

企業におけるデジタルアーカイブの取り組みとその技術的課題

秋元良仁^{†1} 斎藤伸雄^{†2}

近年、種々の文化資産をデジタル化し、保存・活用しようとする動きが活発である。企業においては、このような活動を支援するための情報システムの提案、構築、運用が盛んに行われている。本稿では、様々なデジタルアーカイブ向け情報システムを概観し、特に運用上の観点からアーカイブされたデータを長期保存するための技術的課題について考察する。

The Technical Challenges of Undertaking Activity for Digital Archives

RYOJI AKIMOTO^{†1} and NOBUO SAITO^{†2}

Recently, digital archiving technology has progressed. Then, a lot of digital archive system has proposed. In such a situation, it is necessary that the methodology preserve valuable data according to requirements from archiving systems. In this paper, we summarize the current state of digital archive and problem of preservation. Based on it, we discuss the solution to this problem.

1. はじめに

情報技術の急速な進展に伴い、多様なデジタルコンテンツの作成及び利用が容易になりつつある。特に、1990年代以降、インターネットに代表される高速な広域ネットワークの普及を受け、ネットワークを通じて受発信されるデジタルコンテンツは飛躍的に増加している。

このような状況において、デジタルコンテンツを作成、整理、蓄積、配信、利用する場としてデジタルアーカイブに注目が集まっている。

デジタルアーカイブは、以下に示す様々な情報技術を利用した統合情報システムと言える¹⁾。

- 文書・画像・映像・音声等の入力機器や各種ソフトウェアを用いたデジタルコンテンツ作成技術
- データベース化、メタデータの組織化/構造化技術に基づくデジタルコンテンツ管理・保存技術
- Web リソース/メディア・印刷媒体等へのフォーマット変換やデリバリサービスに基

づくデジタルコンテンツ配信技術

- 情報資源発見のための情報収集/検索技術を用いたデジタルコンテンツ利用技術

このような統合情報システムは、企業における資産管理や、多様な Web コンテンツの収集・利用、あるいは図書館や博物館における文化資産の管理や活用等に供される。

(1) 企業におけるアーカイブ

近年では会計不祥事やコンプライアンスの欠如を防止するため、金融商取引法の一部が改正され、俗に J-SOX 法として企業における内部統制の強化が行われている。内部統制においては違法行為や不正防止のため、職務と管理体制を明確化し、業務の処理規定やマニュアル等のコンテンツ管理が不可欠となる。また、万が一情報漏洩等の不祥事が発生した際には、その証拠保全のため、電子メールを長期に渡り保存するメールアーカイブの構築を行っておく必要も生じる。

(2) ウェブアーカイブ

WWW 全体あるいは特定の一部を収集し、含まれるハイパーリンクやマルチメディアデータの整合性を保った状態で言語別・組織別・テーマ別等に再構成して利用に供するものにウェブアーカイブがある。この取り組みは WWW 上に存在する様々な分野の価値ある情報やデータを恒久的に保存し、後世の研究者が現代社会を研究す

^{†1} 凸版印刷株式会社 総合研究所
TOPPAN PRINTING CO., LTD. Technical Research Institute

^{†2} 凸版印刷株式会社 情報コミュニケーション事業本部
TOPPAN PRINTING CO., LTD. Information and Communication Division

るために必要不可欠な情報源として利用できるように整備したデジタルアーカイブと言える。

(3) 文化資産アーカイブ

人間の知的活動の所産である有形無形の文化財、記念物、景観、建造物、図書、歴史史料、公文書等、様々な文化資産をデジタルアーカイブシステムを用いて保存・活用しようとする動きも活発になりつつある。例えば、文化庁と総務省は2003年度から、全国の国公立美術館・博物館が管理する收藏品や史跡、記念物等に関する資料をデジタル化し、容易に検索閲覧できる仮想博物館とも言える文化遺産オンラインをインターネット上に公開している²⁾。また、コンピュータメーカーや印刷会社は各社が持つデジタル技術の応用として、美術館や博物館と共同で、文化財の持つ特徴をイメージや映像の形態で半永久的に保存する試みを始めている³⁾⁴⁾。この他にも、新しい展示形態や美術体験を情報技術を用いて行おうとする試み⁵⁾や、美術館・博物館が持つ研究資源を共有化し、資源の質的・量的な拡充を行うと共に、横断的な研究の促進を行おうとする試みも始まっている⁶⁾。

企業においては、上述のような試みを支援するためのデジタルアーカイブシステムの提案、構築、運用が盛んに行われている。このようなシステムに共通する特徴として、アーカイブデータの長期保存がある。アーカイブデータは、企業における証拠保全のように比較的短期間の保存から、ウェブアーカイブや文化資産アーカイブにおける長期間の保存に至るまで、内容を損なうことなく確実に保存しなければならない。しかし、保存期間の定義や、メディア媒体、マイグレーションやエミュレーション等のソフトウェアを介する保存手法の定義、保存に関するメタデータの定義等、未解決のまま残されている問題も少なくない。

そこで、本稿では、デジタルアーカイブの要素技術となる様々な情報技術を概観し、特に文化資産向けのデジタルアーカイブにおけるアーカイブデータの長期保存に関する技術的課題について考察する。

本稿の構成は以下のとおりである。まず、第2章では、基本的事項としてデジタルアーカイブにおける要素技術について説明し、次に第3章でデジタルアーカイブデータの長期保存に関

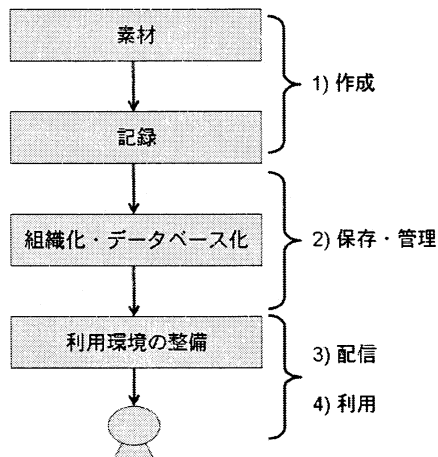


図1 デジタルアーカイブの構築

する技術的課題について述べる。最後に、第4章でまとめと今後の課題について述べる。

2. デジタルアーカイブの要素技術

2.1 デジタルアーカイブの構築

一般に、デジタルアーカイブの構築はデジタルコンテンツの作成、管理・保存、配信、利用の各技術からなる。図1にデジタルアーカイブの構築工程を示し、以下、詳細に説明する。

2.2 デジタルコンテンツの作成

デジタルコンテンツの作成は、以下の2点に分類することができる。

- テキスト・静止画・動画・音声等のデジタル入力機器やソフトウェアを用いるコンテンツ(ポーンデジタル)の作成
- 物理的な素材(文書・絵画・建築物・録音テープ等)に対して、デジタル計測可能な入力機器を用いてデジタル化されるコンテンツの作成

前者は、携帯小説やオンライン小説等の電子書籍、デジタルカメラやペイントツールを用いて作成されるデジタル写真/絵画、デジタル音楽、CG(Computer Graphics)を用いた映画等、芸術作品を制作する段階から各種の入力機器やソフトウェアが用いられる。他方、後者は対象とする物理的な素材に合わせて様々なデジタル化が行われる。

文書/絵画等のデジタル化

一般に、手書きあるいは印字された文書をディ



図2 絵画のデジタル化の例

デジタル化する場合、OCR (Optical Character Reader) を用いて文字を光学的に読み取り、類似辞書とのパターンマッチングにより文字データ化する方式が取られる。この際、対象となる文書が作成された時代に使用されていた文字群を想定し、類似辞書を現代の文字形のパターンから微調整することが重要となる。

文書の中には貴重書も存在し、文字データのみならず、文書そのものの持つデザイン、色彩、文字字形等、文化財的な特徴が重要になる場合もある。そのような場合、絵画のデジタル化と同様の手法が取られる。即ち、絵画(あるいは文書)に負担をかけずかつ忠実に再現できるように撮影機材と撮影条件の選定、画像サイズとファイル形式及びカラープロファイルの選定、撮影画像のノイズ除去や歪み補正、効率的な符号化手法の検討、電子透かしによる不正利用の防止等、様々な処理が行われることになる。図2に凸版印刷が参加した DADDI(Digital Archive through Direct Digital Imaging) プロジェクト^{*1} における絵画デジタル化の例を示す。

音声/映像のデジタル化

欧州連合(EU)の Culture 2000 programme^{*2} は、アナログビデオテープやコンパクトカセット、マイクロフィルムやレコード等のアナログ媒体に記録される情報の保存プロジェクト TAPE(Training for Audiovisual Preservation in Europe) を推進している。

近年では、VHS テープや8ミリビデオテープ等のアナログテープに記録されている映像音声

は、市販のDVDレコーダやカムコーダ、テレビチューナユニット等を用いることで容易にデジタル化することができる。しかし、貴重な音声映像が記録されているテープや、保存状態の良くないテープ等をデジタル化する際には、更に技術的なケアが必要となる。

慶應義塾大学メディアセンターでは、TAPEのガイドラインに基づき、貴重音源のデジタル化を行った報告がある⁷⁾。報告によれば、テープの汚れやリーダの劣化に対してスプライシングの再加工やリーダの交換を施し、テープの再生速度調節やサンプリング周波数や量子化ビット数の選定、AD変換時のファイルフォーマットの選定に至るまで決め細やかなデジタル化を行っている。なお、特筆すべき点として、様々な選定事項はメタデータとして記録しておくことが望ましいとしている。

三次元映像の作成

これまで、CGを用いた数多くの三次元アニメーションコンテンツが作られてきた。近年では、動きのある人間や動植物を対象とし、その動的な変化を記録する三次元ビデオ映像に関する研究が盛んに行われている。例えば、分散配置された複数のカメラで、動的に変化する対象素材を同期撮影し、多視点ビデオデータを得ることも可能となってきた。また、X線やスペクトル分光計測等を用いた計測技術の発達は、高精細かつ高精度に大小様々な素材を記録することを可能にしている。これらの研究は、例えば能や日本舞踊といった日本古来の無形文化財のデジタル化や、遺跡や建築物といった大型の有形文化財のデジタル化という形で成果を挙げつつある⁸⁾。

また、三次元映像化したコンテンツは大規模なシアター用スクリーンや、中小規模のフラットスクリーン、あるいは高解像度ディスプレイ等を用いて、臨場感や没入感等、人間の感覚を刺激する様々な展示形式が考案されている。

図3に凸版印刷が作成した故宮大和殿の三次元映像コンテンツと大型カーブスクリーンを用いたVR(Virtual Reality)シアターの例を示す。

2.3 デジタルコンテンツの保存・管理

様々な手法でデジタル化されるコンテンツの量は膨大なものになる。膨大な量のコンテンツを効率的に管理し、利用に供するには、ファイ

*1 <http://www.dnp.co.jp/artscape/knowledge/it/0207/kageyama.html>

*2 <http://ec.europa.eu/culture/archive/culture2000/cult.2000.en.html>

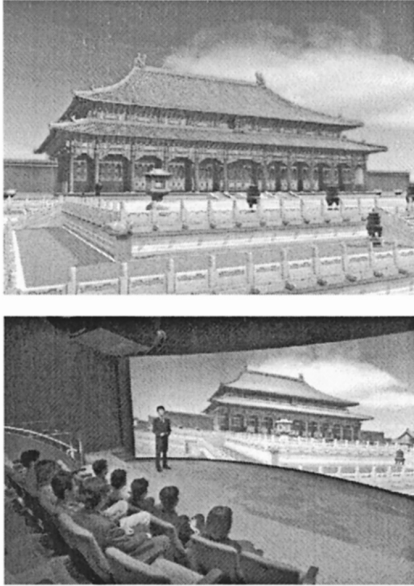


図3 三次元映像(上)とVRシアター(下)(制作・著作: 故宮博物院/凸版印刷)

ルシステムやデータベースシステム等, 何らかのストレージが必要となる。ストレージの設計にはハードウェアの構成やネットワーク技術のみならず, コンテンツに関する情報を記述したデータ(メタデータ)を設計する必要がある。

メタデータは, その記述内容に応じていくつかの階層に分けて考えることができる。即ち,

- **メタデータ実体**
コンテンツに関する属性情報
- **スキーマ**
どのような属性なのか, またどのように属性を与えるのか, 厳密な利用範囲(定義域(Domain)や値域(Range))やデータ型, 他のエレメントとの関連等の定義
- **モデル**
RDFに代表される, セマンティクスを持つグラフ構造による関係性の記述
- **流通形式**
XMLやHTML等, 実際にコンテンツ流通, 配信時に利用される形式

となる。メタデータの設計では, まず対象となるコンテンツの主題(Subject)を分析し, 属性情報を抽出する。次にスキーマにおいて属性情報の詳細な定義を行い, インスタンスはXMLやHTMLのような流通に適した形式で配信され

る。場合によっては, RDF等のモデルを用いてスキーマのバインドや意味の記述が行われる。また, 異種のメタデータ間の連携を必要とする場合, 概念を明確に記述できるオントロジを用いて行われる。

このようなメタデータの設計は様々な分野で行われており, 一部では標準化も進んでいる。例えば, 文化財分野に関するメタデータには, 国際博物館会議(ICOM)が提案しているドメインオントロジを用いた概念の関係性記述モデルであるCIDOC CRM(The International Committee for Documentation Conceptual Reference Model, ISO 21127)や, 東京国立博物館を中心に策定された情報共有の基盤モデルであるミュージアム資料情報構造化モデル, J.Paul Getty博物館が推進するCDWA(Categories for the Description of Works of Art), 構造化された生物標本や観察データの交換用標準であるDarwin Core等がある⁹⁾。

2.4 デジタルコンテンツの配信・利用

様々な手法で作成され, メタデータを用いて整理されたコンテンツは, 情報収集技術, 情報検索技術等を用いてユーザの利用に供される。

情報収集技術

OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)は, メタデータを自動的に収集・交換することを目的として策定されたプロトコルである。ハーベスタと呼ばれるクライアントシステムとリポジトリと呼ばれるサーバシステムからなり, リポジトリはハーベスタのリクエストに応じてメタデータを返す。プロトコルはHTTPをベースとしており, XML形式のレスポンスフォーマットを持つ。フォーマットは汎用的なSimple Dublin Coreが必須で用いられ, 拡張も可能である。

RSS(RDF/Rich Site Summary, Really Simple Syndication)はWebサイトの更新情報を簡潔にまとめ, 配信するための文書フォーマットのことを言う。Weblogによる利用, 新聞社や企業のニュース情報の配信, ラジオ局やテレビ局の音声映像データ公開手法であるポッドキャスト等に利用される。筆者らはかつて, 文化財情報の交換手法としてRSSを用いた情報交換システムについて提案を行ったが¹⁰⁾, 対象とするドメインと交換情報のフォーマットを定義しておけば,

ある程度の情報交換が可能となる。

情報検索技術

情報検索用プロトコルの Z39.50(ISO 23950, JIS X 0806) は、元々図書館等で管理される文献メタデータの情報交換を目的として策定されたプロトコルであり、共通のクエリ形式を使用できるため、ユーザはシステムの違いを意識することなく 1 つのクライアントから複数のシステムリソースを利用できる。近年では、WWW の進展に伴い、HTTP/XML を用いた SRU/SRW(Search/Retrieve via URL, Search/Retrieve Web Service) プロトコルにリプレースするケースも散見される¹¹⁾。前者の SRU は REST(Representational State Transfer) フレームワークに基づくプロトコルであり、後者の SRW は SOAP に基づくプロトコルである。

GETA は文書間や単語間の類似度を計算する汎用の連想検索エンジンである。単語出現マトリックス WAM(Word-Article Matrix) を作成することで、tf/idf 法等に基づく大規模文書のキーワード検索、文書要約等を高速に行うことができる。GETA は様々なサービスで用いられており、前述の文化遺産オンラインや、図書館の文献メタデータ検索サービスである Webcat Plus ^{*1} 等に利用されている。

その他、WWW サーチエンジンも多数考案されているが、本稿では割愛する。

3. アーカイブデータの長期保存

前章で述べたように、デジタルアーカイブシステムは様々な要素技術を組み合わせて構築する統合情報システムである。特に文化資産向けのアーカイブシステムでは、後世の人々が研究活動や教育活動等、様々なシーンで過去のデータを活用するために、あるいは後世に残すべきアーカイブデータを正しく継承していくという意味においても、アーカイブデータを長期に渡り保存する必要があり、またどのように保存するか検討することは非常に重要だと言える。

以下、アーカイブデータの長期保存における問題を概観し、問題の解決について考察する。

3.1 保存対象の範囲

デジタルコンテンツを長期保存する際に重要な点として、再生利用する環境が保たれている

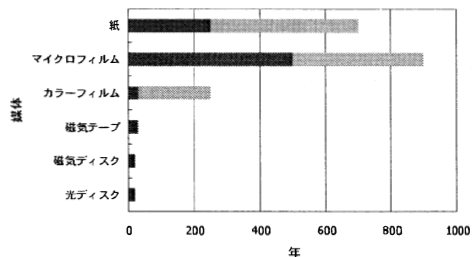


図 4 媒体の寿命

ことが挙げられる。例えば、何らかの画像コンテンツが片面二層の DVD-R に保存されている場合、まずディスクが DVD ドライブに装着され、ドライブがコンピュータに接続され、コンピュータには OS が搭載され、画像ファイルのフォーマットを再生できるソフトウェアがインストールされている状態で初めて閲覧することができる。つまり、画像コンテンツのみならず、再生機器や再生機器を制御するコンピュータ、OS、アプリケーションプログラム等の環境全てを保存の対象として考える必要がある。更に、各要素を詳細に見てみると、再生機器を構成する電子部品や機器を接続するコード類、あるいは再生機器を制御するデバイスドライバは代替/入手可能なものなのか、OS やアプリケーションプログラムは画像コンテンツのフォーマットを再生できるバージョンであるのか等々、様々な技術的な依存関係が存在する。また、これらを全て保存する場合には、アクセスを確保するためのメンテナンスや物理的なスペースの確保等も必要となり、実現は困難を極める。

3.2 寿命の問題

保存の状態や環境にも大きく左右されるが、紙媒体は約 250 年から 700 年、マイクロフィルムは約 500 年から 900 年、カラーフィルムは約 30 年から 250 年程度保存に耐え得るとされている。それに対し、電子媒体の寿命は非常に短い。アレニウス試験等での予測値によれば、磁気テープは約 30 年、磁気ディスクと光ディスクは約 20 年程度と言われている^{*2}。図 4 に媒体毎のおおよその寿命を示す。

寿命に関しては、媒体のみならずコンピュータや外部記録/記憶装置、OS やアプリケーション

*1 <http://webcatplus.nii.ac.jp/>

*2 <http://www.kyokuto-micro.co.jp/media.html>

プログラムにも存在する。例えば、ハードウェアには電子部品等の物理的な寿命が存在する。また、OSも古いバージョンはメーカーのサポートがなく、セキュリティホールが発見された際に対応もできない。アプリケーションプログラムも同様で、特定のOS上でしか動作保証がなされていないため、古いOS上でしか動作しないアプリケーションの場合、併せてOSも保存しておく必要が生じる。また、バージョンが古いために新しい環境で作成されたコンテンツを利用できないといったことも考えられる。

3.3 ソフトウェアを用いた保存手法

保存の範囲や対象の寿命に対して、ソフトウェアを用いた解決のアプローチがある。

3.3.1 マイグレーション

マイグレーションとは、アプリケーションプログラムやプログラム上で利用されるデータを現行の技術環境から他の技術環境に移行・変換する作業を指す。国立国会図書館の報告¹²⁾によれば、CD-ROMからハードディスクへの移行に限定した実験であるものの、市販のマイグレーション用アプリケーションを用いて比較的容易に移行が可能としている。また、長期保存の観点からは、

- プログラムの移行などにより長期間使用可能であること。
- マイグレーション後のファイル形式が長期間使用可能なものであること。マイグレーションプログラムのバージョンアップに伴い使用不可能にならないこと。
- マイグレーション後のファイルがCD上に存在するものとして扱えること。
- コピープロテクトが施されたCDもマイグレーションできること。

をマイグレーション用アプリケーションに求める要件として挙げている。また、環境を完全にコピーする場合を除き、マイグレーションにはデータのエンコードやフォーマット変換が生じる場合がある。このため、XMLやテキストデータ等、特定のプラットフォームに依存しないフォーマットの検討も重要となる。

3.3.2 エミュレーション

エミュレーションとは、ある環境向けに開発されたアプリケーションプログラムを他の環境上で実行させることを指し、エミュレータと呼ばれるソフトウェアを用いて行われる。エミュレー

タは疑似的に他の環境及び環境上で動作するアプリケーションを再現するため、保存対象が技術的な依存関係を持っている場合でもコンテンツの再現をすることができる。

前述の国立国会図書館の報告によれば、市販のエミュレーション用アプリケーションの性能に依存する部分が多いが、再生が困難なコンテンツも多数存在し、現状、長期保存向けには適していないとしている。長期保存の観点からは、

- ハードウェア環境やOS環境に依存しないで旧式ハードウェア環境を再現できること。
- 各ゲストOSに対応した各種デバイスドライバが入手可能で、旧式ハードウェア環境を再現できること。
- エミュレータがバージョンアップされても、旧バージョンがサポートしていた旧式OS環境が新バージョンにおいてもサポートされ続けること。

をエミュレーション用アプリケーションに求める要件として挙げている。報告書では、エミュレータ上で再生用アプリケーションのインストール可否、起動や動作可否、ファイルフォーマット別の再生可否をまとめている。インストール、起動、ボタン押下等のオペレーション、静止画像の再生等はエミュレータ上で行えるが、音声や動画の再生、計算、情報検索、データの記録等はエミュレータで再現できない場合が多い。

3.4 OAIS参照モデルと保存メタデータ

アーカイブデータの長期保存は、前節のようなソフトウェアを用いた解決のアプローチに加え、どのようなコンテンツをどのように保存すべきなのか、保存のためのメタデータ検討やアーカイブシステムの機能検討、更にはアーカイブの戦略や他システムとの連携等、総合的な視野が必要となる。

アーカイブデータの長期保存を総合的に検討するモデルにOAIS参照モデル(Reference Model for an Open Archival Information System, ISO 14721:2003)がある。

図5にOAIS参照モデルに基づく機能要素を示す。モデルでは保存すべき情報と関連するメタデータを合わせて情報パッケージ(IP:Information Package)として取り扱う。IPは処理に応じて3つの形態を持つ。即ち、生産者(Producer)が提出するSIP(Submission In-

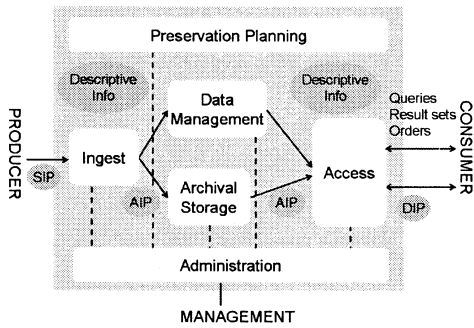


図 5 OAIS : 機能要素

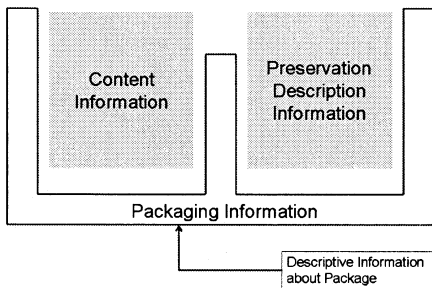


図 6 情報パッケージの構成

formation Package), システムが管理する AIP (Archival Information Package), 利用者に配布される DIP (Dissemination Information Package) である。IP は形態に応じてそれぞれ格納されるメタデータに違いがあるが、基本的には図 6 に示される構成を持つ。

内容情報 (Content Information) は保存対象となるデータオブジェクトと表現情報を持つ。データオブジェクトはコンテンツそのものを指し、表現情報はデータオブジェクトを人間が理解できる形態で提示するための情報を指す。具体的には、データオブジェクトが持つフォーマットや文字コード等が含まれる。表現情報の内容はどこまで含めるが議論を要するが、フォーマットの仕様や文字コード全てを含めると膨大な量となるため、知識ベースを導入し、適宜表現情報を参照できる表現ネットワークを形成すべきであるとされている。

保存記述情報 (Preservation Description Information) は以下の 4 カテゴリが管理される。

- 来歴 (Provenance) : コンテンツの出处や保存に関わる処理の履歴
- コンテキスト (Context) : 情報パッケージ外のオブジェクトとの関係
- 参照 (Reference) : コンテンツを一意に識別するための識別子
- 不変性 (Fixity) : コンテンツを保護するための情報

パッケージ化情報 (Packaging Information) は内容情報と保存記述情報を結びつける情報を持つ。

このように、OAIS 参照モデルでは、アーカイブシステムの機能を検討する上で情報パッケージという形態で保存のためのメタデータ概念が定義されている。いくつかの機関では、モデルに基づくメタデータの策定も始められている。

電子図書館連盟 (Digital Library Federation) では、XML 形式のメタデータ記録方式 METS (Metadata Encoding & Transmission Standard) を策定している。METS は以下に示す 7 セッションからなる標準で、このうち管理的メタデータでは、OAIS 参照モデルの表現情報や保存記述情報の一部を記述することができる。

- METS ヘッド : ヘッド情報
- 記述的メタデータ : リソースに関する記述/ポインタ
- 管理的メタデータ : リソースの来歴, 知的財産権に関する記述
- ファイルセクション : デジタルオブジェクトを構成する全ファイル情報
- 構造マップ : デジタルオブジェクト構造
- 構造リンク : 構造マップに示されたノード間のハイパーリンク
- 動作 : デジタルオブジェクトの動作に関する記述

この他、OAIS 参照モデルを参考にした具体的なメタデータ策定の例として、OCLC (Online Computer Library Center) による保存メタデータの枠組みである PREMIS (Preservation Metadata: Implementation Strategies) がある。

3.5 考察

本章では、アーカイブデータの長期保存に関する問題や様々な解決の取り組みを紹介した。デジタルアーカイブはコンテンツの作成や蓄積の

表 1 電子的にデータベース化された資料の割合

館種	館数	ほぼ 100%	約 75%	約 50%	約 25%	少し	無回答
全体	720 館	40.7%	17.4%	12.9%	11.5%	15.6%	1.9%
総合	53	22.6	9.4	18.9	24.5	22.6	1.9
郷土	74	23.0	16.2	14.9	16.2	27.0	2.7
美術	188	61.7	16.5	9.0	3.7	7.4	1.6
歴史	286	32.9	21.3	12.6	13.3	18.5	1.4
自然史	35	11.4	14.3	22.9	20.0	25.7	5.7
理工	30	50.0	13.3	16.7	13.3	3.3	3.3
動物園	16	68.8	6.3	18.8	0.0	6.3	0.0
水族館	24	66.7	12.5	8.3	4.2	4.2	4.2
植物園	7	28.6	42.9	0.0	14.3	14.3	0.0
動植水	7	85.7	0.0	14.3	0.0	0.0	0.0

みならず、利用されながらも長期的にデータの保存が可能なシステムの構築を行わなければならない。そこで、本節では、これまでの取り組みを踏まえ、デジタルアーカイブシステムにおける長期保存の課題について考察する。

アーカイブデータを長期保存するには以下の3点が重要となる。即ち、(1) 真正性、(2) 継続性、(3) 利用可能性である。特に文化資産の場合、歴史研究や新たな発見において、あるいは復元修復作業において、コンテンツが新しく書き換えられる可能性がある。また、コンテンツは容易に改竄が可能である。コンテンツの書き換えや改竄防止に対処しつつ、情報が正しく再現されていることが重要である。近年では、電子署名の技術が発達してきている。電子署名は認証局を必要とするが、認証局が長期的に存在しなければコンテンツの真正性が保証されないのではその利用は困難となる。認証局を必要としない電子署名技術の確立も併せて検討する必要がある。

後世に真正性の保証されたアーカイブデータを残して行くためには、継続的にデータを作成し、常に利用可能な状態にしておく必要がある。表 1 に博物館における目録台帳の電子化の現状を示す*1。多くの電子化作業が残されている現状を示しているが、日常業務において作業者がインセンティブを感じられ、かつアーカイブに転用可能なデータ作成手法、データ構造の定義を検討することが重要である。

4. ま と め

本稿では、デジタルアーカイブを構成する様々な要素技術を概観し、特に文化資産向けアー

カイブシステムにおけるデータの長期保存に関する技術的課題について考察を行った。ここでは、真正性、継続性、利用可能性を考慮したデータ構造や作成手法を確立することが重要であることがわかった。今後は、より具体的なデータ構造の定義やデータ作成フローの検討を行う予定である。

参 考 文 献

- 1) 後藤: “デジタル・アーカイブ技法と実践 (1) デジタル・アーカイブズと視聴覚教育”, 視聴覚教育, Vol.60, No.2(2006).
- 2) 文化遺産オンライン
<http://bunka.nii.ac.jp/>(参照 2008-08-30)
- 3) 神内ほか: “デジタルイメージシステムの開発とその応用”, 日立論評, Vol. 79, No. 7, pp.27-34(1997).
- 4) 加茂: “デジタルアーカイブと VR 表現”, 情報処理, Vol.43, No.10, pp.1052-1057(2002).
- 5) LOUVRE-DNP MUSEUM LAB
<http://museumlabor.jp/>(参照 2008-08-30)
- 6) 研究資源共有化シンポジウム講演予稿集, 2008-03-14, 人間文化研究機構.
- 7) 鶴岡: “慶應義塾大学メディアセンターでの音源デジタル化の事例と問題点”, 音のデジタル化ワークショップ, 2008-06-20.
- 8) 松山: “3 次元ビデオ”, 日本印刷学会, Vol. 42, No.1, pp.35-41(2005).
- 9) 秋元ほか: “文化財メタデータの現状と課題”, 画像ラボ, Vol.19, No.4, pp.30-36(2008).
- 10) 秋元ほか: “博物館情報の相互利用を目的とした文化財情報システムの提案”, じんもんこん 2004, 2004-12-09,10.
- 11) 江草ほか: “SRU/SRW を用いた教育図書館資料の書誌検索システムの構築”, 情報知識学会誌, Vol. 17, No.2, pp.69-74(2007).
- 12) 電子情報の長期的保存とアクセス手段の確保のための調査報告書, 2004-03, 国立国会図書館.

*1 平成 16 年度実施の博物館総合調査アンケート