

**電子情報保存に係る調査研究
報告書**

平成 15 年 3 月

国立国会図書館

電子情報保存に係る調査研究 報告書 目次

1 . はじめに.....	1
2 . 諸外国の現状.....	3
3 . OAIS 参照モデル.....	24
4 . メタデータ.....	40
5 . 技術動向.....	52
6 . コスト.....	79
7 . 保存媒体、環境.....	99
8 . 情報発信者と図書館の協働.....	118
9 . 終わりに.....	133

1. はじめに

本章では、調査研究の背景と目的及び調査研究の手法について説明する。

1.1. 調査研究の概要

昨今の著しい情報技術の発展により、コンピュータを利用して作成される電子情報の量は、年々飛躍的な増大を続けており、さらにインターネット等のネットワークの急速な普及もこの傾向に拍車をかけている。しかし、流通する電子情報を収集し、後世代に伝えるため保存していく際に、新たな課題に直面する。

例えば、CD や DVD 等の媒体は耐用期間が短く、媒体の劣化により格納している情報が再生できなくなる可能性がある。またインターネット等のネットワーク上の情報は、コピーや改竄が容易であるとともに常に改変・削除が繰り返され、安定した状態ではないため、どの時点で保存したものが原本なのかの特定が困難である。あるいは、電子情報の再生に当たって必要な機器やソフトウェアは、技術進歩による陳腐化も早く、過去に作成した情報を再生する環境が失われてしまうという危険性も挙げられる。

このような認識から、諸外国、特に欧米、オーストラリア等において、電子情報の長期保存に関する各種の調査研究やプロジェクトが実施されている。日々生産され続ける膨大な量の電子情報どのように収集し、集めた情報の長期アクセスを保証していくかという問題が、重要なテーマとして捉えられている。

今回の「電子情報保存に係る調査研究」(以下、「本調査研究」と略記する)では、電子情報保存に関する国内の状況、諸外国における研究動向や技術動向を把握するとともに、今後の取り組みに当たっての整理を行うことを目的としている。ここで行う調査研究の成果を基に、将来的には国立国会図書館における電子情報保存のための技術的、組織的、制度的枠組みづくりについて検討していくことを想定しており、本調査研究はその予備段階と位置づけられるものである。

本調査研究では、以下の調査研究テーマを設定した。

- ・ 諸外国における電子情報保存への取り組み状況
- ・ 技術動向
- ・ OAI参照モデル
- ・ メタデータ
- ・ コスト
- ・ 保存媒体、環境
- ・ 情報発信者と図書館の協働作業
- ・ 国内における電子情報保存の現状把握

本調査研究の実施に当たっては、各テーマについて電子情報保存に係る先進的な取り組みを行っている諸外国の事例調査等が中心となることもあり、インターネット上に公開さ

れている各種研究文献を収集・調査した。なお、調査対象国は時間的な制約もあって 5 カ国（米国、英国、フランス、オランダ、オーストラリア）に限定した。また、国内の電子情報保存の現状を把握するに当たっては、公共機関、民間企業等の関係諸機関に対してアンケート調査を実施した。調査対象文献とアンケートの詳細については巻末の付録に記載した。

2. 諸外国における電子情報保存の現状

本章では、米国、オーストラリア、フランス、英国、オランダの5カ国を取り上げ、各国の電子情報保存への取り組み状況について整理する。

この5カ国は国内における情報通信環境の整備が進んでいる国々であり、フランスを除いては各国内における人口に占めるインターネット普及率がほぼ50%を超えている¹。このような状況を背景に、電子情報の生産・流通が活発であり、電子情報保存に関する問題も顕在化しつつある。

特に米国においては、米国議会図書館がNDIIPP（National Digital Information Infrastructure and Preservation Program）というプロジェクトで、国家レベルで電子情報保存への取り組みを開始した。一方、オーストラリアにおいては国立図書館が中心となって、世界に先駆けて電子情報保存に向けた取り組みを実施している。

ヨーロッパでは、特にオランダにおいて先進的な取り組みがなされている。例えば、1998年から2000年にかけてオランダ国立図書館が中心となり、ヨーロッパ各国の国立図書館が連携したNEDLIB（Networked European Deposit Library）という国際プロジェクトが実施されている。このような研究プロジェクトを基に、オランダでは国立図書館だけでなく、政府関係機関においても電子情報の保存に対する取り組みが積極的になされている。

英国、フランスは、伝統ある国立図書館を保有しており、特に英国図書館では、電子情報保存で先行するオランダ国立図書館と緊密に連携して、積極的な取り組みを行っている。また電子情報保存の啓蒙活動を行う組織も設立されている。

このように諸外国では、電子情報保存に対する課題が認識され始めるとともに、課題解決のための実験プロジェクト等が積極的に行われている。これらの事例及び各国の状況を調査することは、日本における電子情報保存を考える上でも極めて重要である。

¹ 総務省の調査によると、人口当たりのインターネット普及率ではオランダが世界4位（60.8%）、米国が7位（59.1%）、英国が8位（57.2%）、オーストラリアが11位（54.4%）となっている。フランスは国別のインターネット利用者数では9位で1,697万人が利用している。利用者数では1億6千万人で米国が1位となっている。日本は人口当たりの普及率は10位（54.5%）、利用者数では6,942万人で2位となっている

2.1. 諸外国の国立図書館における電子情報保存への取り組み

今回とりあげた各国の国立図書館の大半においては、国内における出版物等を後世に伝えるべき文化遺産と捉えて収集・保存を行ってきたが、近年、著しい増加を見せる電子情報の保存が大きな課題となっている。

オフライン系電子出版物では、CD-ROM、DVD 等のパッケージメディアで作成されたマルチメディア出版物等が増加を見せており、オンライン系ではインターネット等による情報発信が盛んに行われるようになっている。各国の国立図書館では、各種の電子情報についても、紙媒体の出版物と同様に各国の文化の諸側面を表現するものであり、保存していく必要があると考えられている。

このような認識の下に、先に述べた 5 カ国においても国立図書館が中心となって、電子情報の保存に係る取り組みを実施している。

2.1.1. 各国国立図書館の納本制度の枠組みにおける電子情報の収集

各国の国立図書館の大半は納本図書館としての役割を担っており、これまでも各国内で出版される出版物を中心に収集・保存の取り組みが行われてきた。

電子情報の収集・保存については、各国まちまちであるが、以下にオフライン及びオンライン出版物の納本制度による収集状況について簡単に整理する。

	米国議会図書館	オーストラリア 国立図書館	フランス国立図 書館	オランダ国立図 書館 ²	英国図書館
オフライン系 電子出版物	(CD-ROM 等を法定納本)	(任意で CD 等の納本を実施)	(CD-ROM 等の法定納本)	(出版協会と任意納本の協定)	(任意納本制度が存在)
オンライン系 電子出版物	検討中 ³	検討中	検討中	検討中	検討中

表 2-1. 各国の納本制度における電子情報の取り扱い

上記に整理した通り、オフライン系の電子出版物については、各国とも法定納本、任意納本のどちらかで、共通する取り組みが行われている。一方で、オンライン系の電子出版物の法定納本については現在、検討中とするところが大半を占めているのが現状である。以下、各国の納本制度の現状について簡単に整理する。

² オランダ国立図書館においては、納本制度の枠組みよりも、自館の蔵書構築を目的に出版物を収集していく活動に注力している

³ 米国の納本制度では著作権が発生する出版物等についてはすべてが納本対象であるが、オンライン出版物については著作権局の電子登録システムの取り組み等にあわせて、さらに詳細な定義を検討していく方向性にある

米国議会図書館

米国における納本制度は同国の著作権法により規定されており、著作権法上で保護の対象となるもので米国国内において出版されるものについては、最良版⁴2部を3ヶ月以内に米国議会図書館に設置されている著作権局に納本しなければならないとされている。この規定では著作権登録をしない場合にも、作成・出版の段階で著作権法の保護の対象とされるため、納本の義務を負うことになる。出版社等が納本しない場合には罰金を支払うことになっている。納本されたものについては、米国議会図書館が蔵書として管理・保管することになる。

オフライン系の電子出版物である CD-ROM については、現在、納本が義務付けられているが、一方でオンライン系の電子情報については、情報がネットワーク上にあるため、複製権等の各種権利の問題があり、米国議会図書館が納本のために複製を作成することは認められていない。

また著作権局では、著作権局電子登録・記録及び納本システム（CORDS：Copyright Office Electronic Registration, Recordation and Deposit System）が整備されており、インターネットを利用した電子出版物の納本、あるいは電子出版物の登録を行うことが可能である。CORDS は HTML、PDF、MP3 等様々な電子ファイルフォーマットに対応しており、電子出版される学术论文、コンピュータプログラム、電子書籍、音楽ファイル等の納本・登録を意識したシステムになっている。1996 年から試験的な運用が行われており、テストパートナーと協力しながら、徐々に対象となる情報を増やしている。今後も、更なる拡充が予定されている。

オーストラリア国立図書館

オーストラリアは連邦制を採用しており、各州の下に自治体が設置されている。この関係は図書館システムにも反映されており、首都のキャンベラに国立図書館が、各州には各州立図書館と自治体毎に地域図書館が設置されている。

オーストラリアにおいては、法定の電子出版物の納本制度は存在しないが、将来的には法定の制度が設置される見込みである。ただし、現行でも、州によっては電子出版物を納本させているところも存在する。

オーストラリア国立図書館は国立図書館法（National Library Act）によって、紙ベースの出版物等の文化遺産を保持することが義務づけられている。一方で、Screen Sound Australia は映画、音声、放送の保存を義務づけられている。両者は共同で、物理的媒体に格納された電子出版物、オーディオ・ビジュアル関係資料の法定納本に向けた活動を行っている。

⁴ 著作物の「最良版」とは、納本日以前に合衆国内で発行された著作物の版のうち、連邦議会図書館がその著作物の目的に最も適していると判断するものを指す（米国著作権法第1章第101条）。「目的に最も適している」点についての判断基準は、保存性の良さ及び特別機能の付加状況による

フランス国立図書館

フランスでは、法定納本の法律によってフランス国内で製作・流通される出版物については、あらゆる種類のものが納本されることになっている。現在の法定納本に関する法律は 1992 年に制定され、翌 93 年に施行されている。法律では、明確に電子出版物については言及していないが、法律の施行とともに、フランス国内で作成される CD-ROM 等のオフラインの出版物に関しては納本が行われている。現行の法制下では、以下の機関が納本先としての役割を果たしている。

機関名称	概要
フランス国立図書館 (Bibliothèque nationale de France)	すべての出版物、ビデオ、マルチメディア製品を取り扱う義務がある
国立映画技術センター (Le Centre national de la cinématographie)	映画フィルムを取り扱う義務がある
国立オーディオ・ビジュアル研究所 (L'Institut national de l'audiovisuel)	ラジオとテレビ放送を取り扱う義務がある

表 2-2. フランスにおける納本機関

上記以外に、紙の出版物については各地域において納本を受け付ける施設があるほか、19 の地域図書館が納本を受け付けている。しかし、電子出版物の納本だけに特化した地域機関は設置されていない。電子出版物については、2 部をフランス国立図書館に納本することと定められている。

フランスには 3 百万以上のウェブサイトがあると推定されている。2000 年の 7 月にオンラインの電子出版物にまで法定納本を拡充するよう提案がなされたが、現在、法制化については検討が行われている。

オンラインの電子出版物の納本が法制化されると、フランス国内の作成者が製作したウェブサイトについては、納本することが義務づけられることとなる。納本に当たっては、作成者が FTP でファイルを転送する、運搬可能な媒体に記録して提出する、図書館自身がウェブサイトを収集する、といった方法が考えられる。現在検討がなされている法律では、収集は選択的になるのか、網羅的になるのか、あるいは選択は図書館が行うのかといった事柄については定められていないため、今後の議論が待たれる。ただし、議論が多方面にわたるため法制化は 2003 年あるいは 2004 年以降になる見込みである。

オランダ国立図書館

オランダには紙媒体の出版物にも電子出版物にも法定納本制度がない。そのためオランダ国立図書館では、蔵書構築を目的として、独自に出版物を収集するための取り組みが行われている。例えば、Elsevier と Kluwer という 2 つの国際的な出版社と任意納本の協定を結び、1999 年にオランダ出版社協会と任意納本に合意した。しかし、任意納本では元々の出版原稿の作成者と出版社の間での権利関係の調整が十分でないことが多く、制度的に限界があると考えられている。

オランダ国立図書館は、オランダ国内の出版物を収集する役割を担っており、主に書籍と定期刊行物に注力して収集を行っている。この中には CD-ROM 等のマルチメディア出版物も含まれているが、オーディオ・ビジュアル関係資料、映画、放送、データベース、各種ソフトウェア、ゲーム等は含まれていない。同様に、オランダ国立図書館はウェブサイトの選択的な収集・保存についての将来的な実施に関心を持っている。

英国図書館

英国には組織化された納本ネットワークが存在する。英国図書館を中心として、ウェールズ国立図書館、スコットランド国立図書館、オックスフォード大学ボドレー図書館、ケンブリッジ大学図書館、トリニティ大学図書館が連携している。

現在、英国には、電子出版物に関する法定納本制度はないが、将来的には法制化がなされる。英国図書館には、任意納本制度の検討会が設置されており、電子出版物に関する任意の納本制度について、出版関係者等との検討が進められている。おそらく、法定の納本制度が確立された際には、責任が分担され、その中で英国図書館が中心的な役割を果たすものと想定されている。

後述する電子情報保存システム Digital Library Storage の構築は、納本制度の改訂に向けた重要な準備活動になると考えられている。将来的にオンラインの電子情報の選択ガイドラインを考えるに当たっては、権利の保有者からの許諾を得て、ウェブサイトを選択的に収集するウェブアーカイビングプロジェクトが実施されており、この成果が納本制度の改訂に反映されるものと思われる。

任意による電子出版物の納本については、2000 年の 1 月から実施されており、主に物理的な媒体に格納された出版物について収集が行われている。2001 年の 1 月現在で、3,000 の電子出版物が納本されている。これらの中には、国立の地図作成機関である National Mapping Survey が実施した測量成果であるデジタル形式の地質図も含まれている。

任意の納本制度の運用面における課題としては、大容量の商業用データベースをどのように扱うかという点と、出版社（特に中小規模の出版社）が納本に際して付与するメタデータに関する点が挙げられる。メタデータに関しては、出版社が標準化を進めてい

る ONIX スタンドに沿ったメタデータの入力ツールを、英国図書館が一部資金を援助して開発している。このソフトウェアツールの目的は以下の通りである。

- ・ ONIX データベースへの入力、あるいはそのデータベースの管理に携わる人を訓練するためのツール
- ・ ONIX メタデータの入力、維持管理、ONIX メッセージの伝送のためのエクスポートを小規模なユーザー⁵が行えるようにする

2.1.2. 電子情報保存に係る取り組みの状況⁶

ここでは、各国の国立図書館における電子情報保存に対する指針や計画、保存のための情報システムの整備状況等について、国別に見ていく。

米国議会図書館

米国議会図書館は電子情報を使った図書館サービスに早くから力を入れており、90年代の初頭から貴重資料を電子化して公開する”American Memory”プロジェクトを始めとして、様々な取り組みを行っている。

電子情報保存についても、1998年頃から本格的な検討が開始され、2000年には米国議会が電子情報保存のための取り組みである NDIIPP (National Digital Information Infrastructure and Preservation Program) に対して総額約1億ドルの予算が承認されている。この予算の内、2,500万ドルを使用し、電子情報保存のための計画立案が行われ、2003年2月14日に計画書が公開されている⁷。この計画は議会で承認され、残る予算の7,500万ドルを使用して電子情報保存への取り組みが行われる予定である。

NDIIPP では、米国議会図書館が中心となり、連邦政府機関、あるいはそれ以外の機関と共同しつつ電子情報保存のための戦略や実行に取り組んでいくとしている。具体的な協力機関としては、国立医学図書館、国立農業図書館といった国立図書館、RLG (Research Library Group)、OCLC (Online Computer Library Center) といった図書館関係機関、米国商務省、ホワイトハウス等の政府機関、出版社等の民間企業の名前が挙げられている。

また、NDIIPP の計画には、電子情報の長期的な保存と利用のための重要課題として、以下の点が挙げられている。

- ・ 保存すべき電子情報の選択、収集、組織化
- ・ 知的財産権問題

⁵ここでいう「小規模なユーザー」とは、年間100件のレコードを作製し、トータルの管理レコードが1,000を超えないユーザーを想定している

⁶本文における各国の取り組みは2002年3月時点の状況である

⁷ http://www.digitalpreservation.gov/ndiipp/repor/repor_plan.html を参照

- ・ ビジネスモデルの確立
- ・ 標準化の推進とベストプラクティスの確立
- ・ 電子情報保存のためのアーキテクチャー

オーストラリア国立図書館

オーストラリア国立図書館は、冒頭でも述べたとおり、比較的早くから電子情報保存に対する取り組みを行ってきており、経験の蓄積も多い。また、今後の電子情報保存を考える上で重要になるであろう UNESCO の電子情報保存のガイドラインは同国立図書館が中心になって作成している。

オーストラリア国立図書館は、2001 年 7 月に電子情報保存の指針“A Digital Preservation Policy for National Library of Australia”を策定し、効率的な保存活動を行っていくために、他の機関との連携を視野にいれた将来的な方向性を示している。

上記以外で現在実施されている電子情報の保存に係る取り組みとしては、以下のものが挙げられる。

- ・ ウェブアーカイビングプロジェクト PANDRA で収集したコレクションのうち、HTML4.01 を使用しているファイルのフォーマットをマイグレーションするためのテスト
- ・ 既に陳腐化している DOS (Disk Operating System) を活用したエミュレーションに関する試験
- ・ 2001 年に実施したオーストラリアドメインに存在するウェブサイトの一括 (バルク) 収集プロジェクトに関する評価
- ・ 特定のフロッピーディスク、CD からデータを読み出して復元を行うプロジェクト
- ・ 文書の円滑な管理を目的として、Tower システム社の販売している TRIM というソフトウェアを活用し、陳腐化したフォーマットのためのビューワを整備し、古いワープロソフトで作成したファイルの使用について機能評価を実施
- ・ CD-R を保存媒体として位置づけるために徹底した評価の実施
- ・ 複数のソフトウェアレポジトリに関するコンセプト評価及びファイルフォーマットに関する技術動向の調査

電子情報保存を行うための情報システムの整備については、1999 年にデジタル・コレクション・マネジメントシステムに関する仕様書を策定し、意見を求めた上でパートナーとなるベンダーを募集したという経緯がある。しかし、同図書館の仕様を満たすベンダーが現れなかったため、以下に述べる 3 つの分野で、自主開発と外部調達を組み合わせることでシステム整備を実施した。

システムの名称	システムの概要
Digital Object Storage System	オーストラリア国立図書館保有のデジタル情報保存のためのストレージ・システム。多くのコンポーネントから構成されるシステムであり、外部調達によって構築された。2001年6月にサービスを開始し、2006年までの利用を見込んでいる。2002年にディスク容量を4テラバイト、テープ・ストレージを12テラバイトに増設している。
Digital Objects Management System	オーストラリア国立図書館の所蔵する電子化された資料及び元来からデジタルで製作された各種の電子出版物の双方を管理するために、同図書館が構築したシステムである。段階的にサービスが開始されており、将来的には、デジタル化された音声の管理システムや長期保存管理システムが構築される予定である
Digital Archive System	国内のオンライン出版物の分散型アーカイビングを行うPANDRAプロジェクトをサポートするために開発されたソフトウェア。すべての機能はWeb化されているため、関係機関がインターネットを通じてアクセスしやすい仕組みとなっている。Version1は既にリリースされており、ビクトリア州立図書館等から高い評価を受けている。Version2は2002年6月にリリースされた

表 2-3. オーストラリア国立図書館の電子情報保存システムの概要

フランス国立図書館

フランス国立図書館は、米国議会図書館と同様に90年代の前半から貴重資料の電子化に取り組んでおり、作成したデジタル情報はウェブサイトで公開されている。資料の電子化の際には保存を意識して、すべて単一のフォーマットでデータ化され、これらのデータについては、既に1度大規模なマイグレーションが実施されている。現在は、更なる資料の電子化の取り組みとしてビデオ資料の電子化が始められている。

フランス国立図書館では、組織横断的に電子情報の保存を行うためのワーキンググループを設置して検討を開始した。このワーキンググループは、デジタル・ライブラリープロジェクトチームの代表、オーディオ・ビジュアル部門、情報システム部門、保存部門、収集部門からの出席者で構成されている。会合は月に1回開催され、ワーキンググループでは電子情報の長期保存を行うためのアーカイブに関する技術標準であり、ISO規格にもなっているOAIS (Open Archival Information System) 参照モデルを適用するための検討が行われている。さらに、2001年にはフランス国内で電子情報保存の国際会議が開催される等、この分野への取り組みに関心が高まりつつある。

フランス国立図書館の情報基盤の整備は非常に進んでいる。館内に100台のUNIXサ

サーバと 150 台の Windows NT サーバ、そして 3,000 台のクライアントマシンが配備され、それらを 150Mbps の ATM⁸ネットワークで接続し、館外に対してもパリの研究ネットワークを通じて 150Mbps の接続環境を提供している。

しかし、電子情報保存システムを構築していくことは将来的な課題となっており、現段階での電子情報保存に関する取り組みとしては、保有情報を明確化し、マイグレーション等の保存戦略を決定する際に判断材料として利用できるメタデータデータベースの構築が行われている。

オランダ国立図書館

オランダ国立図書館はヨーロッパにおいて、先進的に電子情報保存に取り組んでいる国の一つである。例えば、1998 年からヨーロッパ各国の国立図書館と協働し、電子情報保存を研究するプロジェクトである NEDLIB(Networked European Deposit Library)を中心メンバーとして運営している。

最近のオランダ国立図書館における電子情報保存の取り組みとしては、IBM 社と共同で行った LTP Study (Long Term Preservation Study) が挙げられる⁹。研究の目的は、後述する電子出版物の納本システム DNEP に保存されている電子情報を数百年単位の長期間にわたり保存するために必要な機能を研究することである。研究は 2000 年の 11 月から実施されており、以下の研究テーマを設定している。

研究テーマ	概要
長期保存のための情報システムにおける実装方式	当初の DNEP システムのリリース時には、保存されている情報の再生に必要な技術的な情報の一部のみを記述して保持する機能しか提供されていなかったが、LTP Study では、当初のリリースでは想定されていなかった保存サブシステムの機能定義を行った。研究の結果として、保存されている電子情報の再生に必要なすべての技術的な情報を保持できるよう機能定義がなされた
UVC (Universal Virtual Computer) のコンセプト検証	IBM の Almaden 研究所の研究者 Raymond Lorie が提唱している電子情報保存のアプローチである UVC の仕組みをオランダ国立図書館で利用可能か否かを検討した
大容量データのメディア・マイグレーション	大量のデータを古い保存媒体から新たな媒体に移行する際に直面する課題について、解決策を検討した
真正性	電子出版物の保存の成功を左右する情報の真正性を定義するフレームワークの策定を行った

表 2-4. LTP Study の研究テーマとその概要

⁸ 読みは「イー・ティー・エム」。「非同期転送モード」の略で高速に情報通信を行うための技術をさす。光ファイバーとの併用で 156Mbps の高速転送ができるため、LAN 間接続等大容量の回線が必要ときに利用される

⁹ 約 13 万ユーロ (約 1,700 万円) の調査費用が投じられたとされている。オランダ国立図書館は、この研究以外に電子情報保存に係る情報システムを IBM 社から調達している

LTP Study の報告書は既に完成しており、ウェブで公開されている。さらにオランダ国立図書館は、納本のための情報システムにおける電子ジャーナルと CD-ROM のためのワークフローについても研究を行っている。

また、電子情報保存のための情報システムについても取り組みが行われている。現在は、アクセス系のシステム、特に書誌情報へのウェブからのアクセスに注力して投資がなされている。一方で、電子出版物の納本システムについても段階的な開発が行われている。

1999 年、オランダ国立図書館は、電子出版物のための納本システムをベンダーから調達するための検討を実施し、ストレージと管理機能は既存のベンダーの製品で十分であると結論づけた。しかし、長期保存とアクセスに関する技術については、そのまま使える技術がないと判断し、必要な機能について研究することになった。

2000 年 9 月、オランダ国立図書館はオランダ IBM 社と契約を取り交わし、新たな納本システムの構築に着手した。プロジェクトは DNEP-i (The Deposit of Netherlands Electronic Publications-implementation) と名づけられ、OAIS に準拠した保存システムの構築の他に、長期保存とアクセスのための技術研究として先に述べた LTP Study が実施された。2001 年末、最初のデリバリ&キャプチャシステムのモジュールがリリースされたが、システム全体は 2002 年の 10 月に開発が終了し、現在は試行運用が行われている。

DNEP-i プロジェクトの最終的な成果としては、OAIS に準拠した納本システムを稼働させ、LTP Study の研究成果を基にしたシステムモジュールが追加される見込みである。長期保存のためのシステムモジュールは以下の機能をもつものと想定されている。

- ・ 技術変化によってアクセス不能になる可能性がある電子情報を特定する機能
- ・ マイグレーションやエミュレーションを活用し、上記のような危険性を回避するための戦略を実行するための機能
- ・ デジタル・オブジェクトのアクセスの際に必要な再生環境を確認するためのメタ
- ・ データを生成する機能

今後、これらを実装したシステムの構築が行われる。

英国図書館

英国図書館は、電子情報保存に対する積極的な取り組みを実施しており、英国における電子情報保存の啓蒙団体である Digital Preservation Coalition の創設メンバーでもあり、現在でも重要な役割を担っている。

同図書館は、デジタルコレクション、電子情報保存、他機関との連携の重要性を強調する意味も込めて”Strategic Directions, future Strategy for the BL”を策定している。この文書では、英国国内における出版物の収集に力を入れるとともに、国外の出版物についても収集を強化することが提言されている。これに引き続いて、電子情報保存に係るポリシーが策定されつつある。

2001年には、館内あるいは館外への電子情報保存に関するアドバイスの実施と職員の訓練を目的として、保存部門に Digital Preservation Coordinator が設置されている。また、英国図書館内部で”e-fairs”と呼ばれる会合が組織されており、各種の電子情報保存パイロットプロジェクトのデモンストレーション、広報活動、レクチャー、セミナーの開催、組織横断的な勉強会の開催といった活動が行われている。

英国図書館では、電子情報保存に関する様々なパイロットプロジェクトが実施されている。例えば、出版者からの任意で納本される電子出版物における手続と法定納本の規定の検討、ウェブアーカイビングプロジェクト Domain UK、CD-ROM の納本の手続とコストを検証する CD-ROM デモンストレータ等のプロジェクトが挙げられる。

電子情報保存のための情報システムの整備については、外部調達により IBM 社と 10 年契約を結び Digital Library Store (DLS) を構築する予定である。現在、システム設計を実施しており、実際のシステムの開発に当たっては、オランダ国立図書館と緊密に連携する方向で検討が行われている。また、大容量の画像ファイルの保存装置については、ロンドン大学コンピュータセンターと契約を結び、サービス提供を受けている。

2.2. 諸外国における電子情報保存への取り組みについて

ここでは、各国の国立図書館以外で行われている電子情報保存プロジェクトについて取り上げる。取り上げた事例は、一部で国立図書館が関係するものもあるが、ほとんどは政府機関、あるいは大学等の研究機関によって運営されているものである。

2.2.1. 米国国内における電子情報保存への取り組み

米国には議会図書館以外にも、国立農業図書館、国立医学図書館といった国立図書館が存在するとともに、各大学図書館等においても独自に情報技術を活用した取り組みがなされている。政府機関の電子化に関しては、世界でも有数の先進国であり、連邦政府、各州政府が様々な取り組みを実施している。特に政府機関等は、電子的にやり取りされる情報について保存年限が定められていることもあり、電子情報保存の取り組みが不可欠になっている。

また、米国には図書館関係の OCLC (Online Computer Library Center)、RLG (Research Library Group)、CLIR (Council on Library and Information Resources) といった研究団体が存在し、電子情報保存に対する取り組みを実施している。

以下、米国国内における電子情報保存の取り組みについて述べる。

国立医学図書館

国立医学図書館は、電子情報保存には、情報発信者と情報の受け入れ先の協働作業が必要であると認識しており、“PubMed Central”というサービスを提供している。これは、国立医学図書館が医学関係の学術雑誌と各種情報を保存するためのアーカイブの役割を果たすサービスである。

インターネットアーカイブ

1996年に設立された民間の非営利団体“Internet Archive”は、インターネット上の情報をスナップショットとして収集し、保存する取り組みを行っている。最近では、米国議会図書館、スミソニアン博物館等とも連携して、インターネット上に存在する情報の収集・保存に取り組んでいる。

InterPARES プロジェクト

アーカイブ関係の研究者、コンピュータ研究者、公的なアーカイブ機関の代表、民間企業等が集まり、電子的な記録の長期保存のための技術と方法論を確立し、その知識を基に長期保存のためのモデルや方針、標準等を策定することを目的とするプロジェクト。OAIS (Open Archival Information System) を基礎とした電子情報保存のためのシステムを検討するとともに、記録の真正性、アーカイブの持続性、電子政府・e-Business といった領域における基盤技術に電子情報保存のための技術を統合する方策等について検討を行っている。

大学図書館における電子ジャーナル保存

Andrew W. Mellon 財団の助成を受けて、電子ジャーナルの保存プロジェクトが実施されている。参加しているのは、ニューヨーク公共図書館、コーネル大学、ハーバード大学、マサチューセッツ工科大学、ペンシルバニア大学、スタンフォード大学、イエール大学の図書館である。実際には、イエール、ハーバード、ペンシルバニア大学図書館は、出版社と協調して電子的な学術雑誌を保存する取り組みを行い、コーネル大学図書館、ニューヨーク国立図書館は特定の手法による電子的な学術雑誌の保存方法を、マサチューセッツ工科大学は頻繁に更新が行われる電子的な学術雑誌の保存について検討を行っている。

2.2.2. オーストラリア国内における電子情報保存への取り組み

オーストラリアは人口が比較的少ないにもかかわらず、数多くの分野で国際的にも顕著な情報通信関係プロジェクトが行われている。このように高度な情報通信環境では、送受信される情報も大量にのぼり、これらの情報をアーカイブしていくことはオーストラリアの記録関係機関にとっては大きな課題になっている。また、その取り組みは国際的にも高い評価を受けており、様々な場面でオーストラリアにおいて設定されたガイドライン等が引用されている。

オーストラリア連邦政府は、2001 年まで連邦政府のサービスをすべてオンラインで行う計画を実行していたこともあり、政府の各省庁から情報通信を使ったサービスや情報へのアクセスを推進する動きが盛んに起こっている。また、オーストラリアには、電子情報をアーカイブすることに積極的な取り組みを行っている主体が数多くある。例えば、Monash 大学、ビクトリア公文書館、オーストラリア国立公文書館等が挙げられる。

PANDRA (Preserving and Accessing Networked Documentary Resources of Australia)

オーストラリア国立図書館と関係機関が共同で実施しているオンラインの電子情報を選択的に保存するプロジェクト。1996 年から実施されており、オンラインの電子情報を選択的に収集するための取り組みとして国際的にも評価が高い。国立図書館に中央サーバを配置し、地理的に分散している関係機関が、国立図書館により開発されたソフトウェアを利用して情報の選択・収集・提出に協力している。選択収集のためのガイドラインは国立図書館とオーストラリア州立図書館員協会 (Council of Australian State Librarians: CASL) と協議して作成され、オンラインで公開されている¹⁰。

Australian Digital Resource Identifier (ADRI)

オーストラリア国立図書館が、州立図書館員協会、タスマニア州立図書館と協働して進めている電子出版物の一貫した識別を行うための取り組み。この取り組みが実現され

¹⁰ Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia (<http://pandora.nla.gov.au/selectionguidelines.html>) を参照。また本報告書の巻末の付録に PANDRA プロジェクトで活用されているオンライン電子情報の選択の指針を整理している

た場合、将来的に電子情報にアクセスできる可能性が高まる。

Our Digital Island

タスマニア州立図書館が運営主体となっている選択的収集を行うウェブアーカイビングプロジェクト。

Sound Archives

オーストラリアでは音声アーカイブが長年にわたり整備されており、ここ数年は保存のためにデジタル技術を利用するようになってきている。また、オーストラリア国立図書館と Screen Sound Australia は共同で CD-R や磁気テープのライフレンジをテストする実験を実施している。最近では大容量ストレージ・システムが整備されたこともあり、大容量の音声資料を保管しやすくなっている。今後は CD-R に保管されている資料が大容量ストレージに移行される見込みである。

2.2.3. フランス国内における電子情報保存への取り組み

フランス政府は情報化に多大な投資を行っており、例えば、国立情報技術調査研究所 (Institut National de Recherche en Informatique et Automatique) は W3C (World Wide Web Consortium) の 3 つのセンター機関の一つである。

フランス国内における公文書館は、フランス公文書館によって方針が出されているものの、高度に分散化が進んでおり、国立公文書館も 5 つのセンターに分かれているのが現状である。最近では、分散化している公文書館について、公的なものはその管理を支援し、私的なものについては保護を行うためにフランス公文書館法が定められている。

Perennisation des Informations Numeriques

フランス国立宇宙センターが管理する OAIS に関するフォーラムで 2000 年 6 月に設立された。フォーラムの目的は、OAIS 技術標準の策定に対する貢献、OAIS を用いたシステムの実装に関する議論及び各組織間での情報共有である。参加している機関は以下の通りである。

- フランス公文書館 (Archives de France)
- Archive-17
- フランス国立図書館 (Bibliothèque nationale de France)
- 現代アーカイブセンター (Centre des Archives Contemporaines)
- 国立宇宙センター (Centre National d'Etudes Spatiales)
- 原子力エネルギー委員会 (Commissariat à l'Énergie Atomique)
- Groupe Mederic 社
- 国立オーディオ・ビジュアル研究所 (Institut national de l'audiovisuel)
- パスツール研究所 (Institut Pasteur)

公文書関係

フランス公文書館では電子情報を保存するアーカイブのガイドラインを策定している。フランス公文書館は、国立公文書館を始めとして各地域、各部門等のアーカイブを管轄する役割を担っている。フランス公文書館は国立図書館と協働で政府のウェブサイト情報のアーカイピングについて検討を行っている。先にも述べたとおり、フランスの公文書館は分散化が進んでおり、情報技術を活用したサポートが不足しているため、公文書館と国立図書館が連携する意義は非常に大きい。

国立オーディオ・ビジュアル研究所 (L'Institut National de L'Audiovisuel : INA)

INA はフランスにおけるオーディオ・ビジュアル関係の資料に責任を持つ機関である。納本制度を規定した法律によって、フランス国内の公共放送及び民間放送企業 6 チャンネルと 5 つの公共ラジオ放送からの放送内容の納本を受け付けている。また、INA はこれらの機関から提出されるテレビ・ラジオ放送の内容を基にアーカイブを作成し、維持していく責務を果たすことになっている。INA は放送内容の保存において、電子化を活用しており、元来からデジタル情報として作成された情報の受け入れも著しく増加している。INA はフランスにおけるウェブサイトの保存にも取り組んでおり、国立高等教育機関 (Ecole Nationale Supérieure) とウェブサイトの収集を行うためのソフトウェアであるハーベスタを開発している。

大学等研究機関

フランスの大学では、論文を電子的なフォーマットで提出するための仕組みづくりを行うプロジェクトが始まっている。この仕組みは、Microsoft Word のスタイルシートを提供し、提出された情報を XML にリフォーマットすることができる。現在、このプロジェクトはリヨン大学が中心になって実施されており、長期保存についても検討し始めた。また、大学図書館のための調達コンソーシアムである COURERIN では、自らが購読している電子的な学術雑誌を将来的にもアクセスできるようアーカイブしていくことに関心が高まっている。

2.2.4. オランダ国内における電子情報保存への取り組み

オランダ政府は 2002 年まで、国民と政府のやり取りの 25% を電子的に処理する計画を進めている。これに伴い電子政府化や情報化社会の基盤整備等に対する投資が相次いでいる。

こうした状況を背景に電子情報とビジネスの継続性を意識した取り組みとして、電子情報の永続性を研究するプログラム (Digital Longevity Program) が行われている。このプログラムは、5 つのプロジェクトで構成されており、具体的には、電子情報保存のテスト・ベッド、電子情報保存の啓蒙プロジェクト、省庁間連携プロジェクト等が含まれている。

オランダにおいては1991年から保存に関する“The Delta Plan”と呼ばれる計画が存在し、紙媒体の文書の保存に関するニーズ調査を実施している。97年には、図書館の資料を保存するための“Metamorfoze”と呼ばれる国家プロジェクトがスタートしており、このプロジェクトではオランダ国立図書館が組織している National Preservation Office が中心となっている。具体的には、紙媒体の資料のマイクロフィルム化や脱酸性処理を行うとともに、保存手段の一つとして、電子化した資料の複製物を作成するための技術を調査している。

Digital Longevity プログラム

このプログラムはオランダ政府が実施しているもので、先に述べた通り、5つのプロジェクトにより構成されている。具体的には、電子情報保存のテスト・ベッドや省庁連携で電子情報保存への意識を喚起するプロジェクトが挙げられる。これらの他には、政府における電子情報保存のための中心的なデータベースの整備、記録管理システム、記録の品質に関するもの等が挙げられる。プログラムは電子政府を管轄するために設立された ICTU が運営している。オランダ国立図書館もこのプロジェクトに参加している。また、オランダ公文書館は、政府の各省庁から転送される電子文書を保存するために、一時的にオランダ国立図書館の保存システムを共有するよう調整している。

Digital Preservation Testbed

Digital Longevity プログラムの一環として、オランダ内務省、教育・文化・科学省、国立公文書館が中心となって2000年10月から2003年10月までの3年間にわたるプロジェクトが実施された。このプロジェクトは1999年の Jeff Rothenberg の研究で明らかになった課題を引き続いて研究しているもので、マイグレーション、エミュレーション、XMLの3つのアプローチを取り上げ、テキストドキュメント、表計算ソフトのファイル、eメールメッセージ、様々なサイズ、フォーマット、複雑性を持つデータベースの長期保存を研究するものである。3つのアプローチで、それぞれの保存対象に対する有効性、コスト、アプリケーションの機能を勘案して評価を行っている。このプロジェクトでは、次に挙げる結果を意識して取り組みが進められている。

- ・ 政府の各機関において電子情報保存に採用するアプローチに関するアドバイス
- ・ 特定の環境に適した最良の保存アプローチに関する提案
- ・ 保存に必要な機能面での要件
- ・ それぞれの保存アプローチに係るコストモデル
- ・ 保存アプローチに関する意思決定プロセス
- ・ 法制度面で必要となる検討

現在、本プロジェクトでは、電子情報保存に関する情報源を集約したデータベース、マイグレーションとXMLに関する白書がウェブサイトで公開されている。このプロジェクトは英国の PRO(Public Record Office)と米国の NARA(National Archives and Records Administration)とも連携して進められている。

公文書関係

オランダ国内の公文書館は、教育・文化・科学省によって管轄されている。現状では、連邦政府の公文書館、各州の公文書館、各地方自治体やポルダー¹¹毎の公文書館が存在する。政府は、各地方の公文書館を支援する必要があると認識しており、1995年のアーカイブ法で電子的な文書を公文書として扱うことを決定するとともに、20年以上経過したものについては、中央のアーカイブに転送することになっている。さらに2000年に提出された規定では、文書のフォーマットとメタデータが決められている。

Netherlands Institute for Scientific Information Services (NIWI)

NIWI はオランダ王立芸術科学アカデミーの一機関であり、生物医学、社会科学、歴史、オランダ語の文学といった分野の基礎研究データや研究に必要な情報等を提供している。これに加え、オランダのあらゆる科学分野における研究者のデータベースの提供も行っている。現在、同研究機関では電子情報保存が研究の一分野としてとらえられており、オランダの学術研究コミュニティーにおける研究情報のデジタル・アーカイブ・サービスの事業性を評価する目的で、ADA (Archiving Digital Academic heritage)パイロットプロジェクトが実施されている。このプロジェクトでは事業性の評価の他に、どのようなレベルのサービスが学術研究コミュニティーで必要とされているかを検討するという目的も設定されている。

Roquade プロジェクト

Utrecht 大学図書館、Delft 大学図書館、オランダ科学情報サービス研究所(Netherlands Institute for Scientific Information Services) が連携して実施しているプロジェクトで、学術研究におけるコミュニケーションを向上させるための電子アーカイブを構築することを目的としている。このプロジェクトでは、メタデータの付与、品質管理、保存のための情報基盤に関するコスト試算を実施したところ、年間 5,000 項目を電子アーカイブに受け入れると 1 項目の情報につき 29 ユーロ (約 3,741 円) が必要になるという結果が出ている。

ARNO (The Academic Research in the Netherlands Online) プロジェクト

ARNO プロジェクトは、いくつかの大学が連携し、参加機関の科学研究の成果を共有できるドキュメントサーバを構築するプロジェクトである。参加している機関は、アムステルダム大学、Tilburg 大学、Twente 大学である。このプロジェクトは、オランダ電子出版プロジェクトや Open Archive Initiative で検討している技術が活用されている。

¹¹ ポルダーはオランダの干拓地であり、ひとまとまりの区域を指す

2.2.5.英国国内における電子情報保存への取り組み

英国政府は、“Modernizing Government”というスローガンのもとに、急速に行政サービスの電子化を進めている。2004年に目標を定め、行政サービスのインターネットによる提供と、電子的な記録管理に大きな革命がもたらされると考えられている。

英国には、様々な文化機関が存在し、各々の責任に基づいて電子情報保存に対する取り組みを行っている。これらのプロジェクトの中には、国際的に高い評価を受けているものも数多く存在する。また、英国では、高等教育機関と生涯学習機関における情報技術関係の研究開発に資金を助成する団体として JISC (Joint Information Systems Committee)が設置されている。JISC は、高等教育機関、生涯学習機関が、国内だけでなく国際的にも電子情報保存において大きな役割を果たすべきであると認識しており、各種の電子情報保存プロジェクトに資金援助を行っている。

アーカイブの管理者や大学等の研究機関においては、電子情報保存に対する認識が非常に高まりつつあるものの、一方で、重要な利害関係者、例えば、公共機関の上級官僚、国会議員、資金提供団体、出版社等の電子情報保存に対する認識は著しく低い状態にあるのが現状である。

The Digital Preservation Coalition

Digital Preservation Coalition の設立については、1999年にイングランドの中央部の州ウォリックで開催された電子情報保存のワークショップで初めて検討が行われた。その後、国際的な状況も考慮し、英国の電子情報保存を検討するために2001年の7月に正式な組織として設立された。メンバー構成は、正会員 (Full-member)、準会員 (Associate Member)、関係機関で構成されている。設立後、8ヶ月程度で会員数は19まで増えている。参加機関としては、英国国内の各図書館、公文書館の保存等に関わる部署、出版社団体、研究機関、政府機関、各種のサービスプロバイダーとなっている。

組織の設立には JISC と各参加メンバーが資金を提供した。現在も限られた予算の範囲内で、電子情報保存に対する草の根的な啓蒙活動を実施している。現状で、組織には二つのフォーラムが設置されており、一つは電子情報の長期保存、特に OAIS の技術標準と英国における e-science プログラムについて議論するもので、もう一つはウェブアーカイビングについて議論するものである。

Digital Preservation Coalition における長期的なゴールは以下の通りである：

- ・ 加入メンバーの電子情報保存に関する専門性とプロフェッショナルスキルを向上させることを目的に、新たな研究や実践に関する情報発信を行うこと
- ・ 関係者との議論において、電子情報保存の課題を提示し、問題の重要性を理解してもらうこと
- ・ 電子情報の永続性を保証するための資金源を確保するために、適切で十分な

資金源となる主体と協調して活動すること

- ・ 国際的な研究活動を視野に入れつつ、英国における電子情報保存戦略の発展に資するフォーラムを提供すること
- ・ 電子情報保存のためのサービス、技術、標準を開発・促進すること
- ・ 国家的、あるいは国際的なアライアンスを組織し、産業界や研究機関と協働で電子情報保存に取り組み、経験を共有すること
- ・ 組織としてのゴールを達成するために支援を受け、組織の維持のために資金的な支援を取りつけること

公文書関係

英国における公文書記録は、地方政府よりも中央政府が中心となって行われてきた。更なる法制度の検討が必要なものの、公文書法（Public Records Act）は電子的な記録も含むものと考えられている。国立公文書館（The Public Records Office）は電子情報記録のためのプログラムを実施しており、政府各機関に対して必要なガイダンスとツールキットを提供している。さらに、国立公文書館は e-Preservation の一環として、新たなストレージ・システムの外部調達を行うとともに、大規模な政府データベースの保存にはロンドン大学コンピュータセンターとサービス契約を締結している。

JISC Digital Preservation Focus

JISC は各大学に対して公的な資金を活用した資金援助を行っている。特に近年では、電子情報保存の活動に力を入れており、2000 年 6 月には JISC Digital Preservation Focus を組織するとともに、電子情報保存の戦略やガイドラインを策定している。また、関係諸機関と協調して Digital Preservation Coalition を設立している。

JISC の制定している保存戦略は 3 年毎に改訂が行われている。最初の 3 年間は終了し、既に次の 3 年間のアクションプランが提示されている。これは、更なる資金援助を通じて、高等教育・生涯教育分野における電子情報保存に係る新たな活動を行うものである。

現在、活動の中心は、英国国内において、オンラインのみで公開されている学術研究、論文等の収集・保存である。これらは法定納本制度が拡大されなければ収集・保存されることがないため、現状の納本制度によらない取り組みが必要とされている。

基礎研究データの提供

英国では社会科学分野のデータアーカイブが 1970 年代の初頭に設立されているが、それ以外にも自然環境研究協議会（Natural Environmental Research Council）がデータセンターを設立している。現在は、科学研究の電子化に向けて、大規模なプログラムが実施されているが、電子情報保存の調査研究は重要な研究課題として認識されており、様々な協働研究が実施されている。

例えば、1996年には、JISCによって Arts and Humanities Data Service (AHDS) が設立され、人文科学系のデータと保存サービスを提供する機関となっている。

BBC (British Broadcasting Corporation)

BBC は最も古いテレビ及びラジオの公共放送局の一つであり、巨大なアーカイブを保持している。アーカイブを保持する責任は、BBC の憲章にも記載されているが、アーカイブ内の情報は自社利用のみに限られている。

BBC はオンライン化とデジタル配信に力を入れているが、現状では、膨大なコレクションの内、テレビ放送の 5%程度 (ラジオ放送は 5%以下) がデジタル化されているに過ぎない。最近では、新たに制作される番組のほとんどがデジタルで制作されるとともに、アナログメディアはデジタル化して保存することが重要であると考えられている。また、BBC ではテレビとラジオ放送のためのアーカイブに関する 10 年計画が策定されている。

2.3 電子情報保存の取り組みに関する考え方

電子情報保存について先進的に取り組んでいる国の状況を調査した結果、大半の国では電子情報保存が大きな課題と認識され、積極的な取り組みが行われていることがわかった。しかし、どの国も課題の解決策を模索する途上にある。

各国の電子情報保存への取り組みについては、国内の記録を収集管理していく国立図書館、公文書館、政府機関等で顕著な取り組みが見られる場合が多い。あるいは、電子情報保存を研究対象の一つとして捉え、研究を行う大学等の研究機関の活動にも興味深い取り組みが数多く見受けられる。

とりわけ、各国の国立図書館では、その国々で出版される出版物を保存していくラスト・リゾートとしての性格から発信される電子情報を「文化的遺産」の一部と捉え電子情報に組み込む場合が多く、各図書館にとって電子情報保存は喫緊の課題であるとの共通認識がある。

日本国内に目を転じてみても、今回の調査研究で実施した日本国内の関係諸機関に対するアンケート調査から、電子情報を発信する側の意識として、公的な機関による電子情報の保存を望む声が多い。このような結果に、国立図書館等の機関が電子情報保存を積極的に進めていく必要性が示唆されている。

また、諸外国の現状調査からもわかる通り、電子情報保存は単体の機関で解決できる問題ではなく、公文書館、政府機関、公共図書館等の関係諸機関、あるいは電子情報の発信者とも協調して保存活動を進めていくことが不可欠である。このような協調を進めるに当たっては、本文で述べた英国の Digital Preservation Coalition のような電子情報保存のための啓蒙団体を設立することも、一つの有効な手段である。

一方で、国内にとどまらず先進的に電子情報保存に取り組んでいる諸外国の関係機関と連携して経験の共有を図ることも、効率的な電子情報保存を行うに当たっては重要である。特に、電子情報保存に係る情報システムや各種の保存技術については、諸外国の取り組みが先行しており、それらを継続的に調査研究していくことが必要である。

特に、関係諸機関と協調して保存活動に当たる際には、各種の技術やメタデータ等の標準化、あるいは議論を深めるための共通概念の構築、用語の共有等が重要であり、ある種の「標準」を決めることが必要になる。

また、電子情報の保存は情報技術の進展と密接に関係する問題であり、諸外国で実施されている電子ファイルのフォーマットの調査や保存媒体の寿命の調査等、情報技術に関する動向調査を継続的に実施していくことが求められる。

3. OAIS 参照モデルについて

電子情報の長期保存を行うための情報システムを考える際の指針として、ISO の規格にもなっている OAIS (Open Archival Information Systems) がある。本章では、OAIS の概要、提唱されている概念について整理するとともに、OAIS を採用している電子情報保存を行うアーカイブシステムの事例について触れる。

3.1. OAIS の概要

OAIS は、Consultative Committee for Space Data Systems(以下、CCSDS)により検討されているデジタル・アーカイブ・システムの技術標準である。CCSDS は世界各国の宇宙関係機関が協力して設立され、様々な宇宙関係の機器や情報システム等の標準を検討する団体である。1990 年に ISO TC (Technical Committee) 20 の SC(Sub Committee)、13 のワーキンググループとなっている¹²。

OAIS は、当初、衛星から送られてくるデータ等の宇宙におけるミッションに係るデジタルデータを長期間にわたって保存する目的で検討が行われていた。しかし、最終的に策定された標準は、宇宙関係にとどまらず、デジタルデータの長期保存について幅広い応用を意図したものとなった。1999 年 5 月にドラフト版の CCSDS 勧告の初版が出され、2001 年 7 月に改定がなされ、その後 2002 年 1 月に正式な勧告として採択され ISO14721:2002 となっている。

標準の策定に当たっては、その検討プロセスを広く公開するため、オープンフォーラムにおいて検討がなされている。結果についても、幅広く公開、利用がなされることが意図されており、Web サイト等で公開がなされている。

また、技術的には、ファンクショナルモデルとインターフェイスの定義に DFD(Data Flow Diagram)¹³を、データクラスの定義に UML (Unified Modeling Language)¹⁴を使ってモデリングを行っている¹⁵。

OAIS の標準の策定に当たって使用した用語は、電子情報の長期保存を意識して、「伝統的なアーカイブの用語」、「科学的な研究等に活用するデータセンター」、「デジタル・ライブラリー」といった分野から取り入れられている。OAIS が対象とする読者は長期保存が必要な情報を保持するアーカイブの関係者を始め、以下のようなカテゴリーの人々を想定している。

¹² TC20 は、航空機や宇宙における移動・輸送手段に関する標準を検討する。SC13 は、TC20 の枠組みの中で、宇宙関係のデータと情報伝送システムの検討を行うとされている

¹³ 読みは「データフローダイアグラム」、あるいは「ディー・エフ・ディー」。システム間のデータの流れを示す図。データの発生・吸収・処理・蓄積を行うシステムの間を、データの流れを示す矢印で繋いで作成する。データの流れが明確になることによって、効率化しやすい場所を容易に見出せる等のメリットがある

¹⁴ 読みは「ユーエムエル」。オブジェクト指向のソフトウェア開発における、プログラム設計図の統一表記法。Rational Software 社の Grady Booch 氏、James Rumbaugh 氏、Ivar Jacobson 氏の 3 人によって開発された。従来、オブジェクト指向設計の表記法は 50 以上の規格が乱立していたが、1997 年 11 月に OMG によって UML が標準として認定された。Microsoft 社や IBM 社、Oracle 社、Unisys 社等の大手企業が支持を表明している

¹⁵ ただし、本報告書では、説明の便宜を鑑みて UML と DFD は使用せず、OAIS の機能や情報モデルを簡略化した概念図を用いて説明する

- ・ アーカイブのデザイナー
- ・ アーカイブのユーザー
- ・ アーカイブの管理者
- ・ 標準の策定者

3.2. OAIS の目的

OAIS の技術標準は、電子化された情報の「長期保存」に必要なコンセプトと実際の保存活動に適用するためのフレームワーク¹⁶を提供するものである¹⁷。この文脈における「長期保存」とは、技術が変化していく期間を考慮しても十分に足るほど長い期間のことを指している。

OAIS の技術標準では、OAIS を他のアーカイブから区分するための最低限の要件を定めるとともに、OAIS のフレームワークを用いることで、現在、あるいは将来におけるアーカイブのアーキテクチャーと運用を比較検討することを可能にする。また、このような標準を定めることで、将来的に OAIS と関係する標準を策定するための基礎になる、という側面も持ちあわせている。

ただし、OAIS の技術標準は特定の実装を意図して策定されたものではなく、アーカイブの機能を全般的に表現したコンセプトであると認識しておく必要がある。

3.3. OAIS の要件

OAIS の技術標準では、OAIS に基づくアーカイブとその他のアーカイブを区別するための主な要件として、以下の点を挙げている。

- ・ 情報の Producer と交渉して情報を受け取ること
- ・ 長期間にわたる情報の保存を確約する十分な管理機能を保持すること
- ・ 保存されている情報を理解できる特定のコミュニティ (Designated Communities) を決定できること
- ・ 特定のコミュニティに依存せずに保存されている情報を理解できること
- ・ 想定されるあらゆる偶発的な出来事に対して、保存されている情報を保証するための方針書や手続書を作成しておき、それに従うことができること
- ・ 特定のコミュニティが理解できる形式で、保存されている情報を利用できるようにすること

これらの要件を満たすことで、OAIS を他のアーカイブと区別することが可能となる。

¹⁶ フレームワークとは、「ある特定の環境におかれた実体間の重要な関係性を理解し、特定環境をサポートする標準や仕様を策定するための概念である」と定義される

¹⁷ 後述するように、OAIS のフレームワークの中ではデジタル情報だけでなく、物理的なオブジェクトも考慮に入れた検討を行っている

3.4. OAIS 環境モデル

OAIS の環境モデルは Producer (情報の発信者となる主体)、Management (アーカイブを管理する主体)、Consumer (情報を利用する主体) が、以下の図に示すとおり、それぞれに關係して成立している。

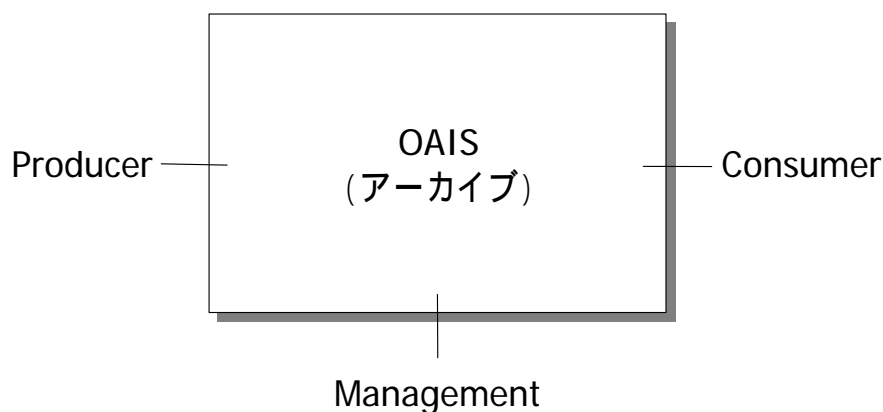


図 3-1. OAIS 環境モデルの概要

それぞれの主体の概要について、以下の表に整理する。

主体の名称	概要
Producer	保存のための情報を OAIS に提供する主体、あるいはそのためのクライアントシステムを指す
Management	数多くの様々な方針のひとつとして、OAIS に関する全般的な方針を定める役割を担う主体を指す
Consumer	OAIS とやり取りを行い、求めている情報を発見し、取得する主体、あるいは同様の役割を担うクライアントシステムを指す

表 3-1. OAIS の環境モデルの関係者

3.5. OAIS の機能モデルについて

OAIS の機能モデルでは、デジタル・アーカイビングにおいて重要な機能を扱っているが、特定の実装に偏らないという OAIS の技術標準の方針から、具体的な実装方式までは立ち入らず、機能の規定を行うにとどまっている。

また、機能モデルの検討に当たっては、分散コンピューティングを念頭において、プロセス間通信、ネームサービス、一時的なストレージの割り当てといった共通のサービスを定めている。以下に、OAIS の機能モデルの概要を図示する。

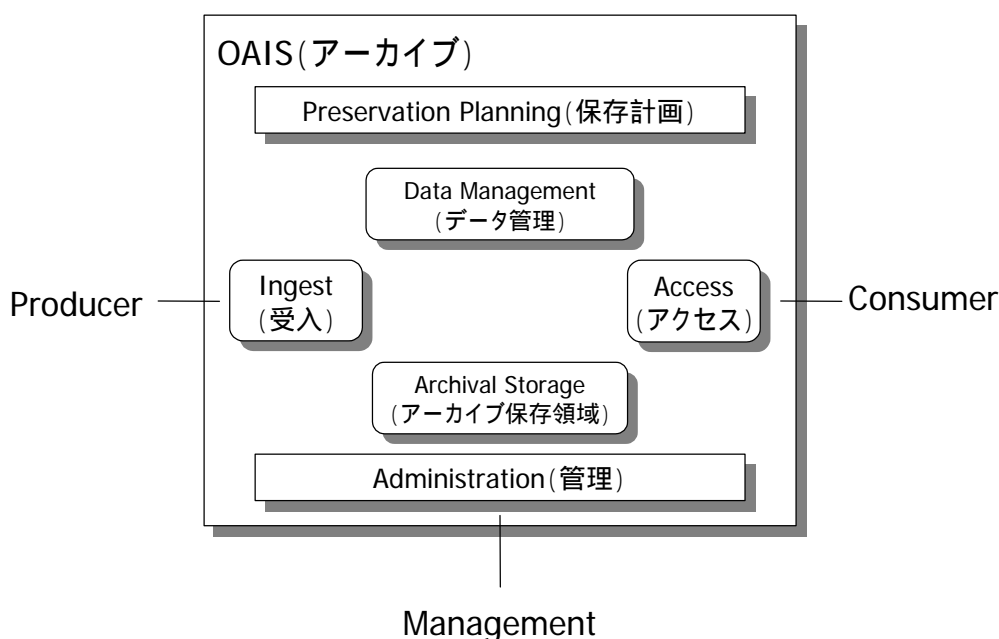


図 3-2. OAIS が規定する機能の概要

以下、それぞれの機能の概要について整理する。

実体の名称	概 要
Ingest (受入)	Producer から SIP (後述する Submission Information Package) を受け付け、受け取った情報をアーカイブ内のストレージや管理のために準備する機能とサービスを提供する
Archival Storage(アーカイブ保存領域)	AIP (後述する Archival Information Package) を保存、管理、引き出す機能とサービスを提供する
Data Management (データ管理)	アーカイブ内の保存対象を特定して存在を示す記述情報 (Descriptive Information) とアーカイブの管理データを管理し、アクセスを行う機能とサービスを提供する
Administration (管理)	アーカイブシステムの全般的な運用を管理する
Preservation Planning (保存計画)	OAIS の環境をモニターし、たとえオリジナルのコンピュータ環境が陳腐化してしまったとしても十分に長い期間にわたり、OAIS に保存されている情報が特定コミュニティのユーザーにアクセス可能であることを保証するように、適切なオプションを提示する
Access (アクセス)	Consumer が、OAIS に保存されている情報の存在を確認するとともに、保存されている場所やアクセスの可能性を決定し、情報を取得する要求を出すことで、実際に情報を取得することを支援する

表 3-2. OAIS 機能モデルにおける機能

3.6. OAIS における「情報」の考え方について

OAIS の技術標準では「情報」に関して、基本的な定義を行い、それを基に様々な情報の変種を設定している。OAIS における情報は、ある特定のデータタイプによって表現されるもの（下図の Data Object を指す）であり、交換可能なあらゆる種類の知識であると定義される。例えば、印刷された書物は「目で見える文字」というデータで表現されており、この印刷物をやり取りすることで情報の交換が可能になる。

このような定義を行った上で、OAIS の技術標準はいくつかの情報に対する考え方を設定する。まず、その全体像について図示する。

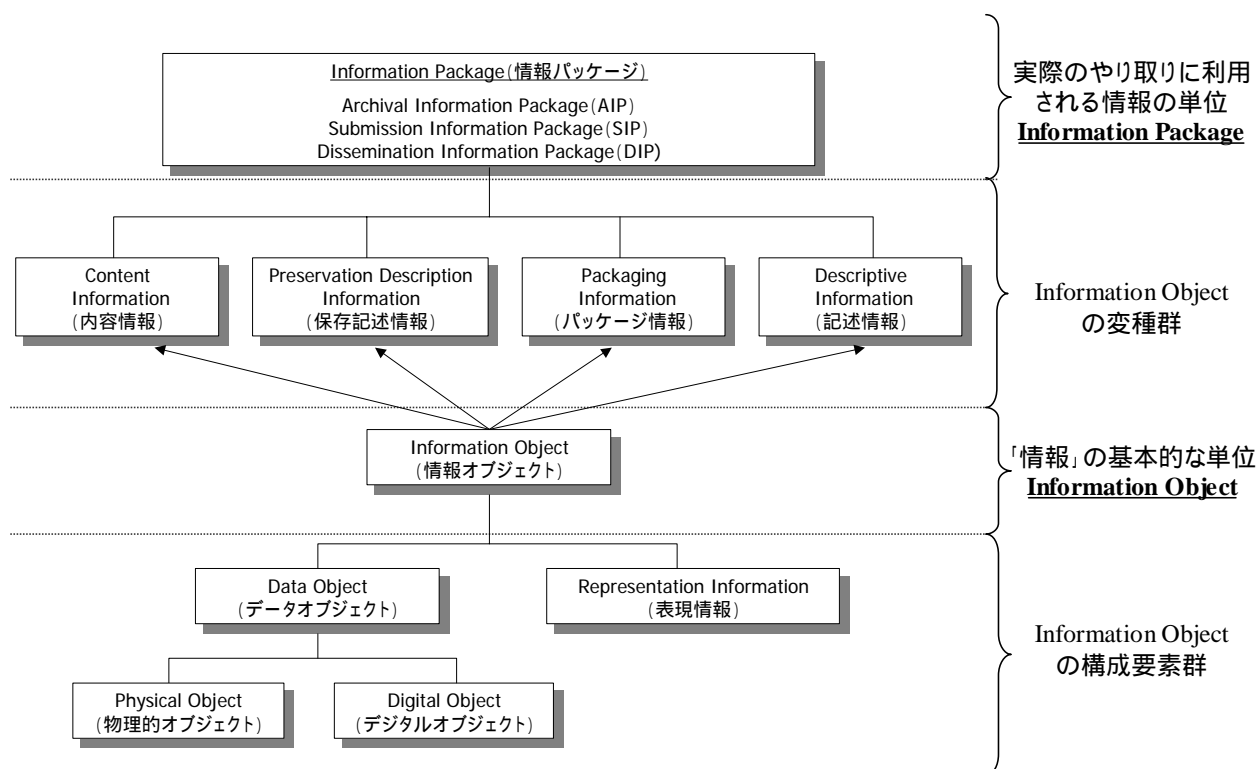


図 3-3. OAIS の情報の全体像

OAIS の情報の基本となるのは Information Object（以下、「情報オブジェクト」と略記する）である。図からもわかるとおり、情報オブジェクトは、Data Object（以下、「データ・オブジェクト」と略記する）と Representation Information（以下、「表現情報」と略記する）からなる。さらにデータ・オブジェクトは Physical Object と Digital Object をその配下に持つ。この二つが並置される意味は、OAIS の考え方がデジタルの情報だけでなく、物理的なオブジェクトの保管にも適用できることを示している。情報オブジェクトは以下に図示する考え方によって説明される。



図 3-4. OAIS における「情報」のコンセプト

先述の通り、情報オブジェクトは、データ・オブジェクトと表現情報から構成されるが、情報オブジェクトは、データ・オブジェクトを表現情報により解釈することで生成されるとされている。本に書かれた文字はデータ・オブジェクトであり、それを目にして解釈することができるのは言語知識という表現情報を人間が持っているからである。これをデジタル情報に置き換えるとデータ・オブジェクトは 0 と 1 の羅列であるビットストリームであり、それを解釈するための文字コード体系等が表現情報となる。

OAIS では、情報オブジェクトに上記で説明した規定を行い、さらにいくつかの情報オブジェクトの変種を規定している。その変種について以下に図示する。

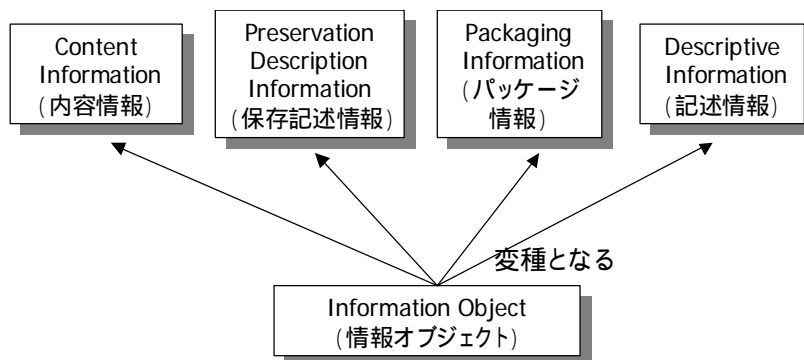


図 3-5 情報オブジェクトの変種

上記のうち、Content Information (以下、「内容情報」と略記する) は保存対象そのものを指し、Preservation Description Information (以下、「保存記述情報」と略記する) は保存に必要な情報を記述する情報オブジェクトの変種である。Packaging Information (以下、「パッケージ情報」と略記する) は、後述するとおり、前者二つをとりまとめ一つの情報パッケージとするための情報であり、この情報パッケージの内容を説明するのが Descriptive Information (以下、「記述情報」と略記する) である。以下に、情報パッケージの概念について図示する。

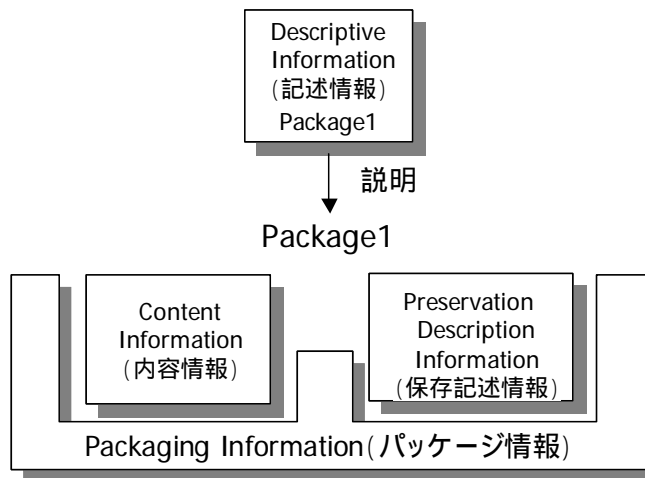


図 3-6. 情報パッケージの概念図

それぞれの情報オブジェクトの変種を組み合わせ、内容情報と保存記述情報をパッケージ情報でくるむかたちで情報パッケージが規定される。この情報パッケージを説明し、アーカイブ内での位置や内容を特定するために記述情報が規定される。このように情報パッケージを規定した上で、図 4-3 の最上段に記した情報パッケージの変種として”Submission Information Package(SIP)”、“Archival Information Package (AIP) ”、“Dissemination Information Package (DIP) ”の 3 種類が規定される。この 3 種類の情報パッケージを先に図 4-2 で触れた OAIS の規定するアーカイブ機能との関係で位置づけると以下の通りとなる。

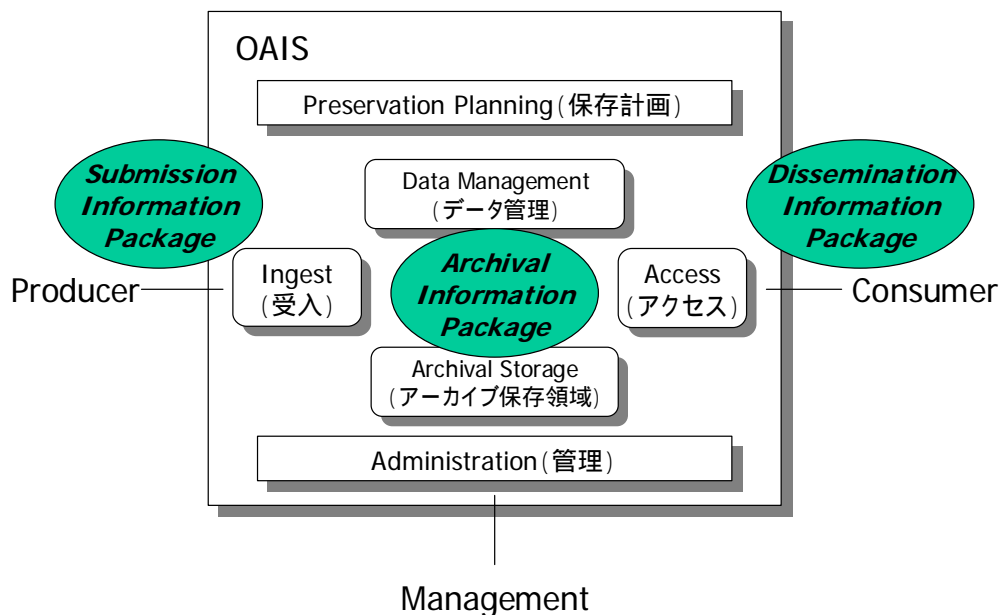


図 3-7. OAIS の機能と情報パッケージの関係

上記の図に基づき、情報パッケージの機能を説明すると以下の通りとなる。

情報パッケージの名称	概要
Submission Information Package (SIP)	Producer と OAIIS 間の交渉において利用され、Producer から OAIIS に送られる情報パッケージ
Archival Information Package (AIP)	情報の保存方法について記した Preservation Description Information を含むもので、保存のための情報パッケージ
Dissemination Information Package (DIP)	ひとつ以上の Archival Information Package を含むもので、OAIIS から Consumer に対して送られる情報パッケージ

表 3-3. 情報パッケージの種別

3.7. 情報の保存について

ここまで OAIIS の「機能」と「情報」の側面について説明してきたが、ここからは OAIIS の技術標準における情報の保存について触れる。OAIIS における情報の保存は「情報自体の保存」と「アクセスの保存」という 2 つの観点から検討されている。大まかに言って「情報自体の保存」については主にマイグレーションの概念を適用し、「アクセスの保存」についてはエミュレーションの概念等を交えて検討がなされている。以下に、情報の保存のために規定されているマイグレーションとアクセスの保存に関する OAIIS の検討内容について概説する。

3.7.1. マイグレーションについて

OAIIS の技術標準における、デジタル情報の長期保存のためのマイグレーションの定義をまとめると以下ようになる。

- ・すべての情報内容の保存に焦点を当てる
- ・新たに実装可能となった情報技術で古い情報技術をおきかえる
- ・移行のあらゆる側面に関する責任やコントロールを OAIIS がすべて行う

このように定義されるマイグレーションを検討しなければならない理由としては、主に以下の点が挙げられる。

- ・情報を保存しているメディアが老朽化する
- ・新たな保存媒体を利用することで、保存にかかる費用対効果を向上させる
- ・Consumer を対象とした新たなサービスへのニーズを満たすために新たな保存媒体を活用する

以下に、OAIIS で検討されているマイグレーションの種別を記す（以下の項目は、情報損失のリスクが低い順に並べている）。

マイグレーションの種別	概要
Refreshment (リフレッシュメント)	内容情報のビットシーケンスの変更を伴わずにメディアの交換を行うこと
Replication (複製)	内容情報のビットシーケンスやパッケージ情報 (Packaging Information) の変更を伴わずに複製を作成すること
Repackaging (リパッケージング)	パッケージ情報のビットシーケンスに多少なりとも変更を伴うマイグレーションの種別のこと
Transformation (変換)	内容情報の変更を伴うマイグレーションであるが、特定のアルゴリズムで変更を復元可能な場合と不可能な場合の二種類がある

表 3-4. マイグレーションの種別

3.7.2. 「アクセス」の保存について

OAIS の技術標準におけるアクセスの保存には、三つのシナリオが想定されている。一つ目は、Application Programming Interfaces (以下、「API」と略記する) を保存すること、二つ目はアクセスのためのソフトウェアのソースコードを保存すること、三つ目はエミュレーションによるアクセスの保存である。

第一の API の保存が想定しているのは、OAIS にアクセスする特定のコミュニティが、OAIS が維持管理するアクセスソフトとしての API を通じて AIP (Archival Information Package) にアクセスするアプリケーションを開発したいと考える場合である。

このようなアクセス手段の提供は、データマイニング等のように、大量のデータを扱うアプリケーションには最適である。API がなければ、Consumer は、AIP として保存されている大量の情報を DIP (Dissemination Information Package) に変換して手に入れることとなるが、このような運用は実質的に不可能である。そのため、API を提供し、Consumer が直接 AIP にアクセスできるようにすることが最適である。

技術環境が変化した場合には、API にラッパーと呼ばれるミドルウェアを追加することでアクセスを保持することができる。

第二のソフトウェアソースコードの保存が想定しているのは、特定のコミュニティがある特定のアプリケーションによって表示される情報の "Look and Feel" (見た目と感覚) を重視している場合である。

概念的な領域を出ないが、OAIS の枠組みにおいては、アプリケーションの変換・提示能力を活用して Consumer が内容情報を閲覧できる環境を提供することを意図している。しかし、技術環境が変化して新たな世代に移った場合は、アクセス環境を提供する

ための新たな方策が必要となる。この際に、ソースコードを入手していればアプリケーションを新たな環境に移植することが可能となるのである。

ただし、このアプローチの実践は、ソースコードの入手の可否に大きく依存する。また、入手が可能であっても、新たな環境への移植の際には、ソフトウェア自体の開発や情報の損失が起こっていないかを確認するために十分なテストが必要となる。そのため費用は多額にのぼると想定される。また、著作権上の問題も不可避である。

第三のエミュレーションによるアクセスの保存は、第二のアプローチで述べたソフトウェアソースコードの保存が困難であるが、特定のコミュニティが情報の”Look and Feel”の保持を必須の条件として挙げる場合である。

エミュレーションは先述の通り、様々な研究が行われている。エミュレーションによって、アクセスのためのソフト、OS、ハードウェアまでも擬似的に環境を再現することができるが、研究としては発展途上段階にあり、今後の更なる取り組みが期待される。

3.8. アーカイブの”Interoperability”（相互運用性）について

OAIS の技術標準では、アーカイブ間の連携に、ユーザーのニーズがあることを想定している。複数のアーカイブ間の相互運用性を検討している。OAIS 環境モデルにおいて、ユーザーが必要とすると想定されるアーカイブ間の連携サービスとしては、以下のものが挙げられる。

ユーザーの種別	必要とするサービス
Consumer	<ul style="list-style-type: none"> ・ OAIS に準拠した複数のアーカイブの情報を横断的に検索するサービス ・ アクセスのための共通的な Package Descriptor スキーマ ・ 共通的な配信のための DIP スキーマ ・ シングルアクセスでグローバルなサイトにアクセスできるサービス
Producer	<ul style="list-style-type: none"> ・ 異なるアーカイブに対して情報を提供する際の共通的な SIP スキーマ ・ 一度ですべての保有情報を提供することが可能なサービス
Manager	<ul style="list-style-type: none"> ・ 高価なハードウェア、ソフトウェア、保存にかかる労力を共有してコストを削減するサービス

表 3-5. ユーザーが求めていると想定されるサービス

上記のユーザーニーズを想定し、OAIS ではアーカイブの連携に 4 つのカテゴリーを想定している。

カテゴリー	概要
Independent (独立)	<ul style="list-style-type: none"> ・ ローカルコミュニティ専用のアーカイブで、アーカイブ間のインタラクションや管理は想定しない

Cooperating (協調)	<ul style="list-style-type: none"> ・ Producer からの情報の提供と、Consumer への情報の配布に関する基準は共通であるが、アクセスには共通の基準を設けない ・ アーカイブの一方が、協調するアーカイブから情報を受け入れることができる場合が想定される (情報を提供するアーカイブが DIP として情報を提供し、受け側のアーカイブが SIP として受け入れる場合を指す)
Federated (統合)	<ul style="list-style-type: none"> ・ 統合に参加しているすべてのアーカイブに対して、共通の手段を用いてアクセスが可能となる ・ 統合に参加するすべてのアーカイブが共通する情報の配布や受け付けの基準はオプションであり、必ずしも求められない
Shared Resources (リソース共有)	<ul style="list-style-type: none"> ・ 主にコスト削減を目的としてアーカイブ間で取り決めを行う ・ 情報の取り込みとストレージ、あるいはアクセスとストレージの間でインターフェイスが必要になる等、アーカイブ内部に適用する共通の基準が必要となる ・ ただし、アーカイブが Consumer に提供する画面等の表示方法は変更されない

表 3-6. アーカイブの連携に関するカテゴリー

このようなアーカイブ間の連携は、相互に意識統一を行うことでも実現可能であるが、一方で、アーカイブの自発的な行動によって達成できるという可能性もあり、多くのアーカイブが「自発的」に共通の基準に従えば、連携が可能になる場合である。

「自発性」によってアーカイブ間の連携を達成するには、連携を構成する各アーカイブの自立性の度合いを検討しなければならない。OAIS の技術標準には、この自立性の度合いに関して三つのレベルを設けている。

- ・ インタラクションがないため、連携もないレベル
- ・ 自アーカイブの自立性をたもったままで、他アーカイブと連携をとるレベル。このレベルでは、連携に参加するために何らかの検討や作業が必要になるが、連携自体については特に意識しなくてもよい
- ・ 契約等の手段により連携を行うレベル。このレベルでは、連携の性質を変える際に交渉が必要になるとともに、後々、連携の性質を変更できるか否かは、交渉の困難さの程度に依存することになる

3.9.OAIS の適用例

OAIS のコンセプトを電子情報保存のための情報システムに適用した主な例として、以下のものが挙げられる。

- **NEDLIB (Networked European Deposit Library)**
オランダ国立図書館が中心となり、ヨーロッパ各国の図書館が電子情報の長期保存について検討を行ったプロジェクト。電子出版物の納本システム(DSEP: Deposit System for Electronic Publications) のシステムアーキテクチャーに OAIS のコンセプトが適用されている
- **オーストラリア国立図書館**
オーストラリア国立図書館における電子情報保存のためのシステムの実装に OAIS が適用されている
- **Cedars (CURL¹⁸ Exemplars in Digital Archives)**
リーズ大学、ケンブリッジ大学、オックスフォード大学が中心となって実施した電子情報の長期保存を目的とする研究プロジェクト。OAIS を適用して、分散型デジタル・アーカイブ・システムのプロトタイプを構築した
- **NSSDC (National Space Science Data Center)**
米国航空宇宙局 (NASA) における電子情報のアーカイブシステムのアーキテクチャーに OAIS が適用されている
- **SIPAD (Systeme d'Information, de Preservation et d'Acces aux Donnees)**
フランス宇宙局のプラズマ物理学用デジタル・アーカイブの基礎的なアーキテクチャーに OAIS が採用されている
- **InterPARES (International Research on Permanent Authentic Records in Electronic Systems)**
ブリティッシュコロンビア大学図書館が中心となり、情報システムにおける記録の真正性を長期的に保証するための研究を行った国際的なプロジェクト。長期保存を行う情報システムの検討の叩き台として OAIS が採用された

¹⁸ Consortium of University Research Libraries の略

3.10. OAIS の概念を適用したシステムの事例

ここでは、OAIS の影響を受けた電子情報の長期保存システムの実例として、MIT (マサチューセッツ工科大学) の DSpace を取り上げる。

3.10.1. MIT の DSpace の概要

DSpace は、MIT (マサチューセッツ工科大学) とヒューレット・パッカード社が共同で開発したデジタル情報の長期保存のための情報システムである。DSpace は、情報中心 (Information-Centric) の考え方をとっている。情報システムは変化しても、「情報」自体はその変化に影響を受けずに保存され続けられるようにしなければならないという考え方である。

DSpace のシステム仕様は、広く一般に公開されており、幅広い業界でアーキテクチャを採用してもらうことを意識している。また、DSpace を構成するソフトウェアは、ダウンロードして利用することが可能であるため、Linux と同様に、オープンソースコミュニティの中で DSpace のソフトをベースに新たなアーカイブソフトウェアが構築される可能性も考えられる。DSpace は、アーキテクチャの適用にかかる障害を低減するため、システムを構成するソフトウェアに、数多くのオープンソースやフリーウェアのソフトが活用されている。また、採用されている各種の技術についても、OAIS や Dublin Core 等の業界標準となっているものを取り入れている。

3.10.2. DSpace のシステムアーキテクチャの概要

DSpace は、システムアーキテクチャとして、Web システムでは標準となっている三層モデルを採用している。以下にシステムアーキテクチャのイメージ図を示す。

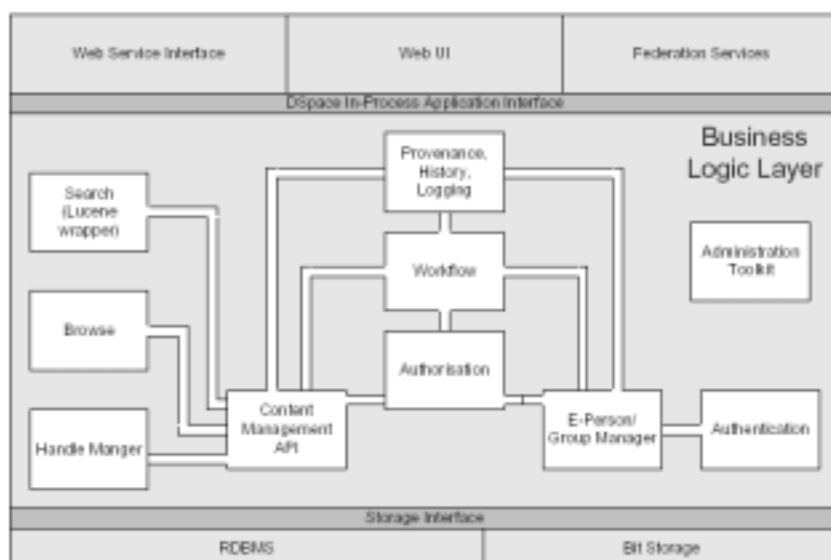


図 3-8. DSpace のシステムアーキテクチャ

それぞれの階層はインターフェイスでつながれており、最も下層がデータの保存・管理を行うストレージ層であり、次に各種の処理を行うビジネスロジック層、そして一番上にユーザーインターフェイス等を提供するサービス層が配置されている。

3.10.3. DSpace が採用している情報モデル

次に、DSpace が採用している情報モデルを以下に記す。

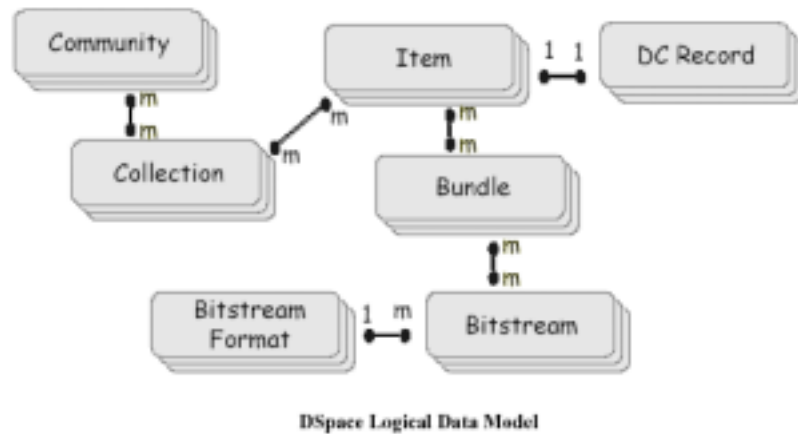


図 3-9. DSpace の情報モデル

情報は「アイテム」として管理されるが、アイテムは複数のコレクションに、コレクションは複数のコミュニティーに対応している。すなわち、アイテムは保存されている対象そのものであり、これが特定のコミュニティー（Designated Community）の興味関心により論理的にコレクションとしてまとめられるということが出来る。コミュニティーは、MIT の各学部、世界中の研究者等を指している。

上図におけるアイテムは、情報の基礎となっている複数のビットストリームをまとめる Bundle と関係づけられるとともに、ユーザーが探索を容易にするためにダブリンコアに準拠したメタデータが付与される。最下層のビットストリームには、OAIS で規定されていたようにデータ・オブジェクトそのものであるビットストリームに表現情報であるフォーマット情報を付与する形態をとっている。

3.10.4. DSpace の機能

DSpace の主なサブシステムについて以下にまとめる。

サブシステムの名称	概要
データベース管理システム (DBMS)	・ オープンソースの DBMS である PostgreSQL ¹⁹ を採用している
ストレージ (Bit stream Storage)	・ シンプルな API でストレージ・デバイスを接続が可能なものを採用している
個人領域 (Personal Workplace)	・ DSpace は OAIS に規定されているユーザー ²⁰ からの情報の提出を重視しており、未提出のものや編集集中のものを個人的に保管しておく領域を提供する
インデックスと検索 (Index and Search)	・ JAVA ベースで開発されたフリーウェアの検索エンジンを採用し、インデックスの追加と検索性能の向上を可能にしている
認証 (Authentication)	・ 個人ベースで認証を行い、細かなアクセス権限を割り振ることで、情報の提出や情報毎のアクセス可否等の管理が可能である
履歴とログ管理 (History and Logging)	・ アーカイブの管理を容易にするために、情報が提出された際等のイベントや、その他の細かなログ情報を管理するサービスを提供する
ユーザーインターフェイス	・ JAVA ベースの Servlet ²¹ や JSP ²² を適用し、拡張性や動的なユーザーインターフェイスの生成を可能とした
配布 (Dissemination)	・ 現在は、サーバー上のファイルをそのまま配信しているが、将来的にはアクセスデバイスの種別を判定し、各デバイスに適した情報を提供できるよう改良する予定である

表 3-7. DSpace の主なサブシステムの概要

¹⁹ 読みは「ポストグ्रेसーケル」。オープンソースのオブジェクトリレーショナルデータベース管理システム(ORDBMS)の一つ。BSD ライセンスによるフリーソフトウェアである。カリフォルニア大学バークレー校で開発されたデータベースシステム POSTGRES を元に、世界中のプログラマたちにより開発された

²⁰ OAIS における Producer (情報の提供者) を指しており、DSpace では、MIT の研究者等を想定している

²¹ 読みは「サープレット」(ジャバサープレットともいう)。Web サーバー上で実行されるモジュール(部品)化された Java プログラムで、サープレットを追加することにより、Web サーバーの機能を拡張することができる。サープレットは Java 言語で記述されているため、特定の OS やハードウェアに依存することがなく、サープレット API を実装したあらゆる Web サーバーで稼働させることができる

²² 読みは「ジェーエスピー」。Java 言語を利用して Web サーバーで動的に Web ページを生成し、クライアントに送信する技術。HTML ファイルの中に Java プログラムを埋め込んでおき、クライアントの要求に応じてプログラムを実行、処理結果のみをクライアントに送信する。結果は通常の HTML 形式になるため、Web ブラウザに特殊な機能を組みこむことなく Web アプリケーションを構築できる

3.11. 電子情報保存に係る情報システムの考え方について

OAIS は各国でも広く認識されており、ISO 規格にも採用される等、非常に重要な概念である。諸外国における電子情報保存に係る各種のプロジェクト、電子情報保存のためのシステムにおいて、OAIS で規定されている用語や概念が「共通の言語」として語られていることからこれは明白である。

一方で、OAIS は電子情報の長期保存を行うためにアーカイブが果たすべき責任、長期保存のコンセプト、アーカイブが備えるべき機能、情報のあり方、情報とアクセスの保存、アーカイブ間の連携のあり方等を規定したものであるが、その大半は OAIS の技術標準の文書でも述べられている通り「抽象的な仕様」であり、「特定の実装方式に偏らない」ものであることを認識しておかなければならない。OAIS は長期保存のための情報システムについて議論する際の共通の基盤でありシステムを構築する際に参考にされる「モデル」である。

このような OAIS の性格に鑑みて、その概念や規定に基づき電子情報保存のための情報システムに係る特定の実装方式を検討する際には、それぞれの主体の実情にあった検討や議論が十分に尽くされる必要がある。フランスにおける取り組みのように、関係諸機関と議論のためのフォーラムを設置して情報共有を行いつつ議論していくことが非常に重要であり、日本においてもこのような取り組みが求められる。

また、OAIS は諸外国において既に電子情報保存のためのデファクトスタンダードとなっており、保存のための情報システムを構築する際にこれを取り入れている事例が数多く見受けられる。このような先行事例を数多く調査し、その利点を抽出した上で、日本の現状にあった保存のための情報システムを構築していくというアプローチが必要である。

4. 電子情報保存のためのメタデータについて

本章では、メタデータの概念整理を踏まえ、電子情報保存に係るメタデータの性質と具体的な保存メタデータの事例、メタデータ付与に関する考え方と具体的なメタデータ付与に係る事例について整理する。

4.1. メタデータの概念整理

メタデータを端的に説明する言葉は、「情報の情報」である。すなわち、メタデータは、ある特定の「情報の内容」、「特徴」、「情報の種別」、「情報の所在」等を示すための情報である。図書館における書誌データはメタデータ的一种である。「本」というある種の情報(の集合)の「分類」や「所在」といった情報を表現するものだからである。

デジタル時代の図書館とメタデータの関係を考える際に、メタデータは情報の探索、発見、保存等に関して大きな役割を果たす。以下に、一般的なメタデータの分類について図示する²³。

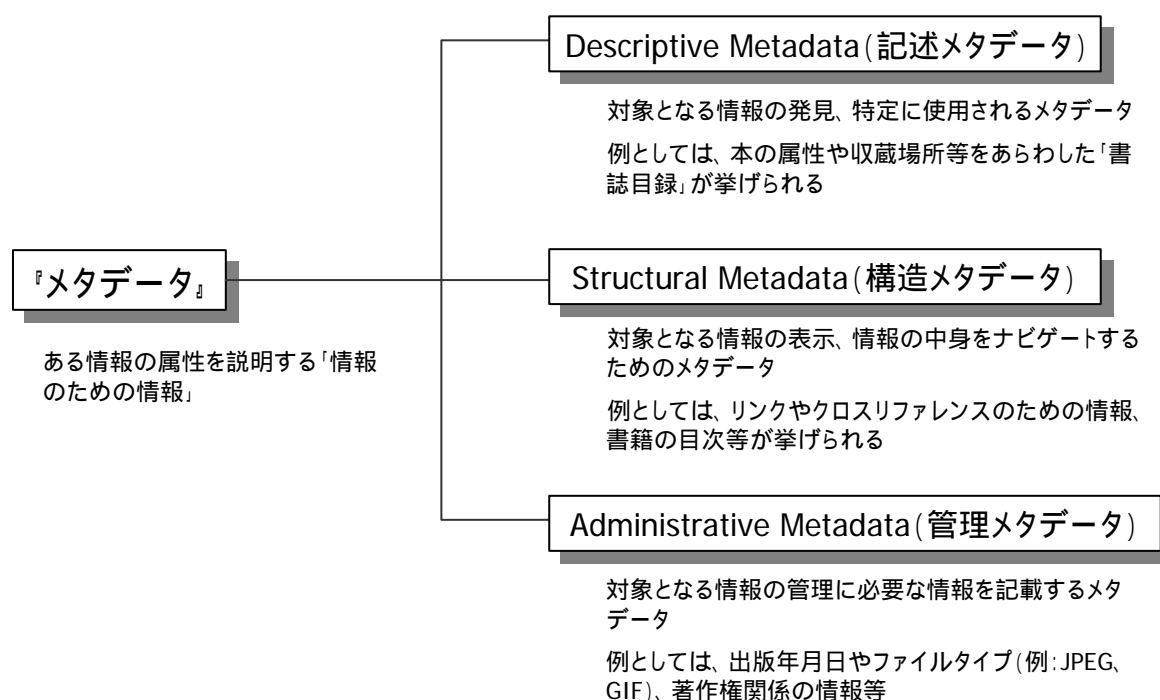


図 4-1. メタデータの概念整理

ただし、上記の分類はあくまでも便宜的なものであり、必ずしも明確に区別する必要はない。なぜなら、実際のメタデータの活用にあたっては、それぞれのメタデータの特質が混在する場合が多いからである。

²³ この分類は、Digital Library Forum System Architecture Committee が行った Making of America (MOA) のテスト・ベッドにおける白書による

4.2. 電子情報保存におけるメタデータの概要

電子情報保存に関するメタデータは、図 5-1 におけるメタデータ概念整理の三番目に記載されている Administrative Metadata (管理用メタデータ) の色彩が強いといえる。電子情報の保存にかかるメタデータの主な特徴をまとめると以下の通りとなる。

- ・ 保存に関する意思決定や、保存のための行動を支援するための技術的な情報
- ・ マイグレーションやエミュレーションといった電子情報保存のための活動を記録するための情報
- ・ 長期間にわたって保存していく電子情報の真正性を保証するための情報
- ・ 保存管理、知的財産権の管理についての注意事項を記載するための情報

電子情報の保存を円滑に行うためには、以上で述べた性質を持つメタデータを整備する必要がある。

4.3. 電子情報保存におけるメタデータの事例

電子情報保存に係るメタデータに関する代表的な取り組みとして、OCLC (Online Computer Library Center)²⁴と RLG (Research Library Group)²⁵が 2000 年から 2001 年に行ったプロジェクトが挙げられる。OCLC と RLG は、OCLC/RLG Working Group on Preservation Metadata (以下、「WG」と略記する) を発足させ、3 つの代表的な電子情報保存のための取り組みの過程で策定されたメタデータの特徴を OAIS の情報モデルを軸に比較検討した。その結果として共通する電子情報保存のためのメタデータの策定を行った。

以下に、WG で取り上げた電子情報保存のための取り組みの概要について整理する。

WG で扱ったプロジェクトの名称	参加国/団体	概要
Cedars	リーズ、ケンブリッジ、オックスフォード大学	電子情報の保存戦略、方法論、課題等を明確化することを目的に、OAIS (Open Archival Information System) をベースに、メタデータ仕様を策定し、分散型アーカイブのプロトタイプシステムを構築した。その他に、知的財産権、コレクションマネジメント、技術戦略等に関する研究を実施した

²⁴ 1967 年に設立された非営利団体。86 の国と地域における 42,289 の図書館が加盟している。世界規模で情報へのアクセスを向上させるとともに、図書館のコストを低減させるために、各図書館やそのユーザーに対するサービスを提供することを目的としている

²⁵ 1974 年に設立された非営利団体。160 に及ぶ世界各国の大学図書館、国立図書館、公文書館が加盟している。研究のためのコレクションとそれらの利用を支援するためのサービスを研究するパイオニア的な団体である

NEDLIB	オランダを中心として、フランス、ノルウェー、フィンランド、ドイツ、ポルトガル、スイス、イタリアの国立図書館等	テストベットや各種のデモンストレーションを通じて、デジタル情報のための納本システムに必要な情報システムの要件、アーキテクチャー等を調査し、各種のガイドラインを制定した
オーストラリア国立図書館	オーストラリア国立図書館	電子情報の保存に先進的な取り組みを行っている国立図書館。電子情報保存のシステムを整備するとともに、各種の電子情報保存のためのガイドラインの策定を行っている

表 4-1. WG で取り扱ったメタデータを策定したプロジェクトの概要

上表に記載した各主体のメタデータの比較は、電子情報保存のための情報システムのコンセプトを定めた OAIIS(Open Archival Information System)の標準仕様に沿って行われた。

OCLC/RLG の保存メタデータの作成において OAIIS を基準とする意義は非常に大きい。OAIIS は電子情報の長期保存にかかるデファクトスタンダードであり、これを基準に各主体のメタデータを比較することで、共通するメタデータ要素を確認できるとともに、不足している部分を補強することができる。結果として、非常に汎用性の高い標準的な電子情報保存のためのメタデータを策定することが可能になる。

4.3.1.各主体の電子情報保存に係るメタデータの比較

まず、各主体におけるメタデータの概要について整理する。

各主体の名称	電子情報保存に係るメタデータの概要
Cedars	<ul style="list-style-type: none"> 本プロジェクトで検討された長期保存のためのプロトタイプシステムに実装する目的でダブリンコアメタデータをベースにメタデータ的设计を行った 特定の電子情報に偏らない抽象度の高い内容で、記述、管理、技術、法的管理の側面にまで踏み込んで、保存している電子情報へのアクセスを容易にするよう設計されている
NEDLIB	<ul style="list-style-type: none"> 激しい技術変化の中で、大量のデータを扱うための必要最小限の核となるメタデータを策定することを目的としている 特に技術の陳腐化への対応をどのように行うかといった部分に焦点を当ててメタデータ的设计が行われている
オーストラリア国立図書館	<ul style="list-style-type: none"> 円滑な電子情報の保存管理を行うことを目的に策定された 情報の発見や探索といった側面には焦点を当てていないが、元々デジタルで作成される情報 (Born Digital) とアナログ媒体からデジタルに変換された情報 (Non-Born Digital) のどちらにも利用できるよう配慮して設計されている

	<ul style="list-style-type: none"> ・ 保存対象とする情報の階層として、コレクション、オブジェクト、サブオブジェクトを設けて、特にオブジェクトへの保存情報の記述を重視している²⁶
--	--

表 4-2. 各主体の電子情報保存にかかるメタデータの特徴

上表からもわかるように、各メタデータは設計の際に重点の置き方が異なっている。しかし、比較に必要な共通点も存在する。以下に、主な共通点についてまとめる。

- ・ 保存に関するメタデータの目的を、アーカイブの管理者が保存活動の意思決定を行う際に、必要十分な情報を記載することとらえている
- ・ 保存対象とする情報に特定の種別を設けていないため、メタデータを利用できる情報の種別が非常に多い
- ・ 電子情報保存にかかる技術的プロセス（マイグレーションやエミュレーション等）について、特定技術への限定を設けていないため、幅広い保存技術の適用が可能である

上記のような特徴を持つ各主体のメタデータについて、OAIS の枠組み、すなわち保存において重要な役割を果たす Content Information（内容情報）と Preservation Description Information(保存記述情報)を軸に、OCLC/RLG ではメタデータの検討を行った（詳細は付録の 3 を参照）。

OCLC/RLG の電子情報保存のためのメタデータは、その策定に当たって、電子情報保存では先進的な取り組みを行っている英国図書館、オーストラリア国立図書館、米国議会図書館等の各国国立図書館、ハーバード大学、コーネル大学、カリフォルニア大学バークレー校等の大学図書館、NEDLIB、Cedars 等の電子情報保存プロジェクト等から広範に参加者をつのり、そのコンセンサスを得て策定がなされた。そのためこれから実施される電子情報保存の取り組みで、保存のためのメタデータを考えるための枠組みとして、活用が図られるものと推測される。

²⁶ コレクションはオブジェクトの集合であり、オブジェクトは保存対象となる個々の情報を指す。サブオブジェクトはオブジェクトを構成するファイル等が挙げられる

4.4. 電子情報保存におけるメタデータ付与について

電子情報保存のためのメタデータは、その要素自体が技術的な性格を帯びているため、通常のメタデータよりも専門的な知識と金銭的、時間的コストが必要になる。以下に、メタデータ付与のためのコスト削減の方策として考えられるトピックスについてまとめる。しかし、メタデータの付与に関するコスト面での研究に関しては、まだ実践的な研究がほとんどなく、明確な基準や示唆が得られていないのが現状であり今後の研究が期待される。

ここでは、電子情報の保存に係るメタデータの付与のためのコスト削減の方策として考えられるポイントをまとめる。

メタデータの共有

電子情報を保存しているレポジトリが、各々が保有するメタデータを共有する、あるいは複製することで、メタデータの作成のコストが大幅に削減できると見込まれる。これは、メタデータの作成にかかるコストを分担するための取り組みと言い換えられる。そのためには作成されたメタデータを広く公開し、アクセス可能にするための取り組みが推進される必要がある。このような取り組みについては、米国において Open Archive Initiative が作成したメタデータ・ハーベスティング・プロトコルが挙げられるが、これについては後述する。

メタデータの共有、複製を実現するには、前節で述べた OCLC/RLG の取り組みにみられるように、非常に汎用性の高い、標準的なメタデータの作成が必要となる。このようなメタデータの作成に当たっては、各メタデータ要素の「記述の粒度」、「入力が必須か、任意か」といった事柄を考慮しなければならない。ただし、メタデータの設計は、保存管理の際の有効性と付与にかかるコストを秤にかけて、十分な検討がなされなければならない。また機種に依存しない汎用性とデータの互換性を確保するために XML で提供されている XML ネームスペース²⁷を活用することも考えられる。

メタデータに XML を活用する動きとしては、世界的なメタデータの標準化団体である Dublin Core Metadata Initiative、米国議会図書館と DLF が中心となって推進する METS²⁸、インターネット技術の標準化を推進する W3C で提唱されている RDF²⁹ といった世界的な動きがみられる。

²⁷ W3C で検討されている XML 関連の技術標準。英語では "Namespace in XML" とされている。XML はユーザーがタグを自由に定義できるため、異なる人々の間で同じ名称のタグが作成されてしまった場合、コンピュータが同名称のタグを識別できない可能性がある。このような不都合を解決するために、同名称のタグがある場合は、XML インスタンス(タグでマークアップした XML の実データ)中に、どの DTD(XML のタグ定義を行う文書)を参照すべきかコンピュータに指示をするネームスペースの宣言が行われる

²⁸ Metadata Encode and Transmission Standard の略。米国議会図書館の Network Development and MARC Standard Office と Digital Library Federation が中心となって標準を策定している

²⁹ Resource Description Framework の略。特定のアプリケーションや知識領域を前提とせずに、相互運用が可能な形で「リソースを記述」するための標準的な枠組みを提供する試み。World Wide Web 上で効果的なデータ交換を行うためのメタデータを提供する。RDF は XML を活用しており、特定のアプリケーションプラットフォームに依存しないため、これを活用することによって、異なるアプリケーションが相互に通信できるようになると考えられている

メタデータの付与の自動化

メタデータの付与を可能な限り、コンピュータで自動処理することでコストの削減を図る方法が考えられる。特に、保存メタデータのうち、人間による高度な判断が不要な項目、例えば、マイグレーションやエミュレーションに必要な情報を記述するメタデータについては、自動化が可能である。

ただし、メタデータの付与を自動化する場合には、メタデータ自体の設計に際して、管理の有効性を保ちつつ、機械的に付与が可能な項目を設計する必要がある。またメタデータの円滑な入力と、管理の確実性を高めるために入力内容の確認を行うツールの開発、あるいはそのようなツールを、ワードプロセッサや電子化ソフトといった、情報の作成を行うためのアプリケーションに組み込むことが必要になる。

情報の提供者との協調作業

情報の提供を受け、それを保存していくアーカイブの立場に立って考えれば、情報の提供者と協力してメタデータの作成にかかるコストを分担するという選択肢もある。

図書館の文脈で考えると、情報の提供を行う出版社等の関係者と議論し、協定等を締結することで、納本制度等により提出される情報に、あらかじめ出版社等がメタデータを付与するという方法もある。こうすることで、図書館側の受け入れ作業の負荷が軽減されるとともに、出版社にとっても情報の確実な保存が可能になる。

4.5. 電子情報の保存に係るメタデータの付与に関する事例

ここではメタデータ付与プロセスに関する事例として、以下のものを取り上げる。

- ・ メタデータを共有するための取り組みである OAI (Open Archive Initiative)
- ・ メタデータ付与の自動化を実践する METAe プロジェクト
- ・ 出版者との協働作業及びメタデータの共有を前提とした NEDLIB システムモデル

以下、順にそれぞれの概要についてまとめる。

4.5.1. OAI (Open Archive Initiative) について

OAI (Open Archive initiative) は、学術研究コミュニティにおける円滑な情報流通を行うために学術論文を配信する Pre-Print / e-Print サーバーの情報共有を目指して、技術標準を策定とそれ普及させるための活動として 1999 年に設立された。設立には、Hervert Van de Sompel、Rick Luce、Paul Ginsparg の 3 人が中心となり、DLF (Digital Library Federation)、CNI (Coalition for Networked Information)、NSF (National Science Foundation) がスポンサーとなっている。

OAI が策定している OAI-PMH (OAI Protocol for Metadata Harvesting) は、ある特定の学術研究コミュニティにとどまらず、デジタル・ライブラリー等の各種のデジタル・

アーカイブに対して適用が可能な技術であると考えられており、米国、ヨーロッパにおいて普及のための活動が展開されている。以下、OAI-PMHの概要について図示する。

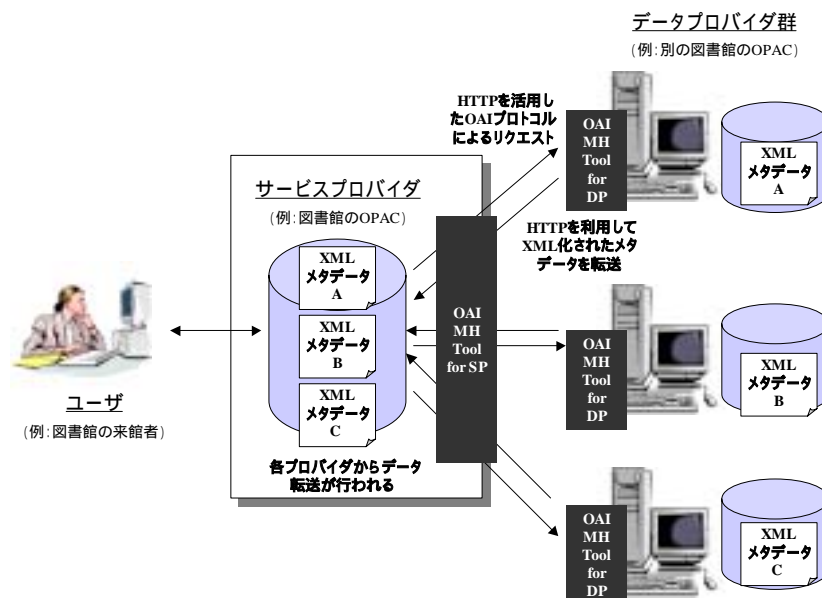


図 4-2. OAI-PMH の概要

上図に示した通り、サービスプロバイダーが OAI-PMH を活用し、HTTP をベースとした通信を行い XML 形式でデータプロバイダーのメタデータを収集し、共有することが可能となる。共有できるメタデータの種類としては、MARC、UNIMARC 等の図書館が活用するメタデータも収集・共有が可能である。

OAI-PMH プロトコルではサービスプロバイダー、データプロバイダー間の通信に際して“Verb”と呼ばれる 6 つのリクエストが使われる。以下にその種別と概要をまとめる。

リクエストの種別	概要
GetRecord	データプロバイダーのレポジトリから個別のメタデータの読み出しを要求する際に利用するリクエスト
Identify	データプロバイダーのレポジトリに関する情報を読み出す際に利用されるリクエスト
ListIdentifier	データプロバイダーのレポジトリから読み出そうとしているメタデータの識別子に関する情報を要求するためのリクエスト
ListMetadataFormats	データプロバイダーのレポジトリがサポートしているメタデータのフォーマットについての情報を要求する際のリクエスト

ListRecord	データプロバイダーのレポジトリから、複数のメタデータを読み出す際に利用されるリクエスト
ListSets	データプロバイダーのレポジトリに格納されているメタデータの階層構造に関する情報を要求するためのリクエスト

表 4-3.OAI-PMH のプロトコルの種類

現在、OAI-PMH は記述メタデータ共有で幅広く使われているがこの仕組みを電子情報保存に応用することで保存のためのメタデータの付与が効率化できる可能性もある。

4.5.2. METAe プロジェクトについて

METAe プロジェクトはヨーロッパ7カ国と米国の大学、図書館、ITベンダー等、総勢14の主体が協力して、2000年から実施しているプロジェクトである。このプロジェクトでは、紙媒体の資料を電子化するプロセスで、メタデータの付与を自動化する試みを実施している。このプロジェクトで作成されたソフトウェアが普及すれば、電子化された資料の相互運用性の確保にもつながる。現在、プロジェクトの終了予定である2003年8月に向けて、メタデータの付与を自動化するツールの作成が行われている。METAe プロジェクトのアプローチは、大まかに以下の2点に絞られる。

- ・ 紙資料の電子化における初期段階で、自動的にメタデータ（記述、構造、管理それぞれのためのメタデータ）を取得する高度な仕組みを構築する
- ・ 電子化の最終段階で、各種の標準に則った Archival Information Package (AIP) を生成し、電子化された資料のアーカイブにおける管理を容易にする

METAe プロジェクトでは、電子化のためのツールを複数作成しているが、メタデータ付与の自動化に寄与するものとして、“METAe Engine”が中心となる。その主な機能は以下のものである。

- ・ 紙の資料を電子化するためのスキャナ
- ・ 取り込んだ画像の処理機能
- ・ 図書館の所有する書誌情報 (MARC21) からデータを取り込み、記述メタデータの付与を行う機能
- ・ 構造メタデータ、管理メタデータを自動的に生成し、付与する機能
- ・ OCR³⁰処理機能

これらの機能を図示すると以下の通りとなる。

³⁰ 読みは「オー・シー・アール」。Optical Character Reader の略。手書き文字や印字された文字を光学的に読み取り、前もって記憶されたパターンとの照合により文字を特定し、文字データを入力する装置。スキャナで読み取った画像から文字を識別して文書に変換する OCR ソフトもある

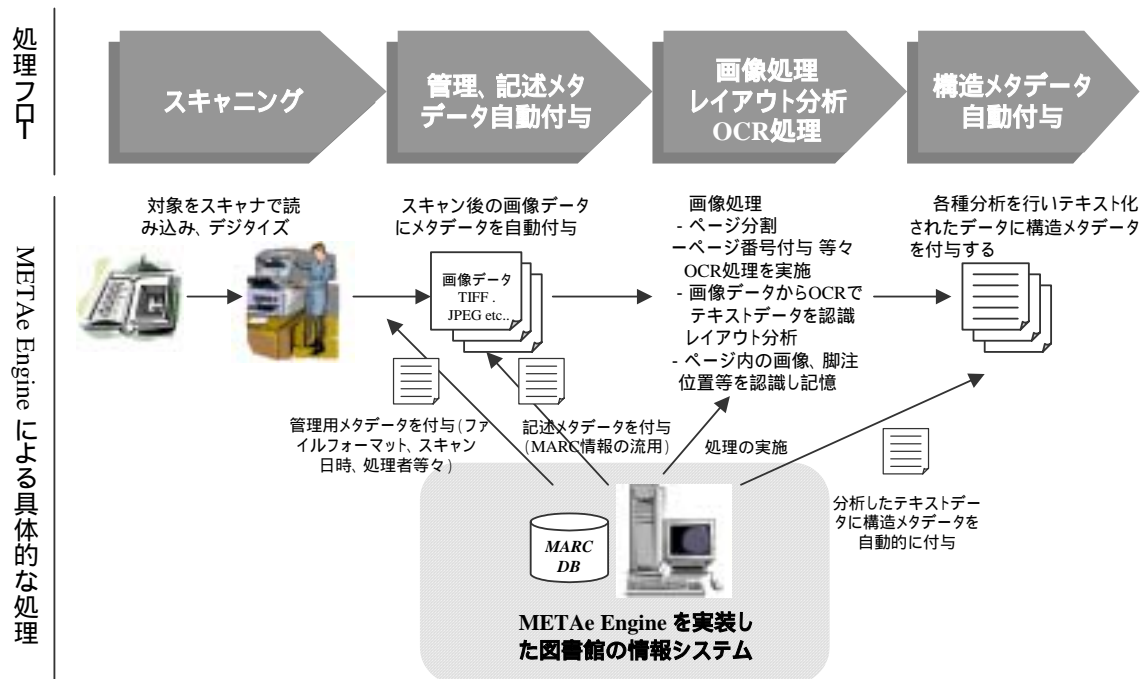


図 4-3. METAE Engine の機能の概要

上図に示した通り、METAE Engine の機能を活用することによって、紙の資料を電子化すると同時に、スキャナの型番や、解像度、処理日時といった管理メタデータを自動的に生成・付与するとともに、書誌データベースから記述メタデータも付与することが可能になる。また、取り込んだ画像は、画像処理、OCR 処理を行うことでテキストの構造の分析が可能となり、構造メタデータの生成・付与も自動化される³¹。このような一連のプロセスを通じて、最終的に生成される電子化された情報にはメタデータが付与されており、アーカイブにおける管理も容易になる。

4.5.3. NEDLIB プロジェクトにおける納本システムのモデルについて

NEDLIB は、オランダ国立図書館を中心に 1998 年から 2000 年にかけて実施されたプロジェクトであり、電子出版物の納本システムのモデルとして DSEP (Deposit System for Electronic Publications) というシステムアーキテクチャーを作成した。DSEP の枠組みでは、メタデータの付与を自動化する、あるいは外部機関が保有する情報を活用することが検討されている。以下に、DSEP における業務フローを図示する。

³¹ METAE プロジェクトで利用されるデータは XML をベースとしており、メタデータについても DLF (Digital Library Federation) と米国議会図書館が中心となって策定を進めている XML スキーマを利用した METS (Metadata Encoding Transmission Standard) を活用している

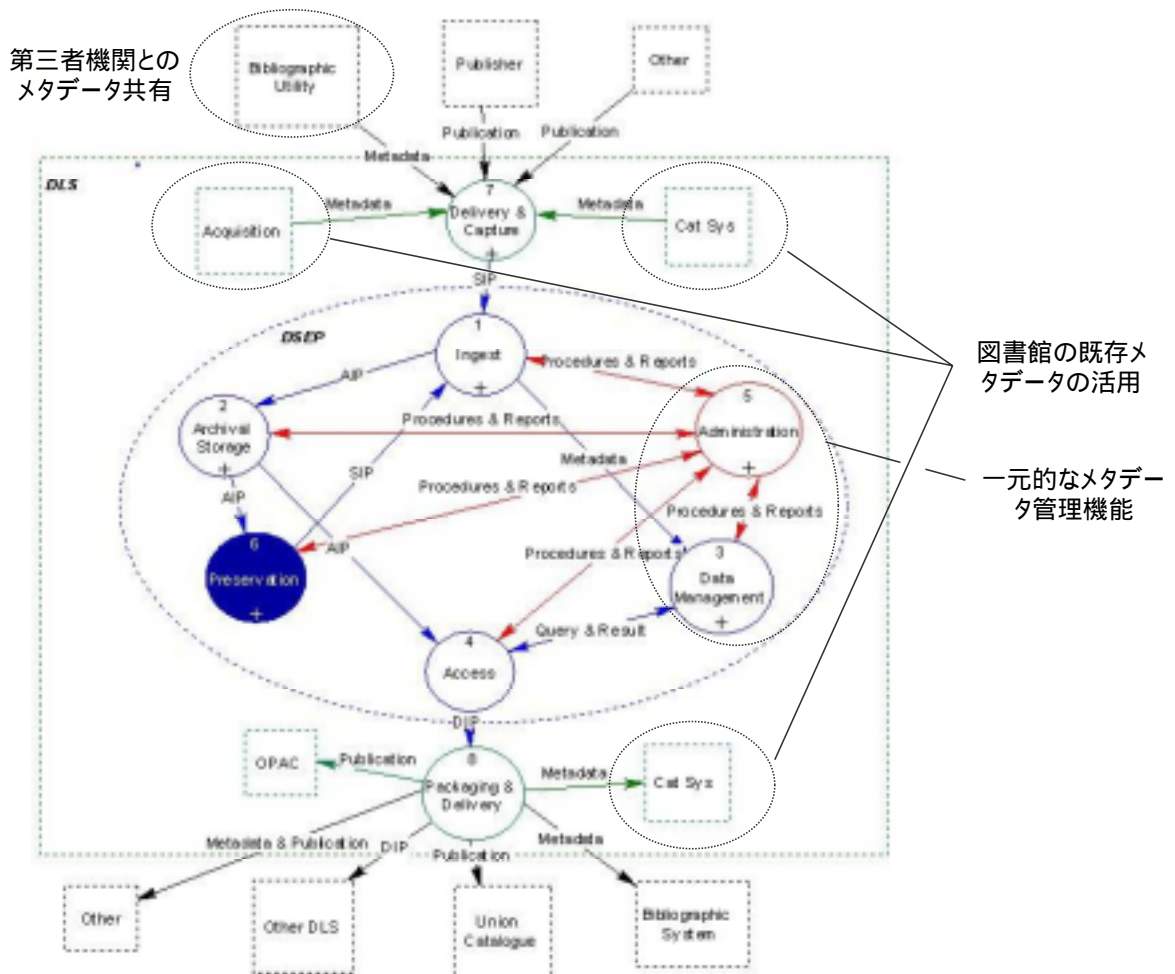


図 4-4. NEDLIB における DSEP システムの業務フローの概念図

DSEP においては、提出される出版物について、出版社自らが電子情報保存のための協働作業の一環としてタイトル情報、出版物内の情報のリンク構造、アクセスの制限、インストールに必要な技術的情報について、付与することを想定している。

DSEP に情報を受け入れる最初の段階、すなわち、上図における”Delivery & Capture”の段階では、提出された出版物に関するメタデータを付与する際に ISBN/ISSN 情報を所有する機関、あるいはそれ以外の書誌情報を提供する機関の保有するメタデータを共有する形で DSEP システムの業務フローが規定されている。

同時に、外部だけでなく、DSEP を導入する図書館自体が所有する全国書誌情報等についても、DSEP の業務フローに組み込み、保存のためのメタデータとして利用することを想定している。さらに、アーカイブの内部で情報を保存する際にも、メタデータの付与プロセスは極力自動化が図られている。例えば、保存活動の一環であるマイグレーションの履歴や、エミュレーションに必要な技術的な情報等は自動的な付与が行われる

よう考慮されている。

出版物の受け入れを行い、AIP (Archival Information Package) として、情報がアーカイブ内で保存されていく過程では、図における”Data Management”が一括してアーカイブ内のメタデータの管理を行うことになっている。この機能では、アーカイブの各機能、それらの業務プロセスから、各種のメタデータを収集、蓄積し管理が行われる。

DSEP では、電子情報の保存にかかるシステムとして、出版物の受け入れから、保存プロセスにいたるまで、トータルにメタデータの管理を効率化させるよう、システムアーキテクチャーの設計が行われている。

4.6. 電子情報保存に係るメタデータの考え方

電子情報保存に当たって、メタデータは非常に重要な役割を果たす。保存されている電子情報にメタデータが付与されていないければ、情報システムの中での位置を特定することすらできず、また情報が発見できたとしても、再生のための環境を記述した情報がなければ、適切な形で再現することができない。

電子情報の保存に当たってはメタデータが必要不可欠であるが、保存対象となる電子情報への付与に当たっては、コストがかかる。特に電子情報保存のためのメタデータは、技術的な内容が多く、それを理解できるスタッフの確保も必要である。

このような課題を解決するためには「メタデータの共有」、「付与の自動化」、「情報提供者との協働作業」が重要になる。また、いずれの場合でも、標準的なメタデータの存在が不可欠であり、その意味では本文で触れた OCLC/RLG の保存メタデータに関する取り組みは重要である。このメタデータは諸外国の有力な電子情報保存の事例研究をベースに、OAIS の概念を反映し、世界各国の主要な電子情報保存の実施主体からコンセンサスを得て策定がなされているからである。

日本国内における電子情報保存のためのメタデータを検討するに当たっても、このような先行事例を勘案しつつ、関係主体を巻き込んだ標準化の議論が必要である。その際には、どこまでの詳細なレベルで議論するかという点も大きな課題として挙げられるが、先に述べた OCLC/RLG の取り組みにも見られるように、特定の技術や情報種別によらない汎用的な性格を持つメタデータを策定する必要がある。

5.技術動向

本章では電子情報保存に係る技術動向として、エミュレーション、マイグレーション、XML を取り上げて、諸外国における先進的な電子情報保存プロジェクトにおけるそれぞれの技術の活用方法について述べる。

5.1.電子情報保存に係る技術の概要

ここでは、エミュレーション、マイグレーション、XML について、各々の特徴について整理する。

5.1.1.エミュレーションの概要

「エミュレーション」はソフトウェアを活用し、ハードウェアやソフトウェアの環境を模倣することで、旧来の技術環境を新たな技術環境で稼働させることを可能とする技術を指す。

この技術を活用することで、技術の陳腐化によって失われていく可能性のある電子情報を長期間にわたって再生することが可能となり、オリジナルの電子情報の機能とアクセスを保持していくための手段とすることができる。電子情報を再生するためのすべてのハードウェアとソフトウェアを動作可能な状態で保持しつづけていくことは、どのような機関であろうともコスト面、管理面のいずれかにおいて不可能だからである。

後述するマイグレーションと比較した場合のメリットは、オリジナルのデータを変換しなくても済む点である。技術環境をソフトウェアで模倣することで、オリジナルの電子情報の真正性と「見た目と感覚 ("Look and Feel")」を保持することが可能となる。

また、効率性の点でもメリットがある。適切なメタデータとソフトウェアが、共に保存されれば、メディアリフレッシュ以外の活動は特に必要とならないためである。また、一つのエミュレータで、同じ動作環境を必要とする複数の電子情報を再生することができるのも、効率的とされるゆえんである。

5.1.2.マイグレーションの概要

1996 年の CPA/RLG (「保存とアクセス委員会及びリサーチ・ライブラリー・グループ」) のレポートによると、マイグレーションとは、以下のように規定されている。

- ・ 「あるソフト/ハード環境からその他のソフト/ハード環境へ、あるいはコンピュータ・テクノロジーのある世代の技術環境から次の技術環境へ、デジタル素材を一時的に移行するためにデザインされ組織化された一連のタスク」

このような考え方は、概念としては幅広く受け入れられている。さらに、紙やマイクロフィルムへ移行する、代替物を作成する、ソフトウェアに依存しないフォーマットでデータを保存する等の場合も含めて、マイグレーションという場合もあるが必ずしも統

一的な見解があるわけではない³²。

後述する CAMiLEON プロジェクトでは、2001 年にマイグレーションに関するディスカッションペーパーが策定されており³³、ここではマイグレーションの概念を詳細化し、OAIS の規定を考慮したマイグレーションの 4 つのカテゴリーを挙げている。以下にそれを整理する。

- ・ リフレッシュメント
- ・ レプリケーション
- ・ リパッケージング
- ・ トランスフォーメーション

まず、リフレッシュメントとは、デジタル・オブジェクトのバイト・ストリームについて信頼できるコピーを維持することである。また、レプリケーションとリパッケージングは、OAIS アーカイブの中で取り扱いやすいパッケージにして維持することを指す。トランスフォーメーションは、デジタル・オブジェクトのバイト・ストリームを修正することを指している。また、さらに以下の二つの区分を設けてマイグレーションの概念を整理している（詳しくは第 3 章の表 3-4 を参照）

- ・ リバーシブル・マイグレーション（可逆性あり）
- ・ ノン・リバーシブル・マイグレーション（可逆性なし）

これは、マイグレーションを行った電子情報をもとに戻せるかどうかという視点からの分類である。

5.1.3.XML の概要

XML (eXtensible Markup Language) は、インターネット関連の技術標準である SGML³⁴ (情報の構造を記述する言語) と HTML (ウェブサイトの文書の表示方法を記述する言語) の双方の特性を受け継いで開発された言語で、インターネット技術の標準化団体である W3C³⁵ で仕様が検討されているオープンな標準技術である。

XML は、「タグ」を利用して情報を「構造化」して表現する。例えば、受注伝票であれば、以下のように示すことができる³⁶。

³²Margaret Hedstrom の分類による

³³Migration –a CAMiLEON discussion paper において、デジタルデータの長期保存に関する問題が取り上げられた

³⁴読みは「エス・ジー・エム・エル」。文書の論理構造、意味構造を記述する言語。文書のタイトル、段落、脚注等の構成要素をタグで表記することにより、文書の体裁を論理的な構造で記述することが可能となり、コンピュータの機種やソフトウェアに依存しないメリットがある

³⁵読みは「ダブリュー・スリー・シー」。WWW で利用される技術の標準化をすすめる団体。WWW 技術に関わりの深い企業、大学・研究所、個人等が集まって、1994 年 10 月に発足した。マサチューセッツ工科大学計算機科学研究所 (MIT/LCS)、日本の慶應義塾大学 SFC 研究所(Keio-SFC)等がホスト機関として W3C を共同運営している

³⁶実際のデータの記述に当たっては、さらに複雑な手続きが必要だが、説明の便宜上、単純化して例を作成している

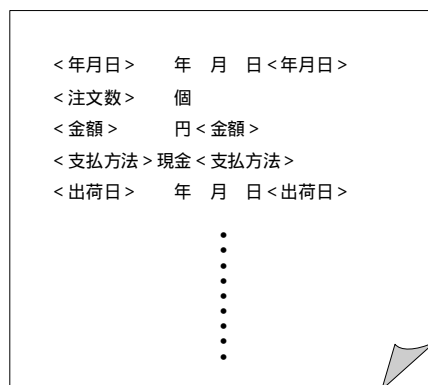


図 5-1. XML のインスタンス³⁷例 (イメージ図)

HTML も情報の表示に「タグ」を活用していたが、ユーザーが「タグ」を自由に定義することができなかった。一方、XML ではユーザーが「タグ」を自由に定義することができるため、「タグ」で囲まれる情報の「意味」をユーザーのニーズに応じて表現することが可能になる。

また、XML は、「タグ」の定義をまとめた DTD (Document Type Definition) と実際にタグを使って書かれたインスタンス文書で構成される。インスタンス文書の解析には、DTD を参照するため、保存内容の情報と、それを復元・解釈するためのコンテキスト情報を分離することが可能となる。一方で、XML はインスタンスと DTD のみでは画面に表示ができないデータである。そのため、表示には文書のレイアウトやフォントや色等を規定した XSL(eXtensible Style Language)が必要になる。DTD、XML インスタンス、XSL の関係を模式図として示すと以下の通りとなる。

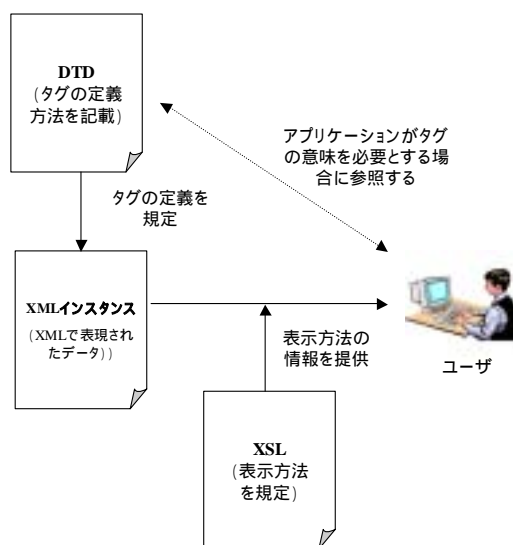


図 5-2. DTD、XML インスタンス、XSL の関係を表す模式図

³⁷ XML では、タグ付けされた実際のデータをインスタンスという

上図に示したように XML を用いることで、実際のデータとそれを解釈する、あるいは表示するためのコンテキスト情報を分離して管理することが可能となる。

XML は特定の技術プラットフォームに依存しない技術標準であるという性格から、電子情報の長期保存において様々な役割を果たしている。例えば、次のような応用の仕方である。

- ・ 保存のためのメタデータ
- ・ 保存対象そのものとしてのファイルフォーマット
- ・ エンキャプスレーション³⁸

XML は、インターネットの技術標準の策定を行う W3C において検討がなされている技術標準である。技術標準の策定に当たっては、異質な技術環境での相互運用性を保障するよう検討が行われている。また、W3C は情報技術の業界に大きな影響力を持つ組織であるため、主要なソフトウェアベンダーの大半がこの規格をサポートする方向にある。このような背景により、電子情報保存の文脈においても、現段階で異質な技術環境、あるいは将来的に出現が予想される技術環境において、XML がサポートされる可能性が高く、電子情報の長期保存に資する技術になると予想される。また、タグを付与することでデータ表現を行うため、0 と 1 の数値で表現されるバイナリコードよりも、視認性が高く、データの記述に当たっても理解がしやすいという利点がある。

さらに、XML 自体は、コンピュータの取り扱うデータとしては基本とも言えるテキストデータとして記述されるため、特定の技術プラットフォームに依存する割合が非常に低く、将来的にも永続的な活用が期待できる。一方で、新たな技術標準であるため、導入を行うユーザーにとっては、これまでのワークフローを変化させる必要があり、XML の導入に消極的になる場合があるとともに、XML 自体がテキストファイル形式であるため、ファイル容量が膨大になる場合があるといったデメリットも予想される。

³⁸ 保存しているデータの復元に利用するコンテキスト情報を XML で記述し、保存対象と対象となる電子情報を復元するためのコンテキストを与える情報をカプセル化して長期保存を行うことを指す

5.2. 調査対象プロジェクトについて

ここでは下表に掲載した 5 つの海外プロジェクトを取り上げ、先に説明したエミュレーション、マイグレーション、XML の活用について確認する。それぞれのプロジェクトが適用した技術のうち、今回の調査で取り上げるのは以下の通りである（点線の丸は各プロジェクトで検討や実験を行っているが、今回の調査では主要な調査対象としていないことを示す）。

	プロジェクト名				
	NEDLIB (1998.1-2000.12)	CAMiLEON (1999.10-2003.9)	Cedars (1998.4-2002.3)	DNEP (2000.9-2002.10)	Digital Preservation Testbed (2000.10-2003.10)
エミュレーション	39				○
マイグレーション	○				○
XML	-	-	○	○	

表 5-1. 調査対象プロジェクトと適用した技術

5.2.1. 調査対象プロジェクトの概要

今回調査対象とするプロジェクトは英国とオランダの事例を中心に 5 つを選出した。このうち、Cedars と CAMiLEON は姉妹プロジェクトであり、後者で検討されたエミュレーションの手法が前者に適用されている。また、NEDLIB については、オランダ国立図書館が中心となって実施したプロジェクトで、その結果の多くは、引き続いて実施された DNEP プロジェクトにも受け継がれている⁴⁰。

NEDLIB (Net worked European Deposit Library)

オランダ国立図書館を中心に、フランス、ノルウェー、フィンランド、ドイツ、ポルトガル、スイス、イタリアといったヨーロッパ地域の 8 つの国立図書館、3 つの主要な出版者、公文書館等が協力して、1998 年 1 月から 2000 年 12 月まで実施されたプロジェクトである。オンライン、オフライン（物理的媒体）、双方の電子出版物に対するアクセス保証を実現するための方法を調査し、ヨーロッパにおけるネットワーク化された納本図書館のインフラストラクチャを構築することを目的としていた。IT ベンダーがテスト・ベッドやデモンストレーションに協力し、納本システムに必要な機能要件、電子出版物の取り扱いプロセス、データモデリング、システムアーキテクチャー等を調査し

³⁹ Cedars のエミュレーションは、CAMiLEON において研究されたものを長期保存のための戦略として適用している。従って、Cedars のエミュレーションと CAMiLEON のエミュレーションは同等のものとする

⁴⁰ オランダ図書館は 1997 年より、IBM のデジタル図書館製品をベースに電子情報保存の取り組みを行っていた

た結果、各種のガイドライン、デモンストレーションシステム、小規模な納本システムに必要なソフトウェアを開発した。

CAMiLEON (Creative Archiving at Michigan and Leeds Emulating the Old on the New)

米国のミシガン大学と英国のリーズ大学が中心となって1999年10月より実施しているプロジェクト。英国側のプロジェクトは2002年12月に終了し、米国側のプロジェクトは2003年9月に終了する予定である。エミュレーションを活用することで、電子情報の見た目と感覚(“Look and Feel”)を保持することが可能か、また、エミュレーションが長期保存のための技術戦略として適切か否かを検討することを目的としている。これらの目的を検証するために、エミュレーションのためのツールを開発し、ユーザーによる評価、投資対効果分析、コレクションマネジメントのための課題等を検証した。

Cedars (CURL Exemplars in Digital Archives)

CURL (The Consortium of University Research Libraries) の全般的な指示のもとで eLib プログラムの一環として1998年4月から2002年3月まで実施されたプロジェクトである。プロジェクトの実施に当たっては、リーズ大学、ケンブリッジ大学、オックスフォード大学が協力した。本プロジェクトの目的は、電子情報保存の戦略、方法論、課題等を明らかにすることであった。この目的を視野にいれて、OAIS をベースに、メタデータ仕様を定めて分散型アーカイブのプロトタイプシステムを構築し、各種の仕様書を作成した。この他にも、知的財産権の問題、コレクション・マネジメント、技術戦略等に関する研究を実施した。

DNEP (Deposit for Dutch Electronic Publications)

2000年9月から2002年10月までオランダ国立図書館とIBM社が共同で実施した、電子情報の長期保存システムの実装と研究に関するプロジェクトである。2000年以前より、オランダ国立図書館とIBMは共同で、同図書館における電子情報保存システム (DIAS: Digital Information Archiving System) の研究を行ってきたが、その実装と平行して、真正性、保存のためのサブシステム、UVC (Universal Virtual Computer) のコンセプト検証、メディア・マイグレーション、Web出版物の6つの領域について研究を行った。オランダ国立図書館では、この研究結果を考慮してDIASの実装を2002年秋に完了し、現在はバージョンアップのための更なる研究を行っている。

Digital Preservation Test bed

オランダ内務省 (Ministry of Interior) と教育・文化・科学省、国立公文書館が中心となって、2000年10月から実施している政府関係情報の長期保存研究を目的としたプロジェクトである。2003年10月のプロジェクト完了を目指して現在も作業が進行中であ

る。このプロジェクトでは、エミュレーション、マイグレーション、XML の三つのアプローチから、異なるファイルフォーマット情報とアクセスの長期保存、システムの技術・システムの機能要件、保存にかかるコストモデル、デジタル・アーカイブ内の記録に関する規制や法制化の検討を行うことを目的としている。保存対象のデータとして、テキスト文書、e-mail、表計算ソフトのファイル、データベースを挙げており、必要に応じて、これらの保存対象にエミュレーション、マイグレーション、XML の技術を適用し、電子情報の長期保存に関する検討を行っている。現在、実験が完了段階に近い事例として、XML を活用した e-mail 保存のための実験を今回の調査対象としている。

5.3.各プロジェクトにおける技術の適用状況

ここでは、エミュレーション、マイグレーション、XML のそれぞれについて、先に説明した諸外国における電子情報保存プロジェクトでの適用状況について述べることにする。

5.3.1. 各プロジェクトにおけるエミュレーションの活用状況

ここでは、NEDLIB における実験を踏まえ、CAMiLEON と Cedars におけるエミュレーションを取り上げることとする。また、本節の最後には、DNEP で扱われたエミュレーションから派生する概念である Universal Virtual Computer (UVC) について述べる。

NEDLIB におけるエミュレーションの概要

NEDLIB におけるエミュレーションに関する実験は、本技術の有力な研究者のひとりである Jeff Rothenberg が中心となって、1999 年に行われた。

ここで取り上げられたエミュレーションの技術は、NEDLIB において検討されている OASIS に準拠した DSEP (Deposit System for Electronic Publications) という電子出版物の納本システムのアーキテクチャーに組み込む際の、実効性を確認する目的で検討されていた。しかし、本技術に関する実験は 1999 年以降、2000 ~ 2001 年にかけて行われる予定であったが、プロジェクト自体は 2000 年で終了している。

NEDLIB におけるエミュレーションの理論的背景

エミュレーションを行う対象には「ハードウェア」、「プログラム」、「オペレーティングシステム」といった三つのレベルが存在する。NEDLIB では、このうちハードウェアのエミュレーションに焦点を当てている。

ハードウェアをエミュレートすることにより、実物のハードウェアの保存を不要にするとともに、保存しておいたオリジナルのソフトウェアをハードウェア・エミュレータ上で再生するのである。

本プロジェクトにおけるエミュレーションの概念を以下に図示する。

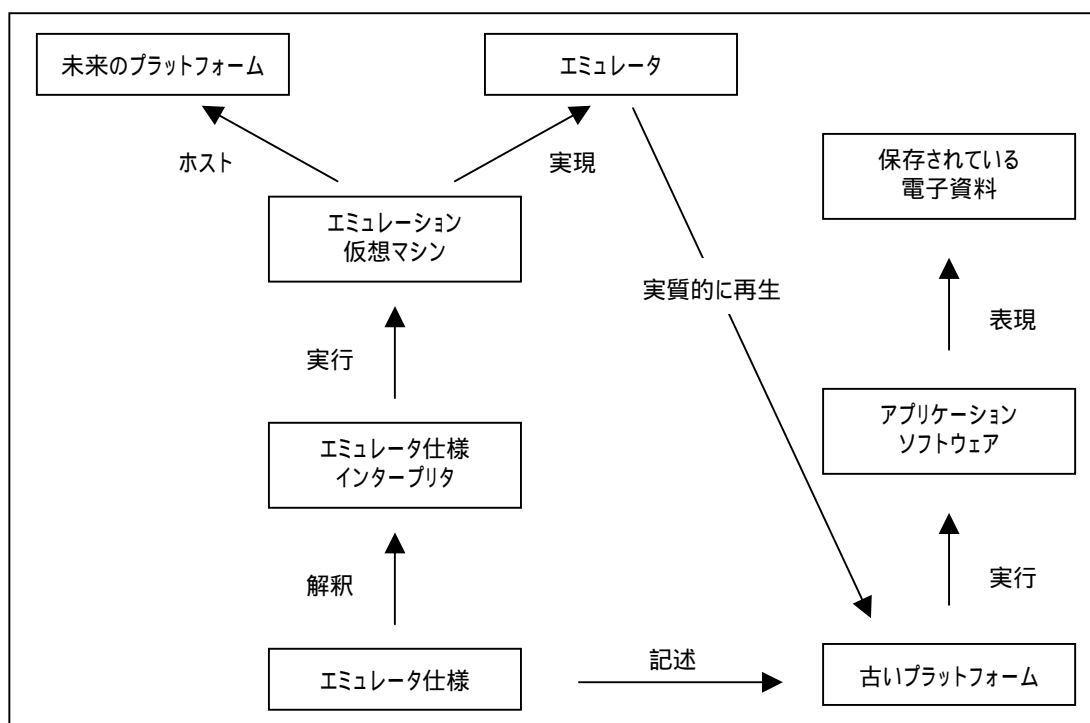


図 5-3. NEDLIB におけるエミュレーションの概念

上図で示した通り、本プロジェクトのエミュレーションの概念では、「エミュレータ仕様」、「エミュレータ仕様インタプリタ」、「エミュレーション仮想マシン」により、エミュレーションを行うこととしている。

この概念では、「エミュレータ仕様」に旧式の保存すべきプラットフォームの仕様を記述する。これを将来の技術環境上で構築される「エミュレータ仕様インタプリタ」で読み出すことにより、旧式のハードウェアの技術環境を読み解く。そして、読み解いた情報をベースに「エミュレーション仮想マシン」がその仕様を実現してエミュレータとなるのである。

ここで再現されるエミュレータ上で、旧来から保存していたアプリケーションと、保存している電子情報を再生することで、将来の環境において、旧環境を表現することが可能になると考えている。

NEDLIB におけるエミュレーションの実験

1999 年の実験においては、第一段階として、すべてのプロセスを単純化し、資金面及び時間的制約を考慮した上で、非常に限定された環境で実験が行われた。以下にエミュレーションの概念を規定した実験において、技術以外に変数となる条件を整理する。

技術以外の変数となりうる条件	実験での設定内容
保存する電子出版物の種類	オランダ国立図書館が独自に選択したサンプル資料
電子出版物の種類に応じた保存基準	-
保存基準の妥当性を評価するテスト	-
電子出版物の閲覧等を行うために必要なソフトウェアの種類	オランダ国立図書館が選択したサンプル資料に対応するソフトウェア
ソフトウェアに応じたシステム要件等	オランダ国立図書館及びR A N D社で対応可能な範囲
エミュレーションを実現するまでのスケジュール	短期：現行システムでの実現可能性を探索する
エミュレーションツール、エミュレータ作成ツール等	フリーウェア又は市販品
エミュレーションを実行/表示できるハードウェア	オランダ国立図書館及びR A N D社で対応可能な範囲
技術以外の要素（例：ソフトウェアの知的所有権等）	影響は最小限とみなす

表 5-2. NEDLIB のエミュレーション実験における条件変数

上記の実験変数を踏まえ、複数の CD-ROM 出版物とオンラインジャーナルを対象に市販のエミュレータを活用し、Macintosh 上で Windows 環境を再現する実験が実施された。実験では、Macintosh 上に Windows 環境を再現する市販エミュレータをインストールし、これを実際の Windows マシンと併置することで、両方のハードウェア上で、再生の際に Windows 環境が必要等キュメントが正しく表示されるかを検証したのみにとどまった。結果的には、ほとんど問題がなく表示がなされた。

NEDLIB におけるエミュレーション実験から得られる示唆

本プロジェクトにおける実験は、非常に限定された環境において実施されたものであるため、特に目立った示唆はない。しかし、Rothenberg が提示したエミュレーションの概念自体は、非常に示唆に富むものであり、他の電子情報保存プロジェクトに多大な影響を与えた。

CAMiLEON と Cedars におけるエミュレーション

CAMiLEON と Cedars は姉妹プロジェクトとして運営されており、CAMiLEON で研究された成果が、Cedars で検討される分散型アーカイブのプロトタイプに反映されている。ここでは、エミュレーションについて中心的に研究を行った CAMiLEON の実験を取り上げる。

CAMiLEON におけるエミュレーションの概要

CAMiLEON におけるエミュレーションは、古い技術環境を新しい技術環境で再生することで、電子情報の「見た目と感覚 (“Look and Feel”）」を長期間にわたって保持することを可能にする技術と位置付けられる。基本的には、先に触れた Rothenberg の理論と同様で、旧来の技術環境を、将来の技術環境で再現するためのエミュレータを利用する形をとる。

この概念に加えて、CAMiLEON では、エミュレーションを電子情報保存のための単独の戦略として扱うのではなく、マイグレーションと組み合わせて利用することが、コスト効率の高い保存戦略を打ち立てることになるとしている。CAMiLEON において検討されたエミュレーションとマイグレーションを組み合わせた “Migration on Request” という概念については、マイグレーションの項で後述する。まずは、CAMiLEON におけるエミュレーションの概念について整理する。

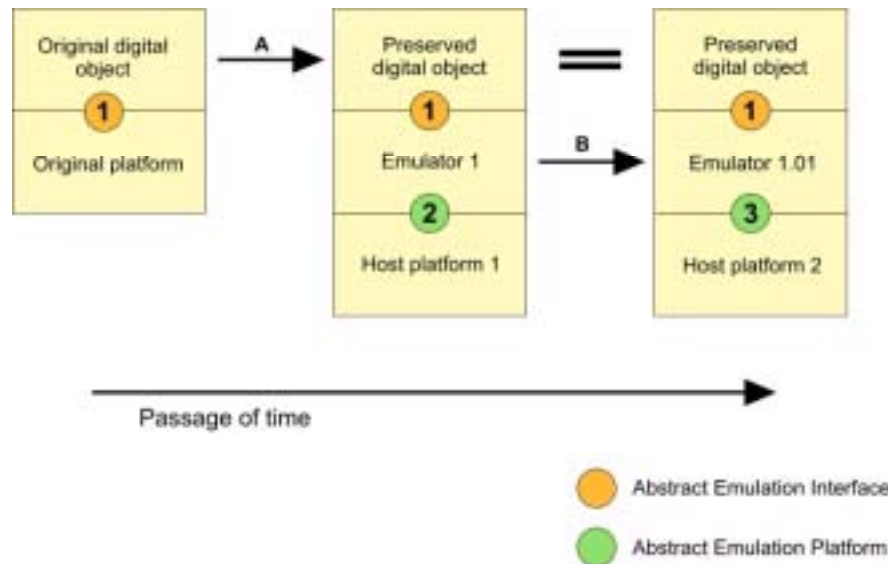


図 5-4. CAMiLEON におけるエミュレーションの概念

上図の説明を以下に記す。

- ・ 矢印 A は、オリジナルの電子オブジェクトを、保存のための電子オブジェクトに変換する⁴¹ことを示す。保存された電子オブジェクトはエミュレータ 1 によって実行される。
- ・ 等号は、保存されている電子オブジェクトが同一のものであることを示す。
- ・ 矢印 B は、プラットフォーム 1 が旧式化し、違うプラットフォーム上で更新されたエミュレータによって、保存された電子オブジェクトが実行されることを示す。

⁴¹ Cedars の Digital Preservation Strategy においては、保存対象とすべき電子情報を媒体から引き離し、バイトストリーム (バイト単位の情報) と、そのバイトストリームを読み出した時に解釈するための情報に分けて保存することを推奨している

- ・ インターフェイス 1 (上図の) は、オリジナルの電子オブジェクトの操作に必要な API (Application Program Interface) を指す。エミュレータ 1 は、このインターフェイスをエミュレーションによって仮想的に再現するため、将来的にも同一のものが使われる。
- ・ インターフェイス 2 (上図) は、オリジナルの電子オブジェクトと再生のためのプラットフォームをつなぐ役割を担う。
- ・ インターフェイス 3 (上図) は、技術環境の変化にともない、新たな環境で保存されている電子情報を再生するために、 と同じ役割を担う。

から に変化する際の改造が少ないか、あるいは が永続的に利用できるものであれば、エミュレータへの改造が非常に少なくすむため (これを想定して、図では Emulator のバージョンが 1.01 となっている) 非常にコスト効率の高い電子情報の保存が可能になる。従って、 の選択が非常に重要になる。

エミュレーションレベルの選択

先に述べたとおり、エミュレーションを行う際にはインターフェイス(図 2 の と) の選択が極めて重要となる。 を選ぶ際の基準は以下の通りに定められている。

- ・ 複雑すぎないこと
- ・ 選択したインターフェイスのドキュメントが確保できること
- ・ 周辺機器の仮想的なマッピングが容易であること
- ・ 保存している情報の特性の重要な部分を保持できること

例えば PC 上で動作する Windows アプリケーションであれば、ハードウェア、PC BIOS⁴²、Windows API⁴³が選択肢となるが、上記の基準に照らして、Windows API を選択することを推奨している。次に を選ぶ際の選択基準として以下の事項が定められている。

- ・ 標準的なソフトウェア開発の方法論に基づいてエミュレータが作成されること (十分なドキュメンテーション、有益なコメント、適切なコード構造の使用が含まれる)
- ・ 使用するプログラム言語が C 言語のサブセットで書かれていること⁴⁴

⁴² 読みは「バイオス」。Basic Input / Output System の略で、コンピュータに接続されたディスクドライブ、キーボード、ビデオカード等の周辺機器を制御するプログラム群。これらの機器に対する基本的な入出力手段を OS やアプリケーションソフトに対して提供することからこう呼ばれる

⁴³ あるプラットフォーム(OS やミドルウェア)向けのソフトウェアを開発する際に使用できる命令や関数の集合のこと。また、それらを利用するためのプログラム上の手続きを定めた規約の集合。個々のソフトウェアの開発者がソフトウェアの持つすべての機能をプログラミングするのは困難で無駄が多いため、多くのソフトウェアが共通して利用する機能は、OS やミドルウェア等の形でまとめて提供されている

⁴⁴ 1972年に米国AT&T社のベル研究所でD. M. Ritchie氏とB. W. Kernighan氏によって開発されたプログラミング言語。1986年に米国規格協会(ANSI)によって標準化され、国際標準化機構(ISO)や日本工業規格(JIS)にも標準として採用されている。Cの拡張版であるC++言語(シープラス・プラス言語)とともに、現在もっとも普及しているプログラミング言語である。特定のプラットフォームに依存した部分を言語から切り離しているため、移植性の高いプログラムを記述することができる。もともとシステムの記述用に開発されたので、ハードウェアよりの低水準な処理を記述することもできる。実際、UNIXはC言語で記述されている。未来のプログラミング言語にもC言語のセマンティクスが引き継がれることを想定して、ここではC言語が選択されている

- ・ 非標準コードに対応ができること(非標準コードに対応する機構はモジュール化して独立性を高めるとともに、仕様について十分にドキュメンテーションされる必要がある)

CAMiLEON におけるエミュレーションの実験

CAMiLEON プロジェクトでは、エミュレーションを用いた実験として、1986年に英国のBBCが開発したDomesday⁴⁵プロジェクトという電子資料の「救出」作業を実施した。BBC Domesdayは、2枚組みの両面レーザーディスクで、1086年当時の英国の生活を記録したデジタル・アーカイブである。当時の最先端のハードウェア、ソフトウェアを用いた画期的なマルチメディアシステムで、再生にはオリジナルのハードウェアとソフトウェアが必要とされていた。



図 5-5. BBC Domesday システム

各種機器	概要
メディア	両面レーザービジョンディスク
ハードウェア	BBC マスタ マイクロコンピュータ、又は Research Machines Nimbus コンピュータ
周辺機器	キーボード、トラックボール、LV ディスク ROM、SCSI カード
ソフトウェア	ソフトウェア言語 BCPL によるプログラム
データ	画像データはアナログ・フォーマットによって1画像1トラックで格納 動画ビデオも同様 テキストや統計、マッピング・データ等はデジタル・フォーマットで格納

表 5-3. BBC Domesday の各種機器の概要

⁴⁵ 1086年イングランド王ウィリアム1世が徴税の必要上作成させた土地台帳。ほぼ全国土の都市・荘園ごとに、面積、市民・農民数から犁の数に至るまでくわしく記録されている。BBCのDomesdayプロジェクトは、ウィリアム1世の土地台帳の完成から900周年を記念して行われた

BBC Domesday システムは、メディア、再生装置、ソフトウェアに問題を抱えていた。メディア自体は、物理的には数百年の寿命を持っているが、これまでの使用によって、今後、読み取りエラーが発生することが予想されていた。再生装置についても、独自仕様のものであるため、ハードウェアの修理・交換が非常に困難になっていた。さらに、ソフトウェアは、BBC Domesday に固有のソフトウェアで、内容のナビゲーション、閲覧、クロスリファレンス等を提供しており、テキスト、イラスト、写真の表示機能、歩行シミュレーションや各種動画の表示等の機能を提供していた。

CAMiLEON プロジェクトでは、BBC Domesday の再現を取り上げた理由として以下の点を挙げている。

- ・ 複雑でインタラクティブなオブジェクトであるため、既存の保存戦略（マイグレーション）を用いて全体をそのまま保存することは困難である
- ・ 複数の周辺機器に依存しているため、周辺機器のエミュレーションに関連する探求が必要とされる
- ・ BBC Domesday と同じ技術を用いた資料が他にも制作されており、それらへの適用によって開発する技術の費用対効果が高まる
- ・ BBC Domesday は、その規模、技術的複雑性、知的所有権の処理において資料保存の難関である

実験の結果

2002 年 12 月 2 日、CAMiLEON プロジェクトの実施主体である英国のリーズ大学において、BBC Domesday の再現デモンストレーションが行われたが、その内容はほとんど問題なく再現できたようである。古いディスクから再現された内容は、CAMiLEON の終了後に、エミュレータとともに Public Record Office に納入される予定である⁴⁶。

DNEP におけるエミュレーション

DNEP では、IBM 研究所の Raymond Lorie が考案した Universal Virtual Computer（以下、「UVC」と略記する）のコンセプトの有効性を検証している。UVC はエミュレーションから派生する概念であり、旧来の技術環境を新しい環境で再生するために、UVC を活用するというものである。DNEP では、PDF ファイルの再生を用いた実験を行い、UVC のコンセプトを検証して、DIAS への将来的な実装の検討を行っている。

⁴⁶ CAMiLEON のウェブサイト (<http://www.si.umich.edu/CAMiLEON/>) によると、2002 年 12 月に最終レポートが提出される予定である旨の告知があるが、2003 年 1 月 12 日現在、公開はされていない

UVC の理論的背景

UVC は、オリジナルのデジタルデータの他に、そのデータを再現するためのプログラムである“プログラム P”を保存することを推奨する。プログラム P の機能によって、ビットの集合からデータを読み出し、情報を理解することができるのである。また、プログラム P は、特定の技術とは独立した形で存在するため、新たな技術環境への移行もスムーズに行うことができる。

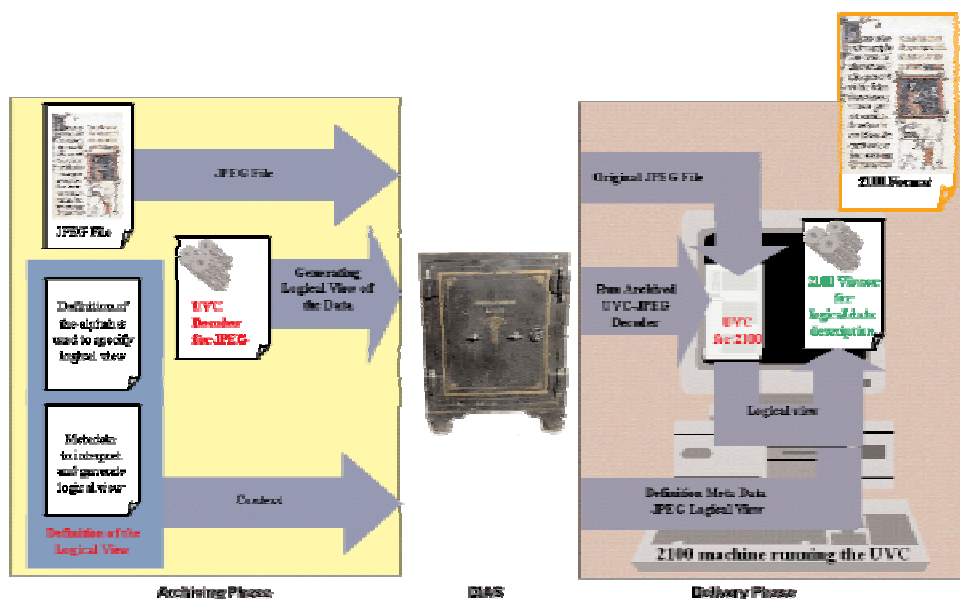


図 5-6. UVC の概念 (JPEG を例にしたイメージ図)

上の図でプログラム P に当たるのは、JPEG 用の UVC デコーダ (UVC Decoder For JPEG) である。概念的には、UVC デコーダは非常に簡略化された命令セットで設計されているため、比較的容易に将来環境で、UVC エミュレータを作成することができるとされている。情報の保存に際しては、現行の環境下で、UVC デコーダ以外に、保存対象であるオリジナルの JPEG ファイル、そして UVC デコーダに保存対象を解釈するための手がかりを与えるコンテキスト情報 (上図の Definition of the Logical View) を保存しておくことが必要となる。

将来の技術環境で構築される UVC エミュレータを用いることで、UVC デコーダを再生し、保存対象のコンテキスト情報を解釈し、オリジナルの JPEG ファイルを再現することが可能になる。こうすることで、例えば、上の図で示されたような 2100 年の技術環境においても、古い技術環境の情報を再生できるとされている。

また、JPEG のみならず、GIF や TIFF 等といったファイルを再生するためのデコーダと情報の解釈に必要なコンテキストを保存すれば、様々なデータタイプでも、将来環境で構築される UVC エミュレータで再生することができるのである。

保存対象となるオリジナルのデジタル情報に特定のアプリケーションが必要な場合も、UVC のコンセプトが適用可能であると考えられている。この場合は、UVC をハードウェアエミュレータとして機能させることとなる。この概念について以下に図示する。

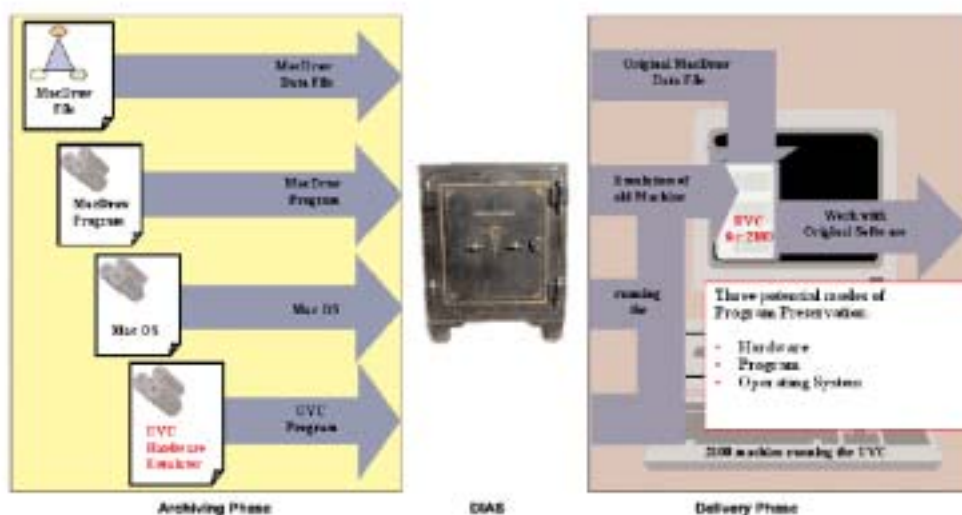


図 5-7. UVC を活用したハードウェアエミュレーションのイメージ

上の図で示された通り、保存の際に、オリジナルのファイル（例：MacDraw⁴⁷ファイル）、特定のアプリケーション（例：MacDraw プログラム）、OS（例：MacOS）を保存し、これらを動作させるための UVC ハードウェアエミュレータを作成しておく。これを将来環境で構築される UVC エミュレータで再現することで、旧来の技術環境の再生が可能になるというものである⁴⁸。

UVC コンセプトの検証実験

DNEP プロジェクトでは、PDF ファイルを用いて UVC のコンセプトを検証する実験が行われた。具体的には、PDF ファイルを HTML 形式に変換し、テキスト、画像 (JPEG)、ブックマーク構造を抽出し、オリジナルの情報の再生に必要なコンテキスト情報を定義するとともに、UVC デコーダを作成し、オリジナルの情報を再生可能かどうか検証するというものであった。実際に、UVC エミュレータを作成し、保存対象としている UVC デコーダ等を介して情報の再生が可能であったが、JPEG 画像については、画像圧縮技術の関係で正しく再現ができなかったため、画像変換を行うことで対処した。この実験の結果から、UVC のコンセプト自体は、ある程度の有効性が検証されたといわれている。ただし、JPEG 画像の再現の問題や、さらに UVC のコンセプトを高度化するために、保存データの圧縮技術等に関する検討を行う予定であり、更なる技術検証が行われる見込みである。

⁴⁷ マッキントッシュ用のグラフィックアプリケーション及びそのアプリケーションで利用するファイル。最古参のドロー系アプリケーションで、グラフィックを数値として扱い、ドットとして扱うものよりも滑らかな描画が可能であるという特徴をもつ

⁴⁸ UVC を活用したハードウェアエミュレーションと Rothenberg が提案したエミュレーションの違いは、将来環境においてハードウェアエミュレータを改めて構築する必要がない点である

5.3.2.各プロジェクトにおけるマイグレーションの活用状況

マイグレーションについては、CAMiLEON、Cedars、DNEP プロジェクトで検討されたものについて取り上げる。

CEDERS、CAMiLEON におけるマイグレーションの概念

デジタル情報の保存後に、再度そのデジタル情報を利用する方法に関しては、デジタル保存の大きな課題である。これは従来から「マイグレーション」と呼ばれる概念が論点となる。「マイグレーション」自体は幅広い概念であり、解釈にばらつきがあったが、これを分類したものが OAIIS によるマイグレーションの4つのカテゴリーである。

- ・ Refreshment (リフレッシュメント)
 - 内容情報のビットシーケンスの変更を伴わずにメディアの交換を行うこと
- ・ Replication (複製)
 - 内容情報のビットシーケンスやパッケージ情報 (Packaging Information) の変更を伴わずに複製を作成すること
- ・ Repackaging (リパッケージング)
 - パッケージ情報のビットシーケンスに多少なりとも変更を伴うマイグレーションの種別のこと
- ・ Transformation (変換)
 - 内容情報の変更を伴うマイグレーションであるが、特定のアルゴリズムで変更復元可能な場合と不可能な場合の二種類がある

伝統的に「マイグレーション」といった場合、オリジナル・オブジェクトにおけるデータフォーマットが陳腐化した場合、マイグレーションツールを利用して、データ変換を行い、現在のプラットフォームにあわせたデータフォーマットに合わせていくという概念であった。またこの場合、技術進歩に伴って、普及しているプラットフォームが変化した場合、その都度マイグレーションをかける必要がある。

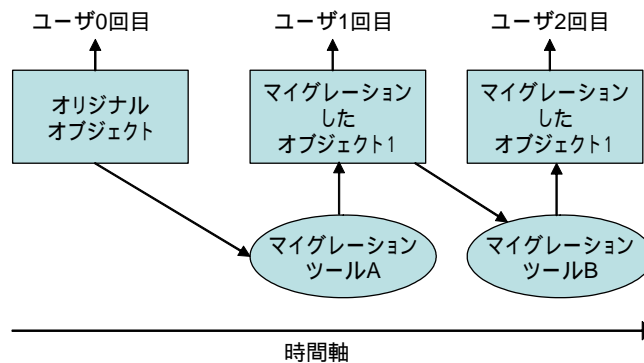


図 5-8. 伝統的なマイグレーションの概念

この場合、上述のリフレッシュメント及び複製においては、オリジナル・オブジェクトの内容情報にそれほど影響が出ないが、リパッケージング及び変換においては、技術的な移行という点に重点を置くと、内容情報の変質が問題となる。

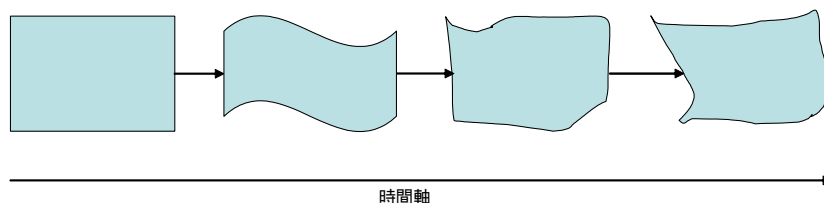


図 5-9. 伝統的なマイグレーションによるデータの変形

伝統的なマイグレーションは、短期的な保存に適しているかもしれないが、長期保存という側面からみた場合、データ変形が想定されるような戦略は好ましくないと考えられる。

ここで、新たな概念として CAMiLEON プロジェクト及び Cedars で取り上げられたのが、マイグレーション・オン・リクエスト (Migration on Request) である。これは、保存対象であるオリジナル・オブジェクトの内容情報の変質を極力減少させるという考え方である。

まず、オリジナル・オブジェクトの保存という側面では、Cedars における概念を採用している。Cedars では、保存対象の電子情報を、コンピュータで扱う情報の基本単位であるバイト・ストリームに変換する方法をとる⁴⁹。例えば、CD-ROM に格納された情報は、その内容を CD-ROM から引き出して情報システムに保存する。引き出した情報の内容は、非常に汎用性が高いと考えられるバイト情報に変換される⁵⁰。

変換されたオリジナル・オブジェクトは、バイトのみの情報であるため、情報復元のためのメタデータを付加しなければならない。例えば、保存対象となる情報が HTML であ

⁴⁹ 実際には、オリジナル・オブジェクトを保存のための保存デジタルオブジェクト (Preserved digital object) に変換するプロセスを実施する。このプロセスを経るため、厳密には 100% オリジナルの情報が保存されるわけではないといえる。このプロセスでは保存対象となるオリジナル・オブジェクトのどの部分を優先的に保存すべきかという "Significant Properties" という概念が適用される。これは、保存対象の重要な特質を保存することを優先し、あまり重要でない部分については保存の優先度を落とすというものである。例えば、先の CD-ROM の例では、CD-ROM という媒体は Significant Properties とは見なされないため、保存対象とされず、情報の知的内容のみが情報システムに移行されるということである。

⁵⁰ Cedars の電子情報保存プロトタイプシステムでは、バイトストリームを保管するためのファイルフォーマットとして UNIX で使用される ".tar" ファイルを使用する。".tar" ファイルはフォーマット情報が十分に文書化されており、読み書きできるソフトウェアもフリーウェアとして広く普及している。このような標準的なフォーマットを利用することで、情報を読み出せなくなるリスクを低減している。

れば、保存対象の知的内容そのものはバイト・ストリームに変換し、これを解釈するために HTML の読み解き方の情報を付加し保存するのである。再生の際には、バイト・ストリーム（OAIS でいう”Content Information”）を HTML の読み解き方の情報（OAIS でいう”Representation Information”）を使って解釈することで再現が可能となる。

「保存対象をバイト・ストリームに変換する」という方法の優位性は、バイトがコンピュータで扱う情報の基本単位であるという部分にある⁵¹。バイトという考え方はコンピュータの黎明期から使われてきた情報の基本単位であり、コンピュータの誕生から 50 年以上経つ現在まで引き継がれているものである。

今後、万が一にも現在のコンピュータの根底を覆す大きな技術パラダイムの転換があったとしても、その転換は徐々に進み、その間に情報損失の少ない移行手段が考えられる可能性が高い。あるいは、パラダイムの転換が急激に起こったとしても、既にある大量のバイト資産を捨てること考えにくく、何らかの移行手段が提示される可能性が高く、その時点で次世代の技術環境に適合が可能になる。

この保存概念をもとに、CAMILEON ではマイグレーション・オン・リクエストを技術面からの補足を行うための検討を行っている。

マイグレーション・オン・リクエストの大きな特徴は、オリジナルのバイト・ストリームを維持しつつ、マイグレーションを行っていく点にある。ここでは、伝統的なマイグレーションのように、大量のデジタル・オブジェクトを一括移行するのではなく、利用の都度、データ変換ツールを利用して再現するという考え方を採用している。これにより、データ変換の際に小さなエラーはあるものの、従来のようなマイグレーションの繰り返しによる変形の影響は少ない。

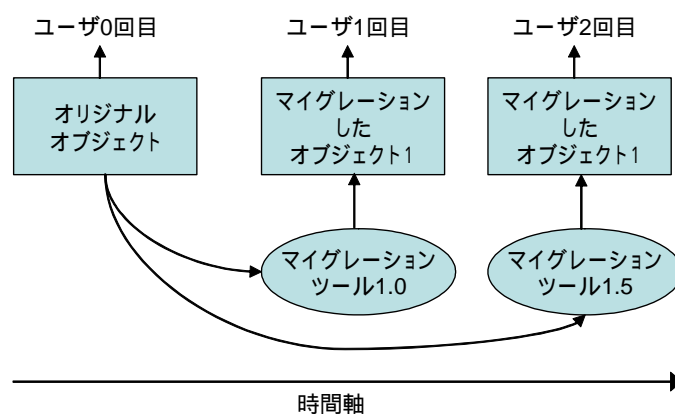


図 5-10. マイグレーション・オン・リクエストの概念

⁵¹ さらに下位の情報の単位としてビットがあるが、コンピュータのほとんどは 8 ビットを 1 バイトとして扱っており、ストレージに保存する際にもバイトの単位で保存される

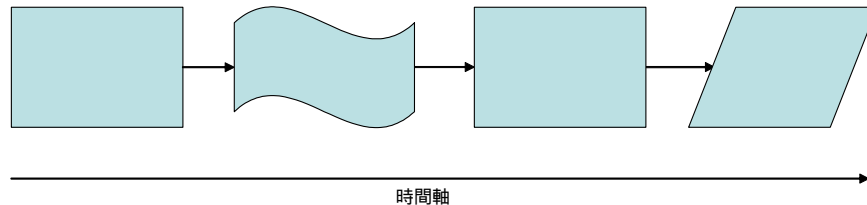


図 5-11. マイグレーション・オン・リクエストによるデータの変形

マイグレーション・オン・リクエストでは従来のマイグレーションにはない五つのメリットがある。

- ・ 従来のマイグレーションのようにデータ変形/変質を伴わないため、正確なデータの再現が可能である：通常のマイグレーションはデータ変換を繰り返す
- ・ オリジナルデータを残したままデータ変換を行うため、マイグレーション実施の際に、情報損失の程度をオリジナルデータと比較・検証可能である：通常のマイグレーションでは、データ変換を繰り返した後にオリジナルデータは残さない場合が多い
- ・ マイグレーションツールは「呼び出しに応じて」配備すればよいので、大量データを保存する場合には、大規模なコスト削減が可能である：通常のマイグレーションは保存データを一括して変換することが多い
- ・ オリジナルの情報を保持しておくため「原本性」の考え方が単純になる：通常のマイグレーションでは変換が繰り返されるため、変換後の複製と原本が異なる場合が多く原本性等の問題が複雑になる
- ・ 将来的に登場する技術を活用することで、オリジナルの情報を高精度に再現できる可能性が高まる：通常のマイグレーションはデータ変形/変質を伴い、データの欠損が起こる確率が高く、オリジナル情報を再現できない可能性が高い

マイグレーションの適応に向けたテスト

実際の適応に向けて、3つのフォーマットを用いたテストを行った。利用したフォーマットは、ベクター図形に対応する WMF⁵²、Draw⁵³、SVG⁵⁴であった。また、マイグレーションツールは、モジュール方式で構築した。

まず、異なる機能がそれぞれのフォーマットにインプットされ、中間層に線、楕円、

⁵²Windows Meta File の略。マイクロソフト社が開発した Windows 標準のベクトル形式グラフィックフォーマット

⁵³ マイクロプロセッサの設計様式の一つである RISC 方式(個々の命令を簡略化し並行して複数の命令を処理する方式、ワークステーション用 CPU にはこの型のプロセッサが多い)の OS で利用するアプリケーションのファイル

⁵⁴XML ベースの 2D ベクター画像記述言語。2001 年 9 月に W3C 勧告として公開。画像を点の集合体ではなく線や面等の図形の集合体として扱うため見る環境に応じた最適な表示が可能。アニメーション機能などもサポートしている

ポリゴンなどの要素として戻ってくる。この構造は、アウトプットモジュールに引き継がれるが、アウトプットフォーマットによっては、サポートされない要素を含みやすい。従って、サポートされていない要素を変換して、フォーマットの操作が可能な機能構造を引き継ぐようにする。

このテストの結果、いずれのフォーマットにおいてもオリジナルの図形を示すための方法が異なっていた。楕円形を例にとると、WMF の場合は、楕円形と長方形、Draw は線とカーブ、SVG は丸い辺として認識されていた。

ここでは WMF は、余分な情報がアウトプットされやすく適当とはいえなかった。現時点では、ベクター図形のマイグレーションツールにおいて、中間フォーマットとして活用するには SVG が基本になると考えられる。

しかし、現時点ではファイルフォーマットの進化を踏まえるなどさらなる検討が必要であり、現在でも CAMiLEON プロジェクトにおいては、マイグレーションに関するコスト及び技術の可能性が議論されている。

DNEP におけるメディア・マイグレーション

DNEP プロジェクトでは、たいていの納本システムが想定するテラバイト級のデータをマイグレーションする際の手順や、DIAS⁵⁵の実装にかかる課題等について検討を行っている。

このような検討を行った理由としては、膨大な電子情報のマイグレーションには、膨大な時間を要する点が挙げられる。例えば、磁気テープに 100 テラバイトのデジタルデータが格納されているとして、メディアの陳腐化により、これを光媒体のストレージに移行する場合を考えると、最近の光媒体のストレージは、1 秒間に 4 メガバイト程度の書き込みスピードがあると仮定して、マイグレーションの完了に 290 日⁵⁶が必要となるのである。このような膨大な時間を要するマイグレーションの処理を如何に効率的に行うかといった課題について DNEP プロジェクトで検討がされている。

ストレージ・パフォーマンス・モデル

膨大な量のデジタルデータを移行する場合、この活動をサポートするプロセスが必要であり、ストレージ・デバイス等の個々のライフサイクルを視野に入れなければならない。そのためには、ストレージ・パフォーマンス・モデルを構築する必要があり、これを確立することで以下の事柄が明確になる。

⁵⁵ Digital Information Archiving System の略。オランダ国立図書館のデジタルライブラリーシステムであり、2002 年秋に第一バージョンがサービス開始となった。DIAS では、磁気ディスク、磁気テープ、光ディスクライブラリを使用している

⁵⁶ 100 テラバイトは、100,000,000 メガバイトであり、これを 1 秒間に 4 メガバイトの割合で移行するとして、2,500 万秒かかる。これを時間に直すと 6,944 時間であり、1 日は 24 時間であるから、約 289.35 日かかることになる

- ・ストレージメディアの特徴
- ・それぞれのストレージのライフレンジ
- ・メディア・マイグレーションやリフレッシュメントを行うことが可能な時間

これらを明らかにするためのパフォーマンスモデルは以下の数式で示される。

$$\text{MEL} > \frac{\left[\frac{(\# \text{objects} \times \text{average}(\text{object size}))}{\text{MRR}} \right] + \left[\frac{(\# \text{objects} \times \text{average}(\text{object size}))}{\text{MWR}} \right] + (\# \text{objects} \times \text{DOP})}{(\text{MMW} \times \text{OTI})}$$

図 5-12. メディア・マイグレーションインディケータ

各項目の名称	各項目の概要
MEL (Medium Expected Lifetime : メディアの予想される寿命)	メディアが電子保存システムの中で機能し続けると予想される時間
MWR (Medium Write Rate : メディアへの書き込みレート)	特定の時間内にメディアに書き込みができる最大バイト量
MRR (Medium Read Rate : メディアからの読み出しレート)	特定の時間内にメディアから読み出しができる最大バイト量
DOP (Digital Object Processing : デジタル・オブジェクトの処理時間)	デジタル・オブジェクトを処理するための時間
OTI (Operational Time Interval : 一回の運用サイクルの時間)	一回の運用サイクルの時間 (例 : 日、週、月、年等)
MMW (Medium Migration Window : マイグレーション等にさくことができる運用時間中の割合)	一回の運用サイクルのうちで、メディアリフレッシュメントやマイグレーションを行うことができる時間の割合

表 5-4. ストレージ・パフォーマンス・モデルの各要素の意味⁵⁷

この計算式の各要素に数値を入力して計算することで、電子情報保存のためのシステムがメディア・マイグレーションを行うための能力を知ることが可能となる。この数式の要素の中で、鍵になる重要なものは、MMW (Medium Migration Window)、マイグレートされるデジタル・オブジェクトの数 (# Object : Number of Objects) 及びその容量である。これらの数に大きく影響を受けながら、電子情報保存のためのシステムが保有するストレージのパフォーマンスを検討することが可能となる。

⁵⁷この表の項目以外に、MDT (Medium Decay Time : メディアの陳腐化にかかる時間)、すなわち読み出し、書き込みエラーがなく正常にメディアが働き続ける時間を指す指標が存在する

メディア・マイグレーションのための戦略

DNEP プロジェクトでは、デジタルデータのマイグレーションを行うための戦略を定めており、これを用いることによって、マイグレーションの時間を短縮化することが可能であるとしている。その戦略を図示すると下記の通りとなる。

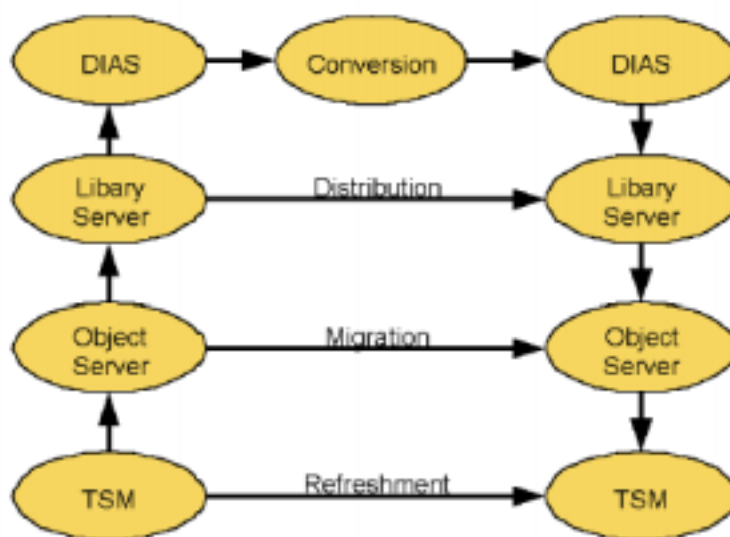


図 5-13. DNEP におけるメディア・マイグレーション戦略

上図は、DIAS のシステムアーキテクチャーのうちで、デジタルデータの長期保存に大きな役割を果たすコンテンツマネージャの概念的な構成を図示したものである。コンテンツマネージャはライブラリサーバー (Library Server)、オブジェクトサーバー (Object Server)、チボリストレージマネージャ (TSM) から構成されている。まず、それぞれの機能の役割について概説する。

コンテンツマネージャの機能	概要
DIAS	<ul style="list-style-type: none"> 保存対象を受け付ける時のインターフェイスとなる
ライブラリサーバー	<ul style="list-style-type: none"> 保存されている情報のインデックス、ユーザー認証、アクセス制限等を行うためのカタログ情報を保持する 検索機能を提供し、保存対象の位置の特定を支援する 保存対象に対してセキュリティを維持したアクセスを提供する オブジェクトサーバーと交信を行う

オブジェクトサーバー	<ul style="list-style-type: none"> ・ ステージングサーバー⁵⁸として機能する ・ パージャー (Purger) ⁵⁹として機能する ・ マイグレーションの管理を行う⁶⁰
TSM ⁶¹	<ul style="list-style-type: none"> ・ 実際のストレージ機器へのアクセスを管理する

表.5-5 DIAS におけるコンテンツマネージャの機能

これらのコンテンツマネージャの機能に応じて、メディア・マイグレーション・プロセスを迂回することによって、すべて DIAS インターフェイスを介してマイグレーションするよりも、時間の短縮が図れる。図 3-13 に示したマイグレーション戦略の概要を以下にまとめる。

⁵⁸ ステージングサーバーは、保存対象を一時的に保存する役割を担う。通常は、ステージングサーバーから、コピーを行い、ストレージエリアに保存を行う

⁵⁹ ステージングサーバー上にある情報のアクセス履歴を管理し、頻繁にアクセスがあるものについては、ステージングサーバー上に残してアクセス時間の短縮を図る機能を持つ

⁶⁰ ストレージ階層構造に応じて、マイグレーション時にストレージから保存対象を読み出す役割を担う（実質的には、TSM が大本のストレージ管理を行うため、マイグレーションを管理することになる）

⁶¹ Tivoli Storage Manager は IBM のストレージ管理用の製品である

マイグレーション戦略	概要
Conversion (変換)	保存対象のオブジェクトである AIP (Archival Information Package) に変換を伴う場合に行われる
Distribution (配信)	オブジェクトサーバー間で、AIP が移転される場合に行われる。ライブラリサーバーが、移転される AIP と移転先の間を管理し、一貫性を取れるよう配慮する
Migration (移行)	ディスクドライブからテープドライブに AIP を移動するといった場合に行われ、通常は、AIP の変換は伴わない
Refreshment (リフレッシュメント)	古いテープドライブから新しいテープドライブに保存対象が移動される (すなわち、保存対象が保持されているのと同じ種類のストレージ・デバイスに AIP が移動される) 場合に行われる。本プロセスは TSM によって管理される

表 5-6. メディア・マイグレーション戦略の概要

上記のメディア・マイグレーションの種別に応じて、通常の情報取得、読み出しプロセスを迂回して、マイグレーションの時間を短縮することが可能になる。また、これ以外の方策として、ストレージ等のハードウェアの増強や、負荷分散装置による負荷分散、ネットワーク帯域の拡充等もあわせて行うことによって、円滑なマイグレーションが行われるとされている。

5.3.3. XML

ここではオランダ内務省が中心となって行った Digital Preservation Testbed プロジェクトにおける XML の活用について整理する。

Digital Preservation Testbed における XML を活用した e-mail の保存の実験

本プロジェクトにおいては、e-mail を複数の情報の集合体ととらえている。すなわち、e-mail はメッセージが経由してきたネットワークノードや文字コードの情報を記述したヘッダー情報、メッセージの本文、添付ファイル等で構成されるということである。このように e-mail を情報コンポーネントの集合体ととらえ、以下に述べる保存アプローチが採られている。

アプローチ		概要	結果
後から一括して e-mail の保存処理を行う	All-in-one アプローチ	<ul style="list-style-type: none"> ・ JAVA ベースのツールで e-mail を XML に自動変換し Oracle のデータベースに格納する ・ ヘッダー情報、メッセージ情報、添付ファイル情報のそれぞれに XML のタグを振って処理する 	<ul style="list-style-type: none"> ・ メッセージ部分が通常のテキストファイルであれば、問題なく保存が可能 ・ メッセージ部分が HTML 形式になっている場合に正しく XML 変換が行えなえず、保存ができないという問題点がある
	Split -File アプローチ	<ul style="list-style-type: none"> ・ JAVA ベースのツールで e-mail を XML に自動変換し、Oracle のデータベースに格納 ・ e-mail のヘッダー情報のみ XML 化し、それ以外⁶²は元のままの状態に保存する 	<ul style="list-style-type: none"> ・ 文書やメールの背景画像を添付している HTML 形式等を使ったメールを保存することが可能 ・ メッセージ部分をネイティブフォーマットのまま保存するため、情報の真正性を保持することが可能
実運用にあわせて e-mail の保存処理を行う		<ul style="list-style-type: none"> ・ Microsoft Outlook2000 をベースに add-in ソフトを開発 ・ e-mail を送信する際に情報の送信者にメタデータを付与させる ・ e-mail を XML 形式に変換してサーバーに一括保存する 	<ul style="list-style-type: none"> ・ ヘッダー情報以外に保存に必要なメタデータをユーザー自身に付与させるため、情報管理がしやすい ・ サーバーに XML 形式で情報が保存できるため、長期保存が可能になる

表 5-7 . Digital Preservation Testbed における e-mail の保存実験の概要

本プロジェクトの実験では、All-in-one アプローチの課題点を解消するために Split-file アプローチが行われ、さらに e-mail の管理を容易にするため、保存に必要なメタデータをユーザー（送信者）に付与させる Add-in ソフトを活用したアプローチが取られた。Add-in ソフトを活用した e-mail 保存のアプローチの処理概要を以下にまとめる。

⁶² 「それ以外」のコンポーネントには、e-mail のメッセージに付与するメタデータ、メッセージの本文、添付ファイル等がある。また XML 化されるヘッダー情報のうち、コンテンツタイプヘッダー（メッセージのエンコーディングの方法を規定している）とメッセージ本文の情報は緊密にリンクが保持されることとなっている

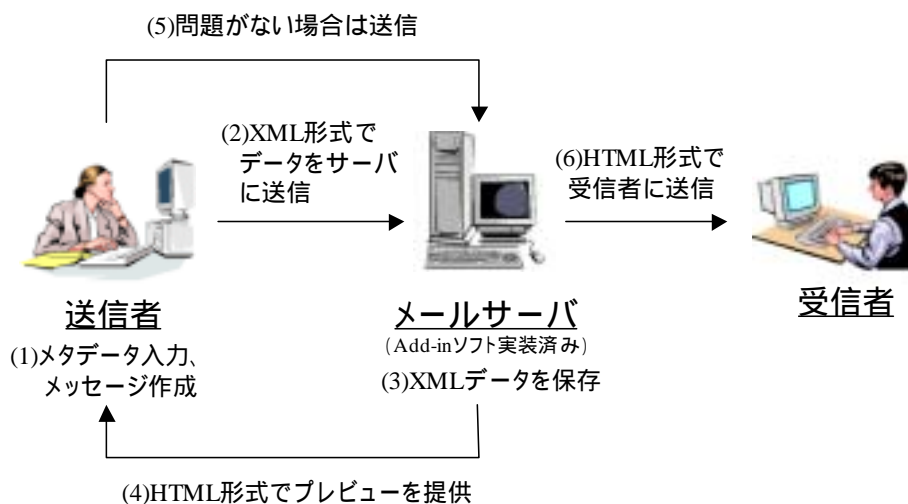


図 5-14. Digital Preservation Testbed における e-mail 保存の業務フロー

上記の業務フローを経て、サーバーにおいて XML 形式のデータが保存されることとなる。この XML データには、ユーザーが付与した保存のためのメタデータが付与されているため、情報の管理が容易になる。

実験の結果について

XML による e-mail の保存は、その業務フローも含めて有効性が検証されたと本プロジェクトではとらえており、今後は、第二世代の Add-in ソフトウェアを作成し、更なる実験が行われる予定である。この方法を他の組織に応用することも考えられるが、その場合には、保存のためのメタデータの形式等を考慮することが必要になる。

5.4. 電子情報保存に係る技術動向の考え方について

諸外国における電子情報保存のための技術研究には、非常に先進的な取り組みが多い半で、そのほとんどが研究の途上にある。

保存対象となる電子情報の見た目と感覚（"Look and Feel"）を保持することは、その時代の文化遺産でもある電子情報を後世に伝えるために重要であり、エミュレーションやマイグレーション等の技術を活用することが非常に有効である。しかし、後述するようにこれらの技術には多大なコストを発生する場合もある。

コスト要素の把握のために各種の実験を十分に行うとともに、どのような情報をどの程度のレベルで保存するか判断する必要がある。しかし、情報の価値は時代によって変わるものであり、何を「価値がある」と決めるかは非常に難しい問題である。

電子情報保存のための技術としては、本文で触れた XML のように高度に標準化が進み、再生において特定のハードウェアやソフトウェアに依存しない技術が有望である。今後も、XML は電子情報保存において重要な役割を果たすと考えられるが、同様に CAMiLEON プロジェクトにおける Migration on Request、エミュレーションの例のように、オリジナルの電子情報をそのまま保存し、将来の環境で活用することも一つの方法であろう。オリジナルの情報が改変されず保存されていれば、将来的に発展する情報技術によって完全な形で情報が再現できる可能性があるからである。

そのためにはオリジナルの電子情報を再現するために必要なハードウェア環境、ソフトウェア環境、データフォーマット等といったいわゆる「表現情報（Representation Information）」を十分に記述し、保存対象に付加することが必要であり、これは後述するメタデータの取り組みと関係が深い。

再生環境の保持のためには、第 2 章で触れたオーストラリア国立図書館の取り組みに見られる、電子情報の再生のためのソフトウェア・レポジトリの検討も必要である。各国が共同で電子情報の再生のために必要なソフトウェア・レポジトリを構築し、それをネットワークで共有すれば、再生環境の保存のための責任と費用の分散化が図られ、コスト効率の高い電子情報保存が可能となる。このような考え方を基に、諸外国では電子情報保存の活動における「協調」の重要性が説かれている。

また、本文で触れた DNEP プロジェクトのメディア・マイグレーションのように年を追うごとに増大する電子情報を如何に効率的に保存し続けていくかといった課題に対する方策の検討も重要である。そのためには、前述した保存のための情報システムを視野に入れてつつ検討が行われなければならない。

今後の電子情報保存における技術の研究に当たっては、先行する諸外国の動向を把握し、経験の共有化を積極的に図るとともに、各種の実験プロジェクトを実施しつつ日本国内の現状にあった保存技術を選択・確立することが重要である。

6. コストについて

6.1. 議論の背景

電子情報保存に関する様々な取り組みが行われてくるにつれて、技術問題に付随してコスト問題が議論されるようになってきた。NEDLIB では OAIS モデルの利用検討が行われ、CEDARS では技術面に関してテストサイトを持つ等の取り組みが行われているが、そうした技術の検討後に「実際に電子情報保存にかかるコストはどこが持つのか」という問題が生じている。

6.2 . 保存コストに対する考え方

6.2.1 . 保存コスト検討の必要性

電子情報保存を行う際に予算が十分につくことはまれであるので、オペレーションコストは十分に考慮しなければならない。例えばオランダでは、納本は任意であるため、電子情報保存には公式の予算は付いていない状況である⁶³。

コストに関する問題でキーワードとなるのが、「コスト・イフェクティブ」という言葉であるが、これまでのところ明確な基準は現れていない。しかし、電子情報保存に関わる論点のひとつとして保存コストについての調査研究は行われていて、既に報告が提出されているものもある。

保存コストへの関心が広まったのは、2001年8月のRLGレポートによるところが大きい⁶⁴。このレポートでは、電子情報は紙媒体と異なり、特にメンテナンス及びアクセスという点で保存コストがかかると指摘している。また、費用認識に当たり、紙媒体とは異なったアプローチの必要性も述べられている。

6.2.2. 保存コストに当たっての認識

紙の印刷物においては、物理的な保管がそのまま資料の保存を意味するのに対して、電子情報の場合、データの保管とあわせて資料データから内容を読み出すための手段・設備を持ってなおかつそれを維持していかなければならない。

何を保存し何を保存しないのかという資料自体の選択や、それをどのように保存するのかという手段の選択は、紙の資料を保存するのに比べて、はるかに多く検討すべき要素がある。

電子情報の保存コストは紙の資料の保存以上に継続的な資源の投入と取り組みとが求められる。

⁶³ R.Dekker らの An electronic archive for academic communities による

⁶⁴ R.Dekker らの An electronic archive for academic communities による

従来と異なる保存コスト

複雑な電子オブジェクトを長期保存するコストに関して、すべてが明確化になっているわけではないが、既存資料に対するよりも継続的な資源投入が必要になるというのが図書館関係者に共通の理解である。

紙媒体の保存においては保管環境の整備コストが中心となるのに対して、電子情報の保存では定期的なデータチェックコスト、アクセスの維持コスト等が中心となる。アクセスの維持コストとは、電子情報にマイグレーション等を行って技術変化に対応させつつ、常にアクセス可能な状態を維持するためのコストである。さらに電子情報の保存には、期間やスケジュールの異なる複数の技術戦略が必要な場合もある。

保存戦略は、デジタル技術の急速な変化に対応させていかなければならない。保存活動に関連するすべての事象について明確なコスト予想を行うことは困難だが、対処は必要となっている。

保存コストの要因の見極め

電子情報保存にどのくらいのコストがかかるのかははっきりしていないため、保存に関する責任の所在が不明確になっている。何にコストがかかるのか、どこを節減すればいいのかを理解した後に保存作業を行うことができればよいのだが、電子情報の保存には正確な情報を待つ時間的な余裕はない。

重要なのは、何に最もコストがかかるのか、それはどのようなコストなのかを、実践の中で把握することである。そこにスケールメリットが見いだせる可能性がある。関係各機関との協力により削減できるコストもあるだろうし、納本フォーマットの特定やメタデータの記述方法を整備することによってコスト削減を図ることも可能である。

また、電子情報の保存を図書館の通常業務に組み込めば、必要なスキルと知識を組織内で早く効果的に得られるであろう。共同納本ネットワークとサービスのモデルを研究開発することも考えられる。

6.3 . 保存コストにおけるフレームワーク

実際に電子情報の保存コストに関してフレームワークの整理を行い、実務に携わる人々が検討すべき点を明らかにしている例として、以下にリーズ大学とミシガン大学が共同で実施している電子情報保存研究プロジェクトである Cedars、オーストラリア国立図書館が開設している電子情報保存のためのサブジェクトゲートウェイである PADI 及び英国の学術研究団体と公文書関係の組織が連携して行った JISC/NPO の例を取り上げる。

6.3.1. Cedars の例

Cedars では、保存コストをさまざまな観点から検討しており、具体的な事例をあげて保存コストをどのように検討すべきかを研究している。

保存コストをめぐる側面

保存コストを検討するに当たって、問題となる側面がある。電子情報を保存するに当たり、従来の紙保存と異なる考え方をしなくてはならない側面である。

時間的側面

紙の印刷資料においては、資料収集を行った後ある程度の年月が経ってから、資料保存の問題が現実的になる。つまり、資料収集と保存の間である程度の時間的な猶予があるといえる。これに対し、電子情報については、収集後に保存が問題となるまで時間的な間隔が非常に短い。

例えば、紙の印刷資料を電子化する場合、どのフォーマットで電子化するかという選択は、その資料をどのように保存するのかという選択でもある。同様に、資料保存の手段を選択することは、その資料にどうやって将来アクセスするのかという選択に影響を与える。この意味で、保存コストのある部分は資料作成の時点で決まってくる。

戦略的側面

電子情報の保存のために採用する戦略も、保存コストに大きな影響を与える。例えば、保存戦略としてマイグレーションを採用すれば、保存資料が必要とするソフトウェアがアップデートされる度に、移行作業を行うことになる。

資料価値の考え方

ある電子情報に、どの程度保存する価値があるのかという点については、コレクション全体の方向性に照らし合わせた資料選択が必要になる。

費用対効果と資料価値

電子情報の保存コストの費用対効果はトレードオフの関係にある。つまり、電子情報であるがゆえ完全性を求めればコストは上昇し、コストを抑えれば資料の完全性において妥協しなければならない。このような点を踏まえると、保存機関が電子情報の保存戦略を策定する場合、費用対効果を考慮すべきといえる。自機関の電子

情報は資料としてどの程度価値を持つのか、またその価値に見合った保存を行うにはどの程度のコストがかかるのか検討しておかなくてはならない。その意味で、まず電子情報保存の対象となる資料選択を行う基準が、改めて検討されなければならないだろう。

資料価値の判断

まず、資料選択において重要なのは、選択がコレクション・マネジメント・ポリシーと一致していることである。加えて、その資料の保存に必要な技術的要件を合わせて検討しなければならない。

コレクション・マネジメント・ポリシーのうち、電子情報保存は一部にすぎない。しかし、電子情報の作成・収集及び保存は密接に結びついている。このため、作成・収集の段階で、資料の長期的な保存価値とコレクション全体との適合性を判断する必要がある。

さらにこの資料を保存して、データ内容を再現しつつ、資料へのアクセスを確保するためには、必要とされる技術的要件について判断しなければならない。ここでいう資料へのアクセスレベルは様々である。

どのレベルのアクセスを確保するのかという判断は、電子情報の「重要な属性⁶⁵」(significant properties)に拠るといえる。つまり「重要な属性」に対する判断によって、保存すべき情報量やメタデータが決定される。例えば、法的な資料には厳格な保存が求められるというような場合である。

⁶⁵「重要な属性」は技術問題のみを示すわけではない。図書館員や専門家の判断により利用者ニーズを考慮した属性をも含んでいる。例えば、電子ジャーナルの保存に当たって「テキストの中身が重要であるためどのような形式であれ、テキストが読める環境を維持する」と保存関係者が判断した場合、それは当該電子ジャーナル保存における「重要な属性」となる

OAIS のアーカイブモデルを用いたコストの考察

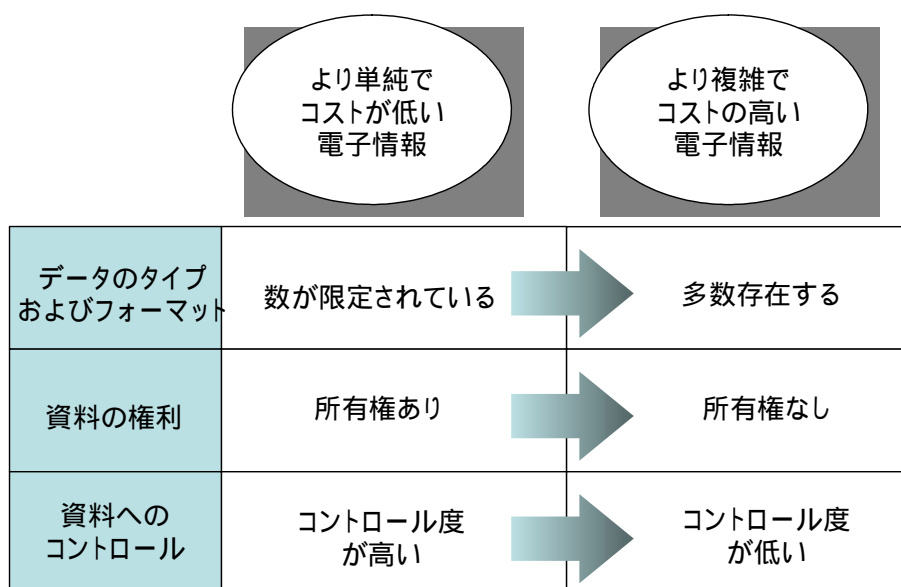


図 6-1. OAIS リファレンスモデルを利用した場合の保存コスト

データのタイプ及びフォーマット

保存コストの観点からは、コンテンツの多様性も含めて、データタイプ及びフォーマットの種類が重要な要素となる。

テキストとイメージ画像から構成される単純な情報保存では、フォーマットは限定される。データベース、GIS アプリケーション、複数の画像・動画によるインタラクティブ・マルチメディア等の場合は、多数のデータタイプやフォーマットを用いており、複雑な情報保存である。

この場合、サポートすべきデータタイプやフォーマットの種類が増えれば、情報保存のコストも上昇する。つまり技術的に行うべき対処が増えると、金銭的な負担も上昇する。また、電子情報にアクセスし、再現する際、品質が高く、種類が多くなれば、技術的な複雑性が増すため、保存コストは上昇する。

資料の権利

情報保存に関する権利を、どこが保有するかによってコストは変動する。自らが電子情報の所有権を持っていればコストは自らの実施範囲内で納まるが、データの所有権がない場合は、権利者との交渉作業が必要となる。これに伴って、権利関係の処理にかかるコストが生じる。

資料のコントロール

電子情報を提出させる際等において、どの程度資料をコントロールできるかによって、コストは変動する。

例えば、そのデータのフォーマットを指定する権利があれば、対応すべきフォーマットを限定して受け入れることができるのに対し、指定する権利がない場合は資料提供者の使用するフォーマットを受け入れなければならない、コストが生じる。

保存コストにおける要素

OAIS におけるアーカイブ類型からは 3 つの問題点が明らかとなったが、この問題点を念頭に置きつつ、保存コストにおける要素を分解すると以下の 9 つになる。

保存対象の選択

資料の長期保存の選択に関しては、「コレクションマネージャー」と「システムマネージャー」の協同作業が求められる。「コレクションマネージャー」は、例えば図書館員や（各分野の）専門家であり、資料の利用方法やコレクション全体にとっての資料価値等についてアドバイスを行う。「システムマネージャー」は資料保存に要求される技術的問題やメタデータ作成等に関するコストについて検討を行う。

保存のための権利交渉

交渉担当者の稼働時間や契約書の作成や調整等の人件費が含まれる。また、保存資料に必要なソフトウェアや技術的構成要素等に関連する権利調整も含まれる。経験的には長期間の作業となる場合が多い。

保存資料へのアクセス権交渉

すべての資料に当てはまるわけではないが、エンドユーザーからの資料へのアクセスを短期的なものとしている場合に必要となる交渉と調整である。

保存戦略の適応

保存資料への長期的なアクセスを確保するために必要な技術を確認するのに用いられる時間的コストも含む。この作業には、デジタル資料の「重要な属性」を確定する検討作業が必要になる。コスト要素にはソフトウェアやハードウェアの購入又は設計費用が含まれる場合もある。

AIP⁶⁶の品質確保と確認

納入された電子情報の完全性がチェックされなければならない。保存対象の電子情報に付随するドキュメントの確認作業も含まれ、多くの場合かなりの人的資源が必要となる。

⁶⁶AIP (Archival Information Package) とは、OAIS リファレンスモデルにおける用語である。内容情報 (Content Information) とそれをサポートする保存記述情報 (Preservation Description Information) を含む情報のまとまりを指す

メタデータの作成

保存対象の資料に関するドキュメントを元にして、公開に必要とされるメタデータを作成するコストである。メタデータは、電子情報自体の機能を既存のコレクションマネジメントとどのように統合させるかという点を考慮して選択・開発がなされる。メタデータには資料の権利関係についての情報も記載される。

ファイルの保管

ハードウェアとソフトウェアの維持及び購入を含むコストであり、定期的なファイルの検証及び保管メディアの検証も含まれる。

電子情報保存の管理

電子情報保存の管理とは電子情報保存の機能として考えられている資料収集・保管・公開/普及における一般的な管理に加えて、技術開発や法的処理の費用をも含んでいる。電子情報保存のポリシーを変更した場合、システムの変更費用が加わる場合もある。

一般的なコスト

人件費（給与、間接費用、トレーニング、スキル向上等）や、保険、施設の維持費等が該当する。

これらのコスト要素の流れは、以下ようになる。

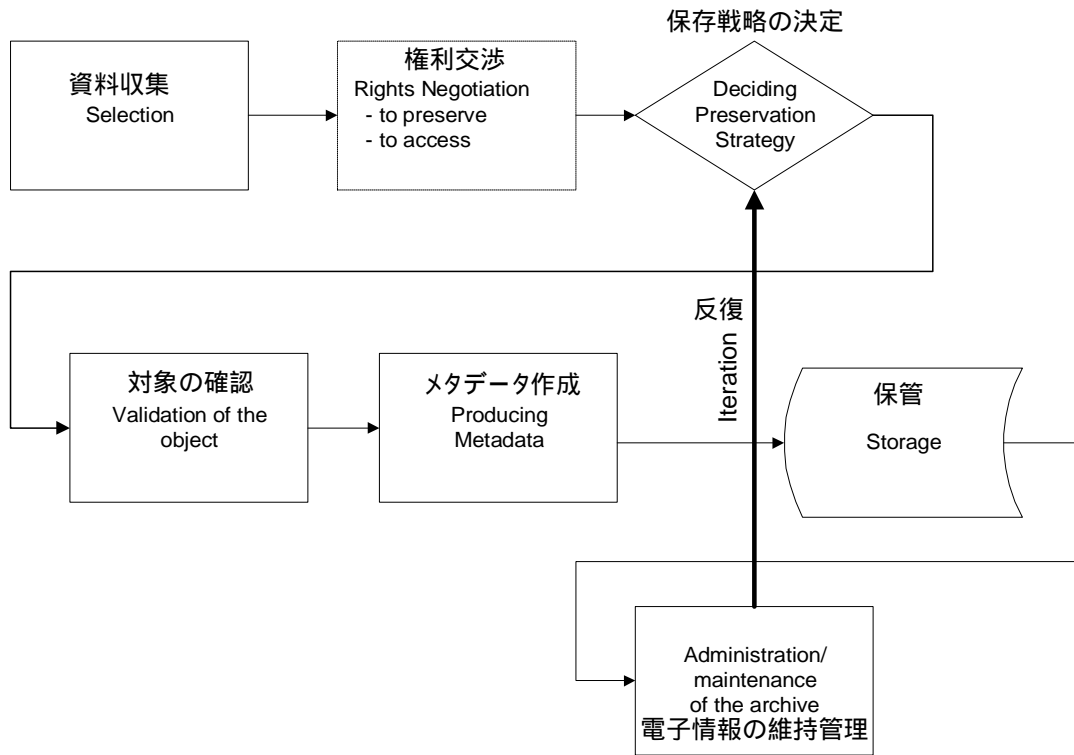


図 6-2. 電子情報の保存コスト要素

実際のコスト要素の構成については電子情報の種類によって異なるだろう。例えば資料についての権利を保有している電子情報であれば権利関係の交渉は必要なくなる。

また、「電子情報の維持管理」から「保存戦略の決定」に向かって矢印が反復(iteration)を示しているのは、対象となる電子情報(object)に対して最適な保存戦略が長期的には変化することを意味している。

保存コスト軽減の原則

電子情報の保存コストは多くの要素からなる場合が多いが、以下の2事項が保存原則として挙げられる；

電子情報の保存は、資料(資源)の寿命があるうちに、可能な限り早く検討されなければならない。理想的には、資料の制作時点あるいは収集時点で保存がなされるのが望ましい。早期の保存活動はコストも抑制する。また、他の電子情報と共同して収集・保存する等によって、作業の効率化を図ることが可能となるため、コスト要素の負担を軽減することができる。

6.3.2. PADI の例

Cedars 同様、PADI においても保存コストへの関心が高まっており、まず領域を設定して保存コストを算定する方法を薦めている。領域設定では Cedars のコスト要素と重複する箇所も多い。

電子情報の保存コストに係る領域

PADI では、保存コストに係る領域を以下のように設定している。

電子情報のストレージ

記憶装置のコストは年々低下しているが、特にサーバーでの保管を行う場合はかなりの費用がかかっている。

メタデータの作成

従来のカタログ化作業においてもコストがかかっていたが、メタデータの作成においては既存のケースと同等かそれ以上の作業コストが見込まれる。

権利関係の交渉と法的処理

すべての資料に関して、著作権等の権利関係の交渉や法的な対応を行わなければならないためコストがかかる。

資源のメンテナンス

定期的なデータチェックを行い、データの完全性やアクセスを確保しなければならないためコストがかかる。

ファイルのマイグレーション

あるファイルを移行させるには、時間及び金額のかかる作業であり、コスト領域として考慮しなければならない。

エンドユーザーへの情報提供

資料へのアクセスを承認したり、使いやすいユーザーインターフェイスの開発・維持することは、アクセスの確保を含めて、電子情報の提供の一環として、コストがかかる領域である。

上記領域について、さらに資料の管理という点に着目し、作業項目の単位で考慮した場合には、以下の 8 点が挙げられる。

- ・ データの作成
- ・ データの選択
- ・ データの評価
- ・ データマネジメント
- ・ 資源の公開

- ・ データの使用
- ・ データ保存
- ・ 権利関係の交渉

試算

電子情報の長期保存に関するコストの実例はまだほとんど見当たらない。しかし、PADI による独自試算による、オンライン出版物の初版資料を収集する作業量は、印刷資料に比べて5倍の作業量が見込まれている。

電子出版物を25年間保存する費用は、管理費用も含め、印刷出版物に比べ20倍の費用が見込まれるという試算もある。ただし、この試算では年間に500～1000タイトルの新刊書を想定しているが、新刊が1万タイトル以上の場合は5倍程度になるとしている。

また電子情報の保存は急速なコンピュータ技術の変化に伴い、その時々標準技術に合わせた対応(マイグレーション等)を5年ごとに行わなければならないとの見解もある。

6.3.3. JISC/NPO 研究の例

次に、英国図書館が中心となった電子情報保存に係る JISC と NPO の協同研究における保存コストの研究事例を紹介する。

電子情報のライフサイクル

この協同研究においては、電子情報の保存コストはコレクション全体の管理と密接に関連しており、保存コストだけを単純に切り出すことは困難かつ不適切であるという視点に立ち、電子情報のライフサイクル（寿命）を軸にした検討を行っている。具体的にこの研究では、AHDS (Arts and Humanities Data Service) の枠組みをベースにし、電子情報のライフサイクルに関して大きく 4 つの段階を区分している。

1	データ作成	電子情報が作成される際に行われる決定事項。これらは多くの場合、コレクション管理者のコントロール外で行われるが、その後の利用手段を選択するに当たり大きな影響を与える。
2	データ管理と保存	
	データ獲得、維持または処分	電子情報の内容、有用性、ユーザーへの配慮等の決定；データの管理、カタログ化、アクセス確保や保存に当たっての扱い易さに影響を与える。
	データ構造	電子情報のフォーマット、圧縮、エンコードに関わる。
	データ記述とドキュメンテーション	電子情報の構造、内容、出自、履歴について、ドキュメント化が行われている度合い。
	データ保管	電子情報を保管するために利用されるハードウェアと保存媒体。
	データ保存	時間や技術の変化、磁気媒体の劣化によるデータ破壊から電子情報の内容を保護すること。データの種類や構造によって、適切な戦略は異なる。保存要件は、電子資料の構造化、付随ドキュメント、保管方法、確認方法に影響を受ける。また、エンドユーザーが電子情報へアクセスする条件及び方法にも影響を受ける。
3	データ使用	電子情報の提供及び使用方法に関する決定。これは電子情報の作成状況に依存しており、管理方法に影響を与える。
4	権利関係の交渉	知的所有権、データ保護、データ機密問題、データ獲得、配布時のライセンス問題や実行手順を開発する必要がある。

表 6-1. 電子情報のライフサイクル

これらのライフサイクルに関連する要素や課題は相互の関連性が強い。例えば、コレクションに電子情報を加える、印刷資料の電子化を行う等の決定を下す場合、その内容とフォーマットの決定を伴う。

内容とフォーマットの決定は、電子情報がどのように管理・保管できるのか、長期保存が可能なのか、エンドユーザーにコピーを配布できるのかという問題に大きな影響を

与えることになる。

電子情報保存の7つの領域

前述の電子情報ライフサイクルをフレームワークとして、研究報告では保存コストを以下に示す7つの領域に区分しており、これは Cedars 等で取り上げられた要素とほぼ同様である。

1. データ制作
2. データの選択と評価
3. データ管理
 - データ・ドキュメンテーション
 - データ検証：査定、複写、保存媒体の回復
 - データ構造の変換
 - データ保存装置（ストレージ）
4. 資料の公開
5. データの使用
6. データの保存
7. 権利関係の調整

7領域から構築したコストモデル

電子情報のライフサイクルから考えられた7領域に関して、検討すべき要素とその保存コストを組み込んで、コストモデルを検討する。

コストモデルの概要

表2で示したコストモデルを元に、それぞれのコスト要素、実行時に留意すべきポイントを概観する。その目的は以下の通りである。

- ・ それぞれの領域で行われる実行課題を明確にすること
- ・ 各実行課題に伴う主要コストを特定しコストの低減、上昇の要因を示すこと
- ・ 保存に直接・間接に関係するコストを示すこと

	領域	考慮する要素	保存コスト
1	データ作成	<ul style="list-style-type: none">• コントロールのレベル• 総原価• 「最良事例（Best Practice）」のコスト• データセンターコスト	<ul style="list-style-type: none">• 「最良事例」の実行• 電子情報のクリーニング

2	データ選択/評価	<ul style="list-style-type: none"> コレクション・ポリシー 評価のための技術的/実践的価値 複雑性 	<ul style="list-style-type: none"> 実践的及び技術的評価
3	データ管理	<ul style="list-style-type: none"> ドキュメンテーション データ検証(査定、複写、回復) データ構造変換 データストレージ 「データ考古学(Data Archaeology)」 	<ul style="list-style-type: none"> ドキュメンテーションの読解/編集/管理 検証(査定、複写、回復) データ構造変換 オフラインの電子情報保管 媒体からのファイル復旧: 出所が不明なファイルの出所解読
4	資料の公開	<ul style="list-style-type: none"> オンライン資料探索エージェント ゲートウェイ オンラインカタログ 	<ul style="list-style-type: none"> 保存後の公開コスト
5	データの使用	<ul style="list-style-type: none"> オンラインアクセス 電子配布メディア 	<ul style="list-style-type: none"> 保存後の使用コスト
6	データの保存	<ul style="list-style-type: none"> 技術保存 	<ul style="list-style-type: none"> 旧式のハードウェア/ソフトウェアの保存 第三者機関サービスコスト
		<ul style="list-style-type: none"> 技術エミュレーション 	<ul style="list-style-type: none"> 第三者機関サービスコスト
		<ul style="list-style-type: none"> デジタル情報マイグレーション 	<ul style="list-style-type: none"> 媒体変更(レコーディング、電子情報保管) 後方互換性⁶⁷(セットアップ、実行、チェック) 動作性(セットアップ、実行、チェック) 標準形式への変換(標準フォームの合意、テスト、実行、テスト解除)
7	権利関係の交渉	<ul style="list-style-type: none"> 使用制限の可能性 納本者/ユーザーの権利 	<ul style="list-style-type: none"> 許可手続き、管理、証明コスト

表 6-2. コストモデル

1. データ作成コスト

電子情報の作成過程で、例えば納本者に対して、どの程度のコントロールが行えるかは、電子情報保存の運営主体によって異なる。しかしデータ作成時点でできる限りの適切なコントロールを行うことが望ましい。この領域での管理は全体コストに

⁶⁷ソフトウェア等において、古いバージョンと同じコードを新しいバージョンでも利用できること

大きな影響を与えるからである。

例えば、電子情報の内容検証やドキュメンテーションを確実にできるのは作成者(= 納本者)である。作成段階で、提出資料に不備があった場合は、その後電子情報保存側(データセンター)で「クリーンアップ」を行わなければならない。

この場合、資料データ、ドキュメンテーション、付帯情報等が失われているため過去に遡って「クリーンアップ」を行うと仮定した場合、完全な資料を受け取った場合の10倍のコストが見込まれる。

データ収集において、最良事例を見つけて実施すること(ベストプラクティス)はその後の管理を容易にし、コストの低減にもつながる。データ作成段階におけるコストは大きく二つの領域にまとめられる。

- ・ 納本者に対して適切な納本作業を奨励・啓蒙するコスト
これは例えば、Web への納本ガイダンスの掲示や説明会の開催等である。データ不備を予防する何らかの活動が求められる。
- ・ データ作成(納本)段階での誤りの修正や不適切な手順の修正を行うコスト
第一次的に行う修正(ファーストコスト)のかけ方が不十分であると、その後のクリーンアップ作業等(セカンドコスト)への負担が大きくなってしまう。電子情報をベースにした場合、第二次コストは第一次コストを遥かに上回る。電子情報のクリーンアップにかかる費用は無制限にもなりうる。これらを抑制するには三つの方法がある。
 - データセンターの予算が固定されている場合、「クリーンアップ」活動に配分する予算を最大化する
 - 基本的な標準事項を設定する
 - 電子情報が提出に関する最低条件のドキュメンテーション等を満たさない場合は、受け取り拒否を明記する

2. データ選択/評価コスト

電子情報の選択/評価は収集機関のコレクション・ポリシーに準拠する。技術的/実践的な評価基準に基づいて資料評価を行うことは更なる保存コストの問題と関連している。つまりコストが発生すると思われる点について評価を行っておくわけである。

資料の管理、カタログ化のしやすさ、エンドユーザーからのアクセスの容易さやデータセンターでの保存のしやすさ等が対象となるだろう。この領域のコストは資料のサイズやデータの複雑性、ドキュメンテーションの充実度によるところが大きい。

3. データ管理コスト

- ・ドキュメンテーション

資料のドキュメントを確認し、編集や追加を行う。必要であれば、ユーザーがドキュメントを利用可能にし、ドキュメントのアップデートを行う。ドキュメントの提供が不十分であると、資料のテストを行い追加のドキュメンテーションを行わなければならない、コスト増の要因になる。また、ドキュメント自体が十分であっても、読解と研究にコストが発生する場合もある。

- ・検証

データ検証は資料の完全性を確保する手順である。具体的には以下の三つの作業がある。

- ・電子情報がドキュメントに示される通りの完全性を持っているか：査定
- ・メディアが破壊されたりダメージを受けたりする場合のバックアップを確保する：複写。
- ・媒体品質の低下によって閲覧できないことを防ぐ：回復（リフレッシュング）

これらは相互に関連するプロセスであり、電子情報のライフタイムで何度も繰り返され、その度にコストが発生する。

- ・データ構造変換

データ構造は電子情報がどのようにフォーマット、圧縮、エンコードされているかに関わる。オリジナルの電子情報がどのようなデータ構造を持つのかは、将来のストレージやアクセスに関するコストのレベルを特定するのに大きな影響を与える。資料がデータセンターのサポートしていない特定のアプリケーションとフォーマットで納本された、長期保存のために標準フォーマットへの変換が必要になる場合はコストが増大する。

- ・データストレージ

データストレージの選択は、利用可能な資料とそのデータ量、データの使用方法と保存方法によって影響を受ける。コスト特定の観点からはオンラインとオフラインのストレージを区別する必要がある。オンラインのストレージにおいては、ユーザーのアクセス要件とデータセンターの管理要件を満たすことが求められる。一方、オフラインのストレージでは特にデータ保存に特化した要件を満たすことが求められる。

- ・データ考古学

データ考古学は不完全なデータを回復するプロセスや出自の不明なファイル进行分析する場合に求められる。例えば、破損データの復旧や陳腐化したシステムを再び活用する場合等が含まれる。

4. 資料公開コスト

資料公開コストは、特定の電子情報の情報をエンドユーザーが入手する場合に発生するコストである。これには、オンライン資料探索エージェント、ゲートウェイ、オンラインカタログ等の情報探索ツールの開発が含まれる。

これらのツールにかかるコストは多くの場合、資料の保存に直接関わるものではない。しかし、その資料の活用が頻繁な時期を越えて、電子情報として保管された後に、保管された対象を探索して再び利用するコストは、保存コストに含まれる。

5. データ使用コスト

電子情報をエンドユーザーに提供するコスト。このコストは資料の保管構造やユーザーニーズ、資料提供がオンライン又はオフラインで行われるかによって変化する。しかしながら、ユーザーへの資料提供コストは、提供が電子情報から直接行われる場合を除いて、資料保存に直結するコストではない。

6. データ保存コスト

データ保存コストは、どんな保存戦略を採用するかによって異なるため、いくつかのケースを示す。

- ・ 技術的保存

技術的保存を戦略として採用した場合、電子情報の作成とアクセスに用いられたオリジナルのアプリケーションプログラム、OS、ハードウェアプラットフォームを保存することになる。

保存コストには2つのタイプがある：

- ・ 古いハードウェア及びソフトウェアの保持コスト

新たなハードウェアとソフトウェアの環境に移行した後、古い環境でしか動作しないアプリケーションを持っている場合、マイグレーションでカバーするまでの間、古いハードウェアとソフトウェアを保持しなければならない。

- ・ データセンターにおける環境適応コスト

新しい環境に移行した後、自データセンターが保有する環境では動作しない電子情報があるとする。この対処のために、古いハードウェアとソフトウェアを保有する第三者機関を探してプログラムとデータを読み込まねばならない。そして、自データセンターの環境にマイグレートする費用を支払うことになりコストが発生する。

- ・ エミュレーション

この戦略を採用した場合、現在及び将来のプラットフォームで動作するエミュ

レータプログラムの開発コストが発生する。実際には、必要なハードウェア/ソフトウェア環境をエミュレートする設備を持つ第三者機関への利用料金あるいは開発費用という形で発生する。

・マイグレーション

デジタル情報のマイグレーション技術を用いたデータ保存は複数のタスクから構成される。

媒体変更：

- ・新媒体への電子情報の変換。この場合は記録（レコーディング）を行う。
- ・新媒体の管理。インデックス化、電子情報の保管、アクセス設備の整備が含まれる。
- ・新媒体の配布用コピー制作。保存の直接経費には含まれないが、必要な作業である。

後方互換性の検証：

- ・これは、アプリケーションのバージョンアップに伴うファイルアップグレード等をさす。例えば、バージョン X のアプリケーションで作成されていたファイルを、バージョン X+1 のアプリケーションのファイルにアップグレードするという場合には後方互換性の検証を行うこととなる。この場合、旧バージョンのファイル読み込みと新フォーマットでのファイル保存手順は、マクロプログラムを作成して自動的に行うことが可能で、このタスクでは、以下の作業コストが発生する。

- ・代表的な資料を用いた互換性のテスト
- ・作業プログラムのセットアップと実行
- ・新しいファイルのデータ検証
- ・旧バージョンのファイルの削除（これは必要に応じて行う）

標準形式への変換：

このタスクは標準フォーマットへの変換である。作業的には互換性検証と同様であるが、標準フォーマットに何を採用するか決定が必要になる。初期テストが完了した後は、変換作業は自動化可能である。例として、ポストスクリプトファイルから PDF ファイルへの変換等が含まれる。このタスクの作業コストは以下のとおりである。

- ・採用する標準フォーマットへの合意形成
- ・特定の資料における互換性のテスト
- ・バッチプロセスとして変換を実行
- ・変換後の資料サンプルのテスト
- ・旧バージョンのファイルの削除（これは必要に応じて行う）

・新ファイルのコピー

7. 権利関係の交渉コスト

電子情報へのアクセスや使用に関して、納本者の知的所有権、データ保護や機密性等の法的問題を交渉する領域もコスト要因となる。

資料に付与された権利は、単に資料へのアクセス方法だけでなく、第三者機関によって資料を保存することが法的に認められるかということも規定することとなる。この権利関係交渉コストは、電子情報の保存コストのうち最も金額の高い領域となるだろう。

6.3.4.保存コストに関するマネジメント:『電子情報の保存マネジメント ハンドブック』から

英国の電子情報保存の啓蒙団体である Digital Preservation Coalition では、電子情報の保存に当たって、マネジメントという観点からハンドブックを作成しているが、ここにおいて保存コストをどのように取り上げているかを見てみることにする。

電子情報の保存マネジメント ハンドブック(Preservation Management of Digital Materials; A Handbook) では、保存コストの課題として、アクセスの問題、組織運営や関係機関との協力について取り上げている。

このハンドブックでは主として、その資料がどのように作成されたかにかかわらず、重要な電子情報へのアクセスを長期間にわたって保存するコストを検討している。

アクセスコストの重要性

電子情報の保存とは長期にわたってアクセスを確保することである。従って、アクセス保存に密接に関連する保存コストを、アクセスコストと切り離して考えることは不可能だ。技術開発が進み、ユーザーの要求するサービスが多様になるにつれて、アクセスコストも増大する。

電子情報の作成・収集・アクセスを可能とするため、保存機関におけるインフラ構築の初期投資は重要である。しかし、その重要性があまり認識されていないのが現状であるため、長期保存計画の着手段階から、この初期投資を確保する手段を講じておくべきといえる。

印刷資料よりも電子情報の保存の方がコストかかる理由は、以下の四つの要因に帰することができる。

- ・ 技術変化に対応し続けることが避けられずその変化は潜在的に無期限であること
- ・ 資料に関して、出版者やデータ制作者との許諾契約を結ぶ際、規格統一や標準が

ないこと。このため情報保存にスケールメリットを活かすのが困難になる。

- ・ 特定の電子情報については、技術変化で本質的な情報が失われないという信頼性と精度の高さを併せ持つ再現手段が開発されていない状況であること。
- ・ コスト削減のための明確な方法を確立しない限り、新たに発生する保存コストはこれまで計上された保存コストに上積みされるのみとなってしまう。

納本図書館のような電子情報と従来の資料の双方に責任を担う機関は、全く異なる二種類の資料について、保存要件を公平に満たしながらバランスをとるという、非常に困難な課題に対処しなければならない。

このような機関においては、短期的にアクセス確保を行うというよりも、長期保存に優先度を持たせる可能性が高い。短期及び長期の保存の両方に共通するコストモデルがあり、なおかつそれが長期保存にとって最も費用対効果が高いというのが理想的であろう。

組織的な対応

アナログ資料を電子化したものであるか、もともと電子的に作成されたものであるかにかかわらず、電子化コストと電子情報の管理コストを区別することが重要である。この意味で、電子情報の作成コストは、最終的には長期保存に役立つであろうコスト要素に含める必要がある。

専門的技術やノウハウ、長期保存に適切なスキルを持つスタッフを採用し、人材開発を行うことは、技術変化のスピードが速く、求められるスキルの範囲が広いことから困難なものとなっている。

また、技術コストと設備やコストについてまだ十分に定義されていないため、保存を行う機関は、長期的コストについて明確なアイデアを持たないまま、電子情報という資産を保護する原則を明確化しなくてはならない。このため、将来計画がより冒険的となり、保存機関はより明確なコストの見通しを持たずに不本意ながら対策を講じる必要に迫られている。

費用対効果を考慮した保存戦略の検討は、保存機関内部だけでなく外部関係者にも良い経験となる。なぜなら、外部関係者は今後発生する作業の見通しを立てられるからである。また、費用対効果を意識していれば、保存機関はコストが生じる実務について十分に検討する。これは、保存機関が実務を理解する良い機会となり、外部関係者と実務知識の共有を図ることができる。保存という同じ試みに直面する人々が理解を深めることは、コミュニケーション不足によるリスクを軽減し、保存戦略を効果的に実現するための基礎となる。

6.4. 電子情報保存に係るコストの考え方について

紙媒体を保存する場合と異なり、電子情報保存のコストは予測が困難であるが、長期保存を行うシステム等についてコストが必要となるため、非常に高額となるのが現状である。特に、オランダ国立図書館のように出版社等の任意による納本を実施しているところにとって、コストは重要な問題となっている。しかしながら、保存コストの算定に関しては、電子情報の保存機関が採用する保存戦略に応じて異なってくるため、一概に類型化できない。

Cedars 等の海外プロジェクトにおいては、保存コストのフレームワークを策定する試みを行っている。具体的には、まずコストが想定される領域を抽出して、保存コストの要素を抽出する方法が一般的に採られている。

コストの考え方の一例として、英国の JISC/NPO が抽出した 7 領域がある。そこには、データ作成、データの選択と評価、データ管理、資料の公開、データの使用、データの保存、権利関係の調整が挙げられている。また、保存コストの要素に関しては、資料自体のコントロール度、データ検証、採用技術等が挙げられている。

いずれの領域、要素においても、電子情報の保存におけるシステム関連費用の重要性は、広く認識されている。オランダを始めとして、データ管理という観点から、エミュレーション、マイグレーション等保存戦略によるコストの変動についての議論が行われている。

日本国内において長期保存を行っていく場合、特に押さえておくべき領域は、データ作成及び権利関係の調整といえるだろう。諸外国における長期保存プロジェクトでも、まず長期保存に取り組むに当たって、この二点について、特に事前の対策が必要と認識されている。

データ作成は、本来であれば保存機関が関与するプロセスではないが、ここに対するコミットの度合いが将来的な保存に係るコストを大幅に削減する。つまり、保存機関が初期段階からコミットすることにより、採用技術等の保存戦略に添ったデータ品質を高め、データ管理を容易にすることができる。これは、既に述べたシステム関連費用に影響する分野である。

また、権利関係の調整も同様に事前の対策が必要となる。権利関係のコストとして一般的に捉えられている著作権費用以外にもコストは発生する。例えば、アクセス権の制限内容によっては、保存機関に費用負担がかかる場合もある。また、著作権等に関する議論は、専門家による交渉が必要なケースも多く人件費に配慮する必要もある。

このように保存コストについての考え方として、長期保存に直接関わる技術/システム費用はもとより、将来的にコスト要因となる領域からまずコスト認識を行い、長期的に「コスト・イフェクティブ」な環境を作り上げることが重要である。

7. 記憶装置の役割

本章では、コンピュータにおける記憶装置の位置付けを整理するとともに、電子情報保存を考えるに当たって必要となる保存媒体及び保存環境について整理する。

7.1. コンピュータの構成と記憶装置の位置付け

コンピュータは、処理装置、入力装置、出力装置、主記憶装置、外部記憶装置の 5 つを基本的な構成要素としている。これらの関係を図示すると以下ようになる。

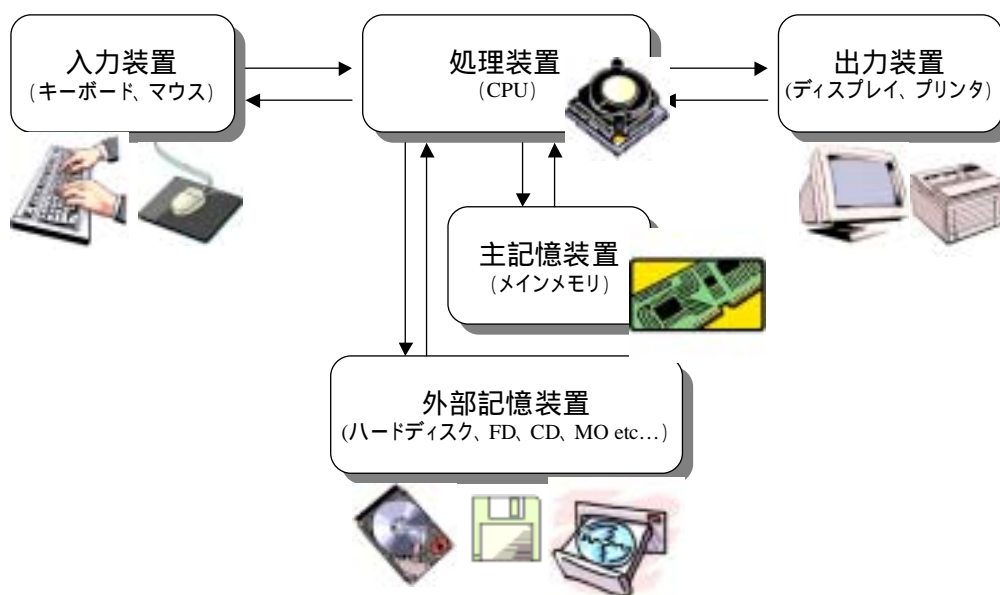


図 7-1. コンピュータの構成要素の整理

図 7-1 に示したように、コンピュータの記憶装置には主記憶装置と外部記憶装置がある。主記憶装置は一般にメインメモリと呼ばれ、処理装置である CPU に読み込まれるデータやプログラムを格納する領域であり、非常に高速なアクセスが求められる。一方、外部記憶装置は、ハードディスクを始め、FD、CD、MO 等、様々な形態・記憶容量のものがあり、メインメモリと比べてアクセス速度は低速であるものの、記憶容量は非常に大きいという特性を持っている。

7.2. 記憶装置の種別と長期保存に必要な記憶装置

前節で述べたように、コンピュータの記憶装置は主記憶装置と外部記憶装置の二種類に分けられるが、その役割や性能は異なっており、一般的には性能・価格を軸にとって配列すると階層構造を成すと考えられている。以下に、記憶装置の階層構造について整理する。

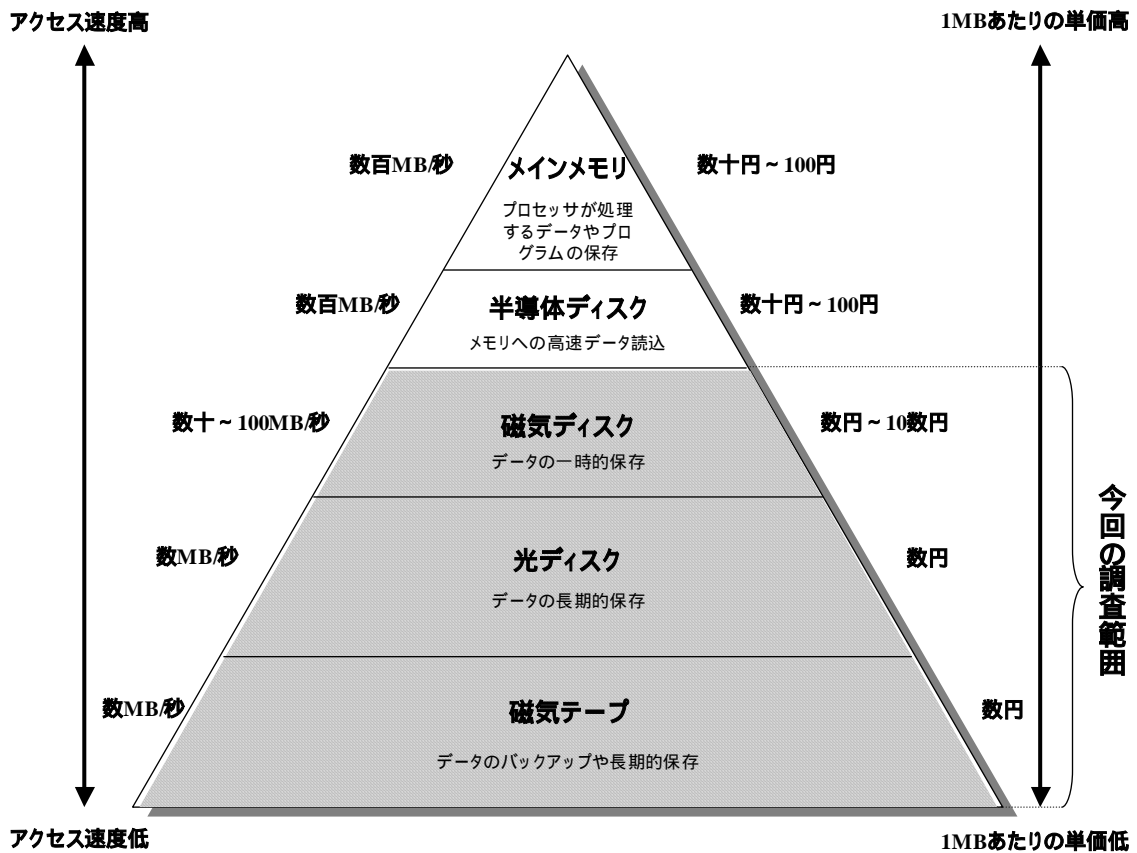


図 7-2. 記憶装置の階層構造

図 7-1. で述べた主記憶装置となるメインメモリは最もアクセス速度が速く、単価が高いメモリを使用している。一般的には、上図の階層構造の下部に向かうにつれ、外部記憶装置としての利用に適した記憶装置となる。また、昨今では、磁気テープのアクセス速度の高速化にともない、一部の機器においては光ディスクよりも 1 MB あたりの単価が低く、高速なアクセス速度を実現している製品も登場しており、ピラミッド内部で逆転が起きている状況も見受けられる。同様に、磁気ディスクにおいても、技術の高度化により、1MB あたりの単価が低廉化する傾向が見られている。今後も、技術進歩にともなって、アクセス速度の高速化、記憶媒体の低廉化が進むものと予想される。

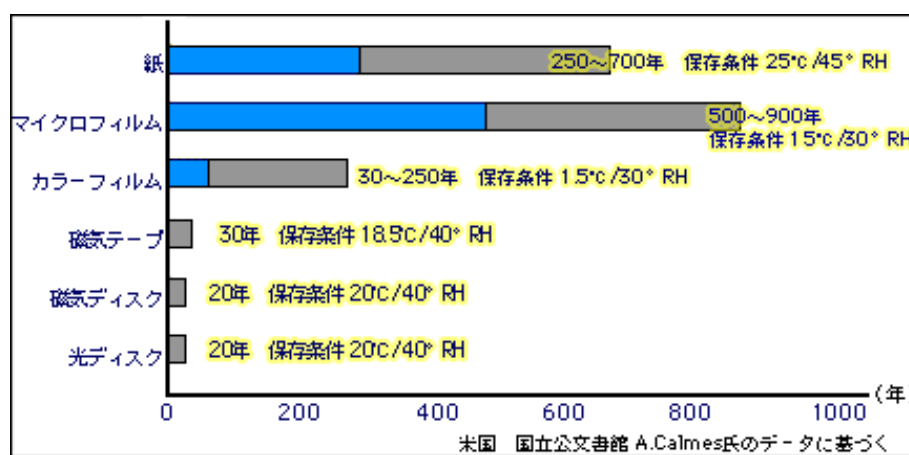
電子情報の長期保存を行うための媒体としては、メインメモリに比べて比較的低速で大容量の光ディスク、磁気テープが適切である。また、今後、価格の低廉化と技術の高度化によって、磁気ディスクを長期保存のための補完的な保存デバイスとして利用できる可能性もある。

7.3. 保存媒体について

本節では、各種媒体のライフレンジに関して整理を行うとともに、光ディスク、磁気テープ、ハードディスクを活用したストレージ技術の現状と今後の動向について見ていくこととする。

7.3.1. 保存媒体の種別によるライフレンジ

ここでは先に述べた光ディスク、磁気テープ、磁気ディスクを活用したハードディスクのライフレンジについて、長期保存には最適とされるマイクロフィルム及び紙媒体と比較しつつ見ていく。



出典：株式会社極東マイクロの Web サイト (<http://www.kyokuto-micro.co.jp/media.html>)

図 7-3. 各媒体のライフレンジ予測⁶⁸

図からもわかるとおり、媒体のライフレンジ（一定の保存条件下でメーカーが保障する期待寿命）は、マイクロフィルムが圧倒的な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ライフレンジは20年から30年にとどまっている。これら、各媒体のライフレンジは後述するとおり、保存環境や使用状況等の外的な要因によって異なる。

デジタル媒体については、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高いといえる。

また、上図のライフレンジについては、保存媒体を発売するメーカーによって、保証するライフレンジが異なっていることも多く、例えば、マイクロフィルムであっても媒

⁶⁸ グラフの色が変化している部分は、各種情報源によるライフレンジの異なりをあらわしており、濃い色は一般的に言われている寿命であり、薄い部分はメーカーや研究者等の間で見解に相違があることを示している

体の寿命は 100 年程度とする規格が存在する⁶⁹。

7.3.2. 保存媒体の種別と代替技術の動向

ここでは、光ディスク、磁気テープ、磁気ディスクを活用したストレージ技術、マイクロフィルムを活用したデジタル情報の保存技術の動向について見ていく。

光ディスクの技術動向

光ディスクについては、CD、MO⁷⁰等も技術向上を見せているが、ここでは、最近、進歩が著しい DVD を中心に扱う。DVD 系の媒体は CD と同様に、再生専用の「ROM」、一度だけデータを書き込める「R」、データの書込み・消去が何度でも可能な「RW」規格がある。記憶方式と記憶容量は規格に関わらず、下記の通りである。

記憶方式	片面 1 層	片面 2 層	両面 1 層	両面 2 層
記憶容量	4.7GB	8.5GB	9.4GB	17GB

表 7-1. DVD の記憶方式と記憶容量(12cm ディスクの場合)

ただし、DVD 系の媒体は、DVD フォーラム⁷¹と DVD + RW Alliance⁷²という二つの勢力に参加する各企業によって規格対立が激しく、今後存続していく方式を見極めるために、継続的な動向把握が必要になる。以下に現行で普及し始めている DVD の各規格について整理する

規格名称	特徴	備考
DVD-ROM	読み出し専用の DVD で、CD という CD-ROM に当たる。コンピュータ用ソフトウェアの配布や、映画等映像作品の配布に利用される	この規格は DVD-Video/Audio の親規格に当たる
DVD-R	一度だけ書きこみが行える追記型 DVD。DVD-ROM や DVD-RAM 等他の DVD 規格とも互換性がある	DVD フォーラムが提唱している規格
DVD-RW	書き換え可能型 DVD の規格。映像記録用途にしたもの	DVD フォーラムによって規格化された。下記で整理している DVD + RW と互換性はない

⁶⁹ 図 3 の紙媒体のライフレンジについても、和紙や酸性質が多量に含まれる用紙まで、多種多様の紙が存在し、酸性質の多い紙については、数十年で劣化することもある

⁷⁰ 読みは「エム・オー」。Magnetic Optical Disk の略。磁気記憶方式に光学技術を併用した書き換え可能な記憶装置。書き込み時はあらかじめレーザー光を照射してからデータを磁気的に書き込むので、記憶の高密度化が可能。読み出し時はレーザー光のみを用いるため、高速にデータを読み出すことができる。容量は 128MB、230MB、540MB、640MB のものが一般的だが、ASMO (アスモ：Advanced Storage Magneto Optical、5GB 以上の記録が可能) や GIGAMO (ギガモ：GIGA byte Magneto Optical、1.3GB の記憶容量を持つ) 等 1GB を超える大容量をサポートしている規格もある

⁷¹ DVD の規格制定や、普及に向けた広報活動を行う業界団体。ハードウェア、ソフトウェア、マルチメディアあるいは放送の各方面から、日本国内で 80 社以上、全世界で 200 社以上が加盟している。規格制定団体としては珍しく日本企業が主導権を握る(初代議長は東芝)団体で、事務局は東京に置かれている

⁷² DVD+RW の推進を図る団体。参加企業はソニー、リコーの他、蘭 Philips、仏 Thomson Multimedia、米 Hewlett Packard、Dell Computer、Verbatim 等で構成される

DVD-RAM	読み書き・消去のいずれも可能な DVD で、コンピュータ用の大容量記憶装置として期待されている	DVD フォーラムによって規格化された。DVD-RAM はコンピュータデータの記録に用いられる
DVD+RW	読みは「ディーバイディープラスアールダブリュ」。書き換え可能な DVD の規格。DVD-ROM と DVD-RAM との再生互換性が低いことが策定理由のひとつ。DVD-RAM や DVD-RW と比べて DVD-ROM と仕様上の互換性が高く、多くの DVD-ROM ドライブで読み出せることがメリットとされている	ソニー、Philips Electronics 社、Hewlett Packard 社の 3 社が策定した規格。実際の製品ベースではリコーが強く推進している。DVD-RW とは互換性がない
DVD+R	読みは「ディーバイディープラスアール」。追記型 DVD の規格。2001 年に規格は制定されていたが、市場導入は 2002 年 4 月から。DVD フォーラムが提唱している DVD-R と比較して DVD-ROM に近い形での記録を行うため、規格レベルでは DVD-R より高い再生互換性を持つとされている。しかし実際は DVD-R より後発だったため、ほとんどの市販 DVD プレーヤーが DVD-R に対応済みであり、実際の再生互換性は DVD-R より劣る	DVD+RW Alliance が提唱している規格。DVD+RW より後発

表 7-2. DVD の規格の整理

表からもわかるように書き換え可能型で DVD-RW と DVD+RW が、追記型で DVD-R と DVD+R が競合しているのがわかる。このような規格対立の中で、昨今では DVD フォーラムによって DVD Multi (DVD マルチ) 規格が定められ、これにより、DVD-ROM/-R/-RW/-RAM のすべての規格について再生・記録が可能なドライブが出荷されるに至っている。また、DVD の流れをくむ次世代光ディスク規格においても同様の規格対立が起こる模様である。次世代光ディスクは、波長の短い青紫色レーザーを活用し、現行世代の DVD よりも、更なる記憶量の大容量化が図られている。以下に、現行世代の DVD-ROM と比較しつつ、AOD (Advanced Optical Disk) 及び Blu-Ray Disk といった次世代光ディスク規格の概要についてまとめる。

規格名称		現行 DVD-ROM	Advanced Optical Disk (AOD)	Blu-Ray Disk
記憶容量	再生専用	両面 1 層 : 9.4GB 両面 2 層 : 17GB	片面 1 層 : 15GB 片面 2 層 : 30GB	なし
	書き換え	なし	片面 1 層 : 20GB 片面 2 層 : 40GB	23.3/25/27GB
規格策定企業		DVD フォーラム	DVD フォーラム	ソニー、松下、日立等 9 社
データ転送レート		1.4MB/秒	4.5MB/秒	4.5MB/秒

表 7-3. 現行 DVD-ROM と次世代光ディスクの概要

次世代光ディスクである AOD と Blu-Ray Disk の間には互換性がなく、今後の規格対立は避けられない状況との見方が業界では一般的である。また、Blu-Ray Disk については、現在、書き換え用の規格しか整備されておらず、今後、再生専用等の規格が登場するものと思われる。

次世代光ディスクの次を行く、次々世代の技術として研究がなされている技術としては「ホログラフィック記録再生」が挙げられる。この技術は、一枚で数百 GB の記録容量を持ち、12.5MB/秒以上のデータ転送速度を実現するといわれている。現在は研究段階であるが、一部には 2005 年ごろに民生機器を発売するという計画を持つ企業も現れている。しかし、ホログラフィック記憶・再生技術は 50 年近くも研究が行われており、昨今の材料関係技術の発展により、ようやく実現に向けて足がかりが得られたといわれているのが現状である。今後、実用化には更なる時間が必要になる。

磁気テープの技術動向

磁気テープは、記憶単位当たりのコストが低く、情報システムの世界では、データのバックアップや長期保存に従来から利用されてきた媒体である。近年の IT 技術の急速な進歩により、技術革新が進んだことと、情報システムで扱われるデータが大容量化していることを背景に、磁気テープ媒体においても、著しい大容量化とデータ読み出し速度の高速化を実現している。磁気テープの方式には、「ヘリカルスキャン方式」と「リニアスキャン方式」の二種類があり、前者はテープがヘッドに巻きついて読込を行う方式であり、後者は、テープがヘッドに巻きつかずに通過し、テープ表面はヘッド以外と全く接触しないで読み出しが可能である。

以下に、両方式の概要と主な対応テープの規格をまとめる。

方式の名称	特 徴	対応テープ規格
ヘリカルスキャン方式	<ul style="list-style-type: none"> リニアスキャンに比べてヘッドのテープ上での移動が遅いため走行が安定する 個々のデータセットの取り出しを行えるマルチポジションローディングに対応 	<ul style="list-style-type: none"> AIT DDS
リニアスキャン方式	<ul style="list-style-type: none"> テープヘッド、テープ自体の消耗や損傷、汚れが少ない 	<ul style="list-style-type: none"> DLT LTO

表 7-4. 「ヘリカルスキャン方式」と「リニアスキャン方式」の特徴と対応規格

次に、上図で述べた対応テープ規格に関する容量とデータ転送速度についてまとめる。各規格の末尾に付く番号は世代番号であり、数が大きいものほど新しい規格である。

方式	規格名称	非圧縮時最大容量	最大データ転送速度
ヘリカルスキャン	AIT1	25GB	3MB/秒
	AIT2	50GB	6MB/秒
	DDS2	4GB	0.36～0.75MB/秒
	DDS3	12GB	0.72～1.5MB/秒
	DDS4	20GB	1～3MB/秒
リニアスキャン	DLT1	40GB	3MB/秒
	DLT8000	40GB	6MB/秒

表 7-5. 各規格のデータ記憶容量と転送速度

今後も情報技術の進展にあわせて、磁気テープの記憶容量とデータ転送速度は増加していくであろう。以下に、主な規格を使った磁気テープの今後の大容量化の計画についてまとめる。ヘリカルスキャン方式では DDS4 以降の大容量化は行われなため、AITのみを取り上げ、リニアスキャン方式では DLT の後継規格である SDLT、及び LTO 方式を活用した Ultrium を取り上げることとする。

規格名称	2003 年	2004 年	2005 年	2006 年	2007 年
AIT	200GB 24MB/秒		400GB 48MB/秒		800GB 96MB/秒
LTO (Ultrium)		400GB 80MB/秒		800GB 160MB/秒	
SDLT	320GB 32MB/秒		640GB 50MB/秒以上	1.2TB 100MB/秒以上	

表 7-6. 各規格の大容量化の計画

上表からもわかるとおり、今後、磁気テープは 2 年で容量・アクセス速度を倍増させる計画で各メーカーが開発を進めている。

磁気ディスクを活用したストレージ技術

磁気ディスクを活用したハードディスクについては、技術発展により、信頼性と記憶単位あたりの単価が低下しており、長期保存のための各種媒体を補完する役割として活用することができる。

大規模情報システムでは、ハードディスクを活用したストレージ・システムが急速な進歩を遂げている。技術トレンドは、従来のようにサーバーに直接に接続する Direct Attached Storage(DAS)の形式から、ストレージ機器自体をサーバーから切り離して、独自のネットワークを構築するネットワーク・ストレージが主流になりつつある。

このようなネットワーク・ストレージへの変化の過程で注目を集めている技術が SAN (Storage Area Network)⁷³である。SAN は 1997 年ごろより、提唱され始めた技術であり、「ディスク装置やテープ装置等のストレージ・デバイスを対象に LAN とは独立したネットワークを構成する技術」と定義される。LAN がサーバーやクライアントといったコンピュータ同士のネットワークである一方で、SAN はストレージ機器のみで構成されるネットワークである。以下に、DAS ストレージを活用したネットワークと SAN を活用したネットワークの図を併置する。

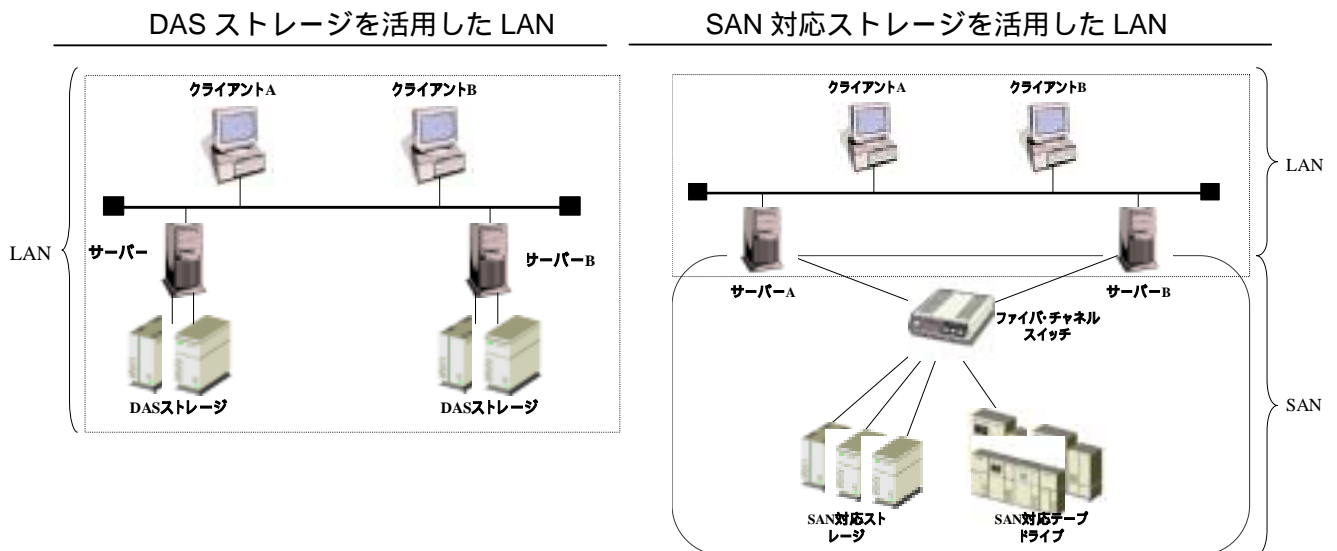


図 7-4. DAS ストレージの LAN と SAN (Storage Area Network) の比較イメージ

図示したとおり、SAN は LAN と別に、ストレージ機器のみのネットワークを構成する。ストレージとサーバーはデータ転送速度を高めるためにファイバー・チャネルで接続が行われ、ファイバー・チャネルスイッチを介してストレージ機器同士を接続する⁷⁴。

⁷³ 一般的にネットワーク・ストレージとして普及しつつあるものには、SAN の他に NAS (Network Attached Storage : ナス) がある。NAS は、プロセッサを搭載したファイルサーバーであり、非常に簡易な設定で LAN にストレージを追加し、ファイルサーバーとして活用できるため、近年、各種情報システムにおいて導入が進みつつある

⁷⁴ SAN 対応のストレージ機器にはファイバチャネルインターフェイスが必要になる

ファイバー・チャンネルはサーバー間、あるいはサーバーと周辺機器を接続するための技術規格であり、最大データ転送量は 2Gbit/秒という高速性能を持っている。従来の LAN で活用されるイーサネット規格では 100M~1G bit/秒、通常のストレージの接続に利用されるインターフェイスである SCSI は数百 M bit/秒であることから比較すると、かなり高速であるといえる。ただし、ファイバー・チャンネルケーブルの最大延長は 10km までであり、ビルの構内等での利用に不便はないが、遠隔地間の広域にわたる SAN を構築できないことが課題である。また、ファイバー・チャンネルをつなぐファイバー・チャンネルスイッチは安いものでも 1 台で数百万円と高価な上、SAN の構築・運用に特定のノウハウが必要になるという課題も指摘されている。

ここまで説明したファイバー・チャンネルを活用した SAN は FC-SAN と呼ぶが、上で述べた課題を解決するための代替技術として、最近では、IP ネットワークを活用して SAN を構築するための技術が登場している。このような SAN の規格は、ファイバー・チャンネルを活用した FC-SAN に対して、IP-SAN と呼ばれている。現在、IP-SAN と呼ばれる技術には「iSCSI」、「FCIP」、「iFCP」といった規格が存在している。この中で最も実用化に近いといわれているのは、「iSCSI」である。

「iSCSI」は米 IBM と米シスコシステムズによって開発された技術規格で、パソコンの周辺装置を接続する「SCSI⁷⁵」と呼ばれるインターフェイスのデータをインターネット/イントラネットで活用される通信プロトコル TCP/IP⁷⁶を使って通信するものである。「iSCSI」を活用することによって、通常のコンピュータでは標準的なインターフェイスとなっている SCSI からデータを読み出し、インターネット/イントラネット上で利用される標準的な通信プロトコルである TCP/IP を使って、距離を気にすることなく、ネットワーク上で通信が可能になる。

「iSCSI」は、SCSI や TCP/IP といった標準技術を活用しているため、FC-SAN に比べ、特別なノウハウは必要なく、既に安価に市販されている技術を用いて高速なストレージネットワークを構築できるというメリットがある。FC-SAN の課題点を解決する代替技術として非常に有望だが、「iSCSI」では、現時点で FC-SAN に比べ通信速度が劣るという欠点がある。しかし、今後のギガビットイーサネットの発展とともにこの問題も解決されるものと予想される。いずれの技術も、ネットワーク・ストレージに通常の IP ネットワークの技術を適用するものであり、業界ではファイバー・チャンネルを使った FC-SAN から IP ベースの IP-SAN に向かうといった考え方が主流となっている。

⁷⁵ 読みは「スカジー」。Small Computer System Interface の略。パソコン本体と周辺機器の接続方法の取り決め。米国規格協会(ANSI)によって規格化されている。現在では汎用性や性能が大幅に強化された後継規格、SCSI-2 や SCSI-3 が普及している

⁷⁶ インターネットやイントラネットで標準的に使われるプロトコル。米国防総省が、核攻撃で部分的に破壊されても全体が停止することのないコンピュータネットワークを開発する過程で生まれた。UNIX に標準で実装されたため急速に普及し、現在世界で最も普及している。OSI 参照モデルでは IP が第 3 層(ネットワーク層)、TCP が第 4 層(トランスポート層)にあたり、HTTP や FTP 等の基盤となるプロトコルである

マイクロフィルムを活用したデジタル情報の保存技術

先に述べた各媒体のライフレンジからも明らかなように、長期保存を行う際の媒体としてはマイクロフィルムが突出した媒体寿命を持っている。このような特徴を活かして、デジタル情報もマイクロフィルムで保存する動きが出ている。

具体的な製品は、コダックや富士フィルムといったフィルムメーカーにより、ドキュメントソリューションの一環として確立している。以下に、デジタル情報をマイクロフィルム化して保存する技術の概要について図示する。

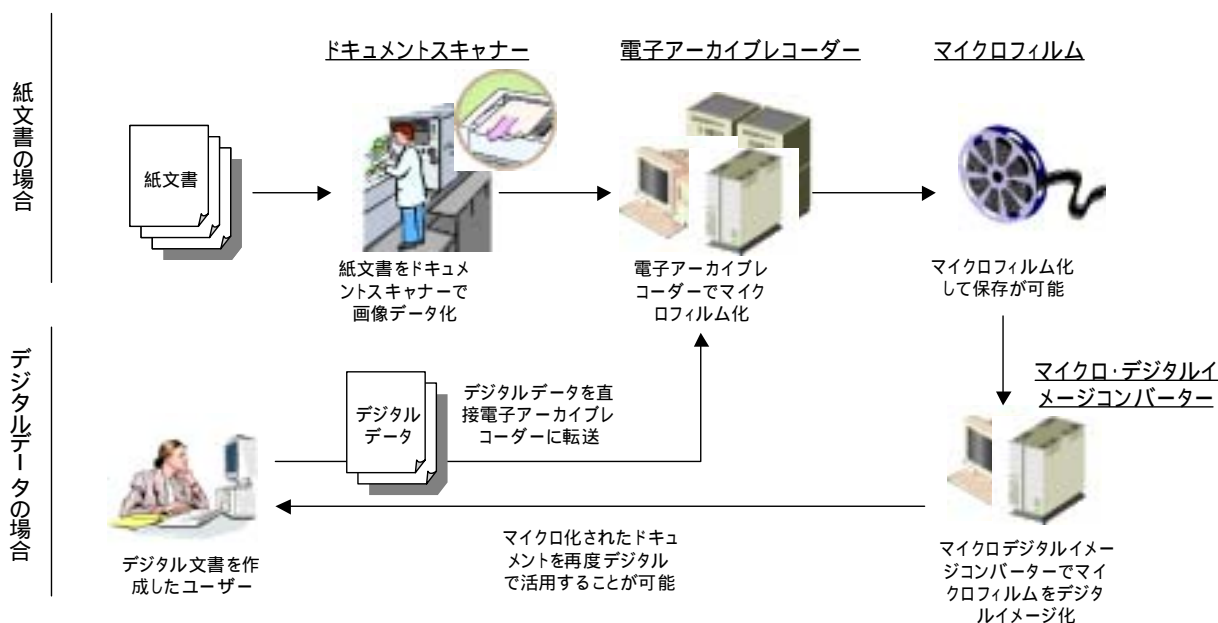


図 7-5. マイクロフィルム化されたドキュメントをデジタルで活用する

上の図で示した通り、紙文書はドキュメントスキャナーで電子化し、元々デジタルで作成される文書については直接に、電子アーカイブレコーダという機器を使って、ドキュメントのマイクロフィルム化を簡単に行うことができるというものである。一方で、マイクロフィルム化されたドキュメントは、マイクロ・デジタルイメージコンバーターを使って、再度デジタルデータとして利用することができる。

マイクロフィルム化されたデータは、そのまま長期保存が可能である。また、マイクロフィルム自体は光に透かすことで内容がある程度判別可能であるため、デジタルデータとは違い、再生装置がなくとも見ることができる。デジタルデータと比べて、再生装置の陳腐化によるリスクを低減しうる。

マイクロフィルム化することでデータの長期保存が可能となることからして、静的な電子出版物の保存には非常に有効なソリューションだが、音声や動画を収録した動的なマルチメディアコンテンツ等については、いわゆる「見た目と感覚 ("Look and Feel")」を保持するために更なる技術検証が必要である。

7.4. 保存環境について

本節では、保存媒体が環境から受ける影響を考えるとともに、データのバックアップや保管庫の施設の面について見ていく。

7.4.1. 保存に最適な環境とは

電子媒体は保存される環境によって、そのライフレンジが著しく異なる。以下に、保存環境別に見た各媒体のライフレンジについて整理する。

媒体	10 湿度 25%	15 湿度 30%	20 湿度 40%	25 湿度 50%	28 湿度 50%
カートリッジ 式磁気テープ	75 年	40 年	15 年	3 年	1 年
DVD	75 年	40 年	20 年	10 年	2 年
CD-ROM	30 年	15 年	3 年	9 ヶ月	3 ヶ月

出典：Preservation Management of Digital Materials A Handbook

表 7-7. 環境が与える各媒体のライフレンジへの影響予測⁷⁷

上記のライフレンジは、汚染物質・紫外線・強力な磁場による影響を排除した上で、気温と湿度を変化させて実験した場合の値である。環境要件が変化する場合、あるいはこの実験で使用した以外のメーカーの媒体を使用した場合には耐用年数が異なる可能性があるが、上表からも明らかなように、電子情報を保存する媒体には、安定した低温・低湿の環境が適している。

このような温度・湿度条件を考慮した上で、本表の出所となっている”Preservation Management of Digital Materials A: Handbook”では、保存媒体の取り扱い上の注意点として下記の事項を推薦している。

(保存媒体に関する注意事項)

- ・ 媒体をどのように扱うか管理手順を設定し、文書化しておくこと
- ・ 媒体を直射日光にあてないこと
- ・ 保存対象の媒体には通常のケースのほかに付加的な補強を施しておくこと
- ・ 磁気媒体は、強力な磁場を持つ場所には保存しないこと
- ・ 磁気媒体は、50mm 以上の間隔を置いた上で、ケースに入れて輸送すること
- ・ 光ディスクにはラベルを貼ったり、ペンで文字を書いたりしないこと
- ・ 保存には可能な限り高品質な媒体を使用すること

上記で述べた条件以外にも、電子情報の長期保存を考える場合は、バックアップを保

⁷⁷ 媒体のライフレンジについては、各メーカー及び各研究者で見解に相違があるため、7.3.1.の図 7-3.各媒体のライフレンジ予測で掲載した数値と異なっている。また、本表は JISC、英国図書館等が出版した Preservation Management of Digital Materials A Handbook からの引用であるが、表は米国 National Media Laboratory、英国 Public Record Office 等の研究を参考に作成されている

持することや保管庫の施設に対する考え方も重要になる。電子情報保存の問題を体系的に扱った同書では、具体的に考慮すべき点として、以下の項目が挙げられている。

(バックアップに関する考え方)

- ・ それぞれに異なるメーカーから媒体を購入し、コピーを作成しておくこと(ただし、できるだけ標準的な媒体を採用すること)
- ・ 異なる媒体・ソフトウェア環境でコピーを作成しておくこと
- ・ 決められた期間で媒体の移し換え等を実施すること
- ・ 保存対象の情報がコピーされた場合は、コピーに関する情報を文書化すること
- ・ バックアップ時に万が一情報が失われた場合に備えて、オリジナルフォーマットの情報を常に保存しておくこと
- ・ 定期的に媒体の可読性を確認すること(自動化されたストレージ・システムであれば、設定次第で定期的にチェックを行うことが可能である)
- ・ チェックサムを活用して保存対象のデータの一貫性を確認すること
- ・ 情報の真正性を担保するための仕組みを採用すること
- ・ バックアップデータは災害時に備えて、現行の保管庫から距離的に離れた場所にも保管しておくこと

(保管庫の施設等に関する考え方)

- ・ 保管庫周辺は禁煙とし塵や埃、その他汚染物質を排除し、常時環境を監視すること
- ・ 保管庫を設ける施設は火災や洪水といった自然災害を受けない場所に建設する、あるいは自然災害を防ぐ対策を施すこと
- ・ 災害時復旧手順を文書化し、スタッフを訓練しておくこと
- ・ 保管庫への入退室の権限の管理、情報へのアクセス管理を実施すること

デジタルデータの長期保存については、上記の事柄を考慮しつつ各種の手順を定めたガイドラインが必要になる。

7.5. 長期保存に係る保存媒体、保存環境に関する事例

本節ではここまで説明してきた内容を踏まえ、保存媒体に関する事例を見ていくこととする。始めに、媒体の取り扱いに関する文書化された方針の事例として米国議会図書館を取り上げる。次に、SAN 等のネットワーク・ストレージを活用した電子情報保存システムとしてオランダ国立図書館を取り上げる。最後に、バックアップの考え方や障害時復旧手順を定めた英国のデータアーカイブの事例を取り上げる。

7.5.1. 米国議会図書館の媒体の取り扱いに関する方針

米国議会図書館においては、シリンダ、ディスク、テープに記録された情報を取り扱う際の規則(“Cylinder, Disc and Tape Care in a Nutshell”)が定められている。この規則は、「取り扱い(Handling)」、「保管(Storage)」、「クリーニング(Cleaning)」、「再生装置のメンテナンス(Playback Equipment Maintenance)」の4つから構成されてい

るが、「クリーニング」については、主に薬品等の説明なのでここでは除外する。

以下から、それぞれの規則について表形式でまとめていく。

取り扱いに関する事項	
一般的事項	<ul style="list-style-type: none"> 各媒体の録音面には直接触れないこと 媒体を扱う前には手を洗うこと
ディスクに関する事項	<ul style="list-style-type: none"> レコードは端からラベルを貼った部分のみに触れて扱うこと コンパクトディスクは端と中央の穴のみに触れて扱うこと
オープンリールテープに関する事項	<ul style="list-style-type: none"> リールの端と中央部分のみに触れて扱うこと テープが傷むのでリールの端を圧迫しないこと
カートリッジ式テープに関する事項	<ul style="list-style-type: none"> 外側のケースのみに触れること 決して録音面には触れないこと
シリンダに関する事項	<ul style="list-style-type: none"> 中指と人差し指を中央の穴に入れ、シリンダが抜け落ちないように指を静かに広げて支えること シリンダの溝には決して触れないこと。特にワックスシリンダはカビに弱いので注意すること 手で触れた際にワックスが溶け出す恐れがあるので、ワックスシリンダは使用前に室温にしておくこと

表 7-8. 媒体の取り扱いに関する事項

保存に関する事項	
一般的事項	<ul style="list-style-type: none"> すべての媒体は端を上にして保管すること。決して平らに寝かせて保管してはならない
環境に関する事項	<ul style="list-style-type: none"> 保管場所はすべて清浄に保たれていること
中期的な保存に関する事項（最低でも 10 年を想定）	<ul style="list-style-type: none"> 保管場所は、24 時間常時、温度約 18 度～21 度、湿度 45%～50%に保たれていること 基準からの逸脱は温度約 ±5 度、湿度約 10%の範囲を超えてはならない 直射日光は避け、フロアライトもシールドする
長期的な保存に関する事項（永続的な価値をもつ対象を想定）	<ul style="list-style-type: none"> 保管場所は、24 時間常時、温度約 7 度～9 度（磁気テープについては破損を防ぐために必ず 7 度以下にすること）あるいはそれ以下の温度とし、湿度については磁気テープでは 20%～30%、それ以外の媒体には 45%～50%に保つこと 基準からの逸脱は温度で約 ±2.5 度、湿度で 5%の範囲を超えてはならない 室内は、入室するとき以外は暗室とし、紫外線を発生する光源は使用しないこと

磁気テープの消磁対策に関する事項	<ul style="list-style-type: none"> ・ 磁気の消失はそれほど多発しないが、万が一のために消磁の可能性のある場所は避けること（例えば、大音量のスピーカーの近く等） ・ 磁気や熱を発生する物体にたてかけて磁気テープを保管しないこと ・ 電気モータを使用する機械を磁気テープの保管庫の隣に設置しないこと
保管棚に関する事項	<ul style="list-style-type: none"> ・ 保存する媒体が大量になると相当な重量になるため棚が壊れないよう頑丈なつくりを採用すること
ディスクの保管に関する事項	<ul style="list-style-type: none"> ・ 垂直に立てて保管すること ・ ディスクを保管する棚はディスクをすべて覆い、各々を約 10～15cm 離して保管するのが望ましい
オープンリールテープに関する事項	<ul style="list-style-type: none"> ・ 箱は垂直に立てて保管し、倒れないようにブックエンド等で支えること
カートリッジ式テープに関する事項	<ul style="list-style-type: none"> ・ 垂直に立てて保管し、決して平らに寝かせたままにしないこと
シリンダに関する事項	<ul style="list-style-type: none"> ・ 垂直に立てて保管すること
テープの巻き戻しに関する事項	<ul style="list-style-type: none"> ・ 早送りや巻き戻しの途中状態で保管しないこと ・ テープを使用するときはすべて再生し、巻き戻ししないことが望ましい

表 7-9. 媒体の保管に関する事項

再生装置のメンテナンスに関する事項	
再生装置のメンテナンスに係る事項	<ul style="list-style-type: none"> ・ テープヘッドやガイドは念入りにクリーニングを行っておくこと ・ テープヘッドの消磁は定期的実施すること ・ 光ディスクのヘッドは定期的清掃すること ・ 最低月に一回は再生装置の設定を確認すること

表 7-10. 再生装置のメンテナンスに関する事項

7.5.2. オランダ国立図書館における電子情報保存システム

オランダ国立図書館における電子情報保存システム”DIAS (Digital Information Archiving System)”は、IBM と共同で開発されているシステムである。オランダ国立図書館と IBM は、システムの実装を進める一方で、LTP Study という電子情報の長期保存のための取り組みを実施しており、その成果が徐々にシステムに反映されつつある。以下に、DIAS システムにおける全体構成図を記す。

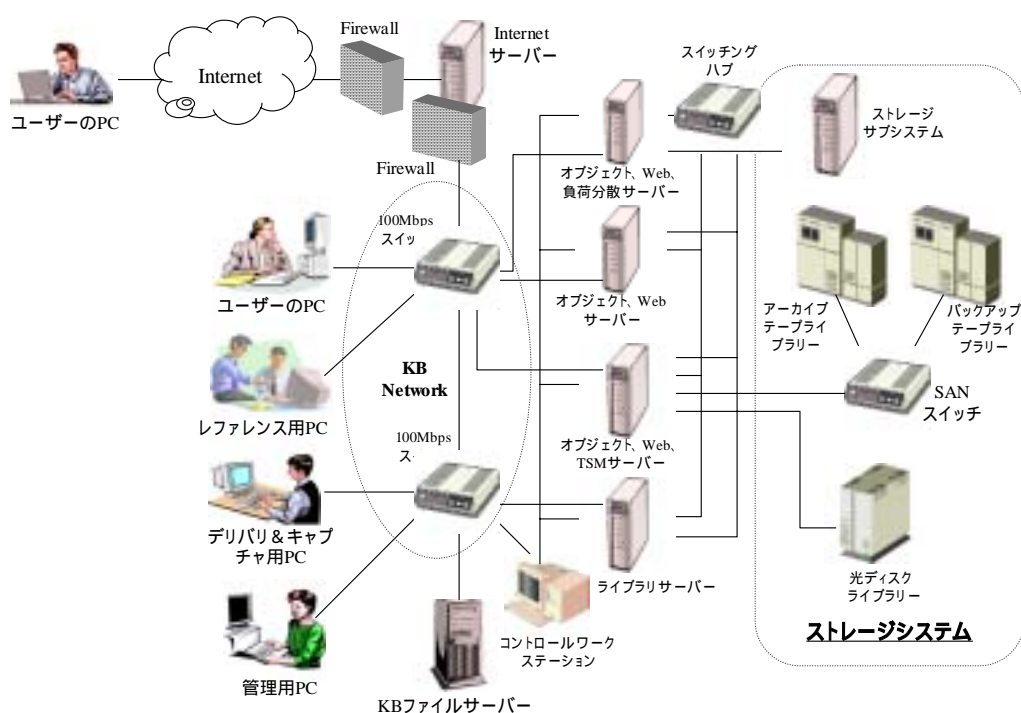


図 7-6. DIAS のシステム構成イメージ図

DIAS ではストレージ・デバイスに SAN を使ったテープライブラリーを使用し、アーカイブ用とバックアップ用に配置している。また、光ディスクライブラリー装置⁷⁸も利用されている。以下、DIAS で使われているストレージ・デバイスの性能についてまとめる。

テープドライブ : Tape Drive 3950E (IBM 社製)	
テープ容量	20/40GB
データ転送レート	14MB/秒
テープ挿入時間	25 秒
テープ取り出し時間	10 秒(推定)
テープ半分の読み出しの所要時間	1.5 分
テープ半分の巻き戻しの所要時間	20 秒(推定)
テープロボットの動作時間	7 秒

表 7-11. DIAS のテープドライブの性能

⁷⁸ 光ディスクライブラリーは大量の記憶用光ディスクを操作する装置である

光ディスクドライブ : Optical Library 3995 C68 (IBM 社製)	
ディスク容量	5.2GB
ドライブ搭載量	6 台
データ転送レート	4.6MB/秒
ディスク挿入時間	5.5 秒
ディスク取り出し時間	3 秒
カートリッジ移動時間	3 秒
平均探索時間	120 ミリ秒

表 7-12. DIAS の光ディスクドライブの性能

DIAS では上記の内、テープドライブを SAN でつなぎ、高速で信頼性の高いストレージネットワークを構築するとともに、データ管理ソフトウェアを利用し、適切なタイミングでバックアップを確保する等、堅牢な構成の電子情報保存システムになっている。

7.5.3. 英国のデータアーカイブにおける障害対策

英国のデータアーカイブは、社会科学分野の国立データセンターであり、Economic and Social Research Council と Joint Information Systems Committee (JISC) が出資して設立された機関である。このデータアーカイブは 125,000 のファイルからなる 4,000 以上のデータセットを保有している。

データアーカイブにおけるストレージ・システムは HSM (Hierarchical Storage Management System : 階層型ストレージ管理システム) を基礎としている。システムで保存されるデータの大半はテープデバイスに保管されているが、ディスクキャッシュ⁷⁹を活用し、アクセス頻度が高い情報への高速なアクセスを提供している。また、システムが自動的に判断し、テープから必要なデータをディスクキャッシュに呼び出し、情報へのすばやいアクセスを提供する機能も備えられている。

データアーカイブでは、バックアップの取得と障害復旧時の手順が厳密に定められている。まず、バックアップの取得に関する事項を整理する。データアーカイブでは、最低でも以下の 4 つのバックアップを保持するよう決められている。

バックアップ媒体	概要
メインコピー	HSM システム内に保存されているコピー
シャドウコピー	アップデートされたデータは必ずひとつは別のテープデバイスにコピーが作成される

⁷⁹ ハードディスク等アクセスが低速な記憶装置に記録されているデータの一部を高速な半導体メモリに写し、高速化する技術。また、そのために使われるメモリ上の領域。使用頻度の高いデータをメモリに置いておけば、データの読み出し要求があった時にいちいちハードディスクからデータを読み込む必要がないため、読み込み速度が飛躍的に向上する

CD コピー	保存手順の一環として CD にデータをコピーする。データは研究対象ごとにまとめられ Zip 形式で圧縮したファイルが保存される。既に保存している研究データがアップデートされた場合は、新たな世代のデータとして CD に追記される
オフサイト・ニアラインコピー	データアーカイブのあるエセックス地方に災害が発生した場合のためのコピーを作成する

表 7-13. 英国データアーカイブのバックアップ方法

英国データアーカイブでは上記のバックアップを行う一方で、様々なレベルで発生することが予想される障害時の復旧手順について、以下のように定めている。

障害のレベル	概要	対策
ファイルの破壊	ファイルの情報がデータ処理の際に判読不能となる場合を指す	<ul style="list-style-type: none"> 情報の提供者に再度提供を依頼する シャドウコピーを活用した復旧を行う
読み出し不能のファイル	媒体の故障でファイルの読み出しが不可能になる場合を指す	<ul style="list-style-type: none"> まず、障害がテープ全体に起こっているかを確認し、全体的な問題の場合は、復旧処理を行う テープ全体にわたる障害でない場合は、問題のあるファイルをシャドウコピーで置き換える
媒体の破壊	情報を保管している媒体自体が破壊される場合を指す	<ul style="list-style-type: none"> テープが全て記録済みで、読み出し専用であり、バックアップが利用できる場合は、バックアップから新たなテープを作成する バックアップが利用できない場合は、シャドウコピーを利用して新たなテープを作成する
メインコピー・シャドウコピーの破壊	メインコピー、シャドウコピーを格納している装置の故障及びバックアップの読み出しが不能となる場合を指す	<ul style="list-style-type: none"> 定期的な保守を行っているためほとんど起こらないトラブルであるが、発生した場合は、バックアップ CD からデータを復元する

データの完全な破壊	データセンターがあるエセックス大学にあるコンピュータ等が大規模災害等により破壊されてしまった場合を指す	<ul style="list-style-type: none"> ・ HSM システムを再構築し、ULCC (University of London Computer Center) に保管してあるオフサイトコピーを用いてデータを復旧する
-----------	---	---

表 7-14. 英国データアーカイブにおける障害復旧手順

上記のように大規模アーカイブでは、定期的なバックアップ規則及び体系的な障害復旧手順が必要になる。

7.6. 電子情報保存に当たっての保存媒体、保存環境の考え方

電子情報の保存に当たっては、適切な保存媒体を選択することが不可欠であるが、本文でも触れた通り、保存媒体における技術の進歩や記憶容量の増加が著しく、その動向を把握しておくことが重要である。

保存媒体の選択に当たっては、光ディスクや磁気テープが有用である。特に磁気テープについては、大容量化が進むとともに読み出し速度の高速化も図られており、大量のデータを保存する際には有効な媒体である。また、本文でも述べた SAN(Storage Area Network)等の技術については、DAS(Direct Attached Storage)に比べストレージ・デバイスを有効利用できることから機器の削減も可能である反面で、ストレージの増設も容易なため、時間とともにデータが増大していく図書館等における電子情報の保存には適しているといえる。また、SAN はストレージを複数使用するのでシステムの冗長化が可能であり、一部のストレージ・システムが利用不能になった場合にも、他のシステムで代替が可能である。従ってデータ保管の安全性の面から考えても有用である。

各種の保存媒体のライフレンジについては、本文でも触れた通り、メーカーや研究者によって認識に違いがある。諸外国においては、第 2 章で述べた通り、オーストラリア国立図書館等で CD-R のライフレンジに関する研究が行われているが、日本においてもメーカーや各種の研究機関の数値を参考に、自らも経験を積み重ね、保存媒体のライフレンジを見極めていくことも必要である。

保存媒体のライフレンジについては、本文でも指摘したとおり、保存環境から大きな影響を受ける。媒体の保存に当たっては適切にコントロールされた環境が必要であり、特に温度・湿度等に気を配った保存環境が求められる。また、万一、保存している電子情報が消失した場合に備えて、バックアップ等の危機回避手段を講じておくことが必要である。

媒体の保存は電子情報保存における活動の重要な一部であり、媒体の取り扱い、バックアップ、障害回避等の手順について文書化するとともに、環境を適切に管理し、保存を担当する職員の訓練を常に実施していく必要がある。

8. 情報発信者と図書館の協働作業

本章では、電子情報保存における出版社等の情報発信者と図書館の協働作業の必要性、その際の検討事項とアプローチに関する留意点を整理するとともに、協働作業の事例として CENL（Conference of European National Librarians：ヨーロッパ国立図書館員会議）と FEP（Federation of European Publishers：ヨーロッパ出版連盟）の任意納本による協定及び 2001 年に行われた Andrew W. Mellon 財団の資金助成による電子ジャーナルの保存プロジェクトを取り上げる。

8.1. 情報発信者と図書館の協働作業の必要性

ここでは、電子情報保存に当たって、情報発信者と図書館が協働作業を行うことのメリットを整理するとともに、協働作業が保存活動に与える影響について整理する。

電子情報保存に当たっては、保存対象とする電子情報のライフサイクルの早い段階、すなわち情報発信者が電子情報を作成する段階から取り組みを実施する必要がある。この取り組みを先送りしてしまうと、その後の際にコストや手間が余計にかかる。しかし、多くの電子情報保存のためのプログラムでは、情報の作成社側と受け入れ側が別の場合が多い。

電子情報保存に当たって、保存対象となる電子情報のライフサイクルの早い段階から協働作業を行うことには、次のようなメリットが挙げられる。

（保存を実施する図書館等の機関におけるメリット）

- ・ 情報発信者との適切なコンタクトポイントを確立することができる
- ・ 保存対象とする情報を特定し、アーカイブに移転させる作業が容易になる
- ・ 電子情報の作者が保存の意志を持った段階で保存活動を実施する場合に比べ、電子情報が利用できるようになると同時に確実な保存ができる
- ・ 電子情報の作成者との間で、保存の際に取り扱う電子情報のデータフォーマットを削減する等合意を得ることができるので、保存の際の労力を低減することができる
- ・ 保存対象となる電子情報に関する情報を数多く手に入れることが可能なため、保存対象への理解を深めることができる
- ・ ビデオ、DVD、CD、カセットテープといったオーディオビジュアル関係資料によく見られるように、形態の異なる同内容の資料がある場合に、どれを保存するか選びやすくなる

上記に述べたとおり、情報の受け手である図書館等の機関にとって、電子情報のライフサイクルの早い段階から保存を行うことは、保存活動の効率を高めることにつながる。一方で、情報の発信元となる主体にとってのメリットとしては、以下のものが挙げられる。

（保存対象となる情報を提供する情報発信者におけるメリット）

- ・ 情報を提供する主体である政府や企業が電子情報の作成と管理の過程に保存に必要な要件を組み込むことができるため、作成した電子情報を安全に保存することが可能となる

- ・ 長期保存されるという意識を持つため、自身が作成する情報を貴重なものと感じるようになる。例えば、学術研究のコミュニティでは電子情報による出版に対する地位が向上する可能性もある
- ・ 保存を実施する主体と情報の作成者が緊密に連携することで、特定の企業の戦略に偏らないオープンなソフトウェアの利用が促進される可能性がある。このような環境が整備されることでさらに協調的な電子情報保存が可能になる

情報発信者と図書館等の機関の協働作業が成立すれば、双方にメリットがあるだけでなく、保存活動の質を高めることもできる。具体的な影響としては以下の点が挙げられる。

(情報発信者と図書館の協働作業による保存活動への好影響)

- ・ 保存活動における役割と責任が明確になる
- ・ 情報を提供する各コミュニティが保有している電子情報の種別を明確にすることができる
- ・ 情報を提供する人々に、自身が作成している電子情報の現在及び将来にわたる価値を理解させることができる
- ・ 電子情報の作成者が、各々が作成する情報の特徴や構成要素に関するアドバイスを提示してくれる場合があるため、保存に対するニーズを理解することができ、質の高い保存を実現することが可能である
- ・ 保存されている情報を利用する人々が、どのような技術環境を保有しているか理解を深めることができるため、利用に際して十分なサービスを提供することができる
- ・ アクセス制限のため、通常なら収集することができない情報(インターネット上でパスワードが設定されているサイト、ウェブロボットでは収集できないデータベース等を保有するサイト)を作成者から提出してもらうことができる。同様に、コピーガード等がかけられている CD-ROM 等についても、このような仕組みを組み込まない形で提出してもらうことができるため、保存活動の効率が高まる

8.2. 情報発信者と図書館の協働作業における検討事項

ここでは、情報発信者と図書館が協働作業を行うに当たって検討すべき事項について大まかに整理する。

情報発信者と図書館が協働作業を行う際の利益を最大化するには、数多くの論点が想定される。例を挙げるなら保存対象のプライオリティー付け、情報提供者からの許諾、情報の受け入れ側が提供する専門性、情報発信者と図書館の現状の関係等がある。

情報発信者と図書館が協働作業を行うために考慮すべき事項を大まかに整理すると以下の通りとなる。

(情報発信者と図書館の協働作業において考慮すべき事項)

- ・ 保存に係るスタッフの体制、文書化された手順、保存のためのガイドライン等、電子情報保存を容易にするための組織的な課題に関すること
- ・ 電子情報の保存に係るガイドラインやマニュアルに応じた活動ができるようスタッフを十分に訓練し、動機付けを行うこと
- ・ 知的財産権等を保護するために付与されているコピーガード等のアクセス制限について、保存の際にはこれらを無効にするような取り決めを行うこと
- ・ 電子情報保存のための情報システムに係るシステムデザイン等を含む全般的な計画に関連する事柄。この計画には、利用するファイルフォーマット、ファイル名の付け方、メタデータ、保存されている情報のバージョン管理等に関する事項が含まれる

上で述べた情報システムに関する事項は、保存活動の重要なポイントとなるため、その基本的な考え方について簡単に整理する。

保存の際に利用するデータフォーマットの選択は非常に重要である。選択されるデータフォーマットは幅広く受け入れられており、十分に標準化が進んでいるものでなければならない。その点では、オープンソースや仕様が公開されているデータフォーマットであり、極力シンプルなものが望ましい。このようなデータの例としては、XML が挙げられる。

利用するデータフォーマットを決めた上で、保存時のデータの妥当性に係る検証、一貫性のある明確なファイル名の付与に関する手順が重要になる。ファイルの識別には一貫した識別子 (Persistent Identifier) と情報システム内でのファイルの移動を記録し、移動した情報へのアクセス要求をリダイレクトする仕組みを整備する必要がある。また、保存されている情報を利用する際には、コピーを利用者に配信し、保存対象そのものが利用によって破損しないよう保護しなければならない。

メタデータについては、情報の作成段階で MARC やダブリンコアといった標準化の進んだものを情報発見のためのメタデータとして付与するとともに、保存する過程で対象となる情報の技術的な性質を記述するメタデータを付与していくというやり方が望ましい。記述する技術的な内容としては、アクセスに必要なソフトウェア、ライフサイクルにおける改変履歴等が考えられる。改変履歴を保持し続けていくことは情報の真正性を保持することにもつながるため、非常に重要であると言える。メタデータの保存については、保存対象となる情報パッケージに埋め込む形でも、メタデータファイルを別に作成し保存対象にリンク付けする形でもよい。

保存の初期段階においては、再生に必要なすべてのソフトウェアや特殊なハードウェアを保存していくことも有効である。このような方策は長期保存には不向きであるが、短期的には有効であると考えられる。

8.3. 情報発信者と図書館の協働作業のアプローチの考え方

ここでは情報発信者と図書館の協働作業を進める際の具体的なアプローチの考え方について整理する。

協働作業を始める第一段階としては、協働体制に関与する主体を特定することが重要である。関連する主体を特定した上で協議を行うことが、議論の有効性を決めると言える。図書館が協議を行う際の相手になると想定されるのは、出版協会等の特定コミュニティの代表や有力な個人である。協議においては数多くの課題点が指摘されるとともに、多くの時間を要することになる。このような協議を通じて、法制度等の側面について十分に議論し、協定を締結するのである。

協議を行っていく上で注意しなければならないのは、情報発信者となる主体の性質について理解しておくことである。情報を作成・公開する側は、電子情報を作成する際に極めて多様なアプローチ・技術を採用しているとともに、各主体の規模、長期保存における課題認識や知識は異なっている場合がほとんどである。

このような性質を持つ様々な主体を取りまとめて協議を行うことは、困難を極める作業になる。すべての関係主体の興味・関心を探り入れて合意点を見出すことも重要であるが、実際にアプローチするときには、数多くの情報発信者の内、特定のコミュニティに絞って協議を重ねておくことが有効である。保存活動の最終局面では、幅広い関係者のニーズを満足させる協定が必要であるが、初期段階で特定コミュニティと議論した際の経験や成果が役立つと考えられる。

また、協議を行う際には、情報の受け入れ側である図書館から情報の発信元に対して、十分なコミュニケーションを試みる必要がある。図書館側から、保存活動における情報発信者の役割について意識を喚起するためである。

このような協議を踏まえて、それぞれの役割・責任、協力の範囲を定義したガイドラインを策定することが保存活動の円滑化に役立つ。情報の発信元となる各主体は、文書化されたガイドラインに従って、それぞれの主体に特有の意見を伝えることができるとともに、情報を作成するスタッフの訓練等を行うことができる。

ガイドラインを策定した後も、長期保存に係る新たな基準を策定するために、主要な関係者と議論の機会を持つことが有効である。このような機会を通じて、関係者の意見を吸収し、情報発信者と図書館の利害調整を図ることで持続的な関係を保つことができる。特に情報発信者の知的財産権等の問題や、保存に係るコスト負担には十分な配慮をする必要がある。情報発信者にとって、電子情報の長期保存という活動は、通常業務とは関わりが薄い場合がほとんどである。そのため、保存に際して、過大な負担をかけ続けてしまうと、協働作業を維持することが困難になる。情報の発信元から十分なフィードバックを受けつつ、利害調整を図ることは重要である。

協働作業による保存活動が円滑に行われ続けられ、情報発信者から協働作業に対する更なる協力を引き出すことができる。

8.4. 情報発信者と図書館の協働作業における事例

ここでは情報発信者と図書館の協働作業に関する具体的な事例として、CENL (Conference of European National Librarians : ヨーロッパ国立図書館員会議) と FEP (Federation of European Publishers : ヨーロッパ出版連盟) の電子出版物の任意納本の協定モデル及び 2001 年に Andrew. W Mellon 財団の助成により実施された電子ジャーナルのアーカイビングプロジェクトについて取り上げる。

8.4.1. CENL/FEP の電子出版物の任意納本協定について

CENL は加盟する各国における電子出版物の任意納本協定の策定を支援する目的で、協定のモデル案となる "Statement on the Development and Establishment of Codes of Practice for the Voluntary Deposit of Electronic Publication" を 2002 年に策定した。この協定モデルの策定に当たっては、ヨーロッパにおける出版社の代表である FEP と 2 年間に及ぶ議論を重ねられ、図書館と出版社の双方の利害調整に基づく協定モデルが策定された。

この協定モデルはあくまでもモデル案であり、実際の適用に当たっては各国の事情を勘案して記載されている内容を変更することが推奨されている。協定モデルの改変にあたっては、CENL が協力して議論を進めることが想定されている。

協定モデルの策定に当たっては、既に電子出版物の任意納本の協定が提示されているドイツ、英国、オランダの例を参考にしている。また、この協定モデルはオフライン・オンラインの双方の電子情報について定めているが、オンラインの電子出版物については、別記を設けてさらなる検討を行うものとしている。

"Statement on the Development and Establishment of Codes of Practice for the Voluntary Deposit of Electronic Publication" の概要

出版を行う地域

- ・ 紙媒体の出版物には伝統的に出版地という概念がある。これはオフラインの電子出版物については当てはまるものの、オンラインの電子出版物については特別な規定が必要になる場合がある
- ・ ある特定の国において利用できるようになった出版物はその国の国立図書館に納められるべきである。ここでいう出版物の「出版者 (Publisher)」とは「出版物を公にするために出版あるいは公開を行うもの」と定義される。この定義によれば、最初に外国で出版され、その後、特定の国内で出版されたものについても納本の対象となる。出版地について何らかの問題がある場合は、出版者の所在地等により解決を図ることとする
- ・ ここでいう「出版物 (Publication)」とは「情報や知的活動の成果を公開する行為」と定義さ

	<p>れる。この定義によると、インターネット等の媒体で利用できる情報や知的活動の成果も出版物となる。一方で、イントラネットや個人的な目的で利用される媒体を活用するものは出版物とみなされない</p> <ul style="list-style-type: none"> ・ 納本の目的は、特定の国における出版物を収集して保存していくことにある
	<p>出版物の形態</p> <ul style="list-style-type: none"> ・ この協定モデルはすべての電子媒体に格納された非印刷物に適用される。すなわち、オンラインで公開されるものの他に、磁気テープ・磁気ディスク・CD-ROM、DVD等の光ディスク媒体で公開されるものを指す ・ オンラインで公開される電子出版物や常に更新されるデータベースの納本については、特定の国内における協議が必要となる。オンラインの出版物については保存の際の容量等に関する問題が予想されるが、納本や保存の手順について十分な実験等を実施してから保存活動を行うことを推奨する ・ ある特定の電子情報には、保存対象そのものの情報以外に再生等のためにソフトウェアが必要になる場合がある。出版者側が必要なソフトウェアのライセンスを納本先に変わって取得することが推奨されるが、任意納本の枠組みの中では必ずしも義務付ける必要はない
	<p>納本される出版物の内容</p> <ul style="list-style-type: none"> ・ 本協定モデルでは映画、音声、電子地図等の GIS 製品は対象外としている。これらは各国において既に別の枠組みで収集が行われている。本協定モデルを採用する国で、これらの対象を収集する場合には、協定モデルの見直しが必要である ・ 本協定モデルはエンターテイメント用の電子出版物については、十分に規定していない。このような情報を納本させる場合には、モデル協定を変更する必要がある ・ この協定モデルは無料、有料、双方の電子情報に適用される
	<p>納本における例外</p> <ul style="list-style-type: none"> ・ 同じ出版者から同じ内容の紙媒体の出版物が出されている場合は納本対象外とする（ただし、検索等の機能等の面で収集することが有効な場合は納本対象とする） ・ ある特定の組織内での利用を目的として出版される出版物は対象外とする ・ 特定の納本機関で既に対象外とされている出版物（例：コンピュータソフトウェア、コンピュータゲーム等）
	<p>納本される際のデータフォーマット</p> <ul style="list-style-type: none"> ・ マイクロフィルム、マイクロフィッシュ等の媒体についても納本を行う ・ 電子出版物は利用者が利用できる形態、すなわち、必要なソフトウェア、マニュアルやその他の資料を添付すること。様々なデータフォーマットで出版されているものについてはPC/AT互換機に対応したものを納本することを推奨する。その他のデータフォーマットに関する合意については図書館等の納本機関と出版者の間で協議を行うこと
	<p>出版物が納本される場所</p> <ul style="list-style-type: none"> ・ 紙媒体の出版物における法定納本制度の規定と同様の条件を適用すべきである。商業面・経済面・技術面で検討が必要な場合は出版者等と納本機関が協議を行うことを推奨する
	<p>納本部数</p> <ul style="list-style-type: none"> ・ 最小限の部数、特にオフラインのパッケージ媒体については、一部を国立図書館に納本する。複数の納本機関がある場合、あるいは紙媒体の出版物の納本制度等に特別な規定がある場合は

	それを適用することも検討する
	<p>納本される出版物のアクセスに関する合意</p> <ul style="list-style-type: none"> ・ 想定されるアクセスレベルには以下のものが挙げられる <ol style="list-style-type: none"> 1) 保存対象の情報が保管されている納本機関の構内で、利用が認められたユーザーが定められた端末でイントラネットを通じて、1度に1人利用できるようにするレベル 2) 納本対象が1部しか納本されない場合で、納本機関が複数ある、あるいは納本機関は一つでも建物が地理的に分散している場合は、高セキュリティなイントラネットを通じて、指定された端末で利用が許可されたユーザーが1度に1人だけ閲覧することを可能にするレベル 3) ネットワークに接続されていないスタンドアロンの端末で1度に1人のユーザーが利用可能とするレベル 4) 納本機関の保管庫に保存対象が保管されているものの、直接的なアクセスは認めないレベル。この場合、納本機関の構内に設置される指定された端末から、利用を認められたユーザーが、出版者側のサーバーにアクセスする形態をとる <p>複数の図書館で資料を共用する、あるいは貸借する場合には、その対象を納本した出版者との合意に基づき、利用料・ロイヤリティ等の支払いを行うことが適切である</p> <p>保存資料へのアクセスは、権利所有者の権利を侵害しないよう制限を設ける必要がある。1度に1人といったアクセス制限を行う場合には、技術的な工夫が設けられなければならないが、このような措置を設けることは納本機関側の義務である</p> <p>特定のカテゴリーの出版物で、一定期間のアクセス制限に納本機関が合意する場合には、納本する出版者等のニーズを満たすような技術的な措置を講じる必要がある</p>
	<p>納本されている電子出版物のハードコピーの作成</p> <ul style="list-style-type: none"> ・ 紙媒体等へのハードコピーが可能な情報の場合、ハードコピーが可能なページ数について、紙媒体の資料と同様の制限を設ける必要がある ・ データベース等の特定の出版物については出版者と協議の上でハードコピー作成数の上限を決める必要がある。上限を協議する際には、図書資料としての利用を考慮に入れるとともに、紙媒体の資料における上限と整合性のとれる形で協議を行う必要がある
	<p>納本されている電子出版物のダウンロードと情報の保存</p> <ul style="list-style-type: none"> ・ 保存されている資料のダウンロードや情報の保存については情報の提供者とのライセンス契約に基づくこととする。出版者とライセンスに関する協議を行う際には、紙媒体の資料における規定を十分に考慮するものとする
	<p>保存目的でのコピーの作成</p> <ul style="list-style-type: none"> ・ CD-ROM 等のオフライン系の媒体については将来的なアクセスを保証することができないため、長期保存を目的に格納されている情報の内容を別の媒体、あるいは新たな技術環境に移行することを出版社等の情報提供者は認める必要がある。そのため、納本される情報はコピーガード等の機能を無効にして納本される必要がある。コピーされた情報はユーザーに提供されるとともにマスターコピーの代わりとして保存されることとなる。アクセスに関する制限はコピーされる前の情報と同等とする ・ 一定期間を経過した後、閲覧不能にする機能等が組み込まれている場合には、このような機能を無効にしておかなければならない(このような権利はあくまでも納本機関に対してのみ)

	<p>認められるもので、エンドユーザーに認められるわけではない)。ほとんどの場合、期限切れとなった情報は新たなバージョンと交換がなされるが、納本機関はすべてのバージョンを保存すべきである</p>
	<p>納本協定の発効</p>
	<ul style="list-style-type: none"> 協定の発効は、協議を行う国の定めるところにより決定される。紙媒体の納本協定が存在しない等の理由で、出版者等及び図書館等の納本機関の間に認識のギャップが存在する場合には、本協定モデルに基づく規定により、協定の発効前の電子出版物についても納本を行うことを推奨する
	<p>オンラインの電子出版物に関する別記事項</p>
	<ul style="list-style-type: none"> ドイツ、オランダ、英国等でオンラインの電子出版物の納本に関する提案が行われているが、出版者側は、任意納本の枠組みにおける出版物の定義や実行面、技術面、資金面等について考慮を要するとのコメントが出されているところである 法定納本機関がすべての種類のオンライン出版物を保存することは不可能であるし、保存のための調整も非常に複雑になると考えられる。数多くの議論や研究が行われているにもかかわらず電子情報保存の長期保存に対する解決策が見出されていないことも課題である。しかし、オンラインの電子情報の保存に関する問題については、着実に解決に向けて動いている オンラインの電子情報がその商業的なライフサイクルを越えて長期保存されなければならないというのは既に広く認識されており、そのような活動を行うのは納本機関が適切であると考えられている。最近では CD-ROM 等の媒体よりもインターネット等のオンラインメディアを使うことが出版のトレンドとなっている。CD-ROM 等の出版物についてもオンラインを補足するコンテンツを準備する機会も増えており、これらのコンテンツは常にアップデートされている。これらの保存の問題も検討の視野に入れなければならない。 オンラインの電子情報の任意納本に関する規定を盛り込むことが必要であることは認識されているが、オンラインのみ、あるいはオフラインとオンラインのハイブリッド出版物の納本について情報発信者と法定納本機関が協力して実験を行うことを推奨する。このような協働作業を通じて課題点などを十分に検証する必要がある オンラインのみ、あるいはオンラインとオフラインのハイブリッドといった様々な種類の出版物（最初は電子ジャーナル等がよいと考えられるが）について、それらの保存、アクセス制限等の事項を検討するための実験を実施することを推奨する。これらの実験を通じて、出版物の種別毎の課題、データフォーマット等の技術的な課題を明らかにしていくことが有効である ヨーロッパにおいては各国の国立図書館が Cedars プロジェクト等において特定のオンライン電子出版物の保存やアクセス等について実験を行っている。数多くの出版者が情報の受け入れ側の国立図書館と数多くの協議を行っている 上記の活動は納本機関となる図書館等と出版者の通常の購読関係、あるいは通常の図書館サービスとは別の活動として行われている

表 8-1. CENL と FEP の電子出版物納本協定モデルの概要について

8.4.2. Mellon e-journal Archiving Project について

2001年、Andrew W. Mellon 財団は、複数の大学図書館等に対して電子ジャーナルの保存プロジェクトに関する資金助成を行った。このプロジェクトでは、各図書館等がそれぞれのテーマに沿って実験を実施した。以下にプロジェクトの概要を整理する。

プロジェクトの観点	参加主体
情報発信者となる出版者と図書館が協働して電子ジャーナルの保存を検討する観点	【グループ1】 ・ ハーバード大学 ・ Wiley 社 ・ Blackwell 社 ・ シカゴ大学出版社 【グループ2】 ・ イェール大学 ・ エルゼビアサイエンス社 【グループ3】 ・ ペンシルバニア大学 ・ オックスフォード大学出版社 ・ ケンブリッジ大学出版社
電子ジャーナルの主題から保存を検討する観点	・ コーネル大学（農業関係雑誌） ・ ニューヨーク公共図書館（公演芸術関係雑誌）
動的な内容変化を伴う電子ジャーナルを保存する観点	・ マサチューセッツ工科大学

表 8-2. Mellon e-Journal Archiving Project の概要

ここでは、情報発信者と図書館の協働作業に関する事例として、ハーバード大学と三つの出版社とのプロジェクトについて取り上げ、協働作業の一環として検討された事項について整理する。

ハーバード大学の e-Journal アーカイビングプロジェクトの概要

ハーバード大学の電子ジャーナルアーカイビングプロジェクトは 2001 年 1 月から 2002 年 3 月にかけて実施された。プロジェクトは、ハーバード大学図書館の他に Wiley 社、Blackwell 社、シカゴ大学出版社の 3 社をパートナーとして進められ、これらの出版社が出版する合計で 1,137 に及ぶ電子ジャーナルをハーバード大学図書館が保存していくための検討が行われた。

この 1 年間のプロジェクトでは電子ジャーナルを保存するアーカイブを構築するために種々の検討が行われたが、次のようなテーマが設定された。

- ・ ビジネス面・技術面を考慮した上で、保存すべき電子ジャーナルの構成要素を検討

- ・ 納本メカニズムとデータフォーマットに関する検討
- ・ 保存されている情報へのアクセス・コントロールに関する検討
- ・ 長期保存のためのガイドラインに関する検討
- ・ アーカイブの運用とメンテナンスに関する検討
- ・ アーカイブを持続するための資金面・運用面でのモデルに関する検討

検討は1年間という短期間であったため、すべての事柄について最終的な回答が出されたわけではない。しかし、出版社と協働作業を行うために必要な項目についてはアーカイブ構築を視野に入れた実践的な検討が行われた。

ハーバード大学のプロジェクトにおける検討事項

ここでは、先に述べた検討事項の内、具体的な内容が検討された主なものとして「ビジネス面・技術面を考慮した上で保存対象とする電子ジャーナルの情報内容」、「納本メカニズムとデータフォーマット」、「保存されている情報へのアクセス・コントロール」について整理する。

アーカイブにおける情報の内容に関する検討事項

電子ジャーナルを保存する際には、単純に記事を保存すればよいと考えがちだが、電子ジャーナル自体は多岐にわたる要素から構成されているため、容易に保存できるものではないと考えられる。例えば、電子ジャーナルには序文、著作権情報、ISSN 番号、編集委員会等といった雑誌の冒頭における情報、あるいは記事・ケースレポート・コメント・コラム・編集後記等といった本文の情報、といったように多様な構成要素が挙げられる。また、広告、他の文書へのリンク、本文の記事を補足するために付属されているデータセットや音声ファイル等も電子ジャーナルを構成する重要な要素となっている。

このように多様な情報構成を持つ電子ジャーナルを保存するために、ハーバード大学のプロジェクトでは21の電子ジャーナルを選択し、情報の構造等を調査するとともに、どのような構成要素を保存すべきかユーザー調査を実施した。ユーザー調査では、ハーバード大学の各学部に対して電子ジャーナルのどの部分の見た目と感覚をどの程度まで保存すべきか質問票を送付して保存対象とする構成要素に優先順位を付けた。

上記の調査を基にして、電子ジャーナルの構成要素を保存対象とする情報及び対象外とする情報に整理した。以下にその分類を記載する。

対象分類	概 要
保存対象とする情報	<ul style="list-style-type: none"> ・ 記事本文（記事を構成する図表等も含む） ・ 補足資料（著者が論文の補足情報として提供するデータセット、音声・画像ファイル、プログラム等を含む） ・ 著者が提供する参考文献 ・ 外部リソースへのリンク

	<ul style="list-style-type: none"> ・ 要約 ・ 目次 ・ 納本されていないものに関する情報⁸⁰ ・ 記事以外の文書（コラム、編集後記、投稿等の情報等） ・ 書誌情報（記事、その他文書に付するメタデータも含む） ・ 編集委員会に関する情報 ・ 編集者に関する情報 ・ ディスカッションリスト（特定の記事に関してディスカッションを行うオンラインの掲示板） ・ 著作権関係の情報 ・ 編集方針 ・ レビューリスト ・ 電子ジャーナルの内容に関する説明 ・ 同じ内容の紙媒体版のカバー画像
保存対象外とする情報	<ul style="list-style-type: none"> ・ 著作権の移転、原稿の準備や提出に関する情報 ・ 購読に関する情報 ・ 広告 ・ ビジネス関係の情報（バックナンバーの販売に関する情報、購読者サービスに関する情報等を含む） ・ その他付加情報（求職情報等を含む）

表 8-3. 保存対象となる情報、対象外となる情報の概要

納本メカニズムとデータフォーマットについて

保存のためのアーカイブシステムは、ハーバード大学におけるデジタル・ライブラリー計画で整備された既存の情報システム基盤を活用することとした。このシステム基盤は3年計画で整備された Digital Repository Service(DRS)を核にして構成されている。オラクル社のデータベースソフトを基礎として、24万件の情報が管理用のメタデータとともに保存されている⁸¹。

この情報システム基盤を活用して、保存のためのサブシステムを OAIS(Open Archival Information System)の技術標準に基づいて構築することが検討された。OAISには受入(Ingest)、データ管理(Data Management)、アーカイブ保存領域(Archival Storage)、保存計画(Preservation Planning)、アクセス(Access)、管理(Administration)の6つの機能が規定されている⁸²。

アーカイブシステムは自動化を図ることが可能である。例えば、出版社の登録や認証、

⁸⁰ 例えば、紙媒体の論文では記載されていた内容が電子版の論文では含まれていないといった事柄について記載する

⁸¹ 情報システムに保存されている情報は120ギガバイトに及ぶ

⁸² 詳細については第3章を参照

保存対象の情報である Submission Information Package (SIP) の提出、情報の受け入れと妥当性のチェック、保存時点での SIP から Archival Information Package (AIP) への変換、マイグレーションの実施、定期的なレポート、アクセス要求への対応等の自動化は十分に可能である。しかし、これらの自動化は、情報の発信者から提出される SIP の質に大きく依存する。そのため、情報発信者とアーカイブとなる図書館との協働作業により、SIP の作成に関する基準を策定することが有効であると考えられた。協議の結果、ハーバード大学と協調している三つの出版社は、アーカイブにおける共通のデータフォーマットの採用に合意している。将来的には、ハーバード大学がアーカイブシステムに関わる主体に対して標準に係る文書やツール等、必要な支援を行うことが検討されている。

SIP を作成する具体的な手続としては、XML を基にした METS (Metadata Encoding and Transmission Standard) を活用して保存対象の技術関係情報を記載したメタデータを付与することとなっている。メタデータは電子ジャーナルのタイトルのレベル、出版年次別のレベル、記事のレベルといった階層構造が規定されており、SIP の作成に係るドラフト版の様子がインターネットで公開されている⁸³。

保存に際して採用しているデータフォーマットは標準的なものであり、例えば、メタデータとテキスト情報には XML、ラスター画像には TIFF⁸⁴、ベクトル画像には SVG 等が採用されている。これら以外のデータフォーマットについては、情報の受け入れの時点で標準的なものに変換することとしている。情報発信者となる出版社等はこれらのフォーマットを念頭において電子ジャーナルを作成することとなっている。

PDF ファイルについても、特定企業の戦略に左右される危険性はあるものの、保存対象として受け入れることにしている。PDF ドキュメントの作成の際には、電子ジャーナルの出版社が遵守すべきルールを設けて、PDF で提出してもよいデータに制限を設ける等、アーカイブ側で受け入れるデータを均質なものにするよう配慮がなされている。

保存されている情報へのアクセス・コントロール

アーカイブに保存されている電子ジャーナルに対するアクセスに関する事項は、出版社との協議において、課題となっていた点である。アクセスに関する課題は、大まかに分けて、「アクセスできるユーザーは誰か」、「どのような状況でアクセスが認められるのか」といった点に分けられる。

アクセスを認めるユーザーについては、ハーバード大学側は、情報の管理を目的に行うアクセス以外にも、閲覧を認める方向で検討していた。しかし、パートナーとなる出

⁸³ Harvard University Library Submission Information Package (SIP) Specification Version 1.0 DRAFT (2001 December) http://www.diglib.org/preserve/harvard_sip10.pdf を参照

⁸⁴ 読みは「ティフ」。Aldus 社と Microsoft 社によって開発された画像データのフォーマット。1 枚の画像データを、解像度や色数、符号化方式の異なるいろいろな形式で一つのファイルにまとめて格納できるため、比較的应用ソフトウェアに依存しない画像用のデータフォーマットである

版社側から、ユーザーを直接的に自社のサーバーにアクセスさせるようにしたいという要望や、不正アクセスに対する監視機能、ハーバード大学にライセンスしていない電子ジャーナルに対するアクセスへの懸念が寄せられた。

協議の結果、ダークアーカイブと呼ばれる期間を設け、この期間中はアーカイブシステムの運用者や外部の監査人によるアクセスのみを認めることとした。また、購読していることを合法的に証明できるユーザーに対しては、購読している雑誌に限りアクセスを認めることとした。

先に述べたダークアーカイブの期間は一般的なユーザーのアクセスが許されていない状態であるが、トリガーイベントと呼ばれる契機を境にしてアクセス制限が取り除かれる方向で議論が重ねられた。トリガーイベントは、情報の発信元となる出版社、論文の著作者、購読者等の利害に関係する概念であるため、今後もさらなる議論が必要であるが、保存されている電子ジャーナルにアクセスする権利を定めるために重要な概念となる。以下にトリガーイベントの考え方について整理する。

- ・ 保存されている電子ジャーナルの特定の号・記事が、出版元あるいは他の主体からもオンラインで利用できなくなった場合
- ・ 保存されている電子ジャーナルの出版社が、特定の電子ジャーナルの出版権利を他の主体に移転した場合
- ・ 保存されている電子ジャーナル毎に出版社が設定する特定の年限を越えた場合
- ・ 保存されている電子ジャーナルが絶版になった場合
- ・ 保存されている電子ジャーナルに特定の権利が主張されない状況になった場合

ここで述べたトリガーイベントに関する考え方は、議論の過程でも種々の変更が行われたが、今後もさらに議論が深められる予定である。

上記で述べたアクセスに関する問題の他にも、ハーバード大学のプロジェクトではアーカイブの運営に関する費用の問題が検討された。具体的な内容としては、アーカイブの運営費用をコントロールするためにいくつかの方策を検討したことと、情報の発信元である出版社からアーカイブに関する費用の拠出を検討したことが挙げられる。

アーカイブの運営費用をコントロールするための方策として検討されたのは、以下の点である。

- ・ アーカイブのシステム費用を低減させるため既存のデジタル・ライブラリー環境を活用すること。既存のデジタル・ライブラリーシステムと技術基盤を共有するとともに、既存の組織でシステムの運用を行うことを検討した
- ・ アーカイブシステムの自動化を進めることで、保存活動に要する人件費を削減することを検討した
- ・ 多大な開発費用を要するアーカイブシステムの機能開発を制限する。具体的には、購

読管理システム等の開発を行わないこととした

また、ハーバード大学と出版社の協働作業の過程では、アーカイブシステムの運用費用を情報発信者となる出版社から提供してもらうことが検討された。アーカイブに拠出する運用費用の資金源としては、保存されている情報を提供することでサービスの対価を得ることも考えられるが、この方策は採られなかった。ハーバード大学のプロジェクトでは、出版社側が自身の情報をハーバード大学が再販売して対価を得ることを嫌ったこと、保存活動以外にサービスという多大な労力を要する活動が追加されることが懸念されたからである。

協議を行った結果、出版社から年間の保存活動の手数料としてアーカイブの運営費用が支払われることが議論された。この保存活動の手数料を拠出するための方策として、出版社は著者から費用を徴収する、学術研究団体から資金助成を受ける、あるいは購読者から費用を徴収するといった内容が検討された。出版社が拠出する保存活動に係る手数料は、以下の二つの要素からなると考えられている。

- ・ 日々提供される電子ジャーナルの受け入れ手続、品質管理、保存準備等の活動に対する対価
- ・ 長期保存のための活動に要する費用を賄うための基金への寄付

上記の考え方にに基づき、ハーバード大学の電子ジャーナル保存プロジェクトでは、アーカイブ運営のための資金源を確保する方策が検討された。

8.5. 情報発信者と図書館の協働作業に関する考え方

本章では情報発信者となる出版社等の関係者と情報の受け入れ側となる図書館との協働について整理した。本章の前半でも整理した通り、情報発信者と図書館が協働作業を行うことは、双方にとってメリットが見出せるとともに、アーカイブにおける保存活動の質にも好影響をもたらすと考えられる。

円滑な協働作業を進めるにあたっては関係者と十分な協議が必要であるが、取り組むべき課題は多岐にわたる。特に重要と考えられるのは、情報発信者と図書館の間で交わされるやり取りの過程における共通認識・手順の確立と、アーカイブに保存している情報の利用を想定した場合の情報発信者への経済的な影響である。

情報発信者と図書館の間での共通認識・手順の確立については、電子情報の保存に係る課題認識を深め、情報の発信元と図書館の間に協力関係が築けるよう相互に十分な意見交換を行うことが重要である。このような議論を重ね、情報発信者の現状を理解した上で、効率的な保存を行えるよう情報の提供を受ける際のデータフォーマットや情報の提出に当たって「標準」となる手順を確立することが重要である。標準の確立は、情報発信者と図書館とのやり取りに続く保存活動の自動化を可能にする側面も持ち合わせているため、十分な検討がなされなければならない。

本文で登場した CENL と FEP の協定モデル、ハーバード大学の電子ジャーナルアーカイビングプロジェクトの双方において、保存資料へのアクセスは大きな課題となっていた。保存資料のアクセス方法によっては、情報発信者に経済的な打撃が及ぶことも予想されるからである。保存、そして先に続く利用を考えるに当たっては、情報発信者の経済的利益を阻害しない仕組みの構築が必要になる。この問題には著作権等の知的財産権の問題も影響してくるので、十分留意する必要がある。

本文に登場した二つのプロジェクトにおける情報発信者と図書館の協働作業に関する議論は、CENL と FEP の協定モデルが 2 年、ハーバード大学のプロジェクトは 1 年の時間を要した。大きな前進は見られたものの、双方とも完全な結論に至ることはできなかった。情報発信者と図書館の協働作業における議論は時間をかけた上で、双方の利害を十分に調整する必要があることが示唆されている。

9 終わりに

電子情報の長期にわたる保存とアクセスの保証という課題について、主に海外における取り組みを見てきた。取り上げた対象は限られたものであり、また、それぞれのテーマの表層をなぞった程度で、いわば序章にすぎないものだが、この調査研究のプロセスを経て明らかになったことがある。電子情報を保存していくための唯一万能の技術やシステムは、まだ存在しないということ。ただし諸外国においてはこの課題の重さへの認識が高く、多様なアプローチによる研究、実験、プロジェクト等が活発に行われているということである。日本では一部を除いて、そうした活動を見ることはできない。

唯一万能の方法はないとはいえ、海外での様々な取り組みを通して、いくつかの共通の認識が生まれつつある。その一つとして、電子情報の保存においては、単独の機関がすべての責任を負うことは考えられず、コミュニティやメディアの種類によって責任を分担し、協力関係の中で保存活動を行う必要があるという見方が挙げられる。ネットワークを介して流通する電子情報をとってみても、ウェブサイト、電子ジャーナル、電子書籍、デジタル音源、デジタル動画、デジタルテレビ・ラジオ等、それぞれ制作者や形態が従来のメディアとは異なり、その保存に当たっても異なるアプローチが求められるものが多い。これらのすべてを、例えば国の一機関が単独で保存していくことは、自国の電子情報に限っても不可能である。ここに協力関係の必要が浮かび上がるのだが、では、どうやってそうした関係を築いていくかが今後の課題になる。

本論で見てきたように、協力関係は次の四つのレベルに分けることができる。

- ・ 電子情報の制作機関及び制作者と保存を行う機関の協力関係
- ・ 研究開発機関及び研究開発者と保存を行う機関の協力関係
- ・ 電子情報の保存を行う機関相互の協力関係
- ・ 電子情報の提供を受けるコミュニティと保存を行う機関の協力関係

今回の調査研究は三ヵ年計画の初年度の取り組みとして行われた。今後、この計画を推進するに当たって、上に挙げた協力関係機関の特定と具体的な提携への準備が一つの柱になる。

本調査研究の初年度作業として実施予定であった課題のうちで積み残した課題もある。

当初、日本国内で流通する電子情報の種別、発信元、流通量を調査し、国立国会図書館が優先的に保存すべき電子情報についての課題整理を行うことを予定した。しかし調査研究を進める中で、電子情報のカテゴライズと量、内容の把握が、予想していたよりも困難な作業であることがわかった。全体の見取り図を描くための有効な手段を見出せないまま、調査研究を終了した。本論でも触れた通り、長期にわたって保存する電子情報を選択する際、この見取り図は判断材料として不可欠のものであり、今後の課題としたい。ただし、そうした情報をもとに、保存する電子情報の選択方針を策定するには時間がかかる。選択の方針は先に述べた他機関との協力関係のあり方によって大きく左右されるからである。

その間に多くの貴重な電子情報が失われていく。国立国会図書館には既に、近代デジタル・ライブラリーを初めとする紙媒体から電子化した膨大な資料群や平成 14 年度から行っている「インターネット資源選択的蓄積実験事業」(Web Archiving Project 略称 WARP) で収集されたウェブ情報等のコレクションがある。また CD や DVD 等、パッケージに納められた電子情報は、1998 年に納本の対象資料として、収集保存されている。それ以前のフロッピーディスクやレーザーディスクも所蔵されている。平成 15 年度には当館が所蔵するそれらの電子情報を調査し、現状の把握を行う予定である。またデータの消失の恐れがある一部の電子情報を使用して、保存のテストを行うための環境を検討していくことも予定している。

公的機関による電子情報の長期保存を必要とする声は多い。電子情報の保存は優れて分散的なアプローチを要する作業だが、同時に少数の機関が主導して、実践と協働を促すことが求められている。国立国会図書館もその役割の一端を担い、デジタル化された文化の記憶を次世代へ繋ぐための第一歩を踏み出すための素材として本報告書をまとめた。

最後に、調査研究の一環として、二回の調査研究会議を開催し、公共機関・民間機関へのアンケートを実施した。会議での議論、アンケートの回答は本報告書の一部に反映させた。本報告書刊行に当たって、皆様方の御協力に感謝したい。

本調査研究の実務的作業は、国立国会図書館との契約に基づき、株式会社NTTデータが行いました。