

「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－内閣府・国立公文書館の取り組みから－ (国立公文書館中島康比古講師) について

水野 京子
国立公文書館

はじめに

内閣府及び国立公文書館は、2005年度より、電子媒体による公文書等の移管・保存・利用についての検討を重ねており、その成果は、公文書等の管理、保存及び利用に関する懇談会による「中間段階における集中管理及び電子媒体による管理・移管・保存に関する報告書」（2006年6月）および、国立公文書館による「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」（2006年3月）などにまとめられ、公表されてきた。本講義では、これまでの成果を振り返りつつ、電子公文書等の長期にわたる保存、利用等に関する基本的な考え方などが紹介された。以下は、受講者の一人として理解した範囲内で講義概要を報告する。

1. 内閣府懇談会による検討

1.1 検討の経緯

2005年6月、電子媒体による公文書等の長期保存に関する本格的な調査研究を行うために、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が設置され、検討が行われた。この成果は、内閣府の公文書等の管理、保存及び利用に関する懇談会により、さらに総合的に検討が加えられ、「中間段階における集中管理及び電子媒体による管理・移管・保存に

関する報告書」として取りまとめられた。

1.2 9つの「基本的視点」

報告書では、電子媒体による公文書等の管理・移管・保存のあり方に関して、9つの基本的視点が提示されている。

- 1) 電子公文書等の適切な管理・移管・保存等に積極的に取り組み、公文書等の記録としての価値を維持しつつ、その多様性、利便性等を将来の国民に受け継いでいく必要がある。
- 2) 電子公文書等は、基本的には電子媒体のまま保存することが適当である。
- 3) 電子公文書等の長期保存に当たっては、記録としての価値を維持するのに不可欠な「エッセンス」のみを保存することが適当である。
- 4) 長期的な安定性を重視しつつ、現時点で可能な方法で電子公文書等の保存に着手する必要がある。
- 5) 長期保存の安定性・効率性等の観点から、作成時から歴史資料としての保存・利用段階までのライフサイクル全体の管理を行う必要がある。メタデータ、フォーマットについては、作成前から適切な対応をとる必要がある。
- 6) 歴史資料として保存・利用の対象となる電子公文書等は、保存期間満了前の可能な限り早期に、媒体を問わず同一の基準に基づいて、選定する必要がある。
- 7) 歴史資料として保存・利用の対象となる電子公文書等は、保存期間満了前の可能な限り

水野 京子 (みずの きょうこ)
独立行政法人国立公文書館公文書専門員

早期に、一定の集中管理下で長期保存上の措置を講じる必要がある。

- 8) 行政文書が情報公開・個人情報保護法の適用を受けるところから、これらの法制と電子公文書等の保存期間満了前における集中管理等の関係を整理する必要がある。
- 9) ウェブ上の電子公文書等の移管は、国立国会図書館とは別個に、各府省等から独自に移管を受けることが適当である。

1.3 「視点」から学ぶべきこと

上記視点から学ぶべき重要な点としては、次のものがあげられる。

第一に、電子文書を非電子媒体に変換せずに、電子媒体のまま保存するという点である。報告書では、電子文書に付されていたメタデータや記録の構造などの機能が非電子媒体への変換によって失われるおそれがある、ということが指摘されている。例えば、電子決裁文書に許認可の申請書が添付されており、さらに、その申請書に添付文書が付いているというような、添付文書によって元の文書が補足されている場合においては、電子文書を非電子文書に置き換えることにより、文書における記録の構造データは失われる。また、このような記録の構造は、記録を生み出した業務の構造を背景にしたものである。許認可関係の文書の場合は、申請から許可に至るプロセスの構造を文書という形で固定化させているが、これは、業務の構造をエビデンス（証拠）として残すものである。

デジタルネットワークにおける古典的な著作である、Nicholas Negroponte の『Being Digital』（『ビーイング・デジタル』新装版、アスキー、2001年）は、アトムからビットへ変革することによって、我々の社会、経済活動などがどのように変化するのかということについて論じているが、この中で FAX にまつわる興味深いエピソードを紹介している。Negroponte は、「ファックスの普及は一步後退でしかなく、情

報システムの全体像の中で深刻な汚点となっている」（前掲257頁）と述べている。FAX が普及した当時、欧米では既に PC が普及しており、PC で作成した文書を FAX で送信するというような現象がおこっていた。PC で作成された文書自体はマシンリーダブル（機械可読）であり、E-mail で送付すれば、ファイル量も小さく簡単に送れ、さらに、受信先でマシンリーダビリティを生かしてテキスト検索などが可能であったが、FAX のために、故意に非電子に置き換えることによって、文書が本来持っていたマシンリーダビリティを失うこととなってしまった。

現在、国立公文書館では、デジタルアーカイブにおいて、紙文書を電子化してインターネット上で公開している。今後、国立公文書館では、ポーンデジタルの文書は、電子文書の特性を生かし、デジタルアーカイブに搭載して公開していくと考えられるが、これは、電子文書の利点であるマシンリーダビリティを維持し、非電子化という作業を伴わない経済的な方法である。

第二は、記録の内容、コンテキスト、構造、機能等と媒体を明確に区別するという点があげられる。媒体変換は今や必須であるが、記録と媒体は本来的には別のものと考えべきである。文書において究極的に何を残すのかを選ぶ場合、多くは、媒体そのものではなく、有用な情報だけを抜き出して保存するという方法が選択されるだろう。例えば、CD R を原本とした場合は、CD R から情報が読み出せなくなったとしても永久に CD R を保存しなくてはならなくなる。しかし、記録を原本とすれば、CD R には利用価値はなくなり、場合によっては破棄という選択肢もあり得る。このように、記録と媒体は本来的に別のものとした上で、電子文書は、電子による保存を行うという原則を打ち立てるべきである。

第三は、メタデータ項目を適切に定義し、用語集の整備等により、記述内容標準化を図る必

要についてである。メタデータの項目を定義しても、実際に使用する用語が不統一ならば意味をなさない可能性がある。よって、用語を整備するために、記述内容の標準化をはからなくてはならない。さらに、報告書では、メタデータを自動または容易に付与できるシステムを作ることが望ましいとされている。

第四は、既に作成済の電子文書については、特定システムへの依存度が低い長期保存に適したフォーマットに変換して保存することが望ましいという点である。電子文書を管理・保存する場合、管理上の困難さやコスト面において、データ量よりもフォーマットの乱立が問題となる。したがって、フォーマットの数を絞り込み、長期保存に適し、かつ、特定のシステムへの依存度が低いフォーマットへ変換して保存することが望ましい。現時点では、長期保存に適した絶対的なフォーマットは存在しないが、未来には、今は特定のソフトでしか可読不可能な形式でも、全ての形式を可読可能なソフトが開発されるかもしれない。また、未来に可読不可能になったとしても、作成時のフォーマットをエビデンスとして記録しておくべきであろう。

第五には、今後作成される文書は長期保存に適したフォーマットでの作成が望まれる点あげられる。長期保存に適したフォーマットとして現在は PDF/A を想定しているが、PDF/A が絶対的に安全な形式とはいえない。しかし、PDF/A は ISO で標準化された、国際コミュニティの中で共有化されているフォーマットなので、次の標準フォーマットに変換する場合、その成り立ちが尊重される可能性が高く、これは、フォーマット変換のサイクルを延ばすことにもつながる。フォーマット変換の頻度が頻繁になれば、データ破損の危険性も高くなり、コストもかかる。したがって、現時点で最も長期保存に適したフォーマットでの作成を選択すべきである。

続いて第六は、記録としての価値を維持する

ために不可欠な「エッセンス」の保存が適切であること、及び、記録の内容および文脈や背景といったコンテキストは、エッセンスとして保存する必要があるが、構造、機能等は、記録様式等に応じてエッセンスを特定し、記録することが適切である点あげられる。報告書では、記録様式ごとにエッセンスの考え方が例示されているが、定義付けは行っておらず、記録様式ごとにエッセンスの範囲を特定することが提言されている。例えば、ワープロソフトで作成した文書を PDF に変換して保存することは、エッセンスの保存にあたるが、これは紙媒体による保存でも同様である。印刷したものを正本として保存する場合、文書作成時に加えられた頁番号追加機能は保存時には失われているからである。

エッセンスとは何か。例えば、講義においてパワーポイントで示した「エッセンスは何か」という言葉は、他よりも文字のポイントを大きくしている。これを PDF 化しても「エッセンスとは何か」を強調したという意図は残るだろう。このように、機能の部分は注記することにより残すことができる。しかし、これは手間や時間の問題から、ある程度記録様式ごとにグルーピングしてやるべきことである。エッセンスについて、講師は、試みとして、2007年の EASTICA のカントリーレポート、2008年の ICA 大会における「日本における電子政府化の親展と電子記録長期保存に向けた取り組み」についての報告（『アーカイブズ』34、2008年、37～42頁）の中で「電子記録の完全性、真正性及び信頼性を損なうことなく、利用可能性・アクセス可能性・理解可能性などを長期的に維持できることが技術的に裏付けられている範囲」という定義付けを行った。

2. 国際公文書館会議 (ICA) の取り組み

2.1 長期保存の目的と記録の保持すべき属性

以下には、2005年の報告書の内容に基づいて

ICA の取り組みについて紹介する。ICA が行った電子文書の長期保存についての調査において比較的まとまっているものは、1997年の「アーカイブズの観点から見る電子記録管理ガイド」(ICA 報告書 8)、2005年の電子環境における現用記録委員会「電子記録：アーキビストのためのワークブック」(ICA 報告書16) の2本であるが、今回は、2005年の報告書の中の「第五章 長期保存」を主として紹介する。

ICA の報告書では、「長期」とは、その記録を作成するシステム(ハードウェアやソフトウェア)の寿命より長期-現在の技術革新の速度では一般的には5年以上)-であることと定義されている。その上で、行動や決定の有無を具体的には決めず証拠としての記録の価値を保存する、及び、記録自体または記録が内包している情報を再利用することが、長期保存の目的としてあげられている。そして、目的に応じた形で電子記録を長期保存する場合における、5点の「記録の保持すべき属性」があげられている。

1点目は「真正性」である。報告書では、記録が求められるときに作成・取得されたこと、作成プロセスが求められるプロセスであったこと、記録がその一部となることを求めるシステムの一部であること、システムが一部となっているため、記録が改変されていないことを証明すること、と説明されている。これは例えば、電子文書の保存においては、個別の文書の中の電子署名やタイムスタンプの長期的な確認を担保できないので、その真正性の保障を基本的にはしない、ということである。電子署名やタイムスタンプについてビットレベルまで100年後も証明することは不可能であり、現時点では、そこまでの厳密性は必要とされていないだろう。基本的に電子署名などは、電子情報の真正性を証明するためのものであり、これらが流通している間はよいが、閉じられたシステムの中で保存するのであれば、個別の証明ではなく、システム全体の安定性を証明したほうが、一定程度

の必要性は成り立つと考えられる。

2点目は「完全性」である。これは、一連の文書に一つの一件文書も加除されていないことを証明するという意味である。これは、逆に考えれば、残ってはならないものが何も存在しないということが証明されなければならない、ということだともいえる。

3点目は「アクセス可能性、理解可能性」である。記録がアクセス可能であり、記録が理解可能な形態に変換できることを保証する、というものであり、いわゆるマシンリーダブルのことである。

4点目は「処理可能性」である。記録は操作、選択、表示できなければならない。

最後は、「潜在的な再利用可能性」である。記録から情報を引出せたり、最新の情報処理システムと相互運用し合えなくてはならない。

以上のように5つの属性が定義されている。しかし、記録やアーカイブズ、文書管理は、その必要性に応じて社会的・技術的な合意が必要であり、現時点では絶対的基準が存在しているわけでないことが付け加えられる。

2.2 長期保存の方法と長期保存フォーマットのあり方

さらに、ICA の報告書は、上記のような長期保存の目的や属性について定義付けした上で、記録の見かけは変化するが、博物館が望むようなスタイルで、旧式のコンピュータソフトウェアやハードウェアの動作や外見を保存するのではなく、記録の根本的な性質や、証拠としての価値や、記録の中の情報を保存するという、長期保存の方法についての考え方を示している。記録を長期保存する場合、長期保存のための原則や方針を定め、これに基づいた記録作成のためのメタデータの標準化や、新技術の採用による個別システムの立ち上げなどを行い、記録を管理していくが、管理においては「変化」を想定して行うことが重要である。しかし、全ての

変化を受け入れることはできないので、その許容範囲を検討する必要がある。また、管理を行うためのツールとしてメタデータが考えられるが、メタデータを体系化していくことも必要である。

こうした長期保存方法についての考え方に基づき、ICA 報告書は、長期保存フォーマットのありかたについて、次の点をあげている。

- 1) 全ての情報、または元の記録のなかで有意義とみなされる情報相互の関係を表現できる。
- 2) 国際・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。
- 3) 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。
- 4) アクセスの目的に直接的に利用可能であるか、利用可能な方法に変換できるか。
- 5) 特定のソフトウェア環境、ハードウェア環境に依存していない。
- 6) 元のフォーマットから保存用フォーマットに自動変換できる。必要に応じて、変換時の問題やエラーを自動で検出・報告する機能がある。
- 7) (必要に応じて) 保存用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変換できる。

このように7点があげられているが、これら全てを満たすフォーマットはおそらく存在しないと考えられる。特に最後の点を満たす場合は、長期保存としての属性が失われることもあり得るように思われる。よって、これらの属性をどの範囲まで求めるのかということは、自身の必要性にしたがって考えるべきであり、様々な制約の中で最善の措置を講じ、その時々で決定したルールを公開していく必要があるだろう。

2.3 メタデータの類型

さて、ICA の報告書では、メタデータについて、ISO15489という記録管理の国際標準の

定義を引用して「記録のコンテキスト (背景・状況・説明)、内容、構造および、ある機関の記録の管理について説明したデータ」とされている。そして、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つに類型化している。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る。記録が作成された元々の目的のために必要なメタデータ」、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メタデータ」、技術的メタデータは、「記録の理解や処理に必要なメタデータ」である。三点目の技術的メタデータは、電子文書においては、使用したフォーマットや文字情報などの情報を示している。

3. 電子記録の管理とメタデータ

ISO15489においてメタデータは、上記のように定義されるが、さらに、ISO15489の考え方に基づき記録の管理を行うためにメタデータを定義した国際規格である、ISO23081がある。ISO23081では、メタデータを、記録の内容、構造及び見かけのほか、記録が作成又は捕捉される業務コンテキストを記録したメタデータと、記録の管理のほか、記録が作成後利用される業務プロセス (記録の内容、構造及び見かけにもたらされるあらゆる変化を含む) を記録したメタデータの二つに類型化している。この二つの類型を、前述のICAのメタデータの類型化にあてはめると、前者のメタデータは記録管理メタデータ、後者のメタデータはアーカイバル・メタデータといえる。

記録・文書管理におけるメタデータを考えた場合、記録そのもの、或いは、記録が作成されたコンテキストや構造を記述するメタデータのほかに、管理プロセスを記録するメタデータも必要である。これは、管理プロセスを記録するメタデータによって、文書が適切に管理されて

いたことが証明されるからである。このように記録を管理するためには、プロセス管理のためのメタデータを付与する必要があるが、そのメタデータも記録として管理するものである。さらに、記録を管理するメタデータにも、新たにメタデータを付与して管理するということになる。つまり、記録とプロセスがあり、それを管理するメタデータがあり、さらに、そのメタデータを記録と同じように管理するためのメタデータと、そのメタデータの管理プロセスがある... というような連続した構造が成り立っている。こうした構造は理論上無限に構築できるが、メタ・メタ・メタレベル程度でも、かなり安定性の高いシステムを作ることができる。

しかし、文書を記録としてではなく情報として保存するだけなら、上記のような厳密なメタデータの管理は必要ない。ISO15489においても、所謂アーカイブズ機関で保存されている資料には、上記のような厳密なメタデータの管理を直接には適用しないとしている。これは、実際にアーカイブズ機関で保存されている資料が多様なため、こうした考えが馴染まないことがあげられるが、もし文書に証拠としての性質を一定程度確保しようとするなら、ISO15489の適用に関わらず、こうした考え方を念頭に置く必要があるだろう。

最後に、電子文書の長期保存については、「できることから始める」ということがあげられる。これは、例えば、電子文書の長期保存について自己及び周囲の理解を深めることなどの試みのことである。2005年のICAの報告書でも「アーキビストの啓発戦略」に一章が割かれているが、電子記録の長期保存の目的を果たすためには、周囲の理解を広げることが重要である。

しかし、電子文書の保存については、現時点

で絶対的に正しい方法があるわけではない。昨年アメリカで刊行された『Electronic Records in Manuscript Repository』(Elizabeth H Dow. Scarecrow Press. 2009)は、個人寄贈文書を保存しているような比較的小規模の文書館のアーキビストが電子文書にどう立ち向かうのかということを描いたものだが、その中には、小規模な機関であれば電子文書を電子のまま保存するというのを金科玉条のように考える必要はないとしており、電子文書を非電子媒体にして保存するというのも選択肢の一つであると述べている。このように、電子文書を証拠として保存するということが可能であれば、様々な方法があり得る。現時点で確実にいえることは、完璧なソリューションはないが、世界各国がそれぞれに試みており、そうした世界的な動向の中で日本も立ち後れることなく取り組むべきだということである。

おわりに

以上に、講義の概要を紹介した。約100分に及んだ講義は、非常に充実した内容であり、数々の例示を交えながら丁寧な解説がなされたことにより、内閣府及び国立公文書館における電子媒体による公文書等の適切な移管・保存・利用に向けての取り組みについて、理解を深めることができた。本講義で紹介された電子文書の記録保存に対する考え方は、講義後行われたグループ討論においても、電子文書の記録・保存を検討する上での共通概念として受けとめられていた。アーカイブ機関の現状は様々であるが、本講義を踏まえ、電子文書の記録・保存に関して「できること」を検討し、それぞれの環境に即した取り組みを「始める」ことの重要性を感じた。