

**電子媒体による公文書等の
適切な移管・保存・利用に向けて
- 調査研究報告書 -**

平成18年3月

独立行政法人国立公文書館

序

独立行政法人国立公文書館(以下「国立公文書館」という。)は、平成17年4月1日から平成22年3月31日までの期間の中期目標において、「電子媒体の公文書等の効率的な管理・保存に向け、最適な保存媒体と管理方策等についての検討を行い、結論を得ること」とした。この中期目標を達成するための中期計画においては、「電子媒体の公文書等の効率的な管理・保存に向け最適な保存媒体と管理方策等についての検討を平成17年度より行い、平成18年度を目途に結論を得ること」とした。

そこで、平成17年度においては、「電子媒体の公文書等の効率的な管理・保存に向け最適な保存媒体と管理方策等について検討する」との年度計画を設定した。また、平成17年5月には、内閣府に「電子媒体による公文書等の管理・移管・保存のあり方に関する研究会」(以下「研究会」という。)が設置された。国立公文書館では、研究会における議論の動向も踏まえつつ、年度計画に基づいて、電子媒体による公文書等の管理・移管・保存等について調査研究を実施することとした。

電子媒体による公文書等の効率的な管理・保存を実現するためには、電子媒体の公文書等の移管を受ける国立公文書館における管理方策や保存媒体等について技術的な検討を行う必要がある。だが、それにとどまらず、情報技術の発達及び電子政府化の急速な進展や電子媒体による公文書等の増加の影響を見極め、これに対応した公文書等の作成、管理、移管及び利用のあり方全般を抜本的に検討する必要がある。

このような認識に立脚して、電子媒体による公文書等の効率的な管理・保存に向けた最適な保存媒体と管理方策等の検討に資するため、公文書等の移管後の国立公文書館における管理方策及び体制等だけでなく、電子媒体による公文書等のライフサイクルを作成段階から一般的に管理するための技術動向及び課題等について調査研究を行った。

具体的には、我が国における電子政府化の進展状況、電子媒体による公文書等の適切な管理・移管・保存・公開に係る技術動向と課題、ウェブページの移管・保存に係る技術動向等のほか、諸外国における取組み状況等について、公表されている文献、ウェブ情報及び関係機関等に対するヒアリング等により情報を収集し、調査研究を行った。その成果として、本報告書は作成された。

なお、本調査研究の実施にあたっては、国立公文書館に総務課長を長とするプロジェクト・チームを組織した。また、主に我が国及び諸外国における技術動向等に関しては、株式会社NTTデータ経営研究所に情報収集を委託した。

本調査研究の成果をもとに、国立公文書館は中期目標及び中期計画の実現を図ることとしている。

また、本報告書が、国及び地方の各機関のみならず、広く電子文書等の長期的な管理・保存・利用等に関心を有する方々にとって、議論や検討の有益な素材となれば幸いである。

目次

1.	はじめに.....	5
1-1	調査の背景.....	5
1-1-1	国立公文書館の取組み.....	5
1-1-2	政府による取組みの強化.....	5
1-1-3	電子政府基盤(インフラ)の充実.....	6
1-1-4	技術革新.....	6
1-2	電子公文書等の移管・保存・利用の意義.....	8
1-2-1	国立公文書館から見た電子公文書等の移管・保存・利用の意義.....	8
1-2-2	政府から見た電子公文書等の移管・保存・利用の意義.....	9
1-2-3	国民から見た電子公文書等の移管・保存・利用の意義.....	9
1-2-4	国立公文書館職員及び国の各機関職員から見た電子公文書等の移管・保存・利用の意義	9
1-3	今回の調査実施について.....	10
1-3-1	電子的に保存された文書の位置付け.....	10
1-3-2	調査目的.....	10
1-3-3	調査対象.....	10
1-3-4	活動内容.....	11
1-4	ライフサイクルと電子化.....	13
1-4-1	紙文書のライフサイクル.....	14
1-4-2	電子文書のライフサイクル.....	14
1-4-3	ライフサイクルと評価選別.....	15
1-5	メタデータ体系の整備.....	16
1-5-1	メタデータとは.....	16
1-5-2	国立公文書館におけるメタデータの必要性.....	16
2.	電子政府化の進展.....	17
2-1	府省庁の文書管理の現状.....	17
2-1-1	行政機関情報公開法と文書管理.....	17
2-1-2	文書管理規則.....	17
2-1-3	行政文書の安全な保存に関連する政府の取組み.....	24
2-1-4	総合文書ファイル管理システム(e-Gov).....	25
2-2	各府省における電子文書化の進捗.....	28
2-2-1	行政事務の電子化状況.....	28
2-2-2	行政文書の電子化率.....	29
2-2-3	海外の取組みも踏まえた今後のあり方.....	29
2-3	電子公文書基盤.....	30
2-3-1	霞が関WAN.....	30
2-3-2	イントラネット.....	32
2-3-3	各府省文書管理システム.....	35
2-3-4	電子決裁システム.....	36
2-3-5	電子申請.....	36
2-3-6	グループウェア.....	37
2-3-7	データベース.....	37
2-4	電子メール.....	38
2-4-1	位置づけ.....	38

2-4-2	電子メールの種類	38
2-4-3	各府省における電子メール取扱いの現状	39
3.	主要国と日本との比較	40
3-1	諸外国と日本との比較調査の背景	40
3-2	電子政府化への取組み	40
3-2-1	アメリカ	40
3-2-2	カナダ	40
3-2-3	オーストラリア	41
3-2-4	日本	41
3-3	電子文書管理・保存への取組み	42
3-3-1	公文書の定義及び電子公文書の位置づけ	42
3-3-2	行政機関の公文書作成・保存への国立公文書館の関与	43
3-3-3	メタデータの標準化へ向けた取組み	44
3-3-4	電子公文書等の長期保存へ向けた取組み	45
3-3-5	電子メール管理・保存に向けた取組み	45
3-3-6	政府機関のウェブ・アーカイピングの現状	46
4.	電子公文書等の適切な管理・保存のための技術動向と課題	47
4-1	電子公文書等の長期保存にまつわるリスク等	47
4-1-1	電子公文書等を保存し続けることによるリスク	47
4-1-2	媒体・ハードウェアに起因する可読性損失リスク	47
4-1-3	ソフトウェアに起因する可読性損失リスク	47
4-1-4	改ざんの危険性	48
4-1-5	既存技術及びシステムの陳腐化	48
4-2	技術動向及び課題に関する調査項目	49
4-2-1	調査項目の位置づけ	49
4-3	メタデータ	50
4-3-1	メタデータの分類	50
4-3-2	記録管理(recordkeeping)メタデータ	50
4-3-3	アーカイバル・メタデータ	51
4-3-4	技術的メタデータ	53
4-3-5	ダブリン・コア	53
4-3-6	メタデータ体系の比較	54
4-3-7	諸外国のメタデータ事例	55
4-3-8	メタデータに関する課題	58
4-4	移管前の技術	59
4-4-1	ファイル形式の標準化	59
4-4-2	テキストファイル	61
4-4-3	画像ファイル	63
4-4-4	音声ファイル	65
4-4-5	映像ファイル	66
4-5	移管時の技術	68
4-5-1	ファイル形式変換技術	68
4-5-2	海外事例(オーストラリア:XML変換ソフトXena)	68
4-5-3	メタデータの抽出	72
4-5-4	移管時のセキュリティ	72

4-6	移管後の技術(保管)	74
4-6-1	電子公文書等の記録媒体	74
4-6-2	バックアップ	76
4-6-3	保管時のセキュリティ	77
4-6-4	原本性保証、改ざん防止	79
4-7	移管後の技術(利用)	82
4-7-1	マルチビューワ(複数ファイル形式対応ビューワ)	82
4-7-2	エミュレーション	82
4-7-3	全文検索機能	84
4-7-4	利用時のセキュリティ	85
4-8	その他の技術動向	86
4-8-1	文字コードの標準化	86
5.	ウェブページの移管・保存技術の動向	89
5-1	ウェブアーカイブの動向	89
5-1-1	背景	89
5-1-2	政府の認識	89
5-1-3	事例	89
5-1-4	我が国におけるウェブ情報アーカイブ動向	90
5-1-5	諸外国の参考事例	93
5-2	ウェブ上の公文書等の種類	97
5-2-1	国民向けのホームページ	97
5-2-2	政府の内部向けページ(霞が関WAN上)	97
5-2-3	政府の内部向けページ(府省庁内イントラネット上)	98
5-3	ウェブページ保存における諸課題	99
5-3-1	ウェブの動的要素に起因する課題	99
5-3-2	安定した識別子の必要性	100
5-3-3	収集上の課題	101
5-3-4	管理上の課題	103
5-3-5	保存上の課題	104
6.	電子公文書等の適切な移管・保存・利用に向けて	105
6-1	適切な移管の実現に向けて メタデータの標準化	105
6-1-1	記録管理メタデータ付与への対応	105
6-1-2	アーカイバル・メタデータ付与への対応	105
6-1-3	技術情報メタデータ付与への対応	105
6-1-4	メタデータ付与の際の課題	106
6-2	適切な移管の実現に向けて 移管対象の評価選別	108
6-2-1	評価選別方法に関する考え方	108
6-2-2	ウェブページの扱い	109
6-2-3	電子メールの扱い	110
6-3	適切な移管の実現に向けて ファイル形式の標準化	111
6-3-1	ファイル形式に関する考え方	111
6-3-2	テキストファイルの標準化	111
6-3-3	画像、音声、映像ファイルの標準化	112
6-4	適切な移管の実現に向けて 移管方法	113
6-4-1	移送と移管	113

6-4-2	移管(移送)の方法.....	113
6-5	適切な保存の実現に向けて.....	115
6-5-1	長期保存上の留意点.....	115
6-5-2	エッセンスの保存.....	115
6-5-3	安全性確保の手法.....	116
6-5-4	原本性確保技術.....	116
6-5-5	オープン技術.....	116
6-6	適切な利用の実現に向けて.....	117
6-6-1	電子公文書等の利用・公開における考え方.....	117
6-6-2	利用システムの概要.....	117

1. はじめに

1-1 調査の背景

1-1-1 国立公文書館の取組み

国立公文書館は、平成13年4月1日から平成17年3月31日までの第1期中期計画において、国民に対して提供するサービスその他の業務の質の向上等为目标として掲げ、歴史公文書等の受入れ、保存、利用について、歴史公文書等の重要性の評価や、劣化要因に応じた保存対策の実施、展示会の開催やホームページの充実等を実施してきた。

平成17年4月1日から平成22年3月31日までの第2期中期計画においても、基本的な方向性は変わらず、国民に対して提供するサービスその他の業務の質の向上等为目标として、多様な取組みを行っているところである。その一つとして、電子媒体の公文書等(以下、「電子公文書等」という。)の効率的な管理・保存に向け最適な保存媒体と管理方策等について検討を行うこととしている。本調査研究は、この検討に資するために実施するものである。

1-1-2 政府による取組みの強化

内閣官房長官のもとで開催された「公文書等の適切な管理、保存及び利用に関する懇談会」(以下、「懇談会」という。)が平成16年6月に取りまとめた報告書「公文書等の適切な管理、保存及び利用のための体制整備について - 未来に残す歴史的文書・アーカイブズの充実に向けて -」(以下、「懇談会報告書」という。)において、米国をはじめとする諸外国とは国立公文書館(及びそれに相当する施設)が持つ機能の多様性や現用文書の集中管理の仕組み、情報公開法や文書管理法といった法制度の整備の状況等が異なるとはいえ、歴史的に重要な公文書等を適切に保存し、国民に利用に供することを保障する体制の整備が今後の課題とされた。

一方、政府内でも電子政府化への取組が急速に進んでいることから、政府において作成される公文書等について、紙媒体によるものから電子媒体によるものへの移行が今後急速に進むことが予測されている。紙媒体からの電子媒体への移行は、文書の管理のあり方等について、抜本的な見直しを求めるパラダイム・シフトとなる可能性があるが、現行の電子政府構築計画には、歴史的に重要な価値をもつ電子媒体による公文書等を将来の国民に遺すという視点が十分には反映されていないのではないかと指摘もある。

以上の実態を踏まえ、懇談会においては、現在の国立公文書館が行っている公文書の移管に関する制度上のアドバイスや移管後の公文書等の管理だけでなく、公文書等の作成、管理、移管、廃棄等のライフサイクルの全体像を視野に入れた総合的な取組みが求められるとされたところである。

懇談会報告書に基づいて、平成17年5月から内閣府に2つの研究会が開催され、公文書等の中間段階における集中管理の仕組み及び電子公文書等の管理・移管・保存のあり方について、さらに議論を深めることとなった。このうち、「電子媒体による公文書等の管理・移管・保存の

あり方に関する研究会」(以下、「研究会」という。)においては、現在の移管等の仕組みが主に紙媒体による公文書等の管理・移管・保存等を前提としていることから、現在の制度にとらわれずに、電子公文書等の適切な管理・移管・保存のあるべき姿について検討が重ねられた。

1-1-3 電子政府基盤（インフラ）の充実

ICA(国際公文書館会議:International Council on Archives)報告書 16「電子記録:アーキビストのためのワークブック」(2005年)(以下、「ICA 報告書 16」という。)において、電子形態で作成され現用記録の量が世界的に増加しているとの指摘がなされているが、我が国においても電子政府基盤の整備によって、国及び地方の各機関において電子的に作成される文書・情報の量が増加している。

(1) e-Japan 戦略

平成 13 年(2001 年)1月に IT 戦略本部が決定した「e-Japan 戦略」は、IT 戦略を国家的戦略の一つとして位置づけたもので、現在にいたる官民の IT 化推進の起源となっている。

e-Japan 戦略はいくつかの基本戦略で構成されている。電子政府については、「2003 年までに、行政(国・地方公共団体)内部の電子化、官民接点のオンライン化、行政情報のインターネット公開・利用促進、地方公共団体の取組み支援等を推進し、電子情報を紙情報と同等に扱う行政を実現し、幅広い国民・事業者の IT 化を促す」とこととされ、政府が関わる申請・届出手続きの多くが電子的に行うことが認められるとともに、電子申請システムをはじめとする各種電子政府基盤インフラが各府省において急速に整備された。

(2) 課題

一方、電子政府基盤インフラの活用については、各府省内という組織の枠内での取組みが中心となっており、情報システムの整備は、各府省が個別に進めている状況にある。

すなわち、文書管理システムをはじめとする関連 IT インフラについては、各府省庁が充実を図ってきたという経緯(注:政府にて最適化計画の一つとして文書管理システムについても全府省横断的に見直す動きがスタートしたところ)から、多くの行政文書が「電子政府基盤インフラ」上で作成されているといえるにもかかわらず、組織間の横の連携が不十分であること、また、電子政府基盤が必ずしも電子公文書等の長期保存上の要件を十分に満たしているとは言えない点が課題となっている。

そういった状況を踏まえ、現在の IT 戦略本部の活動の中心も、これまで整備してきた IT インフラが十分に機能しているかという検証作業、国民も幅広く巻き込んだ利活用の充実施策や障壁の除去といった段階へと移行している。

1-1-4 技術革新

日本では、安価かつ広帯域のブロードバンド基盤の整備により、オンライン上で大容量の電子データを配信・受信することが幅広く可能になった。ICA 報告書 8「アーカイブズの観点から見る電子記録管理ガイド」(1997 年)(以下、「ICA 報告書 8」という。)では、電気通信手段によ

る送付への期待として、「アーカイブズ等の記録提供組織がインターネットかダイヤルアップ接続によるデジタル通信を利用できれば、この方式はデジタル記録媒体へのコピーによる提供方式とほとんど変わらないものになるう」としたが、その後インターネットの主たる利用環境は、「ダイヤルアップ接続」からブロードバンドでの常時接続へと移行した。

技術革新の結果、従来パソコンで作成したにもかかわらず紙で保存されてきた文書についても、保存まで全て電子で完結させたほうが望ましいのではないかという考え方も生じてくる。例えば、研究会(第2回)においても、委員より「データとしては最初の段階で電子である以上、電子のまま保存して残すことで業務の効率化が図られるのでは」との指摘を受けたところである。

一方、公文書の移管・保存という視点からは、国の行政機関等が事務事業の遂行過程で作成するものであるため、文書上に重要度・機密度の高い情報が記録されていることが多いと想定される。したがって、移管や保存については、単に低コストで効率的に行えるというだけでなく、長期間にわたり文書の真正性、信頼性、完全性を確保するために適切な対応をとらなければならないということに留意する必要がある。

1-2 電子公文書等の移管・保存・利用の意義

電子公文書等の移管・保存・利用を行うにあたっては、それ自体を目的とするのではなく、その意義についての的確な理解に基づいて、行う必要がある。

ここでは、電子公文書等の作成を行なう政府、電子公文書等の移管・保存の責務を担う国立公文書館及び保存された電子公文書等を検索・利用する一般の国民の 3 つの視点から電子公文書等の移管、保存及び利用の意義について考える。

1-2-1 国立公文書館から見た電子公文書等の移管・保存・利用の意義

ICA 報告書 8 において、アーカイブ機能の定義は、「長期保存すべき記録を特定し、保護・保存するという目標を達成する為に寄与し、必要とされ、また、そのような記録を確実にアクセス可能で理解可能なものにする一群の関連活動」とされている。

従来、国立公文書館は、「非現用文書」のアーカイブ機能を主に担ってきた。だが、アーカイブの対象が電子にまで広がることによって、非現用化後のプロセスのみに関与するだけでは不十分であり、現用文書時点から政府及び政府職員を巻き込んだかたちで取組みを進めて行く必要が生じてくる。ICA 報告書 8 では、アーキビストが電子文書の作成以前のプロセスに関与しなければ電子化の効果は十分に発揮できないとしている。また、研究会においても、電子的な移管及び保存のための現用段階からの一定の管理の必要性について委員より指摘があったところである。

また、電子公文書等の移管の実現によって国立公文書館に移管される公文書等の質量両面での充実が期待される。というのも、電子公文書等は、紙媒体によるものと比べ、同時かつ大量の処理が容易であることから、電子公文書等の管理・移管・保存等に係るシステムやメタデータ体系等が適切に構築・運用されれば、行政機関にとって国立公文書館への同時かつ大量の移管が容易になるためである。さらに、電子公文書等は、紙媒体と比べて、音声、画像、映像、ウェブ等多様な記録様式が存在するので、紙媒体による公文書等のみを移管・保存してきた従来の枠組みでは想定されなかったような多様な内容の電子公文書等が移管・保存されることとなる可能性がある。

一方、国立公文書館としては、内閣総理大臣を通じて、国の機関から歴史資料として重要な公文書等の移管を受けるだけでなく、移管を受けた公文書等を保存し、一般の利用に供するという役割も果たしている。所蔵する公文書等の利用に関しては、平成17年4月からは、国立公文書館が提供する「デジタルアーカイブ・システム」により、目録データベースの検索や資料のデジタル画像情報の利用をインターネット上で 24 時間可能になっている。現在の「デジタルアーカイブ・システム」では所蔵する紙媒体による公文書等をマイクロフィルム等で撮影した上でデジタル画像化しているが、今後電子媒体により作成された公文書等が電子媒体のまま移管されることによって、電子媒体のまま利用に供するというサービスが可能となるであろう。

1-2-2 政府から見た電子公文書等の移管・保存・利用の意義

電子政府化の進展により、政府の内部事務に情報システムが浸透し、日々事務の電子化が進んでいる。また、電子化された政府の情報は、ウェブを通じ、国民に対して 24 時間提供可能となった。今後は、全ての行政文書中に占める電子媒体による行政文書の比率は次第に高くなり、行政事務の効率化が図られていくものと想定される。

その場合、行政文書の移管・保存についても、電子媒体によるものを紙等非電子媒体に変換してから移管するのは効率的ではない。電子媒体によるものは電子のまま移管できる仕組みを整えることによって、政府にとっても業務の効率化の面で大きなメリットが期待される。

1-2-3 国民から見た電子公文書等の移管・保存・利用の意義

公文書等が紙媒体であった場合、インターネット上で公文書等の所在情報や画像の検索を行うことができたとしても、あくまでも原本そのものではないことから、公文書等の原本を閲覧するためには、来館する必要がある。

これが、電子媒体となることによって、国民は、インターネットを通じて、24 時間いつでもどこでも公文書等の情報を検索するだけでなく、閲覧等も行うことが可能となり、現在よりも身近に公文書等に触れられるようになる。また、電子公文書等は、紙媒体によるものと比較して、データベースや音声・映像といった多様な記録様式が想定されるので、それらを利用に供した場合、データベースの分析や音声・映像情報の解析など、利用の目的や用途が広がっていくことも想定される。

1-2-4 国立公文書館職員及び国の各機関職員から見た電子公文書等の移管・保存・利用の意義

国立公文書館職員から見ると、公文書等が電子媒体により移管・保存・利用されることにより、従来のモノとしての紙媒体による資料の物理的な管理と比較して、全般的な業務の効率化が期待される。

一方、国の各機関の職員から見ると、電子媒体により公文書等を移管することができれば、仮に行政文書としての保存期間満了前に各機関の庁舎等から一定の集中管理の下に移したとしても、コピーを簡単に作れるという電子文書の特質から、国の各機関職員による当該文書利用への障害は少なくなる。

1-3 今回の調査実施について

1-3-1 電子的に保存された文書の位置付け

「文書」の定義については、事務の運営に必要な一切の図画及び書類(電磁的記録)を含むとされていることから、電子媒体によるものも文書の中に含まれている。また、「行政文書」についても、行政機関の保有する情報の公開に関する法律(以下「行政機関情報公開法」)第2条第2項にて、「電磁的記録を含む」という規定があることから、電子媒体も行政文書の中に含まれているとされている。

他方、ICA 報告書 8 の中では、「機関や個人の活動の開始時、実施時、完了時に作成または受領され、その活動に証拠を与えるに足る内容、コンテキスト(文脈、背景)、構造から成る、記録された情報」として記録の定義としている。

1-3-2 調査目的

今回の調査は、多様な電子公文書等を国立公文書館にいかにより電子的に移管し、保存するかという観点から、その実現方法・スキーム及び関連情報等について、主に技術面から整理するものである。また、電子的に移管する際の技術上、運用上の課題にはどのようなものがあるか、どのような技術やソリューションによって解決可能か、当該技術をどのように適用することが望ましいかについても整理する。

なお、単に技術的にあるべき姿のみを論じるのみではなく、現実的な導入可能性についても考慮し、国立公文書館や国の各機関等多様な主体の関与が想定される中で、組織的な課題とその対処策についても整理した。

1-3-3 調査対象

純粋な技術面からの全体的動向について最新情報を把握するとともに、電子文書や電子情報等の保存に関わる内外関係機関(国立国会図書館を含む。)の動き、電子公文書等のライフサイクルの観点から、行政文書の作成・管理を行う各府省における電子公文書等の作成や保存の現状についても整理した。

(1) テキスト文書、映像、音声等

移管対象となる公文書等としては、「記録の媒体については、文書、地図、図画、フィルム、音声記録、磁気テープ、レーザーディスク等そのいかんを問わない」という制度的な位置づけ(「公文書館法の解釈の要旨」(1988、総理府))にもかかわらず、今までのところ、移管されているのは、紙媒体のものがほとんどである。

だが、今後は、電子媒体による移管・保存等の取組みを行うことにより、多様な形態の公文書等が移管されることが考えられる。例として、各種審議会での審議の様子は、従来は紙媒体による議事録等で残されるのみであったが、電子媒体による移管・保存等を行うことにより、音声ファイルや映像ファイルとして電子的に幅広く提供することが可能となるであろう。

このような観点から、映像、音声等を調査対象とした（現に、衆議院や参議院の審議については、既にインターネット上で審議映像が公開されており、審議終了後も過去の審議映像を見ることができる。）。

(2) ウェブ情報

電子政府基盤の普及により、霞が関 WAN や各府省内イントラネット上のインターフェースとして重要な位置を占めるウェブ経由で提供される情報、あるいは、各府省のホームページから政府外に対してインターネットで発信されるホームページ上の情報についても、重要な公文書等と位置づけることができるのではないかとこの観点から調査対象とした。

1-3-4 活動内容

今回の調査報告を取りまとめる過程で、以下の活動を行った。それぞれの活動内容については、報告書の中で動向や事例として必要に応じて盛り込むこととした。

(1) 海外事例

調査対象としたのは、アメリカ、カナダ、オーストラリアの3カ国である。これらの国々では、単に電子公文書等の移管・保存・利用に関する取組みが、ファイル・フォーマットやメタデータ等の標準化など、技術的な側面において進んでいるだけでなく、国立公文書館又はそれに相当する組織が国家の行政組織全体に対するガバナンス機能を果たすなど制度的な位置づけなどにおいても見るべき点が多いことが分かった。

今回の調査報告書においては、我が国の行政機関が行っている電子公文書作成を、国立公文書館への移管という面から側面支援する際の参考として、各国の取り組み内容を先進事例として調査した。また、オーストラリアについては、研究会委員等による視察が2005年10月末に行なわれ、その結果が第5回研究会資料として取りまとめられていることから、これを参考とした。

(2) 関係府省庁へのインタビュー

電子公文書等の作成主体となる各府省がどのようなシステム（ハード、ソフト、ネットワーク等）を用いているのかについての検討は、電子公文書等の移管・保存・利用等を考える上で不可欠である。だが、関連する情報は従来公表しているレベルでは必ずしも十分とはいえないことから、関係府省にインタビューを行うこととした。

(3) XML 変換ソフトウェアの試用と評価

各府省が作成している電子公文書等は様々なアプリケーション、文書管理システム等の環境で構築されている。だが、公文書等の長期的な保存、再現（利用）環境の確保や国民への利用提供といった視点から考えた場合、電子的に移管・保存等する際に公文書ができる限り標準化されていることが望ましい。

その際の方法論の一つとして、各府省が作成、保存した様々なファイル形式の公文書の本文及びファイル属性等のデータをXML化することによって、電子公文書の移管、保存、再現（利用）を容易にすることが考えられる。

オーストラリア国立公文書館（NAA）においては、オフィスドキュメント、html や PDF ファイルをXML化するアプリケーション「Xena」が開発・更新されている。本報告書においては、

Xena の試用結果を報告するとともに、日本の電子公文書等の移管・保存等への活用の可能性について得られた示唆について報告する。

(4) その他

関連する技術動向の内容が多岐にわたることから、調査対象分野については必要に応じ当該分野におけるインタビューを行うことで、最新の動向や最新技術の理想と現実のギャップといった、電子公文書の移管、保存の実現の際に有効性があると思われるノウハウをできる限り抽出した。

1-4 ライフサイクルと電子化

従来より、アーカイブの活動が始まるのは(ライフサイクルの起点)、一般的には長期保存すべき記録の作成段階であるとされ、ライフサイクルはその後、後期段階である保存・利用まで続くとされてきた。ところが、電子環境においては、ライフサイクルの起点は記録の作成よりもさらに前の段階からスタートするとの指摘がなされている(ICA 報告書 8)。

また、保存については、3つのフェーズに分類が可能としている。第1段階では意思決定や業務処理を支える手段として作成者に利用されるための保存があり、次に、その他の活動あるいはアカウントビリティ要件(情報公開請求等)を満たすための保存がある。最後に、法的・文化的目的や研究上の目的のために非現用文書として保存されることとなる。

一方、電子文書のライフサイクルについては、二つの大きな問題点があるとされる。まず、媒体の安定性によるもので、電子文書を記録する媒体は、寿命が短いもので数年、長いもので数十年といわれている。そのため、マイグレーション等の適切な措置を施さない限り、その媒体に記録されている情報が失われてしまうおそれがある。次に、媒体と情報の分離が可能であるため、不正な上書きといった、真正性や信頼性面でのリスクが大きいとされている。

諸外国及び我が国の公文書等のライフサイクル管理の現状としては、研究会(第2回)で事務局より指摘されているとおり、作成から選別を受けて、最終的に公文書館に移管されるまでのライフサイクル全体について、調査対象国(アメリカ、カナダ、オーストラリア)では公文書館が法的な権限を持ち深く関与している一方、日本については、国立公文書館は現用段階の公文書等を管理するための法的権限もなく、実際上も関与はしていない。

各国公文書館

- アメリカ: アメリカ国立公文書記録管理局 (NARA: National Archives and Records Administration)
- カナダ: カナダ国立図書館公文書館(LAC: Library and Archives Canada)
- オーストラリア: オーストラリア国立公文書館 (NAA: National Archives of Australia)

1-4-1 紙文書のライフサイクル

ICA 報告書 8 の中で、紙文書と電子文書のライフサイクル面での差異が整理されている。概して、紙文書のライフサイクルは各プロセスが明確に分離・独立した存在であり、電子文書のライフサイクルと比べ単純である(以下は同報告書を再整理したもの)。

(1) 記録媒体と記録内容の一体性

紙の記録の記録内容は記録媒体(紙などの記録媒体)に記録され、記録内容を媒体から切り離すことはできない。

(2) 作成と証拠性の一体化

紙の環境では、内容、コンテキスト、構造は必然的に記録の保存媒体(通常は紙)と一体化している。したがって、記録作成が決定されると、証拠作成という目標が達成される。

(3) 保存要件

紙の場合の保存は、損傷を避けるために物理的単位(用紙の枚数、冊数等)をできるだけ良好な条件で保存し、損傷が発生した場合、または発生した時には損傷を修復することを意味しており、「安定性」を前提とした適宜の対応である。

(4) 機能の分散化

紙をベースとする従来環境では、アーカイブ機能が分散化されており、業務遂行上の責任は、記録作成者、記録登録者、記録管理者およびアーキビストを含む(が限定はされない)多くの者に割り当てられ、互いに独立している。

1-4-2 電子文書のライフサイクル

電子文書のライフサイクルは、文書作成・管理システムの設計段階での長期保存要件検討の必要性をはじめ、全体として、紙の場合と比べ課題への対応が前倒しになっている。

研究会(第4回)でも、移管時期について、「(保存)期間満了前に移管できれば、長期保存上の措置というものは統一的に講じることができるため、早い時期の移管が望ましいのでは」といった議論が行われたのも、こういった電子文書のライフサイクルの特性が背景にある。

(1) 構想段階

記録作成の条件が設計段階で電子システムに組み込まれていなければ記録の作成はできないため、電子環境では記録のライフサイクルの開始時点を記録作成前の段階(構想段階)に遡って延長しなければならない(ICA 報告書 8)。

まず、システムの設計と仕様策定の段階から、記録の内容、コンテキスト、構造に信頼性を持たせ、長期保存の対象となる電子記録が特定・保存することができるための機能要件を盛り込んでおくべきであり、新たな主体(情報管理者、システム設計者)がシステム設計のプロセスに参加する必要が生じる。

このため、アーキビストが、真正性と信頼性と保存性確保のため、このフェーズから積極的な役割を果たすべきだとしており、ICA 報告書 8 では、システムの設計者、開発者および記録作成者に情報・指導・指示を与える権限をアーキビストに与えてはどうかとしている。

(2) 作成段階

ICA 報告書 8 では、作成段階独自の課題について明確な定義は行っていないが、電子文書の作成についてはあくまで構想段階とセットでとらえられなければならないとしている。すなわち、あくまで記録作成の条件が設計段階で電子システムに組み込まれていなければ、適切な記録の作成はできないということである。

(3) 保存・維持段階

電子的に保存する際に、アーキビストが策定する基準に基づいて保存を行なうとともに、アーカイブ機関がその適切な実施を計画的に監視する必要性が ICA 報告書 8 で指摘されている。したがって、記録の保存、記録へのアクセス、記録の利用を担当するアーカイブ機関以外の者の業務を監視する手順を確立することが今後の課題である。

また、記録媒体だけを保存しても十分ではなく、電子環境では記録媒体から独立している記録の内容、コンテキスト、構造が維持されなければならないのではないかという論点がある(対応策の一つとして、エッセンスのみを保存することが考えられる(第6章参照))。

今後の解決策の導入が成功するか否かは、公的文書の保存の責務を負っている政府と政府機関が記録の保存要件を必要かつ優先的な事項と見なし、マイグレーション等に要する費用を負担するかどうかによって左右される面が大きいとしている。

1-4-3 ライフサイクルと評価選別

移管対象の公文書の評価選別を行う時期もライフサイクルと密接に関連してくる。研究会の中の議論でも、評価・選別はライフサイクルの初期に行うことが望ましいとの方向性が示されていた。

これは、どの記録管理システムに長期保存すべき記録が含まれ、システム中のどの記録に長期保存価値があるかといった評価選別機能は、電子文書の作成・管理システムの設計段階で既にシステムに組み込まれなければ、評価選別が人による作業に依存することとなり、評価選別が質量ともに不十分となり、また高コストになるためである。また、評価選別が保存・維持段階で行なわれるとなれば、適切な記録が作成されない、選別結果の妥当性を検証することができないなどの課題を抱えることになる。

1-5 メタデータ体系の整備

1-5-1 メタデータとは

メタデータはいわば「データについてのデータ」ともいえるもので、電子記録に関する技術データ、記録の構成と内部構造、記録の追加・削除・変更に適用される規則、またはコンテンツの解釈について記述するデータ等を指すものである。ISO15489-1によれば、「コンテキスト(背景・状況・環境)、内容、構造及びある期間の記録の管理について説明したデータ」と定義されている。

メタデータがアーカイビングに必要である理由は、記録を理解・利用可能な記録にする際に記録のコンテキストと構造に関するデータが不可欠なためである。適切なメタデータ体系を整備し、適用させることによって、検索・ユーザビリティ・真正性・信頼性・維持管理・保存・評価選別などの様々な目的に役立てることができる。

メタデータは時間の経過に応じて変更されることもあるが、必要に応じて、記録とそのメタデータは変更されないよう保護することができる。

1-5-2 国立公文書館におけるメタデータの必要性

特に電子公文書等のように膨大な数の電子ファイルの管理が予想される場合、保存後の管理・閲覧を考慮して文書構造を検討する必要がある。もし、移管された電子公文書等にメタデータが付与されていないならば、ファイル名だけで必要な文書を探さなければならない。ファイル名のみから文書の内容を詳細に推測することは難しく、たとえ検索機能を利用したとしても、必要な文書を見つけるまでには、多大な時間を要してしまう。

しかし、移管された電子公文書に作成日や作成者、分類情報などの情報がメタデータとして付与されることで、文書ファイルを1つずつ開くことなく、作成者や作成日といった情報を把握することが可能となり、電子公文書の管理や利用を効率化することができる。

2. 電子政府化の進展

2-1 府省庁の文書管理の現状

ICA 報告書 16 において、「政府は電子記録管理の準備をあらゆる面で整えているということはない。つまり、真正性、信頼性、完全性、有用性を完備した記録の作成体制を整えてはいない。」と、政府の既存の電子記録管理に関するシステムが独自仕様・独自ファイル形式を持つゆえの対応の不備を指摘している。我が国について上記見解が当てはまるか否かについては、電子政府インフラのあり方及び文書管理の現状を把握することで判明する。

2-1-1 行政機関情報公開法と文書管理

行政機関情報公開法と各府省の文書管理は密接な関係にあるが、電子公文書等の移管、保存と関係が深いのは、文書管理規則等と(メタデータ基盤となる)行政文書ファイル管理簿である。

(1) 文書管理規則

行政機関情報公開法第 22 条においては、行政機関の長が行政文書を適正に管理するために、行政文書の管理に関する定め(文書管理規則等)を設け、行政文書の分類、作成、保存及び廃棄に関する基準その他の行政文書の管理に関する必要な事項について定める必要があると規定されている。

行政機関情報公開法の施行を踏まえて、各府省では文書管理規則等を定めている。

(2) 行政文書ファイル管理簿

行政機関情報公開法では、国民からの行政文書の開示請求に行政機関が適切に対応できるよう、行政文書の管理について定めており、同法施行令において、「行政文書ファイル及び行政文書(単独で管理することが適当なものであって、保存期間が 1 年以上のものに限る。)の管理を適切に行うため、これらの名称その他の必要な事項(不開示情報に該当するものを除く。)を記載した帳簿を磁気ディスクをもって調製する」とこととされている(行政文書ファイル管理簿の作成)。

2-1-2 文書管理規則

(1) 概要

行政機関情報公開法施行令第 16 条においては、「行政文書の管理に関する定め」が満たすべき要件が次のとおり規定されている。

- 行政文書分類の基準
- 文書の作成
- 専用の場所に保存
- 文書保存期間の基準
- 文書保存期間満了日までの保存

- 文書保存期間の延長
- 文書の移管、廃棄
- 文書保存期間前の特別の廃棄
- 「行政文書ファイル管理簿」の調製
- 文書管理者

各府省が定めている文書管理規則等は、上記施行令第 16 条の規定のほか、「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)を受けて、共通の枠組みによって作成されていることから、規定の内容も類似している点が多いのが特徴となっている。

出典:「情報公開法における文書管理」(平成 17 年 5 月 27 日 総務省行政管理局)

(2) 目的

文書管理規則等を定める目的として、文書管理事務の適正化、効率化、迅速な処理等を冒頭に掲げているものが多い。ただし、府省によって、その規定内容は異なっている。

- 「文書の管理に関する統一基準その他基本事項を定め、事務の適正かつ効率的な遂行に資する」(内閣府)
- 「行政の事務を適正にし、その効率的な運営を図る」(財務省)
- 「文書の適正かつ迅速な処理を図る」(文部科学省)
- 「事務処理の適正及び効率化を図る」(環境省、厚生労働省)
- 「事務処理の適正化及び能率の向上」(国土交通省)
- 「行政文書を適正に管理」(法務省)

(3) 対象文書

文書管理規則等は、各府省の行政文書の管理に関して定めたものである。そういった点で制度上は幅広い文書が対象として含まれているが、研究会(第 4 回)での事務局説明によると、文書管理規則等は主に決裁文書を第一ターゲットとして運用のあり方を規定したもので、それ以外の例えば会議資料のような決裁を経ない文書については、運用レベルに任されているというのが現状とのことである。

(4) 文書管理規則等の特色

今回の調査では、全府省庁の文書管理規則等について規程の内容や特色を主に行政文書の作成、移管といった面から調査した。基本として、各府省庁の行政文書管理規則等はいずれも共通のガイドラインに沿って策定されているものの、規定の内容について異なる点も何点か見受けられた。

ア. 文書の作成義務

意思決定に当たっては、文書(図画及び電磁的記録を含む)を作成することとされているが、例外事項として、意思決定と同時に文書を作成することが困難な場合と、処理に係る事案が軽微な場合とされている。

前者(意思決定と同時に作成が困難)については、研究会にて政府側出席者からの「会議等での意思決定については同時での文書作成がむずかしい」との発言があった

ことから、会議での決定内容が含まれているようである。後者(軽微な事案)については、具体的に何を指しているのか具体的定義は難しい。

イ. 文書の作成方法

一部省庁(例として、環境省)においては、行政文書を作成するときの文字の使用方法等について以下のような規定が別途設けられていた。

- 常用漢字表、現代仮名遣いに従うこと
- 分かりやすい用字用語で適切、簡潔に記載すること
- 外国語による文書には、訳文又は訳文の要領をつけること

また、総務省では、文書の作成方法の標準化を進めるため、一定の書式もしくは文案で統一化が可能なものについては、省内で協議の上例文として登録を受けることができるとしている。

ウ. 正本及び写し

正本と写しに関する規定を設けているところがある。例えば厚生労働省では「文書は原本と写しとの区別を明確にしなければならない」としている。

写しの作成理由には業務上様々なものがあると思われるが、電子的な安全性(バックアップ)を目的とする場合もある。文部科学省では「行政文書の保存に当たっては、その必要に応じ、当該行政文書に代えて、内容を同じくする同一又は他の媒体の行政文書を作成することができる」としている。

また、保存期間についても文書管理規則上の規定はあくまで原本に限られたものであるとし、写しについてはそれ以前での廃棄を認めている(例:「行政文書の写しは、前2条の規定に関わらず、適宜整理することができる」(厚生労働省))。

エ. 行政文書ファイル

行政文書ファイルについては、行政文書(保存期間1年以上のものに限る)を一定のグループとしてまとめて管理するものであるが、単位については文書管理規則上明確に規定されているわけではない。

ただし、事務的な効率性等の面から、「一つの行政文書ファイルに属する行政文書の数が過度に多くならないようにする」(国土交通省)、「調整の上、分割し、又は統合することができる」(国土交通省)などとしている例が見られる。

また、行政文書ファイルの分類方法を定義した「分類基準表」を府省庁で、大分類・中分類・小分類の別で設けている。ただし、各分類の定義については、各府省庁が独自に定めている。ファイル名については「標準行政文書ファイル名」が定義されているところもあるが、この「標準文書ファイル名」がどの程度遵守されているかについての実態は不明である。

なお、紙媒体による行政文書ファイルの区分方法については、第3回研究会で政府出席者からの説明によると、紙ベースの文書を綴じるファイルの簿冊単位でファイル名がつけられているとともに、幾つかの大分類、中分類、小分類という分類のルールに従って名称がつけられているとのことであった。

オ. 移管対象の手続き及び保存期間の延長

各府省庁とも、国立公文書館への移管は、当該文書の主管課(起案文書の起案課)が大臣官房等の取りまとめ責任者(総括文書管理者)に移管を依頼することとなっている。

行政文書の(当初設定した)保存期間満了後の保存期間の延長については、国立公文書館への移管時期とも密接に関わってくる。各府省庁の行政文書管理規則においては、次のように、明示的な延長理由(各府省庁とも、ほぼ同じ内容)が規定されている。

- 現に監督、検査の対象
- 現に係属している訴訟関連
- 現に係属している不服申し立て関連
- 開示請求関連

以上のほか、「職務の遂行上必要があると認めるとき」、「主管課長がなお保存を要すると認める場合」といったケースで「文書管理者(課長級)」が保存期間を延長できるとしている。後者の規定については、規定が抽象的であり適用の範囲が必ずしも明らかであるとはいえず、今後の移管のあり方の検討の過程で議論を呼ぶ可能性も否定できない。

カ. 体制

行政文書管理の実務責任を負う文書管理者(課長級)及び文書管理担当者は、課等の部署単位で設けられており、文書分類基準や行政文書ファイル管理簿を部署単位で作成する権限を有している。もっとも、全省レベルで文書管理者を指導する役職(主任文書管理者、総括文書管理者)も存在するが、兼職であることから実質上の権限は課レベルに委ねられていると思われる。

(5) 文書管理規則等に見る電子文書の特色

全府省庁の文書管理規則について、電子文書の管理面から調査した。紙と電子という媒体の性質の違いから、管理方法による違いが見受けられた。

ア. 原本の修正について

今回各文書管理規則を調べたところ、「電子的方式による起案文書の修正は、起案文書の特定の箇所に修正すべき内容を入力し、記録することにより行うものとする」(国土交通省)、「決裁文書の修正は、加除訂正によって行うものとし」(厚生労働省)とされており、制度上は最終的な決裁文書の保存を行うことによって、そこにいたるまでの修正の過程も同時に(電子的に)保存することも可能である。

ただし、実態として、紙媒体の場合には、該当部分の差し替えで対応し得ること、また電子決裁システムでも、上書き保存(変更前の情報が見えない)機能やファイルの差し替え機能が提供されている可能性もあり、最終決裁の文書の電子保存のみをもって、全ての履歴を確保できるかどうかについては不明である。

なお、厚生労働省の文書管理規則では、例外措置として、「修正により決裁文書の内容が判読し難くなったときは、当該修正があった後の文書を作成し、当該決裁文書に添えて決裁を受けることができる」としている。

イ. 紙媒体と電子媒体による文書が併存している場合の原本の扱いについて

研究会において、紙媒体と電子媒体による文書が並存している場合に、どちらが原本となるのか明らかでなくなるのではないかと問題提起がなされたが、この点については、金融庁が「行政文書を電磁的記録により保存する場合には、内容が同一の紙媒体の行政文書については破棄することができる」という規定を設けている。また、第2回研究会での内閣府報告によると、内閣府では基本的に紙媒体を原本にする例が多いが、他省庁では紙媒体と電子媒体を同じ文書で両方同時に保存している例もあり、紙媒体、電子媒体による文書のうちのいずれを原本とするかは択一的ではないとのことである。

ウ. 電子行政文書の保存形態

各府省庁の文書管理規則等は、大別して「集中管理型」と「分散管理型」とに分けて考えることができる。このことは、各府省庁における電子行政文書の保存形態においても反映されているようである。

「分散管理型」の場合は、各課室レベルで電子公文書を保存している。国土交通省(主務課ごとにホストコンピュータで管理されている磁気媒体又はサーバの共用部分等において、電子行政文書ファイルを作成して保存する)が代表的である。

「集中管理型」の場合は、全省的に統一した保存を行っている。環境省(文書の処理にあたっては、電子文書化することが困難な場合を除き、原則として総合文書管理システムを利用するよう努めるものとする)が代表的である。

エ. 電子文書安全性確保のための措置

一部の府省庁では、電子文書安全性確保のための措置の必要性について言及している。

- 「行政文書は、保存期間が満了するまでの間、改ざん、漏えい等を防止する措置を講ずるとともに、必要に応じ記録媒体の変換等の措置を講ずること」(厚生労働省)
- 「マイクロフィルム文書は、決裁文書に代えて保存することができる」(文部科学省。なお、同様の規定が総務省にもある。)
- 「保存期間が満了する日まで、必要に応じ記録媒体の変換を行うなどにより、適切かつ確実に利用できる方式で保存しなければならない」(防衛庁)

また、文部科学省の文書管理規則では、アクセス管理に関する規定が設けられており、「総括文書管理者は、システムにアクセスしようとする者を、ユーザ ID 及びパスワードにより識別し、認証するものとする。」、「システムへのアクセスの状況を監視し、記録するものとする」、「職員は外部から入手した電子文書の利用に際しては、必ずウイルスチェックを受けなければならない」などとしている。

オ. 受信文書への電子署名

一部省庁では、電子署名に関する規定がある。

- 「受信した電子文書に電子署名がある場合は、当該電子署名の検証を行うこと」(環境省)
- 「受信した電子文書に公印の電子署名があるときは、当該電子署名の検証を行うこと」(文部科学省)

一般的に電子署名については署名付文書の検証が中心で、幅広く電子署名を行うこととされているわけではないが、電子署名法(電子署名及び認証業務に関する法律)の主管省庁でもある総務省については、「発送する文書が電子文書である場合にあっては、電子署名を行なうものとする(省内に発送する文書その他部長が電子署名を要しないと認めたものを除く)」と、電子署名の必要性を高めた規定を設けている。

(6) 保存期間

行政文書の保存期間は、文書管理規則等によって文書の類型ごとに保存期間が定められている。類型ごとの最低保存期間は、概ね次のとおりである。

文書区分		保存期間
1	法律又は政令の制定、改正又は廃止その他の案件を閣議にかけるための決裁文書	30年
	省令、その他の規則の制定、改正又は廃止のための決裁文書	
	行政手続法(平成5年法律第88号)第2条第3号に規定する許認可等をするための決裁文書であって、当該許認可等の効果が30年間存続するもの	
	国又は行政機関を当事者とする訴訟の判決書	
	国有財産法(昭和23年法律大73号)第32条に規定する台帳	
	決裁文書の管理を行うための帳簿	
2	公印の制定、改正又は廃止を行うための決裁文書	10年
	内閣府設置法(昭和22年法律第5号)第37条若しくは第54条又は国家行政組織法(昭和23年法律第120号)第8条の機関の答申、建議又は意見が記録されたもの	
	行政手続法第5条第1項の審査基準、同法第12条第1項の処分基準その他の法令の解釈又は運用の基準を決定するための決裁文書	
	許認可等をするための決裁文書であって、当該許認可等の効果が10年間存続するもの	
	不服申立てに対する裁決又は決定その他の処分を行うための決裁文書	
3	栄典又は表彰を行うための決裁文書	5年
	法律又はこれに基づく命令により作成すべきものとされる事務及び事業の基本計画書若しくは年度計画書又はこれらに基づく実績報告書	
	独立行政法人、特殊法人、認可法人又は民法(明治29年法律第89号)第34条の規定により設立された法人の業務の実績報告書	
	許認可等をするための決裁文書であって、当該許認可等の効果が5年間存続するもの	
	予算決算及び会計令(昭和22年勅令第165号)第22条に規定する書類又はその写し	
4	取得した文書の管理を行うための帳簿又は行政文書の廃棄若しくは移管の状況が記録された帳簿(情報公開法施行令第16条第1項第9号の記録を含む。)	3年
	許認可等をするための決裁文書であって、当該許認可等の効果が3年間存続するもの	
	所管行政上の定型的な事務に係る意思決定を行うための決裁文書	
5	調査又は研究の結果が記録されたもの	1年
	許認可等をするための決裁文書	
	所管行政上の軽易な事項に係る意思決定を行うための決裁文書	
6	所管行政に係る確認を行うための決裁文書	事務処理上必要な期間(1年未満)

図表 2-1 行政文書の最低保存期間基準(抜粋)

保存期間満了後の保存期間の延長の実態については、統計的なデータはない。ただし、保存期間満了に伴う歴史公文書等の移管に関連して実施された各府省職員へのアンケートによると、移管後の文書の公開の扱いや担当職員による文書へのアクセス等の懸念から、国立公文書館へ移管するのではなく、保存期間を延長して各府省が管理し続けることが少なくないとのことである(「歴史資料として重要な公文書等の移管に関するアンケート調査」(平成16年4月 内閣府))

(7) 今後の課題

電子公文書等の管理、移管、保存のあるべき姿を考える場合、現状の文書管理規則等が抱える問題として、次のような点を指摘することができる。

ア. 文書ファイルの明確化

文書管理規則等においては、管理される単位は「行政文書」となっているが、そのカバーする範囲については、各府省側に委ねられており、その実態は明らかではない。また、今回の調査では、文書ファイルがどのような構成(例えば、決裁文書以外の構成(議事録、参考資料等が添付されているか否かなど))となっているか明らかになっておらず、国立公文書館から見て、個々の行政文書ファイルに移管、保存価値のある行政文書が網羅されているかは明らかでない。

イ. 分散管理型への対応

各府省において定められている文書管理規則等の規定からは、システム整備面を始めて、集中管理型(環境省等)と分散管理型(国土交通省等)とに類型化することができるが、特に分散管理の場合、文書分類や作成者等の項目において、文書作成部局職員に運用が委ねられ、メタデータの体系化・標準化等の際に支障が生じる可能性がある。

ウ. 長期間の保存、管理に伴うリスク

文書管理の運用は、文書の保存期間中一定レベルに保たれる必要があるが、電子文書の保存・管理に特化した技術及び運用の観点に立った場合、そのような面での詳細な規定は、文部科学省を除いて、見当たらなかった。文部科学省では、文書管理規則の中で、システムへのアクセス管理やウィルス対策に関する規定が示されていた。

エ. 国立公文書館への移管に関するルールに関する懸念

文書管理規則等は、あくまでも各府省の事務の適切かつ能率的な遂行に資するとともに、行政機関情報公開法の適正かつ円滑な運用に資する目的から定められている。そのため、国立公文書館への移管に関するルールについては、国立公文書館への移管、移管した文書についての行政文書ファイル管理簿への記載など、最低限の項目が設けられているのみであり、移管後の保存や検索、利活用等までは、必ずしも視野に入っていないといえよう。

また、保存期間満了後の行政文書の扱いとしては、保存期間の延長、国立公文書館への移管若しくは廃棄の3通りがあるが、現在のルールではその選択は各府省側に委ねられている。そのため、国立公文書館側が歴史資料として価値があり移管の対象であると考えた公文書等についても、各府省側の判断により保存期間の延長や廃棄の可能性がある。

オ. 府省間で異なる電子公文書等のファイル形式

調査の対象とはなっていないが、現在の文書管理システムや各種電子文書作成のアプリケーションは各府省に、更新時期も含め一任されている。

したがって、行政事務等の電子化は進んでいるが、電子公文書等の国立公文書館への移管、保存にあたって、ファイル形式の統一や標準化を図ろうとした場合、現状では互換性や長期間経過後の再現性の確保について、課題があるものと思われる。

カ. 個人単位での保存、管理の実態

文書管理規則等上、各課等の文書管理者が行政文書ファイルの管理を部署単位で行っていることについては説明したが、その他個人的に作成しローカルの端末上で保存している電子ファイルが実態は多数を占めているのではないかとされている。

行政文書の電子化のメリットとしては、安全性や共有、検索の利便性といったものが一般的にはあげられているが、個人端末上で保存されている文書については、そういった利便性は極めて限られたものとなることから、今後各府省が電子情報の共有という観点でどのような取組みを行っていくかが課題といえよう。

キ. 紙文書の並存により生じる課題

各府省庁に保存されている現用文書の電子文書ファイル数は、全現用文書ファイル数の4%程度である(第6回研究会資料・各省庁現用文書推計調査結果概要(速報)より)。このことから、公文書館への移管が想定される、非現用化に近い文書類については紙が中心になるものと思われる。また、行政文書ファイルに紙と電子が混ざっている場合、紙を電子化することが難しいため、保存の実態は紙が中心になってしまう可能性がある。

2-1-3 行政文書の安全な保存に関連する政府の取組み

文書管理規則等の中では、電子公文書等の安全な保存については一般的な記述が目につく。一方、政府全体のレベルでの電子公文書等の安全な保存を考える上で関連がある動きとして、情報システムのセキュリティに関するガイドラインや基準が決定されている。

(1) 情報セキュリティポリシーに関するガイドライン(平成12年)

内閣官房副長官(事務)を議長とし、高度情報通信社会推進本部に設置された全省庁の局長級を構成員とする会議母体である情報セキュリティ対策推進会議が平成12年(2000年)7月「情報セキュリティポリシーに関するガイドライン」を決定した。同ガイドラインは、各府省が情報セキュリティポリシーを策定するにあたって参考に資するため、情報セキュリティポリシーに関する基本的考え方、策定、運用及び見直し方法について記述したものである。

(2) 政府機関の情報セキュリティ対策のための統一基準(平成17年)

昨今、国内の企業等においては重要情報の漏洩問題が相次ぐなど、情報セキュリティ関連の事故が多発している状況等を受け、政府のIT戦略本部の下に設置された情報セキュリティ政策会議において、平成17年9月、政府機関全体として統一的にとるべき対策のうち、緊急性の高いものについて、「政府機関の情報セキュリティのための統一基準(2005年項目限定版)」が決定された。さらに、平成17年(2005年)12月に、システムの開発・整備に関する対策項目などを追加した「政府機関の情報セキュリティ対策のための統一基準(2005年12月版[全体版初版])」が決定された。

「政府機関の情報セキュリティ対策のための統一基準(2005年12月版[全体版初版])」は、各府省庁の情報セキュリティ対策内容の整合化・共通化を促進するために、各府省庁がとるべき情報セキュリティ対策を定めたものであり、情報セキュリティ対策を「組織と体制の構築」、「情報についての対策」、「セキュリティ要件の明確化に基づく対策」、「情報システムについての対策」、「個別事項についての対策」に分類して対策基準を定めている。今後は、各府省庁において、当該基準に沿ってセキュリティ対策が更新されていくと想定される。

2-1-4 総合文書ファイル管理システム (e-Gov)

(1) 役割

総合文書ファイル管理システムでは、各府省が管理する行政文書ファイルの情報を国民がインターネットを通じて検索できるように、全府省庁の行政文書ファイル管理簿を、インターネット上の総合窓口 (e-Gov) を通じて、提供している (<http://files.e-gov.go.jp/servlet/Fsearch>)。

総務省行政管理局によると、行政文書ファイル管理簿については、平成 18 年 3 月末現在約 1765 万件のデータが登録されている。

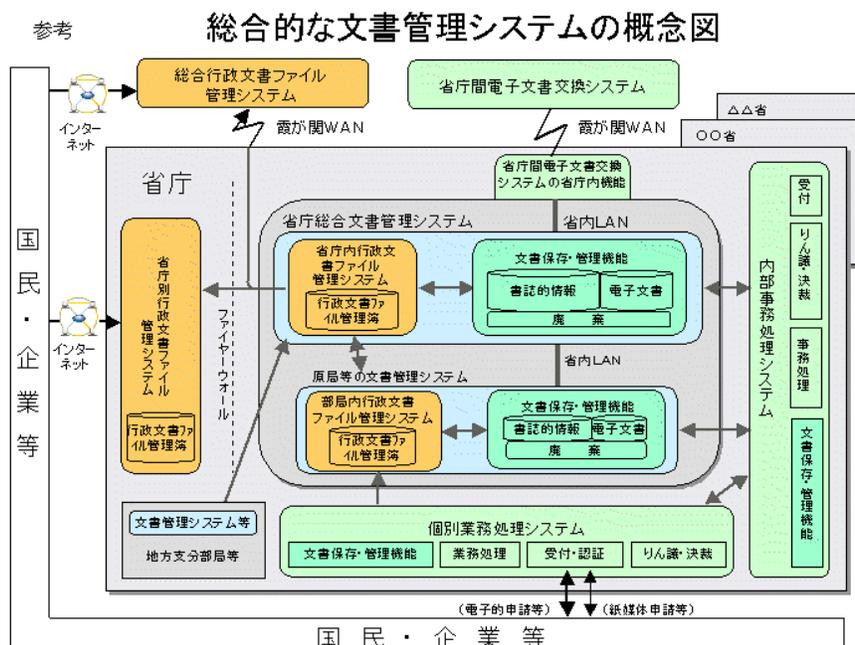
(2) システムの概念

総合文書ファイル管理システムは行政文書を管理対象としており、書庫等で管理されている紙文書や磁気ディスク等で管理されている電子文書等、あるいは「行政機関情報公開法施行令」(平成 12 年政令第 41 号)第 13 条第 2 項第 1 号に規定する行政文書ファイルを体系的に整理したものである。電子文書については電子文書本体及び書誌の情報、電子文書以外についてはその書誌的信息、行政文書ファイルについては書誌的信息を基にした情報を基に総合的に管理するものとしている。

府省庁間等で流通する電子公文書については、「電子公文書の文書型定義 (DTD) の統一的な仕様」(平成 10 年 3 月 31 日各省庁事務連絡会議了承、行政情報システム各省庁連絡会議幹事会了承)に定める標準 DTD (Document Type Definition) を活用するものとしている。

また、電子化方法についても、以下のとおり規定を設けている。

- 本文、表、XML タグ、図、表及び写真の標題、注書等はテキスト方式により電子化
- 図、写真等その他テキスト形式による電子化が困難なものは、国際的な標準の形式である Windows BMP、JPEG 等によりイメージデータとして電子化
- その他必要に応じ、添付ファイルとしてアプリケーションソフトによる電子化も可能
- 1バイト文字の英数字及び記号は、JIS X0201 を使用
- 2 バイト文字は、JIS X0208 を使用し、漢字については、JIS 第 1 水準漢字及び JIS 第 2 水準漢字
- UCS (ISO/IEC10646-1, JIS X0221) 及び JIS X0213 については、今後の各省庁の使用状況等を踏まえ、その扱いを検討
- 文書の保存、交換の際の符号化方式には、シフト符号化表現 (シフト JIS) を使用
- 利用頻度が高い外字については、第2に定める標準 DTD に定義する。



図表 2-2 総合的な文書管理システムの概念図
 (<http://www.soumu.go.jp/gyoukan/kanri/sogobun.htm>)

(3) 統一されたデータ項目

各府省の行政文書ファイル管理簿について、「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)で定められているのと同様、総合文書ファイル管理システムにおいても、各府省庁の提出したデータは所定の項目に基づいて記述されている(4-3-2 参照)。

(4) 運用ルール

主管である総務省行政管理局関係者へのインタビュー結果によると、行政文書ファイル管理簿の情報の元データは、各府省のシステムで管理されていることから、総合文書ファイル管理システム(e-gov)では、各府省から定期的に電子媒体で送られてくる上記データ項目に基づくデータの内容チェックを行った後、そのまま吸い上げる形で、データの登録・更新・削除を行っている。

従って、e-Gov に関する運用ルールは特に設けられておらず、各府省との行政文書ファイルに関するデータのやり取りに関する最小限の取り決めがあるだけとのことである。

ア. データの追加、変更、削除

行政文書ファイル管理簿情報の整備、変更、追加、削除等を行う際、関係する府省が管理簿情報の差分もしくは全部を送付し総務省側はそれを反映している。削除されたデータのアーカイビングは特段行われていない。

イ. データの送付方法

送付方法は、霞が関 WAN 上の電子メール/FTP 経由、もしくは磁気テープ等による媒体経由(オンラインではない)。データ形式は、CSV 形式のテキストデータを採用している。

(5) 今後の予定

行政文書の管理については、現在、政府の情報システムに関する最適化計画の対象の一つとなっており、平成 18 年度中に、現在各府省に存在する公表用のウェブシステムを、総合文書ファイル管理システム(e-Gov)一つに統一することとされている。すなわち、「各府省が行政文書ファイル管理システムを整備して行っている行政文書ファイル管理簿の公表については、総務省が e-Gov において運用している総合行政文書ファイル管理システムを利用して行うこととし、各府省は別紙 2(略)の見直し計画に沿って見直しを行う」(行政情報の電子的提供業務及び電子申請等受付業務の業務・システム最適化計画(平成 17 年 8 月 24 日 各府省情報化統括責任者(CIO)連絡会議決定)こととされている。

今後も各府省ごとにデータ管理を行うという仕組み自体は変わらないことから、各府省が有する行政文書ファイル管理簿情報のデータベースについては今後そのまま残ることとなるが、各府省が(インターネット経由での)公表用として提供しているウェブシステムはなくなることとなっている。

(6) 電子公文書の移管、保存から見た課題

e-Gov の抱える課題としては、省庁の行政文書管理、保存が抱える課題と共通点が多いとともに、窓口のみの機能を提供することによる課題も存在している。

ア. 体系化されていない記述用語

記述情報については、既に説明したとおり、項目レベルでの定義はされているが、項目名の定義については各府省庁に完全に委ねられているため、仮に全府省の登録済み行政文書目録を基にして記述用語の体系化を図ろうとしても、そのような試みは実を結ばない可能性が高い。

記述用語体系化の試みとしては、オーストラリア NAA では、シソーラス(記述用語集)を発行し、各政府機関が自由に使えるようにしている。

イ. 文書内容の特定が困難

全府省の行政文書の目録管理情報として、今後の電子公文書管理と密接な関係にあると思われるが、現時点では、e-Gov 上で管理されている記述情報から個別文書の具体的内容を特定することは、きわめて困難である。

ウ. 情報の追加、保存、更新、削除が各府省庁に依存

目次情報については既に説明したとおり項目レベルでの定義はされているが、項目のネーミングについては各府省庁に完全に委ねられているため、仮に全府省庁の登録済み行政文書目録を作成しようとしても、体系がバラバラなため辞書にはならないという欠点がある。

エ. アーカイブしていない

各府省に委ねている削除について、削除の実態が管理されていないことと、削除後のデータがアーカイブされていないということがある。従って、非現用の公文書の移管、保存の対象としようとしても、e-Gov 上に残っている情報だけでは全ての確認がむずかしいという事態が生じるリスクがある。

2-2 各府省における電子文書化の進捗

2-2-1 行政事務の電子化状況

総務省は、平成 11 年(1999 年)12 月高度情報通信社会推進本部で決定された「行政事務のペーパーレス化(電子化)に関する行動計画」に基づき、平成 12 年度～平成 14 年度の 3 か年計画で各府省内の行政事務の電子化への取り組みに関する調査を実施した。また、平成 17 年度に内閣府においても各府省の現用文書の保管、管理状況について調査を実施した。

(1) 電子化進捗度

行動計画の最終年度に当たる平成 14 年(2002 年)度の各府省における電子化の実施状況は以下のとおりであり、平成 12 年度から 3 か年にわたり取り組んできた行政事務の電子化は、実施に当たり電子化が困難なものを除き、電子化対象 57 事務(文書)は、ほぼ電子化(一部電子化を含む)されたとしている。

		平成12年度	平成13年度	平成14年度
電子化されている		72.10%	81.00%	94.50%
	すべて電子化	20.20%	24.40%	36.50%
	一部電子化	51.90%	56.60%	58.00%
電子化されていない		27.90%	19.00%	5.50%

図表 2-4 府省全体での電子化状況

	平成12年度	平成13年度	平成14年度
総務・文書部門	75.50%	84.10%	90.40%
協議・調整部門	84.90%	86.30%	97.70%
秘書・人事・給与等部門	69.40%	80.90%	95.00%
会計・経理・予算部門	51.80%	62.20%	89.70%
部門共通	83.80%	92.20%	98.70%

図表 2-5 事務(文書)分野別電子化の実施状況

(2) 媒体別の保存割合

職員一人当たりの媒体別情報保存割合は(平成 15 年 3 月末時点)、紙による保存が約 3.2 万枚(4.5%)、電子媒体による保存が約 67.2 万枚(95.5%)と、平成 14 年 3 月末時点(紙 12.9%、電子媒体 87.1%)と比べ、電子による保存の割合が大幅に増加している。ただし、保存情報の絶対量自体が増加しており、紙による保存も絶対量は増加している。

また、電子的保存の種別を見ると、磁気ディスク(LAN 上のサーバ)が電子的保存全体の大部分を占めている。

2-2-2 行政文書の電子化率

紙媒体と電子媒体による行政文書の作成比率を行政文書ファイル管理簿上で見ると、ペーパーレス化(電子化)の進捗状況のデータと違った数値を示している。これは、行政文書ファイル管理簿上の情報には、現在保存期間中の昔作成された文書や受領した(原本が紙の)文書も含まれているため、また、行政文書ファイルは最終的な保存状態に関するものであるためである。

今回サンプルとして、総務省の行政文書ファイル管理簿ベースで、電子媒体で保存されているものと紙媒体で保存されているものの割合を調査してみた(調査日 2006年1月31日)。総務省の行政文書ファイル管理システム(<http://www.soumu.go.jp/jyoho/kanribo.html>)から、「媒体の種類」が「電子」及び「紙」であるものを検索した。その結果、総務省の行政文書ファイルは全部で 25万5861件であったが、そのうち電子媒体により保存されているものは、1万7613件(6.8%)であった。

ちなみに、第6回研究会資料「各省庁現用文書推計調査結果概要(平成16年作成・保存分を調査)」によると、18省庁で1年間に作成されている電子文書ファイル数は約1.4万件で、全行政文書ファイル数の9%を占めている。また、18省庁に保存されている電子文書ファイル数は4.3万件で、全現用文書ファイル数の4%を占めている。保存期間別にみると、電子文書の作成及び保存状況においては、「5年保存」以下の相対的に保存期間の短い文書が多いことが示された。

2-2-3 海外の取組みも踏まえた今後のあり方

電子化に伴う文書管理については、単に文書管理ルールを定めるだけでは不十分で、電子媒体による文書の特性を踏まえた独自の詳細なガイドラインが必要となるであろう。

(1) カナダの事例

カナダでは、「Records/Document/Information Management System (RDIMS)」において、電子公文書のライフサイクルという視点から、公文書情報の作成、保存から再現にいたるまでの技術的なアーキテクチャ、ユーザの立場に立った要件(インターフェース等)、機能要件等を充足するソフトウェアを開発した。政府機関は、これを購入して電子公文書の作成・管理等に活用している。これは、90年代後半に作成された枠組みで、既に成熟したものであるが、今後は職員の理解等いかに効果的なメカニズムにするかが視野に入れているようである。

(2) オーストラリアの事例

「DIRKS Manual」が取りまとめられ、電子記録の作成・管理等に係る要件全般、現行システムの評価、レコードキーピング戦略のあり方、システムのデザイン、システム運用・レビュー等についてまとめた、行政機関に対して活用される包括的なマニュアルが策定された。

オーストラリアの各行政機関では、電子公文書管理に対する評価、戦略立案等の場での活用が可能なものになっているとのことである。

2-3 電子公文書基盤

電子公文書基盤とは、電子公文書等が作成、管理、保存、利用等される情報システムの土台となる部分を指す。電子公文書等が作成・利用される範囲は広く、公文書等が作成される各府省のイントラネットの範囲だけでなく、移管の際には、各府省間を結ぶ霞ヶ関 WAN 等のネットワークも関係してくる。また、イントラネット内においても、各府省の文書管理システムだけでなく、場合によっては、グループウェアシステム等にも電子公文書等が作成等されていることもある。

ここでは、電子公文書基盤である情報システムの現状を整理し、電子公文書等を移管する場合の課題等を検討する。

2-3-1 霞が関 WAN

霞が関 WAN は各府省の LAN を相互に接続する政府内専用ネットワークで、電子メールや霞が関 WAN インフォメーションボード、電子文書交換システム等が接続されている。また、霞が関 WAN では計 26 のサービスが有料で提供されている。

(1) 霞が関 WAN について

霞が関 WAN は、政府機関内における情報の円滑な流通・共有を図るため、各府省の LAN を相互に接続する政府内専用ネットワークで、平成 9 年(1997 年)1 月から運用を開始しており、社団法人行政情報システム研究所にて管理・運用している。

現在のところ 30 機関で利用されている。この中には、各府省と密接な業務上の関連がある組織、例えば、国立公文書館のほか、国家公務員共済組合連合会や独立行政法人国立印刷局なども含まれている。

(2) 霞が関 WAN の利用状況等

霞が関 WAN を利用する主なシステムとしては、以下が挙げられる。今後、人事・給与等業務・システム、共済業務・システム等でも利用される予定である。

ア. 霞が関 WAN を利用している主なシステム等

- 電子メール
- 霞が関 WAN インフォメーションボード(掲示板、バーチャルフォーラム、法令等各種協議)
- 共通報検索システム(法令検索、国会会議録検索、閣議情報検索)
- 歳入金電子納付システム
- 審議会委員等兼職データベース
- 情報システム統一研修オンラインシステム

イ. 霞が関 WAN の利用頻度(平成 16 年度)

霞が関 WAN の利用頻度としては、単純比較はできないが、件数で言えば電子メールが最も多くなっている。

- 電子メール: 944 万件(うち地方公共団体との間: 327 万件)
- 電子掲示板: 129 万アクセス(うち地方公共団体との共用掲示板: 13 万アクセス)

ウ. 霞が関 WAN の電子メール

システム開発当初は、SMTP を用いた場合、電子メールの到達確認が出来ない等の問題があったため、X400 を採用した。しかし、外部メールは SMTP によっているため、外部 - 内部間のメール受渡しについては、逐一プロトコルの変換をかける必要があり非効率である。そのため、現在は、SMTP を用いた到達確認の技術も確立しているところから、全て SMTP に統一することにした。個々の電子メールの管理については、各府省が行っているため、霞が関 WAN としては特に電子メールを保存するようなことは行っていない。

エ. 霞が関 WAN インフォメーションボード(バーチャルフォーラム、掲示板、法令協議等)

「霞が関 WAN インフォメーションボード」は、霞が関 WAN 内に設けられたウェブによるインフォメーションサービスであり、以下の機能を有する。

- ニュース関連: 新着情報、メンテナンス情報、運用センターからのお知らせ、霞が関 WAN 関連情報、電子文書交換情報、LGWAN との相互接続関係情報
- 掲示板関連: 各利用機関の掲示板、各種施策等の掲示板、掲示板内検索
- フォーラム関連: バーチャルフォーラム、法令協議システム
- その他: リンク集、意見・要望

内閣府では、バーチャルフォーラムの 6 個のフォーラムを活用しており、中でも内閣府本府協議フォーラムについては、各省庁との法令協議に活用されているとのことである。(第 2 回研究会)

特にバーチャルフォーラム、法令協議システムについては、成立前の法令等のテーマの協議を行うものであるが、管理者が運用センターに申し込むことにより開設され、開設期限が過ぎると自動的に削除される。各コンテンツについては、可用性確保のためバックアップは取得しているものの、情報の保存目的でのアーカイブは特にしていないとのことである。

なお、霞が関 WAN 上のウェブシステムは、「霞が関 WAN インフォメーションボード」のみであり、その他には存在しない。

オ. 電子文書交換システム

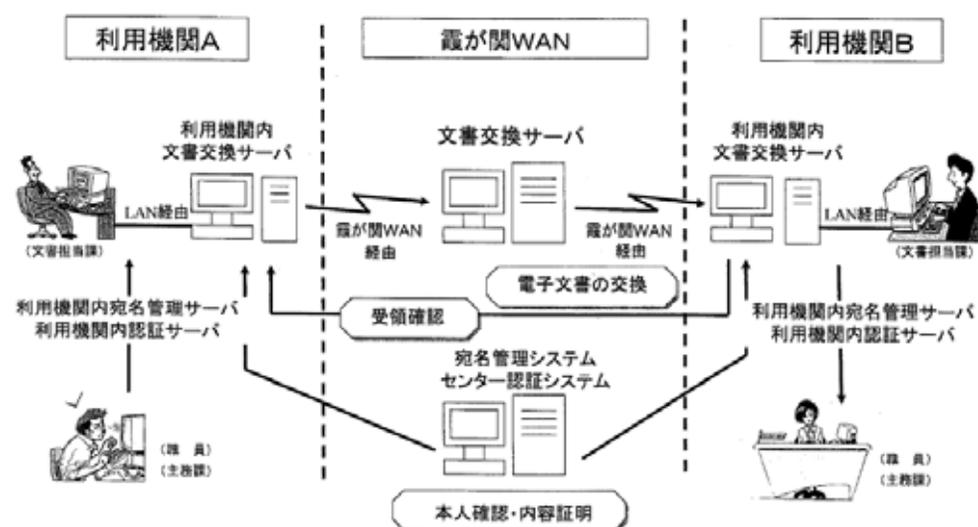
電子文書交換システムとは、各府省の文書取扱主任間の電子文書交換について、暗号化通信により、非改ざんの信憑性を担保するシステムである。

実際の電子文書は、各府省のサーバにあり、電子文書交換システムでは、その中継のみを担当している。各府省の文書取扱主任の情報をディレクトリで管理しており、それぞれに暗号化キーの入った IC カードを渡している。

交換する電子文書については、XML で鑑(かがみ)を作成し、それに添付ファイルを付ける形で交換している(XML については、平成 14 年から採用)。なお、文書交換を行ったログは残るが、どのような文書を交換したかといった情報までは管理していない。また、情報の保存目的でのアーカイブは行っていない。

システムは、国際標準規格を採用、認証証明書の形式は、ITU-T(国際電気通信連合電気通信標準化部門)の X.509、宛名管理機能は、ITU-T の X.500、電子文書交換機能は ITU-T の X.400 に準拠しており、送受信システム、文書交換システム、認証システム、宛名管理システム等のサブシステムから構成されている。

総務省による説明では(第 1 回研究会)、電子文書交換システムで交換された電子文書は原本性を持つという位置づけがされており利用価値の高い一方、現段階では同システムの利用率はあまり高くないということである。



図表 2-6 電子公文書交換システムのイメージ(総務省行政管理局提供資料)

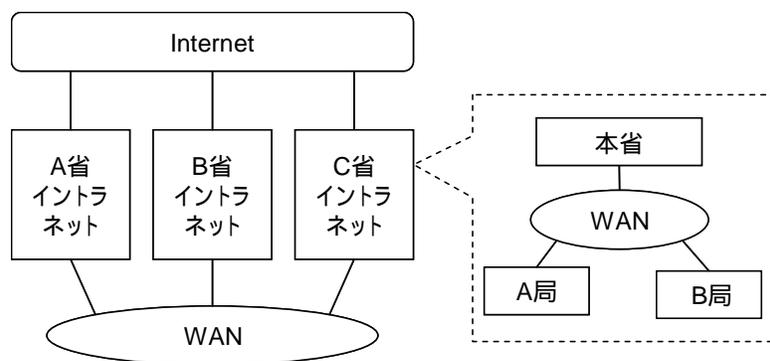
2-3-2 イン트라ネット

各府省のイントラネットには、職員が利用する PC やメールシステム、ファイルサーバ等の様々なシステムが接続されており、インターネットの接続も可能である。

(1) ネットワーク概要

各府省のネットワークは、それぞれインターネットへの出口を持っており、例えば、国民向けのホームページの発信や職員のインターネット閲覧は、この出口を通じて行われる。また、各府省は WAN の接続口を持っており、府省間でやり取りされる情報はこの WAN を通じてやり取りすることができるネットワーク構成である。

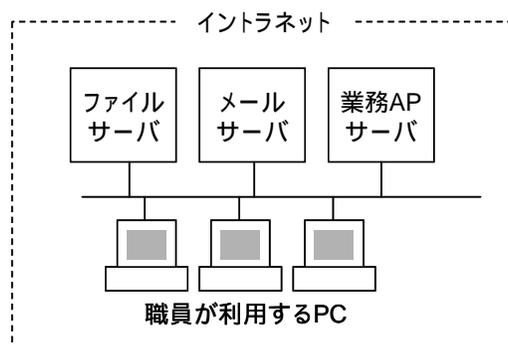
また、各府省は様々な地方支分部局等を有しているが、こうした離れた場所に対しても、一般的にはネットワークで結ばれており、本省とのデータのやり取りを行うことが可能となっている。



図表 2-7 イン트라ネットとWANの関係

(2) 各府省イントラネット

各府省のイントラネットには職員が利用する PC を始め、メールシステム、業務システム、ファイルサーバ等の様々なシステムが接続されている。そのため、職員は PC に導入されている Word や Excel などのアプリケーションソフトウェアを使って作成したファイルをファイルサーバに保存することや、メールに添付して関係者に送付することが可能となる。また、グループウェアシステムを導入している場合には、掲示板やスケジュールを利用した情報共有が可能となる。



図表 2-8 イン트라ネットの構成

ア. PC

Microsoft Office (Word, Excel, Powerpoint など)、一太郎、Internet Explorer などのアプリケーションソフトウェアが府省単位で導入されており、職員が一定のアクセス権限のもと利用することが可能になっている

イ. ファイルサーバ

ファイルサーバとは、自身の管理している記憶装置をネットワーク上の他のコンピュータと共有し、外部から利用できるようにするコンピュータのことで、ファイルサーバ上にあるファイルは(許可されていれば)誰でも他のコンピュータから読みこんだり書きこんだりできるため、データの一括管理が可能になる。

ファイルサーバの導入によりコンピュータ間でのファイル移動の煩雑さが解消され、また個々のコンピュータに保存されているファイルの整合をとる必要もなくなる。ネットワー

クに接続されていれば、他人が途中まで編集したデータにその場で手を加えて完成させる、といったことも可能になる。

各府省の現場では、部署単位、業務単位、地方組織単位等様々な単位でサーバが設けられているようである。

ウ. メールシステム

メールシステムとは、インターネット上に常に接続され、自ネットワーク内のユーザの電子メールの送信や受信を行なうコンピュータのことである。他ネットワークのユーザに電子メールを送信する SMTP サーバと、自ネットワーク内のユーザあてに送られてきた電子メールを保存し、ユーザからの受信要求に対応する POP3 サーバや IMAP4 サーバなどを合わせた呼び方でもあり、通常、府省単位で設けられている。

エ. グループウェアシステム

グループウェアシステムとは、LAN を活用して情報共有やコミュニケーションの効率化をはかり、グループによる協調作業を支援するソフトウェアの総称である。組織の円滑な運営を IT 面から支援するものであり、各府省でも採用している製品や詳細の機能の違いはともかく導入は進んできている。

オ. 文書管理システム

文書管理システムとは、官公庁や企業などの組織体において、文書を統合的に管理するシステムのことである。管理対象となる文書の内容・性格によりさまざまなシステムが存在し、ワークフローシステムやグループウェアなどと連動して利用されたり、あるいはこれらのシステムが文書管理の機能を持っていたりする。

行政文書の電子的作成の基盤ともなっており、単に文書ファイルのみならず、作成者、作成日、分類といった属性までも管理している。ただし、製品に依存する機能や構造もあることから、現時点では、各府省間での互換性まで保証されているわけではない。

カ. 各業務システム

各業務に合わせたシステムで、人事給与・会計・施設管理他各府省庁の共通業務あるいは個別業務に併せて多岐にわたる。

2-3-3 各府省文書管理システム

平成 12 年(2000 年)の各省庁事務連絡会議で、「総合的な文書管理システムの整備について」という申し合わせがなされ、文書管理システムを構築し、日常の文書管理を容易にしようという方向性が固まった。すなわち、単に文書を電子的に作成するだけでなく、電子文書作成時に文書分類、保存期間、保存場所等の設定ができるシステムを構築し、更に電子決裁や府省庁間での電子文書交換の実現までも視野に入れたシステムの整備が進められることになった(第 2 回研究会での事務局説明)。

各府省では、行政機関が作成した文書を組織的に管理するために「総合的な文書管理システム」の整備を進めている。

(1) 総合的な文書管理システムの概要

各府省の総合的な文書管理システムの概要を、ある省(以下「A 省」という。)の関係者へ実施したインタビュー結果をもとに、紹介する。

ア. 背景、目的、概要

平成 12 年の「行政文書の管理方針に関するガイドライン」に基づき、A 省では総合的な文書管理システムを導入した。行政文書は紙文書も電子文書も全て総合的な文書管理システムで管理している。

電子文書は総合的な文書管理システム内で電子文書本体を管理しており、紙文書は総合的な文書管理システム内で記述情報を管理している。e-Gov で管理されている記述情報にあわせて行政文書を管理している。なお、行政文書の定義は、行政機関情報公開法をベースに作成した行政文書管理規程で定められている。

イ. 利用の現状

全職員が総合的な文書管理システムを利用できる環境にあるため、総合的な文書管理システムに対する行政文書の登録は行政文書を作成した職員自身が行っている。職員は総合的な文書管理システム内で電子文書を作成することができる。また、既に電子文書として存在しているファイルをシステムに添付することも可能である。

行政文書を登録する際、「分類」等の文書管理情報は、総合的な文書管理システムで選択形式となっている。なお、「分類」の選択項目は各課室で登録することができる。文書管理に用いているファイル名は職員に一任している。

総合的な文書管理システムに登録した行政文書は保存期間を設定することができ、設定した保存期間を満了すると廃棄等の処理がされる。文書は完全廃棄であり、アーカイブはしていない。一方で、記述情報は 5 年間の保存が義務付けられているため、設定した保存期間が過ぎても総合的な文書管理システム内で管理をしている。

ウ. システムのセキュリティ対策

総合的な文書管理システムは、管理している行政文書に対して、「課室内」や「省内」といったアクセス権(参照権・編集権)を付与して、セキュリティを確保している。

A 省の総合的な文書管理システムはクライアント・サーバ型で構築されているため、クライアント PC にインストールされているグループウェアがバージョンアップされた段階で、総合的な文書管理システムの改修を実施する必要が生じる。

2-3-4 電子決裁システム

行政文書のうち決裁文書が占める位置づけが非常に高いことから、電子決裁システムの現状把握は、電子公文書等の移管、保存等のあり方を検討する場合の重要なテーマである。

(1) 機能、役割

電子決裁システムでは、例えば、決裁の際の電子文書は決裁完了時点でシステムに自動登録され、行政文書ファイル管理簿に登録されることとなっている。現に、総務省では少なくとも決裁手続関連文書は随時登録され、また決裁手続を必要としない事務文書も作成担当者による随時登録が可能となっている(第2回研究会事務局説明より)。

(2) 整備状況

電子決裁インフラについては、府省間で導入進捗度の違いがあるようである。研究会での内閣府関係者の説明によると、総務省ほかの一部省庁の一部の非常に定型性の高い業務(例えば旅行の決裁)について電子決裁システムが導入されている一方、内閣府では電子決裁システムは導入されていないようである。

(3) 先進事例(総務省)

第1回研究会において事務局より、総務省の電子決裁の普及状況が報告されている。それによると、年々、電子決裁率は増加している状況にある。

- 平成15年度:電子化率 8%(8万件的決裁総数のうち電子決裁システムを用いたものは7,100件)
- 平成16年度:電子化率 19.1%(8万9,000件的決裁総数のうち電子決裁システムを用いたものは1万7,000件)
- 2005年(6月9日)現在:電子化率 35.5%(1万6,000件的決裁総数のうち電子決裁システムを用いたものは5,900件)

2-3-5 電子申請

政府の行政文書の電子化の鍵の一つとして、外部からの電子申請の進捗がある。すなわち、申請の際に申請者から受け取る書類の電子化が進むことによって、行政文書の電子化も進むということである。

(1) 普及状況

現在電子申請が可能な手続きの電子化率は、極めて限られた部分を除いてはまだ1%を下回る厳しい状況が続いているとのことである。ただし、電子化が進んでいる手続きも一部存在しており、税関関係や特許関係の手続きの電子申請率については9割を超えている模様である(第1回研究会での事務局説明)。

(2) 課題

電子申請が導入されたからといって、政府内部での業務の電子化に必ずしもつながるわけではない。すなわち、第2回研究会での事務局説明によると、電子決裁システムが存在

していない府省庁では、仮に電子申請を受理した場合でも、内部での決裁の過程では一度紙に印刷し、紙で決裁を行った上で、最終的な回答部分だけ電子文書で行なっているとのことである。

2-3-6 グループウェア

グループウェアとは、グループ活動の全体的な作業効率の向上を支援するソフトウェアの総称で、電子メール、スケジュール、メンバー間で議論を行うための電子会議室や掲示板等の機能がある。

各府省庁ではロータスノート、TeamWare、サイボウズガルーン等のグループウェアソフトを導入しており(第6回研究会資料・各省庁現用文書推計調査結果概要)、グループウェアの一機能として文書管理システムを構築している事例もある。

2-3-7 データベース

ICA 報告書 16 において、「データベースは、長期保存の措置を必要とする記録を生成する、非常に良く見られるアプリケーションである。データベースは、登録台帳、特定事例の記録、記録簿、目録のような媒体の記録システムの形態と同等のことが多い。しかし、コンピュータでのデータベースは紙で作成された記録のシステムよりも、遥に複雑であり、多くの種類の情報が含まれており、様々な種類の情報間の相互関係も複雑である」とされているとおり、データベースが政府の行政で果たしている役割、抱えている情報量は非常に大きいため、移管の対象として議論の的の一つになる。

データベース内のデータは、それ単独では意味を成さないことがある。データベースから必要なデータを取り込み、表示するようなシステムが存在して初めてデータを利用することができる場合もある。

データベースに対する今後の扱いについては研究会の中でも活発な議論が行われた。すなわち、データベースそのものを電子公文書という形で残しておく必要があるのか、あるいは、データベースからアウトプットしての出力を行った上で、紙もしくはイメージファイルとして保存すべきかというものである(第4回研究会での委員発言)。

2-4 電子メール

2-4-1 位置づけ

電子公文書等の管理・移管・保存について考える場合、行政が作成する電子文書のひとつとして、先に示したウェブページの他にも、電子メールを挙げることができる。

電子メールは、政府の職員同士もしくは職員と外部との通信手段として用いられるコミュニケーションツールである。重要度は内容によって様々であるが、中には行政に関する重要な情報が掲載されている場合も想定される。そのような場合には、歴史資料として重要な公文書として移管・保存の対象とも成り得るものである。一方、研究会(第2回)の中で、組織共用されていない個人的なメールが行政文書に当たるかどうかという議論が委員の間で交わされた。

2-4-2 電子メールの種類

電子メールは、その用途・内容により、事業活動や政策進展の証拠となるものと、会議の通知や連絡等の事業の背景となる情報に大きく分けることができる。どちらも公文書とみなすことができるが、後者は長期保存の価値が低いものも多数含まれているものと思われる。

分類	概要
事業活動や政策進展の証拠となるもの	事業活動に関する内容が色濃く文書上に記載されており、政策の進展状況を把握することができる電子メール
事業の背景となる情報	会議の通知や連絡等、職員間のコミュニケーションツールとして使用された電子メール

図表 2-10 電子メールの分類

このような認識は、オーストラリア NAA の「E-mail is a record!」等、今回調査対象とした諸外国でも共通しており、あくまで前者を保存の対象とし、後者については、直接的な利用目的を達成した後は、積極的に削除することが推奨されている。

その他電子メールに関しては調査対象各国とも基本的に法レベルで電子メールが公的記録であると位置づけている。

(1) 海外事例 オーストラリア

オーストラリア国立公文書館の「E-mail is a record!」では、文書管理システムと電子メールのリンクを義務付けている。また、「ビジネスの活動や政策の進展の証拠となるもの」と「会議の通知や連絡等のビジネスの背景となる情報」とに区別して、それぞれの扱いの違いを明確化している。

- ビジネスの活動や政策の進展の証拠となるもの： オーストラリア国立公文書館の最終処分基準に従って維持しなければならない。
- 会議の通知や連絡等のビジネスの背景となる情報： 1983 年のアーカイブズ法の規定に基づいて廃棄・無効にする。

ソフトやハードが更新された場合でも、マイグレーション等の措置を講じることにより、引き続き電子メールの保存義務を負うこととしている。各省に以下に関するメール管理ガイドラインの策定を課している。

- 電子メールシステムの所有者とトランザクションログ
- 電子メールシステムの利用状況とアクセス権
- アーカイブズ法に従ってどのように処理をしたかメッセージの重要性を決定した責任者
- 記録管理システムの中に電子メールを組み込む作業の責任者
- 電子メールシステムのバックアップ責任者
- セキュリティやアクセス要件

(2) 海外事例 カナダ

カナダ LAC の「E-mail Management in the Government of Canada」において、電子メールは原則として記録であることを明記している。記録である電子メールと、(個人間のコミュニケーションや一時的な情報等の)非記録とを取り扱い上明確に区分し、後者は速やかに削除するべきであるとしている。カナダの場合、行政の活動、業務、意思決定の過程などのメールは記録であると定めており、自宅で送信、受信した電子メールも記録としている。

また、電子メールは受け取ったままの状態でおかなければならないとともに、行政文書管理システムとの融合が必要としている。電子メールの管理は組織の方針やガイドライン、手続きに基づいて行われなければならないとしている。

2-4-3 各府省における電子メール取扱いの現状

我が国の各府省では、電子メールの取扱いについて、文書管理規則等上に特段の規定を設けていない。霞が関 WAN 上では膨大な数の電子メールが流通しているが、霞が関 WAN 自体が電子メールシステムに参与しているわけではないようだ。また、各府省においても、保存容量による制限以外については、規定等に基づく管理は行われていないようである。

第 2 回研究会における総務省からの説明によれば、メールの保存に関するルールは特に存在せず、メールの削除については本人が自由に適宜実施しており、保存可能なメールの容量に限界があることから、各自が通常古いメールから削除しているのではないかとのことである。

電子メール機能は Notes(グループウェア)を利用している省もある。今回調査した結果によると、職員のメール受信ボックスは容量が 50MB に制限されており、制限容量をオーバーするとメールの送信が不可能になるという運用を行っている。また、メールの受信ボックスから行政文書に値するメールについては、文書管理システムに移すことで文書管理が可能となる。一方、行政文書でないメールは保存対象とは考えておらず、各職員の運用に任される。

3. 主要国と日本との比較

3-1 諸外国と日本との比較調査の背景

第2章においては、諸外国の事例も参照しつつ、我が国における電子政府化の進展状況、公文書基盤の現状等について説明したが、平成17年8月の内閣府の懇談会(第10回)における研究会の検討状況に関する配布資料の中で、電子文書管理・保存への取組み状況について、我が国と諸外国との比較が行われている。本章では、同資料及び懇談会、研究会における事務局側説明等に基づいて、電子政府化及び電子文書管理・保存への取組みの現状について、国立公文書館(及びそれに相当する機関)の関与の態様等を含め、アメリカ、カナダ、オーストラリアと我が国との比較を行う。

なお、欧米、主要諸外国の公文書館の職員数は、数百人から1000人を超える規模である一方、日本は40人強である(第3回研究会での事務局報告)。

3-2 電子政府化への取組み

3-2-1 アメリカ

OMB(Office of Management and Budget:行政管理予算局)を責任省庁とし、電子政府戦略(E-Government Strategy)が推進されている。

電子政府戦略(E-Government Strategy)は、連邦政府の効率性及び国民に対する応答性並びに公共サービスの向上を加速するための計画で、2003年までに省庁横断のチームが5分野に分類された24の電子政府イニシアティブに取り組むとしている。その中で国立公文書館関連の内容として、24の電子政府イニシアティブのうち、「政府内部の効率性と有効性向上」に分類された中に、NARA(国立公文書記録管理局)が担当する「電子記録管理」がある。

「電子記録管理」は、全ての政府機関において適用可能な、電子記録管理に必要なガイダンス等を提供するほか、多様なフォーマットの電子記録を将来の政府及び国民の利用に供するためにNARAに移管するためのツールを開発することを目標として設定している。

3-2-2 カナダ

Treasury Board of Canada Secretariat/Chief Information Officer Branch(国家予算委員会事務局 CIO 部)を責任省庁とし、コネクティング・カナディアンズ(Connecting Canadians)が推進されている。コネクティング・カナディアンズ(Connecting Canadians)は、行政のサービス及び情報をインターネットで提供することを目標とする「ガバメント・オンライン(Government On-Line)」等の6つの計画から構成されている。「ガバメント・オンライン(Government On-Line)」は、5ヵ年計画で、実現に向けて3段階の目標がある。

- 2000 年未まで:全ての省庁をオンライン化。
- 2005 年未まで:政府が提供する全ての行政サービスをオンライン化。
- 具体的な期限なし:省庁を横断した行政サービスの提供。

その中で、国家予算委員会事務局は、「ガバメント・オンライン」が目標とする行政のサービス及び情報のオンラインでの提供及びその充実には、情報管理 (Information Management) との連携が必要であるとしているが、政府の情報管理を所掌するのは、国家予算委員会事務局 CIO 部、公共事業サービス省及び LAC となっている。LAC は、情報管理に関する各種の標準等を発行している。

3-2-3 オーストラリア

AGIMO (Australian Government Information Management Office: 政府情報管理局) を責任省庁とし、ガバメント・オンライン戦略 (Government Online Strategy) が推進されている。

その中で、「適切な行政サービスの全てを 2001 年未までにオンライン化する」という目標を掲げ、8 項目の優先的戦略 (「認証、メタデータ規格、電子出版と記録保管ガイドライン、アクセシビリティ、プライバシー、セキュリティ等に関する条件整備」他 7 項目) が示されている。計画の具体的な時期は不明である。

- 2002 年 2 月: 「適切な行政サービス全てのオンラインサービス化」という目標の達成を宣言。
- 2002 年 11 月: より広範囲の統合された電子政府を実現するための戦略として、6 項目から構成される「よりよいサービス、よりよい政府 (Better Services, Better Government)」を発表。

また、電子政府の実行計画の 1 つとして、インターネット上の様々な情報を容易に検索できるようにするため、AGLS メタデータ標準 (The Australian Government Locator Service (AGLS) metadata standard) を、NAA (オーストラリア国立公文書館) が開発している。

3-2-4 日本

高度情報通信ネットワーク社会推進戦略本部 (IT 戦略本部) を中心とし、e-Japan 戦略が推進されている。e-Japan 戦略は、すべての国民が情報通信技術 (IT) を積極的に活用し、その恩恵を最大限に享受できる知識創発型社会の実現に向け、早急に革命的かつ現実的な対応を行わなければならないとし、市場原理に基づき民間が最大限に活力を発揮できる環境を整備し、5 年以内に世界最先端の IT 国家となることを目指す、とするものである。

その中で、2003 年までに、行政 (国・地方公共団体) 内部の電子化、官民接点のオンライン化、行政情報のインターネット公開・利用促進、地方公共団体の取組み支援等を推進し、電子情報を紙情報と同等に扱う行政を実現し、幅広い国民・事業者の IT 化を促すものとされている。

また、国立公文書館関連の事項としては、「e-Japan 重点計画 2002」において、重要な公文書等歴史資料等に関する情報を地理的な制約を受けずに入手・利用できる環境を整備することが明記されている。

3-3 電子文書管理・保存への取組み

3-3-1 公文書の定義及び電子公文書の位置づけ

(1) アメリカ

合衆国法典と連邦規則に、「記録」(Records)及び「電子記録」(Electronic record)についての規定がある。

- 「記録」(Records):「書籍、書類、地図、写真、機械により判読可能な資料、その他の記録資料で、物理的形態や性質を問わず、米国政府が業務の過程で、組織の機能・政策・意思決定・手続・事務処理等の証拠として作成又は收受したもの」(合衆国法典(44USC,3301))
- 「電子記録」(Electronic record):「コンピュータのみよって利用可能な形態で記録され合衆国法典(44USC,3301)の記録の定義を満たすあらゆる情報」(連邦規則 36CFR1234)

なお、連邦規則 36CFR1234 は、電子記録の作成・保存・利用・最終処分に係る基本要件を規定している。

(2) カナダ

国立図書館公文書館法に、「記録」(Records)及び「電子記録」(Electronic record)についての記載がある。

- 「記録」(Record):「媒体や様式を問わず、出版物以外のあらゆる記録資料」
- 「政府記録」(Government record):「政府機関の管理下にある記録」(国立図書館公文書館法(2004))

(3) オーストラリア

公文書法に、「記録」(Records)及び「連邦記録」についての記載がある。

- 「記録」(Record):「(手書きあるいは印刷物を含む)文書あるいは(録音、コード化した記録装置、磁気テープ・ディスク、マイクロフォーム、写真、フィルム、地図、図面もしくは模型、絵その他の描画・グラフィック作品等の)物で、当該文書等に含まれるもしくは当該文書等から導き出すことが可能な情報もしくは事柄を理由に、あるいは何らかの事象、個人、状況等と関係があるという理由で、保管されているあるいは保管されてきたもの」
- 「連邦記録」:「オーストラリア連邦もしくは連邦機関が所有する記録」(公文書法(1983))

(4) 日本

公文書館法及び国立公文書館法に、「公文書等」についての記載がある。

- 「公文書等」:「国又は地方公共団体が保管する公文書その他の記録(現用のものを除く)」(公文書館法(1987)、国立公文書館法(2000))

また、「記録の媒体については、文書、地図、図面、フィルム、音声記録、磁気テープ、レーザーディスク等そのいかなを問わない」(「公文書館法の解釈の要旨」(1988、総理府))とあり、電子媒体によるものも含まれるものと解釈される。

3-3-2 行政機関の公文書作成・保存への国立公文書館の関与

(1) アメリカ

NARA は、政府機関が政策及び業務を十分かつ適切に記録するために、政府機関に対して指導と助言を与えなければならない。また、NARA は、公文書の保存、廃棄のスケジュール「Record Schedule」の作成指導権限を有する。行政機関による連邦記録の処分にあたっては、公文書記録管理局長官の許可が必要である。

また、ERM (エレクトロニック・レコード・マネジメント) すなわち電子記録管理に関して NARA は非常に深く関与している。これは ERM が、紙ベースの文書管理及び記録管理のかかわり方の延長線上で行われているためである。

ERA (Electronic Records Archives) プロジェクトでは、文書記録に加え、地図、音声、画像といったすべてのポーンデジタルの記録を同一フォーマットで保存していくということを目指している。例えば、1960 年代につくられたベトナム戦争時の電子データの地図も現在でも見られるような形で、将来見られるような形で保存していく、また、文書に関しても、現在の文書を将来的に可読できる状態で保存していくことを目指している。

(2) カナダ

政府機関記録の処分には、国立図書館公文書館長の許可が必要である。

国家予算委員会事務局が 2003 年 5 月に発表した情報管理戦略(「Information Management Policy」)においては、国立公文書館は、継続的保存価値を有する政府記録の選定・収集・保存、記録処分規準(Records Disposition Authorities)の発行、記録・情報ライフサイクル管理に関する指導・助言、標準・ガイドライン等の策定等を行うこととされている。

LAC (国立図書館公文書館) は、WEB 上で、記録・情報の作成・保存・管理等に関するガイド・ツール等を提供している。また、LAC では、公文書管理・記録管理についても新しい考え方を導入している。それは、「記録」(レコード)という枠にとらわれず、図書・雑誌等の出版物等を含めて、広く「情報」(インフォメーション)という概念に基づいて、「情報」の総合的管理を行っていかうという考え方である。

(3) オーストラリア

連邦記録の最終処分(国立公文書館への移管又は廃棄)には NARA の承認が必要である。

NAA は、Web 上に、記録管理の方法論である DIRKS を始めとする 2,000 ページを超えるガイドライン、標準、アドバイスおよび教材を政府向けに提供している。

DIRKS には、紙でも電子でも媒体を問わず、公文書の作成・管理を行う為の仕組みをどのように設計・導入していくかが記載されている。公文書の評価選別の際、これをツールとして用いなければ公文書の廃棄はさせないと、国立公文書館の方から宣言したもので、ISO15489 のモデルともなっている。

公文書の記録管理におけるモデルケースとしてオーストラリアは最先端であり、紙文書だけでなく電子文書に関しても、かなり早い時期から取組みを行っている。

(4) 日本

行政機関における公文書等の作成・管理等について、国立公文書館の制度的関与はない。

3-3-3 メタデータの標準化へ向けた取り組み

(1) アメリカ

NARA は、Federal Register (いわゆる官報) に掲載する文書について、メタデータに相当する文書属性や構造を定義した「Federal Register Document Drafting Handbook」を発行している。また、NARA が承認した国防総省電子記録管理アプリケーション設計基準 (DoD 5015.2-STD) には、メタデータ標準が盛り込まれている (詳細は 4-3-7 諸外国のメタデータ事例参照)。

(2) カナダ

国家予算委員会や議会図書館、LAC 等が組織する連邦図書館評議会 (Council of Federal Libraries) が Web 情報のメタデータに関するガイド (「Government of Canada Metadata Implementation Guide on Web Resources」) を発行している。同ガイドでは、メタデータ構造に加え、一部の用語の定義を行っている。また、ガバメント・オンラインのメタデータ標準が国家予算委員会によって発行されている (詳細は 4-3-7 諸外国のメタデータ事例参照)。

(3) オーストラリア

NAA は、政府機関における記録管理のためのメタデータ標準 (「Recordkeeping Metadata Standard for Commonwealth Agencies」) 及び記録の分類・標題・インデックス等の一貫性を確保するための用語集 (「Keyword AAA」) 等が発行している。

なお、インターネット上の様々な情報を容易に検索できるようにするため、AGLS メタデータ標準 (The Australian Government Locator Service (AGLS) metadata standard) は、NAA によって開発された (詳細は 4-3-7 諸外国のメタデータ事例参照)。

(4) 日本

ネットワーク上のデータベースとして整備するものとされている「行政文書ファイル管理簿」について、文書分類・行政文書ファイル名・作成者・作成時期・媒体の種類等の記載すべき事項が「行政文書の管理方策に関するガイドライン (平成 12 年各省庁事務連絡会議申合せ)」 (以下「ガイドライン」という) で定められている。また、「電子公文書の文書型定義

の統一的な仕様」が平成 10 年 3 月行政情報システム各省庁連絡会議幹事会了承として出され、平成 14 年 2 月には行政情報化推進各省庁連絡会議幹事会了承として改定されている。

なお、上記のガイドライン等の策定等について、国立公文書館の関与はない。

3-3-4 電子公文書等の長期保存へ向けた取り組み

(1) アメリカ

電子政府イニシアティブの一つである「電子記録管理」において、NARA は、国防総省電子記録管理アプリケーション設計基準 (DoD 5015.2-STD) の全政府機関での採用を承認したほか、NARA への電子記録移管のための指針を発行している。

一方、NARA は、電子記録を永久に保存し、オンラインで利用に供することを目的とする ERA プロジェクトにも着手している。ERA は、2007 年試験運用開始、2011 年本格運用開始をめざしている。

(2) カナダ

LAC は、記録情報ライフサイクルマネジメントに関する全般的ガイドラインのほか、連邦政府内部における電子情報の相互運用性向上を目的として、文書・画像・映像・データベース等個別の記録様式ごとに電子情報の推奨規格等を定めたガイドライン(「Guidelines for Computer File Types, Interchange Formats and Information Standards」)等を発行している。

(3) オーストラリア

NAA は、政府機関に対し、電子記録の作成・管理・保存・利用・最終処分等に関する包括的ガイドライン(Digital Recordkeeping Guideline)草案を発表している(2004 年 5 月)。

電子記録の移管については、NAA は、オリジナルフォーマット及び XML へ変換したものの両方を受け入れている。

(4) 日本

「別の記録媒体への変更、データ・ファイル形式の変更、定期的なバック・アップ等の措置を講ずること」が「ガイドライン」に定められている。

3-3-5 電子メール管理・保存に向けた取り組み

(1) アメリカ

電子記録の作成・管理等に係る要件を規定する連邦規則 36CFR1234 は、電子メールの作成・管理等に係る政府機関の責任及び管理要件等を規定している。

また、NARA は、電子政府イニシアティブの一つである「電子記録管理」において、電子メールの NARA への移管に係る基準を発行している。

(2) カナダ

LAC は、電子メールを国立図書館公文書館法上の Record であると位置づけ、電子メール管理のためのガイドライン(「E-mail Management in the Government of Canada」)を発行している。

(3) オーストラリア

NAA は、電子メールも公文書法上の連邦記録に該当するとして、電子メールの管理・保存等に関するアドバイスを発表している(「Email is a record!」(1998年作成、2004年改訂))。

(4) 日本

特に取組みは見られない。

3-3-6 政府機関のウェブ・アーカイピングの現状

(1) アメリカ

合衆国法典の「記録」の定義に合致するウェブ・コンテンツ記録のうち、永久保存価値があると評価されたものは、他の記録と同様に、政府機関から NARA へ移管されることとされており、NARA は、2004年にウェブ・コンテンツ記録の移管に係る指針を発行している。また、NARA は、2005年1月時点における政府機関のウェブ情報の収集を行い、ウェブ上で利用に供している。

一方、議会図書館は、国民の関心の高いテーマ関連のページを少数選択してアーカイブしている(MINERVA)。

(2) カナダ

LAC は、官民のオンライン出版物の Web アーカイブを「Electronic Collection」として提供している。現在の提供タイトル数は、1万3千件を超える。

(3) オーストラリア

政府機関のウェブ・アーカイピングについて、NAA は、国立図書館及び NOIE (National Office for the Information Economy (2004年4月から AGIMO に組織改編))とともに、役割分担の取り決めを行い、共同で政府機関向けにガイド(「Keeping Government Publications Online」)を発行している(2002年7月)。

同取り決めでは、NAA は、イントラネット上の文書、ウェブ上での業務の記録、電子メール及び公開されたウェブのスナップショットを保存するものとされており、国立図書館は、リムーバブル・メディアで供給される電子出版物及びパブリック・コメントを求めるためにウェブに掲載された草案を保存するものとされている。また、インターネットのみで利用可能なオンライン出版物は、NAA 又は国立図書館が保存することとなっている。

(4) 日本

国立国会図書館が実験事業として「インターネット資源選択的蓄積事業(WARP:Web ARchiving Project)」を推進している。

4. 電子公文書等の適切な管理・保存のための技術動向と課題

4-1 電子公文書等の長期保存にまつわるリスク等

4-1-1 電子公文書等を保存し続けることによるリスク

電子公文書等に対して、作成された時点の状態のまま適切な対応を行わず保存し続けていた場合、将来的に電子公文書等を閲覧できなくなる、あるいは保存した情報の安全性が損なわれるといった事態が生じる可能性がある。

長期保存の目的を達成するためには、電子記録は真正性、完全性、アクセス可能性、理解可能性、処理可能性、潜在的な再利用可能性を備えることについて確認するとともに、外部に説明できなければならない(ICA 報告書 16)。

特に、その可読性が失われることは、公文書等の記録としての存在意義を無くしてしまうことに等しく、恒久的に閲覧できることが必要である。つまり、恒久的な文書の可読性を確保するような対応が求められる。

4-1-2 媒体・ハードウェアに起因する可読性損失リスク

作成された電子公文書等を記録する電子媒体を、作成時の状態で、そのまま保存し続けた場合、媒体の寿命や再生環境の損失により、電子公文書等の可読性を失うリスクを抱えることとなる。また再生向けハードの陳腐化によるリスクも抱えている。

(1) 媒体の寿命

磁気テープ、磁気ディスク、光ディスク等の電子媒体の寿命は 20 年から 30 年にとどまっている。そのため、電子公文書等を記録している電子媒体が寿命に達した場合、記録情報の損失を起こす可能性がある。

(2) 再生装置の陳腐化

電子媒体(磁気テープ、磁気ディスク、光ディスク)については、再生装置を利用することが前提となるため、再生装置の陳腐化により、再生環境が失われる可能性がある。

4-1-3 ソフトウェアに起因する可読性損失リスク

電子公文書等を閲覧及び編集するためのソフトウェア自体が入手出来なくなること、当該ソフトウェアを入手できたとしても、そのソフトウェアを動かすための OS 環境が無くなっているといった事態が、長期間経過後には起こることが想定される。

(1) ソフトウェアの入手困難

一般的に、電子文書は関連付けされたアプリケーションを利用することによって閲覧や編集ができる。しかし、当該アプリケーションが存在しなければ、そのファイルを開くこともできない。10年以上保存する電子公文書等の場合、それが作成された当時のソフトウェアを継続的に入手できる可能性は低い。特に、特定ベンダ(ソフトウェア製品を販売する会社)の仕様に基づいて作成された電子公文書等の場合、製造・販売当時のソフトウェアが長期間経過後販売・サポート終了になっている可能性も高い。

(2) ソフトウェアを動かすための OS 環境の損失

一般的に、ソフトウェアは特定の OS(種類・バージョン)上で動作する。そのため、現在の OS 上で 10 年以上前に存在したソフトウェアが稼働できる可能性は高くない。例えば、Windows95 に対応して作られたソフトウェアが、そのまま WindowsXP 上で稼働する保証は無く、そのソフトウェアが特定ベンダの製品である場合は、持続的な可読性の確保が難しい。

(3) ソフトウェア機能の陳腐化、非互換性

ソフトウェアのある特定の機能は、異なるベンダの製品間に加え、同一製品でさえもバージョンアップ(新製品販売)ごとに変化し、その結果古いバージョンの文書の特定部分を適切に処理できないといった場合がある。例えば、グラフィックスや注釈の埋め込み、データの基本構造の変更などのために、文書とメタデータの間の重要なリンクが損なわれる恐れがある。

4-1-4 改ざんの危険性

電子公文書等を長期間保存することで、外部からの進入等による文書に対する改ざんの危険性が高まる。文書が改ざんされることにより、公文書等の原本性を喪失させてしまうリスクを抱えることになる。

4-1-5 既存技術及びシステムの陳腐化

技術革新のスピードの速さから、旧来のシステムあるいはファイル形式で作成されたものが新たな環境下では機能しないというリスクがある。国立公文書館による電子公文書等の保存については、保存のタイムスパンが非常に長くなることから重大な課題となる。

これについては、ICA 報告書 16 でも、「システムには、記録管理機能が不備なものがあり、その結果、そのシステムの記録の質が、長期保存する意味をほとんど失うほど落ちてしまうこともある。その他の既存システムでも、保存不可能なフォーマットで記録を作成したり、記録を長期保存用のフォーマットへ変換できなかつたり、変換できてもコストがかかりすぎたりするものがある」と指摘している。

本章でもファイル形式について説明するが、特定のソフトウェアやハードウェアについては、単一のもを推薦することについては、現時点では、長期的な保存の安定性に鑑みて、大きなリスクがある。

4-2 技術動向及び課題に関する調査項目

4-2-1 調査項目の位置づけ

まず、メタデータ体系について調査・整理した。次に、移管(保存期間満了前に一定の集中管理の下に移す「移送」を含む。この節において、以下同じ。)前に関わる技術、移管時の技術、移管後の技術(保管)、移管後の技術(利用)にフェーズを分けて、それらのフェーズごとに対応する技術動向を調査した。

(1) メタデータ

長期にわたり電子公文書等を維持管理していくためには、適切なメタデータを付与して管理することが必要である。本調査では、メタデータのデータ項目に関する調査を行い、情報共有目的で国際的に幅広く用いられているダブリン・コアのデータ体系と比較するとともに、将来課題について整理した。

(2) 移管前の技術

電子公文書等は、文書ファイルに加え、音声ファイルや画像ファイルなども今後活用が広まっていくことが考えられる。しかし、これら全てのファイル形式に対して、移管後の再現性を確保することは困難である。そこで、ファイル形式の標準化に向けた課題の整理と海外事例について調査を行った。

(3) 移管時の技術

移管前にいくつかの推奨ファイル形式に絞り込まれた電子公文書等は、移送後、長期可読性を確保するために標準化したファイル形式に変換しなければならない。そこで、ファイル形式変換技術における課題の整理と、オーストラリア国立公文書館が開発したファイル形式変換ソフトウェアである Xena についてテスト・評価を行った。

(4) 移管後の技術(保管)

国立公文書館へ移管された電子公文書等は、データの損失がないように十分な安全性を確保する必要がある。特に、近年はウィルス等のセキュリティ上の脅威が多く存在している。そこで、公文書等を保管する保存媒体、データの損失を防止するバックアップ機能、ウィルス等の脅威から守るためのセキュリティ対策、原本を改ざんの危険性から守る原本性確保機能について、現状の課題整理と活用方法について調査を行った。

(5) 移管後の技術(利用)

「記録保存の第一の理由を、記録の証拠としての価値を保存することと考える。記録保存の目的としてもう一つ考えられるのは、記録自体、また、記録が包含している情報を再利用することである」(ICA 報告書 16)とあるとおり、適切な記録から必要な情報を抽出、利用するためには、検索機能の整備は不可欠である。また、移管された公文書等が利用者にとって永続的に閲覧できるような技術環境が必要となる。

そこで、現状における検索機能や可読性確保技術の特徴を整理したうえで、機能の実現に向けた課題を考える必要がある。

(6) その他

電子公文書等を扱う場合、その保存や利用のために構築・運用するシステムで使用できない文字があった場合、電子公文書等が記録する情報を正確に表現することが出来なくなる可能性がある。そこで、外字に関する現状の課題を整理した。

4-3 メタデータ

長期にわたり電子公文書を利用可能とするには、適切なメタデータ体系を整備し、適用する必要がある。ここでは、メタデータを、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つに分類し、データ項目等を調査した。また、国際標準であるダブリン・コアのメタデータ体系と比較を行った。

4-3-1 メタデータの分類

ICA 報告書 16 によると、電子公文書等の長期保存との関係ではメタデータは以下の3つに分類することが可能である。なお、「事業のコンテキスト情報を示すメタデータ」と「記録を捕捉・保存した後の記録管理を表すメタデータ」との二分法も可能である。

(1) 記録管理(recordkeeping)メタデータ

記録そのものの中から、または記録を作成した組織の中で生じたもの。作成者、作成日、タイトル、機密度、キーワードなどの要素を含んでいることもある。

(2) アーカイバル・メタデータ

記録が最初に作成された後にその記録を管理しやすくするように追加したもの(記録の再確認日、作成元の組織名等)。非現用記録の管理メカニズムの一部として記録を作成した元の組織が行ったり、国立公文書館のような作成された記録を最終的に受入れる組織が行ったりする。

(3) 技術的メタデータ

記録の理解や処理に必要なものである(ファイル形式、バージョン、作成 OS 等)。作成システムから生成されるものもある。技術的メタデータは、長期保存処理の過程の一部として追加される。

4-3-2 記録管理(recordkeeping)メタデータ

記録管理メタデータは、記録そのものの中から、または記録を作成した組織の中で生じており、作成者、作成日、タイトル、機密度、キーワードなどの要素を含んだメタデータである。

現時点で、日本の各府省が管理している文書の記録管理メタデータにあたるものが、行政機関情報公開法に基づいて作成が義務付けられている行政文書ファイル管理簿上のデータ項目である。

「行政文書の管理方策に関するガイドラインについて」においては、行政文書ファイル管理簿は次の事項を記載するものとされている。行政文書の分類の基準は、大分類・中分類・小分類の3段階のツリー構造となっており、行政文書ファイルの適切な保存にも活用できるよう「行政文書分類基準表」を定めるものとされている。

- 文書分類(大分類・中分類・小分類)
- 行政文書ファイル名
- 作成者
- 作成(取得)時期、保存期間、保存期間満了時期
- 媒体の種別
- 保存場所
- 管理担当課・係
- 保存期間満了後の措置結果
- 備考

ただし、各府省では「行政文書の管理方策に関するガイドラインについて」に基づいて行政文書ファイル管理簿を管理しているものの、実際の記載内容を確認したところ、各府省で分類名の記入レベルが異なっている事例が見受けられた。つまり、メタデータが定義されていたとしても、メタデータの中身のレベルにバラつきがある場合は、メタデータとしての役割を果たすことができない懸念があり、このような懸念が生じないようにしていくことが今後の課題として挙げられるであろう。

4-3-3 アーカイバル・メタデータ

アーカイバル・メタデータは、記録が最初に作成された後に、その記録を管理しやすくするように追加したものである。項目としては、例えば、記録の再確認日、作成元の組織名等が挙げられる。非現用記録の管理メカニズムの一部として記録を作成した元の組織が管理する場合や、国立公文書館のような作成された記録を最終的に受入れる組織が管理する場合等が想定される。

アーカイバル・メタデータの例としては、アーカイビング向けの文書記録の目録記述標準の一つであり、アジア歴史資料センターの検索システムで活用されている ISAD(G) や、その応用規格である EAD が挙げられる。今後、このようなアーカイビングに適した体系を、政府による電子文書作成(及びそのためのシステム)に取り込むことが可能かどうかについては、引き続き検討していく必要がある。

(以下の解説及び調査結果は、五島敏芳氏(国文学研究資料館史料館)「国際標準:記録史料記述の一般原則;ISAD(G)」とXMLの利用」を引用した(一部表現改)。)

(1) ISAD(G)

特に各種史料のアーカイビングを目的としている目録記述標準の一つである。5年ごとに見直しが行なわれており、(調査時点で)2000年に第2版が公表されている。

以下の表のとおり、7つの記述エリアと26の記述要素で構成されており、これを用いることで、記録史料に関する情報の検索・交換が容易になる、データの共有が可能になる、様々な機関によるメタデータ記述を標準化させることができるとしている。

エリア	要素
3.1 固有性宣言の領域; 個別情報のエリア	3.1.1 参照記号; レファレンスコード 3.1.2 標題 3.1.3 年代 3.1.4 記述レベル
3.2 脈絡情報の領域	3.2.1 作成者名 3.2.2 管理の/伝記的な歴史 3.2.3 記録史料の歴史 3.2.4 取得あるいは譲渡の直接の源泉
3.3 内容と構造の領域	3.3.1 範囲と内容 3.3.2 追加・廃棄と日程計画の情報 3.3.3 追加受入 3.3.4 整理の体系
3.4 利用可能性(/公開)と利用の状況の領域	3.4.1 利用可能性(/公開)を規定(/規制)する条件 3.4.2 複製を規定(/統制)する条件 3.4.3 史料実体の言語/書体 3.4.4 物的特徴と技術的要件 3.4.5 検索手段
3.5 類縁史料の領域	3.5.1 原本の存在と所在 3.5.2 複製の存在と所在 3.5.3 関係する記述単位 3.5.4 出版についての注記
3.6 備考の領域	3.6.1 備考
3.7 記述制御の領域	3.7.1 記録史料管理者の注記(/アーキビストノート) 3.7.2 規則または取り決め 3.7.3 記述の日付

図表 4-1 ISAD(G)第2版の記述要素

では、このISAD(G)は、我が国における電子公文書等の移管、保存等に活用できるだろうか。現在の行政文書ファイル管理簿の大分類、中分類、小分類という分類体系は、ISAD(G)でも同じ(fonds-series-item)であり、個別メタデータの精度を除けば体系的には類似する部分もある。

一方、現状の行政文書ファイル管理簿より、ISAD(G)は、データ項目数が非常に多く、かつ、日常的に行政文書を作成する各府省職員にとって入力容易でないと思われる項目もある。例えば、記述レベル、管理の歴史、記録史料の歴史等については、事前に十分なガイドライン等を作成することが必要と考えられる。

(2) EAD

EADとは、ISAD(G)の記述要素を実現するためのSGML/XMLの応用規格といえるもので、アーカイビング用途で利用される。EADを用いることによって、ISAD上の各項目の詳細までをタグで定義することが可能である(表組みや字体等)。

タグの作成については、Archives Hub (<http://www.archiveshub.ac.uk>)上にて、ホームページ上のフォームにISAD(G)上の記述要素に沿った入力をしてボタンを押すだけでEAD化を行なうツールが提供されているとのことであり、将来的に利用するとした場合にこのツール上で文書の作成者が必要項目を入力すればメタデータ標準に沿ったメタデータを付加できることとなる。

なお、五島氏の報告によると、EAD による記述は ISAD(G)の記述を実現しているものの、一部記述単位のずれが生じているとのことであるが、このような簡易な手法が既に提供されていることは ISAD(G)の普及に資するのではないかとしている。

4-3-4 技術的メタデータ

電子公文書等では様々なデジタル媒体や形式で保存されたものが移管される。それゆえ、電子公文書等の可読性を恒久的に確保するためには、電子公文書等の作成環境や閲覧環境を把握する必要がある。そのため、電子公文書等の移管・保存においては、電子公文書等の作成環境や再現環境を把握するために、記述情報だけでなく、ファイル形式、バージョン、作成 OS 等といった技術情報をメタデータとして付与する必要がある。

例えば、比較的汎用性が高いとされている PDF ファイルにも、バージョンが複数存在している。国立公文書館に移管される電子公文書等が PDF ファイルで作成・保存等されている場合、これらのバージョン情報が必ずしも PDF ファイルのプロパティ情報に含まれているとは限らない。また、映像ファイルや音声ファイルの場合、多数のファイル形式があり、それぞれにバージョンが存在している。こうした情報は、電子ファイルの拡張子だけでは必ずしも判断できない。そのため、技術的メタデータとして、作成環境や閲覧環境を把握し、管理する必要が生じるわけである。

現状では、ダブリン・コアにおいては、技術情報に関するメタデータが整備されているが、日本の行政文書ファイル管理簿上のメタデータには、情報公開請求への対応を第一としていることもあり、技術情報は含まれていない。

(1) ダブリン・コアで定義される技術情報

ダブリン・コアの 15 の基本要素のうち「Format」において、以下の事項を文書に記載することとされている。(http://dublincore.org/documents/dcmi-terms/)

- 電子情報の物理的形式(メディアタイプやデータサイズ)
- 電子情報の表示・操作に必要なソフトウェア
- 電子情報の表示に必要なハードウェア

4-3-5 ダブリン・コア

ダブリン・コアは、メタデータ標準の 1 つとして情報検索のために開発されたもので、インターネット上で国際的に幅広く利用されている。

ダブリン・コアにおいては、以下のように、DCMES(ディーシーエムイーエス; Dublin Core Metadata Element Set)と呼ばれる 15 の基本要素を定義している。

要素タイプ名	要素タイプ名 (英語)	定義とコメント
タイトル	Title	情報資源に与えられた名前。
著者あるいは作者	Creator	情報資源の創造に主たる責任を持つ人や組織。
主題	Subject	情報資源の主題あるいは内容を説明するキーワード。
内容記述	Description	情報資源の内容に関する説明や抄録。
公開者	Publisher	情報資源を現在の形で利用可能にした組織。出版者など。
寄与者	Contributor	「著者あるいは作者」以外で、情報資源の創造に寄与した人または組織。
日付	Date	情報資源が作成された、または有効になった日付。
資源タイプ	Type	テキスト、イメージなど、情報資源の種類。
フォーマット	Format	情報資源のデータ形式。
資源識別子	Identifier	URI、ISBNなどの、当該情報資源を一意に識別するための文字列または番号。
情報源	Source	情報資源を作り出す元になった別の情報資源に関する情報。出処。
言語	Language	情報資源を記述するために用いられている言語。
関係	Relation	他の情報資源との関係。
時間的・空間的範囲	Coverage	情報資源の地理的または時間的特性。
権利関係	Rights	権利や利用条件に関する記述へのリンク。

図表 4-2 ダブリン・コアの 15 の基本要素 (DCMES)
(<http://www.asahi-net.or.jp/~ax2s-kmtn/internet/search.html>)

4-3-6 メタデータ体系の比較

ダブリン・コア、行政文書ファイル管理簿のメタデータ、ISAD(G)の各項目が、どの程度関連しているかについて整理をおこなった。

ダブリン・コア	行政文書ファイル管理簿のメタデータ	ISAD(G)
	文書分類(大分類・中分類・小分類)	整理の体系
資源識別子		参照記号;レファレンスコード
タイトル	行政文書ファイル名	標題
著者あるいは作者	作成者	作成者名
主題		年代 記述レベル
内容記述		範囲と内容
公開者	管理担当課、係	記録史料管理者の注記(/アーキビストノート)
寄与者		
日付	作成(取得)時期	記述の日付
資源タイプ	媒体の種別	
フォーマット		物的特徴と技術的要件
言語		史料実体の言語/書体
情報源		管理の/伝記的な歴史 記録史料の歴史 取得あるいは譲渡の直接の源泉
関係		原本の存在と所在 複製の存在と所在 関係する記述単位
時間的・空間的範囲	保存期間 保存期間満了時期 保存期間満了後の措置結果 保存場所	追加・廃棄と日程計画の情報 追加受入
権利関係		利用可能性(/公開)を規定(/規制)する条件 複製を規定(/統制)する条件 出版についての注記 規則または取り決め
		検索手段
		備考

図表 4-3 ダブリン・コア等と行政文書ファイル管理簿のメタデータとの比較

結果として、ダブリン・コアで提示されている基本項目が、その他のメタデータ体系では、より細かに提示されている様子である。

ちなみに、国立公文書館が平成 17 年 4 月から提供しているデジタルアーカイブ・システムでは所蔵公文書等の目録情報の管理のためにダブリン・コアベースのメタデータを既に付与している。また、アジア歴史資料センターの検索システムには、コンテンツを検索できるようにするための基礎となるメタデータが付与 (ISAD(G)と Dublin Core を合わせた目録システム) されている。

ただし、これらのメタデータは政府等から電子的に提供されたものではなく、国立公文書館側で所蔵公文書等から独自に設定・入力したものである点で、電子公文書等の移管を前提とした (政府主体での) メタデータ付与とは異なるものである。

4-3-7 諸外国のメタデータ事例

諸外国では、既にメタデータの指針が示されているケースが見受けられる。アメリカ、オーストラリア、カナダの 3 国を例に示す。

(1) アメリカ

「DEPARTMENT OF DEFENSE DISCOVERY METADATA SPECIFICATION (DDMS)

」(VERSION1.3 2005 年 Deputy Assistant Secretary of Defense)により、メタデータの指針が示されている。

また、NARA は、電子政府イニシアティブの一つである「電子記録管理」において、国防総省電子記録管理アプリケーション設計基準 (DoD 5015.2-STD) の全政府機関での採用を承認している。

Core Layer Category Set	Primary Category	Obligation
The <u>Security</u> elements enable the description of security classification and related fields	Security	Mandatory
The <u>Resource</u> elements enable the description of maintenance and administration information	Title	Mandatory
	Identifier	Mandatory
	Creator	Mandatory
	Publisher	Optional
	Contributor	Optional
	Date	Optional
	Rights	Optional
	Language	Optional
	Type	Optional
	Source	Optional
The <u>Summary Content</u> elements enable the description of concepts and topics	Subject	Mandatory
	Geospatial Coverage	Mandatory unless not Applicable
	Temporal Coverage	Mandatory unless not Applicable
	Virtual Coverage	Optional
The <u>Format</u> elements enable the description of physical attributes of the asset	Description	Optional
	Format	Optional

図表 4-4 DDMS Primary Category Sets

(http://diides.ncr.disa.mil/mdreg/uploads/DDMS-v1_3.doc?CFID=606963&CFTOKEN=48085387)

(2) オーストラリア

「The Australian Government Locator Service(AGLS) metadata standard」(「AGLS Metadata Element Set」)(2002年 National Archives of Australia)により、メタデータの指針が示されている。

ウェブ等のインターネット上の情報が主な対象とされているが、オフラインの情報への適用も可能とされている。

Elements	Element refinements
Creator	
Publisher	
Contributor	<i>Role of contributor should appear as an element refinement if required</i>
Availability	
Title	alternative
Subject	
Date	created modified valid issued
Identifier	
Rights	
Description	
Source	
Language	
Relation	isPartOf/hasPart isVersionOf/hasVersion isFormatOf/hasFormat references/isReferencedBy isBasedOn/isBasisFor isRequiredBy/requires isreplacedBy/replaces
Coverage	temporal spatial jurisdiction postcode
Function	
Type	category aggregationLevel documentType serviceType
Format	extent medium
Audience	
Mandate	act regulation case

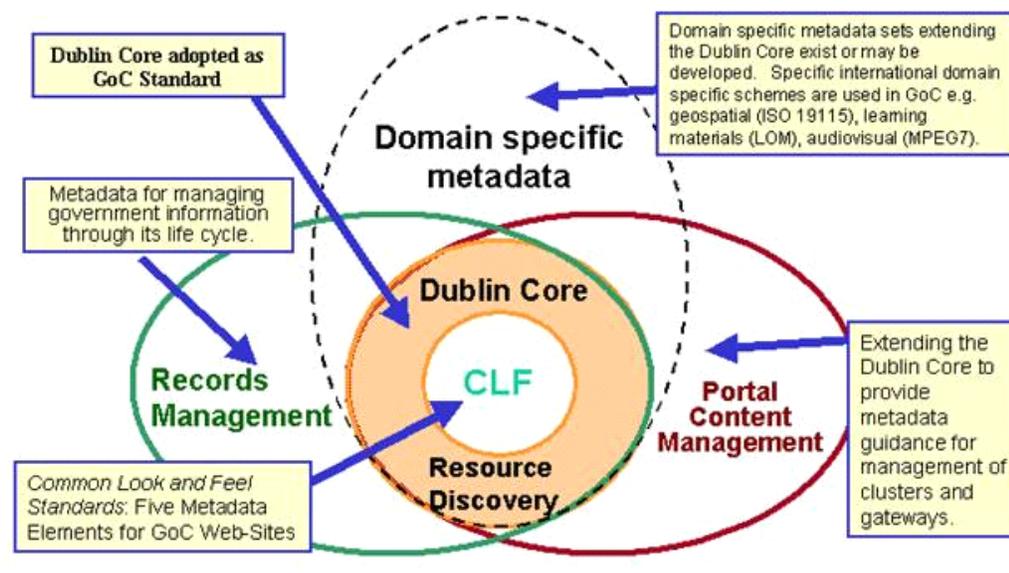
図表 4-5 AGLS Qualifier Summary

(http://www.naa.gov.au/recordkeeping/gov_online/agls/AGLS_reference_description.pdf)

(3) カナダ

GoCにおけるメタデータの開発戦略を示す指針として、「The Government of Canada (GoC) Metadata Framework」(Information Management Resource Centre (IMRC))が示されている。

5つの基本的な必須項目(Common Look and Feel metadata elements)として、creator, date, language, subject, titleが挙げられており、これとGoCの基本フレームとして据えられたDublin Coreを中心として、拡張を図っていく指針が提示されている。



図表 4-6 The Government of Canada Metadata Framework
(http://www.tbs-sct.gc.ca/im-gi/meta/frame-cadre_e.asp)

また、「Government of Canada Metadata Implementation Guide on Web Resources」により、Web情報のメタデータに関するガイドが示されている。

Category	Elements
Common Look and Feel Required Elements	Creator
	Date
	Language
	Subject
	Title
Optional Elements	Audience
	Coverage
	Description
	Format
	Type

図表 4-7 GoC Metadata Elements on Web Resources
(<http://www.collectionscanada.ca/obj/s37/f2/s37-4000-e.pdf>)

4-3-8 メタデータに関する課題

ICA 報告書 16 にて、「記録及び記録に関するメタデータの、作成及び取り込みに関する規則を、活動の証拠を記録する必要がある全業務処理を律する手順に組み込むことが望ましい」とされているとおり、メタデータについてはその構成、論理体系とともに付与に関するルールの明確化も重要な課題である。

本来であれば、構想段階から適切なメタデータが自動的に付与されるようなシステム設計が望ましいが、政府の電子公文書等の移管、保存の際は、メタデータの自動付与が困難と思われるもの(文書の内容に起因するもの(文書名や分類等))が残らざるを得ず、そのようなものについては人による作業により文書の内容から類推したうえで付与することとなる。

そのため、全ての電子公文書等に対して、前述のメタデータを付与しようとした場合、政府の多くの職員の一人一人の作業に依存するゆえに生じる、統一性やレベルの確保といった点の大きなリスクを抱えている。メタデータが定義されていたとしても、個々の作成者が独自にメタデータを付与しては、メタデータ情報の統一性が損なわれてしまう。例えば、「所属」というメタデータがあった場合、同じ課室に所属していたとしても、ある人は「総務部総務課」まで記入し、ある人は「総務部」までしか記入しないことが考えられる。こうした場合、総務部というキーワードで文書を分類・抽出した際、同じ所属として表示されないことになる。すなわち、いかにメタデータの内容を統一化、管理できるかが重要となる。

(1) 新規に作成される電子公文書等

新規に作成される電子公文書等を移管する際、各府省側のメタデータの付与を手作業で行なう場合には、現状の文書管理業務よりも作業負荷が大きくなることは否めない。

ただし、技術メタデータについては、例えば今後の文書管理システム上の機能要件として位置づけるとともに、各府省庁内で利用されるパソコンのソフトウェア環境を統一することによって付与に係る負担を軽減させることが可能である。

(2) 既存の電子公文書等

既存の電子公文書等については、既に作成環境が異なることから、一定のメタデータ体系との整合性を図ることは、新規文書の場合よりも、困難といえる。一定のメタデータ体系に完全に統合させようとする場合には、例えばマイグレーションを行なう際にファイル形式を標準化するといったことが考えられる。ただしこの方法の場合、マイグレーションの際に(例えば罫線のようなものについて)完全性が損なわれるリスクが存在する。

4-4 移管前の技術

電子公文書等は、文字データで構成されるいわゆる文書ファイルに加え、音声ファイル、画像ファイル、映像ファイルなども今後作成・活用が広まっていくと想定される。ただし、ファイル形式が多岐に渡ることから、これら全てのファイル形式に対して、再現性を確保することは物理上困難である。

一方、電子公文書等の作成段階から、全てのファイル形式を標準化することは難しい。したがって、電子公文書等の移管前の段階で、利用するファイルの形式を、他国の取り組み事例も参考に、いくつかの推奨ファイル形式に絞り込むことが考えられる。

4-4-1 ファイル形式の標準化

ファイル形式の標準化の最も望ましい姿としては、国際標準もしくは国内標準に規定されているフォーマットを使用することである。次いで望ましいのは、一般的に利用できる任意の標準で規定された、特許権やライセンス供与の制限を受けないフォーマットである。もっとも、標準フォーマットは、通常多くのソフトウェアメーカーが、保存時のファイル形式を変換したり、あるいは変換したりするといった機能によってサポートするようになってきている。

(1) ファイル形式について

ファイル形式とは、コンピュータ上で利用するファイルの保存形式を指し、コンピュータで扱う文書、画像、動画などのさまざまなファイルを特定の利用方法やソフトウェアで共通に扱うための形式や規格のことをいう。

ファイル形式を汎用性に注目して、以下の3つに分類することができる。

(<http://ja.wikipedia.org/wiki/ファイルフォーマット>)

ア. 標準化されたフォーマット

国際的な機関で制定されたフォーマット。ファイルフォーマットの例としては、JPEG、MPEG などがある。標準化機関の例としては、ISO (国際標準化機構)、JIS(日本工業規格)、W3C(World Wide ウェブ Consortium)などがある。

イ. 事実上、業界標準となっているフォーマット

下の特定のオペレーティングシステム(OS)やアプリケーションに依存するフォーマットに含まれるが、広く使用され事実上標準(デファクトスタンダード)となっているもの。例としては、GIF などがある。

ウ. 特定のオペレーティングシステム(OS)やアプリケーションに依存するフォーマット

各種 OS やアプリケーションのデータファイル。例としては、BMP、RTF(ともにWindows)、PDF(米 Adobe 社)など多数ある。

(2) 推奨するファイル形式の条件

ICA 報告書 16 ではフォーマットが満たすべき条件として以下をリストアップしている。

- 全ての情報のまたは元の記録の中で有意義とみなされる情報相互の関係を表現できる
- 国際標準、国内標準、または広く一般に利用できる標準で規定されている
- 長期にわたる利用可能性や、幅広い適用が証明されている
- アクセスの目的に直接利用可能であるか、利用可能なフォーマットに変換可能である
- 特定のソフトウェア環境、ハードウェア環境に依存しない
- 元のフォーマットから保存用のフォーマットに自動的に変換できる。必要に応じて、変換の問題やエラーを自動で検出・報告する機能がある
- (必要に応じて)保存用のフォーマットから元の記録作成システムや現在の記録作成システムで使用しているフォーマットへ自動的に変換できる

(3) 海外事例 アメリカ NARA

NARA の「Transfer of permanent E-records to NARA」において、政府機関から NARA に対して、電子公文書を送付する際の媒体やファイルフォーマットに関する規定が示されている。

- メールメッセージと添付ファイルを移管する場合のフォーマット
- 紙の文書をスキャンし、イメージ文書として移管する際のフォーマット
- PDF 形式で移管する場合のフォーマット
- NIST(National Institute for Standards and Technology)に登録されているアーカイブ用のメタデータ
- デジタル写真の記録媒体を移管する際のフォーマット
- 地理情報システム(GIS)用のデジタルデータを移行する際のフォーマット
- ウェブコンテンツの記録を移行する際のフォーマット

(4) 海外事例 カナダ

Guidelines for Computer File Types, Interchange Formats and Information Technology において、様々な形態のファイル形式について、推奨と容認を示しており、ベンダ依存のリスクがあるものは、仮に普及が進んでいても容認レベルに止めている。

- 文字コード (推薦)ASCII、UNICODE、EBCDIC
- オーディオ (推薦)WAV、AIFF (容認)MPEG-1、MIDI、RealAudio
- 画像 (推薦)ITU-T、PNG、TIFF (容認)Gif、JPEG、JFIF
- 動画 (推薦)MPEG-2 (容認)AVI、MPEG-4、MOV、RM
- ドキュメント (推薦)XML、HTML、XHTML、SGML (容認).txt、.doc、PDF、.wpd
- 電子メール (推薦)MIME

4-4-2 テキストファイル

テキストファイルとは文字データだけで構成されたファイルで、どんな機種のコピュータでも共通して利用できる数少ないファイル形式の一つである。ただし、テキストファイルは純粋に文字データのみで構成され、修飾情報や埋め込まれた画像など文字以外のデータは表現することができないため、マイクロソフトの Word などで作成された文書の情報をどのような形で保持するかが課題となっている。

(1) カナダの取組み

カナダ LAC では、「Guidelines for Computer File Types, Interchange Formats and Information Standards」のなかで、テキストデータの扱いについていくつかのファイル形式に対する見解をまとめている。

できる限り標準化されているファイル形式で作成することを推奨するとともに、移管後の利便性まで視野に入れたものとなっている

ア. 推奨する (Recommended) 作成・保存・移管方法

- XML: 柔軟性が高く、「内容」を記述するのに適切である。従って、検索する際にも有効なツールとして活躍できる。W3C という国際的組織がサポートしており、主要なベンダの製品も XML をサポートしている。
- XHTML: HTML の変形である。W3C の推薦も 2001 年に受けており、グラフィックスや電子商取引等でのタグとして利用されている。現在はバージョン 2.0 が出ているが、安定的かどうか不明なので、1.0 と 1.1 のみ推奨する。
- HTML: ウェブ上の文書を定義する際の標準的なフォーマットである。ハイパーテキストを定義するのに用いられ、異なるコンピュータ・プラットフォーム間でも互換性がある。バージョン 4.1 を推奨している。
- SGML: ISO8879 で国際標準として認められている。文章のコンテンツと構造を定義するために用いられる。名前とデータを紐付ける役割を果たすタグが使われており、SGML を用いているアプリケーションはプラットフォームに依存しない。

イ. 容認する (Accepted) 作成・保存・移管方法

- テキストファイル(.txt) : 文章の構造は定義されていないテキストだけのファイルである。ISO/IEC 8859-1 に基づいている
- マイクロソフトワード(.doc) : マイクロソフトのワードは世界で最も使われているワープロアプリケーションである。従って、国際標準ではないものの「デファクト」である。
- PDF : PDF はアドビによって作られた、ドキュメントを幅広く配布、閲覧させるためのデファクトなフォーマットである。PDF の利点としてはプラットフォームに依存しない点があるものの、特許や著作権といった面でのリスクがある。国際標準化しようという取組みがいろいろ進んでおり、PDF/A はその代表的なものであるが、LAC としては PDF/A が ISO 標準となるかどうか、注意深く見守るものとする。

(2) PDF

PDF は電子文書のテキストデータを埋め込むとともに、文書の見た目(画像データ)も保存できる。また、最近では各種の電子ファイルを自動で PDF に変換するソフトウェアも販売されており、PDF 変換の手間が大幅に削減されている。

一方、一部標準化の動きは進んでいるものの、ベンダに依存している面もある点が課題となっている。

ア. PDF の標準化動向 (PDF/A)

PDF の標準化は、PDF の相互運用性や将来の商品開発、信頼性の強化によって、公的な記録の保存や、電子的移管を進めるため、政府機関や業界団体、標準化組織、個々の企業によって進められている。具体的には、OASIS、ISO、W3C、AIIM、XBRL、FDA、NFF、ANSI、XML.Gov といった組織が影響力を持っている。

PDF/A は、モノクロとカラーが混在する文書を電子データとして長期保存するための仕様である。行政機関、規制産業、企業、教育機関、図書館から、ビジネス上、技術上の要求に基づいている。現在、AIIM(the Association for Information and Image Management) インターナショナル、NPES (the Association for Suppliers of Printing ,Publishing and Converting Technologies) によって策定作業が進展している。

PDF/E は、エンジニアリング文書を高い信頼性を維持した状態で交換したいというニーズに対応したものである。知的財産の安全な配布や、文書交換や変更管理における信頼性といったメリットのほか、紙文書を排することによるコスト効果もある。

(<http://www.itmedia.co.jp/enterprise/0405/16/epn03.html>)

イ. 海外事例 アメリカ NARA

アメリカ NARA の「Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records: Records in portable document format (PDF)」において、政府機関から NARA に対して、電子公文書を送付する際の PDF ファイルの要件を定めている。

- PDF ファイルのバージョン(バージョン 1.0 から 1.4)
- セキュリティ要件(セキュリティ上の設定をオフにする)
- 元の電子ファイルから PDF ファイルに変換する場合の要件(元の電子ファイルから PDF ファイルにフォントを埋め込んでおく)
- スキャンした文書から PDF に変換した場合の要件(PDF ファイルにテキストを埋め込むこと(原本を OCR 化する))
- 技術メタデータ情報の提出
- オペレーティングシステム(及びバージョン)、閲覧用のソフトウェア(及びバージョン)を明確にしておくこと

(3) テキストファイルの推奨ファイル形式について

カナダ LAC の取組みにもあるとおり、データとしての移行や利活用を前提とした XML のようなものが今後の中心となると思われる。一方、アプリケーションの形態をとる PDF や Word については、ベンダに依存する点をリスクとして認識しているわけであるが、現に政

府にはそれらで作成した電子ファイルが多数あると想定されることから、ISO 化の動向を踏まえて積極的に対応する可能性がある点は注目すべきである。

4-4-3 画像ファイル

一般的に利用されているファイル形式が限定されているものの、画像ファイルも様々な形式が存在している。今後は電子公文書等を作成する段階で画像ファイルの形式を絞り込んでいくことが望ましい。

(1) 画像ファイルの種類

画像形式を大きく分けると、画像を点の集まりとして記録する「ビットマップグラフィックス」と、画像を線や面などの図形の集まりとして記録する「ベクターグラフィックス」の2種類になる。おおむね、インターネット上で使用されているのは処理が簡単なビットマップ、編集集中の画像を記録する用途には劣化が起こらないベクター、という住み分けが行われている。

現在インターネット上でよく使用されている画像形式は、PNG、JPEG、GIFの3種類(すべてビットマップ方式)。グラフィックス表示ができるウェブブラウザは、これらの画像形式の表示に対応しているため、ブラウザの画面上に画像を表示することができる。

この他にも、ローカルマシン上では、BMP や TIFF などの画像形式が使われているほか、グラフィックスソフトはそれぞれ独自の画像形式を使用して独特の機能に対応するなど、用途に応じて非常に多くの画像形式が使い分けられている。

PNG	PNGはネットワークに最適な画像ファイルへの需要に応じて提唱されたもので、特にWWWでの利用に焦点を置き、GIFを発展させた形で定義されている。GIFと同様に可逆方式でデータを圧縮しますが、圧縮効率はGIFより高い。最高48bitのRGBカラーと16bitのグレースケール、アルファチャンネル(透明度を指定できる)をサポートしており、表現力の点でGIFよりも優れている。またネットワークでの利用を考慮して、GIFよりも美しくインターレース表示できるようになっている。
JPEG	静止画像データの圧縮方式の一つで非可逆である。1990年にISO10918-1(ITU-TS T.81)として規格化されている。圧縮の際に、若干の画質劣化を許容する(一部のデータを切り捨てる)方式と、まったく劣化のない方式を選ぶことができ、許容する場合はどの程度劣化させるかを指定することができる(圧縮率は約1/10～1/100)。スキャナで取り込んだ画像や、ビデオから取り込んだ画像などの自然画像をデジタル化した画像の圧縮保存に適しており、画像によっては、圧縮によって画像のサイズを非圧縮の状態の数分の1から数十分の1にできる。このため、WWWのインラインイメージなどのネットワークを通じた画像転送ではGIFとともに標準的に用いられている。今後長期保存の際に適用する場合、方式により圧縮率等にばらつきはあり完全に統一された規格でない点が課題となる。
JPEG2000	画像符号化の標準化を行なうISOとITU-TSの共同組織であるJPEGによって、2001年1月に規格化された。ウェーブレット変換(ウェーブレット関数により画像全体を周波数帯域に分けた縦横それぞれの周波数成分を量子化・符号化して圧縮する方式)で変換するため、JPEGでは高圧縮率(低画質)で保存したときに目立っていたノイズが、JPEG2000では発生しない。また、「電子透かし」の挿入、圧縮する際の画質・ファイルサイズ等の詳細設定や全く損失のない可逆圧縮の選択も可能である。
GIF	GIFはネットワークを通じた画像転送を目的として開発された画像形式であり、WWWのインラインイメージとしてJPEGとともに標準的に用いられる。256色までの画像を保存することができ、JPEGが苦手なイラストやアイコンなどの保存に向いている。動画を保存できるアニメーションGIFや、透明色を指定して背景イメージと重ね合わせることができるトランスペアレントGIF、全体をダウンロードしなくてもイメージの確認ができるインターレースGIFなどの拡張仕様がある。
TIFF	ビットマップ画像をさまざまなコンピュータ間で交換することを目的として開発された形式で、RGB、CMYK、グレースケールという3種類の色表現で保存できる。基本的には非圧縮のフォーマットで、画像をスキャナで取り込むときや、画像の編集、加工を行う際、特に複数のアプリケーションの間でやりとりする際に一般的に使われる形式である。
BMP	BMPはWindowsでの標準画像フォーマットであり、IBM PC互換機などで標準的に使われるファイル形式である。

図表 4-8 画像ファイル形式の特色
(<http://e-words.jp/w/E794BBE5838FE5BDA2E5BC8F.html>)

なお、TIFF ファイルフォーマットについては、ICA 報告書 16 でも言及されており、デジタル静止画像の理想的な作成・保存方法論として認識されてきたものの、非常に容量が大きいこと、また、簡単に処理できるソフトウェアを多くのユーザが持ち合わせていないことから、物理的性質から理想的なフォーマットとはみなされていないとのことである。

ア. JPEG2000

JPEG2000 は、JPEG 規格を策定する JPEG(Joint Photographic Experts Group)が、2001 年に規格化した画像圧縮ファイル形式である。従来の JPEG で生じていた輪郭部分のノイズなどが生じにくいいため、JPEG2000 を利用すると圧縮率を高めても品位の低下が抑えられる。国立公文書館のデジタルアーカイブ・システムのファイル形式としても活用されている。また、国立国会図書館近代デジタル・ライブラリーの画像提供形式としても採用された。

イ. PDF

画像ファイルを PDF 変換するも可能である。バージョン 1.5 以降の PDF (Acrobat Reader6 以降で表示可能な形式) では、ソフトウェアの中で画像を JPEG2000 に変換することができる。画像ファイルを PDF ファイルに変換する場合、一般的に「圧縮率」「解像度」という二つのパラメータを設定する。圧縮率を高くするとファイルのサイズは小さくなるが画質は落ちることとなり、解像度を高くすると、拡大しても画像がきれいに表示されるが、ファイルのサイズは大きくなるという性質がある。



図表 4-9 PDF の画像設定 (画面は Adobe Acrobat Standard7.0)

(2) 海外事例 アメリカ NARA

アメリカ NARA の「Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records: Scanned Images of Textual Records」において、政府機関から NARA に対して、電子公文書を送付する際のイメージファイルの形式を定めている。

ア. 画像ファイル形式

TIFF、GIF、BIIF、PNG を対象としている

イ. 画像の解像度

- モノクロ: 300 ~ 600ppi
- グレースケール: 300 ~ 400ppi
- カラー: 300 ~ 400ppi

ウ. 技術メタ情報の提出内容

- イメージシステム(ソフトウェアやストレージシステム)のバージョン
- オペレーティングシステムのバージョン
- アプリケーションのバージョン
- 画像の質(画素数など)
- 文書数
- 画像数
- メディア内のファイルディレクトリ
- 画面イメージのヘッダー

(3) 画像ファイルの推奨ファイル形式について

画像ファイルは動作環境がテキストファイルに比べて限定されているため、アメリカ NARA の取組みを参考に、利用する画像をいくつかに限定していくことが望ましい。また、最近ではウェブコンテンツに画像が組み込まれていることが多いため、今後のニーズや技術動向を踏まえ、当面は複数の推奨ファイル形式を提示することとなる。

4-4-4 音声ファイル

音声ファイルには様々な形式が存在しており、これら全ての形式に対して持続的に再現環境を確保し続けていくことは難しい。電子公文書等において音声ファイルの数は多くないと想定されるが、今後は電子公文書等を作成する段階で音声ファイルの形式を絞り込んでいくことが望ましい。

(1) 音声ファイルの種類

現在、音声ファイルには様々な形式があり、ファイルの圧縮形式も数多く存在する。このうち、ウェブで広く使用されている音声ファイル形式としては、主に UNIX 系で採用されている AU、Microsoft 社と IBM 社が共同開発した WAVE などがある。また、圧縮形式としては、フラウンホーファ IIS 社がアルゴリズムを開発した MP3 をはじめ、Microsoft 社の WMA などがある。

AU	「Audio」の略称で、UNIXで標準的なサウンドフォーマット。MacやWindowsでもサポートされている形式で、プログラミング言語のJavaでも標準でサポートされており、幅広く利用されている。
WAVE	Windows標準の音声ファイルの形式で、「WAVE形式」などとも呼ばれる。音声信号をデジタルデータに変換したものを記録するための保存形式などを規定している。圧縮(符号化)方式については規定しておらず、任意のものを利用することができる。
MP3 (MPEG Audio Layer-3)	映像データ圧縮方式として知られるMPEG-1で利用される音声圧縮方式。人間の感じ取りにくい部分のデータを間引くことによって高い圧縮率を得る非可逆圧縮方式を採用している。数字が大きい程圧縮率が高く、MP3の場合CDに近い音質を保ったまま、データ量を概ね1/11に圧縮可能(この他にもLayer-1、Layer-2の音声圧縮方式があり、圧縮率はそれぞれ約1/4、約1/7)である。現在最も普及している音声圧縮方式であるが、著作権保護のためのシステムを持っていないため、不正コピーの問題を抱えている。
MP3PRO	最新のMP3圧縮形式で品質をそのままに圧縮効率を約2倍に高めている。(MP3PROの64kbpsエンコーディングレートでもMP3の場合の128kbpsエンコーディングレートと同様のパフォーマンスが得られる。)
WMA (Windows Media Audio)	マイクロソフト社の音声圧縮フォーマット。圧縮率が高い場合の音質に優れており、比較的低速な通信回線を経由して音声配信できる。著作権保護のために暗号化を行うことが可能で、不正なコピーの防止や有料コンテンツとして配信ができる。人間の感じ取りにくい部分のデータを間引くことによって高い圧縮率を得る非可逆圧縮方式を採用している。

図表 4-10 音楽ファイル形式の特色

(2) 音声ファイルの推奨ファイル形式について

音声ファイルの形式は、過去のものを含めると、前述の 5 種類以上のものが存在している。また、今後の情報技術の発展により、現在一般的に広まっているファイル形式が数十年後に残っている可能性については議論のあるところである。

一方で、最近ではウェブコンテンツに音声や音楽を組み込んでいるものもいくつか存在しており、今後のニーズや技術動向を踏まえ、当面は、複数の推奨ファイル形式を提示することとなる。

4-4-5 映像ファイル

映像ファイルには様々な形式が存在しており、これら全ての形式に対して持続的に再現環境を確保し続けていくことは難しい。電子公文書等において映像ファイルの数は多くないと想定されるため、今後は電子公文書等を作成する段階で映像ファイルの形式を絞り込んでいくことが望ましい。

(1) 映像ファイルの種類

動画のデータ形式についても、さまざまなものがある。コンピュータ上で再生できる動画形式の代表的なものに、QuickTime 形式、Windows Media 形式、Real Player 形式、MPEG 形式がある。すべて圧縮形式である。これらの形式のなかでも、さらに細かい形式の違いによる分類があったりして、場合によっては、手持ちのアプリケーションで再生できないこともある。

(<http://hwb.ecc.u-tokyo.ac.jp/current/4857422FA5DEA5EBA5C1A5E1A5C7A5A3A5A22FC6B0B2E8A4CEA5C7A1BCA5BFB7C1BCB0.html>)

QuickTime形式 (拡張子:mov)	QuickTime形式は、Apple社のQuickTime Playerというプレーヤ(再生アプリケーション)で再生することができる。Windows用にもQuickTime Playerがある。QuickTime Playerは、ほかにもいろいろな形式のデータを再生することができる。 1998年2月には、国際標準化機構(ISO)によって、動画フォーマットの国際標準MPEG-4規格のファイルフォーマットとしてQuickTimeが採用された。
Windows Media形式(拡張子:wmv)	Windows Media形式は、名前のとおり、Microsoft社のWindows Media Playerで再生できる。Mac OS X用のWindows Media Playerもある。Windows Media Playerも、いろいろな形式のデータを再生することができ、iMac端末にもMac OS X用のWindows Media Playerがインストールされている。
Real Player形式	RealPlayer形式は、RealNetworks社のRealPlayerというアプリケーションで、再生する動画の形式で、基本的なプレーヤは、RealNetworks社のウェブサイトにおいて、無料で配布されている。ストリーミングとは、動画をダウンロードしながら再生すること。ダウンロードし終わらなくても再生できるという利点がある。RealPlayer形式はストリーミングに向けた形式である。
MPEG形式	MPEG形式は、ISOにより定められた動画形式で、MPEG-1やMPEG-2やMPEG-4などがある。MPEG-2は、DVD再生にも使われている圧縮形式である。MPEG-4形式(拡張子がmp4のもの)はQuickTime Playerやその他のプレーヤで見ることができるが、コーデック(動画や音声ファイルを変換するためのソフトウェア)の違いなどからMPEG-4を名乗っていても再生互換が必ずしも取れていない。

図表 4-11 映像ファイル特色

(2) 映像ファイルの推奨ファイル形式について

映像ファイルの形式は、過去のものを含めると、前述の 4 種類以上のものが存在している。また、今後の情報技術の発展により、現在一般的に広まっているファイル形式が数十年後に残っている可能性についても、音声ファイルと同様議論のあるところである。今後は、電子公文書等の作成段階で推奨する映像ファイル形式を絞り込むことが望ましい。

4-5 移管時の技術

移管前にいくつかの推奨ファイル形式に絞りこまれた電子公文書等は、移管時に、長期可読性を確保するために標準化した(定められた)ファイル形式に変換される。ファイル形式変換技術の海外における具体的な事例としては、様々なファイル形式から XML 形式のデータに変換するツールとしてオーストラリア国立公文書館(NAA)が開発した「Xena」がある。そこで、本調査では、「Xena」の評価を行った。

また、移管時の技術として、各府省庁において、あらゆる形式で管理しているメタデータデータベースから XML 形式でデータを出力する機能について調査した。

4-5-1 ファイル形式変換技術

ファイル形式変換技術とは、あるファイル形式で記録されているファイルを、別の形式に変換して記録するための技術である。

(1) 移管された電子公文書等のファイル形式を変換する

ファイル形式変換技術を利用して、移管された電子公文書を標準化された別の形式に変換するものである。

ICA 報告書 16 において、記録の保存組織にとって電子記録管理の経費を最も効率よく節減する方法は、記録の作成段階から記録システムを標準化することであるとしている。そのためには、少なくとも記録やそのメタデータを現代の標準(例えば、構造化情報のための XML)に準拠したフォーマットに変換しなければならないが、これらの作業は効率面でも、正確性という面でも自動的に行われることが望ましい。

(2) ファイル形式変換技術の課題

平成 16 年度に国立国会図書館が行った「電子情報の長期保存とアクセス手段確保のための調査報告書」におけるテスト結果によると、変換および再生確認ができた資料は 15 点中 2 点と少なかった。その理由として、HTML などの複数のファイルが相互に関連性を持ちながらコンテンツとして再生されなければならないような資料は、ファイル単位の変換ではなく電子資料単位で複数ファイルの関連性が維持されるように変換を行う必要があったためである。

このように、ファイル形式変換技術は、テキスト文書に対しては非常に有効だが、複数ファイルの関連性があるウェブコンテンツ等については利用するのが難しい技術といえよう。

4-5-2 海外事例(オーストラリア:XML 変換ソフト Xena)

ファイル形式変換技術の一つとして、NAA が開発した XML 変換ソフトウェアの Xena がある。Xena はハードウェアや OS に依存しない JAVA プログラムで作成されており、インターネット上のサイトから無償でダウンロードできる。本調査(2005 年 10 月実施)では Xena.2.0.0 を利用

し、現在多くの人々が利用している Microsoft Office 文書、HTML 文書、テキスト文書、画像、PDF ファイルを変換対象として評価を行った。

(1) Xena について

Xena は様々なファイル形式で作成された電子的記録を保存用のフォーマットに変換することを目的として NAA が開発したものである。

ア. Xena がサポートするファイル形式

Xena は、次のファイル形式を XML 形式に変換することができるとしている。

- Microsoft Word, Excel, Powerpoint
- OpenOffice.org Writer, Calc, and Impress
- RTF
- PST, TRIM, MBOX, IMAP (e-mail 用のファイル形式)
- CSV
- Relational Database
- JPG, GIF, TIFF, PNG BMP (画像ファイル形式)
- HTML
- Web Sites
- Plaintext

(<http://xena.sourceforge.net>)

イ. Xena の稼動環境

Xena を利用するにあたり、JRE (Java Runtime Environment)、OpenOffice を別途導入し、Java プログラムの差分情報 (パッチ) を適用する必要がある。

- JRE (Java Runtime Environment) : JRE とは、java 言語で開発されたソフトウェアを実行するために必要なソフトウェアのセットのことをいう。JRE は「<http://java.sun.com/j2se/1.4.2/download.html>」等から無償でダウンロードできる。
- Open Office の導入 : OpenOffice とは、オープンソースのオフィスソフトの一つで、Microsoft 社の「Microsoft Office」と高い互換性があり、Linux や Windows など様々な OS で動作する。「<http://ja.openoffice.org/1.1.4/>」から無償でダウンロードできる。
- パッチ (プログラムの差分情報) の適用 : Xena をインストールしたフォルダに、ダウンロードした中に含まれるパッチ (プログラム差分情報) をコピーする。

(2) 本調査におけるテスト環境

本調査においては、Xena を一般的なパソコンにインストールしてテストを行った。Xena を利用するにあたり、JRE (version 1.4.2)、OpenOffice (version 1.1.4) を別途導入した。導入後のソフトウェア設定を含めると、環境設定には半日時間程度を要した。

テスト環境に用いたパソコンのスペックを以下に示す。

- CPU : インテル(R) Pentium(R) M 753 MHz

- メインメモリ:512 MB
- ハードディスクドライブ:40 GB
- OS:Microsoft(R) Windows(R) XP Professional

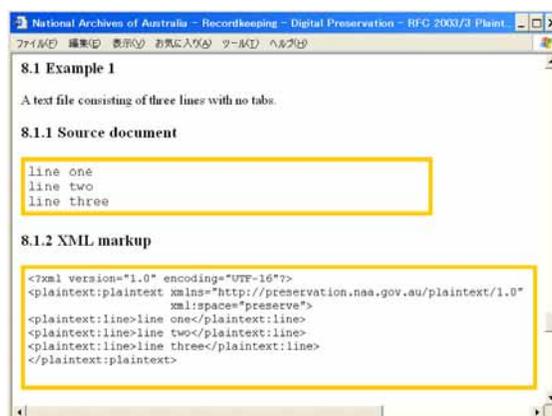
今回のテストは、WindowsXP がプレインストールされた PC 上でテストを行ったが、新たにインストール・設定が必要となった Xena、JRE、OpenOffice はいずれも OS やハードに依存しないアプリケーションであった。

(3) 機能概要

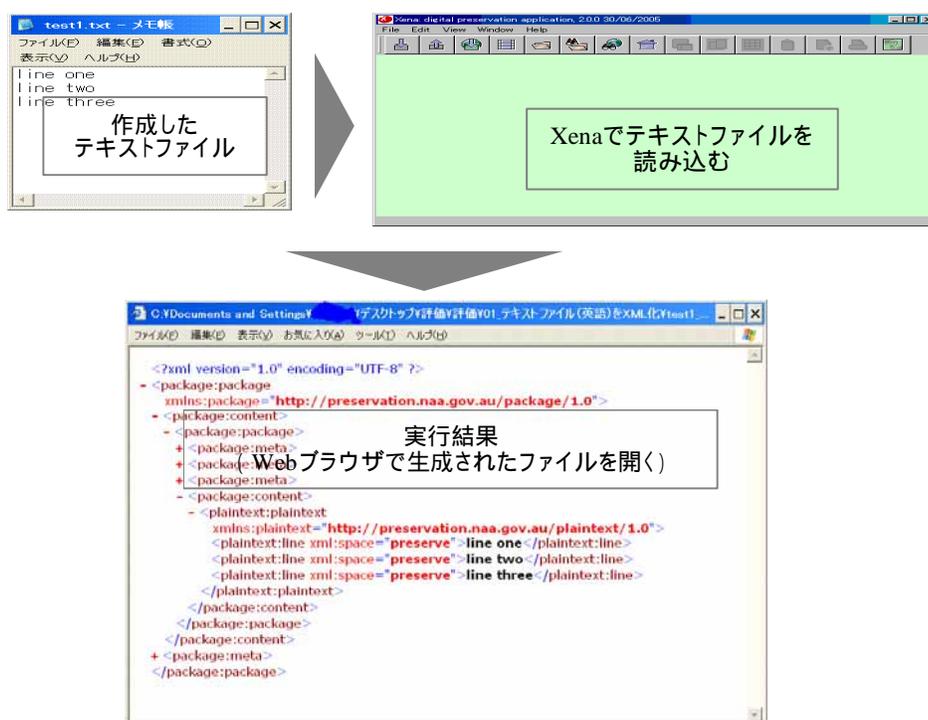
Xena は、変換対象となるファイルを読み込んだ上で、NAA が定義した XML フォーマット(タグの定義等)に基づいて、XML に変換したファイルを提供するものである

また、ビューワ機能もついており、これは XML に変換されたファイルを画面上で(アプリケーションによる)イメージ画面として再現したものである。なお、今回のテスト結果によれば、このビューワ機能は不完全といえる。というのも、XML ファイル自体は日本語も含め作成されていたとしても、ビューワ上での日本語表示機能は備わっていなかったりするためである。

なお、作成された XML ファイルの内容を確認する場合、XML ファイルをブラウザ (InternetExplorer や Netscape)で開くことによって、NAA のタグの規則等に基づいて XML で記述されていることを確認することができる。なお、生成された XML ファイルを再度アプリケーション側に戻して表示させる機能の存在については現時点では確認できていない



図表 4-12 NAA が定める XML データフォーマット



図表 4-13 Xena によるファイル変換の流れ

(4) テスト結果

本調査においては、Xena がサポートするファイル形式のうち、国内で主に利用されている Microsoft Word、Excel、PowerPoint に加え、PDF、JPEG、HTML、Plaintext (TEXT) の各種のファイル (形式) が XML に適切に変換できるかテストを実施した。

テスト結果一覧のとおり、XML 自体は HTML 全般と PDF・テキストの日本語を除き作成できた。またビューアについては、日本語の表示が出来なかった。なお今後 HTML の変換については近いうちに対応が行われる可能性がある。

テスト方法	・National Archives of Australiaが開発したXML変換ソフト「Xena」を利用して、対象ソフトウェア別に動作確認を行った。 テストは、「英語で記述されたファイル」と「日本語で記述されたファイル」の2種類に対して、「Xenaの変換後の状態が表示されるか」、「XML変換したファイルが生成されるか」を確認した。 ・XenaとOpenOfficeとの互換性を調べるため、Microsoft Officeソフトウェアで作られたファイルが、OpenOfficeソフトウェアで開くことができるか確認を行った
テスト環境	ソフトウェア環境: Xena 2.0.0、J2RE 1.4.2.09 (Java環境)、OpenOffice.org 1.1.9、プラットフォーム: Microsoft Windows XP

ソフトウェア	ファイルの変換		ビューア		OpenOffice		備考
	英語	日本語	英語	日本語	英語	日本語	
Microsoft Office	Word			×			
	Excel			×			
	PowerPoint			×			
PDF		(PDFの作成方法等によって動作が異なる模様)		(PDFの作成方法等によって動作が異なる模様)	(PDFファイルを開くことができない)	(PDFファイルを開くことができない)	
JPG							
HTML	×	×	×	×			
Text		(「文字を推測できない」とエラーが出る)		(「文字を推測できない」とエラーが出る)			

：正常動作、：種類によって動作が異なる、×：表示されない/変換されない

図表 4-14 テスト結果一覧

(5) 評価結果

何点かのテストでエラーが表示された。今回のエラーの原因がプログラム側に起因するものかハード側に起因するものか詳細は不明だが、PDF や JPEG 等の XML 変換の際には一定の負荷がかかっていることが想定される。

また、国立公文書館職員の PC 端末での利用を前提とした場合、JRE 等のダウンロードやプログラムの実行によって他のプログラム等に影響が出ないとは限らないため注意が必要である。また、PDF や JPEG について、XML ファイルの作成自体は成功したが、XML ファイルが画像情報をテキストデータでどの程度正確に反映しているかは分からない。というのも、XML ファイルから JPEG や PDF 等に復元する機能は提供されていないためである。

メタデータについては XML 化させて移行させるのが、XML の標準化が進んでいる点からも効果的であろう(電子公文書上のテキストデータの XML 化についても有力な選択肢として残っている)。

4-5-3 メタデータの抽出

行政文書のメタデータは各府省のデータベース上で管理されている。移管時には、データベース内にある当該文書のメタデータを抽出することが必要といえよう。各府省において管理しているメタデータを XML という共通形式で抽出することで、移管された公文書等の情報を一元的に管理することができる。XML 文書がビジネスのデータ交換手段に利用される例が増えてきたため、それを格納するデータベース製品も、XML 文書によるデータの入出力を可能にするべく対応している。

ICA 報告書 16 においても、システムの将来あるアプローチの一つは、文章をマークアップ言語(例えば XML)で記述することである。マークアップ言語を用いるアプローチは記録管理とアーカイブ保存の双方に利点があるため、アーキビストは、このような構想を支持すべきであるとしている。

4-5-4 移管時のセキュリティ

「政府機関の情報セキュリティ対策のための統一基準(内閣官房情報セキュリティセンター、2005 年 12 月)」において情報を移送等する際に守るべき基準が定められている。ただし、基準の対象は「機密情報」が中心であり、これらの基準がそのまま電子公文書等の移管時に適用されるとはいえない。

(1) 政府機関の情報セキュリティ対策のための統一基準

「政府機関の情報セキュリティ対策のための統一基準(内閣官房情報セキュリティセンター、2005 年 12 月)」では、行政事務遂行のために情報を移送する際に発生する恐れがある当該情報の漏えい、滅失、き損及び改ざん等の対策として以下の基準を定めている。

ア. 情報の移送に関する許可及び届出

- 行政事務従事者は、機密性 3 情報(行政事務で取り扱う情報のうち、秘密文書に相当する機密性を要する情報)を移送する場合には、課室情報セキュリティ責任者の許可を得ること
- 行政事務従事者は、機密性 2 情報(行政事務で取り扱う情報のうち、秘密文書に相当する機密性は要しないが、直ちに一般に公表することを前提としていない情報)を移送する場合には、課室情報セキュリティ責任者に届け出ること

イ. 情報の送信と運搬の選択

- 行政事務従事者は、要機密情報を移送する場合には、安全確保に留意して、送信又は運搬のいずれによるかを決定し、課室情報セキュリティ責任者に届け出ること

ウ. 移送手段の選択

- 行政事務従事者は、要機密情報を移送する場合には、安全確保に留意して、当該要機密情報の移送手段を決定し、課室情報セキュリティ責任者に届け出ること

エ. 書面に記載された情報の保護対策

- 行政事務従事者は、要機密情報が記載された書面を運搬する場合には、情報の格付けに応じて、安全確保のための適切な措置を講ずること

オ. オ. 電磁的記録の保護対策

- 行政事務従事者は、要機密情報である電磁的記録を移送する場合には、パスワードを用いて保護する必要性の有無を検討し、必要があると認めたときは、情報にパスワードを設定すること
- 行政事務従事者は、要機密情報である電磁的記録を移送する場合には、暗号化を行う必要性の有無を検討し、必要があると認めたときは、情報を暗号化すること
- 行政事務従事者は、要機密情報である電磁的記録を移送する場合には、必要な強度の暗号化に加えて、複数の情報に分割してそれぞれ異なる移送経路を用いること

(<http://www.bits.go.jp/active/general/pdf/k303-052c.pdf>)

(<http://www.bits.go.jp/active/general/pdf/k303-052.pdf>)

4-6 移管後の技術（保管）

移管された電子公文書等については、データの損失がないよう安全に管理しなければならない。

ICA 報告書 16 にもあるとおり、記録は、業務活動の遂行の過程においてその証拠として作成され、受け取られ、利用される。業務の継続的遂行をささえ、規制環境を遵守し、必要な説明責任を提供する為には、組織は、必要な期間中真正で信頼でき、利用できる記録を作成し、維持し、それらの記録の完全性を保護することが望ましい。

一方、電子公文書等が記録される電子媒体は、紙・マイクロフィルムよりも寿命が短く、環境変化に対してやや脆弱な側面がある。現に、小さな災害や遠隔地での出来事、電源装置の破壊といった場合でさえ電子媒体に損害が与えられる可能性がある。

したがって、移管後も、定期的にシステムのセキュリティ対策やデータのバックアップをとることは、データ保護の観点から必須の作業となる。中でもセキュリティ対策に関しては、日頃から情報システムに関する危険性を把握したうえで適切に対処しておかなければならない。

4-6-1 電子公文書等の記録媒体

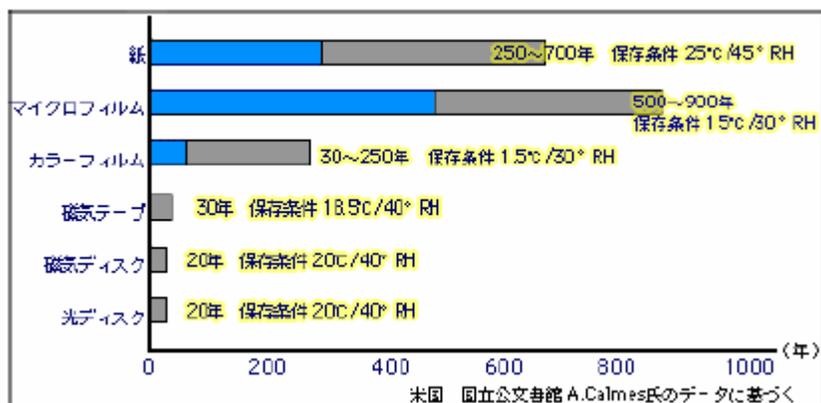
電子情報が記録される媒体は様々あるが、期待寿命や電子情報の書込みや複製し易さ等の特徴に違いがある。電子公文書等の移管及び保存の運用方法に鑑みて、最適な記録媒体を選択する必要がある。

(1) 媒体の寿命

媒体の寿命とは、一定の保存条件下においてメーカーが保証する期待寿命のことをいう。光ディスク、磁気テープ、磁気ディスクの寿命は以下のとおりである。媒体の寿命を見ると、マイクロフィルムが最も長寿命であることがわかる。一方で、磁気テープ、磁気ディスク、光ディスク等のデジタル媒体の寿命は20年から30年にとどまっている。

なお、デジタル媒体(磁気テープ、磁気ディスク、光ディスク)については、再生装置を利用することが前提となるため、技術変化の速さを考慮すると、仮に媒体が存続したとしても、再生装置の陳腐化により、再生環境が失われる可能性がある。したがって、可視性のある紙媒体やマイクロフィルムに比べて情報損失のリスクがさらに高いといえる。

また、上記の寿命は、保存された環境によって大きく異なってくる。環境による保存期間の差は下図のとおり10倍以上にもわたることから、適切な環境での保存及びそれに向けた設備投資が大切である。



図表 4-15 各媒体の寿命予測図

(出典: <http://www.kyokuto-micro.co.jp/media.html>)

(2) 媒体の特徴

保存媒体は、大別して、紙とフィルムとデジタル媒体の 3 つに分けられる。フィルムの場合、長期保存が可能だが、画像として保存されるためテキスト情報を保持することができない。一方、デジタル媒体は媒体の寿命が短いものの、電子情報の複製がし易く、電子公文書等のテキスト情報をそのまま保存することができる。

ア. 紙

紙資料は保存場所をとり、資料を探すのに時間がかかるが、汚れや傷で読み難くなくても情報は残る。

イ. フィルム系(マイクロフィルム、カラーフィルム)

媒体の寿命が長いが、全て画像としての保存となってしまう(テキスト情報を保持できない)ため、媒体に保存されている情報の検索に向かない。

ウ. デジタル媒体(磁気テープ、磁気ディスク、光ディスク)

傷や故障により情報が無くなる可能性があるが、電子情報の書込みや複製が簡単で、情報を検索し易い。

(3) 電子公文書等の長期保存における媒体の選択と課題

電子公文書等を長期保存することを考えると、媒体の寿命が長いマイクロフィルムが好ましいように思える。しかし、電子公文書等の移管頻度や利用時の利便性を考慮に入れると、画像しか保存できないマイクロフィルムよりも、書込み・管理が簡単で、テキスト情報や動画等も保存が可能なデジタル媒体が適しているといえる。

4-6-2 バックアップ

移管された電子公文書等は、あらゆるデータ消失の可能性から防ぐ必要がある。バックアップとはデータ保護のため、別のメディアにデータの複製を取っておく技術である。

(1) バックアップの必要性

電子情報の場合、人為的ミス、コンピュータウイルス、機器類の故障、事故、天災、テロ行為などさまざまな理由でハードディスク等に保存されているデータが消失する可能性がある。バックアップとはデータ保護技術の1つで、データ全体または変更分をテープやディスクに対して定期的に保存することをいう。

また、ICA 報告書 16 においても、大半のコンピュータの保存媒体の寿命は、理想的な保存条件の下にあるいくつかの種類では5年以上となり、理想とは程遠い温度、湿度、照明のレベルの下にあると、多くの場合5年より短くなるという現状を踏まえ、多数の複製を利用することによる電子記録の保護を推奨している。

(2) バックアップの方法

バックアップを実施する場合、データを圧縮してバックアップすることも考えられるが、大容量のデータを圧縮する行為に長い時間が必要となる(データを戻すときも解凍が必要となる)ので、データを圧縮してバックアップは行わないのが一般的である。

以上のことから、データのバックアップは、フルバックアップと差分若しくは増分バックアップを組み合わせて行うことが多い。

バックアップの方法にはフルバックアップ、差分バックアップ(ディファレンシャルバックアップ)、増分バックアップ(インクリメンタルバックアップ)の3種類がある。電子公文書等が移管された際に、移管分の情報を差分若しくは増分バックアップする。保管システムの改良を行うなど、データ損失の可能性がある作業の前には、フルバックアップを取得するなどの運用が考えられる。

(<http://www.itmedia.co.jp/enterprise/articles/0408/18/news060.html>)

ア. フルバックアップ

フルバックアップとは、対象のファイルシステム全体を取得するバックアップのことである。全ファイルがバックアップ対象となるためバックアップには時間がかかるものの、それ自身が完全なデータを保持していることから、データの保護強度はきわめて高い。通常は、十分なバックアップウィンドウを確保できる日に採用される。

イ. 差分バックアップ

差分バックアップはフルバックアップに対しての変更分をバックアップするもので、フルバックアップと併用する形がとられる。差分バックアップの場合は、日を追うごとにバックアップの時間が長くなる傾向にある。ただし、フルバックアップからの変更分を直接保持しているため、フルバックアップと最新の差分バックアップのデータさえ用意すれば、何かあった場合も最新の状態にまで復元できる。

ウ. 増分バックアップ

増分バックアップは前回のバックアップ時に対しての変更分をバックアップするもので、フルバックアップと併用する形がとられる。増分バックアップは、日々のバックアップ時間

は短縮できるものの、復元の手順が複雑になりやすい。例えば、フルバックアップの翌日より4日間の増分バックアップを行った場合、データの復元にはフルバックアップのデータに加え、4日分の増分バックアップのデータを必要とする。

(3) 災害対策

災害などによって生じたシステム障害を復旧させる作業をディザスタ・リカバリーという。データのバックアップは、ハードの障害やソフトのバグ、人的な操作ミスなど日常の運用に備えたものであるのに対し、ディザスタ・リカバリーは、バックアップしたデータを遠隔地に保存しておく。近年、ネットワークのコストが安くなっているため、遠隔地にあるディスクにオンライン接続して、バックアップのデータを保存する方法に切り替える企業等も増えている。

(4) 電子公文書等の保管におけるバックアップの活用と課題

移管後の電子公文書等は、他に代替するものがないため、バックアップ運用は不可欠であるといえよう。電子公文書等の移管に合わせて、バックアップの取得時期を決めることが必要である。電子公文書等は、重要な公的記録であることから、バックアップにおいては災害対策も視野に入れることが考えられるが、現時点においては今後の検討課題といえよう。

4-6-3 保管時のセキュリティ

電子公文書等を含む電子情報の保管に関するシステムのセキュリティとしては、政府機関の情報セキュリティ対策のための統一基準(内閣官房情報セキュリティセンター、2005年12月)や記録管理の国際標準であるISO15489が出されている。電子公文書等の保管を行ううえでも、これらに準拠した構成にする必要がある。

また、緻密な温度湿度管理や嚴重な防災対策を実施できる施設、設備を整備する必要がある。比較的短期かつ小さな温度・湿度の上昇、あるいは電圧の変化等でも電子媒体がほぼすべて解読不可能になるというおそれがある。

(1) 政府機関の情報セキュリティ対策のための統一基準

「政府機関の情報セキュリティ対策のための統一基準(内閣官房情報セキュリティセンター、2005年12月)」では、情報の保存を続ける限り発生する恐れがある、当該情報の漏えい、滅失、き損及び改ざん等の対策として以下の基準を定めている。

ア. 格付けに応じた情報の保存

- 情報システムセキュリティ責任者は、電子計算機に保存された要保護情報について、適切なアクセス制御を行うこと。
- 行政事務従事者は、情報の格付けに応じて、情報が保存された外部記録媒体を適切に管理すること
- 行政事務従事者は、情報システムに入力された情報若しくは情報システムから出力した情報を記載した書面のうち要機密情報を記載した書面、又は重要な設計書を適切に管理すること
- 行政事務従事者は、要機密情報を電子計算機又は外部記録媒体に保存する場合には、暗号化を行う必要性の有無を検討し、必要があると認めるときは、情報を暗号化すること

- 行政事務従事者は、要保全情報を電子計算機又は外部記録媒体に保存する場合には、電子署名の付与を行う必要性の有無を検討し、必要があると認めるときは、情報に電子署名を付与すること
 - 行政事務従事者は、要保全情報若しくは要安定情報である電磁的記録又は重要な設計書について、バックアップ又は複写の必要性の有無を検討し、必要があると認めるときは、そのバックアップ又は複写を取得すること
 - 情報システムセキュリティ責任者は、要保全情報若しくは要安定情報である電磁的記録のバックアップ又は重要な設計書の複写の保管について、災害等への対策の必要性を検討し、必要があると認めるときは、同時被災等しないための適切な措置を講ずること
- イ. 情報の保存期間
- 行政事務従事者は、電子計算機又は外部記録媒体に保存された情報の保存期間が定められている場合には、当該情報を保存期間が満了する日まで保存し、保存期間を延長する必要性がない場合は、速やかに消去すること
- (<http://www.bits.go.jp/active/general/pdf/k303-052c.pdf>)
(<http://www.bits.go.jp/active/general/pdf/k303-052.pdf>)

(2) ISO15489 について

ISO15489-2(4.2.5.2)にセキュリティ強化のためのステップが明記されている。以下の7つのステップを踏むことで、保管された情報が取り込まれたり、記録されたりしたときに、警告や制限を自動的に知らせることができるとしている。

- 記録や組織情報へのアクセスを規定し、施行されている法律を明らかにする
- 個人や商業上の秘密の漏洩等のリスクの大きさを明らかにする
- 組織におけるセキュリティの課題を明らかにする
- 発生確率や損害の大きさに応じて、セキュリティ違反に関するリスクの大きさをランク付けする
- 組織のビジネス活動におけるセキュリティの課題やリスクの大きさをマッピングする
- 最もリスクが大きいものから低いものに対する制限のレベルを作成する
- リスクの順序などに従って、ビジネス活動の記録分類に制限のレベルを割り当てる
- 記録を利用するシソーラスやシステムに対して、割り当てた制限のレベルを関連づける

(3) 電子公文書等保管システムのセキュリティ

電子公文書等の保管システムには、移管された電子公文書等の原本が保管されることとなる。そのため、保管システムは厳重に管理する必要がある。

保管システムの管理を厳重に行うためには、当該システムに対するアクセス権やアクセスログの管理を行うことが必須である。システムのログを管理することで、誰がアクセスしたのかを特定できるようにする。アクセスだけでなく、どのような処理を行ったか(例えば、廃棄等)も記録することで、保管システム及び電子公文書等に対するアクセスを厳重に管理することが可能となる。

また、電子公文書等を保管システムに保存する前には、長期間のウィルスチェック(検疫)を義務づけるとともに、保管システムに対しても、ウィルス対策ソフトを導入することが不

可欠である。万が一、ウイルス感染した電子公文書等が保管システムに保存されてしまった場合、他の電子公文書等に感染するおそれがあるためである。

なお、オーストラリア NAA では、電子公文書等を保管するシステムは外部から切り離されたネットワーク上に設置されており、外部からの攻撃を遮断している。保管システムも外部アクセスから切り離されたネットワーク構成を検討する必要がある。

4-6-4 原本性保証、改ざん防止

ISO15489-1(7-2 項)では、真正性、信頼性、完全性、利用可能性という記録の 4 つの特徴について言及がある。

そのうち、真正性、信頼性、完全性については、作成されたときのデジタル情報がその後不正な改変をされていないかを、デジタルで検証する方法がある。すなわち、電子署名やタイムスタンプといった技術である。しかし、現時点では、電子署名やタイムスタンプの保証期限が過ぎると再度付与する必要がある等、長期にわたって有効性を検証し続けることが困難である。また、特定のベンダの技術に依存するリスクも無視できない。

(1) 電子認証・電子署名

電子署名とは、当該情報が当該措置を行った者の作成に係るものであることを示すためのものであり、当該情報について改変が行われていないかどうかを確認することができるものをいう。保管されている電子公文書等に対して、電子署名を付与しておくことで、文書の改ざんを防止することができる。だが、電子公文書等の保存期間よりも電子署名の有効期限が短いことが多い。したがって、現時点においては、署名延長技術をはじめとする技術動向を引き続き注視する必要がある。

ア. 電子署名

「電子署名」とは、電磁的記録(電子的方式、磁気的方式その他人の知覚によっては認識することができない方式で作られる記録であって、電子計算機による情報処理の用に供されるものをいう。以下同じ。)に記録することができる情報について行われる措置である(電子署名法(電子署名及び認証業務に関する法律)第 2 条第 1 項)

イ. 認証業務

「認証業務」とは、自らが行う電子署名についてその業務を利用する者(以下「利用者」という。)その他の者の求めに応じ、当該利用者が電子署名を行ったものであることを確認するために用いられる事項が当該利用者に係るものであることを証明する業務をいう。(電子署名法(電子署名及び認証業務に関する法律)第 2 条第 2 項)

ウ. 特定認証業務

「特定認証業務」とは、電子署名のうち、その方式に応じて本人だけが行うことができるものとして主務省令で定める基準に適合するものについて行われる認証業務をいう。(電子署名法(電子署名及び認証業務に関する法律)第 2 条第 3 項)

エ. 電子署名延長技術

電子署名技術を単独で、長期間の原本性証明に活用するには様々な課題があるとされている。まず、電子証明書が証明可能な期間は、暗号化の際に用いる秘密鍵の安定

性の観点から3年程度までとされている。将来の高性能のコンピュータの出現により、現在の暗号化のアルゴリズムが脆弱化し、暗号が解読される可能性を完全には否定できないためである。

また、電子証明書を発行し、証明を行う認証機関が、合併・吸収・倒産等により消失した場合には、たとえ証明書有効期間内であっても、署名の検証が行われない場合がある。

これらに対しては、一部民間企業から、署名延長システムが開発されたところである(三菱電機、三菱 Misty Guard 署名延長システム)。同製品は、署名された電子文書に関して、署名の有効性を検証できる客観的なデータをそろえ、その電子文書に対し、そのデジタル署名の有効性が失われる前に署名(タイムスタンプ)をしている。三菱電機ホームページによると、標準技術への準拠をうたい文句としており、電子商取引推進協議会(ECOM)が平成15年(2003年)3月に発表した電子署名文書長期保存に関するガイドラインで推奨する、RFC3126 フォーマットに準拠した業界初の製品とのことである。

(2) タイムスタンプ

タイムスタンプとは、第三者機関により電子データに対して正確な日時情報を付与し、その時点での電子データの存在証明と非改ざん証明を行う仕組みあるいは技術をいう。

タイムスタンプにより、文書が存在した時刻を証明すること、改ざんが行われていないことを確認することができるが、電子署名を利用したタイムスタンプの場合、電子署名と同様にタイムスタンプの有効性が失われる前に、証明書の有効期間を再延長する必要があり、引き続き技術動向を注視する必要があると思われる。

ア. タイムスタンプについて

電子文書が「誰が」「いつ」「何を」「どこで」「誰に」作成されたか、特に、電子文書の内容としてみた場合には、「いつ」と「何を」が正しく証明されることが重要となる。(「いつ」とは作成日時であり、これは日時によって文書の位置づけが違ってくるため。「何を」は内容そのもの。)

現在上記の技術的な方法で最も活用されているのが、「タイムスタンプ」であり、「いつ」「何を」作成したかを、デジタル上での日付・時刻入りの「はんこ」を押すことによって、技術的に証明するものである。

タイムスタンプは、電子文書をはじめとするデジタル情報について、その内容が一定の時間に存在したことを示す「存在証明」及び、その時間以降に改ざんされていないことの「非改ざん証明」の二つの役割を果たす。タイムスタンプが押された電子文書を受け取った場合、受取人はスタンプ上の時刻・日付を確認し、「その時刻以降には文書はかいつ改ざんされていない」ことが分かる。

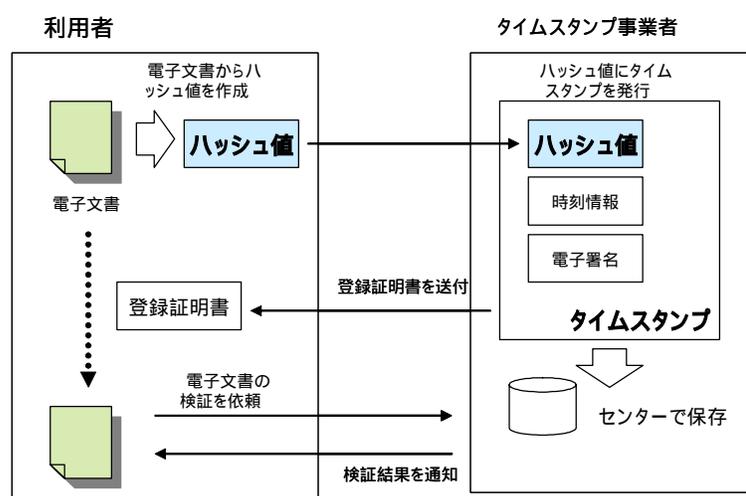
イ. タイムスタンプの検証

タイムスタンプを施したデータや電子文書が、本当にその時に生成され、その時の状態を維持しているかを確認することをタイムスタンプの検証という。検証方法は前述のプロトコルにより違いがある。

ウ. 認定タイムスタンプ

タイムスタンプの機能としては、正確な時刻を記録するという役割に加え、郵便物の消印のような「御墨付き」的な信頼性を制度的に付与することも期待されている。

これについては、2005年4月から施行された「e-文書法」において対象とする電子化書類のうち一部(税務関連)でタイムスタンプの付与が制度化された際に、認定制度が設けられることとなった。認定制度は財団法人日本データ通信協会が提供しているもので、同認定を通過した民間のタイムスタンプを利用した場合には、制度上「タイムスタンプ」として御墨付きを与えられることとなる。



図表 4-16 タイムスタンプイメージ

(3) 電子透かしについて

電子透かしとは、コンピュータで扱う、画像、動画、音楽などに、目や耳では識別できないようなデータを埋め込み不正コピーなどを検出する技術のことである。

画像や動画、音声などのマルチメディアデータに、画質や音質にはほとんど影響を与えずに固有の識別符号を埋め込み、所有権の所在を主張しようと企図したものである。これは、不正コピー者に無言の圧力を与え、あるいは著作権侵害に対する検証の道具として利用されることになる。即ち、「透かし」データを埋め込まれた画像などのデータは、一見すると元のデータと変わりないように見えるが、専用の電子透かし検出ソフトに読み込ませると、作者名やコピー回数などの埋め込まれた情報が表示される。

これにより、不正コピーやデータの改ざんなどを防ぐことができる。さらに、データの改ざんの事実だけでなく改ざんされた個所も具体的に特定できるようにすることも可能である。

電子公文書等は国民一般にとって有用な知的資産であることから、流用を防ぐためには電子透かしを付与しておくことが考えられる。改ざん位置検出を行う一つの方法として利用するという目的もあり得る。

ただし、電子署名と同様に、特定のベンダの技術に依存する面があるため、長期間にわたる検証可能性をどのように確保するかという課題がある。したがって、引き続き技術動向を注視する必要があると思われる。

4-7 移管後の技術（利用）

4-7-1 マルチビューワ（複数ファイル形式対応ビューワ）

マルチビューワは、1つのソフトウェアで複数のファイル形式の閲覧が可能なソフトウェアである。マルチビューワで対象としているファイル形式であれば、後述するエミュレータ環境を構築するよりも簡単に再生環境を構築することができる。

(1) マルチビューワについて

マルチビューワとは、複数のファイル形式に対応したビューワをいう。一般的に、文字の大きさやフォント、字飾りなどの情報を埋め込むことができるリッチテキスト(Word や Excel など)は、専用のアプリケーションソフトを持っていなければ、作成した文字の大きさや字飾りなどの情報を正確に表現することができない。例えば、Windows 系の OS にはそのファイルの種類を表す拡張子(例: .doc)がある。拡張子(例: .doc)とアプリケーションソフト(Word)が対応付けられているため、当該ファイルをダブルクリックするだけでアプリケーションソフトを利用して開くことができる。

マルチビューワは、ファイル形式をファイル内容・構造から判別するため、ファイル名に拡張子(.doc)がなくても、マルチビューワで対象としているファイル形式であればファイルを開くことができる。

現在、マルチビューワは様々なソフトウェアが販売されており、MS-Office ドキュメントから、DOS ワープロまで、幅広い文書形式に対応している物もある。また、マルチビューワは、それぞれのワープロソフトがなくても、文書内容を表示・印刷することができる。

(2) 移管された電子公文書等の利用段階におけるマルチビューワの活用と課題

マルチビューワは、後述するエミュレータに比べ、実現しやすく、維持費用がかからずに再現環境を提供できる技術といえる。エミュレータの場合、ハードウェアが変わる度に仮想環境のシステムを構築する必要があるが、マルチビューワの場合はソフトウェアを導入するだけで過去のファイルを閲覧することができるためである。

ただし、画像や動画、一部の HTML(複数ファイルに関連付けられているもの)については、マルチビューワで閲覧することが難しい。また、マルチビューワが永続的に過去のファイル構造を対象とするとは限らない点が課題といえよう。

4-7-2 エミュレーション

エミュレーションとは、あるコンピュータシステム上で、ほかのハードウェア用のアプリケーションを仮想的に動かす技術のことをいう。古い OS に対応したソフトウェアが新しい OS で使えない場合、エミュレータを利用することで、過去の OS やソフトウェアを最新の PC 上で動かすことができる。なお、エミュレータとはエミュレーション技術を提供するソフトウェアのことをさす。

移管された電子公文書等を閲覧する際、ファイルの標準化(ファイル形式変換)やマルチビューワ等で再現環境を実現することができなかった場合、仮想的に擬似環境を実現して電子公文書等の閲覧を可能にするのが、エミュレーションである。

エミュレーションは試験途中の技術であるが、エミュレーション技術によって最新型のコンピュータで、旧型ソフトウェア・アプリケーションを動かすことが可能となる。ある特定のコンテキストにおいては実用的な解決策であることが実証されているとのことである（ICA 報告書 16）。そういった面で有力ではあるが、長期的に有効かどうか断言できるまでにはいたっていないため、他の手段と並行して進めていくことになる。

(1) エミュレーション(エミュレータ)の特徴

平成 16 年度に国立国会図書館が行った「電子情報の長期保存とアクセス手段確保のための調査報告書」では、「ハードウェアや OS が移り変わっても、エミュレータのみ作成すればそれまでのアプリケーション・ソフトウェアや OS を利用することが可能となる。このようなエミュレータは多数作成され、頒布されている。しかし、独自に作成するためには技術力が必要であり、完全に旧式環境を再現するとは限らない。」とされている(同報告書 P97)。

(http://www.ndl.go.jp/jp/aboutus/report_2004.pdf)

エミュレータは、複数の OS を同時に使うことになるため、非力なハードでは利用できない。また、OS やアプリケーションのライセンスは仮想環境で扱う分も必要であるため、エミュレータ環境を構築する度にライセンス費用が別途必要となる。

(2) 海外におけるエミュレータの活用(オーストラリア)

オーストラリア NAA では、電子情報の保存のガイドラインとして、「エミュレーション」も「マイグレーション」等と組み合わせて利用していくべきと記されているが、具体的なシステム構成等についての記述は無い。

(<http://www.naa.gov.au/recordkeeping/er/guidelines/10-preservation.html>)

エミュレータ機能を提供する代表的製品である「VMWare」の実績は国内及び海外でいくつが存在する。しかし、電子情報の長期保存のためより、むしろ古い OS で開発した独自アプリケーションを改修せずに新しいハードウェアで動かすことを目的としたものが多くなっている。

(3) 移管された電子公文書等の利用段階におけるエミュレータの活用と課題

国立公文書館への移管対象となる電子公文書等の全てが「標準化されたファイル形式」で保存できるとは限らない。例外的に、作成当時のシステム環境を再現しなければ、閲覧・利用できない電子公文書等が存在した場合、エミュレータを用いて電子公文書等を再現することになるだろう。

しかし、エミュレータを導入しているコンピュータシステムのハードウェアに寿命がある限り、1 度作った仮想環境が永久に残るわけではない。結局、ハードウェアが変わるたびに、エミュレータで再度同じ環境を構築する必要があり、半永久的な保存を行う使命を帯びている国立公文書館における電子公文書等の利用段階で、継続的にエミュレータを利用することは、不可能ではないにせよきわめて困難なことである。また、エミュレーション実現のためには、過去に利用された可能性のあるソフトウェアを全て用意しておく必要があるため、維持管理に費用がかかるといった課題もある。

4-7-3 全文検索機能

文書を検索するためには、従来すべての記事に対してキーワードを振る必要があり、検索する場合、このキーワードと書誌データと呼ばれる雑誌名、著者名、タイトルなどを基に行っていた。近年は能力も強力になり、電子化された記事のなかを高速に直接全部調べて検索できる全文検索システムが実用化されている。電子公文書等の利用段階で検索機能を実現する場合に、テキストデータをどのレベルまで利用するかという議論はあるが、技術動向を把握することは必要であろう

(<http://www.gengokk.co.jp/zenbun.htm>)

(1) 概要

全文検索システムとは、文書に含まれる全ての情報を対象に検索する方法をいう。キーワード検索は、あらかじめキーワード付け文書に対して、単語やその論理式を入力して検索する方法であるのに対し、全文検索では、あらかじめキーワードづけされていない文書を対象に検索を行うことができる。

全文検索機能は読み取ったテキストデータを細かい単位に分けてインデックスを作成しており、「形態素解析」と「N-gram」による方法がある。しかし、単語ごとに「分かち書き」する英語に比べて、日本語は文を単語ごとに切り分ける処理が必要となるため、いかに意味のある言葉として抽出するかが課題となっている

ア. 形態素解析

形態素解析は、辞書を使って意味を持つ最小の単位(形態素)に文字列を区切る方法。検索ノイズを減らすことができるが、区切り方が異なると、同じ文字列でも検索できなくなってしまうことがある。例:「東京都」を「東・京都」とするか、「東京・都」と区切るか

イ. N-gram

N-gram は、ある長さで文字列をぶつ切りにする方法。検索漏れを防ぐことができるが、意味のない文字列が一致することがあるため、検索ノイズが発生しやすい。

(2) 電子公文書等の利用段階における全文検索技術の活用と課題

ICA 報告書 16 では、XML のような現在の標準フォーマットに準拠したフォーマットにフラットファイルの情報を記録すると、情報検索方法の選択肢は非常に拡大するとしているが、テキスト検索については、メタデータベースではカバーすることが難しい。

その点で、(精度が多少劣ったとしても)全文検索システムによる対応は利用者の検索効率を高めるための効果的な機能として参考になる。もっとも、全文検索を実現する場合の前提として、移管される電子公文書等にテキストデータが埋め込まれていることが必要であるため、仮に画像データのみを利用者に提供する場合には全文検索を行なうことはできない。

(参 考) 日 経 BP ウ ェ ブ 記 事
<http://arena.nikkeibp.co.jp/qa/other/20020603/100595/>

4-7-4 利用時のセキュリティ

情報の利用(提供)に関するシステムのセキュリティとしては、政府機関の情報セキュリティ対策のための統一基準(内閣官房情報セキュリティセンター、2005年12月)が出されている。万が一、公開情報が改ざんされた場合の対策を考えておく必要がある。

(1) 政府機関の情報セキュリティ対策のための統一基準

府省庁外に情報を提供する場合、情報の不適切な取扱いにより、当該情報の漏えい又は不適切情報の保存を続ける恐れがある。「政府機関の情報セキュリティ対策のための統一基準(内閣官房情報セキュリティセンター、2005年12月)」では、その対策として、情報提供及び情報の公表に関する基準を定めている。

- 行政事務従事者は、情報を公表する場合には、当該情報が機密性 1 情報に格付けされるものであることを確認すること(機密性 1 情報とは、秘密文書に相当する機密性はなく、一般に公表することを前提とする情報をいう)
- 行政事務従事者は、電磁的記録を公表する場合には、当該情報の付加情報等からの不用意な情報漏えいを防止するための措置を採ること
(<http://www.bits.go.jp/active/general/pdf/k303-052c.pdf>)
(<http://www.bits.go.jp/active/general/pdf/k303-052.pdf>)

(2) 電子公文書等利用システムのセキュリティ

電子公文書等の利用システムには、エッセンスだけ抽出して保存された電子公文書等の原本の複製物が保存されている。そのため、万が一、公開されている情報が改ざんされたとしても、再度、保管システムから電子公文書等の原本の複製物を利用システムに移すことで公開情報を復元させることができる。したがって、システムの復旧やデータの復元等が速やかに行えるような体制を整える必要がある。

4-8 その他の技術動向

4-8-1 文字コードの標準化

現状の技術では、外字の再現を行うのは、文字を表現する OS やアプリケーションが未対応となる可能性が高く難しい。そこで、作成したアプリケーションに関係なく閲覧できるファイル形式を活用することも考えられる。

(1) 外字について

外字については、行政文書においては、特に申請届出等の場合に用いられる機会が多いが、これらの外字に関する情報が電子的な長期保存の対象となった場合、再現する際に、外字の再現が引き続き可能となっているかどうかという問題がある。

この点について、本来であれば、標準的な規格やガイドラインを政府が作成することが望ましいが、現在は策定されておらず、ベンダ各社が独自に外字の定義を行い、対応する製品を販売しているのが現状である。

このような現状から、日本経済団体連合会(日本経団連)においても、平成 15 年 3 月に要望書「新IT戦略に関する提言」を出している。また、政府のIT戦略会議においても、平成 15 年 7 月の第 17 回会合において、ある委員より「今、中央省庁、あるいは地方公共団体で認められる外字というのはバラバラであり、民間企業にとってはこれがシステムロード、あるいはシステムコスト上の大きな負担の原因となっている。」とのコメントが出されている。

(2) 電子公文書等の移管・保存における外字の課題

電子公文書等を移管・保存する際、外字の再現は重要な課題である。電子公文書等としては、「電子公文書の文書型定義(DTD)の統一的な仕様」に定義があるものの、今後、外字の扱いの単純化等が社会的に行われる必要性について検討することも考えられる。

(3) 文字コードに関する関連機関の動向

ア. 日本工業規格(JIS)について

日本工業規格(JIS)は、工業標準化法に基づき設置された審議会である「日本工業標準調査会」によって設定、改正等が審議されている。

「日本工業標準調査会」は、工業標準化法に基づいて経済産業省に設置されている審議会で、工業標準化全般に関する調査・審議を行っており、情報技術分野においては、下部組織にあたる「標準部会情報技術専門委員会」で審議されている。

イ. JIS 規格について

日本では JIS により、複数の文字コード関連の規格が定められている(以下は主なもの)

- JIS X0201 は、7 ビット及び 8 ビットの情報交換用符号化文字集合で、英数記号といわゆる半角カナを表現することができる

- JIS X0208 は、7ビット及び8ビットの2バイト情報交換用符号化漢字集合で、いわゆる全角英数、ひらがな、記号、第一水準、第二水準の漢字などを表現することができる
 - JIS X0221-1 は、国際符号化文字集合(UCS)と呼ばれ、世界各国の文字約10万文字を表現することができる。Unicode3.0とほぼ同じである
- (出典：日経コンピュータ 2004.6.14)

ウ. 文字コードに関する JIS 規格の近年の動き

情報機器の急速な普及で表外漢字が使用される機会が多くなり、「一般の書籍で用いられている字体がワープロ等で打ち出せない」、「どの字体を標準字体とするか」という問題が生じていた。経済産業省でも、「表外漢字自体表」(常用漢字に含まれない漢字)の社会的影響を考慮し、2001年から文字コード規格の見直しに着手し、2004年2月に「JIS X 0213」が改定された。

(http://www.jsa.or.jp/domestic/instac/h13reports/jcs_houkoku.htm)

(<http://www.meti.go.jp/kohosys/press/0004964/>)

エ. 政府の IT 政策における文字コード標準化について

IT 政策パッケージ 2005 において、文字コード標準化の推進とデータベースの運用(経済産業省及び関係府省)が示されており、「住基ネット統一文字、戸籍統一文字の重複関係を整理、体系化したデータベースを2005年末までに開発する。また、現在構築中の登記統一文字について、2006年度末までに、このデータベースとの整合性を図ることとする。さらに、データベースの具体的な運用方法について関係府省で検討を行い、2005年末までのできる限り早期に結論を得る」としている。

(<http://www.kantei.go.jp/jp/singi/it2/kettei/050224/050224pac.html>)

(4) UNICODE について

UNICODE とは、Apple 社、IBM 社、Microsoft 社など米国の情報関連企業が中心となって提唱し、1993年に国際標準化機構(ISO)で ISO/IEC 10646 の一部(UCS-2)として標準化された文字コード体系のことで、すべての文字を16ビット(2バイト)で表現し、1つの文字コード体系で多国語処理を可能にしようとする。世界の主要な言語のほとんどの文字を収録している。

(<http://e-words.jp/w/Unicode.html>)

現在、Unicode は追加の漢字を登録するため、結局4バイトに拡張され、現在は約10万文字を収録している。

ア. 課題 1: 漢字表現の課題

2バイト表記では最大65536文字しか収録できないため、中国、日本、韓国の漢字が収まりきらず、3カ国の漢字で字形がほぼ同じものを統合して収録した。その結果、文字コードは統一できたが、Unicode 単体では文字列が日本語か中国語か韓国語かを区別する手がかりがない。

また、国や地域ごとに、漢字の字形が微妙に異なっている事実を無視して、同じものとして、扱っているという問題もある。

イ. 課題 2: 完全対応しているソフトウェアの少なさ

Unicode は近年 Windows や MacOS の内部コードに採用され普及が進んだ。Windows は OS レベルで Unicode に対応している (Office 製品も Unicode に対応している) とのことだが、実際には対応が不完全な部分も残っている模様であり、その実際には Unicode 対応といっても完全に対応しているものは少ないようである。

(http://www.jagat.or.jp/story_memo_view.asp?StoryID=8154)

また、各種文字コードとの相互変換において完全でない部分も多いとのことである。

(<http://support.microsoft.com/default.aspx?scid=kb;ja;JP170559>)

5. ウェブページの移管・保存技術の動向

5-1 ウェブアーカイブの動向

5-1-1 背景

電子公文書等の管理・移管・保存について考える場合、国の機関が作成する電子公文書等の一つとして、ウェブページの存在が挙げられるが、特に今回の調査では、これまでも注目されてきた国民向けの公開ウェブページに加え、行政内部向けのページにも着目することとした。

今後は、国立公文書館の役割や、一般の電子公文書等の移管との重複等の可能性等を踏まえつつ、何をどのように移管するかが論点となる。

5-1-2 政府の認識

様々な情報がウェブ経由で政府から提供されていることもあり、政府もウェブ情報の重要性について認識を高めている。今回の研究会(第1回)においても、事務局から「行政機関の中でウェブページの重要性が増えている。したがって、国の機関が作成するウェブページ上で開示される情報を国の機関の活動の証拠として保存していく必要がある」との問題提起がなされたところである。

しかしながら、これまでウェブページは国立公文書館に移管された実績はなく、また特段保存に関する配慮もなされていない状況である。背景として、移管の仕組みがウェブ上の公文書等を前提としてないことが指摘されている。

5-1-3 事例

国立公文書館法に基づく移管制度の一環として、国立公文書館におけるウェブ上の公文書等の移管・保存を検討するにあたっては、まず、我が国及び諸外国におけるウェブ情報のアーカイブ化の動向を把握する必要がある。

諸外国の状況を見ると、アメリカの「インターネット・アーカイブ」や「MINERVA」、オーストラリアの「PANDORA」など、実現形式は若干異なるものの、ウェブ情報のアーカイブが実運用されている。

一方、日本でも、平成15年(2003年)よりe-Japan戦略によってウェブ情報のアーカイブ化が提唱されており、現在国立国会図書館を中心にウェブ情報アーカイブの実現に向けて検討が進められているところである。

5-1-4 我が国におけるウェブ情報アーカイブ動向

我が国のウェブ情報アーカイブの状況としては、平成 15 年に e-Japan 戦略で重点施策として取り上げられ、総務省の実証実験や国立国会図書館のインターネット資源選択的蓄積実験事業(WARP)などの実験事業を経て、ウェブ情報アーカイブの技術的な基礎が固められてきたところである。

行政のウェブ情報については、インターネット上のホームページが主な収集対象とされており、イントラネット上にある行政内部向けのページは特に対象とされていない。

(1) e-Japan 戦略

平成 15 年の「e-Japan 戦略」にて、ウェブ情報のデジタル化・アーカイブ化が取り上げられたのをはじめ、平成 16 年の「e-Japan 戦略 加速化パッケージ」及び「e-Japan 重点計画 2004」にて、国立国会図書館のウェブページ・アーカイブを活用した政府各機関ホームページの長期的保存がうたわれている。

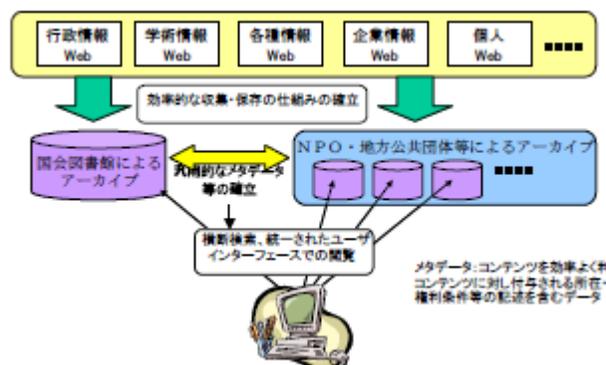
資料の題目	Web情報アーカイブ関連記載の抜粋
e-Japan戦略 (平成15年7月2日 IT戦略本部)	放送・出版等のコンテンツや、美術館・博物館や図書館等の所蔵品、Web 情報、特色のある文化等のデジタル化・アーカイブ化、および国内外への発信を推進する。
e-Japan重点計画- 2003 (平成15年8月8日 IT戦略本部)	デジタルアーカイブ化の推進(内閣府、総務省、文部科学省及び関係府省):2005 年度までに、放送・出版、映画等のコンテンツや、美術館・博物館、図書館等の所蔵品、Web 情報、地域文化、アジア諸国との関係に関わる重要な公文書等について、デジタル化・アーカイブ化を推進し、インターネットを通じて国内外に情報提供が行われるよう必要な措置を講ずる。
e-Japan 戦略 加速化パッケージ (平成16年2月6日 IT戦略本部)	政府コンテンツのデジタルアーカイブ構築と一般利用の拡大:国立国会図書館における政府刊行物アーカイブ(文書や記録を電子的に集積し保管する書庫)構築及び同図書館のウェブページ・アーカイブを活用した政府各機関ホームページの長期的保存により、国等の有するコンテンツの利用機会の拡大と保存を図るため、同図書館も参加した連絡会議を設置し、アーカイブの構築や公開に関するルールの明確化など、同図書館への協力体制を2004 年度中に確立する。(内閣官房及び全府省)
e-Japan重点計画 - 2004 (平成16年6月15日 IT戦略本部)	政府コンテンツのデジタルアーカイブ構築と一般利用の拡大(内閣官房及び全府省):国立国会図書館における政府刊行物アーカイブ(文書や記録を電子的に集積し保管する書庫)構築及び同図書館のウェブページ・アーカイブを活用した政府各機関ホームページの長期的保存により、国等の有するコンテンツの利用機会の拡大と保存を図るため、同図書館も参加した連絡会議を設置し、アーカイブの構築や公開に関するルールの明確化など、同図書館への協力体制を2004年度中に確立する。また、同連絡会議の場において、国立国会図書館で検討しているアーカイブの統合ポータルサイトとの連携のあり方についても検討する。 デジタルアーカイブ化及び国内外への発信:デジタルアーカイブ化の推進(内閣府、総務省、文部科学省、経済産業省及び関係府省)2005年度までに、放送・出版、映画等のコンテンツや、美術館・博物館、図書館等の所蔵品、Web情報、地域文化、アジア諸国との関係に関わる重要な公文書等について、デジタル化・アーカイブ化を推進し、インターネットを通じて国内外に情報提供が行われるよう必要な措置を講ずる。
IT政策パッケージ- 2005 (平成17年2月24日 IT戦略本部)	政府のデジタルコンテンツのアーカイブ化の推進(内閣官房及び全府省):国立国会図書館におけるネットワーク系電子出版物の収集やデジタルアーカイブの統合ポータルサイトの構築等の取り組みを活用し、国等の有するデジタルコンテンツのアーカイブ化を一層強化するため、デジタルアーカイブの推進に関する関係省庁連絡会議において、政府等のデジタルアーカイブ構築・運用に関する基本方針を2005年中に策定する。

図表 5-1 e-Japan 戦略関連資料に見るウェブ情報アーカイブの記載

(2) 総務省:ウェブ情報のアーカイブ化促進に資する技術の開発・実証

総務省は、平成 16 年度及び平成 17 年度において、ウェブ情報のアーカイブ化や利活用を促進するための技術・仕組みに関する実証実験事業を行っている。国立国会図書館とも連携しており、主に以下の観点からの実証実験を実施している。

- ウェブ情報アーカイブの組織化及び大規模アーカイブの保存機能の開発・実証
- ウェブ情報アーカイブ間の連携・横断検索のための汎用的技術の開発・実証
- ウェブ情報の収集・保存・検索するための汎用的なメタデータ等の確立



図表 5-2 ウェブ情報のアーカイブ化促進に資する技術の開発・実証イメージ
 (出典:ウェブ情報のアーカイブ化促進に資する技術の開発・実証説明資料(総務省))

(3) 国立国会図書館:インターネット資源選択的蓄積実験事業(WARP)

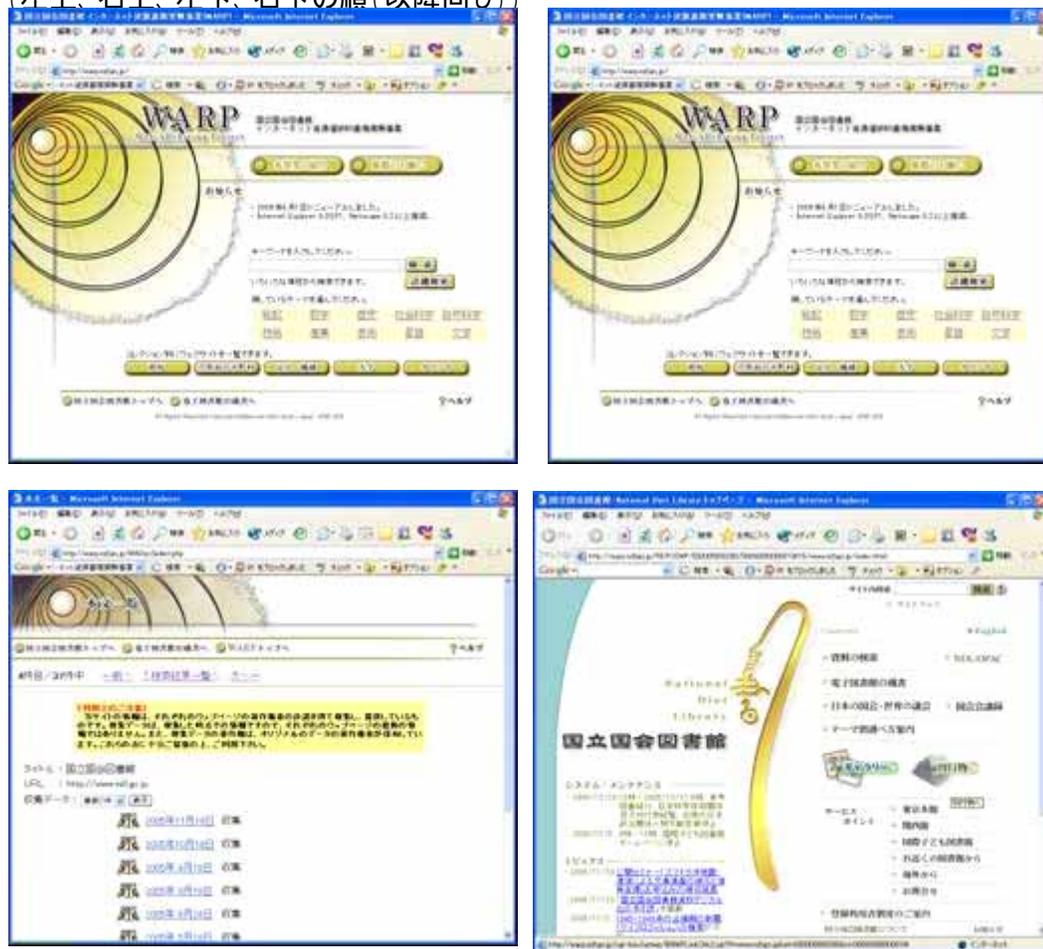
WARP は、インターネット資源選択的蓄積実験事業(WARP:Web ARchiving Project)の略称で、国立国会図書館が平成 14 年度から実施しているウェブ情報アーカイブの実験事業である。インターネット上にある有用な情報資源の多くが、頻繁に更新・削除され日々失われていくことを問題視し、ウェブ情報を文化資産として将来の世代のために保存していこうという趣旨で実施されている。

ウェブ情報の収集対象としては、中央省庁、地方自治体、大学等、公的機関のホームページの他、電子雑誌の収集も合わせて行っている。また、情報の収集に際しては、選択もしくは推薦されたウェブ情報について、WARP にて技術的な収集の可否等を確認し、当該情報の管理者に収集許諾を得た後、ロボットによって収集を行うという手順を踏んで、活動を進めている。

資源タイプ	コレクション	概数	内容
サイト	政府	約40機関	中央省庁、立法機関のホームページ
	都道府県	協力機関 約1,300	都道府県のホームページ
	合併前の市町村		合併を予定している市町村、および合併協議会のホームページ
	法人・機構		組織変更によりリニューアルされる各種法人のホームページ
	大学		国立大学法人化前の国立大学のホームページ
	イベント		国際的・文化的イベントのホームページ
電子雑誌		約1,500 タイトル (国内)	同一のタイトルのもとに、終期を予定せず、巻次・年月次等の表示を伴って、継続的に発行される電子情報。インターネット上で無料で公開されている電子雑誌。

図表 5-3 WARP の情報収集対象
 (出典:WARP のホームページ <http://warp.ndl.go.jp> 等を参考に作成)

(左上、右上、左下、右下の順(以降同じ))



政府、合併前の市町村、法人・機構、大学、イベントの別を入力
 対象組織を選択
 対象日付を選択
 アーカイブされたウェブ情報を表示

図表 5-4 WARP のサイト画面イメージ

(4) 国立国会図書館: インターネット情報の収集・利用に関する制度化

日本国内で発行されるネットワーク系電子出版物を納本制度に組み入れることについて、国立国会図書館長から納本制度審議会に対して平成 14 年 3 月に諮問が行われ、これに対する答申が平成 16 年 12 月になされた。

納本制度とは別の制度による必要がある等とする本答申を受けて、国立国会図書館では、平成 17 年 4 月に「インターネット情報の収集・利用に関する制度化の考え方」を作成し、意見募集を行った上で、平成 17 年 6 月に改訂版を公表した。

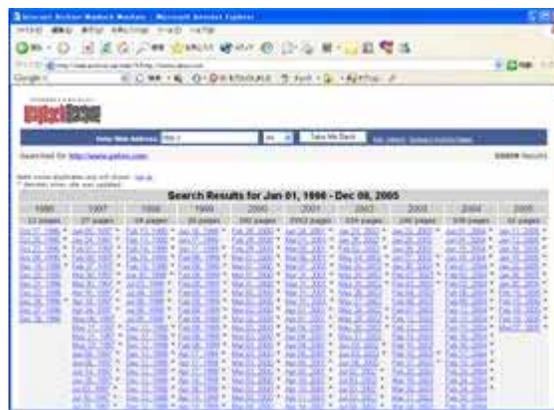
5-1-5 諸外国の参考事例

アメリカのインターネット・アーカイブ、MINERVA、オーストラリアの PANDORA など、諸外国ではウェブ情報のアーカイブが既実践されている状況にある。国際インターネット保存コンソーシアム(IIPC)も設立され、今後は国際間の技術協力及び標準化を行いながら、アーカイブ間の連携の方向にむかっていくものと考えられる。

(1) アメリカ:インターネット・アーカイブ(<http://www.archive.org>)

1996年(平成8年)に設立されたサンフランシスコのNPO団体インターネット・アーカイブ社が、すべての人智への開かれたアクセスという理念で、世界上のウェブ情報の収集・保存・公開に取り組んでいる。

登録拒否しているサイト以外の全ての公開サイトを対象とし、クローラーによるバルク収集を実施しており、保存しているページ数は2005年(平成17年)時点で400億ページに達している。保存した情報は、公開サイト「Wayback machine」上で、誰でも自由にアクセスが可能となっている。また、動画、音楽、書籍のアーカイブにも積極的に取り組んでいる。



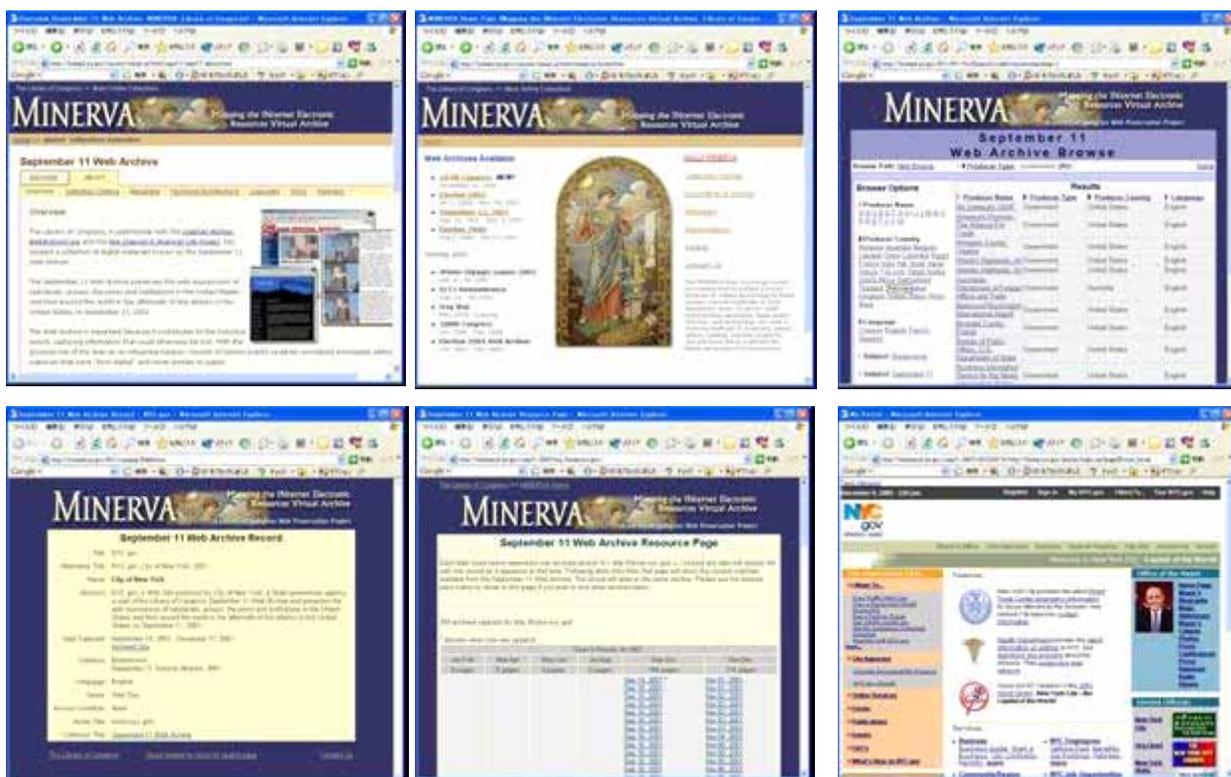
閲覧したい URL を入力
 閲覧したい日付を選択
 アーカイブされたウェブ情報を表示

図表 5-5 Way back machine のサイト画面イメージ

(2) アメリカ:MINERVA (http://www.loc.gov/minerva)

アメリカ議会図書館は、9/11 テロ、イラク戦争、オリンピック等のテーマ毎にウェブ情報の選択収集を行っている。プロジェクト名の MINERVA は、Mapping the INternet Electronic Resources Virtual Archive の略である。

同図書館がテーマ毎に選択基準を満たしたサイトを選定、サイトの管理者と許諾協定を結び、収集を行っている。収集した情報については、図書館内での閲覧が基本であるが、一部テーマについてはサイト上での公開も行っている。



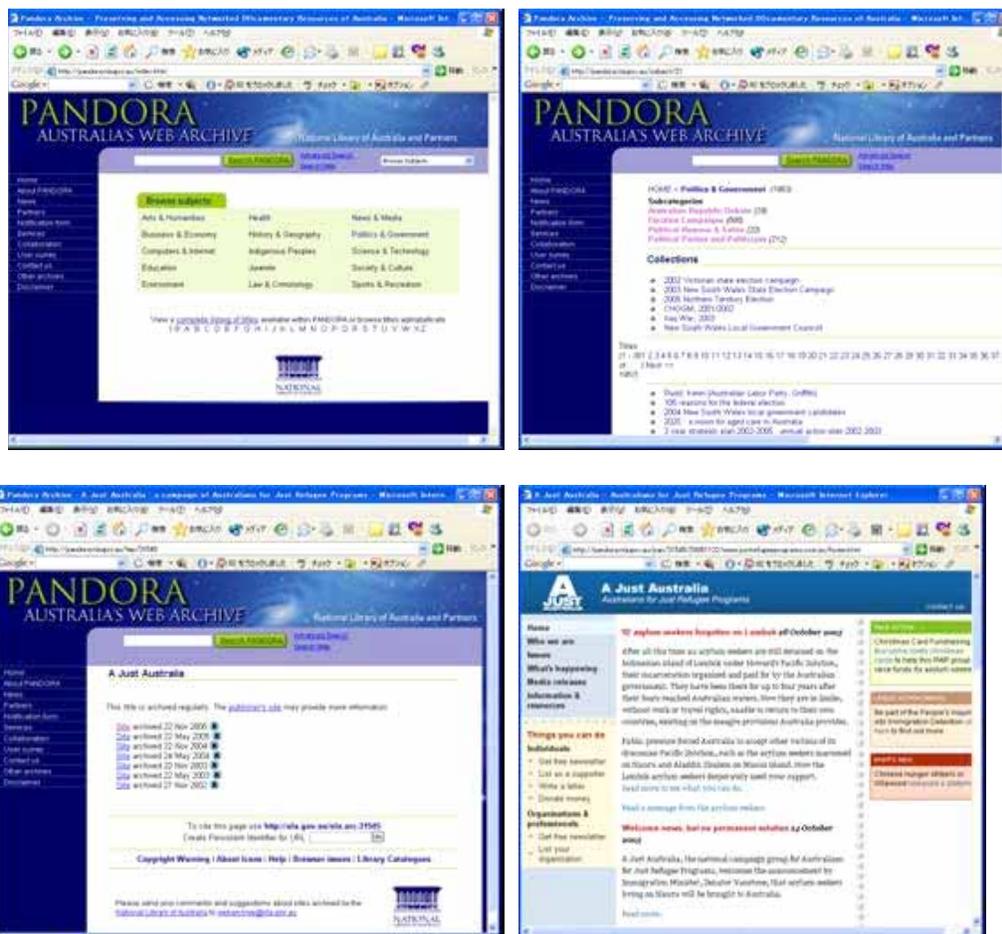
- 閲覧したいテーマを選択
- テーマの概要を表示
- 作成者等、アーカイブの分類を選択
- アーカイブの記述情報を表示
- 閲覧したい日付を選択
- アーカイブされたウェブ情報を表示

図表 5-6 MINERVA のサイト画面イメージ

(3) オーストラリア:PANDORA(<http://www.archive.org>)

オーストラリア国立図書館は、PANDORA というプロジェクト名で、オーストラリアに関する情報やオーストラリアの社会・文化・科学・経済等に関するオーストラリア人の著作などを対象とした、ウェブ情報の選択収集事業に取り組んでいる。

同図書館が、国家にとって重要なサイトとしての基準に沿ったサイトを選択し、サイト管理者の許諾を得た後、州立図書館等とも連携しながら、収集活動を行っている。また、収集した情報については、ウェブサイト上で公開しており、自由にアクセスすることができる。



検索もしくは分類を選択
 閲覧したいウェブ情報を選択
 閲覧したい日付を選択
 アーカイブされたウェブ情報を表示

図表 5-7 PANDORA のサイト画面イメージ

なお、国立図書館は、PANDORA プロジェクトにおけるオンライン政府出版物の特定が、AGLSメタデータ標準の策定によって容易になったとの見解を示している。

AGLS メタデータ標準(The Australian Government Locator Service Metadata Standard)は、インターネット上の情報を容易に検索・利用できるようにするために国立公文書館が策定したもので、ダブリン・コア・メタデータ・エレメント・セットに基づき、一部拡張して策定されている。ダブリン・コアをベースとした理由については、メタデータの作成と実装の容易さが第一の目的とのことである(大部分のダブリン・コア・メタデータは、文書作成時に作成が可能)。

AGLSメタデータは、ウェブページ上の公文書等を長期保存目的で取り込む際に公文書等自体と共に取り込むべきメタデータの一部となっている。

(4) オーストラリアのイントラネット上のウェブ情報収集

オーストラリア NAA では、ウェブ情報の移管のターゲットにイントラネット上のウェブ情報を含めて考えている。

すなわち、国立図書館との棲み分けもあり、イントラネット上の情報やウェブ上の事務処理に係る公文書等は、利用者や配布先が限定されるという理由によって出版物ではないという位置づけをして、国立図書館ではなく、NAA が移管・保存の対象としているのである。一方、国立図書館の側でも、PANDORA プロジェクトで収集するウェブ上の出版物を非常に限定的に選択しているようである。

なお、NAA のウェブアーカイビング・ポリシーは、政府機関のウェブ(公開用、機関内部用ともに)上に掲載される情報は公文書であり、各機関はその公文書を特定・捕捉し、必要とされる期間保存しなければならないという原則に基づいて、ウェブ上の公文書は公文書最終処分規準に従って捕捉・保存・最終処分が行われなければならないというものである。

(5) 国際インターネット保存コンソーシアム(IIPC)

インターネット上の知識及び情報の収集、保存、活用に向けた国際交流及び協力促進の目的で、国際インターネット保存コンソーシアム(International Internet Preservation Consortium)が、2003年に設立された。活動目標は、主に以下の3点である。

- 世界中のインターネット情報の大量収集と、時間を越えた長期保存の可能化
- インターネット情報保存に必要な共通ツール、技術、標準の開発及び活用の促進
- 世界の国立図書館がインターネット情報保存に取り組む際の支援

フランス国立図書館の主導の下、オーストラリア、カナダ、デンマーク、フィンランド、アイスランド、イタリア、ノルウェー、スウェーデン、イギリスの各国立図書館、米国議会図書館、インターネット・アーカイブ社が参加しており、日本の国立国会図書館は、現在のところ未参加であるが、今後ウェブ情報アーカイブの国際連携を図っていく上で、重要な位置を占めると考えられる活動である。

5-2 ウェブ上の公文書等の種類

国の機関が作成等するウェブ上の公文書等は、インターネット経由で提供している国民向けのホームページと、霞が関 WAN 及び各府省庁内イントラネット上で運用している政府の内部向けページに、大きく分けることができる。

5-2-1 国民向けのホームページ

平成17年度に行われた移管基準の改正により、広報紙・PR用パンフレット、ポスター、ビデオ等の広報資料が新たに移管対象に加えられた。今後は、主にホームページ上で提供されている保存期間1年未満の広報資料を移管対象とするか、また、移管対象とする場合どのように移管するか等が検討されなければならないだろう。

(1) 重要性

インターネット上で提供されている、各府省のホームページ上の情報については、国民に行政に関する重要な情報を伝える役割を持っている。

(2) 運用の現状

第2回研究会での内閣府の報告によると、部局で公表される情報のほとんどがホームページに掲載されているとのことである。特に、内閣府の場合、特命大臣ごとにページを持って掲載しているとのことである。

ホームページに掲載する情報の内部決裁については、掲載の内容、掲載期間等について担当部局による決裁が必要となるとともに、官房(広報担当)による審査も行われているとのことである。このような取扱い、内部での決裁処理は他省庁でも同様と思われる。

また、ホームページの保存の現状としては、府省庁側は特段アーカイブの措置を取っていないようである。背景としては、ホームページ上の情報は適宜更新されること、ホームページ上で公表される情報は別途内部で紙媒体により文書化(審議会の報告、報道発表、記者会見録等)していること、全てを府省庁側で保存するのは保存容量面から困難なこと等が挙げられる。

一方、国立国会図書館が、「インターネット資源選択的蓄積実験事業(WARP)」におけるウェブ情報収集の取組みの中で、各府省庁のホームページの保存についても一部実施しており、また、「インターネット情報の収集・利用に関する制度化の考え方」(改訂版)により、将来的なインターネット情報の収集・保存に向けた考え方を示している状況にある。

5-2-2 政府の内部向けページ(霞が関 WAN 上)

(1) 移管・保存の意義

霞が関 WAN 上では、文書交換システムや電子メールシステムと並んで、ウェブシステムである霞が関 WAN インフォメーションボードが、アプリケーションサービスとして提供されている。

霞が関WAN上の電子掲示板や法令協議等には、長期的に保存すべき価値をもつ重要な公文書等と認められるものが含まれている可能性もあると思われる。

(2) 運用の現状

現状は情報提供期間経過後(例:掲示板の場合、掲載時に掲載期限が設定されるようである)に自動的に削除される形となっており、削除後の情報がアーカイブされることは特になく、削除された情報は散逸してしまうおそれがある点が課題である。

5-2-3 政府の内部向けページ(府省庁内イントラネット上)

(1) 位置づけ

府省庁内イントラネット上で提供されている情報については、府省庁により多少の違いはあると思われるが、研修情報、福利厚生情報、会計情報等、基本的には内部職員向けの事務情報が主であるとのことである。

(2) 運用の現状

府省庁内のイントラネットについては、各府省庁にその敷設、利活用に関する権限が委ねられており、ウェブ情報についても、提供されている情報、利用者(本省及び地方支分部局)、システム環境(システム構成やデータベース設計)は多様である。そのため、移管・保存の状況についても各府省庁により様々であるものと想定される。

5-3 ウェブページ保存における諸課題

ウェブページの保存を実施するにあたり、解決の求められる諸課題を整理する。総じて、ウェブページの保存、殊に公開ウェブ上の情報の保存は、先に事例を示したとおり、既の実施されている仕組みである。そのため、技術的な課題はあるものの、今後の課題としては、保存価値という観点から、収集の頻度や対象といったものをどのように定めるかといった点が中心となる。

5-3-1 ウェブの動的要素に起因する課題

(1) ファイル転送

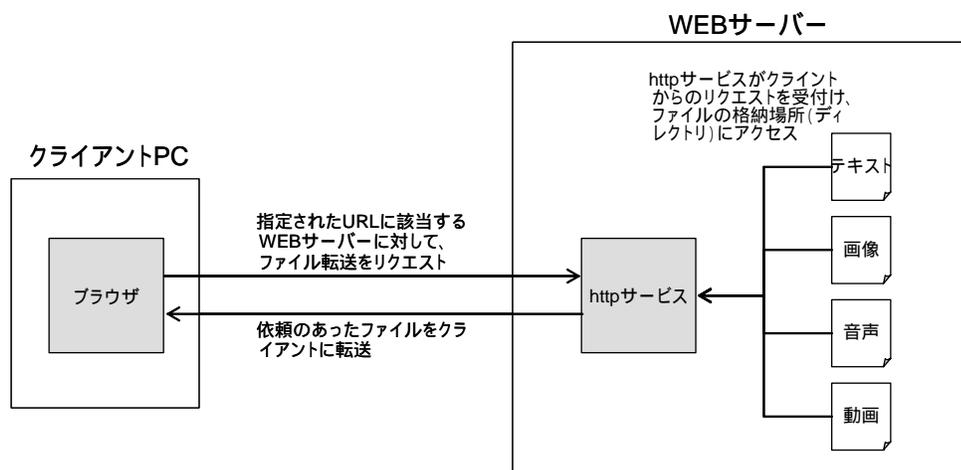
http によるファイル転送を行う場合、ユーザがクライアント PC のブラウザソフトでホームページのリンクを押すと、ブラウザは指定された URL に該当するウェブサーバの http サービスに対してファイル取得のリクエストを発行する。

http サービスは、リクエストを受け付けると、サーバ上のディレクトリに格納されたテキスト、画像、音声、動画等のファイルにアクセスし、該当するファイルが存在する場合には、当該ファイルをブラウザへと転送する。

その後、ブラウザは、サーバから転送されたファイルを表示するが、そこまでの流れを踏まえてはじめて、ユーザはネットワークを介したファイル視聴を行うことが可能となる。

また、その場合、テキスト、画像、音声、動画等といったファイル種別の違いは、ファイル転送の仕組み上は特に意識されず、汎用的に対応可能な仕組みとなっている。

ファイルの種別は、クライアント PC 側で認識されており、ユーザが転送されたファイルを開こうとすると、そのファイル種別に適したビューワーソフトが勝手に立ち上がる。この仕組みによりユーザはファイル種別を意識せずにファイル視聴することが可能となる。



図表 5-8 ファイル転送のイメージ

(2) まとめ

ICA 報告書 16 においては、「長期保存記録と廃棄処分記録との間の関連付けを慎重に検証する」必要がある事を指摘している。なぜならば、「新しい情報システムでは、リンク情報が多く含まれ、リンク情報が削除されてしまうと、長期保存記録の真正性と利用可能性が著しく損なわれてしまうからである」。また、「ウェブサイトの変化の速度は速く、文書間の関連付けや連関は明示的に維持されなければならない。多くのウェブサイトには従来の文書形式ではほとんど表示できない、読者との相互作用が可能な要素や、文書の集合体でなく、データベースから引き出される要素も含まれていることが多い」とウェブ情報が単独で存在せず、他のページあるいは様々なプログラムに依存することによる課題を提起している。

研究会でも、ウェブページのリンク切れや特定ソフトに依存した動画がページに埋め込まれているといった場合への対応が求められるとの指摘があった(第4回)。

なお、オーストラリア NAA が開発した Xena においても、現在 html 用のプラグ・イン・ソフトは、HTML ページを単独で標準化(ノーマライズ)する機能しかなく、関連づけられている画像等のリソースを HTML ページと一体的に標準化することはできないようである。

以上からも、ウェブサイトのリンク情報や動的な機能性の保存は複雑な問題であり、ウェブアーカイブを実施するにあたっては、政府サイトの現状を調査し、適切な対応を取っていく必要がある。

5-3-2 安定した識別子の必要性

(1) 必要理由

ウェブ上の文書にアクセスする場合、通常は URL を用いることとなるが、URL というのはあくまで場所に対してつけられた名前に過ぎないことから、URL 保存だけでは、ウェブアーカイブとしては不十分であるとの指摘もある。すなわち、URL の場合、内容と識別子が必ずしも一致しないので、内容を識別するための何らかの仕組みがなければ、真正な内容の保存を長期にわたって行うことはできない。

(2) URN(Uniform Resource Name)について

研究会(第5回)でも、出版業界で用いられている「Digital Object Identifier」のような一意の識別子を、行政のウェブ情報についても振ることができればよいのではないかという委員からの指摘があったが、何らかの永続的な安定した識別子を振ることが必要になる可能性がある。

ここで、インターネット上の情報資源の保存場所を管理するのではなく、資源それぞれに一意的な名前を付与することで、情報を永続的に識別しようとする考え方である URN(Uniform Resource Name)を紹介する。

ア. 概要

URN は、RFC(Request For Comment: インターネットに関する技術標準を定める団体 IETF が正式に発行した文書)2396 によりルール化されている。

基本的なスキーム: <URN> ::= "urn:" <NID> ":" <NSS>

<URN>: URN スキームによる記述であることを示す。

<NID>: 名前空間を示す。

<NSS>: 名前空間に固有の、リソースを識別する文字列。

イ. 名前空間について

情報資源に付与される名前は、複数存在する名前空間のいずれかに属することで、その一意性を担保される。上記<NID>の部分を IANA (Internet Assigned Number Authority: インターネット上で利用されるアドレス資源の標準化や割り当てを行なっていた組織) 登録することとなり、<NSS>の部分をその登録主体が管理することで、一意性が担保される。

名前空間については、現時点では複数の標準があり、統一はされていない。名前空間の例としては、OASIS 空間(OASIS: Organization for the Advancement of Structured Information Standards: ビジネスにおける情報交換のための技術標準を策定する国際的な非営利組織)、公開識別子空間、IETF 空間(IETF: Internet Engineering Task Force: TCP/IP などのインターネットで利用される技術を標準化する組織)、ISBN 空間(ISBN: International Standard Book Numbers: 書籍のための国際識別コード)等がある。

ウ. 概観

URN は現状では、まだ発展段階の概念である。新たに管理コストが生じるため、手軽さを特徴として発展してきたインターネットの考え方と相反している点が課題である。

今後、ウェブ上の公文書等の移管への適用を考えた場合、現状では管理運用の方法が未定のため、適用が難しい。仮に、国の機関がホームページを作成した段階から当該ページを公文書等として管理する、もしくは国立公文書館に移管した後に国立公文書館所蔵資料の整理体系に合わせて整理するという運用によりウェブ情報を一定の体系に合わせて管理することが可能ならば、適用もあり得るであろう。

5-3-3 収集上の課題

研究会(第2回)において、委員から、公文書館制度に基づく移管・保存という観点からは、組織として重要な公文書等は、できるだけ抜け落ちがないように収集すべきではないかとの指摘があった。

ウェブ情報の収集方法については、選択収集とバルク収集があるが、いずれにせよ、重要度の高いものを漏れなく厳選して収集・保存するという点から、収集方法を選択する必要がある。

(1) 収集方法(選択収集とバルク収集)

選択収集は、収集対象とするウェブページを職員が選定し、サイトの管理者に承諾を取った上で、職員が手作業で(もしくは自動収集ロボットによって)収集作業を実施する方法である。

職員の選定作業が介在するので、収集したウェブページの収集単位は必然的に細くなり、また承諾を取ることでサイト管理者の理解も得られやすいのが長所である。逆に、選定作業に人手がかかり、収集したウェブページが限定的であるという短所もある。

一方、バルク収集は、収集対象の選定を行わず、クローラーと呼ばれる自動収集ロボットによって、全てのサイトを網羅的に集めてくる方法である。

集まるサイトの網羅性は高くなり、クローラーまかせなので人手がかからないというのが長所であるが、サイト管理者への周知等への配慮が必要である。

	概要	ページの質	ページの量	人手	承諾
選択収集	職員が対象を選択して収集	収集単位が細かい	限定的で少ない	人手がかかる	一般的に承諾を得たもののみ収集
バルク収集	クローラーが自動で網羅的に収集	収集単位が大きく網羅的	網羅的で多い	人手がかからない	一般的に無承諾の場合が多い

図表 5-9 選択収集・バルク収集の比較

(2) 収集対象

収集を行う目的、収集の承諾有無、技術的な収集可否、保存する際の容量といった問題から、収集のターゲットをどこに据えるかをできるだけ明確にしておくことで、運用の負担が軽減されることとなる。

一般的には、絞り込みの観点としては、対象とする機関、ドメイン、ファイル形式、プロトコルなどが挙げられる。

収集対象検討の観点	収集対象(案)
対象機関	基本的に国
ドメイン	一般国民向けに公開されたページ(インターネットのgoドメイン上のページ)及び行政の内部向けに作成されたページ(各府省庁内のページ、霞ヶ関WAN及びLG-WAN上のページ等)
ファイル形式	ウェブ(html等)、文書(pdf等)、静止画、動画、音声等
プロトコル	http,https,ftp等、公衆送信プロトコル中心

図表 5-10 収集対象検討の観点

機関及びドメインについては、公文書としての収集対象となる行政のホームページが、一般国民向けに公開されたページと、行政の内部向けに作成されたページ(各府省庁内のページ、霞ヶ関WAN及びLG-WAN上のページ等)に分類できる点に留意が必要である。

ファイル形式については、基本的に全てを対象とするのが望ましいが、データベース等のいわゆる深層ウェブやストリーム形式については技術的に難しい部分もあり、優先順位を下げるのがよいと思われる。

プロトコルについては、まずはウェブページとして一般的なhttp、https、ftpといった公衆送信プロトコルとするのが適当でとおもわれる。

(3) 収集頻度

研究会(第1回)でも、委員よりウェブページが頻繁に更新されるという特質を持つこと、どの時期に移管するのが課題でないかという指摘があった。

ウェブページ保存の価値を高める意味では、収集頻度が高い方がのぞましい。とはいえ、ウェブページの収集にあたっては諸々のコストが発生するため、収集の目的に応じて適切な収集頻度(一定期間毎に収集を行う間隔)を設定する必要があると思われる。

分類	概要
保存容量の増加	収集したWebページは基本的に累積保存することとなるため、収集頻度を高めるに従って保存容量が増加し、ハードディスク等のコストが増すこととなる
収集作業・処理の高速化	収集頻度を高めるためには、収集が手作業であれば作業速度の向上と作業回数の増加により負担増となり、自動収集であればクローラーの処理性能向上を図る必要が生じる
サイト側の負担増	収集頻度を高めるため、サイトへのアクセスが増加すると、サイト側の処理負担を高めることとなる。サイトによっては、ネットワーク業者との契約がアクセス頻度によって支払額が変動する形態となっている場合も有り、コスト負担に直結する場合も生じ得る

図表 5-11 収集頻度を高めるにあたり生じる主な問題点

国立公文書館への歴史公文書等の移管は、現在年に一度である。だが、ウェブページの更新は日々行われているため、年に1度の移管では不足とも考えられる。制度的には年に1度としつつ、実際の収集作業の頻度は年に複数回設定することも可能であろう。また、保存の価値を高めるためには収集頻度を高めたいが、収集コストの増加との二律背反が生じる。今後はウェブページ保存の費用対効果を考慮しながら、然るべき収集頻度を設定する必要がある。

(4) オーストラリア NAA における収集

NAA における取得の頻度は、政府機関ごとに異なるようである。基本的には、ウェブ上における政府機関の活動に関するリスクの評価に基づいて、取得頻度が決められるようである。また、情報の内容(国の重要施策に関すること、国民的関心の高い行事・事件・事故等)に基づいて行う等の選択肢も設けられているとのことである。

5-3-4 管理上の課題

(1) メタデータの標準化

ウェブ情報アーカイブ間の横断検索を将来的に考慮した場合、メタデータの標準を揃えておく必要がある。

これについては、ウェブ情報独自のメタデータを用意するのではなく、その他の文書と同様のメタデータ体系とすることが望ましい。

その際には、その他の文書と同様、記録管理メタデータ、アーカイバル・メタデータ、技術的メタデータといった要素に留意する必要がある。

(2) 時系列での管理

ウェブページの保存を行う場合、保存する情報が累積されていくこととなるため、同一のサイトの情報を時系列で管理していく必要が生じる。つまり、データベースでの管理の際、サイトの URL と日付をキーとして管理することとなる。

また、リンク情報への加工も必要となる。これは、表示したウェブページ上のリンクを単純に押下すると、保存してある情報ではなく現在のサイトのページへとジャンプしてしまうためである。

	URL
通常のURL	http://yahoo.com/
インターネットアーカイブ内のURL	http://Web.archive.org/Web/20050307212714/http://yahoo.com/

図表 5-12 インターネットアーカイブの保存ページの URL 例

国立公文書館における保存にあたっては、時系列での管理のほか、保存した情報内におけるリンクの完結の仕組みが求められる。米国のインターネットアーカイブのような、ウェブ情報アーカイブを運営しているサイトについては、基本的に同種の構造を保持しているため、それらを参考とすることが考えられる。

5-3-5 保存上の課題

(1) 大容量保存対策

平成 15 年版情報通信白書(総務省)によれば、平成 14 年末における我が国のインターネット上のコンテンツの総データ量(JP ドメイン)は約 10 テラバイトに及びとのことである。これらを時系列に従い累積保存していく場合、その保存容量は途方もない大きさととなる。

技術的には、大容量保存の実現は可能であるものの、ストレージの容量を維持し、処理スピードを保っていくためには、当然それに見合ったコストが発生することとなる。

そのため、先に示したウェブページ保存の対象範囲や収集頻度の絞り込み、重複ページ保存の回避、コンテンツのアクセス頻度による処理系統の差別化等といった対策を行い、コスト管理にも配慮していく必要が大きい。

国立公文書館においても、この問題は同様であり、保存の必要のあるウェブページを見極め、必要かつ十分な保存容量を持つシステムとすることが重要である。

(2) 長期再現性の確保

ウェブページは当然何らかのツールでの表示・再現を前提に作成されているわけだが、長い年月の間にそのツールの仕様がバージョンアップによって変更されたり、もしくはツールそのものがなくなってしまうことも考えられる。

例えば、マイクロソフト社の Internet Explorer の特定バージョンでの表示に合わせて作成されたページは多い。今後 Internet Explorer がバージョンアップした際に上位互換としなかった場合や、Internet Explorer そのものがなくなった場合、当該ページを適切に表示する術はなくなってしまう。

その場合、情報のエミュレーションやマイグレーション等、何らかの対策を講じていく必要が生じる。

6. 電子公文書等の適切な移管・保存・利用に向けて

6-1 適切な移管の実現に向けて メタデータの標準化

「コンテキスト(背景・状況・環境)、内容、構造及びある期間の記録の管理について説明したデータ」であるメタデータ(タイトル、作成者名、作成日、利用履歴)を電子公文書に付与する必要がある。

メタデータの付与については、電子公文書等が移管される前に、自動又は職員の操作により容易に付与されるようにすることが望ましい。また、用語集の整備等により、メタデータの記述内容の標準化を可能な限り図っていくことが必要であろう。

6-1-1 記録管理メタデータ付与への対応

記録管理メタデータについては、現在の行政文書ファイル管理簿上の分類項目を基盤とすることが可能であると考えられる。そのため、記録管理メタデータは、行政文書ファイル管理簿と同じように、各府省庁側で付与することが望ましいといえよう。今後は、メタデータ用語集等を整備することにより、標準化を進めていくことが期待される。

6-1-2 アーカイバル・メタデータ付与への対応

ISAD(G)のようなアーカイバル・メタデータには、公文書等の長期保存のために求められるメタデータ項目が多く含まれるものの、第4章で調査したとおり、記録管理メタデータと共通するものも少なくない。

今後は、記録管理メタデータと共通するものについては、それらとのデータの共有化を図りつつ、アーカイバル・メタデータの付与については、各府省庁で共通するものは、記述用語集等の整備等により、各府省庁による付与を促し、それ以外のは国立公文書館側で付与することが考えられる。

6-1-3 技術情報メタデータ付与への対応

技術情報メタデータは、恒久的な可読性の確保や国際標準の観点からも、メタデータとして適切な定義を行った上で付与する必要があるが、現状の行政文書ファイル管理簿のメタデータに対して、そのまま技術情報を加えることはできない。

理由としては、現在の行政文書ファイル管理簿は、情報公開への対応等を目的に作成されており、行政文書の集合体である行政文書ファイル単位で管理されているが、この行政文書ファイルは、技術情報メタデータを付与すべき技術的な管理の単位と異なっていると考えられるためである。

(1) 対応方針

移管対象となる電子公文書等に対しては、行政文書ファイル管理簿のメタデータに、文書一覧と技術情報(ファイル形式)といった項目を加えることが考えられる(下図)。行政文書ファイルの中に含まれる文書一覧を管理することで、各府省庁側における情報公開の迅速な対応にも効果を発揮すると思われる。

大分類	中分類	小分類	ファイル名	作成者	-	文書一覧	技術情報
AAA	AA	A	ファイルA			文書1	PDF, xx
						文書2	PDF, xx
						画像1	JPG, xx
						映像1	MPEG-2, xx

図表 6-1 技術情報を管理するための行政文書ファイル管理簿のメタデータ

(2) 付加項目

ダブリン・コアやアメリカの事例を見ても分かるとおり、技術情報メタデータには、ファイル形式に加え、保存メディア(媒体)、データサイズ、作成・保存時のハードウェア、OS、アプリケーションのバージョンなど、多様な項目が含まれる。

しかし、現時点において、我が国においては、技術情報メタデータが整備されていないことから、データ項目をできるだけ絞り込んだ上で稼働させることが望ましい。そういった点から、当面は上図のとおりファイル形式のみを対象とすることが考えられる。

6-1-4 メタデータ付与の際の課題

電子公文書等の移管、保存を適切かつ円滑に行うためには、主に政府側の各府省庁(各課)の文書作成者及び文書管理担当者が文書作成後速やかにメタデータを付与し、移管することが望ましい。今後は、メタデータ用語集の策定や既存の文書管理システムの活用などにより、メタデータ付与への対応によって発生する作業負荷が過大なものにならないように配慮する必要がある。

(1) メタデータ用語集の策定

メタデータ構造の定義に加え、例えば文書の分類(Subject)等については、辞書として一覧を定義しておき、その中から選択して入力することでデータベースがより使いやすくなる。オーストラリアでは、「Keyword AAA」として、各府省庁の用語辞書と融合が可能な包括的な用語集を作成し、幅広い行政活動や媒体に適用が可能となっている。

なお、ある省では、辞書的なものを提供して職員の入力を支援している。すなわち、メタデータの内容が予めシステムに組み込まれており、プルダウンで選択できる。また、登録されていないメタデータについては各課内で登録することも可能とのことである。

今後、メタデータ用語集を政府全般に適用する場合は、各府省から文書分類表などの情報を提供してもらい、国立公文書館などの機関が用語集の整備をサポートすることも考えられる。

(2) 文書管理システムの記述情報の出力

現在、行政文書の記述情報は各府省庁の文書管理システムで管理されている。管理方法は各府省庁によって、リレーショナルデータベースや Notes データベースなど異なっ

いると想定されるが、電子公文書等の移管時には、共通形式でメタデータがやり取りされる必要がある。XML はインターネットの標準として W3C より勧告されたものであり汎用性が高いため、文書管理システムから出力される記述情報は XML データ形式とすることが適切であろう。

6-2 適切な移管の実現に向けて 移管対象の評価選別

公文書等がカバーする範囲は広範であるとともに、ウェブページや電子メールなど、その記録様式も非常に多様である。公文書館制度においては、国立公文書館の使命は、公文書等のうち、将来の国民に受け継いでいくべき価値があるもののみを選別した上で、それを移管・保存し、利用に供していくことにある。

このような認識に基づいて、電子公文書等のうち、移管・保存の対象となるものをどのように評価選別するかについて、電子公文書等の保存等の長期的安定性・効率性等の観点を踏まえて、検討した。

6-2-1 評価選別方法に関する考え方

(1) 評価選別時期

研究会では、既に作成された電子公文書等のうち、移管・保存の対象となるものについては、それを可能な限り早期に集中管理の下に移送し、長期保存上の措置を講じることが望ましいとし、そのためには、そのような措置を講じる対象となる電子公文書等を特定する必要があるとした。そのためには、電子公文書等の作成から一定期間経過後、または作成時、更には作成前に、移管・保存の対象となるものが評価選別されていなければならないとの考え方が示されている。

今回の調査の結果でも、各種システムやファイルの標準化や公文書等の集中管理の必要性、ライフサイクルにおける早期の対応の必要性等が明らかになっており、評価選別は、当面は作成後速やかとしつつ、中長期的には当該選別が電子公文書等の管理システムの一機能として盛り込まれることで、作成前からの評価選別が可能となる環境が整うことが望ましい。

(2) 評価選別方法

個々の文書を作成する前から電子公文書等の評価選別を可能にする方法としては、カナダ、オーストラリア等で行われている、マクロ評価選別法の考え方を取り入れることが考えられる。これは、組織が行っている業務というものを、一番大きなところで、「機能」、「その機能を構成している活動」、「その活動を構成している事務処理」といった三段階に分け、その重要度に応じて、それら「機能」・「活動」・「事務処理」の過程で作成される公文書等の評価選別を行うというものである。つまり、公文書等自体よりも、その公文書等が作成されるコンテキスト(文脈・背景)である「機能」・「活動」・「事務処理」を評価選別していくという考え方である。

このようなマクロ評価選別法による移管基準の精緻化及び保存期間満了前の評価選別の実施にあたっては、国立公文書館が評価選別に必要なマニュアル等を作成する一方、各府省も自らの業務等について国立公文書館に必要な情報を提供するなど、緊密な協力をを行うことが必要となるのではないかと考えられる。

6-2-2 ウェブページの扱い

歴史資料として重要な公文書等と認められるものは、インターネット及びイントラネット上のものについても、国立公文書館が各府省から移管を受けることは、国立公文書館法の規定に照らして、当然である。また、移管・保存の対象となる公文書等の評価選別は、媒体や記録様式等の如何を問わず、原則として同一の基準に基づいて行われる必要がある。

ただし、今回の調査で、ウェブには一般的な電子公文書等とは構造的に異なった面があること、また、ウェブ上の公文書のみを考える場合にも、イントラネットとインターネット上の情報では異なる点があることから、それらを踏まえた対応が必要となる。

(1) ウェブページの種別の対応

ア. 国民向けホームページ

インターネット上で提供されている、各府省庁のホームページ上の情報については、国民に様々な点から関わりのある情報であり、重要なものについては何らかの形で保存されることが望ましい。

ただし、ホームページ上の情報は、情報量が多く、更新頻度が極めて高いこと、またホームページ上からではなく他の方法によって移管することが可能であるものも多数存在することなどを踏まえ、研究会でも問題提起されたとおり、国家的行事や歴史的事件、事故等に関するものから優先的に集めることは、有力な選択肢の一つであると考えられる。

イ. イン트라ネット(霞が関 WAN、府省内イントラネット)

イントラネット上の情報は、法令協議のような政府の政策決定プロセスに関するような重要な情報が数多く含まれていると言われる。それらの中には、歴史資料として移管・保存の対象となる公文書等も含まれている可能性があると考えられる。

一方、霞が関 WAN 上のイントラネット情報は、提供期間経過後(例: 掲示板の場合、掲載時に掲載期限が設定されるようである)自動的に削除されるとともに、削除後の情報はアーカイブされていないことが、今回の調査で明らかとなった。また、各府省庁にその敷設、利活用に関する権限が委ねられており、ウェブ情報についても、提供されている情報、利用者(本省及び地方支分部局)、システム環境(システム構成やデータベース設計)が多様であることも明らかとなった。

このような運用の現状に鑑みて、歴史資料として移管・保存の対象となる公文書等も含まれている可能性があることについて、政府には早急に理解を深めてもらう必要があるのではないかと考えられる。

(2) 収集方法

ウェブページを収集する方法としては、バルク収集等を活用してポータルページ配下の全てのページを収集する方法と、各政策テーマ等に関して関連するページを個別に選択収集する方法を取ることが考えられる。

いずれを採用するにしても、収集する頻度をどの程度にするか、対象をどの層までにするかといった点が論点となるが、現段階では、何階層目まで、何日単位で収集するかといった点について定量的に定義することは難しい。

国民向けに公開されたウェブページに関しては、内閣府又は国立公文書館が収集することが可能だが、イントラネット上にあるウェブページを外部の機関から収集するにはアクセス権等の課題があることから、各府省庁による移管が適していると考えられる。

(3) 技術面での留意点

リンクが貼られているハイパーテキストを移管しようとした場合、ウェブ上の公文書等で画像・音声・動画等が埋め込まれていることがある場合がある。特に画像については、政府の報道発表資料に図解として掲載されている事例が多数見受けられるようである。

これらの画像については、本質的部分として保存されることが望ましいが、埋め込まれている画像はそもそもウェブサーバの先に存在しているものであり、単にリンク情報を保存するだけでは(リンク切れ等により)適切に保存できない可能性がある。そのような場合、別途画像を取り込まなければならない点については留意する必要がある。

また、ウェブ上の公文書の構造が複雑である場合、記録様式が多様である場合、特定のソフトウェアへの依存性が高い場合といったものについては、単に機械的に収集しただけではエッセンスを正しく(長期にわたって)再現できない可能性があり、長期の保存・利用を確保する方法を確立する必要がある。

最後に、ウェブページは更新が早いことから、保存するデータ量の増大が急速に進み、利便性(パフォーマンス)が損なわれる可能性もあるので、バランスが必要とされる。

6-2-3 電子メールの扱い

今回の調査で、海外主要国では電子メールについて分類を設けた上で、事業活動や政策形成の証拠となるもののみを公文書等として管理・保存等の対象としていることが明らかとなった。日本においても、電子メールについては、別途判断ルールを設け、最終的な利活用の視点から適切なものを優先的に保存することが想定されることから、同様のルールを当てはめることになる。

電子メールの保存はあくまで公文書の一部として行うものであることから、ウェブと同様に、収集以降の管理のフェーズについては、その他の電子文書と同じ扱いとなる。

ここで、単にテキストデータのみで構成される電子メールについては特段問題ないが、電子メール上に圧縮ファイルが添付されている場合やデータベースへのリンクが提供されているといった、利活用を踏まえた場合に何点が課題が残っている。これらについては今後技術・運用面からの検証が必要となる。

6-3 適切な移管の実現に向けて ファイル形式の標準化

各府省で作成される電子公文書等のうち、既に作成されたものについては、可能な限り早期に標準化されたファイル形式に変換して保存することが適切であると考えられる。テキストファイルについては、JPEG2000等の画像形式やXML等のテキスト形式への変換が考えられる。一方、ファイル形式の変換が難しい音声や映像ファイルについては、作成段階で複数の推奨ファイルに絞り込むことが必要になるといえる。

6-3-1 ファイル形式に関する考え方

電子公文書等を作成する各府省側で、作成時から一つのファイル形式に統一することは、省庁が依拠する電子公文書等の作成・管理環境が多様であることから、かなりの困難が伴うものと想定される。したがって、作成段階では、一つのファイル形式に統一するのではなく、いくつかのオプションを示す「ゆるやかな標準化」に止まるものと考えられる。一方、側では、電子公文書等の保存の長期的な安定性等を図る観点から、移管(又は移送)を受けた電子公文書等を標準化したファイル形式に変換して保存することが想定される。

しかし、画像、音声、映像等のファイル形式は、幅広く流通してはいるものの一つの長期安定的な標準仕様を採用することは現時点では難しい。したがって、既に作成された電子公文書等については、特定のハード・ソフト等のシステムへの依存度が低く比較的長期に保存が可能と考えられる複数のフォーマットを選択肢として示すことが考えられる。ただし、中長期的には、国内外の標準化動向を注視しつつ、政府及び国立公文書館において、フォーマットの標準化へ向けた取組みの必要性が高まることが想定される。

6-3-2 テキストファイルの標準化

各府省庁側で作成される文書は、MS-Word や MS-Excel 等ベンダ依存性が高いことから、恒久的に可読性を確保するのが容易ではない。そこで、特定のベンダに依存しないファイル形式に変換することが必要となる。

ただし、ある標準化した形式に変換する場合、本章の冒頭でも述べたが、テキストファイルをイメージとして保存するという考え方と、標準化されたテキストデータとして保存するという考え方がある。

イメージとして保存するという考え方に拠るのであれば、JPEG2000 は国立公文書館が既にデジタルアーカイブ・システムで使用していることから、他に選択肢を与えつつも、有力な選択肢の一つとなるであろう。一方、テキストデータとして保存するという考え方をとる場合は、XML への変換に加え、LAC が実際に行っているように、ベンダに依存しているもののデファクト・スタンダード(事実上の標準)となっている MS - Word や PDF のようなフォーマットを選択肢として認めるといことも考えられる。

6-3-3 画像、音声、映像ファイルの標準化

画像ファイル、音声ファイル、映像ファイルについては、文書作成段階から定められたファイル形式であることが望ましい。だが、現時点では一つのファイル形式に絞り込むことが難しいため、複数のファイル形式を推奨形式として選定することが適切であるといえよう。その際、電子公文書等を作成する各府省において、各職員に対する周知が必要となるほか、定められたファイル形式以外のソフトウェアを職員の PC に導入しないような運用上の制限が必要となるだろう。

また、ファイル形式変換技術を用いて、過去に作成された電子公文書等については、標準化されたファイル形式に変換する必要がある。ただし、全てがファイル形式変換技術で文書の再現性を確保できるわけではないため、それらの電子公文書等に対する可読性確保の方法も検討していく必要がある。

6-4 適切な移管の実現に向けて 移管方法

研究会においては、電子公文書等については、当該公文書等の作成後可能な限り早期に、全府省横断的な集中管理の下に移し、ファイル形式の変換や媒体変換等長期保存上の措置を講じることが適切であるとの議論の方向性が示された。

可能な限り早期に長期保存上の措置を講じるため、「移送」という概念が提示されている。これは、国立公文書館法の規定に基づく保存期間満了後の移管を待たずに、各府省の委託等により、保存期間満了前に公文書等を集中管理の下に移すことをいう。保存期間満了前、すなわち、「現用」段階で移送を行うことにより、長期保存上の措置の統一性・安定性・効率性等の確保が期待できる。

また、電子公文書等を紙媒体やマイクロフィルム等非電子媒体に変換して移管することは、(文書構造等)重要なデータが損なわれるリスクがあることから、過渡的措置としては選択肢となり得るものの、好ましくないとの方向性も示されている。

6-4-1 移送と移管

移送とは、保存期間満了前の公文書等を集中管理の下に移すことをいう。電子公文書等の場合は、現用段階で原本が府省庁側にあるが、電子ファイルの複製物を内閣府又は国立公文書館側に移行するものである。この場合、移送を受けた内閣府又は国立公文書館側でファイルフォーマットの標準化等の作業を進めることができる。また、当該公文書等を作成した各府省においても、原本を引き続き業務に利用できるほか、仮に各府省が保管する原本の可読性が失われた場合には、内閣府又は国立公文書館が保管する複製物又はその複製物を長期保存用の標準フォーマットに変換したもから複製物を作り、各府省に提供することも考えられる。

一方、保存期間が満了となり、非現用文書になった時点で、法に基づく歴史公文書等の移管が行われる。移管後は、国立公文書館側にある電子ファイルを原本と位置づける一方、各府省においては、保存期間満了に伴い、廃棄(消去)することとなる。

6-4-2 移管(移送)の方法

各府省で作成された電子公文書等を移管(移送)する方法については、電子データの特性を踏まえて、オンライン移管を行うか、あるいはテープやDVDのようなモノとしての電子媒体によるオフライン移管を行うか、また、オンラインについても、電子データの送信を行うか、管理権限等の移行のみによるか、研究会で議論がなされたところである。

(1) オフライン

「目に見える」モノとしての電子媒体による移送は、各府省及び内閣府・国立公文書館の双方の職員にとってイメージしやすく、導入が容易であると考えられるため、当面の措置としては、適当な方法である。また、技術的にも、各府省側の保存サーバ上から、バックアップ

ソフト等を利用して、テープや DVD 等のモノとしての電子媒体に自動的に保存することができることから、各府省側の負担は大きいとはいえない。

ただし、この方法の場合、移管を行なう側、受ける側とも運搬等物理的な作業が発生するため、一定程度の工数とコストが発生することは避けられない。また、内閣府又は国立公文書館は、移送を受けた後、速やかに長期保存用のサーバ等に電子公文書を書き込み、検疫(ウイルス対策)等の長期保存上の措置を講じることが必要である。

なお、オーストラリアにおける電子公文書の移管では、物理媒体を通して移管された電子公文書について、約 1 ヶ月間にわたりウイルスチェックを行っている。

(<http://www8.cao.go.jp/chosei/koubun/kenkyu/02denshi/05/youshi.pdf>)

(2) オンライン

オンラインによる移管(移送)は、各府省側のインフラ整備がオフラインに比べて大規模になること、オンラインで接続されることでセキュリティ上のリスクが生じること、従来の作業手順が大幅に変わる事等から、早期に実施できるものではないが、移管による工数が小さいことは事実であり、中長期的には効率性等の観点から見て高いのは間違いない。

ただし、移送時のセキュリティや安全性を確保する必要がある、「目に見えない」ので移送の明確なルール化やプロセス管理が不可欠である等、モノとしての電子媒体による移送に比して、克服すべき課題が多い点については留意する必要がある。

6-5 適切な保存の実現に向けて

6-5-1 長期保存上の留意点

長期的な安定性を重視しつつ、基本的には電子公文書等は電子媒体のまま保存することが適当であると考えられる。ただし、可能な限り早期に、一定の集中管理下で長期保存のために必要な措置を講じる必要がある。

電子公文書等の適切な保存環境を確保し、緻密な温湿度管理や厳重な防災対策を実施するためのガイドライン等や電子公文書を適切に保存できる施設・設備を整備する等が指摘されているが、ここでは、オンライン移管のときのサーバの安全性及びオフライン移管のときの電子媒体の安全性を長期に確保するために必要な対応を示す。なお、記録がきわめて重要である場合には、記録のバックアップを解読・表示できる第 2 システムを離れた保存装置の区域に装備する必要性も考えられる。(ICA 報告書 16)。

6-5-2 エッセンスの保存

電子公文書等に含まれる全ての情報(例えば統計データを算定する場合の数式や埋め込まれているアプリケーションのプログラム)を保存するという選択肢も考えられるが、持続的に完全な形で保存することは難しい。したがって、研究会の議論の中では、電子公文書等の本質的部分 = エッセンスを特定し、保存することが適当であるとの方向性が示されている。

例えば、「ELECTRONIC RECORDS ARCHIVES INTRODUCTION TO PRESERVATION AND ACCESS LEVELS CONCEPTS」(2005年 NARA ERA PMO)によれば、NARA の目標は電子的な記録の“essential properties”を保存することである、との記載がある。また、“essential properties”は、記録の content: 内容、structure: 構造、context: 背景、presentation: 表現を含むものとされている。

今後は、電子公文書等のエッセンスを記録様式等に応じて特定することが検討課題となる。

(1) 見た目の保存

現在国立公文書館が提供している「デジタルアーカイブ・システム」や「デジタルギャラリー」においては、いずれも、画像レベルで紙媒体による歴史公文書等の原本の「見た目」に近いものがインターネット上で提供されている。したがって、これと同様のレベルで、電子公文書等のエッセンスを保存し利用に供することができれば、利用者のニーズは十分満たされると考えられる。

(2) テキストのみの配列の保存

研究会において、「エッセンス」に関するもう一つの選択肢として、テキストの配列のみを保存するといった選択肢が示されている。これは、電子メールの保存を行う場合等に活用可能ではないかと思われる。この場合、添付ファイルやデータベースへのリンクといったも

のをどのように扱うかが問題となる。だが、これらについては、「エッセンス」の対象外として一般的に認知されるのではないか。

また、テキストのみの配列の提供を行う場合の課題として、昔のワープロソフトで作成された電子ファイルがテキスト形式で読み出すことが可能かどうか、公文書として一定の重要性を示す決裁印や手書きの情報をテキスト化させることが可能かどうかという点があげられる。

6-5-3 安全性確保の手法

移管後、保存システムにある電子公文書等は原本となるため、あらゆる消失の危険性に対して、対策を施す必要がある。安全性確保の手法として、電子公文書等を保存するハードディスクの故障対策と、災害や人為的ミス等によるデータ消失対策を2段階で利用することが考えられる。

(1) 電子公文書等を保管するハードディスクの故障対策

ハードディスクの安全性を確保するためには、複数のハードディスクに同じデータを保存しておくことが重要である。ハードディスクは機械である以上、少なからず故障の危険性が存在する。そのため、同じデータを常に2つ以上のハードディスクに保存しておくことで情報の損失を防ぐことが好ましい。具体的技術としては、2つ以上のハードディスクにデータを保存する方法として、ミラーリングという手法がある。

(2) 災害や人為的ミスによるデータ消失の防止

2台のハードディスクに電子公文書等を保存しておくことで、ディスクの故障に対する脅威からは守られるが、地震や人為的ミスにより、二つのディスクが同時に故障してしまうことも考えられる。そのため、電子公文書等のバックアップ媒体を離れた区域で保管することも考えられるが、今後の検討事項である。

6-5-4 原本性確保技術

電子認証、電子署名等については、現在の技術では、電子公文書等を将来の世代にわたって国民の利用に供していくという使命を有する国立公文書館側のニーズを満たすレベルまではいたっていない。すなわち、国際的に標準化している技術レベルでは、電子署名の安定性は数年レベルといわれており、それ以上の対応となるとベンダ依存のリスクを大きく抱えてしまう。したがって、当面はこうした原本性確保技術に頼らず、運用ルールの徹底により、電子公文書等の原本性を保護する努力を図ることとする。なお、技術の進展については注意深く見守る必要がある。

6-5-5 オープン技術

電子公文書等の長期保存に伴う技術面でのリスクについては、ベンダに依存しない技術の標準化が進むことで一部軽減される。今後は、システムの設計フェーズにいたった場合には、オープン技術を使うことが適当か、専門的な検討を行なうこととなる。

6-6 適切な利用の実現に向けて

国立公文書館の使命は、政府から移管された歴史公文書等を保存し、広く国民に利用に供することにある。

移管された電子公文書等のエッセンスを保存し、その複製を作成してインターネット等により提供することで、従来、国立公文書館まで足を運ばなければ得られなかった情報が、自宅や職場などにいながら閲覧することができる。一方、移管された電子公文書等の原本については、一般利用者がアクセスする複製とは分離して保管しておくことで、安全性を維持する必要がある。

6-6-1 電子公文書等の利用・公開における考え方

ブロードバンドサービスの普及など、我が国内におけるインターネット利用環境が飛躍的に向上していることなどを背景として、国立公文書館においても、電子公文書等の移管が開始される際には、国民に対するサービスを一層向上させることが求められているといえよう。

国立公文書館では、既に、インターネットを通じ、「いつでも」「どこでも」「だれもが」「自由に」「無料で」、所蔵歴史公文書等の目録情報の検索と資料原本のデジタル画像情報の利用が可能な「デジタルアーカイブ」を提供している。

電子公文書等の移管についても、移管後の電子公文書等をインターネットにより利用に供することが考えられる。これにより、24 時間の情報提供が可能となり、どこからでも自由に情報にアクセスすることができる。

6-6-2 利用システムの概要

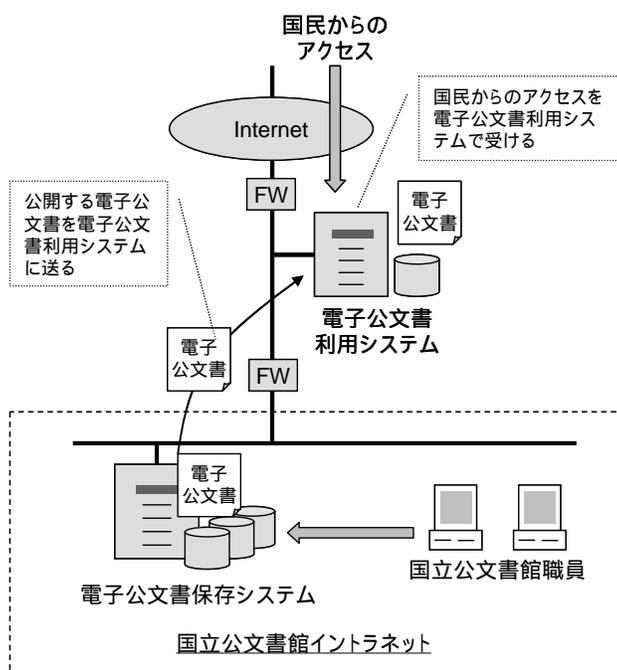
オーストラリア国立公文書館では、移管後の電子公文書等は、当該公文書等を作成後30年経過した時点で、公文書法(Archives Act 1983)に基づいた公開が行われる。つまり、いわゆる「30年原則」に基づいて、公文書館における公文書等の公開が行われている。そのため、オフラインにより物理媒体を用いて移管した電子公文書等を、いわば「凍結保存」し、利用時の提供方法等については、現時点では具体的検討等は行っていないようである。

一方、我が国では、「30年原則」が消滅しており、現用段階の公文書等の保存期間が3年、5年のものでも、移管後、11か月以内に国立公文書館は原則として一般の利用に供することとしている。したがって、電子公文書等を移管後、適切に保存すると同時に早期に公開できるようにしなければならない。

ICA 報告書 16 で「記録の保存と利用の双方の目的に元のシステムを利用するという選択は出来ないのが普通である。記録の保存には別のシステムを実装する必要があり、記録を作成したシステムから今後保存する予定のシステムへ記録を移動させるメカニズムが必要となる。」とあり、保存システムと利用システムは別個に対応することが必要と考えられる。

電子公文書の原本を保管する保存システムにアクセスできる者は限定されていなければならない。保存システムを一般の利用者がアクセスできない箇所(例えば、国立公文書館イントラネット)に分離しておくことで、電子公文書等の原本に対するセキュリティの脅威から守ることができる。一方、国民に対して利用に供する電子公文書等は、国立公文書館のイントラネットの外部に専用のシステム(ここでは、利用システムという)を設け、保存システムからエッセンスの複製をおく。

このシステム構成をとることにより、イントラネットの外部に設置されている利用システムが、万が一、悪意のあるユーザ等に入り込まれ、公開されているエッセンスの複製が改ざんされたとしても、保存システム自体がネットワーク上分離しているため、電子公文書等の原本を守ることができる。



図表 6-7 電子公文書利用システムの概要(案)