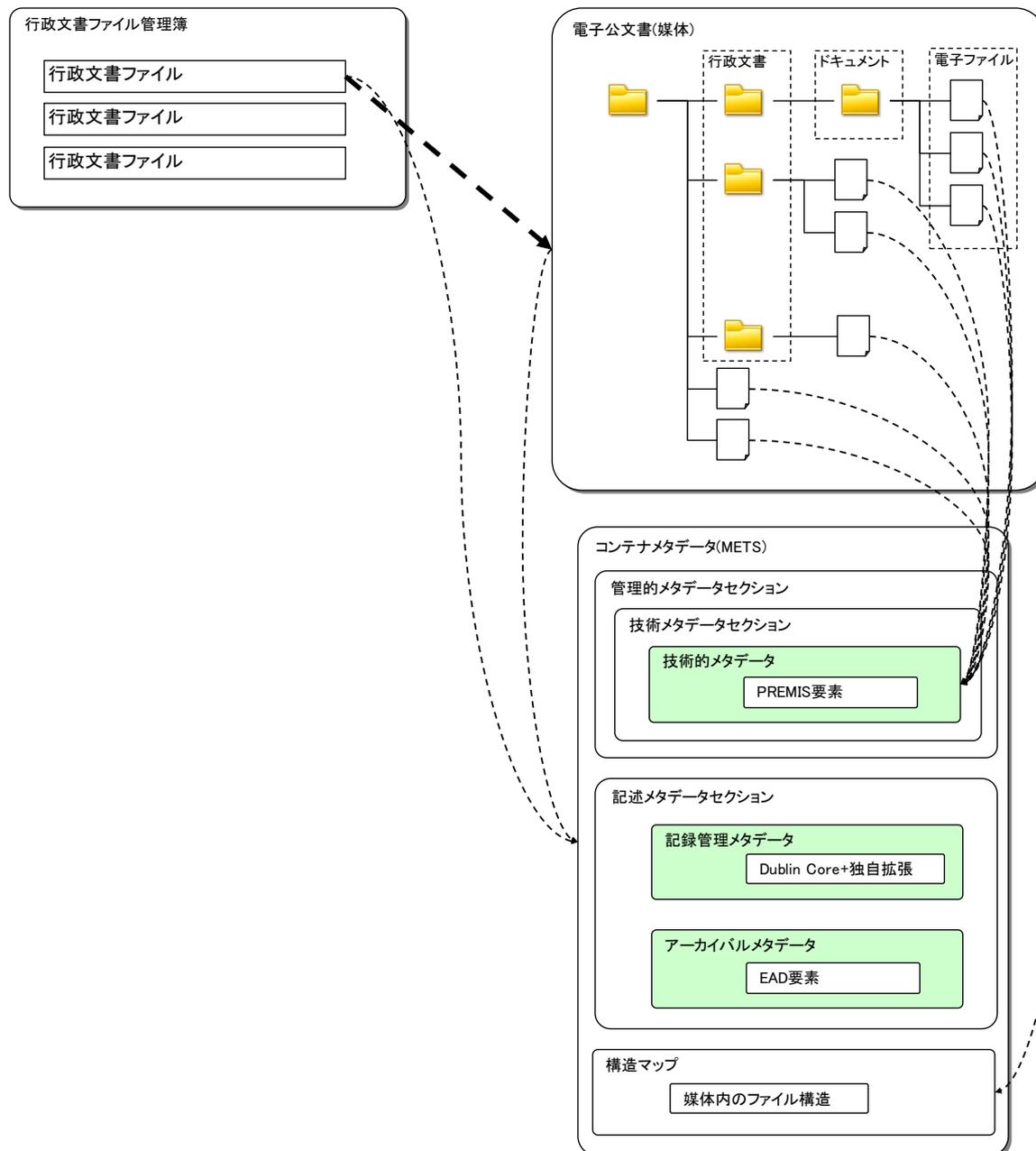


図 3-5 電子公文書の移管におけるメタデータと電子ファイルの対応関係

### 3.3.4. 記録管理メタデータ

#### (1) 記録管理メタデータ項目

記録管理メタデータは、主に行政文書ファイル管理簿を基にしたデータから構成される。

記録管理メタデータ項目については、Dublin Core の要素と対応する項目は Dublin Core 要素を用い、それ以外の項目については独自要素を用いて表現することとした。

本調査では、平成 19 年度までの調査研究での検討課題等も参考とし、データ項目を再検討した上で策定した。以下に記録管理メタデータ項目の一覧を示す。

表 3-14 記録管理メタデータ項目

No.	項目名	要素(dc:は Dublin Core、それ以外は独自要素)	対応メタデータ項目	説明
1	文書 ID	/recordkeepingMD/dc:identifier	21	文書のシステム識別番号
2	タイトル	/recordkeepingMD/dc:title	2.1,2.2	文書の簿冊名または件名
3	作成者	/recordkeepingMD/dc:creator	3	文書の作成した所属または個人
4	行政文書ファイル管理簿大分類	/recordkeepingMD/classification /class @level="1"	1.1	文書の分類(大)
5	行政文書ファイル管理簿中分類	/recordkeepingMD/classification /class @level="2"	1.2	文書の分類(中)
6	行政文書ファイル管理簿小分類	/recordkeepingMD/classification /class @level="3"	1.3	文書の分類(小)
7	移管元府省名	/recordkeepingMD/preservation/ transferredFrom	11.3	文書の移管元府省名(コード)
8	保存期間	/recordkeepingMD/preservation/ term	5.1	文書の作成時期から満了時期までの期間
9	保存期間満了時期	/recordkeepingMD/preservation/ endDate	5.2	文書の一番遅い保存期間満了時期
10	保存場所	/recordkeepingMD/preservation/ location	5.3	文書の保存場所
11	保存時期満了時の措置結果	/recordkeepingMD/preservation/ resultAfterExpire	5.4	文書の保存期間満了期の措置結果
12	管理担当課・係	/recordkeepingMD/dc:publisher	7	文書の管理担当課・係
13	備考	/recordkeepingMD/note	8	文書の保存期間満了期の措置結果の詳細

No.	項目名	要素(dc:は Dublin Core、それ以外は 独自要素)	対応メ タ デ ー タ 項 目	説明
14	各府省個別情 報	/recordkeepingMD/dc:description	9	文書の所属する府省庁で 記載した情報
15	行政文書ファイ ル管理簿番号	/recordkeepingMD/dc:identifier @type="行政文書ファイル管理簿 番号"	12.1	行政文書ファイル管理簿の 識別番号(キー)
16	受入文書番号	/recordkeepingMD/dc:identifier @type="受入文書番号"	12.2	移管後の管理番号
17	言語	/recordkeepingMD/dc:language	13	文書が書かれた言語コード
18	関連資料、参 考文献	/recordkeepingMD/relatedItem	14	文書の関連リソース
19	権利関係	/recordkeepingMD/dc:rights	15	文書の権利関係
20	公開に関する 移管元の可否	/recordkeepingMD/accessRestrict/v alue	16.5	移管元省庁における公開 可否の判断結果
21	公開に関する 移管元の可否 (否の理由)	/recordkeepingMD/accessRestrict/r eason	16.6	移管元省庁において公開 が否である場合の理由
22	文書分類	/recordkeepingMD/docClass	1.4	No.6 の下層の文書の分類
23	作成年月日	/recordkeepingMD/dc:date	4.1	文書の作成日情報
24	元ファイル名	/recordkeepingMD/orgFileName	18.2.9	文書のファイル名

対応メタデータ項目は、「エラー! 参照元が見つかりません。」のNo.を表す。

## (2) データ項目の策定について

記録管理メタデータ項目については、メタデータの国際標準である Dublin Core の要素と行政文書ファイル管理簿の項目を組み合わせで策定した。

### 3.3.5. 技術的メタデータ

#### (1) 技術的メタデータ項目

技術的メタデータは、電子ファイル単位の技術情報を基にしたデータから構成される。

技術的メタデータ項目については、PREMIS 要素で表現することとし、平成 19 年度までの調査研究での検討課題等も参考としながらデータ項目を策定した。以下にその表を示す。

表 3-15 技術的メタデータ項目

No.	項目名	PREMIS2.0 の要素	対応メタデータ項目	説明
1	オブジェクト ID:種別	/premis/object xsi:type="file"/objectIdentifier/objectIdentifierType	-	対象の種別
2	オブジェクト ID:値	/premis/object xsi:type="file"/objectIdentifier/objectIdentifierValue	18.1	ファイル名
3	保存レベル:値	/premis/object xsi:type="file"/preservationLevel/preservationLevelValue	-	ファイル種別
4	メッセージダイジェストアルゴリズム	/premis/object xsi:type="file"/objectCharacteristics/fixity/messageDigestAlgorithm	-	メッセージダイジェストのアルゴリズム
5	メッセージダイジェスト	/premis/object xsi:type="file"/objectCharacteristics/fixity/messageDigest	18.2.7 、 18.4.7	メッセージダイジェストの出力値
6	サイズ	/premis/object xsi:type="file"/objectCharacteristics/size	18.2.8 、 18.4.8 、 18.4.14	ファイルサイズ
7	フォーマット名	/premis/object xsi:type="file"/objectCharacteristics/format/formatDesignation/formatName	18.2.5 、 18.4.5	ファイルのフォーマット名
8	フォーマットバージョン	/premis/object xsi:type="file"/objectCharacteristics/format/formatDesignation/formatVersion	18.2.6 、 18.4.6	ファイルのフォーマットバージョン

No.	項目名	PREMIS2.0 の要素	対応メタデータ項目	説明
9	フォーマット注記	/premis/object xsi:type="file"/objectCharacteristics/format/formatNote	-	ファイルのフォーマットにおける補足情報
10	作成アプリケーション名	/premis/object xsi:type="file"/objectCharacteristics/creatingApplication/creatingApplicationName	18.2.3 18.4.3	ファイルを作成したアプリケーション名
11	作成アプリケーションバージョン	/premis/object xsi:type="file"/objectCharacteristics/creatingApplication/creatingApplicationVersion	18.2.4 18.4.4	ファイルを作成したアプリケーションバージョン
12	作成日付	/premis/object xsi:type="file"/objectCharacteristics/creatingApplication/dateCreatedByApplication	18.3.3 18.5.3	ファイルを作成した日付
13	制限種別	/premis/object xsi:type="file"/objectCharacteristics/inhibitors/inhibitorType	-	使用した暗号方法
14	制限対象	/premis/object xsi:type="file"/objectCharacteristics/inhibitors/inhibitorTarget	-	守られている内容(コンテンツまたは機能)
15	オリジナルファイル名	/premis/object xsi:type="file"/originalName	18.2.9	ファイルのオリジナル名
16	ストレージ場所:種別	/premis/object xsi:type="file"/storage/contentLocation/contentLocationType	-	ロケーションスキーマ
17	ストレージ場所:値	/premis/object xsi:type="file"/storage/contentLocation/contentLocationValue	18.2.12 18.2.13 18.4.12	ロケーションの値
18	ストレージ:媒体	/premis/object xsi:type="file"/storage/storageMedium	-	ストアされている物理的媒体
19	依存アイテム名	/premis/object xsi:type="file"/environment/dependency/dependencyName	18.2.14 18.4.15	必須な添付物

No.	項目名	PREMIS2.0 の要素	対応メタデータ項目	説明
20	ソフトウェア名	/premis/object xsi:type="file"/environment/software/swName	18.2.1 、 18.4.1	再生に必要なソフトウェア名
21	ソフトウェアバージョン	/premis/object xsi:type="file"/environment/software/swVersion	18.2.2 、 18.4.2	再生に必要なソフトウェアバージョン
22	ソフトウェア種別	/premis/object xsi:type="file"/environment/software/swType	-	再生に必要なソフトウェアのタイプ
23	署名者	/premis/object xsi:type="file"/signatureInformation/signature/signer	18.3.4	署名した個人、責任権利者

## (2) データ項目の策定について

PREMIS の仕様にある要素は非常に多岐にわたるため、技術的メタデータ項目の検討にあたっては Mandatory(必須)とされている項目のうち、必要と考えられる項目のみに限定した。従って、個々の要素としては PREMIS から流用しているが、技術的メタデータ全体としては PREMIS スキーマに準じたものではない点に注意されたい。

## (3) 今後の検討課題

上記の通り、今回対象としなかった PREMIS 要素についても更なる調査・検討が望まれる。

また、No.13,14 の制限については、PREMIS の定義:アクセス・使用・移動などの制限を意味し、項目としてとりあげたが、今回のプロトタイプシステムではその内容に該当する暗号化を行っていない。また、No.19 の依存アイテム名には、ファイルのアプリケーションで使用しているフォントを対応させたが、No.20～23 には OS 情報を対応させていてそれと同じ environment の要素に対応させてよいものかどうか検討が必要な項目もある。

### 3.3.6. アーカイバルメタデータ

#### (1) アーカイバルメタデータ項目

アーカイバルメタデータは、主に行政文書ファイル管理簿を基にした記録管理メタデータの情報と、アーカイブにおいて必要とされるメタデータから構成される。

アーカイバルメタデータ項目については、EAD により表現することとし、その項目を選定した。また、国立公文書館デジタルアーカイブ・システムにおける EAD 定義も参考とし、必要と思われる項目を追加している。

平成 19 年度までの調査研究での検討課題等も参考としながらデータ項目を策定した。以下にその表を示す。

表 3-16 アーカイバルメタデータ項目

No.	項目名	要素 ※EAD の /ead/archdesc/dsc 要素以降を記述する	対応メタデータ項目	説明
1	タイトル	/c/did/unittitle @label=" 簿冊 標題" @label=" 件名"	2.1.2.2	文書の簿冊名または件名
2	作成者名称	/c/did/origination @label=" 作成 部局"	3	文書の作成した所属または個人
3	作成年月日	/c/did/unitdate @label=" 作成年 月日" normal="*"	4.1	文書の作成日情報
4	言語	/c/did/langmaterial/language @langcode="*"	13	文書が書かれた言語コード
5	関連資料、参考文献	/c/relatedmaterial/p	14	文書の関連リソース
6	文書番号	/c/scopecontent/note @label=" 文書番号"/p	12.2	移管後の管理番号
7	受入年月日	/c/acqinfo/p[1]/date @type=" 受 入年月日" @normal="*"	11.2	移管時の年月日
8	受入方法	/c/acqinfo/p[1]	11.1	移管方法(コード) 01=移管
9	移管年度/年月日	/c/acqinfo/p[2]/date @type=" 移 管年度" @normal="*"	11.6、11.7	移管時の年度(コード)
10	記述レベル	/c @level="*"	22.7 の値により自動付与	簿冊か件名の識別 簿冊 =file 件名=item
11	資料内容	/c/scopecontent/note @label=" 資料内容"/p	10	文書の内容説明
12	レファレンス・コード	/c/did/unitid @label="レファレンス・コード"	22.1	移管後の管理番号

No.	項目名	要素 ※EAD の/ead/archdesc/dsc 要素以降を記述する	対応メタデータ項目	説明
13	データ種別	/c/did/materialspec @label="データ種別"	-	文書のデータ種別 1=公文書
14	目録種別	/c/did/materialspec @label="目録種別"	22.7	文書の目録種別 簿冊=2 件名=3
15	簿冊番号1	/c/did/materialspec @label="簿冊番号1"	22.2	文書の簿冊番号 1
16	簿冊番号2	/c/did/materialspec @label="簿冊番号2"	22.3	文書の簿冊番号 2
17	簿冊番号枝番	/c/did/materialspec @label="簿冊番号枝番"	22.4	文書の枝番号
18	件名の表示	/c/did/materialspec @label="件名の表示"	-	件名を表示状況 0=表示
19	件名番号	/c/did/materialspec @label="件名番号"	22.5	文書が件名の時の番号
20	件名番号 SEQ	/c/did/materialspec @label="件名事項 SEQ"	22.6	文書が件名の時の番号 SEQ
21	公開状況	/c/accessrestrict @type="公開状況"/p	16.1	文書の公開状況 02=公開保留
22	公開年月日	/c/accessrestrict @type="公開状況"/p/date @type="公開年月日"	16.2	文書の公開年月日
23	審査状況	/c/accessrestrict @type="審査状況"/p	16.3	文書の公開判定の情報 01=要審査
24	審査年月日	/c/accessrestrict @type="審査状況"/p/date @type="審査年月日"	16.4	文書の公開判定年月日
25	保管履歴情報	/c/custodhist/p	19	保管履歴情報
26	利用履歴情報	/c/bioghist/p	20	利用履歴情報
27	文書 ID	/c/did/unitid @label="文書 ID"	21	文書のシステム識別番号
28	保存場所	/c/did/physloc @label="保存場所"	11.5	移管後の文書の保存場所
29	移管元府省名	/c/acqinfo/p[2]/corpname @role="移管省庁"	11.3	文書の移管元府省名(コード)
30	移管時管理担当部局	/c/acqinfo/p[2]/corpname @role="移管時管理担当部局"	11.4	文書の管理担当課・係
31	媒体	/c/did/physdesc/genreform	6	文書の媒体種別

No.	項目名	要素 ※EAD の /ead/archdesc/dsc 要素以降を記述する	対応メタデータ項目	説明
		@type="媒体"		
32	文書分類;大分類	/c/did/materialspec @label="大分類"	1.1	文書の分類(大)
33	文書分類;中分類	/c/did/materialspec @label="中分類"	1.2	文書の分類(中)
34	文書分類;小分類	/c/did/materialspec @label="小分類"	1.3	文書の分類(小)

## (2) データ項目の策定について

アーカイバルメタデータ項目の策定にあたり、平成 19 年度までの調査研究に挙げられておらず、今回の調査において追加した項目及びその理由は以下の通りである。

4	言語	記録管理メタデータのみとなっていたが、アーカイバルメタデータとしても設定。
5	関連資料、参考文献	記録管理メタデータのみとなっていたが、アーカイバルメタデータとしても設定。
9	移管年度/年月日	移管年度を追加。 ただし値は移管年月日より自動的に設定する。
14	目録種別	デジタルアーカイブ・システム連携で必要となるため追加。
15	簿冊番号1	デジタルアーカイブ・システム連携で必要となるため追加。
16	簿冊番号2	デジタルアーカイブ・システム連携で必要となるため追加。
17	簿冊番号枝番	デジタルアーカイブ・システム連携で必要となるため追加。
18	件名の表示	デジタルアーカイブ・システム連携で必要となるため追加。
19	件名番号	デジタルアーカイブ・システム連携で必要となるため追加。
20	件名番号 SEQ	デジタルアーカイブ・システム連携で必要となるため追加。
27	文書 ID	システム管理上必要と考えられるため追加。
32	文書分類;大分類	記録管理メタデータのみとなっていたが、アーカイバルメタデータとしても設定。
33	文書分類;中分類	記録管理メタデータのみとなっていたが、アーカイバルメタデータとしても設定。
34	文書分類;小分類	記録管理メタデータのみとなっていたが、アーカイバルメタデータとしても設定。

## (3) 今後の検討課題

3.3.3 項でも述べた通り、本調査における電子公文書の階層構造は行政文書ファイル管理簿の大分類、中分類、小分類の 3 階層を前提としているが、デジタルアーカイブ・システムでは目録情報の階層として簿冊と件名という構造も存在している。簿冊と件名は、紙媒体での公文書ではそれぞれ図 3-4 における

行政文書とドキュメントに該当する概念と言えるが、電子公文書の場合にも同様に適用できるのか、という点は十分に検討すべきである。

### 3.3.7. コンテナメタデータ

#### (1) コンテナメタデータ項目

コンテナメタデータは、記録管理メタデータ、アーカイバルメタデータ、技術的メタデータの全てのメタデータを包含し、1 件の電子公文書に対応するメタデータの単位である。

コンテナメタデータについては、METS により表現することとし、METS における構造との対応は以下の通りとした。

表 3-17 コンテナメタデータ項目(構造)

No.	項目名	項目名(METS)	要素名 ※METS 1.7 に準拠する
1	コンテナメタデータ		/mets
2	METS ヘッダ	METS ヘッダ	/mets/metsHdr
3	コンテナメタデータ作成日時	作成日付	/mets/metsHdr @CREATEDATE
4	コンテナメタデータ最終更新日時	最終更新日付	/mets/metsHdr @LASTMODDATE
5	記録管理メタデータ	記述メタデータセクション	/mets/dmdSec
6	ID	ID	/mets/dmdSec @ID
7		メタデータ部	/mets/dmdSec/mdWrap
8		メタデータ種別	/mets/dmdSec/mdWrap @MDTYPE
9		XML データ	/mets/dmdSec/mdWrap/xmlData
10	アーカイバルメタデータ	記述メタデータセクション	/mets/dmdSec
11	ID	ID	/mets/dmdSec @ID
12		メタデータ部	/mets/dmdSec/mdWrap
13		メタデータ種別	/mets/dmdSec/mdWrap @MDTYPE
14		XML データ	/mets/dmdSec/mdWrap/xmlData
15	管理メタデータセクション	管理メタデータセクション	/mets/amdSec
16	技術的メタデータ	技術的メタデータ	/mets/amdSec/techMD
17	ID	ID	/mets/amdSec/techMD @ID
18		メタデータ部	/mets/amdSec/techMD/mdWrap
19		メタデータ種別	/mets/amdSec/techMD/mdWrap @MDTYPE
20		XML データ	/mets/amdSec/techMD/mdWrap/xmlData
21	ファイルセクション	ファイルセクション	/mets/fileSec
22	ファイルグループ	ファイルグループ	/mets/fileSec/fileGrp
23	ファイル	ファイル	/mets/fileSec/fileGrp/file
24	ファイル ID	ID	/mets/fileSec/fileGrp/file @ID

No.	項目名	項目名(METS)	要素名 ※METS 1.7 に準拠する
25	ファイルの管理メタデータ		/mets/fileSec/fileGrp/file @ADMID
26	ファイル場所	ファイル場所	/mets/fileSec/fileGrp/file/FLocat @xlink:href
27		場所タイプ	/mets/fileSec/fileGrp/file/Flocat @LOCTYPE
28	媒体上のファイル構造	構造マップ	/mets/structMap
29	フォルダ構成		/mets/structMap/div @TYPE="folder"
30	フォルダ名		/mets/structMap/div @LABEL
31	ファイル		/mets/structMap/div/fptr @FILEID

(2) 今後の検討課題

本調査におけるメタデータスキーマとしての METS 要素では、表 3-5 に挙げている記述単位のうち、構造リンクと動作セクションは使用していないが、これらの記述も含めるべきかどうか、METS の適用事例を踏まえつつ今後検討が必要と思われる。

## 4. 受入・保存管理機能

### 4.1. 調査概要

本項では、電子公文書等の受入・保存管理に必要な、検疫・媒体変換、フォーマット変換、アーカイバルメタデータ作成、保存の各機能について、各機能の評価、連携方法、実運用における課題と対策を調査する。

調査にあたって構築したプロトタイプシステムにおける、各機能間の関連図を図 4-1 に示す。太線で囲った部分が受入・保存管理機能の範囲である。

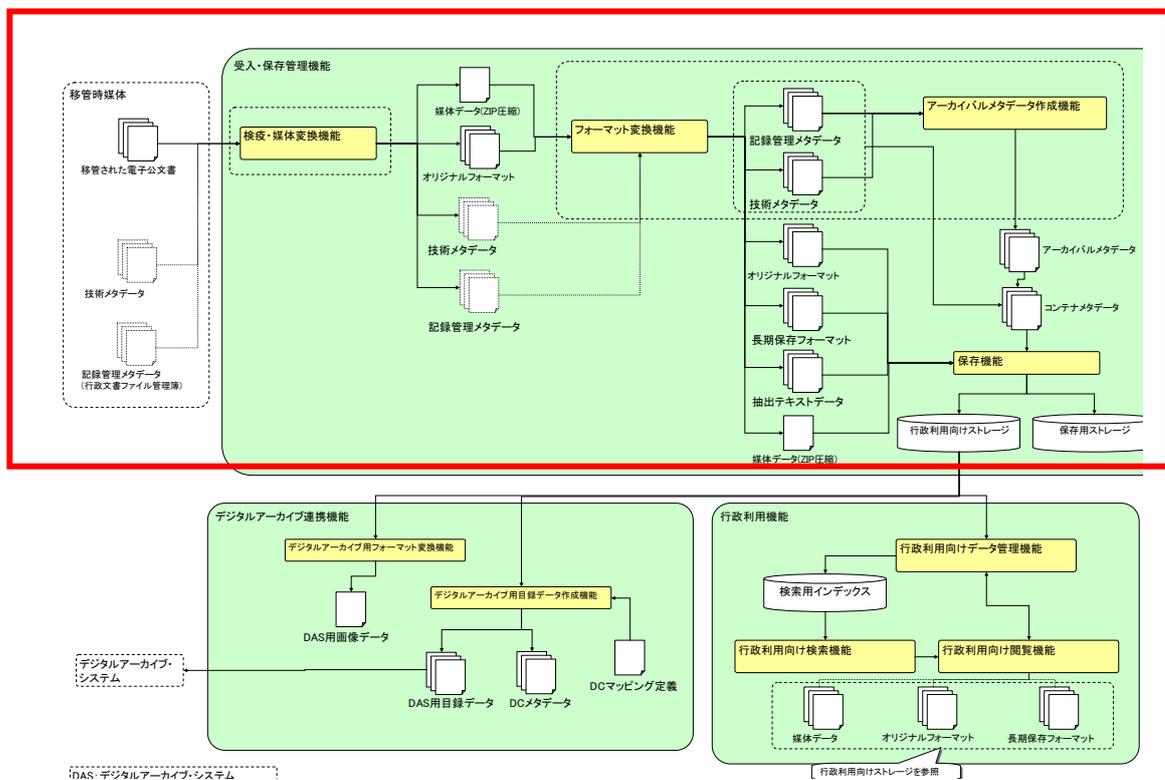


図 4-1 プロトタイプシステムにおける機能関連図

## 4.2. 検疫・媒体変換機能実証結果報告

### 4.2.1. プロトタイプシステムにおける機能概要

検疫・媒体変換機能は、受入対象となる電子公文書等の格納された媒体に対してウイルスチェックを行い、ウイルスに感染していない場合は検疫済データフォルダにコピーする。ウイルスが検出された場合は処理を中断し、ログファイルにメッセージを出力する。

また、媒体からのコピーを行う際に、媒体全体または特定のフォルダのみを媒体データとして ZIP 形式で圧縮し、検疫済みデータフォルダに出力する。

プロトタイプシステムでは、媒体・フォーマット変換サーバ(OS:Windows Server 2003)上で検疫・媒体変換機能を実装し、稼働させた。媒体・フォーマット変換サーバは移管媒体のウイルスチェックを行うため、ウイルスに感染したファイルが含まれていた場合、ネットワーク上の他のサーバ機器に感染が拡大する恐れがある。そのため、セキュリティの観点から媒体・フォーマット変換サーバについてはネットワークに接続せず、スタンドアロンの状態とした。

### 4.2.2. ウイルスチェック

プロトタイプシステムでは、ウイルスチェックソフトウェアとして「エフセキュア アンチウイルス Windows サーバ」を導入した。

#### (1) ウイルスチェックソフトの機能

「エフセキュア アンチウイルス Windows サーバ」の機能は以下の通りである。

- リアルタイム自動保護
- ルートキット検知
- 一元集中管理

プロトタイプシステムでは、バッチ処理により移管媒体から自動的に処理を行う仕組みとしているため、コマンドラインからのウイルスチェックを実行することも必要な要件となった。「エフセキュア アンチウイルス Windows サーバ」にはコマンドラインによるウイルスチェックの機能がある。

また、スタンドアロン環境で稼働させるため、ウイルス定義(パターン)ファイルの更新がネットワークを経由せずに実行することも要件である。プロトタイプシステムでは、エフセキュア社のサイトから最新のアップデートプログラムをダウンロードし、それを MO 等のメディア経由で媒体・フォーマット変換サーバにコピーすることで、パターンファイルを最新の状態とした。

#### (2) ウイルスチェックソフトの対応フォーマット

「エフセキュア アンチウイルス Windows サーバ」では、チェック対象となるフォーマット(ファイル形式)の制限は特に存在しない。ただし、データベース(MySQL, PostgreSQL, Oracle, SQLServer など)ファイルについては、現状感染するウイルスが存在しないことからウイルス検査を行わない仕様となっている<sup>19</sup>。

---

<sup>19</sup> [http://www.f-secure.co.jp/support/html/products\\_all\\_00023.html](http://www.f-secure.co.jp/support/html/products_all_00023.html)

(3) 移管媒体・フォーマットごとのウイルスチェック処理結果

プロトタイプシステム上でのウイルスチェック処理にかかる時間を、対象となるファイルのサイズ及び移管媒体の種類ごとに計測した結果を表 4-2 に示す。対象としたファイルのファイル形式、サイズは表 4-1 に示した。

表 4-1 対象ファイル形式及びサイズ

100KB 未満	Excel2003 形式	49,664 バイト
100～500KB	Excel2003 形式	175,104 バイト
500KB～1MB	Excel2003 形式	691,712 バイト
1MB 以上	Word2003 形式	1,148,928 バイト
10MB 以上	GZIP 圧縮形式	21,935,943 バイト

表 4-2 移管媒体・フォーマットごとのウイルスチェック処理時間(単位:秒)

媒体	FD	MO	CD-R	DVD-R
100KB 未満	6	1 秒未満	1 秒未満	1 秒未満
100～500KB	8	1 秒未満	1 秒未満	1 秒未満
500KB～1MB	23	1 秒未満	1	1
1MB 以上	32	1	2	2
10MB 以上	-	1 秒未満	1	1 秒未満

FD を除く媒体では、媒体の違いによる差は余り見られなかった。10MB 以上のファイルがそれ以下のファイルサイズの場合よりも処理時間が短いのは、圧縮された個別のファイルのチェックが省略されているためと考えられる。本項における検証趣旨からは外れるが、圧縮ファイルの場合、パスワードにより暗号化されている場合もあること、移管時にどのような内容のファイルが含まれているのか確認しづらいことから、受入対象のフォーマットとするか検討が必要である。

(4) 移管媒体からのコピー処理結果

プロトタイプシステム上で、移管時媒体からサーバ上にファイルをコピーする処理にかかる時間を、想定される移管媒体の種類ごとに計測した結果を表 4-3 に示す。対象としたファイルは表 4-1 と同じである。

表 4-3 移管時媒体からのコピー処理時間(単位:秒)

媒体	FD	MO	CD-R	DVD-R
100KB 未満	1 秒未満	1 秒未満	1 秒未満	1 秒未満
100～500KB	9	1 秒未満	1 秒未満	1 秒未満
500KB～1MB	22	1 秒未満	2	1 秒未満
1MB 以上	34	1 秒未満	2	1
10MB 以上	-	9	17	7

媒体からのコピーについては、各媒体のドライブの読み書き速度に応じた結果となった。

本システムの運用を想定した場合、FDについてはデータ容量の少なさとそれによるシステム運用者の作業量の増加が懸念されるため、受入対象とはするが極力使用しないことが望ましい。

## 4.3. フォーマット変換機能実証結果報告

### 4.3.1. プロトタイプシステムにおける機能概要

#### (1) フォーマット変換機能概要

検済データフォルダのオリジナルフォーマットからファイル形式を判別し、標準フォーマットの場合は長期保存フォーマットへの変換を行う。標準フォーマットでない場合はオリジナルフォーマットをそのまま使用する。

文書フォーマット、表計算フォーマット、プレゼンテーションフォーマットの場合はテキストデータ抽出を行う。

オリジナルフォーマットのファイルのプロパティ情報から、技術的メタデータの自動生成を行う。

行政文書ファイル管理簿のデータから記録管理メタデータの自動生成を行う。

オリジナルフォーマット、長期保存フォーマット、技術的メタデータ、記録管理メタデータ、抽出テキストデータ、媒体データ(移管時媒体またはフォルダに含まれている全ての電子ファイルを、フォルダ構成も含めて ZIP 形式で圧縮したもの)を変換済データフォルダに移動する。

変換済データフォルダのファイルをコピー用媒体に移動し、アーカイバルメタデータ作成機能に連携する。

#### (2) 長期保存フォーマットへの変換

##### (ア) PDF/A への変換について

一般的に、アプリケーション上で文書作成、表計算、プレゼンテーション等のフォーマットから PDF へ変換を行う方式としては

- ① 各フォーマットに対応したアプリケーションが PostScript 形式のデータを生成し、PostScript 形式のデータを PDF に変換する
  - ② 各アプリケーションの印刷機能で、PDF 生成用のプリンタドライバを指定して PDF に出力する
  - ③ プリンタドライバを経由せず、各アプリケーションが直接 PDF データを出力する
- の 3 通りの方式がある。

各種フォーマットから PDF への変換をサーバ側で自動的に行うシステムを考える場合、最も望ましい方式は③である。しかし、③の方式は、各アプリケーション側の実装に依存するものであり、またフォーマットの仕様が公開されていない場合はアプリケーションベンダーにしか内部構造は分からず、他のベンダーでは該当フォーマットのデータから PDF へ変換するロジックを作ることすら困難となる(リバースエンジニアリングによりデータを解析するという方法もあるが、法律上の問題が含まれる場合がある)。

①、②は Adobe Acrobat 等多くの PDF 生成ツールでも使われている方法である。ただし、アプリケーションの印刷を呼び出す必要があり、その制御がサーバ側の処理にはあまり適していない(アプリケーション自体が外部からの制御に対応している必要がある)こと、大量のファイルを PDF に変換する場合にアプリケーション起動のオーバーヘッドにより処理時間がかかることがデメリットとなる。

##### (イ) プロトタイプシステムにおける PDF/A への変換の実現方式

プロトタイプシステムの構築及び検証にあたっては、上記の③については時間的な制約から実現が困

難であると判断し、Adobe Acrobat 9 Standard を印刷ドライバとして設定し、②の方式によりサーバ上で PDF/A への変換を実現することとした。

(3) フォーマット変換ソフトの機能

プロトタイプシステムでは、フォーマット変換を実現するために下記のソフトウェアを使用した。各ソフトウェアのバージョン、処理方針についても示す。

(ア) 長期保存フォーマット(PDF/A)への変換

アプリケーション(バージョン)	Adobe Acrobat 9 Standard
変換元ファイルフォーマット	文書作成、表計算、プレゼンテーション
処理方針	変換元フォーマットの各ファイルを、対応アプリケーションの印刷機能呼び出し、印刷ドライバを Acrobat とすることで PDF/A を出力する。
OS	Windows Server 2003
備考	Adobe Acrobat 9 Standard の印刷ドライバを利用し、コマンドライン(WSH: Windows スクリプティングホスト)からアプリケーションを操作して PDF/A 印刷を実行する。

(イ) 長期保存フォーマット(JPEG 2000)への変換

アプリケーション(バージョン)	imagemagick(6.4.8-1)
変換元ファイルフォーマット	画像
処理方針	コマンドラインから変換を実行する。
OS	Windows Server 2003
備考	JPEG 2000 の変換・圧縮パラメータは、設定ファイルにより定義できることとする。

(ウ) 長期保存フォーマット(MP3)への変換

アプリケーション 1(バージョン)	Wavext(11.0.3)
アプリケーション 2(バージョン)	lame(3.97)
変換元ファイルフォーマット	音楽・音声

処理方針	<ul style="list-style-type: none"> <li>・wma ファイルの場合 <ul style="list-style-type: none"> <li>① wma ファイルを「Wavext」を利用し、wav ファイルに変換する。</li> <li>② wav ファイルを「lame」を利用し、mp3 ファイルに変換する。</li> </ul> </li> <li>・wav ファイルの場合 <ul style="list-style-type: none"> <li>① wav ファイルを「lame」を利用し、mp3 ファイルに変換する。</li> </ul> </li> </ul>
OS	Windows Server 2003
備考	<p>WMA 形式についてはアプリケーション 1 により一度 WAV 形式に変換してから MP3 形式への変換を行う。</p> <p>MP3 への変換パラメータは、設定ファイルにより定義できることとする。</p>

(エ) 長期保存フォーマット(MPEG2)への変換

アプリケーション(バージョン)	FFMpeg (Revision 15986)
変換元ファイルフォーマット	映像
処理方針	コマンドラインから実行する。
OS	Windows Server 2003
備考	MPEG2 への変換パラメータは、設定ファイルにより定義できることとする。

(4) フォーマット変換ソフトの対応フォーマット

上記のソフトウェアについて、対応する入出力フォーマットを以下に示す。

(ア) Adobe Acrobat 9 Standard

入力フォーマット	※プロトタイプシステムでは、プリンタドライバとして使用するためアプリケーションに印刷機能が実装されていれば入力とすることが可能。
出力フォーマット	PDF 1.3~1.7 (設定による) PDF/A-1b

※アプリケーションとして対応する入出力フォーマットの詳細は Adobe Systems 社サイト ([http://www.adobe.com/jp/support/kb/cs/7/cs\\_7671\\_ja-jp.html](http://www.adobe.com/jp/support/kb/cs/7/cs_7671_ja-jp.html))を参照。

(イ) imagemagick

入力フォーマット	JPEG JPEG 2000*1 GIF TIFF*2 BMP その他
出力フォーマット	JPEG JPEG 2000*1 GIF TIFF*2 BMP その他

\*1) Jasper ライブラリが必要となる。

\*2) libtiff ライブラリが必要となる。

※上記以外にアプリケーションとして対応する入出力フォーマットの詳細は `imagemagick` サイト (<http://www.imagemagick.org/www/formats.html>)を参照。

(ウ) wavext

入力フォーマット	WAV MP3 WMA
出力フォーマット	WAV

(エ) lame

入力フォーマット	WAV
出力フォーマット	MP3

(オ) FFMpeg

入力フォーマット	QuickTime(.mov) Windows Media(.wmv、.avi) RealPlayer(.rm) MPEG(.mpg)
出力フォーマット	MPEG2

※上記以外にアプリケーションとして対応する入出力フォーマットの詳細は「`ffmpeg.exe -formats`」コマンドを実行することで確認できる。

## 4.3.2. フォーマット変換検証結果

### (1) 長期保存フォーマットへの変換処理時間

プロトタイプシステム上で、上記の標準フォーマットのファイルから長期保存フォーマットへの変換を行い、フォーマット及びファイルサイズごとにかかる処理時間、変換後のファイルサイズの検証を行った。

なお、以下の表では「-」は該当ファイルなし、「N/A」は変換エラー等により計測不可であったことを表す。

#### (ア) 文書作成フォーマット

##### ① OASYS

OASYS フォーマット(拡張子 OA2)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	2,725	7.141	5,408
100～500KB	102,226	13.016	85,640
500KB～1MB	-	-	-
1MB 以上	-	-	-

##### ② 一太郎

一太郎フォーマット(拡張子 JTD)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	20,992	33.500	5,547
100～500KB	275,456	9.109	721,145
500KB～1MB	664,064	7.922	1,055,101
1MB 以上	1,578,496	37.922	2,333,290

##### ③ Word2003

Word2003 フォーマット(拡張子 DOC)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	32,256	1.875	11,365
100～500KB	273,408	2.141	257,368
500KB～1MB	662,016	10.297	1,120,122
1MB 以上	-	-	-

④ Word2007

Word2007 フォーマット(拡張子 DOCX)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	13,091	6.109	6,809
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

⑤ Word2007(マクロ有効)

Word2007(マクロ有効)フォーマット(拡張子 DOCM)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。今回は変換に失敗しているが、これはフォーマット変換機能から印刷制御をする際に、保存の確認ダイアログが表示されたためにアプリケーションの終了を判断できなかったことが原因である。本形式の場合に必ずこの現象が発生するわけではないため、印刷処理を行うことにより作成したファイルの情報が変更されたと Word アプリケーション側で判断していると考えられるが、今回のプロトタイプシステムと同様の方式の場合は終了時のダイアログ判定等の処理が必要になる。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	39,628	N/A	N/A
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

⑥ PDF

PDF フォーマット(拡張子 PDF)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。100KB 未満のファイルの変換後のサイズが 30 倍近くになっているが、これは元の PDF がスキャンされた白黒 2 値画像であったものが、変換後にグレースケールとなったためにファイルサイズが増大していると考えられる。また、元ファイルではテキストとして選択できた文字列が変換後は画像データとなっていたが、今回の調査においては設定等により回避できるかは分からなかった。画像の色変換設定も含め、今後の検討課題として挙げておく。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	46,960	41.250	1,345,246
100～500KB	256,016	5.875	526,335
500KB～1MB	653,050	7.922	1,055,101
1MB 以上	1,594,112	37.922	2,333,290

⑦ OpenOffice

OpenOffice フォーマット(拡張子 SXW)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	10,146	2.250	5,664
100～500KB	268,676	3.094	251,713
500KB～1MB	668,207	4.203	1,043,720
1MB 以上	-	-	-

⑧ OpenOffice(ODF)

OpenOffice(ODF)フォーマット(拡張子 ODT)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。拡張子 SXW の場合とほぼ同様の結果となっている。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	10,321	2.703	5,664
100～500KB	268,809	2.484	251,713
500KB～1MB	668,337	4.641	1,043,720
1MB 以上	-	-	-

OpenOffice では、1.x フォーマットと ODF ではファイルサイズでは差異が見られず、変換時間についても同程度であった。

(イ) 表計算フォーマット

① Excel2003

Excel2003 フォーマット(拡張子 XLS)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

100KB 以上のファイルで長期保存のファイルサイズに差が見られないのは、複数シートのうち最後のシートのみしか出力されていないためである。複数シートであっても全シートが出力される場合もあり、フォーマット変換機能から印刷制御を行った場合に Acrobat 側の処理がシート何枚かをまとめてファイルに出力する、という処理を繰り返し実行していることが原因と考えられる。Excel の場合は 1 シートずつを個別に PDF/A ファイルとして出力し、全シートの出力が完了した段階で 1 ファイルに結合する、といった処理が必要になる。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	49,664	5.875	6,277
100～500KB	175,104	22.594	443,926

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
500KB～1MB	691,712	79.969	472,691
1MB 以上	1,293,824	155.813	472,686

## ② Excel2007

Excel2007 フォーマット(拡張子 XLSX)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	30,797	3.750	620,009
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

## ③ Excel2007(マクロ有効)

Excel2007(マクロ有効)フォーマット(拡張子 XLSM)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。変換失敗となっているが、これは Word2007(マクロ有効)の場合と同様の問題が発生したためである。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	36,743	N/A	N/A
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

## ④ OpenOffice

OpenOffice フォーマット(拡張子 SXC)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	29,371	4.000	95,919
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

## ⑤ OpenOffice(ODF)

OpenOffice(ODF)フォーマット(拡張子 ODT)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	29,715	4.781	95,913
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

OpenOffice では Writer の場合と同様に、1.x フォーマットと ODF ではファイルサイズでは差異が見られず、変換時間についても同程度であった。

#### (ウ) プレゼンテーションフォーマット

##### ① PowerPoint2003

PowerPoint2003 フォーマット(拡張子 PPT)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	84,480	8.844	298,921
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

##### ② PowerPoint2007

PowerPoint2007 フォーマット(拡張子 PPTX)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	119,564	8.578	300,548
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

##### ③ PowerPoint2007(マクロ有効)

PowerPoint2007 フォーマット(拡張子 PPTM)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	125,072	5.813	300,538
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

#### ④ OpenOffice

OpenOffice フォーマット(拡張子 SXI)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	33,756	6.250	307,350
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

#### ⑤ OpenOffice(ODF)

OpenOffice フォーマット(拡張子 ODP)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	39,369	5.688	341,517
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

### (エ) 画像フォーマット

#### ① JPEG

JPEG フォーマット(拡張子 JPG)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	77,425	0.391	48,706
100～500KB	130,149	0.375	168,128
500KB～1MB	-	-	-
1MB 以上	-	-	-

#### ② JPEG 2000

JPEG 2000 フォーマット(拡張子 JP2)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	48,706	0.422	190,934

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100～500KB	183,790	0.641	183,793
500KB～1MB	-	-	-
1MB 以上	-	-	-

### ③ GIF

GIF フォーマット(拡張子 GIF)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	86,733	0.422	207,678
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

### ④ TIFF

TIFF フォーマット(拡張子 TIF)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	-	-	-
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	1,350,008	0.328	183,847

### ⑤ BMP

BMP フォーマット(拡張子 BMP)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	-	-	-
100～500KB	-	-	-
500KB～1MB	766,554	0.359	352,005
1MB 以上	1,346,456	0.313	183,847

## (オ) 音声・音楽フォーマット

### ① WAVE

WAVE フォーマット(拡張子 WAV)からの長期保存フォーマットへの変換処理時間、変換後の長期保存

フォーマットファイルのサイズは以下の通りとなった。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	-	-	-
100～500KB	291,346	0.141	26,957
500KB～1MB	-	-	-
1MB 以上	-	-	-

## ② MP3

MP3フォーマット(拡張子 MP3)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

100KB 未満のファイルは元ファイルが 128Kbps だったため、長期保存(256Kbps)への変換によりおおよそ倍のサイズとなった。1MB 以上のファイルは元ファイルが 32Kbps だったが、256Kbps への変換ではエラーとなったため、32Kbps のままとしたためファイルサイズにはほとんど差が見られない。エラーとなった原因は、元ファイルのサンプリング周波数が 11.025kHz(MPEG-2.5)であり、256Kbps に対応していないためと考えられる。長期保存を 256Kbps 以上とする場合、サンプリング周波数を 32kHz 以上とする必要がある。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	53,499	0.422	109,504
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	1,472,640	6.625	1,477,484

## ③ WMA

WMAフォーマット(拡張子 WMA)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。第1段階の WMA から WAV への変換でエラーが発生しているが、正常に変換できるファイルもある。両者のファイルを比較したところ、オーディオコーデックのバージョンが異なっていた。今回のプロトタイプシステムでは Windows Media Audio 9.1 には対応できなかった。標準フォーマットとする形式については、対象とするオーディオコーデックのバージョンも明確にしておく必要がある。

エラー: Windows Media Audio 9.1 48 kbps, 44 kHz, mono 1-pass CBR

成功: Windows Media Audio V8 128 kbps, 44 kHz, stereo

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	27,747	N/A	N/A
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	-	-	-

## (カ) 映像フォーマット

### ① QuickTime

QuickTime フォーマット(拡張子 MOV)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

ただし、元ファイルのビデオコーデックが MPEG-4 で、フレームレートが 1/10 fps の場合は、MPEG-2 で 1/10 fps がサポートされていないためエラーとなるケースがあった。標準フォーマットとする形式については、ビデオコーデック及び fps 値も明確にしておく必要がある。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	-	-	-
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	2,892,349	2.045	3,582,309

### ② Windows Media

Windows Media フォーマット(拡張子 WMV)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。

100KB 未満のファイルでは変換エラーとなったが、これは音声(WMA)と同様に、コーデックのバージョンが「Windows Media Video 9」となっているためと考えられる。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	113,878	N/A	N/A
100～500KB	-	-	-
500KB～1MB	-	-	-
1MB 以上	6,197,721	8.281	17,115,136
10MB 以上	20,846,137	25.125	29,102,080

### ③ RealPlayer

RealPlayer フォーマット(拡張子 RM)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。いずれの場合も変換エラーとなったが、100KB 未満のファイルについては QuickTime と同様にフレームレートの問題があったため、500KB～1MB のファイルについては拡張子が「.rm」であるが音声しか含まれていないためである。処理においては拡張子だけでなく、音声のみか映像も含まれるかをチェックすることが必要と考えられる。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	17,846	N/A	N/A
100～500KB	-	-	-

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
500KB～1MB	589,708	N/A	N/A
1MB 以上	-	-	-

#### ④ MPEG-2

MPEG-2 フォーマット(拡張子 RM)からの長期保存フォーマットへの変換処理時間、変換後の長期保存フォーマットファイルのサイズは以下の通りとなった。いずれの場合も変換エラーとなったが、100KB 未満のファイルについては QuickTime と同様にフレームレートの問題があったため、500KB～1MB のファイルについては拡張子が「.rm」であるが音声しか含まれていないためである。処理においては拡張子だけでなく、音声のみか映像も含まれるかをチェックすることが必要となる。

サイズ分類	元ファイルサイズ(バイト)	変換時間(秒)	長期保存ファイルサイズ(バイト)
100KB 未満	-	-	-
100～500KB	-	-	-
500KB～1MB	527,548	0.210	196,608
1MB 以上	-	-	-

(2) 変換の適正状況

上記の標準フォーマット及び変換された長期保存フォーマットの電子ファイルを対応アプリケーションで開き、変換が適正に行われているかの検証を行った。

各フォーマットにおける変換の適性状況評価の観点は以下の通りである。「○」の機能等を各フォーマットでの検証対象とした。

検証機能等観点	文書作成	表計算	プレゼンテーション
フォント	○	○	○
版情報	○	-	-
ヘッダー、フッター	○	○	○
計算式・関数	-	○	-
複数シートの出力	-	○	-
オブジェクトの動作設定	-	-	○
アニメーション設定	-	-	○
変更履歴	○	○	○
コメント	○	○	○
オブジェクト埋め込み	○	○	○
ハイパーリンク	○	○	○
マクロ(フォーム部品)	○	○	○

検証機能等観点	画像	音声・音楽	映像
ビットレート	-	○	○
チャンネル数	-	○	-
コピーガード/著作権保護	-	○	○
透過	○	-	-
色数	○	-	-
マルチページ	○	-	-
インターレース	○	-	-

閲覧に使用したアプリケーションは以下の通りである。OS はすべて Windows XP(SP2)である。

【標準フォーマット】

フォーマット分類	フォーマット	閲覧アプリケーションまたはプラグイン
文書作成フォーマット	OASYS	OASYS V10
	一太郎 8-12	一太郎 2008
	Word97-2003	Microsoft Word 2003
	Word2007	Microsoft Word 2007
	PDF1.x、PDF/A	Adobe Reader 9
	OpenOffice Writer	OpenOffice Writer 3

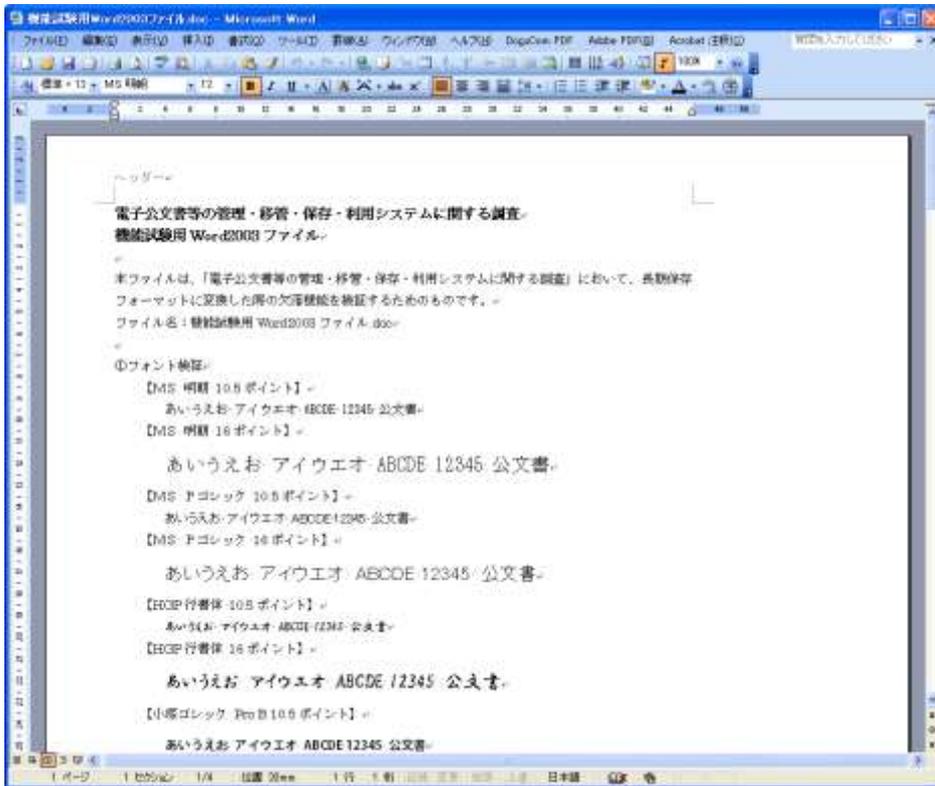
フォーマット分類	フォーマット	閲覧アプリケーションまたはプラグイン
表計算フォーマット	Excel97-2003	Microsoft Excel 2003
	Excel2007	Microsoft Excel 2007
	OpenOffice Calc	OpenOffice Calc 3
プレゼンテーションフォーマット	PowerPoint97-2003	Microsoft PowerPoint 2003
	PowerPoint2007	Microsoft PowerPoint 2007
	OpenOffice Impress	OpenOffice Impress 3
画像	JPEG	Microsoft Office Picture Manager
	JPEG 2000	LizardTech Express View Browser Plug-in 4
	GIF	Microsoft Office Picture Manager
	TIFF	Windows Picture and Fax Viewer
	BMP	Microsoft Office Picture Manager
音声・音楽	WAVE	Windows Media Player 11
	MP3	Windows Media Player 11
	WMA	Windows Media Player 11
映像	QuickTime	QuickTime Player
	Windows Media Video	Windows Media Player 11
	RealPlayer	Real Player
	MPEG	Windows Media Player 11

【長期保存フォーマット】

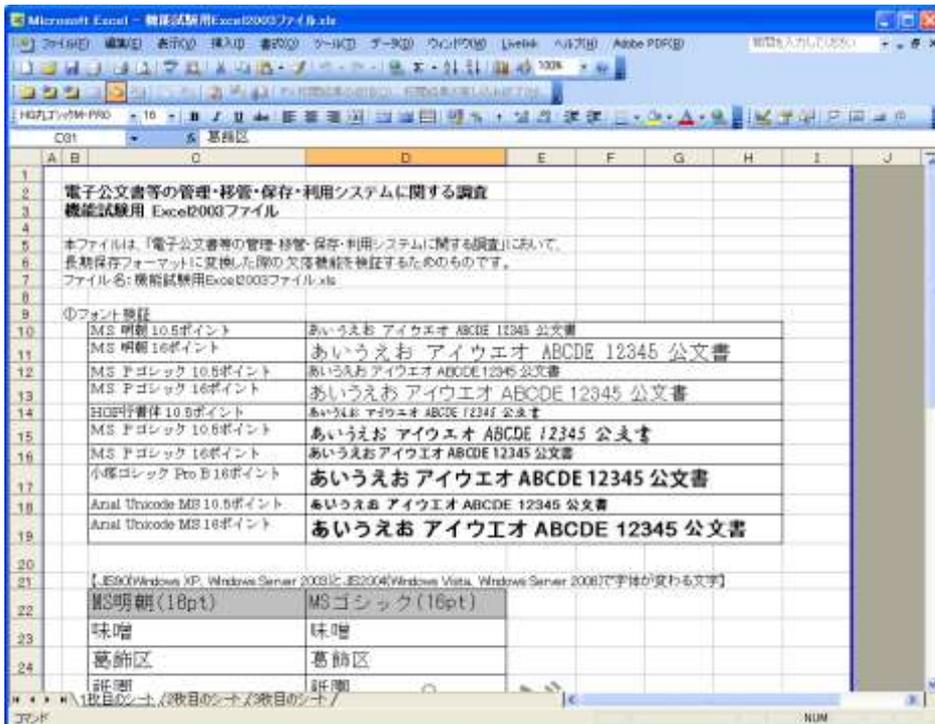
フォーマット	フォーマット	対応アプリケーションまたはプラグイン
長期保存フォーマット	PDF/A	Adobe Reader 9

検証のために作成したサンプルファイル(文書作成、表計算、プレゼンテーション)の例を以下に示す。

① Word2003 形式



② Excel2003 形式





(ア) 文書作成フォーマットでの検証結果

検証機能	文書作成フォーマット					
	OASYS	一太郎	Word2003	Word2007	PDF	OpenOffice
フォント	○	△	△	△	○	×
版情報	-	※1	×	※1	-	×
ヘッダー、フッター	-	○	○	○	-	○
変更履歴、コメント	-	×	×	×	-	×
オブジェクト埋め込み	-	×	×	×	-	×
ハイパーリンク	-	×	×	×	×	×
マクロ(フォーム部品)	-	×	×	×	-	×
外字	-	×	×	×	-	×
レイアウト	○	○	○	×	×	○

※1 一太郎、Word2007 には版情報機能を確認できなかった。

フォントについて、一太郎、Word2003、Word2007 では微妙な字体の差異があった。

変更履歴、コメントは機能としては欠落するものの、元ファイル上で残ったままの場合、印刷時のイメージとして PDF/A に出力されてしまうため、移管時には注意が必要である。Word2007 の場合、コメント箇所が色違いで表示されるものの、コメント内容は PDF/A には出力されなかった。

オブジェクト埋め込みはテキスト及び画像として出力され、元のオブジェクトの再現や動作は実行できなかった。

ハイパーリンクの部分は単なる文字列として出力されるが、PDF では http://」で始まる文字列の部分がリンクとして処理されている。これは変換時または閲覧ソフトウェア上で自動的に判断して制御を行っている可能性が高い。

マクロ、フォーム部品は画像としての表示となり、マクロを動作させることはできない。

外字についてはいずれのフォーマットにおいても、作成環境以外では表示できなかった。

レイアウトは、Word2007において、2つのオブジェクトが重なる場合に一方のオブジェクトの位置が大きくなるという例があった。また、PDF でページごとに縦・横のレイアウトが設定されていても、PDF/A に出力されたときにすべて同じ向きとなってしまった。

(イ) 表計算フォーマットでの検証結果

検証機能	表計算フォーマット		
	Excel2003	Excel2007	OpenOffice
フォント	○	○	△
計算式・関数	×	×	×
複数シート	△	△	○
ヘッダー、フッター	○	○	○
コメント	×	×	-
オブジェクト埋め込み	×	×	-
ハイパーリンク	×	×	×
マクロ(フォーム部品)	×	×	×
外字	×	×	×
レイアウト	×	×	×

フォントについて、OpenOffice では一部フォントの字体(文字の太さ)において見た目で見え方が分かる程度の差異があった。また、また、文書作成フォーマットと同様に微妙な字体の差異があった。

計算式・関数は結果としての値は出力されるのみで、元の計算式や関数は欠落する。

複数シートについては Excel の場合に全シートが出力されないケースがあった。これは前述の通り、変換時の制御に問題があるためと考えられる。

オブジェクト埋め込みは文書作成フォーマットと同様に、テキスト及び画像として出力され、元のオブジェクトの再現や動作は実行できなかった。ハイパーリンク、マクロ、フォーム部品、外字についても文書作成と同様となった。

レイアウトについては、いずれのフォーマットにおいても、印刷レイアウトでセル内の文字が切れてしまう場合に、PDF/A に出力したときに同様の文字欠けが発生した。また、この状態でテキスト抽出を行うと欠けた文字が出力されない。

改ページが設定されていない場合も、改行位置が自動的に設定されてしまうため、PDF/A での出力イメージが大幅に異なる(ページ数が大幅に増える)ケースが見られた。

(ウ) プレゼンテーションフォーマットでの検証結果

検証機能	プレゼンテーションフォーマット		
	PowerPoint2003	PowerPoint2007	OpenOffice
フォント	○	○	×
オブジェクトの動作設定	×	×	×
アニメーション設定	×	×	×
ヘッダー、フッター	○	○	○
コメント	×	×	※2
オブジェクト埋め込み	×	×	×
ハイパーリンク	×	×	×
マクロ(フォーム部品)	×	×	×
外字	×	×	×
レイアウト	△	△	△

※2 OpenOffice Impress ではコメント機能を確認できなかった。

フォントについて、OpenOffice では一部フォントの字体(文字の太さ)において見た目で見え方が分かる程度の差異があった。また、文書作成フォーマットと同様に微妙な字体の差異があった。

オブジェクトの動作設定、アニメーション設定は動作完了時の状態で出力されるのみで、機能としては再現されない。

コメントについては文書作成フォーマットと同様に PDF/A に出力されてしまうため、注意が必要である。

オブジェクト埋め込み、ハイパーリンク、マクロ、フォーム部品、外字は文書作成と同様の結果となった。

レイアウトについては、印刷イメージとなるため作成アプリケーションで閲覧した場合と比べて周囲に余白が設定されてしまう。

(エ) 画像フォーマットでの検証結果

画像フォーマットについて、閲覧アプリケーションを用いて検証した結果は以下の通りである。

変換前フォーマット	①BMP	②BMP
変換結果	成功	成功
再生状況	○	○
欠落機能	なし	なし

変換前フォーマット	③GIF(透過)	④GIF(透過)	⑤GIF(透過、インターレース)
変換結果	成功	成功	成功
再生状況	○	○	○
欠落機能	なし	なし	なし

変換前フォーマット	⑥JPG	⑦JPG	⑧JPG(プログレッシブ)
変換結果	成功	成功	成功
再生状況	○	○	○
欠落機能	なし	なし	なし

変換前フォーマット	⑨TIF(LZW 圧縮)	⑩TIF	⑪TIF(マルチページ)
変換結果	失敗	成功	成功(※)
再生状況	—	○	△
欠落機能	—	なし	なし

変換前フォーマット	⑫JPEG2000
変換結果	成功
再生状況	○
欠落機能	なし

長期保存フォーマット(JPEG 2000)への変換においては、LZW で圧縮された TIFF の場合に変換できなかった。また、マルチページ TIFF については変換は可能なものの、1 ページが 1 ファイルとして出力されるため、プロトタイプシステムでは扱うことができなかった。マルチページ TIFF の場合は PDF として移管されることが望ましい。

なお、画質については見た目上の違いはほとんど見られなかった。

(オ) 音声・音楽フォーマットでの検証結果

音声・音楽フォーマットについて、閲覧アプリケーションを用いて検証した結果は以下の通りである。

変換前フォーマット	①MP3	②MP3	③WAVE	④WMA
変換結果	成功	成功	成功	失敗
再生状況	○	全体的に聞き 苦しい	○	—
欠落機能	ビットレート	なし	ビットレート	—

①及び③のビットレートは半分以下になった。②はビットレートが同じであるが、実際に聞いてみたときに音質の劣化が確認された。

WMA は変換に失敗しているが、原因については 4.3.2 (1)(オ)で述べた通りである。

音声・音楽の変換については内部のコーデックやパラメータ等、さらに考慮する必要があるといえる。

(カ) 映像フォーマットでの検証結果

映像フォーマットについて、閲覧アプリケーションを用いて検証した結果は以下の通りである。

変換前フォーマット	①MPEG	②MOV	③RM
変換結果	失敗	失敗	失敗
再生状況	—	—	—
欠落機能	—	—	—

変換前フォーマット	④WMV	⑤WMV	⑥WMV
変換結果	成功	失敗	成功
再生状況	○	—	△
欠落機能	なし	—	音声は問題ないが、ブロックの残像が絶えず残る

①～③、⑤ではいずれも変換に失敗しているが、原因については 4.3.2 (1)(カ)で述べた通りである。成功した映像はいずれもオリジナルフォーマットよりもビットレートが高い数値となっていた。また、⑥では映像にブロックノイズが多発しており、画質上の問題が見られた。映像についても内部のコーデックやパラメータ等、さらに考慮する必要があるといえる。

### (3) テキストデータ抽出

プロトタイプシステムでは、長期保存フォーマットのうち PDF/A に変換されたものについて、テキスト抽出処理を行う仕組みとした。

抽出されたテキストデータの出力状況は、おおよそテキスト情報として格納されている部分が出力されていたが、文字列が重複して出力されたり、不自然な位置に改行が挿入される(英単語の途中で改行される)というケースが見られた。このようなケースでは、行政利用の検索機能に影響を及ぼす可能性が高いため、テキストデータの正規化処理について検討が必要であると思われる。

なお、ファイルサイズごとのテキスト抽出処理時間を測定した結果は以下の通りであった。

サイズ分類	長期保存ファイルサイズ (バイト)	変換時間(秒)	テキストファイルサイズ (バイト)
100KB 未満	95,913	1.000	9,401
100～500KB	300,548	0.422	5,500
500KB～1MB	544,547	0.891	104,375
1MB 以上	1,055,101	0.297	13,723

### (4) 外字の扱い

移管時の電子ファイルに、ローカル環境で作成した外字が含まれている場合、フォーマット変換時に該当文字は正しく表示されない文字に変換されてしまう。

上:ローカル環境でのオリジナルフォーマット(Word2003)で表示される外字

下:サーバ上で長期保存フォーマット(PDF/A)に変換されたファイルを開いたときに表示される外字

Windows XP の「外字エディタ」で作成した外字→𠄎

Windows XP の「外字エディタ」で作成した外字→・

また、テキストデータ抽出時に抽出ソフトウェアによっては変換エラーや文字化けが発生する可能性がある。

外字については JIS 化に向けて検討されている「文字図形識別番号」を用いて、作成時に共通のコードとして埋め込まれることが望ましいが、現状では各府省で独自の外字取り扱い方針を定めている場合や、方針が決まっていない状況であるため、技術面・制度面の両方で外字に関する統一的な環境が整備されることが期待される。

### 4.3.3. フォーマット変換のまとめ

プロトタイプシステムにおける検証結果を以下にまとめる。

#### (1) オリジナルフォーマットからの欠落機能

- オフィス系フォーマットでは、マクロやフォームといった機能が使えなくなる(再現できない)。
- ハイパーリンクの位置がオリジナルと異なる場所になる場合がある。ただし、これは PDF/A 変換用(または閲覧用の)ソフトウェア設定で制御される部分でもあり、オリジナルで意図していたリンク表示が再現されない。
- アプリケーションとともにインストールされるフォントなどは、完全に同一のフォントでは再現されない。PDF/A 変換時にフォント埋め込みを設定することで再現できる可能性はあるが、デジタルデータとして埋め込まれるフォントの著作権の問題がある。
- 外字が再現されない。
- レイアウトが崩れる場合がある。
- 音声・音楽、映像については対応コーデックや、画質や音質を維持するためのパラメータを今後検討する必要がある。RealPlayer(RealMedia)形式は、今回の調査でも変換に失敗しており、実運用においても対応が難しい可能性がある。

#### (2) 移管時に留意すべき事項

- PDF/A については印刷イメージで変換されるため、コメントや変更履歴など、意図しない情報が出力される可能性がある。移管時点でこれらの情報を削除する、というルールや運用方針を定めておく必要がある。
- 表計算では、印刷範囲を適切に設定していない場合に正しいレイアウトとならない。また、印刷イメージとして PDF/A に変換されるため、セル内での文字切れにも注意する必要がある。
- セキュリティ機能(パスワードによる保護)が有効になっている場合、長期保存への変換が行えない。また、オリジナルフォーマットで保存しても、閲覧時にパスワードを求められるため、作成者以外に該当ファイルを開くことがほぼ不可能となる。移管時にはセキュリティ設定を解除しておく必要がある。
- 作成時の環境と閲覧環境(OS、アプリケーション、インストールされているフォントセットのバージョン)が異なる場合に、同一文字でも表示される字体が変わる場合がある(JIS90 と JIS2004 の問題)。

## 4.4. アーカイバルメタデータ作成機能実証結果報告

### 4.4.1. メタデータの自動生成

#### (1) 記録管理メタデータ

##### (ア) 自動生成

記録管理メタデータの生成は、主に行政文書ファイル管理簿より抽出して自動的に作成、その他の項目はアーカイバルメタデータ編集画面上で入力する。行政文書ファイル管理簿より、記録管理メタデータとしては、「3.3.1 メタデータ項目の定義」のうち、No.2～15,21の項目を取得した。No.1「文書ID」は、検疫・媒体変換機能を実行する段階で、システム内でユニークとなるように値を生成し、設定している。

##### (イ) 自動生成された記録管理メタデータの検証

「3.3.4 記録管理メタデータ」に示した記録管理メタデータ項目のうち、今回のプロトタイプシステムで自動生成される項目について検証した結果を表 4-4 にまとめた。

表 4-4 自動生成された記録管理メタデータの検証結果

No.	項目名	データ出自と設定方法 *1	検証結果 (編集画面)	検証結果説明
1	文書 ID	検疫・媒体変換機能で生成	○	変更更新不可確認
2	タイトル	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
3	作成者	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
4	文書分類; 大分類	R(変更不可)	○	R と同一を表示 変更更新不可確認
5	文書分類; 中分類	R(変更不可)	○	R と同一を表示 変更更新不可確認
6	文書分類; 小分類	R(変更不可)	○	R と同一を表示 変更更新不可確認
7	移管元府省名	R(変更不可)	○	R と同一を表示 変更更新不可確認
8	保存期間	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
9	保存期間満了時期	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
10	保存場所	R(変更可)	○	R と同一を表示

No.	項目名	データ出自と設定方法 *1	検証結果 (編集画面)	検証結果説明
				変更更新時: 変更データを表示
11	保存時期満了時の措置結果	R(変更可)	○	Rと同一を表示 変更更新時: 変更データを表示
12	管理担当課・係	R(変更可)	○	Rと同一を表示 変更更新時: 変更データを表示
13	備考	R(変更可)	○	Rと同一を表示 変更更新時: 変更データを表示
15	行政文書ファイル管理簿番号	R(変更可)	○	Rと同一を表示 変更更新時: 変更データを表示
17	言語	初期値 JPN(変更可)	○	固定値 JPN を表示 変更更新時: 変更データ(現在は eng のみ)を表示
23	作成年月日	R(変更可)	○	Rと同一を表示 変更更新時: 変更データを表示

\*1 行政文書ファイル管理簿は R No.は「3.3.4 記録管理メタデータ」と対応している。

#### (ウ) 自動生成されない記録管理メタデータ項目

自動生成されない記録管理メタデータ項目を表 4-5 にまとめた。これらはプロトタイプシステムでは初期値なしとして扱った。

表 4-5 自動生成されない記録管理メタデータ項目

No.	項目名	理由
14	各府省個別情報	行政文書ファイル管理簿上には存在しないため。
16	受入文書番号	移管後に国立公文書館で付与することが想定されている値のため。
18	関連資料、参考文献	移管後に国立公文書館で付与することが想定されている値のため。
19	権利関係	移管後に国立公文書館で付与することが想定されている値のため。
20	公開に関する移管元の可否	移管元からの情報として想定されているが、行政文書ファイル管理簿には存在しないため。
21	公開に関する移管元の可否(否の理由)	移管元からの情報として想定されているが、行政文書ファイル管理簿には存在しないため。
22	文書分類	行政文書ファイル管理簿上には存在しないため。
24	元ファイル名	プロトタイプシステムでは該当項目が存在しないため。

No.は「3.3.4 記録管理メタデータ」と対応している。

(エ) 自動生成されない記録管理メタデータ項目の作成方法

「3.3.4 記録管理メタデータ」の項目のうち、No.22「文書分類」、No.24「元ファイル名」を除いた項目は、記録管理メタデータ編集画面にて入力・変更可能であることが確認できた。No.24 の元ファイル名は、今回のプロトタイプシステムでは1 媒体＝電子公文書 1 件としているため、この項目の対応データがない。

(2) 技術的メタデータ

(ア) 自動生成

技術的メタデータの生成は、検疫・媒体変換機能、フォーマット変換機能で電子ファイル単位のプロパティから抽出された情報と、検疫結果や長期保存フォーマットの形式、作成アプリケーション等の本システム上で判断可能である情報から自動生成して作成する。

(イ) 自動生成された技術的メタデータの検証

「3.3.5 技術的メタデータ」に示した技術的メタデータ項目のうち、今回のプロトタイプシステムで自動生成される項目について検証した結果を表 4-6 に示す。

表 4-6 自動生成された技術的メタデータの検証結果

No.	項目名	データ出自と設定方法	検証結果*1	検証結果説明
1	オブジェクト ID: 種別	自動設定	○	非表示 変更更新不可確認
2	オブジェクト ID: 値	検疫・媒体変換機能で生成	○	表示 変更更新不可確認
3	保存レベル: 値	ファイルのタイプにより自動 設定	○	非表示 変更更新不可確認
4	メッセージダイジ ェストアルゴリズム	固定値: MD5	○	表示 変更更新不可確認
5	メッセージダイジ ェスト	検疫・媒体変換機能、フォー ーマット変換機能で生成	○	表示 変更更新不可確認
6	サイズ	プロパティより取得	○	表示 変更更新不可確認
7	フォーマット名	拡張子から判断して設定	△	同一拡張子で形式が異なる場合(.doc 等)は判 断できない。 変更更新時: 変更データを表示
8	フォーマットパー ジョン	拡張子から判断して設定	△	同一拡張子でバージョンが異なる場合(PDF 等)は判断できない。 変更更新時: 変更データを表示
10	作成アプリケー ション名	拡張子から判断して設定 フォーマット変換時はシステ ム上のアプリケーション名を 設定	△	オリジナルフォーマットについては作成アプリ ケーション名を厳密に判断できない。 変更更新時: 変更データを表示
11	作成アプリケー ションバージョン	拡張子から判断して設定 フォーマット変換時はシステ	△	オリジナルフォーマットについては作成アプリ ケーション名を厳密に判断できない。

No.	項目名	データ出自と設定方法	検証結果*1	検証結果説明
		ム上のアプリケーションバージョンを設定		変更更新時: 変更データを表示
12	作成日付	プロパティより取得	○	表示 ただし、平成 19 年度までの調査結果より、作成日付の正確性については疑念が提起されている。 変更更新不可確認
15	オリジナルファイル名	プロパティより取得	○	表示 変更更新不可確認
16	ストレージ場所: 種別	固定値: storageMedium	○	表示 変更更新不可確認
17	ストレージ場所: 値	固定値	○	表示 変更更新不可確認
18	ストレージ: 媒体	システム側で判断して自動設定	○	表示 変更更新不可確認
20	ソフトウェア名	オリジナルフォーマットについては値無し フォーマット変換時はシステム上の OS 名を設定	△	オリジナルフォーマットの OS は判断できない。 変更更新時: 変更データを表示
21	ソフトウェアバージョン	オリジナルフォーマットについては値無し フォーマット変換時はシステム上の OS バージョンを設定	△	オリジナルフォーマットの OS は判断できない。 変更更新時: 変更データを表示
22	ソフトウェア種別	No.20 の内容により自動設定	△	非表示 変更更新不可確認

\*1 編集画面に表示しないものはコンテナメタデータ(XML)への出力を確認

#### (ウ) 自動生成されない技術的メタデータ項目

今回のプロトタイプシステムで自動生成されなかった技術的メタデータ項目を表 4-7 にまとめた。

No.13「制限種別」、14「制限対象」については、プロトタイプシステムでは行政利用機能におけるアクセス制限のための項目としているため、自動生成の対象外とした。

No.20、21、22(ファイルの作成された OS 情報)については、オリジナルフォーマットの場合は取得することができなかったが、長期保存フォーマットへの変換時はシステム側で自動的に設定できる。

表 4-7 自動生成されない技術的メタデータ項目

No.	項目名	理由
9	フォーマット注記	編集画面上での入力を想定しているため。

No.	項目名	理由
13	制限種別	編集画面上での入力を想定しているため。
14	制限対象	編集画面上での入力を想定しているため。
19	依存アイテム名	オリジナルフォーマット内で使用されているフォント情報を想定しているが、プロトタイプシステムではフォント情報を自動的に取得できなかった。
23	署名者	プロトタイプシステムでは自動取得の対象外とした。

(3) アーカイバルメタデータ

(ア) 自動生成

アーカイバルメタデータの生成は、記録管理メタデータから一部抽出して作成、その他の項目はシステム上で自動生成するか、アーカイバルメタデータ編集画面上で入力する。

行政文書ファイル管理簿(記録管理メタデータ)より、アーカイバルメタデータとしては、「3.3.6 アーカイバルメタデータ」のうち No.1～3、29～34 の項目を取得した。

(イ) 自動生成されたアーカイバルメタデータの検証

「3.3.6 アーカイバルメタデータ」に示したアーカイバルメタデータ項目のうち、今回のプロトタイプシステムで自動生成される項目について検証した結果を表 4-8 に示す。

表 4-8 自動生成されたアーカイバルメタデータの検証結果

No.	項目名	データ出自と設定方法 *1	検証結果 (編集画面)	検証結果説明 *1
1	タイトル	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
2	作成者	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
3	作成年月日	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
4	言語	初期値 JPN(変更可)	○	JPN を表示 変更更新時: 変更データ(現在は eng のみ)を表示
7	受入年月日	初期値は検疫・媒体変換 実行時の年月日(変更可)	○	初期値を表示 変更更新時: 変更データを表示
8	受入方法	初期値は 01(移管) (変更可)	○	初期値を表示 変更更新時: 変更データを表示
9	移管年度	初期値は編集時年度 (No.7 の値により自動設定)	○	247:** (元号コード 3 桁(平成=247)+「:」 十年 2 桁)を表示 変更更新時: 変更データ(平成 20 年なら 247:20)を表示
10	記述レベル	初期値は file (No.14 変更 で自動変更)	○	File を表示 変更更新時: 変更データを表示
12	レファレンス・コード	初期値は自動設定(関連 データ変更で自動変更)	○	No.15 + No.16 + No.17 を表示 変更更新時: No.10 が File の場合 No.9 変更により No.15 + No.16 + No.17

No.	項目名	データ出自と設定方法 *1	検証結果 (編集画面)	検証結果説明 *1
				を表示 No.10 が item の場合 No.9 変更により No.15 + No.16 + No.17 + “-” + No.19 + No.20 を表示
13	データ種別	初期値は 1(変更可)	○	初期値を表示 変更更新時: 変更データを表示
14	目録種別	初期値は 2(変更可)	○	初期値を表示 変更更新時: 変更データを表示
15	簿冊番号1	自動設定(移管年度、府 省庁略称は取得)	○	元号文字(「平」)+移管年度(全角数字 2 文字)+府省庁略称 を表示 変更更新時: No.9 変更時、変更データ を表示
16	簿冊番号2	自動設定(移管年度、移 管元府省は取得)	○	移管元府省・移管年度ごとに、1 から始 まる連番(5桁 0埋め) 変更更新時: No.9 変更時、変更データ を表示
17	簿冊番号枝番	初期値は 100(変更可)	○	初期値を表示 変更更新時: 変更データを表示
18	件名の表示	初期値は 0(変更可)	○	初期値を表示 変更更新時: 変更データを表示
21	公開状況	初期値 02(変更可)	○	初期値を表示 変更更新時: 変更データを表示
22	公開年月日	初期値は編集時年月日 (変更可)	○	初期値を表示 変更更新時: 変更データを表示
23	審査状況	初期値 01(変更可)	○	初期値を表示 変更更新時: 変更データを表示
24	審査年月日	初期値は編集時年月日	○	初期値を表示 変更更新時: 変更データを表示
27	文書 ID	記録管理メタデータより 取得	○	取得した ID を表示 変更更新不可確認
28	保存場所	R(変更可)	○	R と同一を表示 変更更新時: 変更データを表示
29	移管元府省名	R(変更不可)	○	R と同一を表示 変更更新不可確認
30	移管時管理担当 部局	R(変更可)	○	R と同一を表示(管理担当課・係と同じ 値) 変更更新時: 変更データを表示

No.	項目名	データ出自と設定方法 *1	検証結果 (編集画面)	検証結果説明 *1
31	媒体	R(変更可)	○	Rと同一を表示 変更更新時:変更データを表示
32	文書分類:大分類	R(変更不可)	○	Rと同一を表示 変更更新不可確認
33	文書分類:中分類	R(変更不可)	○	Rと同一を表示 変更更新不可確認
34	文書分類:小分類	R(変更不可)	○	Rと同一を表示 変更更新不可確認

No.は「3.3.6 アーカイバルメタデータ」と対応

\*1 行政文書ファイル管理簿は R

#### (ウ) 自動生成されないアーカイバルメタデータ項目

今回のプロトタイプシステムで自動生成されなかったアーカイバルメタデータ項目を表 4-9 にまとめた。

表 4-9 自動生成されないアーカイバルメタデータ項目

No.	項目名	理由
5	関連資料、参考文献	編集画面上での入力を想定しているため。
6	文書番号	編集画面上での入力を想定しているため。
11	資料内容	編集画面上での入力を想定しているため。
19	件名番号	プロトタイプシステムではすべて簿冊として扱うこととしているため。
20	件名番号 SEQ	プロトタイプシステムではすべて簿冊として扱うこととしているため。
25	保管履歴情報	マイグレーションやデータ更新により情報を格納することが想定される。プロトタイプシステムでは項目のみ定義している。
26	利用履歴情報	プロトタイプシステムでは項目のみ定義している。

#### (エ) 自動生成されないアーカイバルメタデータ項目の作成方法

表 4-9 のすべての項目について、アーカイバルメタデータ編集画面にて入力・変更可能であることを確認した。

No.1 のタイトルが簿冊なのか件名なのかの判断は、行政文書ファイル管理簿のレコードと電子公文書データの紐付けによって変わってくる。現状では機械的に判断することが困難であるため、プロトタイプシステムではタイトル=簿冊とした。

(4) コンテナメタデータ

(ア) 自動生成

コンテナメタデータは、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データを元に METS 形式の XML として出力される。

(イ) 自動生成されたコンテナメタデータの検証

「3.3.7 コンテナメタデータ」に示したコンテナメタデータ項目のうち、自動生成される項目について検証した結果を表 4-10 に示す。コンテナメタデータは他のメタデータからシステム上で生成するため、自動生成不可能な項目は存在しなかった。

表 4-10 自動生成されたコンテナメタデータの検証結果

No.	項目名	データ出自と設定方法 *1	検証結果	検証結果説明
1	コンテナメタデータ	自動設定	○	<mets>を表示
2	METS ヘッダ	自動設定	○	<metsHdr>を表示
3	コンテナメタデータ作成日時	コンテナメタデータを作成した日時を自動設定	○	<metsHdr CREATEDATE=" 作成年月日" >を表示
4	コンテナメタデータ最終更新日時	コンテナメタデータ最終更新日時を自動設定	○	<metsHdr CREATEDATE="最終更新年月日" >を表示
5	記述メタデータセクションヘッダ	自動設定	○	<dmdSec>を表示
6	記録管理メタデータ ID	「R」+文書 ID を設定	○	<dmdSec ID="R 文書 ID">を表示
7	メタデータ部	自動設定	○	<mdWrap>を表示
8	メタデータ種別	自動設定	○	<mdWrap @MDTYPE="OTHER">を表示
9	XML データ	自動設定	○	<xmlData> を表示
10	記述メタデータセクションヘッダ	自動設定	○	<dmdSec>を表示
11	アーカイバルメタデータ	「A」+文書 ID を設定	○	<dmdSec ID="A 文書 ID">を表示

No.	項目名	データ出自と設定方法 *1	検証結果	検証結果説明
	ID			
12	メタデータ部	自動設定	○	<mdWrap>を表示
13	メタデータ種別	自動設定	○	<mdWrap @MDTYPE="EAD">を表示
14	XML データ	自動設定	○	<xmlData> を表示
15	管理メタデータセクション	自動設定	○	<amdSec>を表示
16	技術的メタデータヘッダ	自動設定	○	<techMD>を表示
17	技術的メタデータID	技術的メタデータのファイルIDより設定	○	<techMD ID="*ファイルID">を表示 * オリジナルフォーマット=TO 長期保存フォーマット=TP 抽出テキスト=TE
18	メタデータ部	自動設定	○	<mdWrap>を表示
19	メタデータ種別	自動設定	○	<mdWrap @MDTYPE="PREMIS">を表示
20	XML データ	自動設定	○	<xmlData> を表示
21	ファイルセクション	自動設定	○	<fileSec> を表示
22	ファイルグループ	自動設定	○	<fileGrp> を表示
23	ファイル	自動設定	○	<file>を表示
24	ファイルID	技術的メタデータのファイルIDより設定	○	<file ID="*ファイルID">を表示 * オリジナルフォーマット=FO 長期保存フォーマット=FP 抽出テキスト=FE
25	ファイルの管理メタデータ	技術的メタデータのファイルIDを取得し自動設定	○	<file ADMID="*ファイルID">を表示 * オリジナルフォーマット=TO 長期保存フォーマット=TP 抽出テキスト=TE
26	ファイル場所	ファイルのストレージ値を取得し自動設定	○	<xlink:href="保存用ストレージの場所">を表示
27	場所タイプ	自動設定	○	"OTHER"を表示

No.	項目名	データ出自と設定方法 *1	検証結果	検証結果説明
28	媒体上のファイル構造	自動設定	○	<structMap>を表示
29	フォルダ構成	媒体上のファイル構成を取得し自動設定	○	<div TYPE="folder">を表示
30	フォルダ名	媒体上のフォルダ名を取得し自動設定	○	<div TYPE="folder" LABEL="フォルダ名">を表示
31	ファイル	ファイル ID を取得し自動設定	○	<fptr FILEID="該当ファイルのファイルID(No.24)">を表示

## 4.4.2. プロトタイプシステムにおける機能概要

### (1) 概要

フォーマット変換機能において作成されている文書 ID データを取り込み、登録用データフォルダに格納する。登録用データフォルダの各種メタデータ(CSV 形式)は RDBMS に格納する。登録用データフォルダに格納されたその他のデータを作業用フォルダにコピーする。RDBMS に格納された各メタデータは、編集画面にて入力または変更を行う。RDBMS への格納まではバッチ処理となる。

#### (ア) アーカイバルメタデータ自動作成処理

登録用データフォルダのメタデータ(記録管理メタデータ、技術的メタデータ)のファイルを読み込み、RDBMS に格納する。同時に各種電子ファイル(長期保存フォーマット・抽出テキストデータ・オリジナルフォーマット・媒体データ)を作業用データフォルダに格納する。

#### (イ) アーカイバルメタデータ編集処理

RDBMS に格納された記録管理メタデータ、アーカイバルメタデータ、技術的メタデータを画面にて編集する。

#### (ウ) 確定処理

アーカイバルメタデータ編集画面で確定を実行した場合、RDBMS の確定フラグを立てる。これによって以降の保存機能を実行できるようになる。

## (2) 画面説明

プロタイプシステムにおける、メタデータ編集の流れと各画面について以降に説明する。

### (ア) ログイン画面

認証可能なログイン ID とパスワードを入力すると、アーカイバルメタデータ編集画面の初期画面(府省庁一覧)に遷移する。



### (イ) アーカイバルメタデータ編集画面(府省庁一覧)

府省庁一覧が表示されるので、府省庁を選択するとアーカイバルメタデータ編集画面(文書 ID 一覧)に遷移する。



### (ウ) アーカイバルメタデータ編集画面(文書 ID 一覧)

前画面でクリックした府省庁の文書 ID 一覧が表示される。ここで、編集する電子公文書の文書 ID を選択すると編集画面が別ウィンドウで表示される。また、キーワードによる検索も可能である。



#### (エ) アーカイバルメタデータ編集画面(文書 ID データ一覧)

前画面で選択された文書 ID の電子公文書についてのメタデータが表示される。ここでは記録管理メタデータ、アーカイバルメタデータの入力・変更が可能である。また、この画面から電子ファイル一覧表示画面に遷移することも可能である。





#### 4.4.3. プロトタイプシステムにおける操作性評価

本評価は、委員会及び各府省向けデモンストレーションの結果を踏まえた評価を実施したものである。

##### (1) アーカイバルメタデータ編集画面(府省庁一覧)

編集可能な文書データが所属する府省庁名のみをリンクさせ、またアルファベット順に表示させているが、実際に国立公文書館職員が操作する場合、作業が効率的になる表示順序がある場合もあるため、検討が必要と思われる。

##### (2) アーカイバルメタデータ編集画面(文書 ID 一覧)

下位階層に表示される分類データを2列で表示しているため、件数が多くなるとその後が続く文書 ID をスクロールしないと見られなくなる。下位階層をメニュー表示にするなどの工夫が必要と思われる。

##### (3) アーカイバルメタデータ編集画面(文書 ID データ一覧)

この画面では、2つのメタデータ(記録管理、アーカイバル)をタブ別に入力・更新するが、2 つにまたがる項目は記録管理タブに記載し、アーカイバルのほうでは表示のみにしている。それが見た目上ですぐに判断できるようなインタフェースとすべきである。

編集画面でのデータ入力にあたっては、電子ファイルの内容を確認しながら入力作業を行うことになるので、電子ファイルの別画面での表示等に工夫が必要である。

また、受入方法・データ種別・件名の表示・公開状況・審査状況 等のコードで入力する項目に関しては、プルダウン等による選択式とし、コードに対応する値も表示したほうが分かりやすい。

##### (4) アーカイバルメタデータ編集画面(ファイル一覧)

ここでは、前画面でクリックした文書 ID のファイルに関する情報を得られるが、ファイル名のリンクや“ファイル確認”のリンクでは何ができるのか一見して不明なため、別途ヘルプ等を参照しなくてもある程度内容が分かるような説明があることが望ましい。

##### (5) ファイル情報編集画面

ここでは、各ファイルの技術的メタデータを入力・更新できる。ファイルから取得できなかった情報やここで入力するものが多いと思われるが、入力に関して形式を考慮しないといけないと思われるもの(例えば、フォーマットバージョン)に関する説明がない。また、全体的に何を入力したらよいのかが分かりにくいいため、項目ごとに説明を追加することが望ましい。

編集画面でのデータ入力にあたっては、電子ファイルの内容を確認しながら入力作業を行うことになるので、電子ファイルの別画面での表示等に工夫が必要である。

## 4.5. 保存機能実証結果報告

### 4.5.1. プロトタイプシステムにおける機能概要

#### (1) 概要

RDBMS に格納されているデータを元に、文書 ID のコンテナメタデータを作成し、アーカイバルメタデータ作成処理によって作業用フォルダに作成された文書 ID データ(長期保存フォーマット、オリジナルフォーマット、抽出テキストデータ、媒体データ)とともに保存用ストレージに格納する。アーカイバルメタデータ編集画面にて対象文書 ID の確定処理を行った後続いて実行される。

#### (2) 保存条件判定処理

対象の文書 ID のコンテナメタデータ作成が成功し、かつ、対象の文書 ID データ(概要参照)が作業用フォルダに格納されていることが保存条件となる。

#### (3) ストレージ保存処理

(2)の条件を満たした場合は保存用フォルダに格納、作業用フォルダのデータは削除する。条件に満たない場合は、保存不可フォルダに格納する。

### 4.5.2. データ容量

#### (1) 標準フォーマット、長期保存フォーマット、抽出テキストデータのサイズ

保存用ストレージに格納されるサイズを測定した結果を以下に示す。

なお、対象となるデータはテスト用に作成した 1 万件の電子公文書で、Excel2003 形式、GIF 形式、OpenOffice Writer(ODT)、JPEG 形式画像を電子ファイルとして含むものとした。(電子公文書 1 件についていずれかの電子ファイルが 1 個添付される形)。電子ファイルはすべて同じファイルのため、1 件あたりの平均サイズの数字はあくまで参考である。

表 4-11 標準フォーマットのデータ容量

ファイル種別	サイズ(バイト)	
標準フォーマット	10000 件	386,372,885
	1 件当たり平均	<b>38,637</b>

上記の標準フォーマットから変換された長期保存フォーマットの容量は以下の通りである。

表 4-12 長期保存フォーマットのデータ容量

ファイル種別		サイズ(バイト)	
長期保存フォーマット	①PDF/A	総数:1429	585,789,328
		1件当たり平均	<b>409,930</b>
	②JPEG2000	総数:2857	75,262,452
		1件当たり平均	<b>26,343</b>
	③MP3	総数:1429	76,450,071
		1件当たり平均	<b>53,499</b>
	①~③	1件当たり平均	<b>129,046</b>

上記の長期保存フォーマットから抽出されたテキストデータのサイズは以下の通りである。

表 4-13 抽出テキストデータのデータ容量

ファイル種別	サイズ(バイト)	
抽出テキストデータ	総数:1429	251,504
	1件当たり平均	<b>176</b>

(2) コンテナメタデータのサイズ

保存用ストレージに格納されている全文書 ID のコンテナメタデータのサイズを計測した結果を、表 4-14 に示す。

表 4-14 コンテナメタデータのデータ容量

ファイル種別	サイズ(バイト)	
コンテナメタデータ	10000 件	58,089,872
	1件当たり平均	<b>5,809</b>

(3) 媒体データのサイズ

保存用ストレージに格納されている全文書 ID の媒体データのサイズを計測した結果を、表 4-15 に示す。

表 4-15 媒体データのデータ容量

ファイル種別	サイズ(バイト)	
媒体データ	10000 件	347,414,477
	1件当たり平均	<b>34,741</b>

#### (4) ストレージ容量

前述の結果を踏まえて、保存用ストレージの容量を想定し計算した結果を、以下に示す。

保存用ストレージに格納されている1件の電子公文書(文書ID)フォルダを構成しているファイルは、(1)の5種類と(2)、(3)であるから、

(1)の平均 + (2)の平均 + (3)の平均 = 電子公文書1件のデータ平均容量

(1)の平均 (バイト)	+	(2)の平均 (バイト)	+	(3)の平均 (バイト)	=	電子公文書1件のデータ平均容 量(バイト)
129,046	+	5,809	+	34,741	=	169,596

ここで、本システムで対象とする電子公文書の件数として、現在国立公文書館デジタルアーカイブ・システムに登録されている公文書目録データの件数(約300万件)から、5年間の運用とした場合に年間40万件の電子公文書が移管されると仮定して、300万+5×40万=500万件の電子公文書が移管されると想定した。この場合、

電子公文書1件のデータ平均容量:  $169,596 \times 5,000,000 = 847,980,000,000$  バイト

となり、約848GBのストレージ容量が必要となる。

データ容量は、あくまでもファイル構成やファイル容量に依存するため、この結果は1つの目安である。今回のデータでは、電子公文書1件のフォルダを構成している電子ファイルが1個であり、各ファイル容量があまり大きなものではなかったこと、それに伴い媒体データも大きくならなかったため、基準となる数字としてはかなり低めに見ておく必要がある。

#### (5) データベース容量

(4)までと同一の前提条件において、必要となるデータベース容量について試算を行った。

データベース容量の算出基準として、

データベース容量 = 全テーブルのレコードサイズ

であり、前述の条件では、テーブルの1レコードとはすべてのテーブルにおいて電子公文書1件にあたるため、

データベース容量 = 全テーブルの1レコードサイズ×レコード件数  
= 全テーブルの1レコードサイズ×データ件数

により算出できる。

今回のプロトタイプシステムで定義しているテーブルについて、テーブルごとのフィールドサイズの合計は表4-16の通りである。

表 4-16 テーブルごとのレコードサイズ一覧

テーブル名	ヘッダサイズ (バイト)	1フィールドサイズ× フィールド数	レコードサイズ (バイト)	レコード数
-------	-----------------	----------------------	------------------	-------

ファイル情報管理	48	36,092	36,140	1
文書情報管理	40	6,432	6,472	1
メタデータ情報管理	44	33,930	33,974	1
文書情報履歴管理	40	787	827	8
ファイル情報履歴管理	40	803	843	8
計 レコードサイズ×レコード数			<b>89,946 バイト</b>	

表 4-16 のうち、文書情報履歴管理とファイル情報履歴管理テーブルのレコード数については、処理履歴を保存しているため、処理ごとに 1~2 レコード作成される。よって、レコード数は 2 レコード×4 処理 = 8 とした。

総データ件数を 5,000,000 と想定した場合、

データベース容量: $89,946 \times 5,000,000 = 449,730,000,000$ バイト
---

となり、約 450GB となる。

上記のデータベース容量は、データ容量と同様にファイル数に依存するため、この結果は 1 つの目安である。今回のデータでは、電子公文書 1 件のフォルダを構成している電子ファイルを 1 個としたため、

ファイル情報管理のレコード数:  $1(n)\text{ファイル} \times 1 = 1$

ファイル情報履歴管理のレコード数:  $1(n)\text{ファイル} \times 8 = 8$

としている。電子公文書 1 件のフォルダに含まれる電子ファイル数が増えた場合、上記の  $n$ =ファイル数となりその分レコードが増えることになる。

実際のシステム運用を想定下場合、受け入れる電子公文書の電子ファイルの種類や構成は大きく変わることが予想されるため、システム構築前に容量の試算が必要である。

### 4.5.3. セキュリティ

#### (1) 改ざん防止機能

電子公文書の受入・保存において、電子ファイルの原本性、真正性を如何にして確保するかという問題は、平成 19 年度までの調査においても指摘されている課題である。

今回のプロトタイプシステムでは、各電子ファイルについて MD5 チェックサムを付与することにより、仮に行政利用から電子ファイルが流出し、内容が改ざんされたとしても長期保存されているメタデータに格納された MD5 チェックサムと照合することで改ざんの有無をある程度までは検出できる仕組みとした。長期保存フォーマットのうち、PDF/A については電子署名機能が可能な仕様となっており、認証局を設置・連携することで電子署名を埋め込むことは可能だが、標準フォーマット以外の場合は適用できず、認証局自体が長期的に維持される保証が無いことが問題となる。また、メタデータ項目として署名情報の項目を定義したが、具体的な署名情報の格納方法までは調査ができていないため、今後の検討課題となった。

## (2) 通信暗号化による転送速度検証

プロトタイプシステムのサーバ上で、ファイルサイズ約 1.5GB(1,625,425,920 バイト)のファイルを、FTP(バイナリモード)で転送した場合と SFTP で転送した場合の速度を、それぞれ送出時、受信時に測定し、1 秒あたりのスループットを算出した。

なお、測定は 1000BASE-T でスイッチングハブを経由して接続されたサーバ間で行った。

	転送時間(秒)	スループット(Mbps)
FTP(送出)	70	177.2
FTP(受信)	67	185.1
SFTP(送出)	67	185.1
SFTP(受信)	66	187.9

結果として、暗号化を行っているはずの SFTP の方が若干スループットが高いという結果であった。これはネットワーク上の実効転送速度の上限に近い数字で転送が行われたためと推測される。

一般的には暗号化処理のオーバーヘッドが発生する SFTP の方がスループットは低くなるが、ハードウェア上で暗号化を行う機器を使用するなどの方法により、その差を小さくすることが可能である。

## (3) 完全性の確保のための課題

保存用ストレージは電子公文書の完全性を担保するための重要な要素であり、長期に渡り正常稼動する信頼性、耐障害性が最も求められる部分である。保存用ストレージでは、自己診断機能等のチェック機能は必須であると考えられる。

さらに、保存用ストレージを配置する場所についても、物理的なアクセスを制限するためのセキュリティの仕組みや運用方法を検討する必要がある。その一方で、事故や天災などの不可抗力に備えて保存用ストレージの複製またはテープ等のメディアにバックアップされたデータを別の場所に保管/設置する、といったことも検討されるべきである。

## 4.6. 結論

### 4.6.1. メタデータの定義

電子公文書のメタデータとして、記録管理メタデータ、アーカイバルメタデータ、技術的メタデータの3つの区分を定義し、これらすべてを包含するメタデータをコンテナメタデータとして定義した。

メタデータの策定方針は以下の通りである。

- 記録管理メタデータは Dublin Core を基本とし、必要な要素を独自拡張要素として追加した。
- アーカイバルメタデータは EAD 要素を基本とした。
- 技術的メタデータは PREMIS 要素を基本とした。
- コンテナメタデータは3つのメタデータを METS の各セクションに対応させて表現する。

これらのメタデータと電子公文書の対応、及び OAIS 参照モデルとの対応関係を図 4-2 に示す。

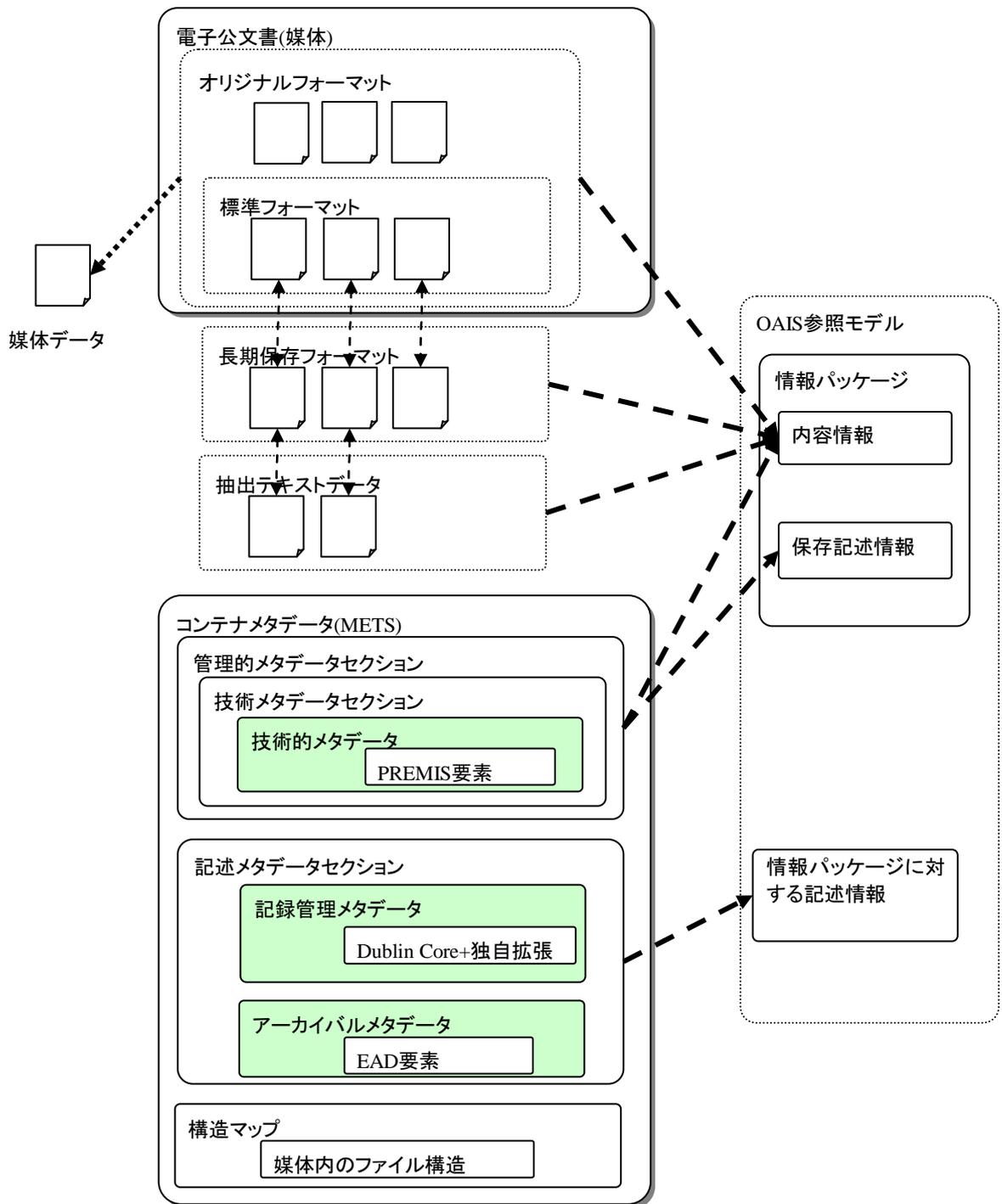


図 4-2 メタデータと電子公文書、及び OAIS 参照モデルとの対応関係

#### 4.6.2. 記録管理メタデータの自動生成

記録管理メタデータは行政文書ファイル管理簿から生成することで、おおよその項目を自動的に生成することが可能である。ただし、保存期間は任意の文字列として設定されている場合があるため、システム上で処理する場合に注意が必要である。

自動生成できない項目としては言語、関連資料、参考文献、権利関係などがある。

#### 4.6.3. 技術的メタデータの自動生成

技術的メタデータについては、ファイル名やファイルサイズなどの情報はファイルのプロパティ情報から自動生成可能である。長期保存フォーマット、抽出テキストデータなど、システム上で生成するファイルについては OS、作成アプリケーション、ウイルスチェック情報などの情報を自動生成することが可能であるが、オリジナルフォーマットについては自動生成(取得)ができない。また、フォント情報も自動抽出することができるフォーマットが限定されるため、これらの情報は移管時に付与されていることが望ましい。

#### 4.6.4. アーカイバルメタデータの自動生成

アーカイバルメタデータは移管後の管理のために利用される情報であるため、自動生成される項目は記録管理メタデータと重複する項目のみである。ただし、移管年月日、移管年度、移管元府省名などの情報は行政文書ファイル管理簿の情報と処理日付から自動生成することが可能である。

#### 4.6.5. 保存機能

保存機能は、オリジナルフォーマット、長期保存フォーマット、抽出テキストデータ、コンテナメタデータを長期保存のためにストレージ上に保存する機能である。

##### (1) 必要なディスク容量

電子データ(オリジナルフォーマット)1 ファイルにつき、長期保存フォーマットと抽出テキストデータの両方が生成される場合、必要となる容量はオリジナルフォーマットのサイズの 2 倍程度を想定しておく必要がある。1 件の電子公文書で生成されるコンテナメタデータは、およそ 30KB+(ファイル数×1KB)程度である。さらに媒体データを保存対象とする場合、オリジナルフォーマットのサイズ合計と同程度の容量が必要となる。

長期保存のために必要なディスク容量は、想定される移管電子公文書件数と、1 件当たりの電子ファイル数より算出する。

##### (2) 保存用ストレージの要件

長期保存を目的とした場合、保存用ストレージは耐障害性を考慮したミラーリングやストライピングが為されていること、自己診断による障害の早期発見が行えることが必須である。また、改ざん防止機能により、システム外の不正なデータ操作を防止し、データの完全性を維持する必要がある。



# 5. 行政利用機能

## 5.1. 調査概要

本項では、電子公文書等の行政利用向けデータの管理、検索、閲覧の各機能について、各機能の評価、連携方法、実運用における課題と対策を調査する。

調査にあたって構築したプロトタイプにおける、各機能間の関連図を図 5-1 に示す。太線で囲った部分が行政利用機能の範囲である。

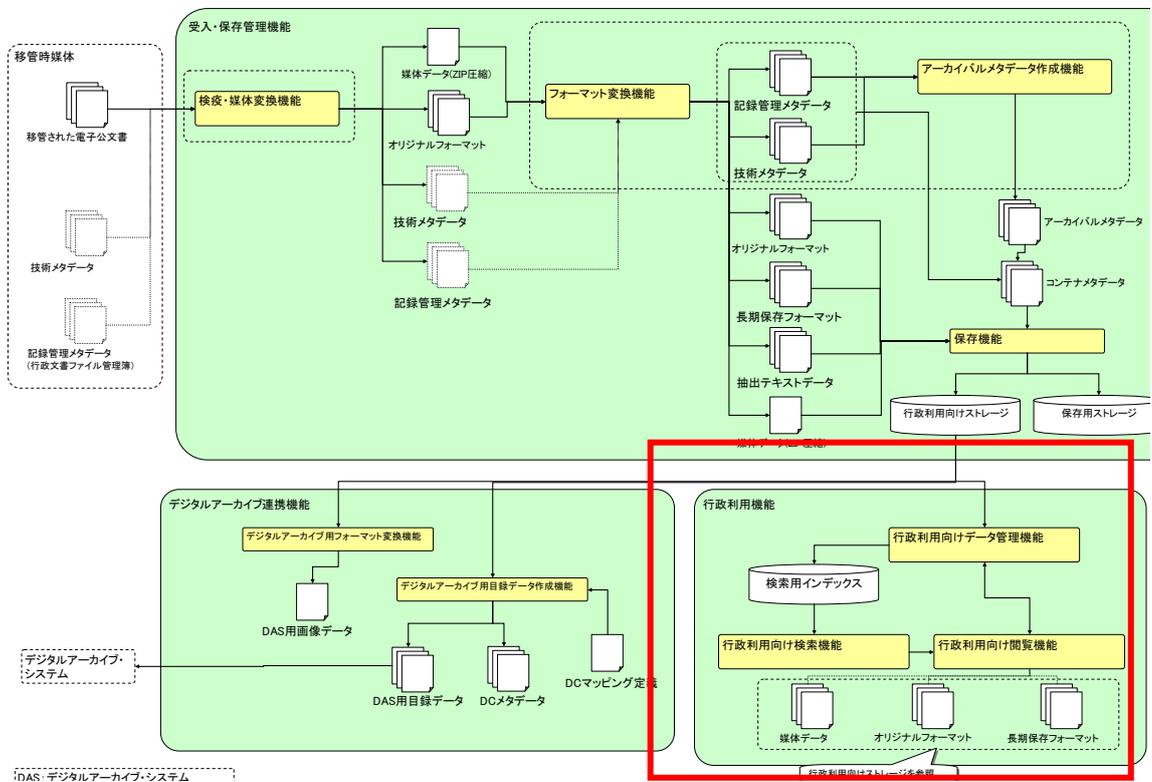


図 5-1 プロトタイプシステムにおける機能関連図

## 5.2. 行政利用向けデータ管理機能実証結果報告

### 5.2.1. プロトタイプシステムにおける機能概要

#### (1) 保存機能からのデータ移行方式

保存用ストレージに格納されている電子公文書ファイル(オリジナルフォーマット、長期保存フォーマット、抽出テキストデータ、媒体データ)及びコンテナメタデータを行政利用ストレージにコピーする。

#### (2) データ管理方式

実際のシステムでは保存用ストレージは行政利用とは切り離された環境であるが、プロトタイプシステムでは、行政利用向けストレージは保存用ストレージとネットワーク上で参照できる構成とする。

#### (3) 検索インデックス

行政利用向けストレージにコピーされたコンテナメタデータ及び抽出テキストデータから、検索用データファイルを生成し、それを元に検索用インデックスの生成を行う。

検索用インデックスに登録するメタデータは、コンテナメタデータのうち記録管理メタデータの部分、及び技術的メタデータのうちオリジナルファイル名のみとする。

### 5.2.2. 保存機能からのデータ移行、管理

#### (1) データ移行における処理時間

保存用ストレージから行政利用ストレージにデータをコピーする処理時間を、対象データ件数を 10,000 件として計測を行った結果を表 5-1 に示す。

表 5-1 行政利用向けストレージコピー処理時間計測結果

対象データ件数	ファイルサイズ(合計)	処理時間
10,000	1,453,180(KB)	15分 12秒 (912秒)

上記の結果より、1件あたりの平均コピー所要時間は 0.0912 秒、1秒あたりのコピー量(スループット)は約 2MB(1,593KB)である。この数字は、検証用サーバ上でローカルディスクから外付けディスクへ 1.6GB のファイルをコピーするのにかかる時間が 33 秒であったため、そのスループット値(1 秒あたり約 50MB)と比較すると遅くなっているが、これは行政利用向けストレージへのコピーにおいて、対象となる電子公文書1件ごとのフォルダを含めてコピーされているため、単一ファイルのコピーを行う場合の数値よりも低くなっていると考えられる。

### 5.2.3. 検索インデックス

#### (1) メタデータサイズと検索インデックス作成処理時間

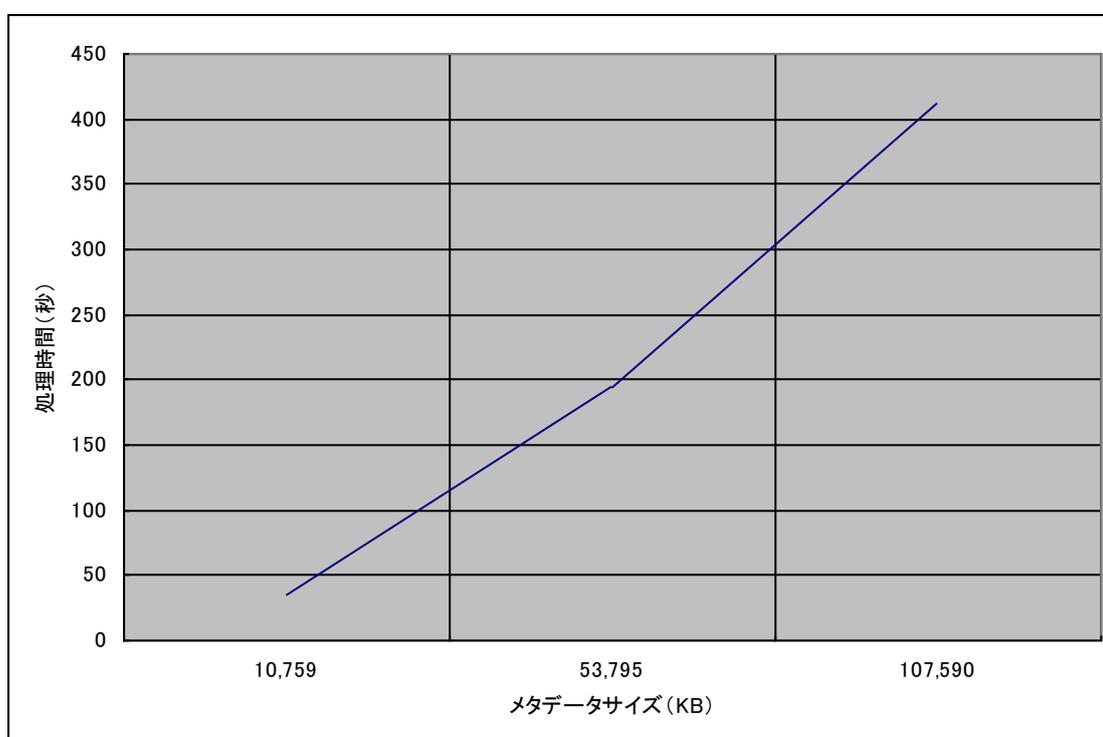
行政利用向けストレージにコピーされたコンテナメタデータ及び抽出テキストデータから、検索用データファイルを生成し、それを元に検索用インデックスの生成処理時間の計測結果を表 5-2 に示す。

なお、メタデータサイズは、下記に示す式で算出した値である。

$$\text{メタデータサイズ} = \text{対象データ件数} * \text{対象データごとのコンテナメタデータのファイルサイズ}$$

表 5-2 メタデータサイズによる検索インデックス作成処理時間計測結果

対象データ件数	メタデータサイズ(合計)	処理時間
10,000	10,759(KB)	35 秒
50,000	53,795(KB)	3 分 14 秒 (194 秒)
100,000	107,590(KB)	6 分 53 秒 (413 秒)



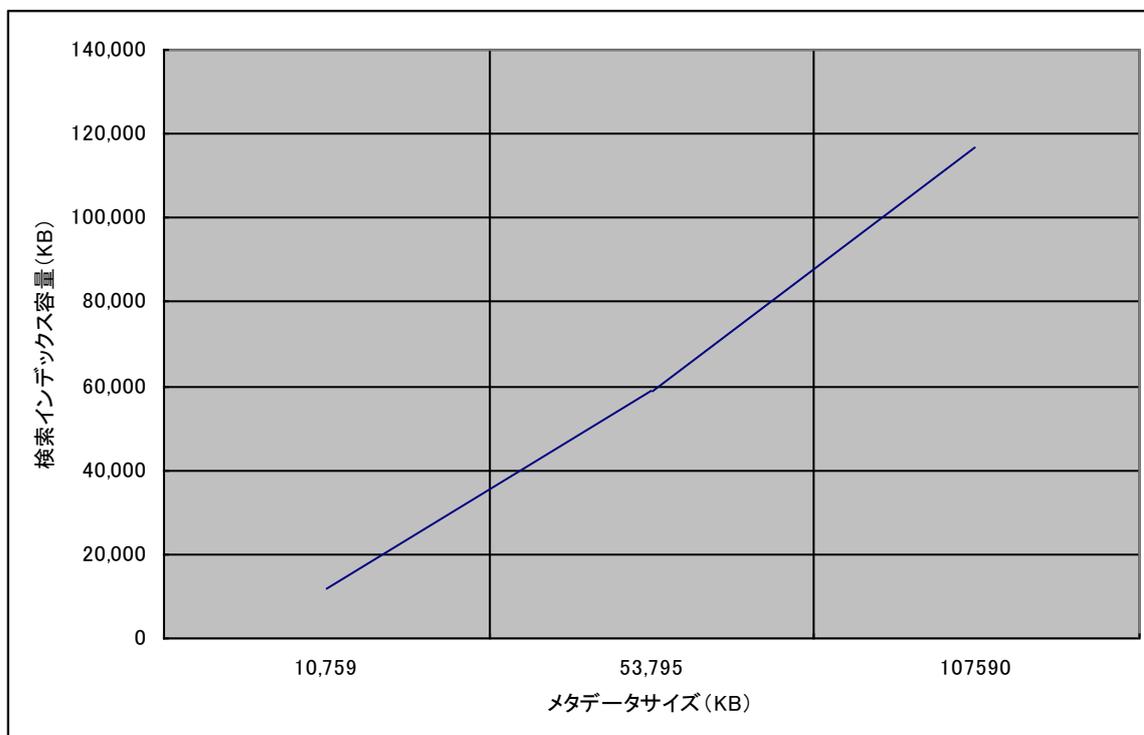
今回の検証ではテスト的にデータをコピーすることで件数を最大 10 万件まで増やしているため、件数とメタデータサイズの増加比率は一定となっている。メタデータサイズとインデックス作成時間は上記のグラフの通り正比例の関係にあるため、本システムの構築にあたってはデータサイズの増加に合わせてインデックス作成を分散・並列処理できる仕組みが必要になると考えられる。

## (2) メタデータサイズと検索インデックス容量

メタデータサイズによる検索インデックス容量の測定結果を表 5-3 に示す。

表 5-3 メタデータサイズによる検索インデックス容量の測定結果

対象データ件数	メタデータサイズ	検索インデックス容量
10,000	10,759(KB)	31,346(KB)
50,000	53,795(KB)	78,326(KB)
100,000	107,590(KB)	156,652(KB)



検証の結果、メタデータサイズと検索インデックスの容量においても比例関係が見られるが、メタデータサイズが大きくなるにつれて検索インデックスの増加比率は小さくなる傾向が見られる。

本システムの構築にあたっては必要となるストレージの容量を算出する際に、メタデータサイズの 2 倍のインデックス容量を見越しておく必要がある。

#### 5.2.4. プロトタイプシステムにおける操作性評価

プロトタイプシステムでは、行政利用向けデータ管理機能における処理はすべてバッチ処理として実行される(ユーザインタフェースを持たない)ため、操作性の評価は行わない。

## 5.3. 行政利用向け検索機能実証結果報告

### 5.3.1. プロトタイプシステムにおける機能概要

#### (1) 検索機能

##### (ア) ログイン処理

利用者がWebブラウザで入力したユーザID、パスワードにより認証を行い、行政利用向け検索機能の利用可否を判断する。ログイン成功時は、該当ユーザIDに対する部局・課の情報を取得する。

##### (イ) 検索処理

利用者がWebブラウザで入力した条件から検索エンジン用の問い合わせ(クエリー)を生成し、検索を実行する。

検索結果はXML形式で取得する。

##### (ウ) 検索結果変換処理

検索エンジンから返されるXML形式のデータに対して、XSLTスタイルシートを適用することでHTML形式に変換し、利用者に返却する。

### 5.3.2. 検索機能

#### (1) ユーザインタフェース要件

##### (ア) ログイン画面

ユーザID、パスワードによるログイン画面



##### (イ) 任意文字列検索画面

キーワードによるメタデータ及び抽出テキストデータの全文検索を行う画面。

ログイン直後は、この画面を表示する。

タブをクリックすることで詳細検索画面、階層検索画面に切り替えることができる。



#### (ウ) 詳細検索画面

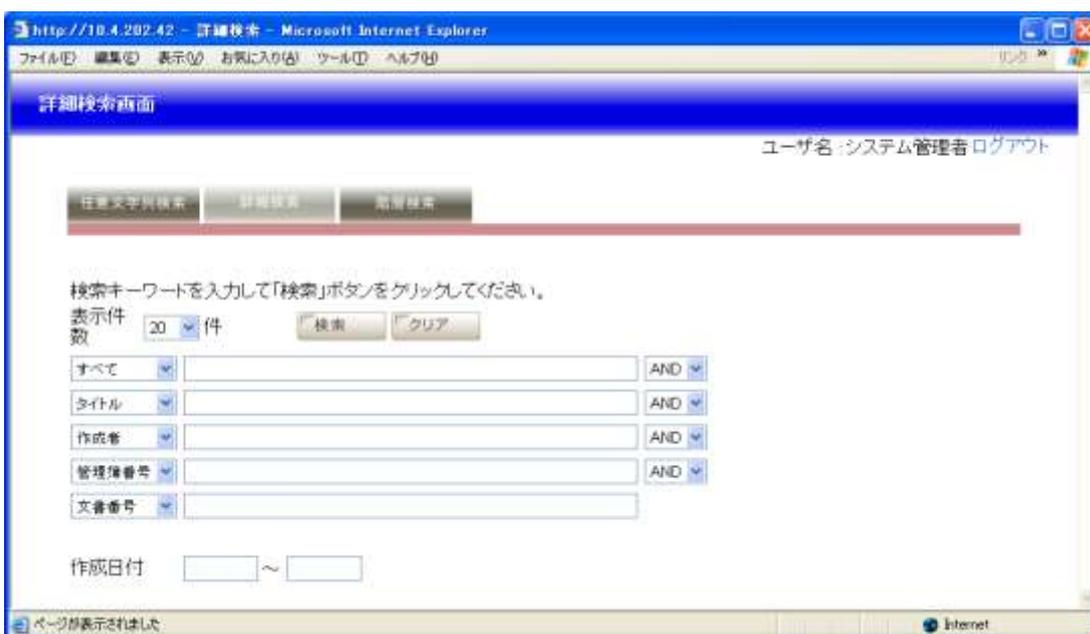
メタデータ項目(または抽出テキスト)とキーワードを指定して検索を行う画面。

検索対象として選択できる項目は以下の通りである。

- すべて
- タイトル
- 作成者
- 管理簿番号
- 受入文書番号
- 備考
- ファイル名
- テキスト
- 作成日付

項目間の論理演算子として AND、OR、NOT を指定できる。

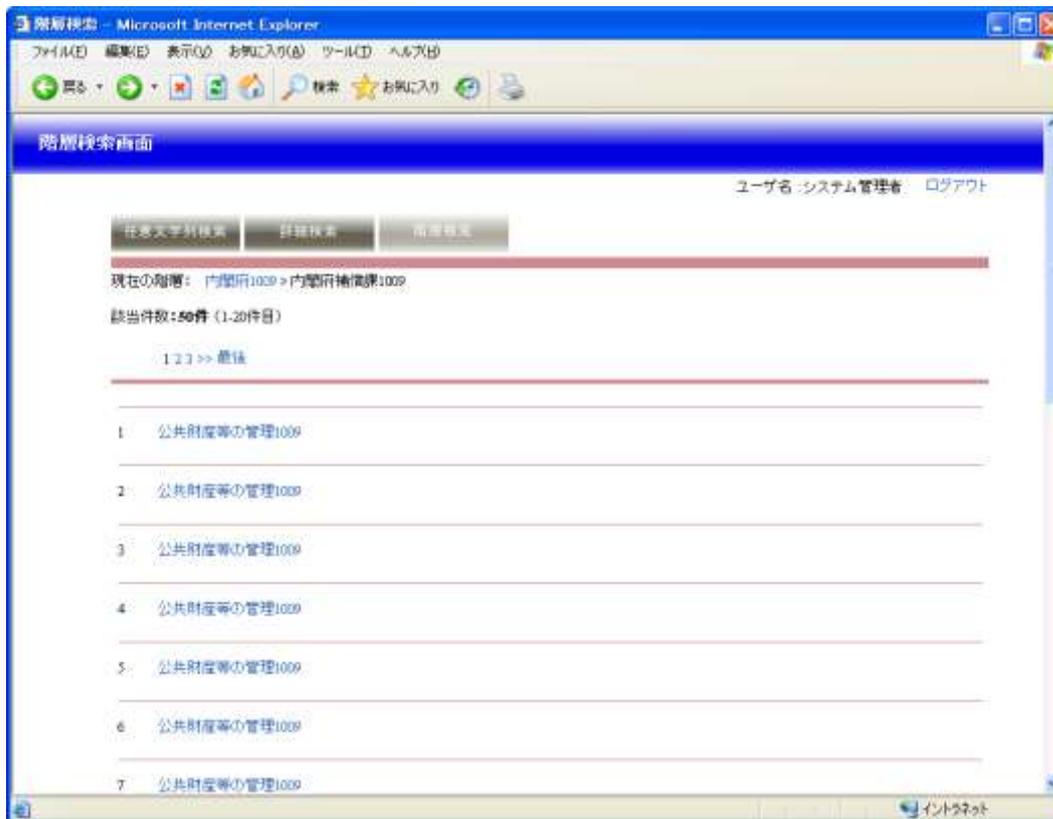
タブをクリックすることで任意文字列検索画面、階層検索画面に切り替えることができる。



## (エ) 階層検索画面

分類ごとに階層的に検索を行う画面。

タブをクリックすることで任意文字列検索画面、詳細検索画面に切り替えることができる。



## (オ) 検索結果一覧画面

指定された条件による検索結果一覧を表示する画面。

文書タイトル、移管元府省名、作成者等の概要情報と、階層情報も併せて表示される。階層情報からは該当階層の階層検索画面に遷移することができる。

また、ファイル一覧に遷移して電子ファイルの閲覧が可能である。



#### (カ) 詳細表示画面

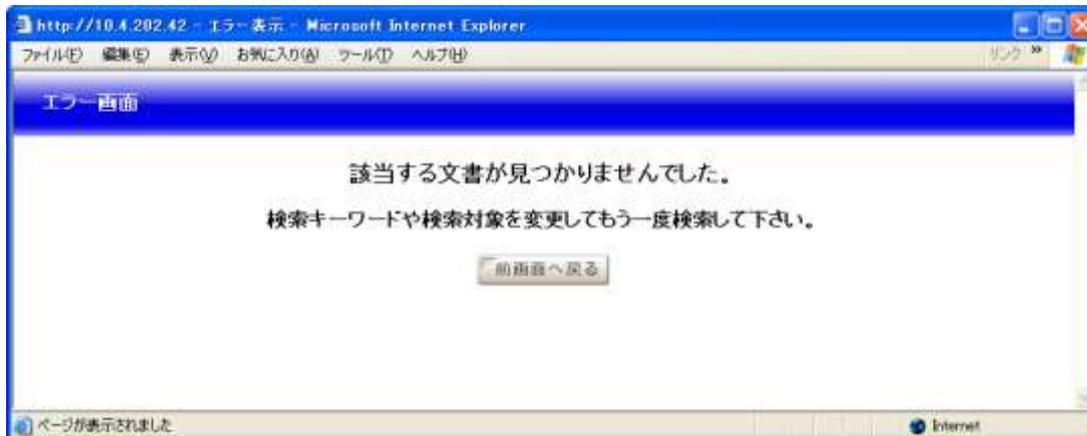
指定された公文書のメタデータ詳細を表示する画面。

画面上部の「記録管理」、「アーカイバル」タブで、表示する情報を記録管理メタデータの内容、アーカイバルメタデータの内容と切り替えることができる(下図は記録管理メタデータの内容)。



(キ) エラー画面

各画面よりエラーが発生した場合は、この画面に遷移する。



(2) 検索インデックス容量と検索処理時間

検索インデックス容量別の、行政利用向け検索処理時間の計測結果を表 5-4 に示す。

表 5-4 検索インデックス容量による検索処理時間の計測結果

対象データ件数	検索インデックス容量	処理時間
10,000	31,346(KB)	1 秒未満
50,000	78,326(KB)	1 秒未満
100,000	156,652(KB)	1 秒未満

検索方法は、以下の 4 パターンで行ったがすべて 1 秒未満で検索結果一覧画面が表示された。

1. 任意文字列検索画面での検索を行う。
2. 詳細検索画面で「すべて」、「作成者」、「タイトル」、「ファイル名」、「テキスト」を指定し、AND 条件で検索を行う。
3. 2 を OR 条件で検索を行う。
4. 2 を NOT 条件で検索を行う。

今回の検証ではデータ件数を最大 10 万件としたが、検索処理のパフォーマンス上の影響は見られなかった。検索エンジンの性能による面もあるが、将来的に本システムの稼働・運用が始まってから、電子公文書として受入が予想される件数が数百万件を超えることを想定し、拡張性のあるソフトウェア・ハードウェア構成を検討する必要がある。

### 5.3.3. プロトタイプシステムにおける操作性評価

詳細検索画面における検索項目は、今回は、「すべて」、「タイトル」、「作成者」、「管理簿番号」、「受入文書番号」、「備考」、「ファイル名」、「テキスト」、「作成日付」の 9 種類としているが、府省ごとに検索時にキーとして指定する項目は異なることが予想されるため、今回用意した 9 種より増える可能性がある。

検索結果一覧画面、詳細表示画面に表示する項目も、同様のことが考えられる。検索一覧画面のソート機能については、プロトタイプシステムでは、1 項目に対してのソートしか行えない。検索一覧画面に表示される項目が増加した場合、1項目のソートでは、要望どおりのソートが必ずしも行えるとはいえないため、ソート項目も複数指定できるように改善し、ソートすべき項目も検討する必要がある。

## 5.4. 行政利用向け閲覧機能実証結果報告

### 5.4.1. プロトタイプシステムにおける機能概要

#### (1) 閲覧機能

閲覧機能は行政利用向け検索機能から呼び出され、電子公文書等を構成する電子ファイルを検閲するための機能である。

#### (ア)ログイン処理

利用者がWebブラウザで入力したユーザID、パスワードにより認証を行い、行政利用向け閲覧機能の利用可否を判断する。

ログイン成功時は、該当ユーザIDに対応する部局・課の情報を取得する。行政利用向け検索機能でログイン済みの場合、認証情報を引き継ぎ、認証処理を行わない。

#### (イ)ファイル一覧取得処理

文書IDをキーとしてテーブルの検索を行い、該当するファイル一覧を取得する。

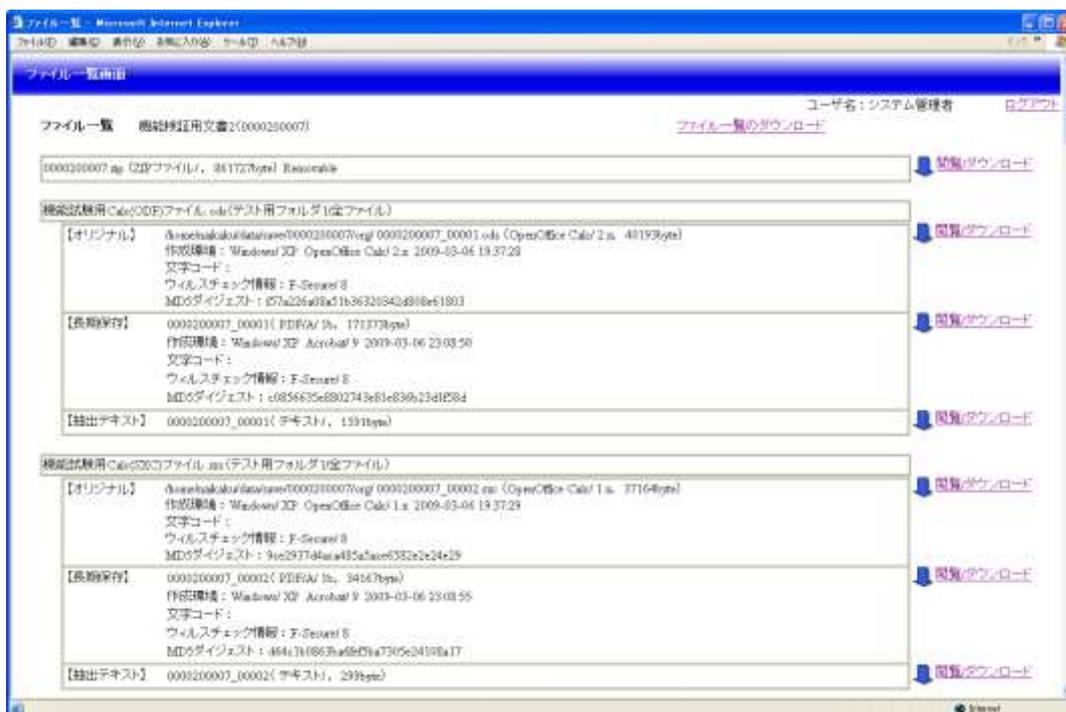
ファイル一覧はXML形式で出力し、XSLTスタイルシートを適用することでHTML形式に変換し、利用者に返却する。

また、ファイル一覧画面から「ファイル一覧のダウンロード」がクリックされた場合には、該当電子公文書等の技術的メタデータをCSV形式に変換して利用者に提供する。

ファイル一覧画面を下記に示す。

なお、ここで提供されるデータはオリジナルフォーマット、長期保存フォーマット、抽出テキストデータ、媒体データである。それぞれのフォーマット(データ)について、媒体上のファイルパス、元ファイル名、フォーマット、ファイルサイズ、作成時タイムスタンプ、MD5ダイジェスト等の情報が表示される。

「閲覧/ダウンロード」をクリックすることで、各フォーマットに対応したファイル取得処理が呼び出される。



#### (ウ)ファイル取得処理

ファイル一覧画面から各フォーマットの「閲覧ダウンロード」がクリックされた場合に、ファイル ID 及びファイル形式をキーとしてテーブルの検索を行い、電子公文書ファイルのデータファイルを行政利用向けストレージから取得し、データを利用者の Web ブラウザに送信(ダウンロード)する。

#### (エ)媒体データ取得処理

ファイル一覧画面から媒体データの「閲覧ダウンロード」がクリックされた場合に、文書 ID をキーとしてテーブルの検索を行い、該当する媒体データ(ZIP 圧縮形式)を利用者の Web ブラウザに送信する。

### 5.4.2. 閲覧機能検証

#### (1) フォーマットごとの対応ソフトウェア調査

「ファイルフォーマット調査報告」で定義した標準フォーマット及び長期保存フォーマットについて、行政利用向け閲覧機能で利用者が各ファイルを閲覧するために利用可能なソフトウェアについて調査を行った。なお、以下に挙げたアプリケーションまたはプラグインについては、2009 年 2 月時点において比較的入手が容易なもの、フォーマット自体の作成アプリケーションであるため再現性の高いものを中心としているが、対応するソフトウェアをすべて記載しているわけではない。また、今後のバージョンアップ等により対応状況が変わる可能性もある点に注意されたい。

まず、標準フォーマットについて、閲覧機能で使用可能(閲覧可能)なアプリケーションまたはプラグインは表 5-5 の通りである。

表 5-5 標準フォーマットの対応ソフトウェア

NO	フォーマット分類	フォーマット	対応アプリケーションまたはプラグイン
1.	文書作成フォーマット	OASYS	OASYS V10

NO	フォーマット分類	フォーマット	対応アプリケーションまたはプラグイン
		一太郎 8-12	一太郎 8 以降、一太郎ビューア 5.x 以降
		Word97-2003	Microsoft Word 97/98、2000、2002、2003、2007(互換モード)
		Word2007	Microsoft Word 2007
		PDF1.x、PDF/A	Adobe Reader 9 以降
		Open Office Writer	Open Office Write 3
2.	表計算フォーマット	Excel97-2003	Microsoft Excel 97、2000、2002、2003、2007(互換モード)
		Excel2007	Microsoft Excel 2007
		Open Office Calc	Open Office Calc 3
3.	プレゼンテーションフォーマット	PowerPoint97-2003	Microsoft PowerPoint 97、2000、2002、2003、2007(互換モード)
		PowerPoint2007	Microsoft PowerPoint 2007
		Open Office Impress	Open Office Impress 3
4.	画像	JPEG	Web ブラウザ(Internet Explorer 5.x 以降、FireFox 1.x 以降)
		JPEG 2000	LizardTech Express View Browser Plug-in 4
		GIF	Web ブラウザ(Internet Explorer 5.x 以降、Fire Fox 1.x 以降)
		TIFF	Microsoft Office Document Imaging、Windows Picture and Fax Viewer、Paint
		BMP	Windows Picture and Fax Viewer、Paint
5.	音声・音楽	WAVE	Windows Media Player 11、Real Player
		MP3	Windows Media Player 11、Real Player
		WMA	Windows Media Player 11
6.	映像	QuickTime	QuickTime Player
		Windows Media Video	Windows Media Player 11
		RealPlayer	Real Player
		MPEG	Windows Media Player 11、Real Player、QuickTime Player

次に、長期保存フォーマットデータに対して、閲覧機能で使用可能(閲覧可能)なアプリケーションまたはプラグインは表 5-6 の通りである。

表 5-6 長期保存フォーマットの対応ソフトウェア

NO	フォーマット分類	フォーマット名	対応アプリケーションまたはプラグイン
1.	文書	PDF/A	Adobe Reader 9 以降
2.	画像	JPEG 2000	LizardTech Express View Browser Plug-in 4
3.	音声・音楽	MP3	Windows Media Player 11、Real Player
4.	映像	MPEG	Windows Media Player 11、Real Player、QuickTime Player

(2) フォーマット、対応ソフトウェアごとの処理時間

上記の標準フォーマット及び長期保存フォーマットに対応したアプリケーションまたはプラグインを用いて、検証システム上で閲覧を行うまでの時間(ファイルのダウンロードを開始してから、アプリケーションが起動するまでの時間)を検証した結果を以下に示す。

なお、検証は100BASE-Tにより接続されたローカルエリアネットワーク上に配置されたクライアントPC上で、検証サーバを対象として行った。

表 5-7 標準フォーマットごとの閲覧処理結果

NO	フォーマット分類	フォーマット	ファイルサイズ	処理時間
1.	文書作成フォーマット	OASYS	3(KB)	1秒
			102(KB)	2秒
		一太郎 8-12	21(KB)	2秒
			275(KB)	2秒
			664(KB)	2秒
			1,578(KB)	3秒
		Word97-2003	32(KB)	1秒
			273(KB)	1秒
			335(KB)	2秒
			662(KB)	2秒
		Word2007	34(KB)	3秒
			40(KB)	5秒
		PDF1.x、PDF/A	47(KB)	1秒
			256(KB)	1秒
			653(KB)	1秒
		Open Office Writer	10(KB)	1秒
48(KB)	1秒			
88(KB)	3秒			
269(KB)	2秒			
2.	表計算フォーマット	Excel97-2003	50(KB)	1秒
			175(KB)	1秒
			692(KB)	1秒
			1,294(KB)	2秒
		Excel2007	31(KB)	2秒
			37(KB)	2秒
		Open Office Calc	30(KB)	1秒
3.	プレゼンテーションフォーマット	PowerPoint97-2003	20(KB)	2秒
			84(KB)	2秒

NO	フォーマット分類	フォーマット	ファイルサイズ	処理時間
		PowerPoint2007	120(KB)	2 秒
			125(KB)	2 秒
		Open Office Impress	14(KB)	3 秒
			34(KB)	3 秒
			39(KB)	4 秒
4.	画像	JPEG	77(KB)	1 秒未満
			106(KB)	1 秒未満
			130(KB)	1 秒未満
		JPEG 2000	49(KB)	3 秒
			184(KB)	3 秒
		GIF	11(KB)	1 秒未満
			87(KB)	1 秒未満
			93(KB)	1 秒未満
		TIFF	806(KB)	1 秒未満
			1,350(KB)	1 秒未満
			2,706(KB)	1 秒未満
		BMP	767(KB)	1 秒未満
			1,346(KB)	1 秒未満
5.	音声・音楽	WAVE	291(KB)	2 秒
		MP3	53(KB)	1 秒
			1,473(KB)	2 秒
		WMA	28(KB)	1 秒
6.	映像	QuickTime	3,284(KB)	4 秒
		Windows Media Video	114(KB)	2 秒
			6,198(KB)	2 秒
			20,846(KB)	3 秒
		RealPlayer	18(KB)	2 秒
			58(KB)	2 秒
			560(KB)	2 秒
MPEG	528(KB)	4 秒		

表 5-8 長期保存フォーマットの閲覧処理結果

NO	フォーマット分類	フォーマット	ファイルサイズ	処理時間
1.	文書	PDF/A	86(KB)	1 秒
			1,214(KB)	1 秒
			1,120(KB)	1 秒
			1,345(KB)	1 秒

NO	フォーマット分類	フォーマット	ファイルサイズ	処理時間
			1,044(KB)	1 秒
			473(KB)	1 秒
			620(KB)	1 秒
			96(KB)	1 秒
			436(KB)	1 秒
			301(KB)	1 秒
			438(KB)	1 秒
2.	画像	JPEG 2000	44(KB)	3 秒
			168(KB)	3 秒
			184(KB)	3 秒
			191(KB)	3 秒
			208(KB)	3 秒
			352(KB)	3 秒
3.	音声・音楽	MP3	27(KB)	1 秒
			1,477(KB)	1 秒
4.	映像	MPEG	1,157(KB)	1 秒
			17,115(KB)	3 秒
			29,102(KB)	5 秒

標準フォーマットにおいては、表 5-7 に示すとおり、対象ファイルのバイト数に比例して処理時間が増大している。但し、Open Office Writer で作成した 269KB のファイルを表示するのに要した処理時間が 2 秒なのに対して、88KB のファイルを開くのに要した処理時間が 3 秒である。これは、アプリケーションを起動する時間の影響だと思われる。

長期保存フォーマットにおいては、表 5-8 に示すとおり、PDF/A は約 1 秒、JPEG2000 は約 3 秒、MP3 は約 1 秒、そして、MPEG はファイルサイズに比例するという結果が得られた。MPEG 以外は処理時間がほぼ一定になったのは、各フォーマット表示用アプリケーションがバックグラウンドで読み込み処理を行っているためであると思われる。

プロトタイプシステムでは処理時間についての問題は特に見受けられなかったが、本システムでは、回線速度、ネットワークの混雑による影響、大容量ファイルダウンロードで処理時間の増大が考えられるので考慮する必要がある。

#### 5.4.3. プロトタイプシステムにおける操作性評価

ファイル一覧画面は、プロトタイプシステムでは該当するファイル情報を全て一画面に表示しているが、1 文書に対してファイルが多数存在する場合もあるため、検索結果一覧画面のように、表示件数指定機能を追加する必要がある。

また、ファイル一覧画面では、一文書全体の階層構成が把握しづらいため、ツリー表示できる機能、さら

には、ツリー情報からファイル一覧画面を表示できる機能を追加することによって操作性の向上が見込める。

## 5.5. 業務管理方法

行政利用機能(行政利用向け検索機能、行政利用向け閲覧機能)において必要となる業務管理方法について検討した結果を以下に示す。

### 5.5.1. 利用者権限、アクセス権限

今回のプロトタイプシステムでは、内部で簡易的なユーザ管理機能を実装し、ユーザID及びパスワード、所属組織を管理する仕組みとした。

行政利用機能ではユーザ認証の仕組みが必須であり、また各府省(組織)から移管された電子公文書の検索・閲覧は移管元組織に所属している行政利用者のみがアクセスできるように制限を行う必要がある。

本システムの構築にあたり、利用者の管理、アクセス権限の管理のために以下の課題を検討しなければならない。

#### (1) 利用者管理の実現方法

利用者管理については、対象となる行政利用者が各府省に存在し、またその数も多くなることが想定されるため、国立公文書館ですべてを管理することは現実的に難しいと考えられる。

この問題を解決するためには、各府省の職員に対する共通の認証システムが整備され、そのシステムと連携することが期待される。総務省では職員等利用者共通認証基盤(GIMA)の整備が進められており、それとの連携実現も選択肢の一つとして検討すべきであろう。

#### (2) 行政利用者の異動、組織変更への対応

検索・閲覧の対象を移管元組織に所属する行政利用者に制限することについて、技術的な問題はクリアされているといえるが、府省内の異動や、組織変更が発生した場合にどのように処理するかはまた別の問題として検討されなければならない。

例えば、行政利用者の異動履歴や組織変更履歴をデータベースとして管理しておき、過去の組織に遡って権限のチェックを行う仕組みを実装することは可能だが、現在所属していない組織の情報にもアクセスを許可すべきなのか、またどの時点までの履歴を有効とするか、といった点は実際に行政利用の主体となる各府省の意見も踏まえながらルール化されるべきである。

### 5.5.2. 利用統計データ

行政利用向け検索機能、行政利用向け閲覧機能では、検索・閲覧の各要求について、ユーザID、IPアドレス、アクセス日時、対象データ等の利用情報をログとして保管しておく必要がある。

これらのログは行政利用機能の利用動向把握に役立つと同時に、不正アクセスや必要以上のデータ閲覧(ダウンロード)が行われていないか、といったセキュリティ面での管理情報としても利用すべきである。

## 5.6. 結論

### 5.6.1. 行政利用向けデータ管理機能

#### (1) 保存機能からのデータ移行・管理

プロトタイプシステムでは、行政利用向けストレージは保存用ストレージとネットワーク上で参照できる構成とし、ネットワーク上で物理的なファイルコピーを行う方式としたが、実際のシステムでは保存用ストレージは行政利用とは切り離された環境であるため、別の媒体を介したデータコピーを行う方式とする。

行政利用にコピーするデータは、保存用ストレージ上の全データを対象とすると容量が大きすぎることに、コピー処理に膨大な時間がかかることから、前回コピー時からの更新分のみを対象とできる仕組みが望ましい。同時に、MD5 チェックサム値を用いて、行政利用向けストレージにコピーされたファイルが一致するかのチェックを行うことが推奨される。

また、コピーに使用する媒体は、USB 等で接続可能な外付け HDD (移動可能な大容量媒体) とし、1TB 程度の容量を持つことが望ましい。この作業で発生する人的コスト、さらにはセキュリティについて留意しなければならない。

### 5.6.2. 行政利用向け検索機能

#### (1) 検索機能

行政利用向け検索機能では、任意文字列検索、詳細検索、階層検索の 3 方式の検索用インタフェースを用意した。詳細検索では項目を指定することが可能であるが、指定できる項目については実際に検索を行う各府省の職員からの要望を踏まえて検討する必要がある。

検索結果一覧では主要なメタデータ項目と階層情報が表示されるが、この項目も各府省の職員からの要望を踏まえて検討する必要がある。検索機能では全角文字と半角文字、アルファベットの太文字や小文字といった文字を正規化することで、検索時の網羅性を向上させる仕組みも必要である。

#### (2) 検索インデックスに含まれるメタデータについて

プロトタイプシステムでは、検索インデックスに含まれる情報は記録管理メタデータと技術的メタデータの一部、抽出テキストデータとした。これはコンテナメタデータ全体を検索インデックスに登録する場合に処理時間と検索インデックスのサイズが大きくなってしまふことによる。ただし、行政利用においてアーカイバルメタデータの検索の必要性が高いようであれば、コンテナメタデータ全体を検索できるようにしておくことが望ましい。

行政利用サービスを継続的に提供するためにも、検索インデックスを二重化するなどの方式により、更新時の検索レスポンス低下やサービス停止時間を避けることが必要である。

#### (3) セキュリティ

行政利用向け検索機能は、メタデータの検索のみとはいえ、各府省の行政利用者により提供される機能であるため、通信は暗号化されることが必須要件と言える。ネットワークとして霞ヶ関 WAN 内のみを対象とするのであれば、霞ヶ関 WAN との通信制御や帯域利用についても調整が必要となる。

### 5.6.3. 行政利用向け閲覧機能

#### (1) 閲覧機能

閲覧機能では、オリジナルフォーマット、長期保存フォーマット、媒体データのダウンロードが可能である。オリジナルフォーマットのうち標準フォーマットについては、原則として作成アプリケーションが必要となるが、無償のビューアや互換性の高いアプリケーション(MS-Office と OpenOffice など)を利用することで閲覧が可能となる場合もある。ただし、オリジナルフォーマットでのレイアウト等が完全に維持されない場合もある点に留意が必要である。

映像(動画)や音声、媒体データをダウンロードする場合、データサイズが数十 MB を超えるケースもあるため、行政利用サービスについてはバックボーンとなるネットワーク帯域に留意する必要がある。

#### (2) アクセス管理

行政利用向け閲覧では、該当電子公文書の作成(移管)元組織のみが閲覧できるように制限を設ける必要がある。プロトタイプシステムでは内部で利用者管理を行い、所属組織の情報をメタデータに付与することで閲覧制限を行う仕組みとしたが、実際のシステム運用では、国立公文書館が各府省の行政利用者を個別に管理することは難しいため、全府省の統一的な利用者認証システムが構築された場合、その機能と連携することも視野に入れる必要がある。

#### (3) セキュリティ

行政利用向け閲覧機能においても、行政利用向け検索機能で指摘したセキュリティについて配慮する必要がある。



# 6. デジタルアーカイブ連携機能

## 6.1. 調査概要

本項では、行政利用機能で管理される電子公文書等を、国立公文書館デジタルアーカイブで一般の利用に供するために必要となる、デジタルアーカイブ用フォーマット変換、デジタルアーカイブ用目録データ作成の各機能について、実稼動環境等での検証等を踏まえて、各機能の評価、連携方法、実運用における課題と対策を調査する。

調査にあたって構築したプロトタイプシステムにおける、各機能間の関連図を図 5-1 に示す。太線で囲った部分がデジタルアーカイブ連携機能の範囲である。

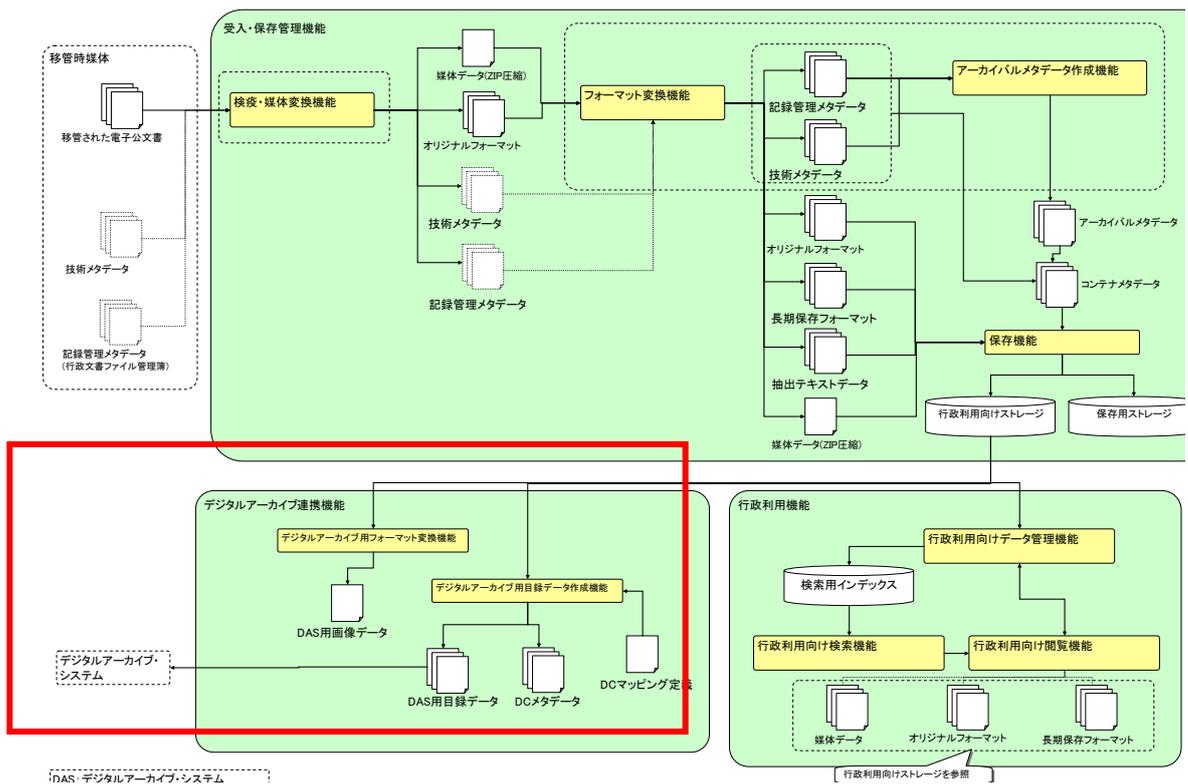


図 6-1 プロトタイプシステムにおける機能関連図

## 6.2. デジタルアーカイブ用フォーマット変換機能実証結果報告

### 6.2.1. プロトタイプシステムにおける機能概要

#### (1) フォーマット変換方式

プロトタイプシステムでは、長期保存フォーマットからデジタルアーカイブ用フォーマットへの変換は、文書フォーマット、表計算フォーマット、プレゼンテーションフォーマット、画像フォーマットのみを対象とした(長期保存フォーマットとして JPEG 2000 または PDF/A として出力されている場合は変換を行わず、コピーのみとする)。本機能はバッチ処理として実行される。

### 6.2.2. フォーマット変換

#### (1) 変換処理時間

プロトタイプシステムでは、デジタルアーカイブ用フォーマット変換の対象となるデータはすべて長期保存フォーマットに変換済みのデータとなるため、変換処理時間については特に測定を行わず、ファイルコピーにかかる時間を測定した。

コピー対象となるデータについては、1万件分の電子公文書(それぞれ JPEG 2000 または PDF/A 形式のファイルが1ファイルずつ存在)を対象とした。

対象データ件数	ファイルサイズ(合計)	処理時間
10,000	799,812 (KB)	2分45秒 (165秒)

上記の結果より、1件あたりの平均コピー所要時間は0.0165秒、1秒あたりのコピー量(スループット)は約5MB(4,847KB)である。この数字は、検証用サーバ上でローカルディスクから外付けディスクへ1.6GBのファイルをコピーするのにかかる時間が33秒であったため、そのスループット値(1秒あたり約50MB)と比較すると遅くなっているが、これはデジタルアーカイブ用フォーマットのコピーにおいて、対象となる電子公文書1件ごとのフォルダを含めてコピーされているため、単一ファイルのコピーを行う場合の数値よりも低くなっていると考えられる。

フォーマット変換が必要な場合、受入・保存管理機能で検証したフォーマット変換機能の処理時間と同様の時間がかかるものと推測される。

#### (2) 変換結果検証

変換結果の検証については、上記の通り長期保存フォーマットに変換されたデータをそのまま利用しているため、受入・保存管理機能における検証結果を参照されたい。

#### (3) マスキング処理

国立公文書館デジタルアーカイブで一般の利用に供するにあたって、電子公文書の内容に当館利用規則で利用を制限する場合がある。そのような状況に対応するためにデジタルアーカイブ用フォーマットについて、該当情報をマスキングするための実現方式を調査した。

## (ア)PDF/A でのマスキング

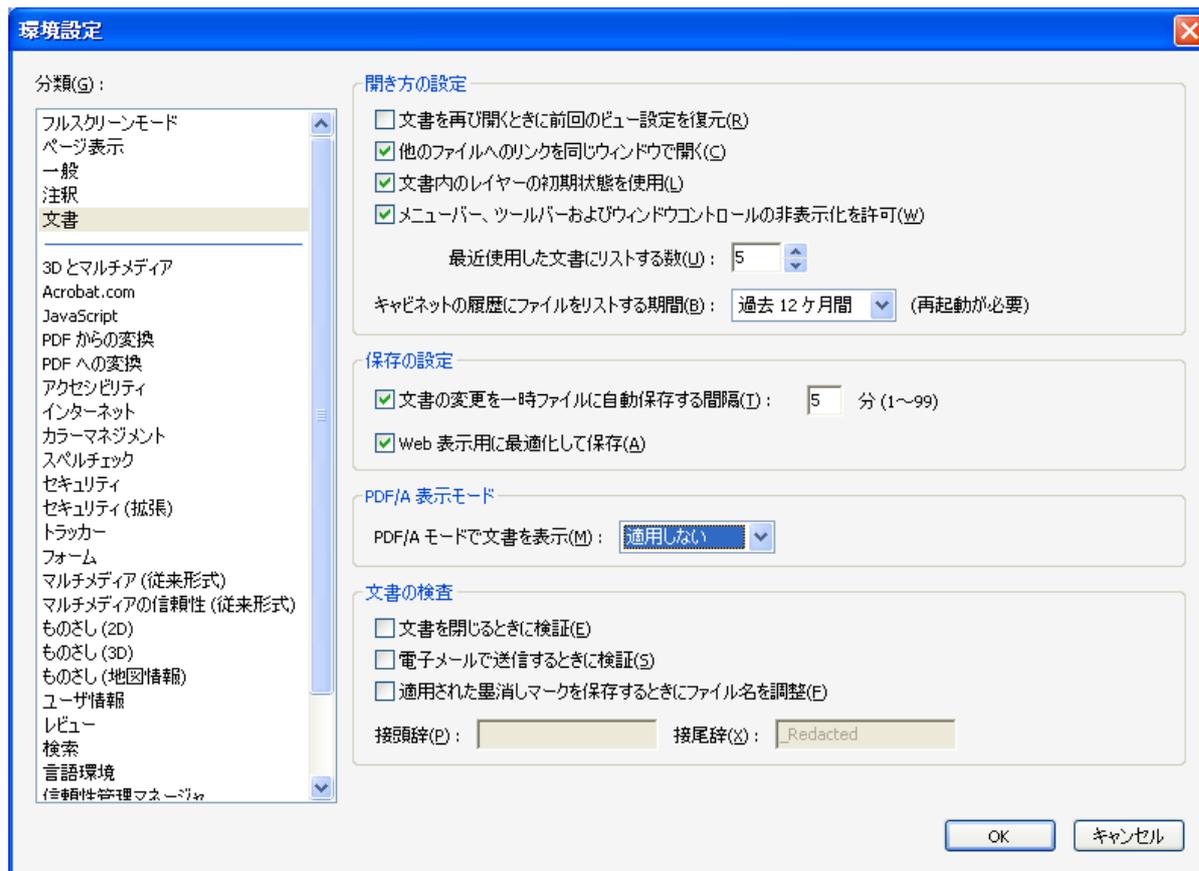
PDF/A 形式のデータに対するマスキングの実現性について、以下の手順で検証を行った。

使用ソフトウェア: Adobe Acrobat 9 Standard (バージョン 9.0.0)

### (手順 1)

Acrobat 9 を起動し、「環境設定」で「分類」オプションの「PDF/A 表示モード」の設定を「適用しない」にする。

※PDF/A 表示モードでは編集ツールによるデータ編集が行えないため。



### (手順 2)

PDF/A 形式のファイルを Acrobat 9 で開き、「ツール」→「高度な編集」→「TouchUp テキストツール」を選択する。

### (手順 3)

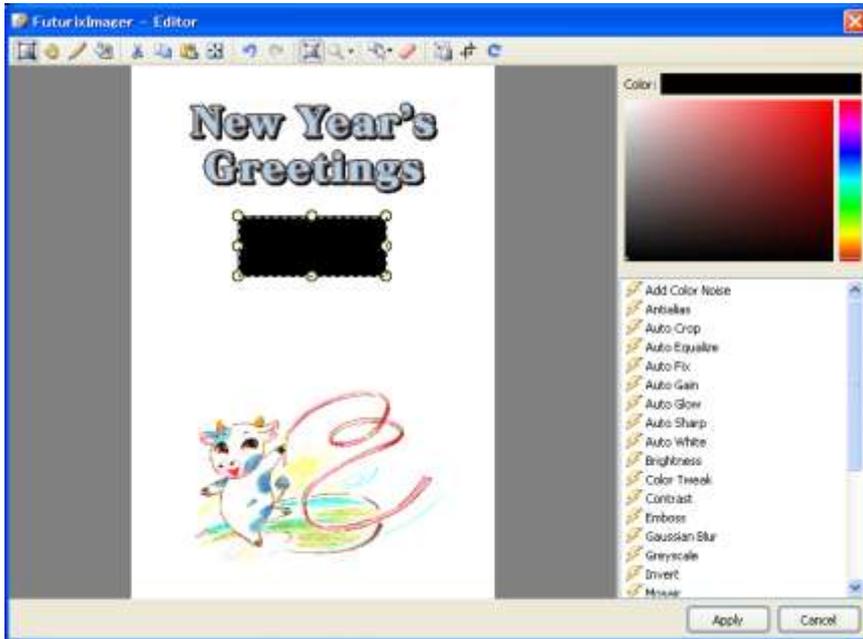
マスキング対象となる箇所をマウスでドラッグして選択し、置き換える文字列を入力する。





(手順 3)

Cut ボタンをクリックすると選択範囲が切り取られるので、「Apply」をクリックし、編集内容を保存する。



(手順 4)

手順 3 で保存したファイルを FuturixImager で開くと、マスキングした箇所が黒く切り取られている。



この状態では再度編集状態で開いても、マスキングされた箇所を再現することはできない。

結論として、JPEG 2000 の場合は、対応した画像編集ソフトを使って該当箇所を塗りつぶすことにより、マスキングを実現可能であるといえる。

### 6.2.3. プロトタイプシステムにおける操作性評価

プロトタイプシステムでは、デジタルアーカイブ用フォーマット変換機能はバッチ処理として実行される(ユーザインタフェースを持たない)ため、操作性の評価は行なわない。

## 6.3. デジタルアーカイブ用目録データ作成機能実証結果報告

### 6.3.1. プロトタイプシステムにおける機能概要

#### (1) 目録データ変換

デジタルアーカイブへの提供が必要となる電子公文書等の一覧を取得し、対象電子公文書等のコンテナメタデータを行政利用向けストレージから取得する。その中のアーカイバルメタデータ部分を抽出し、デジタルアーカイブ用目録データ(EAD)へのマッピング定義に従って変換を行った上で、デジタルアーカイブ連携データフォルダに出力する。

また、同時にコンテナメタデータからデジタルアーカイブ・システム連携用の Dublin Core メタデータへの変換を行い、デジタルアーカイブ連携データフォルダに出力する。

デジタルアーカイブ用目録データ変換、Dublin Core メタデータ変換はバッチ処理として実行する。

プロトタイプシステムでは、デジタルアーカイブ用目録データ変換、Dublin Core メタデータ変換はコンテナメタデータに対して XSLT スタイルシートを適用することで変換する方式とした。

#### (2) デジタルアーカイブ用目録データ編集機能

標準的な Web ブラウザ上で、出力されたデジタルアーカイブ用目録データの内容を編集画面から呼び出し、内容を編集可能とする。

デジタルアーカイブ用目録データ編集はユーザ ID とパスワードによる認証を経た利用者のみが利用できる。

目録データ編集画面までの遷移は、基本的にアーカイバルメタデータ編集画面における遷移と同じである。デジタルアーカイブ用目録データ編集ではメタデータの確定という処理が存在しないため、状態や確定済み件数の表示は行わない。

## (ア) デジタルアーカイブ用目録データ編集画面

階層	トップ > 内閣府 > 内閣府 > 内閣府情報課
文書ID	000001035
作成日時	2009-03-09 18:44:23
DA用目録データ    Dublin Coreメタデータ <a href="#">DA用フォーマットファイル一覧を表示</a>	
必須項目	
タイトル (*)	文庫ファイル0
作成者 (*)	内閣府情報課0
作成年月日 (*)	19980212
言語 (*)	JPN
資料内容 (*)	JPN
レファレンスコード (*)	平21門府00001100
データ種別 (*)	1
目録種別 (*)	2
簿前番号1 (*)	平21門府
簿前番号2 (*)	00001
簿前番号桁数 (*)	100

デジタルアーカイブ用目録データ編集画面では、変換されたデジタルアーカイブ用目録データの情報を編集することができます。

タブをクリックすることでDublin Core メタデータ表示画面への切り替えが行える。また、「デジタルアーカイブ用フォーマットファイル一覧を表示」をクリックするとデジタルアーカイブ用フォーマットファイル一覧画面を表示する。

## (イ) Dublin Core メタデータ表示画面

階層	トップ > 内閣府 > 内閣府 > 内閣府情報課
文書ID	000001035
作成日時	2009-03-09 18:44:23
DA用目録データ    Dublin Coreメタデータ <a href="#">DA用フォーマットファイル一覧を表示</a>	
必須項目	
タイトル	文庫ファイル0
作成者	内閣府情報課0
作成年月日	19980212
言語	JPN
資料内容	JPN
レファレンスコード	平21門府00001100
データ種別	1
目録種別	2

Dublin Core メタデータ表示画面では、デジタルアーカイブ用目録データから変換された Dublin Core メタデータの内容を表示する。Dublin Core メタデータはデジタルアーカイブ用目録データとの整合性を保つため

に編集不可とし、表示のみである。

(ウ)デジタルアーカイブ用フォーマットファイル一覧表示画面



デジタルアーカイブ用フォーマットファイル一覧表示画面では、デジタルアーカイブ用フォーマットに変換された電子ファイルの一覧を表示する。「閲覧/ダウンロード」をクリックすると該当ファイルのダウンロードを行い、ファイル内容を閲覧することができる。

## 6.3.2. 目録データ変換

### (1) デジタルアーカイブ用目録項目へのマッピング定義

プロトタイプシステムでは、メタデータ項目とデジタルアーカイブ用目録項目とのマッピングを表 6-1 の通り定義した。

「対応メタデータ項目 No.」はメタデータ項目定義における番号を表し、「備考」にはマッピングを行うにあたっての変換規則等を記載している。

表 6-1 メタデータ項目のデジタルアーカイブ用目録項目へのマッピング

No.	項目名	デジタルアーカイブ用目録項目要素 ※EAD の /ead/archdesc/dsc 要素以降を 記述	対応メタデータ 項目 No.	備考
1	タイトル(簿冊名または件名)	/c/did/unittitle @label="簿冊標題" "@label="件名"	2.1 または 2.2	簿冊・件名により、label 属性の値を以下の通り記述する。 簿冊 : label="簿冊標題" 件名 : label="件名"
2	レファレンス・コード	/c/did/unitid @label="レファレンス・コード"	22.1	※DAS 固有情報 簿冊の場合 : 簿冊番号1 + 簿冊番号2 + 簿冊番号枝番 件名の場合 : 簿冊番号1 + 簿冊番号2 + 簿冊番号枝番 + 「-」 + 件名番号 + 件名事項 SEQ
3	文書 ID	/c/did/unitid @label="文書 ID"	21	システムで自動付与
4	作成年月日	/c/did/unitdate @label="作成年月日" normal="*"	4.1	normal 属性に 「YYYYMMDD/YYYYMMDD」で開始・終了を記述する。 閏月フラグがセットされている場合、unitdate 要素の PCDATA 内(年月日の先頭)に「[閏]」を記述する。
5	記述レベル	/c @level="*"	22.7 の値により 自動付与	c 要素の level 属性に、以下の値を設定する。 簿冊 : level="file" 件名 : level="item" 件名の場合は、簿冊の c 要素にネストする。
6	作成者名称	/c/did/origination @label="作成部局"	3	
7	受入方法	/c/acqinfo/p[1]	11.1	受入方法コードを設定。 簿冊の場合のみ付与する。

No.	項目名	デジタルアーカイブ用目録項目要素 ※EAD の/ead/archdesc/dsc 要素以降を 記述	対応メタデータ 項目 No.	備考
8	受入年月日	/c/acqinfo/p[1]/date @type="受入年月日" @normal="*"	11.2	受入年月日を normal 属性に「YYYYMMDD」で記述する。Date 要素の PCDATA にも同じ値を記述する。 簿冊の場合のみ付与する。
9	保存場所	/c/did/physloc @label="保存場所"	11.5	
10	移管元府省名	/c/acqinfo/p[2]/corpname @role="移管省庁"	11.3	移管元省庁コードを設定。 簿冊の場合のみ付与する。
11	移管時管理担当部局	/c/acqinfo/p[2]/corpname @role="移管時管理担当部局"	11.4	簿冊の場合のみ付与する。
12	移管年度/年月日	/c/acqinfo/p[2]/date @type="移管年度" @normal="*"	11.6、11.7	移管年度は date 要素の PCDATA として記述する。移管年度の元号はコードで記述する。 Normal 属性に移管年月日を「YYYYMMDD」で記述する。 簿冊の場合のみ付与する。
13	資料内容	/c/scopecontent/note @label="資料内容"/p	10	
14	関連資料、参考文献	/c/relatedmaterial/p	14	
15	文書番号	/c/scopecontent/note @label="文書番号"/p	12.2	受入文書番号
16	公開状況	/c/accessrestrict @type="公開状況"/p	16.1	公開状況コードを設定。
17	公開年月日	/c/accessrestrict @type="公開状況"/p/date @type="公開年月日"	16.2	公開年月日を date 要素の PCDATA として「YYYYMMDD」で記述する。 簿冊の場合のみ付与する。
18	審査状況	/c/accessrestrict @type="審査状況"/p	16.3	審査状況コードを設定。
19	審査年月日	/c/accessrestrict @type="審査状況"/p/date @type="審査年月日"	16.4	審査年月日を date 要素の PCDATA として「YYYYMMDD」で記述する。
20	言語	/c/did/langmaterial/language @langcode="*"	13	langcode 属性に ISO-639-2b のコードを記述する。また、対応するコードを PCDATA に記述する。 簿冊の場合のみ付与する。
21	媒体	/c/did/physdesc/genreform @type="媒体"	6	媒体コードを記述する。

No.	項目名	デジタルアーカイブ用目録項目要素 ※EAD の/ead/archdesc/dsc 要素以降を 記述	対応メタデータ 項目 No.	備考
		”		
22	保管履歴情報	/c/custodhist/p	19	
23	利用履歴情報	/c/bioghist/p	20	
24	文書分類:大分類	/c/did/materialspec @label=“大分類”	1.1	
25	文書分類:中分類	/c/did/materialspec @label=“中分類”	1.2	
26	文書分類:小分類	/c/did/materialspec @label=“小分類”	1.3	
27	データ種別	/c/did/materialspec @label=“データ種別”		※DAS 固有情報 コード「1」で固定。
28	目録種別	/c/did/materialspec @label=“目録種別”	22.7	※DAS 固有情報 簿冊の場合はコード「2」、件名の 場合はコード「3」で固定。
29	簿冊番号1	/c/did/materialspec @label=“簿冊番号1”	22.2	※DAS 固有情報 簿冊の場合のみ付与する。
30	簿冊番号2	/c/did/materialspec @label=“簿冊番号2”	22.3	※DAS 固有情報 簿冊の場合のみ付与する。
31	簿冊番号枝番	/c/did/materialspec @label=“簿冊番号枝 番”	22.4	※DAS 固有情報 簿冊の場合のみ付与する。
32	件名の表示	/c/did/materialspec @label=“件名の表示”		※DAS 固有情報 コード「0」で固定。 簿冊の場合のみ付与する。
33	件名番号	/c/did/materialspec @label=“件名番号”	22.5	※DAS 固有情報 件名の場合のみ付与する。
34	件名事項 SEQ	/c/did/materialspec @label=“件名事項 SEQ”	22.6	※DAS 固有情報 件名の場合のみ付与する。
35	オリジナルフォーマット情報	/c/did/daogrp @title=“オリジナルフォー マット”/daodesc	18.1、18.2.1、 18.2.2、18.2.5、 18.2.6、18.2.8、 18.3.3	オリジナルフォーマットの技術的 情報を提供する。
36	デジタルアーカイブ用フォーマット情報	/c/did/daogrp @title=“デジタルアーカイブ 用フォーマット”/daodesc	18.1、18.4.1、 18.4.2、18.4.5、 18.4.6、18.4.8、 18.4.9、18.4.8.12	デジタルアーカイブ用フォーマット の技術的情報を提供する。

※ DAS…デジタルアーカイブ・システム

※ 対応メタデータ項目は、「エラー! 参照元が見つかりません。」のNo.を表す。

No.35「オリジナルフォーマット情報」、No.36「デジタルアーカイブ用フォーマット情報」については、技術的

メタデータの情報から生成することになるが、個別の項目に該当する EAD の要素が存在しないため、どのように表現するかは今後検討が必要である。

## (2) Dublin Core へのマッピング定義

表 6-1 に定義した項目のうち、Dublin Core メタデータ要素とのマッピングを表 6-2 の通り定義した。「No.」及び「項目名」は表 6-1 における No. に対応している。

なお、プロトタイプシステムにおける Dublin Core メタデータへのマッピング定義に際しては、国立公文書館デジタルアーカイブ・システムにおける EAD から Dublin Core へのマッピング定義を参考とした。

**表 6-2 デジタルアーカイブ用目録項目から Dublin Core 要素へのマッピング定義**

No.	項目名	Dublin Core マッピング
1	タイトル(簿冊名または件名)	title
2	レファレンス・コード	identifier
3	文書 ID	identifier
4	作成年月日	date
6	作成者名称	creator
13	資料内容	description
20	言語	language
21	媒体	format
27	データ種別	subject
28	目録種別	subject

## (3) 目録データ変換処理時間

1 万件分のコンテナメタデータを対象として、デジタルアーカイブ用目録データ及び Dublin Core メタデータへの変換処理にかかる時間を検証用サーバ上で測定した。なお、下記の処理時間はデジタルアーカイブ用目録データ、Dublin Core メタデータの両方の変換処理時間の合計である。

対象データ件数	処理時間
10,000	13 分 42 秒 (822 秒)

上記の結果より、1 件あたりの平均変換処理時間は 0.082 秒である。1 秒あたりでは約 12.2 件、1 時間で約 43,000 件の変換処理が行える計算となる。

### 6.3.3. 横断検索

#### (1) 横断検索における検索処理時間

上記の 10,000 件の Dublin Core メタデータを用いて検索用インデックスを生成し、Z39.50 ターゲット及び SRW/SRU サービス、OAI-PMH リポジトリを稼動させた上で、横断検索における検索処理時間を検証した。

検証にはインフォコム社製「GlobalFinder」を用い、上記の Z39.50 ターゲットを検索対象とする設定を追加して Web ブラウザ上から検索を行った。

(GlobalFinder 検索画面)



※検索条件としてキーワード「内閣府補償課」を指定して検索を実行。

(GlobalFinder データベース別検索結果画面)



※Dublin Core メタデータのインデックスで、10,000 件ヒットしている。

(GlobalFinder 検索結果一覧画面)

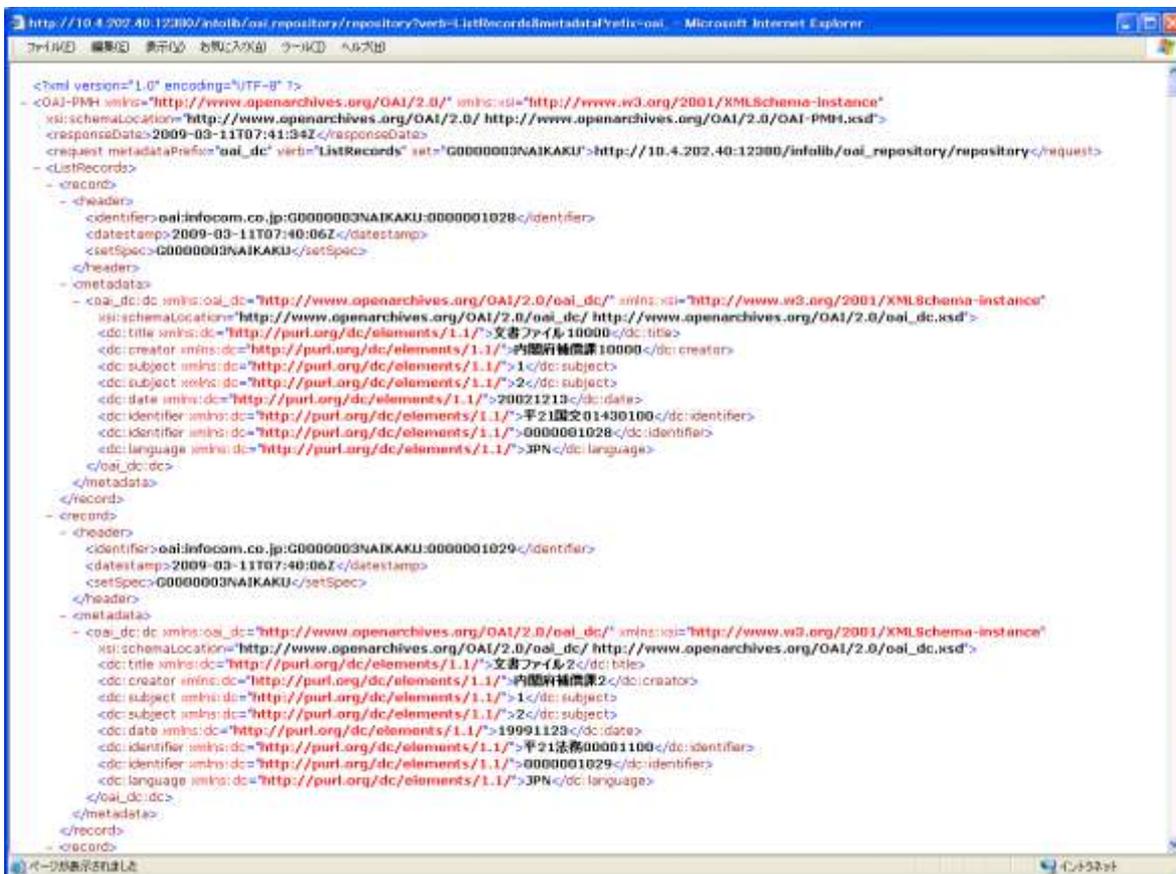


※Dublin Core メタデータの検索結果一覧を表示。

検索処理時間は 1 秒未満であり、非常に良好な結果が得られた。対象となった Dublin Core メタデータがコンテナメタデータと比較して非常に小さいデータサイズであるが、バックエンドで動いている検索エンジン

は行政利用機能と同じエンジンであるため、検索処理のパフォーマンスは行政利用機能と同程度の性能であるといえる。

OAI-PMH のリポジトリについては、直接横断検索の対象とすることはできないため、Web ブラウザからデータを収穫した画面イメージを以下に示す。



(2) 行政利用向け検索機能との比較評価

行政利用機能における行政利用向け検索機能と、横断検索における Z39.50、SRW/SRU、OAI-PMH の各横断検索機能との比較評価結果を以下表に示す。

表 6-3 行政利用向け検索機能と各横断検索機能との比較評価

評価項目	行政利用向け検索機能	Z39.50	SRW/SRU	OAI-PMH
検索性能	検索性能はバックエンドの検索エンジンの性能に依存する。	検索性能はバックエンドの検索エンジンの性能に依存する。 TCP/IP 上のプロトコルに変換されるため、若干のオーバーヘッドが発生することが予想される。	検索性能はバックエンドの検索エンジンの性能に依存する。	検索用のプロトコルではないが、収穫条件を指定された場合の処理はバックエンドの RDBMS、検索エンジンの性能に依存する。
検索機能	フロントエンドのアプリケーションをカスタ	ターゲット自体の検索機能としては Z39.50 のプロ	検索機能には SRW/SRU のプロトコ	OAI-PMH の仕様制限があるため、柔軟な条

評価項目	行政利用向け検索機能	Z39.50	SRW/SRU	OAI-PMH
	マイズすることにより、多様な機能の追加・実装が可能である。	トコル仕様による制限がある。	ル仕様による制限があるが、HTTP ベースのプロトコルであるため外部機能との連携が行いやすいメリットがある。	件による検索は行えない。
表示機能	フロントエンドのアプリケーションをカスタマイズすることにより、多様な表示方式を実現することができる。	Z39.50 のプロトコル仕様に従った表示形式となり、表示機能の拡張は行いにくい。	返戻されるデータフォーマットは任意であるため、表示機能の拡張やカスタマイズの実装は容易である。	収穫されるメタデータフォーマットに制限がある。

OAI-PMH についてはメタデータ収穫(ハーベスティング)のためのプロトコルであるため、他の検索プロトコルと同列に比較することは難しいが、収穫したメタデータを Z39.50、SRW/SRU で検索できる形式に変換する仕組みを導入することで、横断検索のための 1 方式として捉えることも可能である。

### (3) 横断検索における Dublin Core メタデータの有効性

Dublin Core メタデータを対象とした横断検索は、外部機関との連携という観点では非常に有効性の高い機能であるが、本来の電子公文書の長期保存のためのメタデータ(コンテナメタデータ)と比較すると項目が簡略化されており、利用者が求める情報に辿り着けないケースも考えられるため、今後再検討が必要であろう。Dublin Core メタデータへのマッピング対象項目として、公開状況や審査状況を加えることにより、外部からの検索のみで該当電子公文書が公開されているかを確認することができるため、さらに横断検索の有効性が高まるのではないかと考えられる。

行政利用とは異なる利用目的となるため、Dublin Core メタデータの活用方法については国立公文書館デジタルアーカイブ・システムでのサービスのあり方、方向性も含めて、今後検討すべき点である。

## 6.3.4. プロトタイプシステムにおける操作性評価

デジタルアーカイブ目録データ編集画面までの遷移画面は、アーカイバルメタデータ編集画面とほぼ同様のため、評価も同様となる。

デジタルアーカイブ用目録データ編集では、システム上メタデータとは別のデータに対する編集作業となり、アーカイバルメタデータ編集画面で修正した内容はデジタルアーカイブ用目録データには自動的に反映されない。利用者の観点からはその違いが分かりにくいと、説明があるほうが望ましいといえる。同様に Dublin Core メタデータ表示画面は整合性維持のため表示のみとなっている点も説明が必要と思われる。

なお、本システム稼動に向けては、一般の利用に供されている電子公文書についてアーカイバルメタデータ編集画面等での修正内容について、デジタルアーカイブ側の目録情報に反映させる方法等を検討する必要がある。

## 6.4. デジタルアーカイブ・システム変更調査報告

### 6.4.1. デジタルアーカイブ・システム側の変更点

今回のプロトタイプシステム上でのデジタルアーカイブ連携機能の検証結果を踏まえ、本システムの稼動に向けて国立公文書館デジタルアーカイブ・システム側での変更が望まれる、または検討すべき点を以下にまとめる。

#### (1) 電子公文書における簿冊・件名の扱い

プロトタイプシステムでは、移管時のメタデータとして行政文書ファイル管理簿の情報を基礎としているが、受入・保存管理機能の調査・検証においても触れている通り、電子公文書1件が行政文書ファイル1レコードという対応となっている。実際には行政文書ファイルの下には行政文書という単位があるはずであるが、現状では移管時の媒体・データ構成について明確なルールが存在しないため、システム上で機械的に行政文書単位に分割することが難しい。

このことはデジタルアーカイブ連携において、現在のデジタルアーカイブ・システムでの公文書の簿冊・件名という構造を再現することができないという問題と結び付くものである。ただし、紙媒体での移管を前提としている現在のデジタルアーカイブ・システムの簿冊・件名という概念を、電子公文書の移管においてもそのまま適用できるか、という点については今後検討すべきであると考えられる。

デジタルアーカイブ連携機能では便宜的に全てのデータを簿冊として扱うこととしたが、編集画面上で件名への分割機能を実装することは技術的に難しいことではない。しかし、件名分割の単位を機械的に判断できない場合、その判断及び作業は国立公文書館職員が行うことになり、大きな負担増となることが懸念される。

#### (2) 資料群階層情報の付与

プロトタイプシステムでは、公文書の階層性を現用文書管理のための行政文書ファイル管理簿の大分類・中分類・小分類により再現しているが、デジタルアーカイブ・システムでは、資料の内容・特性等も考慮して、利用者の利便性に配慮した独自の階層構造となっている。このため、連携データ作成時に資料群階層の再付与が必要となる。

資料群の再付与は、現行の一括変換機能で対応可能であるが、将来的には、デジタルアーカイブ連携機能のデータ出力時に出力先資料群を設定できる等の仕組みを実装することが考えられる。

#### (3) 目録データ項目の追加

表 6-1 が示すように、デジタルアーカイブ・システムでの目録データ項目の追加が必要である。特に移管された電子公文書とデジタルアーカイブ・システム上の目録とを結び付けるための重要な情報である文書IDについては、デジタルアーカイブ・システム側でのユニークキーとの関連等にも配慮して付与方法等を検討する必要がある。

#### (4) 長期保存フォーマット(PDF/A)への対応

現在の国立公文書館デジタルアーカイブ・システムでは、公文書のデジタル画像情報は JPEG 2000 形式

で提供することが前提となっており、PDF/A 形式のデータをそのまま登録・公開するための仕組みが存在しない。デジタルアーカイブ・システム側で PDF/A の登録・公開を可能とするための対応が今後必要となる。

PDF/A をデジタルアーカイブ・システム側でも保持することは管理上の煩雑さ、コスト増を招く恐れがあるため、デジタルアーカイブ用フォーマットの電子ファイルを公開する仕組みを別に用意し、URL のみで連携させることも検討の余地はあろう。

#### (5) 検索結果表示画面等について

今回の調査においては検討対象外であるが、電子公文書については、電子であることにより提供すべき情報が項目として追加されている。紙媒体と電子媒体を一体として検索し、その結果を表示する場合のデジタルアーカイブ・システムでの表示画面については、わかり易いレイアウト等、表示方法について検討する必要がある。

## 6.5. 結論

### 6.5.1. デジタルアーカイブ用フォーマット変換機能

#### (1) デジタルアーカイブ用フォーマットへの変換

現在の国立公文書館デジタルアーカイブ・システムでは、画像データフォーマットとして JPEG 2000 と PDF が採用されており、いずれも長期保存フォーマットであるため、プロトタイプシステムでは実質的なフォーマット変換は行わないこととした。

しかし、今後国立公文書館デジタルアーカイブの画像データフォーマットが変更される可能性も考慮した仕組みとする必要がある。

#### (2) マスキング処理

国立公文書館デジタルアーカイブでの一般利用にあたっては、電子公文書の内容に公開に不適切な情報(個人情報など)が含まれている場合がある。デジタルアーカイブ用フォーマットについて、そのような情報をマスキングするための実現方式を調査した。

JPEG 2000 の場合は、ビットマップデータの画像であるため、対応した画像編集ソフトを使って該当箇所を塗りつぶす、切り取るなどの操作を行うことにより、マスキングを実現可能である。

PDF の場合は、表面的に該当箇所を塗りつぶしたとしても、内部で保持されているテキスト情報が保持されたままとってしまう問題がある。最新のアプリケーションでは、テキスト情報も含めてデータ上から消去する機能が提供されているもの (Adobe Acrobat 9 Professional の「墨消し」機能など) もあるが、見た目上で容易に確認できないため、作業には注意が必要となるため、マスキングが必要なケースでは PDF ではなく画像フォーマットとして処理する方が望ましいと考えられる。

いずれのフォーマットについてもマスキングについて技術的な課題はクリアされているといえるが、本システムの運用にあたっては作業及び確認手順、必要となるツール・ソフトウェアの選定基準といったルールを国立公文書館において明確にしておく必要がある。

### 6.5.2. デジタルアーカイブ用目録データ作成機能

#### (1) 目録データの自動生成

デジタルアーカイブ用目録データの自動生成は、アーカイバルメタデータにおいて項目が網羅されており、基本的に自動生成は可能である。デジタルアーカイブにおけるレファレンス・コードも、移管元府省、移管年度から自動的に生成することが可能であるが、デジタルアーカイブとの不整合や重複が発生しない仕組みとする必要がある。

#### (2) Dublin Core へのマッピング定義

プロトタイプシステムでは、Dublin Core へのマッピングについては国立公文書館デジタルアーカイブと同じになるよう定義した。

#### (3) 目録データの扱い

デジタルアーカイブ用目録データは GUI 上で変更することが可能であるが、それにより受入・保存機能(または行政利用機能)で管理されるアーカイバルメタデータとの差異が発生してしまう。デジタルアーカイブ用目録データの位置付けについては検討が必要である。

## 7. 業務実施要件報告

本章では、今回の調査研究の結果を踏まえて、「電子公文書等の管理・移管・保存・利用システム」の構築及び運用に向けての業務実施要件をまとめる。

### 7.1. 業務フロー

電子公文書等の移管から、各府省での検索・閲覧(行政利用)までの想定される業務フローを図 7-1 に示す。

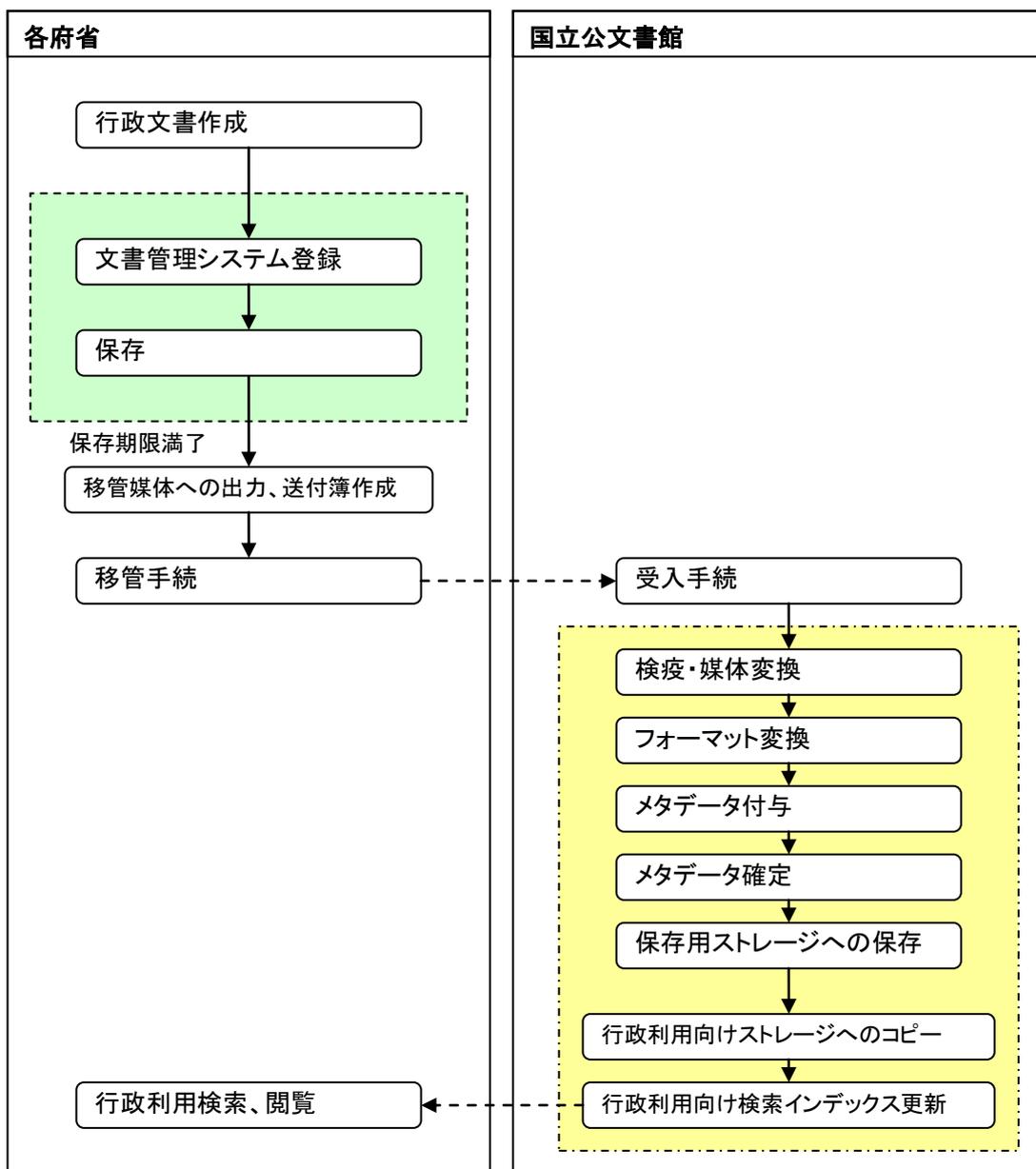


図 7-1 電子公文書等の移管から行政利用までの想定業務フロー

図 7-1 の業務フローでは、各府省からの電子公文書等の移管から、国立公文書館での受入・保存、各府省での行政利用までは一つのルートのみを前提としている。各府省から移管される電子公文書自体が改版されることはないはずだが、受入後のフォーマット変換における変換ミスや技術的なトラブル、メタデータ付与における誤入力等があった場合に、各府省からの指摘や相談を受ける窓口を国立公文書館側に設け、それらに対応するための再変換やメタデータ修正といった作業も業務フローに含めて考える必要があろう。

次に、国立公文書館での保存プロセスにおいて今後想定される業務フローを図 7-2 に示す。

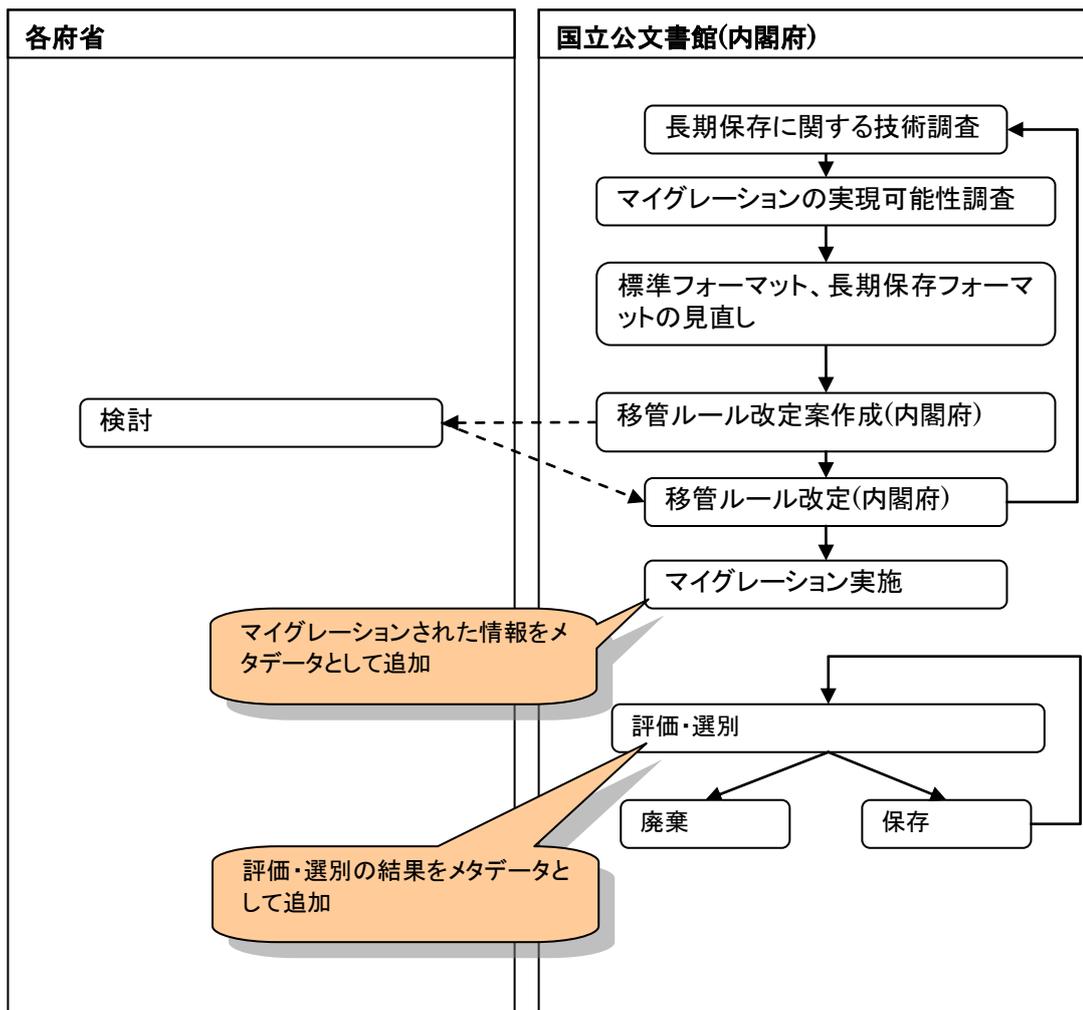


図 7-2 電子公文書等の保存プロセスにおける想定業務フロー

今回の調査では長期保存のプロセスまでは対象としなかったが、保存プロセスにおける業務フローとしては、長期保存に関する技術調査を継続的に実施し、移管ルールとして定める標準フォーマット、長期保存フォーマットがその時点で長期保存、行政利用、一般利用のそれぞれの観点において妥当であるかの判断を行うことが必要である。仮にそれまで標準フォーマットあるいは長期保存フォーマットとして採用していたフォーマットの作成・閲覧が近い将来において困難となることが予想されるのであれば、代替となるフォーマットを速やかに選定し、マイグレーションが技術的に可能であるかを検討しなければならない。その上で各府省の意見も集約し、移管ルールの改定とマイグレーションの実施が求められる。マイグレーションに関する

情報もメタデータとして追加していく必要がある。

さらに、保存プロセスにおいては長期保存されている電子公文書等の評価・選別も必要となる。記録管理プロセスとして考えれば評価・選別の結果もメタデータとして保存されるべきである。廃棄が決定された場合でも、メタデータは維持されることが期待されるが、電子ファイルの情報を削除するか否かは電子公文書の概念や考え方と合わせて今後の議論が待たれるところである。

最後に、保存から一般利用者への公開までにおいて想定される業務フローを図 7-3 に示す。

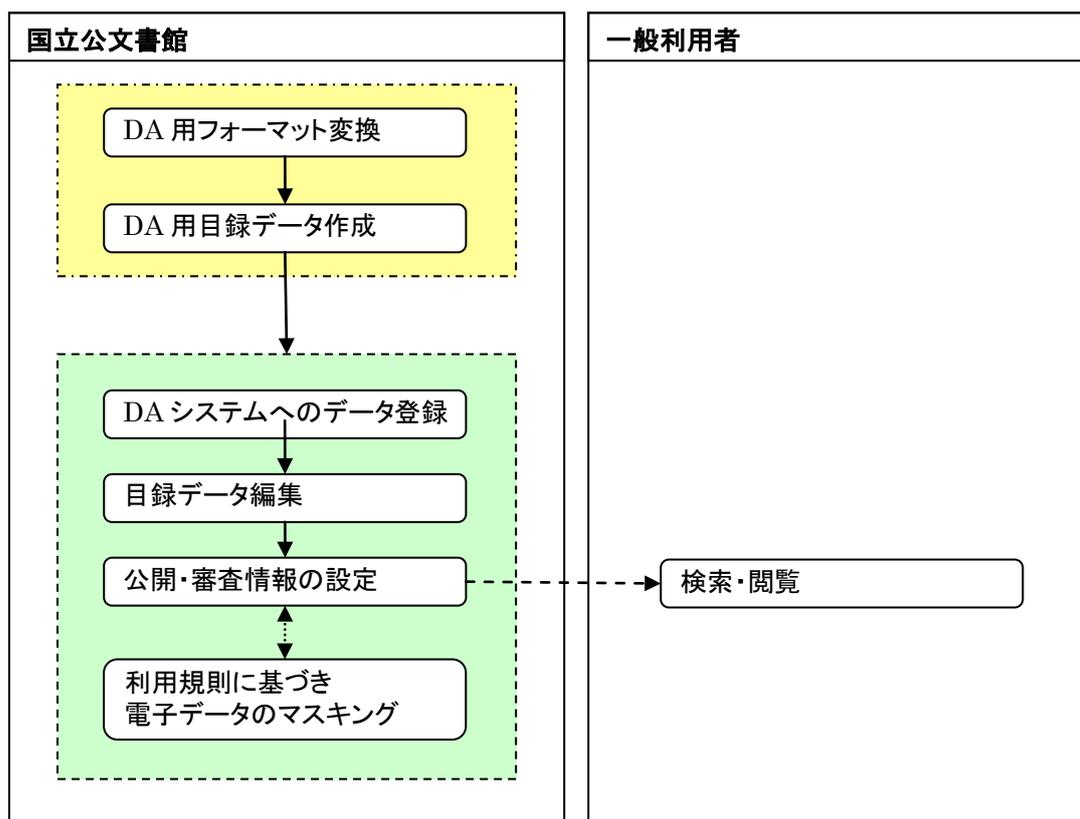


図 7-3 電子公文書等の保存から一般の利用に供するまでの想定業務フロー

電子公文書の一般の利用は、国立公文書館デジタルアーカイブ・システム(以下「DAS」と記載する)において実現されるため、今回の調査では DAS との連携のためのデータ出力(フォーマット変換、目録データ作成)とマスクングの実現のための問題点の調査までを範囲とした。

DAS による、一般利用者への提供にあたっては、国立公文書館利用規則上一般の利用を制限する情報についてマスクングを実施し、提供する流れとなる。紙媒体等ではマスクングされたコピーなどで提供されるため、原本はそのまま残るが、電子公文書等ではマスクングされた情報が公開となった時点で長期保存フォーマットから再度 DAS 用フォーマットへの変換を行う必要がある。

## 7.2. 業務実施にあたっての問題・課題

### 7.2.1. 電子公文書における「原本」の概念

各府省向けのデモンストレーションにおいて、移管された電子ファイルの原本性、真正性をどのように担保するのか、という意見が多く挙げられた。

電子公文書(狭義にはそれを構成する電子ファイル)においては、「原本」が何を指すのかという点についての議論に明確な回答が出されていないのが現状である。紙媒体では例えば押印されているものが原本である、と考えることができる。一方で電子公文書においては、今回のプロトタイプシステムにおいては「オリジナルフォーマット」「長期保存フォーマット」が存在しており、いずれに原本性を認めるかという問題がある。また、電子ファイルは容易にコピーを作ることができ、移管元においても複数のコピーが存在するようなケースもありうる。これらの場合において原本性をどのように考えるかは、引き続き検討と意見の統一が求められるところである。

### 7.2.2. 受入の前提条件(電子公文書等の受入のためのガイドライン案)

電子公文書等の移管・受入をスムーズに実現するために、以下のガイドラインを提案する。

#### (1) 移管・受入対象フォーマット

- 移管・受入対象となる電子ファイルの種類については原則制限を設けない。ただし、圧縮ファイル形式は対象外とする。
- 標準フォーマット以外の電子ファイルも移管対象となるが、長期保存の観点から標準フォーマットに準じて作成することが推奨される。

#### (2) 移管時の電子媒体

- 移管される媒体は以下のいずれかとする。いずれもファイルシステムは NTFS とする。
  - CD-R、CD-ROM、CD-RW
  - DVD±R、DVD±RW、DVD-RAM
  - MO
  - USB 接続のハードディスク等の記録媒体(ただしセキュリティの観点よりフラッシュメモリは対象外とする)
- 以下は移管対象とするが、容量及び読み書き速度の面から推奨しない。
  - FD(フロッピーディスク)

#### (3) 移管時のメタデータ

- 移管時の最低限のメタデータとして、行政文書ファイル管理簿のデータを添付する。
- 「一元的な文書管理システム」が導入を想定し、その連携方法も今後検討する。

#### (4) 電子媒体におけるフォルダ及びファイル構成

- 移管される電子媒体内のフォルダ構成は、以下の規則に準じた構成とすることが望ましい。

第 1 階層	第 2 階層	第 3 階層	説明
行政文書ファイル レベル			行政文書ファイル 1 件に相当する。フォルダ名と行政文書ファイル名が一致することが望ましい。
	行政文書レベル		行政文書 1 件に相当する。フォルダ名と行政文書名が一致することが望ましい。
		文書(ドキュメント) レベル	いわゆる公文書における「件名」に相当する。

- 移管時媒体の最上位フォルダ(ルートフォルダ)の直下に第 1 階層を配置する。第 1、第 2、第 3 階層は複数存在(繰り返し)してもよい。
- 第 3 階層までは第 4 階層以降については特に制限を設けない。
- 各電子ファイル名については、同一階層内で順序性の再現が必要な場合は先頭に連番を振ることが望ましい。

#### (5) ファイルのセキュリティ設定

- 標準フォーマットとして移管される電子ファイルについては、セキュリティ設定(パスワードによる保護)を解除した状態とする。  
※セキュリティ設定が行われているかどうかを移管時に自動的にチェックし、移管元にリストとして提供できるような仕組みを本システム側で用意することが望ましい。
- その他のフォーマットについても、パスワードによる保護や暗号化がされているファイルは移管対象外とする。

#### (6) 受入可能なフォント

- 長期保存の観点から、少なくとも標準フォーマットのうち文書系フォーマットとして作成される電子ファイルについては、OS 標準のフォントのみを使用することが望ましい。
- それ以外のフォントを使用する場合は、使用されているフォント情報を移管と同時に提供するように各府省に協力を求めることが望ましい。

#### (7) 外字の扱い

- 外字については、全府省共通の外字管理システムが導入されることが望まれるが、今後数年のうちに実現される可能性は極めて低い状況である。
- 外字については文字図形番号を文書内に埋め込む方式を推奨する。記述方式は EGIX に準拠することが望ましいが、文書の見読性が損なわれないよう、各府省の意見も踏まえた上で最適な記述方式を今後検討する。
- 独自の外字を文書内に埋め込む場合は、外字図形情報を画像データとして同時に添付する等のルールを今後検討すべきである。

#### (8) 紙媒体から電子化された電子公文書等について

- 電子公文書の範囲には、ポーンデジタルだけでなく、紙媒体の文書をスキャンして電子化するケースも想定される。スキャンされたデータを受け入れる場合は、見読性を向上させるために文書のフォントサイズ、ルビのサイズ、ドット密度、圧縮率などを決めておく必要がある。
- 以下の項目について、e-文書法において省令として定めている府省もあるため、それら府省の意見も踏まえた上で一定の基準を設けることを提案する。
  - 文書作成時の最小フォントサイズ
  - 文書作成時のルビのサイズ
  - スキャン時のドット密度
  - スキャン時の圧縮率

# 8. 全体総括

## 8.1. 今年度の調査結果のまとめ

### 8.1.1. 今年度構築したプロトタイプシステムについて

今回の調査研究の実施にあたり構築したプロトタイプシステムの概要図を図 8-1 に示す。

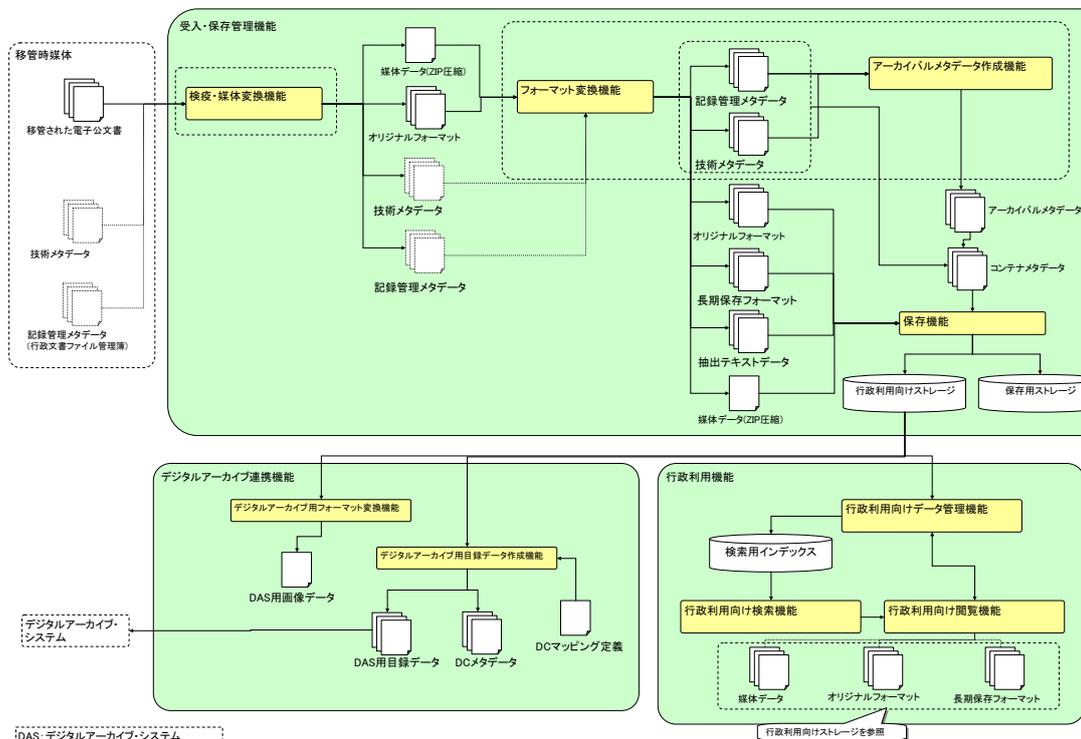


図 8-1 プロトタイプシステム概要図

プロトタイプシステムは受入・保存管理機能、行政利用機能、デジタルアーカイブ連携機能の 3 つの機能により構成される。

受入・保存管理機能では、電子公文書等の受入・保存管理、メタデータ管理を行う。受入・保存機能は表 8-1 に示す 4 つの下位機能により構成される。

表 8-1 受入・保存管理機能の下位機能

下位機能名称	説明
検疫・媒体変換機能	受入対象となる電子公文書等の格納された媒体に対してウイルスチェックを行い、システムへのウイルスの侵入を防御する。また、媒体からシステム上の一時媒体(ディスク)への電子ファイルのコピーを行

下位機能名称	説明
	う。
フォーマット変換機能	<p>検疫・媒体変換機能を通過した電子公文書(オリジナルフォーマット)から、長期保存フォーマットへの変換を行う。文書フォーマット、表計算フォーマット、プレゼンテーションフォーマットの場合はテキストデータ抽出を行う。</p> <p>同時にオリジナルフォーマットのファイル情報から技術的メタデータの自動生成、行政文書ファイル管理簿のデータから記録管理メタデータの自動生成を行う。</p>
アーカイバルメタデータ作成機能	<p>技術的メタデータ、記録管理メタデータを RDBMS に登録するとともに、アーカイバルメタデータを自動生成して RDBMS に登録する。</p> <p>また、標準的な Web ブラウザ上で、登録された技術的メタデータ、記録管理メタデータ、アーカイバルメタデータの内容を編集画面から呼び出し、編集を行える。</p>
保存機能	<p>作業中の各種メタデータ、ファイルを確定することにより、保存用ストレージへのデータ移動(出力)を実行する。</p>

行政利用機能では、受入・保存管理機能で保存された電子公文書等及びメタデータの検索・閲覧のための機能を提供する。

行政利用機能は表 8-2 に示す 3 つの下位機能により構成される。

表 8-2 行政利用機能の下位機能

下位機能名称	説明
行政利用向けデータ管理機能	<p>保存用ストレージに格納されたコンテナメタデータと各ファイルを行政利用向けストレージにコピーする。</p> <p>また、行政利用向けストレージに格納されたコンテナメタデータと抽出テキストデータから、行政利用向け検索機能で使用する検索用インデックスの生成を行う。</p>
行政利用向け検索機能	<p>Web ブラウザ上で、行政利用者が指定した条件で検索を行い、検索結果一覧と詳細情報を提供する。</p> <p>検索画面として、</p> <ul style="list-style-type: none"> <li>● 任意文字列検索画面(抽出テキストデータ及びコンテナメタデータすべてを対象とした全文検索)</li> <li>● 詳細検索画面(コンテナメタデータのうち、特定の項目のみを対象とした検索)</li> <li>● 階層検索画面(階層をツリー状に表示し、展開することで該当階層に属する公文書を検索)</li> </ul> <p>の 3 種類が提供される。</p>
行政利用向け閲覧機能	<p>Web ブラウザ上で、行政利用向け検索機能の検索結果一覧、詳細</p>

下位機能名称	説明
	表示から、行政利用者が指定した電子公文書等のファイル一覧、ファイルデータを提供する。

デジタルアーカイブ連携機能では、行政利用機能で管理される電子公文書等を、国立公文書館デジタルアーカイブで一般の利用に供するためにフォーマット変換や目録データ生成を行う。

デジタルアーカイブ連携機能は表 8-3 に示す 2 つの下位機能により構成される。

表 8-3 デジタルアーカイブ連携機能の下位機能

下位機能名称	説明
デジタルアーカイブ用フォーマット変換機能	行政利用向けストレージに格納された長期保存フォーマットから、デジタルアーカイブ用フォーマットへの変換を行う。
デジタルアーカイブ用目録データ作成機能	電子公文書等のコンテナメタデータを行政利用向けストレージから取得してデジタルアーカイブ用目録データ(EAD)及び Dublin Core メタデータへの変換を行う。 また、Web ブラウザ上で、デジタルアーカイブ用目録データの内容を編集画面から呼び出し、内容の編集を行える。

### 8.1.2. プロトタイプシステムによる検証結果のまとめ

前項で説明したプロトタイプシステムを用いて行った各種検証の結果の概略を以下にまとめる。

#### (1) 受入・保存管理機能検証結果

受入・保存管理機能においては、まずプロトタイプシステムで用いるメタデータ項目及びスキーマの定義を行い、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各メタデータについて、自動生成が可能であるかの検証を行った。結果として、記録管理メタデータは行政文書ファイル管理簿の情報から取得できる項目については自動生成が可能であった。技術的メタデータはファイルのプロパティ情報から一部は自動的に生成できるが、プロパティから取得できない情報については移管時にメタデータとして付与されることが望ましい。アーカイバルメタデータについては移管後の管理のための情報であるため、自動生成できない情報は多いが、記録管理メタデータから流用することである程度カバーできることが分かった。

フォーマット変換機能については、オリジナルフォーマットから長期保存フォーマットへ変換した場合に欠落する機能やレイアウトへの影響を検証した。結果として、文書フォーマットでは一部機能が利用できなくなるケースや、オリジナルフォーマットのフォントが完全に再現されなかったり、レイアウトが崩れたりするケースがあることが分かった。音声・音楽、映像については変換に失敗するケースがあり、技術革新の目覚ましい領域においては、長期保存フォーマットの選定が難しい。

#### (2) 行政利用機能検証結果

行政利用向けデータ管理機能では、保存用ストレージに格納されたデータを、行政利用向けストレージにコピーする際の方法と、セキュリティ面での検討が必要である。

行政利用向け検索機能では、任意文字列検索、詳細検索、階層検索の3方式の検索用インタフェースを用意した。詳細検索で指定できる項目や、検索結果一覧での表示項目は各府省の職員からの要望を踏まえて検討する必要がある。検索機能では文字の正規化により検索時の網羅性を向上させたり、検索インデックスを二重化するなどの方式により、更新時の検索レスポンス低下やサービス停止時間を避ける仕組みを考慮しなければいけないことが分かった。

閲覧機能では、オリジナルフォーマット、長期保存フォーマットのダウンロードに加えて、移管時の媒体におけるフォルダ構成などに依存するファイルを再生するための手段として、媒体データのダウンロードも可能な仕組みとした。

行政利用向け検索・閲覧では、該当電子公文書の作成(移管)元組織のみが閲覧できるように制限を設ける必要がある。プロトタイプシステムでは内部で利用者管理を行い、所属組織の情報をメタデータに付与することで閲覧制限を行う仕組みとしたが、実際のシステム運用では、国立公文書館が各府省の行政利用者を個別に管理することは難しいため、全府省の統一的な利用者認証システムと連携できることが望ましいといえる。

### (3) デジタルアーカイブ連携機能

デジタルアーカイブ用フォーマット変換機能については、電子公文書の内容に公開に不適切な情報(個人情報など)が含まれている場合に情報をマスキングするための実現方式を調査した。プロトタイプシステムでデジタルアーカイブ用フォーマットとしたPDF/AとJPEG 2000のいずれについても、技術的な課題はクリアされているが、システムの運用にあたっては作業及び確認手順、必要となるツール・ソフトウェアの選定基準といったルールを明確にしておく必要があることが分かった。

デジタルアーカイブ用目録データの自動生成では、アーカイバルメタデータにおいて国立公文書館デジタルアーカイブ・システムで使用される項目がおおよそ網羅されているため、基本的に自動生成は可能であるといえる。しかし公文書の階層情報の差異をどのようにして吸収するか、また、デジタルアーカイブ・システムで公開される目録情報と、受入・保存対象として管理されるメタデータについて、不整合を生じることなくスムーズに連携させる仕組みなどは、将来的に解決すべき課題である。

### 8.1.3. 電子公文書のフォーマット

今回の調査研究におけるプロトタイプによる総合的検証では、記録様式ごとに、受入れ時の「標準フォーマット」又はオリジナルフォーマット及び保存のための「長期保存フォーマット」を設定した。したがって、1つの電子ファイルにつき、「標準フォーマット」又はオリジナルフォーマットのほか、「長期保存フォーマット」(変換可能な場合)及びテキスト抽出データを電子公文書等の移管・保存・利用システムの管理対象とする想定で検証を進めることとなった。つまり、1つの電子ファイルにつき、最大で3つのフォーマットを保存する想定である。

3つのフォーマットを保存する意義・趣旨は何であろうか。一般的に言って、複数のフォーマットを保存する意味はリスクの分散である。例えば、「標準フォーマット」又はオリジナルフォーマットが長期に保存できなかった場合でも、「長期保存フォーマット」によって保存・利用の可能性を維持できる。逆に、「長期保存フォーマット」が長期に保存できなかったとしても、「標準フォーマット」又はオリジナルフォーマットによって保存・利用の可能性を維持できる。むしろ、絶対的に保存・利用の可能性が維持できるとは言えないが、複数のフォーマットを保存することが保存・利用の可能性を喪失するリスクを低減することにつながる。

では、複数のフォーマットを保存するという観点とは別に、「標準フォーマット」又はオリジナルフォーマット、「長期保存フォーマット」及び抽出テキストデータの保存は、それぞれ、どのような意味を持っているか。

「標準フォーマット」又はオリジナルフォーマットは、まさに、電子公文書等のオリジナルとして作成時の意図を最もよく再現するフォーマットであり、電子公文書等の作成者の意図を示す証拠として保存する意味がある。ただし、オリジナルフォーマットの中には特定のOS、アプリケーション等のシステムへの依存性が高いものもあり、長期的には、依存するシステムの陳腐化によって利用できなくなるおそれがあることに留意する必要がある。見方を変えれば、オリジナルフォーマット自体がそのような依存性が低い標準的フォーマットである方が望ましいとも言える。

次に、「長期保存フォーマット」は、長期保存の高度の安定性・効率性等の確保に配慮した標準的フォーマットを選定できれば、長期の保存・利用の可能性を高めることが出来る。また、今回の調査研究においては、ワープロ・フォーマット、表計算フォーマット及びプレゼンテーション・フォーマットの電子公文書等を単一の「長期保存フォーマット」へ変換した。このような方法を採用することができれば、フォーマットの種類を絞り込むことによって、多様なフォーマットを保存・管理することで生じるリスク、コスト等を低減することが可能になる。ただし、「長期保存フォーマット」への変換によってオリジナルフォーマットで実現していた機能等が損なわれるおそれがある。したがって、「長期保存フォーマット」としては、電子公文書等の記録としての価値を維持するのに不可欠な「エッセンス」を保存することができるものを選定しなければならない。

これらとは別に、テキスト抽出データは、電子公文書等の本体というよりはむしろ検索用のメタデータの一つとして保存するという点に意味を見出すことができる。

なお、「長期保存フォーマット」への変換は電子公文書等の保存期間満了時の移管した後に国立公文書館が行うというのが、今回の調査研究の前提であった。だが、「長期保存フォーマット」への変換を移管前に行うということも考えられる。特に、ワープロ等のアプリケーションで作成し、当該アプリケーションでの編集可能性を維持する必要がない場合については、現用段階から、あらかじめ固定的な「長期保存フォーマット」に変換して管理すれば、電子公文書等の意図せざる又は不正な改変等のリスクを低減することが可能になるというメリットもある。

以上の技術的な検討とは別に、複数のフォーマットを法制度上でどのように位置付けるかという問題もあ

る。例えば、「標準フォーマット」又はオリジナルフォーマットも「長期保存フォーマット」も電子公文書等の原本として扱うのであれば、移管を受けた国立公文書館は永久的に複数のフォーマットを保存し続ける責任を負うことになる。また、一般の利用に供する場合は、保存システム内からコピー（複製物）を抽出して、一般利用に適する形式（フォーマット、メタデータ等）に変換することが考えられるが、その際、「標準フォーマット」又はオリジナルフォーマットから変換するのか、それとも、「長期保存フォーマット」から変換するのかといった問題も生じると考えられる。ただし、法制度上のフォーマットの位置づけは、ある種の「社会的合意」形成が必要であり、技術論だけで片づけることはできない。

#### 8.1.4. メタデータスキーマ

今回の調査研究では、電子公文書等をシステム上で管理するためのメタデータについて、メタデータの類型ごとに既存の標準等をモジュール的に採用し（一部独自定義を含む。）組み合わせ活用するという考え方に基いてスキーマを作成した。具体的には、メタデータ全体を包括するメタデータであるコンテナ・メタデータとして METS を、電子公文書等のファイル形式、作成アプリケーション等を記述する技術的メタデータとして PREMIS を使用した。また、電子公文書等の作成日、作成者、タイトル等作成機関で生成される記録管理メタデータとして、行政文書ファイル管理簿の項目等をダブリン・コア (Dublin Core) にマッピングしつつ、一部独自拡張したメタデータを使用した。さらに、電子公文書等の移管後に国立公文書館で長期に保存し一般の利用に供するために必要なアーカイバルメタデータとして、EAD を採用した。

従来、主に図書館の世界で電子情報をインターネット等で共有する際の情報の検索・特定の効率向上、相互運用性 (interoperability) 確保等のために、各種のメタデータスキーマが策定されてきた。また、電子オブジェクトの保存の観点から策定されたメタデータスキーマも存在する。

しかしながら、電子記録の長期保存という分野自体が世界的に見てもまだ「揺籃期」にある現時点では、電子公文書等を長期に安定的に保存し一般の利用に供するために必要と考えられるエレメントを網羅し、かつ相互運用性等にも配慮した単一のメタデータスキーマは存在していない。したがって、平成 23 年度から運用開始を予定している電子公文書等の移管・保存・利用システムにおいては、今回の調査研究と同様のアプローチにより、複数のメタデータスキーマをモジュール的に組み合わせ、必要に応じて独自定義エレメントを追加して、独自のスキーマを策定することが適切であろう。ただし、この場合、メタデータスキーマ自体を適切に定義すると共に、管理対象である電子公文書等と同様に、メタデータスキーマも長期保存し、メタデータ自体の長期のアクセス可能性、理解可能性等を保持する必要がある。

なお、今回の調査研究の対象に含まれていなかった要素として、電子公文書等を 30 年、50 年、100 年、500 年等保存を継続していく過程では、例えば、ある長期保存フォーマットから別の長期保存フォーマットへ変換する必要が確実に生じる。また、今のところ、長期保存フォーマットへの変換は移管・受入れ後に国立公文書館が実施する想定であるが、将来的には作成省庁が事前に変換するようになることも考えられる。さらに、法的枠組みの変化に伴う組織改編等により、「国立公文書館」という名称が変更される等の事態も想定されないわけではない。このような将来の枠組みの変化や保存プロセスの継続による作業の累積などに対応可能なメタデータスキーマを策定する必要がある。

## 8.2. 次年度以降の検討課題

### (1) 移管媒体、フォルダ構成についてのルール化

現状では、国立公文書館に移管される電子公文書等の媒体や電子ファイル、フォルダ構成についての取り決めが存在しないため、システム稼動に向けたルール化が望まれるが、長期的な媒体規格への柔軟な対応、各府省のコンセンサスを得ることが重要である。

### (2) 文書作成時のガイドライン

各府省において文書を作成するにあたり、標準フォーマットやファイルセキュリティ設定、フォントの使用方針といったガイドラインの制定と周知、運用の徹底が望まれる。

例えば、文書ファイルがパスワードにより保護されている場合、システム上では長期保存フォーマットへの変換が行えず、保存することができない。行政利用向け閲覧でもパスワードにより保護された状態で提供されるため、実質的に内容を閲覧することはおろか、エッセンスの抽出も不可能となる。

### (3) 非標準フォーマットの電子公文書等の移管

標準フォーマットではないファイルフォーマットについては、フォーマットに関わらずオリジナルフォーマットを長期保存としたが、プログラムやデータベース等のファイルでは、ファイル単体では利用できないケースもある。そのようなケースを想定し、電子公文書ごとの媒体データ(媒体に含まれるすべてまたは一部の電子ファイル)を、フォルダ構成も含めて保持しておくことが推奨される。

Web サイトの場合は、外部サイトを参照しているバナー画像やリンクについて、サイト自体の見読性を優先する(見た目の再現を優先する)か、完全性を優先する(本来サイト上に配置されているファイルのみを保存しておく)かについては今後の議論が待たれるところである。メールについては移管対象となる範囲、添付ファイルの扱いなど、技術面・制度面の両方の観点から検討すべきである。

### (4) 原本性の考え方

プロトタイプシステムでは、保存対象となる電子ファイルはオリジナルフォーマット、長期保存フォーマット、抽出テキストデータ、媒体データの最大 4 種類となる。長期保存の観点から見ると、各データを保管するためのディスクストレージの容量もさることながら、1 つの電子公文書ファイルについて複数のデータを維持・管理していくことのコストやについての検討が必要である。

### (5) フォント

フォントの問題は文書作成、フォーマット変換、閲覧の各段階において、OS やアプリケーションといった環境の違いにより発生しうる。移管時のフォーマットを PDF 形式とし、フォントデータを埋め込むことで環境に依存せず文字図形を適切に再現することは可能であるが、フォント自体の著作権の問題をクリアしなければならない。

文書作成ガイドラインにおける標準フォント使用の徹底や、フォント情報を一元的に管理・配信する統合フォントサーバの運用など、制度面での整備が望まれる。

(6) 文書作成時のメタデータ付与

現状では、移管時の電子公文書等に関するメタデータとして利用できる情報は行政文書ファイル管理簿のみである。行政文書ファイル管理簿のデータ項目は記録管理メタデータまたはアーカイバルメタデータとして使用するには必ずしも十分ではない。現在導入が進められている「一元的な文書管理システム」から電子公文書の移管が実施されるようになれば、より適切なメタデータを自動的に付与することができる可能性がある。

(7) 行政文書ファイル管理簿と媒体(または電子ファイル)の関連性

移管時の媒体において、1 件の電子公文書が媒体中の一部のデータ(フォルダ)のみを対象としている(すなわち、1 つの媒体の中に複数の電子公文書が含まれる)場合があるが、行政文書ファイル管理簿ではレコードに対するファイル(またはフォルダ)の情報が存在しないため、それらの関連性を機械的に判断することができない。

上記の一元的な文書管理システムの導入により改善が望まれる。

(8) 公文書の持つ階層性

公文書ではファイル、簿冊、件名といった構造上の階層性と、移管元府省内での組織構造の階層性が含まれていると考えられる。電子公文書においては、電子ファイルのフォルダ階層も重要となる。

行政文書ファイル管理簿では、大分類・中分類・小分類という 3 段階の分類情報が存在し、これを元にある程度までは自動的に階層情報をメタデータに付与することが可能である。しかし、分類情報は組織階層を表す場合もあれば、純粋な文書の分類となっている場合もある。この階層を決定する基準と、文書作成時に付与する方法も決めておく必要がある。